



Periodisme de dades  
Carlos Alonso  
18/05/2012



Periodisme de dades

Introducció

Què és el periodisme de dades

Una mica d'Història

Una proposta de model i procés periodístic

Cas pràctic: Ràtios de mortalitat

Cas pràctic: Víctimes d'ETA

Treballant amb dades

Por on començar: fons de dades

Treballar, netejar i seleccionar dades

Què podem trobar visualitzant dades

Categories de visualitzacions: com elegir un gràfic?

I ara?

Què has de saber?

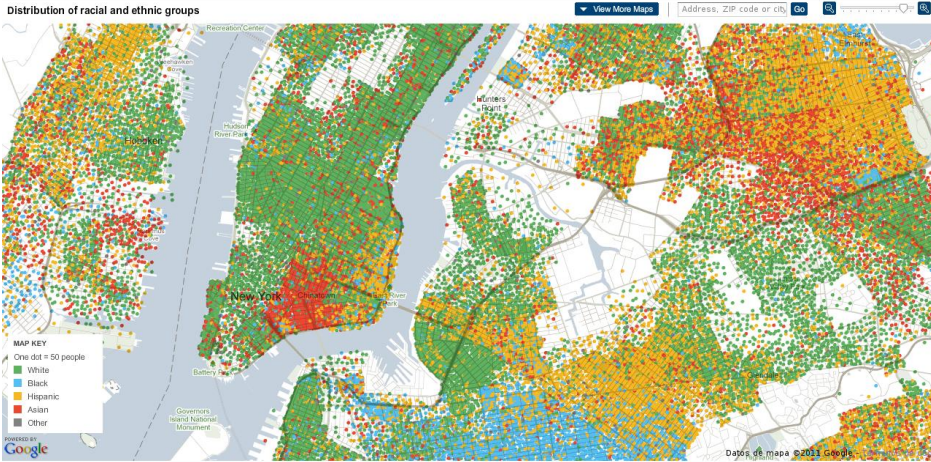
On continuar aprenent?

	A	B	C	D	E	G	I	K	L	BP	BS	BV	BY	CB	CG
1	Zip Code /Territory	Name	Land Area (Square Miles)	Population	Population per Square Mile	% of Population that is Rural	Number of Vehicles per Square Mile	Number of Vehicles per Occupied Housing Unit	Number of Vehicles per Occupied Housing Unit	% of Housing Units without a Mortgage	% of Sample Households with Social Security Income	% of Sample Households with Public Assistance Income	% of Sample Population with Income in 1999 below Poverty Level	% of Labor Force that is Unemployed	
31	48618	Coleman	93.0	5,256	56.5	100.0%	44.9	2.1	1.73	41.2%	29.2%	3.0%	9.8%	6.8%	
32	48620	Edenwille	0.6	36	60.0	100.0%	56.7	1.8	3.59	100.0%	41.2%	0.0%	16.7%	0.0%	
33	48628	Hope	42.4	2,011	47.4	100.0%	39.3	2.0	2.14	32.1%	33.1%	1.3%	5.3%	4.4%	
34	48640	Midland	103.6	33,085	319.4	36.1%	228.2	1.9	2.28	30.8%	23.7%	3.1%	8.3%	5.2%	
35	48657	Sanford	69.3	8,053	116.2	62.0%	91.5	2.0	0.79	2.04	32.7%	2.5%	7.7%	6.6%	
36	106	All	308.9	48,441	156.8	50.1%	116.1	1.9	0.74	2.21	32.1%	24.9%	2.9%	8.2%	5.6%
37															
38	48613	Bentley	40.4	1,096	27.1	100.0%	21.2	2.2	0.78	2.50	37.9%	32.5%	4.2%	10.6%	4.8%
39	48631	Kawlawlin	32.8	4,506	137.4	93.7%	106.2	2.1	0.77	1.88	42.6%	31.2%	3.0%	4.8%	5.1%
40	48634	Litwood	48.0	4,732	98.6	76.2%	73.1	2.1	0.74	1.94	39.6%	27.7%	2.9%	7.7%	7.5%
41	48650	Pinconning	96.9	7,589	78.3	100.0%	59.4	2.0	0.76	1.77	40.5%	26.6%	3.2%	9.7%	8.4%
42	48747	Munger	32.1	1,537	47.9	100.0%	37.3	2.0	0.78	1.83	41.0%	33.5%	2.1%	6.7%	5.8%
43	107	All	250.2	19,460	77.6	92.8%	59.1	2.0	0.76	1.84	40.7%	29.5%	3.0%	7.9%	7.0%
44															
45	48415	Birch Run	64.6	9,592	146.5	97.5%	111.2	2.0	0.75	1.68	34.8%	22.8%	2.1%	6.2%	6.4%
46	48417	Burt	25.2	3,195	126.8	100.0%	97.3	2.0	0.77	1.72	32.7%	19.0%	3.3%	9.5%	5.7%
47	48609	Saginaw	39.1	12,231	312.8	37.5%	249.6	2.0	0.79	1.69	29.6%	20.1%	0.8%	3.6%	5.2%
48	48614	Brant	36.4	1,341	36.8	100.0%	27.8	2.1	0.75	1.76	29.1%	30.5%	5.9%	12.1%	6.5%
49	48616	Chesaning	79.5	7,614	95.8	100.0%	74.7	2.0	0.78	1.63	38.4%	29.1%	1.9%	6.2%	5.9%
50	48623	Freeland	76.1	12,507	164.3	57.3%	114.4	2.1	0.69	1.84	30.8%	22.8%	1.9%	4.8%	3.9%
51	48626	Henlock	61.9	5,834	94.2	100.0%	77.7	2.1	0.77	1.86	32.5%	23.3%	1.2%	6.3%	4.8%
52	48637	Merrill	71.9	3,444	47.9	100.0%	77.7	2.1	0.78	1.79	37.0%	29.4%	2.4%	7.6%	4.1%
53	48649	Oakley	26.4	1,681	63.7	100.0%	77.7	2.1	0.78	2.21	41.6%	29.6%	1.9%	4.6%	6.8%
54	48655	Saint Charles	103.7	6,672	64.3	100.0%	77.7	2.1	0.77	1.75	34.9%	29.0%	3.4%	8.4%	6.6%
55	48734	Franklinnuth	49.7	7,377	148.3	37.6%	113.3	1.8	0.75	1.57	38.2%	34.9%	0.7%	4.7%	3.3%
56	108	All	625.5	71,488	114.3	75.1%	114.3	2.1	0.75	1.72	33.7%	27.0%	1.8%	5.9%	5.2%

Mapping America: Every City, Every Block

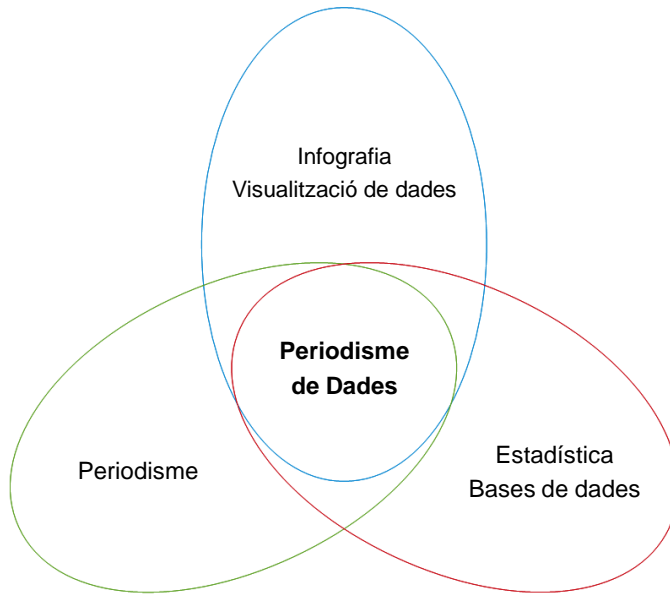
Find something interesting? Share this view on [Twitter](#) or [Facebook](#)

Browse local data from the Census Bureau's American Community Survey, based on samples from 2005 to 2009. Because these figures are based on samples, they are subject to a margin of error, particularly in places with a low population, and are best regarded as estimates. [View Readers Maps \(40\)](#)

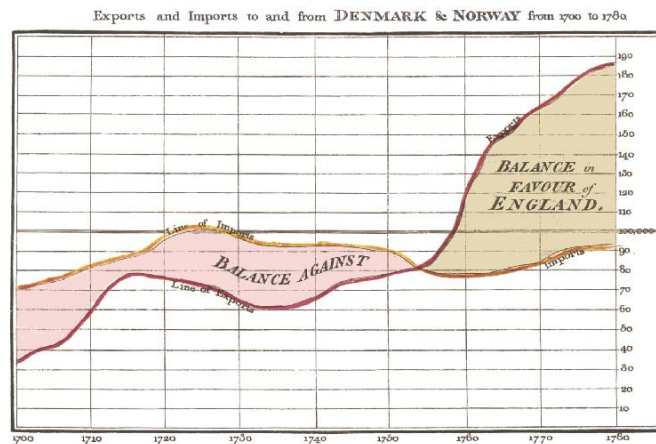


By MATTHEW BLOCH, SHAN CARTER and ALAN McLEAN | Source: 2005-9 American Community Survey, Census Bureau, socialexplorer.com

Note: Dots are evenly distributed across each Census tract or county. Dollar amounts are adjusted for inflation.

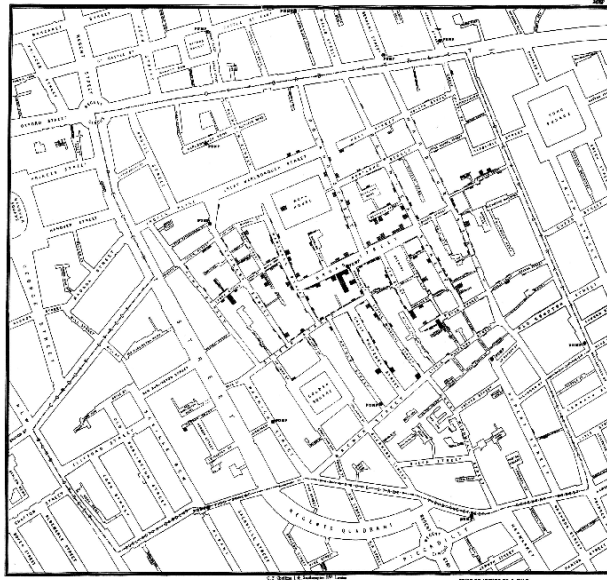


William Playfair - 1786

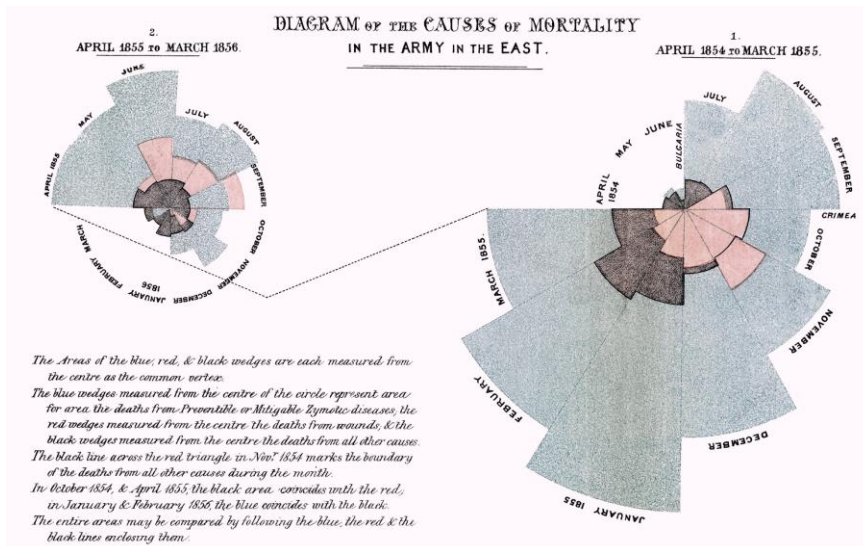


*The Bottom line is divided into Years, the Right hand line into 10000 each.  
Published at the Age of Reason, 1787. Playfair.  
Scale weight 100, Broad, London.*

John Snow - 1854



Florence Nightingale - 1858





## CBS i l'UNIVAC 1 - 1952



## Phillip Meyer - 1967

## The Non-Rioters: A Hopeful Majority

BY PHILIP MEYER

Despite the recent weeks of rioting, Detroit's Negro community is basically calm and committed to traditional American values.

The rioters were a small and desiant minority—protesting not only against society in general, but against prevailing Negro social standards.

These are the basic findings from a survey research project in the central riot areas of east and west Detroit. The study, based on interviews with a representative sample of 417 Negroes, was sponsored by the Urban League with assistance from the Free Press. Two University of Michigan faculty members acted as consultants.

The findings bear out, to a surprising degree, many of the views of Detroit race relations that were widely held before the riots: Detroit Negroes agree that they are as well off as or in better position than Negroes in other northern cities.

They have many problems but these are not, for the most part, overwhelming. And in seeking solutions, most Detroit Negroes look to the established social structure rather than the revolutionary ideology of black nationalism.

THE PEOPLE  
BEYOND 12TH STREET



Then why the riot?

In the first place, it does not take many clankers to create a riot. Historians have estimated that only a third of the colonists supported the American Revolution when it began.

Persons in the Detroit area were asked, with a guarantee that they would remain anonymous, to assess their own extent of participation in the riot. Twelve percent of the total sample (16 percent of those who replied to the question) admitted some degree of active participation.

Rioters stand in sharp contrast to majority of Detroit Negroes. Page 1B.

Police brutality cited as the No. 1 cause of Negro discontent. Page 4B.

Many rioters were settling old grievances. Page 4B.

This is consistent with other estimates that most riots have the active support of only 10 to 15 percent of the Negro community.

It is possible that the true rate of riot involvement in Detroit was somewhat higher, due to a reluctance on the part of some to admit participation or to talk to interviewers.

However, the interviewers—all Negroes—reported that most of their subjects were cooperative. For purposes of comparison, the study took all self-descriptions of riot activity at face value.

There was not a basic or lasting breakdown of respect for law and order. The vast majority of Negroes interviewed think of looting, burning and shooting as crime. They favor fines or jail for looters and jail for the more serious offenses.

Only a few Detroit Negroes see themselves as worse off in terms of education, income and jobs than Negroes in other northern cities.

Of those who see any difference at all, far more see Detroit Negroes as better off—25 percent of those surveyed think Detroit Negroes have more education compared to 10 percent who think they have less.

Forty-six percent think Detroit Negroes have more income; only eight percent think they have less. Thirty-eight percent think Detroit Negroes have worse jobs.

There are grievances, of course. And these grievances are related to rioting. But the pattern suggests that they did not constitute an explosive potential that demanded release. The riot could well have been an ill-fated accident.

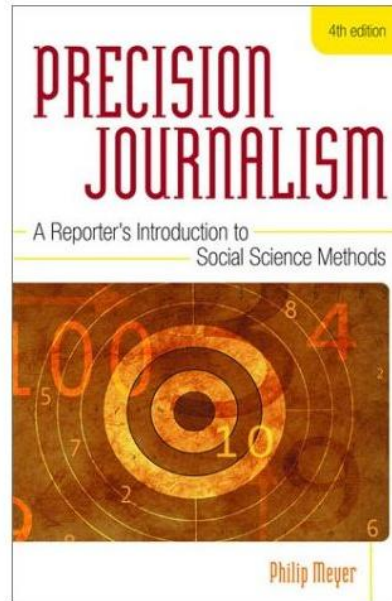
These points are explored more fully in the article beginning on Page 1B of today's Free Press.





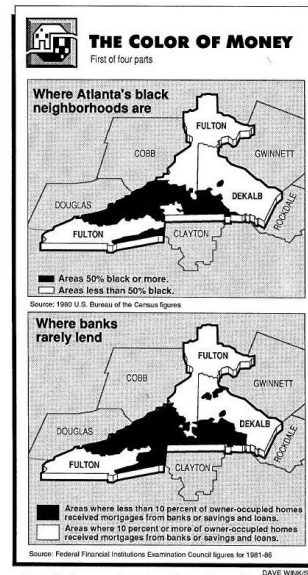
Una mica d'Història

Phillip Meyer - 1973

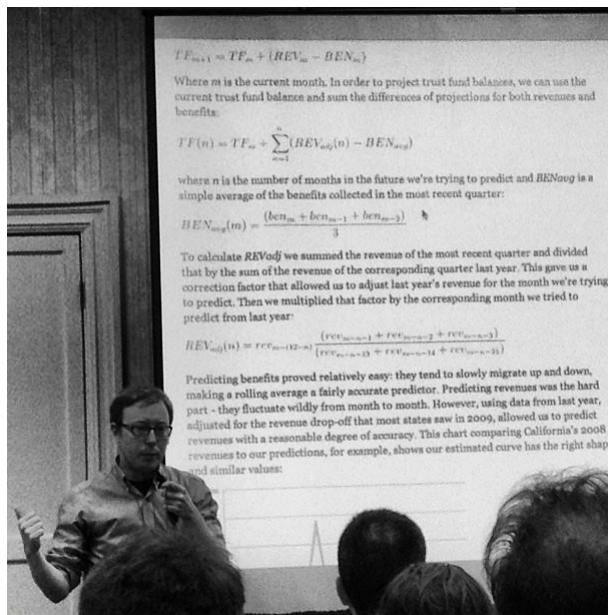


Una mica d'Història

Bill Dedman - 1989



NICAR - 1989



Introducció

Què és el periodisme de dades

Una mica d'Història

Una proposta de model i procés periodístic

Cas pràctic: Ràtios de mortalitat

Cas pràctic: Víctimes d'ETA

Treballant amb dades

Por on començar: fons de dades

Treballar, netejar i seleccionar dades

Què podem trobar visualitzant dades

Categories de visualitzacions: com elegir un gràfic?

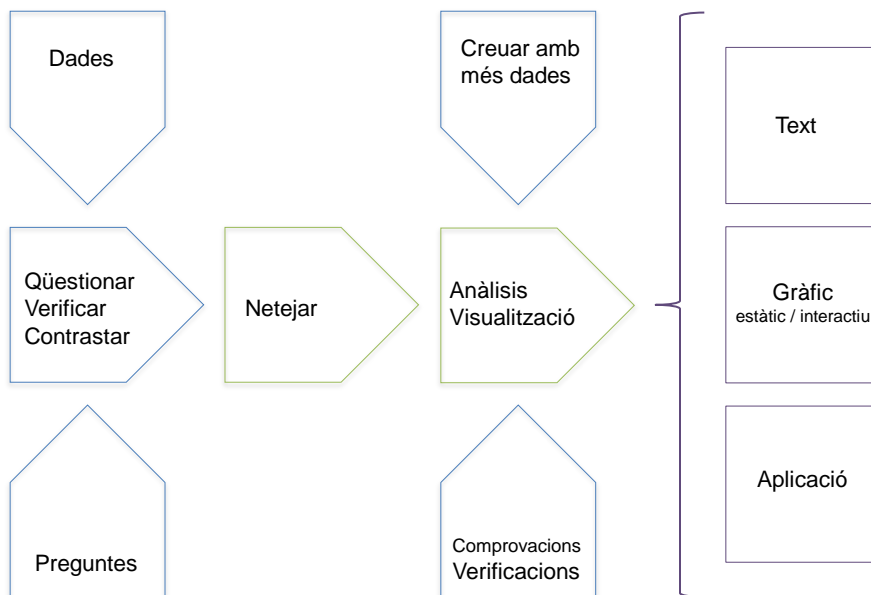
I ara?

Què has de saber?

On continuar aprenent?



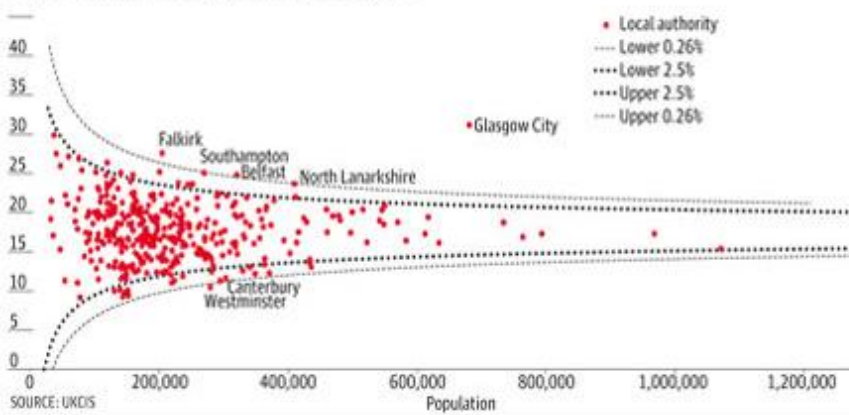
## Una proposta de model i procés periodístic



## Una proposta de model i procés periodístic

### Bowel cancer mortality

By UK local authority, deaths per 100,000



<a href="#">Ministerio del Interior</a>	829
<a href="#">Fundación Víctimas del Terrorismo</a>	828
<a href="#">Asociación Víctimas del Terrorismo</a>	858
<a href="#">Colectivo Víctimas del Terrorismo en el País Vasco</a>	952
<a href="#">El País</a>	829
<a href="#">El Mundo</a>	864
<a href="#">El Correo</a>	858
<a href="#">Diario Vasco</a>	829
<a href="#">ABC</a>	857
<a href="#">La Voz de Galicia</a>	829
<a href="#">El Periódico de Catalunya</a>	829
<a href="#">La Vanguardia (1 / 2)</a>	829 / 858
<a href="#">Wikipedia (ES)</a>	839
<a href="#">Wikipedia (EN)</a>	829

TERRORISMO

## Las cifras del terrorismo de ETA

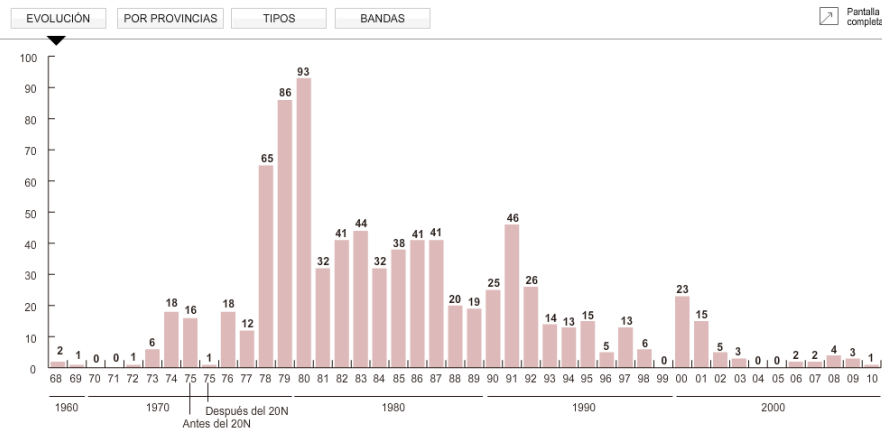
01/08/2011 | Factoría de Multimedia

Evolución de los asesinados por la banda terrorista ETA, distribución por provincias, tipos de atentado y comparativa con las cifras de otras bandas terroristas en España.

[deja tu comentario](#) [ 41 ]

[f Share](#) 54

2







Introducció

Què és el periodisme de dades

Una mica d'Història

Una proposta de model i procés periodístic

Cas pràctic: Ràtios de mortalitat

Cas pràctic: Víctimes d'ETA

Treballant amb dades

Por on començar: fons de dades

Treballar, netejar i seleccionar dades

Què podem trobar visualitzant dades

Categories de visualitzacions: com elegir un gràfic?

I ara?

Què has de saber?

On continuar aprenent?



## Fons de dades

1. Repositoris públics
2. Repositoris comercials
3. Cerca avançada
4. Scraping & APIs
5. Petició directa
6. Creació pròpia



## Fons de dades: repositoris públics

### Institucions públiques:

#opendata

Govern d'Espanya  
 Generalitat de Catalunya  
 Ajuntament de Barcelona  
 (n'hi ha més!)

Departaments d'Estadística

Generalitat de Catalunya - IDESCAT

Govern d'Espanya – INE

Unió Europea – Eurostat

Departaments Cartogràfics

Ministeris, Òrgans Reguladors, altres empreses i institucions

### Iniciatives privades:

CKAN

Publicdata.eu

Open Knowledge Foundation



## Fons de dades: repositoris comercials

Infochimps

Timetric

Google Public Data Explorer

Amazon Web Services Public Data



## Fons de dades: cerca avançada

Paràmetres de cerca avançada en Google

`filetype:xls site:domain.tld`

Motors de cerca especialitzats





## Fons de dades: scraping & APIs

- Scraperwiki
- OutWit
- Scripts (PHP, Python...)
- Yahoo! Pipes
- Google Spreadsheets + Fórmules



## Fons de dades: demanar accés

L'Avantprojecte de Llei de Transparència, Accés a la Informació Pública y Bon Govern és clarament deficient:

- Dret limitat
- Excepcions
- Silenci administratiu negatiu
- No hi ha sancions
- No hi ha òrgan independent

- AccessInfo
- Tu Derecho a Saber
- Civio
- Pro Bono Público
- Abredatos



## Fons de dades: creació pròpia

Bases de dades de creació pròpia

Col·laboració amb l'audiència – Crowdsourcing

Experiències:

Adoptaunsenador

El Periódico

BTV



## Treballar amb dades: netejar i processar

Netejar

Cel·les i columnes buides o innecessàries

Ordenar

Separar o unificar: cel·les amb diversos valors

Normalitzar

St. Andreu != Sant Andreu != S. Andreu

Caràcters estranys

Ajustar per inflació

Formatejar

Text, número, dates

Format d'arxiu: CSV Vs XLS, Shape Vs KML



## Treballar amb dades: netejar i processar

### Verificar

- Comparar amb altres sets de dades
- Definició de les dades
- Qualitat de les dades
- Grandària de la mostra
- Esbiaix de les dades
- Llei de Benford

### Calcular

- Canvi en dades absolutes Vs Relatives
- Càlculs estadístics bàsics: mitjana, mediana, moda
- Valors atípics
- Rellevància estadística
- Agrupar
- Filtrar

### Eines:

- Fulla de càlcul
- R
- Google Refine
- Data Wrangler



## Què podem trobar visualitzant dades

Tendències

Patrons

Anomalies

Connexions

Correlacions

Comparacions

Relacions

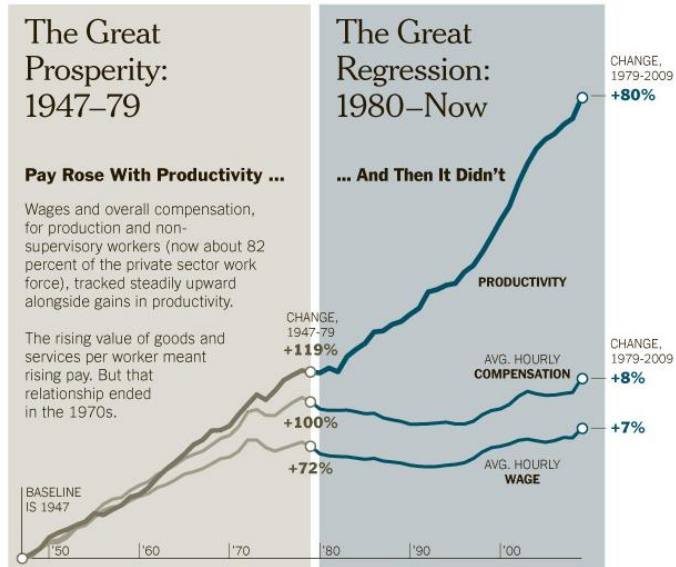
Jerarquia

Localitzacions

Processos

## Què podem trobar visualitzant dades

### Tendències



## Què podem trobar visualitzant dades

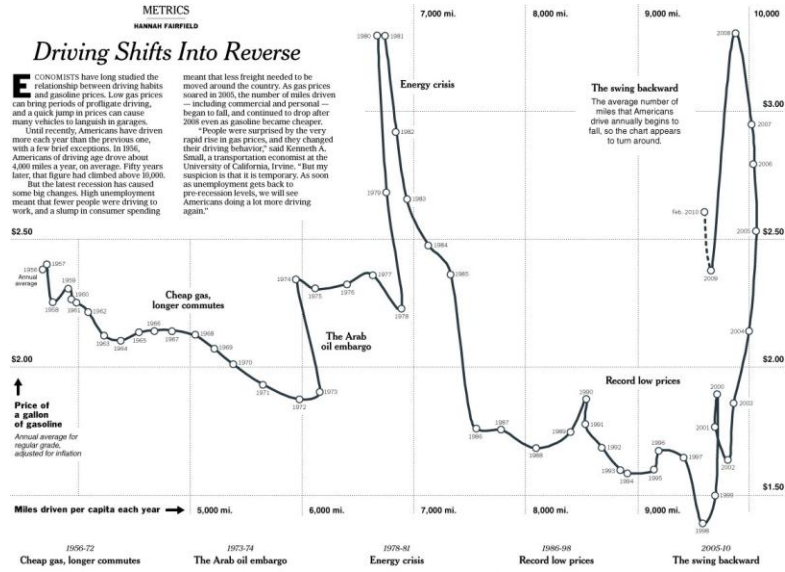
### Patrons





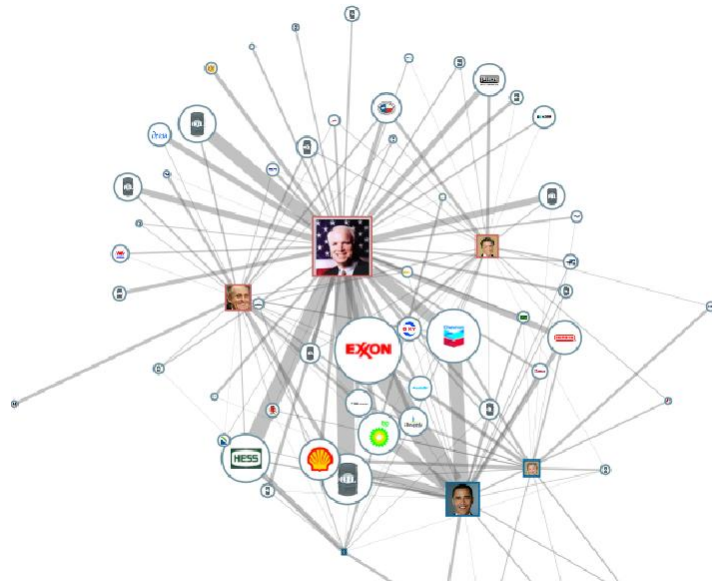
# Què podem trobar visualitzant dades

## Anomalies



# Què podem trobar visualitzant dades

## Connexions

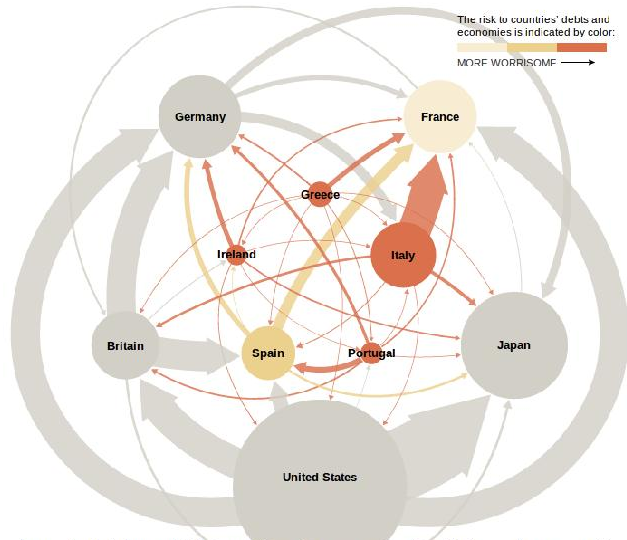






# Què podem trobar visualitzant dades

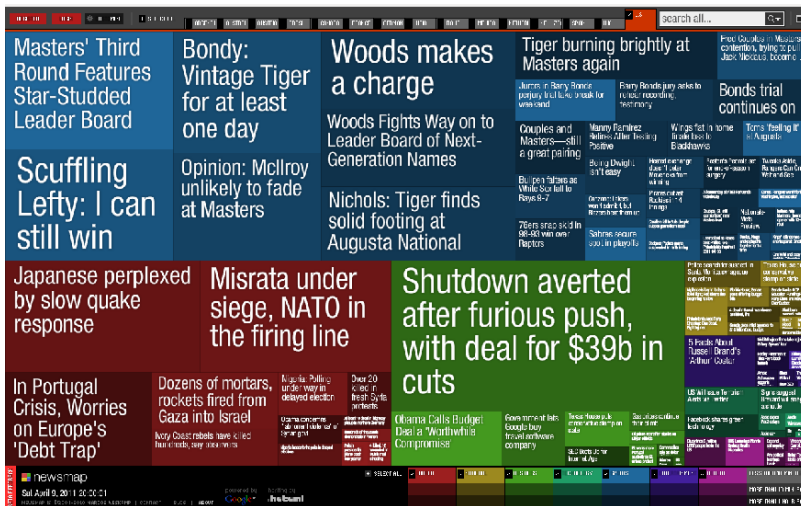
## Relacions



Arrows show imbalances of debt exposure between borrowers in one country and banks in another; arrows point from debtors to their bank creditors. Arrow width's are proportional to the balance of money owed. For example, French borrowers owe Italian banks \$50.5 billion, Italian borrowers owe French banks \$116.4 billion. The difference — their imbalance — shows France's banking system more exposed to Italian debtors by about \$365.8 billion.

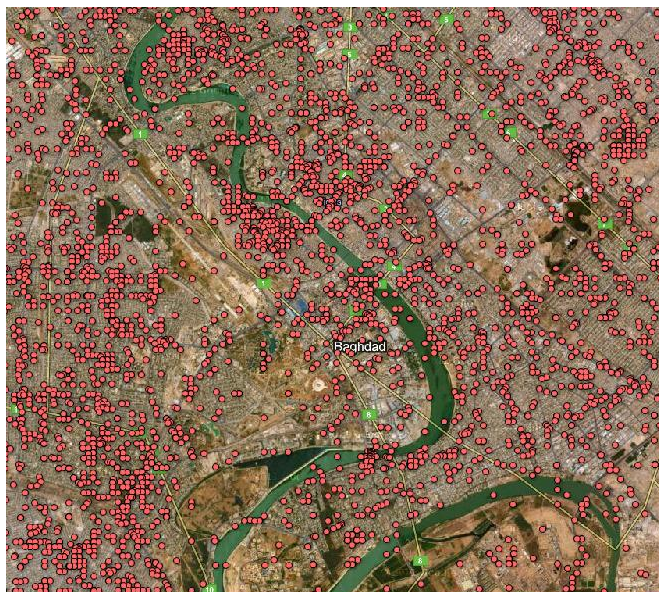
# Què podem trobar visualitzant dades

## Jerarquia



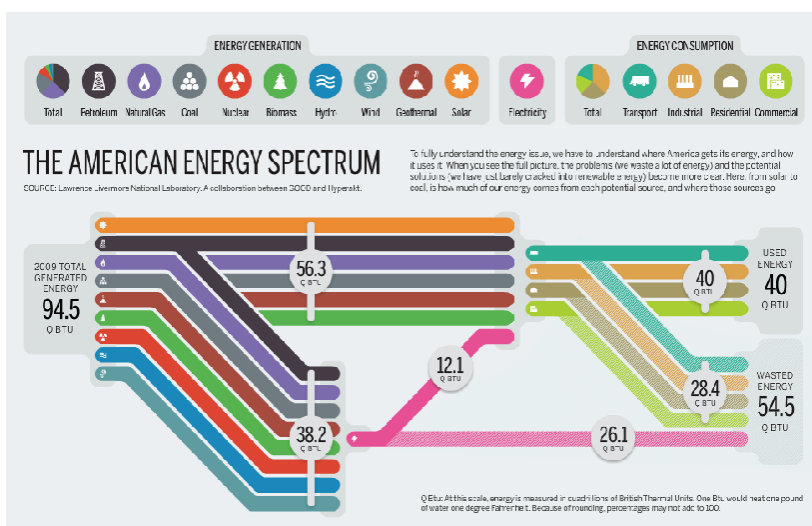
# Què podem trobar visualitzant dades

## Localitzacions



# Què podem trobar visualitzant dades

## Processos





## Com elegir un gràfic?

Quin tipus de dades tenim?  
 Quines variables volem mostrar?  
 Quantes variables volem mostrar?  
 Tenim informació temporal?  
 Tenim informació geogràfica?

Què volem representar?  
 Comparacions generals / Específiques  
 Quin tipus de relació?  
 Relació / Comparació / Distribució / Composició

Possibles distorsions:  
 Escala  
 3D  
 Àrea  
 Agrupació de dades



## I ara? Què has de saber?

### Periodista

Estadística  
 Fulls de càlcul  
 Bases de dades

### Periodista de dades

R  
 Google Refine  
 Fusion Tables  
 CartoDB  
 Impure / Quadrigram  
 Tableau

### Visualitzador de dades

HTML5 + \*CSS  
 Javascript  
 Python  
 SQL  
 Processing  
 Dibuix Vectorial  
 Disseny Gràfic  
 Psicologia Cognitiva



## I ara? On continuar aprenent?

Data Driven Journalism

<http://datadrivenjournalism.net>

Grup de Periodisme de Dades del Medialab-Prado

<http://medialab-prado.es>

NICAR

<http://www.ire.org/nicar/>

Malofiej

<http://www.snd-e.com/es/malofiej>

Alberto Cairo

<https://twitter.com/#!/albertocairo>

Malaprensa

<http://www.malaprensa.com>

Mar Cabra

<http://twitter.com/#!/cabralens>

Visualising Data

<http://www.visualisingdata.com>

Simon Rogers

<http://www.guardian.co.uk/profile/simonrogers>



## Gràcies

[carlos.alonso.jimenez@gmail.com](mailto:carlos.alonso.jimenez@gmail.com)

[@carlosalonso](#)

[delicious.com/carlosalonso](http://delicious.com/carlosalonso)

Gràcies!