

## Chapter 2 Notes (Data)

- Recall we defined **data** as numbers or characteristics with some specific meaning or context. We also discussed the importance of providing context for data to make it clear what the data represent.
- Via Blackboard, you provided data on several **variables** (defined shortly), such as:
  1. Class (Freshmen, Sophomore, Junior, Senior, other)
  2. Major (business, biology, many others)
  3. Gender
  4. Longest distance traveled on foot
  5. Number of text messages sent yesterday

### The Five W's, and How

- As suggested in the text, whenever you are working with data, you should play the role of a news reporter, and try to identify the “*Who, What, Where, When, and Why*” of the collected data, and often most importantly: *How?*
- Mentally answering these questions can help identify aspects of the data which are missing or not reported. Key information is often omitted from a description of data, making it more difficult to analyze.

Example: In the student information data, identify the following:

1. Who?
2. What?
3. Where?
4. When?
5. Why?
6. How?

### Some Terminology Defined

- The “Who” in a data set are called the **observational units** or **cases** of the data.  
Example: If we record the undergraduate enrollment number for each U.S. university, the universities are the observational units or cases.
- When working with people, the cases are often called **subjects**, **respondents**, or **participants**.
- A **variable** is a characteristic recorded about each case or observational unit. The value of a variable *varies* from case to case (the **What’s**). Typically, variables will be assigned a number or category.
- Data then are **values** taken on by some variable. Name some variables considered in the class data.

- There are two basic types of variables (or data) encountered:
  1. **Categorical** (also sometimes called qualitative or nominal)  
Categorical variables record into which of several categories a case falls. Examples?

A categorical variable with exactly two categories (such as sex) is said to be a **binary** variable.

2. **Quantitative** (also sometimes called measurement or scale)  
Quantitative variables take on numerical values for which arithmetic comparisons make sense. Examples?

To better understand fire ignitions and burn intensity, researchers randomly select trees in the Lubrecht Experimental Forest, take some needles, and record the following:

Variable	Type
DBH (Diameter Breast Height) of the tree	
Time to ignition (in a burn chamber)	
Moisture content before ignition	
Mountain Pine Bark Beetle?	
Fire Danger Rating ( 1=Low, 2=Moderate, 3=High, 4=Very High)	

- In this last example, although a *number* is assigned to the fire danger rating variable, one must be cautious in viewing this as a quantitative variable. If we do view it as quantitative, what would this mean?
- A variable which is categorical, but has a natural ordering (such as the fire danger variable above) is known as an **ordinal** variable. We will not consider methods for ordinal variables in this course.
- One more definition: A **statistic** is a numerical *summary* of data. Typically, information is collected on the observational units (such as students in this class), and summarized by statistics such as the percentage of this class who prefer boating (favorite sport).
- Summaries of data can be verbal, visual, and numerical. Chapters 3-7 focus on techniques for summarizing both quantitative and categorical data.

Some Examples to Put it All Together: For the following example, identify the who, what, where, when, why, & how of the data, the observational units, and the variable(s) being measured. For each identified variable, give its type (quantitative or categorical) and units if applicable.

Example 1: *Consumer Reports* published an article evaluating refrigerators. It listed 41 models, giving the brand, cost, size (cu ft), type (such as freezer top), estimated annual energy cost, and overall rating (good, excellent, etc.).

Who?

What?

Where?

When?

Why?

How?

Observational Units?

Variable(s) and Type(s)?

Example 2: People who get lost in the desert, mountains, or woods often seem to wander in circles rather than walk in straight lines. To see whether people naturally walk in circles in the absence of visual clues, researcher Andrea Axtell tested 32 people on a football field. One at a time, they stood at the center of one goal line, were blindfolded, and then tried to walk to the other goal line. She recorded each individual's sex, height, handedness, the number of yards each was able to walk before going out of bounds, and whether each wandered off course to the right or to the left. No one made it to the far end of the field without crossing one of the sidelines. [Stats No. 39, Winter 2004]

Who?

What?

Where?

When?

Why?

How?

Observational Units?

Variable(s) and Type(s)?