

How to data 101

How to scrape, clean, and visualise data.

Scrape. theory + python
Clean. python
Visualise. theory + rawGraphs+

Data Visualization

Data Visualization

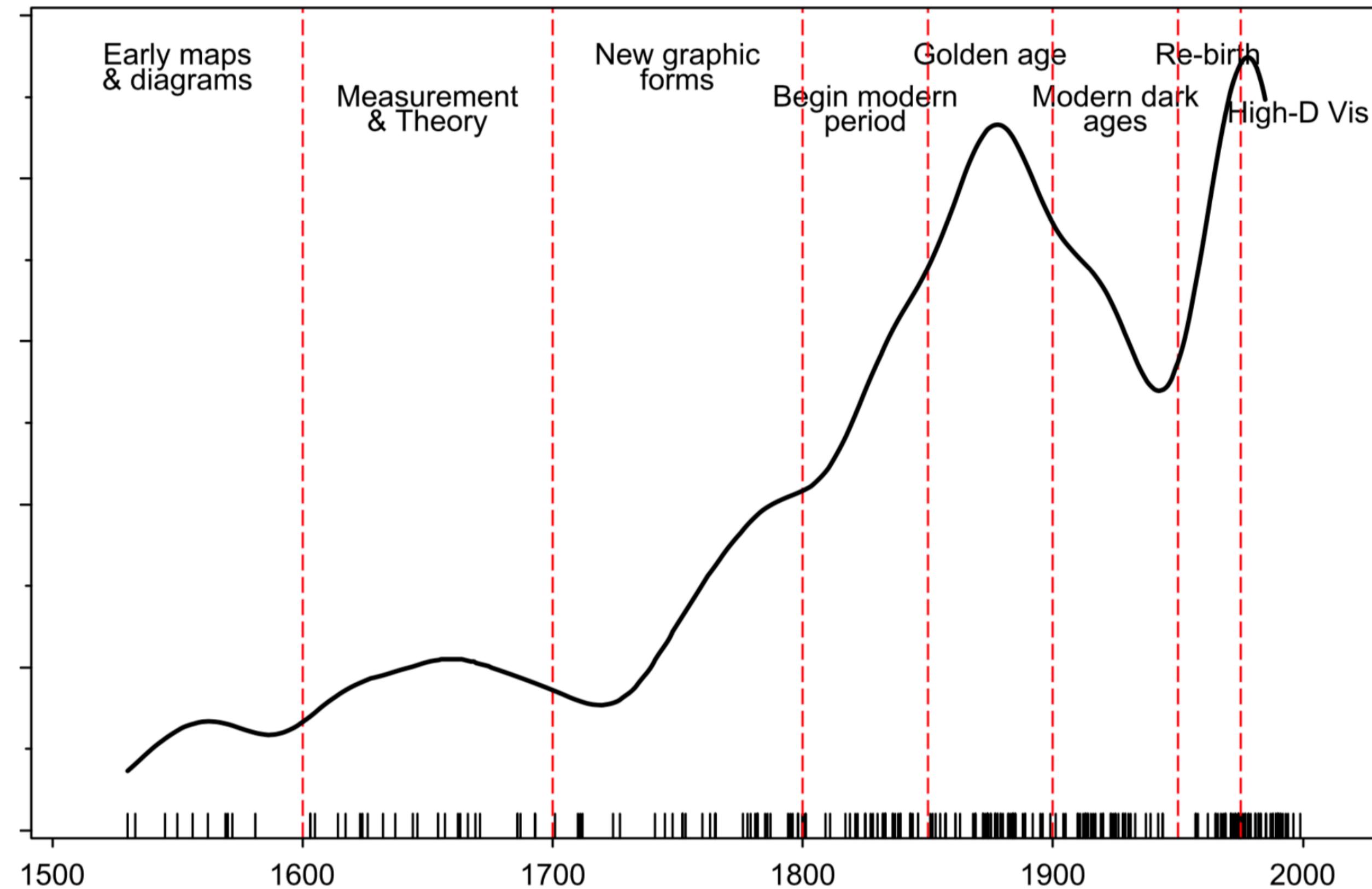
“The use of computer-supported, interactive visual representations of data to amplify cognition”¹

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”²

1 Card, S. K., Mackinlay, J. D., and Shneiderman, B. (1999). Using Vision to Think, chapter 1: Information Visualization, pages 1–34. Morgan Kaufmann.

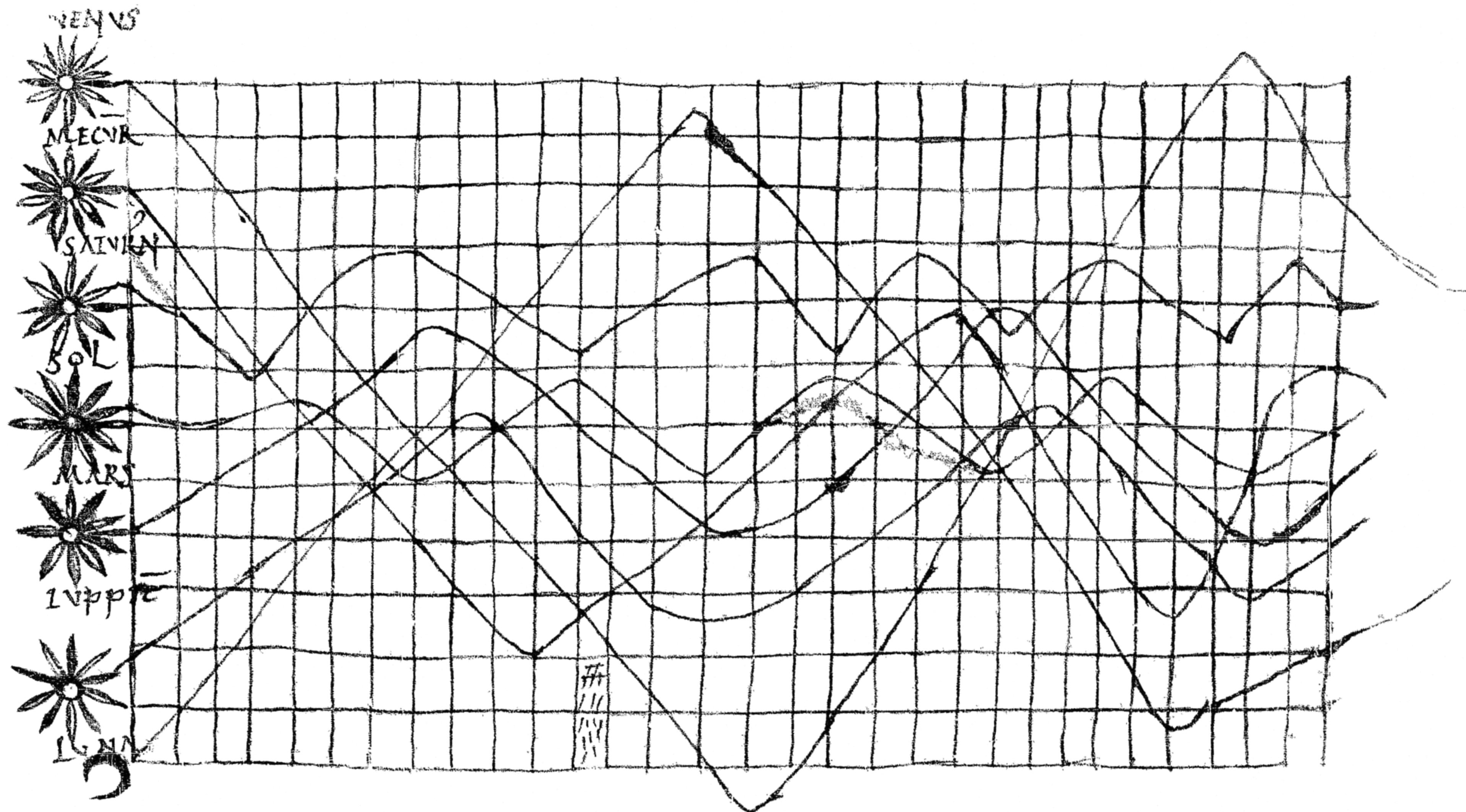
2 Munzner, T., (2014). Visualization Analysis and Design, A K Peters/CRC Press.

A short history

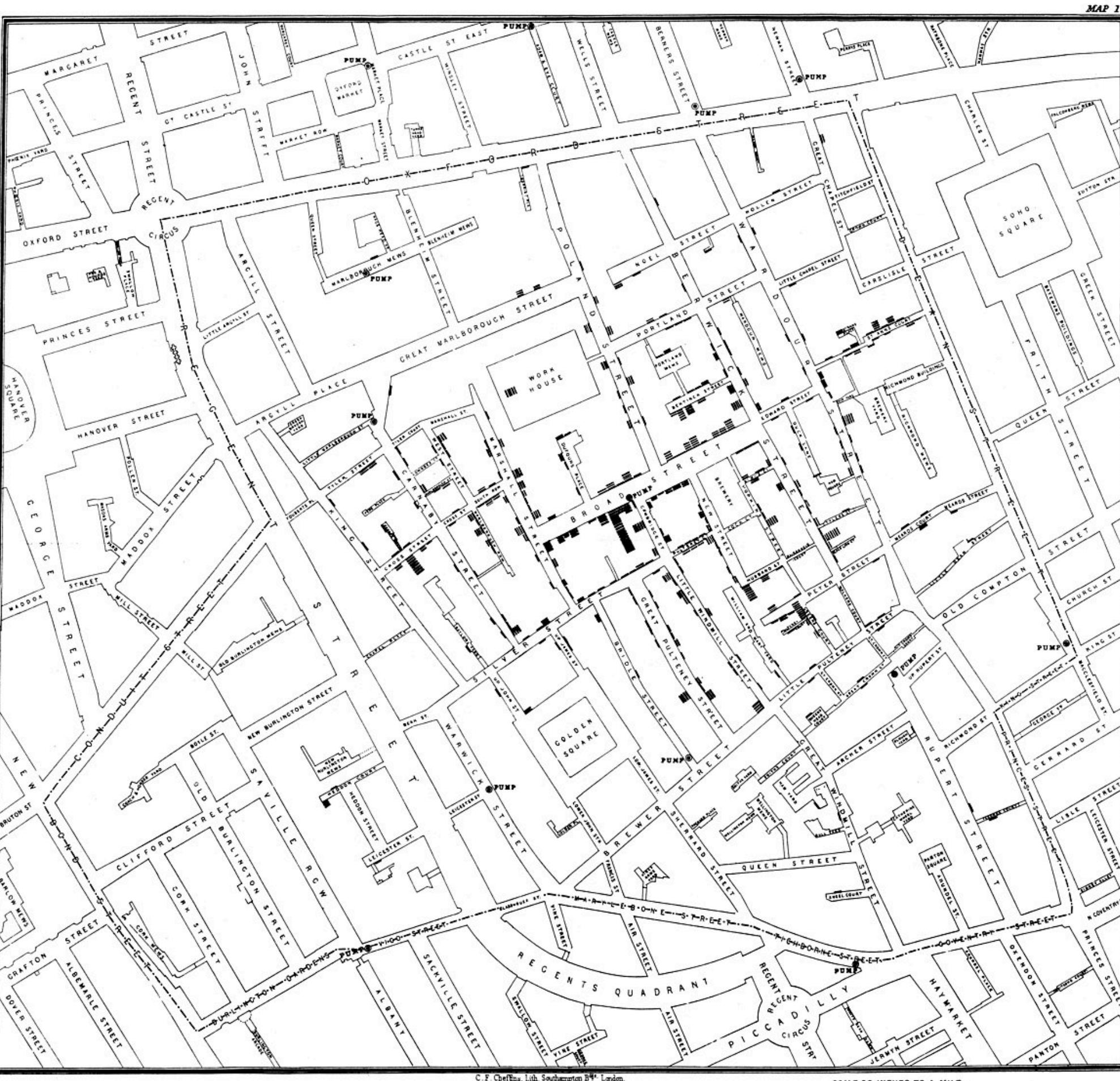


Friendly, M. (2008). Handbook of data visualization, chapter A brief history of data visualization, pages 15–56. Handbooks of Computational Statistics. Springer

A short history

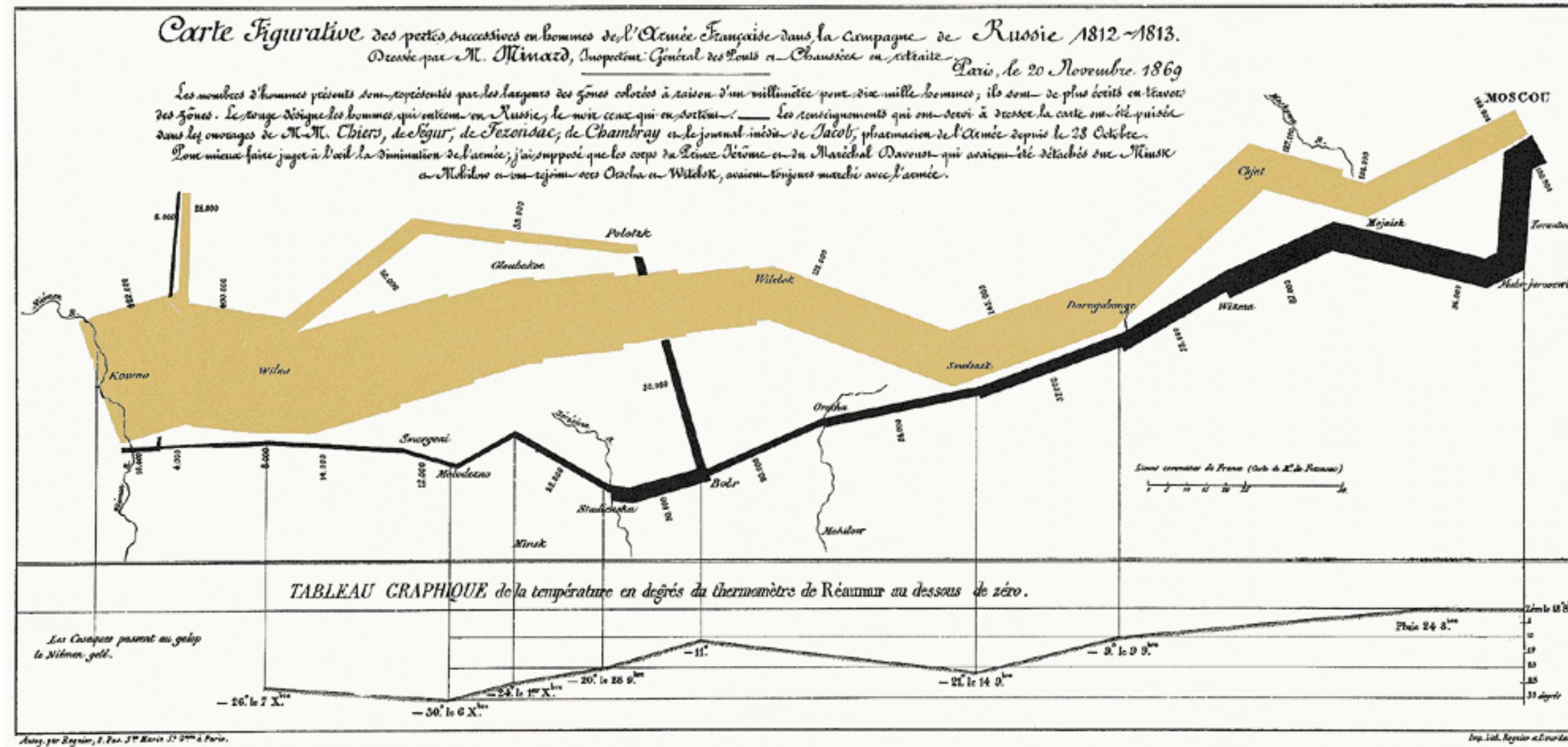


A short history



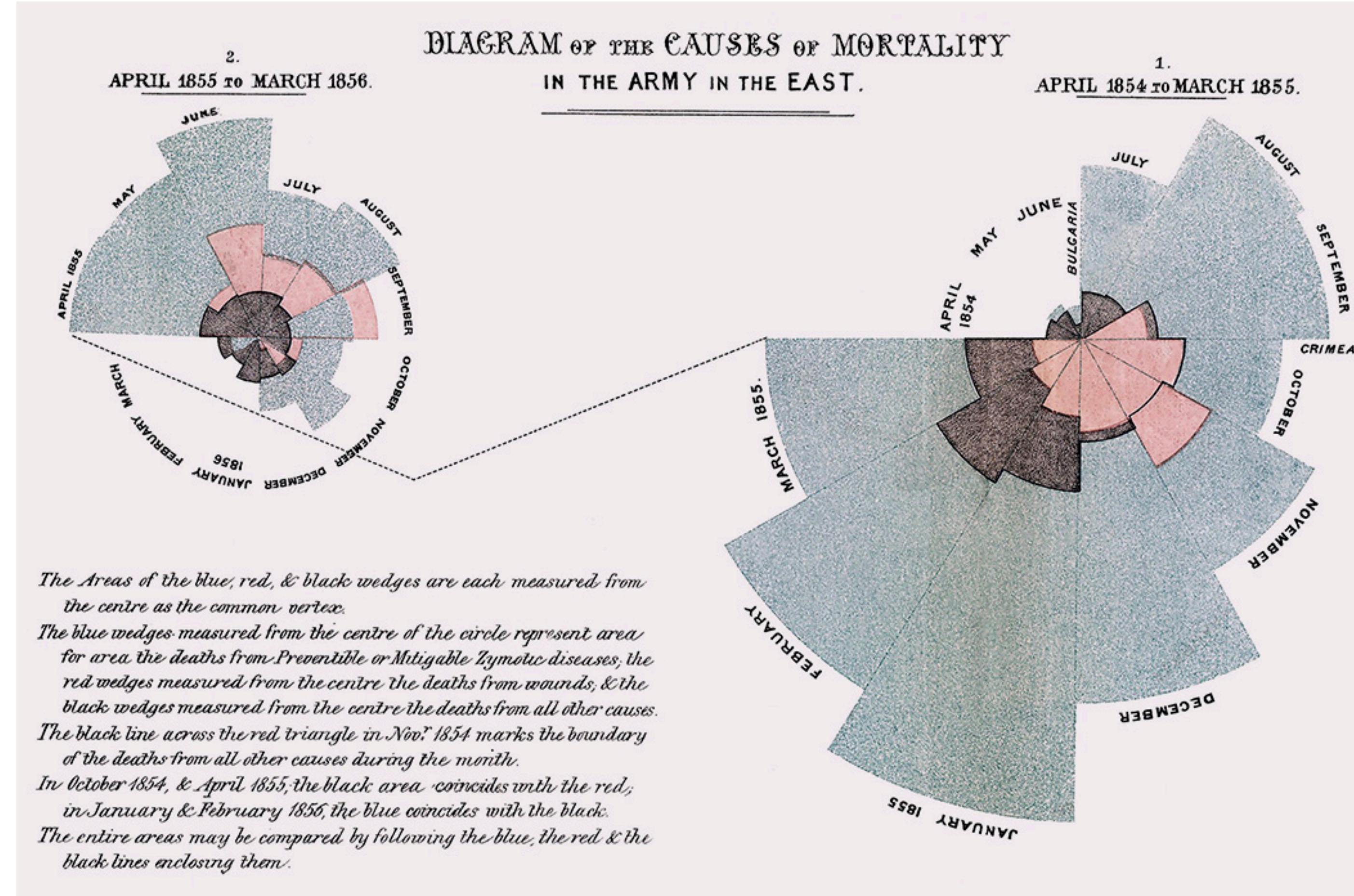
Cholera map of central London by John Snow, 1854

A short history



Napoleon's March by Minard, 1869

A short history



Florence Nightingale's Diagram of the causes of mortality in the army in the East, 1858



A short history



Nightingale

TOPICS IN DATA VIZ

HOW TO

CAREER ADVICE

HISTORIC DATA VIZ

OUTLIER

MORE INFO

ABOUT

Historic Data Visualization

Discussion of data visualization and information design crafted before 1990

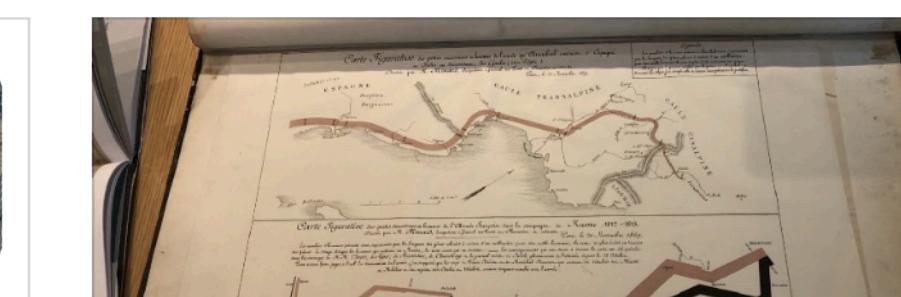
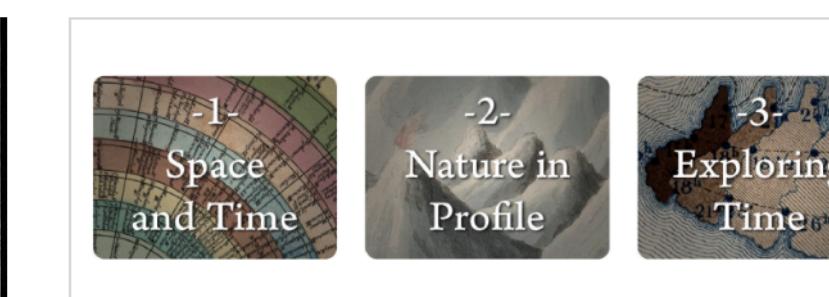
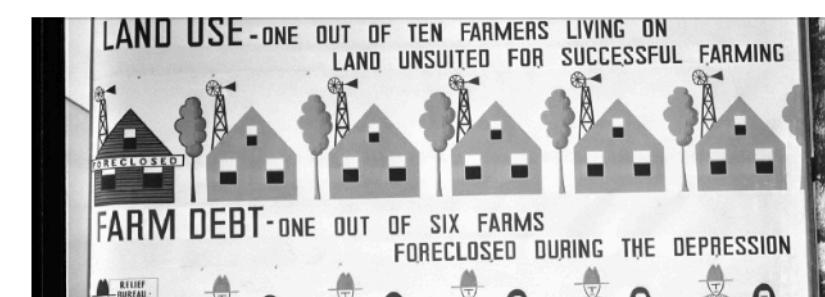


War and Human Nature

Revisiting a pamphlet from 1946 that helps explain what is happening in America right now



Jason Forrest
Jan 19 · 9 min read ★



<https://medium.com/nightingale/historicdv/home>

What is it for?

Communication.

Insights.

What else?

Entertainment.

Supporting cognition

Increases resources: working memory, parallel processing

Reduce search: high data density, including “overview”

Pattern recognition: abstraction and aggregation

Perceptual inferences: making possible relationships obvious

Perceptual monitoring: revealing trends and changes

Interactive manipulation: adjusting parameters, filters and views

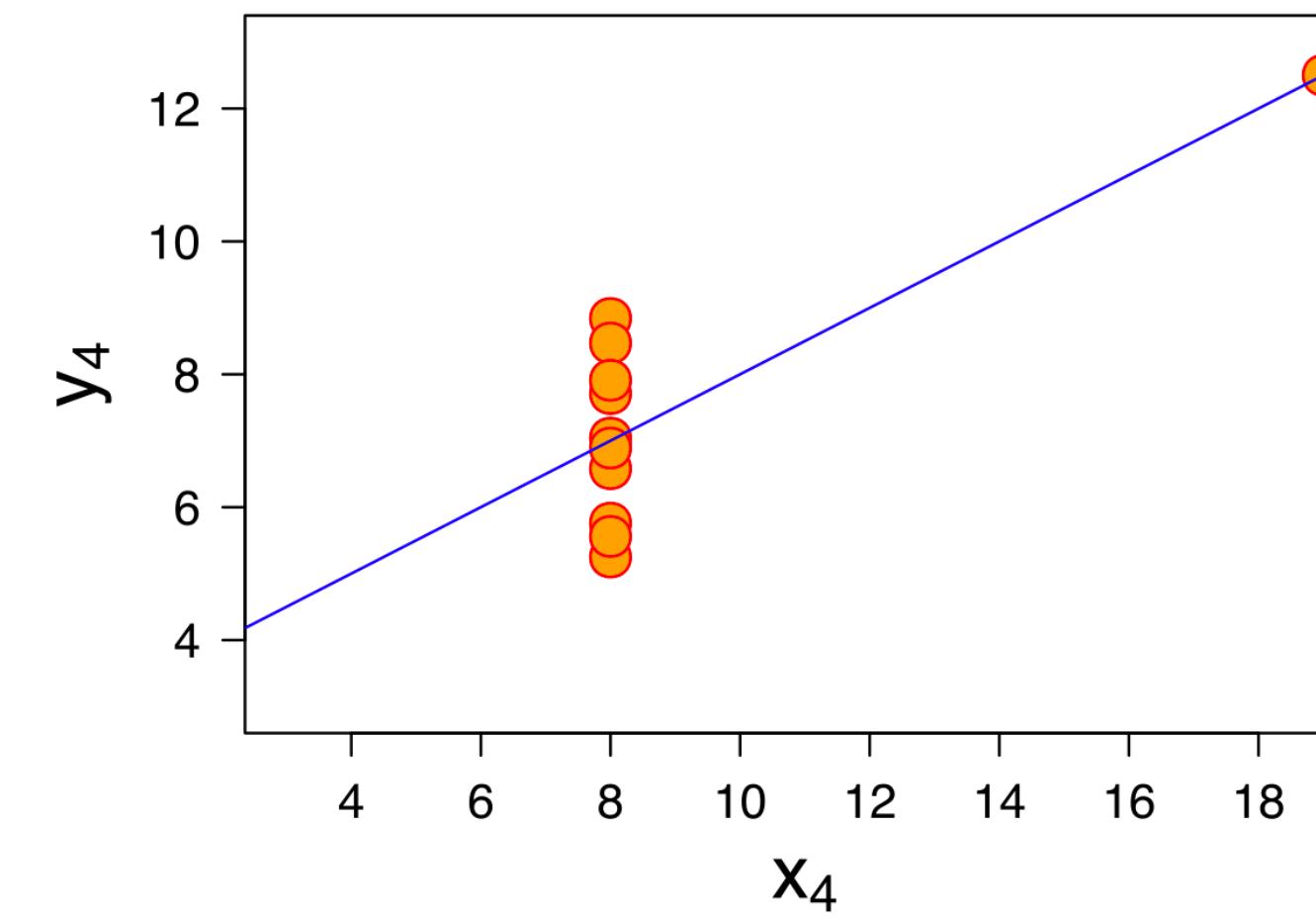
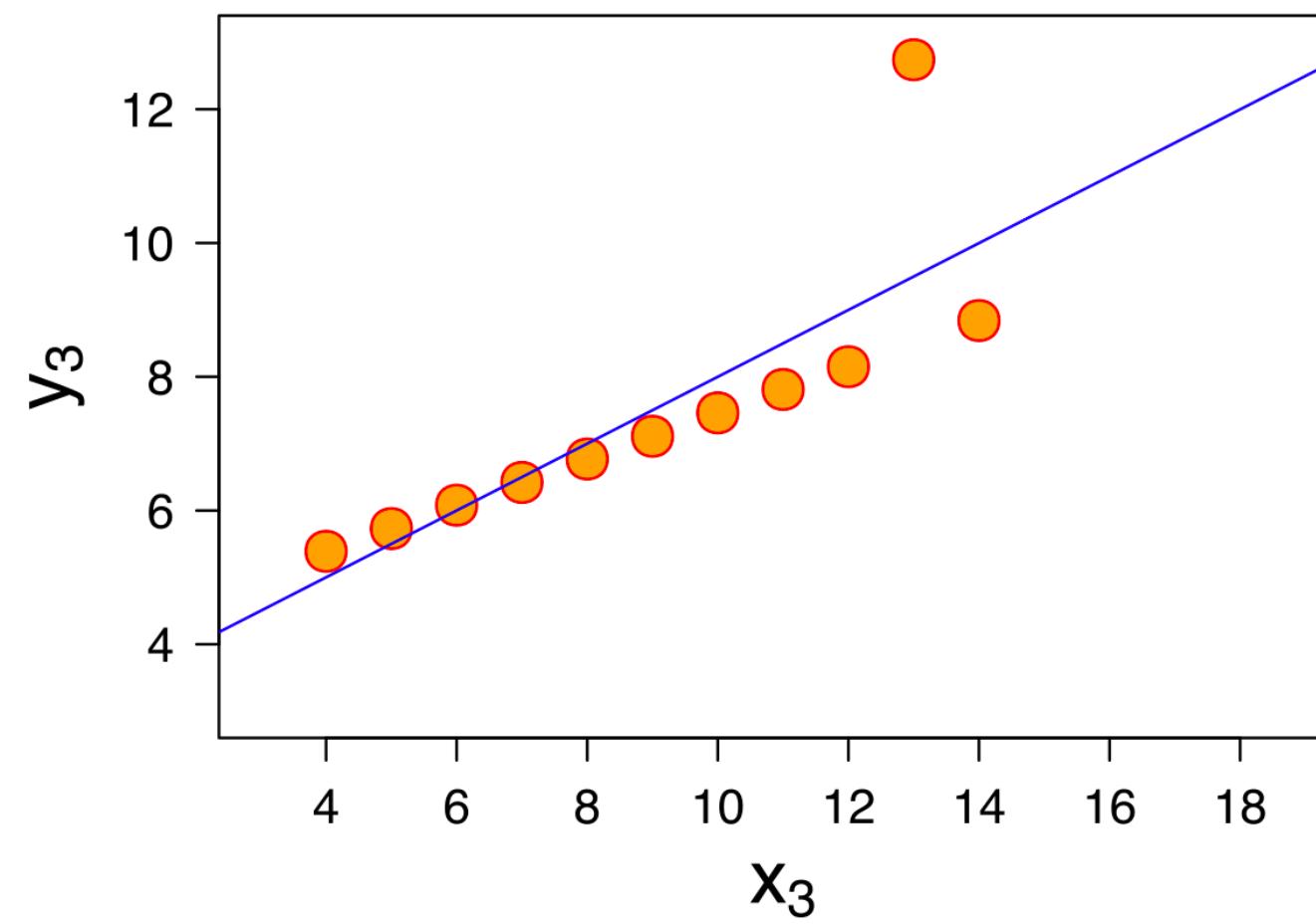
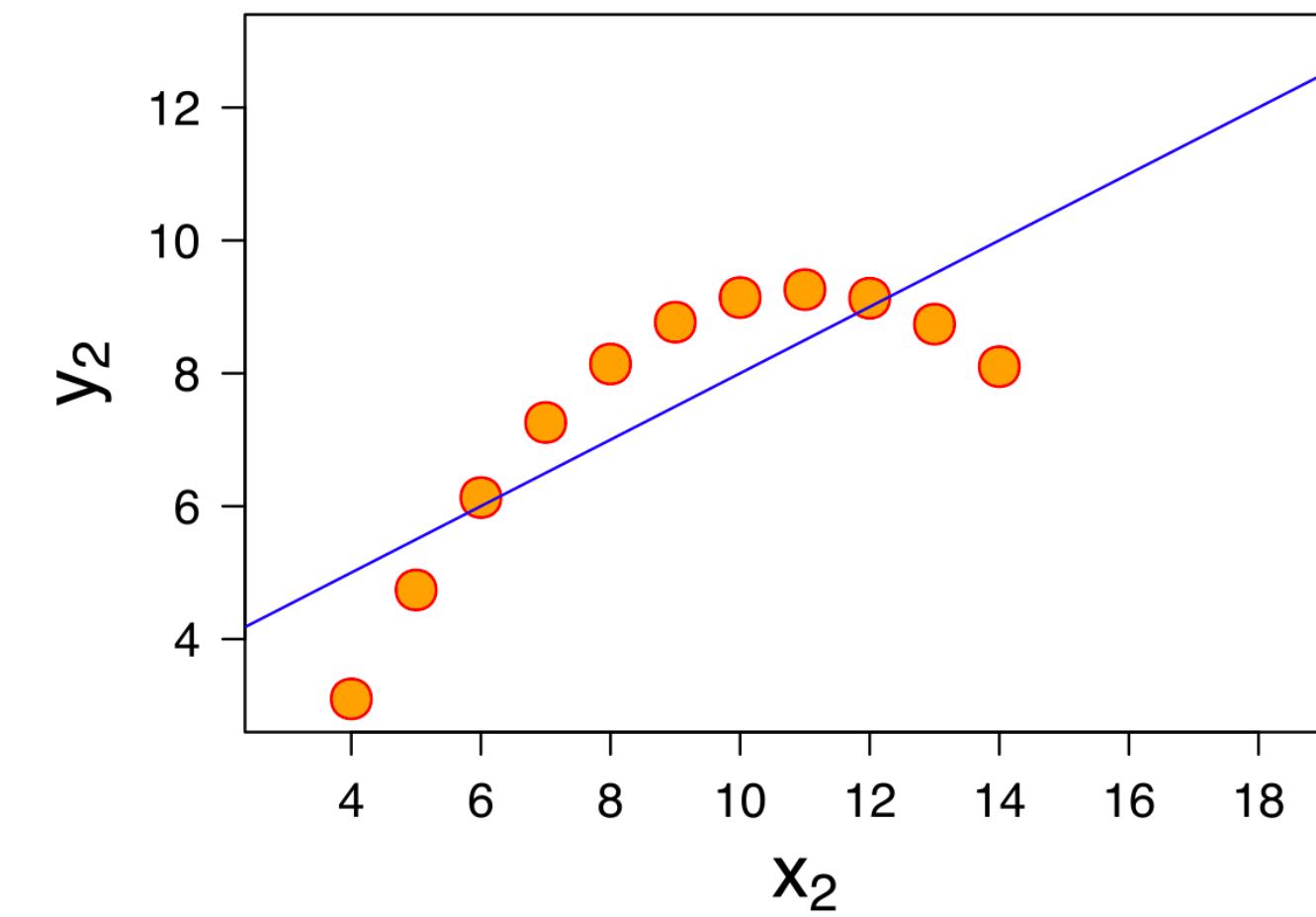
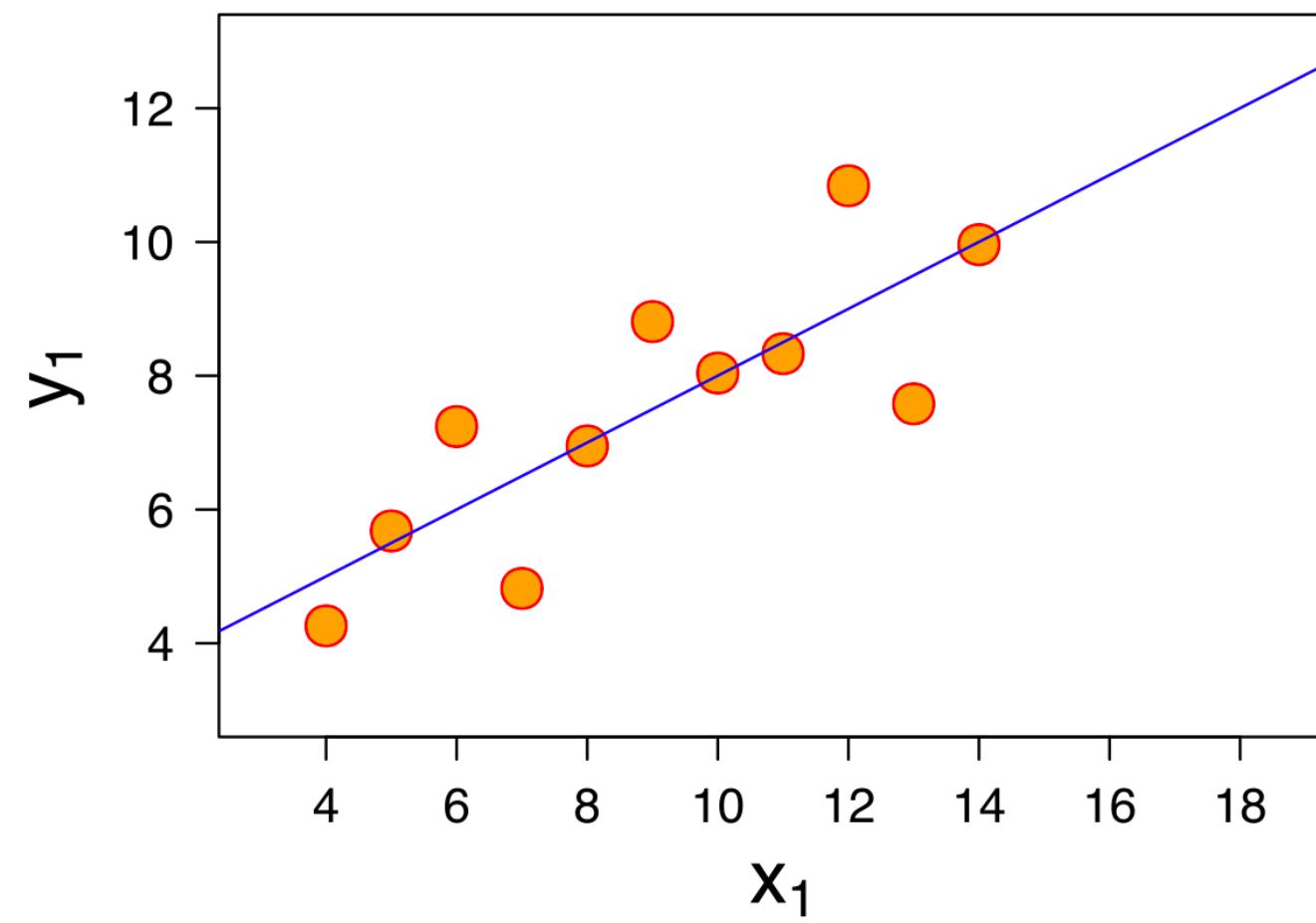
Anscombe's Quartet

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

Statistical description

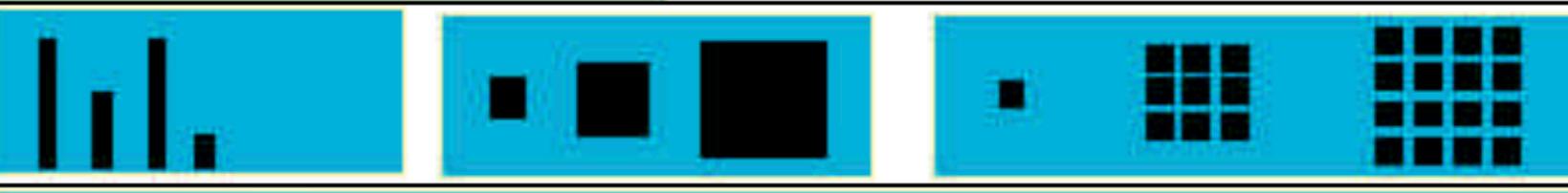
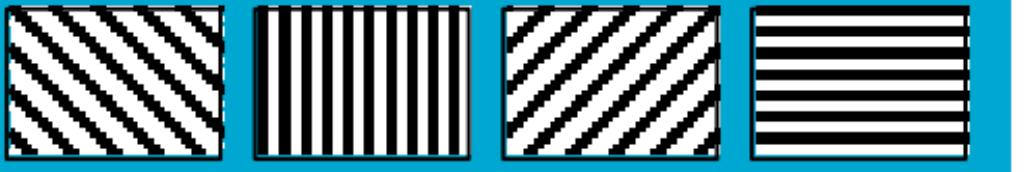
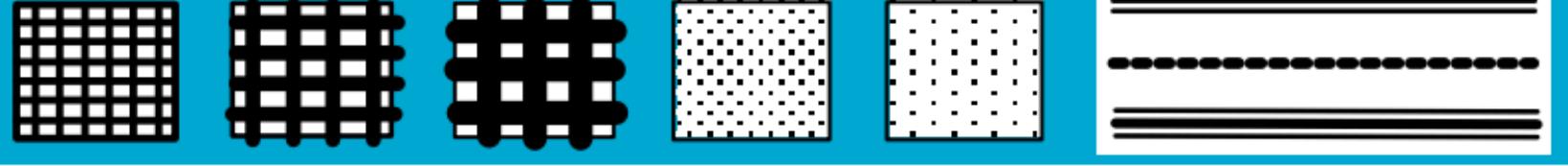
Mean of x: 9
Variance of x: 11
Mean of y: 7.5
Variance of y: 4.12
Correlation: 0.816
Regression line y=.5x+3

Anscombe's Quartet



Visual Variables

Position	X-Y location
Size	Location
Value	Length or area brightness
Texture	Visual pattern
Colour	Hue & Saturation
Orientation	Alignment
Shape	Visual form
Animation	Change visual vars

Bertin's Original Visual Variables	
Position changes in the x, y location	
Size change in length, area or repetition	
Shape infinite number of shapes	
Value changes from light to dark	
Colour changes in hue at a given value	
Orientation changes in alignment	
Texture variation in 'grain'	

Bertin, J. (1983). Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press.

Carpendale, S. (2003). Considering visual variables as a basis for information visualisation.
Technical Report 2001-693-16, Department of Computer science, University of Calgary.

Visualization types





10 minutes break :)

Hands on tutorials.

<https://github.com/uclab-potsdam/filmuni-workshop/>

15 minutes break :)

Next week