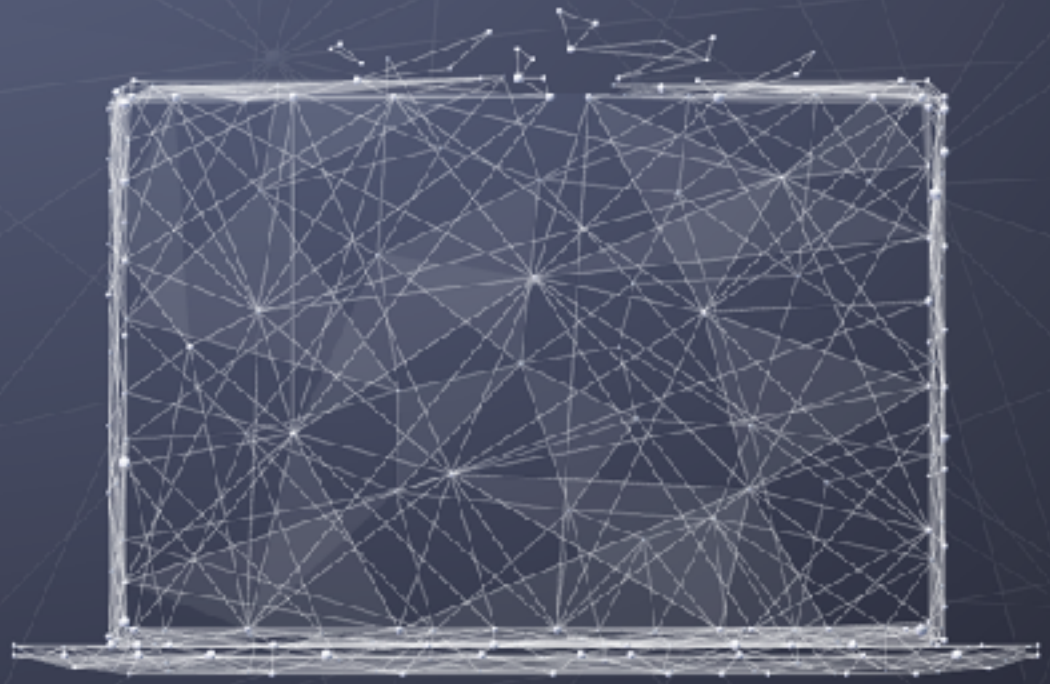
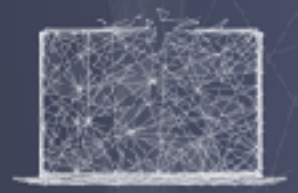
Data Science
Data Engineering I
Introduction to visualization

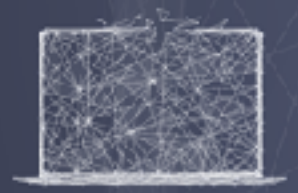


PURDUE UNIVERSITY | College of Science



Introduction to visualization

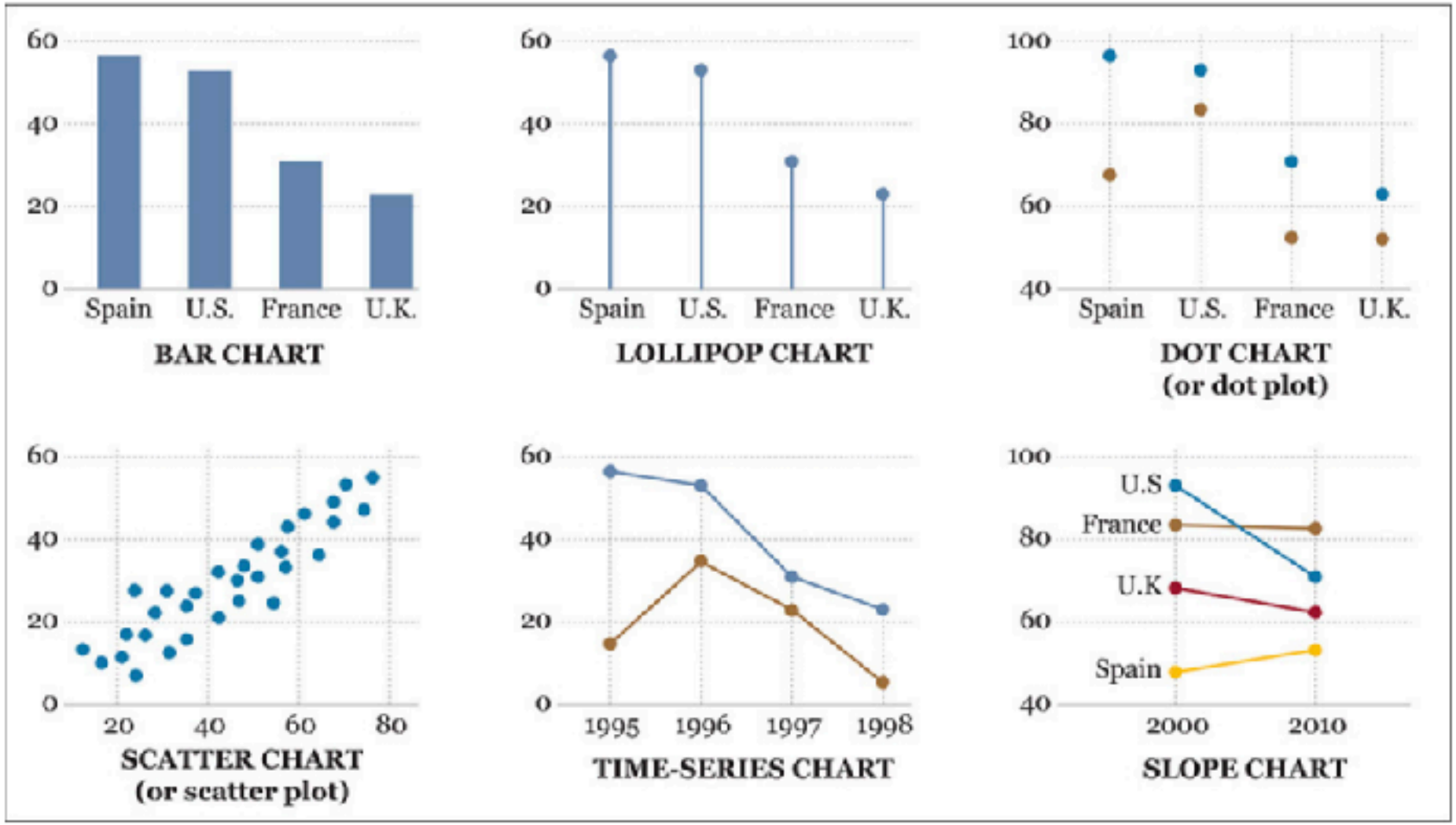
- Visualization: any kind of visual representation of information designed to enable communication, analysis, discovery, exploration, etc.
- Good data visualizations communicate information by:
- Providing reliable information
- Visually encoding it so relevant patterns become noticeable
- Organizing it in a way that enables some exploration
- Presenting it in an attractive manner



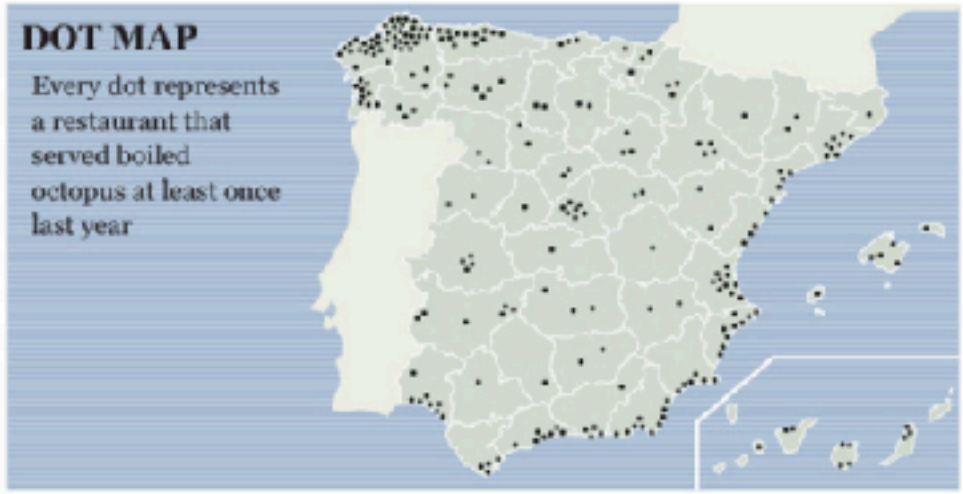
Types of visualizations

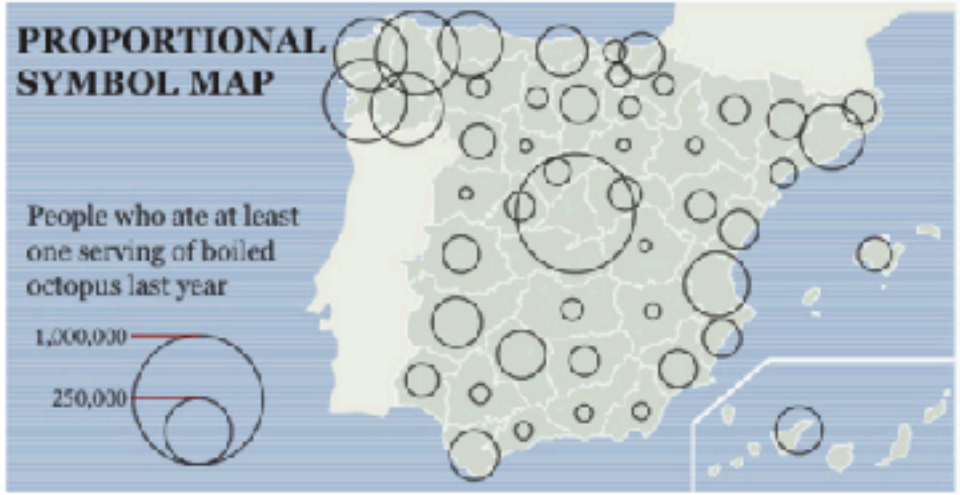
- Chart: display in which data are encoded with symbols that have different shapes, colors, or proportions
- Map: depiction of a geographical area or a representation of data that pertains to that area
- Infographic: multi-section visual representation of information intended to communicate one or more specific messages

Examples of charts

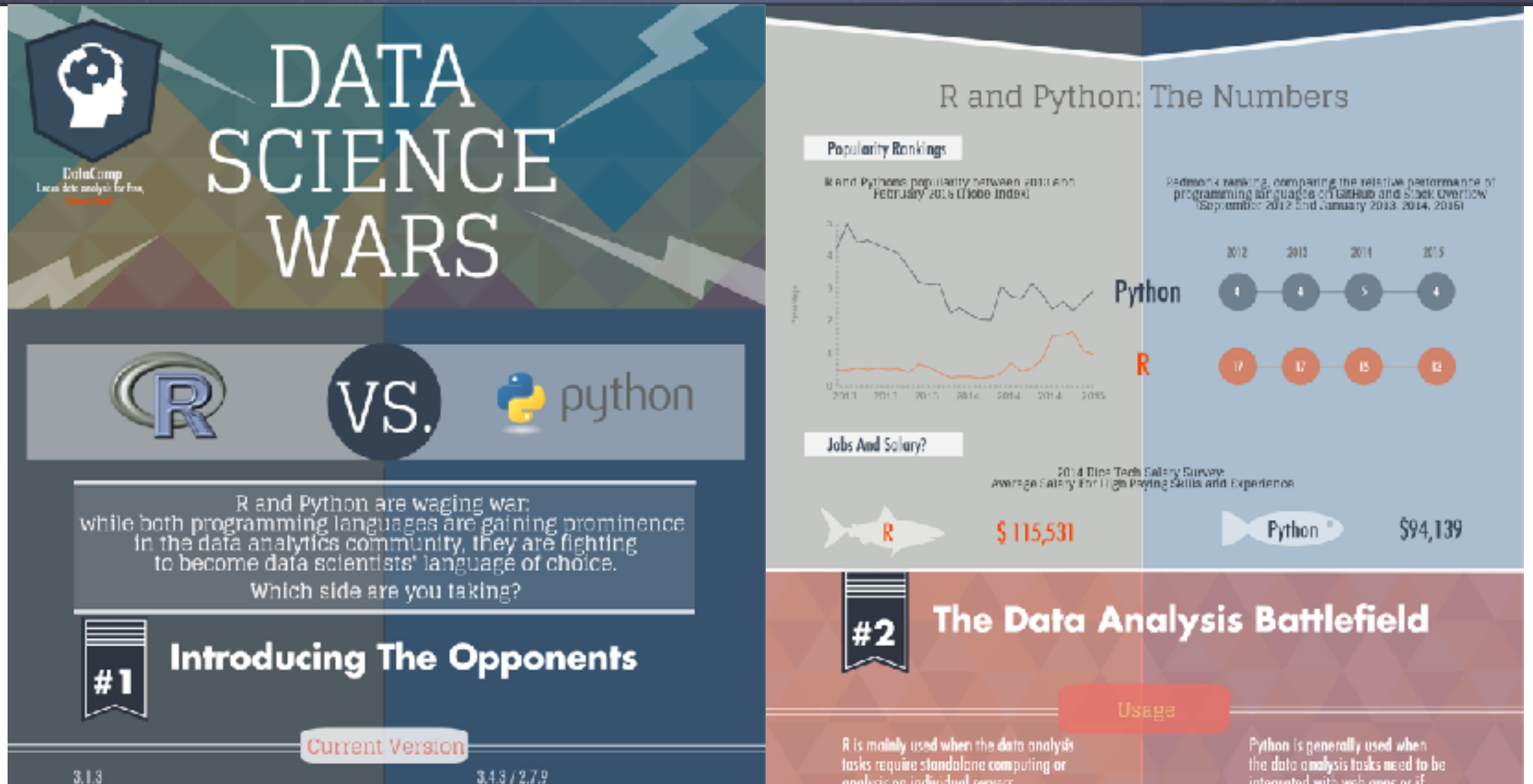


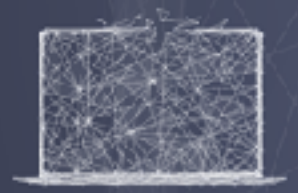
Examples of data maps





Example infographic





Qualities of great visualizations

- Truthful: based on thorough and honest research
- Functional: accurate description of data that lets people do meaningful operations based on it
- Beautiful: intriguing and aesthetically pleasing (for intended audience)
- Insightful: reveals evidence that we would have a hard time seeing otherwise

Example: Visualization success story

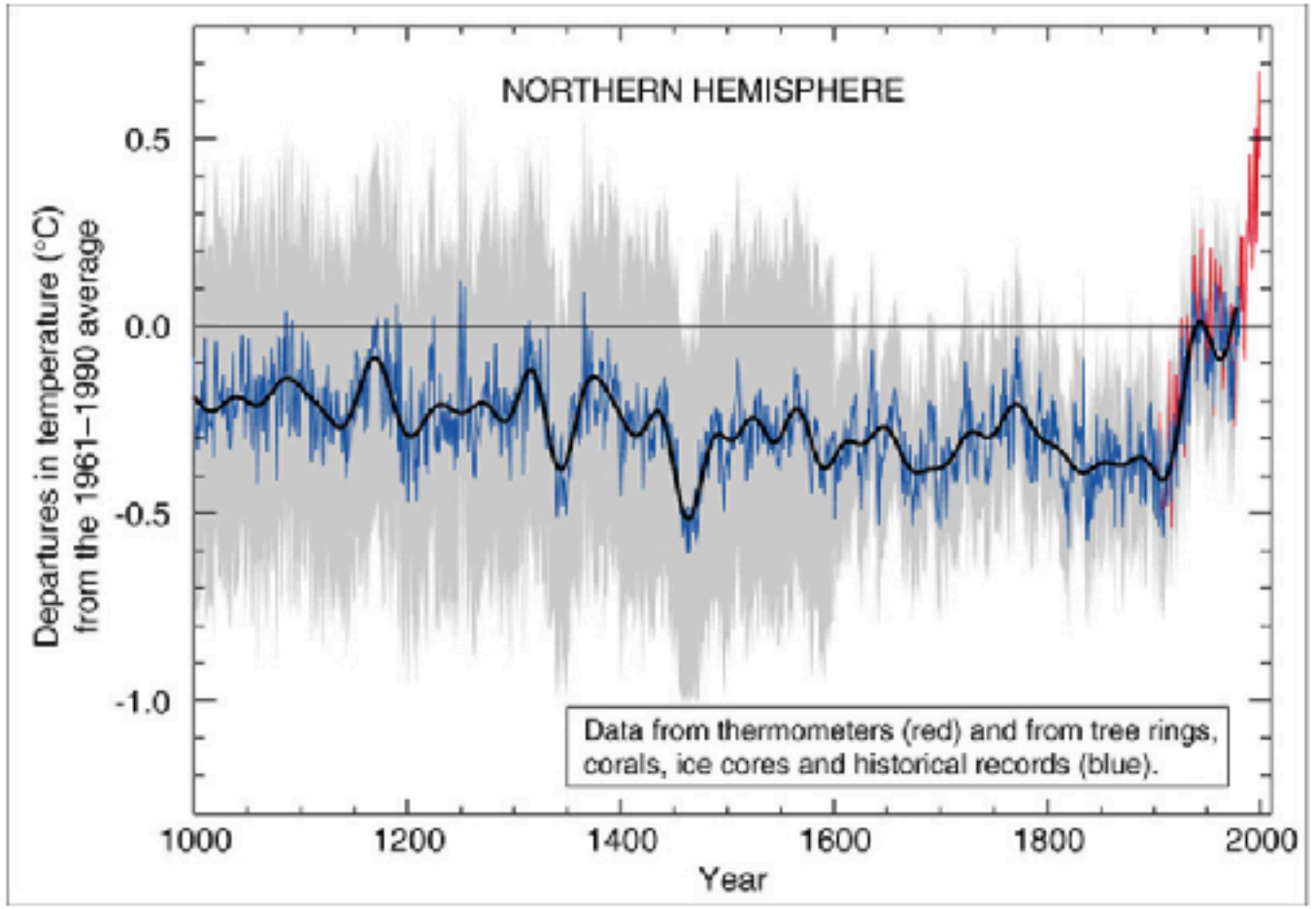
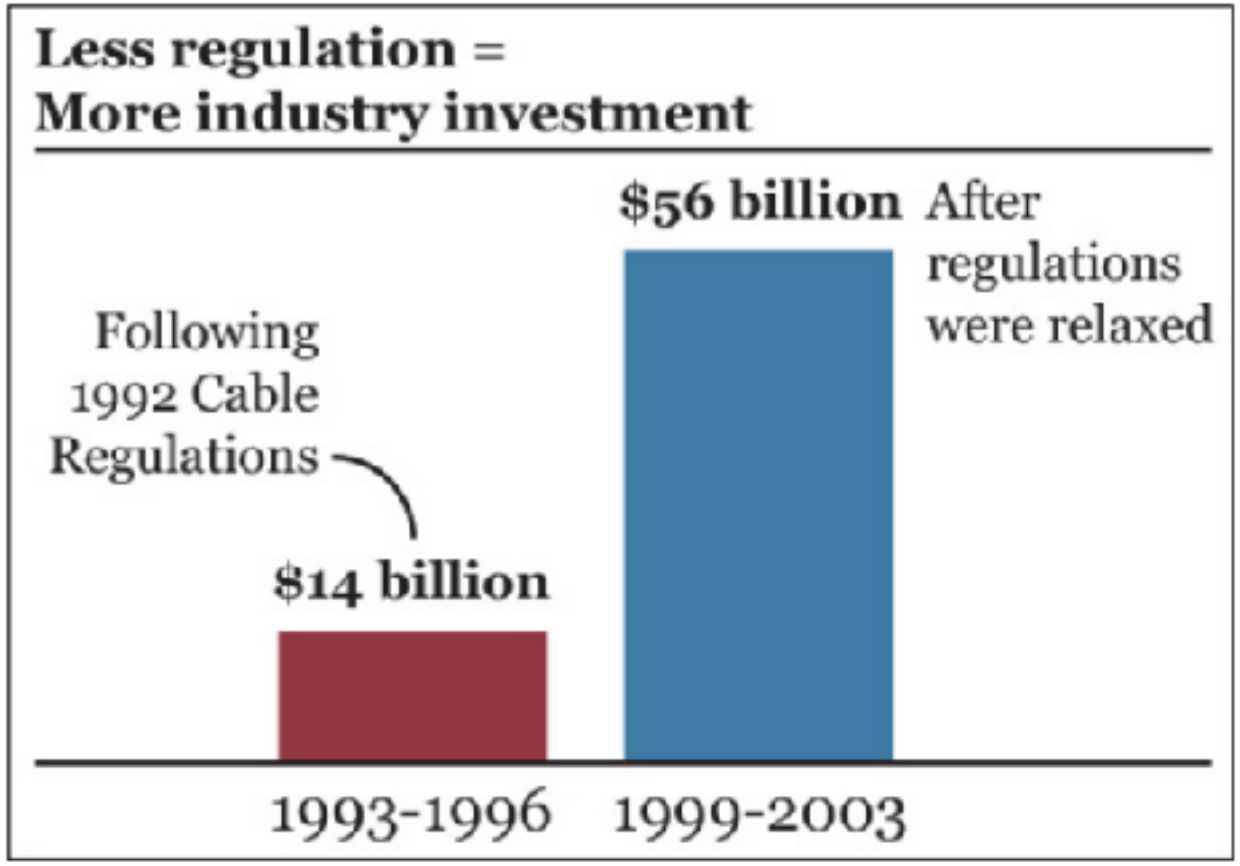


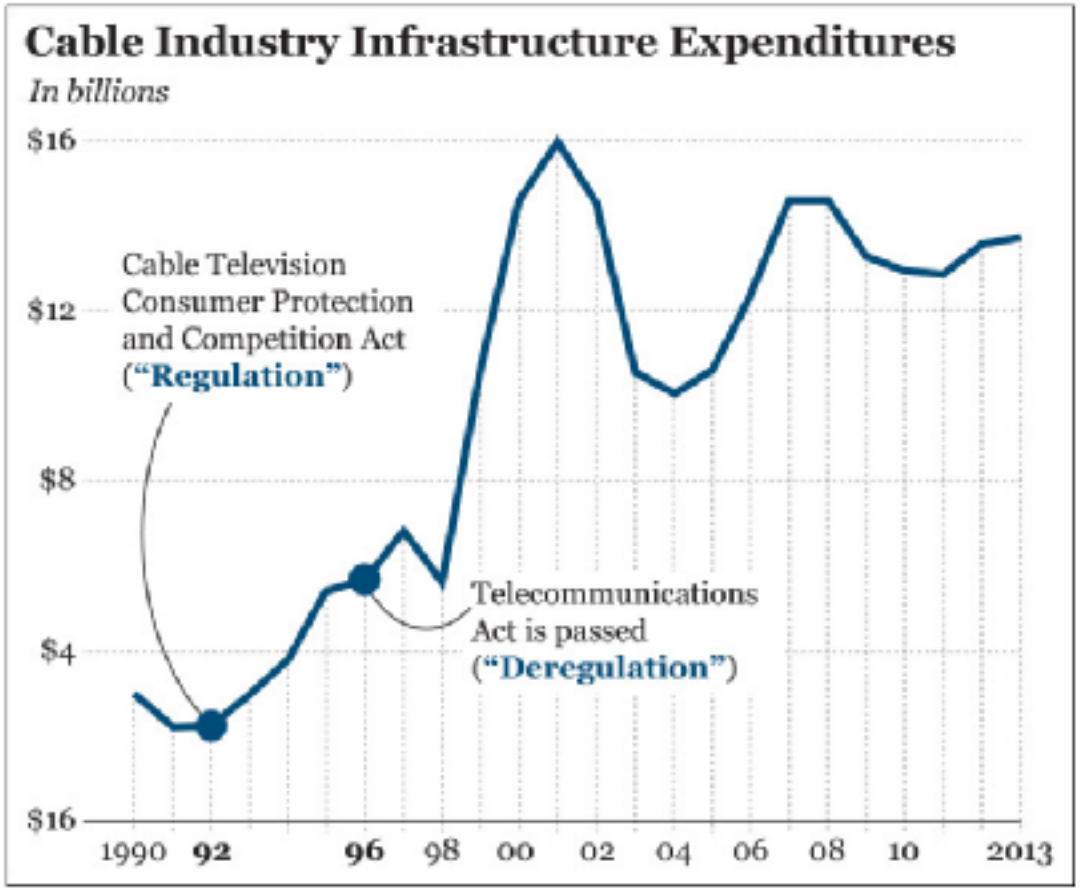
Figure 2.1 The hockey stick chart. Summary For Policymakers of the 2001 Third Assessment Report of the Intergovernmental Panel on Climate Change.

Truthful: What is wrong with this chart?



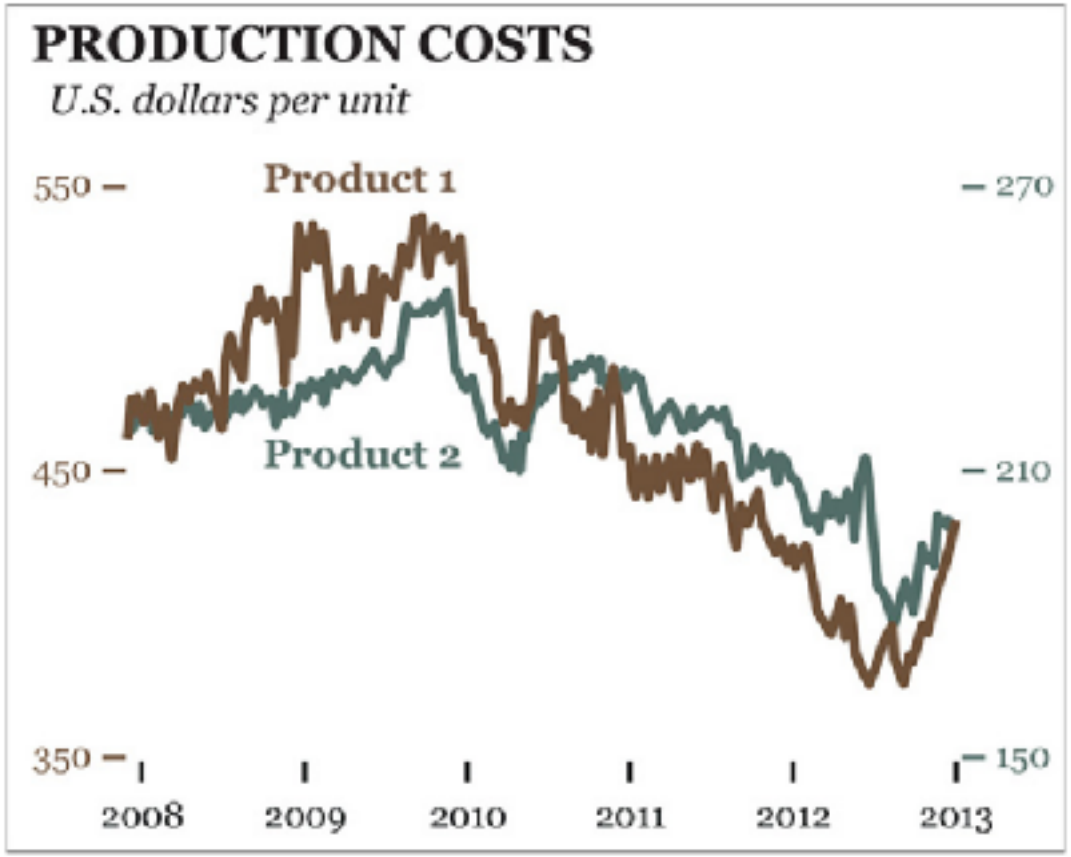
Start to read visualizations, don't just look at them

Conclusion changes with more information



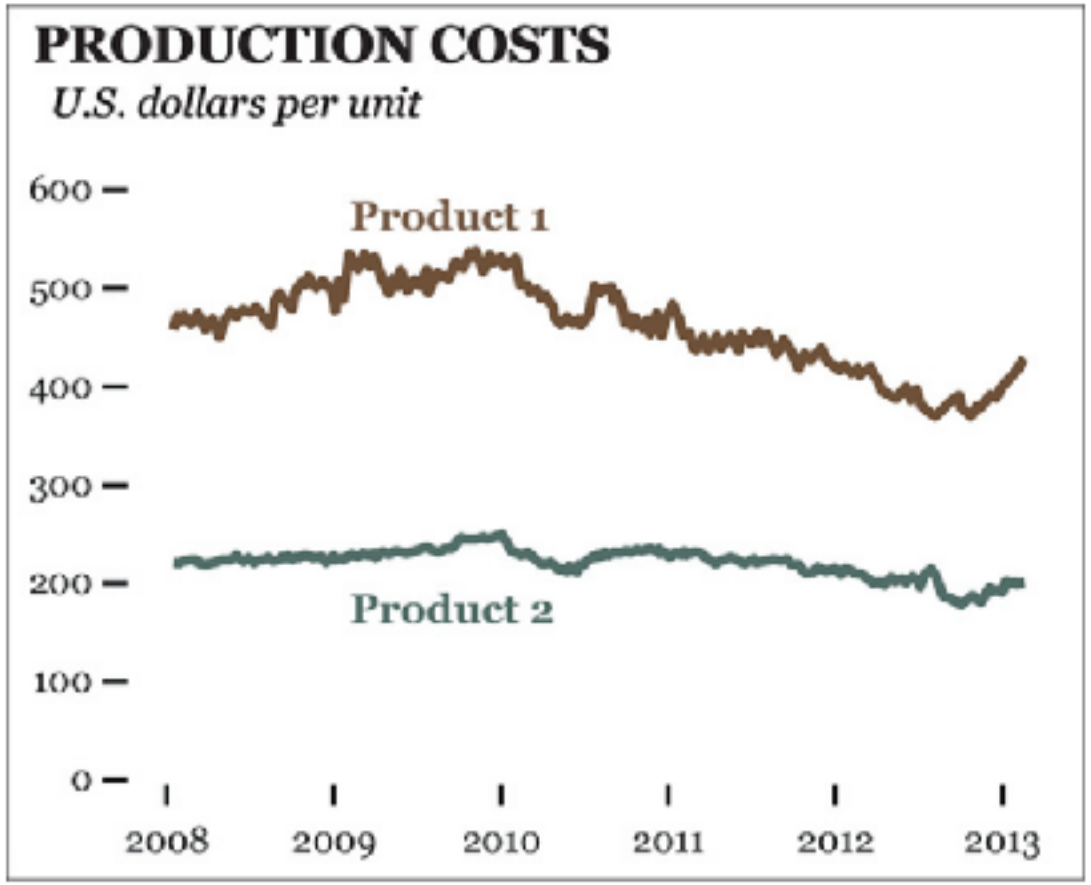
If someone hides data from you, there's probably a reason

But truth is relative…

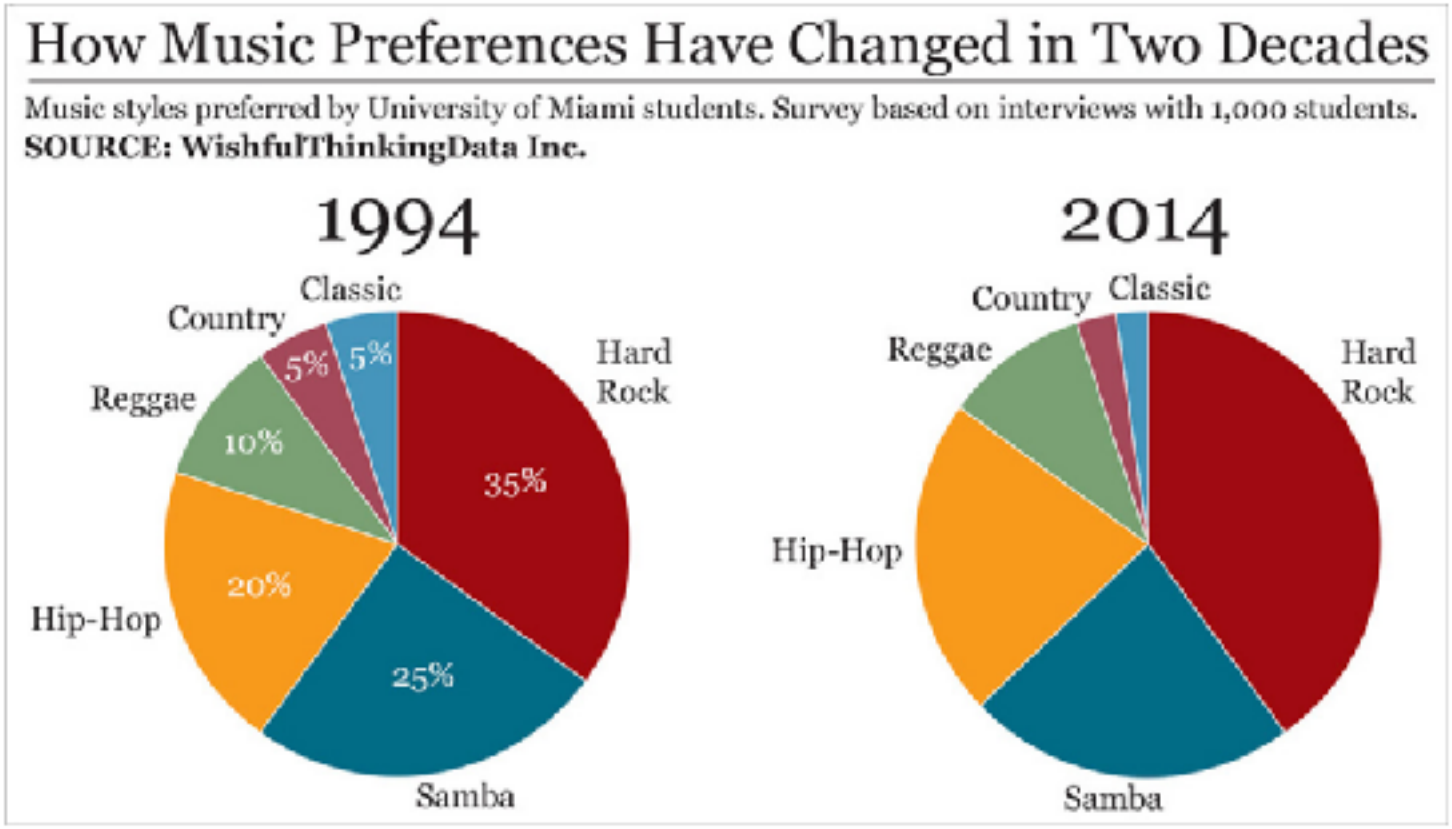


…and interacts with other design choices

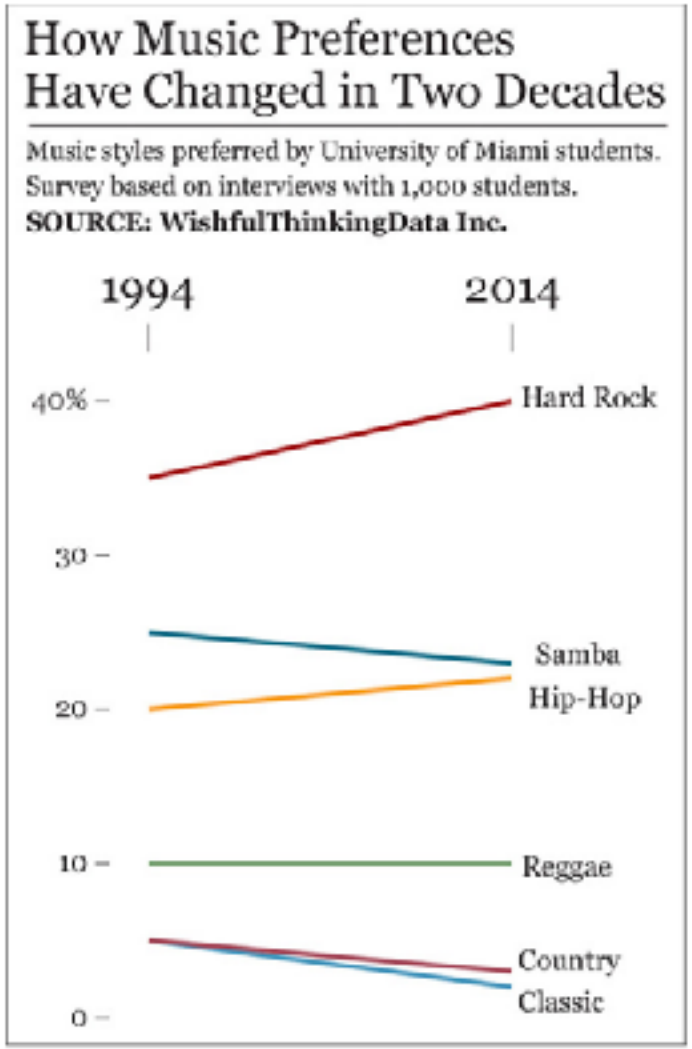
Same data, now on same scale



Functional: Purpose should guide design decisions



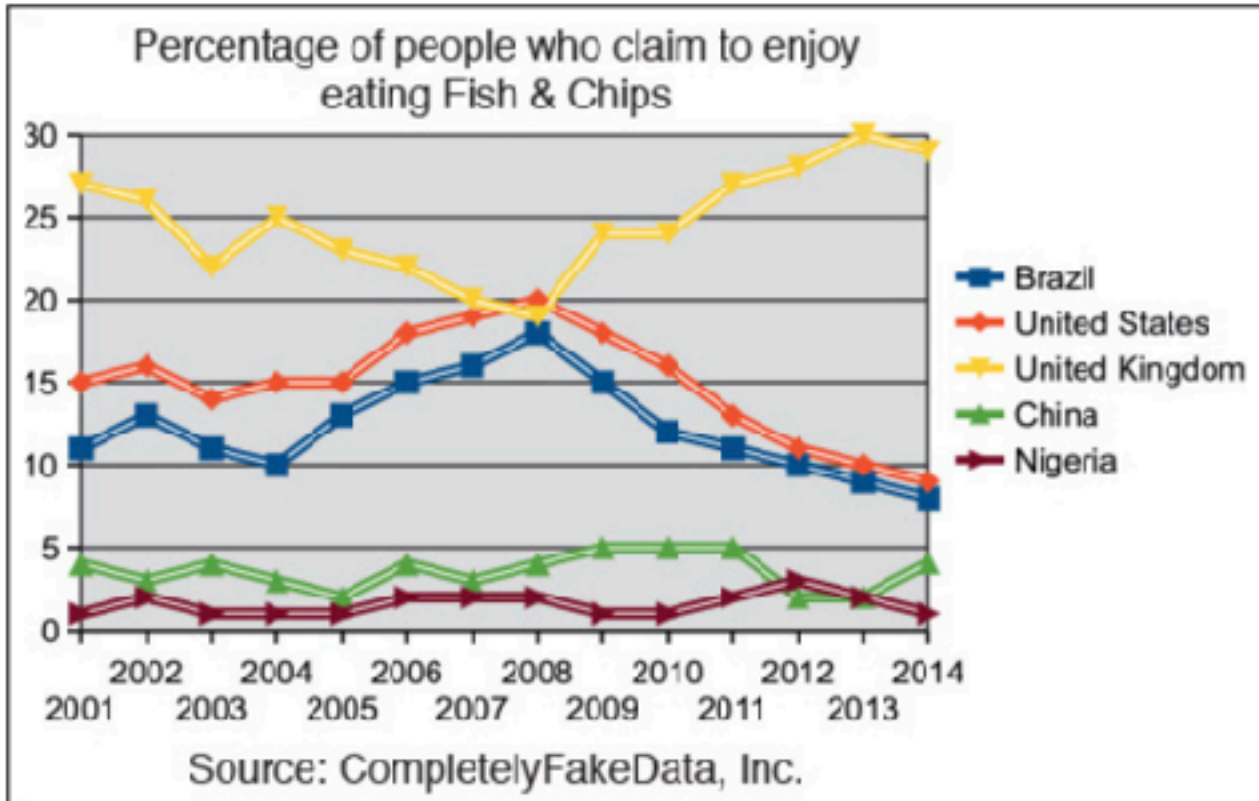
Functional: Purpose should guide design decisions

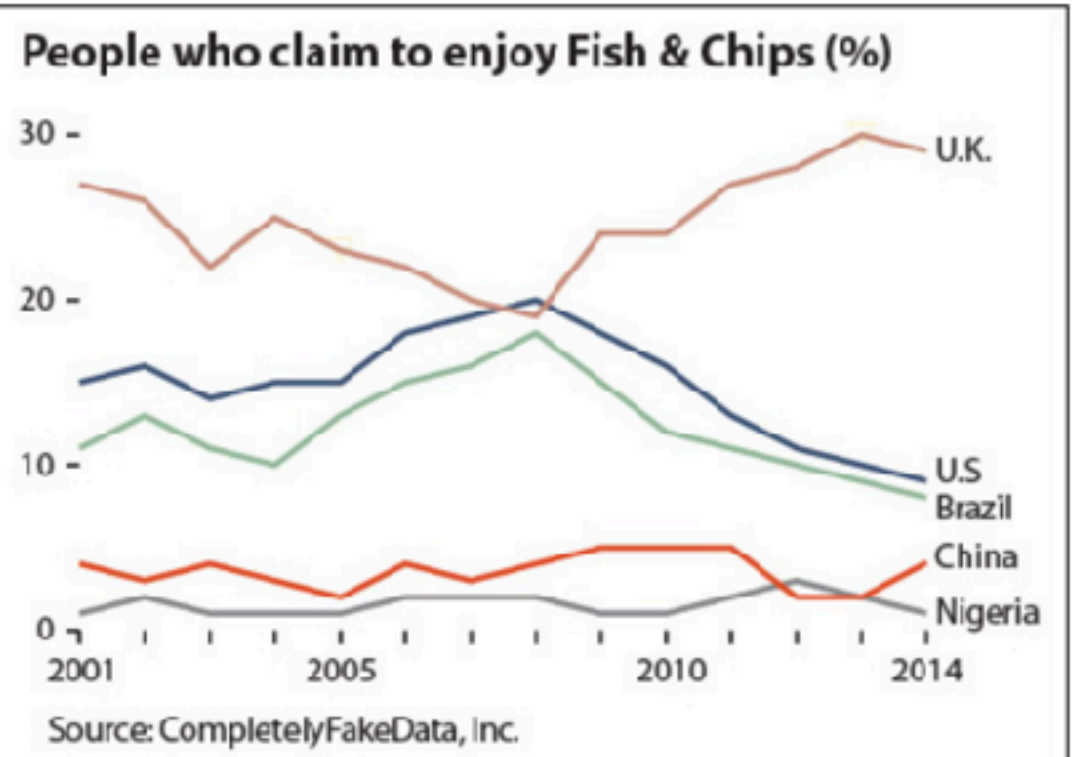


Slope chart is more effective to represent change between two points in time

PURDUE UNIVERSITY | College of Science

Beautiful: Which chart is more appealing?





Insightful: Spontaneous eureka

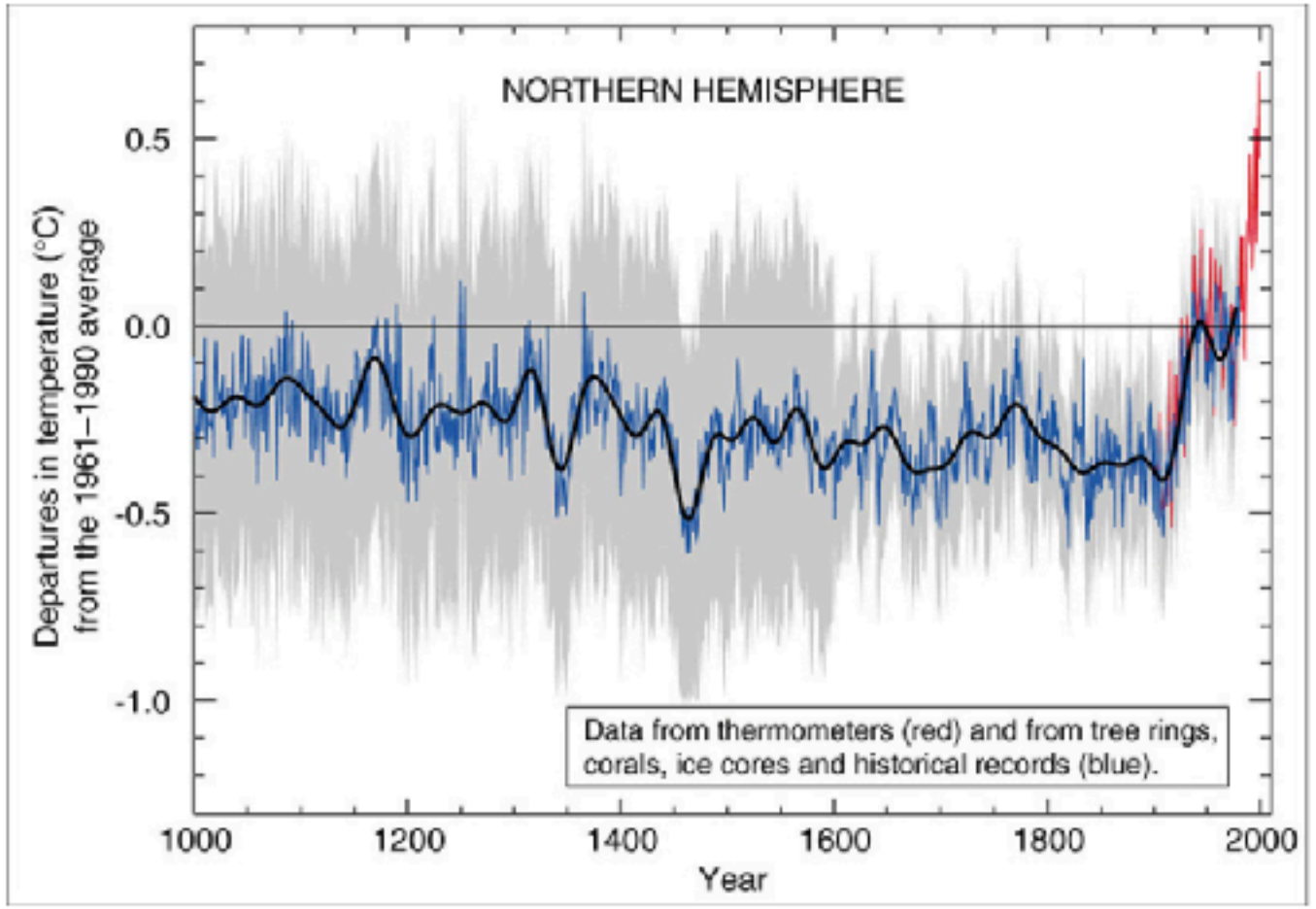
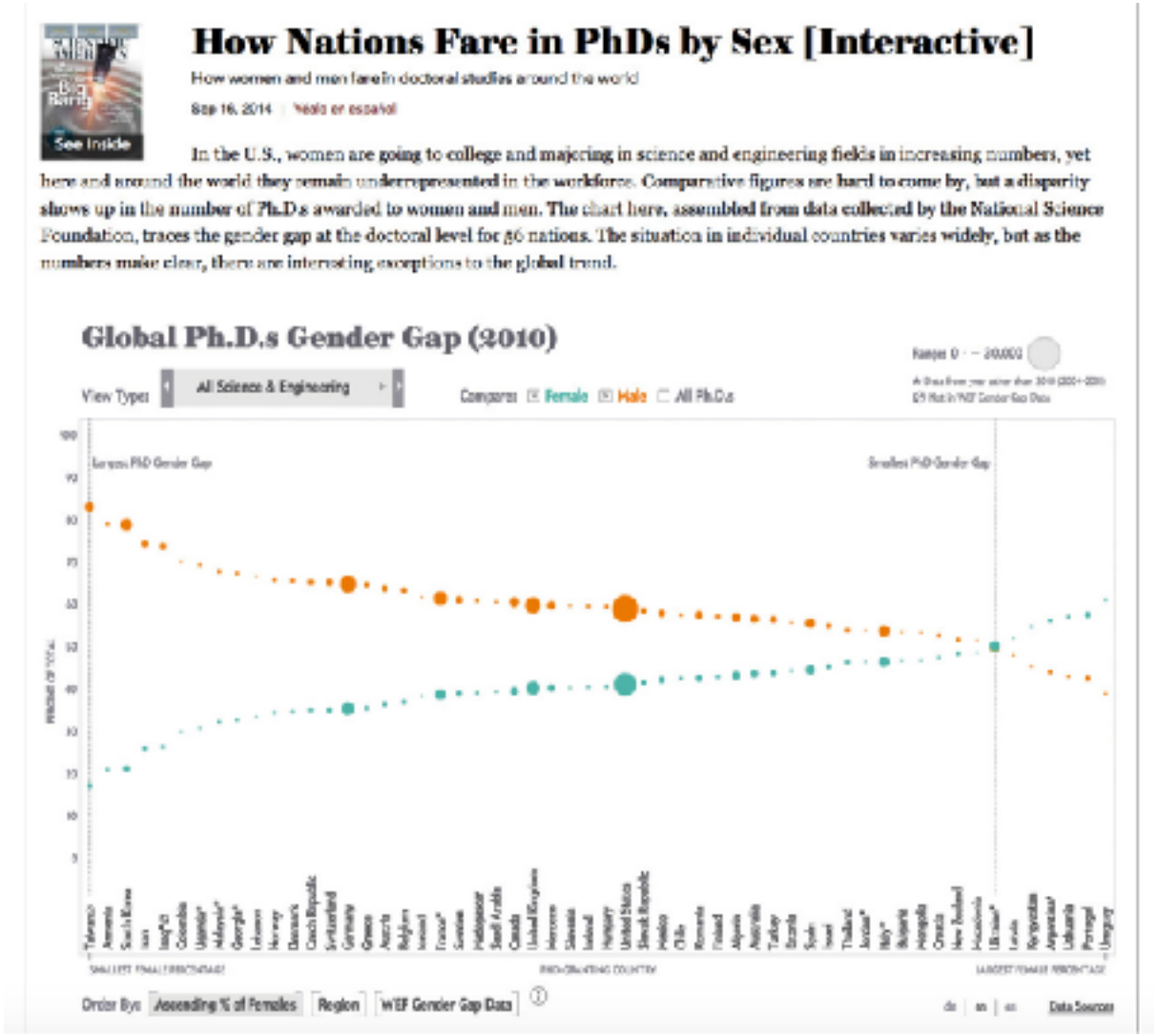


Figure 2.1 The hockey stick chart. Summary For Policymakers of the 2001 Third Assessment Report of the Intergovernmental Panel on Climate Change.

Insightful: Knowledge building

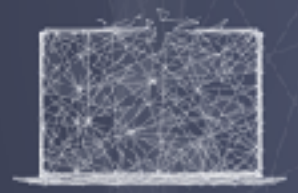


How Nations Fare in PhDs by Sex [Interactive]

How women and men fare in doctoral studies around the world

Sep 16, 2014 | Texto en español

In the U.S., women are going to college and majoring in science and engineering fields in increasing numbers, yet here and around the world they remain underrepresented in the workforce. Comparative figures are hard to come by, but a disparity shows up in the number of Ph.D.s awarded to women and men. The chart here, assembled from data collected by the National Science Foundation, traces the gender gap at the doctoral level for 36 nations. The situation in individual countries varies widely, but as the numbers make clear, there are interesting exceptions to the global trend.



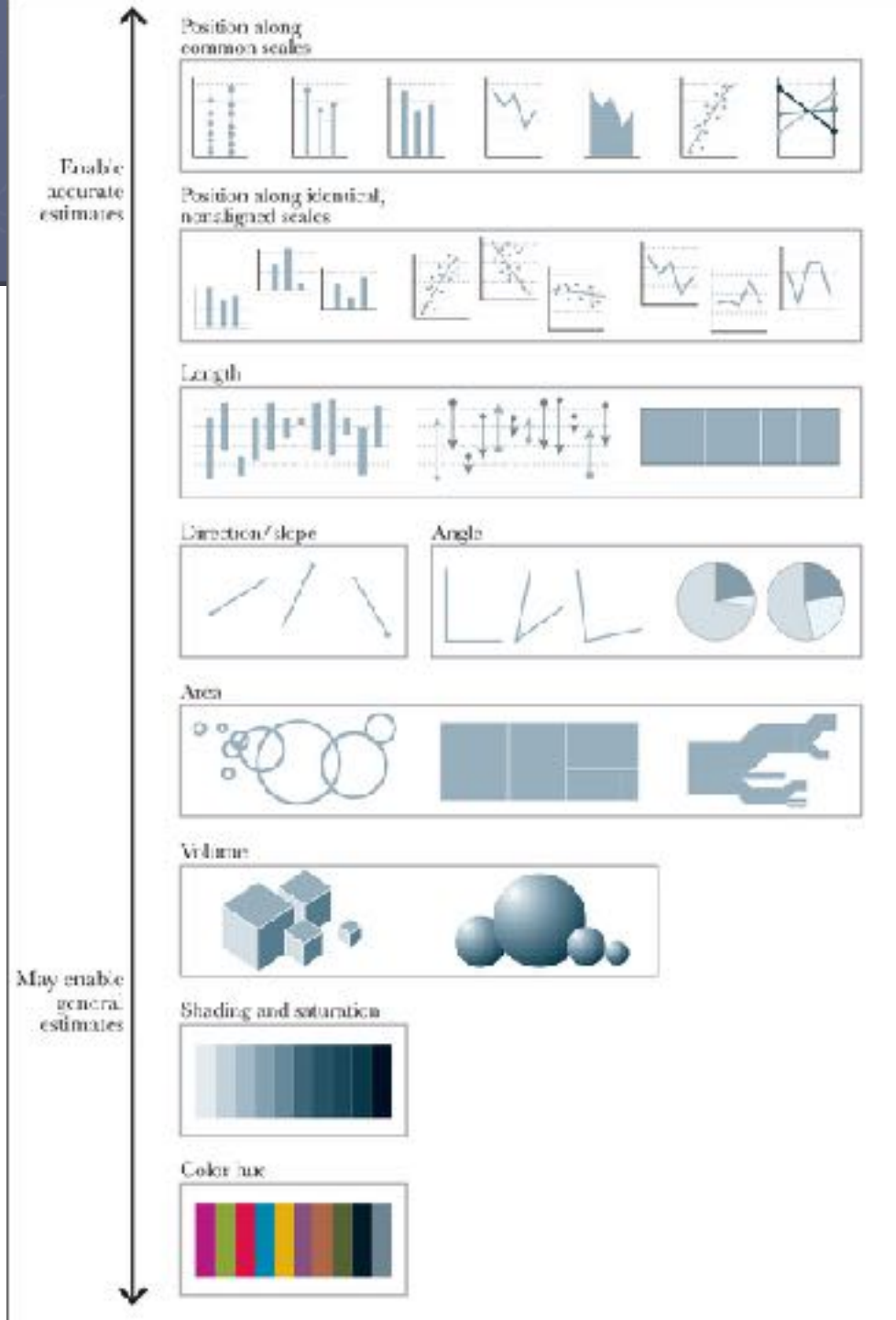
How to choose the best graphic form for your data

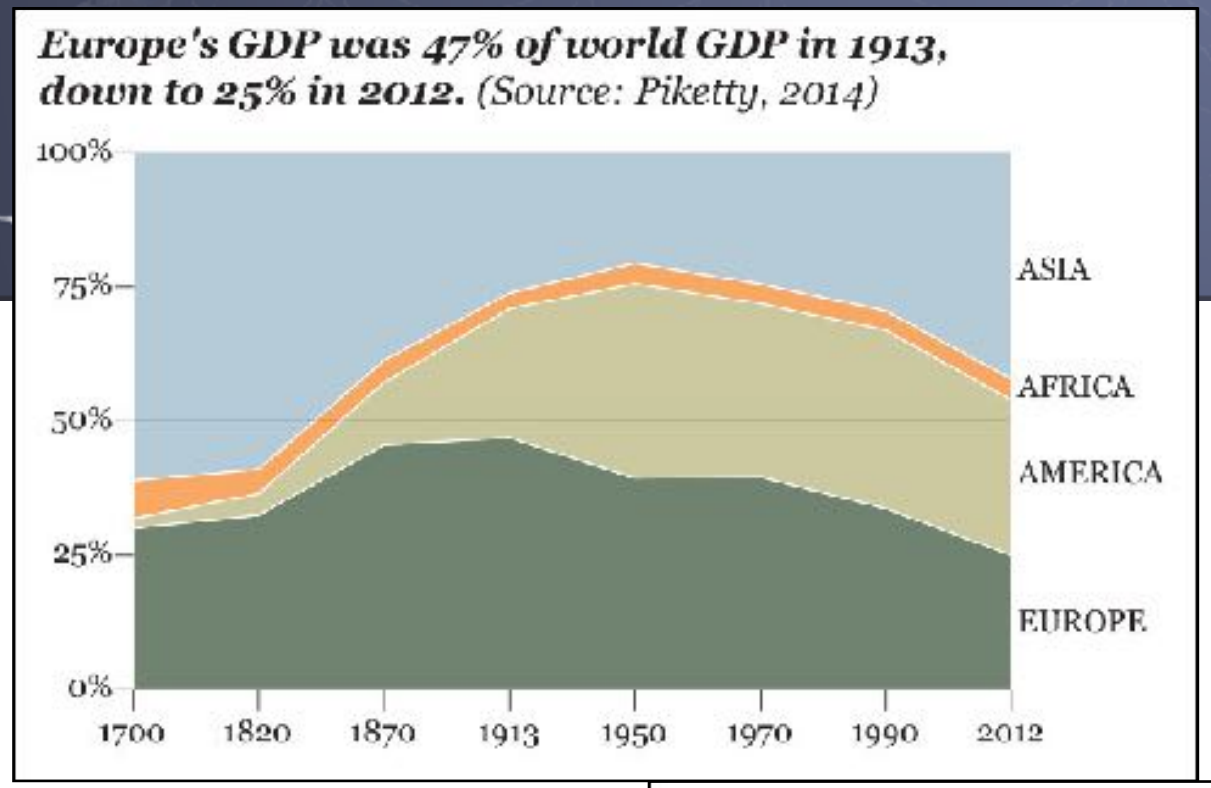
- Think about the task(s) you want to enable
- Try different graphic forms
- Arrange the components of the graph to make as easy as possible to extract meaning from it
- Test the outcomes yourself or people representative of your target audience

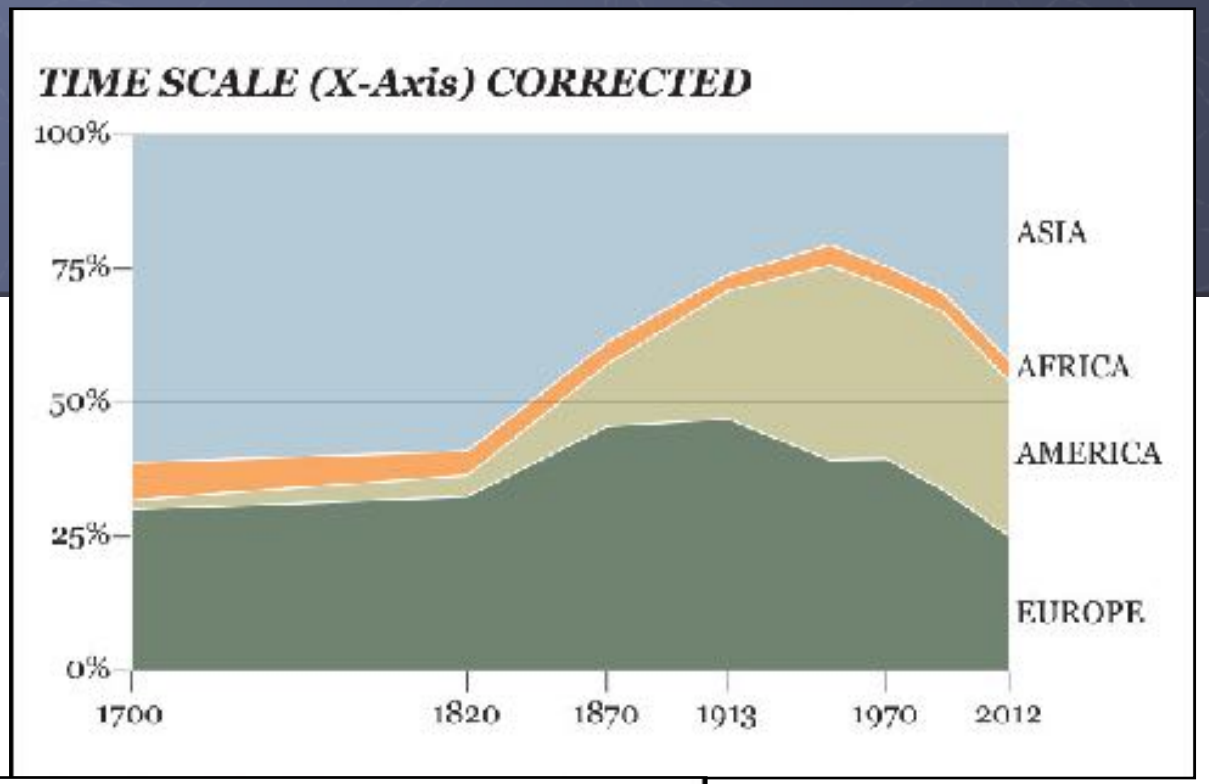
Graphical perception

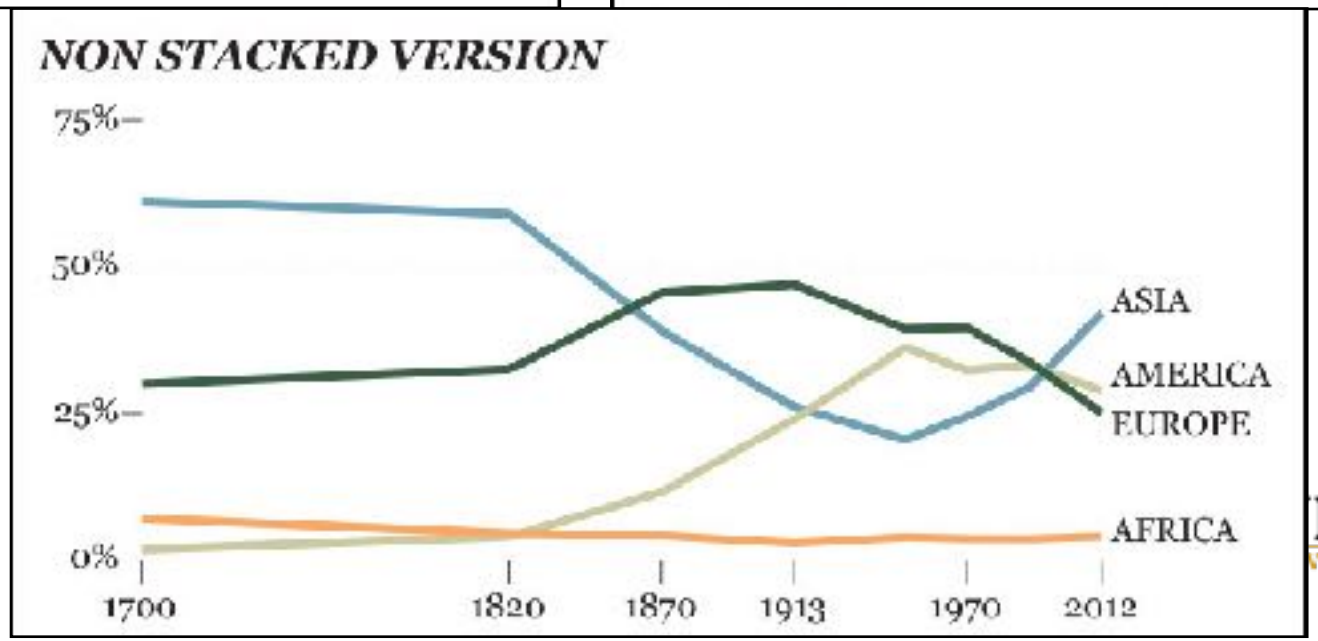
Scale of elementary perceptual tasks (Cleveland & McGill)

- User performs one or more of the associated mental-visual tasks to extract values represented in the plot
- To create a successful chart, construct it based on tasks as high in the hierarchy as possible

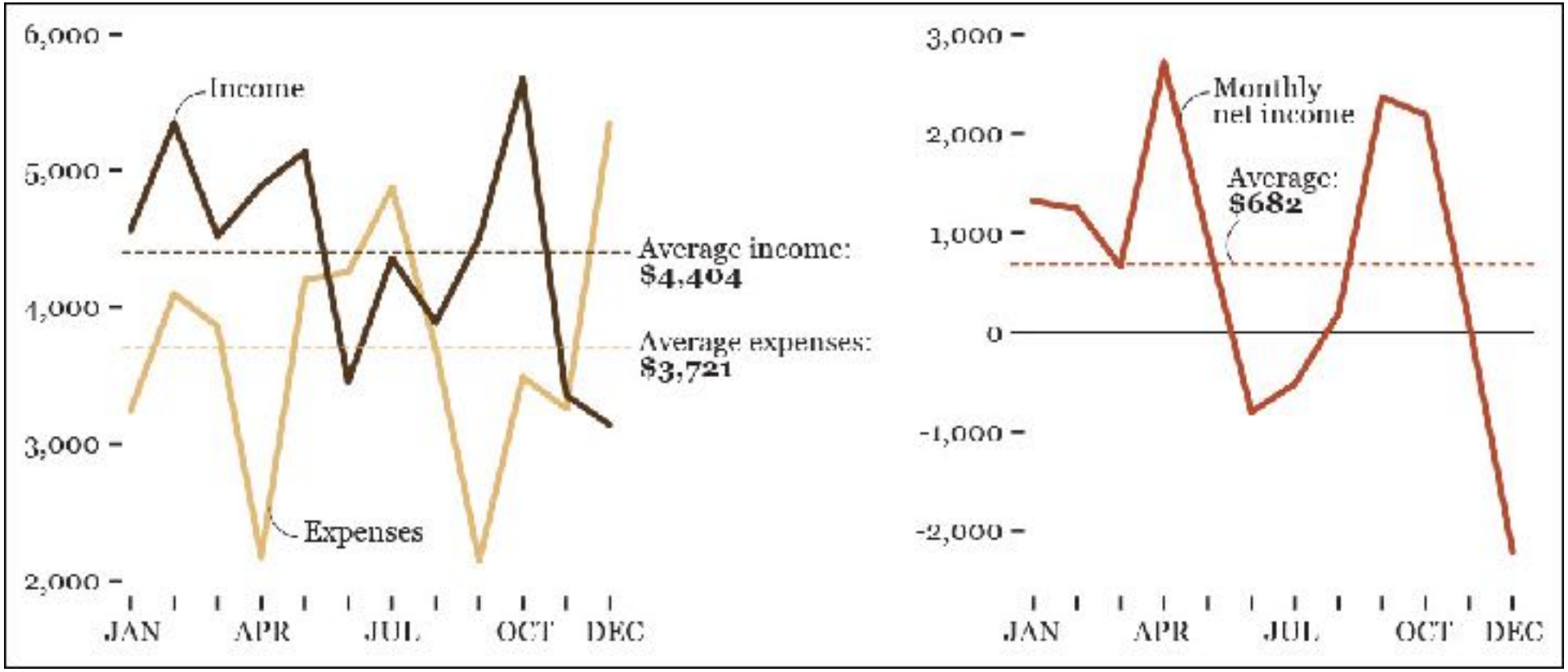



Europe's GDP was 47% of world GDP in 1913, down to 25% in 2012. (Source: Piketty, 2014)
100%
75%
50%
25%
0%
ASIA
AFRICA
AMERICA
EUROPE
1700
1820
1870
1913
1950
1970
1990
2012


TIME SCALE (X-Axis) CORRECTED
100%
75%
50%
25%
0%
ASIA
AFRICA
AMERICA
EUROPE
1700
1820
1870
1913
1970
2012

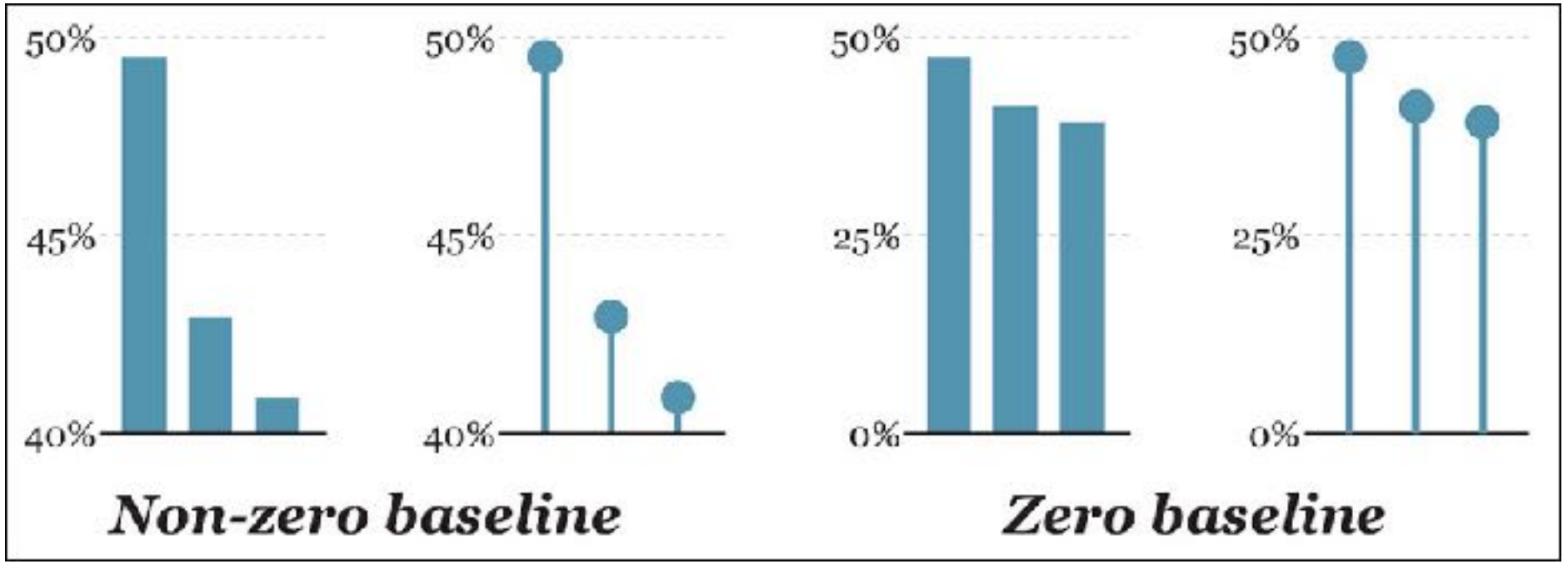

NON STACKED VERSION
75%
50%
25%
0%
ASIA
AMERICA
EUROPE
AFRICA
1700
1820
1870
1913
1970
2012

Which chart is better?



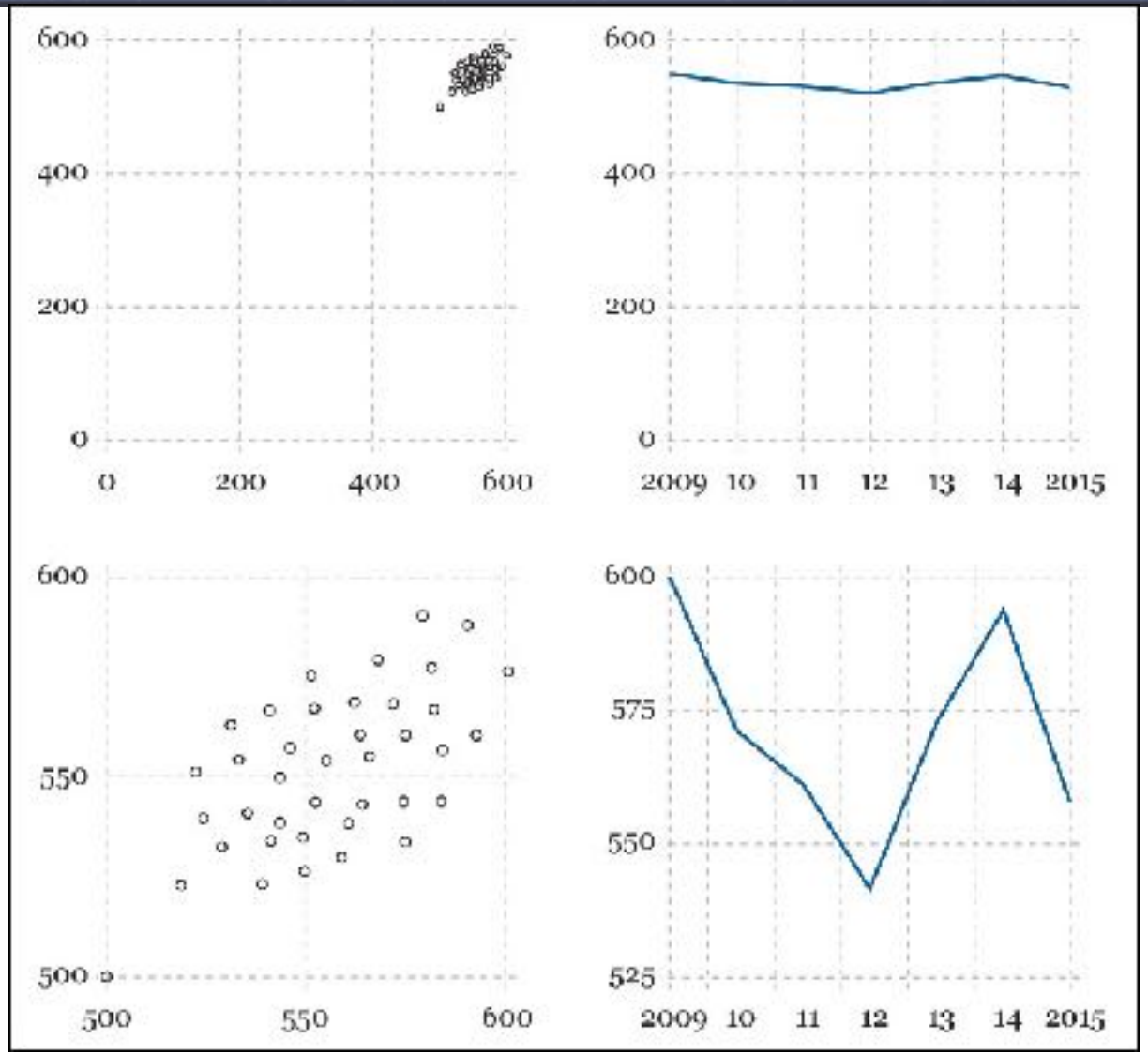
Depends on whether you want to emphasize income vs. expenses or if you want to emphasize the difference between them

Baselines and scales of plots



Should you always include 0 in the axis as a baseline?

Use logical and meaningful baselines



PURDUE UNIVERSITY | College of Science