Chapter 4 – Analyzing Skewed Quantitative Data

Introduction: In chapter 3, we focused on analyzing bell shaped (normal) data, but many data sets are not bell shaped. How do we analyze quantitative data when it is not bell shaped?

The main issue is that the mean and standard deviations are not accurate and should not be used in the analysis. Then what statistics should we use?

We will be introducing a new kind of graph that is specially designed for analyzing skewed data. It is called the “box and whisker plot” or “box plot” for short.

When data sets are not bell shaped, we will focus on the median, quartiles, interquartile range and boxplots to measure center and spread. Quartiles are more accurate because they are based on the order of the numbers instead of distances and so are not as affected by the skewed shape and extremely unusual values.
Section 4A – Review of Shapes and Centers, Creating Histograms and Dot Plots with Technology

Let us start by reviewing how to create dot plots and histograms with technology and determine the shape of a quantitative data set.

Here are the directions for making dot plots and histograms in Statcato.

Making a dot plot in Statcato:  Graph => Dot plot => Pick a column => OK

Making a histogram in Statcato:  Graph => Histogram => Pick a column => Chose number of bins => OK

Example 1

Use the women’s weight data (in pounds) from the Health data and Statcato to create a dot plot, histogram and boxplot. What is the shape of the data set?

Plugging in the women’s weight data into Statcato, we can use the directions above to create the following graphs.

![Dot Plot for Women's Weights in Pounds](image1)

![Histogram of Women's Weights in Pounds](image2)
What is the shape of women’s weight data?

*Though you can determine the shape from a dot plot and box plot, I prefer to look at the histogram when I judge the shape. Notice the highest bar is on the far left with a long tail to the right. Therefore, this data is skewed right.*

Are the mean and standard deviation accurate measures of center and spread for this data? No. *The mean and standard deviation are only accurate for bell shaped data sets.*

So what measure of center (average) should we use? **Median average**

**Center Principle for Quantitative Data**

If a data set is bell shaped (normal), the mean average is usually an accurate measure of center and we should use the mean as the average for the data set.

If a data set has a skewed shape, the median average is usually the most accurate measure of center and we should use the median as the average for the data set.

*Note: If a data set is not skewed, but just has an unusual shape like uniform, use the median also. Do not use the mean unless it is bell shaped. The mode is sometimes used as the center for bimodal or multimodal shaped data, since it can have multiple values and represent each hill in the data. That is why it is called bi-modal or multi-modal.*

---

**Problem Set Section 4A**

Directions: Copy and paste the following columns of data from the Bear data into Statcato. For each column of data, create a dot plot and a histogram. Draw a rough sketch of the graphs on a piece of paper or save them on a word document. Determine the shape of the data set and if you should use the mean or median as the center (average) for the data.

1. Bear Ages in Months
2. Bear Head Length in Inches
3. Bear Head Width in Inches
4. Bear Neck Circumference in Inches
Section 4B – Understanding the Median Average

In the last section, we said that we should use the median as our center and average, when data is skewed or not bell shaped, but what is the median? Why is it more accurate than the mean for skewed data?

Let us see if we can get a better understanding of the median average.

Definition of the Median average: The median average is the center of the data when the values are put in order from smallest to largest. The median is also called the “50th percentile” since approximately 50% of the numbers in the data set will be greater than the median and 50% of the numbers in the data set will be less than the median. Think of the median as a marker that divides the data in half. That is why it is often called the true center of the data.

Skewed data tends to have extremely unusual values. These unusual values (outliers) are very far from the mean. That is why the mean and standard deviation (typical distance from the mean) are not accurate for skewed data. The median is based on how many numbers are in the data set (frequency) and the order of the numbers. If the highest value were 40 or 4000, it would not change the median.
How to Calculate the Median Average
As with all statistics, rely on technology to calculate. No statistician calculates the median by hand, especially for large data sets. All of them use statistics software or computer software programs. To get a better understanding of the median, we will look at a couple examples where we calculate the median with small data sets.

To calculate the median, put the data in order from smallest to largest. Computer programs like excel can sort the data for you if you do not want to put it in order. Once the data is in order, you will look for the center of the data.

Odd Number of Values: If you have an odd number of values in the data set, then your median will be the number in the exact middle of the data when it is in order. Suppose we have 17 numbers in the data set. Then our median would be the ninth number in the data set. That would give us eight numbers below the median and eight numbers above the median. Remember the median separates the data into two equal groups.

Even Number of Values: If you have an even number of values in the data set, then your median will not be a value in the data set. The median will be half way in between the two numbers in the middle. Suppose you have 26 numbers in the data set. Then the median will be half way between the 13th and 14th numbers in the data set. That way thirteen numbers will be below the median and thirteen numbers will be above the median. If you cannot think of what half way in between would be, you could use the following formula.

Median (even # of data values) = (first number in middle + second number in middle) / 2

Example 1
Find the median for the brick weight (in kilograms) data from last chapter.

4.7 , 6.2 , 3.3 , 5.1 , 2.9 , 7.4 , 4.5

The first thing to notice is that the data is not in order. It needs to be put in order before we can find the median.

Data in order:
2.9, 3.3, 4.5, 4.7, 5.1, 6.2, 7.4

Since there are seven numbers in the data set. The fourth number (4.7) will be the median.

Median Average = 4.7 kilograms

Notice there are three numbers in the data set greater than the median (5.1, 6.2 and 7.4) and there are three numbers in the data set less than the median (2.9, 3.3 and 4.5).
Example 2
Let us look at a second example.

Here are the yearly salaries in thousands of dollars for employees from a small company.

36.5, 51.2, 47.9, 44.1, 37.2, 39.6, 41.8, 45.4, 43.2, 253.5

(This last salary of 253.5 thousand dollars was the CEO of the company.)

Remember to put the numbers in order first.

Yearly Salary Data in order:

36.5, 37.2, 39.6, 41.8, 43.2, 44.1, 45.4, 47.9, 51.2, 253.5

Since there are ten numbers (even), the median will not be a number in the data set. It will be half way between the two middle numbers that can divide the data in half. The two numbers in the middle are 43.2 and 44.1 thousand dollars.

Median Average = \frac{(43.2 + 44.1)}{2} = \frac{87.3}{2} = 43.65

Notice again that there are five numbers above the median (44.1, 45.4, 47.9, 51.2, 253.5) and five numbers below the median (36.5, 37.2, 39.6, 41.8, 43.2). The data has been split in half.

This is a good example to explain why the median is a better average than the mean. The CEO is a large unusual value in the data set, making the data very skewed right. Let us compare the mean and median averages.

Mean Average =

\frac{(36.5 + 37.2 + 39.6 + 41.8 + 43.2 + 44.1 + 45.4 + 47.9 + 51.2 + 253.5)}{10} = \frac{640.4}{10} = 64.04 \text{ thousand dollars.}

Median Average =

\frac{(43.2 + 44.1)}{2} = \frac{87.3}{2} = 43.65 \text{ thousand dollars}

Notice no one in the company makes 64 thousand dollars. The mean is not a good average for this data. The median however is very accurate. Many people in the company make around 43 or 44 thousand dollars. Recently, companies have been using the median average as their “average salary” on their websites for this very reason.
Calculating the median average with technology
All statistics software programs can calculate the median. This is a much better way to find the median, especially if you have larger data sets.

Remember the steps to calculating statistics with Statcato.

**To calculate statistics with Statcato:** Statistics => Basic Statistics => Descriptive Statistics => Pick a column of data => Pick what statistics you want to calculate (Median) => OK

**Example 3**
In the last section, we saw that the weights of forty women had a skewed right shape. Calculating the mean and median with Statcato gives us the following.
Notice that since the data is skewed right, the mean has been pulled in the direction of the skew. In other words, the mean average weight of the women is not very accurate and is too large. The median average of 135.8 pounds is a much more accurate average weight.

Note: In a skewed left data set, the mean will also be pulled in the direction of the skew. This will make the mean average too small. You can often get a good idea of the shape of a data set by just looking at the mean and median.

**Bell Shaped Data:** The mean and median are very close (both are accurate).

**Skewed Right Data:** The mean is significantly larger than the median. Only the median is accurate.

**Skewed Left Data:** The mean is significantly smaller than the median. Only the median is accurate.

---

**Problem Set Section 4B**

1. Put each of the following data sets in order from smallest to largest. Then calculate the median average (50th percentile).

   A. 5, 7, 8, 8, 9, 11, 14, 16, 17, 19, 21, 25, 26, 29, 31, 33, 36

   B. 2.1, 3.8, 5.1, 6.9, 7.2, 10.4, 11.3, 14.7, 15.1, 16.0
C.  31, 34, 41, 52, 68, 71, 79, 83, 88, 90, 103

D.  150, 152, 154, 155, 157, 159, 163, 164, 165

E.  7.5, 2.3, 4.6, 1.9, 2.8, 9.4, 8.3, 6.1

F.  21, 29, 23, 26, 25, 19, 28, 31, 32, 20, 18

2. Use Statcato and the bear data to calculate median average of the following data sets.

   Median average for Bear Ages = _____________ Months
   Median average for Bear Head Length = _____________ Inches
   Median average for Bear Head Width = _____________ Inches
   Median average for Bear Neck Circumference = _____________ Inches
   Median average for Bear Length = _____________ Inches
   Median average for Bear Chest Size = _____________ Inches
   Median average for Bear Weight = _____________ Pounds
Section 4C – Understanding Spread for Skewed Data, Quartiles, Interquartile Range (IQR), and the Five Number Summary

We have now seen that when data is skewed, we should use the median average as our measure of center and average.

The median is actually a type of quartile. Quartile analysis is an important part of understanding skewed data.

**Definition of Quartiles:** The quartiles are three numbers that break the data into four equal groups. Think of them as three fences that separate the data into quarters when the data is in order.

- **First Quartile (Q1):** This statistic is also called the 25th percentile and is the number that approximately 25% of the data is less than and 75% of the data is greater than.
- **Second Quartile (Q2):** This statistic is also called the Median or the 50th percentile and is the number that approximately 50% of the data is less than and 50% of the data is greater than.
- **Third Quartile (Q3):** This statistic is also called the 75th percentile and is the number that approximately 75% of the data is less than and 25% of the data is greater than.

When a data set is skewed, we said that we should not use the standard deviation to measure spread. So what measure of spread should we use for skewed data? In bell shaped data, typical values are closer to the center. The empirical rule implies that for bell shaped data, about 68% is typical. In skewed data, the data is more spread out with less values being typical. For skewed data, we look for the middle 50% of the data for typical values. This is called the interquartile range.

**Definition of Interquartile Range (IQR):** The interquartile range is how far typical values are from each other in a skewed data set. IQR is the length of the middle 50% of the data values and is calculated by taking the difference between the third quartile (Q3) and the first quartile (Q1).
Interquartile range formula: \( IQR = Q3 - Q1 \)

**Example 1**
Different statistics programs sometimes give slightly different values for the quartiles. People sometimes overemphasize these differences. The key is to remember the idea, finding three fences that separate the data into four equal groups. Each quarter should have approximately the same number of values in that group.

Let us calculate the three quartiles and the interquartile range (IQR) for the following data set.

Texas Horned Lizard Lengths in Centimeters (cm)
6.7, 10.4, 7.8, 4.7, 5.5, 5.8, 7.2, 7.5, 7.1, 6.3, 6.6, 7.4

To calculate quartiles, we need to put the numbers in order first.

Texas Horned Lizard Length Data in order:
4.7, 5.5, 5.8, 6.3, 6.6, 6.7, 7.1, 7.2, 7.4, 7.5, 7.8, 10.4

Now that the data is in order, think about quartering the data. Since there are 12 values in the data set, we should have \( \frac{12}{4} = 3 \) values in each quarter. Therefore, the quartiles should be placed between every three numbers.

\[
4.7, 5.5, 5.8 | 6.3, 6.6, 6.7 | 7.1, 7.2, 7.4 | 7.5, 7.8, 10.4
\]

Therefore, Q1 should be half way between 5.8 and 6.3

\[
Q1 = \frac{(5.8 + 6.3)}{2} = 6.05
\]

Therefore, 25% of the horned lizards had a length less than 6.05 cm.

Q2 (median) should be half way between 6.7 and 7.1

\[
Q2 (median) = \frac{(6.7 + 7.1)}{2} = 6.9
\]

Therefore, 50% of the horned lizards had a length less than 6.9 cm.

Q3 should be half way between 7.4 and 7.5

\[
Q3 = \frac{(7.4 + 7.5)}{2} = 7.45
\]

Therefore, 75% of the horned lizards had a length less than 7.45 cm.
This is how to think about quartiles. Notice we did not need a formula or fancy procedure to find the quartiles. We just needed to separate the data into four groups.

What about the interquartile range (IQR) for this data set?

\[ \text{IQR} = Q_3 - Q_1 = 7.45 - 6.05 = 1.40 \text{ cm} \]

Because of the one unusually large lizard of 10.4 cm, this data is probably skewed right. It is difficult to tell the shape of small data sets. Here is a dot plot of the data. It does look like much of the data is bunched up on the left and the one unusual value on the right gives a longer tail to the right. IQR is the best measure of spread for skewed data. In this case, the IQR tells us that the middle 50% of the data values (typical values) were within 1.40 cm from each other.

Example 2
There is some debate about how to calculate quartiles when the frequency is not divisible by four. This is especially true when there is an add number of data values. Remember if there is an odd number of data values, the median (Q2) is an actual number in the data set. This is where the debate sets in. Some computer programs include the median (Q2) in the calculation of the Q1 and Q3. Other programs leave the median out of the calculation. This next example highlights this difference.

Let us look at the weight of the bricks in kilograms from the previous section. Here is the data in order.

2.9, 3.3, 4.5, 4.7, 5.1, 6.2, 7.4

To find the quartiles, always start by finding the median (second quartile). In the previous section, we learned that since the number of values is odd, Q2 would be the middle number of 4.7 kg. To find the first quartile (Q1), find the median of the bottom half of the data and to find the third quartile (Q3), find the median of the top half of the data.
Here is where the debate comes in. Some computer programs include the 4.7 in the top and bottom half of the data and some do not.

Case 1: Not including the median in the Q1 and Q3 calculation. *(Note: Statcato does not include the Q2 in the top and bottom half of the data.)*

Q1 = median of bottom half of the data (all numbers below 4.7) = 3.3 kg

There are only three numbers below Q2. The median of 2.9, 3.3, and 4.5 would be just 3.3 since this is an odd number of values and that is the number in the middle.

Q3 = median of top half of the data (all numbers above 4.7) = 6.2 kg

There are only three numbers above Q2. The median of 5.1, 6.2, and 7.4 would be just 6.2 since this is an odd number of values and that is the number in the middle.

Hence, Statcato would calculate the following for this data:

Median = 4.7 kg, Q1 = 3.3 kg, Q3 = 6.2 kg and IQR = Q3-Q1 = 6.2 – 3.3 = 2.9 kg

Case 2: If Q2 (median) is an actual value in the data set, then some computer programs include the median in the Q1 and Q3 calculation. *(Note: Statcato does not calculate this way, but some other programs do.)*

Q1 = median of bottom half of the data (all numbers below 4.7 and including 4.7) = 3.9 kg

If we include the median in the bottom half, then there would be four numbers. The median of 2.9, 3.3, 4.5, and 4.7 would be just half way between 3.3 and 4.5 since this is an even number of values and there are two numbers in the middle.

Q3 = median of top half of the data (all numbers above 4.7 and including 4.7) = 5.65 kg

If we include the median in the top half, then there would be four numbers. The median of 4.7, 5.1, 6.2, and 7.4 would be half way between 5.1 and 6.2 since this is an even number of values and there are two numbers in the middle.

Hence, we may get the following statistics for this data:

Median = 4.7 kg, Q1 = 3.9 kg, Q3 = 5.65 kg and IQR = Q3-Q1 = 5.65 – 3.3 = 2.9 kg

**Take away**

This difference in how quartiles are calculated is not something to dwell on. In a data set with ten thousand numbers, the quartiles will be about the same no matter what program you are using. In small data sets like the previous example, there can be some discrepancy, but it is not something to worry about.
Again, the key is to explain the meaning of statistics like median, Q1, Q3 and IQR. Use technology to calculate the statistics. Take whatever value the program gives and use it. It matters very little if Q1 came out to be 78.4 degrees Fahrenheit or 78.3 degrees Fahrenheit. The important thing is to be able to explain that approximately 25% of the values in the data were less than Q1.

Calculating quartiles and IQR with technology
In large data sets, it is virtually impossible to calculate quartiles or any statistic for that matter with a calculator or by hand. We are now living in the age of “big data” where data sets often have hundreds of thousands of values or even millions of values. These data sets are so large that even programs like Statcato cannot calculate them. Surely, we cannot calculate the graphs and statistics we need from big data with a calculator. That is why it is so vital for data analysts to learn how to use statistics software like Statcato. You can always adapt to a larger program later if you need it, but the functionality of these programs are very similar.

To calculate statistics with Statcato: Statistics => Basic Statistics => Descriptive Statistics => Pick a column of data => Pick what statistics you want to calculate (Median, Q1, Q3, IQR) => OK

Example 3
Calculate the median, Q1, Q3 and IQR for women’s pulse rates in beats per minute (BPM) from the health data. Then write a sentence to explain each of these statistics and what it tells you about the data in context.

Putting the data into Statcato, we got the following graph and statistics.

![Histogram of Women's Pulse Rates in Beats Per Minute (BPM)](image)

Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>IQR</th>
</tr>
</thead>
<tbody>
<tr>
<td>C9 Women Pulse (Beats per min)</td>
<td>68.0</td>
<td>74.0</td>
<td>80.0</td>
<td>12.0</td>
</tr>
</tbody>
</table>
Notice first that this data is skewed right. So the median average of 74 beats per minute (BPM) is the best measure of center for the data and the IQR of 12.0 BPM is the best measure of spread for this data.

Sentence for the Median: The center of the data is the median of 74 BPM. This also tells us that the average pulse rates for these women is 74 BPM.

Sentence for Q1: Approximately 25% of the women in the data set had a pulse rate below 68 BPM.

Sentence for Q3: Approximately 75% of the women in the data set had a pulse rate below 80 BPM.

Sentence for IQR: The best measure of spread for this data was the interquartile range (IQR) of 12.0 BPM. This also tells us that typical women in the data set had a pulse rate within 12.0 BPM of each other.

Note: Do not confuse the interpretation of standard deviation with that of interquartile range. They are both measures of spread but they are different. The standard deviation is a measure of how far typical values are from the mean. However, IQR is not the distance from the median. IQR is the distance between Q1 and Q3. In other words, IQR is how far typical values may be from each other.

Quartiles are not based on distances like the mean and standard deviation. They are based on how many numbers are in the data set and the order of the numbers.

Five Number Summary
The quartiles belong to a group of numbers called the five number summary. We will see in the next section that the five number summary is the basis for making a box and whisker plot (boxplot).

Five number Summary: Minimum, Q1, Median, Q3, and Maximum

Example 3 continued
We can calculate the five number summary for the women’s pulse data in the last example by simply adding the minimum and maximum to the list of statistics for Statcato to find.

Descriptive Statistics
Variable | Q1 | Median | Q3 | IQR  
--- | --- | --- | --- | ---  
C9 Women Pulse (Beats per min) | 68.0 | 74.0 | 80.0 | 12.0  

Variable | Min | Max  
--- | --- | ---  
C9 Women Pulse (Beats per min) | 60.0 | 124.0  

Therefore, the five number summary of these women’s pulse rates is

Minimum = 60 BPM
Q1 = 68.0 BPM
Median = 74.0 BPM
Q3 = 80.0 BPM
Maximum = 124 BPM

**Problem Set Section 4C**  
1. Put each of the following data sets in order from smallest to largest. Calculate the median ($Q_2$), the first quartile ($Q_1$), the third quartile ($Q_3$) and the Interquartile Range (IQR). Give the five number summary for the data set (Minimum, $Q_1$, $Q_2$, $Q_3$ and Maximum).

   A. 5, 7, 8, 8, 9, 11, 14, 16, 17, 19, 21, 25, 26, 29, 31, 33, 36

   B. 2.1, 3.8, 5.1, 6.9, 7.2, 10.4, 11.3, 14.7, 15.1, 16.0

   C. 31, 34, 41, 52, 68, 71, 79, 83, 88, 90, 103
2. Use Statcato and the bear data to calculate minimum, maximum, median, Q1, Q3, and IQR for the following data sets. Then give the “five number summary” for each data set.

A. Bear Ages in Months

B. Bear Head Length in Inches

C. Bear Head Width in Inches

D. Bear Neck Circumference in Inches

E. Bear Length in Inches

F. Bear Chest Size in Inches

G. Bear Weight in Pounds
Section 4D – Box Plots, Typical and Unusual Values for Skewed Data
So far, we have seen that when a data set is skewed, we should use the median as our center and average and the interquartile range (IQR) for the spread. What about finding typical and unusual values for skewed data sets.

Typical Values
As with bell shaped data sets, your measure of spread should give you typical values. For skewed data sets, the interquartile range (IQR) is the best measure of spread. It measures the distance between the third quartile (Q3) and the first quartile (Q1). It is a measure of the
middle 50% of the data values. This is exactly what we want for typical values when the data is not bell shaped. In a skewed data set, typical values fall between the first quartile (Q1) and the third quartile (Q3).

*Typical Values for Skewed Data:* \[ Q_1 \leq \text{Typical Values} \leq Q_3 \]

**Unusual Values**

For bell shaped data sets, unusual values are more than two standard deviations from the mean, but skewed data involves more extreme values and is more spread out. It therefore has a different rule for finding unusual values. As with bell shaped data, let us start with the unusual cutoff values for skewed data.

*Unusually High Cutoff for Skewed Data:* \[ Q_3 + (1.5 \times \text{IQR}) \]

*Unusually Low Cutoff for Skewed Data:* \[ Q_1 - (1.5 \times \text{IQR}) \]

The good news is that the typical and unusual values for skewed data are summarized nicely with a box plot. The box plot is a fabulous graph to look at when your data is skewed.

I like to call the unusual cutoff values the “Box and a Half Rule”, since \(1.5 \times \text{IQR}\) represents the length of a box and a half. So any value in the skewed data that is a box and half from the box is considered unusual.

**Introduction to Box Plots**

Let us look at how box plots work. Remember to use technology when you create a box plot. No statistician, data analyst, or data scientist creates graphs by hand, especially with big data sets.

Let us look at an example where we do make the box plot by hand, just so we can understand the process.
Example 1
Let us look at the Texas Horned Lizard data and create a box plot for the data.

Texas Horned Lizard Length Data in order:
4.7, 5.5, 5.8, 6.3, 6.6, 6.7, 7.1, 7.2, 7.4, 7.5, 7.8, 10.4

A dot plot of the data indicated a skewed right shape. Therefore, this works nicely for a box plot.

In the last section, we calculated the three quartiles and the interquartile range for this data.

4.7, 5.5, 5.8 | 6.3, 6.6, 6.7 | 7.1, 7.2, 7.4 | 7.5, 7.8, 10.4

Q1 | Q2 | Q3

Q1 should be half way between 5.8 and 6.3
Q1 = (5.8 + 6.3) / 2 = 6.05

Q2 (median) should be half way between 6.7 and 7.1
Q2 (median) = (6.7 + 7.1) / 2 = 6.9

Q3 should be half way between 7.4 and 7.5
Q3 = (7.4 + 7.5) / 2 = 7.45

IQR = Q3 – Q1 = 7.45 – 6.05 = 1.40 cm
Typical Values: Since this is skewed data, typical values will fall in between Q1 and Q3. So typical horned lizards in this data set have a length between 6.05 cm and 7.45 cm.

Note: Q1 and Q3 do not accurately represent typical values in bell shaped data. You would need to use the standard deviation and the mean in that case.

Making the box plot
Start by drawing an even number line that goes from the smallest and largest values in the data set. Then draw a box from Q1 to Q3. Draw a line in the box at the median (Q2).

Now we need to calculate the unusual cutoff fences to determine if there are any unusual values (outliers) in the data set.

Unusually High Cutoff for Skewed Data:
Q3 + (1.5 x IQR) = 7.45 + (1.5 x 1.40) = 7.45 + 2.1 ≈ 9.55

Unusually Low Cutoff for Skewed Data:
Q1 – (1.5 x IQR) = 6.05 – (1.5 x 1.40) = 6.05 – 2.1 ≈ 3.95
Let us look at the dot plot again and see if there are any numbers that are 3.95 or lower. We can also look to see if there are any numbers that are 9.55 or higher.

![Dot Plot for Lengths of Texas Horned Lizards in cm](image)

Notice there are no values in the data set that are 3.95 cm or below. That means there are no unusually low values in the data set.

There is one value in the data set that is 9.55 cm or higher. It is the maximum value of the data set 10.4 cm. Therefore, 10.4 cm is an unusually high value in the data set. We need to designate that value as an outlier (unusual). Some people like to draw their outliers with a circle, some draw it with a triangle, and some draw it with a star. I will draw it with a triangle.
Now we need to determine where to draw the whiskers. The whiskers are drawn to the highest and lowest numbers in the data set that are not outliers (not unusual). Be careful. The whiskers are not drawn to the unusual cutoff fences. They must be drawn to numbers that are actually in the data set and are not outliers.

There was no unusually low value in the data set. Therefore, the low whisker on the left should be drawn to the smallest number in the data set, which is 4.7 cm.

There was an unusually high value (outlier) at 10.4 cm. That means we cannot draw the whisker to that value. We must choose a new maximum value in the data set that is not an outlier. Looking at the dot plot, we see that the next biggest number in the data set was 7.8 cm. That is 9.55 cm or below so it is not unusual. We will draw the high whisker (on the right) to 7.8 cm since that is the largest number in the data set that is not an outlier (not unusual).
The whiskers probably get their name because they kind of look like cat whiskers. This is our complete box plot.

Creating Box Plots with Technology
Let us look at how to create box plots with Statcato.

Making a box plot in Statcato:  
**Graph => Histogram => Pick a column =>**
*Chose whether you want the graph horizontally or vertically => OK*
*Right click on the graph and push “zoom out” and then “range axis” to see unusual values (outliers)*

Example
Use the women’s weight data (in pounds) from the Health data and Statcato to create a boxplot. In section 4A, we found the following dot plot and histogram and determined that the shape of the data was skewed right.
Plugging in the women’s weight data into Statcato, we can use the directions above to create a box plot.

Making a box plot can be a little tricky. You want to pay attention to where the highest and lowest values in the data set are. Sometime the default box plot does not show the unusual values (outliers). When I first clicked on box plot, I got the following graph. I made a horizontal box plot, though sometimes you may see them drawn vertically.
Notice if we compare this to the dot plot, the graph is missing a couple high values. If you right click on the graph, then click on zoom out and range axis we see the unusual values drawn as circles or triangles.

The box plot is a fabulous graph for analyzing skewed data. The box itself tells us where typical values in the data set are. The whiskers show us the largest and smallest values in the data set that are not unusual. The circles or triangles outside the whiskers are our unusual values in the data. Notice we did not have to calculate the unusual cutoff fences. The computer does it automatically. If we hold our cursor on the circle and triangle, Statcato can tell us what our unusual values (outliers) are.

The circle and the triangle represents unusual values in the data. The triangle is a designation as an extremely unusual value. Notice if you hold your cursor on the circle in Statcato, it tells you that this weight was 238.4 pounds. Holding your cursor on the triangle tells us that this unusual value was 255.9 pounds.
Holding your curser on the box itself gives you a lot of information. Statcato gives you the five number summary. It is important to realize that the five number summary is not including the unusual values (outliers). It gives the minimum value that is not an outlier (whisker), the maximum value that is not an outlier (whisker), Q1, Median, and Q3.

The dot inside the box is the mean. Most box plots do not tell you the mean. Remember the mean is not accurate for skewed data. Think of it as a way to compare the mean and the median and to tell how far off the mean is. Remember the following principles.

*If the mean is significantly greater than the median, the data is probably skewed right.*

*If the mean is significantly less than the median, the data is probably skewed left.*

*If the mean and median are close, the data is close to bell shaped.*

Box plots are often drawn vertically. Here is the same graph drawn vertically. Again, do not forget to right click on the graph and zoom out the range axis. If you do not, you will not see the unusual values (outliers). Notice the quantitative scale is now on the left.

**Interpreting Box Plots**

Remember to use technology to create graphs and find statistics. The important part is being able to interpret what the graph and statistics are telling us.
Box plots are often used to compare quantitative data from different groups. We call this kind of graph a “side by side” box plots. These graphs can be drawn horizontally or vertically. The following example was found from the bear data. The bear weights were separated into three groups depending on what time of the year the measurements were taken (Spring, Summer or Fall). The box on top is describing bears measured in spring (April – July), the box in the middle is describing bears measured in summer (August – September) and the box on the bottom is describing bears measured in fall (October – November).

Graphs like this give us a lot of information and are the foundation of ANOVA testing later on.

Which group of bears had the highest average weight and what was the highest average weight?

Notice the lines inside the box are the medians, which are very accurate measures of center. The group whose median line is farthest to the right is the fall bears (October-November). Holding the curser on the box tells us the median average weight for the bears measured in fall was 225 pounds.
Which group of bears had the most typical spread (variability) in their weights? What was the largest typical spread? Find the two values that typical values fall in between for the largest spread group.

Typical spread is the length of your box. So which group had the longest box? We can see it is the middle group of bears measured in summer (August – September). Holding your cursor on the box tells us that in fact typical bears measured in summer had a weight between 147 pounds (Q1) and 217.5 pounds (Q3). So the typical spread for the summer group can be measured with $IQR = Q3 - Q1 = 70.5$ pounds.

Problem Set Section 4D

1. Find the median, Q1, Q3, IQR and the unusual cutoff fences and use them to construct a box plot for the following data sets. Be sure to identify the outliers (unusual values) and draw the whiskers to the largest and smallest numbers in the data set that are not outliers. The data sets are already in order.

   A. 4, 5, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 30, 32, 33, 51

   B. 31, 32, 33, 34, 35, 36, 37, 55

   C. 6.4, 10.6, 10.8, 10.9, 11.0, 11.2, 11.3, 11.4, 11.6, 11.7, 11.9, 15.1

2. Copy and paste the following columns of data from the Bear data into Statcato. For each column of data, create a box plot. Make sure to zoom out the axis on the box plot so that you can see the highest and lowest values in the data set. Draw a rough sketch of the box plot on a piece of paper or save it on a word document. Determine the two values that typical values fall in between and list all unusually high values and unusually low values.

   A. Bear Ages in Months
   B. Bear Head Width in Inches
   C. Bear Length in Inches
   D. Bear Chest Size in Inches
   E. Bear Weight in Pounds
Let us look again at the side-by-side box plot describing the weights of bears measured at different times of the year. The box on top is describing bears measured in spring (April – July), the box in the middle is describing bears measured in summer (August – September) and the box on the bottom is describing bears measured in fall (October – November).

3. Which group of bears had the lowest average weight? What was the lowest average for the three groups?

4. Which group of bears had the lowest typical spread (variability) in their weights? (These are the bears whose weights were most consistent.) What was the smallest typical spread? Find the two values that typical values fall in between for the smallest spread group.

5. Were there any unusual values in any of the three groups (yes or no)? If so, what group had an unusual value? Was it unusually high or unusually low? Give the weight of the bear or bears that were considered unusual.
Section 4E – Summary Report Paragraph for Skewed Data

In our last chapter, we discussed how to summarize the key features of a data set with a summary report paragraph.

Quantitative Data Analysis Summary

- What is the data measuring? What are the units?
- How many numbers are in the data set? *(Frequency “N” or Sample Size)*
- What is the shape of the data? *(This will be skewed right or skewed left in this section.)*
- What is the best measure of center? What is the average? *(If the data is skewed, these should both be the median average. Write a sentence to explain the median average.)*
- What is the best measure of spread? *(If the data set is skewed, this should be the interquartile range (IQR). Write a sentence to explain the IQR.)*
- Find two numbers that typical values fall in between. *(If the data is skewed, then typical values will fall between Q1 and Q3.)*

\[ Q_1 \leq \text{Typical Values} \leq Q_3 \]
• Find any unusual values (outliers) in the data set. *(For skewed data, you do not have to calculate the outlier fences, box plots do this automatically. We simply need to create box plot with technology and then see if there are any unusual values outside the whiskers.)*

Example 1
Let us look at some skewed data and write the summary report paragraph.

The following statistics and graphs come from the social media time data (in minutes) from the Fall 2015 Math 075 Survey data. Students were asked how many minutes a day they spend on social media.

We started by checking the shape with a dot plot and histogram.
We see that the data is strongly skewed to the right. That means we should not use the mean and standard deviation. We should instead focus on the median, Q1, Q3 and IQR. It is a good idea to include the minimum and maximum values. We should also make a box plot to check for unusual values.
C1 Approximately how many minutes a day, on average, do you spend on social media? | 20.0 | 60.0 | 110.0 | 90.0

Variable | Min | Max
--- | --- | ---
C1 Approximately how many minutes a day, on average, do you spend on social media? | 0 | 700.0

Math 075 Students Social Media Data Summary Report Paragraph: This data describes the time in minutes per day that Math 075 students in the Fall 2015 semester spend on social media. There were a total of 481 student social media times in the data set. The shape of the data set was very skewed right. The best measure of center was the median average of 60 minutes. The average amount of time that these students spend on social media each day is 60 minutes (1 hour). The best measure of spread was the interquartile range (IQR) of 90 minutes. So typical times spent on social media were within 90 minutes of each other. In fact, the typical amount of time spent on social media by these students is between 20 minutes (Q1) and 110 minutes (Q3). There were no unusually low values in the data. The smallest amount spent on social media was 0 minutes, but the box plot indicated that this was not unusual. There was several unusually high values (high outliers) in the data set. The unusually high values were 240 min, 270 min, 300 min, 360 min, 400 min, 420 min, 500 min, and 700 min.
Problem Set Section 4E
1. Answer the following questions:
   a) In a skewed data set, what measure of center should we use? ________________
   b) In a skewed data set, what measure of average should we use? ________________
   c) In a skewed data set, what measure of spread should we use? ________________
   d) In a skewed data set, what are the two numbers that typical values fall in between?
      ________________ ≤ typical values ≤ ________________
   e) In a skewed data set, approximately what percentage is typical? ________________
   f) In a skewed data set, how do we find unusual values (outliers)?
   g) Explain how we can use a Box Plot to find the unusually high and unusually low values in
      the data set.

Directions: Now analyze the following data sets. Open “Bear” data and the “Health” data from
my website www.matt-teachout.org. (Look under “Int Alg for Stats” and then the “data sets”
tab.) Use Statcato to create a histogram and dot plot to verify that each data set is skewed.
Then use Statcato to find the median, Q1, Q3, IQR, Minimum and Maximum. Give the best
center (median), average (median), and spread (IQR). Give two numbers that typical values are
in between (Q1 and Q3). Use Statcato to create a box plot. Make sure right click and zoom out
the range axis so that you can see the unusual values. Identify any numbers in the data set that
are unusually high or unusually low. Now write the summary report paragraph for the data set.
2. Bear Ages in Months

What is the data measuring and what are the units?

How many numbers are in the data set?

What is the shape of the data set?

Minimum = _________________

Maximum = ________________

Median = _________________

Q1 = _________________

Q3 = _________________

IQR = _________________

______________ ≤ typical values ≤ ________________

List all the numbers in this data set that are unusually high. *(If there were no unusually high values, write “none”)* _________________

List all the numbers in this data set that are unusually low. *(If there were no unusually low values, write “none”)* _________________

Now write a summary report paragraph for this data set.

3. Bear Weight in Pounds

What is the data measuring and what are the units?

How many numbers are in the data set?

What is the shape of the data set?

Minimum = _________________
Maximum = ________________
Median = ________________
Q1 = ________________
Q3 = ________________
IQR = ________________

_____________ ≤ typical values ≤ ________________

List all the numbers in this data set that are unusually high. *(If there were no unusually high values, write “none”)* ___________________________

List all the numbers in this data set that are unusually low. *(If there were no unusually low values, write “none”)* ___________________________

Now write a summary report paragraph for this data set.

4. Women’s Age in Years

What is the data measuring and what are the units?

How many numbers are in the data set?

What is the shape of the data set?

Minimum = ________________

Maximum = ________________

Median = ________________

Q1 = ________________

Q3 = ________________
IQR = ___________________

_______________ ≤ typical values ≤ ________________

List all the numbers in this data set that are unusually high. *(If there were no unusually high values, write “none”)* ______________________________

List all the numbers in this data set that are unusually low. *(If there were no unusually low values, write “none”)* ______________________________

Now write a summary report paragraph for this data set.

5. Women’s Systolic Blood Pressure (millimeters of mercury (mm of Hg))

What is the data measuring and what are the units?

How many numbers are in the data set?

What is the shape of the data set?

Minimum = _________________

Maximum = ________________

Median = _________________

Q1 = _________________

Q3 = ________________

IQR = _________________

_______________ ≤ typical values ≤ ________________

List all the numbers in this data set that are unusually high. *(If there were no unusually high values, write “none”)* ______________________________

List all the numbers in this data set that are unusually low. *(If there were no unusually low values, write “none”)* ______________________________
Now write a summary report paragraph for this data set.

6. Men’s Diastolic Blood Pressure (millimeters of mercury (mm of Hg))

What is the data measuring and what are the units?
How many numbers are in the data set?
What is the shape of the data set?
Minimum = _________________
Maximum = ________________
Median = _________________
Q1 = _________________
Q3 = _________________
IQR = _________________

_____________ ≤ typical values ≤ _________________

List all the numbers in this data set that are unusually high. *(If there were no unusually high values, write “none”)* _________________

List all the numbers in this data set that are unusually low. *(If there were no unusually low values, write “none”)* _________________

Now write a summary report paragraph for this data set.

7. Women’s Cholesterol in milligrams per deciliter

What is the data measuring and what are the units?
How many numbers are in the data set?
What is the shape of the data set?

Minimum = _________________

Maximum = _________________

Median = _________________

Q1 = _________________

Q3 = _________________

IQR = _________________

_______________ ≤ typical values ≤ _________________

List all the numbers in this data set that are unusually high. *(If there were no unusually high values, write “none”)* _________________

List all the numbers in this data set that are unusually low. *(If there were no unusually low values, write “none”)* _________________

Now write a summary report paragraph for this data set.
Section 4F – Measures of Center, Spread and Position

Though the mean, median, standard deviation and IQR are used most often in data analysis, there are many different types of statistics that can be used to dig deeper into the data. We will not be covering these statistics in depth, but it is good to at least have an idea of what they measure.

**Measures of Center**

**Mean Average**: The balancing point in terms of distances. The measure of center or average used when a data set is bell shaped (normal).

**Median Average**: The center of the data in terms of order. Also called the second quartile (Q2) or the 50th percentile. Approximately 50% of the data will be less than the median and 50% will be above the median. This is the measure of center or average used when a data set is skewed (not bell shaped).

**Mode**: The number that occurs most often in a data set. Data sets may have no mode, one mode, or multiple modes. It is also sometimes used in bimodal or multimodal data.

**Midrange**: A quick measure of center that is usually not very accurate, but can be calculated quickly without a computer. \((\text{Max} + \text{Min}) / 2\)

**Measures of Spread**

**Standard Deviation**: How far typical values are from the mean in a bell shaped data set. It is the most accurate measure of spread for bell shaped data. If you add and subtract the mean and standard deviation, you get two numbers that typical values in a bell shaped data set fall in between. It can also be used to find unusual values in bell shaped data. Should not be used unless the data is bell shaped.

**Variance**: The standard deviation squared. A measure of spread used in ANOVA testing. Only accurate when the data is bell shaped.

**Range**: A quick measure of spread that is not very accurate. It is based on unusual values and does not measure typical values in the data set. It can be calculated quickly without a computer. \((\text{Max} – \text{Min})\)
Interquartile range (IQR): How far typical values are from each other in a skewed data set. Measures the length of the middle 50% of the data. It is the most accurate measure of spread for skewed data sets. Should not be used when data is bell shaped. (Q3-Q1)

**Measures of Position**

**Minimum**: The smallest number in the data set. Is sometimes classified as an unusual value (outlier).

**Maximum**: The largest number in the data set. Is sometimes classified as an unusual value (outlier).

**First Quartile (Q1)**: The number that approximately 25% of the data is less than and 75% of the data is greater than. Used for finding typical values for skewed data sets.

**Third Quartile (Q3)**: The number that approximately 75% of the data is less than and 25% of the data is greater than. Used for finding typical values for skewed data sets.

**Frequency or Sample Size (N)**
The frequency or sample size of a data set (N) is not a measure of center, spread or position, but is important bit of information. It tells us how many numbers are in the data set.

---

**Problem Set Section 4F**

1. For each of the following statistics, classify it as a measure of center, spread or position.
   
   a) Q1
   b) Mean
   c) Variance
   d) Midrange
   e) Standard Deviation
   f) Minimum
   g) Q3
   h) Mode
   i) IQR
   j) Median
   k) Range
   l) Maximum
2. The following statistics were created from some weekly salary data in dollars from people living in Victoria, Australia. Write a sentence or two explaining the meaning of each of these statistics in context.

**Descriptive Statistics**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>C9 Victoria</td>
<td>1149.050</td>
<td>516.553</td>
<td>266826.719</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>IQR</th>
<th>Mode</th>
<th>N for mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>C9 Victoria</td>
<td>703.45</td>
<td>1015.74</td>
<td>1496.11</td>
<td>792.660</td>
<td>*</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>C9 Victoria</td>
<td>371.57</td>
<td>2396.28</td>
<td>2024.710</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>N total</th>
</tr>
</thead>
<tbody>
<tr>
<td>C9 Victoria</td>
<td>35</td>
</tr>
</tbody>
</table>
Chapter 4 Review
Here is a list of important ideas in this chapter.

- Be able to distinguish between categorical data and quantitative (numerical measurement) data.
- Be able to create histograms and dot plots with technology and find the shape of a quantitative data set.
- Be able to find the five number summary (minimum, Q1, median, Q3, Maximum) with a calculator and with technology. Also find the interquartile range (IQR) and the total frequency (N).
- Write sentences to explain Q1, Median, Q3 and IQR.
- The interquartile range (IQR) tells us the maximum distance that typical values are from each other in a skewed data set. It measures the spread for the middle 50% and is the most accurate spread for skewed data sets.
- The first quartile Q1 is a divider that about 25% of the data values are less than and about 75% of the data values are greater than.
- The third quartile Q3 is a divider that about 75% of the data values are less than and about 25% of the data values are greater than.
- A center gives an average value for the data set is usually close to the highest bar or bars in the histogram.
- If a data set is skewed, we should use the median average as our measure of center and our average for the data set.
- A measure of spread or variability tells us how spread out the data set is. The more spread out the data is, the less consistent the data is and the harder it is to predict. A small amount of spread tells us that the data is more consistent and easier to predict.
- If a data set is skewed, we should use the interquartile range (IQR) as our measure of spread for the data set. If a data set is bell shaped, then we should not use the IQR.
- For Skewed Data: \( Q_1 \leq \text{Typical Values} \leq Q_3 \)
- Unusually High Cutoff for Skewed Data: \( Q_3 + (1.5 \times \text{IQR}) \) (Automatically calculated in a box plot)
- Unusually Low Cutoff for Skewed Data: \( Q_1 - (1.5 \times \text{IQR}) \) (Automatically calculated in a box plot)
- Be able to read and use a box plot to understand quartiles and percentages and identify unusual values in the data set.
- Be able to write a summary report paragraph summarizing the key characteristics of a skewed quantitative data set.
- Be able to classify various statistics as a measure of center, spread or position.
Problem Set Chapter 4 Review
Directions: Give the shape of each of the following graphs from the men’s health data. Then decide what the best measure of center and spread would be. (Mean/standard deviation or median/IQR?)

1. Men’s Pulse Rate in Beats per Minute (BPM)
   Shape = ________________________      Mean/Stand Dev OR Median/IQR?

2. Men’s Diastolic Blood Pressure in Millimeters of Mercury (mm of Hg)
   Shape = ________________________      Mean/Stand Dev OR Median/IQR?

3. Men’s Heights (inches)
   Shape = ________________________      Mean/Stand Dev OR Median/IQR?
4. Calculate the Median, Q1, Q3 and IQR for the following data. The 16 numbers are already in order. Show work and put your answers in the spaces below.

17, 19, 20, 26, 28, 31, 35, 37, 41, 43, 44, 48, 51, 53, 55, 62

Median Average = _________________

Q1 = _________________

Q3 = _________________

IQR = Q3-Q1 = _________________

5. Interquartile Range (IQR) is an important measure of spread or variability in statistics. Give the basic definition of IQR.

6. How can we tell if we should use the median and IQR as our center and spread?

Look at the following Histogram, Box Plot and summary statistics of the women’s cholesterol data and answer the following questions.

![Histogram of Women's Cholesterol in mg per dL](image)
<table>
<thead>
<tr>
<th>Women Cholesterol (mg per dL)</th>
<th>Q1</th>
<th>Median</th>
<th>Q3</th>
<th>IQR</th>
<th>Min</th>
<th>Max</th>
<th>N total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>124.5</td>
<td>215.0</td>
<td>311.25</td>
<td>186.75</td>
<td>44.0</td>
<td>920.0</td>
<td>38</td>
</tr>
</tbody>
</table>

7. What is this data measuring? ________________

8. What are the units for the data set? ________________

9. What is the shape of the data set? ________________

10. How many numbers are in the data set? ________________

11. Are the median and IQR accurate for this data? (Yes or No) ________________
12. What is the average cholesterol for these women? (Give a number) (No calculation needed)
Average Cholesterol = _____________________

13. How far are typical values in the data set from each other? (Give a number) (No calculation needed)
Average distance typical values are from each other = _________________

14. Find two numbers that typical values fall in between and put your answer below. (No calculation needed)
_____________________ ≤ typical cholesterol for these women ≤ _______________________

15. Are there any unusually low values (low outliers) in the data set (yes or no)?
__________________

16. Are there any unusually high values (high outliers) in the data set (yes or no)?
__________________

17. List all unusual values (outliers) in the data set.
   Give the actual numbers, not a cutoff point. ________________________________

(For #18-22, refer to the boxplot.)
18. What percent of these women had a cholesterol below 311.25? _______________

19. What percent of these women had a cholesterol below 124.5? _______________

20. What percent of these women had a cholesterol higher than 215? _______________
21. What was the largest value in the data set that was not an outlier (not unusual)? _________________

22. True or False? There were more numbers in the data set greater than 311.25 than there were numbers in the data set less than 124.5.

Project Chapter 4 – Skewed Data Analysis Group Poster

Directions: The class will be separated into groups. Each group is required to pick a “team name” for their group and analyze one skewed quantitative data set from the math 075-survey data fall 2015, create a poster summarizing their findings, and present the poster to other students in the class.

Each group will have a different topic and will pick one of the following data sets from the math 075 survey data fall 2015 to present it to their classmates: Hours work per week, Hours sleep per night, Hours of exercise per week, Number of Minutes to get to school, College GPA, Number of Units completed at COC, Average cell phone bill per month, Dollars spent on a meal when eat out, Number of times eat at restaurant or fast food per week, Number of U.S. states visited, Number of minutes spent on social media.

The Poster Should Have

- Group/Team Name
- First and Last Name of each team members on the poster
- Why is this data important or interesting to your group?
- Graph: Histogram and Boxplot (Include outlier fences and outliers in Boxplot.)
- Measures of Center (Give the #’s): mean, median, mode, midrange = (max + min)/2
- Measures of Spread (Give the #’s): standard deviation, IQR, range, variance
- Measures of Position (Give the #’s): min, max, Q1, Q3
- What is the data measuring?
- What are the units?
- How many numbers are in the data set: sample size (n)
- Shape
- **Best Center** (*Median - Give the #*)
- **Average** (*Median - Give the #*)
- **Best Spread** (*IQR - Give the #*)
- **Two numbers that typical values fall in between** (*Q1 and Q3, Give the #'s*)
- **Unusually high values in the data (high outliers)** (*Actual unusual # in the data set. Use the boxplot.*)
- **Unusually low values in the data (low outliers)** (*Actual unusual # in the data set. Use the boxplot.*)
- **Decorate Poster**

**Presentation**

*Make sure each person on the team understands the poster and can present your findings. Bring your poster to a designated presentation area in the classroom and hang or tape your poster to a wall. One person at a time will present the poster. We will then rotate so that each member of the team gets to present. Everyone else will listen to presentations and give feedback. (Be Nice!)*