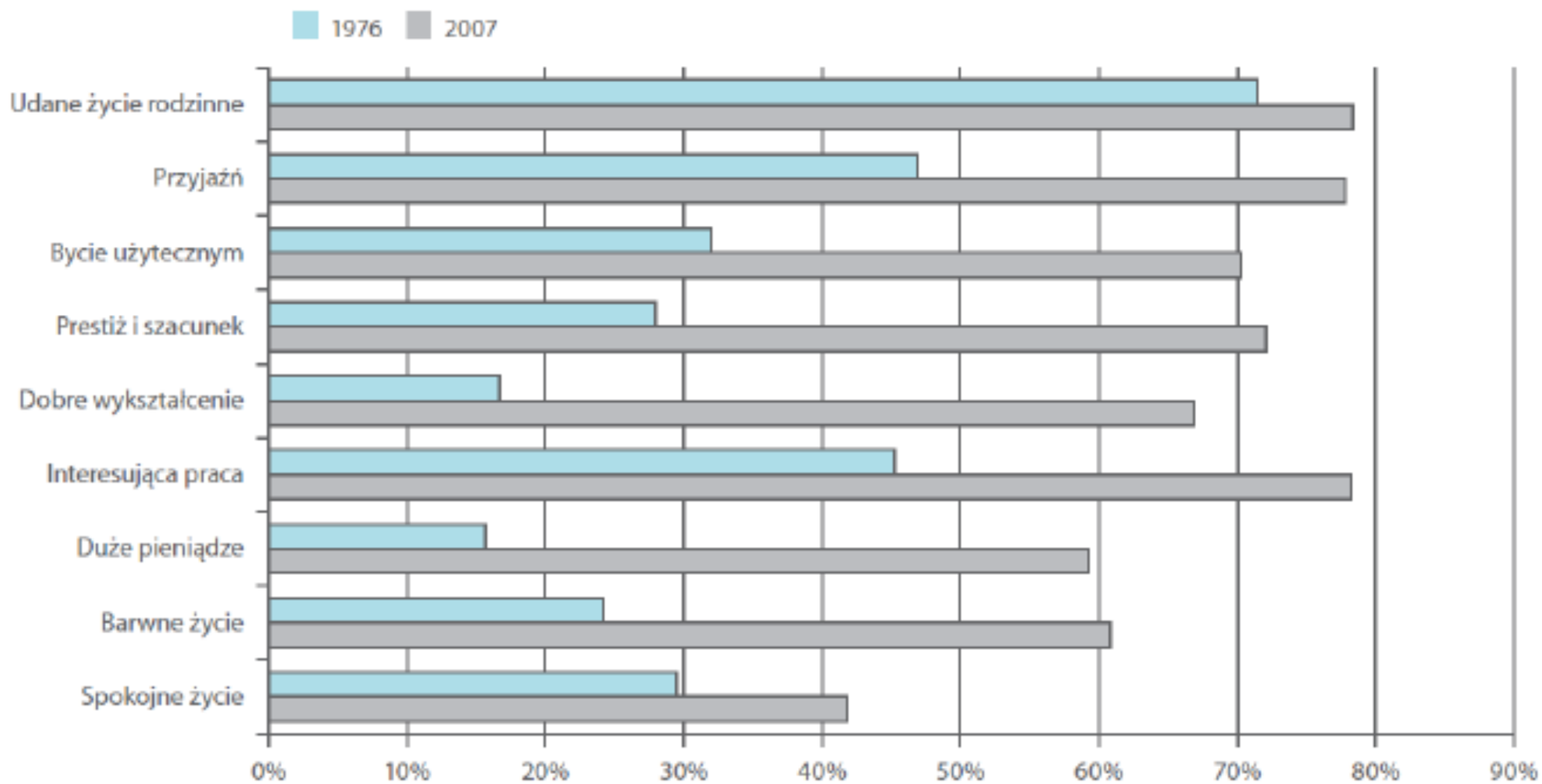
Grammar of Graphics

Statistics II
Part 1: Exploration

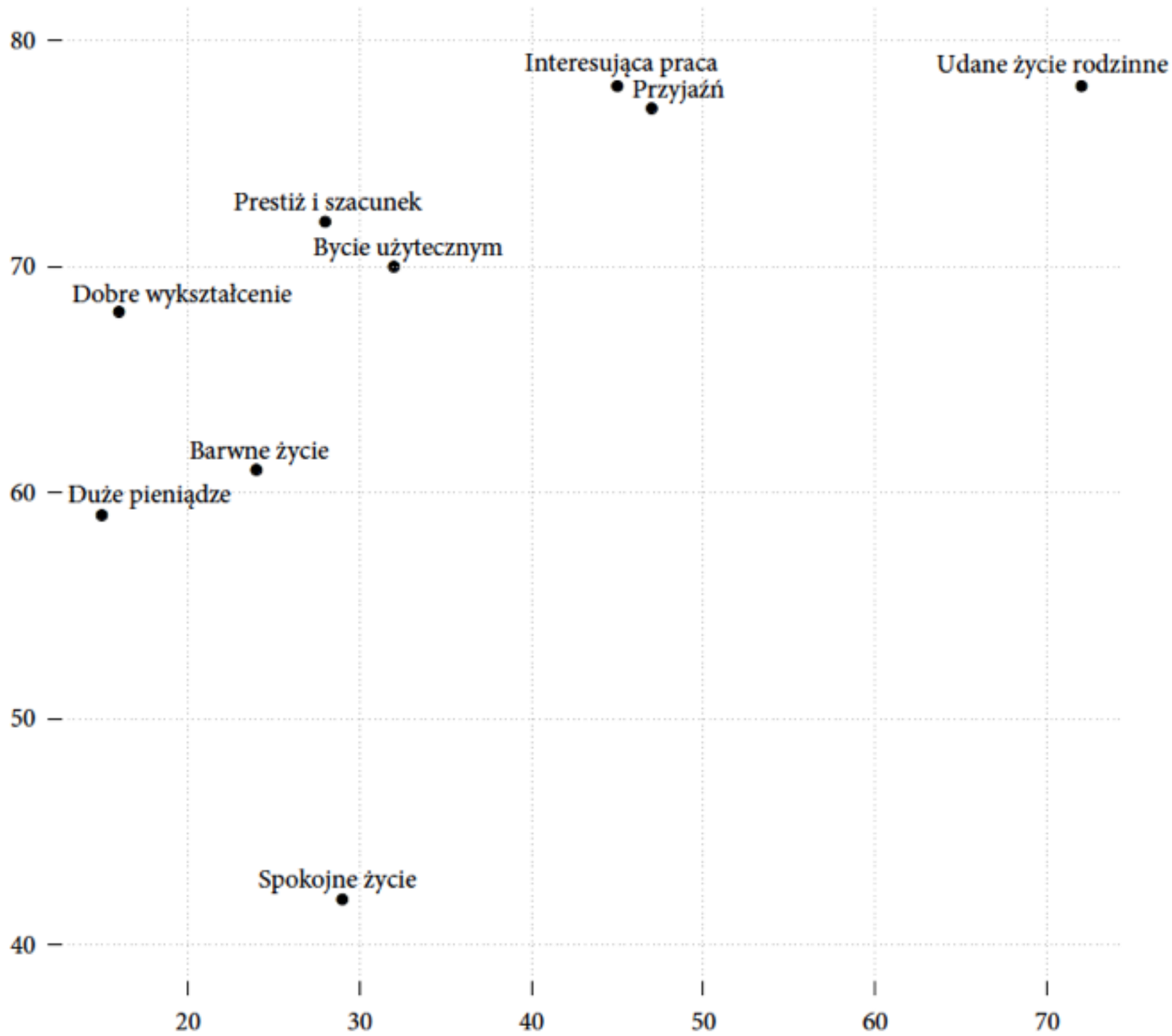
Myth of a single story

Two generations - one question:
what is important in life?



Źródło: Badania warszawsko-kieleckie S. Nowaka (lata 70. XX w.), badania własne: *„Porzucona generacja" – ścieżki edukacyjne i wchodzenie w dorosłość* (N = 1096).

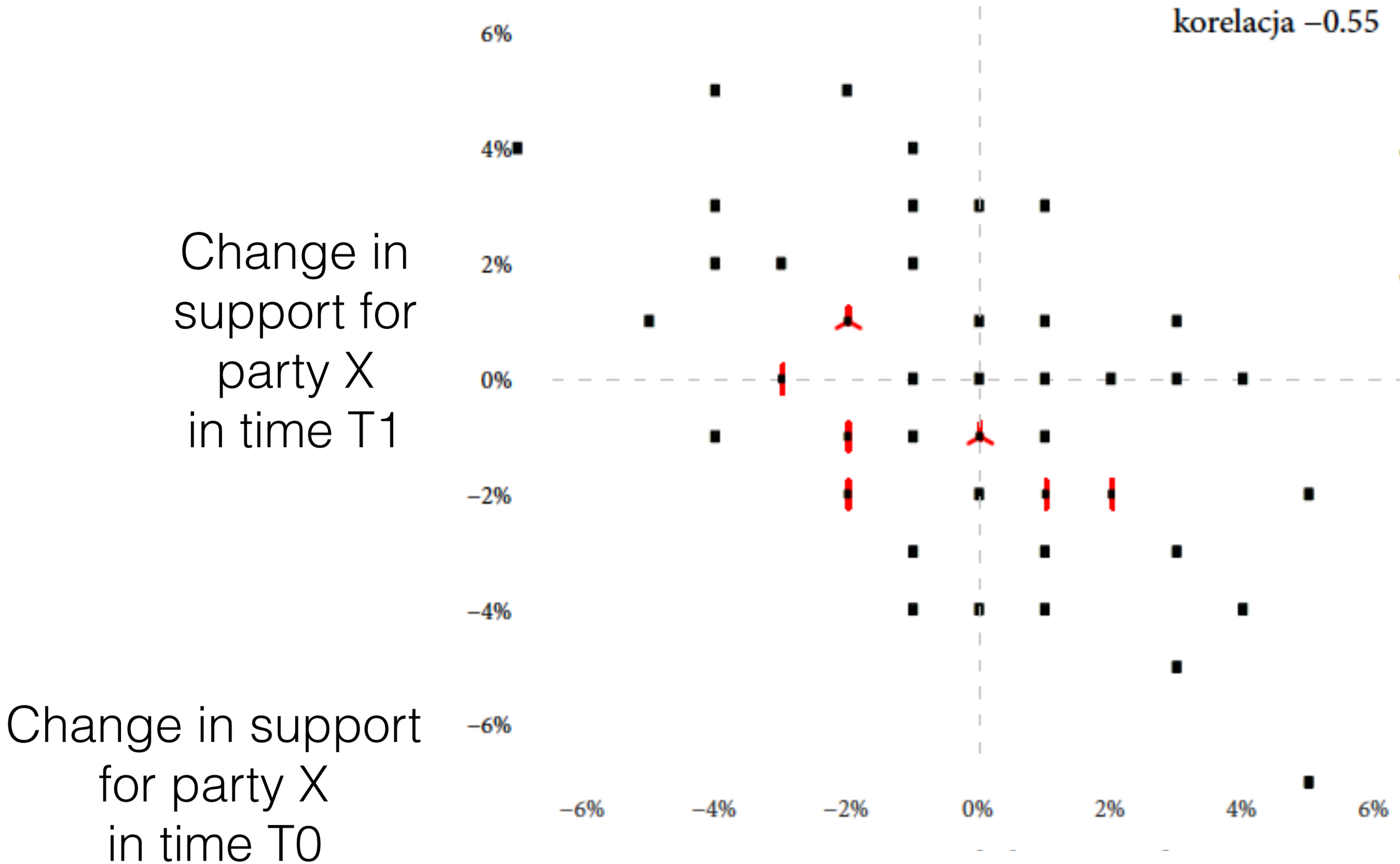
Same data - different perspectives



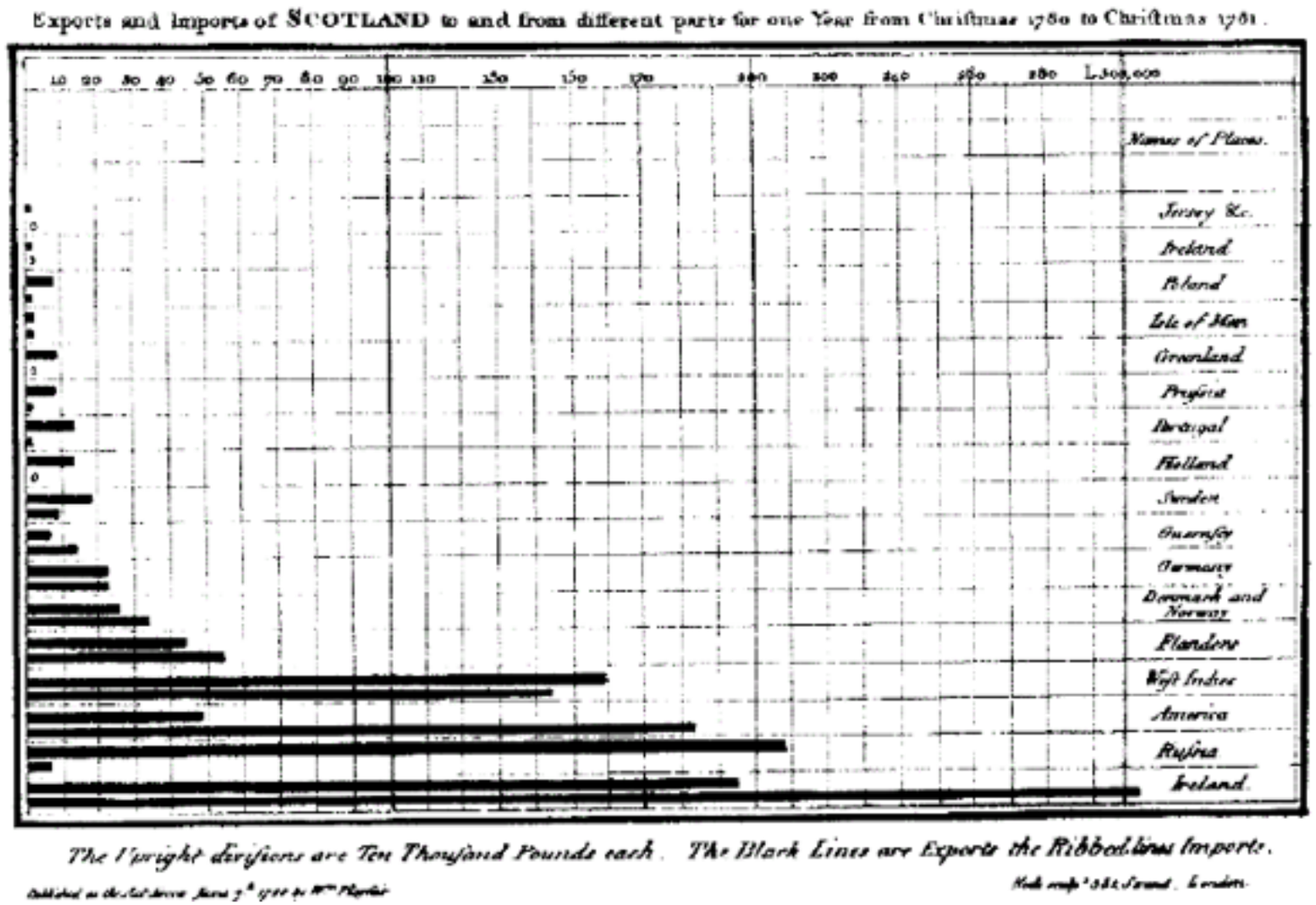
Answers in 2007

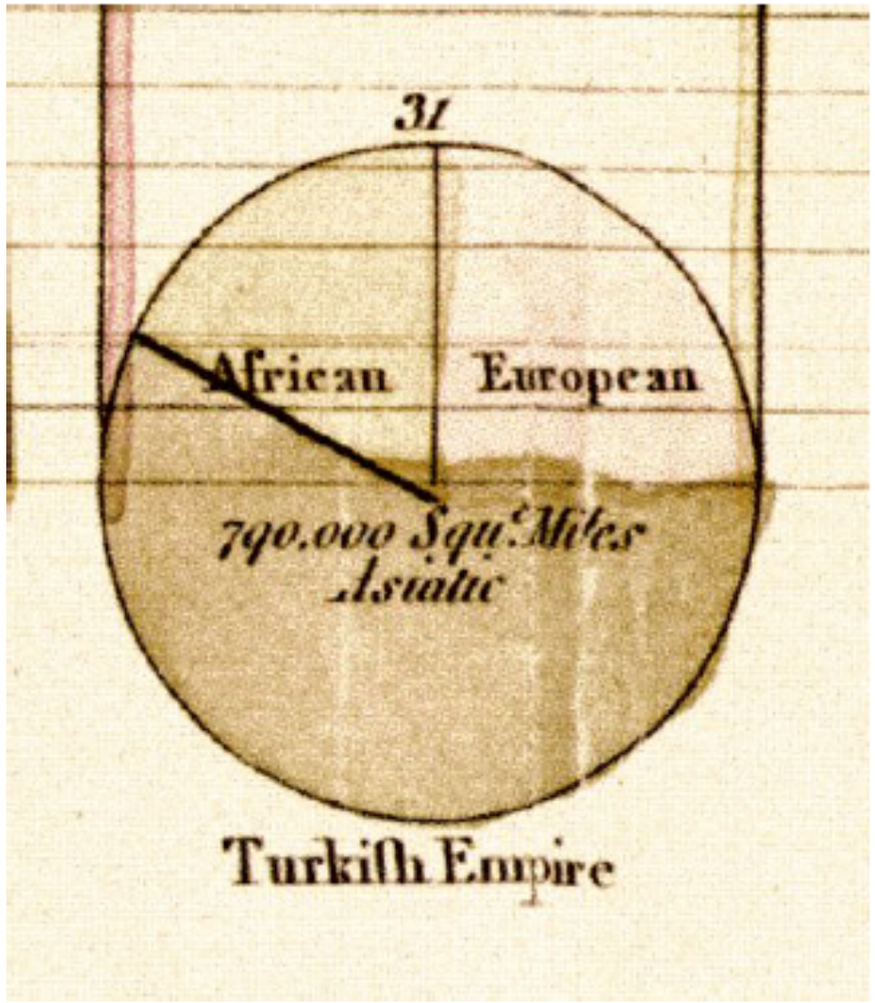
Answers in 1976

Is there any story?



History of statistical graphics

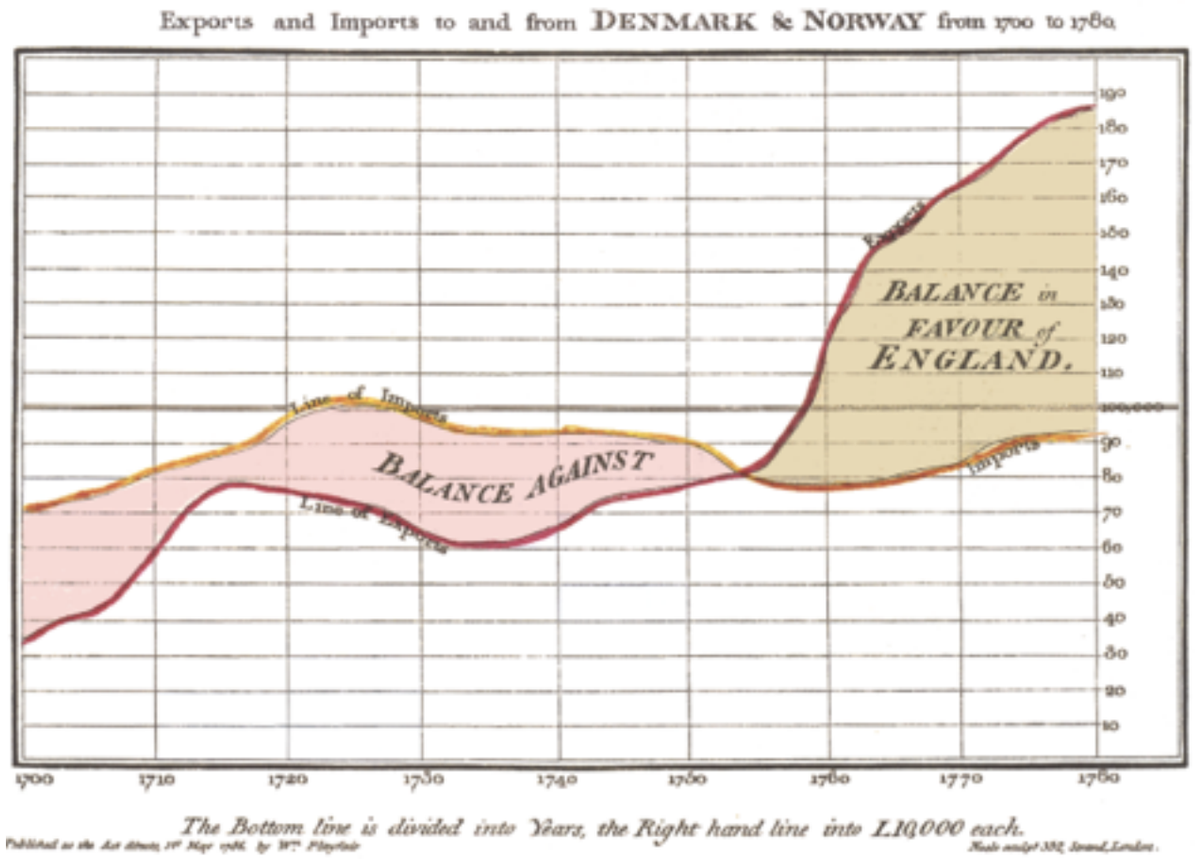




Oldest modern statistical graphics

William Playfair (1786)

The Commercial and Political Atlas



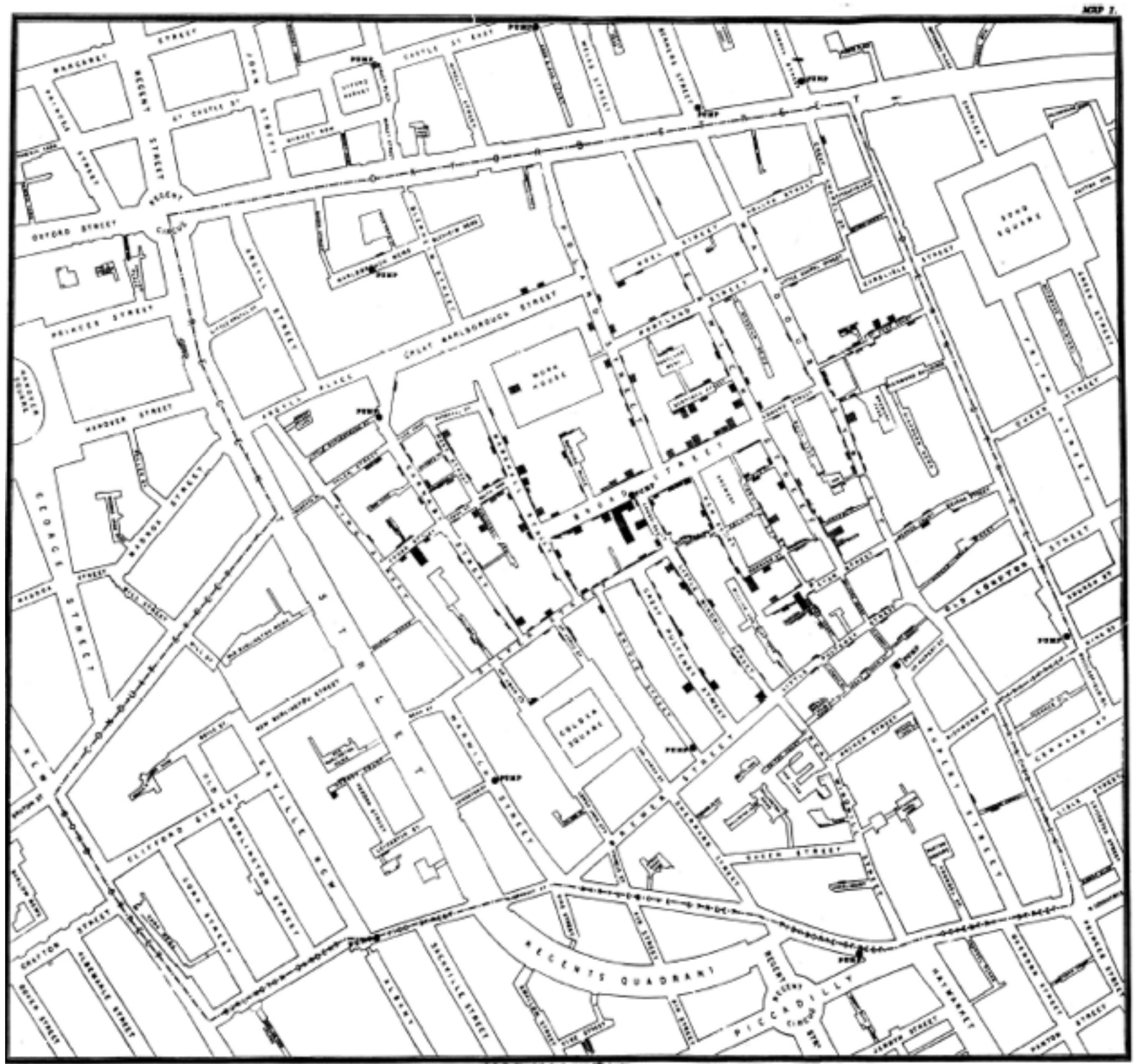
Most known graphics

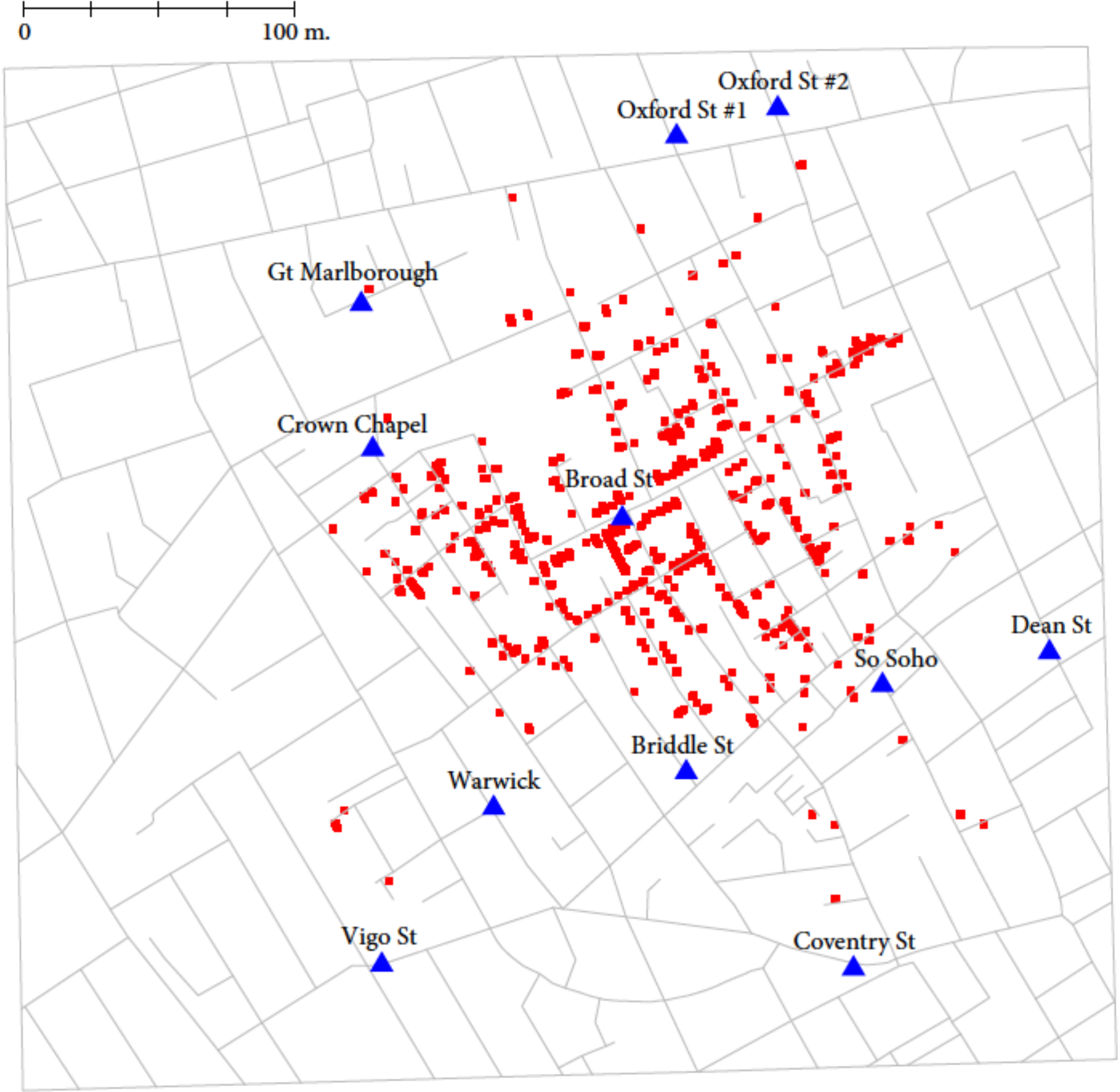
John Snow

(1854)

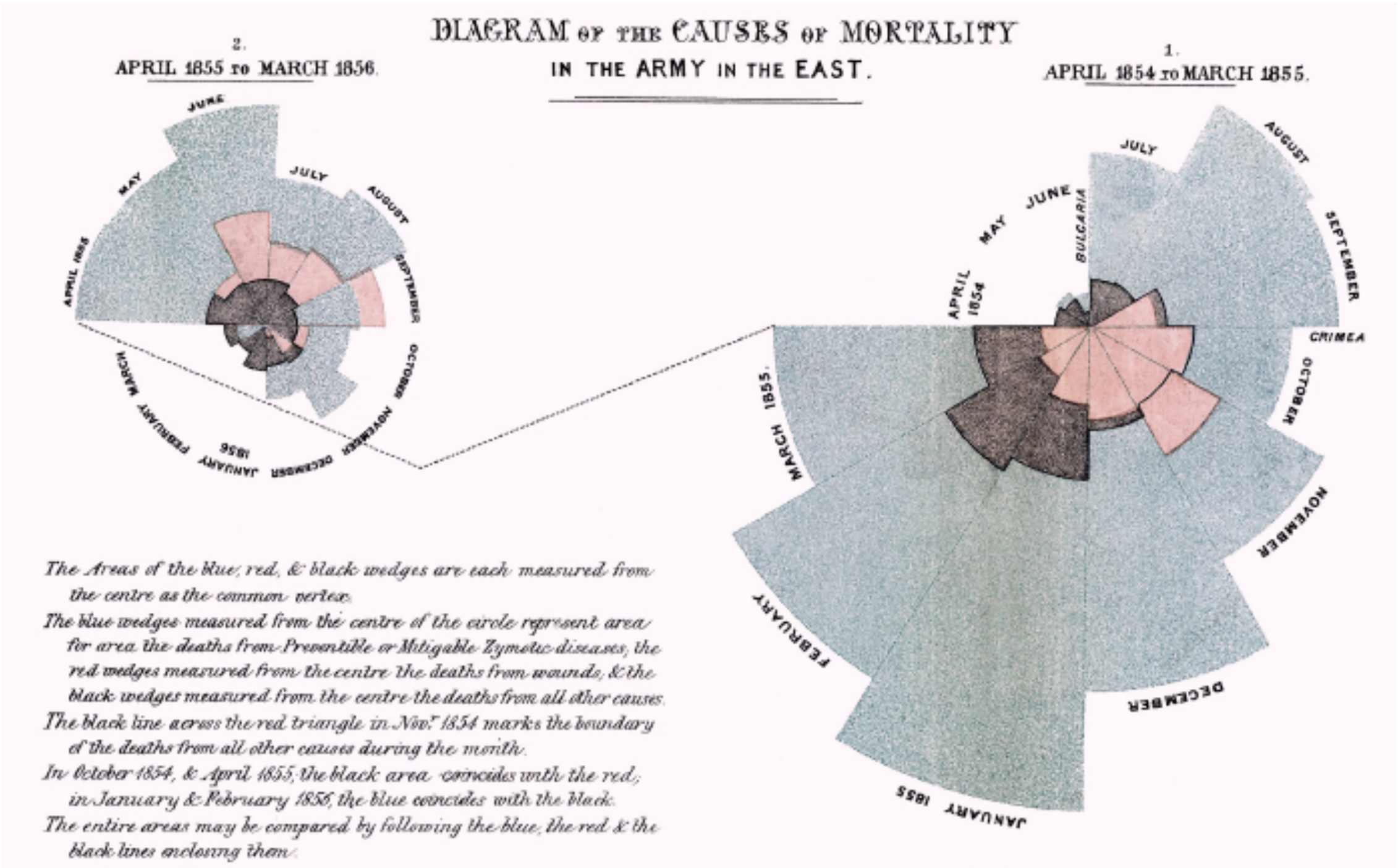
cholera cases in the London epidemic

drawn by Charles Cheffins

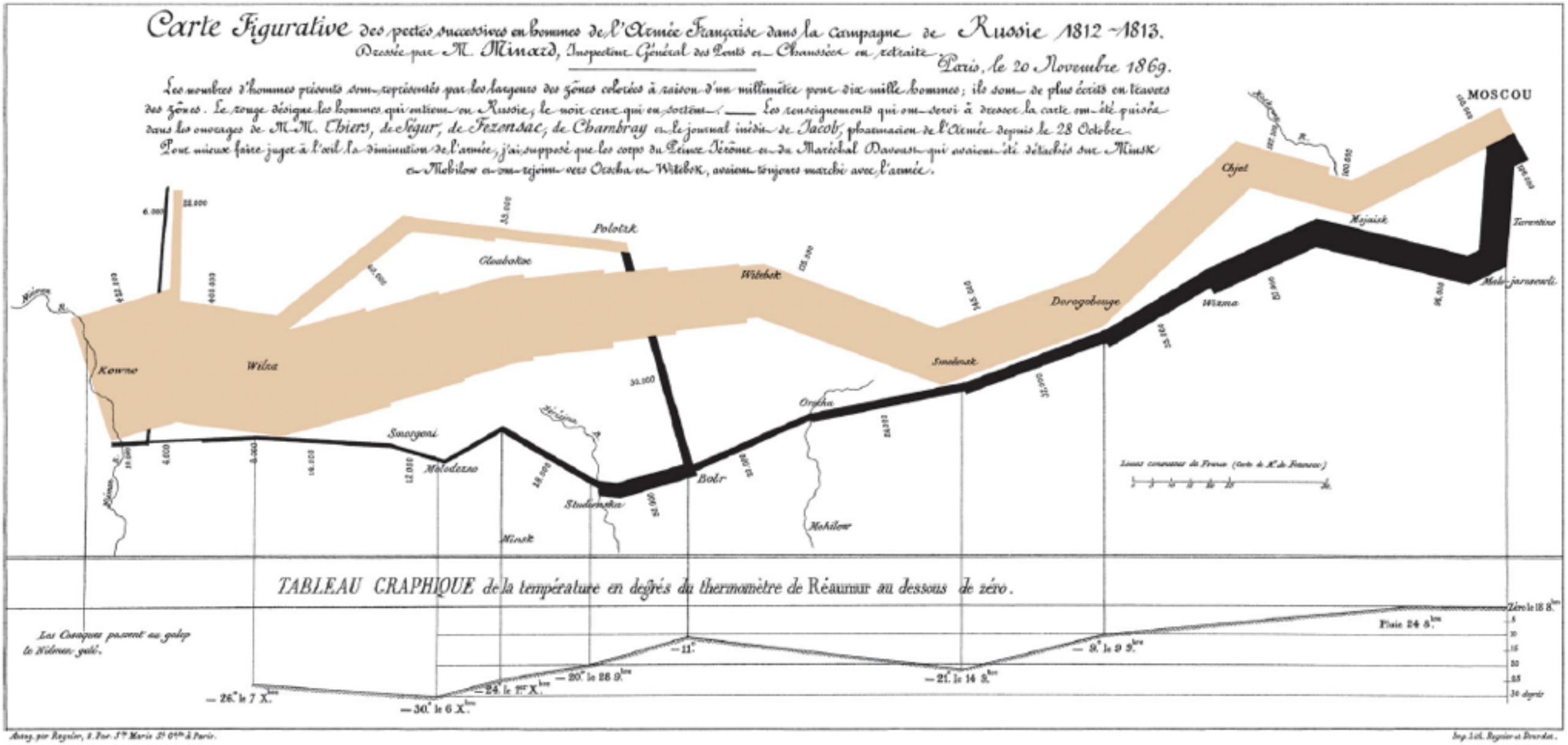



0
100 m.
Oxford St #2
Oxford St #1
Gt Marlborough
Crown Chapel
Broad St
Dean St
So Soho
Briddle St
Warwick
Vigo St
Coventry St

Florence Nightingale (1859) coxcombs

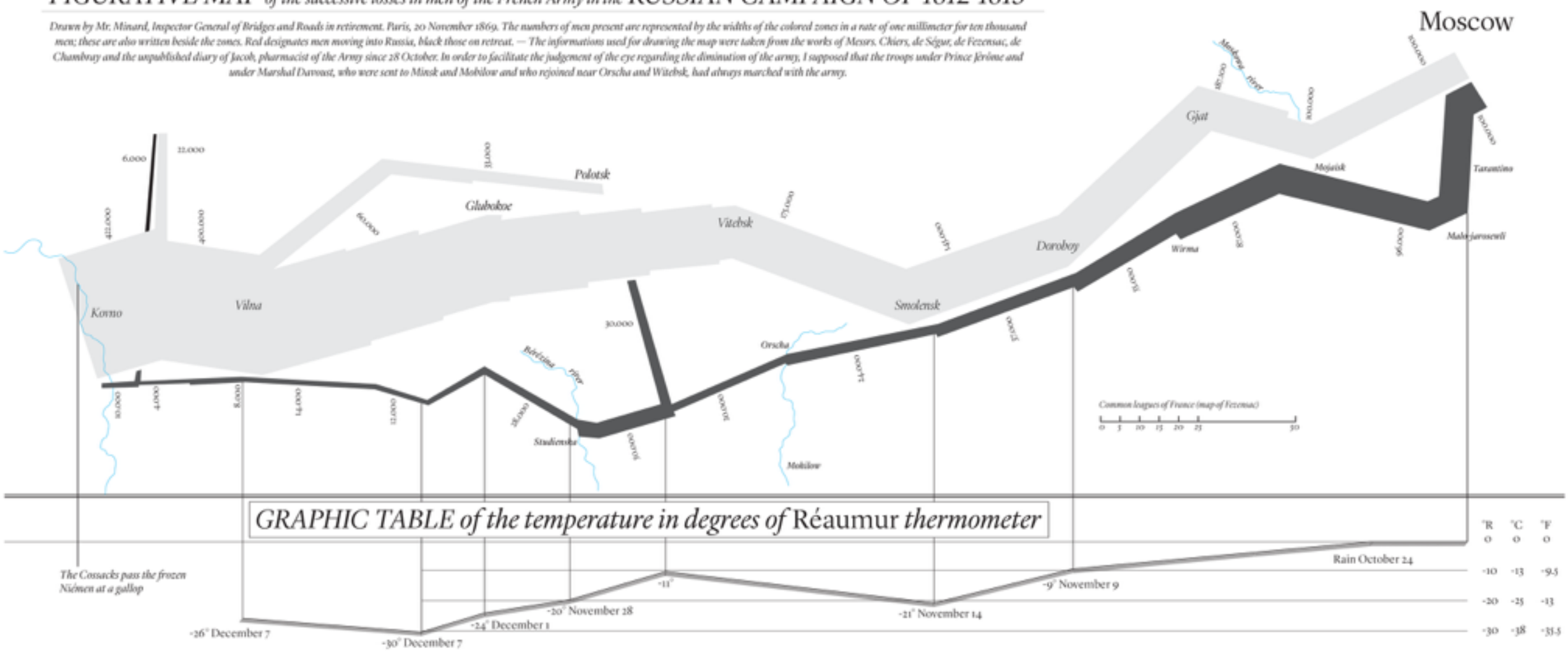


Charles Minard (1869)
French Army in the Russian campaign

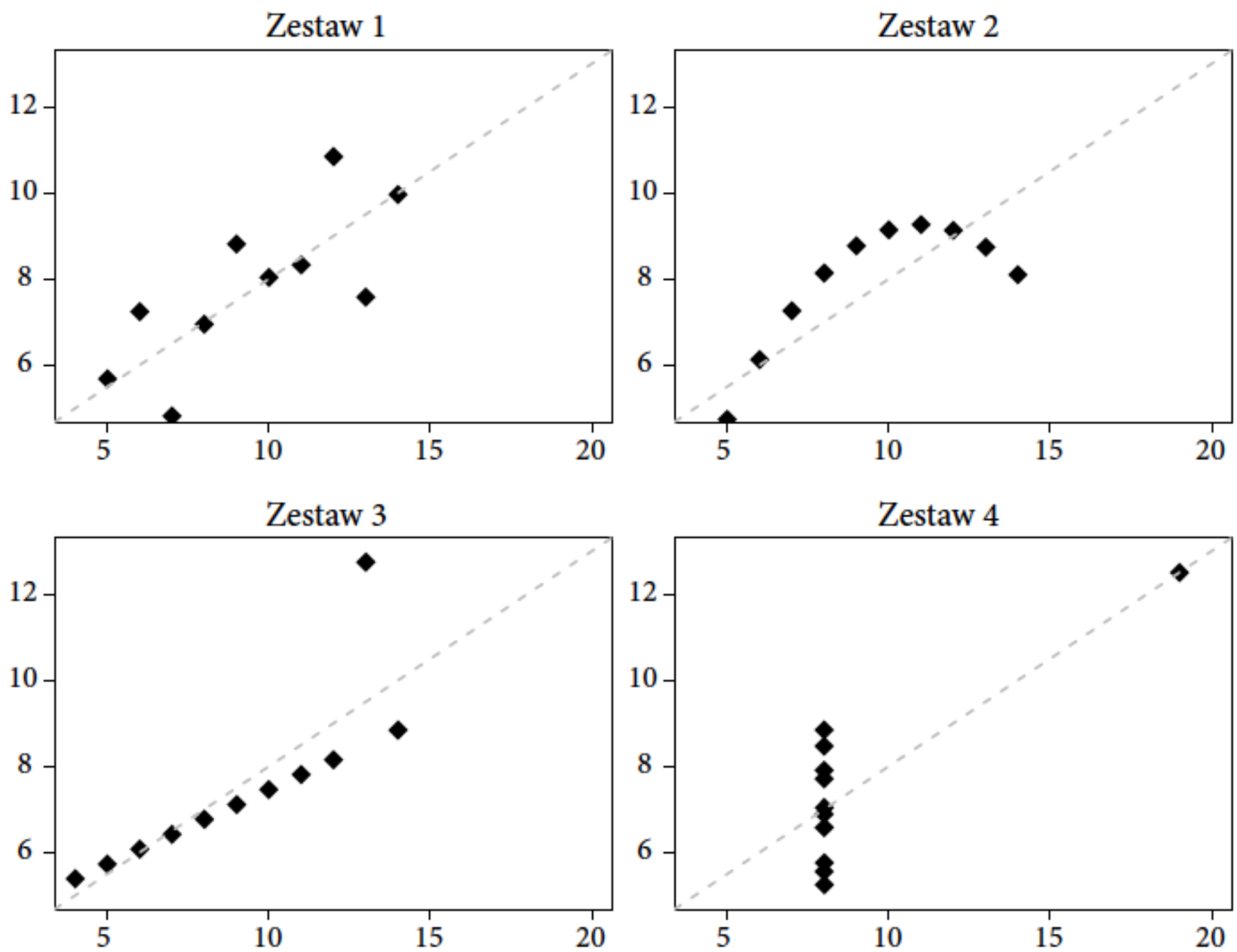


FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

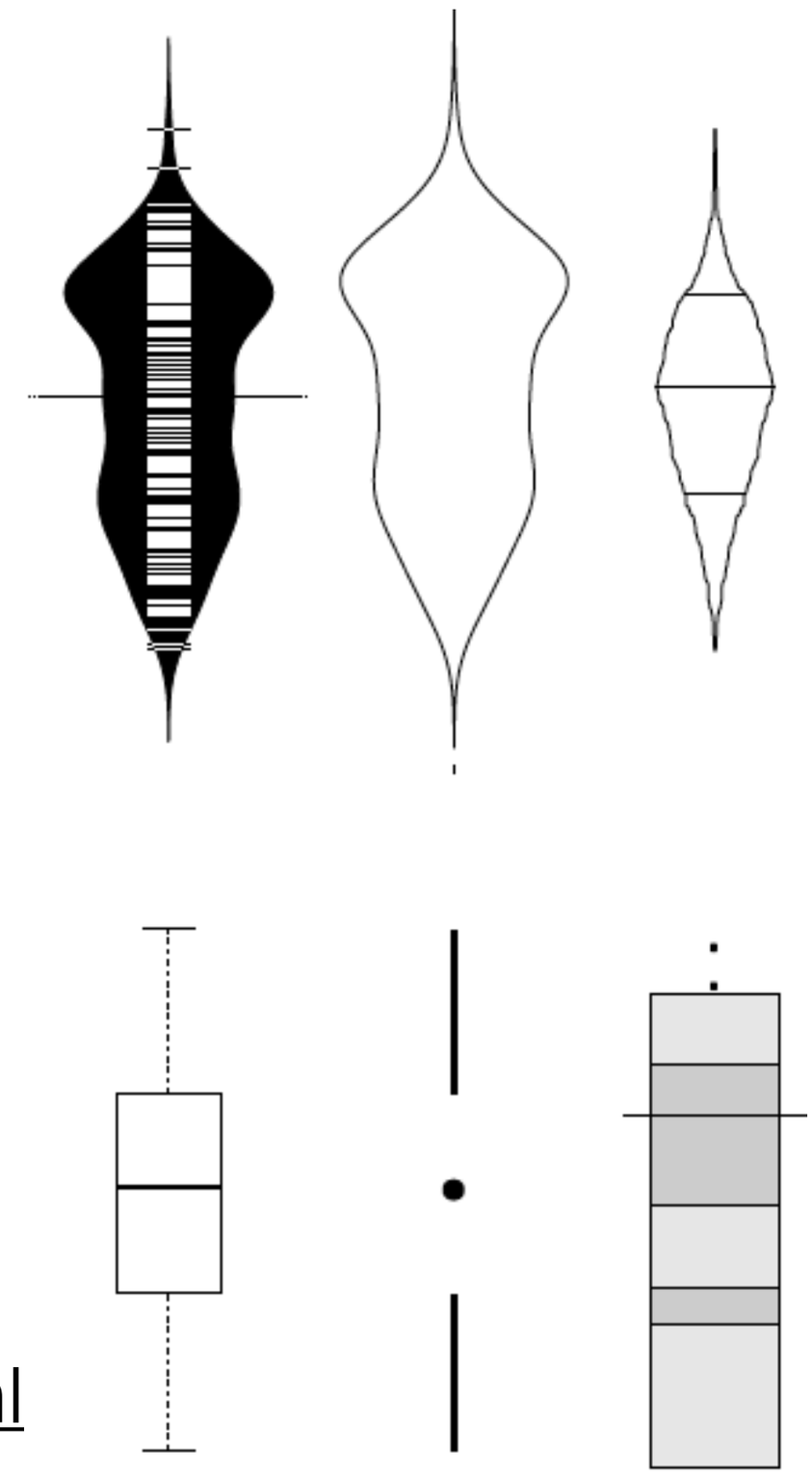
Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement. Paris, 20 November 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Meurs. Chiers, de Ségur, de Fezensac, de Chambray and the unpublished diary of Jacob, pharmacist of the Army since 28 October. In order to facilitate the judgement of the eye regarding the diminution of the army, I supposed that the troops under Prince Jérôme and under Marshal Davoust, who were sent to Minsk and Mobilow and who rejoined near Orscha and Witebsk, had always marched with the army.



Francis Anscombe (1973) anscombe's quartet



Tukey's boxplots and some variations



Invented by John Tukey (1977)

Exploratory Data Analysis

Addison-Wesley.

Hadley Wickham, Lisa Stryjewski.

40 years of boxplots.

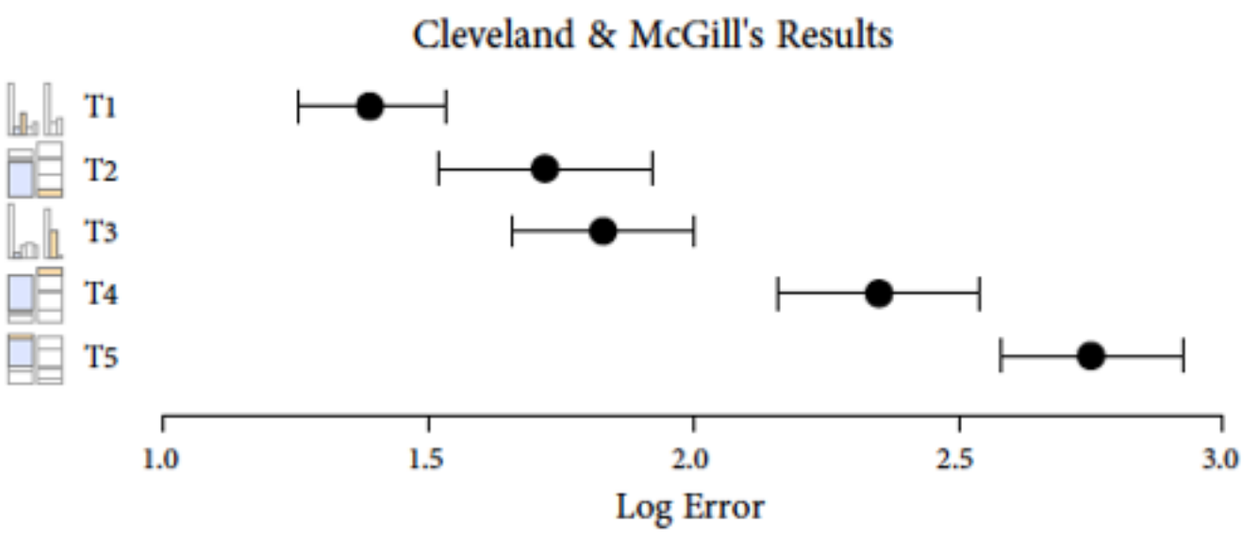
http://vita.had.co.nz/papers/boxplots.html

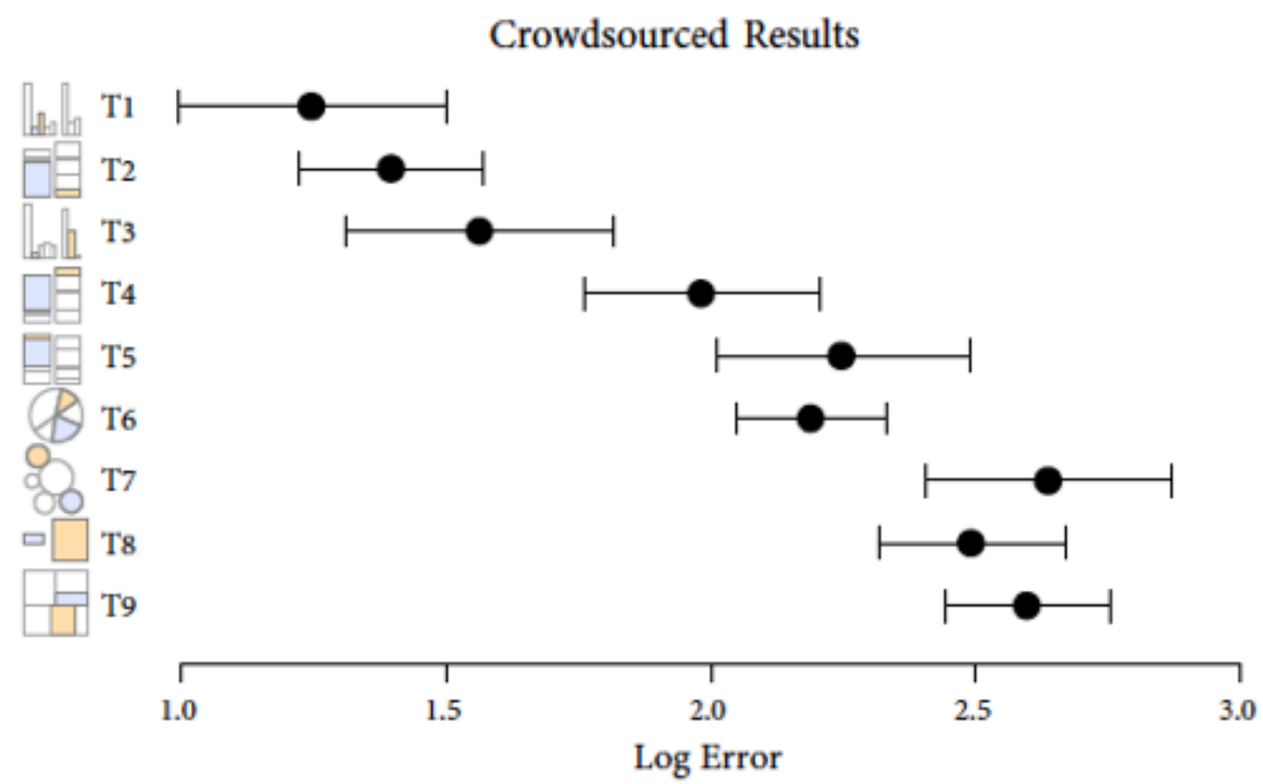
What works and what does not?

William Cleveland and Robert McGill

Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods

JASA 1984

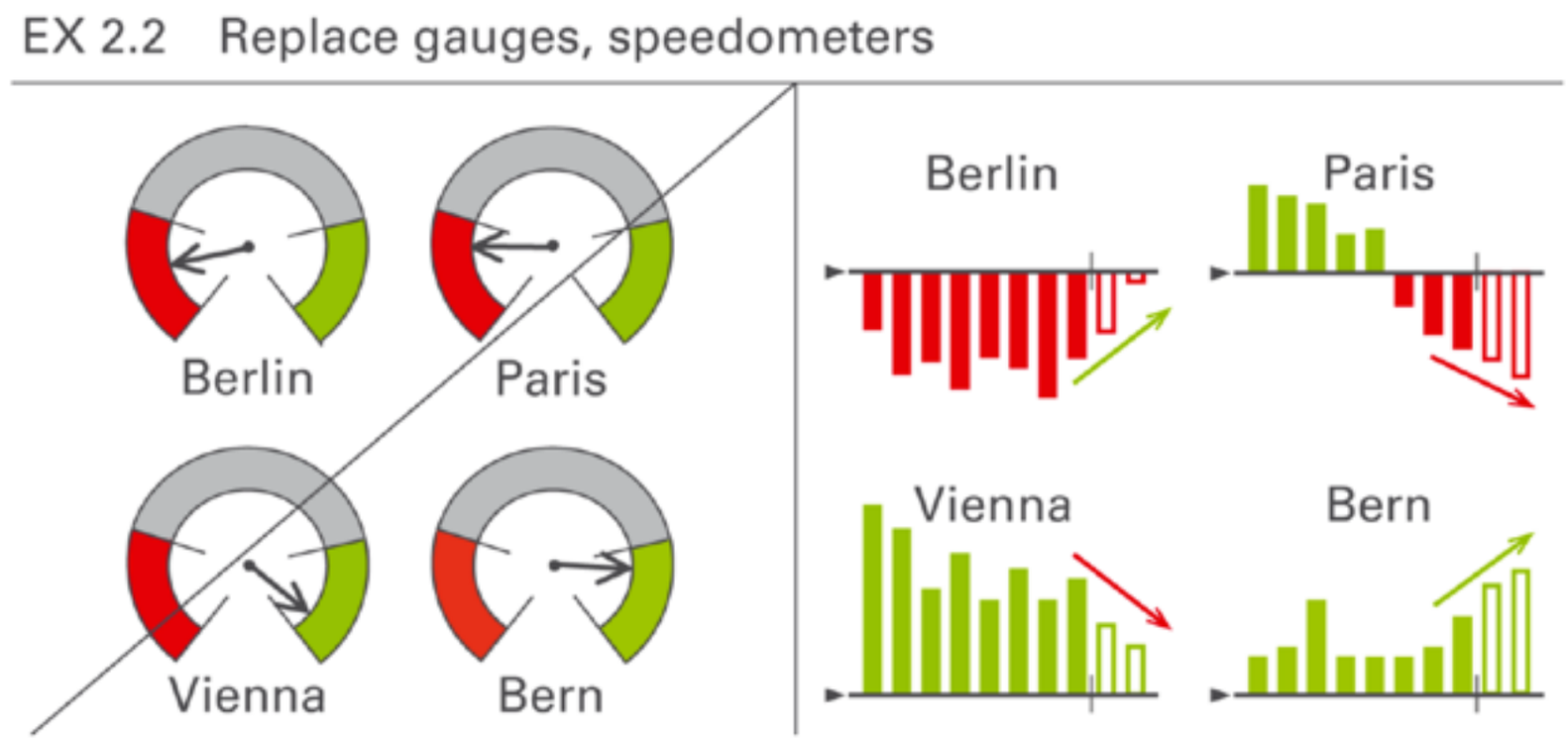




And now we even have some standards

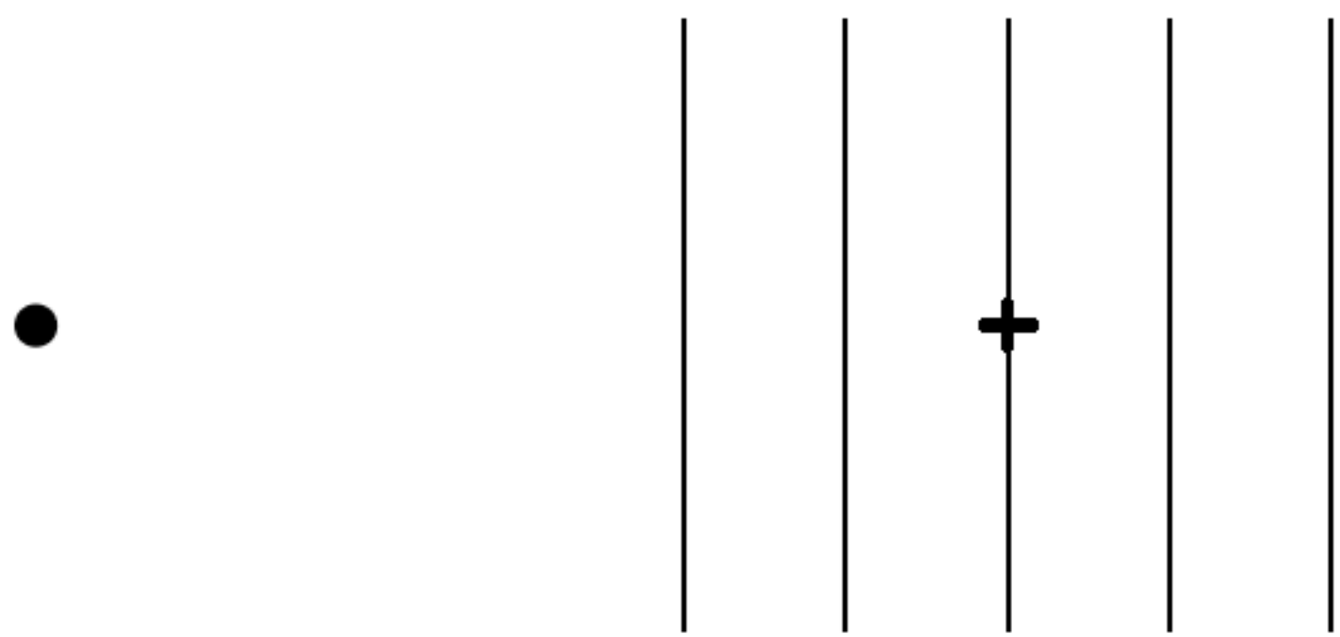
International Business Communication Standards (IBCS)

https://www.ibcs-a.org/

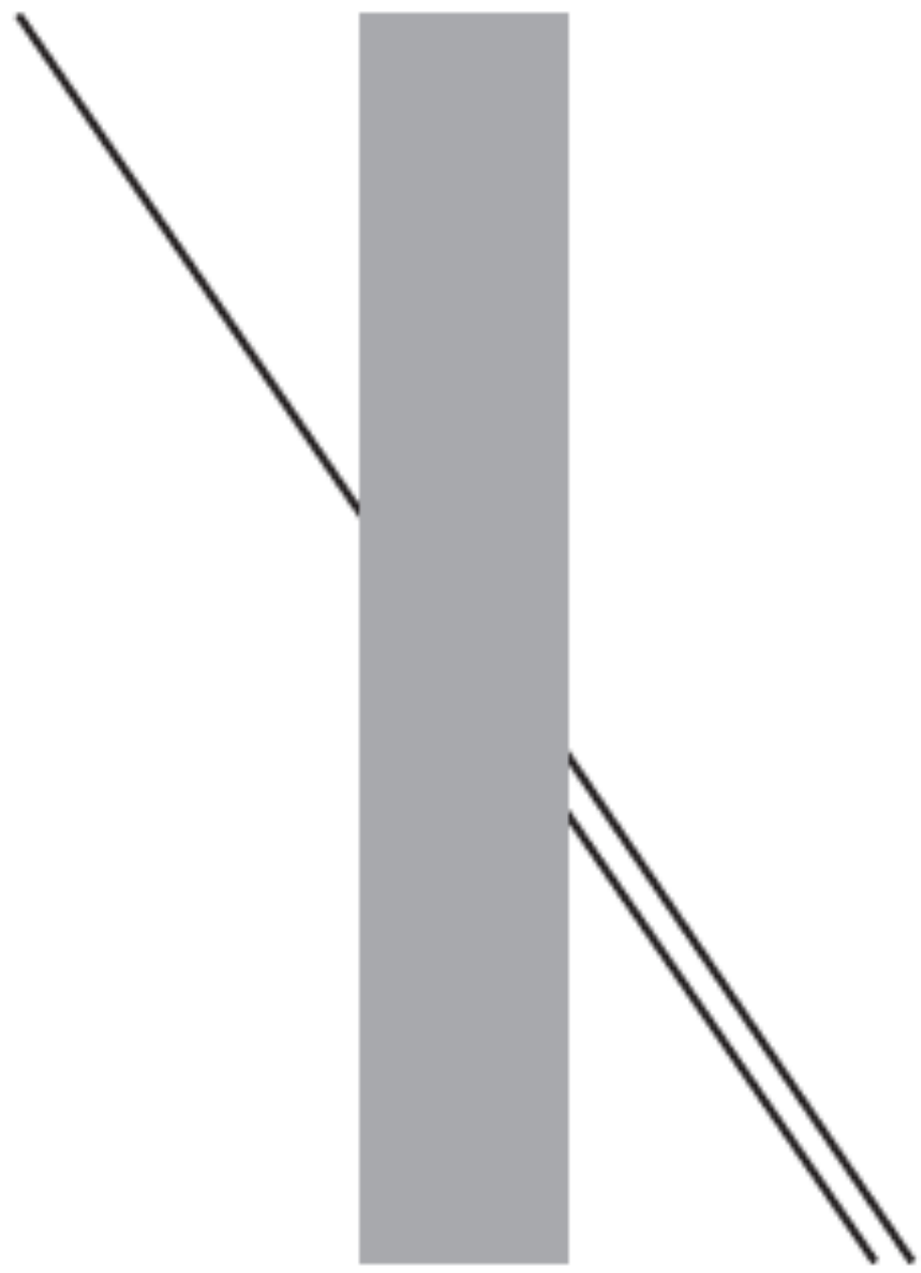


Why it’s so hard?

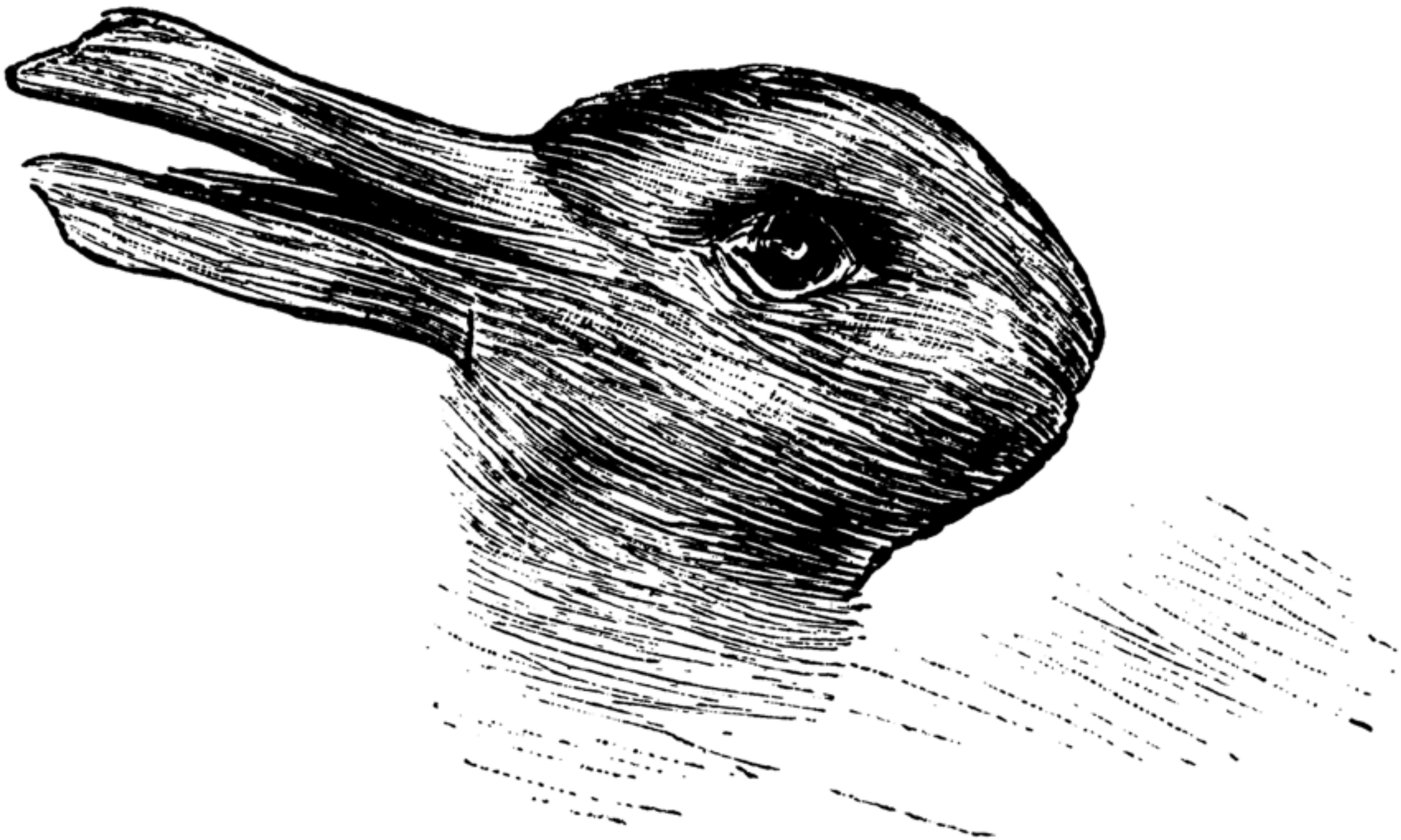
Have you heard about the blind spot?



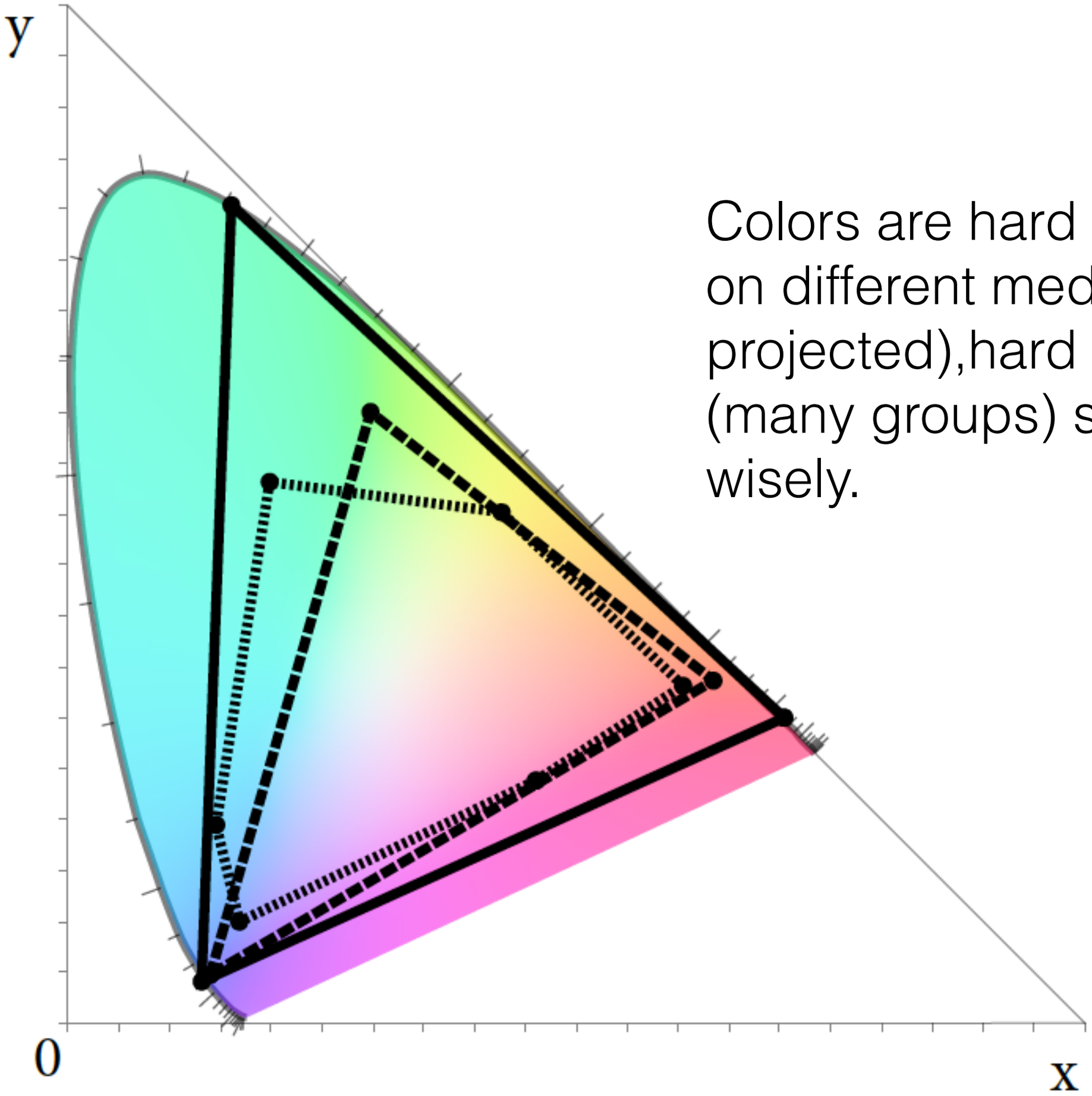
Johann Poggendorff Ilussion (1796–1877)
and there is much more illusions like this



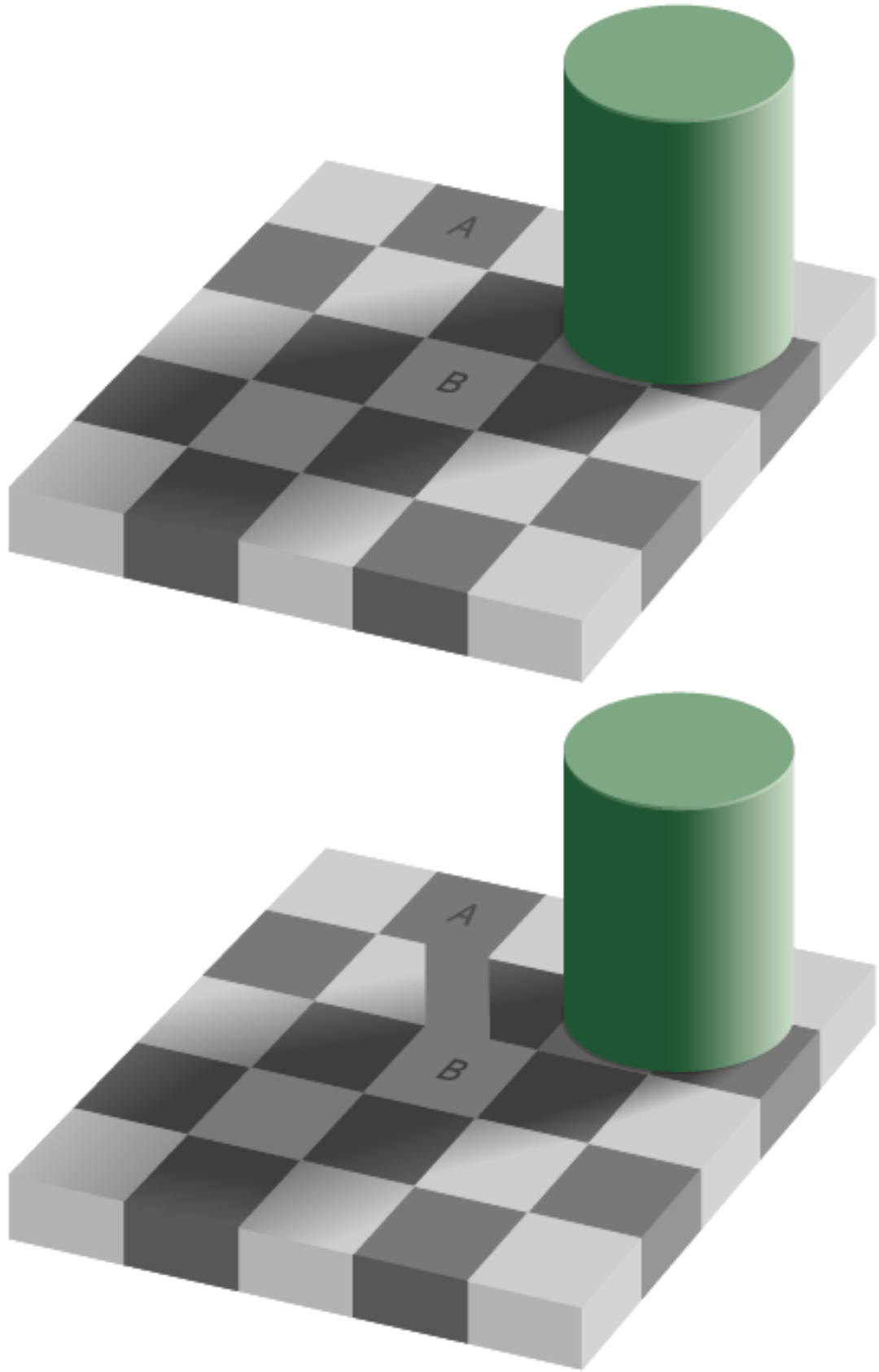
Ambiguous images
Your plot may look like that for other people



Eye see just a small fraction of an image

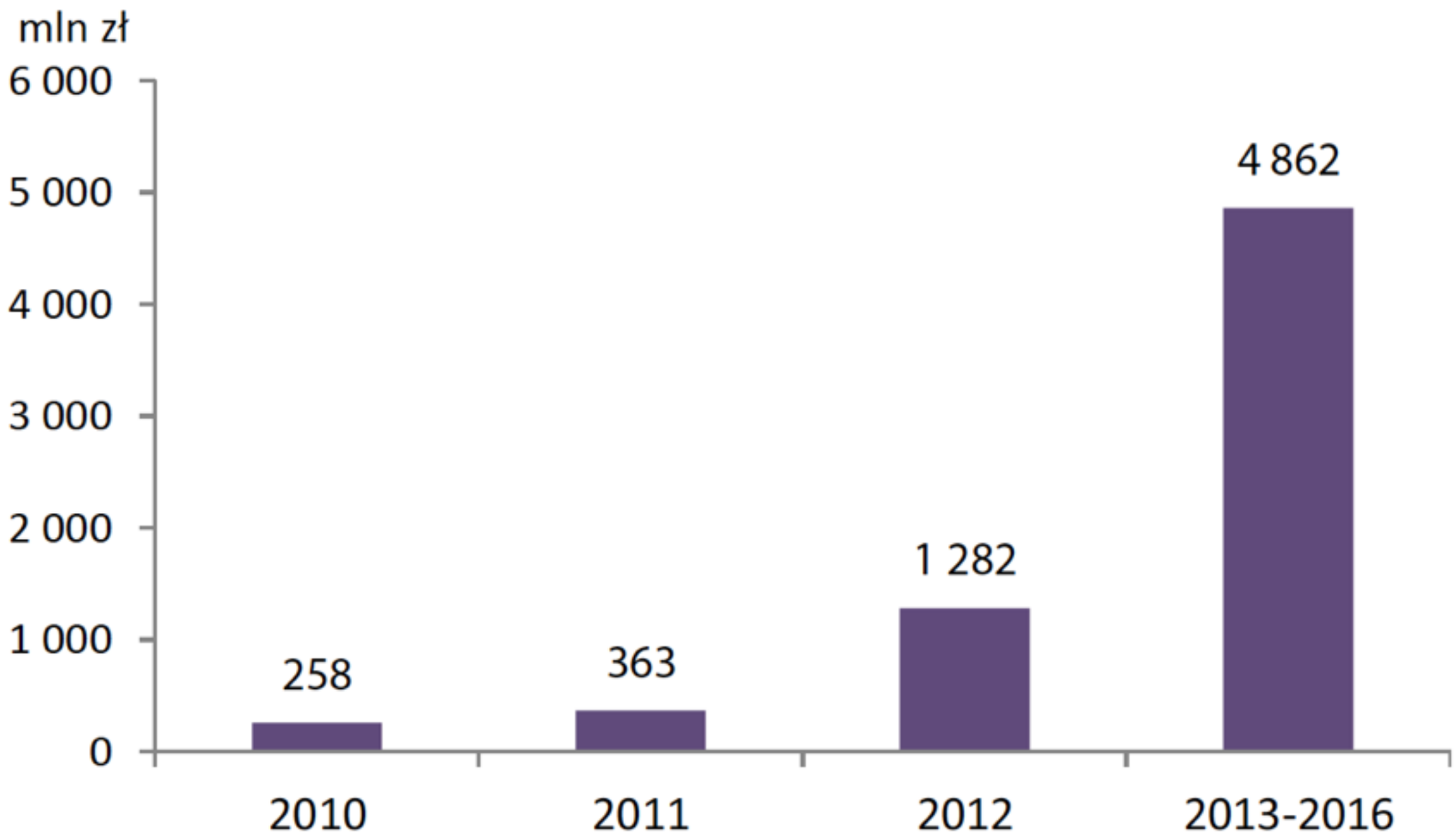


Colors are hard to reproduce on different media (printed / projected),hard to distinguish (many groups) so choose them wisely.


A
B
A
B

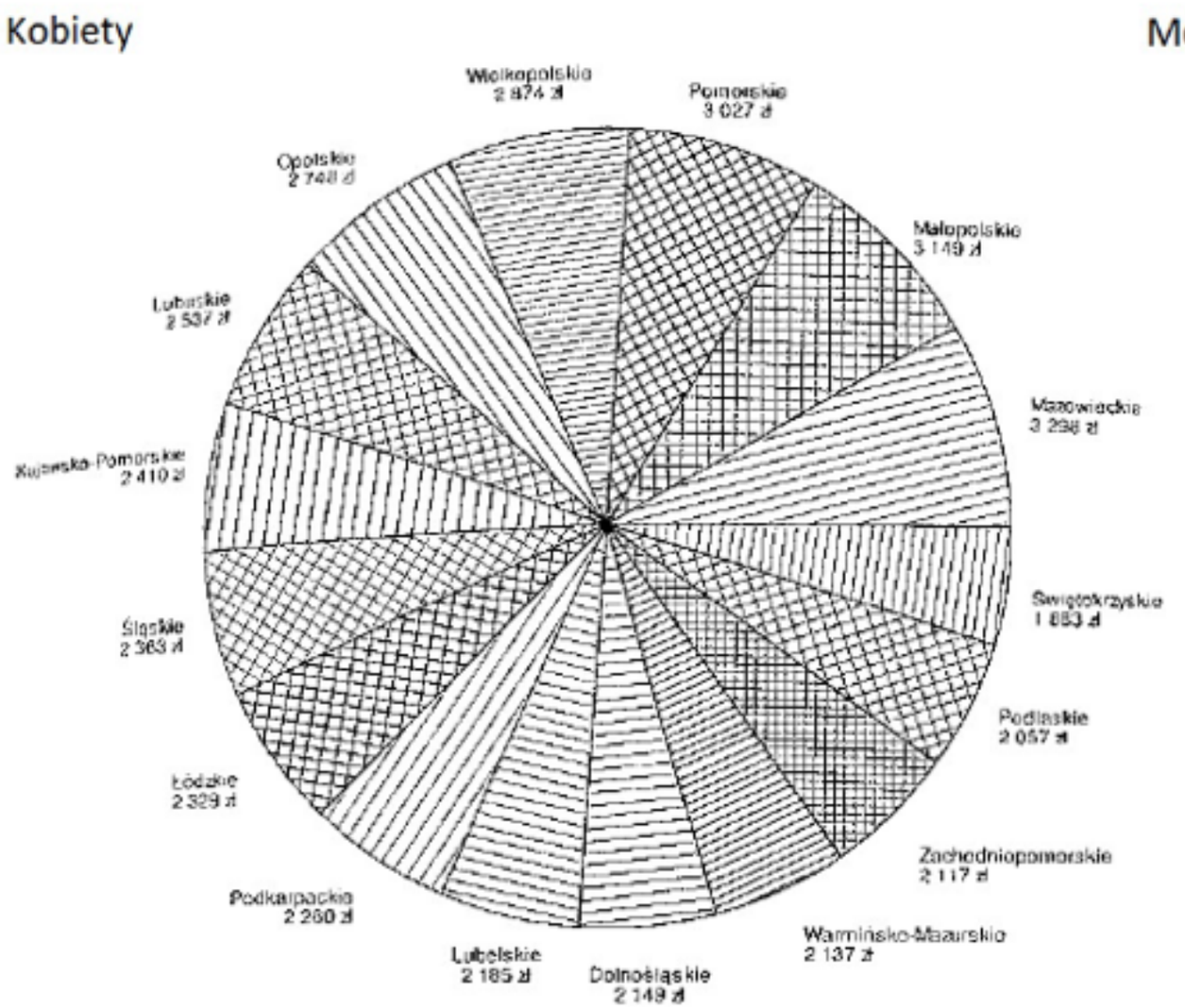
Avoid junk charts

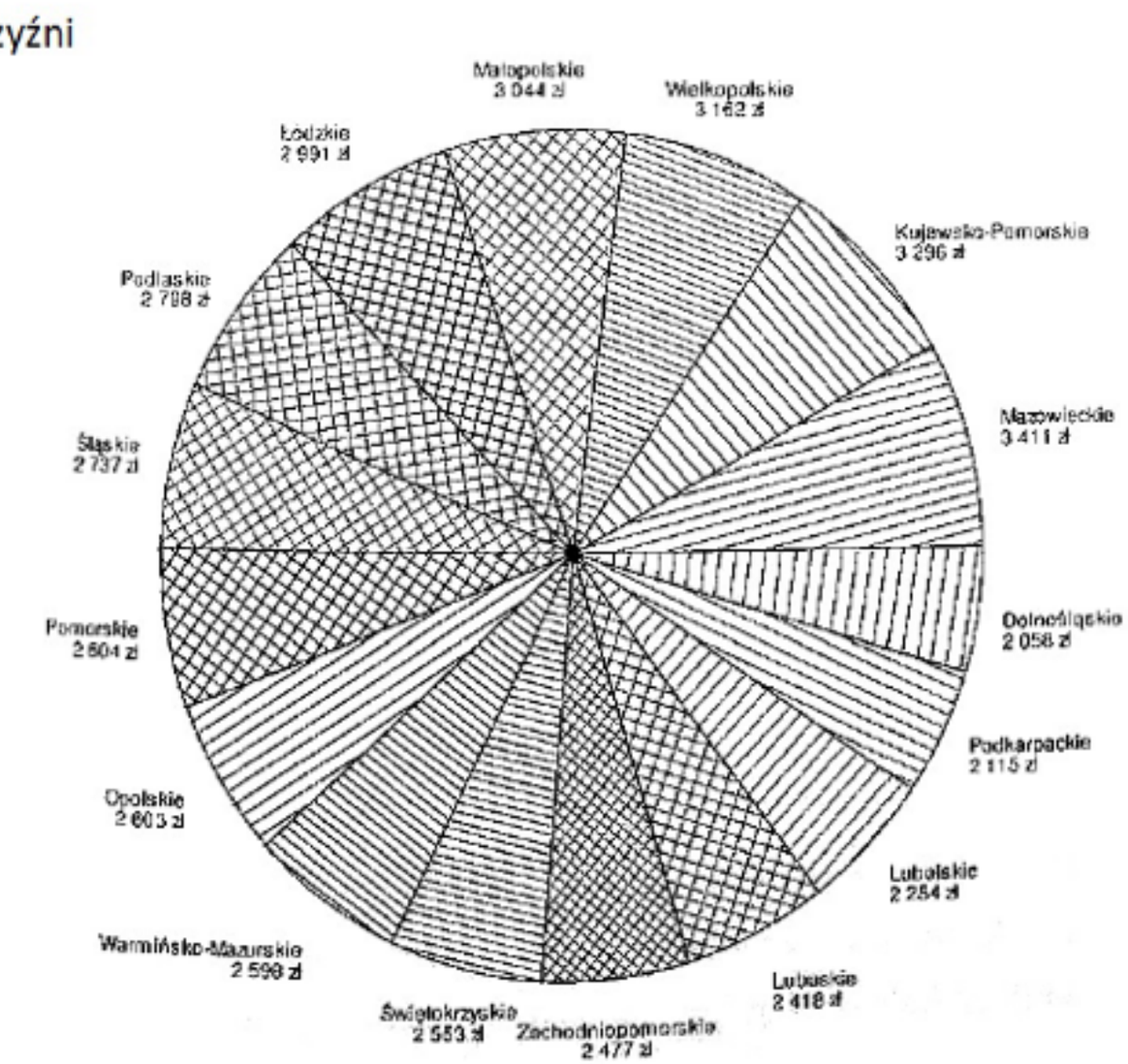
Spendings on R&D



Price index - is it a fast progression?

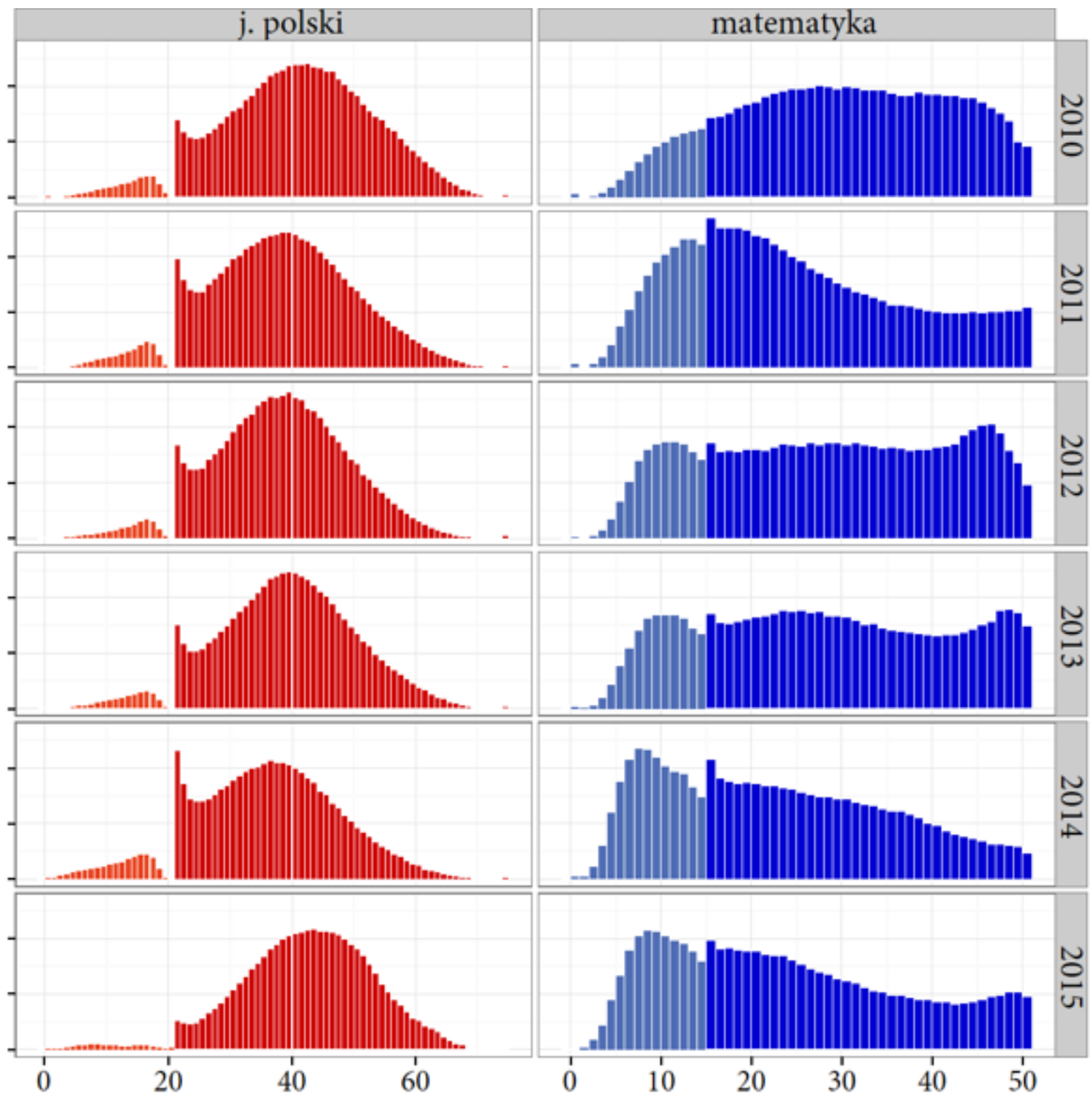
Structure of earnings, man vs. woman in different regions





Do you really need a plot?

Sometimes you do



Lot's of problems with graphical presentation.

Use good tools to avoid most of them.
There are many good packages in R for graphics.

We are going to begin with ggplot2.
Flexible and elegant engine for data visualisation.

Docs:
http://ggplot2.org/

Examples:
http://bit.ly/1nW9Cvz