

Organizing and Summarizing Data

Key Definitions

- **Frequency Distribution:** This lists each category of data and how often they occur.
- **Relative Frequency:** The percent of observations within the one of the categories. This is found by dividing the frequency by the sum of all the frequencies.
- **Relative Frequency Distribution:** This lists each category of data with the relative frequency.
- **Classes:** Classes can be qualitative or quantitative. The qualitative classes use names to distinguish each class. These are typically known as categories. The quantitative classes can be either discrete or continuous. For most discrete classes, each class will be given its own number. This is true if there are only a few different numbers such as 1-8. For large discrete values or continuous values, you need to find a class width as well as upper and lower class limits. They will hold multiple numbers in the classes.
- **Upper Class Limit:** This is the largest number in the class.
- **Lower Class Limit:** This is the smallest number in the class.
- **Pie Chart:** A chart used to show the relative frequency of qualitative data.
- **Histogram:** A graph used to represent quantitative data.
- **Skewed Left:** This is a graph where the tail to the left side of the peak is longer than the tail to the right side of the peak (the highest point on the graph is on the left).
- **Skewed Right:** This is a graph where the tail to the right side of the peak in the graph is longer than the tail to the left side of the peak (the highest point on the graph is on the right).
- **Uniform:** This is a graph where the frequency is spread out evenly throughout the data collected. A uniform graph is a symmetric graph. When you split the graph down the middle, each side is a reflection of the other.
- **Bell-Shaped:** This is a graph where the highest frequency is in the middle and the left and right frequencies slope downwards creating a bell-shape. A bell-shaped graph is also a symmetric graph. When you split the graph down the middle, each side is a reflection of the other.

Relative Frequency Distribution

- **How to find the Relative Frequency:** To find the relative frequency, you take the frequency (number of responses) of the category and divide it by the sum of all the frequencies (the entire sample).
- **The Distribution for Discrete or Qualitative Data:** The distribution is shown in a table format. Each column of the table refers to the class, frequency, and relative frequency. The rows in the first column will be classes. The classes could be either numbers or titles (such as a black car). These individual numbers or titles are called the categories. The frequency refers to how many time the

number or title appeared in the collection of data. The relative frequency is the percentage of how often the category appeared in the data.

- **Example of a Relative Frequency Distribution for Discrete or Qualitative Data:**

A company surveyed 30 people in the morning about how they felt about the service offered in the store. The following is the relative frequency distribution made based on their findings.

Categories	Frequency	Relative Frequency
Very Good	8	$\frac{8}{30} = .2667 = 26.67\%$
Good	10	$\frac{10}{30} = 0.3333 = 33.33\%$
No Opinion	4	$\frac{4}{30} = 0.1333 = 13.33\%$
Bad	6	$\frac{6}{30} = 0.2 = 20\%$
Very Bad	2	$\frac{2}{30} = 0.0667 = 6.67\%$

- **Classes for Continuous, Quantitative Data:** Class for continuous data need to have class width. To find the class width, we first must chose the amount of classes to have in the relative frequency distribution. When we have determined that, you use the following formula:

$$\text{Class Width} \approx \frac{\text{largest data value} - \text{smallest data value}}{\text{number of classes}}$$

If you are given a decimal, round up to the nearest whole number. To find the class, you take all the data between the lower class limit and the upper class limit. To find the upper class limit, you take the lower class limit and add the class width to it. From here if you have decimals, you lower the added number by whatever the decimal place is so classes don't overlap. If you have whole numbers, you subtract 1 from it so the classes don't overlap. From there, you have created the classes.

- **The Distribution for Continuous, Quantitative Data:** The distribution is displayed in a table. Each column of the table refers to the class, frequency, and relative frequency. The rows in the first column will be classes (see above). The frequency refers to how many time the number or response appeared in the collection of data. The relative frequency is the percentage of how often the category appeared in the data.
- **Example of a Relative Frequency Distribution for Continuous, Quantitative Data:**
A fast food location looked at the average amount of money customers spent at their location in a half-hour period. They collected 20 different pieces of data. The numbers they collected were \$4.50, \$1.25, \$3.75, \$2.50, \$3.50, \$8.50, \$6.25, \$7.75, \$9.50, \$11.75, \$2.50, \$10.00, \$11.00, \$5.00, \$8.25, \$5.00, \$9.75, \$2.50, \$3.75, and \$6.50. The owner wants 4 classes to be used when making this distribution. Find the relative frequency distribution.
First, we find the class width. The largest value is \$11.75 and the smallest value is \$1.25. We then plug it into our formula:

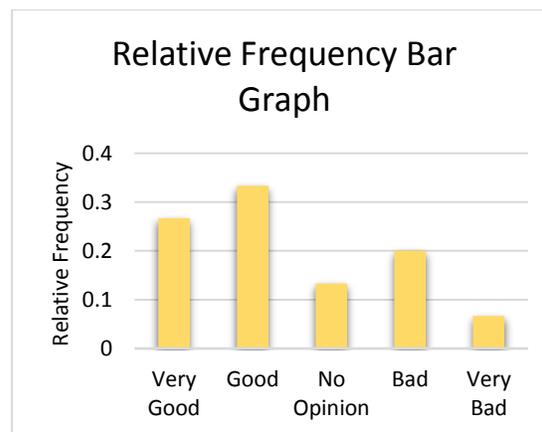
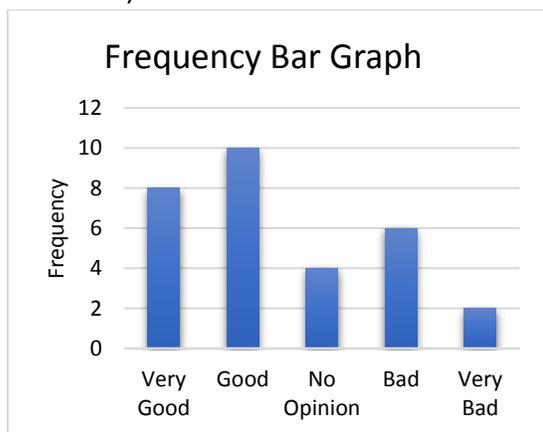
$$\text{Class width} \approx \frac{11.75 - 1.25}{4} \approx 2.625 \approx 3$$

To find our first lower class limit, we look at the smallest value which is \$1.25. We want our lower class limit to be a whole number. So, we round ours down to \$1. Now to find the first upper class limit, we take our lower class limit, \$1, and add the class width to it. We originally get $1 + 3 = 4$, but we cannot have overlap. Therefore, our first upper class limit is \$3.99. We do this continuously until we have made each class. We, then, find the frequency and relative frequency for each class. The following is the relative frequency distribution made based on their findings.

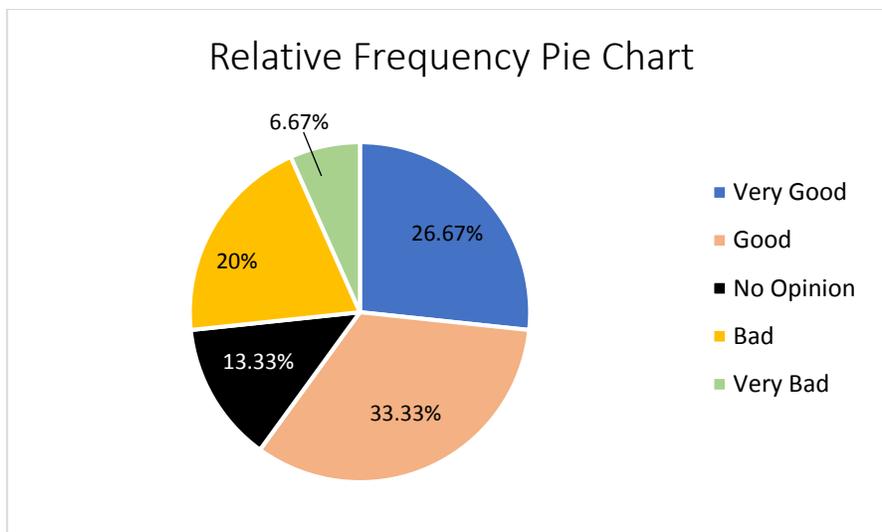
Classes	Frequency	Relative Frequency
\$1.00 - \$3.99	7	$\frac{7}{20} = .35 = 35\%$
\$4.00 - \$6.99	5	$\frac{5}{20} = .25 = 25\%$
\$7.00 - \$9.99	5	$\frac{5}{20} = .25 = 25\%$
\$10.00 - \$12.99	3	$\frac{3}{20} = .15 = 15\%$

Common Graphs

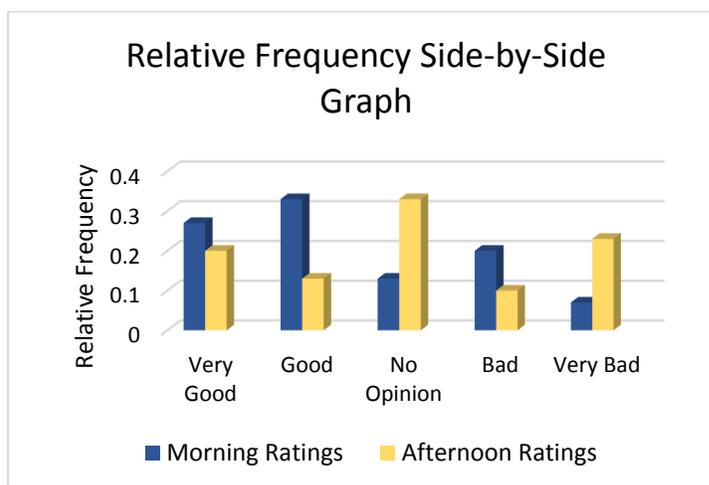
- Bar Graph:** The bar graph can be constructed from the relative frequency distribution in two ways. The first way it can be constructed is by using the frequency to determine the height of each bar and the categories to distinguish each bar (see the figure below on the left using the data from the qualitative, relative frequency distribution). The second way it can be constructed is by using the relative frequency distribution to determine the height of each bar and the categories to distinguish each bar (see the figure below on the right using the data from the qualitative, relative frequency distribution).



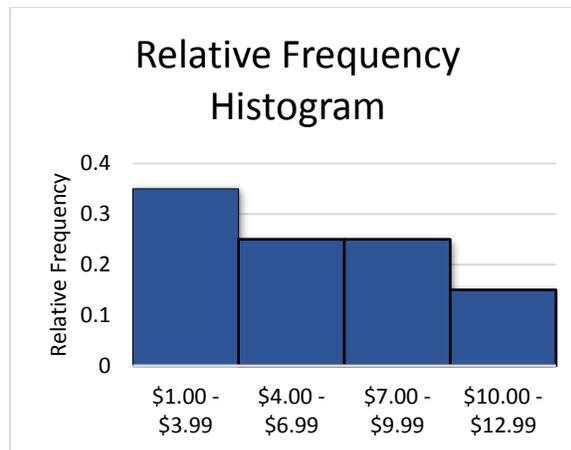
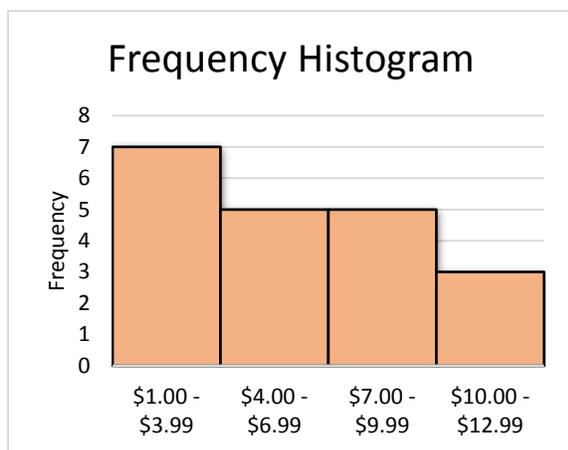
- Pie Chart:** A pie chart is constructed from a relative frequency distribution. It splits up the circle into the different categories of the relative frequency distribution by each categories' relative frequency. The pie chart gives us a visual of the popularity of each category among the data. An example of a pie chart is below using the data from the qualitative, relative frequency distribution.



- Side-by-Side Graph:** The side-by-side graph is a visual representation of the comparisons of two different data sets that works like the bar graph. The side-by-side graph is used for relative frequency only since you can have different population or sample sizes. The height of the graph represents the relative frequency and each bar represents a different category. In the example below, we use the data from the qualitative, relative frequency distribution (represented in blue), and new data found in the afternoon where the relative frequency for each is represented in yellow down below.

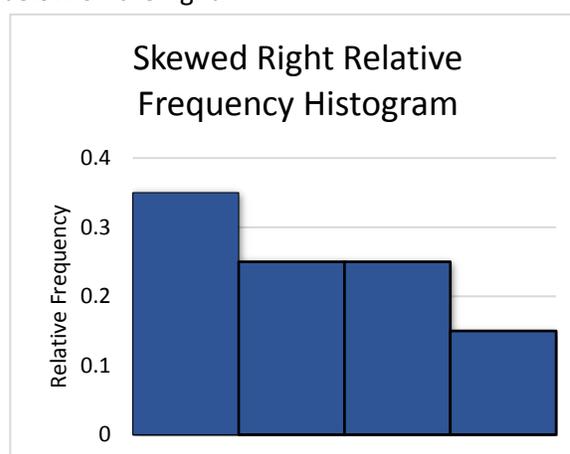
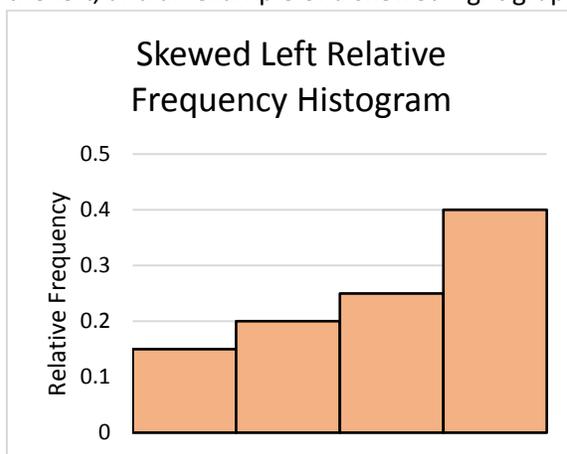


- Histogram:** A histogram is made by drawing rectangles which represent the classes. The height of the rectangles represents the frequency or relative frequency of the classes depending on whether you are making a frequency histogram or relative frequency histogram. The width of each rectangle is the same since all class widths are the same. Notice that a histogram is very similar to a bar graph, but it has one defining feature. There is no space between the bars. For a graph to be a histogram, there must be no spaces between the rectangles. Below are two examples that use the relative frequency distribution for continuous, quantitative data. You can make histograms for discrete relative frequency distributions, but they must be quantitative.

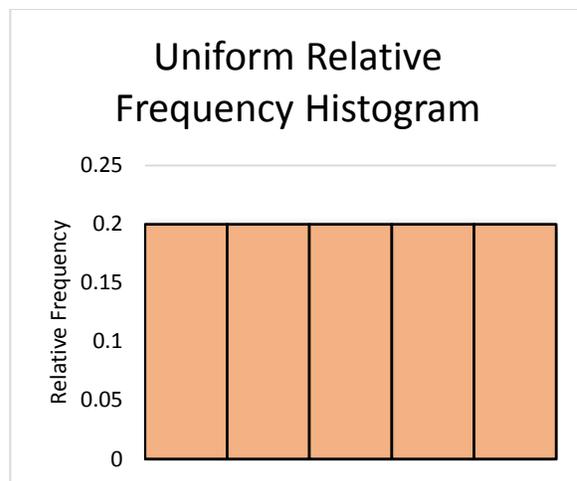
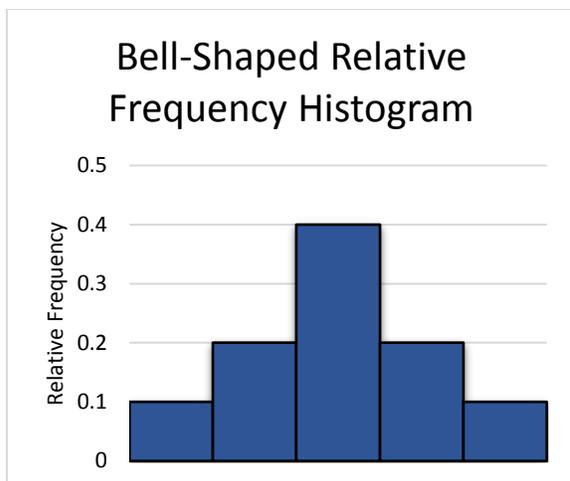


Graph Distributions

- Skewed Left and Skewed Right:** If the tail to the left side of the highest point on the graph is considerably longer than the tail to the right side of the highest point, the graph is skewed left. If the tail to the right side of the highest point on the graph is considerably longer than the tail to the left side of the highest point, the graph is skewed right. An example of a skewed left graph is below on the left, and an example of a skewed right graph is below on the right.



- Uniform and Bell-Shaped:** A bell-shaped graph will have the highest point in the middle of the graph and slope downwards on both sides. This creates a bell shape if you were to draw a line on the graph. An example of a bell-shaped graph is below to the left. A uniform graph has the same frequencies across each category or class. An example of a uniform graph is below to the right.



Common Plots

- **Stem-and-Leaf Plot:** This is a way to represent quantitative data. For the stem-and-leaf plot, we create a “stem” and “leaves”. To create the stem, we use the digits to the left of the rightmost digit in each number. To create a leaf, we use the rightmost digit of each number. From there, we organize the numbers into having one stem for the digits to the left and having multiple leaves.
- **Example of Steam-and-Leaf:**
A clothing store collects offered a survey to all their clients. In that survey, they collected the data of the different ages that shop at their store. This is the data they collected:

31, 45, 35, 37, 41, 28, 48, 25, 34, 31, 40, 29, 33, 34, and 38.

To create a stem and leaf plot, we first must separate each number into a stem and a leaf. For 31, the digits to the left of the rightmost digit is 3 and the rightmost digit is 1. Therefore, the stem is 3 and the lead is 1. Now, we do this for the next number which is 45. The stem is 4 and the leaf is 5. We continue this until we have separated each number into a stem and a leaf. Next we organize them in the plot. We do this by making a line straight between stems and leaves. The stems go on the left side of the line and we do not repeat any stems that are the same. The leaves go on the right side of the line (aligned with the correct stem), and we put each leaf in ascending numerical order. As we can see in our data, the number 34 appears twice. On our stem-and-leaf plot, we have 3 once on the left side but 4 twice. This is because we do not repeat the stem, but we must repeat the leaves if there are more than one of them.

```

2 | 5 8 9
3 | 1 1 3 4 4 5 7 8
4 | 0 1 5 8

```

We can tell by looking at the stem and leaf that a majority of the people who shop at the store are in their 30s. We can also tell each age by looking at the stem and then the leaf.

- **Time-Series Plot:** A time-series plot is used if you have collected data over different time periods. The plot is made by plotting the time along the x-axis (the horizontal axis) and the other data on the y-axis (vertical axis). Then, you connect the dots with line segments. We only use this plot for frequency.
- **Example of a Time-Series Plot:** A small boutique keeps track of how many products have been sold every month for a year to see when they sell the most items. They have recorded these results:

Month	Items Sold
January	400
February	325
March	700
April	550
May	750
June	650
July	450
August	800
September	650
October	400
November	725
December	900

We now need to plot this data on a graph. For the first entry, we go across to where the horizontal axis says January and then up until we reach 400. We plot our point there. Then, we do this for each point until all points have been plotted. Next, we draw a line between each point. Afterwards, the time-series plot will look like the one below.

