INTRODUCTION TO DATA VISUALIZATION

CSE 551

CSE 551

Serves as a cumulative and hands-on experience course in a semester

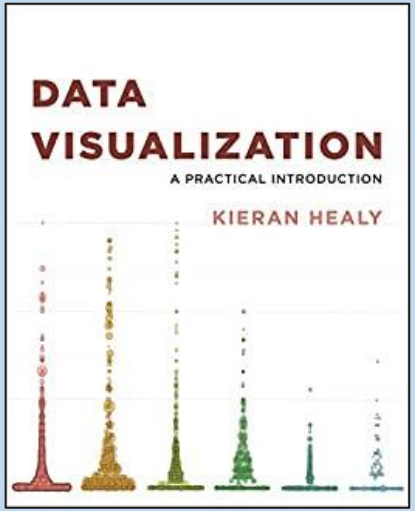
Goal: understand, process, and create visual representations from datasets

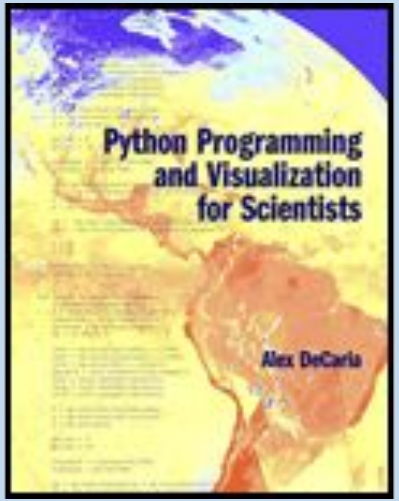
Complementary activities:

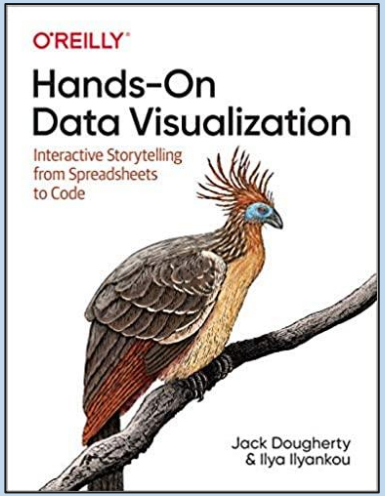
- Data collection and preprocessing
- Environment setup (e.g., Anaconda, R/Rshiny, and various packages)
- Source code control (e.g., git/github)
- Project management
- Presentations

CSE 551 OBJECTIVES

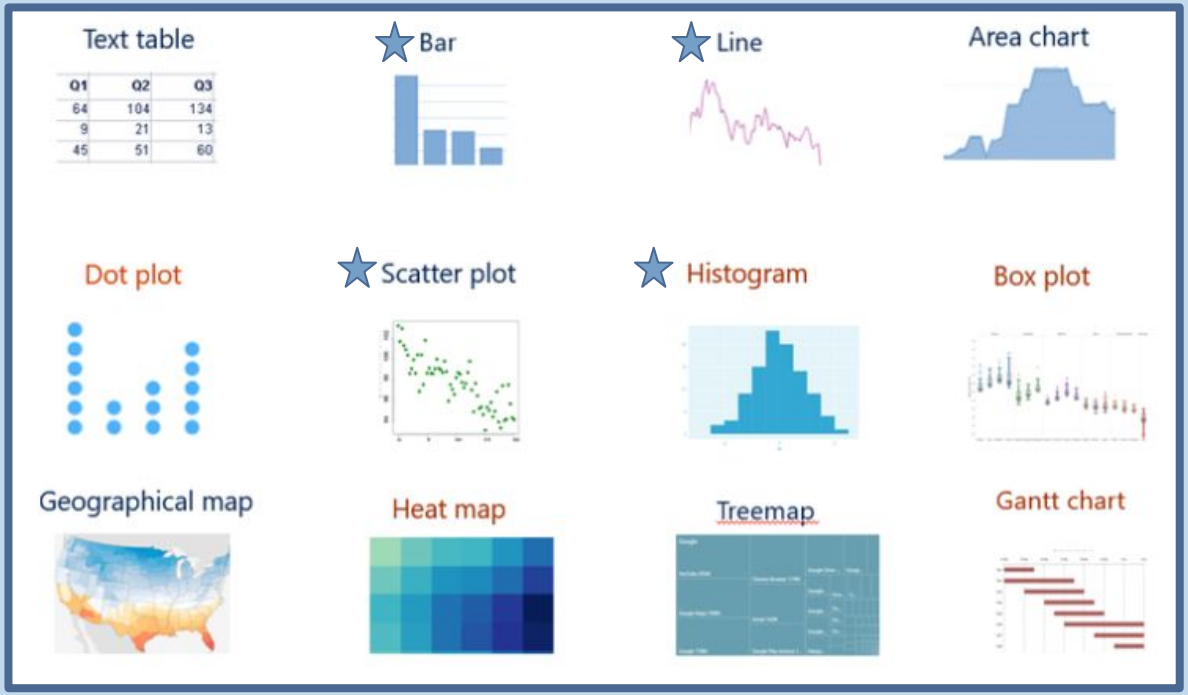
- Turning data into **actionable format** for **visualization**
- Understand the **ideas and methods** of data visualization in a sensible, **comprehensible**, reproducible way
- Create meaningful **visualizations** using **Python** from datasets
- Produce compelling **visual documentation** to support **analytics**







USEFUL BOOKS FOR THIS COURSE



12 CORE PLOTS

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

Three key questions:

- **Why computers?** Computers have a large amount of memory storage and can finish more tasks much faster than humans can manually.
- **Why people?** While certain tasks can be fully automated, analysis questions are often ill-specified. People are needed to ask the correct question, detect patterns, conduct analyses, debug, and check for errors to verify and build trust in the model.
- **Why visual?** Our eyes are drawn to colors and patterns and we can see trends and outliers quickly. Thus, we can build upon pre existing hypotheses. It is storytelling with a purpose.

HUMAN LIMITATIONS

- **Visual Working Memory (VWM) is small**
  - **VWM: a core cognitive function in which we perceive the identity of objects and perceive where objects are located in space at any given point in time to help focus attention**
  - **The capacity of visual working memory for adults is limited to approximately 4 visual times**
- **Change blindness: large changes go unnoticed when we are working on something else in our view**

The change blindness phenomenon can be demonstrated relatively easily by repeatedly presenting two pictures (let us call them picture A and picture B) one after the other. Picture B is the same as picture A except a single change has been made to it. For example, in a driving scene, picture B might be the same as picture A except that a car or pedestrian has been removed from the image. In a typical change blindness demonstration, picture A will be presented followed by a blank screen and then picture B followed by a blank screen, and so on (e.g., see Rensink et al., 1997). The interleaved blank screens simulate and have the same effect as making an eye blink (or eye movement) by masking the transients between the two images that would normally indicate the location of a change. This picture A–blank–picture B–blank sequence repeats and the task is to try to find the difference between the two pictures. Typical findings show that people are exceptionally bad at spotting the difference between the two images even with prolonged viewing times (Rensink et al., 1997).

DATA VISUALIZATION: BENEFITS AND DISADVANTAGES

ADVANTAGES

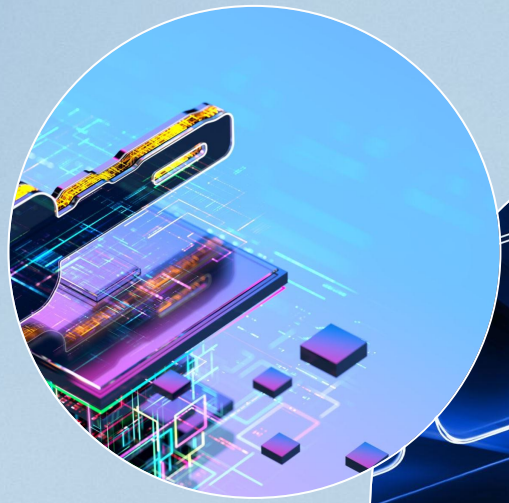
- Easily share information
- Interactively explore opportunities
- Visualize patterns and relationships

if we can see something, we internalize it quickly

DISADVANTAGES

- Biased or inaccurate information
- Correlation does not always mean causation
- Core messages can get lost in translation

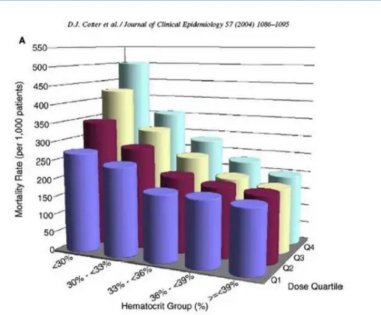
prevalent when designed wrong or with challenging data points



WHY IS EFFECTIVENESS IMPORTANT?

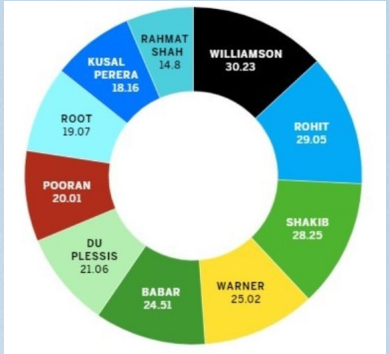
Even when we have the best intentions in communicating data through visuals, they don't always turn out the way we want. Sometimes it's the result of a mislabeled axis or a poor choice of color. Other times, we may choose the wrong type of chart entirely. Whatever the case, a bad data visualization can derail the message we want to communicate to the audience — or lead them to draw inaccurate conclusions.

KEY THINGS TO NOTE..



don’t use 3D charts

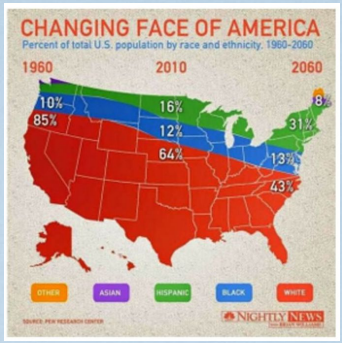
While they may look flashy, they rarely do a good job of communicating important information to your audience. Past the cool factor, 3D bar charts are often hard to read, making them more trouble than they’re worth.



use pie charts wisely

Pie charts are fairly narrow in their usage. But they’re inherently meant to add up to a sum percentage. This one doesn’t follow that general rule. In fact, the numbers don’t appear to add up to any sensible total.

KEY THINGS TO NOTE..



don't use misleading graphs

While interesting to look at, using a map doesn't help you understand what's going on here. For example, the fact that Oklahoma is red or that other states are partially different colors doesn't really mean anything. In addition, the racial and ethnic differences have no correlation to the states.

don't over-graph

This chart is too difficult to read. It should be broken up into a few separate visuals. Only visualize necessary and important variables and don't be afraid to make multiple graphs.