EGR 1330 Computational Thinking with Data Science

Data Display

Outline

- Charts: line, bar, histogram, scatter plot.
- Pandas library
- Matplotlib library

Objective

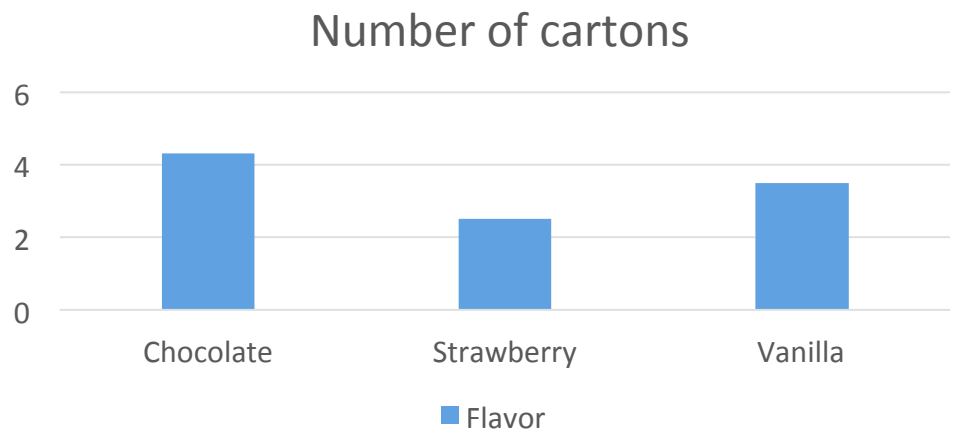
- Be able to visualize data with different types of charts using Pandas or Matplotlib.

Categorical Distributions

- Data come in many forms that are not numerical. Data can be pieces of music, or places on a map. They can also be categories into which you can place individuals:
  - The individuals are cartons of ice-cream, and the variable is the flavor in the carto
  - The individuals are professional basketball players, and the variable is the player's team.

Bar charts

- Used to represent graphical representation for categorical distributions.
- The bars are equally spaced and equally wide. The length of each bar is proportional to the frequency of the corresponding category.

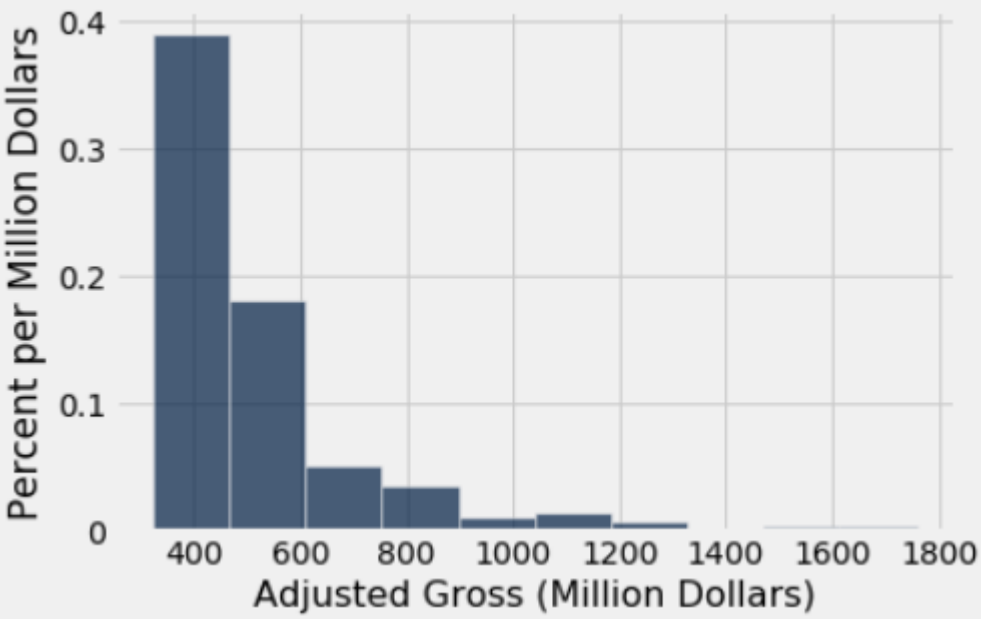


Live example

- Find a simple categorical dataset and create a bar chart using Pandas and Matplotlib.

Histogram

- A **histogram** is an approximate representation of the distribution of numerical data.
- Each bar is a contiguous intervals called *bin.*

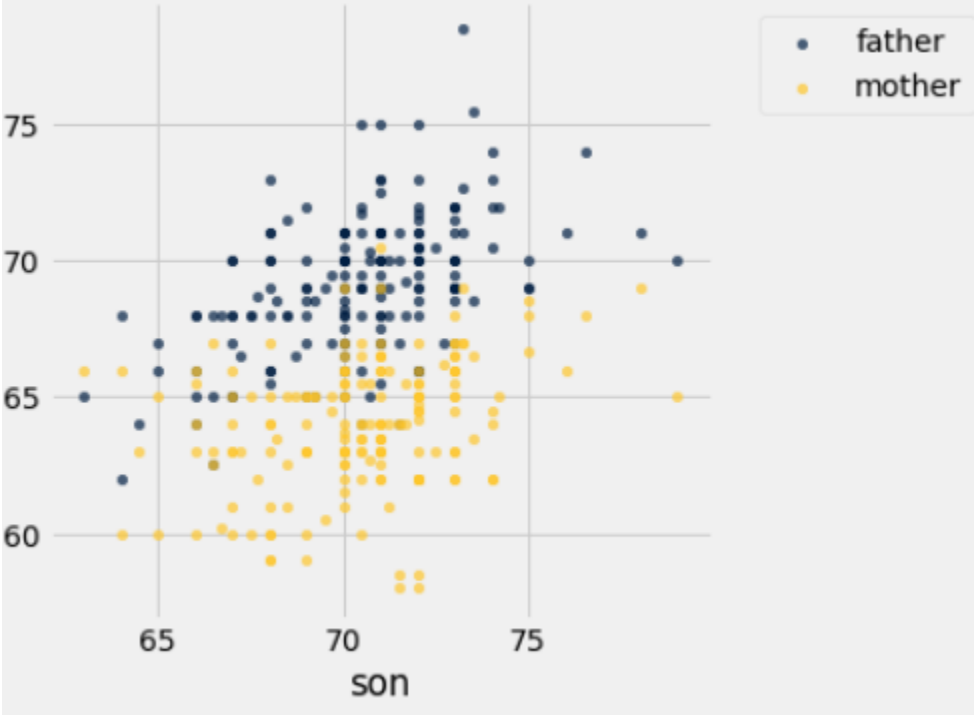


Live example

- Use top-movies dataset and create a histogram chart using Pandas and Matplotlib.

Scatter Plot

- A **scatter plot** is a type of plot using Cartesian coordinates to display values for typically two variables for a set of data.

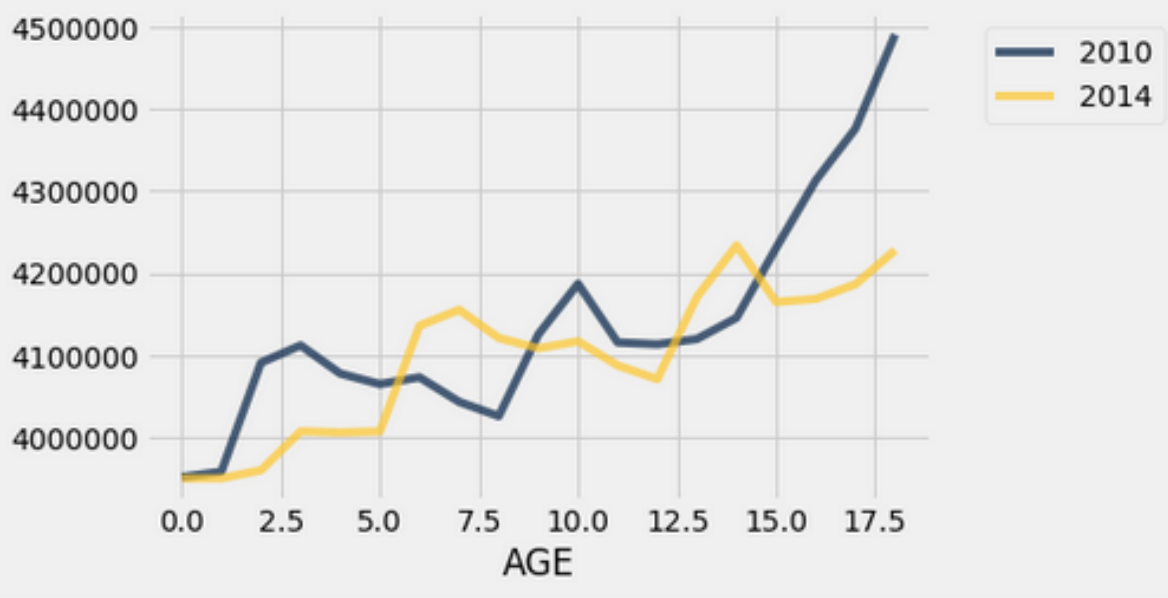


Live example

- Use galton-subset dataset and create a scatter plot using Pandas and Matplotlib.

Line Chart

A line chart or line plot or line graph or curve chart is a type of chart which displays information as a series of data points connected by straight line segments.



Live example

- Use census dataset and create a line chart using Pandas and Matplotlib.