Data Visualization

Data Management and Visualization

SoftEng
http://softeng.polito.it

Version 1.2.4

Definition

Encode data through visual features in order to convey useful information

Motivation

Information retrieval

- An experiment showed information pieces to two groups of subjects
- After 3 days
  - Text alone: 10% correct retrieval
  - Text + visuals: 65% correct retrieval

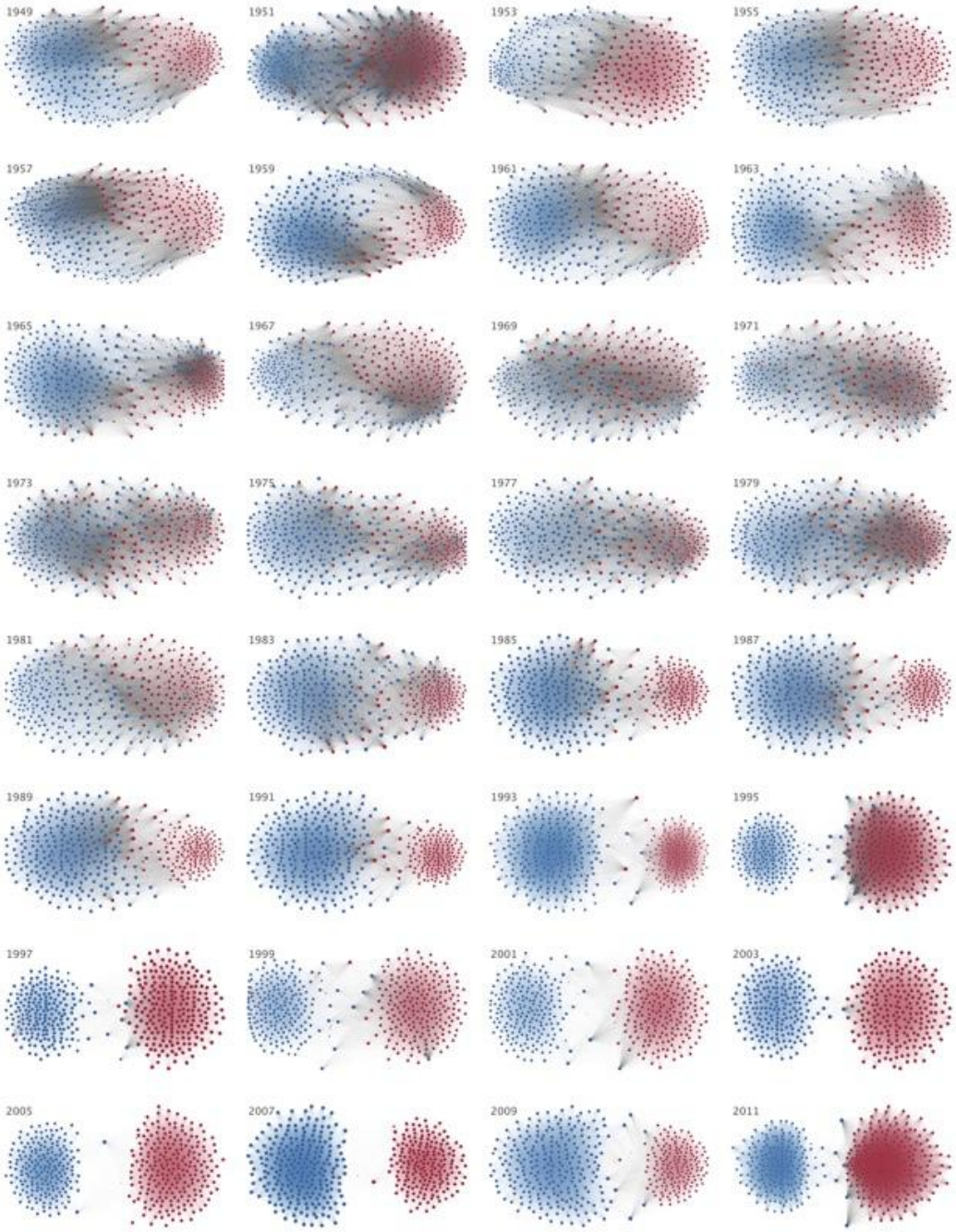
[John Medina, Brain Rules, 2008]

Motivation

Information retrieval

Information density

- In principle every single pixel in an image could encode a datum
  - Screen (1024x768) ~ 1 M pixels
  - 1 M characters ~ 250 pages



http://www.mamartino.com/projects/rise_of_partisanship/index.html

Motivation

Information retrieval

Information density

Information context

Visualization compares multiple values and puts the information into context.

A single number means nothing.

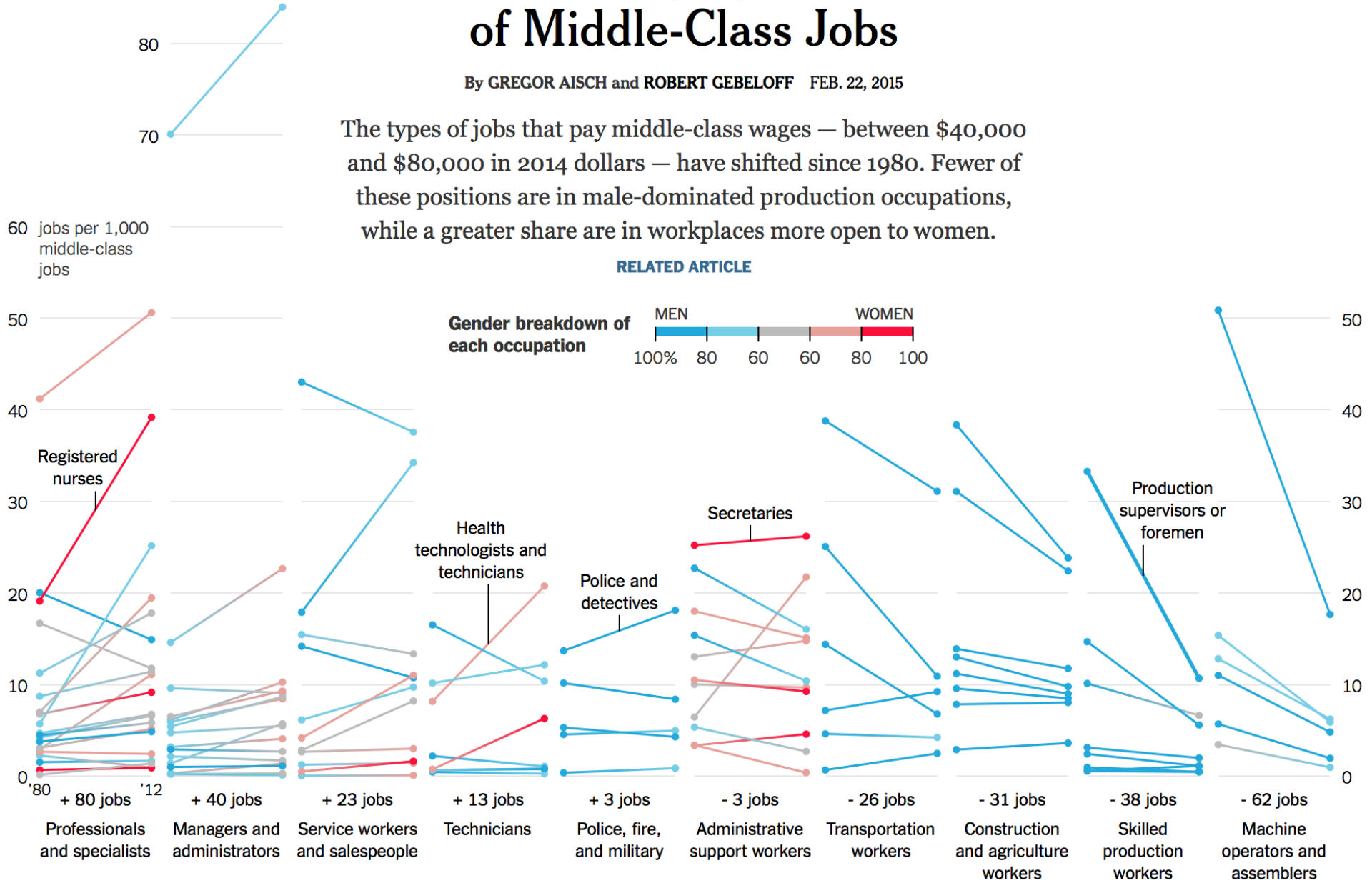
[Randy Krum presentation at Malofiej 23 (March 2015)]

The Changing Nature of Middle-Class Jobs

By GREGOR AISCH and ROBERT GEBELOFF FEB. 22, 2015

The types of jobs that pay middle-class wages — between $40,000 and $80,000 in 2014 dollars — have shifted since 1980. Fewer of these positions are in male-dominated production occupations, while a greater share are in workplaces more open to women.

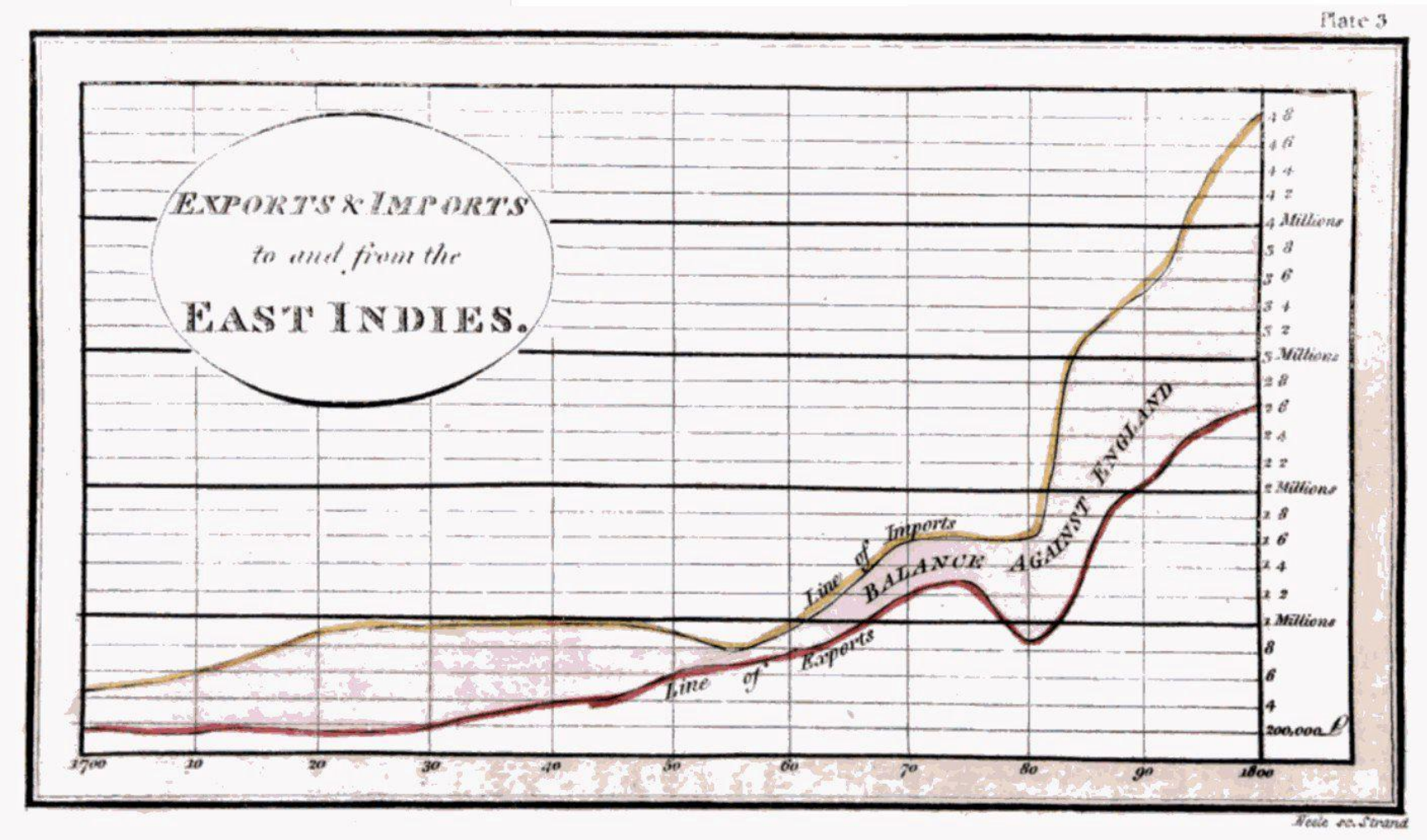
RELATED ARTICLE



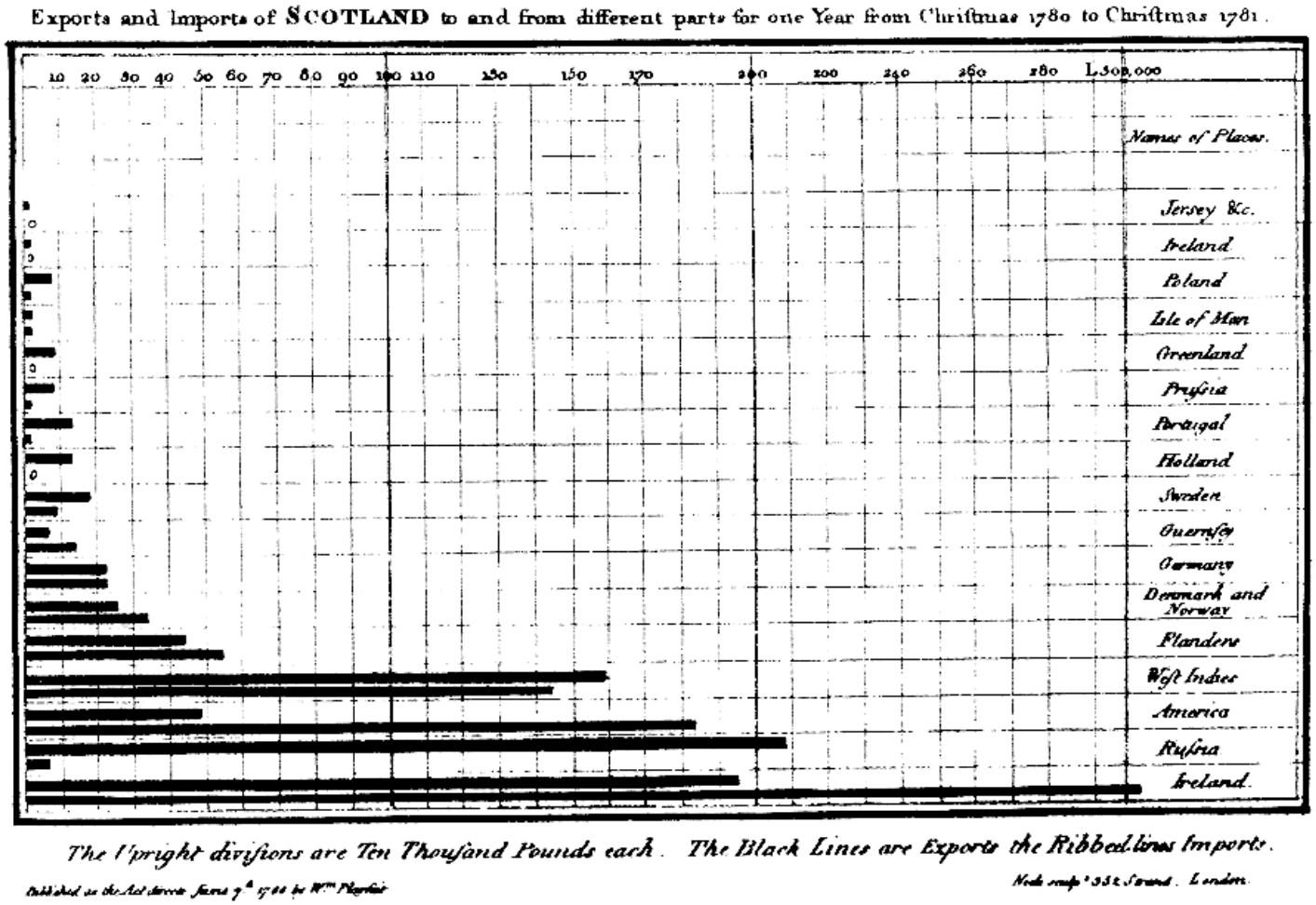
http://www.nytimes.com/interactive/2015/02/23/business/economy/the-changing-nature-of-middle-class-jobs.html

HISTORY

William Playfair – 1786

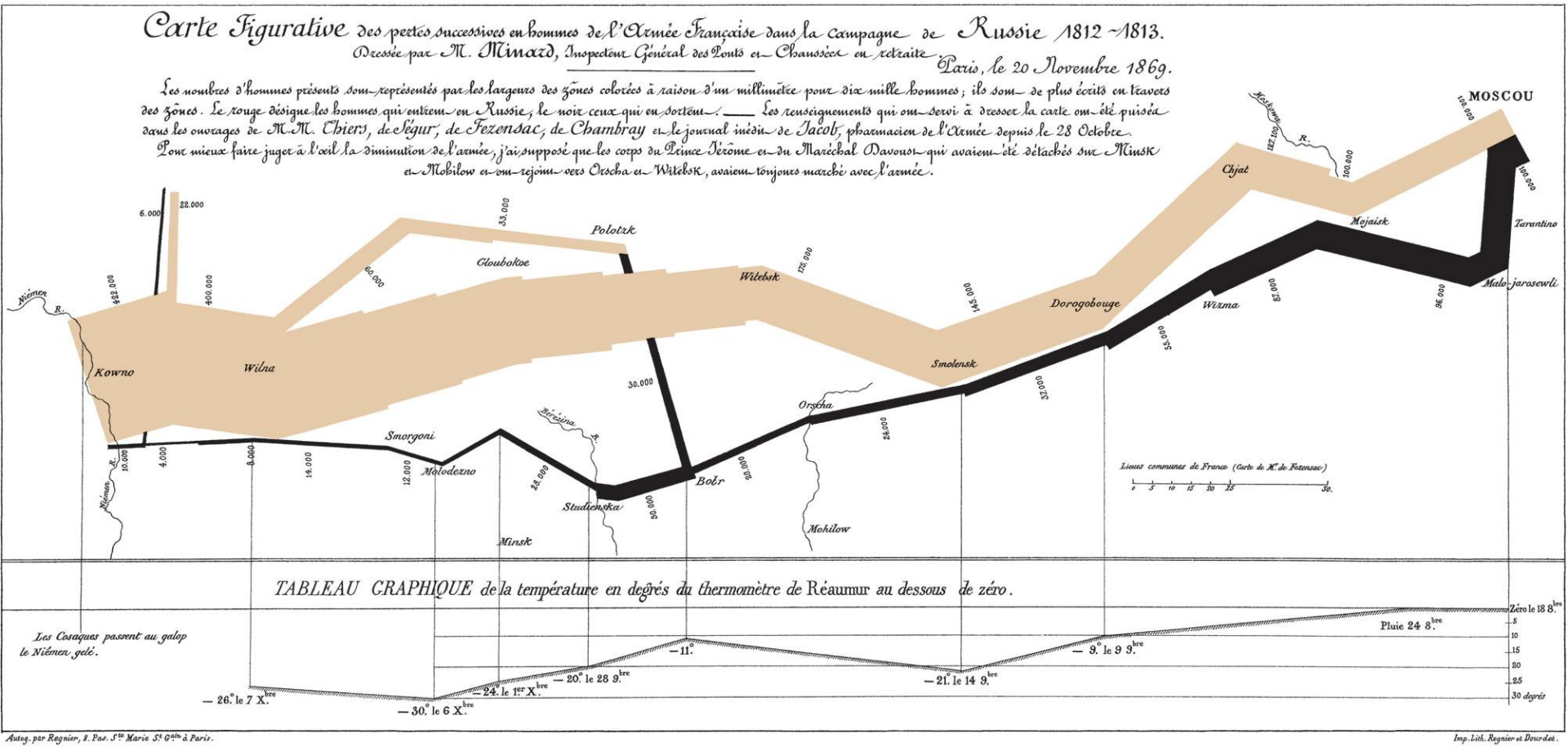


W.Playfair, The Commercial and Political Atlas, London 1786



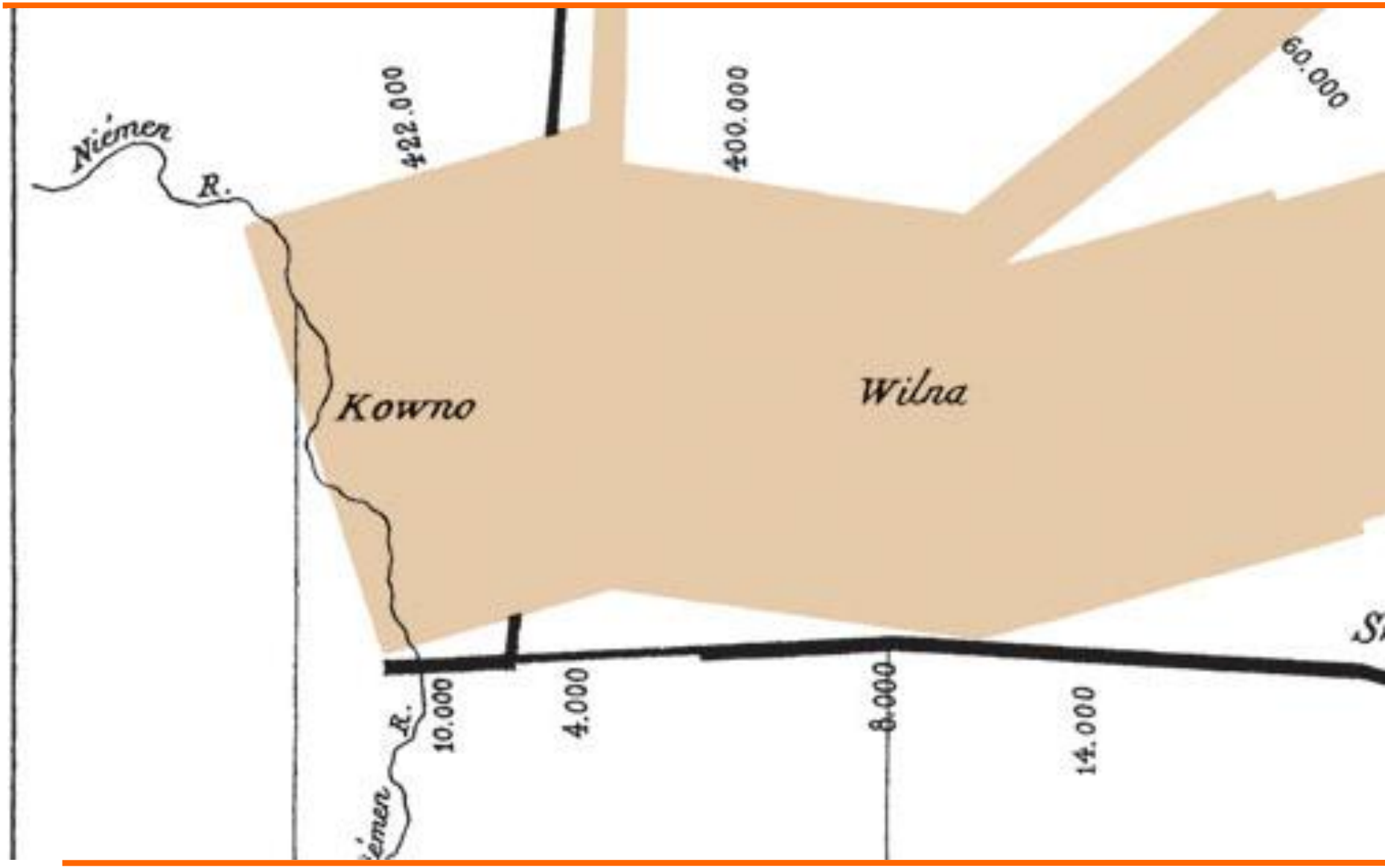
W.Playfair, The Commercial and Political Atlas, London 1786

Charles Joseph Minard – 1869

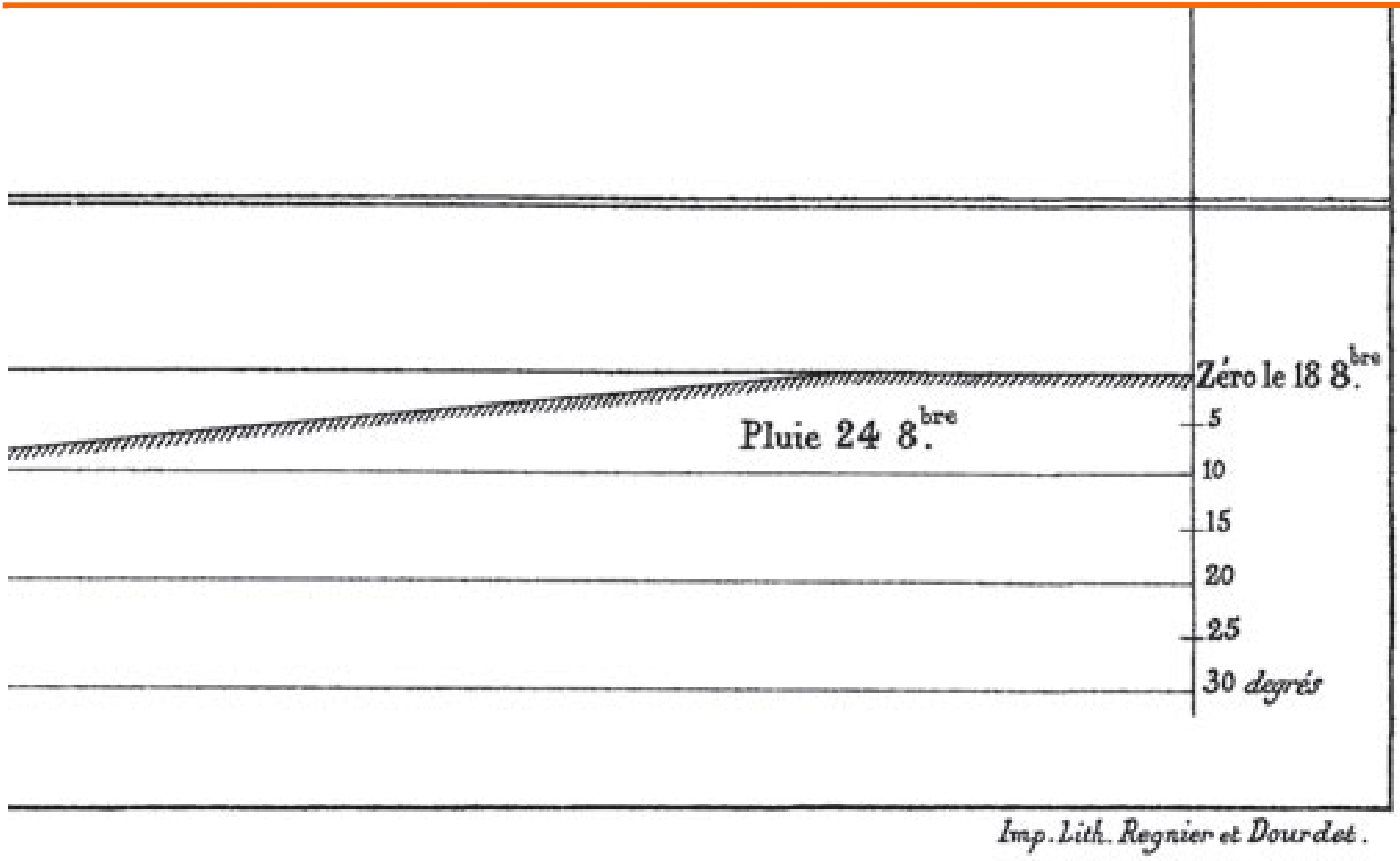


https://en.wikipedia.org/wiki/Charles_Joseph_Minard

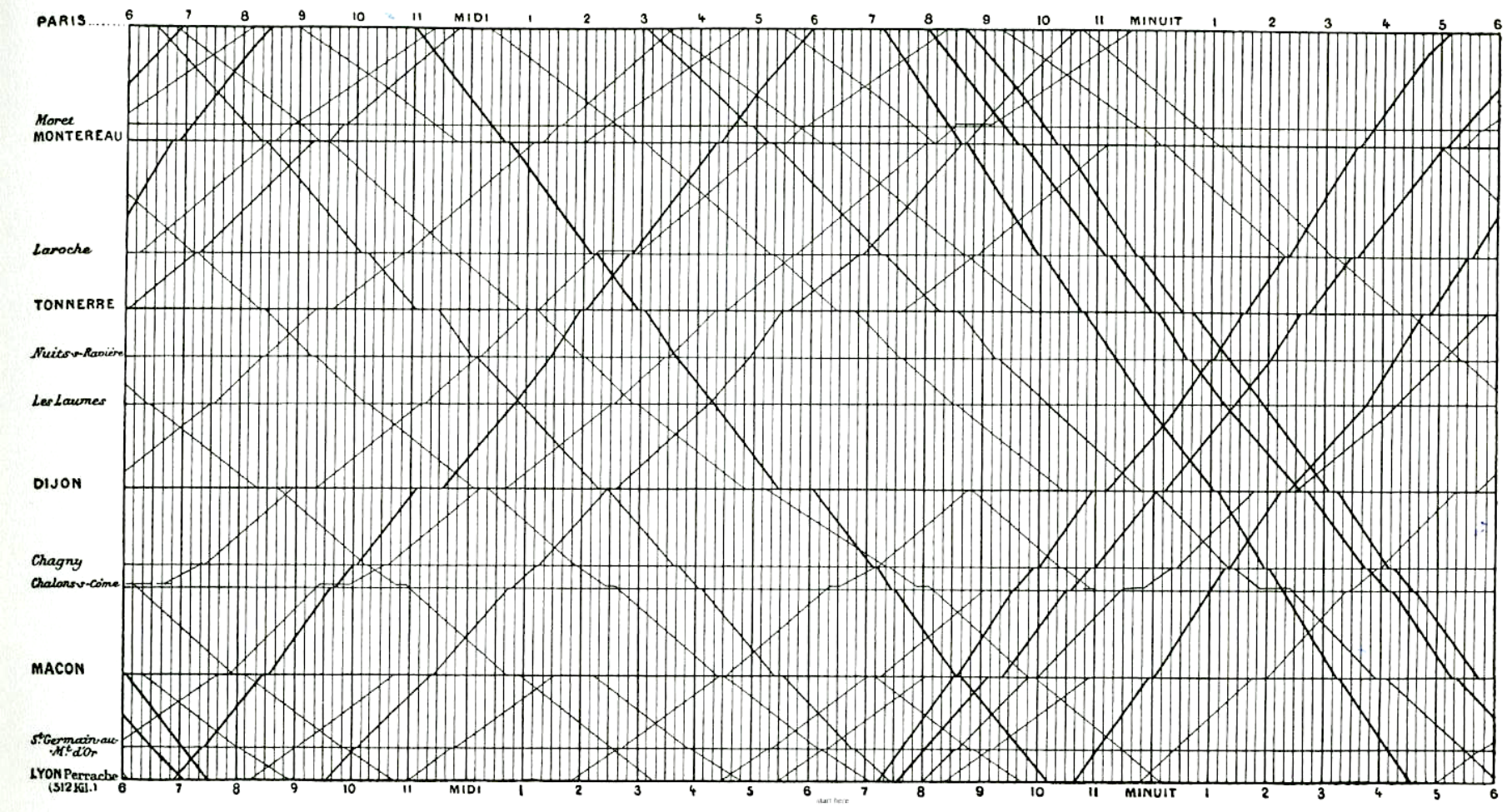
Numbers and direction



Temperature



Étienne-Jules Marey – 1885



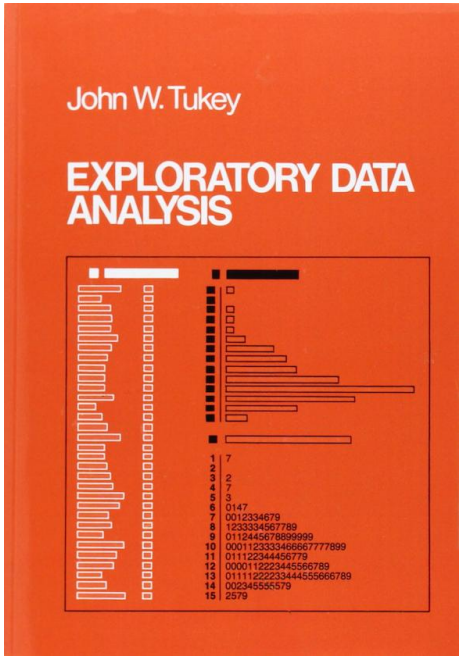
La Méthode graphique dans les sciences expérimentales et principalement en physiologie et en medicine, 1885

https://archive.org/details/lamthodegraphiq00maregoog

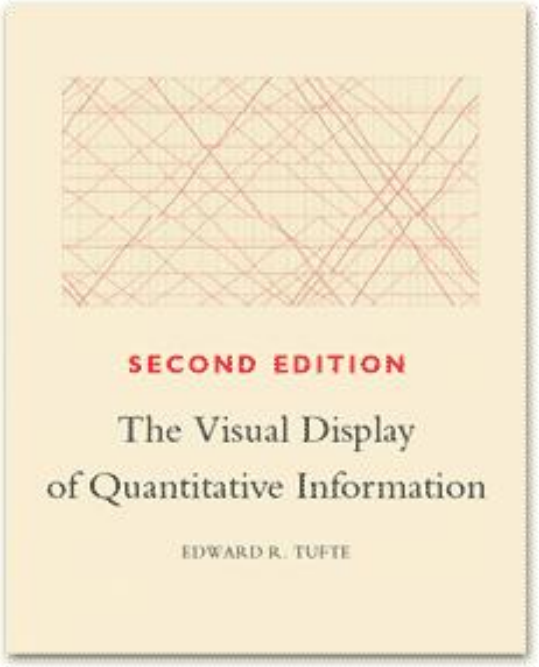
XX Century

- http://www.datavis.ca/milestones/

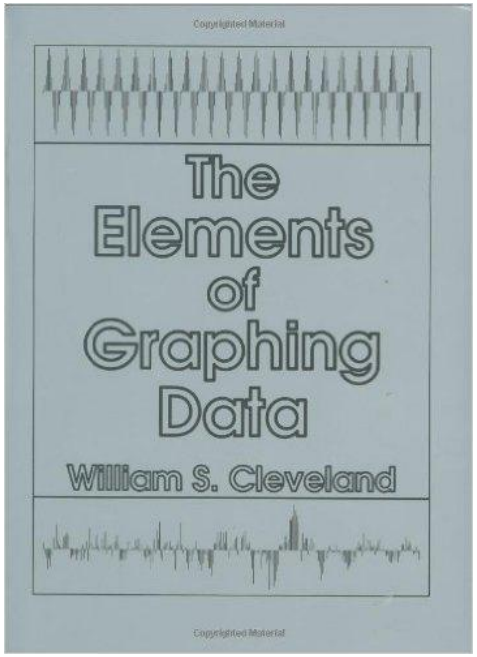
1977



1983



1985



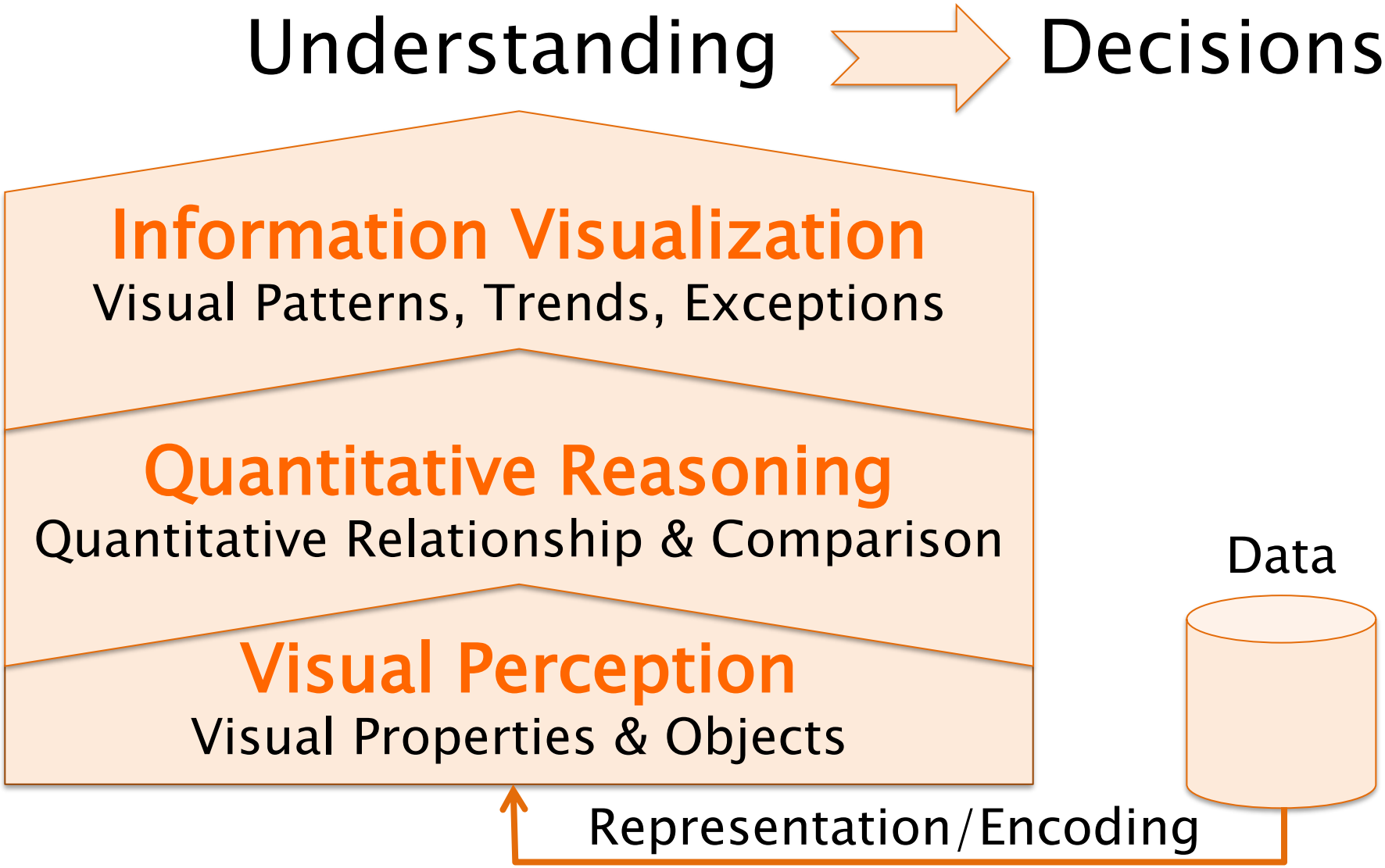
INFORMATION VISUALIZATION

Information visualization

The use of computer-supported, interactive, visual representations of abstract data to amplify cognition

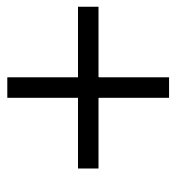
Readings in Information Visualization: Using Vision to Think.
S.K.Card, J.D.Mackinlay, and B.Shneiderman, Academic Press, 1999

Overview


Understanding
Decisions
Information Visualization
Visual Patterns, Trends, Exceptions
Quantitative Reasoning
Quantitative Relationship & Comparison
Visual Perception
Visual Properties & Objects
Data
Representation/Encoding

Quantitative message

- Quantitative values
  - Express measures



- Categories
  - Identify what entities the values refer to
  - Define groups of entities

Understanding tasks

- Variation within quantitative measures
  - Distribution
  - Deviation
  - Correlation
- Variation within category
  - Ranking
  - Part-to-whole
  - Time
  - Space
- Multivariate relations

Visualization instruments

- Tables
  - Textual information
- Graphs
  - Visual information

Tables

- Main features
  - Data arranged in rows and columns
  - Data encoded as text
- Strengths
  - Easy look-up of values
  - Precise values
    - Allow selected comparisons
  - Several units of measure are possible

Graphs

- Main features
  - One or more axes delineate the display area where values are shown
  - Values encoded as visual objects in relation to axes
  - Axes provide scales
    - Assign values and labels to visual objects
    - Both categorical and quantitative
- Strengths
  - Overall shape of data (holistic)

Graphs

- Show
  - Trend
    - Pattern of change over time
  - Comparison of subsets
    - Overall
    - Spot similarities and differences
  - Highlight exceptions
- Display relationships among multiple quantitative values by giving them shape

In general

Use tables to	Use graphs to
Look up individual values	Focus on the shape of values
Compare individual values	Reveal relationships among multiple values
Precise values are required	
There is more than one unit of measure	

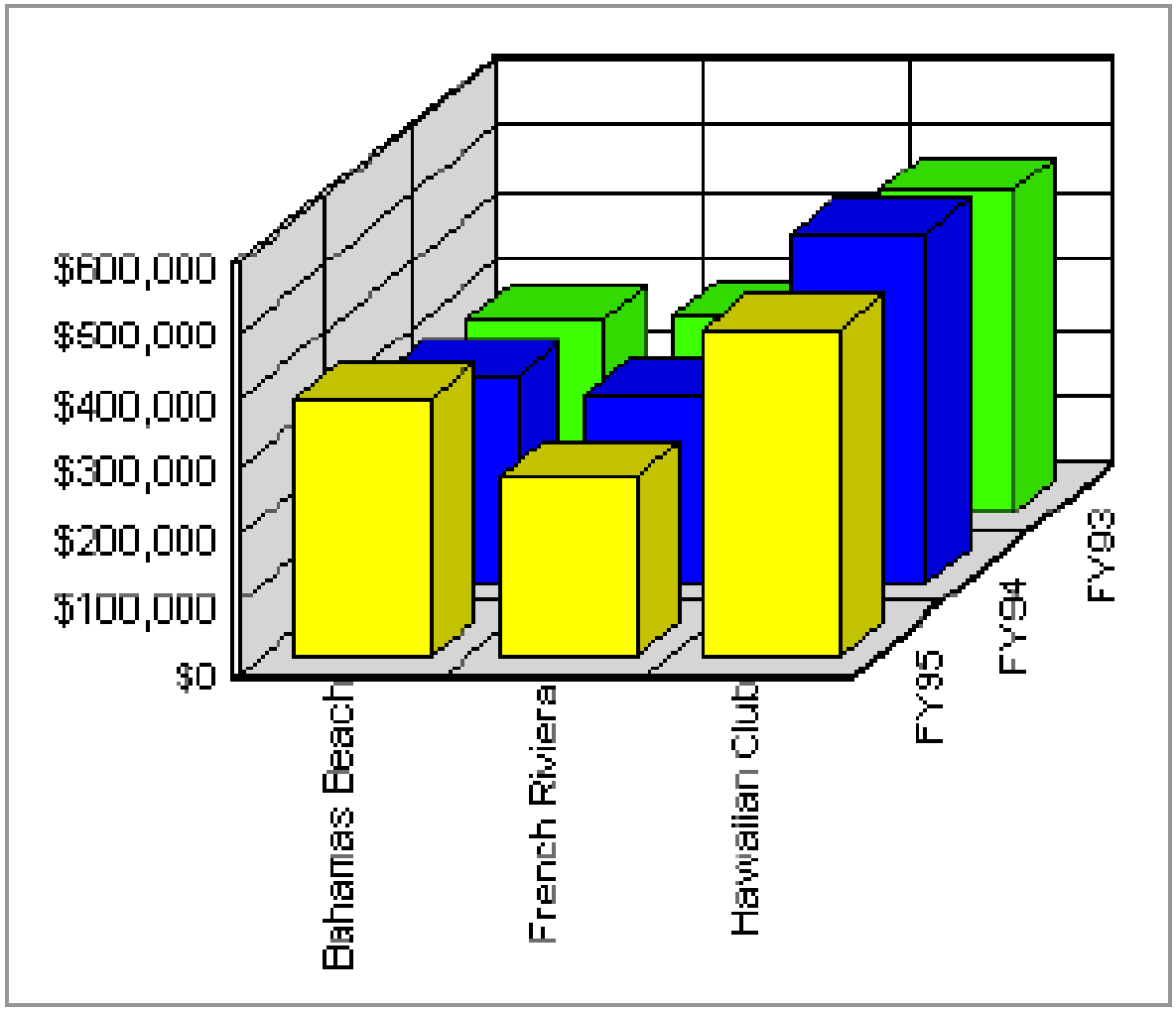
EXAMPLES

Good and Poor visualization

- Like good writing, good graphical displays of data communicate ideas with clarity, precision, and efficiency.
- Like poor writing, bad graphical displays distort or obscure the data, make it harder to understand or compare, or otherwise thwart the communicative effect which the graph should convey.

Friendly, Michael, and Daniel J. Denis. (2001)
"Milestones in the history of thematic cartography, statistical graphics, and data visualization."
http://www. datavis. ca/milestones

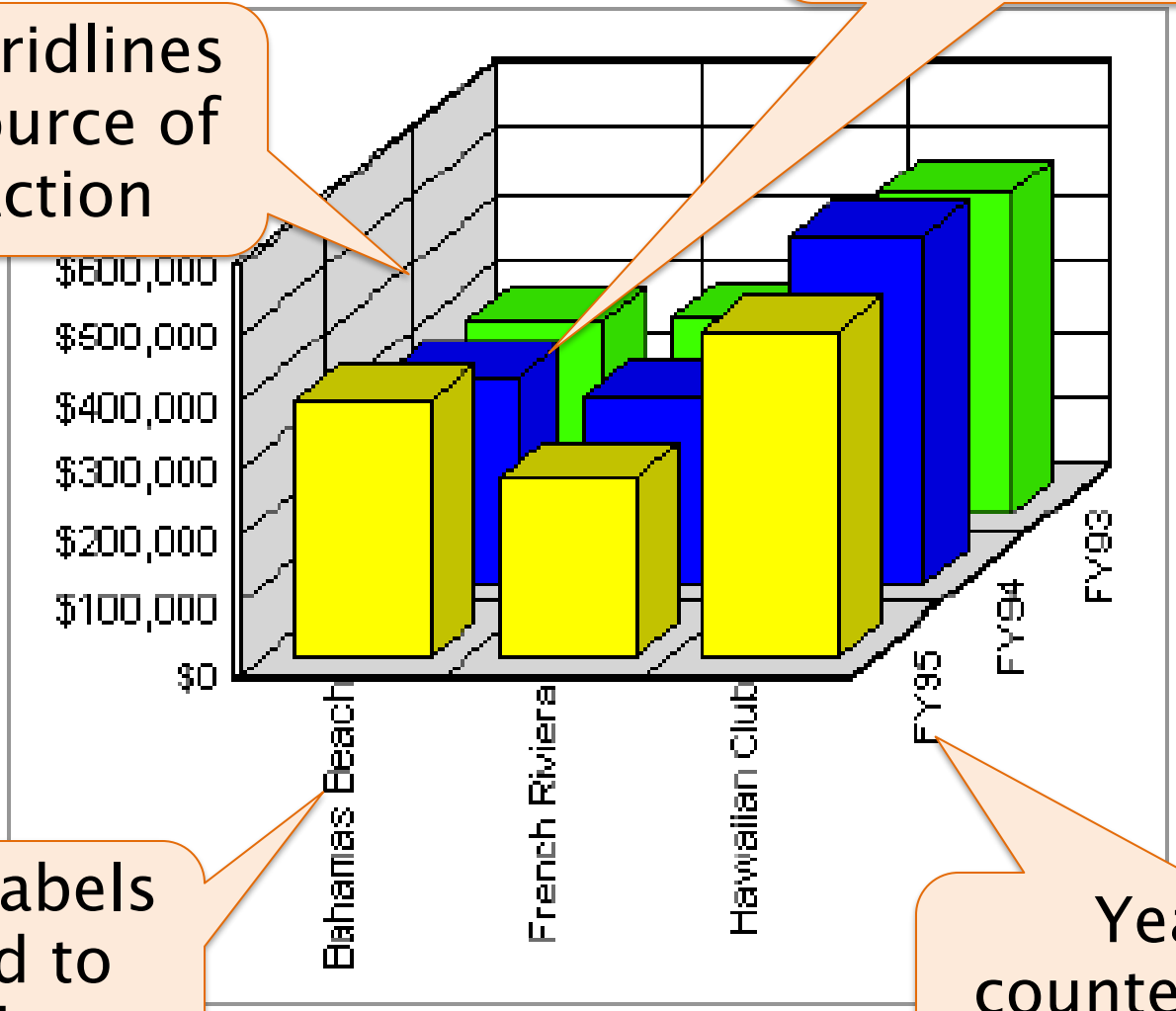
A bar graph


$600,000
$500,000
$400,000
$300,000
$200,000
$100,000
$0
Bahamas Beach
French Riviera
Hawaiian Club
FY95
FY94
FY93

A bad bar graph

3D bars are impossible to read

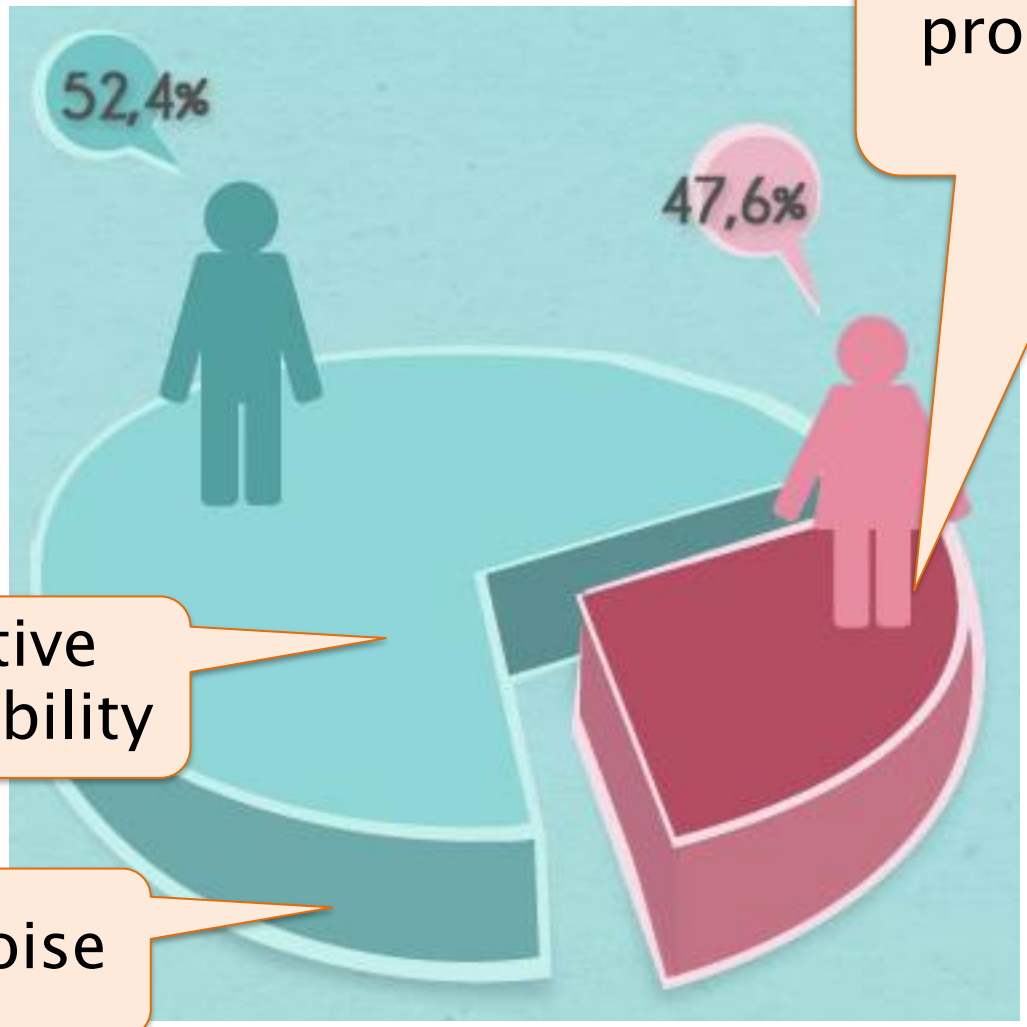
Heavy gridlines are a source of distraction



Vertical labels are hard to read

Years run counterintuitively from back to front

A pie chart

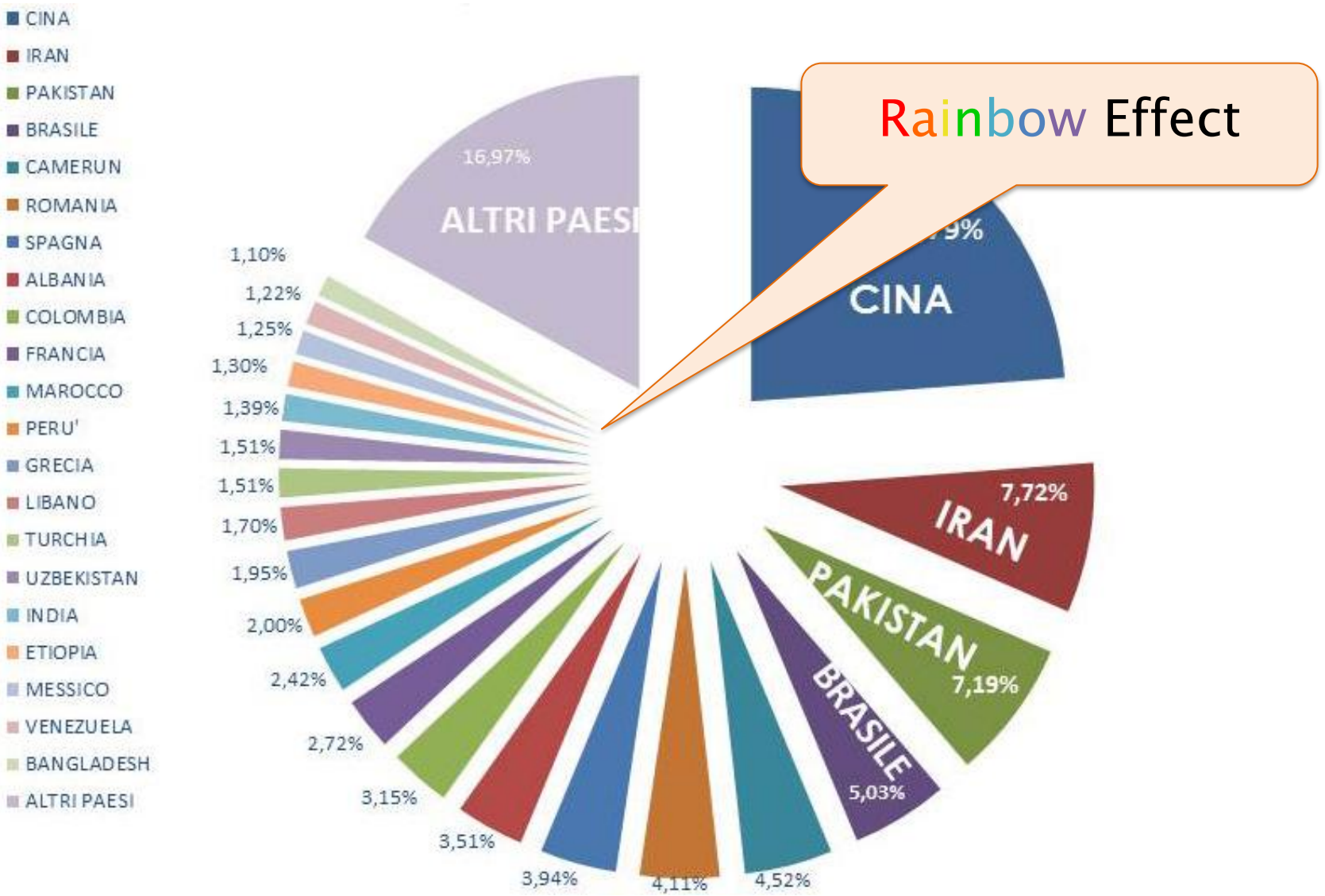


Size of slices not proportional to values

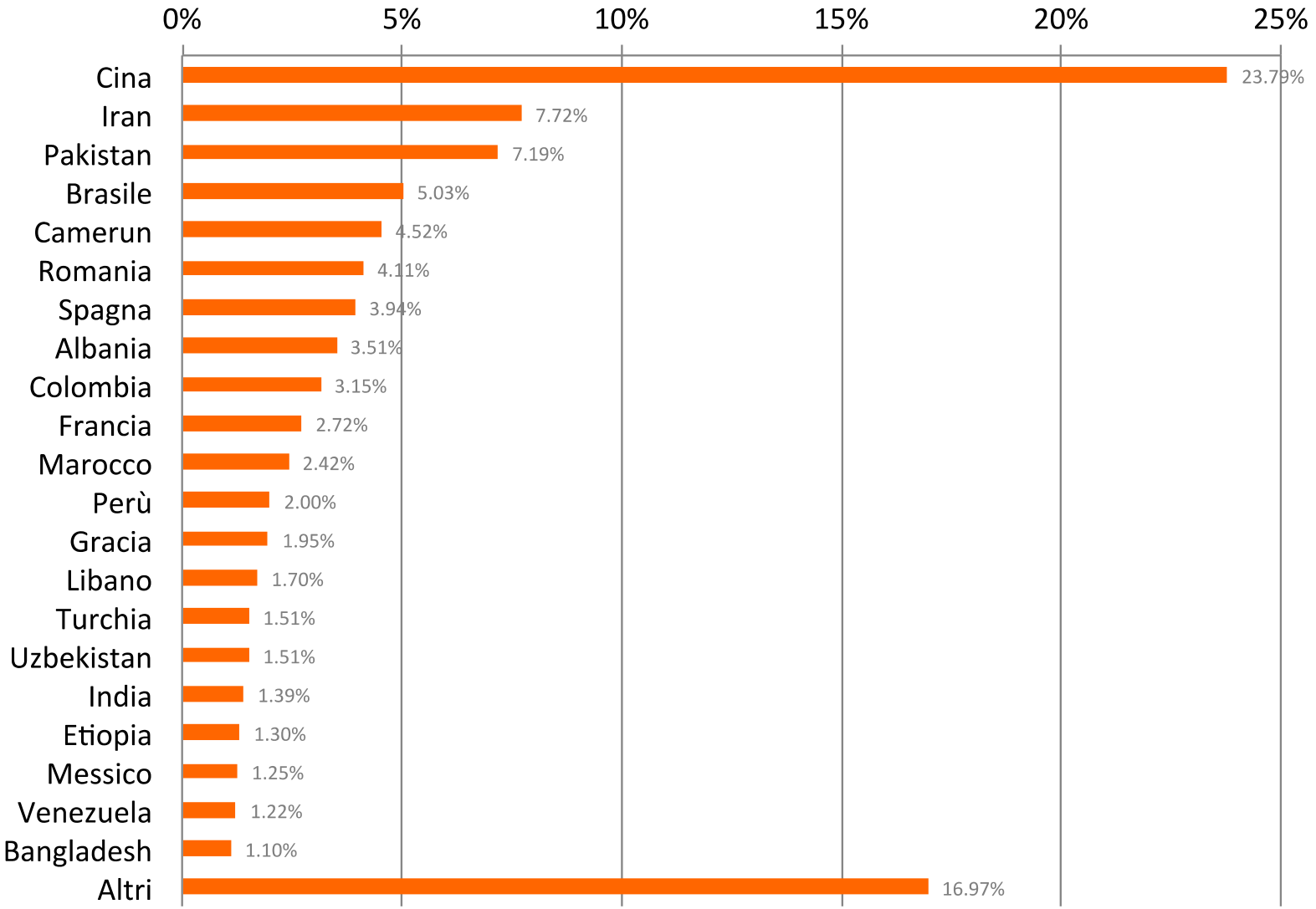
3D Perspective worsen readability

Depth add noise

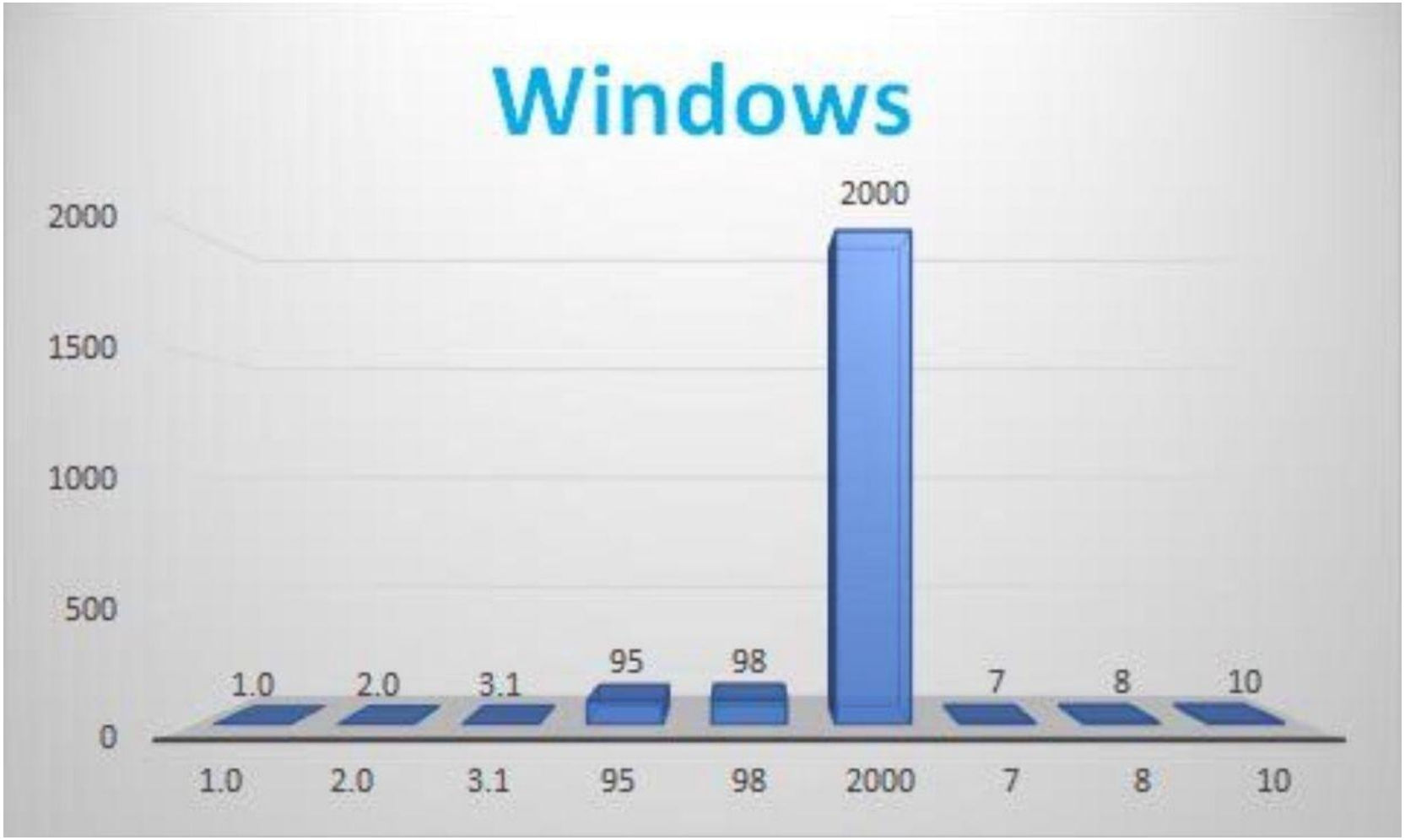
Pie chart (original)



Bar chart (redesign)

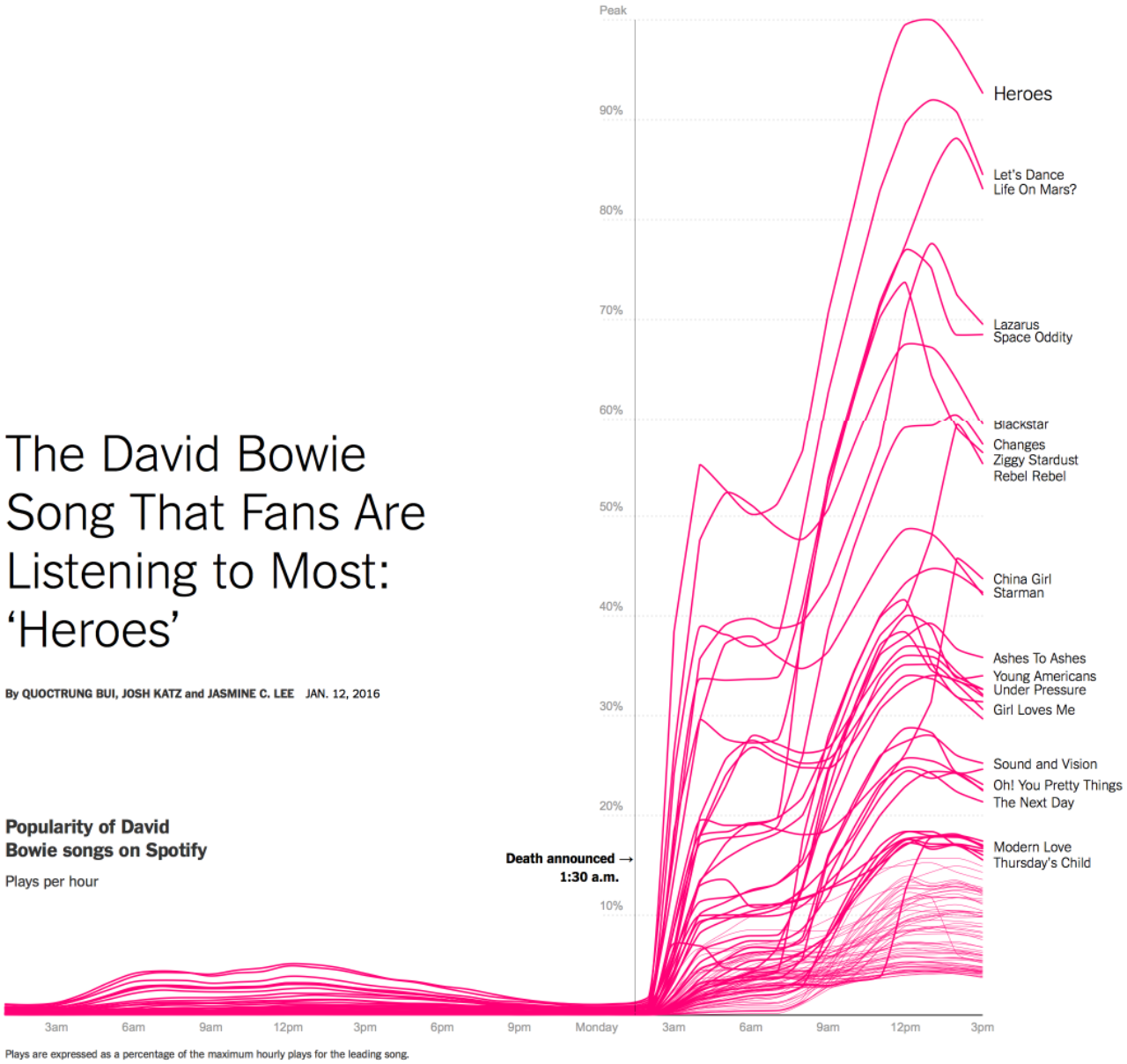


Meaningless Data



The David Bowie Song That Fans Are Listening to Most: ‘Heroes’

By QUOCTRUNG BUI, JOSH KATZ and JASMINE C. LEE JAN. 12, 2016

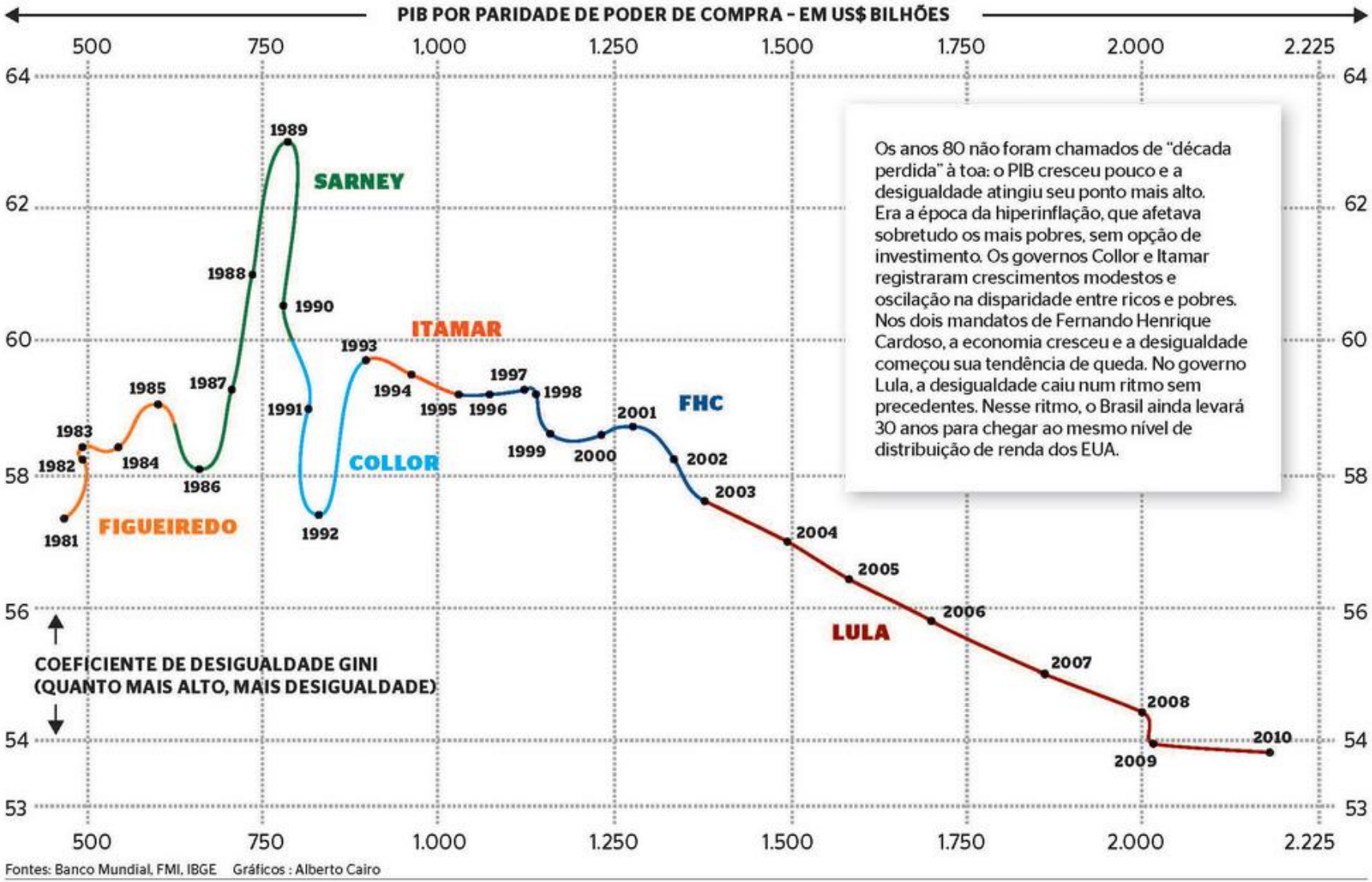


Plays are expressed as a percentage of the maximum hourly plays for the leading song.

http://www.nytimes.com/interactive/2016/01/12/upshot/david-bowie-songs-that-fans-are-listening-most-heroes-starman-major-tom.html

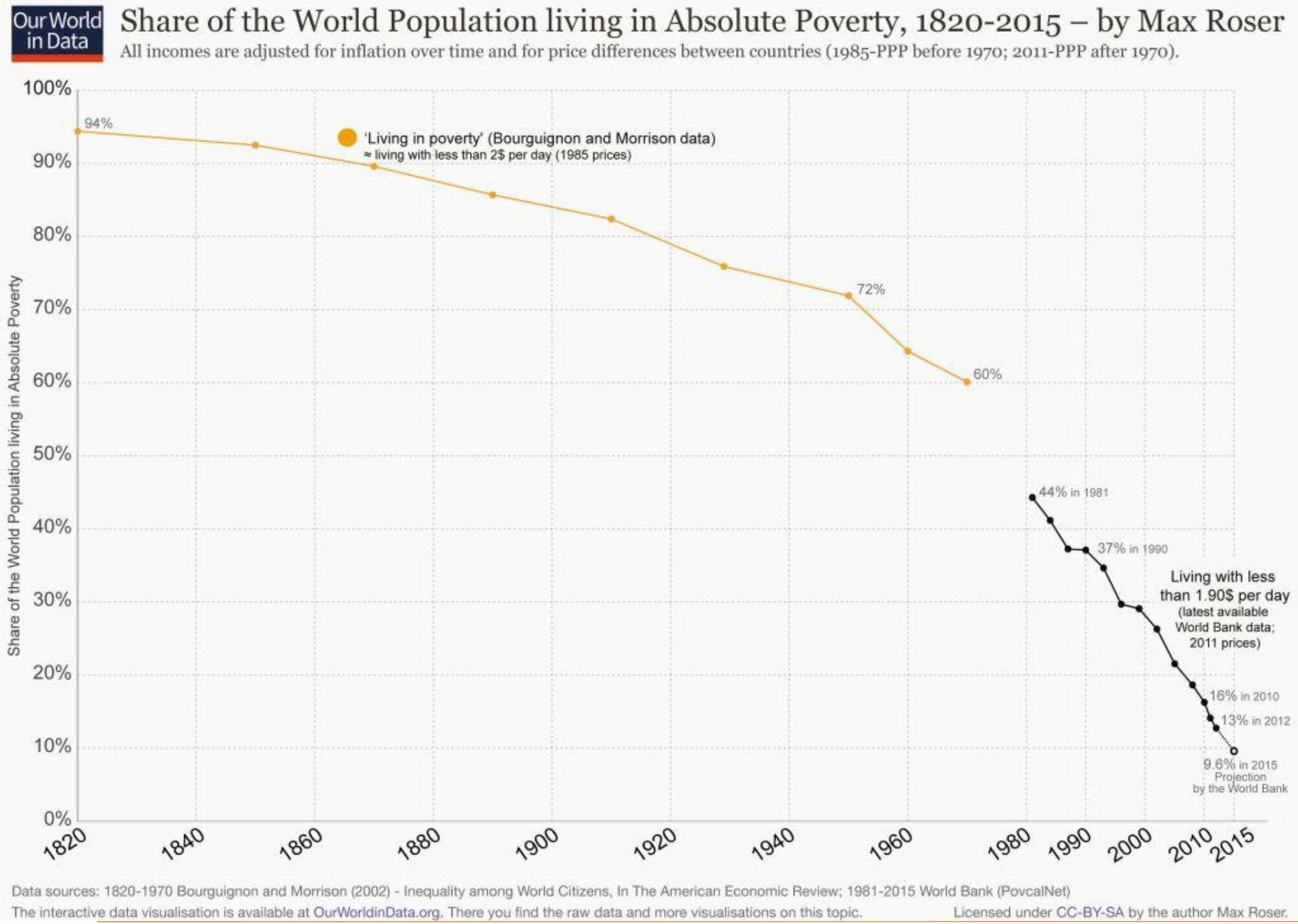
Quando o PIB cresce, nem sempre a desigualdade cai

O gráfico abaixo mostra o avanço do PIB comparado à evolução da desigualdade no Brasil desde 1980. Nem sempre o crescimento econômico levou a uma redução proporcional na disparidade de renda entre os mais pobres e os mais ricos



Fontes: Banco Mundial, FMI, IBGE Gráficos : Alberto Cairo

http://www.visualisingdata.com/2011/08/data-visualisation-stories-from-brazil-by-alberto-cairo/



http://ourworldindata.org/data/growth-and-distribution-of-prosperity/world-poverty/

Hans Rosling (1948–2017)

- 200 Countries, 200 Years, 4 Minutes
  - The Joy of Stats – BBC 4
    - https://www.youtube.com/watch?v=jbkSRLYSojo

References

- The picture superiority effect in associative recognition.
  - https://www.ncbi.nlm.nih.gov/pubmed/18927048
- John W. Tuckey, 1977. Exploratory Data Analysis, Pearson.
- Edward R. Tufte, 1983. The Visual Display of Quantitative Information. Graphics Press.
- William S. Cleveland, 1994, The Elements of Graphing Data, Hobart Press
- S.K.Card, J.D.Mackinlay, and B.Shneiderman. Readings in Information Visualization: Using Vision to Think. Academic Press, 1999