

# Chapter 1: Collecting and Analyzing Data

## Vocabulary

Data: Information in all forms.

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Bias: When data does not represent the population.

**Introduction:** The goal of collecting and analyzing data is to understand the world around us. How data is collected is very important. The goal of collecting data is to get “unbiased” data that represents the population. Analyzing biased data may result in incorrect conclusions and lead to a misguided view of the world around us. It is also important to have a goal in mind when you collect data. Are we trying to find a population percentage from categorical data or a population average from quantitative data? Are we trying to show that two variables are related or are we trying to show cause and effect? Data needs to be collected differently depending on what goal you have in mind.

---

## Section 1A – Two Types of Data – Categorical and Quantitative

One of the most important factors when analyzing data is to determine what type of data you have and how many variables you are analyzing. Let us start with the type of data.

There are two general types of data, categorical and quantitative.

### Categorical Data

Categorical data (or qualitative data) are generally labels that tell us something about the people or objects in the data set. For example, what country do they live in, what is the person’s occupation, or what kind of pet they have?

Usually categorical data is made up of words (do you smoke - yes or no), but occasionally a number can be used as a category. For example, a zip code can be used instead of the place a person lives. The numbers “1” and “2” may be used instead of yes and no.

### Quantitative Data

Quantitative data are numbers that measure or count something. They usually have units and taking an average makes sense. For example: a list of people’s heights in inches, or their weights in kilograms, or a list of how many dogs are there in various animal shelters across Los Angeles. Notice in each of these cases the data is numerical and an average seems appropriate in the context. We can find the average height, the average weight, or the average number of dogs in animal shelters in Los Angeles.

### Numbers used as categories

Remember, not all numeric data is quantitative. Ask yourself if the numbers are measuring or counting something and if an average would make sense. For example, a list of people’s zip codes are numbers but an average zip code would not really tell us anything. In addition, identity numbers like hospital ID numbers, student ID numbers or social security numbers are not measuring anything and an average would not make sense in the context so they are not quantitative.

---

