Periodisme de dades

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

1. Introducció: El periodisme, el periodisme d'investigació i el periodisme de dades

1.1. Fiscalitzar el poder

1.1.1. Introducció

1.1.2. Independència

1.1.3. Vigilar el poder

1.2. El periodisme d'investigació

1.3. El periodisme de dades

1.3.1. Presentar les dades al públic

2. Aspectes jurídics del periodisme de dades

Introducció

2.1. El govern obert

2.1.1. Introducció

2.1.2. La transparència pública

2.1.3. La participació i la col·laboració ciutadana

2.2. Llibertat d'informació i d'expressió

2.2.1. Introducció

2.2.2. La llibertat d'informació

2.2.3. La llibertat d'expressió

2.2.4. Límits a la llibertat d'informació i d'expressió

2.3. L'exercici de la professió periodística i l'autorregulació deontològica

Bibliografia

3. Fonts d'informació i mecanismes d'obtenció de dades

3.1. Les fonts periodístiques

3.1.1. Introducció

3.1.2. Tipologia de les fonts d'informació

3.2. Introducció: Les dades com a font d'informació

3.3. Bases de dades disponibles

3.3.1. Introducció

3.3.2. El moviment Open Data

3.3.3. Recull de fonts

3.4. La llei de transparència

3.4.1. Introducció

3.4.2. La llei de transparència a Espanya i a Catalunya

3.4.3. Com formular una petició d'informació

3.5. La mineria de dades

3.5.1. La mineria

3.5.2. Eines

4. Anàlisi de dades

4.1. Eines per a l'anàlisi de dades

4.1.1. Gestors de fulls de càlcul

4.1.2. Eines avançades

4.2. La importància d'ordenar bé un full de càlcul

4.2.1. L'estructura de les dades

4.2.2. Netejar un full de càlcul

4.2.3. Un exemple pràctic

4.3. Eines i mecanismes d'anàlisi

4.3.1. Importar dades

4.3.2. Filtrar i ordenar

4.3.3. Les taules dinàmiques

4.3.4. Filtres condicionals i altres opcions avançades

4.3.5. Utilitzar descriptors estadístics típics

4.4. L'anàlisi visual

4.4.1. Introducció

4.4.2. Un exemple pràctic amb Microsoft Excel

4.5. Com trobar notícies a partir de l'anàlisi

4.5.1. Introducció

4.5.2. L'anàlisi de dades com a element conductor de la notícia: el *Data Driven Journalism*

4.5.3. Les dades com a suport de la notícia

4.5.4. Les dades com a punt de partida d'una investigació

4.5.5. La visualització com a centre de la notícia

5. Redacció i visualitatzació

5.1. La construcció de la notícia: la teoria de les 6w, la piràmide invertida i les entradetes

5.2. Tipologia de peces periodístiques: els gèneres periodístics

5.2.1. Introducció

5.2.2. Gènere informatiu

5.2.3. Gènere interpretatiu

5.2.4. Els gèneres d'opinió

6. Tres exercicis de periodisme de dades

Introducció

6.1. Twitter: captura de dades i anàlisi de xarxes

6.1.1 Introducció

6.1.2. Context

6.1.3. Conceptes bàsics o previs

6.1.4. Preguntes a resoldre
6.1.5. Informació complementària
6.2. Airbnb: anàlisi geoespacial
6.2.1. Introducció
6.2.2. Context
6.2.3. Conceptes bàsics o previs
6.2.4. Preguntes a resoldre
6.2.5. Informació complementària
6.3. Educació pública-privada: obrint dades obertes
6.3.1. Introducció
6.3.2. Context
6.3.3. Conceptes bàsics o previs
6.3.4. Preguntes a resoldre
6.3.5. Informació complementària
Crèdits

1. El periodisme, el periodisme d'investigació i el periodisme de dades

1.1. Fiscalitzar el poder

1.1.1. Introducció

> "El periodisme és allò que algú no vol que se sàpiga, la resta són relacions públiques", George Orwell.

El periodisme consisteix en proporcionar al ciutadà la informació que necessita per a ser lliure i capaç de governar-se a si mateix. Per complir aquest objectiu, ha de ser fidel als 9 elements que estipulava Bill Kovach i Tom Rosenstiel:

1. La primera obligació del periodisme és la veritat
2. Deu lleialtat als ciutadans
3. Tota informació s'ha de verificar
4. Ha de mantenir la seva independència respecte a aquells dels quals informa
5. Ha d'exercir un control independent del poder
6. Ha d'oferir espai al públic per la crítica i el comentari. Interacció i feedback
7. Ha d'esforçar-se perquè el que explica sigui suggerent i rellevant
8. Les informacions han de ser exhaustives i proporcionades
9. Ha de respectar la consciència individual dels seus professionals (la clàusula d'objecció de consciència).

Ser imparcial o neutral no és un principi essencial del periodisme, però sí que ho és que el periodista compleixi amb el seu compromís amb els ciutadans i proporcioni un fòrum públic on impulsar el debat. En aquest apartat, ens centrarem en el quart i cinquè element del llistat de Kovach i Rosenstiel.

1. El periodisme, el periodisme d'investigació i el periodisme de dades

1.1. Fiscalitzar el poder

1.1.2. Independència

> "El procés del periodisme té les seves limitacions. L'escassetat de temps i la freqüent manca de disponibilitat de la informació són dues dificultats endèmiques de la professió, com també ho són els errors que cometen els periodistes quan treballen sotmesos a pressions. Però el bon periodisme també xoca amb limitacions creades pels propis periodistes, així com pels propietaris i els que controlen els mitjans".
>
> David Randall

Tots els propietaris dels mitjans de comunicació fan bandera dels conceptes veritat, honestedat i transparència, però en general utilitzen les plataformes informatives per fer negoci o com a canal de propaganda. Aquesta lògica es fa encara més visible en contextos de convulsió política. La promoció de les pròpies idees i l'exclusió de les idees adverses, la informació esbiaixada i intencional, adaptada a un punt de vista o uns interessos empresarials determinats, són algunes de les males praxis que impregnen la història del periodisme.

En l'actualitat és habitual que els mitjans siguin propietat de grans empreses. Els seus requeriments propagandístics acostumen a limitar-se a donar suport a un determinat partit polític, mimant els polítics que puguin afavorir-los o atacant els que no. També realçant històries còmplices o encobridores dels seus interessos corporatius. La seva prioritat és maximitzar el marge de benefici i fer-ho molt ràpid si el seu deute està sent finançat per agents externs. Les prioritats dels propietaris probablement han imposat més limitacions a la manera com es practica periodisme que qualsevol altre factor en els darrers anys.

Al conjunt d'Europa existeix un problema en aquest sentit. Segons un informe publicat el 2017 pel Centre pel Pluralisme i la Llibertat als Mitjans, de l'Institut Universitari Europeu, les principals amenaces al pluralisme mediàtic són l'excessiva concentració dels mitjans en poques mans, la poca diversitat de punts de vista a les notícies, la manca de transparència sobre els interessos empresarials dels mitjans i una independència editorial vulnerable, subjecte a influències polítiques.

En el cas espanyol, si bé existeix un risc mitjà pel que fa a la manca de pluralisme, un dels resultats més preocupants és la concentració de la propietat dels mitjans espanyols, que presenta un risc del 71%. També és greu el risc relatiu a la independència política (54%). Concretament, la independència dels mitjans públics pel que fa a la governança i al finançament té un risc del 83%, el que vol dir que la ingerència política als mitjans espanyols és habitual. A més, l'estudi també indica que la influència del partit al govern sobre la direcció de les redaccions s'ha multiplicat en els darrers anys. L'autonomia editorial i els intents de controlar políticament els mitjans mostren senyals d'alarma.

1. El periodisme, el periodisme d'investigació i el periodisme de dades

1.1. Fiscalitzar el poder

1.1.3. Vigilar el poder

Aquesta funció es veu amenaçada per l'acumulació empresarial que pot destruir la independència que els mitjans necessiten per exercir el seu paper de vigilants. El principi de vigilància i control va més enllà de la fiscalització a un govern i s'estén a totes les institucions poderoses de la societat. El periodisme ha de vigilar els escassos poderosos d'una societat en representació dels molts que no ho són, per així evitar tiranies. La finalitat de la funció de vigilància no és només aconseguir que l'acció del poder sigui transparent, sinó donar a conèixer les conseqüències de l'acció d'aquest poder i aconseguir que siguin comprensibles. Això implica que els mitjans sàpiguen reconèixer on actua el poder de manera eficaç i on no. Aquesta vocació del periodisme és l'origen del periodisme d'investigació.

En el periodisme d'investigació original són els propis reporters els que busquen, investiguen i documenten una informació que els usuaris desconeixen. Freqüentment, aquest tipus de reportatges ocasionen una investigació pública i oficial dels subjectes o activitats que protagonitzen la informació i constitueixen l'exemple clàssic de la tasca de pressió dels mitjans sobre les autoritats en benefici del ciutadà.

El periodisme d'investigació interpretatiu és el resultat d'una curosa reflexió i anàlisi d'una idea i d'una meticulosa cerca de dades que donen peu a un context informatiu nou i més complet capaç d'oferir al ciutadà una millor comprensió dels fets. En general afecta temes o successos més complexos que els que donen lloc al reportatge clàssic, perquè no només destapa una informació nova sinó que també ofereix una nova versió d'aquesta informació.

La tercera modalitat de periodisme d'investigació és la que es dedica a investigacions que altres instàncies ja han posat en marxa. En aquest cas, el reportatge parteix d'un descobriment o una filtració referents a una investigació oficial que altres, normalment organismes governamentals, ja han començat. El risc d'aquest mètode de recopilar informació és que depèn en gran mesura del rigor i escepticisme del reporter en qüestió. Garanteix als subjectes entrevistats, que en general són fonts oficials, un valuós fòrum en què pot propagar acusacions sense que recaigui en ell cap responsabilitat pública, i aquesta, precisament, és una de les febleses d'aquest tipus de reportatges. Normalment els reporters implicats coneixen només una part de la investigació, de manera que el seu treball, més que vigilar les institucions poderoses es pot convertir en una eina en les seves mans.

Tots els reportatges requereixen certa activitat investigadora, però el que coneixem com a periodisme d'investigació té, a més, certa dimensió moral. Compromet al lector, l'anima a emetre un judici sobre l'assumpte que denuncia i dóna per suposat que el mitjà informatiu que el publica el considera important, mereixedor d'un esforç especial. Un dels problemes que sorgeix amb aquest model de periodisme és que el mitjà informatiu s'implica en una informació sobre la base que hi ha hagut algun acte de mala fe. Per això ha estat denominat com a "periodisme fiscalitzador". Com que pot alterar el curs dels esdeveniments o pot afectar negativament la reputació d'una persona, una empresa o una institució, la tasca del reporter implica una gran responsabilitat, no només a l'hora de verificar els fets sinó també de compartir dades sobre la naturalesa de les fonts informatives.

La revolució de les noves tecnologies i la nova organització econòmica que ha generat creen noves oportunitats, però també amenacen una premsa vigilant i independent. Els mitjans de comunicació digital permeten que la informació viatgi amb major facilitat i rapidesa, però condueixen a la creació de grups de comunicació internacionals que posen en perill el propi concepte d'Estat-Nació. L'experiència ens diu que és molt possible que amb el temps, a mesura que els directius dels informatius formats en aquest nou entorn transnacional vagin assolint quotes de major responsabilitat, als mitjans de comunicació els sembli que ja no val la pena exercir la funció de vigilància i control sobre la seva pròpia empresa matriu. La teoria de la premsa lliure, que deia que era necessària una veu independent i capaç que s'ocupés de vigilar la influència de les institucions més poderoses de la societat, està en perill.

Però més important que el control efectiu que poden exercir els vigilants dels mitjans de comunicació és el fet que les noves tecnologies proporcionen eines que ofereixen la possibilitat a quasi tots els ciutadans de vigilar la premsa consolidada. Les institucions dedicades al periodisme independent indiquen de quina manera les noves tecnologies podrien reorganitzar els mecanismes de producció i difusió de notícies. Potencialment això suposa un desafiament als canals d'informació tradicionals i dóna peu a pensar que si els mitjans de comunicació convencionals abandonen el seu paper de vigilància, altres podrien assumir-lo. Més enllà de la visió utòpica que pot oferir la tecnologia, hi ha qüestions econòmiques més pràctiques que no troben resposta encara. Pot alguna de les noves plataformes o dels nous mitjans malanomenats alternatius o independents, suscitar l'interès d'una part significativa de l'opinió pública? I, en cas afirmatiu, acabaran sent absorbits per les grans corporacions i cauran, per tant, en la cultura empresarial de la qual els seus fundadors volien escapar?

Els departaments d'informatius dels grans mitjans ofereixen el model més clar amb què comptem de la nova organització econòmica de la informació. Van ser les primeres institucions periodístiques en veure's sotmeses a les grans corporacions amb interessos fora del periodisme. Tot i així, a mesura que començaven a patir una pèrdua estructural d'audiència a favor de les noves tecnologies, van anar abandonant la seva funció de vigilància i inclinant-se per un model informatiu més proper a l'entreteniment.

Aquest model, que és l'única evidència que tenim dels efectes de l'assimilació global dels mitjans informatius, desperta seriosos dubtes de que el periodisme d'investigació es mantingui al nivell que va assolir la segona meitat del segle xx.

La funció de vigilància i control és com qualsevol altra. Requereix de facultats especials, un temperament especial. A més, exigeix un compromís seriós amb les fonts, el desig d'ocupar-se d'assumptes d'importància i una premsa independent de qualsevol interès que no sigui el del consumidor final de la notícia. El principi de vigilància i control s'enfronta avui a més reptes que mai però, a la vegada, la naturalesa expansiva del periodisme com a fòrum públic ha provocat una nova onada de periodisme interpretatiu i opinatiu que fa més crítica la necessitat d'un periodisme vigilant més seriós i actiu. En el segle xxi, els mitjans han d'exercir vigilància no només sobre el govern sinó també i especialment sobre el món empresarial i l'àmbit cada vegada major d'activitats sense ànim de lucre.

1. El periodisme, el periodisme d'investigació i el periodisme de dades

1.2. El periodisme d'investigació

> "La raó més habitual del fracàs d'una història no és una mala redacció, les cites manipulades o una pobre elaboració, sinó una investigació inadequada".
>
> David Randall.

El periodisme de investigació té uns quants trets diferencials:

1) Originalitat de la investigació

El periodisme consisteix en realitzar investigacions originals, moltes vegades fent servir materials en brut. Es pot fer mitjançant una amplia sèrie d'entrevistes o bé per mitjà de xifres. Molts dels reportatges són conseqüència d'un treball meticulós del periodista analitzant la connexió entre dades que no s'havien investigat prèviament

2) Informacions relacionades amb negligències o irregularitats de les que no es tenen proves

Dedicar-hi temps si es tracta d'una qüestió d'interès pels lectors, assumint que potser la investigació no és fructífera i que "no hi ha tema". El periodisme d'investigació comença amb la sospita que en un tema determinat hi ha les senyals d'una bona història. El primer que cal fer és posar a la balança si la història mereix l'esforç i el temps requerits per tirar-la endavant. Aquesta classe de peces es poden realitzar sobre pràcticament qualsevol àmbit de la vida pública, però hi ha dues categories que són especialment fecundes: les activitats que estan allunyades de l'escrutini del públic i les persones o entitats amb rellevància pública.

Les empreses i entitats financeres, que generalment han estat poc fiscalitzades, són un terreny molt fèrtil per la indagació periodística.

> "Si posem la lupa sobre un pla d'inversions recent llançat amb un gran desplegament publicitari, podem apostar-hi el sou a que traurem a la llum alguna informació confidencial".

El mateix passa amb alguns organismes, institucions i empreses que desenvolupen una activitat que ha estat molt poc controlada per part dels mitjans de comunicació.

Consorci de Turisme de Barcelona

Un exemple d'això és el Consorci de Turisme de Barcelona, una entitat pública-privada el finançament de la qual havia passat molt desapercebut. Fa un parell d'anys es va situar de ple sota els focus mediàtics arran una acció reivindicativa que va convocar l'Assemblea de Barris per un Turisme Sostenible, on precisament denunciava l'opacitat d'aquest consorci. Això va fer que alguns mitjans comencessin a investigar què era exactament aquesta entitat, qui la controlava i com es finançava. Un altre cas similar ha estat el de l'empresa pública Barcelona Regional, constituïda l'any 1993 a partir de l'experiència acumulada amb la transformació olímpica. Durant anys havia operat lliurement sense ser fiscalitzada, però ara està a l'ull de l'huracà arran que una denuncia anònima presentada a la Fiscalia Provincial revelés que els gestors de l'agència havien malversat els seus fons en el període de mandat convergent. Arran d'aquesta acusació, els mitjans de comunicació van iniciar una investigació paral·lela a la dels Tribunals, l'Ajuntament i el propi Consell d'administració de Barcelona Regional. A partir d'aquí han començat a sortir a la llum contractes irregulars, subvencions injustificades i vincles injustificats amb exrepresentants públics.

Hi ha uns quants **factors que poden simplificar la tasca i millorar l'eficàcia** del reporter:

a) Conèixer la legislació sobre l'accés públic a la informació

A Catalunya i l'Estat espanyol, la Llei de transparència (una eina que encara s'utilitza poc a la majoria de mitjans) permet que els periodistes, i en realitat qualsevol ciutadà, pugui sol·licitar informació pública a l'administració. Hi ha moltes dades públiques (que no publicades) que estan a l'abast de qualsevol persona sempre i quan conegui els procediments per requerir la informació.

Les sol·licituds no sempre són fructíferes, en el sentit que hi ha organismes opacs que no volen oferir informació sobre el seu funcionament, sous de directius, pressupostos, ingressos, etc, i que procuraran bloquejar-nos l'accés. En aquests casos, a

Catalunya, els periodistes podem jugar la carta de la GAIP (*).

Si una institució ens denega una informació acollint-se, per exemple, a la llei de protecció de dades, sempre podem fer una reclamació a aquesta comissió, justificant que, al nostre entendre, el dret a la informació passa per sobre de la negativa de la institució en qüestió. És un recurs que els mitjans de comunicació comencen a utilitzar.

Exemple

Per posar un exemple, La Directa i el CRÍTIC van acollir-se al dret d'accés a la informació pública per saber els diners que havia gastat el Departament d'Interior en armament, munició i altre equipament per als Mossos d'Esquadra. La Direcció General de la Policia va denegar inicialment aquesta informació, però els dos mitjans van recórrer aquesta decisió a la GAIP, que va donar-los la raó i va forçar el Departament a facilitar aquesta informació.

Algunes investigacions han sorgit de documents entregats als periodistes però moltes més s'han posat en marxa quan un reporter ha descobert l'existència de registres i documents a la consulta dels quals hi tenia accés. En general els organismes burocràtics no estaran desitjant donar publicitat a l'existència de determinades fonts de informació i fins i tot, en alguns casos, miraran d'impedir que la gent es prengui la molèstia de consultar-les. És important que els periodistes descobreixin els documents oficials que existeixen, els examinin i els utilitzin en les seves investigacions. També és recomanable fer ús de les fonts de re referència estandarditzades: els llistats de publicacions oficials, els informes legislatius, les llistes d'organismes públics, els registres mercantils o els registres de les organitzacions que reben finançament públic, etc.

Una bona llista de contactes és transcendental a l'hora de fer periodisme de investigació. Són útils per cobrir una amplia gama de temes, orientar les nostres informacions i facilitar-nos la informació. Diversificar-los: tenir fonts només de l'àmbit institucional farà que sempre oferim la versió oficial dels fets. En aquest sentit és imprescindible obrir el focus: què hi tenen a dir els moviments socials? Les plataformes ciutadanes? El moviment veïnal?

Hi ha una dita històrica segons la qual els periodistes som experts en tot i en res. Però el cert és que no som homes ni dones orquestra i els nostres coneixements són limitats. A l'hora de fer alguns reportatges ens podem trobar fàcilment que la temàtica que hem seleccionat s'escapa àmpliament de la nostra zona de confort. Sempre està bé tenir formació en una especialitat concreta a l'hora de cobrir temes d'economia, societat o qüestions jurídiques, però no podem saber-ne de tot. Per això, esdevé especialment important tenir a l'abast fonts especialitzades, que no tinguin interessos específics en el tema al qual donem cobertura, que ens ajudin a entendre i processar millor la informació. En aquest sentit, és especialment útil disposar del telèfon de sociòlegs, juristes, politòlegs o economistes "independents", assumint que sempre tindran la seva visió subjectiva de les dades però que no tinguin una relació directa amb l'objecte d'investigació. Si volem parlar del judici contra una farmacèutica no demanarem a l'advocat de l'empresa en qüestió que ens analitzi la seva validesa!

b) Coneixements d'informàtica (i més)

No només (però també) referida a l'habilitat de rastrejar online amb eficàcia, sinó també a la capacitat de recórrer al software de bases de dades. La informàtica pot modificar la capacitat del periodisme d'investigació, perquè té la capacitat d'anar més enllà de les tradicionals anècdotes i entrevistes i aglutinar una quantitat molt àmplia de proves documentals. [Ampliació a l'apartat "Periodisme de dades"].

Com podem obtenir bon material de cara a una investigació?

En primer lloc, saber què hem de buscar: més enllà de respondre les 5w, ens calen detalls, anècdotes i context. Aquesta informació addicional és el que marca la diferència entre una història ordinària i una bona història.

- Recol·lecció de detalls: important construir una cronologia exhaustiva del que ha passat. No hi ha detall tan insignificant que no valgui la pena recollir, ja que fins i tot els fragments diminuts poden afegir pes a la historia molt per sota del seu calor aparent. En certa manera, el detall és la història en molts casos.

- Anècdotes: Incidents o episodis que puguin resumir-se en una o dues frases.

- El teló de fons: l'entorn, el context i les parts rellevants del tema.

- Perspectiva: el context a vegades pot ser la part vital de la historia, perquè posa els fets o esdeveniments en una perspectiva apropiada.

- Recelar de les fonts i de les pàgines web: s'ha de desenvolupar un radar capaç de detectar els llocs que no són de confiança, plantejant les mateixes preguntes que ens plantejaríem amb una font (qui hi ha darrera? Per què ha publicat aquesta

informació? Què hi inclou? Què omet?).

- Ser específics amb la cerca: mètodes de recerca online com posar cometes en el concepte que es busca, posar els ítems AND o OR. Guies per la recerca online: com searchenginewatch.com o researchbuzz.com.

- Fer servir l'aplicació de cerca avançada, que permet afinar la cerca amb paràmetres o dates.

- Subscriure's a arxius online: per exemple el Google Alerts.

- Fer ús de les hemeroteques.

Com desenvolupar una investigació?

> "Els temes de les investigacions arriben als diaris per les vies més diverses: a partir de les informacions dels contactes, per pur atzar, a partir d'una notícia en principi rutinària i que informacions posteriors n'amplien la magnitud, gràcies a les observacions dels reporters, a partir d'informacions corrents que van complicant-se o en les que es descobreixen interrogants cada vegada més interessants quan comencem a indagar."

Alguns consells:

- Aconseguir i arxivar tots els documents pertinents: Encara que no els fem servir en el moment de rebre'ls, cal guardar una documentació tan ingent com sigui possible i no desprendre's d'ella, perquè mai se sap quan podran ser útils. Mesos després de que hagi arribat a les nostres mans un informe d'aparença innòcua, pot passar alguna cosa que li atorgui una rellevància inesperada.

- Transcriure totes les entrevistes i arxivar-les: Això és particularment important quan el procés de investigació és llarg o quan hi col·laboren diversos reporters. Val la pena intercanviar notes per comprovar si s'ha passat per alt alguna dada d'importància.

- Ser persistents: Quan es llegeix un relat sobre qualsevol investigació el que crida més l'atenció és la persistència dels reporters.

- Cultivar el tracte amb les fonts més ben informades: Si una font estableix certes regles, cal complir-les, però això no vol dir que s'hagi de donar crèdit a totes les seves paraules. S'han de respectar els acords.

- Les informacions que finalment arribin a publicar-se no han de tenir cap punt dèbil: Publicar un article per l'únic motiu que se li ha dedicat molt temps és la millor forma d'abocar-se al desastre. Els directius i reporters hauran d'adoptar en una fase avançada de la investigació la decisió d'anar-la publicant per entregues a mesura que es realitzo o, per contra, reservar les troballes per donar un gran cop d'efecte al final del procés. En aquest segon cas és necessari establir una data tope. D'altra banda, publicar conclusions provisionals pot incitar les fonts a proporcionar més informació o més dades de vital importància.

- Actuar clandestinament: En algunes ocasions excepcionals la única manera d'aconseguir la informació és de forma clandestina, per exemple, quan estem investigant un delicte. Aquesta forma de treballar implica sempre un cert grau d'engany, així que el delicte en qüestió ha de ser suficientment seriós com per a justificar aquesta falta d'honradesa. En segon lloc, els riscos físics en tal situació poden ser immensos i perdurar fins i tot temps després de publicar la nostra història. En tercer lloc, si estem investigant activitats criminals sota una identitat encoberta, podem veure'ns obligats a participar en elles, el que fa que els nostres actes resultin encara més difícils, si no impossibles de defensar. A vegades els periodistes han fet servir falses identitats amb el propòsit fonamental de deixar al descobert el tracte o maltracte rebut per diverses víctimes de la societat, com els sense sostre, els malalts psíquics o els drogoaddictes. Aquesta tàctica requereix dots interpretatives i a vegades, fins i tot una disfressa.

1. El periodisme, el periodisme d'investigació i el periodisme de dades

1.3. El periodisme de dades

1.3.1. Presentar les dades al públic

1.3.1.1. Visualitzar les dades o no

Hi ha ocasions en que les dades poden explicar una història millor que les paraules o les fotografies. Per això, termes com "aplicació de notícies" i "visualització de dades" han adquirit l'estatus de paraules clau en tantes redaccions els últims temps. També promou l'interès la gran recapta de noves eines i tecnologies (sovint, gratuïtes) destinades a ajudar fins i tot al periodista menys dotat tècnicament a convertir dades en una presentació visual de la història.

Eines om Google Fusion Tables, Many Eyes, Tableau, Dipty i altres fan més fàcil que mai crear mapes, quadres, gràfics i fins i tot aplicacions amb dades que fins ara eren domini d'especialistes. La qüestió ara pels periodistes no és tan si poden convertir els seus conjunts de dades en una visualització sinó si els convé fer-ho. Una mala visualització de les dades és pitjor en molts sentits que cap visualització.

1) Utilització dels gràfics amb moviment

Amb un guió ajustat, animacions ben cronometrades i explicacions clares, els gràfics amb moviment poden donar vida a nombres o idees complexes, orientant el seu públic. Els recursos visuals explicatius amb la veu en off ofereixen una manera poderosa i memorable d'explicar una història.

En la majoria d'ocasions el nostre flux de treball comença amb l'Excel. És una manera fàcil de descobrir si hi ha alguna cosa interessant a les dades. Si és així, hem de buscar la manera de visualitzar-les i després escriurem el text que ho acompanyi. El text ha de servir per analitzar i aportar context a les dades, que no es quedi amb una simple visualització de la relació entre diverses variables: el que no ens digui la infografia, ho hem d'explicar nosaltres.

2) Publicar les dades

Alguns consells a l'hora de publicar les dades és utilitzant una visualització que permeti la descàrrega fàcil del conjunt de dades. Els lectors poden explorar-les darrera les històries interactuant amb les visualitzacions o fent servir les dades d'altres maneres. Això és important perquè augmenta la transparència i permet el feedback del lector.

3) Donar accés a les dades

Així els lectors poden assegurar-se que no manipulem les dades per arribar a conclusions forçades. A banda, permetre l'accés a les dades fa que hi hagi investigadors que puguin reutilitzar-los.

4) Humanitzar les dades

Les dades estan associades a la vida de gent real, no són nombres dissaciats flotant en el buit, sinó medicions de coses tangibles. Quan els abordem hem de considerar els sistemes del món real d'on provenen. Els periodistes tenen un rol molt important en treure a la llum aquesta humanitat inherent a les dades. Al fer-ho, tenen el poder de canviar la comprensió del públic tan de les dades com dels sistemes dels que van emergir els números.

1.3.1.2. La visualització en la recerca periodística

En la fase de buscar informació, les visualitzacions poden:

- Ajudar-nos a identificar temes.
- Excloure allò que no ens interessa.
- Identificar errors en les dades.
- Trobar exemples típics.

En la fase d'edició, podem:

- Il·lustrar un argument de una història d'una manera més convincent.

- Treure informació tècnica innecessària de la prosa.

- En particular, quan són interactius i permeten l'exploració, ofereixen transparència respecte al procés d'informació.

Consells sobre l'ús de visualitzacions:

- Fer servir petits múltiples per orientar-se ràpidament en un conjunt de dades gran. Els mapes amb petits múltiples poden ser una manera de mostrar temps i lloc en una imatge estàtica que és fàcil de comparar, a vegades més fàcil que la versió interactiva.

- Mirar les dades, remirar-les i tornar-les a mirar.

- utilitzar xifres crues, percentatges i índexs.

- Pensar com combinar les dades i si ens interessa fer-ho. Per exemple, si parlem de criminalitat, podem fer una infografia establint diferents paràmetres i els podem visualitzar tots a la vegada o un a un, fent servir el mateix patró gràfic però refrescant les variables cada vegada que punxem un ítem diferent.

1) No donar res per suposat

Quan analitzem les dades ens podem trobar amb registres que no semblen correctes, o bé perquè no entenem què signifiquen o bé per errors informàtics, de transcripció o del tipus que sigui. Si volem publicar alguna cosa basada en aquestes dades el primer que cal fer és resoldre les qüestions que no ens quedn clares. Si nosaltres no som capaços d'entendre el que expliquem, portarem a conclusions erronies als nostres lectors.

No és inusual que els governs de les diferents administracions donin plantilles de càlcul plenes d'errors i és també fàcil confondres amb els termes oficials que es fan servir per denominar una partida pressupostària, o les sigles que s'utilitzen per fer referència a un cos intern, o un departament, o el que sigui. El primer que hem de fer és tornar a mirar-nos la documentació i veure si, efectivament, existeix un problema en la versió original de les dades. Si tot el que hem fet nosaltres està bé, llavors és el moment de fer trucades. No tots els errors són importants. En els registres de campanyes electorals és comú que hi hagi diversos centenars de codis postals que no existen en una base de dades de 100.000 registres. Sempre que no siguin tots en la mateixa ciutat o estiguin relacionats amb un mateix candidat, el registre ocasional equivocat simplement no importa. La pregunta que ens hem de fer és: "Si utilitzem aquesta informació, els lectors tindrien una visió encertada del que diuen les dades?"

2) No obsesionar-se amb la precisió

La contracara de no fer prou preguntes és obsessionar-se amb la precisió. Els gràfics exploratoris han de ser correctes però no convé preocupar-se si no sumen exactament el 100% o si ens falta alguna informació per a tenir tot el relat complet, perquè aquest gràfic ha de servir per mostrar-nos tendències generals i què és el que hem de buscar per estructurar la nostra història.

3) Cronologies

Fer servir Excel, un document word o eines especials com el TimeFlow.

4) Publicació de les dades

L'esforç de recolecció de les dades ha de coincidir amb el que requereix el gràfic interactiu:Si volem que els lectors puguin explorar les dades de la manera que vulguin, hem de preveure la correcció d'errors, el control de dades i l'edició del text.

5) Dissenyar pensant en dos tipus de lectors

Una gràfica (sigui interactiva o un element estàtic que acompanya el nostre text) ha de ser fàcil d'entendre a simple vista però ha d'aportar prou elements com per a que resulti interresant per al lector més exigent.Si apostem per una visualització interactiva, ens hem d'assegurar que els lectors obtinguin alguna cosa més que una sola xifra i que puguin, per exemple, creuar dades, establir correlacions per arribar a les seves propies conclusions.

6) Transmetre una idea

Assegurar-nos, en primer lloc, que tenim una informació que volem que la gent vegi i després, decidir quina idea volem destacar al gràfic. En molts casos això vol dir eliminar informació encara que Internet permeti que la posem tota. A menys que el nostre principal objectiu sigui la transparència de la nostra activitat periodística, la major part dels detalls que hem recollit a la cronolgia no són gaire rellevants i poden resultar aborrits o confusos. De fet, si volem mostrar totes les dades i explicar al lector com les hem obtingut, hi ha altres maneres de fer-ho que no impliquen saturar-lo amb una visualització que reculli masses ítems. Les xarxes socials són una bona eina per a fer aquest exercici de transparència: podem crear grups al Facebook, per exemple, on fem posts donant constància del procés de recollida de la informació i fins i tot facilitar totes les taules "brutes" que continguin aquella informació que hem deixat fora de la visualització.

1.3.1.3. Les visualitzacions per a narrar històries

La visualització de les dades no només pot ser cridanera i atraure l'atenció del lector, també aprofita el fet que la meitat del cervell humà està dedicat a processar informació visual. Una visualització de dades ben dissenyada pot oferir als que la veuen una impressió immediata i profunda i anar al gra de la qüestió sense enredar-se. Però a diferència d'altres mitjans visuals com la fotografia i el vídeo (eines que, per cert, no hem de descartar fer servir de manera complementària al nostre gràfic i text), la visualització de dades també mostra fets mesurables.

El periodisme de dades en l'era dels mitjans fets a mida de públics amb punts de vista particulars, ofereix l'oportunitat de narrar històries orientades principalment per fets. De la mateixa manera que altres formes de periodisme narratiu, la visualització de dades pot ser efectiva tant per presentar notícies noves –transmetent ràpidament nova informació com la ubicació d'un accident o el nombre de víctimes– com articles de fons, on es pot aprofundir en un tema i oferir una nova perspectiva.

Exemple

Un exemple seria el gràfic interactiu publicat pel *New York Times* a finals del 2009, que permetia als lectors analitzar la taxa de desocupació per grups de població als Estats Units per mitjà de filtres demogràfics i educatius. Una bona visualització es queda al cap del lector i deixa un model mental durador d'un fet, una tendència o un procés.

1.3.1.4. Exemples en què la visualització de dades pot ser útil

1) Mostrar el canvi al llarg del temps: Per exemple, el creixement de la població xinesa des de 1960 és un bon exemple. Però les visualitzacions de dades també poden mostrar de manera molt poderosa el canvi al llarg del temps a través d'altres formes gràfiques. Per exemple, la participació en el mercat de les principals aerolínies nord-americanes durant diverses dècades de concentració en el sector.

2) Comparar valors: La visualització de dades també és útil quan es tracta d'ajudar els lectors a comparar dos o més valors discrets, com per exemple quan el National Geographic, fent servir un quadre molt minimalista, va mostrar quant majors són les probabilitats de morir de malaltia coronaria o infart que en accidents d'aviació o per picades d'avella. Així mostrava les probabilitats relatives de diferents causes de mort.

3) Dissenyar amb gràfiques: Els diagrames de flux també codifiquen informació en les línies de connexió, generalment d'acord al gruix i/o al color de les mateixes. Per exemple, amb l'Eurozona en crisi i diversos membres incapacitats per fer front al deute, The New York Times va buscar donar una explicació al conjunt de deutes que vincula els membres de la UE amb els seus socis comercials a l'altre costat de l'Atlàntic i a Àsia.

4) Explorar grans bases de dades: A vegades la visualització de dades és molt efectiva per mostrar informació familiar des d'un angle totalment nou, però què succeeix quan es té informació nova que la gent vol buscar? La era de les dades porta a descobriments nous sorprenents quasi tots els dies, com l'anàlisi de fotos de Flickr d'Eric Fischer. Aquestes bases de dades són encara més poderoses quan els usuaris poden explorar la informació i arribar allà on els resulti més rellevant.

Bases de dades interactives

Un exemple és la base de dades interactiva del *The New York Times* que permet als usuaris explorar les 100 pel·lícules més llogades en 12 zones metropolitanes dels Estats Units. Un altre és el de La BBC, que per la seva banda, va fer una investigació dels morts a les carreteres de Gran Bretanya entre 1999 i 2010.

5) Quan no hauríem de fer servir la visualització de dades: La visualització de dades efectiva depèn d'explicar amb informació neta, bona, precisa i significativa. Així com moltes cites, dades i descripcions alimenten el bon periodisme narratiu, la visualització de dades és tan bona com les dades que l'alimenten.

A vegades les dades per si soles no narren una historia de la millor manera possible. Per bé que un quadre simple que il·lustri una tendència o una estadística pot ser útil, una narrativa que relati les conseqüències d'una qüestió pot tenir més impacte pel lector.

- **Quan tenim poques dades:** Si les dades varien poc i no mostren una conclusió clara. Potser en aquest punt ens adonem que, o bé ens cal més investigació i anàlisi que ens permeti fer una bona visualització o bé és millor publicar-les en format escrit, i apuntar cap a una tendència determinada tot alertant que no és definitiva.

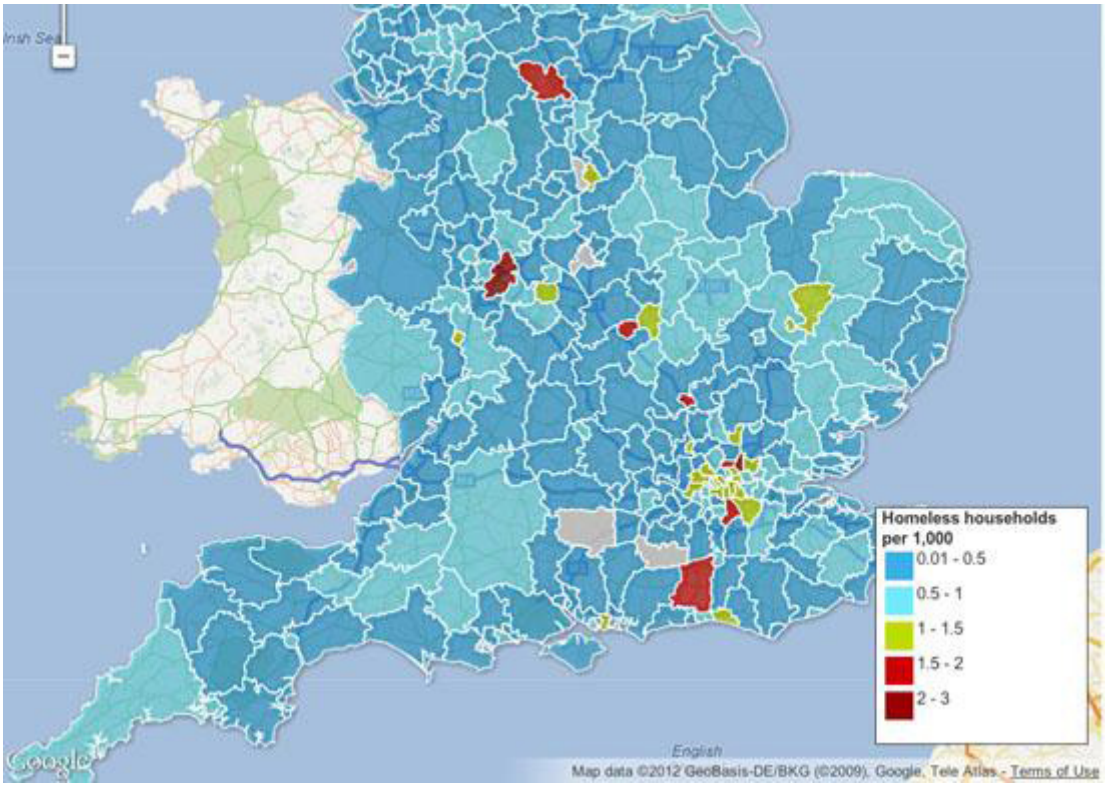
- **Quan amb una taula n'hi ha prou:** És net, de ràpida lectura i no crea expectatives no realistes d'una història. Suposa una forma ràpida i fàcil de presentar dades senzilles.

1.3.1.5. Eines per a fer les nostres pròpies visualitzacions de dades

Està clar que hi ha equips gràfics i interactius increïbles per a fer grans investigacions, però per la tasca diària, sovint n'hi ha prou amb eines a les quals tothom hi té accés. Aquestes són algunes de les més comunes:

1) Google Fusion Tables: Aquesta base de dades i eina de mapeig online ha esdevingut la primera elecció per produir mapes ràpids i detallats, especialment aquells que requereixen zoom. El fem servir per produir mapes com el d'Irak o mapes de fronteres sobre els sense sostre que adjuntem tot seguit.

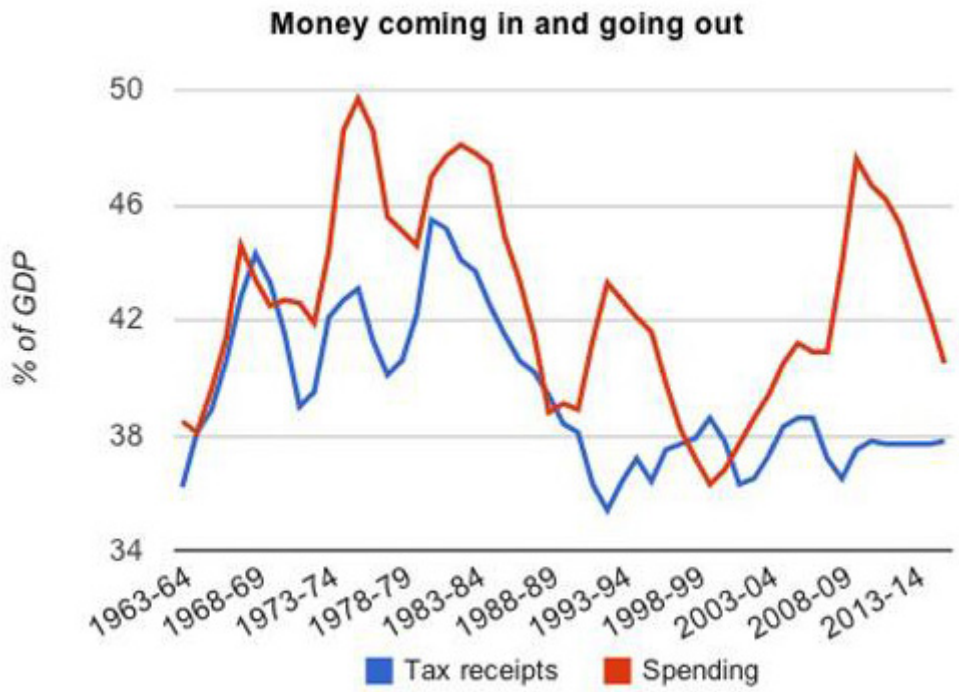
Els registres de guerra de WikiLeaks (The Guardian)



Mapa interactiu de persones sense llar (The Guardian)

2) Tableau Public: Es fa servir sobretot quan hem d'unir diferents tipus de quadres, com en aquest mapa de taxes impositives màximes a tot el món, o com a explorador de dades. La Nación, un mitjà d'Argentina, ha construit tota la seva operació de periodisme de dades amb el Tableau.

3) Gràfics amb Google Spreadsheets: Poden generar alguns gràfics com les bombolles animades que fa servir el Gapminder. És similar a fer gràfics amb l'Excel, en el sentit que se seleccionen les dades i es fa clic al widget de gràfics. També val la pena explorar les opcions de personalització: es pot canviar el color, les escales i els encapçalaments. És útil, sobretot, per a gràfics petits i els gràfics de corbes inclouen, a més, opcions per a fer-hi anotacions.



Despesa pública i impostos del Regne Unit (The Guardian)

4) Datamarket: Eina pràctica per visualitzar xifres i, sobretot, dades de séries temporals.

5) Chartsbin: És una eina per crear mapamundis en els quals es pot fer clic.

6) iCharts: Especialitzat en petits widgets de gràfics.

7) Geoccomons: Ofereix dades de fronteres per crear mapes globals i locals.

8) Pikctochart.com: Ofereix plantilles per aquestes visualitzacions de text i xifres.

9) Infogram: És una eina per visualitzar dades simples. Ofereix gràfics lineals, de barres i circulars, i per fer-los servir n'hi ha prou amb introduir el seguit de dades que volem representar gràficament en l'eix de les files i de les columnes.

2. Aspectes jurídics del periodisme de dades

Introducció

En aquest apartat es fa una aproximació als principals aspectes jurídics que cal tenir en compte quan es duu a terme periodisme de dades.

El periodisme de dades està estretament vinculat a l'accessibilitat i a la disponibilitat de dades que produeixen les administracions públiques, i que difonen en el marc de les polítiques de govern obert i de les lleis de transparència. A mesura que les administracions públiques s'obren a la ciutadania i expliquen millor el que fan, els periodistes també tenen més facilitat per a accedir a dades que són una font important a l'hora de dur a terme l'activitat professional.

Des d'aquesta perspectiva, la principal qüestió que es tracta en aquest apartat és la vinculada a l'accés a les dades que produeixen les administracions públiques, que, en l'actualitat, constitueix una de les principals fonts utilitzades en el periodisme de dades. La legislació vigent preveu diferents mecanismes pels quals els periodistes poden accedir a les dades de les administracions públiques, i determina uns límits que seran analitzats.

També es farà referència a altres qüestions que emmarquen l'activitat dels periodistes de dades, com són el fonament de l'activitat periodística basat en les llibertats d'informació i d'expressió, i les normes que regeixen l'activitat periodística.

Els objectius són els següents:

1. Identificar les normes que regulen la difusió, l'accés i la reutilització de les dades que produeixen les administracions públiques.

2. Conèixer els fonaments de la llibertat d'expressió i la llibertat d'informació, i els seus límits.

3. Contextualitzar les normes que regulen l'exercici de la professió periodística.

2. Aspectes jurídics del periodisme de dades

2.1. El govern obert

2.1.1. Introducció

El govern obert fa referència als mecanismes que permeten que les administracions públiques i la ciutadania puguin treballar plegats en la generació de valor públic (*).

El govern obert persegueix enfortir les relacions entre les administracions públiques i la ciutadania mitjançant la transparència de les administracions públiques i la participació i la col·laboració ciutadana amb l'ús intensiu dels mitjans electrònics. Per diferents mecanismes, les administracions públiques informen la ciutadania sobre la seva activitat i n'escolten les preferències i les opinions implicant-la en el disseny i la implementació de les polítiques públiques. El govern obert es basa en l'ús intensiu de la tecnologia i implica l'impuls de la innovació en les administracions públiques.

Així, doncs, el govern obert és transparent, és a dir, facilita a la ciutadania informació sobre les activitats que duen a terme les administracions públiques, les decisions que prenen i com les prenen, i com es gestionen els recursos públics.

El govern obert (*) també és participatiu, és a dir, escolta les necessitats que expressa la ciutadania i en facilita la implicació en la presa de decisions públiques per mitjà d'espais que permetin el diàleg i la deliberació.

La legislació vigent recull els principis de transparència i de participació. Algunes normes fins i tot fan referència al govern obert i preveuen mecanismes específics. Un exemple el trobem en la Llei 19/2014, de 29 de desembre, de transparència, accés a la informació pública i bon govern de Catalunya, que defineix el govern obert (*) com «les mesures per a establir una relació i un diàleg permanents i bidireccionals entre l'Administració i les persones a l'hora de definir i d'aplicar les polítiques públiques, i per a introduir i desenvolupar instruments de participació i de col·laboració ciutadana en els assumptes públics».

2. Aspectes jurídics del periodisme de dades

2.1. El govern obert

2.1.2. La transparència pública

2.1.2.1. Introducció

La transparència és el primer pilar del govern obert, i persegueix que la ciutadania pugui tenir un coneixement efectiu de l'activitat de les administracions públiques.

La transparència facilita que la ciutadania pugui participar en les decisions públiques, i alhora és un mecanisme mitjançant el qual les administracions públiques poden retre comptes a la ciutadania perquè pugui controlar com gestionen els recursos públics o com prenen les decisions. La transparència és un mecanisme que pot enfortir la integritat pública i esdevenir un fre a les irregularitats i a la corrupció. En última instància, la transparència permet incrementar la confiança pública i l'afecció ciutadana per les institucions públiques, la qual cosa ha d'enfortir-ne la legitimitat.

En aquest sentit s'expressa el preàmbul de la Llei 19/2013, de 9 de desembre, de transparència, accés a la informació pública i bon govern (LTAIBG):

> «Només quan l'acció dels responsables públics se sotmet a escrutini, quan els ciutadans poden conèixer com es prenen les decisions que els afecten, com es gestionen els fons públics, o sota quins criteris actuen les nostres institucions, podem parlar de l'inici d'un procés en què els poders públics comencen a respondre una societat que és crítica, exigent i que demana participació dels poders públics.»

La transparència és un principi jurídic recollit en diverses normes. Amb caràcter general, podem referir-nos a la Llei 40/2015, d'1 d'octubre, de règim jurídic del sector públic, que recull, entre els principis generals, el principi de transparència de l'actuació administrativa (art. 3). A més, altres normes preveuen el principi de transparència com a principi que ha de guiar l'activitat dels càrrecs i dels treballadors públics. En aquest sentit, l'LTAIBG estableix com a principi de bon govern dels alts càrrecs que actuïn amb transparència en la gestió dels afers públics. Igualment, el Reial decret legislatiu 5/2015, de 30 d'octubre, pel qual s'aprova el text refós de la Llei de l'Estatut bàsic de l'empleat públic, també recull la transparència entre els principis que guien l'activitat dels treballadors públics.

L'LTAIBG i, en general, les normes autonòmiques de transparència defineixen i regulen diferents mecanismes per mitjà dels quals es canalitza la transparència (*):

- la difusió d'informació pública (transparència activa),
- l'accés a la informació pública (transparència passiva), i
- la reutilització de la informació pública (transparència col·laborativa).

2.1.2.2. La difusió d'informació pública

La difusió d'informació pública consisteix a posar a disposició de la ciutadania la informació per part de les administracions públiques.

Les lleis de transparència preveuen que les administracions públiques han de difondre la informació per mitjà de pàgines web i seus electròniques. La majoria d'aquestes lleis han previst que la informació s'ha de difondre per un portal de transparència.

Les lleis de transparència també defineixen quina informació han de difondre les administracions públiques pels portals de transparència. Si agafem com a exemple l'LTAIBG, les administracions públiques han de difondre la següent informació:

- Informació institucional, organitzativa i de planificació. Les administracions públiques han de difondre informació sobre les funcions que duen a terme, la normativa que els sigui aplicable, així com la seva estructura organitzativa. També han de publicar informació relativa als plans i als programes anuals i plurianuals.
- Informació de rellevància jurídica. Les administracions públiques han de publicar, entre d'altres informacions de rellevància jurídica, les directrius, les instruccions, els acords, les circulars; els avantprojectes de llei i els projectes de decrets legislatius i de reglaments, o les memòries i els informes que conformin els expedients d'elaboració dels textos normatius.

- Informació econòmica, pressupostària i estadística. Les administracions públiques han de difondre la informació sobre tots els contractes, els convenis subscrits, les subvencions i les ajudes públiques concedides, els pressupostos, els comptes anuals, les retribucions percebudes pels alts càrrecs, les resolucions d'autorització o de reconeixement de comptabilitat que afectin els treballadors públics; i les declaracions anuals de bens i activitats dels representants locals, o la informació relativa al grau de compliment i qualitat dels serveis públics.

Cal entendre aquesta informació com un mínim, que pot ser ampliat per cada administració pública (art. 5.2 LTAIBG). De fet, les lleis autonòmiques de transparència han ampliat significativament la informació que han de difondre les administracions públiques (*).

Índex de transparència

En els darrers anys s'han impulsat diversos índexs per a mesurar el nivell de transparència de les administracions públiques.

Transparencia Internacional España elabora l'Índex de transparència de les Comunitats Autònomes (INCAU), que valora el nivell de transparència de les administracions autonòmiques i el compliment de les seves obligacions, previstes en la legislació bàsica de transparència. Aquesta ONG també elabora l'Índex de transparència dels Ajuntaments, que mesura la transparència dels 110 ajuntaments més grans d'Espanya (ITA).

El Laboratori de Periodisme i Comunicació per a la Ciutadania Plural, de la Universitat Autònoma de Barcelona, elabora el mapa InfoParticipa, per mitjà del qual s'avalua el nivell de transparència dels ajuntaments.

La informació que difonen les administracions públiques pels portals de transparència ha de ser de qualitat, per tal que pugui acomplir les finalitats previstes. D'acord amb la legislació vigent (art. 5 LTAIBG) la informació que es difongui ha d'ésser:

- clara, estructurada i entenedora per als interessats,
- en formats reutilitzables,
- accessible d'acord amb el principi d'accessibilitat universal i disseny per a tothom,
- interoperable,
- identificable i localitzable,
- comprensible, i
- d'accés fàcil i gratuït.

2.1.2.3. L'accés a la informació pública

L'accés a la informació consisteix en la sol·licitud que fa una persona a una administració pública perquè mostrin una determinada informació o en facilitin una còpia. La legislació vigent reconeix el dret d'accés com un dret subjectiu (*) (art. 12 LTAIBG).

D'acord amb l'LTAIBG (art. 13), «s'entenen per informació pública els continguts o els documents, sigui quin en sigui el format o suport, que estiguin en poder d'algun dels subjectes inclosos en l'àmbit d'aplicació d'aquest títol, i que hagin estat elaborats o adquirits en l'exercici de les seves funcions».

Qualsevol persona pot demanar l'accés a la informació pública mitjançant una sol·licitud, que cal adreçar al titular de l'òrgan administratiu o entitat que tingui la informació, i que es pot presentar per qualsevol mitjà que permeti tenir constància de la identitat del sol·licitant (inclòs l'electrònic amb l'ús de la signatura electrònica), de la informació que sol·licita, de l'adreça de contacte –preferentment electrònica– i de la modalitat que prefereixi per a accedir a la informació. El sol·licitant no està obligat a motivar-ne la sol·licitud, tot i que pot exposar els motius pels quals demana la informació, però l'absència de motiu no podrà ser causa de rebuig de la sol·licitud.

L'accés a la informació de l'Administració de la Generalitat

L'anàlisi del període comprès entre l'1 de juliol de 2016 i el 31 de desembre de 2016 del dret d'accés a la informació pública, en l'àmbit de l'Administració de la Generalitat de Catalunya (*) i el seu sector públic (març 2017), mostra que es van formular 399 sol·licitud d'accés (83% en format electrònic), essent els temes més sol·licitats els referents a l'àmbit de salut (150), d'administració local (27) i d'ensenyament (25). Del total de sol·licituds, el 64% van ésser estimades totalment, i el 16% de manera parcial. Finalment, el nombre de recursos i de reclamacions sobre el total de 399 sol·licituds presentades va ser del 30,58%.

Les administracions públiques poden no admetre, mitjançant resolució motivada, les sol·licituds d'accés a la informació si concorre alguna de les següents causes (*):

a) Que facin referència a informació en curs d'elaboració o de publicació general.

b) Que estiguin referides a informació de caràcter auxiliar o de suport, com la continguda en notes, esborranys, opinions, resums, comunicacions i informes interns, o entre òrgans o entitats administratives.

c) Que siguin relatives a informació per a la divulgació de la qual calgui una acció prèvia de reelaboració.

d) Que estiguin adreçades a un òrgan que no disposi de la informació quan se'n desconegui el competent.

e) Que siguin manifestament repetitives o tinguin un caràcter abusiu no justificat amb la finalitat de transparència d'aquesta llei.

Si hi ha drets o interessos de terceres persones afectats per la sol·licitud d'accés a la informació se'ls concedirà un termini de quinze dies perquè puguin fer les al·legacions que considerin oportunes.

El procediment s'ha de resoldre i notificar en el termini d'un mes des de la recepció de la sol·licitud per part de l'òrgan competent per a fer-ho. En el cas que la resolució en denegui l'accés, en concedeixi l'accés parcial, o la informació es faciliti en un format diferent del demanat, s'haurà de motivar. Si transcorre el termini d'un mes sense que s'hagi dictat i notificat la resolució, caldrà entendre que ha estat desestimada (*).

Una de les novetats de l'LTAIBG és la creació d'un organisme independent encarregat de garantir el compliment de les obligacions de transparència activa i passiva. En particular, aquesta norma preveu que les persones poden interposar, amb caràcter potestatiu i previ a la impugnació davant dels tribunals, una reclamació al Consejo de Transparencia y Buen Gobierno (*) o òrgan autonòmic independent que determini cada comunitat autònoma. Aquesta reclamació es pot interposar en el termini d'un mes. El termini per a resoldre i notificar la resolució és de tres mesos, i un cop transcorreguts sense resolució, la reclamació s'entendrà desestimada (*).

El Consejo de Transparencia y Buen Gobierno

El Consejo de Transparencia y Buen Gobierno es un organisme públic amb personalitat jurídica pròpia i plena capacitat d'obrar, que actua amb autonomia i plena independència en el compliment de les seves finalitats. D'acord amb l'LTAIBG, el Consejo de Transparencia y Buen Gobierno té com a finalitats promoure la transparència de l'activitat pública, vetllar pel compliment de les obligacions de publicitat, salvaguardar l'exercici del dret d'accés a la informació pública i garantir l'observança de les disposicions de bon govern.

L'LTAIBG també disposa que les comunitats autònomes poden crear un òrgan independent al qual atribuir la competència per a resoldre les reclamacions contra les resolucions relatives al dret d'accés a la informació pública de les administracions públiques del territori. Tanmateix, les comunitats autònomes poden atribuir, mitjançant l'oportú conveni, la competència per a resoldre la reclamació al Consejo de Transparencia y Buen Gobierno. La majoria de les comunitats autònomes han creat, amb configuracions i atribucions diverses, òrgans o organismes independents per a la garantia de la transparència pública.

2.1.2.4. Les dades obertes i la seva reutilització

La reutilització consisteix en l'ús, per part de qualsevol persona, de la informació pública amb finalitats comercials o no comercials (*).

La Llei 37/2007, 16 de novembre, sobre la reutilització de la informació del sector públic (LRISP), preveu que les administracions públiques poden autoritzar la reutilització de la informació que està sota el seu poder sigui quin en sigui el suport i la forma d'expressió. L'LRISP únicament exclou de la reutilització els documents que no siguin accessibles, aquells el coneixement dels quals requereixi una legitimació especial, i els continguts públics creats, per exemple, per entitats de radiodifusió o per entitat educatives.

L'LRISP preveu tres modalitats de reutilització:

- La reutilització sense subjecció a condicions.
- La reutilització amb subjecció a condicions, establertes en llicències tipus, que han d'estar disponibles en format digital i han de poder processar-se de forma electrònica.
- La reutilització amb sol·licitud prèvia, que pot incorporar condicions en una llicència. Aquesta llicència ha de reflectir, com a mínim, la informació relativa a la finalitat concreta, comercial o no comercial, per la qual es concedeix la reutilització, la durada de la llicència, les obligacions del beneficiari i de l'organisme que ho concedeix, les responsabilitats d'ús i les modalitats financeres, per la qual cosa cal indicar el caràcter gratuït o, si escau, la contraprestació econòmica aplicable.

Les condicions que s'incorporin en les llicències de reutilització han de ser clares, justes i transparents; no han de restringir les possibilitats de reutilització, ni limitar-ne la competència; no han d'ésser discriminatòries per a categories comparables de reutilització, i no han de contenir acords exclusius (*).

L'LRISP disposa que la reutilització de la informació es pot sotmetre a les següents condicions generals:

a) Que el contingut de la informació, incloent-hi les metadades, no sigui alterat.

b) Que no es desnaturalitzi el sentit de la informació.

c) Que se citi la font.

d) Que es mencioni la data de l'última actualització.

e) Quan la informació contingui dades de caràcter personal, la finalitat o les finalitats concretes per a les quals és possible la reutilització futura de les dades.

f) Quan la informació, tot i ser facilitada de forma dissociada, contingués prou elements que poguessin permetre la identificació dels interessats en el procés de reutilització (*), la prohibició de revertir el procediment de dissociació mitjançant l'addició de noves dades obtingudes d'altres fonts.

La reutilització de la informació es pot veure facilitada per l'obertura de les dades públiques (*).

L'obertura de les dades públiques afavoreix la reutilització de la informació pública en facilitar la creació d'informació i l'oferta de nous serveis de valor afegit basats en les dades difoses per les administracions públiques. En aquest sentit, la Comissió Europea afirma que:

> «[...] L'impuls cap a la utilització de dades obertes està agafant força en diversos estats membres, que adopten aquest concepte per raons de transparència, eficiència administrativa i potencial econòmic de la reutilització.»
>
> Comissió Europea (2011)

L'obertura de les dades públiques consisteix a eliminar els obstacles tècnics o jurídics que poden dificultar la reutilització de la informació pública. En efecte, les dades públiques que es posen a disposició de la ciutadania han de complir determinades característiques que persegueixen garantir que es puguin reutilitzar fàcilment per a generar noves informacions o serveis que, en última instància, poden incrementar la transparència de les administracions públiques. A més, el compliment de les obligacions de transparència exigeix un canvi en la forma com es gestiona, emmagatzema i recupera la informació per a facilitar no només una major eficàcia i eficiència administrativa, sinó per a contribuir en l'increment de la transparència pública.

Les dades obertes es caracteritzen, des d'un punt de vista tècnic, per poder processar-se electrònicament i, des del punt de vista jurídic, per l'absència de restriccions que en limiten la reutilització (propietat intel·lectual, dades personals, etc.). D'acord amb la Sunlight Foundation, les dades obertes són aquelles que compleixen les següents condicions:

- Dades completes. Les dades han de ser tan completes com sigui possible, i no poden estar afectades per limitacions relatives a la privacitat, la seguretat o qualsevol altra restricció.

- Dades primàries. Les dades obertes han de ser dades originals. Cal facilitar informació sobre com s'han obtingut i on són els documents originals, de manera que els usuaris puguin comprovar si les dades han estat seleccionades i emmagatzemades adequadament.

- Dades oportunes. Les dades s'han de facilitar de forma ràpida, tan bon punt es recullin. En la mesura del possible cal facilitar les dades en temps real, de manera que se'n pugui maximitzar el valor. Cal donar prioritat a la difusió de les dades que siguin peribles (*time sensitive*).

- Dades accessibles de forma física i electrònica. Les dades han de ser tan accessibles com sigui possible, tant per mitjans físics com electrònics, per a evitar els desplaçaments o la necessitat de sol·licitar formalment l'accés a la informació.

- Dades processables electrònicament. Les dades han d'estar disponibles en formats d'ús generalitzat i que facilitin el processament automàtic.

- Dades no discriminatòries. Les dades han d'estar disponibles per a tothom, sense que calgui fer una sol·licitud o qualsevol altre tràmit.

- Dades que fan servir estàndards oberts en la difusió.

- Dades reutilitzables sense llicència. L'ús de les dades no s'ha de sotmetre a cap regulació que en restringeixi la reutilització excepte, de forma raonable, quan hi hagi aspectes relatius a la privacitat o la seguretat.

- Dades permanents.

- Dades sense costos. No s'ha d'establir una contraprestació per a poder accedir o reutilitzar les dades i, en cas que s'estableixi, la seva quantificació s'ha de basar en costos marginals i no en costos totals.

L'LRISP (art. 5 Llei 37/2007), després de la modificació operada per la Llei 18/2015, de 9 de juliol, disposa 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».

D'aquesta manera, l'LRISP preveu que, quan sigui possible, les administracions públiques han de donar les dades obertes per a facilitar-ne la reutilització. Tanmateix, tot i la formulació d'aquest precepte, és important tenir present que si bé les administracions públiques han de mirar de difondre les dades en format obert i promoure que posar-les a disposició per a la seva reutilització es faci per mitjans electrònics, i mitjançant plataforma multicanal, no tenen l'obligació de difondre-les en aquest format.

Diverses normes autonòmiques preveuen que les administracions públiques han de promoure la difusió de les dades públiques en formats oberts (Castella – la Manxa, Galícia, Aragó, Múrcia, Andalusia i Extremadura). Així mateix, diverses normes autonòmiques disposen la creació de portals de dades obertes o de seccions específiques a l'efecte en els portals de transparència, amb la finalitat d'afavorir-ne la reutilització (Castella – la Manxa i Aragó). Altres normes també preveuen la creació de catàlegs d'informació reutilitzable (Galícia, i la Rioja).

Portal de dades obertes de l'Ajuntament de Barcelona

L'Ajuntament de Barcelona difon 420 conjunts de dades (*datasets*) en 20 formats diferents que en faciliten la consulta i la reutilització, com ara el Bicing, el cens d'activitats comercials o els equipaments escolars (conjunts més descarregats).

L'LRISP estableix diferents infraccions per l'incompliment del que es disposa en l'àmbit de l'Administració General de l'Estat, com ara la desnaturalització del sentit de la informació; l'alteració del contingut de la informació per a la reutilització; i la reutilització de documentació sense haver obtingut la corresponent llicència, en els casos en què sigui requerida o incompleixi les condicions imposades en la llicència o la normativa reguladora aplicable. L'LRISP preveu que, per la comissió de les anteriors infraccions, es poden imposar diferents sancions de multa a més de la prohibició de reutilitzar documents sotmesos a llicència durant un període d'entre 1 i 5 anys, i amb la revocació de la llicència concedida. Aquest règim sancionador únicament és aplicable en relació amb l'Administració General de l'Estat.

2.1.2.5. Els límits de la transparència pública. La protecció de dades personals

De vegades, el coneixement públic de determinada informació pública pot generar danys o perjudicis a determinats bens, drets i interessos, la protecció dels quals és necessària. Per a evitar-ho, l'LTAIBG disposa que la transparència pot ser limitada quan el coneixement o l'accés a la informació signifiqui un perjudici a una sèrie de bens o drets (*) com:

- la seguretat nacional,
- la defensa,
- les relacions exteriors,
- la seguretat pública,
- la prevenció, investigació i sanció dels il·lícits penals, administratius o disciplinaris,
- la igualtat de les parts en els processos judicials i la tutela judicial efectiva,
- les funcions administratives de vigilància, inspecció i control,
- els interessos econòmics i comercials,
- la política econòmica i monetària,
- el secret professional i la propietat intel·lectual i industrial,
- la garantia de la confidencialitat o el secret requerit en processos de presa de decisió, i la protecció del medi ambient.

Aquests límits són aplicables tant en la difusió de la informació pública, com en l'accés a la informació i, fins i tot, en la reutilització. En efecte, l'article 5.3 LTAIBG disposa que «són aplicables, si escau, els límits al dret d'accés a la informació pública que preveu l'article 14 i, especialment, el derivat de la protecció de dades de caràcter personal, que regula l'article 15». Així mateix, l'article 14 estableix que «el dret d'accés podrà ser limitat quan accedir a la informació suposi un perjudici» per a determinats bens i drets. Finalment, l'LRISP preveu que no serà aplicable en els documents sobre els quals hi hagi prohibicions o limitacions en el dret d'accés en virtut del que preveu l'LTAIBG, així com «en els documents als quals no pugui accedir o l'accés estigui limitat en virtut de règims d'accés per motius de protecció de les dades personals, de conformitat amb la normativa vigent i les parts de documents accessibles, en virtut d'aquests règims que continguin dades personals i la reutilització s'hagi definit per llei com incompatible amb la legislació relativa a la protecció de les persones físiques, pel que fa al tractament de les dades personals».

Amb caràcter general cal advertir que la simple concurrència d'un d'aquests bens o drets en la informació que es vulgui difondre, o a la qual una persona vulgui accedir, no en determina la impossibilitat del coneixement públic. Sobre això cal tenir en compte els criteris que es deriven del que preveu l'LTAIBG.

En primer lloc, les administracions públiques no disposen de potestat discrecional per a denegar o limitar l'accés a la informació pública.

En segon lloc, l'aplicació dels límits ha de ser motivada. Així ho assenyala amb caràcter bàsic l'LTAIBG quan afirma que «seran motivades les resolucions que deneguin l'accés, les que concedeixin l'accés parcial o mitjançant una modalitat diferent de la sol·licitada, i les que permetin l'accés quan hagi hagut oposició d'un tercer».

En tercer lloc, la concurrència d'un interès o d'un dret protegit per un límit no significa obligatòriament la denegació de l'accés a la informació pública. En efecte, l'LTAIBG preveu que l'accés podrà ser limitat, però no cal que ho sigui, ni que ho sigui directament.

En quart lloc, cal que la difusió o l'accés a la informació signifiqui un perjudici concret, definit i avaluable. Per a poder determinar el perjudici que el coneixement de la informació pugui generar, així com la concurrència d'un interès que justifiqui la difusió o l'accés a la informació pública, les administracions públiques poden utilitzar diferents instruments.

D'una banda, el test del dany per mitjà del qual l'Administració pública ha de valorar el perjudici, dany o fins i tot el risc que l'accés pot ocasionar al be o interès protegit pel límit. D'altra banda, el test de la ponderació d'interessos mitjançant el qual l'Administració pública ha de ponderar si es dona un interès públic o privat superior que hi justifiqui l'accés.

L'LTAIBG disposa que quan l'aplicació d'algun dels límits que preveu no afecti la totalitat de la informació, es concedirà l'accés parcial, havent omès prèviament la informació afectada pel límit, llevat que en resulti una informació distorsionada o que no tingui sentit, i s'indicarà al sol·licitant que part de la informació ha estat omesa.

A més dels límits que s'han citat anteriorment, l'LTAIBG també disposa que quan la informació sol·licitada conté dades personals, l'accés a la informació pot ser limitat si concorren determinades circumstàncies o es compleixen determinats criteris. Amb aquesta finalitat, les dades personals es classifiquen en tres grups que tenen diferent règim de transparència (*).

1) La transparència de la informació que conté dades personals especialment protegides

L'LTAIBG preveu dos règims diferents.

En primer lloc, quan la informació sol·licitada conté dades especialment protegides relatives a la ideologia, l'afiliació sindical, la religió o les creences l'accés, únicament es pot autoritzar en cas que es tingui el consentiment exprés i per escrit de l'afectat, tret que l'afectat hagi fet manifestament públiques les dades abans que es demanés l'accés.

En segon lloc, quan la informació inclou dades especialment protegides sobre l'origen racional, la salut o la vida sexual o dades relatives a la comissió d'infraccions penals o administratives que no comportin l'amonestació pública a l'infractor, l'accés només es pot autoritzar en cas que es tingui el consentiment de l'afectat, o si està emparat per una norma amb rang de llei.

2) La transparència de la informació que conté dades personals

Quan la informació sol·licitada no contingui dades especialment protegides, l'òrgan al qual s'adreci la sol·licitud ha de concedir-ne l'accés, després de ponderar de manera prou raonada l'interès públic en la divulgació de la informació i els drets dels afectats, les dades dels quals apareguin en la informació sol·licitada.

Per tal de dur a terme aquesta ponderació, l'LTAIBG preveu els següents criteris:

a) El menor perjudici als afectats derivat del transcurs dels terminis previstos a la Llei 16/1985, de 25 de juny, del patrimoni històric espanyol (vint-i-cinc anys des de la mort de la persona afectada, si se'n coneix la data, o cinquanta anys des de la data del document).

b) La justificació pels sol·licitants de la petició en l'exercici d'un dret, o el fet que tinguin la condició d'investigadors i motivin l'accés a finalitats històriques, científiques o estadístiques.

c) El menor perjudici dels drets dels afectats en cas que els documents únicament continguin dades de caràcter merament identificatiu.

d) La garantia més gran dels drets dels afectats en cas que les dades contingudes en el document puguin afectar-ne la intimitat o la seguretat, o es refereixin a menors d'edat.

3) La transparència de la informació que conté dades personals merament identificatives

Quan la informació a la qual es vol accedir, difondre o reutilitzar només contingui dades identificatives relacionades amb l'organització, el funcionament o l'activitat pública de l'òrgan (nom i cognoms, DNI, càrrec, telèfon, adreça de correu electrònic, etc.) seran accessibles llevat en el cas concret que prevalgui la protecció de dades personals o altres drets constitucionalment protegits sobre l'interès públic en la divulgació que ho impedeixi.

Quan l'Administració pública, amb caràcter previ a la difusió o a l'accés, ha anonimitzat el document o dissociat les dades personals, no s'aplicaran els criteris que s'han comentat anteriorment, sempre que s'impedeixi la identificació de les persones afectades.

Orientacions sobre protecció de dades en la reutilització de la informació del sector públic

L'Agència Espanyola de Protecció de Dades proposa que abans de decidir facilitar la reutilització de la informació cal examinar els riscos per als interessats i les mesures que es poden utilitzar per a minimitzar-los. En aquest sentit, en els casos en què es puguin derivar riscos per a la protecció de les dades personals, cal anonimitzar la informació essent

conscients de les possibilitats de reidentificar les persones, i de la necessitat de complementar l'anonimització amb compromisos jurídicament vinculants per llicències específiques que en facin referència.

La finalitat del procés d'anonimització és eliminar o reduir al mínim els riscos de reidentificació de les dades anonimitzades, i mantenir-ne la veracitat dels resultats del tractament, és a dir, a més d'evitar la identificació de les persones, les dades anonimitzades han de garantir que qualsevol operació o tractament que pugui dur-se a terme posteriorment a l'anonimització no comporta una distorsió de les dades reals. D'acord amb l'AEPD, cal garantir la irreversibilitat de l'anonimització, i per a això hauran de valorar tant les fonts d'informació disponibles en els diferents mitjans i, en particular, mitjançant internet, com la tecnologia aplicable en els processos d'anonimització i en els de reidentificació. L'anonimització es pot dur a terme amb diferents tècniques que convé utilitzar de manera combinada per a reduir els riscos d'identificació.

Finalment, cal tenir present que la normativa de protecció de dades personals és aplicable al tractament posterior de les obtingudes per mitjà de l'exercici del dret d'accés.

2. Aspectes jurídics del periodisme de dades

2.1. El govern obert

2.1.3. La participació i la col·laboració ciutadana

La participació és el segon pilar del govern obert, i consisteix en la intervenció de la ciutadania en el disseny, l'execució, el seguiment i el control de l'activitat de les administracions públiques. La participació també facilita la col·laboració ciutadana en el desenvolupament d'aquestes activitats, el foment de la integritat pública i la prevenció dels conflictes d'interessos, i la lluita contra la corrupció.

La participació i la col·laboració es pot dur a terme mitjançant diferents aplicacions (*).

La participació i la col·laboració ciutadana (*) adquireix un paper important quan parlem de periodisme de dades.

Els professionals del periodisme que han accedit a les dades poder col·laborar amb les administracions públiques en la difusió.

«¿Quién cobra la obra?»

És un projecte impulsat per la Fundación Ciudadana Civio, en el qual s'han analitzat tots els contractes d'obra pública publicats en el BOE entre els anys 2009 i 2015 (en total, 8.058 contractes). L'elaboració del projecte ha tingut força dificultats pel baix nivell de reusabilitat de la informació analitzada, que ha implicat que algunes de les tasques s'hagin hagut de fer de forma manual. El projecte permet visualitzar de forma gràfica les constructores adjudicatàries, les obres públiques, els contractes adjudicats, així com cercar dades i comparar-les.

També poden col·laborar amb les administracions públiques analitzant les dades, la qual cosa pot permetre tenir un millor coneixement del funcionament de les administracions públiques, de com gasten els diners o de com adjudiquen els contractes públics.

Development Gateway

L'ONG Development Gateway ha desenvolupat un tauler de comandament basat en programari lliure i en l'estàndard de dades de contractació oberta, que sobre la base d'unes alertes (*red flags*), prèviament definides, permet identificar la possible existència d'irregularitats o de casos de corrupció, o els riscos que hi ha perquè puguin ser analitzats o investigats. Aquest sistema també permet identificar les irregularitats més freqüents en una determinada administració pública en matèria de contractació pública.

Finalment, els periodistes també poden col·laborar informant o denunciant les irregularitats que detectin per l'anàlisi que hagin fet de les dades a les quals hagin pogut accedir.

Plataforma Xnet

En l'actualitat hi ha aplicacions que permeten la informació o denúncia d'irregularitats i casos de corrupció per mitjans electrònics. Com a exemple, la «Bústia d'Xnet» permet que qualsevol persona que tingui coneixement d'una irregularitat o d'un cas de corrupció ho pugui comunicar de manera confidencial perquè s'analitzi si pot tenir una sortida jurídica, i perquè es faci arribar a periodistes que col·laboren en la plataforma.

2. Aspectes jurídics del periodisme de dades

2.2. Llibertat d'informació i d'expressió

2.2.1. Introducció

La Constitució espanyola reconeix la llibertat d'expressió i la llibertat d'informació com a drets fonamentals (art. 20).

La llibertat d'expressió i la llibertat d'informació en els tractats internacionals

La llibertat d'expressió i la llibertat d'informació estan reconegudes en diversos tractats internacionals. Aquests, més enllà del valor intrínsec, són un instrument per a la interpretació de les normes que regulen els drets fonamentals (art. 10.2 CE (*)).

Pacte Internacional dels Drets Civils i Polítics (art. 19)

1) Ningú no podrà ésser molestat a causa de les seves opinions.

2) Tota persona té dret a la llibertat d'expressió; aquest dret comprèn la llibertat de buscar, rebre i difondre informacions i idees i de tot tipus, sense consideració de fronteres, tant oralment, per escrit o en forma impresa o artística, o per qualsevol altre procediment que esculli.

3) L'exercici del dret previst en el paràgraf 2 d'aquest article comporta deures i responsabilitats especials. Per això, pot estar subjecte a algunes restriccions que hauran d'estar expressament fixades per la llei, i siguin necessàries per a:

a) Assegurar el respecte als drets o a la reputació d'altres.

b) La protecció de la seguretat nacional, l'ordre públic o la salut o la moral públiques.

Conveni Europeu de Drets Humans (art. 10)

Tota persona té dret a la llibertat d'expressió. Aquest dret comprèn la llibertat d'opinió i la llibertat de rebre o de comunicar informacions o idees sense la ingerència de les autoritats públiques i sense consideració de fronteres. El present article no impedeix que els estats sotmetin les empreses de radiodifusió, de cinematografia o de televisió a un règim d'autorització prèvia.

Carta dels Drets Fonamentals de la Unió Europea (art. 11)

1) Tota persona té dret a la llibertat d'expressió. Aquest dret comprèn la llibertat d'opinió i la llibertat de rebre o comunicar informacions o idees sense que pugui haver ingerència de les autoritats públiques i sense consideració de fronteres.

2) Es respecten la llibertat dels mitjans de comunicació i el pluralisme.

La llibertat d'expressió i la llibertat d'informació estan estretament vinculades a la «llibertat de comunicar el producte d'una tasca intel·lectual prèvia» i a la «llibertat de les persones a manifestar-se en la societat per mitjà de les idees», tot i que tenen un contingut autònom i diferenciat.

En aquest sentit, el Tribunal Constitucional s'ha manifestat en nombroses ocasions considerant que la llibertat d'expressió i la llibertat d'informació són dos drets diferents i autònoms, tot i que estretament vinculats en origen. La llibertat d'expressió fa referència a la lliure difusió de pensaments, d'idees i d'opinions, mentre que la llibertat d'informació es refereix a la manifestació de fets o de successos noticiables respecte al qual es pugui demostrar la veracitat.

> «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 novembre

La llibertat d'expressió i la llibertat d'informació contribueixen a la formació de l'opinió pública lliure de la qual són el fonament, i alhora són el motiu per a l'exercici d'altres drets fonamentals, com ara el dret a la participació política i d'altres inherents al funcionament del sistema democràtic (STC 159/1986, de 12 de desembre).

Des del punt de vista de les garanties, la llibertat d'expressió i la llibertat d'informació, com a drets fonamentals que són, tenen reconeguts diferents mecanismes de garantia en la Constitució espanyola.

D'una banda, l'eficàcia directa d'aquests drets vinculants en tots els poders públics. Qualsevol ciutadà pot reclamar la tutela judicial d'aquests drets, que pot canalitzar-se per un procediment preferent i sumari davant dels tribunals de justícia i, de manera extraordinària, mitjançant un recurs d'empara davant del Tribunal Constitucional.

D'altra banda, la regulació d'aquests drets cal fer-la per mitjà d'una llei orgànica, és a dir, una llei aprovada per les Corts Generals per majoria absoluta dels diputats.

La llibertat d'expressió i la llibertat d'informació no poden ésser restringides mitjançant cap tipus de censura prèvia (art. 20.2 CE). D'acord amb el Tribunal Constitucional (STC 187/1999, de 25 d'octubre), «per censura prèvia ha de tenir-se qualsevol mesura limitativa de l'elaboració o de la difusió d'una obra de l'esperit, que consisteixi en el sotmetiment a un examen previ del contingut per part d'un poder públic, i la finalitat del qual sigui enjudiciar l'obra en qüestió d'acord amb uns valors abstractes i restrictius de la llibertat, de manera que s'atorgui el plàcet a la publicació de l'obra si s'acomoda al judici del censor, i es negui en cas contrari».

2. Aspectes jurídics del periodisme de dades

2.2. Llibertat d'informació i d'expressió

2.2.2. La llibertat d'informació

2.2.2.1. Introducció

L'article 20 CE reconeix i protegeix el dret de comunicar o de rebre lliurement informació veraç per qualsevol mitjà de difusió. Qualsevol persona, en el marc de la seva autonomia personal, pot ser titular del dret a la llibertat d'informació. Per tant, no només els periodistes són titulars d'aquest dret, tot i que són els qui més sovint l'exerceixen.

Com es desprèn de la formulació, aquest dret es basa en la veracitat de la informació. La llibertat d'informació «empara la transmissió de fets verídics que, alhora, poden ser considerats noticiables, és a dir, que revesteixin interès o rellevància pública per a la societat». El Tribunal Constitucional, en la seva jurisprudència, reconeix l'estreta connexió entre la llibertat d'informació i la llibertat de formació de l'opinió pública, per la qual cosa ha de prevaldre per sobre d'altres drets amb els quals pugui entrar en conflicte, sempre que la informació que es transmet sigui veraç i rellevant.

2.2.2.2. Veracitat

La veracitat no vol dir veritat o realitat incontrovertible dels fets, sinó que fa referència a la diligència del periodista per a esbrinar els fets sobre els quals versa la informació, i d'acord amb allò que és exigible a un professional de la informació que exclou rumors o invencions. Tampoc no vol dir que la informació no pugui ser errònia, sinó que hagi estat contrastada prèviament amb dades objectives o, si escau, que es rectifiqui ràpidament.

En diverses sentències, el Tribunal Constitucional ha perfilat l'abast d'aquest terme i ha identificat els elements que han de concórrer per a poder acreditar la veracitat. En efecte, la jurisprudència constitucional ha indicat que la veracitat està vinculada a la possibilitat de contrastar la informació que es difon, a la diligència del periodista per a esbrinar la informació, a l'actitud positiva respecte a la veritat, i a l'activitat del periodista per a comprovar de manera raonable els fets que es difondran.

> «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 trasmiten 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 setembre

Un altre aspecte rellevant és el relatiu a la neutralitat amb què el periodista tracta la informació.

> «Para que pueda hablarse de reportaje neutral han de concurrir los siguientes requisitos:
>
> a) 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)».

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 febrer

2.2.2.3. Rellevància pública

No n'hi ha prou amb la veracitat perquè la informació que es comuniqui es consideri una manifestació de la llibertat d'informació. La llibertat d'informació està més estretament vinculada a la rellevància pública que a la informació, és a dir, relacionada amb assumptes públics d'interès general. Així, doncs, la transmissió de rumors no està emparada pel dret a la llibertat d'informació.

Cal recordar que la llibertat d'informació està relacionada amb la formació de l'opinió pública. Per aquest motiu, la informació ha de ser d'interès per a la ciutadania, la qual cosa és diferent de la simple satisfacció de la curiositat humana per a conèixer la vida d'altres (STC 83/2002, de 22 d'abril).

D'aquesta manera, el periodista, abans de comunicar la informació, ha de valorar si aquesta pot tenir o no interès per a la ciutadania.

2. Aspectes jurídics del periodisme de dades

2.2. Llibertat d'informació i d'expressió

2.2.3. La llibertat d'expressió

La Constitució espanyola (art. 20.1.a) defineix la llibertat d'expressió com el dret «a expressar i difondre lliurement els pensaments, les idees i les opinions mitjançant la paraula, l'escriptura o qualsevol altre mitjà de reproducció».

La llibertat d'expressió té com a objecte garantir que les persones puguin exterioritzar o transmetre els seus pensaments, idees, opinions o judicis de valor.

La llibertat d'expressió està conformada per dos elements.

En primer lloc, la creació de pensaments, idees o opinions com a manifestació de la capacitat humana. Aquestes creences, raonaments o idees, en la mesura que són pròpies de cada persona, i per tant poden ser subjectives, no han de ser veraces ni cal acreditar-ne la veracitat. No cal que se'n demostri l'exactitud, i per tant, quan algú exerceix la llibertat d'expressió no se li demana que en demostri la veracitat, o mostri diligència per a esbrinar-la, com succeeix en el cas de la llibertat d'informació.

En segon lloc, el mitjà pel qual es manifesten. Com es desprèn de la lectura de la Constitució espanyola, les persones poden expressar les seves idees o pensaments per qualsevol mitjà, fins i tot del llenguatge no verbal.

L'exercici de la llibertat d'expressió està estretament vinculat a altres drets, com ara la llibertat ideològica (art. 16.1 CE), el dret a rebre informació (art. 20.1.d CE), i la llibertat d'expressió activa.

En primer lloc, la llibertat ideològica és prèvia a la llibertat d'expressió i és l'origen de les idees i les opinions de les persones.

En segon lloc, el dret a rebre informació veraç per qualsevol mitjà permet a les persones formar-se una opinió.

En tercer lloc, la llibertat d'expressió activa és més àmplia que la llibertat d'informació, que, com s'ha vist, està vinculada a la veracitat i a l'interès públic de la informació, aspectes que no han de concórrer quan es parla de la llibertat d'expressió.

2. Aspectes jurídics del periodisme de dades

2.2. Llibertat d'informació i d'expressió

2.2.4. Límits a la llibertat d'informació i d'expressió

2.2.4.1. Introducció

La llibertat d'informació i d'expressió poden ésser limitats per a protegir altres drets o bens. D'acord amb la Constitució espanyola, la llibertat d'informació i d'expressió tenen el límit en el respecte als drets reconeguts en el títol primer, en els preceptes de les lleis que el desenvolupin i, especialment, en el dret a l'honor, a la intimitat, a la imatge, i a la protecció de la joventut i de la infància (art. 20.4 CE). A més, hi ha lleis que també preveuen límits per a protegir bens, com ara la seguretat pública.

2.2.4.2. Els drets de personalitat com a límit de les llibertats d'informació i d'expressió

La Constitució espanyola reconeix la dignitat de la persona, els drets inviolables que li són inherents, i el lliure desenvolupament de la personalitat (art. 10.1 CE). També garanteix el dret a l'honor, a la intimitat personal i familiar i a la imatge (art. 18.1 CE).

Tots aquests aspectes constitueixen límits de les llibertats d'informació i d'expressió. Tanmateix, no són límits absoluts. La llibertat d'expressió o la llibertat d'informació poden prevaldre respecte al dret a la protecció de l'honor, la intimitat o la imatge en determinades circumstàncies, com ara, quan es derivi de la llei, per exemple, perquè estigui autoritzat en la llei mateixa, s'autoritzi per una autoritat pública d'acord amb el que preveu la llei, prevalgui un interès històric, científic o cultural rellevant, o per actes propis o per consentiment (*).

Atès el caràcter relatiu dels drets, tant el Tribunal Constitucional com el Tribunal Suprem han identificat uns criteris per a dur a terme la ponderació entre els diversos drets. En primer lloc, cal determinar quin dret s'està exercint, és a dir, si el conflicte és fruit de l'exercici de la llibertat d'expressió o de la llibertat d'informació. En segon lloc, cal valorar si el dret en qüestió s'ha exercit correctament o s'ha excedit en els límits establerts. En tercer lloc, cal determinar si en l'exercici de la llibertat d'informació o d'expressió s'han superat els límits permesos, de manera que hagi de prevaldre el dret a l'honor, a la intimitat o a la imatge.

> «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 juny de 1995

a) El dret a l'honor (*). L'honor és un concepte que depèn de les circumstàncies socials i culturals de cada moment, i que està vinculat a la reputació o la fama que cada persona té davant dels altres. El Tribunal Constitucional ha anat perfilant amb la jurisprudència uns criteris que han de tenir-se en compte per a perfilar aquest concepte, com ara la forma i el context en què s'emmarquen, la inexistència de deshonor pels actes, el prestigi professional o social, o els actes contraris a la dignitat 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 d'octubre

La LOPC considera intromissions il·legítimes en el dret a l'honor «la imputació de fets o la manifestació de judicis de valor mitjançant accions o expressions que de qualsevol manera lesionin la dignitat d'una altra persona, menyscabant-ne la fama, o atemptant contra la seva estimació» (art. 7.7 LOPC). Per això, quan una notícia imputi una persona d'un fet denigrant o difamant

es considerarà il·legítima si és falsa, encara que sigui d'interès públic o verídica. D'altra banda, si les manifestacions dutes a terme en l'exercici de la llibertat d'expressió són innecessàries per a l'exposició de les idees, degraden la reputació d'una persona en qüestions que no tenen interès públic, o simplement són insults o injúries, també es poden considerar com a il·legítimes.

> «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. 1ª CE están excluidas las expresiones absolutamente vejatorias; es decir, aquéllas 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 novembre

> «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 juny

b) El dret a la intimitat. La jurisprudència del Tribunal Constitucional també ha perfilat el concepte d'intimitat, que té per objectiu:

> «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 éstos 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 d'abril.

Poden considerar-se actes d'intromissió en la intimitat l'ús d'aparells d'escolta, de filmació o de dispositius òptics que permetin enregistrar, reproduir o conèixer la vida íntima de les persones, i la divulgació de dades relatives a la vida privada de les persones quan això n'afecti la reputació o bon nom, o la revelació de dades conegudes en l'exercici de l'activitat professional.

La intimitat també és objecte de protecció penal mitjançant delictes com el de descobriment o el de revelació de secrets (art. 197 CP (*)).

c) Dret a la imatge. El dret a la imatge té un abast més ampli que el dret a la intimitat, tot i que, de vegades, pot estar-hi vinculat.

> «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 març

Aquest dret faculta les persones a decidir sobre la difusió pública de la seva imatge, a impedir que una tercera persona pugui captar, reproduir o difondre-la i, per tant, cal considerar com a formes d'intromissió qualsevol captació, reproducció o difusió mitjançant una fotografia, un vídeo o qualsevol altre mitjà, de la imatge d'una persona en llocs o moments de la vida privada, o fins i tot fora d'ella, excepte en els casos previstos en la LOPC.

En particular, aquesta Llei orgànica preveu que es tracti de persones que exerceixin un càrrec públic o una professió de notorietat o projecció pública, i la imatge es capti durant un acte públic o en llocs oberts al públic, la utilització de la caricatura d'aquestes

persones o informació gràfica sobre un succés o fet públic, quan la imatge d'una persona determinada aparegui com a merament accessòria.

2.2.4.3. Els límits legals de les llibertats d'informació i d'expressió

Com ja s'ha avançat, a més de la protecció d'altres drets fonamentals, la Constitució espanyola també reconeix que la llibertat d'expressió i d'informació tenen com a límit altres drets i interessos definits per les lleis.

a) Protecció del menor

La Llei orgànica 1/1996, de 15 de gener, de protecció jurídica del menor disposa que prevaldrà l'interès superior dels menors per sobre de qualsevol altre interès legítim que pugui concórrer (art. 2.1). Això es tradueix, entre d'altres mesures, en la intervenció del Ministeri fiscal en la difusió d'informació, la utilització d'imatges o noms de menors en mitjans de comunicació que pugui implicar una intromissió il·legítima en la seva intimitat, honra o reputació, o sigui contrària als seus interessos, fins i tot quan hi hagi el consentiment del menor o dels representants legals (art. 4).

b) Dret de rectificació

La Llei orgànica 2/1984, de 26 de març, reguladora del dret de rectificació disposa que tota persona té dret a rectificar la informació difosa per qualsevol mitjà de comunicació i relativa a fets que tinguin relació amb ella, quan consideri que són inexactes o que la divulgació pot causar-li un perjudici.

Aquest dret s'exerceix mitjançant un escrit de rectificació que cal enviar al director del mitjà de comunicació en el termini de set dies des de la publicació de la informació, un escrit de rectificació que haurà de ser publicat en el termini de tres dies. Si en aquest termini no es publica la rectificació, es podrà exercir una acció de rectificació davant del jutjat de primera instància.

> «La inserción de la réplica sólo 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) Els secrets oficials

La Llei 9/1968, de 5 d'abril, sobre secrets oficials disposa que poden ser declarades matèries classificades els assumptes, actes, documents, informacions, dades i objectes el coneixement dels quals, per persones no autoritzades, pot perjudicar o posar en risc la seguretat i la defensa de l'Estat (art. 2 LSO).

La declaració d'una matèria com a classificada implica que només podrà ésser coneguda pels òrgans i les persones degudament facultats. A més, també determina la prohibició que les persones que no estiguin autoritzades puguin accedir o circular pels llocs o les zones on hi hagi matèries classificades (art. 8 LSO).

L'LSO estableix que la persona que tingui coneixement d'una matèria classificada està obligada a mantenir-la en secret i lliurar-la a l'autoritat corresponent. Quan es consideri que una matèria classificada pot arribar a coneixements d'un mitjà de comunicació, se'ls notificarà la circumstància als efectes oportuns (art. 9 LSO).

Finalment, cal recordar que el Codi penal tipifica com a delicte la revelació de secrets oficials (art. 13 LSO, en relació amb l'article del CP).

2. Aspectes jurídics del periodisme de dades

2.3. L'exercici de la professió periodística i l'autorregulació deontològica

L'exercici de la professió periodística està regulada per diferents normes.

En primer lloc, l'exercici de la professió periodística està regulada per la Constitució espanyola i per diverses normes que la regulen. La Constitució preveu que la llei regularà el dret a la clàusula de consciència, i també al secret professional en l'exercici de la llibertat d'informació (art. 20.1.d CE).

Pel que fa a la clàusula de consciència, la Llei orgànica 2/1997, de 19 de juny, reconeix que és un dret constitucional dels professionals de la informació que té per objecte garantir la independència en l'exercici de l'activitat professional (art. 1 Llei orgànica 2/1997). D'acord amb la Llei orgànica, els periodistes poden sol·licitar la rescissió dels contractes amb les empreses per a les quals treballin en les circumstàncies previstes (per canvi substancial d'orientació informativa o de línia ideològica, i per trasllat a un altre mitjà que signifiqui una ruptura d'orientació informativa o de línia ideològica). A més, els periodistes poden negar-se a participar en l'elaboració d'informacions que siguin contràries als principis ètics de la comunicació.

> «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 sólo el reconocimiento del derecho a la información cómo 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 novembre

Pel que fa al secret professional, és un dret dels periodistes, no un deure. És a dir, els periodistes no estan obligats jurídicament (*) a guardar secret sobre les dades que coneguin durant l'exercici de la professió, com succeeix, per exemple, amb els advocats o els metges. No obstant això, des del punt de vista ètic no és correcte desvelar la identitat de les fonts que aporten informació compromesa. En qualsevol cas, el secret professional inclou el dret a no revelar la identitat de les fonts de la informació.

En segon lloc, les normes que adopten els periodistes. En efecte, els periodistes com a col·lectiu adopten codis ètics que defineixen els valors i els principis que regeixen l'exercici de la seva activitat. Aquests codis poden anar acompanyats de la creació d'òrgans encarregats de vetllar pel seu compliment i d'impulsar les mesures previstes en el cas d'incompliment.

Codi Deontològic de la professió periodística a Catalunya

El Codi Deontològic de la professió periodística a Catalunya fou adoptat l'any 1992 i ha estat objecte de diverses actualitzacions (la darrera, l'any 2016).

Està integrat pels següents criteris:

1. Informar de manera acurada i precisa.
2. Evitar perjudicis per informacions sense prou fonament.
3. Rectificar les informacions incorrectes.

4. Utilitzar mètodes lícits i dignes per a obtenir informació.

5. Citar les fonts i preservar el secret professional.

6. Conciliar els drets individuals amb el dret del públic de saber.

7. Evitar el conflicte d'interessos.

8. No utilitzar en benefici propi.

9. Respectar el dret a la privacitat.

10. Salvaguardar la presumpció d'innocència.

11. Protegir els drets dels menors.

12. Respectar la dignitat de les persones i la seva integritat física i moral.

També formen part del codi diversos annexos en els quals s'inclouen recomanacions en relació amb la manipulació d'imatges, d'internet, del plagi, del tractament dels conflictes bèl·lics o armats, de l'ús del terme «il·legals» (referit a persones), i de la cita de nacionalitats i d'ètnies.

El Consell de la Informació de Catalunya (CIC) és l'organisme independent encarregat de vetllar pel compliment del Codi Deontològic de la professió periodística a Catalunya. Està integrat per periodistes i representants d'altres sectors professionals.

Bibliografia

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, pàg. 770–792).

Cerrillo i Martínez, A. (2014). «Los principios de datos abiertos en la legislación española». *Revista de Internet, Derecho y Política* (núm. 19, pàg. 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, pàg. 67-85).

Diversos autors (2017). *Los límites al derecho de acceso a la información pública*. Madrid: Instituto Nacional de Administración Pública.

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, José 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, José 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.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.1. Les fonts periodístiques

3.1.1. Introducció

Hem de ser tan incansables en la recol·lecció de fonts potencials com respecte a la informació en sí.

> "Malgrat que les relacions amb les fonts són, en la seva majoria, pura rutina, meres transaccions en les que tots surten guanyant, en ocasions ens veiem involucrats en un joc d'ingeni altament competitiu amb les fonts per assegurar-nos de que arribi al carrer la versió millor i més completa de la història, enlloc de la que els afavoreix".

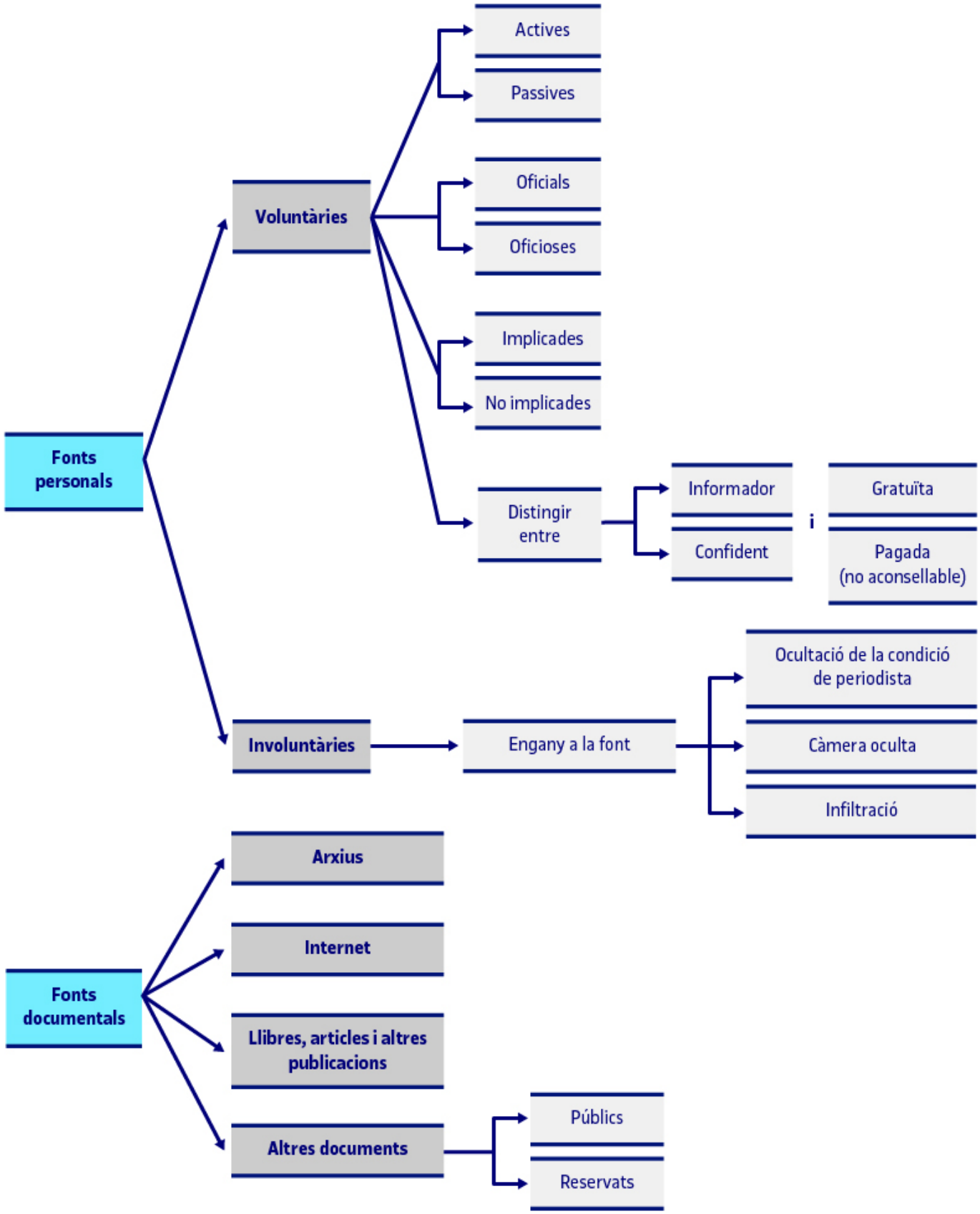
3. Fonts d'informació i mecanismes d'obtenció de dades

3.1. Les fonts periodístiques

3.1.2. Tipologia de les fonts d'informació

A continuació es presenta una classificació de les fonts d'informació. Tota la informació sbre la tipologia de les fonts s'ha extret de Javier Chicote (*).

Tipologia de les fonts d'informació en el periodisme d'investigació



Classificació de les fonts d'informació en el periodisme d'investigació.
Font: J. Chicote (2006).

3.1.2.1. Les fonts personals voluntàries

Són les fonts més familiars i ho abasteixen tot, des del portaveu oficial o els funcionaris i policies als que veiem amb regularitat, fins la persona amb què parlem potser una sola vegada. Si hi ha una regla d'or per l'èxit en les investigacions és no tenir mai por a preguntar. El pitjor que podem esperar és una negativa i, en canvi, en podem treure molt rèdit.

1) Segons la seva **activitat**, la font pot ser:

- Activa: Pren la iniciativa de posar-se en contacte amb el periodista i desitja transmetre la informació que posseeix.
- Passiva: No prenen la iniciativa de subministrar la informació, sinó que és el periodista qui s'hi posa en contacte.
- Oficials: Si es tracta de l'oficina de premsa o d'una veu autoritzada per una institució per proporcionar informació.
- Oficioses: També són individus que ocupen llocs de responsabilitat i que proporcionen informació. La principal diferència respecte les oficials és que aquestes segones no acostumen a acceptar que el periodista les citi com a font, volen mantenir-se en l'anonimat o l'off the record. Per norma general, les fonts oficials no acostumen a ser comprovades pel periodista mentre que les oficioses ho solen ser molt més.

2) Segons la **relació amb el fet** que es vol investigar:

- Implicades: Testimonis o protagonistes dels fets i, per tant, s'hi veuen afectats de manera directa. Dins aquesta tipologia hi ha les fonts favorables, que són aquelles que es veuen beneficiades per la publicació d'una investigació sobre els fets, i les fonts desfavorables, que són les que es veuen perjudicades si els fets adquireixen una rellevància púbica. Han de ser comprovades perquè la tendència serà sempre justificar la seva actuació o simplement negar-la, en cas que sigui incriminatòria.
- No implicades/ Alienes: Si bé aquestes fonts no tenen relació directa amb els fets que es volen investigar, poden oferir dades que enriqueixin la investigació. D'una banda hi ha les fonts tècniques, que són les que tenen una visió especialitzada sobre el tema i poden enriquir i aclarir la investigació; i de l'altra, les fonts neutrals, que no tenen relació directa amb els fets però també hi aporten dades, ja que poden conèixer la gent involucrada.

Per Chicote, cal distingir entre confident i informant. El confident té una relació habitual amb el periodista producte d'un tracte personal i prolongat, que genera una comunicació bidireccional, més fluida i abundant. Aquesta relació es basa en l'interès, l'amistat, el respecte mutu, els diners o qualsevol altre motiu. L'informant, en canvi, manté una relació ocasional amb el periodista, tot i que també pot ser motivada per diners o ser una filtració, és a dir, una informació que a la font li interessa propagar.

També cal distingir les fonts d'informació gratuïtes, les que no demanen una contraprestació per la informació facilitada, i les pagades, l'interès econòmic de les quals pot posar en tela de judici la fiabilitat del seu testimoni.

3.1.2.2. Les fonts personals involuntàries

No totes les fonts personals són voluntàries. Les involuntàries són aquelles que proporcionen informació sense saber com ni per a què, o el que és el mateix, són fonts a les quals el periodista enganya per obtenir informació. Això passa sempre que el periodista no s'identifica com a tal. Un exemple d'aquesta classe de fonts són aquelles que s'enfronten, sense saber-ho, a una càmara oculta. Aquest tipus de pràctiques per obtenir informació són reprovables des del punt de vista deontològic i ètic, i per això són molt difícils de justificar. Al respecte, Chicote considera que només es recorrerà a la càmera oculta "en virtut d'un bé major. El periodista pot enganyar un traficant de drogues per desemascarar la seva organització mafiosa i fer un bé a la societat, però per contra, no és lícit ocultar la condició d'informador per descobrir la vida privada d'un personatge".

3.1.2.3. Les fonts documentals

Són els arxius públics o privats, l'Internet, les bibliografies, les hemeroteques i, en definitiva, tota classe de dades a les quals es pugui accedir i tinguin relació amb el tema investigat. Moltes d'aquestes dades o l'accés a les mateixes pot ser proporcionat per les fonts personals, però en qualsevol cas es tracta d'infromacions que generalment són públiques o semipúbliques. Algunes de les plataformes documentals recents més conegudes són Wikileaks, Contrast o whistelblower. Tot i així, existeix una amplitud de recursos i bases de dades online on obtenir informació, com ja vam veure en el subapartat 1.3. Periodisme de dades.

Per la seva banda, David Randall estableix un seguit de pautes generals a l'hora de tractar les fonts (*); una proposta complementària a la classificació de Javier Chicote.

1) Fonts humanes

Per norma general, les fonts s'han d'identificar. Si s'hi neguen, podem esgrimir l'argument que la importància d'una notícia depèn de que es pugui verificar gràcies a que es coneix la seva procedència. Si malgrat tot una font decideix no identificar-se però nosaltres volem citar les seves declaracions, podem mirar de pactar una fórmula d'identificació genèrica, com per exemple "fonts

del Departament d'interior". Tot i així, el secret professional és un dret del periodista, a la vegada que un deure que garanteix la confidencialitat de les fonts d'informació. Per tant, el periodista garantirà el dret de les seves fonts informatives a mantenir l'anonimat si així ho han sol·licitat. No obstant, aquest deure professional podrà cedir excepcionalment en el supòsit que consti fefaentment que la font ha falsejat de manera conscient la informació o quan revelar la font sigui l'únic mitjà per evitar un dany greu i imminent a les persones.

2) Fonts oficials

Que siguin fonts autoritzades no sempre és sinònim de que siguin fonts ben informades quan es tracta de departaments de relacions públiques i premsa. Algunes coneixen bé les seves respectives organitzacions i estan en condicions d'aportar detalls. Altres són simples intermediaris, que transmeten les nostres preguntes a un funcionari i tornen amb respostes, amb la impossibilitat de formular noves qüestions al respecte. Per això convé cultivar els contactes amb funcionaris disposats a donar-nos informació directament, enlloc de la filtrada a través del seu encarregat de relacions públiques. Hi ha hagut moltes lamentacions per part dels periodistes respecte a la perversitat dels "experts en relacions públiques", coneguts com a *spin doctors*, de la "gestió de notícies".

a) *Spin doctors*. Poc a poc ha anat designant quasi tots els que s'ocupen de les relacions públiques o la gestió de la imatge. Un veritable spin doctor és algú procedent del món polític, que actua com a portaveu en cap de premsa o bé un ajudant amb prou estatus. A més de les seves declaracions rutinàries i conferències de premsa, s'ocupa d'interpretar els fets en favor del seu cap, intenta anticipar-se a les males notícies (i distreure l'atenció d'elles) i s'esforça per gestionar les expectatives de manera que beneficiï la seva causa. Els autèntics spin doctors estan a prop del seu cap, i tot el que diuen per publicar posseeix la major autoritat.

b) Gestió de les notícies. Les organitzacions tenen una gran quantitat de declaracions i informacions que volen donar a conèixer. No és particularment escandalós que escullin la manera i el moment més avantatjós per a fer-ho. De fet, és freqüent que aquesta "gestió de les notícies" sigui d'ajuda, perquè assegura que els periodistes disposin de temps suficient per redactar la informació per les seves edicions o evitar un conflicte entre la publicació d'un informe i un esdeveniment de importància. Però precisament perquè les organitzacions tenen interessos, poden actuar de tres maneres:

- Fer que l'hora de la declaració coincideixi amb el pitjor moment del cicle informatiu. D'aquesta manera és fàcil que la convocatòria sigui ignorada o tractada amb indulgència.

- Filtrar l'aspecte principal a un mitjà simpatitzant amb l'esperança que, quan els altres rebin la informació a través dels canals oficials, ja tinguin establerta la seva agenda.

- Endolcir la píndola amb algun fragment de bones notícies que després es converteix en l'argument principal de la convocatòria.

La única garantia contra aquesta i altres formes de gestió informativa és utilitzar el nostre ingeni.

3) Fonts no autoritzades

Es tracta d'algú no autoritzat per transmetre'ns la informació, des d'un denunciant ben situat en una organització a una persona sense connexió oficial amb la informació. Poden filtrar documents o simplement suggerir-nos que excavem en una àrea determinada. La informació així obtinguda serà desmentida, com a mínim en un primer moment, i és quasi segur que les nostres fonts no voldran que se les mencioni. Per tractar amb elles s'ha de tenir presents diverses coses:

- Preguntar-nos quins són els seus motius: Estan els que busquen perjudicar un rival polític, comercial o personal, impulsar una determinada causa, venjar-se o provocar problemes sense més. El desig de venjança d'una persona pot resultar d'utilitat, però és millor saber què pretén des del principi.

- Preguntar-nos (i esbrinar!) l'altra cara de la història: La informació que ens donen fonts polèmiques no acostuma a ser tan simple com volen que creiem. Estalviarem temps si descobrim l'altra cara, o algun matís de la informació que hauríem de conèixer. Tot el que publiquem s'ha de contrastar prèviament i per explicar una història de forma veraç és necessari conèixer les diferents versions de les persones que hi estan implicades.

- Demanar documents sempre que sigui possible: S'ha de demanar a les fonts si hi ha alguna documentació que recolzi el que ens estan explicant. Si no ens la volen donar, demanarem fotocòpies i en cas que també s'hi neguin, demanarem de llegir la documentació davant seu. Si obtenim els documents d'una manera extraoficial, és convenient assessorar-nos amb fonts expertes per assegurar-nos de la seva validesa.

Randall destaca els següents **consells bàsics** (amb aportacions de l'autora):

1) Ser just amb les fonts: Tractar-les amb justícia i no assetjar-les, enganyar-les o intimidar-les. El fonamental és ser justos. Si s'està criticant o acusant una font de joc brut, no només l'hem d'informar sinó també hem de donar-li l'oportunitat de replicar.

2) Aprovació del text: Mostrar-li a la font l'article acabat abans de la seva publicació és com convidar-la a que el censuri. La tasca del periodista consisteix en redactar informació precisa, no quelcom que pugui constituir una base de negociacions amb l'implicat. Tot i així, unes comprovacions sensates no representen una aprovació del text. De fet, quan la nostra història va sobre qüestions amb les quals no estem familiaritzats pot ser fins i tot un exercici de responsabilitat enviar part o la totalitat de l'article a un expert, per garantir que la història estigui lliure d'errors. En ocasions molt concretes, fins i tot és justificable enviar parcialment o totalment el contingut a una o vàries de les fonts citades, sobretot si el seu relat aborda qüestions particularment sensibles. Aquesta pràctica dóna peu a que les fonts puguin alterar el seu testimoni, així que no pot convertir-se en la dinàmica habitual i s'ha de delimitar. No és obligatori permetre aquesta revisió, depèn sobretot de l'empatia del periodista amb les fonts utilitzades.

Exemple

Cito un exemple. CRÍTIC va publicar un reportatge sobre la violència sexual envers les dones en el qual hi apareixia una font que havia patit una agressió d'aquest tipus en primera persona i estava disposada a explicar-ho. A l'hora de recollir el seu testimoni el mitjà va decidir entrevistar-la primer sense gravar el seu relat, per pactar amb la font quina informació volia que el reportatge inclogués i quina no. A l'hora de transcriure l'entrevista i redactar el tema, se li va enviar a la font els fragments on apareixia el seu relat dels fets. Es va tenir aquesta "deferència" amb la font perquè es va fer valer la gravetat del seu testimoni i les dificultats afegides d'expressar-lo en un mitjà. La voluntat del mitjà era assegurar-se el testimoni que se sentia còmode amb la manera com es recollien les seves paraules. CRÍTIC podia haver prescindit d'enviar-li aquests fragments d'informació a la font? Sí, la nostra autoritat com a periodistes ens legitima. Per tant, com apuntàvem a dalt, depèn de l'empatia, la sensibilitat i el compromís del periodista amb la font.

3) Retractació: Molt sovint una font diu una cosa a un periodista i acte seguit, lamenta haver-ho dit. En aquestes ocasions, hem de permetre que corregeixin les seves declaracions a menys que tinguem excel·lents motius per pensar que la voluntat de la font és obstruir la investigació. Si és així, deixarem constància tant de la seva primera declaració com de la segona, amb recursos del tipus: "En una primera entrevista, la persona 'x' explicava a aquest mitjà 'la primera versió', però també matisa que/ es retracta dient que 'la segona versió'". Així ens curem en salut sense faltar a la veritat. En última instància, si no tenim clar si publicar la primera versió o deixar-la fora del text, ho consultarem amb l'editor o el director del mitjà.

Ara bé, com a periodistes sempre hem de respectar "l'off the record", tot i que convé evitar-lo tan com sigui possible. A més, la norma a seguir és que una informació només es considerarà "off the record" quan les dues parts ho hagin acordat així des del principi. Hi ha molt poques ocasions on estigui justificat publicar una informació quan la font ha precisat que ens l'explica fora de l'entrevista. D'altra banda cal saber utilitzar l'off the record per tirar del fil. A vegades les fonts ens revelen una informació "fora de gravació" que ens pot servir per iniciar una investigació, tot i que no citem la persona que ens ha donat aquesta informació.

4) Temptativa d'impedir l'accés: A vegades, la cobertura de la premsa disgusta tan a les organitzacions que ens tallen tot accés a la informació. S'hauria d'informar els lectors de qualsevol intent de manipulació i fomentar la resposta d'altres mitjans.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.2. Introducció: Les dades com a font d'informació

Un conjunt de dades o un *data set* –el terme anglosaxó per a denominar-ho– és un recull d'informació estructurada en columnes i files que pot ser processada i analitzada. Una font periodística, més que qualsevol altra, pot revelar informació clau i context sobre una realitat concreta que vulguem cobrir periodísticament. El més important és fer les preguntes adequades i conèixer les eines que permeten «entrevistar» un full de càlcul per tal que les conclusions extretes puguin presentar un fet noticiable.

Periodisme de dades o dades amb periodisme. L'escenari tecnològic ha obert un ventall molt ampli de possibilitats de fonts d'informació que fa trenta anys era impensable tenir a l'abast. Parlem de dades emmagatzemades en fulls de càlcul, i de dades que, tot i no estar ordenades en files i columnes, processant-les en un full de càlcul es converteixen en un material potencialment publicable.

El periodisme de dades és el periodisme de precisió de tota la vida, que en el context actual se serveix d'unes eines informàtiques d'anàlisi i de visualització que permeten analitzar i extreure conclusions d'un volum de dades que fa un temps hagués estat inimaginable, així com poder mostrar-les al lector.

La delinqüència en una ciutat, les hectàrees d'un bosc cremat, les epidèmies, les votacions parlamentàries i el resultat d'unes eleccions possiblement no són més que números. És cert, però també són xifres que parlen de la realitat política, social i econòmica que ens envolta, i que com a periodistes volem explicar en forma d'històries en un diari, televisió o ràdio.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.3. Bases de dades disponibles

3.3.1. Introducció

Abans de començar a abordar qualsevol tema, és recomanable preguntar-se quines dades hi ha disponibles sobre la qüestió que interessa cobrir. Conèixer les fonts públiques disponibles facilitarà i accelerarà moltíssim el procés de cerca de les fonts.

La quantitat d'informació pública i disponible en format lliure no ha deixat de créixer en els últims anys. Aquest creixement ha tingut molta relació amb l'esforç que s'ha fet des de diferents organitzacions en favor de la transparència i l'obertura de dades (*open data*).

3. Fonts d'informació i mecanismes d'obtenció de dades

3.3. Bases de dades disponibles

3.3.2. El moviment Open Data

3.3.2.1. Introducció

El concepte de dades obertes és una pràctica que busca que la informació pública estigui disponible i en formats oberts o reutilitzables (Excel o CSV, entre d'altres). Sense restriccions d'autor, lliure de patents i sense cap tipus de mecanisme de control, parteix d'una base similar de la qual sorgeixen altres moviments tecnològics com l'*Open Source* (codi obert) i el programari lliure.

Tant a Catalunya com a escala estatal, el moviment Open Data ha estat integrat, entre d'altres perfils professionals, per periodistes, que han vist en les dades obertes amplíssimes possibilitats periodístiques. També han estat promotors d'aquest canvi perfils com els enginyers, les entitats públiques, els desenvolupadors i els investigadors.

En aquest sentit, les administracions públiques han hagut de córrer per a fer la renovació de formats i posar a disposició de la ciutadania, i en formats oberts, la informació pública. En els darrers anys la informació disponible ha augmentat de forma exponencial tant en el pla estatal com autonòmic i municipal. I, malgrat que encara queda molta feina per a fer en tots els estrats de les institucions, actualment ja hi ha repositoris públics de dades que val la pena tenir en compte.

3.3.2.2. Catalunya

1) Dades obertes Gencat

És el portal de la Generalitat de Catalunya, gestionat des de la Secretaria de Transparència i Govern obert del Departament d'Afers i Relacions Institucionals i Exteriors i Transparència, en el qual hi ha publicades les dades de caràcter públic, amb l'objectiu de fomentar l'ús i la reutilització de la informació provinent de l'administració. Les dades provenen de diferents organismes públics de la Generalitat, estan agrupades en el catàleg de dades i disponibles en diferents formats, la majoria dels quals són estàndard, de manera que es poden reutilitzar fàcilment.

2) Dades obertes de la Diputació de Barcelona

En aquest portal, la Diputació de Barcelona ofereix un catàleg de dades públiques susceptibles de ser reutilitzades per la ciutadania i les empreses. Tracta des de qüestions més municipals fins a contractació en l'àmbit de la diputació, i és un magnífic recull del que passa en el pla de la província de Barcelona.

3) Transparència de la Generalitat de Catalunya

Portal que recull tots els documents relacionats amb l'exercici de la transparència en els diferents Departaments de la Generalitat. Per a una correcta utilització, cal navegar per les pàgines dels departaments per a veure quines dades hi ha publicades, i com s'estructuren en cada cas.

4) Institut d'Estadística de Catalunya (IDESCAT)

Recull de dades estadístiques de Catalunya, sobretot de l'àmbit municipal. Hi ha moltes més dades que a l'Instituto Nacional de Estadística (INE). Cal destacar que les dades municipals de l'IDESCAT són molt més detallades de les que hi ha a l'INE. Per exemple, l'aplicació «Catalunya en Xifres» permet desglossar moltes variables de l'àmbit municipal.

5) Centre d'Estudis d'Opinió

Recull de les dades dels estudis d'opinió a Catalunya. No té res a veure amb l'homònim espanyol, el Centro de Investigaciones Sociológicas. No comparteixen estudis ni variables, i serveix únicament en cas que l'objectiu periodístic se centri en territori català.

6) Open Data Barcelona

Recull de dades obertes de l'Ajuntament de Barcelona. En els últims mesos el portal ha renovat molt el conjunt de dades obertes, les ha actualitzat i n'ha carregat de noves.

7) Dades obertes Girona

Recull de dades obertes de l'Ajuntament de Girona que tracta des del nombre de parades de taxis fins a l'enllumenat públic. La quantitat d'àmbits que s'aborden és innumerable.

3.3.2.3. Espanya

En el pla estatal cal tenir en compte que per a totes aquelles competències descentralitzades a les comunitats autònomes, no hi ha dades centralitzades i homogeneïtzades. Això pot fer que, en alguns casos, hi hagi buits d'informació.

És el cas, per exemple, de les dades d'ajuda a la dependència. Cada comunitat autònoma és l'encarregada de posar a disposició del ciutadà la informació respecte a les ajudes. Si volem dur a terme un treball periodístic sobre l'aplicació de la llei de la dependència, ens trobarem que no només hem d'anar a buscar les dades en disset fonts diferents, sinó que, a més a més, aquestes estan organitzades de disset maneres diferents i, que moltes vegades, ni tan sols estan homogeneïtzades.

De totes maneres, tret dels àmbits que són competència de les comunitats autònomes, l'Estat té a la seva disposició una quantitat molt extensa d'informació i de dades que ens pot ser molt útil com a periodistes.

1) Govern Obert d'Espanya

Conjunt de dades obertes dels diferents ministeris, municipis i institucions municipals i autonòmiques. En aquest portal es carreguen de forma automàtica els *datasets* públics de les diferents administracions de l'Estat. Té l'inconvenient que els filtres de càrrega són lents, però serveix per a situar quantes dades hi ha sobre un tema concret i anar fins a la font original.

2) Institut Nacional d'Estadística

Recull de dades estadístiques d'Espanya. Des de variables econòmiques, demogràfiques fins a estils de vida. En l'àmbit periodístic són especialment interessants les «microdades», desglossades per municipi, que permeten aprofundir molt en l'anàlisi. En general és fàcil que qualsevol tema que es vulgui abordar tingui dades disponibles a l'INE, per tant, està bé fer-hi un cop d'ull abans de posar-se a escriure.

3) Ministeri de l'Interior

Conjunt de dades estadístiques que són competència d'aquest ministeri: institucions penitenciàries, trànsit, asil, protecció civil, terrorisme i seguretat ciutadana. La publicació de les dades va amb un any d'endarreriment.

4) Ministeri de Sanitat, Serveis Socials i Igualtat

Conjunt de dades estadístiques que són competència d'aquest ministeri: violència masclista, inversió en sanitat, recursos per a la dependència. Cal tenir en compte, de totes maneres, que tant en el pla sanitari com de serveis socials les competències estan descentralitzades a les comunitats i, per tant, pot ser que hi hagi una manca d'informació homogeneïtzada de tot Espanya.

5) Ministeri de Defensa

Conjunt de dades estadístiques respecte a tot l'aparell militar de l'Estat i política de defensa. Estan lleugerament desactualitzades.

6) Ministeri d'Agricultura i Pesca, Alimentació i Medi Ambient

Conjunt de dades estadístiques que són competència d'aquest ministeri: pesca, agricultura, desenvolupament rural, ramaderia, observatori de preus, baròmetre del clima, consum alimentari, indústries i medi ambient.

7) Ministeri d'Economia, Indústria i Competitivitat

Dades macroeconòmiques i recull d'altres portals d'interès relacionats amb les competències d'aquest ministeri.

8) Ministeri de Foment

Conjunt de dades estadístiques que són competència d'aquest ministeri: construcció, mobilitat, habitatge, actuacions urbanístiques, transport i infraestructures estatals.

9) Ministeri d'Educació, Cultura i Esport

Conjunt de dades estadístiques que són competència d'aquest ministeri: ensenyament universitari, educació primària i secundària, museus, teatres, cinemes, esports i competicions.

10) Ministeri d'Energia, Turisme i Agenda Digital

Conjunt de dades estadístiques que són competència d'aquest ministeri: principals indicadors conjunturals com la producció industrial, el consum d'energia, i l'Índex de Preus al Consum (IPC).

11) Poder judicial

Recull de dades estadístiques del poder judicial: activitats dels òrgans, resolucions dels diferents tribunals, dades estadístiques dels jutjats penals, civils i laborals.

12) Recaptació i estadístiques del sistema tributari espanyol

Conjunt de dades estadístiques del sistema tributari: recaptació de l'estat amb detall de cada comunitat, indicadors fiscals, estadístiques sobre l'Impost sobre la Renda de les Persones Físiques (IRPF) i estadístiques cadastrals.

13) DataComex

Recull de dades estadístiques del comerç exterior de l'Estat: tipus de mercaderies, comerç entre països. Dades detallades per país, mes, any, lliurament i tipus de producte.

14) Banc d'Espanya

Recull de dades estadístiques del Banc d'Espanya: indicadors econòmics, fons d'inversió, dades històriques relacionades amb el sector financer, balanços, enquesta financera de les famílies, préstecs bancaris i estadístiques del sector exterior.

15) Centre d'Investigacions Sociològiques (CIS)

Recull de l'evolució del Baròmetre d'Opinió de l'Estat, i estudis concrets sobre temes específics que tenen rellevància en un moment determinat.

3.3.2.4. La Unió Europea

1) Eurostat

Dades homogeneïtzades del conjunt de països de la Unió Europea referents a qualsevol àmbit: economia, política exterior, medi ambient, producció industrial, condicions de vida, agricultura, transport i demografia, entre molts altres aspectes. És una font imprescindible a l'hora d'agafar referències sobre un tema.

2) Dades econòmiques de la Comissió Europea

Dades homogeneïtzades sobre l'economia de la Unió Europea: estadístiques macroeconòmiques, financeres, principals indicadors de la zona euro, índexs monetaris, evolució de preus i competitivitat.

3) Dades de salut pública de la Comissió Europea

Dades sobre indicadors de salut pública a escala europea, desenvolupament sostenible, lesions, demografia, alimentació i condicions de vida.

3.3.2.5. Dades globals

1) Les Nacions Unides

a) Alt Comissionat de les Nacions Unides per als Refugiats (ACNUR)

La millor base de dades actual sobre els moviments migratoris, sol·licituds d'asil acceptades, denegades, en espera, i tot desglossat per país d'entrada i de petició. Inclou també informació estadística detallada sobre els llocs que estan en estat d'emergència.

b) UN Comtrade

Accés gratuït a dades detallades del comerç mundial. L'UN Comtrade és un dipòsit de les estadístiques oficials del comerç internacional i les taules analítiques rellevants. Totes les dades són accessibles per mitjà d'una API pròpia.

c) Estadístiques de gènere

Pretén ser una guia nacional i internacional d'estadístiques de gènere. És un conjunt de 52 indicadors quantitatius i 11 variables qualitatives sobre temes rellevants relacionats amb la igualtat de gènere i l'apoderament femení.

d) Objectius per al desenvolupament sostenible

Aquesta plataforma proporciona accés a dades recopilades pel Sistema de les Nacions Unides.

e) UN Data

Recopilatori de dades de tots els temes que es tracten a les Nacions Unides. La pàgina web està una mica desactualitzada, i això pot dificultar la cerca de documents.

f) FAOSTAT Data

Recopilatori de dades que utilitzen les Nacions Unides destinada a l'alimentació i a l'agricultura: comerç d'aliments, producció, malnutrició, evolució dels preus i fam, entre moltes d'altres dades.

g) AIDSinfo

Recull de dades de l'impacte de la SIDA arreu del món, amb detall de cada país, que inclou mapes, dades sectorialitzades per sexe, edat i condicions socioeconòmiques. També permet l'anàlisi en un pla interestatal.

h) Food Security

Les dades de classificació de la seguretat alimentària de FEWS NET, que es remunten al mes de juny de 2009, estan disponibles per a la descàrrega com a fitxers i imatges de formularis geogràfics regionals. Aquestes dades es publiquen cada quatre mesos.

2) Organitzacions internacionals

a) World Bank Open Data

La iniciativa Open Data del Banc Mundial té com a objectiu proporcionar a tots els usuaris accés a dades del banc. El catàleg de dades és una llista de conjunts de dades disponibles del Banc Mundial, que inclouen bases de dades, taules preformatades, informes i altres recursos.

b) CIA World Factbook

La secció de la pàgina de la CIA, The World Factbook, proporciona informació sobre temes relacionats amb la història, la gent, el govern, l'economia, la geografia, les comunicacions, el transport, els assumptes militars i els assumptes transnacionals per a 267 entitats mundials.

c) Our World in Data

Pàgina web de la Universitat d'Oxford que pretén mostrar com les condicions de vida estan canviant. Té com a objectiu donar una visió global, i mostrar canvis a llarg termini. Proporciona tant visualitzacions de dades com conjunts de dades reutilitzables.

d) The Humanitarian Data Exchange

La plataforma d'intercanvi d'informació humanitària (HDX) és una plataforma oberta per a compartir dades que es iniciar el juliol de 2014. L'objectiu d'HDX és fer que les dades humanitàries siguin fàcils de trobar i d'utilitzar per a l'anàlisi. La col·lecció de dades ha rebut l'accés d'usuaris de més de 200 països i territoris.

e) Fons Monetari Internacional

Dades financeres de la gran majoria de països: crèdits concedits per part del Fons Monetari Internacional i la situació fiscal, entre moltes d'altres dades macroeconòmiques.

3) Bases de dades per àmbits

a) Flight Radar

El Flightradar24 és un servei global de seguiment de vols que proporciona informació en temps real i històric sobre milers d'avions arreu del món.

b) Aviation Safety Network

Base de dades que acumula informació sobre tots els accidents d'avió que hi ha hagut. S'actualitza al moment i és molt completa, amb un Excel per a cada any des del 1919.

c) Marine Traffic Database

Visualitzacions i dades interactives sobre el trànsit de vaixells arreu del món. Informació que es pot desglossar per port, vessaments i dades històriques detallades.

d) Historical Index of Economic Liberty (HIEL)

Conjunt de dades històriques sobre economia desglossades per any, per regió i per país. Les dades es poden descarregar en format reutilitzable i la recollida compta amb una metodologia molt rigorosa.

e) Consortium of Spatial Information

Informació i mapes sobre climatologia global del planeta Terra. Conté dades històriques i amb formats que permeten la visualització amb eines de tractament cartogràfic.

f) Code for Africa – Source Africa

Source Africa és el repositori principal d'Àfrica de «documents processables», és a dir, documents que el públic necessita per a ajudar a prendre decisions més informades i que pretén ajudar els organitzadors civils, mitjans de comunicació, ONG o organitzacions de justícia social a demostrar el seu reportatge. També inclou documents que han estat suprimits o censurats en altres llocs. És una plataforma gestionada per Code for Africa.

g) TeleGeography

És una empresa privada que es dedica a generar mètriques detallades que no estan disponibles en cap altre font. També proporcionen dades obertes geolocalitzades de la xarxa de telecomunicacions global i de la indústria, en general.

h) Google Trends

Tendències de cerques en línia a Google segmentades a escala global, per país i per llengua. També permet buscar una paraula concreta i veure quin ha estat el nombre de cerques en el buscador.

i) Wayback Machine

És una llibreria oberta formada per milers de llibres, pel·lícules, programari lliure, música i pàgines web.

j) SIPRI Military Expenditure Database

Es tracta d'una base de dades pública que conté la despesa militar de tots els estats durant el període comprès entre els anys 1949 i 2016.

k) Global Terrorism Database

És una base de dades de lliure accés en la qual es registren tots els atemptats arreu del món. Per a cada un es desglossa la informació sobre la localització, el nombre de víctimes, l'organització perpetradora i la data, entre d'altres detalls.

l) Foreign Trade – U.S.

Aquesta pàgina web registra tot el comerç exterior dels Estats Units. És una base de dades d'accés lliure que permet filtrar la informació per país i extreure-la en el format desitjat.

m) CERAC

En l'apartat de recursos d'aquesta pàgina web hi ha un llistat de bases de dades dels conflictes que hi ha arreu del món, amb els enllaços corresponents.

n) SICE

Dades sobre comerç entre els països que conformen l'Organització dels Estats Americans (OEA). Inclou acords de comerç i volum segmentat per països i tipus de producte.

o) Asian Development Bank

Recull d'estadístiques relacionades amb el desenvolupament dels països asiàtics: dades econòmiques, socials i indicadors financers relacionats amb la tasca de Banc Asiàtic de Desenvolupament.

p) Inter-American Development Bank

3. Fonts d'informació i mecanismes d'obtenció de dades

3.4. La llei de transparència

3.4.1. Introducció

Si cap de les fonts públiques disponibles reuneix les dades desitjades sempre queda l'alternativa de formular una petició d'informació pública.

Es tracta d'un mecanisme que disposa l'Estat per tal de demanar accés a qualsevol informació pública que estigui en mans de les institucions. És un mecanisme que a Catalunya es regeix per la Llei 19/2014, de 29 de desembre, de transparència, accés a la informació pública i bon govern, i a Espanya per la Llei 19/2013, de 9 de desembre.

Ambdues lleis consideren informació pública tots aquells continguts o documents, sigui quin sigui el format o el suport, que estan en mans de les administracions públiques.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.4. La llei de transparència

3.4.2. La llei de transparència a Espanya i a Catalunya

La Generalitat de Catalunya es va afegir a la transparència un any després de l'entrada en vigor de la Llei de transparència i bon govern d'Espanya, l'LTBG, que va entrar en vigor el desembre de 2013. Espanya ja va fer tard, tan tard, que quan es va decidir a fer-ho era l'únic país europeu de més d'un milió d'habitants que encara no tenia una llei que reconegués el dret d'accés a la informació.

La pressió exercida per part de la Coalició ProAcceso, una organització formada per una seixantena d'organitzacions i actors de la societat civil, va ser fonamental perquè finalment el govern popular comencés a redactar el projecte de llei.

Actualment, tant la legislació catalana com l'espanyola tenen algunes mancances. En concret, una que afecta el més bàsic: Espanya no reconeix el dret a la informació pública com un dret fonamental.

Per tant, per a formular una petició, tant si és en el portal de transparència català com en el de l'Estat, la persona ha d'identificar-se prèviament com a ciutadà. En el cas de Catalunya, es pot fer mitjançant el certificat digital –la signatura electrònica estatal– o amb la contrasenya IdCAT Mòbil.

Tant l'un com l'altre han de aconseguir-se de forma presencial anant, en el primer cas, a les dependències de l'Administració General de l'Estat i, en el segon, a una de les 38 oficines acreditades per la Generalitat.

Tot i que no hi ha dades disponibles sobre quin percentatge de la població disposa de certificat electrònic ni d'IdCAT Mòbil, la complexitat de la identificació xoca amb un dels fonaments del dret d'accés a la informació: fer peticions hauria de ser un procés senzill, ràpid i gratuït. Gratuït ho és, però no és ni senzill ni ràpid, un dels principis bàsics que l'organització ProAcceso reclama per al correcte funcionament de la transparència.

Un cop feta la petició, l'administració competent té un termini màxim de 30 dies per a respondre, que en el cas del conjunt estatal es pot endarrerir fins a 15 dies més.

Un dels grans avantatges que ofereix la llei catalana respecte a l'espanyola és el silenci administratiu positiu. Si l'administració pública catalana no respon al ciutadà, automàticament se li han de facilitar les dades requerides. En canvi, en el cas de la legislació espanyola, aquest silenci s'interpretarà automàticament com una denegació.

Si la petició d'informació és denegada o les dades obtingudes no són satisfactòries, es pot recórrer a un òrgan independent que avaluï de forma externa si s'ha de facilitar, o no, la informació requerida. En aquest sentit, la llei catalana torna a ser més favorable per al ciutadà.

La Comissió de Garantia del Dret d'Accés a la Informació Pública (GAIP), l'òrgan que vetlla pel compliment del dret i l'encarregat de resoldre totes les reclamacions, és elegit pel Ple del Parlament, i els membres tenen dedicació exclusiva per a dur a terme aquesta tasca. A diferència del Consell de Transparència i Bon Govern, l'organisme homòleg estatal, els membres exerceixen altres funcions més enllà de la pertinença a aquest ens, i són triats pel govern.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.4. La llei de transparència

3.4.3. Com formular una petició d'informació

3.4.3.1. Concreció

Perquè una sol·licitud es pugui resoldre amb èxit és imprescindible ser molt concrets a l'hora de demanar la informació: quines dades concretes es demanen, les variables determinades, durant quin període de temps i per a quin àmbit geogràfic.

Exemple

Si demanem el nombre d'accidents a les carreteres catalanes des de l'any 2000 fins al 2017, és molt fàcil que ens acabin lliurant un full de càlcul amb un nombre d'accidents per any. Per tant, si es pretén anar un pas més enllà en l'anàlisi, caldrà especificar exactament quina informació volem de cada accident, com per exemple, la localització (latitud i longitud), el nombre de morts, el nombre de ferits, el nombre de vehicles accidentats i l'hora de l'accident. També és important concretar tant per a quin període de temps requerim la informació, com l'àmbit territorial en el qual estem interessats.

Un altre element que pot facilitar encara més la tasca és ajustar la petició a la forma com la institució pública té emmagatzemades les dades. Les notes de premsa, els anuaris i els comunicats donen pistes de la manera com cada institució té emmagatzemades i estructurades les bases de dades. Donar un cop d'ull a tota la documentació abans de formular la petició pot ajudar a encertar el tir i augmentar les possibilitats d'èxit.

3.4.3.2. Format

De la mateixa manera, també és imprescindible especificar el format mitjançant el qual es desitja rebre les dades. Les administracions públiques sovint estan acostumades a treballar amb formats tancats i no reutilitzables com ara el PDF o el PDF d'imatge. Aquests tipus de formats dificulten i alenteixen la tasca periodística d'analitzar les dades i extreure conclusions, per això és del tot recomanable especificar el format en la petició d'informació.

3.4.3.3. Basada en la llei

Tot i que, en cap cas, no cal justificar el motiu pel qual demanem la informació, sí que va bé mencionar l'article de la llei que engloba que les dades que demanem estan emparades pel dret d'accés a la informació pública. Tot i que no és obligatori, el coneixement de la llei pot augmentar les possibilitats d'èxit de la petició.

Exemple

En el cas de la llei estatal, la petició es podria formular de la següent manera:

«Es tracta d'una petició que pretén aconseguir sota l'empara de la llei estatal l'evolució del preu del transport públic a la ciutat de Sevilla.»

En el cas de la legislació catalana, un bon exemple podria ser:

«Es tracta d'una sol·licitud d'informació dirigida al Departament de Salut, que pretén obtenir el registre de contractació públic de material sanitari dels anys anteriors al 2011. Dels anys posteriors a aquesta data, el registre de contractació pública conté tota la informació pertinent a la contractació major efectuada per qualsevol dels departaments de la Generalitat de Catalunya.»

3.4.3.4. La conveniència de dividir, o no, una petició

Segons el volum de dades que es vulguin demanar, cal considerar si convé, o no, dividir la petició en dues o més sol·licituds diferents. Per exemple, si les dades que interessen fan referència als sous que reben els alts càrrecs de quatre empreses públiques diferents, serà més adequat formular una petició per cadascuna de les empreses, en comptes d'una de sola.

En aquest cas, el motiu rau en el fet que cada empresa té una gestió i unes especificacions diferents i, per tant, és fàcil que en no poder accedir a les dades d'una empresa concreta, se'ns denegui l'accés a totes. Si fem una petició per cada empresa, serem més concrets i més directes sobre la informació que sol·licitem, i serà més difícil que la institució pública la denegui per omissió.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.5. La mineria de dades

3.5.1. La mineria

Fins ara hem fet un repàs de les bases de dades disponibles i dels mecanismes d'accés a la informació pública que hi ha actualment. Tanmateix, les possibilitats de treballar amb dades s'estenen més enllà de les vies esmentades.

Hi ha multitud de conjunts de dades, que tot i no estar estructurats com a tals, ens poden ser útils a l'hora de dur a terme un projecte periodístic. És el que es coneix com a *scrapping* o mineria de dades.

Es tracta d'una tècnica per a descarregar, de forma automatitzada, informació emmagatzemada en pàgines web per a estructurar-la en un full de càlcul o base de dades. Un exemple seria el cas de les dades dels parlamentaris del Congrés de Diputats. L'estructura de la pàgina web permet la descàrrega automàtica de les dades en format de full de càlcul sense dependre que la institució pública la posi al servei de la ciutadania.

Per tal de dur a terme aquestes descàrregues, normalment s'utilitzen llenguatges de programació com l'R o el Python. Tenint en compte que la gran majoria dels i de les periodistes no tenen prou coneixements per a fer servir aquestes eines, hi ha altres alternatives com ara aplicacions informàtiques que d'una manera més senzilla, tot i que amb certes limitacions, permeten que el periodista pugui autogestionar les descàrregues.

És el cas d'eines com Import.io, o els fulls de càlcul de Google.

3. Fonts d'informació i mecanismes d'obtenció de dades

3.5. La mineria de dades

3.5.2. Eines

Hi ha una sèrie d'eines i d'aplicacions informàtiques gratuïtes que permeten de manera senzilla poder descarregar informació d'una pàgina web i emmagatzemar-la en fulls de càlcul de forma automàtica. Tot seguit fem una llista d'algunes de les més utilitzades per la seva practicitat i senzillesa:

1) Import.io

És una eina en línia que, sense necessitat de descarregar en el dispositiu, permet de manera fàcil i senzilla la descàrrega de dades. Només requereix la creació d'un compte de manera gratuïta i connexió a internet per a poder utilitzar-la. El format gratuït accepta fins a cent descàrregues automatitzades al mes. Si se'n volen efectuar més, caldrà esperar al mes següent, o subscriure's a la versió *premium*.

2) Full de càlcul de Google

L'eina de gestió de fulls de càlcul de Google conté fórmules preparades per a descarregar contingut emmagatzemat en les pàgines web, sigui quin sigui el format d'origen: XML, HTML, CSV o TSV. Es tracta d'una sèrie de fórmules que s'han d'escriure en la primera casella en la qual es vol que s'importin les dades. Per al seu correcte funcionament cal seguir la sintaxi de la fórmula:

- IMPORTXML (url, xpath_query)
- IMPORTHTML (url, query, index)
- IMPORTDATA (url)

3) Web Scraper

Es tracta d'un complement del navegador Google Chrome que permet automatitzar la descàrrega d'informació de les pàgines web. Per a poder utilitzar-lo, primer cal instal·lar el complement en el navegador. A continuació, apareixerà una icona a la part superior dreta de la pantalla. Cada cop que l'haguem d'utilitzar, farem clic a la icona amb la pàgina web oberta. A partir d'aquí, només caldrà seguir els passos que indica el complement.

4) ParseHub

És una aplicació informàtica que es pot descarregar tant en PC com en dispositius Macintosh, i que permet de manera senzilla la descàrrega automàtica de dades. Es tracta d'una eina gratuïta que, un cop après el funcionament, permet de manera eficaç aconseguir el seu propòsit. Actualment és una de les eines més potents, ja que sigui quina sigui l'estructura de la pàgina, acostuma a extreure amb èxit les dades.

4. Anàlisi de dades

4.1. Eines per a l'anàlisi de dades

4.1.1. Gestors de fulls de càlcul

4.1.1.1. Introducció

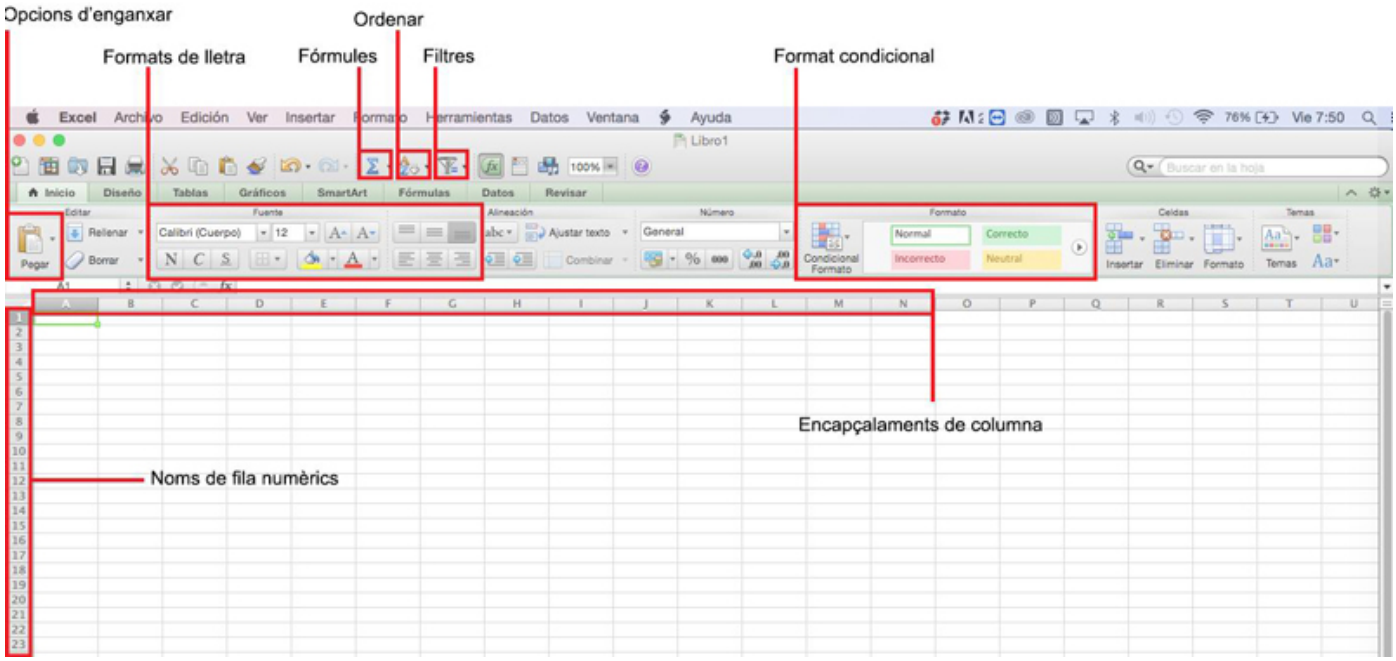
Hi ha eines diverses per a tractar l'anàlisi de dades. En línies generals, les dades que haurem de tractar estan emmagatzemades en fulls de càlcul que poden analitzar-se mitjançant programes de gestió de fulls de càlcul com ara Microsoft Excel, Google Spreadsheets o Open Calc, la versió de programari lliure.

En els següents apartats, s'explicarà per tutorials i per captures de pantalla com gestionar una anàlisi de dades amb l'eina Microsoft Excel, ja que és la més comú i potent.

4.1.1.2. Microsoft Excel

És un programa de gestió de fulls de càlcul que forma part del paquet Office. Microsoft va llançar la primera versió d'aquesta eina l'any 1985, i des d'aleshores s'ha popularitzat com l'eina per excel·lència de gestió de base de dades. La seva interfície és intuïtiva i coneguda per milers d'usuaris arreu del món.

També serà el programa que utilitzarem com a eix central per a veure com netejar i analitzar un conjunt de dades. Tots i cadascun dels passos que es detallen a continuació se seguiran mitjançant el programa Microsoft Excel. De moment, però, cal fer una presentació oficial de la interfície:

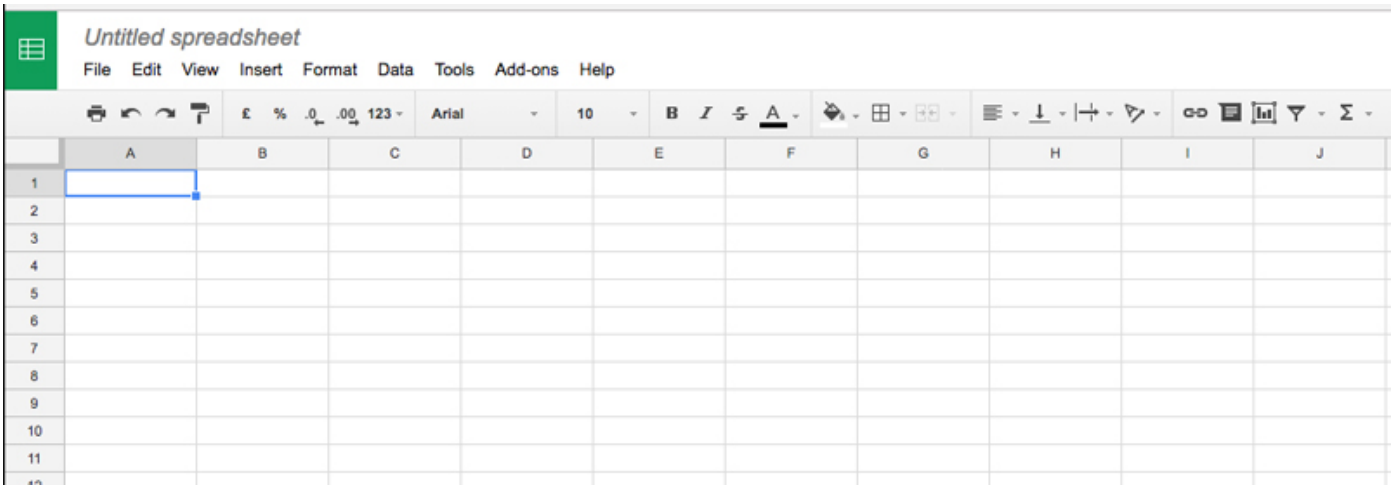


Interfície de Microsoft Excel

4.1.1.3. Google Spreadsheets

Creat l'any 2006 per Google, és l'homòleg del Microsoft Excel. Tot i que la interfície és menys coneguda i afegeix variacions respecte al programa de Microsoft, té l'avantatge de ser un programa gratuït al qual tothom pot accedir per mitjà d'un compte de Google. A més a més, l'aplicació en línia també permet que diferents usuaris puguin treballar amb el mateix full de càlcul alhora.

La contrapartida és que accepta un número de files considerablement inferior al Microsoft Excel, de manera que per a grans fulls de càlcul no hi ha una altra alternativa que utilitzar l'Excel.



Interfície de Google Spreadsheets

4.1.1.4. LibreOffice Calc

Es tracta d'un programa de gestió de fulls de càlcul de lliure utilització. Com tot el paquet de programari lliure, el LibreOffice està disponible per a totes les plataformes: Mac, Microsoft i Linux, entre les més conegudes. Tot i que ha millorat molt respecte a les versions anteriors, a l'hora d'utilitzar determinades fórmules hi ha certes limitacions, ja que no ofereix el mateix ventall que Microsoft Excel o Google Spreadsheets.

Aquests són els programes de gestió de fulls de càlcul més utilitzats avui dia. Ara bé, quan som davant d'un conjunt de dades molt més gran emmagatzemat en una base de dades, haurem d'optar per eines més completes que permetin gestionar un volum més gran.

4. Anàlisi de dades

4.1. Eines per a l'anàlisi de dades

4.1.2. Eines avançades

4.1.2.1. SQL (Structured Query Language)

Es tracta d'un llenguatge que permet modificar, consultar i realitzar operacions sobre bases de dades relacionals. Respecte a d'altres llenguatges, té una sintaxi senzilla i intuïtiva. Es pot fer funcionar directament des de la consola o terminal de l'ordinador, o per mitjà d'algun programa que doti el llenguatge d'una interfície que permeti treballar-hi de manera més senzilla.

4.1.2.2. Python

És un altre llenguatge de programació que també permet treballar amb bases de dades. Més enllà de l'anàlisi de bases de dades, té molts altres usos, i això el converteix en un llenguatge molt més potent i útil que l'SQL. Tot i així, la sintaxi és molt més complexa, i aprendre'n el funcionament requereix més especialització per part del periodista.

4.1.2.3. R

Aquest tercer llenguatge de programació, originalment parteix del camp de l'estadística, per això s'utilitza molt tant en l'àmbit acadèmic com en el periodístic. Igual que Python, és una eina que ofereix unes funcionalitats que van més enllà de l'accés a les bases de dades: també permet fer anàlisis visuals, elaborar gràfics, anar un pas més enllà en mètriques estadístiques i en l'automatització de descàrregues de dades.

4. Anàlisi de dades

4.2. La importància d'ordenar bé un full de càlcul

4.2.1. L'estructura de les dades

L'estructura del *data set* o conjunt de dades ha de respondre únicament a la voluntat periodística. Abans de començar a recollir les dades és imprescindible plantejar-se quines preguntes volem que respongui el conjunt de la informació. Segons «l'entrevista» que vulguem fer en el full de càlcul, caldrà incloure uns camps o uns altres, i estructurar les files i les columnes d'una manera o d'una altra. Es tracta d'un plantejament previ que cal fer tant si les dades han estat obtingudes per una font tercera, per exemple l'Administració pública, com si han estat recollides manualment pel periodista.

En el cas de dades provinents de fonts terceres, a més a més, caldrà tenir en compte que sovint estan estructurades segons els interessos i l'objectiu de la font, no des d'un punt periodístic. Per això també és important replantejar-se si és pertinent, o no, reestructurar les columnes i les files en funció de les necessitats periodístiques, i no només això, sinó que un cop comprovat que tots els camps siguin correctes, s'haurà de «netejar» i disposar la informació de forma adequada per a poder analitzar-la correctament.

4. Anàlisi de dades

4.2. La importància d'ordenar bé un full de càlcul

4.2.2. Netejar un full de càlcul

Netejar unes dades vol dir posar-les de forma adequada per tal que es puguin analitzar mitjançant programes de gestió de fulls de càlcul, com ara el Microsoft Excel o el Google Spreadsheets, o de gestió de base de dades com l'SQL o el Microsoft Access.

Procedir directament a l'anàlisi sense haver fet una neteja de les dades pot conduir el periodista a extreure conclusions del tot errònies. A l'hora d'efectuar aquesta neteja hi ha algunes comprovacions que pràcticament es poden aplicar en la totalitat dels casos.

- Eliminar tots els elements complementaris: files, columnes i imatges incrustades extres. Quan s'utilitzen dades de l'Administració pública sovint s'incrusta la imatge de la entitat corresponent i es posen capçaleres extres en les files superiors. Per a poder fer una anàlisi per mitjà de qualsevol programa caldrà que en la primera fila hi hagi la capçalera de cada columna, és a dir, el nom del camp.

- Comprovar que els registres d'una mateixa columna no estan escrits de maneres diferents. Això ens podria portar a fer recomptes erronis. Aquesta comprovació es pot fer amb una taula dinàmica, que veurem més endavant com es fa, o utilitzant programes específics de neteja de dades com ara el Google Refine (*).

- Comprovar que no hi hagi espais extres dins de les cel·les que formen la taula. Per a eliminar els espais complementaris en un gestor de fulls de càlcul, tant si és el Microsoft Excel o el Google Spreadsheets, caldrà fer-ho amb la fórmula =ESPACIOS, o =TRIM, si el programa està en anglès.

- Que no hi hagi diferències de minúscules i de majúscules. Precisament per aquesta raó, és una bona idea canviar a majúscules tots aquells camps que pretenguin ser cercables per la formula d'Excel =MAYUSC, o =UPPER, si el programa està en anglès.

- Comprovar que tots els números estiguin en el mateix format. Per a fer-ho, caldrà seleccionar la columna, i en la pestanya «format», fer clic a l'opció de «cel·les». S'obrirà un menú emergent i, en la primera pestanya, la de «número», escollirem el format idoni per a totes les columnes que contenen xifres.

- Abans de fer qualsevol modificació de l'estructura, cal afegir un nombre únic per a cada registre. D'aquesta manera, tindrem un valor únic per a cada fila i no ens confondrem en el cas que *a posteriori* es faci algun duplicat sense voler.

4. Anàlisi de dades

4.2. La importància d'ordenar bé un full de càlcul

4.2.3. Un exemple pràctic

Per a aquest exemple, podeu treballar amb l'arxiu: Adjudicacions Brut

Per a entendre millor com s'ha d'estructurar i de netejar un conjunt de dades, val més que seguim un exemple pràctic. Imaginem que volem dur a terme una anàlisi de la licitació de contractes de tots els departaments de la Generalitat de Catalunya durant la segona quinzena del mes de desembre de 2016. L'objectiu és esbrinar la següent informació:

- Quina és l'empresa que s'endú un import més gran per contractes de cada departament?
- Quina és la mitjana dels imports de les licitacions de subministrament per departament?
- El nombre de contractes que estan adjudicats a més d'una empresa.

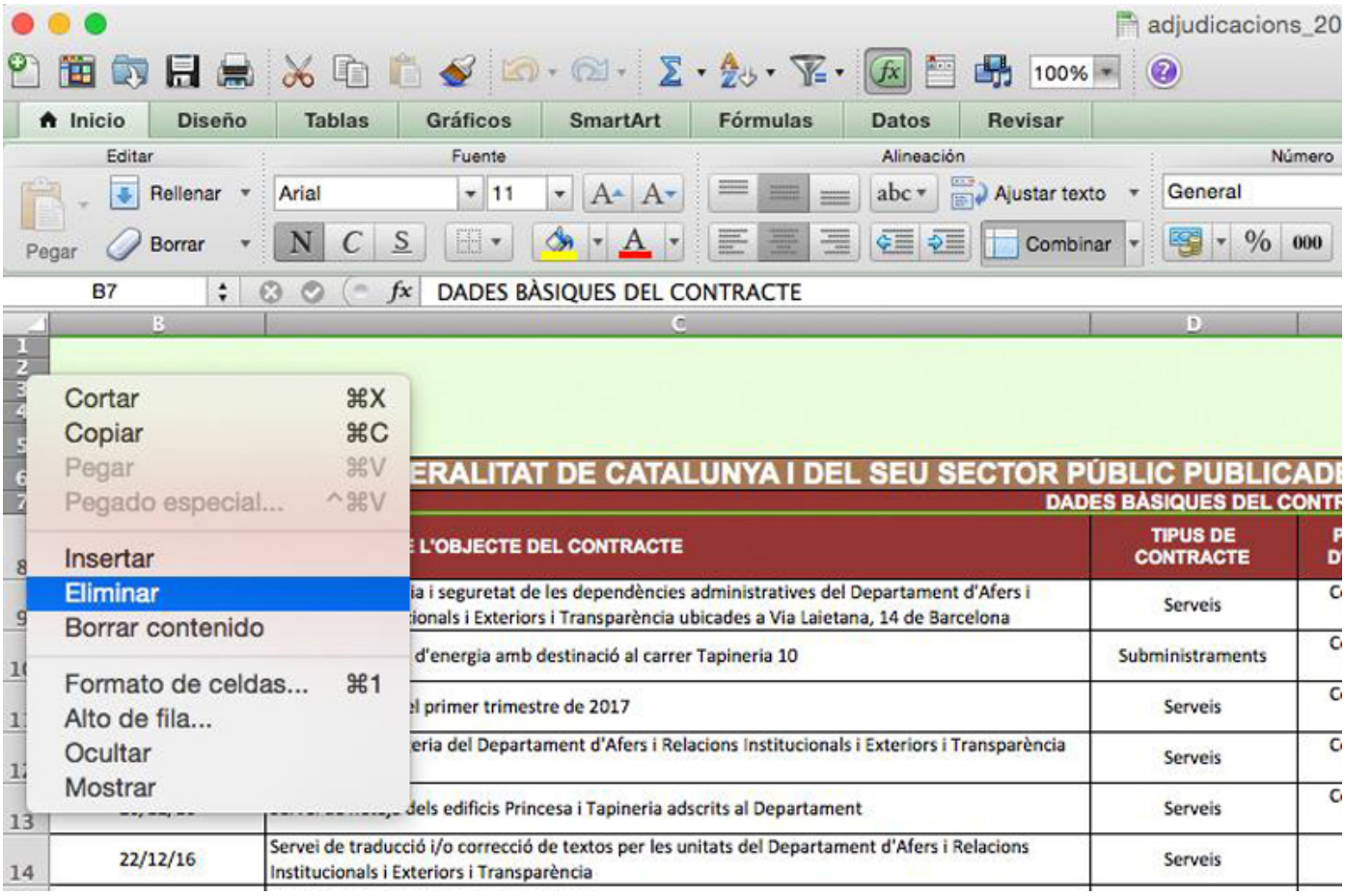
Sabem amb antelació que el registre de contractació pública conté tota la informació respecte a la contractació major de la Generalitat de Catalunya estructurada en fulls de càlculs de forma quinzenal. Es pot accedir a l'enllaç pel portal de transparència de l'administració catalana. Per dur a terme aquest exercici, descarregarem l'arxiu «.xls» corresponent a la primera i segona quinzena de desembre de 2016, el període de temps en el qual volem centrar el focus periodístic.

Tot just obrir els Excel podem veure la informació corresponent a tots els departaments de la Generalitat estructurada amb els següents camps:

- **Organització de la contractació:** departament o organisme que convoca el concurs públic.
- **Data d'adjudicació:** dia, mes i any de l'adjudicació. Com que el full de càlcul descarregat correspon a la segona quinzena de desembre de 2016, només constaran adjudicacions fetes durant aquest període de temps.
- **Descripció de l'objecte:** necessitat que pretén cobrir el contracte, és a dir, motiu pel qual es convoca.
- **Tipus de contracte:** tipologia segons l'objecte del contracte, que pot ser de serveis, de subministraments o d'obres.
- **Procediment de l'adjudicació:** procés que s'ha seguit per tal de licitar i adjudicar el contracte: obert, negociat, etc.
- **Valor estimat del contracte:** import en el qual està valorat el concurs en el moment de la licitació.
- **Import d'adjudicació**: import final de l'adjudicació del contracte. Unitats en euros.
- **Empresa adjudicatària:** nom de l'empresa o les empreses que han guanyat el concurs. Per a cada concurs, hi ha tantes files com adjudicataris finals que han guanyat el concurs.
- **Durada:** període de temps que durarà el contracte.

1) Eliminar files sobrants

Abans de fer qualsevol pas, primer caldrà eliminar la imatge incrustada i les files de més que encapçalen el full de càlcul. Per a fer-ho, seleccionarem les files amb el cursor, farem clic amb el botó dret del ratolí i desplegarem el menú per tal d'eliminar-les. Farem el mateix per a les files inferiors, que indiquen el total de dades i una especificació metodològica sobre les dades.



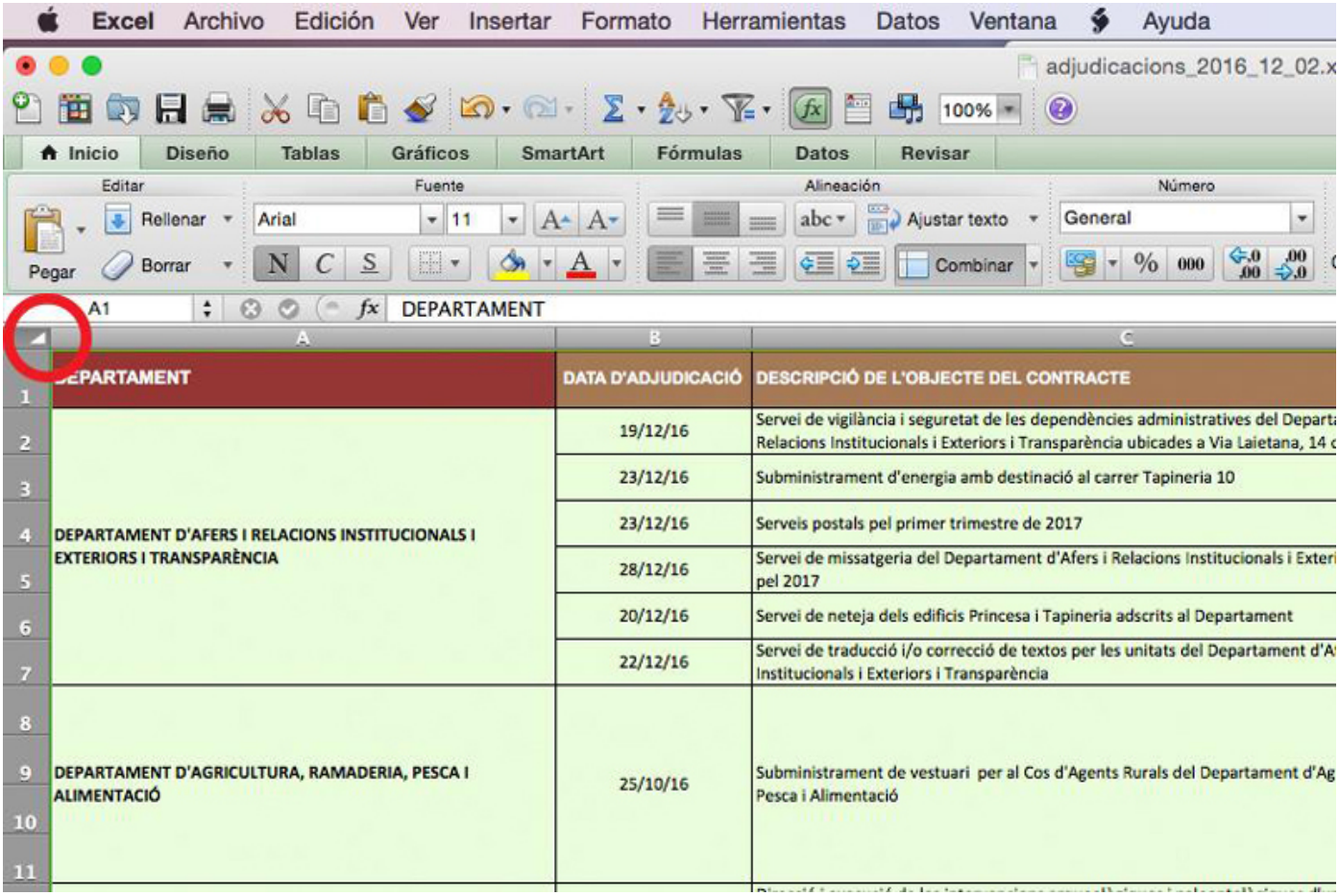
Eliminació de files sobrants

2) Descombinar les cel·les

Si fem un repàs molt ràpid al full de càlcul, veiem que pràcticament en totes les columnes hi ha cel·les combinades i, per tant, ens serà impossible aplicar els filtres per a ordenar de major a menor.

Per a fer-ho, primer haurem de descombinar les cel·les i decidir de quina manera estructurem la informació.

Per a descombinar-les seleccionarem tot el conjunt de dades d'Excel fent clic en el quadrant superior esquerre, tal com indica la següent imatge, o anant columna per columna.



Descombinació de cel·les (primer pas)

Un cop seleccionat tot el conjunt de la informació, anirem a la pestanya «format», i dins d'aquesta, a «cel·les». En el menú emergent que apareixerà, hem de fer clic en la pestanya «alineació» i desmarcar la casella inferior de «combinar cel·les» tal com es mostra a continuació.

Descombinació de cel·les (segon pas)

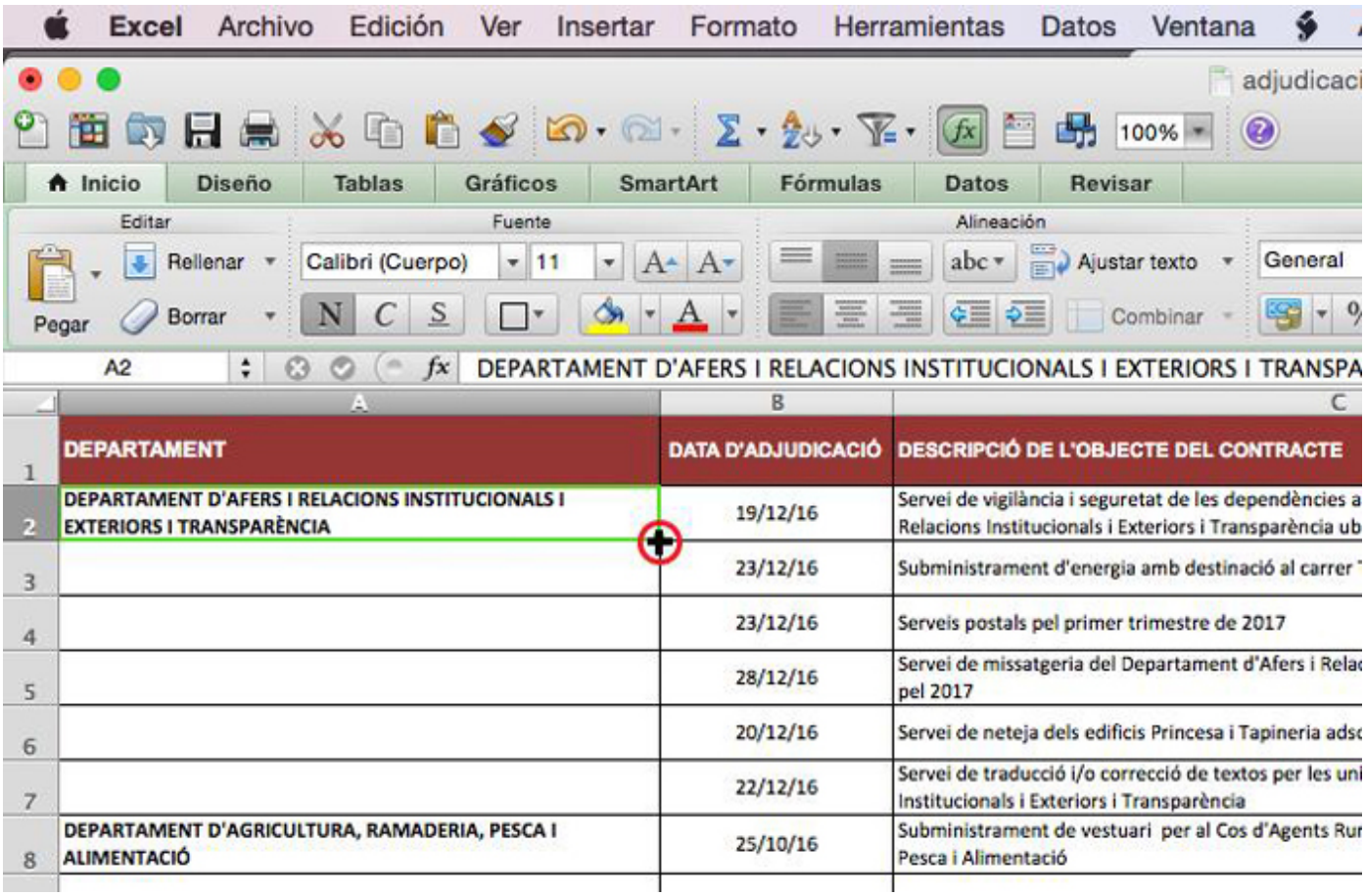
Un cop efectuada aquesta operació, totes les cel·les quedaran automàticament dividides de la següent manera:

Cel·les descombinades

El problema és que després d'aquest canvi, les cel·les descombinades han quedat buides de contingut. En dividir-les ha quedat la informació que pertoca només a la cel·la superior. Per tant, el que haurem de fer a continuació és copiar el contingut a la resta de cel·les. Tot i que hi ha macros, una opció avançada del programa, pràctiques per a Microsoft Excel, que permeten automatitzar aquest procés, es tracta d'un procediment més complex que no és vàlid per a tots els fulls de càlcul. En canvi, el sistema manual sempre serà una alternativa viable.

3) Entendre l'estructura i omplir el contingut amb cura

Així, doncs, per a copiar el contingut de forma manual ens situarem a la part inferior esquerra de la casella que conté la informació que volem copiar. Quan aparegui una creu negra, arrosseguem la creu fins l'última casella en la qual volem que es copiï la dada, i automàticament la informació quedarà copiada. Si en comptes d'arrossegar fem doble clic, aleshores el contingut es copiarà a totes les cel·les buides consecutives que hi hagi a continuació de la mateixa columna.



Còpia de contingut de forma manual

Per a navegar de manera més directa i senzilla per la taula sense haver d'utilitzar la barra de la dreta, es pot fer servir la tecla «Control» i les fletxes. En cas de Macintosh, la combinació de tecles serà «Comando» i, de nou, les fletxes.

Abans de començar a fer la feina de copiar i enganxar, és important donar un cop d'ull a tot el conjunt de dades per a entendre com està estructurat. Cal omplir la informació columna per columna sense caure en errors.

A simple vista localitzem que la primera columna, la de «departament», serveix al mateix temps com a «encapçalament» i com a «registre». És a dir, que hi ha licitacions que les fa el departament, d'altres que les fan altres organismes que conformen el departament, i d'altres casos en què el departament no fa cap licitació i simplement l'anoten com a encapçalament. És el cas, per exemple, de la fila 60.

Per a dur a terme el nostre propòsit periodístic ens interessa que hi hagi tant la columna del departament, com la de l'organisme concret que fa la licitació. A més a més, per a poder analitzar-ho bé no pot haver-hi registres buits que només serveixin com a encapçalaments.

Per tant, crearem una columna nova que es dirà «Departament» i reanomenarem la que teníem en segon lloc com a «Organisme Licitador». Tot seguit, copiarem en la primera columna el nom del departament que fa la licitació, i en la segona, l'organisme concret que fa la licitació: si és el departament mateix, copiarem el nom del departament i, si és un organisme més petit que pertany al departament, aleshores copiarem el nom de l'organisme.

Els departaments de Governació, Administracions Públiques i Habitatge, el de Justícia, i el de Territori i Sostenibilitat no han fet cap licitació i, per tant, només serveixen com a encapçalaments. En aquests tres casos eliminarem la fila buida i copiarem el nom

del departament a les files inferiors, a les quals hi ha les licitacions dels òrgans que sí que han licitat i depenen d'aquests departaments.

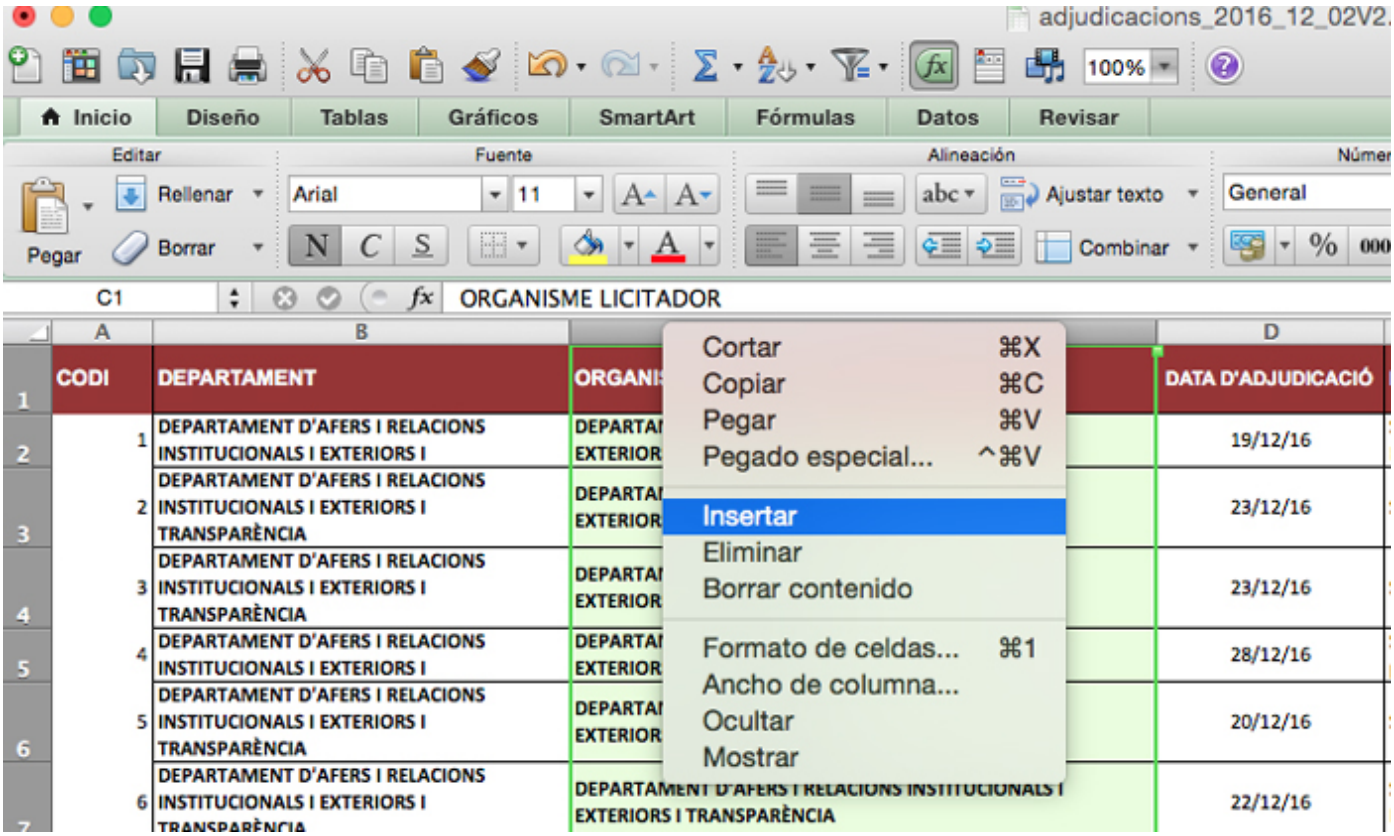
Abans de començar l'anàlisi, cal afegir una columna que identifiqui cada contracte amb un registre únic, per exemple, un numèric. Necessitem aquesta columna perquè cada fila no és un contracte, sinó una adjudicació. De manera que si un contracte té més d'un adjudicatari, el contingut estarà acumulat en files diverses segons el nombre d'empreses adjudicatàries. Per a solucionar aquest aspecte, crearem una columna nova amb un codi únic per a cada contracte. El valor adjudicat a cada contracte, doncs, es repetirà en els casos en què un contracte hagi estat adjudicat a més d'una empresa.

Una vegada hem creat aquesta columna, acabarem de copiar la informació per a la resta de camps. Igual que hem fet amb les dues primeres columnes, arrossegarem la informació de la primera cel·la a la resta de caselles duplicades. Compte a arrossegar i no copiar, ja que hi ha valors com les dates o els números que poden augmentar en lloc de copiar-se exactament. En aquests casos haurem de copiar la informació i no arrossegar-la.

4) Homogeneïtzar els camps

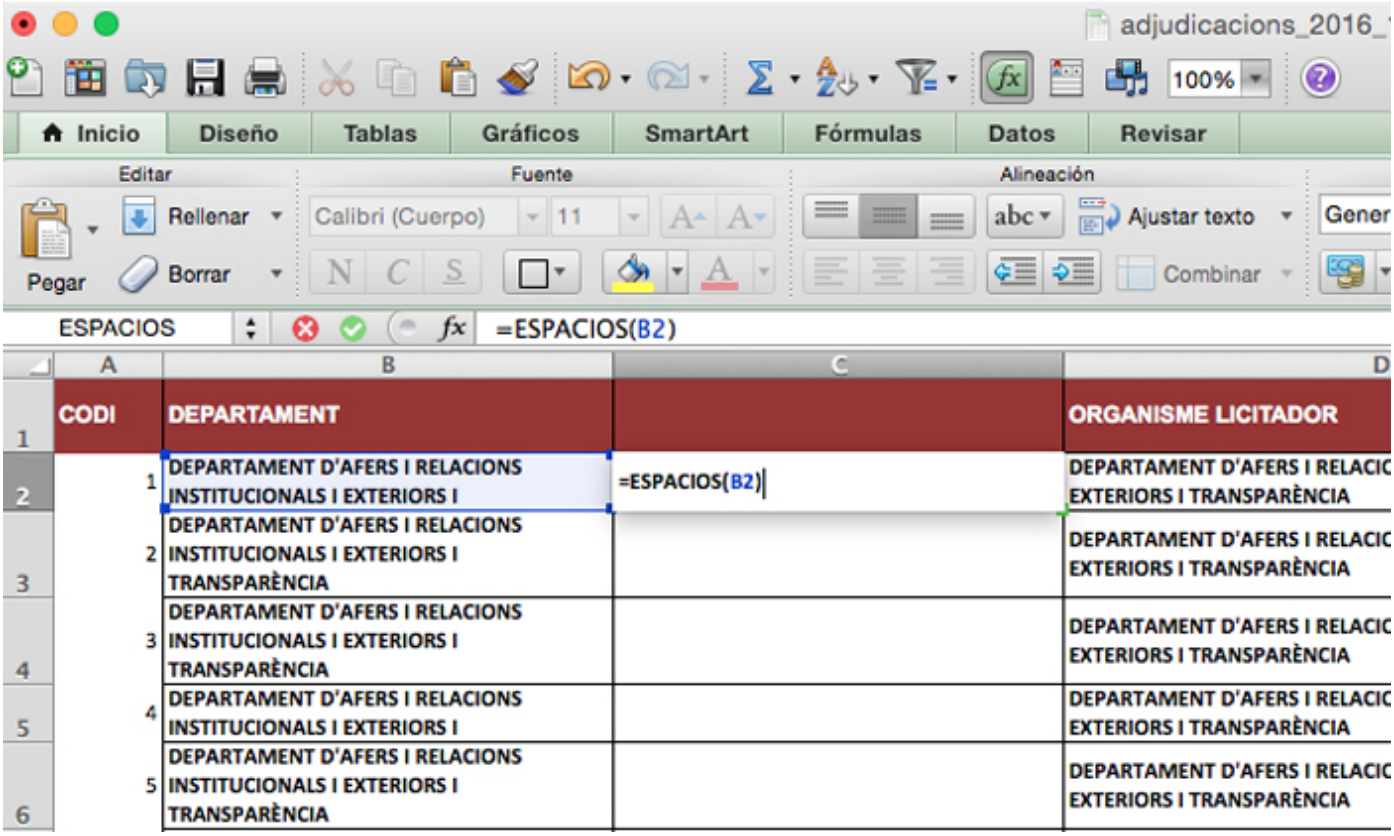
Una vegada modificada l'estructura podrem procedir a fer la neteja habitual amb les fórmules que hem vist en l'apartat anterior, i que serveixen per a homogeneïtzar tots els camps. Així, doncs, aplicarem les fórmules corresponents per a treure els espais excedents i per a col·locar totes les lletres en majúscula.

Per a fer-ho, crearem una columna nova al costat de la que volem netejar fent un clic amb el botó de la dreta sobre la columna.



Homogeneïtzar els camps (primer pas)

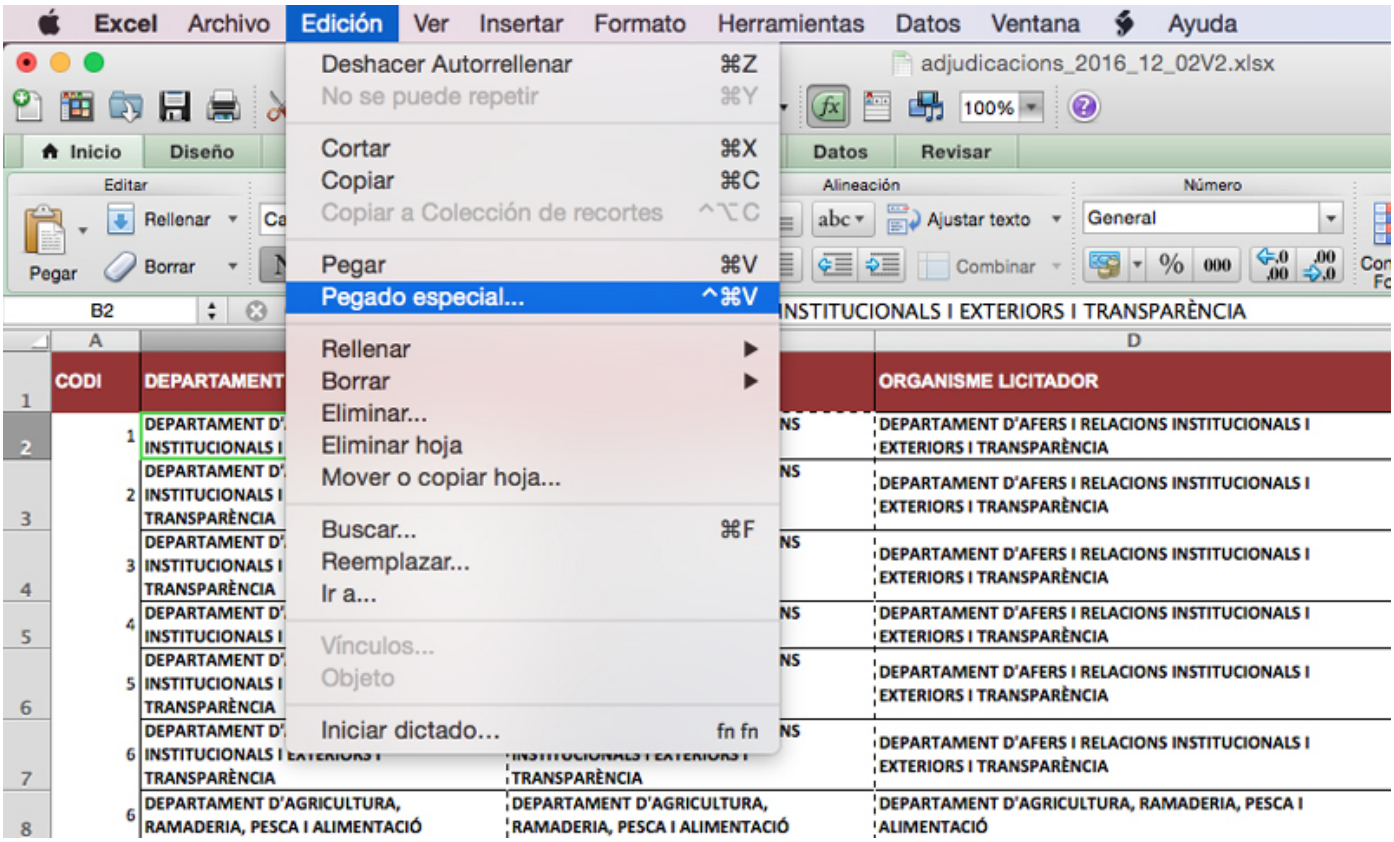
Un cop creada la columna, aplicarem la fórmula corresponent, tant si és =ESPACIOS, o =MAYUSC, i l'arrossegarem al llarg de tota la columna.



Homogeneïtzar els camps (segon pas)

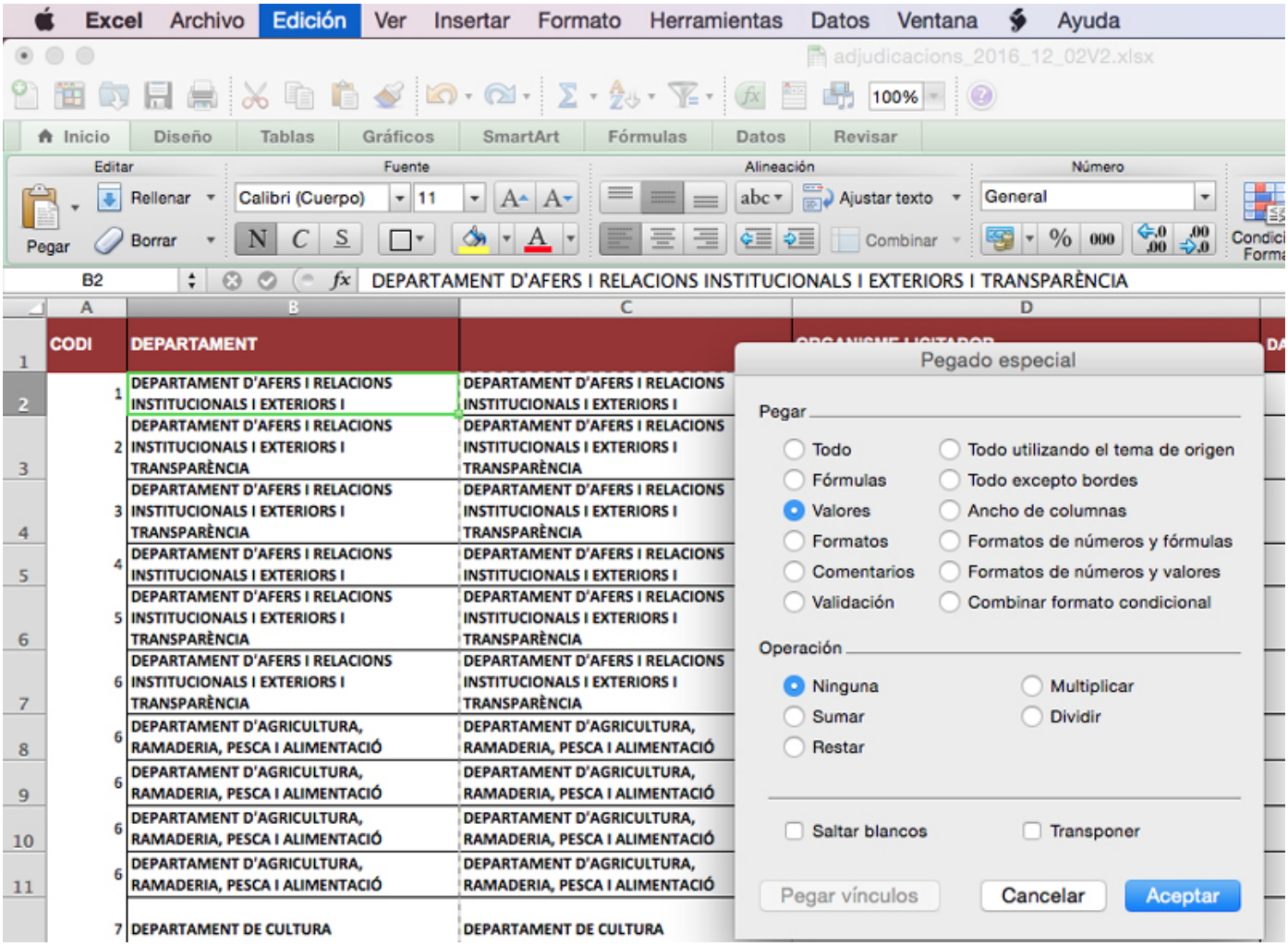
Per a poder treballar millor *a posteriori* amb el full de càlcul, és convenient eliminar les fórmules un cop aplicades i, per tant, treballar només amb els valors. Per a fer-ho, copiarem tota la columna sobre la qual estem aplicant la fórmula, en aquest cas la columna C.

Amb la informació copiada, ens situarem sobre la columna que conté la informació original, la B, i anirem a la pestanya «edició» per a fer un clic sobre l'opció «enganxament especial».



Homogeneïtzar els camps (tercer pas)

Per defecte, s'obrirà un menú que tindrà marcada l'opció d'enganxar-ho «tot». Si no volem copiar també les fórmules, haurem de canviar l'opció determinada per defecte per «valors», tal com es mostra en la següent captura de pantalla.



Homogeneïtzar els camps (quart pas)

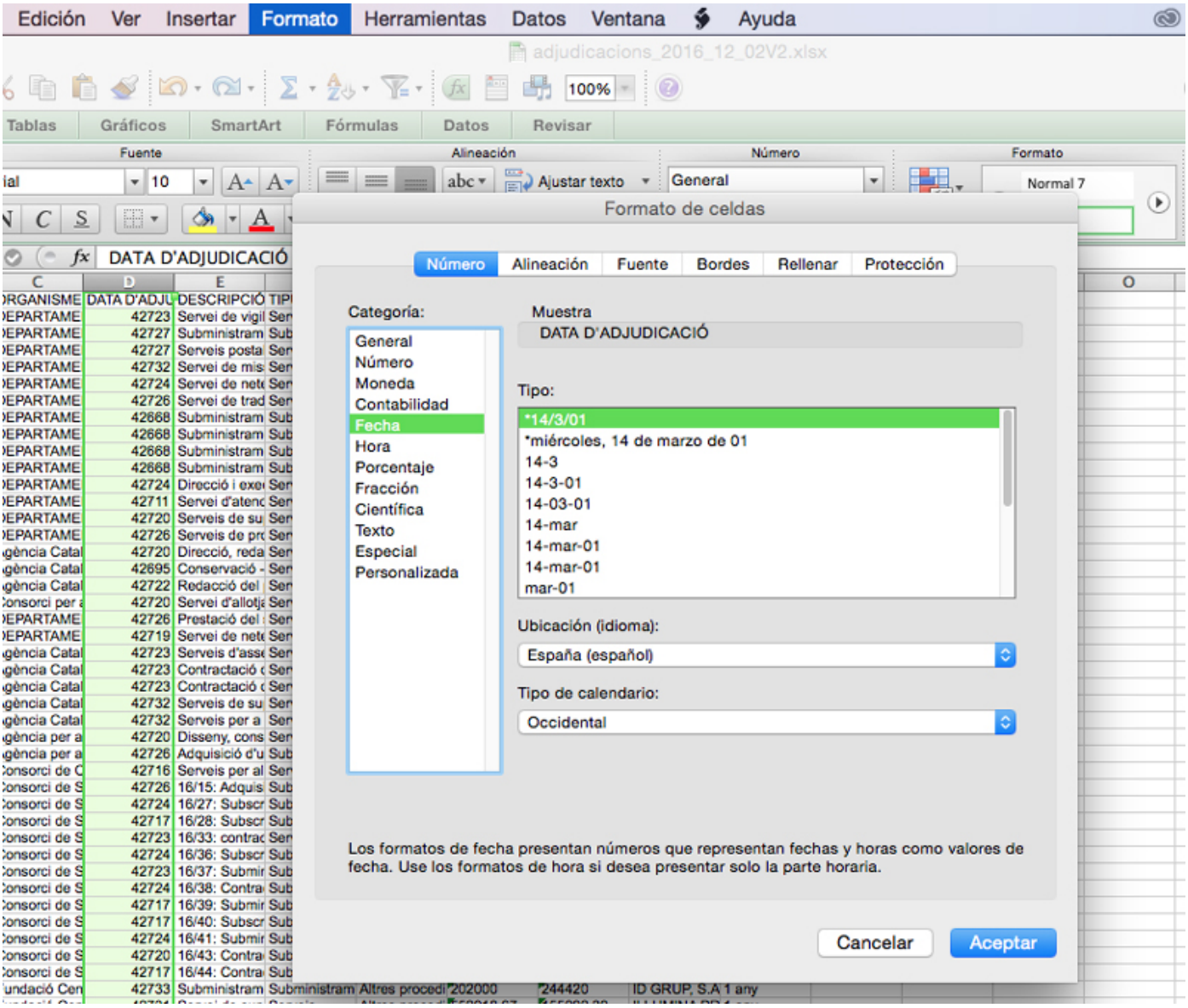
Una vegada copiada la informació de la forma correcta, podrem procedir a eliminar la columna de la fórmula. Aquest procés el repetirem amb totes les columnes, fins que tot el full de càlcul estigui a punt per a analitzar.

Una altra manera més ràpida d'aplicar aquesta fórmula, és aplicant les fórmules en un full nou, en lloc d'una columna nova. D'aquesta manera aplicarem la fórmula desitjada (=MAYUSC o =ESPACIOS) a totes les columnes alhora. Tot i que a la llarga és la forma més efectiva, quan estem començant, anar pas per pas va bé per tal d'evitar confusions.

5) Comprovar que la feina de neteja s'ha fet correctament

Després d'haver fet la neteja, i abans de passar a l'anàlisi, caldrà comprovar que no s'han comès errors mentre es picava durant la neteja. Alguns errors comuns són, per exemple, haver passat per alt copiar i enganxar alguna cel·la. Per a detectar de manera ràpida aquest error, el més senzill és navegar pel full de càlcul amb la tecla «CONTROL + fletxes», o «CMD + fletxes» en cas de treballar en Mac.

D'altra banda, també caldrà comprovar que el format de les columnes és correcte. Sobretot a l'hora de fer l'enganxament especial d'únicament els «valors», ja que en fer-ho perdem el format atribuït a cada columna. Per a modificar el format de les dades caldrà seleccionar la columna, i en la pestanya «format» fer clic en l'opció de «cel·les», seleccionant el format adient per a cadascuna de les columnes. Per exemple, en el cas de la columna D, això és traduirà en el format de data format per dia, mes i any.



Comprovació de formats

Ara sí, una vegada fetes les comprovacions i haver comprovat que no hi ha errors de neteja, podrem procedir a analitzar les dades.

Per a anar bé, el resultat final net hauria de quedar com l'arxiu adjunt següent: Adjudicacions_Net.

4. Anàlisi de dades

4.3. Eines i mecanismes d'anàlisi

4.3.1. Importar dades

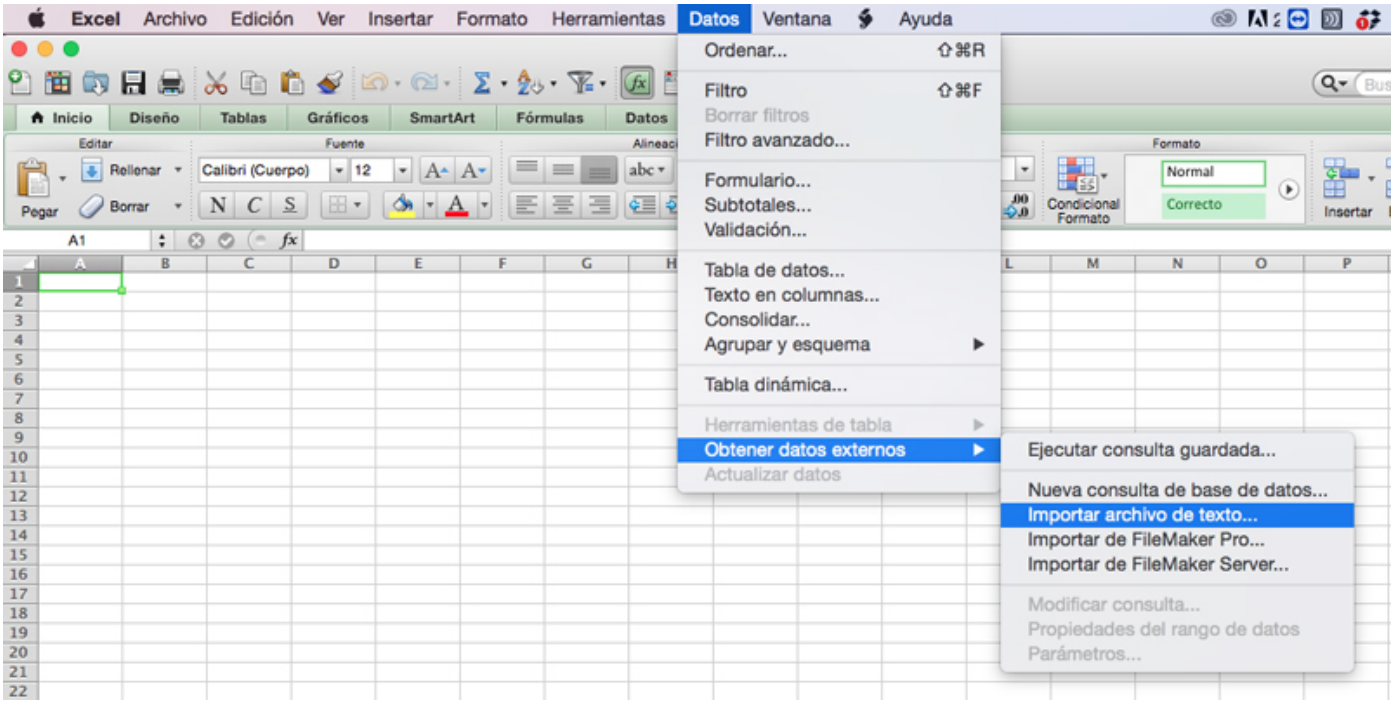
A l'hora de treballar amb dades exportades d'altres bases de dades més grans, ens podem trobar tres tipus de formats: XLS o XLSX, segons la versió, TSV, i CSV.

El CSV (*) és un arxiu que s'ordena en files i en columnes, en el qual la primera fila correspon a l'encapçalament de cada columna, i cada fila a un registre. Cada camp (columna) està separat per comes. Més enllà de l'arxiu d'Excel, és el tipus de format més fàcil de trobar.

El TSV (*) és un arxiu estructurat de la mateixa manera que un CSV, però el separador en aquest cas és el tabulador i no la coma.

En el cas dels arxius XLS o XLSX, els arxius de Microsoft Excel, només cal obrir-los amb qualsevol dels programes de gestió de fulls de càlcul que hem vist anteriorment i començar a treballar-hi. En canvi, en el cas del CSV o el TSV, haurem de seguir uns passos específics, similars per a tots dos formats, per a poder treballar amb l'arxiu.

Amb el Microsoft Excel obert, caldrà anar a la pestanya «dades», i dins d'aquesta, a «obtenir dades externes» i a «importar arxiu de text». Aleshores, s'obrirà un menú pel qual podrem seleccionar l'arxiu desitjat i començar el procés per a dur a terme la importació.

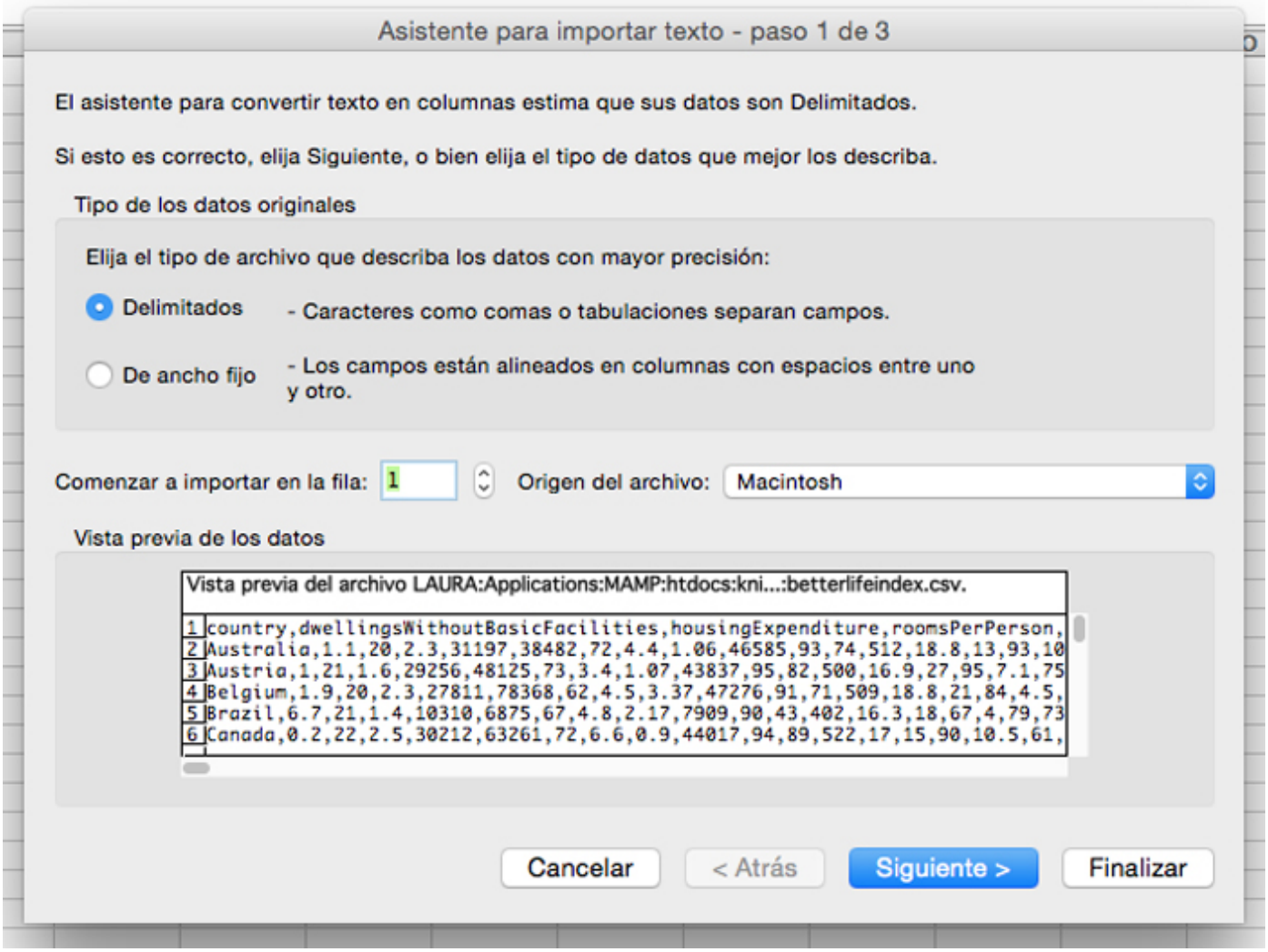


Importar dades (primer pas)

El primer que ens demana l'assistent d'importació d'arxiu de text és saber de quina manera estan delimitades les files: si per un valor separador, una coma o un tabulador, per exemple, o per una mida determinada. Marcarem la primera opció: «caràcters com comes i tabulacions separen els camps».

Abans de passar a la pantalla següent haurem de marcar la fila a partir de la qual es vol iniciar la importació. Generalment, en aquesta part ens interessa mantenir el valor per defecte: que comenci la importació des de la fila 1.

Per últim, caldrà marcar el tipus de codificació amb què està guardat l'arxiu amb el qual volem treballar en el desplegable que hi ha en la part inferior dreta del menú emergent. Normalment utilitzarem les codificacions universals: UTF-8 o UTF-16. Tot i així, de vegades ens podem trobar que segons com hagi estat desat l'arxiu, hi hagi alguns caràcters especials en el cas del català, que no acaba de llegir correctament. És el cas de la «ç» o dels accents oberts. Per tant, abans de continuar amb aquest tipus de codificació, caldrà donar un cop d'ull al previsualitzador per a assegurar-se que llegeix correctament l'arxiu. Si les codificacions UTF no descodifiquen correctament, es podrà provar amb la Windows 1250, Windows 1251, Windows 1252 o la Macintosh. És important, però, que abans de passar a la següent pantalla, els caràcters es previsualitzin correctament.



Importar dades (segon pas)

Una vegada fets aquests passos, passarem a la següent pantalla. En aquest cas, l'objectiu és senzill: seleccionar el valor que separa les columnes. En el cas del CSV és una coma, i en el cas del TSV, un tabulador.

Un cop fet aquest pas, podrem continuar endavant. La resta de pantalles no requereixen cap intervenció, per tant farem clic al botó «següent» fins que les dades ja estiguin definitivament importades i visualitzem a la pantalla de l'ordinador un full de càlcul com qualsevol altre, separat en les columnes corresponents.

Evidentment hi ha milers d'altres formats, alguns molt comuns també en la pràctica del periodisme de dades i les visualitzacions interactives, però són una tipologia d'arxius que veurem en altres casos aplicats.

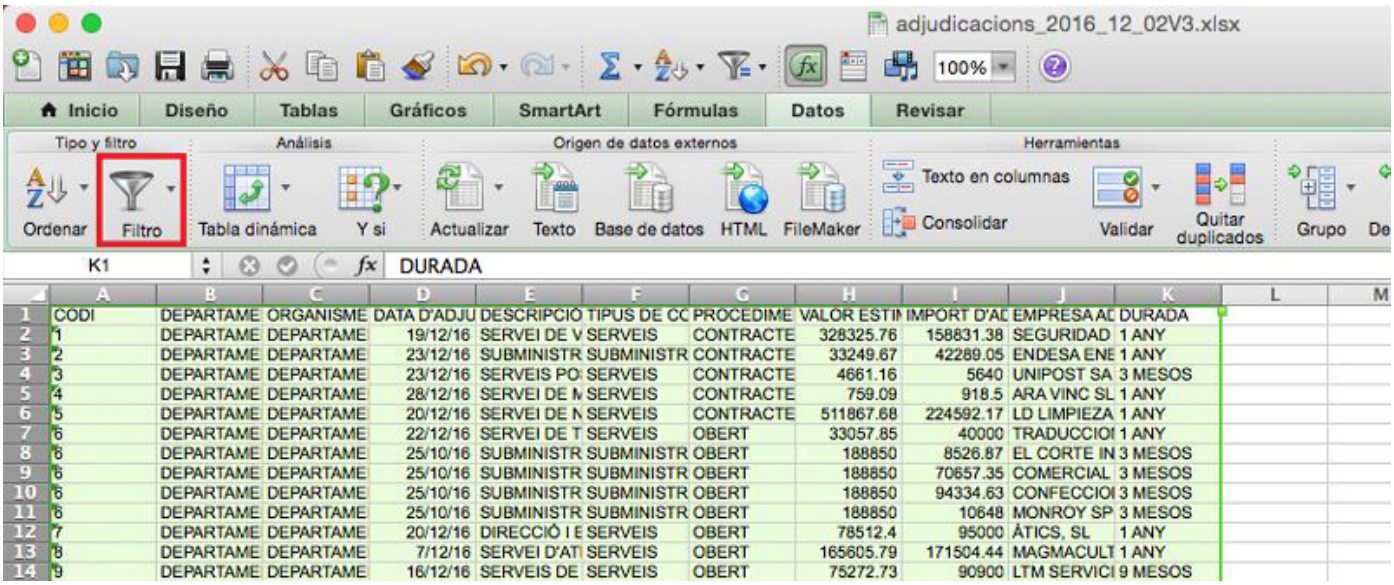
4. Anàlisi de dades

4.3. Eines i mecanismes d'anàlisi

4.3.2. Filtrar i ordenar

El primer pas a l'hora de començar a treballar amb les dades ja netes serà posar les opcions per a ordenar de menor a major. És molt important que prèviament les dades hagin estat tractades i col·locades en la disposició correcta. Procedir al filtratge de dades sense que hagin estat tractades prèviament pot portar a errors que ens faran extreure conclusions equivocades.

Per tant, i seguint amb l'exemple de l'adjudicació dels concursos públics: seleccionarem les columnes que contenen dades, farem clic en el menú de «dades», i marcarem l'opció «filtre».

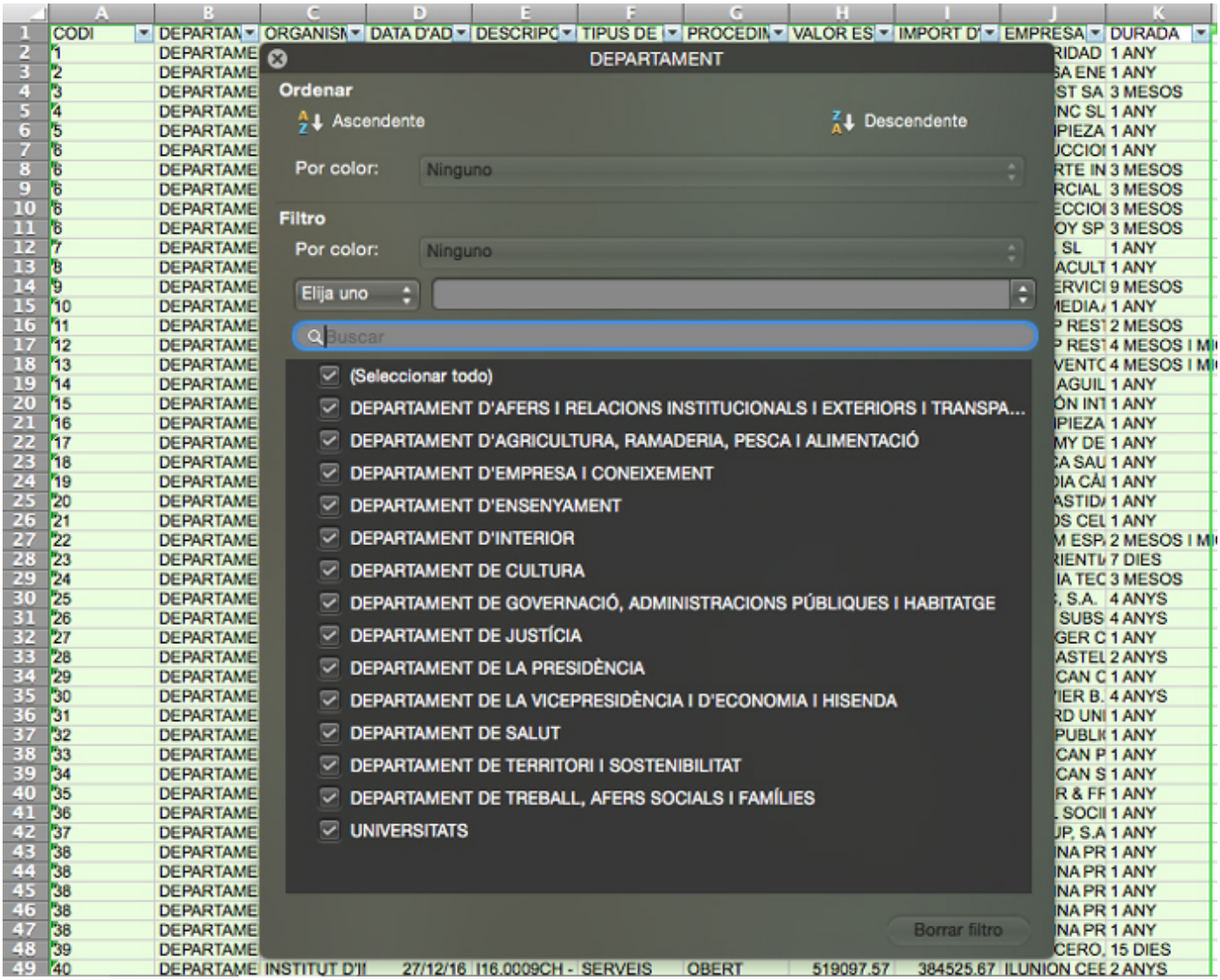


	A	B	C	D	E	F	G	H	I	J	K
1	CODI	DEPARTAME	ORGANISME	DATA D'ADJU	DESCRIPCIÓ	TIPUS DE CC	PROCEDIME	VALOR ESTIN	IMPORT D'AD	EMPRESA AD	DURADA
2	1	DEPARTAME	DEPARTAME	19/12/16	SERVEI DE V	SERVEIS	CONTRACTE	328325.76	158831.38	SEGURIDAD	1 ANY
3	2	DEPARTAME	DEPARTAME	23/12/16	SUBMINISTR	SUBMINISTR	CONTRACTE	33249.67	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 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'ATI	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

Aplicació de filtres

D'aquesta manera, apareixerà en la primera fila de cada columna una fletxa que en fer-hi clic ens permetrà les següents opcions:

- Ordenar de més a menys, i de menys a més, tant en ordre alfabètic com numèric.
- Ordenar per color, en el cas que hi hagi caselles d'un color determinat.
- Aplicar filtres de coincidència. És a dir, escollir visualitzar només les opcions determinades que ens puguin interessar, tant si és un valor concret, si són valors que siguin superiors a una xifra determinada, o que comencin o acabin per un caràcter concret.



Aplicació de filtres de coincidència

Però més enllà de veure màxims, mínims i filtrar per valors concrets, les opcions d'anàlisi per mitjà dels filtres són certament limitades.

4. Anàlisi de dades

4.3. Eines i mecanismes d'anàlisi

4.3.3. Les taules dinàmiques

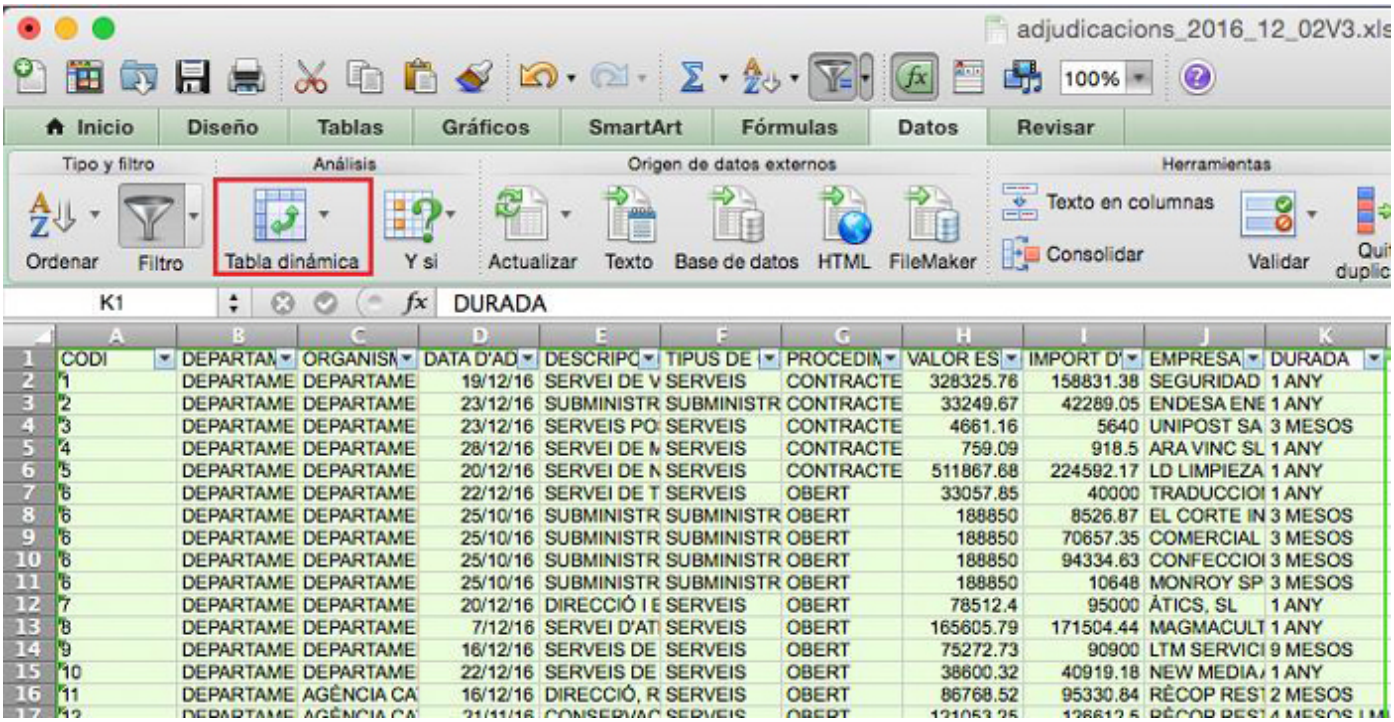
En aquest cas treballarem amb l'arxiu: Adjudicacions_TaulaDinamica.

4.3.3.1. Les taules

L'eina per excel·lència a l'hora de dur a terme una anàlisi de dades amb un gestor de fulls de càlcul és la taula dinàmica. La taula dinàmica és una de les funcions bàsiques d'Excel que permet reordenar les dades com més interessi per a poder analitzar-les des d'angles diferents.

Ara seguirem pas per pas la creació d'una taula dinàmica per a entendre'n més el funcionament, i per a fer-ho, recuperem de nou l'exemple de les adjudicacions públiques.

Amb l'arxiu obert, haurem de seleccionar de nou les columnes del *data set*, i en el menú de «dades» fer clic a l'opció «taula dinàmica» de la següent manera:

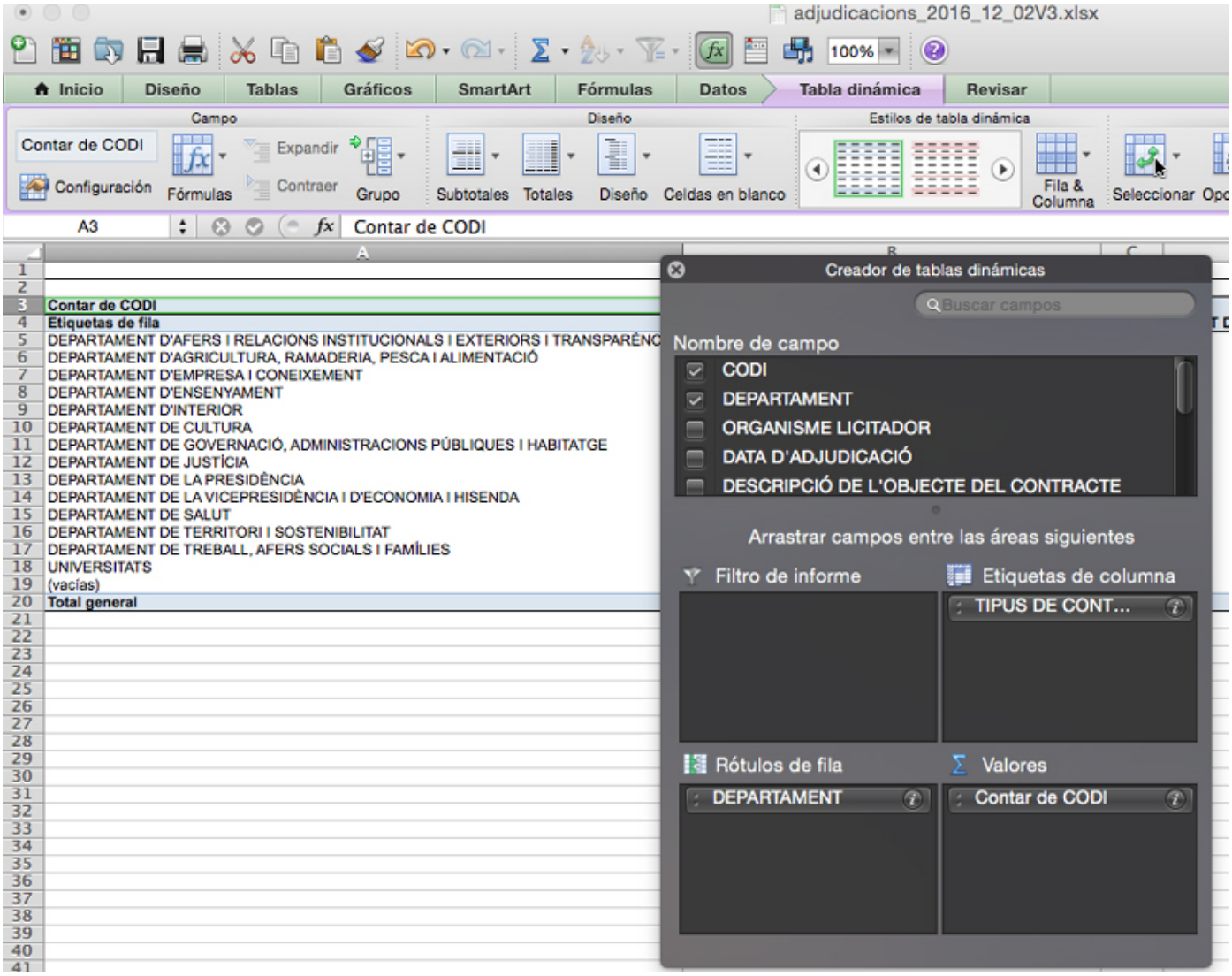


	A	B	C	D	E	F	G	H	I	J	K
1	CODI	DEPARTAM	ORGANISM	DATA D'AD	DESCRIPC	TIPUS DE	PROCEDIM	VALOR ES	IMPORT D'	EMPRESA	DURADA
2	1	DEPARTAME	DEPARTAME	19/12/16	SERVEI DE V	SERVEIS	CONTRACTE	328325.76	158831.38	SEGURIDAD	1 ANY
3	2	DEPARTAME	DEPARTAME	23/12/16	SUBMINISTR	SUBMINISTR	CONTRACTE	33249.67	42289.05	ENDESA 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 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'ATI	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	2 MESOS

Creació d'una taula dinàmica

Un cop fem clic a la creació d'una nova taula dinàmica, es crearà automàticament una nova pàgina en el full de càlcul. Pot ser que per defecte, la taula estigui mig plena o completament buida, però abans de procedir a l'anàlisi, caldrà entendre bé per a què serveix cada opció.

D'entrada veurem com a la part superior dreta del full de càlcul hi ha la taula dinàmica, i a l'esquerra hi ha el «creador de taules dinàmiques», que és l'eina que utilitzarem per a afegir o treure camps de la taula dinàmica.



Visualització de la taula dinàmica a la dreta i el creador de taules dinàmiques a l'esquerra

Vegem, opció per opció, per a què serveix cadascun dels espais.

- **Nom del camp**: en aquest apartat, si hem fet bé els passos, hauríem de trobar el nom de totes les columnes de la nostra taula original. En el cas del full de càlcul d'adjudicacions, la taula s'ha ordenat correctament i en aquest apartat podem veure les columnes del conjunt de dades (codi, departament, organisme licitador, etc.)

- **Filtre d'informe**: aquí arrossegarem els camps sobre els quals ens interessi aplicar filtres sobre les dades.

- **Etiquetes de columna**: en aquest apartat arrossegarem els camps que vulguem situar com a columnes de la nova taula que estem creant a partir de la inicial.

- **Rètols de fila**: és el mateix que en el cas anterior, però per a les files. Aquí hi arrossegarem els camps que volem que es visualitzin en les files.

- **Valors**: un cop tinguem la taula estructurada en files i en columnes, hem de decidir quins valors volem mostrar. Per a fer-ho, haurem de llançar els valors en aquest apartat.

Per a introduir o treure els camps en cadascuna de les opcions, caldrà arrossegar les columnes de l'apartat «Nom de camps» a l'apartat que correspongui o interessi. En el cas que la taula ja vingui mig plena per defecte o que vulguem modificar una taula ja realitzada, per a treure els camps llançats haurem d'arrossegar-los de cadascuna de les opcions fora del generador de taules dinàmiques.

A mesura que arrosseguem per a afegir o treure camps, veurem com la taula del full d'Excel es modifica segons el que estiguem fent.

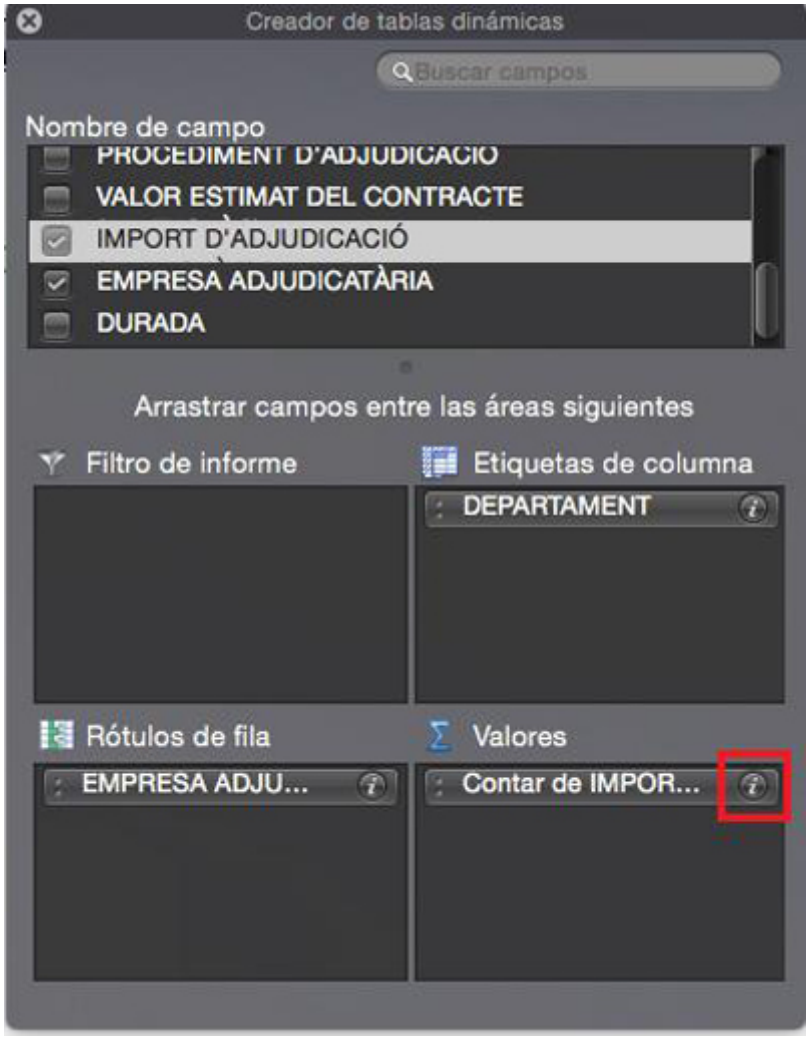
4.3.3.2. Un exemple pràctic

Provem, doncs, de respondre les preguntes inicials seguint l'exemple que hem estat prenent fins ara, el de l'Excel de les adjudicacions de contractes públics.

a) Quina és l'empresa que s'endú més contractes en cada departament?

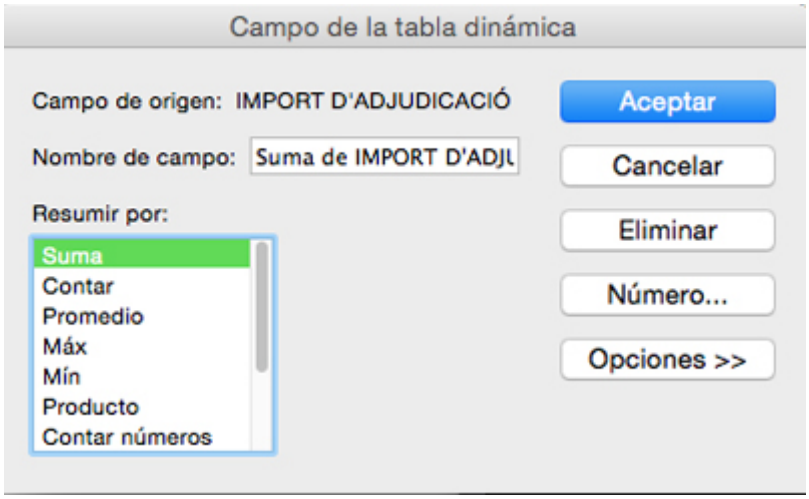
Per a respondre aquesta pregunta haurem de fer una taula dinàmica en la qual, en les files (o rètols de fila) hi hagi el nom d'«empresa adjudicatària», en les columnes (o camps de fila) hi hagi el «departament», i en el camp de valors hauríem de llançar l'«import d'adjudicació».

A l'hora de llançar els valors, és fàcil que per defecte compti els registres en comptes de sumar-los. Per a modificar aquesta opció caldrà fer clic al botó de la «i» que hi ha dins del camp que hem llançat.



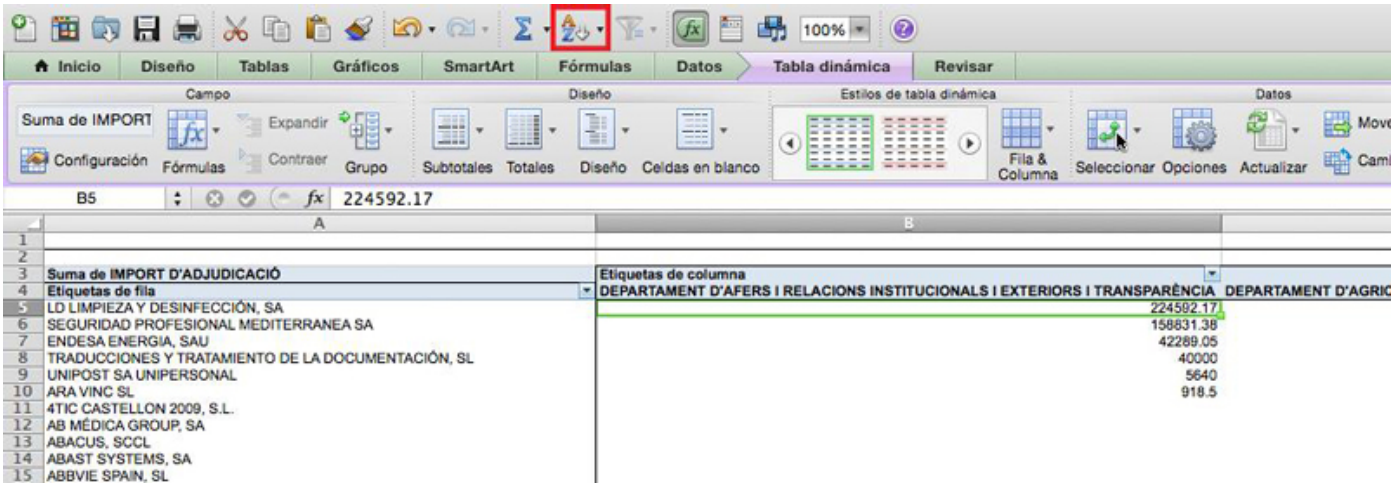
Creació d'una taula dinàmica

Aleshores, sortirà un menú emergent que ens permetrà canviar l'opció «contar» per moltes altres, també la de «sumar», que és la que ens interessa aquesta vegada.



Menú emergent per crear taules dinàmiques

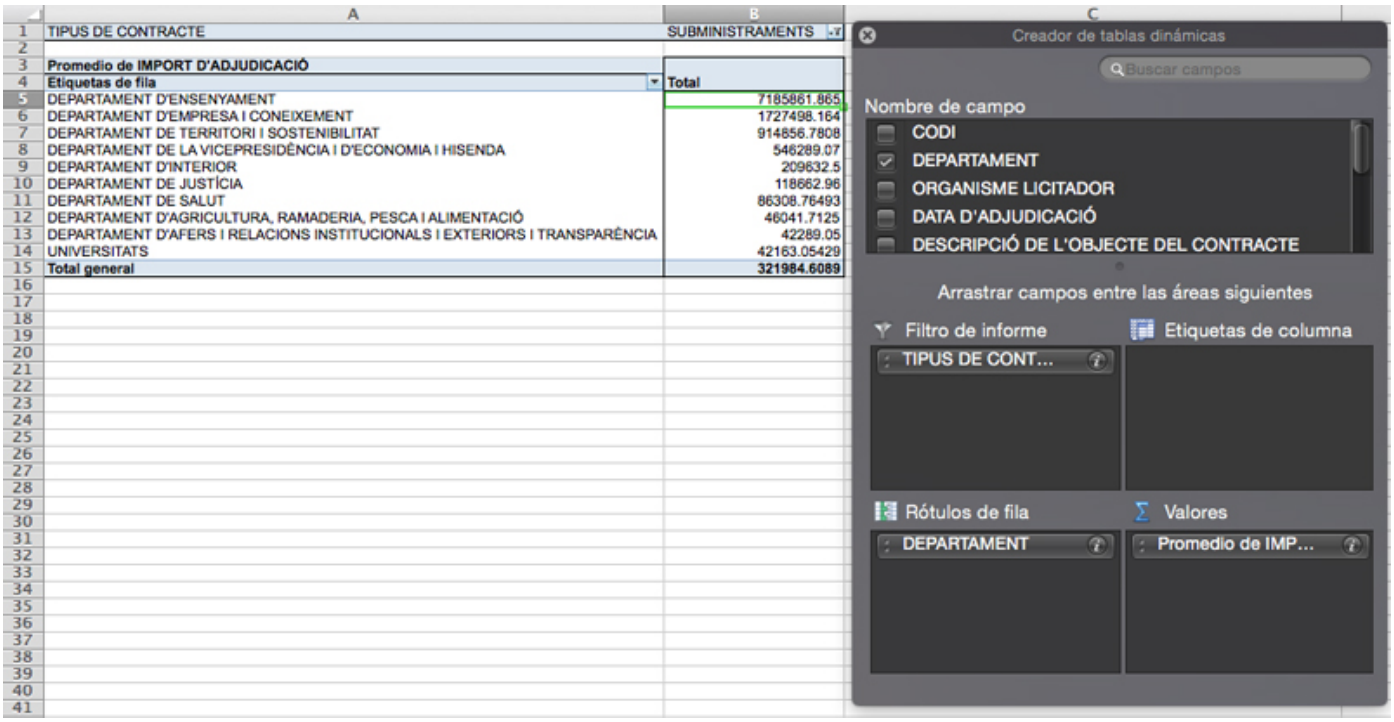
D'aquesta manera ens quedarà una taula en la qual s'agregarà l'import d'adjudicació per a cada empresa, en la qual les columnes siguin el departament i en el camp de valors hauríem de llançar el valor. Per a veure els màxims i els mínims de cada departament, haurem de situar-nos a la primera fila de cada columna i fer clic al botó que ens permet ordenar de més a menys, o de menys a més, segons el nombre de vegades que fem clic. Ho haurem de fer amb cadascuna de les columnes per a saber l'empresa que més adjudicacions s'ha endut en cada departament.



Taula dinàmica resultant

b) Quina és la mitjana dels imports de les licitacions de subministrament per departament?

Per a resoldre aquesta pregunta, haurem de fer una taula dinàmica en la qual les files corresponguin als departaments, i els valors corresponguin a l'«import de l'adjudicació». En l'apartat de filtres triarem l'opció «Tipus de Contracte» per a escollir únicament aquells que pertanyin a contractes de subministrament. Finalment, per tal que ens surti directament la mitjana, i no el recompte o la suma, farem clic en la «i» que apareix a la dreta de valors, i marcarem l'opció «mitjana» en el menú emergent. Finalment, la taula i el generador quedaran de la següent manera:



Creació d'una nova taula dinàmica

Ara, suposem que ens crida l'atenció el resultat del Departament d'Ensenyament, el més elevat de tots pel que fa a la mitjana. Si fem doble clic a la columna dels valors, la que està assenyalada en color verd a l'anterior captura de pantalla, ens apareixeran en un nou full tots els valors que estan vinculats a aquest valor per tal que puguem navegar-hi i explorar-los més detalladament si ens interessa.

4. Anàlisi de dades

4.3. Eines i mecanismes d'anàlisi

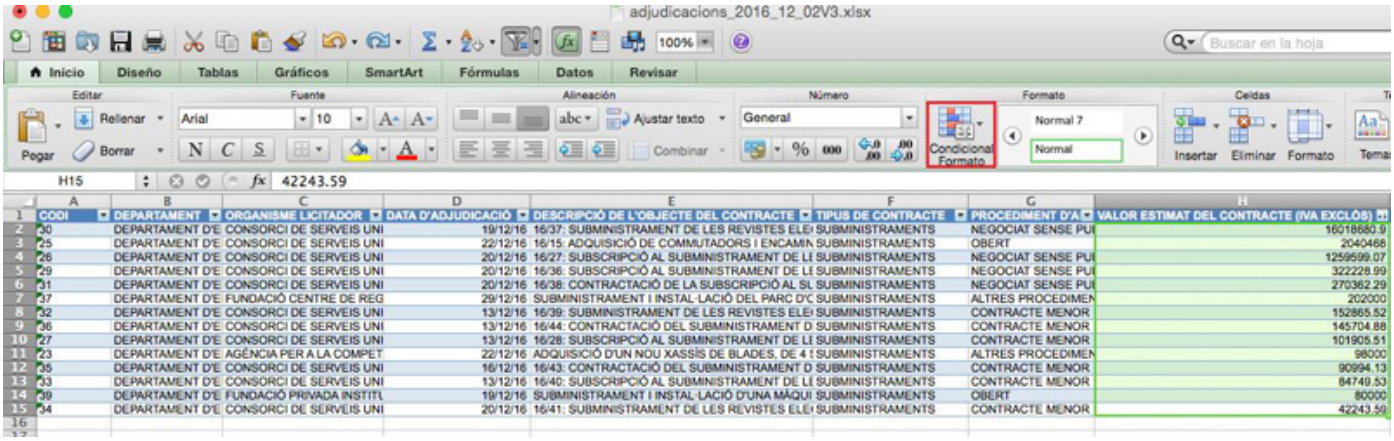
4.3.4. Filtres condicionals i altres opcions avançades

Però més enllà de les taules dinàmiques, també hi ha altres factors d'anàlisi que es poden aplicar a qualsevol taula, sigui dinàmica, o no.

El format condicional permet aplicar colors a les cel·les segons condicions específiques, com ara valors duplicats, rangs de major a menor, i ressaltar valors concrets. L'avantatge d'aquesta prestació és que si els valors canvien, les regles marcades també es reajusten.

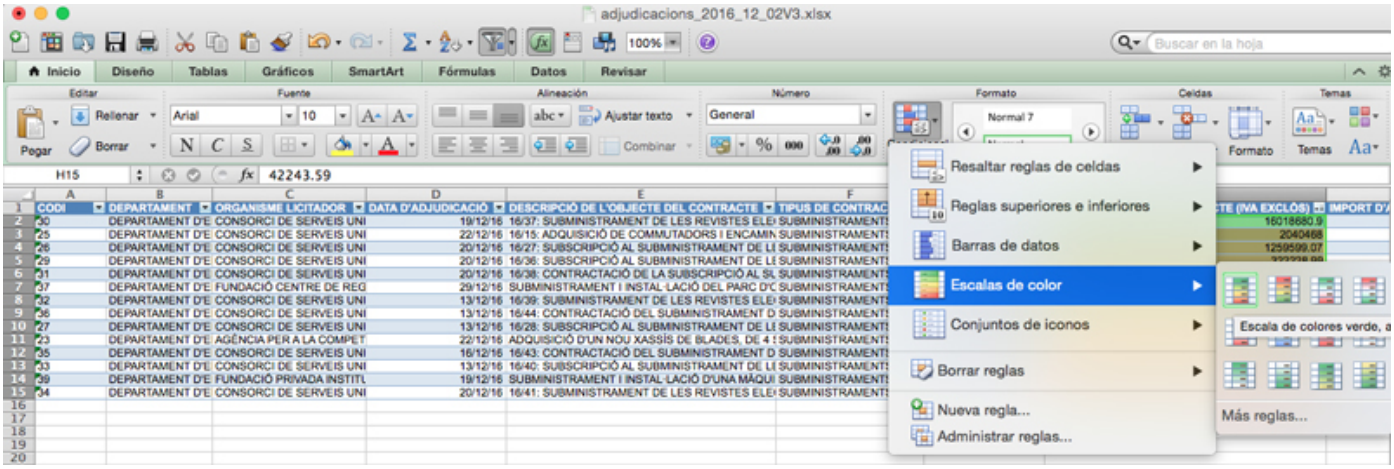
Continuant amb l'exemple que hem treballat fins ara, veure'm com podem fer-ho. Imaginem-nos que ens interessa aprofundir en els contractes del Departament d'Empresa i Coneixement efectuats durant aquest període, i concretament ens interessa veure d'una manera més visual el rang de preus del «valor estimat del contracte».

Per a fer-ho, haurem de seleccionar la columna de valor que ens interessa i obrir el menú de «condicional formato», situat a la barra d'eines.



Visualització de l'aplicació de filtres condicionals (primer pas)

Amb el menú desplegat, podrem seleccionar l'opció que més ens interessi. En aquest cas, una senzilla escala de color ja ens permetrà veure quin és el rang de preus, quins estan més distanciats, i quins són més similars.



Visualització de l'aplicació de filtres condicionals (segon pas)

Una vegada seleccionada l'opció, la columna ens quedarà pintada automàticament. Si esborrem el contingut de la cel·la, el filtre de color continuarà apareixent. Si en algun moment volem esborrar les regles de colors, caldrà tornar al menú «condicional formato» i fer clic a l'opció «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 EXCLÒS)
2	30	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	19/12/16	16/37: SUBMINISTRAMENT DE LES REVISTES ELE	SUBMINISTRAMENTS	NEGOCIAT SENSE PUI	16018680.9
3	25	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	22/12/16	16/15: ADQUISICIÓ DE COMMUTADORS I ENCAMIN	SUBMINISTRAMENTS	OBERT	2040468
4	26	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	20/12/16	16/27: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	NEGOCIAT SENSE PUI	1259599.07
5	29	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	20/12/16	16/36: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	NEGOCIAT SENSE PUI	322228.99
6	31	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	20/12/16	16/38: CONTRACTACIÓ DE LA SUBSCRIPCIÓ AL SU	SUBMINISTRAMENTS	NEGOCIAT SENSE PUI	270362.29
7	37	DEPARTAMENT DE	FUNDACIÓ CENTRE DE REG	29/12/16	SUBMINISTRAMENT I INSTAL·LACIÓ DEL PARC D'C	SUBMINISTRAMENTS	ALTRES PROCEDIME	202000
8	22	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	13/12/16	16/39: SUBMINISTRAMENT DE LES REVISTES ELE	SUBMINISTRAMENTS	CONTRACTE MENOR	152865.52
9	36	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	13/12/16	16/44: CONTRACTACIÓ DEL SUBMINISTRAMENT D	SUBMINISTRAMENTS	CONTRACTE MENOR	145704.88
10	27	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	13/12/16	16/28: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	CONTRACTE MENOR	101905.51
11	23	DEPARTAMENT DE	AGÈNCIA PER A LA COMPET	22/12/16	ADQUISICIÓ D'UN NOU XASSÍS DE BLADES, DE 4	SUBMINISTRAMENTS	ALTRES PROCEDIME	98000
12	35	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	16/12/16	16/43: CONTRACTACIÓ DEL SUBMINISTRAMENT D	SUBMINISTRAMENTS	CONTRACTE MENOR	90994.13
13	33	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	13/12/16	16/40: SUBSCRIPCIÓ AL SUBMINISTRAMENT DE LI	SUBMINISTRAMENTS	CONTRACTE MENOR	84749.53
14	39	DEPARTAMENT DE	FUNDACIÓ PRIVADA INSTITU	19/12/16	SUBMINISTRAMENT I INSTAL·LACIÓ D'UNA MÀQUI	SUBMINISTRAMENTS	OBERT	60000
15	34	DEPARTAMENT DE	CONSORCI DE SERVEIS UNI	20/12/16	16/41: SUBMINISTRAMENT DE LES REVISTES ELE	SUBMINISTRAMENTS	CONTRACTE MENOR	42243.59

Visualització de l'aplicació de filtres condicionals (tercer pas)

Aquest només és un exemple de filtre de colors que ens convé i ens interessa aquesta vegada. De totes maneres, per a altres vegades ens pot interessar més crear regles específiques per a buscar o ressaltar un valor determinat. La resta de possibilitats estan totes juntes en el menú de «condicional formato».

4. Anàlisi de dades

4.3. Eines i mecanismes d'anàlisi

4.3.5. Utilitzar descriptors estadístics típics

4.3.5.1. Introducció

En aquest subapartat treballarem amb l'arxiu: Formules.

A l'hora de treballar amb l'anàlisi de dades, hi ha una sèrie de fórmules i descriptors estadístics bàsics que s'utilitzen pràcticament cada vegada que es busquen històries entre un conjunt d'informació emmagatzemada en un full de càlcul.

L'estructura de les fórmules d'Excel és força bàsica:

a) Sempre es comença amb un signe =, que l'Excel interpreta com que a continuació hi haurà una fórmula.

b) En segon lloc, escriurem el nom de la fórmula. Si no és que estem fent un càlcul personalitzat, les fórmules ja estan registrades en el programa i, per tant, de seguida que comencem a escriure-la el programa ens farà suggeriments.

c) Després del nom de la fórmula s'obre parèntesi per a:

- Indicar el rang de dades sobre el qual es vol aplicar el càlcul. És el cas de les fórmules més senzilles, en les quals es podrà indicar un rang concret o unes columnes senceres concretes. Per exemple, de la casella A2 a l'A34, i ho escriuríem (A2:A34), i de la columna D a la F, escrit així (D:F).

- Complir l'estructura requerida per l'Excel, en el cas de les fórmules amb una sintaxi més complexa, perquè la fórmula es resolgui amb èxit.

Les fórmules es poden arrossegar i aplicar de manera senzilla en totes les files d'una columna. Igual que hem vist en apartats anteriors, ens haurem de posicionar a l'extrem dret inferior de la cel·la, i fer doble clic o arrossegar fins l'última casella a la qual vulguem arrossegar la cel·la, i els valors de les caselles seleccionades aniran canviant automàticament. Si hi hagués alguna casella fixa que volguéssim mantenir, podríem fer-ho amb el símbol $.

Exemple

Per exemple, si volem fixar la casella A24, si escrivim $A24 fixarem la columna, però no la fila; si escrivim A$24 fixarem la fila, però no la columna; i si escrivim A24, la fixarem del tot.

Per a poder seguir amb més detall les fórmules presentades a continuació, les trobareu posades en pràctica en l'Excel adjunt d'aquest apartat. Aquest arxiu fa referència al nombre de matriculats en el curs de 4t. d'ESO a tots els instituts públics de Catalunya per a l'any 2012 i 2013. Trobareu dos fulls, un per a les fórmules bàsiques i un segon per a les avançades.

4.3.5.2. Fórmules bàsiques

1) Sumar

Permet sumar un rang de dades numèriques. Per a aplicar-la cal escriure =SUM(), o =SUMA(), depenent de l'idioma del programa, i seleccionar dins del parèntesi el rang de dades que es volen sumar.

2) Comptar

Permet comptar el nombre de valors numèrics que hi ha en un rang de dades amb la fórmula =CONTAR(), o =COUNT(), segons l'idioma.

En canvi, si el que ens interessa és comptar valors numèrics i alfanumèrics, utilitzarem la fórmula =CONTAR.A(), o =COUNTA.

3) Mitjana

Calcula la mitjana d'un rang determinat de dades. La fórmula és =AVERAGE(), o =PROMEDIO(), segons l'idioma preestablert del programa.

4) Mediana

És un càlcul que ordena de major a menor tot el rang de dades seleccionat, i retorna aquella dada que hi ha exactament en el punt mig de tot el rang. Per a aplicar-la s'utilitza =MEDIAN, o =MEDIANA.

5) Moda

La moda és una mètrica que representa el valor més repetit entre un rang determinat de dades. La fórmula per a calcular-la és =MODE(), si l'Excel és en anglès, o =MODA() i =MODA.UNO(), segons la versió en castellà.

6) Valor màxim

Per a trobar el valor més alt d'un rang de dades determinat, la fórmula és =MAX, tant en la versió en castellà com en anglès.

7) Valor mínim

Per a trobar el valor més alt d'un rang de dades determinat, la fórmula és =MIN, tant en la versió en castellà com en anglès.

8) Percentatge sobre el total

Els percentatges són valuosos perquè aporten informació valuosa sobre el context que envolta una dada concreta. Si volem calcular el percentatge d'un valor respecte al total de la variable, utilitzarem la regla de tres.

Exemple

Per exemple, tornant al cas dels instituts públics, si volem saber quin percentatge representen el nombre d'estudiants de 4t. d'ESO del municipi d'Alcanar, 74, respecte al total matriculats a Catalunya l'any 2012, un total de 17.635, el càlcul que haurem de fer serà: total alumnes d'Alcanar/total alumnes de Catalunya *100.

Cal tenir en compte que si ja hem multiplicat per cent, no caldrà especificar que els valors són percentatge en el format de número. És important recordar-ho, perquè si canviem el format de número a percentatge, automàticament els valors es multiplicaran per cent una altra vegada.

4.3.5.3. Fórmules avançades

1) Comptar els valors en blanc

Si volem comptar el nombre de cel·les en blanc que hi ha en un rang determinat, la fórmula serà =CONTAR.BLANCO(), per a la versió en castellà, i =COUNTBLANK(), per a la versió en anglès.

2) Comptar condicional

La següent fórmula serveix per a comptar només aquells valors del rang que compleixin una condició: =CONTAR.SI(), o =COUNTIF().

Exemple

Per exemple, si volem comptar tots aquells valors d'una columna que són més grans de 15, la fórmula correcta seria: =CONTAR.SI(B:B;“>15”).

3) Sumar condicional

Funciona de la mateixa manera que l'operació anterior, però en aquest cas, suma. Per tant, només suma aquells valors del rang que compleixin una condició: =SUMAR.SI(), o =SUMIF().

Exemple

Per exemple, si volem sumar tots aquells valors d'una columna que són més grans de 15, la fórmula correcta seria: =SUMAR.SI(B:B;">15").

4) Veure si un registre compleix dues o més condicions

Aquesta operació permet veure si una fila compleix condicions en columnes diferents. La fórmula per a aplicar és =Y(), en castellà, i =AND(), en anglès.

Exemple

Per exemple, tornant al cas dels instituts, imaginem-nos que volem saber quins van tenir més de 50 matriculats a 4t. d'ESO l'any 2012 de la totalitat d'instituts, més d'una cinquantena l'any 2013 i que, a més a més, el creixement d'un any a l'altre va ser positiu. Per a veure l'aplicació de la fórmula podeu obrir l'arxiu adjunt i fer clic en el segon full anomenat «fórmules avançades». En la columna G hi ha la següent fórmula: =Y(E2>50 ";F2>50;F2>0).

D'aquesta manera, si en la fila 2, la que correspon a l'únic institut d'Alcanar, els alumnes de 4t. d'ESO de l'any 2014 són més de 50 (columna E), l'any 2012 també són més d'una cinquantena (columna F), i el creixement (columna G) d'un any respecte a l'altre ha estat positiu, el resultat definitiu serà «VERDADERO». En canvi, el resultat serà «FALSO» tant si no en compleix cap, o només en compleix alguna.

Amb la creu negra que apareix a l'extrem inferior dret de la cel·la, arrossegarem la fórmula a la resta de la columna.

5) Veure si un registre compleix alguna condició

Aquesta fórmula té un funcionament pràcticament igual a l'anterior, amb l'única diferència que en aquesta versió indicarà «VERDADERO» quan compleixi alguna de les condicions especificades, i només indicarà «FALSO» si no en compleix cap.

Per a aplicar-la haurem d'utilitzar =O(), en la versió castellana, i =OR() en l'anglesa. En l'exemple anterior quedaria formulada de la següent manera en la columna H: =Y(E2>50 ";F2>50;F2>0).

6) Condicional

La funció =SI() o IF() permet veure si una cel·la compleix una condició concreta en una columna, i en el cas que sí o que no, poder customitzar la resposta.

Exemple

Tornem a l'exemple dels alumnes d'ESO. Per exemple si l'únic institut d'Alcanar va tenir més de 30 alumnes de batxillerat l'any 2012, voldrem que la cel·la indiqui que és «dins de la ràtio», i si no els ha tingut, que aparegui «fora de la ràtio».

La fórmula correcta serà: =SI(D2>30; "dins la ràtio"; "fora la ràtio").

7) Traslladar valors per una columna en comú

Aquesta fórmula serveix per a trobar valors iguals que hi ha en fulls d'Excel separats. Per a entendre-ho millor, caldrà obrir l'Excel adjunt del nombre d'inscrits en els instituts de Catalunya i veure l'operació completa en la columna E del full de «fórmules avançades». Imaginem-nos que volem unir el full de «matriculats 2012» amb el de «matriculats 2013». Per a copiar les xifres d'una manera senzilla, ràpida i sense caure en errors, utilitzarem la fórmula =BUSCARV(), o =VLOOKUP().

8) Variació percentual

La variació percentual és un càlcul que ens permetrà saber quina ha estat la variació al llarg del temps d'una variable determinada.

Es calcula de la següent manera:

(Valor Final – Valor Inicial)/Valor Inicial *100

Exemple

Per exemple, si el nombre d'alumnes matriculats a 4t. d'ESO en els instituts públics de Catalunya l'any 2012 era de 17.635, i el 2014 era de 17.436, la variació percentual haurà estat de –1,2%. Per a veure amb més detall el càlcul, aneu directament a la columna F del segon full, anomenat «fórmules avançades».

4. Anàlisi de dades

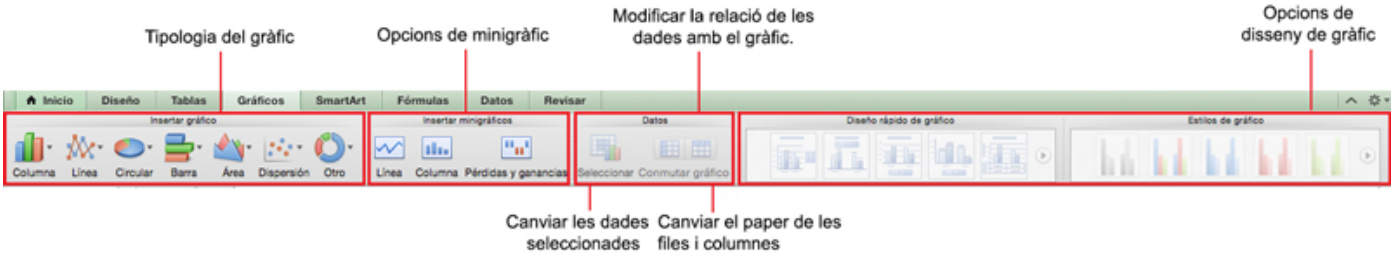
4.4. L'anàlisi visual

4.4.1. Introducció

En aquesta unitat treballarem amb l'arxiu: Grafics.

Ara per ara hem vist com fer una anàlisi de dades mitjançant taules dinàmiques, càlculs estadístics i fórmules específiques, però hi ha un altre tipus d'anàlisi que fins ara no hem tractat: l'anàlisi visual. Fer gràfics com a mètode d'exploració de dades ens pot permetre trobar tendències i patrons que d'altra banda serien difícils de localitzar.

Hi ha programes específics, com Tableau o Quadrigram que permeten un ventall molt ampli de gràfics amb els quals es poden explorar les dades d'una manera diferent. Però sense anar més lluny, l'Excel ofereix unes possibilitats de gràfics amb els quals també es pot explorar el conjunt de dades des d'una altra perspectiva. Abans de veure'n el funcionament, donem un cop d'ull a la interfície gràfica d'aquesta part del Microsoft Excel.



Interfície gràfica del software Excel

El primer bloc, el de tipologia de gràfic, fa referència al tipus de visualització que volem dur a terme: barres, línies, colors, àrees, dispersió. La idoneïtat de cada gràfic es tractarà específicament en l'assignatura de visualització.

El segon bloc és l'espai reservat per als «minigràfics», que no treballarem en profunditat ja que les opcions dels gràfics del primer bloc ja contenen totes les opcions que podem necessitar per a fer una anàlisi visual.

El tercer bloc serveix per a modificar la relació de les dades amb el gràfic, és a dir, per a ampliar o restringir les files i les columnes seleccionades, i per a intercanviar el valor que tenen en els gràfics. Finalment, el darrer bloc serveix per a modificar qüestions estilístiques i de disseny un cop tenim fets els gràfics.

4. Anàlisi de dades

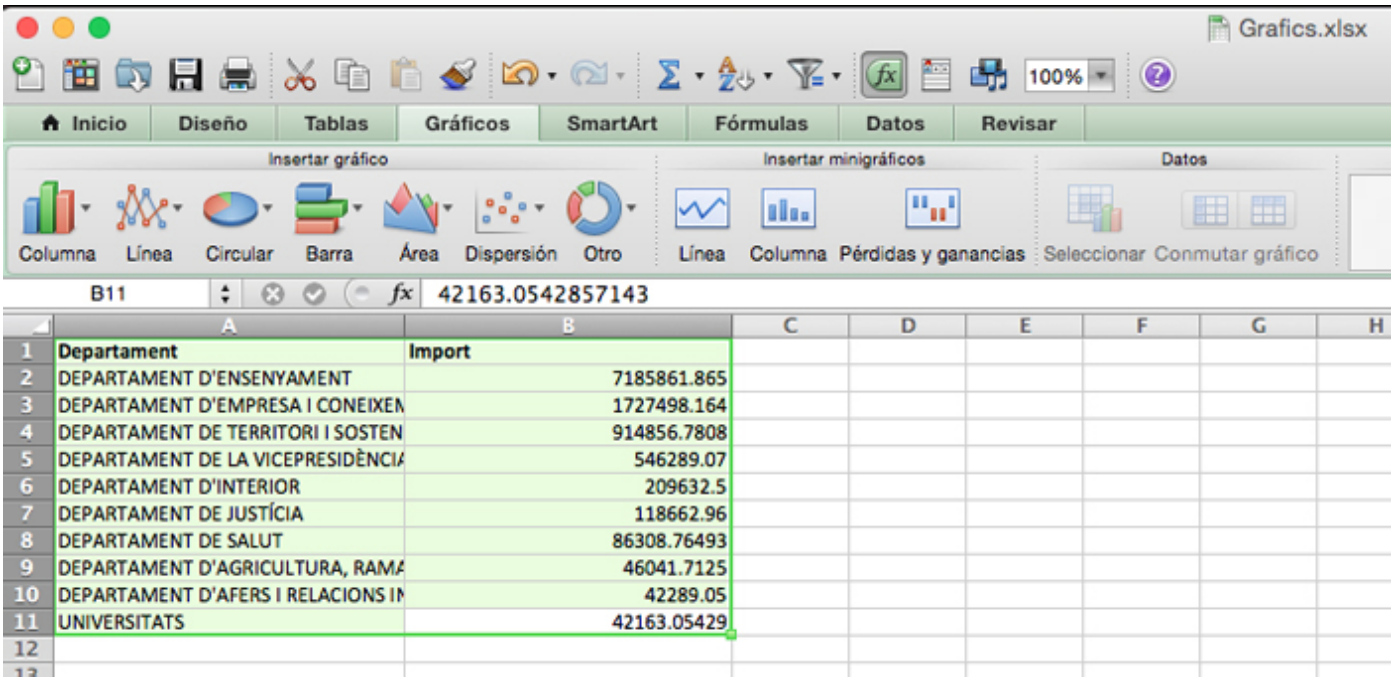
4.4. L'anàlisi visual

4.4.2. Un exemple pràctic amb Microsoft Excel

Per a poder fer un gràfic amb Excel cal tenir en un full a part, netes i seleccionades, les dades que pretenem representar. Si tornem a l'exemple de les adjudicacions, un cop hem elaborat la taula que ens permet veure quin import en adjudicacions s'ha gastat cada departament, també podem veure quin ha estat el que més i quin el que menys ha gastat, si ordenem la taula de menor a major. Tot i així, a simple vista ens costa veure quina és la relació entre totes les xifres, és a dir, si són totes similars o si n'hi ha alguna que es dispara respecte a la resta.

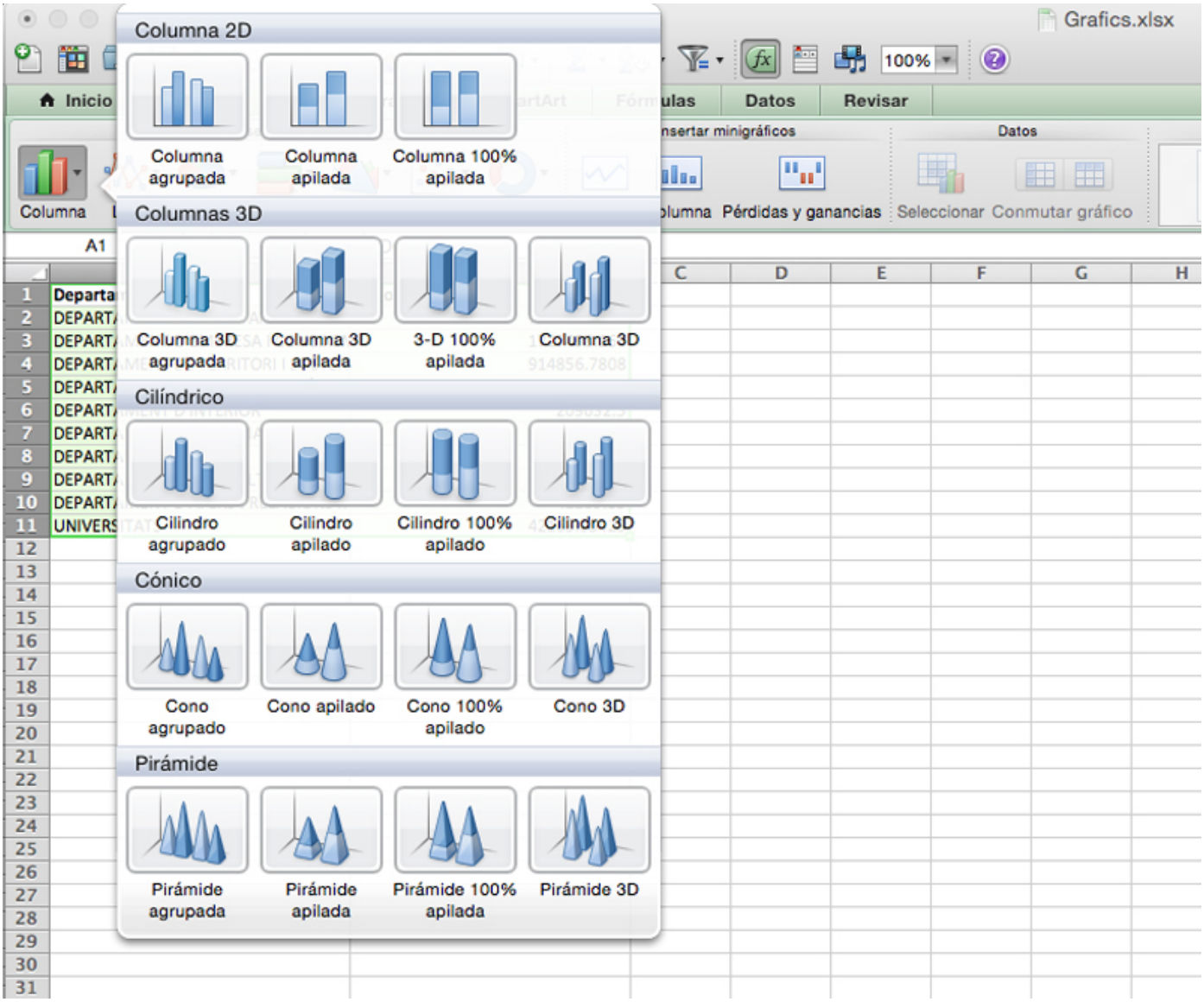
L'anàlisi visual ens permetrà veure d'una manera senzilla i ràpida la relació que tenen les diferents xifres entre elles.

Per a poder fer el gràfic, el primer que hem de fer és seleccionar les dades que volem representar, i fer clic en la pestanya «gràfics».



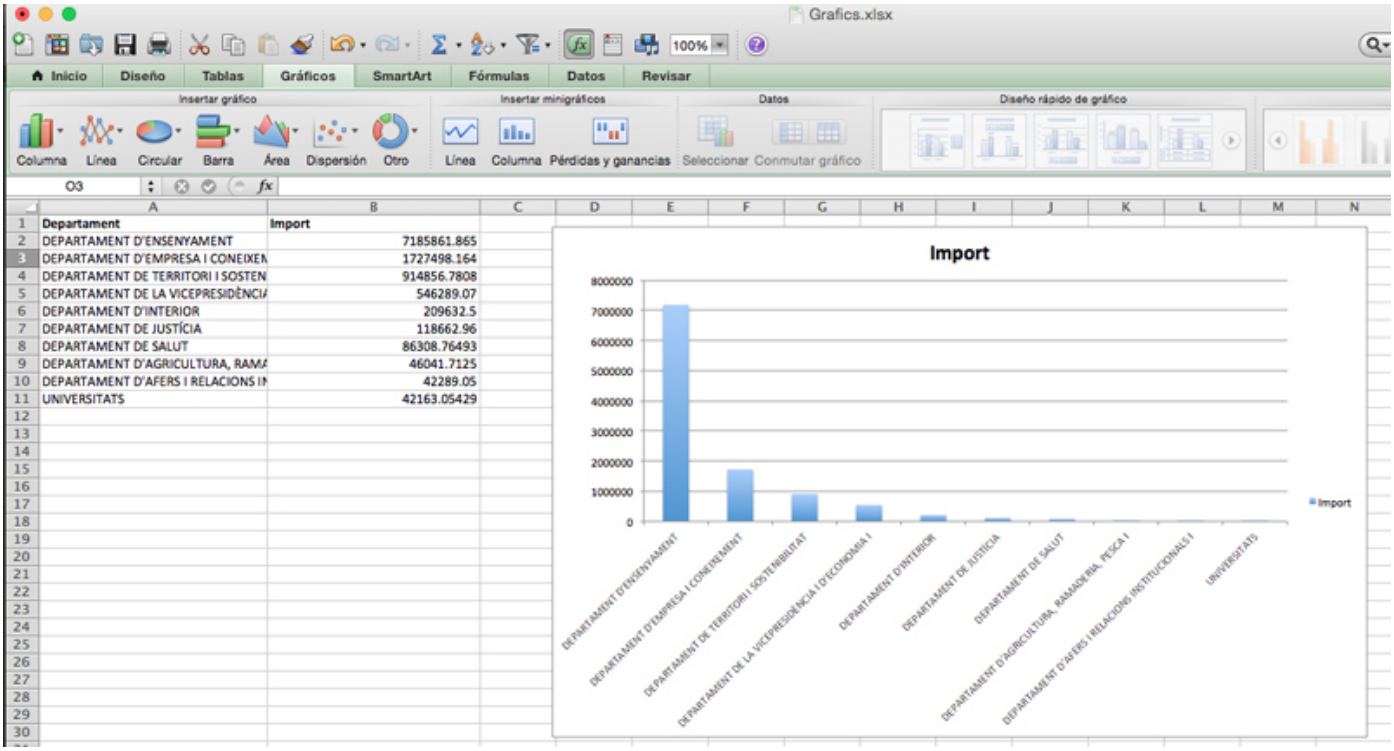
Visualització de la selecció de dades per gràfica

Una vegada situats al damunt, haurem de triar el gràfic pertinent per a representar les dades. En aquest cas, el més indicat serà un gràfic de columnes o de barres en el qual cadascun dels departaments sigui una barra o una columna. Això ens permetrà poder veure com són de més grans les xifres sobre la resta, i a l'inrevés. En aquest cas, escollim elaborar un gràfic de columnes.



Selecció de la tipologia de visualització gràfica

Dins de l'opció de la tipologia de gràfics de columnes, ens apareixen d'una banda opcions purament estilístiques: columnes, columnes 3D, cilindres, cons i piràmides. Si deixem de banda aquesta part més estètica, que en el cas de l'anàlisi visual no aporta res, seleccionarem qualsevol de les tres opcions de columnes 2D. Com que només estem visualitzant una variable, el resultat serà igual cliquem el que cliquem: un gràfic en el qual cada departament està representat en una columna. En el cas que hi hagués dues variables, sí que hauríem de filar més prim a l'hora d'escollir un gràfic o un altre, i decidir si volem les variables en barres una al costat de l'altra o apilades. D'aquesta manera, ja tindrem el gràfic elaborat.



Visualització del gràfic elaborat

Sense fer cap retoc estilístic, el gràfic hauria de sortir per defecte tal com es mostra en la captura de pantalla anterior. Si fem clic sobre el gràfic, ens apareixeran habilitades les opcions dels blocs tres i quatre que hem vist al principi d'aquest apartat: relació amb les dades i el disseny dels gràfics.

Si el gràfic no s'ha mostrat tal com desitjàvem, podrem fer variacions amb l'opció de «dades», ja que ens permetrà modificar la relació que tenen les dades amb la visualització. Pel que fa a les opcions estilístiques, no val la pena entretenir-s'hi gaire. Es tracta de veure què és el que ens revelen les dades, no de fer un gràfic publicable. A l'hora de fer gràfics per a publicar hi ha moltes altres eines que ens permetran fer visualitzacions úniques per a cada notícia, segons els requeriments de la història que vulguem explicar. Ara per ara, només estem fent un repàs a l'apartat de gràfics de l'Excel per a veure com podem enriquir-nos-en i aprofundir en l'anàlisi de dades.

Així, doncs, amb un simple gràfic com el que hem elaborat a continuació, veiem com hi ha una xifra que és molt més elevada respecte a la resta, les adjudicacions fetes pel Departament d'Ensenyament, que es disparen amb diferència respecte a la resta. En canvi, tots els altres departaments mantenen uns imports més similars d'adjudicacions.

Ara veurem què ha passat amb les adjudicacions del departament d'Ensenyament. Es tracta d'un contracte molt elevat o de diversos contractes que tenen un import molt alt? L'anàlisi visual d'aquesta part la trobareu en l'exemple 2 del mateix arxiu d'Excel.

En l'apartat de dades tenim la taula original filtrada pels contractes del Departament d'Ensenyament. Veiem que només n'hi ha dos, i són tan elevats perquè estan licitant, per mitjà de dos contractes, el subministrament de gas natural per a tots els centres d'ensenyament públic. Tot i que en aquest cas tindria poc sentit fer una anàlisi visual, perquè només hi ha dues xifres i és fàcil veure la relació que tenen, farem un gràfic ràpid a tall d'exemple.

De nou, seleccionarem les dades i clicarem l'opció de gràfic que més ens convingui. En aquest cas podria ser un gràfic de barres, columnes o circular. D'aquesta manera tindrem un gràfic que ens permet veure d'una manera ràpida la relació entre ambdues variables.

4. Anàlisi de dades

4.5. Com trobar notícies a partir de l'anàlisi

4.5.1. Introducció

Les dades poden tenir papers molt diferents a l'hora d'elaborar una notícia. Poden ser el suport d'una història que ja tenim localitzada, la base d'una investigació o l'element conductor d'una notícia. Vegem diferents exemples de notícies publicades en diaris generalistes en els quals les dades tenen papers diferents.

4. Anàlisi de dades

4.5. Com trobar notícies a partir de l'anàlisi

4.5.2. L'anàlisi de dades com a element conductor de la notícia: el *Data Driven Journalism*

1) *EL CONFIDENCIAL* | No solo se equivoca Errejón: 164 diputados han roto la disciplina de voto en la legislatura

Aquesta notícia de l'equip de dades d'*El Confidencial* se centra en els errors de les votacions en el Congrés dels Diputats. Durant gairebé cinc mesos, els periodistes es van dedicar a anotar les votacions de cadascun dels diputats cada cop que hi havia votacions a l'hemicicle.

LA MITAD DE LAS VOTACIONES, CON ALGÚN ERROR

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

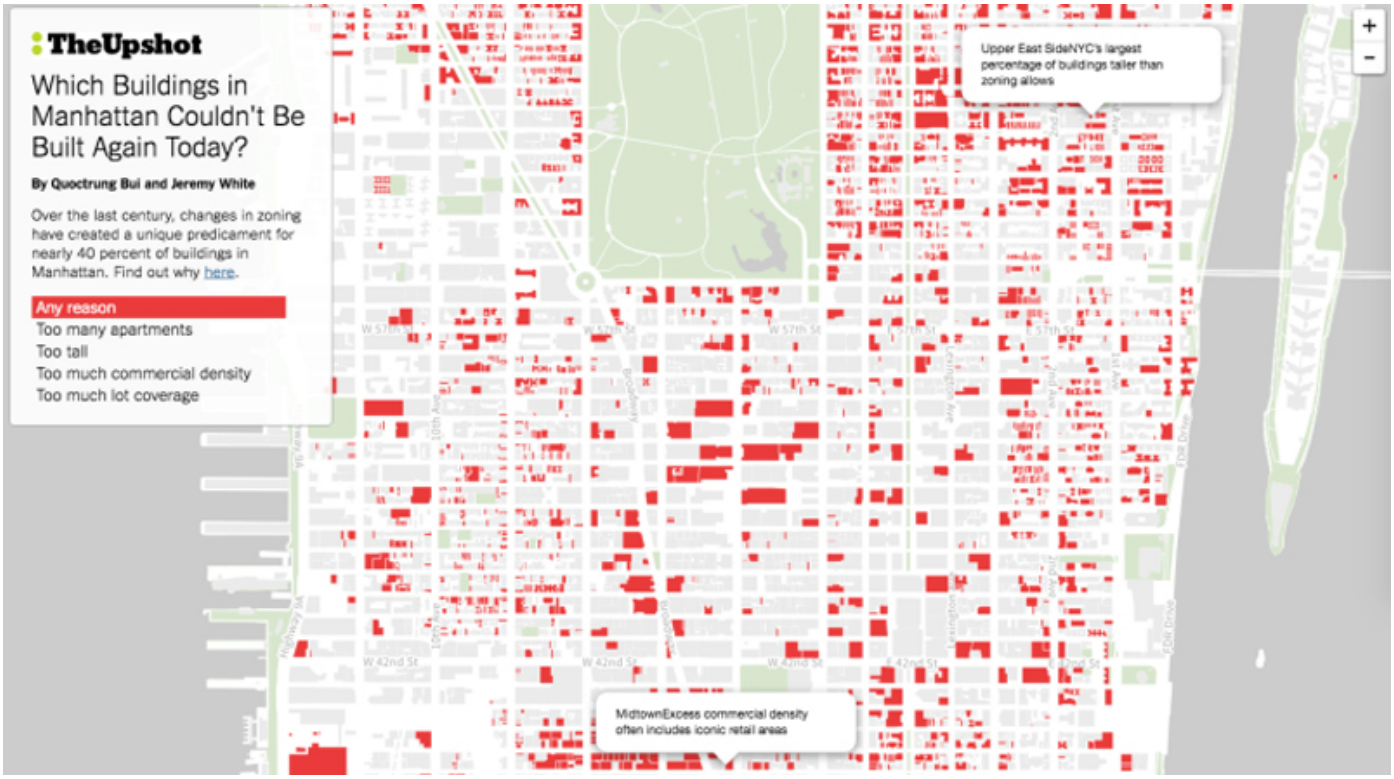
Los errores en las votaciones están a la orden del día en el Congreso. En cinco meses se han dado 274 casos de 'transfuguismo' parlamentario a la hora de pulsar el botón correcto

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

Gràcies a la constància, aquesta notícia va poder veure la llum. Es tracta d'un exemple en el qual les dades són l'element central de la notícia perquè se situen en el centre i es converteixen en el valor afegit informatiu de l'element que articula la publicació.

2) *THE NEW YORK TIMES* | 40 Percent of the Buildings in Manhattan Could Not Be Built Today

Amb la legislació vigent a la mà, i amb una base de dades sobre els edificis que hi ha a Manhattan, els periodistes del *New York Times* van esbrinar que el 40% de les edificacions construïdes avui dia no estarien permeses.



3) *LA VANGUARDIA* | La cúpula de TMB cobijó a tres directivos 'doscientosmileuristas' durante la crisis

Aquesta vegada, mitjançant una petició d'informació pública, es van obtenir les dades referents als salaris dels alts directius de la cúpula de Transports Metropolitans de Barcelona des del període de la crisi fins a l'actualitat. L'anàlisi de les dades rebudes i el creuament de dues informacions diferents va ser clau per a poder desenvolupar la notícia *a posteriori*.

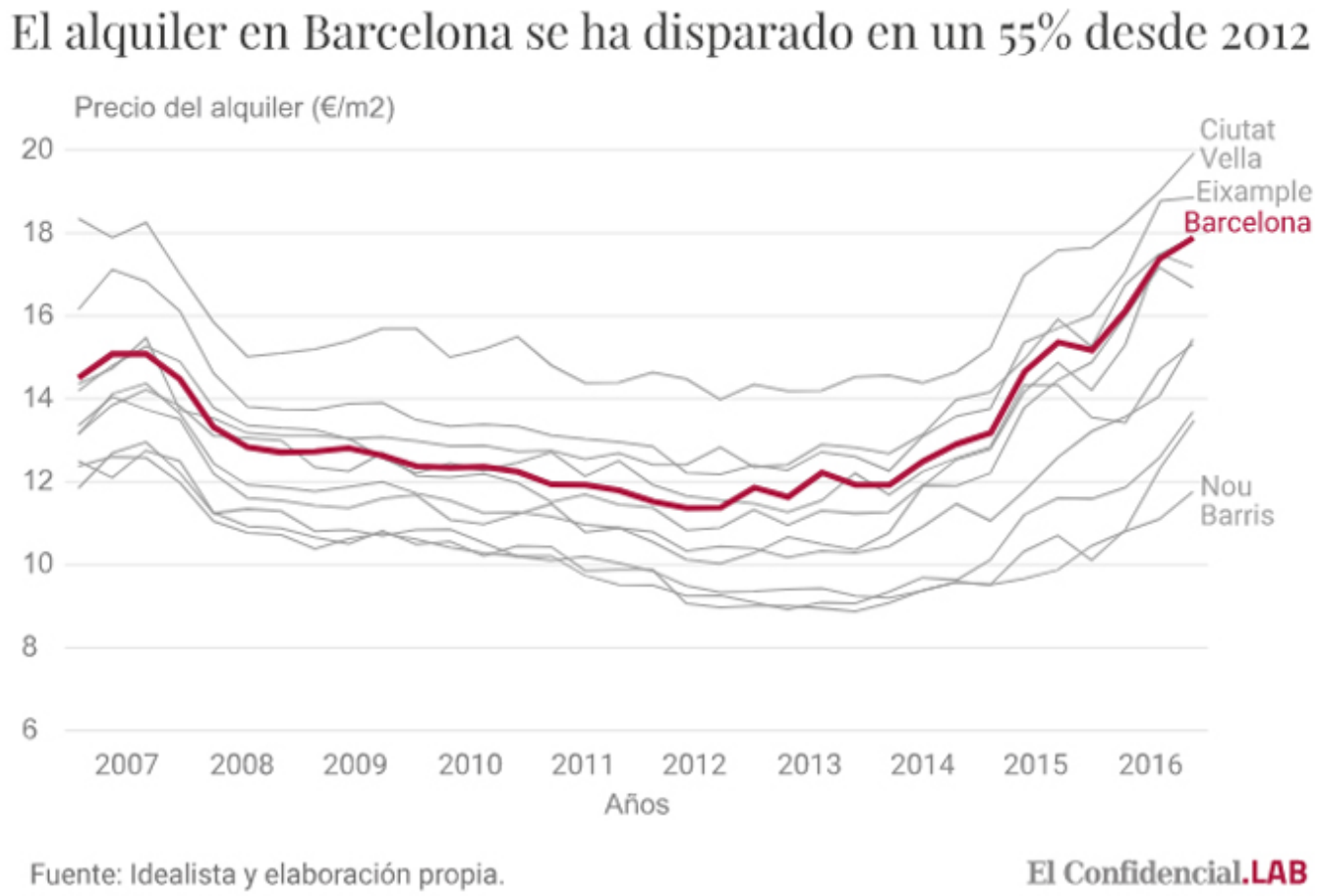
4. Anàlisi de dades

4.5. Com trobar notícies a partir de l'anàlisi

4.5.3. Les dades com a suport de la notícia

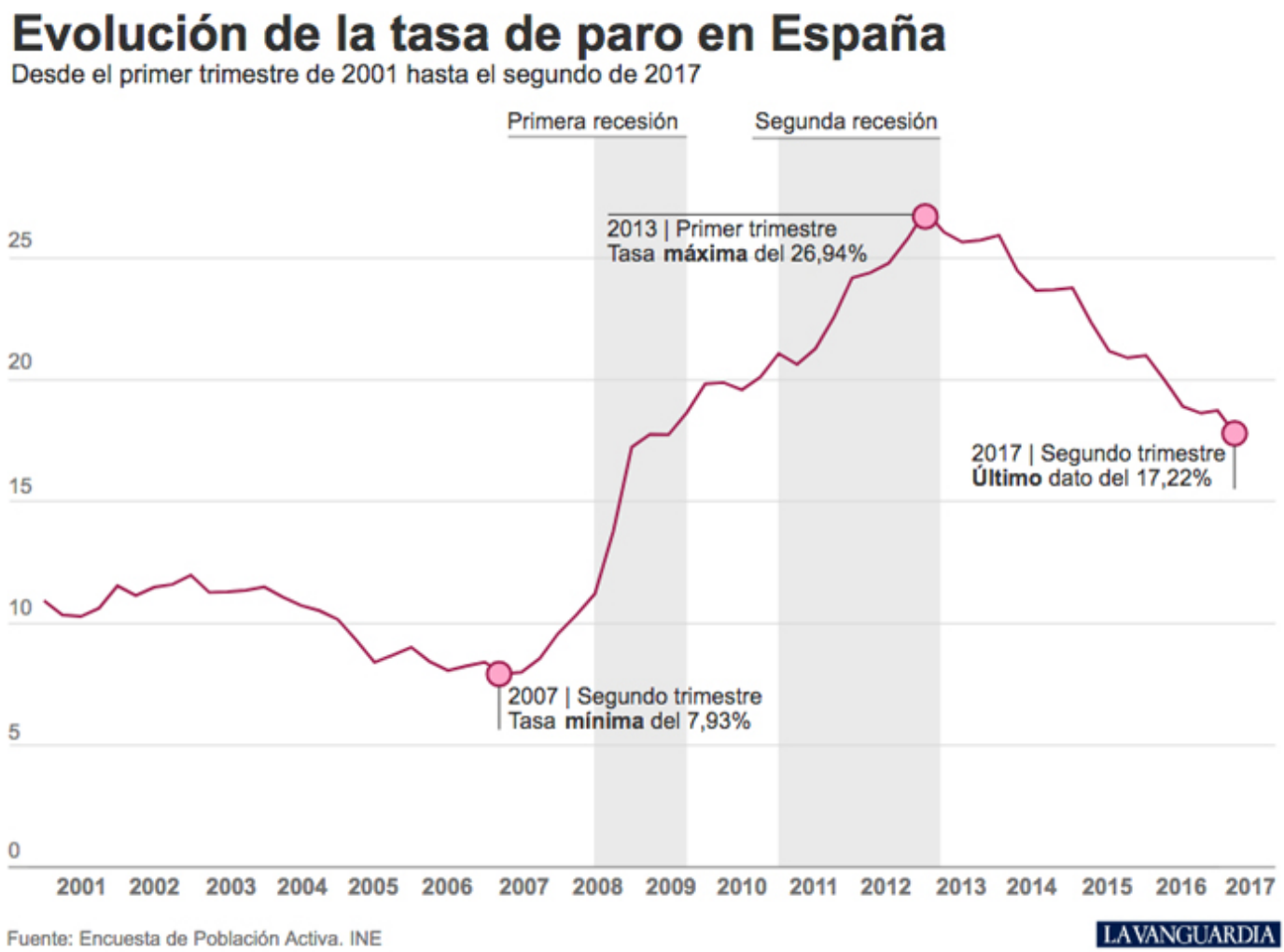
1) *EL CONFIDENCIAL* | Mil euros por 28 m² en Vallecas: el alquiler enloquece y supera las cotas pre-crisis

En aquest cas, les dades sobre la evolució del preu de la llar, tant a Barcelona com a Madrid, i la seva anàlisi i visualització, donen un valor afegit a una notícia. La doten de context i de maneres gràfiques per a entendre el moment actual d'una situació determinada, en aquest cas, l'augment del preu de la llar.



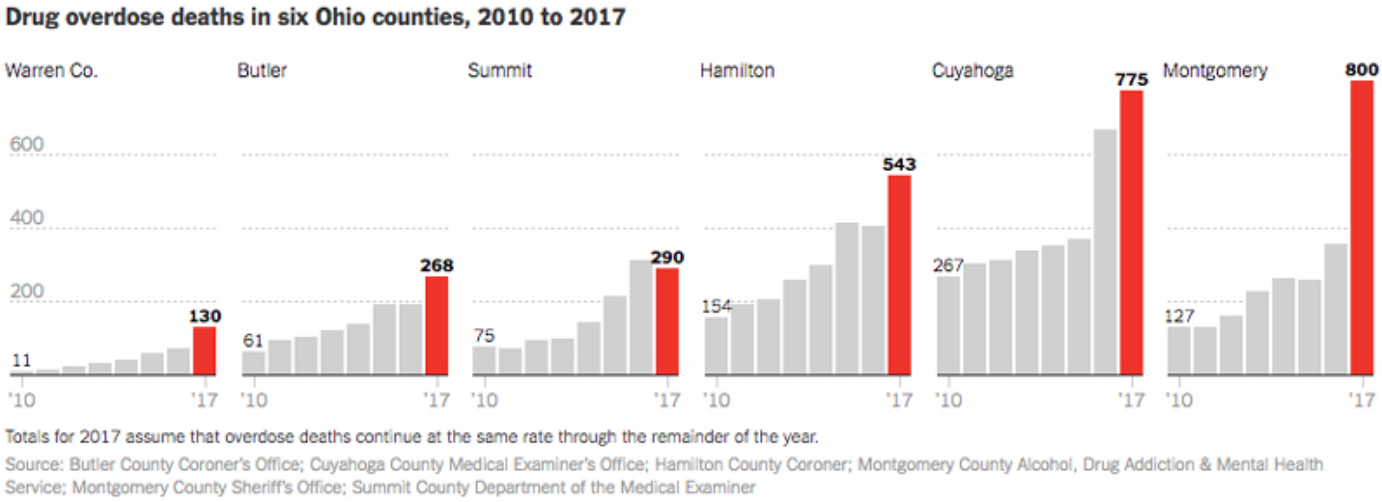
2) *LA VANGUARDIA* | España asienta una salida de la crisis que no logra frenar los desequilibrios

Aquest és un altre exemple de com les dades i la seva visualització poden ajudar a donar suport i fer més comprensible una notícia, i a dotar-la d'un contingut més rigorós.



3) *THE NEW YORK TIMES* | Les morts pel consum de drogues als Estats Units creixen més que mai

Aquesta notícia exemplifica amb gràfics i visualitzacions com el consum de drogues torna a causar estralls als Estats Units. El tractament que s'ha fet de les dades, i la manera de mostrar-les, permet al lector entendre encara millor el fenomen que es vol explicar.



4. Anàlisi de dades

4.5. Com trobar notícies a partir de l'anàlisi

4.5.4. Les dades com a punt de partida d'una investigació

1) *PROPUBLICA* | Killing Colorado

És una investigació periodística feta per l'organització nord-americana *ProPublica* amb l'objectiu de denunciar l'estat deplorable del riu Colorado per culpa de la intervenció humana sense mesura. ProPublica és una organització sense ànim de lucre que compta amb una unitat molt potent de periodistes de dades que treballen per a tirar endavant investigacions periodístiques.

2) *EL MUNDO* | Football Leaks

Es tracta d'una investigació duta a terme per EIC, la xarxa europea de periodistes d'investigació, que a l'Estat espanyol duia a terme el diari *El Mundo*. Aquesta investigació va destapar la cara oculta del futbol: l'entramat financer creat per tal que els futbolistes mileuristes eludeixin impostos davant d'Hisenda.

3) *EL CONFIDENCIAL* i LA SEXTA | Papers de Panamà

La major filtració de la història no s'hauria pogut gestionar si no fos gràcies a la utilització d'eines de tractament de bases de dades. El Consorci Internacional de Periodistes d'Investigació (*) va tenir accés a milers d'arxius procedents d'un bufet d'advocats

panameny, el despatx de Mossack Fonseca, que es dedicava a gestionar societats *offshore* de ciutadans procedents de països d'arreu del món.

4. Anàlisi de dades

4.5. Com trobar notícies a partir de l'anàlisi

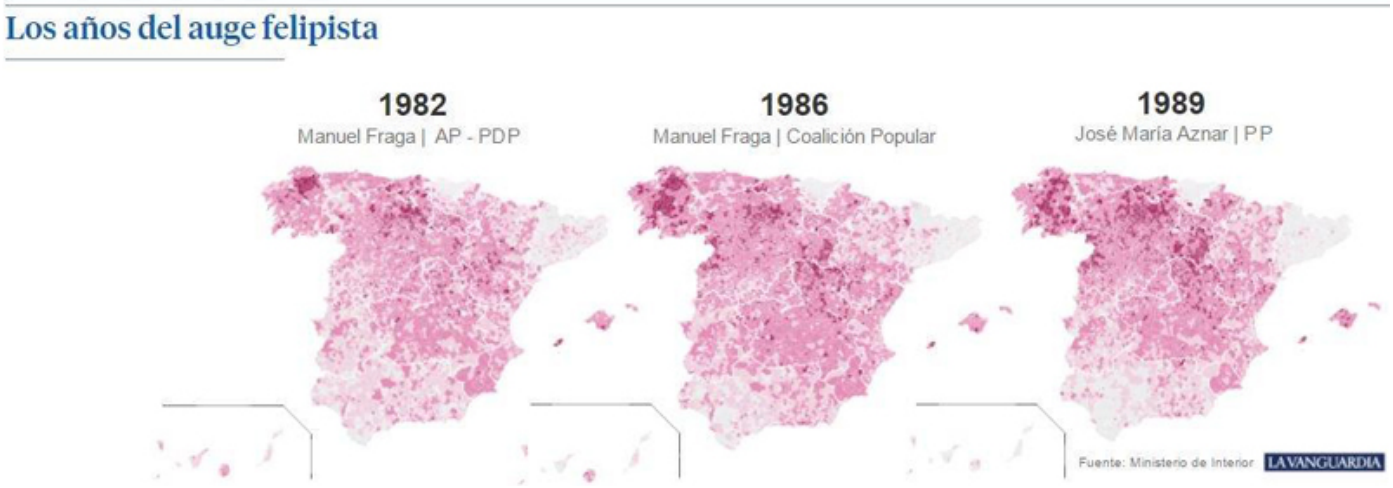
4.5.5. La visualització com a centre de la notícia

1) *EL CONFIDENCIAL* | Radiografia sobre els feminicidis

En aquest monogràfic especial, l'equip de dades del diari *El Confidencial* va recopilar dades sobre tots els feminicidis que hi havia hagut a Espanya des de la implantació de la llei de la violència de gènere. En aquest cas, la visualització interactiva té un paper central i indiscutible en la presentació del tema i del tractament.

2) *LA VANGUARDIA* | Los 'segundones' de la democracia

Aquest recull de mapes fa un repàs sobre el recorregut de la segona força, és a dir els perdedors, en cadascun dels comicis de la democràcia espanyola. En aquest cas, de nou, les visualitzacions prenen un paper central en l'element informatiu de la notícia.



3) *ELDIARIO.ES* | Més de 14.000 personas han muerto en el Mediterráneo desde 2014

Aquest reportatge utilitza les dades sobre el recompte de morts al mar Mediterrani per a fer una aproximació al desastre humanitari de les costes mediterrànies. Les dades, i sobretot les visualitzacions, són un element clau ja que se situen en el centre de l'element informatiu i aporten una manera diferent d'entendre els números, però, sobretot, també el context.

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

5. Redacció i visualitatzació

5.1. La construcció de la notícia: la teoria de les 6w, la piràmide invertida i les entradetes

La teoria de les 6W es basa en les sis preguntes que tota informació, particularment aquella que tingui caràcter noticiós, ha de poder respondre amb la finalitat que el lector sàpiga de manera clara i completa què s'està explicant. Aquestes preguntes són, en la seva expressió anglesa: who? (qui o a qui?), what? (què?), when? (quan?), where? (on?), why? (qui?) i how? (com?).

A l'hora de redactar una notícia els mitjans de comunicació escrits han optat per respondre aquestes preguntes seguint l'esquema de la piràmide invertida, o el que és el mateix, el primer paràgraf de la informació (a partir d'ara, entradeta o *lead*) s'han de resoldre les sis preguntes o, com a mínim, aquella que resulti més important per focalitzar el lector i, a mesura que avança el cos de la peça, s'han de contestar d'una manera més o menys explícita la resta. Aquesta forma d'organitzar la notícia ha estat la més estesa i encara avui és la que fan servir prioritàriament molts mitjans escrits, perquè és una manera senzilla i eficaç de jerarquitzar la informació, i garanteix una comprensió immediata per part del lector encara que no arribi a llegir el text sencer.

Piràmide invertida, nivell bàsic d'utilització

El lead té una rellevància cabdal. Va seguit d'una ampliació de les respostes a les 6W i de la presentació de la informació de context en forma decreixent (situant sempre el més important en els primers paràgrafs i el menys important, en els darrers). En un moment d'explosió dels suports periodístics i de boom informatiu, d'aquest primer paràgraf depèn en bona mesura que el lector s'aferri al text o que el descarti. Per això, David Randall, proposava un seguit de claus per redactar una bona entradeta:

1. Directa, concisa i inquietant: Suprimir tot allò que la recarregui així com càrrecs i noms que es podrien deixar pel segon paràgraf, en cas que no siguin necessaris.

2. L'entradeta ha de funcionar de manera independent.

3. No començar la informació amb una clàusula subordinada, del tipus: "Malgrat que el nombre d'homicidis augmenta..."

4. No començar una història amb nombres en dígits, escriure'ls en lletres o buscar una altra manera de començar si la xifra és molt gran.

5. No començar una informació amb noms d'organismes oficials, potser millor començar amb una altra fórmula, com per exemple: "experts alerten que..."

6. Evitar, en la majoria dels casos, començar amb cites (pot confondre el lector si no identifiquem qui parla immediatament).

La majoria de persones que visiten una notícia a la web centren la seva màxima atenció en la part superior de la peça informativa i l'escanegen en forma d'"F", un esquema compatible amb el de la piràmide invertida. Així ho demostren els estudis d'eyetracking.

La transformació de la piràmide invertida a l'era digital

Amb la crisi del periodisme imprès i l'explosió d'Internet, la majoria de lectors fan servir ara l'ordinador, els dispositius mòbils o les tauletes per llegir les notícies. Per aquesta raó, la piràmide invertida s'ha hagut de recontextualizar per adaptar-se a la hipertextualitat, multimedialitat i interactivitat pròpies d'Internet. Segons el periodista Guillermo Franco (*), en l'escriptura per Internet és important captar l'atenció dels usuaris brindant-los informació a l'instant, en el que ha anomenat "una piràmide invertida horitzontal". Aquest concepte suggereix posar a l'esquerra de l'oració les paraules més importants de la notícia. Hi ha tres nivells d'ús de la piràmide invertida quan s'escriu una notícia per un mitjà digital.

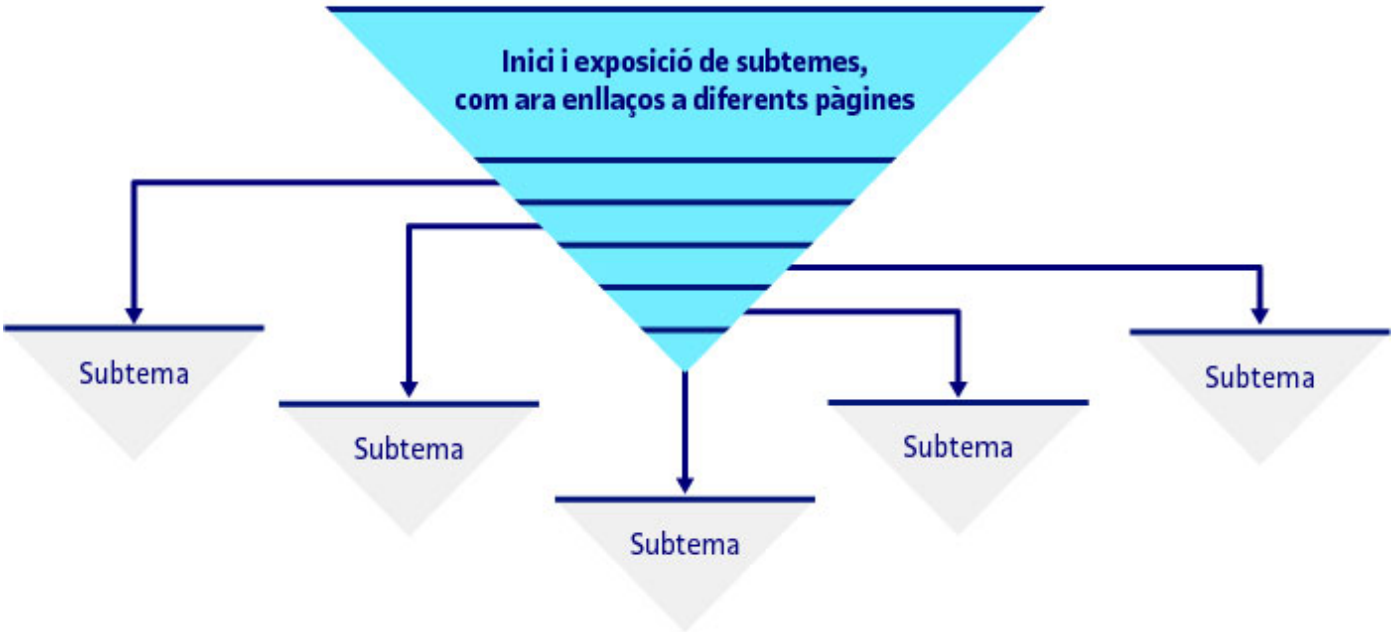
1. El lineal o bàsic: Començar el text amb el més rellevant – les 6W – i anar presentant la resta de paràgrafs en funció de la seva importància per entendre el tema. Això ha d'aparèixer en una mateixa pàgina web.

2. L'estructura: En aquest segon nivell el periodista ha de jerarquitzar la informació, creant un primer paràgraf en què presenti les dades i després, mostrant un llistat amb els intertítols en els quals es divideix la informació.

3. Els subtemes: El darrer nivell és la creació d'un contingut dividit en subtemes que ja no estarien en la mateixa pàgina web, sinó en diferents pàgines vinculades al lloc web inicial. A la pàgina inicial s'ha de presentar el lead i després s'exposen els subtítols que enllacen a l'ampliació del contingut. Quan es clica sobre cada un d'aquests subtítols s'obra una pàgina nova i el lector pot veure un contingut autònom i complet que també pot tenir altres enllaços per accedir a una altra part del contingut. A aquestes estructures Guillermo Franco les denomina "Piràmides Flotants".

Piràmide invertida, tercer nivell d'utilització

Piràmides flotants situades en diferents pàgines web.
L'usuari escull la ruta de navegació i construeix la seva piràmide invertida
a partir de la presentació i de l'exposició del tema



L'adaptació de la redacció i la transformació de la piràmide invertida amb el boom de l'Internet i el sorgiment de nous mitjans digitals fa que sigui necessari introduir el concepte de l'hipertext. L'hipertext configura un nou llenguatge per a comunicar-nos, construït a partir de frases curtes, amb enllaços per aprofundir i amb elements multimèdia per enriquir la informació. Al mateix temps això representa:

- Una nova manera de llegir la informació: La lectura ha deixat de ser passiva per convertir-se en activa, la manera en que l'usuari llegeix és totalment diferent. Abans rebia indicacions dels creadors de continguts sobre el que havia de llegir i com fer-ho. Ara, el lector és un individu lliure i sense lligams, que pot navegar com vulgui.

- Una nova manera d'organitzar i presentar la informació: No hi ha límits d'espai per presentar la informació. L'organització de les dades en una xarxa de nodes interconnectats mitjançant enllaços o hipervincles permet aprofundir de manera quasi infinita.

- Enllaços: Els hipervincles són una part essencial de l'hipertext i mitjançant clics aconsegueixen portar l'usuari a contextualitzar la informació, no només amb publicacions relacionades, sinó amb qualsevol dada possible que pugui aportar alguna cosa més a la lectura.

- Les respostes a les preguntes de les 6W exigien la contextualització del fet narrat, el que implicava la necessitat d'un espai físic important on s'hi acumulava la informació. Si al lector només li interessaven algunes d'aquestes dades de context podia veure's saturat per la multitud d'informació i acabar desconnectant del text. Amb l'hipertext, per contra, la contextualització es resol d'una manera molt més eficaç, perquè els hipervincles generen la possibilitat d'anar saltant per les dades relacionades amb la informació.

Malgrat que el format web ofereix eines que possibiliten una millor organització de la infromació, el periodisme no ha de perdre de vista que no tots els textos necessiten una estructura hipertextual, i el seu abús també és un error. El primordial és que el periodista sigui conscient que ha d'escriure sempre pensant en el lector, perquè és ell qui finalment valida la seva existència quan decideix llegir o navegar pel contingut.

També hi ha un altre tipus de piràmide vinculada al periodisme de dades, creada per Paul Bradshaw (*) (2017), que descriu els passos per convertir la informació bruta a la que accedeix el periodista fins a un producte que pugui ser publicable en un mitjà digital. La lògica en aquest cas és diferent, i té més a veure amb la relació entre el redactor i la informació que recull per a elaborar un reportatge fent servir dades.

1. Compilar la informació.

2. Netejar-la fins a deixar allò més rellevant.

3. Contextualitzar les dades.

4. Combinar-les per, finalment, publicar la informació.

Font: Paul Bradshaw (2017).

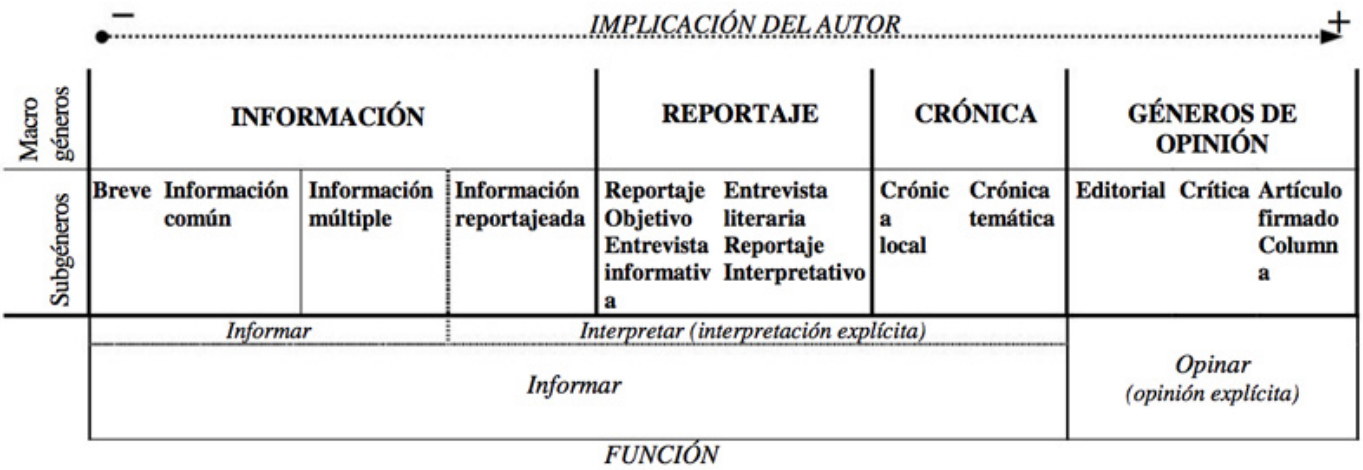
5. Redacció i visualitatzació

5.2. Tipologia de peces periodístiques: els gèneres periodístics

5.2.1. Introducció

En termes generals, existeixen tres gèneres periodístics que es divideixen en tres grans blocs, en funció de la implicació de l'autor amb el text: el gènere informatiu, el gènere interpretatiu i l'opinió. Aquests gèneres també impliquen un menor o major grau de creativitat i de subjectivitat. A continuació fem un recull dels principals trets característics dels tres gèneres i subgèneres que identifica Sonia F. Parrat (*) (2008).

CLASIFICACIÓN DE LOS GÉNEROS PERIODÍSTICOS



Font: Sonia F. Parrat (2008)

5. Redacció i visualitatzació

5.2. Tipologia de peces periodístiques: els gèneres periodístics

5.2.2. Gènere informatiu

5.2.2.1. Introducció

Idealment, el paper del redactor en els gèneres informatius o noticiosos ha de ser el d'un mer observador de la realitat. Té una implicació mínima en el fet noticiós i la seva tasca consisteix en descriure els fets de la manera més objectiva possible (assumint que l'objectivitat no existeix). Se li exigeix veracitat, contrastar la informació i aportar-hi pluralitat. Aquesta darrera exigència és bàsica: mai s'ha de donar per vàlida ni recollir en exclusiva ni la versió oficial dels fets ni una única veu; per a redactar una bona informació el redactor s'ha d'encarregar que la peça reculli fidelment les diferents "cares de la veritat".

A grans trets, totes les informacions contenen els següents elements:

- Estan encapçalades per un titular (que a vegades va acompanyat d'un avantítol, per situar geogràficament o temàticament la informació; i/o un subtítol, que acostuma a complementar el titular). El títol de la informació ha de ser breu –el nombre de paraules depèn del mitjà en què es treballi, però mai convé superar el parell de línies–, clar i atractiu. la seva funció és la d'aportar un resum de la informació que trobarem al cos de la informació.

- El lead o entradeta.

- El cos, que desenvolupa i completa la informació seguint l'esquema de la piràmide invertida, és a dir, explicant els fets en ordre d'importància decreixent, deixant les dades més secundàries pel final. La informació es divideix en paràgrafs que han d'estar interrelacionats temàticament entre si, però que han de ser independents a nivell gramatical. En el cos del text s'amplia la informació del lead i se n'aporta de nova, s'afegeixen declaracions de les persones implicades amb la notícia i s'hi aporta context.

- La notícia és la peça que caracteritza aquest gènere, malgrat que la seva extensió, contingut o enfocament pot variar, creant diferents tipologies de peces informatives, com veurem tot seguit.

5.2.2.2. La notícia

> "La funció d'un diari consisteix en descobrir informació nova sobre assumptes d'interès general i transmetre-la als lectors amb la major rapidesa i precisió possibles, de manera veraç i equànime".
>
> David Randall

> "La funció primordial d'un diari és recollir notícies i, fins i tot a risc de la seva pròpia existència, garantir que es transmetin sense desvirtuar-les. La veritat no es pot deformar amb el que es diu o no es diu, ni amb la manera de presentar-ho".

Això ho escrivia, per la seva banda, el periodista C. P. Scott, director del *Manchester Guardian* el 1921. A aquesta cita, hi afegia: "cadascú és lliure de comentar el que vulgui, però els fets són sagrats".

Les notícies són l'element del que es nodreixen els mitjans de comunicació, i tenen les següents característiques:

- Són quelcom inusual. "Que un gos mossegui un home no és notícia; ho és que un home mossegui un gos". Aquest és el clàssic exemple que citen els professors de primer curs a les facultats de periodisme, però les notícies són més que això.

- Tenen valor informatiu perquè aporten novetat, coses que la gent no ha escoltat abans.

- És fonamental que recullin temes d'interès públic: No només temes que tinguin un impacte en la vida pública sinó també temes que siguin d'interès per al públic.

Pel que fa a les tipologies, es distingeixen quatre varietats relacionades amb el grau de profunditat de la notícia o de la peça informativa:

1) El breu: Una informació curta d'un assumpte sense gran importància que s'insereix en un racó de la pàgina del diari o del web, amb un titular poc elaborat i sense fotografia. La seva funció és la d'informar sobre un fet d'actualitat i el seu redactor es distancia al màxim del que està explicant. Sovint es firma com a "redacció" i moltes vegades la notícia prové d'una agència d'informació, com l'ACN o l'agència Efe, per exemple.

2) Peça informativa bàsica: Tracta d'un assumpte concret de l'actualitat. No compta amb cap afegit ni cap text complementari però porta, sovint, una fotografia o una infografia. La seva funció també és purament informativa i la implicació de l'autor és mínima.

3) Informació múltiple: Diverses informacions connectades entre si que es publiquen en una mateixa pàgina en un diari imprès, o sota un element gràfic que indiqui que estan interrelacionades, en cas que es publiquin en un mitjà digital. Compleix una funció informativa i el grau d'implicació de l'autor és baix però més gran que en les anteriors peces.

4) Informació reportatjada: Es tracta d'un text informatiu redactat amb un enfocament més pròxim al reportatge que ha agafat presència a la premsa per intentar captar l'atenció del lector. No s'ha de confondre aquestes peces amb els reportatges, perquè les informacions reportatjades comuniquen fets d'estricta actualitat i la implicació de l'autor és menor que en el cas del reportatge, tot i que a vegades conté certes dosis d'interpretació. Una peça d'aquest tipus acostuma a anar acompanyada de notícies complementàries, que hi aporten noves dades, o d'un text interpretatiu, una petita entrevista o fins i tot un text d'opinió. També l'acompanyen fotografies o infografies.

5. Redacció i visualitatzació

5.2. Tipologia de peces periodístiques: els gèneres periodístics

5.2.3. Gènere interpretatiu

5.2.3.1. El reportatge

> "El reportatge és el gènere periodístic més complet: a més de incloure altres gèneres – crònica, entrevista, notícia – pot tenir com a antecedent una notícia i ho elabora un periodista cada vegada més especialitzat. Però té una avantatge respecte a la notícia, i és que està deslligat de l'estricta actualitat diària".
>
> Randall (2008)

Com assenyalava David Randall, el reportatge pot sorgir arrel d'un fet noticiós que s'amplia o s'aprofundeix, però també a partir de qüestions d'actualitat o d'interès general però no necessàriament noticiosos. No segueix estrictament les mateixes pautes de la notícia, sinó que s'apropa més a l'estil literari i conté detalls descriptius, frases de testimonis i vivències pròpies.

Els reportatges ocupen més espai que una notícia. El seu contingut no deixa de ser informació (això és, probablement, el que més el distingeix d'un text literari), però així com en les notícies hi ha un motlle definit, en els reportatges no hi ha fórmules determinades. El reporter escull la forma i el to que millor s'acomoda a l'assumpte del que parla, i gaudeix d'una major llibertat estructural i expressiva. Acostuma a anar acompanyat de fotografies o infografies.

El primer paràgraf no té interès informatiu però ha de ser atractiu perquè de la manera com està redactat depèn que el lector llegeixi la resta del text. Les característiques del lead també existeixen però es col·loquen on més li convé al redactor. Per triar la temàtica del reportatge, en certa manera hem de deixar-nos influenciar pel clima d'opinió. Abans de redactar-lo, és convenient elaborar un esquema simple que ens facilitarà el procés d'elaboració del text.

1) Hipòtesi de partida: Creiem que existeix un conflicte, perquè els mitjans/ plataformes ciutadanes/ grups de pressió/ administracions en parlen. Per exemple: A Barcelona s'ha disparat el preu de l'habitatge. S'està forjant una segona bombolla immobiliària que comporta un procés de gentrificació o substitució social que obliga molts dels seus habitants a abandonar els seus barris.

2) Triar les fonts: Per poder recollir la informació ens hem de preguntar qui ho denuncia? (font 1) –plataformes i moviment veïnal–; a qui beneficia aquest conflicte? (font 2) –sobretot als grans tenidors, com les entitats financeres, les immobiliàries o els fons voltor–; qui pot posar context a la història? (font 3) –sociòleg, expert en urbanisme–; en quin context legal s'emmarca la història? (font 4) –un jurista especialitzat en la LAU (la Llei d'arrendaments urbans) pot ser un bon contacte–; i per últim, a qui perjudica aquest conflicte? (font 5) –un testimoni: aportar la visió humana és sempre interessant perquè el lector pugui empatitzar amb la situació. Podríem ampliar el nombre de fonts, però hem de saber tallar a temps per no crear monstres que després no es llegeixi ningú.

El reportatge es caracteritza pels següents trets:

- La seva funció és bàsicament la d'un segon nivell d'informació, és a dir, la interpretació. No conté opinió sinó que mostra els fets i els explica. Exposa dades seleccionades per l'autor i explicades amb un cert grau d'intencionalitat, per això no es pot considerar informació en sentit estricta.

- Hi ha quatre passos bàsics en l'elaboració d'un reportatge: elecció del tema, acumulació de tota la informació possible recorrent a tot tipus de fonts, decisió de l'enfocament que li donarem i les tècniques que s'aplicaran en el procés de redacció.

- L'autor té una gran llibertat que li permet escollir l'enfocament, l'estil i l'estructura que consideri més convenient per al seu text. Especialment en el gènere interpretatiu, és desitjable que tingui un principi captivador, un desenvolupament interessant i un final concret. Es redacta combinant la narració amb les descripcions.

- Els seus objectius són: captar l'atenció del lector, mantenir l'interès i estimular el desig de seguir llegint.

Tipologies de reportatges:

1) Reportatge objectiu (o estàndard): encara que la seva funció bàsica és interpretativa i, encara que el grau d'implicació de l'autor és major que en la informació, acostuma a estar escrit amb un estil poc creatiu, no molt extens, aprofundeix poc i gairebé

no analitza els fets. Acostuma a denominar-se *perfil* al reportatge que se centra en una persona concreta, i *necrològica,* al que es publica després de la mort d'un personatge conegut i fa un repàs dels fets més significatius de la seva vida.

2) Reportatge interpretatiu (o reportatge en profunditat): Aporta antecedents, contextualitza, analitza els fets fins a arribar al fons i preveu possibles conseqüències. La seva funció és interpretativa, però està exempta de les valoracions que conté la crònica o l'article d'opinió. El reporter analitza i el lector treu les seves pròpies conclusions.

5.2.3.2. L'anàlisi

La base d'una anàlisi és la selecció d'un fet d'actualitat que ha estat notícia i que requereix una explicació posterior més seriosa, detallada i especialitzada, de manera que el seu component interpretatiu és alt. Acostuma a encarregar-se a persones que no són necessàriament periodistes però sí expertes en la matèria, i el seu mèrit radica més en la capacitat d'aprofundir i analitzar qüestions complexes que en la seva qualitat literària.

5.2.3.3. L'entrevista

Abans que res, l'entrevista no es refereix al procés d'obtenció d'informació per part del periodista, el qual es posa en contacte amb la persona que li ha de facilitar aquesta informació, sinó al text final, resultat de posar per escrit el que l'entrevistat ens ha explicat. Però una entrevista és molt més complexe que formular un parell de preguntes al final d'una roda de premsa. Perquè el contingut sigui de qualitat (i en especial, si volem obtenir una informació personal de l'entrevistat que vagi més enllà de la seva opinió sobre un tema concret) hem de crear un clima de confiança i de confort. Ha d'existir un acord previ entre entrevistador i entrevistat per pactar aquesta trobada.

Una entrevista és un relat de la conversa entre dues o més persones, una de les quals, assumint el rol d'entrevistador, n'és l'autor. Aquestes són les principals característiques de les entrevistes:

- Conté un cert grau d'interpretació i implicació del periodista, que és major que quan redacta una informació, perquè el redactor no parteix d'uns fets aliens sinó que decideix a qui entrevistar i les preguntes que li formularà. Alguns pensen que l'elecció d'aquestes preguntes pot condicionar les respostes i el to general de l'entrevista. Quan transcrivim les respostes hem de preservar el seu contingut, no en el sentit que no podem editar-les, sinó que hem de procurar que la nostra intervenció no alteri el seu sentit, ni la intencionalitat, ni l'ambient dialèctic en que es van produir.

- L'autor de l'entrevista ha de saber què desitja obtenir abans de realitzar-la, fet que exigeix una recopilació prèvia de documentació sobre l'entrevistat i la preparació de les preguntes en funció de la informació que es vol recollir.

- En el transcurs de l'entrevista, les preguntes han de ser el més concretes i directes possibles.

- L'entrevistador ha de ser capaç de crear un bon ambient però no parlar massa, per obtenir la major informació possible del protagonista (aprendre a jugar amb els silencis).

- També ha de tenir la capacitat de veure, observar, escoltar, improvisar i percebre tot allò que un expressa sense paraules, amb gestos, moviments, to, etc. És el que s'acostuma a denominar informació subliminal.

- L'objectiu últim que persegueix l'autor és fer que el lector se senti quasi testimoni de la conversa.

La Doctora en periodisme Montserrat Quesada proposa una classificació de les entrevistes en dues grans tipologies:

1) Entrevista informativa (o objectiva): Recull les declaracions que persones implicades fan sobre un fet d'actualitat, o bé personatges coneguts dels quals volem obtenir idees sobre la seva activitat professional, social, política o artística, no sobre aspectes de la seva vida íntima. Això passa quan el personatge ens interessa per la seva especialitat, no com a personatge en si. El periodista, en aquesta classe d'entrevistes, es limita a transcriure la conversa mitjançant el sistema de pregunta-resposta, sense comentaris ni interpretacions.

Aquest tipus de entrevista té una subtipologia, la **conversa objectiva**: és un subtipus de la modalitat anterior que consisteix en transcriure una conversa prèviament gravada que es dóna entre dos personatges sobre una mateixa qüestió, als quals se'ls deixa parlar lliurement sense que el reporter hi intervingui.

2) Entrevista literària/ Entrevista perfil: A cavall entre la informació i la interpretació, en aquesta peça traslladem les idees d'un personatge, tot i que passaran pel filtre del propi periodista. L'objectiu d'aquestes peces és donar a conèixer la personalitat de l'entrevistat per mitjà d'un llenguatge més literari. No recorrem a la pregunta – resposta, sinó que reproduïm les declaracions de l'entrevistat entre cometes i alternades amb descripcions i explicacions sobre el personatge, la seva vida i la seva actitud durant l'entrevista. És una peça més atemporal.

5.2.3.4. La crònica

La crònica és un dels subgèneres més difícils de definir perquè comparteix trets amb altres subgèneres. En un intent d'afinar al màxim una descripció de la crònica, el periodista Manuel Graña, ja a mitjans del segle passat, va assenyalar algunes de les principals característiques de la crònica.

- Exclou tot el que sigui simple informació, és a dir, un relat impersonal. El que distingeix la crònica de la mera informació és precisament l'element personal, perquè l'autor es permet la llicència de fer comentaris, ampliar la informació i ordenar els fets a la seva manera. Encara que la crònica sigui informativa, hi ha marge perquè el reporter hi posi el seu estil propi.

- En comparació amb altres textos més especialitzats, la crònica periodística és lleugera, decorada amb recursos literaris però sense molts tecnicismes, breu i ordenada.

- Condensa l'element informatiu i l'element literari o interpretatiu.

- Conté les respostes a les 6W però se'ls atorga un caràcter més retòric. La informació està explicada amb més llibertat estilística i un vocabulari més ric que la informació.

- Admet una gran varietat de fons i forma, de manera que es podrien enumerar molts subtipus de cròniques, com l'artística, la biogràfica, la personal, la descriptiva...

Anys més tard, el catedràtic i periodista Gonzalo Martín Vivaldi va proposar una nova definició de la crònica, segons la qual es tracta "d'una informació interpretativa i valorativa de fets noticiosos, actualts o actualitzats, on es narra quelcom al mateix temps que es jutja el que s'està narrant". Li atribuïa les següents característiques:

- Estil personal però sense desvirtuar els fets ni deformar la realitat.

- Llibertat de recursos estilístics: comparació, metàfora, ironia, anècdota, però sempre amb claredat comunicativa i concisió.

- Format narratiu-informatiu, sense preocupar-se per la piràmide invertida.

- El subjecte-protagonista són els fets noticiosos, així com la interpretació que en fa el cronista.

- El tema és la notícia radiografiada, o el que és el mateix, no només l'anatomia del fet noticiós sinó també la seva psicologia, les seves implicacions.

- Vivaldi hi afegia una modalitat anomenada *croniqueta,* que definia com una peça petita dedicada a quelcom quotidià i aparentment intranscendent.

A dia d'avui, a l'Estat espanyol, s'entén per crònica "la narració temporal d'un esdeveniment" o, en paraules del Doctor en periodisme Manuel Bernal:

> "Una informació de fets noticiosos, transcorreguts en un període de temps, per un cronista que els ha viscut com a testimoni o investigador i fins i tot com a protagonista i que, al mateix temps, els narra, analitza i interpreta mitjançant una explicació personal".

Sense abandonar les característiques senyalades pels altres autors, Bernal hi afegeix nous trets distintius a la crònica:

- Conté certs elements d'informació perquè sempre té una base informativa – uns fets noticiosos que donen peu a escriure una crònica sobre allò que ha succeït – i també del reportatge interpretatiu, perquè aprofundeix i analitza, està escrit amb un estil personal i creatiu però no recarregat, tot i que la implicació de l'autor és encara major perquè participa de l'acció com a testimoni directe i valora els fets a partir d'allò que ha vist i dels coneixements que té. Però la seva funció principal no és opinar sinó informar sobre una qüestió interpretant-la segons les seves impressions.

- Té una relació temporal clara, el que vol dir que el fet narrat ha de tenir una certa continuïtat perquè el cronista explicarà tot el que hagi succeït entre dos moments concrets (al principi i al final dels fets en qüestió).

- Acostuma a tenir una extensió considerable.

- Estructuralment, sovint s'inicia amb un plantejament del que ha succeït seguit d'una interpretació dels fets (explicació, anàlisi i valoració) i, finalment, una conclusió que sintetitzi el resultat d'allò analitzat.

Es poden diferenciar dos tipus de cròniques:

1) Crònica local: El seu autor acostuma a ser un periodista desplaçat permanentment o de manera temporal amb els corresponsals coneixedors de l'actualitat informativa d'altres països o ciutats, o els enviats especials que cobreixen determinats

esdeveniments puntuals, com conflictes bèl·lics, processos judicials, debats parlamentaris, cimeres o grans esdeveniments esportius. Tot i així, avui moltes de les cròniques provinents de l'estranger que es publiquen als diaris locals acostumen a ser més informatives que interpretatives, ja que els reporters es limiten a informar "objectivament" del que passa sense entrar en valoracions que requeririen un coneixement molt més profund i, en conseqüència, molt més temps per a elaborar-les.

2) Crònica temàtica: El propi text informa sobre un fet a la vegada que el valora. L'autor acostuma a ser un periodista especialitzat en cobrir informativament determinats fets relacionats amb l'esport, la tauromàquia, esdeveniments de la societat o culturals, per exemple. Els presencia de principi a fi i això li permet narrar-los i fer arribar al lector les seves impressions del que ha presenciat.

5. Redacció i visualitatzació

5.2. Tipologia de peces periodístiques: els gèneres periodístics

5.2.4. Els gèneres d'opinió

Comentar és una activitat complementària a la de redactar per informar sobre l'actualitat. Els textos d'opinió que es publiquen als mitjans contenen punts de vista respecte als fets d'actualitat que hem conegut a través de la resta del diari. Aquest caràcter de complementarietat podria fer-nos pensar que el comentarista du a terme un rol secundari respecte a l'informador, però no és així. De fet, segons el periodista Héctor Borrat:

> "Comentar un tema és conferir-li un rang més elevat que el d'aquells temes que només són narrats i dedicar-li una editorial és assignar-li el més alt rang".

Quasi tots els mitjans compten amb una secció dedicada a l'opinió. Quan es púbica un text d'opinió a les pàgines dedicades a la informació, es diferencia de la resta col·locant-lo en un requadre o amb una tipologia diferent, un fet que fa saber al lector que està davant un text amb una finalitat diferent a la resta. I si en alguna cosa coincideixen tots els textos opinatius és que els seus autors han de basar les seves opinions en dades correctes i sense manipulacions. Per a fer-ho es requereix un profund coneixement de l'actualitat i un bon domini del llenguatge.

5.2.4.1. L'editorial

L'editorial és un subgènere de l'opinió que té les següents característiques:

- El seu objecte és, sobretot, comentar les notícies però sense prescindir dels elements informatius, que donen peu a l'opinió.
- Generalment no es firma.
- No és l'autor qui opina, afirma o nega, acusa o enalteix, sinó el conjunt del mitjà com a empresa.
- Qui redacta l'editorial guanya en autoritat per la influencia que té en la societat l'opinió del mitjà com a institució.
- Respecte l'estil: El propòsit de l'article ha de ser evident i ben definit.
- A l'articulista se li demana narració ordenada o exposició lògica, molta informació, molta memòria, molta dialèctica i poca emotivitat agressiva.
- L'elocució ha de ser senzilla i sense pretensions d'art literari o purisme exagerat.
- Reflecteix la postura del mitjà respecte a un assumpte.
- Acostuma a escriure'l el director del mitjà.

El que dirigeix l'opinió dels lectors del mitjà no és l'editorial en si, sinó el conjunt de les notícies publicades. El que faria l'editorial seria donar un sentit als fets, que en aquest cas s'exposen amb el propòsit més o menys manifest de portar el lector a una convicció. Existeixen diferents modalitats d'editorials:

5.2.4.2. La columna

La temàtica que es tracta en les columnes i el seu to té a veure amb el sentit i l'estil personal del columnista: a vegades és una anàlisi profund, d'altres, un material satíric, irònic o humorístic i d'altres, un simple comentari personal. La columna té una periodicitat i lloc fixes en els mitjans i el seu autor és un col·laborador fixe que opina sobre qüestions d'actualitat, dotant-les d'un caràcter personal. A vegades l'autor és una persona coneguda de qui sovint interessa més la seva figura que la seva opinió. Hi ha dues tipologies de columnes:

- La columna d'actualitat: el columnista opina sobre qüestions de l'actualitat més immediata.
- La columna personal: l'autor fa servir un estil molt més personal per opinar sobre temes més lleugers i no necessàriament vinculats a l'actualitat informativa.

5.2.4.3. La crítica

És una anàlisi valoratiu de la producció artística (en el camp del teatre, la música, l'art, la literatura, el cine, etc,) propi de la secció cultural d'un mitjà i està escrit per persones amb un alt grau d'especialització. Les seves principals característiques són:

- La funció del crític és apreciar la relació entre el propòsit de l'autor de l'obra i els resultats obtinguts, amb la finalitat d'ajudar al lector a entendre l'obra i orientar-lo.

- L'autor acostuma a ser especialista en la matèria que ha de valorar. El seu comentari ha de ser fruit d'una anàlisi objectiva, detallada i fonamentada.

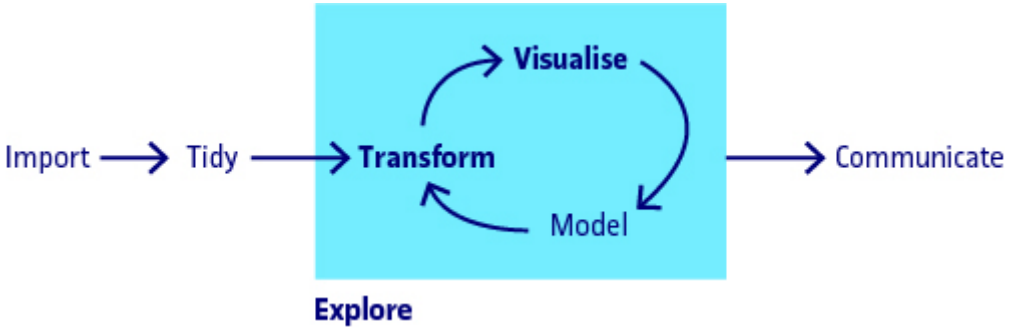
- L'autor, que també ha estat lector o espectador de l'obra, acostuma a començar la seva crítica informant el lector sobre qui és l'autor, quines altres obres ha creat i el que suposa la seva nova obra. A continuació descriu l'obra i finalment, a partir dels seus coneixements com a expert, emet unes valoracions raonades.

6. Tres exercicis de periodisme de dades

Introducció

Aquesta sèrie de tres exercicis posa en pràctica allò que s'ha après en aquesta assignatura i altres assignatures del màster (obtenció, anàlisi i visualització de dades i redacció). Troba les dades i, mitjançant la seva anàlisi i visualització, crea el fil argumental que les expliqui.

El Consell de Redacció del mitjà pel qual treballes ha acceptat la proposta de reportatge/recerca que vas fer fa unes setmanes. Ara cal provar amb dades les hipòtesis que vas plantejar i explicar-les de forma atractiva i intel·ligible en una història periodística.



Font: Garrett Grolemund i Hadley Wickham.

Per a cadascun dels encàrrecs, hauràs de:

1. Aconseguir les dades.
2. Aplicar mecanismes de selecció i filtratge de les dades (neteja i ordena, genera noves variables).
3. Analitzar – visualitzar (i tornar a analitzar, procés reiteratiu) fins a trobar les visualitzacions i els arguments que expliquin les dades.
4. Trobar el fil argumental a partir de l'anàlisi de les dades.

Aquest és l'esquema dels passos que cal seguir per desenvolupar qualsevol projecte d'anàlisi de dades. Com s'indica en el gràfic, el procés no és lineal. En aquesta sèrie d'exercicis es mostra una aproximació a cadascun dels passos per treballar les diferents fases d'un projecte de dades mitjançant la pràctica. Els casos inclouen l'obtenció de dades de diferents fonts (API, fonts oficials i no oficials, PDF) i el seu tractament i visualització mitjançant tot tipus de programes (anàlisis de xarxes, cartografies, visualitzacions interactives). Per facilitar l'accés a aquests exercicis, s'han seleccionat eines de programari obert o lliure (Gephi, Python, D3) i que funcionen en totes les plataformes (Linux, Windows i Mac).

Si decideixes desenvolupar codi per a algun dels projectes, et recomanem que utilitzis un sistema de control de versions (per exemple, Git) que et permeti revisar l'evolució del teu treball i col·laborar més fàcilment amb altres persones. Documenta amb rigor totes les manipulacions que facis a les dades per després poder revisar la teva feina.

Nota

En els següents exercicis se suggereixen una sèrie d'eines a tall d'exemple. Si et resulta més senzill o vols explorar altres camins, pots utilitzar-ne d'altres. Les preguntes que es formulen també són orientatives; pots formular-ne d'altres que et semblin més interessants i intentar resoldre-les.

6. Tres exercicis de periodisme de dades

6.1. Twitter: captura de dades i anàlisi de xarxes

6.1.1 Introducció

Et truquen de la redacció, has de fer una peça ràpida sobre una cosa de Twitter. Et diuen que «les xarxes estan cremant» i et sona a acudit, però no et queda més remei que capbussar-te en un tema del moment (en anglès, *trending topic*) i produir alguna cosa que valgui la pena llegir: Qui hi ha darrere d'aquest tema del moment? Com va sorgir? Potser fins i tot aprens alguna cosa pel camí.

Nota

Aquest exercici està basat en els exemples, el codi i les presentacions de Mariluz Congosto (blog: http://barriblog.com, Twitter: @congosto).

Coses que pots aprendre: capturar piulades i perfils a Twitter per mitjà de la seva API, analitzar les dades capturades mitjançant grafs amb el programa Gephi.

6. Tres exercicis de periodisme de dades

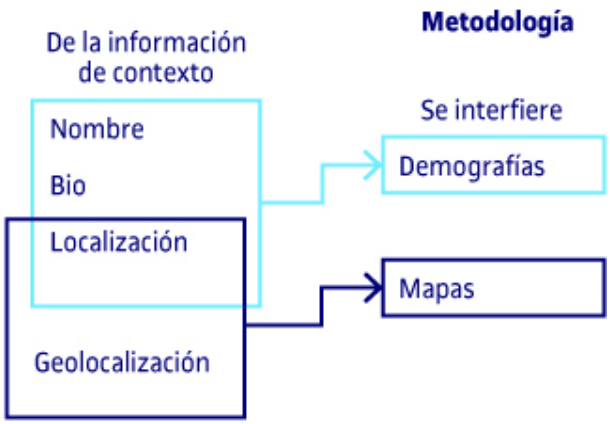
6.1. Twitter: captura de dades i anàlisi de xarxes

6.1.2. Context

Les xarxes socials són un lloc important dins l'ecosistema de mitjans, on es distribueixen (i de vegades es creen) les notícies. D'entre elles, Twitter s'ha configurat com la plataforma de *microblogging* més popular. S'ha convertit en un lloc on informar-se (i de vegades desinformar-se). L'obertura de l'API de Twitter ha permès un accés directe a les fonts d'informació que aquest tipus de plataformes acumula (l'altre cas és Facebook, que és molt més tancada a l'hora de permetre l'accés a la seva informació). Una API és una forma estructurada d'accedir a la informació que conté una biblioteca. Analistes, acadèmics, activistes i periodistes la usen per accedir a les dades que, si no fos per l'API, haurien de comprar o raspar (en anglès, *scrap*). Hi ha moltíssima literatura sobre aquest tema i multitud d'articles acadèmics des de la sociologia, la politologia, les ciències de la comunicació i moltes altres disciplines que utilitzen Twitter com a cas d'estudi per la seva facilitat a l'hora d'obtenir les dades. Però deixem aquestes discussions acadèmiques de banda:

La teva cap t'ha encarregat una peça curta sobre un dels temes del moment del dia a Twitter. Qui està publicant piulades sobre el tema? Pots fer una història amb això més enllà de seleccionar les piulades més repiulades? Multitud de notícies parlen de les ingerències a Twitter per part de Rússia o Veneçuela. Com fan les anàlisis? Com capturen les dades?

D'acord, hem dit que ara Twitter ha facilitat la descàrrega de piulades, però per on hem de començar? Per a aquest exercici, hem pensat a utilitzar T-hoarder, un programari lliure que funciona en Python i que està pensat per ajudar-te a interactuar amb l'API d'una manera senzilla, sense necessitat de tenir coneixements de programació. Amb T-hoarder podràs obtenir informació de Twitter a través de la seva API (que en realitat conté tres API: REST, *search* i *streaming*) i processar la informació obtinguda per fer visualitzacions. Segons el tipus de dades que obtinguem, podrem generar un tipus de visualitzacions o d'altres. Amb els usuaris, amb els seus seguidors i amb les seves repiulades (RT) i esments, podem obtenir relacions entre usuaris per fer gràfics de xarxes (grafs). D'altra banda, si analitzem la localització de les piulades podem fer mapes.

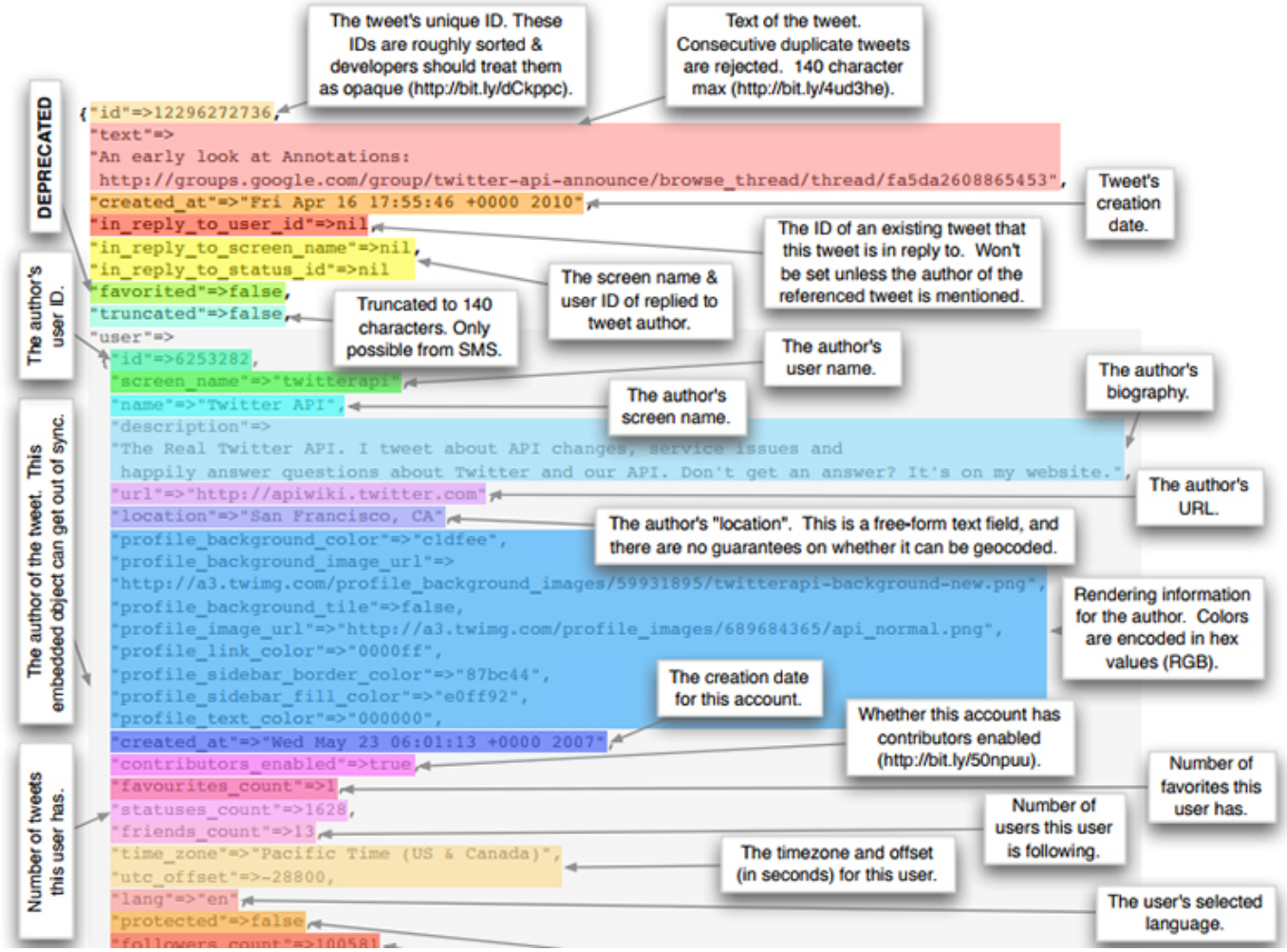


6. Tres exercicis de periodisme de dades

6.1. Twitter: captura de dades i anàlisi de xarxes

6.1.3. Conceptes bàsics o previs

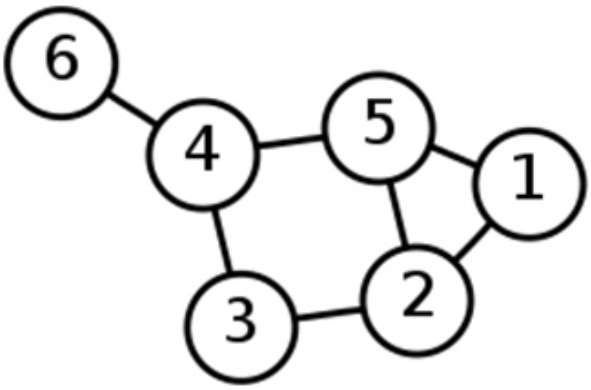
Anatomia d'una piulada. Una piulada (en anglès, *tweet*) és un missatge que conté molta informació associada (metadades) que no veiem des del web o des d'una aplicació.

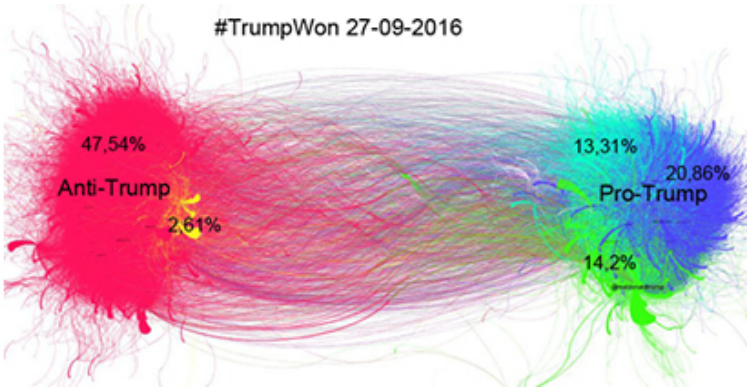


Font: Raffi Kriokorian.

Què és una API? Una API (acrònim d'*application programming interface*) és la forma estructurada d'accedir i interactuar amb una aplicació o web.

Què és un graf? Un graf és un conjunt d'objectes relacionats. Els objectes es solen anomenar «nodes» o «vèrtexs», mentre que la relació entre dos objectes s'anomena «aresta». Se solen representar com a diagrames de punts o cercles connectats per línies. Vegem un parell de grafs, un de senzill (a l'esquerra) i una xarxa tan densa que no s'hi representen els nodes, sinó que només s'hi representen les arestes:





Wikipedia (esquerra); Mariluz Congosto (dreta).

Què és T-hoarder? És un programari lliure desenvolupat en Python per Mariluz Congosto que permet interactuar d'una forma senzilla amb l'API de Twitter per obtenir informació de forma automatitzada i processar-la per facilitar la seva anàlisi. Genera arxius .gdf, que serveixen per fer grafs amb el programa Gephi. Si tens problemes per executar Python en el teu ordinador, T-hoarder inclou una màquina virtual que permet treballar-hi sense necessitat d'instal·lar res més. T-hoarder disposa de dues versions: T-hoarder i T-hoarder_kit.

- t-hoarder_kit (*) serveix per analitzar HT o coses que han succeït. Té més eines per analitzar usuaris, generar grafs i processar informació.

- T-hoarder serveix per monitoritzar coses que succeiran, com ara unes eleccions, una convocatòria, un esdeveniment, etc. És més adient per emmagatzemar informació de llarga durada.

Què és una máquina virtual. Una màquina virtual és un programari que simula un sistema de computació i pot executar programes com si fos una computadora real.

Què és Python? Ja ho saps, però t'ho recordem: és un llenguatge de programació que pren el nom dels humoristes britànics Monty Python.

Què és Gephi. És un programari de codi obert escrit en Java que permet analitzar i visualitzar xarxes.

6. Tres exercicis de periodisme de dades

6.1. Twitter: captura de dades i anàlisi de xarxes

6.1.4. Preguntes a resoldre

6.1.4.1. Qui hi ha darrere d'un *trending topic*?

Podries dir si és dirigit o espontani? Quines comunitats el recolzen? Hi ha xarxes de *bots*?

a. Quina anàlisi o visualització respon la pregunta?

A l'hora d'analitzar una gran quantitat de piulades, un graf (un conjunt de nodes i arestes) és la manera més eficaç i senzilla per entendre com es configura una xarxa. En aquest tipus de gràfics, els usuaris es representen com a nodes i les repiulades (la republicació d'un missatge d'un altre usuari) es representen com les arestes que relacionen dos nodes. Si l'usuari A repiula un missatge de l'usuari B, hi haurà una línia (aresta) que anirà del punt A al punt B. Les repiulades són el mètode més usat, però també es poden utilitzar-se les mencions o les respostes. Una altra manera d'analitzar les dades és representar les relacions dels usuaris que han usat aquesta etiqueta segons si es segueixen entre ells.

b. Quines dades necessito i on les puc trobar?

Per poder analitzar les piulades d'una etiqueta en concret, en primer lloc cal obtenir-les. El mètode més senzill per capturar aquestes piulades és usar l'API de Twitter (Tot allò que sempre va voler saber de l'API de Twitter i mai es va atrevir a preguntar). Per fer aquest exercici hem usat l'etiqueta #xxxxx, i durant xxxx hores hem capturat xxx piulades. Existeixen altres opcions per analitzar una etiqueta en directe, com ara l'eina Flocker, que descarrega les piulades en directe i dibuixa el graf que es formen.

```
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://
```

Piulada capturada per T-hoarder en un arxiu de text pla. Metadades separades per tabuladors

Per facilitar el procés de captura amb l'API utilitzarem T-hoarder, el programari lliure que ha desenvolupat Mariluz Congosto. Es tracta d'una aplicació lleugera que corre en Python i que permet interactuar amb l'API de Twitter d'una manera senzilla per obtenir piulades i altres informacions (perfils d'usuaris, seguidors, etc.) i guardar-les en formats fàcilment reutilitzables (arxius de text pla). T-hoarder es pot utilitzar de dues maneres:

- Usant el codi disponible a Github, executant Python directament a la consola. Accedeix al manual d'instal·lació per Windows i Linux/Unix per més detalls.

- Mitjançant una màquina virtual que permet accedir a un entorn en el que es pot usar T-hoarder i Gephi. És necessari usar el programari Virtual Box per fer funcionar la màquina virtual. Descarrega la màquina virtual (3,9GB) que ve preinstalada amb T-hoarder_kit i Geph: https://www.dropbox.com/s/18gt69suptie5hw/ubuntumate1602_taller.ova (usuari: taller; contrasenya: tallerdatos).

Escriptori de la màquina virtual

A la màquina virtual només cal fer doble clic en la icona de T-hoarder per arrencar el programa.

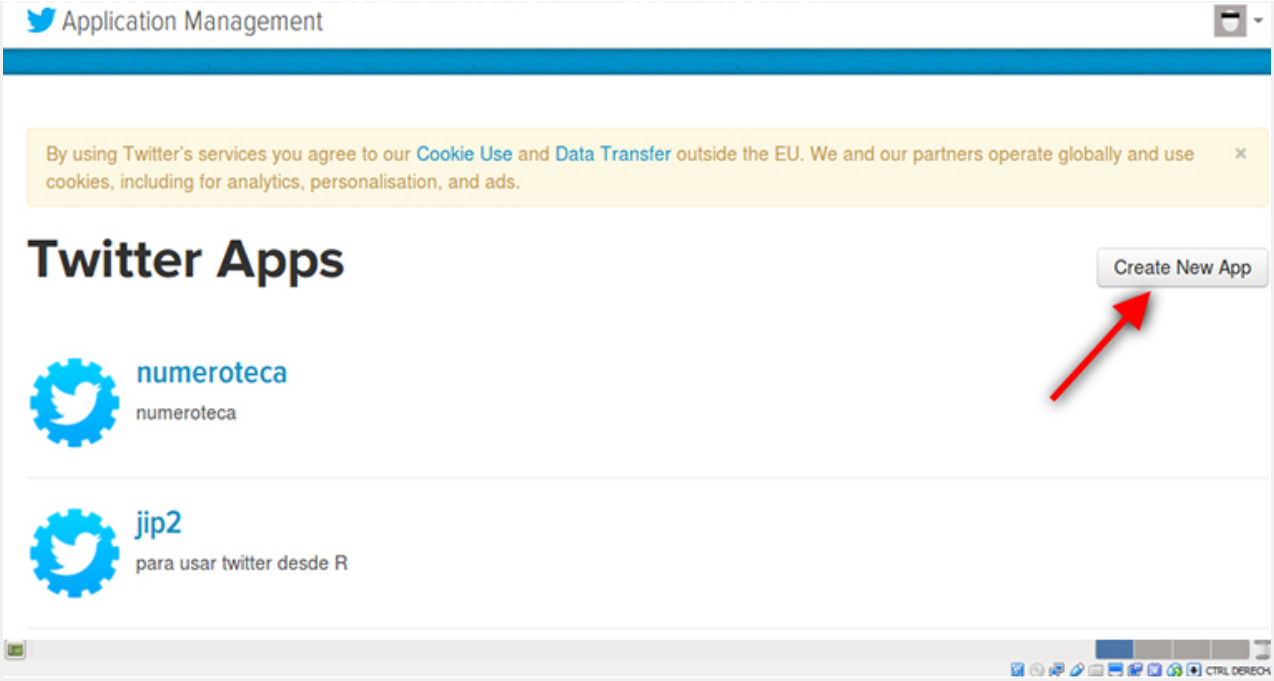
Si executes directament t-hoarder_kit en la teva consola, ves al directori `/scripts/` on hagis baixat t-hoarder_kit i executa `./t_hoarder_kit.sh`.

En ambdós casos et demanarà on és el fitxer *application keys*. Vegem com generar-lo:

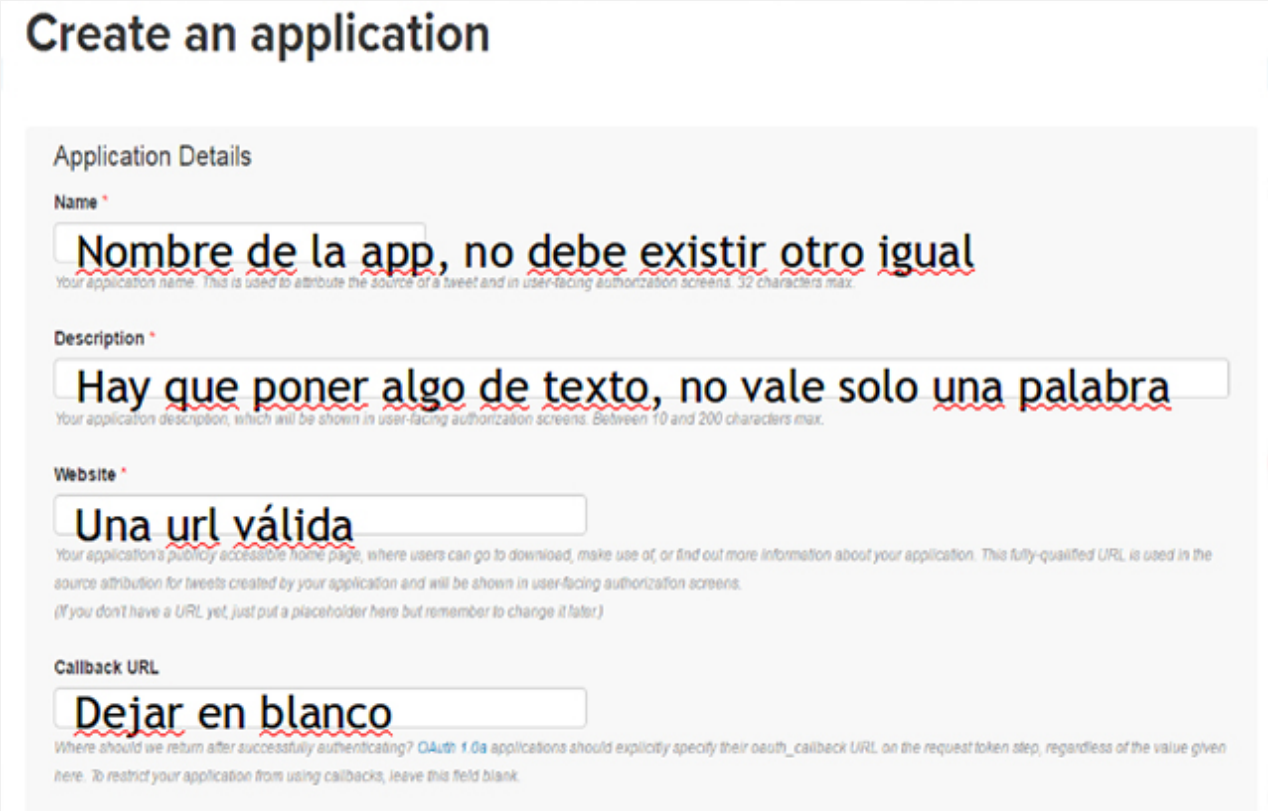
```
numeroteca@cinco ~/sites/t-hoarder_kit/scripts (master *=) $ ./t_hoarder_kit.sh
----------------------------------------
------> Welcome to t-hoarder kit <------
----------------------------------------
t-hoarder_kit is up to date
working in /home/numeroteca/sites/t-hoarder_kit

----------------------------------------
------>    Environment  data     <------
----------------------------------------
Enter the file name with the application keys: forthoarder.key
Enter a twitter user: numeroteca
Enter experiment name: pruebas
```

Generar Keys en Twitter Apps

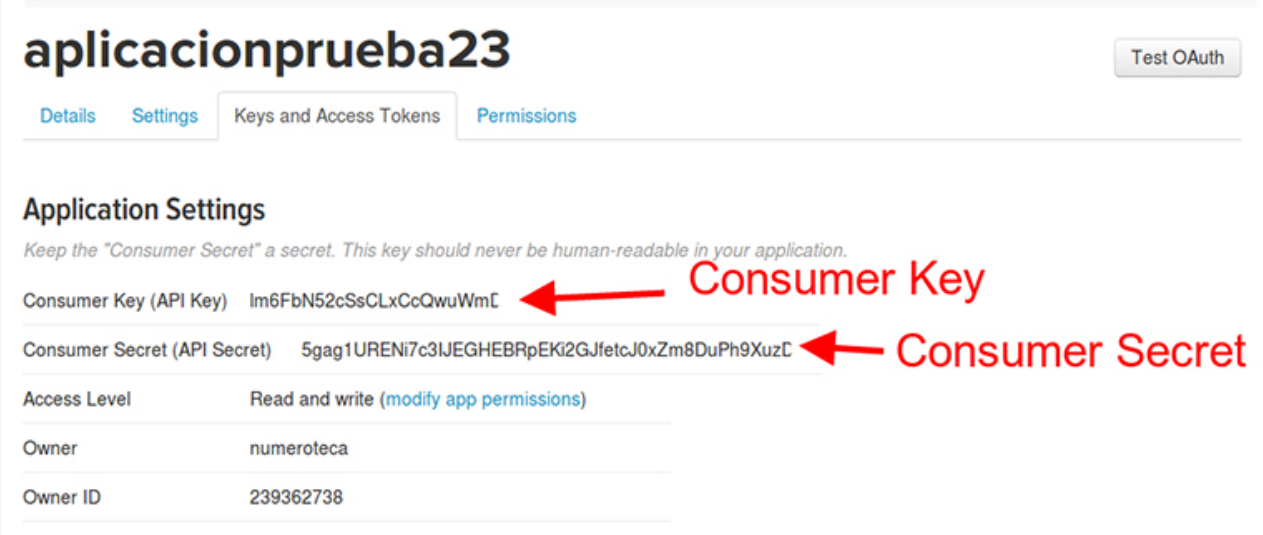
Cal crear una aplicació que ens permeti accedir a l'API de Twitter. Ves a Twitter Apps per crear una nova aplicació: https://apps.twitter.com. Per poder accedir-hi cal crear un usuari de Twitter.



Omple les dades de l'aplicació segons s'indica a la imatge:



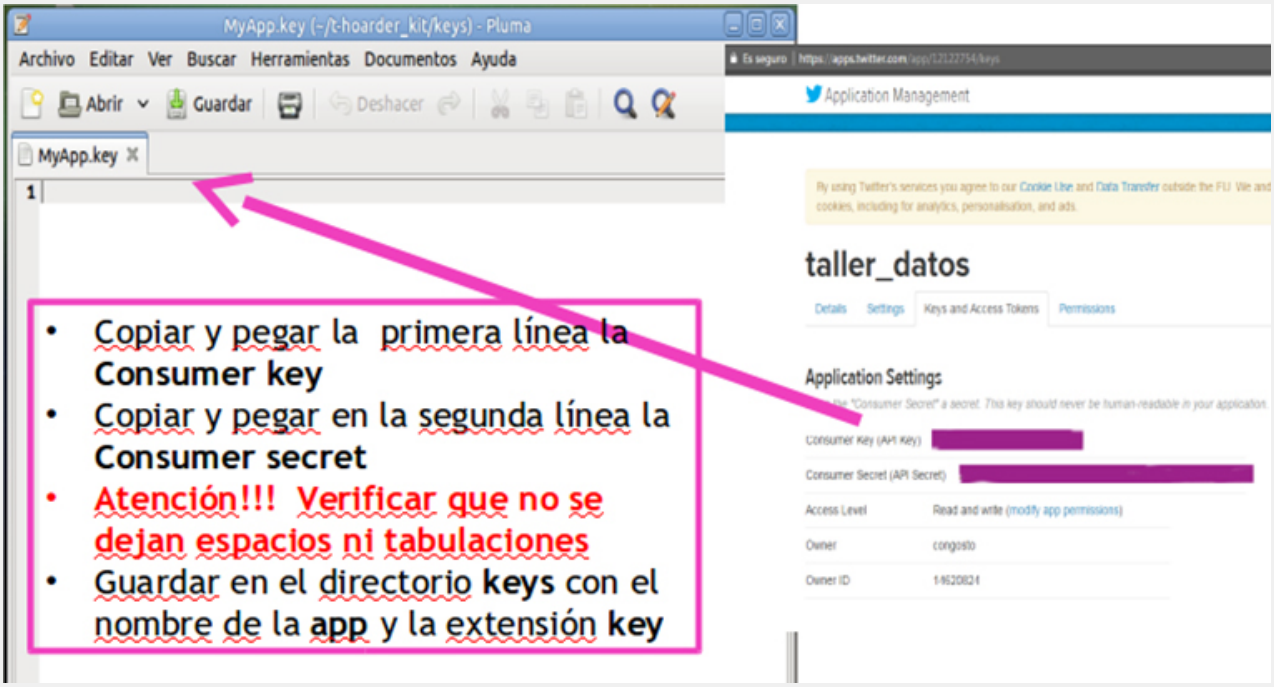
A la pestanya *Keys and Access Tokens*, copia la *Consumer Key* i el *Consumer Secret*:



Crea un arxiu amb el nom `forthoarder.key` que inclogui la *Consumer Key* (*API Key*) en la primera línia i el *Consumer Secret* (*API Secret*) en la segona. El resultat serà semblant a això:

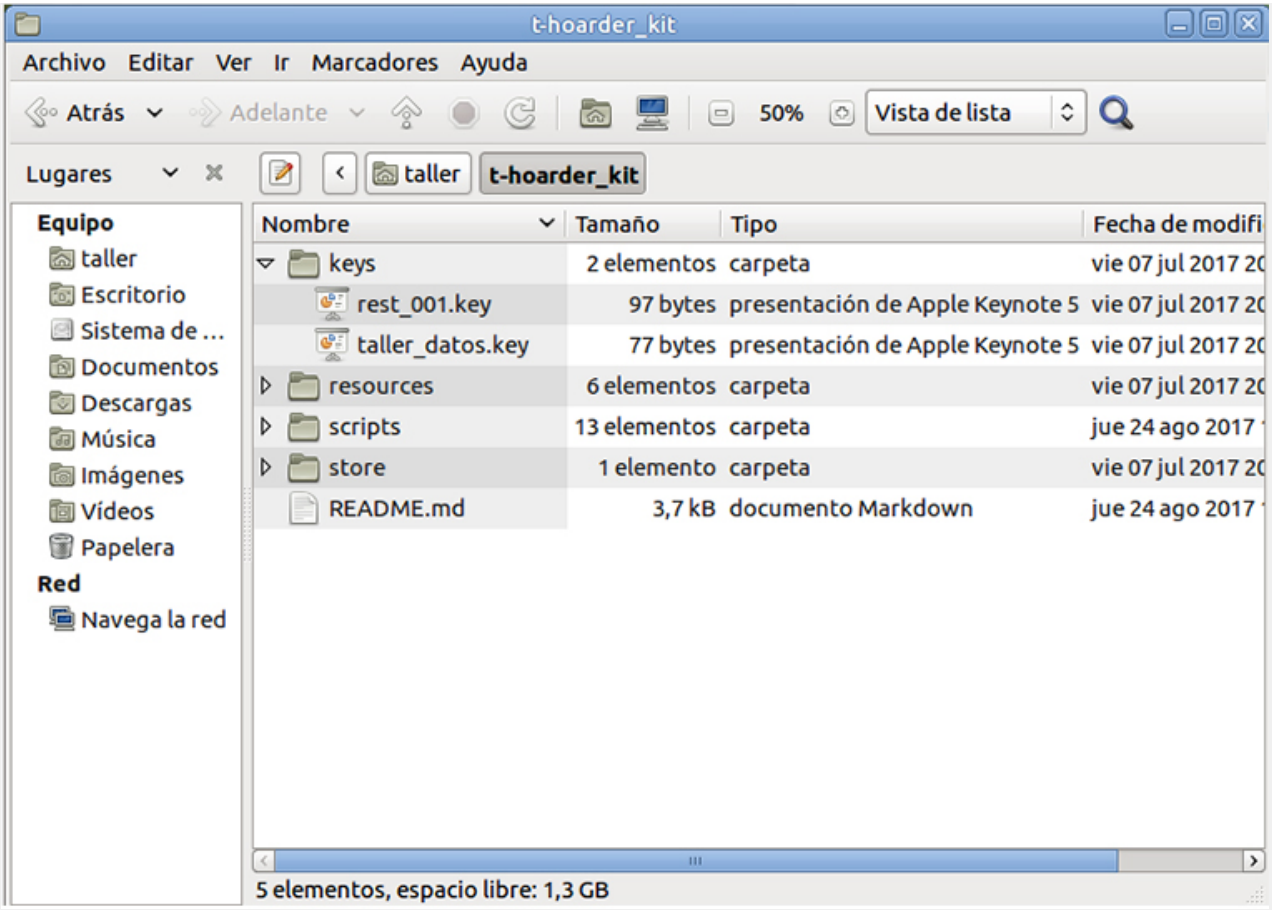
```
W960clipfwefVxEdg
jKt9GVQ3BHOq0Jp4hHIOlB6npoD8n5feL89ruE
```

I guarda'l amb el nom `forthoarder.key` en el directori `/keys/` de la teva instal·lació de T-hoarder. Utilitza una aplicació d'edició de text pla (Notepad, Gedit, etc.).



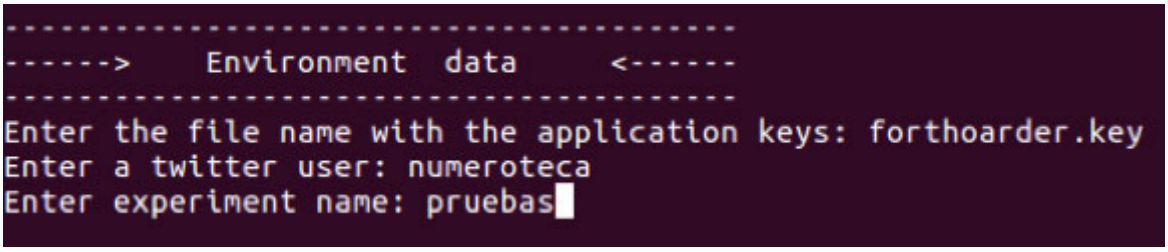
Ja ho tens tot llest per al següent pas.

L'estructura de carpetes hauria de quedar així:



En aquest cas, les claus creades són `rest_001.key` i `taller_datos.key`.

A la consola, l'aplicació et farà preguntes i hauràs d'introduir-hi les les dades. Les dades que demana a l'inici són per al context del treball: es necessita una *app*, un usuari Twitter i un directori de treball prèviament creat per deixar-hi els resultats.



1. Enter the file name with the application keys: `forthoarder.key`.
2. Enter a twitter user: `mi_user`.
3. Enter experiment name: `pruebas`.

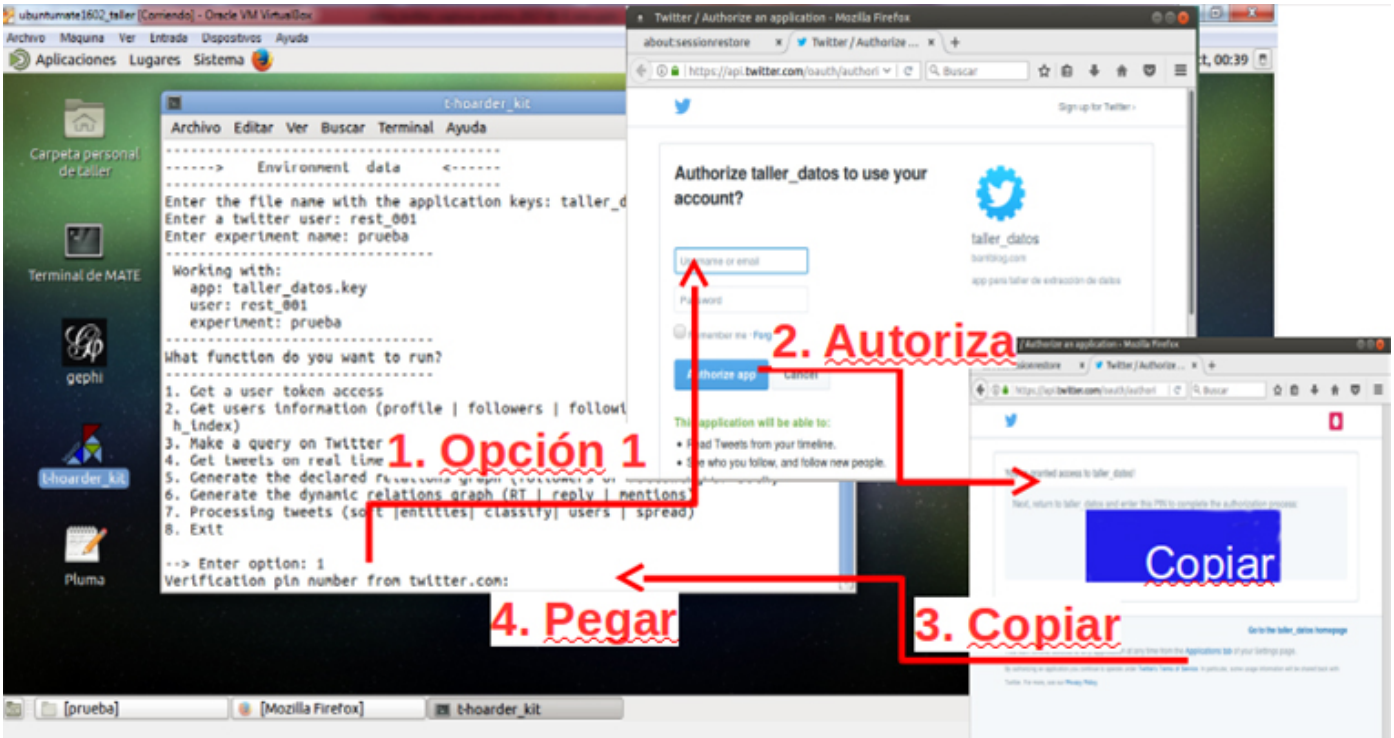
Si estàs utilitzant la màquina virtual, posa com a usuari `mi_user`.

On `forthoarder.key` és l'arxiu amb les nostres claus, `numeroteca` és l'usuari amb el que has creat l'aplicació, i `taller` és el directori que has de crear a la carpeta `/store/`. Pots tenir tants experiments com vulguis, cadascun en el seu directori corresponent.

Si tot ha anat bé, hauríem de veure el menú següent:

```
--------------------------------
 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
```

Seleccionem l'opció 1 per generar el permís del nostre T-hoarder_kit des de la nostra màquina. Això obrirà el navegador que tinguem assignat per defecte. Introduïm el nostre usuari i la nostra contrasenya, i obtindrem un nombre de 7 xifres que hem de copiar i enganxar a l'aplicació:



Si tot ha anat bé, a la consola hi veurem alguna cosa semblant a això:

```
Verification pin number from twitter.com: 6347239
Access token generado con éxito. Guardado en keys/numeroteca.key
```

I s'haurà generat un arxiu `numeroteca.key` en el nostre directori `/keys`.

Ja està tot llest per començar a capturar dades.

> **Nota**
>
> T-hoarder no captura tota la informació de les piulades (veure documentació, anatomia d'una piulada) sinó «id, user, ... [completar]» i l'emmagatzema en arxius de text pla (.txt). Altres programaris emmagatzemen la informació en bases

de dades (Postgres, SQL...), la qual cosa dificulta el seu tractament i la capacitat de compartir-la de forma ràpida i senzilla.

Tria l'opció `3. Make a query on` Twitter per descarregar piulades d'una cerca. Per exemple, si volem descarregar retroactivament l'etiqueta #parlament (abans, crea l'arxiu buit `parlament-tweets-query.txt` a la carpeta de l'experiment):

```
--> 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/proves/parlament-tweets-query.txt
>>>>/home/numeroteca/sites/t-hoarder_kit/store/proves//parlament-tweets-query.txt file
exists, do you want to append more tweets (y/n)?y
```

Si tot va bé, veuràs alguna cosa semblant a això, l'*script* que informa que les piulades s'estan capturant bé:

```
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 es vol parar la captura de les piulades, prémer `ctrl + c`.

Si obrim el fitxer `store/proves/parlament-tweets-query.txt` podrem veure les piulades capturats.

Una altra opció és descarregar les piulades en temps real amb la *streaming* API. Per fer-ho, usa l'opció `4. Get tweets on real` time:

```
--> Enter option: 4
Enter input file name with the keywords separated by , : palabras_clave.txt
```

Per fer la cerca de piulades cal un fitxer de text pla amb les paraules clau. Per exemple, si vols buscar l'etiqueta #parlament, el fitxer només contindrà #parlament. Si vols fer cerques d'altres paraules, separa-les amb comes. Prova-ho a Twitter sur bús

c. Quines eines he d'utilitzar per preparar les dades?

Quan ja haguem obtingut les piulades, necessitarem convertir-les a un format interpretable per Gephi, un programari per generar grafs que és lliure i està molt estès. Una vegada més, T-hoarder ens facilita convertir una sèrie de piulades en un arxiu .gdf.

Abans de processar les dades podem netejar-les i filtrar-les de falsos positius. Per exemple, si estem capturant piulades sobre el Parlament de Catalunya, podrem filtrar les piulades que no facin referència a aquest tema. Amb eines com ara R o Vi es poden eliminar, per exemple, totes les piulades que continguin una paraula en concret.

Què es pot fer amb la llista de piulades?

A partir d'una llista de piulades es poden extreure moltes anàlisis: els usuaris més repiulats, el nombre de piulades per minut, hora o dia, etc. A partir de les piulades es poden obtenir els ID dels usuaris, i amb els usuaris es poden calcular les seves relacions (*follower-following*). Compte: per obtenir els *following* de cada usuari, l'API et dona els *following* de 60 usuaris cada hora!

Segons com es processin les piulades, es poden obtenir dos tipus de grafs:

- De relacions declarades (*following* i *followers* entre els usuaris de l'etiqueta).
- De relacions dinàmiques (els que es repiulen).

Ambdues són interessants a l'hora d'analitzar qui hi ha darrere d'una etiqueta, però per realitzar el primer graf (relacions declarades) és necessari obtenir noves dades: a qui segueixen els usuaris que usen aquesta etiqueta? [Veure la pregunta següent].

Vegem pas a pas com generar el .gdf amb T-hoarder:_kit. Selecciona l'opció `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): parlament-
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 introduït a `Introduce top` size és per «podar» el graf. Genera un subgraf amb els *n* nodes més actius i rellevants. Això és útil quan el graf és molt gran i no hi cap a Gephi (és massa gran i l'ordinador no pot manejar-lo). En aquests casos podem generar un graf més petit que no és el mateix però és representatiu del graf original.

Si has analitzat una llista de piulades i les relacions establertes són els seus RT (repiulades), obtindràs dos fitxers:

- Els *topRT* : `xxxxxxxx_top_RT.gdf`. És el graf podat amb els *n* usuaris (en el nostre cas, 100) més repiulats.
- Tots els RT: `xxxxxxxx_all_RT.gdf`.

Està tot llest per usar aquests fitxers a Gephi.

d. Quines eines es poden utilitzar per produir l'anàlisi o la visualització?

Quan tinguem l'arxiu .gdf podrem fer el penúltim pas: crear el graf.

Com fer un graf amb Gephi pas a pas

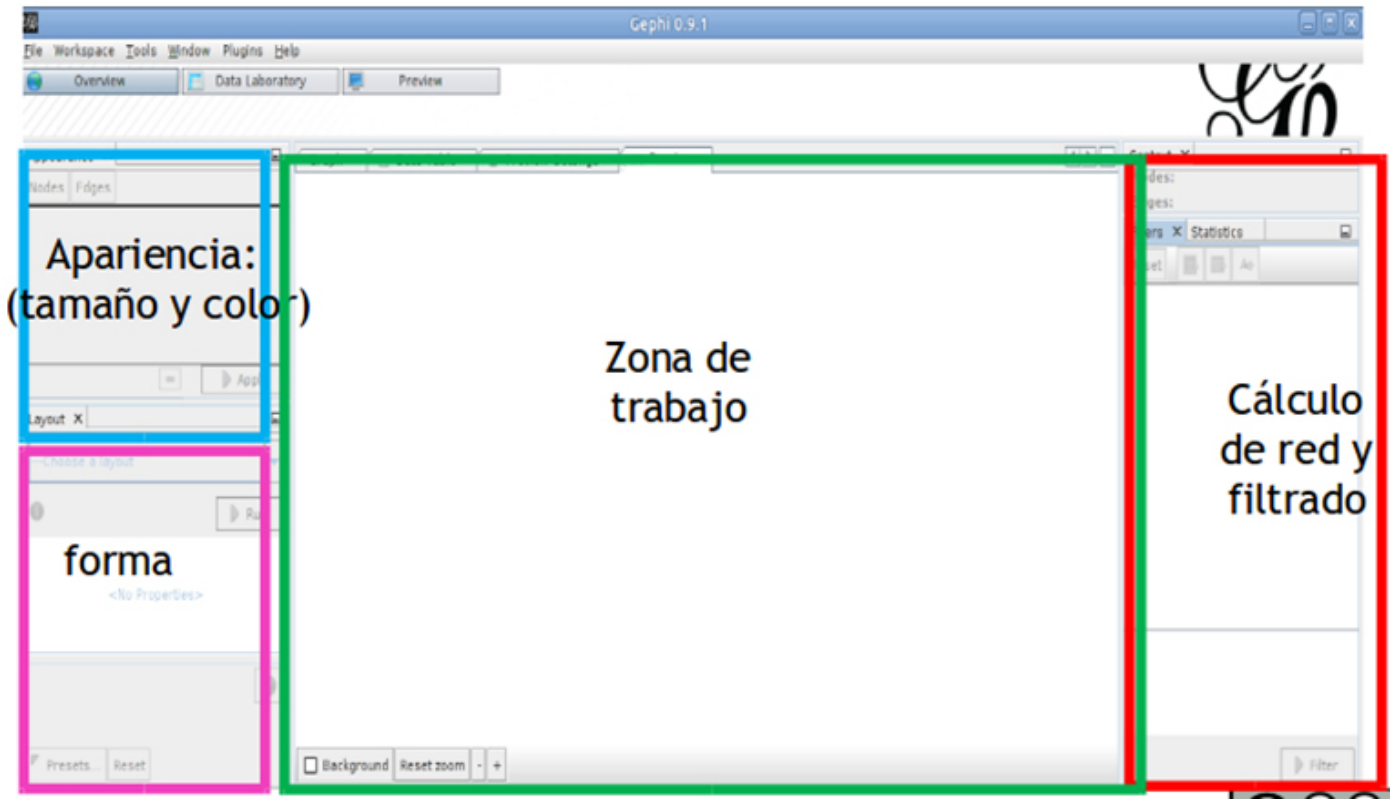
0) Instal·la Gephi i obre'l

Descarrega https://gephi.org/users/download/.

És un programa que funciona amb Java. Més que instal·lar-lo, cal executar-lo cada vegada. Des de la línia de comandaments, ves al directori `/bin` que has descarregat i executa `./gephi`. El més segur és que per fer-ho hagis de donar permisos d'execució a l'arxiu.

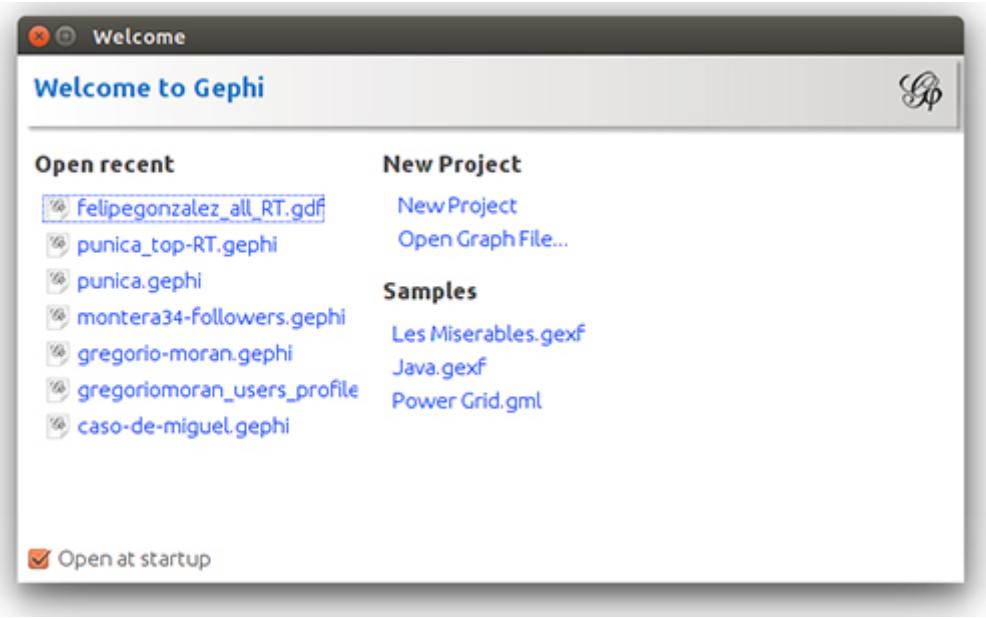
```
numeroteca@cinco ~/Downloads/gephi-0.9.2/bin $ ./gephi
```

Si fas servir la màquina virtual, simplement fes clic en la icona de Gephi que hi ha a l'escriptori.

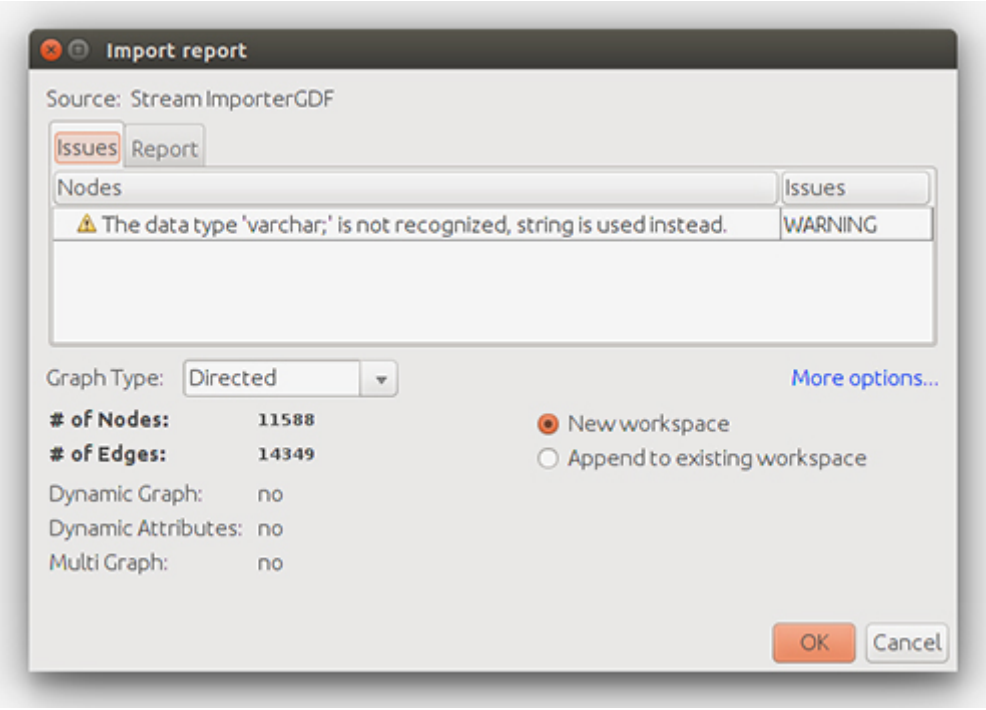


1) Genera un graf

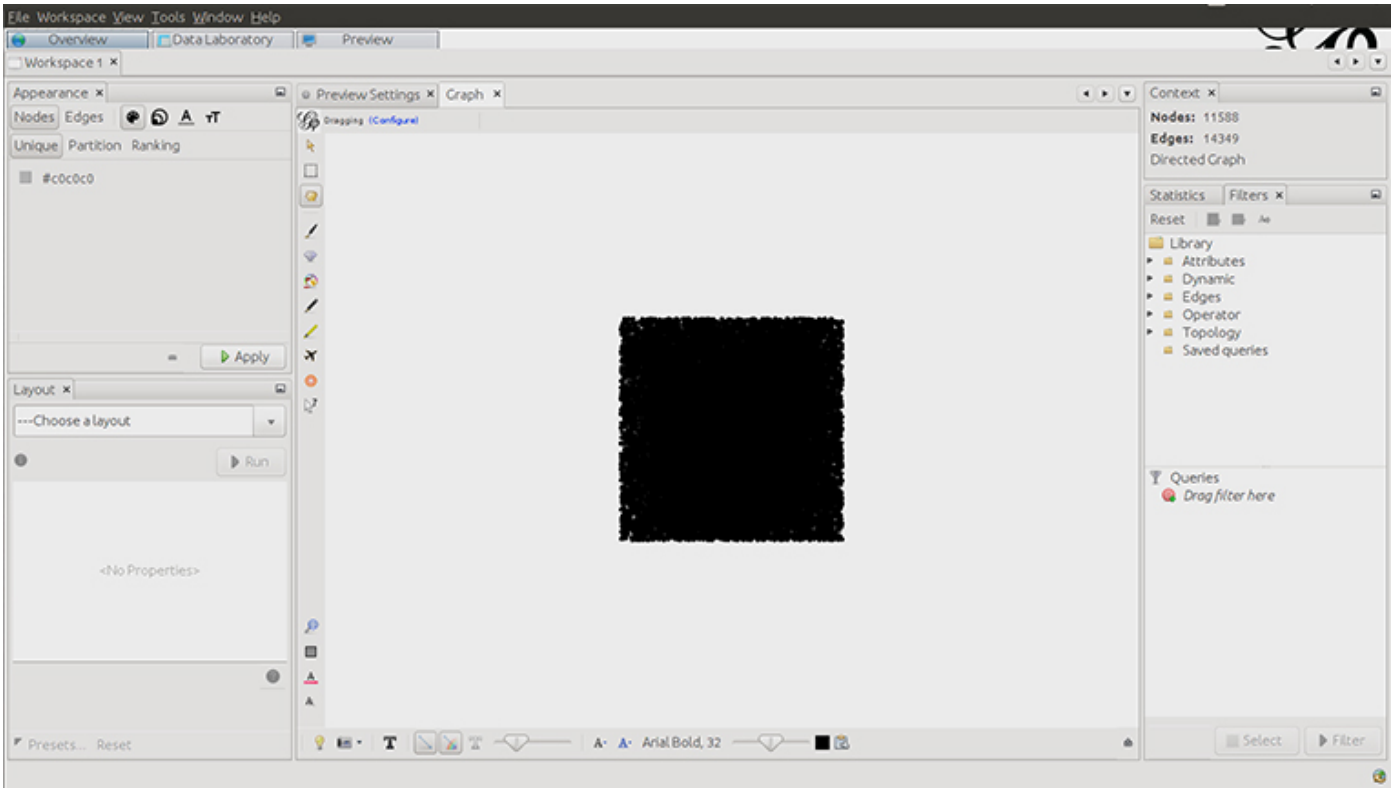
Obre l'arxiu xxxxxxxx_all_RT.gdf que has generat en el pas anterior.



Et preguntarà si el graf és dirigit, no dirigit o mixt (*graph style*). Escollirem dirigit, atès que els RT tenen adreça: estableixen una aresta (fletxa) del node de l'usuari que repiula a l'autor original de la piulada.

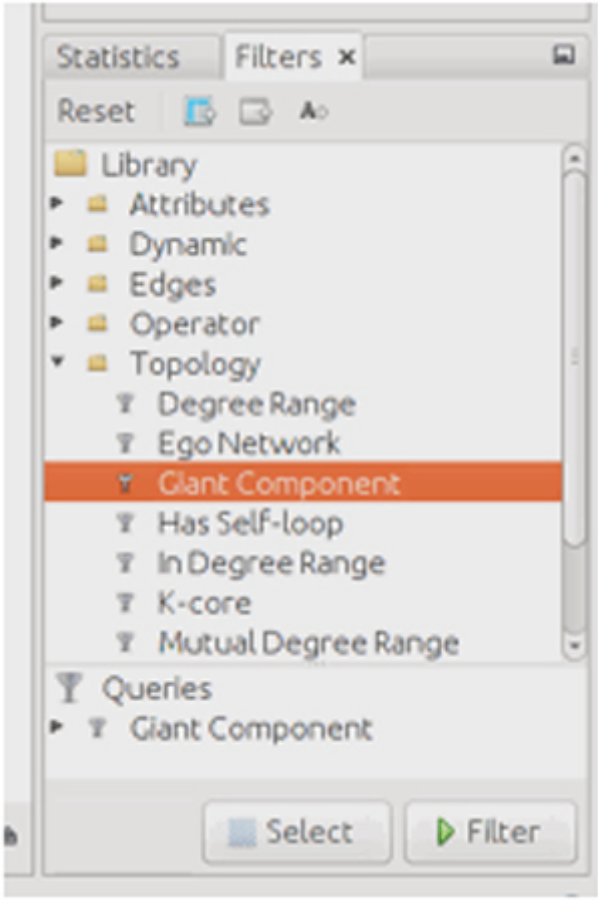


Això importa tots els nodes a l'espai de treball.



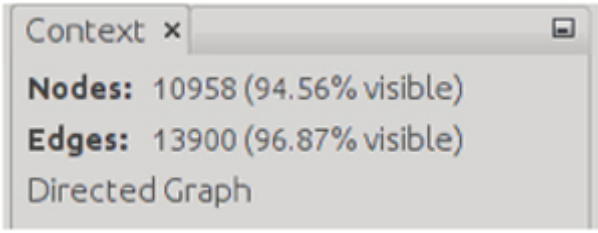
Els nodes creats recentment sempre tenen aquesta forma quadrada.

A la secció *Statistics* de la dreta, en el directori `Library -> Topology`, seleccionem amb doble clic *Giant component*.

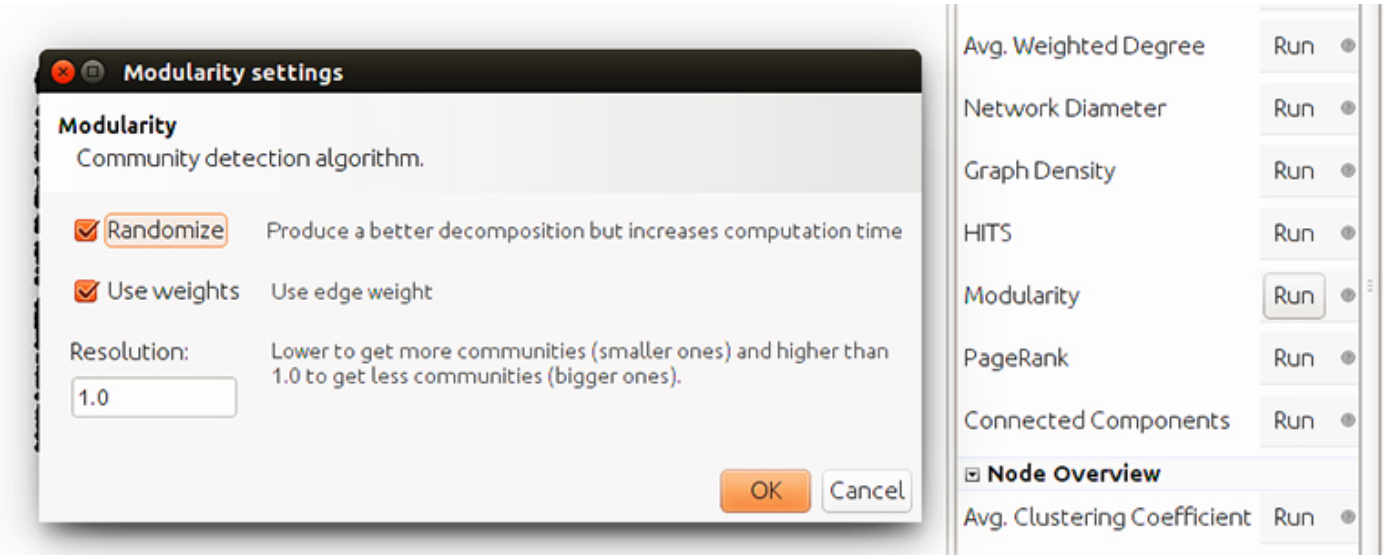


Ara el filtre *Giant component* és a la llista de *Queries*. Si el seleccionem i fem clic a *Filter*, apliquem aquest filtre a les nostres dades.

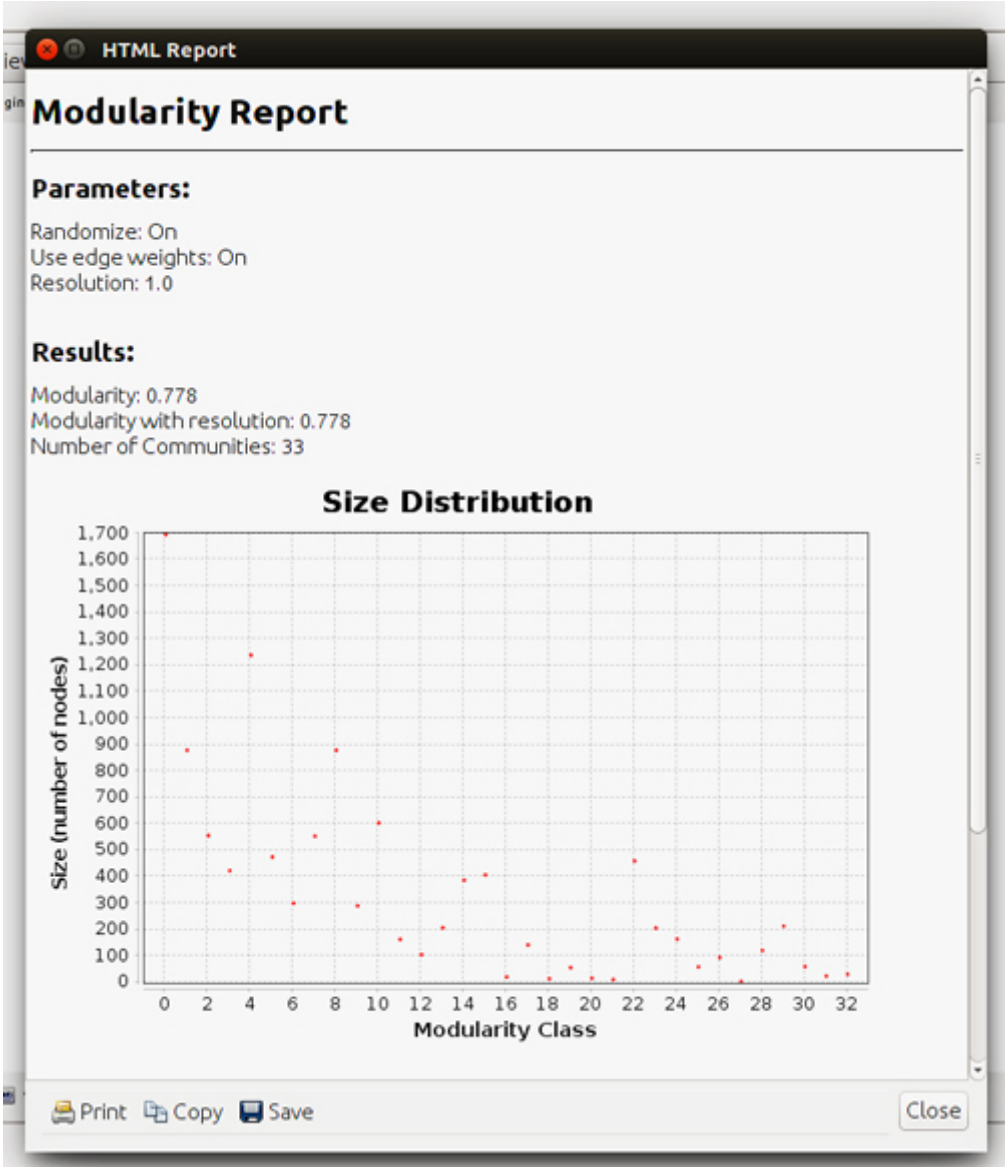
Vegeu que a dalt a la dreta hem reduït el nombre de nodes i arestes seleccionades:



Ara passem a la pestanya *Statistics* i executem el càlcul de *Modularity* (modularitat de la xarxa) fent clic a *Run*:



Obtenim el resultat següent:


HTML Report
Modularity Report
Parameters:
Randomize: On
Use edge weights: On
Resolution: 1.0
Results:
Modularity: 0.778
Modularity with resolution: 0.778
Number of Communities: 33
Size Distribution
Size (number of nodes)
1,700
1,600
1,500
1,400
1,300
1,200
1,100
1,000
900
800
700
600
500
400
300
200
100
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32
Modularity Class
Print
Copy
Save
Close

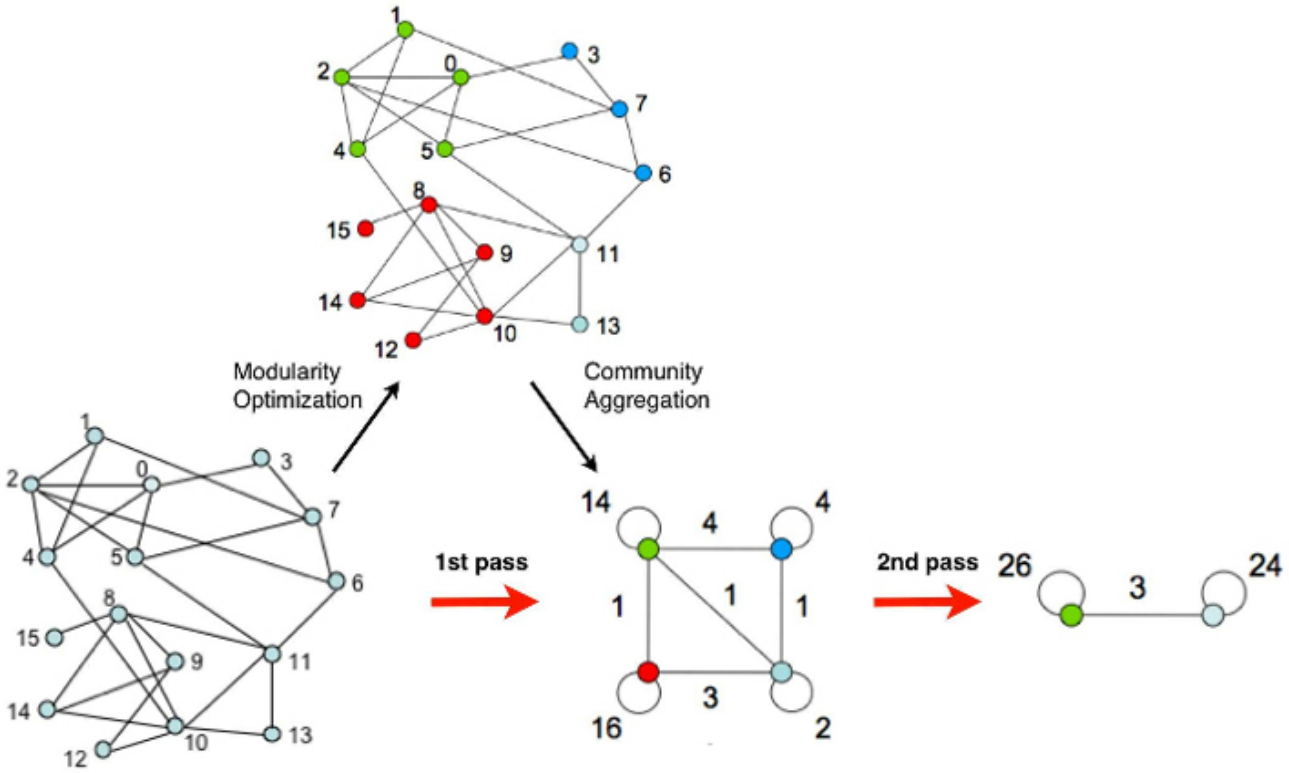
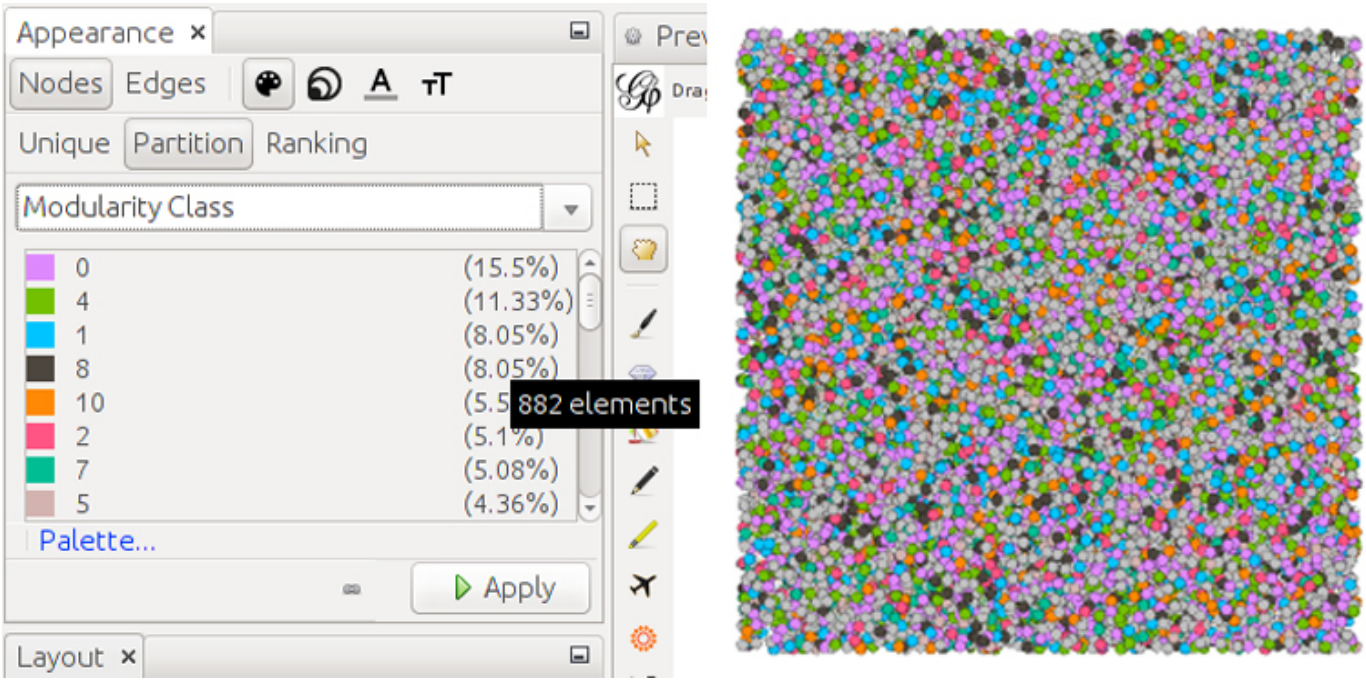


Figure 1. Visualization of the steps of our algorithm. Each pass is made of two phases: one where modularity is optimized by allowing only local changes of communities; one where the found communities are aggregated in order to build a new network of communities. The passes are repeated iteratively until no increase of modularity is possible.

Font: 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.

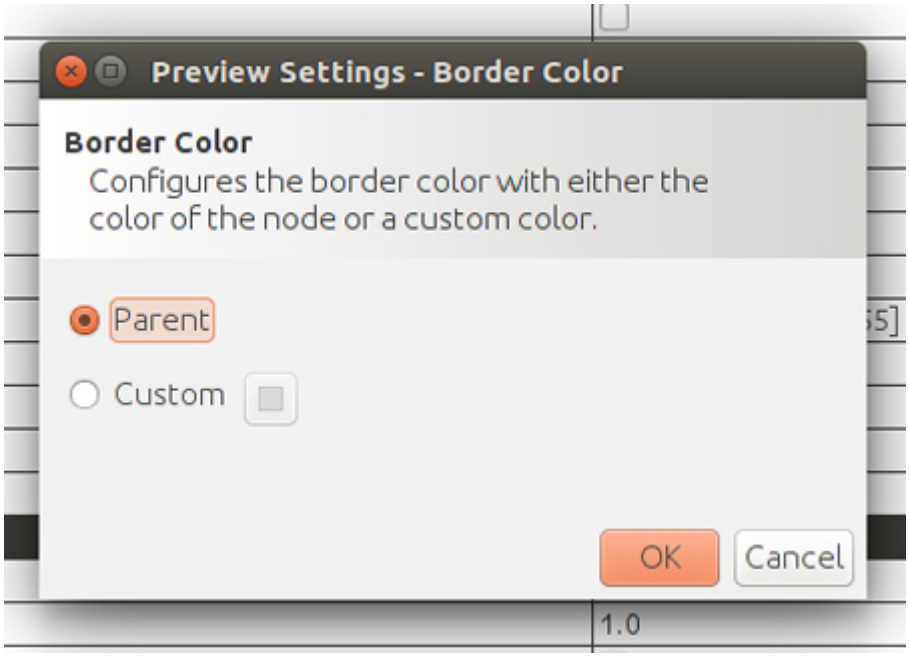
Ara hem de seleccionar com es pinten els nodes. Triem acolorir-los segons la modularitat que acabem de calcular (si anem a la manera de visualització de dades veurem que s'ha afegit una columna amb la modularitat). A la pestanya `Appearance -> Nodes [paleta pintor] -> Partition`, selecciona `Modularity Class`.



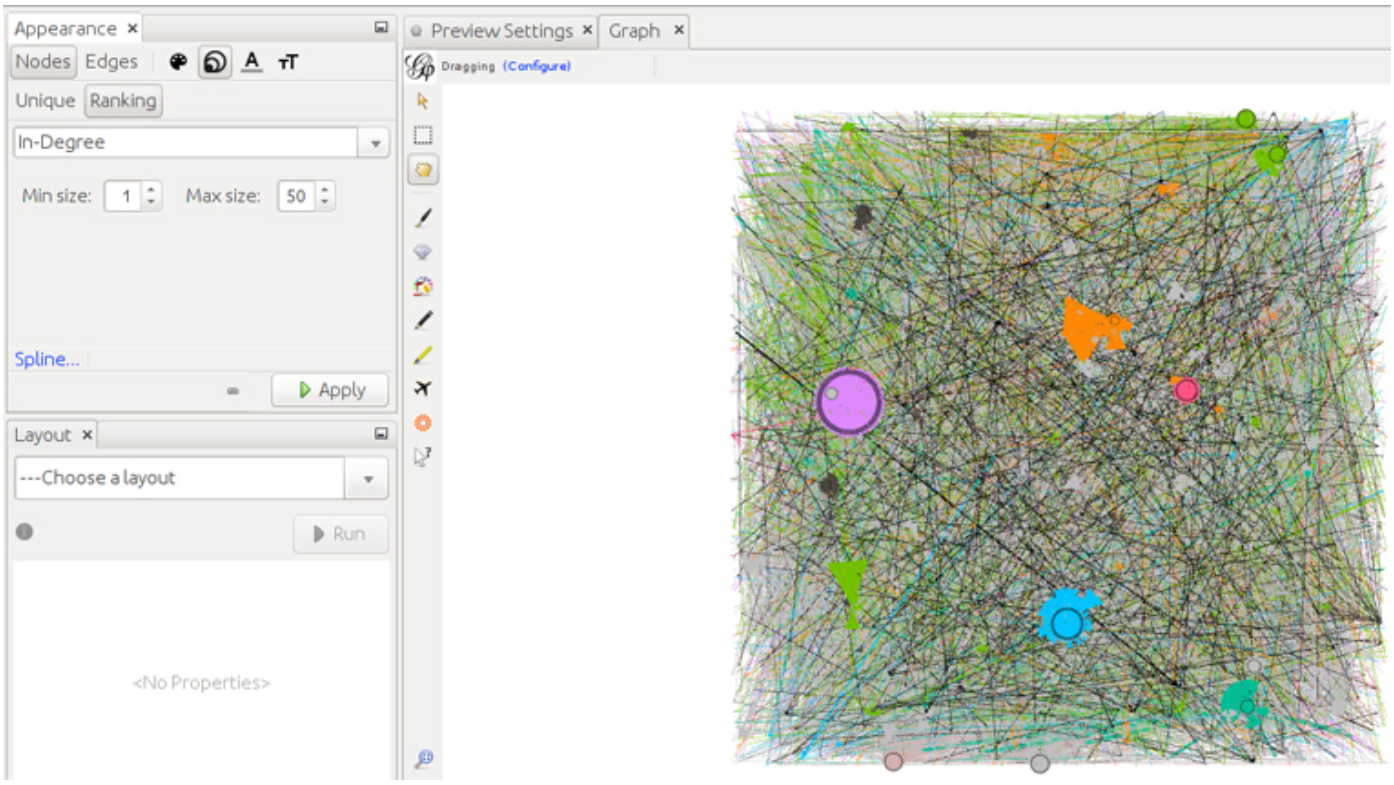
Pots canviar el color de cada categoria prement en el color de la llegenda.

Ara els nodes estan acolorits en funció de la modularitat calculada.

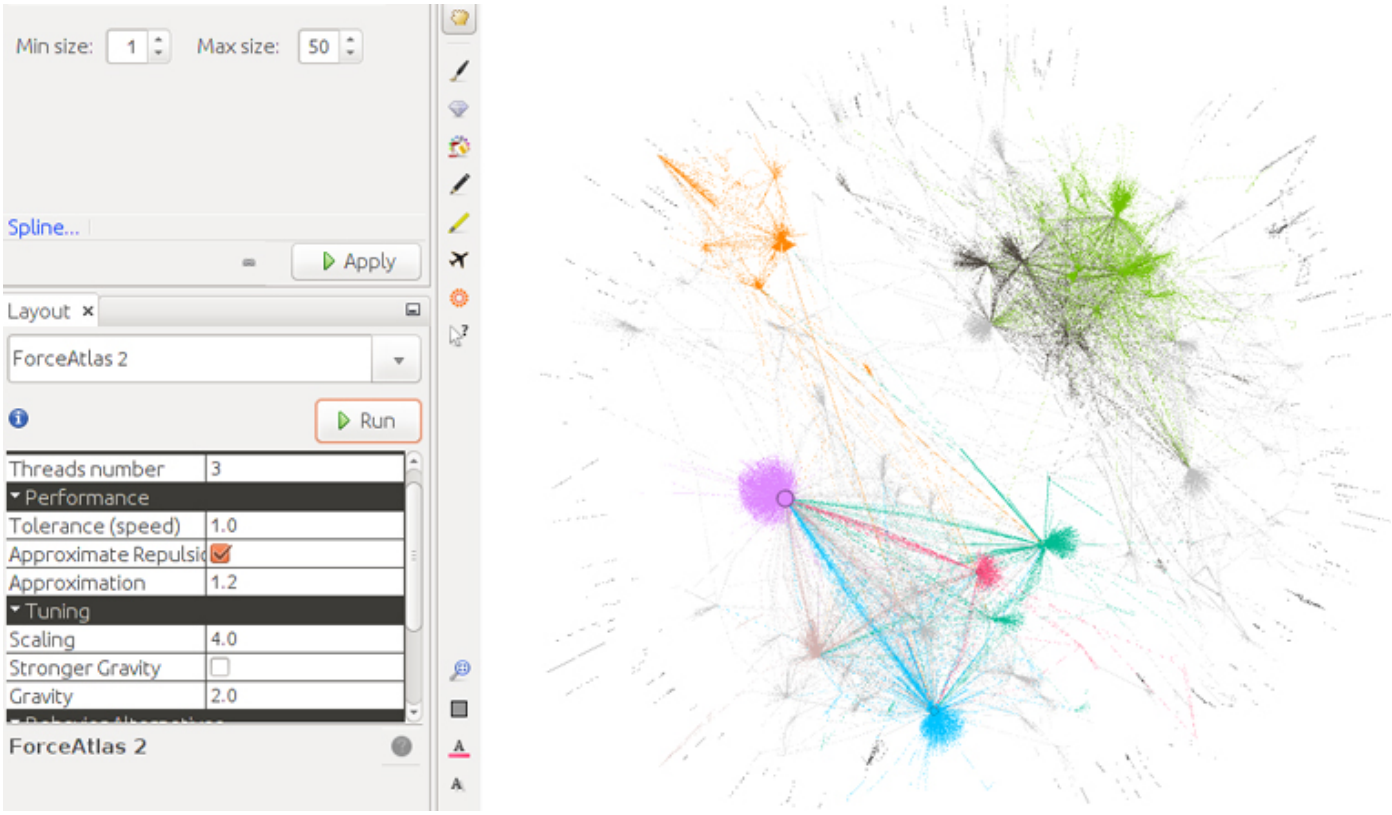
A `Preview Settings ->Settings`, selecciona l'opció *Parent* a *Border color* dels nodes:



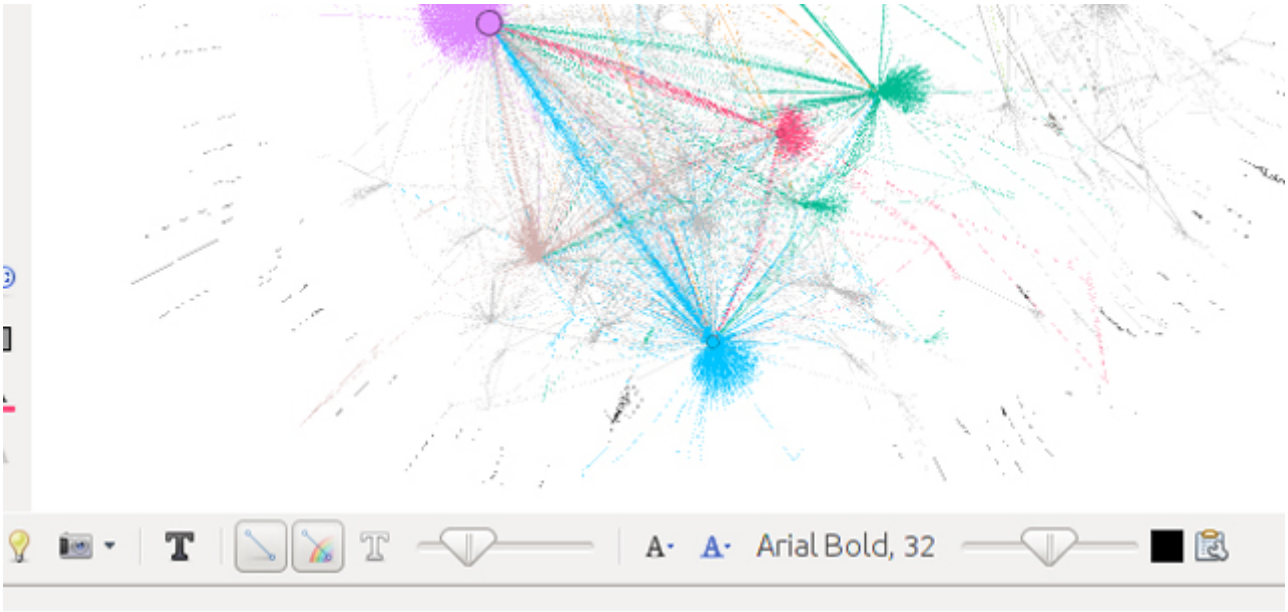
Ara modifiquem la grandària dels nodes segons del seu *In-degree*, que és el nombre d'arestes que reben, això és, el nombre de RT que han rebut les seves piulades. Per fer-ho seleccionem `Apearance -> Nodes -> [tamaño] -> Ranking` i seleccionem `In-Degree`.

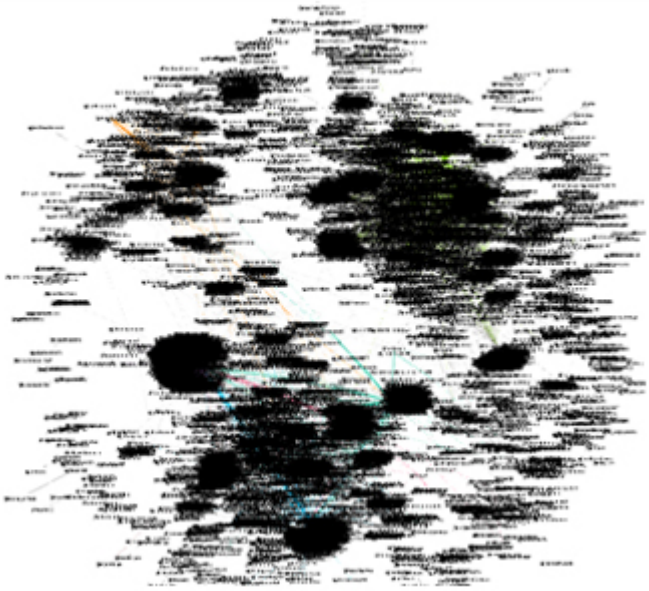


Ara, a *Layout*, seleccionem *Forceatlas 2* per tornar a ubicar la posició dels nodes i aconseguir que s'agrupin en relació amb les comunitats detectades de modularitat.

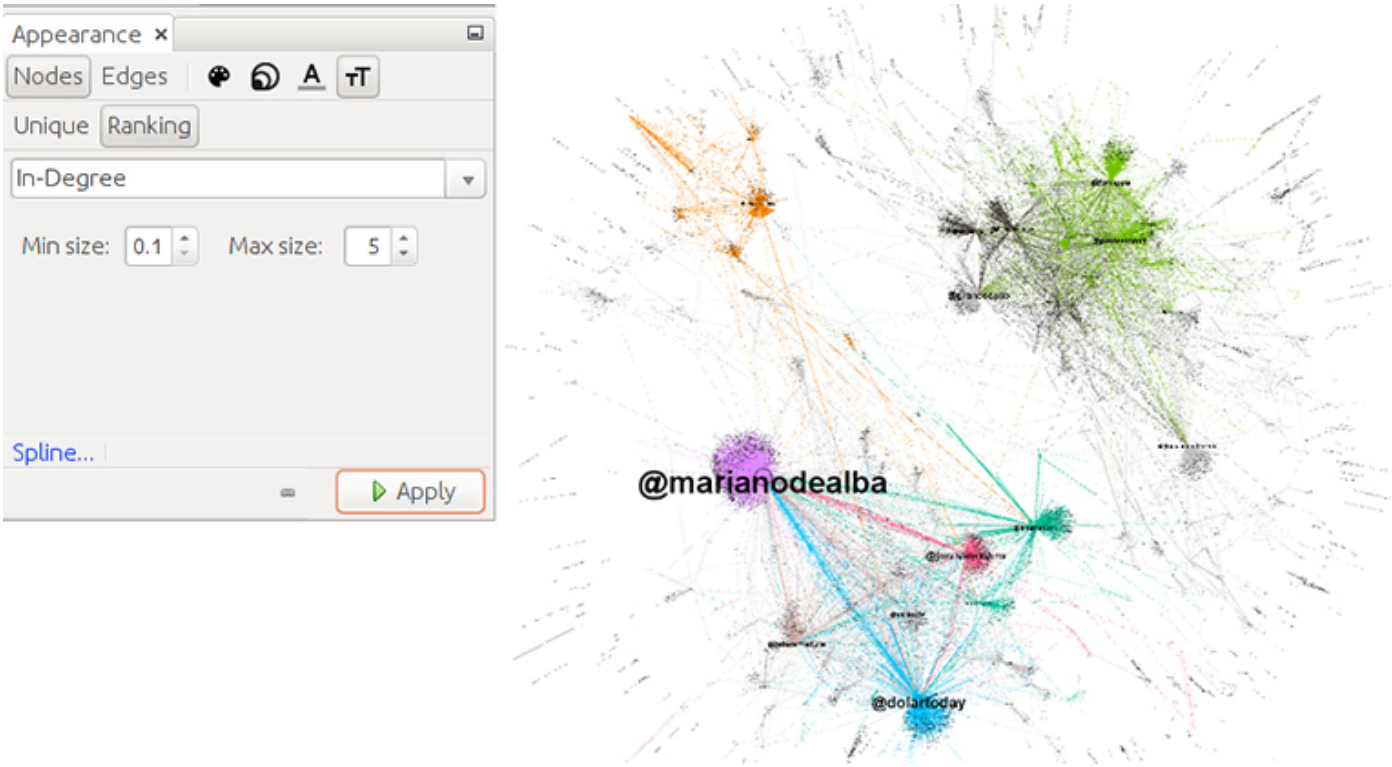


La *T* negra ens permet activar les etiquetes (*labels*) dels nodes.

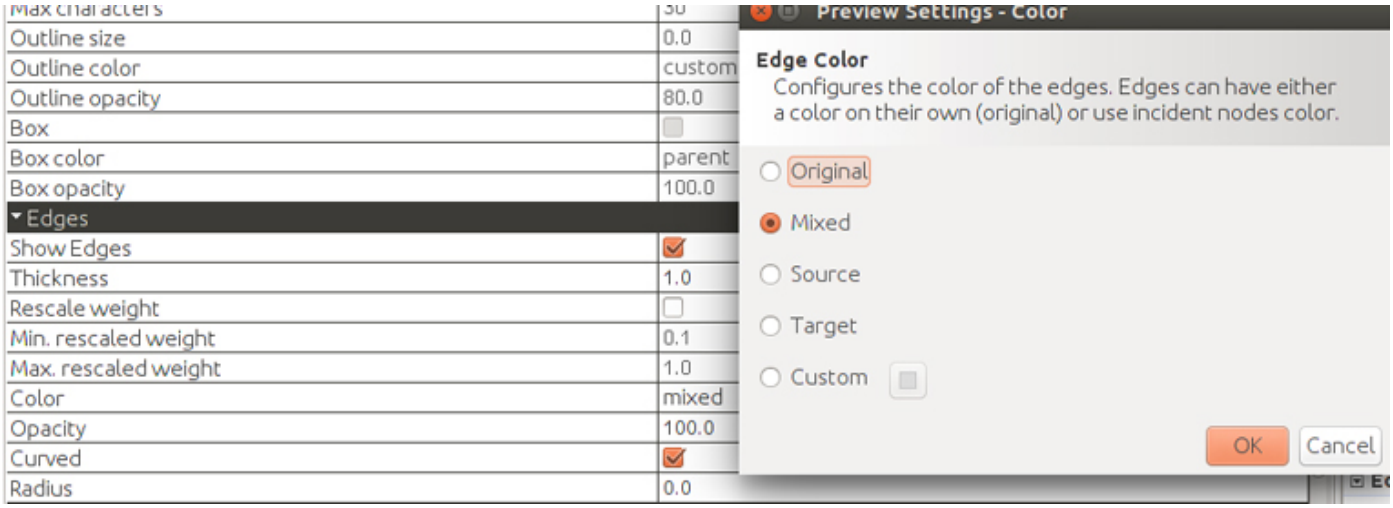




L'opció de grandària de text d'`Appearance -> Nodes -> [tamaño letras] -> Ranking` ens permet seleccionar la grandària mínima i màxima de les etiquetes dels nodes en funció, una vegada més, de la variable *In-degree*. Selecciona 1 de mínim i 50 de màxim. A *Preview settings*, selecciona una tipografia de 5 punts de grandària.



Podem controlar el color i el gruix de les arestes des de *Preview settings*. Triem *Source* per veure d'on vénen els RT, és a dir, el color de l'aresta serà el de l'usuari que repiula.

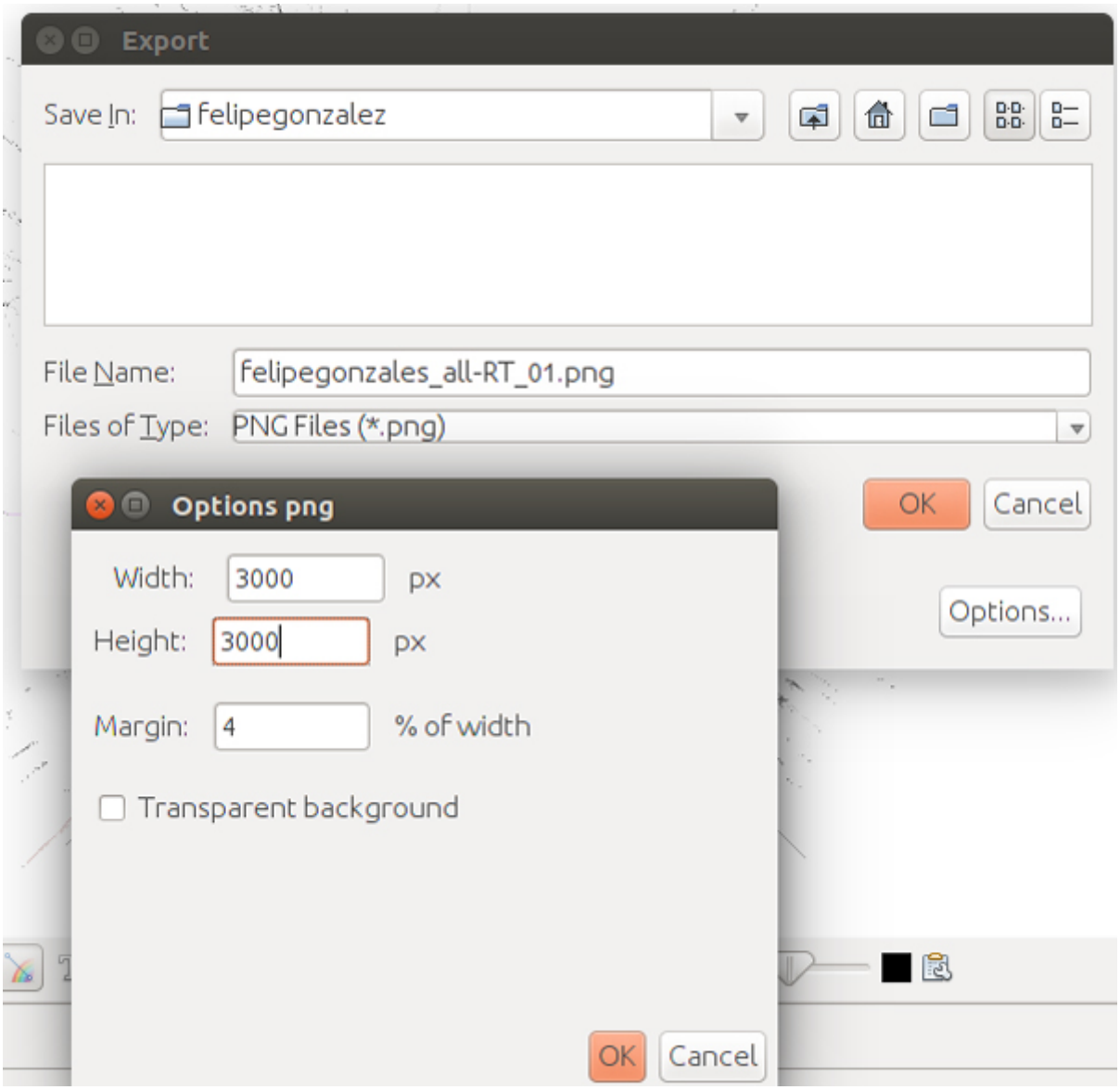


Aquestes són les opcions:

- 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.

De vegades, quan hi ha molts RT d'un node a un altre, Gephi no pot manejar-ho bé i pinta triangles o cercles. Això es pot resoldre marcant «reescalar pesos».

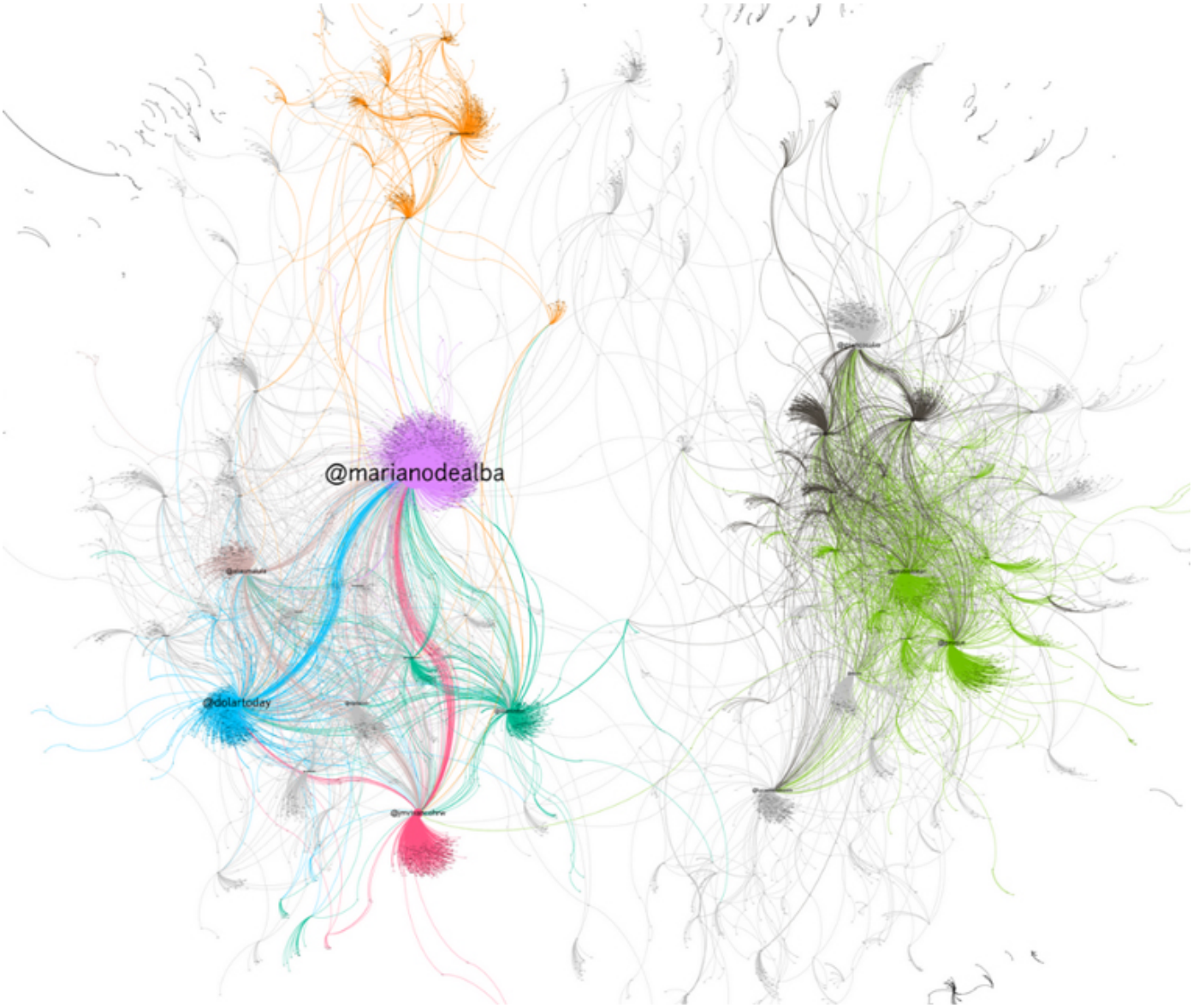
Ha arribat el moment d'exportar: anem al menú `File -> Export -> SVG/PNG/PDF file...` i seleccionem si volem un arxiu vectorial (PDF o SVG) o una imatge rasteritzada, de píxels (PNG). A les opcions podem canviar la grandària i el marge de la imatge.



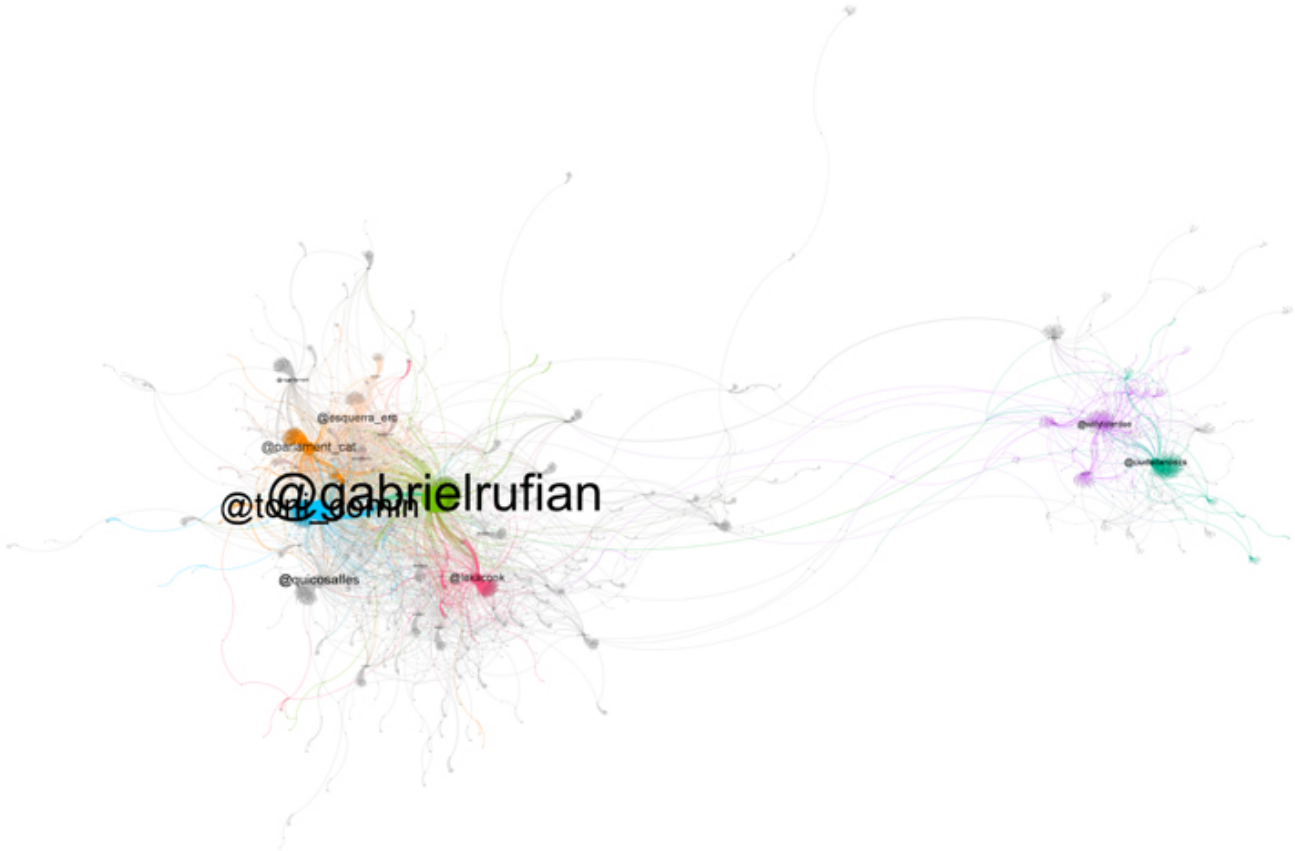
Haurem de realitzar el procés d'exportació repetides vegades fins que ajustem la grandària dels nodes i les etiquetes que desitgem.

Resultats

Graf de la cerca «Felipe González»: es representen els usuaris d'aquestes piulades com a nodes, mentre que les relacions representades com a vèrtexs són els RT entre ells. Piulades descarregades amb *search query* des de 2018-01-09 00:49:41 fins a 2018-01-16 12:43:37: un total de 17.597 piulades.



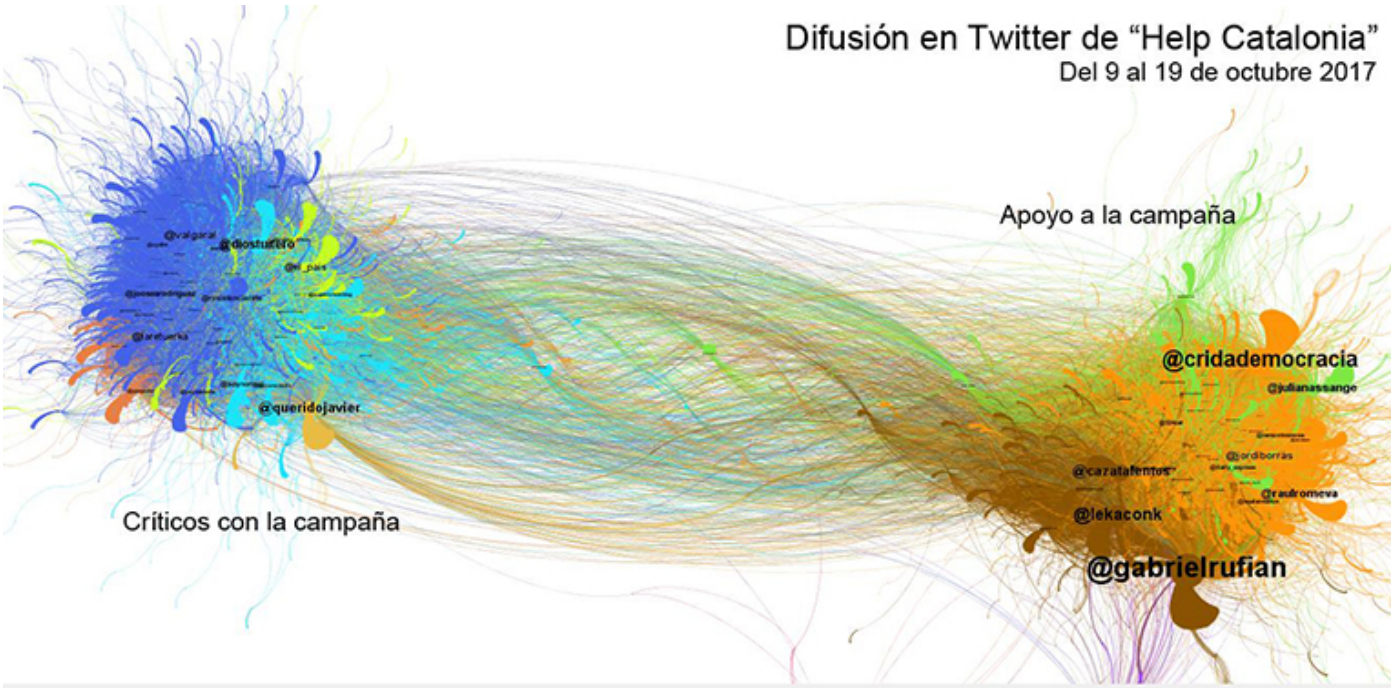
Graf de l'etiqueta #parlament: es representen els usuaris d'aquestes piulades com a nodes, mentre que les relacions representades com a vèrtexs són els RT entre ells. Piulades descarregades amb T-hoarder, amb *search query* des de 2018-01-17 13:20:53 fins a 2018-01-17 17:20:34 i amb *streaming* des de 2018-01-17 18:20:50.308562 fins a 2018-01-17 19:11:03: un total de 14.779 piulades.



e. Com presento la informació per fer un relat intel·ligible?

Es probable que el gràfic resultant no sigui prou explicatiu per respondre la pregunta inicial. Un graf sense una llegenda que expliqui els colors, sense un text que l'acompanyi o sense un titular o un destacat que l'expliqui no és una resposta vàlida. Els grafs s'utilitzen cada vegada més en la premsa, però és necessari aclarir què representen els punts i línies que estem veient. Aquest aclariment pot servir a tall d'exemple:

> «Els grafs s'han generat prenent les repiulades com a element de connexió dels usuaris. Cada node és un usuari i cada arc és una connexió entre l'usuari que repiula i l'usuari repiulat. La grandària dels nodes és directament proporcional al nombre de RT rebuts».
>
> Mariluz Congosto (2016).



Exemple de graf amb títol, descripció d'usuaris i període. Difusió de la campanya «Help Catalonia» a Twitter
Font: http://www.barriblog.com/2017/10/hilo-difusion-la-campana-help-catalonia-twitter/.

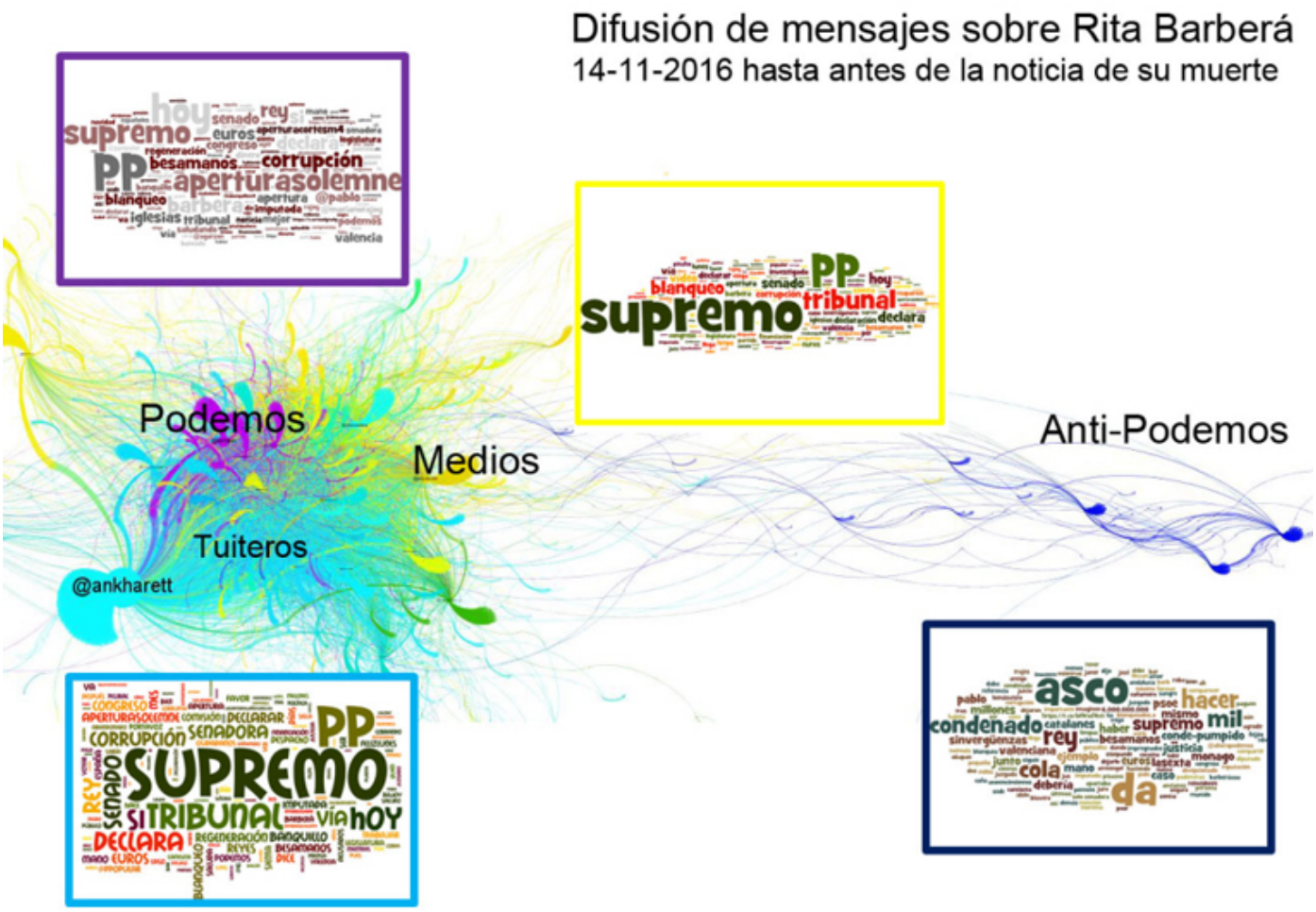
Per fer-ho, hem d'explicar el graf o els grafs resultants i acompanyar-los de tots aquests elements. També hem d'afegir-hi documentació sobre com s'han obtingut les dades i com s'han realitzat els gràfics. Si has capturat dades d'un període de temps llarg, és possible que necessitis realitzar diferents grafs segons les fases de propagació de l'etiqueta.

Pots usar els filtres que ofereix Gephi per mostrar les diferents categories (`modularity`) que s'usen per acolorir els gràfics. Això et permetrà analitzar la composició d'aquestes xarxes i analitzar de què parla cadascuna d'elles. Aquest tipus de filtres permet activar només els usuaris d'una o més categories.

Pas a pas:

- Tria un títol.
- Escriu un paràgraf explicatiu/introductori.
- Afegeix gràfics que l'acompanyin.

Per a cadascun dels grups detectats (*modularity*) podem calcular les paraules més usades (bigrames i trigrames) amb databasic.io i/o utilitzar-les per fer un rànquing o un núvol. T-hoarder també ofereix la possibilitat de calcular les paraules més populars amb l'opció `7. Processing tweets (Entities)`. A partir de llista de tuits, les paraules associades a cada grup poden ajudar a entendre el graf:



Núvol de paraules associades a cada categoria d'usuaris després de la mort de Rita Barberá
Font: http://www.barriblog.com/2016/11/despues-la-muerte-rita-barbera-twitter/.

Exemples d'anàlisis amb aquesta metodologia:

- Així es es va fer viral Tabarnia a les xarxes socials. Mariluz Congosto, 2018.
- Difusió de la campanya «Help Catalonia» en Twitter.

Nota

Pots descarregar el .gdf per replicar aquest exercici:

- helpcatalonia_all_RT.gdf.
- helpcatalonia_most_RT.gdf.

6.1.4.2. Com es relacionen els seguidors d'un compte de Twitter?

a. Quina anàlisi o visualització respon a la pregunta?

Un graf.

b. Quines dades necessito i on les puc trobar?

Crea un llistat amb els usuaris que han publicat una etiqueta determinada (pregunta anterior): com s'obté aquesta llista? Amb l'opció `7.Processing tweets` i usant les opcions `entity` (que extreu els usuaris més repiulats, els més esmentats i els més actius), T-hoarder_kit et permet obtenir un fitxer amb la llista d'usuaris. Cal netejar el fitxer generat, ja que té altres variables associades i el que necessitem és un fitxer amb cada nom d'usuari en una línia. El següent pas consisteix a obtenir els perfils (*profiles*) d'aquests usuaris: `2.Get users information`, escriu l'opció `profile`. Això genera un arxiu amb el nom `xxxxx_profiles.txt` que conté tots els perfils.

A partir d'aquesta llista de perfils es poden calcular les seves relacions (*follower–following*). Atenció: a l'hora d'obtenir els *followers-following* de cada usuari, l'API els proporciona a una velocitat de 60 usuaris cada hora. Per tant, trigarem 16,6 hores a per analitzar la relació entre 1.000 usuaris. Amb l'arxiu .gdf que hem generat passem al següent pas. Per tant, abans de començar a veure les relacions entre un grup d'usuaris d'una etiqueta determinada, podem seleccionar els usuaris segons un criteri (activitat o rellevància) per fer llistes d'usuaris més petites i, per exemple, estudiar com estan relacionats els 100 usuaris més esmentats o els 100 usuaris més actius.

Has de fer un graf de relacions declarades (*following* i *followers* entre els usuaris de l'etiqueta).

c. Quines eines de d'utilitzar per preparar les dades?

No aplicable. L'arxiu .gdf que has generat ja és vàlid per al següent pas.

d. Quines eines he d'utilitzar per crear l'anàlisi o la visualització?

Per generar el graf seguim el procediment descrit en l'exercici anterior, amb la diferència que ara tindrem molts menys nodes que en un graf de RT. El *layout Forceatlas* funciona millor per a menys nodes que el *Forceatlas2*.

e. Com presento la informació per fer un relat intel·ligible?

De la mateixa manera que a l'exercici anterior.

Exemple d'anàlisi amb una metodologia similar

- Digital Humanities on Twitter, a small-world? Martin Grandjean, 2015.
- Gregorio Morán, una columna vacía. Mariluz Congosto, 2017.

Pots descarregar el .gdf amb les relacions entre usuaris per replicar l'exercici: gregoriomoran_users_profiles.gdf.

6. Tres exercicis de periodisme de dades

6.1. Twitter: captura de dades i anàlisi de xarxes

6.1.5. Informació complementària

- Animació en un mapa Torque amb CARTO: exemple.
- Obtenció de piulades i graf amb http://flocker.outliers.es/.
- Tutorials de Gephi:
  - https://www.slideshare.net/gephi/gephi-quick-start.
  - https://www.slideshare.net/gephi/gephi-tutorial-visualization.
- Article sobre T-hoarder: 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, pàg. 28-39).

6. Tres exercicis de periodisme de dades

6.2. Efecte Airbnb: anàlisi geoespacial

6.2.1. Introducció

Últimament es parla molt de la presència d'Airbnb a la teva ciutat. Hi ha organitzacions ciutadanes que es comencen a moure perquè als seus barris hi ha cada vegada més turistes, i l'ajuntament està pensant a regular la situació. El diari digital pel qual treballes ha decidit treure'n l'entrellat amb una anàlisi gràfica de les dades disponibles. Quin és l'impacte d'Airbnb a les diferents zones de la ciutat? És similar al d'altres ciutats?

Nota

Aquest exercici està basat en la sèrie de tallers d'anàlisi de dades del projecte Efecto Airbnb de Montera34 i en la sèrie de manuals sobre anàlisis de dades d'Airbnb.

Coses que pots aprendre: on i com aconseguir dades d'Airbnb i d'altres plataformes turístiques, netejar i preparar un conjunt de dades amb OpenRefine, produir una cartografia digital amb el servei CARTO.

6. Tres exercicis de periodisme de dades

6.2. Efecte Airbnb: anàlisi geoespacial

6.2.2. Context

L'aparició a internet de plataformes de lloguer d'allotjaments *peer to peer* està provocant canvis d'ús ràpids als centres urbans amb més atractiu turístic. Aquestes plataformes, amb Airbnb al capdavant, han estat acusades de contribuir a l'alça en els preus immobiliaris i de ser un element disruptiu en les comunitats locals. D'altra banda, es defensa que la suposada tendència gentrificadora del lloc web és exagerada.

6. Tres exercicis de periodisme de dades

6.2. Efecte Airbnb: anàlisi geoespacial

6.2.3. Conceptes bàsics o previs

Què és una API? Una API (de l'acrònim d'*application programming interface*) és la forma estructurada d'accedir i interactuar amb una aplicació o un lloc web.

Terminologia de les dades d'Airbnb. Airbnb utilitza una terminologia pròpia en la seva estructura de dades. És convenient conèixer-la abans d'explorar les seves dades: els anuncis es diuen *listings*, mentre que els usuaris són *hosts*.

Què és Carto.com? És un servei web disponible a https://carto.com que permet produir mapes i altres visualitzacions a partir d'un conjunt de dades. Carto.com és una instal·lació del programari CartoDB, que és de codi obert i està disponible lliurement: https://github.com/CartoDB/cartodb.

Què és OpenRefine? És un programari de codi obert que permet netejar i formatar un conjunt de dades mitjançant una interfície gràfica. Es pot descarregar a la web http://openrefine.org/.

Què és CSV? *Comma Separated Values* (CSV) és un format de text pla en el que cada línia equival a un registre i els diferents valors d'un registre se separen pel mateix caràcter, tot i que, contra tot pronòstic, aquest caràcter no té per què ser una coma, sinó que pot ser qualsevol caràcter.

Què és geoJSON? És la variació del format JSON que sorgeix en afegir-li informació de posició geogràfica, és a dir, coordenades de posició.

Què és un mapa de coropletes o mapa coroplètic? Segons Viquipèdia:

> «És un mapa temàtic en el qual les superfícies són ensombrides o patronades en proporció al mesurament de la variable estadística que es mostra al mapa, com per exemple, la densitat de població o els ingressos per càpita. Els mapes coroplètics proporcionen una visió ràpida de la variació sobre una superfície geogràfica o mostra el nivell de variabilitat dins una regió».

Què és un *scraper*? És un petit programari que permet recopilar automàticament les dades contingudes en un lloc web.

6. Tres exercicis de periodisme de dades

6.2. Efecte Airbnb: anàlisi geoespacial

6.2.4. Preguntes a resoldre

Les preguntes de l'exercici s'han respost per a la ciutat de Madrid, però el procediment és vàlid per a qualsevol altra ciutat. T'animem a que apliquis aquest exercici a la teva ciutat.

6.2.4.1. Quin és l'impacte d'Airbnb a les diferents zones de la ciutat?

Considerarem com a zones els districtes de Madrid.

a. Quina anàlisi o visualització respon a la pregunta?

Dues anàlisis complementàries:

- Un mapa on es vegin els contorns dels barris i els pisos d'Airbnb representats per punts. Les àrees poden tenir un codi de color que indiqui la quantitat de pisos d'Airbnb en aquesta zona o la proporció d'allotjaments d'Airbnb anunciats respecte del total d'habitatges. Aquesta segona opció, la proporcionalitat, és més indicada per a un mapa de coropletes.

- Un diagrama de barres que representi la quantitat de pisos Airbnb per districte.

b. Quines dades necessito i on les puc trobar?

Per calcular el nombre d'allotjaments d'Airbnb per districte necessitaràs el **llistat d'allotjaments d'Airbnb** geolocalitzats (amb latitud i longitud) i **els contorns dels districtes**. Per calcular la proporció d'allotjaments d'Airbnb respecte del total d'habitatges també necessitaràs el **nombre d'habitatges de cada barri**. Vegem com podem aconseguir cadascuna d'aquestes dades:

Llistat d'allotjaments d'Airbnb

Airbnb no té una API i, per tant, no resulta senzill interactuar o descarregar-se les dades de la plataforma. La manera de fer-ho és programar un *scraper* que recorri la web d'Airbnb descarregant les dades que ens interessen. Tot i que seria molt interessant fer-ho, en aquest exercici no ho farem.

L'alternativa és posar-se en contacte amb algú que ja hagi programat un *scraper* per al lloc web d'Airbnb i que ens faciliti les dades. Aquí teniu diversos llocs web que proporcionen dades d'Airbnb, i també us animem a que busqueu les vostres pròpies fonts:

- Inside Airbnb. Inside Airbnb és una pàgina alimentada per Murray Coix, que ha programat el seu propi *scraper* per obtenir dades d'Airbnb. Disposa de dades de desenes de ciutats. Si la ciutat que busques no apareix a la llista o si les dades no estan actualitzades, pots contactar directament amb Murray.

- Data Hippo. Data Hippo és una plataforma per allotjar dades de plataformes de lloguer turístic de ciutats d'Espanya i Portugal. Disposa de dades d'Airbnb, Homeaway i Housetrip per regions, províncies i ciutats. A les principals ciutats turístiques, disposa de dades per barris o districtes. Pots posar-te en contacte amb la gent de Data Hippo per mitjà del seu compte de twitter.

L'elecció de la font de les dades és crucial: d'aquesta elecció dependrà l'estructura de dades, que és diferent per a cada font i depèn de com de recents i completes siguin les dades. Abans de descarregar un conjunt de dades, convé consultar-ne la data de recol·lecció i els camps que conté.

Nota

En aquest exercici utilitzarem les dades de Data Hippo, que són del mes d'octubre del 2017 i, per tant, més recents que les d'Inside Airbnb, i que contenen la informació necessària per fer l'estudi per districtes.

Les dades de Madrid es poden consultar aquí: http://datahippo.org/en/region/599232178a46554f807aec1b/.

I es poden descarregar en format CSV:
http://datahippo.org/media/regions/ea554b7f-0516-4d1c-91e6-883567800e08/599232178a46554f807aec1b_airbnb.csv

Contorns dels districtes

Els contorns dels districtes de Madrid també els podem obtenir de Data Hippo en format geoJSON.

L'arxiu de Data Hippo també conté el total d'allotjaments Airbnb que hi ha a cada districte. En la secció següent veurem que amb aquest fitxer en tenim prou per realitzar l'exercici i que no ens farà falta el fitxer anterior. Tanmateix, recollirem el procediment complet per si les dades les dades de Data Hippo no estan disponibles.

A Data Hippo hi ha dos fitxers geoJSON per a Madrid: `geojson_region.json`, que conté el contorn del municipi, i `geojson_subregions.json`, que conté els contorns dels districtes. Pots accedir a aquests arxius des de la secció del municipi de Madrid.

Sempre és recomanable obtenir les dades de la font institucional (la font primigènia), que en aquest cas és l'Ajuntament de Madrid. Els contorns dels districtes també estan disponibles al portal web de l'Ajuntament de Madrid. Vénen en un *shapefile* (conjunt de fitxers que es pot baixar en format comprimit .zip), així que hauràs de transformar el seu sistema de projecció ETRS89 al CRS EPSG:4326 – WGS 84. Abans, Carto.com no acceptava altres CRS diferents de l'estàndard EPSG:4326 – WGS 84; ara, però, sí que n'accepta. Des de QGIS pots obrir l'arxiu que has baixat com a capa vectorial i guardar-lo com un nou arxiu amb el CRS que triïs (en aquest manual trobaràs una descripció més detallada).

Habitatges per districte

El nombre total d'habitatges per districte i barri el podem obtenir del portal web de dades obertes de l'Ajuntament de Madrid.

c. Quines eines puc utilitzar per preparar les dades?

Nombre de pisos d'Airbnb per districte

Les dades d'Airbnb que es poden aconseguir a Data Hippo ja estan segmentades per districte, així que la dada que necessitem ja és al conjunt de dades.

Si no la tinguéssim, la podríem obtenir mitjançant el procediment següent: podem calcular el nombre d'allotjaments per districte explicant quants punts hi ha dins del contorn de cada districte. Aquesta és l'operació que Data Hippo ha fet amb les dades.

La ubicació dels allotjaments d'Airbnb és aproximada. Segons el mateix lloc 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».

Per tant, si optem pel mètode d'explicar punts en contorns, hem de tenir en compte que alguns allotjaments estan situats fora del districte que els correspon i, per tant, els nombres inclouran una imprecisió.

Bola extra: Una manera de corregir aquest error és aconseguir dades d'Airbnb que incloguin el codi postal de cada allotjament (Data Hippo no el proporciona, però Inside Airbnb sí que ho fa). Llavors es podria fer una taula d'equivalència entre els districtes i els codis postals. Amb aquesta taula podríem recórrer el llistat d'allotjaments i afegir-hi una columna amb el districte. Aquesta unió de taules es pot realitzar amb QGIS o amb una consulta SQL.

Contorn de districtes

Els contorns que ens hem descarregat de Data Hippo estan preparats per ser utilitzats.

Habitatges per districte

L'arxiu que hem baixat del portal de dades de l'Ajuntament de Madrid és un full de càlcul amb una pestanya per a cada districte i moltes dades per a cada districte i cada barri. Cal anar a buscar la dada de quantitat d'habitatges a la cel·la 239I de cadascuna de les pestanyes. Es podria automatitzar l'obtenció d'aquesta dada per a cada pestanya, però com que hi ha tants pocs districtes no paga la pena fer-ho.

Enganxem aquí la taula amb el total d'habitatges a cada districte, i d'aquesta manera t'estalviem aquest pas tan avorrit:

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 continuació tens les mateixes dades però sense tildes i en format CSV, ja preparada per a 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. Quines eines puc utilitzar per produir la visualització?

El mapa de coropletes dels districtes es pot realitzar amb diferents eines. Una possibilitat és utilitzar QGIS, que ens permet fer el càlcul de punts (allotjaments) dins d'un polígon (districte) i produir un paquet que inclogui tot allò necessari (fitxers amb les

dades, llibreries Javascript i fitxer HTML) per publicar les dades en un lloc web. Una altra opció és utilitzar el servei Carto.com, que ens permet fer el mateix que QGIS i després generar un codi per incrustar el mapa en un altre lloc web.

> **Nota**
>
> Atès que estàs produint una visualització per incloure en el lloc web del mitjà digital on treballes, et recomanem que utilitzis Carto.com, ja que el codi que genera et permetrà incrustar fàcilment el mapa, com si fos un article normal.

Nombre de punts per districte

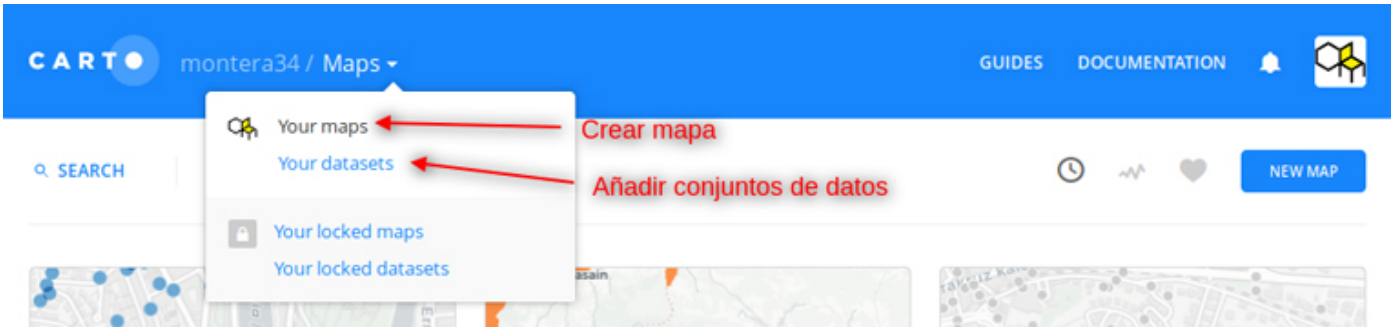
Com comentàvem a la secció anterior, el *dataset* de Data Hippo inclou el nombre d'allotjaments d'Airbnb per districte a la `airbnb_aptms`.

En cas que no el tinguem, cal seguir els passos que es mostren a continuació per obtenir-lo amb Carto.com.

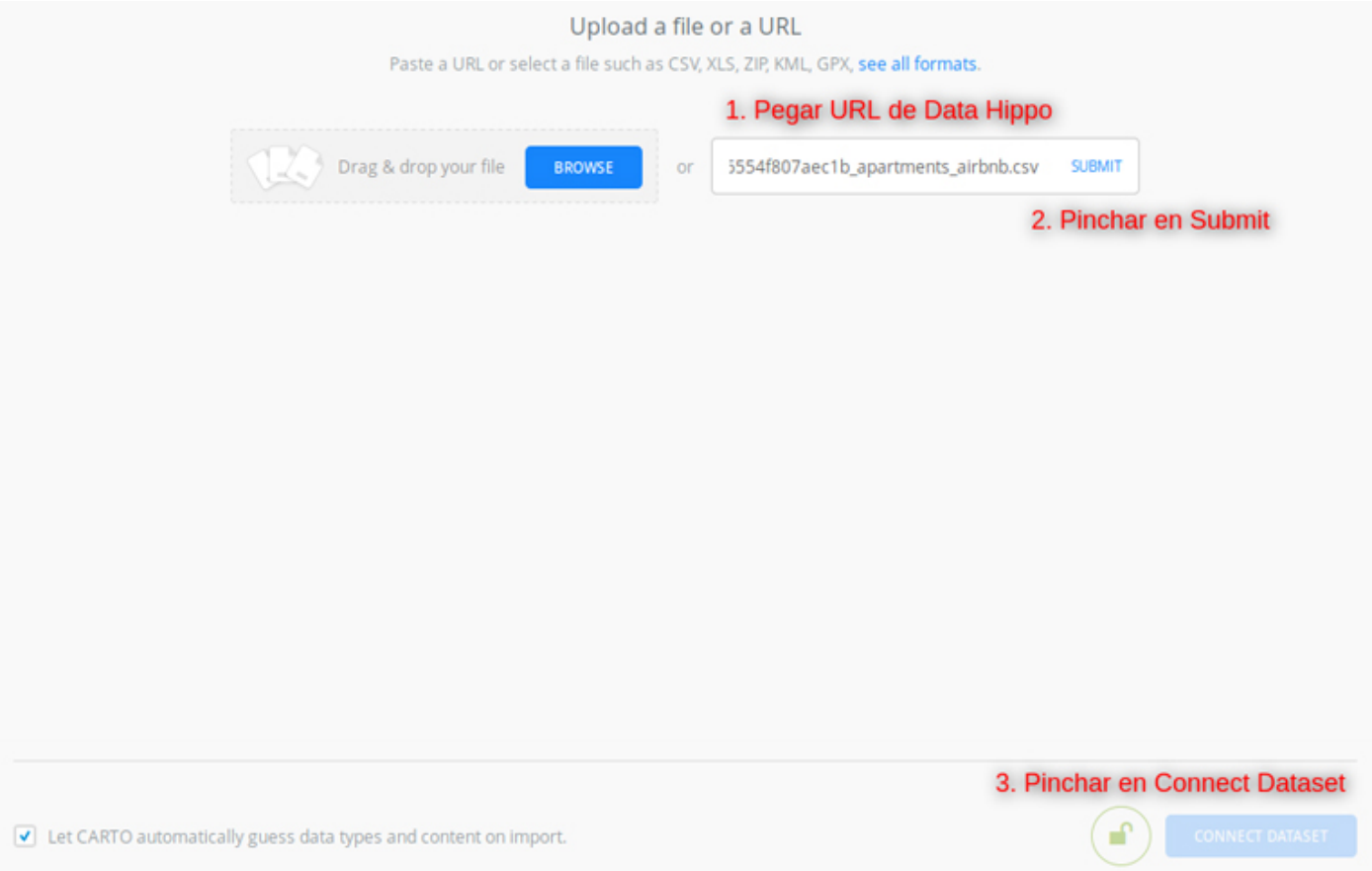
> **Nota**
>
> Durant l'elaboració d'aquests exercicis, Carto.com s'ha convertit en un servei de pagament. Per utilitzar-lo pots sol·licitar un accés d'estudiant (Student and Educator Accounts). També hem desenvolupat el mateix exercici amb QGIS. Està disponible al manual ¿Cómo se concentra la oferta por barrios de Airbnb? de la wiki de Montera34.

1) Inicia sessió amb el teu compte de Carto.com. Si no tens compte, dóna't d'alta.

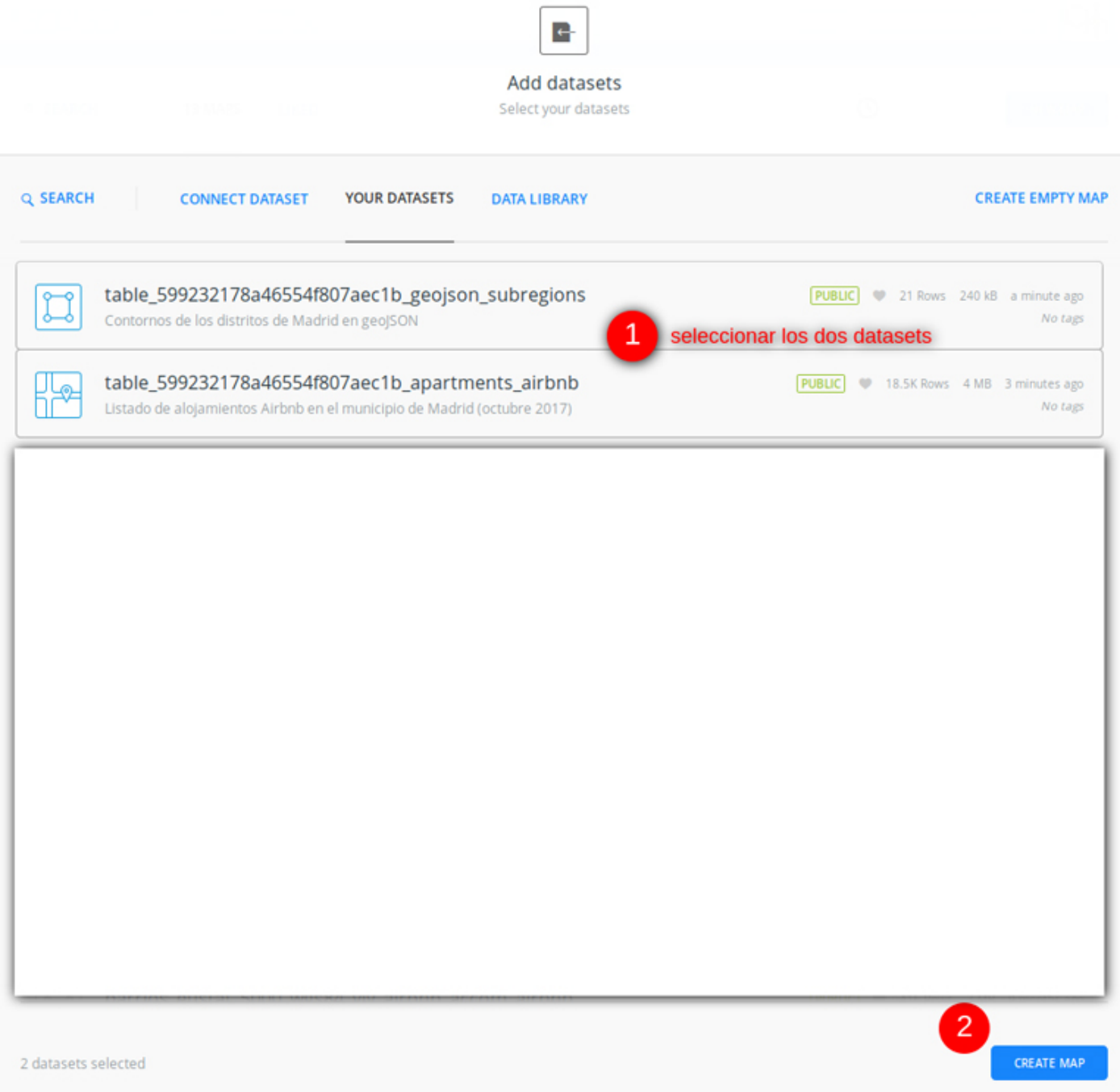
2) Afegeix com a nous *datasets* l'arxiu dels allotjaments i l'arxiu amb els contorns dels districtes.



No fa falta que baixis els conjunts de dades de Data Hippo; si proporcioneses les URL als arxius, Carto.com crearà els *datasets*:



3) Crea un mapa i afegeix com a capes els arxius que acabes de pujar:



4) Situa't a la capa dels contorns i fes clic a `Analysis`. Afegeix una nova anàlisi. Selecciona la manera `Intersect second layer. Intersect with a second layer and calculate aggregations on the fly`.

En el menú desplegat, selecciona la capa dels contorns com `Aggregate intersection` i la dels allotjaments com `Intersect Layer`.

A `Operation`, l'opció `COUNT` ha d'estar seleccionada.

Fes clic a `Done`. Això generarà una nova columna anomenada `count_values` a la base de dades.

5) Entra al mode taula amb el commutador de la cantonada inferior dreta del mapa i canvia el nom de la columna `count_values` per `alojamientos_airbnb`.

6) Ara ja podem produir un primer mapa de coropletes que acoloreixi els districtes en funció del nombre d'allotjaments d'Airbnb. Per fer-ho, anem a la capa de contorns i, a la pestanya `Style`, editem l'estil `Color` i assignem la columna que conté el conjunt d'allotjaments, `alojamientos_airbnb`, com a valor a l'opció `By value`.

Proporció d'allotjaments d'Airbnb per districte

Per fer aquest pas necessitem unir informació que es troba a fitxers diferents. En concret, necessitem fer la següent equació per a cada districte:

```
alojamientos de airbnb del distrito / viviendas totales del distrito
```

Per poder relacionar fàcilment aquests dos valors per a cada districte, hem d'incloure els totals d'habitatge en el *dataset* dels contorns. Una opció és fer-ho manualment, ja que són pocs valors. Tanmateix, també recollim l'opció semiautomàtica per a casos en què hi hagi una gran quantitat de dades. Ho farem amb una consulta SQL a Carto.com.

1) Creem un nou *dataset* amb el llistat d'habitatges totals per districte. Anotem el nom del *dataset* i canviem el nom de la columna que conté els districtes a `districte`, i de la columna que conté el nombre total d'habitatges a `total`.

2) Creem dues noves columnes en el *dataset* de contorns: `total_viviendas` i `percent_turisticas`. Per a ambdues columnes, seleccionem `number` com a tipus de dada.

3) Ens hem d'assegurar que el nom dels districtes està escrit exactament igual en ambdós *datasets*: els mateixos caràcters accentuats, els mateixos espais, les mateixes majúscules, etc.

4) Tornem al mapa i entrem a la capa de contorns. A la pestanya `Data`, a baix, canviem a la consola SQL amb el commutador situat a la part inferior de la barra lateral esquerra.

5) Escrivim la següent 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.distrito
```

que, traduït al llenguatge comú, significa: actualitza (UPDATE) el *dataset* de contorns i posa (SET) el nombre d'habitatges per districte de la taula d'habitatges per districte (el nom de la columna és total) (FROM) a les files on coincideixi la columna de districte del *dataset* de contorns amb la columna de districte del *dataset* de total d'habitatges (WHERE). Per això és important que siguem consistents a l'hora d'escriure els noms dels districtes.

D'aquesta manera tindrem les dades d'habitatges per districte a la taula de contorns de districtes.

Ara només ens queda calcular la ràtio d'allotjaments d'Airbnb per cada 100 habitatges. Podem fer-ho amb una altra consulta SQL:

```
UPDATE nombre_dataset_contornos SET percent_turisticas = airbnb_aptms * 100 /
total_viviendas
```

Per arrodonir els decimals i evitar resultats com ara 8,21439204893, fem una altra consulta i la restringim a dos decimals:

```
UPDATE nombre_dataset_contornos SET percent_turisticas = round( CAST(percent_turisticas as
numeric), 2 )
```

Visualitzacions

1) Mapa de coropletes. Ara només ens falta fer el mapa de coropletes utilitzant la variable `percent_turisticas`, tal com hem descrit a l'apartat anterior. És fàcil.

Per fer el mapa públic només falta prémer el botó *Publish* i, a continuació, copiar el codi per incrustar el mapa a qualsevol lloc web.

2) Gràfic de barres. Podem baixar les dades que hem calculat a Carto.com i utilitzar-les amb qualsevol programa de fulls de càlcul per generar un gràfic de barres.

Per baixar les dades, ves al mode *Data view* (taula) de la capa de contorns. En el menú `Edit > Export Layer` es pot baixar la capa en format CSV.

Exemple

Aquí tens un exemple d'un gràfic de barres doble amb la ràtio d'allotjaments d'Airbnb per cada 100 habitatges i la quantitat d'allotjaments per barris de diferents ciutats. Està desenvolupat amb la llibreria de JavaScript D3js en el projecte Efecto Airbnb de Montera34. El codi font està disponible.

e. Com presento la informació per fer un relat intel·ligible?

Quan tinguem les dades de les ràtios per districte ja podrem comparar uns valors amb d'altres. Amb la representació en el mapa de coropletes podem comparar la distribució espacial (és més difícil apreciar la diferència de valors associats a colors), mentre que en el gràfic de barres podem comparar la presència d'Airbnb a cada barri. També podem publicar una taula amb els districtes on hi ha més presència d'Airbnb.

Per completar el reportatge també pots analitzar el tipus d'allotjament que s'ofereix a cadascun dels districtes (`private_room` o `entire_home`, dins de la variable `room_type`) o el total de places (`variable capacity`).

El procés per calcular la ràtio d'allotjaments d'Airbnb per cada 100 habitatges és anàleg. Busca els contorns i les dades de cada districte i repeteix el procés.

6. Tres exercicis de periodisme de dades

6.2. Efecte Airbnb: anàlisi geoespacial

6.2.5. Informació complementària

- Articles de premsa que estudien l'impacte d'Airbnb:
  - «Dormir en el limbo: radiografía de Airbnb». Reportatge de *El Español*.
  - Estudi sobre Mallorca amb dades d'Inside Airbnb.
  - «Airbnb vs. Berlin». Informe gràfic de la situació del lloguer de pisos d'Airbnb a Berlín.
- Altres: Pla Especial Urbanístic d'Allotjament Turístic (PEUAT) de Barcelona. Explicació gràfica de les mesures de zonificació engegades a Barcelona per controlar els efectes dels pisos turístics als diferents barris.

6. Tres exercicis de periodisme de dades

6.3. Educació pública-privada: obrint dades obertes

6.3.1. Introducció

Ha sorgit una urgència per a l'especial sobre educació. Una vegada més, la teva sèrie de reportatges haurà d'esperar. Et demanen que, en dos dies, enllesteixis un article per a l'edició d'Euskadi sobre les diferències entre l'educació pública i l'educació privada. Les dades vénen en un informe en format PDF i, per tant, has de passar dades a net per poder fer alguna cosa amb elles. No és la millor situació, però almenys tens les dades.

Nota

Aquest exercici està basat en l'anàlisi de la segregació escolar a Euskadi.

Aprenentatges: rescatar dades de taules incloses en PDF, unir dades espacials i taules de dades, fer mapes de coropletes, generar mapes interactius en D3.

6. Tres exercicis de periodisme de dades

6.3. Educació pública-privada: obrint dades obertes

6.3.2. Context

A Euskadi hi ha una de les majors xarxes educatives privades-concertades de tot Europa, amb una quota del 48,9% de l'alumnat. També encapçala el llistat de les comunitats autònomes d'Espanya quant a percentatge d'alumnat en xarxa concertada, amb un 48,1%. Aquesta realitat on l'alumnat està repartit gairebé a parts iguals entre les xarxes públiques i concertades (a Euskadi, amb prou feines hi ha centres privats) ha donat lloc a fortes diferències en el tipus d'alumnes escolaritzats a una xarxa i a l'altra. Es parla de «segregació escolar»: determinats sectors socials no es troben homogèniament distribuïts a les xarxes. Tenint en compte que la xarxa concertada està completament subvencionada per l'Estat, hauria de tenir els mateixos criteris d'admissió que la pública. L'objectiu d'aquest exercici és verificar amb dades si hi ha segregació escolar a Euskadi i, si és així, quin grau de segregació escolar hi ha.

Les dades disponibles provenen de l'informe «La educación en Euskadi 2013-2015», del Consejo Escolar de Euskadi: un document en PDF que inclou dades en taules i inclou gràfics de línies, barres i mapes. Com és habitual, lamentablement les dades no estan en formats reutilitzables ni es troben a cap catàleg de dades obertes. No obstant això, com que les taules estan disponibles, pots accedir a les dades i convertir-les a CSV copiant-les i enganxant-les a un full de càlcul. El nivell de desagregació de les dades és el de zona escolar (no estan disponibles les dades de cada centre escolar).

Cada administració (local, nacional, etc.) té dades que permeten fer un estudi similar al que proposem en aquest exemple de l'educació a Euskadi.

6. Tres exercicis de periodisme de dades

6.3. Educació pública-privada: obrint dades obertes

6.3.3. Conceptes bàsics o previs

Què és GIS? GIS és l'àcrònim de «sistema d'informació geogràfica» (en anglès, *geographic information system*), que es defineix com el conjunt d'eines que permeten capturar, emmagatzemar, analitzar i presentar dades geogràfiques.

Què és QGIS? Un programari lliure de GIS que permet visualitzar, editar i analitzar informació geoespacial.

Què és GeoJSON? És un format estàndard obert (un tipus de fitxer amb extensió .geojson) dissenyat per representar elements geogràfics senzills juntament amb els seus atributs no espacials. Està basat en Javascript Object Notation (JSON) i pot estar compost de punts, línies i polígons.

```
{ "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": {
```

Què és TopoJSON? Un arxiu de tipus TopoJSON (amb extensió .json) és un tipus de format de GeoJSON en el qual les geometries estan compostes per segments anomenats arcs. Si dos o més contorns comparteixen una línia, com és el cas d'una frontera entre dos països, aquest segment compartit només queda definit una vegada. Això redueix la redundància i la grandària del fitxer. Aquest tipus de definició fa que aquests fitxers defineixin una topologia, i no simplement elements per separat. Això, entre altres coses, permet fer cartogrames.

6. Tres exercicis de periodisme de dades

6.3. Educació pública-privada: obrint dades obertes

6.3.4. Preguntes a resoldre

6.3.4.1. Com es reparteix l'alumnat entre la xarxa pública i la privada a Euskadi? Hi ha segregació escolar?

a. Quina anàlisi o visualització respon a la pregunta?

Volem estudiar quin tipus d'alumnat pertany a cada xarxa per investigar si hi ha segregació escolar. És molt probable que l'exploració de les dades ens porti a preguntar-nos el perquè de les diferències tan grans entre els indicadors de la xarxa pública i els de la privada: aquestes diferències són homogènies arreu del territori? Pots explorar si hi ha diferències per província o per nivell socioeconòmic.

Passem a realitzar diverses anàlisis i visualitzacions:

- Una sèrie de mapes de coropletes que mostrin els contorns de les zones escolars, on el color de cada zona representi el percentatge de l'alumnat de cada xarxa (pública o privada) que té una beca de material escolar o una beca de menjador. Aquests indicadors són la millor manera que tenim per estudiar el perfil socioeconòmic dels alumnes de cada zona escolar, ja que no disposem de les dades desagregades de l'ISEC (l'índex socioeconòmic i cultural) recopilades a partir de les respostes recollides en proves d'avaluació. Una alta presència d'alumnat becat indicarà que aquesta zona escolar i aquesta xarxa escolar tenen un nivell socioeconòmic menor. També utilitzarem el percentatge d'alumnat estranger, que sol indicar el nombre d'immigrants, que acostumen a tenir un nivell socioeconòmic més baix (el 63% de l'alumnat estranger prové d'Amèrica del Sud i d'Àsia).

- Gràfics que mostrin les diferències entre els indicadors a les zones escolars.

b. Quines dades necessito i on les puc trobar?

Per poder representar les dades en un mapa cal obtenir els **indicadors** de cada zona escolar i els **contorns** d'aquestes zones escolars. Vegem com ho podem fer.

Indicadores

Necessitem dades al més desagregades possible de la xarxa pública i de la xarxa privada. Les dades disponibles provenen de l'informe del Consejo Escolar de Euskadi «La educación en Euskadi 2013-2015», que està en format PDF. Les dades més detallades es donen per zones escolars, que de vegades abasten diversos districtes o municipis sencers. No hi ha dades disponibles per centres escolars. En els annexos 1.18, 1.19 i 1.21 de l'informe podem trobar diverses taules amb les dades que ens interessen.

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

Contorns de zones escolars

Lamentablement, els contorns de les zones escolars no estan disponibles enlloc (fins i tot la definició de la composició de vegades és imprecisa, sobretot en el cas de Vitoria-Gasteiz). És per això que, per facilitar l'exercici, hem creat un fitxer .geojson amb els contorns de les zones escolars d'Euskadi. Està fet a partir de la fusió de dos fitxers: el de barris i el de municipis d'Euskadi.

c. Quines eines puc utilitzar per preparar les dades?

Indicadors

Com hem assenyalat, les dades dels indicadors estan disponibles en els annexos del PDF de l'informe del Consejo Escolar de Euskadi. Lamentablement, com sovint succeeix amb aquest tipus de dades, no és possible obtenir-les en formats reutilitzables (fulls de càlcul, .csv, .json), i és necessari recompondre les dades per poder-les utilitzar. Per fer-ho, crearem una nova taula que inclogui una zona escolar a cada fila i cadascuna de les variables que ens interessen a cada columna. Si no volem transcriure un a un els nombres podem utilitzar eines automatitzades, com ara el programari lliure Tabula, o bé mètodes semiautomàtics, com ara la meravellosa funció de retallar i enganxar i altres trucs.

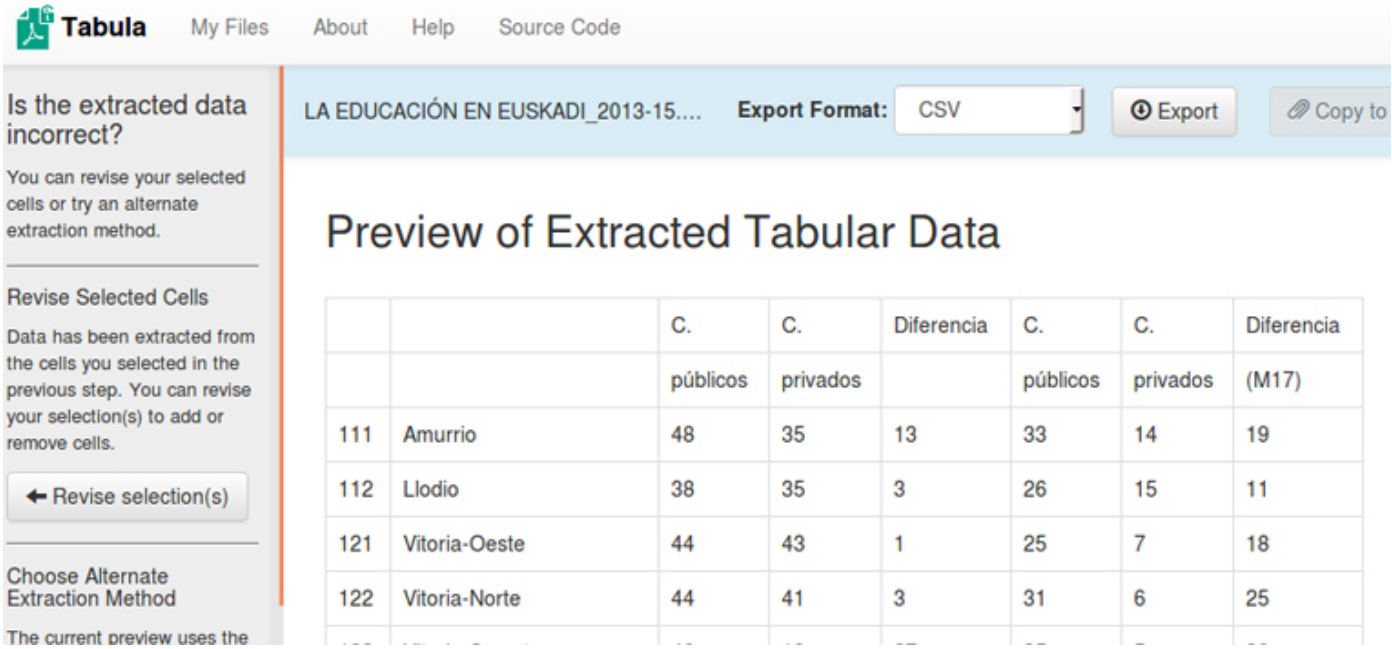
		% becarios material escolar			% becarios comedor		
		C. públicos	C. privados	Diferencia	C. públicos	C. privados	Diferencia (M17)
111	Amurrio	48	35	13	33	14	19
112	Llodio	38	35	3	26	15	11
121	Vitoria-Oeste	44	43	1	25	7	18

Taula amb cel·les combinades a la capçalera

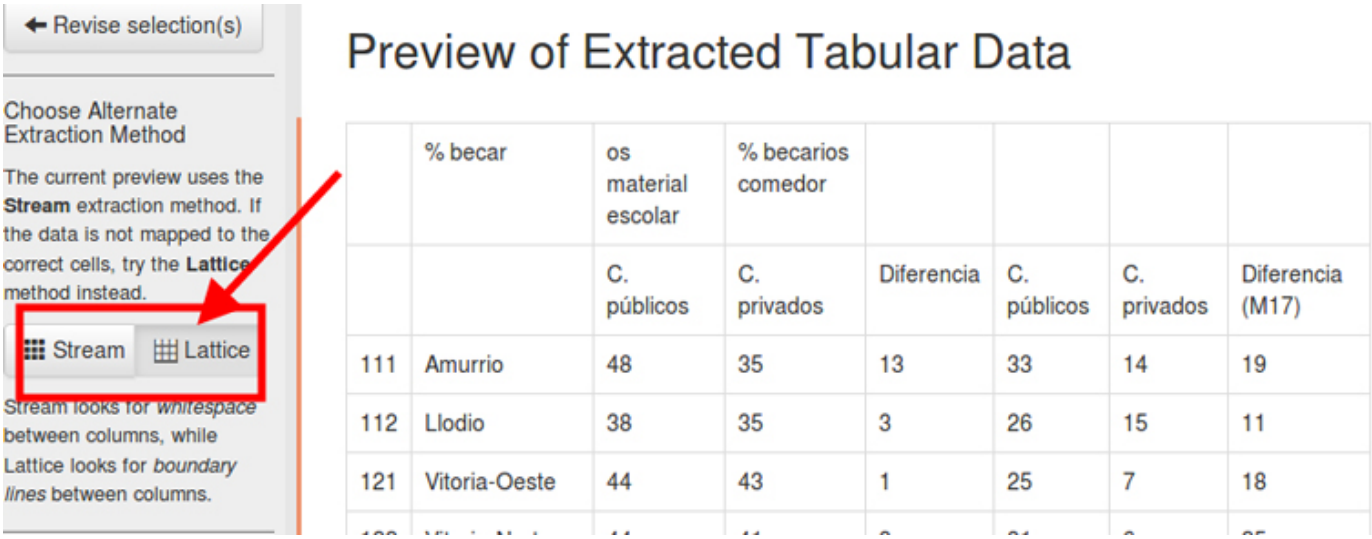
Tabula és un programa fàcil d'utilitzar. Primer, cal importar el PDF que conté les taules i, a continuació, cal seleccionar les taules que es volen exportar. Funciona bé per a taules que no tenen cel·les combinades, com és el cas de moltes de les capçaleres de les taules dels annexos. Perquè detecti bé les columnes, hem de deixar fora els encapçalats que contenen cel·les combinades, tal com s'indica en aquest gràfic:

		% 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 fem clic a `Preview & Export Extracted Data` obtindrem la taula:



Ara podem exportar-la a diversos formats (CSV, TSV, JSON) o copiar-la i enganxar-la directament al nostre programa de full de càlcul preferit. Com que moltes de les taules es divideixen en diverses pàgines, n'hi ha prou de copiar i enganxar les taules en el mateix document. Atenció: atès que els noms de les columnes no s'han importat bé, necessitaran ser retocats i caldrà afegir-hi l'indicador corresponent. Nota: Tabula permet dos mètodes d'extracció. D'altra banda, Lattice dona millors resultats per a aquest tipus de taules.



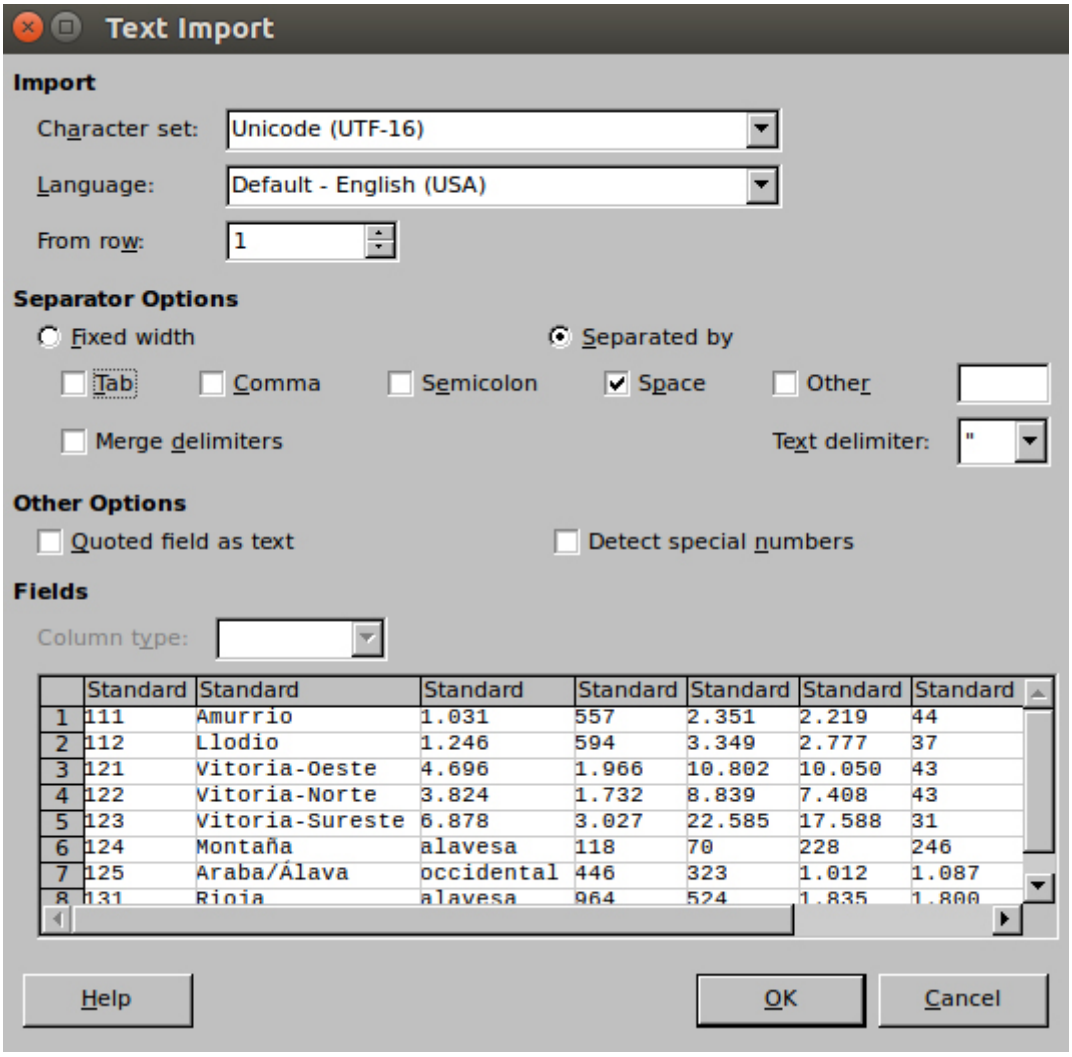
Vegem com quedaria el full de càlcul que estem completant:

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

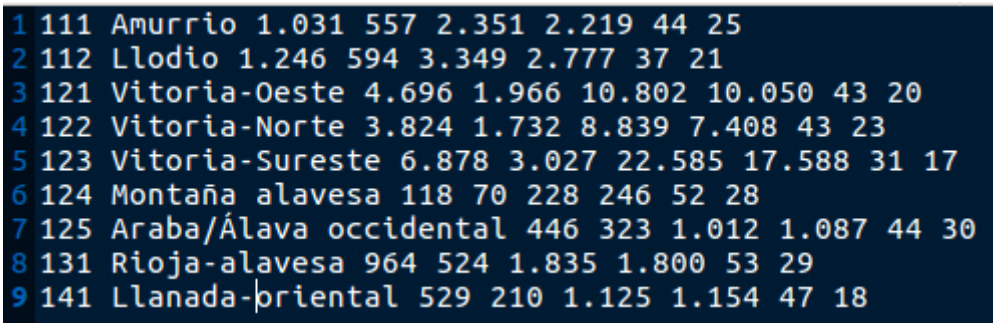
A la capçalera de cada columna hem afegit la procedència de les dades, el curs acadèmic i el nom de la variable sense espais i amb guió baix. Guardarem el fitxer o l'exportarem en format .csv o .tsv i eliminarem les primeres tres files de l'encapçalament, deixant que la primera sigui el nom de les variables:

```
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
```

L'altra manera d'extreure les dades del PDF és retallant-les i enganxant-les en un full de càlcul per poder tenir tots els indicadors en un únic fitxer de text pla. Si selecciones i copies el text d'una taula del PDF i l'enganxes en un full de càlcul, com ara Libre Office o Excel, segurament el programa et preguntarà amb quins delimitadors estan separades les columnes (en aquest cas, estan separades amb espais). Tal com veuràs, els noms de zones que porten espai, com ara Montaña alavesa, acostumen a donar problemes.



La solució passa per enganxar primer el text en un editor de text pla, editar els noms de les zones perquè no tinguin espais (Montaña-alavesa), copiar-los i tornar-los a enganxar en el full de càlcul, que ara no donarà problemes.



Pots baixar directament el CSV amb les variables extretes, però si fas això no aprendràs a usar Tabula.

Llistat de noms de les variables i la seva descripció:

Variables	Nom de les variables
alum_bec_mat_escolar	nombre alumnat becats material escolar
alum_bec_comedor	nombre alumnat becats menjador
alum_mat_escolar_total	nombre alumnat material escolar total
alum_comedor_total	nombre alumnat menjador total
perc_material_m15	% alumnat beca material escolar
perc_comedor_m16	% alumnat beca menjador

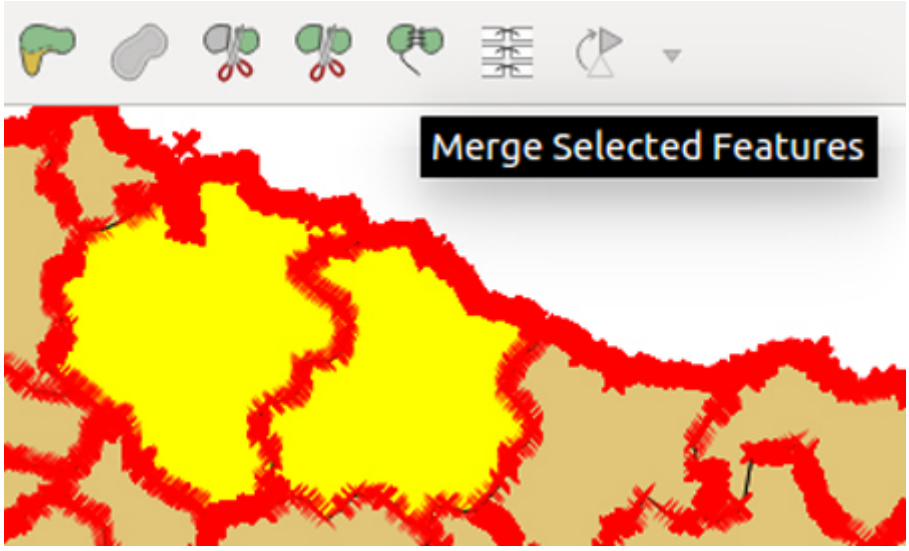
perc_bec_mat_escolar_pub	% alumnat becats material escolar pública
perc_bec_mat_escolar_priv	% alumnat becats material escolar privada
dif_perc_bec_mat_escolar_dif	diferencia % alumnat becats material escolar entre pública i privada
perc_bec_comedor_pub	% alumnat becats menjador pública
perc_bec_comedor_priv	% alumnat becats menjador privada
dif_perc_bec_comedor	diferencia % alumnat becats menjador entre pública i privada
total_alumnado	total alumnat
perc_alum_ext_publi	% alumnat estranger pública
perc_alum_ext_priv	% alumnat estranger privada
perc_alum_ext_todos	% alumnat estranger tots
indice_desigualdad_ext	índex desigualtat estrangers (% publica / % privada)
alum_ext_total	nombre alumnat estranger total
alum_ext_publi	nombre alumnat estranger pública
alum_ext_kristau_eskola	nombre alumnat estranger a xarxa Kristau Eskola
alum_ext_partaide	nombre alumnat estranger a xarxa Partaide
alum_ext_otros_privados	nombre alumnat estranger en altres privats
perc_reparto_alum_ext_publi	% repartiment alumnat estranger en pública
perc_reparto_alum_ext_privada	% repartiment alumnat estranger en privada

Contorns

Tal com hem comentat a l'apartat anterior, hem generat amb QGIS el fitxer dels contorns sobre la base de la informació disponible. Us estalviem aquest pas que consumeix molt temps i queda fora de l'abast d'aquest exercici. Hem indicat la composició de les zones en aquest arxiu TSV, a la columna `municipis_barris`. Els contorns de les zones han estat construïts a partir de:

- Límits municipals.
- Límits de barris d'Euskadi (font original i modificada amb projecció WGS84).

Per unir les àrees es pot utilitzar l'eina *Merge Selected Features*. Després de crear les àrees, s'afegeix a cadascuna d'elles les dades associades d'ID de zona i nom.



Icones d'eines per manipular àrees en QGIS

Aquí es poden baixar els contorns de les zones escolars a Euskadi:

- GeoJSON.
- Shapefile.

Associar indicadors als contorns

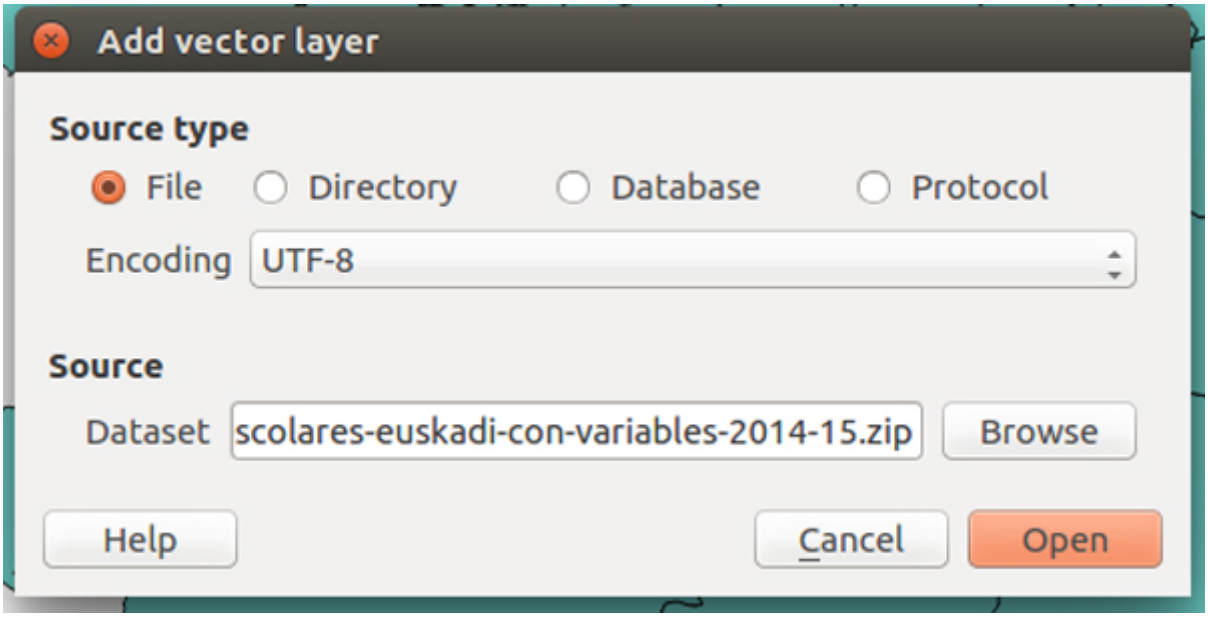
El següent pas consisteix a associar els indicadors al fitxer de contorns de les zones escolars.

En l'exercici anterior, el d'Airbnb, hem realitzat una acció similar amb Carto.com. Ara realitzarem aquesta acció amb QGIS. Per poder unir el fitxer d'indicadors i el de contorns, necessitem que ambdós tinguin almenys una columna en comú. En el nostre cas, utilitzarem la variable «zona_id» de tots dos fitxers, que conté el número que identifica la zona escolar.

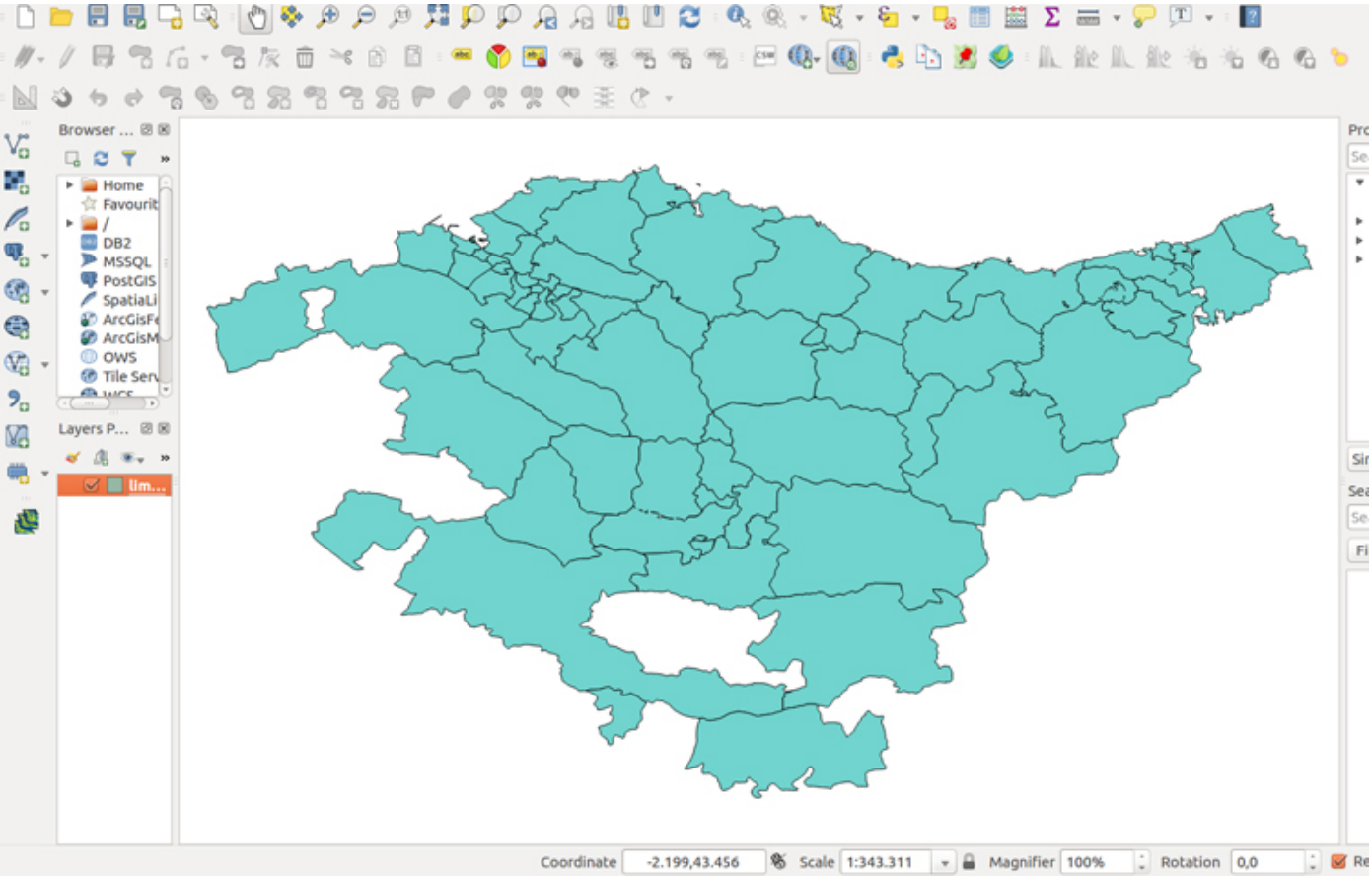
Per poder unir aquests dos fitxers necessitem importar-los a QGIS. Fem-ho.

1) Obre QGIS i crea un nou projecte.

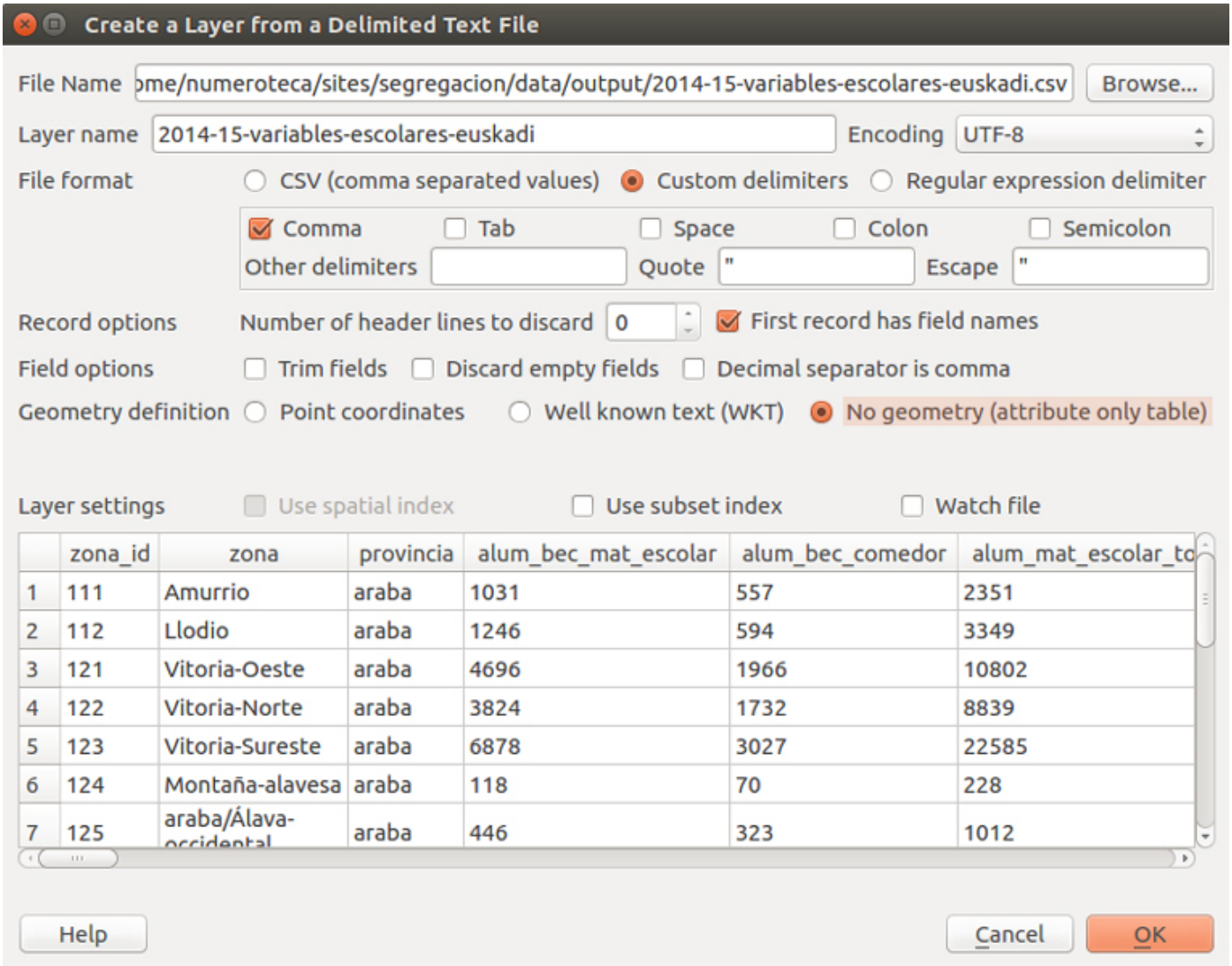
2) Importa el fitxer comprimit (.zip) *shapefile* des de `Layers > Add Layer > Add Vector Layer` i utilitza les opcions per defecte.



Aquest és el resultat:

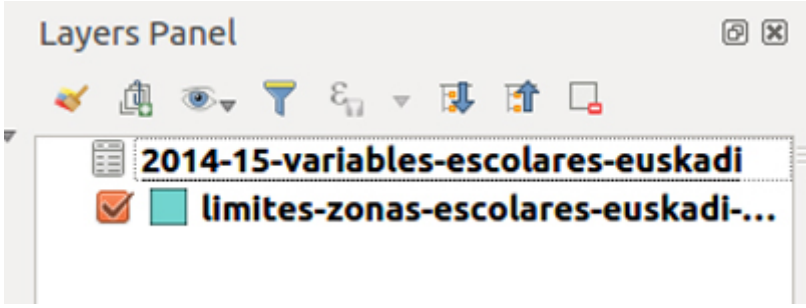


3) Importem el CSV amb els indicadors `Layers > Add Layer > Add Delimited Text Layer`. Cal indicar que els valors estan separats per comes, que la primera fila conté els noms de les variables i que la capa no conté geometria (és a dir, informació geoespacial).



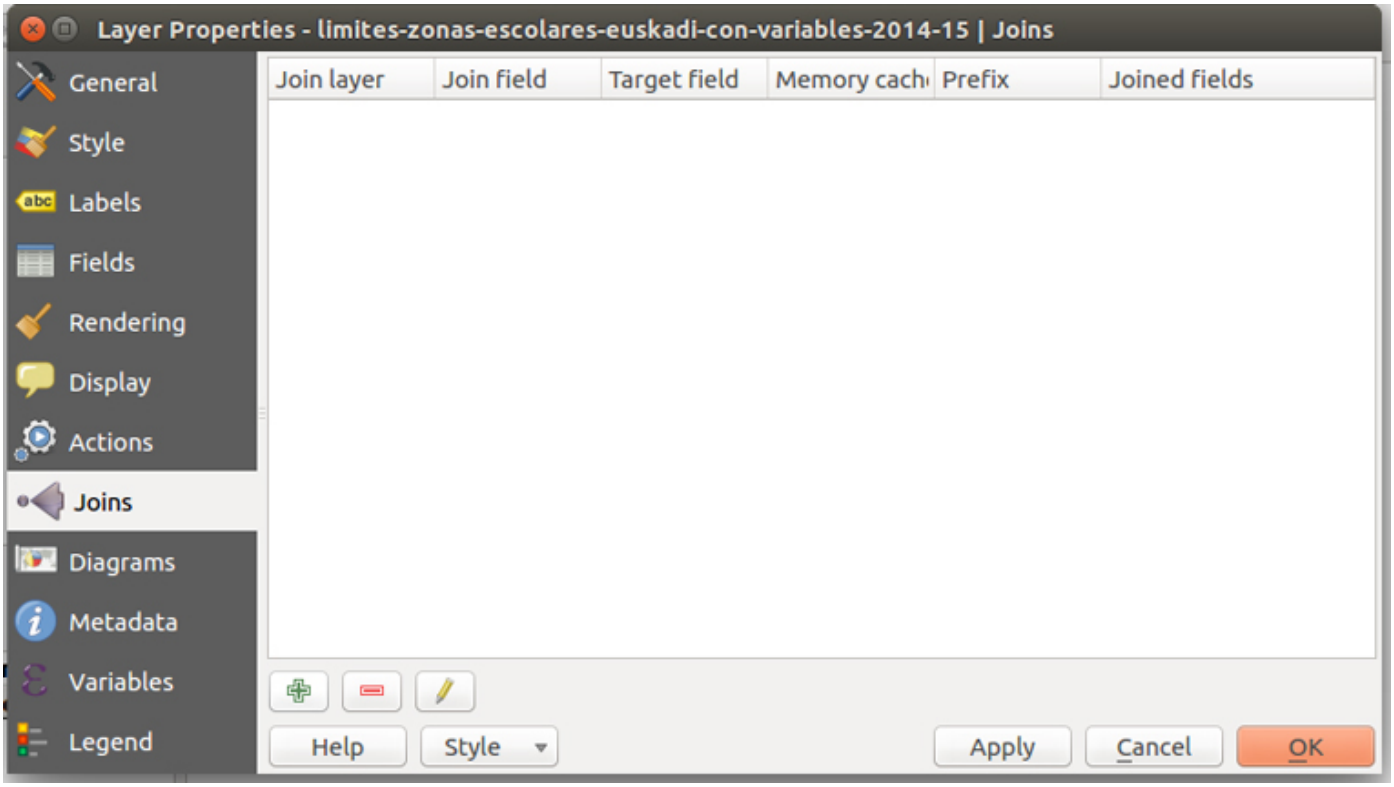
	zona_id	zona	provincia	alum_bec_mat_escolar	alum_bec_comedor	alum_mat_escolar_to
1	111	Amurrio	araba	1031	557	2351
2	112	Llodio	araba	1246	594	3349
3	121	Vitoria-Oeste	araba	4696	1966	10802
4	122	Vitoria-Norte	araba	3824	1732	8839
5	123	Vitoria-Sureste	araba	6878	3027	22585
6	124	Montaña-alavesa	araba	118	70	228
7	125	araba/Álava-occidental	araba	446	323	1012

En el panell de capes hauríem de veure les dues capes:

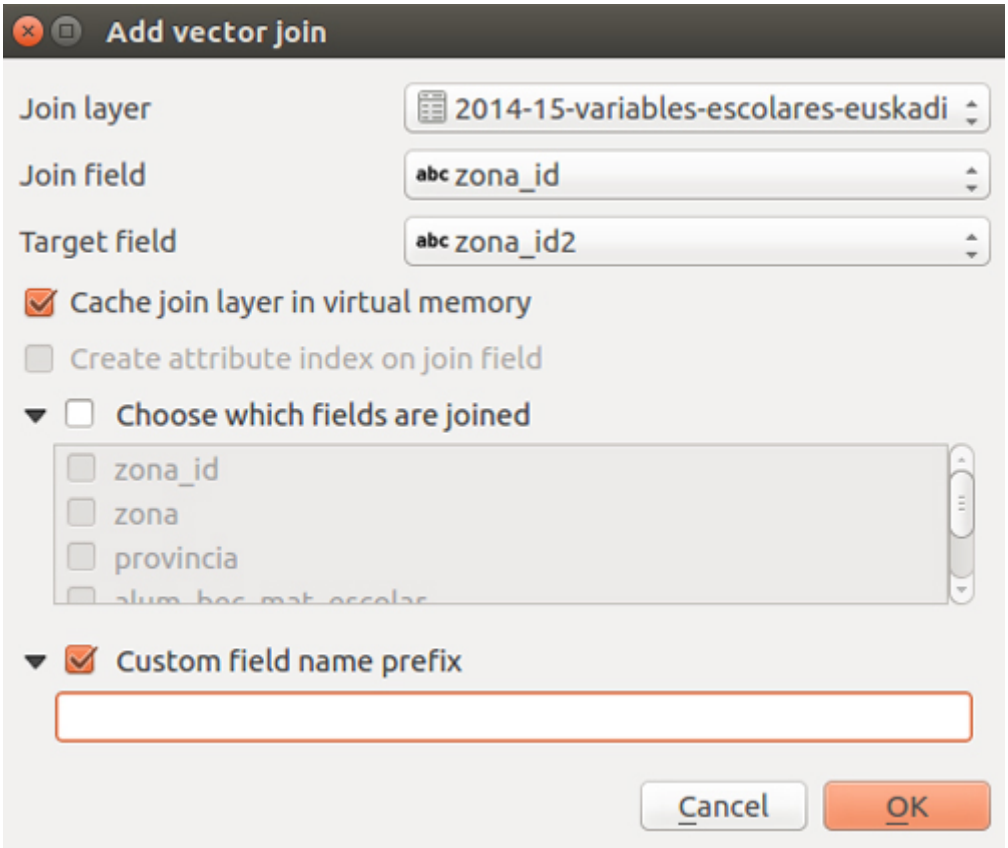


El símbol de la capa de dalt indica que conté text, mentre que la segona capa conté informació geoespacial.

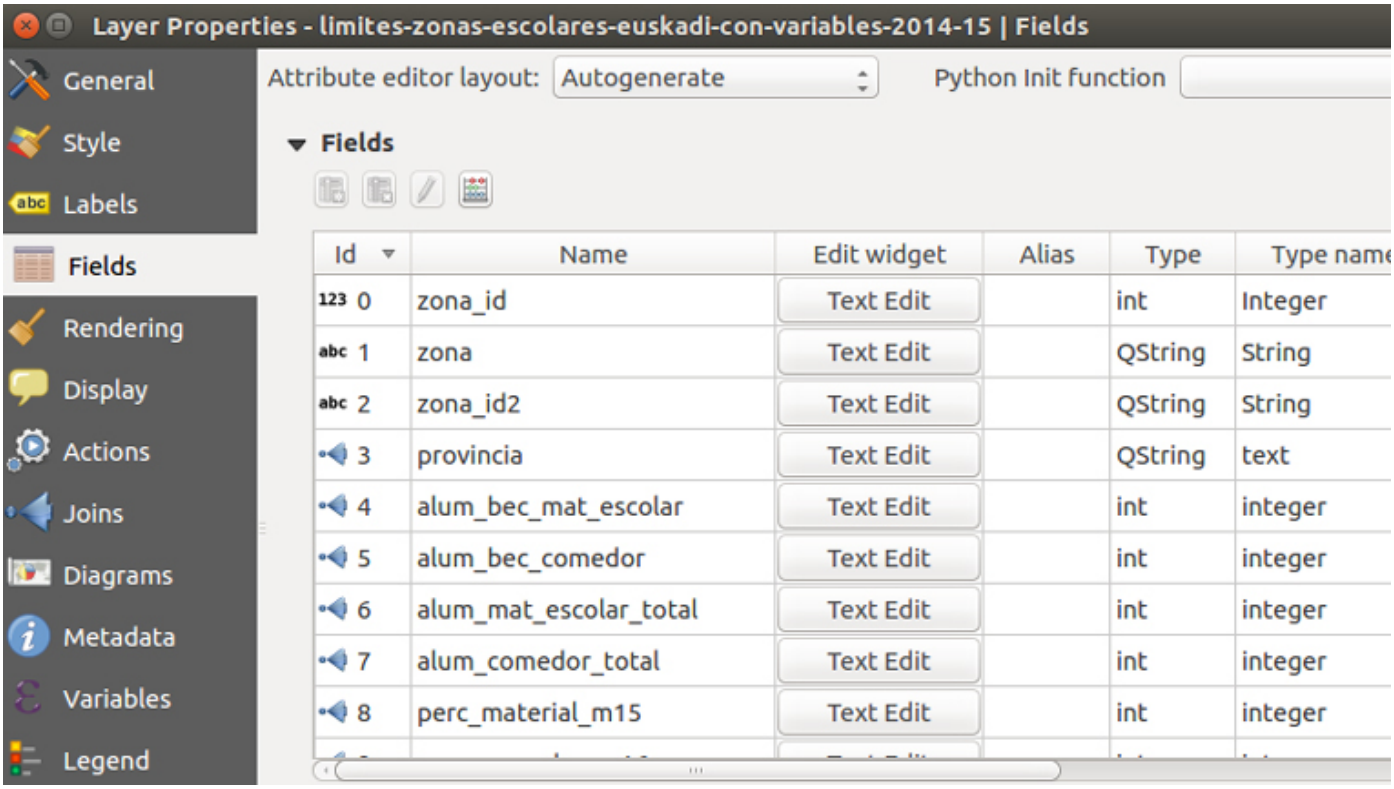
4) Clica amb el botó dret la capa dels contorns i selecciona `Propiedades`. Des del menú de `Propiedades`, selecciona `Joins > +`.



Selecciona la columna en comú, és a dir, la `zona_id` d'una capa i la `zona_id2` de l'altra. Deixa l'apartat `Custom field name prefix` buit.



A la pestanya `Properties > Fields` veuràs que les noves variables s'han afegit al fitxer de contorns:

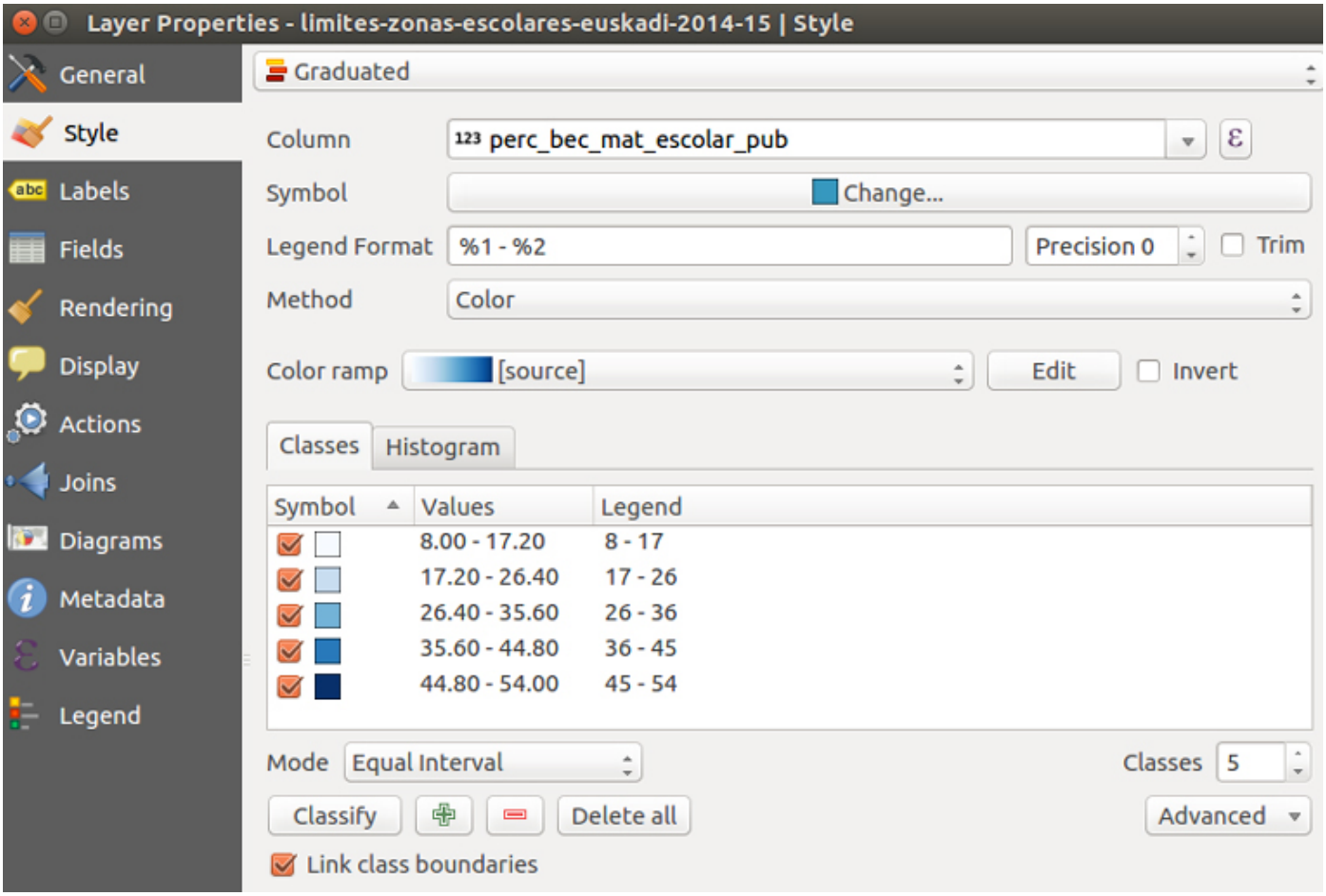


d. Quines eines puc utilitzar per produir l'anàlisi o la visualització?

Per produir els mapes de cadascuna de les variables, podem utilitzar QGIS o Carto.com, entre altres eines. Atenció: si volem comparar mapes de coropletes, hem de seleccionar una escala de color amb els mateixos rangs per als diferents mapes.

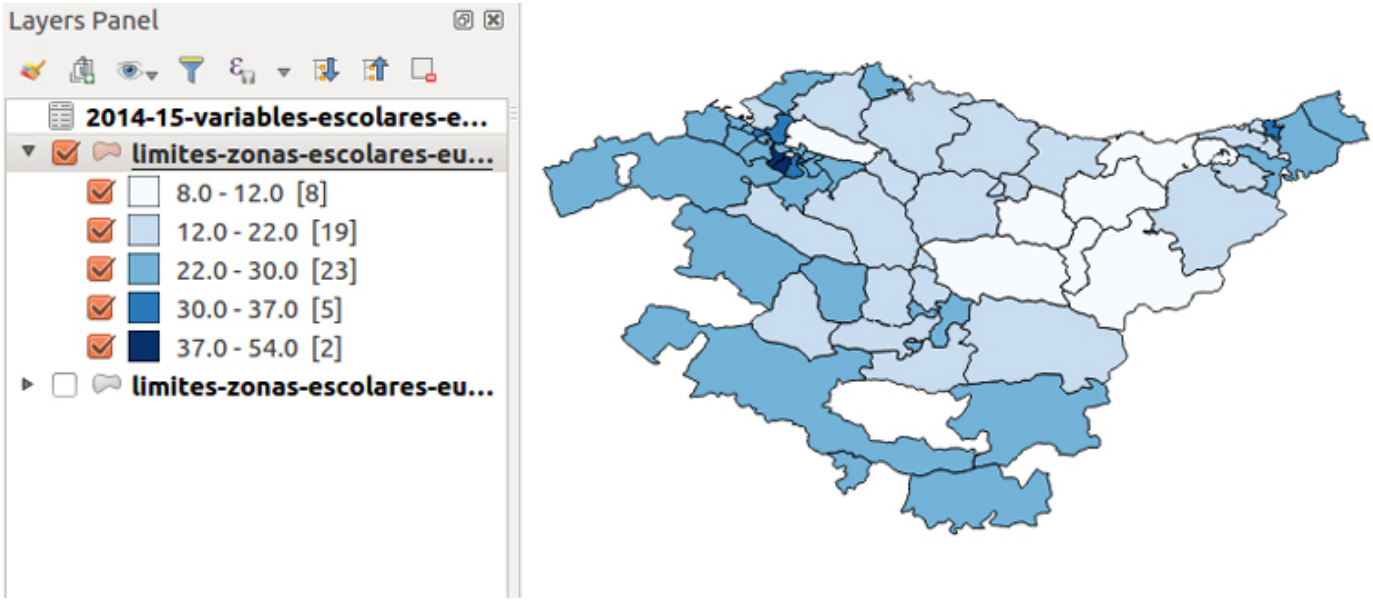
Per produir tots els mapes de manera automatitzada podem utilitzar llibreria de Javascript D3js.

5) Seguint amb els passos anteriors, ara és el moment de donar color a les àrees en funció de la variable que vulguem. Ho podem fer des de `Properties > Style`. Selecciona `Graduated` i selecciona la variable a la columna. En aquest cas, tria el percentatge d'alumnes amb beca de material escolar (`perc_bec_mat_escolar`) i fes clic a `Classify` per crear el rangs que assignen els colors. A `Legend Format`, tria `Precision 0` per controlar el nombre de decimals que apareixen a la llegenda. Des de `Mode` pots triar la manera de partir dels rangs. Deixa-ho en `Equal Interval`. Pots veure com es fa aquesta partició utilitzant la pestanya `Histogram`.

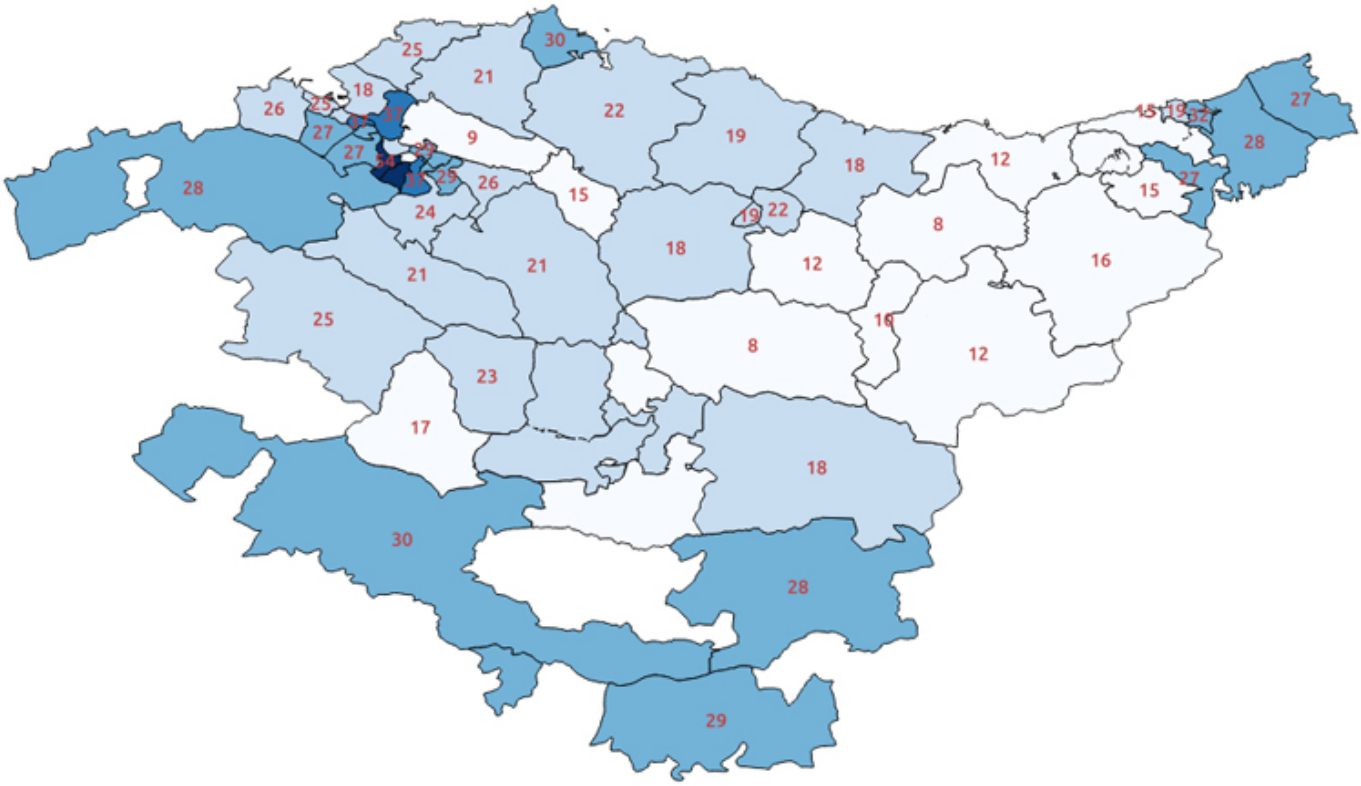


Si fas doble clic en els valors de la llegenda, pots canviar-los.

Ara ja tens el mapa de coropletes, i a la llegenda pots veure els valors associats a cadascun dels colors.

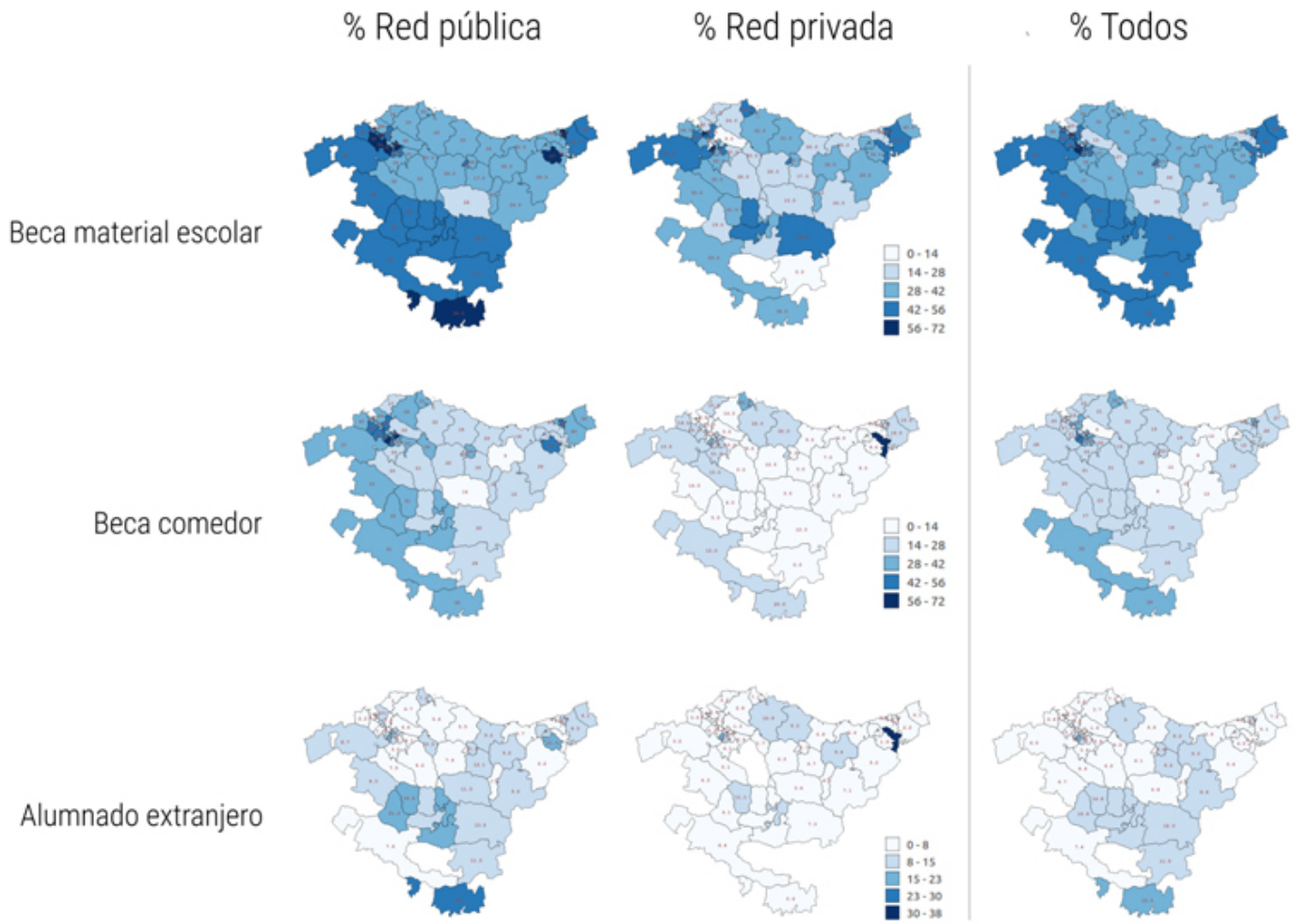


6) A la pestanya `Properties > Label` pots afegir un text a cada contorn (per exemple, el valor representat pel color).



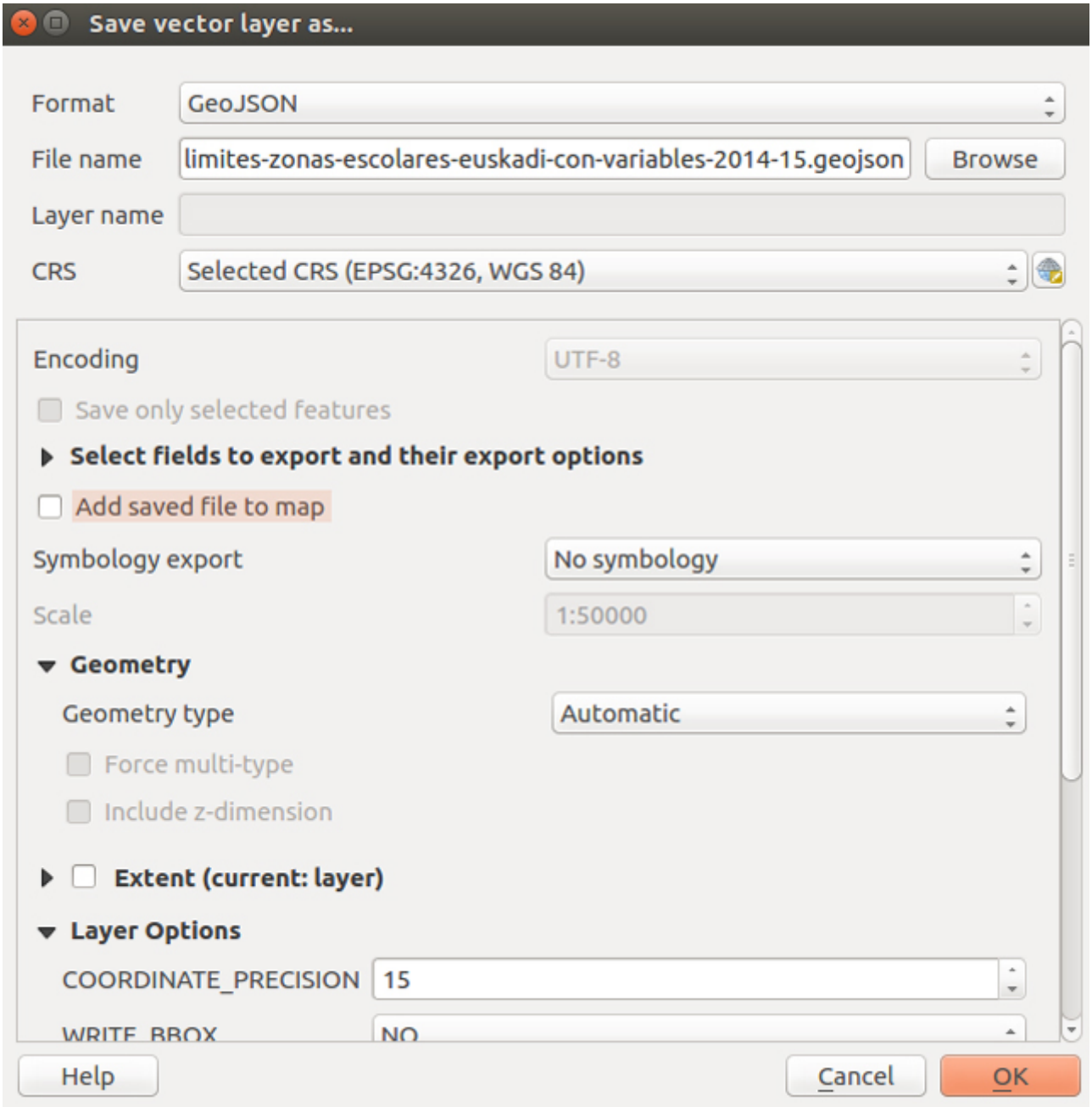
Si canviem la variable que utilitzem per representar a la capa `Properties > Style > Column`, podem generar un mapa per a cada variable.

7) Des de QGIS podem exportar els mapes generats a una imatge (PNG) o un vector (PDF, SVG). Això ens permet crear diversos mapes, escales i llegendes. Aquí tens un tutorial per fer un mapa amb múltiples vistes. La següent composició està feta mitjançant captures de pantalla de QGIS i maquetada amb Inkscape:



Aquest tipus de visualització (que en anglès s'anomena *small multiple*, va ser popularitzat per Edward Tufte i compara diferents variables d'una sèrie de dades sobre gràfics de la mateixa escala) ja és suficient per començar a veure variacions espacials dels indicadors entre la xarxa pública i la privada. A partir d'aquí poden sorgir noves preguntes.

8) Abans d'acabar amb QGIS cal que guardem el projecte i la capa en format geojson. Per fer-ho, fes clic amb el botó dret sobre el nom de la capa a la llista de capes i selecciona `Save vector layer as`.



Si vols, aquí pots baixar l'arxiu .geojson dels contorns de les zones escolars amb totes les variables associades.

Generar un mapa o uns mapes amb D3

Per anar un pas més enllà quant a l'automatització del procés de generació dels gràfics, utilitzarem la llibreria de Javascript D3. És relativament senzilla d'utilitzar i requereix poc codi. A més, podrem crear gràfics interactius per millorar l'experiència d'usuari.

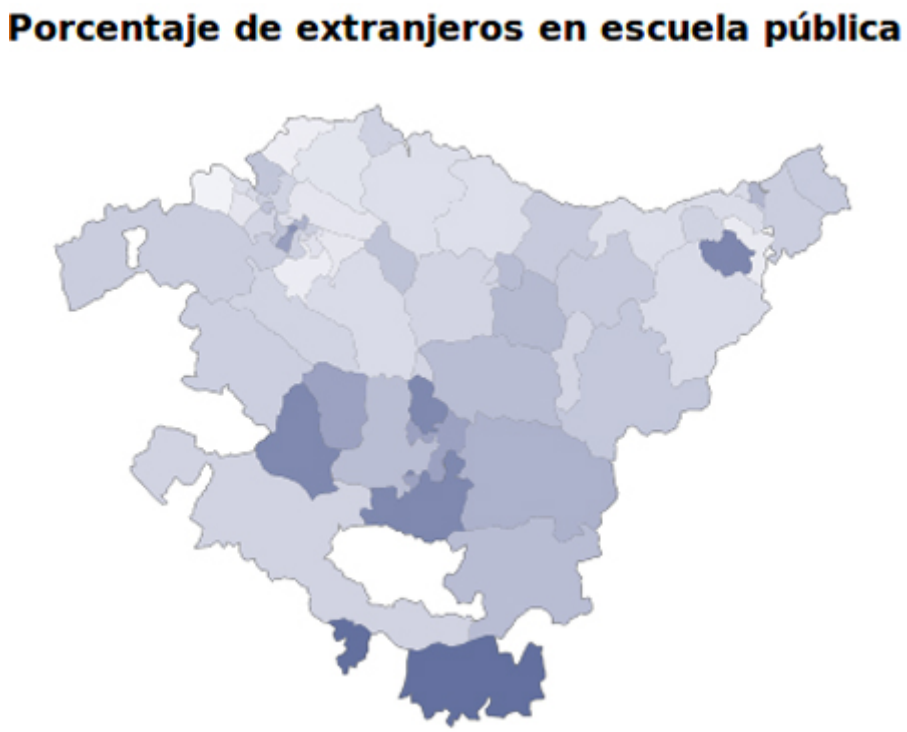
1) El primer que hem de fer és **simplificar les línies que componen** el fitxer dels contorns per alleugerir la seva grandària (pesa més de 7MB) i accelerar el període de càrrega de la visualització. També convertirem el fitxer .geojson en un fitxer de tipus **topoJSON**. Aquest pas redueix la grandària dels arxius fins a un 80% sense reduir el detall dels contorns.

Es pot convertir el fitxer en un fitxer topoJSON amb el paquet d'NPM topoJSON des de la línia de comandos, o bé utilitzar l'eina en línia Mapshaper. Mapshaper permet pujar fitxers geoJSON, corregir-hi errors, simplificar-ne el traçat i guardar-los en format topoJSON amb una interfície gràfica.

El fitxer resultant és aquest: https://github.com/montera34/segregacion/blob/master/data/output/limites-zonas-escolares-euskadi-con-variables-2014-15_simplify3.json.

2) Produir un mapa de coropletes d'una sola variable.

A partir del fitxer topoJSON, que ja conté els contorns de les zones escolars i les dades associades, és relativament senzill mostrar les dades en un mapa. Descarrega el .zip amb els fitxers, descomprimeix-lo i obre el fitxer `index.html` des d'un navegador web. Hauries de veure això:



Nota per Windows

Perquè D3 funcioni en local amb el sistema operatiu Windows, és necessari executar això a la línia d'ordres (com recomanen en aquesta resposta d'Stackoverflow):

```
> python -m SimpleHTTPServer 8888
> http://localhost

> python3 -m http.server 8888
> http://localhost
```

Recomanem utilitzar el navegador Firefox.

Ara explicarem les parts més importants del codi que hi ha a `index.html` i que generen el gràfic. Tant si tens coneixements d'HTML , CSS i JavaScript com si no, pensem que podràs entendre part del codi i manipular certes variables per modificar-ne la grandària, el color i la variable representada.

Primer, crida les llibreries de JavaScript D3js i Topjson per fer funcionar la visualització:

```
<script src="https://d3js.org/d3.v4.min.js"></script>
<script src="https://d3js.org/topojson.v1.min.js"></script>
```

Definim les dimensions del mapa en píxels:

```
var width = 400, height = 350;
```

Afegim l'SVG on es dibuixarà el mapa:

```
var svg = d3.select("body").append("svg")
  .attr("width", width)
  .attr("height", height)
  .style("background-color","#EFEFEF")
```

Definim l'escala de color:

```
var color = d3.scale.linear() // seleccionem una escala lineal
                 .domain([0, 38]) // valors mínim i màxim
                 .range(['#fff','#053874']) //color per al mínim és blanc (#FFF) i blau
(#053874) per al màxim
```

Selecciona el tipus de projecció del mapa (Mercator), el nivell de zoom i el centre del mapa:

```
var projection = d3.geoMercator() // projecció mercator
         .scale(12000) // zoom level
         .center([-2.6,42.9]) // center of map
         .translate([width / 2, height / 2]);
```

D'altra banda, cridem el fitxer topoJSON per utilitzar les dades que conté:

```
d3.json("data/output/limites-zonas-escolares-euskadi-con-variables-2014-
15_simplify3.json", function(error, euskadi) {
```

Creem una variable que contingui totes les zones. Per fer-ho, accedim al fitxer JSON que hem anomenat «euskadi» a la línia anterior i accedim a `objects` i `zonas`.

```
var subunits = topojson.feature(euskadi, euskadi.objects.zonas);
```

Ara falta dibuixar el mapa:

```
svg.append("path")
  .datum(subunits)
  .attr("d", path)
  .attr("stroke", "grey");
```

Dibuixa els contorns de les zones i acoloreix-les (`atributo fill`) segons una variable (`perc_alum_ext_publi`):

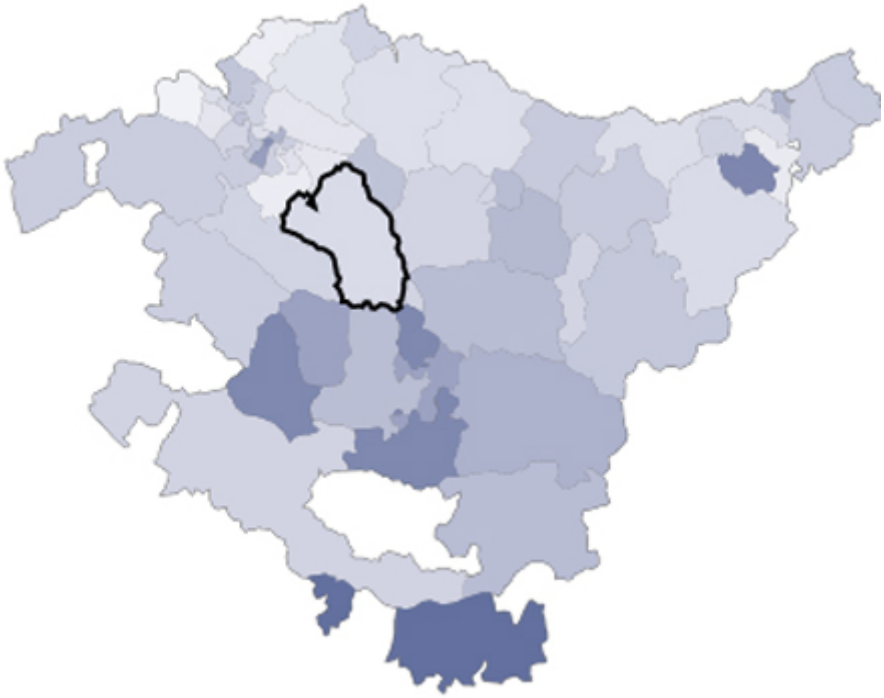
```
svg.selectAll(".subunit")
     .data(subunits.features)
```

```
    .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); }) //
acoloreix les àrees segons la variable "perc_alum_ext_publi"
     .attr("d", path)
```

Si canviem la variable `perc_alum_ext_publi` per una altra, com ara `perc_alum_ext_priv`, acolorirem el mapa segons un altre indicador. Inclou la variable en la línia `.attr("fill", function(d) { return olor(d.properties.perc_alum_ext_priv); }).` Atenció: hauràs de canviar el domini de l'escala `.domain([0, 38])` si vols canviar els seus valors màxims i mínims. Els valors que quedin fora del domini amb els contorns de color negre (per fer la prova, canvieu el domini de `[0, 38]` a `[0, 4]`.

Per afegir-hi interactivitat, cal afegir tot allò que està indicat en vermell:

```
// 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); }) //
acoloreix les àrees segons la variable "perc_alum_ext_publi"
  .attr("d", path)
   .on("mousemove", showTooltip) // AFEGIR ESDEVENIMENT SHOW TOOLTIP
   .on("mouseout", hideTooltip); // OCULTAR TOOLTIP

// Funció per 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)
  }
  // Funció per ocultar el tooltip
  function hideTooltip(d) {
    tooltip.style("opacity", 0)
  }
```

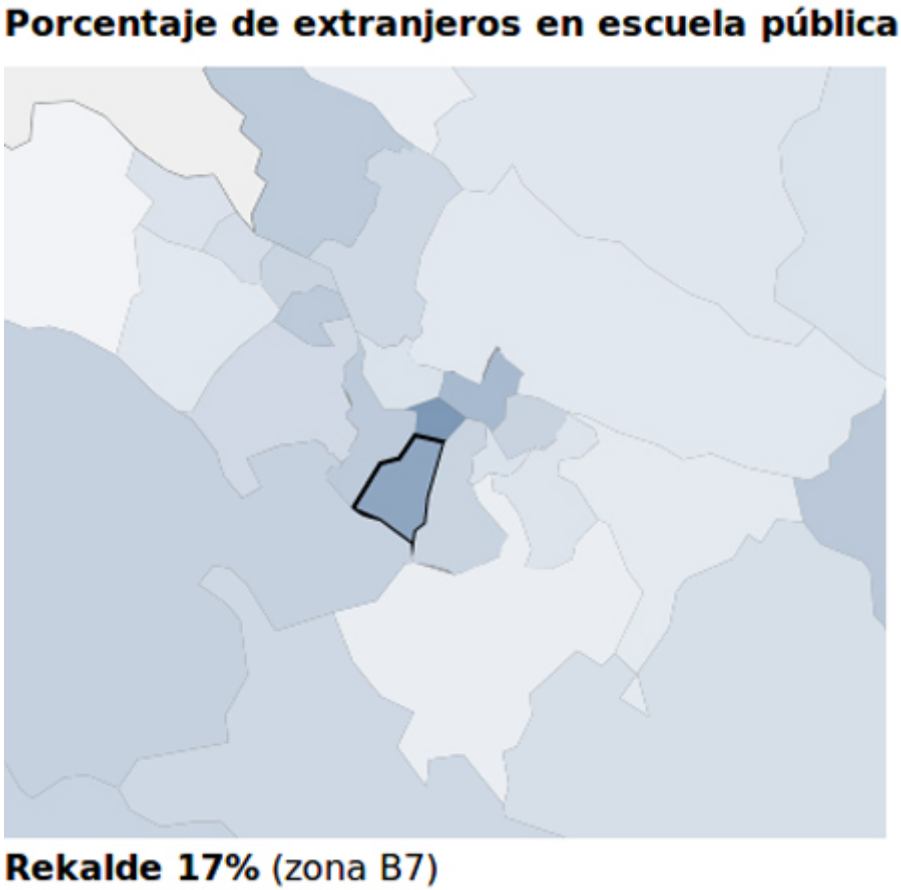


Quan passem el cursor per sobre, se'ns indica que aquesta zona és Igorre, amb un 6,2% d'alumnat estranger a la xarxa pública.

Si es canvia el zoom i el centre del mapa, per exemple, per a Bilbao:

```
var projection = d3.geoMercator()
 .scale(60000) // note change of zoom
 .center([-2.9331,43.2521])
 .translate([mapWidth / 2, mapHeight / 2]);
```

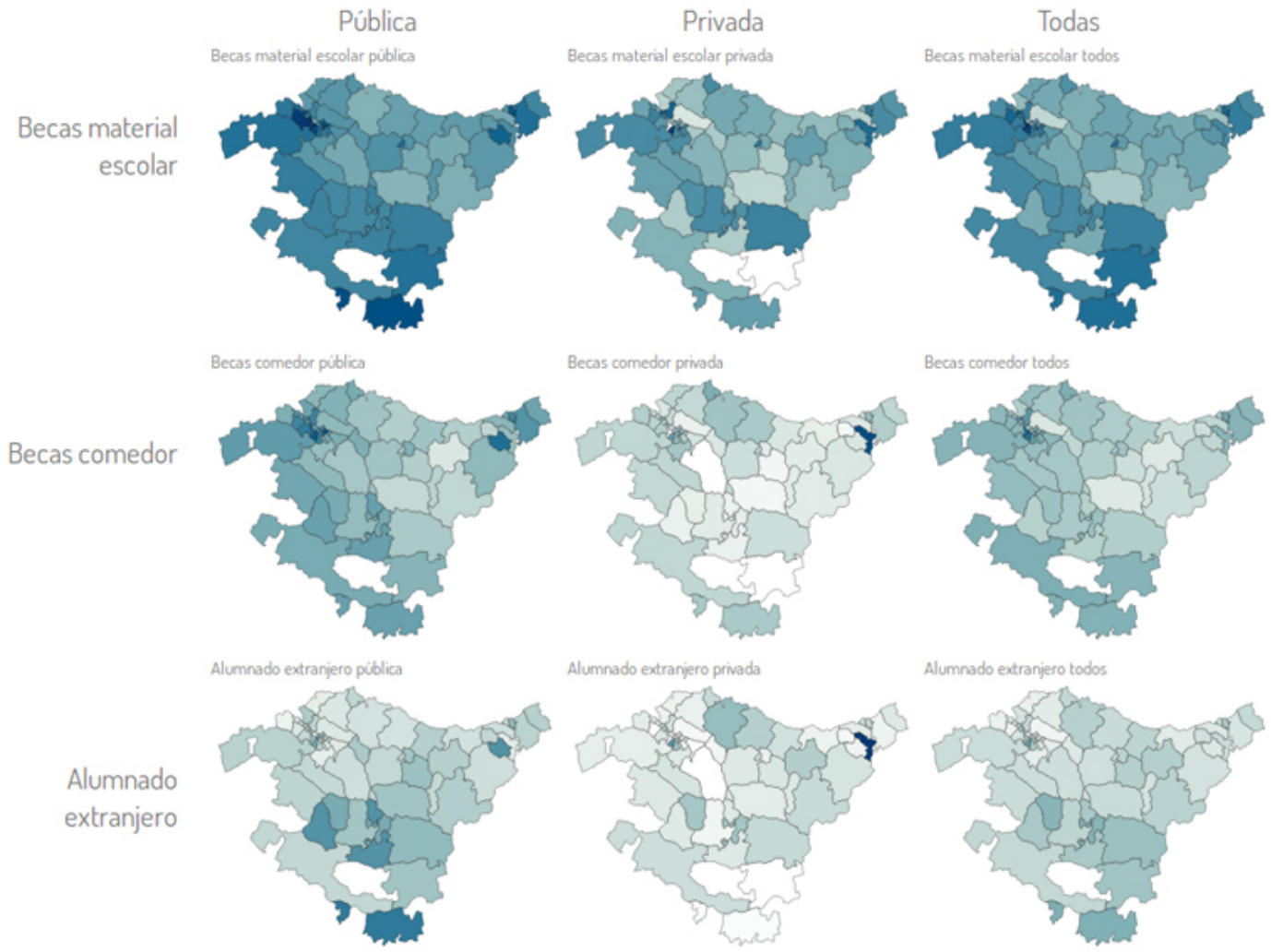
Es mostra un zoom de la zona de Bilbao:



Pots veure el codi complet i la visualització aquí: http://bl.ocks.org/numeroteca/33325b3234175c5450b856026c62a920 y https://gist.github.com/numeroteca/5764a860ecc0f7c4d421b0ad00bd6188.

3) Generar múltiples mapes interactius, cadascun amb la seva pròpia variable.

Per generar múltiples mapes hem de replicar el mapa del pas anterior. Baixa el .zip amb els fitxers, descomprimeix-lo i obre el fitxer `index.html` des d'un navegador web. Hauries de veure això:



Pots veure el codi complet i la visualització.

Primer, hem de definir dues escales de color diferents, perquè els dominis (el valor mínim i el valor màxim de cada variable) són diferents:

```
// Escala de color per a beques
var color = d3.scaleLinear().domain([0, 72]).range(['#fff','#053874'])
// Escala de color estrangers
var color2 = d3.scaleLinear().domain([0, 38]).range(['#fff','#053874'])
```

En una matriu, definim les variables que volem que apareguin:

```
// Array to list variables to display in map and display name of variables
var mapasdef = [
  ["Beques material escolar pública","perc_bec_mat_escolar_pub"],
  ["Beques material escolar privada","perc_bec_mat_escolar_priv"],
  ["Beques material escolar tots","perc_material_m15"],
  ["Beques menjador pública","perc_bec_comedor_pub"],
  ["Beques menjador privada","perc_bec_comedor_priv"],
  ["Beques menjador tots ","perc_comedor_m16"],
  ["Alumnat estranger pública","perc_alum_ext_publi"],
  ["Alumnat estranger privada","perc_alum_ext_priv"],
  ["Alumnat estranger tots ","perc_alum_ext_todos"]
]
```

Crea un div (contenidor) per a cada element de la matriu que acabem 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 els contenidors `párrafo` (p) i `span`, que allotjaran el nom de la variable que el mapa representa i el número:

```
divs.append('p').text(function(d){ return d[0]; });
divs.select('p').append('span').text(" ")
```

Afegeix un contenidor SVG per a cada mapa:

```
var zonas = divs.append('svg').attr('width',mapWidth).attr('height',mapHeight);
```

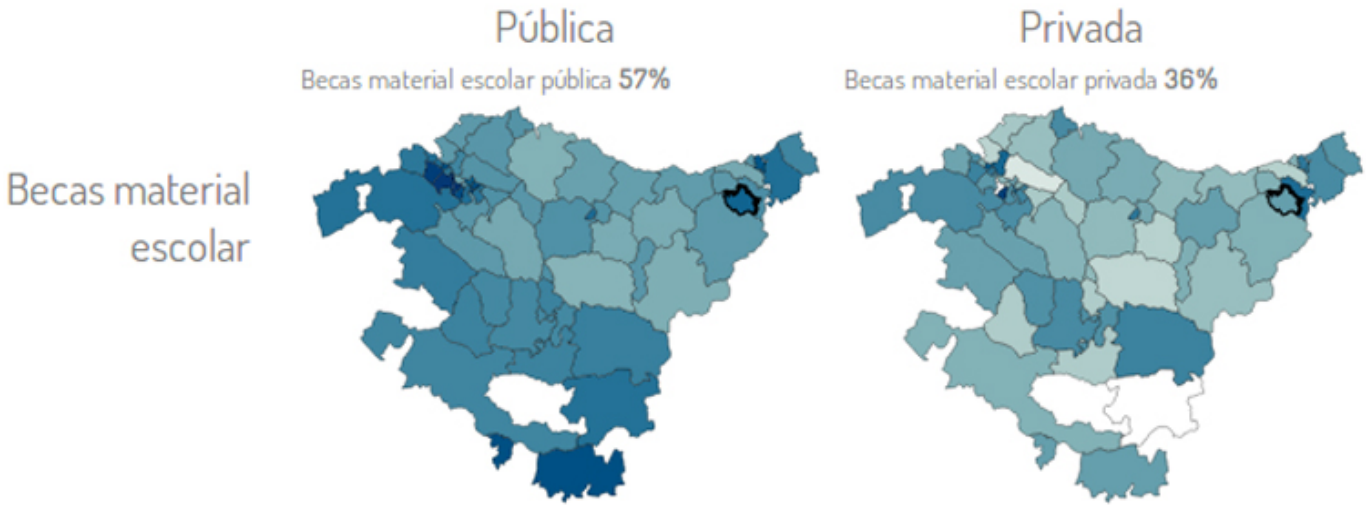
Insereix a cadascun dels SVG que hem creat en la línia anterior els contorns de totes les zones i els hi aplica un color segons l'escala de color que hem definit.

```
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 afegim el codi que hi ha a continuació aconseguirem que s'escrigui el percentatge a la capçalera 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");
})
```

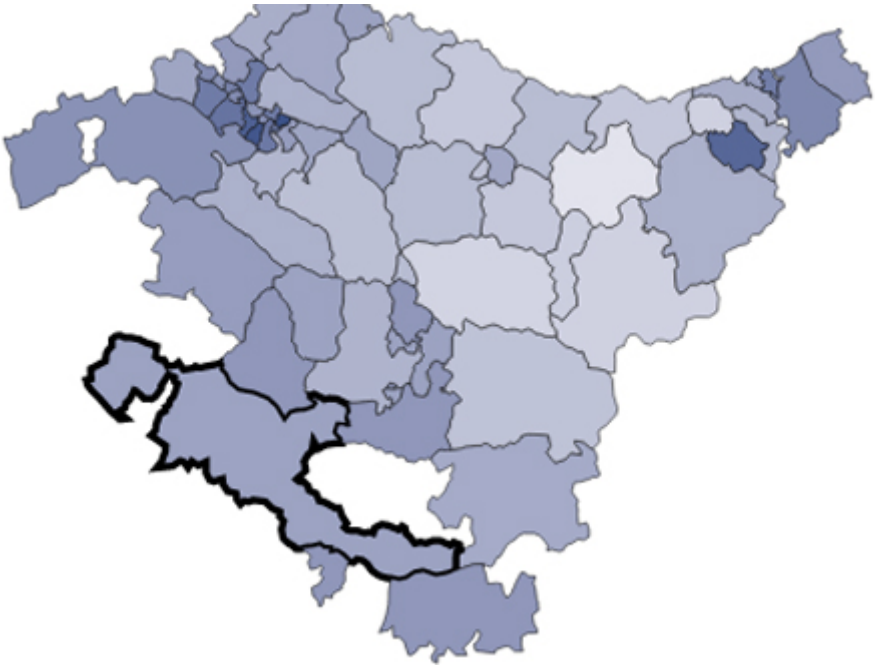
```
.on('select', function(self) {
  var geoData = self.data(); // creates variable with data contained in the zone (zona
escolar) where the cursor is on
  //looks for the "span" element in the div and writes inside the percentage of that
variable
  self.node().parentNode.parentNode.getElementsByTagName('span')[0].innerHTML = " " +
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 funció utilitzada per "notificar" als otros path que comparteixen class amb el que
està assenyalat pel cursor
  function notify(selector, eventName) {
        d3.selectAll(selector)._groups[0].forEach(function(el, i) {
        var shape = d3.select(el);
        shape.on(eventName)(shape);
        });
  }
```



Tot i que el codi de D3 pugui resultar complicat si no estàs familiaritzat amb JavaScript i HTML, pots intentar modificar certs elements per aconseguir els canvis desitjats: un canvi de la grandària del mapa, un canvi del nivell de zoom i un canvi dels colors. Cal fer proves.

Altres formes de representar: els mapes no són l'única opció

Els mapes solen ser l'opció preferida per representar regions, però no tenen perquè ser sempre la millor opció. S'ha escrit molta teoria sobre aquest tema, però n'hi ha prou amb esmentar que per a l'ull és difícil distingir les variacions de color, i la possibilitat de fer-ho depèn molt dels colors del voltant per avaluar com de fosc o de clar és un color determinat. A més, les regions amb poc alumnat però molt extenses queden sobrerepresentades en el mapa. És el cas, per exemple, de la zona escolar «Álava occidental», una de les més extenses però la segona amb menys alumnat (752).

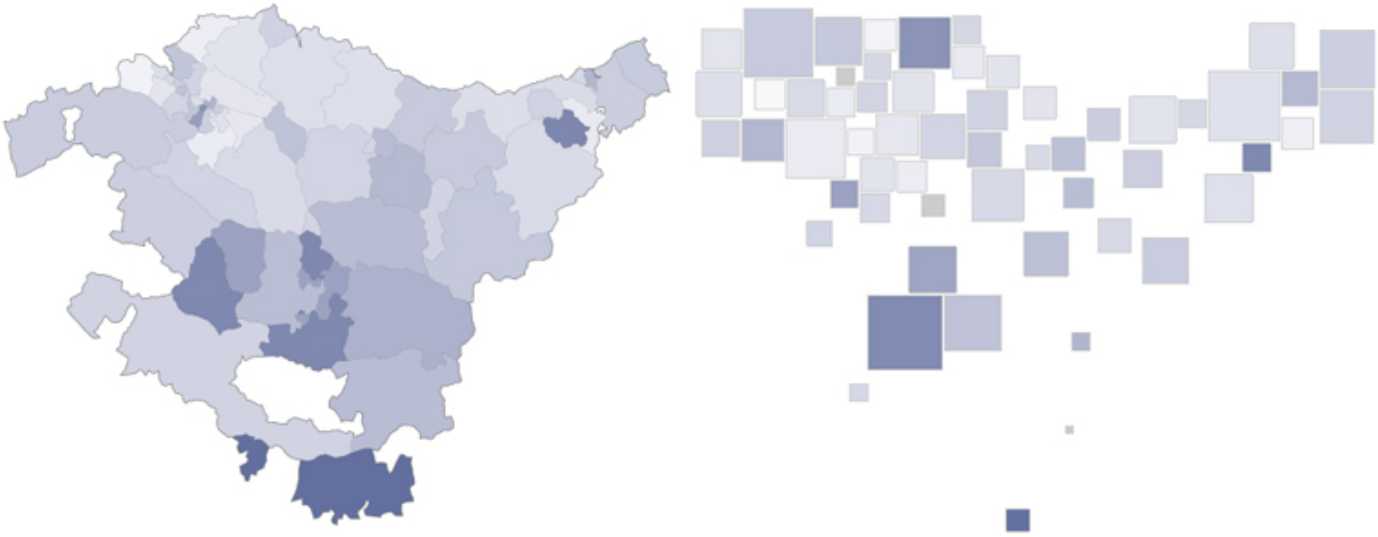


Es marca la zona escolar Álava occidental.

Visualment, tindrà molta més presencia en el mapa que algunes zones de Bilbao, com ara Abando, que té 6.054 alumnes, però és gairebé invisible degut a la seva grandària en el mapa.

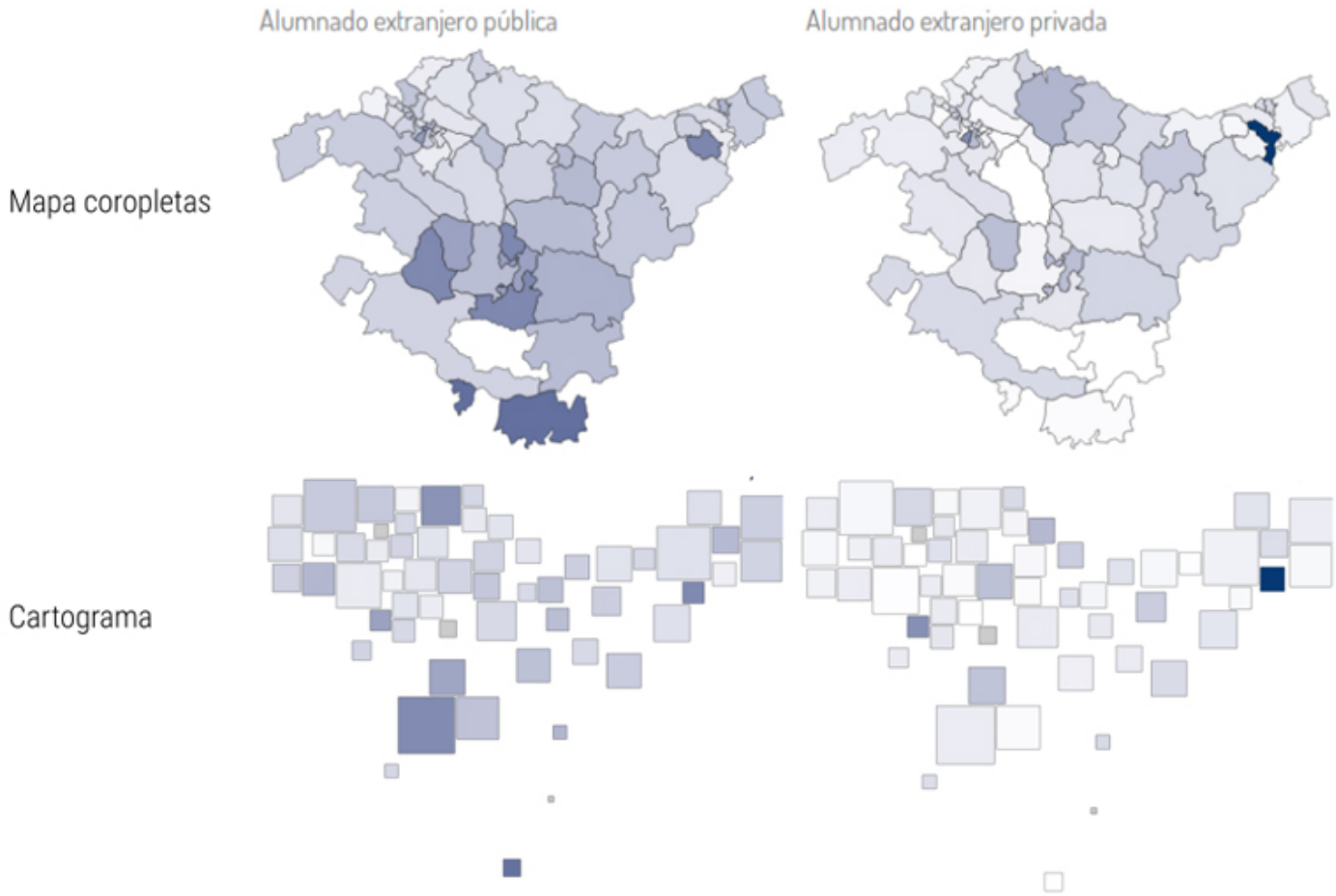
Cartogrames

Per corregir aquest tipus de desigualtats visuals existeixen un altre tipus de representacions: els cartogrames. En aquest tipus de cartograma es representa cada zona escolar amb un quadrat. L'àrea d'aquest quadrat és proporcional a l'alumnat que té la zona, i el seu color varia en funció, en aquest cas, del percentatge d'una de les variables, com en els mapes anteriors. Els cartogrames, però, tenen la dificultat de mantenir la forma geomètrica de les zones. Això fa que el mapa perdi la seva llegibilitat geogràfica, la qual cosa pot desorientar a qui vegi la visualització. L'objectiu d'aquest tipus de cartogrames (ja que hi ha cartogrames contigus que intenten deformar al menys possible les regions) és ser el més



Mapa de coropletes (esquerra). Cartograma (dreta; la grandària de les zones ve determinada per la quantitat d'alumnat)

No hi ha un mètode bo ni un de dolent; cadascun d'ells serveix per explorar les dades d'una manera diferent. Aquí repetim la comparativa entre la xarxa pública i la xarxa privada amb mapes de coropletes i amb un cartograma.



La visualització i el codi (basat en cartogrames de Martín González) per realitzar aquests cartogrames estan disponibles. Degut a la seva extensió, no comentarem el codi, però us animem a utilitzar-lo per generar els vostres propis cartogrames.

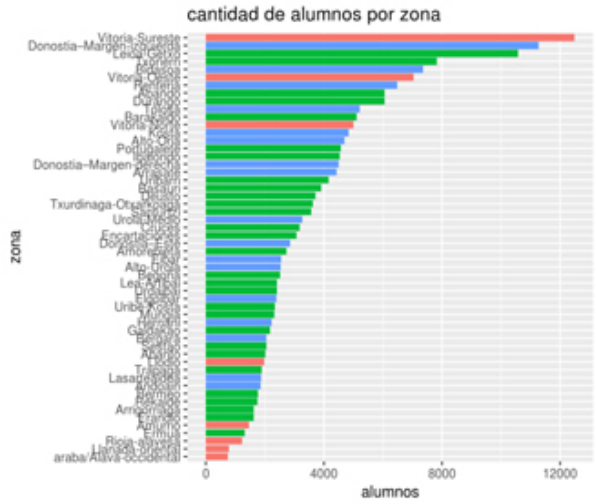
Explorar les dades i les seves correlacions amb R

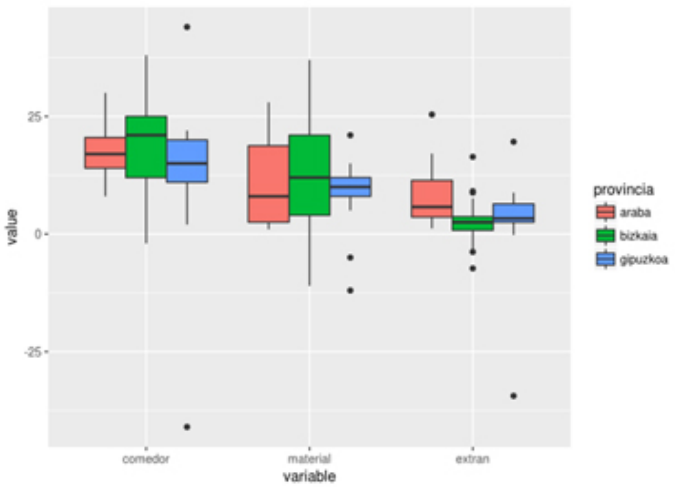
Després de realitzar l'exploració espacial de les dades (és a dir, d'establir les diferències entre les xarxes en l'àmbit geogràfic), proposem explorar les diferències i les relacions entre els indicadors d'altres formes.

Suggerim l'ús d'R amb RStudio (RStudio) per fer un primer apropament a les dades.

1) Preanàlisis

Pots descarregar el fitxer amb aquestes preanàlisis a segregacion-analisis.Rmd per executar-lo amb TStudio, o bé pots baixar directament el fitxer segregacion-analisis.html, que no necessita R per funcionar, i obrir-lo amb un navegador.





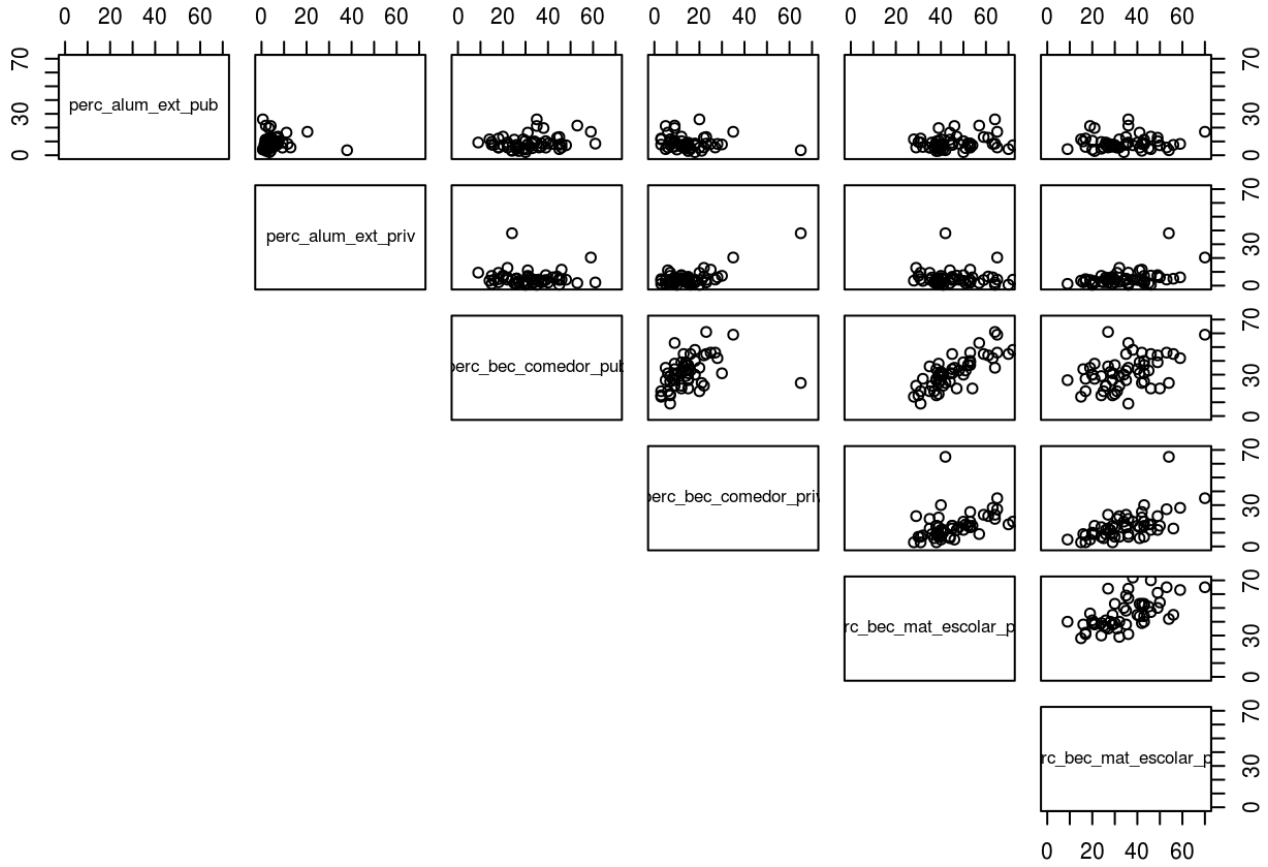
També pots accedir a aquesta preanàlisi en línia.

2) Correlacions

Ara explorarem si existeixen correlacions entre les variables estudiades: % beques de material escolar i menjador, % de presència d'alumnat estranger a les xarxes pública i privada dins de cadascuna de les zones escolars. El codi i les visualitzacions estan disponibles.

Tenim aquesta primera visualització amb `pairs`, que conté molt poc codi:

```
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_be

        xlim=c(0,70),ylim = c(0,70),
        lower.panel = abline(a = 0,b = 1)
```



S'hi aprecien correlacions a nivell visual entre els indicadors:

- Beques material escolar i Beques menjador a xarxa pública.
- Beques material escolar i Beques menjador a xarxa privada.

A més de millorar la visualització, la llibreria GGally calcula les correlacions per a totes les zones i per a cadascuna de les províncies, i genera histogrames i diagrames de dispersió separant cadascuna de les zones escolars per colors, depenent de la província:

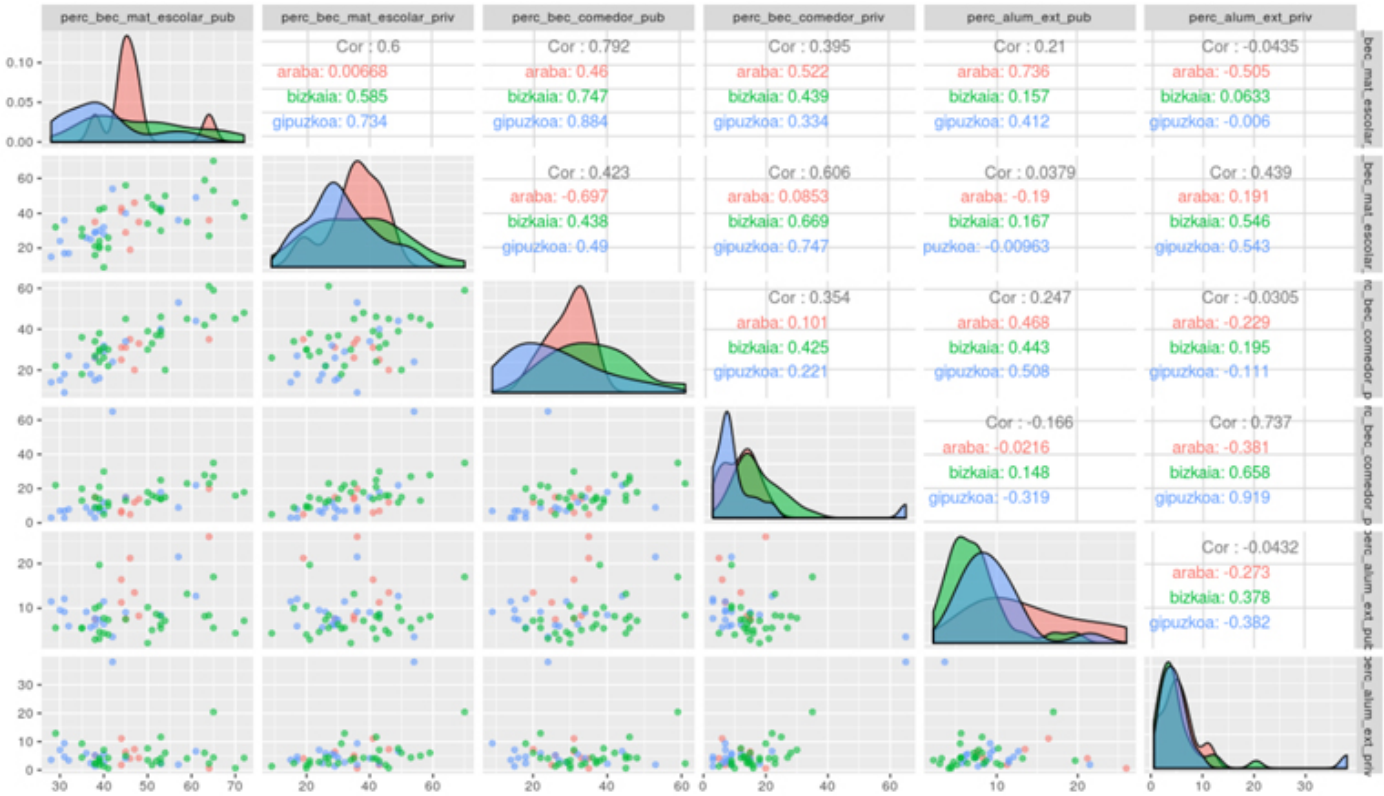
```
library(GGally)
ggpairs(segreg,
  columns = c(10,11,13,14,17,18),
```

```
    title = "buscando correlaciones",
    aes(colour = provincia, alpha = 0.4))
```

Amb tot i això, abans de computar correlacions és millor deixar fora les zones escolars de Montaña alavesa, Igorre i Basurto-Zorroza, atès que no tenen oferta privada:

```
segreg = segreg[c(-6,-33,-57),] #elimina els que no tenen xarxa privada
attach(segreglimpio)
```

I tornem a calcular:



Les correlacions més fortes entre les variables una vegada calculades són:

- Beques material escolar i Beques menjador en xarxa pública **0,792** (Bizkaia 0,747, Gipuzkoa **0,884**, Araba 0,46).
- Beques menjador i Alumnat estranger a xarxa privada **0,737** (Bizkaia 0,658, Gipuzkoa **0,919**, Araba -0,381).
- Beques material escolar i Beques menjador a xarxa privada **0,606** (Bizkaia 0,669, Gipuzkoa **0,747**, Araba 0,085).

Així, els valors més alts són sempre els de la província de Gipuzkoa.

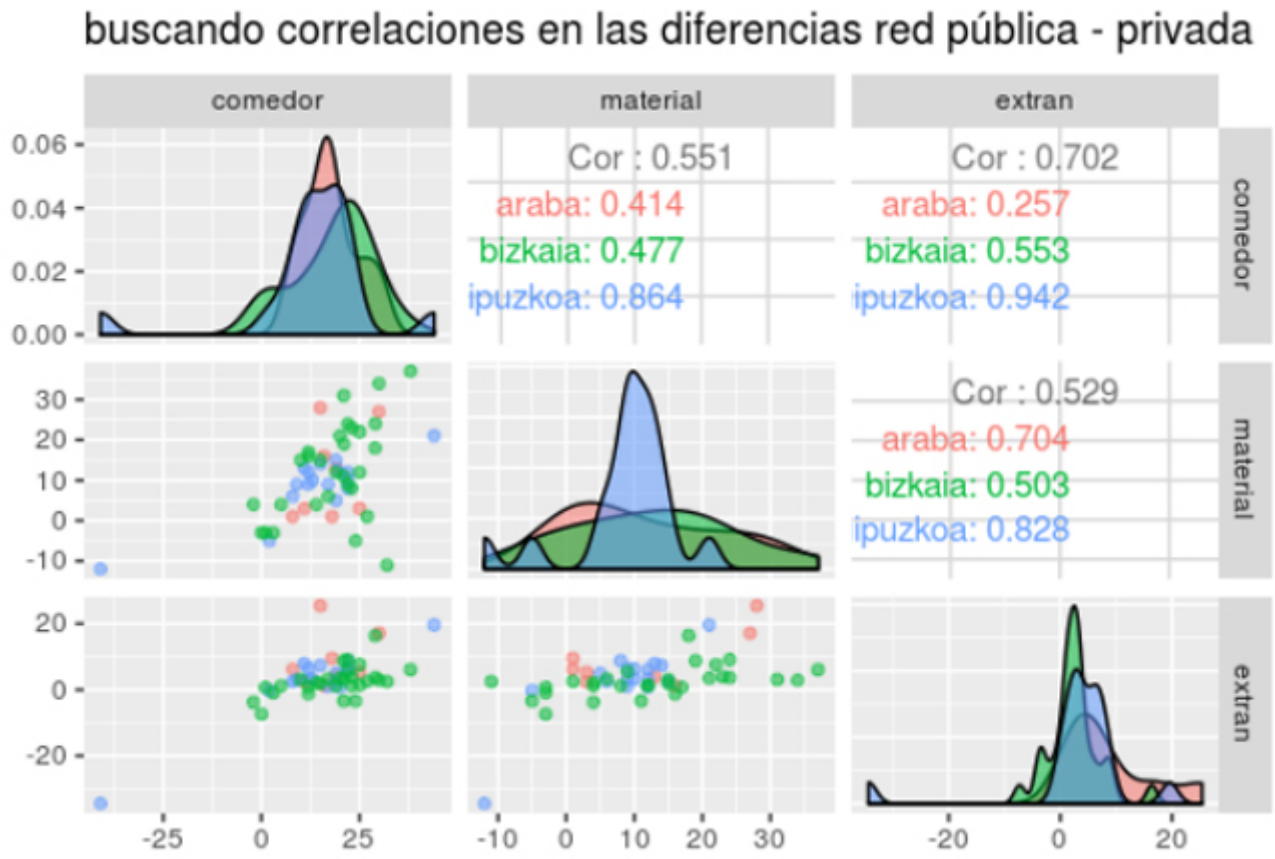
Aquesta correlació indica que hi ha una correlació relativament forta entre el percentatge d'alumnat becat en menjador i el percentatge becat amb material escolar a la xarxa pública. Això reforça la idea que les beques són una mesura del nivell socioeconòmic i indica certa homogeneïtat dins de cada zona escolar, que és major a la xarxa pública (0,792) que a la privada (0,606). També hi ha una forta correlació entre el percentatge de becaris de menjador i l'alumnat estranger a la xarxa privada (0,737). Què ens indica aquesta correlació? Que la presència d'alumnat estranger és un indicador de nivell socioeconòmic? Aquesta correlació és molt més baixa (0,247) a la xarxa pública.

Per acabar aquesta exploració ens centrarem en veure si hi ha correlacions entre **les diferències entre la xarxa pública i la xarxa privada** per a cada indicador i zona. Per fer-ho, generem tres noves variables:

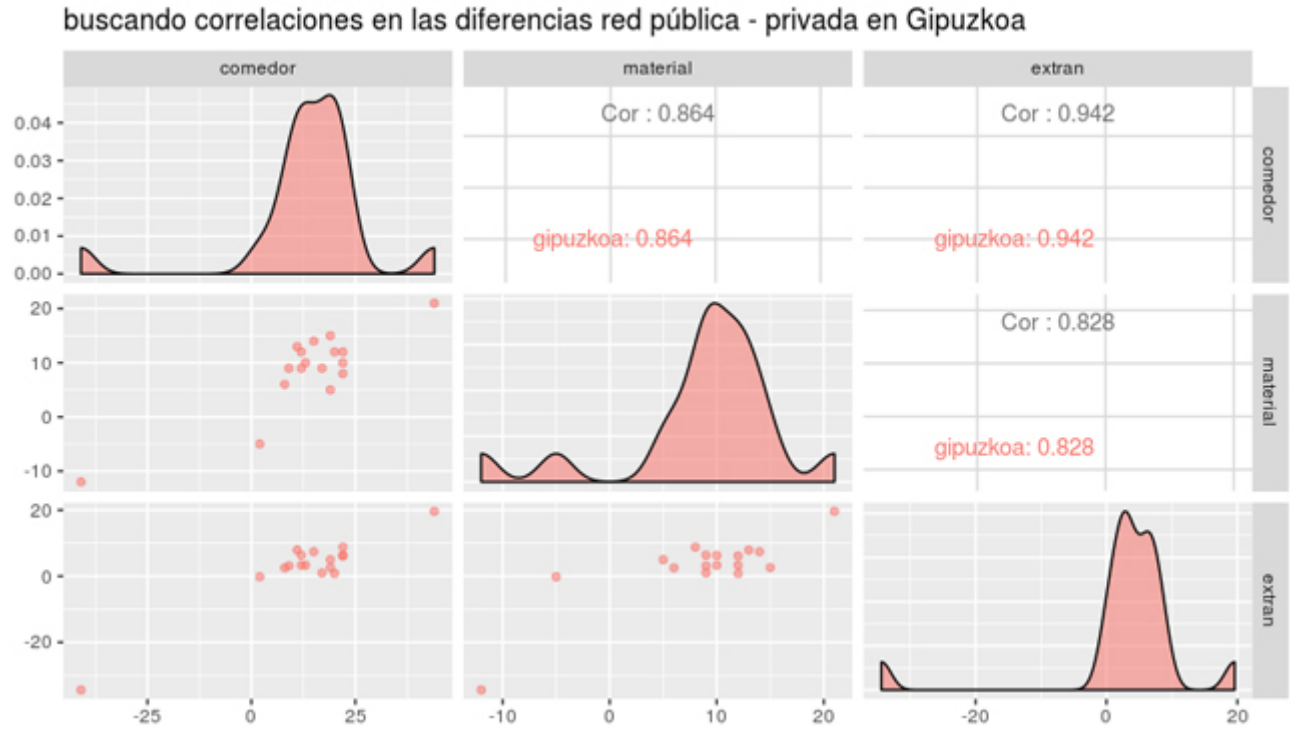
- menjador = % beca menjador a xarxa pública – % beca menjador a xarxa privada.
- material = % beca material escolar a xarxa pública – % beca material escolar a xarxa privada.

- extran = % alumnat estranger a xarxa pública – % alumnat estranger a xarxa 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)

segregdif = data.frame(comedor,material,extran,provincia,zona)
colnames(segregdif) = c("comedor","material","extran","provincia","zona")
ggpairs(segregdif,
        title = "buscant correlacions en les diferències xarxa pública - privada",
        columns = c(1:3),
        aes(colour = provincia, alpha = 0.4))
```



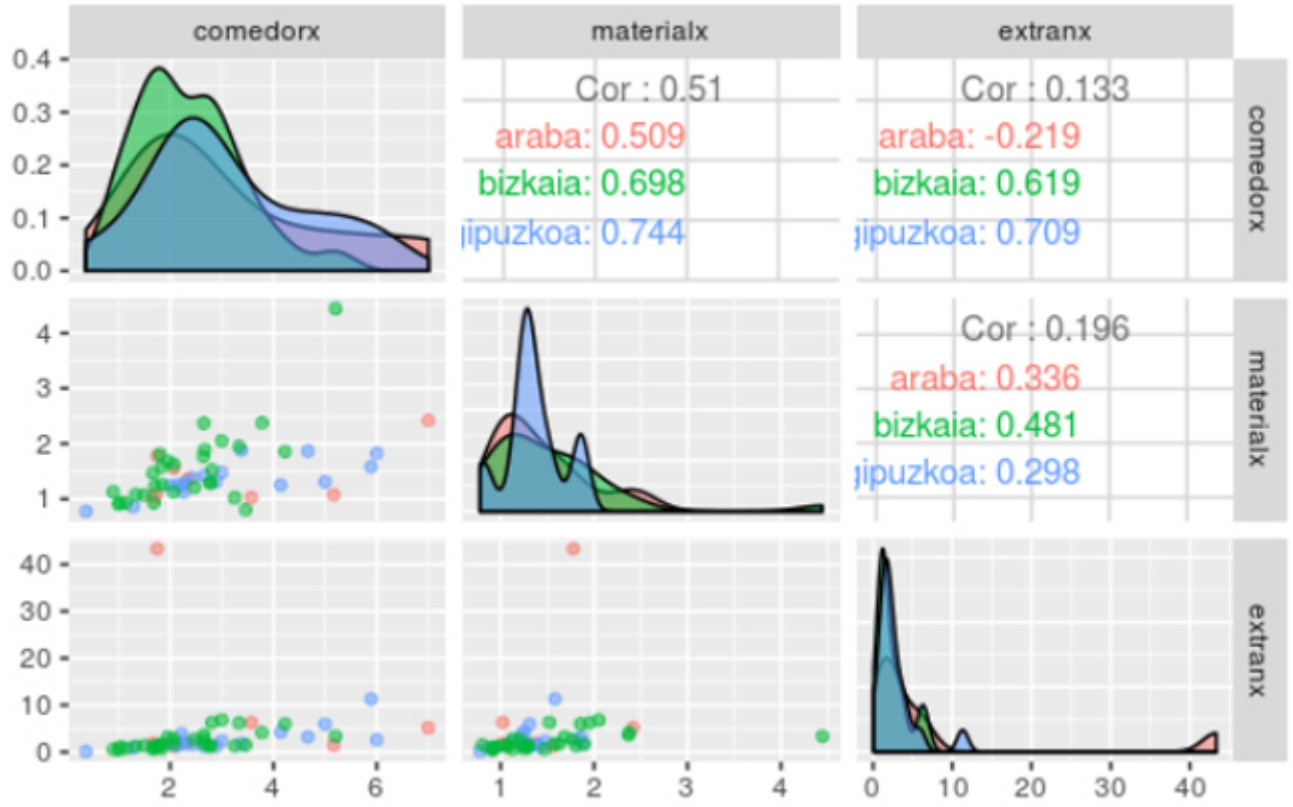
Obtenim valors moderadament forts (0,551, 0,702, 0,529), i criden l'atenció els valors molt alts de tots els indicadors de diferència (0,864, 0,942 i 0,828) a Gipuzkoa:



I si calculem els quocients entre les variables (de fet, l'informe recull aquest càlcul per calcular el grau de desequilibri):

- comedorx = % beca menjador a xarxa pública / % beca menjador a xarxa privada
- materialx = % beca material escolar a xarxa pública / % beca material escolar a xarxa privada
- extranx = % alumnat estranger a xarxa pública / % alumnat estranger a xarxa privada

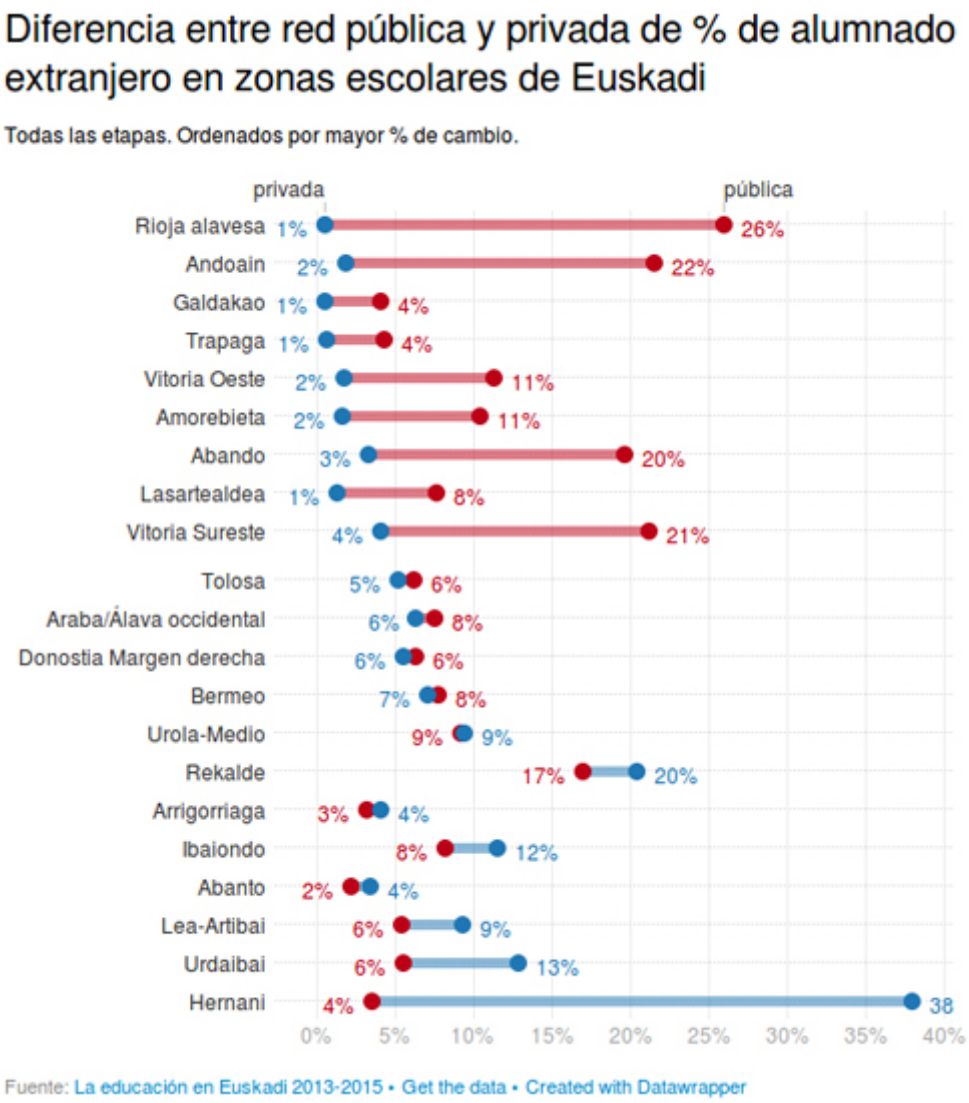
```
comedorx = data.frame(perc_bec_comedor_pub / perc_bec_comedor_priv)
materialx = data.frame(perc_bec_mat_escolar_pub / perc_bec_mat_escolar_priv)
extranx  = data.frame(perc_alum_ext_pub / perc_alum_ext_priv)

segregdif = data.frame(comedor,material,extran,comedorx,materialx,extranx,provincia,zona)
colnames(segregdif) =
c("comedor","material","extran","comedorx","materialx","extranx","provincia","zona")
ggpairs(segregdif,
        title = "buscant correlacions",
        columns = c(4:6),
        aes(colour = provincia, alpha = 0.4))
```



En general, no s'aprecien correlacions fortes (0,51, 0,133, 0,196), tot i que sí que n'hi ha entre els quocients de beques de material amb menjador i alumnat estranger amb menjador per a Gipuzkoa (0,744 i 0,709, respectivament) i Bizkaia (0,698 i 0,619, respectivament). És possible que aquestes dades indiquin que a Araba hi ha una història per explorar, ja que aquesta zona mostra la correlació més pobra (fins i tot és negativa) entre el quocient d'estrangers i beques de menjador. Quina història hi ha a les dades de Gipuzkoa?

Mostrar les diferències

Aquestes exploracions ens ajuden a entendre les dades, però és possible que siguin massa complexes com per publicar-les en un mitjà generalista. Ara que ha hem identificat que les **diferències públic-privat per a cada zona** són una de les variables a representar, suggerim el següent tipus de gràfic, que les representa d'una forma clara. Per generar-lo hem utilitzat Datawrapper (http://datawrapper.de/), una eina en línia que permet pujar les dades, escollir el tipus de visualització i inserir el gràfic allà on vulguem:



Pots veure el en datawrapper.

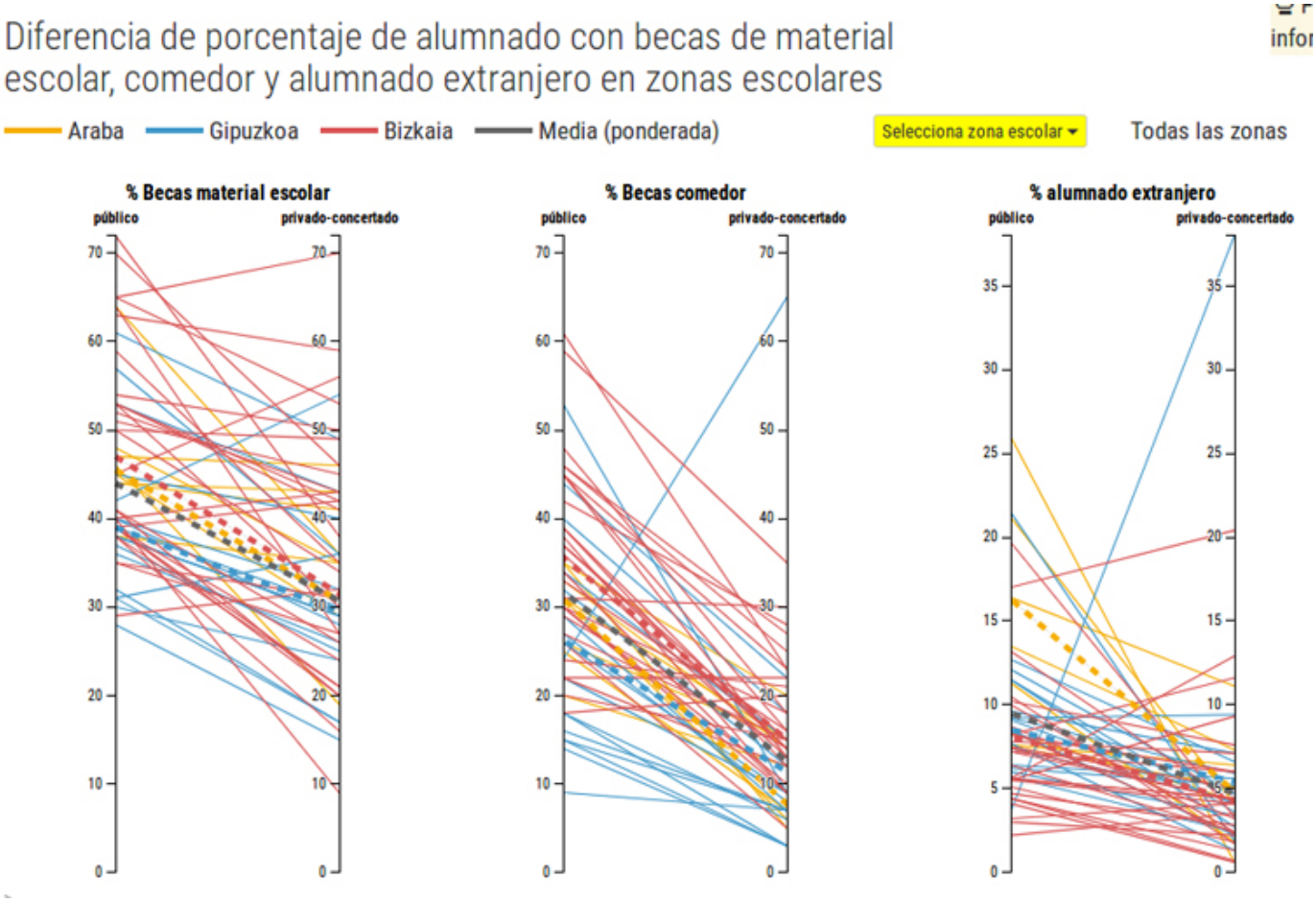
Aquest gràfic ens sembla molt interessant, ja que ofereix els valors de cada zona i fa èmfasi en la quantitat de la diferència. El color de la línia varia segons si el major percentatge és de la xarxa pública (vermell) o de la xarxa privada (blau).

e. Com puc presentar la informació per fer un relat intel·ligible?

Quan hagis recopilat i analitzat tots els gràfics realitzats, és moment de seleccionar els gràfics que vols utilitzar i pensar quina història vols explicar. És molt probable que cap dels gràfics et serveixi tal com és ara. Per tant, hauràs de modificar els gràfics o buscar noves variables. Aquest exercici ha intentat ser un exemple d'exploració que et permeti començar a fer noves preguntes.

Quines històries amaguen les diferències entre la xarxa pública i la privada? Què ocorre a la província d'Araba? Per què Hernani té uns valors tan diferents dels de la resta de zones escolars? Es pot investigar si hi ha una relació entre les zones escolars i el tipus de context (urbà o rural).

Finalment, aquesta altra visualització expressa la diferència amb el pendent de la línia que va del valor de l'esquerra (públic) al de la dreta (privat). Ho fa per a cada zona i per a cada indicador, i es basa en el tipus de gràfic de coordenades paral·leles:



Vegeu el gràfic en manera interactiva i accediu al codi.

Per a saber més sobre aquest gràfic de coordenades paral·leles pots llegir aquest article.

6. Tres exercicis de periodisme de dades

6.3. Educació pública-privada: obrint dades obertes

6.3.5. Informació complementària

- Cartogramas:
  - The State of the Art in Cartograms.
  - Evaluating Cartogram Effectiveness.
  - Algorithms for Cartogram Computatio

Crèdits

Periodisme de dades, coord Núria Ferran-Ferrer (2018). UOC. http://periodisme-dades.uoc.edu/

- 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/
- Cerrillo, A. (2018). "Aspectes jurídics del periodisme de dades". A Núria Ferran-Ferrer (coord.), *Periodisme de dades*. UOC http://periodisme-dades.uoc.edu/
- 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/
- Aragó, L. (2018). "Anàlisi de dades". A Núria Ferran-Ferrer (coord.), *Periodisme de dades*. UOC http://periodisme-dades.uoc.edu/

Autors:

Laura Aznar Llucià. Sociòloga i periodista. Redactora a Crític. @LauraAzLlu

Articles: http://www.elcritic.cat/?s=laura+aznar&lang=ca

Agustí Cerrillo. Catedràtic de Dret administratiu. Universitat Oberta de Catalunya

Laura Aragó. Periodista de dades, investigació i visualització. La Vanguardia. @arago_la

Articles: http://www.lavanguardia.com/autores/laura-arago.html

Coordinació:

Núria Ferran-Ferrer. Professora als Estudis de Ciències de la Informació i la Comunicació a la Universitat Oberta de Catalunya. @nferranf

(*) Contingut disponible només en web