POLS 5377 Scope & Method of Political Science

Week 7 Data Presentation

Data presentation

Healey. (2016) *Statistics: A Tool for Social Research,* Chapter 2

Key Questions:

- What are the most commonly used graphs for data presentation?
- How do we decide which graph we should use?
- How can graphs be misused, and how do we avoid it?

Outline

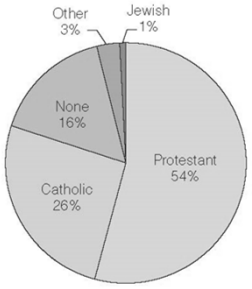
- Pie charts
- Bar charts
- Histograms
- Line charts
- Using Graphs Appropriately

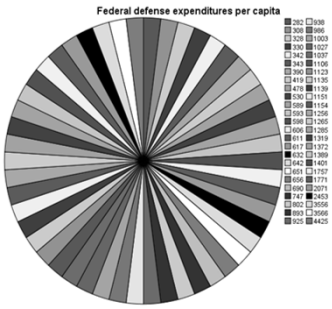
Using Graphs to Present Data

- Pie charts, bar charts, histograms, and line charts present frequency distributions graphically
- Graphs are commonly used for presenting “pictures” of research results
- Graphs and charts are very useful ways to display the overall shape of a distribution

Pie Chart

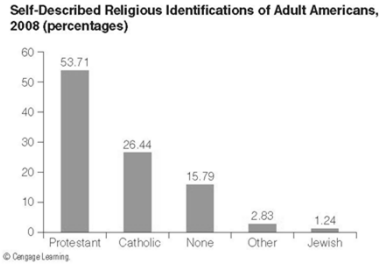
- **Pie charts** are useful for categorical variables with only a few categories. The pie is divided into segments which are proportional in size to the percentage of cases in each category.
- When more than five, may consider to use other types of chart.

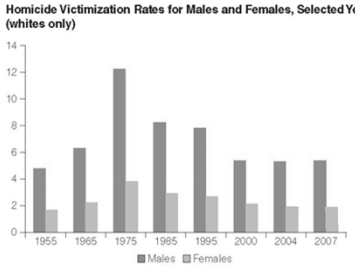




Bar Chart

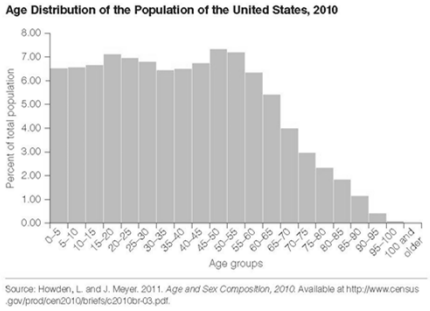
- **Bar** charts are useful for *categorical* variables. The categories are represented by bars. The height of these bars corresponds to the number or percentage of cases in each category.





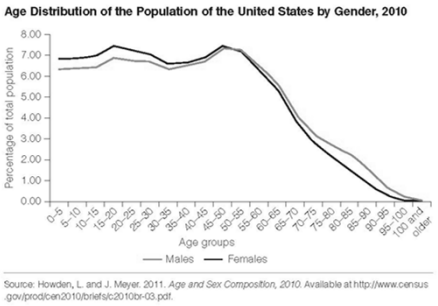
Histograms

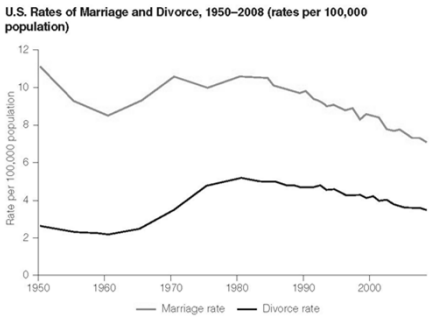
- Histogram: most appropriate for continuous *interval-ratio level* variables, but can be used for discrete interval-ratio level variables
- Look like bar charts, but categories (or scores) of the variable border each other



Line Charts

- Sometimes called frequency polygons
- Constructed similarly to a histogram, except graph a dot at each category's midpoint and then connect the dots
- Especially appropriate for continuous interval-ratio level variables, but can be used for discrete interval-ratio level variables



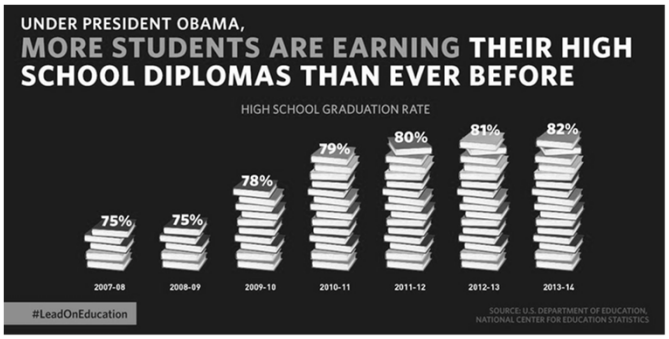


Using Graphs Appropriately

- Because graphs are powerful and effective tool to deliver messages, they can also be misused by the presenter or misleading to the audience.
- We need to be extra cautious when presenting or reading graphs.
- Here are some examples to show the importance of using graphs appropriately:

Using Graphs Appropriately

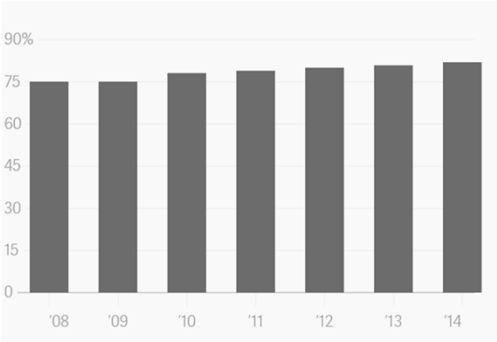
- In 2015, the White House tweeted: “Good news: America’s high school graduation rate has increased to an all-time high.” The tweet included this chart:



Source: http://qz.com/580859/the-most-misleading-charts-of-2015-fixed/

Using Graphs Appropriately

- The chart is misleading, because:
  - The illustration does not represent the data in appropriate scale.
  - The column chart didn't start the vertical axis at zero



Source: http://qz.com/580859/the-most-misleading-charts-of-2015-fixed/

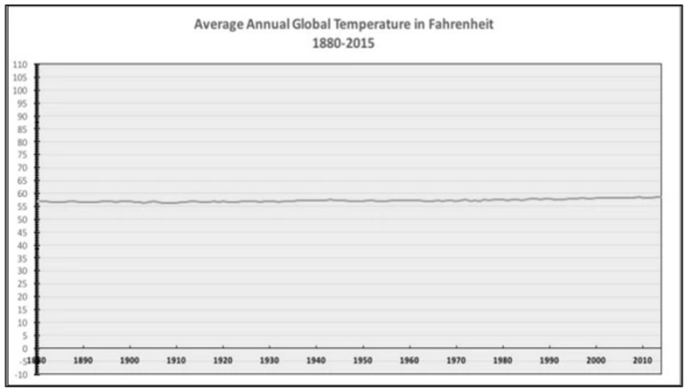
Using Graphs Appropriately

- To emphasize the change of rates, we can present the data with the y-axis starts from a higher value. But it is better to include more data points to give the audience a good reference.

Source: http://qz.com/580859/the-most-misleading-charts-of-2015-fixed/

Using Graphs Appropriately

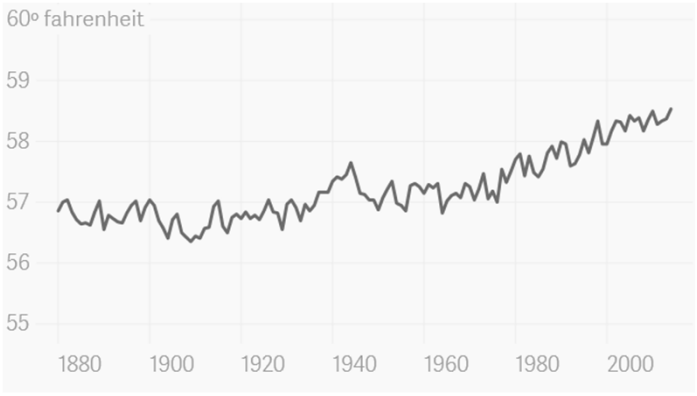
- Does Global Warming exist? National Review tweeted this line chart in December 2015.



Source: https://twitter.com/NRO/status/676516015078039556/photo/1?ref_src=twsrc%5Etfw

Using Graphs Appropriately

- If we want to accurately represent the change of global average temperature since 1880, we can start the y-axis at a higher number:



Source: http://qz.com/580859/the-most-misleading-charts-of-2015-fixed/

Using Graphs Appropriately

- Watch this video to learn more about Misleading Graphs: https://www.youtube.com/watch?v=GEsjMbTPiZM
- Common misleading practices:
  - Vertical axis does not start from zero (it is appropriate under certain circumstances)
  - Using illustrations or icons
  - 3D graphs
  - Uneven interval
  - Missing data points
  - Wrong information

Dealing with Data

- As a analyst, when you get a dataset, first thing you should do: screen the dataset by using the descriptive statistics and check the following common problems:
  - Improbable values
    - Check the values that outside of expected range, for example, the age of respondents that is 3, 4, 5, or 99, 107…etc.
  - missing values
    - Check the blanks that should not be existed
  - outliers
  - inconsistent spelling

After this lecture:

You should be able to:

- Distinguish different types of graphs used in descriptive statistics
- Be aware of the misleading graphs
- Be a conscious data consumer and responsible data presenter