



2.2 Presenting Quantitative Data Graphically

2.2 Presenting Quantitative Data Graphically

Upon completion of this section, you should be able to

- » Create and interpret Histograms
- » Create and interpret frequency polygons
- » Create and interpret basic stem and leaf displays
- » Identify common graphical mistakes with graphical representation of quantitative data

Presenting Quantitative Data Graphically

Quantitative, or numerical, data can be expressed as a quantity or a measurement. Height, weight, response time, subjective rating of pain, temperature, and score on an exam are all examples of quantitative data values. Quantitative data is distinguished from categorical (sometimes called qualitative) data such as favorite color, religion, city of birth, and favorite sport in which there is no ordering or measuring involved.

We will start by seeing what happens if we use the same approach to quantitative data as we did with categorical (qualitative) data. In our first example we will create a frequency table as we did with categorical data in the last section.

Example 1

A teacher records scores on a 20-point quiz for the 30 students in their class. The scores are:

19, 20, 18, 18, 17, 18, 19, 17, 20, 18, 20, 16, 20, 15, 17, 12, 18, 19, 18, 19, 17, 20, 18, 16, 15, 18, 20, 5, 0, 0

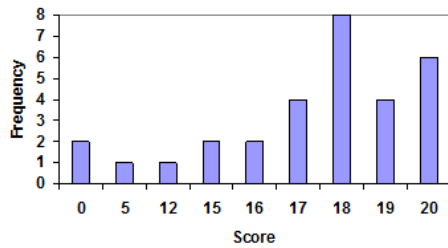
These scores could be summarized into a frequency table by grouping like values. Use this frequency table to also construct a bar graph.

Quiz Scores

Score	Frequency
0	2

Score	Frequency
5	1
12	1
15	2
16	2
17	4
18	8
19	4
20	6

It would be possible to create a standard bar chart, like we did for categorical data:

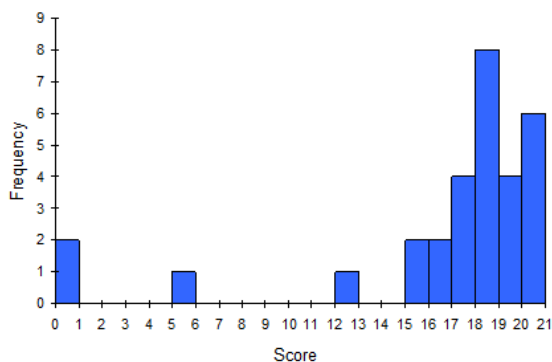


However, since the scores are numerical values, this chart doesn't really make sense; the first and second bars are five values apart, while the later bars are only one value apart. It is not appropriate to create a bar graph in this case.

[Video Explanation of why bar graphs may not be appropriate with quantitative data.](#) (cc)

For the above example it would be better to treat the horizontal axis as a number line. This way when a bar is added at the appropriate location we can see the gap that exists between values. Not including the gap hides the information from the viewer and is misleading. By adding the structure on the horizontal axis (number line) we are using what is called a **histogram**.

For the values in Example 1 above, a histogram would look like:

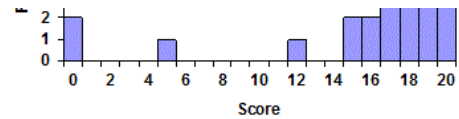


Notice that in the histogram, a bar represents values on the horizontal axis from that on the left hand-side of the bar up to, but not including, the value on the right hand side of the bar. Some people choose to have bars start at $\frac{1}{2}$ values to avoid this ambiguity.

Unfortunately, not a lot of common software packages can correctly graph a histogram directly from data provided. The programs all



require you to give input on the number of class intervals and where to start the first interval. About the best you can do in Excel or Word is a bar (column) graph with no gap between the bars.



If we have a large number of widely varying data values, creating a frequency table that lists every possible value as a category would lead to an exceptionally long frequency table, and probably would not reveal any patterns. For this reason, it is common with quantitative data to group data into **class intervals**.

Histograms

A **histogram** is like a bar graph, but where the horizontal axis is a number line. Each bar represents a **class interval** or range of numeric values for the data observations to fall in. Each data value fits into a single class (no overlap).

In general, we define class intervals so that:

- » Each interval is equal in width.
- » There should be between 5 and 20 classes, typically, depending upon the number of data we're working with.

How to Construct a Histogram

1. Construct a frequency table based on the class intervals
 - a. Determine the number of class intervals to be used. Typically a number between 5 and 20 is appropriate, but it can depend on the data and some trial and error may be involved.
 - b. Calculate the class width by taking the range of the data and dividing by the number of class intervals you identified above

$$\text{Class Width} = \frac{\text{Max Data Value} - \text{Min Data Value}}{\text{Number of Class Intervals}}$$

At this point you may need to round the class width value. For a general rule of thumb you should round up based on the precision of the data values (if the data has two decimal places, then round up to two decimal places).

- c. Construct a table with three columns: Lower class limit, upper class limit, and frequency.

Lower Class Limit	Upper Class Limit	Frequency

- d. Next add Min data value into the first entry for the Lower Class Limit column.
- e. Fill in the rest of the lower class limits by adding the class width to each previous lower class limit value. Continue to do this until you reach the number of class intervals identified in step 1 or the next lower class limit you would enter is larger than the maximum data value.
- f. Fill in the Upper Class limits. The upper class limit is the biggest number below the next lower class limit with the same accuracy as the class width.
- g. Find the frequency for each class interval defined by the lower and upper class limits. If done correctly the total number in the frequency column will add up to the total number of data values.

2. Graph the histogram

- a. First draw the coordinate system ensuring the y-axis starts at 0 and the goes to at least the maximum value in the data set. The x-axis can be truncated to start at the first lower class limit

maximum value in the data set. The x-axis can be truncated to start at the first lower class limit and extends out to at least the last upper class limit value. It is helpful to have just the tick marks representing the lower class limits on the x-axis.

- For each class interval draw a bar going up to the frequency given in the table created. The bar for each class interval starts at the lower class limit and ends at the next lower class limit leaving no gaps between bars unless there is a class interval with a frequency of 0.
- Add labels for the axis and relabel bars if needed (a common approach is to use the midpoint for the class intervals to be used as the label).

When using software the approach may vary with the information limited above. Typically you will need to predefine the lower class limits or upper class limits. The counting can then be done through the software as well as the construction of the graph. The examples below will go through the method as identified above, but the output of the graph from the histogram will be done through Google Sheets or Excel.

Example 2

We will revisit our data from exercise 1 where a teacher records scores on a 20-point quiz for the 30 students in their class. The scores are:

19, 20, 18, 18, 17, 18, 19, 17, 20, 18, 20, 16, 20, 15, 17, 12, 18, 19, 18, 19, 17, 20, 18, 16, 15, 18, 20, 5, 0, 0

Create a histogram for the data using 6, 8, and then 10 class intervals.

We will go into detail for constructing the frequency table for the 6 class intervals, but only provide short summary for the 8 and 10.

First step is to create the frequency table using 6 class intervals. To get started we find the class width. The Max value in the data is 20 and the Min value is 0. $\text{Class Width} = \frac{20-0}{6} \approx 3.33$. We will round this up to 4 (since all our data are integers).

Next construct the frequency table starting with the Min value of 0 for the first lower class limit:

Lower Class Limit	Upper Class Limit	Frequency
0		

We now fill in the rest of the lower class limits by adding 4 (the class width) to the values.

Lower Class Limit	Upper Class Limit	Frequency
0		
4		
8		
12		
16		
20		

Since our data is expressed as an integer we will fill in the upper class limits by using the largest integer below the next lower class limit. The first upper class limit would be 3 since it is the largest value below the next lower class limit found

Stem-and-Leaf Plot

A stem-and-Leaf plot is a visual table that splits the data into a "stem" (first digit(s) of a data value) and a "leaf" (typically the last digit of the data value). The stems are written in a vertical column from smallest to largest. A vertical bar is drawn to the right of the stem. The leaves are then entered to the right of the stem in an increasing order.

Directions:

- Each of the data values will have two parts: left = "stem," right = "leaf"
 - » Example for a data value of 45, "4" would be the stem and "5" will be the leaf.
- Align all stems in vertical column in increasing order with a vertical line to their right (remove any duplicate stems and add any stem that appears to be missing in that ordered list - we want to see gaps in the data distribution).
- Place all leaves with same stem on the same row as the stem. Rearrange in an increasing order and keep equal spacing between leaves for all rows so that values align in a column.
 - » It is important to align the leaves in columns so that we may actually compare the length of each row directly and not worry about data points be clumped together to make two rows appear the same length yet actually have a different amount of data points.
- Label the plot.

Programs like Excel and Google Sheets do not have an automatic way to create a stem-and-leaf plot without typically using a third party solution. Thankfully it is not difficult to style the spreadsheet in either program to look like a stem-and-leaf plot. You can also do this in a Word or Google Doc document as well by using a table and only showing particular borders. In the example below we will go through the process of creating a stem-and-leaf plot using our quiz score data from earlier.

Example 4

A teacher records scores on a 20-point quiz for the 30 students in their class. The scores are:

19, 20, 18, 18, 17, 18, 19, 17, 20, 18, 20, 16, 20, 15, 17, 12, 18, 19, 18, 19, 17, 20, 18, 16, 15, 18, 20, 5, 0, 0

Create a stem-and-leaf plot of the quiz scores.

We will start by identifying how to break apart the data. Since that data starts at 0 and has a max of 20 and are all integers it would make sense to break apart the stems as the value in the tens place. This leaves the leaf as the value in the ones place. For instance 17 would create a stem of "1" and a leaf of "7" while the number 5 would give a stem of "0" and a leaf of "5".

The Stems are 0, 1, and 2. Start by putting those values into a vertical column and add a line to the right of them as shown below:

Quiz Scores

Stem	Leaf
0	
1	
2	

Now that we put in the stems it is time to place all the data in the table by adding the leaves in an increasing order. Make sure to repeat values as the length of each stem row is like a histogram bar showing the frequency.

Quiz Scores

Stem	Leaf
0	0 0 5

0	0 0 5
Stem	Leaf
1	2 5 5 6 6 7 7 7 7 8 8 8 8 8 8 8 8 9 9 9
2	0 0 0 0 0 0

In the next example pay attention to what we do for a missing stem. The missing stem helps show what is happening in the data by identifying the gap that exists between values.

Example 5

The following is a list of student ages in a MAT 142 course from Fall of 2020 at the start of the semester: 17 18 18 18 18 18 18 18 18 18 18 19 19 19 19 19 20 20 20 22 28 31 55

Create a stem-and-leaf plot of the ages.

We can naturally break up each age into a stem and leaf by using the tens place for the stem and the ones place as the leaf. Start with creating a column for stems. Notice how we include the missing "4" stem between 3 and 5.

Student Ages in MAT 142

Stem	Leaf
------	------

1	
2	
3	
4	
5	

Next add in the leaves behind each stem. Again make sure the leaves align in columns of equal spacing.

Student Ages in MAT 142

Stem	Leaf
1	7 8 8 8 8 8 8 8 8 8 9 9 9 9 9
2	0 0 0 2 8
3	1
4	
5	5

Try it Now 2

The following is some data from past Oscar-winning Best Actors (males).

32 37 36 32 51 53 33 61 35 45 76 37 42
40 32 60 38 56 48 48 62 43 42 44 41 56
39 46 31 47 46 40 36 39 43 60 55 40 45

80	50	50	90	70	70	100	60	70	80	70	50
90	100	80	70	30	80	80	70	100	60	60	50

5. Studies are often done by pharmaceutical companies to determine the effectiveness of a treatment program. Suppose that a new AIDS antibody drug is currently under study. It is given to patients once the AIDS symptoms have revealed themselves. Of interest is the average length of time in months patients live once starting the treatment. Two researchers each follow a different set of 40 AIDS patients from the start of treatment until their deaths. The following data (in months) are collected.

Researcher 1: 3; 4; 11; 15; 16; 17; 22; 44; 37; 16; 14; 24; 25; 15; 26; 27; 33; 29; 35; 44; 13; 21; 22; 10; 12; 8; 40; 32; 26; 27; 31; 34; 29; 17; 8; 24; 18; 47; 33; 34

Researcher 2: 3; 14; 11; 5; 16; 17; 28; 41; 31; 18; 14; 14; 26; 25; 21; 22; 31; 2; 35; 44; 23; 21; 21; 16; 12; 18; 41; 22; 16; 25; 33; 34; 29; 13; 18; 24; 23; 42; 33; 29

Create a comparative histogram of the data for the researchers:

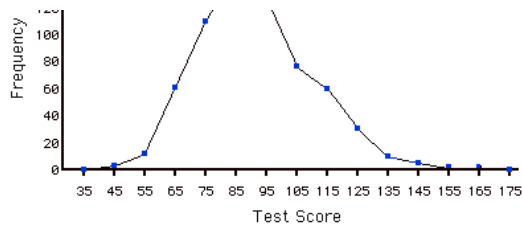
6. An experiment compared the ability of three groups of participants to remember briefly-presented chess positions. The data are shown below. The numbers represent the total number of pieces correctly remembered from **three** chess positions. Create a histogram for each of the player types using the same class intervals (start the first class interval as 20 - 29) for these three groups. What can you say about the differences between these groups from the Histograms?

Non-players	Beginners	Tournament players
22.1	32.5	40.1
22.3	37.1	45.6
26.2	39.1	51.2
29.6	40.5	56.4
30.1	41.2	56.7
31.7	45.5	58.1
33.5	51.3	71.1
38.9	52.6	74.9
39.7	55.7	75.9
43.2	55.9	80.3
43.2	57.7	85.3

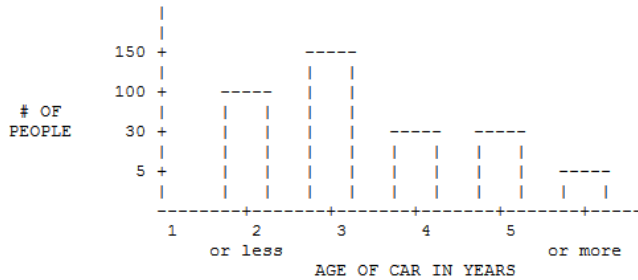
7. Using the same data from the above problem about chess players create frequency polygons for each of the player types and graph them on the same set of axis. What advantages do you see for using the frequency polygon over the histograms to compare each of the player types ability to remember chess pieces?
8. Based on the frequency polygon displayed below, the most common test grade was around what score? Explain.

150
140
120





9. The histogram shown below is misleading. Explain why.



10. A random sample of students were selected from a local college and their age was recorded. The ages for the 25 students were: 17, 18, 18, 18, 18, 19, 19, 19, 19, 19, 20, 20, 20, 21, 21, 22, 22, 23, 25, 26, 29, 30, 32, 38, 51. Create a Stem and Leaf plot of the ages.

Answers to Select Homework Exercises

1. Categorical Variables: Bar Graph; Pareto Chart; Pie Chart. Quantitative Variable: Histogram, Pie chart (sometimes), Frequency Polygon, Stem and Leaf Plot.

lower class limit. The first upper class limit would be 3 since it is the largest value below the next lower class limit found on the second row of lower class limit values. The remaining upper class limit values can be found by adding the class width to the previous upper class limit value.

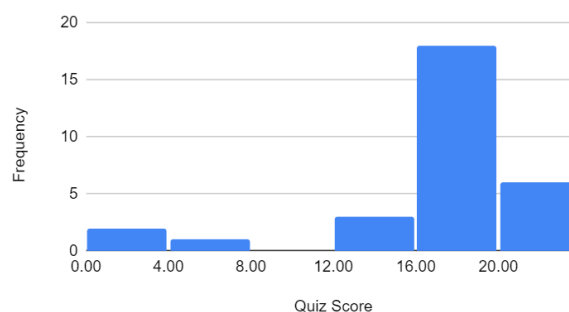
Lower Class Limit	Upper Class Limit	Frequency
0	3	
4	7	
8	11	
12	15	
16	19	
20	23	

Now the final step for constructing the table is getting the frequency. It may help to keep a tally marker and mark off each data value in a systematic way. Double check you counted all the data by adding up the frequency and see you have the same total as the number of data values you started with.

Lower Class Limit	Upper Class Limit	Frequency
0	3	2
4	7	1
8	11	0
12	15	3
16	19	18
20	23	6

To construct the Histogram start with labeling the lower class limits on the horizontal axis and going from 0 to 20 on the vertical axis (technically you can go up to 18 as a max). For each class interval you will draw a bar going up to the frequency height. Typically you want the bars to touch each other to show there is no artificial gaps in the data trends, but some programs will put in a very small gap between bars like shown below made from Google Sheets.

Quiz Scores with 6 class intervals



Quiz score with 6 class intervals: [Google Sheet for Example 2](#)

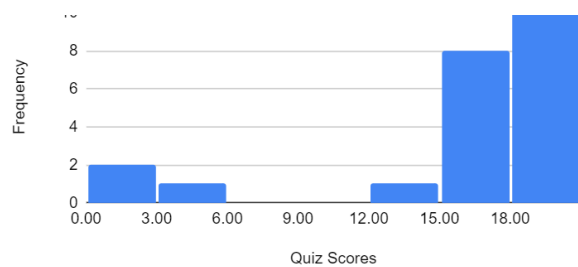
Below you will see the tables and histograms for class intervals of size 8 and 10.

Quiz Scores with 8 class intervals

Quiz Scores with 8 class intervals



Lower Class Limit	Upper Class Limit	Frequency
0	2	2
3	5	1
6	8	0
9	11	0
12	14	1
15	17	8
18	20	18

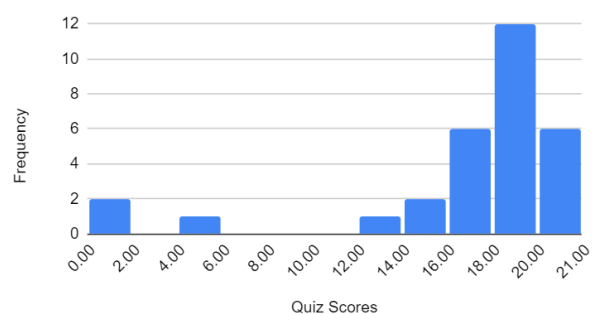


Quiz score with 8 class intervals: [Google Sheet for Example 2](#)

Quiz Scores with 10 class intervals

Lower Class Limit	Upper Class Limit	Frequency
0	1	2
2	3	0
4	5	1
6	7	0
8	9	0
10	11	0
12	13	1
14	15	2
16	17	6
18	19	12
20	21	6

Quiz Scores with 10 class intervals



It would be better to adjust the graph horizontal axis to remove the 21 to avoid confusion with the context of the application (where 20 was the maximum score), but sometimes the tool we use for the construction has limitations and edits like this may not be possible. It is best to include descriptions with any graphics to help avoid confusion (so in this case adding something about the maximum score being 20 would be helpful).

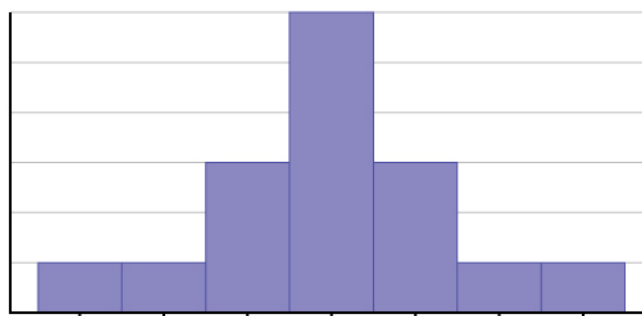
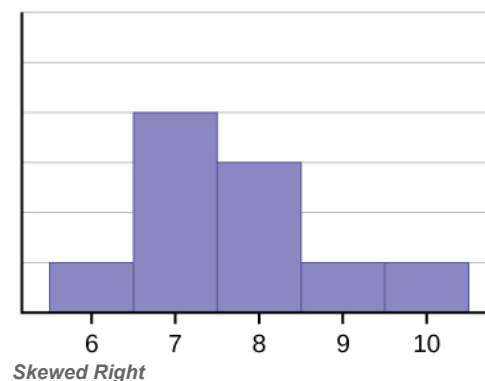
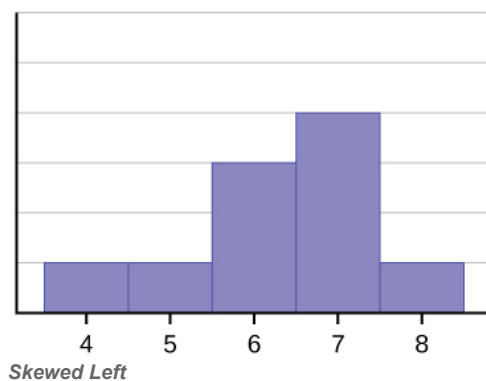
When we look at the three histograms a slightly different visual display of the data is seen. Does that mean any of them are right or wrong? No. There is no right histogram for a given data set, but there are good ways to go about constructing them. Ensuring the each class interval has an equal width, not distorting the graph through manipulation of the vertical axis, using between 5 and 20 class intervals (typically more class intervals for larger data sets, but trial and error is needed to help determine what number of classes to use).

In many software packages, you can create a graph similar to a histogram by putting the class intervals as the labels on a bar (column) chart, but you need to make sure that you include any bars with 0 frequency and close the gap between the bars.

What should we notice?

After the construction of a histogram there are some questions that should be answered by the researcher when describing the histogram:

- » What are the maximum and minimum potential values of the data set?
- » How many peaks does the histogram have (a peak can be any bar that has a shorter bar to the left and right or a bar at the left or right hand end that is taller than the adjacent bar)? Where are those locations?
- » Are there any gaps in the histogram (places where a bar has a height of 0) and where are they located?
- » Does the histogram have a general shape you can describe (uniform, symmetric, skewed left or right, bell-shaped)?



Example 3

We will revisit our data from exercise 1 and 2 where a teacher records scores on a 20-point quiz for the 30 students in their class ignoring we have the original data available to answer some questions.

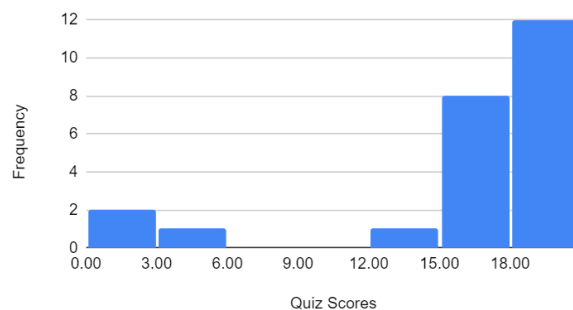
Answer the following based on the histogram constructed using 8 class intervals (also shown below):

- » What are the maximum and minimum potential values of the data set?
- » How many peaks does the histogram have (a peak can be any bar that has a shorter bar to the left and right or a bar at the left or right hand end that is taller than the adjacent bar)? Where are those locations?
- » Are there any gaps in the histogram (places where a bar has a height of 0) and where are they located?
- » Does the histogram have a general shape you can describe (uniform, symmetric, skewed left or right, bell-shaped)?

Quiz Scores with 8 class intervals

Lower Class Limit	Upper Class Limit	Frequency
0	2	2
3	5	1
6	8	0
9	11	0
12	14	1
15	17	8
18	20	18

Quiz Scores with 8 class intervals



Quiz score with 8 class intervals: [Google Sheet for Example 3](#)

From the histogram we can see that the first bar starts at 0 and has a frequency of 2. This tells us we have a potential value in the data to be 0 and that would be the smallest or minimum value. The last bar starting at 18 and knowing it ends at 20 lets us know we have a potential maximum value of 20.

There is a peak on the right hand side on the class interval from 18 to 20. The peak reaches a height of 12. Another potential minor peak happens at the lowest class interval of 0 to 2 with a height of 2.

There is a large gap in the middle of the histogram in the class intervals starting at 6 and ending at 11.

The general shape of the histogram can be described as being skewed left as the peak is on the far right and as you move left the data trails off, but another minor peak occurs at the very start with a gap between the two sides.

Now it is your turn to construct a histogram in the TryItNow below.

Try it Now 1

The total cost of textbooks for the term was collected from 36 students. Create a histogram for this data. Use 6 class

intervals. Describe the results.

Textbook Prices

\$140	\$160	\$160	\$165	\$180	\$220	\$235	\$240	\$250	\$260	\$280	\$285
\$285	\$285	\$290	\$300	\$300	\$305	\$310	\$310	\$315	\$315	\$320	\$320
\$330	\$340	\$345	\$350	\$355	\$360	\$360	\$380	\$395	\$420	\$460	\$460

- ANSWER

With 6 class intervals given to us we start by finding the class width: $\text{Class Width} = \frac{460-140}{6} \approx 53.33$. The textbook cost is given to us in integers, so we round this up to 54.

Starting with 140 for the first lower class limit we can now construct the frequency table. Add 54 to the first class lower limit and continue to add 54 to the previous lower class limit to fill out the first column.

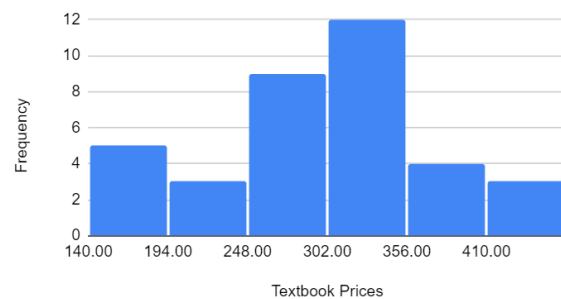
To get the upper class limit we find the highest number below the next lower class limit, so for the first lower class limit we look at the 2nd row of the lower class limits and see the value is 194, which means the first upper class limit will be 193 (since our data precision is in integers). Add the class width to the upper class limit values to fill down

that column. After those are done we count how many data values fall into each interval. The finished table is below and was used to construct the histogram.

Textbook Prices

Lower Class Limit	Upper Class Limit	Frequency
140	193	5
194	247	3
248	301	9
302	355	12
356	409	4
410	463	3

Histogram of Textbook Prices



Quiz score with 8 class intervals: [Google Sheet for TryItNow 1](#)

In the histogram and data for textbook prices we see the minimum value is \$140 with a maximum value of \$460. The histogram shows a peak that occurs between \$302 and \$356 with a frequency of 12 (out of the 36 data values).

There are no gaps shown on this histogram. The overall shape appears to be symmetric about the middle where the peak occurs, but there is a small minor peak on the left hand side between \$140 and \$194.

Stem-and-Leaf Plots

In some cases when the data set is relatively small it is easier to visualize the entire data set and distribution in a single display. We will look at one way to do that using what is called a stem and leaf plot. In a stem and leaf plot we will be able to see the distribution of the data AND we will be able to show the entire data set. The stem-and-leaf conveys the impressions that a histogram would without going through the process for creating one (it is typically just as easy to create if doing both by hand). The benefit is that a stem and leaf plot also preserves the exact data values.

Create a stem-and-leaf plot for the data. Is there anything we should be concerned about after the creation?

- ANSWER

Each value is a two digit number (if someone were to be age 8 we would write 8 as 08 to put it into the same form). The first digit is the "stem" and second digit of each number is a "leaf". We create the plot by listing all possible values of "stems" in a vertical column starting with the lowest and ending at the highest value (if there is a gap we put the value of the missing stem in the place, but will have no leafs attached to it on the right hand side). Each "leaf" is then placed to the right of its stem in an increasing order (best to start by looking for any leafs with a value of 0 and add it to the right for each zero then continue on to the ones, twos, etc.... When done you have a visualization that shows you all the data values along with their distribution (more powerful then just looking at the raw data by itself).

Male Oscar Winners

Stem	Leaf
3	1 2 2 2 3 5 6 6 7 7 8 9 9
4	0 0 0 1 2 2 3 3 4 5 5 6 6 7 8 8

5	1 3 5 6 6
6	0 0 1 2
7	6

The plot allows us to see a visual distribution of the ages - from here we can try to make sense of the age of male oscar winners. For instance we can say it appears you have a better chance of winning if you are in your 30's or 40's. The thing to watch out for in this type of graph (and histograms in general) is that you lose track of a possible variation in results over time (we do not know if the trend of the ages was changing over time or if this distribution is held constant over time).

Identify common graphical mistakes with graphical representation of quantitative data

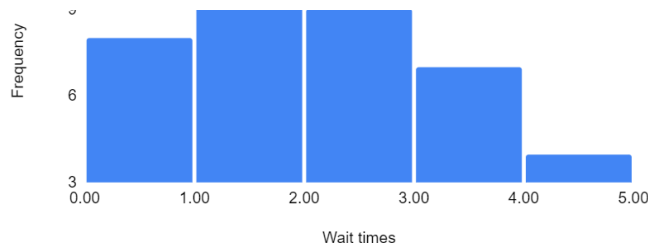
In the instructions for histogram and stem-and-leaf plot construction you were given details to help avoid distorting the data visually. We will look at some examples where an error was made in a graph for quantitative data and discuss the importance of being able to identify the error and what to watch out for when creating your own graphical display for data. Much of what we learn in the examples can be extended to graphs that we did not introduce in this section (a few examples below are also of graphs we did not review on how to make, but follow similar structures to our work from above).

Example 6

In the histogram below we have the wait times (to the nearest tenth a minute) recorded at a local fast food drive thru on a randomly selected day. The general manager is making a presentation to the regional manager about these wait times. Identify any cause for concern for how the data was presented in the histogram.

Wait Times Drive Thru

12

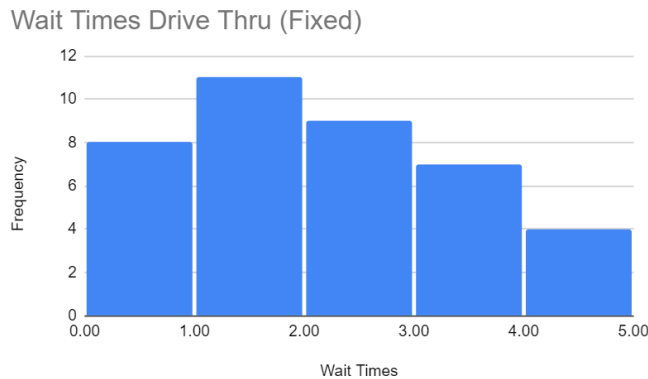


Drive Thru Wait Times [Google Sheet for Example 6](#)

The first concern we can identify is that the vertical axis begins at 3 instead of zero. Visually it would appear the wait times for 3 to 4 minutes is about half the wait times from 1 to 2 minutes, but that is not what the actual data shows. The wait times really about two thirds instead of half if you extend the vertical axis down to 0.

A minor point of concern is not having any grid lines and tick marks for the vertical axis for the height of each bar. Without some visual aid it can be challenging to read the direct height from the vertical axis on each bar. Without the tick marks it is also challenging to know which value goes to a bar height.

Below is the adjusted histogram.



Drive Thru Wait Times (Fixed) [Google Sheet for Example 6](#)

Looking at the fixed histogram it shows how the shorter wait times were not as high as previously seen when compared to the longer wait times. In this example we set up the scenario in a more nefarious way, but typically you may see those truncated graphs appear unintentionally. The problem though is that it can change the perception or feeling on a topic if shown that way. This is not isolated just to histograms as that change in the vertical axis can happen with other types of charts or graphs used to display data.

In the next example we will look at a chart that is similar to histograms, but instead uses a frequency polygon. A **frequency polygon** starts out like a histogram, but instead of drawing a bar, a point is placed in the midpoint of each interval at height equal to the frequency. Typically the points are connected with straight lines to emphasize the distribution of the data. When the horizontal axis represents time it is typically called a **Time Series** graph.

Example 7

In the graph below the unemployment rate is shown for four quarters from Dec 7 till June 10. Examine the graph and identify any concerns about how it was constructed.



The title is the first misleading piece of information on this chart. The data presented does not represent the (gross) job loss for the given quarter, but rather the total unemployed at that time.

The horizontal axis is missing quarters between the start of Dec 07 and June 2010. This is an example of **cherry picking data** as only those selected quarters were used. If they used all the quarters would the chart give a different interpretation? Possibly.

Related to the above is that the horizontal axis uses different time spans between the data points. When looking at the last two dates listed of March 09 and June 10 we can count 15 months between them, but the corresponding pair right before is a shorter span with a longer gap having 9 months.

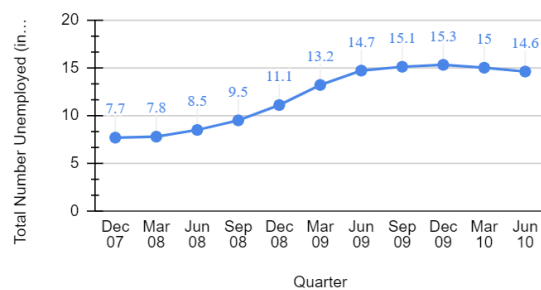
The vertical axis does not start at 0. This visually distorts the difference in values and makes the doubling of unemployment between Dec 07 and June 10 seem larger.

Take a look below at the adjusted chart including all the quarters starting at Dec 7 and extending through June 2010. The table of data is also provided for reference.

Unemployment Numbers

Quarter	Unemployed (in Millions)
Dec 07	7.7
Mar 08	7.8
Jun 08	8.5
Sep 08	9.5
Dec 08	11.1
Mar 09	13.2
Jun 09	14.7
Sep 09	15.1
Dec 09	15.3
Mar 10	15
Jun 10	14.6

Total Number Unemployed (in millions) vs....



Quiz score with 8 class intervals: [Google Sheet for Example 7](#)

What should we watch out for in a graph?

- » Look at the vertical axis. Does it start with 0? If not why is that?
- » Does it look like the data was cherry picked (look for non-uniform horizontal values in a time series, or missing values in the time series where data points should be)?
- » Does the vertical and horizontal axis keep a consistent scale (does not jump values)?
- » Were graphics used to display a bar? If so did the volume or area stay relative to data values?

- » For histograms did it add large gaps between bars that should be connected?
- » For histograms did it use consistent class widths?
- » Does the title match up with the data?
- » On a stem-and-leaf plot is it missing any stems (gaps in stem values)?

Exercises:

- A student collected data about the type of cola students liked to drink. In the process he had people select out of four different brands (Pepsi, Coke, Sprite, Diet Cola). Would it be appropriate to graph the frequency with a histogram? Why or why not.

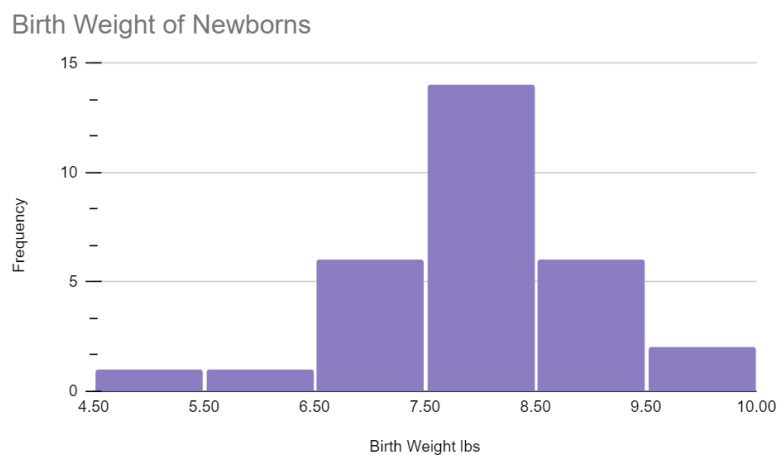
+ ANSWER

- What is the difference between a bar graph and a histogram?

+ ANSWER

- The weights of newborn babies are given on the histogram below. If the the newborn weighs between 6.5 lb and 8.5 lb they are considered healthy. If the baby weights below 5.5 lbs we classify the baby as a low birthweight (some low birthweight babies are healthy, but others can have health problems that could be serious).

- What percentage of these newborns were healthy?
- What percentage of these newborns are considered to have a low birthweight?



Exercise 1: Newborn baby weights. [Google Sheet for Exercise 1](#)

+ ANSWER

- The table below shows the number of hours worked for the 12 employees at a local restaurant.
 - Construct a histogram using 10 class intervals.
 - Construct a histogram of the data using an appropriate class width (make sure to describe your class width).
 - Construct a Stem and Leaf plot of the data.
 - Compare the two displays of the data and discuss which one you prefer and why.