Visual Mappings

James EAGAN

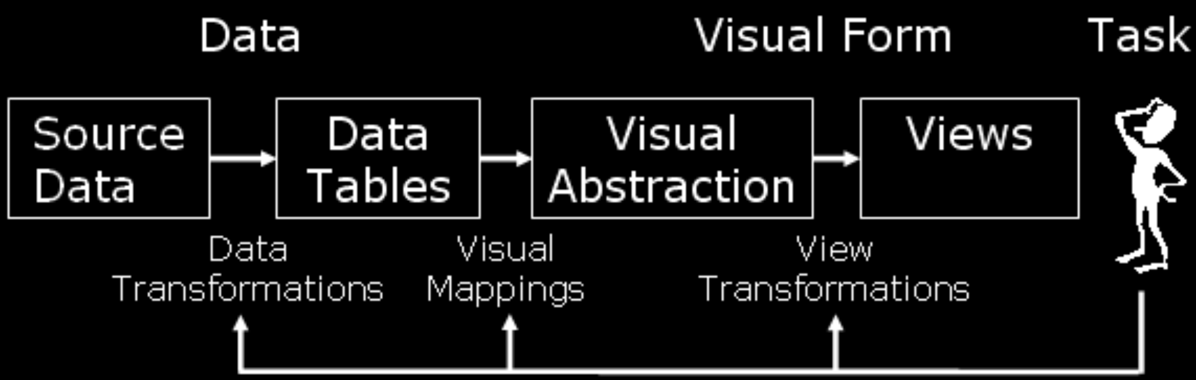
TELECOM Paris

Includes slides adapted from John Stasko (Georgia Tech), Petra Isenberg & Jean-Daniel Fekete (INRIA), Nadia Boukhelifa (INRA), Chris North (Virginia Tech), Tamara Munzner (UBC)

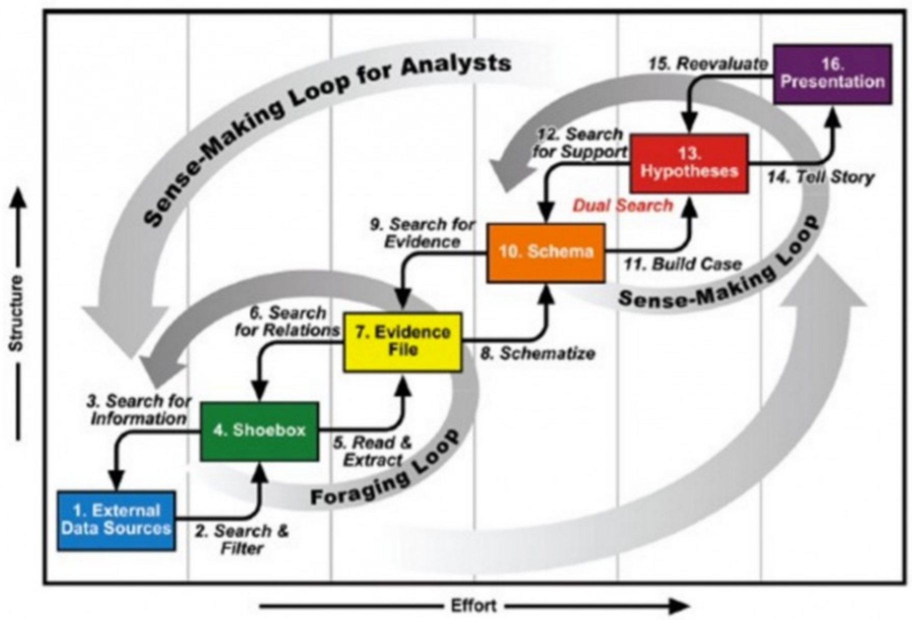
Updated: April 2020

How do we show the data?

InfoVis Pipeline



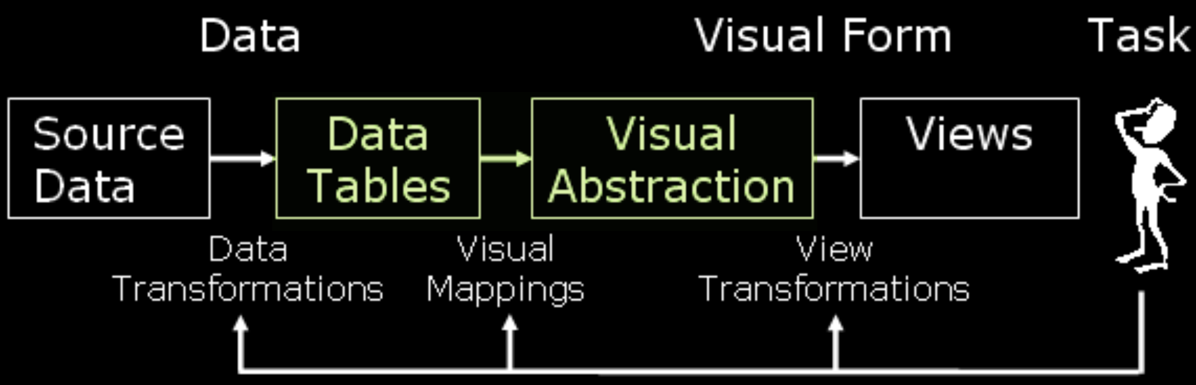
Pirolli-Card Sensemaking Loop



Model: Pirolli, Peter, and Stuart Card. "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis." *Proceedings of International Conference on Intelligence Analysis.* Vol. 5. McLean, VA: Mitre, 2005.

Image source: Thomas, James J. and Kristin A. Cook "Illuminating The Path" (2005): pp. 44

InfoVis Pipeline



Map *data* to a *representation*

- Data is abstract
- Representation is more conceptual
- Use a space
- Create an *implantation* of *data* into the space

Visual Structures

- Composed of
  - Spatial substrate
  - Marks
  - Graphical properties of marks

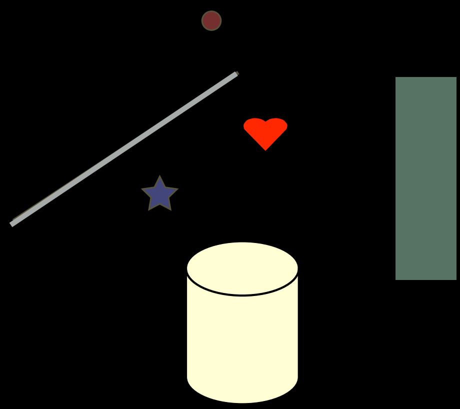
[Bertin, Sémiologie Graphique 1967]

Space

- Visually dominant
- Often put axes on space to assist
- Use techniques of
  composition, alignment, folding,
  recursion, overloading to
  1. increase use of space
  2. do data encodings

Marks

- Things that occur in space
  - Points
  - Lines
  - Areas
  - Volumes



Graphical Properties

also called **Visual Variables**

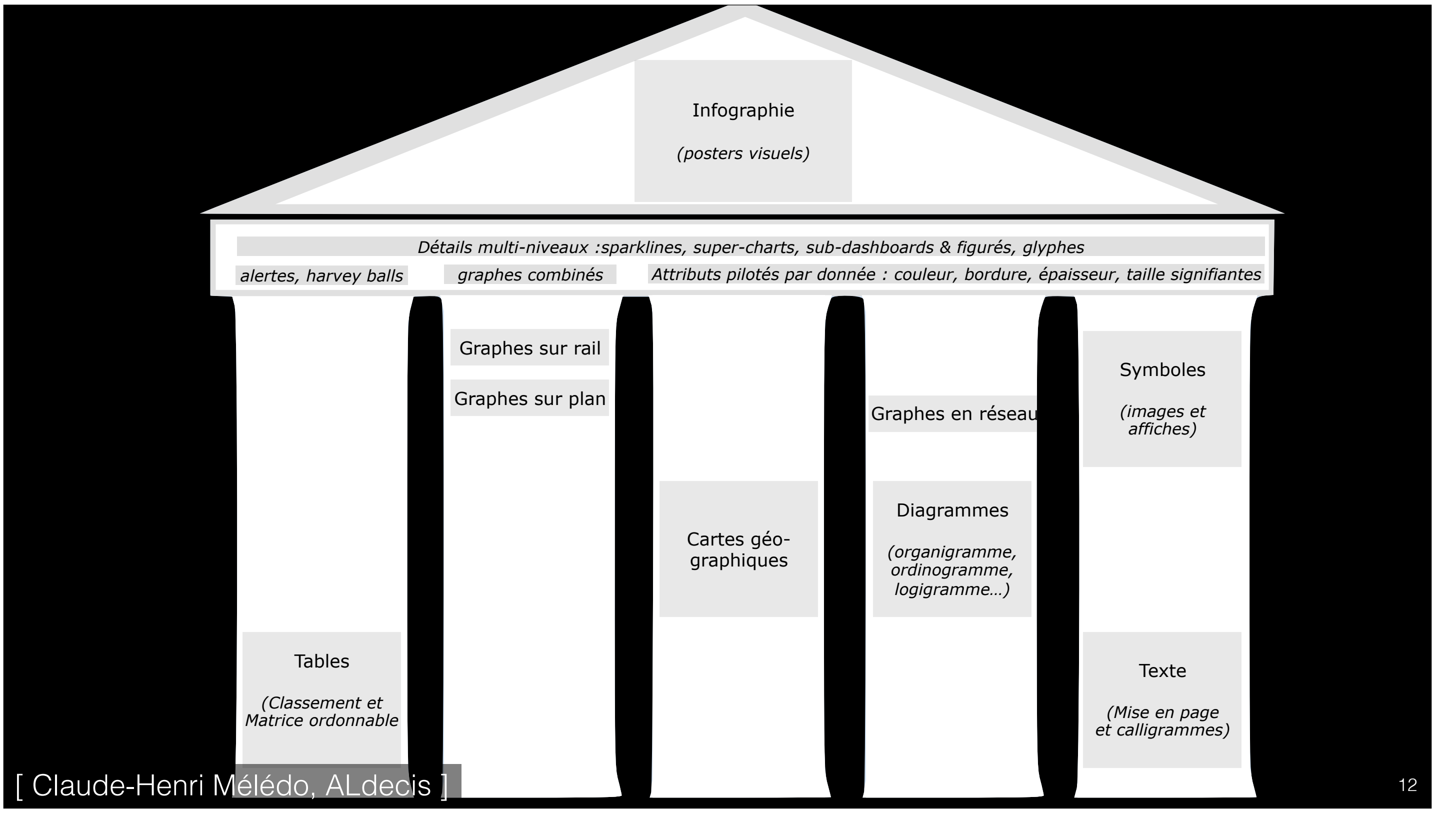
	Spatial properties	Object properties
Expressing extent	position size	greyscale
Differentiating marks	orientation	color shape texture

There are only 5 graphics

(sort of…)

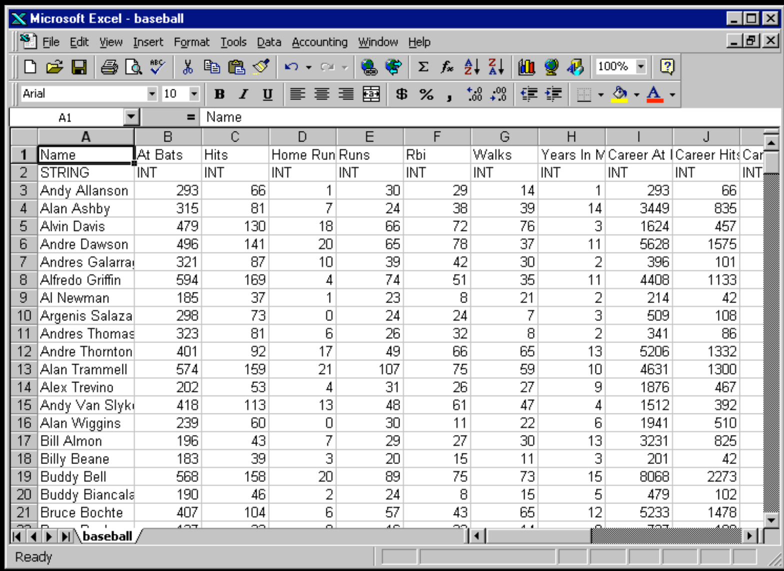
11

11



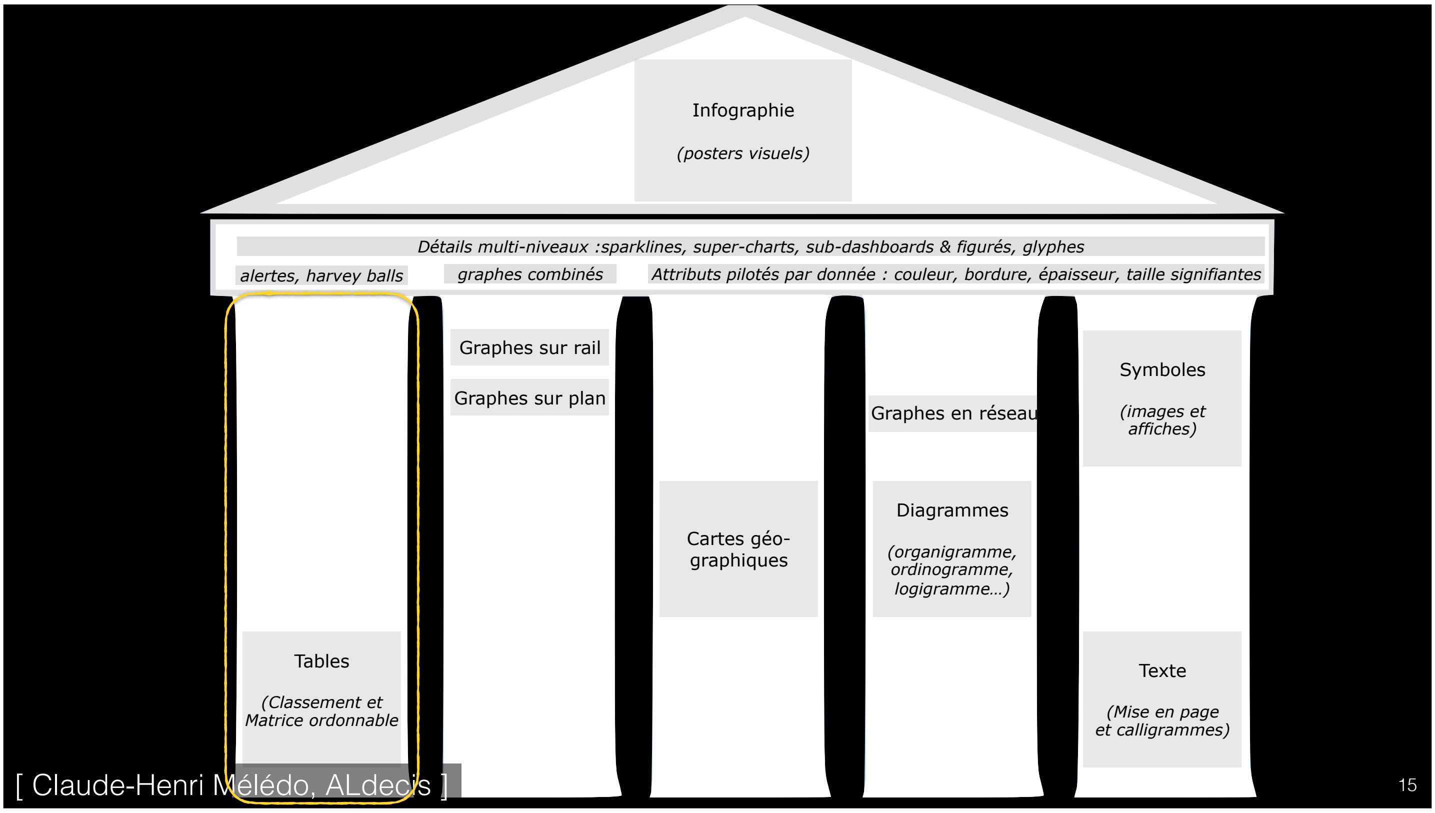
12

Tables

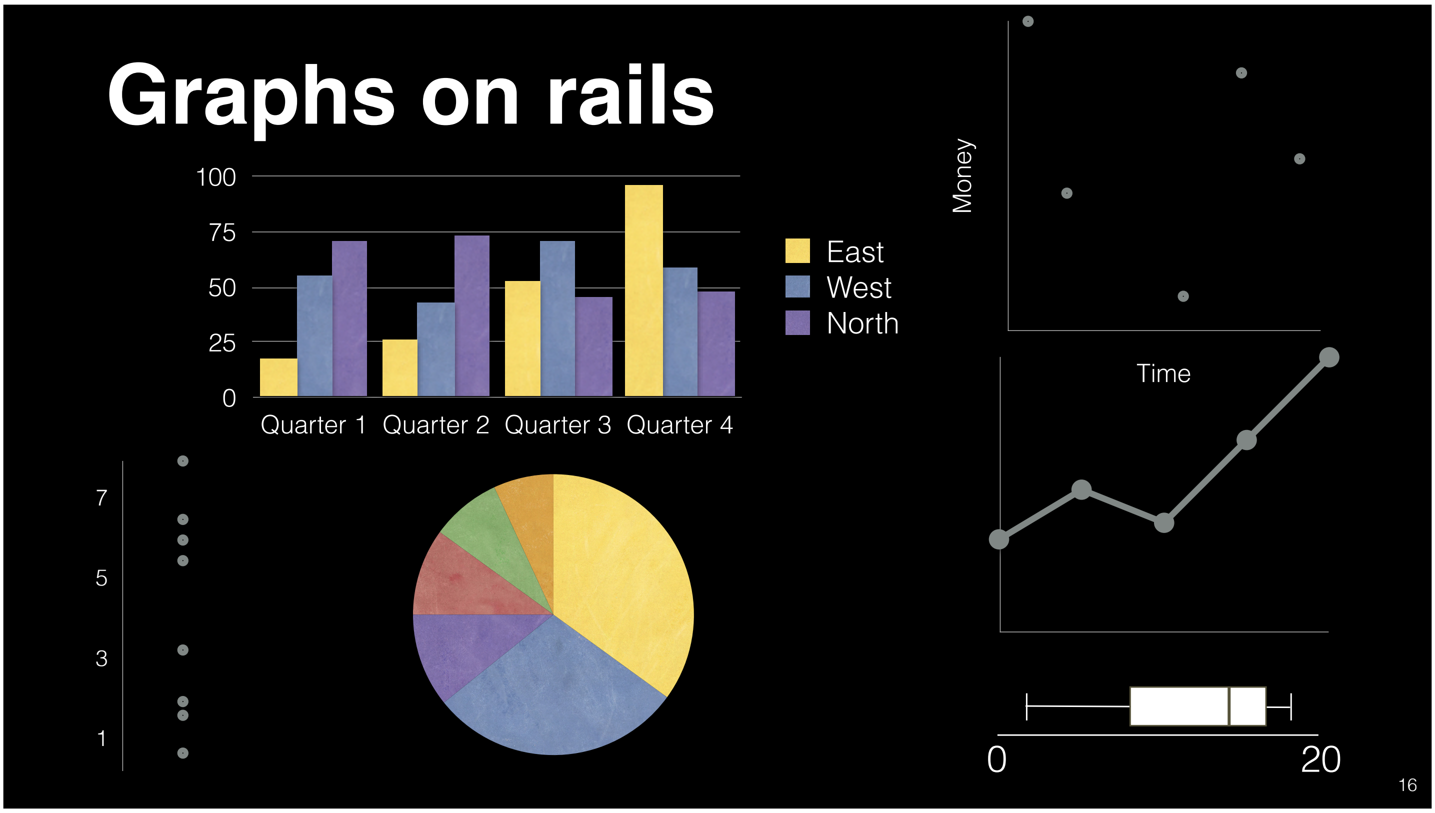


	A	B	C	D	E	F	G	H	I	J	
1	Name	At Bats	Hits	Home Run	Runs	Rbi	Walks	Years In M	Career At B	Career Hits	Car
2	STRING	INT	INT	INT	INT	INT	INT	INT	INT	INT	INT
3	Andy Allanson	293	66	1	30	29	14	1	293	66	
4	Alan Ashby	315	81	7	24	38	39	14	3449	835	
5	Alvin Davis	479	130	18	66	72	76	3	1624	457	
6	Andre Dawson	496	141	20	65	78	37	11	5628	1575	
7	Andres Galarra	321	87	10	39	42	30	2	396	101	
8	Alfredo Griffin	594	169	4	74	51	35	11	4408	1133	
9	Al Newman	185	37	1	23	8	21	2	214	42	
10	Argenis Salaza	298	73	0	24	24	7	3	509	108	
11	Andres Thomas	323	81	6	26	32	8	2	341	86	
12	Andre Thornton	401	92	17	49	66	65	13	5206	1332	
13	Alan Trammell	574	159	21	107	75	59	10	4631	1300	
14	Alex Trevino	202	53	4	31	26	27	9	1876	467	
15	Andy Van Slyke	418	113	13	48	61	47	4	1512	392	
16	Alan Wiggins	239	60	0	30	11	22	6	1941	510	
17	Bill Almon	196	43	7	29	27	30	13	3231	825	
18	Billy Beane	183	39	3	20	15	11	3	201	42	
19	Buddy Bell	568	158	20	89	75	73	15	8068	2273	
20	Buddy Biancala	190	46	2	24	8	15	5	479	102	
21	Bruce Bochte	407	104	6	57	43	65	12	5233	1478	

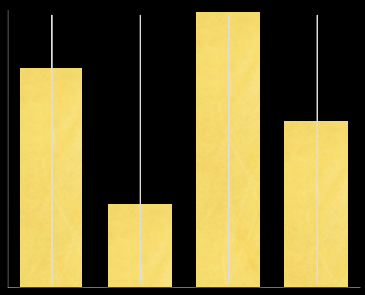
Bullet Train Schedule


Infographie
(posters visuels)
Détails multi-niveaux :sparklines, super-charts, sub-dashboards & figurés, glyphes
alertes, harvey balls
graphes combinés
Attributs pilotés par donnée : couleur, bordure, épaisseur, taille signifiantes
Graphes sur rail
Graphes sur plan
Graphes en réseau
Symboles
(images et affiches)
Cartes géo-graphiques
Diagrammes
(organigramme, ordinogramme, logigramme…)
Tables
(Classement et Matrice ordonnable
Texte
(Mise en page et calligrammes)
[Claude-Henri Mélédo, ALdecis]
15

15

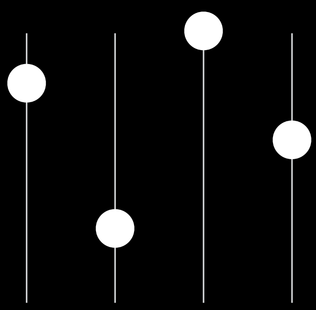

Graphs on rails
100
75
50
25
0
Quarter 1
Quarter 2
Quarter 3
Quarter 4
East
West
North
Money
Time
7
5
3
1
0
20
16

Graphs on rails



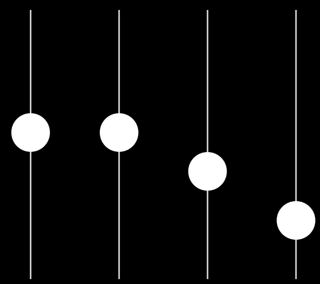
17

Graphs on rails



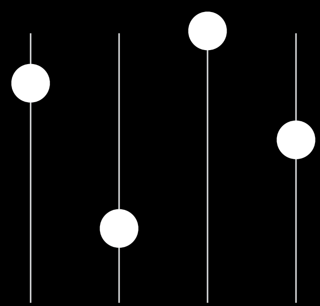
18

Graphs on rails



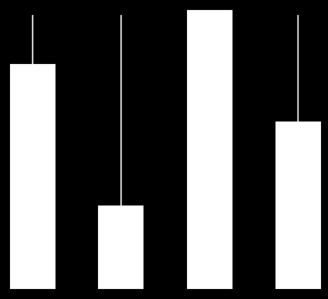
19

Graphs on rails



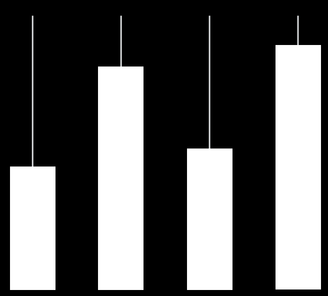
20

Graphs on rails



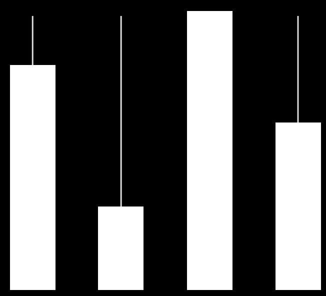
21

Graphs on rails



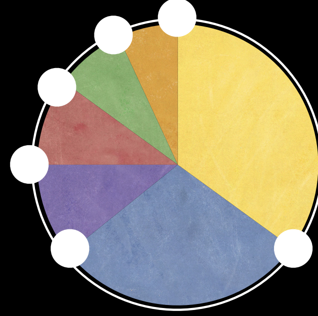
22

Graphs on rails



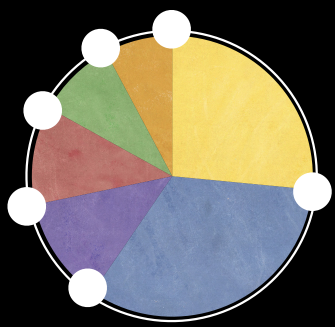
23

Graphs on rails



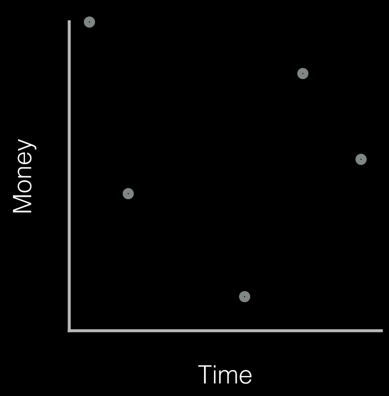
24

Graphs on rails



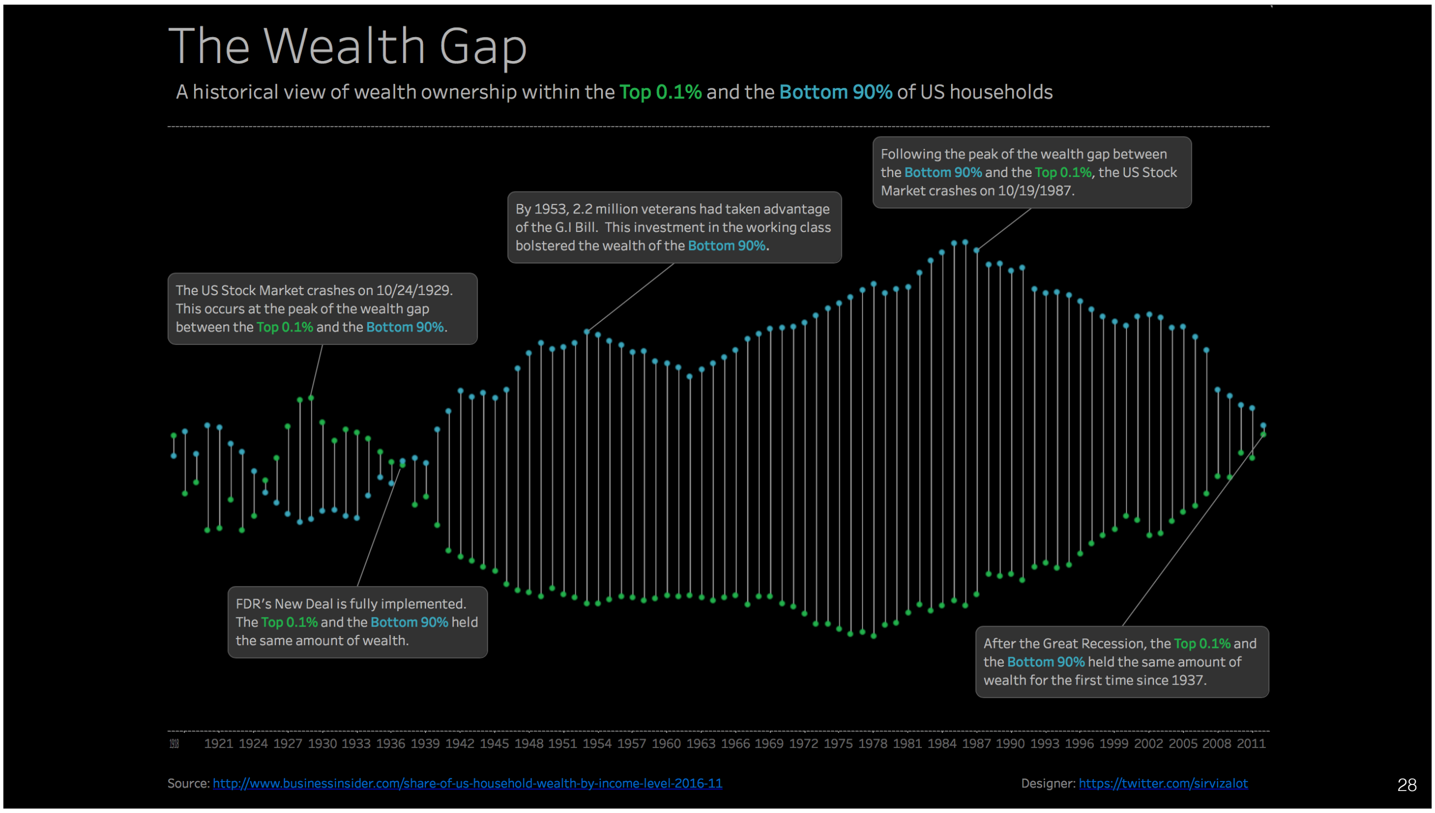
25

Graphs on the plane

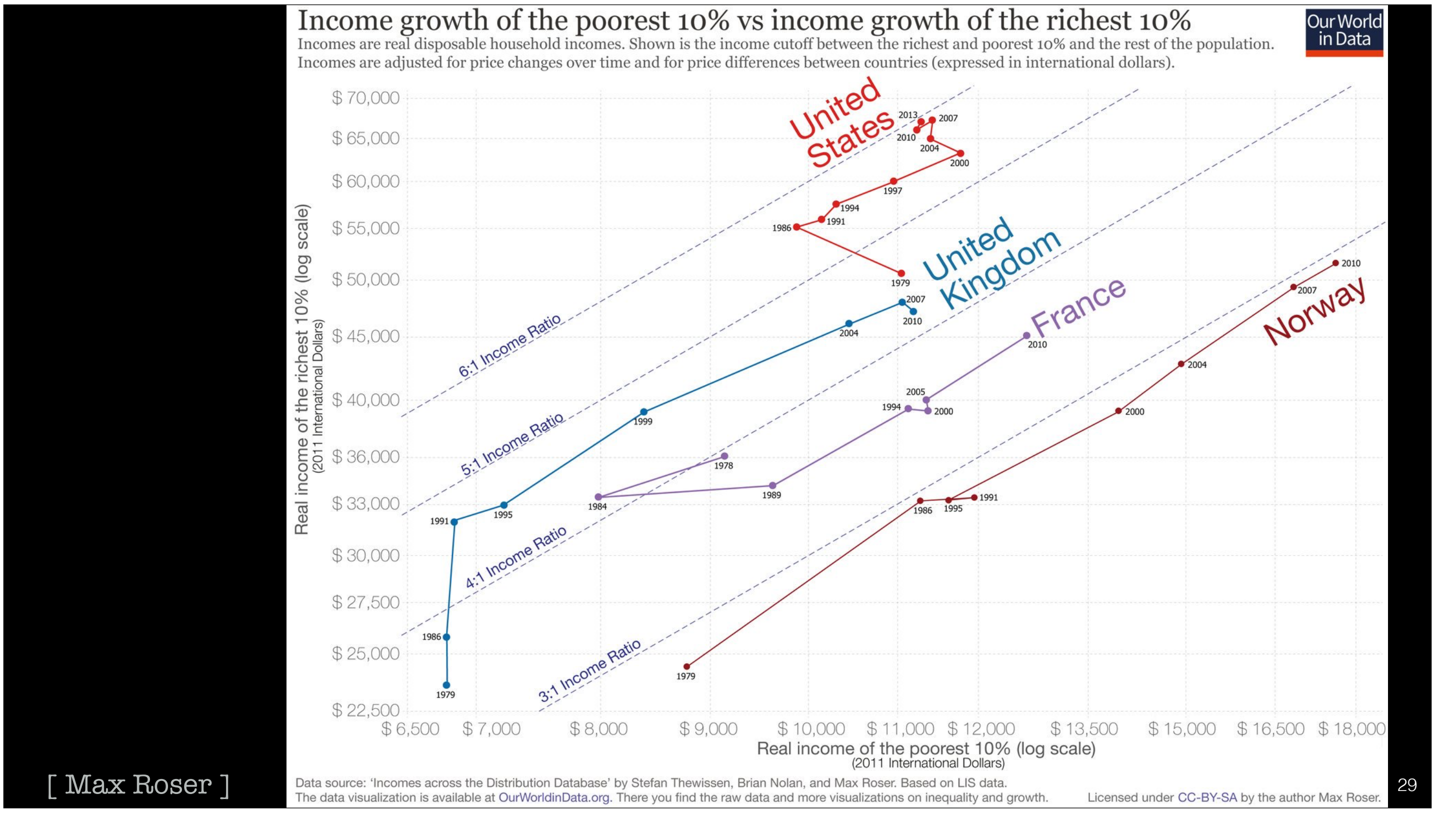


26

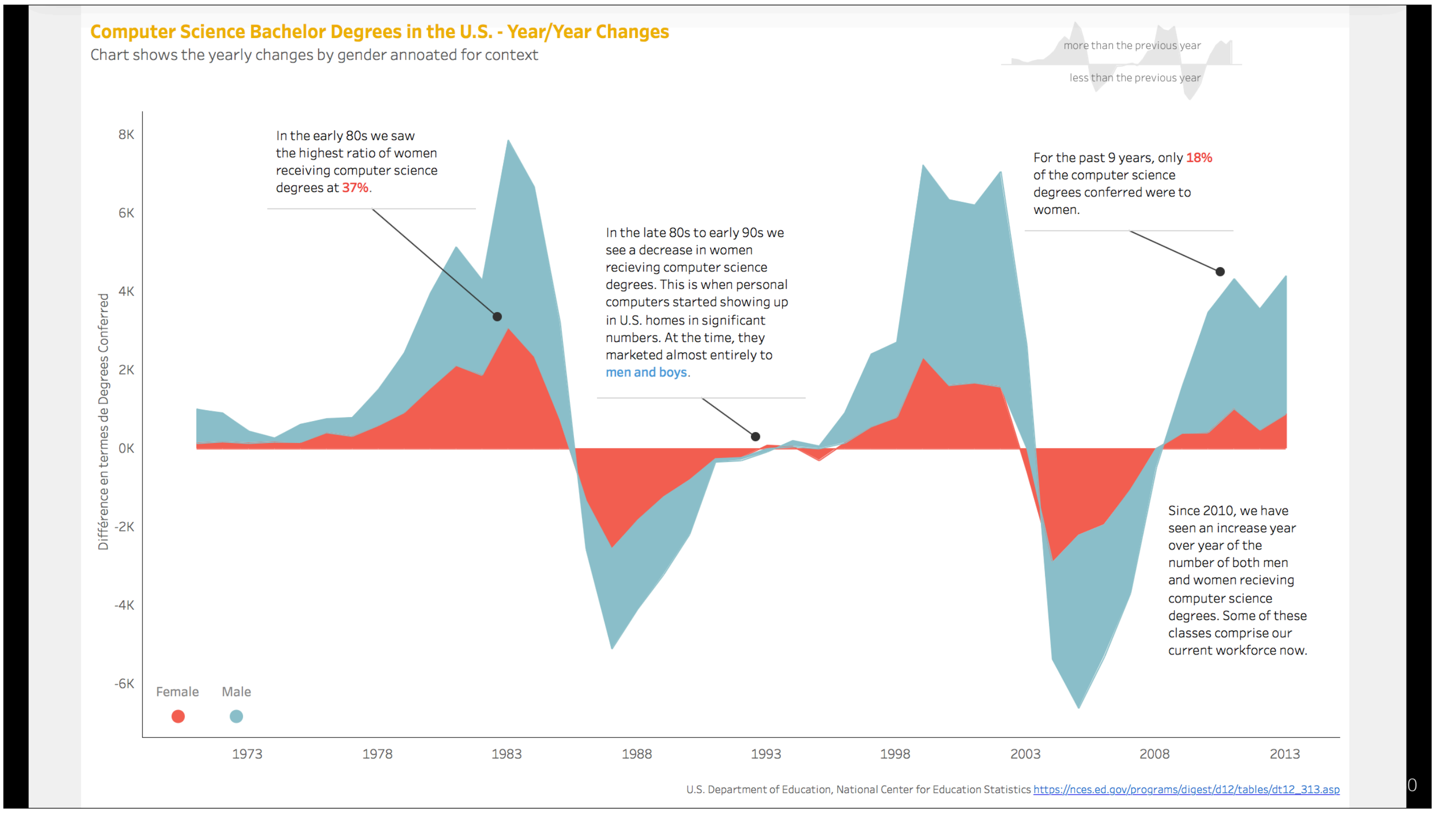
Stock Markets
CAC 40 Index
CAC:IND
+ Watchlist
5,342.71
-1.55
-0.03%
As of 7:58 AM EST 2/27/2018
Before it's here, it's on the Bloomberg Terminal.
Learn More
1D 1M 1Y 5Y
Time Frame
Add Comparison
Indicators
Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar
27/03/2020
5,500.00
5,000.00
4,500.00
4,000.00
05/06 5,483.52
5,400.00
5,200.00
5,000.00
4,800.00
100 M
0
-100
07/06



The Wealth Gap
A historical view of wealth ownership within the Top 0.1% and the Bottom 90% of US households
Following the peak of the wealth gap between the Bottom 90% and the Top 0.1%, the US Stock Market crashes on 10/19/1987.
By 1953, 2.2 million veterans had taken advantage of the G.I Bill. This investment in the working class bolstered the wealth of the Bottom 90%.
The US Stock Market crashes on 10/24/1929. This occurs at the peak of the wealth gap between the Top 0.1% and the Bottom 90%.
FDR's New Deal is fully implemented. The Top 0.1% and the Bottom 90% held the same amount of wealth.
After the Great Recession, the Top 0.1% and the Bottom 90% held the same amount of wealth for the first time since 1937.
1921 1924 1927 1930 1933 1936 1939 1942 1945 1948 1951 1954 1957 1960 1963 1966 1969 1972 1975 1978 1981 1984 1987 1990 1993 1996 1999 2002 2005 2008 2011
Source: http://www.businessinsider.com/share-of-us-household-wealth-by-income-level-2016-11
Designer: https://twitter.com/sirvizalot
28

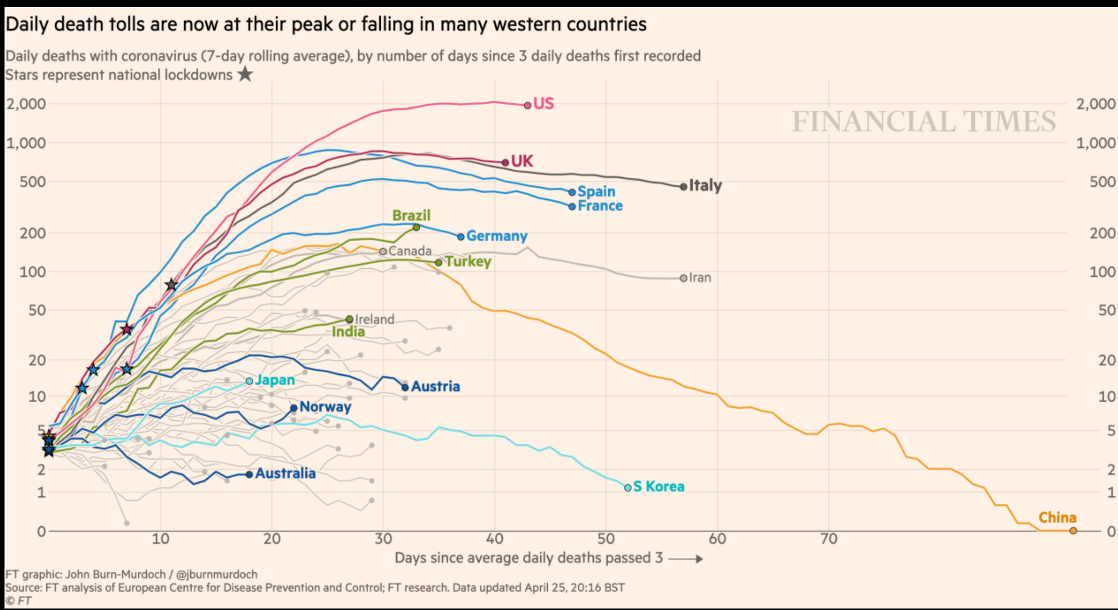

Income growth of the poorest 10% vs income growth of the richest 10%
Incomes are real disposable household incomes. Shown is the income cutoff between the richest and poorest 10% and the rest of the population.
Incomes are adjusted for price changes over time and for price differences between countries (expressed in international dollars).
Our World in Data
Real income of the richest 10% (log scale)
(2011 International Dollars)
$ 70,000
$ 65,000
$ 60,000
$ 55,000
$ 50,000
$ 45,000
$ 40,000
$ 36,000
$ 33,000
$ 30,000
$ 27,500
$ 25,000
$ 22,500
$ 6,500
$ 7,000
$ 8,000
$ 9,000
$ 10,000
$ 11,000
$ 12,000
$ 13,500
$ 15,000
$ 16,500
$ 18,000
Real income of the poorest 10% (log scale)
(2011 International Dollars)
6:1 Income Ratio
5:1 Income Ratio
4:1 Income Ratio
3:1 Income Ratio
United States
United Kingdom
France
Norway
[Max Roser]
Data source: 'Incomes across the Distribution Database' by Stefan Thewissen, Brian Nolan, and Max Roser. Based on LIS data.
The data visualization is available at OurWorldinData.org. There you find the raw data and more visualizations on inequality and growth.
Licensed under CC-BY-SA by the author Max Roser.
29

29


Computer Science Bachelor Degrees in the U.S. - Year/Year Changes
Chart shows the yearly changes by gender annotated for context
more than the previous year
less than the previous year
Différence en termes de Degrees Conferred
8K
6K
4K
2K
0K
-2K
-4K
-6K
In the early 80s we saw the highest ratio of women receiving computer science degrees at 37%.
In the late 80s to early 90s we see a decrease in women recieving computer science degrees. This is when personal computers started showing up in U.S. homes in significant numbers. At the time, they marketed almost entirely to men and boys.
For the past 9 years, only 18% of the computer science degrees conferred were to women.
Since 2010, we have seen an increase year over year of the number of both men and women recieving computer science degrees. Some of these classes comprise our current workforce now.
Female
Male
1973
1978
1983
1988
1993
1998
2003
2008
2013
U.S. Department of Education, National Center for Education Statistics https://nces.ed.gov/programs/digest/d12/tables/dt12_313.asp

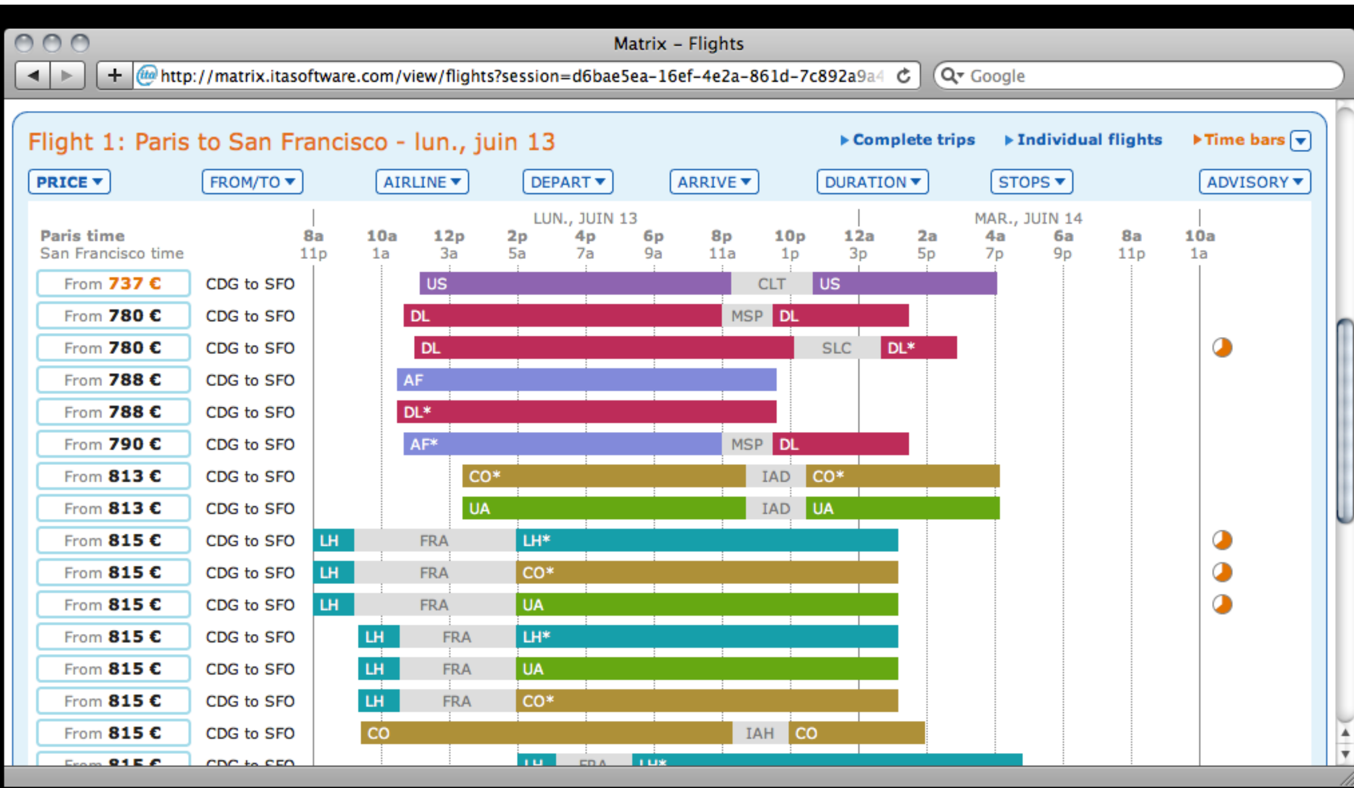
30

Line chart

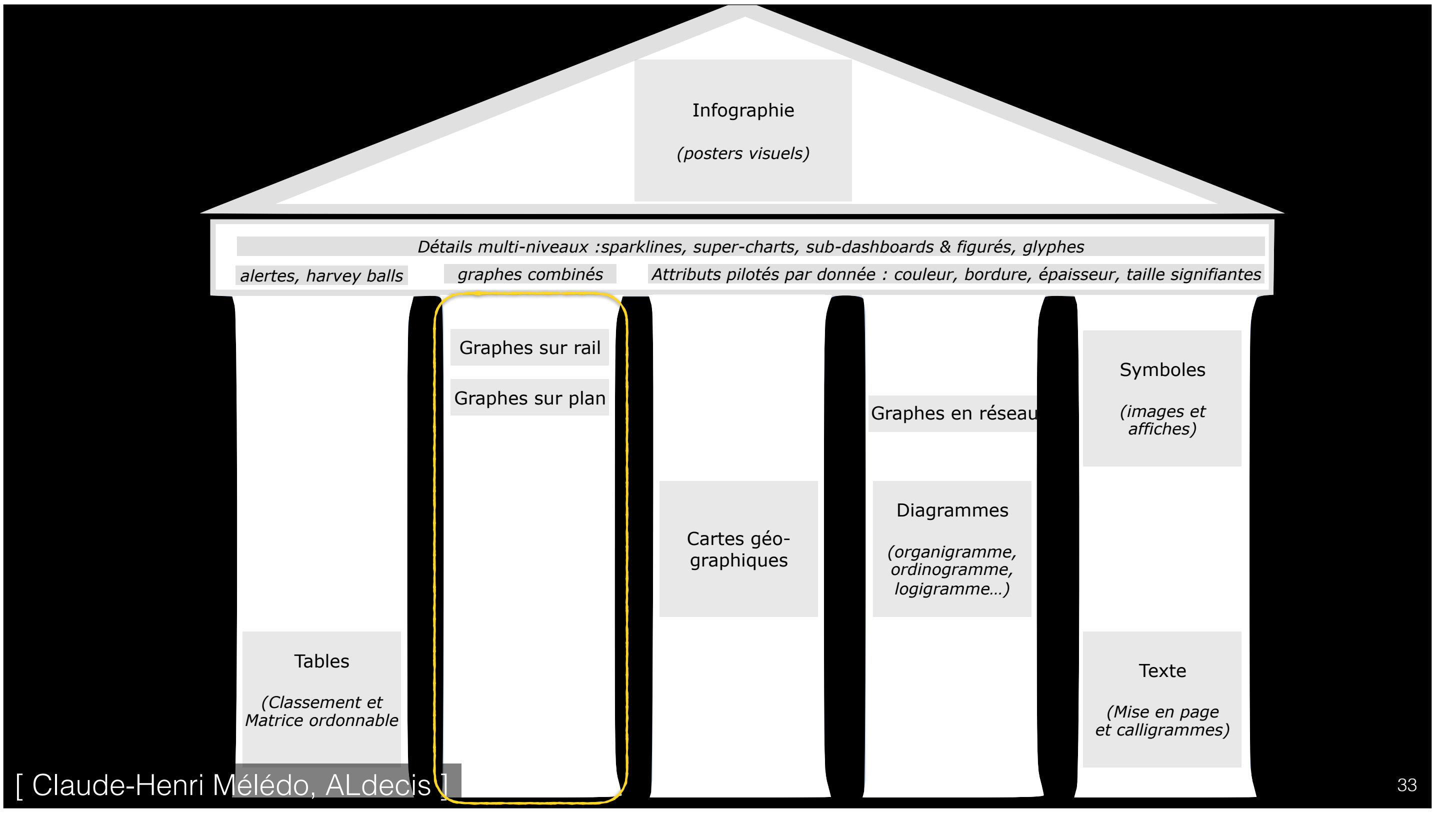


[Financial Times, 27/04/2020]

31

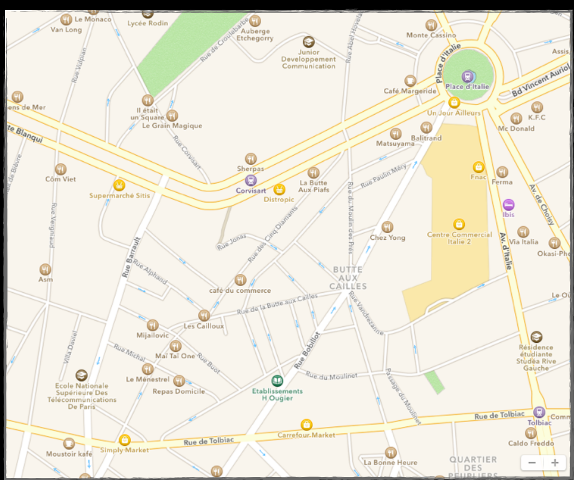


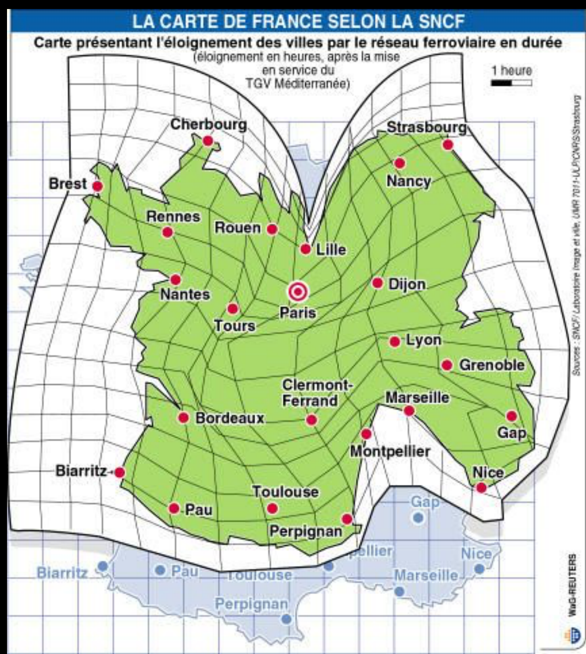
32



33

Geospatial maps

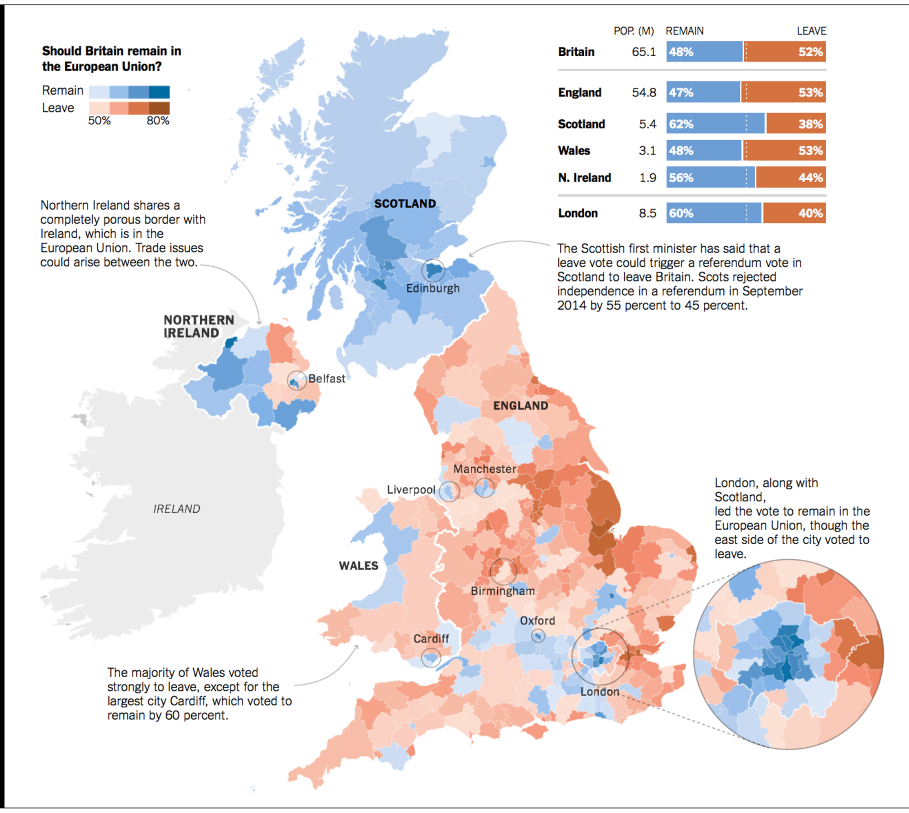




34

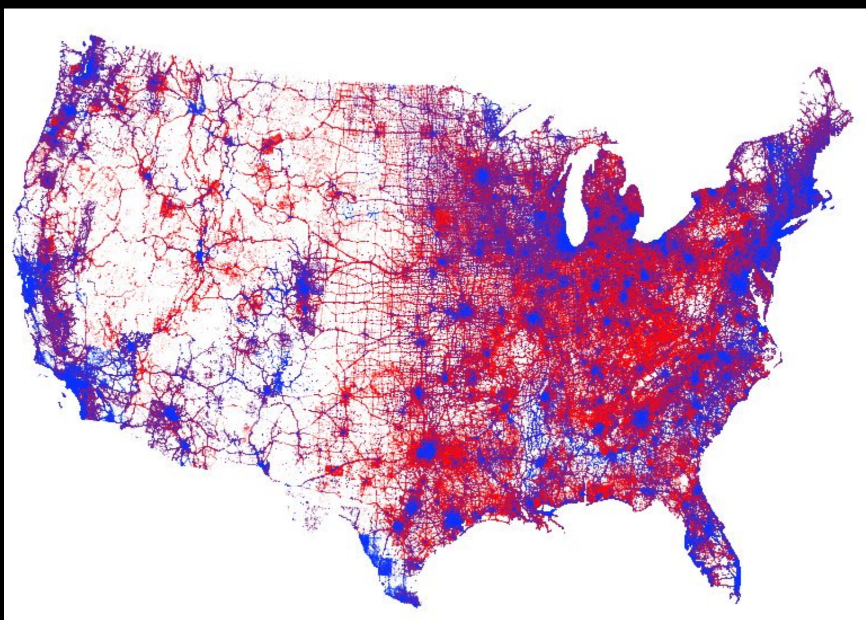
34

Brexit



35

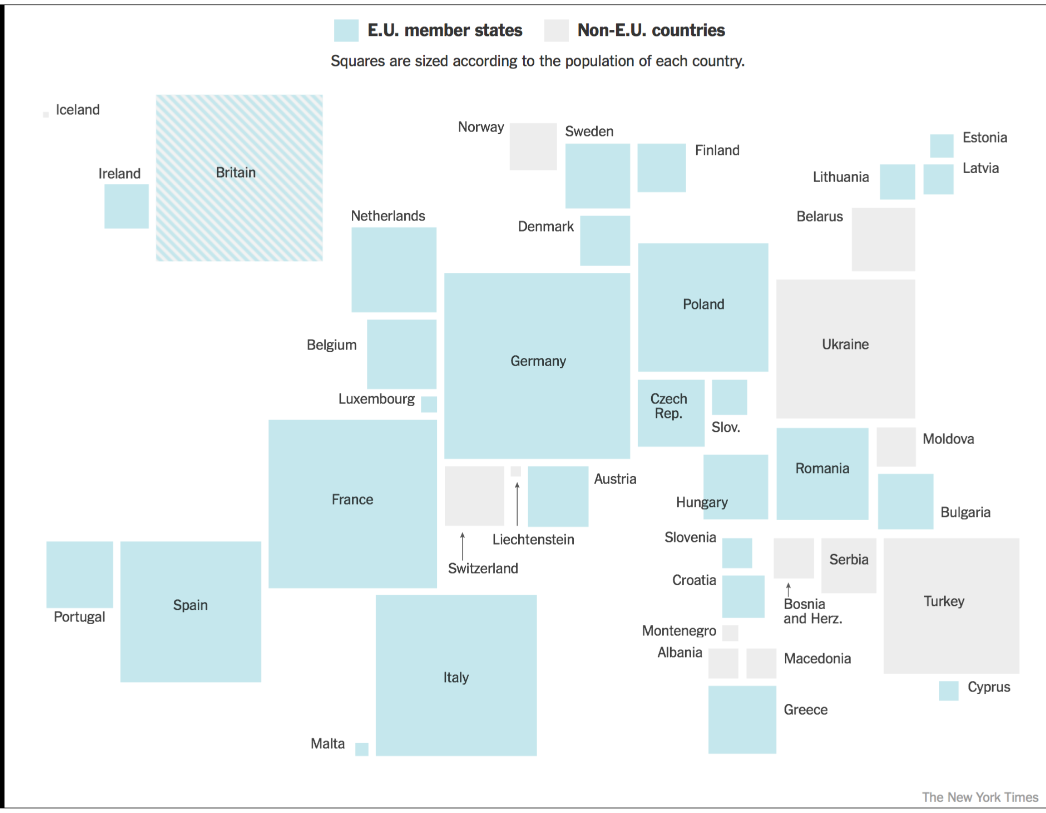
35



36

36

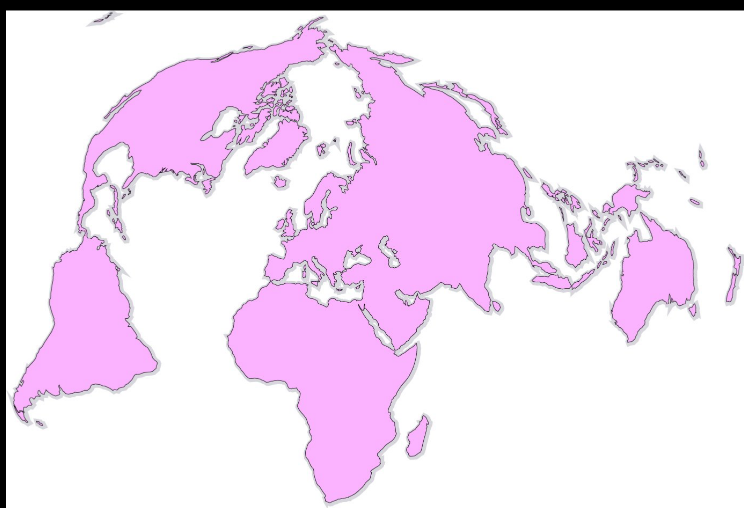
E.U.



E.U. member states
Non-E.U. countries
Squares are sized according to the population of each country.
Iceland
Ireland
Britain
Norway
Sweden
Finland
Estonia
Latvia
Lithuania
Belarus
Netherlands
Denmark
Poland
Ukraine
Belgium
Germany
Luxembourg
Czech Rep.
Slov.
Moldova
Romania
France
Austria
Hungary
Bulgaria
Liechtenstein
Switzerland
Slovenia
Serbia
Croatia
Turkey
Bosnia and Herz.
Portugal
Spain
Montenegro
Albania
Macedonia
Italy
Cyprus
Greece
Malta
The New York Times

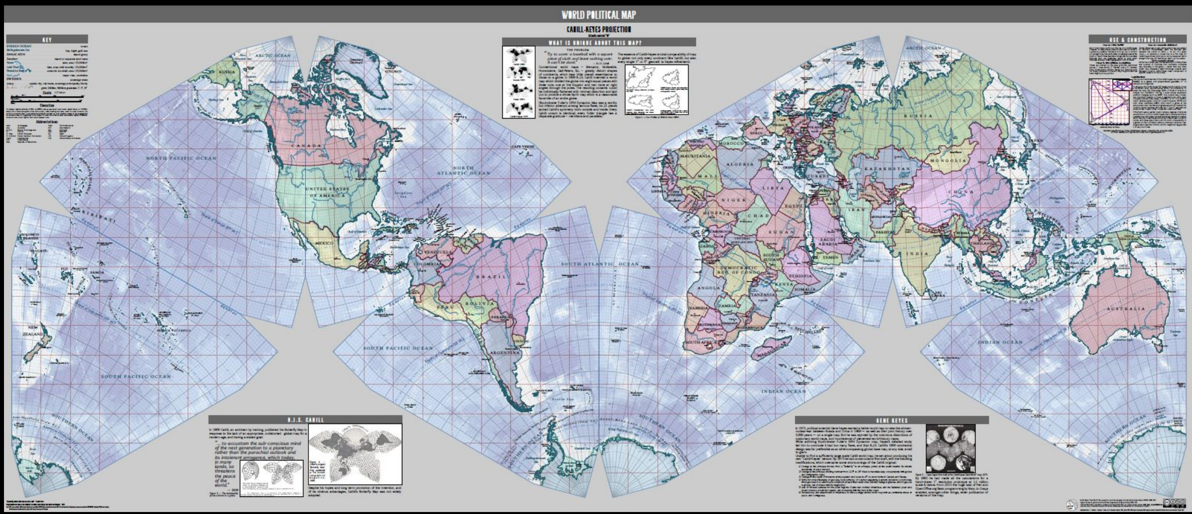
37

Détroit de Béring
OCÉAN ARCTIQUE
Projection de Spilhaus
Quand les océans sont au centre du monde
AMÉRIQUE DU NORD
EUROPE
ASIE
ASIE
OCÉAN ATLANTIQUE
AFRIQUE
OCÉAN INDIEN
AMÉRIQUE DU SUD
Antarctique
ASIE
OCÉAN PACIFIQUE
Détroit de Béring
Le Cartographe
www.le-cartographe.net
OCÉAN PACIFIQUE
AMÉRIQUE DU SUD
AMÉRIQUE DU NORD



38

Duncan-Webb



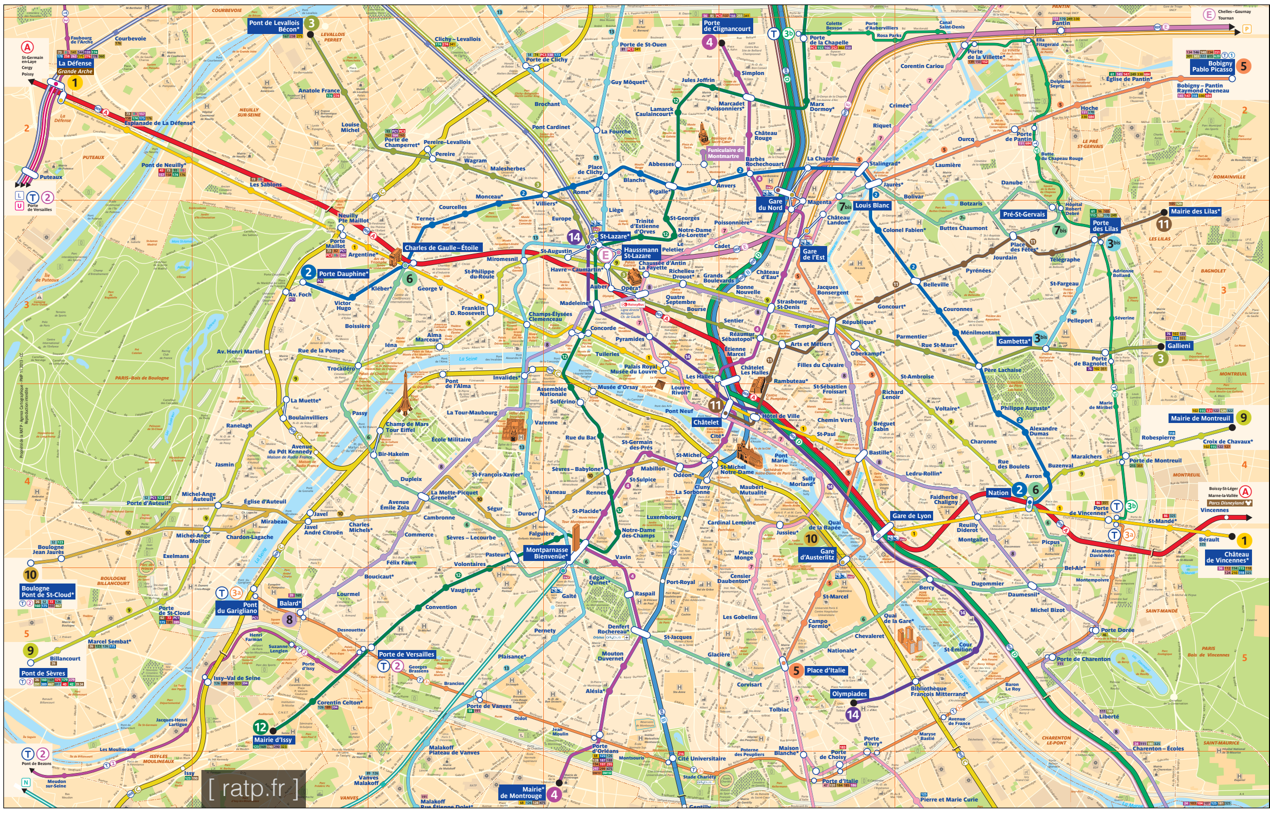
[Source : Wired]

Country Music

Figure 14. States Mentioned in Country-Music Lyrics

Source: Ben Marsh, "A Rose-Colored Map," *Harper's*, July 1977, 80. Used by permission.

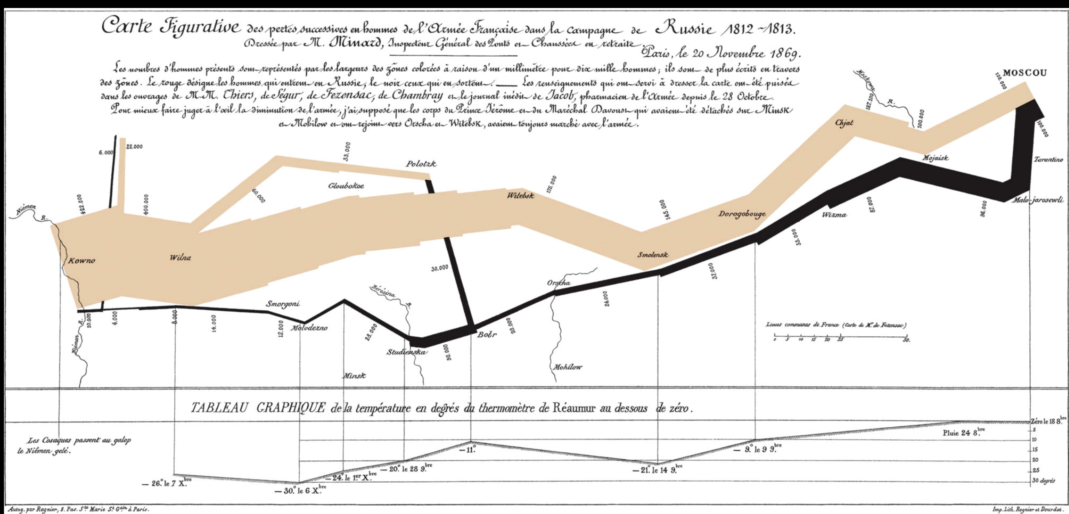
Note: The size of each state is proportional to the number of times it is mentioned.



41

Napolean's March

From E. Tufte
The Visual Display of Quantitative Information



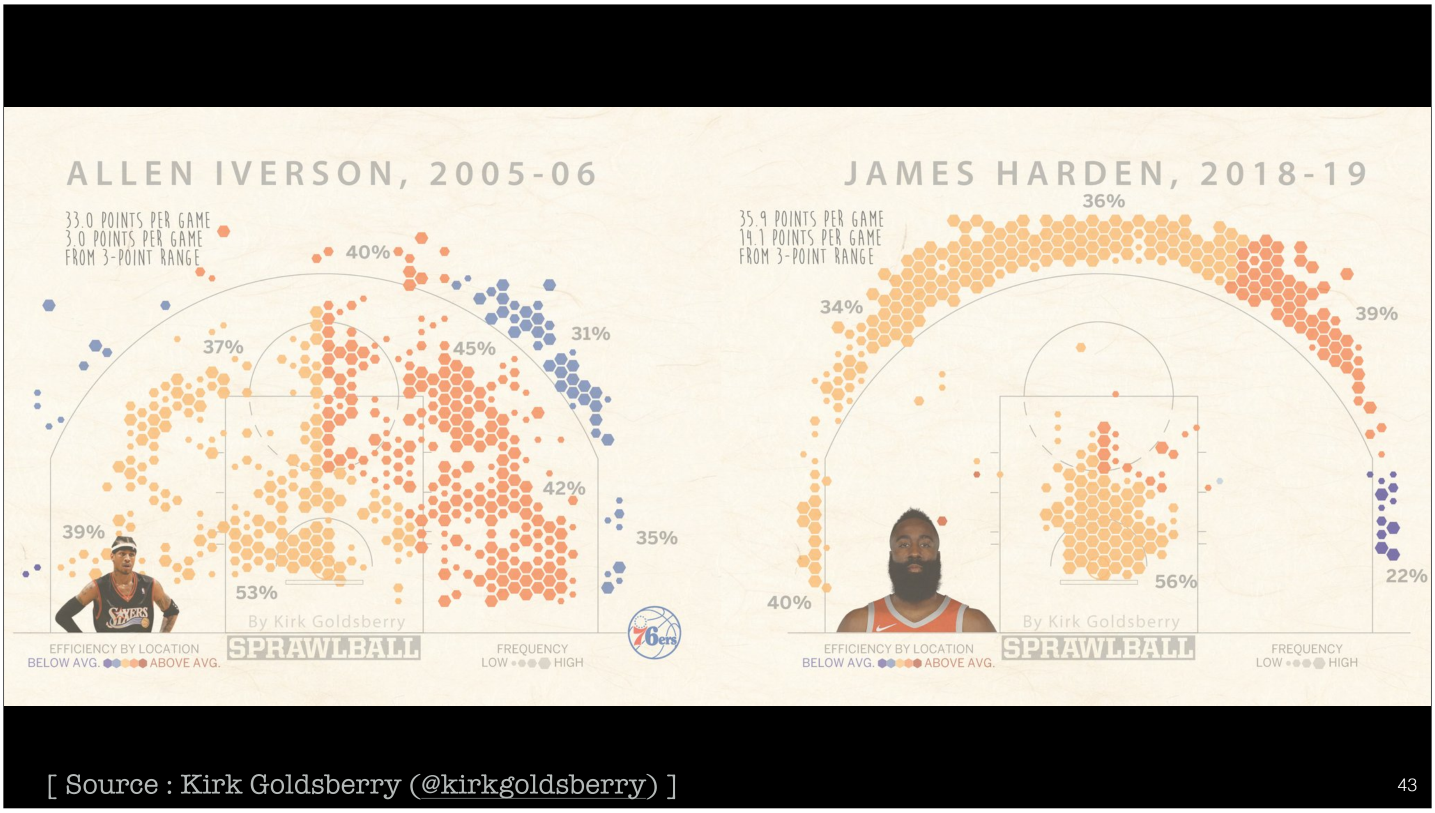
Graphic by Minard

size of army
direction

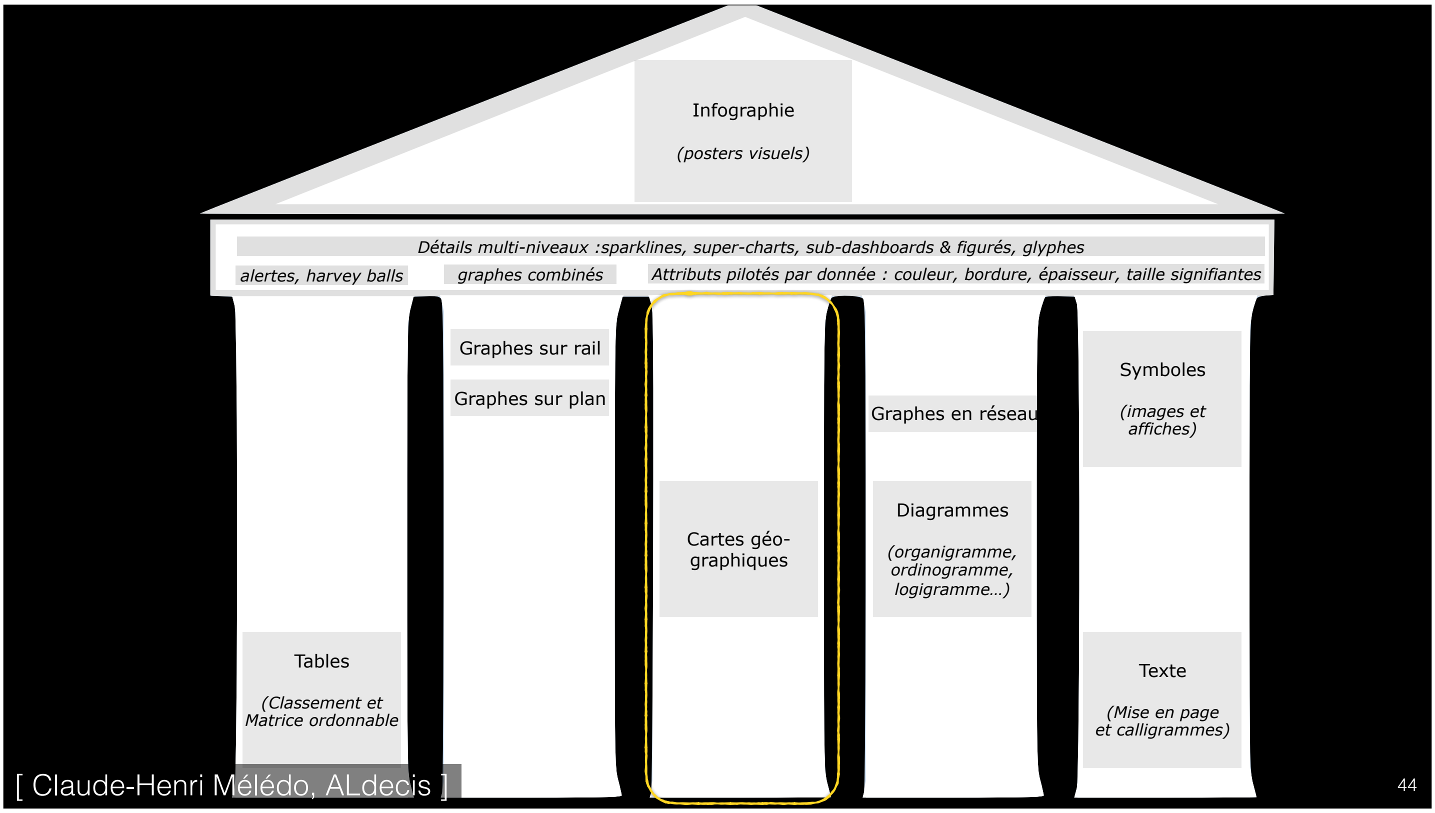
latitude
longitude

temperature
date

42

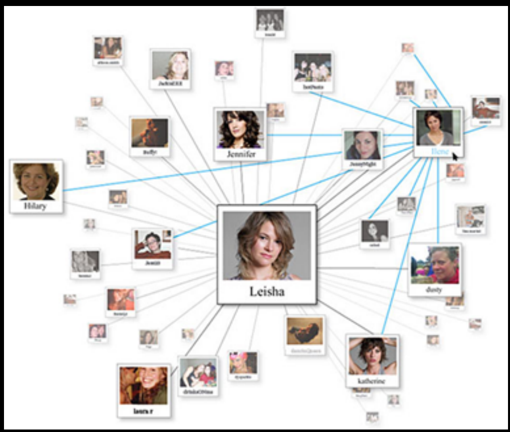

ALLEN IVERSON, 2005-06
33.0 POINTS PER GAME
3.0 POINTS PER GAME
FROM 3-POINT RANGE
40%
37%
45%
31%
42%
35%
39%
53%
By Kirk Goldsberry
EFFICIENCY BY LOCATION
BELOW AVG. ABOVE AVG.
SPRAWLBALL
FREQUENCY
LOW HIGH
JAMES HARDEN, 2018-19
35.9 POINTS PER GAME
14.1 POINTS PER GAME
FROM 3-POINT RANGE
36%
34%
39%
40%
56%
22%
By Kirk Goldsberry
EFFICIENCY BY LOCATION
BELOW AVG. ABOVE AVG.
SPRAWLBALL
FREQUENCY
LOW HIGH
[Source : Kirk Goldsberry (@kirkgoldsberry)]
43

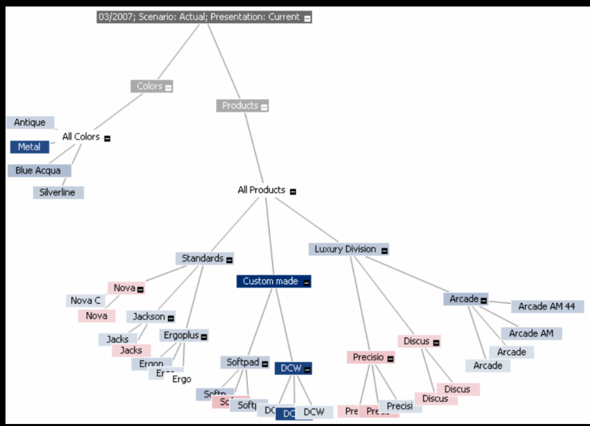
43


Infographie
(posters visuels)
Détails multi-niveaux :sparklines, super-charts, sub-dashboards & figurés, glyphes
alertes, harvey balls
graphes combinés
Attributs pilotés par donnée : couleur, bordure, épaisseur, taille signifiantes
Graphes sur rail
Graphes sur plan
Cartes géo-graphiques
Graphes en réseau
Diagrammes
(organigramme, ordinogramme, logigramme...)
Symboles
(images et affiches)
Tables
(Classement et Matrice ordonnable
Texte
(Mise en page et calligrammes)
[Claude-Henri Mélédo, ALdecis]
44

44

Networks & Diagrams

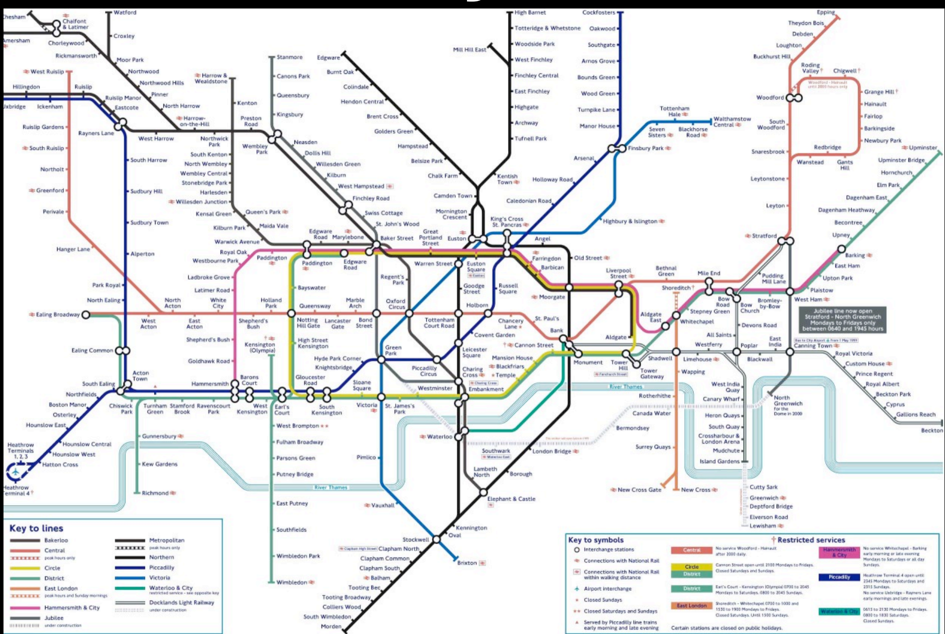




45

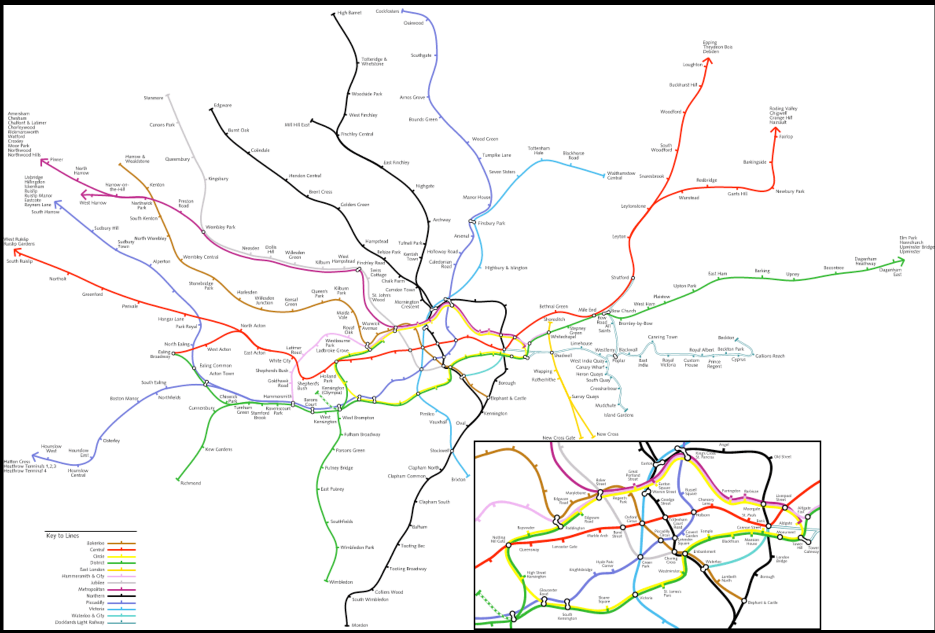
London Subway

www.thetube.com



46

True Geography



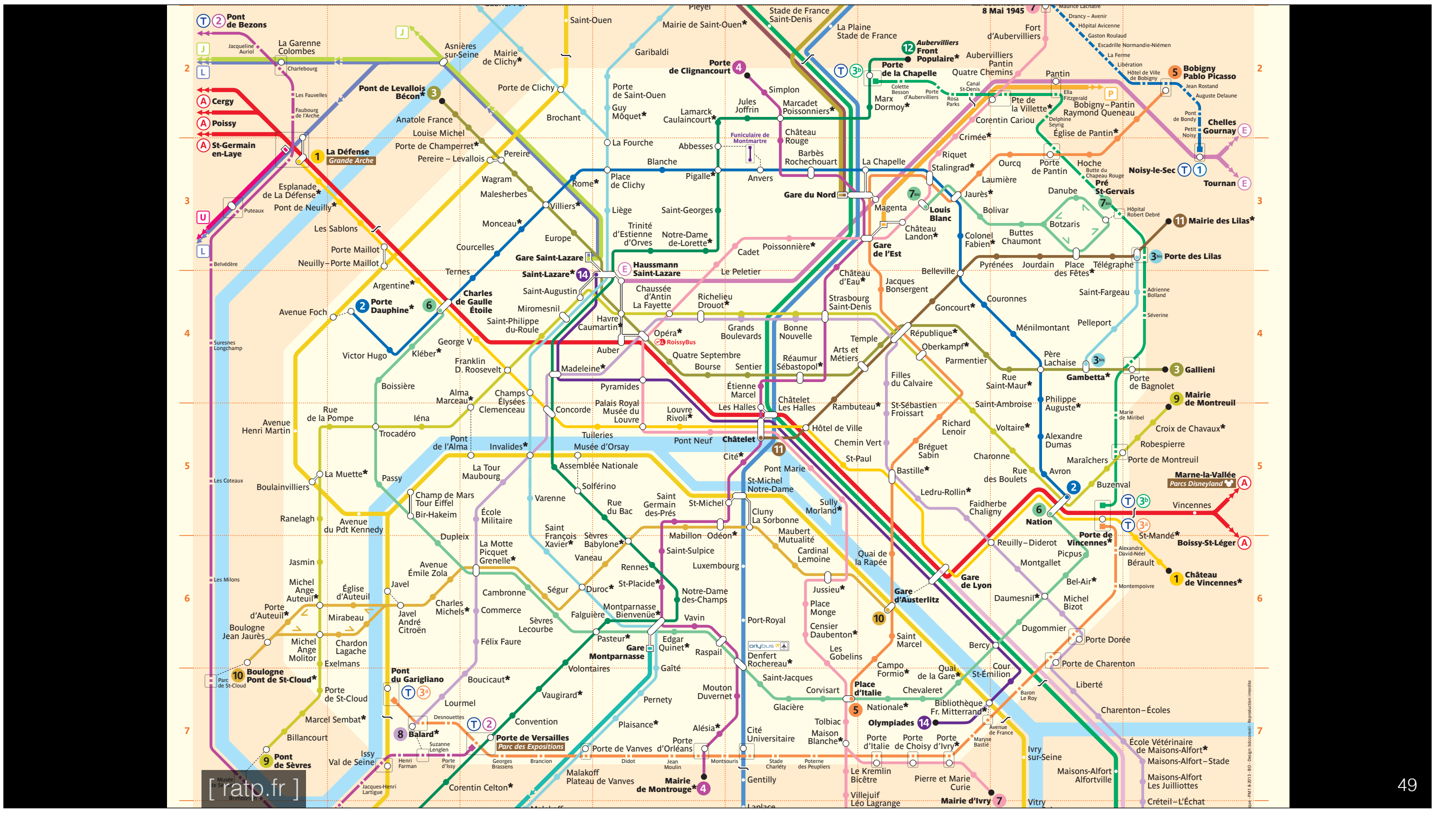
www.kottke.org/plus/misc/images/tubegeo.gif

47

Easy Walking Lines Added

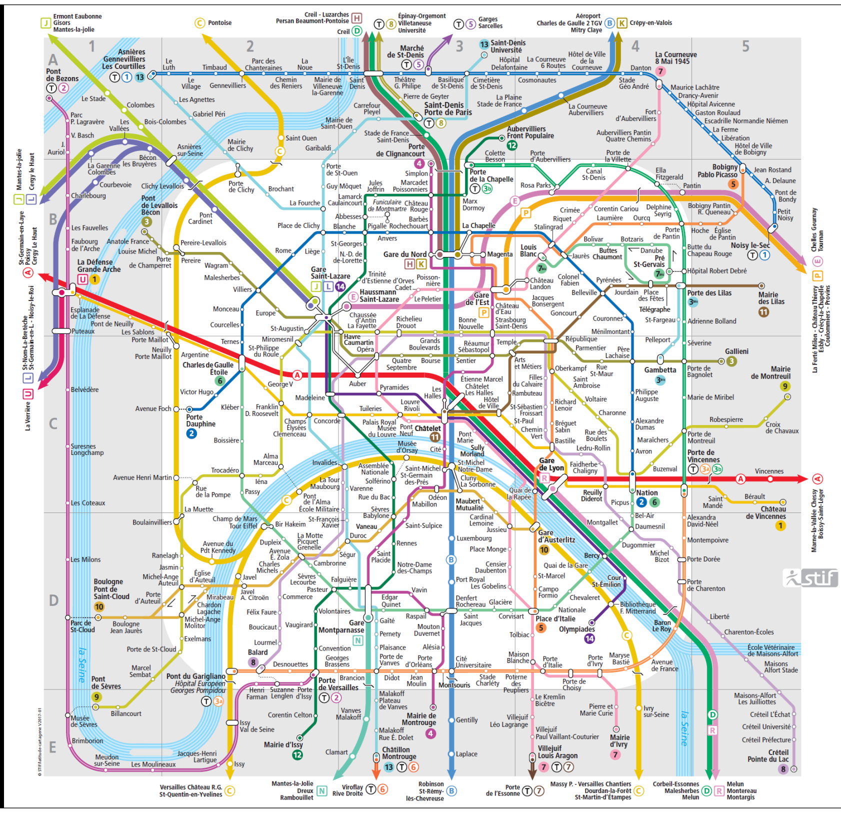
rodcorp.typepad.com/photos/art_2003/tube_walklines_final_lmfaint.html

48



49

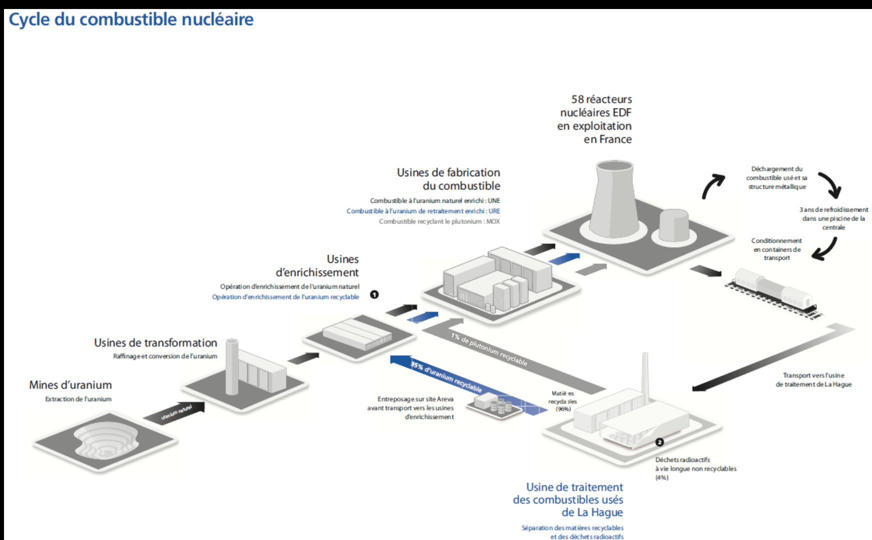
50



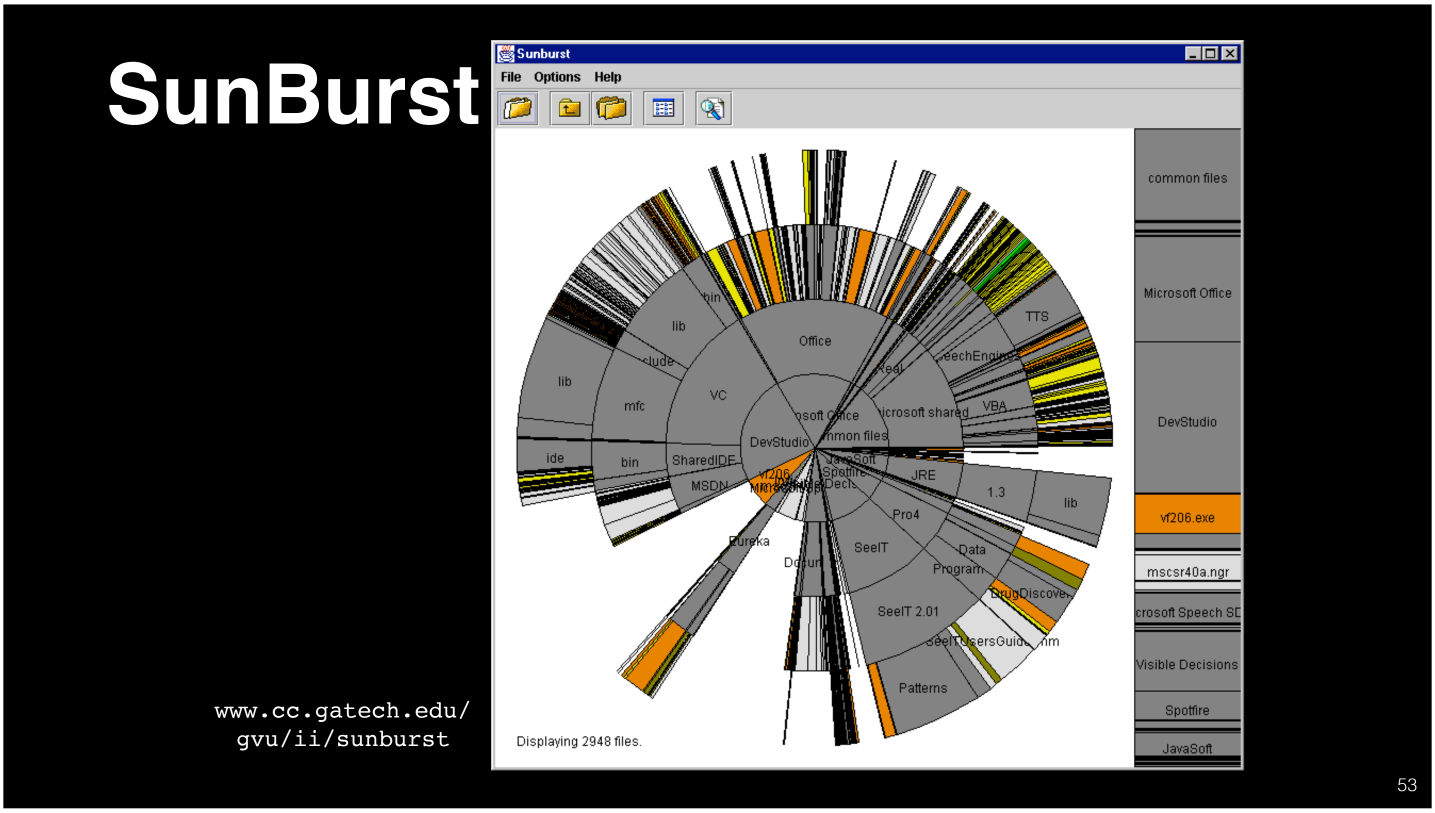
51

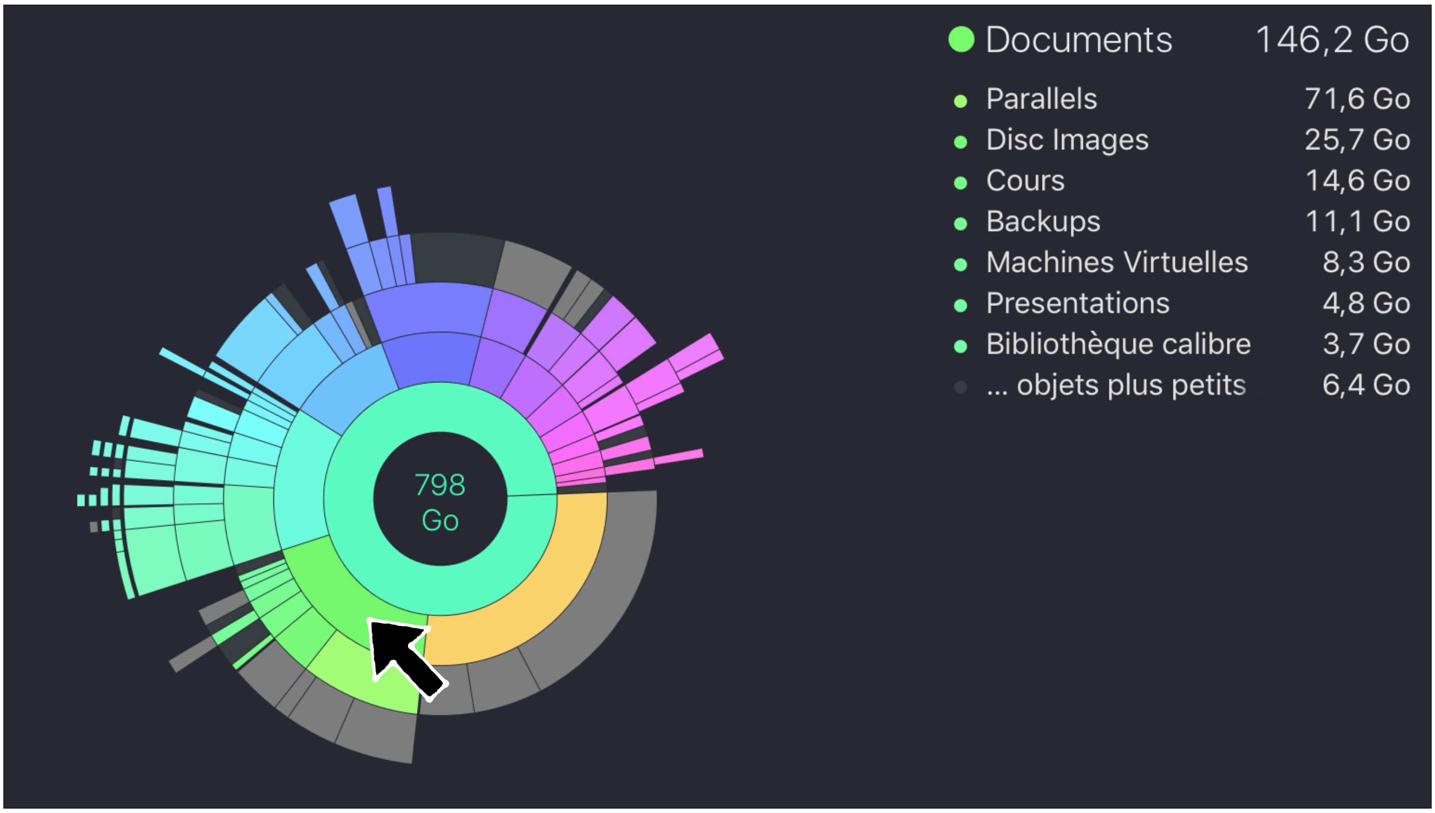
51

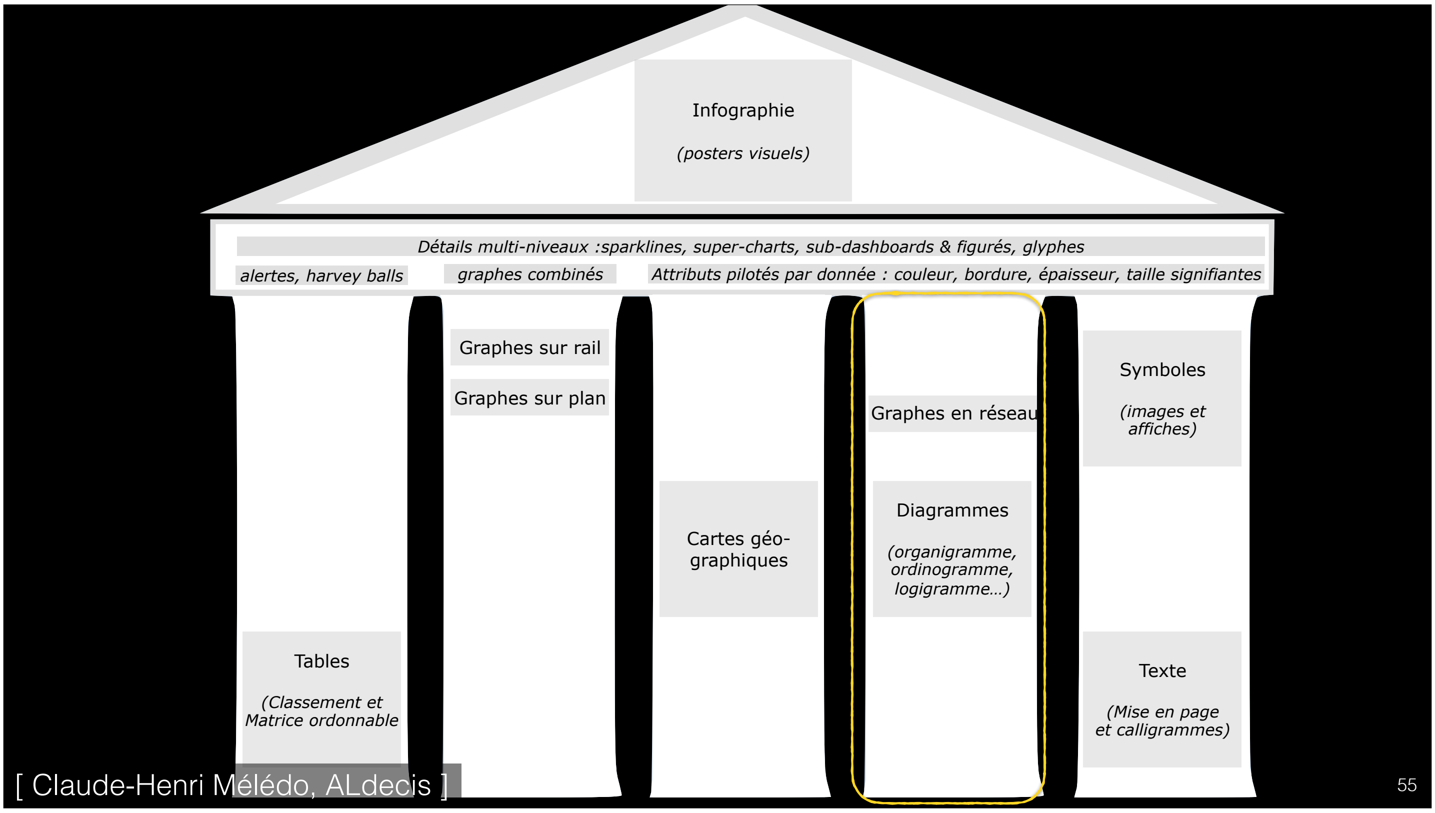
Networks & Diagrams



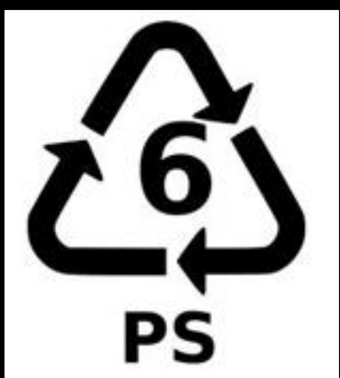
52

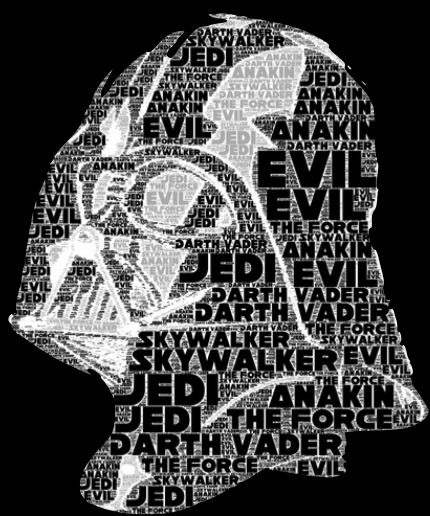

SunBurst
Sunburst
File Options Help
Office
lib
bin
mfc
ide
VC
DevStudio
SharedIDE
MSDN
TTS
VBA
JRE
1.3
lib
Pro4
SeeIT
Data
Program
SeeIT 2.01
Patterns
Eureka
common files
Microsoft Office
DevStudio
vf206.exe
mscsr40a.ngr
Visible Decisions
Spotfire
JavaSoft
Displaying 2948 files.
www.cc.gatech.edu/
gvu/ii/sunburst
53


798
Go
Documents 146,2 Go
Parallels 71,6 Go
Disc Images 25,7 Go
Cours 14,6 Go
Backups 11,1 Go
Machines Virtuelles 8,3 Go
Presentations 4,8 Go
Bibliothèque calibre 3,7 Go
... objets plus petits 6,4 Go

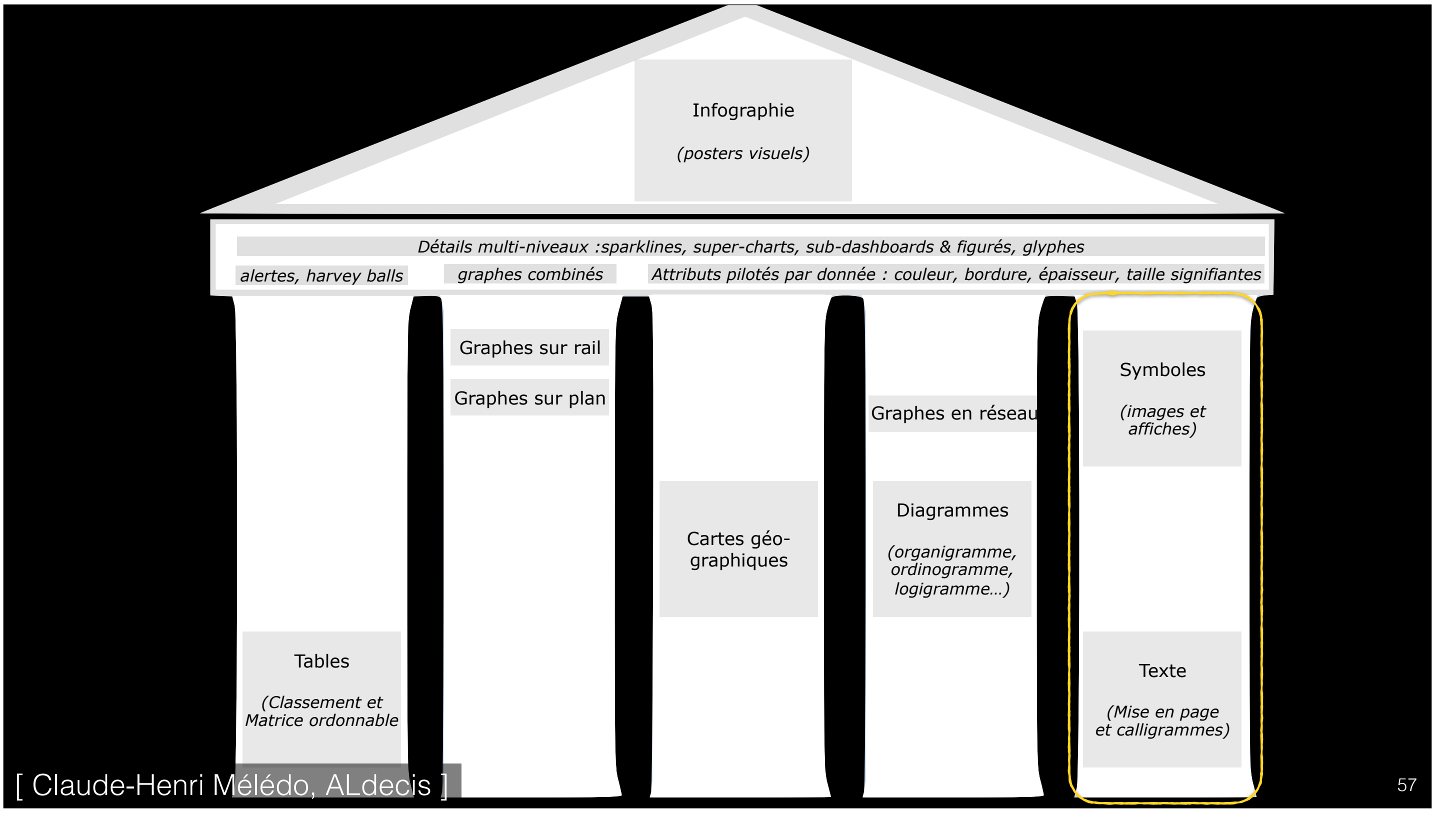


Conceptual Images

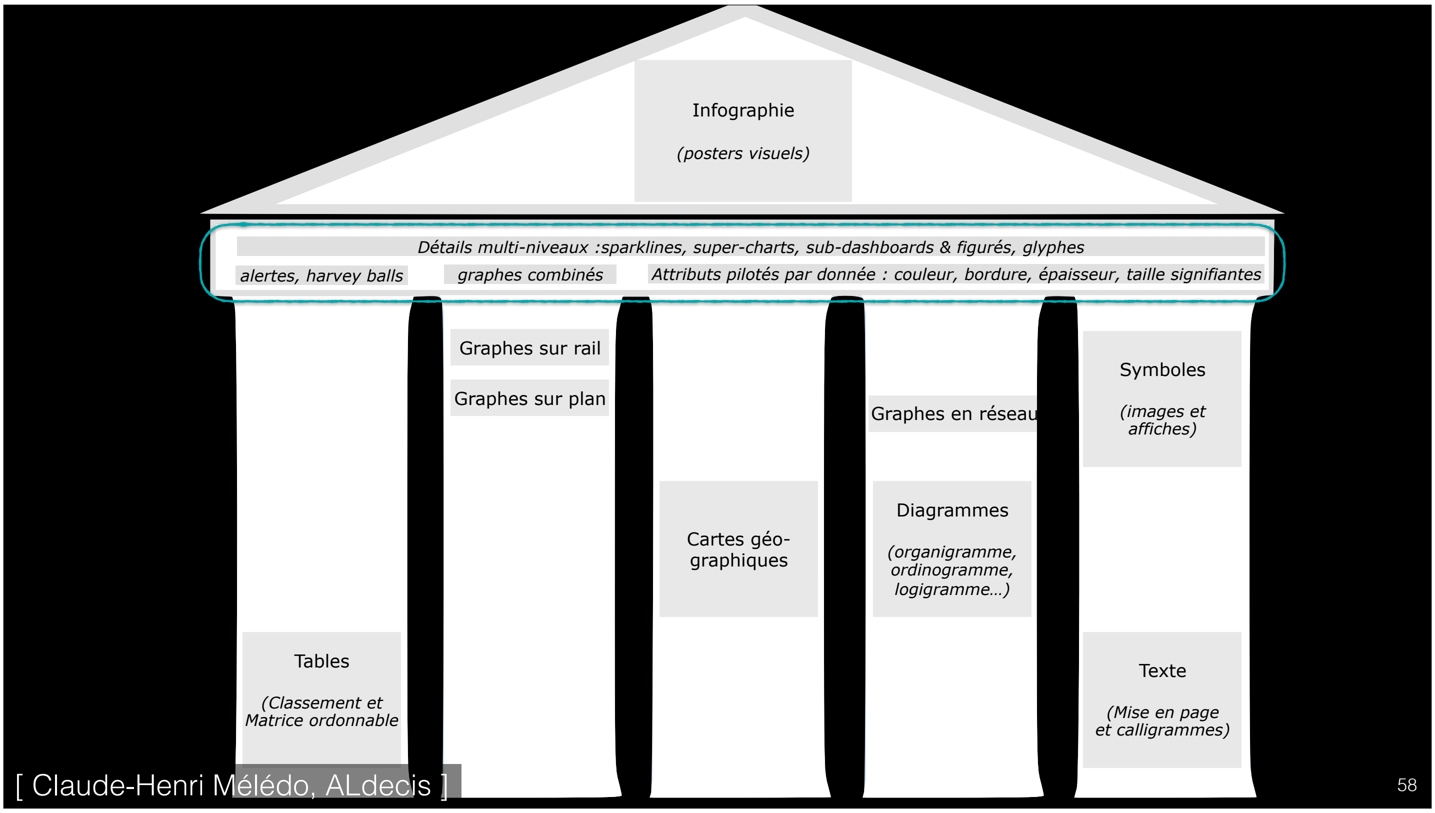




56

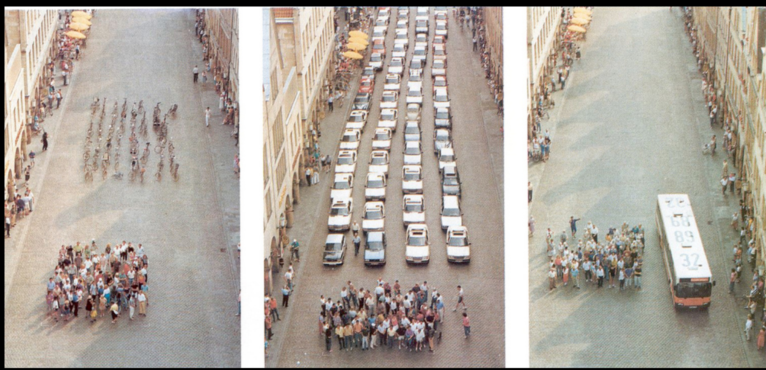

Infographie
(posters visuels)
Détails multi-niveaux :sparklines, super-charts, sub-dashboards & figurés, glyphes
alertes, harvey balls
graphes combinés
Attributs pilotés par donnée : couleur, bordure, épaisseur, taille signifiantes
Graphes sur rail
Graphes sur plan
Graphes en réseau
Symboles
(images et affiches)
Cartes géo-graphiques
Diagrammes
(organigramme, ordinogramme, logigramme…)
Tables
(Classement et Matrice ordonnable
Texte
(Mise en page et calligrammes)
[Claude-Henri Mélédo, ALdecis]
57

57


Infographie
(posters visuels)
Détails multi-niveaux :sparklines, super-charts, sub-dashboards & figurés, glyphes
alertes, harvey balls
graphes combinés
Attributs pilotés par donnée : couleur, bordure, épaisseur, taille signifiantes
Graphes sur rail
Graphes sur plan
Graphes en réseau
Symboles
(images et affiches)
Cartes géo-graphiques
Diagrammes
(organigramme, ordinogramme, logigramme…)
Tables
(Classement et Matrice ordonnable
Texte
(Mise en page et calligrammes)
[Claude-Henri Mélédo, ALdecis]
58

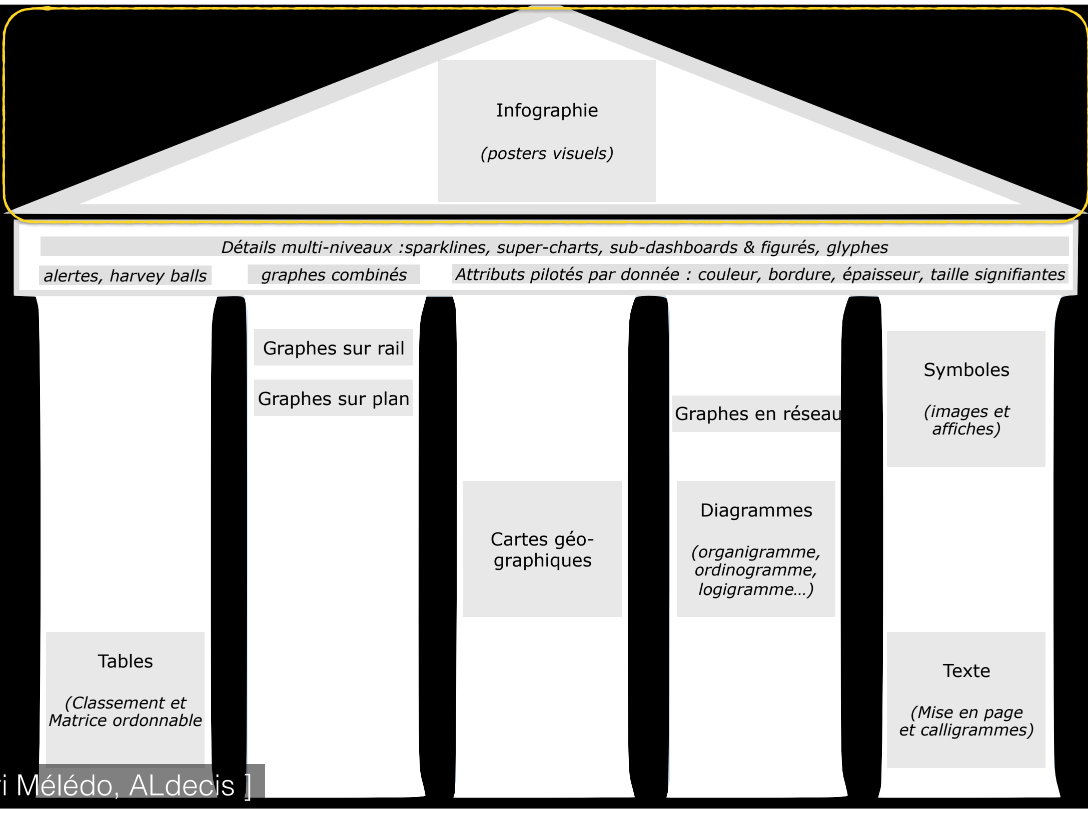
58

72 people



[www.visualnews.com]

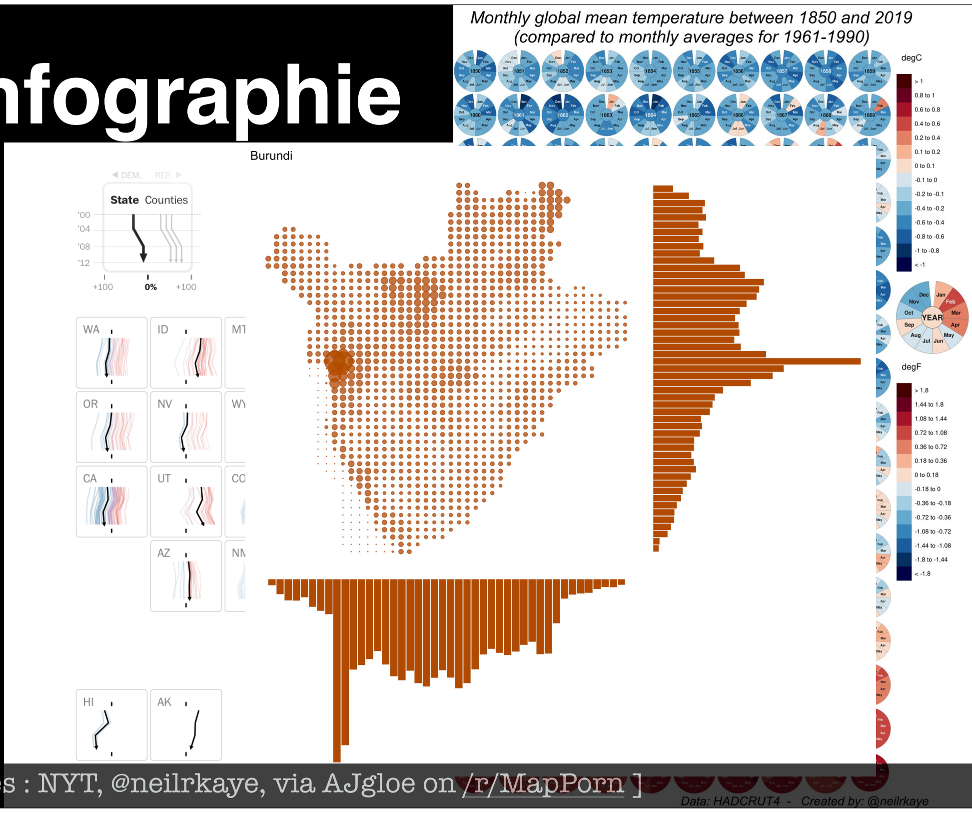
59



[Claude-Henri Mélédo, ALdecis]

60

Infographie



[Sources : NYT, @neilrkaye, via AJgloe on /r/MapPorn]

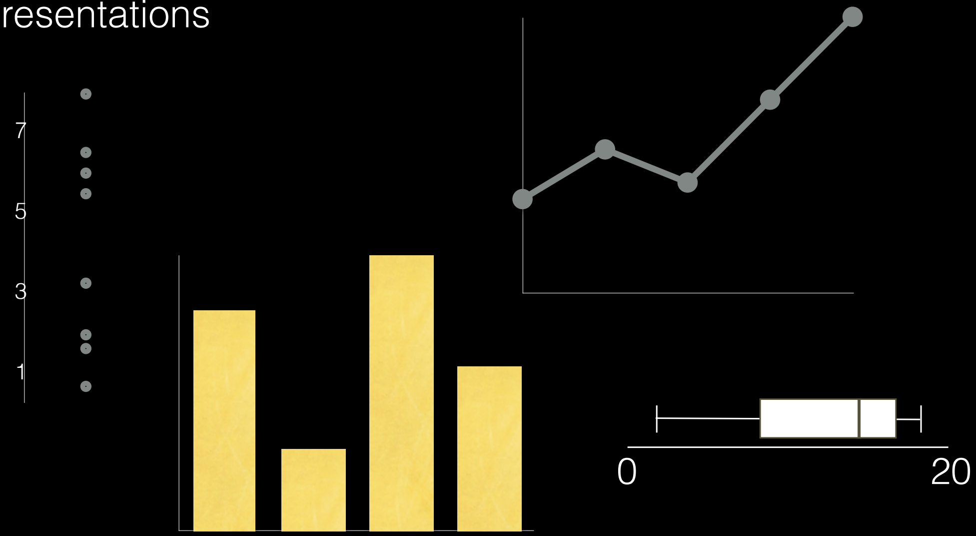
Back to Data

- What were the different types of data sets?
- Number of variables per class
  - 1 – Univariate data
  - 2 – Bivariate data
  - 3 – Trivariate data
  - >3 – Hypervariate data
- But also data types
  - Temporal
  - Geospatial
  - Relational (e.g. Trees, Graphs)
  - Textual

63

Univariate Data

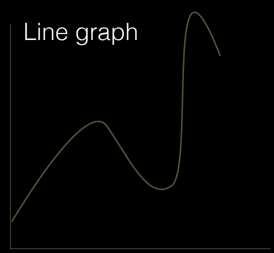
- Representations

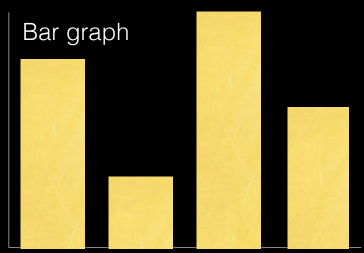


64

What Goes Where?

- In univariate representations, we often think of the data case as being shown along one dimension, and the value in another





- Y-axis is quantitative variable
- See changes over consecutive values

- Y-axis is quantitative variable
- Compare relative point values

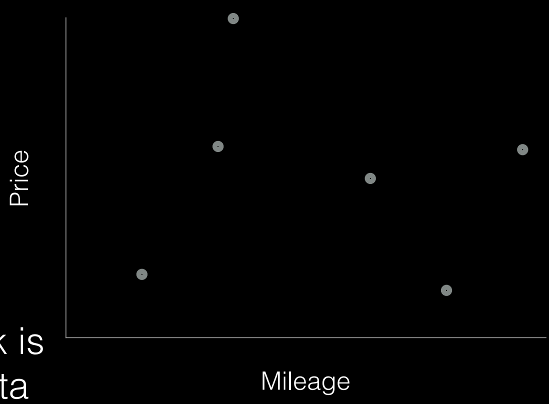
Alternative View

- We may think of graph as representing independent (data case) and dependent (value) variables
- Guideline:
  - Independent vs. dependent variables
    - Put independent on x-axis
    - See resultant dependent variables along y-axis

Bivariate Data

- Representations

Scatterplot is common

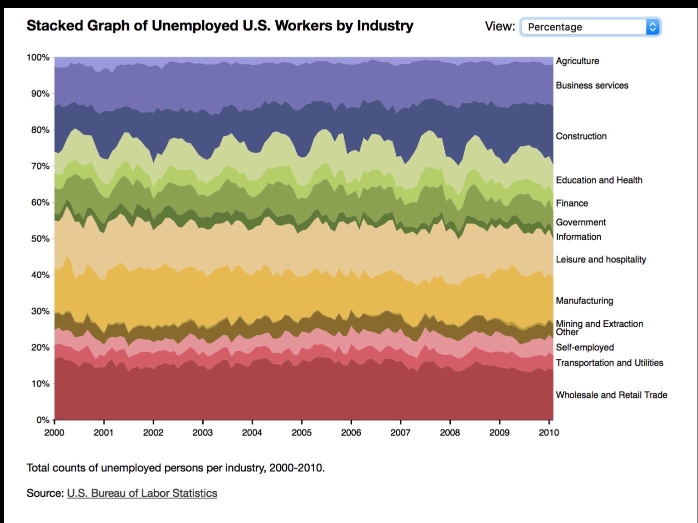


Each mark is now a data case

Two variables, want to see a relationship

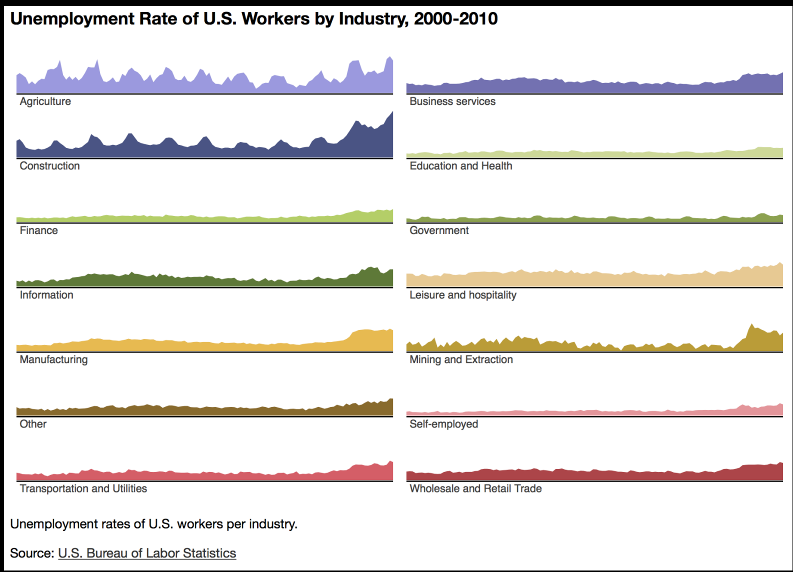
Is there a linear, curved, or random relationship?

Time Series



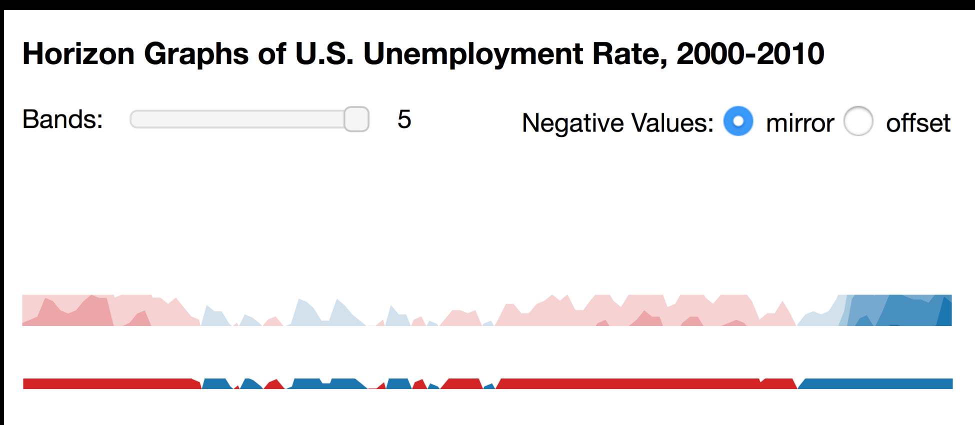
[Source: Heer et al., CACM 2010]

Time Series (Small multiples)



[Source: Heer et al., CACM 2010]

Horizon Graphs

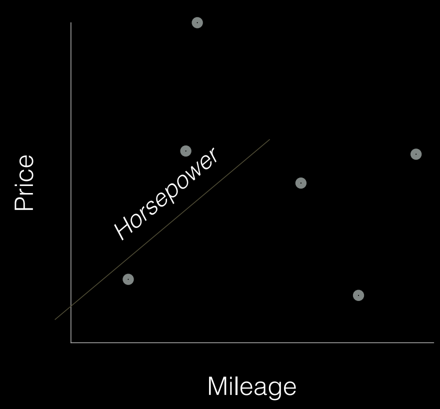


[Source: Heer et al., CACM 2010]

Trivariate Data

- Representations

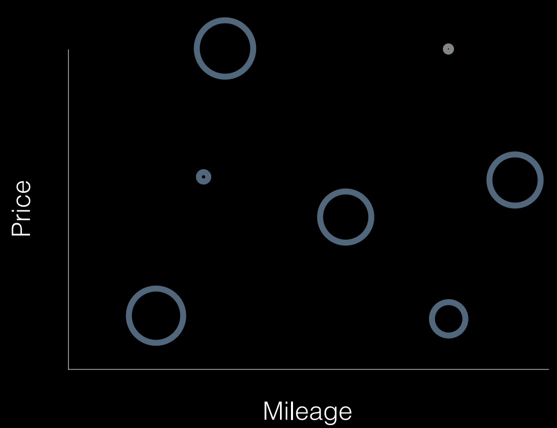
3D scatterplot is possible



71

Alternative Representation

- Representations

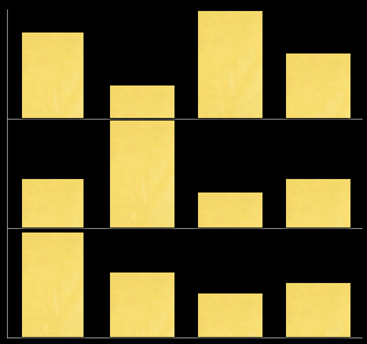


Still use 2D but have mark property represent third variable

72

Alternative Representation

- Representations



Represent each variable
in its own explicit way

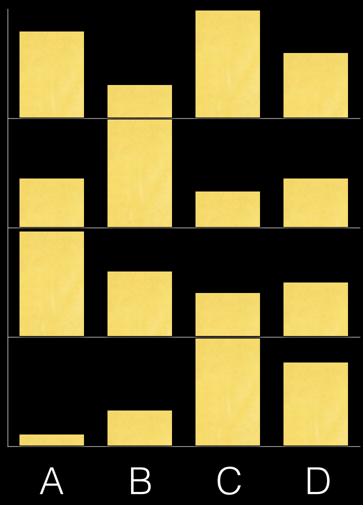
Hypervariate Data

- Ahhh, the tough one
- Number of well-known visualization techniques exist for data sets of 1-3 dimensions
    - line graphs, bar graphs, scatter plots OK
    - We see a 3-D world (4-D with time)
- What about data sets with more than 3 variables?
    - Often the interesting, challenging ones

Multiple Views

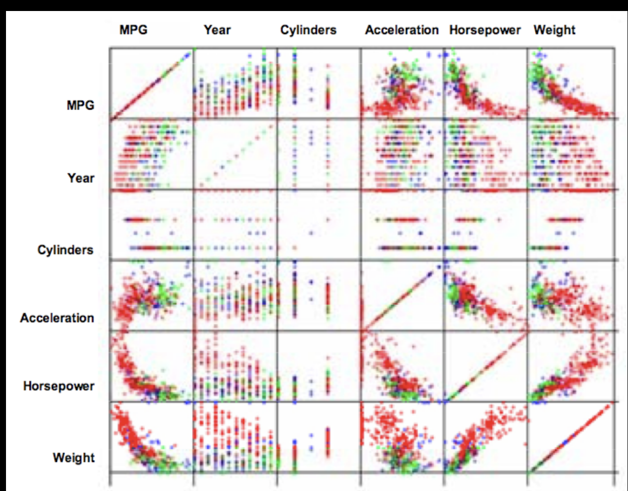
- Give each variable its own display

	A	B	C	D
1	8	3	10	6
2	4	9	3	4
3	10	6	4	5
4	1	3	9	7



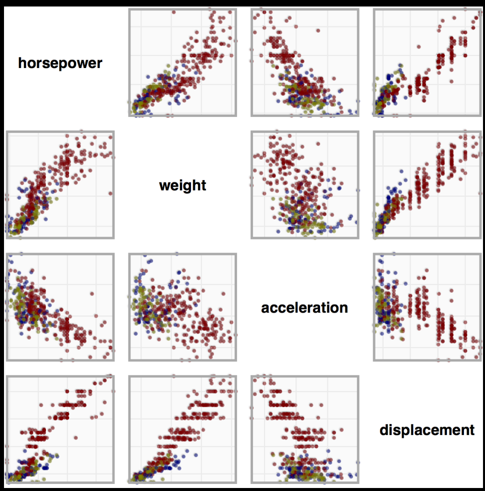
Scatterplot Matrix

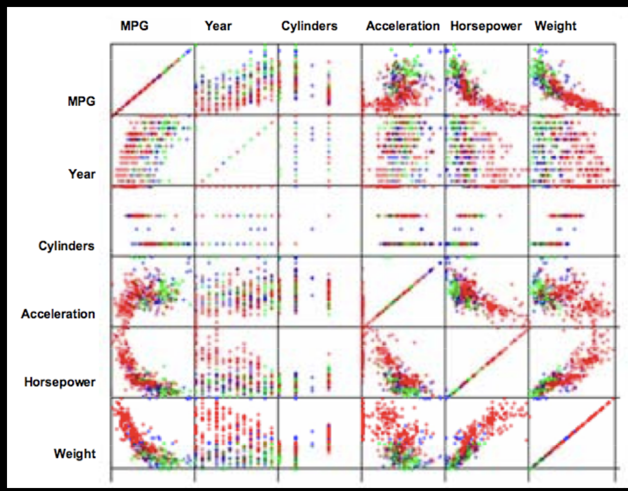
- Represent each possible pair of variables in their own 2-D scatterplot



- Useful for what?
- Misses what?

Scatterplot Matrix





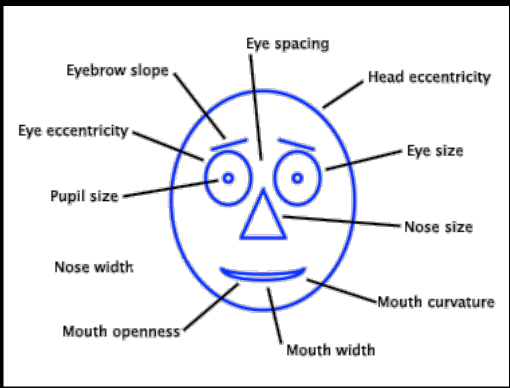
77

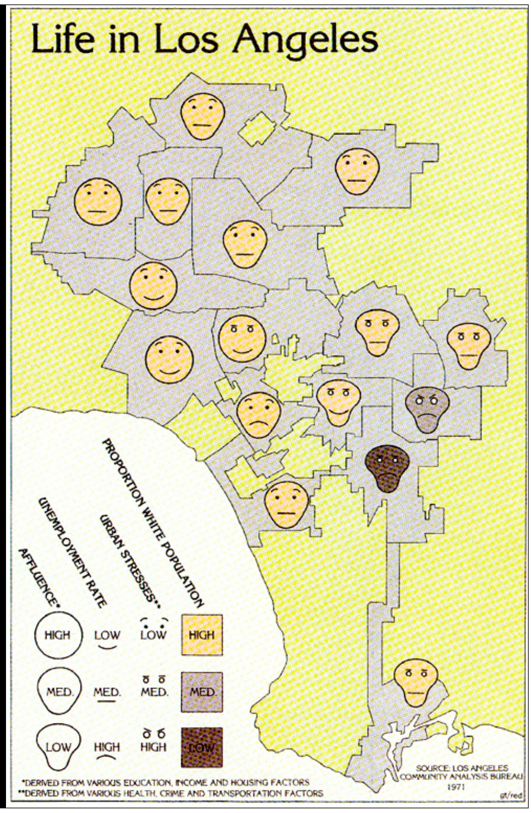
Small Multiples

[FT, 27/04/2020]

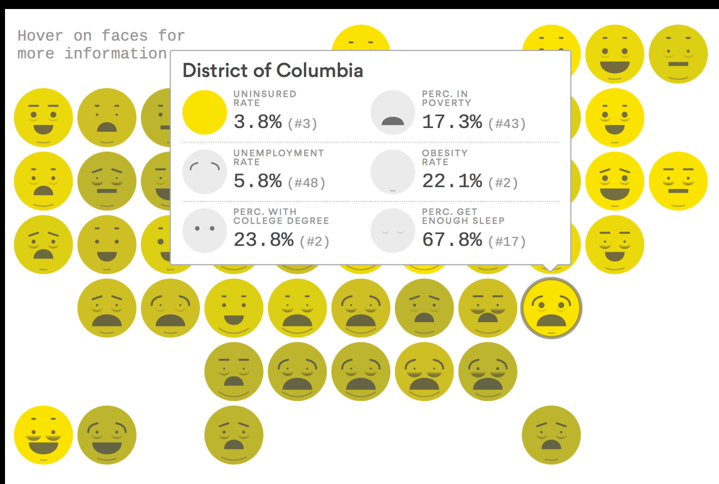
Chernoff Faces

- Encode different variables' values in characteristics of human face





Emoji Chernoff Faces



[https://www.axios.com/an-emoji-built-from-data-for-every-state-2408885674.html]

So far…

- We examined a number of tried-and-true techniques/visualizations for presenting multivariate (typically ≤3) data sets
    - Hinted at how to go above 3 dimensions

81

More Dimensions

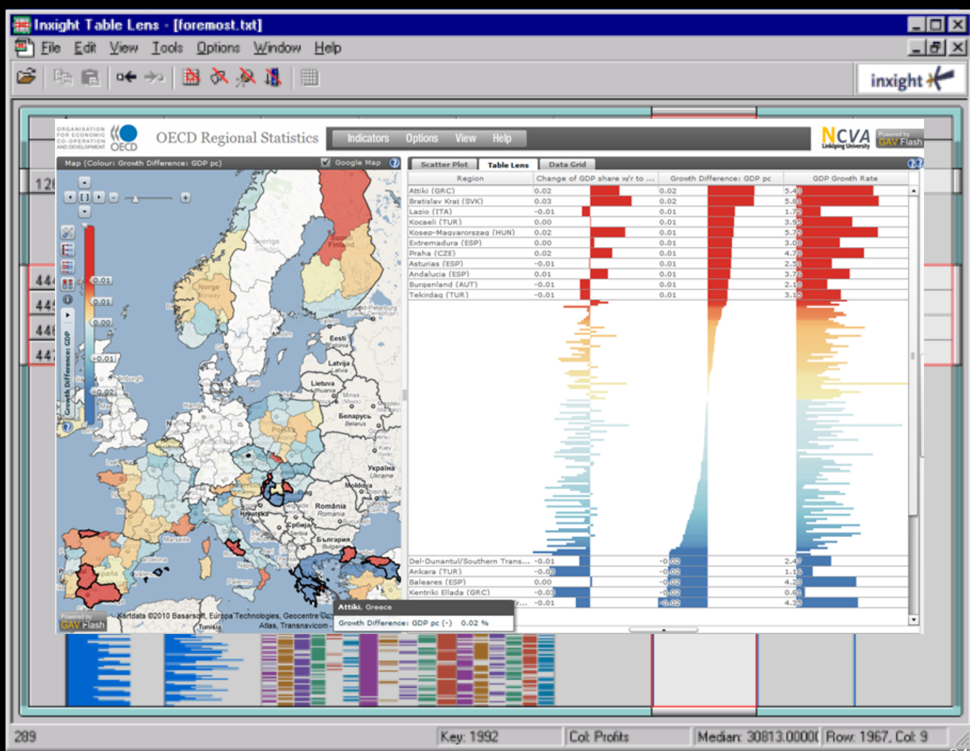
- Fundamentally, we have 2 display dimensions
- For data sets with >2 variables, we must project data down to 2D
- Come up with visual mapping that locates each dimension into 2D plane
- Computer graphics 3D→2D projections

82

Wait a moment...

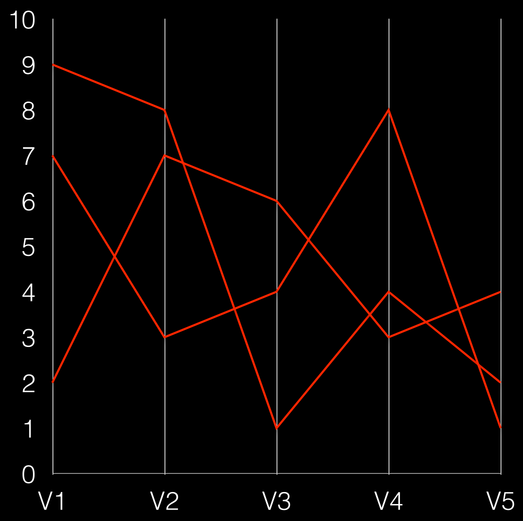
- A spreadsheet already does that
  - Each variable is positioned into a column
  - Data cases in rows
  - This is a projection (mapping)

Table Lens



[Rao & Card, CHI '94]

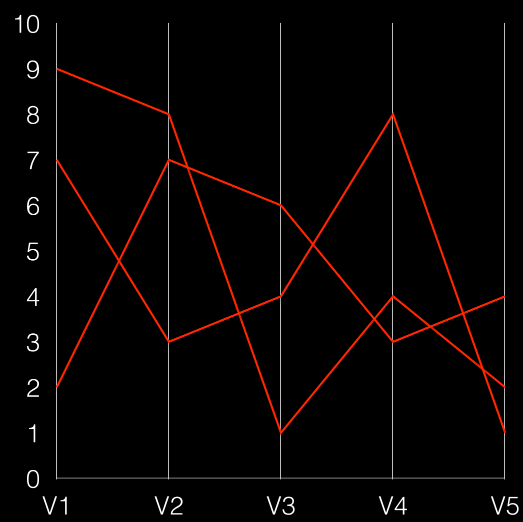
Parallel Coordinates



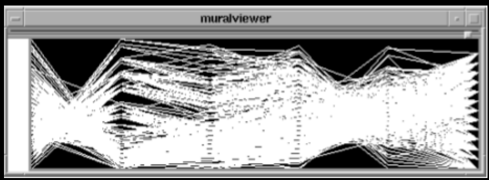
	V1	V2	V3	V4	V5
C1	7	3	4	8	1
C2	2	7	6	3	4
C3	9	8	1	4	2

Parallel Coordinates

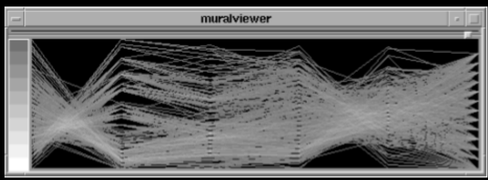
- Encode variables along a horizontal row
- Vertical line specifies different values that variable can take
- Data point represented as a polyline



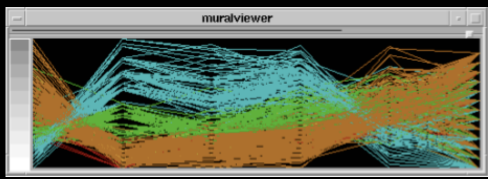
Parallel Coordinates Example



Basic



Greyscale

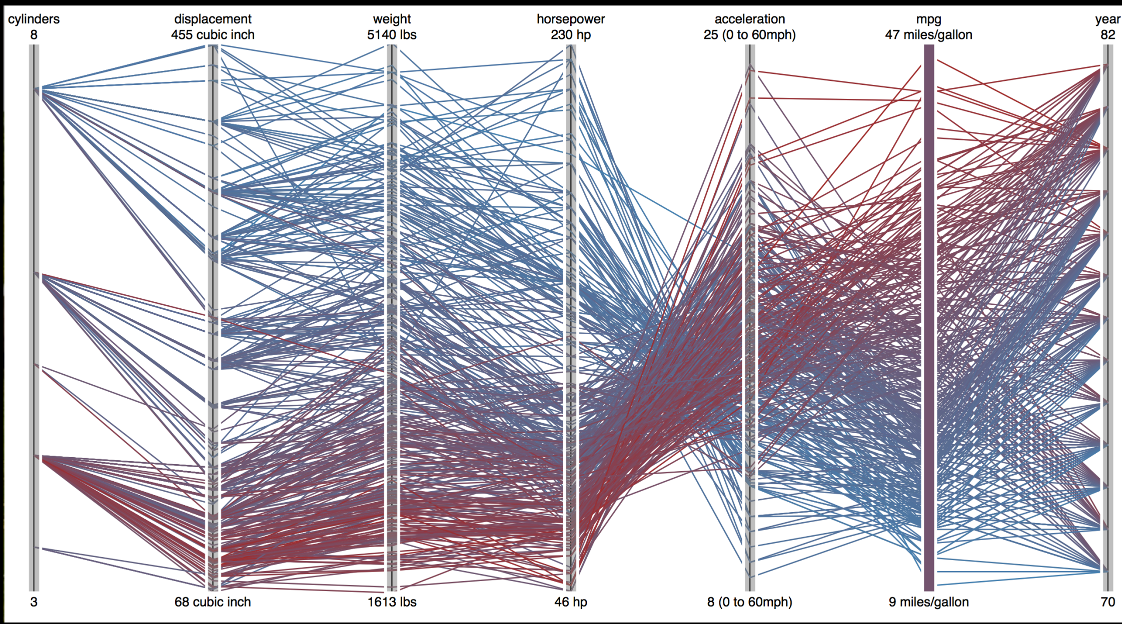


Color

Issue

- Different variables can have values taking on quite different ranges
- Must normalize all down (e.g., f(x): $\mathbb{N} \rightarrow [0, 1]$)

Parallel Coordinates



[Source: Heer et al., CACM 2010]

Example

- VLSI chip manufacture
- Want high quality chips (high speed) and a high yield batch (% of useful chips)
- Able to track defects
- Hypothesis: No defects gives desired chip types
- 473 batches of data

The Data

- 16 variables
- X1 — yield
- X2 — quality
- X3–X12 - # defects (inverted)
- X13–X16 - physical parameters

91

Parallel Coordinates

yield & quality

defects

parameters

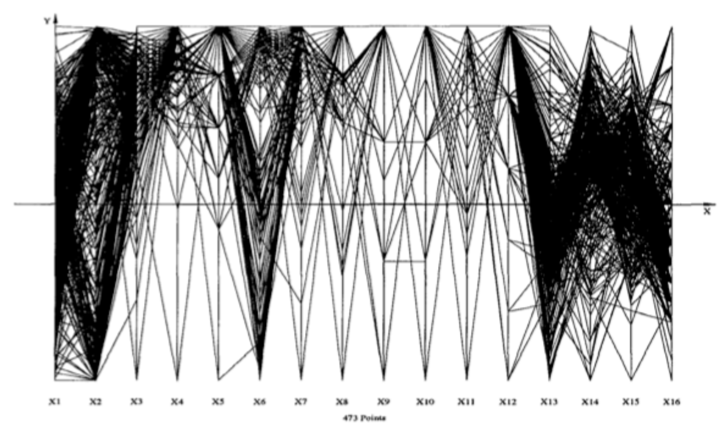


Figure 1: The full dataset consisting of 473 batches

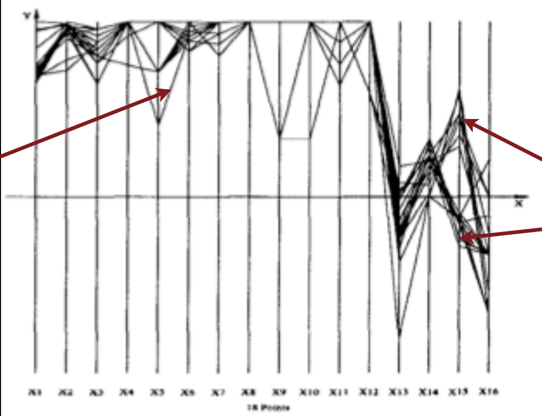
Yikes!
But not
that bad

Distributions
x1 - normal
x2 - bipolar

92

Top Yield & Quality

defects



split

Figure 2: The batches high in Yield, $X1$, and Quality, $X2$.

93

Minimal Defects

Not the highest yields and quality

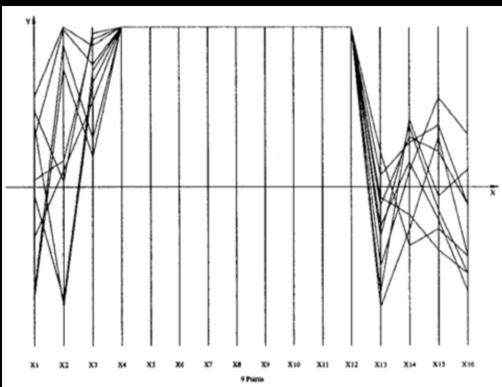


Figure 3: The batches with zero in 9 out of the ten defect types.

94

Best Yields

Appears that
some defects
are necessary
to produce
the best chips

Non-intuitive!

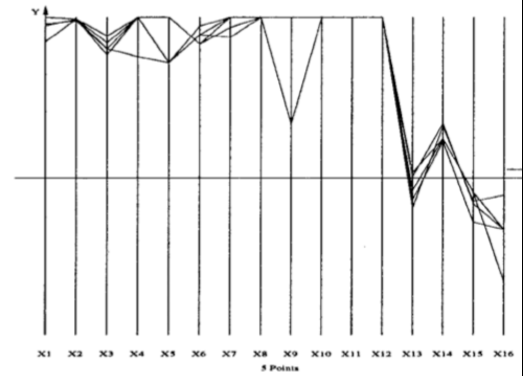
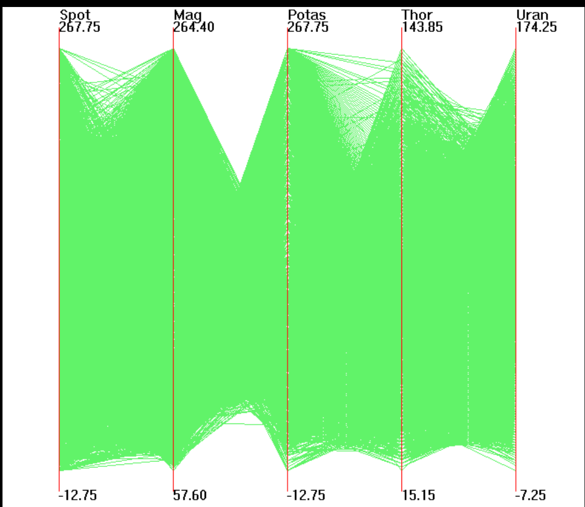


Figure 6: Batches with the highest Yields do not have the lowest defects in $X3$ and $X6$.

95

Challenges

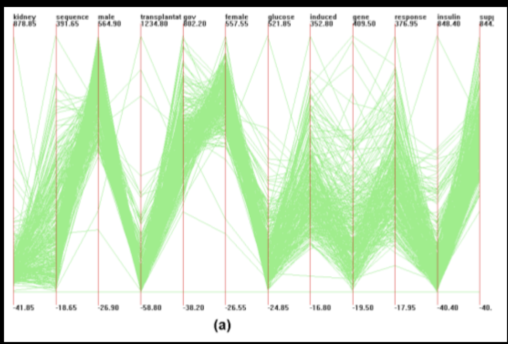


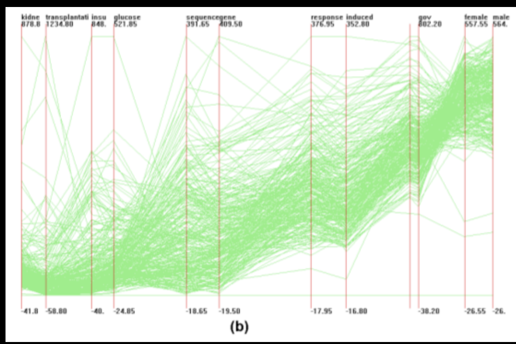
Out5d dataset (5 dimensions, 16384 data items)

96

Dimensional Reordering

Which dimensions are most like each other?





Same dimensions ordered by similarity

97

Dimensional Reordering

Can you reduce clutter and highlight other interesting features in data?

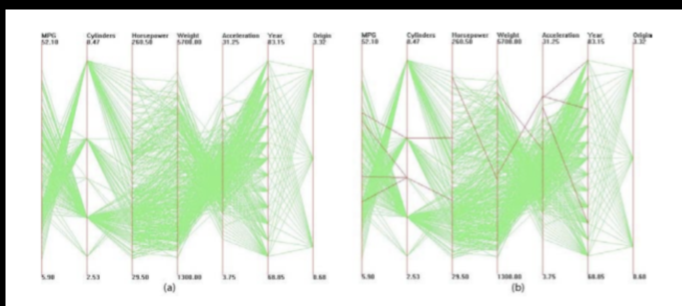
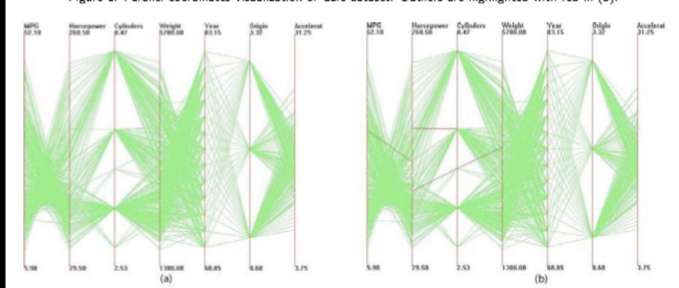
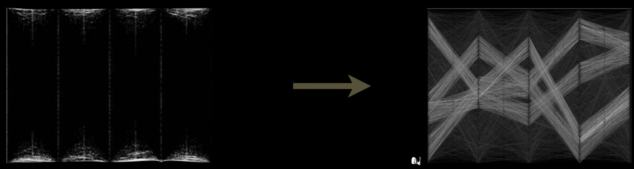


Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).

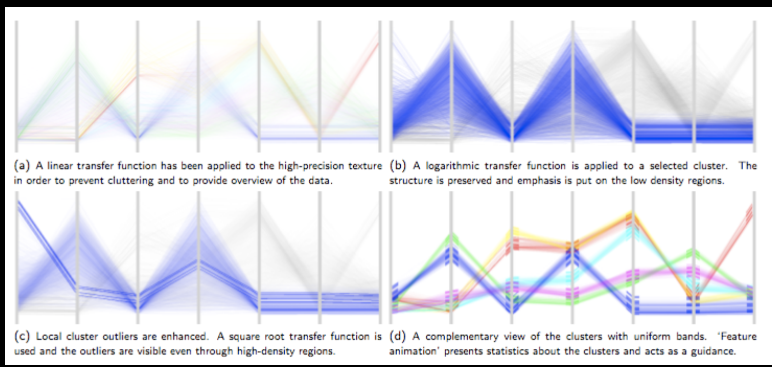


98

Reducing Density



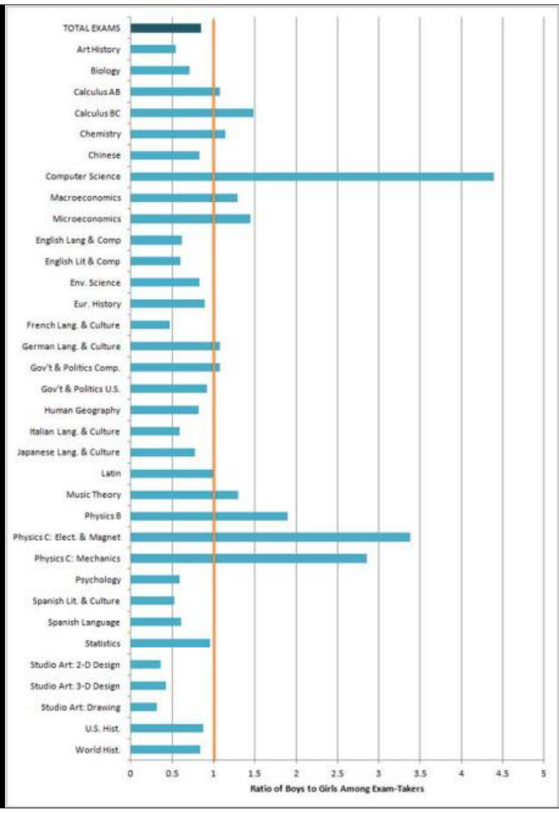
[Artero et al, InfoVis 2004]



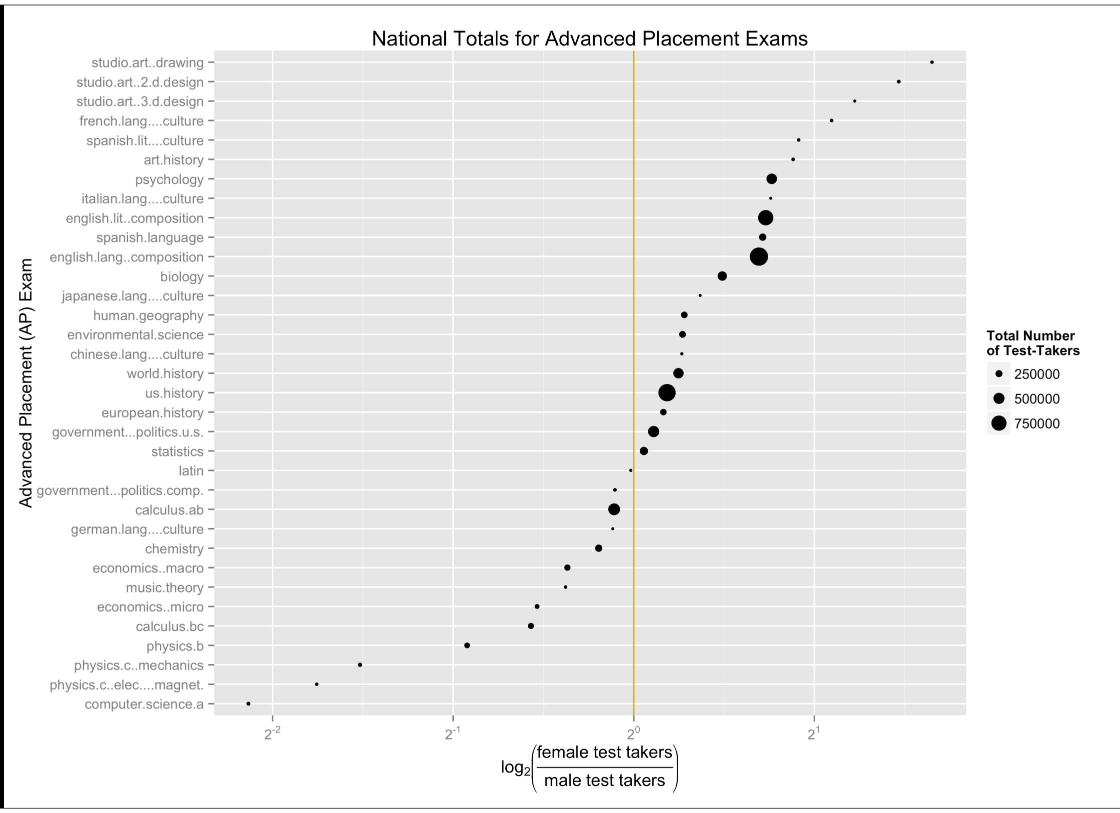
[Johansson et al, InfoVis 2005]

Critique du moment

Ratio of boys to girls among exam takers



101



102

Next time: Tasks