

Guia de visualització de dades



Generalitat de Catalunya

Avís legal:



Aquesta obra està subjecta a una llicència Reconeixement 4.0 de Creative Commons. Se'n permet la reproducció, la distribució, la comunicació pública i la transformació per generar una obra derivada, sense cap restricció sempre que se'n citi el titular dels drets (Generalitat de Catalunya). La llicència completa es pot consultar a:

<https://creativecommons.org/licenses/by/4.0>

Primera edició: novembre de 2018

ISBN: 978-84-393-9734-2

Guia de visualització de dades

1. Introducció a la guia	5
2. Què és la visualització de dades	10
2.1. Visualitzacions per explorar	12
2.2. Visualitzacions per analitzar	13
2.2.1. Quadre de comandament	14
2.2.2. Aplicació d'analítica visual	14
2.3. Visualitzacions per explicar	15
2.3.1. Infografia	16
2.3.2. Narrativa per desplaçament	17
2.3.3. Presentació	17
2.3.4. Vídeo	20
3. Metodologia	21
3.1. Fase 1 - Estratègia	21
3.1.1. Recerca i anàlisi	21
3.1.2. Objectius	23
3.1.3. Indicadors	25
3.2. Fase 2 - Dades	27
3.2.1. Obtenció	27
3.2.2. Formatació i neteja	27
3.2.3. Processament	28
3.3. Fase 3 - Disseny	28
3.3.1. Esbossar	28
3.3.2. Prototipar	29
3.3.3. Finalitzar	30
4. Gràfics	31
4.1. Comparacions	32
4.1.1. Gràfic de barres	33
4.1.2. Gràfic de barres agrupades	37
4.1.3. Gràfic de barres apilades	38
4.1.4. Gràfic de radar	41
4.1.5. Gràfic d'intensitat de colors	43
4.1.6. Gràfic de marques	45
4.1.7. Gràfic múltiple	47

4.2. Tendències	48
4.2.1. Gràfic de línies	49
4.2.2. Gràfic de pendents	50
4.2.3. Minigràfic	51
4.3. Mapes	52
4.3.1. Mapa de coropletes	53
4.3.2. Mapa de símbols proporcionals	54
4.4. Parts d'un total	55
4.4.1. Gràfic de sectors	56
4.4.2. Pictograma	60
4.4.3. Mapa d'arbre	61
4.5. Distribucions	62
4.5.1. Histograma	63
4.5.2. Diagrama de caixa	64
4.6. Correlacions	67
4.6.1. Gràfic de dispersió	68
4.6.2. Gràfic de bombolles	72
4.6.3. Coordenades paral·leles	74
4.7. Connexions, relacions i xarxes	76
4.7.1. Diagrama de nodes i arestes	77
4.7.2. Diagrama de Sankey	78
5. Principis del disseny	80
5.1. Pensament visual	80
5.2. Color	82
5.3. Forma	84
5.4. Interacció	85
5.5. Recomanacions generals	87
5.5.1. Menys és més	87
5.5.2. Equilibri entre funcionalitat i estètica	87
5.5.3. La forma segueix la necessitat	88
5.5.4. Text	89
5.5.5. Maquetació	90
6. Bibliografia	91
7. Annexos	92
7.1. Glossari	92
7.2. Eines	93

1. Introducció a la guia

Actualment es generen dades sobre pràcticament qualsevol fenomen humà o natural. Àmbits tan diversos com la salut, l'esport, la política, l'educació o el màrqueting, per citar-ne només alguns, fan servir les dades per entendre millor la realitat. Per tant, cada cop més persones necessiten desenvolupar competències per treballar amb dades i entendre'n el significat i impacte. En canvi, un estudi recent estima que només el 17% dels ciutadans europeus estan preparats per fer servir dades¹.

Aquesta guia vol ajudar tothom a treballar amb dades, en particular a explorar-les, analitzar-les i explicar-les gràficament. Es fa servir el marc de treball de la visualització de dades, una disciplina que explica com tractar la informació de manera visual, i que ofereix dos grans avantatges. En primer lloc, el llenguatge visual és el més adequat per fer accessibles les dades a públics no especialistes. En segon lloc, la visualització facilita destacar en un context de sobrecàrrega informativa i, per tant, ajuda a fer arribar un missatge a un públic determinat.

Ara bé, un gràfic mal dissenyat pot donar lloc a una interpretació incorrecta de les dades i, per tant, provocar una mala decisió. Així mateix, un gràfic correcte que respon una pregunta irrelevant no tindrà cap impacte. És per això que en aquesta guia es tracten tant els aspectes de disseny com els metodològics, per tal d'assegurar que es formulen les preguntes adequades i que es trien els indicadors pertinents.

A més, es fa referència a les pautes per a l'elaboració de gràfics, taules i infografies que s'han d'harmonitzar amb el disseny de la resta de suports comunicatius de la Generalitat de Catalunya. La identitat corporativa és un element bàsic de la imatge que la institució projecta a la ciutadania i que identifica l'oficialitat i veritat de les dades que es presenten.

La guia es divideix en quatre grans blocs: què es la visualització de dades, metodologia, gràfics i principis del disseny. Tot i que la guia es pot llegir de manera lineal, hi haurà lectors que la podran fer servir com a referència en moments puntuals. Per aquest motiu, a continuació es presenten quatre infografies que resumeixen els quatre grans blocs de contingut. Aquestes infografies funcionen com un índex interactiu per accedir a aquells continguts de la guia que es poden consultar puntualment.

¹ Qlik. Data Equality Campaign. <<http://dataequality.org/>> [recurs electrònic]. [Consultat: novembre 2017]

Què és la visualització de dades

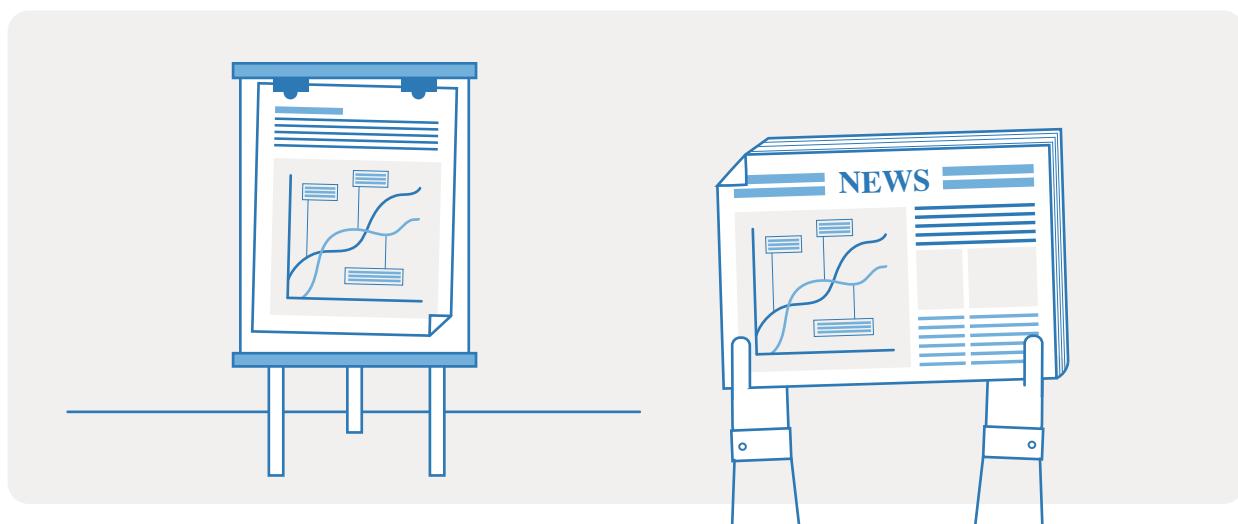
ÉS EXPLORAR unes dades per tal de respondre preguntes o fer-ne de noves.



ÉS ANALITZAR patrons, relacions i valors atípics entre les dades.

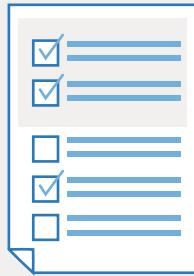


ÉS EXPLICAR una història amb dades.



Com s'elabora una visualització

DEFININT una estratègia



① Fer recerca

② Definir objectius

③ Definir indicadors

PREPARANT unes dades



A	X	
B	X	X
C	X	X
D	X	X



① Obtenir les dades

② Donar forma i netejar

③ Processar

DISSENYANT uns gràfics



① Esbossar

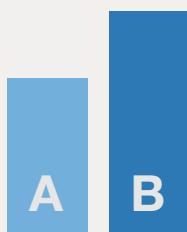
② Prototipar

③ Finalitzar

Gràfics

Comparacions

Per comparar diferents variables o categories entre si.



Tendències

Per entendre l'evolució temporal de variables quantitatives.



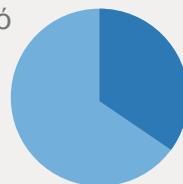
Mapes

Per trobar patrons geogràfics.



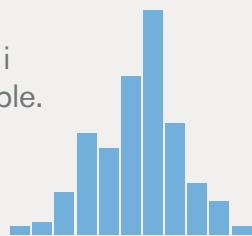
Parts d'un total

Per entendre la contribució de diferents categories a un total.



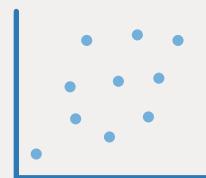
Distribucions

Per entendre la forma i propietats d'una variable.



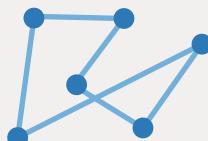
Correlacions

Per entendre la relació entre diferents variables.



Connexions, relacions i xarxes

Per entendre la relació entre els elements d'un conjunt de dades.



Principis de disseny

Pensament visual

Com aprofitar les capacitats cognitives de l'ésser humà.



Color

Com utilitzar el color per comunicar dades adequadament.



Forma

Quins atributs existeixen per codificar les dades mitjançant la forma.



Interacció

Com utilitzar la interacció per facilitar l'exploració i la comprensió.



Recomanacions generals

Altres consideracions per desenvolupar bones visualitzacions.



2. Què és la visualització de dades

La visualització de dades és una disciplina molt recent i encara està en procés de consolidació. Colin Ware va ser un dels primers investigadors a oferir una definició de visualització de dades prou acotada i, a la vegada, prou àmplia per no quedar desfasada enfront dels constants avanços tecnològics en el tractament i representació de dades. Segons Ware, la visualització és:

“la representació gràfica de dades o conceptes, que té com a resultat una imatge mental o un artefacte extern que ajuda a la presa de decisions”²

Així doncs, la visualització de dades té dos components principals:

- és una representació gràfica de dades, però també de conceptes;
- ajuda a prendre decisions.

Encara que per a Ware, la visualització és la representació gràfica de dades o conceptes, en aquesta guia s'utilitza la paraula visualització per fer referència només a la visualització de dades.

A nivell formal, i de manera molt general, es poden distingir dos tipus de visualitzacions: les estàtiques i les interactives.

Les visualitzacions estàtiques s'acostumen a utilitzar per comunicar unes dades que s'han analitzat prèviament. Aquest tipus de visualitzacions ajuden a explicar patrons i tendències generals.

En la següent imatge es pot veure una visualització en la qual es mostra l'evolució de la taxa d'atur als Estats Units en forma de gràfic de línies.

² Ware, Colin. *Information Visualization, Third Edition: Perception for Design*. Morgan Kaufmann, 2012

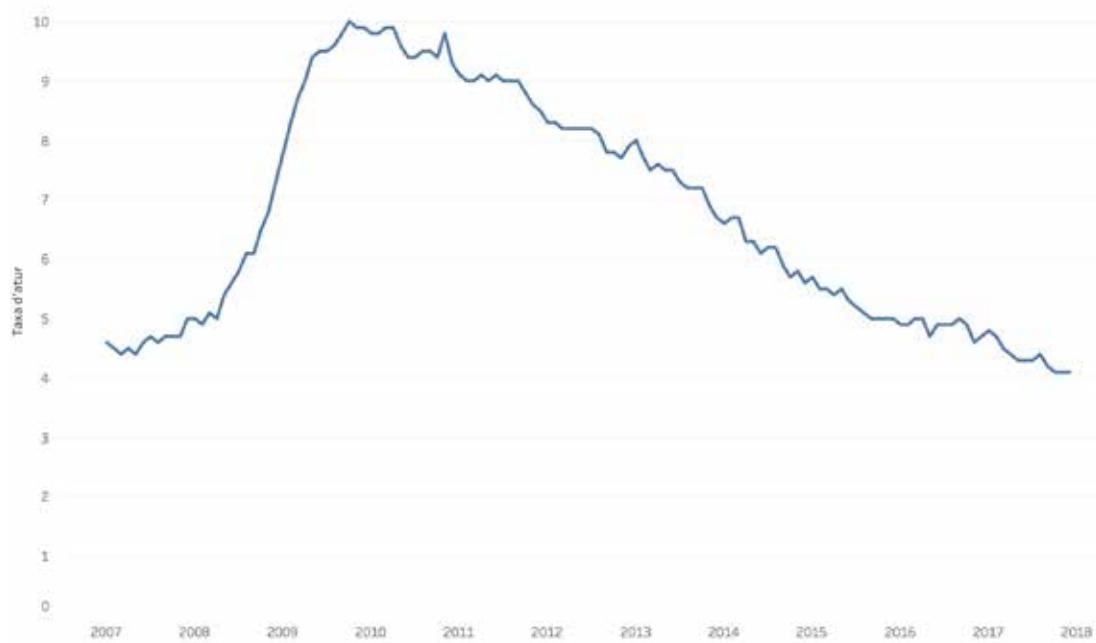


Figura 2.1. Evolució de la taxa d'atur als Estats Units.

Les visualitzacions interactives s'utilitzen perquè l'usuari pugui interactuar amb les dades. Per exemple, la visualització interactiva de The New York Times que referenciem a continuació permet que l'usuari dibuixi la seva estimació per a una sèrie d'indicadors econòmics i socials dels Estats Units:

<https://www.nytimes.com/interactive/2017/01/15/us/politics/you-draw-obama-legacy.html>.

Gràcies a la interacció, aconseguim centrar l'atenció de l'usuari en les dades. En aquest cas, el fem pensar sobre l'evolució de certs indicadors sota la presidència de George Bush i de Barack Obama. Un gràfic estàtic podria comunicar el mateix, però segurament no aconseguiria el mateix grau d'implicació del lector.

D'altra banda, la visualització té tres objectius principals:

- explorar
- analitzar
- explicar

A continuació, expliquem aquests objectius detalladament, així com els formats més adients per cadascun d'ells.

2.1. Visualitzacions per explorar

La naturalesa bidireccional dels mitjans digitals facilita que l'usuari pugui interaccionar amb les dades. Així doncs, en comptes d'explicar una història amb dades, es pot convertir el lector en protagonista. Facilita que el lector interrogui les dades, de manera que l'ajudin a respondre inquietuds particulars i preguntes adaptades al seu context. En aquests casos, la visualització ajuda a explorar un conjunt de dades.

Així doncs, una visualització de dades no només ajuda a resoldre dubtes, sinó que convida a plantejar-se preguntes que ni tan sols s'havien imaginat. Les visualitzacions exploratives són un bon exemple d'això: faciliten la interacció del lector, contribueixen a obrir un debat sobre un tema i responen moltes preguntes, alhora que en generen de noves.

Les visualitzacions de dades exploratives solen requerir un equip de desenvolupament multidisciplinari, capaç de pensar una història, recollir i analitzar les dades i dissenyar interfícies interactives. Eines com Tableau faciliten el desenvolupament d'aplicacions interactives de dades, però el format web (mitjançant llibreries com D3) és el més flexible i eficient, i el que s'adapta millor a diferents dispositius.

En el següent exemple, en comptes d'ofrir unes estadístiques mitjanes del rendiment del jugador de bàsquet Stephen Curry, es va optar per mostrar totes les dades dels llançaments. El resultat és una aplicació que permet veure l'evolució temporal dels punts que va marcar, en quins partits en va marcar més, i la localització d'aquests dins de la pista. Podem filtrar per diferents partits, moments del joc o tipus de tirs. Amb aquesta visualització no es pretén respondre cap pregunta concreta; simplement es tracta de poder explorar un conjunt de dades.

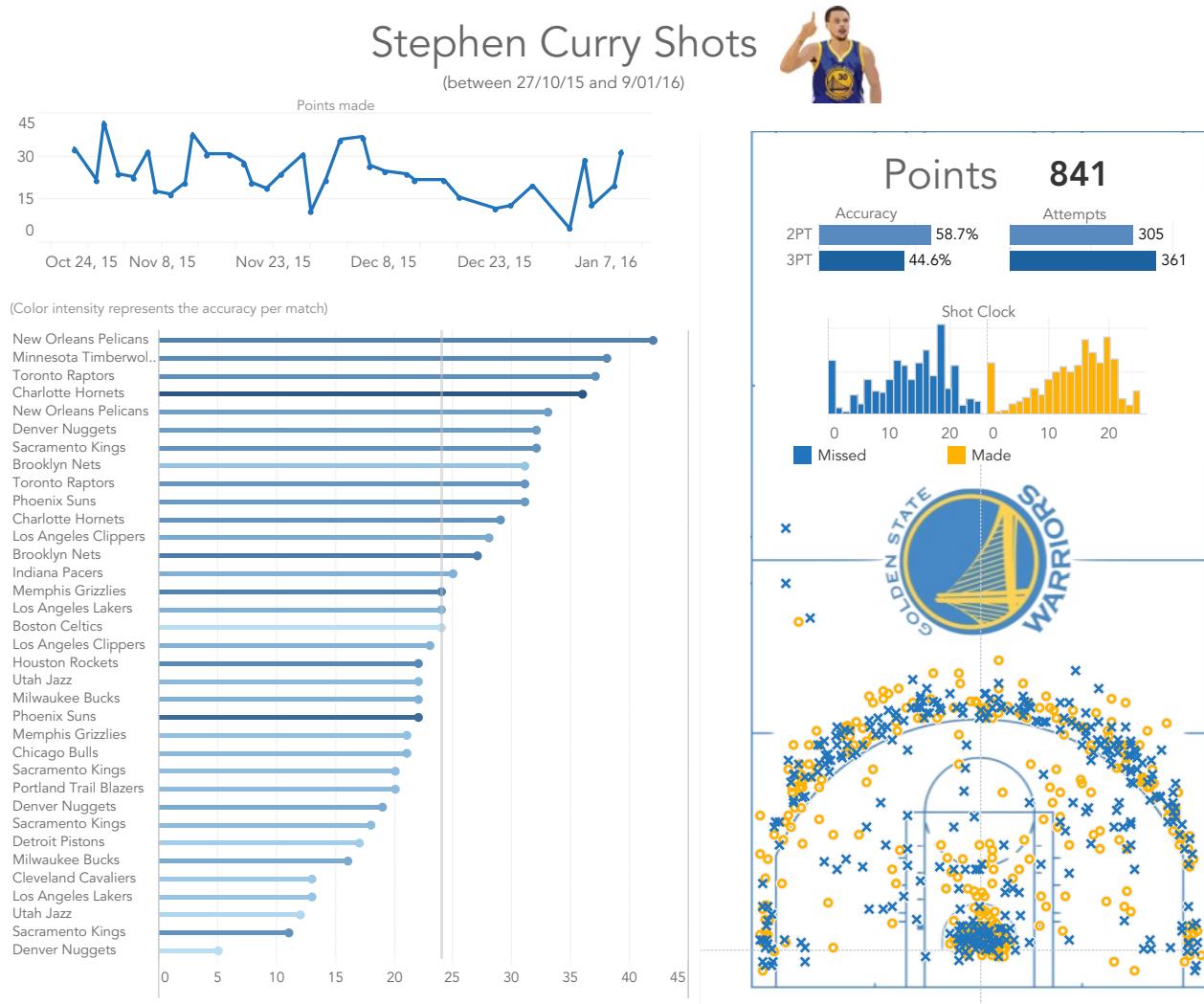


Figura 2.2. Visualització dels llançaments efectuats pel jugador Stephen Curry. Imatge cedida per OneTandem.

Font: <https://public.tableau.com/profile/onetandem#!/vizhome/curry/StephenCurryShots>

2.2. Visualitzacions per analitzar

A banda d'explicar una història o de permetre l'exploració d'un conjunt de dades, les visualitzacions també són útils a l'hora d'analitzar aquestes dades. Per tant, primer es pot fer servir la visualització de dades per analitzar un conjunt de dades i extreure'n una sèrie de conclusions, i després per explicar una història o permetre'n la seva exploració segons una sèrie de paràmetres.

Tot i que les visualitzacions interactives han arribat al gran públic gràcies al periodisme de dades, el seu ús principal continua sent el tractament i anàlisi de problemes complexos, en camps tan diversos com l'exèrcit, la medicina, la biologia, les polítiques públiques o les xarxes socials. Són especialment útils quan la complexitat del problema que cal tractar o la quantitat de dades excedeix les capacitats de les visualitzacions i tècniques d'anàlisi tradicionals.

2.2.1. Quadre de comandament

El quadre de comandament és una eina amb què es pot controlar l'estat d'un sistema. Per exemple, un quadre de comandament financer permet controlar els ingressos i les despeses d'una organització, per diferents departaments i segments de productes i clients.

Com a tret diferencial, un quadre de comandament va dirigit a usuaris experts en la temàtica que cobreix, que, alhora, són persones amb capacitat d'alterar el funcionament del sistema sobre el qual informa. Per tant, és molt important entendre els objectius i els coneixements de l'usuari que l'utilitzarà.

2.2.2. Aplicació d'analítica visual

Una aplicació d'analítica visual requereix unes capacitats superiors d'anàlisi i exploració respecte d'un quadre de comandament.

Parlarem d'aplicacions d'analítica visual sempre que es puguin fer servir tècniques estadístiques per analitzar les dades. Per exemple:

- facilitar mesures com la mitjana, la mediana o la correlació;
- agrupar elements mitjançant tècniques d'anàlisi de clústers³;
- construir models predictius que facilitin respondre què passarà, i no només què ha passat segons diferents escenaris.

El següent exemple és una aplicació d'analítica visual que permet contextualitzar els rànquings de diferents universitats, analitzar quins factors contribueixen a la seva posició, i quines correlacions hi ha entre els diferents indicadors. No és una visualització adequada per comunicar un rànquing d'universitats (per a això n'hi hauria prou amb un **gràfic de barres**). En canvi, és l'adequada per tal que una universitat pugui entendre per què ocupa una posició determinada i què pot fer per millorar-la.

³ L'anàlisi de clústers és una tècnica d'anàlisi de dades amb què es pot segmentar, de manera automàtica, un conjunt de dades en subgrups (anomenats clústers) que tinguin valors similars a les seves variables.



Figura 2.3. Aplicació analítica per analitzar rànquings d'universitats. Imatge cedida per l'empresa SIRIS Academic. Font: <http://university-analytics.com/rankings/#/explorer/ARWU/2017>.

2.3. Visualitzacions per explicar

Internet, les xarxes socials i les dades obertes han democratitzat l'accés a les dades. En aquest context, han sorgit iniciatives com el periodisme de dades, enfocat a generar notícies utilitzant precisament les dades com a font principal d'informació.

La visualització és vital per comunicar i ajudar a entendre adequadament les dades en el context d'una història o propòsit determinat. Si se'n fa una bona anàlisi, però no se sap explicar, no servirà de res. Però el contrari també és cert. Es pot crear una visualització de dades formalment perfecta, però si l'anàlisi que se'n fa o la història que s'explica no tenen importància en un context determinat, la visualització tampoc no servirà de res. Així mateix, cal tenir present que la visualització pot servir per aclarir el significat de les dades, però també pot servir per enganyar el lector. És per això que tenir coneixements de visualització de dades avui en dia és indispensable si s'aspira a entendre millor el món i la societat que ens envolten.

D'altra banda, la visualització de dades fa que es destaquen com a comunicador en un context de saturació informativa. Hi ha més dades que mai, però no es dediquen prou esforços a refinar-les i a oferir una informació rellevant i de qualitat. La metodologia de la visualització de dades ens obliga a pensar què es volen comunicar amb aquestes, què és important i com s'ha de fer. A la vegada, s'aconsegueix comunicar els continguts de manera visualment més atractiva i en formats més adequats per propagar-los a les xarxes socials.

A continuació, es presenta una introducció als principals formats que utilitza la visualització per explicar històries basades en dades.

2.3.1. Infografia

Les infografies es van popularitzar en revistes i diaris com una manera de comunicar temes complexos al públic en general. Amb les xarxes socials, les infografies han incrementat de manera exponencial la seva audiència per tres motius:

- el component visual ajuda que destaquen i això fa que siguin més compartides i que es propaguin ràpidament per les xarxes;
- estan dissenyades per ser consumides en poc temps i, per tant, s'adapten a les pautes de consum ràpid d'informació de les xarxes socials;
- no solen requerir coneixements previs del tema tractat i, per tant, s'adrecen a un públic ampli.

Un molt bon exemple d'infografia és el següent treball sobre la balena, de Jaime Serra:

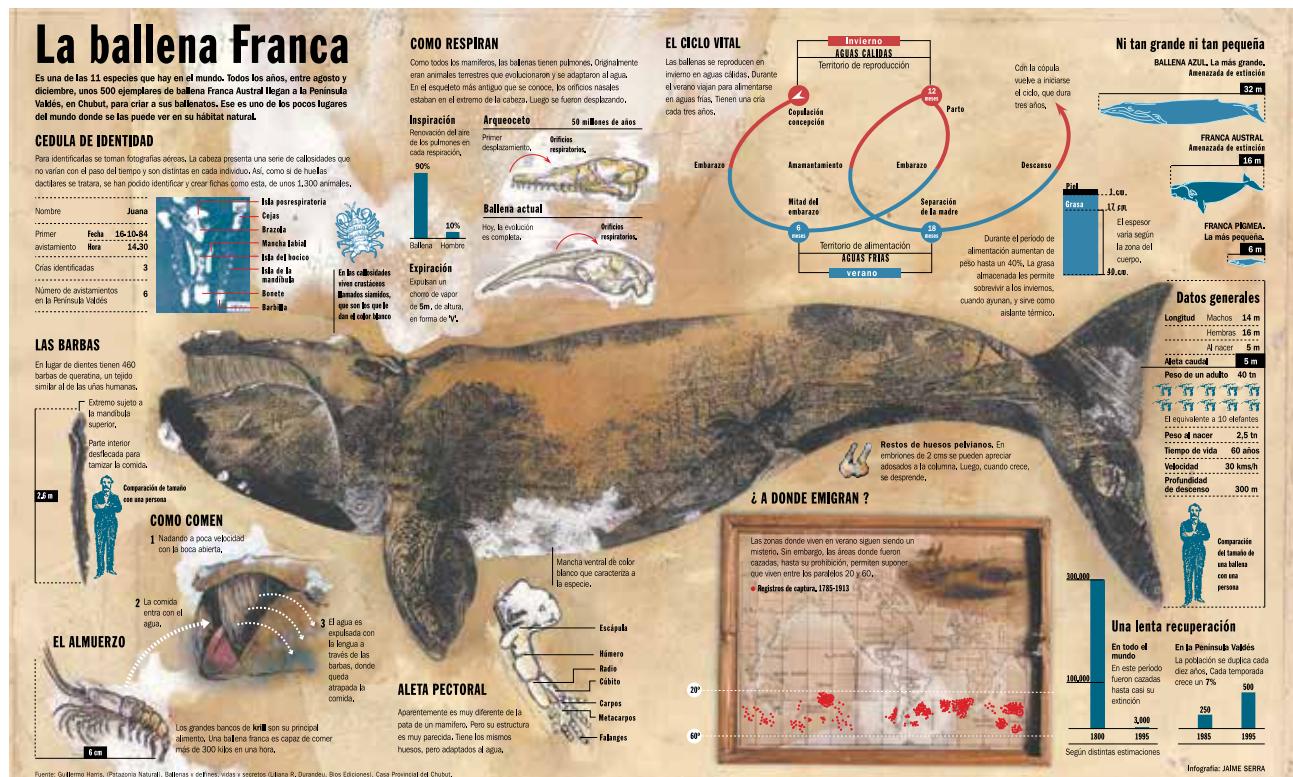


Figura 2.4. "La ballena franca". Infografia cedida per Jaime Serra.

2.3.2. Narrativa per desplaçament

La narrativa per desplaçament (scrollytelling), és una forma de narració interactiva on l'usuari ha de desplaçar-se (fent servir la barra de desplaçament, amb el ratolí, o arrossegant el dit en un dispositiu tàctil) per obtenir més informació. Es fa servir per explicar històries des de múltiples punts de vista i/o amb múltiples elements, com ara gràfics, vídeos, animacions, fotografia i text. La interacció facilita que la història aparegui en l'ordre requerit i que es gaudeixi de l'atenció de l'usuari.

Un bon exemple d'aquest format és la narrativa per desplaçament feta per Bloomberg que explica els motius de l'escalfament global. Aquesta narrativa es pot consultar a:

<https://www.bloomberg.com/graphics/2015-whats-warming-the-world/>.

La narrativa per desplaçament és un format purament pensat per a pàgines web, generalment associat a grans mitjans de comunicació digitals. Es pot emmarcar dins de la tradició de la narrativa interactiva⁴. La seva elaboració requereix molts recursos perquè necessita la col·laboració de perfils diversos: guionistes, dissenyadors, programadors, fotògrafs, realitzadors, etc.

2.3.3. Presentació

Les presentacions amb eines com PowerPoint o similars s'han convertit en un format habitual de comunicació en les organitzacions. Com que soLEN presentar la informació de manera concisa, les visualitzacions de dades hi juguen un paper important.

En general, es recomana incorporar només una idea a cada diapositiva. Aquesta idea es pot transmetre amb una frase, una imatge, o una visualització de dades, i ha d'acompanyar el discurs de manera natural. La regla de limitar-nos a una idea per diapositiva ens convida a introduir la visualització de dades en diferents passos.

A continuació es reproduceix el cas presentat per Cole Nussbaumer, especialista en *storytelling* amb dades, en el seu llibre i blog *Storytelling with data*⁵ (<http://www.storytellingwithdata.com/>).

⁴ Wikipedia, Interactive Storytelling. <https://ca.wikipedia.org/wiki/Projecci%C3%B3_cartogr%C3%A0fica> [Consulta: novembre 2017]

⁵ Nussbaumer, Cole. *A Data Visualization Guide for Business Professionals*. Wiley, 2015

En el cas que s'analitza, es veu l'evolució del nombre d'usuaris actius d'una aplicació de mòbil anomenada Moonville. La forma habitual de presentar les dades és la següent:

Moonville: active users over time



Figura 2.5. Visualització utilitzada per mostrar l'evolució dels usuaris actius d'una aplicació mòbil. Imatge cedida per Cole Nussbaumer.

Per comunicar les dades amb una presentació en PowerPoint o similar, Cole Nussbaumer proposa descompondre el gràfic en una sèrie de diapositives. En primer lloc, començaria amb aquesta imatge per situar el problema:

Moonville: active users over time



Figura 2.6. Imatge inicial recomanada per explicar el problema. Imatge cedida per Cole Nussbaumer.

Després, s'aniria presentant l'evolució amb la següent sèrie de diapositives, cada una acompanyada del discurs pertinent:

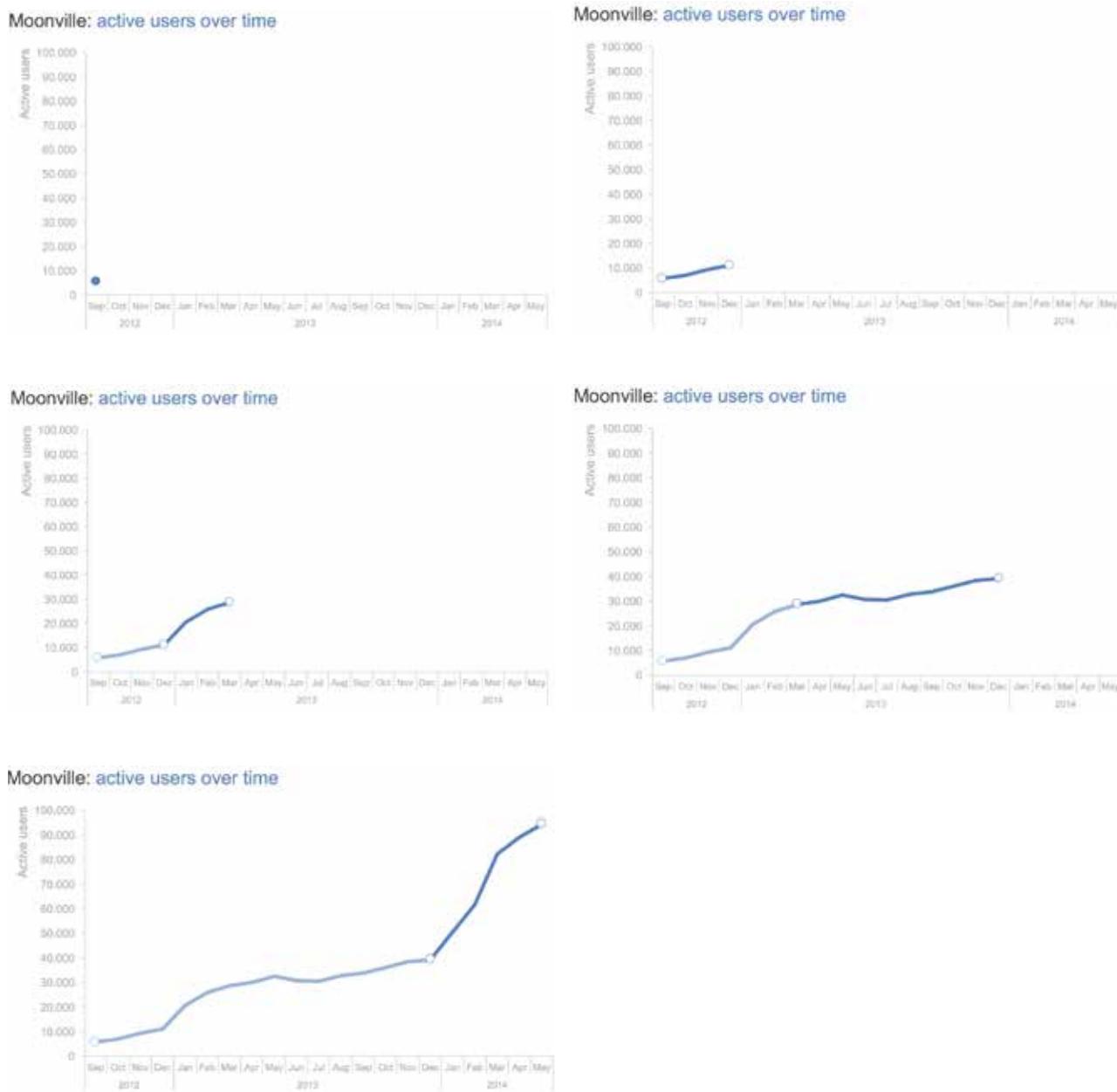


Figura 2.7. Sèrie d'imatges utilitzades per explicar i presentar l'evolució dels usuaris actius en una aplicació mòbil. Imatges cedides per Cole Nussbaumer.

Finalment, es pot presentar una transparència que resumeixi tot el discurs. Aquesta serà la diapositiva que es pot incloure en un fitxer PowerPoint (o similar) que s'envii a l'audiència per correu electrònic o que es lliuri imprès.

Moonville: active users over time



Figura 2.8. Visualització final que resumeix tot el discurs fet prèviament. Imatge cedida per Cole Nussbaumer.

Amb aquest exemple, es mostra que el repte d'una presentació és doble: conèixer el tema a tractar i saber explicar-lo. Per això, a banda de ser bons dissenyadors i oradors, és vital conèixer l'audiència. Per exemple, si es prepara una presentació per a una classe cal esbrinar què volen aprendre els alumnes i quin és el nivell de coneixements sobre el tema que es tractarà. Es pot dissenyar una presentació molt completa, però si no respon al nivell i objectius de l'audiència, no funcionarà. Un error habitual és fer servir terminologia que l'audiència no coneix o bé visualitzacions de dades massa complexes.

2.3.4. Vídeo

El vídeo va pel camí de convertir-se en el format per excel·lència de la xarxa. La visualització de dades, tan lligada a Internet, no és aliena a aquest fet. Per exemple, els gràfics en moviment (*motion charts*) s'han convertit en un mecanisme habitual per transmetre dades de forma animada. Els GIF també ocupen un espai important en xarxes de consum ràpid d'informació com Twitter. Així mateix, nombrosos canals de Youtube i mitjans digitals aposten per formats de vídeo que incorporen dades i infografia de manera innovadora.

Un mitjà que aposta per aquest format és Vox (<https://www.vox.com/>). En el seu canal de Youtube (<https://www.youtube.com/user/voxdotcom>) es pot veure una bona mostra de vídeos que fan ús de la visualització de dades, com per exemple el titulat “The real reason American health care is so expensive” https://www.youtube.com/watch?v=k1vE_LVBx4s.

3. Metodologia

En aquest capítol es descriu el procés per desenvolupar una visualització de dades, que consta de tres fases:

- Fase 1: estratègia
- Fase 2: dades
- Fase 3: disseny

3.1. Fase 1 - Estratègia

L'estratègia fa referència a tot allò que ens condueix a definir els objectius de la visualització de dades. En particular, es tracten els següents punts:

- Recerca i ànalisi
- Objectius
- Indicadors

3.1.1. Recerca i ànalisi

Per elaborar una visualització sobre un tema en particular cal conèixer-lo en profunditat. Així es pot tractar des de diferents angles i assegurar que es disposa de les dades adequades per mostrar una bona ànalisi. Per això, sovint cal fer recerca del tema que es vol tractar. Per fer-ho adequadament, convé investigar els següents aspectes:

- Temàtica
- Audiència
- Ànalisi

Temàtica

Tot projecte de visualització de dades gira al voltant d'una temàtica en particular. Per exemple, si es volen visualitzar dades sobre la contaminació a Catalunya, es pot definir la temàtica com el medi ambient.

El primer pas és plantejar la temàtica des de diferents punts de vista. L'esquema 5W1H⁶ pot ajudar. Aplicat a l'exemple de la contaminació, aquest esquema plantejaria les següents preguntes:

- Què - Què vol dir tenir un problema de contaminació?
- Qui - Qui n'és responsable / Qui la pateix?
- On - On es pateix més la contaminació? On han arreglat el problema?
- Quan - Quan es pateix més la contaminació? Quina evolució hi ha hagut en els darrers anys?
- Per què - Per què varien els nivells de contaminació? Per què poden suposar un problema?
- Com - Com es pot reduir / augmentar la contaminació? Com es regula la contaminació?

Un cop definits un o més angles per tractar la temàtica, és recomanable consultar publicacions i experts en aquesta àrea, així com altres visualitzacions de dades que ja s'hagin fet, per assegurar-se que no es repeteix la mateixa feina.

Finalment, és útil trobar el testimoni d'implicats en el tema. Per exemple, en el cas de la contaminació, seria molt enriquidor incorporar l'opinió de persones afectades pel problema. Les dades ajuden a quantificar un problema, però els testimonis personals ajuden a fer-lo més proper i, per tant, augmenta la probabilitat de causar un major impacte en el consumidor de la visualització.

Audiència

Per audiència s'entén qui farà servir la visualització de dades. Dissenyar una bona visualització de dades és similar al disseny d'una aplicació i, per tant, saber quins són els coneixements i les motivacions de l'usuari és vital per assegurar l'èxit de la visualització.

Per conèixer l'audiència, es recomana respondre aquestes preguntes:

Relació

- la coneixem personalment?
- ens coneix?
- és homogènia o heterogènia?

Coneixements sobre el tema

- ens reconeix com a autoritat en el tema?
- quins coneixements té del tema?
- quina experiència té en l'ús de les dades?

⁶ Wikipedia, Five Ws. <https://en.wikipedia.org/wiki/Five_Ws> [Consulta: 20 novembre 2017]

Context d'ús

- quin dispositiu o dispositius farà servir per consultar la visualització?
- quins objectius té?
- quines accions esperem que dugui a terme?
- quant de temps té per consultar la visualització?
- per quins canals arribarem a l'audiència?

En definitiva, es tracta d'entendre el model mental dels usuaris per poder dissenyar productes adaptats als seus objectius, capacitats i context d'ús.

Anàlisi

Un altre punt a tenir en compte a l'hora de definir l'estratègia és l'anàlisi de les dades. I és que en el procés d'exploració i anàlisi d'un conjunt de dades, sovint es troba la idea d'una visualització. També pot passar que les dades demostrin que la idea inicial no és vàlida, o que és irrellevant.

Per aquesta primera anàlisi de les dades, es recomana fer les següents exploracions preliminars:

- Distribució dels principals indicadors, amb histogrames i diagrames de caixa: això ajudarà a trobar valors atípics, que sovint amaguen històries interessants.
- Tendències dels principals indicadors: es detectarà si hi ha grans canvis en determinats moments.
- Correlació entre diferents indicadors: la ment tendeix a establir relacions entre fenòmens, per tant, les històries que expliquen relacions funcionen molt bé.

3.1.2. Objectius

Ara que ja es coneix la temàtica i l'audiència, és el moment de definir els objectius de la visualització de dades. En general, es diu que els objectius estan definits si es poden respondre clarament les següents preguntes:

- Com volem ajudar l'usuari?
- Per què?
- Com ho avaluarem?

Per fer-ho més entenedor, cal imaginar que es planteja, per exemple, una visualització de dades sobre la contaminació a Catalunya.

Com volem ajudar l'usuari?

Tal com s'ha explicat en el capítol d'introducció, les visualitzacions de dades poden ajudar a explorar, analitzar i/o explicar un tema. Per tant, el primer que cal decidir és com s'ajuda l'usuari, és a dir:

- Es vol ajudar a explorar unes dades sobre un tema?
- Es vol ajudar a analitzar les dades sobre un tema?
- Es vol explicar un tema?

En l'exemple de la visualització sobre la contaminació a Catalunya, es podria respondre de tres maneres ben diferents la pregunta “Com es vol ajudar l'usuari? ”:

- Es volen facilitar a l'usuari les dades de l'evolució de la contaminació per comarca, perquè sigui capaç d'explorar com han variat a la zona on viu i treballa.
- Es volen facilitar a l'usuari dades sobre la contaminació, densitat de població, indústria i trànsit, així com mecanismes per establir correlacions entre diferents indicadors, per poder analitzar quins factors contribueixen més a la presència de contaminació en diferents zones del país.
- Es vol explicar a l'usuari quant ha disminuït la contaminació a Catalunya en els darrers deu anys, i quins n'han estat els principals motius.

Les tres respostes poden donar lloc a bones visualitzacions de dades, però segurament no podran resoldre amb una mateixa visualització de dades. És a dir, es pot enfocar un mateix tema des de múltiples punts de vista, i per a cada punt de vista hi haurà diferents maneres de visualitzar les dades.

Per què?

Aquesta pregunta pretén fer-nos reflexionar sobre la importància i l'abast del que s'està fent i, per tant, els recursos que s'hi haurien de dedicar. També provoca que es comenci a pensar en el format i els canals de comunicació.

Tornant al cas de la contaminació a Catalunya, es podrien definir els següents perquès:

- sensibilitzar la població sobre els efectes nocius de la contaminació per a la salut;
- ajudar els tècnics mediambientals a planificar millor les polítiques d'actuació.

En el primer cas, es vol comunicar amb la població en general, que no té coneixements avançats d'anàlisi i interpretació de dades. Per tant, una sèrie de visualitzacions de dades en format GIF accompanyades d'una campanya en xarxes socials, segurament seran més efectives que una sofisticada visualització interactiva. En canvi, en el segon cas, seria millor desenvolupar una visualització interactiva amb múltiples indicadors i eines d'anàlisi que facilitin la interpretació de les dades de la millor manera possible, d'acord amb les competències d'un tècnic.

Com s'avaluarà?

Es tracta de definir unes mètriques per avaluar si s'ha aconseguit ajudar l'usuari adequadament, de manera que s'assoleixin els objectius plantejats a la pregunta “per què? ”.

Seguint amb l'exemple d'abans, si es vol sensibilitzar la població sobre els efectes de la contaminació, es pot mesurar el percentatge de ciutadans que són capaços de respondre una sèrie de preguntes sobre la contaminació abans i després de consultar la visualització de dades.

3.1.3. Indicadors

Sovint, les visualitzacions de dades no assoleixen els objectius plantejats perquè tracten unes dades que no són rellevants o que no capten tots els angles del problema. Per això és vital definir uns indicadors adequats. En general, es poden definir tres tipus d'indicadors:

- Volum
- Qualitat
- Context

Per exemple, el nombre d'accidents de trànsit (indicador de volum) no és un bon indicador per comparar-lo entre províncies, tot i que es faci amb un gràfic adequat com el següent:



Figura 3.1. Nombre d'accidents per província.

En canvi, un indicador com la taxa d'accidents per habitant (indicador de qualitat) és més adient, ja que és independent de la població de la comunitat i, per tant, es poden comparar. En el gràfic següent es pot veure que Ceuta i Melilla passen a estar a la primera posició de la classificació i que, per exemple, les illes Balears se situen per sobre de Madrid.

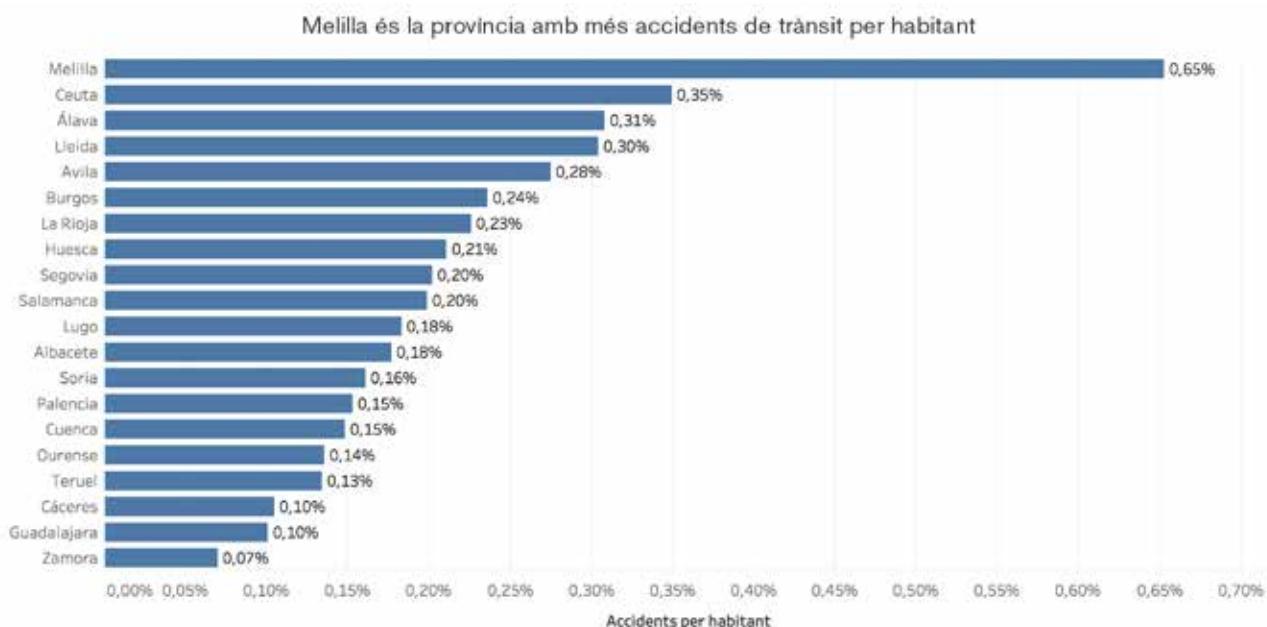


Figura 3.2. Nombre d'accidents per habitant i província.

Finalment, disposar de la tendència de la taxa d'accidents en els darrers anys (indicador contextual) ajudarà a interpretar els indicadors. Per exemple, una taxa del 0,2% seria positiva si fa cinc anys era del 0,5%, però negativa si era del 0,1%.

A banda de fer servir els indicadors adequats, cal tenir present el nivell d'agregació de les dades. Per exemple, si es vol analitzar la tendència anual dels accidents de trànsit, cal decidir si es volen mostrar les dades per hores, per dies o per mesos. Cada nivell d'agregació ens permetrà arribar a diferents conclusions i requerirà un tipus de gràfic diferent.

En resum, la fase d'estrategia finalitza quan:

- Es coneix el tema a tractar i l'audiència a la qual s'adreça
- Es defineix com es vol ajudar l'usuari i per què;
- Es defineixen uns indicadors adequats (volum, qualitat, context).

Es recomana tenir sempre presents aquests punts durant el posterior procés de disseny de la visualització de dades, per assegurar que s'assoleixen els objectius marcats inicialment.

3.2. Fase 2 - Dades

Tot i que no és l'objectiu d'aquesta guia fer una introducció al tractament de les dades, cal ser conscient que en molts casos s'han de treballar per poder disposar-les en un format i estructura adequats.

Sovint es diu que la informació són dades posades en context o dades organitzades i estructurades.

A continuació, s'esmenten els passos que cal tenir en compte.

3.2.1. Obtenció

El primer pas per poder treballar una visualització serà obtenir les dades necessàries, definides durant el procés d'estratègia. Sovint aquestes ja estan donades, però altres vegades cal cercar-les i extreure-les de bases de dades, o implementar sistemes que les generin.

3.2.2. Formatació i neteja

Un cop que es disposa de les dades, cal validar-ne el contingut, assegurar-se que no contenen errors i que el format és adequat i consistent per al seu consum. Sovint contenen errades o valors poc coherents que cal revisar. Els valors decimals expressats amb comes i alhora amb punts en són un exemple típic.

Caldrà, doncs, en la mesura del possible, repassar cadascun dels registres de les nostres dades per validar que tots els valors estan ben escrits i tenen sentit.

Es pot parlar de dos tipus de formats:

- **Format general de les dades:** hi ha una infinitat de formats que expressen dades de manera estructurada. Els fulls de càcul, que contenen files i columnes, són un exemple d'organització de dades que es poden expressar en diferents formats (XLS, CSV, etc.). Un altre format força utilitzat, però alhora més avançat, és el format JSON. Es tria el format de les dades en funció de les dades de què es disposen i en funció de les opcions d'importació de què disposi el programari que s'utilitzi per treballar-hi.
- **Format de cadascun dels indicadors:** és important comprovar que cadascun dels indicadors segueix un format consistent i, preferiblement, estàndard. Per exemple, si hi ha una columna que expressa una data, és important que tots els valors d'aquesta segueixin el mateix format. Per exemple, dia/mes/any (15/10/1981). En aquest cas, el més important és que hi hagi consistència, de manera que tots els valors de l'indicador estiguin expressats de la mateixa manera.

3.2.3. Processament

El processament de les dades és, per si mateix, una disciplina pròpia. En funció del volum i del que es vulgui analitzar, es podran executar algorismes molt complexos que podran donar molt més valor a les dades. No obstant això, normalment cal disposar de grans coneixements de matemàtiques, estadística i programació per poder-los implementar i utilitzar.

Finalment, un cop acabada la visualització (un cop fet el procés de disseny descrit en la següent secció), cal ser conscients que és important ser metòdics a l'hora d'emmagatzemar les nostres dades. En grans institucions és desitjable disposar d'un repositori de dades únic i estàndard on es puguin desar els arxius i documentar-los.

En qualsevol cas, és recomanable desar els fitxers finals on hi hagi les dades en un lloc fàcilment identifiable. A més a més, també és una bona idea acompañar-los amb un petit document de text que descrigui les dades, la seva procedència, com s'han processat i la data d'obtenció. Aquests fitxers seran molt útils si es volen reutilitzar les dades en el futur o si eventualment es volen fer públiques.

3.3. Fase 3 - Disseny

El disseny fa referència a tot allò que ens permet aconseguir els objectius definits a l'estrategia. El resultat és el producte final, que s'ha d'avaluar en funció de com ajuda a complir aquests objectius.

El procés de disseny que es presenta és iteratiu i basat en la metodologia de disseny centrat en l'usuari. En particular, se segueixen aquestes fases:

- Esbossar
- Prototipar
- Finalitzar

3.3.1. Esbossar

Aquesta fase consisteix a elaborar esboçaments amb l'objectiu de descobrir maneres de representar les dades d'acord amb l'estrategia definida. També és habitual referir-s'hi amb el terme anglès *sketching*.

En particular, la fase d'esbós és molt útil per definir l'aspecte de la visualització de dades, que ha de ser coherent amb els objectius definits a l'estrategia. Quan es fan els primers esboçaments, es decideixen el tipus de gràfics que es faran servir, quins textos han d'acompanyar-los i com s'ha de maquetar.

Generalment, en aquesta fase es treballa amb paper i bolígraf per tenir el màxim de llibertat a l'hora de pensar la visualització de dades. Així mateix, es recomana treballar amb llibretes de diferents mides, que ajudin a pensar dissenys per a pantalla d'ordinador, per a tauleta i per a mòbil.

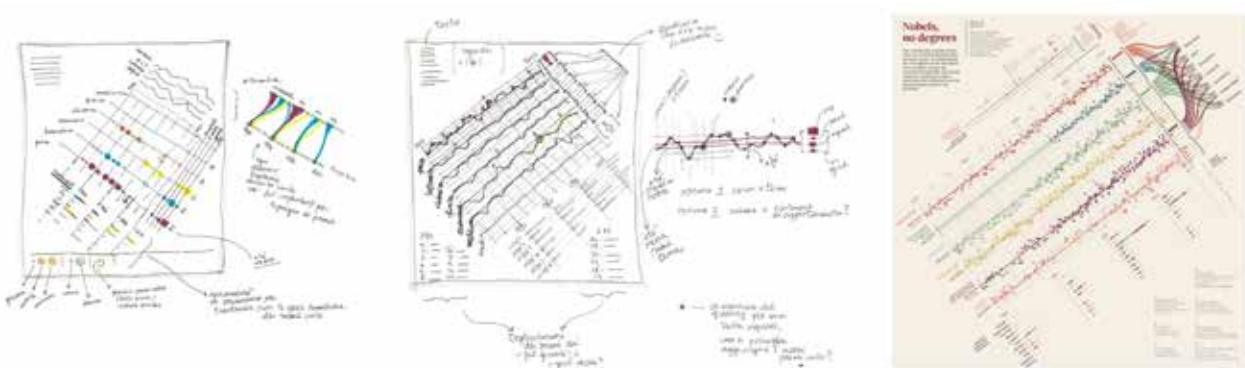


Figura 3.3. Esbossos presentats a l'article “Sketching with Data Opens the Mind’s Eye”, de Giorgia Lupi. Imatge cedida per l'autora. <https://medium.com/accurat-studio/sketching-with-data-opens-the-mind-s-eye-92d78554565>

3.3.2. Prototipar

En aquesta fase es fan servir les dades reals per veure si els esbossos continuen sent vàlids. A vegades, algunes visualitzacions triades no són les adequades. En aquest cas, cal tornar a treballar amb paper i bolígraf per obtenir noves idees. És a dir, durant el prototipatge, s'alterna l'ús de paper i bolígraf amb l'ús d'eines digitals de prototipatge per treballar amb dades reals.

Un cop que es té una primera versió del prototip, s'ha de presentar a l'usuari final amb l'objectiu de validar si respon a les seves necessitats. Si no es coneix l'usuari final, convé triar un grup d'usuaris representatius i fer un test d'usuari, semblant als que es fan per avaluar la usabilitat d'una aplicació. Es considera que un test amb 5 usuaris ajuda a detectar gran part dels problemes d'un disseny.

És important centrar el test amb usuaris en els objectius del disseny. Per exemple, si es pregunta “què et sembla aquest quadre de comandament?” el més habitual serà que l'usuari parli de si li agrada o no, basant-se en la seva experiència particular. L'estètica és el més fàcil de jutjar, però no és l'element més important d'un prototip. En canvi, es recomana recordar a l'usuari els objectius de la visualització de dades que se li presenta i fer-li preguntes per esbrinar si és capaç d'assolir els objectius marcats.

Una altra tècnica consisteix a presentar la visualització de dades a l'usuari i deixar-lo interaccionar durant un temps, sense cap interferència per part nostra. Finalment, se li pregunta què ha après amb la visualització i com l'ha fet servir per entendre què funciona i què no funciona.

Un cop que es coneixen els aspectes a millorar, s'elabora un nou prototip i es repeteix el test amb uns altres usuaris. És recomanable fer un parell d'iteracions, abans de passar a la fase final.

3.3.3. Finalitzar

En aquesta fase es transforma el prototip en el producte final. És el moment de tractar aspectes com:

- La redacció de textos finals
- L'adaptació a la guia d'estil requerida (colors, tipografies, espais, logotips, etc.)
- L'elaboració de mecanismes d'ajuda

A més a més, un cop que el producte està acabat, cal fer una revisió de qualitat, és a dir:

- Assegurar que el disseny funciona correctament en tots els dispositius on caldrà fer-lo servir, amb diferents resolucions i mides de pantalla.
- Comprovar cada funció i assegurar-se que no dona cap error (particularment, els filtres i altres elements d'interacció).
- Comprovar que les dades són correctes i fixar-se especialment que no apareguin resultats inesperats per a determinades combinacions de filtres o elements interactius.

Malgrat que ja s'hagi finalitzat el disseny, encara no s'haurà fet servir en entorns reals. Per tant, és probable que sorgeixin nous problemes i que facin falta altres millors. També és probable que els requeriments inicials puguin canviar al cap d'un temps. És per això que es recomana tenir present que un disseny mai serà definitiu i que cal reservar recursos per adaptar-lo a diferents circumstàncies.

Per assegurar-se que el disseny evoluciona, es recomana el següent:

- Fer trobades periòdiques amb l'usuari o usuaris de la visualització de dades.
- Analitzar periòdicament les mètriques de rendiment i l'ús de la visualització de dades.

4. Gràfics

Per visualitzar dades, ja sigui per explorar-les, analitzar-les o explicar-les, hi ha infinites possibilitats. Tot i això, hi ha un conjunt de gràfics considerats estàndard, com per exemple els gràfics de barres o els gràfics de línies. En aquest capítol, es classifiquen aquests gràfics en 7 categories diferents, en funció del que es pot aconseguir:

- comparacions
- tendències
- mapes
- parts d'un total
- distribucions
- correlacions
- connexions, relacions i xarxes

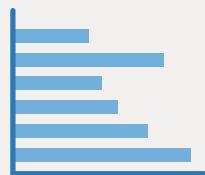
A continuació, es presenten les diferents visualitzacions de cada categoria. Tot i això, cal tenir en compte que sempre hi pot haver diverses opcions vàlides per representar un conjunt de dades. S'haurà de valorar què es vol representar i com són les dades per acabar de triar quina és la millor visualització a utilitzar.

4.1. Comparacions

Cal triar un gràfic amb què es puguin fer comparacions quan es volen comparar diferents variables entre si, quan vulguem comparar el valor d'una o més categories entre elles, o quan es volen comparar valors per tal d'establir un rànquing (és a dir, quan es vol saber quin element d'un conjunt de dades és el primer, quin és el segon, i quina diferència de valor hi ha entre els uns i els altres).

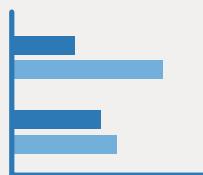
Gràfic de barres

Per comparar un conjunt de valors i establir un rànquing.



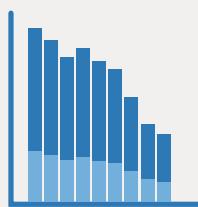
Gràfic de barres agrupades

Per comparar el valor dels diferents segments que formen part de les nostres categories.



Gràfic de barres apilades

Per comparar el valor total de la suma dels segments que formen cadascuna de les barres.



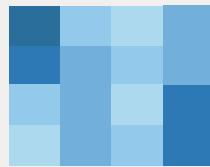
Gràfic de radar

Per entendre els valors d'un element en diferents variables alhora.



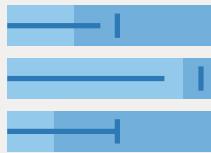
Gràfic d'intensitat de colors

Per descobrir la variabilitat d'un conjunt de variables.



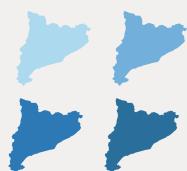
Gràfic de marques

Per avaluar un valor concret respecte al que s'esperava i respecte a uns rangs de qualitat establerts.



Gràfic múltiple

Per representar dades de diferents categories que no es poden veure correctament en un únic gràfic.



4.1.1. Gràfic de barres

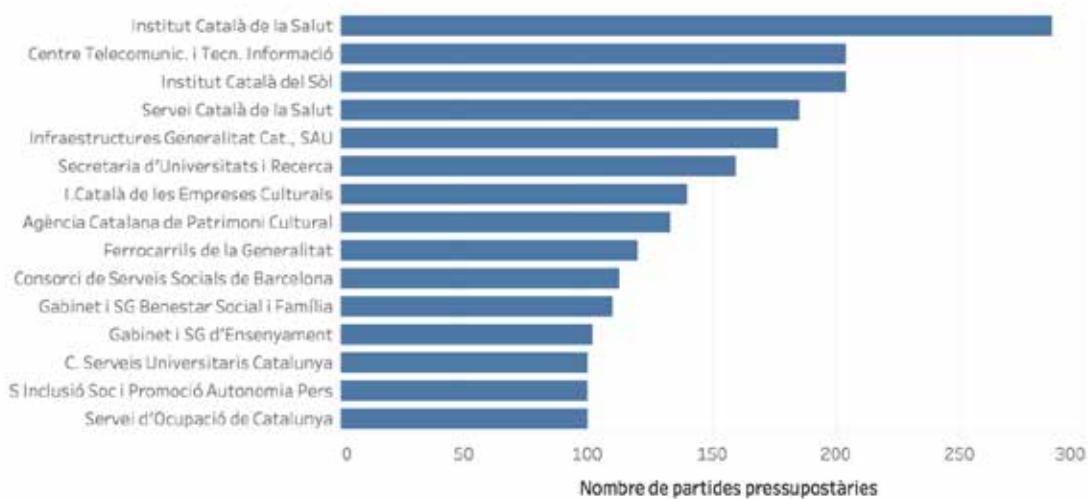


Figura 4.1. Exemple dels 15 serveis amb major nombre de partides pressupostàries durant el 2015.

Font: Pressupostos de la Generalitat de Catalunya 2015

(http://economia.gencat.cat/ca/70_ambits_actuacio/pressupostos/els-pressupostos-de-la-generalitat-de-catalunya-per-al-2015/).

Aquest gràfic està format per dos eixos: un de quantitatiu, que mostra l'escala dels valors que es representen; i un de textual, que representa la categoria a la qual pertanyen les dades representades. A l'eix textual s'hi situen un conjunt de barres, la longitud de les quals codifica el valor de cada categoria.

Quan s'ha d'utilitzar?

Quan es vol comparar un conjunt de valors i establir un rànquing. Com que l'ull humà és molt bo comparant distàncies d'elements que estan situats sobre un mateix eix, el gràfic de barres no només descobreix els valors més alts i els més baixos, sinó que també ajuda a tenir una intuïció força acurada de la diferència que hi ha entre uns valors i uns altres.

Recomanacions

Es recomana ordenar les categories de més gran a més petita, o a l'inrevés. Això facilita la comunicació del rànquing de categories.

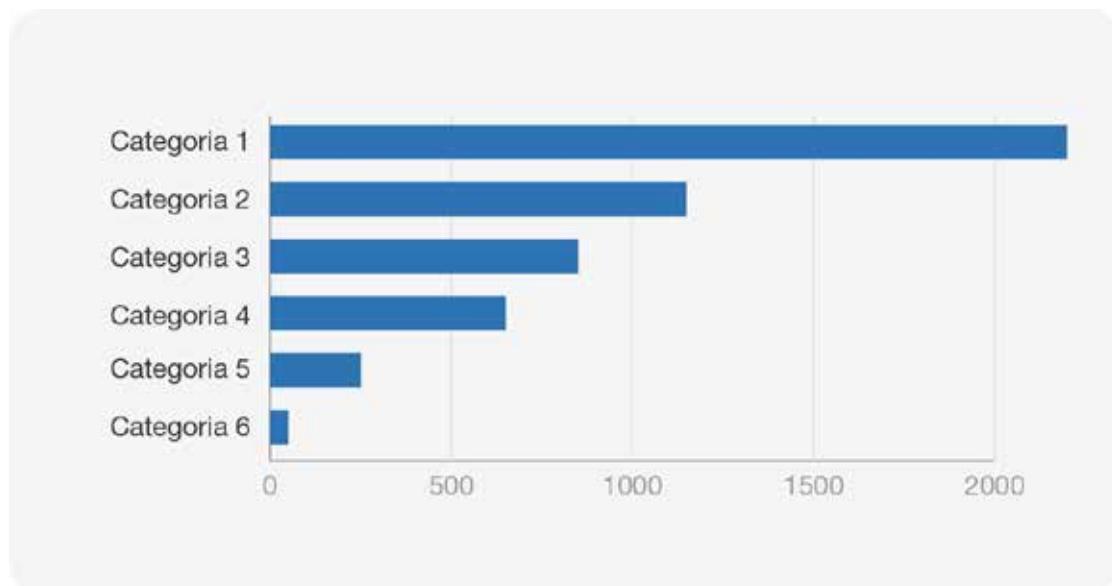


Figura 4.2. Exemple d'un gràfic de barres ordenat per facilitar la comunicació del rànquing de les categories.

Quan es representen dades temporals (i les categories són hores, dies, mesos, etc.) cal ordenar l'eix de manera temporal. En aquest cas, es recomana la disposició vertical de les barres.

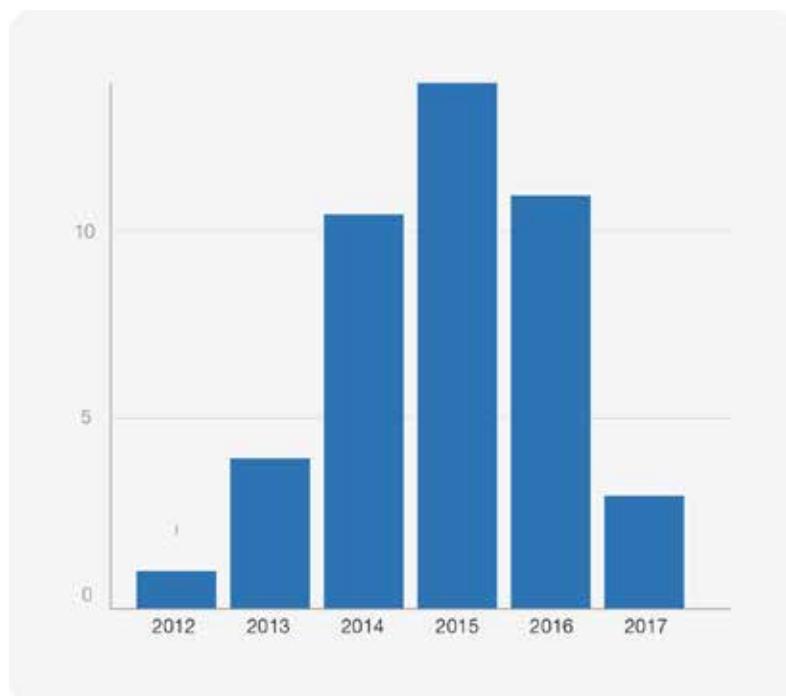


Figura 4.3. En cas de disposar de categories que tinguin un ordre temporal, serà millor utilitzar el gràfic de barres en vertical.

L'eix quantitatiu ha de començar sempre al zero, perquè, si no, les diferències de valors es magnifiquen i confonen el lector.



Figura 4.4. L'eix quantitatiu dels gràfics de barres sempre ha de començar al 0.

Quan els noms de les categories són llargs, és convenient orientar-lo horitzontalment per facilitar-ne la lectura.

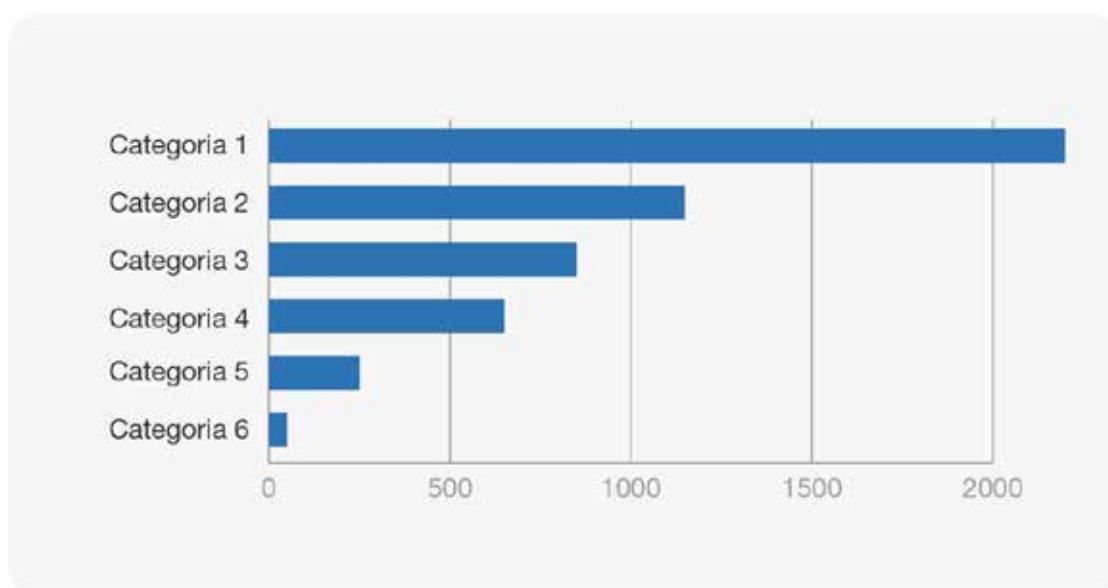


Figura 4.5. Exemple de gràfic de barres horitzontals per facilitar la lectura del nom de les categories.

Una altra pràctica interessant és afegir una línia de referència al gràfic de barres. Aquesta línia pot representar la mitjana, la mediana, els objectius o qualsevol altre valor que ajudi a descobrir en quant s'han superat o quant s'està per sota d'aquests.

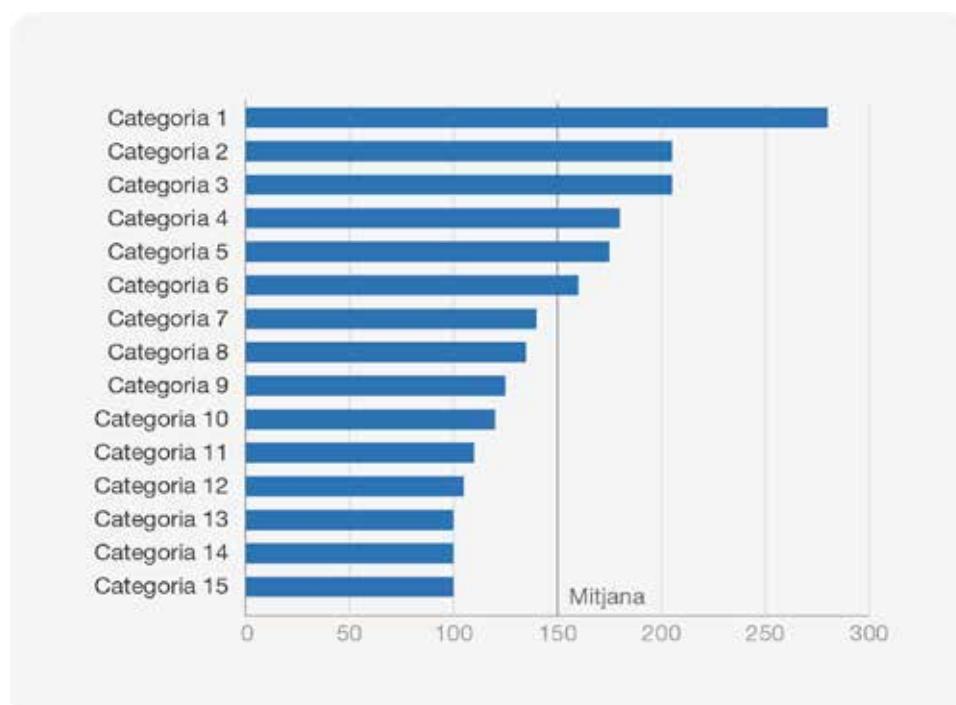


Figura 4.6. Exemple de gràfic de barres que mostra la mitjana.

Sovint es vol comparar un dels elements amb la resta. En aquests casos, és interessant ressaltar la barra que li correspon amb un color diferent, de manera que sigui més fàcil situar-lo i, per tant, comparar-lo amb la resta d'elements.

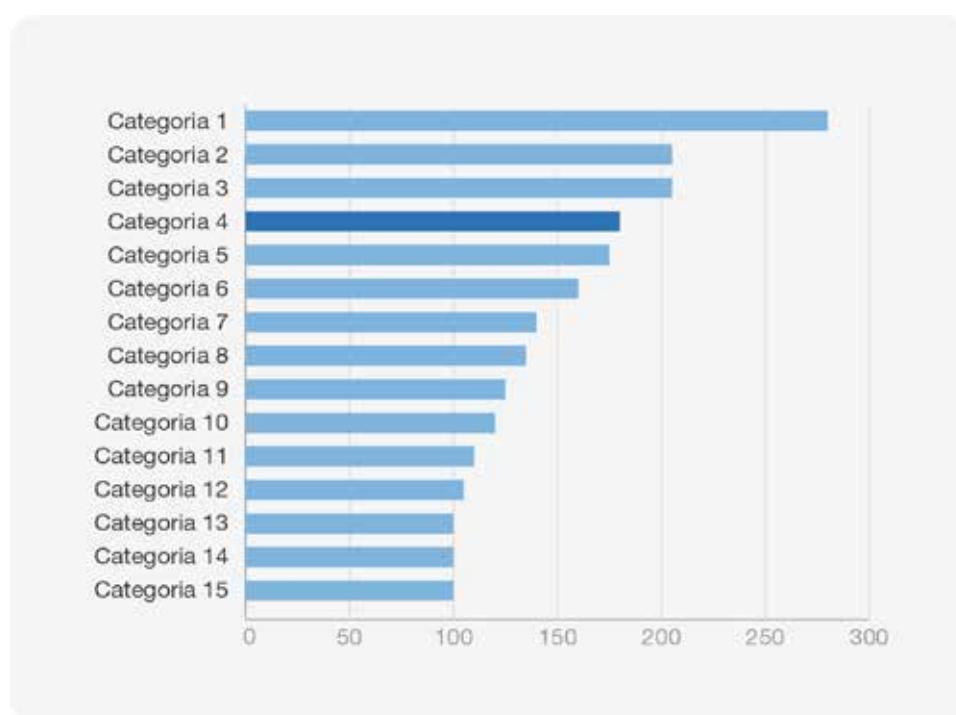


Figura 4.7. Exemple de gràfic de barres on es ressalta la categoria que es vol comparar amb les altres.

4.1.2. Gràfic de barres agrupades

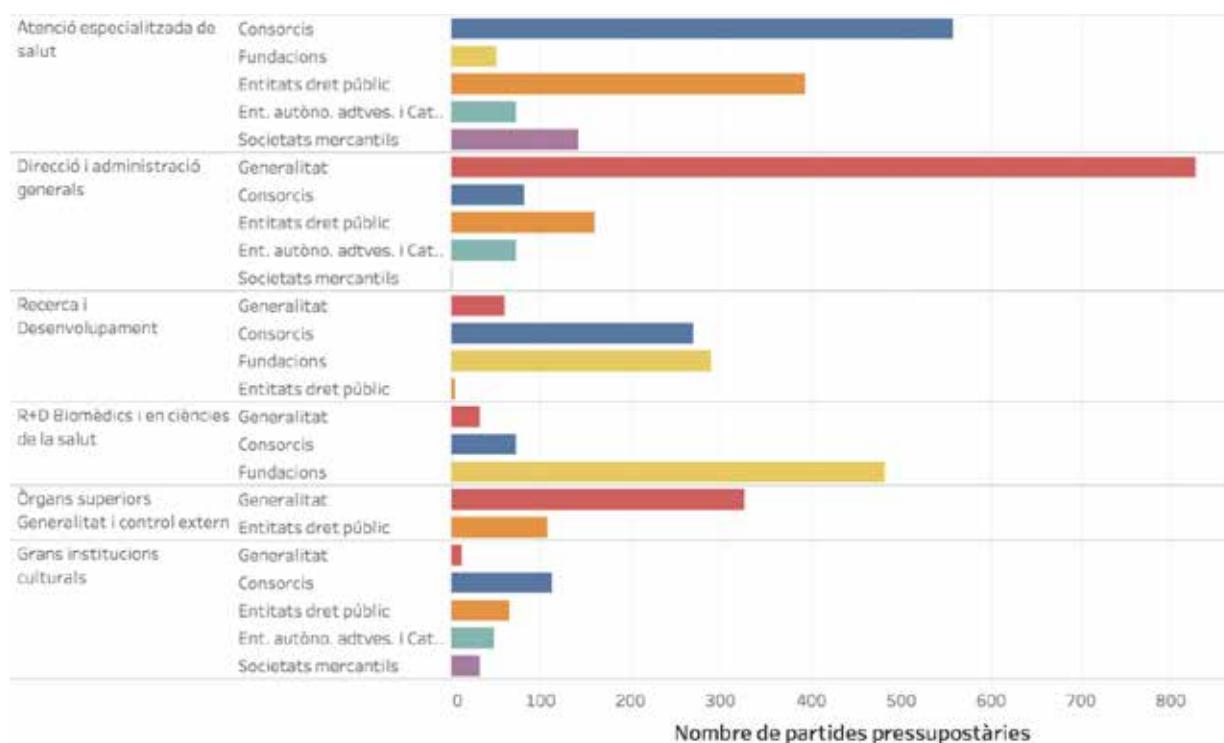


Figura 4.8. Nombre de partides pressupostàries per categoria i subcategoria.

Aquest gràfic és una versió del gràfic de barres estàndard. En aquest cas, consisteix en un conjunt de gràfics de barres disposats sobre un mateix eix.

Quan s'ha d'utilitzar?

Quan es vol comparar el valor dels diferents segments que formen part de les categories seleccionades. Aquest gràfic permet entendre la descomposició d'una categoria en els diferents segments que la componen (per exemple, com es reparteixen les partides pressupostàries de Recerca i Desenvolupament entre diferents subsectors, com es pot veure a l'exemple) i, alhora, es poden comparar les subcategories per diferents categories (per exemple, per a quin programa rep més partides el subsector de Fundacions).

Recomanacions

Per a aquest gràfic cal tenir en compte les mateixes recomanacions que les del gràfic de barres estàndard. Però és important no sobrecarregar-lo amb massa categories, ja que pot acabar generant problemes de comprensió. En casos així, és recomanable convertir el gràfic en diferents gràfics de barres estàndard separats.

4.1.3. Gràfic de barres apilades

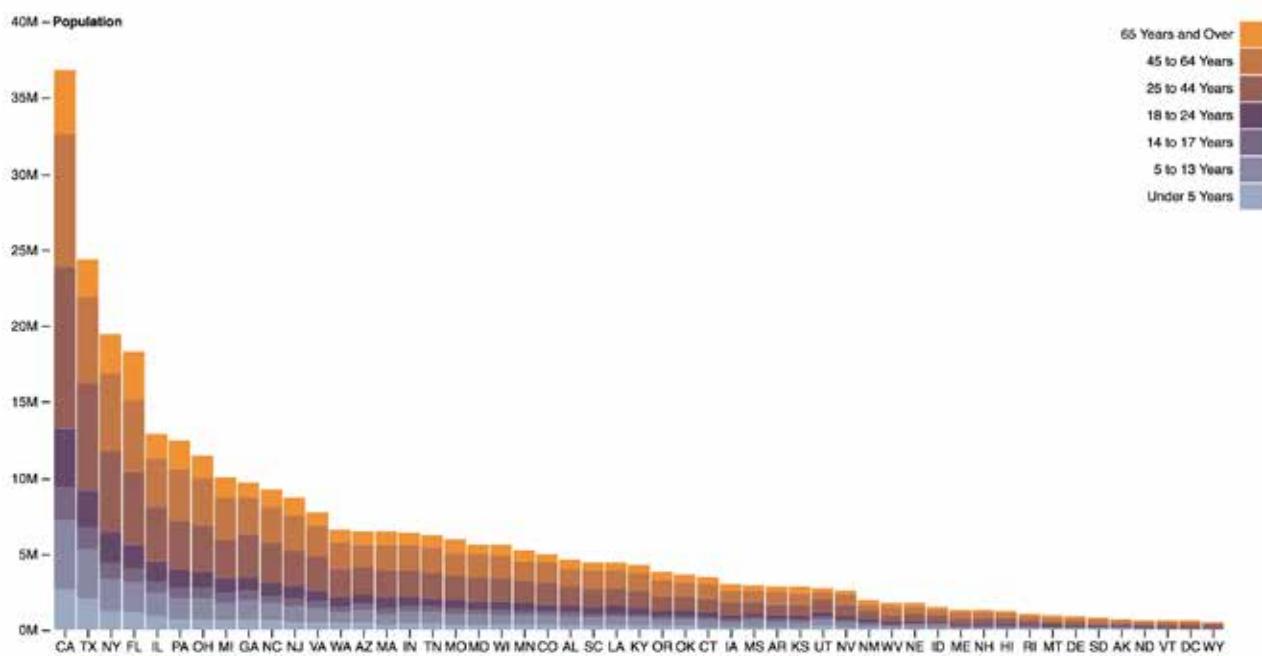


Figura 4.9. Població dels estats americans segmentats per rangs d'edat. Font: <https://bl.ocks.org/mbostock/3886208>.

Aquest tipus de gràfic representa una altra extensió del gràfic de barres estàndard. En aquest cas, es divideix cadascuna de les barres en funció dels segments que la componen.

Quan s'ha d'utilitzar?

El gràfic de barres apilades s'utilitza quan es vol comparar el valor total de la suma dels segments que formen cadascuna de les barres. Alhora, donen informació sobre com són de grans aquests segments.

Quan les barres apilades sumen un 100%, o sigui que cada barra segmentada ocupa tota l'alçada de la representació, el gràfic es pot considerar un gràfic que permet representar parts d'un total.

Recomanacions

Si es vol comparar la composició de les barres entre si, aquest gràfic no és una bona opció ja que, com es pot apreciar en la següent imatge, és difícil comparar el nombre de partides presupostàries destinades a cada subsector.

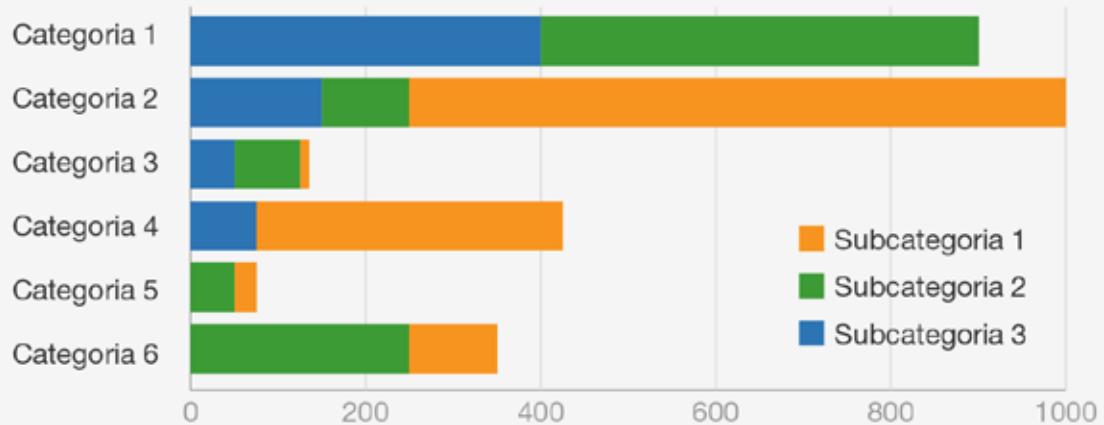


Figura 4.10. Exemple de barres apilades que representen una mala pràctica, ja que costa comparar els valors de les subcategories 1 i 2.

En aquest cas, es recomana descompondre el gràfic en diferents gràfics de barres, com es pot veure en el següent exemple.

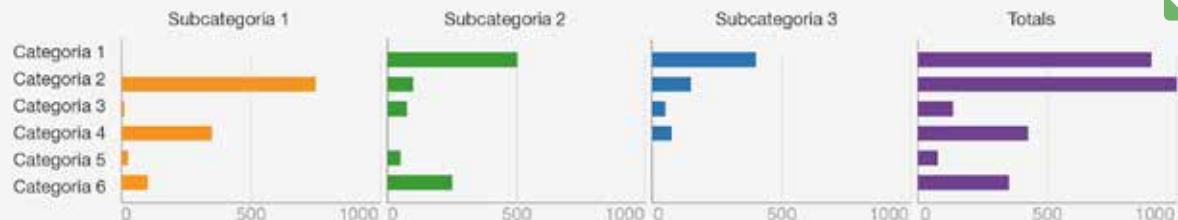


Figura 4.11. Una alternativa al gràfic de barres apilades és separar cada subcategoria en diferents gràfics de barres.

El nombre de segments que s'utilitzin per apilar les barres és molt important. En el cas d'utilitzar-ne només dos, i que a sobre formin part d'un total, del 100%, es podran obtenir dos gràfics complementaris que poden ser molt aclaridors.

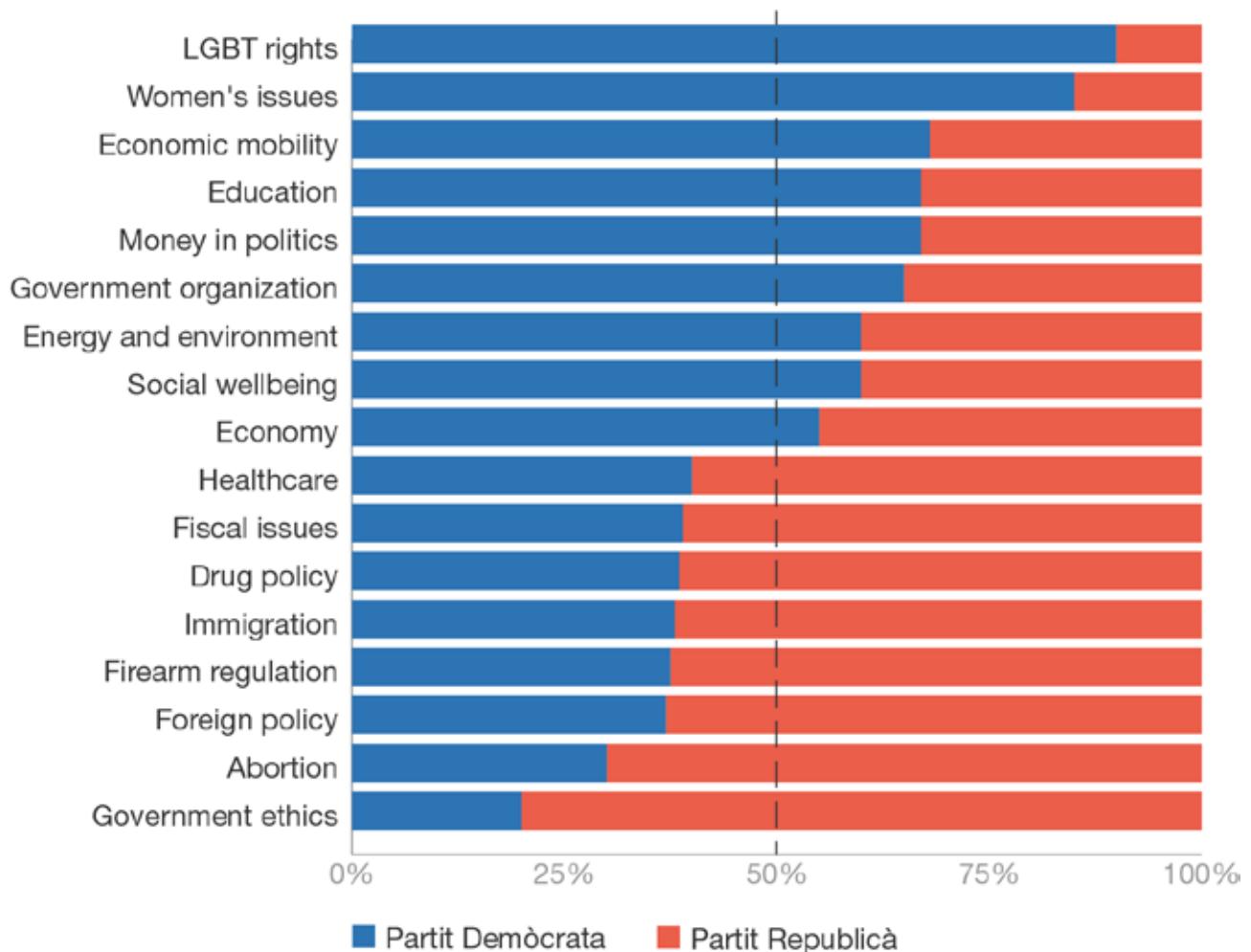


Figura 4.12. Adaptació del gràfic que mostra el percentatge de publicacions de facebook dedicades a diferents temes d'actualitat pel Partit Demòcrata i el Republicà dels Estats Units. Font: <https://www.facebook.com/notes/10152581594083859/>.

4.1.4. Gràfic de radar

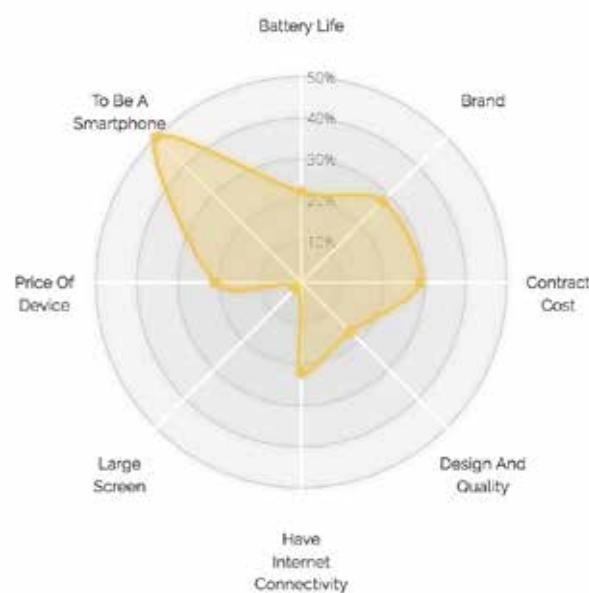


Figura 4.13. Exemple de gràfic de radar que representa les característiques més valorades en un telèfon.
Font: <http://bl.ocks.org/nbremer/21746a9668ffdf6d8242>

Els gràfics de radar mostren de manera circular un conjunt d'eixos, cadascun relatiu a una variable del nostre conjunt de dades. El valor que pren cada variable se situa en un punt del seu eix corresponent i, després, es traça una línia que uneix aquests punts. Sovint, s'acoloreix l'àrea que crea el polígon resultant.

El següent diagrama detalla l'estructura d'un gràfic de radar:



Figura 4.14. Parts que formen un gràfic de radar.

En essència, aquest gràfic converteix un conjunt de valors en una forma intuitiva i fàcil d'interpretar.

Quan s'ha d'utilitzar?

Aquest tipus de gràfic s'utilitza quan es volen entendre els valors d'un element en diferents variables alhora. Per exemple, es podria representar el rendiment d'un jugador de futbol a partir de diferents variables. En el cas que volguéssim comparar diferents jugadors, podríem fer servir un gràfic de radar per cada jugador.

Recomanacions

S'ha d'evitar mostrar massa variables alhora, de manera que no es generin formes massa angulades o estranyes, dificultant la lectura i la comparació entre diferents gràfics.

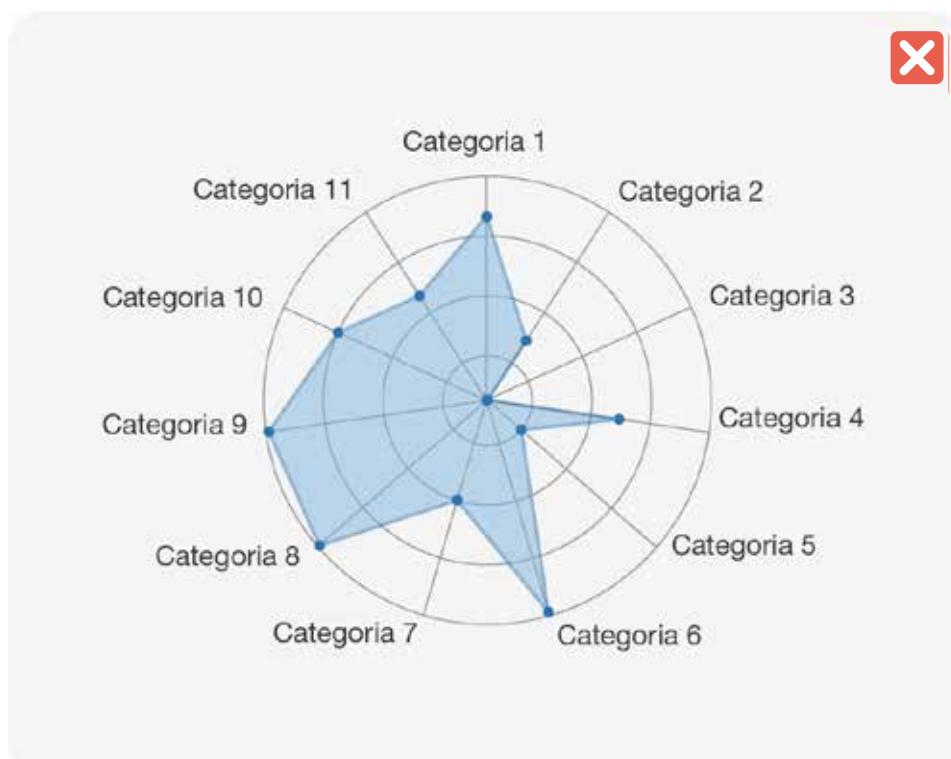


Figura 4.15. Exemple de gràfic de radar amb massa variables on es generen polígons massa angulats poc reconeixibles.

Aquest tipus de gràfic és especialment útil quan l'ordre en què se situen les diferents variables té algun significat. Per exemple, en ànalisis futbolístiques, les variables relacionades amb l'atac d'un equip (nombre de gols, xuts a porteria, etc.) se situen en els eixos que queden a la part superior del gràfic, mentre que aquelles variables de caire defensiu (nombre d'aturades, faltes, intercepcions de pilota, etc.) se situen a la part inferior. D'aquesta manera, un gràfic de radar que mostri un cercle quasi perfecte, indicarà un equip que és molt bo tant en atac com en defensa. Si, per contra, el gràfic és més complet a la part superior, indicarà que l'equip és millor en atac que en defensa.

Aquest gràfic no permet representar variables que puguin tenir valors negatius i positius.

4.1.5. Gràfic d'intensitat de colors



Figura 4.16. Gràfic d'intensitats de colors que representa el nombre d'infectats per malària en diferents països. Imatge cedida per OneTandem. Font: https://public.tableau.com/views/malaria_5/Dashboard1.

Els gràfics d'intensitat de colors (*heatmap*) són una evolució de les taules. En lloc de representar els valors utilitzant nombres, aquests es representen mitjançant la intensitat del color de la cel·la que ocupen.

Quan s'ha d'utilitzar?

Quan es vol descobrir la variabilitat d'un conjunt de variables. El gràfic d'intensitats de colors també ajuda a revelar patrons quan mostra si diferents variables es comporten igual i, fins i tot, pot ajudar a trobar correlacions entre elles.

Recomanacions

Ordenar les files i les columnes del gràfic d'intensitats de colors segons un criteri establert pot ser molt interessant per facilitar el descobriment d'elements que tinguin dades similars. Seguint l'exemple anterior, caldria ordenar els països per nombre d'infectats i els anys seguint un ordre temporal.

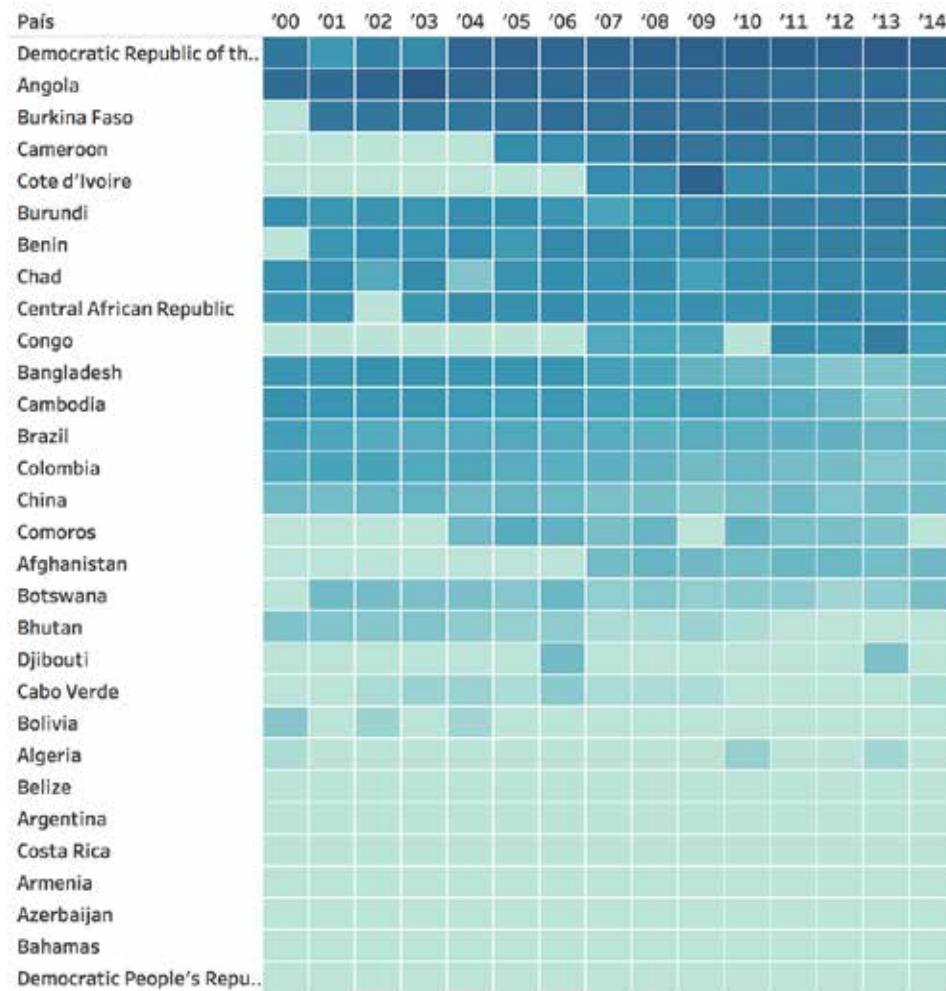


Figura 4.17. Gràfic d'intensitats de colors amb l'evolució d'infectats per malària en alguns països del món, ordenats segons la mitjana anual d'infectats per país. Adaptació de la imatge cedida per OneTandem.

Font: https://public.tableau.com/views/malaria_5/Dashboard1.

Si el gràfic d'intensitats de colors s'utilitza per a un conjunt de dades on les seves variables tenen escales diferents, és aconsellable normalitzar-les, de manera que totes passin a tenir el mateix rang de valors. En aquest article es representen diferents indicadors de jugadors de l'NBA: <http://flowingdata.com/2010/01/21/how-to-make-a-heatmap-a-quick-and-easy-solution/>.

4.1.6. Gràfic de marques

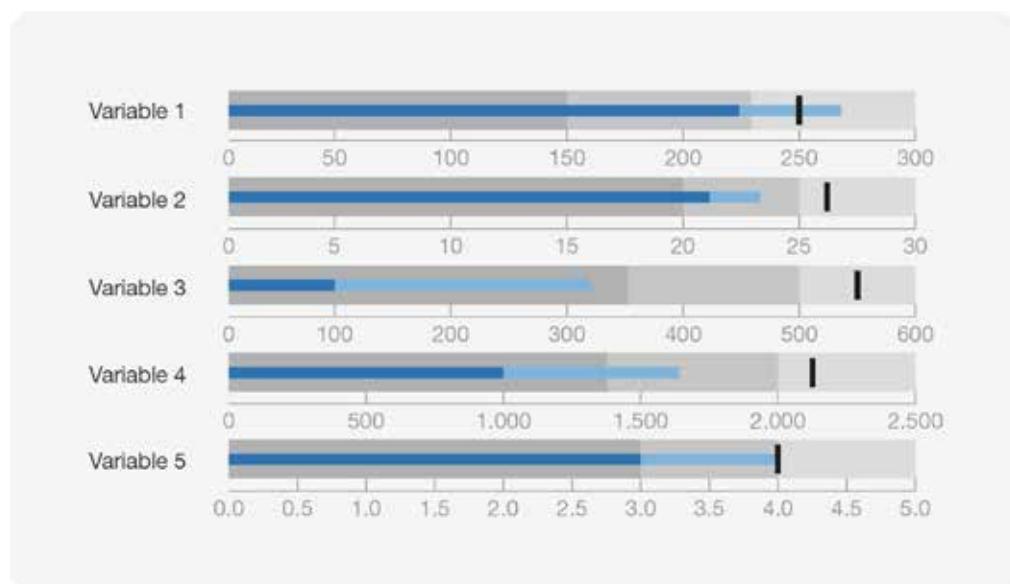


Figura 4.18. Exemple d'un gràfic de marques mostrant cinc categories diferents.

Els gràfics de marques (*bullet chart*) són una modificació del gràfic de barres, però amb dos elements que ajuden a proporcionar context:

- Una marca vertical que representa un valor que permet comparar la variable representada (per exemple, si la barra indica vendes, la marca vertical podria indicar l'objectiu de vendes).
- Un fons amb àrees de diferents tonalitats, que indiquin els rangs de valors “acceptables”, “bons” o “dolents” que pot prendre la variable en qüestió.

El següent diagrama detalla l'estructura d'un gràfic de marques:

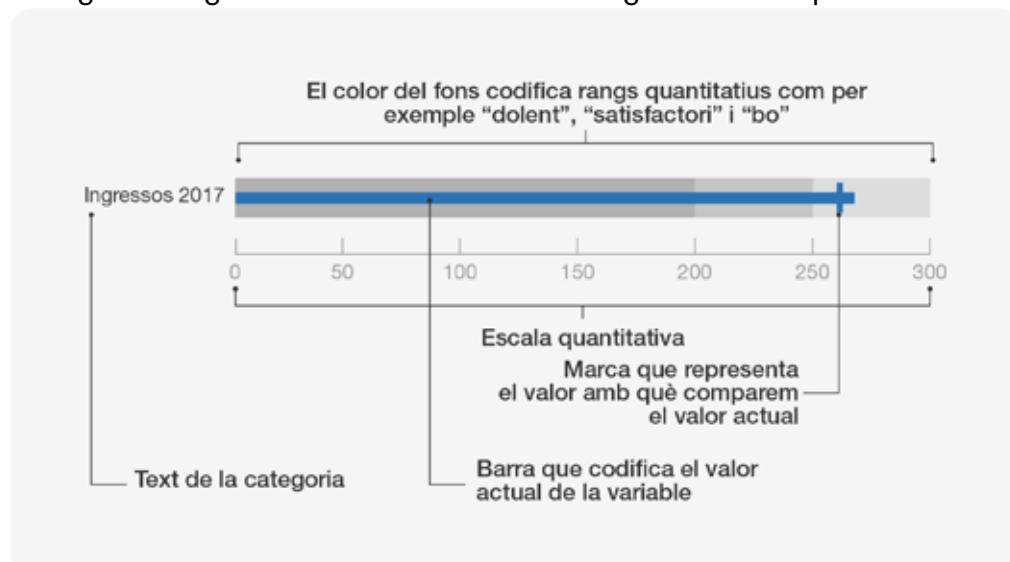


Figura 4.19. Parts que formen un diagrama de marques.

Quan s'ha d'utilitzar?

Els gràfics de marques s'utilitzen quan es vol avaluar un valor concret respecte del que s'esperava i respecte d'uns rangs de qualitat establerts.

La particularitat del gràfic de marques és que permet no només avaluar el valor de la variable representada, sinó comparar-la amb un valor esperat, de manera que contextualitza el valor representat i ajuda a la seva interpretació. De la mateixa manera, les bandes pintades del fons també ajuden a saber si el valor es troba dins d'un rang identificat com a acceptable, bo o dolent.

Recomanacions

Sovint l'objectiu que es marca per a un indicador és anual (per exemple, objectiu de vendes anuals). Per tant, si es fa aquest gràfic, per exemple, a meitat d'any, el més freqüent és no arribar a l'objectiu. Per això, hi ha una versió del gràfic que mostra el valor actual (blau fosc a l'exemple inicial) i el valor esperat (blau clar).

4.1.7. Gràfic múltiple

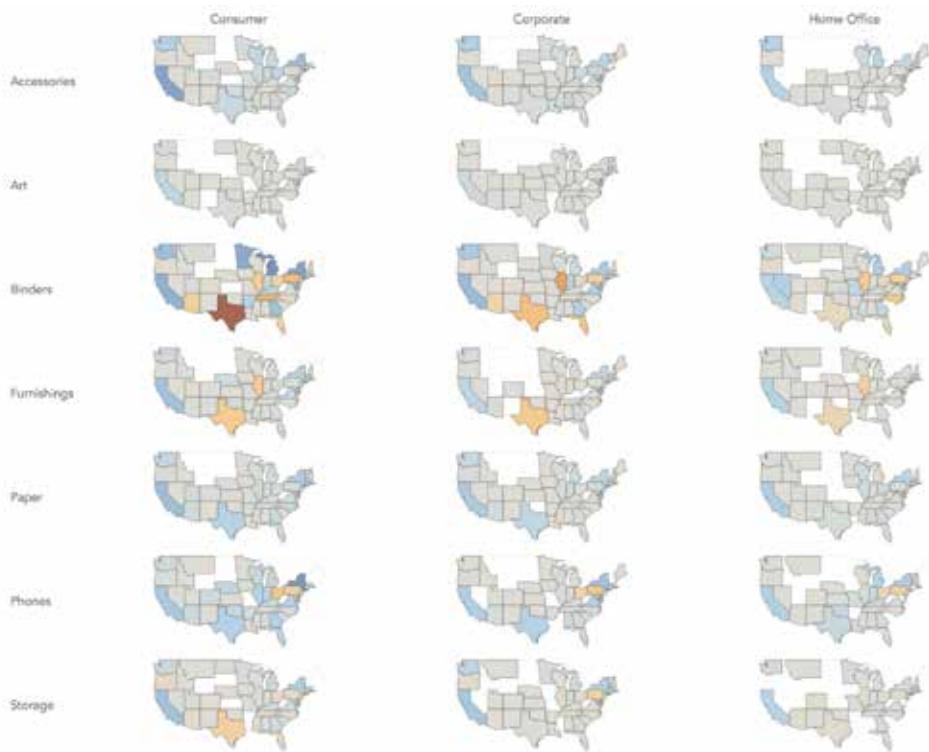


Figura 4.20. Exemple de l'ús del gràfic múltiple: una cadena de botigues de material d'oficina fa servir aquesta tècnica per comparar els beneficis de cada estat dels Estats Units per diferents categories de producte i pels seus tres tipus de clients.

Més que un tipus de gràfic, el concepte del gràfic múltiple (*small multiples*) fa referència a una tècnica de composició de gràfics. Concretament, aquesta tècnica proposa utilitzar el mateix gràfic repetidament, per mostrar el comportament d'una mateixa variable per diferents categories. D'aquesta manera, gràcies a la repetició i al fet que els gràfics estan molt junts, es facilita la comparació entre ells.

Quan s'ha d'utilitzar?

Quan representar totes les dades de les diferents categories en un únic gràfic genera representacions complexes. Això acostuma a passar quan hi ha massa categories que dificulten la comprensió.

També es pot fer servir el gràfic múltiple per representar dades en el temps i mostrar diferents gràfics per a cada moment determinat.

Recomanacions

És important que el tipus de gràfic utilitzat sigui prou clar perquè s'entengui el seu significat tot i la seva petita mida.

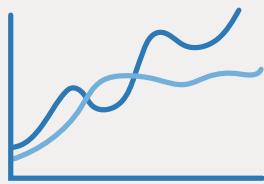
També cal ser coherent en l'ús dels rangs als eixos i als colors, de manera que sigui fàcil comparar tots els gràfics entre si.

4.2. Tendències

Els gràfics de tendència permeten veure l'evolució temporal d'una o més variables quantitatives. És un dels tipus d'anàlisi més comuns ja que, sovint, no només interessa veure el valor que té un element en una variable, sinó que volem entendre quin comportament ha tingut al llarg del temps.

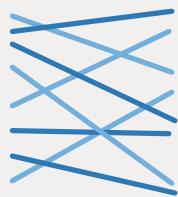
Gràfic de línies

Per veure l'evolució temporal d'una o més variables.



Gràfic de pendents

Per comparar l'abans i el després d'una variable en varíes categories.



Mini-gràfic de línia

Per contextualitzar un indicador.



4.2.1. Gràfic de línies



Figura 4.21. Exemple de gràfic de línies que representa l'evolució del total de vendes de dos productes d'una empresa.

El gràfic de línies permet representar el canvi d'una variable al llarg del temps. La variable temporal o contínua se situa a l'eix x , i es construeix disposant una sèrie de punts connectats per una línia segons l'alçada que marca la variable col·locada a l'eix y .

Un gràfic de línies pot contenir una o més línies, de manera que permet comparar l'evolució de dades de diferents categories, sovint representades amb colors diferents.

En cas que només hi hagi una línia, l'àrea que queda entre aquesta i l'eix x es pot pintar d'un color concret; així obtindrem el que s'anomena un gràfic d'àrea.

Quan s'ha d'utilitzar?

Quan volem veure l'evolució temporal d'una o més variables.

Recomanacions

Cal ser molt curós amb el rang de valors utilitzat en l'eix y , ja que es corre el risc d'emfatitzar massa els canvis de la variable en el temps. Aquest efecte es pot observar en els gràfics següents, que mostren l'evolució del nombre d'infectats per malària al llarg del temps. El gràfic de l'esquerra té com a valor mínim de l'eix y el 0, mentre que el de la dreta comença a 60.000 infectats. Com es pot observar, en el gràfic de la dreta els pendents de les línies són molt més pronunciats i, per tant, donen una sensació de grans canvis.

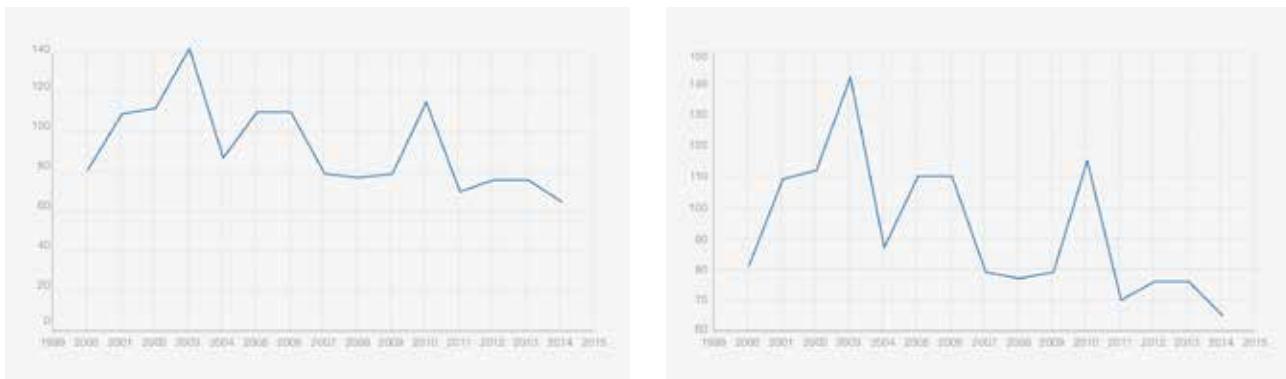


Figura 4.22. Exemple de dos gràfics de barres amb diferents valors a l'eix y.

En general, és una bona idea afegir línies de referència en forma de retícula que ajudin a estimar els valors de les variables de cada punt.

El gràfic de línies també fa que es pugui mostrar més d'una línia alhora, com s'ha vist al primer exemple. No obstant això, la representació de massa línies pot reduir considerablement la llegibilitat del gràfic.

4.2.2. Gràfic de pendents

Com ha variat la classificació de la lliga entre 2015-16 i 2016-17?

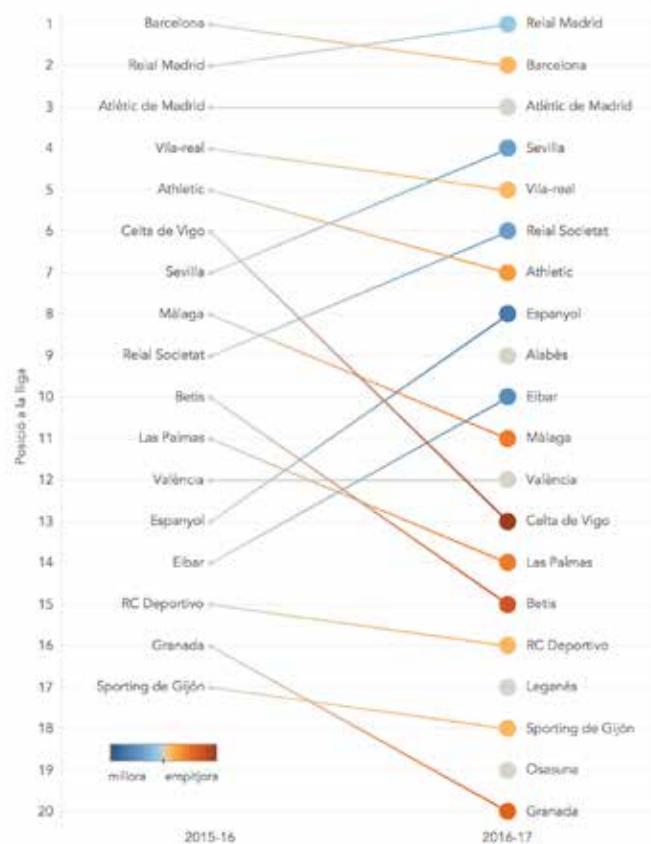


Figura 4.23. Exemple de gràfic de pendents que mostra la classificació dels diferents equips de la primera divisió de futbol en les temporades 2015-16 i 2016-17.

El gràfic de pendents (*slope chart*) és un gràfic de línies amb només dos punts i està pensat per mostrar, a simple vista, l'abans i el després d'una variable concreta. Generalment, un gràfic de pendents acostuma a contenir més d'una línia per facilitar la comparació entre els valors de diferents categories.

Quan s'ha d'utilitzar?

Quan es vol comparar l'abans i el després d'una variable en diverses categories. A l'exemple anterior es pot veure la diferència de classificació dels diferents equips de la lliga espanyola en les temporades 2015-2016 i 2016-17.

Recomanacions

Perquè el gràfic de pendents sigui coherent, els valors màxims i mínims de cadascun dels eixos ha de ser el mateix.

4.2.3. Minigràfic



Figura 4.24. Minigràfics utilitzats a la capçalera d'un quadre de comandament que mostren l'evolució dels indicadors principals d'una empresa.

Un minigràfic (*sparkline*) és un gràfic dissenyat per donar context a una xifra. Generalment s'utilitzen en quadres de comandament i es posen al costat del valor d'un indicador clau del sistema, de manera que han de ser gràfics més aviat petits.

Tot i que inicialment els minigràfics eren bàsicament petites línies de tendència que ajudaven a entendre l'evolució d'una determinada mètrica, el concepte s'ha estès en els darrers anys i ha permès la utilització d'altres tipus de gràfics, com per exemple els gràfics de barres.

Quan s'ha d'utilitzar?

Quan es vol contextualitzar un indicador. Generalment s'utilitzen per representar patrons o tendències en el temps, utilitzant un espai petit al costat de l'indicador.

Recomanacions

Com a norma general, l'escala dels eixos d'un minigràfic no es mostra, ja que l'objectiu és poder veure el comportament del valor mostrat a grans trets.

4.3. Mapes

Els mapes són un dels tipus de gràfics més utilitzats. La seva principal funció és ajudar a trobar patrons geogràfics en les dades.

A més a més, els mapes són projeccions que intenten representar de la manera més fidedigna possible l'esfera de la Terra en dues dimensions. Per això hi ha moltes projeccions matemàtiques, cadascuna amb les seves problemàtiques particulars, que possibiliten la generació dels mapes que fem servir avui dia. En general, els programes que ens permeten mostrar dades sobre mapes ja porten incorporats mapes que utilitzen unes projeccions que deformen al mínim la realitat⁷.

Mapa de corolletes

Per trobar patrons geogràfics a les dades que estan categoritzades per zones o regions.



Mapa de símbols proporcionals

Per trobar patrons geogràfics en unes dades que contenen punts geogràfics específics amb latitud i longitud.



⁷ Més informació sobre les projeccions de mapes: Wiquipèdia, Projecció cartogràfica.
<https://ca.wikipedia.org/wiki/Projecci%C3%B3_cartogr%C3%A0fica> [Consulta: 21 novembre 2017]

4.3.1. Mapa de coropletes

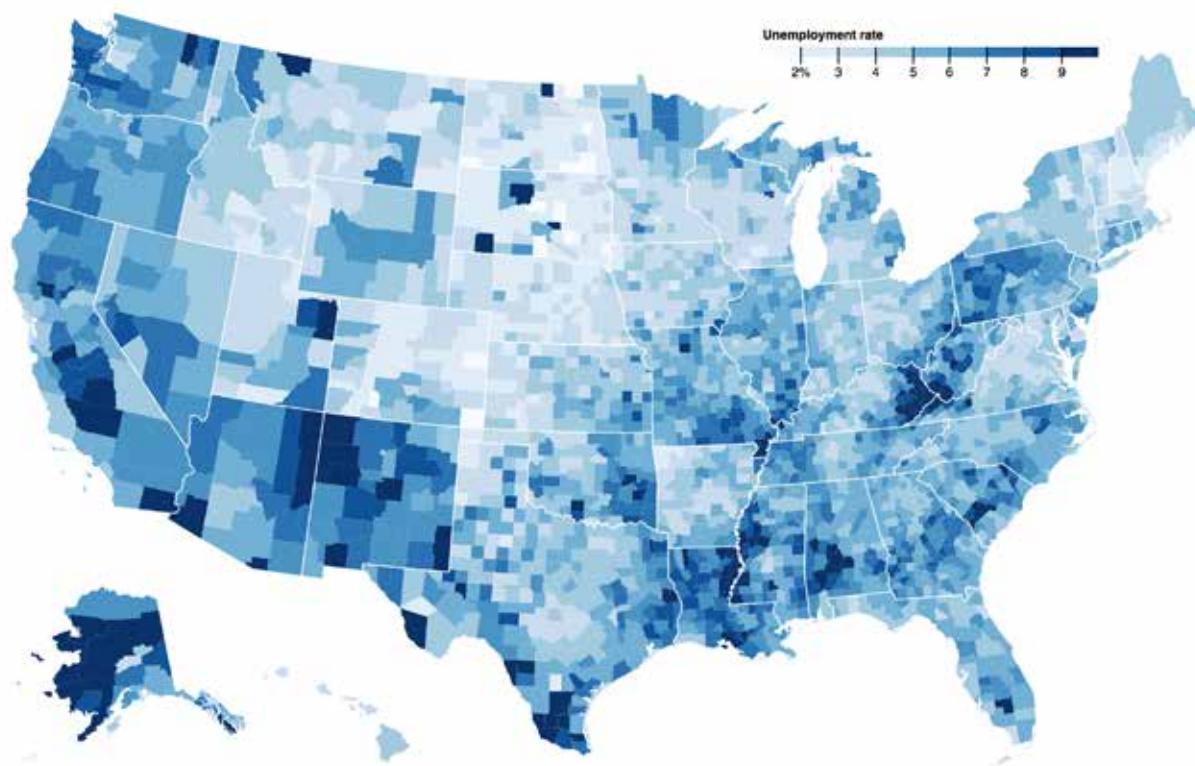


Figura 4.25. Mapa de coropletes que representa la taxa d'atur als Estats Units l'agost del 2016.
Font: <https://bl.ocks.org/mbostock/4060606>.

Els mapes de coropletes mostren els valors d'una variable sobre un mapa tot pintant les àrees de cada regió afectada d'un color determinat. Els colors s'utilitzen per representar una variable numèrica o bé per representar la pertinença d'una regió a una categoria concreta. En funció de l'objectiu que se cerqui s'utilitzarà una escala de colors o una altra.

Quan s'ha d'utilitzar?

Quan es volen trobar patrons geogràfics a les dades que estan categoritzades per zones o regions.

Recomanacions

En general, un mapa s'ha d'utilitzar per resoldre dubtes de caire geogràfic. Per tant, un mapa de coropletes no serà l'adequat si el que es desitja és veure un rànquing de les diferents regions.

4.3.2. Mapa de símbols proporcionals

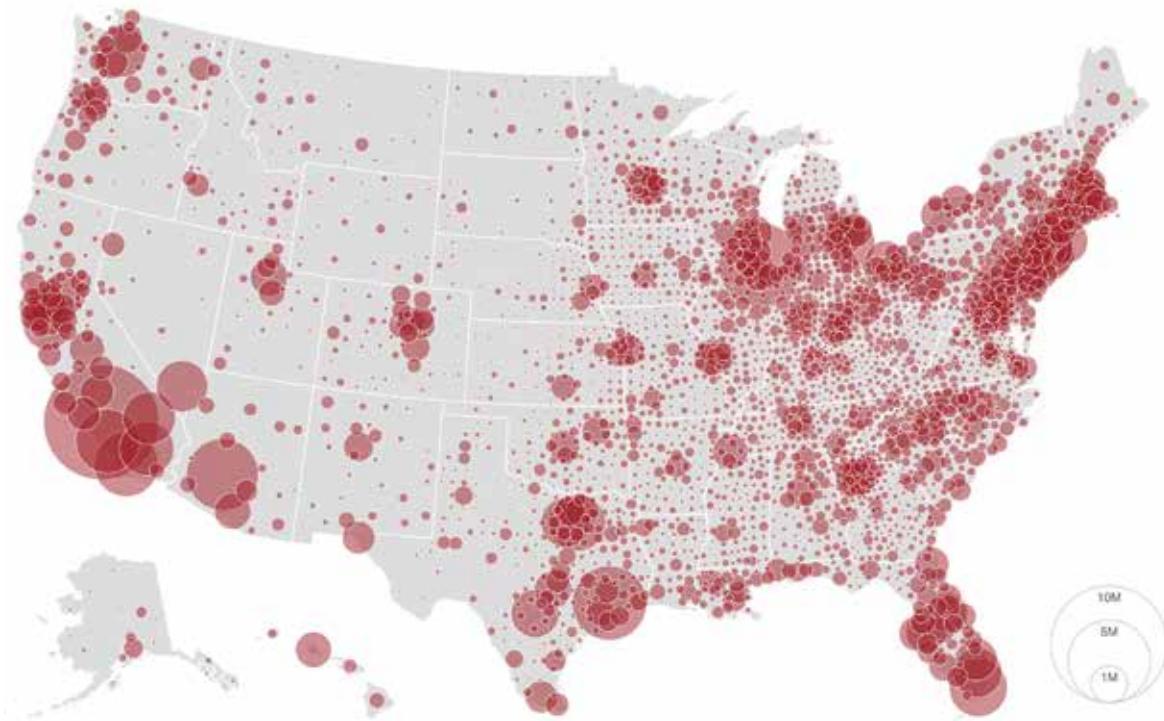


Figura 4.26. Mapa de símbols proporcionals que mostra el nombre de persones censades als Estats Units.
Font: <https://bostocks.org/mike/bubble-map/>.

El mapa de símbols proporcionals consisteix en un mapa on se situa una icona o símbol, generalment un cercle, de mida proporcional a la variable que es representa a sobre del centre de la regió a la qual correspon.

A diferència del mapa de coropletes, el mapa de símbols proporcionals permet representar dades geogràfiques sense necessitat que representin una àrea política. Alhora, es pot utilitzar el color per afegir una segona variable al gràfic, o per emfatitzar la variable utilitzada en la mida.

Quan s'ha d'utilitzar?

Quan se cerquen patrons geogràfics en unes dades que contenen punts geogràfics específics amb latitud i longitud.

Recomanacions

Com amb els mapes de coropletes, un mapa de símbols proporcionals s'ha d'utilitzar únicament quan es volen trobar patrons geogràfics, i no quan es vol crear un rànquing de llocs geogràfics a partir d'una variable.

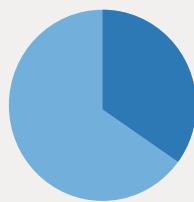
Cal tenir en compte que, tal i com es pot veure a l'exemple, sovint aquest tipus de mapes generen un encavalcament entre els símbols utilitzats per representar els valors de la variable seleccionada. En aquests casos, cal utilitzar la transparència per facilitar la lectura del gràfic i identificar els símbols que queden els uns a sota dels altres.

4.4. Parts d'un total

Analitzar dades que formen part d'un total és una tasca que fem molt sovint. Per exemple, quan volem entendre la contribució de cada departament al total de despeses d'una organització, o quan volem analitzar les respistes d'una enquesta. A continuació, descrivim una sèrie de gràfics que són molt útils per comunicar aquest tipus d'informació.

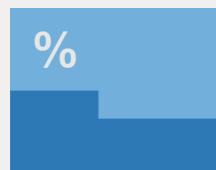
Gràfic de sectors

Per mostrar com es distribueixen proporcionalment les nostres dades.



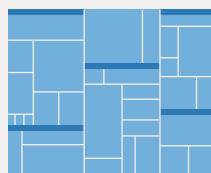
Pictograma

Per expressar un valor i alhora oferir un context que ens permeti identificar si és gran o petit.



Mapa d'arbre

Per representar dades jeràrquiques i comparar una o dues variables entre els diferents elements.



4.4.1. Gràfic de sectors

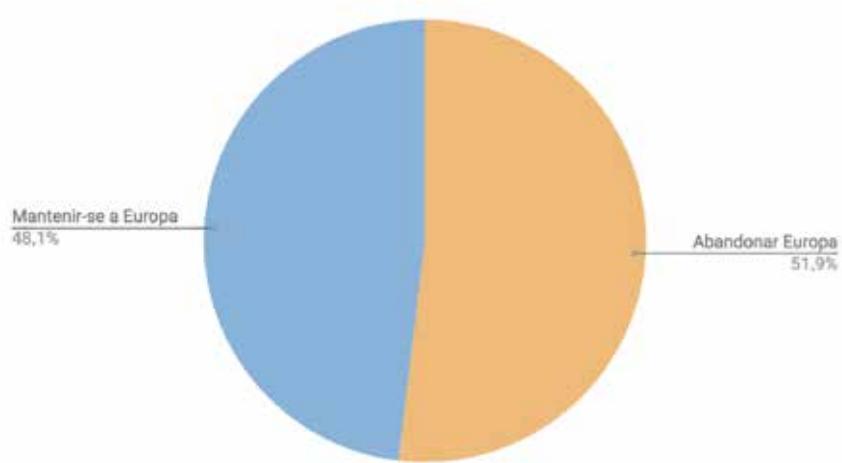


Figura 4.27. Resultat de la votació del Brexit al Regne Unit.

També anomenat *gràfic de pastís*, el gràfic de sectors utilitza una representació circular que es divideix en sectors proporcionals als valors de cada categoria representada.

Quan s'ha d'utilitzar?

El gràfic de sectors s'utilitza quan es vol mostrar com es distribueixen proporcionalment les dades. És un gràfic ideal quan hi ha dos valors, com per exemple el nombre de persones que contesten *sí* o *no* a una pregunta.

També és especialment útil quan es vol comparar un valor amb la suma dels altres, com es pot apreciar en la següent imatge, on queda clar que el sector taronja és més gros que la suma de la resta de valors.

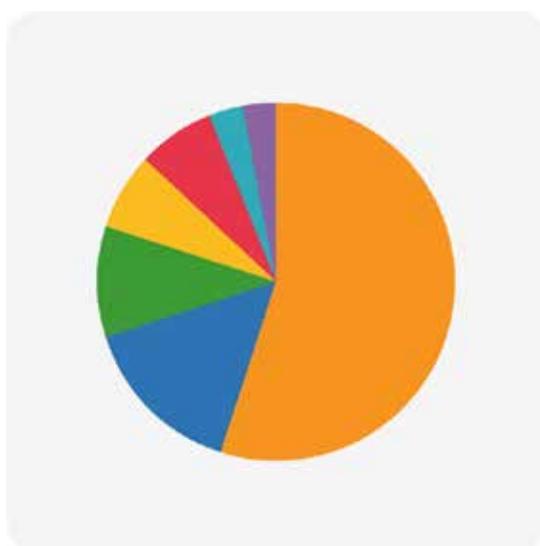
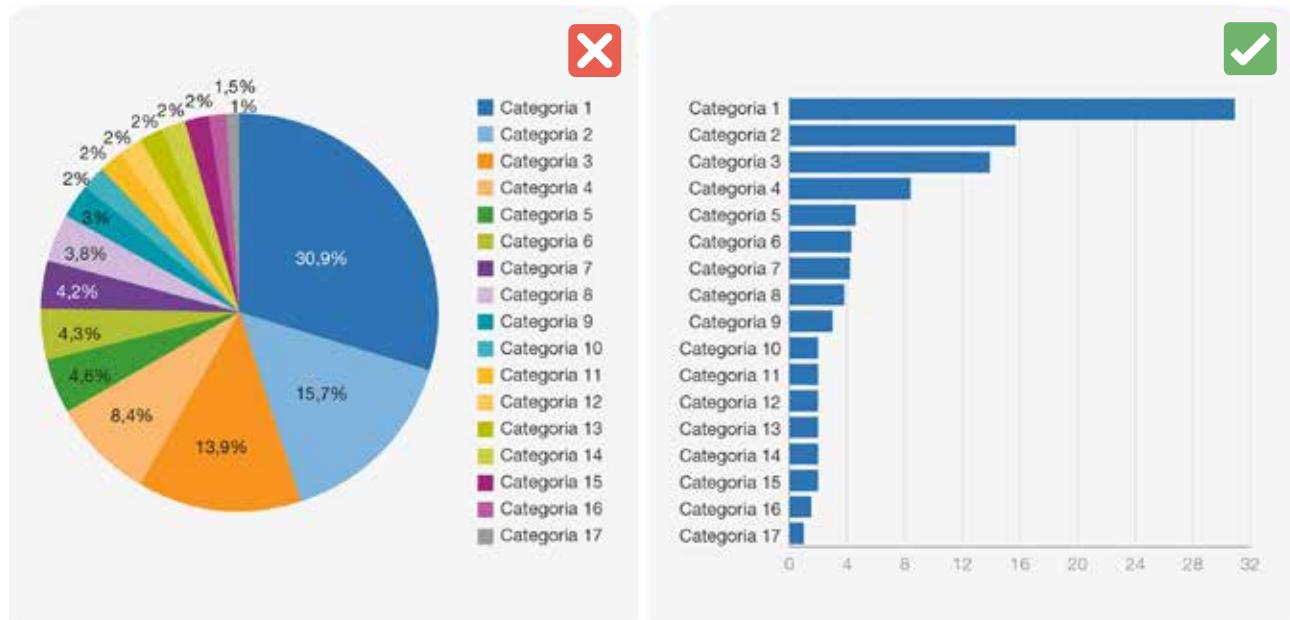


Figura 4.28. En aquest gràfic de sectors podem veure com la categoria taronja ocupa més de la meitat del total, una xifra a la qual no arriben la resta de sectors junts.

Recomanacions

Excepte en els casos esmentats abans, no és recomanable utilitzar el gràfic de sectors per representar més de dues o tres parts d'un total. En aquest cas, és molt millor fer servir un gràfic de barres.



Figures 4.29. Els gràfics de sectors amb moltes divisions no són bones representacions ja que confonen i no ajuden a entendre la diferència entre els valors. Font: <http://www.conceptdraw.com/solution-park/resource/images/solutions/pie-charts/Graphs-and-Charts-Pie-Chart-Business-Report.png>.

L'ús del 3D és especialment popular per a aquest tipus de gràfic; tot i això, aquesta tècnica encara dificulta més la seva comprensió per culpa de la perspectiva.

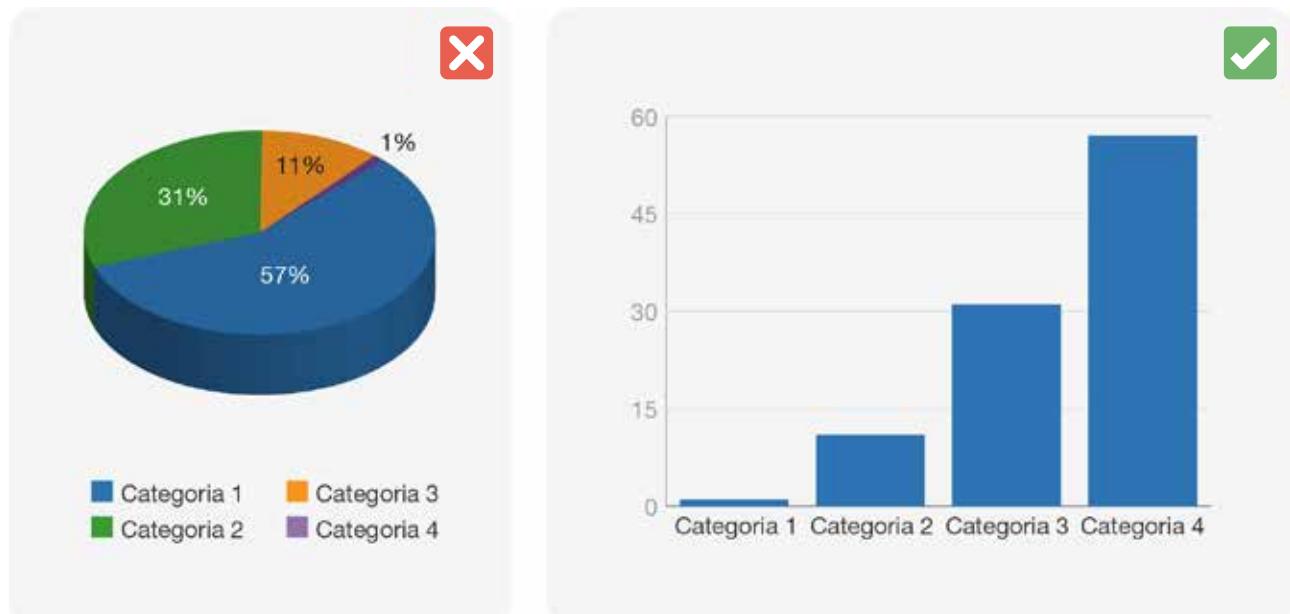


Figura 4.30. Gràfic de sectors en 3D representat amb un gràfic de barres. Adaptació de la imatge cedida per Ann K. Emery. Font: <http://annkemery.com/pie-chart-guidelines/>.

Si es tenen dades ordinals (que segueixen un ordre seqüencial), és preferible fer servir un gràfic de barres. Així, es pot jugar amb els extrems per mostrar la gradació (en l'exemple, des de l'extrem de gran acord a l'extrem de gran desacord).



Figura 4.31. Dades ordinals representades amb un gràfic de barres apilades. Adaptació de la imatge cedida per Ann K. Emery.
Font: <http://annkemery.com/pie-chart-guidelines/>.

Si es tenen dades que segueixen un ordre temporal, és preferible fer servir un gràfic de línies o de barres.

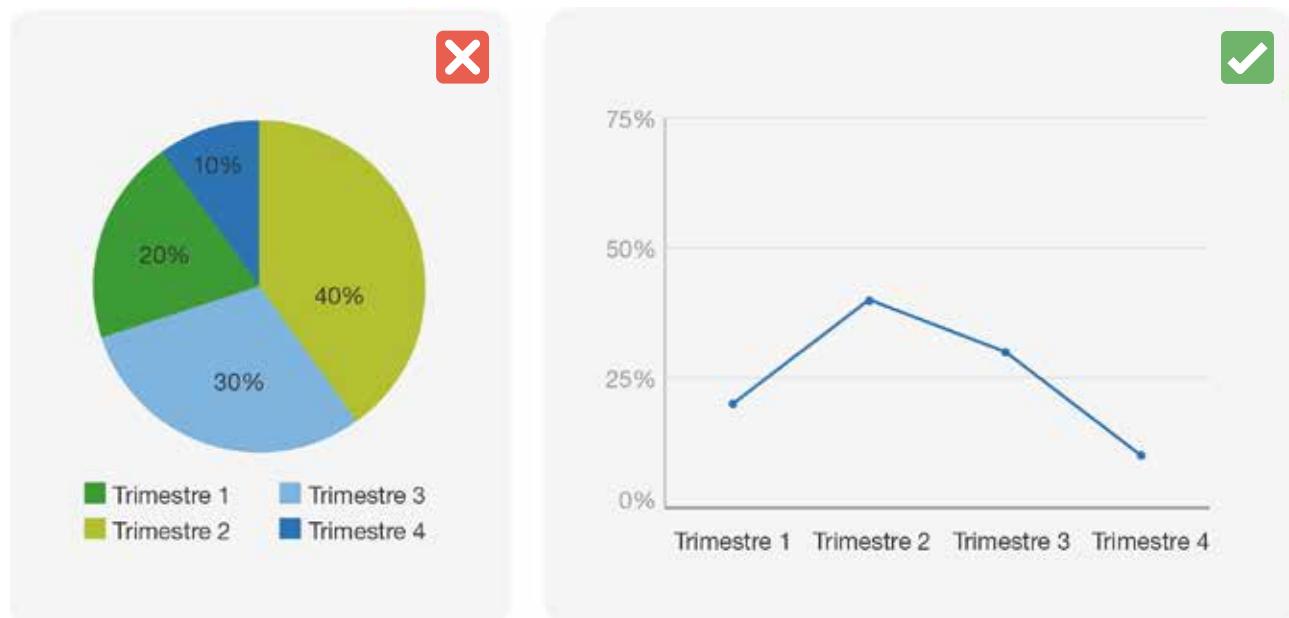


Figura 4.32. És millor representar les dades temporals amb un gràfic de línia. Adaptació de la imatge cedida per Ann K. Emery.
Font: <http://annkemery.com/pie-chart-guidelines/>.

Si s'han de comparar diferents categories, novament és preferible fer servir barres, que faciliten la comparació de la seva composició, ja que es poden ordenar visualment.

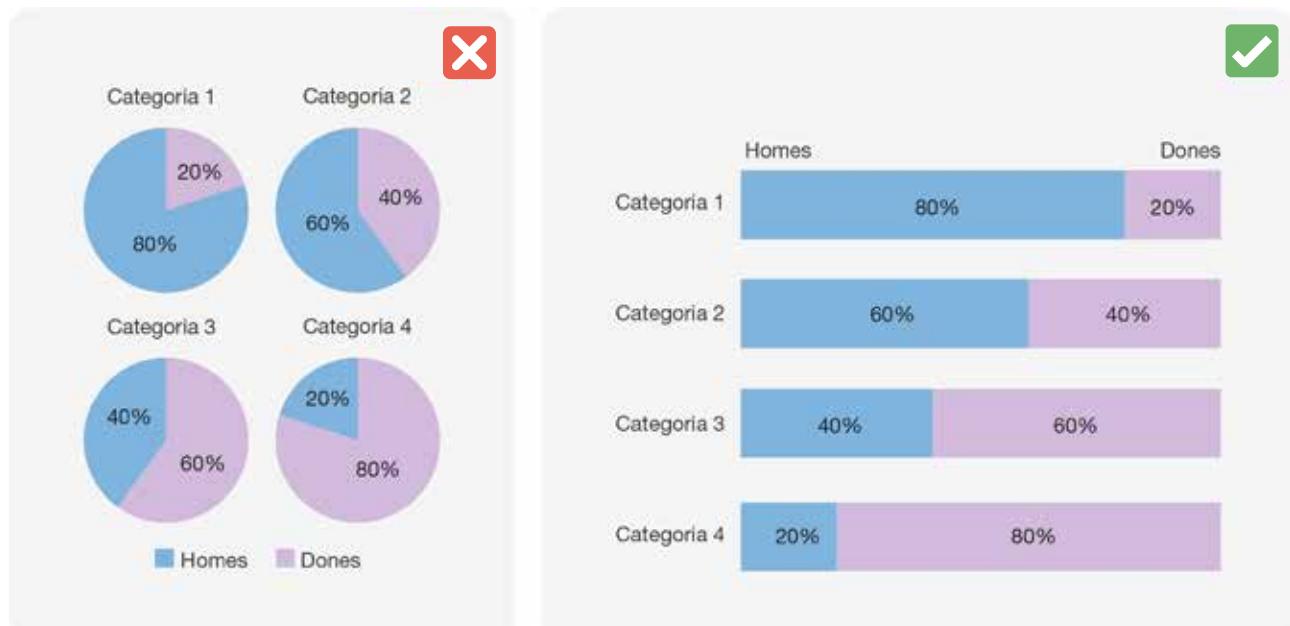


Figura 4.33. És preferible utilitzar barres per comparar valors a través de diverses categories. Adaptació de la imatge cedida per Ann K. Emery. Font: <http://annkemery.com/pie-chart-guidelines/>.

Una alternativa al gràfic de sectors és l'anomenat gràfic de corona. Aquest es construeix exactament igual que el de sectors, però afegint-hi un forat al mig on es pot mostrar un valor. Aquesta variació es fa servir sovint per mostrar el progrés d'un valor.

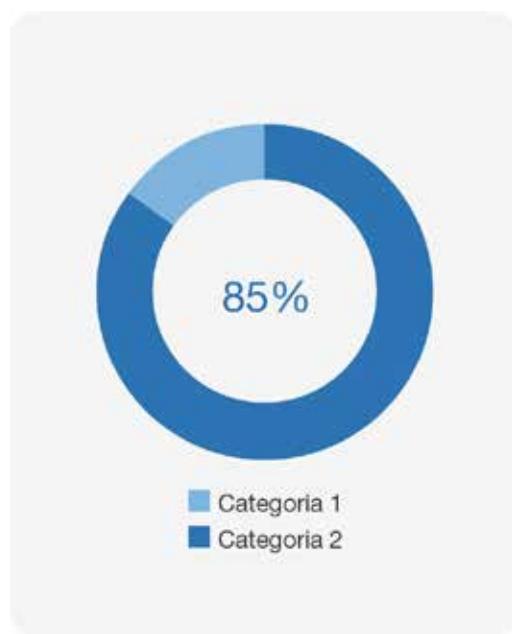


Figura 4.34. Exemple de gràfic de corona.

4.4.2. Pictograma

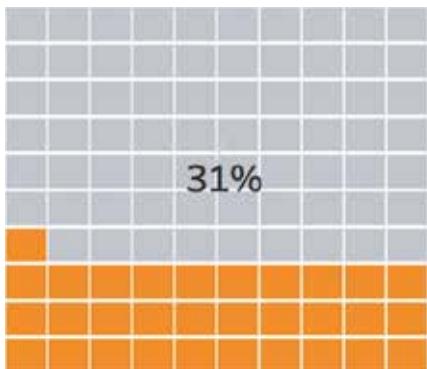


Figura 4.35. Exemple de pictograma.

Aquest gràfic s'utilitza per expressar un valor concret, generalment un percentatge. S'utilitza en forma de quadrat o de rectangle. El quadrat exterior representa el valor màxim, i el nombre de quadrats pintats d'un altre color representa el valor actual.

Quan s'ha d'utilitzar?

Quan es vol expressar un valor i alhora oferir un context que permeti identificar si és gran o petit. Per exemple, en el següent gràfic es pot veure el percentatge d'exportacions de cava per cada continent. És fàcil veure que les exportacions es concentren a Europa i que els altres continents poden representar una gran oportunitat comercial.

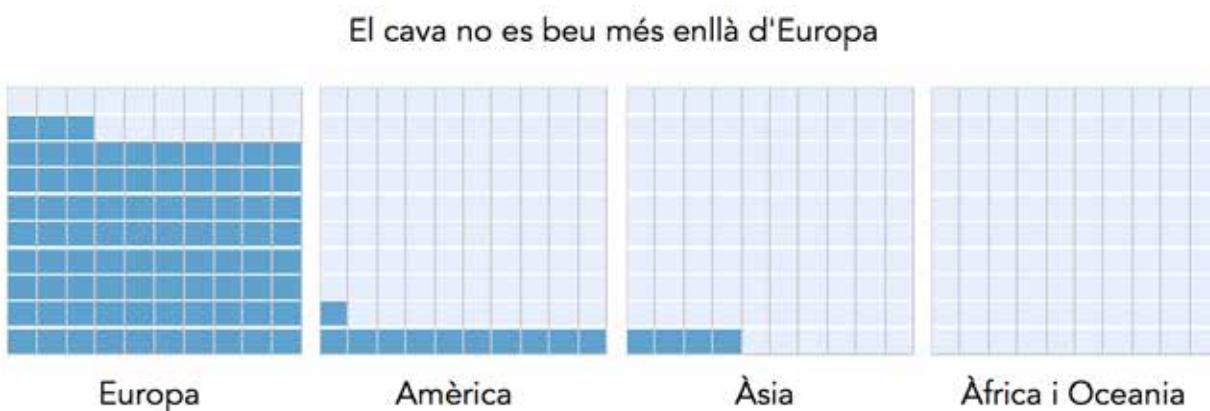


Figura 4.36. Percentatge de població que beu cava als diferents continents. Visualització feta per OneTandem.
Font: <https://public.tableau.com/profile/onetandem#/vizhome/cava/cavaexports>.

Els pictogrames són una bona alternativa gràfica per mostrar un únic nombre, especialment quan es tracta d'un indicador principal que va del 0% al 100%.

Recomanacions

Són òptims per representar indicadors amb un valor màxim de 100, per exemple, els percentatges que mai superaran el 100%.

Si és important detallar els decimals de l'indicador, no són una bona alternativa perquè cada quadrat de color representa una unitat.

4.4.3. Mapa d'arbre

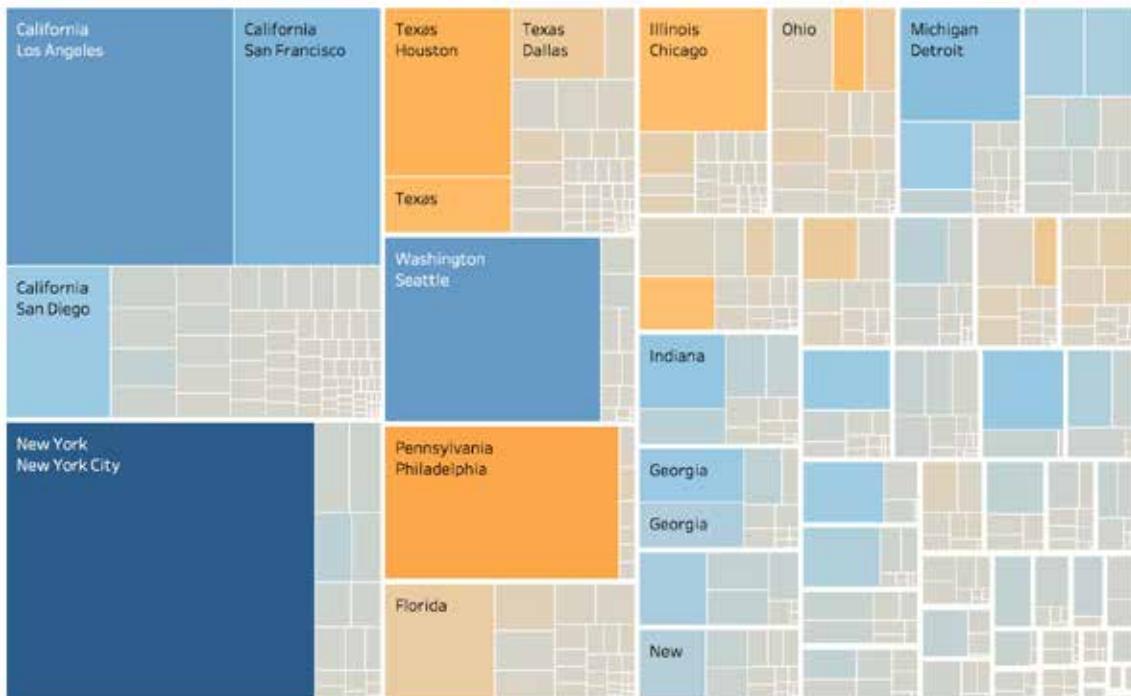


Figura 4.37. Mapa d'arbre que representa el volum de vendes (representat amb la mida) i el benefici (representat amb el color) d'una empresa fictícia a diferents ciutats americanes agrupades per estat.

El mapa d'arbre (*treemap*) permet veure una agrupació jeràrquica de valors. Es construeix dividint un rectangle en rectangles més petits, on la mida d'aquests i el seu color donen una informació determinada.

Quan s'ha d'utilitzar?

Quan les dades tenen una jerarquia i es vol comparar una o dues variables entre els diferents elements de les dades. A l'exemple, podem veure la jerarquia que s'estableix entre els estats i les ciutats a on una empresa ven productes. Amb aquest mapa d'arbre es compara el volum de vendes a cadascuna de les ciutats, representat mitjançant la mida dels rectangles. El color mostra el benefici que s'obté a cada ciutat (taronja per a beneficis negatius, i blau per a beneficis positius).

Recomanacions

La mida dels rectangles fa que no es vegin fàcilment les diferències entre valors molt semblants. Per tant, aquest gràfic no és adequat per identificar quins són els valors més grans (o sigui, per representar un rànquing).

El mapa d'arbre és especialment útil en versió interactiva. Això fa, per exemple, que es mostri el nom de cada rectangle quan s'hi passa per sobre amb el cursor. D'aquesta manera s'evita un dels problemes principals d'aquest gràfic: que si el quadrat és massa petit, no es veu el nom de la categoria que representa.

4.5. Distribucions

Abans d'abordar l'anàlisi i la comunicació d'un conjunt de dades, convé entendre la distribució de les diferents variables. La distribució ajudarà a comprendre la forma i les propietats principals d'una variable. Així mateix, es detecta si hi ha una distribució normal, de llarga cua o més aviat aleatòria dels valors. Això ens indicarà, per exemple, si és adequat fer servir mesures estadístiques com la mitjana o la mediana.

Histograma

Per entendre com estan distribuïts els valors d'una variable.

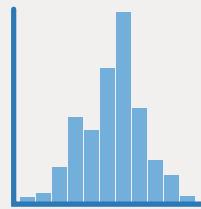
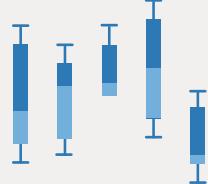


Diagrama de caixa

Per comparar la distribució de diferents variables o la distribució d'una d'elles en diferents categories.



4.5.1. Histograma

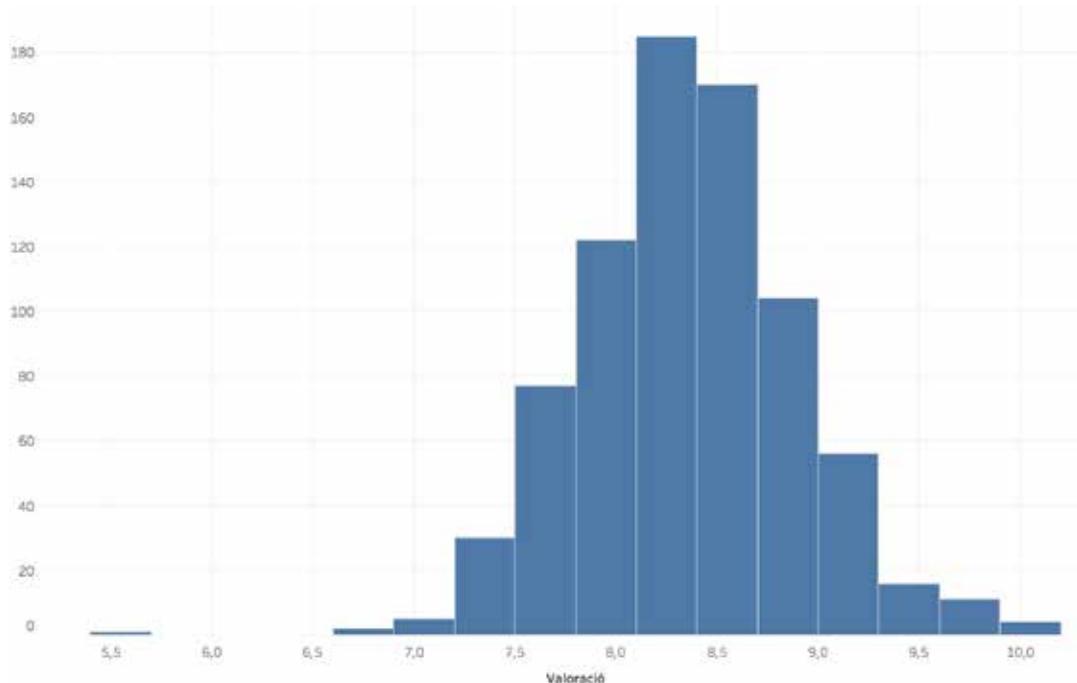


Figura 4.38. Histograma d'una distribució normal que mostra les valoracions d'un conjunt de sèries de televisió. Visualització feta per OneTandem. Font: <https://public.tableau.com/views/series2017/Dashboard1>.

Un histograma és un gràfic de barres que mostra la distribució de valors d'una variable. L'alçada de cada barra representa la freqüència d'aparició de valors dins del rang que defineix cada barra.

Quan s'ha d'utilitzar?

Quan es vol entendre com estan distribuïts els valors d'una variable. En particular, quan els valors es distribueixen de manera similar al voltant de la mitjana, com a l'exemple, es parla d'una distribució normal. Si, en canvi, la forma dibuixada per l'histograma no té forma de campana, i el valor màxim se situa en un dels extrems, es parla d'una distribució asimètrica.

Recomanacions

Per convenció, les barres dels histogrames no s'han de separar entre si.

La forma d'un histograma dependrà del rang de valors que representa cada barra. Aquest rang és important ja que és el que ens permetrà agrupar els valors i, per tant, acabarà definint la forma de l'histograma. Generalment, els programes decideixen automàticament la mida d'aquest rang en funció del conjunt de barres, tot i que es pot editar. A continuació es mostren dues distribucions de les mateixes dades, utilitzant dos rangs diferents per agrupar els valors del conjunt de dades. El de l'esquerra fa servir un rang de 0,5, mentre que el de la dreta en fa servir un de 0,25.

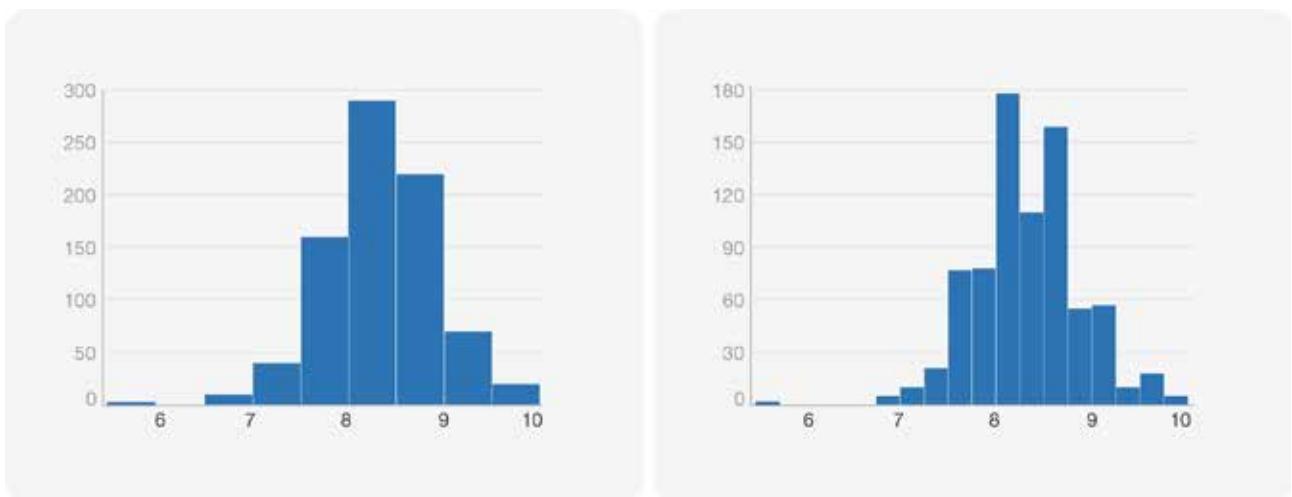


Figura 4.39. Dos histogràmes sobre les mateixes dades que utilitzen rangs diferents per agrupar les dades.

No hi ha una fórmula exacta i ideal per calcular aquest rang, de manera que serà feina de l'analista de dades provar-ne de diferents per fer-se una idea de com es distribueixen les dades.

4.5.2. Diagrama de caixa

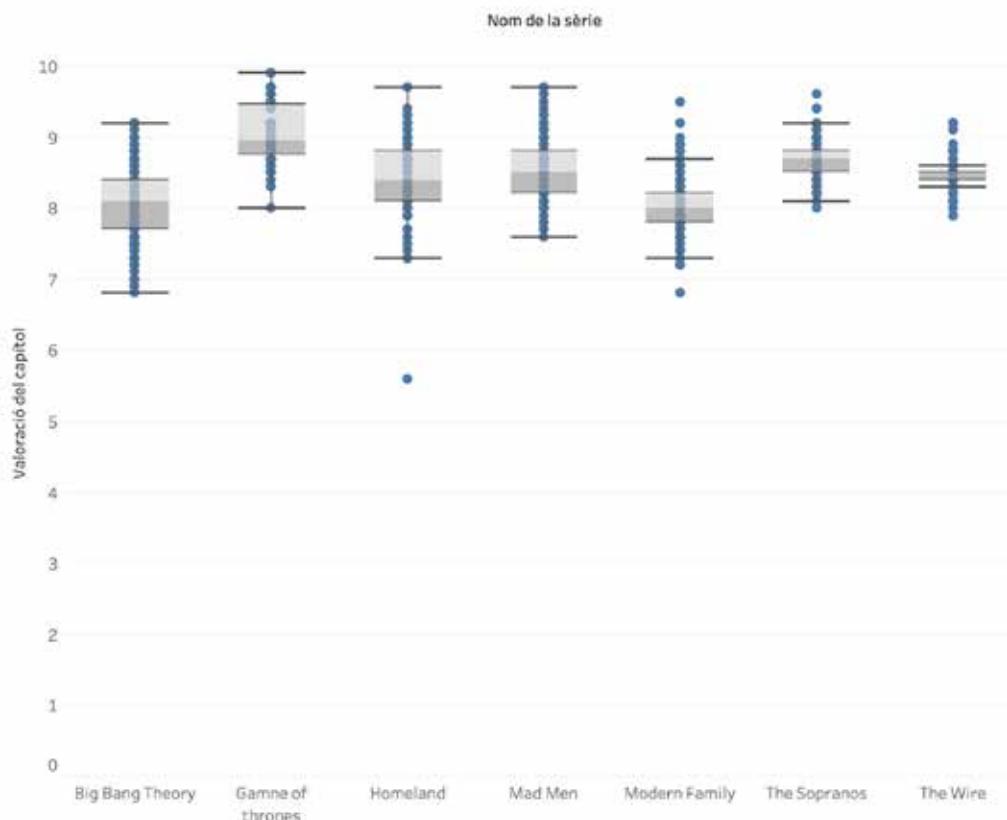


Figura 4.40. Diagrama de caixa de les valoracions dels capítols de sèries de televisió (dades extretes de www.imdb.com). Cada punt és un capítol diferent. Gràfic elaborat per OneTandem.

Un diagrama de caixa (*box plot*) permet representar la distribució dels valors d'una variable i, opcionalment, comparar-los per diferents categories.

En el següent diagrama s'explica el significat dels elements d'un diagrama de caixa. La caixa està dividida per una línia que representa la mediana. Els extrems de la caixa corresponen al primer i al tercer quartil. De la caixa, emergeixen unes línies anomenades “bigotis”, que poden calcular-se de diferents maneres. Tots els punts que es trobin més enllà dels bigotis es consideren valors atípics.

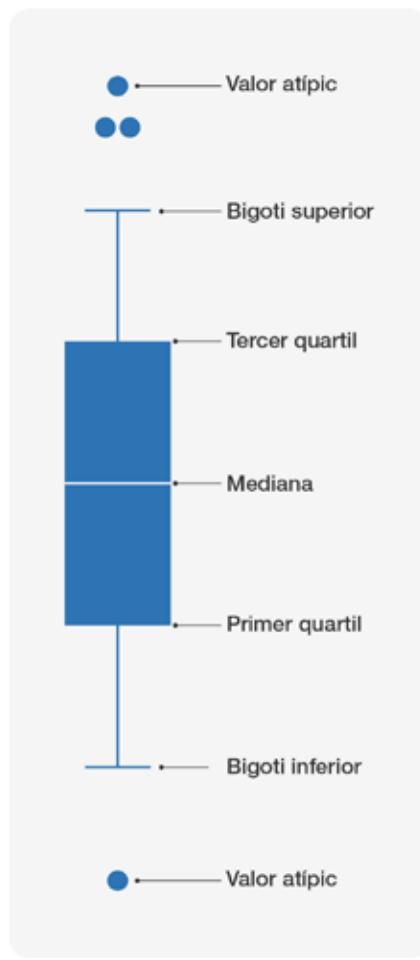


Figura 4.41. Esquema d'un diagrama de caixa.

Quan s'ha d'utilitzar?

Quan volem comparar la distribució de diferents variables o la distribució d'una variable en diferents categories.

El diagrama de caixa ajuda a entendre la forma de les dades, un pas previ en tota anàlisi de dades. Si la línia de la mediana està situada al centre de la caixa (com a la sèrie *MadMen* a la imatge), voldrà dir que hi ha una distribució simètrica (sovint, una distribució “normal”). En cas contrari, voldrà dir que hi ha una distribució asimètrica (veure [Histogrames](#)).

D'altra banda, si la caixa és molt petita, vol dir que els valors de la variable estan poc dispersos (com en el cas de *The Wire* a l'exemple). Dit d'una altra manera, és fàcil predir un valor aproximat de la variable, ja que sempre presenta valors molt similars.

El diagrama de caixa també és idoni per detectar valors atípics, que són aquells punts que es troben més enllà de les línies horizontals situades als extrems de cada caixa o bigotis (se'n pot apreciar un de molt important a la sèrie *Homeland*). Aquests valors poden ser fruit d'errors de mesura i, per tant, és important descartar-los per evitar arribar a conclusions errònies (per això l'ús del diagrama de caixa és comú en àmbits de recerca científica).

Sovint, però, els valors atípics indiquen que alguna cosa està passant. Caldrà analitzar les dades més a fons, o trobar-ne de noves, perquè segurament estem a les portes d'una troballa interessant.

Recomanacions

Per fer un bon ús d'aquest gràfic cal conèixer conceptes fonamentals de l'estadística descriptiva, en particular, de la mediana i els quartils. Per tant, no és un gràfic adient per a tots els públics.

De la mateixa manera que els gràfics de barres, els diagrames de caixa poden ser horizontals o verticals. Generalment es fan verticals, però si els noms de les categories són molt llargs, caldrà fer-los horizontals perquè hi càpiguen còmodament.

S'anomena bigotis a les línies horizontals que se situen als extrems del diagrama de caixa. Aquestes línies ajuden a identificar valors atípics. Hi ha diferents convencions a l'hora de situar aquestes línies:

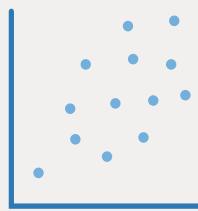
- Situar-les a 1,5 vegades la distància interquartil, a partir del primer i tercer quartil respectivament
- Situar-les als percentils 9 i 91 de les dades
- Situar-les als percentils 2 i 98 de les dades

4.6. Correlacions

Els gràfics de correlacions ajuden a entendre la relació entre diferents variables. Per exemple, permeten veure si el preu d'un producte està relacionat amb la seva demanda. Tot i això, convé tenir present que la correlació de dues variables no implica necessàriament que l'una sigui causa de l'altra.

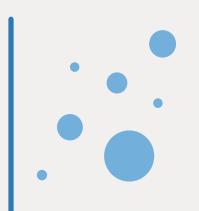
Gràfic de dispersió

Per veure la relació que existeix entre dues variables.



Gràfic de bombolles

Quan es vol gaudir del que ofereix un gràfic de dispersió i, alhora, es desitja tenir més informació sobre cada un dels elements representats.



Coordenades paral·leles

Per explorar conjunts de dades multidimensionals.



4.6.1. Gràfic de dispersió

Relació entre benefici i vendes



Figura 4.42. Gràfic de dispersió que mostra les vendes i el benefici per producte. Cada punt és un producte.

És un tipus de gràfic que permet representar els valors de dues variables sobre dos eixos de coordenades x i y . Cada element es representa com un punt situat a l'espai en funció dels seus valors en cada un dels eixos.

Quan s'ha d'utilitzar?

Quan es vol veure la relació que hi ha entre dues variables.

El gràfic de dispersió també és útil per detectar elements atípics, que no segueixen la mateixa relació de correlació que la resta de punts.

Recomanacions

Per analitzar un gràfic de dispersió, cal tenir en compte el següent:

- **La forma:** si els punts se situen al voltant d'una recta, vol dir que les variables estan correlacionades linealment. Si els punts es distribueixen al voltant d'una recta de pendent positiu, vol dir que estan positivament correlacionats (és a dir, quan augmenta el valor de x , també ho fa el de y , com a l'exemple principal). En cas contrari, estan negativament correlacionats. D'altra banda, si no es distribueixen al voltant de cap línia, no estan correlacionats. A part de la correlació lineal, expressada com una recta, també podria existir correlació exponencial, logarítmica o polinomial.

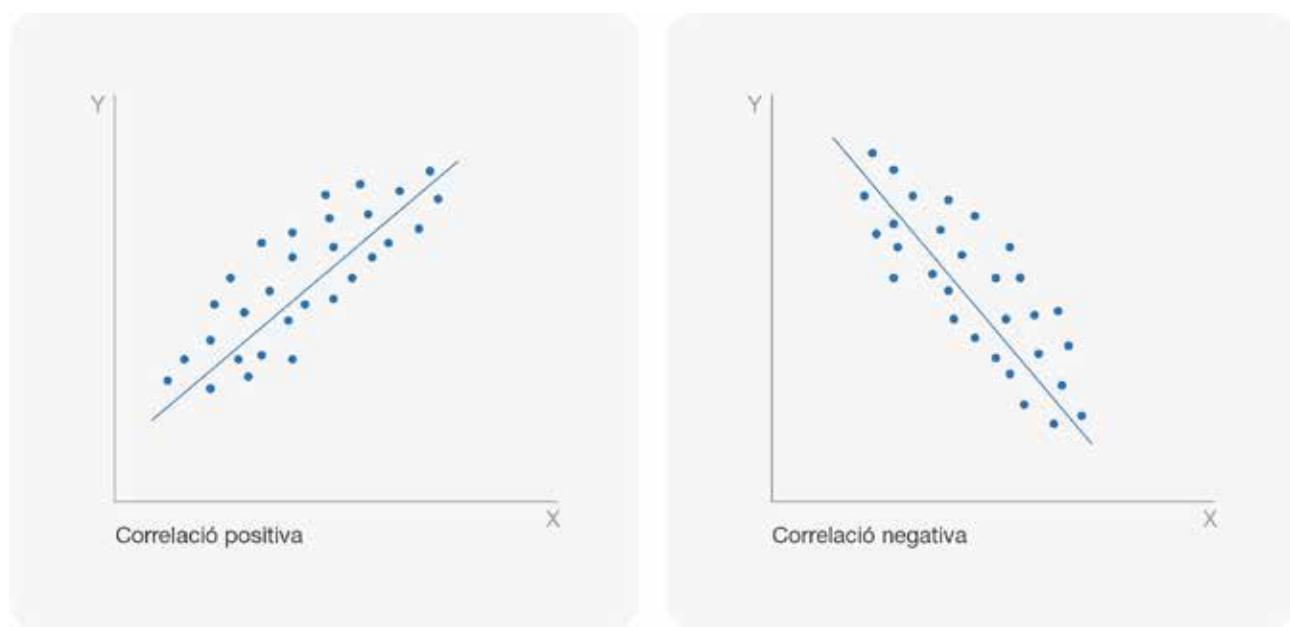


Figura 4.43. Exemples de correlació positiva i negativa.

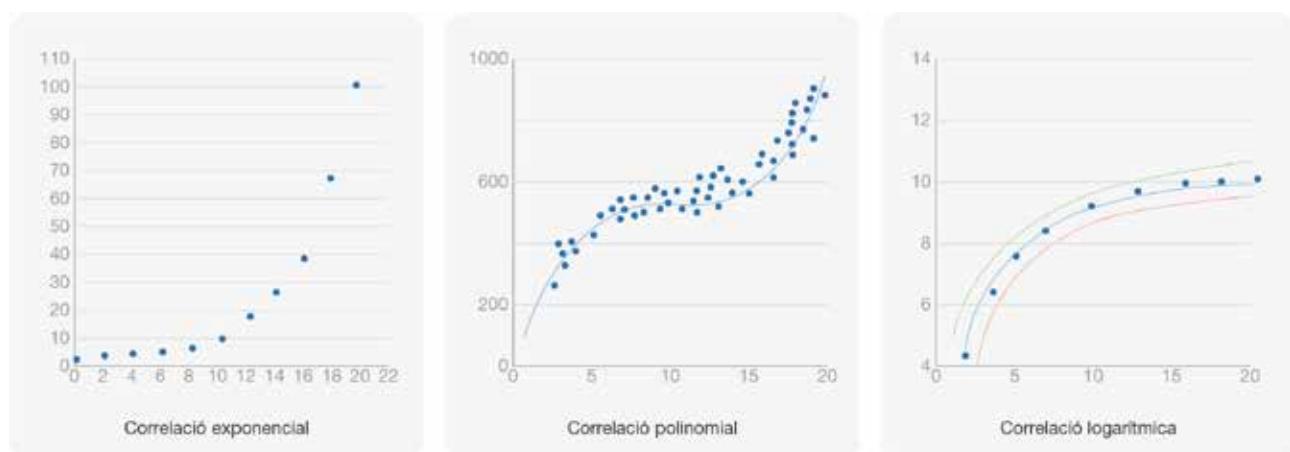


Figura 4.44. D'esquerra a dreta, correlació exponencial, polinomial i logarítmica.

- **Les agrupacions:** si molts punts s'agrupen al voltant de valors similars, vol dir que tenen característiques similars (en anglès s'anomena cluster una agrupació d'elements).

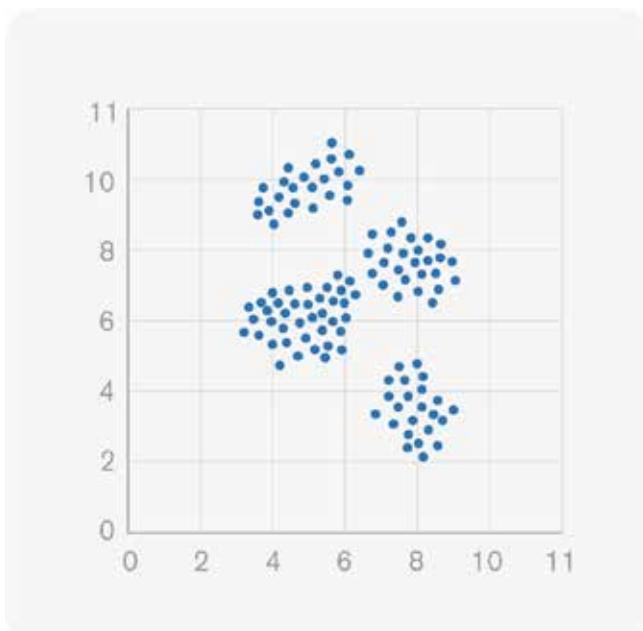


Figura 4.45. Exemple de gràfic de dispersió que permet veure agrupacions.

- **Els valors atípics** són aquells valors que queden fora de la forma principal que s'aprecia en un gràfic de dispersions i corresponen a excepcions que caldrà saber explicar o bé eliminar del gràfic per no distorsionar la seva forma. En la següent figura corresponen als punts blaus.

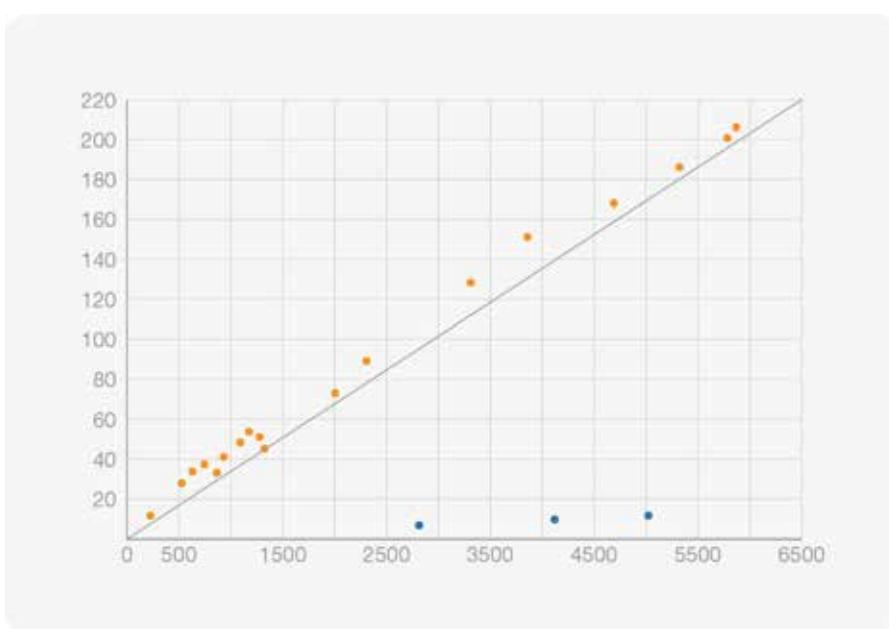


Figura 4.46. Exemple de gràfic de dispersió que permet veure, en blau, valors atípics.

- **Els quadrants:** si dividim el gràfic en quatre quadrants, es poden veure quatre agrupacions d'elements molt diferenciades, en particular els valors alts per l'eix *x* i l'eix *y*, els que tenen valors baixos pels dos eixos, i els que tenen valors baixos per un eix i alts per l'altre, tal com es pot veure a la següent imatge.

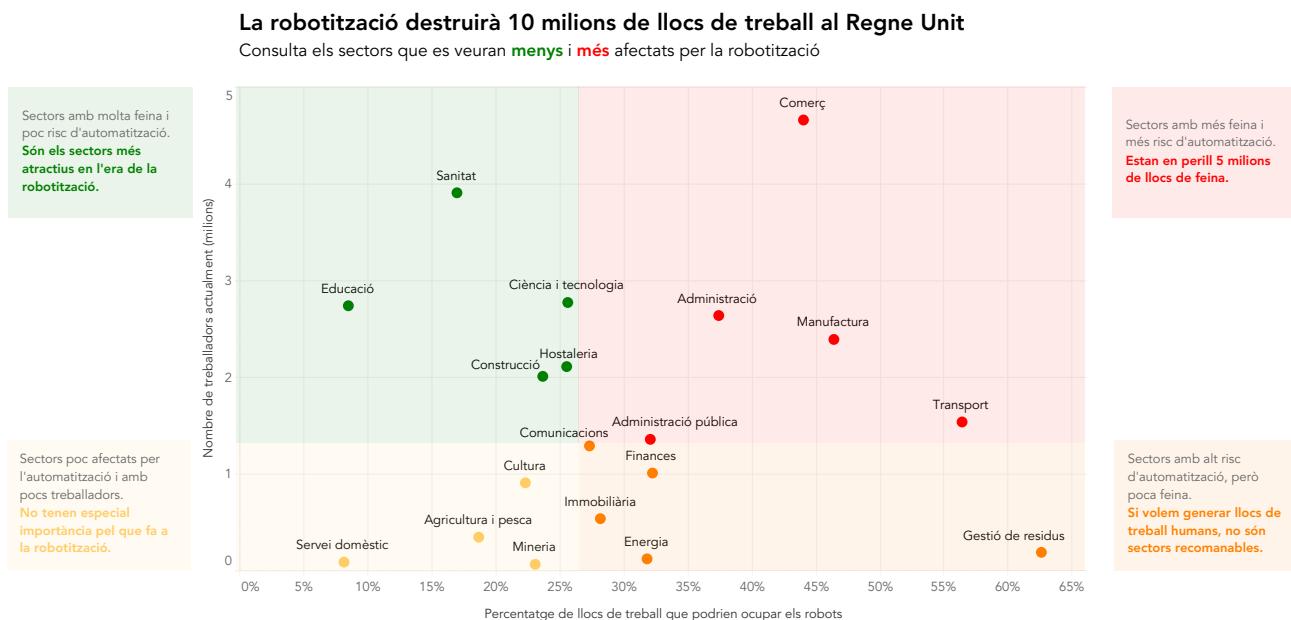


Figura 4.47. La robotització i el mercat laboral. Imatge cedida per OneTandem.

Font: <http://onetandem.com/blog/la-robotizacion-mercado-laboral/>.

Finalment, cal tenir en compte que afegir línies de referència en forma de retícula als gràfics de dispersió és molt útil perquè ajuden a estimar els valors de les variables de cada punt.

4.6.2. Gràfic de bombolles

Relació entre benefici, vendes i descompte

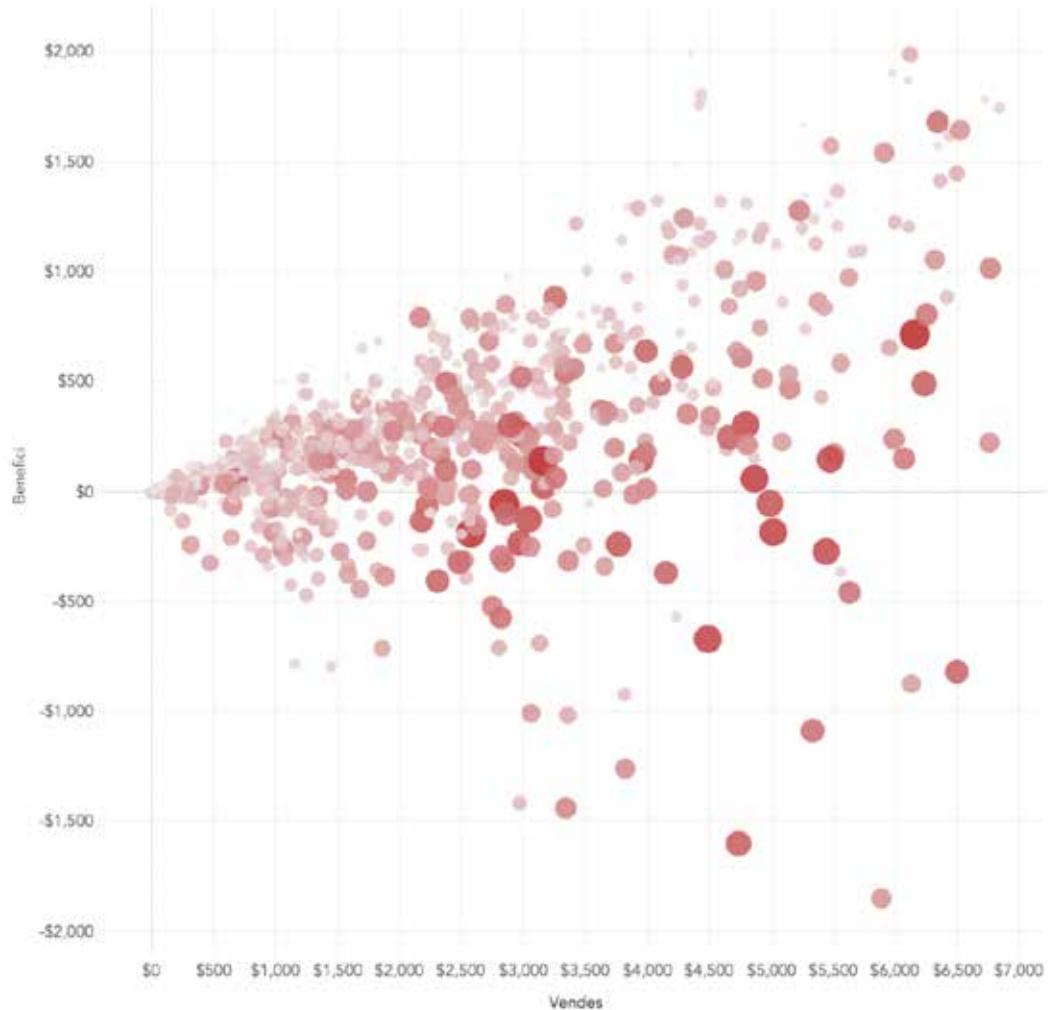


Figura 4.48. Gràfic de bombolles que mostra la relació entre benefici, vendes i descompte per a un conjunt de productes. La mida i color de la bombolla representa la magnitud del descompte. Es pot veure com els productes amb benefici negatiu no sempre són els que tenen més descompte.

El gràfic de bombolles és un gràfic de dispersió en què l'àrea i el color dels punts poden representar variables addicionals. D'aquesta manera, es poden representar fins a 5 variables diferents alhora:

- la posició segons l'eix x
- la posició segons l'eix y
- el color del punt
- la mida del punt
- l'animació per representar l'evolució temporal de cada punt.

Quan s'ha d'utilitzar?

Quan es vol aprofitar el que ofereix un gràfic de dispersió i, alhora, es vol tenir més informació sobre cada un dels elements representats utilitzant la mida, el color i/o l'animació.

Recomanacions

A banda de les recomanacions per al gràfic de dispersió, en el cas del gràfic de bombolles és important decidir quina variable s'utilitza per a cadascun dels atributs visuals utilitzats (la posició marcada pels eixos x i y , la mida i el color).

Les variables principals de l'anàlisi se situen al llarg de l'eix x i de l'eix y per veure més fàcilment si estan relacionades entre si o per veure si hi ha elements que s'agrupen al voltant de certs valors.

La mida s'acostuma a utilitzar per representar una variable de la qual es vol veure quin és l'element més gran.

S'utilitza el color per distingir les categories a les quals pertany cada element o per comparar valors grans i petits d'una altra variable (en aquest cas caldrà utilitzar una escala de colors numèrica).

I, finalment, l'animació es fa servir per representar l'evolució temporal dels valors dels elements, de manera que es pugui veure com els punts es mouen amunt o avall, a la dreta o a l'esquerra, a mesura que passa el temps.

4.6.3. Coordenades paral·leles

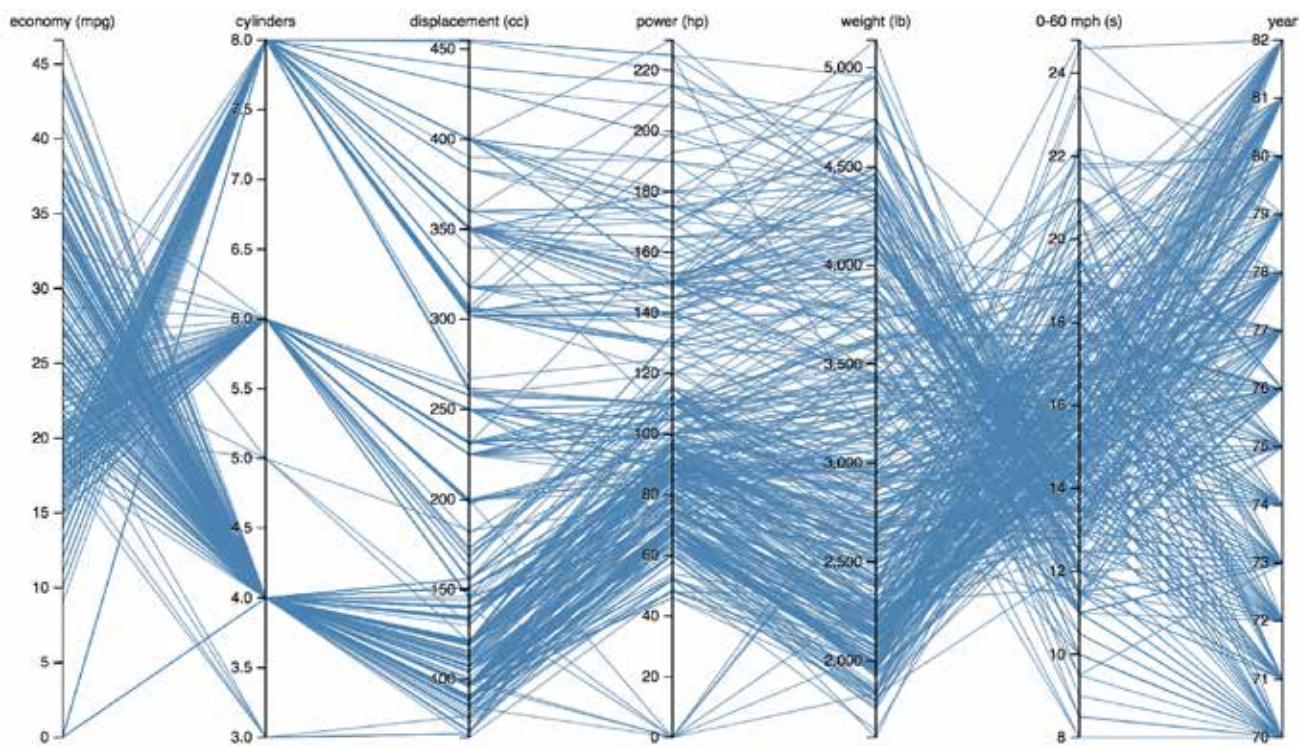


Figura 4.49. Comparació de models de cotxe a partir de diferents característiques.

Font: <https://bl.ocks.org/mbostock/7586334>.

Les coordenades paral·leles permeten explorar les relacions entre diferents variables per a un conjunt d'elements. Cada variable es representa amb un eix vertical, i els elements del conjunt de dades es representen unint els punts que indiquen el valor de cada variable en cada eix.

Quan s'ha d'utilitzar?

Quan es volen explorar conjunts de dades multidimensionals. En particular, amb aquest gràfic es poden descobrir tendències o valors atípics a través de l'ús de filtres interactius.

Recomanacions

Les coordenades paral·leles són especialment útils en la seva versió interactiva. La interacció que més sovint s'utilitza és la de filtrat, que fa que l'usuari seleccioni un rang de valors en un eix i veure així com es comporten en la resta de variables. En la següent imatge, per exemple, s'han filtrat els elements per un rang determinat de la variable *weight (lb)*. Entre d'altres coses, es pot veure que aquests elements acostumen a tenir valors baixos de la variable *economy (mpg)* i, per tant, és possible que hi hagi una relació inversa entre aquestes dues variables.

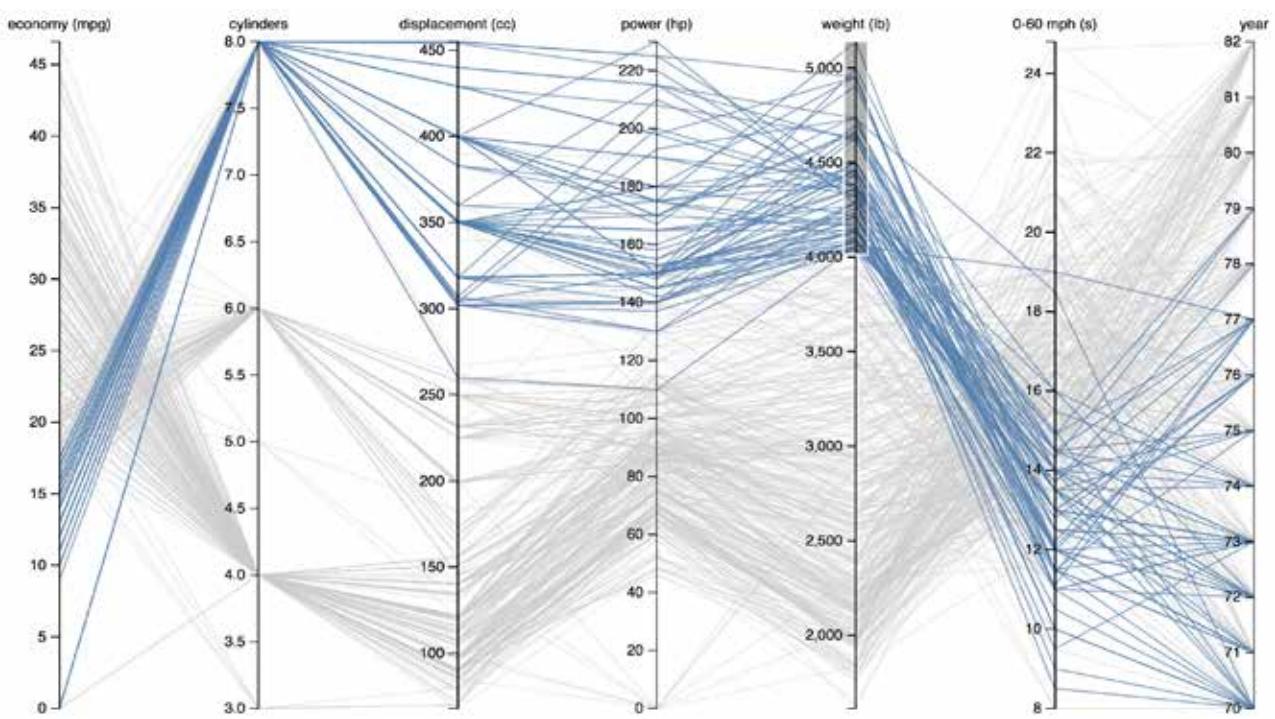


Figura 4.50. Comparació de models de cotxe a partir de diferents característiques filtrat per la variable pes (*weight*).
Font: <https://bl.ocks.org/mbostock/7586334>.

4.7. Connexions, relacions i xarxes

Els gràfics de connexions, relacions i xarxes ajuden a entendre la relació entre els elements d'un conjunt de dades. Per exemple, les relacions entre membres d'una xarxa social.

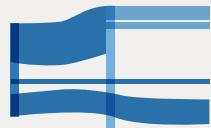
Diagrama node-aresta

Per detectar patrons que sorgeixen de les relacions entre els elements del conjunt de dades.



Diagrama de Sankey

Per representar el canvi d'una variable a través del flux entre diferents estats.



4.7.1. Diagrama de nodes i arestes

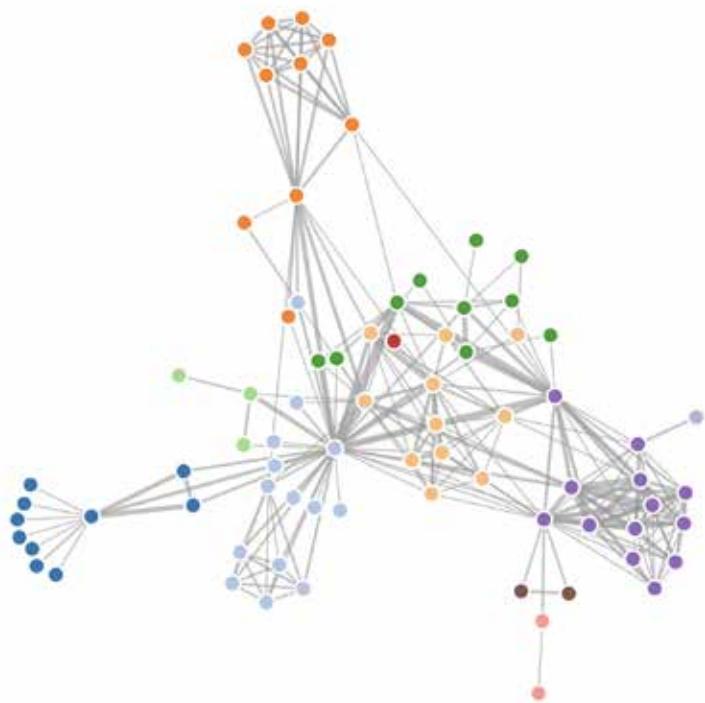


Figura 4.51. Xarxa de personatges a l'obra *Els miserables*. Cada unió indica que apareixen a la mateixa escena.
Font: <https://bl.ocks.org/mbostock/4062045>.

Els diagrames de nodes i arestes utilitzen cercles per representar elements, que s'anomenen nodes, i arestes que es representen com a línies entre nodes i que indiquen una relació entre ells.

Les connexions es poden establir partint de diferents factors, com per exemple, relacions d'amistat entre membres d'una xarxa social, o personatges que apareixen a les mateixes escenes en una pel·lícula. Poden tenir una direcció que indica la naturalesa de la relació i que es representa amb arestes que acaben en forma de fletxa. Per exemple, en una xarxa com Facebook la relació entre persones és bidireccional, mentre que a Twitter, no (l'usuari A pot seguir l'usuari B, però no cal que l'usuari B segueixi l'usuari A).

L'entitat matemàtica que sorgeix d'establir la connexió entre nodes a través d'arestes s'anomena "graf".

Quan s'ha d'utilitzar?

Quan es volen detectar patrons que sorgeixen de les relacions entre els elements del conjunt de dades. En general, ens ajuda a trobar elements molt connectats o elements que queden més aïllats dins la xarxa. A l'exemple es poden veure fins a vuit grups diferents de personatges a *Els miserables*. Alguns queden barrejats al centre de la xarxa mentre que d'altres mostren comunitats molt clares i revelen que aquests personatges no interactuen amb altres personatges de l'obra.

Recomanacions

La complexitat principal recau en la manera de situar els nodes a l'espai. Generalment cal seguir els següents principis:

- Minimitzar els encavalcaments entre nodes i els creuaments entre arestes.
- Incloure tota la xarxa a la mateixa pantalla.
- Afavorir el descobriment de grups d'elements similars (clústers).

El compliment d'aquests principis sol aconseguir-se aplicant els anomenats *algorismes dirigits per forces* que, segons l'estructura de la xarxa, decideixen la millor posició per a cada node.

A més a més, sovint s'utilitzen variables del conjunt de dades per modificar la mida o el color dels nodes i/o les arestes.

4.7.2. Diagrama de Sankey

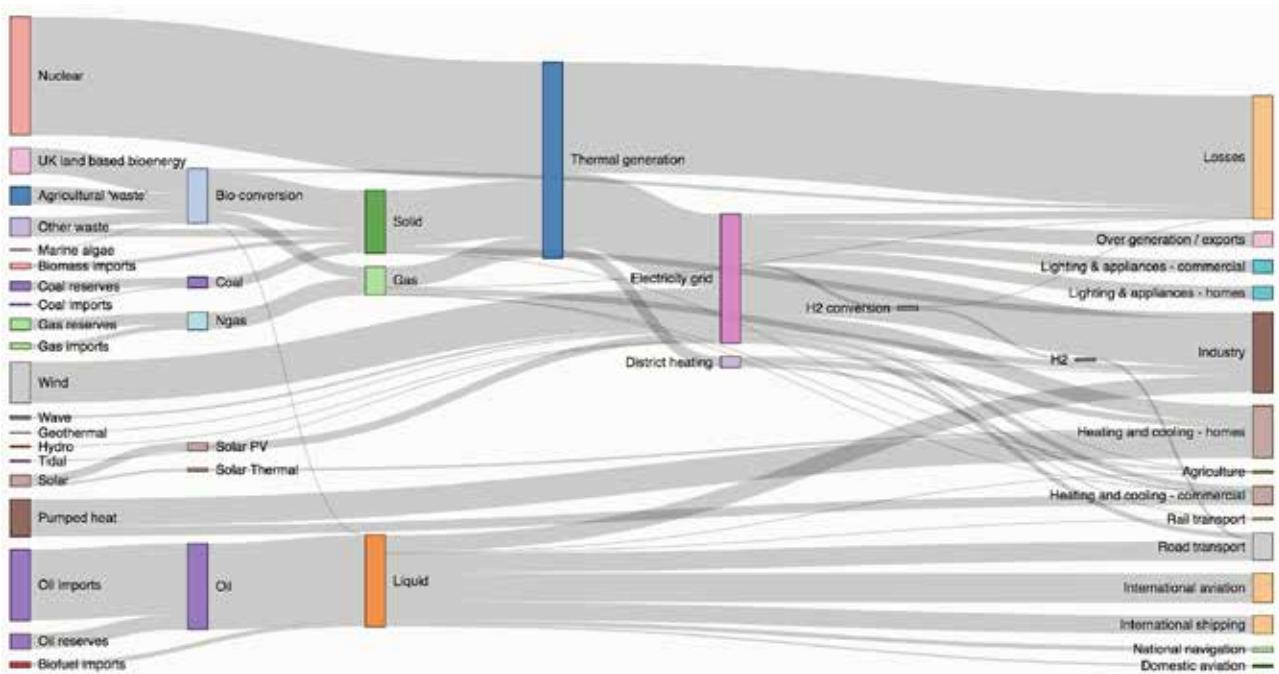


Figura 4.52. Diagrama de Sankey que mostra la cadena de producció i consum d'energia elèctrica.
Font: <https://bost.ocks.org/mike/sankey/>.

El diagrama de Sankey mostra diferents categories o estats a través dels quals una variable va canviant de valor. També mostra que una categoria és la suma dels valors de diferents categories, tal com s'aprecia a l'exemple.

Quan s'ha d'utilitzar?

Quan es vol representar el canvi d'una variable a través del flux entre diferents estats. El diagrama de Sankey mostra com una variable es va descomponer en els diferents passos i/o categories. Per exemple, quan es representa la cadena de producció i consum d'energia elèctrica, com es pot veure a la imatge superior.

Recomanacions

El més important a l'hora de crear un diagrama de Sankey és decidir en quin ordre vertical se situen les diferents categories que es representen en cada un dels estats. De l'elecció de la posició resultaran diferents graus d'encavalcament entre arestes que dificultaran més o menys la lectura del gràfic.

5. Principis del disseny

A l' hora de desenvolupar una visualització de dades, a banda de fer una anàlisi de les dades, cal pensar la millor manera de representar-les. És per això que convé tenir uns fonaments dels principis del disseny, que ens permetran entendre per què prenem certes decisions de representació visual i, per tant, podrem actuar en conseqüència. En aquest capítol tractarem els principals aspectes del disseny que cal tenir presents en la visualització de dades.

Prèviament, s'ha de tenir en compte que les visualitzacions de dades de la Generalitat de Catalunya s'elaboren d'acord amb la normativa d'identitat corporativa. Per tant, els departaments, organismes autònoms adscrits i les empreses públiques que en depenen han d'aplicar les pautes següents, definides al web Identitat corporativa:

- Gràfics i taules <http://identitatcorporativa.gencat.cat/ca/aplicaciones/grafics-i-taules/>
- Infografies <http://identitatcorporativa.gencat.cat/ca/aplicaciones/infografies/>

5.1. Pensament visual

La vista és el nostre sentit més important. Un 70% dels òrgans receptors del cos humà estan concentrats als ulls. A més a més, els sensors visuals estan connectats directament amb el sistema cognitiu. Per tant, si entenem el funcionament de la percepció visual, podrem dissenyar visualitzacions que aprofitin al màxim les nostres capacitats cognitives. Veiem-ho amb un exemple.

La següent taula de valors representa les vendes trimestrals d'una empresa a dos segments de clients (petita empresa i gran empresa).

	2014 T1	2014 T2	2014 T3	2014 T4	2015 T1	2015 T2	2015 T3	2015 T4	2016 T1	2016 T2
Gran empresa	\$13.992	\$27.245	\$32.993	\$54.206	\$22.280	\$18.980	\$35.252	\$52.245	\$26.578	\$42.418
Petita empresa	\$21.965	\$47.852	\$96.104	\$100.177	\$40.963	\$56.330	\$79.399	\$89.844	\$46.672	\$70.139

Figura 5.1. La taula mostra les vendes trimestrals d'una empresa a clients de petita empresa i de gran empresa.

La taula ens permet localitzar de manera senzilla un valor concret. Però si volem extreure més informació, podem optar per la següent visualització:

Vendes per trimestre per clients de petita i gran empresa

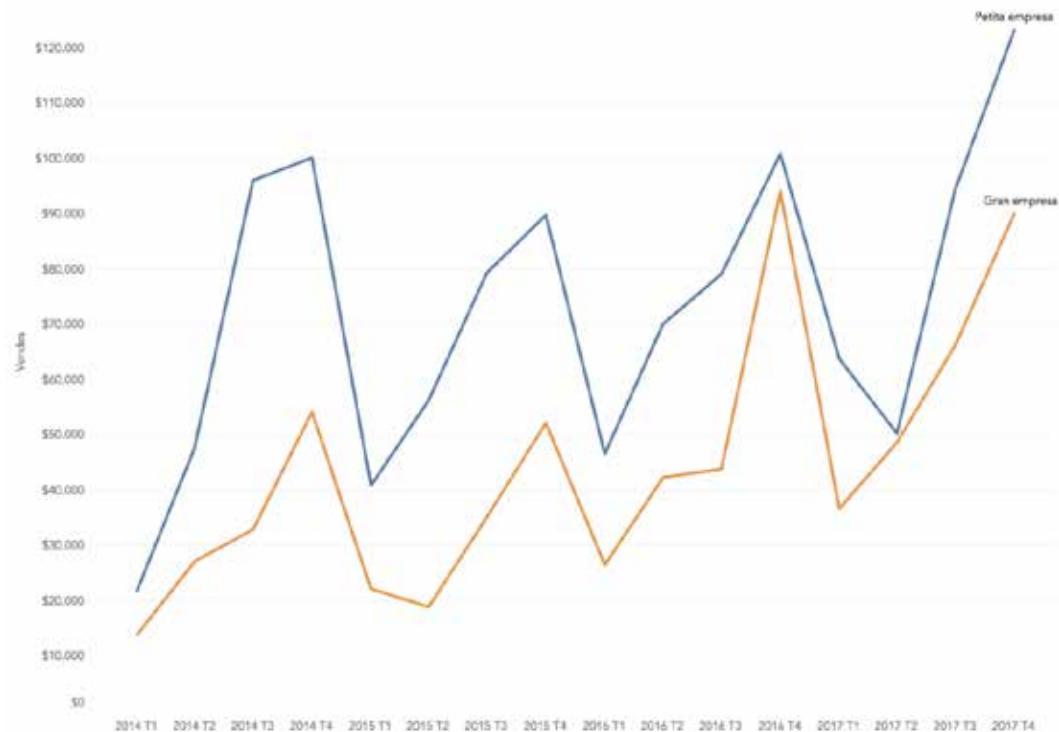


Figura 5.2. El gràfic de línies mostra les vendes trimestrals d'una empresa a clients de petita empresa i de gran empresa.

Gràcies a la representació amb un gràfic de línies podem veure la tendència cíclica i globalment ascendent de les vendes. Per què un gràfic de línies ho permet i una taula no? Per entendre-ho, cal tenir presents els tres tipus de memòria que actuen al cervell:

- memòria icònica
- memòria a llarg plaç
- memòria de treball

La memòria icònica s'ocupa de la informació que rep el cervell procedent dels receptors visuals. Es processa ràpidament i sense que en siguem conscients: en diem processament "preatentiu". No pretén l'anàlisi exhaustiva de tot el que veiem, sinó que extreu un subconjunt d'elements rellevants com el color o la forma.

La memòria a llarg plaç és aquella d'accés més lent, que ens permet emmagatzemar una informació que hem tractat molt o que hem estudiat detingudament. Per exemple, la clau d'accés al nostre correu electrònic, o el material que hem estudiat per un examen.

La memòria de treball processa informació proporcionada per la memòria icònica i la de llarg plaç. Té una capacitat limitada: diferents estudis⁸ diuen que pot arribar a emmagatzemar entre 5 i 9 elements (en funció de la persona i el tipus de tasca que estigui realitzant). Aquesta limitació fa que entenguem millor les dades de l'exemple quan aquestes estan codificades en un gràfic de línies (un element) enllot d'estar codificades en una taula (tants elements diferents com valors).

Així doncs, a l'hora de fer bones visualitzacions de dades, caldrà:

- triar els atributs visuals adequats per aprofitar la capacitat de processament preatentiu de la memòria icònica (en particular, color i forma)
- respectar les limitacions de la memòria de treball
- ser consistent amb l'ús de la maquetació i format dels elements per afavorir els mecanismes de la memòria a llarg plaç.

5.2. Color

El color és un atribut visual preatentiu (processat per la memòria icònica), i el més utilitzat juntament amb la forma. Per utilitzar-lo per comunicar informació cal distingir entre dades quantitatives o bé dades categòriques.

En el cas de disposar de dades quantitatives, tenim dues opcions:

- utilitzar una escala on s'utilitzi un únic color amb diferents saturacions, tal i com es pot veure a la figura 5.3
- en cas que disposem d'un valor central, com per exemple el zero, utilitzar una escala dicotòmica, com podem veure a la figura 5.4

100	100	100	100	100	75.3
41.8	82.8	79.8	71.1	70.9	51.9
68.4	80.7	60.6	73.6	61.5	67.1
66.8	79.4	65.3	67.5	68.1	55.9
79.1	96.6	50.8	56.2	66.5	55.2
52.1	88.5	57.1	46.2	44.2	68.1
48.5	66.7	49.3	60	44.9	100

Figura 5.3. Escala d'un únic color, que fa servir la saturació per distingir elements.

⁸ Miller, George Armitage. *The magical number seven, plus or minus two: Some limits on our capacity for processing information.* Psychological Review, 1956



Figura 5.4. Escala dicotòmica de tres colors, amb diferents saturacions.

Tot i que l'ús del color és una manera de codificar les dades molt intuïtiva, no és una bona elecció quan es necessita saber quant més gran és un valor respecte un altre. Per exemple, podem percebre que un color és més fosc que un altre, però ens resulta pràcticament impossible percebre que un color és dues vegades més fosc que un altre.

En el cas de disposar de dades categòriques, el color ens servirà per distingir les categories entre sí. Per tant, caldrà cercar colors que siguin ben diferents.

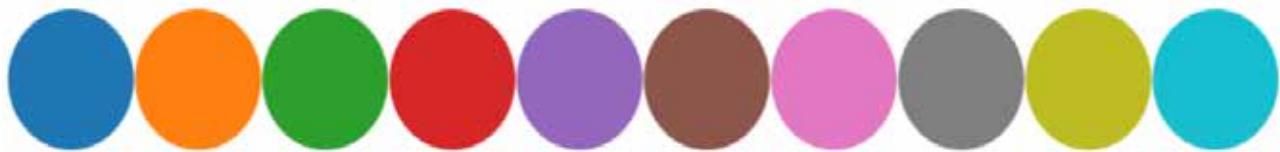


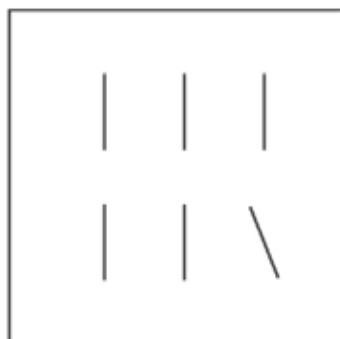
Figura 5.5. Escala de colors categòrica proporcionada per la llibreria D3js. Font: <https://d3js.org/>

Encara que existeixi un gran nombre de colors, és difícil trobar més de deu colors que siguin marcadament diferents entre sí. Per tant, el color no és un bon atribut visual en cas que tinguem massa categories per representar. Així mateix, recordem que la memòria de treball té dificultats treballant amb més de 5 a 9 elements diferents (en particular, més de 5 a 9 colors). En aquests casos, caldrà crear una agrupació alternativa de dades, o bé fer servir mecanismes de filtratge que ens permetin anar de la globalitat a la particularitat de manera interactiva.

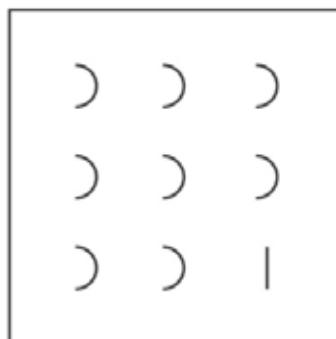
Finalment, cal tenir en compte que aproximadament d'un 5% a un 10% de la població pateix daltonisme i, per tant, té dificultats per distingir el vermell i el verd. Existeixen escales de colors especialment pensades per a aquestes persones, que recomanem fer servir.

5.3. Forma

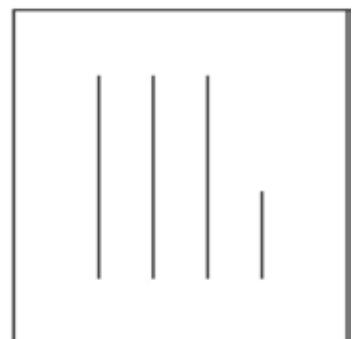
La forma és l'altre atribut visual preatentiu més utilitzat, a banda del color. La següent imatge mostra diferents maneres de fer servir la forma:



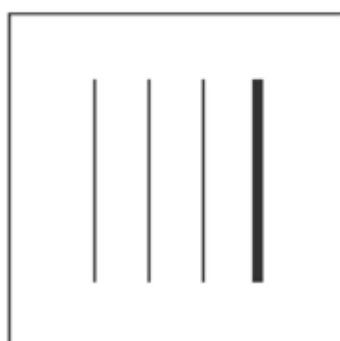
Orientació i coliniaritat



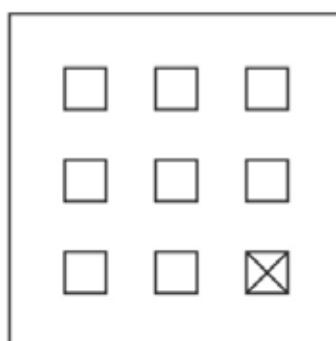
Curvatura



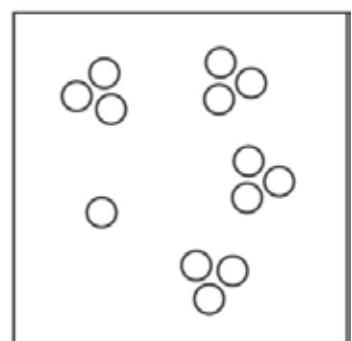
Llargada



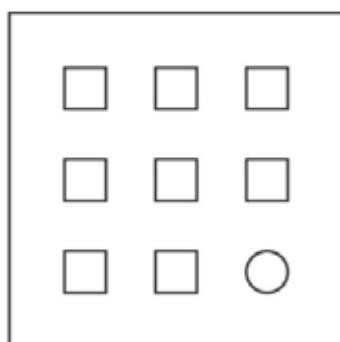
Amplada



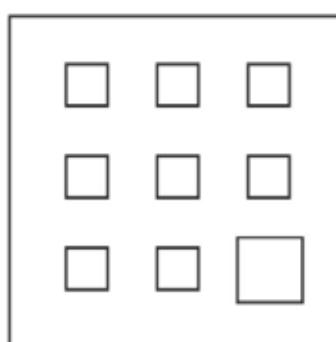
Amb marques



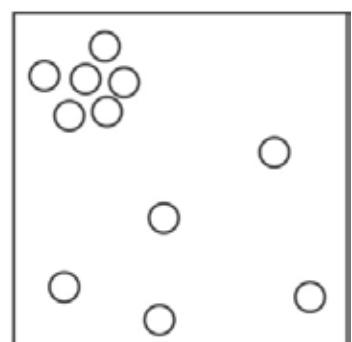
Numeració



Formes diferents



Mida



Agrupació espacial

Figura 5.6. "Diferents maneres d'utilitzar l'atribut preatentiu de la forma".

La longitud és l'atribut que el nostre cervell processa més fàcilment. Per això els gràfics de barres són tan populars. La mida també s'utilitza bastant, per exemple en un gràfic de bombolles. L'agrupació espacial funciona molt bé en un gràfic de dispersió on puguem detectar un grup d'elements clarament diferenciats de la resta. La resta d'atributs es fan servir en menor grau, però també ens poden resultar útils en certes situacions.

5.4. Interacció

Els elements d'interacció ens permeten adaptar la visualització de dades a unes necessitats informatives concretes. Ben Shneiderman va definir l'any 1996 el mantra de la cerca visual d'informació (*Information Seeking Mantra*) com: “mostrar una perspectiva general, fer zoom i filtrar, i després proporcionar detalls sota demanda”. Aquesta manera d'entendre la visualització requereix l'ús de la interacció. A continuació, expliquem els tres tipus d'interaccions més comuns.

Descripcions emergents

Les anomenades descripcions emergents (*tooltips*), són petites caixes amb text i/o visualitzacions que apareixen quan seleccionem un element de la nostra visualització. Són el tipus d'interacció bàsica per proporcionar detalls sobre un element pel qual s'interessa l'usuari.

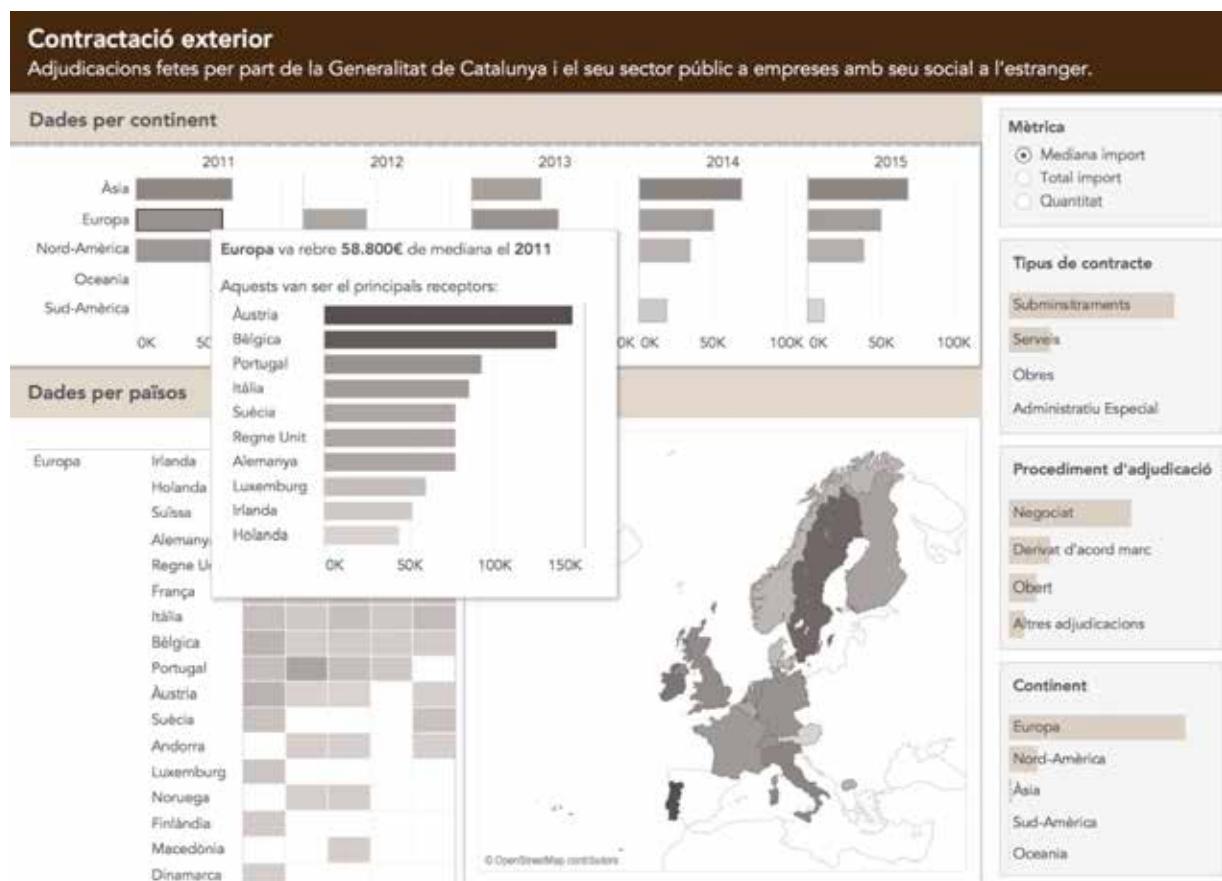


Figura 5.7. Quan passem per sobre la barra “Europa”, una descripció emergent ens mostra un gràfic de barres amb el detall dels països receptors.

Enllaç i marcatge

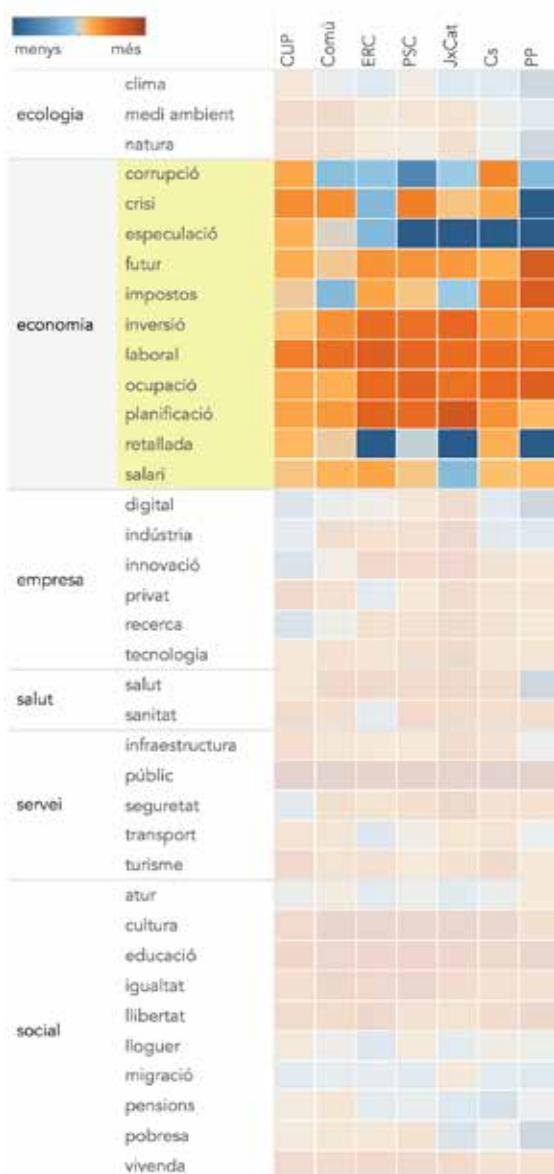
El marcatge (*brushing*) és una tècnica d'interacció que permet que l'usuari seleccioni un conjunt d'elements de la visualització i que fa que quedin ressaltats en una o més visualitzacions. Quan hi ha més d'una visualització sobre les mateixes dades disponible, el fet que totes elles ressaltin els mateixos elements s'anomena “enllaç” (*linking*).

A la següent imatge podem veure un exemple de visualització que utilitza la tècnica de l'enllaç i marcatge. A la visualització titulada *De què parlen els partits del 21-D* es representa, mitjançant un gràfic d'intensitats per colors, la freqüència amb que cada partit polític utilitza una paraula en el seu programa electoral. La intensitat dels colors ens ajuda a veure si una paraula s'ha utilitzat més o menys. Paral·lelament, uns gràfics de barres mostren la utilització de les diferents paraules ordenades de més a menys utilitzades. Com es pot apreciar a la imatge, quan es ressalta un conjunt de paraules, aquestes es marquen també als gràfics de barres i ens permeten veure si aquestes són les més o menys utilitzades a cada programa electoral.

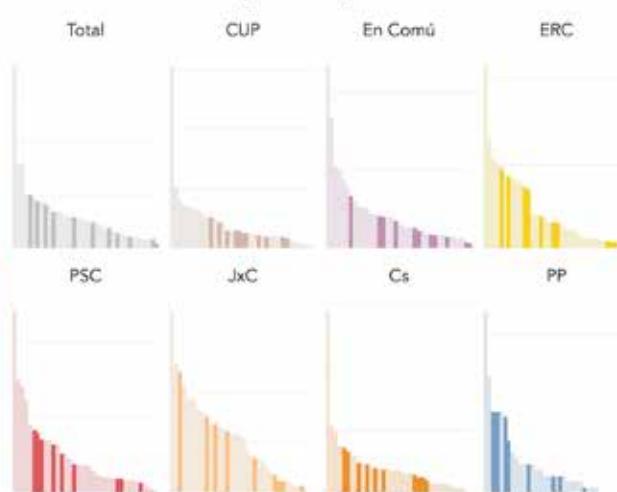
De què parlen els partits del 21-D?

Visualitzem el nombre de mèncions de diferents conceptes en els seus programes

Quins termes es mencionen més?



Quina és la distribució per cada partit?



Limitacions de l'anàlisi

Aquesta visualització de dades és útil per obtenir una panoràmica general dels programes dels diferents partits polítics. Ara bé, no sempre ho farà de manera acurada. Que un partit mencioni poc un terme no vol dir necessàriament que no li doni importància. Per exemple, un partit potser menciona poc el terme "corrupció", però les seves polítiques són molt fermes contra la corrupció.

Per entendre bé els programes, cal llegir-los. Per això, aquí us oferim enllaços a cada programa dels partits polítics que es presenten el 21D. ...

[CUP](#) [Comú](#) [ERC](#) [PSC](#)
[JxCat](#) [Cs](#) [PP](#)

Metodologia

Per un recull de 45 paraules, hem comptat quantes vegades es mencionen a cada programa dels partits polítics que es presenten el 21-D. Hem ponderat els resultats per tal de poder comparar independentment de la longitud dels programes. Gràcies a <https://twitter.com/danirdlpz> per la recopilació i tractament de les dades.

Visualització realitzada per **OneTandem**

ONETANDEM

Figura 5.8. Exemple d'enllaç i marcatge utilitzat a la visualització titulada “De què parlen els partits del 21-D” feta per l’empresa OneTandem. Font: <https://public.tableau.com/profile/onetandem#!/vizhome/ProgrameseleccionsGeneralitatCatalunya21D/he-atmap>

5.5. Recomanacions generals

A continuació es descriuen una sèrie d'aspectes a tenir en compte a l'hora de dissenyar les nostres visualitzacions.

5.5.1. Menys és més

Edward Tufte és un dels autors més respectats de la visualització de dades. El seu llibre *The visual display of quantitative information*⁹ es considera tot un clàssic de la disciplina. En aquest llibre, descriu l'excel·lència gràfica de la següent manera:

“Per damunt de tot, cal representar dades (...). L'excel·lència gràfica és aquella que dona a l'usuari el major nombre d'idees, en el període més curt de temps, utilitzant la mínima quantitat de tinta, en l'espai més petit possible”

El següent GIF animat mostra clarament com aplicar aquest principi al disseny de gràfics:

<http://www.darkhorseanalytics.com/blog/data-looks-better-naked>

5.5.2. Equilibri entre funcionalitat i estètica

L'equilibri entre funcionalitat i estètica està supeditat als objectius de la visualització. Generalment, si busquem la complicitat del públic general, l'estètica pot arribar a prevale per sobre de la funcionalitat, sempre i quan siguem rigurosos. Per exemple, el Better Life Index de l'OCDE, mostrat en la figura següent, és un exemple de visualització on l'estètica cerca la complicitat de l'usuari, per sobre de la funcionalitat analítica.

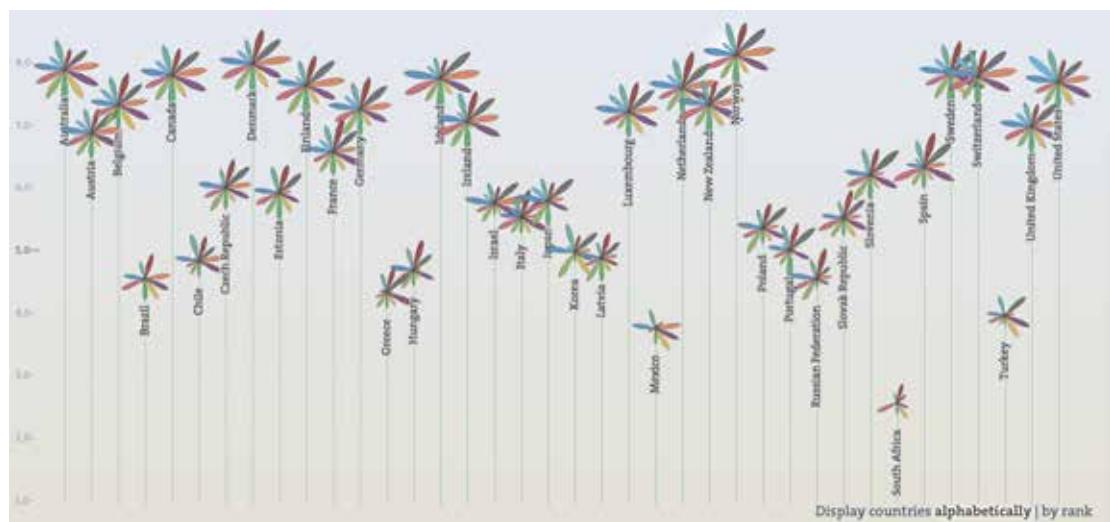


Figura 5.9. Captura parcial de la visualització del “Better Life Index” de l'OCDE. Font: <http://www.oecdbetterlifeindex.org/>

Tanmateix, és molt important entendre que l'objectiu principal de la visualització de dades en un entorn analític ha de ser generar coneixement.

⁹ Tufte, Edward R. *The visual display of quantitative information*. Graphic Press, 1983

5.5.3. La forma segueix la necessitat

La forma triada dependrà sempre de l'objectiu informatiu. No podem separar la forma de la necessitat. Alberto Cairo posa un bon exemple en el seu llibre *The Functional Art*, en el qual es mostra la visualització següent:

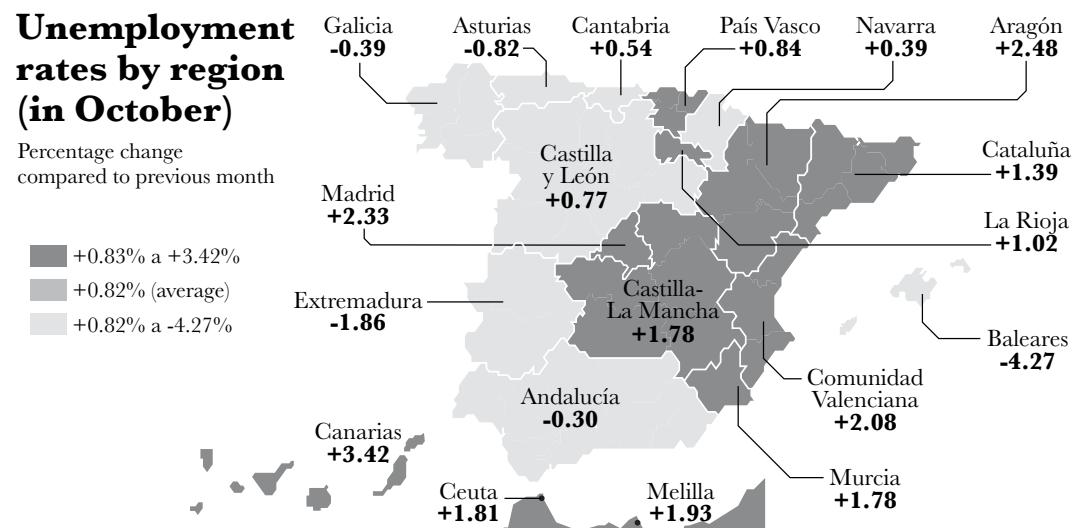


Figura 5.10. Representació de la taxa d'atur per comunitat autònoma. Imatge cedida per Alberto Cairo del seu llibre *El Arte Funcional*¹⁰.

L'objectiu de la visualització és descobrir quines comunitats autònomes han millorat més i quines estan tenint més problemes en comparació amb el mes anterior. Tanmateix, la forma triada fa que aquesta tasca sigui complicada, ja que només s'han utilitzat tres gradacions de color i, dins de les comunitats amb un mateix color, no queda més remei que inspeccionar els noms, intentar memoritzar-los i després mirar d'inferir l'ordre dels valors.

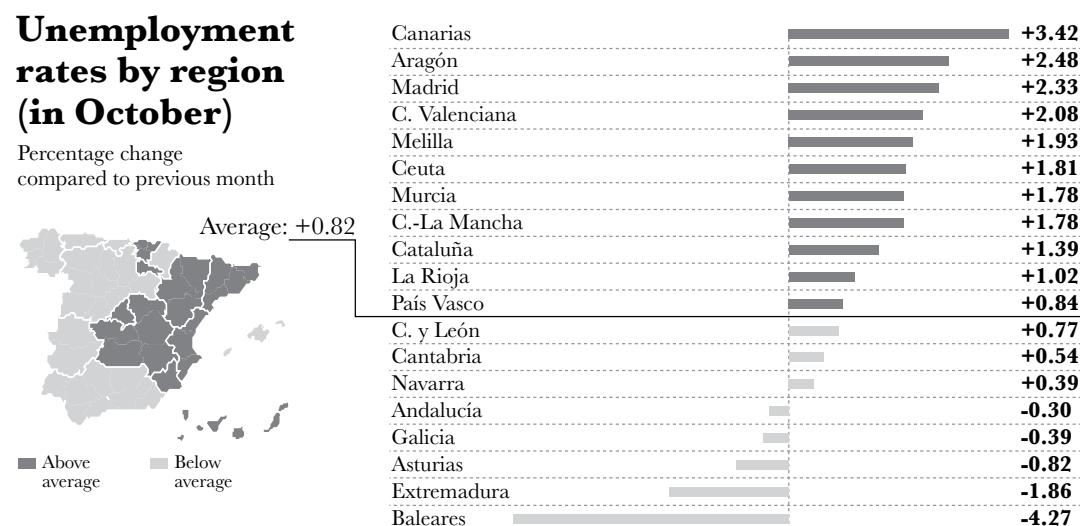


Figura 5.11. Redisseny plantejat per Alberto Cairo. Imatge cedida pel mateix autor del seu llibre *El Arte Funcional*.

¹⁰ Cairo, Alberto. *El arte funcional. Infografía y visualización de información*. Alamut 2011.

El redisseny del mapa proposa la utilització d'un gràfic de barres que permet ordenar fàcilment les comunitats autònomes, segons el canvi en la taxa d'atur. El mapa es manté, però no per mostrar valors numèrics, sinó per evidenciar patrons geogràfics (que hauria de ser sempre la utilitat d'un mapa).

5.5.4. Text

A vegades, en una visualització de dades, no es dona al text la importància que mereix. Per exemple, un titular adequat pot ser la clau per atraure l'atenció del lector. Així mateix, els textos que acompanyen els gràfics en forma de titolars o anotacions són necessaris per assegurar-ne la correcta comprensió. Fins i tot, el text pot ser l'element principal de la visualització. Ho podem veure clarament amb un exemple.

El gràfic següent mostra el nombre de passes que ha realitzat una persona al llarg d'un dia.

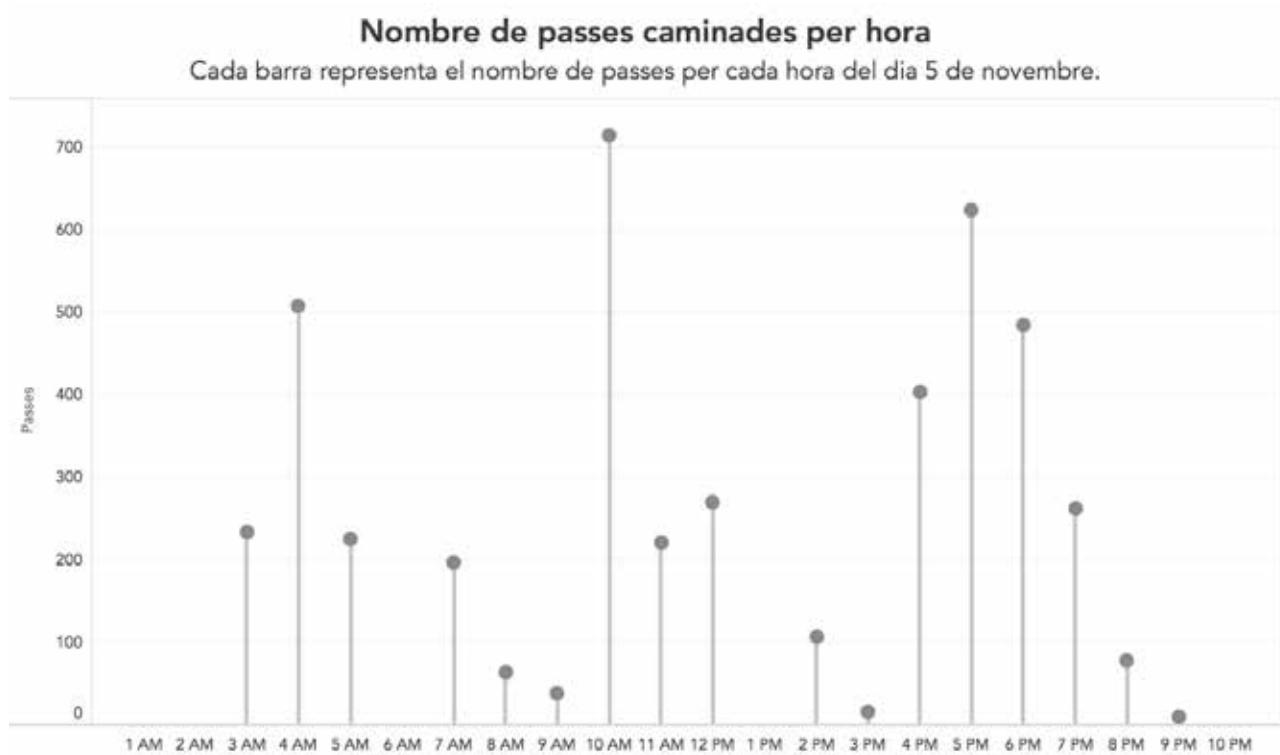


Figura 5.12. Nombre de passes en un dia.

Es fa servir un text genèric que descriu el contingut del gràfic. El text no és incorrecte, però no ajuda a entendre el significat de les dades. Fixeu-vos com canvia el mateix gràfic, fent servir altres textos:

El camí de la vida

La història del dia en què **va néixer el meu fill Teo**, explicada a través de les passes que vaig fer cada hora.

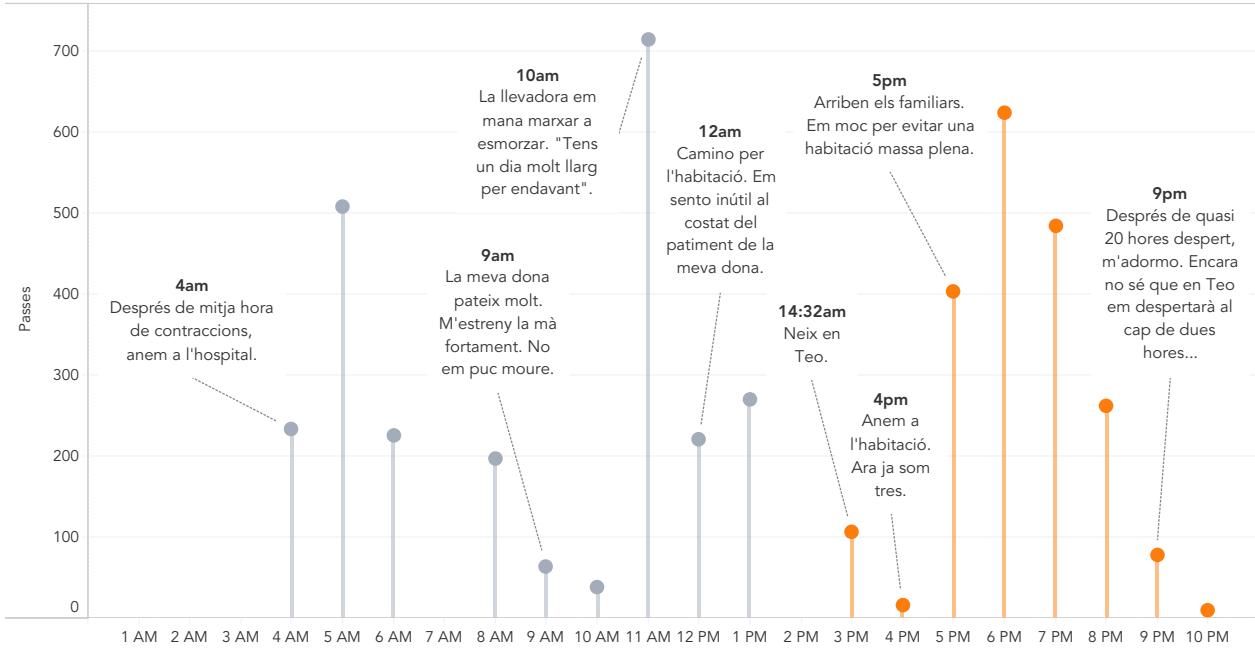


Figura 5.13. “El camí de la vida”. Imatge cedida per OneTandem.

Un gràfic simple esdevé una història personal, fins i tot emotiva. Les dades són exactament les mateixes, però, gràcies al text, les situem en context, entenem què hi ha darrera i quin significat tenen. Com veiem, en una visualització, el text pot ser fins i tot més important que les pròpies dades numèriques.

5.5.5. Maquetació

Quan parlem de maquetació ens referim a la disposició dels diferents elements de la visualització de dades: gràfics, textos, imatges, filtres, etc. En general, recomanem reservar espai per a:

- la informació sobre la metodologia i les fonts de dades, que generalment anirà al peu de la visualització
- els filtres i selectors, que generalment aniran agrupats en una columna a la dreta o l'esquerra, o en una fila a la part superior
- els mecanismes d'ajuda, que generalment aniran al costat de cada gràfic (en aquest cas també es pot situar una icona d'ajuda general a la cantonada superior dreta).

També convé recordar que generalment el públic llegeix de dalt a baix i d'esquerra a dreta i, per tant, la informació més important s'ha de situar a la cantonada superior esquerra.

6. Bibliografia

Qlik. *Data Equality Campaign*. <<http://dataequality.org/>> [recurs electrònic]. [Consultat: novembre 2017].

Ware, Colin. *Information Visualization, Third Edition: Perception for Design*. Morgan Kaufmann, 2012

Wikipedia. *Interactive Storytelling*.

<https://ca.wikipedia.org/wiki/Projecci%C3%B3_cartogr%C3%A0fica> [Consulta: novembre 2017]

Nussbaumer, Cole. *A Data Visualization Guide for Business Professionals*. Wiley, 2015

Wikipedia, *Five Ws*. <https://en.wikipedia.org/wiki/Five_Ws> [Consulta: novembre 2017]

Wikipedia, *Projecció cartogràfica*.

<https://ca.wikipedia.org/wiki/Projecci%C3%B3_cartogr%C3%A0fica> [Consulta: novembre 2017]

Miller, George Armitage. *The magical number seven, plus or minus two: Some limits on our capacity for processing information*. Psychological Review, 1956

Tufte, Eduard R. *The visual display of quantitative information*. Graphic Press, 1983

Cairo, Alberto. *El arte funcional. Infografía y visualización de información*. Alamut 2011.

A continuació citem alguns llibres que poden ser d'utilitat per aprofundir en el coneixement de la matèria de la visualització de dades.

Few, Stephen. *Now You See It*. Analytics Press, 2009.

Meirelles, Isabel. *Design for Information*. Rockport Publishers, 2013.

Cotgreave, Andy; Shaffer, Jeffrey; Wexler, Steve. *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business*. Wiley, 2017.

Cairo, Alberto. *The Truthful Art*. [s. n.] 2016.

7. Annexos

7.1. Glossari

Anàlisi de clústers: L'anàlisi de clústers és una tècnica d'anàlisi de dades que permet, de manera automàtica, segmentar un conjunt de dades en subgrups (anomenats clústers) que tinguin valors similars a les seves variables.

Dimensió: Aquelles variables de tipus textual o categòric que ens permeten segmentar les nostres dades.

Indicador: També anomenat mètrica. Un indicador és una variable de les dades de tipus numèric que pot ser mesurada i agregada.

Quartil: Els quartils d'un conjunt ordenat de dades són els tres punts de tall que divideixen el conjunt de dades en quatre grups de la mateixa mida

(definició de Viquipèdia: <https://ca.wikipedia.org/wiki/Quartil>).

Variable: Tot atribut de les nostres dades, ja sigui numèric o de text. Per exemple, en un full de càcul, les variables del nostre conjunt de dades són les columnes.

Trobareu més termes relacionats amb aquesta matèria a la Terminologia de la visualització de dades del Termcat: <http://gen.cat/termvisdades>

7.2. Eines

La disciplina de la visualització de dades està en ple auge. Per això constantment apareixen eines que faciliten el procés de creació de visualitzacions. A continuació es presenten les eines més interessants actualment.

Tableau (<https://www.tableau.com/>): Tableau s'ha convertit en l'eina referent en el món de la visualització de dades, especialment quan s'utilitza en els camps de la intel·ligència de negoci o l'analítica visual. Tableau possibilita canviar el tipus de visualització amb un sol clic, i facilita la creació de gràfics interactius. A més a més, l'eina facilita la connexió a múltiples fonts de dades que van des dels fulls de càlcul més senzills, a bases de dades, o a plataformes de dades massives.

Tableau es pot utilitzar en les fases de prototipatge i de finalització de les visualitzacions, però també pot ser una bona eina per explorar visualment les dades en la fase d'estratègia.

Tot i que Tableau és fàcil d'utilitzar per crear gràfics senzills, la corba d'aprenentatge és força gran quan es volen fer visualitzacions avançades. Per últim, és interessant saber que, malgrat que és una eina de pagament, també té una versió gratuïta (Tableau Public <https://public.tableau.com/s/>).

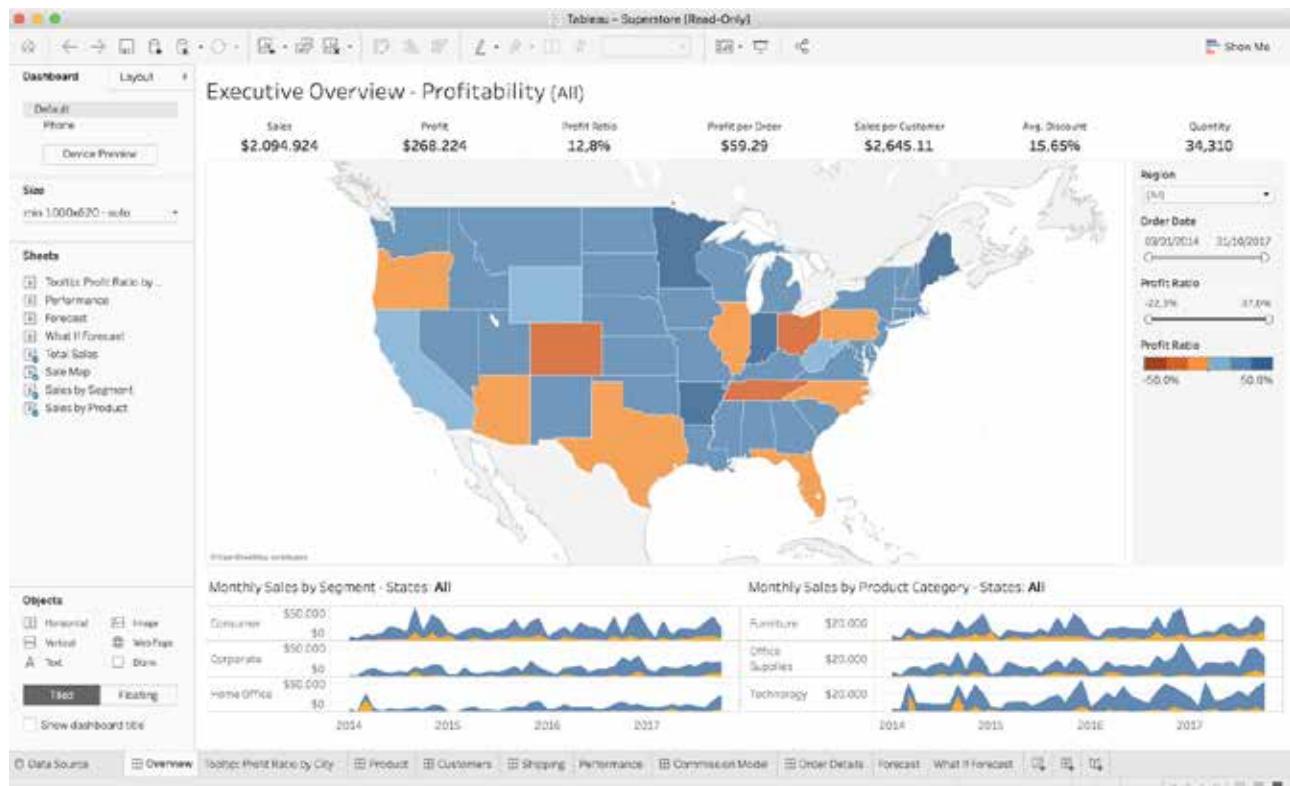


Figura 7.5. Captura de pantalla d'un dels dashboards que proporciona Tableau en la seva versió Desktop.

DataWrapper (<https://www.datawrapper.de/>): Aquesta eina ha guanyat darrerament molta popularitat. Tot i ser molt més senzilla que Tableau, permet pujar les dades en un full de càlcul i triar entre una gran varietat de gràfics. Cadascun disposa d'una sèrie d'opcions per personalitzar-los i adequar-los a les nostres necessitats.

En tractar-se una eina web, DataWrapper permet incrustar de manera molt senzilla les visualitzacions que es creen en una pàgina web o blog. Tanmateix, cal tenir en compte que DataWrapper és una eina de pagament.

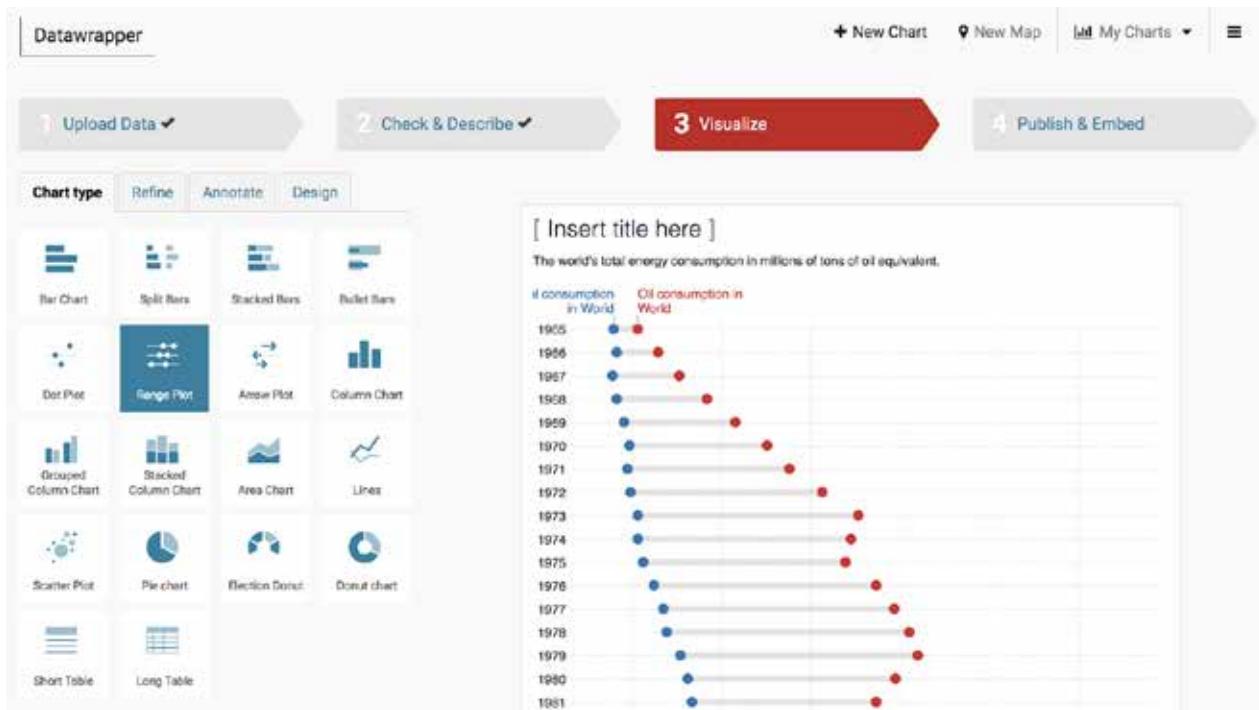


Figura 7.6. Captura de pantalla de DataWrapper.

Adobe Illustrator (<https://www.adobe.com/es/products/illustrator.html>): Aquesta popular eina d'il·lustració s'utilitza molt per esbossar, prototipar i finalitzar visualitzacions. Sovint s'utilitza per afegir els darrers detalls a les visualitzacions fetes amb altres eines. No obstant això, la corba d'aprenentatge és molt gran. Adobe Illustrator és una eina de pagament.

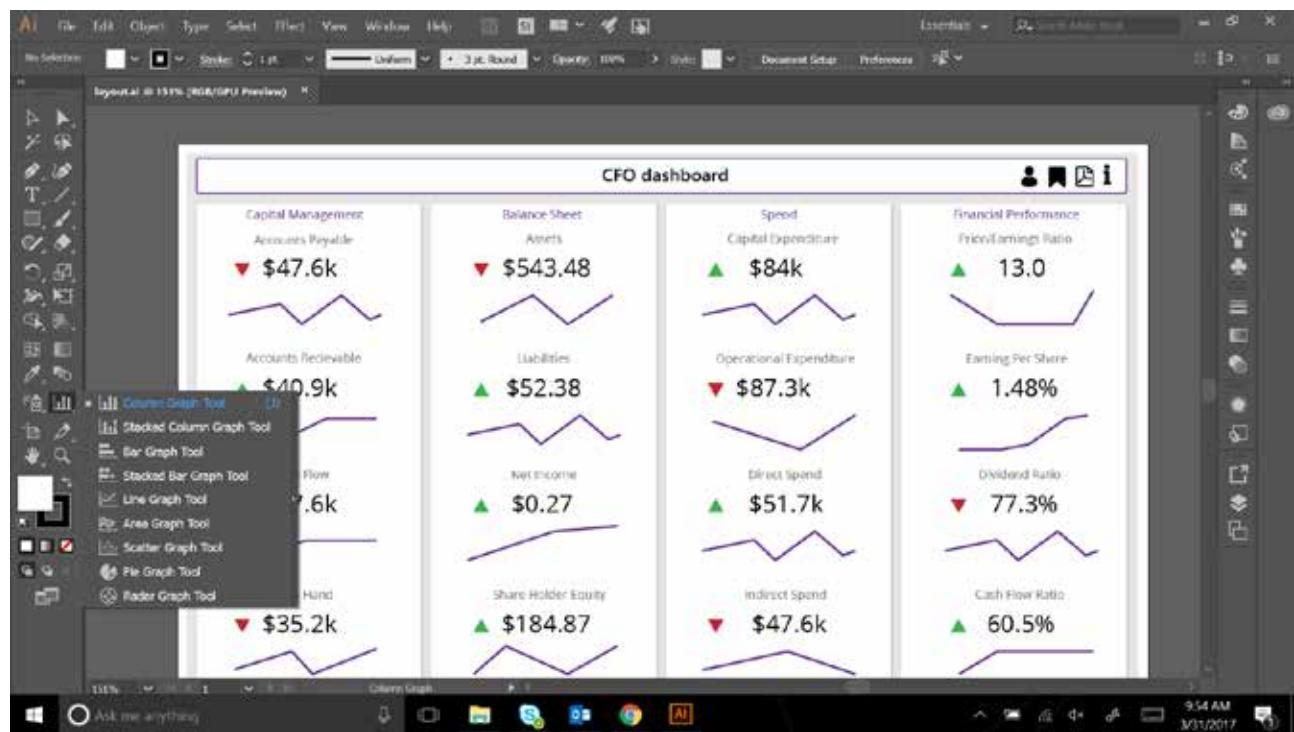


Figura 7.7. Captura de pantalla d'un quadre de comandament fet amb Adobe Illustrator.
Font: https://cdn-images-1.medium.com/max/1600/1*h7ol80RsF6ajio5yboEAJg.png.

SketchApp (<https://www.sketchapp.com/>): SketchApp s'ha convertit en una alternativa a l'Adobe Illustrator que fan servir els dissenyadors d'interfícies d'usuari i els visualitzadors de dades. És molt més senzilla d'utilitzar que l'Adobe Illustrator i resulta especialment adequada per a la fase de prototipatge de la nostra visualització. SketchApp és una eina de pagament.



Figura 7.8. Captura de pantalla d'un quadre de comandament prototipat amb SketchApp.

Font: <https://www.sketchappsources.com/resources/source-image/collection-of-charts-bykova.png>.

Microsoft Excel (<https://products.office.com/en/excel>): Aquest popular full de càlcul encara és una eina referent en el món de l'anàlisi de dades. Històricament ha facilitat la creació de tota mena de gràfics, tot i que a poc a poc s'han anat descobrint mancances en el camp de la visualització. En les darreres versions s'ha anat adaptant i cada vegada proporciona més varietat de visualitzacions modernes.

Microsoft Excel és una eina adequada en la fase de prototipatge i en la de finalització. És de pagament.

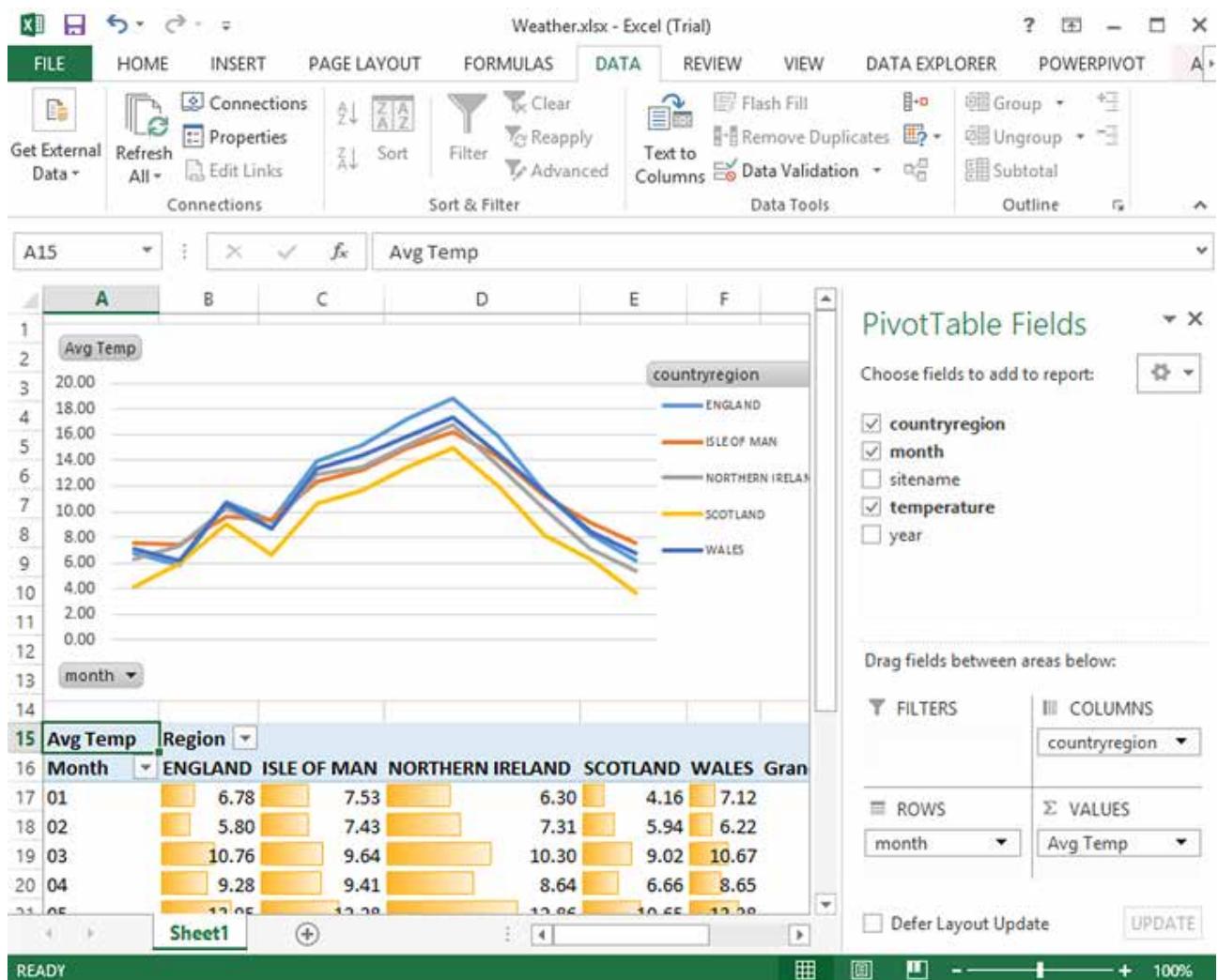


Figura 7.9. Captura de pantalla d'un quadre de comandament desenvolupat amb Microsoft Excel.
Font: <https://msdn.microsoft.com/en-us/library/dn749860.aspx>.

Microsoft PowerBI (<https://powerbi.microsoft.com/es-es/>): Juntament amb Tableau, és una de les eines de visualització de dades més populars en els sectors de la intel·ligència de negoci. El fet d'estar integrada amb la resta d'eines de Microsoft per processar grans volums de dades està fent que molts usuaris la triïn per desenvolupar les seves visualitzacions i quadres de comandament.

Un dels punts forts de Microsoft Power BI és que permet la creació de quadres de comandament de manera molt ràpida.

Microsoft PowerBI es pot fer servir a les etapes de prototipatge i de finalització. L'eina disposa de plans d'utilització, un d'ells és gratuït.

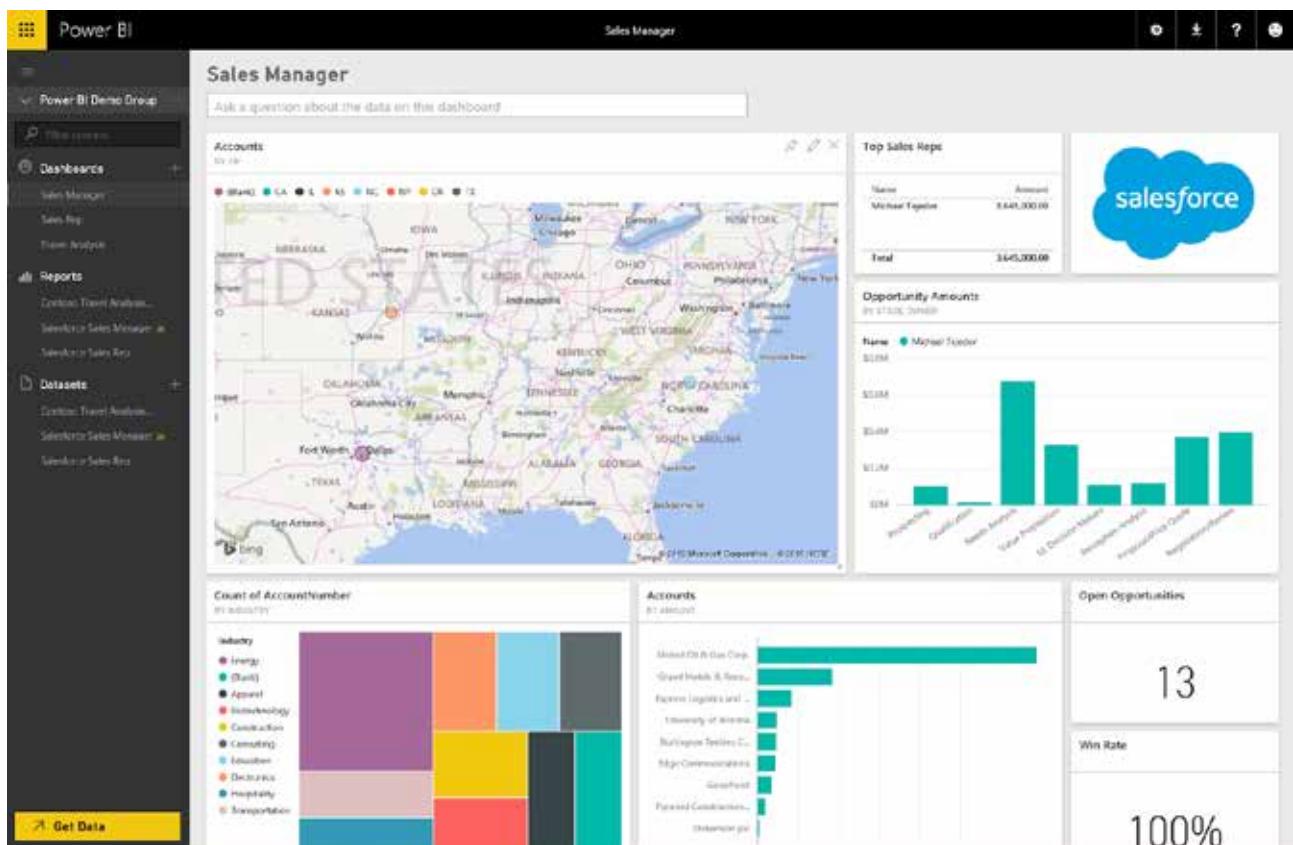


Figura 7.10. Captura de pantalla d'un quadre de comandament fet amb Microsoft PowerBI.

Font: <https://blogs.microsoft.com/blog/2015/09/18/microsoft-expands-computer-science-education-partnership-with-salesforce-and-your-mind-with-hacking-mars-design-challenge-weekend-reading-sept-18-edition/>.

Instamaps (<https://www.instamaps.cat>): És una eina gratuïta per fer mapes dirigida a usuaris no experts; els mapes es poden publicar a internet de forma ràpida i senzilla, en obert o privadament, dins de blogs o pàgines web.

Pot posicionar elements puntuals com una plaça, una font o un comerç, itineraris com ara rutes o curses o bé assenyalar polígons com una zona industrial o una àrea d'estudi.

Instamaps pot fer servir informació continguda en arxius geogràfics (shp, gpx, kml, etc.) o altres arxius (xlsx, csv, taules de Google, etc.), així com arxius i serveis provinents de portals de dades obertes, geoserveis wms o altra informació en línia.

Instamaps és una plataforma oberta a tothom creada per l'Institut Cartogràfic i Geològic de Catalunya (<https://www.icgc.cat>).

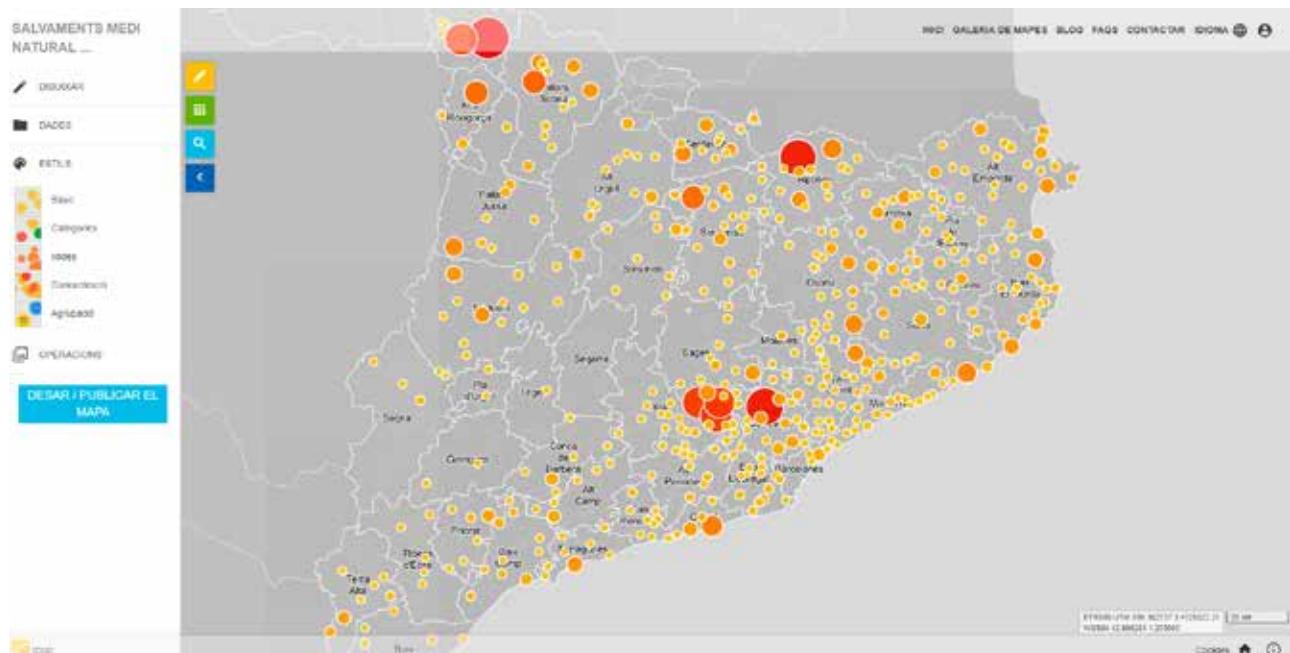


Figura 7.11. Captura de pantalla del mapa de Salvaments del Medi Natural

Font: https://www.instamaps.cat/instavisor/498de3e1152b0d1d5ab20cc6ba5a8956/SALVAMENTS_MEDI_NATURAL_2017.html#8/41.882/1.208

D3 (<https://d3js.org/>) : Tot i no ser una eina, cal esmentar D3, una llibreria feta amb el llenguatge de programació Javascript. Aquesta llibreria és una de les més utilitzades pels professionals del sector. Com que no és una eina, sinó una llibreria que facilita la programació de visualitzacions per al web, no ofereix límits a l'hora de crear visualitzacions i interaccions que vagin més enllà de les solucions més comunes i estàndards.

Aquesta llibreria és de codi obert, de manera que es pot utilitzar sense cap cost.



Figura 7.12. Captura d'un collage fet amb D3 que mostra visualitzacions desenvolupades amb aquesta llibreria.
Font: <https://d3js.org/>.

R (<https://www.r-project.org/>) : A cavall entre les llibreries de programació i els programaris més avançats, R permet fer anàlisis de dades molt complexes i generar visualitzacions gràcies a paquets com ggplot (<https://www.statmethods.net/advgraphs/ggplot2.html>) o Shiny (<https://shiny.rstudio.com/>). Concretament, Amb Shiny es poden crear visualitzacions interactives amb una sintaxi semblant a l'HTML (llenguatge de programació per crear pàgines web). Resulta força complexa i és utilitzada bàsicament per perfils tècnics.

R és útil en les fases de prototipatge i de finalització, però també pot ser una bona eina per explorar visualment les dades en la fase d'estratègia. Un dels principals avantatges d'R és que és una eina de programari lliure.

