Data Science Toolbox

Types of Questions:

1. Descriptive Analysis
   1. Describe a set of data
      1. The first kind of data analysis performed
      2. Commonly applied to census data
      3. The description and interpretation are different steps
      4. Descriptions can usually be generalized without additional statistical modeling
2. Exploratory Analysis
   1. Find relationships you didn’t know about
      1. Exploratory models are good for discovering new connections
      2. They are also useful for defining future studies
      3. Exploratory analysis are usually not the final say
      4. Exploratory analysis alone should not be used for generalizing/predicting
      5. Correlation does not imply causation
3. Inferential Analysis
   1. Use a relatively small sample of data to say something about a bigger population
      1. Inference is usually the goal of statistical models
      2. Inference involves estimating both the quantity you care about and your uncertainty about your estimate
      3. Inference depends heavily on both the population and the sampling scheme
4. Predictive Analysis
   1. To use the data on some objects to predict values for another object
      1. If x predicts y it does not mean that x causes y
      2. Accurate prediction depends heavily on measuring the right variables
5. Causal Analysis
   1. To find out what happens to one variable when you make another variable change
      1. Usually randomized studies are required to identify causation
6. Mechanistic Analysis
   1. Understand the exact changes in variables that lead to changes in other variables for individual objects
      1. Incredibly hard to infer, except in simple situations