

# Periodismo de datos

Autores: Agustí Cerrillo Martínez, Laura Aznar, Laura Aragó, Alfonso Sánchez Uzábal y Pablo Rey Mazón  
Coordinadora: Núria Ferran Ferrer

## 1. Introducción: el periodismo, el periodismo de investigación y el periodismo de datos

- 1.1. Fiscalizar el poder
  - 1.1.1. Introducción
  - 1.1.2. Independencia
  - 1.1.3. Vigilar el poder
- 1.2. El periodismo de investigación
- 1.3. El periodismo de datos
  - 1.3.1. Presentar los datos al público

## 2. Aspectos jurídicos del periodismo de datos

- Introducción
- 2.1. El gobierno abierto
  - 2.1.1. Introducción
  - 2.1.2. La transparencia pública
  - 2.1.3. La participación y la colaboración ciudadana
- 2.2. Libertad de información y de expresión
  - 2.2.1. Introducción
  - 2.2.2. La libertad de información
  - 2.2.3. La libertad de expresión
  - 2.2.4. Límites a la libertad de información y de expresión
- 2.3. El ejercicio de la profesión periodística y la autorregulación deontológica

Bibliografía

## 3. Fuentes de información y mecanismos de obtención de datos

- 3.1. Las fuentes periodísticas
  - 3.1.1. Introducción
  - 3.1.2. Tipología de las fuentes de información
- 3.2. Introducción: los datos como fuente de información
- 3.3. Bases de datos disponibles
  - 3.3.1. Introducción
  - 3.3.2. El movimiento *Open Data*
  - 3.3.3. Recopilación de fuentes
- 3.4. La Ley de transparencia
  - 3.4.1. Introducción
  - 3.4.2. La Ley de transparencia en España y en Cataluña

3.4.3. Cómo formular una petición de información

3.5. La minería de datos

3.5.1. La minería

3.5.2. Herramientas

#### **4. Análisis de datos**

4.1. Herramientas para el análisis de datos

4.1.1. Gestores de hojas de cálculo

4.1.2. Herramientas avanzadas

4.2. La importancia de ordenar bien una hoja de cálculo

4.2.1. La estructura de los datos

4.2.2. Limpiar una hoja de cálculo

4.2.3. Un ejemplo práctico

4.3. Herramientas y mecanismos de análisis

4.3.1. Importar datos

4.3.2. Filtrar y ordenar

4.3.3. Las tablas dinámicas

4.3.4. Filtros condicionales y otras opciones avanzadas

4.3.5. Utilizar descriptores estadísticos típicos

4.4. El análisis visual

4.4.1. Introducción

4.4.2. Un ejemplo práctico con Microsoft Excel

4.5. Cómo encontrar noticias a partir del análisis

4.5.1. Introducción

4.5.2. El análisis de datos como elemento conductor de la noticia: el *Data Driven Journalism*

4.5.3. Los datos como apoyo de la noticia

4.5.4. Los datos como punto de partida de una investigación

4.5.5. La visualización como centro de la noticia

#### **5. Redacción y visualización**

5.1. La construcción de la noticia: la teoría de las 6 W, la pirámide invertida y las entradillas

5.2. Tipología de piezas periodísticas: los géneros periodísticos

5.2.1. Introducción

5.2.2. Género informativo

5.2.3. Género interpretativo

5.2.4. Los géneros de opinión

#### **6. Tres ejercicios de periodismo de datos**

Introducción

6.1. Twitter: captura de datos y análisis de redes

6.1.1. Introducción

6.1.2. Contexto

6.1.3. Conceptos básicos o previos

6.1.4. Preguntas a resolver

6.1.5. Información complementaria

6.2. Airbnb: análisis geoespacial

6.2.1. Introducción

6.2.2. Contexto

6.2.3. Conceptos básicos o previos

6.2.4. Preguntas a resolver

6.2.5. Información complementaria

6.3. Educación pública-privada: abriendo datos abiertos

6.3.1. Introducción

6.3.2. Contexto

6.3.3. Conceptos básicos o previos

6.3.4. Preguntas a resolver

6.3.5. Información complementaria

**Créditos**

---

# 1. El periodismo, el periodismo de investigación y el periodismo de datos

## 1.1. Fiscalizar el poder

### 1.1.1. Introducción



«El periodismo es aquello que alguien no quiere que se sepa, el resto son relaciones públicas.»

George Orwell

El periodismo consiste en proporcionar al ciudadano la información que necesita para ser libre y capaz de gobernarse a sí mismo. Para cumplir este objetivo, debe ser fiel a los nueve elementos que estipulaban Bill Kovach y Tom Rosenstiel:

1. La primera obligación del periodismo es la verdad.
2. Debe lealtad a los ciudadanos.
3. Toda información se tiene que verificar.
4. Debe mantener su independencia respecto a aquellos de los que informa.
5. Tiene que ejercer un control independiente del poder.
6. Debe ofrecer espacio al público para la crítica y el comentario. Interacción y *feedback*.
7. Tiene que esforzarse para que lo que explica sea sugerente y relevante.
8. Las informaciones deben ser exhaustivas y proporcionadas.
9. Tiene que respetar la conciencia individual de sus profesionales (la cláusula de objeción de conciencia).

Ser imparcial o neutral no es un principio esencial del periodismo, pero sí lo es el hecho de que el periodista cumpla con su compromiso con los ciudadanos y proporcione un foro público donde impulsar el debate. En este apartado, nos centraremos en el cuarto y quinto elemento del listado de Kovach y Rosenstiel.

# 1. El periodismo, el periodismo de investigación y el periodismo de datos

## 1.1. Fiscalizar el poder

### 1.1.2. Independencia

«El proceso del periodismo tiene sus limitaciones. La escasez de tiempo y la frecuente falta de disponibilidad de la información son dos dificultades endémicas de la profesión, como también lo son los errores que cometen los periodistas cuando trabajan sometidos a presiones. Sin embargo, el buen periodismo también choca con limitaciones creadas por los propios periodistas, así como por los propietarios y los que controlan los medios.»

David Randall

Todos los propietarios de los medios de comunicación hacen bandera de los conceptos de verdad, honestidad y transparencia, pero en general utilizan las plataformas informativas para hacer negocio o como canal de propaganda. Esta lógica se hace todavía más visible en contextos de convulsión política. La promoción de las propias ideas y la exclusión de las ideas adversas, la información sesgada e intencional, adaptada a un punto de vista o unos intereses empresariales determinados, son algunas de las malas praxis que impregnan la historia del periodismo.

En la actualidad, es habitual que los medios sean propiedad de grandes empresas. Sus requerimientos propagandísticos acostumbran a limitarse a apoyar a un determinado partido político, mimando a los políticos que puedan favorecerlos o atacando a los que no. También realizando historias cómplices o encubridoras de sus intereses corporativos. Su prioridad es maximizar el margen de beneficio y hacerlo muy rápido si su deuda está siendo financiada por agentes externos. Las prioridades de los propietarios probablemente han impuesto más limitaciones a la manera en que se practica periodismo que cualquier otro factor en los últimos años.

En el conjunto de Europa, existe un problema en este sentido. Según un informe publicado en el 2017 por el Centro por el Pluralismo y la Libertad en los Medios, del Instituto Universitario Europeo, las principales amenazas al pluralismo mediático son la excesiva concentración de los medios en pocas manos, la poca diversidad de puntos de vista en las noticias, la falta de transparencia sobre los intereses empresariales de los medios y una independencia editorial vulnerable, sujeta a influencias políticas.

En el caso español, si bien existe un riesgo medio en cuanto a la falta de pluralismo, uno de los resultados más preocupantes es la concentración de la propiedad de los medios españoles, que presenta un riesgo del 71 %. También es grave el riesgo relativo a la independencia política (54 %). Concretamente, la independencia de los medios públicos en cuanto a la gobernanza y a la financiación tiene un riesgo del 83 %, lo que quiere decir que la injerencia política en los medios españoles es habitual. Además, el estudio también indica que la influencia del partido en el Gobierno sobre la dirección de las redacciones se ha multiplicado en los últimos años. La autonomía editorial y los intentos de controlar políticamente los medios muestran señales de alarma.

# 1. El periodismo, el periodismo de investigación y el periodismo de datos

## 1.1. Fiscalizar el poder

### 1.1.3. Vigilar el poder

Esta función se ve amenazada por la acumulación empresarial, que puede destruir la independencia que los medios necesitan para ejercer su papel de vigilantes. El principio de vigilancia y control va más allá de la fiscalización a un Gobierno, y se extiende a todas las instituciones poderosas de la sociedad. El periodismo tiene que vigilar a los escasos poderosos de una sociedad, en representación de los muchos que no lo son, para así evitar tiranías. La finalidad de la función de vigilancia no es solo conseguir que la acción del poder sea transparente, sino dar a conocer las consecuencias de la acción de este poder y conseguir que sean comprensibles. Esto implica que los medios sepan reconocer dónde actúa el poder de manera eficaz y dónde no. Esta vocación del periodismo es el origen del periodismo de investigación.

En el periodismo de investigación original, son los propios reporteros los que buscan, investigan y documentan una información que los usuarios desconocen. Con frecuencia, este tipo de reportajes ocasionan una investigación pública y oficial de los sujetos o actividades que protagonizan la información, y constituyen el ejemplo clásico de la tarea de presión de los medios sobre las autoridades, en beneficio del ciudadano.

El periodismo de investigación interpretativo es el resultado de una cuidadosa reflexión y análisis de una idea y de una metódica búsqueda de datos que dan pie a un contexto informativo nuevo y más completo, capaz de ofrecer al ciudadano una mejor comprensión de los hechos. En general, afecta a temas o sucesos más complejos que los que dan lugar al reportaje clásico, porque no solo destapa una información nueva, sino que también ofrece una nueva versión de esta información.

La tercera modalidad de periodismo de investigación es la que se dedica a investigaciones que otras instancias ya han puesto en marcha. En este caso, el reportaje parte de un descubrimiento o una filtración referentes a una investigación oficial que otros, normalmente organismos gubernamentales, ya han empezado. El riesgo de este método de recopilar información es que depende en gran medida del rigor y escepticismo del reportero en cuestión. Garantiza a los sujetos entrevistados, que en general son fuentes oficiales, un valioso foro en el que pueden propagar acusaciones sin que recaiga en ellos ninguna responsabilidad pública, y esta, precisamente, es una de las debilidades de este tipo de reportajes. Normalmente, los reporteros implicados conocen solo una parte de la investigación, de forma que su trabajo, más que vigilar las instituciones poderosas, se puede convertir en una herramienta en sus manos.

Todos los reportajes requieren cierta actividad investigadora, pero lo que conocemos como periodismo de investigación tiene, además, cierta dimensión moral. Compromete al lector, lo anima a emitir un juicio sobre el asunto que denuncia y da por supuesto que el medio informativo que lo publica lo considera importante, merecedor de un esfuerzo especial. Uno de los problemas que surgen con este modelo de periodismo es que el medio informativo se implica en una información sobre la base de que ha habido algún acto de mala fe. Por eso ha sido denominado como «periodismo fiscalizador». Dado que puede alterar el curso de los acontecimientos o puede afectar negativamente a la reputación de una persona, una empresa o una institución, la tarea del reportero implica una gran responsabilidad, no solo a la hora de verificar los hechos, sino también de compartir datos sobre la naturaleza de las fuentes informativas.

La revolución de las nuevas tecnologías y la nueva organización económica que ha generado crean nuevas oportunidades, pero también amenazan a una prensa vigilante e independiente. Los medios digitales permiten que la información viaje con mayor facilidad y rapidez, pero conducen a la creación de grupos de comunicación internacionales que ponen en peligro el propio concepto de Estado-nación. La experiencia nos dice que es muy posible que con el tiempo, a medida que los directivos de los informativos formados en este nuevo entorno transnacional vayan logrando cuotas de mayor responsabilidad, a los medios de comunicación les parezca que ya no merece la pena ejercer la función de vigilancia y control sobre su propia empresa matriz. La teoría de la prensa libre, que decía que era necesaria una voz independiente y capaz que se ocupara de vigilar la influencia de las instituciones más poderosas de la sociedad, está en peligro.

Sin embargo, más importante que el control efectivo que pueden ejercer los vigilantes de los medios de comunicación es el hecho de que las nuevas tecnologías proporcionan herramientas que ofrecen la posibilidad a casi todos los ciudadanos de vigilar la prensa consolidada. Las instituciones dedicadas al periodismo independiente indican de qué manera las nuevas tecnologías podrían reorganizar los mecanismos de producción y difusión de noticias. Potencialmente, esto supone un desafío a los canales de información tradicionales y da pie a pensar que si los medios convencionales abandonan su papel de vigilancia, otros podrían asumirlo. Más allá de la visión utópica que puede ofrecer la tecnología, hay cuestiones económicas más prácticas que no encuentran respuesta todavía. ¿Puede alguna de las nuevas plataformas o de los nuevos medios mal llamados alternativos o independientes suscitar el interés de una parte significativa de la opinión pública? Y, en caso afirmativo, ¿acabarán siendo absorbidos por las grandes corporaciones y caerán, por lo tanto, en la cultura empresarial de la que sus fundadores querían escapar?

Los departamentos de informativos de los grandes medios ofrecen el modelo más claro con el que contamos de la nueva organización económica de la información. Fueron las primeras instituciones periodísticas en verse sometidas a las grandes

corporaciones con intereses fuera del periodismo. Aun así, a medida que empezaban a sufrir una pérdida estructural de audiencia a favor de las nuevas tecnologías, fueron abandonando su función de vigilancia e inclinándose por un modelo informativo más cercano al entretenimiento. Este modelo, que es la única evidencia que tenemos de los efectos de la asimilación global de los medios informativos, despierta serias dudas de que el periodismo de investigación se mantenga en el nivel que logró en la segunda mitad del siglo xx.

La función de vigilancia y control es como cualquier otra. Requiere de facultades especiales, un temperamento especial. Además, exige un compromiso serio con las fuentes, el deseo de ocuparse de asuntos de importancia y una prensa independiente de cualquier interés que no sea el del consumidor final de la noticia. El principio de vigilancia y control se enfrenta hoy a más retos que nunca pero, al mismo tiempo, la naturaleza expansiva del periodismo como foro público ha provocado una nueva ola de periodismo interpretativo y «opinativo» que hace más crítica la necesidad de un periodismo vigilante más serio y activo. En el siglo xxi, los medios tienen que ejercer vigilancia no solo sobre el Gobierno, sino también, y especialmente, sobre el mundo empresarial y el ámbito cada vez mayor de actividades sin ánimo de lucro.

# 1. El periodismo, el periodismo de investigación y el periodismo de datos

## 1.2. El periodismo de investigación

«La razón más habitual del fracaso de una historia no es una mala redacción, las citas manipuladas o una pobre elaboración, sino una investigación inadecuada.»

David Randall

El periodismo de investigación tiene algunos rasgos diferenciales:

### 1) Originalidad de la investigación

El periodismo consiste en hacer investigaciones originales, muchas veces usando materiales en bruto. Se puede llevar a cabo mediante una amplia serie de entrevistas o bien por medio de cifras. Muchos de los reportajes son consecuencia de un trabajo meticuloso del periodista, analizando la conexión entre datos que no se habían investigado previamente.

### 2) Informaciones relacionadas con negligencias o irregularidades de las que no se tienen pruebas

Hay que dedicarle tiempo si se trata de una cuestión de interés para los lectores, asumiendo que quizá la investigación no es fructífera y que «no hay tema». El periodismo de investigación empieza con la sospecha de que en un tema determinado están las señales de una buena historia. Lo primero que hay que hacer es poner en la balanza si la historia merece el esfuerzo y el tiempo requeridos para echarla adelante. Esta clase de piezas se pueden hacer sobre prácticamente cualquier ámbito de la vida pública, pero hay dos categorías que resultan especialmente fecundas: las actividades que están alejadas del escrutinio del público y las personas o entidades con relevancia pública.

Las empresas y entidades financieras, que generalmente han sido poco fiscalizadas, son un terreno muy fértil para la indagación periodística.

«Si ponemos la lupa sobre un plan de inversiones reciente lanzado con un gran despliegue publicitario, podemos apostar el sueldo a que sacaremos a la luz alguna información confidencial.»

Lo mismo pasa con algunos organismos, instituciones y empresas que desarrollan una actividad que ha sido muy poco controlada por parte de los medios de comunicación.

## Consortio de Turismo de Barcelona

Un ejemplo de esto es el Consorcio de Turismo de Barcelona, una entidad público-privada cuya financiación había pasado muy desapercibida. Hace un par de años, se situó de pleno bajo los focos mediáticos a raíz de una acción reivindicativa que convocó la Asamblea de Barrios por un Turismo Sostenible, que precisamente denunciaba la opacidad de este consorcio. Esto hizo que algunos medios empezaran a investigar qué era exactamente esta entidad, quién la controlaba y cómo se financiaba. Otro caso similar ha sido el de la empresa pública Barcelona Regional, constituida en 1993 a partir de la experiencia acumulada con la transformación olímpica. Durante años había operado libremente, sin ser fiscalizada, pero ahora está en el ojo del huracán a partir de que una denuncia anónima presentada a la fiscalía provincial revelase que los gestores de la agencia habían malversado sus fondos en el periodo de mandato convergente. A raíz de esta acusación, los medios iniciaron una investigación paralela a la de los tribunales, el Ayuntamiento y el propio Consejo de Administración de Barcelona Regional. A partir de aquí, han empezado a salir a la luz contratos irregulares, subvenciones y vínculos injustificados con exrepresentantes públicos.

Hay algunos **factores que pueden simplificar la tarea y mejorar la eficacia** del reportero:

### a) Conocer la legislación sobre el acceso público a la información

En Cataluña y el Estado español, la Ley de transparencia (una herramienta que todavía se utiliza poco en la mayoría de los medios) permite que los periodistas, y en realidad cualquier ciudadano, puedan solicitar información pública a la Administración. Hay

muchos datos públicos (que no publicados) que están al alcance de cualquier persona, siempre y cuando conozca los procedimientos para requerir la información.

Las solicitudes no siempre son fructíferas, en el sentido de que hay organismos opacos que no quieren ofrecer información sobre su funcionamiento, sueldos de directivos, presupuestos, ingresos, etc., y que procurarán bloquearnos el acceso. En estos casos, en Cataluña, los periodistas podemos jugar la carta de la [GAIP \(\\*\)](#).

Si una institución nos deniega una información acogiéndose, por ejemplo, a la Ley de protección de datos, siempre podemos hacer una reclamación a esta comisión, justificando que, a nuestro entender, el derecho a la información pasa por encima de la negativa de la institución en cuestión. Es un recurso que los medios empiezan a utilizar.

## Ejemplo

Por poner un ejemplo, *La Directa* y *Crític* se acogieron al derecho de acceso a la información pública para saber el dinero que había gastado el Departamento de Interior en armamento, munición y otro equipamiento para los Mossos d'Esquadra. La Dirección General de la Policía denegó inicialmente esta información, pero los dos medios recurrieron esta decisión a la GAIP, que les dio la razón y forzó al Departamento a facilitar esta información.

Algunas investigaciones han surgido de documentos entregados a los periodistas, pero muchas más se han puesto en marcha cuando un reportero ha descubierto la existencia de registros y documentos a cuya consulta tenía acceso. En general, los organismos burocráticos no estarán deseando dar publicidad a la existencia de determinadas fuentes de información e incluso, en algunos casos, intentarán impedir que la gente se tome la molestia de consultarlas. Es importante que los periodistas descubran los documentos oficiales que existen, los examinen y los utilicen en sus investigaciones. También es recomendable hacer uso de las fuentes de referencia estandarizadas: los listados de publicaciones oficiales, los informes legislativos, las listas de organismos públicos, los registros mercantiles o los registros de las organizaciones que reciben financiación pública, etc.

Una buena lista de contactos es trascendental a la hora de hacer periodismo de investigación. Son útiles para cubrir una amplia gama de temas, orientar nuestras informaciones y facilitarnos la información. Diversificarlos: tener fuentes solo del ámbito institucional hará que siempre ofrezcamos la versión oficial de los hechos. En este sentido, es imprescindible abrir el foco: ¿qué tienen que decir los movimientos sociales? ¿Las plataformas ciudadanas? ¿El movimiento vecinal?

Hay un dicho histórico según el cual los periodistas somos expertos en todo y en nada. No obstante, lo cierto es que no somos hombres ni mujeres orquesta, y nuestros conocimientos son limitados. A la hora de hacer algunos reportajes, nos podemos encontrar fácilmente con el hecho de que la temática que hemos seleccionado se escapa ampliamente de nuestra zona de confort. Siempre está bien tener formación en una especialidad concreta a la hora de cubrir temas de economía, sociedad o cuestiones jurídicas, pero no podemos saber de todo. Por eso, se hace especialmente importante tener al alcance fuentes especializadas, que no tengan intereses específicos en el tema al que damos cobertura, que nos ayuden a entender y procesar mejor la información. En este sentido, es especialmente útil disponer del teléfono de sociólogos, juristas, politólogos o economistas «independientes», asumiendo que siempre tendrán su visión subjetiva de los datos, pero que no tengan una relación directa con el objeto de investigación. Si queremos hablar del juicio contra una farmacéutica, ¿no pediremos al abogado de la empresa en cuestión que nos analice su validez!

### b) Conocimientos de informática (y más)

No solo (pero también) referida a la habilidad de rastrear en línea con eficacia, sino también a la capacidad de recurrir al software de bases de datos. La informática puede modificar la capacidad del periodismo de investigación, porque tiene la capacidad de ir más allá de las tradicionales anécdotas y entrevistas y aglutinar una cantidad muy amplia de pruebas documentales. [Ampliación al apartado «Periodismo de datos».]

¿Cómo podemos obtener buen material de cara a una investigación?

En primer lugar, saber qué tenemos que buscar: más allá de responder las 6 W, nos hacen falta detalles, anécdotas y contexto. Esta información adicional es lo que marca la diferencia entre una historia ordinaria y una buena historia.

- Recolección de detalles: es importante construir una cronología exhaustiva de lo que ha pasado. No hay detalle tan insignificante que no valga la pena recoger, puesto que incluso los fragmentos diminutos pueden añadir peso a la historia muy por debajo de su calor aparente. En cierto modo, el detalle es la historia en muchos casos.
- Anécdotas: incidentes o episodios que puedan resumirse en una o dos frases.

- El telón de fondo: el entorno, el contexto y las partes relevantes del tema.
- Perspectiva: el contexto a veces puede ser la parte vital de la historia, porque pone los hechos o acontecimientos en una perspectiva apropiada.
- Recelar de las fuentes y de las páginas web: se tiene que desarrollar un radar capaz de detectar los lugares que no son de confianza, y plantearnos las mismas preguntas que nos haríamos con una fuente (¿quién está detrás?, ¿por qué ha publicado esta información?, ¿qué incluye?, ¿qué omite?).
- Ser específicos con la búsqueda: métodos de búsqueda en línea como poner comillas en el concepto que se busca, poner los ítems AND u OR. Guías para la búsqueda en línea: [searchenginewatch.com](http://searchenginewatch.com) o [researchbuzz.com](http://researchbuzz.com).
- Usar la aplicación de búsqueda avanzada, que permite afinar la búsqueda con parámetros o fechas.
- Suscribirse a archivos en línea: por ejemplo, [Google Alerts](https://www.google.com/alerts).
- Hacer uso de las hemerotecas.

## ¿Cómo desarrollar una investigación?

«Los temas de las investigaciones llegan a los diarios por las vías más diversas: a partir de las informaciones de los contactos, por puro azar, a partir de una noticia en principio rutinaria y cuya magnitud amplían informaciones posteriores, gracias a las observaciones de los reporteros, a partir de informaciones corrientes que van complicándose o en las que se descubren interrogantes cada vez más interesantes cuando empezamos a indagar.»

Algunos consejos:

- Conseguir y archivar todos los documentos pertinentes: aunque no los usemos en el momento de recibirlos, hay que guardar una documentación tan ingente como sea posible y no desprenderse de ella, porque nunca se sabe cuándo podrá ser útil. Meses después de que haya llegado a nuestras manos un informe de apariencia inocua, puede pasar algo que le otorgue una relevancia inesperada.
- Transcribir todas las entrevistas y archivarlas: esto es particularmente importante cuando el proceso de investigación es largo o cuando colaboran varios reporteros. Merece la pena intercambiar notas para comprobar si se ha pasado por alto algún dato de importancia.
- Ser persistentes: cuando se lee un relato sobre cualquier investigación, lo que más llama la atención es la persistencia de los reporteros.
- Cultivar el trato con las fuentes mejor informadas: si una fuente establece ciertas reglas, hay que cumplirlas, pero esto no quiere decir que se tenga que dar crédito a todas sus palabras. Se tienen que respetar los acuerdos.
- Las informaciones que finalmente lleguen a publicarse no han de tener ningún punto débil: publicar un artículo por el único motivo de que se le ha dedicado mucho tiempo es la mejor forma de abocarse al desastre. Los directivos y reporteros tendrán que adoptar en una fase avanzada de la investigación la decisión de ir la publicando por entregas a medida que se lleva a cabo o, por el contrario, reservar los hallazgos para dar un gran golpe de efecto al final del proceso. En este segundo caso, es necesario establecer una fecha tope. Por otro lado, publicar conclusiones provisionales puede incitar a las fuentes a proporcionar más información o más datos de vital importancia.
- Actuar clandestinamente: en algunas ocasiones excepcionales, la única manera de conseguir la información es de forma clandestina, por ejemplo, cuando estamos investigando un delito. Esta forma de trabajar implica siempre un cierto grado de engaño, así que el delito en cuestión tiene que ser suficientemente serio como para justificar esta falta de honradez. En segundo lugar, los riesgos físicos en tal situación pueden ser inmensos y perdurar incluso tiempo después de haber publicado nuestra historia. En tercer lugar, si estamos investigando actividades criminales bajo una identidad encubierta, podemos vernos obligados a participar en ellas, lo que hace que nuestros actos resulten todavía más difíciles, si no imposibles de defender. A veces, los periodistas han usado falsas identidades con el propósito fundamental de dejar a cuerpo descubierto el trato o maltrato recibido por varias víctimas de la sociedad, como los sintecho, los enfermos psíquicos o los drogadictos. Esta táctica requiere dotes interpretativas y, a veces, incluso un disfraz.

# 1. El periodismo, el periodismo de investigación y el periodismo de datos

## 1.3. El periodismo de datos

### 1.3.1. Presentar los datos al público

#### 1.3.1.1. Visualizar los datos o no

Hay ocasiones en las cuales los datos pueden explicar mejor una historia que las palabras o las fotografías. Por eso, términos como «aplicación de noticias» y «visualización de datos» han adquirido el estatus de palabras clave en tantas redacciones en los últimos tiempos. También promueve el interés la gran colección de nuevas herramientas y tecnologías (a menudo, gratuitas) destinadas a ayudar incluso al periodista menos dotado técnicamente a convertir datos en una presentación visual de la historia.

Herramientas como Google Fusion Tables, Many Eyes, Tableau, Dipty y otras hacen más fácil que nunca crear mapas, cuadros, gráficos e incluso aplicaciones con datos que hasta ahora eran dominio de especialistas. La cuestión ahora para los periodistas no es tanto si pueden convertir sus conjuntos de datos en una visualización, sino si les conviene hacerlo. Una mala visualización de los datos es peor en muchos sentidos que ninguna visualización.

##### 1) Utilización de los gráficos con movimiento

Con un guion ajustado, animaciones muy cronometradas y explicaciones claras, los gráficos con movimiento pueden dar vida a números o ideas complejas, y orientar a su público. Los recursos visuales explicativos con la voz en *off* ofrecen una manera poderosa y memorable de explicar una historia.

En la mayoría de las ocasiones, nuestro flujo de trabajo empieza con Excel. Es una manera fácil de descubrir si hay algo interesante en los datos. Si es así, tenemos que buscar la manera de visualizarlos y después escribiremos el texto que los acompañe. El texto tiene que servir para analizar y aportar contexto a los datos, que no se quede con una simple visualización de la relación entre distintas variables: lo que no nos diga la infografía, lo tenemos que explicar nosotros.

##### 2) Publicar los datos

Un consejo a la hora de publicar los datos es utilizar una visualización que permita la descarga fácil del conjunto de datos. Los lectores pueden explorarlos tras las historias, interactuando con las visualizaciones o usando los datos de otras maneras. Esto es importante porque aumenta la transparencia y permite el *feedback* del lector.

##### 3) Dar acceso a los datos

De este modo, los lectores pueden asegurarse de que no manipulamos los datos para llegar a conclusiones forzadas. Aparte, permitir el acceso a los datos hace que haya investigadores que puedan reutilizarlos.

##### 4) Humanizar los datos

Los datos están asociados a la vida de gente real, no son números disociados flotando en el vacío, sino mediciones de cosas tangibles. Cuando los abordamos, tenemos que considerar los sistemas del mundo real de donde provienen. Los periodistas tienen un rol muy importante de sacar a la luz esta humanidad inherente a los datos. Al hacerlo, tienen el poder de cambiar la comprensión del público tanto de los datos como de los sistemas de los que emergieron los números.

#### 1.3.1.2. La visualización en la investigación periodística

En la fase de buscar información, las visualizaciones pueden:

- Ayudarnos a identificar temas.
- Excluir lo que no nos interesa.
- Identificar errores en los datos.
- Encontrar ejemplos típicos.

En la fase de edición, podemos:

- Ilustrar un argumento de una historia de una manera más convincente.

- Sacar información técnica innecesaria de la prosa.
- En particular, cuando son interactivos y permiten la exploración, ofrecen transparencia respecto al proceso de información.

Consejos sobre el uso de visualizaciones:

- Usar pequeños múltiplos para orientarse rápidamente en un conjunto de datos grande. Los mapas con pequeños múltiplos pueden ser una manera de mostrar tiempo y lugar en una imagen estática que es fácil de comparar, a veces más fácil que la versión interactiva.
- Mirar los datos, remirarlos y volverlos a mirar.
- Utilizar cifras crudas, porcentajes e índices.
- Pensar cómo combinar los datos y si nos interesa hacerlo. Por ejemplo, si hablamos de criminalidad, podemos hacer una infografía estableciendo diferentes parámetros y los podemos visualizar todos a la vez o uno a uno, usando el mismo patrón gráfico, pero refrescando las variables cada vez que hacemos clic en un ítem diferente.

### **1) No dar nada por supuesto**

Cuando analizamos los datos, nos podemos encontrar con registros que no parecen correctos, o bien porque no entendemos qué significan o bien por errores informáticos, de transcripción o del tipo que sea. Si queremos publicar algo basado en estos datos, lo primero que hay que hacer es resolver las cuestiones que no nos quedan claras. Si nosotros no somos capaces de entender lo que explicamos, llevaremos a conclusiones erróneas a nuestros lectores.

No es inusual que los gobiernos de las diferentes administraciones den plantillas de cálculo llenas de errores, y también es fácil confundirse con los términos oficiales que se usan para denominar una partida presupuestaria, o las siglas que se utilizan para hacer referencia a un cuerpo interno, o un departamento, o lo que sea. Lo primero que tenemos que hacer es volver a revisar la documentación y ver si, efectivamente, existe un problema en la versión original de los datos. Si todo lo que hemos hecho nosotros está bien, entonces es el momento de hacer llamadas. No todos los errores son importantes. En los registros de campañas electorales, es común que haya varios centenares de códigos postales que no existen en una base de datos de 100.000 registros. Siempre que no sean todos en la misma ciudad o estén relacionados con un mismo candidato, el registro ocasional equivocado simplemente no importa. La pregunta que nos tenemos que hacer es: «Si utilizamos esta información, ¿los lectores tendrían una visión acertada de lo que dicen los datos?».

### **2) No obsesionarse con la precisión**

La contracara de no hacer suficientes preguntas es obsesionarse con la precisión. Los gráficos exploratorios tienen que ser correctos, pero no conviene preocuparse si no suman exactamente el 100 % o si nos falta alguna información para tener todo el relato completo, porque este gráfico ha de servir para mostrarnos tendencias generales y qué es lo que tenemos que buscar para estructurar nuestra historia.

### **3) Cronologías**

Usar Excel, un documento Word o herramientas especiales como [TimeFlow](#).

### **4) Publicación de los datos**

El esfuerzo de recolección de los datos tiene que coincidir con lo que requiere el gráfico interactivo: si queremos que los lectores puedan explorar los datos de la manera que deseen, tenemos que prever la corrección de errores, el control de datos y la edición del texto.

### **5) Diseñar pensando en dos tipos de lectores**

Una gráfica (ya sea interactiva o un elemento estático que acompaña nuestro texto) tiene que ser fácil de entender a simple vista, pero debe aportar suficientes elementos como para que resulte interesante para el lector más exigente. Si apostamos por una visualización interactiva, nos tenemos que asegurar de que los lectores obtengan algo más que una sola cifra y que puedan, por ejemplo, cruzar datos, establecer correlaciones para llegar a sus propias conclusiones.

### **6) Transmitir una idea**

En primer lugar, debemos asegurarnos de que tenemos una información que queremos que la gente vea y, después, decidir qué idea queremos destacar en el gráfico. En muchos casos, esto significa eliminar información, aunque internet permita que la pongamos toda. A menos que nuestro principal objetivo sea la transparencia de nuestra actividad periodística, la mayor parte de los detalles que hemos recogido en la cronología no son muy relevantes, y pueden resultar aburridos o confusos. De hecho, si queremos mostrar todos los datos y explicar al lector como los hemos obtenido, hay otras maneras de hacerlo que no implican

saturarlo con una visualización que recoja demasiados ítems. Las redes sociales son una buena herramienta para hacer este ejercicio de transparencia: podemos crear grupos en Facebook, por ejemplo, en los que hagamos publicaciones en las que demos constancia del proceso de recogida de la información e incluso facilitemos todas las tablas «brutas» que contengan aquella información que hemos dejado fuera de la visualización.

### 1.3.1.3. Las visualizaciones para narrar historias

La visualización de los datos no solo puede ser llamativa y atraer la atención del lector, también aprovecha el hecho de que la mitad del cerebro humano está dedicado a procesar información visual. Una visualización de datos bien diseñada puede ofrecer a los que la ven una impresión inmediata y profunda, e ir al grano de la cuestión sin enredarse. Sin embargo, a diferencia de otros medios visuales como la fotografía y el vídeo (herramientas que, por cierto, no tenemos que descartar usar de manera complementaria a nuestro gráfico y texto), la visualización de datos también muestra hechos medibles.

El periodismo de datos, en la era de los medios hechos a medida de públicos con puntos de vista particulares, ofrece la oportunidad de narrar historias orientadas principalmente por hechos. Del mismo modo que otras formas de periodismo narrativo, la visualización de datos puede ser efectiva tanto para presentar noticias nuevas –transmitiendo rápidamente nueva información como la ubicación de un accidente o el número de víctimas– como artículos de fondo, en los que se puede profundizar en un tema y ofrecer una nueva perspectiva.

#### Ejemplo

Un ejemplo sería el [gráfico interactivo](#) publicado por *The New York Times* a finales del 2009, que permitía a los lectores analizar la tasa de desempleo por grupos de población en Estados Unidos por medio de filtros demográficos y educativos. Una buena visualización se queda en la cabeza del lector y deja un modelo mental duradero de un hecho, una tendencia o un proceso.

### 1.3.1.4. Ejemplos en los que la visualización de datos puede ser útil

**1) Mostrar el cambio a lo largo del tiempo:** por ejemplo, el crecimiento de la población china desde 1960 es un buen ejemplo de ello. Sin embargo, las visualizaciones de datos también pueden mostrar de manera muy poderosa el cambio a lo largo del tiempo a través de otras formas gráficas. Por ejemplo, la participación en [el mercado de las principales aerolíneas](#) norteamericanas durante varias décadas de concentración en el sector.

**2) Comparar valores:** la visualización de datos también es útil cuando se trata de ayudar a los lectores a comparar dos o más valores discretos, como por ejemplo cuando *National Geographic*, utilizando un [cuadro muy minimalista](#), mostró lo mayores que son las probabilidades de morir de enfermedad coronaria o infarto que en accidentes de aviación o por picaduras de abeja. Así, mostraba las probabilidades relativas de diferentes causas de muerte.

**3) Diseñar con gráficas:** los diagramas de flujo también codifican información en las líneas de conexión, generalmente según el grosor y/o el color de las mismas. Por ejemplo, con la eurozona en crisis y varios miembros incapacitados para hacer frente a la deuda, *The New York Times* buscó dar una explicación al [conjunto de deudas](#) que vincula a los miembros de la UE con sus socios comerciales al otro lado del Atlántico y en Asia.

**4) Explorar grandes bases de datos:** a veces, la visualización de datos es muy efectiva para mostrar información familiar desde un ángulo totalmente nuevo. Sin embargo, ¿qué sucede cuando se tiene información nueva que la gente quiere buscar? La era de los datos lleva a nuevos y sorprendentes descubrimientos casi todos los días, como el análisis de [fotos de Flickr de Eric Fischer](#). Estas bases de datos son todavía más poderosas cuando los usuarios pueden explorar la información y llegar allí donde les resulte más relevante.

#### Bases de datos interactivas

Un ejemplo es la [base de datos interactiva](#) de *The New York Times*, que permite a los usuarios explorar las 100 películas más alquiladas en 12 zonas metropolitanas de Estados Unidos. Otro es el de la BBC que, por su parte, llevó a cabo una investigación de los [muertos en las carreteras de Gran Bretaña entre 1999 y el 2010](#).

**5) Cuándo no tendríamos que usar la visualización de datos:** la visualización de datos efectiva depende de contar con información limpia, buena, precisa y significativa. Así como muchas citas, datos y descripciones alimentan el buen periodismo narrativo, la visualización de datos es tan buena como los datos que la alimentan.

A veces, los datos por sí solos no narran una historia de la mejor manera posible. Si bien un cuadro simple que ilustre una tendencia o una estadística puede ser útil, una narrativa que relate las consecuencias de una cuestión puede tener más impacto para el lector.

- **Cuando tenemos pocos datos:** si los datos varían poco y no muestran una conclusión clara. Quizá en este punto nos demos cuenta de que, o bien necesitamos más investigación y análisis que nos permita hacer una buena visualización, o bien es mejor publicarlos en formato escrito, y apuntar hacia una tendencia determinada, alertando de que no es definitiva.
- **Cuando con una tabla es suficiente:** es limpio, de rápida lectura y no crea expectativas no realistas de una historia. Supone una forma rápida y fácil de presentar datos sencillos.

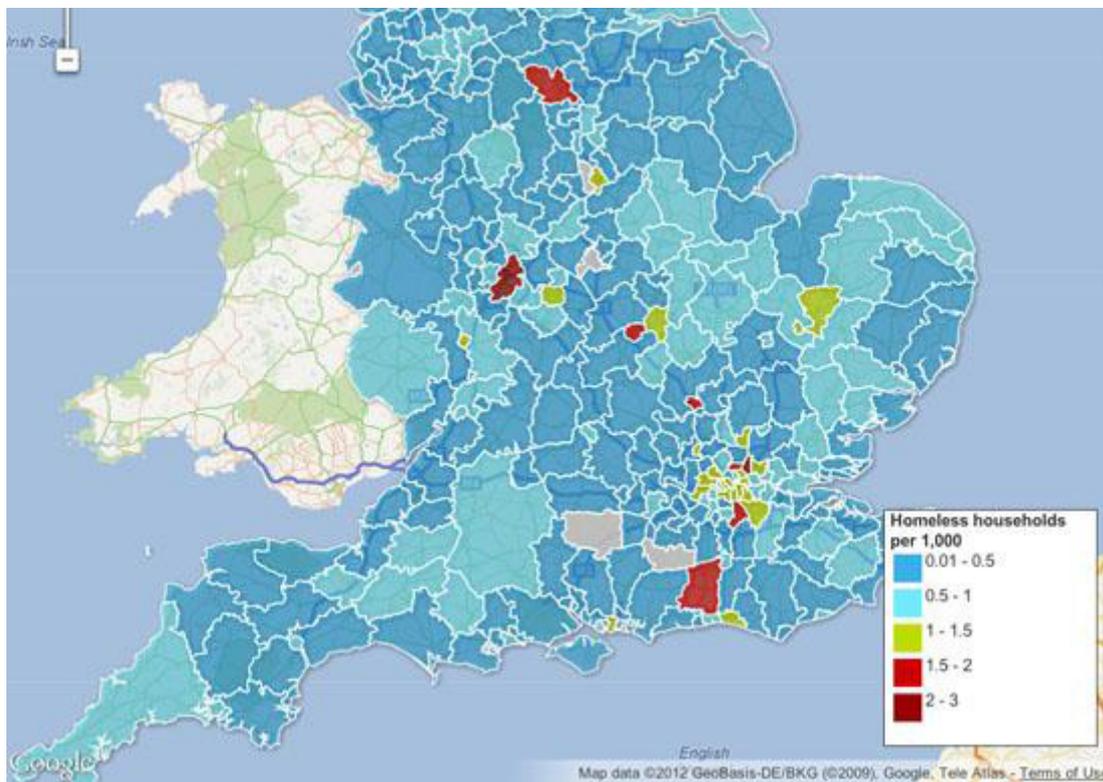
### 1.3.1.5. Herramientas para hacer nuestras propias visualizaciones de datos

Está claro que hay equipos gráficos e interactivos increíbles para hacer grandes investigaciones, pero para la tarea diaria, a menudo es suficiente con herramientas a las que todo el mundo tiene acceso. Estas son algunas de las más comunes:

**1) Google Fusion Tables:** esta [base de datos y herramienta de mapeo en línea](#) se ha convertido en la primera elección para producir mapas rápidos y detallados, especialmente aquellos que requieren *zoom*. La usamos para producir mapas como el de Irak o mapas de fronteras sobre los sintecho, que adjuntamos a continuación.



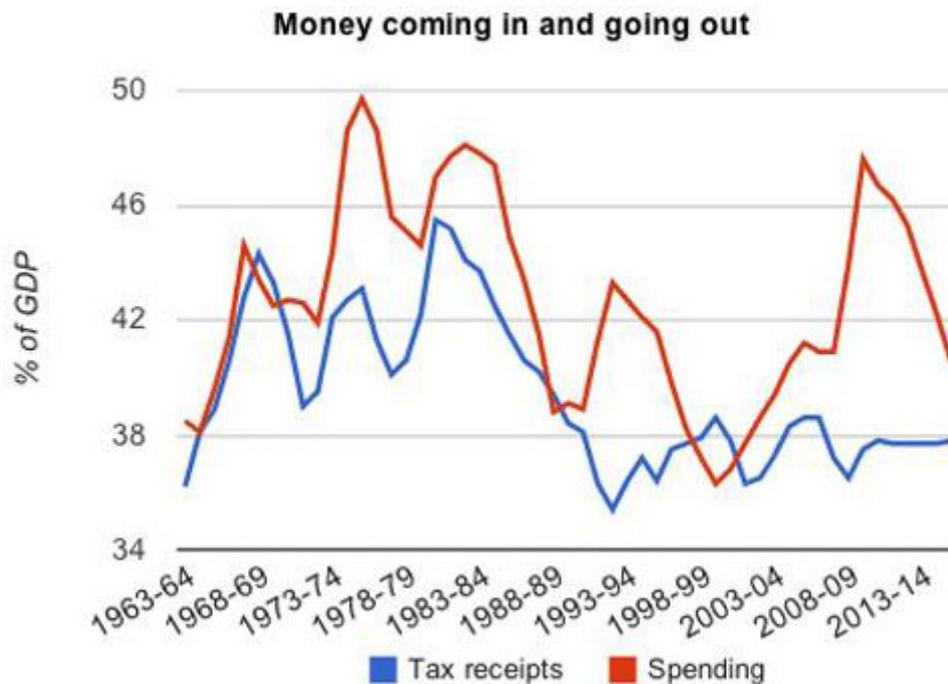
Los registros de guerra de Wikileaks (The Guardian)



Mapa interactivo de personas sin hogar (The Guardian)

**2) Tableau Public:** se usa sobre todo cuando tenemos que unir diferentes tipos de cuadros, como en este [mapa de tasas impositivas](#) máximas en todo el mundo, o como explorador de datos. *La Nación*, un medio de Argentina, ha construido toda su [operación de periodismo de datos](#) con Tableau.

**3) Gráficos con Google Spreadsheet:** puede generar algunos gráficos, como las burbujas animadas que utiliza [Gapminder](#). Es similar a hacer gráficos con Excel, en el sentido de que se seleccionan los datos y se hace clic en el *widget* de gráficos. También merece la pena explorar las opciones de personalización: se puede cambiar el color, las escalas y los encabezados. Es útil, sobre todo, para gráficos pequeños, y los gráficos de curvas incluyen, además, opciones para hacer anotaciones.



Gasto público e impuestos del Reino Unido (The Guardian)

**4) Datamarket:** herramienta práctica para visualizar cifras y, sobre todo, datos de series temporales.

**5) Chartsbin:** es una herramienta para crear mapamundis en los que se puede hacer clic.

**6) iCharts:** especializado en pequeños *widgets* de gráficos.

7) **Geocommons**: ofrece datos de fronteras para crear mapas globales y locales.

8) **Piktochart.com**: ofrece plantillas para estas visualizaciones de texto y cifras.

9) **Infogram**: es una herramienta para visualizar datos simples. Ofrece gráficos lineales, de barras y circulares, y para usarlos basta con introducir la serie de datos que queremos representar gráficamente en el eje de las filas y de las columnas.

## 2. Aspectos jurídicos del periodismo de datos

### Introducción

En este apartado, se hace una aproximación a los principales aspectos jurídicos que hay que tener en cuenta cuando se lleva a cabo periodismo de datos.

El periodismo de datos está estrechamente vinculado a la accesibilidad y a la disponibilidad de datos que producen las administraciones públicas, y que difunden en el marco de las políticas de gobierno abierto y de las leyes de transparencia. A medida que las administraciones públicas se abren a la ciudadanía y explican mejor lo que hacen, los periodistas también tienen más facilidad para acceder a datos que son una fuente importante a la hora de llevar a cabo la actividad profesional.

Desde esta perspectiva, la principal cuestión que se trata en este apartado es la vinculada al acceso a los datos que producen las administraciones públicas, que, en la actualidad, constituyen una de las principales fuentes utilizadas en el periodismo de datos. La legislación vigente prevé diferentes mecanismos por los cuales los periodistas pueden acceder a los datos de las administraciones públicas, y determina unos límites que serán analizados.

También se hará referencia a otras cuestiones que enmarcan la actividad de los periodistas de datos, como son el fundamento de la actividad periodística, basada en las libertades de información y de expresión, y las normas que rigen la actividad periodística.

Los objetivos son los siguientes:

1. Identificar las normas que regulan la difusión, el acceso y la reutilización de los datos que producen las administraciones públicas.
2. Conocer los fundamentos de la libertad de expresión y la libertad de información, así como sus límites.
3. Contextualizar las normas que regulan el ejercicio de la profesión periodística.

## 2. Aspectos jurídicos del periodismo de datos

### 2.1. El gobierno abierto

#### 2.1.1. Introducción

El gobierno abierto hace referencia a los mecanismos que permiten que las administraciones públicas y la ciudadanía puedan trabajar juntos en la generación de valor público (\*).

El gobierno abierto persigue fortalecer las relaciones entre las administraciones públicas y la ciudadanía, mediante la transparencia de las administraciones públicas y la participación y la colaboración ciudadana con el uso intensivo de los medios electrónicos. Por diferentes mecanismos, las administraciones públicas informan a la ciudadanía sobre su actividad y escuchan sus preferencias y sus opiniones, implicándola en el diseño y la implementación de las políticas públicas. El gobierno abierto se basa en el uso intensivo de la tecnología e implica el impulso de la innovación en las administraciones públicas.

Así pues, el gobierno abierto es transparente, es decir, facilita a la ciudadanía información sobre las actividades que llevan a cabo las administraciones públicas, las decisiones que toman y cómo las toman, y cómo se gestionan los recursos públicos.

El gobierno abierto (\*) también es participativo, es decir, escucha las necesidades que expresa la ciudadanía y facilita la implicación en la toma de decisiones públicas por medio de espacios que permitan el diálogo y la deliberación.

La legislación vigente recoge los principios de transparencia y de participación. Algunas normas, incluso, hacen referencia al gobierno abierto y prevén mecanismos específicos. Un ejemplo lo encontramos en la Ley 19/2014, de 29 de diciembre, de transparencia, acceso a la información pública y buen gobierno de Cataluña, que define el gobierno abierto (\*) como «las medidas para establecer una relación y un diálogo permanentes y bidireccionales entre la Administración y las personas a la hora de definir y de aplicar las políticas públicas, y para introducir y desarrollar instrumentos de participación y de colaboración ciudadana en los asuntos públicos».

## 2. Aspectos jurídicos del periodismo de datos

### 2.1. El gobierno abierto

#### 2.1.2. La transparencia pública

##### 2.1.2.1. Introducción

La transparencia es el primer pilar del gobierno abierto, y persigue que la ciudadanía pueda tener un conocimiento efectivo de la actividad de las administraciones públicas.

La transparencia facilita que la ciudadanía pueda participar en las decisiones públicas, y a la vez es un mecanismo mediante el cual las administraciones públicas pueden rendir cuentas a la ciudadanía para que pueda controlar cómo gestionan los recursos públicos o cómo toman las decisiones. La transparencia es un mecanismo que puede fortalecer la integridad pública y convertirse en un freno a las irregularidades y a la corrupción. En última instancia, la transparencia permite incrementar la confianza pública y la afección ciudadana por las instituciones públicas, lo cual tiene que fortalecer su legitimidad.

En este sentido se expresa el preámbulo de la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno (LTAIBG):

«Solo cuando la acción de los responsables públicos se somete a escrutinio, cuando los ciudadanos pueden conocer cómo se toman las decisiones que les afectan, cómo se gestionan los fondos públicos, o bajo qué criterios actúan nuestras instituciones, podemos hablar del inicio de un proceso en el que los poderes públicos empiezan a responder a una sociedad que es crítica, exigente y que pide participación de los poderes públicos.»

La transparencia es un principio jurídico recogido en varias normas. A todos los efectos, podemos referirnos a la Ley 40/2015, de 1 de octubre, de régimen jurídico del sector público, que recoge, entre los principios generales, el principio de transparencia de la actuación administrativa (art. 3). Además, otras normas prevén el principio de transparencia como principio que tiene que guiar la actividad de los cargos y de los trabajadores públicos. En este sentido, la LTAIBG establece como principio de buen gobierno de los altos cargos que actúen con transparencia en la gestión de los asuntos públicos. Igualmente, el Real decreto legislativo 5/2015, de 30 de octubre, por el cual se aprueba el texto refundido de la Ley del Estatuto básico del empleado público, también recoge la transparencia entre los principios que guían la actividad de los trabajadores públicos.

La LTAIBG y, en general, las normas autonómicas de transparencia definen y regulan diferentes mecanismos por medio de los cuales se canaliza la transparencia (\*):

- La difusión de información pública (transparencia activa).
- El acceso a la información pública (transparencia pasiva).
- La reutilización de la información pública (transparencia colaborativa).

##### 2.1.2.2. La difusión de información pública

La difusión de información pública consiste en poner a disposición de la ciudadanía la información por parte de las administraciones públicas.

Las leyes de transparencia prevén que las administraciones públicas tienen que difundir la información por medio de páginas web y sedes electrónicas. La mayoría de estas leyes han previsto que la información se tiene que difundir por un portal de transparencia.

Las leyes de transparencia también definen qué información tienen que difundir las administraciones públicas por los portales de transparencia. Si tomamos como ejemplo la LTAIBG, las administraciones públicas deben difundir la información siguiente:

- Información institucional, organizativa y de planificación. Las administraciones públicas tienen que difundir información sobre las funciones que llevan a cabo, la normativa que les sea aplicable, así como su estructura organizativa. También han de publicar información relativa a los planes y a los programas anuales y plurianuales.
- Información de relevancia jurídica. Las administraciones públicas tienen que publicar, entre otras informaciones de relevancia jurídica, las directrices, las instrucciones, los acuerdos, las circulares; los anteproyectos de ley y los proyectos de decretos

legislativos y de reglamentos, o las memorias y los informes que conformen los expedientes de elaboración de los textos normativos.

- Información económica, presupuestaria y estadística. Las administraciones públicas deben difundir la información sobre todos los contratos, los convenios suscritos, las subvenciones y las ayudas públicas concedidas, los presupuestos, las cuentas anuales, las retribuciones percibidas por los altos cargos, las resoluciones de autorización o de reconocimiento de contabilidad que afecten a los trabajadores públicos; y las declaraciones anuales de bienes y actividades de los representantes locales, o la información relativa al grado de cumplimiento y calidad de los servicios públicos.

Hay que entender esta información como un mínimo, que puede ser ampliado por cada Administración pública (art. 5.2 LTAIBG). De hecho, las leyes autonómicas de transparencia han ampliado de manera significativa la información que deben difundir las administraciones públicas (\*).

## Índice de transparencia

En los últimos años, se han impulsado diferentes índices para medir el nivel de transparencia de las administraciones públicas.

Transparencia Internacional España elabora el índice de transparencia de las comunidades autónomas ([INCAU](#)), que valora el nivel de transparencia de las administraciones autonómicas y el cumplimiento de sus obligaciones, previstas en la legislación básica de transparencia. Esta ONG también elabora el índice de transparencia de los ayuntamientos, que mide la transparencia de los 110 ayuntamientos más grandes de España ([ITA](#)).

El Laboratorio de Periodismo y Comunicación para la Ciudadanía Plural, de la Universidad Autónoma de Barcelona, elabora el mapa [InfoParticipa](#), por medio del cual se evalúa el nivel de transparencia de los ayuntamientos.

La información que difunden las administraciones públicas por los portales de transparencia tiene que ser de calidad, para que pueda cumplir las finalidades previstas. De acuerdo con la legislación vigente (art. 5 LTAIBG), la información que se difunda debe ser:

- clara, estructurada y comprensible para los interesados,
- en formatos reutilizables,
- accesible de acuerdo con el principio de accesibilidad universal y diseño para todo el mundo,
- interoperable,
- identificable y localizable,
- comprensible, y
- de acceso fácil y gratuito.

### 2.1.2.3. El acceso a la información pública

El acceso a la información consiste en la solicitud que hace una persona a una Administración pública para que muestre una determinada información o facilite una copia de la misma. La legislación vigente reconoce el derecho de acceso como un derecho subjetivo (\*) (art. 12 LTAIBG).

De acuerdo con la LTAIBG (art. 13), «se entienden por información pública los contenidos o los documentos, sea cual sea el formato o apoyo, que estén en poder de alguno de los sujetos incluidos en el ámbito de aplicación de este título, y que hayan sido elaborados o adquiridos en el ejercicio de sus funciones».

Cualquier persona puede pedir el acceso a la información pública mediante una solicitud, que hay que dirigir al titular del órgano administrativo o entidad que tenga la información, y que se puede presentar por cualquier medio que permita tener constancia de la identidad del solicitante (incluido el electrónico, con el uso de la firma electrónica), de la información que solicita, de la dirección de contacto –preferentemente electrónica– y de la modalidad que prefiera para acceder a la información. El solicitante

no está obligado a motivar la solicitud, a pesar de que puede exponer los motivos por los cuales pide la información, pero la ausencia de motivo no podrá ser causa de rechazo de la solicitud.

## El acceso a la información de la Administración de la Generalitat

El análisis del periodo comprendido entre el 1 de julio del 2016 y el 31 de diciembre del 2016 del derecho de acceso a la información pública, en el ámbito de la Administración de la Generalitat de Cataluña (\*) y su sector público (marzo 2017), muestra que se formularon 399 solicitudes de acceso (83 % en formato electrónico), y los temas más solicitados fueron los referentes al ámbito de salud (150), de Administración local (27) y de enseñanza (25). Del total de solicitudes, el 64 % fueron estimadas totalmente, y el 16 %, de manera parcial. Finalmente, el número de recursos y de reclamaciones sobre el total de 399 solicitudes presentadas fue del 30,58 %.

Las administraciones públicas pueden no admitir, mediante resolución motivada, las solicitudes de acceso a la información si concurre alguna de las siguientes causas (\*):

- a) Que hagan referencia a información en curso de elaboración o de publicación general.
- b) Que estén referidas a información de carácter auxiliar o de apoyo, como la contenida en notas, borradores, opiniones, resúmenes, comunicaciones e informes internos, o entre órganos o entidades administrativas.
- c) Que sean relativas a información para la divulgación de la que se necesite una acción previa de reelaboración.
- d) Que estén dirigidas a un órgano que no disponga de la información cuando se desconozca el competente.
- e) Que sean manifiestamente repetitivas o tengan un carácter abusivo y no justificado con la finalidad de transparencia de esta ley.

Si hay derechos o intereses de terceras personas afectados por la solicitud de acceso a la información, se les concederá un plazo de quince días para que puedan hacer las alegaciones que consideren oportunas.

El procedimiento se tiene que resolver y notificar en el plazo de un mes desde la recepción de la solicitud por parte del órgano competente para hacerlo. En el supuesto de que la resolución deniegue su acceso, conceda su acceso parcial, o la información se facilite en un formato diferente del pedido, se tendrá que motivar. Si transcurre el plazo de un mes sin que se haya dictado y notificado la resolución, habrá que entender que ha sido desestimada (\*).

Una de las novedades de la LTAIBG es la creación de un organismo independiente encargado de garantizar el cumplimiento de las obligaciones de transparencia activa y pasiva. En particular, esta norma prevé que las personas pueden interponer, con carácter potestativo y previo a la impugnación ante los tribunales, una reclamación al Consejo de Transparencia y Buen Gobierno (\*) u órgano autonómico independiente que determine cada comunidad autónoma. Esta reclamación se puede interponer en el plazo de un mes. El plazo para resolver y notificar la resolución es de tres meses y, una vez transcurridos sin resolución, la reclamación se entenderá desestimada (\*).

## El Consejo de Transparencia y Buen Gobierno

El Consejo de Transparencia y Buen Gobierno es un organismo público con personalidad jurídica propia y plena capacidad de obrar, que actúa con autonomía y plena independencia en el cumplimiento de sus finalidades. De acuerdo con la LTAIBG, el Consejo de Transparencia y Buen Gobierno tiene como finalidades promover la transparencia de la actividad pública, velar por el cumplimiento de las obligaciones de publicidad, salvaguardar el ejercicio del derecho de acceso a la información pública y garantizar la observancia de las disposiciones de buen gobierno.

La LTAIBG también dispone que las comunidades autónomas pueden crear un órgano independiente al que atribuir la competencia para resolver las reclamaciones contra las resoluciones relativas al derecho de acceso a la información pública de las administraciones públicas del territorio. Aun así, las comunidades autónomas pueden atribuir, mediante el oportuno convenio, la competencia para resolver la reclamación al Consejo de Transparencia y Buen Gobierno. La mayoría de las comunidades autónomas han creado, con configuraciones y atribuciones diferentes, órganos u organismos independientes para la garantía de la transparencia pública.

## 2.1.2.4. Los datos abiertos y su reutilización

La reutilización consiste en el uso, por parte de cualquier persona, de la información pública con finalidades comerciales o no comerciales (\*).

La Ley 37/2007, 16 de noviembre, sobre la reutilización de la información del sector público (LRISP), prevé que las administraciones públicas pueden autorizar la reutilización de la información que está bajo su poder, sea cual sea su soporte y su forma de expresión. La LRISP únicamente excluye de la reutilización los documentos que no sean accesibles, aquellos cuyo conocimiento requiera una legitimación especial, y los contenidos públicos creados, por ejemplo, por entidades de radiodifusión o por entidades educativas.

La LRISP prevé tres modalidades de reutilización:

- La reutilización sin sujeción a condiciones.
- La reutilización con sujeción a condiciones, establecidas en licencias tipo, que deben estar disponibles en formato digital y poder procesarse de manera electrónica.
- La reutilización con solicitud previa, que puede incorporar condiciones en una licencia. Esta licencia tiene que reflejar, como mínimo, la información relativa a la finalidad concreta, comercial o no comercial, por la cual se concede la reutilización, la duración de la licencia, las obligaciones del beneficiario y del organismo que lo concede, las responsabilidades de uso y las modalidades financieras, por lo cual hay que indicar el carácter gratuito o, si procede, la contraprestación económica aplicable.

Las condiciones que se incorporen en las licencias de reutilización tienen que ser claras, justas y transparentes; no han de restringir las posibilidades de reutilización, ni limitar la competencia; no deben ser discriminatorias para categorías comparables de reutilización, y no han de contener acuerdos exclusivos (\*).

La LRISP dispone que la reutilización de la información se puede someter a las condiciones generales siguientes:

- a) Que el contenido de la información, incluyendo metadatos, no sea alterado.
- b) Que no se desnaturalice el sentido de la información.
- c) Que se cite la fuente.
- d) Que se mencione la fecha de la última actualización.
- f) Cuando la información contenga datos de carácter personal, la finalidad o las finalidades concretas para las cuales es posible la reutilización futura de los datos.

Cuando la información, pese a ser facilitada de forma disociada, contuviera suficientes elementos que pudiesen permitir la identificación de los interesados en el proceso de reutilización (\*), la prohibición de revertir el procedimiento de disociación mediante la adición de nuevos datos obtenidos otras fuentes.

La reutilización de la información puede verse facilitada por la apertura de los datos públicos (\*).

La apertura de los datos públicos favorece la reutilización de la información pública, al facilitar la creación de información y la oferta de nuevos servicios de valor añadido basados en los datos difundidos por las administraciones públicas. En este sentido, la Comisión Europea afirma lo siguiente:

«[...] El impulso hacia la utilización de datos abiertos está tomando fuerza en distintos estados miembros, que adoptan este concepto por razones de transparencia, eficiencia administrativa y potencial económico de la reutilización.»

Comisión Europea (2011).

La apertura de los datos públicos consiste en eliminar los obstáculos técnicos o jurídicos que pueden dificultar la reutilización de la información pública. En efecto, los datos públicos que se ponen a disposición de la ciudadanía tienen que cumplir determinadas características que persiguen garantizar que se puedan reutilizar fácilmente para generar nuevas informaciones o servicios que, en última instancia, pueden incrementar la transparencia de las administraciones públicas. Además, el cumplimiento de las obligaciones de transparencia exige un cambio en la forma en que se gestiona, almacena y recupera la información no solo para facilitar una mayor eficacia y eficiencia administrativa, sino para contribuir en el incremento de la transparencia pública.

Los datos abiertos se caracterizan, desde un punto de vista técnico, por poder procesarse electrónicamente y, desde el punto de vista jurídico, por la ausencia de restricciones que limitan su reutilización (propiedad intelectual, datos personales, etc.). De acuerdo con la Sunlight Foundation, los datos abiertos son aquellos que cumplen las siguientes condiciones:

- Datos completos. Los datos tienen que ser tan completos como resulte posible, y no pueden estar afectados por limitaciones relativas a la privacidad, la seguridad o cualquier otra restricción.
- Datos primarios. Los datos abiertos deben ser datos originales. Hay que facilitar información sobre cómo se han obtenido y dónde están los documentos originales, de forma que los usuarios puedan comprobar si los datos han sido seleccionados y almacenados adecuadamente.
- Datos oportunos. Los datos se tienen que facilitar de forma rápida, tan pronto como se recojan. Dentro de lo posible, hay que facilitar los datos en tiempo real, de forma que se pueda maximizar su valor. Hay que dar prioridad a la difusión de los datos que sean perecederos (*time sensitive*).
- Datos accesibles de forma física y electrónica. Los datos tienen que ser tan accesibles como sea posible, tanto por medios físicos como electrónicos, para evitar los desplazamientos o la necesidad de solicitar formalmente el acceso a la información.
- Datos procesables electrónicamente. Los datos deben estar disponibles en formatos de uso generalizado y que faciliten el procesamiento automático.
- Datos no discriminatorios. Los datos deben estar disponibles para todo el mundo, sin que sea necesario hacer una solicitud o cualquier otro trámite.
- Datos que utilizan estándares abiertos en la difusión.
- Datos reutilizables sin licencia. El uso de los datos no se tiene que someter a ninguna regulación que restrinja la reutilización excepto, de forma razonable, cuando haya aspectos relativos a la privacidad o la seguridad.
- Datos permanentes.
- Datos sin costes. No se tiene que establecer una contraprestación para poder acceder o reutilizar los datos y, en caso de que se establezca, su cuantificación se tiene que basar en costes marginales y no en costes totales.

La LRISP (art. 5 Ley 37/2007), después de la modificación operada por la Ley 18/2015, de 9 de julio, dispone que:

«2. Las administraciones y organismos del sector público facilitarán sus documentos en cualquier formato o lengua preexistente, pero también procurarán, siempre que ello sea posible y apropiado, proporcionarlos en formato abierto y legible por máquina conforme a lo previsto en el apartado anterior y conjuntamente con sus metadatos, con los niveles más elevados de precisión y desagregación. Tanto el formato como los metadatos, en la medida de lo posible, deben cumplir estándares y normas formales abiertas. Esto no implicará que las administraciones y organismos del sector público estén obligados a crear documentos, adaptarlos o facilitar extractos de documentos, cuando ello suponga un esfuerzo desproporcionado que conlleve algo más que una simple manipulación.»

De este modo, la LRISP prevé que, cuando sea posible, las administraciones públicas tienen que dar los datos abiertos para facilitar su reutilización. Aun así, pese a la formulación de este precepto, es importante tener presente que si bien las administraciones públicas deben intentar difundir los datos en formato abierto y promover el hecho de que ponerlos a disposición para su reutilización se haga por medios electrónicos, y mediante plataforma multicanal, no tienen la obligación de difundirlos en este formato.

Diferentes normas autonómicas prevén que las administraciones públicas tienen que promover la difusión de los datos públicos en formatos abiertos (Castilla-La Mancha, Galicia, Aragón, Murcia, Andalucía y Extremadura). Asimismo, varias normas autonómicas disponen la creación de portales de datos abiertos o de secciones específicas al efecto en los portales de transparencia, con el fin de favorecer su reutilización (Castilla-La Mancha y Aragón). Otras normas también prevén la creación de catálogos de información reutilizable (Galicia y La Rioja).

El [Ayuntamiento de Barcelona](#) difunde 420 conjuntos de datos (*datasets*) en 20 formatos diferentes que facilitan su consulta y reutilización, como por ejemplo el Bicing, el censo de actividades comerciales o los equipamientos escolares (conjuntos más descargados).

La LRISP establece diferentes infracciones por el incumplimiento de lo que se dispone en el ámbito de la Administración general del Estado, como por ejemplo la desnaturalización del sentido de la información; la alteración del contenido de la información para la reutilización; y la reutilización de documentación sin haber obtenido la correspondiente licencia, en los casos en los que sea requerida o incumpla las condiciones impuestas en la licencia o la normativa reguladora aplicable. La LRISP prevé que, por la comisión de las anteriores infracciones, se pueden imponer diferentes sanciones de multa, además de la prohibición de reutilizar documentos sometidos a licencia durante un periodo de entre 1 y 5 años, y con la revocación de la licencia concedida. Este régimen sancionador únicamente es aplicable en relación con la Administración general del Estado.

### 2.1.2.5. Los límites de la transparencia pública. La protección de datos personales

A veces, el conocimiento público de determinada información pública puede generar daños o perjuicios a determinados bienes, derechos e intereses, cuya protección es necesaria. Para evitarlo, la LTAIBG dispone que la transparencia puede ser limitada cuando el conocimiento o el acceso a la información signifique un perjuicio a una serie de bienes o derechos (\*) como:

- la seguridad nacional,
- la defensa,
- las relaciones exteriores,
- la seguridad pública,
- la prevención, investigación y sanción de los ilícitos penales, administrativos o disciplinarios,
- la igualdad de las partes en los procesos judiciales y la tutela judicial efectiva,
- las funciones administrativas de vigilancia, inspección y control,
- los intereses económicos y comerciales,
- la política económica y monetaria,
- el secreto profesional y la propiedad intelectual industrial,
- la garantía de la confidencialidad o el secreto requerido en procesos de toma de decisión, y la protección del medio ambiente.

Estos límites son aplicables tanto en la difusión de la información pública, como en el acceso a la información e, incluso, en la reutilización. En efecto, el artículo 5.3 LTAIBG dispone que «son aplicables, si procede, los límites al derecho de acceso a la información pública que prevé el artículo 14 y, especialmente, el derivado de la protección de datos de carácter personal, que regula el artículo 15». Asimismo, el artículo 14 establece que «el derecho de acceso podrá ser limitado cuando acceder a la información suponga un perjuicio» para determinados bienes y derechos. Finalmente, la LRISP prevé que no será aplicable en los documentos sobre los cuales haya prohibiciones o limitaciones en el derecho de acceso, en virtud de lo que prevé el LTAIBG, así como «en los documentos a los cuales no pueda acceder o el acceso esté limitado en virtud de regímenes de acceso por motivos de protección de los datos personales, en conformidad con la normativa vigente y las partes de documentos accesibles, en virtud de estos regímenes que contengan datos personales y la reutilización se haya definido por ley como incompatible con la legislación relativa a la protección de las personas físicas, en cuanto al tratamiento de los datos personales».

Con carácter general, hay que advertir que la simple concurrencia de uno de estos bienes o derechos en la información que se quiera difundir, o a la cual una persona quiera acceder, no determina la imposibilidad del conocimiento público. Sobre esto, hay que tener en cuenta los criterios que se derivan de lo que prevé la LTAIBG.

En primer lugar, las administraciones públicas no disponen de potestad discrecional para denegar o limitar el acceso a la información pública.

En segundo lugar, la aplicación de los límites tiene que ser motivada. Así lo señala con carácter básico la LTAIBG, cuando afirma que «serán motivadas las resoluciones que denieguen el acceso, las que concedan el acceso parcial o mediante una modalidad diferente de la solicitada, y las que permitan el acceso cuando haya habido oposición de un tercero».

En tercer lugar, la concurrencia de un interés o de un derecho protegido por un límite no significa obligatoriamente la denegación del acceso a la información pública. En efecto, la LTAIBG prevé que el acceso podrá ser limitado, pero no hace falta que lo sea, ni que lo sea directamente.

En cuarto lugar, hace falta que la difusión o el acceso a la información signifique un perjuicio concreto, definido y evaluable. Para poder determinar el perjuicio que el conocimiento de la información pueda generar, así como la concurrencia de un interés que justifique la difusión o el acceso a la información pública, las administraciones públicas pueden utilizar diferentes instrumentos.

Por un lado, el test del daño por medio del cual la Administración pública tiene que valorar el perjuicio, daño o incluso el riesgo que el acceso puede ocasionar al bien o interés protegido por el límite. Por otro lado, el test de la ponderación de intereses mediante el cual la Administración pública tiene que ponderar si se da un interés público o privado superior que justifique el acceso.

La LTAIBG dispone que cuando la aplicación de alguno de los límites que prevé no afecte a la totalidad de la información, se concederá el acceso parcial, habiendo omitido previamente la información afectada por el límite, salvo que resulte una información distorsionada o que no tenga sentido, y se indicará al solicitante que parte de la información ha sido omitida.

Además de los límites que se han citado anteriormente, la LTAIBG también dispone que cuando la información solicitada contiene datos personales, el acceso a la información puede ser limitado si concurren determinadas circunstancias o se cumplen determinados criterios. Con este objeto, los datos personales se clasifican en tres grupos que tienen diferente régimen de transparencia (\*).

### **1) La transparencia de la información que contiene datos personales especialmente protegidos**

La LTAIBG prevé dos regímenes distintos.

En primer lugar, cuando la información solicitada contiene datos especialmente protegidos relativos a la ideología, la afiliación sindical, la religión o las creencias, el acceso solo se puede autorizar en caso de que se tenga el consentimiento expreso y por escrito del afectado, salvo que este haya hecho manifiestamente públicos los datos antes de que se pidiera el acceso.

En segundo lugar, cuando la información incluye datos especialmente protegidos sobre el origen racial, la salud o la vida sexual o datos relativos a la comisión de infracciones penales o administrativas que no supongan la amonestación pública al infractor, el acceso solo se puede autorizar en caso de que se tenga el consentimiento del afectado, o si está amparado por una norma con rango de ley.

### **2) La transparencia de la información que contiene datos personales**

Cuando la información solicitada no contenga datos especialmente protegidos, el órgano al cual se dirija la solicitud tiene que conceder su acceso, después de ponderar de manera bastante razonada el interés público en la divulgación de la información y los derechos de los afectados, cuyos datos aparezcan en la información solicitada.

Para llevar a cabo esta ponderación, la LTAIBG prevé los criterios siguientes:

- a)** El menor perjuicio a los afectados derivado del transcurso de los plazos previstos en la Ley 16/1985, de 25 de junio, del patrimonio histórico español (veinticinco años desde la muerte de la persona afectada, si se conoce la fecha, o cincuenta años desde la fecha del documento).
- b)** La justificación por los solicitantes de la petición en el ejercicio de un derecho, o el hecho de que tengan la condición de investigadores y motiven el acceso a finalidades históricas, científicas o estadísticas.
- c)** El menor perjuicio de los derechos de los afectados en caso de que los documentos únicamente contengan datos de carácter meramente identificativo.
- d)** La mayor garantía de los derechos de los afectados en caso de que los datos contenidos en el documento puedan afectar a su intimidad o seguridad, o se refieran a menores de edad.

### **3) La transparencia de la información que contiene datos personales meramente identificativos**

Cuando la información a la que se quiere acceder, difundir o reutilizar solo contenga datos identificativos relacionados con la organización, el funcionamiento o la actividad pública del órgano (nombre y apellidos, DNI, cargo, teléfono, dirección de correo electrónico, etc.), serán accesibles salvo en el caso concreto de que prevalezca la protección de datos personales u otros derechos constitucionalmente protegidos sobre el interés público en la divulgación que lo impida.

Cuando la Administración pública, con carácter previo a la difusión o al acceso, ha anonimizado el documento o disociado los datos personales, no se aplicarán los criterios que se han comentado anteriormente, siempre que se impida la identificación de las personas afectadas.

## **Orientaciones sobre protección de datos en la reutilización de la información del sector público**

La Agencia Española de Protección de Datos propone que antes de decidir facilitar la reutilización de la información, hay que examinar los riesgos para los interesados y las medidas que se pueden utilizar para minimizarlos. En este sentido, en los casos en que se puedan derivar riesgos para la protección de los datos personales, es preciso anonimizar la información, siendo conscientes de las posibilidades de reidentificar a las personas, y de la necesidad de complementar la anonimización con compromisos jurídicamente vinculantes por licencias específicas que hagan referencia a ello.

La finalidad del proceso de anonimización es eliminar o reducir al mínimo los riesgos de reidentificación de los datos anonimizados, y mantener la veracidad de los resultados del tratamiento, es decir, además de evitar la identificación de las personas, los datos anonimizados tienen que garantizar que cualquier operación o tratamiento que puedan llevarse a cabo posteriormente a la anonimización no suponen una distorsión de los datos reales. De acuerdo con la AEPD, hay que garantizar la irreversibilidad de la anonimización, y para ello tendrán que valorar tanto las fuentes de información disponibles en los diferentes medios y, en particular, mediante internet, como la tecnología aplicable en los procesos de anonimización y en los de reidentificación. La anonimización se puede llevar a cabo con diferentes técnicas que conviene utilizar de manera combinada para reducir los riesgos de identificación.

Finalmente, hay que tener presente que la normativa de protección de datos personales es aplicable al tratamiento posterior de los obtenidos por medio del ejercicio del derecho de acceso.

## 2. Aspectos jurídicos del periodismo de datos

### 2.1. El gobierno abierto

#### 2.1.3. La participación y la colaboración ciudadana

La participación es el segundo pilar del gobierno abierto, y consiste en la intervención de la ciudadanía en el diseño, la ejecución, el seguimiento y el control de la actividad de las administraciones públicas. La participación también facilita la colaboración ciudadana en el desarrollo de estas actividades, el fomento de la integridad pública y la prevención de los conflictos de intereses, y la lucha contra la corrupción.

La participación y la colaboración se pueden llevar a cabo mediante diferentes aplicaciones (\*).

La participación y la colaboración ciudadana (\*) adquieren un papel importante cuando hablamos de periodismo de datos.

Los profesionales del periodismo que han accedido a los datos pueden colaborar con las administraciones públicas en la difusión.

#### «¿Quién cobra la obra?»

Es un proyecto impulsado por la Fundación Ciudadana Civio, en el que se han analizado todos los contratos de obra pública publicados en el BOE entre los años 2009 y el 2015 (en total, 8.058 contratos). La elaboración del proyecto ha tenido bastantes dificultades por el bajo nivel de reusabilidad de la información analizada, que ha implicado que algunas de las tareas se hayan tenido que hacer de forma manual.

El proyecto permite visualizar de forma gráfica las constructoras adjudicatarias, las obras públicas y los contratos adjudicados, así como buscar datos y compararlos.

También podemos colaborar con las administraciones públicas analizando los datos, lo cual puede permitir un mejor conocimiento del funcionamiento de las administraciones públicas, de cómo gastan el dinero o cómo adjudican los contratos públicos.

#### Development Gateway

La ONG Development Gateway, ha desarrollado un tablero de mando basado en software libre y en el estándar de datos de contratación abierta, que a partir de unas alertas (*red flags*), previamente definidas, permite identificar la posible existencia de irregularidades o de casos de corrupción, o los riesgos que hay para que puedan ser analizados o investigados. Este sistema también permite identificar las irregularidades más frecuentes en una determinada Administración pública en materia de contratación pública.

Finalmente, los periodistas también pueden colaborar informando o denunciando las irregularidades que detecten por el análisis que hayan hecho de los datos a los que hayan podido acceder.

#### Plataforma Xnet

En la actualidad, hay aplicaciones que permiten la información o denuncia de irregularidades y casos de corrupción por medios electrónicos. Como ejemplo, el Buzón de Xnet permite que cualquier persona que tenga conocimiento de una irregularidad o de un caso de corrupción lo pueda comunicar de manera confidencial para que se analice si puede tener una salida jurídica, y para que se haga llegar a periodistas que colaboran en la plataforma.

## 2. Aspectos jurídicos del periodismo de datos

### 2.2. Libertad de información y de expresión

#### 2.2.1. Introducción

La Constitución española reconoce la libertad de expresión y la libertad de información como derechos fundamentales (art. 20).

#### La libertad de expresión y la libertad de información en los tratados internacionales

La libertad de expresión y la libertad de información están reconocidas en varios tratados internacionales. Estos, más allá del valor intrínseco, son un instrumento para la interpretación de las normas que regulan los derechos fundamentales (art. 10.2 CE (\*)).

##### Pacto Internacional de los Derechos Civiles y Políticos (art. 19)

- 1) Nadie podrá ser molestado a causa de sus opiniones.
- 2) Toda persona tiene derecho a la libertad de expresión; este derecho comprende la libertad de buscar, recibir y difundir informaciones e ideas de todo tipo, sin consideración de fronteras, tanto oralmente, por escrito o en forma impresa o artística, o por cualquier otro procedimiento que elija.
- 3) El ejercicio del derecho previsto en el párrafo 2 de este artículo supone deberes y responsabilidades especiales. Por eso, puede estar sujeto a algunas restricciones que tendrán que estar expresamente fijadas por la ley, y que sean necesarias para:
  - a) Asegurar el respeto a los derechos o a la reputación de otros.
  - b) La protección de la seguridad nacional, el orden público o la salud o la moral públicas.

##### Convenio Europeo de Derechos Humanos (art. 10)

Toda persona tiene derecho a la libertad de expresión. Este derecho comprende la libertad de opinión y la libertad de recibir o de comunicar informaciones o ideas sin la injerencia de las autoridades públicas y sin consideración de fronteras. El presente artículo no impide que los estados sometan a las empresas de radiodifusión, de cinematografía o de televisión a un régimen de autorización previa.

##### Carta de los Derechos Fundamentales de la Unión Europea (art. 11)

- 1) Toda persona tiene derecho a la libertad de expresión. Este derecho comprende la libertad de opinión y la libertad de recibir o comunicar informaciones o ideas sin que pueda haber injerencia de las autoridades públicas y sin consideración de fronteras.
- 2) Se respetan la libertad de los medios de comunicación y el pluralismo.

La libertad de expresión y la libertad de información están estrechamente vinculadas a la «libertad de comunicar el producto de una tarea intelectual previa» y a la «libertad de las personas a manifestarse en la sociedad por medio de las ideas», a pesar de que tienen un contenido autónomo y diferenciado.

En este sentido, el Tribunal Constitucional se ha manifestado en numerosas ocasiones considerando que la libertad de expresión y la libertad de información son dos derechos diferentes y autónomos, a pesar de que estrechamente vinculados en origen. La libertad de expresión hace referencia a la libre difusión de pensamientos, de ideas y de opiniones, mientras que la libertad de información se refiere a la manifestación de hechos o de sucesos noticiables cuya veracidad se pueda demostrar.

“ «Nuestra jurisprudencia viene distinguiendo, desde la STC 104/1986, de 17 de julio, entre los derechos que garantizan la libertad de expresión, cuyo objeto son los pensamientos, ideas y opiniones (concepto amplio que incluye las apreciaciones y los juicios de valor) y, por otra parte, el derecho a comunicar información, que se refiere a la difusión de aquellos hechos que merecen ser considerados noticiables. Esta distinción entre pensamientos, ideas y opiniones, de un lado, y comunicación informativa de hechos, de otro, tiene decisiva importancia a la hora de determinar la legitimidad del ejercicio de esas libertades, pues mientras los hechos son susceptibles de prueba, las opiniones o juicios de valor, por su misma naturaleza, no se prestan a una

demostración de exactitud, y ello hace que al que ejercita la libertad de expresión no le sea exigible la prueba de la veracidad o diligencia en su averiguación, que condiciona, en cambio, la legitimidad del derecho de información por expreso mandato constitucional, que ha añadido al término “información”, en el texto del artículo 20.1d CE, el adjetivo “veraz” (STC 4/1996, de 19 de febrero).»

STC 278/2005, de 7 de noviembre.

La libertad de expresión y la libertad de información contribuyen a la formación de la opinión pública libre de la que son el fundamento, y a la vez son el motivo para el ejercicio de otros derechos fundamentales, como por ejemplo el derecho a la participación política y otros inherentes al funcionamiento del sistema democrático (STC 159/1986, de 12 de diciembre).

Desde el punto de vista de las garantías, la libertad de expresión y la libertad de información, como derechos fundamentales que son, tienen reconocidos diferentes mecanismos de garantía en la Constitución española.

Por un lado, la eficacia directa de estos derechos vinculantes en todos los poderes públicos. Cualquier ciudadano puede reclamar la tutela judicial de estos derechos, que puede canalizarse por un procedimiento preferente y sumario ante los tribunales de justicia y, de manera extraordinaria, mediante un recurso de amparo ante el Tribunal Constitucional.

Por otro lado, la regulación de estos derechos hay que hacerla por medio de una ley orgánica, es decir, una ley aprobada por las Cortes Generales por mayoría absoluta de los diputados.

La libertad de expresión y la libertad de información no pueden ser restringidas mediante ningún tipo de censura previa (art. 20.2 CE). De acuerdo con el Tribunal Constitucional (STC 187/1999, de 25 de octubre), «por censura previa debe tenerse cualquier medida limitativa de la elaboración o de la difusión de una obra del espíritu, que consista en el sometimiento a un examen previo del contenido por parte de un poder público, y cuya finalidad sea enjuiciar la obra en cuestión de acuerdo con unos valores abstractos y restrictivos de la libertad, de forma que se otorgue el plázet a la publicación de la obra si se acomoda al juicio del censor, y se niegue en caso contrario».

## 2. Aspectos jurídicos del periodismo de datos

### 2.2. Libertad de información y de expresión

#### 2.2.2. La libertad de información

##### 2.2.2.1. Introducción

El artículo 20 CE reconoce y protege el derecho de comunicar o de recibir libremente información veraz por cualquier medio de difusión. Cualquier persona, en el marco de su autonomía personal, puede ser titular del derecho a la libertad de información. Por lo tanto, no solo los periodistas son titulares de este derecho, a pesar de que son quienes más a menudo lo ejercen.

Como se desprende de la formulación, este derecho se basa en la veracidad de la información. La libertad de información «ampara la transmisión de hechos verídicos que, a la vez, pueden ser considerados noticiables, es decir, que revistan interés o relevancia pública para la sociedad». El Tribunal Constitucional, en su jurisprudencia, reconoce la estrecha conexión entre la libertad de información y la libertad de formación de la opinión pública, por lo cual tiene que prevalecer por encima de otros derechos con los que pueda entrar en conflicto, siempre que la información que se transmite sea veraz y relevante.

##### 2.2.2.2. Veracidad

La veracidad no quiere decir verdad o realidad incontrovertible de los hechos, sino que hace referencia a la diligencia del periodista para averiguar los hechos sobre los que versa la información, y de acuerdo con aquello que es exigible a un profesional de la información que excluye rumores o invenciones. Tampoco quiere decir que la información no pueda ser errónea, sino que haya sido contrastada previamente con datos objetivos o, si procede, que se rectifique rápidamente.

En varias sentencias, el Tribunal Constitucional ha perfilado el alcance de este término y ha identificado los elementos que tienen que concurrir para poder acreditar la veracidad. En efecto, la jurisprudencia constitucional ha indicado que la veracidad está vinculada a la posibilidad de contrastar la información que se difunde, a la diligencia del periodista para averiguar la información, a la actitud positiva respecto a la verdad, y a la actividad del periodista para comprobar de manera razonable los hechos que se difundirán.

“ «Sobre la veracidad de la información, este Tribunal ha establecido una consolidada doctrina según la cual este requisito constitucional “no va dirigido a la exigencia de una rigurosa y total exactitud en el contenido de la información, sino a negar la protección constitucional a los que transmiten como hechos verdaderos, bien simples rumores, carentes de toda constatación, o bien meras invenciones o insinuaciones sin comprobar su realidad mediante las oportunas averiguaciones propias de un profesional diligente; todo ello sin perjuicio de que su total exactitud puede ser controvertida o se incurra en errores circunstanciales que no afecten a la esencia de lo informado [...]. La razón se encuentra en que, como hemos señalado en muchas ocasiones, cuando la Constitución requiere que la información sea “veraz” no está tanto privando de protección a las informaciones que puedan resultar erróneas, como estableciendo un deber de diligencia sobre el informador a quien se le puede y debe exigir que lo que transmite como “hechos” haya sido objeto de previo contraste con datos objetivos [...]. De este modo, el requisito de la veracidad deberá entenderse cumplido en aquellos casos en los que el informador haya realizado, con carácter previo a la difusión de la noticia, una labor de averiguación de los hechos sobre los que versa la información y la referida indagación la haya efectuado con la diligencia que es exigible a un profesional de la información.»

STC 158/2003, de 15 de septiembre

Otro aspecto relevante es el relativo a la neutralidad con la que el periodista trata la información.

“ «Para que pueda hablarse de reportaje neutral han de concurrir los siguientes requisitos:

El objeto de la noticia ha de hallarse constituido por declaraciones que imputan hechos lesivos del honor, pero que han de ser por sí mismas, esto es, como tales declaraciones, noticia y han de ponerse en boca de personas determinadas responsables de ellas (SSTC 41/1994, de 15 de febrero,

FJ 4, y 52/1996, de 26 de marzo FJ 5). De modo que se excluye el reportaje neutral cuando no se determina quién hizo tales declaraciones (STC 190/1996, de 25 de noviembre, FJ 4 b).

El medio informativo ha de ser mero transmisor de tales declaraciones, limitándose a narrarlas sin alterar la importancia que tengan en el conjunto de la noticia (STC 41/1994, de 15 de febrero, FJ 4). De modo que si se reelabora la noticia no hay reportaje neutral (STC 144/1998, de 30 de junio, FJ 5) y tampoco lo hay cuando es el medio el que provoca la noticia, esto es, en el llamado periodismo de investigación (STC 6/1996, de 16 de enero, VP), sino que ha de limitarse a reproducir algo que ya sea, de algún modo, conocido.

Y sobre esta base cuando se reúnen ambas circunstancias la veracidad exigible se limita a la verdad objetiva de la existencia de dichas declaraciones y a la fidelidad a su contenido: si concurren ambas circunstancias el medio ha de quedar exonerado de responsabilidad.»

STC 53/2006, de 27 de febrero

### 2.2.2.3. Relevancia pública

No basta con la veracidad para que la información que se comunique se considere una manifestación de la libertad de información. La libertad de información está más estrechamente vinculada a la relevancia pública que a la información, es decir, relacionada con asuntos públicos de interés general. Así pues, la transmisión de rumores no está amparada por el derecho a la libertad de información.

Hay que recordar que la libertad de información está relacionada con la formación de la opinión pública. Por este motivo, la información tiene que ser de interés para la ciudadanía, lo cual es diferente de la simple satisfacción de la curiosidad humana para conocer la vida de otros (STC 83/2002, de 22 de abril).

De este modo, el periodista, antes de comunicar la información, tiene que valorar si esta puede tener o no interés para la ciudadanía.

## 2. Aspectos jurídicos del periodismo de datos

### 2.2. Libertad de información y de expresión

#### 2.2.3. La libertad de expresión

La Constitución española (art. 20.1.a) define la libertad de expresión como el derecho «a expresar y difundir libremente los pensamientos, las ideas y las opiniones mediante la palabra, la escritura o cualquier otro medio de reproducción».

La libertad de expresión tiene como objeto garantizar que las personas puedan exteriorizar o transmitir sus pensamientos, ideas, opiniones o juicios de valor.

La libertad de expresión está conformada por dos elementos.

En primer lugar, la creación de pensamientos, ideas u opiniones como manifestación de la capacidad humana. Estas creencias, razonamientos o ideas, en la medida en que son propios de cada persona, y por lo tanto pueden ser subjetivos, no tienen que ser veraces ni hay que acreditar su veracidad. No hace falta que se demuestre su exactitud, y por lo tanto, cuando alguien ejerce la libertad de expresión no se le pide que demuestre su veracidad, o muestre diligencia para averiguarla, como sucede en el caso de la libertad de información.

En segundo lugar, el medio por el cual se manifiestan. Como se desprende de la lectura de la Constitución española, las personas pueden expresar sus ideas o pensamientos por cualquier medio, incluso del lenguaje no verbal.

El ejercicio de la libertad de expresión está estrechamente vinculado a otros derechos, como por ejemplo la libertad ideológica (art. 16.1 CE), el derecho a recibir información (art. 20.1.d CE), y la libertad de expresión activa.

En primer lugar, la libertad ideológica es previa a la libertad de expresión y es el origen de las ideas y las opiniones de las personas.

En segundo lugar, el derecho a recibir información veraz por cualquier medio permite a las personas formarse una opinión.

En tercer lugar, la libertad de expresión activa es más amplia que la libertad de información que, como se ha visto, está vinculada a la veracidad y al interés público de la información, aspectos que no tienen que concurrir cuando se habla de la libertad de expresión.

## 2. Aspectos jurídicos del periodismo de datos

### 2.2. Libertad de información y de expresión

#### 2.2.4. Límites a la libertad de información y de expresión

##### 2.2.4.1. Introducción

La libertad de información y la de expresión pueden ser limitadas para proteger otros derechos o bienes. De acuerdo con la Constitución española, la libertad de información y de expresión tienen el límite en el respeto a los derechos reconocidos en el título primero, en los preceptos de las leyes que lo desarrollen y, especialmente, en el derecho al honor, a la intimidad, a la imagen y a la protección de la juventud y de la infancia (art. 20.4 CE). Además, hay leyes que también prevén límites para proteger bienes, como por ejemplo la seguridad pública.

##### 2.2.4.2. Los derechos de personalidad como límite de las libertades de información y de expresión

La Constitución española reconoce la dignidad de la persona, los derechos inviolables que le son inherentes, y el libre desarrollo de la personalidad (art. 10.1 CE). También garantiza el derecho al honor, a la intimidad personal y familiar y a la imagen (art. 18.1 CE).

Todos estos aspectos constituyen límites de las libertades de información y de expresión. Aun así, no son límites absolutos. La libertad de expresión o la libertad de información pueden prevalecer respecto al derecho a la protección del honor, la intimidad o la imagen en determinadas circunstancias, como por ejemplo cuando se derive de la ley, porque esté autorizado en la ley misma, se autorice por una autoridad pública de acuerdo con lo que prevé la ley, prevalezca un interés histórico, científico o cultural relevante, o por actos propios o por consentimiento (\*).

Teniendo en cuenta el carácter relativo de los derechos, tanto el Tribunal Constitucional como el Tribunal Supremo han identificado unos criterios para llevar a cabo la ponderación entre los distintos derechos. En primer lugar, hay que determinar qué derecho se está ejerciendo, es decir, si el conflicto es fruto del ejercicio de la libertad de expresión o de la libertad de información. En segundo lugar, valorar si el derecho en cuestión se ha ejercido correctamente o se ha excedido en los límites establecidos. Y en tercer lugar, determinar si en el ejercicio de la libertad de información o de expresión se han superado los límites permitidos, de forma que tenga que prevalecer el derecho al honor, a la intimidad o a la imagen.



«Lo indicado conduce a contemplar de nuevo el hasta ahora tan trascendente tema de la prioridad o prevalencia de los derechos fundamentales entre sí, a cuyos efectos, conviene insistir, en que tanto la doctrina de esta sala como la del Tribunal Constitucional vienen apuntando en orden a tan discutida cuestión y como punto generalizado de partida, que en principio todos los derechos establecidos en el texto constitucional como fundamentales, por el mero hecho de tal calificación, son en principio iguales, lo que no es obstáculo para reconocer que en ciertas circunstancias haya de otorgarse prevalencia a alguno de ellos sobre otros, atendiendo a los intereses que más directamente tutelen y siempre que en el ejercicio del que en cada caso concreto pueda aparecer como prevalente no se hayan sobrepasado ciertos límites, ya que la delimitación de la colisión en tales supuestos ha de realizarse caso por caso, esto es, sin fijar de modo apriorístico los límites entre ellos.»

STS, de 12 de junio de 1995.

**a) El derecho al honor (\*).** El honor es un concepto que depende de las circunstancias sociales y culturales de cada momento, y que está vinculado a la reputación o la fama que cada persona tiene ante los otros. El Tribunal Constitucional ha ido perfilando con la jurisprudencia unos criterios que hay que tener en cuenta para perfilar este concepto, como por ejemplo la forma y el contexto en el que se enmarcan, la inexistencia de deshonor por los actos, el prestigio profesional o social, o los actos contrarios a la dignidad humana.



«[...] la buena reputación de una persona, protegiéndola frente a expresiones o mensajes que le hagan desmerecer en la consideración ajena al ir en su descrédito o menosprecio o que sean tenidas en el concepto público por afrentosas.»

STC 180/1999, de 11 de octubre.

La LOPC considera intromisiones ilegítimas en el derecho al honor «la imputación de hechos o la manifestación de juicios de valor mediante acciones o expresiones que de cualquier manera lesionen la dignidad de otra persona, menoscabando su fama, o atentando contra su estimación» (art. 7.7 LOPC). Por eso, cuando una noticia impute a una persona de un hecho denigrante o difamante, se considerará ilegítima si es falsa, aunque sea de interés público, o verídica. Por otro lado, si las manifestaciones llevadas a cabo en el ejercicio de la libertad de expresión son innecesarias para la exposición de las ideas, degradan la reputación de una persona en cuestiones que no tienen interés público, o simplemente son insultos o injurias, también se pueden considerar como ilegítimas.

«La Constitución no veda, en cualesquiera circunstancias, el uso de expresiones hirientes, molestas o desabridas, pero de la protección constitucional que otorga el artículo 20. 1.<sup>a</sup> CE están excluidas las expresiones absolutamente vejatorias; es decir, aquellas que, dadas las concretas circunstancias del caso, y al margen de su veracidad o inveracidad, sean ofensivas u oprobiosas y resulten impertinentes para expresar las opiniones o informaciones de que se trate.»

STC 278/2005, de 7 de noviembre.

«La Constitución no reconoce un pretendido derecho al insulto, que sería por lo demás incompatible con la dignidad de la persona que se proclama en el artículo 10.1 del texto fundamental.»

STC 105/1990, de 6 de junio.

**b)** El derecho a la intimidad. La jurisprudencia del Tribunal Constitucional también ha perfilado el concepto de intimidad, que tiene por objetivo:

«[...] garantizar al individuo un ámbito reservado de su vida vinculado con el respeto de su dignidad como persona, frente a la acción y el conocimiento de los demás, sean estos poderes públicos o simples particulares [...]. El derecho a la intimidad garantiza el secreto sobre nuestra propia esfera de vida personal y, por tanto, veda que sean los terceros particulares o poderes públicos quienes decidan cuáles son los contornos de nuestra vida privada.»

STC 83/2002, de 22 de abril.

Pueden considerarse actos de intromisión en la intimidad el uso de aparatos de escucha, de filmación o de dispositivos ópticos que permitan grabar, reproducir o conocer la vida íntima de las personas, y la divulgación de datos relativos a la vida privada de las personas cuando esto afecte a su reputación o buen nombre, o la revelación de datos conocidos en el ejercicio de la actividad profesional.

La intimidad también es objeto de protección penal mediante delitos como el de descubrimiento o el de revelación de secretos (art. 197 CP (\*)).

**c)** Derecho a la imagen. El derecho a la imagen tiene un alcance más amplio que el derecho a la intimidad, aunque, a veces, puede estar vinculado al mismo.

«No cabe desconocer que mediante la captación y publicación de la imagen de una persona puede vulnerarse tanto su derecho al honor, como su derecho a la intimidad. Sin embargo, lo específico del derecho a la propia imagen es la protección frente a las reproducciones de la misma que, afectando a la esfera personal de su titular, no lesionan su buen nombre ni dan a conocer su vida íntima. El derecho a la propia imagen pretende salvaguardar un ámbito propio y reservado, aunque no íntimo, frente a la acción y conocimiento de los demás; un ámbito necesario para poder decidir libremente el desarrollo de la propia personalidad y, en definitiva, un ámbito necesario según las pautas de nuestra cultura para mantener una calidad mínima de vida humana.»

STC 81/2001, de 26 de marzo.

Este derecho faculta a las personas a decidir sobre la difusión pública de su imagen, a impedir que una tercera persona pueda captar, reproducir o difundirla y, por lo tanto, hay que considerar como formas de intromisión cualquier captación, reproducción o difusión mediante una fotografía, un vídeo o cualquier otro medio, de la imagen de una persona en lugares o momentos de la vida privada, o incluso fuera de ella, excepto en los casos previstos en la LOPC.

En particular, esta ley orgánica prevé que se trate de personas que ejerzan un cargo público o una profesión de notoriedad o proyección pública, y la imagen se capte durante un acto público o en lugares abiertos al público, la utilización de la caricatura de estas personas o información gráfica sobre un suceso o hecho público, cuando la imagen de una persona determinada aparezca como meramente accesoria.

### 2.2.4.3. Los límites legales de las libertades de información y de expresión

Como ya se ha avanzado, además de la protección de otros derechos fundamentales, la Constitución española también reconoce que la libertad de expresión y la de información tienen como límite otros derechos e intereses definidos por las leyes.

#### a) Protección del menor

La Ley orgánica 1/1996, de 15 de enero, de protección jurídica del menor dispone que prevalecerá el interés superior de los menores por encima de cualquier otro interés legítimo que pueda concurrir (art. 2.1). Esto se traduce, entre otras medidas, en la intervención del ministerio fiscal en la difusión de información, la utilización de imágenes o nombres de menores en medios de comunicación que puedan implicar una intromisión ilegítima en su intimidad, honra o reputación, o sea contraria a sus intereses, incluso cuando haya el consentimiento del menor o de los representantes legales (art. 4).

#### b) Derecho de rectificación

La Ley orgánica 2/1984, de 26 de marzo, reguladora del derecho de rectificación dispone que toda persona tiene derecho a rectificar la información difundida por cualquier medio de comunicación y relativa a hechos que tengan relación con la misma, cuando considere que es inexacta o que la divulgación puede causarle un perjuicio.

Este derecho se ejerce mediante un escrito de rectificación que hay que enviar al director del medio de comunicación en el plazo de siete días desde la publicación de la información, un escrito de rectificación que tendrá que ser publicado en el plazo de tres días. Si en este plazo no se publica la rectificación, se podrá ejercer una acción de rectificación ante el juzgado de primera instancia.

“ «La inserción de la réplica solo procede en la medida en que se pretenden rectificar hechos y no opiniones, y cuando los hechos publicados afectan perjudicialmente a los intereses del demandante aludido por la información.»

STC 168/1986

#### c) Los secretos oficiales

La Ley 9/1968, de 5 de abril, sobre secretos oficiales dispone que pueden ser declarados materias clasificadas los asuntos, actas, documentos, informaciones, datos y objetos cuyo conocimiento, por personas no autorizadas, puede perjudicar o poner en riesgo la seguridad y la defensa del Estado (art. 2 LSO).

La declaración de una materia como clasificada implica que solo podrá ser conocida por los órganos y las personas debidamente facultados. Además, también determina la prohibición de que las personas que no estén autorizadas puedan acceder o circular por los lugares o las zonas donde haya materias clasificadas (art. 8 LSO).

La LSO establece que la persona que tenga conocimiento de una materia clasificada está obligada a mantenerla en secreto y entregarla a la autoridad correspondiente. Cuando se considere que una materia clasificada puede llegar a conocimiento de un medio de comunicación, se le notificará la circunstancia a los efectos oportunos (art. 9 LSO).

Finalmente, hay que recordar que el Código penal tipifica como delito la revelación de secretos oficiales (art. 13 LSO, en relación con el artículo del CP).

## 2. Aspectos jurídicos del periodismo de datos

### 2.3. El ejercicio de la profesión periodística y la autorregulación deontológica

El ejercicio de la profesión periodística está regulado por diferentes normas.

En primer lugar, el ejercicio de la profesión periodística está regulado por la Constitución española y por varias normas que la regulan. La Constitución prevé que la ley regulará el derecho a la cláusula de conciencia, y también al secreto profesional en el ejercicio de la libertad de información (art. 20.1.d CE).

En cuanto a la cláusula de conciencia, la Ley orgánica 2/1997, de 19 de junio, reconoce que es un derecho constitucional de los profesionales de la información que tiene por objeto garantizar la independencia en el ejercicio de la actividad profesional (art. 1 Ley orgánica 2/1997). De acuerdo con la Ley orgánica, los periodistas pueden solicitar la rescisión de los contratos con las empresas para las que trabajen en las circunstancias previstas (por cambio sustancial de orientación informativa o de línea ideológica, y por traslado a otro medio que signifique una ruptura de orientación informativa o de línea ideológica). Además, los periodistas pueden negarse a participar en la elaboración de informaciones que sean contrarias a los principios éticos de la comunicación.

“ «No es ocioso recordar cómo la progresiva diferenciación de la libertad de información respecto de la de expresión, a medida que la transmisión de los hechos y noticias ha ido adquiriendo históricamente importancia esencial, supuso no solo el reconocimiento del derecho a la información como garantía de una opinión pública libre en un estado democrático, sino la exigencia de evitar que su ejercicio por parte de las empresas de comunicación, generalizadas como medios de transmisión de noticias, pudiera atentar a la finalidad del derecho o a su ejercicio por parte de aquellos profesionales que prestan servicios en ellas, titulares a su vez de la misma libertad de información. Es respecto a dichos profesionales donde encuentra sentido el reconocimiento del derecho a la cláusula de conciencia como garantía de un espacio propio en el ejercicio de aquella libertad frente a la imposición incondicional del de la empresa de comunicación, esto es, frente a lo que históricamente se designaba como “censura interna de la empresa periodística”. Pero también como forma de asegurar la transmisión de toda la información por el profesional del medio, contribuyendo así a preservar el pluralismo que justifica el reconocimiento del derecho, reforzando las oportunidades de formación de una opinión pública no manipulada y paliando el “efecto silenciador” que, por su propia estructura, puede producir el “mercado” de la comunicación.»

STC 199/1999, de 8 de noviembre.

En cuanto al secreto profesional, es un derecho de los periodistas, no un deber. Es decir, los periodistas no están obligados jurídicamente (\*) a guardar secreto sobre los datos que conozcan durante el ejercicio de la profesión, como sucede, por ejemplo, con los abogados o los médicos. Sin embargo, desde el punto de vista ético no es correcto desvelar la identidad de las fuentes que aportan información comprometida. En cualquier caso, el secreto profesional incluye el derecho a no revelar la identidad de las fuentes de la información.

En segundo lugar, las normas que adoptan los periodistas. En efecto, los periodistas como colectivo adoptan códigos éticos que definen los valores y los principios que rigen el ejercicio de su actividad. Estos códigos pueden ir acompañados de la creación de órganos encargados de velar por su cumplimiento y de impulsar las medidas previstas en el caso de incumplimiento.

#### Código deontológico de la profesión periodística en Cataluña

El código deontológico de la profesión periodística en Cataluña fue adoptado en el año 1992 y ha sido objeto de distintas actualizaciones (la última, en el año 2016).

Está integrado por los criterios siguientes:

1. Informar de manera cuidadosa y precisa.
2. Evitar perjuicios por informaciones sin suficiente fundamento.

- 3.** Rectificar las informaciones incorrectas.
- 4.** Utilizar métodos lícitos y dignos para obtener información.
- 5.** Citar las fuentes y preservar el secreto profesional.
- 6.** Conciliar los derechos individuales con el derecho del público de saber.
- 7.** Evitar el conflicto de intereses.
- 8.** No utilizar en beneficio propio.
- 9.** Respetar el derecho a la privacidad.
- 10.** Salvaguardar la presunción de inocencia.
- 11.** Proteger los derechos de los menores.
- 12.** Respetar la dignidad de las personas y su integridad física y moral.

También forman parte del código varios anexos en los que se incluyen recomendaciones en relación con la manipulación de imágenes, de internet, del plagio, del tratamiento de los conflictos bélicos o armados, del uso del término «ilegales» (referido a personas), y de la cita de nacionalidades y de etnias.

El Consejo de la Información de Cataluña (CIC) es el organismo independiente encargado de velar por el cumplimiento del código deontológico de la profesión periodística en Cataluña. Está integrado por periodistas y representantes de otros sectores profesionales.

## Bibliografía

- Cerrillo i Martínez, A. (2012). «The Reuse of Public Sector Information in Europe and Its Impact on Transparency». *European Law Journal* (vol. 18, núm. 6, págs. 770-792).
- Cerrillo i Martínez, A. (2014). «Los principios de datos abiertos en la legislación española». *Revista de Internet, Derecho y Política* (núm. 19, págs. 62-77).
- Cerrillo i Martínez, A. (2016). «La difusión de información pública como instrumento para la prevención de la corrupción: una aproximación desde la legislación autonómica». *Revista catalana de dret públic* (núm. 52, págs. 67-85).
- Fernández Ramos, S. (2017). «La reclamación ante los órganos de garantía del derecho de acceso a la información pública». *Revista General de Derecho Administrativo* (núm. 45).
- Fernández Ramos, S; Pérez Monguió, J. M. (2014). *Transparencia, acceso a la información pública y buen gobierno. Ley 19/2013, de 9 de diciembre*. Pamplona: Ed. Thomson Reuters Aranzadi.
- Fernández Ramos, S.; Pérez Monguió, J. M. (2017). *El Derecho de acceso a la información pública en España*. Pamplona: Ed. Thomson Reuters Aranzadi.
- Guichot Reina, E. (2014). *Transparencia, Acceso a la Información Pública y Buen Gobierno. Estudio de la Ley 19/2013, de 9 de diciembre*. Madrid: Tecnos.
- Jiménez Asensio, R. (2017). «Instituciones de garantía de la transparencia». *El Cronista del Estado Social y Democrático de Derecho* (núm. 68).
- OECD (2010). «Expert meeting on “Building an open and innovative government for better policies and service delivery”». *OECD Guiding Principles for Open and Inclusive Policy Making*. París.
- VV. AA. (2017). *Los límites al derecho de acceso a la información pública*. Madrid: Instituto Nacional de Administración Pública.

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.1. Las fuentes periodísticas

#### 3.1.1. Introducción

Tenemos que ser tan incansables en la recolección de fuentes potenciales como respecto a la información en sí.

“ «A pesar de que las relaciones con las fuentes son, en su mayoría, pura rutina, meras transacciones en las que todos salen ganando, en ocasiones nos vemos involucrados en un juego de ingenio altamente competitivo con las fuentes para asegurarnos de que llegue a la calle la versión mejor y más completa de la historia, en lugar de la que los favorece.»

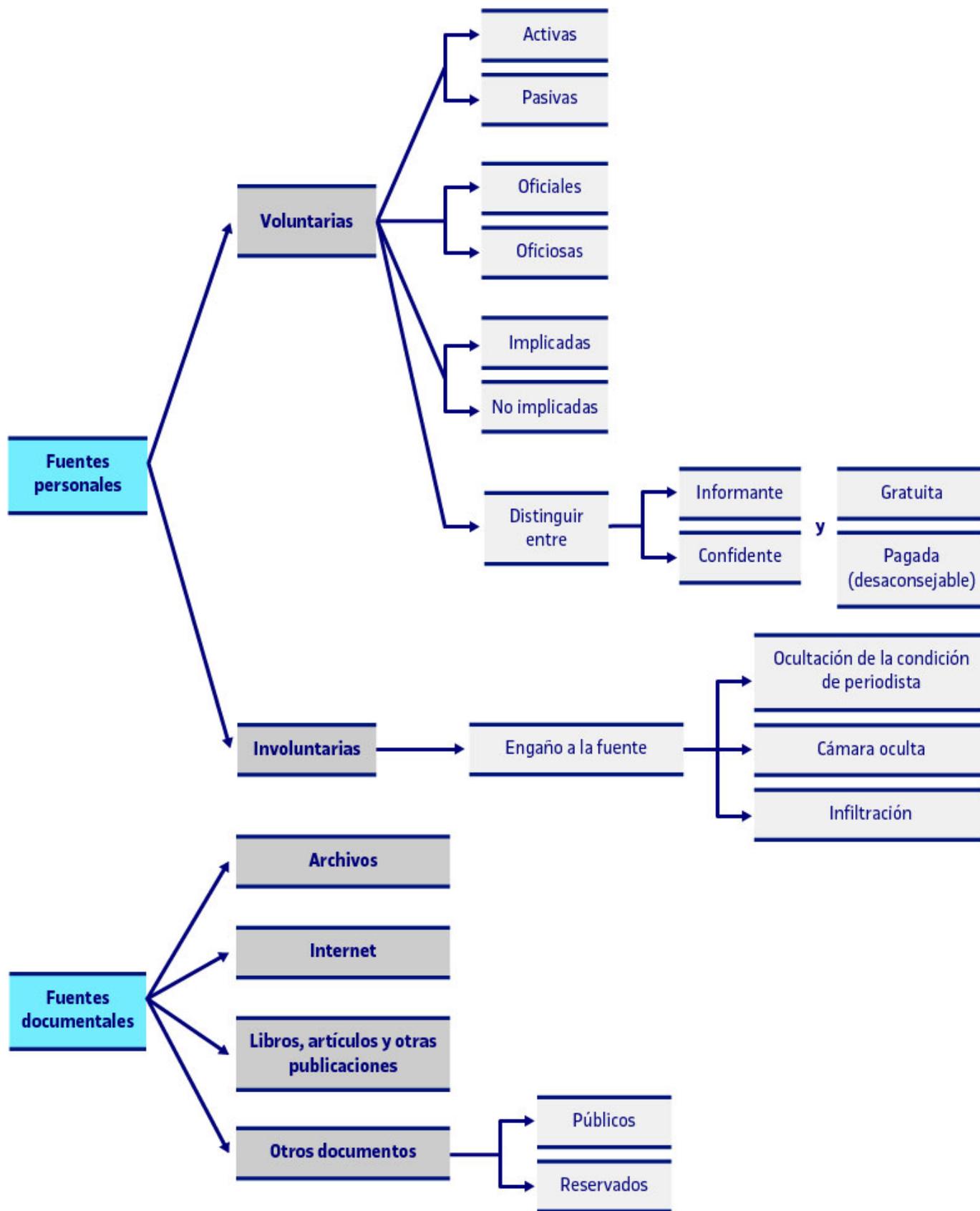
## **3. Fuentes de información y mecanismos de obtención de datos**

### **3.1. Las fuentes periodísticas**

#### 3.1.2. Tipología de las fuentes de información

A continuación, se presenta una clasificación de las fuentes de información. Toda la información sobre la tipología de las fuentes se ha extraído de Javier Chicote (\*).

## Tipología de las fuentes de información en el periodismo



Clasificación de las fuentes de información en el periodismo de investigación  
Fuente: J. Chicote (2006).

### 3.1.2.1. Las fuentes personales voluntarias

Son las fuentes más familiares y lo alcanzan todo, desde el portavoz oficial o los funcionarios y policías a los que vemos con regularidad, hasta la persona con la que hablamos quizá una sola vez. Si hay una regla de oro para el éxito en las investigaciones, es no tener nunca miedo a preguntar. Lo peor que podemos esperar es una negativa y, a cambio, podemos conseguir mucho rédito.

**1) Según su actividad**, la fuente puede ser:

- Activa: toma la iniciativa de ponerse en contacto con el periodista y desea transmitir la información que posee.
- Pasiva: no toma la iniciativa de suministrar la información, sino que es el periodista quien se pone en contacto con ella.
- Oficiales: si se trata de la oficina de prensa o de una voz autorizada por una institución para proporcionar información.
- Oficiosas: también son individuos que ocupan lugares de responsabilidad y que proporcionan información. La principal diferencia respecto a las oficiales es que estas segundas no acostumbran a aceptar que el periodista las cite como fuente, quieren mantenerse en el anonimato u *off the record*. Por norma general, las fuentes oficiales no suelen ser comprobadas por el periodista, mientras que las oficiosas lo acostumbran a ser mucho más.

**2) Según la relación con el hecho** que se quiere investigar:

- Implicadas: testigos o protagonistas de los hechos y, por lo tanto, se ven afectados por los mismos de manera directa. Dentro de esta tipología, encontramos las fuentes favorables, que son aquellas que se ven beneficiadas por la publicación de una investigación sobre los hechos, y las fuentes desfavorables, que son las que se ven perjudicadas si los hechos adquieren una relevancia pública. Tienen que ser comprobadas porque la tendencia será siempre a justificar su actuación o simplemente negarla, en caso de que sea incriminatoria.
- No implicadas/ajenas: si bien estas fuentes no tienen relación directa con los hechos que se quieren investigar, pueden ofrecer datos que enriquezcan la investigación. Por un lado encontramos las fuentes técnicas, que son las que tienen una visión especializada sobre el tema y pueden enriquecer y aclarar la investigación; y por el otro, las fuentes neutrales, que no tienen relación directa con los hechos pero también aportan datos, puesto que pueden conocer a la gente involucrada.

Según Chicote, hay que distinguir entre confidente e informante. El confidente tiene una relación habitual con el periodista producto de un trato personal y prolongado, que genera una comunicación bidireccional, más fluida y abundante. Esta relación se basa en el interés, la amistad, el respeto mutuo, el dinero o cualquier otro motivo. El informante, en cambio, mantiene una relación ocasional con el periodista, a pesar de que también puede ser motivada por dinero o tratarse de una filtración, es decir, una información que a la fuente le interesa propagar.

También hay que distinguir las fuentes de información gratuitas, las que no piden una contraprestación por la información facilitada, y las pagadas, cuyo interés económico puede poner en tela de juicio la fiabilidad de su testigo.

### 3.1.2.2. Las fuentes personales involuntarias

No todas las fuentes personales son voluntarias. Las involuntarias son aquellas que proporcionan información sin saber cómo ni para qué o, lo que es lo mismo, son fuentes a las cuales el periodista engaña para obtener información. Esto pasa siempre que el periodista no se identifica como tal. Un ejemplo de esta clase de fuentes son aquellas que se enfrentan, sin saberlo, a una cámara oculta. Este tipo de prácticas para obtener información son reprobables desde el punto de vista deontológico y ético, y por eso son muy difíciles de justificar. Al respecto, Chicote considera que solo se recurrirá a la cámara oculta «en virtud de un bien mayor. El periodista puede engañar a un traficante de drogas para desenmascarar su organización mafiosa y hacer un bien a la sociedad, pero por el contrario, no es lícito ocultar la condición de informador para descubrir la vida privada de un personaje».

### 3.1.2.3. Las fuentes documentales

Son los archivos públicos o privados, internet, las bibliografías, las hemerotecas y, en definitiva, toda clase de datos a los cuales se pueda acceder y que tengan relación con el tema investigado. Muchos de estos datos, o el acceso a los mismos, pueden ser proporcionados por las fuentes personales, pero en cualquier caso se trata de informaciones generalmente públicas o semipúblicas. Algunas de las plataformas documentales recientes más conocidas son Wikileaks, Contrast o Whistleblower. Aun así, hay una amplitud de recursos y bases de datos en línea en los que obtener información, como ya vimos en el apartado 1.3, «Periodismo de datos».

Por su parte, David Randall establece una serie de pautas generales a la hora de tratar las fuentes (\*); una propuesta complementaria a la clasificación de Javier Chicote.

#### 1) Fuentes humanas

Por norma general, las fuentes se tienen que identificar. Si se niegan, podemos esgrimir el argumento de que la importancia de una noticia depende de que se pueda verificar gracias a que se conoce su procedencia. Si, a pesar de todo, una fuente decide no identificarse, pero nosotros queremos citar sus declaraciones, podemos intentar pactar una fórmula de identificación genérica, como por ejemplo «fuentes del Departamento de Interior». Aun así, el secreto profesional es un derecho del periodista, a la vez que un deber que garantiza la confidencialidad de las fuentes de información. Por lo tanto, el periodista garantizará el derecho de sus fuentes informativas a mantener el anonimato si así lo han solicitado. No obstante, este deber de profesional podrá ceder excepcionalmente en el supuesto de que conste fehacientemente que la fuente ha falseado de manera consciente la información, o cuando revelar la fuente sea el único medio para evitar un daño grave e inminente a las personas.

## 2) Fuentes oficiales

Que se trate de fuentes autorizadas no siempre es sinónimo de que sean fuentes bien informadas cuando son departamentos de relaciones públicas y prensa. Algunas conocen bien sus respectivas organizaciones y están en condiciones de aportar detalles. Otras son simples intermediarios, que transmiten nuestras preguntas a un funcionario y vuelven con respuestas, con la imposibilidad de formular nuevas cuestiones al respecto. Por eso conviene cultivar los contactos con funcionarios dispuestos a darnos información directamente, en lugar de la filtrada a través de su encargado de relaciones públicas. Ha habido muchas lamentaciones por parte de los periodistas respecto a la perversidad de los «expertos en relaciones públicas», conocidos como *spin doctors*, de la «gestión de noticias».

**a) Spin doctors.** Poco a poco, esta expresión ha ido designando a casi todos los que se ocupan de las relaciones públicas o la gestión de la imagen. Un verdadero *spin doctor* es alguien procedente del mundo político, que actúa como portavoz en jefe de prensa o bien un ayudante con suficiente estatus. Además de sus declaraciones rutinarias y conferencias de prensa, se ocupa de interpretar los hechos en favor de su jefe, intenta anticiparse a las malas noticias (y distraer la atención de ellas) y se esfuerza para gestionar las expectativas de forma que beneficien a su causa. Los auténticos *spin doctors* están cerca de su jefe, y todo lo que dicen para publicar posee la mayor autoridad.

**b) Gestión de las noticias.** Las organizaciones tienen una gran cantidad de declaraciones e informaciones que quieren dar a conocer. No es particularmente escandaloso que elijan la manera y el momento más ventajosos para hacerlo. De hecho, es frecuente que esta «gestión de las noticias» sea de ayuda, porque asegura que los periodistas dispongan de tiempo suficiente para redactar la información para sus ediciones o evitar un conflicto entre la publicación de un informe y un acontecimiento de importancia. Sin embargo, precisamente por el hecho de que las organizaciones tienen intereses, pueden actuar de tres maneras:

- Hacer que la hora de la declaración coincida con el peor momento del ciclo informativo. De este modo, es fácil que la convocatoria sea ignorada o tratada con indulgencia.
- Filtrar el aspecto principal a un medio simpatizante, con la esperanza de que, cuando los otros reciban la información a través de los canales oficiales, ya tengan establecida su agenda.
- Endulzar la píldora con algún fragmento de buenas noticias que después se convierte en el argumento principal de la convocatoria.

La única garantía contra esta y otras formas de gestión informativa es utilizar nuestro ingenio.

## 3) Fuentes no autorizadas

Se trata de alguien no autorizado para transmitirnos la información, desde un denunciante bien situado en una organización a una persona sin conexión oficial con la información. Pueden filtrar documentos o, simplemente, sugerirnos que excavemos en un área determinada. La información así obtenida será desmentida, como mínimo en un primer momento, y es casi seguro que nuestras fuentes no querrán que se las mencione. Para tratar con ellas, hay que tener presentes varias cosas:

- Preguntarnos cuáles son sus motivos: están los que buscan perjudicar a un rival político, comercial o personal, impulsar una determinada causa, vengarse o provocar problemas sin más. El deseo de venganza de una persona puede resultar de utilidad, pero es mejor saber qué pretende desde el principio.
- Preguntarnos (¡y averiguar!) la otra cara de la historia: la información que nos dan fuentes polémicas no acostumbra a ser tan simple como quieren que creamos. Ahorraremos tiempo si descubrimos la otra cara, o algún matiz de la información que deberíamos conocer. Todo lo que publicamos se tiene que contrastar previamente, y para explicar una historia de forma veraz es necesario conocer las distintas versiones de las personas que están implicadas.
- Pedir documentos siempre que sea posible: se tiene que pedir a las fuentes si hay alguna documentación que apoye lo que nos están explicando. Si no nos la quieren dar, pediremos fotocopias y, en caso de que también se nieguen, pediremos leer la documentación delante de ellas. Si obtenemos los documentos de una manera extraoficial, es conveniente asesorarnos con fuentes expertas para asegurarnos de su validez.

Randall destaca los siguientes **consejos básicos** (con aportaciones de la autora):

**1) Ser justo con las fuentes:** tratarlas con justicia y no asediarlas, engañarlas o intimidarlas. Lo fundamental es ser justos. Si se está criticando o acusando a una fuente de juego sucio, no solo lo tenemos que informar, sino que también debemos darle la oportunidad de replicar.

**2) Aprobación del texto:** mostrarle a la fuente el artículo acabado antes de su publicación es como invitarla a que lo censure. La tarea del periodista consiste en redactar información precisa, no algo que pueda constituir una base de negociaciones con el implicado. Aun así, unas comprobaciones sensatas no representan una aprobación del texto. De hecho, cuando nuestra historia trata sobre cuestiones con las que no estamos familiarizados, puede ser incluso un ejercicio de responsabilidad enviar parte o la totalidad del artículo a un experto, para garantizar que la historia esté libre de errores. En ocasiones muy concretas, incluso es justificable enviar parcial o totalmente el contenido a una o varias de las fuentes citadas, sobre todo si su relato aborda cuestiones particularmente sensibles. Esta práctica da pie a que las fuentes puedan alterar su testimonio, así que no puede convertirse en la dinámica habitual y se tiene que delimitar. No es obligatorio permitir esta revisión, depende sobre todo de la empatía del periodista con las fuentes utilizadas.

## Ejemplo

Cito un ejemplo. *Crític* publicó un reportaje sobre la violencia sexual hacia las mujeres, en el que aparecía una fuente que había sufrido una agresión de este tipo en primera persona, y estaba dispuesta a explicarlo. A la hora de recoger su testimonio, el medio decidió entrevistarla primero sin grabar su relato, para pactar con la fuente qué información quería que el reportaje incluyera y cuál no. A la hora de transcribir la entrevista y redactar el tema, se le envió a la fuente los fragmentos en los que aparecía su relato de los hechos. Se tuvo esta «deferencia» con la fuente porque se hizo valer la gravedad de su testimonio y las dificultades añadidas de expresarlo en un medio. La voluntad del medio era asegurarse de que el testigo se sentía cómodo con la manera en que se recogían sus palabras. ¿Podría *Crític* haber prescindido de enviarle estos fragmentos de información a la fuente? Sí, nuestra autoridad como periodistas nos legitima. Por lo tanto, como apuntábamos arriba, depende de la empatía, la sensibilidad y el compromiso del periodista con la fuente.

**3) Retracción:** muy a menudo, una fuente dice una cosa a un periodista y a continuación, lamenta haberlo dicho. En estas ocasiones, debemos permitir que corrijan sus declaraciones, a menos que tengamos excelentes motivos para pensar que la voluntad de la fuente es obstruir la investigación. Si es así, dejaremos constancia tanto de su primera declaración como de la segunda, con recursos del tipo: «En una primera entrevista, la persona X explicaba a este medio “la primera versión”, pero también matiza que se retracta diciendo que “la segunda versión”». Así nos curamos en salud, sin faltar a la verdad. En última instancia, si no tenemos claro si publicar la primera versión o dejarla fuera del texto, lo consultaremos con el editor o el director del medio.

Ahora bien, como periodistas, siempre tenemos que respetar el *off the record*, a pesar de que conviene evitarlo tanto como sea posible. Además, la norma que hay que seguir es que una información solo se considerará *off the record* cuando las dos partes lo hayan acordado así desde el principio. Hay muy pocas ocasiones en las que esté justificado publicar una información cuando la fuente ha precisado que nos la explica fuera de la entrevista. Por otro lado, es necesario saber utilizar el *off the record* para tirar del hilo. A veces, las fuentes nos revelan una información «fuera de grabación» que nos puede servir para iniciar una investigación, a pesar de que no citemos a la persona que nos ha dado esta información.

**4) Tentativa de impedir el acceso:** a veces, la cobertura de la prensa disgusta tanto a las organizaciones que nos cortan todo acceso a la información. Se tendría que informar a los lectores de cualquier intento de manipulación y fomentar la respuesta de otros medios.

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.2. Introducción: los datos como fuente de información

Un conjunto de datos o un *data set* –el término anglosajón para denominarlo– es una compilación de información estructurada en columnas y filas, que puede ser procesada y analizada. Una fuente periodística, más que cualquier otra, puede revelar información clave y contexto sobre una realidad concreta que queramos cubrir periodísticamente. Lo más importante es hacer las preguntas adecuadas y conocer las herramientas que permiten «entrevistar» una hoja de cálculo para que las conclusiones extraídas puedan presentar un hecho noticiable.

Periodismo de datos o datos con periodismo. El escenario tecnológico ha abierto un abanico muy amplio de posibilidades de fuentes de información que hace treinta años era impensable tener al alcance. Hablamos de datos almacenados en hojas de cálculo, y de datos que, a pesar de no estar ordenados en filas y columnas, procesándolos en una hoja de cálculo se convierten en un material potencialmente publicable.

El periodismo de datos es el periodismo de precisión de toda la vida, que en el contexto actual se sirve de unas herramientas informáticas de análisis y de visualización que permiten analizar y extraer conclusiones de un volumen de datos que hace un tiempo hubiera sido inimaginable, así como poder mostrarlos al lector.

La delincuencia en una ciudad, las hectáreas de un bosque quemado, las epidemias, las votaciones parlamentarias y el resultado de unas elecciones posiblemente no son más que números. Es cierto, pero también son cifras que hablan de la realidad política, social y económica que nos rodea, y que como periodistas queremos explicar en forma de historias en un diario, televisión o radio.

## **3. Fuentes de información y mecanismos de obtención de datos**

### **3.3. Bases de datos disponibles**

#### **3.3.1. Introducción**

Antes de empezar a abordar cualquier tema, es recomendable preguntarse qué datos hay disponibles sobre la cuestión que interesa cubrir. Conocer las fuentes públicas disponibles facilitará y acelerará muchísimo el proceso de búsqueda de las fuentes.

La cantidad de información pública y disponible en formato libre no ha dejado de crecer en los últimos años. Este crecimiento ha tenido mucha relación con el esfuerzo que se ha hecho desde diferentes organizaciones en favor de la transparencia y la apertura de datos (*open data*).

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.3. Bases de datos disponibles

#### 3.3.2. El movimiento *Open Data*

##### 3.3.2.1. Introducción

El concepto de datos abiertos es una práctica que busca que la información pública esté disponible y en formatos abiertos o reutilizables (Excel o CSV, entre otros). Sin restricciones de autor, libre de patentes y sin ningún tipo de mecanismo de control, parte de una base similar de la que surgen otros movimientos tecnológicos como *Open Source* ('código abierto') y el software libre.

Tanto en Cataluña como a escala estatal, el movimiento *Open Data* ha sido integrado, entre otros perfiles profesionales, por periodistas, que han visto en los datos abiertos amplísimas posibilidades periodísticas. También han sido promotores de este cambio perfiles como los ingenieros, las entidades públicas, los desarrolladores y los investigadores.

En este sentido, las administraciones públicas han tenido que correr para hacer la renovación de formatos y poner a disposición de la ciudadanía, y en formatos abiertos, la información pública. En los últimos años, la información disponible ha aumentado de forma exponencial tanto en el plano estatal como autonómico y municipal. Y, a pesar de que todavía queda mucho trabajo para hacer en todos los estratos de las instituciones, actualmente ya hay repositorios públicos de datos que merece la pena tener en cuenta.

##### 3.3.2.2. Cataluña

###### 1) Datos abiertos Gencat

Es el portal de la Generalitat de Cataluña, gestionado desde la Secretaría de Transparencia y Gobierno Abierto del Departamento de Asuntos y Relaciones Institucionales y Exteriores y Transparencia, en el cual están publicados los datos de carácter público, con el objetivo de fomentar el uso y la reutilización de la información proveniente de la Administración. Los datos provienen de diferentes organismos públicos de la Generalitat, están agrupados en el catálogo de datos y disponibles en diferentes formatos, la mayoría de los cuales son estándar, de forma que se pueden reutilizar fácilmente.

###### 2) Datos abiertos de la Diputación de Barcelona

En este portal, la Diputación de Barcelona ofrece un catálogo de datos públicos susceptibles de ser reutilizados por la ciudadanía y las empresas. Trata desde cuestiones más municipales hasta contratación en el ámbito de la diputación, y es una magnífica compilación de lo que ocurre en el plano de la provincia de Barcelona.

###### 3) Transparencia de la Generalitat de Cataluña

Portal que recoge todos los documentos relacionados con el ejercicio de la transparencia en los diferentes departamentos de la Generalitat. Para una correcta utilización, es preciso navegar por las páginas de los departamentos para ver qué datos hay publicados, y cómo se estructuran en cada caso.

###### 4) Instituto de Estadística de Cataluña (IDESCAT)

Recopilación de datos estadísticos de Cataluña, sobre todo del ámbito municipal. Hay muchos más datos que en el Instituto Nacional de Estadística (INE). Es necesario destacar que los datos municipales del IDESCAT son mucho más detallados que los que hay en el INE. Por ejemplo, la aplicación «Cataluña en Cifras» permite desglosar muchas variables del ámbito municipal.

###### 5) Centro de Estudios de Opinión

Recopilación de los datos de los estudios de opinión en Cataluña. No tiene nada que ver con el homónimo español, el Centro de Investigaciones Sociológicas. No comparten estudios ni variables, y sirve únicamente en caso de que el objetivo periodístico se centre en territorio catalán.

###### 6) Open Data Barcelona

Recopilación de datos abiertos del Ayuntamiento de Barcelona. En los últimos meses, el portal ha renovado mucho el conjunto de datos abiertos, los ha actualizado y ha cargado otros nuevos.

###### 7) Datos abiertos Gerona

Recopilación de datos abiertos del Ayuntamiento de Gerona, que trata desde el número de paradas de taxis hasta el alumbrado público. La cantidad de ámbitos que se abordan es innumerable.

### 3.3.2.3. España

En el plano estatal debemos tener en cuenta que, para todas aquellas competencias descentralizadas a las comunidades autónomas, no hay datos centralizados y homogeneizados. Esto puede hacer que, en algunos casos, haya vacíos de información.

Es el caso, por ejemplo, de los datos de ayuda a la dependencia. Cada comunidad autónoma es la encargada de poner a disposición del ciudadano la información respecto a las ayudas. Si queremos llevar a cabo un trabajo periodístico sobre la aplicación de la Ley de la dependencia, nos encontraremos con el hecho de que no solo tenemos que ir a buscar los datos en diecisiete fuentes distintas, sino que, además, estos están organizados de diecisiete maneras diferentes y que, muchas veces, ni siquiera están homogeneizados.

De todos modos, salvo los ámbitos que son competencia de las comunidades autónomas, el Estado tiene a su disposición una cantidad muy extensa de información y de datos que nos puede ser muy útil como periodistas.

#### **1) Gobierno Abierto de España**

Conjunto de datos abiertos de los diferentes ministerios, municipios e instituciones municipales y autonómicas. En este portal se cargan de forma automática los *datasets* públicos de las diferentes administraciones del Estado. Tiene el inconveniente de que los filtros de carga son lentos, pero sirve para situar cuántos datos hay sobre un tema concreto e ir hasta la fuente original.

#### **2) Instituto Nacional de Estadística**

Recopilación de datos estadísticos de España. Desde variables económicas, demográficas hasta estilos de vida. En el ámbito periodístico son especialmente interesantes los «microdatos», desglosados por municipio, que permiten profundizar mucho en el análisis. En general, es fácil que cualquier tema que se quiera abordar tenga datos disponibles en el INE y, por lo tanto, está bien darles un vistazo antes de ponerse a escribir.

#### **3) Ministerio del Interior**

Conjunto de datos estadísticos que son competencia de este ministerio: instituciones penitenciarias, tráfico, asilo, protección civil, terrorismo y seguridad ciudadana. La publicación de los datos va con un año de atraso.

#### **4) Ministerio de Sanidad, Servicios Sociales e Igualdad**

Conjunto de datos estadísticos que son competencia de este ministerio: violencia machista, inversión en sanidad, recursos para la dependencia. Hay que tener en cuenta, de todos modos, que tanto en el plano sanitario como de servicios sociales las competencias están descentralizadas a las comunidades y, por lo tanto, es posible que haya una falta de información homogeneizada de toda España.

#### **5) Ministerio de Defensa**

Conjunto de datos estadísticos respecto a todo el apartado militar del Estado y política de defensa. Están ligeramente desactualizados.

#### **6) Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente**

Conjunto de datos estadísticos que son competencia de este ministerio: pesca, agricultura, desarrollo rural, ganadería, observatorio de precios, barómetro del clima, consumo alimentario, industrias y medio ambiente.

#### **7) Ministerio de Economía, Industria y Competitividad**

Datos macroeconómicos y recopilación de otros portales de interés relacionados con las competencias de este ministerio.

#### **8) Ministerio de Fomento**

Conjunto de datos estadísticos que son competencia de este ministerio: construcción, movilidad, vivienda, actuaciones urbanísticas, transporte e infraestructuras estatales.

#### **9) Ministerio de Educación, Cultura y Deporte**

Conjunto de datos estadísticos que son competencia de este ministerio: enseñanza universitaria, educación primaria y secundaria, museos, teatros, cines, deportes y competiciones.

### **10) Ministerio de Energía, Turismo y Agenda Digital**

Conjunto de datos estadísticos que son competencia de este ministerio: principales indicadores coyunturales como la producción industrial, el consumo de energía y el índice de precios al consumo (IPC).

### **11) Poder judicial**

Recopilación de datos estadísticos del poder judicial: actividades de los órganos, resoluciones de los diferentes tribunales, datos estadísticos de los juzgados penales, civiles y laborales.

### **12) Recaudación y estadísticas del sistema tributario español**

Conjunto de datos estadísticos del sistema tributario: recaudación del Estado con detalle de cada comunidad, indicadores fiscales, estadísticas sobre el impuesto sobre la renta de las personas físicas (IRPF) y estadísticas catastrales.

### **13) DataComex**

Recopilación de datos estadísticos del comercio exterior del Estado: tipo de mercancías, comercio entre países. Datos detallados por país, mes, año, entrega y tipo de producto.

### **14) Banco de España**

Recopilación de datos estadísticos del Banco de España: indicadores económicos, fondos de inversión, datos históricos relacionados con el sector financiero, balances, encuesta financiera de las familias, préstamos bancarios y estadísticas del sector exterior.

### **15) Centro de Investigaciones Sociológicas (CIS)**

Recopilación de la evolución del barómetro de opinión del Estado, y estudios concretos sobre temas específicos que tienen relevancia en un momento determinado.

## **3.3.2.4. La Unión Europea**

### **1) Eurostat**

Datos homogeneizados del conjunto de países de la Unión Europea referentes a cualquier ámbito: economía, política exterior, medio ambiente, producción industrial, condiciones de vida, agricultura, transporte y demografía, entre otros muchos aspectos. Es una fuente imprescindible a la hora de tomar referencias sobre un tema.

### **2) Datos económicos de la Comisión Europea**

Datos homogeneizados sobre la economía de la Unión Europea: estadísticas macroeconómicas, financieras, principales indicadores de la zona euro, índices monetarios, evolución de precios y competitividad.

### **3) Datos de salud pública de la Comisión Europea**

Datos sobre indicadores de salud pública a escala europea, desarrollo sostenible, lesiones, demografía, alimentación y condiciones de vida.

## **3.3.2.5. Datos globales**

### **1) Las Naciones Unidas**

#### **a) Alto Comisionado de las Naciones Unidas para los Refugiados (ACNUR)**

La mejor base de datos actual sobre los movimientos migratorios, solicitudes de asilo aceptadas, denegadas, en espera, y todo desglosado por país de entrada y de petición. Incluye también información estadística detallada sobre los lugares que se encuentran en estado de emergencia.

#### **b) UN Comtrade**

Acceso gratuito a datos detallados del comercio mundial. El UN Comtrade es un depósito de las estadísticas oficiales del comercio internacional y las tablas analíticas relevantes. Todos los datos son accesibles por medio de una aplicación propia.

#### **c) Estadísticas de género**

Pretende ser una guía nacional e internacional de estadísticas de género. Es un conjunto de 52 indicadores cuantitativos y 11 variables cualitativas sobre temas relevantes relacionados con la igualdad de género y el empoderamiento femenino.

#### **d) Objetivos para el desarrollo sostenible**

Esta plataforma proporciona acceso a datos recopilados por el sistema de las Naciones Unidas.

#### **e) UN Data**

Recopilatorio de datos de todos los temas que se tratan en las Naciones Unidas. La página web está un poco desactualizada, y esto puede dificultar la búsqueda de documentos.

#### **f) FAOSTAT Data**

Recopilatorio de datos que utiliza Naciones Unidas, destinado a la alimentación y a la agricultura: comercio de alimentos, producción, malnutrición, evolución de los precios y hambre, entre muchos otros datos.

#### **g) AIDSinfo**

Recopilación de datos del impacto del sida en todo el mundo, con detalle de cada país, que incluye mapas, datos sectorializados por sexo, edad y condiciones socioeconómicas. También permite el análisis en un plano interestatal.

#### **h) Food Security**

Los datos de clasificación de la seguridad alimentaria de FEWS NET, que se remontan al mes de junio del 2009, están disponibles para la descarga como ficheros e imágenes de formularios geográficos regionales. Estos datos se publican cada cuatro meses.

### **2) Organizaciones internacionales**

#### **a) World Bank Open Data**

La iniciativa Open Data del Banco Mundial tiene como objetivo proporcionar a todos los usuarios acceso a datos del banco. El catálogo de datos es una lista de conjuntos de datos disponibles del Banco Mundial, que incluyen bases de datos, tablas preformatadas, informes y otros recursos.

#### **b) CIA World Factbook**

La sección de la página de la CIA, The World Factbook, proporciona información sobre temas relacionados con la historia, la gente, el Gobierno, la economía, la geografía, las comunicaciones, el transporte, los asuntos militares y los asuntos transnacionales para 267 entidades mundiales.

#### **c) Our World in Data**

Página web de la Universidad de Oxford que pretende mostrar cómo las condiciones de vida están cambiando. Tiene como objetivo dar una visión global, y mostrar cambios a largo plazo. Proporciona tanto visualizaciones de datos como conjuntos de datos reutilizables.

#### **d) The Humanitarian Data Exchange**

La plataforma de intercambio de información humanitaria (HDX) es una plataforma abierta para compartir datos que se inició en julio del 2014. El objetivo de la HDX es hacer que los datos humanitarios sean fáciles de encontrar y de utilizar para el análisis. La colección de datos ha recibido el acceso de usuarios de más de 200 países y territorios.

#### **e) Fondo Monetario Internacional**

Datos financieros de la gran mayoría de los países: créditos concedidos por parte del Fondo Monetario Internacional y la situación fiscal, entre otros muchos datos macroeconómicos.

### **3) Bases de datos por ámbitos**

#### **a) Flight Radar**

El Flightradar24 es un servicio global de seguimiento de vuelos que proporciona información en tiempo real e histórico sobre miles de aviones por todo el mundo.

#### **b) Aviation Safety Network**

Base de datos que acumula información sobre todos los accidentes de avión que ha habido. Se actualiza al momento y es muy completa, con un Excel para cada año desde 1919.

#### **c) Marine Traffic Database**

Visualizaciones y datos interactivos sobre el tráfico de barcos en todo el mundo. Información que se puede desglosar por puerto, derramamientos y datos históricos detallados.

#### **d) Historical Index of Economic Liberty (HIEL)**

Conjunto de datos históricos sobre economía desglosados por año, por región y por país. Los datos se pueden descargar en formato reutilizable y la recogida cuenta con una metodología muy rigurosa.

#### **e) Consortium of Spatial Information**

Información y mapas sobre climatología global del planeta Tierra. Contiene datos históricos y con formatos que permiten la visualización con herramientas de tratamiento cartográfico.

#### **f) Code for Africa – Source Africa**

Source Africa es el repositorio principal de África de «documentos procesables», es decir, documentos que el público necesita para ayudar a tomar decisiones más informadas y que pretende ayudar a los organizadores civiles, medios de comunicación, ONG u organizaciones de justicia social a demostrar su reportaje. También incluye documentos que han sido suprimidos o censurados en otros lugares. Es una plataforma gestionada por Code for Africa.

#### **g) TeleGeography**

Es una empresa privada que se dedica a generar métricas detalladas que no están disponibles en ninguna otra fuente. También proporcionan datos abiertos geolocalizados de la red de telecomunicaciones global y de la industria, en general.

#### **h) Google Trends**

Tendencias de búsquedas en línea en Google segmentadas a escala global, por país y por lengua. También permite buscar una palabra concreta y ver cuál ha sido el número de búsquedas en el buscador.

#### **i) Wayback Machine**

Es una librería abierta formada por miles de libros, películas, software libre, música y páginas web.

#### **j) SIPRI Military Expenditure Database**

Se trata de una base de datos pública que contiene el gasto militar de todos los estados durante el periodo comprendido entre los años 1949 y 2016.

#### **k) Global Terrorism Database**

Es una base de datos de libre acceso en la cual se registran todos los atentados en todo el mundo. Para cada uno, se desglosa la información sobre la localización, el número de víctimas, la organización perpetradora y la fecha, entre otros detalles.

#### **l) Foreign Trade – U.S.**

Esta página web registra todo el comercio exterior de Estados Unidos. Se trata de una base de datos de acceso libre que permite filtrar la información por país y extraerla en el formato deseado.

#### **m) CERAC**

En el apartado de recursos de esta página web, encontramos un listado de bases de datos de los conflictos que hay por todo el mundo, con los enlaces correspondientes.

#### **n) SICE**

Datos sobre comercio entre los países que conforman la Organización de los Estados Americanos (OEA). Incluye acuerdos de comercio y volumen segmentado por países y tipos de producto.

#### **o) Asian Development Bank**

Recopilación de estadísticas relacionadas con el desarrollo de los países asiáticos: datos económicos, sociales e indicadores financieros relacionados con el trabajo del Banco Asiático de Desarrollo.

**p) Inter-American Development Bank**

## **3. Fuentes de información y mecanismos de obtención de datos**

### **3.4. La Ley de transparencia**

#### **3.4.1. Introducción**

Si ninguna de las fuentes públicas disponibles reúne los datos deseados, siempre queda la alternativa de formular una petición de información pública.

Se trata de un mecanismo que dispone el Estado para pedir acceso a cualquier información pública que esté en manos de las instituciones. Es un mecanismo que en Cataluña se rige por la Ley 19/2014, de 29 de diciembre, de transparencia, acceso a la información pública y buen gobierno, y en España por la Ley 19/2013, de 9 de diciembre.

Las dos leyes consideran información pública todos aquellos contenidos o documentos, sea cual sea el formato o soporte, que están en manos de las administraciones públicas.

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.4. La Ley de transparencia

#### 3.4.2. La Ley de transparencia en España y en Cataluña

La Generalitat de Cataluña se añadió a la transparencia un año después de la entrada en vigor de la Ley de transparencia y buen gobierno de España, la LTBG, que entró en vigor en diciembre del 2013. España llegó tarde, tan tarde, que cuando se decidió a hacerlo era el único país europeo de más de un millón de habitantes que aún no tenía una ley que reconociera el derecho de acceso a la información.

La presión ejercida por parte de la Coalición ProAcceso, una organización formada por unas sesenta organizaciones y actores de la sociedad civil, fue fundamental para que finalmente el gobierno popular empezase a redactar el proyecto de ley.

Actualmente, tanto la legislación catalana como la española tienen algunas carencias. En concreto, una que afecta a lo más básico: España no reconoce el derecho a la información pública como un derecho fundamental.

Por lo tanto, para formular una petición, tanto si es en el portal de transparencia catalán como en el del Estado, la persona tiene que identificarse previamente como ciudadano. En el caso de Cataluña, se puede hacer mediante el certificado digital –la firma electrónica estatal– o con la contraseña IdCAT móvil.

Tanto el uno como el otro tienen que conseguirse de forma presencial, yendo, en el primer caso, a las dependencias de la Administración general del Estado y, en el segundo, a una de las 38 oficinas acreditadas por la Generalitat.

A pesar de que no hay datos disponibles sobre qué porcentaje de la población dispone de certificado electrónico ni de IdCAT móvil, la complejidad de la identificación choca con uno de los fundamentos del derecho de acceso a la información: hacer peticiones tendría que ser un proceso sencillo, rápido y gratuito. Gratuito lo es, pero no es ni sencillo ni rápido, uno de los principios básicos que la organización ProAcceso reclama para el correcto funcionamiento de la transparencia.

Una vez hecha la petición, la Administración competente tiene un plazo máximo de 30 días para responder, que en el caso del conjunto estatal se puede atrasar hasta 15 días más.

Una de las grandes ventajas que ofrece la ley catalana respecto a la española es el silencio administrativo positivo. Si la Administración pública catalana no responde al ciudadano, automáticamente se le tienen que facilitar los datos requeridos. En cambio, en el caso de la legislación española, este silencio se interpretará automáticamente como una denegación.

Si la petición de información es denegada o los datos obtenidos no son satisfactorios, se puede recurrir a un órgano independiente que evalúe de forma externa si se tiene que facilitar, o no, la información requerida. En este sentido, la ley catalana vuelve a ser más favorable para el ciudadano.

La Comisión de Garantía del Derecho de Acceso a la Información Pública (GAIP), el órgano que vela por el cumplimiento del derecho y el encargado de resolver todas las reclamaciones, es elegida por el pleno del Parlamento, y los miembros tienen dedicación exclusiva para llevar a cabo esta tarea. A diferencia del Consejo de Transparencia y Buen Gobierno, el organismo homólogo estatal, los miembros ejercen otras funciones más allá de la pertenencia a este ente, y son elegidos por el Gobierno.

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.4. La Ley de transparencia

#### 3.4.3. Cómo formular una petición de información

##### 3.4.3.1. Concreción

Para que una solicitud se pueda resolver con éxito, es imprescindible ser muy concretos a la hora de pedir la información: qué datos concretos se piden, las variables determinadas, durante qué periodo de tiempo y para qué ámbito geográfico.

#### Ejemplo

Si pedimos el número de accidentes en las carreteras catalanas desde el año 2000 hasta el 2017, es muy fácil que nos acaben entregando una hoja de cálculo con un número de accidentes por año. Por lo tanto, si se pretende ir un paso más allá en el análisis, habrá que especificar exactamente qué información queremos de cada accidente, como por ejemplo, la localización (latitud y longitud), el número de muertos, el número de heridos, el número de vehículos accidentados y la hora del accidente. También es importante concretar tanto para qué periodo de tiempo requerimos la información, como el ámbito territorial en el que estamos interesados.

Otro elemento que puede facilitar todavía más la tarea es ajustar la petición a la forma en que la institución pública tiene almacenados los datos. Las notas de prensa, los anuarios y los comunicados dan pistas de la manera en que cada institución tiene almacenadas y estructuradas las bases de datos. Dar un vistazo a toda la documentación antes de formular la petición puede ayudar a acertar el tiro y aumentar las posibilidades de éxito.

##### 3.4.3.2. Formato

Del mismo modo, también es imprescindible especificar el formato mediante el cual se desea recibir los datos. Las administraciones públicas, a menudo, están acostumbradas a trabajar con formatos cerrados y no reutilizables como por ejemplo PDF o PDF de imagen. Estos tipos de formatos dificultan y retrasan la tarea periodística de analizar los datos y extraer conclusiones, y por eso es del todo recomendable especificar el formato en la petición de información.

##### 3.4.3.3. Basada en la ley

A pesar de que en ningún caso hay que justificar el motivo por el que pedimos la información, sí va bien mencionar el artículo de la ley que engloba que los datos que pedimos están amparados por el derecho de acceso a la información pública. A pesar de que no es obligatorio, el conocimiento de la ley puede aumentar las posibilidades de éxito de la petición.

#### Ejemplo

En el caso de la ley estatal, la petición se podría formular de la manera siguiente:

«Se trata de una petición que pretende conseguir bajo el amparo de la ley estatal la evolución del precio del transporte público en la ciudad de Sevilla.»

En el caso de la legislación catalana, un buen ejemplo podría ser:

«Se trata de una solicitud de información dirigida al Departamento de Salud, que pretende obtener el registro de contratación pública de material sanitario de los años anteriores al 2011. De los años posteriores a esta fecha, el registro de contratación pública contiene toda la información pertinente a la contratación mayor efectuada por cualquiera de los departamentos de la Generalitat de Cataluña.»

### 3.4.3.4. La conveniencia de dividir, o no, una petición

Según el volumen de datos que se quieran pedir, hay que considerar si conviene, o no, dividir la petición en dos o más solicitudes diferentes. Por ejemplo, si los datos que interesan hacen referencia a los sueldos que reciben los altos cargos de cuatro empresas públicas diferentes, será más adecuado formular una petición por cada una de las empresas, en vez de una sola.

En este caso, el motivo reside en el hecho de que cada empresa tiene una gestión y unas especificaciones diferentes y, por lo tanto, es fácil que al no poder acceder a los datos de una empresa concreta, se nos deniegue el acceso a todas. Si hacemos una petición para cada empresa, seremos más concretos y más directos sobre la información que solicitamos, y resultará más difícil que la institución pública la deniegue por omisión.

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.5. La minería de datos

#### 3.5.1. La minería

Hasta ahora, hemos hecho un repaso de las bases de datos disponibles y de los mecanismos de acceso a la información pública que hay actualmente. Aun así, las posibilidades de trabajar con datos se extienden más allá de las vías mencionadas.

Hay multitud de conjuntos de datos que, pese a no estar estructurados como tales, nos pueden ser útiles a la hora de llevar a cabo un proyecto periodístico. Es lo que se conoce como *scrapping* o minería de datos.

Se trata de una técnica para descargar, de forma automatizada, información almacenada en páginas web para estructurarla en una hoja de cálculo o base de datos. Un ejemplo sería el caso de los datos de los parlamentarios del Congreso de Diputados. La estructura de la página web permite la descarga automática de los datos en formato de hoja de cálculo, sin depender de que la institución pública la ponga al servicio de la ciudadanía.

Para llevar a cabo estas descargas, normalmente se utilizan lenguajes de programación como R o Python. Teniendo en cuenta que la gran mayoría de los periodistas no tienen suficientes conocimientos para usar estas herramientas, hay otras alternativas como por ejemplo aplicaciones informáticas que, de una manera más sencilla, a pesar de que con ciertas limitaciones, permiten que el periodista pueda autogestionar las descargas.

Es el caso de herramientas como Import.io, o las hojas de cálculo de Google.

## 3. Fuentes de información y mecanismos de obtención de datos

### 3.5. La minería de datos

#### 3.5.2. Herramientas

Hay una serie de herramientas y de aplicaciones informáticas gratuitas que permiten de manera sencilla poder descargar información de una página web y almacenarla en hojas de cálculo de forma automática. A continuación, hacemos una lista de algunas de las más utilizadas por su practicidad y sencillez:

##### 1) Import.io

Es una herramienta en línea que, sin necesidad de descargar en el dispositivo, permite de manera fácil y sencilla la descarga de datos. Solo requiere la creación de una cuenta de manera gratuita y conexión a internet para poder utilizarla. El formato gratuito acepta hasta cien descargas automatizadas al mes. Si se quieren efectuar más, habrá que esperar al mes siguiente, o suscribirse a la versión *premium*.

##### 2) Hoja de cálculo de Google

La herramienta de gestión de hojas de cálculo de Google contiene fórmulas preparadas para descargar contenido almacenado en las páginas web, sea cual sea el formato de origen: XML, HTML, CSV o TSV. Se trata de una serie de fórmulas que se tienen que escribir en la primera casilla en la que se quiere que se importen los datos. Para su correcto funcionamiento, hay que seguir la sintaxis de la fórmula:

- IMPORTXML (url, xpath\_query)
- IMPORTHTML (url, query, index)
- IMPORTDATA (url)

##### 3) Web Scraper

Se trata de un complemento del navegador Google Chrome, que permite automatizar la descarga de información de las páginas web. Para poder utilizarlo, primero hay que instalar el complemento en el navegador. A continuación, aparecerá un icono en la parte superior derecha de la pantalla. Cada vez que lo tengamos que utilizar, haremos clic en el icono con la página web abierta. A partir de aquí, solo habrá que seguir los pasos que indica el complemento.

##### 4) ParseHub

Es una aplicación informática que se puede descargar tanto en PC como en dispositivos Macintosh, y que permite de manera sencilla la descarga automática de datos. Se trata de una herramienta gratuita que, una vez aprendido su funcionamiento, permite de manera eficaz conseguir su propósito. Actualmente es una de las herramientas más potentes, puesto que sea cual sea la estructura de la página, acostumbra a extraer con éxito los datos.

## 4. Análisis de datos

### 4.1. Herramientas para el análisis de datos

#### 4.1.1. Gestores de hojas de cálculo

##### 4.1.1.1. Introducción

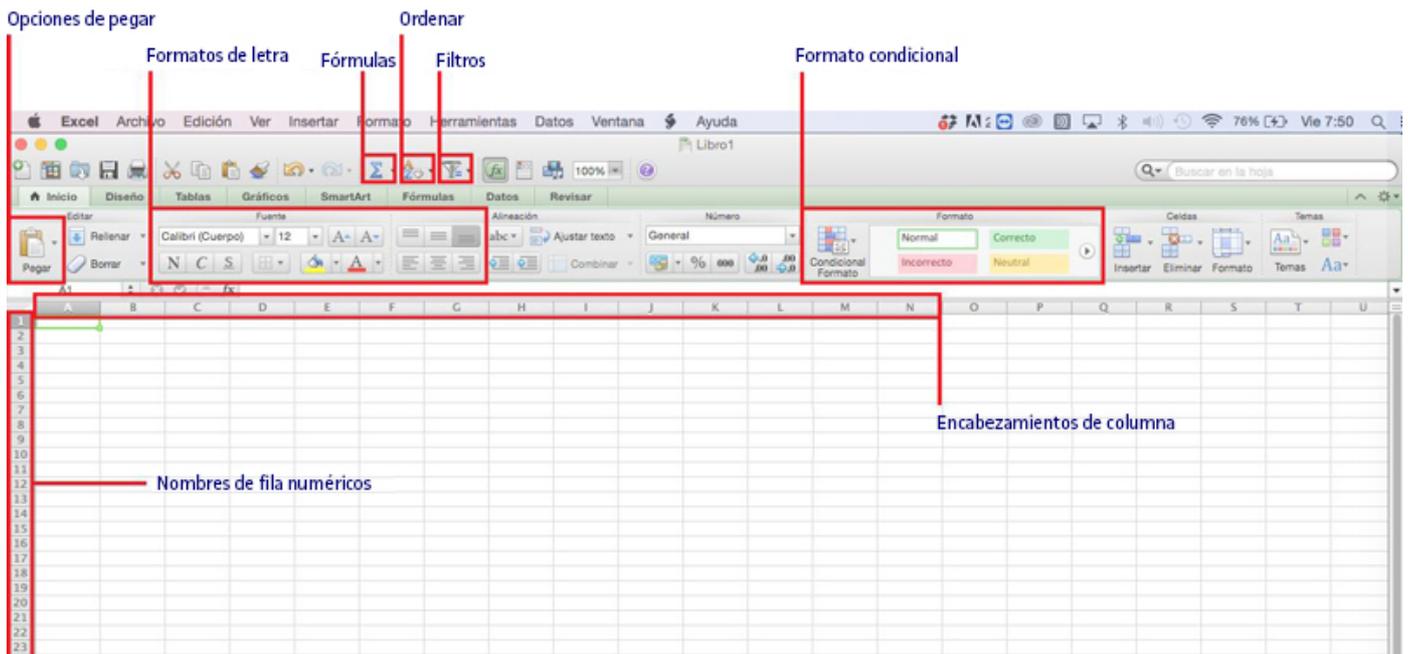
Hay diferentes herramientas para tratar el análisis de datos. En líneas generales, los datos que deberemos tratar están almacenados en hojas de cálculo que pueden analizarse mediante programas de gestión de hojas de cálculo como por ejemplo Microsoft Excel, Google Spreadsheets u Open Calc, la versión de software libre.

En los siguientes apartados, se explicará por tutoriales y por capturas de pantalla cómo gestionar un análisis de datos con la herramienta Microsoft Excel, ya que es la más común y potente.

##### 4.1.1.2. Microsoft Excel

Es un programa de gestión de hojas de cálculo que forma parte del paquete Office. Microsoft lanzó la primera versión de esta herramienta en 1985, y desde entonces se ha popularizado como la herramienta por excelencia de gestión de base de datos. Su interfaz es intuitiva y conocida por miles de usuarios en todo el mundo.

También será el programa que utilizaremos como eje central para ver cómo limpiar y analizar un conjunto de datos. Todos y cada uno de los pasos que se detallan a continuación se seguirán mediante el programa Microsoft Excel. De momento, sin embargo, hay que hacer una presentación oficial de la interfaz:

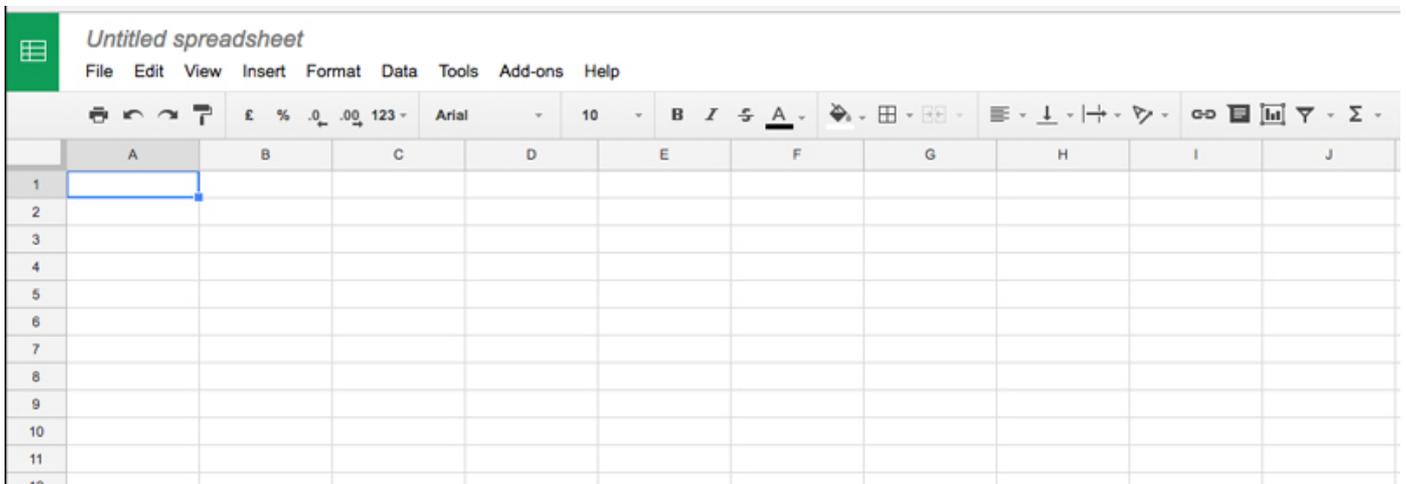


Interfaz de Microsoft Excel

##### 4.1.1.3. Google Spreadsheets

Creado en el 2006 por Google, es el homólogo de Microsoft Excel. A pesar de que la interfaz es menos conocida y añade variaciones respecto al programa de Microsoft, tiene la ventaja de ser un programa gratuito al que todo el mundo puede acceder por medio de una cuenta de Google. Además, la aplicación en línea también permite que diferentes usuarios puedan trabajar con la misma hoja de cálculo a la vez.

La contrapartida es que acepta un número de filas considerablemente inferior a Microsoft Excel, de manera que para grandes hojas de cálculo no hay otra alternativa que utilizar Excel.



Interfaz de Google Spreadsheets

#### 4.1.1.4. LibreOffice Calc

Se trata de un programa de gestión de hojas de cálculo de libre utilización. Como todo el paquete de software libre, LibreOffice está disponible para todas las plataformas: Mac, Microsoft y Linux, entre las más conocidas. A pesar de que ha mejorado mucho respecto a las versiones anteriores, a la hora de utilizar determinadas fórmulas hay ciertas limitaciones, puesto que no ofrece el mismo abanico que Microsoft Excel o Google Spreadsheets.

Estos son los programas de gestión de hojas de cálculo más utilizadas hoy día. Ahora bien, cuando estamos ante un conjunto de datos mucho más grande almacenado en una base de datos, tendremos que optar por herramientas más completas que permitan gestionar un mayor volumen.

## 4. Análisis de datos

### 4.1. Herramientas para el análisis de datos

#### 4.1.2. Herramientas avanzadas

##### 4.1.2.1. SQL (Structured Query Language)

Se trata de un lenguaje que permite modificar, consultar y llevar a cabo operaciones sobre bases de datos relacionales. Respecto a otros lenguajes, tiene una sintaxis sencilla e intuitiva. Se puede hacer funcionar directamente desde la consola o terminal del ordenador, o por medio de algún programa que dote al lenguaje de una interfaz que permita trabajarlo de manera más sencilla.

##### 4.1.2.2. Python

Es otro lenguaje de programación que también permite trabajar con bases de datos. Más allá del análisis de bases de datos, tiene muchos otros usos, y esto lo convierte en un lenguaje mucho más potente y útil que SQL. Aun así, la sintaxis es mucho más compleja, y aprender su funcionamiento requiere más especialización por parte del periodista.

##### 4.1.2.3. R

Este tercer lenguaje de programación, originalmente parte del campo de la estadística, por eso se utiliza mucho tanto en el ámbito académico como en el periodístico. Al igual que Python, es una herramienta que ofrece unas funcionalidades que van más allá del acceso a las bases de datos: también permite hacer análisis visuales, elaborar gráficos, ir un paso más allá en métricas estadísticas y en la automatización de descargas de datos.

## 4. Análisis de datos

### 4.2. La importancia de ordenar bien una hoja de cálculo

#### 4.2.1. La estructura de los datos

La estructura del *data set* o conjunto de datos tiene que responder únicamente a la voluntad periodística. Antes de empezar a recoger los datos, es imprescindible plantearse qué preguntas queremos que responda el conjunto de la información. Según «la entrevista» que queramos hacer en la hoja de cálculo, habrá que incluir unos campos u otros, y estructurar las filas y las columnas de una forma o de otra. Se trata de un planteamiento previo que hay que hacer tanto si los datos han sido obtenidos por una fuente tercera, por ejemplo la Administración pública, como si han sido recogidos manualmente por el periodista.

En el caso de datos provenientes de fuentes terceras, además, habrá que tener en cuenta que a menudo están estructurados según los intereses y el objetivo de la fuente, no desde un punto periodístico. Por eso, también es importante replantearse si es pertinente, o no, reestructurar las columnas y las filas en función de las necesidades periodísticas, y no solo esto, sino que una vez comprobado que todos los campos sean correctos, se tendrá que «limpiar» y disponer la información de forma adecuada para poder analizarla correctamente.

## 4. Análisis de datos

### 4.2. La importancia de ordenar bien una hoja de cálculo

#### 4.2.2. Limpiar una hoja de cálculo

Limpiar unos datos quiere decir ponerlos de forma adecuada para que se puedan analizar mediante programas de gestión de hojas de cálculo, como por ejemplo Microsoft Excel o Google Spreadsheets, o de gestión de base de datos como SQL o Microsoft Access.

Proceder directamente al análisis sin haber hecho una limpieza de los datos puede conducir al periodista a extraer conclusiones del todo erróneas. A la hora de efectuar esta limpieza, hay algunas comprobaciones que prácticamente se pueden aplicar en la totalidad de los casos.

- Eliminar todos los elementos complementarios: filas, columnas e imágenes incrustadas extra. Cuando se utilizan datos de la Administración pública, a menudo se incrusta la imagen de la entidad correspondiente y se ponen cabeceras extra en las filas superiores. Para poder hacer un análisis por medio de cualquier programa, será necesario que en la primera fila haya la cabecera de cada columna, es decir, el nombre del campo.
- Comprobar que los registros de una misma columna no están escritos de maneras diferentes. Esto nos podría llevar a hacer recuentos erróneos. Esta comprobación se puede hacer con una tabla dinámica, que veremos más adelante cómo se hace, o utilizando programas específicos de limpieza de datos como por ejemplo [Google Refine \(\\*\)](#).
- Comprobar que no haya espacios extra dentro de las celdas que forman la tabla. Para eliminar los espacios complementarios en un gestor de hojas de cálculo, tanto si es Microsoft Excel o Google Spreadsheets, será necesario hacerlo con la fórmula =ESPACIOS, o =TRIM, si el programa está en inglés.
- Que no haya diferencias de minúsculas y de mayúsculas. Precisamente por esta razón, es una buena idea cambiar a mayúsculas todos aquellos campos que pretendan ser buscables por la fórmula de Excel =MAYUSC, o =UPPER, si el programa está en inglés.
- Comprobar que todos los números estén en el mismo formato. Para hacerlo, habrá que seleccionar la columna y, en la pestaña «formato», hacer clic en la opción de «celdas». Se abrirá un menú emergente y, en la primera pestaña, la de «número», elegiremos el formato idóneo para todas las columnas que contienen cifras.

Antes de hacer cualquier modificación de la estructura, hay que añadir un número único para cada registro. De este modo, tendremos un valor único para cada fila y no nos confundiremos en el supuesto de que *a posteriori* se haga algún duplicado sin querer.

## 4. Análisis de datos

### 4.2. La importancia de ordenar bien una hoja de cálculo

#### 4.2.3. Un ejemplo práctico

Para este ejemplo, podéis trabajar con el archivo:  [Adjudicacions Brut](#)

Para entender mejor cómo se tiene que estructurar y limpiar un conjunto de datos, es mejor que sigamos un ejemplo práctico. Imaginemos que queremos llevar a cabo un análisis de la licitación de contratos de todos los departamentos de la Generalitat de Cataluña durante la segunda quincena del mes de diciembre del 2016. El objetivo es averiguar la siguiente información:

- ¿Cuál es la empresa que se lleva un mayor importe por contratos de cada departamento?
- ¿Cuál es la media de los importes de las licitaciones de suministro por departamento?
- El número de contratos que están adjudicados a más de una empresa.

Sabemos con antelación que el registro de contratación pública contiene toda la información respecto a la contratación mayor de la Generalitat de Cataluña, estructurada en hojas de cálculo de forma quincenal. Se puede acceder al enlace por el portal de transparencia de la Administración catalana. Para llevar a cabo este ejercicio, descargaremos el archivo XLS correspondiente a la primera y segunda quincena de diciembre del 2016, el periodo de tiempo en el que queremos centrar el foco periodístico.

En cuanto abramos los documentos Excel, podremos ver la información correspondiente a todos los departamentos de la Generalitat, estructurada con los campos siguientes:

- Organización de la contratación: departamento u organismo que convoca el concurso público.
- Fecha de adjudicación: día, mes y año de la adjudicación. Puesto que la hoja de cálculo descargada corresponde a la segunda quincena de diciembre del 2016, solo constarán adjudicaciones hechas durante este periodo de tiempo.
- Descripción del objeto: necesidad que pretende cubrir el contrato, es decir, motivo por el que se convoca.
- Tipo de contrato: tipología según el objeto del contrato, que puede ser de servicios, de suministros o de obras.
- Procedimiento de la adjudicación: proceso que se ha seguido para licitar y adjudicar el contrato: abierto, negociado, etc.
- Valor estimado del contrato: importe en el que está valorado el concurso en el momento de la licitación.
- Importe de adjudicación: importe final de la adjudicación del contrato. Unidades en euros.
- Empresa adjudicataria: nombre de la empresa o las empresas que han ganado el concurso. Para cada concurso, hay tantas filas como adjudicatarios finales que han ganado el concurso.
- Duración: periodo de tiempo que durará el contrato.

#### 1) Eliminar filas sobrantes

Antes de dar cualquier paso, primero será necesario eliminar la imagen incrustada y las filas de más que encabezan la hoja de cálculo. Para hacerlo, seleccionaremos las filas con el cursor, haremos clic con el botón derecho del ratón y desplegaremos el menú para eliminarlas. Haremos lo mismo para las filas inferiores, que indican el total de datos y una especificación metodológica sobre los datos.

The screenshot shows the Excel interface with a context menu open over a table. The table has the following structure:

DADES BÀSIQUES DEL CONTRATE		
L'OBJECTE DEL CONTRACTE	TIPUS DE CONTRACTE	F D
ia i seguretat de les dependències administratives del Departament d'Afers i nacionals i Exteriors i Transparència ubicades a Via Laietana, 14 de Barcelona	Serveis	G
d'energia amb destinació al carrer Tapineria 10	Subministraments	G
el primer trimestre de 2017	Serveis	G
eria del Departament d'Afers i Relacions Institucionals i Exteriors i Transparència	Serveis	G
els edificis Princesa i Tapineria adscrits al Departament	Serveis	G
22/12/16	Servei de traducció i/o correcció de textos per les unitats del Departament d'Afers i Relacions Institucionals i Exteriors i Transparència	

Eliminación de filas sobrantes

## 2) Descombinar las celdas

Si damos un repaso muy rápido a la hoja de cálculo, vemos que prácticamente en todas las columnas hay celdas combinadas y, por lo tanto, nos será imposible aplicar los filtros para ordenar de mayor a menor.

Para hacerlo, primero tendremos que descombinar las celdas y decidir de qué manera estructuramos la información.

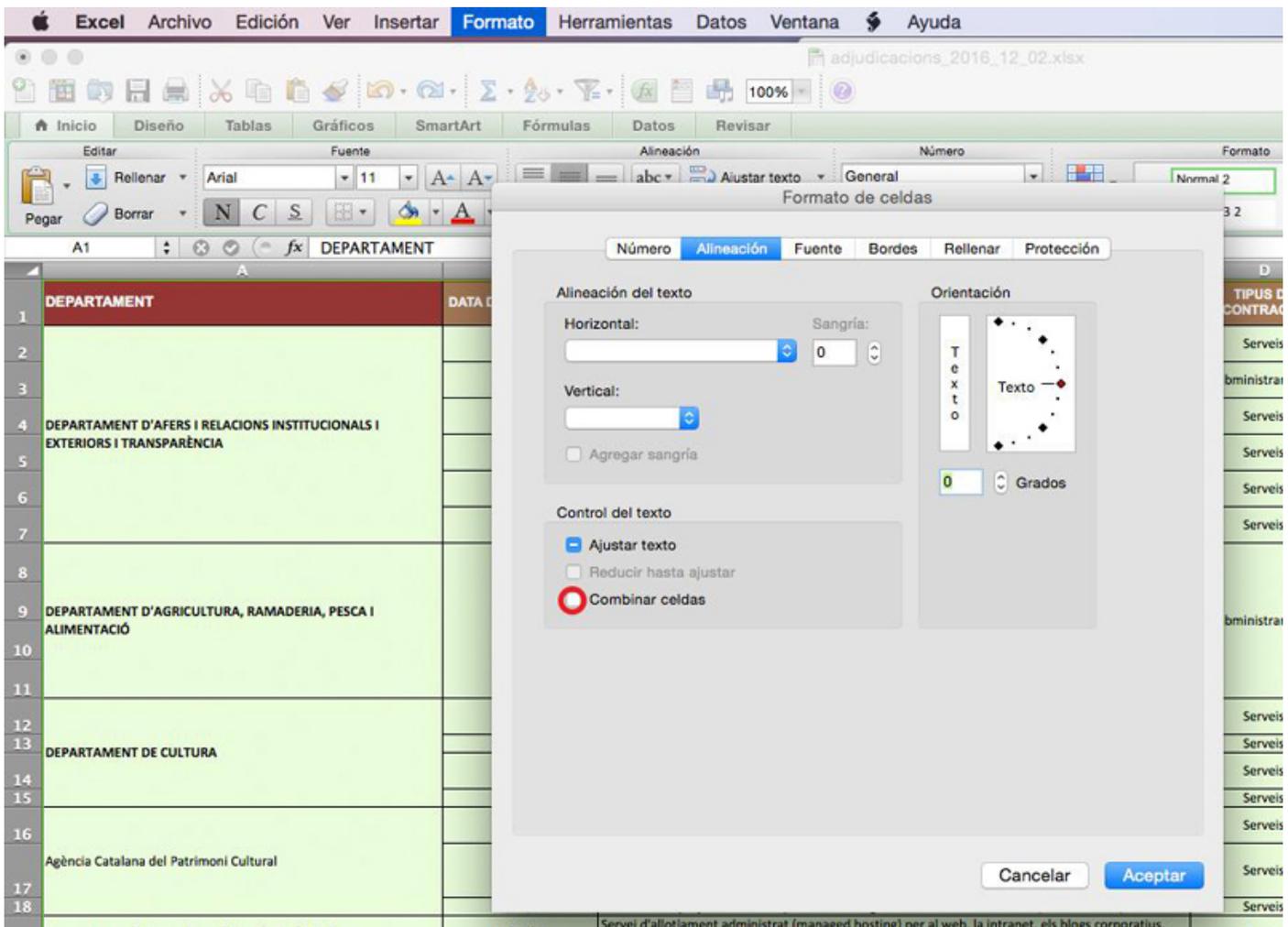
Para descombinarlas, seleccionaremos todo el conjunto de datos de Excel haciendo clic en el cuadrante superior izquierda, tal y como indica la siguiente imagen, o yendo columna por columna.

The screenshot shows the Microsoft Excel interface with the following data in the table:

DEPARTAMENT	DATA D'ADJUDICACIÓ	DESCRIPCIÓ DE L'OBJECTE DEL CONTRACTE
DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA	19/12/16	Servei de vigilància i seguretat de les dependències administratives del Departament d' Afers i Relacions Institucionals i Exteriors i Transparència ubicades a Via Laietana, 14 c
	23/12/16	Subministrament d'energia amb destinació al carrer Tapineria 10
	23/12/16	Serveis postals pel primer trimestre de 2017
	28/12/16	Servei de missatgeria del Departament d' Afers i Relacions Institucionals i Exteriors pel 2017
	20/12/16	Servei de neteja dels edificis Princesa i Tapineria adscrits al Departament
	22/12/16	Servei de traducció i/o correcció de textos per les unitats del Departament d' Afers i Relacions Institucionals i Exteriors i Transparència
DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ	25/10/16	Subministrament de vestuari per al Cos d'Agents Rurals del Departament d' Afers i Relacions Institucionals i Exteriors i Transparència

Descombinación de celdas (primer paso)

Una vez seleccionado todo el conjunto de la información, iremos a la pestaña «Formato», y dentro de esta, a «Celdas». En el menú emergente que aparecerá, tenemos que hacer clic en la pestaña «Alineación» y desmarcar la casilla inferior de «Combinar celdas» tal y como se muestra a continuación.



Descombinación de celdas (segundo paso)

Una vez efectuada esta operación, todas las celdas quedarán automáticamente divididas de la manera siguiente:

The image shows the resulting spreadsheet after unmerging cells. The table has 7 columns: DEPARTAMENT, DATA D'ADJUDICACIÓ, DESCRIPCIÓ DE L'OBJECTE DEL CONTRACTE, TIPUS DE CONTRACTE, PROCEDIMENT D'ADJUDICACIÓ, VALOR ESTIMAT DEL CONTRACTE (IVA exclòs), and IMPORT D'ADJUDICACIÓ (IVA inclòs). The data is as follows:

DEPARTAMENT	DATA D'ADJUDICACIÓ	DESCRIPCIÓ DE L'OBJECTE DEL CONTRACTE	TIPUS DE CONTRACTE	PROCEDIMENT D'ADJUDICACIÓ	VALOR ESTIMAT DEL CONTRACTE (IVA exclòs)	IMPORT D'ADJUDICACIÓ (IVA inclòs)
DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA	19/12/16	Servei de vigilància i seguretat de les dependències administratives del Departament d'Affers i Relacions Institucionals i Exteriors i Transparència ubicades a Via Laretana, 14 de Barcelona	Serveis	Contracte derivat d'acord marc	328.325.76 €	158.831.38 €
	23/12/16	Subministrament d'energia amb destinació al carrer Tapineria 10	Subministraments	Contracte derivat d'acord marc	33.249.67 €	42.289.05 €
	23/12/16	Serveis postals pel primer trimestre de 2017	Serveis	Contracte derivat d'acord marc	4.661.16 €	5.640.00 €
	28/12/16	Servei de missatgeria del Departament d'Affers i Relacions Institucionals i Exteriors i Transparència pel 2017	Serveis	Contracte derivat d'acord marc	759.09 €	918.50 €
	20/12/16	Servei de neteja dels edificis Princesa i Tapineria adscrits al Departament	Serveis	Contracte derivat d'acord marc	511.867.68 €	224.592.17 €
	22/12/16	Servei de traducció i/o correcció de textos per les unitats del Departament d'Affers i Relacions Institucionals i Exteriors i Transparència	Serveis	Obert	33.057.85 €	40.000.00 €
DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ	25/10/16	Subministrament de vestuari per al Cos d'Agents Rurals del Departament d'Agricultura, Ramaderia, Pesca i Alimentació	Subministraments	Obert	188.850.00 €	8.526.87 €
						70.657.35 €
						94.334.63 €
						10.648.00 €
DEPARTAMENT DE CULTURA	20/12/16	Direcció i execució de les intervencions arqueològiques i paleontològiques d'urgència al territori de Catalunya	Serveis	Obert	78.512.40 €	95.000.00 €
	7/12/16	Servei d'atenció al públic i de visites guiades de l'Arts Santa Mònica (ASMI) per al 2017	Serveis	Obert	165.605.79 €	171.594.44 €
	16/12/16	Serveis de suport a la dinamització cultural per a la Biblioteca Pública de Girona (període gener-setembre 2017)	Serveis	Obert	75.272.73 €	90.900.00 €
	22/12/16	Serveis de producció audiovisual de la programació d'activitats de l'Arts Santa Mònica per al 2017	Serveis	Obert	38.600.32 €	40.919.18 €
Agència Catalana del Patrimoni Cultural	16/12/16	Direcció, redacció dels informes i de la memòria i execució de la intervenció arqueològica al teatre romà de Tarragona	Serveis	Obert	86.768.52 €	95.330.84 €
	21/11/16	Conservació - restauració de les restes arqueològiques del Teatre Romà de Tarragona corresponents als sectors de la cavea, l'orchestra, el proscenium i el frons scaenae, així com del seu entorn immediat	Serveis	Obert	121.053.25 €	126.612.50 €
	18/12/16	Redacció del projecte executiu per a la museografia del Castell de Miravet (Ribera d'Ebre)	Serveis	Obert	28.925.62 €	33.250.00 €
Consorti per a la Normalització Lingüística (CPNL)	16/12/16	Servei d'allotjament administrat (managed hosting) per al web, la intranet, els blogs corporatius, Moodle, els wikis i altres	Serveis	Obert	10.780.00 €	4.235.00 €
DEPARTAMENT D'EMPRESA I CONEIXEMENT	22/12/16	Prestació del servei de conducció i manteniment integral, de les dependències i instal·lacions de diversos immobles ocupats pel Departament d'Empresa i Coneixement (lot 1 Barcelona)	Serveis	Obert	439.632.00 €	245.783.90 €
	15/12/16	Servei de neteja de les dependències d'Avinguda del Segre, 7 (Lleida)	Serveis	Contracte derivat d'acord marc	11.484.56 €	12.327.48 €
Agència Catalana de Turisme	19/12/16	Serveis d'assessorament, desenvolupament i execució del Social Media Plan de l'Agència Catalana de Turisme, i més concretament de la gestió de les xarxes socials pel mercat francès per l'any 2017	Serveis	Altres procediments segons instruccions internes	45.000.00 €	54.450.00 €
	19/12/16	Contractació del servei de suport en la implementació d'accions a les xarxes socials del Catalunya Experience online, i la coordinació d'accions on-off	Serveis	Altres procediments segons instruccions internes	28.114.51 €	34.018.56 €

Celdas descombinadas

El problema es que después de este cambio, las celdas descombinadas han quedado vacías de contenido. Al dividir las, ha quedado la información que corresponde solo a la celda superior. Por lo tanto, lo que tendremos que hacer a continuación es copiar el contenido al resto de las celdas. A pesar de que hay macros, una opción avanzada del programa, prácticas para Microsoft Excel, que permiten automatizar este proceso, es un procedimiento más complejo que no resulta válido para todas las hojas de cálculo. En cambio, el sistema manual siempre será una alternativa viable.

### 3) Entender la estructura y llenar el contenido con cuidado

Así pues, para copiar el contenido de forma manual, nos situaremos en la parte inferior izquierda de la casilla que contiene la información que queremos copiar. Cuando aparezca una cruz negra, arrastramos la cruz hasta la última casilla en la que queremos que se copie el dato, y automáticamente la información quedará copiada. Si en vez de arrastrar hacemos doble clic, entonces el contenido se copiará a todas las celdas vacías consecutivas que haya a continuación de la misma columna.

	A	B	C
1	<b>DEPARTAMENT</b>	<b>DATA D'ADJUDICACIÓ</b>	<b>DESCRIPCIÓ DE L'OBJECTE DEL CONTRACTE</b>
2	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA	19/12/16	Servei de vigilància i seguretat de les dependències a Relacions Institucionals i Exteriors i Transparència ub
3		23/12/16	Subministrament d'energia amb destinació al carrer
4		23/12/16	Serveis postals pel primer trimestre de 2017
5		28/12/16	Servei de missatgeria del Departament d'Affers i Relac pel 2017
6		20/12/16	Servei de neteja dels edificis Princesa i Tapineria adsc
7		22/12/16	Servei de traducció i/o correcció de textos per les uni Institucionals i Exteriors i Transparència
8	DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ	25/10/16	Subministrament de vestuari per al Cos d'Agents Rur Pesca i Alimentació

Copia de contenido de forma manual

Para navegar de manera más directa y sencilla por la tabla sin tener que utilizar la barra de la derecha, se puede usar la tecla Control y las flechas. En el caso de Macintosh, la combinación de teclas será Comando y, de nuevo, las flechas.

Antes de empezar a hacer el trabajo de copiar y pegar, es importante dar un vistazo a todo el conjunto de datos para entender cómo está estructurado. Es necesario llenar la información columna por columna, sin caer en errores.

A simple vista, localizamos que la primera columna, la de «departamento», sirve al mismo tiempo como «encabezamiento» y como «registro». Es decir, que hay licitaciones que las hace el departamento, otras que las hacen otros organismos que conforman el departamento, y otros casos en los que el departamento no hace ninguna licitación y simplemente lo anotan como encabezamiento. Es lo que ocurre, por ejemplo, con la fila 60.

Para llevar a cabo nuestro propósito periodístico, nos interesa que haya tanto la columna del departamento, como la del organismo concreto que hace la licitación. Además, para poder analizarlo bien no puede haber registros vacíos que solo sirvan como encabezamientos.

Por lo tanto, crearemos una columna nueva que se llamará «Departamento» y renombraremos la que teníamos en segundo lugar como «Organismo licitador». A continuación, copiaremos en la primera columna el nombre del departamento que hace la licitación, y en la segunda, el organismo concreto que hace la licitación: si es el departamento mismo, copiaremos el nombre del departamento y, si es un organismo más pequeño que pertenece al departamento, entonces copiaremos el nombre del organismo.

Los departamentos de Gobernación, Administraciones Públicas y Vivienda, el de Justicia, y el de Territorio y Sostenibilidad no han hecho ninguna licitación y, por lo tanto, solo sirven como encabezamientos. En estos tres casos, eliminaremos la fila vacía y copiaremos el nombre del departamento en las filas inferiores, en las que están las licitaciones de los órganos que sí han licitado y dependen de estos departamentos.

Antes de empezar el análisis, hay que añadir una columna que identifique cada contrato con un registro único, por ejemplo, uno numérico. Necesitamos esta columna porque cada fila no es un contrato, sino una adjudicación. De forma que si un contrato tiene más de un adjudicatario, el contenido estará acumulado en distintas filas según el número de empresas adjudicatarias. Para solucionar este aspecto, crearemos una columna nueva con un código único para cada contrato. El valor adjudicado a cada contrato, pues, se repetirá en los casos en los que un contrato haya sido adjudicado a más de una empresa.

Una vez que hemos creado esta columna, acabaremos de copiar la información para el resto de los campos. Al igual que hemos hecho con las dos primeras columnas, arrastraremos la información de la primera celda al resto de las casillas duplicadas. Cuidado con arrastrar y no copiar, puesto que hay valores, como las fechas o los números, que pueden aumentar en lugar de copiarse exactamente. En estos casos, tendremos que copiar la información y no arrastrarla.

#### 4) Homogeneizar los campos

Una vez modificada la estructura, podremos proceder a hacer la limpieza habitual con las fórmulas que hemos visto en el apartado anterior, y que sirven para homogeneizar todos los campos. Así pues, aplicaremos las fórmulas correspondientes para quitar los espacios excedentes y colocar todas las letras en mayúscula.

Para hacerlo, crearemos una columna nueva al lado de la que queremos limpiar, haciendo un clic con el botón de la derecha sobre la columna.

	A	B	C	D
			ORGANISME LICITADOR	
	CODI	DEPARTAMENT		DATA D'ADJUDICACIÓ
1		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT EXTERIOR	
2	1	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT EXTERIOR	19/12/16
3	2	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT EXTERIOR	23/12/16
4	3	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT EXTERIOR	23/12/16
5	4	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT EXTERIOR	28/12/16
6	5	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT EXTERIOR	20/12/16
7	6	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	22/12/16

Homogeneizar los campos (primer paso)

Una vez creada la columna, aplicaremos la fórmula correspondiente, tanto si es =ESPACIOS, o =MAYUSC, y la arrastraremos a lo largo de toda la columna.

	A	B	C	D
1	<b>CODI</b>	<b>DEPARTAMENT</b>		<b>ORGANISME LICITADOR</b>
2	1	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I	=ESPACIOS(B2)	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
3	2	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
4	3	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
5	4	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
6	5	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA

Homogeneizar los campos (segundo paso)

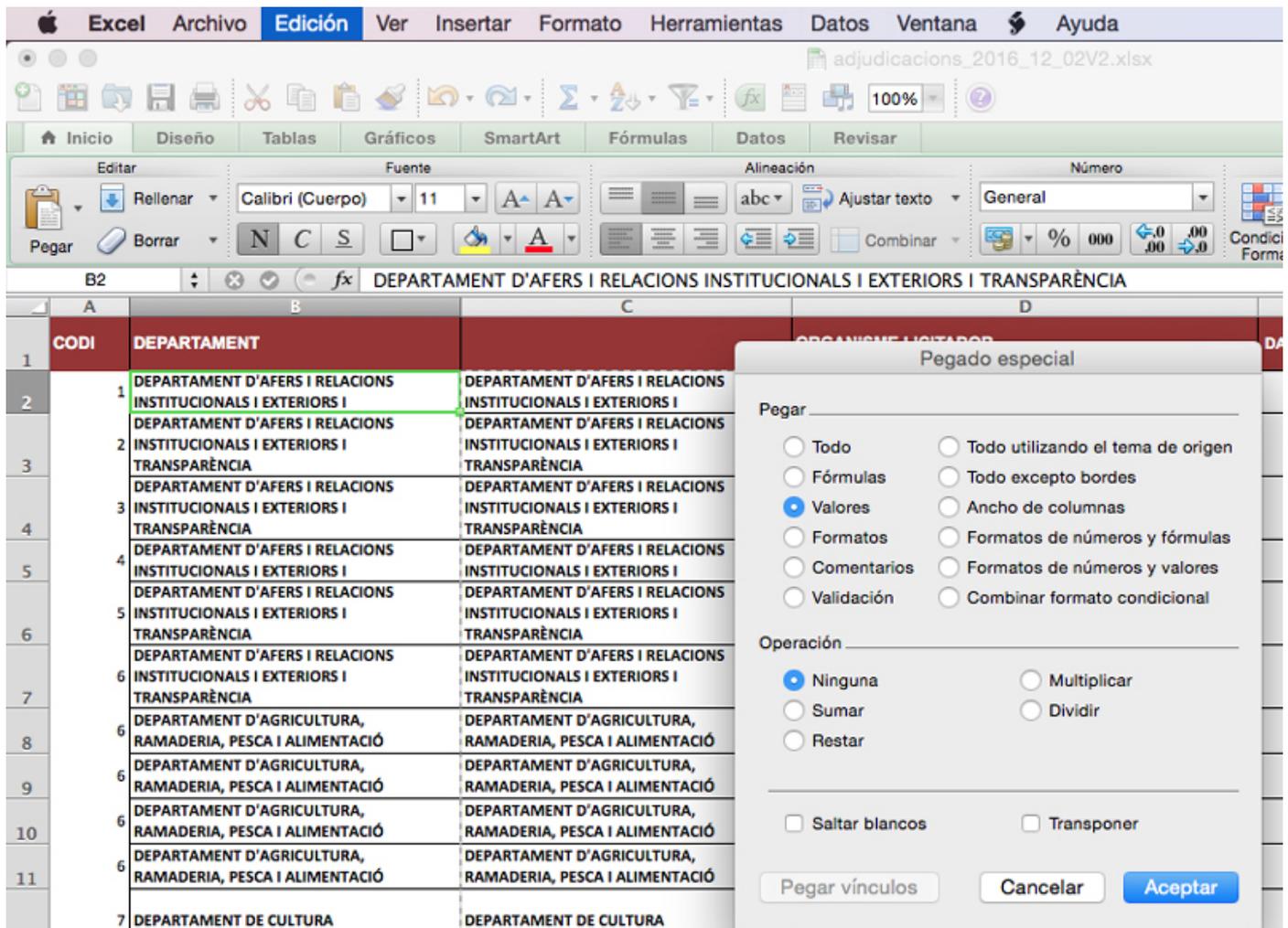
Para poder trabajar mejor *a posteriori* con la hoja de cálculo, es conveniente eliminar las fórmulas una vez aplicadas y, por lo tanto, trabajar solo con los valores. Para hacerlo, copiaremos toda la columna sobre la que estamos aplicando la fórmula, en este caso la columna C.

Con la información copiada, nos situaremos sobre la columna que contiene la información original, la B, e iremos a la pestaña «Edición» para hacer un clic sobre la opción «Pegado especial».

	A	B	C	D
1	<b>CODI</b>	<b>DEPARTAMENT</b>		<b>ORGANISME LICITADOR</b>
2	1	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
3	2	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
4	3	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
5	4	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
6	5	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
7	6	DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA		DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA
8	6	DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ	DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ	DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ

Homogeneizar los campos (tercer paso)

Por defecto, se abrirá un menú que tendrá marcada la opción de pegarlo «todo». Si no queremos copiar también las fórmulas, deberemos cambiar la opción determinada por defecto por «valores», tal y como se muestra en la siguiente captura de pantalla.



Homogeneizar los campos (cuarto paso)

Una vez copiada la información de la forma correcta, podremos proceder a eliminar la columna de la fórmula. Este proceso lo repetiremos con todas las columnas, hasta que toda la hoja de cálculo esté a punto para analizar.

Otra manera más rápida de aplicar esta fórmula consiste en aplicar las fórmulas en una hoja nueva, en lugar de una columna nueva. De este modo, aplicaremos la fórmula deseada (=MAYUSC o =ESPACIOS) a todas las columnas a la vez. A pesar de que a la larga es la forma más efectiva, cuando estamos empezando, ir paso por paso va bien para evitar confusiones.

## 5) Comprobar que la herramienta de limpieza se ha hecho correctamente

Después de haber hecho la limpieza, y antes de pasar al análisis, habrá que comprobar que no se han cometido errores mientras se teclaba durante la limpieza. Uno de los errores comunes es, por ejemplo, haber pasado por alto copiar y pegar alguna celda. Para detectar de manera rápida este error, lo más sencillo es navegar por la hoja de cálculo con la tecla Control + flechas, o Comando + flechas si trabajamos en Mac.

Por otro lado, también habrá que comprobar que el formato de las columnas es correcto. Sobre todo a la hora de hacer el pegado especial de únicamente los «valores», puesto que al hacerlo perdemos el formato atribuido a cada columna. Para modificar el formato de los datos, será necesario seleccionar la columna, y en la pestaña «Formato» hacer clic en la opción «Celdas», seleccionando el formato adecuado para cada una de las columnas. Por ejemplo, en el caso de la columna D, esto se traducirá en el formato de fecha formado por día, mes y año.

Edición Ver Insertar **Formato** Herramientas Datos Ventana Ayuda

adjudicacions\_2016\_12\_02V2.xlsx

100%

Tablas Gráficos SmartArt Fórmulas Datos Revisar

Fuente Alineación Número Formato

Normal 7

**Formato de celdas**

Número Alineación Fuente Bordes Rellenar Protección

Categoría: Muestra  
 General  
 Número  
 Moneda  
 Contabilidad  
**Fecha**  
 Hora  
 Porcentaje  
 Fracción  
 Científica  
 Texto  
 Especial  
 Personalizada

DATA D'ADJUDICACIÓ

Tipo:  
 \*14/3/01  
 \*miércoles, 14 de marzo de 01  
 14-3  
 14-3-01  
 14-03-01  
 14-mar  
 14-mar-01  
 14-mar-01  
 mar-01

Ubicación (idioma):  
 España (español)

Tipo de calendario:  
 Occidental

Los formatos de fecha presentan números que representan fechas y horas como valores de fecha. Use los formatos de hora si desea presentar solo la parte horaria.

Cancelar Aceptar

C	D	E
ORGANISME	DATA D'ADJU	DESCRIPCIÓ TIPI
DEPARTAME	42723	Servei de vigil Ser
DEPARTAME	42727	Subministram Sub
DEPARTAME	42727	Serveis posta Ser
DEPARTAME	42732	Servei de mis Ser
DEPARTAME	42724	Servei de netx Ser
DEPARTAME	42726	Servei de trad Ser
DEPARTAME	42668	Subministram Sub
DEPARTAME	42724	Direcció i exe Ser
DEPARTAME	42711	Servei d'atenc Ser
DEPARTAME	42720	Serveis de su Ser
DEPARTAME	42726	Serveis de pr Ser
gència Catal	42720	Direcció, reda Ser
gència Catal	42695	Conservació - Ser
gència Catal	42722	Redacció del Ser
consorci per a	42720	Servei d'allotj Ser
DEPARTAME	42726	Prestació del Ser
DEPARTAME	42719	Servei de netx Ser
gència Catal	42723	Serveis d'asse Ser
gència Catal	42723	Contractació Ser
gència Catal	42723	Contractació Ser
gència Catal	42732	Serveis de su Ser
gència Catal	42732	Serveis per a Ser
gència per a	42720	Disseny, cons Ser
gència per a	42726	Adquisició d'u Sub
consorci de C	42716	Serveis per al Ser
consorci de S	42726	16/15: Adquis Sub
consorci de S	42724	16/27: Subscr Sub
consorci de S	42717	16/28: Subscr Sub
consorci de S	42723	16/33: contrac Ser
consorci de S	42724	16/36: Subscr Sub
consorci de S	42723	16/37: Submir Sub
consorci de S	42724	16/38: Contrai Sub
consorci de S	42717	16/39: Submir Sub
consorci de S	42717	16/40: Subscr Sub
consorci de S	42724	16/41: Submir Sub
consorci de S	42720	16/43: Contrai Sub
consorci de S	42717	16/44: Contrai Sub
undació Cen	42733	Subministram Subministram Altres procedi 202000

## Comprobación de formatos

Ahora sí, una vez hechas las comprobaciones y tras comprobar que no hay errores de limpieza, podremos proceder a analizar los datos.

Para ir bien, el resultado final limpio tendría que quedar como el archivo adjunto siguiente:  [Adjudicacions\\_Net](#).

## 4. Análisis de datos

### 4.3. Herramientas y mecanismos de análisis

#### 4.3.1. Importar datos

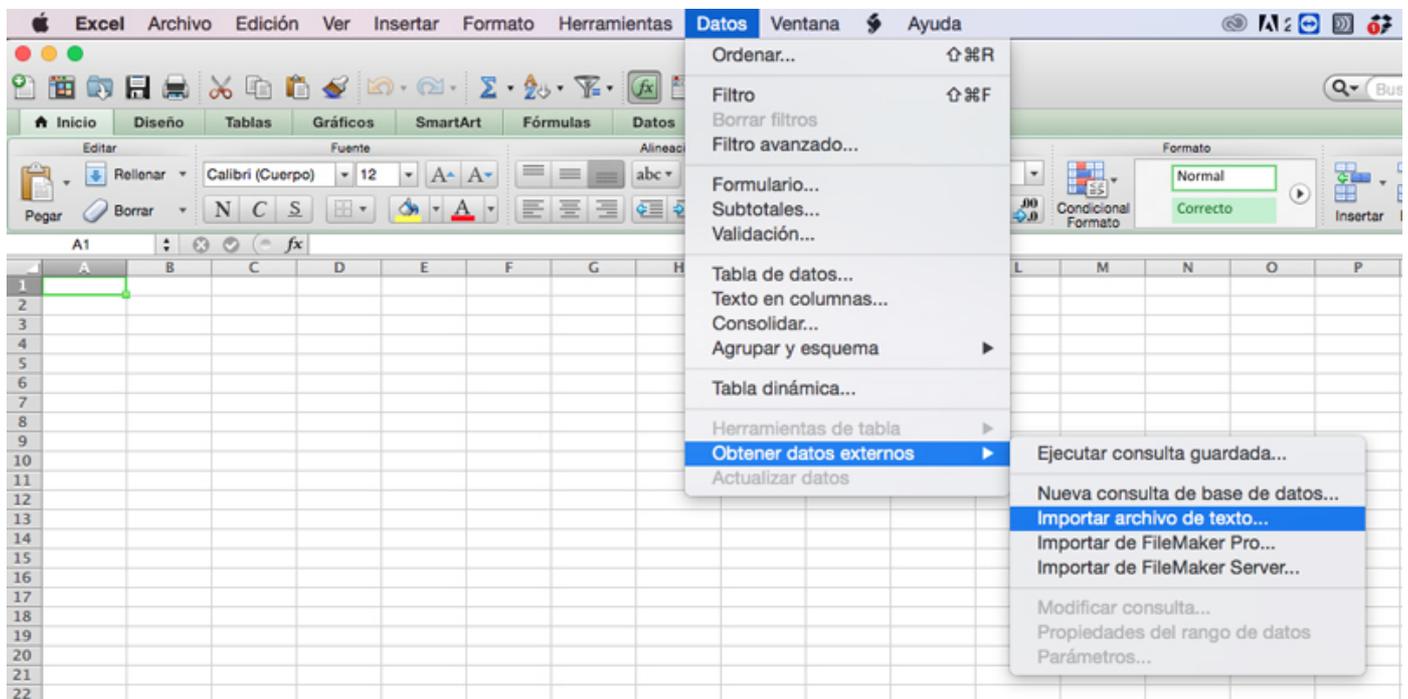
A la hora de trabajar con datos exportados de otras bases de datos más grandes, podemos encontrarnos tres tipos de formatos: XLS o XLSX, según la versión, TSV, y CSV.

El **CSV** (\*) es un archivo que se ordena en filas y en columnas, en el cual la primera fila corresponde al encabezamiento de cada columna, y cada fila a un registro. Cada campo (columna) está separado por comas. Más allá del archivo de Excel, es el tipo de formato más fácil de encontrar.

El **TSV** (\*) es un archivo estructurado de la misma manera que un CSV, pero el separador en este caso es el tabulador, y no la coma.

En el caso de los archivos XLS o XLSX, los archivos de Microsoft Excel solo hay que abrirlos con cualquiera de los programas de gestión de hojas de cálculo que hemos visto anteriormente, y empezar a trabajar. En cambio, en el caso del CSV o el TSV, tendremos que seguir unos pasos específicos, similares para los dos formatos, para poder trabajar con el archivo.

Con Microsoft Excel abierto, habrá que ir a la pestaña «Datos» y, dentro de esta, a «Obtener datos externos» y a «Importar archivo de texto». Entonces, se abrirá un menú por el cual podremos seleccionar el archivo deseado y empezar el proceso para llevar a cabo la importación.

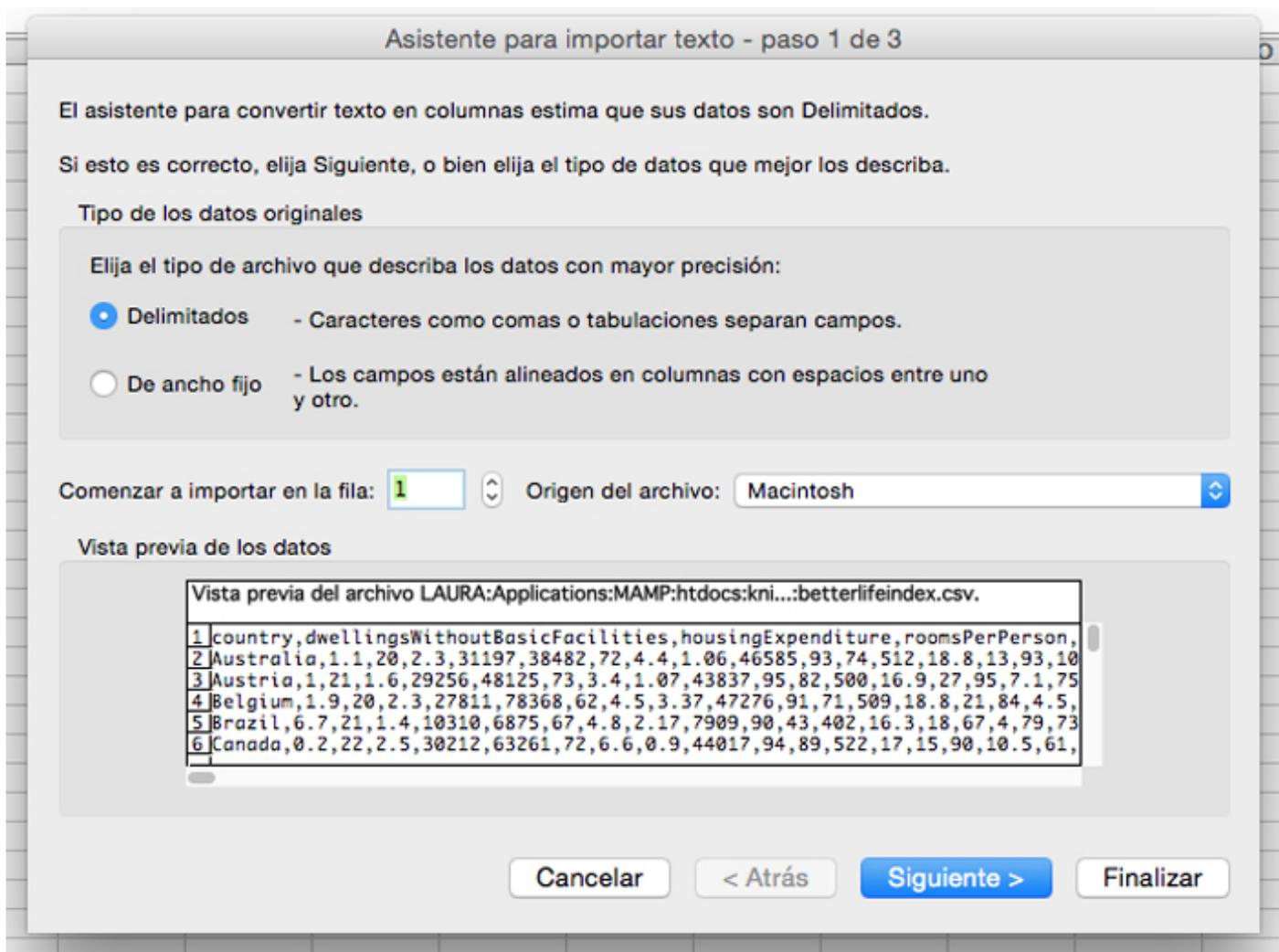


#### Importar datos (primer paso)

Lo primero que nos pide el asistente de importación de archivo de texto es saber de qué manera están delimitadas las filas: si por un valor separador, una coma o un tabulador, por ejemplo, o por un tamaño determinado. Marcaremos la primera opción: «Caracteres como comas y tabulaciones separan los campos».

Antes de pasar a la pantalla siguiente, tendremos que marcar la fila a partir de la cual se quiere iniciar la importación. Generalmente, en esta parte nos interesa mantener el valor por defecto: que empiece la importación desde la fila 1.

Por último, habrá que marcar el tipo de codificación con el que está guardado el archivo con el que queremos trabajar en el desplegable que hay en la parte inferior derecha del menú emergente. Normalmente, utilizaremos las codificaciones universales: UTF-8 o UTF-16. Aun así, a veces nos podemos encontrar con el hecho de que, según cómo haya sido guardado el archivo, haya algunos caracteres especiales en el caso del catalán, que no acaba de leer correctamente. Es lo que ocurre con la «ç» o los acentos abiertos. Por lo tanto, antes de continuar con este tipo de codificación, será necesario dar un vistazo al previsualizador para asegurarse de que lee correctamente el archivo. Si las codificaciones UTF no descodifican correctamente, se podrá probar con la Windows 1250, Windows 1251, Windows 1252 o la Macintosh. Es importante, sin embargo, que antes de pasar a la siguiente pantalla, los caracteres se previsualicen correctamente.



### Importar datos (segundo paso)

Una vez hechos estos pasos, pasaremos a la siguiente pantalla. En este caso, el objetivo es sencillo: seleccionar el valor que separa las columnas. En el caso del CSV, es una coma, y en el caso del TSV, un tabulador.

Cuando demos este paso, podremos continuar adelante. El resto de las pantallas no requieren ninguna intervención. Por lo tanto, haremos clic en el botón «Siguiente» hasta que los datos ya estén definitivamente importados y visualicemos en la pantalla del ordenador una hoja de cálculo como otra cualquiera, separada en las columnas correspondientes.

Evidentemente, hay miles de otros formatos, algunos muy comunes también en la práctica del periodismo de datos y las visualizaciones interactivas, pero son una tipología de archivos que veremos en otros casos aplicados.

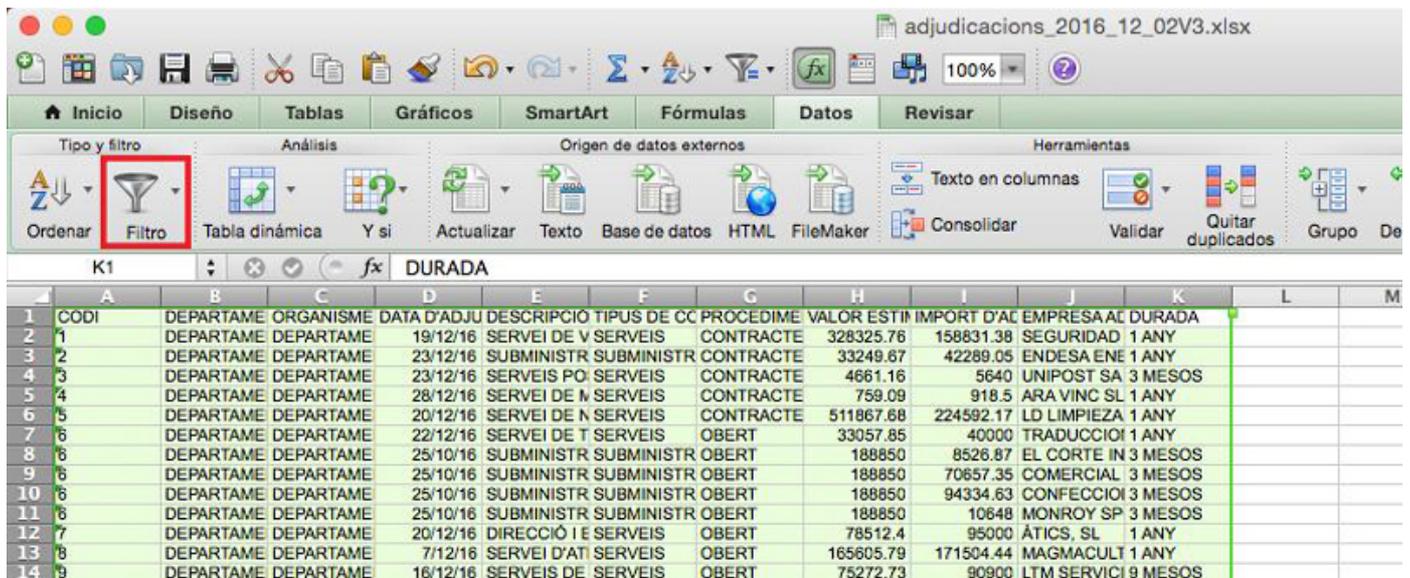
## 4. Análisis de datos

### 4.3. Herramientas y mecanismos de análisis

#### 4.3.2. Filtrar y ordenar

El primer paso a la hora de empezar a trabajar con los datos ya limpios será poner las opciones para ordenar de menor a mayor. Es muy importante que, previamente, los datos hayan sido tratados y colocados en la disposición correcta. Proceder al filtrado de datos sin que hayan sido tratados previamente puede llevar a errores que nos harán extraer conclusiones equivocadas.

Por lo tanto, y siguiendo con el ejemplo de la adjudicación de los concursos públicos: seleccionaremos las columnas que contienen datos, haremos clic en el menú de «Datos», y marcaremos la opción «Filtro».



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	CODI	DEPARTAME	ORGANISME	DATA D'ADJU	DESCRIPCIÓ	TIPUS DE CC	PROCEDIME	VALOR ESTIM	IMPORT D'AC	EMPRESA	AC	DURADA	
2	1	DEPARTAME	DEPARTAME	19/12/16	SERVEI DE V	SERVEIS	CONTRACTE	328325.76	158831.38	SEGURIDAD	1 ANY		
3	2	DEPARTAME	DEPARTAME	23/12/16	SUBMINISTR	SUBMINISTR	CONTRACTE	33249.67	42289.05	ENDESA ENE	1 ANY		
4	3	DEPARTAME	DEPARTAME	23/12/16	SERVEIS PO:	SERVEIS	CONTRACTE	4661.16	5640	UNIPOST SA	3 MESOS		
5	4	DEPARTAME	DEPARTAME	28/12/16	SERVEI DE N	SERVEIS	CONTRACTE	759.09	918.5	ARA VINC SL	1 ANY		
6	5	DEPARTAME	DEPARTAME	20/12/16	SERVEI DE N	SERVEIS	CONTRACTE	511867.68	224592.17	LD LIMPIEZA	1 ANY		
7	6	DEPARTAME	DEPARTAME	22/12/16	SERVEI DE T	SERVEIS	OBERT	33057.85	40000	TRADUCCIOI	1 ANY		
8	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	8526.87	EL CORTE IN	3 MESOS		
9	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	70657.35	COMERCIAL	3 MESOS		
10	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	94334.63	CONFECCIOI	3 MESOS		
11	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	10648	MONROY SP	3 MESOS		
12	7	DEPARTAME	DEPARTAME	20/12/16	DIRECCIÓ I E	SERVEIS	OBERT	78512.4	95000	ÀTICS, SL	1 ANY		
13	8	DEPARTAME	DEPARTAME	7/12/16	SERVEI D'AT	SERVEIS	OBERT	165605.79	171504.44	MAGMACULT	1 ANY		
14	9	DEPARTAME	DEPARTAME	16/12/16	SERVEIS DE	SERVEIS	OBERT	75272.73	90900	LTM SERVICI	9 MESOS		

#### Aplicación de filtros

De este modo, aparecerá en la primera fila de cada columna una flecha que, al hacer clic sobre la misma, nos permitirá las siguientes opciones:

- Ordenar de más a menos, y de menos a más, tanto en orden alfabético como numérico.
- Ordenar por color, en el caso de que haya casillas de un color determinado.
- Aplicar filtros de coincidencia. Es decir, elegir visualizar solo las opciones determinadas que nos puedan interesar, tanto si es un valor concreto, si son valores que sean superiores a una cifra determinada, o que empiecen o acaben por un carácter concreto.

	A	B	C	D	E	F	G	H	I	J	K
1	CODI	DEPARTAM	ORGANIS	DATA D'AD	DESCRIPC	TIPUS DE	PROCEDI	VALOR ES	IMPORT D'	EMPRESA	DURADA
2	1	DEPARTAME									RIDAD 1 ANY
3	2	DEPARTAME									SA ENE 1 ANY
4	3	DEPARTAME									ST SA 3 MESOS
5	4	DEPARTAME									NC SL 1 ANY
6	5	DEPARTAME									PIEZA 1 ANY
7	6	DEPARTAME									JCCIOI 1 ANY
8	6	DEPARTAME									RTE IN 3 MESOS
9	6	DEPARTAME									RCIAL 3 MESOS
10	6	DEPARTAME									ECCIOI 3 MESOS
11	6	DEPARTAME									OY SP 3 MESOS
12	7	DEPARTAME									SL 1 ANY
13	8	DEPARTAME									ACULT 1 ANY
14	9	DEPARTAME									ERVICI 9 MESOS
15	10	DEPARTAME									MEDIA 1 ANY
16	11	DEPARTAME									P RES1 2 MESOS
17	12	DEPARTAME									P RES1 4 MESOS I M
18	13	DEPARTAME									VENTC 4 MESOS I M
19	14	DEPARTAME									AGUIL 1 ANY
20	15	DEPARTAME									ÓN INT 1 ANY
21	16	DEPARTAME									PIEZA 1 ANY
22	17	DEPARTAME									MY DE 1 ANY
23	18	DEPARTAME									SA SAU 1 ANY
24	19	DEPARTAME									DA CAÍ 1 ANY
25	20	DEPARTAME									ASTID/ 1 ANY
26	21	DEPARTAME									DS CEL 1 ANY
27	22	DEPARTAME									M ESP 2 MESOS I M
28	23	DEPARTAME									RIENT/ 7 DIES
29	24	DEPARTAME									IA TEC 3 MESOS
30	25	DEPARTAME									, S.A. 4 ANYS
31	26	DEPARTAME									SUBS 4 ANYS
32	27	DEPARTAME									GER C 1 ANY
33	28	DEPARTAME									ASTEL 2 ANYS
34	29	DEPARTAME									CAN C 1 ANY
35	30	DEPARTAME									IER B. 4 ANYS
36	31	DEPARTAME									RD UNI 1 ANY
37	32	DEPARTAME									PUBLI 1 ANY
38	33	DEPARTAME									CAN P 1 ANY
39	34	DEPARTAME									CAN S 1 ANY
40	35	DEPARTAME									R & FF 1 ANY
41	36	DEPARTAME									SOCII 1 ANY
42	37	DEPARTAME									JP, S.A 1 ANY
43	38	DEPARTAME									NA PR 1 ANY
44	38	DEPARTAME									NA PR 1 ANY
45	38	DEPARTAME									NA PR 1 ANY
46	38	DEPARTAME									NA PR 1 ANY
47	38	DEPARTAME									NA PR 1 ANY
48	39	DEPARTAME									CERO, 15 DIES
49	40	DEPARTAME	INSTITUT D'II	27/12/16	I16.0009CH - SERVEIS	OBERT		519097.57	384525.67	ILUNION CEE	2 ANYS

Aplicación de filtros de coincidencia

Sin embargo, más allá de ver máximos, mínimos y filtrar por valores concretos, las opciones de análisis por medio de los filtros son ciertamente limitadas.

## 4. Análisis de datos

### 4.3. Herramientas y mecanismos de análisis

#### 4.3.3. Las tablas dinámicas

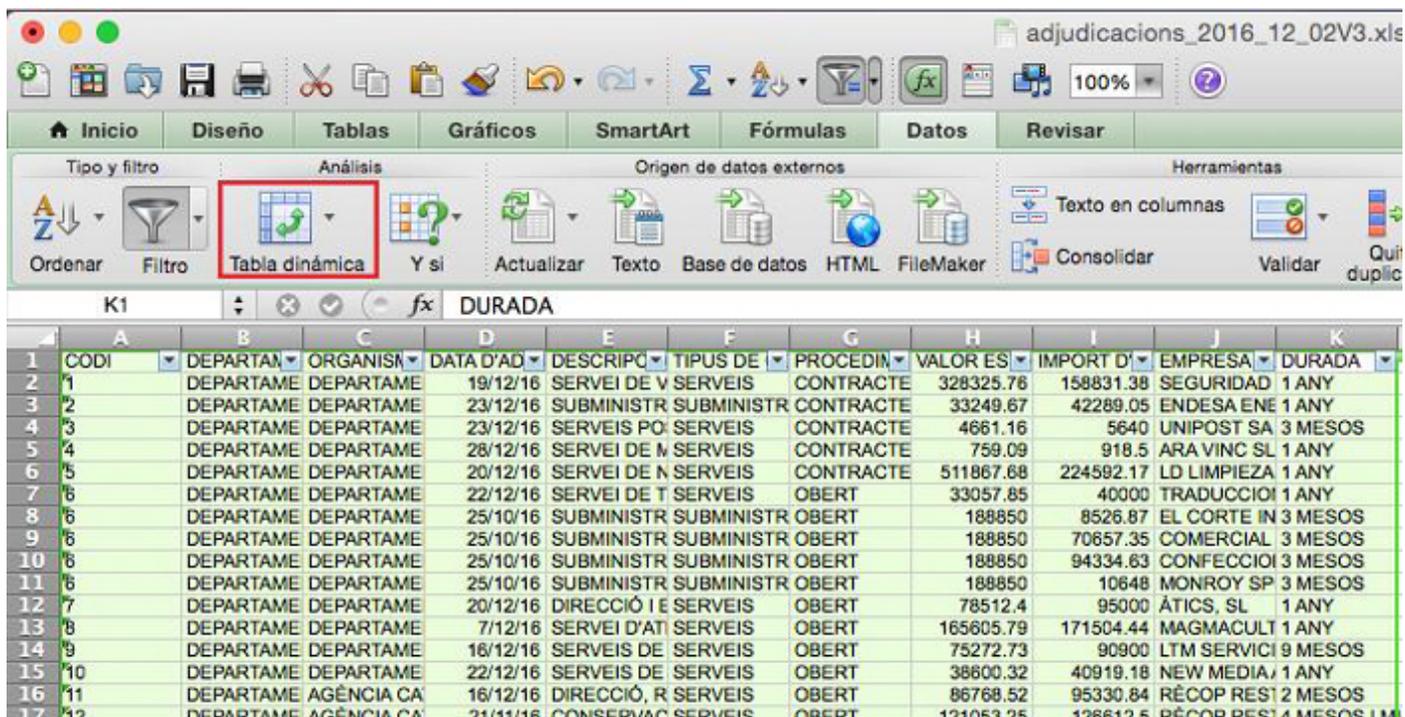
En este caso, trabajaremos con el archivo:  [Adjudicacions\\_TaulaDinamica](#).

##### 4.3.3.1. Las tablas

La herramienta por excelencia a la hora de llevar a cabo un análisis de datos con un gestor de hojas de cálculo es la tabla dinámica. La tabla dinámica es una de las funciones básicas de Excel, y permite reordenar los datos como más interese para poder analizarlos desde ángulos diferentes.

Ahora seguiremos paso a paso la creación de una tabla dinámica para entender más su funcionamiento. Para hacerlo, recuperamos de nuevo el ejemplo de las adjudicaciones públicas.

Con el archivo abierto, tendremos que seleccionar de nuevo las columnas del *data set* y, en el menú de «Datos», hacer clic en la opción «Tabla dinámica» de la manera siguiente:

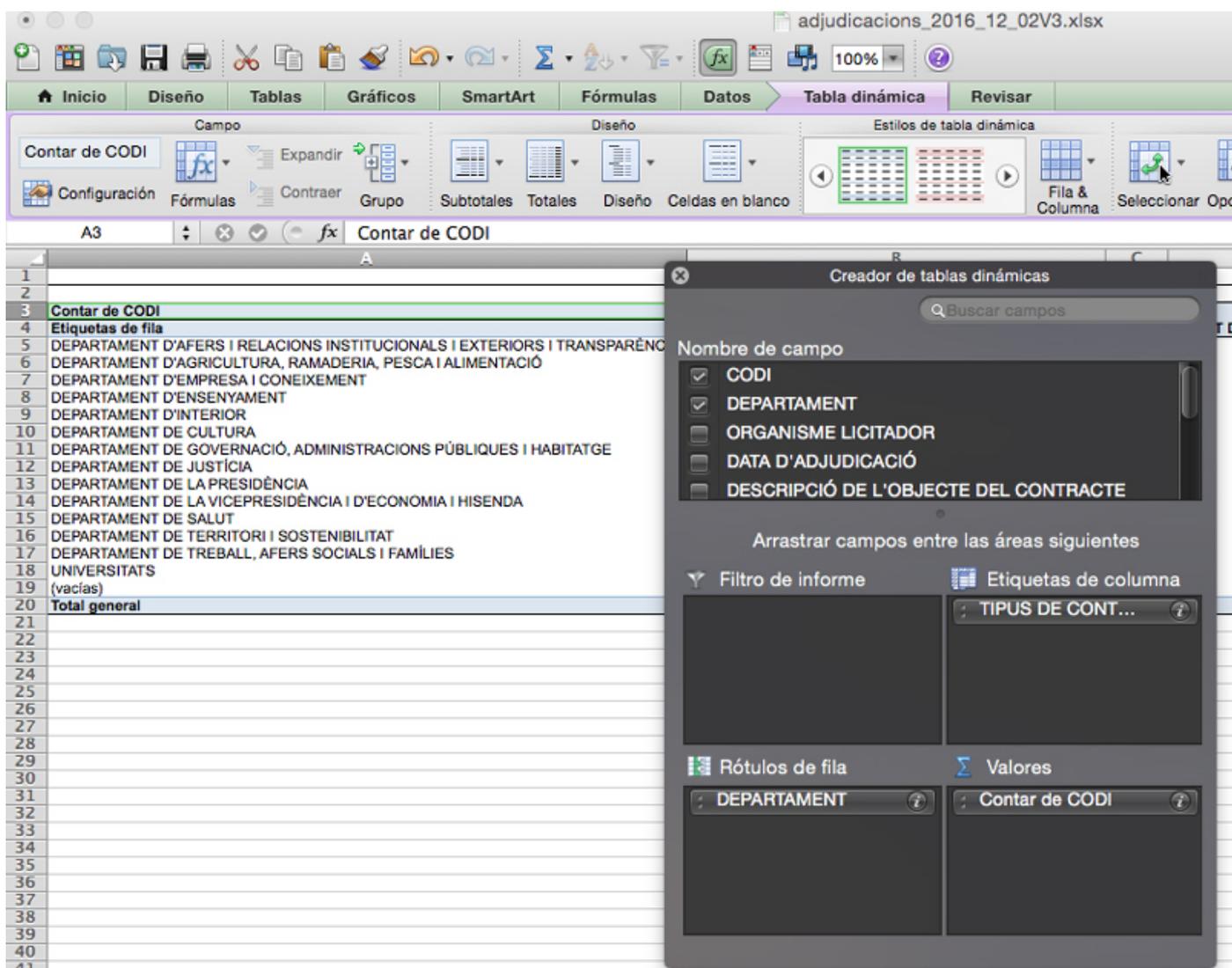


	A	B	C	D	E	F	G	H	I	J	K
1	CODI	DEPARTAM	ORGANISM	DATA D'AD	DESCRIPC	TIPUS DE	PROCEDIN	VALOR ES	IMPORT D'	EMPRESA	DURADA
2	1	DEPARTAME	DEPARTAME	19/12/16	SERVEI DE V	SERVEIS	CONTRACTE	328325.78	158831.38	SEGURIDAD	1 ANY
3	2	DEPARTAME	DEPARTAME	23/12/16	SUBMINISTR	SUBMINISTR	CONTRACTE	33249.67	42289.05	ENDESA ENE	1 ANY
4	3	DEPARTAME	DEPARTAME	23/12/16	SERVEIS PO	SERVEIS	CONTRACTE	4661.16	5640	UNIPOST SA	3 MESOS
5	4	DEPARTAME	DEPARTAME	28/12/16	SERVEI DE M	SERVEIS	CONTRACTE	759.09	918.5	ARA VINC SL	1 ANY
6	5	DEPARTAME	DEPARTAME	20/12/16	SERVEI DE N	SERVEIS	CONTRACTE	511867.68	224592.17	LD LIMPIEZA	1 ANY
7	6	DEPARTAME	DEPARTAME	22/12/16	SERVEI DE T	SERVEIS	OBERT	33057.85	40000	TRADUCCIOI	1 ANY
8	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	8526.87	EL CORTE IN	3 MESOS
9	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	70657.35	COMERCIAL	3 MESOS
10	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	94334.63	CONFECCIOI	3 MESOS
11	6	DEPARTAME	DEPARTAME	25/10/16	SUBMINISTR	SUBMINISTR	OBERT	188850	10648	MONROY SP	3 MESOS
12	7	DEPARTAME	DEPARTAME	20/12/16	DIRECCIÓ I E	SERVEIS	OBERT	78512.4	95000	ÀTICS, SL	1 ANY
13	8	DEPARTAME	DEPARTAME	7/12/16	SERVEI D'AT	SERVEIS	OBERT	165605.79	171504.44	MAGMACULT	1 ANY
14	9	DEPARTAME	DEPARTAME	16/12/16	SERVEIS DE	SERVEIS	OBERT	75272.73	90900	LTM SERVICI	9 MESOS
15	10	DEPARTAME	DEPARTAME	22/12/16	SERVEIS DE	SERVEIS	OBERT	38600.32	40919.18	NEW MEDIA	1 ANY
16	11	DEPARTAME	AGÈNCIA CA	16/12/16	DIRECCIÓ, R	SERVEIS	OBERT	86768.52	95330.84	RÈCOP RES	12 MESOS
17	12	DEPARTAME	AGÈNCIA CA	21/11/16	CONSERVAC	SERVEIS	OBERT	121653.25	138812.5	RÈCOP RES	14 MESOS

##### Creación de una tabla dinámica

Una vez que hacemos clic en la creación de una nueva tabla dinámica, se creará automáticamente una nueva página en la hoja de cálculo. Es posible que, por defecto, la tabla esté medio llena o completamente vacía, pero antes de proceder al análisis, habrá que entender bien para qué sirve cada opción.

De entrada, veremos que en la parte superior derecha de la hoja de cálculo está la tabla dinámica, y a la izquierda, el «Creador de tablas dinámicas», que es la herramienta que utilizaremos para añadir o quitar campos de la tabla dinámica.



Visualización de la tabla dinámica a la derecha y el creador de tablas dinámicas a la izquierda

Veamos, opción por opción, para qué sirve cada uno de los espacios.

- **Nombre del campo:** en este apartado, si hemos hecho bien los pasos, tendríamos que encontrar el nombre de todas las columnas de nuestra tabla original. En el caso de la hoja de cálculo de adjudicaciones, la tabla se ha ordenado correctamente y en este apartado podemos ver las columnas del conjunto de datos (código, departamento, organismo licitador, etc.).
- **Filtro de informe:** aquí arrastraremos los campos sobre los que nos interese aplicar filtros sobre los datos.
- **Etiquetas de columna:** en este apartado, arrastraremos los campos que queramos situar como columnas de la nueva tabla que estamos creando a partir de la inicial.
- **Rótulos de fila:** es lo mismo que en el caso anterior, pero para las filas. Aquí arrastraremos los campos que queremos que se visualicen en las filas.
- **Valores:** una vez que tengamos la tabla estructurada en filas y en columnas, debemos decidir qué valores queremos mostrar. Para hacerlo, tendremos que lanzar los valores en este apartado.

Para introducir o sacar los campos en cada una de las opciones, habrá que arrastrar las columnas del apartado «Nombre de campos» al apartado que corresponda o interese. En el supuesto de que la tabla ya venga medio llena por defecto, o que queramos modificar una tabla ya elaborada, para quitar los campos lanzados tendremos que arrastrarlos de cada una de las opciones fuera del generador de tablas dinámicas.

A medida que arrastramos para añadir o quitar campos, veremos cómo la tabla de la hoja de Excel se modifica según lo que estemos haciendo.

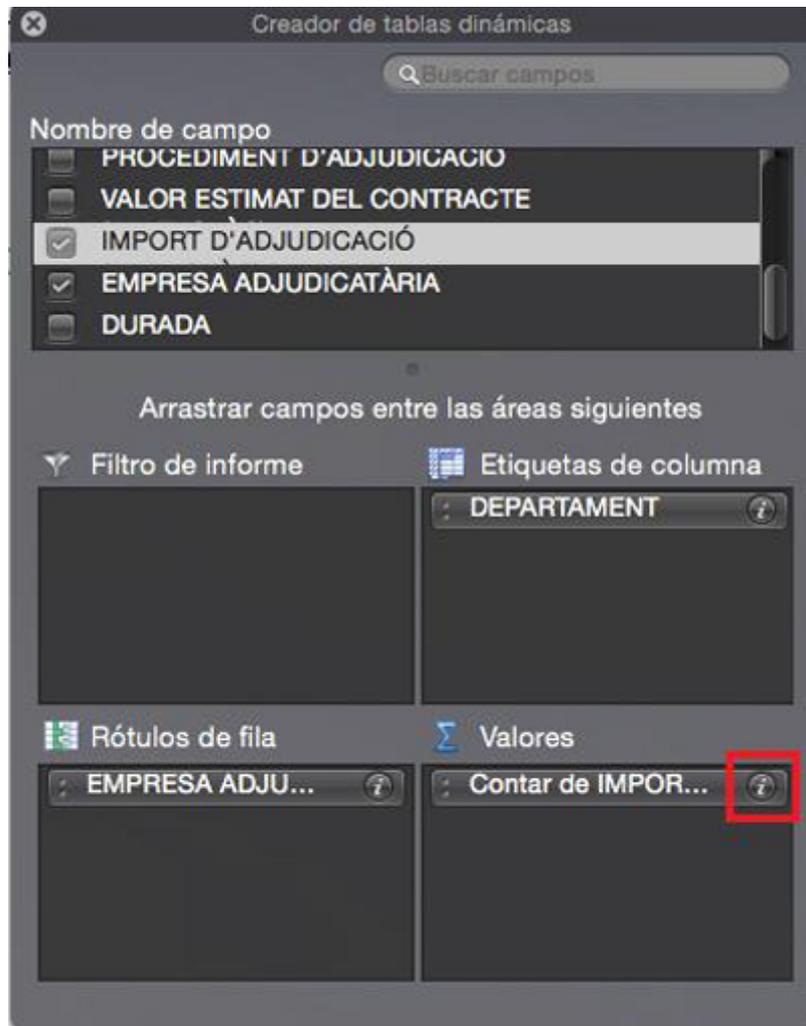
### 4.3.3.2. Un ejemplo práctico

Intentemos, entonces, responder a las preguntas iniciales siguiendo el ejemplo que hemos estado tomando hasta ahora, el del Excel de las adjudicaciones de contratos públicos.

a) ¿Cuál es la empresa que se lleva más contratos en cada departamento?

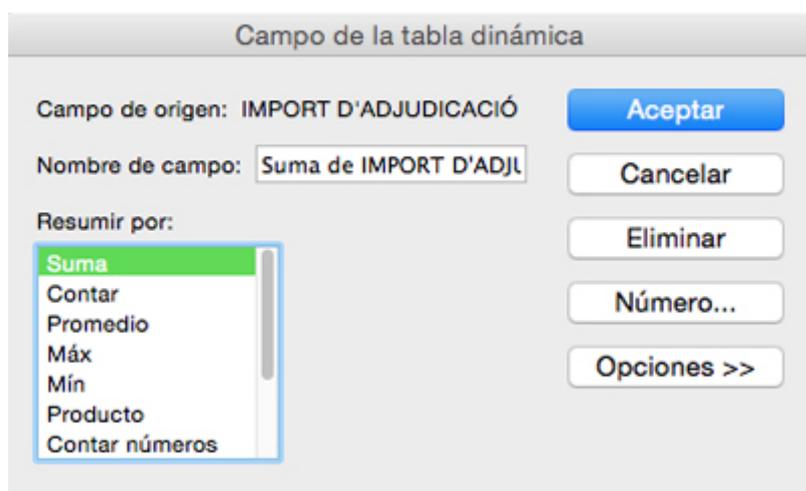
Para responder a esta pregunta, tendremos que hacer una tabla dinámica en la cual, en las filas (o letreros de fila), esté el nombre de «empresa adjudicataria», en las columnas (o campos de fila) esté el «departamento», y en el campo de valores deberíamos lanzar el «importe de adjudicación».

A la hora de lanzar los valores, es fácil que por defecto cuente los registros en lugar de sumarlos. Para modificar esta opción, habrá que hacer clic en el botón de la «i» que hay dentro del campo que hemos lanzado.



Creación de una tabla dinámica

Entonces, saldrá un menú emergente que nos permitirá cambiar la opción «Contar» por muchas otras, también «Sumar», que es la que nos interesa esta vez.



## Menú emergente para crear tablas dinámicas

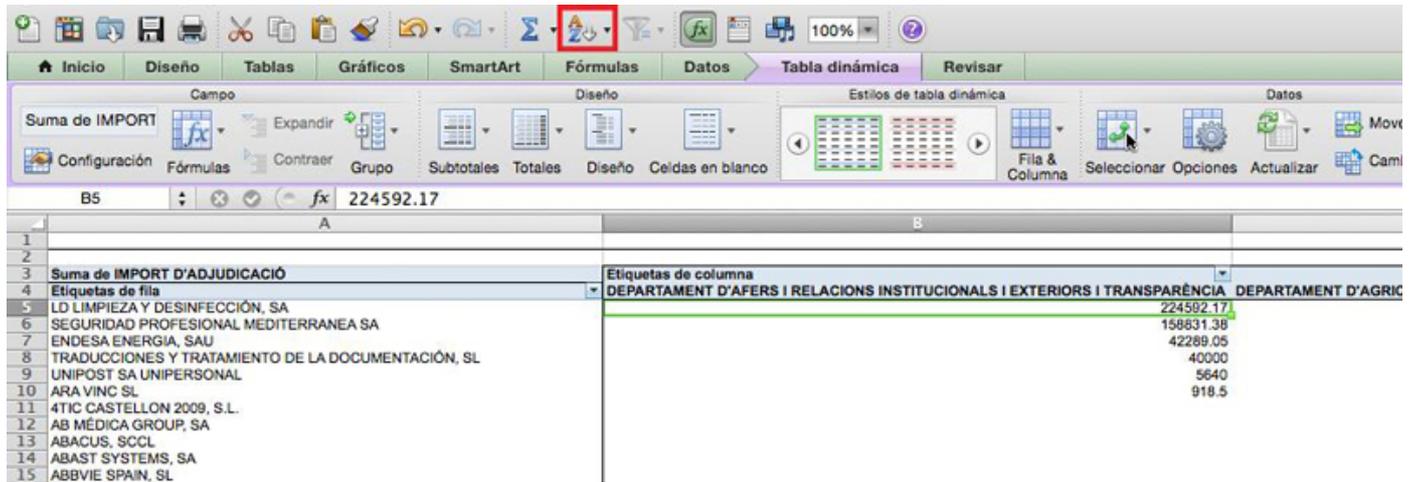


Tabla dinámica resultante

b) ¿Cuál es la media de los importes de las licitaciones de suministro por departamento?

Para resolver esta pregunta, tendremos que hacer una tabla dinámica en la cual las filas correspondan a los departamentos, y los valores correspondan al «importe de la adjudicación». En el apartado de filtros, elegiremos la opción «Tipo de contrato» para elegir únicamente aquellos que pertenezcan a contratos de suministro. Finalmente, para que nos salga directamente la media, y no el recuento o la suma, haremos clic en la «i» que aparece a la derecha de valores, y marcaremos la opción «Media» en el menú emergente. Finalmente, la tabla y el generador quedarán de la siguiente manera:

TIPUS DE CONTRACTE	SUBMINISTRAMENTS
<b>Promedio de IMPORT D'ADJUDICACIÓ</b>	
<b>Etiquetas de fila</b>	<b>Total</b>
DEPARTAMENT D'ENSENYAMENT	7185861.865
DEPARTAMENT D'EMPRESA I CONEIXEMENT	1727498.164
DEPARTAMENT DE TERRITORI I SOSTENIBILITAT	914856.7808
DEPARTAMENT DE LA VICEPRESIDÈNCIA I D'ECONOMIA I HISENDA	546289.07
DEPARTAMENT D'INTERIOR	209632.5
DEPARTAMENT DE JUSTÍCIA	118662.96
DEPARTAMENT DE SALUT	86308.76493
DEPARTAMENT D'AGRICULTURA, RAMADERIA, PESCA I ALIMENTACIÓ	46041.7125
DEPARTAMENT D'AFERS I RELACIONS INSTITUCIONALS I EXTERIORS I TRANSPARÈNCIA	42289.05
UNIVERSITATS	42163.05429
<b>Total general</b>	<b>321984.6089</b>

Crearor de tablas dinámicas

Nombre de campo

- CODI
- DEPARTAMENT
- ORGANISME LICITADOR
- DATA D'ADJUDICACIÓ
- DESCRIPCIÓ DE L'OBJECTE DEL CONTRACTE

Arrastrar campos entre las áreas siguientes

Filtro de informe      Etiquetas de columna

TIPUS DE CONT...      ?

Rótulos de fila      Valores

DEPARTAMENT      ?      Promedio de IMP...      ?

Creación de una nueva tabla dinámica

Ahora, supongamos que nos llama la atención el resultado del Departamento de Enseñanza, el más elevado de todos en cuanto a la media. Si hacemos doble clic en la columna de los valores, la que está señalada en color verde en la anterior captura de pantalla, nos aparecerán en una nueva hoja todos los valores que están vinculados a este valor, para que podamos navegar y explorarlos con más detalle si nos interesa.

## 4. Análisis de datos

### 4.3. Herramientas y mecanismos de análisis

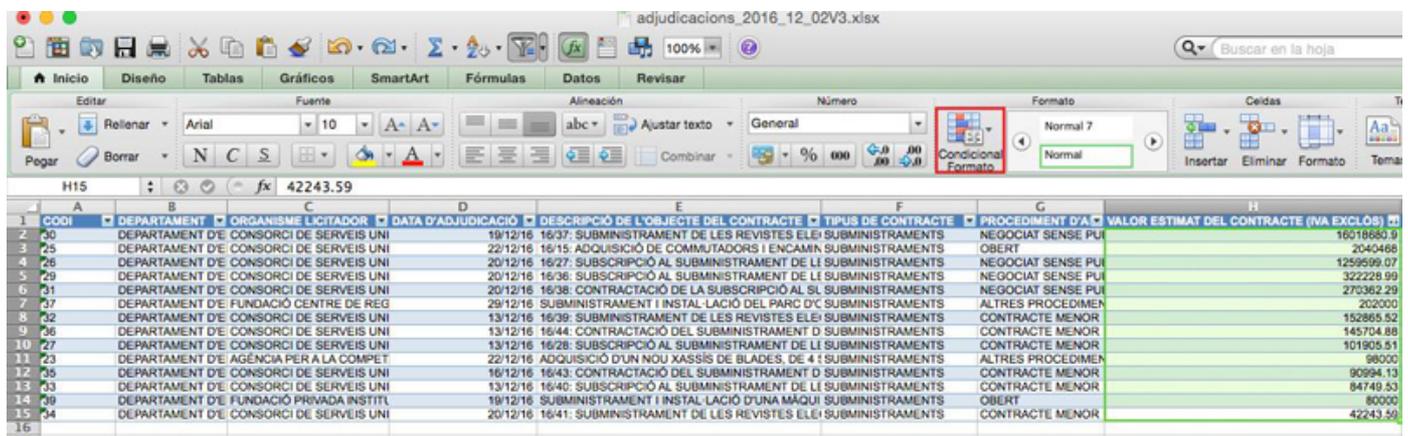
#### 4.3.4. Filtros condicionales y otras opciones avanzadas

Sin embargo, más allá de las tablas dinámicas, también hay otros factores de análisis que se pueden aplicar a cualquier tabla, sea dinámica o no.

El formato condicional permite aplicar colores a las celdas según condiciones específicas, como por ejemplo valores duplicados, rangos de mayor a menor, y resaltar valores concretos. La ventaja de esta prestación es que si los valores cambian, las reglas marcadas también se reajustan.

Continuando con el ejemplo que hemos trabajado hasta ahora, veremos cómo podemos hacerlo. Imaginemos que nos interesa profundizar en los contratos del Departamento de Empresa y Conocimiento efectuados durante este periodo, y concretamente nos interesa ver de una manera más visual el rango de precios del «valor estimado del contrato».

Para hacerlo, tendremos que seleccionar la columna de valor que nos interesa y abrir el menú «Condicional Formato», situado en la barra de herramientas.

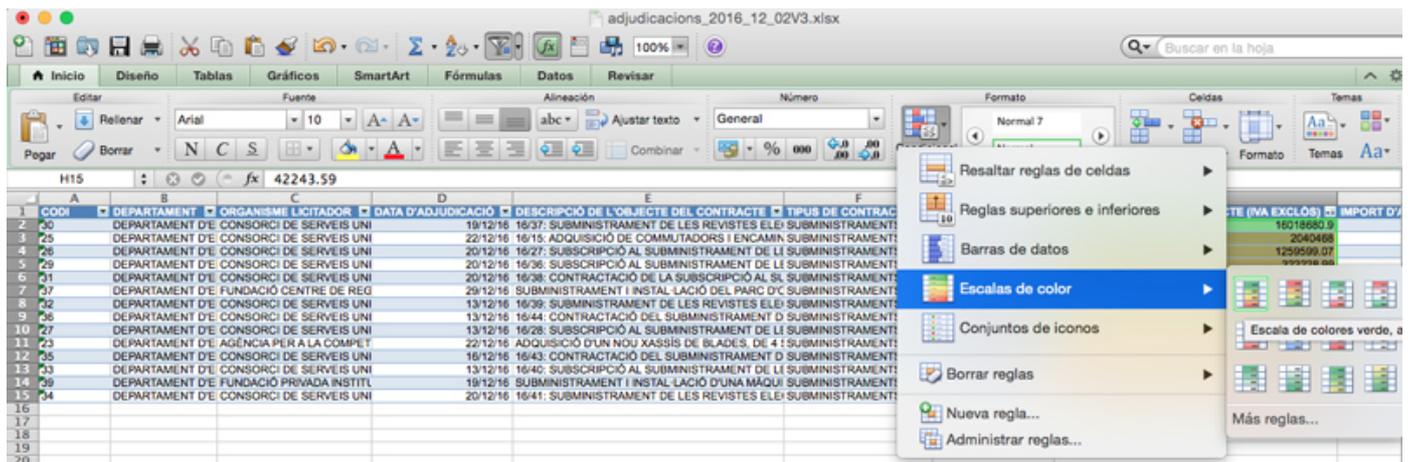


The screenshot shows the Excel interface with the 'Formato' menu open. The 'Condicional Formato' option is highlighted. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H
	COOI	DEPARTAMENT	ORGANISME LICITADOR	DATA D'ADJUDICACIÓ	DESCRIPCIÓ DE L'OBJECTE DEL CONTRATE	TIPUS DE CONTRATE	PROCEDIMENT D'A	VALOR ESTIMAT DEL CONTRATE (IVA EXCLÓS)
2	20	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	19/12/16	16/37: SUBMINISTRAMENT DE LES REVISTES ELEI	SUBMINISTRAMENTS	NEGOCIAT SENSE PU	16018680.19
3	25	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	22/12/16	16/15: ADQUISICIÓ DE COMUTADORS I ENCAMI	SUBMINISTRAMENTS	OBERT	2340486
4	26	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	20/12/16	16/27: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	NEGOCIAT SENSE PU	1259599.07
5	29	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	20/12/16	16/36: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	NEGOCIAT SENSE PU	322228.99
6	31	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	20/12/16	16/38: CONTRACTACIÓ DE LA SUBSCRIPCIÓ AL SI	SUBMINISTRAMENTS	NEGOCIAT SENSE PU	270362.29
7	37	DEPARTAMENT D'E	FUNDACIÓ CENTRE DE REG	29/12/16	SUBMINISTRAMENT I INSTAL·LACIÓ DEL PARC D'C	SUBMINISTRAMENTS	ALTRES PROCEDIMEN	202000
8	32	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	13/12/16	16/39: SUBMINISTRAMENT DE LES REVISTES ELEI	SUBMINISTRAMENTS	CONTRACTE MENOR	152865.52
9	26	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	13/12/16	16/44: CONTRACTACIÓ DEL SUBMINISTRAMENT D	SUBMINISTRAMENTS	CONTRACTE MENOR	145704.88
10	27	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	13/12/16	16/28: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	CONTRACTE MENOR	101905.51
11	23	DEPARTAMENT D'E	AGÈNCIA PER A LA COMPET	22/12/16	ADQUISICIÓ D'UN NOU XASSIS DE BLADES DE 4 I	SUBMINISTRAMENTS	ALTRES PROCEDIMEN	9800
12	25	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	16/12/16	16/43: CONTRACTACIÓ DEL SUBMINISTRAMENT D	SUBMINISTRAMENTS	CONTRACTE MENOR	90994.13
13	33	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	13/12/16	16/40: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	CONTRACTE MENOR	84749.53
14	29	DEPARTAMENT D'E	FUNDACIÓ PRIVADA INSTITL	19/12/16	SUBMINISTRAMENT I INSTAL·LACIÓ D'UNA MÁQUI	SUBMINISTRAMENTS	OBERT	8000
15	34	DEPARTAMENT D'E	CONSORCI DE SERVEIS UNI	20/12/16	16/41: SUBMINISTRAMENT DE LES REVISTES ELEI	SUBMINISTRAMENTS	CONTRACTE MENOR	42243.59

Visualización de la aplicación de filtros condicionales (primer paso)

Con el menú desplegado, podremos seleccionar la opción que más nos interese. En este caso, una sencilla escala de color ya nos permitirá ver cuál es el rango de precios, cuáles están más distanciados, y cuáles son más similares.



The screenshot shows the 'Formato' menu open, with the 'Escala de color' option selected. The spreadsheet data is the same as in the previous image.

Visualización de la aplicación de filtros condicionales (segundo paso)

Una vez seleccionada la opción, la columna nos quedará pintada automáticamente. Si borramos el contenido de la celda, el filtro de color continuará apareciendo. Si en algún momento queremos borrar las reglas de colores, habrá que volver al menú «Condicional Formato» y hacer clic en la opción «Borrar reglas».

	A	B	C	D	E	F	G	H
1	CODI	DEPARTAMENT	ORGANISME LICITADOR	DATA D'ADJUDICACIÓ	DESCRIPCIÓ DE L'OBJECTE DEL CONTRACTE	TIPUS DE CONTRACTE	PROCEDIMENT D'A	VALOR ESTIMAT DEL CONTRACTE (IVA EXCLÓS)
2	30	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		19/12/16	16/37: SUBMINISTRAMENT DE LES REVISTES ELEI SUBMINISTRAMENTS		NEGOCIAT SENSE PUI	16018666.9
3	25	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		22/12/16	16/15: ADQUISICIÓ DE COMMUTADORS I ENCAMIN SUBMINISTRAMENTS		OBERT	2040466
4	28	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		20/12/16	16/27: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LE SUBMINISTRAMENTS		NEGOCIAT SENSE PUI	1259599.07
5	29	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		20/12/16	16/36: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LE SUBMINISTRAMENTS		NEGOCIAT SENSE PUI	322228.99
6	31	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		20/12/16	16/38: CONTRACTACIÓ DE LA SUBSCRIPCIÓ AL SL SUBMINISTRAMENTS		NEGOCIAT SENSE PUI	270362.29
7	37	DEPARTAMENT D'E FUNDACIÓ CENTRE DE REG		29/12/16	SUBMINISTRAMENT I INSTAL·LACIÓ DEL PARC D'C SUBMINISTRAMENTS		ALTRES PROCEDIMEN	202000
8	32	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		13/12/16	16/39: SUBMINISTRAMENT DE LES REVISTES ELEI SUBMINISTRAMENTS		CONTRACTE MENOR	152665.52
9	36	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		13/12/16	16/44: CONTRACTACIÓ DEL SUBMINISTRAMENT D SUBMINISTRAMENTS		CONTRACTE MENOR	145704.86
10	27	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		13/12/16	16/28: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LE SUBMINISTRAMENTS		CONTRACTE MENOR	101905.51
11	23	DEPARTAMENT D'E AGÈNCIA PER A LA COMPET		22/12/16	ADQUISICIÓ D'UN NOU XASSIS DE BLADES, DE 4 I SUBMINISTRAMENTS		ALTRES PROCEDIMEN	98000
12	35	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		16/12/16	16/43: CONTRACTACIÓ DEL SUBMINISTRAMENT D SUBMINISTRAMENTS		CONTRACTE MENOR	90994.13
13	33	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		13/12/16	16/40: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LE SUBMINISTRAMENTS		CONTRACTE MENOR	84749.83
14	39	DEPARTAMENT D'E FUNDACIÓ PRIVADA INSTITU		19/12/16	SUBMINISTRAMENT I INSTAL·LACIÓ D'UNA MÀQUI SUBMINISTRAMENTS		OBERT	80000
15	34	DEPARTAMENT D'E CONSORCI DE SERVEIS UNI		20/12/16	16/41: SUBMINISTRAMENT DE LES REVISTES ELEI SUBMINISTRAMENTS		CONTRACTE MENOR	42243.99

Visualización de la aplicación de filtros condicionales (tercer paso)

Este solo es un ejemplo de filtro de colores que nos conviene y nos interesa esta vez. De todos modos, para otras veces nos puede interesar más crear reglas específicas para buscar o resaltar un valor determinado. El resto de las posibilidades están todas juntas en el menú de «Condiciona l Formato».

## 4. Análisis de datos

### 4.3. Herramientas y mecanismos de análisis

#### 4.3.5. Utilizar descriptores estadísticos típicos

##### 4.3.5.1. Introducción

En este subapartado, trabajaremos con el archivo:  [Formulas](#).

A la hora de trabajar con el análisis de datos, hay una serie de fórmulas y descriptores estadísticos básicos que se utilizan prácticamente cada vez que se buscan historias entre un conjunto de información almacenada en una hoja de cálculo.

La estructura de las fórmulas de Excel es bastante básica:

- a) Siempre se empieza con un signo =, que Excel interpreta como que a continuación habrá una fórmula.
- b) En segundo lugar, escribiremos el nombre de la fórmula. A menos que estemos haciendo un cálculo personalizado, las fórmulas ya están registradas en el programa y, por lo tanto, enseguida que empecemos a escribirla el programa nos hará sugerencias.
- c) Después del nombre de la fórmula, se abre paréntesis para:
  - Indicar el rango de datos sobre el cual se quiere aplicar el cálculo. Es el caso de las fórmulas más sencillas, en las cuales se podrá indicar un rango concreto o unas columnas enteras concretas. Por ejemplo, de la casilla A2 a la A34, y lo escribiríamos (A2:A34), y de la columna D a la F, escrito así (D:F).
  - Cumplir la estructura requerida por Excel, en el caso de las fórmulas con una sintaxis más compleja, para que la fórmula se resuelva con éxito.

Las fórmulas se pueden arrastrar y aplicar de manera sencilla en todas las filas de una columna. Al igual que hemos visto en apartados anteriores, nos tendremos que posicionar en el extremo derecho inferior de la celda, y hacer doble clic o arrastrar hasta la última casilla a la que queramos arrastrar la celda, y los valores de las casillas seleccionadas irán cambiando automáticamente. Si hubiera alguna casilla fija que quisiéramos mantener, podríamos hacerlo con el símbolo \$.

### Ejemplo

Por ejemplo, si queremos fijar la casilla A24, si escribimos \$A24 fijaremos la columna, pero no la fila; si escribimos A\$24 fijaremos la fila, pero no la columna; y si escribimos \$A\$24, la fijaremos del todo.

Para poder seguir con más detalle las fórmulas presentadas a continuación, las encontraréis puestas en práctica en el documento Excel adjunto de este apartado. Este archivo hace referencia al número de matriculados en el curso de 4.º de ESO en todos los institutos públicos de Cataluña para el año 2012 y 2013. Encontraréis dos hojas, una para las fórmulas básicas y una segunda para las avanzadas.

#### 4.3.5.2. Fórmulas básicas

##### 1) Sumar

Permite sumar un rango de datos numéricos. Para aplicarla, hay que escribir =SUM(), o =SUMA(), dependiendo del idioma del programa, y seleccionar dentro del paréntesis el rango de datos que se quieren sumar.

##### 2) Contar

Permite contar el número de valores numéricos que hay en un rango de datos con la fórmula =CONTAR(), o =COUNT(), según el idioma.

En cambio, si lo que nos interesa es contar valores numéricos y alfanuméricos, utilizaremos la fórmula =CONTAR.A(), o =COUNTA.

### 3) Media

Calcula la media de un rango determinado de datos. La fórmula es =AVERAGE(), o =PROMEDIO(), según el idioma preestablecido del programa.

### 4) Mediana

Es un cálculo que ordena de mayor a menor todo el rango de datos seleccionado, y retorna aquel dato que hay exactamente en el punto medio de todo el rango. Para aplicarla, se utiliza =MEDIAN, o =MEDIANA.

### 5) Moda

La moda es una métrica que representa el valor más repetido entre un rango determinado de datos. La fórmula para calcularla es =MODO(), si el documento Excel es en inglés, o =MODA() y =MODA.UNO(), según la versión en castellano.

### 6) Valor máximo

Para encontrar el valor más alto de un rango de datos determinado, la fórmula es =MAX, tanto en la versión en castellano como en inglés.

### 7) Valor mínimo

Para encontrar el valor más bajo de un rango de datos determinado, la fórmula es =MIN, tanto en la versión en castellano como en inglés.

### 8) Porcentaje sobre el total

Los porcentajes son importantes porque aportan información valiosa sobre el contexto que rodea a un dato concreto. Si queremos calcular el porcentaje de un valor respecto al total de la variable, utilizaremos la regla de tres.

## Ejemplo

Por ejemplo, volviendo al caso de los institutos públicos, si queremos saber qué porcentaje representan el número de estudiantes de 4.º de ESO del municipio de Alcanar, 74, respecto al total matriculados en Cataluña en el 2012, un total de 17.635, el cálculo que tendremos que hacer será: total alumnos de Alcanar / total alumnos de Cataluña \* 100.

Hay que tener en cuenta que si ya hemos multiplicado por cien, no habrá que especificar que los valores son porcentaje en el formato de número. Es importante recordarlo, porque si cambiamos el formato de número a porcentaje, automáticamente los valores se multiplicarán por cien otra vez.

## 4.3.5.3. Fórmulas avanzadas

### 1) Contar los valores en blanco

Si queremos contar el número de celdas en blanco que hay en un rango determinado, la fórmula será =CONTAR.BLANCO(), para la versión en castellano, y =COUNTBLANK(), para la versión en inglés.

### 2) Contar condicional

La siguiente fórmula sirve para contar solo aquellos valores del rango que cumplen una condición: =CONTAR.SI(), o =COUNTIF().

## Ejemplo

Por ejemplo, si queremos contar todos aquellos valores de una columna que son mayores de 15, la fórmula correcta sería: =CONTAR.SI(B:B;">15").

### 3) Sumar condicional

Funciona del mismo modo que la operación anterior, pero en este caso, suma. Por lo tanto, solo suman aquellos valores del rango que cumplan una condición: =SUMAR.SI(), o =SUMIF().

#### Ejemplo

Por ejemplo, si queremos sumar todos aquellos valores de una columna que son mayores de 15, la fórmula correcta sería: =SUMAR.SI(B:B;">15").

### 4) Ver si un registro cumple dos o más condiciones

Esta operación permite ver si una fila cumple condiciones en columnas diferentes. La fórmula para aplicar es =Y(), en español, y =AND(), en inglés.

#### Ejemplo

Por ejemplo, volviendo al caso de los institutos, imaginémos que queremos saber cuáles tuvieron más de 50 matriculados en 4.º de ESO en el 2012 de la totalidad de institutos, más de una cincuentena en el 2013 y que, además, el crecimiento de un año al otro fue positivo. Para ver la aplicación de la fórmula, podéis abrir el archivo adjunto y hacer clic en la segunda hoja, denominada «fórmulas avanzadas». En la columna G hay la siguiente fórmula: =Y(E2>50;"F2>50;F2>0).

De este modo, si en la fila 2, la que corresponde al único instituto de Alcanar, los alumnos de 4.º de ESO del año 2014 son más de 50 (columna E), en el 2012 también son más de una cincuentena (columna F), y el crecimiento (columna G) de un año respecto al otro ha sido positivo, el resultado definitivo será «VERDADERO». En cambio, el resultado será «FALSO» tanto si no cumple ninguna, o solo cumple alguna.

Con la cruz negra que aparece en el extremo inferior derecho de la celda, arrastraremos la fórmula al resto de la columna.

### 5) Ver si un registro cumple alguna condición

Esta fórmula tiene un funcionamiento prácticamente igual al anterior, con la única diferencia de que en esta versión indicará «VERDADERO» cuando cumpla alguna de las condiciones especificadas, y solo indicará «FALSO» si no cumple ninguna.

Para aplicarla, tendremos que utilizar =O(), en la versión española, y =OR() en la inglesa. En el ejemplo anterior, quedaría formulada de la siguiente manera en la columna H: =Y(E2>50;"F2>50;F2>0).

### 6) Condicional

La función =SI() o IF() permite ver si una celda cumple una condición concreta en una columna y, en el caso de que sí o de que no, poder personalizar la respuesta.

#### Ejemplo

Volvamos al ejemplo de los alumnos de ESO. Por ejemplo, si el único instituto de Alcanar tuvo más de 30 alumnos de bachillerato en el 2012, queremos que la celda indique que está «dentro de la ratio» y, si no los ha tenido, que aparezca «fuera de la ratio».

La fórmula correcta será: =SI(D2>30; «dentro de la ratio»; «fuera de la ratio»).

## 7) Trasladar valores por una columna en común

Esta fórmula sirve para encontrar valores iguales que hay en hojas de Excel separadas. Para entenderlo mejor, será necesario abrir el Excel adjunto del número de inscritos en los institutos de Cataluña y ver la operación completa en la columna E de la hoja de «fórmulas avanzadas». Imaginémonos que queremos unir la hoja de «matriculados 2012» con la de «matriculados 2013». Para copiar las cifras de una manera sencilla, rápida y sin caer en errores, utilizaremos la fórmula =BUSCARV(), o =VLOOKUP().

## 8) Variación porcentual

La variación porcentual es un cálculo que nos permitirá saber cuál ha sido la variación a lo largo del tiempo de una variable determinada.

Se calcula de la manera siguiente:

$(\text{Valor final} - \text{valor inicial}) / \text{valor inicial} * 100$

### Ejemplo

Por ejemplo, si el número de alumnos matriculados en 4.º de ESO en los institutos públicos de Cataluña en el 2012 era de 17.635, y en el 2014 era de 17.436, la variación porcentual habrá sido de -1,2 %. Para ver con más detalle el cálculo, podéis ir directamente a la columna F de la segunda hoja, denominada «fórmulas avanzadas».

## 4. Análisis de datos

### 4.4. El análisis visual

#### 4.4.1. Introducción

En esta unidad, trabajaremos con el archivo:  [Grafics](#).

Ya hemos visto cómo hacer un análisis de datos mediante tablas dinámicas, cálculos estadísticos y fórmulas específicas, pero hay otro tipo de análisis que hasta ahora no hemos tratado: el análisis visual. Hacer gráficos como método de exploración de datos nos puede permitir encontrar tendencias y patrones que por otro lado serían difíciles de localizar.

Hay programas específicos, como Tableau o Quadrigram, que permiten un abanico muy amplio de gráficos con los cuales se pueden explorar los datos de una manera diferente. Sin embargo, sin ir más lejos, Excel ofrece unas posibilidades de gráficos con los cuales también se puede explorar el conjunto de datos desde otra perspectiva. Antes de observar su funcionamiento, veamos la interfaz gráfica de esta parte de Microsoft Excel.



#### Interfaz gráfica del software Excel

El primer bloque, el de tipología de gráfico, hace referencia al tipo de visualización que queremos llevar a cabo: barras, líneas, colores, áreas, dispersión. La idoneidad de cada gráfico se tratará específicamente en la asignatura de visualización.

El segundo bloque es el espacio reservado para los «minigráficos», que no trabajaremos en profundidad puesto que las opciones de los gráficos del primer bloque ya contienen todas las opciones que podemos necesitar para hacer un análisis visual.

El tercer bloque sirve para modificar la relación de los datos con el gráfico, es decir, para ampliar o restringir las filas y las columnas seleccionadas, y para intercambiar el valor que tienen en los gráficos. Finalmente, el último bloque sirve para modificar cuestiones estilísticas y de diseño, una vez tenemos hechos los gráficos.

## 4. Análisis de datos

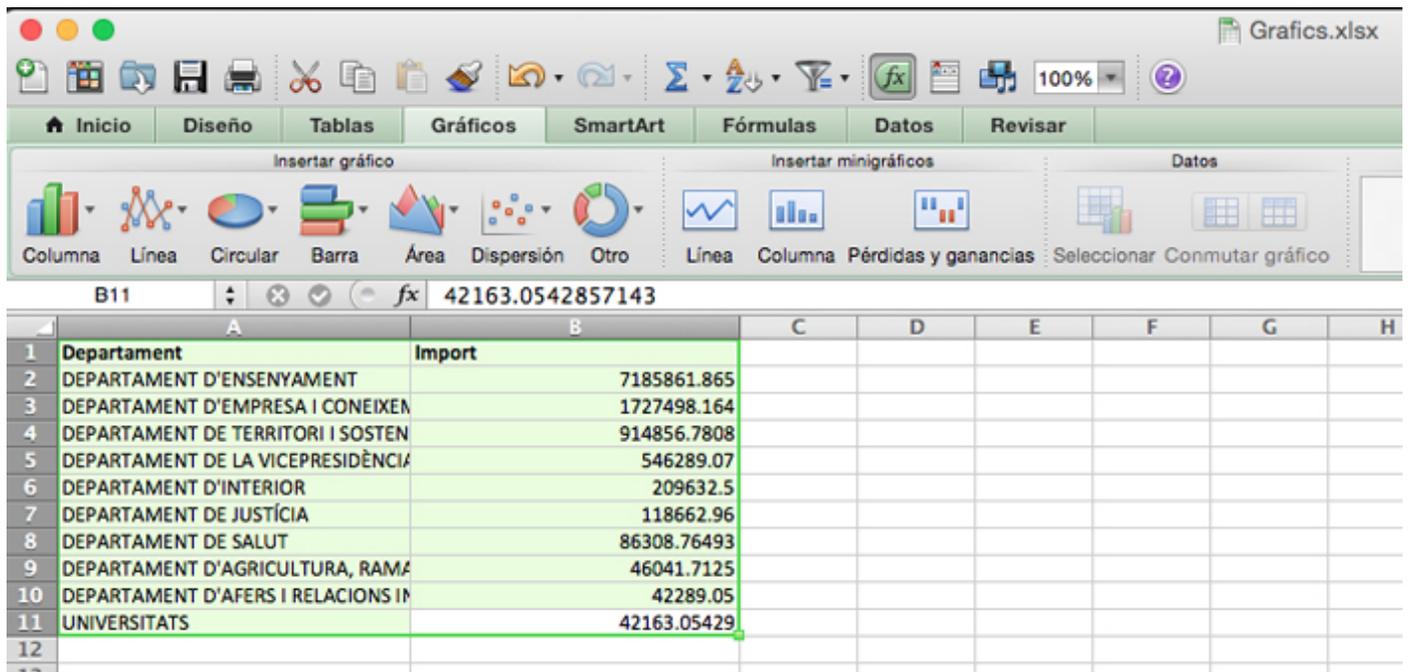
### 4.4. El análisis visual

#### 4.4.2. Un ejemplo práctico con Microsoft Excel

Para poder hacer un gráfico con Excel, hay que tener en una hoja aparte, limpios y seleccionados, los datos que pretendemos representar. Si volvemos al ejemplo de las adjudicaciones, una vez que hemos elaborado la tabla que nos permite ver qué importe en adjudicaciones se ha gastado cada departamento, también podemos ver cuál ha sido el que más y el que menos ha gastado, si ordenamos la tabla de menor a mayor. Aun así, a simple vista nos cuesta ver cuál es la relación entre todas las cifras, es decir, si son todas similares o si hay alguna que se dispara respecto al resto.

El análisis visual nos permitirá ver de una manera sencilla y rápida la relación que tienen las diferentes cifras entre ellas.

Para poder hacer el gráfico, lo primero que tenemos que hacer es seleccionar los datos que queremos representar, y hacer clic en la pestaña «Gráficos».

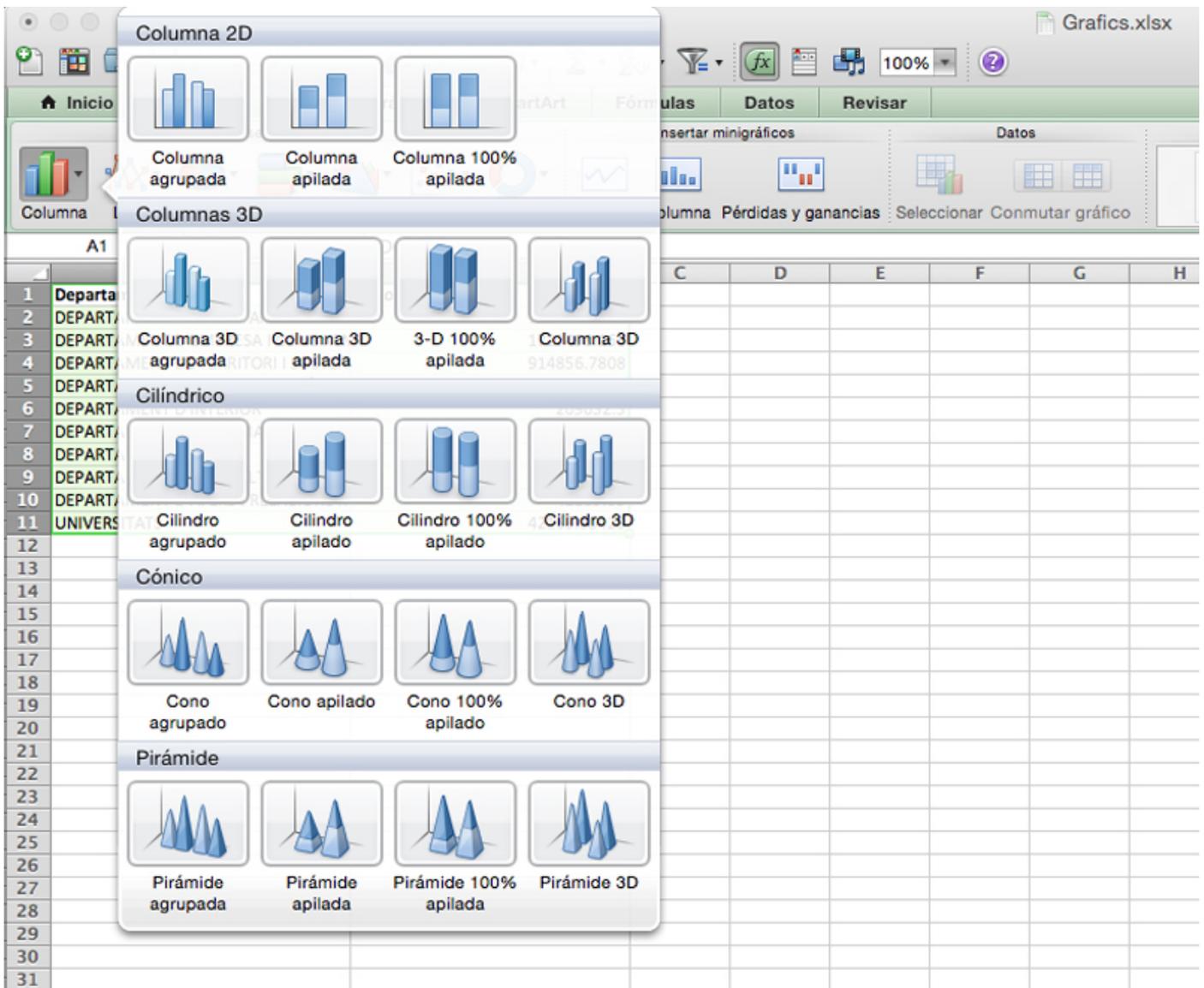


The screenshot shows the Microsoft Excel interface with the 'Gráficos' (Charts) ribbon selected. The ribbon includes options for 'Insertar gráfico' (Column, Line, Circular, Barra, Área, Dispersión, Otro) and 'Insertar minigráficos' (Line, Column, Pérdidas y ganancias). The data table below shows the following information:

	A	B	C	D	E	F	G	H
1	Departament	Import						
2	DEPARTAMENT D'ENSENYAMENT	7185861.865						
3	DEPARTAMENT D'EMPRESA I CONEIXEN	1727498.164						
4	DEPARTAMENT DE TERRITORI I SOSTEN	914856.7808						
5	DEPARTAMENT DE LA VICEPRESIDÈNCIA	546289.07						
6	DEPARTAMENT D'INTERIOR	209632.5						
7	DEPARTAMENT DE JUSTÍCIA	118662.96						
8	DEPARTAMENT DE SALUT	86308.76493						
9	DEPARTAMENT D'AGRICULTURA, RAMA	46041.7125						
10	DEPARTAMENT D'AFERS I RELACIONS IM	42289.05						
11	UNIVERSITATS	42163.05429						
12								
13								

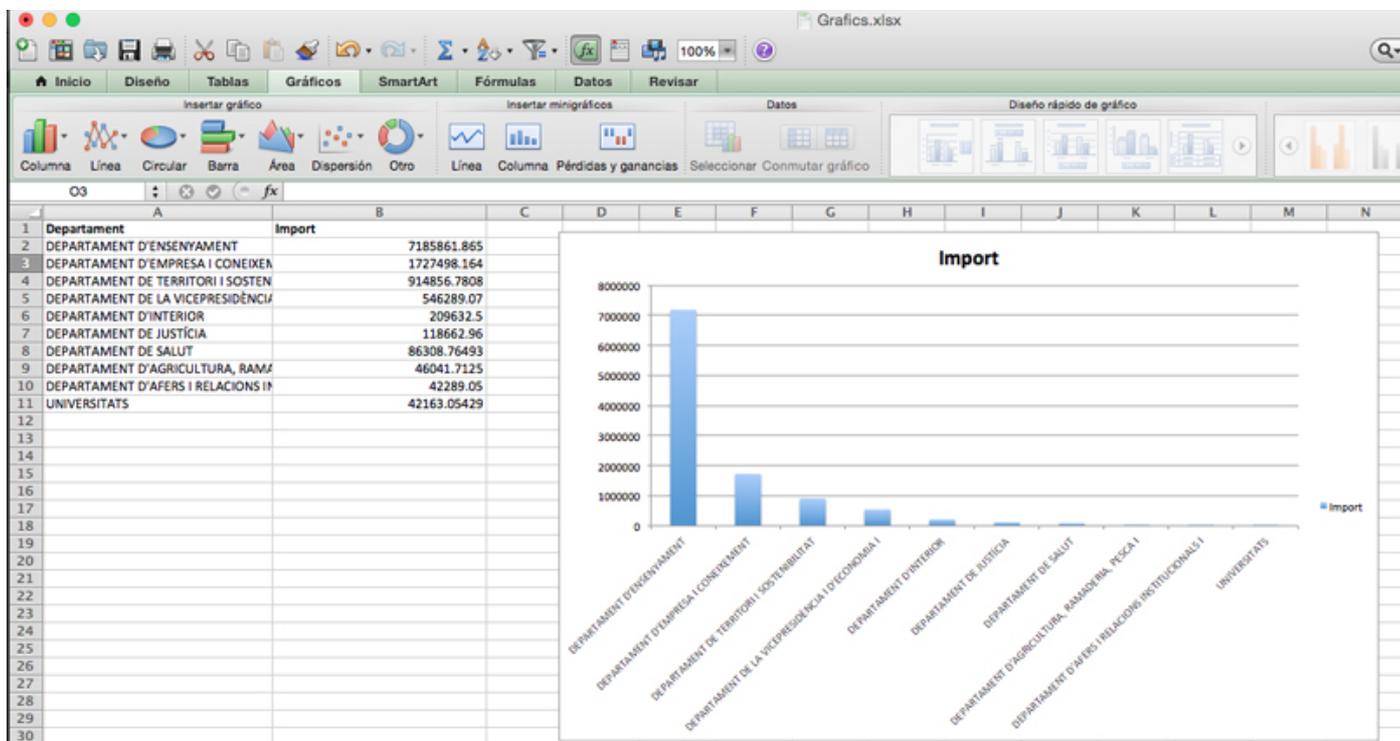
Visualización de la selección de datos por gráfica

Una vez situados encima, tendremos que elegir el gráfico pertinente para representar los datos. En este caso, lo más indicado será un gráfico de columnas o de barras, en el cual cada uno de los departamentos sea una barra o una columna. Esto nos permitirá poder ver lo mayores que son las cifras sobre el resto, y al revés. En este caso, elegimos elaborar un gráfico de columnas.



### Selección de la tipología de visualización gráfica

Dentro de la opción de la tipología de gráficos de columnas, nos aparecen por un lado opciones puramente estéticas: columnas, columnas 3D, cilindros, conos y pirámides. Si dejamos de lado esta parte más estética, que en el caso del análisis visual no aporta nada, seleccionaremos cualquiera de las tres opciones de columnas 2D. Dado que solo estamos visualizando una variable, el resultado será igual sin que importe dónde hagamos clic: un gráfico en el que cada departamento está representado en una columna. En el supuesto de que hubiera dos variables, sí deberíamos hilar más fino a la hora de elegir un gráfico u otro, y decidir si queremos las variables en barras una junto a la otra, o apiladas. De este modo, ya tendremos el gráfico elaborado.



### Visualización del gráfico elaborado

Sin hacer ningún retoque estilístico, el gráfico tendría que salir por defecto tal y como se muestra en la captura de pantalla anterior. Si hacemos clic sobre el gráfico, nos aparecerán habilitadas las opciones de los bloques tres y cuatro que hemos visto al principio de este apartado: relación con los datos y el diseño de los gráficos.

Si el gráfico no se ha mostrado tal y como deseábamos, podremos hacer variaciones con la opción «Datos», puesto que nos permitirá modificar la relación que tienen los datos con la visualización. En cuanto a las opciones estilísticas, no merece la pena entretenerse mucho. Se trata de ver qué es lo que nos revelan los datos, no de hacer un gráfico publicable. A la hora de hacer gráficos para publicar, hay muchas otras herramientas que nos permitirán hacer visualizaciones únicas para cada noticia, según los requerimientos de la historia que queramos explicar. Hoy por hoy, solo estamos haciendo un repaso al apartado de gráficos de Excel para ver cómo podemos enriquecernos del mismo y profundizar en el análisis de datos.

Así pues, con un simple gráfico como el que hemos elaborado a continuación, vemos que hay una cifra que es mucho más elevada respecto al resto, las adjudicaciones hechas por el Departamento de Enseñanza, que se disparan con diferencia a los otros. En cambio, todos los otros departamentos mantienen unos importes más similares de adjudicaciones.

Ahora veremos qué ha pasado con las adjudicaciones del Departamento de Enseñanza. ¿Se trata de un contrato muy elevado, o de varios contratos que tienen un importe muy alto? El análisis visual de esta parte la encontraréis en el ejemplo 2 del mismo archivo de Excel.

En el apartado de datos, tenemos la tabla original filtrada por los contratos del Departamento de Enseñanza. Vemos que solo hay dos, y son tan elevados porque están licitando, por medio de dos contratos, el suministro de gas natural para todos los centros de enseñanza pública. A pesar de que en este caso tendría poco sentido hacer un análisis visual, porque solo hay dos cifras y es fácil ver la relación que tienen, haremos un gráfico rápido a modo de ejemplo.

De nuevo, seleccionaremos los datos y haremos clic en la opción de gráfico que más nos convenga. En este caso, podría ser un gráfico de barras, columnas o circular. De este modo, tendremos un gráfico que nos permite ver de una manera rápida la relación entre las dos variables.

## **4. Análisis de datos**

### **4.5. Cómo encontrar noticias a partir del análisis**

#### **4.5.1. Introducción**

Los datos pueden tener papeles muy diferentes a la hora de elaborar una noticia. Pueden ser el apoyo de una historia que ya tenemos localizada, la base de una investigación o el elemento conductor de una noticia. Veamos diferentes ejemplos de noticias publicadas en diarios generalistas, en los cuales los datos tienen papeles diferentes.

## 4. Análisis de datos

### 4.5. Cómo encontrar noticias a partir del análisis

#### 4.5.2. El análisis de datos como elemento conductor de la noticia: el *Data Driven Journalism*

##### 1) *El Confidencial* | «No solo se equivoca Errejón: 164 diputados han roto la disciplina de voto en la legislatura»

Esta noticia del equipo de datos de *El Confidencial* se centra en los errores de las votaciones en el Congreso de los Diputados. Durante casi cinco meses, los periodistas se dedicaron a anotar las votaciones de cada uno de los diputados cada vez que había votaciones en el hemiciclo.

LA MITAD DE LAS VOTACIONES, CON ALGÚN ERROR

# No solo se equivoca Errejón: 164 diputados han roto la disciplina de voto en la legislatura

Los errores en las votaciones están a la orden del día en el Congreso. En cinco meses se han dado 274 casos de 'transfuguismo' parlamentario a la hora de pulsar el botón correcto

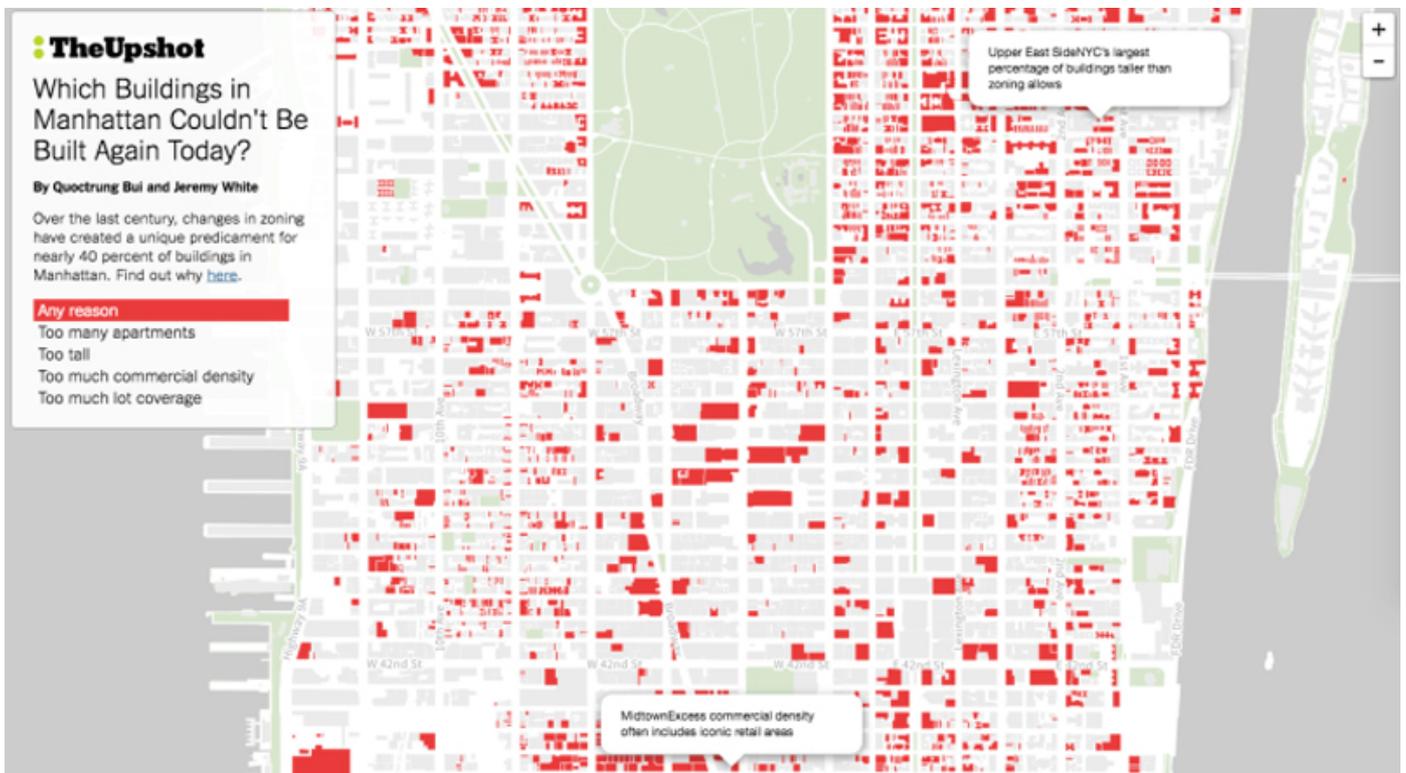


Íñigo Errejón, detrás de Pablo Iglesias durante el debate por el decreto de reforma de la estiba. (EFE)

Gracias a la constancia, esta noticia pudo ver la luz. Se trata de un ejemplo en el que los datos son el elemento central de la noticia porque se sitúan en el centro y se convierten en el valor añadido informativo del elemento que articula la publicación.

##### 2) *THE NEW YORK TIMES* | «El 40 % de los edificios de Manhattan no se podrían construir hoy»

Con la legislación vigente en la mano, y con una base de datos sobre los edificios que hay en Manhattan, los periodistas del *New York Times* averiguaron que el 40 % de las edificaciones construidas hoy día no estarían permitidas.



### 3) *La Vanguardia* | «La cúpula de TMB cobijó a tres directivos “doscientosmilleuristas” durante la crisis»

Esta vez, mediante una petición de información pública, se obtuvieron los datos referentes a los salarios de los altos directivos de la cúpula de Transportes Metropolitanos de Barcelona desde el periodo de la crisis hasta la actualidad. El análisis de los datos recibidos y el cruce de dos informaciones diferentes fueron clave para poder desarrollar la noticia *a posteriori*.



## 4. Análisis de datos

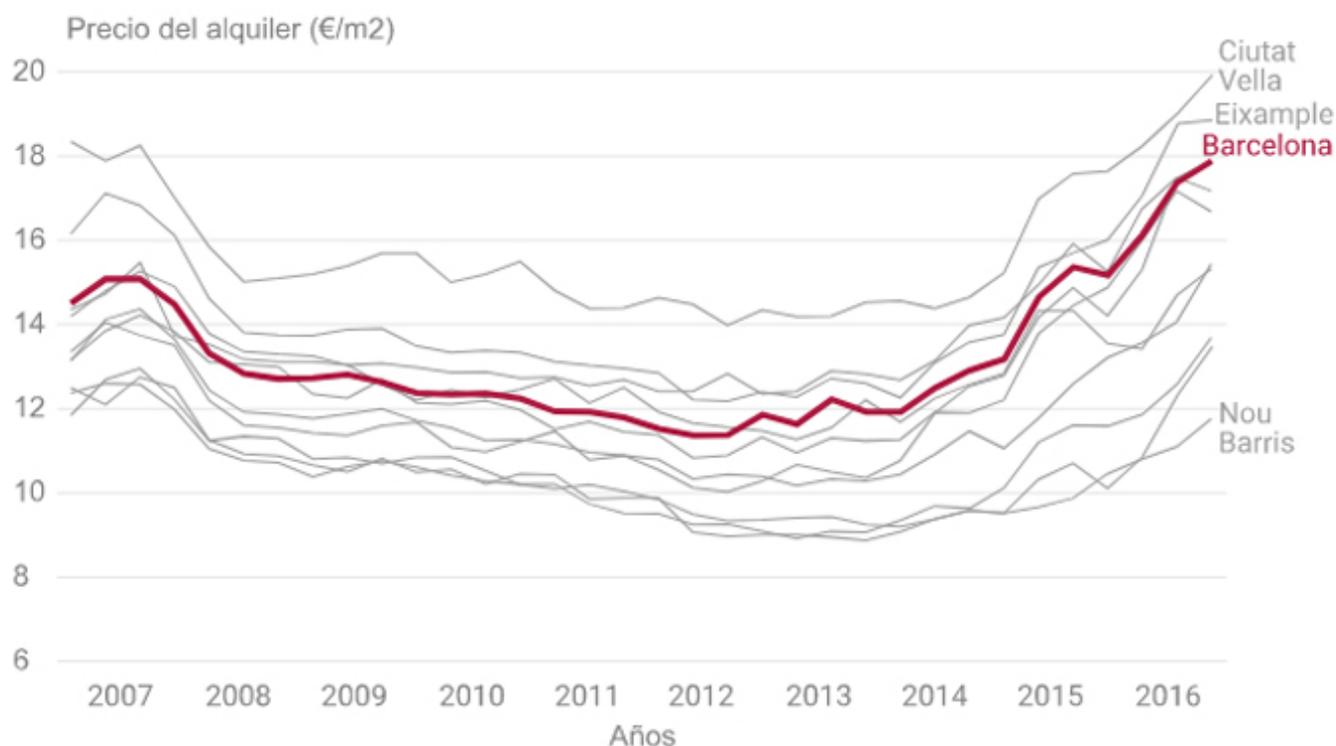
### 4.5. Cómo encontrar noticias a partir del análisis

#### 4.5.3. Los datos como apoyo de la noticia

##### 1) *El Confidencial* | «Mil euros por 28 m<sup>2</sup> en Vallecas: el alquiler enloquece y supera las cotas precrisis»

En este caso, los datos sobre la evolución del precio del hogar, tanto en Barcelona como Madrid, y su análisis y visualización, dan un valor añadido a una noticia. La dotan de contexto y de maneras gráficas para entender el momento actual de una situación determinada, en este caso, el aumento del precio del hogar.

### El alquiler en Barcelona se ha disparado en un 55% desde 2012



Fuente: Idealista y elaboración propia.

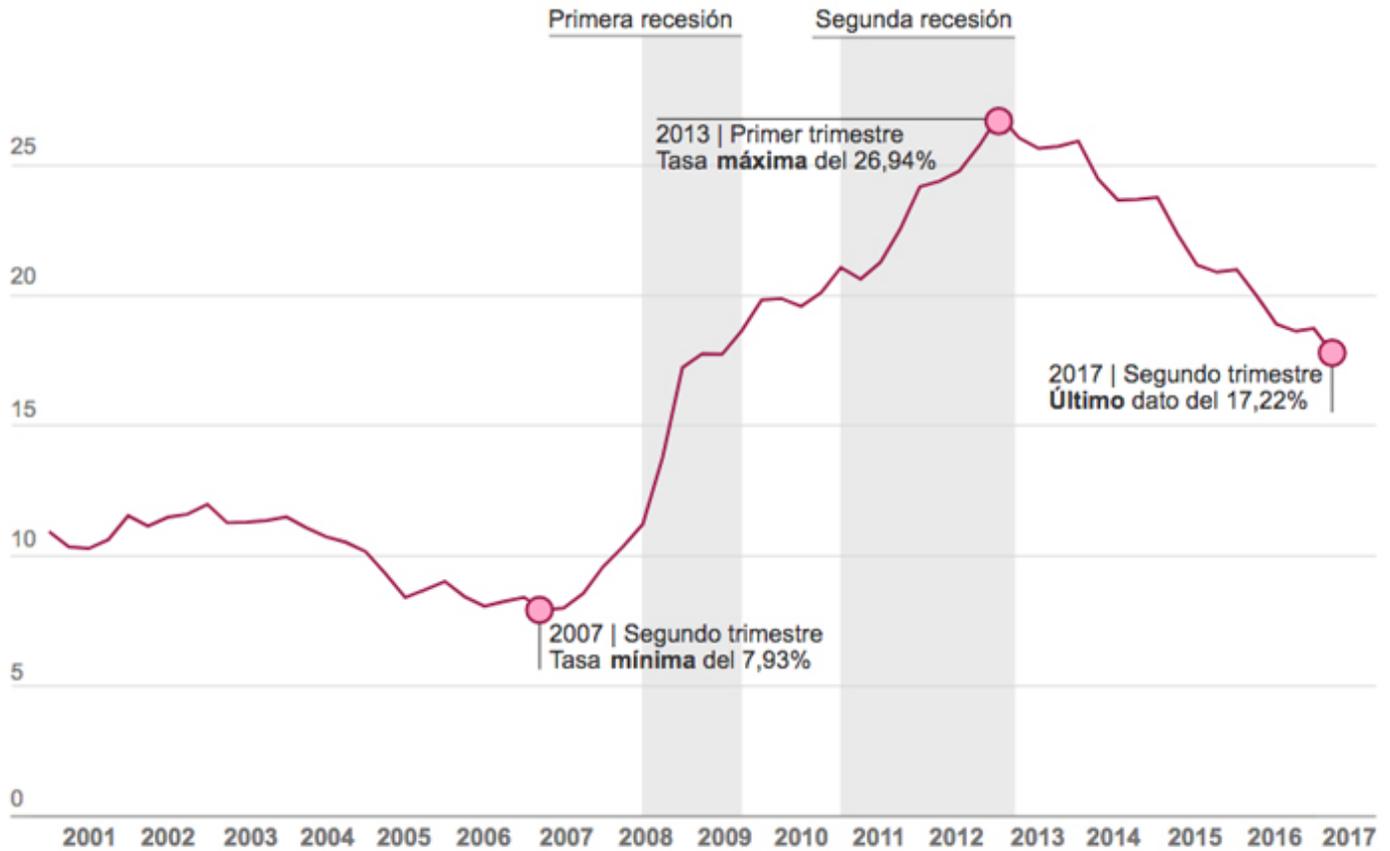
El Confidencial.LAB

##### 2) *La Vanguardia* | «España asienta una salida de la crisis que no logra frenar los desequilibrios»

Este es otro ejemplo de cómo los datos y su visualización pueden ayudar a dar apoyo y hacer más comprensible una noticia, y dotarla de un contenido más riguroso.

# Evolución de la tasa de paro en España

Desde el primer trimestre de 2001 hasta el segundo de 2017



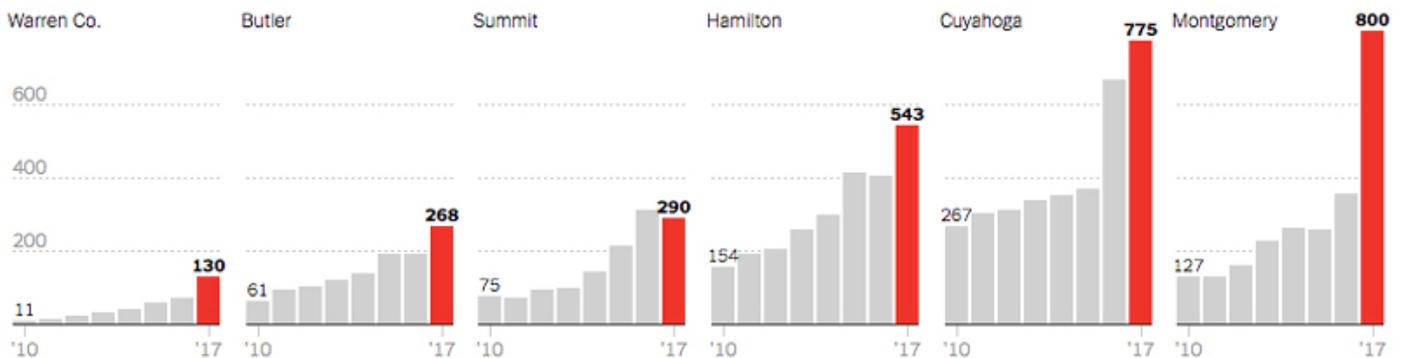
Fuente: Encuesta de Población Activa. INE

LA VANGUARDIA

### 3) The New York Times | «Las muertes por el consumo de drogas en Estados Unidos crecen más que nunca»

Esta noticia ejemplifica, con gráficos y visualizaciones, cómo el consumo de drogas vuelve a causar estragos en Estados Unidos. El tratamiento que se ha hecho de los datos, y la manera de mostrarlos, permite al lector entender todavía mejor el fenómeno que se quiere explicar.

#### Drug overdose deaths in six Ohio counties, 2010 to 2017



Totals for 2017 assume that overdose deaths continue at the same rate through the remainder of the year.

Source: Butler County Coroner's Office; Cuyahoga County Medical Examiner's Office; Hamilton County Coroner; Montgomery County Alcohol, Drug Addiction & Mental Health Service; Montgomery County Sheriff's Office; Summit County Department of the Medical Examiner

## 4. Análisis de datos

### 4.5. Cómo encontrar noticias a partir del análisis

#### 4.5.4. Los datos como punto de partida de una investigación

##### 1) ProPublica | «Killing the Colorado»

Se trata de una investigación periodística llevada a cabo por la organización norteamericana ProPublica, con el objetivo de denunciar el estado deplorable del río Colorado por culpa de la intervención humana sin medida. ProPublica es una organización sin ánimo de lucro, que cuenta con una unidad muy potente de periodistas de datos que trabajan para sacar adelante investigaciones periodísticas.



##### 2) El Mundo | «Football leaks»

Se trata de una investigación hecha por EIC, la red europea de periodistas de investigación, que en el Estado español llevaba a cabo el diario *El Mundo*. Esta investigación destapó la cara oculta del fútbol: el entramado financiero creado para que los futbolistas millonarios eludan impuestos ante Hacienda.



##### 3) El Confidencial y La Sexta | «Papeles de Panamá»

La mayor filtración de la historia no se habría podido gestionar si no fuera gracias a la utilización de herramientas de tratamiento de bases de datos. El Consortio Internacional de Periodistas de Investigación (\*), tuvo acceso a miles de archivos procedentes de

un bufete de abogados panameño, el despacho de Mossack Fonseca, que se dedicaba a gestionar sociedades *offshore* de ciudadanos procedentes de países de todo el mundo.

## 4. Análisis de datos

### 4.5. Cómo encontrar noticias a partir del análisis

#### 4.5.5. La visualización como centro de la noticia

##### 1) *El Confidencial* | «Radiografía sobre los feminicidios»

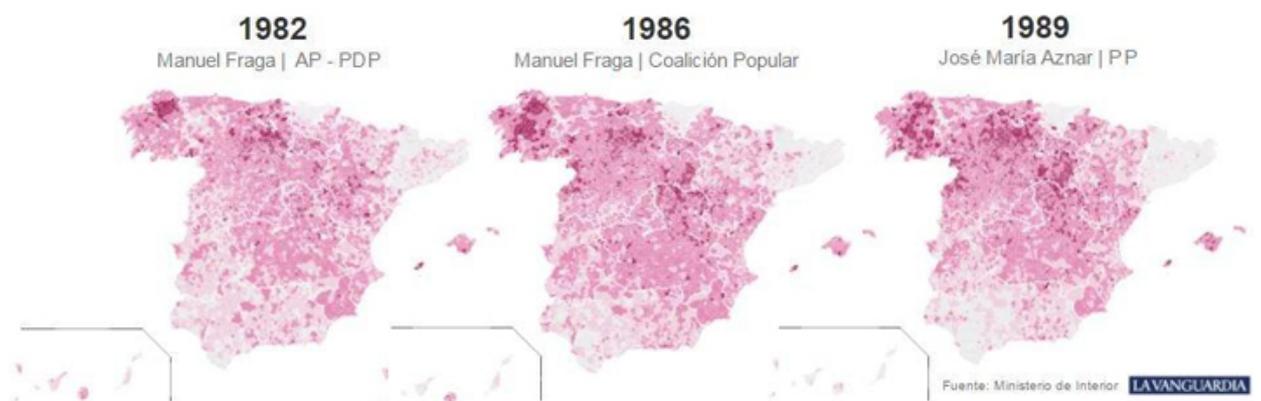
En este monográfico especial, el equipo de datos del diario *El Confidencial* recopiló datos sobre todos los feminicidios que se habían producido en España desde la implantación de la Ley de la violencia de género. En este caso, la visualización interactiva tiene un papel central e indiscutible en la presentación del tema y del tratamiento.



##### 2) *La Vanguardia* | «Los “segundones” de la democracia»

Esta recopilación de mapas hace un repaso sobre el recorrido de la segunda fuerza, es decir, los perdedores, en cada uno de los comicios de la democracia española. En este caso, de nuevo, las visualizaciones toman un papel central en el elemento informativo de la noticia.

### Los años del auge felipista



##### 3) *Eldiario.es* | «Más de 14.000 personas han muerto en el Mediterráneo desde el 2014»

Este reportaje utiliza los datos sobre el recuento de muertos en el mar Mediterráneo para hacer una aproximación al desastre humanitario de las costas mediterráneas. Los datos, y sobre todo las visualizaciones, son un elemento clave, puesto que se sitúan en el centro del elemento informativo y aportan una manera diferente de entender los números pero, sobre todo, también el contexto.



Más de **14.000** personas  
han muerto en el Mediterráneo  
intentando llegar a Europa desde 2014

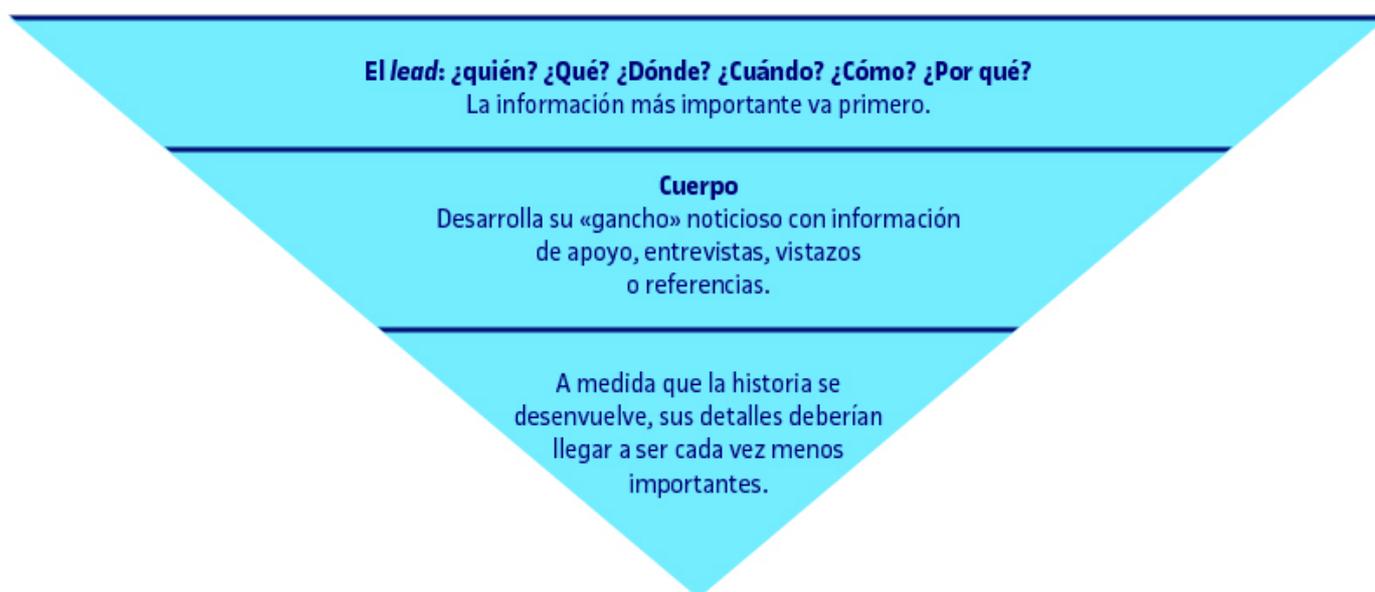
## 5. Redacción y visualización

### 5.1. La construcción de la noticia: la teoría de las 6 W, la pirámide invertida y las entradillas

La teoría de las 6 W se basa en las seis preguntas que toda información, particularmente aquella que tenga carácter noticioso, tiene que poder responder con el fin de que el lector sepa de manera clara y completa qué se está explicando. Estas preguntas son, en su expresión inglesa: *who?* (‘¿quién o a quién?’), *what?* (‘¿qué?’), *when?* (‘¿cuándo?’), *where?* (‘¿dónde?’), *why?* (‘¿quién?’) y *how?* (‘¿cómo?’).

A la hora de redactar una noticia, los medios escritos han optado por responder estas preguntas siguiendo el esquema de la pirámide invertida; o lo que es lo mismo, el primer párrafo de la información (a partir de ahora, entradilla o *lead*) debe resolver las seis preguntas o, como mínimo, aquella que resulte más importante para focalizar al lector y, a medida que avanza el cuerpo de la pieza, se tiene que contestar de una manera más o menos explícita el resto. Esta forma de organizar la noticia ha sido la más extendida, y aún hoy es la que usan prioritariamente muchos medios escritos, porque se trata de una manera sencilla y eficaz de jerarquizar la información, y garantiza una comprensión inmediata por parte del lector, aunque no llegue a leer el texto entero.

#### Pirámide invertida, nivel básico de utilización



El *lead* tiene una relevancia capital. Va seguido de una ampliación de las respuestas a las 6 W y de la presentación de la información de contexto en forma decreciente (situando siempre lo más importante en los primeros párrafos, y lo menos importante, en los últimos). En un momento de explosión de los soportes periodísticos y de *boom* informativo, de este primer párrafo depende en buena medida que el lector se aferre al texto o que lo descarte. Por eso, David Randall proponía una serie de claves para redactar una buena entradilla:

1. Directa, concisa e inquietante: suprimir todo aquello que la recargue, así como cargos y nombres que se podrían dejar para el segundo párrafo, en caso de que no sean necesarios.
2. La entradilla tiene que funcionar de manera independiente.
3. No empezar la información con una cláusula subordinada, del tipo: «A pesar de que el número de homicidios aumenta...».
4. No empezar una historia con números en dígitos, escribirlos en letras o buscar otra manera de empezar si la cifra es muy grande.
5. No empezar una información con nombres de organismos oficiales, quizá mejor empezar con otra fórmula, como por ejemplo: «expertos alertan que...».
6. Evitar, en la mayoría de los casos, empezar con citas (esto puede confundir al lector si no identificamos inmediatamente quién habla).

La mayoría de las personas que visitan una noticia en la web centran su máxima atención en la parte superior de la pieza informativa, y la escanean en forma de F, un esquema compatible con el de la pirámide invertida. Así lo demuestran los estudios

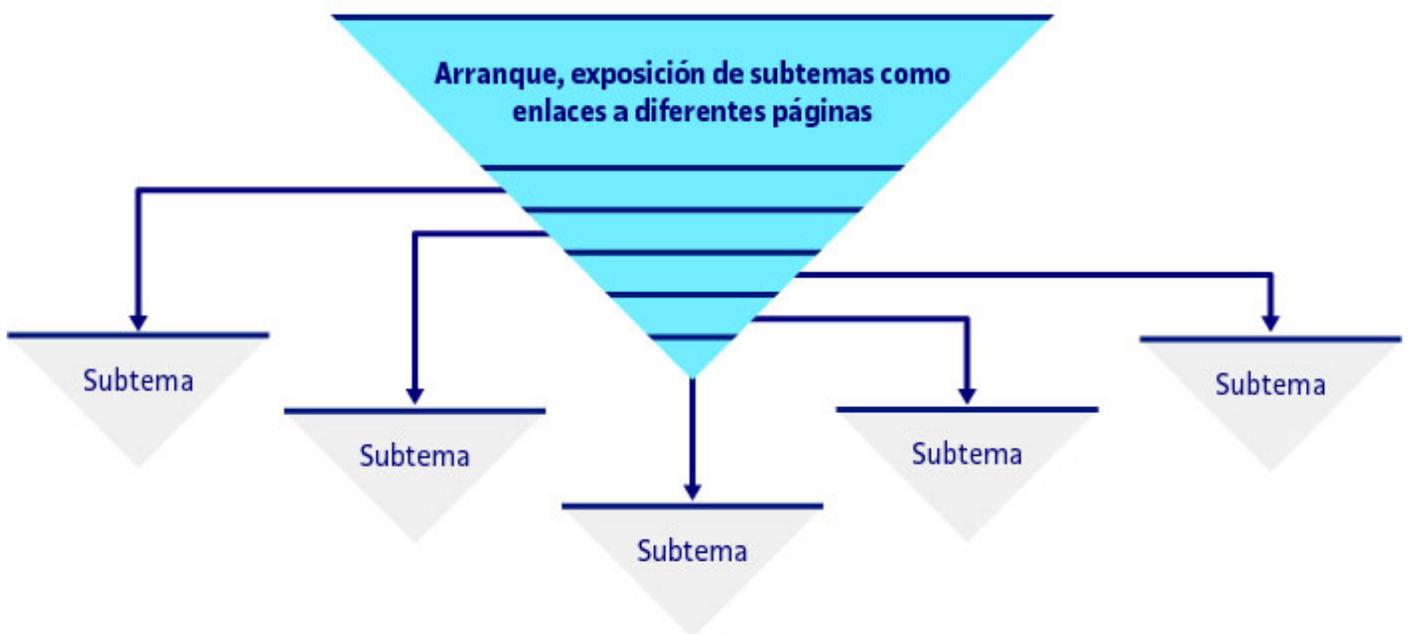
## La transformación de la pirámide invertida en la era digital

Con la crisis del periodismo impreso y la explosión de internet, la mayoría de los lectores usan ahora el ordenador, los dispositivos móviles o las tabletas para leer las noticias. Por esta razón, la pirámide invertida se ha tenido que recontextualizar para adaptarse a la hipertextualidad, multimedialidad e interactividad propias de internet. Según el periodista [Guillermo Franco](#) (\*), en la escritura por internet es importante captar la atención de los usuarios brindándoles información al instante, en lo que ha denominado «una pirámide invertida horizontal». Este concepto sugiere poner a la izquierda de la oración las palabras más importantes de la noticia. Hay tres niveles de uso de la pirámide invertida cuando se escribe una noticia en un medio digital.

1. El lineal o básico: empezar el texto con lo más relevante –las 6 W– e ir presentando el resto de los párrafos en función de su importancia para entender el tema. Esto tiene que aparecer en una misma página web.
2. La estructura: en este segundo nivel, el periodista debe jerarquizar la información, creando un primer párrafo en el que presente los datos y, después, mostrar un listado con los intertítulos en los cuales se divide la información.
3. Los subtemas: el último nivel es la creación de un contenido dividido en subtemas que ya no estarían en la misma página web, sino en diferentes páginas vinculadas al sitio web inicial. En la página inicial, se tiene que presentar el *lead* y después se exponen los subtítulos que enlazan a la ampliación del contenido. Cuando se hace clic sobre cada uno de estos subtítulos, se abre una página nueva y el lector puede ver un contenido autónomo y completo que también puede tener otros enlaces para acceder a otra parte del contenido. A estas estructuras, Guillermo Franco las denomina «pirámides flotantes».

### Pirámide invertida, tercer nivel de utilización

Pirámides flotantes ubicadas en diferentes páginas web.  
El usuario elige su propia ruta de navegación y construye su propia pirámide invertida a partir de la presentación y exposición del tema.



La adaptación de la redacción y la transformación de la pirámide invertida con el *boom* de internet y el surgimiento de nuevos medios digitales hacen que sea necesario introducir el concepto del hipertexto. El hipertexto configura un nuevo lenguaje para comunicarnos, construido a partir de frases cortas, con enlaces para profundizar y con elementos multimedia para enriquecer la información. Al mismo tiempo, esto representa:

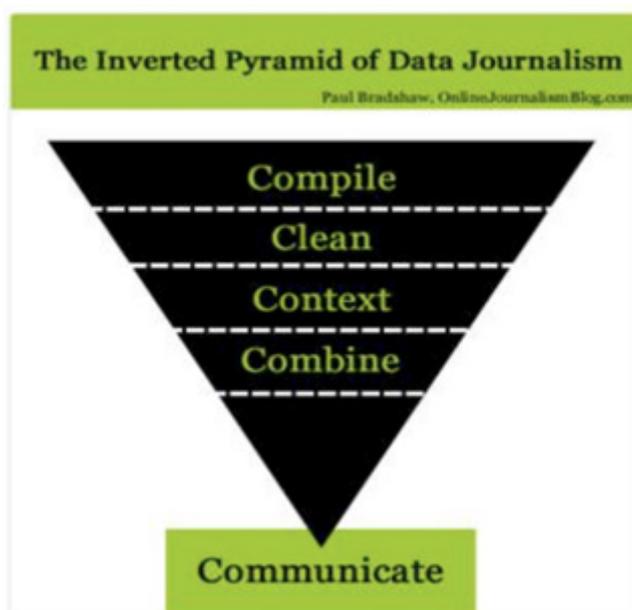
- Una nueva manera de leer la información: la lectura ha dejado de ser pasiva para convertirse en activa, la manera en que el usuario lee es totalmente diferente. Antes recibía indicaciones de los creadores de contenidos sobre lo que tenía que leer y cómo hacerlo. Ahora, el lector es un individuo libre y sin vínculos, que puede navegar como quiera.
- Una nueva manera de organizar y presentar la información: no hay límites de espacio para presentar la información. La organización de los datos en una red de nodos interconectados mediante enlaces o hipervínculos permite profundizar de manera casi infinita.

- Enlaces: los hipervínculos son una parte esencial del hipertexto, y mediante clics consiguen llevar al usuario a contextualizar la información, no solo con publicaciones relacionadas, sino con cualquier dato posible que pueda aportar algo más a la lectura.
- Las respuestas a las preguntas de las 6 W exigían la contextualización del hecho narrado, lo que implicaba la necesidad de un espacio físico importante en el que se acumulaba la información. Si al lector solo le interesaban algunos de estos datos de contexto, podía verse saturado por la multitud de información y acabar desconectando del texto. Con el hipertexto, por el contrario, la contextualización se resuelve de una manera mucho más eficaz, porque los hipervínculos generan la posibilidad de ir saltando por los datos relacionados con la información.

A pesar de que el formato web ofrece herramientas que posibilitan una mejor organización de la información, el periodismo no tiene que perder de vista que no todos los textos necesitan una estructura hipertextual, y su abuso también es un error. Lo primordial es que el periodista sea consciente de que tiene que escribir siempre pensando en el lector, porque es él quien finalmente valida su existencia cuando decide leer o navegar por el contenido.

También hay otro tipo de pirámide vinculada al periodismo de datos, creada por Paul Bradshaw (\*) (2017), que describe los pasos para convertir la información bruta a la que accede el periodista hasta un producto que pueda ser publicable en un medio digital. La lógica en este caso es diferente, y tiene más que ver con la relación entre el redactor y la información que recoge para elaborar un reportaje usando datos.

1. Compilar la información.
2. Limpiarla hasta dejar aquello más relevante.
3. Contextualizar los datos.
4. Combinarlos para, finalmente, publicar la información.



Fuente: Paul Bradshaw (2017).

## 5. Redacción y visualización

### 5.2. Tipología de piezas periodísticas: los géneros periodísticos

#### 5.2.1. Introducción

En términos generales, existen tres géneros periodísticos que se dividen en tres grandes bloques, en función de la implicación del autor con el texto: el género informativo, el género interpretativo y la opinión. Estos géneros también implican un menor o mayor grado de creatividad y de subjetividad. A continuación, hacemos una recopilación de los principales rasgos característicos de los tres géneros y subgéneros que identifica Sonia F. Parrat (\*) (2008).

#### CLASIFICACIÓN DE LOS GÉNEROS PERIODÍSTICOS

		-.....IMPLICACIÓN DEL AUTOR.....+									
Macro géneros	INFORMACIÓN			REPORTAJE		CRÓNICA		GÉNEROS DE OPINIÓN			
Subgéneros	Breve Información común	Información múltiple	Información reportajeada	Reportaje Objetivo	Entrevista literaria	Crónica local	Crónica temática	Editorial	Crítica	Artículo firmado	Columna
<i>Informar</i>			<i>Interpretar (interpretación explícita)</i>					<i>Opinar (opinión explícita)</i>			
<i>Informar</i>											
<b>FUNCIÓN</b>											

Fuente: Sonia F. Parrat (2008).

## 5. Redacción y visualización

### 5.2. Tipología de piezas periodísticas: los géneros periodísticos

#### 5.2.2. Género informativo

##### 5.2.2.1. Introducción

Idealmente, el papel del redactor en los géneros informativos o noticiosos tiene que ser el de un mero observador de la realidad. Tiene una implicación mínima en el hecho noticioso y su tarea consiste en describir los hechos de la manera más objetiva posible (asumiendo que la objetividad no existe). Se le exige veracidad, contrastar la información y aportar pluralidad. Esta última exigencia es básica: nunca se tiene que dar por válida ni recoger en exclusiva ni la versión oficial de los hechos ni una sola voz; para redactar una buena información, el redactor debe encargarse que la pieza recoja fielmente las distintas «caras de la verdad».

A grandes rasgos, todas las informaciones contienen los siguientes elementos:

- Están encabezadas por un titular (que a veces va acompañado de un antetítulo, para situar geográficamente o temáticamente la información; y/o un subtítulo, que suele complementar el titular). El título de la información ha de ser breve –el número de palabras depende del medio en el que se trabaje, pero nunca conviene superar el par de líneas–, claro y atractivo. Su función es la de aportar un resumen de la información, que encontraremos en el cuerpo de la información.
- El *lead* o entradilla.
- El cuerpo, que desarrolla y completa la información siguiendo el esquema de la pirámide invertida, es decir, explicando los hechos en orden de importancia decreciente, dejando los datos más secundarios para el final. La información se divide en párrafos que tienen que estar interrelacionados temáticamente entre sí, pero que deben ser independientes en un ámbito gramatical. En el cuerpo del texto, se amplía la información del *lead* y se aporta otra nueva, se añaden declaraciones de las personas implicadas con la noticia y se aporta contexto.
- La noticia es la pieza que caracteriza este género, a pesar de que su extensión, contenido o enfoque pueden variar, creando diferentes tipologías de piezas informativas, como veremos a continuación.

##### 5.2.2.2. La noticia

«La función de un diario consiste en descubrir información nueva sobre asuntos de interés general y transmitirla a los lectores con la mayor rapidez y precisión posibles, de manera veraz y ecuaníme.»

David Randall

«La función primordial de un diario es recoger noticias e, incluso a riesgo de su propia existencia, garantizar que se transmitan sin desvirtuarlas. La verdad no se puede deformar con lo que se dice o no se dice, ni con la manera de presentarlo.»

Esto lo escribía, por su parte, el periodista C. P. Scott, director del *Manchester Guardian* en 1921. A esta cita, añadía: «cada cual es libre de comentar lo que quiera, pero los hechos son sagrados».

Las noticias son el elemento del que se nutren los medios, y tienen las siguientes características:

- Son algo inusual. «Que un perro muerda a un hombre no es noticia; lo es que un hombre muerda a un perro». Este es el clásico ejemplo que citan los profesores de primer curso en las facultades de periodismo, pero las noticias son más que esto.
- Tienen valor informativo porque aportan novedad, cosas que la gente no ha escuchado antes.
- Es fundamental que recojan temas de interés público: no solo temas que tengan un impacto en la vida pública, sino también temas que sean de interés para el público.

En cuanto a las tipologías, se distinguen cuatro variedades relacionadas con el grado de profundidad de la noticia o de la pieza informativa:

**1) El breve:** una información corta de un asunto sin gran importancia que se inserta en un rincón de la página del diario o de la web, con un titular poco elaborado y sin fotografía. Su función es la de informar sobre un hecho de actualidad, y su redactor se distancia al máximo de lo que está explicando. A menudo se firma como «redacción», y muchas veces la noticia proviene de una agencia de información, como la ACN o la Agencia Efe, por ejemplo.

**2) Pieza informativa básica:** trata de un asunto concreto de la actualidad. No cuenta con ningún añadido ni ningún texto complementario pero lleva, a menudo, una fotografía o una infografía. Su función también es puramente informativa, y la implicación del autor es mínima.

**3) Información múltiple:** varias informaciones conectadas entre sí que se publican en una misma página en un diario impreso, o bajo un elemento gráfico que indique que están interrelacionadas, en caso de que se publiquen en un medio digital. Cumple una función informativa, y el grado de implicación del autor es bajo, pero mayor que en las anteriores piezas.

**4) Información reportajeada:** se trata de un texto informativo redactado con un enfoque más próximo al reportaje, que ha tomado presencia en la prensa para intentar captar la atención del lector. No se tienen que confundir estas piezas con los reportajes, porque las informaciones reportajeadas comunican hechos de estricta actualidad y la implicación del autor es menor que en el caso del reportaje, a pesar de que a veces contiene ciertas dosis de interpretación. Una pieza de este tipo acostumbra a ir acompañada de noticias complementarias, que aportan nuevos datos, o de un texto interpretativo, una pequeña entrevista o incluso un texto de opinión. También la acompañan fotografías o infografías.

## 5. Redacción y visualización

### 5.2. Tipología de piezas periodísticas: los géneros periodísticos

#### 5.2.3. Género interpretativo

##### 5.2.3.1. El reportaje

“ «El reportaje es el género periodístico más completo: además de incluir otros géneros – crónica, entrevista, noticia–, puede tener como antecedente una noticia y lo elabora un periodista cada vez más especializado. Pero tiene una ventaja respecto a la noticia, y es que está desatado de la estricta actualidad diaria.»

Randall (2008).

Como señalaba David Randall, el reportaje puede surgir a raíz de un hecho noticioso que se amplía o se profundiza, pero también a partir de cuestiones de actualidad o de interés general, pero no necesariamente noticiosas. No sigue estrictamente las mismas pautas de la noticia, sino que se acerca más al estilo literario y contiene detalles descriptivos, frases de testigos y vivencias propias.

Los reportajes ocupan más espacio que una noticia. Su contenido no deja de ser información (esto es, probablemente, lo que más lo distingue de un texto literario), pero así como en las noticias hay un molde definido, en los reportajes no existen fórmulas determinadas. El reportero elige la forma y el tono que mejor se acomodan al asunto del que habla, y disfruta de una mayor libertad estructural y expresiva. Suele ir acompañado de fotografías o infografías.

El primer párrafo no tiene interés informativo pero debe ser atractivo, porque de la manera en que esté redactado depende que el lector lea el resto del texto. Las características del *lead* también existen, pero se colocan donde más le conviene al redactor. Para elegir la temática del reportaje, en cierto modo tenemos que dejarnos influenciar por el clima de opinión. Antes de redactarlo, es conveniente elaborar un esquema simple que nos facilitará el proceso de elaboración del texto.

**1) Hipótesis de partida:** creemos que existe un conflicto, porque los medios / plataformas ciudadanas / grupos de presión / administraciones hablan del mismo. Por ejemplo: en Barcelona se ha disparado el precio de la vivienda. Se está forjando una segunda burbuja inmobiliaria que supone un proceso de gentrificación o sustitución social, que obliga a muchos de sus habitantes a abandonar sus barrios.

**2) Elegir las fuentes:** para poder recoger la información, nos tenemos que preguntar: ¿quién lo denuncia? (fuente 1) –plataformas y movimiento vecinal–; ¿a quién le beneficia este conflicto? (fuente 2) –sobre todo a los grandes tenedores, como las entidades financieras, las inmobiliarias o los fondos buitre–; ¿quién puede poner contexto a la historia? (fuente 3) –sociólogo, experto en urbanismo–; ¿en qué contexto legal se enmarca la historia? (fuente 4) –un jurista especializado en la LAU (la Ley de arrendamientos urbanos) puede ser un buen contacto–; y por último, ¿a quién le perjudica este conflicto? (fuente 5) –un testigo: aportar la visión humana es siempre interesante para que el lector pueda empatizar con la situación. Podríamos ampliar el número de fuentes, pero tenemos que saber cortar a tiempo para no crear monstruos que después no lea nadie.

El reportaje se caracteriza por los siguientes rasgos:

- Su función es básicamente la de un segundo nivel de información, es decir, la interpretación. No contiene opinión, sino que muestra los hechos y los explica. Expone datos seleccionados por el autor y explicados con un cierto grado de intencionalidad, por eso no se puede considerar información en sentido estricto.
- Hay cuatro pasos básicos en la elaboración de un reportaje: elección del tema, acumulación de toda la información posible recurriendo a todo tipo de fuentes, decisión del enfoque que le daremos y las técnicas que se aplicarán en el proceso de redacción.
- El autor tiene una gran libertad que le permite elegir el enfoque, el estilo y la estructura que considere más convenientes para su texto. Especialmente en el género interpretativo, es deseable que tenga un principio cautivador, un desarrollo interesante y un final concreto. Se redacta combinando la narración con las descripciones.
- Sus objetivos son captar la atención del lector, mantener el interés y estimular el deseo de seguir leyendo.

Tipologías de reportajes:

**1) Reportaje objetivo** (o estándar): aunque su función básica es interpretativa y, aunque el grado de implicación del autor es mayor que en la información, acostumbra a estar escrito con un estilo poco creativo, no muy extenso, profundiza poco y casi no analiza los hechos. Suele denominarse *perfil* al reportaje que se centra en una persona concreta, y necrológica, al que se publica después de la muerte de un personaje conocido y hace un repaso de los hechos más significativos de su vida.

**2) Reportaje interpretativo** (o reportaje en profundidad): aporta antecedentes, contextualiza, analiza los hechos hasta llegar al fondo y prevé posibles consecuencias. Su función es interpretativa, pero está exenta de las valoraciones que contienen la crónica o el artículo de opinión. El reportero analiza y el lector saca sus propias conclusiones.

### 5.2.3.2. El análisis

La base de un análisis es la selección de un hecho de actualidad que ha sido noticia y que requiere una explicación posterior más seria, detallada y especializada, de forma que su componente interpretativo es alto. Suele encargarse a personas que no son necesariamente periodistas, pero sí expertas en la materia, y su mérito radica más en la capacidad de profundizar y analizar cuestiones complejas que en su calidad literaria.

### 5.2.3.3. La entrevista

Antes que nada, la entrevista no se refiere al proceso de obtención de información por parte del periodista, el cual se pone en contacto con la persona que le tiene que facilitar esta información, sino al texto final, resultado de poner por escrito lo que el entrevistado nos ha explicado. Sin embargo, una entrevista es algo mucho más complejo que formular un par de preguntas al final de una rueda de prensa. Para que el contenido sea de calidad (y en especial, si queremos obtener una información personal del entrevistado que vaya más allá de su opinión sobre un tema concreto), tenemos que crear un clima de confianza y de confort. Tiene que existir un acuerdo previo entre entrevistador y entrevistado para pactar este encuentro.

Una entrevista es un relato de la conversación entre dos o más personas, una de las cuales, asumiendo el rol de entrevistador, es el autor. Estas son las principales características de las entrevistas:

- Contienen un cierto grado de interpretación e implicación del periodista, que es mayor que cuando redacta una información, porque el redactor no parte de unos hechos ajenos, sino que decide a quién entrevistar y las preguntas que le formulará. Algunos piensan que la elección de estas preguntas puede condicionar las respuestas y el tono general de la entrevista. Cuando transcribimos las respuestas tenemos que preservar su contenido, no en el sentido de que no podemos editarlas, sino que debemos procurar que nuestra intervención no altere su sentido, ni la intencionalidad, ni el ambiente dialéctico en el que se produjeron.
- El autor de la entrevista tiene que saber qué desea obtener antes de hacerla, hecho que exige una recopilación previa de documentación sobre el entrevistado y la preparación de las preguntas en función de la información que se quiere recoger.
- En el transcurso de la entrevista, las preguntas han de ser lo más concretas y directas posibles.
- El entrevistador tiene que ser capaz de crear un buen ambiente, pero no hablar demasiado, para obtener la mayor información posible del protagonista (aprender a jugar con los silencios).
- También ha de tener la capacidad de ver, observar, escuchar, improvisar y percibir todo aquello que uno expresa sin palabras, con gestos, movimientos, tono, etc. Es lo que se acostumbra a denominar información subliminal.
- El objetivo último que persigue el autor es hacer que el lector se sienta casi testigo de la conversación.

La doctora en Periodismo Montserrat Quesada propone una clasificación de las entrevistas en dos grandes tipologías:

**1) Entrevista informativa** (u objetiva): incluye las declaraciones que personas implicadas hacen sobre un hecho de actualidad, o bien personajes conocidos sobre cuya actividad profesional, social, política o artística queremos obtener ideas, no sobre aspectos de su vida íntima. Esto pasa cuando el personaje nos interesa por su especialidad, no como personaje en sí. El periodista, en esta clase de entrevistas, se limita a transcribir la conversación mediante el sistema de pregunta-respuesta, sin comentarios ni interpretaciones.

Este tipo de entrevista tiene una subtipología, la **conversación objetiva**: es un subtipo de la modalidad anterior, que consiste en transcribir una conversación previamente grabada que se da entre dos personajes sobre una misma cuestión, a los cuales se les deja hablar libremente sin que el reportero intervenga.

**2) Entrevista literaria / entrevista perfil**: a caballo entre la información y la interpretación, en esta pieza trasladamos las ideas de un personaje, pese a que pasarán por el filtro del propio periodista. El objetivo de estas piezas es dar a conocer la personalidad del entrevistado por medio de un lenguaje más literario. No recurrimos a la pregunta-respuesta, sino que reproducimos las

declaraciones del entrevistado entre comillas y alternadas con descripciones y explicaciones sobre el personaje, su vida y su actitud durante la entrevista. Es una pieza más atemporal.

### 5.2.3.4. La crónica

La crónica es uno de los subgéneros más difíciles de definir, porque comparte rasgos con otros subgéneros. En un intento de afinar al máximo una descripción de la crónica, el periodista Manuel Graña, ya a mediados del siglo pasado, señaló algunas de las principales características de la crónica.

- Excluye todo lo que sea simple información, es decir, un relato impersonal. Lo que distingue a la crónica de la mera información es precisamente el elemento personal, porque el autor se permite la licencia de hacer comentarios, ampliar la información y ordenar los hechos a su manera. Aunque la crónica sea informativa, hay margen para que el reportero le ponga su estilo propio.
- En comparación con otros textos más especializados, la crónica periodística es ligera, decorada con recursos literarios pero sin muchos tecnicismos, breve y ordenada.
- Condensa el elemento informativo y el elemento literario o interpretativo.
- Contiene las respuestas a las 6 W, pero se les otorga un carácter más retórico. La información está explicada con más libertad estilística y un vocabulario más rico que la información.
- Admite una gran variedad de fondo y forma, de modo que se podrían enumerar muchos subtipos de crónicas, como la artística, la biográfica, la personal, la descriptiva, etc.

Años más tarde, el catedrático y periodista Gonzalo Martín Vivaldi propuso una nueva definición de la crónica, según la cual se trata «de una información interpretativa y valorativa de hechos noticiosos, actuales o actualizados, donde se narra algo al mismo tiempo que se juzga lo que se está narrando». Le atribuía las siguientes características:

- Estilo personal, pero sin desvirtuar los hechos ni deformar la realidad.
- Libertad de recursos estilísticos: comparación, metáfora, ironía, anécdota, pero siempre con claridad comunicativa y concisión.
- Formato narrativo-informativo, sin preocuparse por la pirámide invertida.
- El sujeto-protagonista son los hechos noticiosos, así como la interpretación que hace de los mismos el cronista.
- El tema es la noticia radiografiada, o lo que es lo mismo, no solo la anatomía del hecho noticioso sino también su psicología, sus implicaciones.
- Vivaldi añadía una modalidad llamada *croniquilla*, que definía como una pieza pequeña dedicada a algo cotidiano y aparentemente intrascendente.

Hoy día, en el Estado español se entiende por crónica «la narración temporal de un acontecimiento» o, en palabras del doctor en Periodismo Manuel Bernal:

“ «Una información de hechos noticiosos, transcurridos en un periodo de tiempo, por un cronista que los ha vivido como testigo o investigador e incluso como protagonista y que, al mismo tiempo, los narra, analiza e interpreta mediante una explicación personal.»

Sin abandonar las características señaladas por los otros autores, Bernal añade nuevos rasgos distintivos a la crónica:

- Contiene ciertos elementos de información, porque siempre tiene una base informativa –unos hechos noticiosos que dan pie a escribir una crónica sobre lo que ha sucedido–, y también del reportaje interpretativo, porque profundiza y analiza, está escrita con un estilo personal y creativo pero no recargado, a pesar de que la implicación del autor es aún mayor porque participa de la acción como testigo directo y valora los hechos a partir de lo que ha visto y de los conocimientos que tiene. Sin embargo, su función principal no es opinar, sino informar sobre una cuestión, interpretándola según sus impresiones.
- Tiene una relación temporal clara, lo que significa que el hecho narrado ha de tener una cierta continuidad porque el cronista explicará todo lo que haya sucedido entre dos momentos concretos (al principio y al final de los hechos en cuestión).
- Suele tener una extensión considerable.

- Estructuralmente, a menudo se inicia con un planteamiento de lo que ha sucedido, seguido de una interpretación de los hechos (explicación, análisis y valoración) y, finalmente, una conclusión que sintetice el resultado de lo analizado.

Se pueden diferenciar dos tipos de crónicas:

**1) Crónica local:** su autor acostumbra a ser un periodista desplazado permanentemente o de manera temporal, con los corresponsales conocedores de la actualidad informativa de otros países o ciudades, o los enviados especiales que cubren determinados acontecimientos puntuales, como conflictos bélicos, procesos judiciales, debates parlamentarios, cumbres o grandes acontecimientos deportivos. Aun así, hoy muchas de las crónicas provenientes del extranjero que se publican en los diarios locales suelen ser más informativas que interpretativas, puesto que los reporteros se limitan a informar «objetivamente» de lo que pasa, sin entrar en valoraciones que requerirían un conocimiento mucho más profundo y, en consecuencia, mucho más tiempo para elaborarlas.

**2) Crónica temática:** el propio texto informa sobre un hecho al mismo tiempo que lo valora. El autor suele ser un periodista especializado en cubrir informativamente determinados hechos relacionados con el deporte, la tauromaquia, acontecimientos de la sociedad o culturales, por ejemplo. Los presencia de principio a fin, y esto le permite narrarlos y hacer llegar al lector sus impresiones de lo que ha presenciado.

## 5. Redacción y visualización

### 5.2. Tipología de piezas periodísticas: los géneros periodísticos

#### 5.2.4. Los géneros de opinión

Comentar es una actividad complementaria a la de redactar para informar sobre la actualidad. Los textos de opinión que se publican en los medios contienen puntos de vista respecto a los hechos de actualidad que hemos conocido a través del resto del diario. Este carácter de complementariedad podría hacernos pensar que el comentarista lleva a cabo un rol secundario respecto al informador, pero no es así. De hecho, según el periodista Héctor Borrat:

«Comentar un tema es conferirle un rango más elevado que el de aquellos temas que solo son narrados, y dedicarle una editorial es asignarle el más alto rango.»

Casi todos los medios cuentan con una sección dedicada a la opinión. Cuando se publica un texto de opinión en las páginas dedicadas a la información, se diferencia del resto colocándolo en un recuadro o con una tipología diferente, lo que hace saber al lector que está ante un texto con una finalidad diferente del resto. Y si en algo coinciden todos los textos de opinión, es que sus autores tienen que basar sus opiniones en datos correctos y sin manipulaciones. Para hacerlo, se requiere un profundo conocimiento de la actualidad y un buen dominio del lenguaje.

##### 5.2.4.1. La editorial

La editorial es un subgénero de la opinión, que tiene las siguientes características:

- Su objeto es, sobre todo, comentar las noticias, pero sin prescindir de los elementos informativos, que dan pie a la opinión.
- Generalmente no se firma.
- No es el autor quien opina, afirma o niega, acusa o enaltece, sino el conjunto del medio como empresa.
- Quien redacta la editorial gana en autoridad por la influencia que tiene en la sociedad la opinión del medio como institución.
- Respecto al estilo: el propósito del artículo tiene que ser evidente y bien definido.
- Al articulista se le pide narración ordenada o exposición lógica, mucha información, mucha memoria, mucha dialéctica y poca emotividad agresiva.
- La elocución tiene que ser sencilla y sin pretensiones de arte literario o purismo exagerado.
- Refleja la postura del medio respecto a un asunto.
- Suele escribirlo el director del medio.

Lo que dirige la opinión de los lectores del medio no es la editorial en sí, sino el conjunto de las noticias publicadas. Lo que haría la editorial sería dar un sentido a los hechos, que en este caso se exponen con el propósito más o menos manifiesto de llevar al lector a una convicción. Existen diferentes modalidades de editoriales:

##### 5.2.4.2. La columna

La temática que se trata en las columnas y su tono tienen que ver con el sentido y el estilo personal del columnista: a veces es un análisis profundo, otras, un material satírico, irónico o humorístico y otras, un simple comentario personal. La columna tiene una periodicidad y lugar fijos en los medios, y su autor es un colaborador fijo que opina sobre cuestiones de actualidad, dotándolas de un carácter personal. A veces, el autor es una persona conocida cuya figura, con frecuencia, interesa más que su opinión. Hay dos tipologías de columnas:

- La columna de actualidad: el columnista opina sobre cuestiones de la actualidad más inmediata.
- La columna personal: el autor usa un estilo mucho más personal para opinar sobre temas más ligeros y no necesariamente vinculados a la actualidad informativa.

##### 5.2.4.3. La crítica

Es un análisis valorativo de la producción artística (en el campo del teatro, la música, el arte, la literatura, el cine, etc.), propio de la sección cultural de un medio, y está escrito por personas con un alto grado de especialización. Sus principales características son:

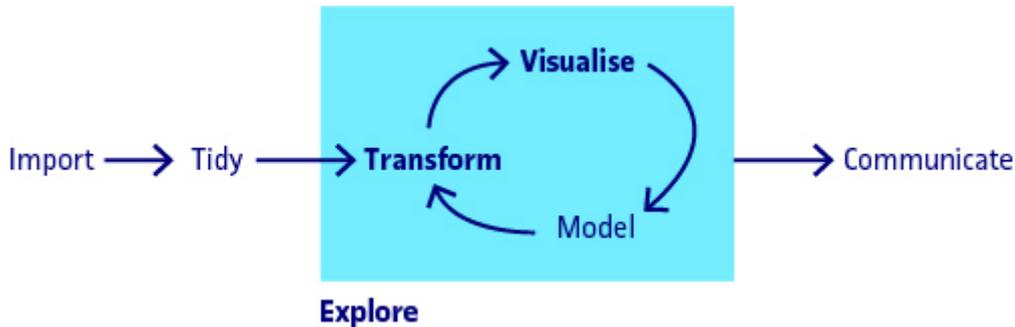
- La función del crítico es apreciar la relación entre el propósito del autor de la obra y los resultados obtenidos, con el fin de ayudar al lector a entender la obra y orientarlo.
- El autor acostumbra a ser especialista en la materia que tiene que valorar. Su comentario debe ser fruto de un análisis objetivo, detallado y fundamentado.
- El autor, que también ha sido lector o espectador de la obra, acostumbra a empezar su crítica informando al lector sobre quién es el autor, qué otras obras ha creado y lo que supone su nueva obra. A continuación, describe la obra y finalmente, a partir de sus conocimientos como experto, emite unas valoraciones razonadas.

## 6. Tres ejercicios de periodismo de datos

### Introducción

Esta serie de tres ejercicios quiere poner en práctica lo aprendido en esta asignatura y en otras del máster (obtención, análisis y visualización de datos y redacción). Encuentra los datos y, mediante su análisis y visualización, crea el hilo argumental que los explique.

El Consejo de Redacción del medio para el que trabajas ha aceptado la propuesta de reportaje/investigación que hiciste hace unas semanas. Ahora hay que probar con datos las hipótesis que planteaste y contarlos de forma atractiva e inteligible en una historia periodística.



Fuente: Garrett Golemund y Hadley Wickham.

En cada uno de los encargos deberás:

- Conseguir los datos.
- Aplicar mecanismos de selección y filtrado de los datos (limpia y ordena, genera nuevas variables).
- Analizar – visualizar (y volver a analizar, proceso reiterativo) hasta dar con las visualizaciones y argumentos que expliquen los datos.
- Encontrar el hilo argumental a partir del análisis de los datos.

Este es el esquema de los pasos que hay que dar para desarrollar cualquier proyecto de análisis de datos. Como se indica en el gráfico, el proceso no es lineal. En esta serie de ejercicios se muestra un acercamiento a cada uno de los pasos para trabajar mediante la práctica las diferentes fases de un proyecto de datos. Los casos incluyen la obtención de datos de diferentes fuentes (API, fuentes oficiales y no oficiales, PDF) y su tratamiento y visualización mediante todo tipo de programas (análisis de redes, cartografías, visualización interactivas). Se han seleccionado herramientas de software abierto o libre (Gephi, Python, D3) y que funcionan en todas las plataformas (Linux, Windows y Mac) para facilitar el acceso a estos ejercicios.

Si vas a desarrollar código para alguno de los proyectos, recomendamos que utilices un sistema de control de versiones como Git, que te permitirá revisar la evolución de tu trabajo y colaborar con otras personas más fácilmente. Documenta con rigor todas las manipulaciones que hagas a los datos para poder luego revisar tu trabajo.

#### Nota

En los siguientes ejercicios se sugieren una serie de herramientas a modo de ejemplo. Utiliza otras si te resulta más sencillo o quieres explorar otros caminos. Las preguntas que se formulan son también orientativas; puedes formular otras que te parezcan más interesantes e intentar resolverlas.

## 6. Tres ejercicios de periodismo de datos

### 6.1. Twitter: captura de datos y análisis de redes

#### 6.1.1. Introducción

Te llaman de la redacción, tienes que hacer una pieza rápida sobre una cosa de Twitter. Que «las redes están ardiendo», dicen, te suena a chiste, pero no te queda otra que ponerte a bucear en un *trending topic* y producir algo que merezca la pena ser leído: ¿quién está detrás?, ¿cómo surgió? A lo mejor incluso aprendes algo por el camino.

#### **Nota**

Este ejercicio está basado en los ejemplos, código y presentaciones de Mariluz Congosto (blog: <http://barriblog.com>, Twitter: [@congosto](https://twitter.com/congosto)).

Cosas que puedes aprender: capturar tuits y perfiles en Twitter a través de su API, analizar los datos capturados mediante grafos con el programa Gephi.

## 6. Tres ejercicios de periodismo de datos

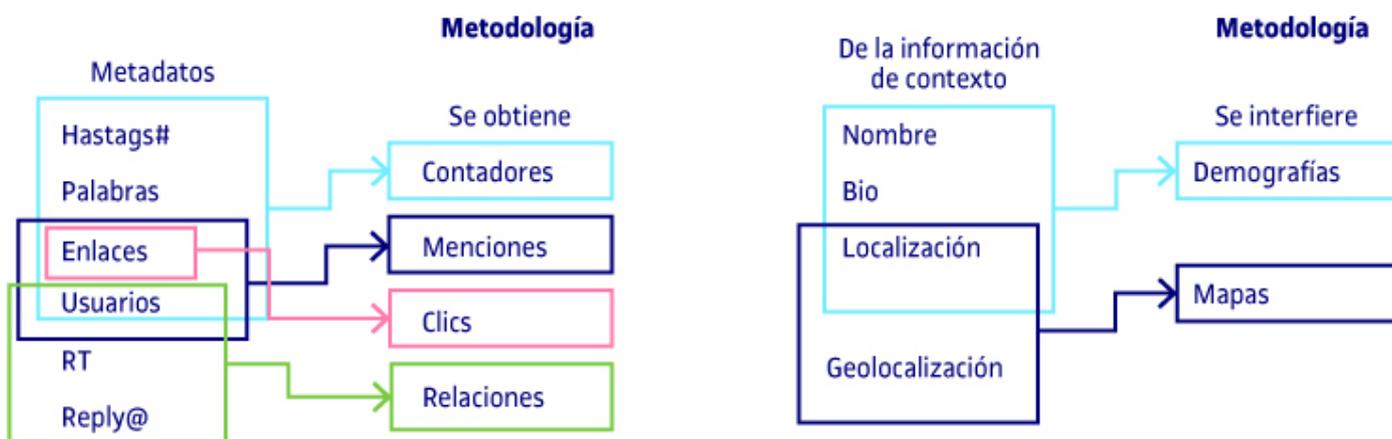
### 6.1. Twitter: captura de datos y análisis de redes

#### 6.1.2. Contexto

Las redes sociales en internet son un lugar importante dentro del ecosistema de medios, donde se distribuyen, y a veces se crean, las noticias. De entre ellas, Twitter se ha configurado como la plataforma de microblogging más popular. Se ha convertido en lugar donde informarse (y a veces desinformarse). La apertura de la [API de Twitter](#) ha permitido un acceso directo a las fuentes de información que este tipo de plataformas acumula (el otro caso es Facebook, que es mucho más cerrada a la hora de acceder a su información). Una API es una forma estructurada de acceder a la información que contiene una biblioteca. Analistas, académicos, activistas y periodistas la usan para acceder a los datos que de otra forma tendrían que comprar o «scrapear». Hay muchísima literatura al respecto y multitud de artículos académicos desde la sociología, la politología, las ciencias de la comunicación y muchas otras disciplinas que utilizan Twitter como caso de estudio por su facilidad a la hora de obtener los datos. Pero dejemos estas discusiones académicas de lado:

Tu jefe te ha encargado preparar una pequeña pieza sobre uno de los *trending topics* de Twitter del día. ¿Quién está publicando sobre el tema? ¿puedes hacer una historia con ello más allá de seleccionar los tuits más retuiteados? Multitud de noticias hablan de las injerencias en Twitter de Rusia o Venezuela. ¿Cómo hacen los análisis?; ¿cómo capturan los datos?

Vale, hemos dicho que ahora Twitter lo ha puesto fácil para descargarse sus tuits, pero ¿por dónde empezar? Para este ejercicio, hemos pensado usar T-hoarder, un software libre que funciona en Python, pensado para ayudarte a interactuar con la API de una manera sencilla, sin necesidad de conocimientos de programación. Con ella podrás obtener información de Twitter a través de su API, que en realidad son tres REST, *search* y *streaming*, y procesar la información obtenida para hacer visualizaciones. Dependiendo del tipo de datos que obtengamos, podremos generar un tipo de visualizaciones u otras. Con los usuarios, con sus *followers* y sus retuits (RT) y menciones, podemos obtener relaciones entre usuarios para hacer gráficos de redes (grafos) o, si analizamos la localización de los tuits, podemos hacer mapas.



## 6. Tres ejercicios de periodismo de datos

### 6.1. Twitter: captura de datos y análisis de redes

#### 6.1.3. Conceptos básicos o previos

**Anatomía de un tuit.** Un tuit (tweet en inglés) es un mensaje que contiene mucha información asociada (metadatos) que no vemos cuando lo vemos en la web o en una aplicación.

```
{
  "id" => 12296272736,
  "text" =>
    "An early look at Annotations:
    http://groups.google.com/group/twitter-api-announce/browse_thread/thread/fa5da2608865453",
  "created_at" => "Fri Apr 16 17:55:46 +0000 2010",
  "in_reply_to_user_id" => nil,
  "in_reply_to_screen_name" => nil,
  "in_reply_to_status_id" => nil,
  "favorited" => false,
  "truncated" => false,
  "user" =>
    {
      "id" => 6253282,
      "screen_name" => "twitterapi",
      "name" => "Twitter API",
      "description" =>
        "The Real Twitter API. I tweet about API changes, service issues and
        happily answer questions about Twitter and our API. Don't get an answer? It's on my website.",
      "url" => "http://apiwiki.twitter.com",
      "location" => "San Francisco, CA",
      "profile_background_color" => "c1dfee",
      "profile_background_image_url" =>
        "http://a3.twimg.com/profile_background_images/59931895/twitterapi-background-new.png",
      "profile_background_tile" => false,
      "profile_image_url" => "http://a3.twimg.com/profile_images/689684365/api_normal.png",
      "profile_link_color" => "0000ff",
      "profile_sidebar_border_color" => "87bc44",
      "profile_sidebar_fill_color" => "e0ff92",
      "profile_text_color" => "000000",
      "created_at" => "Wed May 23 06:01:13 +0000 2007",
      "contributors_enabled" => true,
      "favourites_count" => 1,
      "statuses_count" => 1628,
      "friends_count" => 13,
      "time_zone" => "Pacific Time (US & Canada)",
      "utc_offset" => -28800,
      "lang" => "en",
      "protected" => false,
      "followers_count" => 100581
    }
}
```

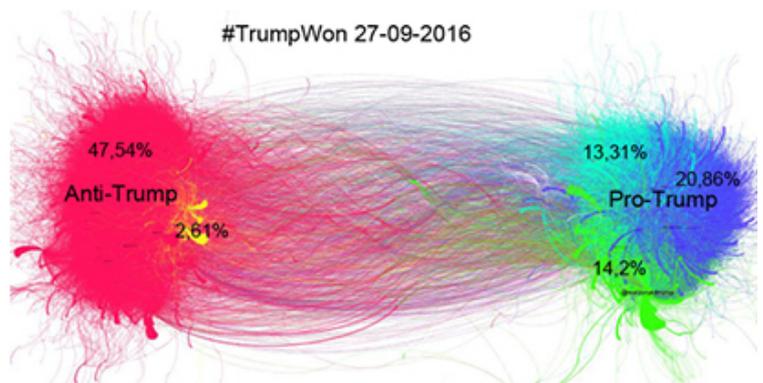
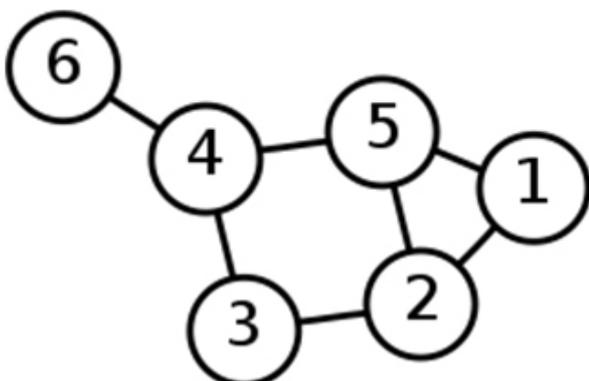
The diagram includes the following callouts:

- DEPRECATED**: The tweet's unique ID. These IDs are roughly sorted & developers should treat them as opaque (<http://bit.ly/dCkppc>).
- Text of the tweet**: Consecutive duplicate tweets are rejected. 140 character max (<http://bit.ly/4ud3he>).
- Tweet's creation date**: The ID of an existing tweet that this tweet is in reply to. Won't be set unless the author of the referenced tweet is mentioned.
- The author's user ID**: The screen name & user ID of replied to tweet author.
- The author's user name**: Truncated to 140 characters. Only possible from SMS.
- The author's biography**: The author's screen name.
- The author's URL**: The author's "location". This is a free-form text field, and there are no guarantees on whether it can be geocoded.
- Rendering information for the author**: Colors are encoded in hex values (RGB).
- Whether this account has contributors enabled**: (<http://bit.ly/50npuu>).
- Number of favorites this user has**: Number of tweets this user has.
- Number of users this user is following**: The creation date for this account.
- The user's selected language**: The timezone and offset (in seconds) for this user.
- The author of the tweet**: This embedded object can get out of sync.
- The author's location**: The author's "location". This is a free-form text field, and there are no guarantees on whether it can be geocoded.
- The author's screen name**: The author's screen name.
- The author's URL**: The author's URL.
- Rendering information for the author**: Colors are encoded in hex values (RGB).
- Whether this account has contributors enabled**: (<http://bit.ly/50npuu>).
- Number of favorites this user has**: Number of tweets this user has.
- Number of users this user is following**: The creation date for this account.
- The user's selected language**: The timezone and offset (in seconds) for this user.

Fuente: [Raffi Krikorian](#).

**Qué es una API.** Una API, abreviada del inglés *application programming interface*, es la forma estructurada de acceder e interactuar con una aplicación o web.

**Qué es un grafo.** Un grafo es un conjunto de objetos en los que algunos de ellos están relacionados. Los objetos se suelen llamar «nodos» o «vértices», y la relación entre dos de ellos, «arista». Se suelen representar como diagramas de puntos o círculos conectados por líneas. Veamos un par de grafos, uno sencillo (izquierda) y una red tan densa en la que solo se representan las aristas y no los nodos:



[Wikipedia](#) (izquierda); [Mariluz Congosto](#) (derecha).

**Qué es T-hoarder.** Es un software libre desarrollado por Mariluz Congosto en Python que sirve para interactuar de una forma sencilla con la API de Twitter, a la hora de obtener información, de forma automatizada, y procesarla para facilitar su análisis. Genera archivos .gdf que sirven para hacer grafos con el programa Gephi. Si tienes problemas para correr Python en tu ordenador, T-hoarder viene con una máquina virtual que permite trabajar con ello sin necesidad de instalar nada. T-hoarder dispone de dos versiones: t-hoarder y t-hoarder\_kit.

- [t-hoarder\\_kit \(\\*\)](#) es para analizar HT o cosas que han sucedido. Tiene más herramientas para analizar usuarios, generar grafos, procesar información.
- t-hoarder es para monitorizar algo que va a suceder, por ejemplo, elecciones, una convocatoria, un evento, etc. Es más para almacenar información de larga duración.

**Qué es una máquina virtual.** Una máquina virtual es un software que simula un sistema de computación y puede ejecutar programas como si fuese una computadora real.

**Qué es Python.** Ya lo sabes, pero te lo recordamos: es un lenguaje de programación que toma el nombre de los humoristas británicos Monty Python.

**Qué es Gephi.** Es un software *open source* escrito en Java para analizar y visualizar redes.

## 6. Tres ejercicios de periodismo de datos

### 6.1. Twitter: captura de datos y análisis de redes

#### 6.1.4. Preguntas a resolver

##### 6.1.4.1. ¿Quiénes están detrás de un *trending topic*?

¿Podrías decir si es dirigido o espontáneo?, ¿qué comunidades lo apoyan?, ¿hay redes de *bots*?

#### a. Qué análisis o visualización responde a la pregunta

A la hora de analizar una gran cantidad de tuits, un grafo (un conjunto de nodos y aristas) es la manera más eficaz y sencilla para entender cómo se configura una red. En este tipo de gráficos se representan los usuarios como los nodos y los retuits (republicación de un mensaje de otro usuario) como las aristas que relacionan un nodo y otro. Si el usuario A retuitea al usuario B, habrá una línea (arista) del punto A al B. Los retuits son el método más usado, pero también pueden usarse las menciones o las respuestas. Otra forma de analizar los datos es representar las relaciones de los usuarios que han usado ese *hashtag* según se sigan unos a otros.

#### b. Qué datos necesito y dónde los encuentro

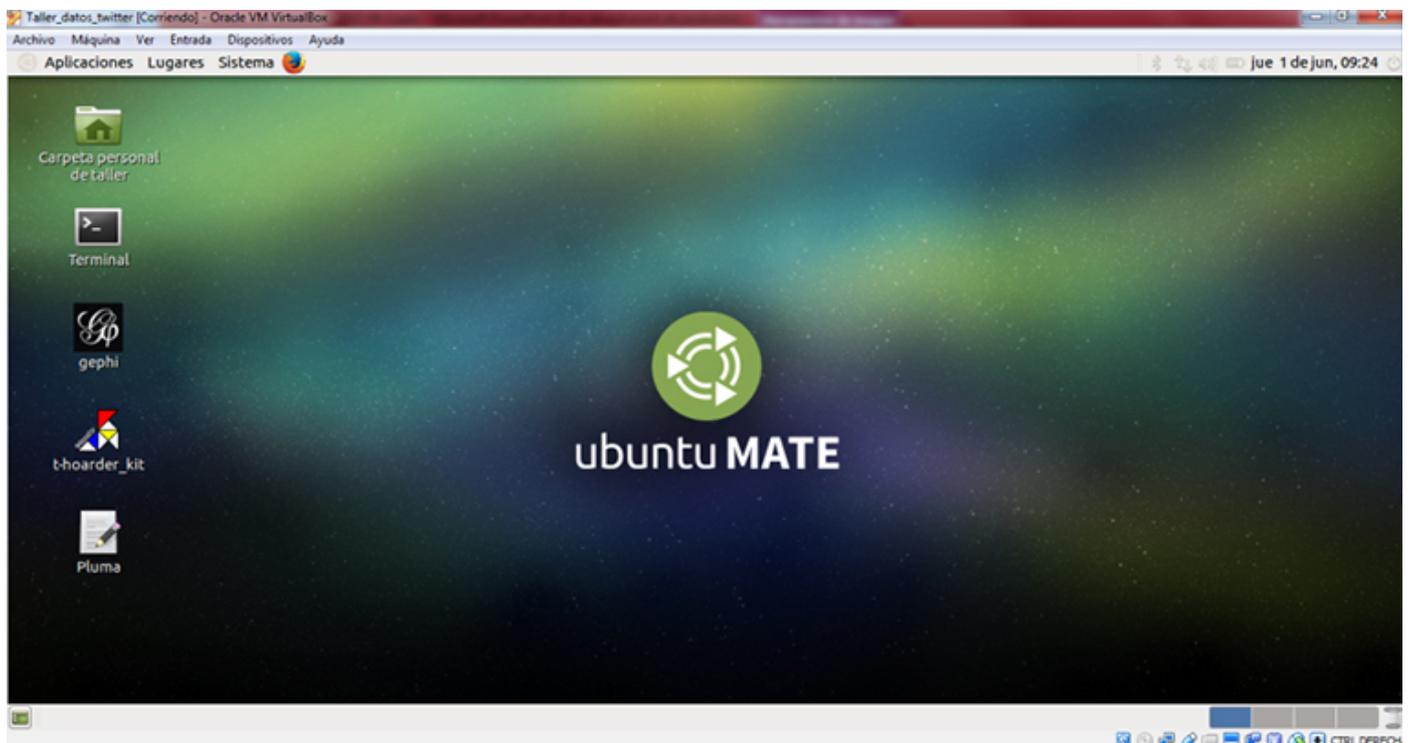
Para poder analizar los tuits de un determinado *hashtag* hace falta, en primer lugar, obtenerlos. Para ello, el método más sencillo es usar la API de Twitter ([Lo que siempre quiso saber del API de Twitter y nunca se atrevió a preguntar](#)) para capturar esos tuits. Para hacer este ejercicio hemos usado el hashtag #xxxxx y durante xxxx horas hemos capturado xxx tuits. Existen otras opciones para analizar un *hashtag* en directo, como la herramienta [Flocker](#), que descarga los tuits en directo y dibuja el grafo que se forma.

```
Cartagena. Licenciado en Periodismo por la Universidad Complutense de
Madrid y Máster ABC-UCM. http://pbs.twimg.com/media/DA_QyLuW0AAoAJO.jpg
photo es
7 869181101377548289 2017-05-29 13:18:17 @azulymarta RT @IdoiaVR: Operación
Púnica: El juez pide al Supremo que impute a la senadora popular Pilar
Barreiro Suma y sigue https://t.co/OU0B6NpIwb via=Twitter for Android
id=226302820 followers=144 following=417 statuses=2737 loc=madrid http://
www.elconfidencial.com/espana/2017-05-29/punica-pp-senadora-pilar-barreiro-
audiencia-nacional-juez-velasco_1389435/ None charo bermudo soy de
sevilla, vivo en madrid, hecho de menos mi familia y mi tierra , tengo 2
hijos maravillosos y un marido que es un sol,mi pasión la vida la gente
intentar None None es
8 869181237390397440 2017-05-29 13:18:50 @Cat72222 https://t.co/QbkpgGbd1r
La futura rana🐸via=Twitter for Android id=4806190881 followers=3035
following=3417 statuses=42479 loc=En algun lugar http://
```

Tuit capturado por T-warder en un archivo de texto plano. Metadata separados por tabuladores

Para facilitar el proceso de captura con la API vamos a usar T-warder, el software libre que ha desarrollado Mariluz Congosto. Se trata de una aplicación ligera que corre en Python y que permite interactuar con la API de Twitter de una forma sencilla para obtener tuits y otras informaciones (perfiles de usuarios, seguidores...) y guardarlos en formatos fácilmente reutilizables (archivos de texto plano). T-warder puede usarse de dos modos:

- Usando el código disponible en [github](#) corriendo Python directamente en la consola. Accede al [manual de instalación](#) para Windows y Linux/Unix para más detalle.
- Mediante una máquina virtual en la que se accede a un entorno en el que se puede usar T-warder y Gephi. Es necesario usar el software [Virtual box](#) para hacer funcionar la máquina virtual. Descarga la máquina virtual (3,9GB) que viene preinstalada con T-warder\_kit y Gephi: [https://www.dropbox.com/s/18gt69suptie5hw/ubuntu1602\\_taller.ova](https://www.dropbox.com/s/18gt69suptie5hw/ubuntu1602_taller.ova) (usuario: taller; contraseña: tallerdatos).



Escritorio de la máquina virtual

En la máquina virtual basta con hacer doble clic en el icono de T-hoarder para arrancar el programa.

Si corres directamente t-hoarder\_kit en tu consola al directorio `/scripts/` donde hayas descargado t-hoarder\_kit y ejecuta `./t_hoarder_kit.sh`.

En ambos casos te pedirá donde el archivo de «application keys»; vamos a ver cómo generarlo:

```
numeroteca@cinco ~/sites/t-hoarder_kit/scripts (master *) $ ./t_hoarder_kit.sh
-----> Welcome to t-hoarder kit <-----
-----
t-hoarder_kit is up to date
working in /home/numeroteca/sites/t-hoarder_kit

-----> Environment data <-----
-----
Enter the file name with the application keys: forthoarder.key
Enter a twitter user: numeroteca
Enter experiment name: pruebas
```

### Generando Keys en Twitter Apps

Es necesario crear una aplicación con la que obtener el acceso a la API de Twitter. Ve a Twitter Apps para crear una nueva aplicación: <https://apps.twitter.com>. Para poder acceder necesitas crear un usuario de Twitter si no lo tienes ya.

By using Twitter's services you agree to our [Cookie Use](#) and [Data Transfer](#) outside the EU. We and our partners operate globally and use cookies, including for analytics, personalisation, and ads. ×

## Twitter Apps

[Create New App](#)

numeroteca  
numeroteca



jip2  
para usar twitter desde R



Rellena los datos de la aplicación según se indica en la imagen:

## Create an application

### Application Details

Name \*

Nombre de la app, no debe existir otro igual

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description \*

Hay que poner algo de texto, no vale solo una palabra

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website \*

Una url válida

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.  
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Dejar en blanco

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth\_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

En la pestaña Keys and Access Tokens, copia el Consumer Key y el Consumer Secret:

# aplicacionprueba23

Test OAuth

Details Settings Keys and Access Tokens Permissions

## Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) Im6FbN52cSsCLxCcQwuWmE

Consumer Key

Consumer Secret (API Secret) 5gag1URENi7c3JEGHEBRpEKi2GJfetcJ0xZm8DuPh9XuzC

Consumer Secret

Access Level Read and write (modify app permissions)

Owner numeroteca

Owner ID 239362738

Creas un archivo de nombre `forthoarder.key` en el que en la primera línea esté la Consumer Key (API Key) y en la segunda el Consumer Secret (API Secret); quedará algo como esto:

```
W960clipfwefVxEdg
jKt9GVQ3BHOq0Jp4hHIOlB6npOD8n5feL89ruE
```

Y sávalo con el nombre `forthoarder.key` en el directorio `/keys/` de tu instalación de `t-hoarder`. Utiliza una aplicación de edición de texto plano (notepad, gedit...).

MyApp.key (-/t-hoarder\_kit/keys) - Pluma

Archivo Editar Ver Buscar Herramientas Documentos Ayuda

MyApp.key x

1

- Copiar y pegar la primera línea la Consumer key
- Copiar y pegar en la segunda línea la Consumer secret
- Atención!!! Verificar que no se dejan espacios ni tabulaciones
- Guardar en el directorio keys con el nombre de la app y la extensión key

Es seguro | https://apps.twitter.com/app/12122751/keys

Application Management

By using Twitter's services you agree to our [Cookie Use](#) and [Data Transfer](#) outside the EU. We use cookies, including for analytics, personalisation, and ads.

### taller\_datos

Details Settings Keys and Access Tokens Permissions

#### Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) [Redacted]

Consumer Secret (API Secret) [Redacted]

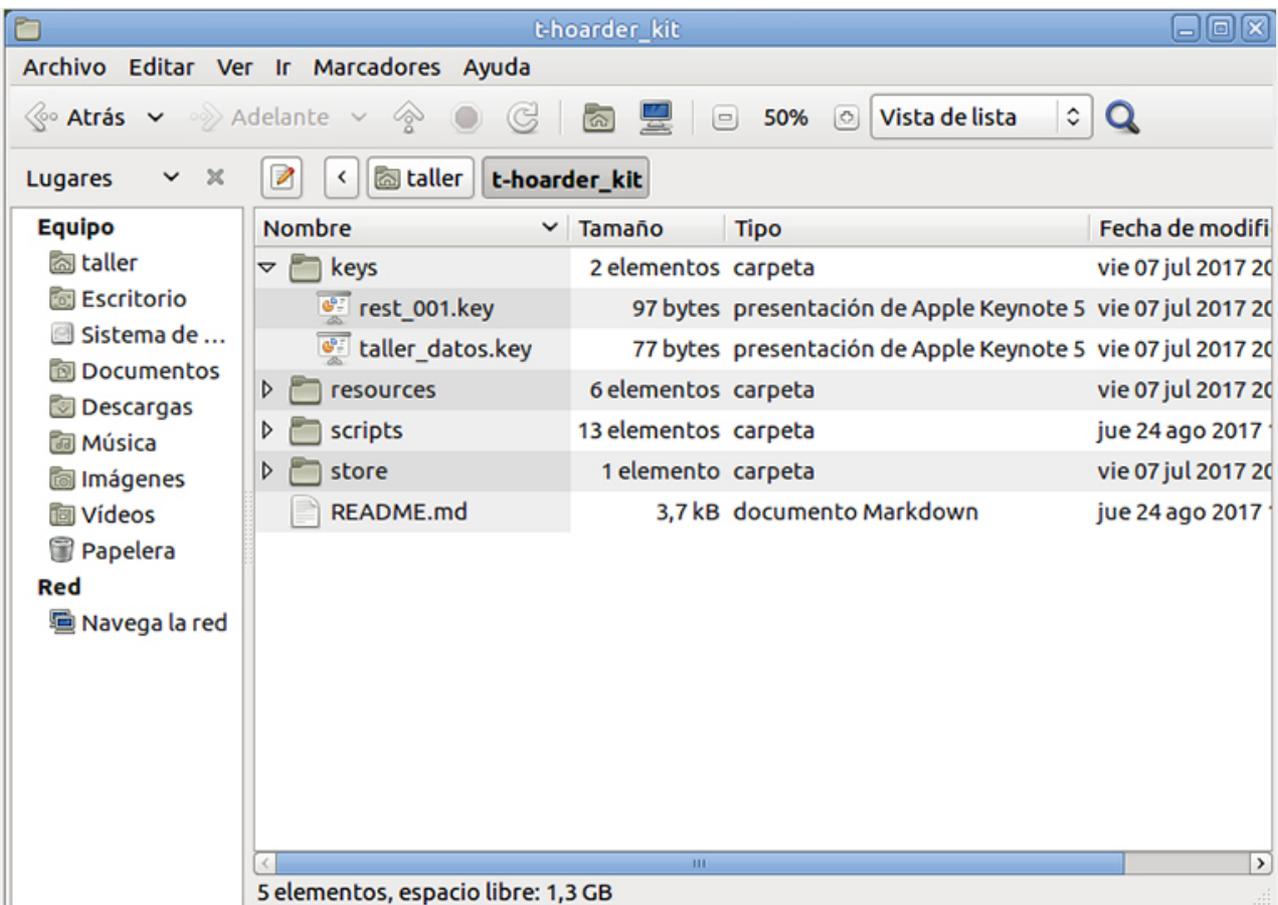
Access Level Read and write (modify app permissions)

Owner congosto

Owner ID 14620824

Ya tienes todo listo para el siguiente paso.

Así debería quedar la estructura de carpetas:



En este caso las claves creadas son `rest_001.key` y `taller_datos.key`.

La aplicación en la consola te va preguntando y tienes que ir dando los datos. Los datos que pide al inicio son para el contexto del trabajo: se necesita una app, un usuario Twitter y un directorio de trabajo previamente creado para dejar los resultados.

```

-----> Environment data <-----
-----
Enter the file name with the application keys: forthoarder.key
Enter a twitter user: numeroteca
Enter experiment name: pruebas

```

1. Enter the file name with the application keys: `forthoarder.key`.
2. Enter a twitter user: `mi_user`.
3. Enter experiment name: `pruebas`.

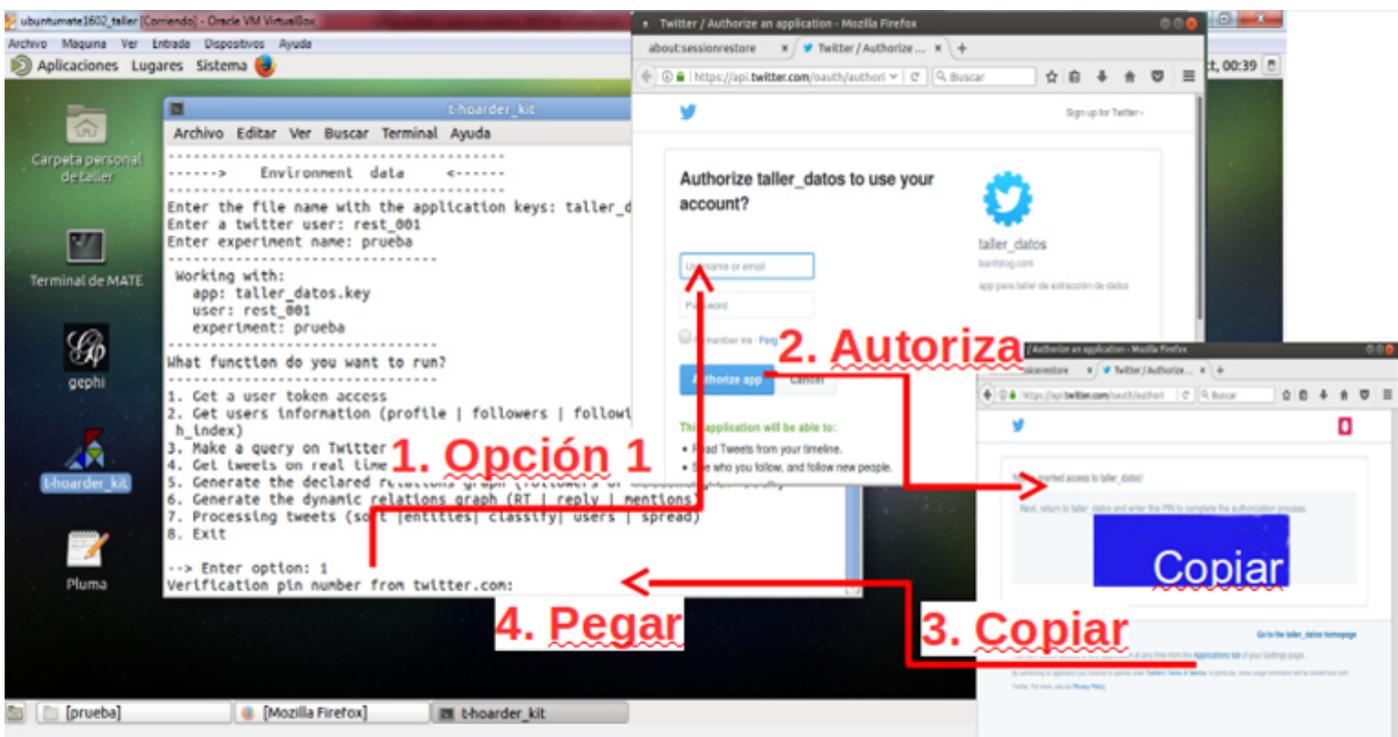
Si estás utilizando la máquina virtual pon como usuario `mi_user`.

Donde `forthoarder.key` es el archivo con nuestras claves, `numeroteca` es el usuario con el que has creado la aplicación, y `taller` es el directorio que debes crear en la carpeta `/store/`. Puedes tener tantos experimentos como quieras; cada uno en su correspondiente directorio.

Si todo ha ido bien, deberíamos ver el siguiente menú:

```
-----  
Working with:  
app: forthoarder.key  
user: numeroteca  
experiment: pruebas  
-----  
What function do you want to run?  
-----  
1. Get a user token access  
2. Get users information (profile | followers | following | relations | tweets | h_index)  
3. Make a query on Twitter  
4. Get tweets on real time  
5. Generate the declared relations graph (followers or following or both)  
6. Generate the dynamic relations graph (RT | reply | mentions)  
7. Processing tweets (sort | entities | classify | users | spread)  
8. Exit  
--> Enter option: 1
```

Seleccionamos la opción 1 para generar el permiso de nuestro t-warder\_kit desde nuestra máquina. Esto abrirá el navegador que tengamos asignado por defecto; metemos nuestro usuario y *password* y nos dará un número de 7 cifras que tenemos que copiar y pegar en la aplicación:



Si todo ha ido bien, veremos algo así en la consola:

```
Verification pin number from twitter.com: 6347239  
Access token generado con éxito. Guardado en keys/numeroteca.key
```

Y se habrá generado un archivo `numeroteca.key` en nuestro directorio `/keys`.

Ya está todo listo para empezar a capturar datos.

#### Nota

T-warder no captura toda la información de los tuits (ver documentación, anatomía de un tuit) sino «id, user, ... [completar]» y la almacena en archivos de texto plano (.txt). Otros software almacenan la información en bases de

datos (postgres, SQL...), lo que dificulta su tratamiento y la capacidad para ser compartida de forma rápida y sencilla.

Elige la opción 3. Make a query on Twitter para descargar tuits de una búsqueda. Por ejemplo, si queremos descargar retroactivamente el *hashtag* #parlament (crea antes en la carpeta del experimento el archivo vacío `parlament-tweets-query.txt`):

```
--> Enter option: 3
Enter a query (allows AND / OR connectors): #parlament
Enter output file name: parlament-tweets-query.txt
/home/numeroteca/sites/t-hoarder_kit/store/pruebas/parlament-tweets-query.txt
>>>/home/numeroteca/sites/t-hoarder_kit/store/pruebas//parlament-tweets-query.txt file
exists, do you want to append more tweets (y/n)?y
```

Si todo va bien verás algo como esto, el script que informa que los tuits se están capturando bien:

```
collected 0
{'u'type': u'Point', u'coordinates': [-3.7058, 40.4203]}
remaing hits 179
collected 100
remaing hits 178
collected 200
remaing hits 177
collected 300
[...]
```

Si se quiere parar la captura de tuits, presionar `ctrl + c`.

Si abrimos el archivo `store/pruebas/parlament-tweets-query.txt` podemos ver los tuits capturados.

Otra opción es descargar los tuits en tiempo real con la streaming API; para ello, usa la opción 4. Get tweets on real time:

```
--> Enter option: 4
Enter input file name with the keywords separated by , : palabras_clave.txt
```

Se necesita un archivo de texto plano con las palabras clave para hacer la búsqueda de tuits. Por ejemplo, si quieres buscar el *hashtag* #parlament, el archivo solamente contendrá #parlament. Si quieres hacer búsquedas de otras palabras, sepáralas por comas. Prueba en Twitter sur bús

### c. Qué herramientas uso para preparar los datos

Una vez que hemos obtenido los tuits, necesitamos convertirlos en un formato interpretable por Gephi, un software libre y muy extendido para generar grafos. T-hoarder de nuevo nos facilita convertir una serie de tuits en un archivo `.gdf`.

Antes de procesar los datos podemos limpiarlos y filtrarlos de falsos positivos; por ejemplo, si estamos capturando tuits sobre el Parlamento catalán, los que no hagan referencia a ese tema. Con herramientas como R o vi se podrían eliminar, por ejemplo, todos los tuits que contengan tal o cual palabra.

## ¿Qué hacer con la lista de tuits?

A partir de una lista de tuits se pueden extraer muchos análisis: usuarios más retuiteados, número de tuits por minuto/hora/día. A partir de los tuits se pueden obtener los id de los usuarios y con los usuarios puedo calcular sus relaciones (*follower-following*). Ojo: para obtener los *following* de cada usuario, ¡la API te da los *following* de 60 usuarios cada hora!

Según se procesen los tuits, se pueden obtener dos tipos de grafos:

- De relaciones declaradas (*following* y *followers* entre los usuarios del *hashtag*).
- De relaciones dinámicas (los que se retuitean).

Ambas son interesantes a la hora de analizar quién está detrás de un *hashtag*, pero para realizar la primera de ellas (relaciones declaradas) es necesario obtener nuevos datos: ¿a quién siguen los usuarios que usan ese *hashtag*? [ver siguiente pregunta].

Veamos paso a paso cómo generar el .gdf con t-hoarder:kit. Selecciona la opción 6. Generate the dynamic relations graph (RT | reply | mentions):

```
--> Enter option: 6
Enter input file name with the tweets (got from a query or in real time): parliament-
tweets-query.txt.log
Enter the relationship type (RT | reply | mention): RT
Introduce top size (100-50000): 100
-----> Extracting relation RT
-----> second pass
format gdf
type: top nodes: 100
generating gdf file
type: all nodes: 5402
generating gdf file
```

El valor introducido en `Introduce top size` es para «podar» el grafo. Genera un subgrafo con los n nodos más activos y relevantes. Es por si el grafo es muy grande y no nos cabe en Gephi (es demasiado grande y el ordenador no puede manejarlo); podemos generar un grafo menor que no es lo mismo pero es representativo.

Si has analizado una lista de tuits y las relaciones establecidas son sus RT (retuit), se obtienen dos archivos:

- Los top RT: `xxxxxxxx_top_RT.gdf`. Es el grafo podado con los n usuarios (100 en nuestro caso) más retuiteados.
- Todos los RT: `xxxxxxxx_all_RT.gdf`.

Está todo listo para usar estos archivos en Gephi.

## d. Qué herramientas uso para producir el análisis o la visualización

Una vez tenemos el archivo .gdf podemos emprender el penúltimo paso: crear el grafo.

### Paso a paso de cómo hacer un grafo con Gephi

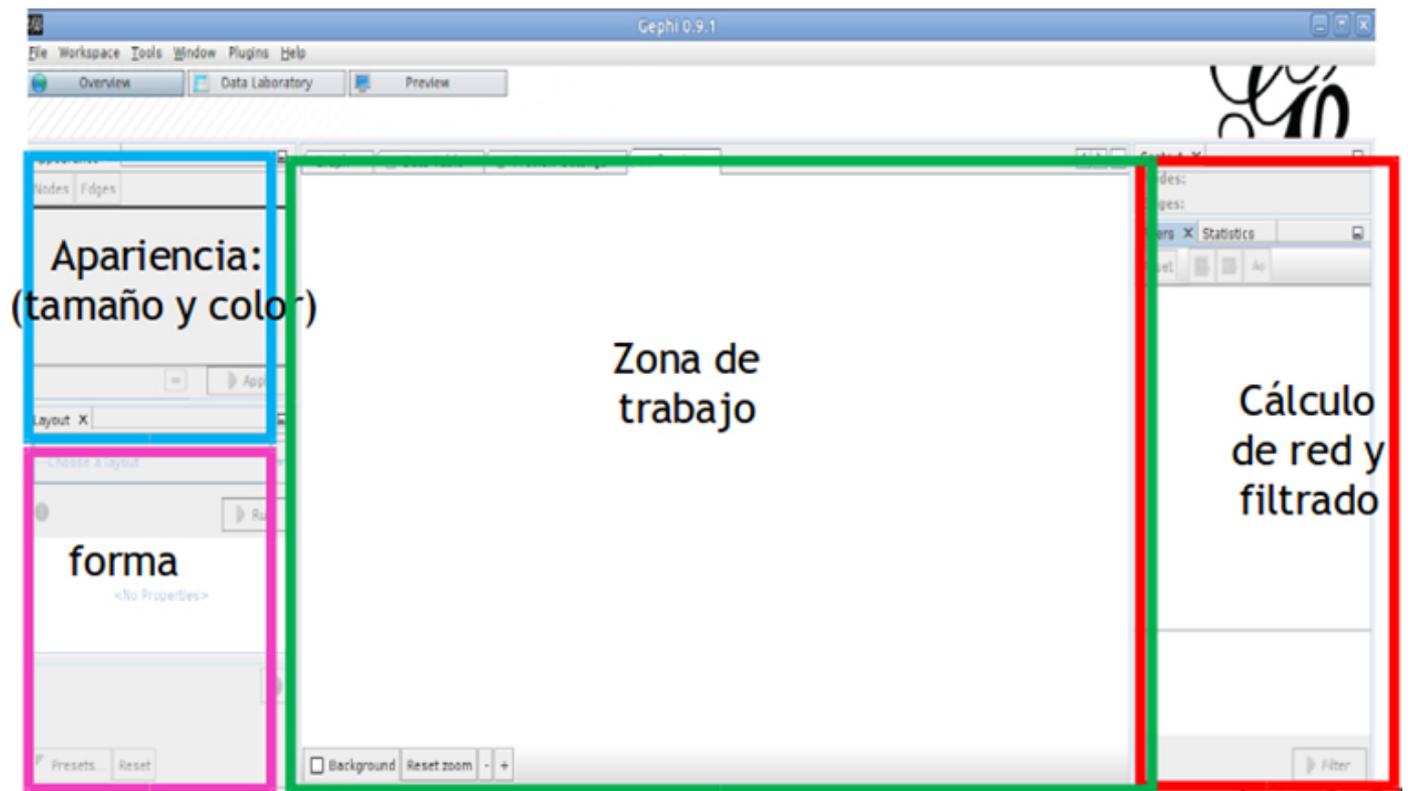
#### 0) Instala y abre Gephi

Descarga <https://gephi.org/users/download/>.

Es un programa que funciona en Java; más que instalarlo, tienes que ejecutarlo cada vez. Desde la línea de comandos ve al directorio `/bin` de los que has descargado y ejecuta `./gephi`. Lo más seguro es que para hacer esto tengas que darle permisos de ejecución al archivo.

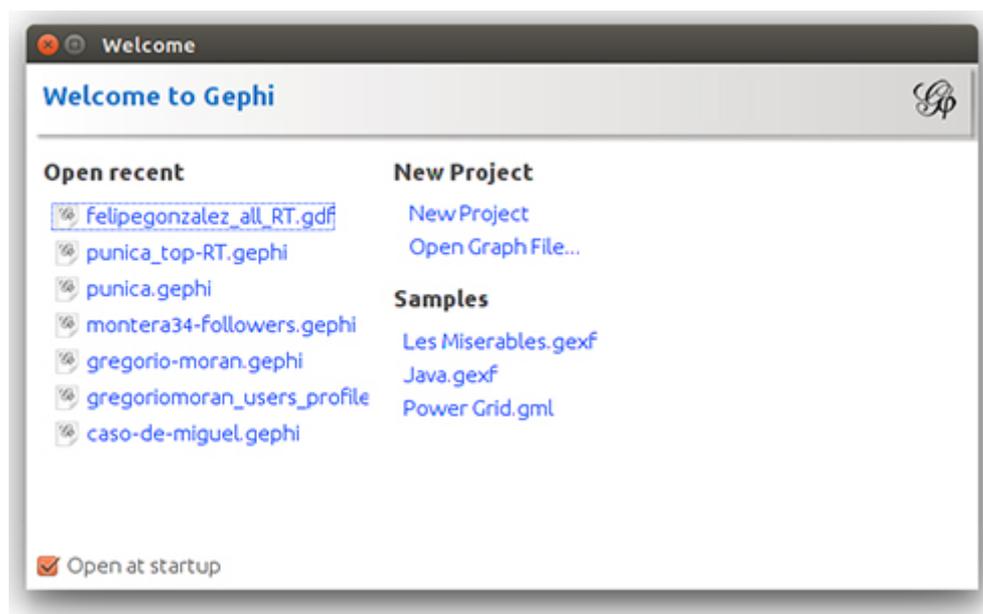
```
numeroteca@cinco ~/Downloads/gephi-0.9.2/bin $ ./gephi
```

Si estás en la máquina virtual, simplemente haz clic en el icono de Gephi del escritorio.

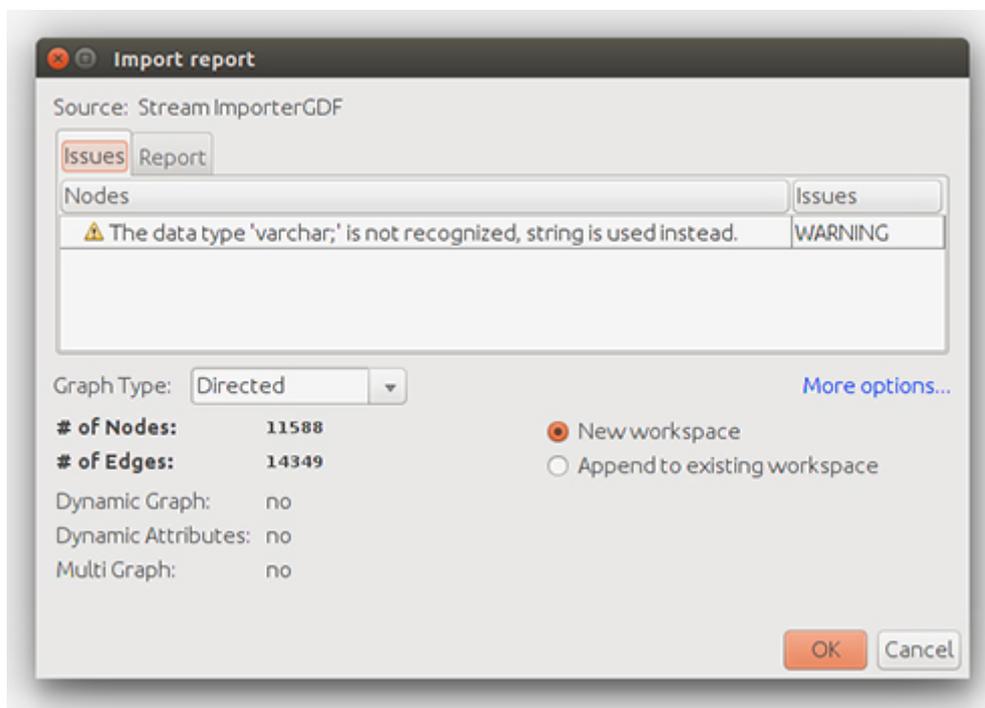


### 1) Genera un grafo

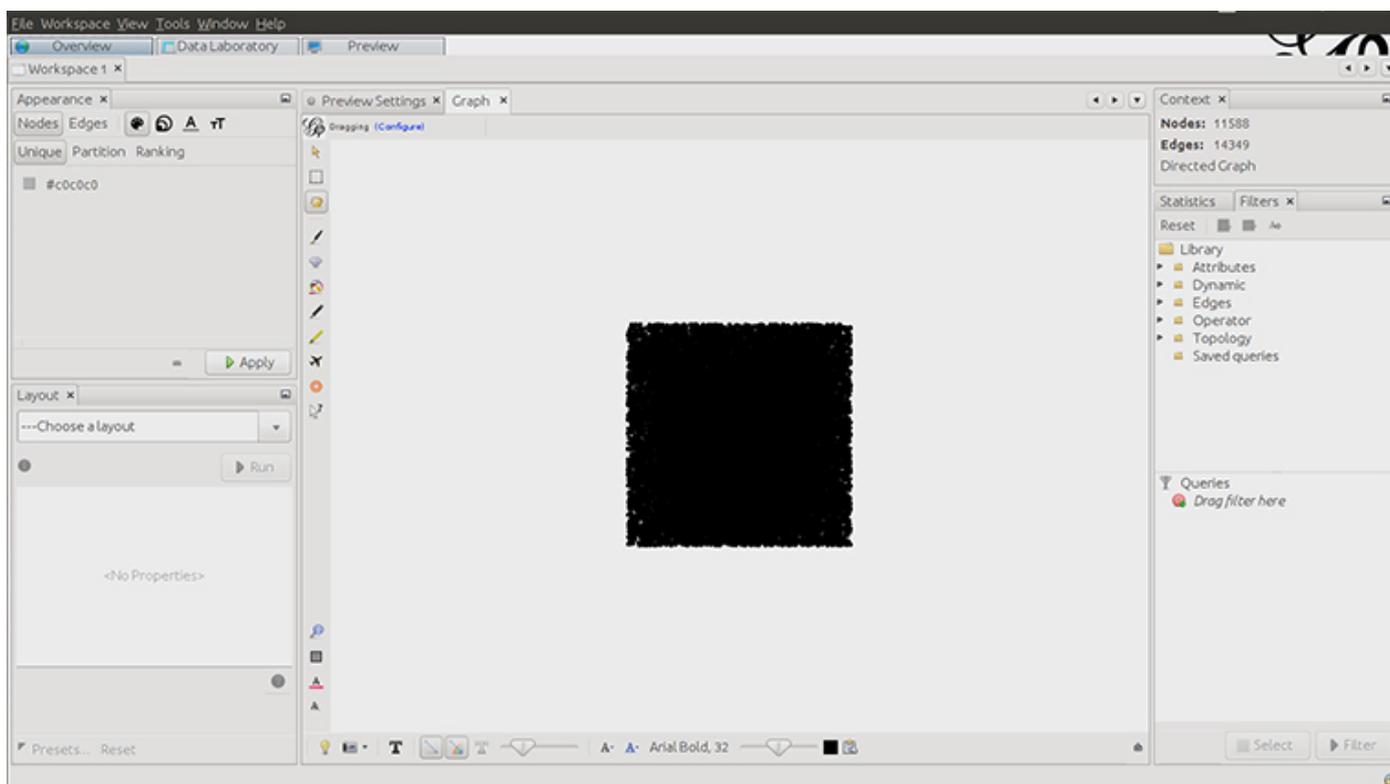
Abre el archivo `xxxxxxx_all_RT.gdf` generado en el paso anterior.



Te preguntará si el grafo es dirigido, no dirigido o mixto («graph style»). Escogemos dirigido, ya que los RT tienen dirección: establecen una arista (flecha) del nodo del usuario que retuitea al originario autor del tuit.

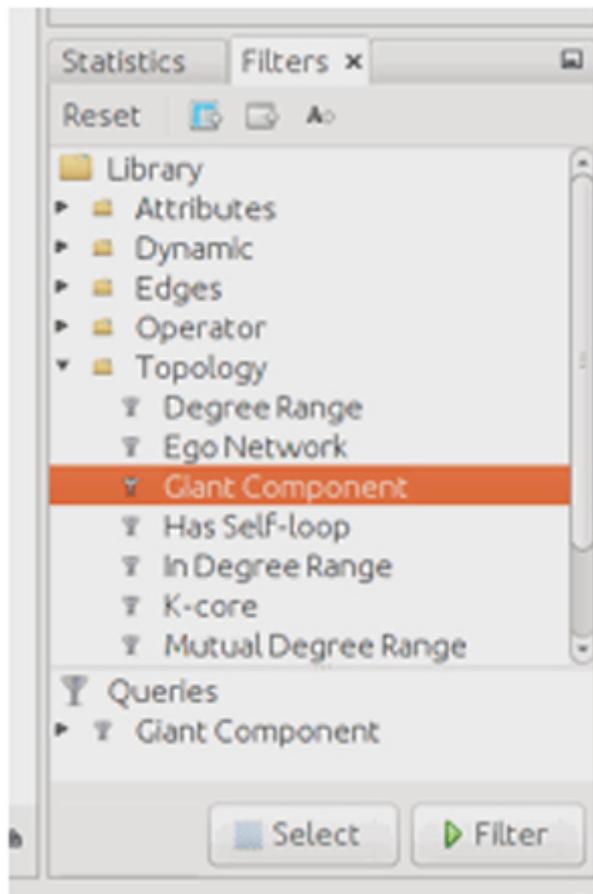


Esto me importa todos los nodos al espacio de trabajo.



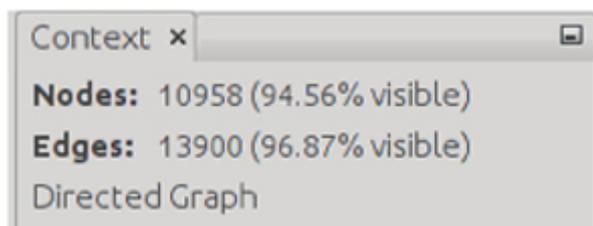
Los nodos recién creados siempre toman esa forma cuadrada.

De la sección «Statistics» de la derecha, en el directorio `Library` -> `Topology` seleccionamos con doble clic «Giant component».

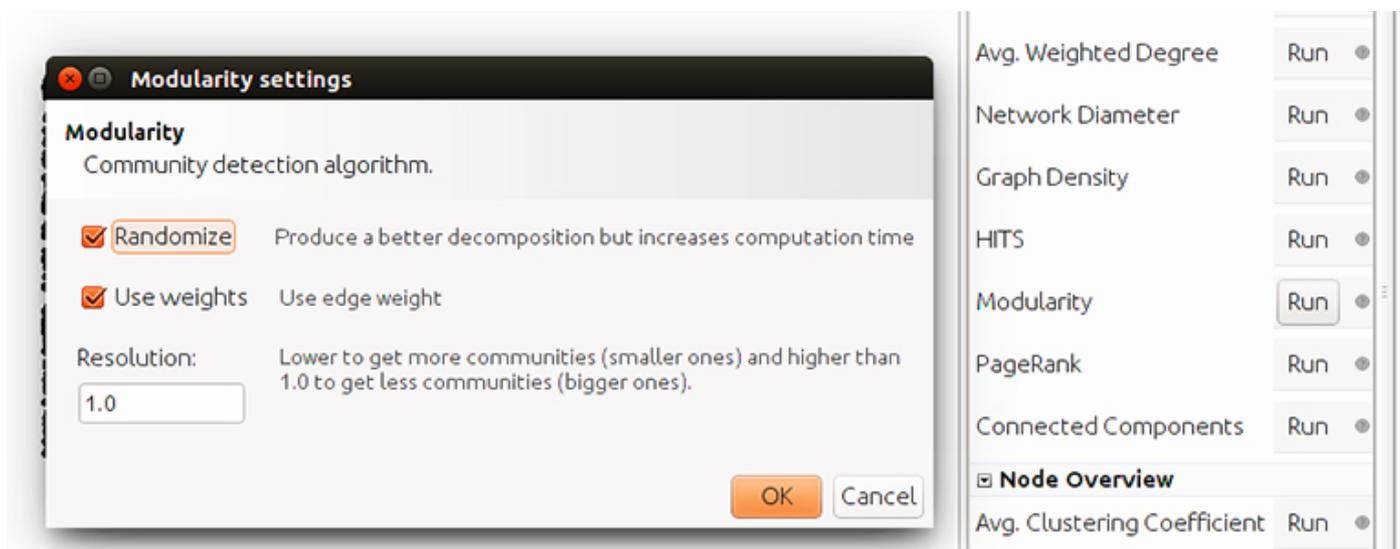


Ahora el filtro «Giant component» está en la lista de Queris. Si lo seleccionamos y hacemos clic en «Filter», estamos aplicando este filtro a nuestros datos.

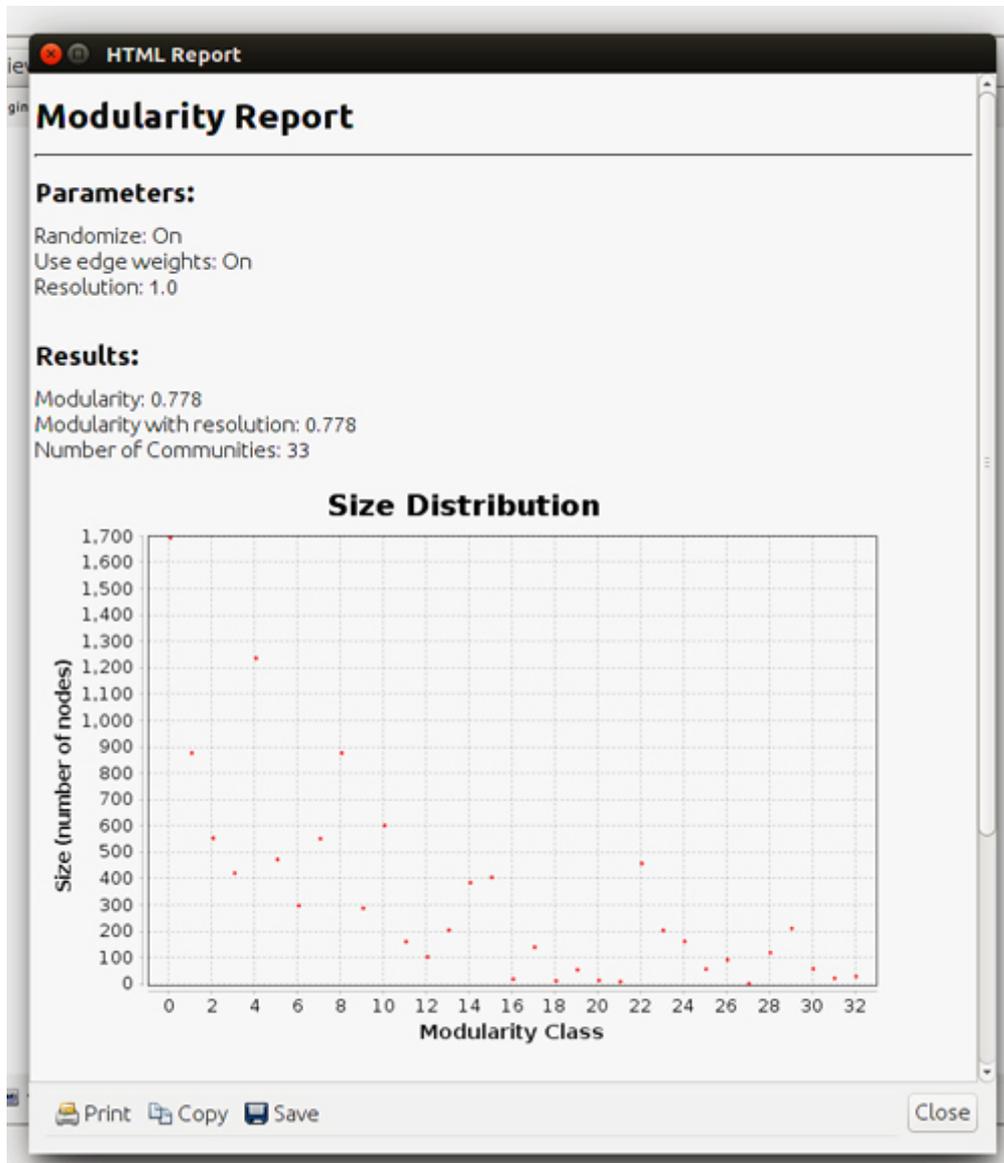
Ver que arriba a la derecha hemos reducido el número de nodos y aristas seleccionadas:

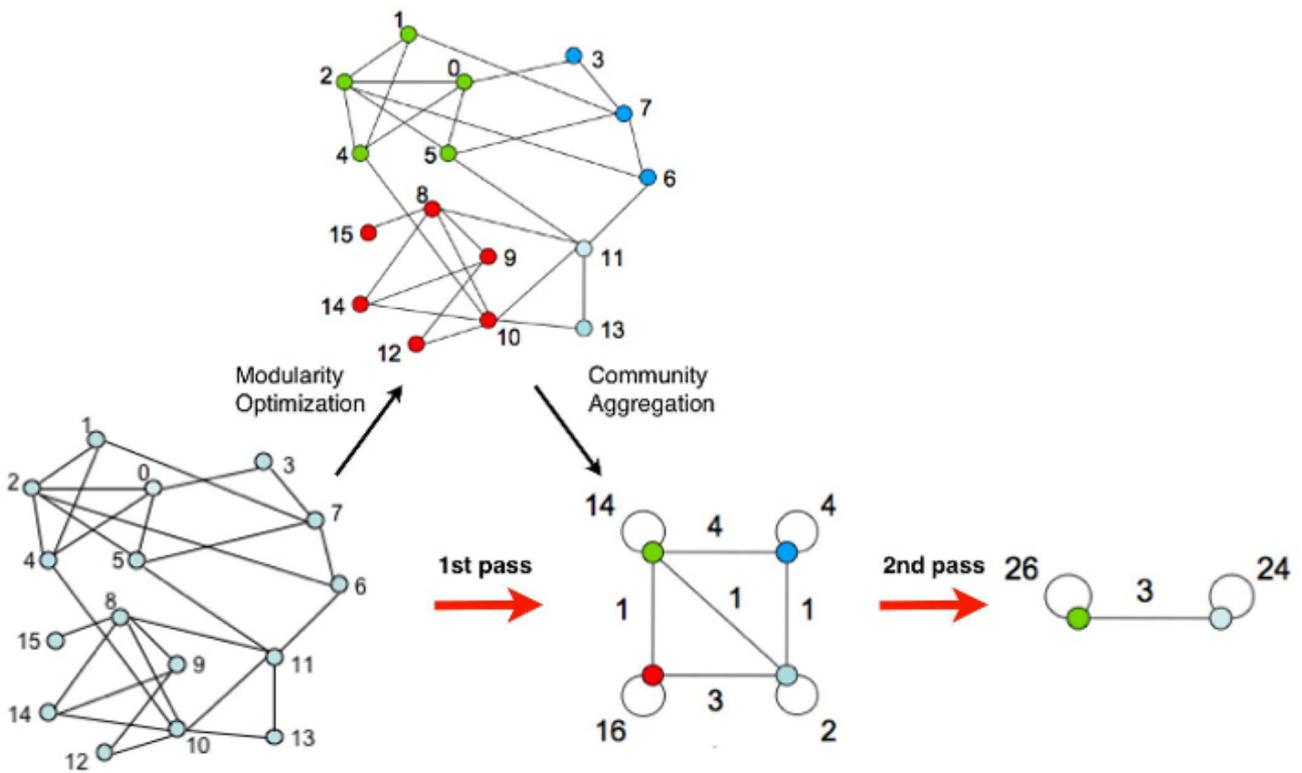


Ahora pasamos a la pestaña «Statistics» y ejecutamos el cálculo de «Modularity» (modularidad de la red) haciendo clic en Run:



Nos da el resultado:





**Figure 1.** Visualization of the steps of our algorithm. Each pass is made of two phases: one where modularity is optimized by allowing only local changes of communities; one where the found communities are aggregated in order to build a new network of communities. The passes are repeated iteratively until no increase of modularity is possible.

Fuente: Algorithm, Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre (2008). «[Fast unfolding of communities in large networks](#)». Journal of Statistical Mechanics: Theory and Experiment (10), P1000.

Ahora tenemos que seleccionar cómo se pintan los nodos. Elegimos colorearlos según la modularidad que acabamos de calcular (si vamos al modo de visualización de datos veremos que se ha añadido una columna con la modularidad). En la pestaña Appearance -> Nodes [paleta pintor] -> Partition selecciona Modularity Class.

The screenshot shows the 'Appearance' panel of a software application. The 'Nodes' tab is active, and the 'Partition' dropdown is set to 'Modularity Class'. A legend lists the following categories and their percentages:

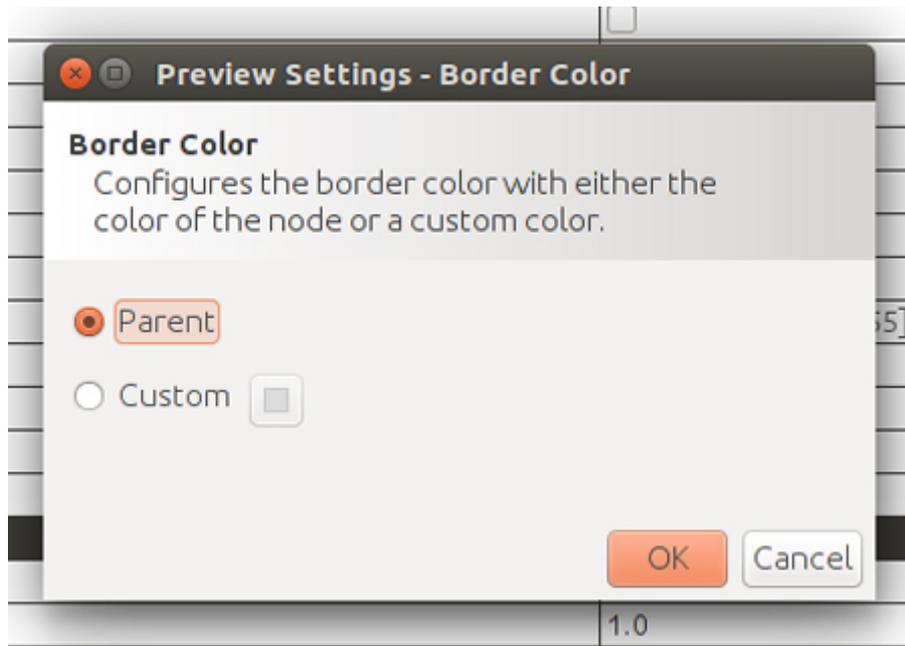
Modularity Class	Percentage
0	(15.5%)
4	(11.33%)
1	(8.05%)
8	(8.05%)
10	(5.5%)
2	(5.1%)
7	(5.08%)
5	(4.36%)

The 'Apply' button is visible at the bottom of the panel. To the right, a large visualization of a network with 882 elements is shown, where nodes are colored according to their modularity class.

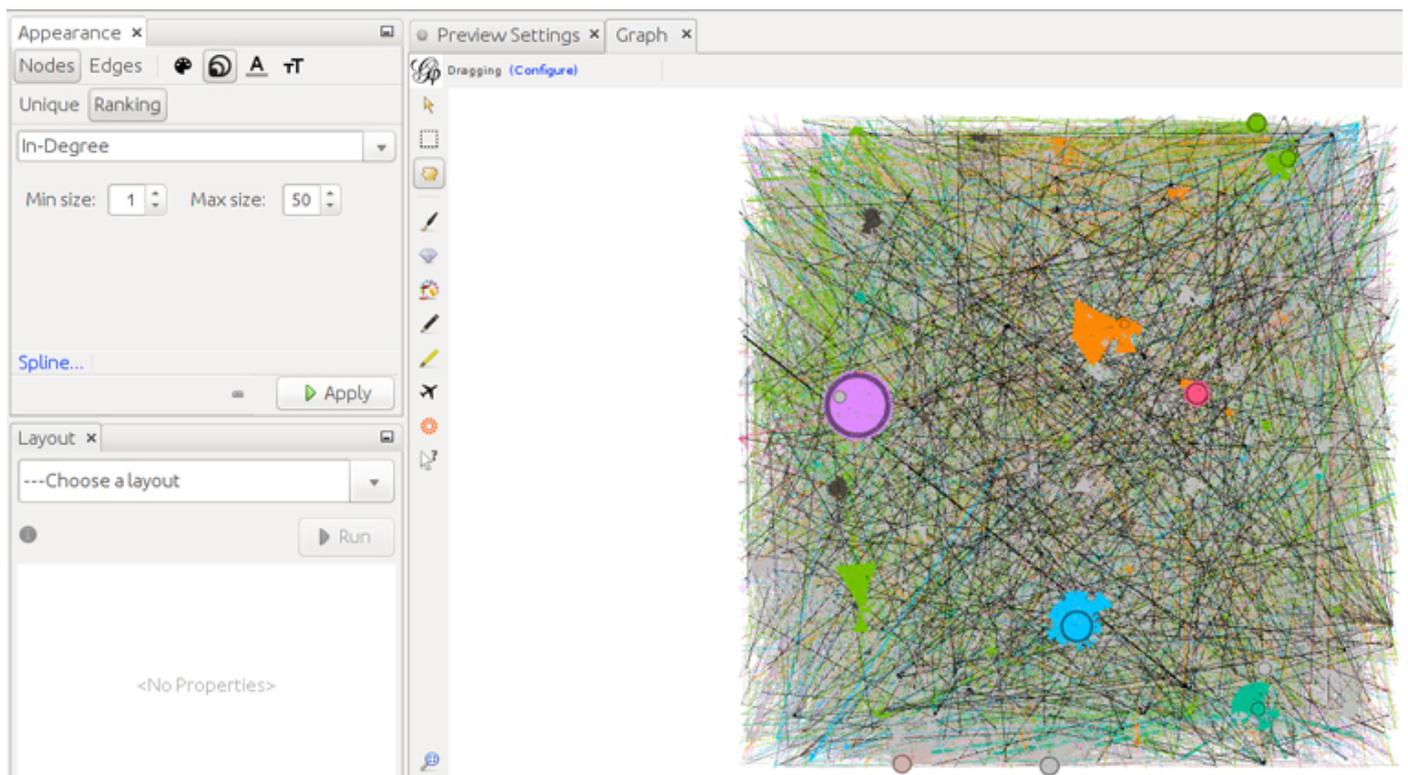
Puedes cambiar el color de cada categoría pulsando en el color de la leyenda.

Ahora los nodos están coloreados en función de la modularidad calculada.

En `Preview Settings` -> `Settings` selecciona la opción «parent» en «Border color» de los nodos:



Ahora modificamos el tamaño de los nodos dependiendo de su «in-degree», que es el número de aristas que reciben, esto es, el número de RT que han recibido sus tuits. Para ello seleccionamos `Appearance` -> `Nodes` -> [tamaño] -> `Ranking` y seleccionamos `In-Degree`.



Ahora en «Layout» seleccionamos «Forceatlas 2» para reubicar la posición de los nodos y conseguir que se agrupen en relación con las comunidades detectadas de «modularidad».

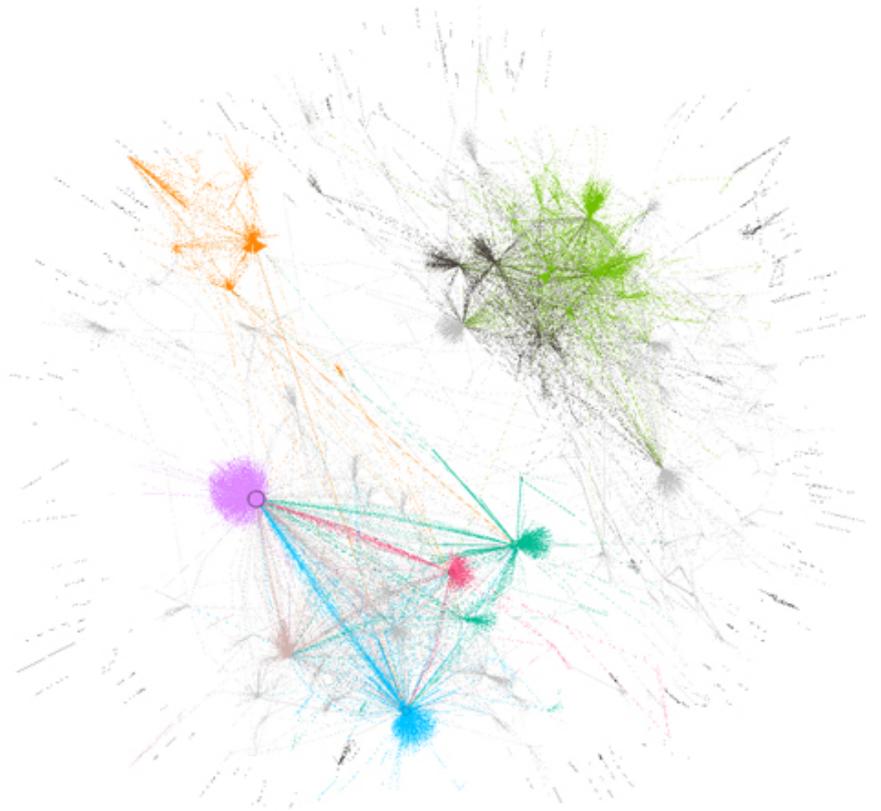
Min size: 1 Max size: 50

Spline... Apply

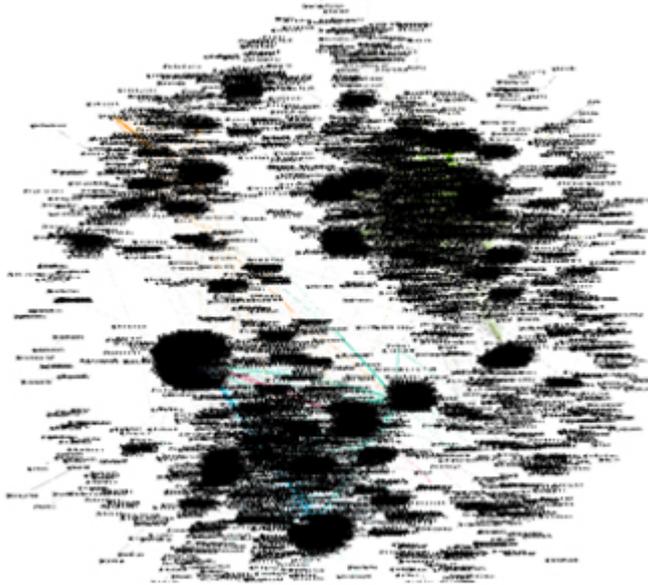
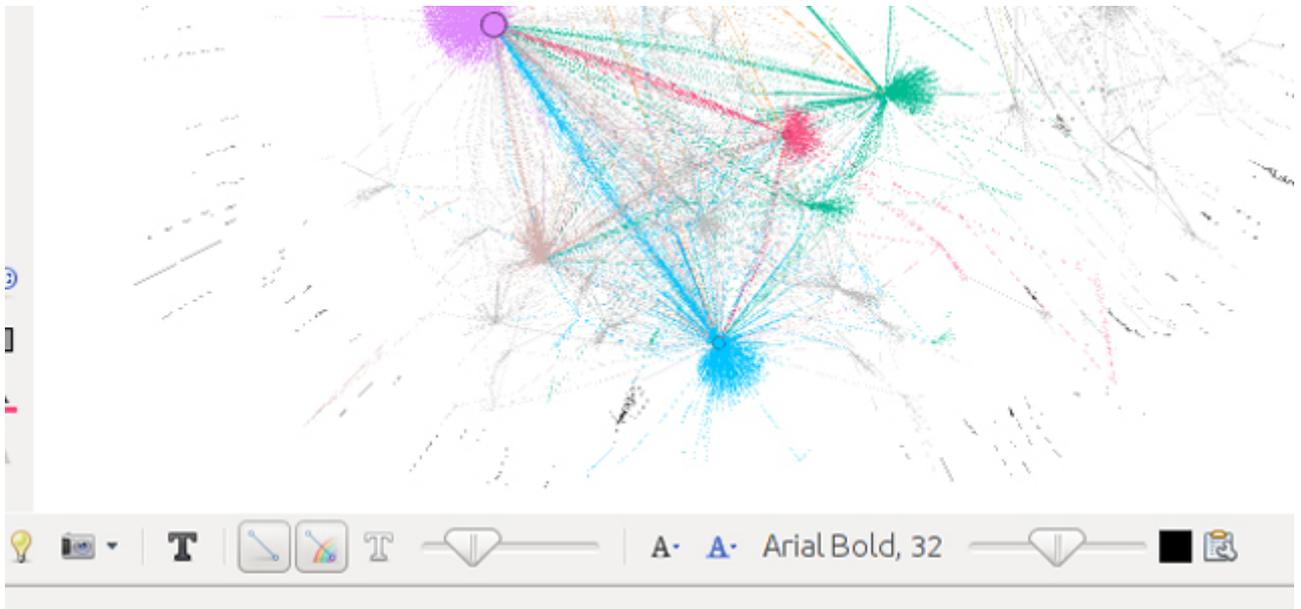
Layout x ForceAtlas 2 Run

Threads number	3
Performance	
Tolerance (speed)	1.0
Approximate Repulsion	<input checked="" type="checkbox"/>
Approximation	1.2
Tuning	
Scaling	4.0
Stronger Gravity	<input type="checkbox"/>
Gravity	2.0
Behavioral Alternatives	

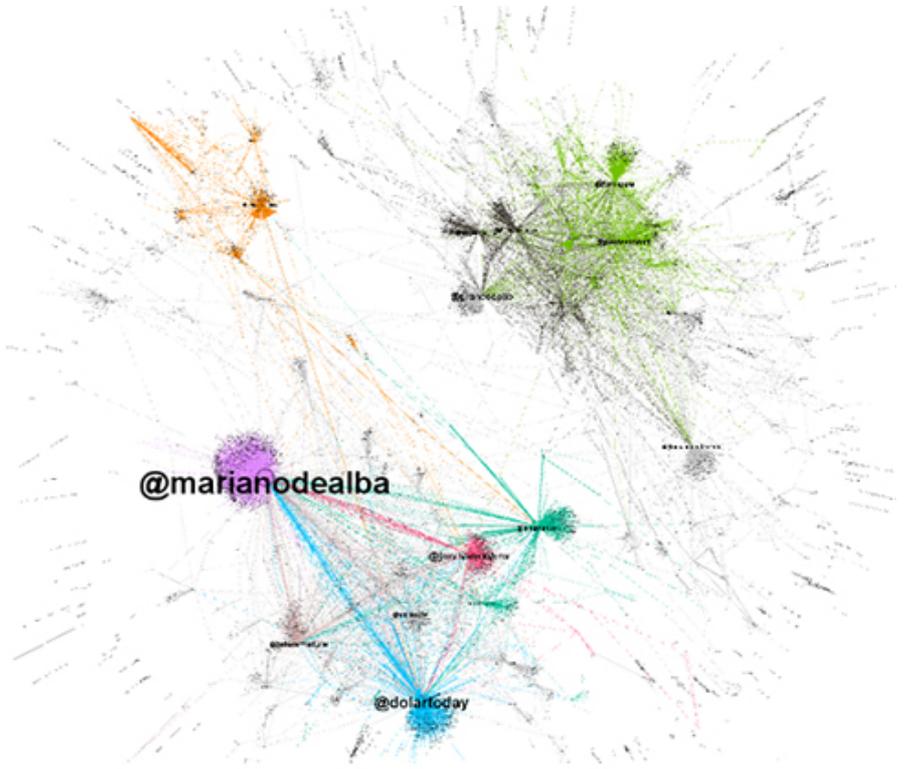
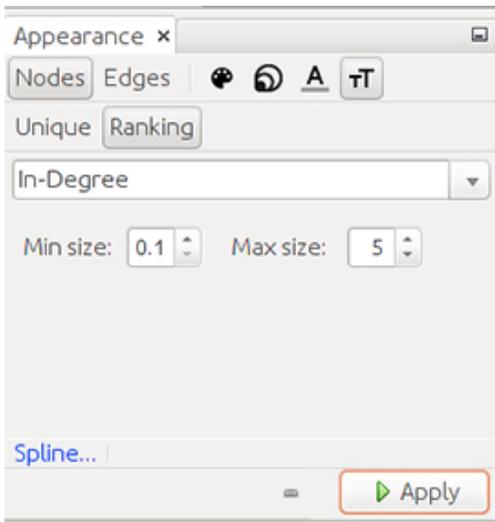
ForceAtlas 2



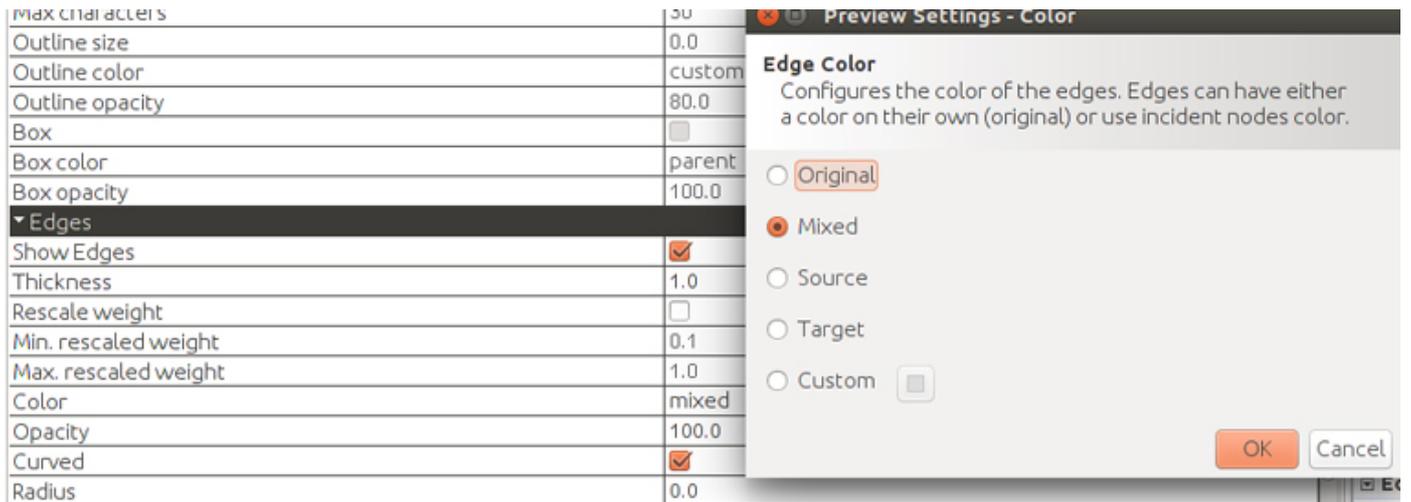
La «T» negra nos permite activar las etiquetas («labels») de los nodos.



La opción de tamaño de texto de Appearance -> Nodes -> [tamaño letras] -> Ranking nos deja seleccionar el tamaño mínimo y máximo de las etiquetas de los nodos en función de la variable, de nuevo, «in-degree». Selecciona de mínimo 1 y 50 de máximo. En Preview settings selecciona una tipografía de 5 puntos de tamaño.



Podemos controlar el color y grosor de las aristas en «Preview» settings. Elegimos «source» para ver de dónde vienen los RT, esto es, el color de la arista será el del retuiteador.

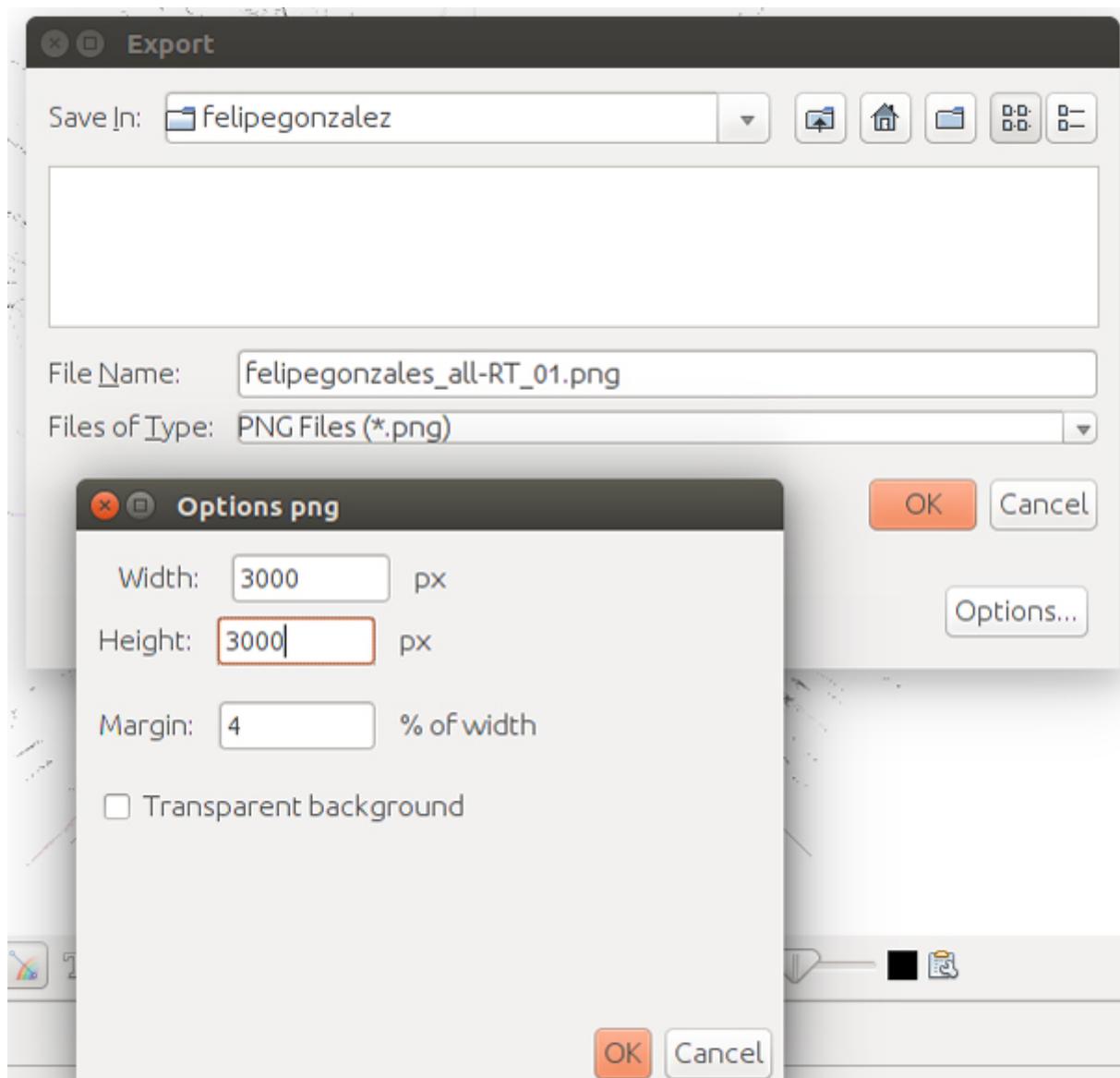


Estas son las opciones:

- SOURCE: The source node's color.
- TARGET: The target node's color.
- MIXED: An average of source and target color.
- CUSTOM: A custom color.
- ORIGINAL: The original edge color, if exists.

A veces, cuando hay muchos RT de un nodo a otro, Gephi no puede manejarlo bien y pinta triángulos o círculos. Se arregla marcando «reescalar pesos».

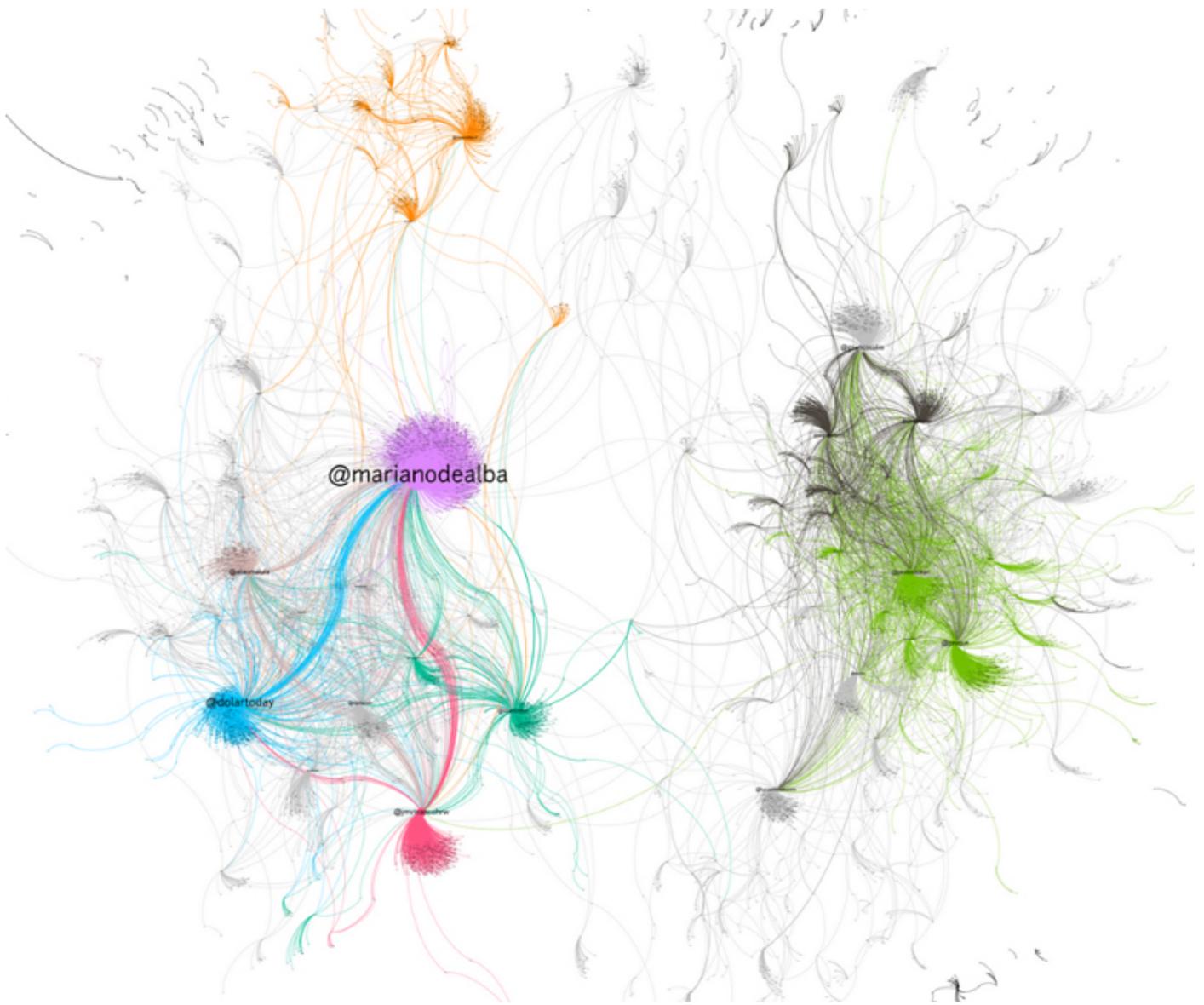
Es el momento de exportar, vamos al menú File -> Export -> SVG/PNG/PDF file... y seleccionamos si queremos un archivo vectorial (PDF o SVG) o imagen raster, de píxeles (PNG). En opciones, podemos cambiar el tamaño de la imagen y el margen.



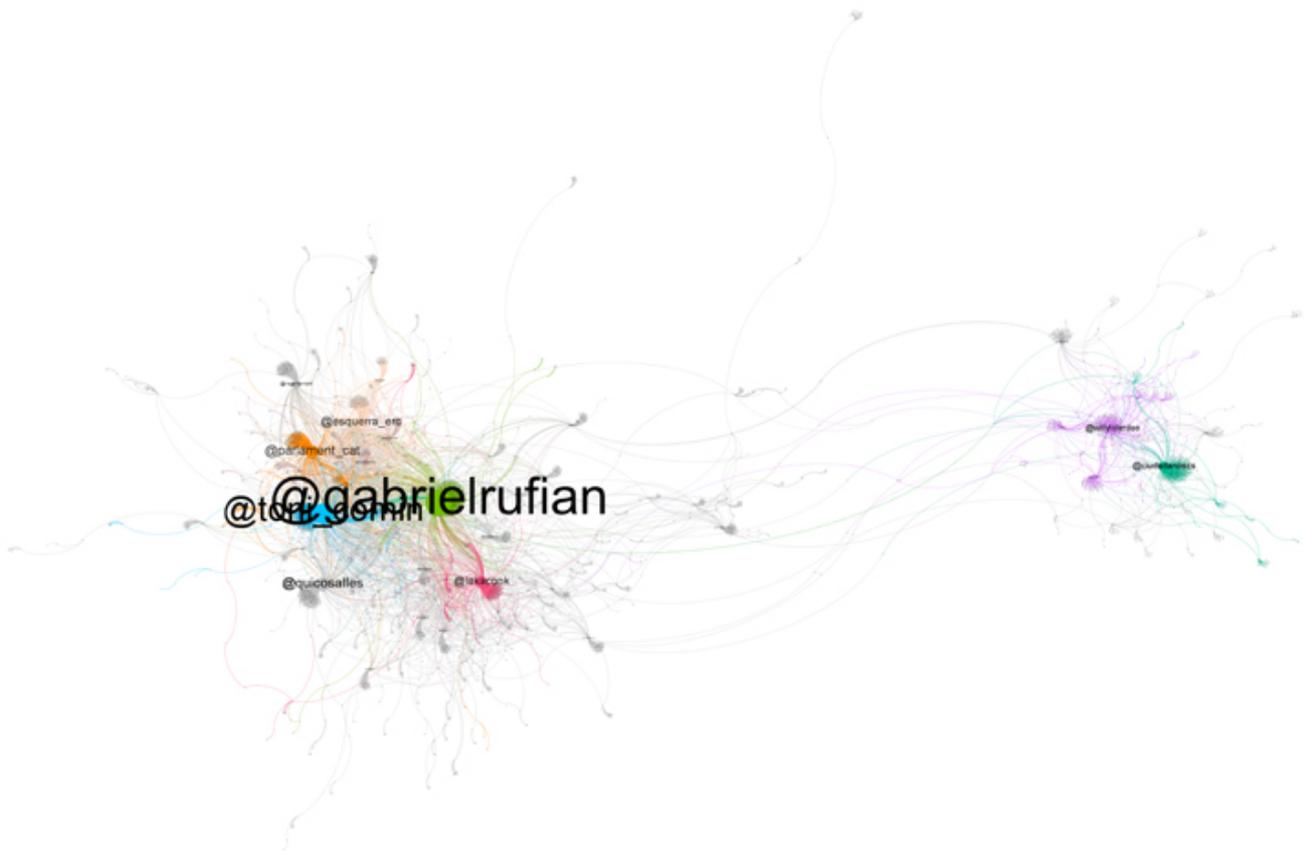
El proceso de exportación tendremos que realizarlo repetidas veces hasta que ajustemos el tamaño de los nodos y las etiquetas que deseemos.

### Resultados

Grafo de la búsqueda «Felipe González»; se representan usuarios en esos tuits como nodos y las relaciones representadas como vértices son los RT entre ellos. Tuits descargados con search query de 2018-01-09 00:49:41 a 2018-01-16 12:43:37 un total de 17.597 tuits.



Grafo del *hashtag* #parlament, se representan usuarios en esos tuits como nodos y la relaciones representadas como vértices son los RT entre ellos. Tuits descargados con t-hoarder con search query de 2018-01-17 13:20:53 a 2018-01-17 17:20:34 y con *streaming* de 2018-01-17 18:20:50.308562 a 2018-01-17 19:11:03, un total de 14.779 tuits.

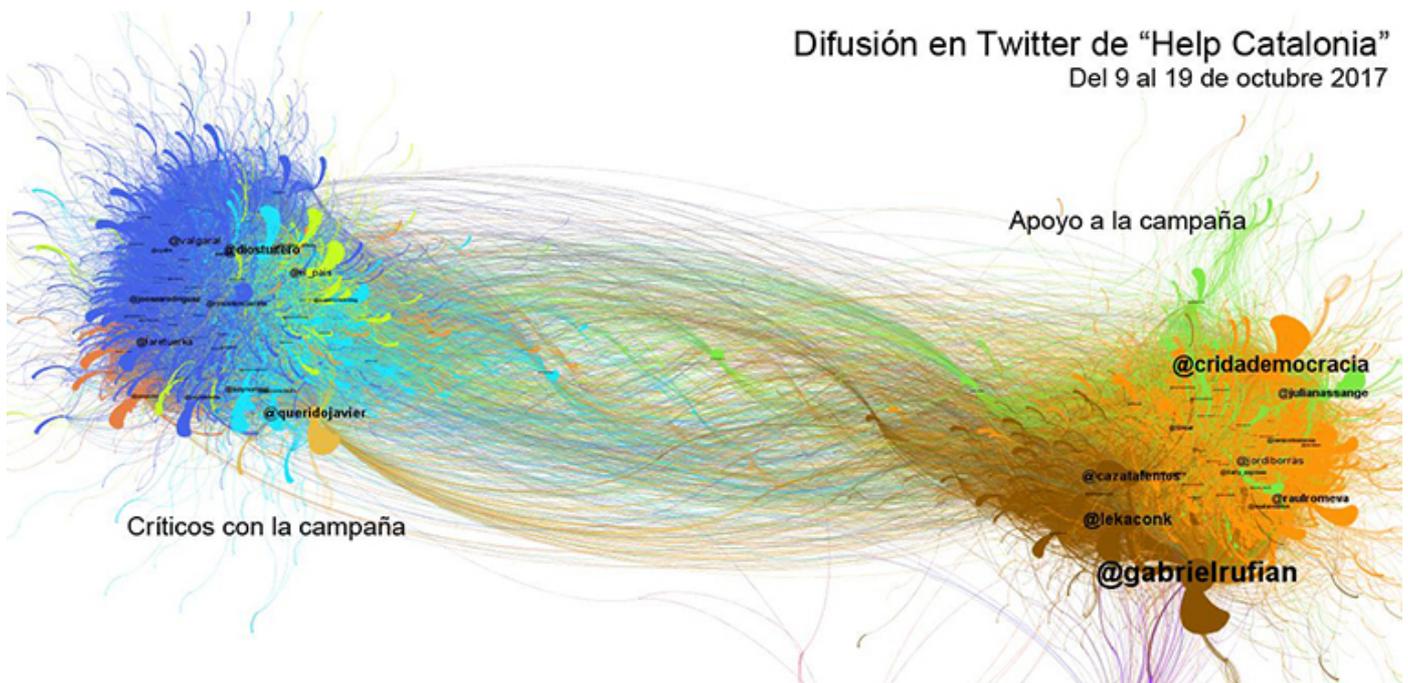


#### e. Cómo presento la información para hacer un relato inteligible

Probablemente, el gráfico resultante no sea lo suficientemente explicativo para responder a la pregunta inicial. Un grafo sin una leyenda que explique los colores, sin un texto que lo acompañe, sin un titular o destacado que lo explique no vale como respuesta. Cada vez se usan más los grafos en la prensa, pero es necesario aclarar qué es lo que representan los puntos y líneas que estamos viendo. Sirva esta aclaración como ejemplo:

“ «Los grafos se han generado tomando los retuits como elemento de conexión de los usuarios. Cada nodo es un usuario y cada arco es una conexión entre retuiteador y retuiteado. El tamaño de los nodos es directamente proporcional al número de RT recibidos».

Mariluz Congosto (2016).



Ejemplo de grafo con título, descripción de usuarios y periodo. Difusión de la campaña «Help Catalonia» en Twitter  
 Fuente: <http://www.barriblog.com/2017/10/hilo-difusion-la-campana-help-catalonia-twitter/>.

Para ello, tenemos que contar y acompañar al grafo o grafos resultantes de todos estos elementos y añadir también la documentación de cómo se han obtenidos los datos y realizado el/los gráficos. Si has capturado datos por un largo periodo de tiempo, quizás necesites realizar diferentes grafos según las fases de propagación del *hashtag*.

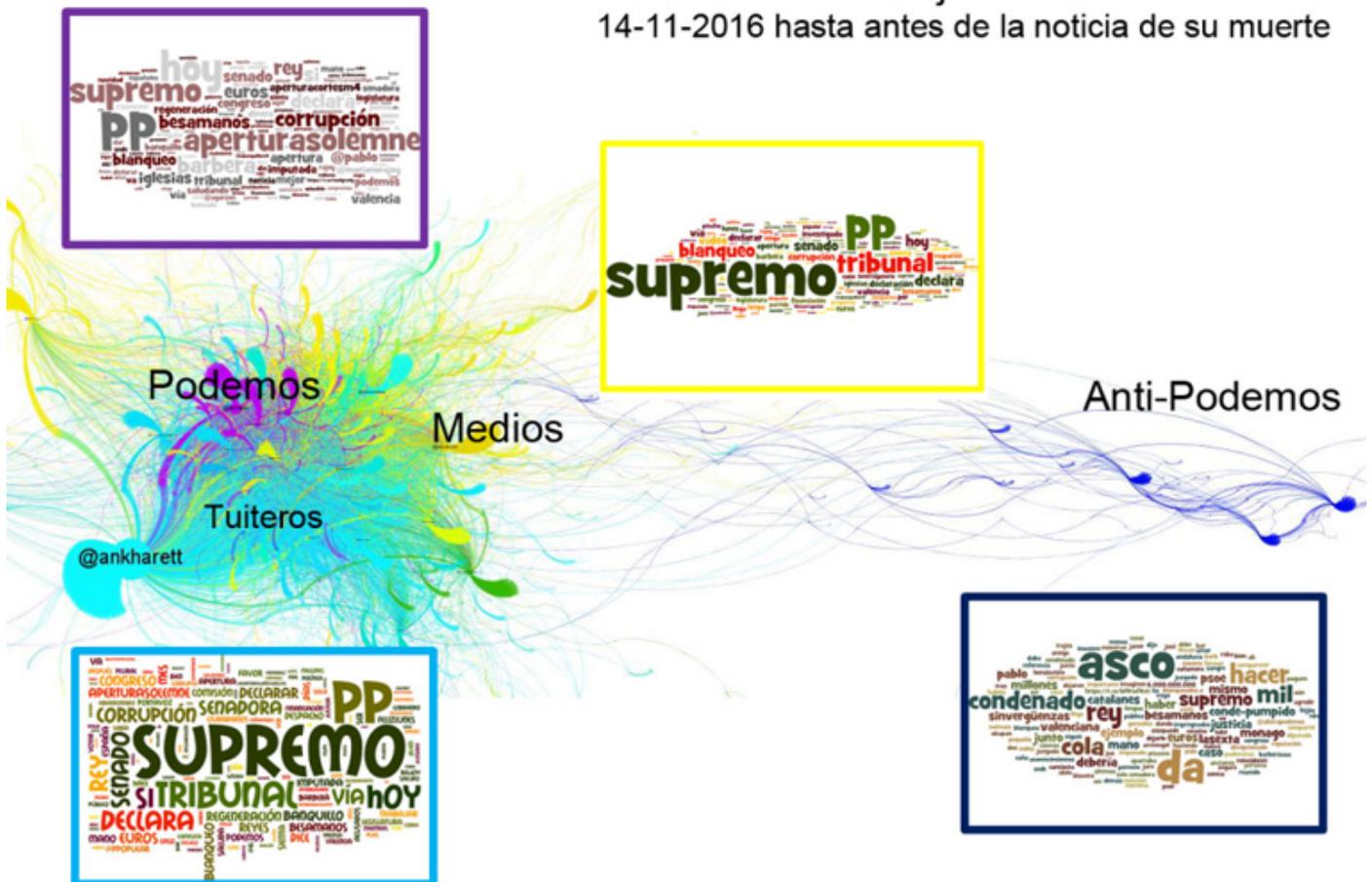
Puedes usar los filtros que ofrece Gephi para mostrar las diferentes categorías (*modularity*) que se usan para colorear los gráficos para analizar la composición de esas redes y de qué habla cada una de ellas. Este tipo de filtros permite dejar solo activos los usuarios de una o más categorías.

Paso a paso:

- Elige título.
- Escribe párrafo explicativo/introductorio.
- Añade gráficos que lo acompañen.

Para cada uno de los grupos detectados (*modularity*), podemos calcular las palabras más usadas (bigramas y trigramas con [databasic.io](http://databasic.io)) y/o hacer un *ranking* o una nube con ellas. T-hoarder ofrece también la posibilidad de calcular las palabras más populares con la opción 7. *Processing tweets (Entities)*. A partir de lista de tuits. Las palabras asociadas a cada grupo pueden ayudar a entender el grafo:

### Difusión de mensajes sobre Rita Barberá 14-11-2016 hasta antes de la noticia de su muerte



Nube de palabras asociadas a cada categoría de usuarios. Después de la muerte de Rita Barberá  
Fuente: <http://www.barriblog.com/2016/11/despues-la-muerte-rita-barbera-twitter/>.

Ejemplos de análisis con esta metodología:

- [Así se viralizó Tabarnia en redes sociales](#). Mariluz Congosto, 2018.
- Difusión de la [campana «Help Catalonia»](#) en Twitter.

**Nota**

Puedes descargar el .gdf para replicar este ejercicio:

- [helpcatalonia\\_all\\_RT.gdf](#).
- [helpcatalonia\\_most\\_RT.gdf](#).

## 6.1.4.2. ¿Cómo se relacionan los seguidores de una cuenta de Twitter?

### a. Qué análisis o visualización responde a la pregunta

Un grafo.

### b. Qué datos necesito y dónde los encuentro

Crea un listado con los usuarios que han publicado un determinado *hashtag* (pregunta anterior): ¿cómo se obtiene esa lista? `T-hoarder_kit` te permite, con la opción `7.Processing tweets` y usando las opciones `entity` (que extrae los usuarios más retuiteados, los más mencionados y los más activos), obtener un fichero con la lista de usuarios. Hay que limpiar el archivo generado ya que tiene asociadas otras variables, y lo que necesitamos es un archivo con cada nombre de usuario en una línea. Lo siguiente es obtener los perfiles (profiles) de esos usuarios: `2.Get users information` escribe opción `profile`. Esto genera un archivo de nombre `xxxxx_profiles.txt` que contiene todos los perfiles.

A partir de esa lista de perfiles se pueden calcular sus relaciones (*follower-following*). Atención: para obtener los *followers-following* de cada usuario, la API los proporciona a una velocidad de 60 usuarios cada hora. Por lo tanto, para analizar la relación entre 1.000 usuarios, nos llevaría 16,6 horas. Con el archivo .gdf generado pasamos al siguiente paso. Por tanto, antes de empezar a ver las relaciones entre un grupo de usuarios de un determinado *hashtag*, podemos seleccionar los usuarios según un criterio (actividad o relevancia) para hacer listas de usuarios más pequeñas y, por ejemplo, estudiar cómo están relacionados los 100 usuarios más mencionados o los 100 usuarios más activos.

Tienes que hacer un grafo de relaciones declaradas (*following* y *followers* entre los usuarios del *hashtag*).

### c. Qué herramientas uso para preparar los datos

No aplicable. El archivo .gdf generado ya es válido para el siguiente paso.

### d. Qué herramientas uso para producir el análisis o la visualización

Para generar el grafo seguimos el procedimiento descrito en el ejercicio anterior, con la diferencia de que ahora tendremos muchos menos nodos que en un grafo de RT. El layout «forceatlas» funciona mejor para menos nodos que el Forceatlas2.

### e. Cómo presento la información para hacer un relato inteligible

Lo mismo del ejercicio anterior.

#### Ejemplo de análisis con metodología similar

- [Digital Humanities on Twitter, a small-world?](#) Martin Grandjean, 2015.
- [Gregorio Morán, una columna vacía](#). Mariluz Congosto, 2017.

Puedes descargar el .gdf con las relaciones entre usuarios para replicar el ejercicio: [gregoriomoran\\_users\\_profiles.gdf](#).

## 6. Tres ejercicios de periodismo de datos

### 6.1. Twitter: captura de datos y análisis de redes

#### 6.1.5. Información complementaria

- Animación en un mapa torque con Carto: [ejemplo](#).
- Obtención de tuits y grafo con <http://flocker.outliers.es/>.
- Tutoriales de Gephi:
  - <https://www.slideshare.net/gephi/gephi-quick-start>.
  - <https://www.slideshare.net/gephi/gephi-tutorial-visualization>.
- Artículo sobre T-warder: M. Congosto, P. Basanta-Val y L. Sánchez-Fernández, (2017). «[T-Hoarder: A framework to process Twitter data streams](#)». *Journal of Network and Computer Applications* (núm. 83, págs. 28-39).

## 6. Tres ejercicios de periodismo de datos

### 6.2. Efecto Airbnb: análisis geoespacial

#### 6.2.1. Introducción

Se está hablando mucho últimamente sobre la presencia de Airbnb en tu ciudad. Hay organizaciones ciudadanas que se empiezan a mover porque en sus barrios hay cada vez más turistas, y el ayuntamiento está pensando en regular la situación. El periódico digital para el que trabajas ha decidido arrojar un poco de luz con un análisis gráfico de los datos disponibles. ¿Cuál es el impacto de Airbnb en las diferentes zonas de la ciudad? ¿Es similar al de otras ciudades?

#### **Nota**

Este ejercicio está basado en la serie de talleres de análisis de datos del proyecto [Efecto Airbnb](#) de Montera34 y en la serie de [manuales sobre análisis de datos de Airbnb](#).

Cosas que puedes aprender: dónde y cómo conseguir datos de Airbnb y otras plataformas turísticas, limpiar y preparar un set de datos con OpenRefine, producir una cartografía digital con el servicio Carto.

## 6. Tres ejercicios de periodismo de datos

### 6.2. Efecto Airbnb: análisis geoespacial

#### 6.2.2. Contexto

La aparición en internet de plataformas de alquiler de alojamientos *peer to peer* está provocando rápidos cambios de uso en los centros urbanos de mayor atractivo turístico. Estas plataformas, con Airbnb a la cabeza, han sido acusadas de contribuir al alza en los precios inmobiliarios y de ser un elemento disruptivo en las comunidades locales. Por otro lado, se defiende que la supuesta tendencia gentrificadora de la web es exagerada.

## 6. Tres ejercicios de periodismo de datos

### 6.2. Efecto Airbnb: análisis geoespacial

#### 6.2.3. Conceptos básicos o previos

**Qué es una API.** Una API, abreviada como API del inglés: *application programming interface*, es la forma estructurada de acceder e interactuar con una aplicación o web.

**Terminología de los datos de Airbnb.** Airbnb utiliza una terminología propia en su estructura de datos. Es conveniente conocerla cuando exploramos sus datos: los anuncios se llaman *listings*; los usuarios son *hosts*.

**Qué es Carto.com.** Es un servicio web disponible en <https://carto.com> que permite producir mapas y otras visualizaciones a partir de un conjunto de datos. Carto.com es una instalación del software Cartodb, que es de código abierto y está disponible libremente: <https://github.com/CartoDB/cartodb>.

**Qué es OpenRefine.** Es un software de código abierto que permite limpiar y formatear un conjunto de datos usando una interfaz gráfica. Se puede descargar en la web <http://openrefine.org/>.

**Qué es CSV.** Comma Separated Values (CSV) es un formato de texto plano en el que cada línea equivale a un registro y los distintos valores de un registro se separan por el mismo carácter; aunque contra todo pronóstico, no tiene por qué ser una coma, sino cualquier carácter.

**Qué es geoJSON.** Es la variación del formato JSON que surge al añadirle información de posición geográfica, coordenadas de posición.

**Qué es un mapa de coropletas o coroplético.** Citando a [Wikipedia](#):

“«Es un mapa temático en el que las regiones se colorean de un motivo que muestra una medida estadística, como puede ser la densidad de población o el ingreso por habitante. Este tipo de mapa facilita la comparación de una medida estadística de una región con la de otra».

**Qué es un scraper.** Es un pequeño programa que permite recopilar automáticamente los datos contenidos en un sitio web.

## 6. Tres ejercicios de periodismo de datos

### 6.2. Efecto Airbnb: análisis geoespacial

#### 6.2.4. Preguntas a resolver

Las preguntas del ejercicio se han respondido para la ciudad de Madrid, pero el procedimiento es válido para cualquier otra ciudad. Te animamos a que apliques este ejercicio en tu ciudad.

##### 6.2.4.1. ¿Cuál es el impacto de Airbnb en las diferentes zonas de la ciudad?

Vamos a considerar como zonas los distritos de Madrid.

###### a. Qué análisis o visualización responde a la pregunta

Dos análisis complementarios:

- Un mapa en el que se vean los contornos de los barrios y los pisos de Airbnb representados por puntos. Las áreas pueden tener un código de color que indique la cantidad de pisos de Airbnb en esa zona o la proporción de alojamientos anunciados de Airbnb respecto del total de viviendas. Esta segunda opción, la proporcionalidad, es más indicada para un mapa de coropletas.
- Un diagrama de barras de la cantidad de pisos Airbnb por distrito.

###### b. Qué datos necesito y dónde los encuentro

Para calcular el número de alojamientos de Airbnb por distrito, necesitarás el **listado de alojamientos de Airbnb** geolocalizados (con latitud y longitud) y **los contornos de distritos**. Para calcular la proporción de alojamientos de Airbnb respecto del total de viviendas necesito, además, el **número de viviendas de cada barrio**. Vamos a ver cómo conseguimos cada uno de estos datos:

###### Listado de alojamientos de Airbnb

Airbnb no tiene una API, así que no resulta sencillo interactuar o descargarse los datos de la plataforma. La manera de hacerlo es programar un scraper que recorra la web de Airbnb descargando los datos que nos interesan. Vamos a dejar esto fuera de este ejercicio, aunque sería muy interesante hacerlo.

La alternativa es ponerse en contacto con alguien que ya haya programado un scraper para la web de Airbnb y que nos facilite los datos. Aquí tienes varios sitios web que proporcionan datos de Airbnb, y os animamos a que busquéis vuestras propias fuentes:

- [Inside Airbnb](#). Inside Airbnb es una página alimentada por Murray Cox, quien ha programado su propio scraper para obtener datos de Airbnb. Dispone de datos de decenas de ciudades. Si la ciudad que buscas no está listada o si los datos no están actualizados, puedes contactar directamente con Murray.
- [Data Hippo](#). Data Hippo es una plataforma para alojar datos de plataformas de alquiler vacacional de ciudades de España y Portugal. Dispone de datos de Airbnb, Homeaway y Housetrup por regiones, provincias y ciudades. En las principales ciudades turísticas, dispone de datos por barrio o distrito. Puedes entrar en contacto con la gente detrás de Data Hippo en su [cuenta de twitter](#).

La elección de la fuente de los datos es crucial: de esta elección va a depender la estructura de datos, que para cada fuente es diferente; depende de lo recientes y completos que sean los datos. Conviene consultar la fecha en la que han sido recolectados y los campos que contiene el conjunto de datos a descargar antes de decidir.

###### Nota

En este ejercicio vamos a usar los datos de Data Hippo más recientes (octubre de 2017) que los de Inside Airbnb y que contienen la información necesaria para hacer el estudio por distritos.

Los datos de Madrid se pueden consultar aquí: <http://datahippo.org/en/region/599232178a46554f807aec1b/>.

Y se pueden descargar en formato CSV:

[http://datahippo.org/media/regions/ea554b7f-0516-4d1c-91e6-883567800e08/599232178a46554f807aec1b\\_airbnb.csv](http://datahippo.org/media/regions/ea554b7f-0516-4d1c-91e6-883567800e08/599232178a46554f807aec1b_airbnb.csv)

## Contornos de los distritos

Los contornos de los distritos de Madrid los podemos obtener también de Data Hippo en [formato geoJSON](#).

El archivo de Data Hippo contiene también el total de alojamientos Airbnb para cada distrito. Veremos en la siguiente sección que, estrictamente, para realizar el ejercicio con este archivo tenemos suficiente: no hace falta el anterior. Aun así, recogemos el procedimiento completo para el caso de no tener disponibles los datos en Data Hippo.

En Data Hippo hay dos archivos geoJSON para Madrid: `geojson_region.json` contiene el contorno del municipio; `geojson_subregions.json` contiene los contornos de los distritos. Puedes acceder a esos archivos cuando navegues a la sección [municipio de Madrid](#).

Siempre es recomendable obtener los datos de la fuente institucional (la fuente primigenia), que en este caso es el Ayuntamiento de Madrid. Los contornos de los distritos están también disponibles en el [portal web del Ayuntamiento de Madrid](#). Vienen en un shapefile (conjunto de ficheros descargable en formato comprimido .zip); tendrás que transformar su sistema de proyección ETRS89 al CRS EPSG:4326 – WGS 84. Antes Carto.com no aceptaba otros CRS distintos del estándar EPSG:4326 – WGS 84; ahora sí que lo acepta. Desde QGIS puedes abrir el archivo que te descargas como capa vectorial y salvar como nuevo archivo con el CRS que elijas (en este [manual](#) puedes ver una descripción más detallada).

## Viviendas por distrito

El número total de viviendas por distrito y barrio lo podemos obtener del [portal web de datos abiertos del Ayuntamiento de Madrid](#).

### c. Qué herramientas uso para preparar los datos

#### Número de pisos Airbnb por distrito

Los datos de Airbnb que se pueden conseguir en Data Hippo ya están segmentados por distrito, así que ya tenemos en el dataset el dato que necesitamos.

Si no lo tuviéramos, el procedimiento para obtenerlo es el siguiente: podemos calcular el número de alojamientos por distrito contando cuántos puntos hay dentro del contorno de cada distrito. Esta es la operación que han hecho con los datos en Data Hippo.

La ubicación de los alojamientos en Airbnb es aproximada. Según la propia web: «Location information for listings are anonymized by Airbnb. In practice, this means the location for a listing on the map, or in the data will be from 0-450 feet (150 metres) of the actual address. Listings in the same building are anonymized by Airbnb individually, and therefore may appear “scattered” in the area surrounding the actual address».

Por tanto, si optamos por el método de contar puntos en contornos, debemos tener en cuenta que algunos alojamientos están ubicados fuera del distrito que les corresponde, con lo cual los números incluirán una imprecisión.

Bola extra: Una manera de corregir este error es conseguir datos de Airbnb que incluyan el código postal para cada alojamiento (Data Hippo no lo proporciona, Inside Airbnb sí). Entonces se podría hacer una tabla de equivalencia entre distritos y códigos postales. Con dicha tabla podríamos recorrer el listado de alojamientos y añadir una columna con el distrito. Esta unión de tablas se puede realizar con QGIS o con una consulta SQL.

## Contorno de distritos

Los contornos que nos hemos descargado de Data Hippo están listos para ser usados.

## Viviendas por distrito

El archivo que nos hemos descargado del portal de datos del Ayuntamiento de Madrid es una hoja de cálculo con una pestaña para cada distrito y muchos datos para cada distrito y cada barrio. El dato de cantidad de viviendas para cada distrito hay que ir a buscarlo en cada una de las pestañas en la celda 239E. Se podría automatizar la obtención de ese dato para cada pestaña, pero como son pocos distritos no merece la pena.

Pegamos aquí la tabla con el total de vivienda para cada distrito y así te ahorramos este aburrido paso:

---

Arganzuela	76260
------------	-------

---

Barajas	20370
Carabanchel	115270
Centro	86795
Chamartín	72120
Chamberí	79780
Ciudad Lineal	101740
Fuencarral- El Pardo	99260
Hortaleza	74930
Latina	115500
Moncloa-Aravaca	55125
Moratalaz	44405
Puente de Vallecas	106585
Retiro	56595
Salamanca	83575
San Blas-Canillejas	67450
Tetuán	80160
Usera	59285
Vicálvaro	30305
Villa de Vallecas	43520
Villaverde	61920

A continuación tienes los mismos datos pero sin tildes y en formato CSV, ya preparada para ser usada:

```

distrito,total
Arganzuela,76260
Barajas,20370
Carabanchel,115270
Centro,86795
Chamartin,72120
Chamberi,79780
Ciudad Lineal,101740
Fuencarral- El Pardo,99260
Hortaleza,74930
Latina,115500
Moncloa-Aravaca,55125
Moratalaz,44405
Puente de Vallecas,106585
Retiro,56595
Salamanca,83575
San Blas -Canillejas,67450
Tetuan,80160
Usera,59285
Vicalvaro,30305
Villa de Vallecas,43520
Villaverde,61920

```

#### **d. Qué herramientas uso para producir la visualización**

El mapa de coropletas de los distritos se puede realizar con diferentes herramientas. Una posibilidad es usar QGIS, que nos permite hacer el cálculo de puntos (alojamientos) dentro de un polígono (distrito), y producir un paquete incluyendo todo lo necesario (archivos con los datos, librerías JavaScript y archivos HTML) para publicar en web. Otra opción es usar el servicio Carto.com, que nos permite hacer lo mismo que QGIS y luego generar un código para incrustar el mapa en otro sitio web.

### Nota

Dado que estás produciendo una visualización para incluir en el sitio web del medio digital en el que trabajas, te recomendamos que utilices Carto.com, ya que con el código que genera podrás incrustar el mapa sin problema como si fuese un artículo normal.

## Número de puntos por distrito

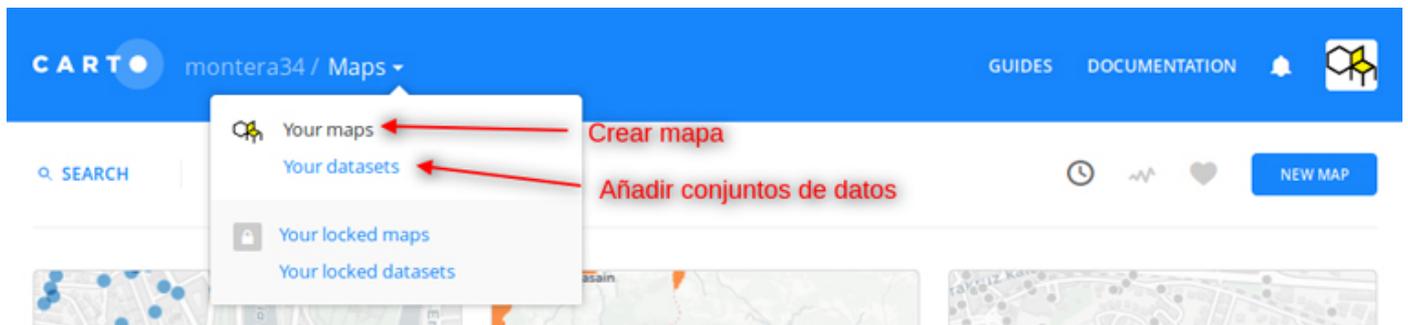
Como comentábamos en la sección anterior, el dataset de Data Hippo ya incluye el número de alojamientos Airbnb por distrito, en la columna `airbnb_aptms`.

En el caso de que no lo tuviéramos, hay que seguir los pasos que se muestran a continuación para obtenerlo con Carto.com.

### Nota

Carto.com se ha convertido mientras elaborábamos estos ejercicios en un servicio de pago. Para poder usarlo puedes solicitar como estudiante acceso ([Student and Educator Accounts](#)). También hemos desarrollado el mismo ejercicio con QGIS disponible en el manual [¿Cómo se concentra la oferta por barrios de Airbnb?](#) de la wiki de Monteras34.

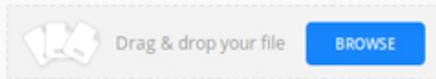
- 1) Inicia sesión con tu cuenta en Carto.com. Si no tienes cuenta, date de alta.
- 2) Añade como nuevos datasets el archivo de los alojamientos y el archivo con los contornos de los distritos.



No hace falta que te descargues los conjuntos de datos de Data Hippo; con proporcionar a las URL a los archivos, Carto.com creará los datasets:

## Upload a file or a URL

Paste a URL or select a file such as CSV, XLS, ZIP, KML, GPX, [see all formats](#).



or

**1. Pegar URL de Data Hippo**

3554f807aec1b\_apartments\_airbnb.csv [SUBMIT](#)

**2. Pinchar en Submit**

**3. Pinchar en Connect Dataset**

Let CARTO automatically guess data types and content on import.



[CONNECT DATASET](#)

**3) Crea un mapa y añade como capas los archivos que acabas de subir:**

  
**Add datasets**  
Select your datasets

SEARCH | CONNECT DATASET | YOUR DATASETS | DATA LIBRARY | CREATE EMPTY MAP

	<b>table_599232178a46554f807aec1b_geojson_subregions</b> Contornos de los distritos de Madrid en geoJSON	<b>PUBLIC</b> ❤️ 21 Rows 240 kB a minute ago No tags
	<b>table_599232178a46554f807aec1b_apartments_airbnb</b> Listado de alojamientos Airbnb en el municipio de Madrid (octubre 2017)	<b>PUBLIC</b> ❤️ 18.5K Rows 4 MB 3 minutes ago No tags

2 datasets selected

**2** CREATE MAP

**1** seleccionar los dos datasets

**4)** Sitúate en la capa de los contornos y pincha en *Analysis*. Añade un nuevo análisis. Selecciona el modo *Intersect second layer*. *Intersect with a second layer and calculate aggregations on the fly*.



En el menú desplegado debes seleccionar la capa de los contornos como `Aggregate intersection`, y la de los alojamientos como `Intersect Layer`.

En `Operation` debe estar seleccionada la opción `COUNT`.

Pincha en `Done`. Eso generará una nueva columna en la base de datos llamada `count_values`.

**5)** Entra en el modo tabla con el conmutador de la esquina inferior derecha del mapa, y cambia el nombre de la columna `count_values` a `alojamientos_airbnb`.

**6)** En este momento, ya podemos producir un primer mapa de coropletas que nos coloree los distritos en función del número de alojamientos Airbnb. Para ello, vamos a la capa de contornos y en la pestaña `Style` editamos el estilo `Color` y asignamos a la opción `By value` como valor la columna que contiene el count de alojamientos, `alojamientos_airbnb`.

### Proporción de alojamientos Airbnb por distrito

Para hacer este paso necesitamos unir información que está en diferentes archivos. En concreto, necesitamos hacer la siguiente ecuación para cada distrito:

```
alojamientos de airbnb del distrito / viviendas totales del distrito
```

Para poder relacionar fácilmente estos dos valores para cada distrito, tenemos que incluir los totales de vivienda en el dataset de los contornos. Una opción es hacerlo manualmente, ya que son pocos valores. Recogemos aquí también la opción semiautomática para casos en los que haya gran cantidad de datos: vamos a hacerlo con una consulta SQL en `Carto.com`.

**1)** Creamos un nuevo dataset con el listado de viviendas totales por distrito. Anotamos el nombre del dataset y cambiamos el nombre de la columna que contiene los distritos a `distrito`, y de la columna que contiene el número total de viviendas a `total`.

**2)** Creamos dos nuevas columnas en el dataset de contornos que llamamos `total_viviendas` y `percent_turisticas`. Seleccionamos como tipo de dato `number` para ambas.

**3)** Nos tenemos que asegurar de que el nombre de los distritos está escrito de igual manera en ambos datasets: mismos caracteres acentuados, mismos espacios, mismas mayúsculas.

**4)** Volvemos al mapa, entramos en la capa de contornos y en la pestaña `Data`. Allí, abajo cambiamos a la consola SQL con el conmutador situado en la parte inferior de la barra lateral izquierda.

**5)** Escribimos la siguiente consulta SQL:

```
UPDATE nombre_dataset_contornos SET total_viviendas =
nombre_dataset_viviendas_distrito.total FROM nombre_dataset_viviendas_distrito WHERE
nombre_dataset_contornos.name = nombre_dataset_viviendas_distrito.districto
```

que, traducido a lenguaje común, significa: actualiza (UPDATE) el dataset de contornos y pon (SET) el número de viviendas por distrito de la tabla de viviendas por distrito (el nombre de la columna es total) (FROM) en las filas donde coincida la columna de distrito del dataset de contornos con la columna de distrito del dataset de total de viviendas (WHERE). Por esto era importante que fueran consistentes los nombres de los distritos.

De este modo tendremos los datos de viviendas por distrito en la tabla de contornos de distritos.

Ya solo nos queda calcular el ratio de alojamientos de Airbnb por cada 100 viviendas. Podemos hacerlo con otra consulta SQL:

```
UPDATE nombre_dataset_contornos SET percent_turisticas = airbnb_aptms * 100 /
total_viviendas
```

Para redondear los decimales y evitar resultados tipo 8,21439204893, hacemos una consulta más restringiendo a dos decimales:

```
UPDATE nombre_dataset_contornos SET percent_turisticas = round( CAST(percent_turisticas as
numeric), 2 )
```

## Visualizaciones

**1) Mapa de coropletas.** Ya solo nos queda hacer el mapa de coropletas utilizando la variable `percent_turistica`, como hemos descrito en el apartado anterior. Fácil.

Para hacer el mapa público solo queda pinchar en el botón Publish. Y luego copiar el código para incrustar el mapa en cualquier web.

**2) Gráfico de barras.** Podemos descargar los datos que hemos calculado en Carto.com y utilizarlos en cualquier programa de hojas de cálculo para generar un gráfico de barras.

Para descargar, ve a modo Data view (tabla) de la capa de contornos y en el menú `Edit > Export Layer` se puede descargar la capa como CSV.

### Ejemplo

Tienes un ejemplo de un [gráfico de barras](#) doble con el ratio de alojamientos de Airbnb por cada 100 viviendas y la cantidad de alojamientos por barrios de diferentes ciudades. Está desarrollado con la librería de JavaScript D3js en el proyecto Efecto Airbnb de Montera34. El código fuente está disponible.

## e. Cómo presento la información para hacer un relato inteligible

Una vez que tenemos los datos de los ratios por distrito, es posible comparar unos y otros valores. Con la representación en el mapa de coropletas podemos comparar la distribución espacial (es más difícil apreciar la diferencia de valores asociados a colores) y en el gráfico de barras podemos comparar la presencia de Airbnb en cada barrio. También puedes publicar una tabla con los distritos con más presencia de Airbnb.

Otras cosas que puedes hacer, para completar el reportaje, es analizar el tipo de alojamiento que se ofrece en cada uno de los distritos (`private_room`, `entire_home` dentro de la variable `room_type`) o el total de plazas (variable `capacity`).

El proceso para calcular el ratio de alojamientos de Airbnb por cada 100 viviendas es análogo. Busca los contornos y datos de cada y repite el proceso.

## 6. Tres ejercicios de periodismo de datos

### 6.2. Efecto Airbnb: análisis geoespacial

#### 6.2.5. Información complementaria

- Artículos de prensa que estudian el impacto de Airbnb:
  - [Dormir en el limbo: radiografía de Airbnb](#). Reportaje de *El Español*.
  - [Estudio sobre Mallorca con datos de Inside Airbnb](#).
  - [Airbnb vs. Berlín](#). Informe gráfico de la situación del alquiler de pisos Airbnb en Berlín.
- Otros: [Plan Especial Urbanístico de Alojamiento Turístico \(PEUAT\)](#) de Barcelona. Explicación gráfica de las medidas de zonificación puestas en marcha en Barcelona para controlar los efectos de los pisos turísticos en los diferentes barrios.

## 6. Tres ejercicios de periodismo de datos

### 6.3. Educación pública-privada: abriendo datos abiertos

#### 6.3.1. Introducción

Ha surgido una urgencia para el especial sobre educación. Tu serie de reportajes tendrá que esperar otra vez. Te piden tener para dentro de dos días un artículo para la edición en Euskadi sobre las diferencias entre la educación pública y privada. Los datos vienen de un informe en... PDF, así que te toca pasar datos a limpio para poder hacer algo con ellos. No es lo mejor, pero por lo menos tienes los datos.

#### **Nota**

Este ejercicio está basado en el [análisis de la segregación escolar en Euskadi](#).

Aprendizajes: rescatar datos de tablas incluidas en PDF, unir datos espaciales y tablas de datos, hacer mapas de coropletas, generar mapas interactivos en D3.

## 6. Tres ejercicios de periodismo de datos

### 6.3. Educación pública-privada: abriendo datos abiertos

#### 6.3.2. Contexto

En Euskadi hay una de las mayores redes educativas privado-concertadas de toda Europa, con una cuota del 48,9 % del alumnado. Encabeza también el listado de las comunidades autónomas en España en cuanto a porcentaje de alumnado en red concertada, con un 48,1 %. Esta realidad en la que el alumnado está repartido casi a partes iguales en las redes públicas y concertadas (apenas hay centros privados en Euskadi) ha dado lugar a fuertes diferencias en el tipo de alumnos que atienden a una y otra red. Se habla de «segregación escolar», en la que determinados sectores sociales no se encuentran homogéneamente distribuidos en las redes. Teniendo en cuenta que la red concertada está enteramente subvencionada por el Estado, debería tener los mismos criterios de admisión que la pública. El objetivo de este ejercicio es verificar con datos si existe y en qué medida segregación escolar en Euskadi.

Los datos disponibles provienen del informe *La educación en Euskadi 2013-2015* del Consejo Escolar de Euskadi: un documento en PDF que incluye datos en tablas e incluye gráficos de líneas, barras y mapas. Como es tristemente habitual, los datos no están en formatos reutilizables, ni se encuentran en ningún catálogo de datos abiertos. Sin embargo, como las tablas están disponibles, puedes acceder a los datos y convertirlos en .csv con una hoja de cálculo a base de copiar y pegar. El nivel de desagregación de los datos es el de zona escolar (no están disponibles los datos de cada centro escolar).

Cada Administración, local, nacional, etc. tiene datos que permiten hacer un estudio similar al propuesto en este ejemplo de *La educación en Euskadi*.

## 6. Tres ejercicios de periodismo de datos

### 6.3. Educación pública-privada: abriendo datos abiertos

#### 6.3.3. Conceptos básicos o previos

**Que es GIS.** De sus siglas en inglés sistema de información geográfica (*geographic information system*), o SIG en castellano, se define como el conjunto de herramientas que permiten capturar, almacenar, analizar y presentar datos geográficos.

**Qué es QGIS.** Un software libre de GIS que permite visualizar, editar y analizar información geoespacial.

**Qué es GeoJSON.** Es un formato estándar abierto (un tipo de archivo con terminación .geojson) diseñado para representar elementos geográficos sencillos, junto con sus atributos no espaciales, basado en JavaScript Object Notation (JSON). Puede estar compuesto de puntos, líneas y polígonos.

```
{ "type": "FeatureCollection",
  "features": [
    { "type": "Feature",
      "geometry": {
        "type": "Point",
        "coordinates": [102.0, 0.5]
      },
      "properties": {
        "prop0": "value0"
      }
    },
    { "type": "Feature",
      "geometry": {
        "type": "LineString",
        "coordinates": [
          [102.0, 0.0], [103.0, 1.0], [104.0, 0.0]
        ]
      },
      "properties": {
```

**Que es TopoJSON.** Un archivo de tipo topoJSON (con terminación .json) es un tipo de formato de GeoJSON en el que las geometrías están compuestas por segmentos llamados arcos (*arcs*). Si dos o más contornos comparten una línea, piensa en la frontera de dos países, ese segmento compartido queda solamente definido una vez, reduciendo la redundancia y el tamaño del archivo. Este tipo de definición resulta en que definen una topología y no simplemente elementos por separado, lo que permite, entre otras cosas, hacer cartogramas.

## 6. Tres ejercicios de periodismo de datos

### 6.3. Educación pública-privada: abriendo datos abiertos

#### 6.3.4. Preguntas a resolver

##### 6.3.4.1. ¿Cómo se reparte el alumnado entre la red pública y la privada en Euskadi? ¿Existe segregación escolar?

###### a. Qué análisis o visualización responde a la pregunta

Queremos estudiar qué tipo de alumnado pertenece a una y otra red y para investigar si existe segregación escolar. La exploración de los datos llevará a preguntarse, seguramente, el porqué de las diferencias tan grandes en los indicadores entre las redes pública y privada: ¿son homogéneas en el territorio esas diferencias? Puedes explorar si hay diferencias por provincia o por nivel socioeconómico.

Vamos a realizar varios análisis y visualizaciones:

- Una serie de mapas de coropletas en el que se vean los contornos de las zonas escolares, donde el color de cada zona representa el porcentaje del alumnado que en cada red (pública o privada) tiene una beca material escolar o una beca de comedor. Estos indicadores son la mejor forma que tenemos para estudiar el perfil socioeconómico de los alumnos en cada zona escolar, ya que no se disponen de los datos desagregados del ISEC, índice socioeconómico y cultural realizado a partir de las respuestas recogidas en pruebas de evaluación. Cuanto más alumnado becario haya, indicará que hay menor nivel socioeconómico en esa zona escolar y en esa red escolar. También vamos a usar el porcentaje de alumnado extranjero, que suele ser indicador del número de inmigrantes, muchas veces con menor nivel socioeconómico (de América del Sur y de Asia proviene el 63 % del alumnado extranjero).
- Gráficos que muestren las diferencias entre los indicadores en las zonas escolares.

###### b. Qué datos necesito y dónde los encuentro

Para poder representar en un mapa los datos, se necesita obtener los **indicadores** de cada zona escolar y los **contornos** de esas zonas escolares. Vamos a ver cómo.

###### Indicadores

Necesitamos datos lo más desagregados posibles de la red pública y la privada. Los datos disponibles provienen del informe en PDF [La educación en Euskadi 2013-2015](#), del Consejo Escolar de Euskadi. Lo más detallados se dan por zonas escolares, que a veces abarcan varios distritos o municipios enteros. No hay datos disponibles por centros escolares. En los anexos 1.18, 1.19 y 1.21 del informe podemos encontrar varias tablas con los datos que nos interesan.

## 1.19. (para mapa 12) Diferencia entre redes educativas en el porcentaje de alumnado becario de material escolar y de comedor, según las zonas escolares. Curso 2014-15

		% becarios material escolar			% becarios comedor		
		C. públicos	C. privados	Diferencia	C. públicos	C. privados	Diferencia (M17)
111	Amurrio	48	35	13	33	14	19
112	Llodio	38	35	3	26	15	11
121	Vitoria-Oeste	44	43	1	25	7	18
122	Vitoria-Norte	44	41	3	31	6	25
123	Vitoria-Sureste	46	19	27	35	5	30
124	Montaña alavesa	52	0	52	28	0	28
125	Araba/Alava occidental	45	29	16	31	15	16
131	Rioja alavesa	64	36	28	35	20	15
141	Llanada oriental	47	46	1	20	12	8

### Contornos de zonas escolares

Lamentablemente, los contornos de las zonas escolares no están disponibles en ningún sitio (incluso la definición de la composición es a veces imprecisa, sobre todo para el caso de Vitoria-Gasteiz). Es por ello que, para facilitar el ejercicio, hemos creado un archivo .geojson con los contornos de las zonas escolares de Euskadi. Está hecho a partir de la fusión de dos archivos: el de barrios y el de municipios de Euskadi.

### c. Qué herramientas uso para preparar los datos

#### Indicadores

Como hemos señalado, los datos de los indicadores están disponibles en los anexos del PDF del informe del Consejo Escolar de Euskadi. Como lamentablemente ocurre muchas veces con este tipo de datos, no es posible obtener los datos en formatos reutilizables (hojas de cálculo, .csv, .json) y es necesario recomponer esos datos para poder usarlos. Para ello, vamos a crear una nueva tabla que incluya en cada fila una zona escolar y en cada columna cada una de las variables que nos interesan. Para no tener que transcribir uno a uno los números, se pueden usar herramientas automatizadas, como el software libre [Tabula](#), o por métodos semiautomáticos usando la maravillosa función del corta-pega y algunos trucos.

		% becarios material escolar			% becarios comedor		
		C. públicos	C. privados	Diferencia	C. públicos	C. privados	Diferencia (M17)
111	Amurrio	48	35	13	33	14	19
112	Llodio	38	35	3	26	15	11
121	Vitoria-Oeste	44	43	1	25	7	18

Tabla con celdas combinadas en la cabecera

**Tabula** es un programa fácil de usar. Primero, hay que importar el PDF que contiene las tablas y luego seleccionar las que se quieren exportar. Funciona bien para tablas que no tienen celdas combinadas, como le ocurren a muchas de las tablas de los anexos en su cabecera. Para que detecte bien las columnas, tenemos que dejar fuera los encabezados de celdas combinadas, como se indica en este gráfico:

		% becarios material escolar			% becarios comedor		
		C. públicos	C. privados	Diferencia	C. públicos	C. privados	Diferencia (M17)
111	Amurrio	48	35	13	33	14	19
112	Llodio	38	35	3	26	15	11
121	Vitoria-Oeste	44	43	1	25	7	18
122	Vitoria-Norte	44	41	3	31	6	25

Si hacemos clic en [Preview & Export Extracted Data](#) se obtiene la tabla:

Tabula My Files About Help Source Code

LA EDUCACIÓN EN EUSKADI\_2013-15... Export Format: CSV Export Copy to

### Preview of Extracted Tabular Data

		C. públicos	C. privados	Diferencia	C. públicos	C. privados	Diferencia (M17)
111	Amurrio	48	35	13	33	14	19
112	Llodio	38	35	3	26	15	11
121	Vitoria-Oeste	44	43	1	25	7	18
122	Vitoria-Norte	44	41	3	31	6	25

Is the extracted data incorrect? You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

Choose Alternate Extraction Method The current preview uses the

Ahora se puede exportar a diversos formatos (CSV, TSV, JSON) o directamente copiar y pegar en nuestro programa de hoja de cálculo preferido. Como muchas de las tablas se dividen en varias páginas, basta con copiar y pegar en el mismo documento. Atención: como los nombres de las columnas no se han importado bien, necesitarán ser retocados y añadir a qué indicador se refiere. Nota: Tabula permite dos métodos de extracción; Lattice da mejores resultados para este tipo de tablas.

← Revise selection(s)

### Preview of Extracted Tabular Data

	% becar	os material escolar	% becarios comedor				
		C. públicos	C. privados	Diferencia	C. públicos	C. privados	Diferencia (M17)
111	Amurrio	48	35	13	33	14	19
112	Llodio	38	35	3	26	15	11
121	Vitoria-Oeste	44	43	1	25	7	18

Choose Alternate Extraction Method The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

Stream Lattice

Stream looks for *whitespace* between columns, while Lattice looks for *boundary lines* between columns.

Mostramos cómo quedaría la hoja de cálculo que vamos componiendo:

A	B	H	I	K	M
who		Todas las etapas	Todas las etapas	Todas las etapas	Todas las etapas
source		anezo-1.18	anezo-1.18	anexo-1.19	anexo-1.19
date		c.2014-2015	c.2014-2015	c.2014-2015	c.2014-2015
zona-id	zona	perc_material_m15	perc_comedor_m16	perc_bec_mat_esc	perc_bec_mat_escolar_
111	Amurrio	44	25	48	35
112	Llodio	37	21	38	35
121	Vitoria-Oeste	43	20	44	43
122	Vitoria-Norte	43	23	44	41
123	Vitoria-Sureste	31	17	46	19

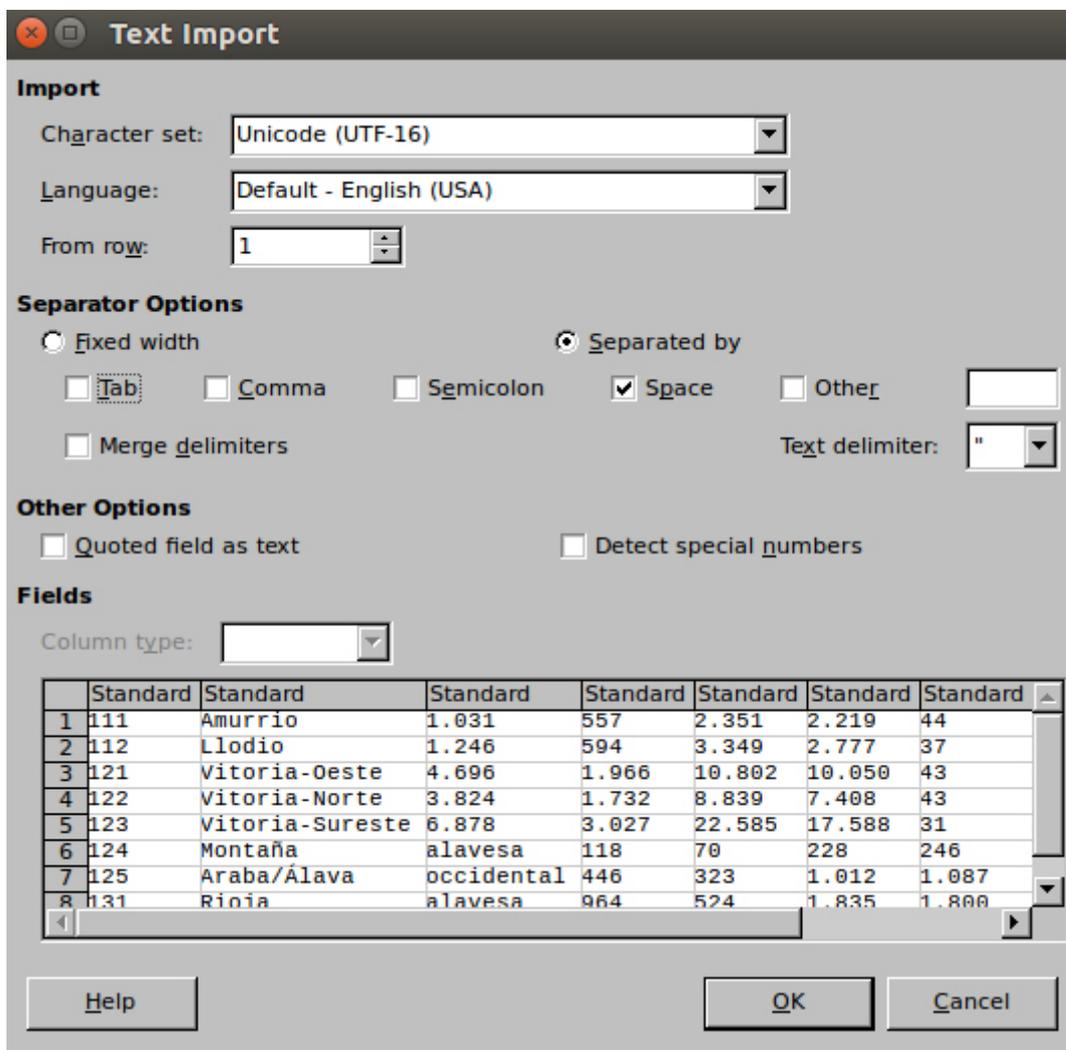
Hemos añadido en la cabecera de cada columna la procedencia de los datos, curso académico y el nombre de la variable sin espacios y con guion bajo. Lo salvaremos o exportaremos como .csv o .tsv y eliminaremos las primeras tres filas del encabezamiento dejando que la primera sea el nombre de las variables:

```

1 zona-id,zona,total-alumnado,perc_alum_ext_pu
  partaide,alum-ext-otros-privados,perc_repart
2 111,Amurrio,1467,8.3,4.3,6.7,1.9,98,73,15,16
3 112,Llodio,1971,7.5,5.1,6.4,1.5,126,80,44,2
4 121,Vitoria-Oeste,7032,11.3,1.8,8.2,6.3,578
5 122,Vitoria-Norte,5002,16.4,11.1,14.6,1.5,73
6 123,Vitoria-Sureste,12489,21.2,4.1,10.6,5.2
7 124,Montaña-alavesa,156,11.5,,11.5,,18,18,,
8 125,Araba/Álava-occidental,752,7.6,6.4,7.4,1

```

La otra forma de extraer los datos del PDF es mediante cortar y pegar en una hoja de cálculo para poder tener en un único fichero de texto plano todos los indicadores. Si seleccionas y copias el texto de una tabla del PDF y lo pegas en una hoja de cálculo, como Libre Office o Excel, seguramente te preguntará con qué delimitadores están separadas las columnas (espacios en este caso); verás que da problemas en los nombres de zonas que llevan espacio como «Montaña-alavesa».



La solución pasa por pegar primero el texto en un editor de texto plano, editar los nombres de las zonas para que no tengan espacios «Montaña-alavesa», copiar y pegar de nuevo en la hoja de cálculo, que ahora no dará problemas.

```

1 111 Amurrio 1.031 557 2.351 2.219 44 25
2 112 Llodio 1.246 594 3.349 2.777 37 21
3 121 Vitoria-Oeste 4.696 1.966 10.802 10.050 43 20
4 122 Vitoria-Norte 3.824 1.732 8.839 7.408 43 23
5 123 Vitoria-Sureste 6.878 3.027 22.585 17.588 31 17
6 124 Montaña alavesa 118 70 228 246 52 28
7 125 Araba/Álava occidental 446 323 1.012 1.087 44 30
8 131 Rioja-alavesa 964 524 1.835 1.800 53 29
9 141 Llanada-oriental 529 210 1.125 1.154 47 18

```

Puedes [descargar directamente el CSV](#) con las variables extraídas, pero no aprenderás a usar Tabula.

Listado de nombres de variables y su descripción:

Variables	Nombre de variables
alum_bec_mat_escolar	n.º alumnado becados material escolar
alum_bec_comedor	n.º alumnado becados comedor
alum_mat_escolar_total	n.º alumnado material escolar total
alum_comedor_total	n.º alumnado comedor total
perc_material_m15	% alumnado beca material escolar
perc_comedor_m16	% alumnado beca comedor

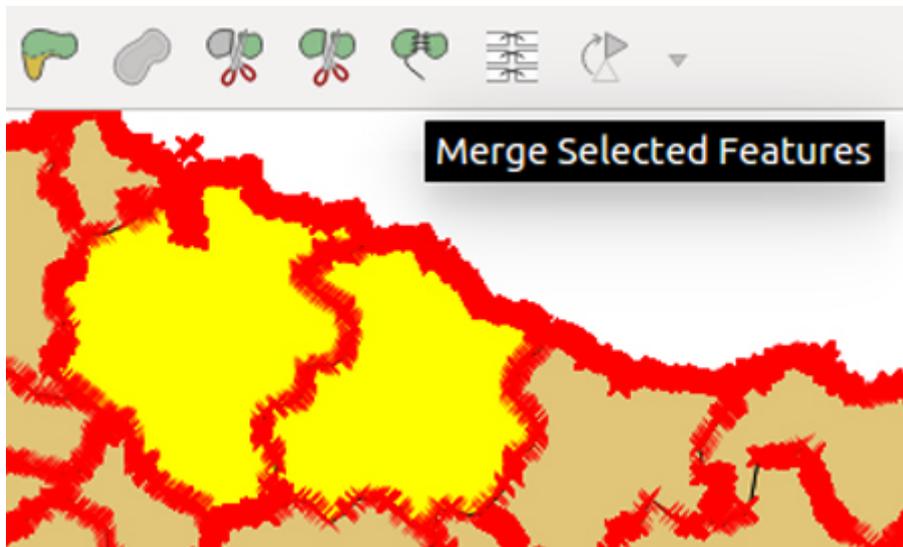
perc_bec_mat_escolar_pub	% alumnado becados material escolar pública
perc_bec_mat_escolar_priv	% alumnado becados material escolar privada
dif_perc_bec_mat_escolar_dif	diferencia % alumnado becados material escolar entre pública y privada
perc_bec_comedor_pub	% alumnado becados comedor pública
perc_bec_comedor_priv	% alumnado becados comedor privada
dif_perc_bec_comedor	diferencia % alumnado becados comedor entre pública y privada
total_alumnado	total alumnado
perc_alum_ext_publi	% alumnado extranjero pública
perc_alum_ext_priv	% alumnado extranjero privada
perc_alum_ext_todos	% alumnado extranjero todos
indice_desigualdad_ext	índice desigualdad extranjeros (% publica / % privada)
alum_ext_total	n.º alumnado extranjero total
alum_ext_publi	n.º alumnado extranjero pública
alum_ext_kristau_eskola	n.º alumnado extranjero en red Kristau Eskola
alum_ext_partaide	n.º alumnado extranjero en red Partaide
alum_ext_otros_privados	n.º alumnado extranjero en otros privados
perc_reparto_alum_ext_publi	% reparto alumnado extranjero en pública
perc_reparto_alum_ext_privada	% reparto alumnado extranjero en privada

## Contornos

Como hemos comentado en el anterior apartado, hemos generado con QGIS el archivo de los contornos sobre la base de la información disponible. Os ahorramos ese paso que consume mucho tiempo y queda fuera del alcance de este ejercicio. Hemos dejado indicado la composición de las zonas en este [archivo TSV](#), en la columna `municipios_barrios`. Los contornos de las zonas han sido construidos a partir de:

- [Límites municipales](#).
- Límites de barrios de Euskadi ([fuente original](#) y [modificada](#) con proyección WGS84).

Usando la herramienta «Merge Selected Features» para unir áreas entre sí. Una vez creadas las áreas, se añaden a cada una los datos asociados de ID de zona y nombre.



Iconos de herramientas para manipular áreas en QGIS

Contornos descargables de las zonas escolares en Euskadi:

- [GeoJSON](#).
- [Shapefile](#).

### Asociar indicadores a los contornos

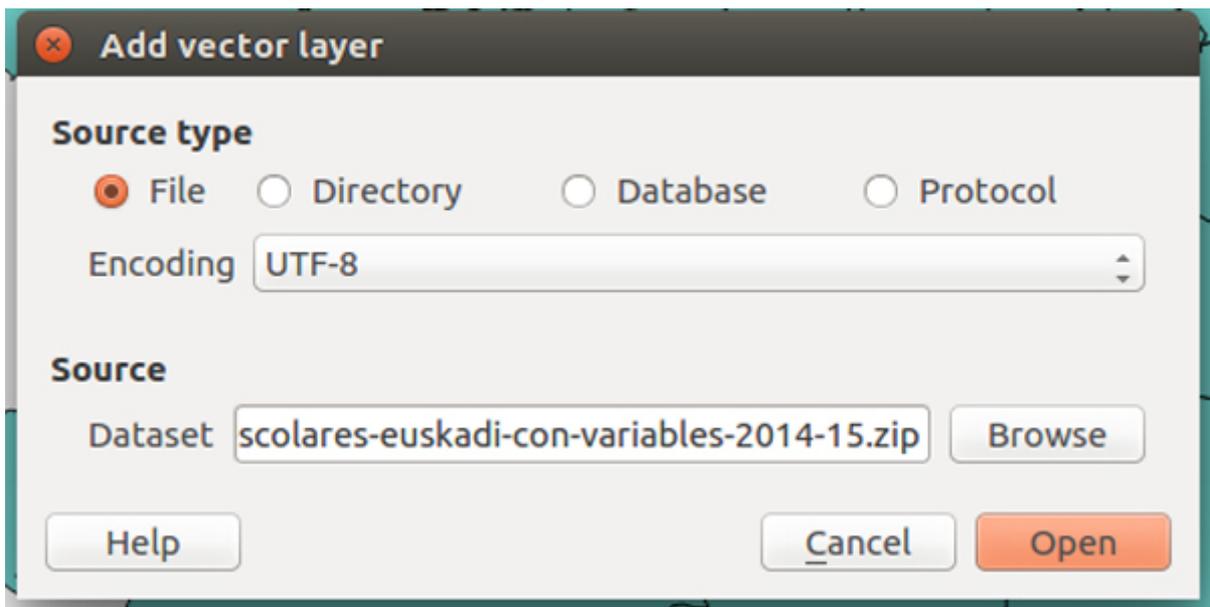
El siguiente paso consiste en asociar los indicadores al archivo de contornos de las zonas escolares.

En el ejercicio anterior de Airbnb hemos realizado algo similar con Carto.com. Esta vez, esta acción la vamos a realizar con QGIS. Para poder unir el archivo de indicadores y los contornos, necesitamos que ambos tengan al menos una columna en común. En nuestro caso, usaremos la variable «zona\_id» de ambos archivos que contiene el número que identifica a la zona escolar.

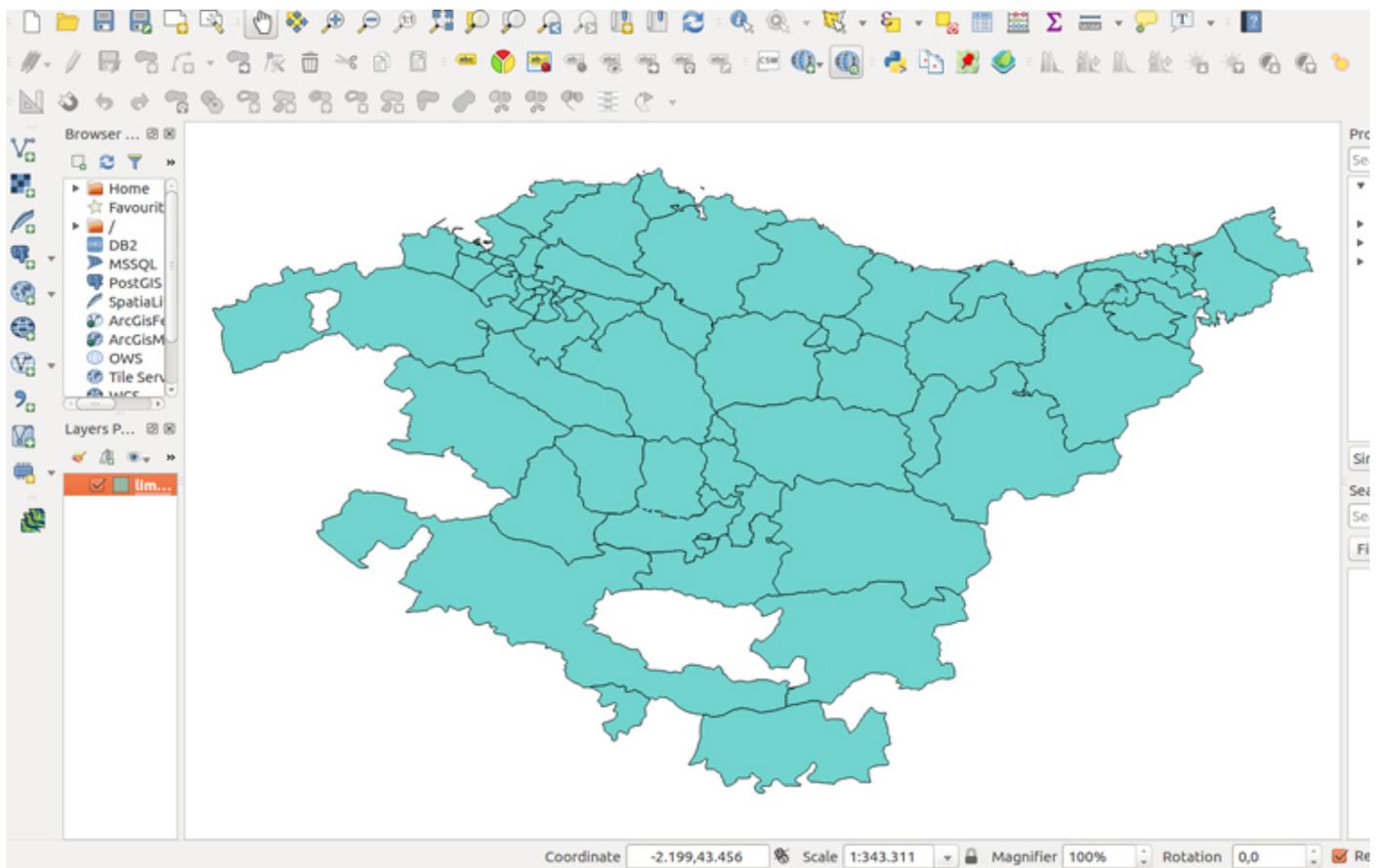
Para poder unir esos dos archivos necesitamos antes importarlos a QGIS. Vamos a ello.

1) Abre QGIS y crea un nuevo proyecto.

2) Importa el archivo shapefile que viene comprimido .zip desde `Layers > Add Layer > Add Vector Layer` usa las opciones por defecto.



Este es el resultado:



**3)** Importamos ahora el CSV con los indicadores **Layers > Add Layer > Add Delimited Text Layer**. Debe estar marcado que está separado por comas, que en la primera fila se encuentran los nombres de las variables y que no tiene geometría (información geoespacial).

**Create a Layer from a Delimited Text File**

File Name

Layer name  Encoding

File format  CSV (comma separated values)  Custom delimiters  Regular expression delimiter

Comma  Tab  Space  Colon  Semicolon

Other delimiters  Quote "  Escape "

Record options Number of header lines to discard   First record has field names

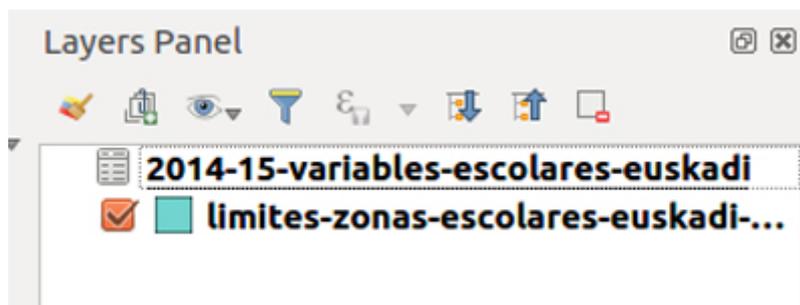
Field options  Trim fields  Discard empty fields  Decimal separator is comma

Geometry definition  Point coordinates  Well known text (WKT)  No geometry (attribute only table)

Layer settings  Use spatial index  Use subset index  Watch file

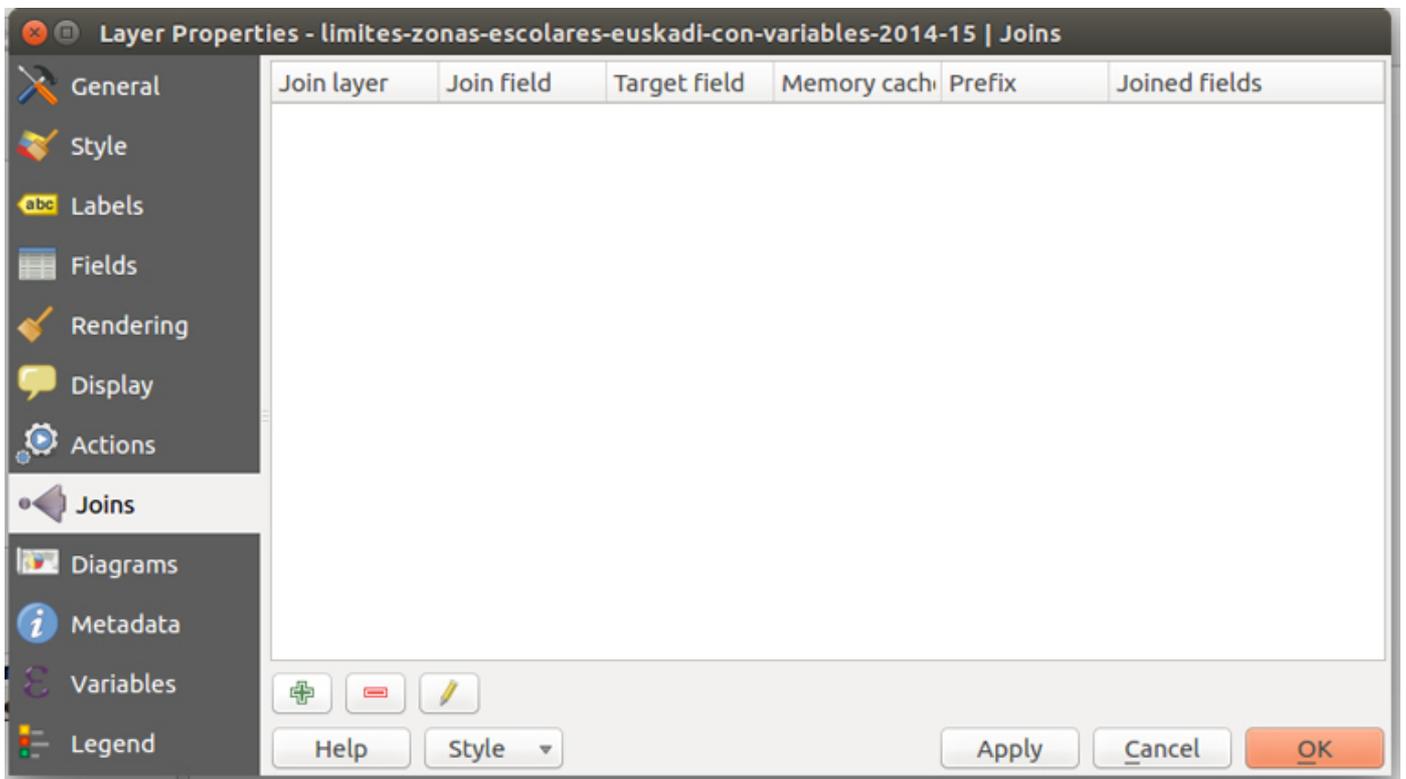
	zona_id	zona	provincia	alum_bec_mat_escolar	alum_bec_comedor	alum_mat_escolar_to
1	111	Amurrio	araba	1031	557	2351
2	112	Llodio	araba	1246	594	3349
3	121	Vitoria-Oeste	araba	4696	1966	10802
4	122	Vitoria-Norte	araba	3824	1732	8839
5	123	Vitoria-Sureste	araba	6878	3027	22585
6	124	Montaña-alavesa	araba	118	70	228
7	125	araba/Álava-occidental	araba	446	323	1012

En el panel de capas deberíamos ver las dos capas:

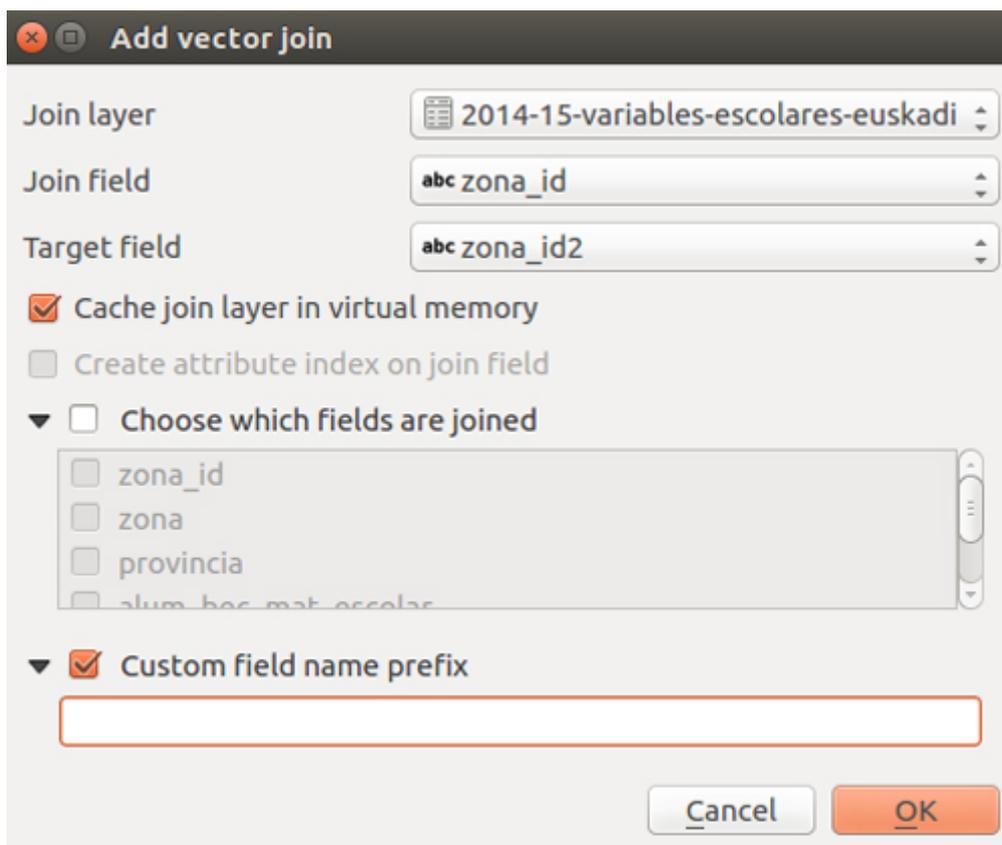


El símbolo de la de arriba indica que es de texto y la segunda que es de información geoespacial.

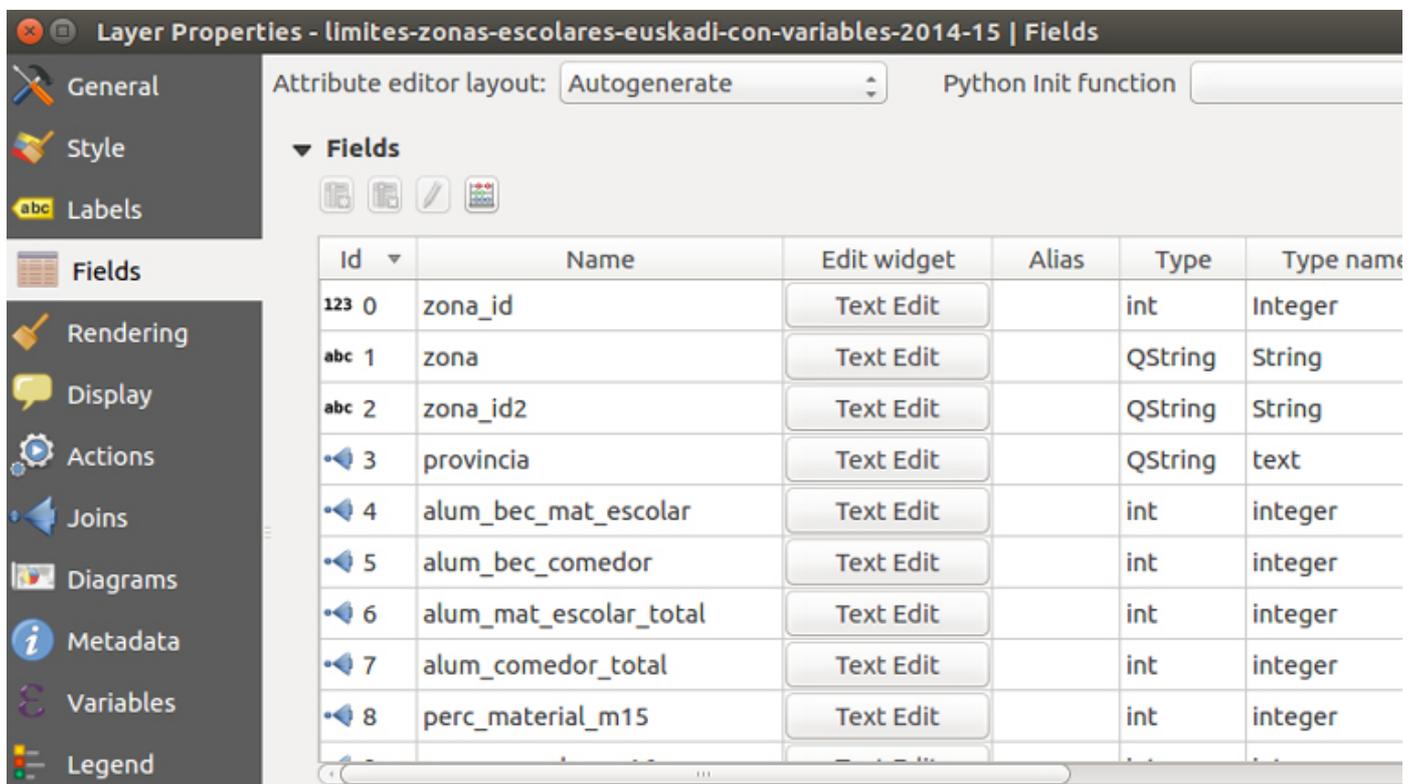
**4)** Hacer clic con botón derecho en la capa de los contornos y selección *Propiedades*. Una vez en el menú de *Propiedades* selecciona *Joins > +*.



Selecciona la columna en común `zona_id` de una capa y `zona_id2` de la otra. En el apartado `Custom field name prefix` déjalo vacío.



En la pestaña `Properties > Fields` verás ahora que las nuevas variables se han añadido al archivo de contornos:



#### d. Qué herramientas uso para producir el análisis o la visualización

Para producir los mapas de cada una de las variables, podemos hacerlos entre otras herramientas con QGIS o Carto.com. Atención: si queremos comparar mapas de coroplemas entre sí, debemos seleccionar una escala de color con los mismos rangos para los diferentes mapas.

Para producir todos los mapas de manera automatizada, podemos hacerlo con la librería de JavaScript D3js.

**5)** Siguiendo con los pasos anteriores, ahora es el momento de dar color a las áreas en función de la variable que queramos en *Properties > Style*. Selecciona *Graduated*, en la *Column* selecciona la variable, en este caso elegimos *Porcentaje de alumnos con beca de material escolar (perc\_bec\_mat\_escolar)*, y haz clic en *Classify* para crear los rangos que asignan los colores. En *Legend Format* elige *Precision 0* para controlar el número de decimales que aparece en la leyenda. En *Mode* puedes elegir el modo de partición de los rangos. Déjalo en *Equal Interval*. Puedes ver cómo se hace esa partición usando la pestaña *Histogram*.

Layer Properties - limites-zonas-escolares-euskadi-2014-15 | Style

General

Style

Column: 123 perc\_bec\_mat\_escolar\_pub

Symbol: Change...

Legend Format: %1 - %2 Precision 0 Trim

Method: Color

Color ramp: [source] Edit Invert

Classes Histogram

Symbol	Values	Legend
<input checked="" type="checkbox"/> [Symbol]	8.00 - 17.20	8 - 17
<input checked="" type="checkbox"/> [Symbol]	17.20 - 26.40	17 - 26
<input checked="" type="checkbox"/> [Symbol]	26.40 - 35.60	26 - 36
<input checked="" type="checkbox"/> [Symbol]	35.60 - 44.80	36 - 45
<input checked="" type="checkbox"/> [Symbol]	44.80 - 54.00	45 - 54

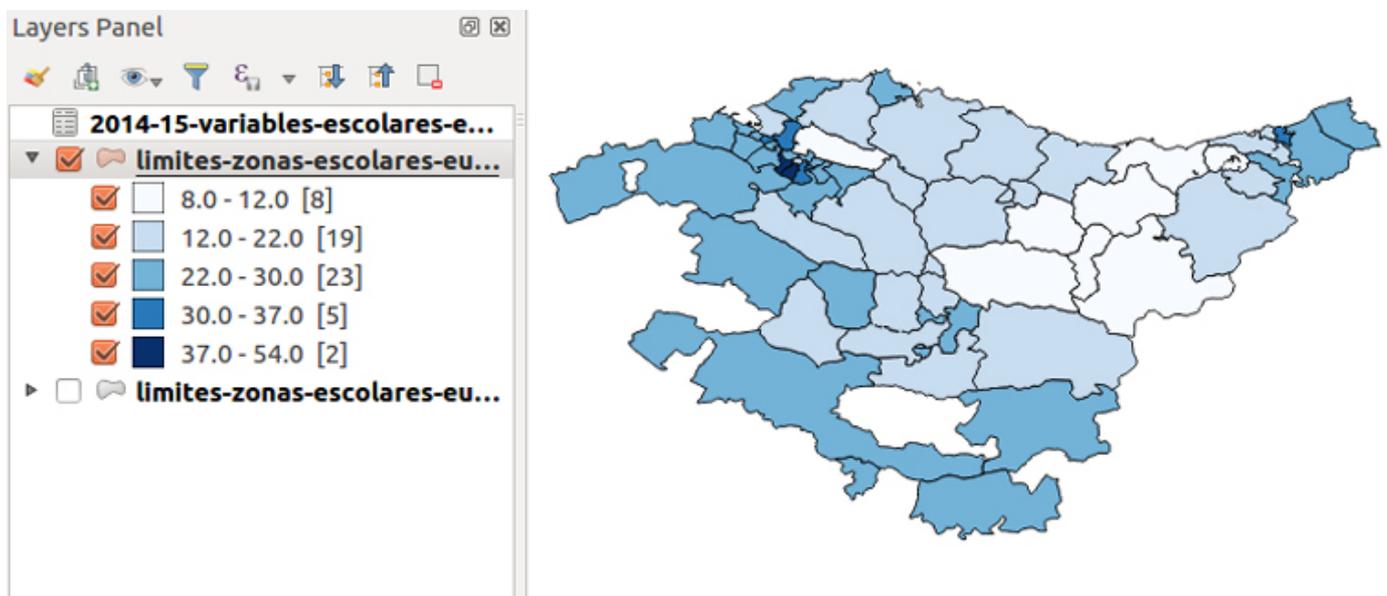
Mode: Equal Interval Classes: 5

Classify + - Delete all Advanced

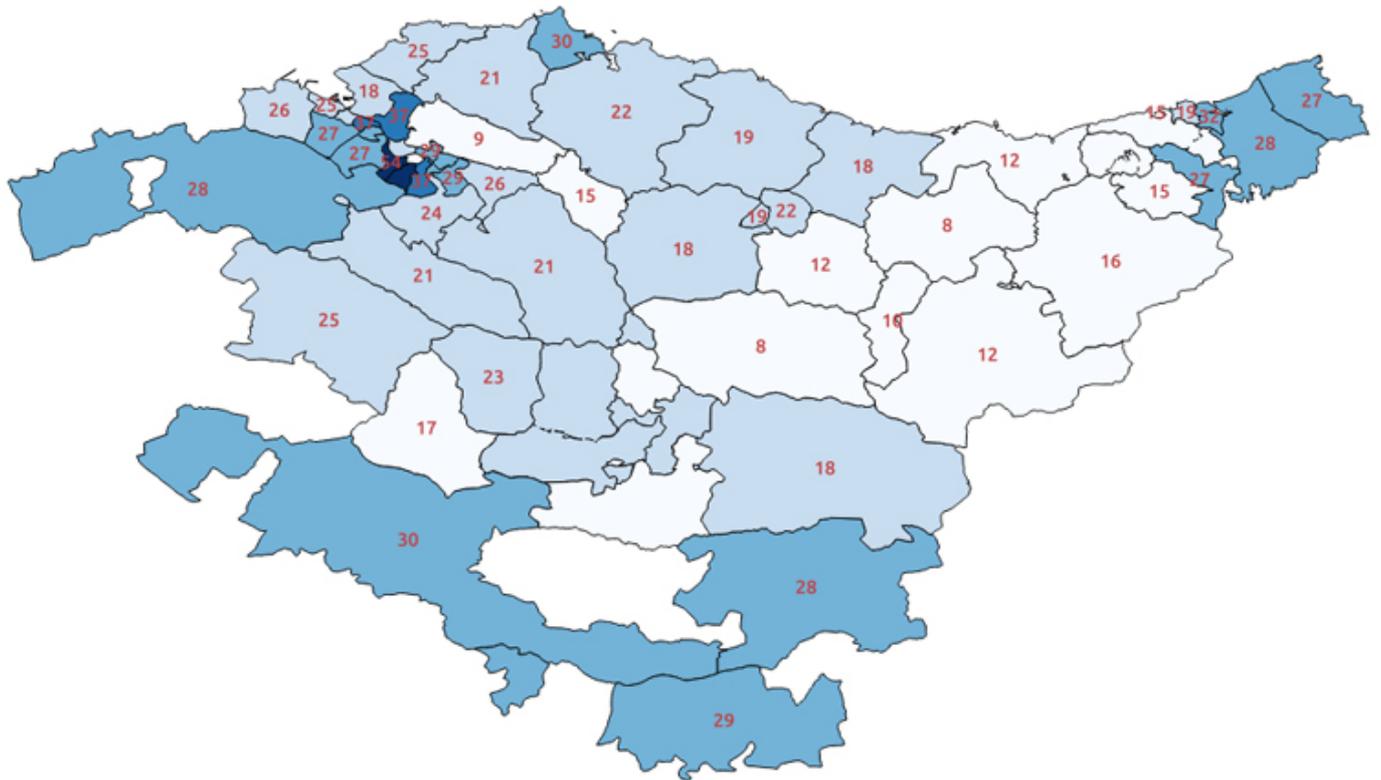
Link class boundaries

Si haces doble clic en los valores de la leyenda, puedes cambiarlos.

Ahora ya tienes el mapa de coropletas, y en la leyenda puedes ver los valores asociados a cada uno de los colores.

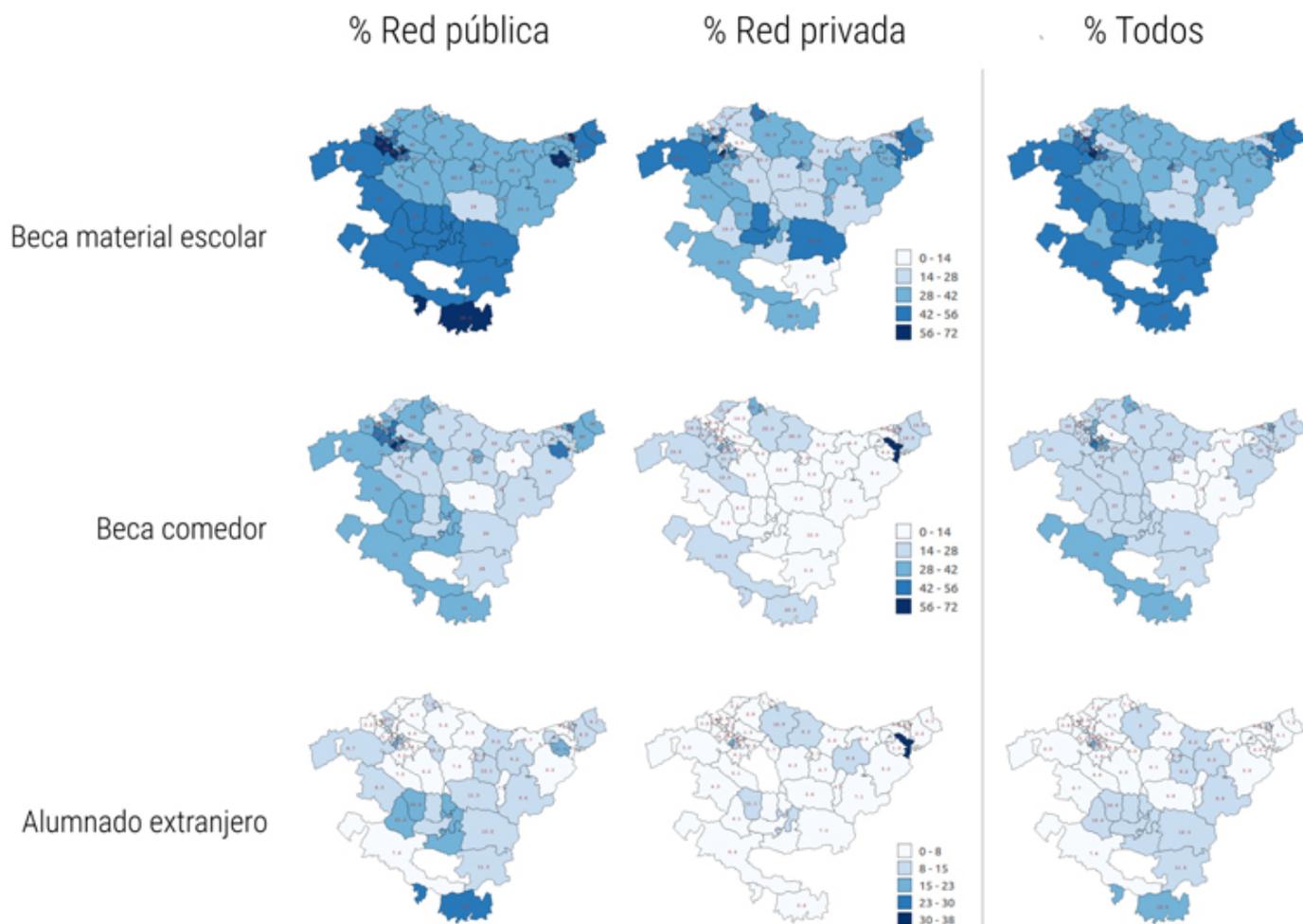


6) En la pestaña Properties > Label puedes añadir un texto a cada contorno; por ejemplo, el valor representado por el color.



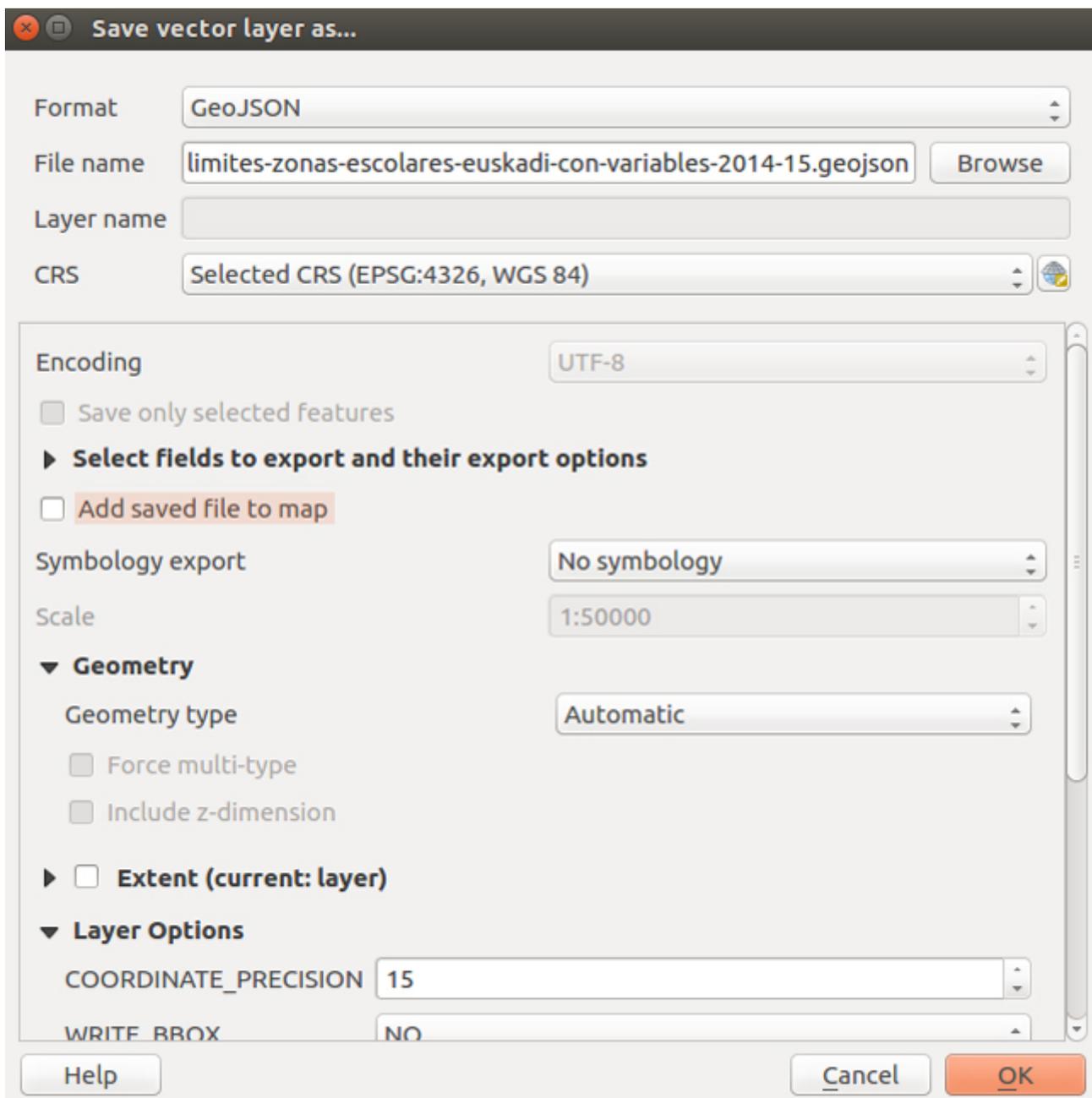
Cambiando la variable que utilizemos para representar en la capa `Properties > Style > Column` podemos generar un mapa para cada variable.

**7)** Desde QGIS podemos exportar los mapas generados a imagen (PNG) o vector (PDF, SVG) componiendo diversos mapas, escalas y leyendas. Aquí tienes un [tutorial para hacer un mapa con múltiples vistas](#). La siguiente composición está hecha mediante capturas de pantalla de QGIS y maquetada con Inkscape:



Este tipo de visualización, llamado *small multiple* en inglés y popularizado por Edward Tufte, que compara diferentes variables de una serie de datos sobre gráficos de la misma escala, ya es suficiente para empezar a ver variaciones espaciales de los indicadores en las redes pública y privada. A partir de aquí pueden surgir nuevas preguntas.

**8)** Antes de terminar con QGIS no hay que olvidar salvar el proyecto y salvar también la capa (botón derecho sobre el nombre de la capa en la lista de capas) `Save vector layer as...` como geojson.



Si quieres puedes [descargar aquí el archivo .geojson](#) de los contornos de las zonas escolares con todas las variables asociadas.

### Generar un/os mapa/s con D3

Para ir un paso más en cuanto a automatización del proceso de generación de los gráficos, usaremos la librería de [JavaScript D3](#). Hacerlo es relativamente sencillo y precisa poco código. Además, podremos crear gráficos interactivos y mejorar con ello la experiencia de usuario.

**1)** Lo primero que tenemos que hacer es **simplificar las líneas que componen** el archivo de los contornos para aligerar su tamaño (pesa más de 7MB) y, con ello, acelerar el periodo de carga de la visualización. También convertiremos el archivo .geojson a un archivo de tipo **topoJSON** que también reduce el tamaño de los archivos hasta un 80 % sin reducir el detalle de los contornos.

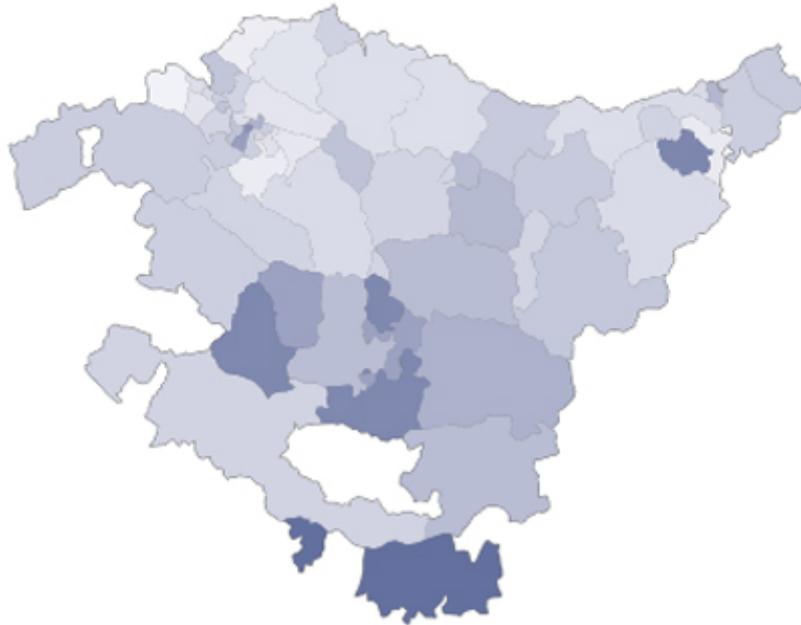
Se puede convertir a topoJSON con el paquete de [NPM topoJSON](#) desde la línea de comandos o usar la herramienta en línea [mapshaper](#). Mapshaper permite subir archivos geoJSON, corregir errores, simplificar su trazado y salvar como topoJSON con una interfaz gráfica.

El archivo resultante es este: [https://github.com/montera34/segregacion/blob/master/data/output/limites-zonas-escolares-euskadi-con-variables-2014-15\\_simplify3.json](https://github.com/montera34/segregacion/blob/master/data/output/limites-zonas-escolares-euskadi-con-variables-2014-15_simplify3.json).

**2)** Producir un mapa de coropletas una sola variable

A partir del topoJSON, archivo que ya contiene los contornos de las zonas escolares y los datos asociados con ellas, es relativamente sencillo mostrar los datos en un mapa. Descarga el [zip](#) con los archivos, descomprímelo y abre el archivo `index.html` desde un navegador web. Deberías ver esto:

## Porcentaje de extranjeros en escuela pública



### Nota para Windows

Para que funcione D3 en local con el sistema operativo Windows, es necesario correr en línea de comandos esto (como recomiendan en esta [respuesta en Stackoverflow](#)):

```
> python -m SimpleHTTPServer 8888
> http://localhost

> python3 -m http.server 8888
> http://localhost
```

Recomendamos usar el navegador Firefox.

Vamos a explicar las partes más importantes del código contenido en `index.html` y que generan el gráfico. Tanto si tienes conocimientos de HTML, CSS y JavaScript como si no, pensamos que podrás entender parte del código y manipular ciertas variables para modificar tamaño, color y variable representada.

Primero, llama a las librerías de JavaScript D3js y Topjson para hacer funcionar la visualización:

```
<script src="https://d3js.org/d3.v4.min.js"></script>
<script src="https://d3js.org/topojson.v1.min.js"></script>
```

Definimos dimensiones del mapa en píxeles:

```
var width = 400, height = 350;
```

Añadimos el SVG donde se dibujará el mapa:

```
var svg = d3.select("body").append("svg")
  .attr("width", width)
  .attr("height", height)
  .style("background-color", "#EFEFEF")
```

Definimos la escala de color:

```
var color = d3.scale.linear() // seleccionamos una escala lineal
  .domain([0, 38]) // valores mínimo y máximo
  .range(['#fff', '#053874']) // color para el mínimo es blanco (#FFF) y azul
  (#053874) para el máximo
```

Selecciona el tipo de proyección del mapa (Mercator), el nivel de zoom y el centro del mapa:

```
var projection = d3.geoMercator() // proyección mercator
  .scale(12000) // zoom level
  .center([-2.6, 42.9]) // center of map
  .translate([width / 2, height / 2]);
```

Por otro lado, llamo al archivo topoJSON para usar los datos que contiene:

```
d3.json("data/output/limites-zonas-escolares-euskadi-con-variables-2014-
15_simplify3.json", function(error, euskadi) {
```

Crea variable que contiene todas las zonas; para ello accedemos al archivo JSON que hemos llamado euskadi en la anterior línea y accedemos a `objects` y `zonas`.

```
var subunits = topojson.feature(euskadi, euskadi.objects.zonas);
```

Ahora queda dibujar el mapa:

```
svg.append("path")
  .datum(subunits)
  .attr("d", path)
  .attr("stroke", "grey");
```

Dibuja los contornos de las zonas y las colorea (atributo `fill`) según una variable (`perc_alum_ext_publi`):

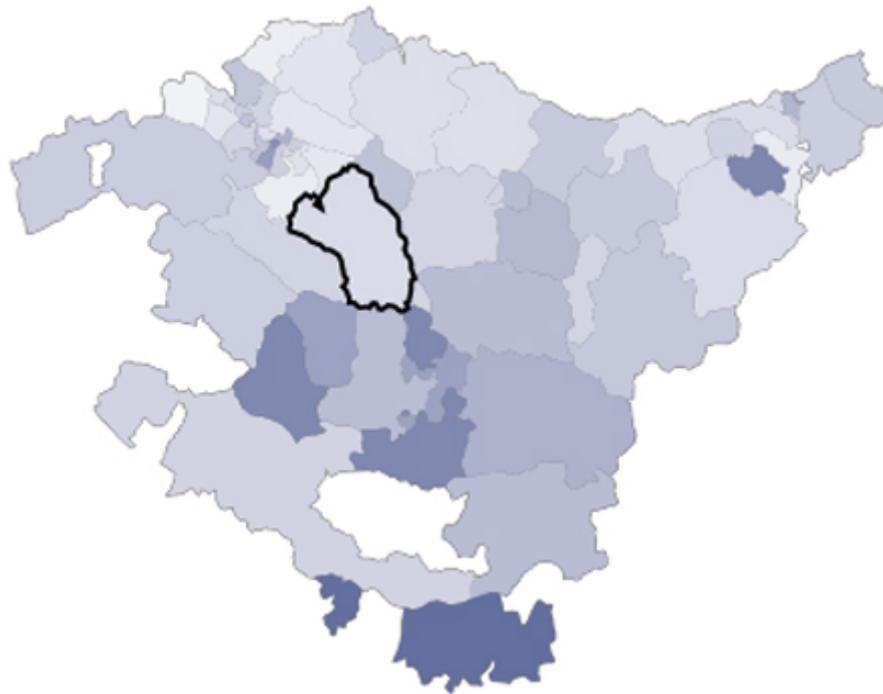
```
svg.selectAll(".subunit")
  .data(subunits.features)
  .enter().append("path")
  .attr("class", function(d) { return "subunit z" + d.properties.zona_id; })
```

```
.attr("fill", function(d) { return color(d.properties.perc_alum_ext_publi); }) //  
colorea las áreas según la variable "perc_alum_ext_publi"  
.attr("d", path)
```

Si cambiamos la variable `perc_alum_ext_publi` por otra como `perc_alum_ext_priv` colorearemos el mapa acorde a otro indicador. Incluye la variable en la línea `.attr("fill", function(d) { return color(d.properties.perc_alum_ext_priv); })`. **Atención: tendrás que cambiar el dominio de la escala `.domain([0, 38])` si quieres cambiar sus valores máximos y mínimos. Los valores que queden fuera del dominio dibujarán los contornos rellenos de color negro (haced la prueba cambiando el dominio de `[0, 38]` a `[0, 4]`).**

Para añadir interactividad, hay que añadir lo indicado en rojo:

```
// Adds tooltip  
var tooltip = d3.select("body")  
  .append("div")  
  .attr("class", "tooltip")  
  [...]  
svg.selectAll(".subunit")  
  .data(subunits.features)  
  .enter().append("path")  
  .attr("class", function(d) { return "subunit z" + d.properties.zona_id; })  
  .attr("fill", function(d) { return color(d.properties.perc_alum_ext_publi); }) //  
colorea las áreas según la variable "perc_alum_ext_publi"  
  .attr("d", path)  
  .on("mousemove", showTooltip) // AÑADIR EVENTO SHOW TOOLTIP  
  .on("mouseout", hideTooltip); // OCULTAR TOOLTIP  
  
// Función para mostrar el tooltip  
function showTooltip(d) {  
  tooltip.html("<div class='table-responsive'><strong>" + d.properties.zona + " " +  
d.properties.perc_alum_ext_publi + "%</strong> (zona " + d.properties.zona_id2 + ")  
</div>").style("opacity", 1)  
  }  
// Función para ocultar el tooltip  
function hideTooltip(d) {  
  tooltip.style("opacity", 0)  
  }  
}
```



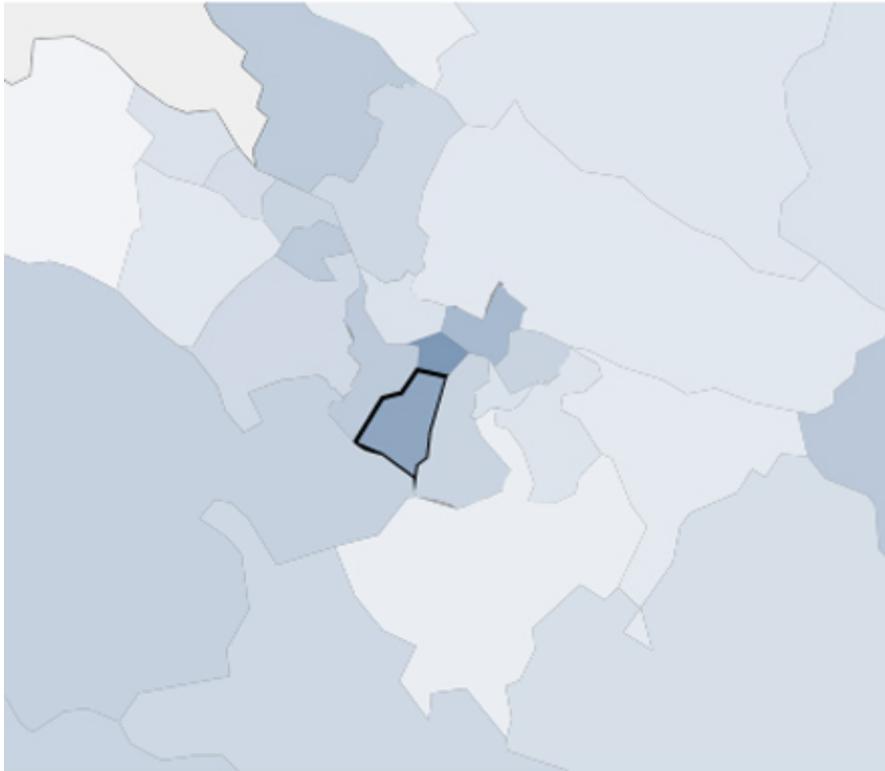
Al pasar el cursor indica que esa zona es Igorre con un 6,2 % de alumnado extranjero en red pública.

Si se cambia el zoom y el centro del mapa, por ejemplo, para Bilbao:

```
var projection = d3.geoMercator()  
  .scale(60000) // note change of zoom  
  .center([-2.9331,43.2521])  
  .translate([mapWidth / 2, mapHeight / 2]);
```

Se muestra un zoom de la zona de Bilbao:

## Porcentaje de extranjeros en escuela pública

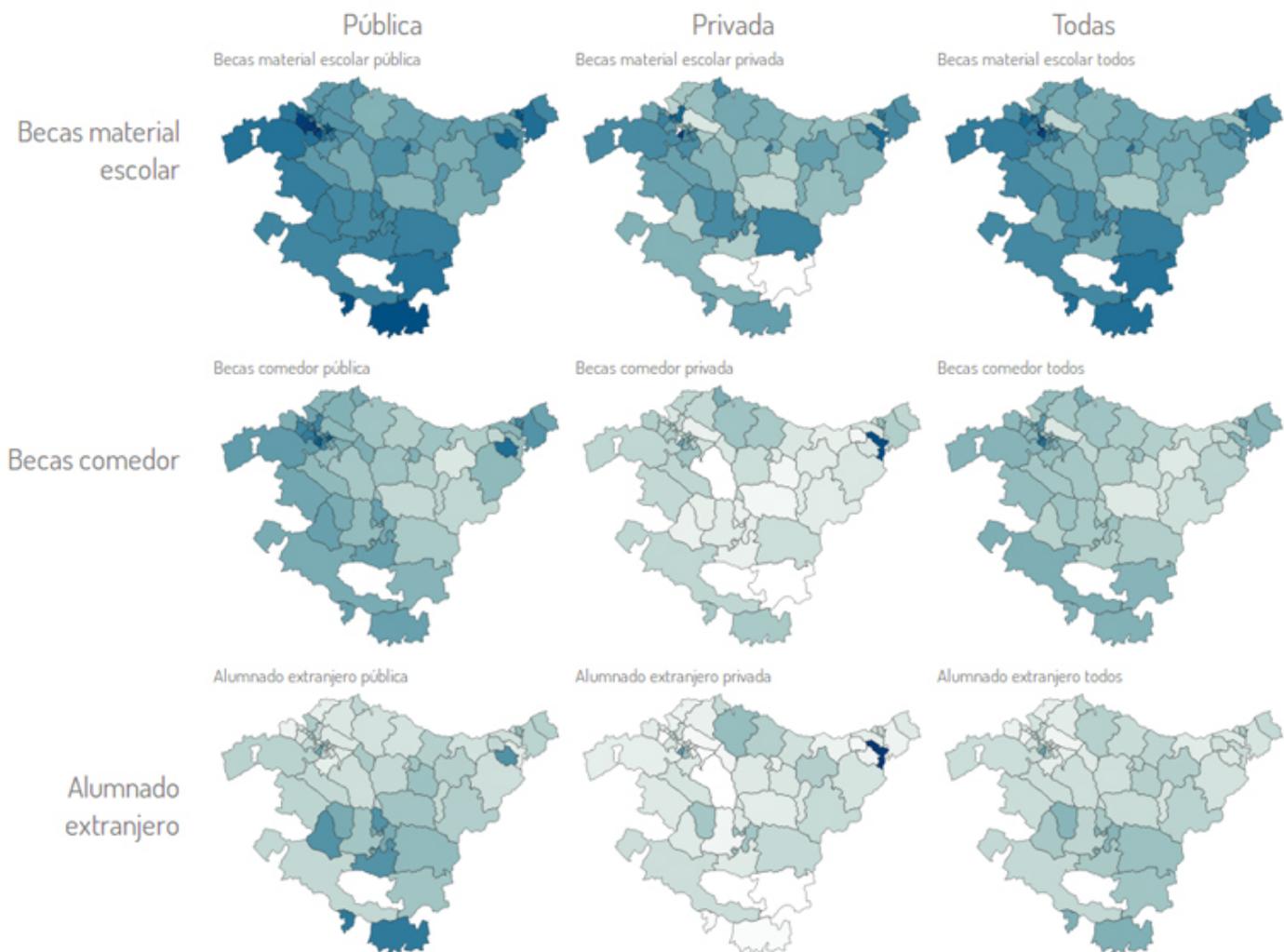


**Rekalde 17% (zona B7)**

Puedes ver el código completo y la visualización aquí: <http://bl.ocks.org/numeroteca/33325b3234175c5450b856026c62a920> y <https://gist.github.com/numeroteca/5764a860ecc0f7c4d421b0ad00bd6188>.

**3)** Generar múltiples mapas interactivos, cada uno con su propia variable.

Para generar múltiples mapas, lo que tenemos que hacer es replicar el mapa del paso anterior. Descarga el [.zip](#) con los archivos, descomprímelo y abre el archivo `index.html` desde un navegador web. Deberías ver esto:



Puedes ver el [código completo](#) y la [visualización](#).

Primero, tenemos que definir dos escalas de color diferentes, porque los dominios (valor mínimo y máximo de cada variable) son diferentes:

```
// Escala de color para becas
var color = d3.scaleLinear().domain([0, 72]).range(['#fff', '#053874'])
// Escala de color extranjeros
var color2 = d3.scaleLinear().domain([0, 38]).range(['#fff', '#053874'])
```

En una matriz definimos las variables que queremos que aparezcan:

```
// Array to list variables to display in map and display name of variables
var mapasdef = [
  ["Becas material escolar pública", "perc_bec_mat_escolar_pub"],
  ["Becas material escolar privada", "perc_bec_mat_escolar_priv"],
  ["Becas material escolar todos", "perc_material_m15"],
  ["Becas comedor pública", "perc_bec_comedor_pub"],
  ["Becas comedor privada", "perc_bec_comedor_priv"],
  ["Becas comedor todos ", "perc_comedor_m16"],
  ["Alumnado extranjero pública", "perc_alum_ext_publi"],
  ["Alumnado extranjero privada", "perc_alum_ext_priv"],
  ["Alumnado extranjero todos ", "perc_alum_ext_todos"]
]
```

Crea un div (contenedor) para cada elemento de la matriz que acabamos de definir:

```
var variableJoin = d3.select('#maps').selectAll('div.map')
  .data(mapasDef);
var divs = variableJoin.enter()
  .append('div')
  .attr('id', function(d){ return 'map_'+d; })
  .attr('class','map')
  .style('float','left')
  .style('width',mapWidth);
```

Crea los contenedores párrafo (p) y span que alojarán el nombre de la variable que el mapa representa y el número:

```
divs.append('p').text(function(d){ return d[0]; });
divs.select('p').append('span').text(" ")
```

Añade un contenedor SVG para cada mapa:

```
var zonas = divs.append('svg').attr('width',mapWidth).attr('height',mapHeight);
```

Inserta en cada uno de los SVG que hemos creado en la línea anterior los contornos de todas las zonas y les aplica un color dependiendo de la escala de color que hemos definido.

```
zonas.each(function(key, i=1){
  d3.select(this).selectAll('path')
    .data(subunits.features)
    .enter().append("path")
    .attr("fill", function(d) { return i > 5 ? color2(d.properties[key[1]]) :
color(d.properties[key[1])); }) //uses two color scales for % Becas (grants) and % of
foreigners
    .attr("d", path)
    .attr("stroke", "black")
    .attr("stroke-width", "0.3px")
})
```

Si añadimos lo siguiente, conseguimos que escriba el porcentaje en la cabecera de cada mapa.

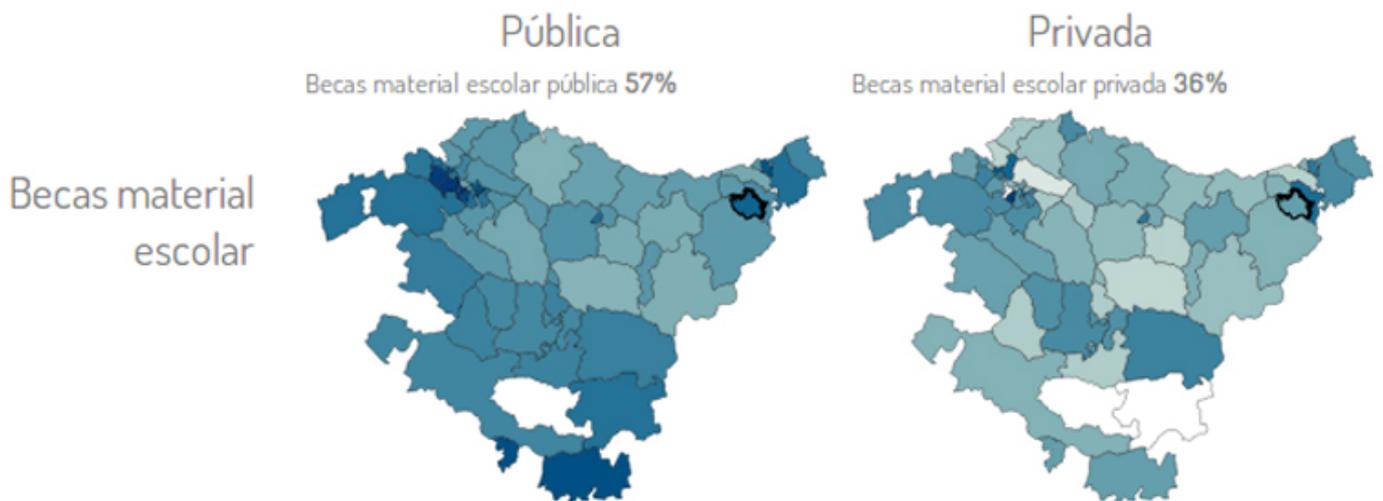
```
.on('mouseenter', function(d, i) { // selects path with same class in other maps and
changes to wider troke
  notify('.z' + d.properties.zona_id2, 'select');
  d3.selectAll('.z' + d.properties.zona_id2).attr("stroke-width", "2px");
})
.on('mouseleave', function(d) {
  notify('.z' + d.properties.zona_id2, 'unselect');
  d3.selectAll('.z' + d.properties.zona_id2).attr("stroke-width", "0.3px");
})
```

```

.on('select', function(self) {
    var geoData = self.data(); // creates variable with data contained in the zone (zona
    escolar) where the cursor is on
    //looks for the "span" element in the div and writes inside the percentage of that
    variable
    self.node().parentNode.parentNode.getElementsByTagName('span')[0].innerHTML = " " +
geoData[0].properties[key[1]]+ "%";
})
.on('unselect', function(self) {
    self.node().parentNode.parentNode.getElementsByTagName('span')[0].innerHTML = "";
});

// Interactivity from https://blog.webkid.io/multiple-maps-d3/
// Una función usada para "notificar" a los otros path que comparten class con el señalado
por el cursor
function notify(selector, eventName) {
    d3.selectAll(selector)._groups[0].forEach(function(el, i) {
        var shape = d3.select(el);
        shape.on(eventName)(shape);
    });
}
}

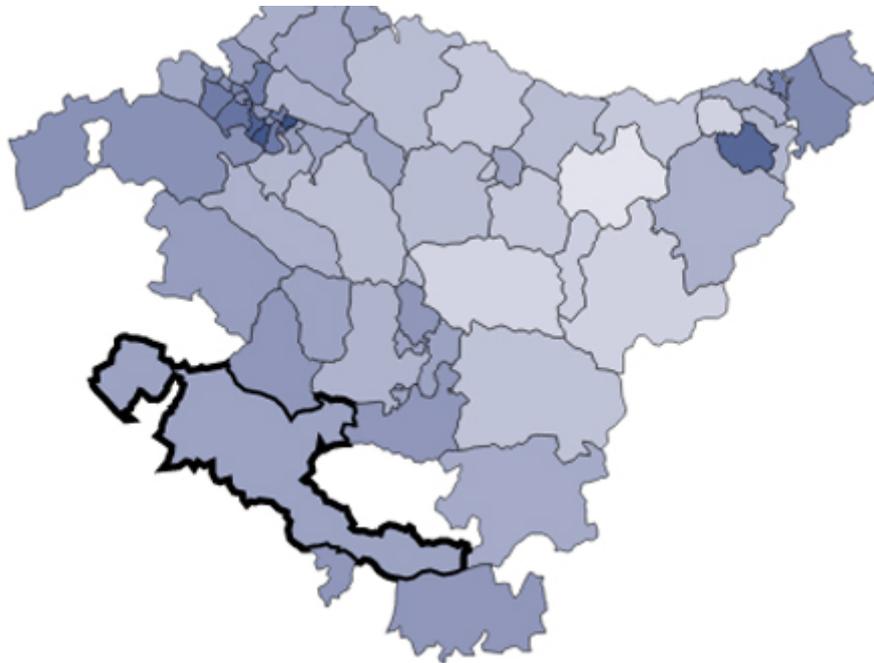
```



Aunque el código de D3 pueda resultar complicado si no estás familiarizado con JavaScript y HTML, creemos que es posible probar a modificar ciertos elementos para conseguir cambios deseados: cambio de tamaño del mapa, de nivel de zoom, de colores. Hay que hacer pruebas.

### Otras formas de representar: los mapas no son la única opción

Los mapas suelen ser la opción preferida para representar regiones, pero no tienen por qué ser siempre las mejores. Hay mucha teoría al respecto, pero baste con mencionar que es difícil para el ojo distinguir las variaciones de color y que depende mucho de los colores que le rodean para evaluar lo oscuro o claro que es. Además, las regiones con poco alumnado pero muy extensas quedan sobrerrepresentadas en el mapa. Es el caso, por ejemplo, de la zona escolar Álava occidental, una de las más extensas pero la segunda con menor cantidad de alumnado (752).

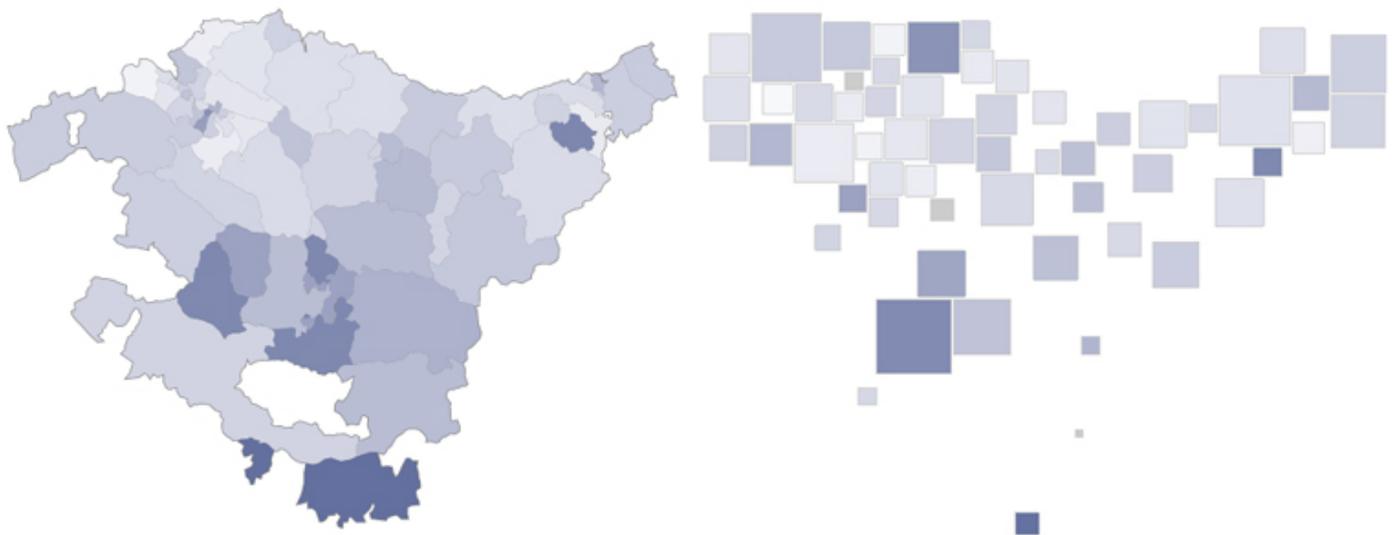


Se marca la zona escolar de Álava occidental.

Visualmente en el mapa tendrá mucha más presencia que algunas zonas de Bilbao como Abando, que tiene 6.054 alumnos, pero casi invisible por su tamaño en el mapa.

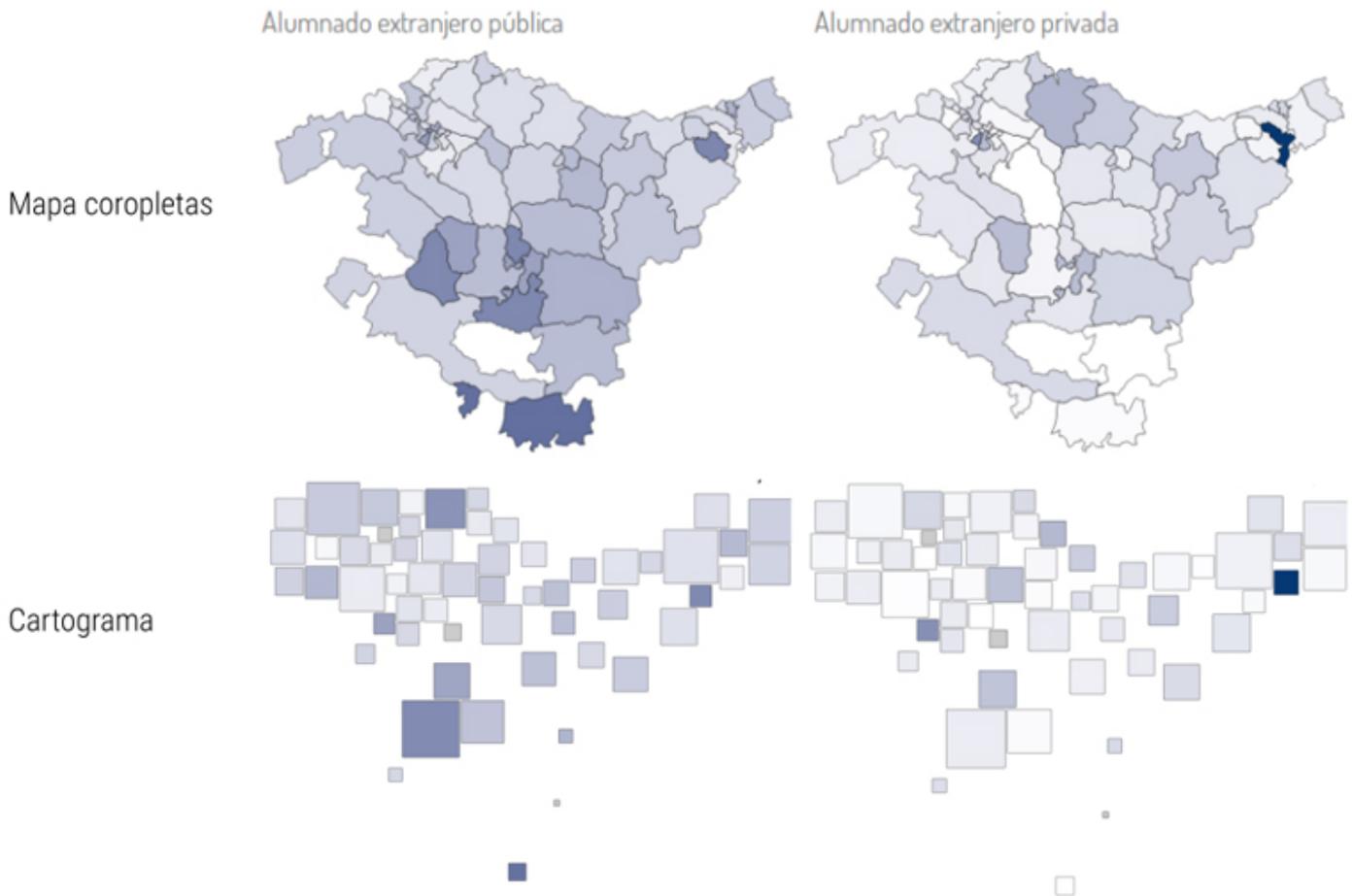
### Cartogramas

Para corregir este tipo de desigualdades visuales, existen otro tipo de representaciones como los cartogramas. En este tipo de cartograma se representa cada zona escolar como un cuadrado, donde su área es proporcional al alumnado que tiene y su color en función, en este caso, del porcentaje de una de las variables, como en los mapas anteriores. Los cartogramas tienen la dificultad de mantener la forma geométrica de las zonas que hace perder al mapa su legibilidad geográfica, lo que puede desorientar a quien vea la visualización. Lo que busca el cartograma de este tipo (ya que hay cartogramas contiguos que intentan deformar lo menos posible las regiones) es ser lo más



Mapa de coropletas (izquierda). Cartograma (derecha; el tamaño de las zonas es en función de la cantidad de alumnado)

No hay una forma buena o mala, cada una sirve para explorar los datos de una forma diferente. Aquí repetimos la comparativa red pública y red privada con mapas de coropletas y cartograma.



La [visualización](#) y el [código](#) (basado en cartogramas de [Martín González](#)) para realizar estos cartogramas están disponibles. Por su extensión no entraremos a comentar el código, pero os animamos a utilizarla para generar vuestros propios cartogramas.

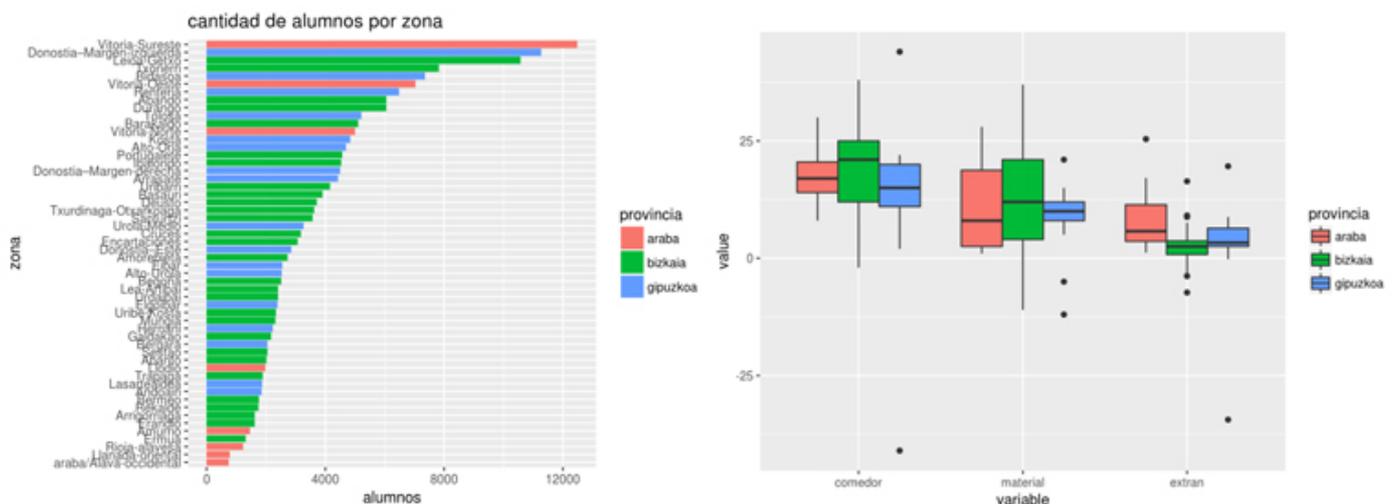
### Explorando los datos y sus correlaciones con R

Una vez realizada la exploración espacial de los datos, es decir, cuáles son las diferencias entre las redes a nivel geográfico, proponemos explorar las diferencias y relaciones entre los indicadores de otras formas.

Sugerimos usar R con RStudio (RStudio) para hacer un primer acercamiento a los datos.

#### 1) Preanálisis

Puedes descargar el archivo con estos preanálisis de [segregacion-analisis.Rmd](#) para hacerlos correr en TStudio o directamente descargar y abrir en un navegador el archivo [segregacion-analisis.html](#), que no necesita R para funcionar.



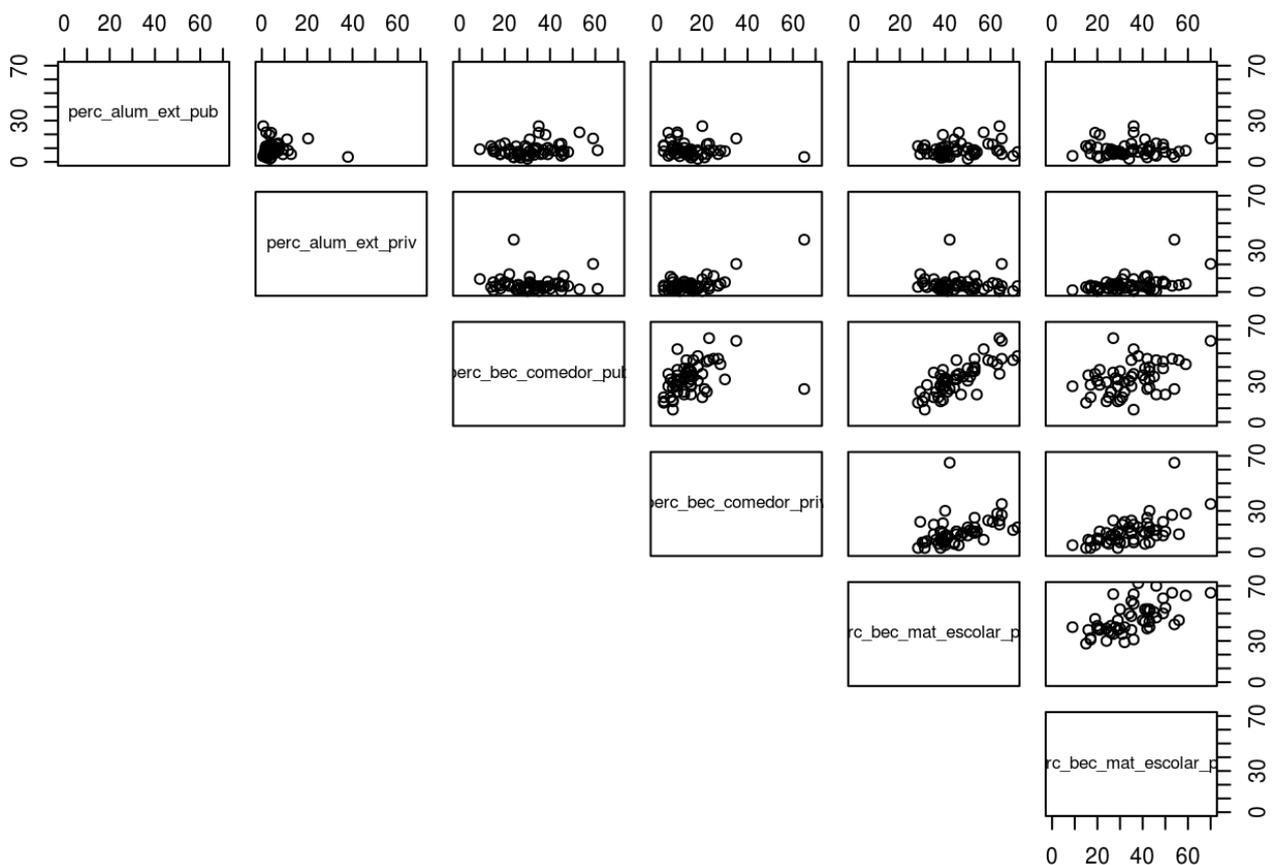
Puedes acceder en línea también a este [preanálisis](#).

#### 2) Correlaciones

Ahora vamos a explorar si existen correlaciones entre las variables estudiadas: % becas de material escolar y comedor, % de presencia de alumnado extranjero en redes pública y privada dentro de cada una de las zonas escolares. El [código](#) y [las visualizaciones](#) están disponibles.

Con muy poco código tenemos esta primera visualización con `pairs`:

```
segreg = read.csv("2014-15-variables-escolares-euskadi.csv")
attach(segreg)
pairs(~perc_alum_ext_pub+perc_alum_ext_priv+perc_bec_comedor_pub+perc_bec_comedor_priv+perc_bec_mat_escolar_p
      xlim=c(0,70),ylim = c(0,70),
      lower.panel = abline(a = 0,b = 1)
```



Donde se aprecian correlaciones a nivel visual entre los indicadores:

- Becas material escolar y Becas comedor en red pública.
- Becas material escolar y Becas comedor en red privada.

La librería `GGally` calcula, además de mejorar la visualización, las correlaciones para todas las zonas y para cada una de las provincias, y genera histogramas y diagramas de dispersión separando por colores, dependiendo de la provincia, cada una de las zonas escolares:

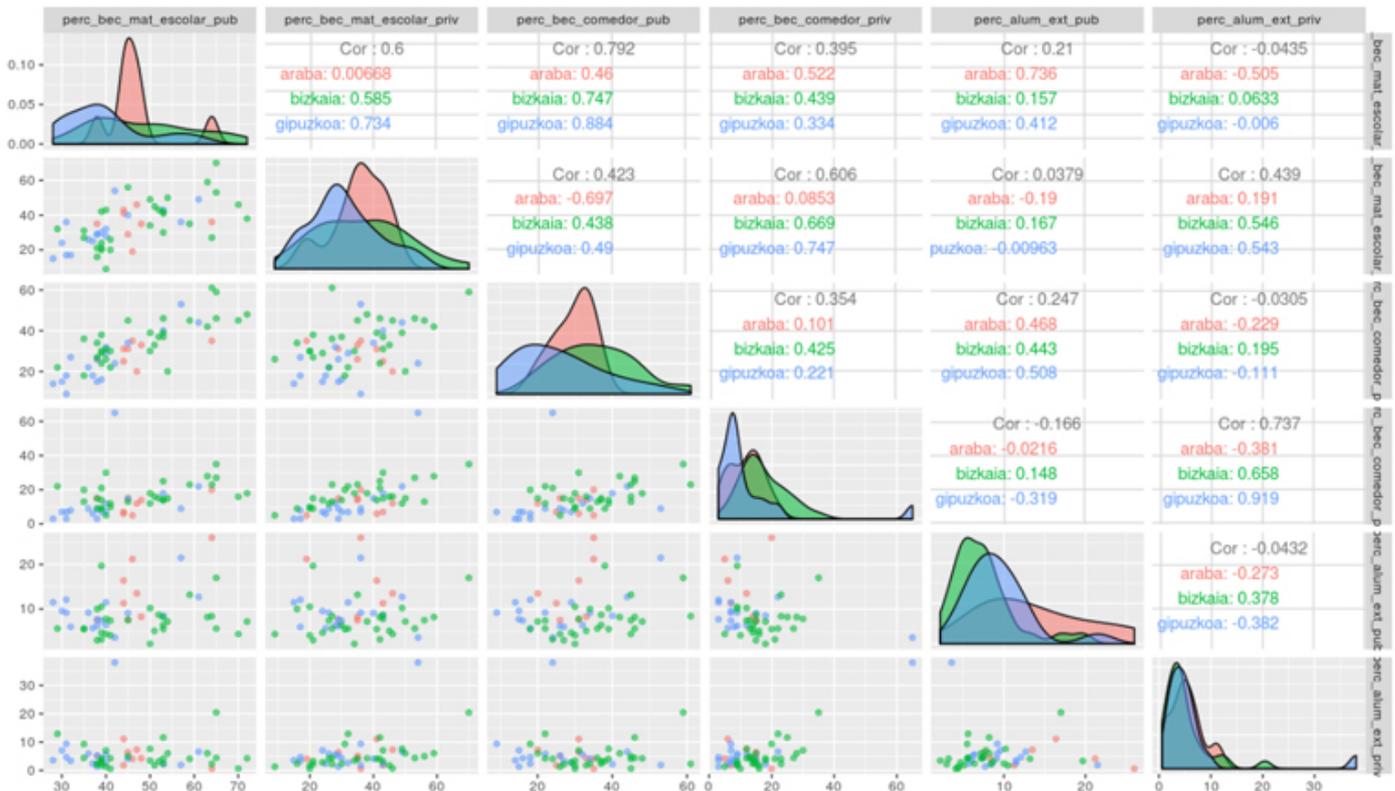
```
library(GGally)
ggpairs(segreg,
        columns = c(10,11,13,14,17,18),
```

```
title = "buscando correlaciones",
aes(colour = provincia, alpha = 0.4))
```

Con todo, es mejor, antes de computar correlaciones, dejar fuera las zonas escolares de Montaña alavesa, Igorre y Basurto-Zorroza, porque no tienen oferta privada:

```
segreg = segreg[c(-6,-33,-57),] #elimina los que no tienen red privada
attach(segreglimpio)
```

Y volvemos a calcular:



Las correlaciones más fuertes entre las variables una vez calculadas son:

- Becas material escolar y Becas comedor en red pública **0,792** (Bizkaia 0,747, Gipuzkoa **0,884**, Araba 0,46).
- Becas comedor y Alumnado extranjero en red privada **0,737** (Bizkaia 0,658, Gipuzkoa **0,919**, Araba -0,381).
- Becas material escolar y Becas comedor en red privada **0,606** (Bizkaia 0,669, Gipuzkoa **0,747**, Araba 0,085).

Así, los valores más altos son siempre los de la provincia de Gipuzkoa.

Esta correlación indica que existe una correlación relativamente fuerte entre el porcentaje de alumnado becado en comedor y el porcentaje becado con material escolar en la red pública. Refuerza la idea de que las becas son una medida del nivel socioeconómico e indica cierta homogeneidad dentro de cada zona escolar, mayor en la red pública (0,792) que en la privada (0,606). También, hay una fuerte correlación entre porcentaje de becarios de comedor y alumnado extranjero en la red privada (0,737). ¿Qué nos indica esta correlación? ¿Que la presencia de alumnado extranjero es indicador de nivel socioeconómico? Esa correlación es mucho más baja (0,247) en la red pública.

Para terminar esta exploración, nos vamos a centrar en ver si existen correlaciones entre **las diferencias entre la red pública y privada** para cada indicador y zona. Para ello, generamos tres nuevas variables:

- comedor = % beca comedor en red pública – % beca comedor en red privada.
- material = % beca material escolar en red pública – % beca material escolar en red privada.

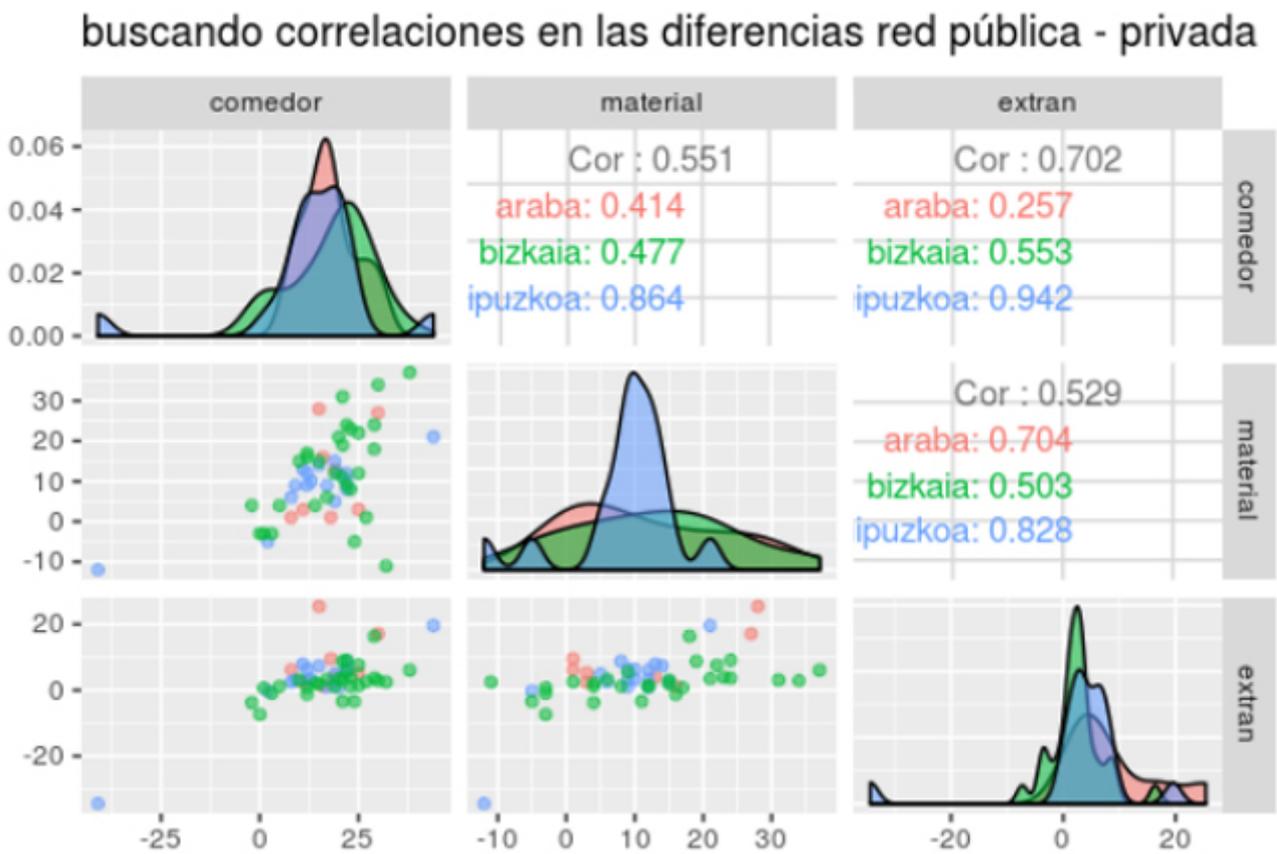
- extran = % alumnado extranjero en red pública – % alumnado extranjero en red privada.

```

comedor = data.frame(perc_bec_comedor_pub - perc_bec_comedor_priv)
material = data.frame(perc_bec_mat_escolar_pub - perc_bec_mat_escolar_priv)
extran = data.frame(perc_alum_ext_pub - perc_alum_ext_priv)

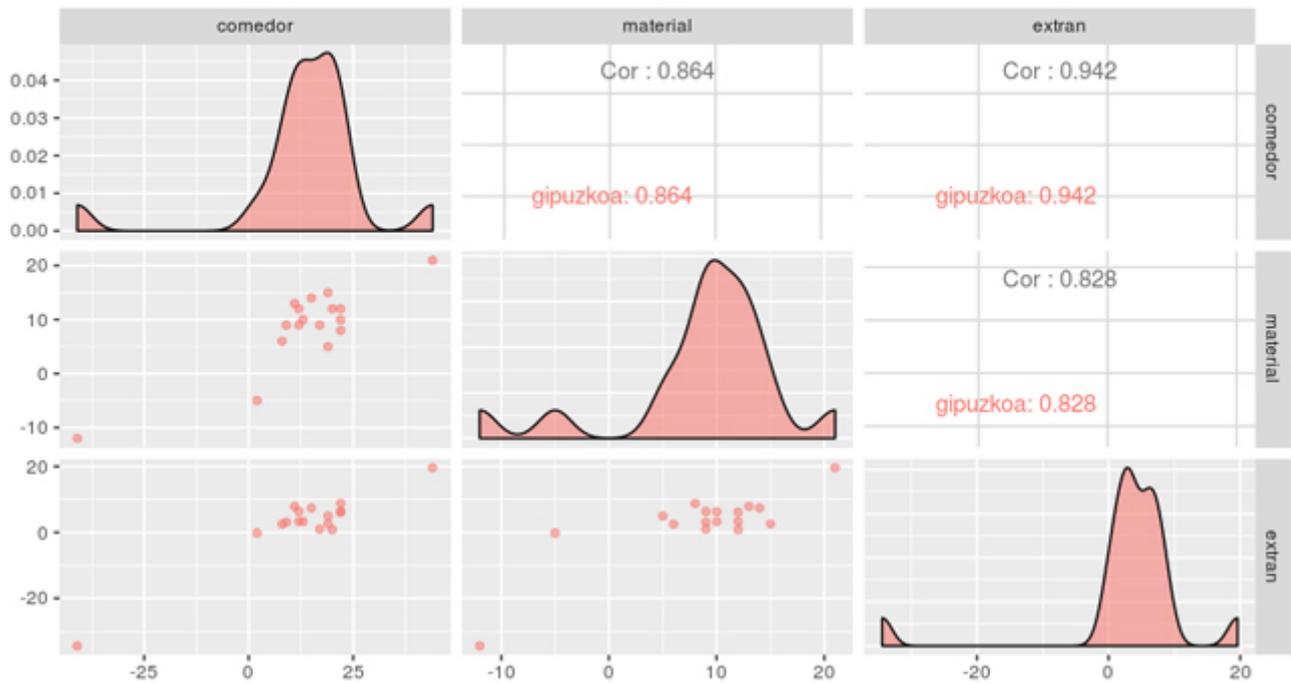
segredif = data.frame(comedor,material,extran,provincia,zona)
colnames(segredif) = c("comedor","material","extran","provincia","zona")
ggpairs(segredif,
  title = "buscando correlaciones en las diferencias red pública - privada",
  columns = c(1:3),
  aes(colour = provincia, alpha = 0.4))

```



Obtenemos valores moderadamente fuertes (0,551, 0,702, 0,529) y llaman la atención los valores muy altos para Gipuzkoa en todos los indicadores de diferencia (0,864, 0,942 y 0,828):

## buscando correlaciones en las diferencias red pública - privada en Gipuzkoa

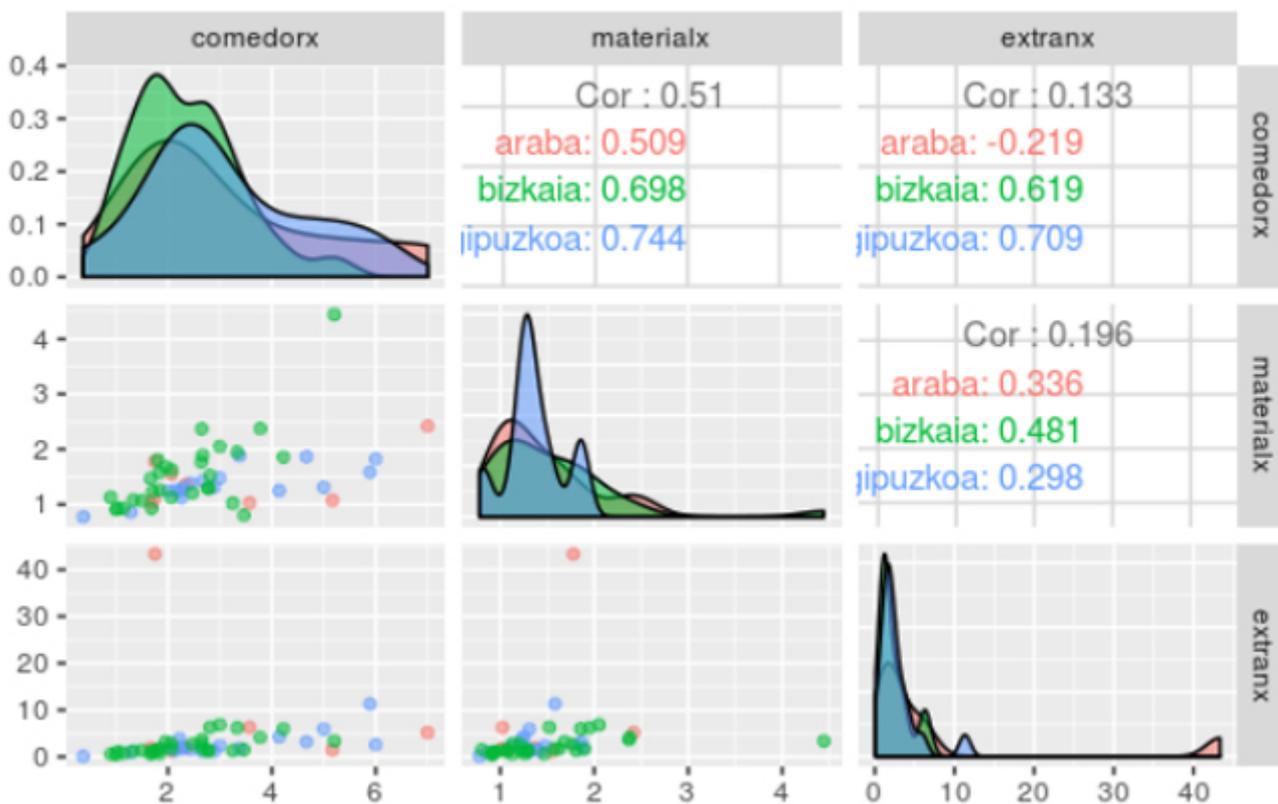


Y si calculamos los cocientes entre las variables (de hecho, el informe recoge ese cálculo para calcular el grado de desequilibrio):

- $\text{comedorx} = \% \text{ beca comedor en red pública} / \% \text{ beca comedor en red privada}$
- $\text{materialx} = \% \text{ beca material escolar en red pública} / \% \text{ beca material escolar en red privada}$
- $\text{extranx} = \% \text{ alumnado extranjero en red pública} / \% \text{ alumnado extranjero en red privada}$

```
comedorx = data.frame(perc_bec_comedor_pub / perc_bec_comedor_priv)
materialx = data.frame(perc_bec_mat_escolar_pub / perc_bec_mat_escolar_priv)
extranx = data.frame(perc_alum_ext_pub / perc_alum_ext_priv)

segredif = data.frame(comedor,material,extran,comedorx,materialx,extranx,provincia,zona)
colnames(segredif) =
c("comedor","material","extran","comedorx","materialx","extranx","provincia","zona")
ggpairs(segredif,
  title = "buscando correlaciones",
  columns = c(4:6),
  aes(colour = provincia, alpha = 0.4))
```



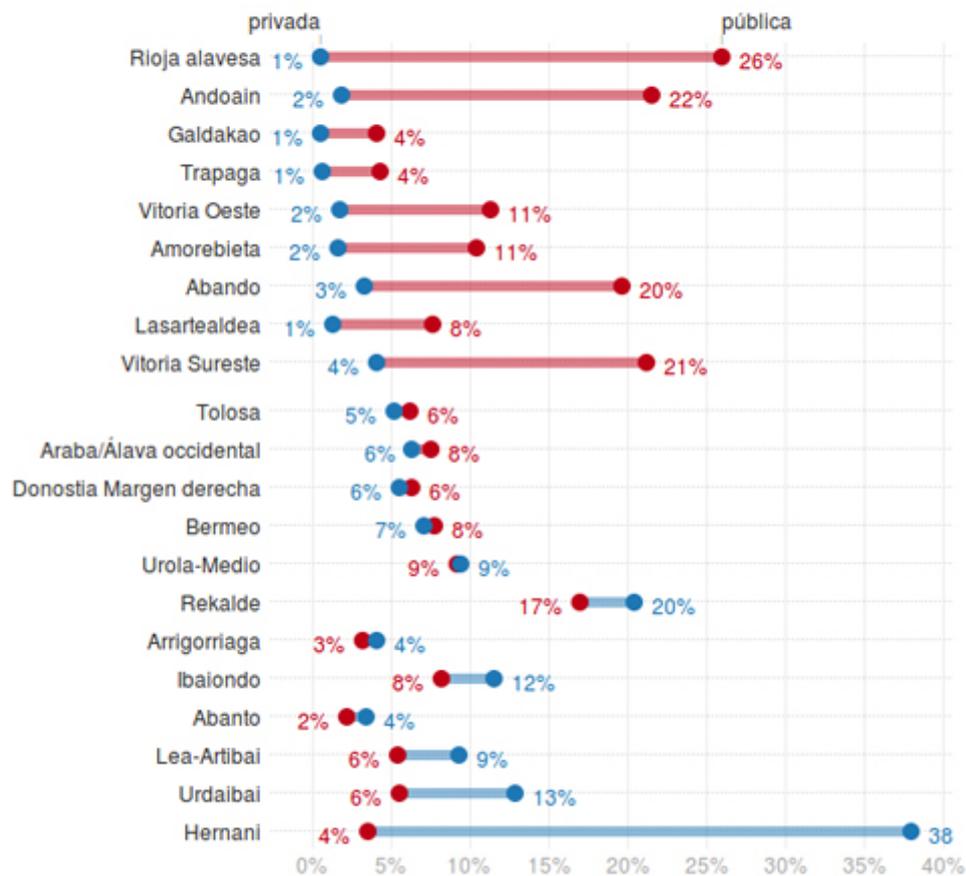
No se aprecian correlaciones fuertes en general (0,51, 0,133, 0,196), aunque sí entre los cocientes de becas de material con comedor y alumnado extranjero con comedor para Gipuzkoa (0,744 y 0,709 respectivamente) y Bizkaia (0,698 y 0,619 respectivamente). Puede haber una historia que explorar a partir de estos datos en Araba, donde la correlación es más pobre e incluso negativa entre el cociente de extranjeros y becas de comedor. ¿Qué historia hay en los datos de Gipuzkoa?

### Mostrar las diferencias

Estas exploraciones nos ayudan a entender los datos, pero quizás puedan ser complejas para publicarlas en un medio generalista para todos los públicos. Una vez que hemos identificado que las **diferencias público-privado para cada zona** son una de las variables a representar, sugerimos el siguiente tipo de gráfico en el que se representa, de forma clara, para una de las variables. Hemos usado para generarlo Datawrapper (<http://datawrapper.de/>) herramienta en línea en la que subes los datos y escoges el tipo de visualización para luego insertarlo donde quieras:

## Diferencia entre red pública y privada de % de alumnado extranjero en zonas escolares de Euskadi

Todas las etapas. Ordenados por mayor % de cambio.



Fuente: La educación en Euskadi 2013-2015 • Get the data • Created with Datawrapper

Puedes ver el gráfico en [datawrapper](#).

Nos parece muy interesante porque ofrece los valores de cada zona y hace particular énfasis en la cantidad de la diferencia. El color de la línea es según si el mayor porcentaje es de la red pública (rojo) o privada (azul).

### e. Cómo presento la información para hacer un relato inteligible

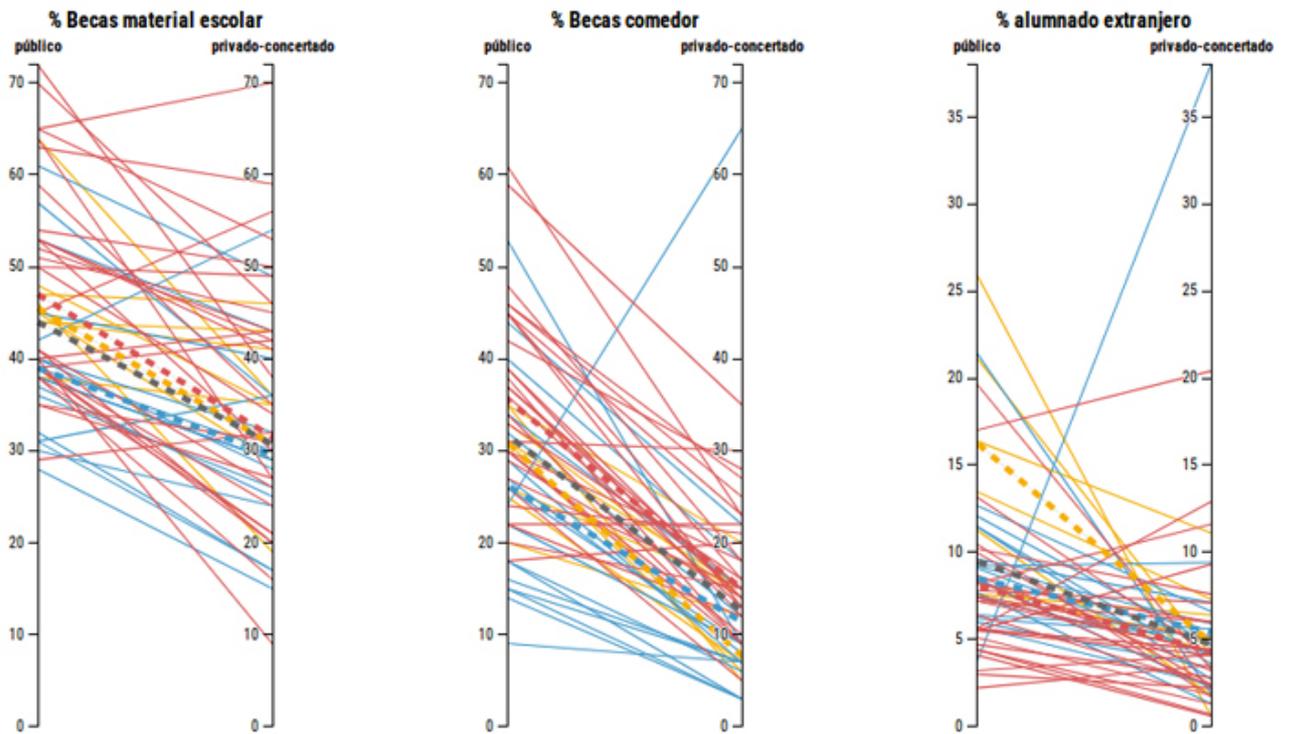
Una vez que recopiles y analices todos los gráficos realizados, es momento de seleccionar cuáles quieres utilizar y pensar qué historia quieres contar. Seguramente, ninguno de los gráficos te valga tal cual y tengas que modificarlos o buscar nuevas variables. Este ejercicio ha intentado ser un ejemplo de exploración con el que empezar a hacer nuevas preguntas.

¿Qué historias esconden las diferencias entre la red pública y la privada? ¿Qué ocurre en la provincia de Araba? ¿Por qué Hernani tienes unos valores tan diferentes al resto de zonas escolares? Se podrían ver si existe relación entre las zonas escolares y el tipo de contexto urbano-rural.

Por último, otra visualización en la que la diferencia se expresa por la pendiente de la línea que va del valor de la izquierda (público) a la derecha (privado) para cada zona y cada indicador, basado en el tipo de gráfico de coordenadas paralelas:

# Diferencia de porcentaje de alumnado con becas de material escolar, comedor y alumnado extranjero en zonas escolares

— Araba 
 — Gipuzkoa 
 — Bizkaia 
 — Media (ponderada)
 Selecciona zona escolar ▾
Todas las zonas



Ver el gráfico en modo interactivo y [accede al código](#).

Para saber más sobre este gráfico de coordenadas paralelas puedes leer este [artículo](#).

## 6. Tres ejercicios de periodismo de datos

### 6.3. Educación pública-privada: abriendo datos abiertos

#### 6.3.5. Información complementaria

- Cartogramas:
  - [The State of the Art in Cartograms.](#)
  - [Evaluating Cartogram Effectiveness.](#)
  - [Algorithms for Cartogram Computatio](#)

## Créditos

Periodismo de datos, coord. Núria Ferran-Ferrer (2018). UOC. <http://periodisme-dades.uoc.edu/es/>

- Aznar, L. (2018). “El periodisme, el periodisme d’investigació i el periodisme de dades”. A Núria Ferran-Ferrer (coord.), *Periodisme de dades*. UOC <http://periodisme-dades.uoc.edu/es/>
- Cerrillo, A. (2018). “Aspectes jurídics del periodisme de dades”. A Núria Ferran-Ferrer (coord.), *Periodisme de dades*. UOC <http://periodisme-dades.uoc.edu/es/>
- Aragó, L. (2018). “Fonts d’informació i mecanismes d’obtenció de dades”. A Núria Ferran-Ferrer (coord.), *Periodisme de dades*. UOC <http://periodisme-dades.uoc.edu/es/>
- Aragó, L. (2018). “Anàlisi de dades”. A Núria Ferran-Ferrer (coord.), *Periodisme de dades*. UOC <http://periodisme-dades.uoc.edu/es/>

### **Autores:**

**Laura Aznar Lluçà**. Sociòloga y periodista. Redactora a Crític. @LauraAzLlu

Artículos: <http://www.elcritic.cat/?s=laura+aznar&lang=ca>

**Agustí Cerrillo**. Catedrático de Derecho administrativo. Universitat Oberta de Catalunya

**Laura Aragó**. Periodista de datos, investigación y visualización. La Vanguardia. @arago\_la

Artículos: <http://www.lavanguardia.com/autores/laura-arago.html>

### **Coordinación:**

**Núria Ferran-Ferrer**. Profesora en los Estudios de Ciencias de la Información y la Comunicación en la Universitat Oberta de Catalunya. @nferranf

(\*) Contenido disponible solo en web