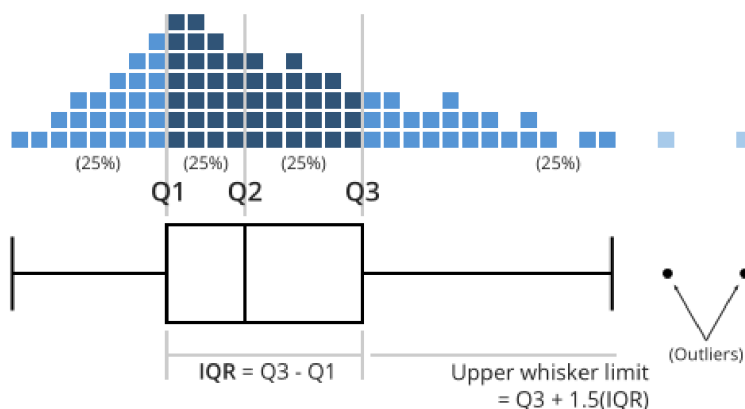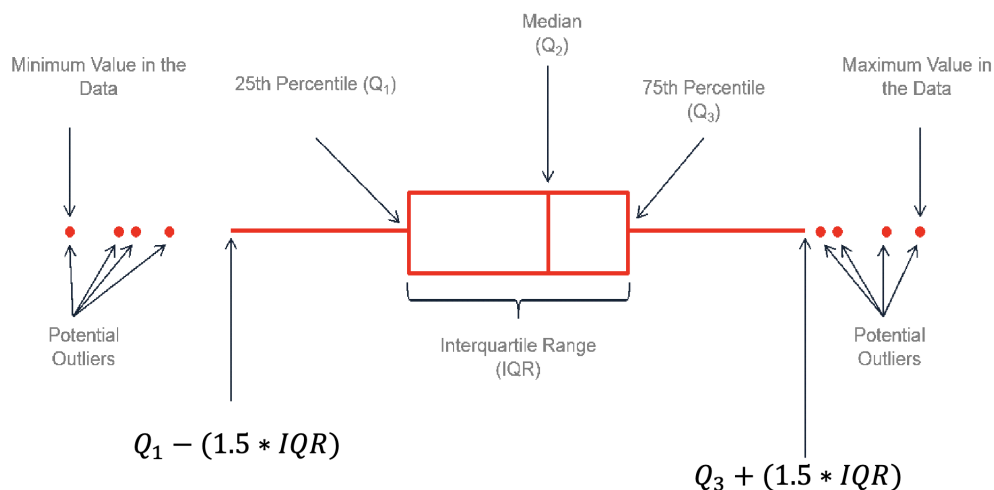# Data Visualization Continued

Visualization with One Quantitative Variable – Box Plot

- Displays a quantitative variable's distribution, while providing access to information such as median, quartiles, interquartile range, outliers, minimum and maximum

- The whiskers end on a data point

- Some data may be lost in box plots, for example, both of the histograms shown below generate the same box plots.
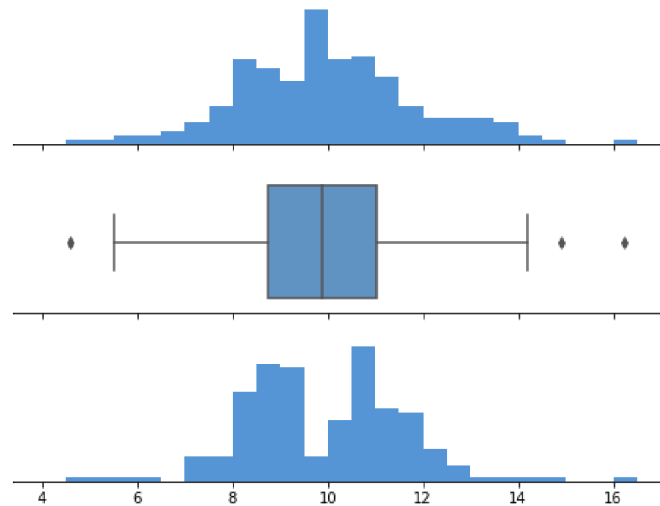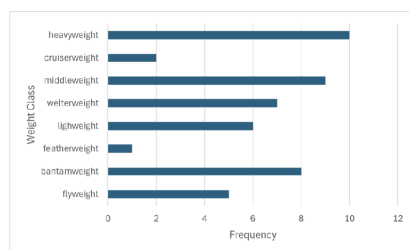
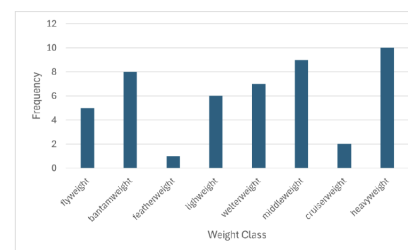- We can plot many box plots on the same graph to allow us to see trends
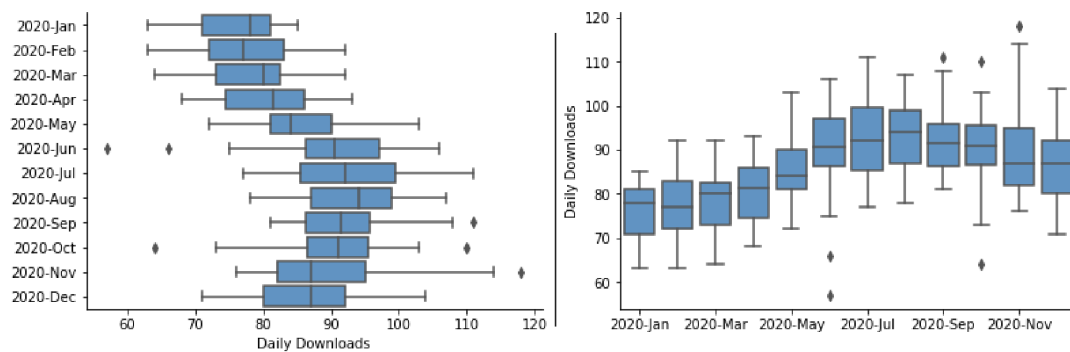
Visualization with One Categorical Variable – Bar / Column Charts

- These charts can display a frequency distribution for a single categorical variable

- They can show the number of instances in each category for a variable
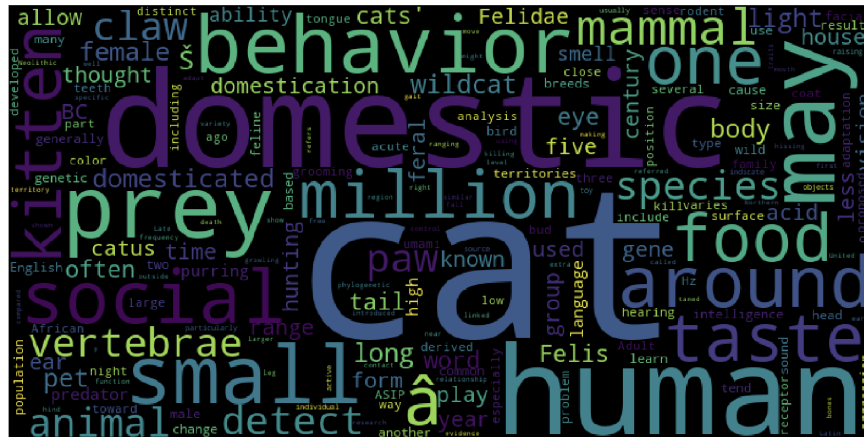

Bar Chart   Column Chart

Visualization with One Categorical Value – Word Cloud

- A collection of words shown in different sizes related to how frequent a word appears in a corpus (collection of text)

- Large words indicate more frequent words

- No axes

- No indication of raw frequency counts, it becomes relative to the frequencies of other words

- To use word clouds, you must install `wordcloud` via python
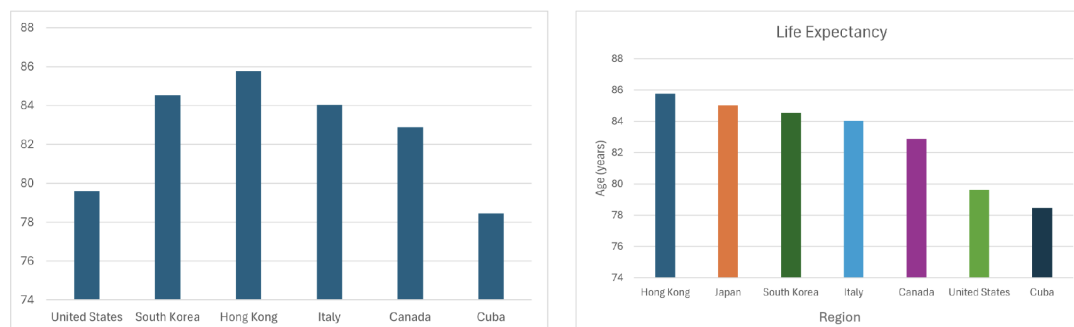

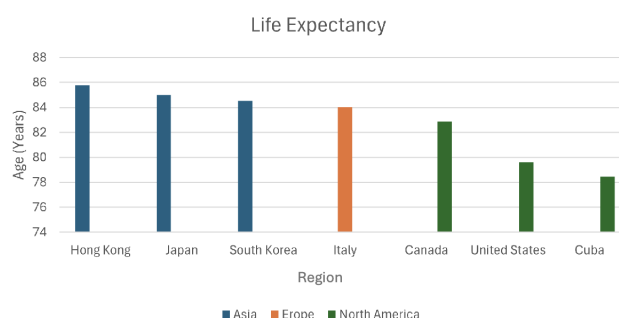
Visualizations with Two Variables

- These allow for more sophisticated graphics

- Shows the relationship between two variables

Visualizations with Two Variables – Bar / Column Charts

- These charts can be used for visualizing values of quantitative variables by category

- Bar and column charts can contain two categorical variables and a qualitative variable

- There are a few ways to visualize additional categorical variables

Visualizations with Two Variables – Tree Map

- Displays hierarchical data, which is data structured with parent-child relationships

- Examples of Hierarchical Data:

    - Continent - Country
    - Manufacturer - Vehicle Model

- Tree maps are divided into the categories from the data and the size of rectangles indicate the value of the data

- Rectangles are grouped in larger rectangles, which captures the hierarchical property of the data

Visualizations with Two Variables – Scatterplot

- Displays the relationship between two quantitative variables

- Response (dependent) variable - often the primary variable of interest that is thought to be affected by one or more explanatory attributes

- Explanatory (independent) variable - a variable that is thought to affect, influence, explain, or be associated with the response variable
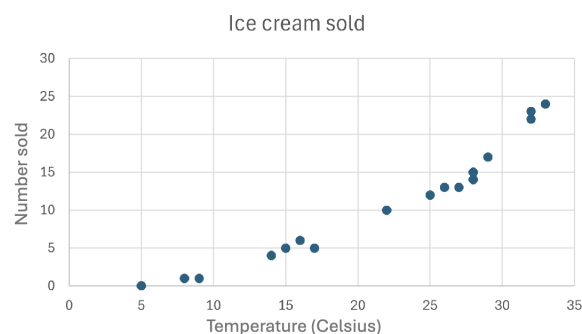
Visualizations with Two Variables – Line Chart

- Displays data as a series of points connected by a series of lines, often showing a value over time