Information Visualisation: Design Examples

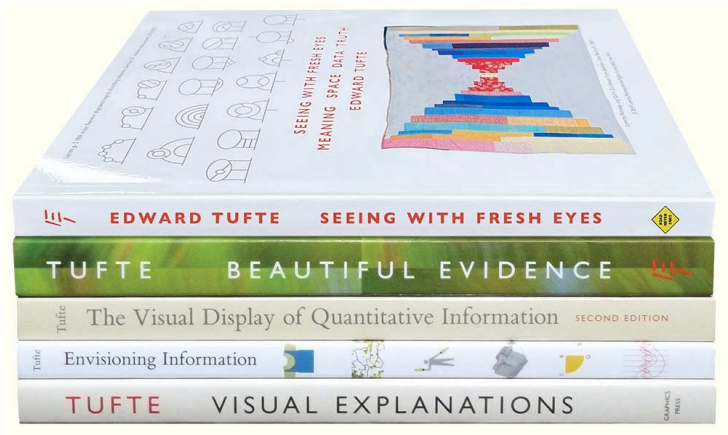
Gallery of Data Visualisation: https://www.datavis.ca/gallery/

Infoviz.info: https://www.infovis.info/

Calling Bullshit book, site and course: https://callingbullshit.org/

Edward Tufte

- ***The Visual Display of Quantitative Information****: "*statistical graphics, charts, tables"
- ***Envisioning Information****: "*high-dimensional complex data"
- ***Visual Explanations: Images and Quantities, Evidence and Narrative:*** "pictures of verbs"
- ***Beautiful Evidence****: "how seeing turns into showing"*
- ***Seeing with Fresh Eyes: Meaning, Space, Data, Truth*** : "deeper level of wisdom and wider realm of inquiry."

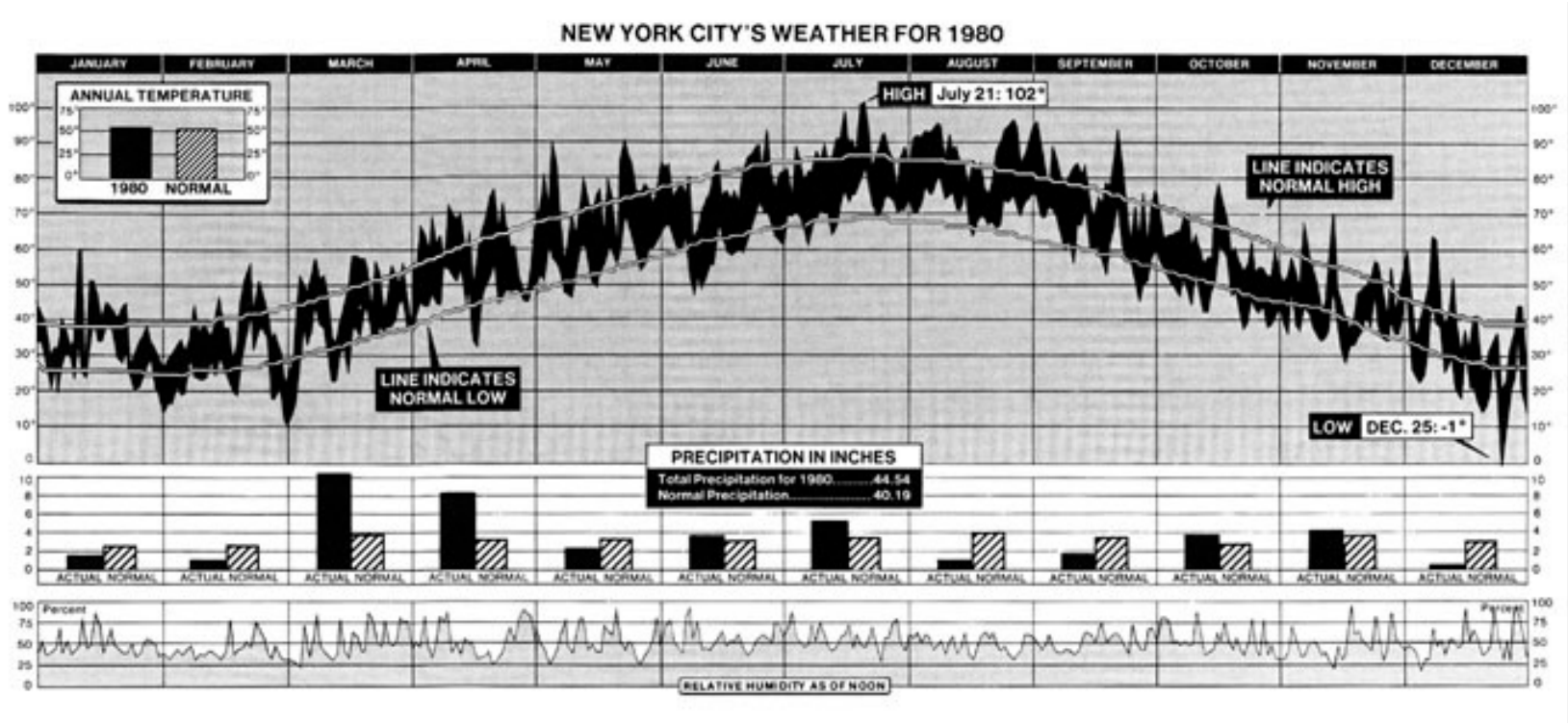


Andy Kirk's blog, with 'best of the visualization web' posting each month (although he's taken a break for a while)

The New York Times' Graphics group is one of the best for 'data journalism', especially in using visualisations to inform and explain.

For thorough yet clear COVID visualisations, it's hard to beat the Financial Times'. Good discussion there of how and why it was made. Andy Kirk did a great podcast with the FT lead, J. Burn-Murdoch

Tamara Munzner's web site, with most of her book, syllabus, videos, etc.

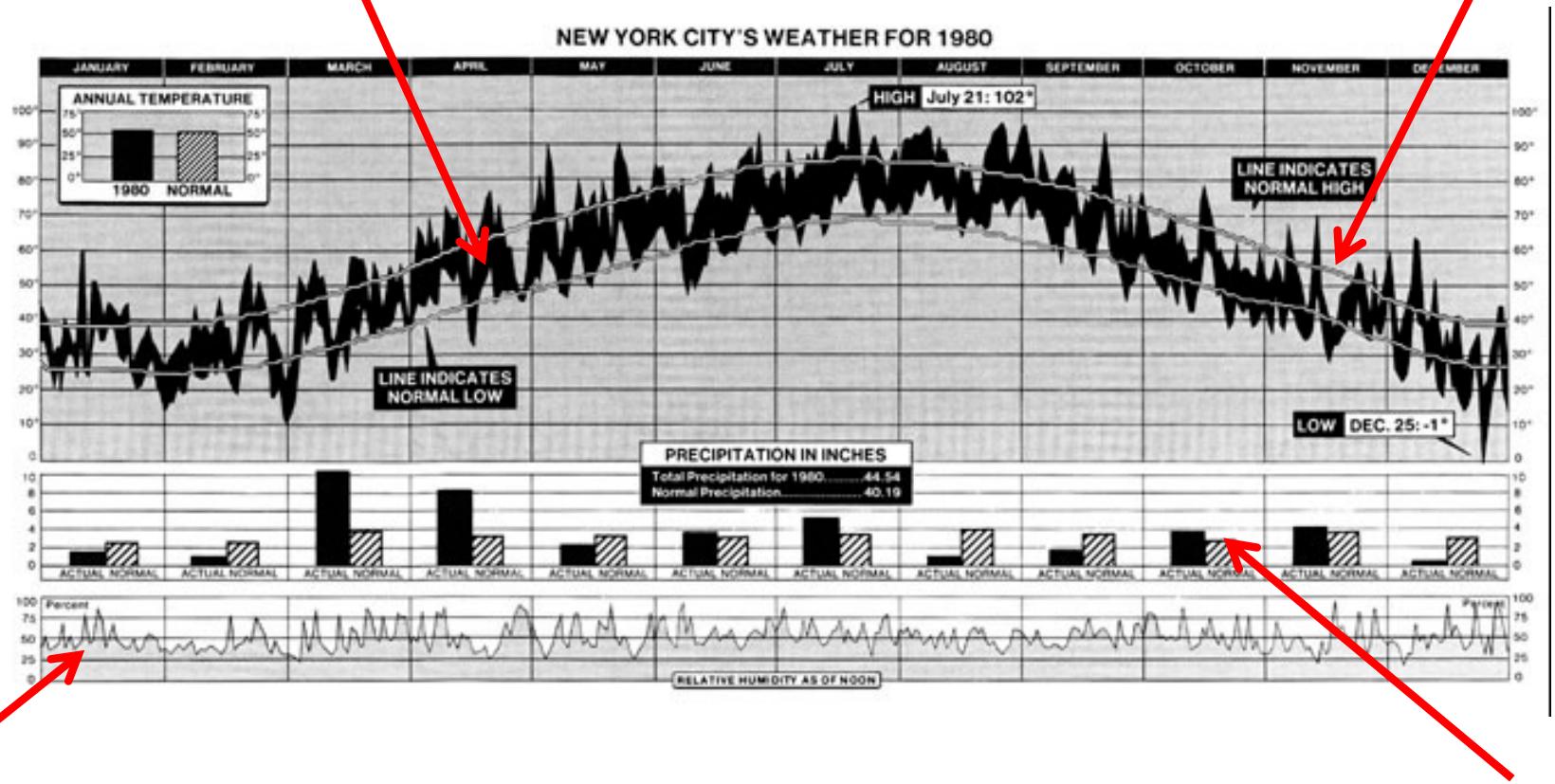


New York Weather in 1980

New York Times (Jan. 11, 1981, p. 32); Tufte (1983, p. 30)

daily temperature
(area filled between high and low)

normal high/low



relative humidity

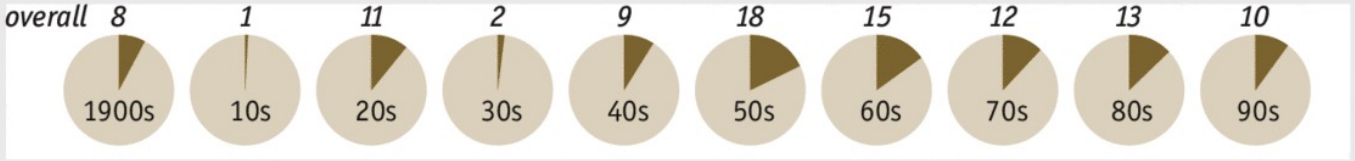
rainfall

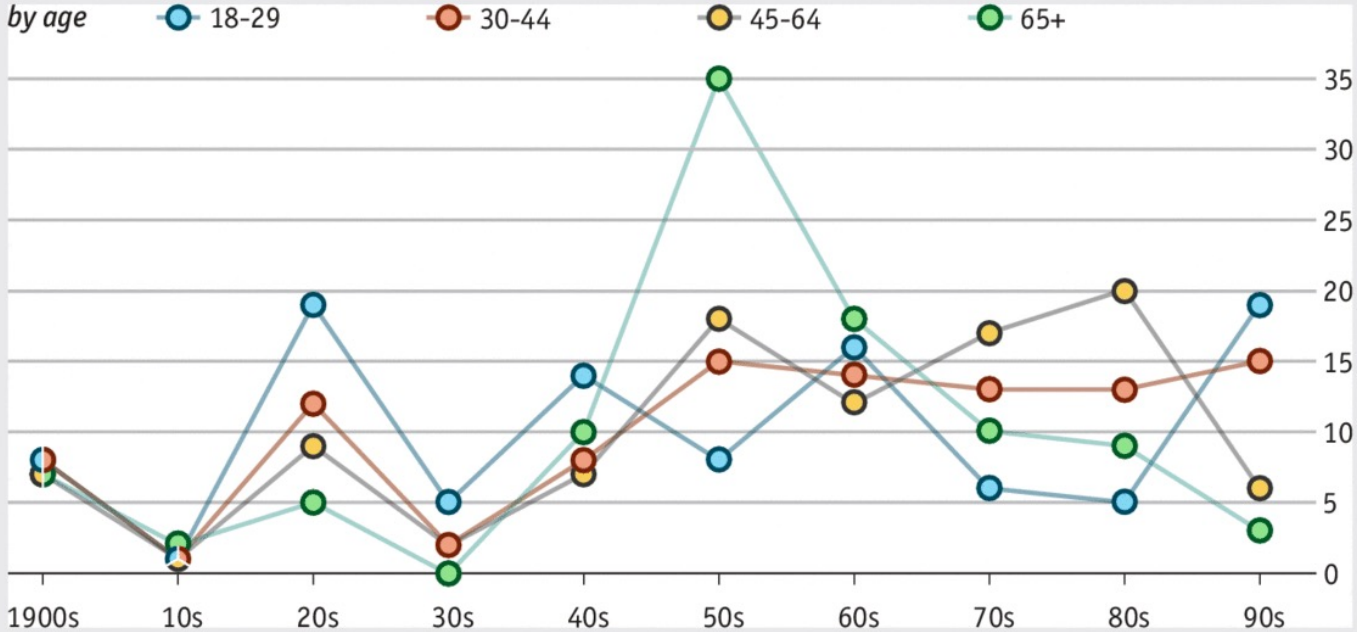
New York Weather in 1980

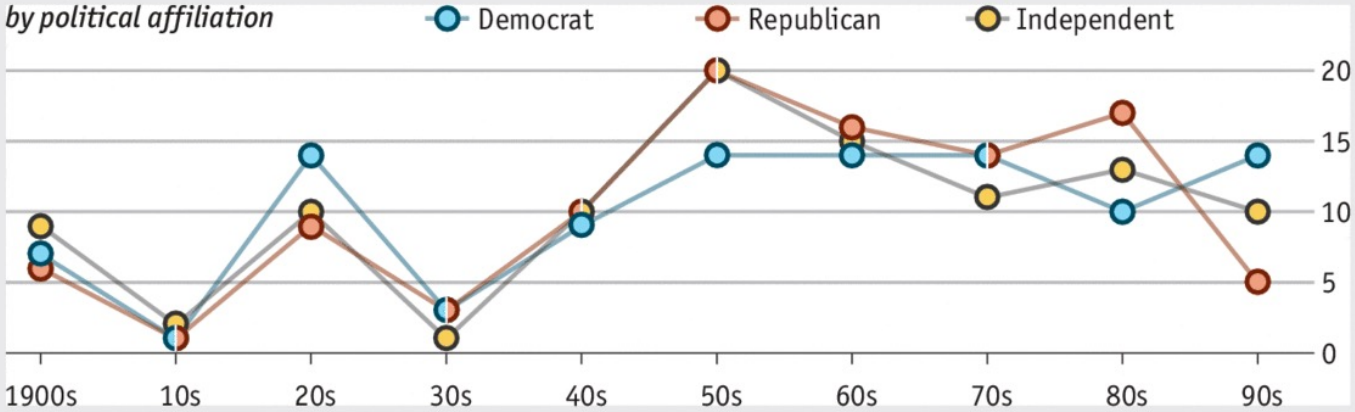
New York Times (Jan. 11, 1981, p. 32); Tufte (1983, p. 30)

Which decade of the 20th century would you most like go back to?

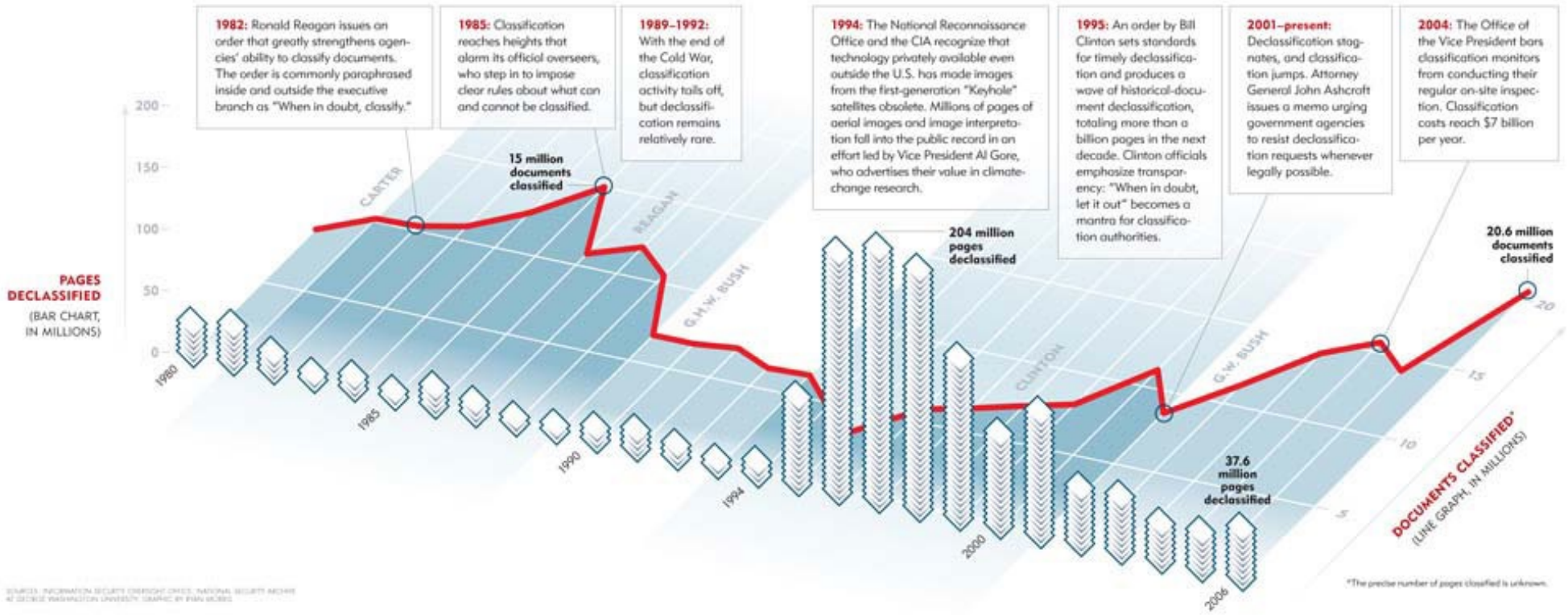
August 9th-10th 2013, *% of Americans responding*:







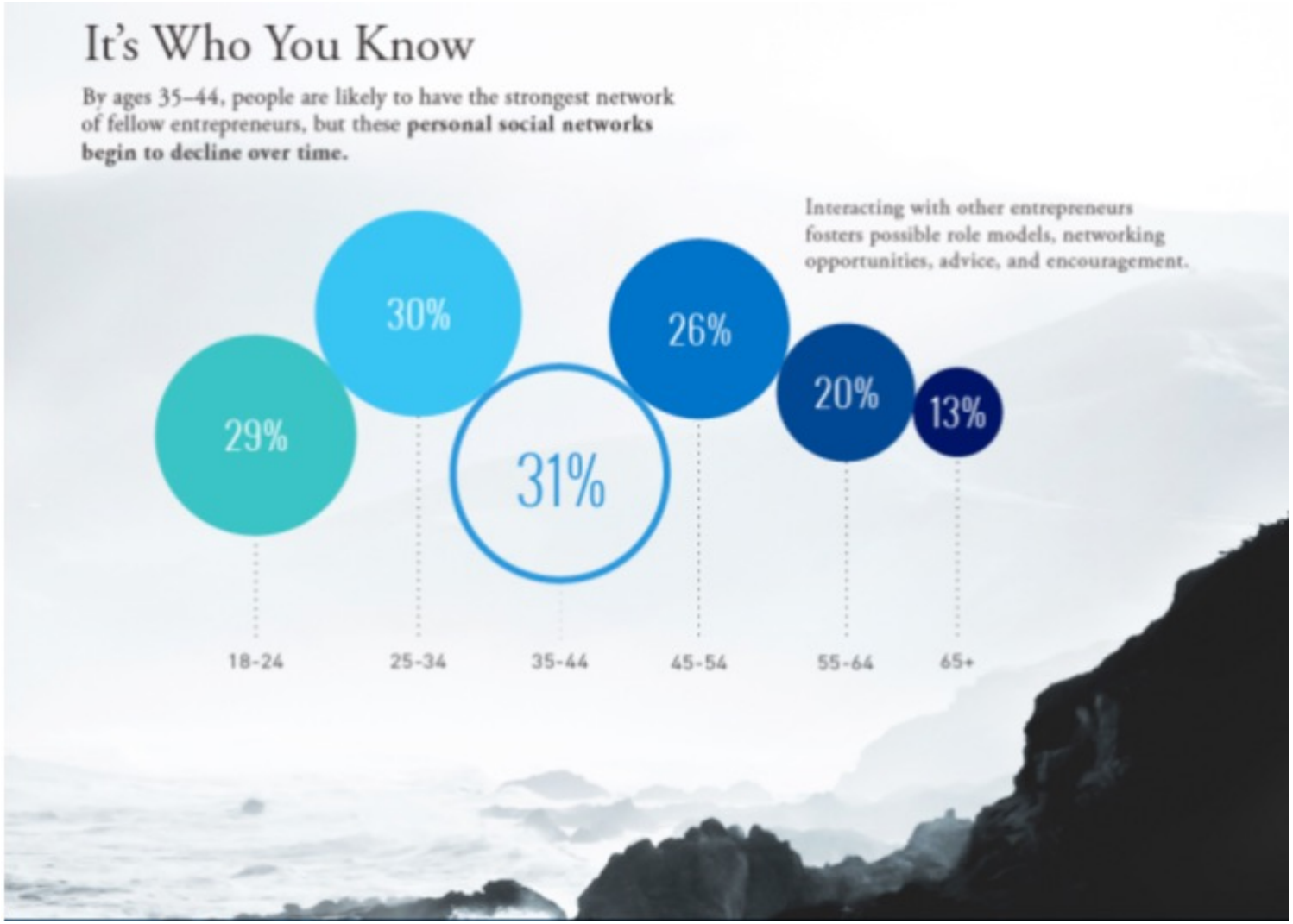
Source: *The Economist*/YouGov

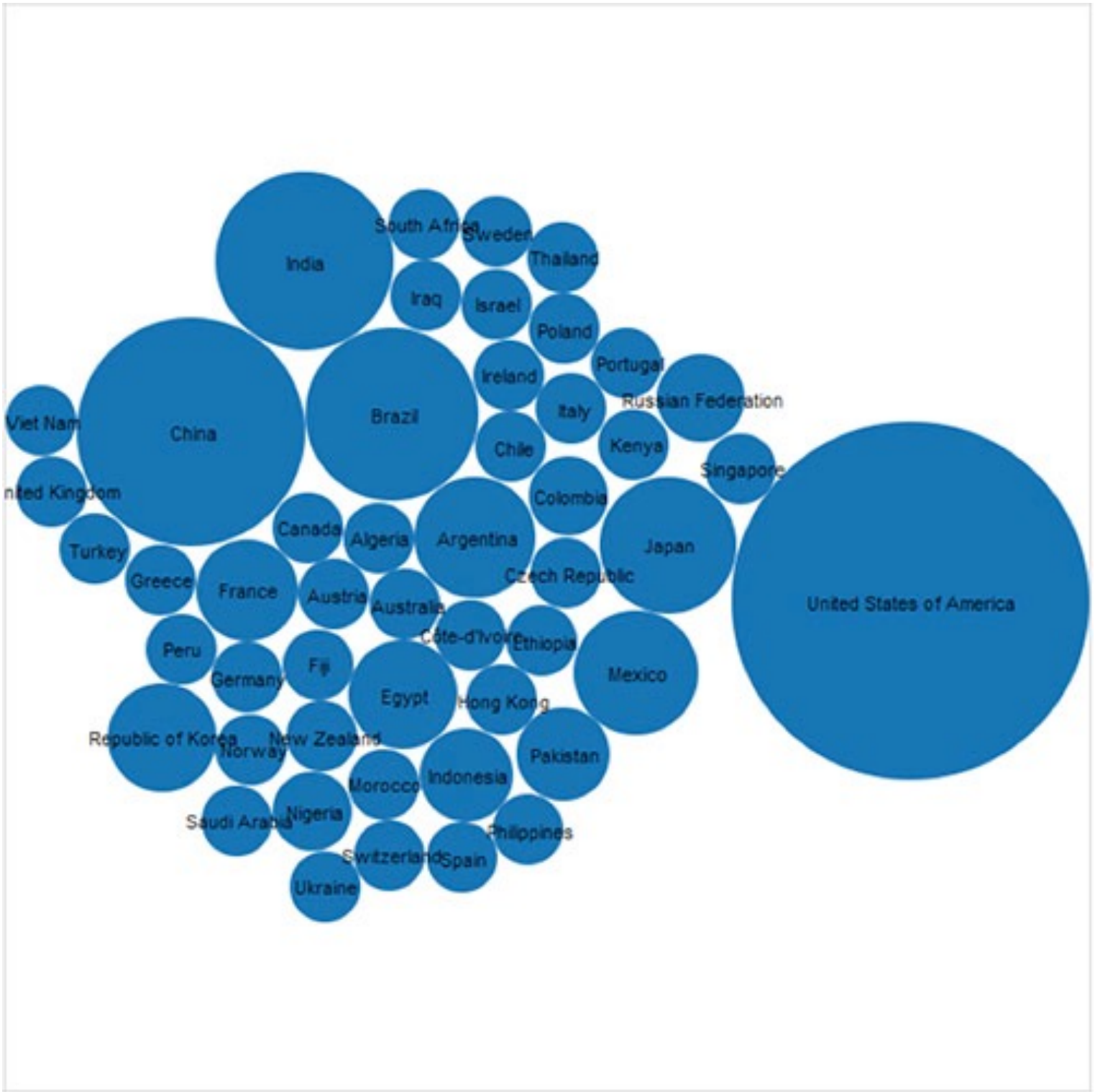


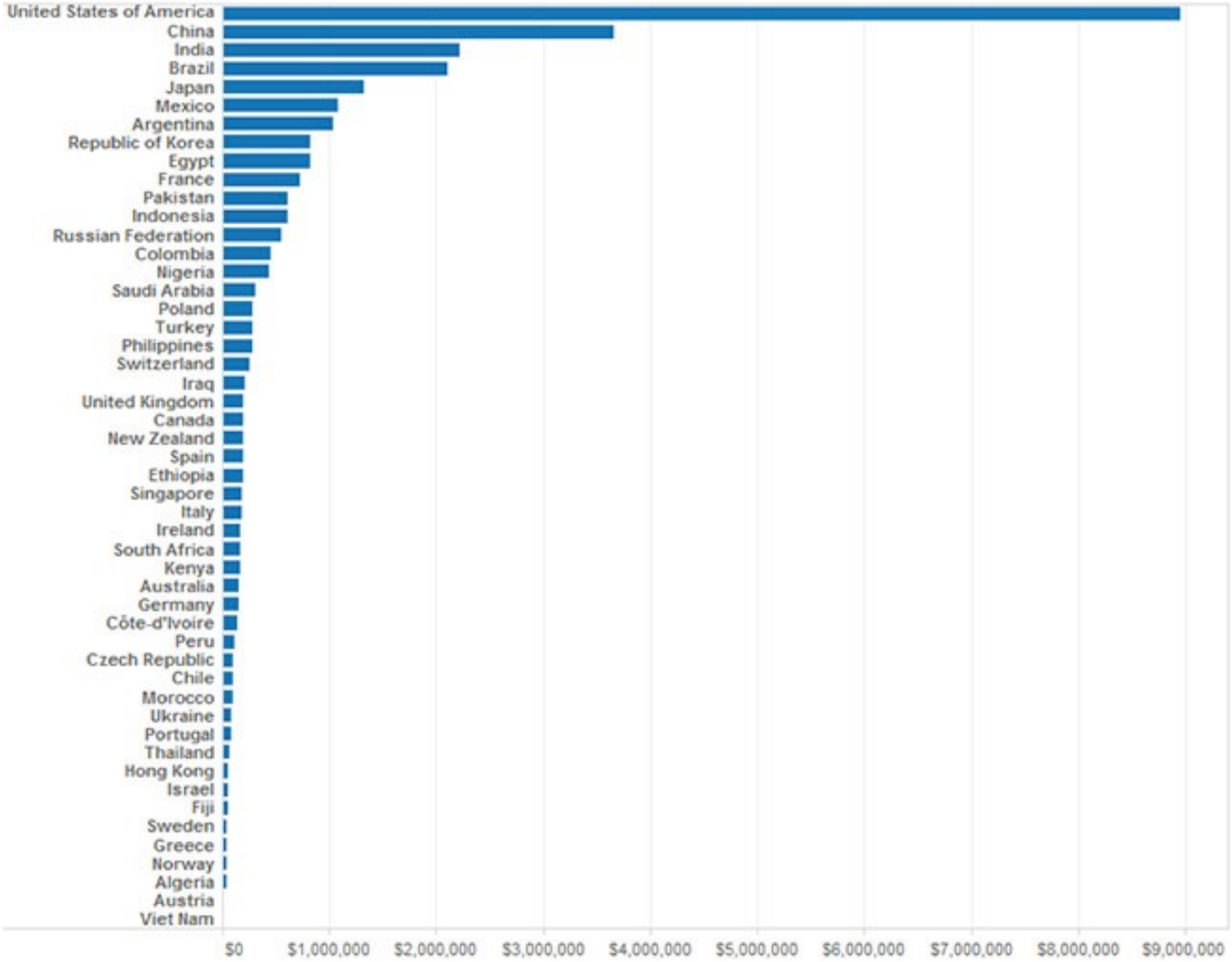
Source: *The Atlantic* 300 no. 2 (September 2007)
Number of Classified U.S. Documents

It's Who You Know

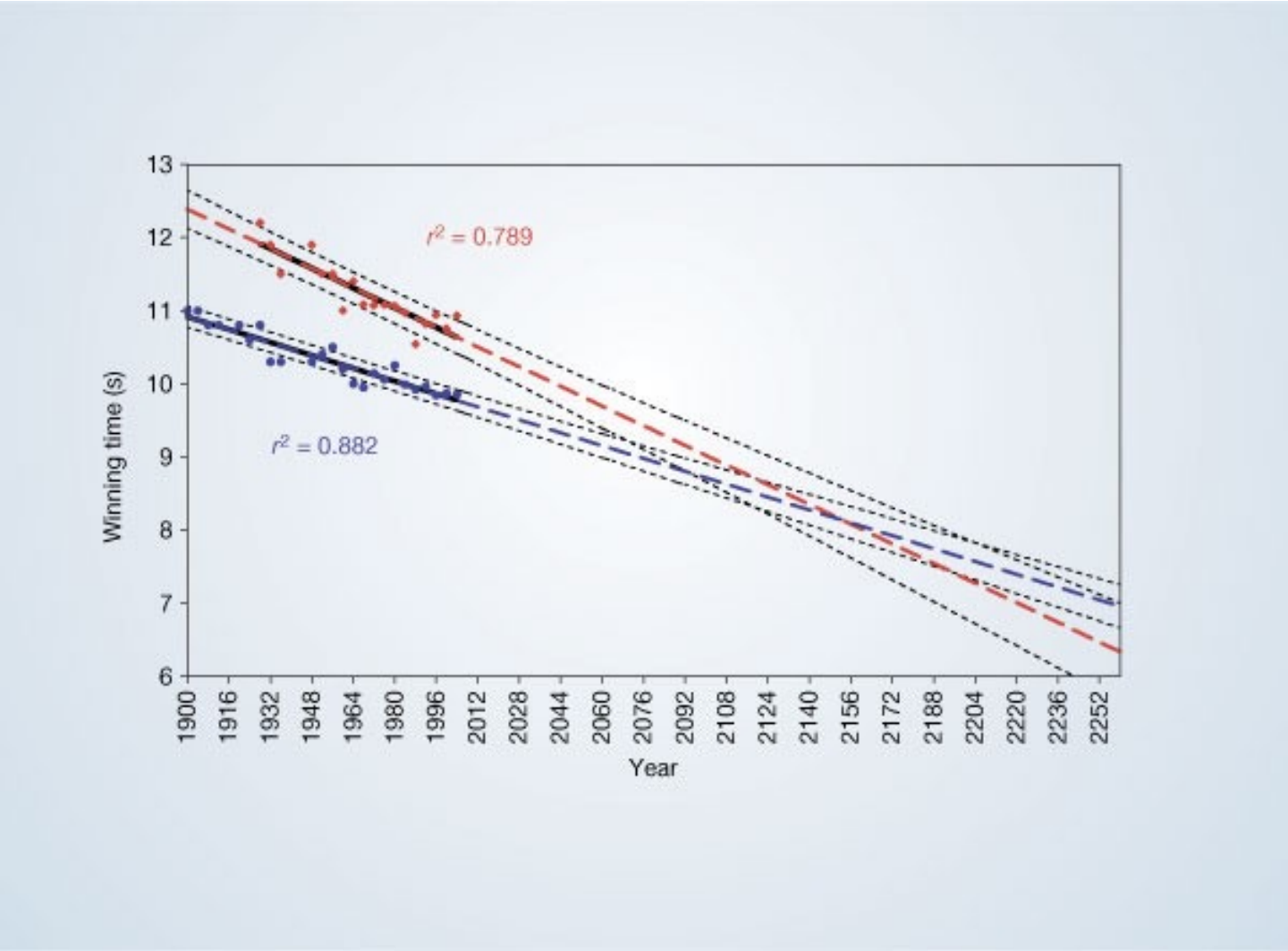
By ages 35–44, people are likely to have the strongest network of fellow entrepreneurs, but these **personal social networks begin to decline over time.**




India
South Africa
Sweden
Thailand
Iraq
Israel
Poland
Portugal
Ireland
Italy
Russian Federation
Viet Nam
China
Brazil
Chile
Kenya
Singapore
nited Kingdom
Colombia
Canada
Algeria
Argentina
Japan
Turkey
Czech Republic
Greece
France
Austria
Australia
United States of America
Côte-d'Ivoire
Ethiopia
Peru
Fiji
Germany
Egypt
Mexico
Hong Kong
Republic of Korea
New Zealand
Norway
Pakistan
Morocco
Indonesia
Nigeria
Saudi Arabia
Philippines
Switzerland
Spain
Ukraine

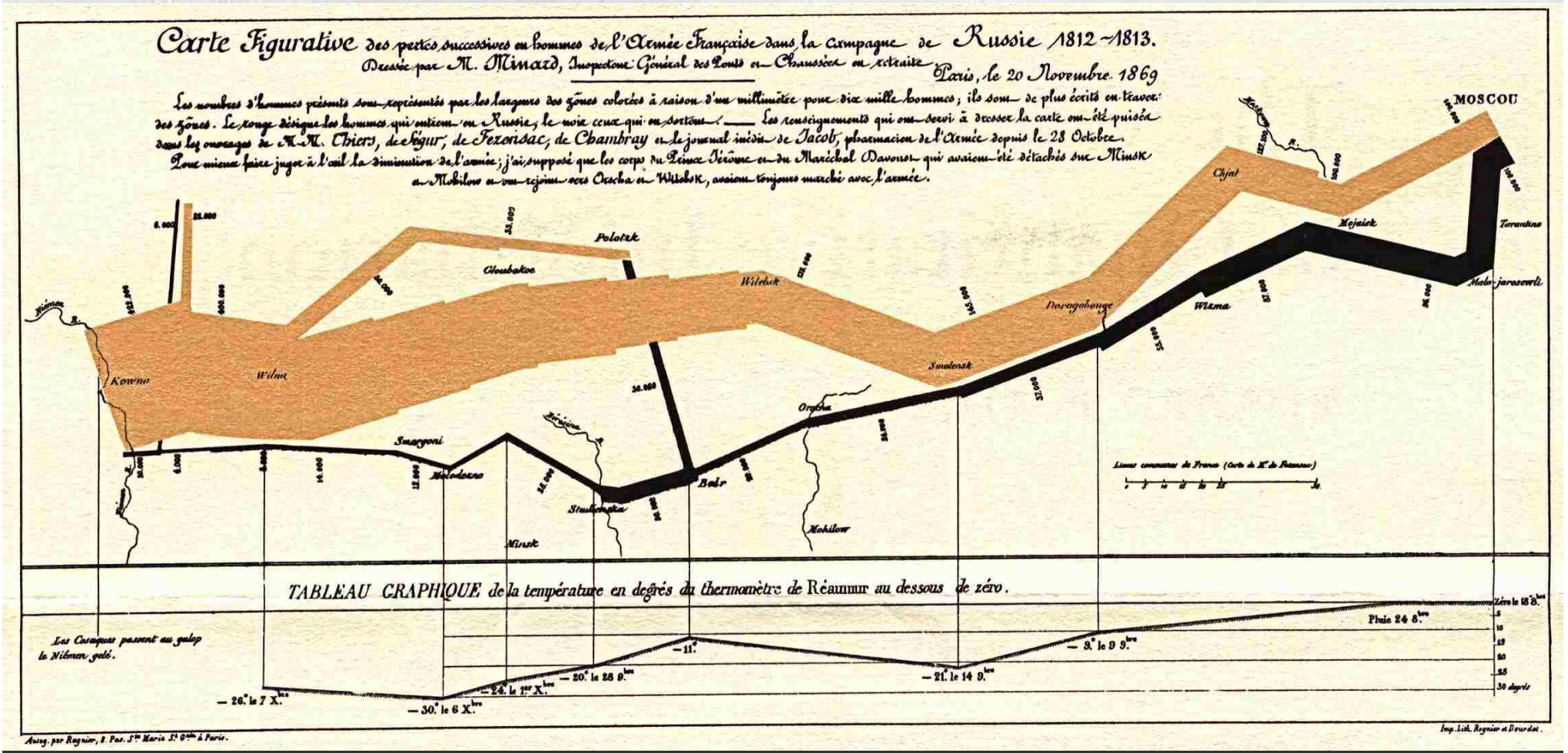

United States of America
China
India
Brazil
Japan
Mexico
Argentina
Republic of Korea
Egypt
France
Pakistan
Indonesia
Russian Federation
Colombia
Nigeria
Saudi Arabia
Poland
Turkey
Philippines
Switzerland
Iraq
United Kingdom
Canada
New Zealand
Spain
Ethiopia
Singapore
Italy
Ireland
South Africa
Kenya
Australia
Germany
Côte-d'Ivoire
Peru
Czech Republic
Chile
Morocco
Ukraine
Portugal
Thailand
Hong Kong
Israel
Fiji
Sweden
Greece
Norway
Algeria
Austria
Viet Nam
$0
$1,000,000
$2,000,000
$3,000,000
$4,000,000
$5,000,000
$6,000,000
$7,000,000
$8,000,000
$9,000,000

“The winning Olympic 100-metre sprint times for men (blue points) and women (red points), with superimposed best-fit linear regression lines (solid black lines) and coefficients of determination.”
Crossover at year 2156 suggests women will run faster than men in 2156 Olympics.

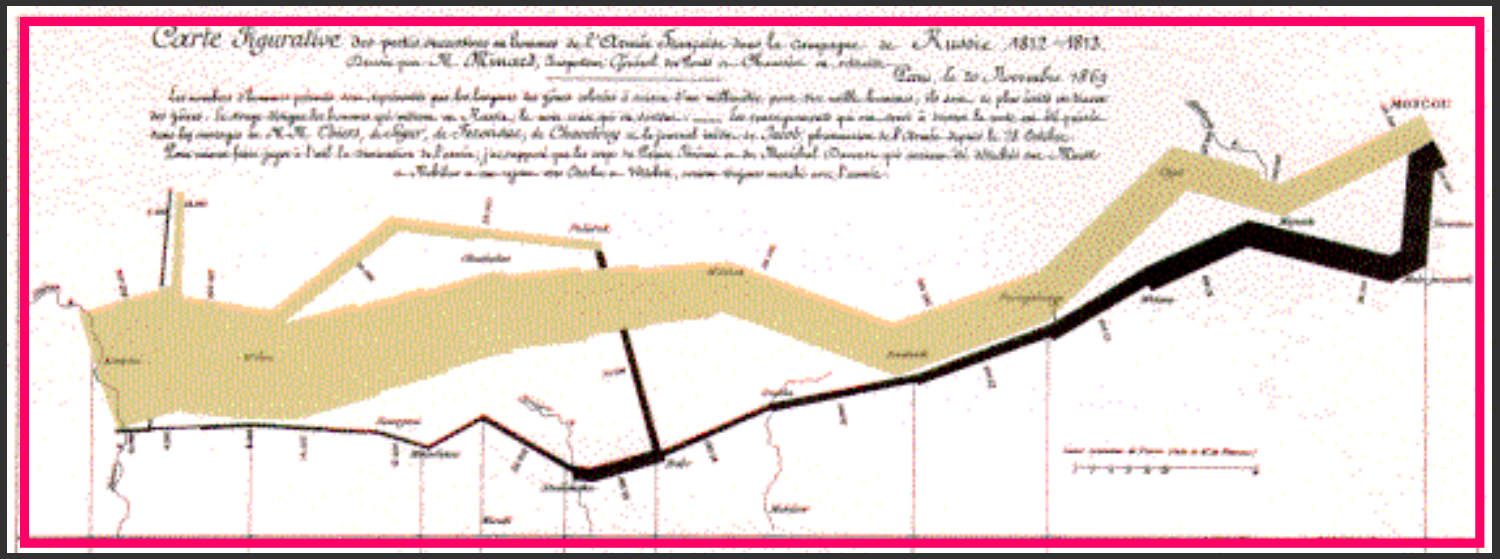


https://doi.org/10.1038/431525a

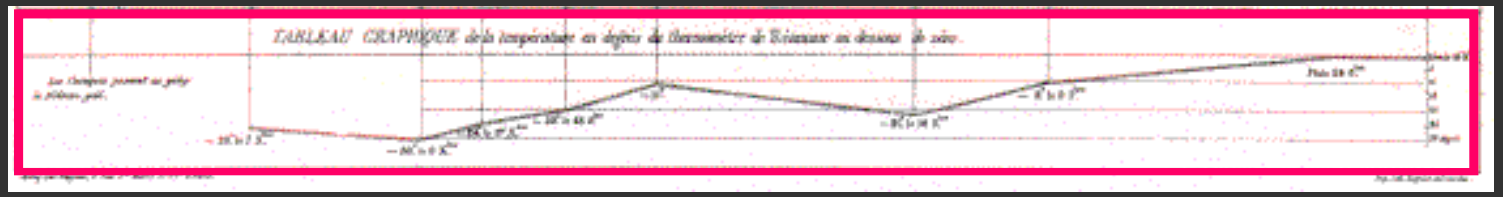
Minard 1869: Napoleon's march



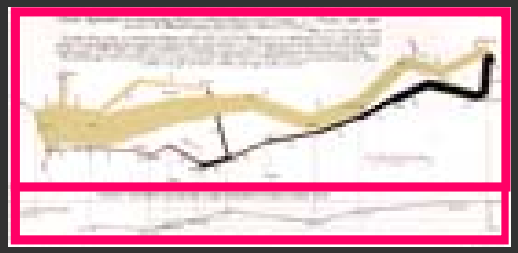
Single axis composition



+



=

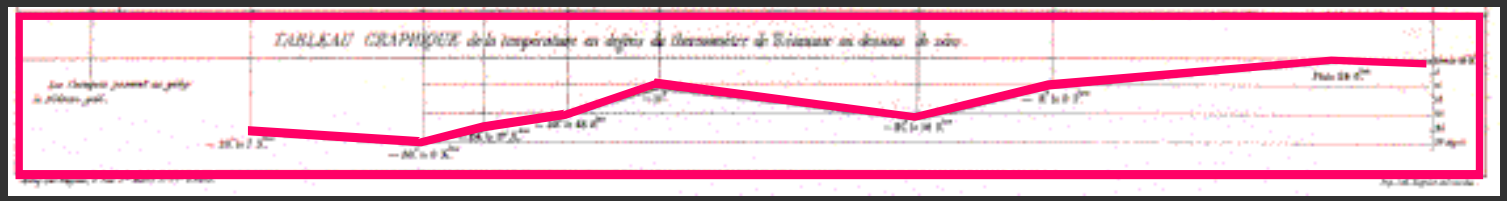


[based on slide from Mackinlay]

Mark composition

y-axis: temperature (Q)

\+ x-axis: longitude (Q) / time (O)

\= 

temp over space/time (Q x Q)

[based on slide from Mackinlay]

Mark composition

y-axis: longitude (Q)

\+ x-axis: latitude (Q)

\+ width: army size (Q)

\+

=

army position (Q x Q) and army size (Q)

[based on slide from Mackinlay]

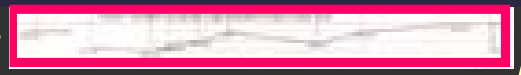
longitude (Q)

latitude (Q)

army size (Q)

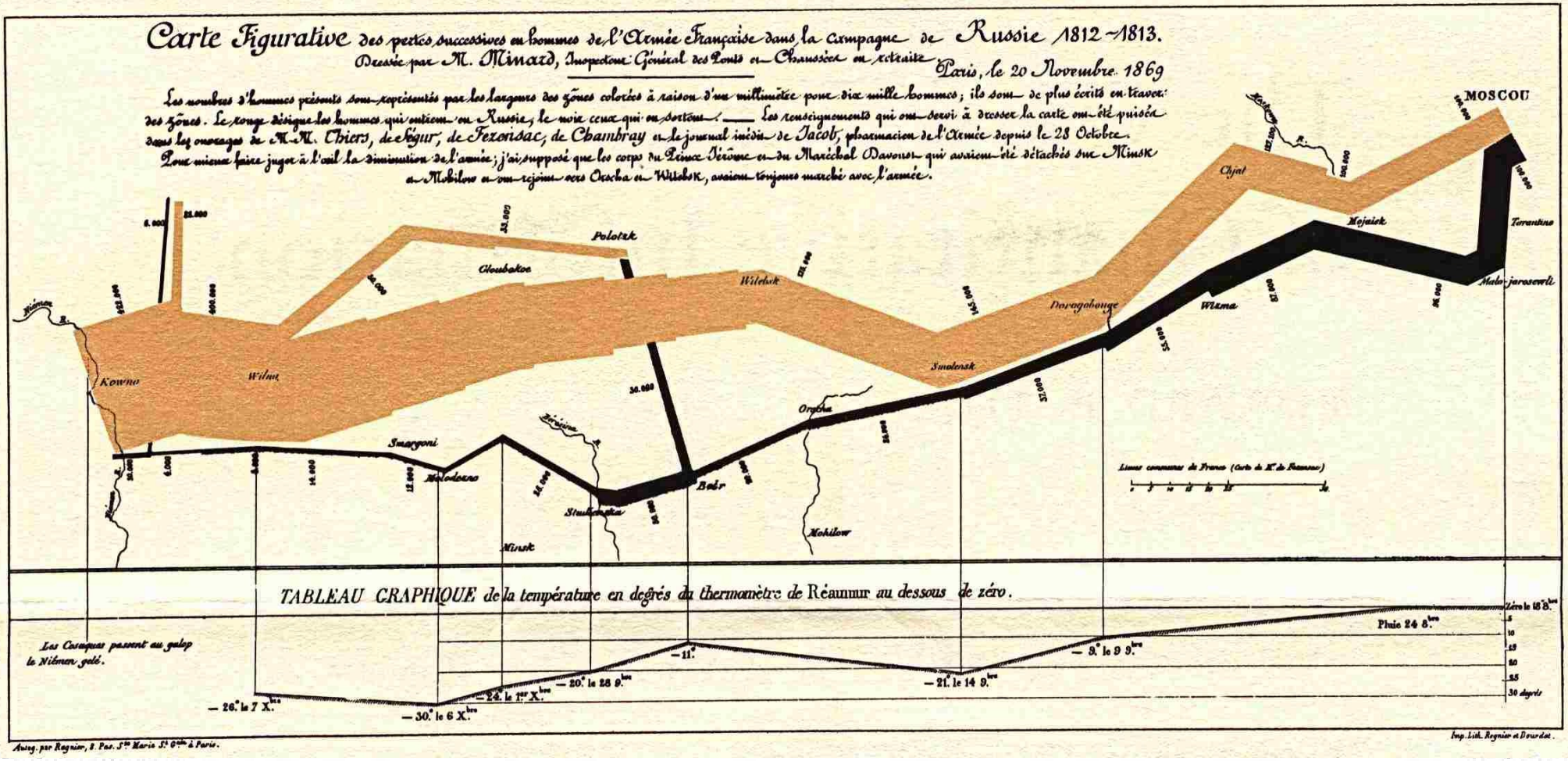
temperature (Q)

latitude (Q) / time (O)



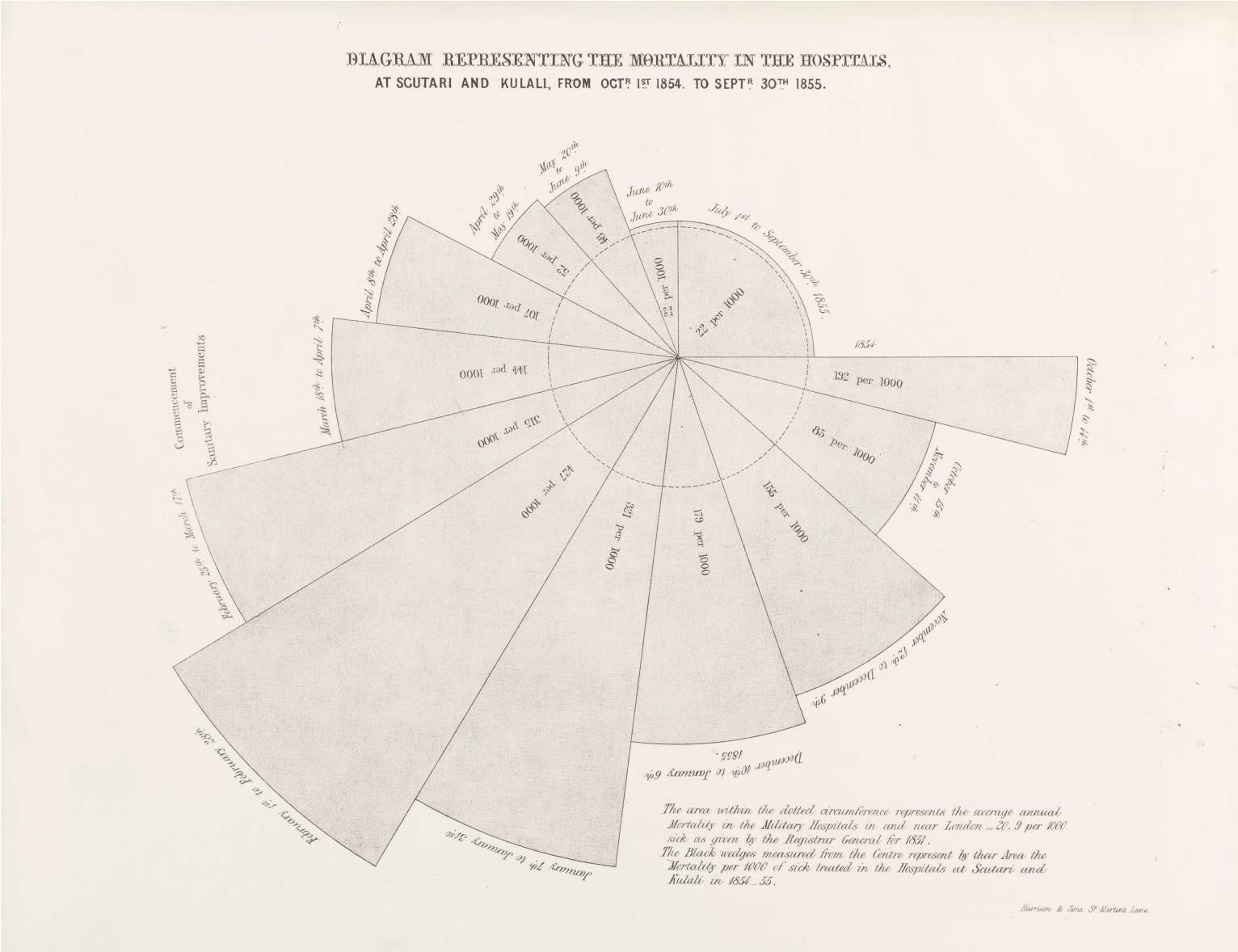
[based on slide from Mackinlay]

Minard 1869: Napoleon's march

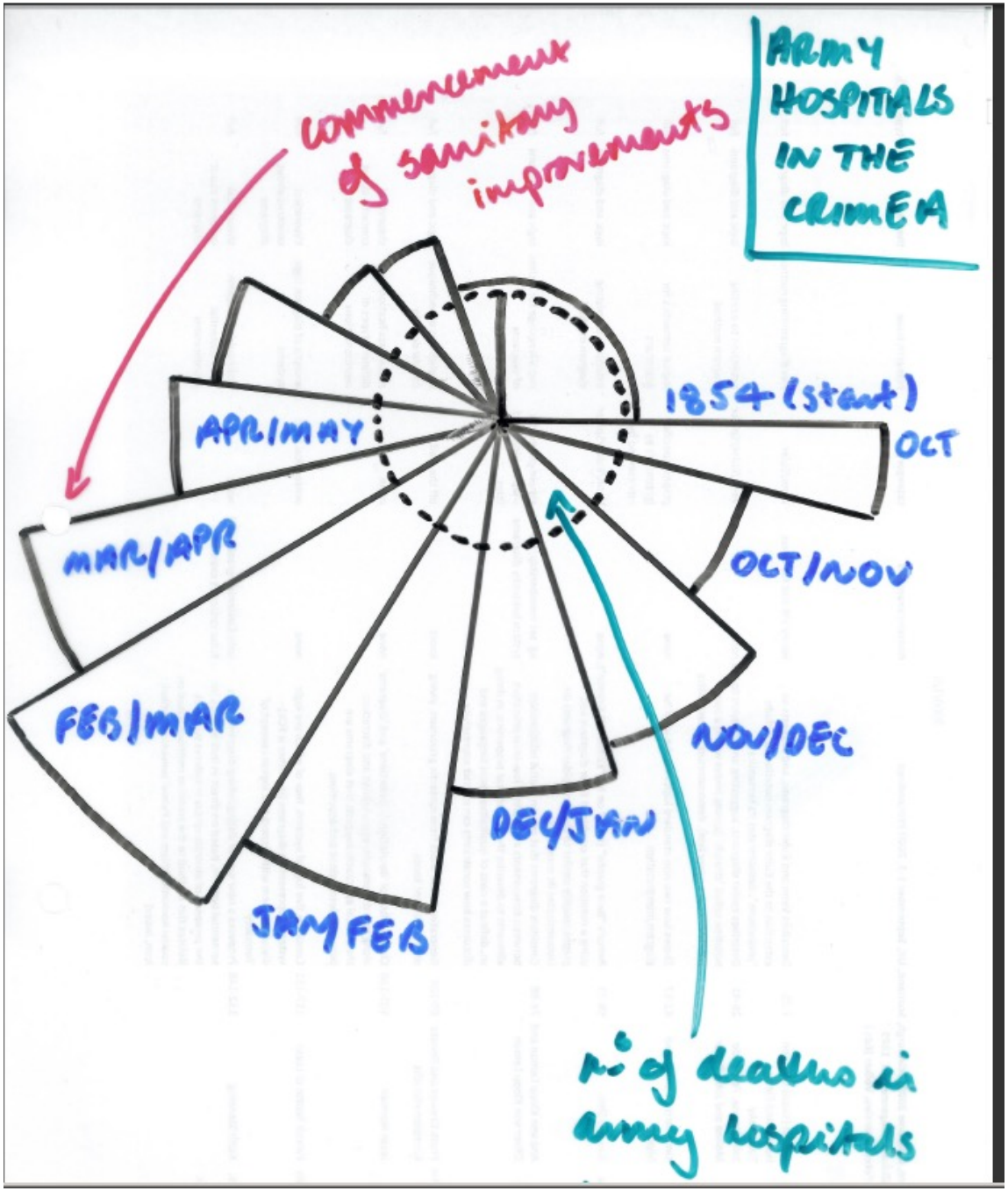


Depicts at least 5 quantitative variables. Any others?

DIAGRAM REPRESENTING THE MORTALITY IN THE HOSPITALS AT SCUTARI AND KULALI FROM OCT^{R} 1ST 1854 TO SEPT^{R} 30TH 1855



commons.wikimedia.org



Effect of Sanitation

Dotted area: *average annual Mortality in the Military Hospitals in and near London – 20.9 per 1000 sick as given by the Registrar General for 1854*

Black wedges*: mortality per 1000 of sick treated in the Hospitals at Scurati and Kumail in 1854-1855*

Data from October 1st 1854 to September 30th 1855

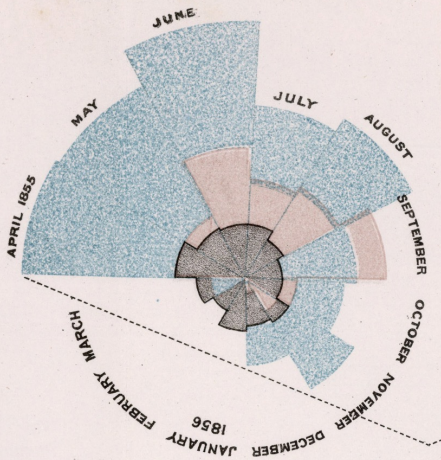
Or we could look at the same data like this…

Oct1- Oct 14	Oct 15– Nov 12	Nov 12– Dec 9	Dec 10– Jan 6	Jan 7– Jan 31	Feb 1 – Feb 28	Feb 29– Mar 17
192	**85**	**155**	**179**	**321**	**427**	**315**

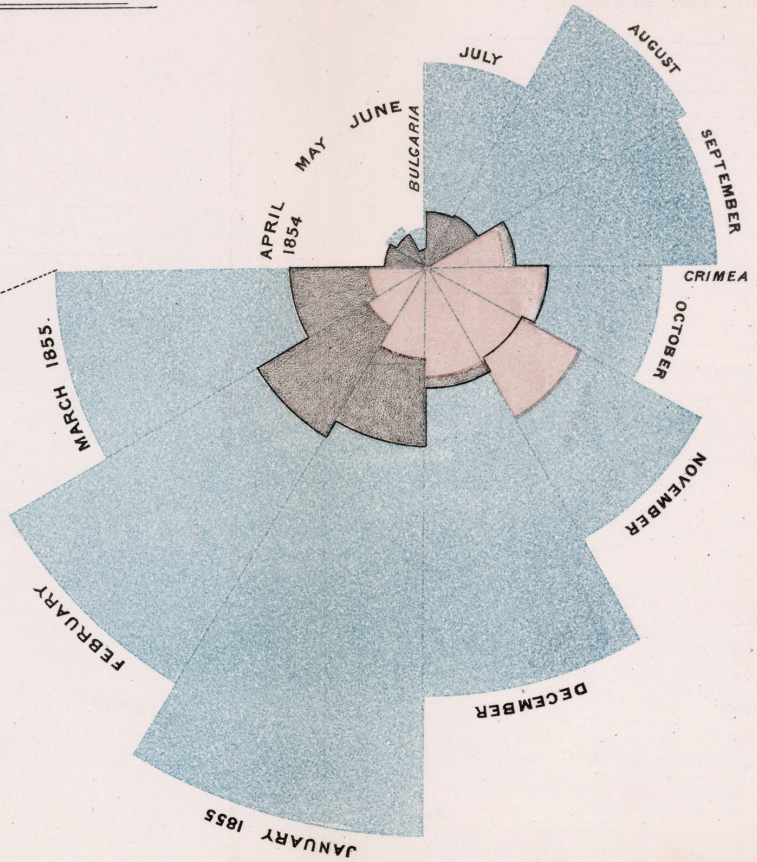
Mar 18 – April 7	April 8– April 28	April 29 – May 19	May 20 – June 9	June 10 – June 30	July 1 – Sept 29
144	**107**	**52**	**48**	**22**	**22**

DIAGRAM of the CAUSES of MORTALITY IN THE ARMY IN THE EAST.

2.
APRIL 1855 to MARCH 1856.



1.
APRIL 1854 to MARCH 1855.



The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

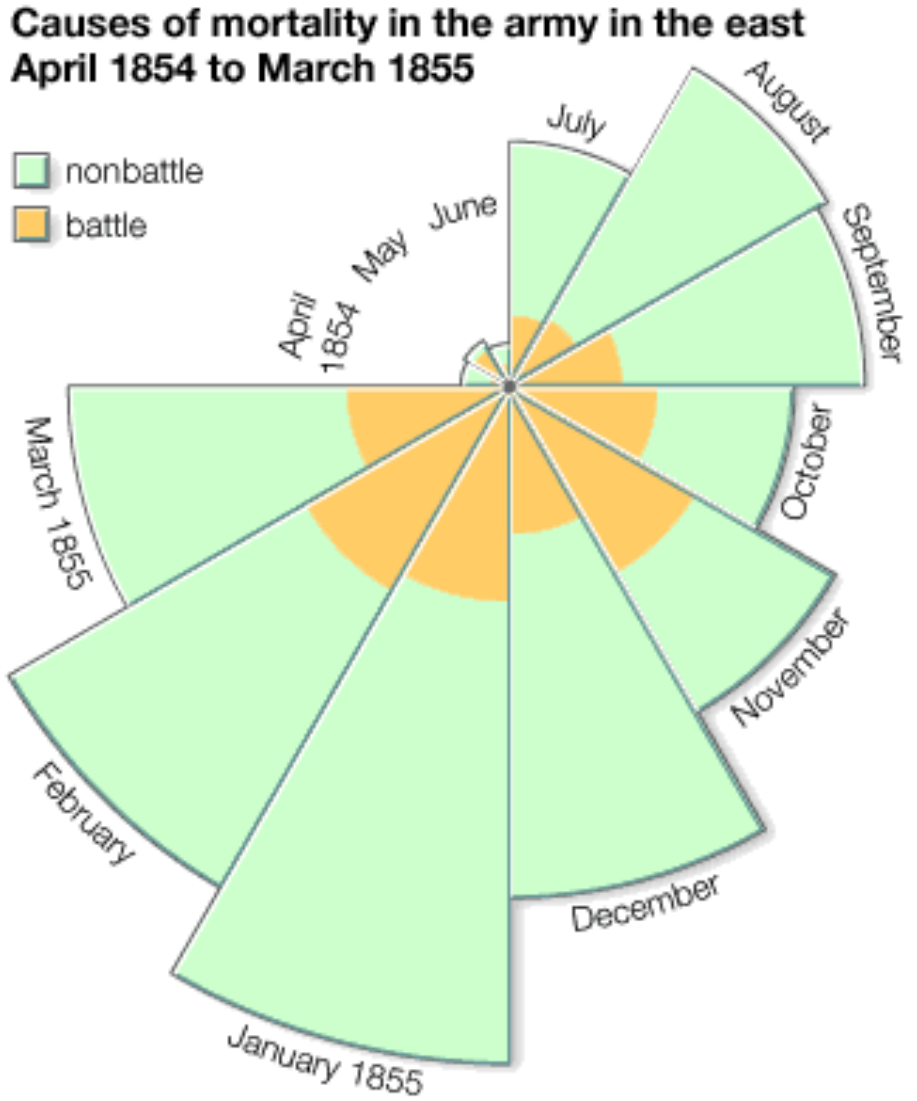
The black line across the red triangle in Nov.r 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red; in January & February 1855, the blue coincides with the black.

The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Harrison & Sons, St. Martin's Lane.

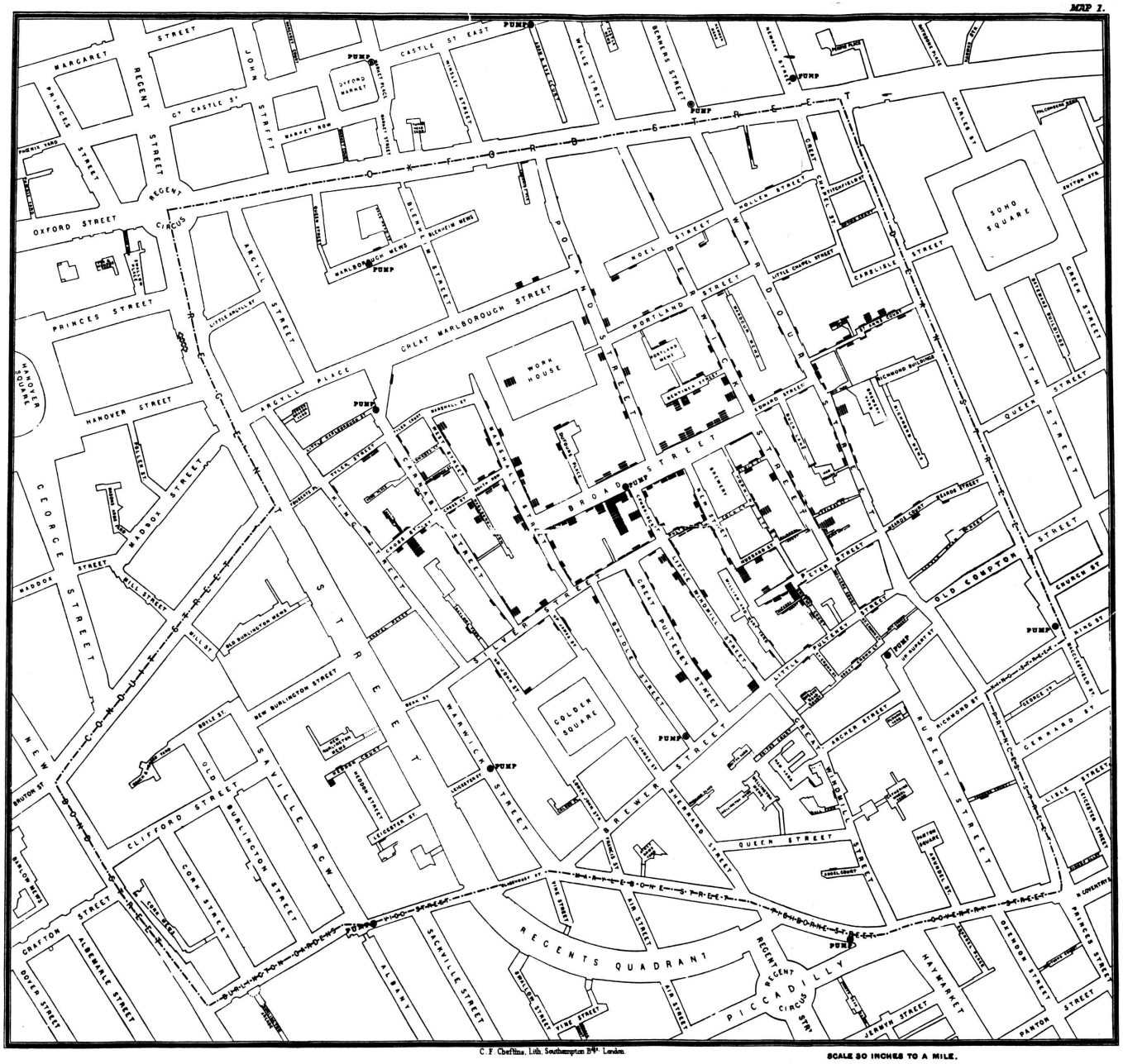
The "Coxcomb" diagram showed how preventable deaths, outside of battle, were overwhelming… and needed action



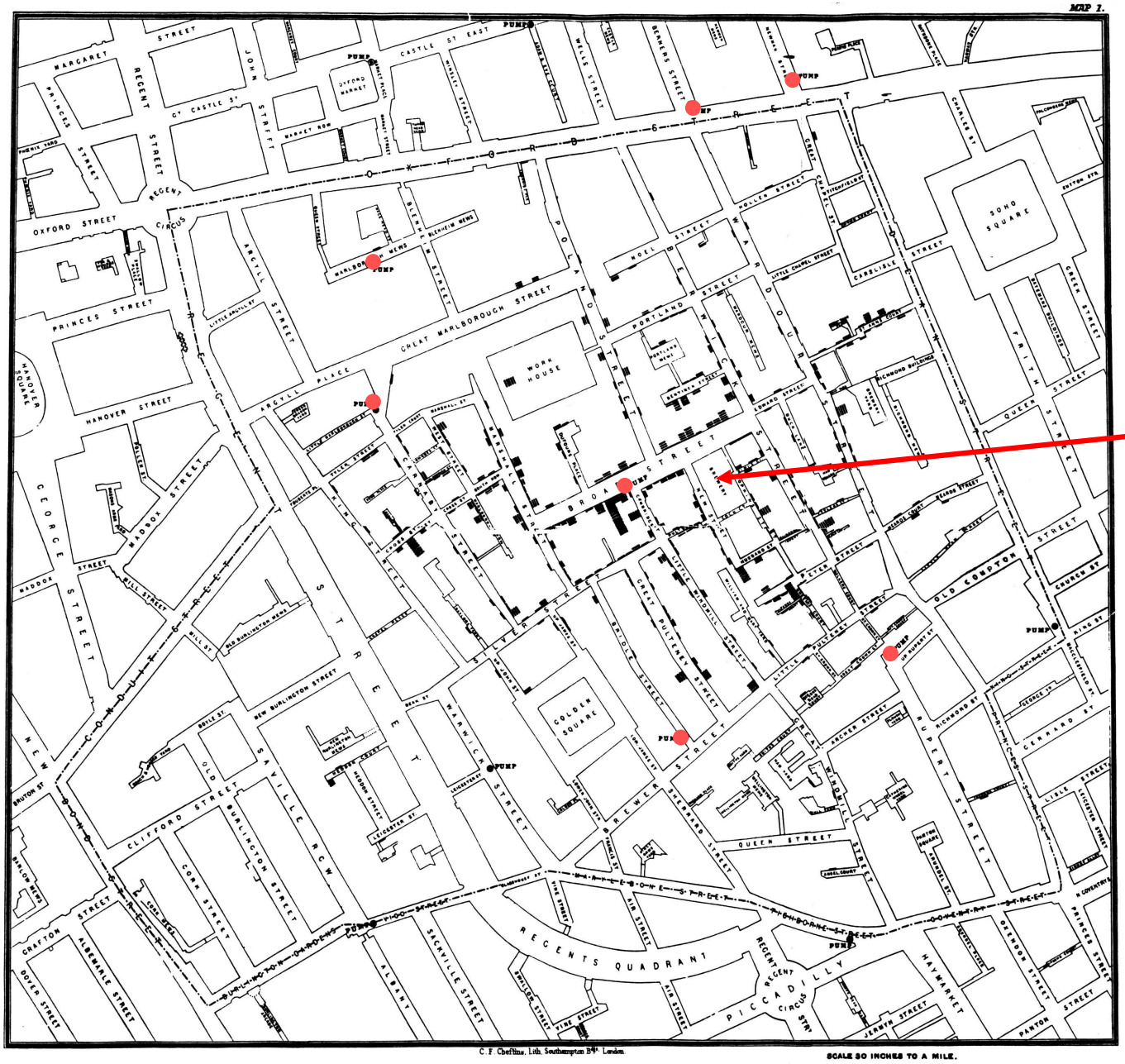
Based on Florence Nightingale's "Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army," 1858.

Preventable deaths

https://kids.britannica.com/students/assembly/view/70822



Snow, J. (1855). On the Mode of Communication of Cholera. (secondary source)



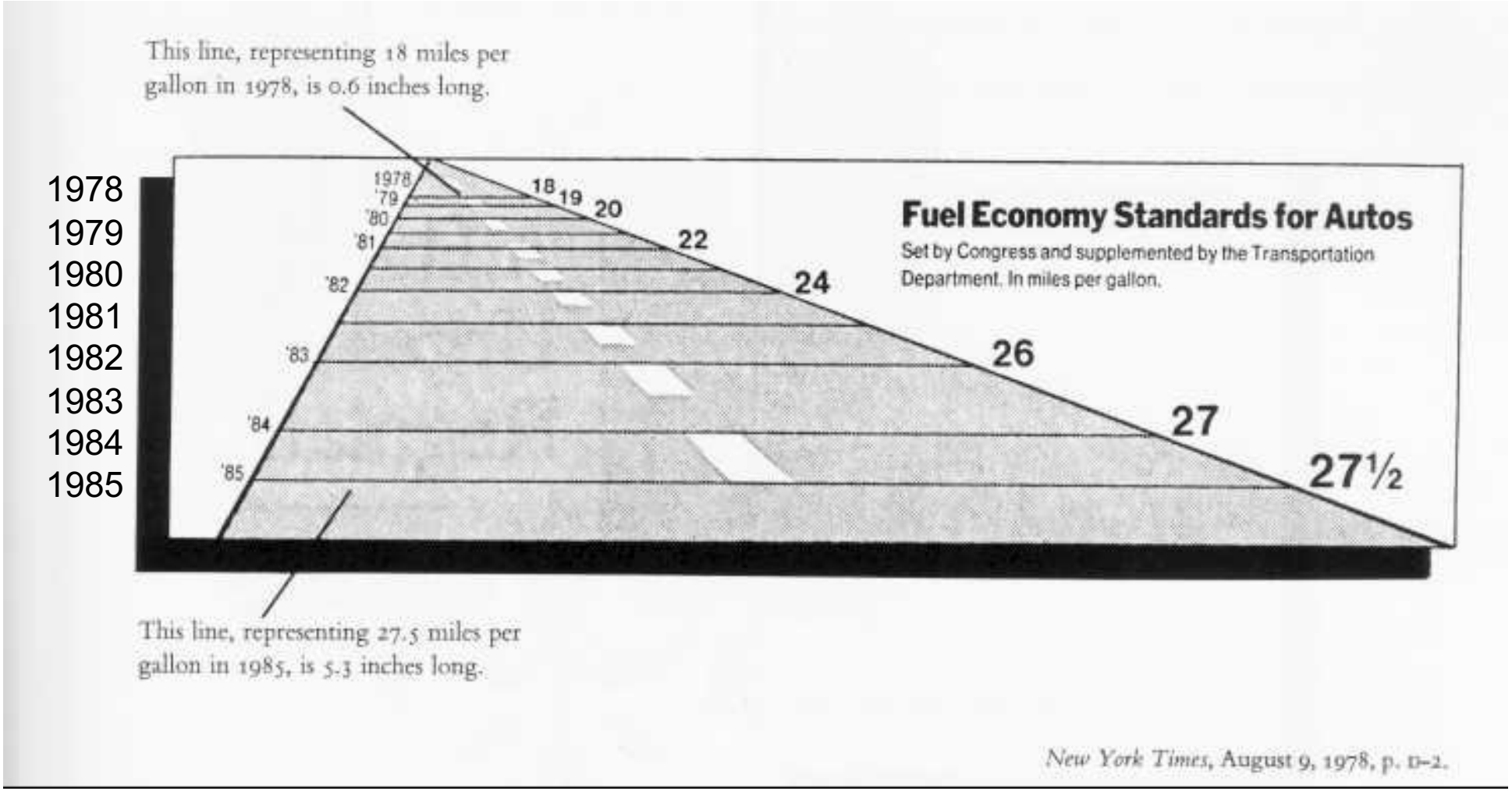
Snow, J. (1855). On the Mode of Communication of Cholera. (secondary source)

The Lie factor

A simple measure or description of how much a chart or visualization is not telling the truth!

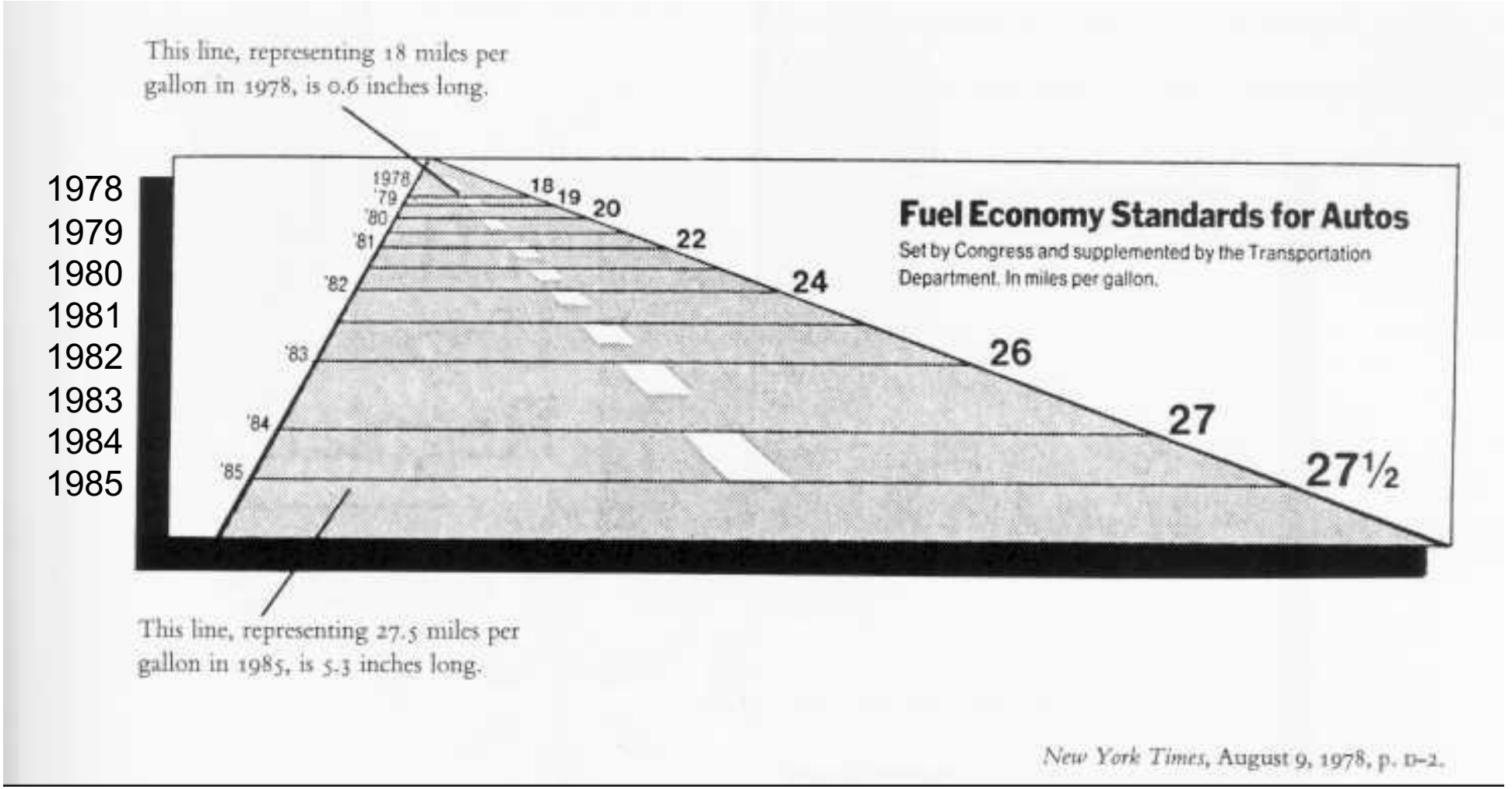
$$\text{Lie factor} = \frac{\text{size of effect shown in visualisation}}{\text{size of effect in data}}$$

Tufte (1983), p. 57

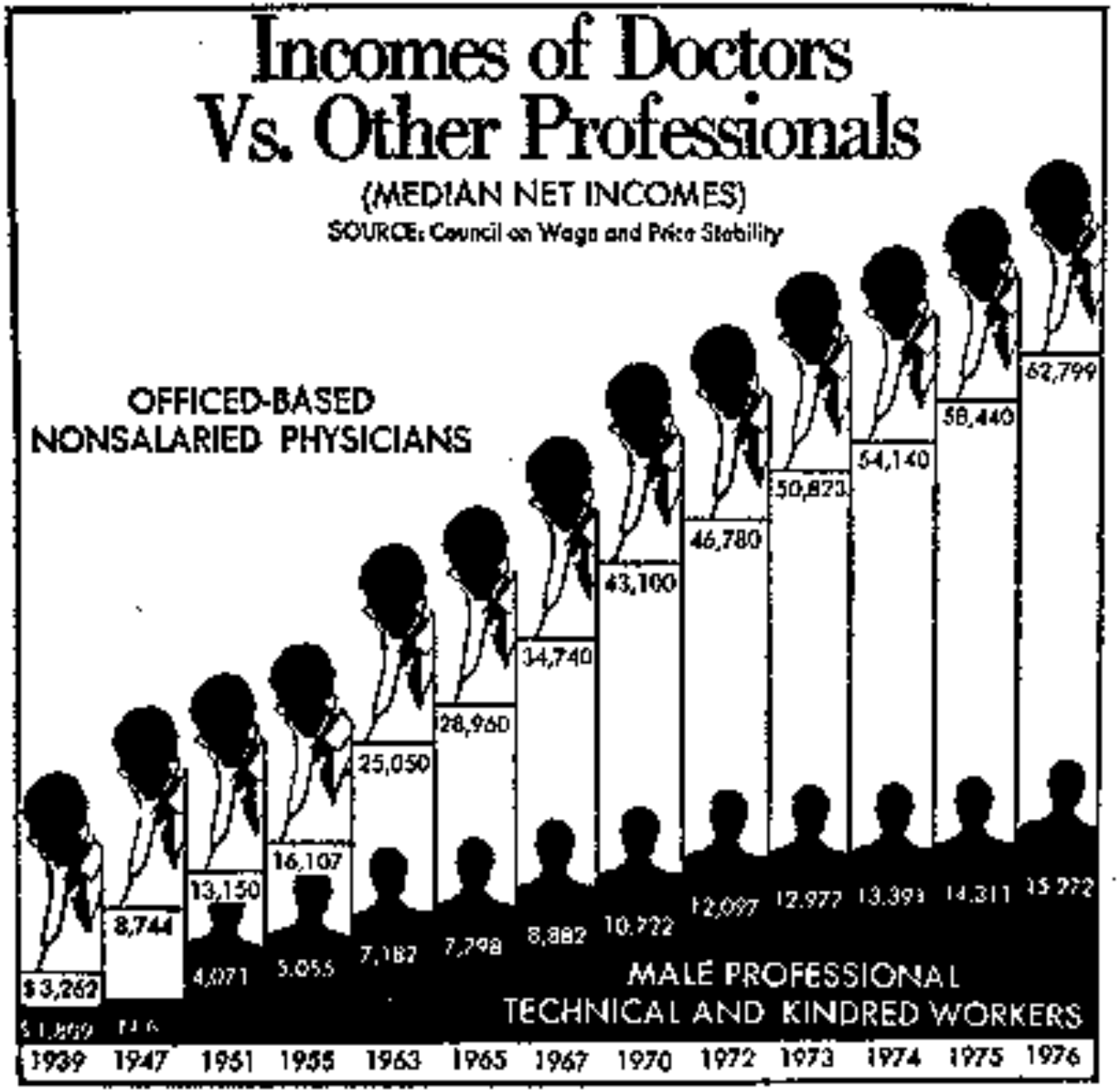


Increase in the value of the fuel economy standard over time
Most recent years first
Suggests a very high increase

Tufte (1983), p. 57



Numerical increase: from 18m/g to 27.5m/g (53%)
Line length increase: from 0.6in to 5.3in (738%)
Suggests increase has been greater than it has
Lie factor = 783/53=14.8

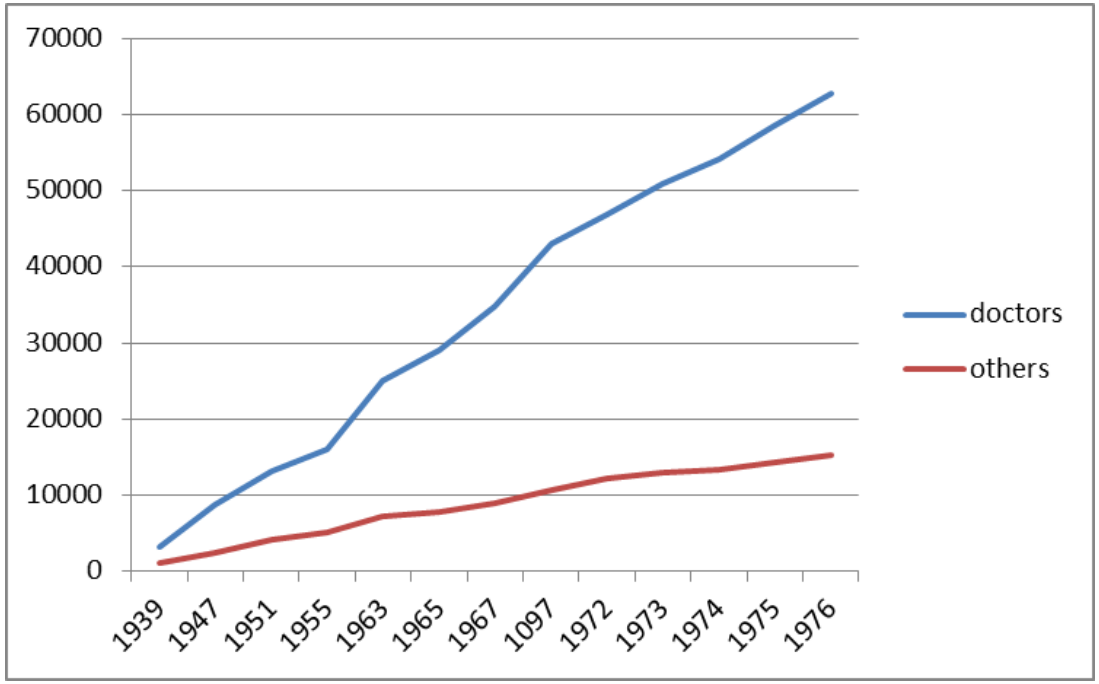


1939-1976

Doctors' salaries

Technicians' salaries

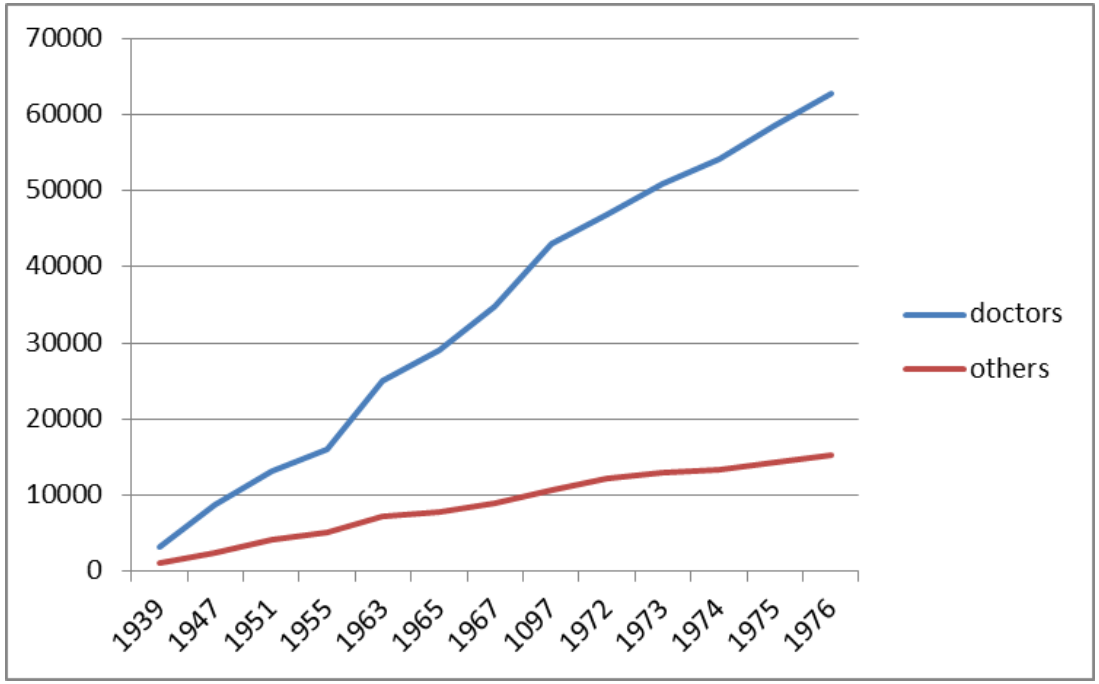
Washington Post, https://www.datavis.ca/gallery/lie-factor.php



Plotting the raw numbers

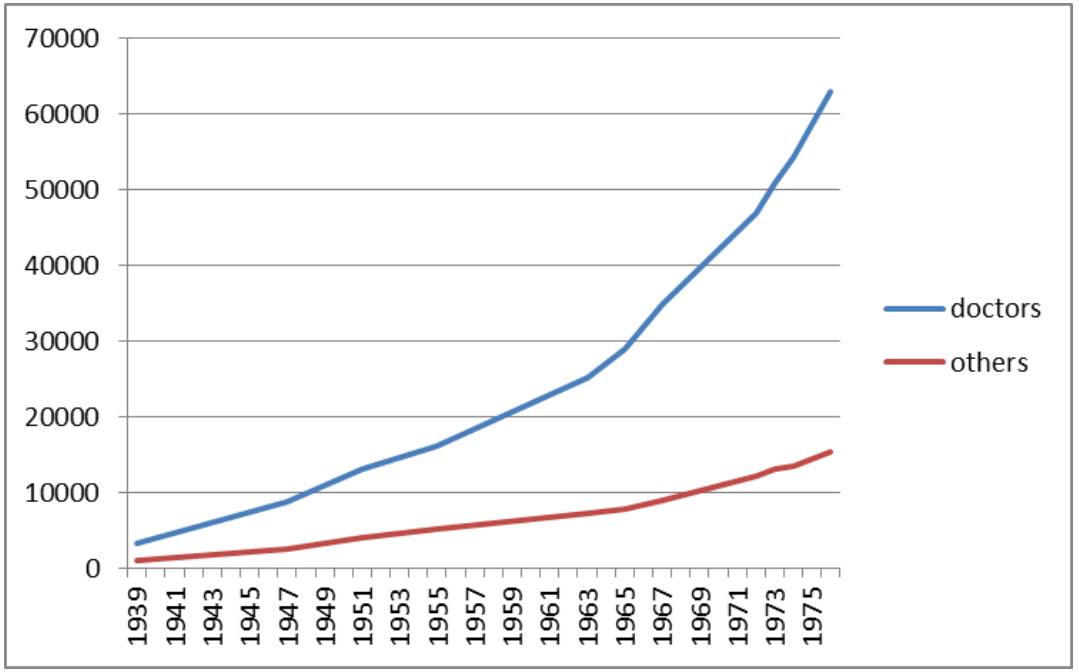
Difference between years on the x-axis:
8,3,4,8,2,2,3,2,1,1,1,1

A “Rubber band scale”



Plotting the raw numbers with the given scale

Linear time scale

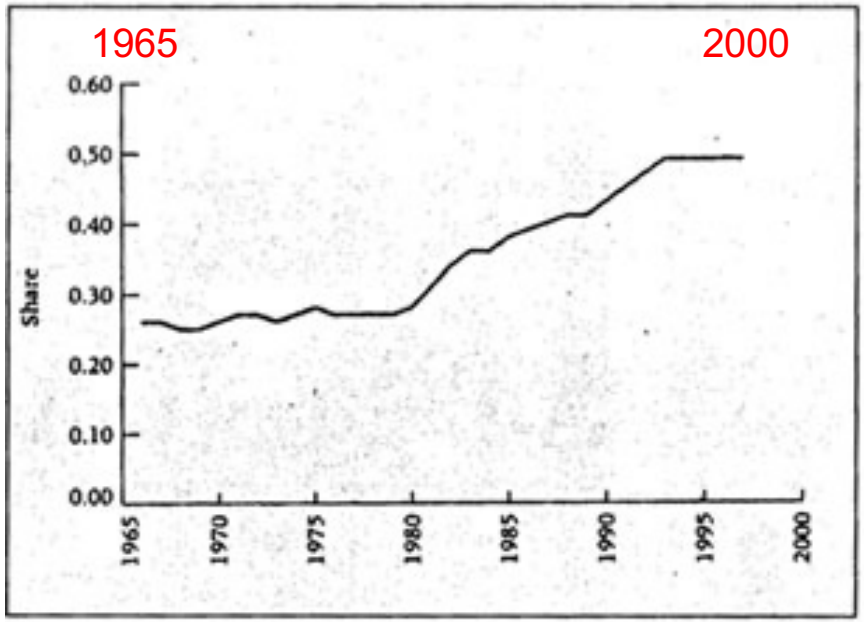


Cornell University's Tuition

Students are paying more for less??

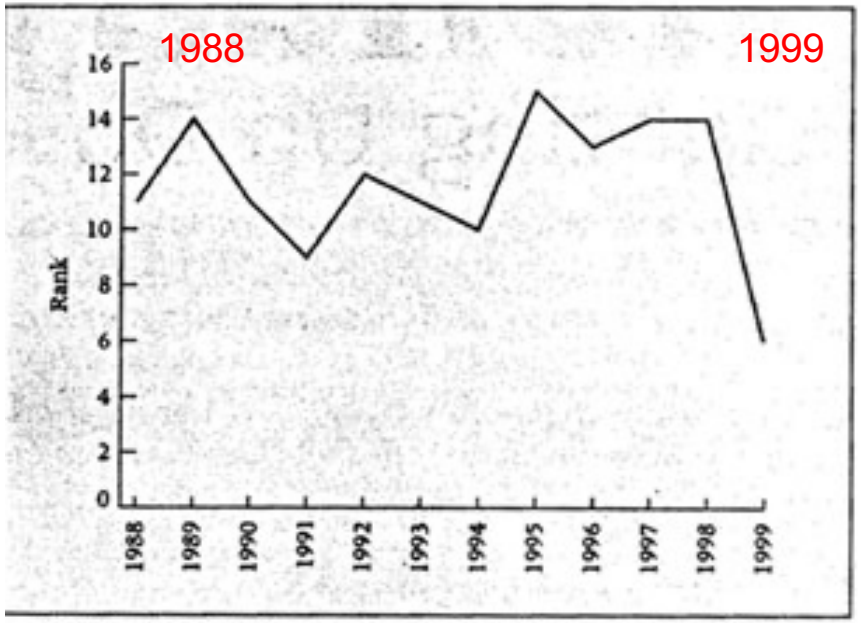
Cornell University's ranking

Ithaca Times, Dec 7, 2000; http://euclid.psych.yorku.ca/SCS/Gallery/context.html



BY THE NUMBERS: OVER 35 YEARS, CORNELL'S TUITION HAS TAKEN AN INCREASINGLY LARGER SHARE OF ITS MEDIAN STUDENT FAMILY INCOME.

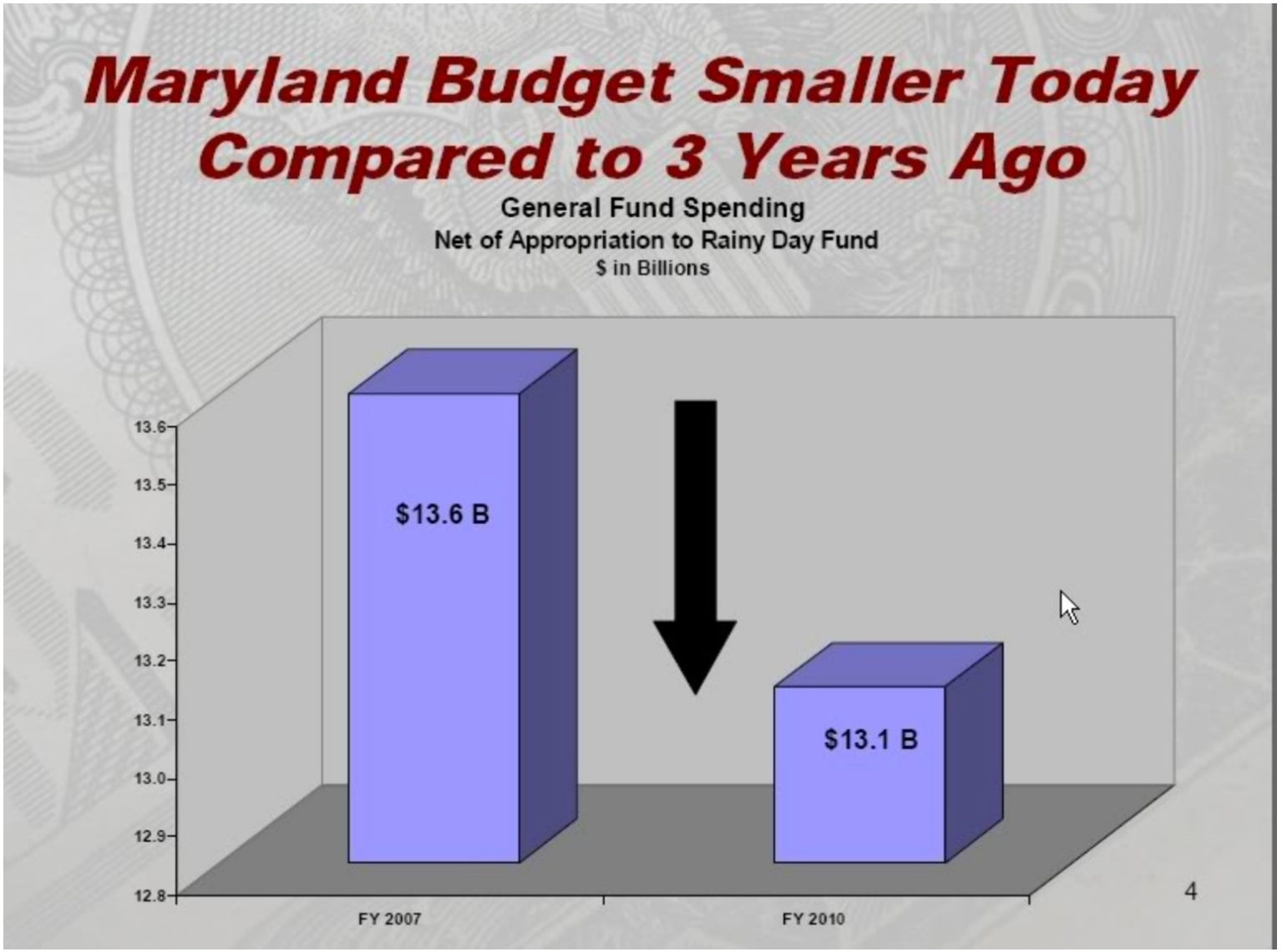
Tuition fees – over 35 years



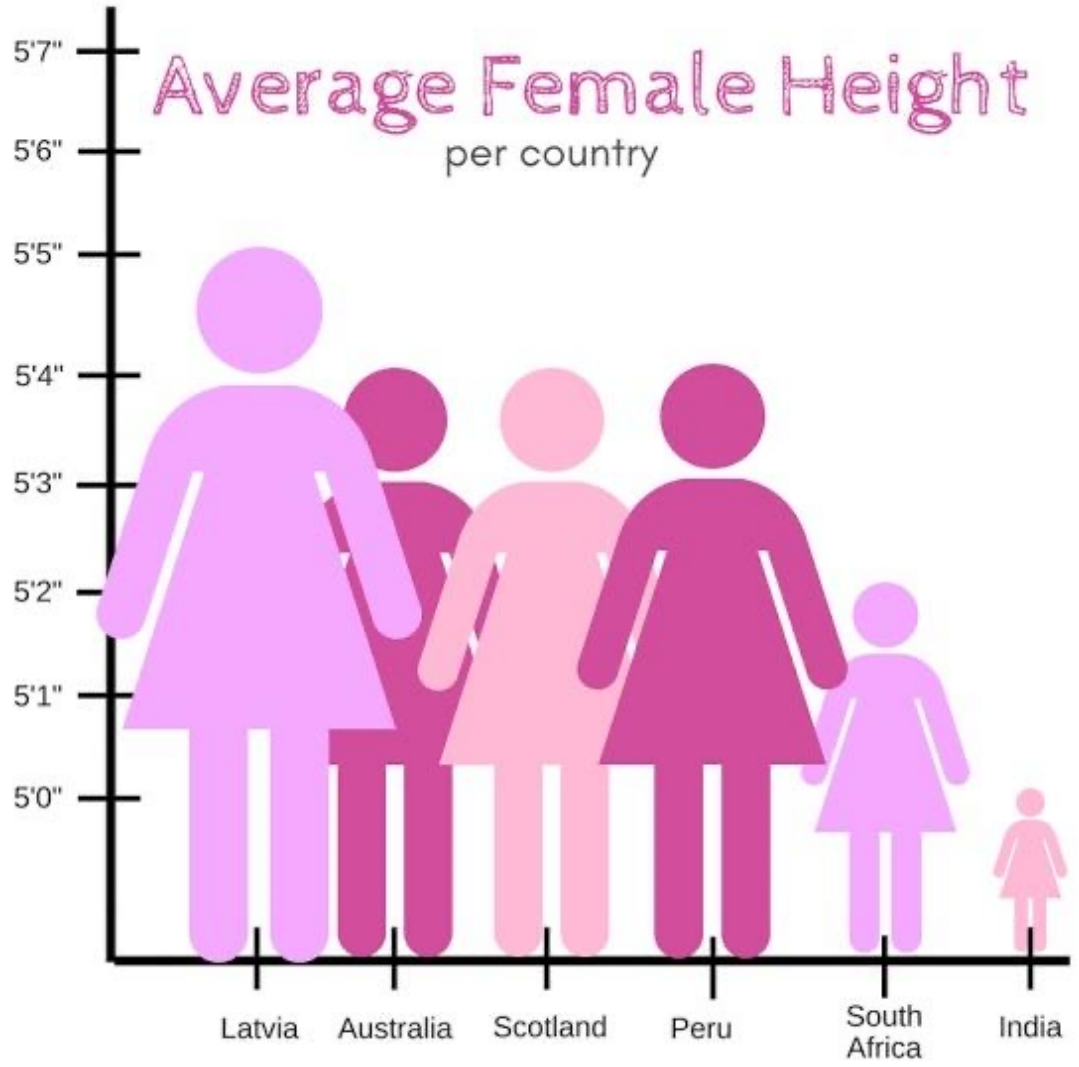
PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN *US NEWS & WORLD REPORT* HAS RISEN AND FALLEN ERRATICALLY.

Ranking – over 11 years

http://euclid.psych.yorku.ca/SCS/Gallery/context.html



http://www.dailykos.com/story/2009/12/12/809947/-Gov-t-Transparency-and-Chartjunk-in-Maryland#



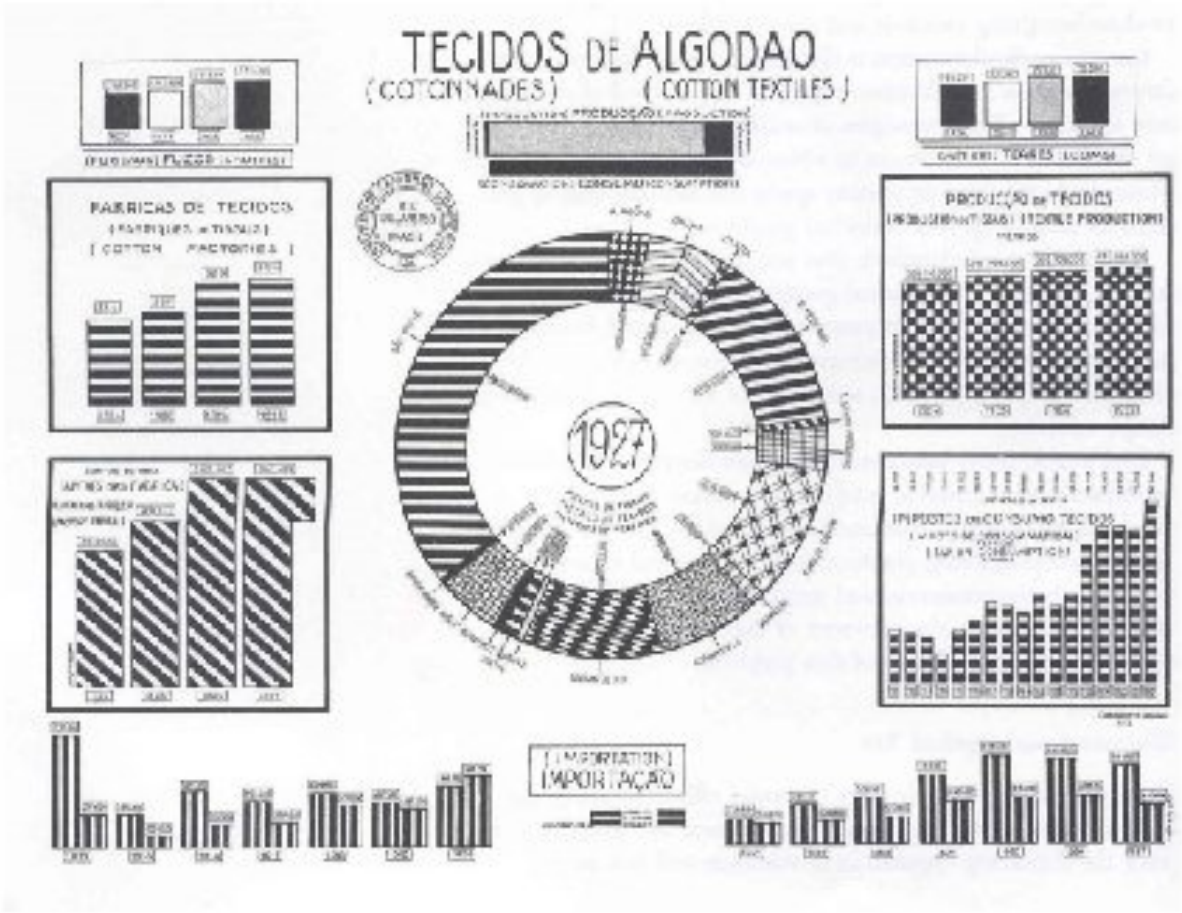
Via https://twitter.com/lizardbill/status/1127005323636686848?s=12

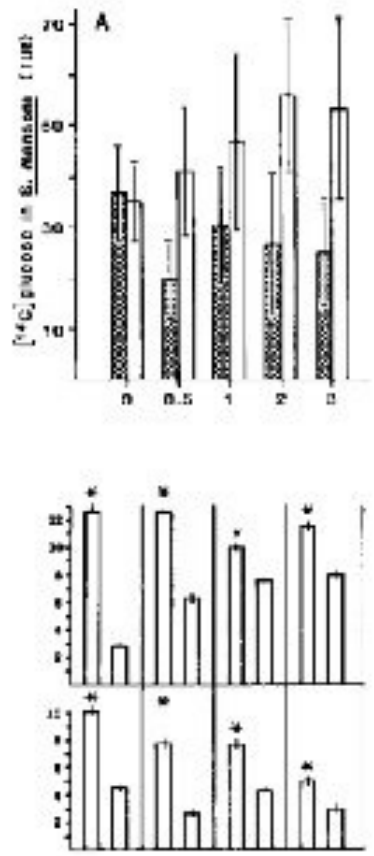
Chart Junk

Stuff that obscures the real information in the interests of 'prettiness'

The use of unnecessary data-ink

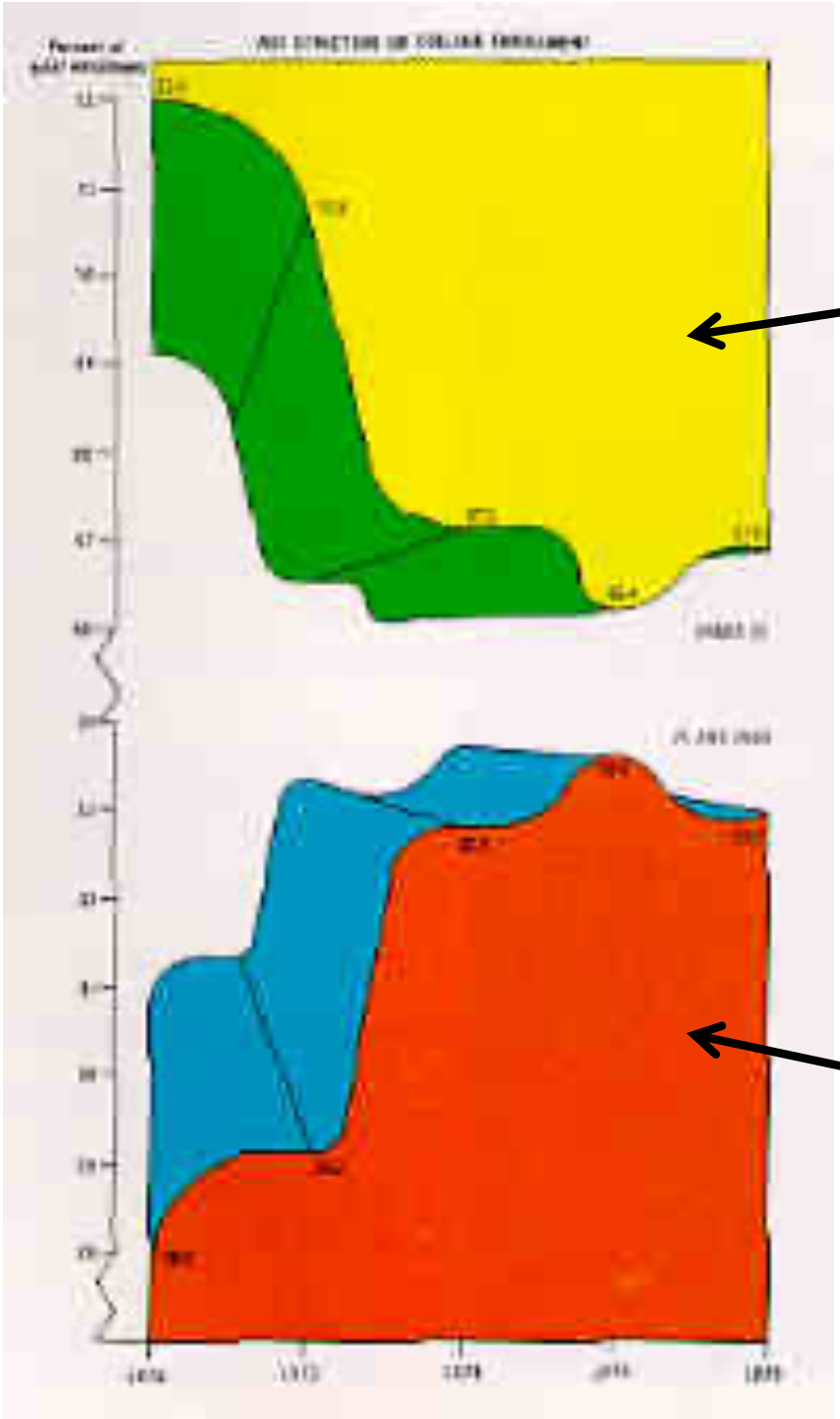
Tufte (1983), p. 41





Moiré effect: "the design interacts with the physiological tremor of the eye to produce the distracting appearance of vibration and movement"

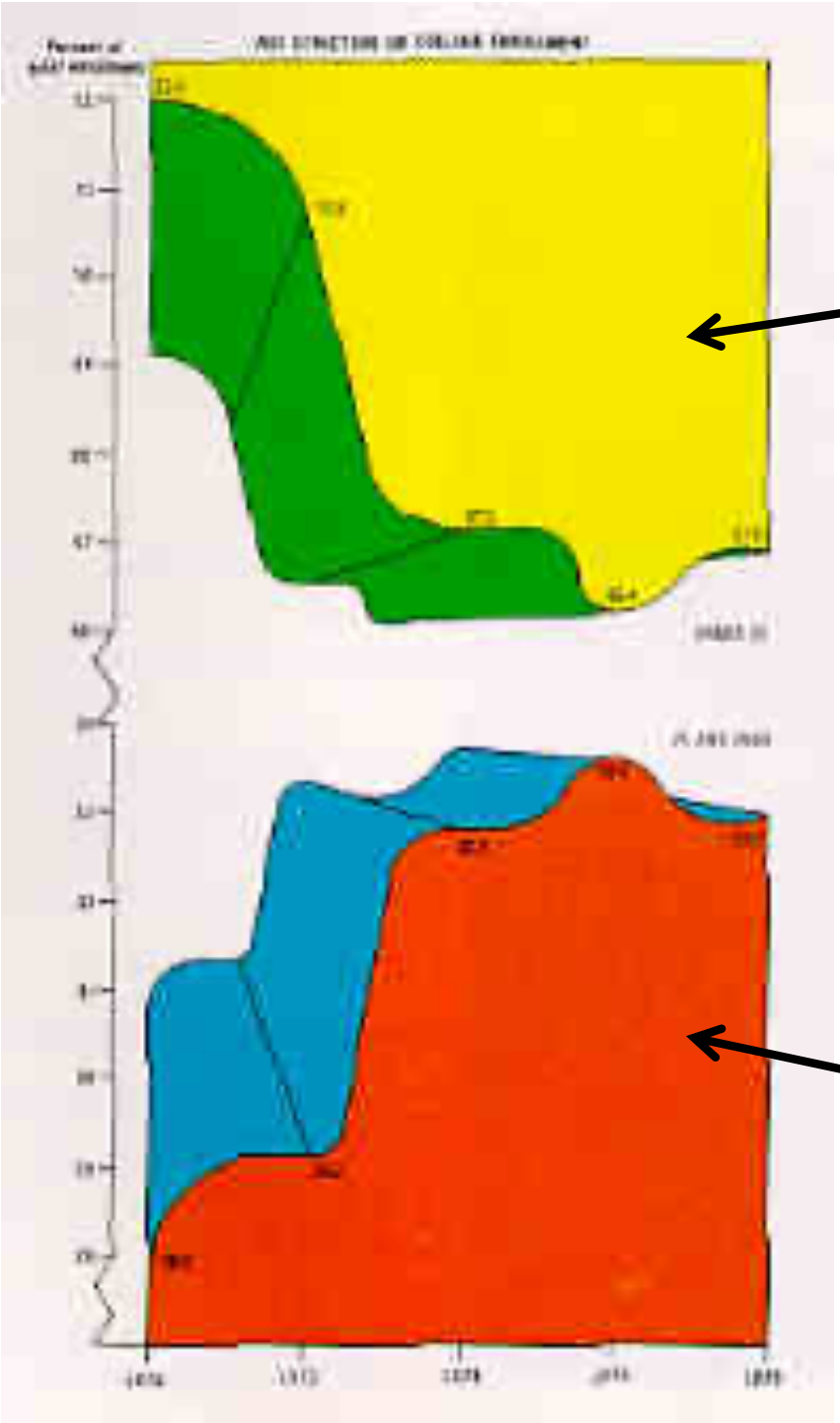
Tufte (1983), p. 107



under 25

Age structure of college enrolment, showing percent of total enrolment

25 and over

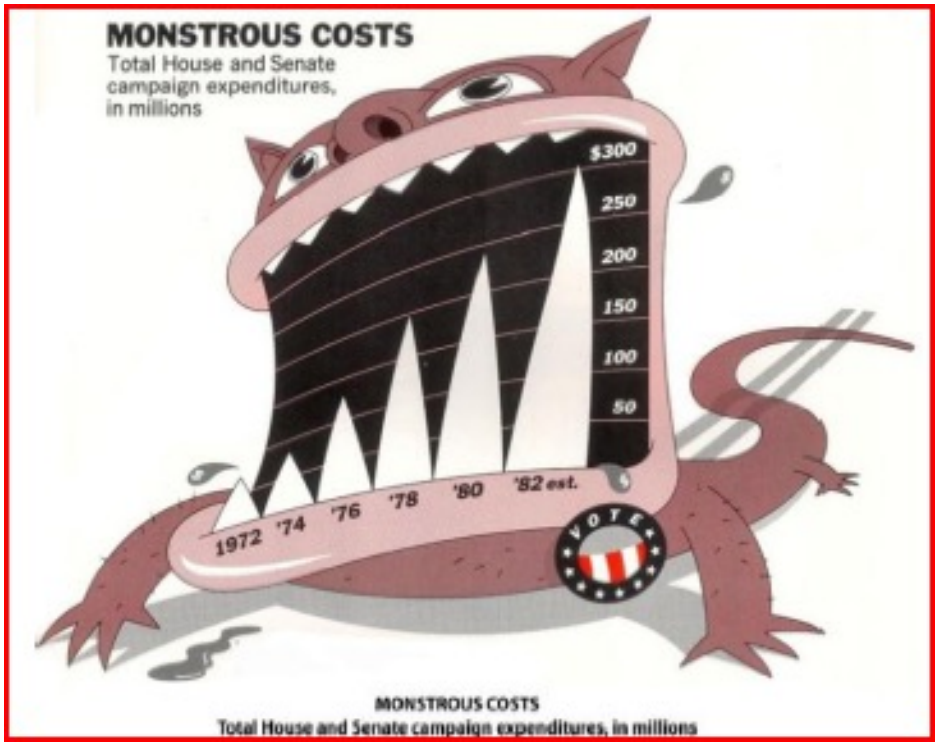


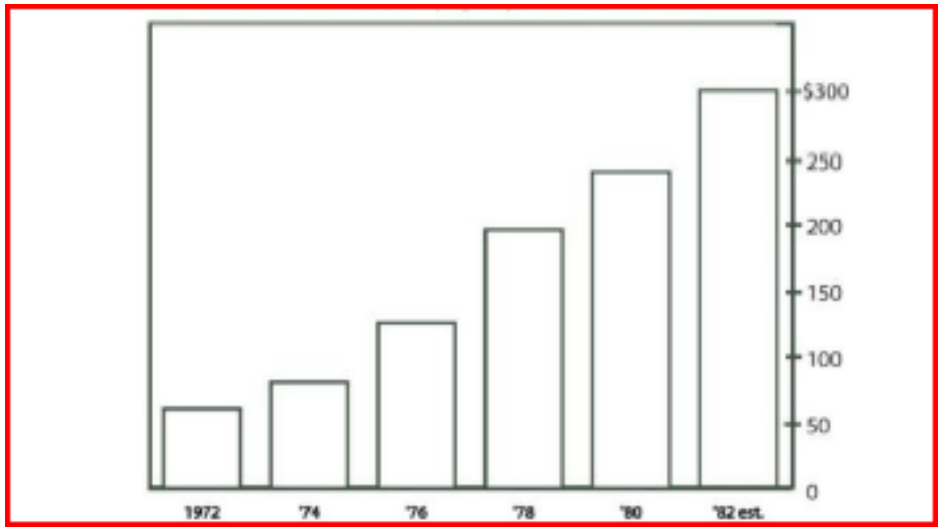
under 25

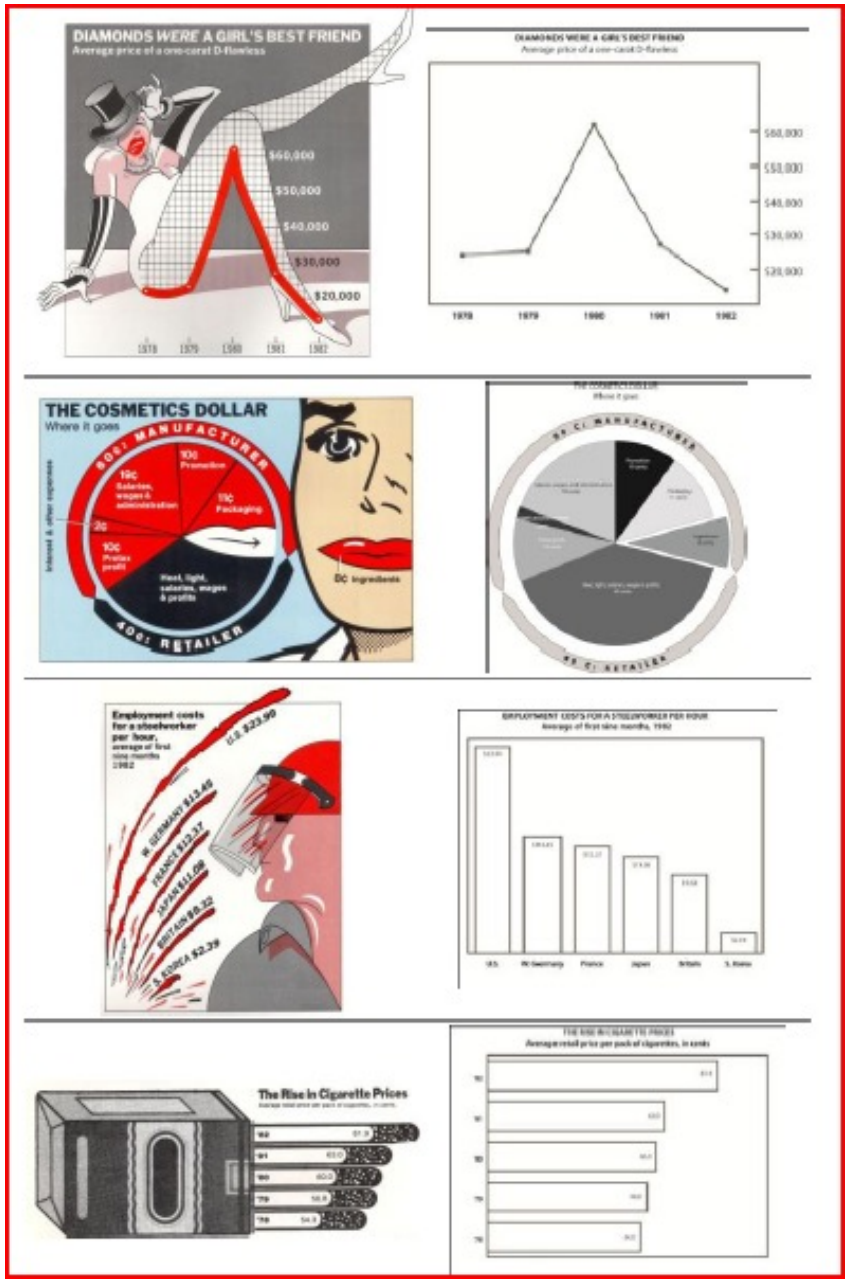
25 and over

Age structure of college enrolment, showing percent of total enrolment

1972	1973	1974	1975	1976
28	29.2	32.8	33.6	33

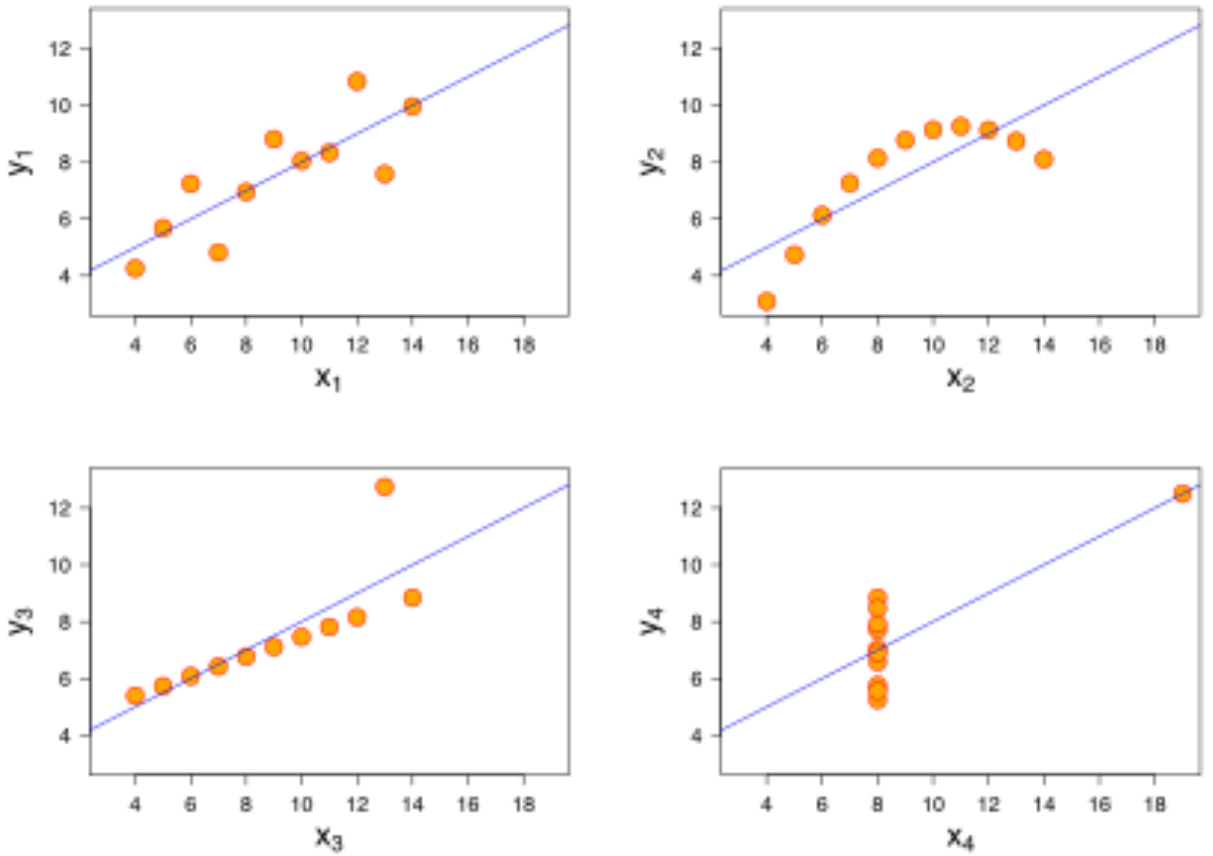




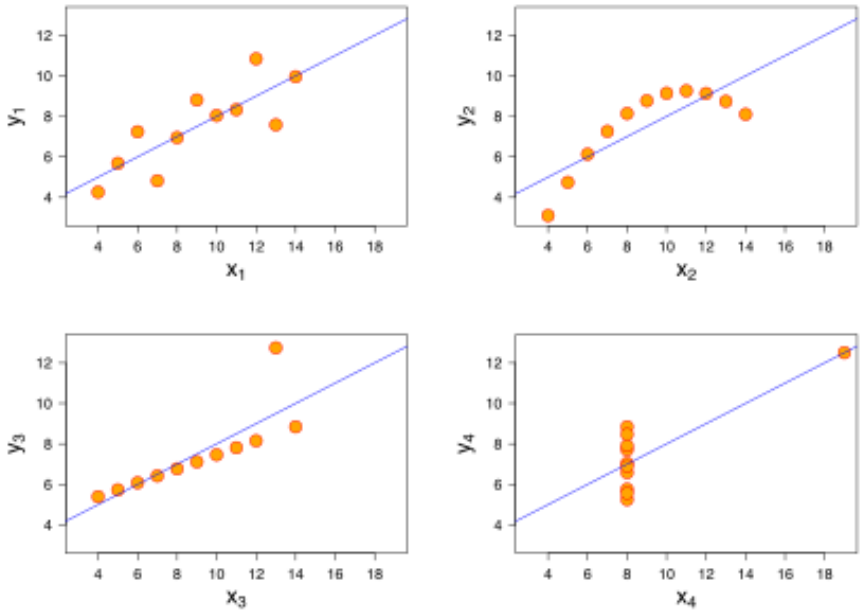


Bateman et al., (2010), “Useful Junk?”, CHI

Anscombe’s quartet



https://en.wikipedia.org/wiki/Anscombe's_quartet



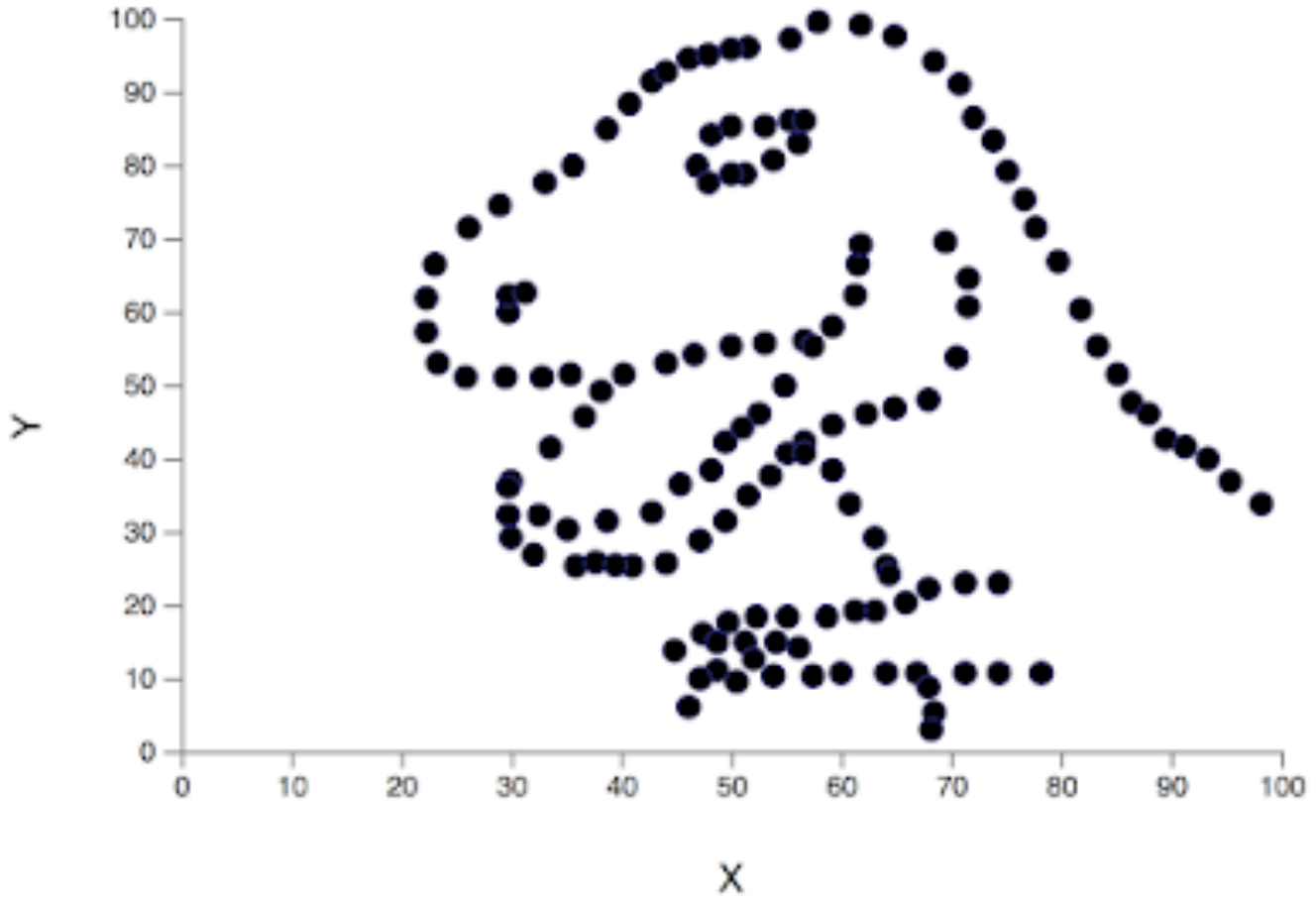
Property	Value	Accuracy
mean (x)	9	exact
variance (x)	11	exact
mean (y)	7.50	to 2 decimal places
variance (y)	4.125	±0.003
correlation between x and y	0.816	to 3 decimal places
linear regression line	y = 3.00 + 0.500x	to 2 and 3 decimal places, resp.
R^2	0.67	to 2 decimal places

Anscombe's quartet

I		II		III		IV	
x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

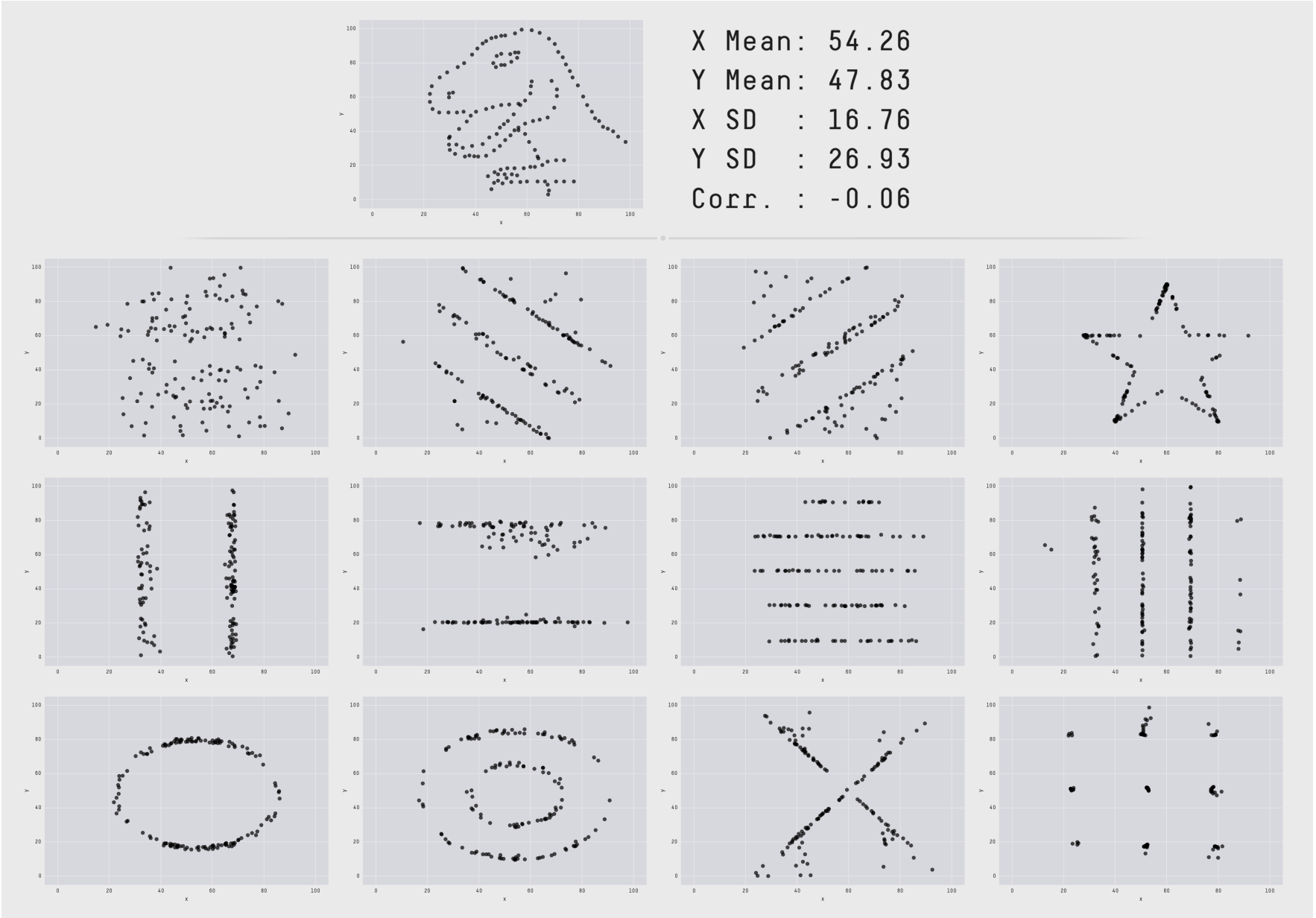
Paper/Site/Video:
‘Same stats, different graphs’

https://www.research.autodesk.com/publications/same-stats-different-graphs/



N = 142 ; X mean = 54.2633 ; X SD = 16.7651 ; Y mean = 47.8323 ; Y SD = 26.9354 ; Pearson correlation = -0.0645

First came a tweet by Alberto Cairo, an InfoVis academic and author, basically saying that one should never trust simple statistics alone… but combine statistics with visualization. The tweet has been deleted, but it is explained by him here.



Matejka, J. & Fitzmaurice, G. (2017). Same stats, different graphs, CHI, 2017

https://www.research.autodesk.com/publications/same-stats-different-graphs/

Information Visualisation Examples