Deck 3

* Data dimensionality
  + The number of variables in your data
* Visualizing data
  + Scatterplots
    - 2d plot
  + Pairplots
    - 2d plot
  + Coplots
    - 3d plot with 2d slices
    - Example
      * Categorical with 2 continuous numerical
      * Penguin flipper length, body mass, and species
      * Flipper length y-axis
      * Body mass x-axis
      * Species are the multiple graphs plotted next to one another to show their difference
    - 2nd example
      * When all continuous numerical, it depends on the algorithm how things are split up
      * With body mass as the category to compare (called a bin), the algorithm of choice has some overlap between mass to have equal an equal number of data points in each plot
  + Univariate data/plots
    - Shows shape of the distribution of the data
    - Sometimes called frequency diagrams
    - Data aggregated into bins
    - Histograms
      * Bill length x- axis frequency y-axis
    - Boxplots
      * Line is median, box then extends to the 25% and 75% percentile
        + Can see data is skewed based on the location of the line, if the median line is not in the center of the box and whiskers are also longer on that side, then it shows data is skewed toward that side
      * Whiskers are special
      * Can show 3d data as long as you have 1 numerical dataset and 2 categorical
        + Bill length, male/female, species