
Bones pràctiques en visualització de dades

PID_00249218

Víctor Pascual Cid

Temps mínim de dedicació recomanat: 2 hores





Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-Compartir igual (BY-SA) v.3.0 Espanya de Creative Commons. Podeu modificar l'obra, reproduir-la, distribuir-la o comunicar-la públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), i sempre que l'obra derivada quedi subjecta a la mateixa llicència que el material original. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-sa/3.0/es/legalcode.ca>

Índex

1. Què és la visualització de dades?	5
2. El procés de visualització de dades	8
3. Objectius de la visualització de dades	10
3.1. Comunicar	10
3.2. Explorar	11
3.3. Dades en context	12
3.4. Trobar patrons i <i>outliers</i>	14
4. Regles per desenvolupar bones visualitzacions	16
4.1. Començar amb preguntes	16
4.2. Gràfics autoexplicatius	16
4.3. La funcionalitat ha de prevaler enfront de l'estètica	17
4.4. Ús de la interacció	19
4.5. La forma segueix la necessitat	21
4.6. Preparar bé les dades	22
Resum	24

1. Què és la visualització de dades?

La visualització de dades, com s'entén avui dia, es va començar a estudiar a finals dels anys noranta. El fet que sigui tan recent fa que estigui en constant evolució i contínuament apareixen definicions noves sobre què és visualització de dades i què no ho és. Colin Ware va definir el concepte de visualització com:

«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».

C. Ware (2012). *Information Visualization, Third Edition: Perception for Design*. San Francisco: Morgan Kaufmann.

De la mateixa manera, Alberto Cairo l'ha definit com:

«la representació visual d'informació dissenyada per permetre la comunicació, l'anàlisi, el descobriment i l'exploració».

A. Cairo (2016). *The Truthful Art*. New Riders.

Dins d'aquest camp, es poden distingir dos tipus de visualitzacions:

1) Visualitzacions estàtiques: la seva funció principal sol ser comunicar unes dades que poden haver estat analitzades prèviament. Aquest tipus de visualitzacions han d'ajudar a descobrir patrons i valors atípics (en anglès, *outliers*) en les dades. A més, són les visualitzacions que poden ser utilitzades en formats físics, com ara diaris i revistes en paper. En la figura 1 es pot veure una visualització feta per al diari *El País* en la qual es mostra l'evolució de la taxa d'atur a Espanya en forma de gràfic de línies i l'evolució del percentatge de població activa en atur representada amb un gràfic de barres.



Figura 1. Evolució del nombre d'aturats i taxa d'atur a Espanya. Font: El País.

2) **Visualitzacions interactives:** són les que permeten als usuaris interactuar amb les dades. Aquesta propietat permet que l'usuari pugui explorar les dades i tingui la llibertat de centrar-se en allò que més li interessa. Per exemple, acostumats a les representacions clàssiques de l'evolució de l'atur, com la que es mostrava en la figura anterior, la visualització interactiva de *The New York Times* de la figura 2 permet a l'usuari triar el sector de població del qual vol veure la taxa de desocupació mitjançant l'ús del menú superior. D'aquesta manera, les visualitzacions interactives permeten el descobriment i l'exploració de les dades, a més de comunicar el resultat d'anàlisis prèvies.

The Jobless Rate for People Like You

Not all groups have felt the recession equally.

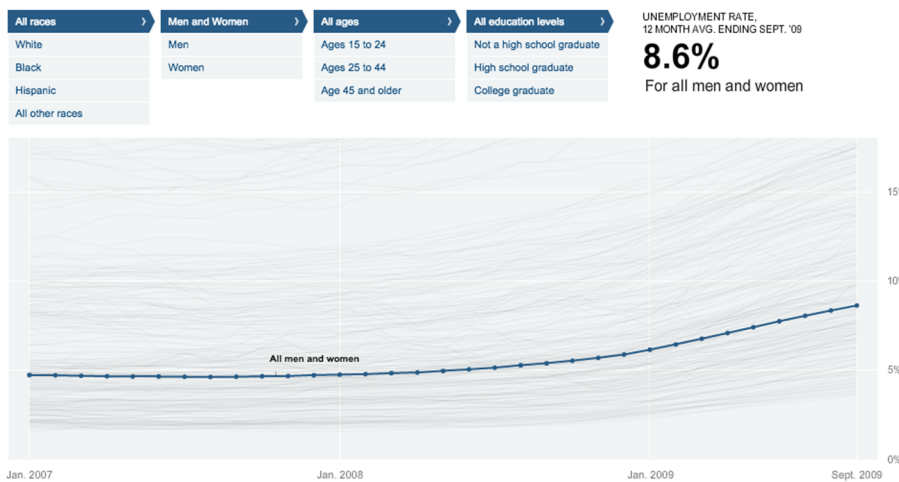
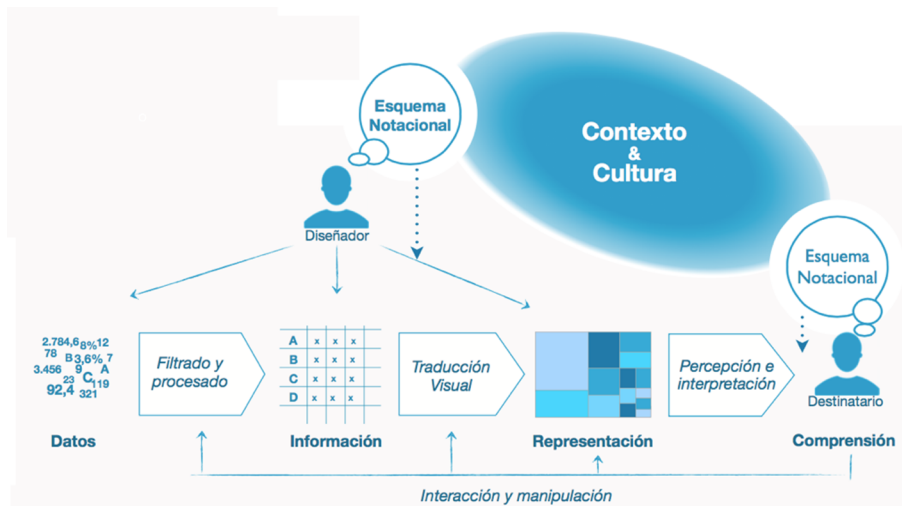


Figura 2. Captura de pantalla de la visualització interactiva de la taxa d'atur als Estats Units. Font: The New York Times.

Triar un tipus de visualització o un altre depèn de l'objectiu que es vol aconseguir. Mentre que les visualitzacions estàtiques poden comunicar dades de manera ràpida i concisa, les interactives permeten a l'usuari explorar-les i jugar-hi.

2. El procés de visualització de dades

La figura 3 representa el diagrama de visualització de dades creat per Juan Carlos Dürsteler i Yuri Engelhardt. Aquest diagrama representa tot el procés que s'ha de fer a l'hora de generar una visualització.



J. C. Dürsteler

Lloc web: <http://infovis.net/>

Y. Engelhardt

Lloc web: <http://yuriengelhardt.com/>

Figura 3. El diagrama de visualització de dades de J. C. Dürsteler i Y. Engelhardt redissenyat per Jaume Pérez.

El diagrama parteix de la idea que es disposa d'unes dades que han estat recollides prèviament. Aquestes dades són filtrades i processades fins a convertir-les en informació, és a dir, fins a tenir-les netes i estructurades. A aquesta informació se li aplica una traducció visual o, dit d'una altra manera, se la representa gràficament. Finalment, és important ser conscients que aquesta representació serà consumida per un usuari destinatari, el qual, òbviament, no ha de ser necessàriament el mateix dissenyador de la visualització. Per aquest motiu, és molt important ser conscients de quin és el perfil del destinatari del disseny de la visualització. No és el mateix publicar una infografia en un diari generalista que generar un *report* per al conseller delegat d'una empresa que necessita prendre una decisió molt ràpidament. Per això serà important ser conscients que l'esquema notacional de cada usuari pot ser diferent, és a dir, que cadascú pot interpretar certes dades de manera diferent. De la mateixa manera, cal tenir en compte que els dissenyadors i els usuaris finals poden compartir o no un mateix context o cultura.

Serà, doncs, de vital importància tenir l'usuari final al cap per evitar possibles errors d'interpretació de la visualització creada. Per a això, s'hauran de fer diferents tests d'usabilitat amb usuaris potencials per assegurar-se de l'adequació del disseny seleccionat. A més, el disseny també haurà de preveure el tipus de dispositiu en el qual s'utilitzarà la representació, ja que no és el mateix disse-

nyar per a una pantalla gran, com la d'un ordinador de sobretaula o un portàtil, que fer-ho per a un dispositiu mòbil molt més petit, on és més difícil fer segons quin tipus d'interaccions.

Finalment, és important emfatitzar que, gràcies a les tecnologies actuals, serà possible desenvolupar sistemes amb els quals l'usuari podrà interactuar i manipular les dades en qualsevol fase del procés. Per exemple, com es veurà més endavant, és molt comú oferir visualitzacions interactives en les quals un usuari pot filtrar les dades utilitzant un menú.

3. Objectius de la visualització de dades

En els apartats anteriors s'ha definit què és la visualització de dades, s'han explicat els diferents tipus de visualitzacions i el procés per dur-les a terme. A continuació ens centrarem en els objectius principals que persegueix una bona representació.

3.1. Comunicar

L'objectiu principal de la visualització de dades és comunicar una idea o el resultat d'una anàlisi feta sobre aquestes dades. En aquest sentit, és interessant ser conscients que les figures 4 i 5 mostren la mateixa informació, però utilitzen diferents representacions. Mentre la taula (figura 4) ajuda a veure el detall concret dels valors, el gràfic de línies (figura 5) permet veure l'evolució temporal del valor de les accions. Totes dues són útils, però el gràfic de línies aporta molta més informació a simple vista. És a dir, el gràfic «comunica millor». Malgrat que és possible triar entre diferents tipus de representacions, amb la visualització de dades es busca seleccionar la més adequada per a transmetre la informació.

Aug 2, 2016	105.60	106.07	104.04	104.07	20,954,862
Aug 1, 2016	104.41	106.15	104.41	106.05	37,499,000
Jul 29, 2016	104.19	104.55	103.68	104.21	27,733,700
Jul 28, 2016	102.83	104.45	102.82	104.34	39,869,800
Jul 27, 2016	104.27	104.35	102.75	102.95	92,344,800
Jul 26, 2016	96.82	97.97	96.42	96.67	56,239,800
Jul 25, 2016	98.25	98.84	96.92	97.34	40,382,900
Jul 22, 2016	99.26	99.30	98.31	98.66	28,313,700
Jul 21, 2016	99.83	101.00	99.13	99.43	32,702,000
Jul 20, 2016	100.00	100.46	99.74	99.96	26,276,000
Jul 19, 2016	99.56	100.00	99.34	99.87	23,779,900
Jul 18, 2016	98.70	100.13	98.60	99.83	36,493,900
Jul 15, 2016	98.92	99.30	98.50	98.78	30,137,000
Jul 14, 2016	97.39	98.99	97.32	98.79	38,919,000
Jul 13, 2016	97.41	97.67	96.84	96.87	25,892,200

Figura 4. Taula de valors de les accions d'una empresa en el NASDAQ extret de *Yahoo! Finance*. L'absència d'elements visuals que cridin l'atenció del sistema visual preatentiu limita la comprensió de les dades.

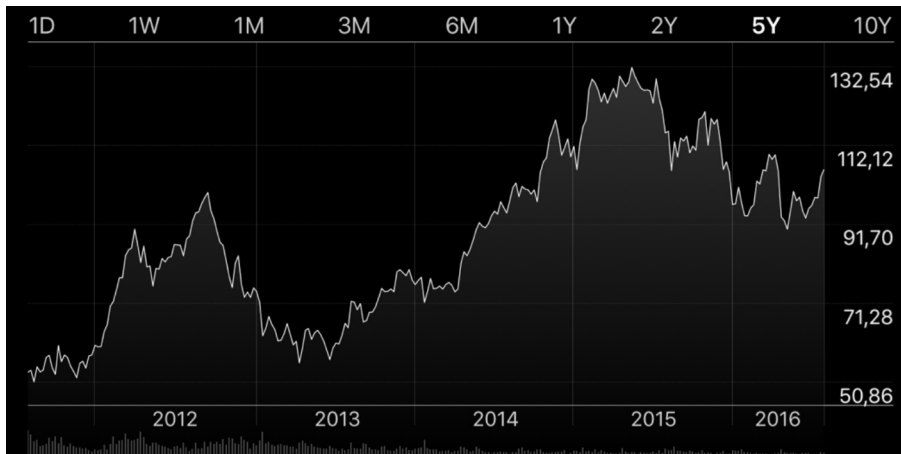


Figura 5. Gràfic de línies que representa l'evolució dels valors de les accions d'una empresa en el NASDAQ de l'aplicació Stock de Yahoo! La codificació de les dades en una línia fa que la memòria de treball processi totes les dades representades en la línia com un sol bloc, la qual cosa en facilita la comprensió.

Aquest fenomen es deu al fet que el cervell humà funciona de manera molt semblant a un ordinador, en el qual es disposa de memòria RAM (la que s'utilitza per carregar programes que són utilitzats al moment, ja que és d'accés ràpid) i del disc dur, que és el magatzem de dades que podem consultar quan vulguem, però que és d'accés més lent. En lloc de memòria RAM, en el cervell humà hi ha l'anomenada *memòria de treball*, mentre que en lloc de disc dur es té l'anomenada *memòria a llarg termini*. Quan s'observa una visualització, el sistema visual humà envia la informació del que es veu en blocs a la memòria de treball. No obstant això, se sol dir que la memòria de treball només pot emmagatzemar, de mitjana, uns set blocs (depenent del tipus d'informació, del tipus d'activitat i d'altres factors). Aquests blocs són seleccionats per la memòria icònica, que és la primera que respon als estímuls visuals (com, per exemple, formes, colors, contrastos, curvatures i grandàries). Quan s'ha assimilat la informació de la memòria de treball, aquesta passa a formar part de la memòria a llarg termini.

Això és el que explica que sigui molt més fàcil entendre les dades mitjançant la figura 5, que codifica tots els nombres d'una columna en una única línia temporal. Tanmateix, en la figura 4 no hi ha cap element que es destaquï, res que ens cridi l'atenció (que activi la nostra memòria icònica) i, per tant, no és possible «carregar» tota aquesta informació en la memòria de treball, cosa que en dificulta la comprensió.

3.2. Explorar

El 1977, l'estadístic John W. Tukey va proposar el concepte d'*anàlisi de dades exploratòria* (sovint anomenada EDA, com a acrònim d'*exploratory data analysis*). La proposta de Tukey consisteix a analitzar les dades mitjançant la realització de successives visualitzacions que ajudin a entendre'n les característiques principals.

Referència bibliogràfica

G. A. Miller (1956). «The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information». *Psychological Review*.

Per saber-ne més

Exploratory data analysis

Ara bé, en els últims vint anys el concepte d'*explorar* ha pres una nova dimensió gràcies als avenços tecnològics. I és que la ràpida capacitat de càlcul dels dispositius d'avui dia permet l'existència de visualitzacions interactives en les quals la representació visual pot canviar de forma o proporcionar nova informació a partir de les interaccions de l'usuari. La potència d'aquest concepte rau en el fet que, mentre s'interactua amb les dades, la visualització també pot mostrar noves perspectives d'aquestes dades. Aquests nous punts de vista poden ajudar al descobriment de fets que ni tan sols s'estaven buscant, la qual cosa la converteix en una eina d'anàlisi de gran capacitat.

Per entendre millor la potència d'aquest concepte, es recomana dedicar uns minuts a explorar la visualització de *The New York Times* «How Different Groups Spend Their Day». Aquesta visualització mostra el percentatge de ciutadans nord-americans que dediquen temps a activitats com dormir, menjar, treballar, etc. al llarg del dia. Abans de començar a utilitzar la visualització, és important pensar què es vol saber sobre les dades que representa: què fan els nord-americans quan surten de treballar?; dormen més els homes que les dones?; com canvien els percentatges entre els qui no tenen fills i els qui en tenen més de dos?

Com a usuaris d'aquesta visualització, quan s'hi hagi interactuat, és important reflexionar sobre totes les altres dades que s'han anat descobrint gràcies a les capacitats interactives del gràfic. Així, és possible suposar que les dades més crucials per a un treball o objectiu concrets estan representades de manera interactiva i també es pot imaginar el coneixement que se'n podria arribar a extreure.

3.3. Dades en context

Ben Shneiderman, el 1996, va proposar l'*information-seeking mantra* o **mantra de la cerca d'informació**:

«Overview first, zoom and filter, then details-on-demand»

B. Shneiderman (1996). «The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations». A: *Proceedings of the IEEE Symposium on Visual Languages*. Washington: IEEE Computer Society Press (pàg. 336-343).

El que proposa Shneiderman amb el seu mantra és que per buscar informació primerament cal proporcionar una vista general de les dades, després cal filtrar-les i obtenir detalls d'allò que més interessa. Representar una vista general de totes les dades permet tenir un context d'aquestes dades, de manera que sigui possible entendre, per exemple, si un valor concret és gran o petit en funció de com és en el conjunt de les dades.

En aquest sentit, el «Billion Dollar-O-Gram» de la figura 6 és una visualització dissenyada per David McCandless amb l'objectiu de fer tangibles grans valors que cada dia es poden veure a la premsa. Sovint passa que quan s'esmenten volums de diners (o de qualsevol altra magnitud) tan grans es fa difícil arribar a

The New York Times

Lloc web: <http://www.nytimes.com/interactive/2009/07/31/business/20080801-metrics-graphic.html>

entendre i comparar quant és poc, molt o moltíssim. En aquesta representació es poden veure conceptes com el que es va predir que costaria la guerra de l'Iraq enfront del que ha costat fins a avui (requadres de color lila a la cantonada superior dreta), o que el valor d'Apple i Google al mercat és similar (quadrats marró fosc i marró clar al centre).



Figura 6. «The Billion Dollar-O-Gram» posa diferents valors en context. Font: <http://www.sickchirpse.com/wp-content/uploads/2013/01/Billion-Dollar-O-Gram-Dave-McCandless-USA-version.jpg>

És molt important ser conscients que aquesta visualització possibilita tenir una vista general de les dades i ajuda a posar en context cadascun dels valors, la qual cosa relativitza alguns d'aquests nombres tan grans. No obstant això, és vital entendre que, com es pot veure amb els quadrats que simbolitzen Apple i Google, aquesta representació no és bona per fer comparacions precises, ja que el cervell humà no està preparat per comparar àrees. Tanmateix, sí que ho està per longituds o distàncies, com es pot observar en el redisseny de la visualització mostrat en la figura 7.

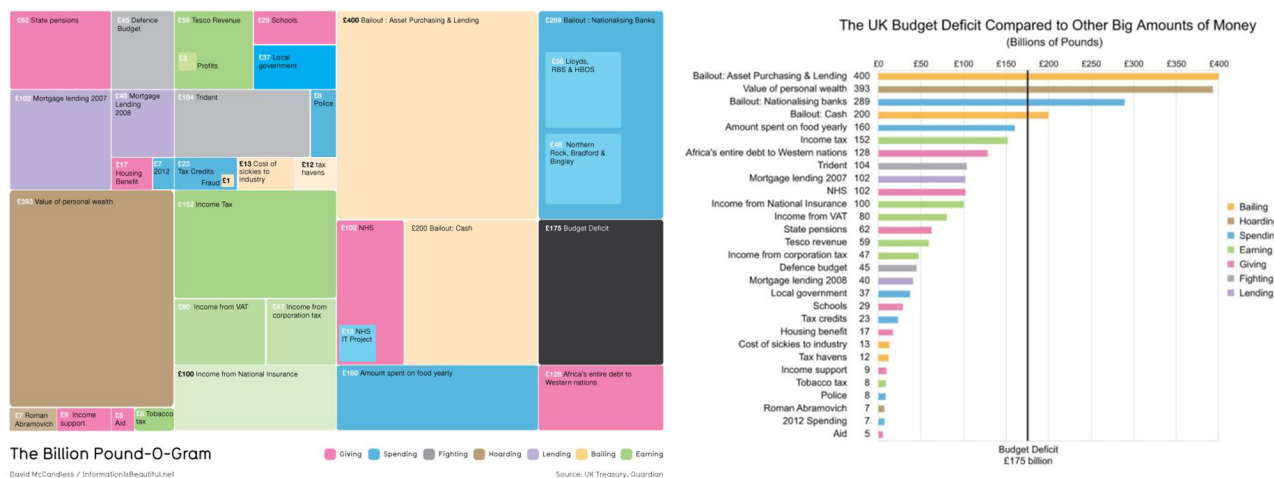


Figura 7. Proposta de redisseny del «Billion Pound-O-Gram» per a facilitar la comparació feta per Stephen Few. La idea consisteix a utilitzar un gràfic de barres horitzontals, que facilita molt la comparació entre els diferents valors.

3.4. Trobar patrons i outliers

Un altre objectiu fonamental de la visualització de dades és facilitar el descobriment de patrons o *outliers* (terme anglès per a referir-se a ‘valors atípics’). El gràfic de la figura 8 n’és un exemple clàssic, en el qual es pot veure l’estacionalitat dels accidents als Estats Units des del 2001 fins al 2011. A més de poder apreciar que els accidents augmenten durant els mesos d’estiu, ja que hi ha un patró anual, també es pot observar un descens del nombre d’accidents a partir del 2008.

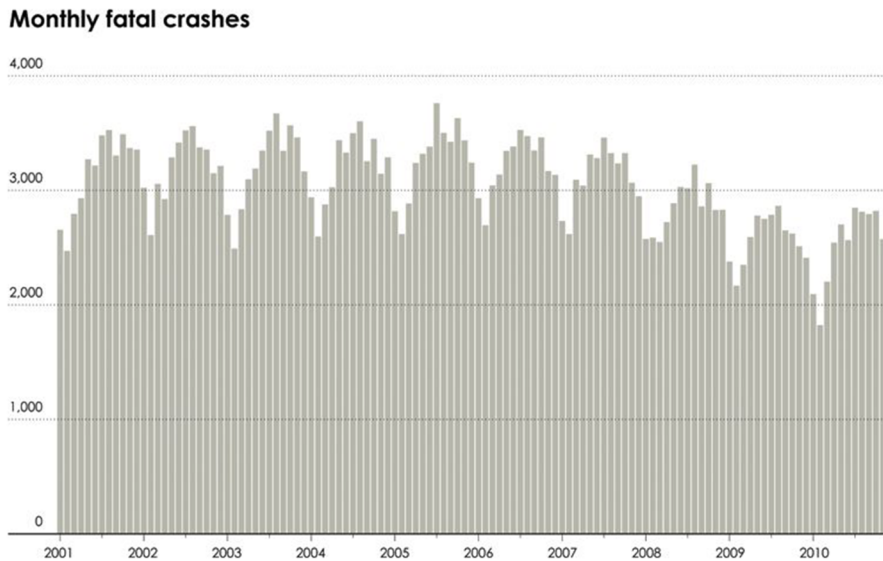


Figura 8. Accidents mortals per mes als Estats Units.
Font: *Data Points: Visualization That Means Something*.

Ara bé, val la pena reflexionar sobre aquest objectiu des d’una perspectiva més oberta. La figura 9 mostra una captura de pantalla presa del projecte «Flight Patterns» d’Aaron Koblin. L’autor disposava de la posició GPS de milers d’avions que sobrevolaven l’espai aeri nord-americà durant diverses hores, i ho va representar per mitjà de punts de color que deixen un petit traç de la seva trajectòria, fent servir colors més clars (és a dir, més blancs) quan l’avió s’acostava a terra.

Referència bibliogràfica

N. Yaw (2013). *Data Points: Visualization That Means Something*. Somerset: Wiley.

Aaron Koblin

Lloc web: <http://www.aaronkoblin.com/>



Figura 9. «Flight Patterns» d'Aaron Koblin. Captura del vídeo. http://www.aaronkoblin.com/work/flightpatterns/FPWeb_Final_3.mov

Gràcies a les codificacions visuals utilitzades en el projecte, aquesta representació revela un clar patró conegut per tothom: el mapa dels Estats Units. I no només això, sinó que també revela la posició dels aeroports dels Estats Units, que són els punts que sobresurten en totes les rutes que s'aprecien en el mapa.

És evident que no calen aquestes dades per descobrir el mapa. No obstant això, és important recordar que aquí no s'està representant un mapa, sinó les trajectòries d'un conjunt d'avions, les quals porten el mapa implícit. La genialitat d'aquest projecte és que **quan es representen les dades adequadament, es propicia el descobriment de patrons ocults.**

4. Regles per desenvolupar bones visualitzacions

Dissenyar una visualització implica decidir quines codificacions visuals s'utilitzaran per representar les dades. Les principals codificacions visuals que hi ha són: posició, forma, color i moviment¹. A continuació es detalla una sèrie de regles que cal tenir en compte per a la tria d'aquests elements. Com es podrà veure, moltes d'aquestes regles són molt senzilles d'aplicar, i generalment es podran implementar mitjançant qualsevol eina que pugui treballar amb dades.

⁽¹⁾Per obtenir-ne més detalls: «Pre-attentive Visual Properties and How to Use Them in Information Visualization».

4.1. Començar amb preguntes

Una visualització pot tenir moltes formes diferents. Per exemple, un conjunt de dades que contingui informació sobre accidents a Espanya pot tenir la localització de l'accident, el nombre d'ocupants del vehicle, el nombre de víctimes o ferits o el tipus de vehicle. És clar, doncs, que en funció del que es vulgui saber (en funció de l'«objectiu informacional») es podran representar aquestes dades en un mapa, en un gràfic de barres que acumuli el nombre d'accidents per nombre d'ocupants, o bé el nombre de ferits o víctimes.

Per tant, tenir preguntes concretes que es vulguin fer a les dades és de vital importància per crear una representació visual que ajudi a respondre-les correctament.

4.2. Gràfics autoexplicatius

En general, és desitjable que l'usuari sigui capaç d'entendre una visualització sense llegir el text que l'acompanya (com en el cas d'una notícia periodística) o les instruccions que presenta. Per a això es disposa d'una eina elemental que millora molt la llegibilitat de les visualitzacions: els **textos** o **etiquetes**.

El gràfic següent, de la figura 10, mostra dues línies temporals: la verda representa el consum d'aigua a la ciutat canadenca d'Edmonton un dia laborable qualsevol, mentre que la blava mostra el consum d'aigua el dia de la final d'hoquei sobre gel dels Jocs Olímpics d'hivern, en la qual participava la selecció nacional del Canadà. Veient el gràfic, i gràcies a les etiquetes, és molt fàcil poder veure com el consum d'aigua està íntimament relacionat amb el curs del partit, ja que en cada final de període (un partit d'hoquei consta de tres períodes) hi ha un repunt en el consum d'aigua (presumiblement perquè els espectadors fan una «parada tècnica» al lavabo, mentre que no la fan durant el partit).

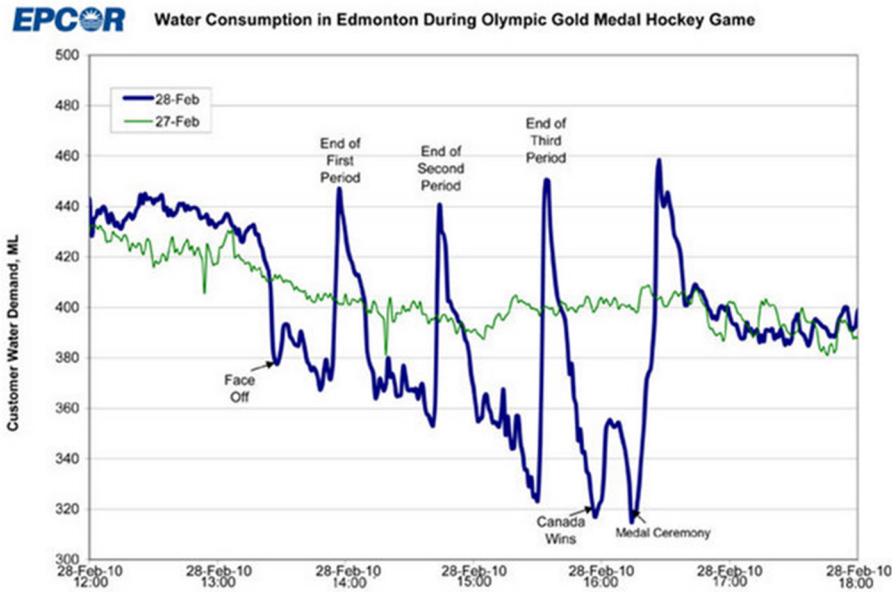


Figura 10. Visualització del consum d'aigua a Edmonton feta per l'empresa EPCOR.
 Font: <http://www.zdnet.com/article/infographic-water-consumption-in-edmonton-canada-during-olympic-gold-medal-hockey-game/>

Gràcies a les etiquetes i al títol de la representació és molt fàcil entendre la història que s'amaga darrere d'aquestes dades. I és que no només s'hi pot veure que, a mesura que avançava el partit, més gent aprofitava (presumiblement) per anar al lavabo, sinó que, a més, hi va haver un període extra de desempat just després del final del tercer període. Finalment, quan el Canadà va guanyar, alguns espectadors van aprofitar per tornar al lavabo just abans de poder veure el lliurament de medalles. Sense les etiquetes, caldria conèixer el context per relacionar el consum amb el partit.

4.3. La funcionalitat ha de prevaler enfront de l'estètica

No hi ha cap dubte que crear visualitzacions boniques té un poder de persuasió molt important. El «Better Life Index» de l'OCDE, mostrat en la figura 11, és un exemple de visualització molt atractiva que representa gràficament l'índex per comparar països creat per aquesta institució. L'ús d'una metàfora (la flor) per a la representació de cada país resulta molt interessant des del punt de vista estètic, tot i que pot presentar problemes pel que fa a la informació que es representa gràficament. Es tracta, de fet, d'una visualització interactiva molt complexa que combina diferents indicadors (un per pètal) i un indicador ponderat (la posició de la flor).

OCDE

Lloc web: <http://www.oecdbetterlifeindex.org/>

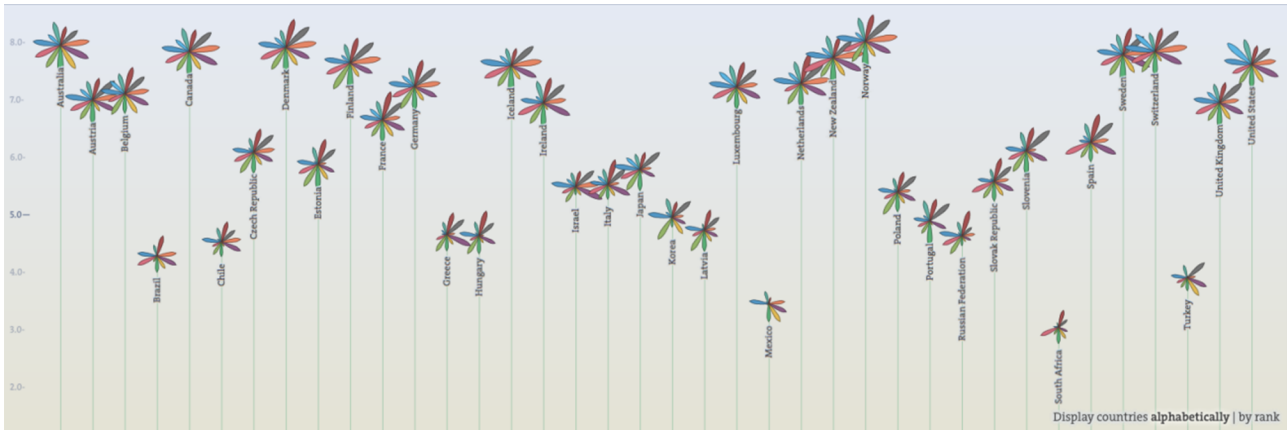


Figura 11. Flors representant barres en el «Better Life Index» de l'OCDE.
 Font: <http://www.oecdbetterlifeindex.org>

Tanmateix, és de vital importància entendre que l'objectiu principal de la visualització de dades, en un entorn analític, ha de ser generar coneixement sobre les dades amb les quals es treballa. Observeu, per exemple, la representació de la figura 12, que mostra els moments més actius a Twitter durant l'any 2011. Quin ha estat el tercer moment més actiu?

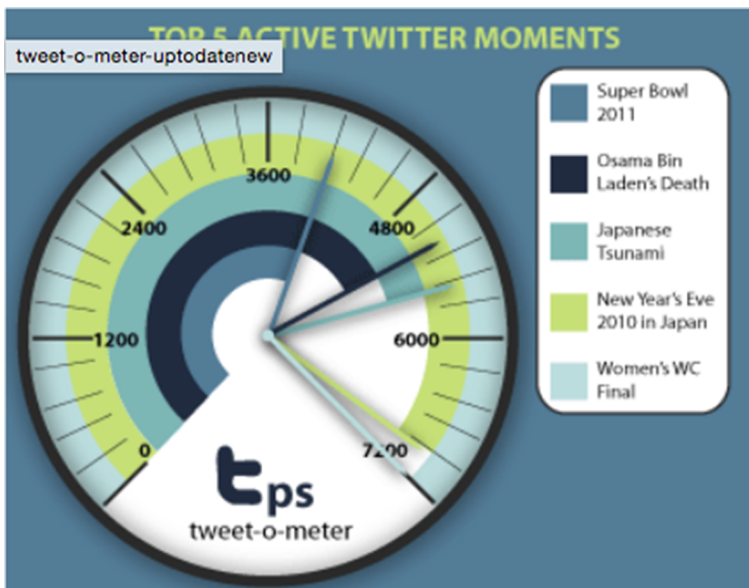


Figura 12. Els cinc moments més actius a Twitter el 2011 representats en una metàfora semblant a un comptakilòmetres.
 Font: <https://www.smashingmagazine.com/2011/10/the-dos-and-donts-of-infographic-design/>

Probablement, per contestar la pregunta s'han necessitat diversos segons, i fins i tot interpretar el gràfic ha pogut ser una mica confús. Això és perquè la representació utilitzada no és la més adequada per a l'«objectiu informacional» que es pretén. Si s'intenta fer el mateix exercici amb el gràfic de la figura 13, es pot observar que resulta molt més senzill.

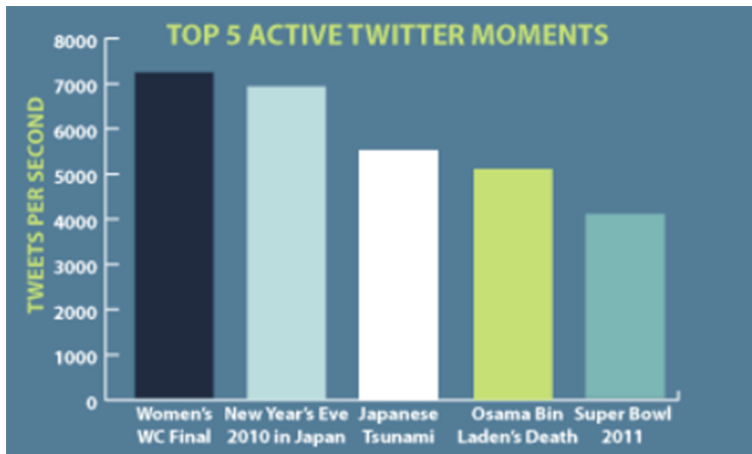


Figura 13. Representació alternativa al comptaquilòmetres que transmet millor la informació.

És important adonar-se que, en totes dues visualitzacions, el valor de piulades per segon s'ha codificat amb una barra. Mentre que en la figura 12 el valor que es representa s'ha codificat mitjançant un angle (com més piulades, més llarga és la barra circular o segment d'arc), en la figura 13 s'ha codificat amb la distància de la barra. D'aquesta manera, la longitud de les barres de la figura 12 no és proporcional als valors que representen², la qual cosa genera confusió. L'ús del color de fons també és un aspecte que s'ha de tenir en compte en aquest cas en concret.

⁽²⁾ Cal recordar que la longitud de la circumferència és $2 \times \pi \times r$. D'aquesta manera, les barres circulars que estiguin a l'exterior tindran una longitud més gran que les interiors donat un mateix angle, d'aquí la falta de proporcionalitat entre els diferents segments d'arc.

4.4. Ús de la interacció

Anteriorment s'ha comentat l'existència de dos tipus de visualitzacions: les estàtiques i les interactives. És molt important ser conscients que l'ús d'interacció pot aportar molts beneficis a una visualització. Una de les limitacions que pateixen els gràfics estàtics és que «només» poden comunicar un conjunt de dades concret d'una manera única, mentre que en els gràfics interactius es pot permetre a l'usuari que decideixi en quines dades fixar-se. A més, hi ha vegades en què la decisió que es pren sobre la forma que tindran les dades implica no ser tan precisos com es voldria en algun aspecte concret. Observeu, per exemple, la visualització de la figura 14, ja utilitzada anteriorment en l'apartat 3.3, però ara plasmada com a visualització estàtica:

Everyone

Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children

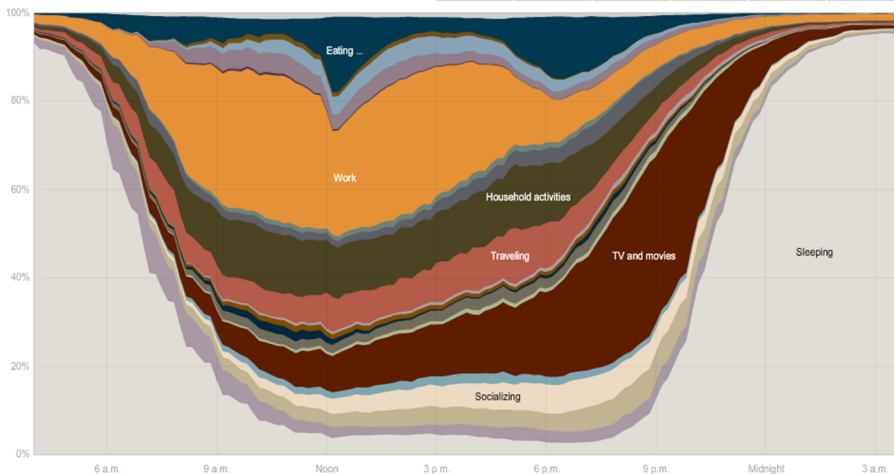


Figura 14. «How Different Groups Spend Their Time». Font: *The New YorkTimes*.

Aquesta representació és un gràfic d'àrees apilades en el qual cada àrea pintada representa un tipus d'activitat que fan els nord-americans durant el dia. L'eix de les x representa el temps, que comença a les quatre del matí, que és quan s'assumeix que comença el dia per a moltes persones i també és el de menor activitat. L'eix de les y és el que ajuda a identificar el percentatge de població que està fent cada activitat, representada per un color diferent. El gràfic és bastant clar i interessant, i gràcies a la interactivitat proporcionada pel menú superior es poden veure les dades relatives al segment de població en el qual s'està interessat, ja que permet canviar entre un sector o un altre fàcilment, i veure, així, els canvis en les diferents activitats.

Fins aquí aquesta visualització és molt bona; ara bé, la tria de la codificació de les dades fa que sigui molt difícil, per exemple, comparar si hi ha més gent que està treballant a les deu del matí que a les dotze del migdia. I és així perquè les àrees de les diferents activitats estan apilades i **no estan referenciades en el zero**. No obstant això, els dissenyadors d'aquesta visualització van crear una altra interacció per solucionar aquest problema: amb un clic a qualsevol àrea, aquesta queda referenciada en el zero i la comparació que es volia fer passa a ser molt més senzilla d'executar, com mostra la figura 15.

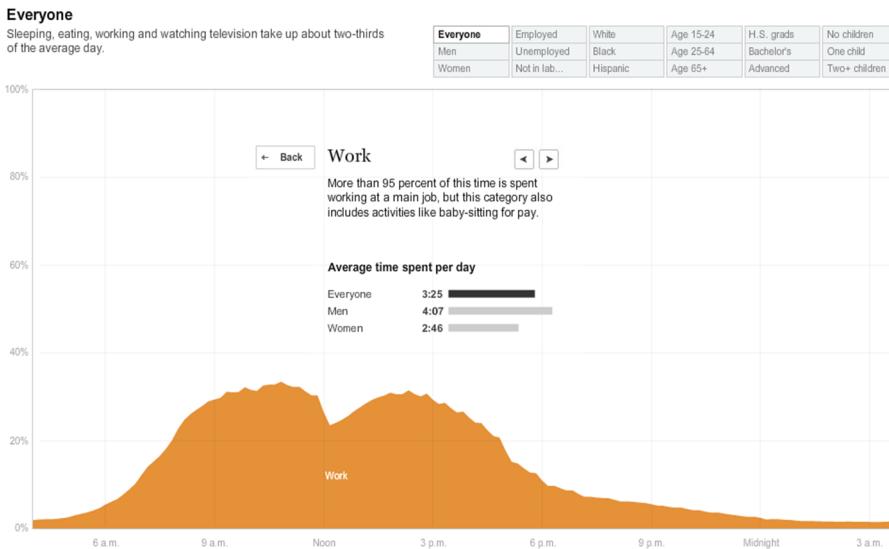


Figura 15. Selecció d'un tipus d'activitat que permet fer un altre tipus d'anàlisi visual mitjançant l'ús de la interacció.

Així doncs, es pot observar que la interacció pot ajudar a solucionar problemes inherents a la codificació visual triada i, així, proporcionar més i millor informació des de diferents perspectives.

4.5. La forma segueix la necessitat

Com ja s'ha comentat amb anterioritat, és important tenir clar quin és l'objectiu de la visualització per poder triar, d'aquesta manera, la codificació visual que hi sigui més adequada. Alberto Cairo posa un bon exemple en el seu llibre *The Functional Art*, en el qual es mostra la visualització de la figura 16 sobre dades d'ocupació.

Alberto Cairo

Lloc web: <http://albertocairo.com/>

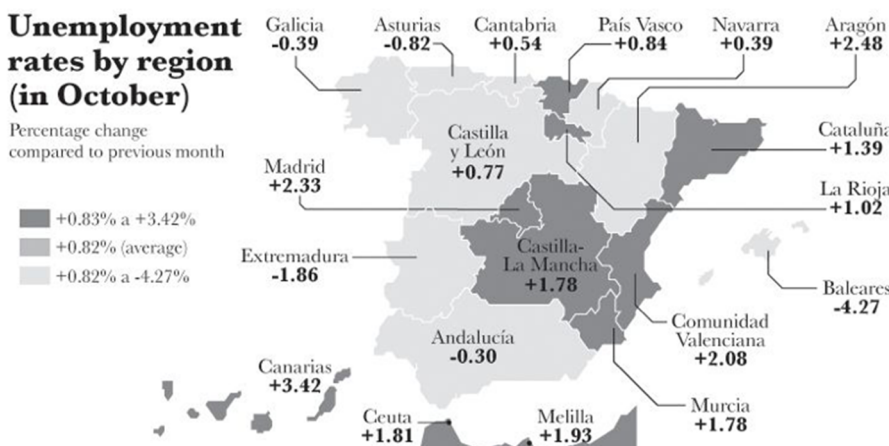


Figura 16. Representació de la taxa d'atur per comunitat autònoma. Font: *The Functional Art*.

Atès el percentatge de canvi en la taxa d'atur per comunitat autònoma respecte al mes anterior, sembla raonable pensar que l'objectiu principal d'una visualització d'aquestes dades hauria de permetre descobrir quines comunitats autònomes han millorat més i quines estan tenint més problemes en comparació amb el mes anterior. Tanmateix, la figura 16 fa que aquesta tasca sigui bastant 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 nombres, intentar memoritzar-los i després intentar inferir l'ordre dels valors. En definitiva, aquesta visualització ajuda més aviat poc a respondre la pregunta més bàsica que s'amaga dins d'aquestes dades.

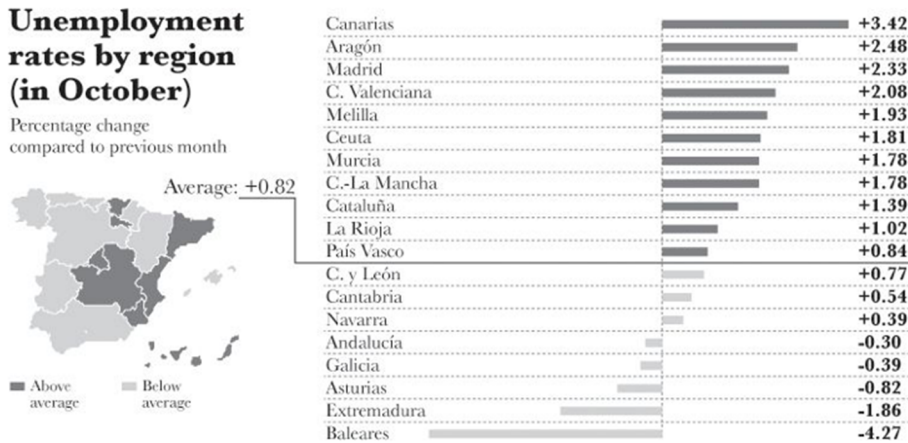


Figura 17. Redisseny de la figura 16 fet per Alberto Cairo.
Font: *The Functional Art*.

La figura 17 mostra el redisseny de la figura 16 fet per l'autor, en el qual proposa la utilització d'un gràfic de barres que, com es pot observar, és molt més clar i comunica millor les dades disponibles que no pas el mapa, que queda com a element que proporciona context geogràfic.

4.6. Preparar bé les dades

Per fer una bona visualització de dades és de vital importància disposar de bones dades. Això significa que s'haurà de treballar amb les dades abans de representar-les, amb la finalitat de poder mostrar el missatge que s'hi amaga darrere. Un exemple clar de preparació de dades és el procés de normalització. En la visualització següent es representa el nombre d'accidents mortals per comunitat autònoma l'any 2014. En la figura 18 es pot veure que Catalunya, Madrid i Andalusia són les tres comunitats amb un nombre d'accidents mortals més elevat. Ara bé, aquestes tres comunitats autònomes són les més poblades, per la qual cosa té sentit que també siguin les que tenen més trànsit i, per tant, més accidents. Aquesta informació és útil, però no permet, per exemple, fer una anàlisi que possibiliti a la Direcció General de Trànsit decidir en quina comunitat autònoma cal aplicar noves mesures per prevenir accidents.

Accidentes por comunidad autónoma en 2014

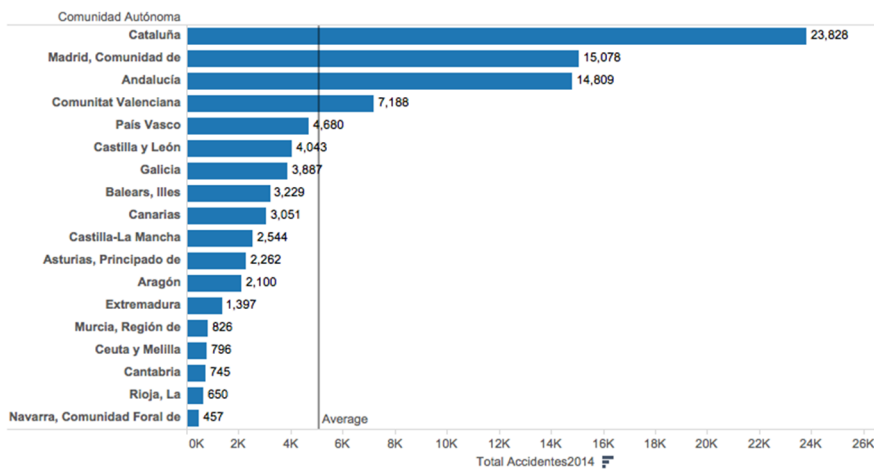


Figura 18. Accidents mortals per comunitat autònoma el 2014.
Font: Elaboració pròpia.

En el gràfic de la figura 19 s’ha aplicat una normalització, calculant el nombre d’accidents per habitant. D’aquesta manera, es pot veure que Ceuta i Melilla³ passen a estar en la primera posició de la classificació, i que, per exemple, les illes Balears se situen per sobre de Madrid.

⁽³⁾També cal tenir en compte que Ceuta i Melilla tenen molt pocs habitants, cosa que implica que la taxa d’accidents per habitant pot variar molt fàcilment.

Accidentes por habitante comunidad autónoma en 2014

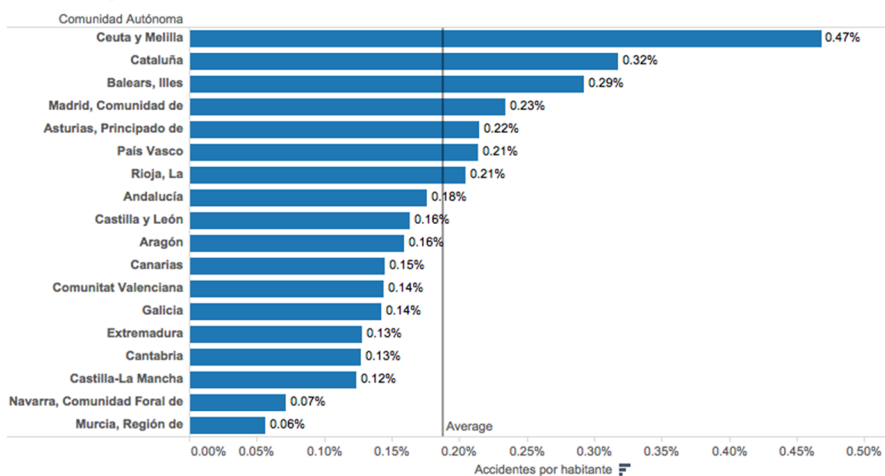


Figura 19. Accidents mortals per habitant per a cada comunitat autònoma el 2014.
Font: Elaboració pròpia.

Resum

A manera de sumari, per fer una bona visualització hi ha una sèrie de regles bàsiques que cal tenir en compte:

- És molt important tenir clar què es vol que expliqui una visualització, qui en serà el consumidor i en quin dispositiu es farà servir.
- Un conjunt de dades pot ser representat de moltes maneres diferents. La missió de l'analista o dissenyador és triar la visualització que ajudi millor a complir amb l'objectiu informacional de l'usuari final.
- Hi ha dos tipus fonamentals de visualitzacions: les estàtiques i les interactives. En general, les primeres se centren més a comunicar dades o fets mitjançant recursos visuals, mentre que les segones permeten una anàlisi de dades més exploratòria.
- Les visualitzacions de dades en un entorn analític tenen com a prioritat ajudar a generar coneixement entorn de les dades, i no la mera generació d'imatges atractives.

Per a aprendre a generar visualitzacions de dades no hi ha cap altra alternativa que practicar amb l'exemple. Com més visualitzacions es facin, més s'aconseguirà interioritzar totes les regles que, al final, determinen la diferència entre una bona i una mala visualització. I, per descomptat, estar al corrent de les noves tendències i eines per al seu desenvolupament. Es tracta, per tant, de seguir l'exemple descrit per Enrico Bertini en forma de cicle continu: aprèn, copia, critica, crea i exposa les teves visualitzacions.

Referència bibliogràfica

Enrico Bertini (2010). «How to Become a Data Visualization Expert: A Recipe».