Chapter 1 Stats Starts Here

Data values, no matter what kind, are useless without their context.

The Five W's for data: Who, What, When, Where, and Why

"How" can also be included in the list as well.

The first step in any data analysis is to know what you are trying to accomplish and what you need to know.

In many instances, all 5 W's are not available or given; "who" and "what" are essential.

Data table:

- rows are the "who"; each row is an individual **case** about the "who".
- columns are the "what"; each column is a **variable** that has been measured.

Types of variables:

1. Categorical (Qualitative) – data that is best described by labels i.e. color, year in school, political preference, make of cars

   - Ordinal – data that is put into some sort of ranking (i.e. 1= Poor, 2 = Fair, 3=Good, 4=Excellent)
   - Identifiers – each case has it's own unique identifier (i.e. Social Security Number, Student ID)

2. Quantitative – numerical data (most often with units)
   i.e. height, GPA, salary

Don't label a variable as categorical or quantitative without thinking about the question you want it to answer. A categorical variable may have numbers associated with it as well, but those numbers don't necessarily make it a quantitative variable.

Population: an entire group of people, objects, items of interest, in which we are interested.

Sample: a subset of the population

Statistics is about using a sample to get a better idea of what the population really looks like.