Lesson 1.1 - February 2, 2021

Introduction

- Notes can be found in `files` tab on Canvas
- Recordings of the lecture can be found on the ConferZoom tab on Canvas
- Take pictures of the take-home part of the examination (Final and Midterms)

Basics

- Two types of statistics: Descriptive and Inferential
- Inferential statistics is using mathematics to try and make some kind of conclusion or larger group
- The main idea of statistics is to make a statement about a large group (population)
  - How will people in California vote in an election?
  - Is that within the realm of chance, or is there something else that is responsible for this generalization
- In practice, it's very difficult to examine or survery every person in a population, so we look at samples
- **Sample**: A small section or subset of the entire population
- If the sample is chosen well in conjunction with several other nice properties, then the ideas noticed in the sample will be similar to that of the entire population
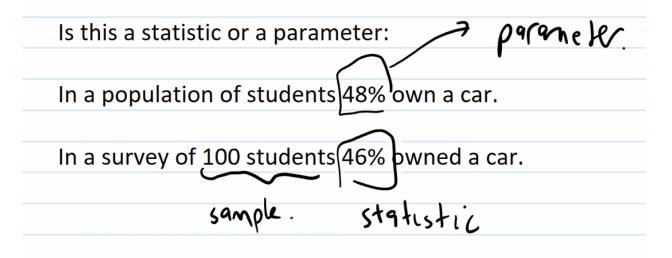
Vocabulary

- **Population**: Large group of objects that we are studying
- **Sample**: Smaller subset of the population that is feasible to work with
- **Data Science**: A field with a similar goal to statistics, but also utilizing computer science
- **Statistic**: A numerical value asssociated with a sample (average, minimum, maximum, etc.)
- **Descriptive Statistics**: Looking at how to use numbers or graphs to describe a sample
- **Inferential Statistics**: Methods that take results for the sample and extend them to the entire population

Sample vs. Population

- Samples have statistics while populations have parameters

- Example:

Is this a statistic or a parameter: → parameter.

In a population of students 48% own a car.

In a survey of 100 students 46% owned a car.

sample.   statistic

Data

- Data is information that we collect about a sample
- Variable is a type of information that we collect (data type equivalent of programming)
- Variables can be either categorical (qualitative) or quantitative
- Data can be stacked or unstacked
  - Stacked means that all data in a specific row will be the same type
  - Unstacked is the opposite of stacked data
- Categorical data can be integers as well, but may be something like a ZIP Code or phone number