

Visual Mappings

James EAGAN



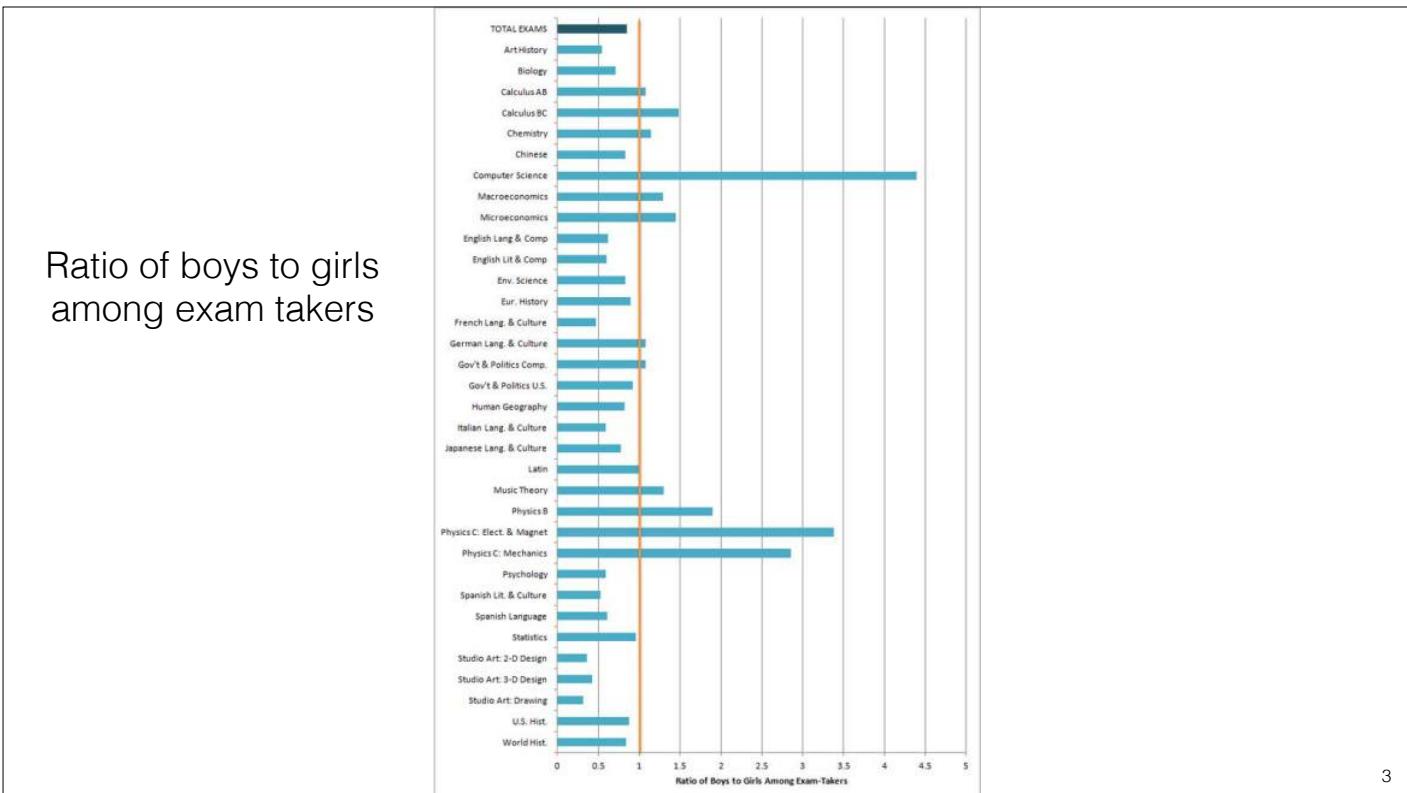
Includes slides adapted from John Stasko
(Georgia Tech), Petra Isenberg & Jean-Daniel
Fekete (INRIA), Chris North (Virginia Tech), Tamara
Munzner (UBC)



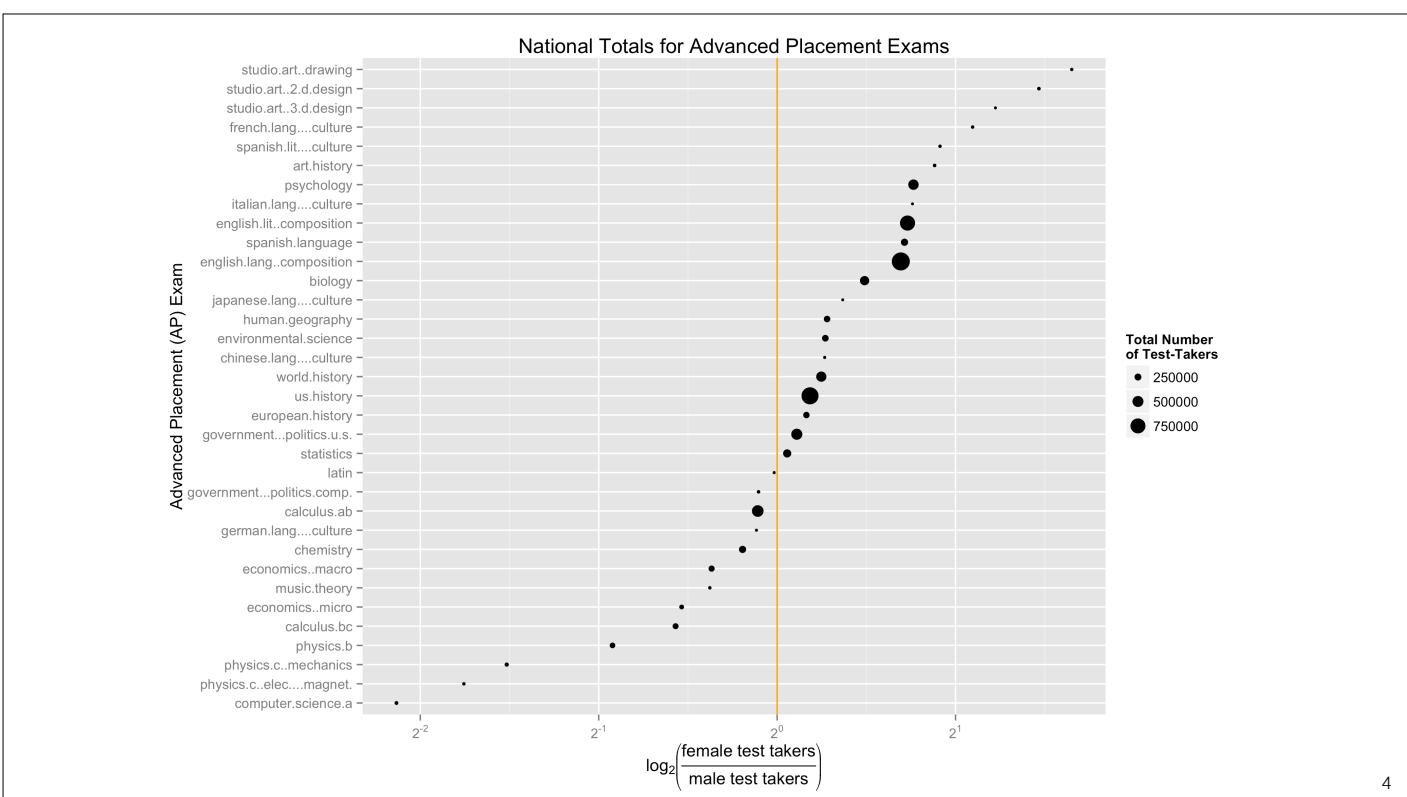
Dernière mise à jour : avril 2019

1

Critique du moment

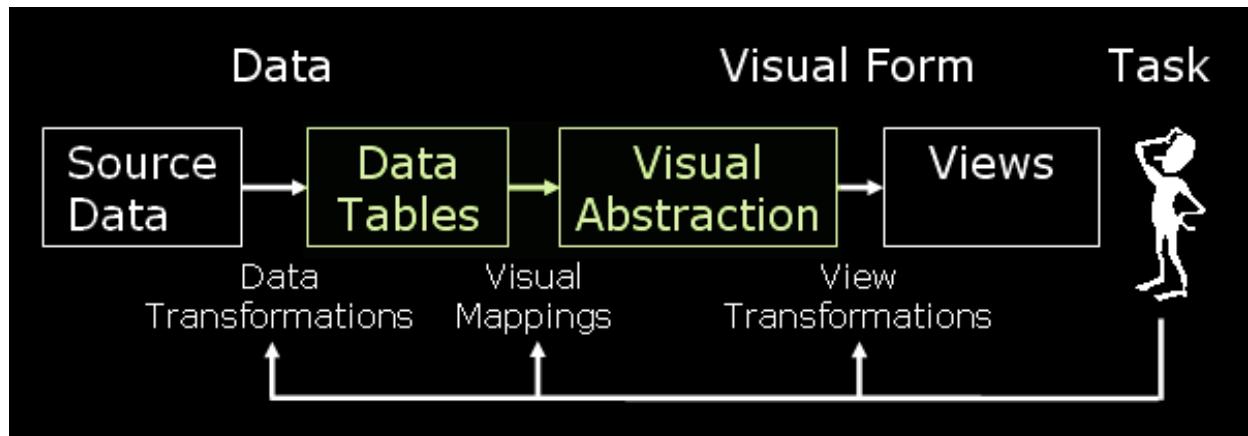


3



4

InfoVis Pipeline



5

5

Map *data* to a *representation*

Data is abstract

Representation is more conceptual

Use a space

Create an *implantation* of *data* into the space

6

6

Visual Structures

Composed of

Spatial substrate

Marks

Graphical properties of marks

[Bertin, Sémiologie Graphique 1967]

7

Space

Visually dominant

Often put axes on space to assist

Use techniques of
composition, alignment, folding,
recursion, overloading to

increase use of space

do data encodings

8

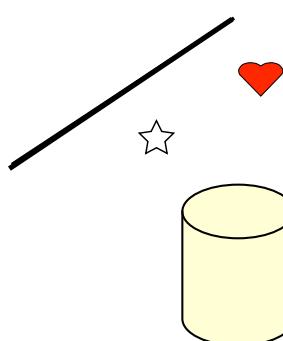
Marks

Things that occur in space

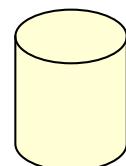
Points



Lines



Areas



Volumes

9

9

Graphical Properties

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

10

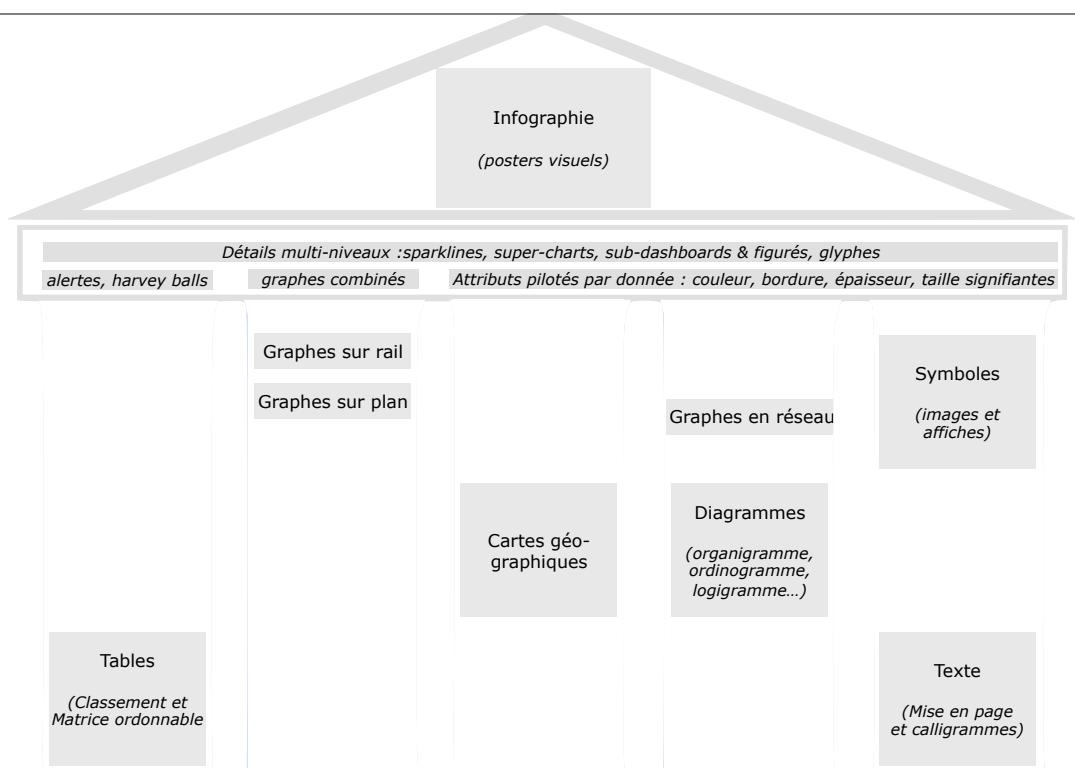
10

There are only 5 graphics

(sort of...)

11

11



12

12

Tables

Microsoft Excel - baseball

File Edit View Insert Format Tools Data Accounting Window Help

Arial 10 B I U \$, .00 .00 100% ?

A1 = Name

	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	Career Hits
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 Galarraga	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 Slyk	418	113	13	48	81	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 Biancalca	190	46	2	24	8	15	5	479	102
21	Bruce Bochte	407	104	6	57	43	65	12	5233	1478

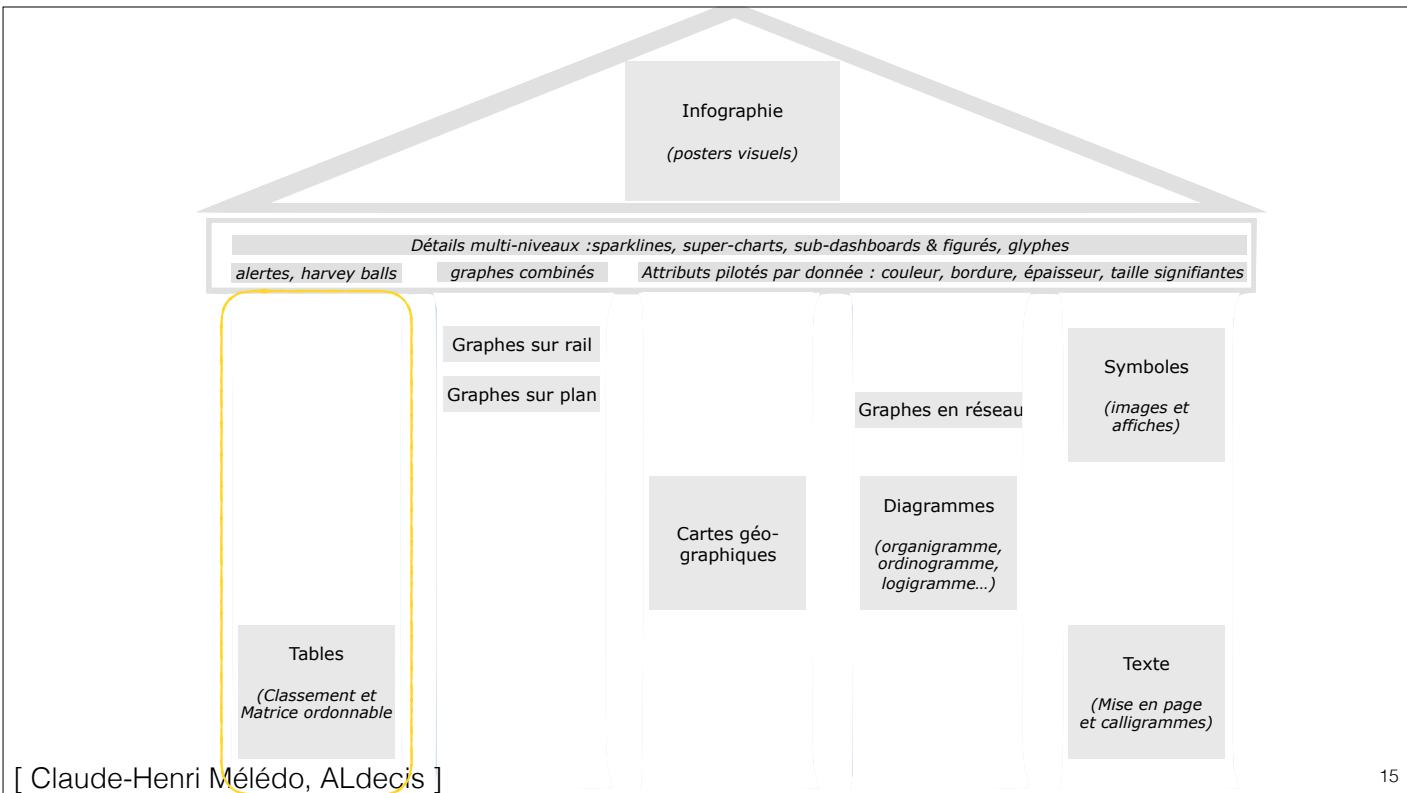
13

13

Bullet Train Schedule

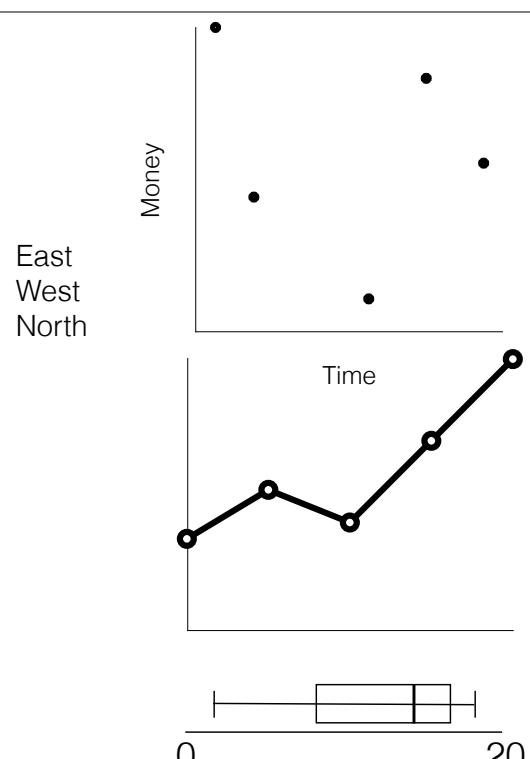
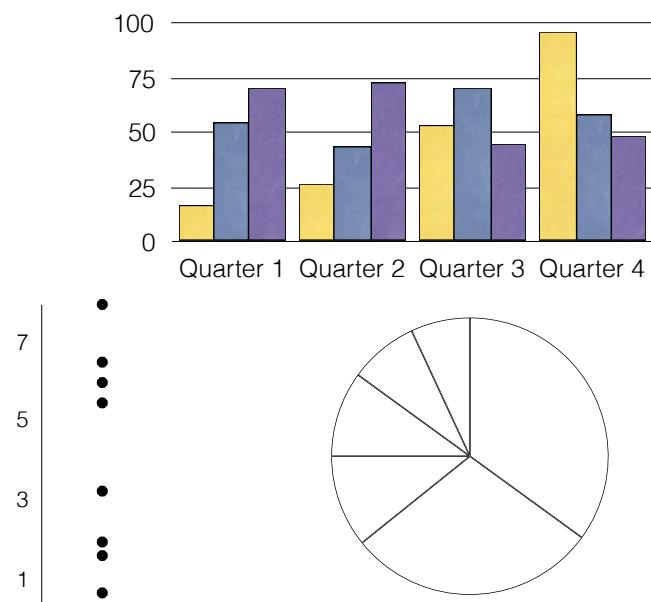


14



15

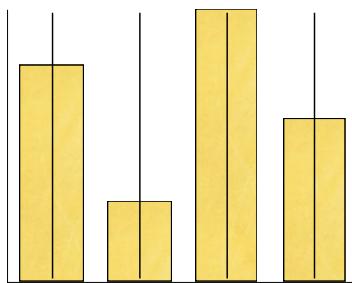
Graphs on rails



16

16

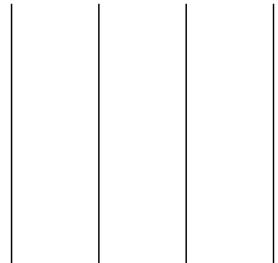
Graphs on rails



17

17

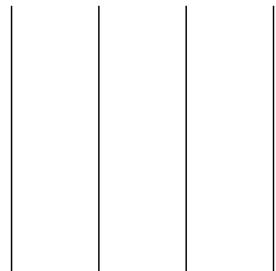
Graphs on rails



18

18

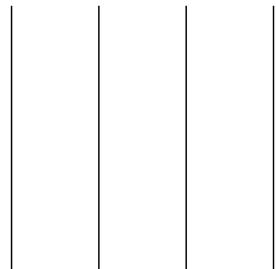
Graphs on rails



19

19

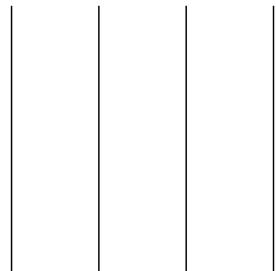
Graphs on rails



20

20

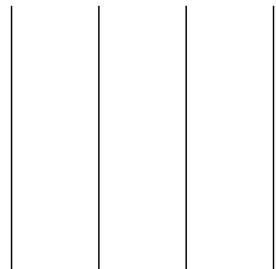
Graphs on rails



21

21

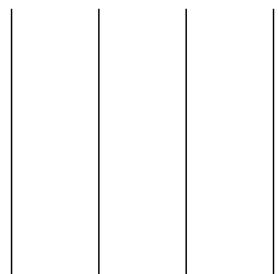
Graphs on rails



22

22

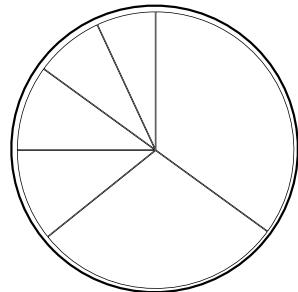
Graphs on rails



23

23

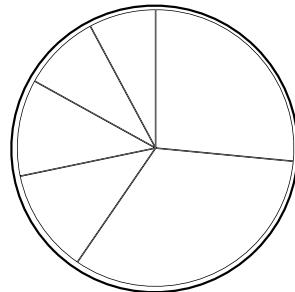
Graphs on rails



24

24

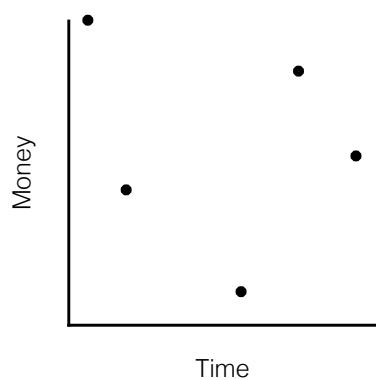
Graphs on rails



25

25

Graphs on the plane

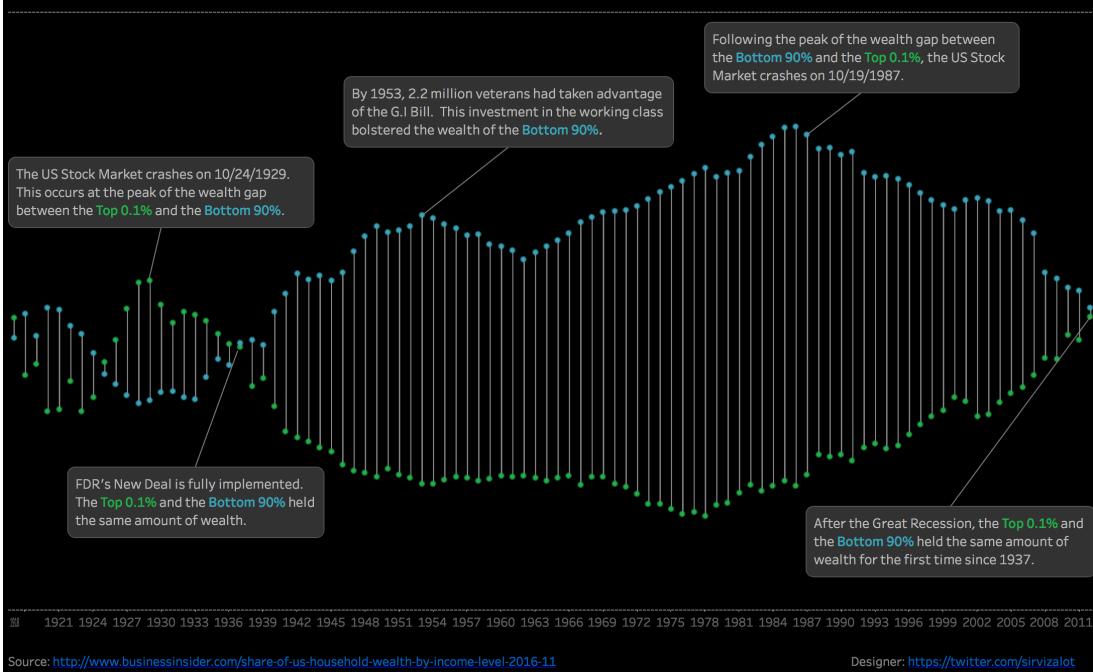


26

26

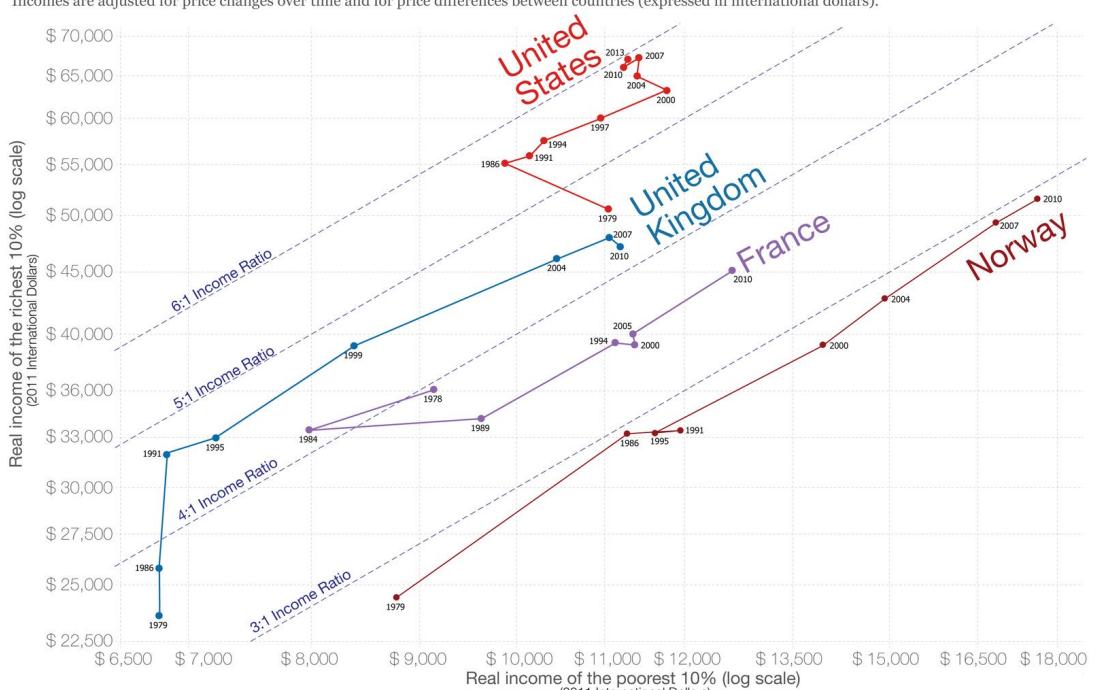
The Wealth Gap

A historical view of wealth ownership within the **Top 0.1%** and the **Bottom 90%** of US households



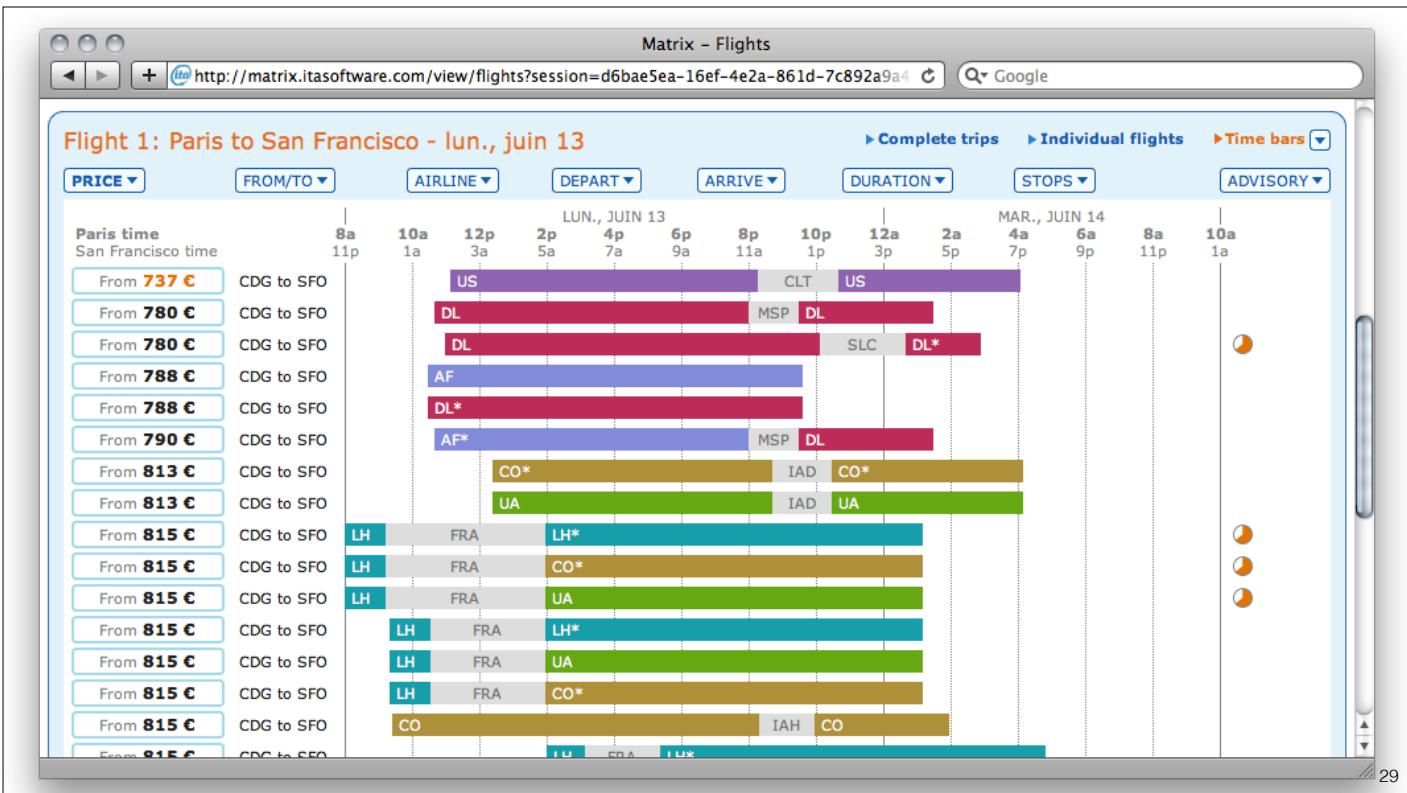
27

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).

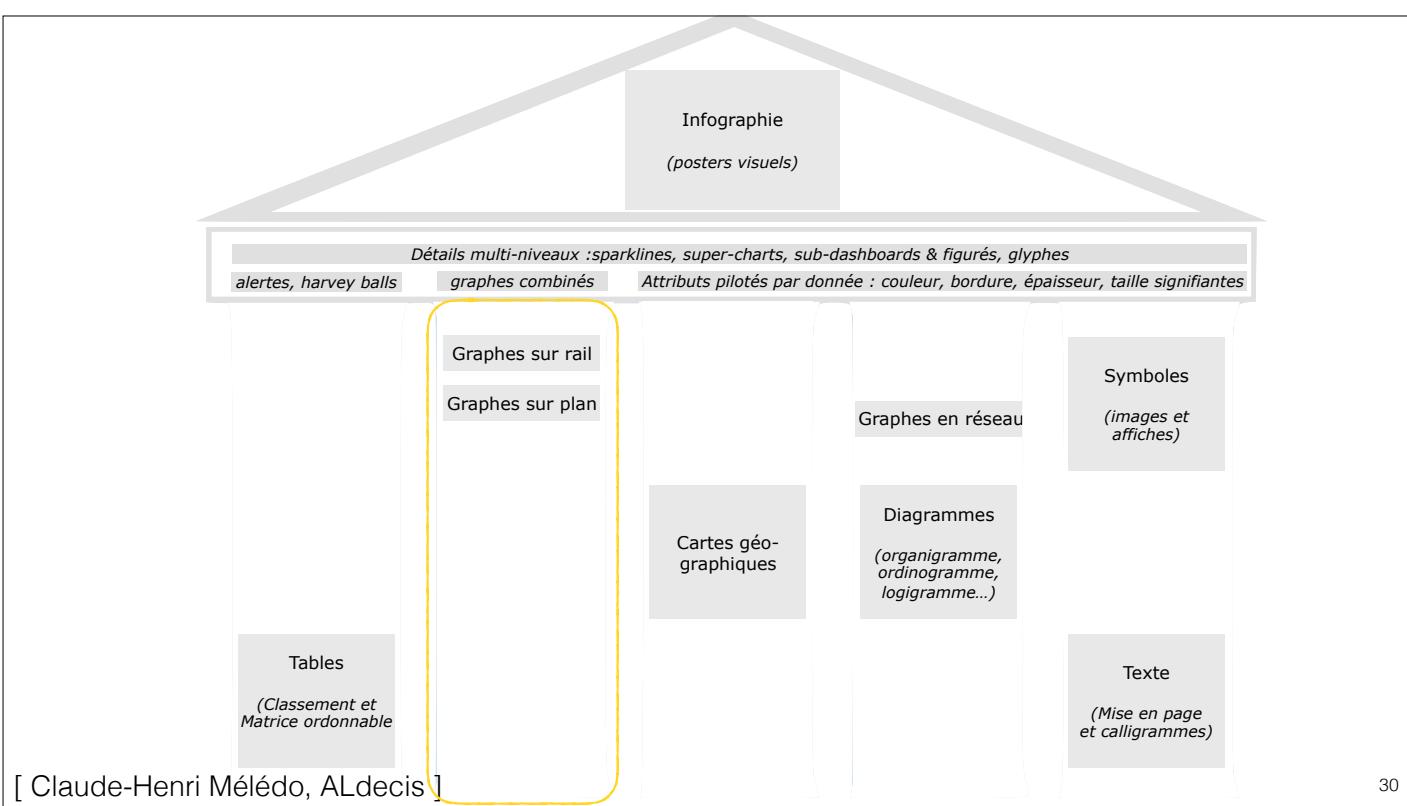


[Max Roser]

28

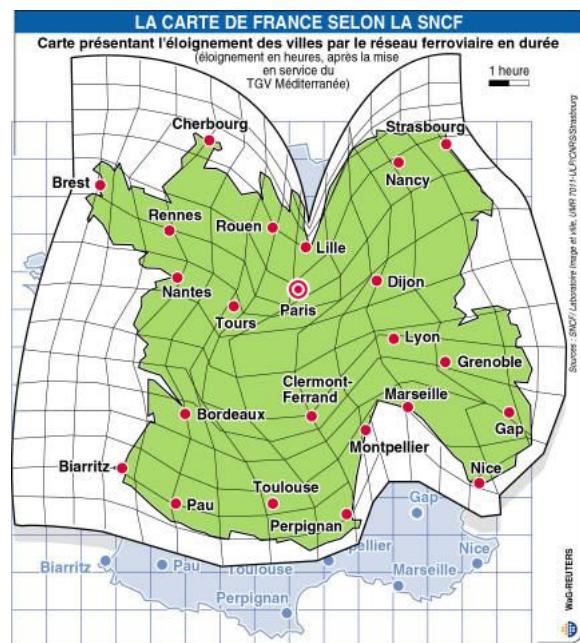
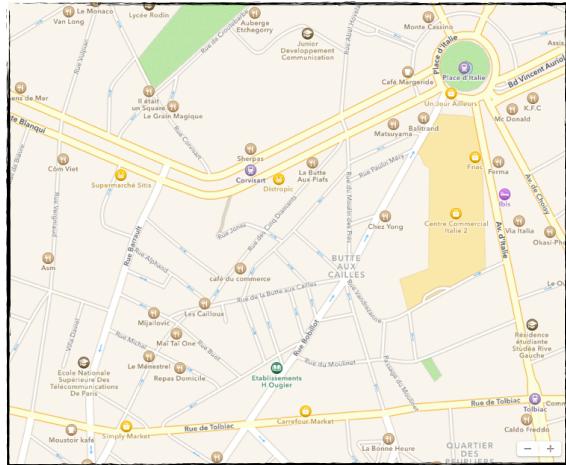


29



30

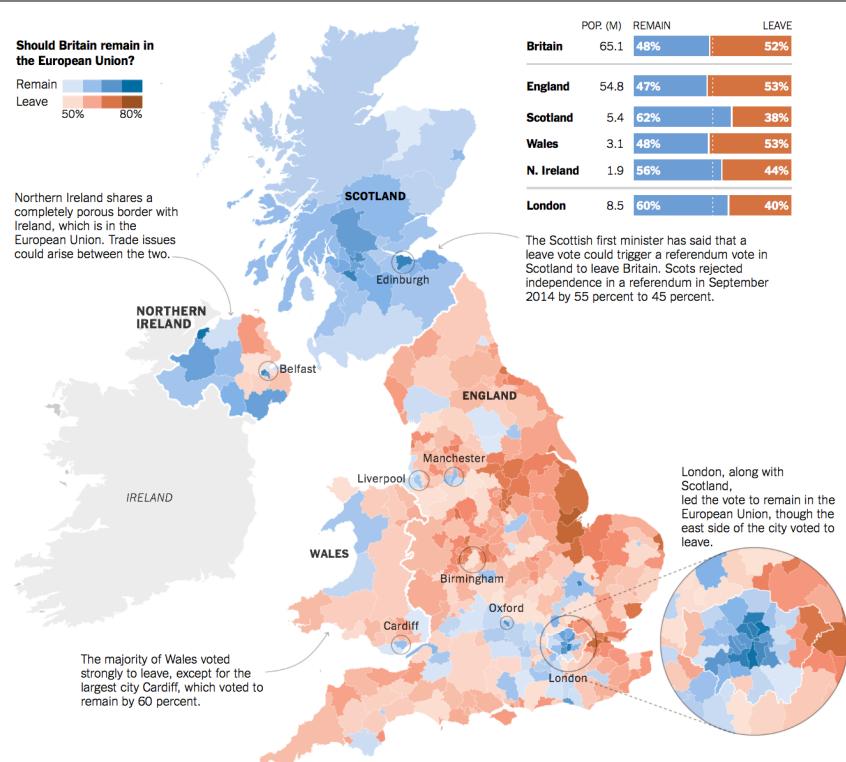
Geospatial maps



31

31

Brexit



32

32

E.U.



33

33

Country Music



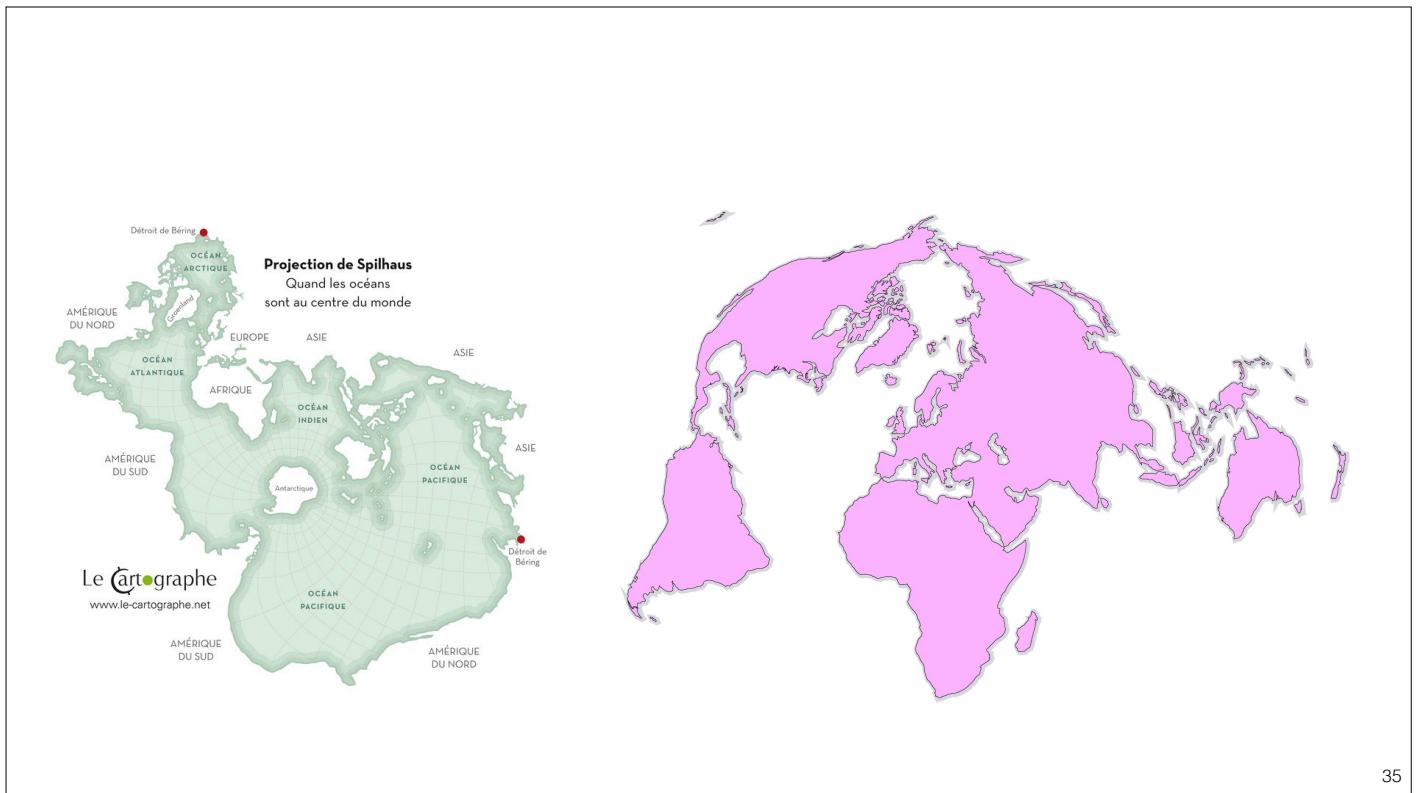
Figure 14. States Mentioned in Country-Music Lyrics

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.

34

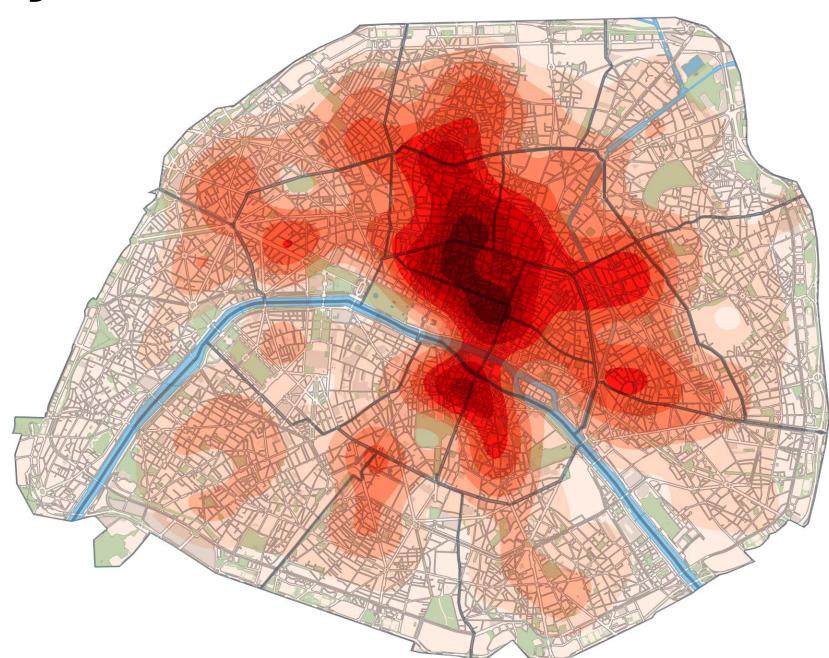
34



35

35

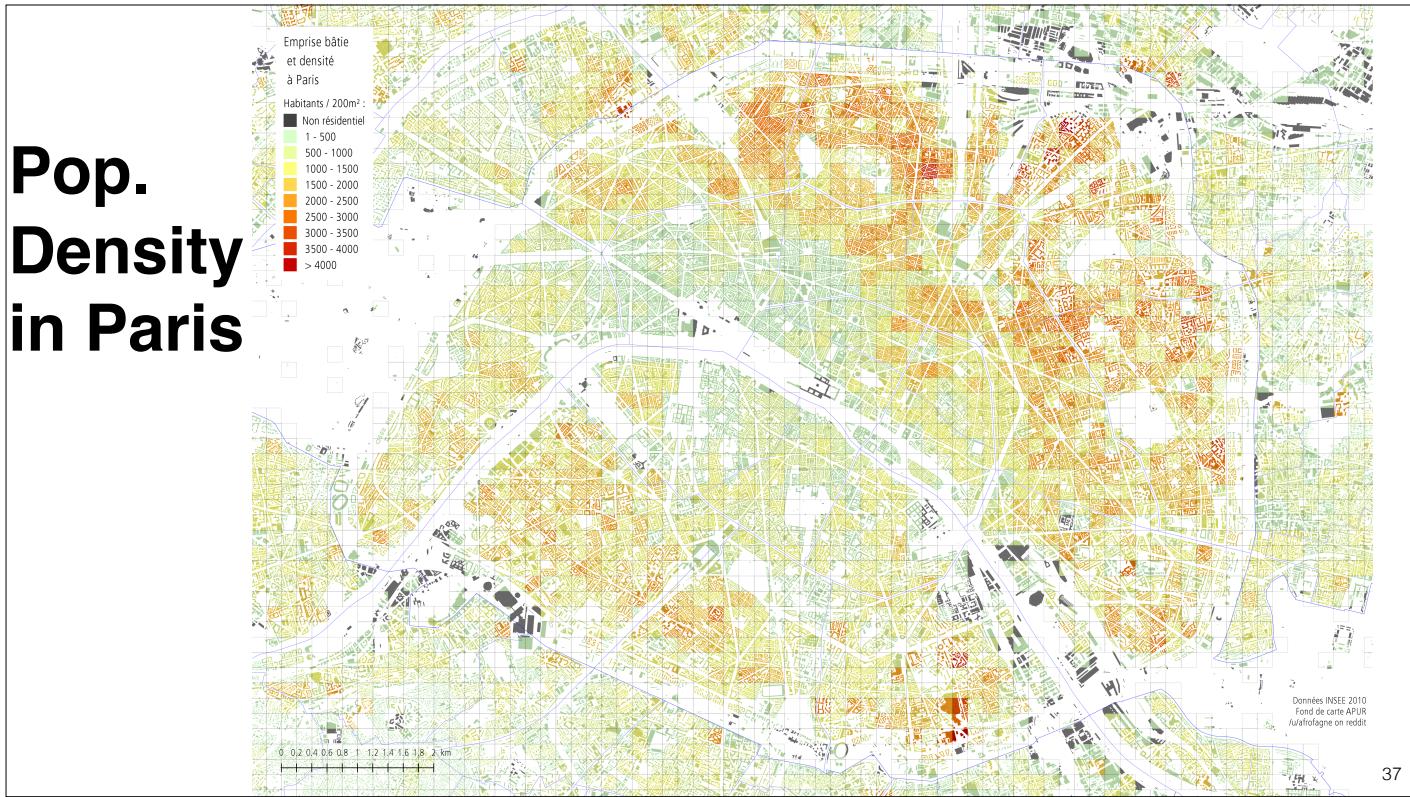
Density of bars & restaurants in Paris



36

36

Pop. Density in Paris



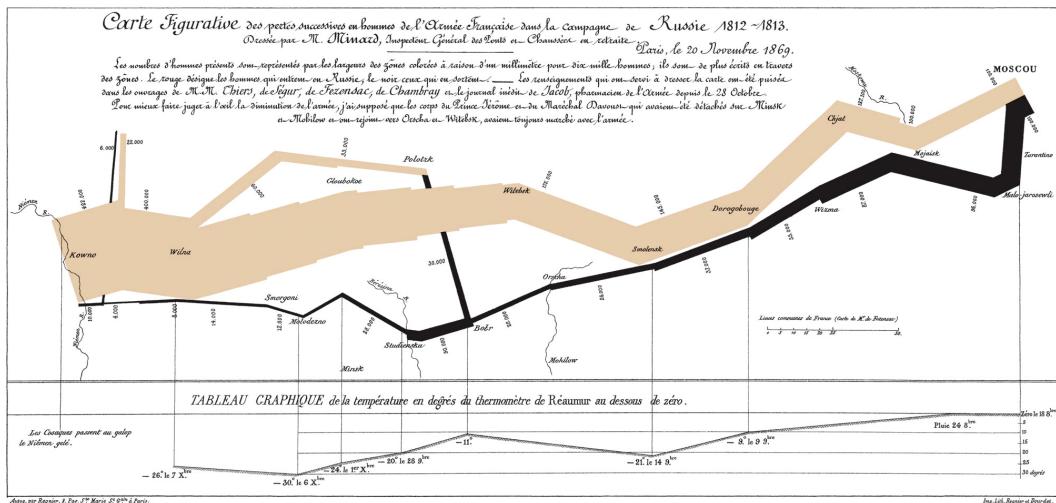
37



38

From E. Tufte
The Visual Display of Quantitative Information

Napoleons's March



Graphic by Minard

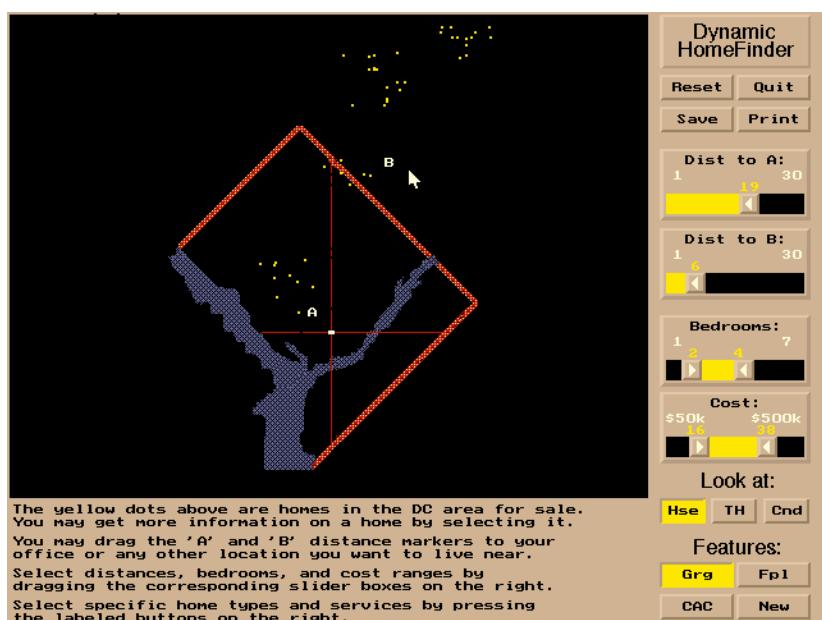
size of army
direction

latitude
longitude temperature
date

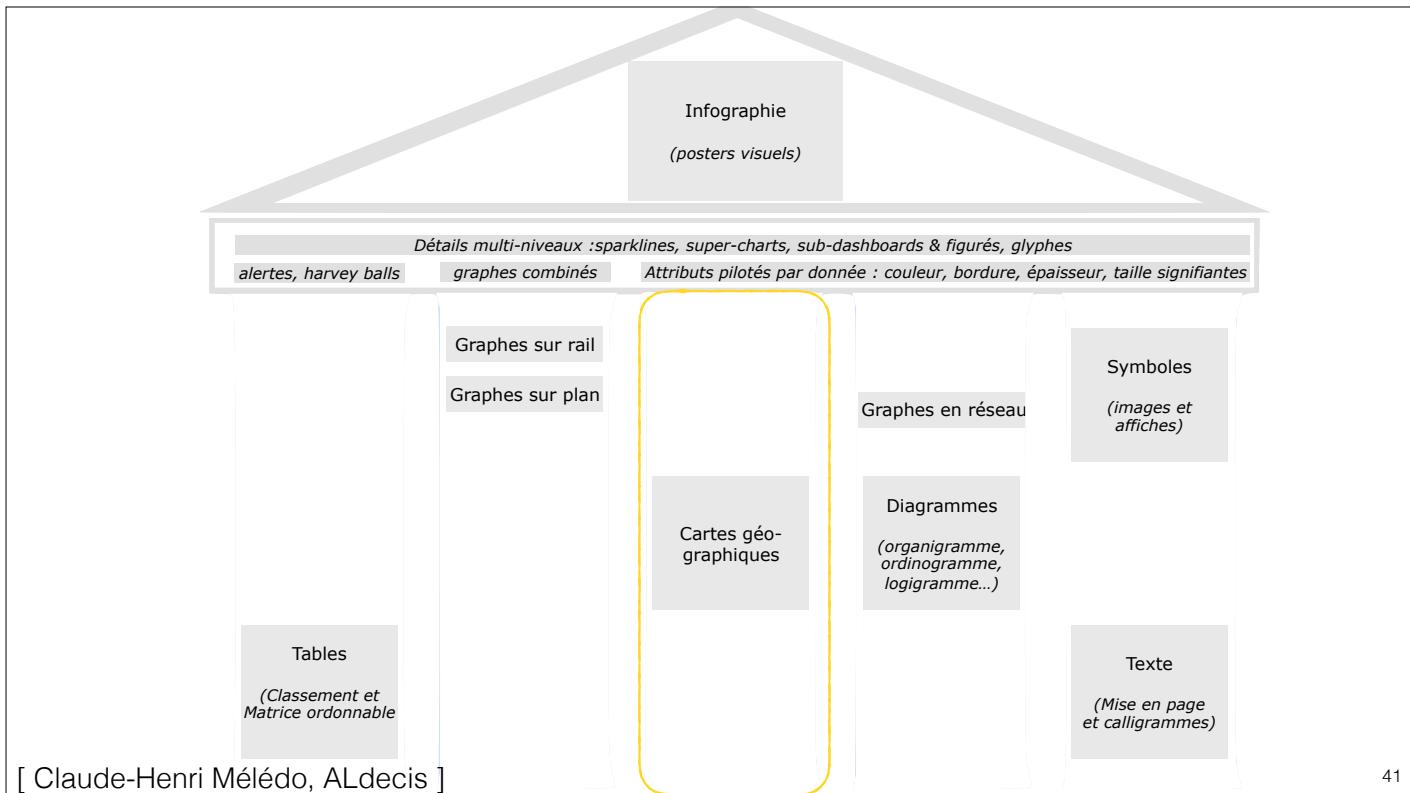
39

HomeFinder

HCIL
U of
Maryland

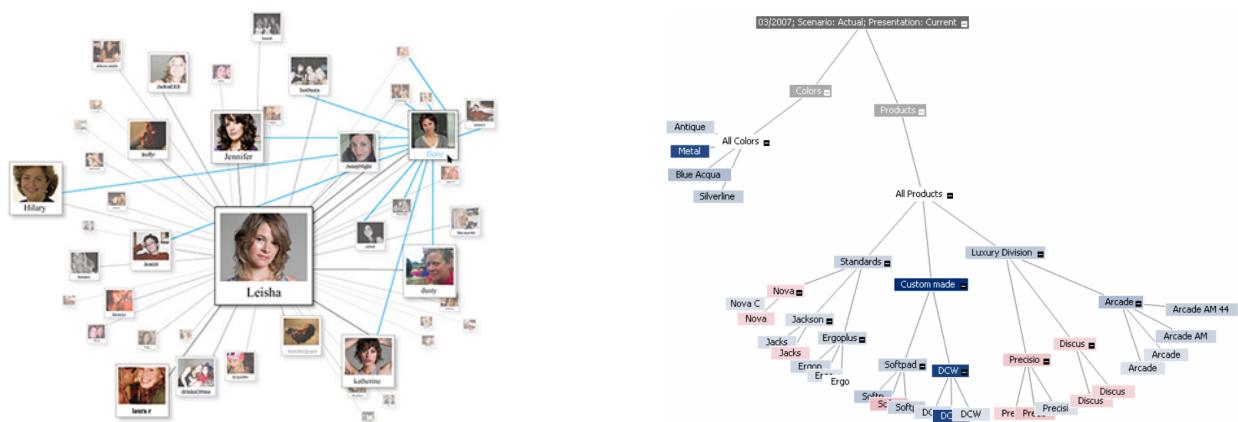


40



41

Networks & Diagrams

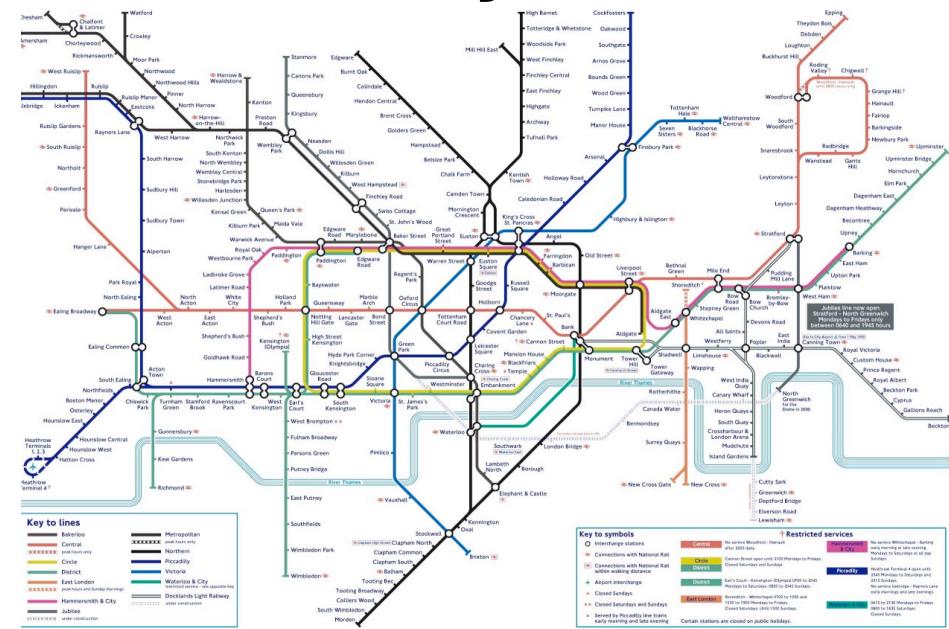


42

42

London Subway

www.thetube.com



43

True Geography



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

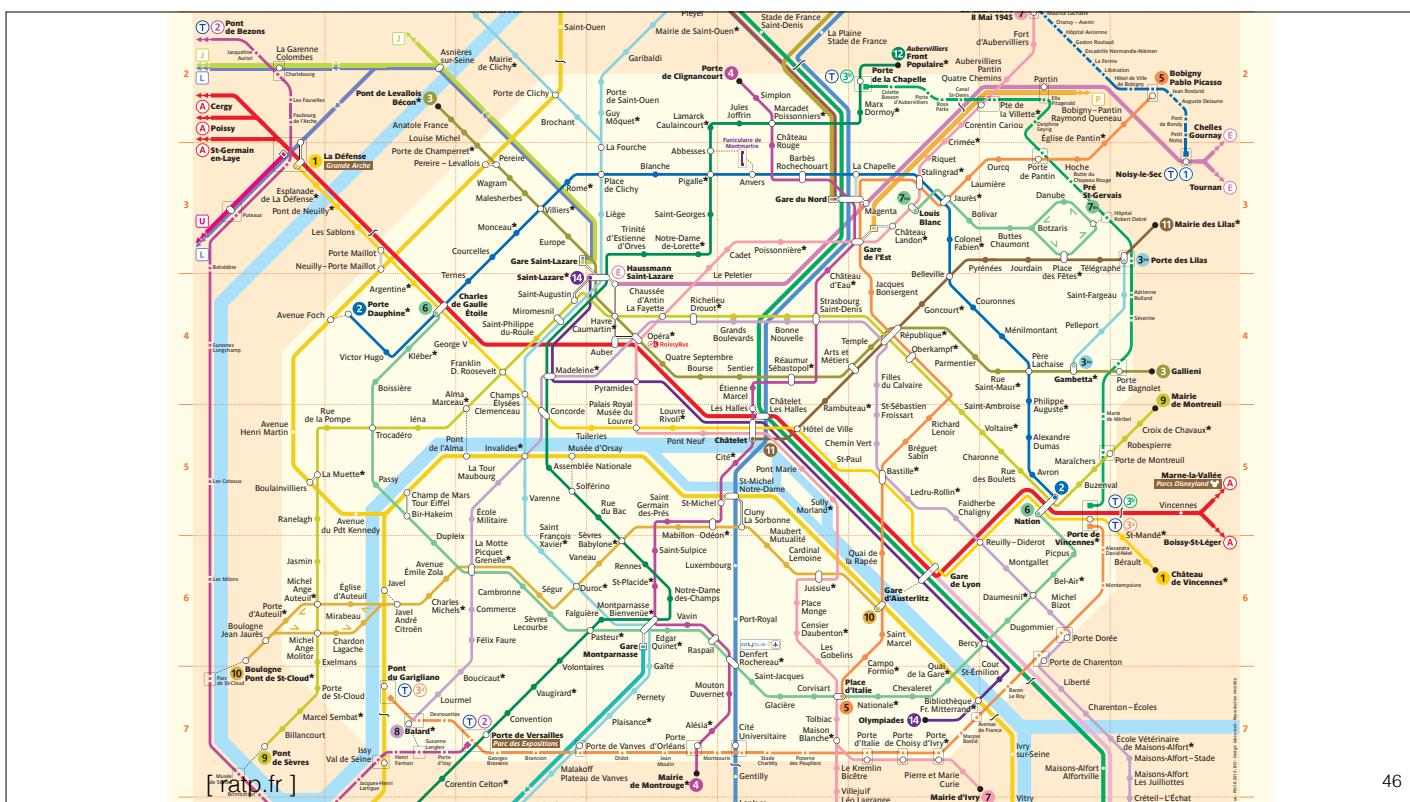
44

Easy Walking Lines Added

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

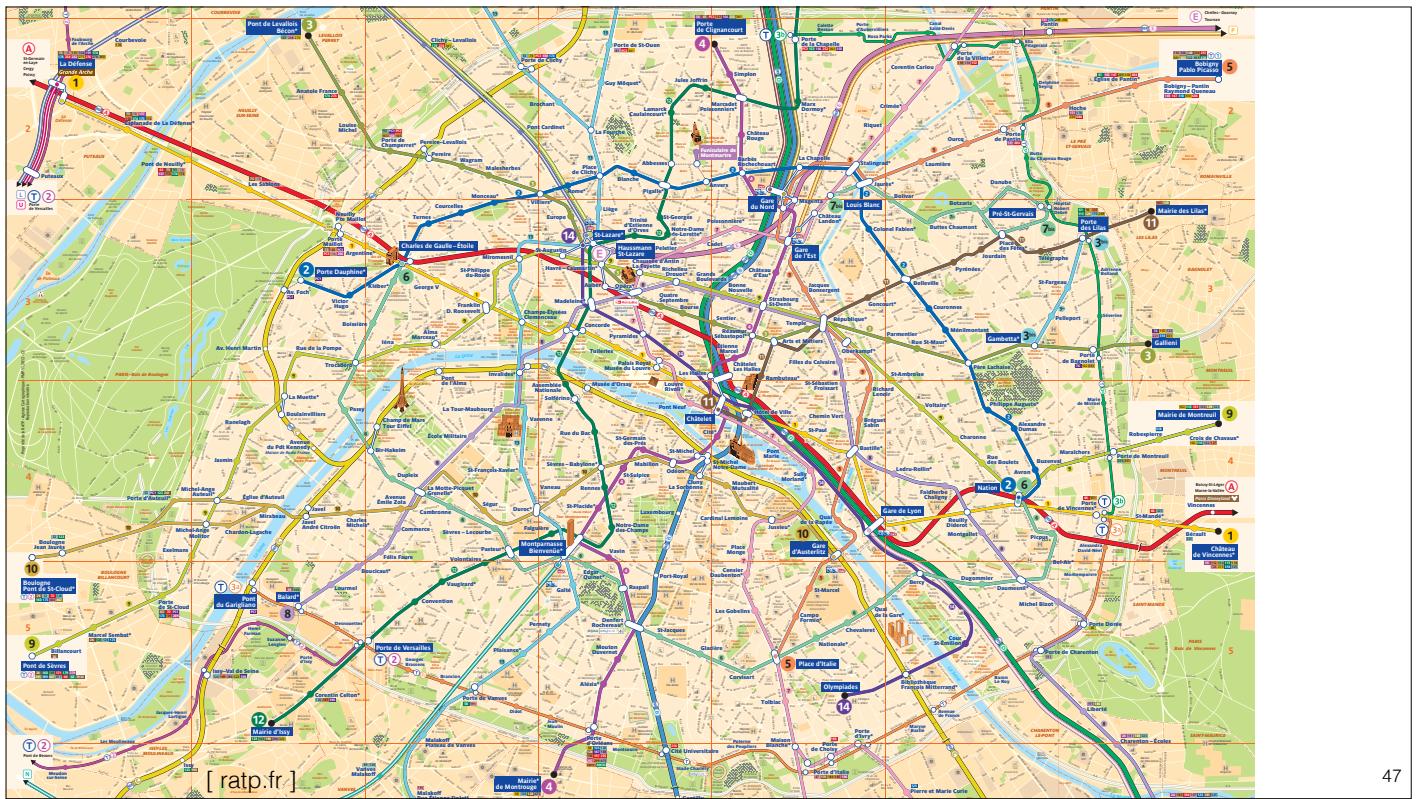
45

45



46

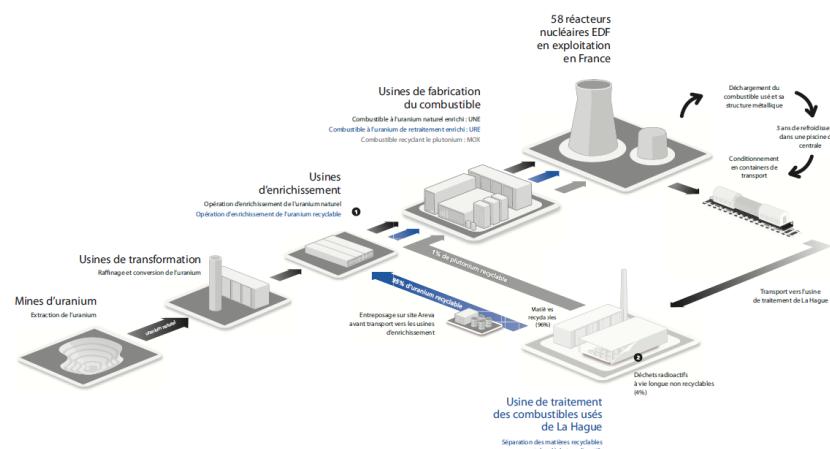
46



47

Networks & Diagrams

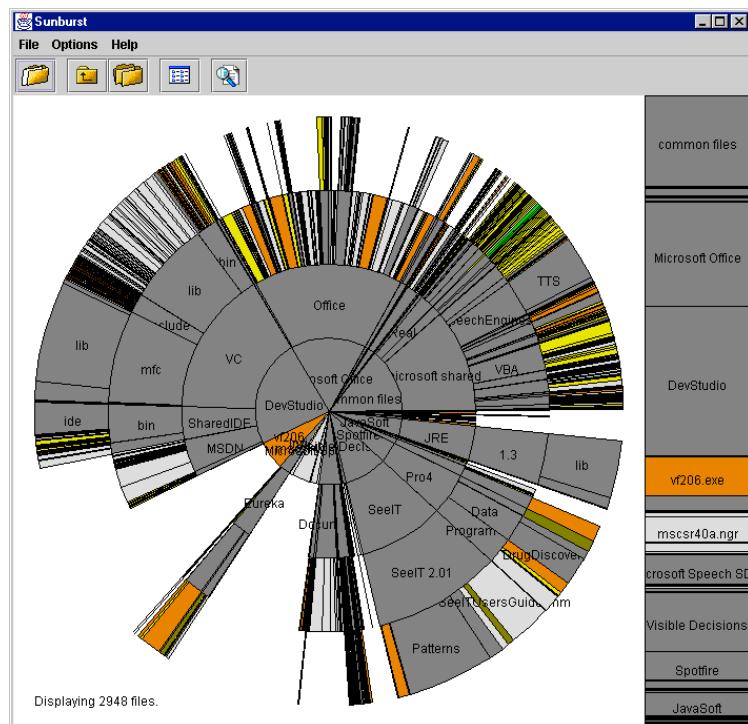
Cycle du combustible nucléaire



48

48

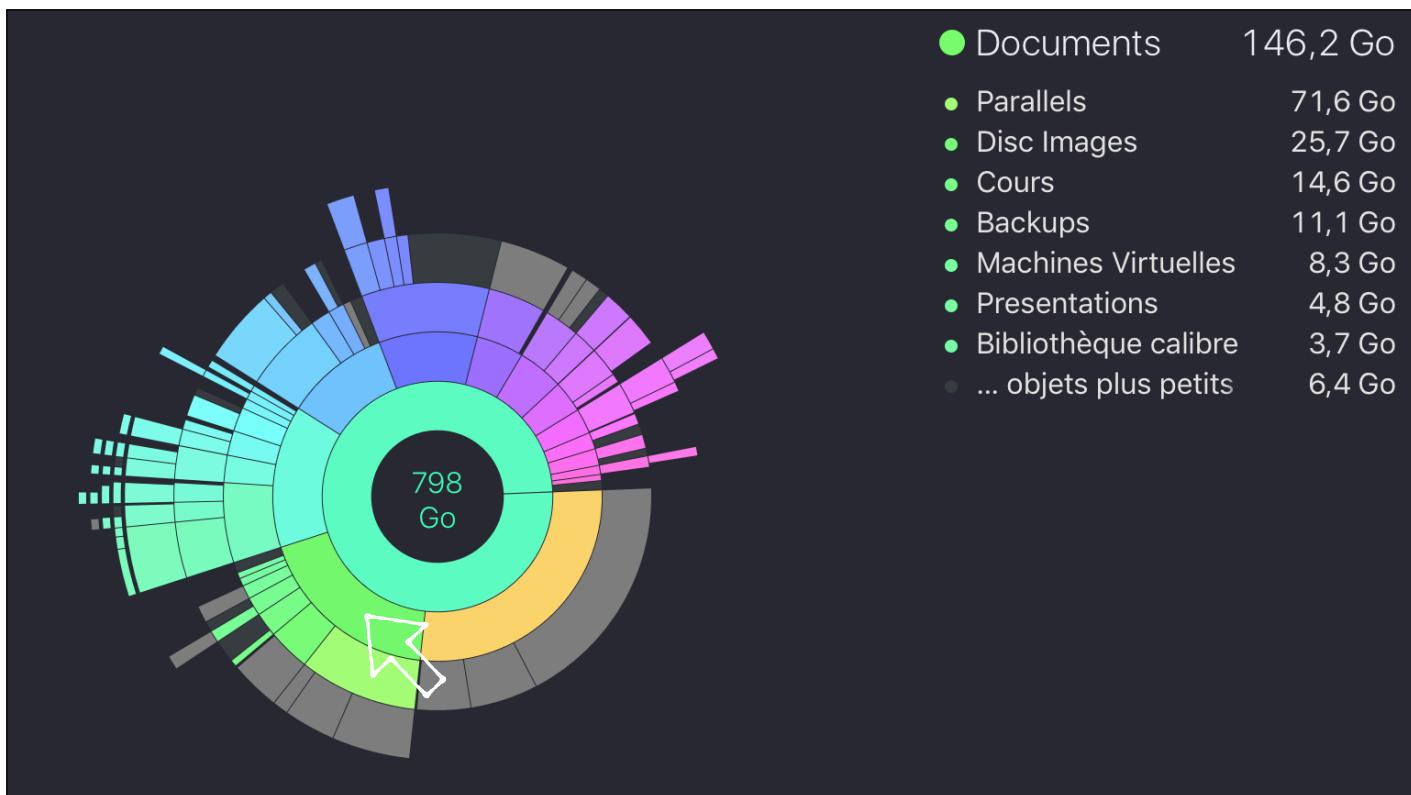
SunBurst



[www.cc.gatech.edu/
gvu/ii/sunburst](http://www.cc.gatech.edu/gvu/ii/sunburst)

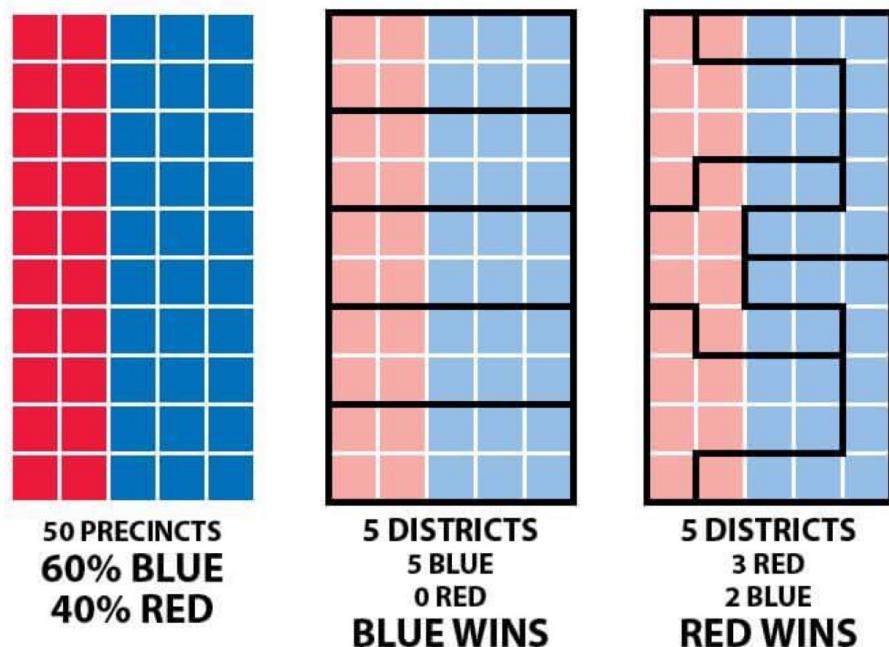
Displaying 2948 files.

49



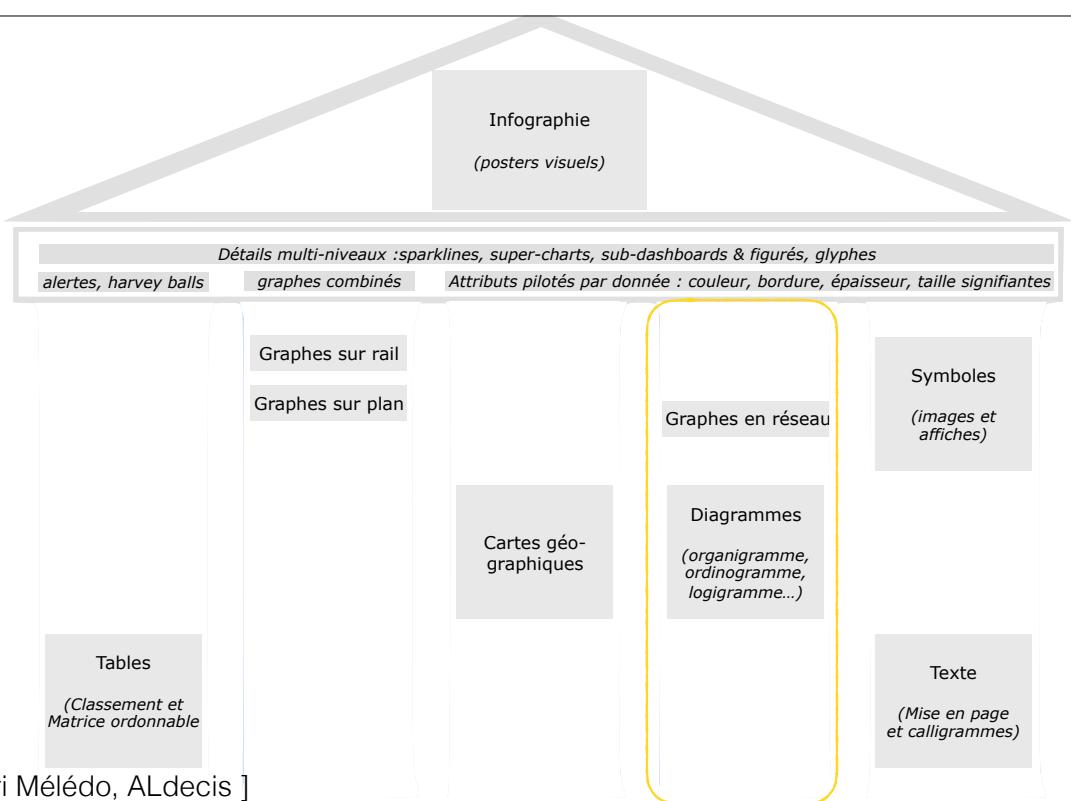
50

HOW TO STEAL AN ELECTION



51

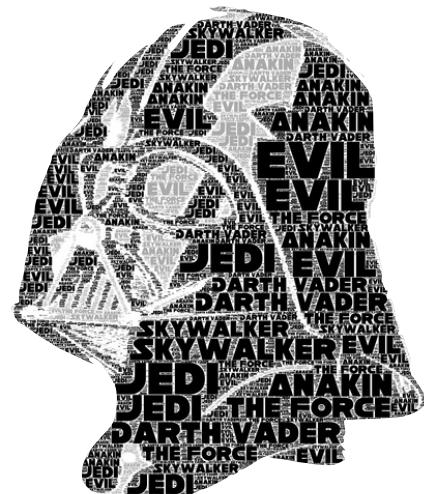
51



52

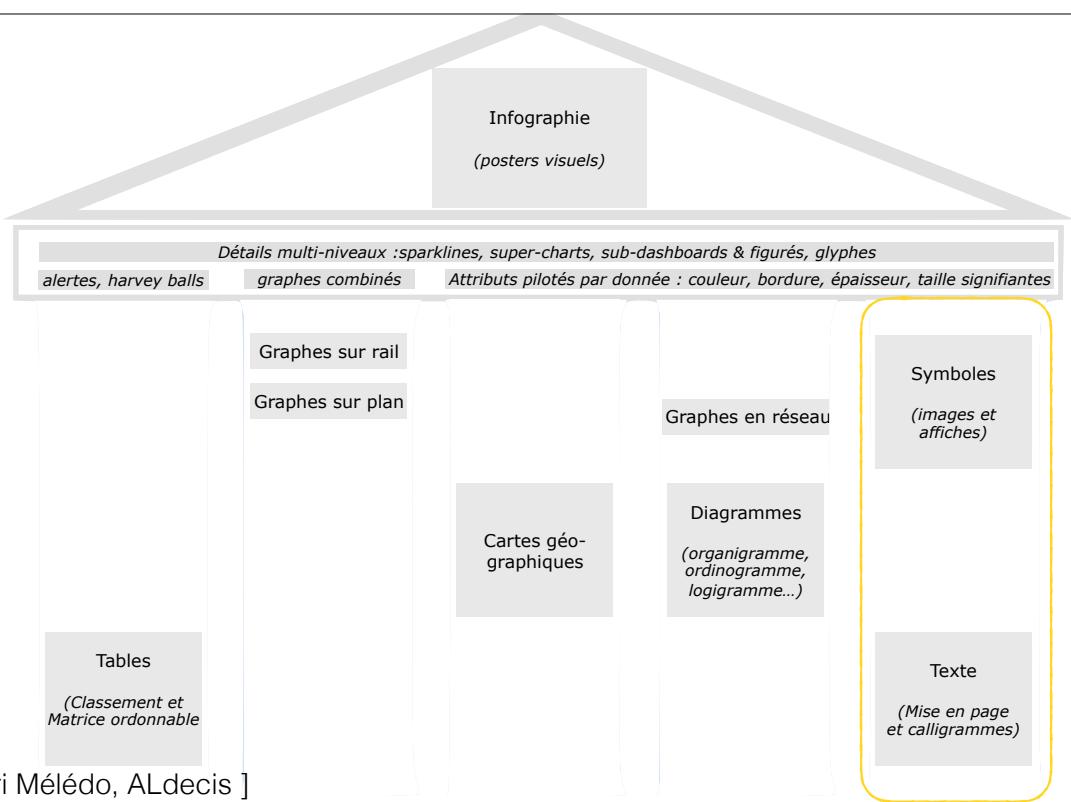
52

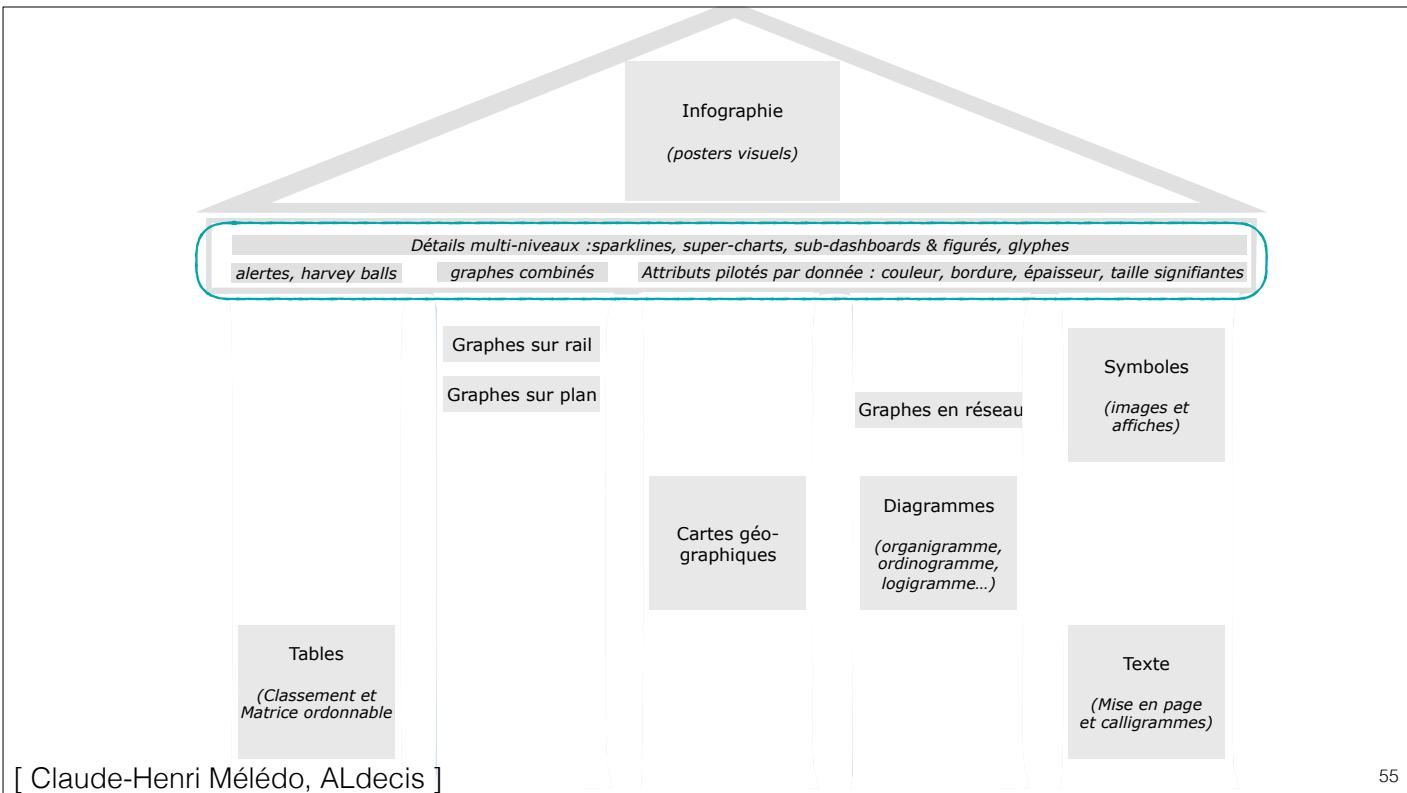
Conceptual Images



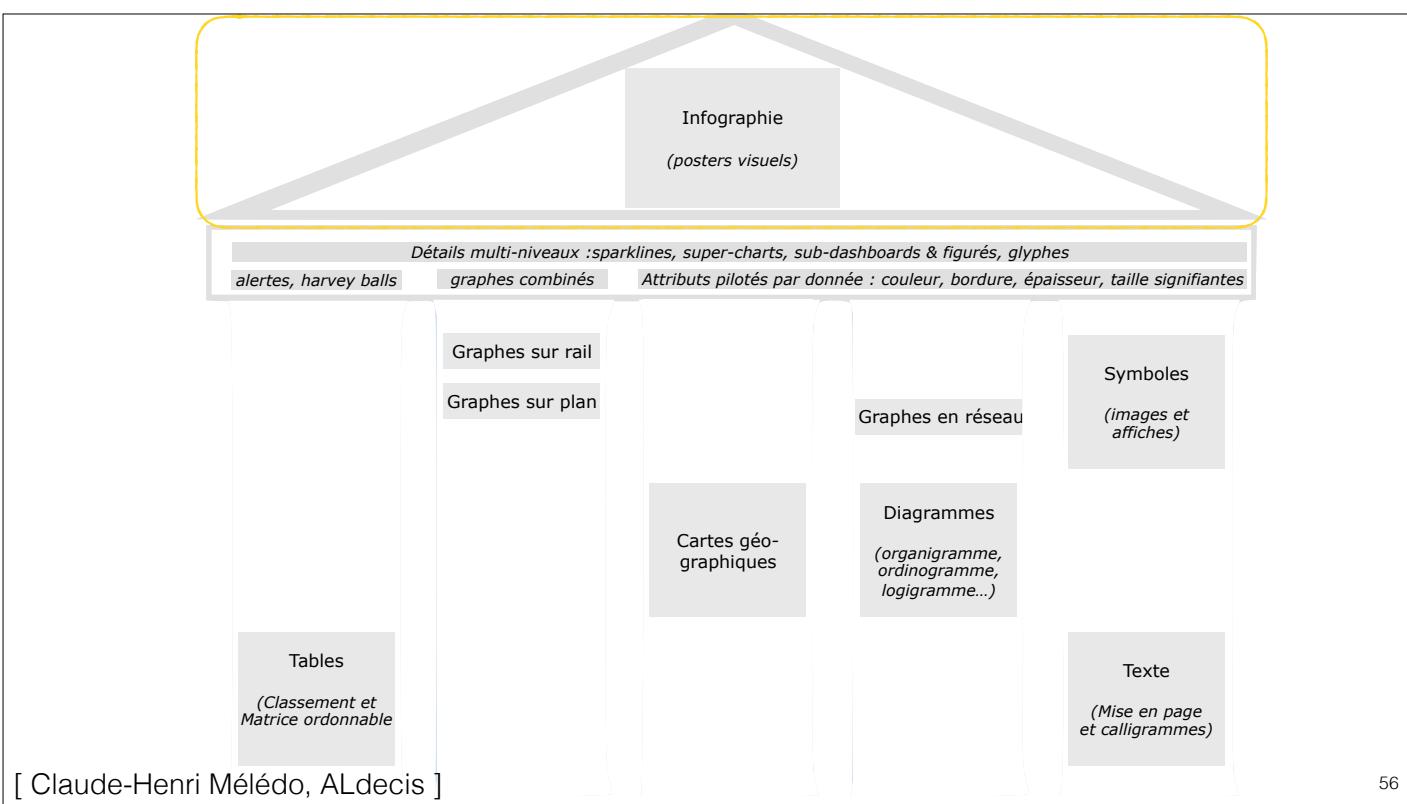
53

53



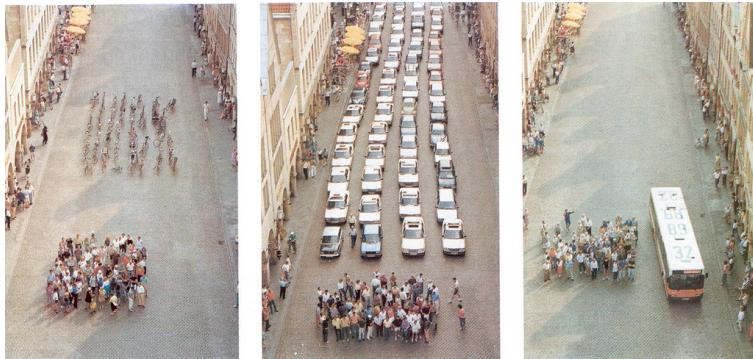


55



56

72 people



[www.visualnews.com]

57

57

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

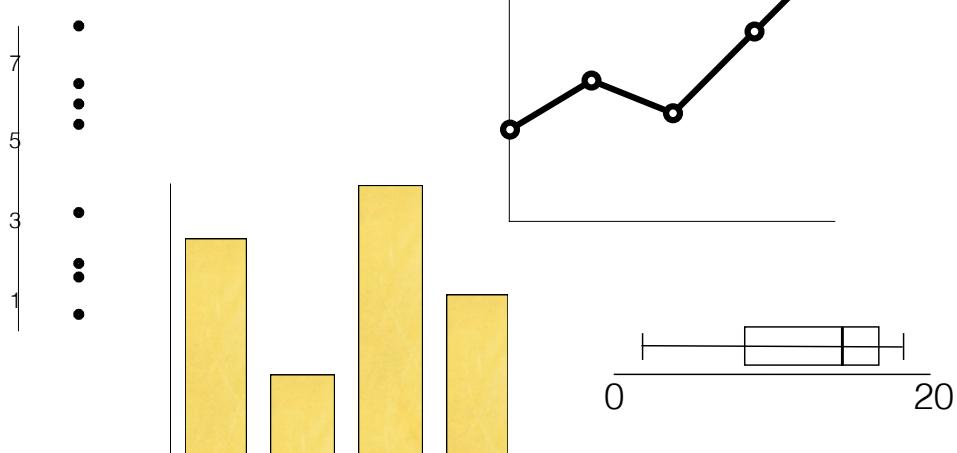
>3 – Hypervariate data

58

58

Univariate Data

Representations

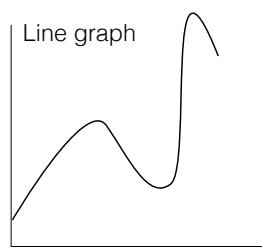


59

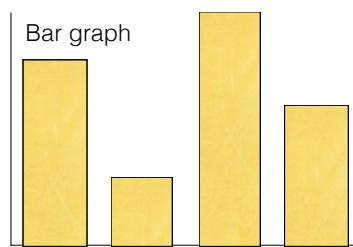
59

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

60

60

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

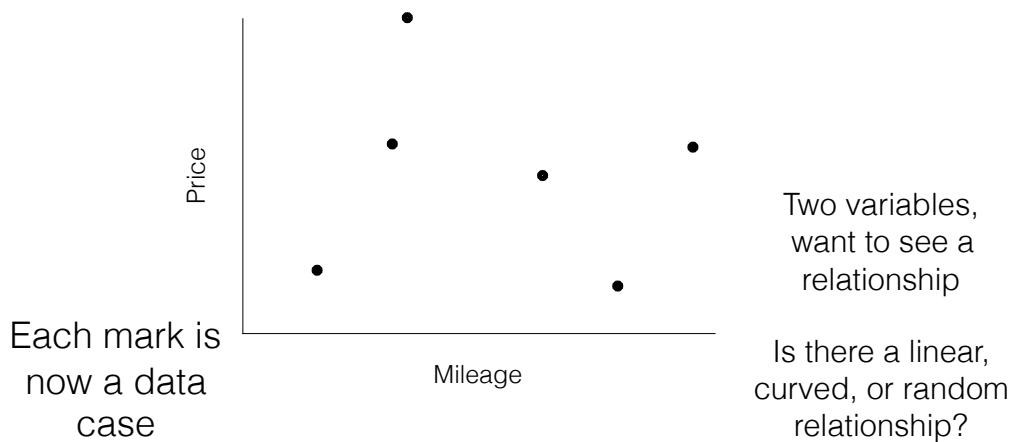
61

61

Bivariate Data

Representations

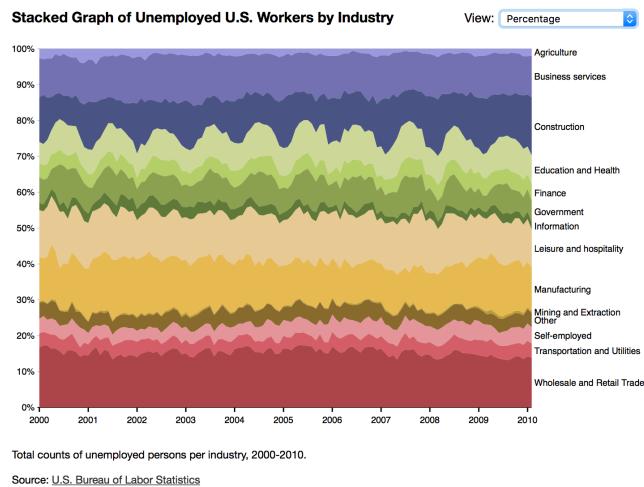
Scatterplot is common



62

62

Time Series

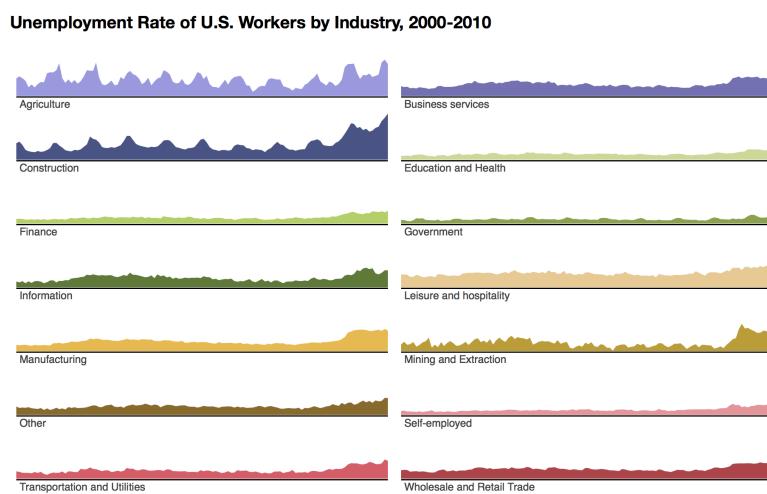


[Source: Heer et al., CACM 2010]

63

63

Time Series (Small multiples)



[Source: Heer et al., CACM 2010]

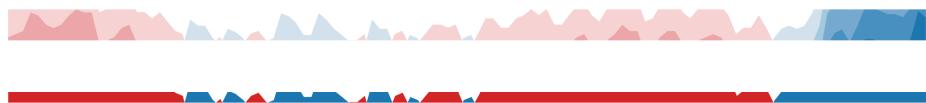
64

64

Horizon Graphs

Horizon Graphs of U.S. Unemployment Rate, 2000-2010

Bands: 5 Negative Values: mirror offset



[Source: Heer et al., CACM 2010]

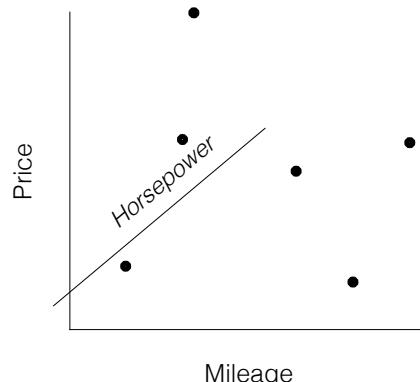
65

65

Trivariate Data

Representations

3D scatterplot is possible

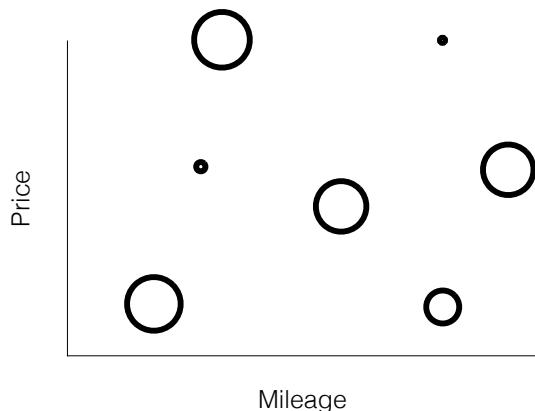


66

66

Alternative Representation

Representations

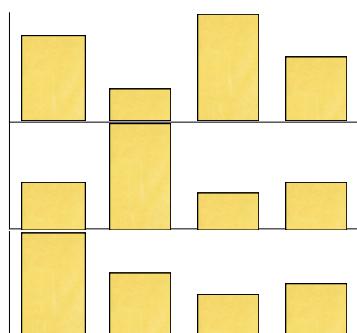


Still use 2D but have
mark property
represent third
variable

67

Alternative Representation

Representations



Represent each variable
in its own explicit way

68

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?

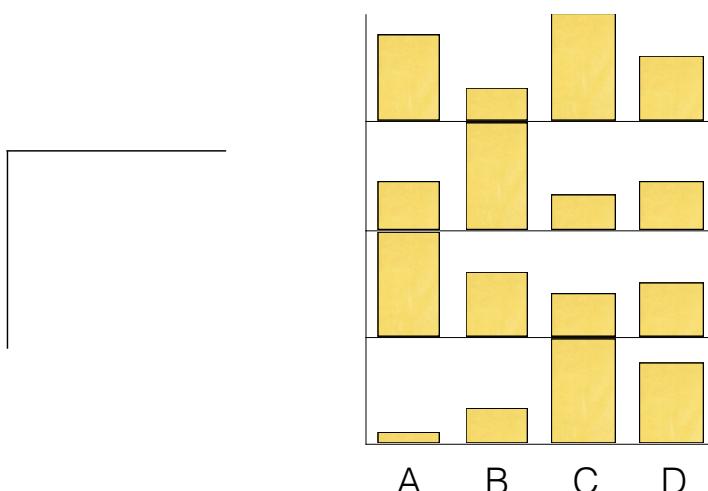
Often the interesting, challenging ones

69

69

Multiple Views

Give each variable its own display



70

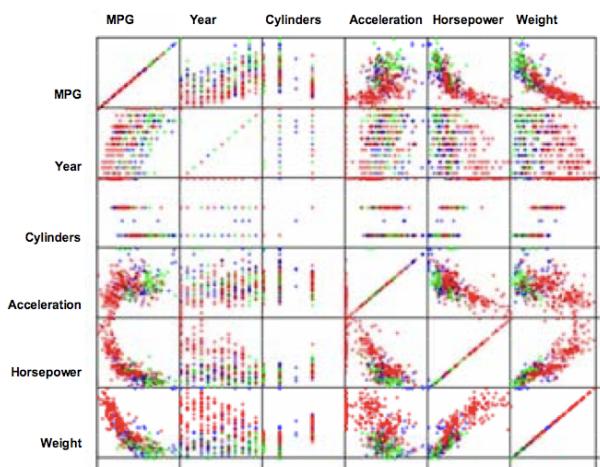
70

Scatterplot Matrix

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

Useful for what?

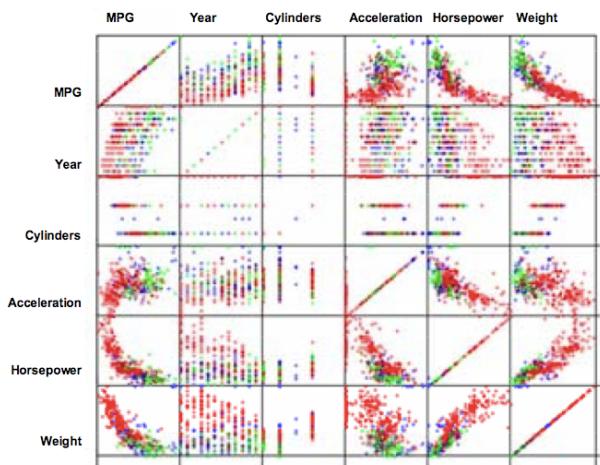
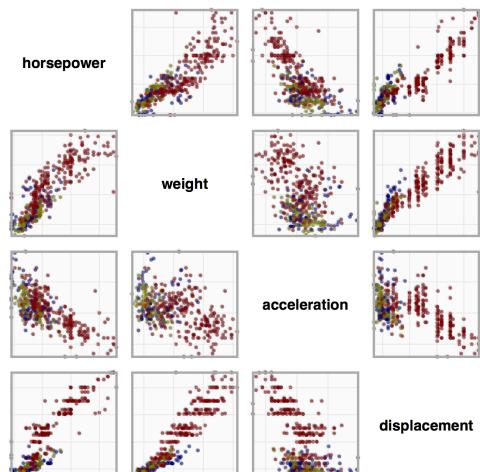
Misses what?



71

71

Scatterplot Matrix

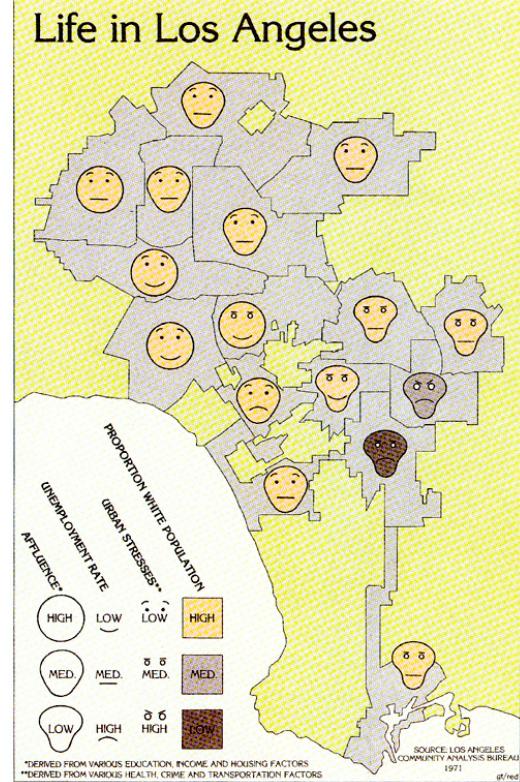
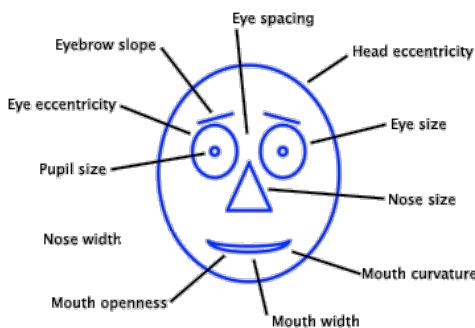


72

72

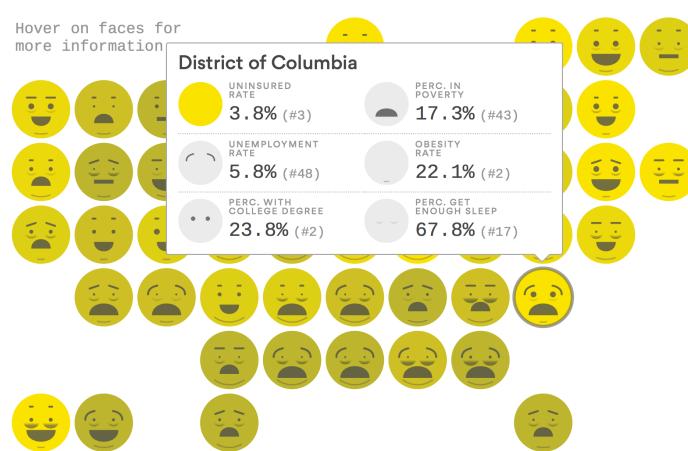
Chernoff Faces

Encode different variables' values in characteristics of human face



73

Emoji Chernoff Faces



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

75

75

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 \rightarrow 2D projections

76

76

Wait a moment...

A spreadsheet already does that

Each variable is positioned into a column

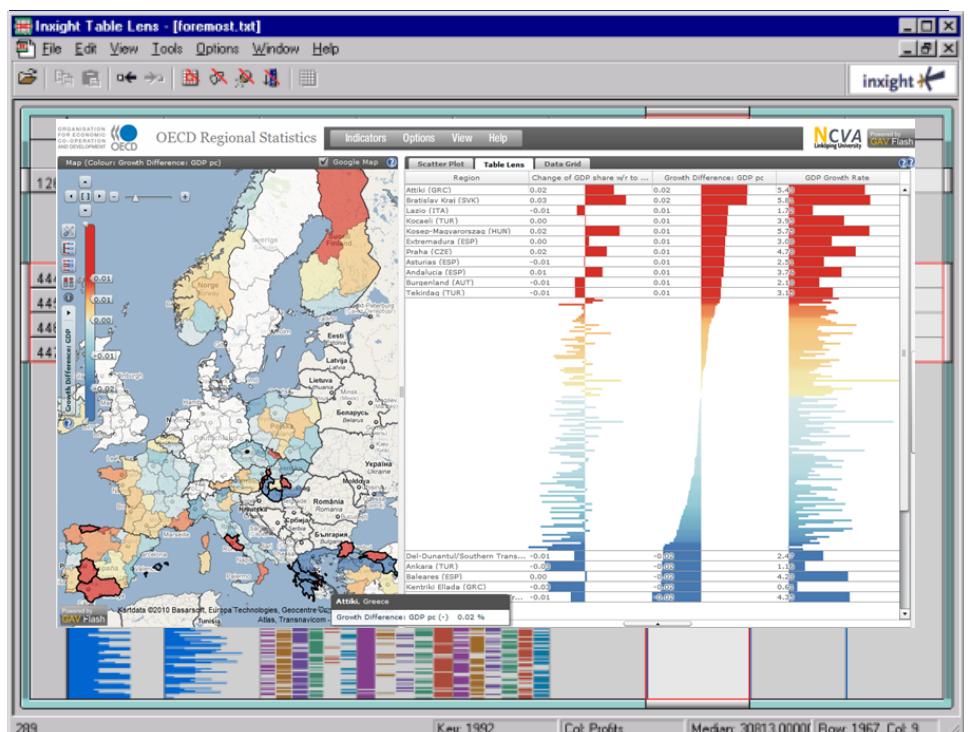
Data cases in rows

This is a projection (mapping)

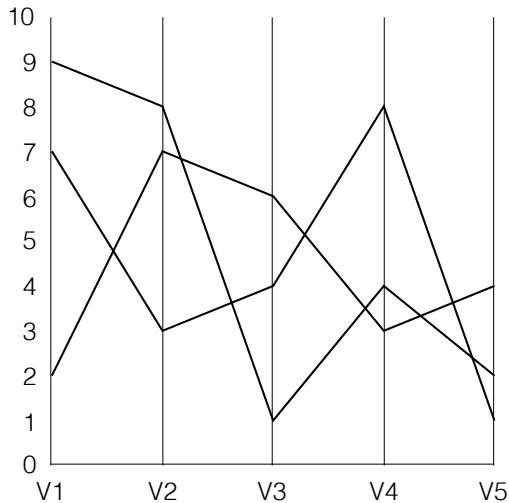
77

77

Table Lens



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

79

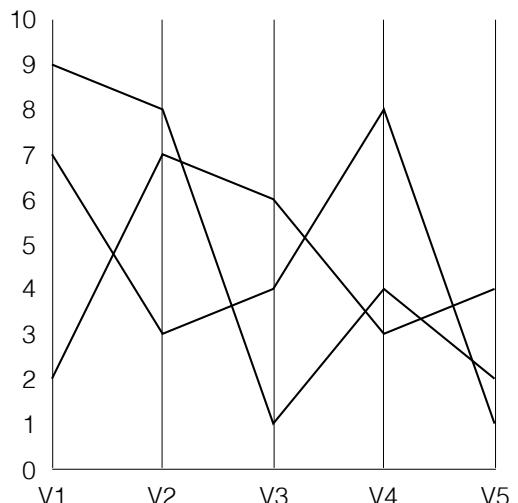
79

Parallel Coordinates

Encode variables along a horizontal row

Vertical line specifies different values that variable can take

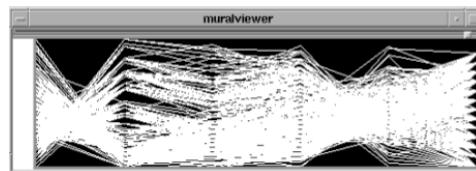
Data point represented as a polyline



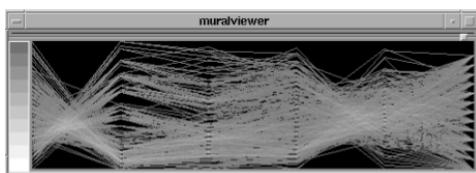
80

80

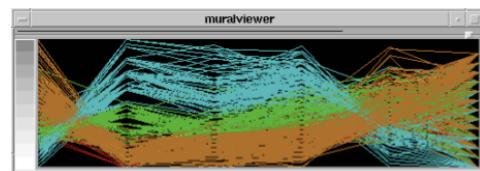
Parallel Coordinates Example



Basic



Greyscale



Color

81

81

Issue

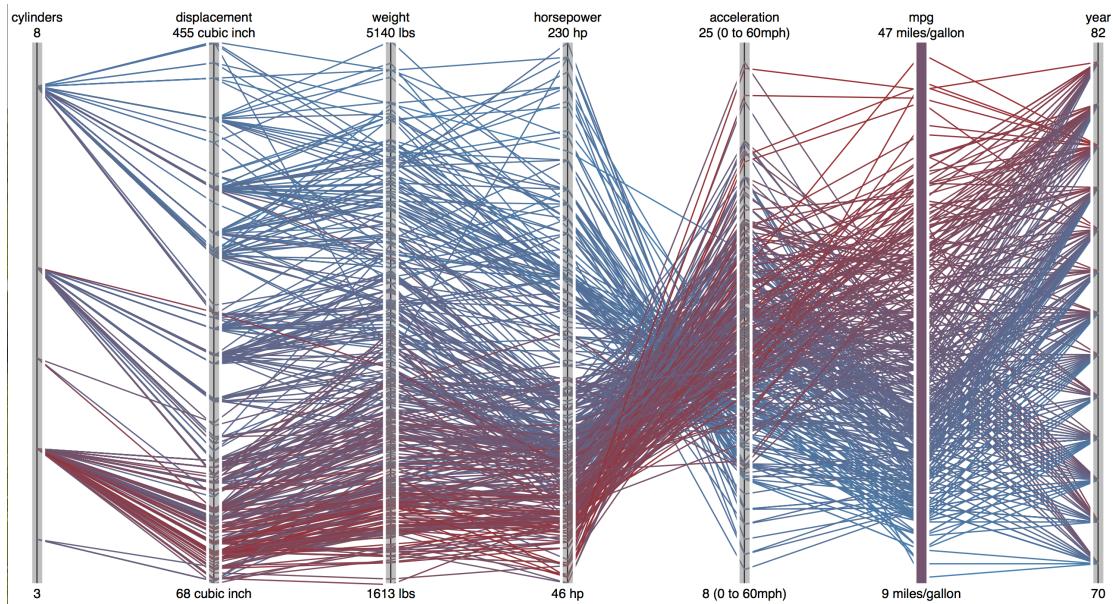
Different variables can have values taking on quite different ranges

Must normalize all down (e.g., $f(x): \mathbb{N} \rightarrow [0, 1]$)

82

82

Parallel Coordinates



[Source: Heer et al., CACM 2010]

83

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

84

The Data

16 variables

X1 — yield

X2 — quality

X3–X12 - # defects (inverted)

X13–X16 - physical parameters

85

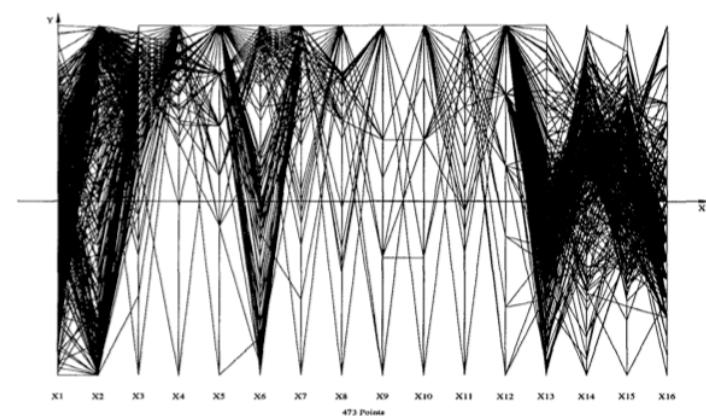
85

Parallel Coordinates

yield &
quality

defects

parameters



Yikes!
But not
that bad

Figure 1: The full dataset consisting of 473 batches

Distributions

x1 - normal

x2 - bipolar

86

86

Top Yield & Quality

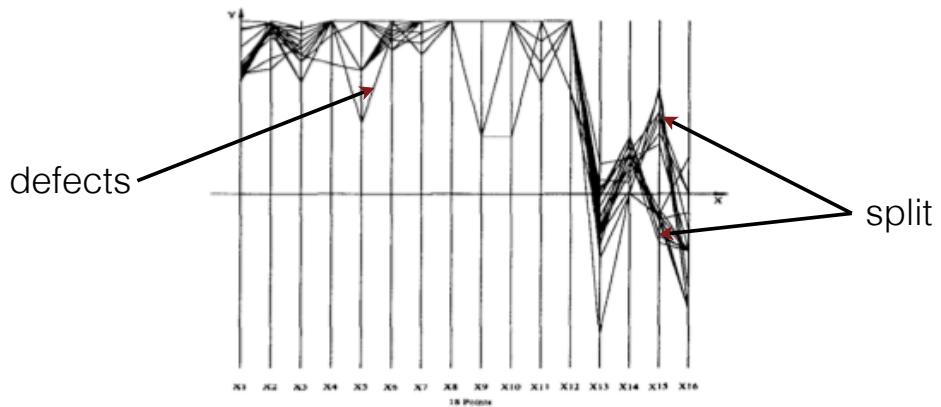


Figure 2: The batches high in Yield, X_1 , and Quality, X_2 .

87

87

Minimal Defects

Not the highest yields and quality

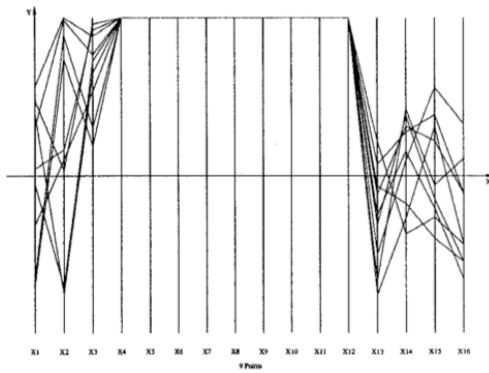


Figure 3: The batches with zero in 9 out of the ten defect types.

88

88

Best Yields

Appears that some defects are necessary to produce the best chips

Non-intuitive!

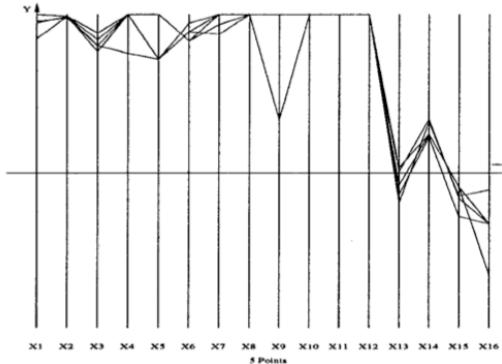
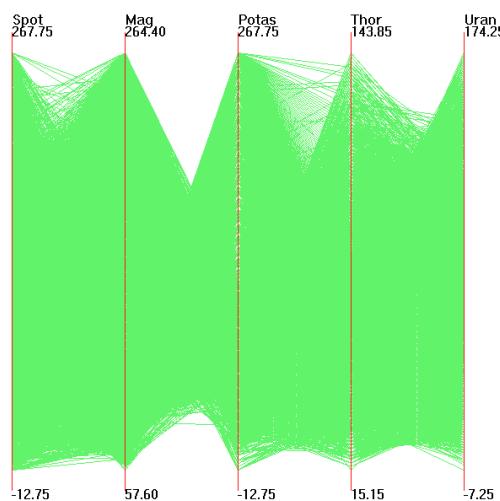


Figure 6: Batches with the highest Yields do not have the lowest defects in X_3 and X_6 .

89

89

Challenges



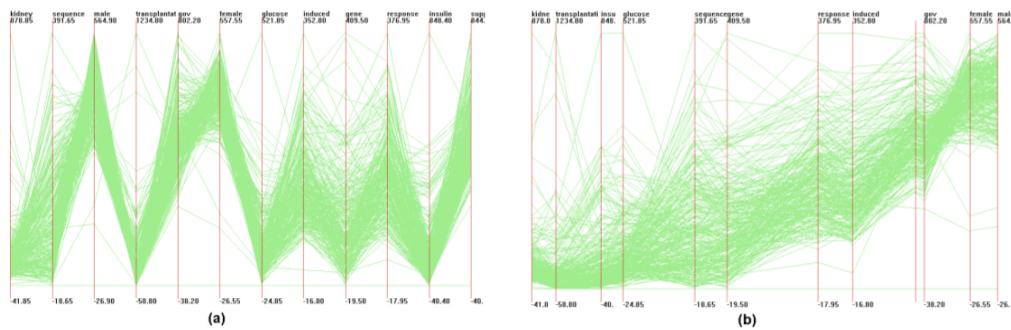
Out5d dataset (5 dimensions, 16384 data items)

90

90

Dimensional Reordering

Which dimensions are most like each other?



Same dimensions ordered by similarity

91

91

Dimensional Reordering

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

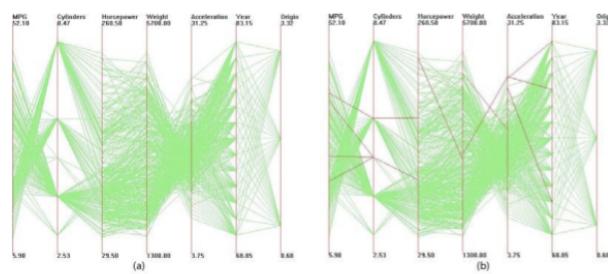
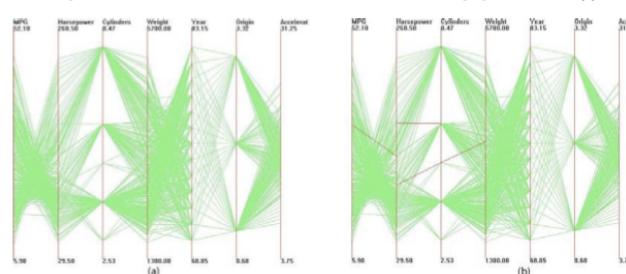


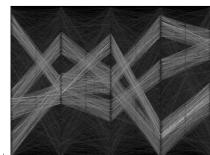
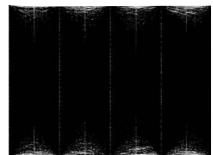
Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).



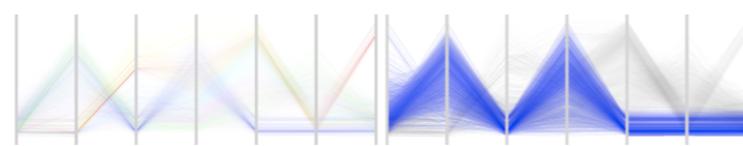
92

92

Reducing Density



[Artero et al, InfoVis 2004]



(a) A linear transfer function has been applied to the high-precision texture in order to prevent cluttering and to provide overview of the data.

(b) A logarithmic transfer function is applied to a selected cluster. The structure is preserved and emphasis is put on the low density regions.

(c) Local cluster outliers are enhanced. A square root transfer function is used and the outliers are visible even through high-density regions.

(d) A complementary view of the clusters with uniform bands. 'Feature animation' presents statistics about the clusters and acts as a guidance.

[Johansson et al, InfoVis 2005]

93

93

Homework

Reading

Table Lens [Rao & Card, CHI '94]

Pixel Bar Charts [Keim et al., InfoVis (1)1, '02]

Munzner, Ch. 5

Project Milestone 1 due at “midnight”

94

94

Next time: Tasks