Before beginning this packet – please check your e-mail from Mrs. Sumner for important links to data needed to complete some of these activities and videos to help you if you are having trouble on an activity!!!!

Also, this packet is scanned to be printed double sided. There may be blank pages intentionally scanned so that the formatting will be correct.

AP Statistics Summer Assignment

Welcome to AP Statistics!! The purpose of this assignment is to get a jump on the school year and make you comfortable exploring and analyzing data. It will take several hours to properly complete, so please allow yourself a few days to do it and don't wait until the last minute! I created it with the feedback of previous AP Stats students right after they took the AP exam, so the information in this packet is information in which they felt would best prepare future AP Stats students. Successful completion of this assignment will help us "hit the ground running" and ensure adequate time in 2nd semester for proper review leading up to the AP Exam (and hopefully fewer assignments over Fall/Thanksgiving/Minimester/Spring Breaks).

This Summer Assignment will consist of 3 parts: Videos, Chapter 1 content, and Calculator Practice.

This assignment will be due on the 1st day of class. It consists of 3 parts

1. Videos: Please go to www.apstatsguy.com and watch the videos named "Summer Video One", "Summer Video Two", "Summer Video Three", and "Summer Video Four". You will need to take notes or write a BRIEF summary for these videos. These will be due on the first day of school and will be worth 20% of your summer assignment grade.

2. Chapter 1: These activities are covering Chapter 1 of your textbook and StatsMedic. You do not need your textbook to complete them. However, you will want to go to www.statsmedic.com, create an account, and you will have access to the electronic copies of the worksheets and answer keys for all the classroom activities this year. You can find these under https://www.statsmedic.com/ced-apstats-unit1. Please note that we will be doing the Parkinson's Disease activity on the first day of school, so please, please, please NO SPOILERS!!!! I have shared the data that has been collected in previous years in a google sheet so that you can complete these activities with data from William & Reed students (some of you may remember contributing to this data collection)!!! Please check your email for a link to a google drive folder that contains this data as well as videos by the StatsMedic creators, Luke Wilcox and Lindsay Gallas. These daily activities along with the quizzes in this packet will be due on the first day of school and will be worth the remaining 80% of your summer assignment grade.

3. The third part of the summer assignment is not for a grade but is HIGHLY recommended. You need to become familiar with your calculator. A TI-84 is preferred and is an investment you will not regret. This calculator will be your BFF in AP Statistics. Please take the time to go though the calculator practice in a separate document.

Also, there are a few specific supplies you will need for class on Day 1:
- Graphing calculator (TI-84 is preferred).
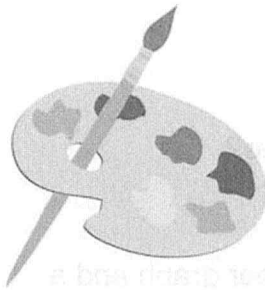- One-inch (or bigger) 3 ring binder
- Colored pens/pencils

I will check emails periodically throughout the summer. If you have questions, please email, but beware that it may take several days for me to get back to you. ksumner@williamANDreed.com

Have a great summer!!!

Mrs. Sumner

Lesson 1.1: How are your favorite classes related?

Is your favorite elective class associated with your favorite core class? Collect class data to see if there is a relationship.

1. Which of the following is your favorite elective class? You must choose only one and mark your choice on the board.

Art	Music	Physical Education	Foreign Language	Technology

2. Identify the individuals and variable?

3. Is the variable categorical or quantitative?

4. Go to stapplet.com to enter the class data. Make a bar graph and a pie chart. Sketch them below.

5. Sometimes it is helpful to investigate more than one variable. Come to the board and put a tally mark where you belong.

Find each of the following:

Core Class

Elective		Math	English
	Art		
	Music		
	P.E.		
	Foreign Lang.		
	Tech.		

% of all students who chose P.E.:

% of all students who chose Math and chose Art:

% of the students who prefer math that chose Tech.

6. How many variables does the table have? Are the variables categorical or quantitative?

7. Which variable would best explain or predict the other variable?

8. Go to stapplet.com and enter the data. Make a side-by-side bar graph and a segmented bar graph. Sketch them below.

9. How do the bars in the side-by-side-bar graph relate to the bars in the segmented bar graph?

10. Is there an association between favorite core subject and favorite elective? If so, describe it.

11. If there was not an association between favorite core subject and favorite elective, what would the graphs look like? Explain.

Name: _____ Hour: _____ Date: _____

Lesson 1.1 – Analyzing Categorical Data

Important Ideas:

Check Your Understanding:

1. Students at a local high school were asked which gaming system they preferred: the Playstation 3, the Xbox 360 or neither. The graph shown at right shows the results. Explain why the graph may be misleading.

Prefered GTA version (percentage)
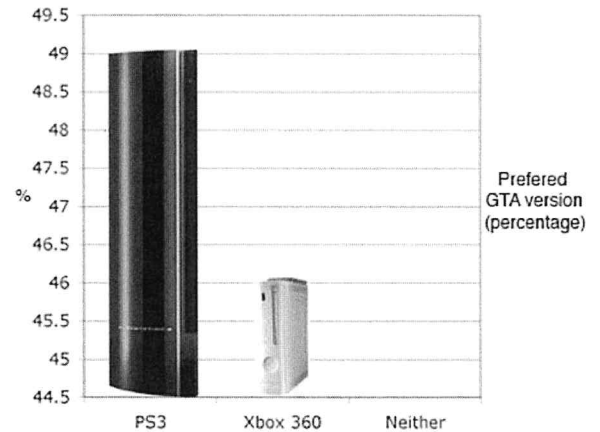
2. An article in the Journal of the American Medical Association reports the results of a study designed to see if the herb St. John's wort is effective in treating moderately severe cases of depression. The study involved 338 patients who were being treated for major depression. The subjects were randomly assigned to receive one of three treatments: St. John's wort, Zoloft (a prescription drug), or placebo (an inactive treatment) for an 8-week period. The two way table summarizes the data from the experiment.

a. What proportion of subjects in the study were randomly assigned to take St. John's wort? Explain why this value makes sense.

| | | Treatment | | |
		St. John's wort	Zoloft	Placebo
Change in depression	Full response	27	27	37
	Partial response	16	26	13
	No response	70	56	66

b. Find the distribution of change in depression for the subjects in this study using relative frequencies.

c. What percent of subjects took Zoloft and showed a full response?

STATS MEDIC

What will be the W&R mascot?

When the school was built in 2014, the school needed to pick a mascot. The principal decided to have the students and teachers vote between three choices: Warriors, rams, or prairie dogs. He took a random sample of students and a random sample of teachers. The results of the surveys are given in the table.

	Warriors	Rams	Prairie Dogs
Teachers	80%	10%	10%
Students	30%	60%	10%

1. Create two bar graphs below to display the results. Use three different colors for the bars.

2. Complete the third graph by taking each bar from the teacher sample and stacking them. Use the colors to mark each section. Do the same for the student sample.
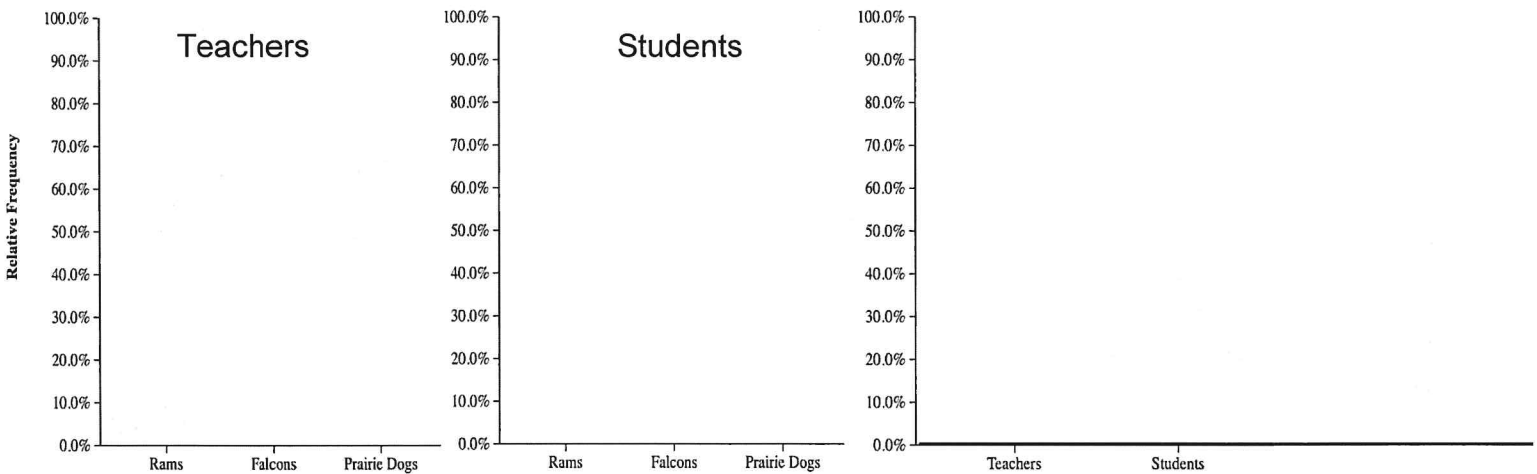
3. According to your displays, which mascot appears to have the most support? Explain.

4. Upon hearing the results of the surveys, the students argued that the decision was incorrect because 100 teachers had been surveyed and 500 students had been surveyed. Use this information to fill in the table below with the number of responses.

	Warriors	Rams	Prairie Dogs
Teachers			
Students			

5. How many times more students were sampled than teachers? _____. How can you update the third graph in #1 to take into account the sample size? Adjust your graph.

6. What should they make the W&R mascot? Explain.

# Topic 2.2 – Representing Two Categorical Variables

Important Ideas:

Check Your Understanding:

The following table gives the result of a random sample of upper level students at Rocky Vista University (the Fighting Prairie Dogs!), along with a mosaic plot.

Employment Status	Grade Level	
	Junior	Senior
Currently working	14	30
Not working but have had a job	22	40
Never had a job	15	10

Employment status

- Never had a job
- Not working but did in the past
- Currently working
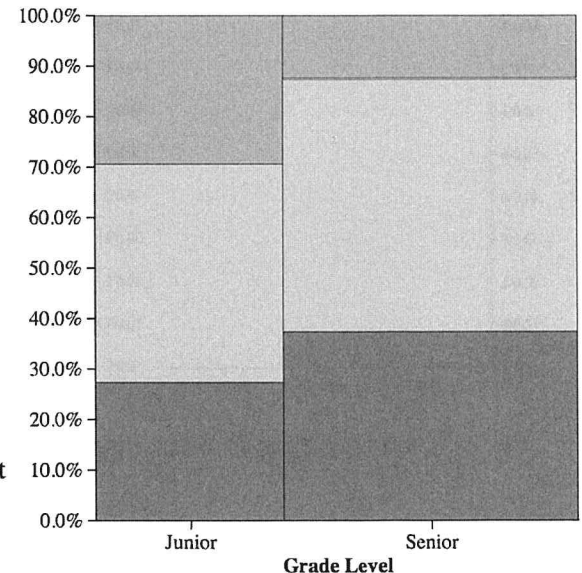
a. Calculate the proportion of Juniors that are currently working, not working but have had a job, and never had a job.

b. Calculate the proportion of Seniors that are currently working, not working but have had a job, and never had a job.

c. Write a few sentences summarizing what the display in part (a) reveals about the association between grade level and job experience for the students in the sample.

Lesson 1.2: How many pairs of shoes do you own?

1. How many pairs of shoes do you own? Record your answer on the board.

2. Is "Number of pairs of shoes" a categorical or quantitative variable?

3. Enter the data at www.stapplet.com. Make a dotplot, stemplot, and histogram and sketch each below.

4. Describe the distribution of the number of pairs of shoes for your class.

   Shape:

   Outliers:

   Center:

   Variability:

5. Which of the three types of display do you prefer? Why?

6. Mrs. ~~Gallas.~~ Sumner wonders if teachers have the same number of pairs of shoes as students. She asked her colleagues to record the number of pairs that they had. The results are below.

<div align="center">

15 8 10 29 14 7 22 35 6 15 13 12 9 10

</div>

7. Enter this data at stapplet.com. Be sure to make 2 groups (students and teachers).

8. Make dotplots, a side-by-side stemplot, and then histograms. Copy one of these graphs below.

9. Compare the distributions of the number of pairs of shoes for students versus teachers. Be sure to address shape, outliers, center, and variability.
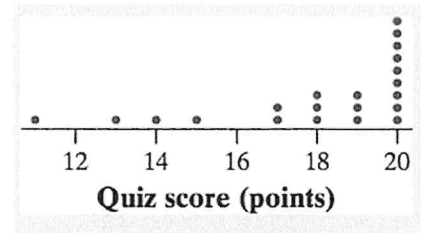
Lesson 1.2 – Displaying Quantitative Data

Important Ideas:

Check Your Understanding:

1. The dotplot displays the scores of 21 statistics students on a 20-point quiz.

(a) What percent of students scored higher than 16 points?

(b) Describe the shape of the distribution.

(c) Are there any potential outliers? Why?

2. Here is a back-to-back stemplot of 19 middle school students' resting pulse rates and their pulse rates after 5 minutes of running.

Write a few sentences comparing the distributions of resting and after-exercise pulse rates.

| Resting		After exercise
9888	6	
8664110	7	
8862	8	6788
60	9	02245899
4	10	044
	11	8
0	12	44
	13	
	14	6

Key: 8|2 is a student whose pulse rate is 82 beats per minute.

1. Mr. Molesky observes a group of monkeys for 24 hours to learn about their behavior. He records how long they slept, how many bananas they ate, gender, age, and the specific breed of monkey.

(a) What are the individuals in this data set?

(b) Identify the variables that were recorded, and indicate whether each one is categorical, quantitative (discrete), or quantitative (continuous).

2. The following bar graph gives the percent of owners of three brands of trucks who are satisfied with their truck. What is wrong with the way information is presented in this graph?
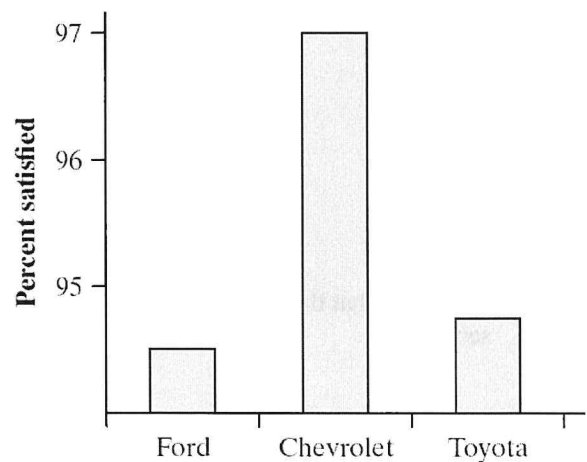
3. A research study asked children which of four different emotions they associated with the color red. The response and gender of each child are given in the following table.

	Male	Female
Joy	28	61
Happiness	20	25
Love	40	80
Anger	18	60

(a) Find the distribution of emotion for each gender using relative frequencies.

(b) Make a segmented bar graph to compare the distributions in part (a).
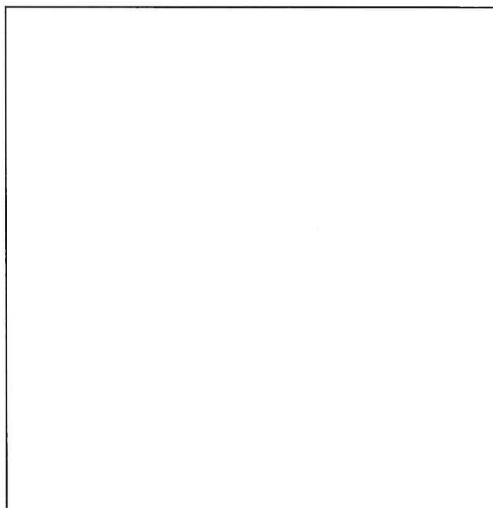
(c) Describe what the graph in (b) reveals about the association between gender and emotion for the students in the sample.

1. A random sample of 20 employees at a large company was selected. Here are the salaries (in thousands of dollars) for these employees during one year.

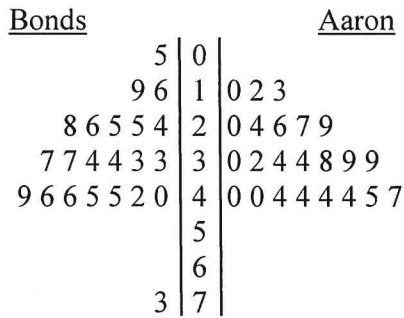| 28	31	34	35	37	41	42	42	42	47
49	51	52	52	60	61	67	72	75	77

(a) Make a dotplot of these data.

(b) What percentage of the sample of employees have a salary of at least $60,000?

2. Professor Windley teaches a statistics class with 136 students. On the first day of class he records how much money each student has in his or her possession (in dollars) during the first class of the semester. The histogram displays the data. Describe the distribution.
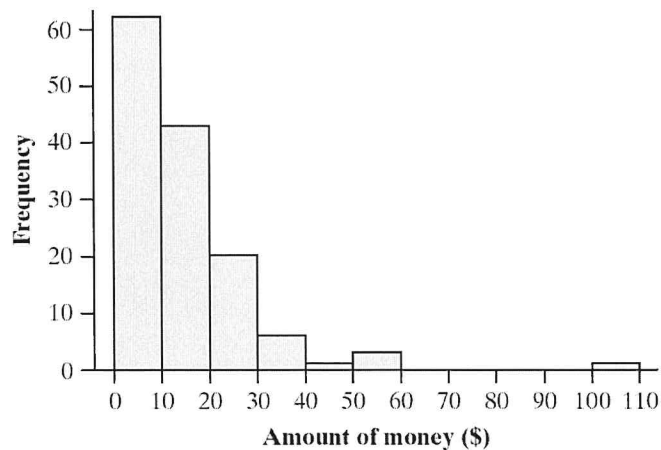
3. On August 7, 2007 Barry Bonds hit his 756th home run, breaking the all-time career home run record, formerly held by Hank Aaron. Does that make Bonds a better home run hitter than Aaron? Let's compare their annual home run production over their entire careers. A side-by-side stemplot is shown below.

Number of Home Runs Per Year

Bonds		Aaron
5	0	
9 6	1	0 2 3
8 6 5 5 4	2	0 4 6 7 9
7 7 4 4 3 3	3	0 2 4 4 8 9 9
9 6 6 5 5 2 0	4	0 0 4 4 4 4 5 7
	5	
	6	
3	7	

Key: 1|4 = 14 home runs

Use the plot to write a few sentences discussing the similarities and differences in the distributions of home runs per year for Bonds and Aaron.

Lesson 1.3: How many colleges are you applying to?

| How many different colleges is your group of 4 applying to? Find the total number of colleges for your whole group. |

1. Record the data for the class here.

2. Calculate the mean and median for the set of data. Compare them.

3. What is the range of the data?

Finding Standard Deviation

4. Finding range is helpful but it does not tell us how spread out the data is between the minimum and maximum. How can we find the *average distance of the values from the mean*?

a. Complete the table.

b. The average you calculated is the average of the **squared distances** from the mean. How do we use this to find the **average distance from the mean**? Find it.

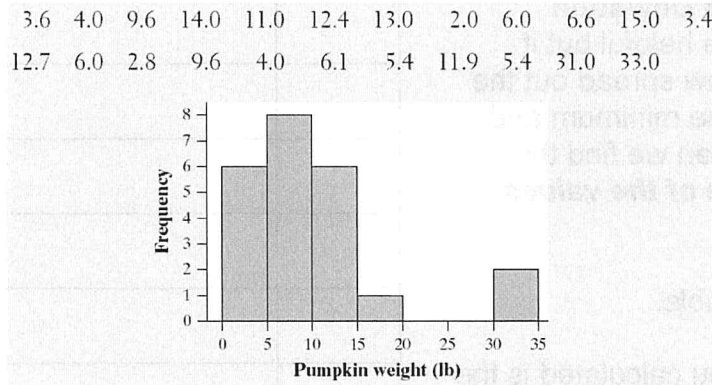| Value	Distance from mean	(Distance from mean)^2
	Total:	
Average (Distance from mean)^2:		

5. Go to stapplet.com. Enter the classroom data and find the summary statistics. Verify our work. How does it compare?

6. We forgot to add Mrs. Sumner. She applied to 20 colleges. Add her to the data set. Calculate the new mean, median and standard deviation using the applet. How does it compare to the original measures? Why do you think this is?

TheStatsMedic

# Lesson 1.3 – Describing Quantitative Data with Numbers

Big Ideas:

Check Your Understanding:

Some students purchased pumpkins for a carving contest. Before the contest began, they weighed the pumpkins. The weights in pounds are shown here, along with a histogram of the data.

| 3.6 | 4.0 | 9.6 | 14.0 | 11.0 | 12.4 | 13.0 | 2.0 | 6.0 | 6.6 | 15.0 | 3.4 |
| 12.7 | 6.0 | 2.8 | 9.6 | 4.0 | 6.1 | 5.4 | 11.9 | 5.4 | 31.0 | 33.0 | |

1. Calculate the mean weight of the pumpkins.

2. Find the median weight of the pumpkins.

3. Would you use the mean or the median to summarize the typical weight of a pumpkin in this contest? Explain.
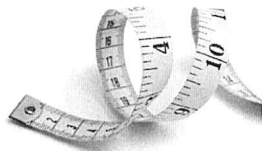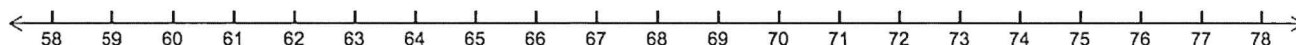
Lesson 1.3: Where Do I Stand?

How does my height compare with other AP Stats students?

In pairs, measure each other's height, rounded to the nearest inch.
Record your height on the dotplot at the front of the room (females use red, males use green).
Make a line at the front of the room, shortest to tallest.
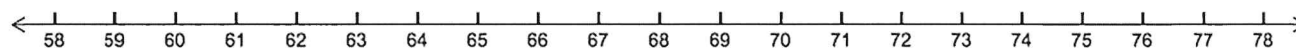
1. Record the dotplot

```
←—+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——→
  58   59   60   61   62   63   64   65   66   67   68   69   70   71   72   73   74   75   76   77   78
```

2. What is the median height? Describe how you found it.

3. What is Q_1 and Q_3? Describe how you found them.

4. Record the following values and then use them to make a boxplot.

Minimum: Q_1:   Median: Q_3:   Maximum:

```
←—+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——+——→
  58   59   60   61   62   63   64   65   66   67   68   69   70   71   72   73   74   75   76   77   78
```

4. The **interquartile range** (or *IQR*) is defined as $Q_3 - Q_1$. Find the *IQR*. Where do you see the *IQR* in the boxplot?

5. An **outlier** is a data value that is way too small or way too big (using the rules below). Are there any outliers? Show your work.
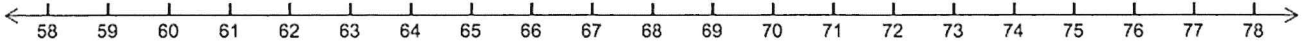
Way too small $< Q_1 - 1.5IQR$

Way too big $> Q_3 + 1.5IQR$

TheStatsMedic

6. Now we will separate our data into two groups, females and males.
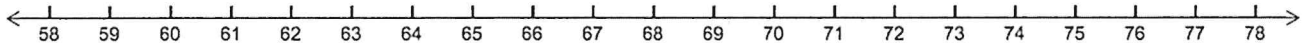
Heights for females – find the following values and then make a boxplot.

Minimum: Q_1:   Median: Q_3:   Maximum:

Heights for males – find the following values and then make a boxplot.

Minimum: Q_1:   Median: Q_3:   Maximum:

Write a few sentences <u>comparing</u> the distribution of heights for girls with the distribution of heights for boys.

TheStatsMedic

# Lesson 1.3 – Describing Quantitative Data with Numbers

Big Ideas:

Check Your Understanding:

Mr. Wilcox is a huge fan of University of Michigan football. His favorite season was the 1997 season (a perfect season!). Here is a back-to-back stemplot of the points scored by the 1997 University of Michigan football team and the archrival Michigan State University football team. Write a few sentences comparing the distributions.

University of Michigan

Michigan State University

Points scored
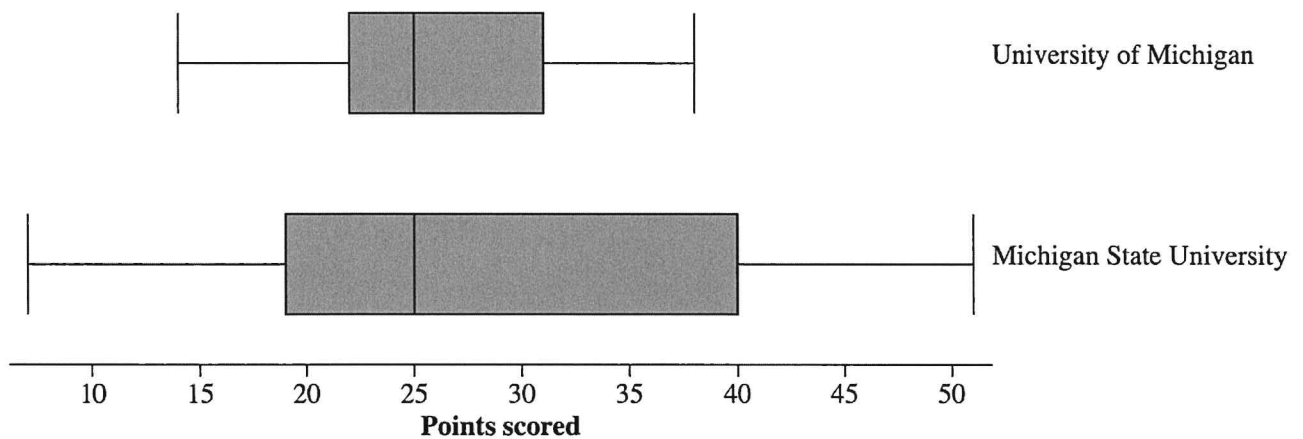
1. A group of 78 third-grade students in a Midwestern elementary school took a "self-concept" test that measured how well they felt about themselves. Higher scores indicate more positive self-concepts. The lowest self-concept scores were 20, 26, 27, 31, and 32. A histogram and some summary statistics from Minitab for these students' self-concept scores are given below.
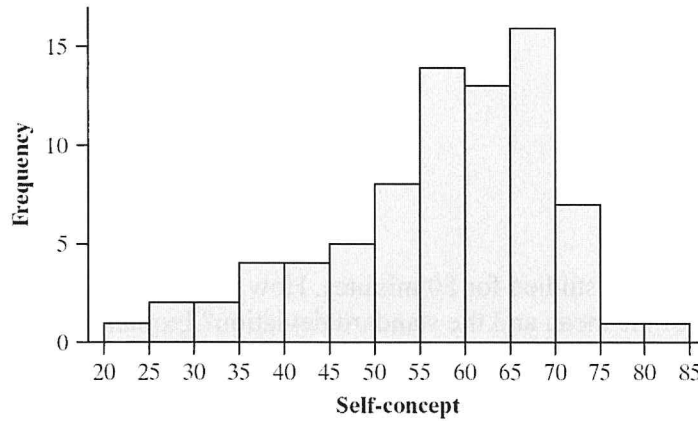
Variable	n	Mean	StDev	Minimum	Q_1	Median	Q_3	Maximum
SelfConc	78	56.85	12.35	20	50	59	65	84

(a) Are there any outliers? Justify your answer.

(b) Draw a boxplot of this distribution.

(c) One of the students had a self-concept score of 32. If this score had been accidently recorded as 22, what effect would this have had on the value of the mean and the median? Justify your answer.

2. Five students reported the amount of time (in minutes) they spent studying for an AP Statistics test the night before the test. The mean of the reported times is 45 minutes and the standard deviation is 10 minutes.

(a) Interpret the standard deviation in context.

(b) A 6th student reported that they studied for 50 minutes. How would the addition of this student to the data set affect the value of the mean and the standard deviation? Explain your answers.

# Who is Baseball's Greatest Home Run Hitter?

VS

Barry Bonds broke Mark McGwire's record when he hit 73 home runs in the 2001 season. How does this accomplishment fit with the rest of Bond's career? Here are Bond's home run counts for the years 1986 to 2007.

16	25	24	19	33	25	34	46	37	34	49	73	46	45	45	5	26	28

1. What display did your group get assigned? _____

2. Create the display on the whiteboard and bring it to the front of the room.

3. Describe the distribution.

4. What are the advantages and disadvantages of your type of display?

5. Below is this distribution of the number of home runs per season for Mark McGwire. Compare this distribution to the one for Barry Bonds.

Number of Home Runs
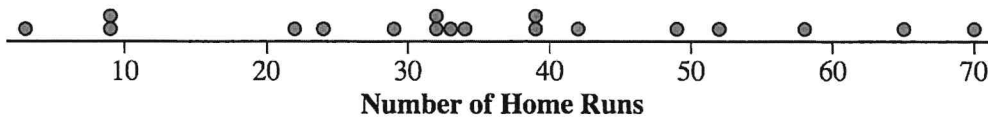
n	mean	SD	min	Q_1	med	Q_3	max
18	35.611	18.69	3	24	33.5	49	70