

Name: _____

Date: _____ Pd: _____

Chapter 2: Data

What are data?

Data are values along with their context. Data can be numbers or labels.

In order to determine the context of data, consider the “W’s”

- **Who** – *the cases (about whom the data was collected). People are referred to as respondents, subjects, or participants, while objects are referred to as experimental units.*
- **What (and in what units)** – *the variables recorded about each individual.*
- **When** – *when the data was collected.*
- **Where** – *where the data was collected.*
- **Why** – *why the data was collected. This can determine whether a variable is treated as categorical or quantitative.*
- **How** – *how the data was collected.*

There are two major ways to treat data: *categorical* and *quantitative*.

- A *categorical variable* names categories and is used to answer questions about how cases fall into those categories. A categorical variable may be comprised of word labels, or it may use numbers as labels.
- A *quantitative variable* is used to answer questions about the quantity of what is being measured. A quantitative variable is comprised of numeric values.

What is a statistic? *A statistic is a numerical summary of data.*

17, 21, 44, 76

Are the numbers listed above data? Data must have *context* to be meaningful. The numbers listed above could be test scores, ages of a group of golfers, or the uniform numbers of the starting backfield on the football team. Without *context*, data cannot be interpreted.

Suppose a Consumer Reports article (published in June 2005) on energy bars gave the brand name, flavor, price, number of calories, and grams of protein and fat. Identify the following:

- **Who:** *energy bars*
- **What:** *brand, flavor, price, calories, protein, fat*
- **When:** *not specified*
- **Where:** *not specified*
- **How:** *not specified (nutrition label? laboratory testing?)*
- **Why:** *to inform potential consumers*

- **Categorical variables:** *brand, flavor*
- **Quantitative variables (with units):** *price (US\$), number of calories (calories), protein (grams), fat (grams)*

A report on the Boston Marathon listed each runner's gender, county, age, and time. Identify the following:

- **Who:** *Boston Marathon runners*
- **What:** *gender, county, age, time*
- **When:** *not specified*
- **Where:** *Boston*
- **How:** *not specified (registration information?)*
- **Why:** *race result reporting*
- **Categorical variables:** *gender, country*
- **Quantitative variables (with units):** *age (years), time (hours, minutes, seconds)*