**Math 124 Lecture 2**

# Where are we at and where are we heading?

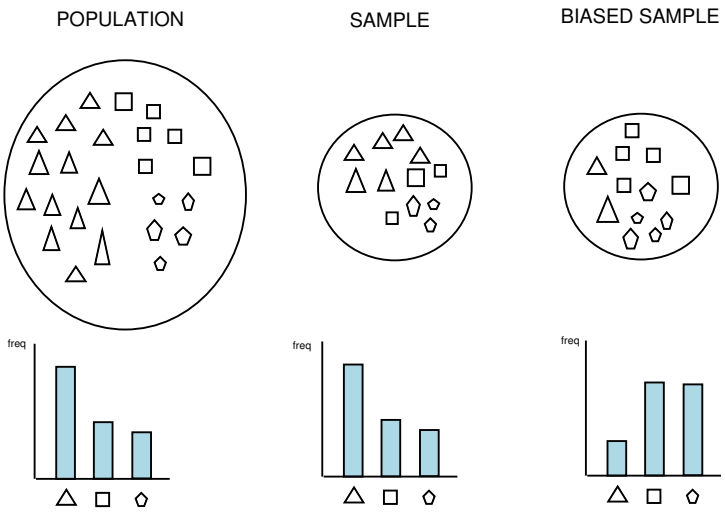
Last time we talked about data and how statistics was about collecting, organizing and understanding data. Today we will expand on these issues and then begin examining how we can look at the distribution of a dataset.

# Samples and Populations

First lets define these two terms:

- **Population** a well defined set of set of objects/units/individuals that we are interested in studying eg All adults in the USA, all widgets produced at a factory in a month, all bottles filled in a day at a soft drink company
- **Sample** a subset of the the population eg the first 1000 people selected by randomly choosing from valid social security numbers, every 10th widget off the assembly line, the bottles filled in the first hour of the day

A study of all the elements of a population is called a *census*. In an ideal world for anything which we wanted to study we could carryout a census, unfortunately this is almost never possible. Why? Cost or time prohibitive or perhaps our measurement method is destructive. So instead we usually instead look at data from samples from the population. One property that we would like is that our sample reflect the population from which it is drawn so that any conclusions we make based on the sample can be generalized to the population. We would say that a sample that does not reflect the general population as biased.

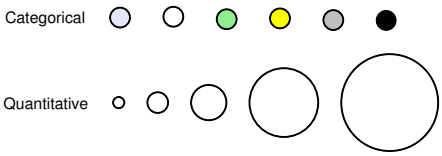



# Variables

- **variable** a characteristic which may change between different objects in a population.

  **Categorical Variable** a variable that places each individual into one of a specified number of categories. This may or may not be ordered. eg high or low, eye color, gender

  **Quantitative Variable** a variable which takes numerical values. Something for which we could add or average. eg heights, volumes, weights




# Two different types of statistical analysis

- **Exploratory Data Analysis** sometimes called *descriptive statistics* is about investigating the values of the measurements of the variables in your dataset. This could involve examining each variable by itself or looking at relationships between variables.

- **Confirmatory Data Analysis** sometimes called *inferential statistics* is the about using information from the sample to draw conclusions about the population from which it is drawn.

## What will we do in this class?

We will investigate both Exploratory Data Analysis (EDA) and Confirmatory Data Analysis. We will spend the next sequence of lectures exploring topics in EDA and leave CDA for later in the semester.