

# Data Visualization

---

## Visualizzazione dell'Informazione Quantitativa

<http://softeng.polito.it/courses/VIQ>



**SoftEng**  
<http://softeng.polito.it>

Version 1.2.2  
© Marco Torchiano, 2020






This work is licensed under the Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International License.

To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-nd/4.0/>.

You are free: to copy, distribute, display, and perform the work

Under the following conditions:

-  **Attribution.** You must attribute the work in the manner specified by the author or licensor.
-  **Non-commercial.** You may not use this work for commercial purposes.
-  **No Derivative Works.** You may not alter, transform, or build upon this work.
  - For any reuse or distribution, you must make clear to others the license terms of this work.
  - Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

# Definition

---

Encode data through visual features in order to convey useful information

# Motivation

---

## Information retrieval

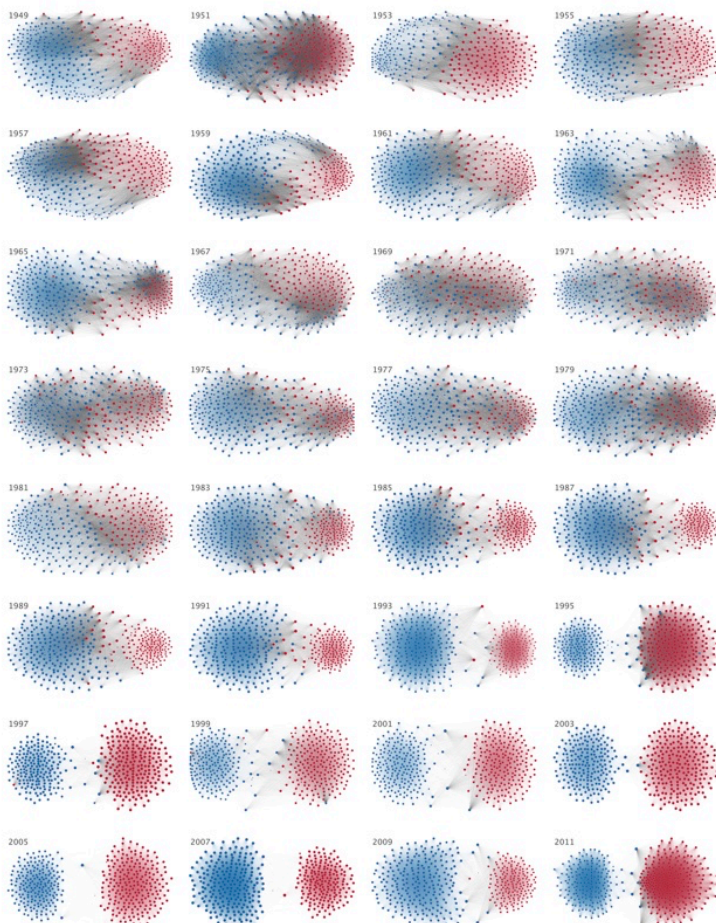
- After 3 days
  - ◆ Text alone: 10%
  - ◆ Text + visuals: 65%

# 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



# 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)]

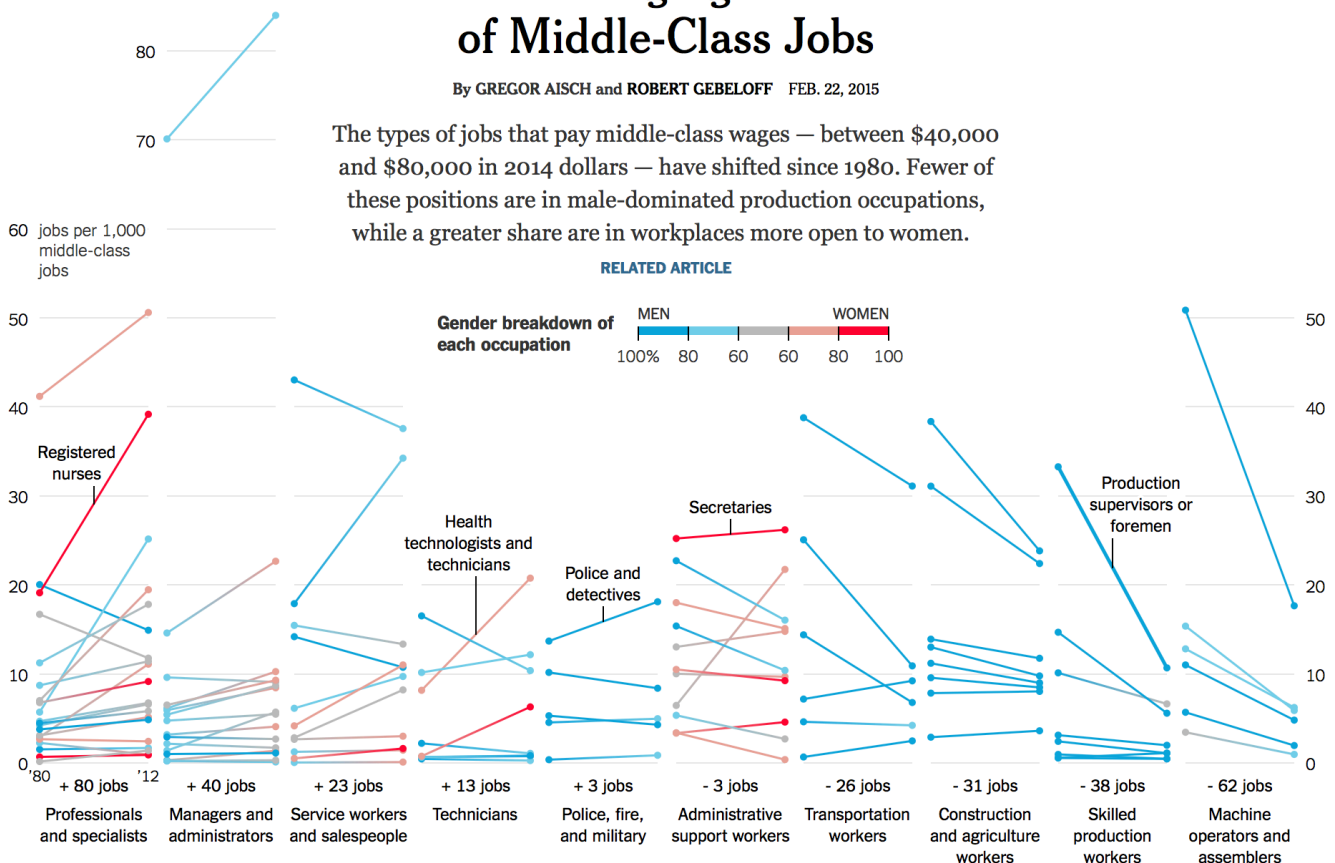
## The Changing Nature of Middle-Class Jobs

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

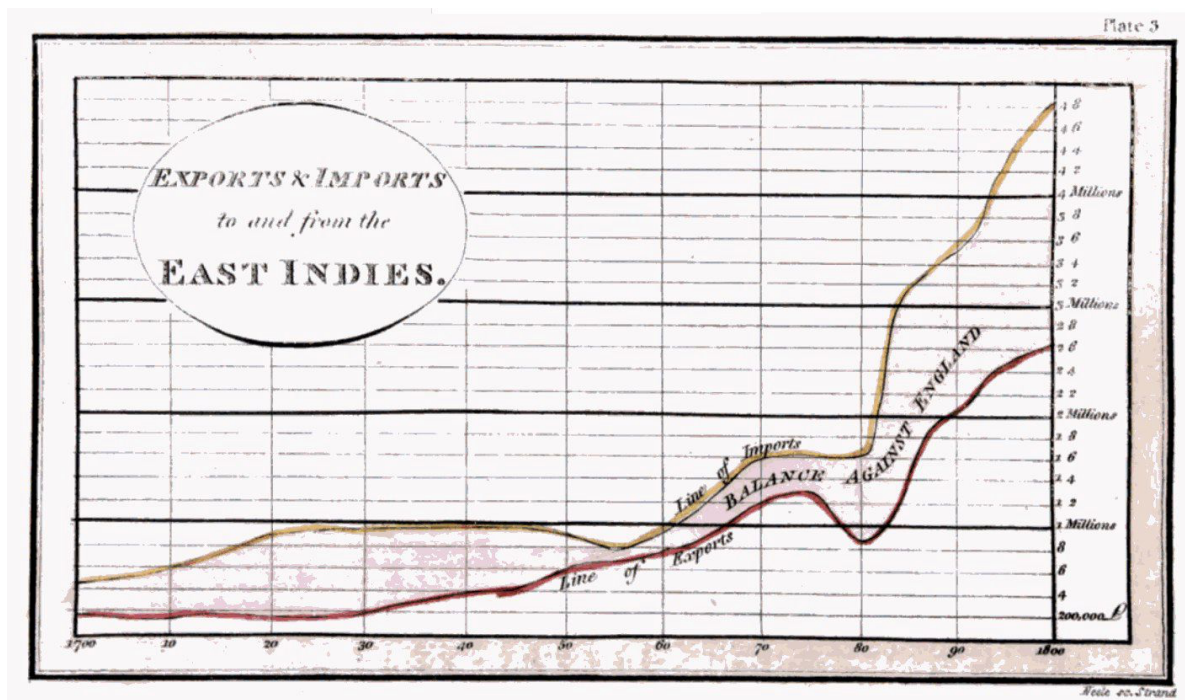
Gender breakdown of each occupation  
MEN 100% 80 60 60 80 100 WOMEN





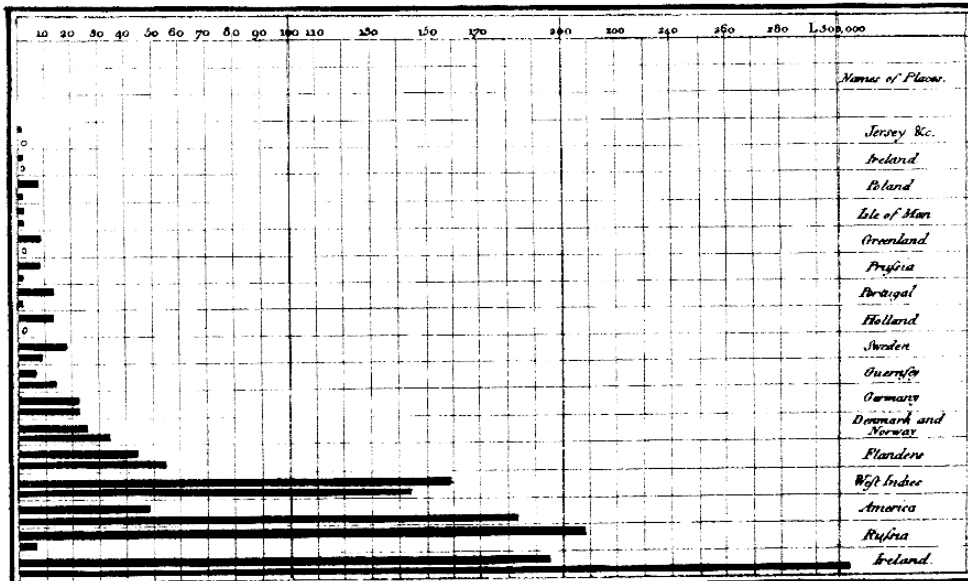
# HISTORY

## William Playfair



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

Exports and imports of SCOTLAND to and from different parts for one Year from Christmas 1780 to Christmas 1781.



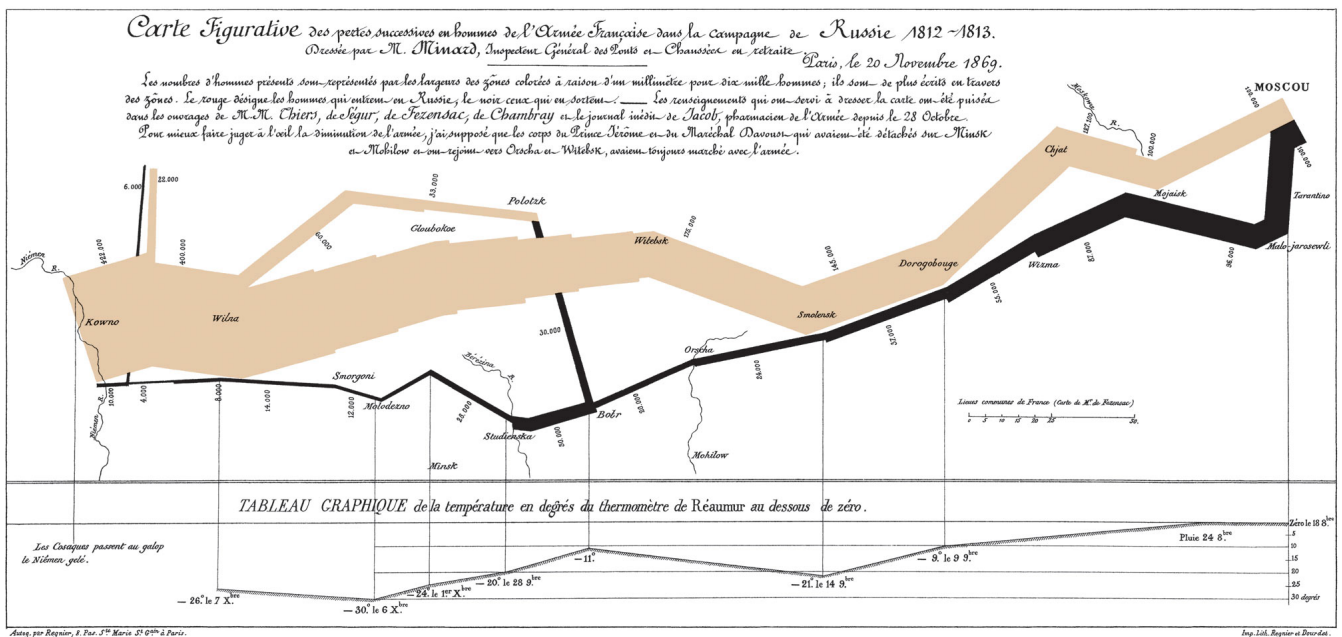
The upright divisions are Ten Thousand Pounds each. The Black Lines are Exports the Red Lines Imports.

Published as directed March 7<sup>th</sup> 1782 by W. Playfair

Made comp<sup>t</sup> 1782, James, London.

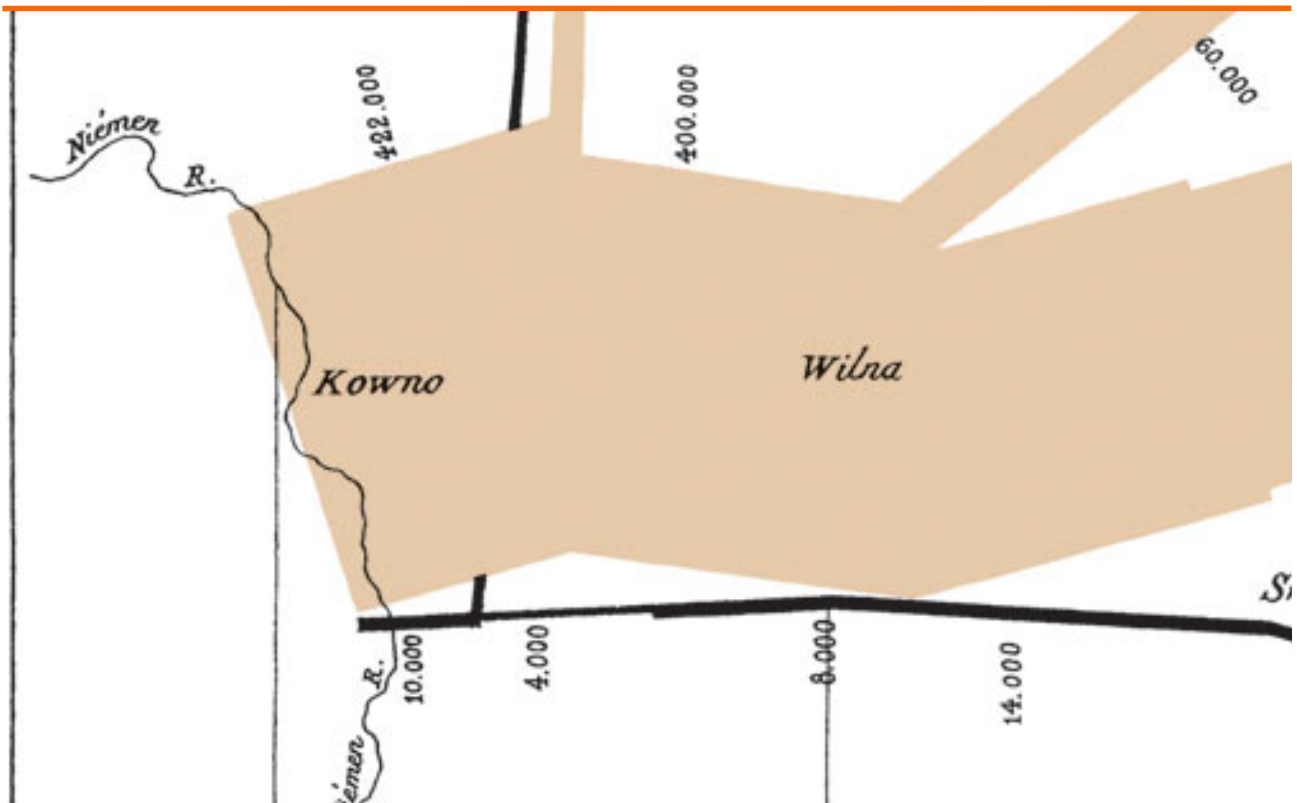
W. Playfair, The Commercial and Political Atlas, London 1786

# Charles Joseph Minard

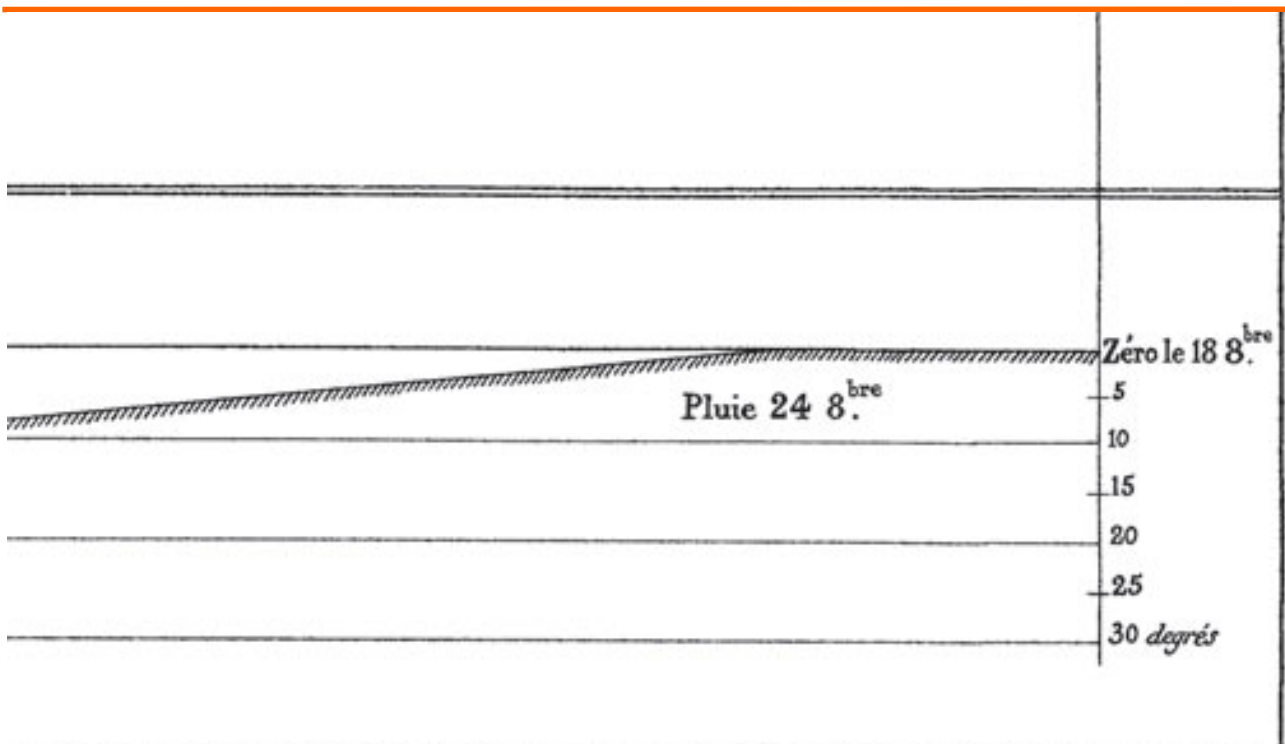


[https://en.wikipedia.org/wiki/Charles\\_Joseph\\_Minard](https://en.wikipedia.org/wiki/Charles_Joseph_Minard)

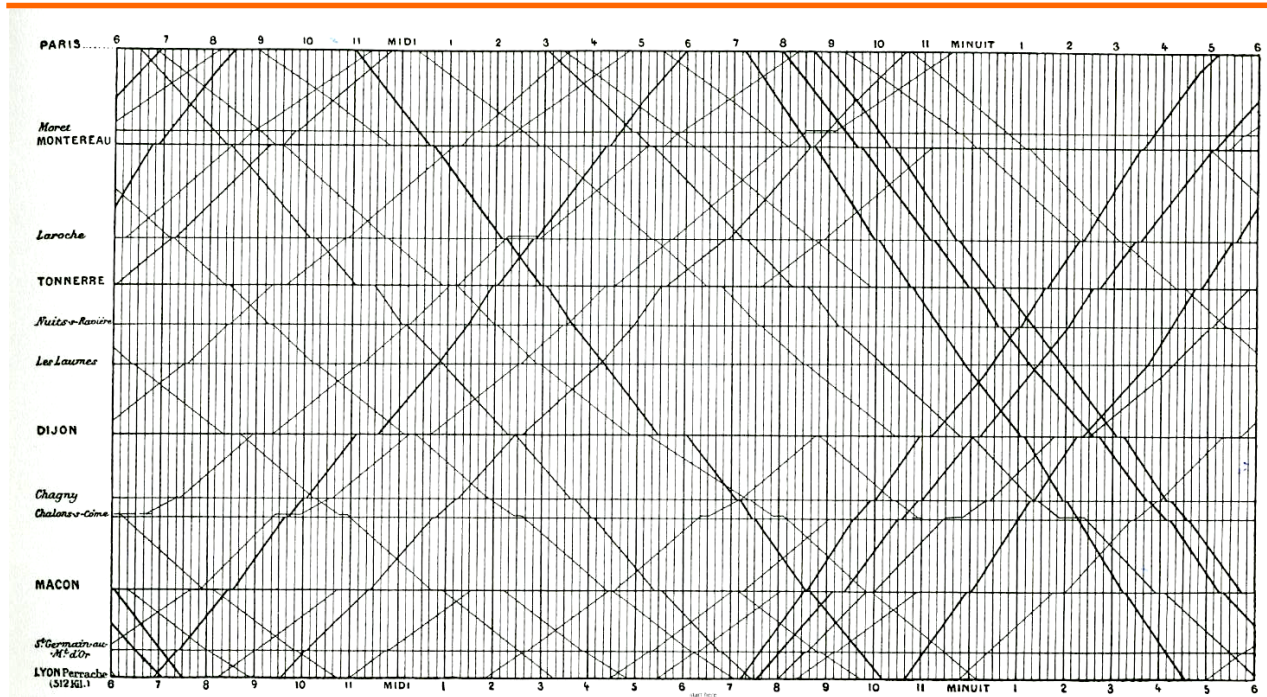
# Numbers and direction



# Temperature



# Étienne-Jules Marey



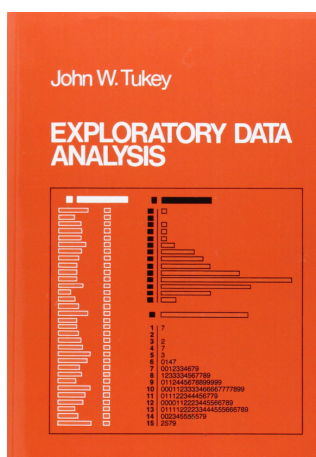
La Méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine, 1885

<https://archive.org/details/lamthodegraphiq00maregoog>

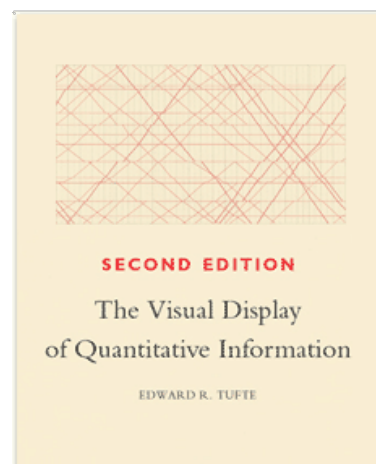
## XX Century

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

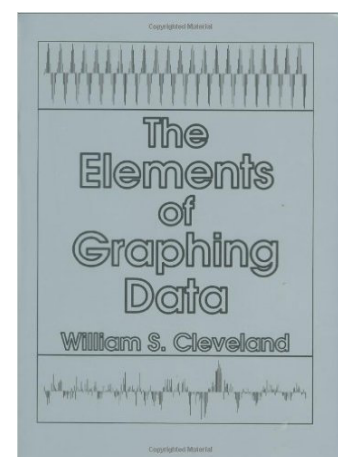
1977



1983



1985



---

# INFORMATION VISUALIZATION

## Information visualization

---

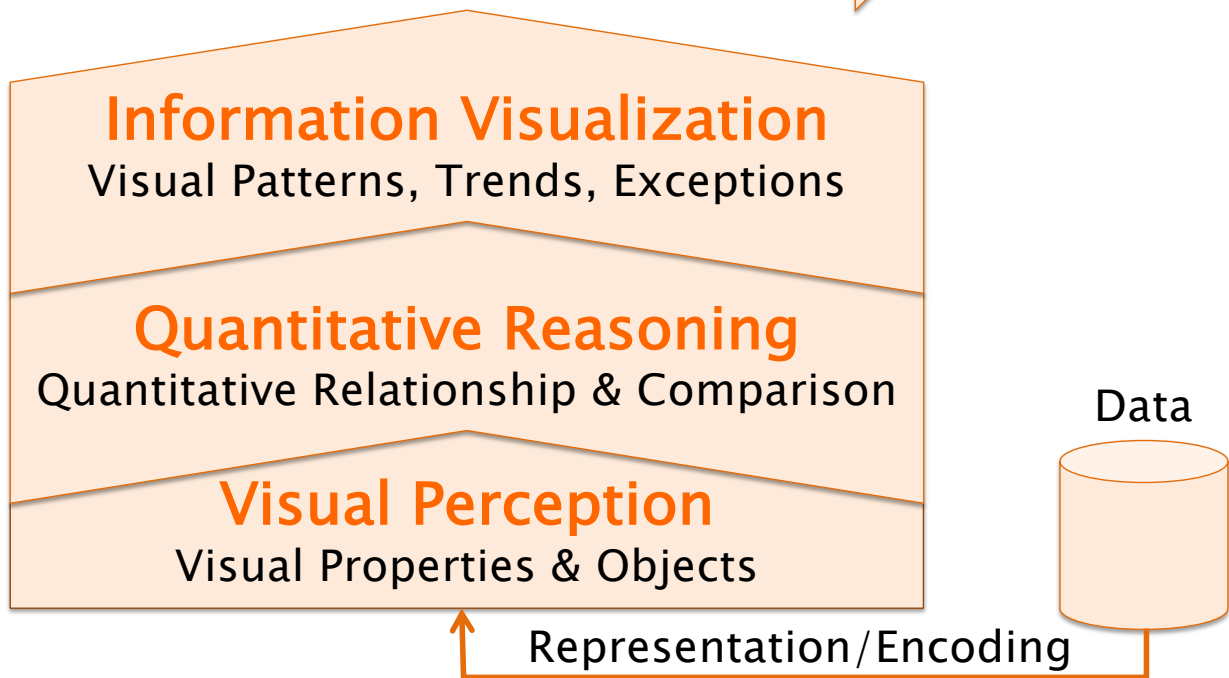
The use of computer-supported, interactive, visual representations of abstract data to amplify cognition



# Overview

---

Understanding  Decisions



**SoftEng**  
http://softeng.polito.it

## Quantitative message

---

- Quantitative values
  - ◆ Express measures

+

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

**SoftEng**  
http://softeng.polito.it

# Understanding tasks

---

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

# 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

Look up individual values

Compare individual values

Precise values are required

There is more than one unit of measure

### Use graphs to

Focus on the shape of values

Reveal relationships among multiple values

---

# EXAMPLES

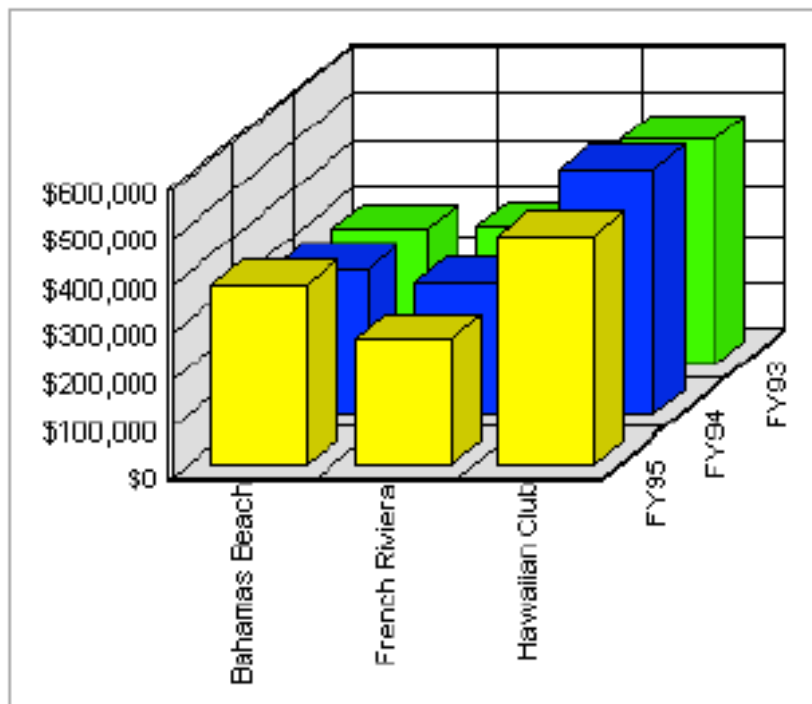
---

## 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.

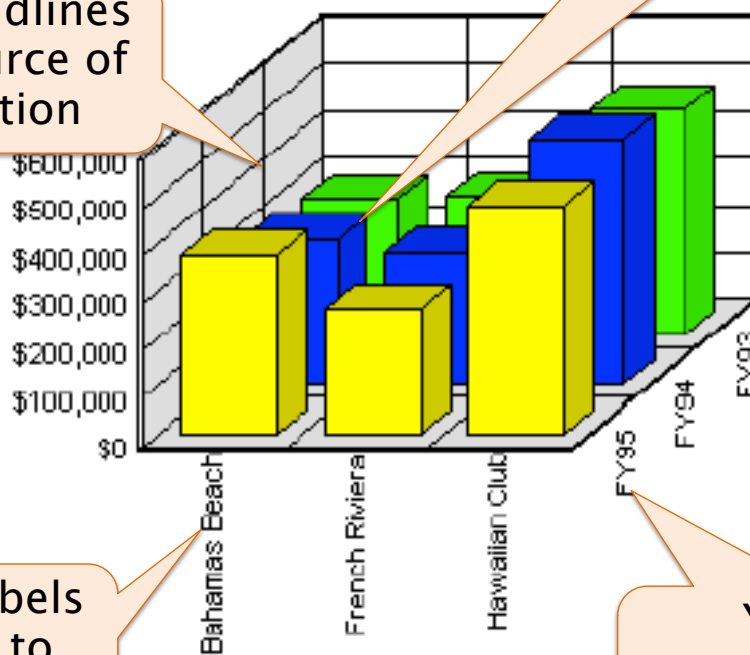
# A bar graph



## A **bad** bar graph

Heavy gridlines are a source of distraction

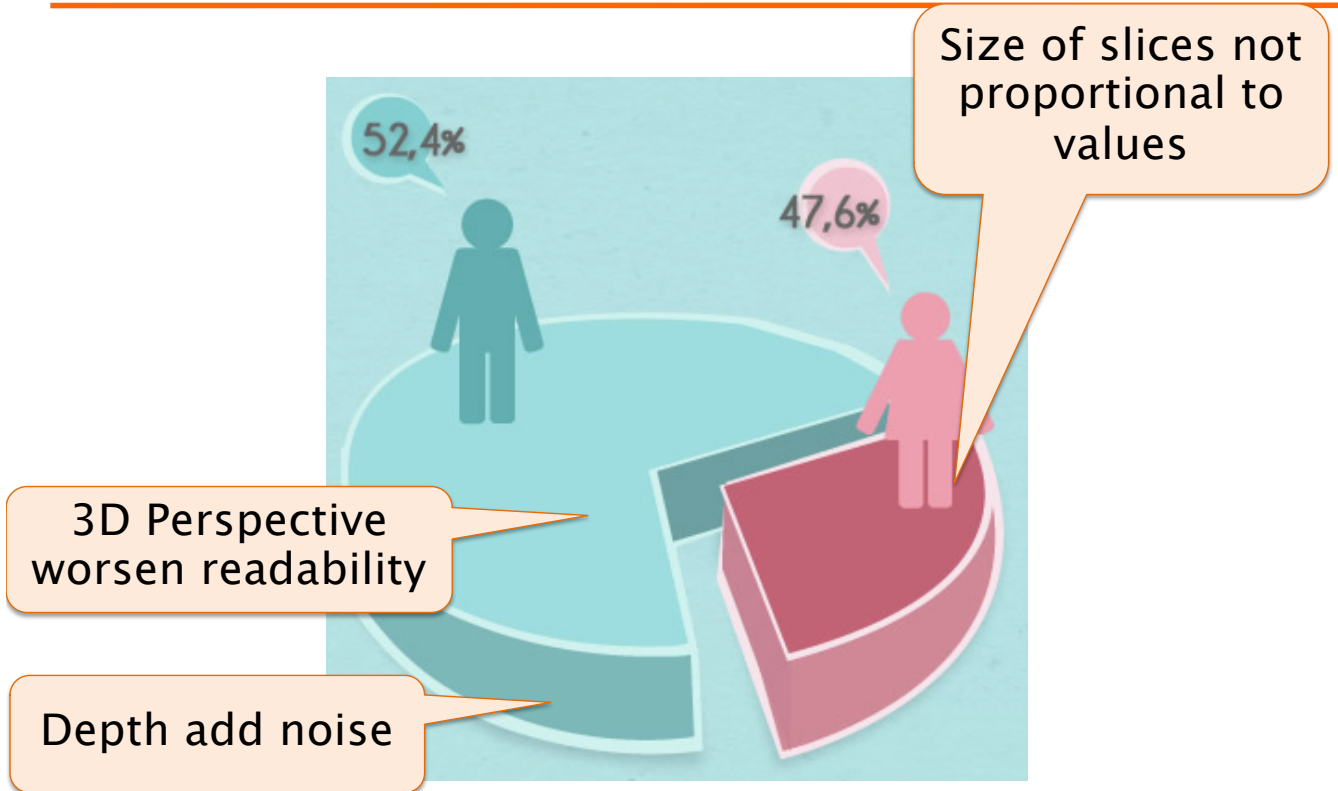
3D bars are impossible to read



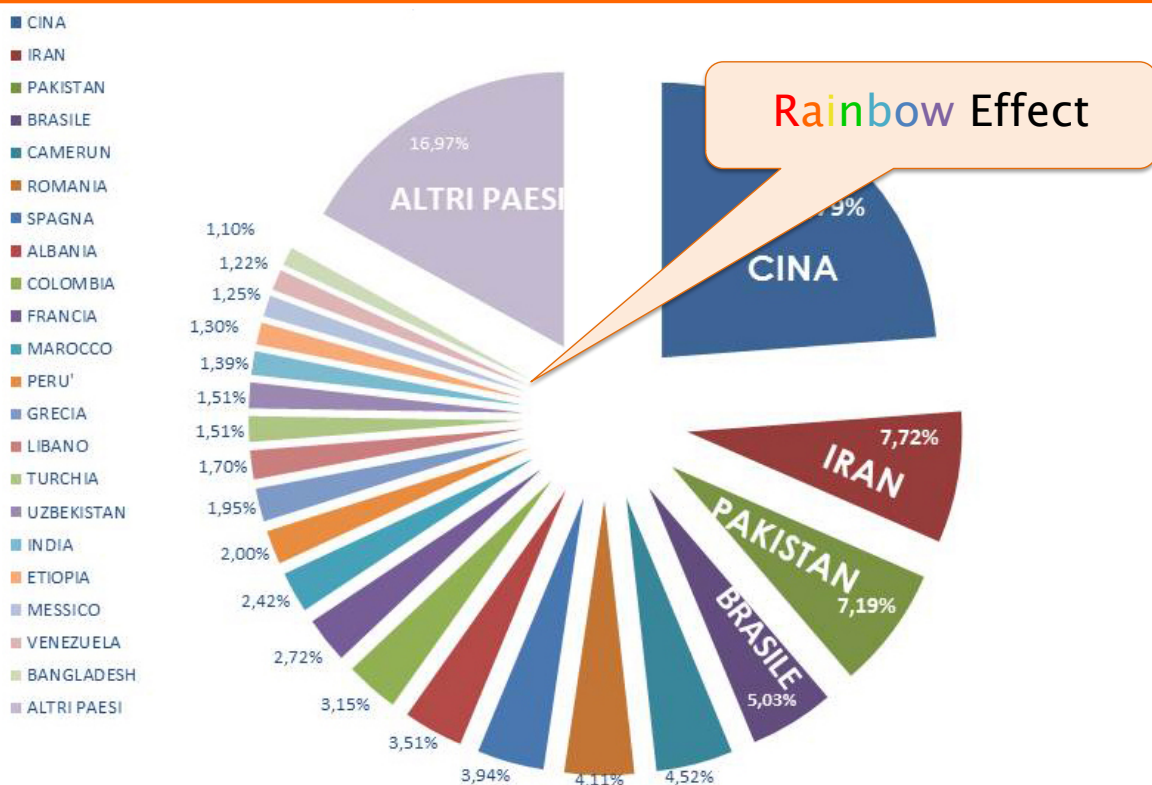
Vertical labels are hard to read

Years run counterintuitively from back to front

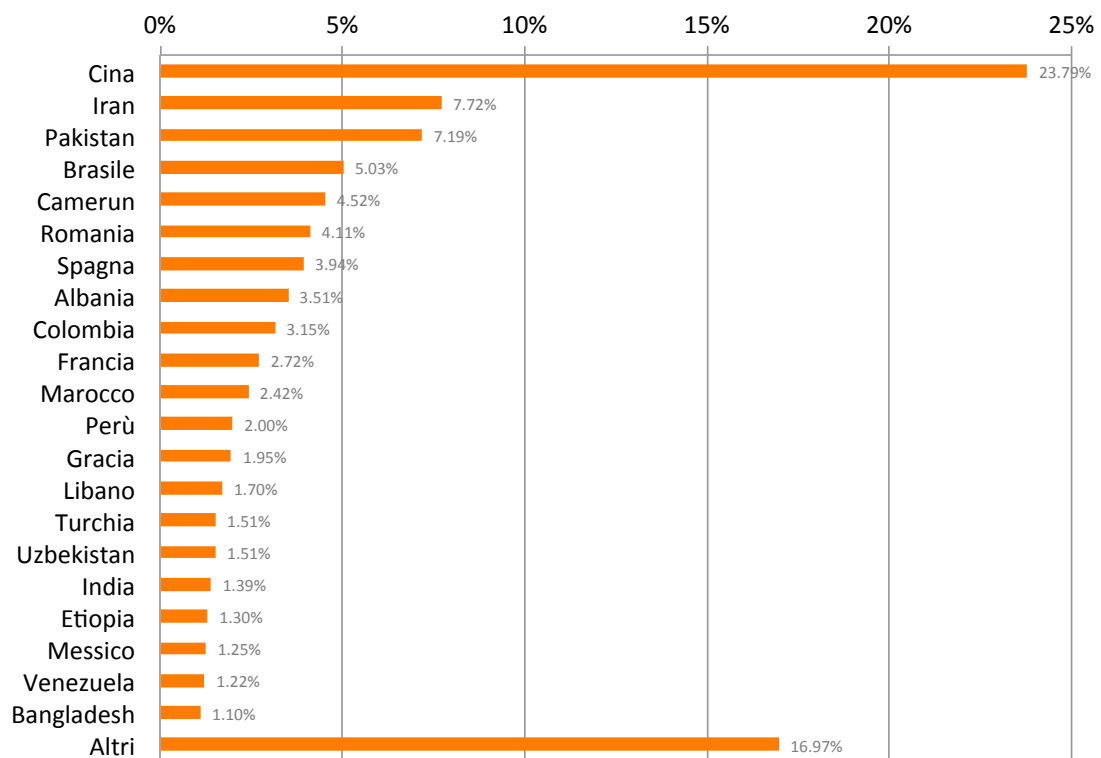
# A pie chart



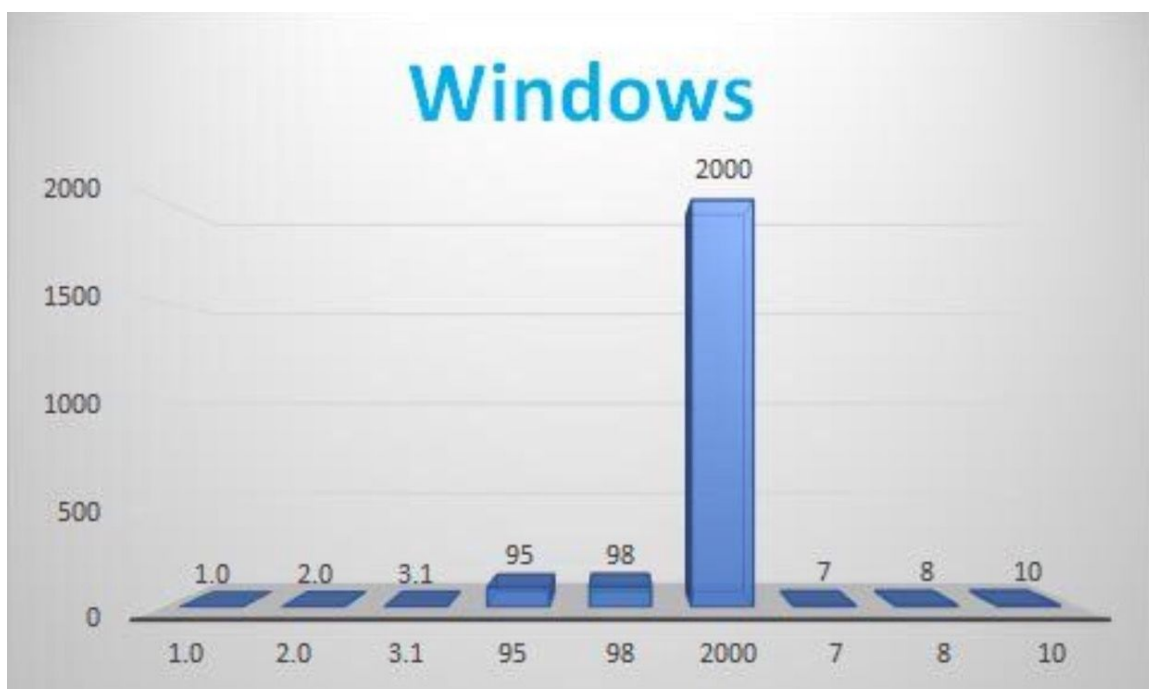
# Pie chart (original)



# Bar chart (redesign)



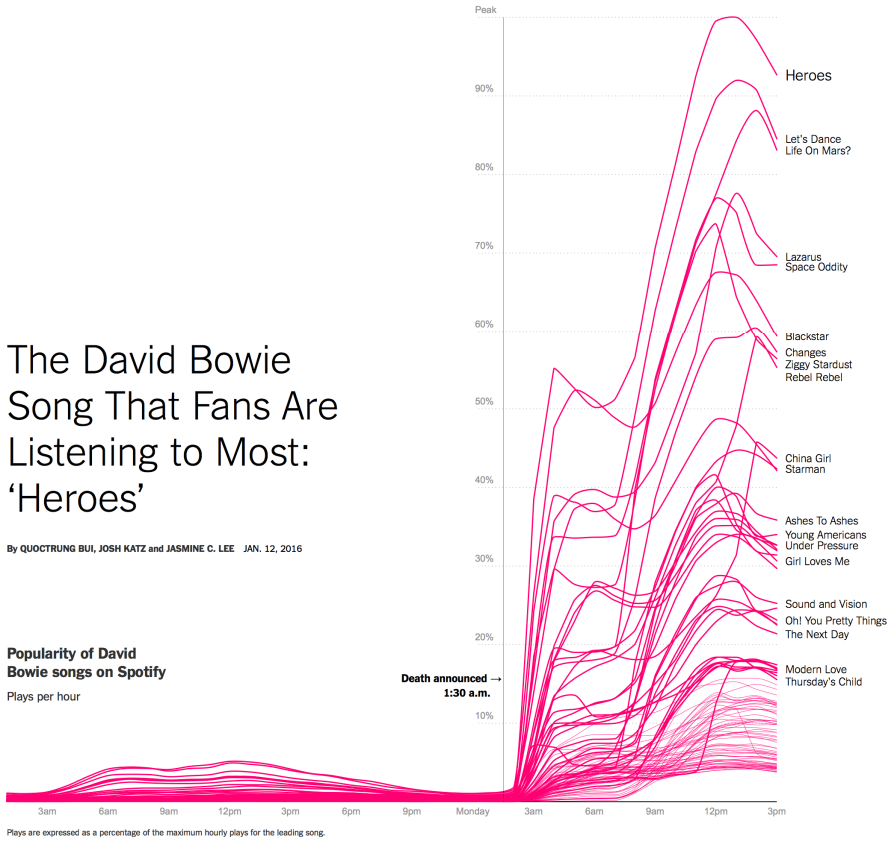
# Meaningless Data



# The David Bowie Song That Fans Are Listening to Most: 'Heroes'

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

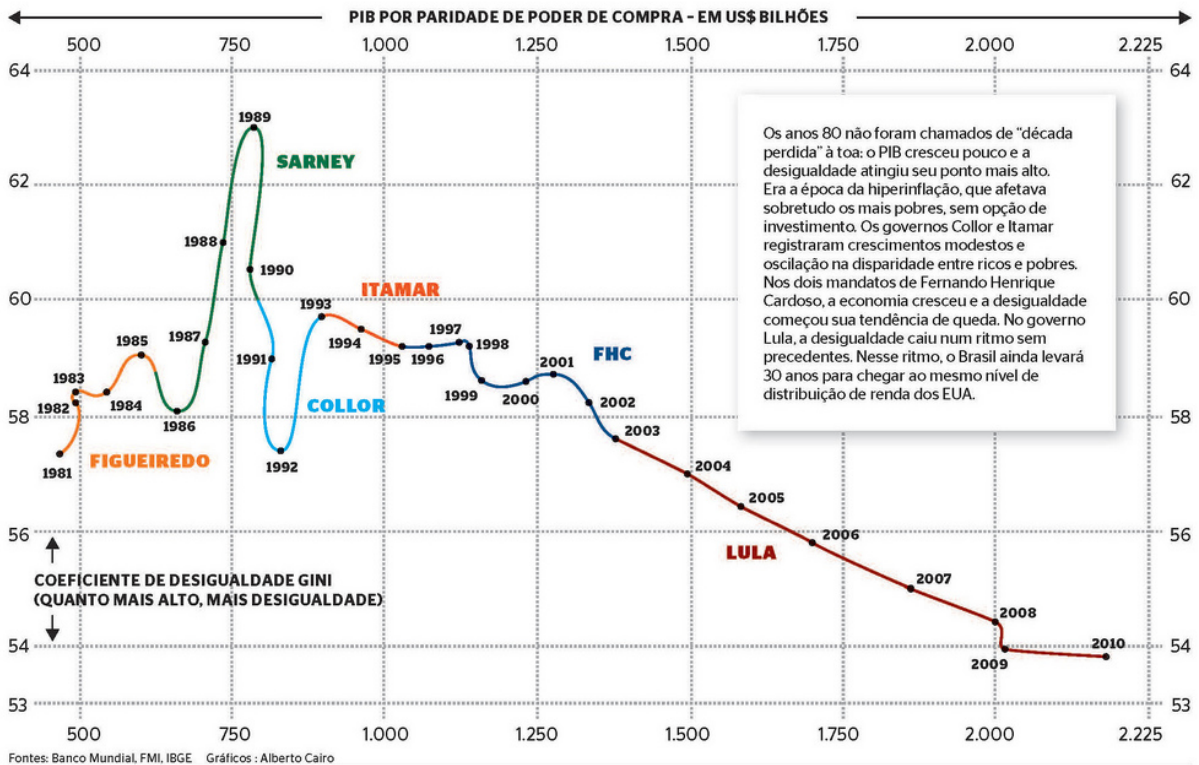
Popularity of David Bowie songs on Spotify  
Plays per hour



<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

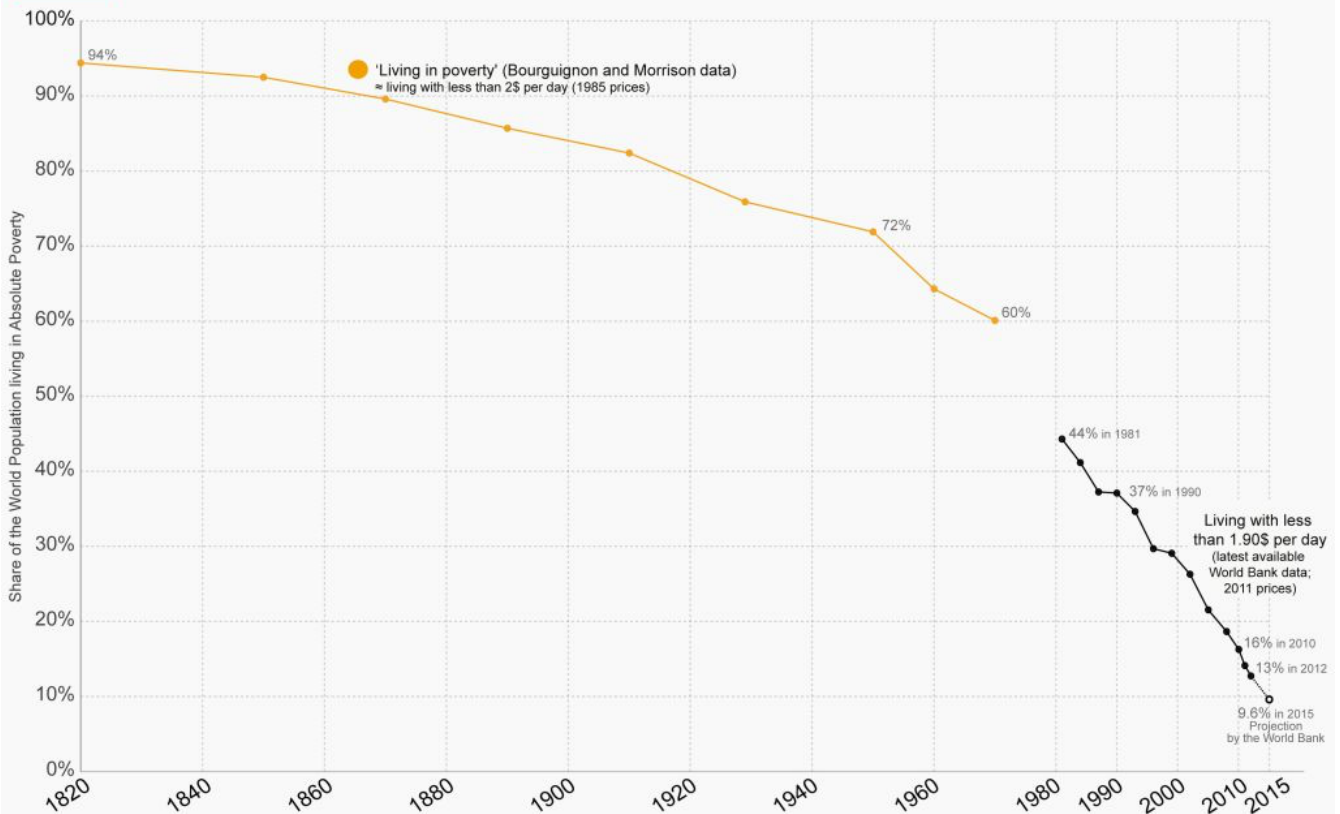
O gráfico abaixo mostra o avanço do PIB comparado à evolução da desigualdade no Brasil desde 1980. Nem sempre o crescimento econômico levou a uma redução proporcional na disparidade de renda entre os mais pobres e os mais ricos



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

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





Data sources: 1820-1970 Bourguignon and Morrison (2002) - Inequality among World Citizens, In The American Economic Review; 1981-2015 World Bank (PovcalNet)

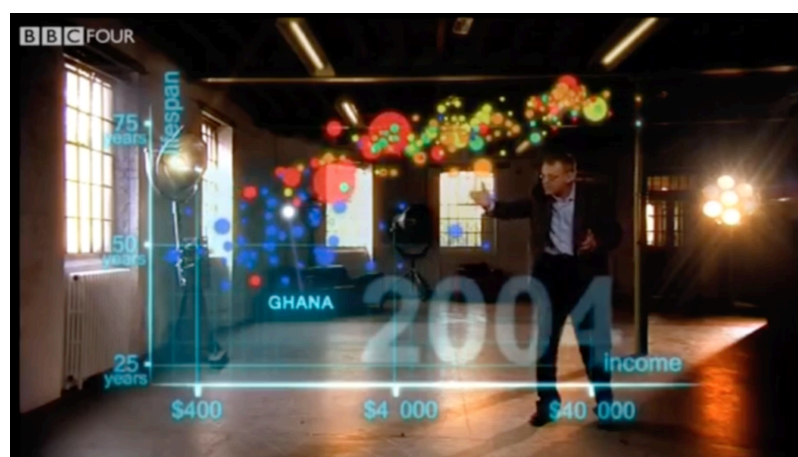
The interactive data visualisation is available at OurWorldInData.org. There you find the raw data and more visualisations on this topic.

Licensed under CC-BY-SA by the author Max Roser.

<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