# Chapter 1

# Representing Data

## Qualitative Data

### Qualitative and Quantitative Data

Data is all around us, and informs much of how our lives run. Let's focus on two specific types of data, **qualitative data** and **qualitative data**.

**Definition 1.1.1** (Qualitative/Quantitative Data)

**Qualitative data** is data which describes qualities or characteristics. Sometimes qualitative data is called **categorical data**.
**Quantitative data** is data which is represented using numbers.

**Example 1.1.2.** Classify the following data from members of our class as qualitative or quantitative. Give a brief explanation for the answer.

(a) Color of hair

(b) Political affiliation

(c) Height

(d) Classification by credit hours

(e) Number of hours taken this semester

**Example 1.1.3.** Come up with three sets of quantitative data and three sets of qualitative data. These should be different than the previous example!

## Representing Qualitative Data

When data is collected, we need a way to communicate the information effectively. One way of communicating qualitative information is by using a **frequency distribution**.

**Definition 1.1.4** (Frequency Distribution)

A **frequency distribution** is a two-column table which collects the number of times a category appears in a set of data.

**Example 1.1.5.** A store employee is doing an inventory of items near the register. They found 51 bags of chips, 62 candy bars, 47 bags of candy, 121 packs of gum, 16 bags of nuts, 20 soft drinks, and 20 bottles of water. Create a frequency table to display the result of the inventory.

**Example 1.1.6.** Find out the state (or country) of birth of the portion of the class your instructor specifies. Once you've collected the data, create a frequency table for the data.

We can also graphically represent qualitative data. This is often done using a **bar graph**.

**Definition 1.1.7** (Bar Graph)

A **bar graph** is a graph that represents category frequencies with a bar whose length is equal to the frequency.

**Example 1.1.8.** Create a bar graph for the data from Example 1.1.2. For convenience, here is the data again: A store employee is doing an inventory of items near the register. They found 51 bags of chips, 62 candy bars, 47 bags of candy, 121 packs of gum, 16 bags of nuts, 20 soft drinks, and 20 bottles of water.

### Bar Graph Features

Bar graphs should have the following features:

- Vertical axis with label
- Horizontal axis with labels
- A consistent scale on the vertical axis
- A descriptive title

### Note

Bar graphs can be oriented vertically (like we just did) or horizontally. If they are oriented horizontally, the vertical and horizontal information flips, but all other features will remain the same.

**Example 1.1.9.** Now create a bar graph for the data you collected about home states/countries. Be sure to include the features listed above!

A special type of bar graph is called a **Pareto chart**.

**Definition 1.1.10** (Pareto Chart)

A **Pareto chart** is a bar graph whose bars are ordered from highest to lowest frequency.

**Example 1.1.11.** A recent Math 1473 class had 15 freshmen, 3 sophomores, 10 juniors, and 8 seniors. Create the Pareto chart for the data.

Sometimes it is helpful to more than one qualtitative variable on a graph. This can be accomplished by creating a **side-by-side bar graph** or a **stacked bar graph**.

**Definition 1.1.12** (Side-by-Side Bar Graph)

A **side-by-side bar graph** is a bar graph which shows two or more bars next to each other

**Example 1.1.13.** The 2020 US Census gave the following data about racial identities in the top three most populous cities in Oklahoma:

| Race | Oklahoma City | Tulsa | Norman |
|---|---|---|---|
| **White only** | 364,706 | 214,012 | 89,283 |
| **Black/African American only** | 95,634 | 61,526 | 6,398 |
| **Asian only** | 31,510 | 14,352 | 5,083 |

Create a side-by-side bar graph to display the data.

Note that in this example, we needed to include a **legend** in order to identify what each bar was telling us.

**Definition 1.1.14** (Stacked Bar Graph)

A **stacked bar graph** is a bar graph in which bars are stacked on top of each other to show relative size.

**Example 1.1.15.** Use a stacked bar graph to display the Census data from Example 1.1.13. Here it is again:

The 2020 US Census gave the following data about racial identities in the top three most populous cities in Oklahoma:

| Race | Oklahoma City | Tulsa | Norman |
|---|---|---|---|
| **White only** | 364,706 | 214,012 | 89,283 |
| **Black/African American only** | 95,634 | 61,526 | 6,398 |
| **Asian only** | 31,510 | 14,352 | 5,083 |

**Question 1.1.16** What are some pros and cons of using a side-by-side bar graph or a stacked bar graph for this data?

Another way of visualizing the data is by using a **pie chart**.

**Definition 1.1.17** (Pie Chart)

A **pie chart** is a circle with wedges for each category whose size corresponds to the relative frequency of the category.

**Example 1.1.18.** Data for the 2022 Democratic Primary for US Senator in Oklahoma is given below:

| **Candidate** | Bollinger | Wade | Horn | Azma | Glenn | Baker |
|---|---|---|---|---|---|---|
| **Votes** | 23,367 | 19,986 | 60,691 | 11,478 | 21,198 | 22,467 |

(a) Create the relative frequency distribution for the data set

(b) Use the relative frequency distribution to create a pie chart for the data.

**Example 1.1.19.**

(a) Take a poll to determine the make of cars that your classmates drive (if they drive). Use the poll data to create a frequency table.

(b) Create the relative frequency table by dividing each frequency by the total responses. Write the relative frequency as a percentage.

(c) Use the relative frequency table to create a pie chart.