10.1: Graphing Data

Types of data:

Qualitative data: Values are categories or labels. A qualitative variable is also known as a categorical variable or a nominal variable.

Quantitative data: Values are numbers that represent the amount of some characteristic. ~~A quantitative variable is also known as a quantitative variable.~~

Examples of qualitative data:

Name
Student ID number
School district
Race
1st Gen College Student status

Examples of quantitative data:

Age
# of hours enrolled
date of 1st enrollment
TSI test scores
SAT

Data is easier to interpret if it is organized into a visual display.
Useful displays for qualitative data include tables, bar graphs, and pie charts.
Useful displays for ~~qualitative~~ quantitative data include tables, line graphs, and histograms.

Organizing qualitative data:

Bar charts (bar graphs):

A *bar graph* is a visual display in which the category names are along the vertical or horizontal axis, and the frequencies (or relative frequencies) are on the other axis.

Note: Category name = value of the qualitative variable

In a *side-by-side bar graph*, the relative frequencies for values of the nominal variable are simultaneously displayed for two or more subgroups (or for two or more years, or for two or more values of some other variable).

Pie charts:

A pie chart is a circle divided into sectors, in which each sector represents a category. For each category, the relative frequency is equal to the ratio of the sector area to the total circle area. (In other words, if one-third of the observations are in a given category, then one-third of the circle area will be in the sector corresponding to that category.)

Example 1: Create a table, bar graph, and pie chart representing the makes of the cars driven by family members of everyone in the class.

Car make	Frequency		Relative Frequency
Honda	III	3	3/36
Dodge	IIII	4	4/36 ≈ 0.111
Chevy	IIII IIII	10	10/36 ≈ 0.278
Ford	IIII IIII	9	9/36 ≈ 0.25
Toyota	IIII I	6	6/36 ≈ 0.167
Nissan	II	2	2/36 ≈ 0.056
Mitsubishi	I	1	1/36 ≈ 0.028
VW	I	1	1/36 ≈ 0.028

Frequency distributions:

A *frequency distribution* lists each category and the number of data points (occurrences) corresponding to that category. When arranged in the form of table, this is called a *frequency table*.

Tables often include the *relative frequency*.

The *relative frequency* of a category is the proportion (or percentage) of the total observations that fall within that category.

$$\text{Relative frequency} = \frac{\text{Frequency}}{\text{Sum of all frequencies}}$$

A *relative frequency distribution* lists the categories along with their relative frequencies. For a probability experiment, the relative frequencies correspond to the probabilities in the probability distribution.

Example 2: Create a relative frequency distribution for the previous example.

Example 3: Use Excel to create a frequency distribution, bar chart and pie chart for the data below.

Automaker	Units Sold in USA (2017)
General Motors	3002241
Ford Motor Company	2575200
Toyota Motor Corporation	2434515
FCA/Chrysler Group	2073073
Nissan Motor Company	1697150
Honda Motor Company	1641429
Hyundai-Kia	1275223
Subaru	647956
Volkswagen Group*	625068
Daimler	375311
BMW-Mini	354110
Mazda	289470
Jaguar-Land Rover	114333
Volvo	81507
Tesla***	55120

Source:
http://www.goodcarbadcar.net/2018/01/u-s-auto-sales-brand-rankings-december-2017-ytd/#2

Organizing quantitative data:

For quantitative data, we often divide the data into groups by using *class intervals*. We can then create a frequency table for the classes. We use a *histogram* to represent the frequencies (and thus the associated probabilities). A histogram is just a particular type of bar graph.

Note:
In a bar graph representing qualitative data, the bars do not touch.
In a histogram representing quantitative data, the bars must touch, because the "area under the curve" is important.

Example 1: A random sample was chosen from among the employees of a large corporation. Their commute times (in hours) from home to work were determined and recorded in the table:

Commute Times							
0.3	0.7	0.2	0.5	0.7	1.2	1.1	0.6
0.6	0.2	1.1	1.1	0.9	0.2	0.4	1.0
1.2	0.9	0.8	0.4	0.6	1.1	0.7	1.2
0.5	1.3	0.7	0.6	1.1	0.8	0.4	0.8

a. Construct a frequency table showing the frequency, relative frequency, cumulative frequency, and relative cumulative frequency. Use intervals of equal width starting with 0.2–0.4 (inclusive).
b. Construct a histogram and a frequency polygon.
c. What is the probability that a person chosen at random from the sample will have a commute of an hour or less?