Algebra Name:_____

Data Analysis Review Packet Date:_____

Section 1: Histograms

1. What kind of data do we need to make a histogram?

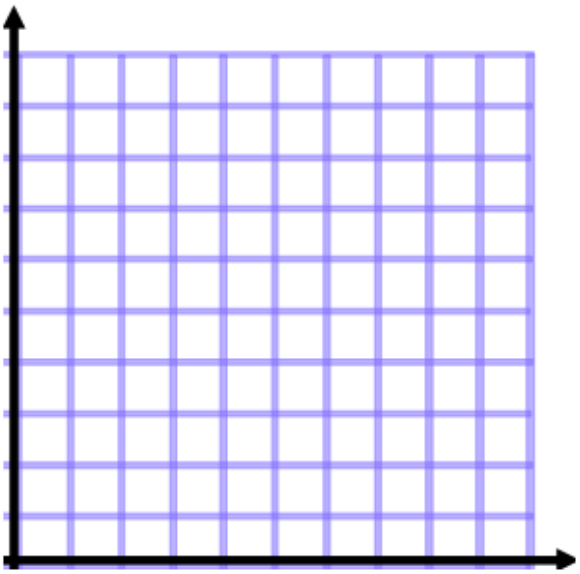
2. What is the definition of the type of data you identified in number 1?

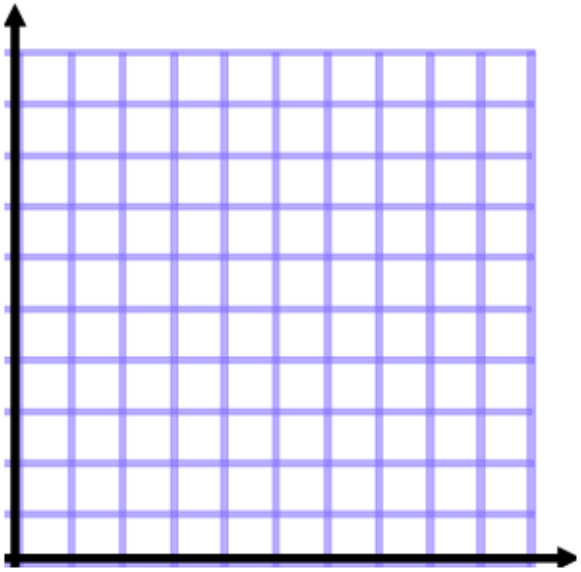
3. The following data set are the number of runs scored by the softball team. Make a table of the data including a tally column, a frequency column and a cumulative frequency column.

4, 10, 8, 0, 2, 4, 5, 7, 0 1, 18, 4, 3, 7, 13, 15, 8, 9, 1, 3, 5, 2,

Runs Scored	Tally	Frequency	Cum. Freq.
0-5			
6-10			
11-15			
16-20			

4. Make a histogram of the data above and a cumulative frequency histogram.

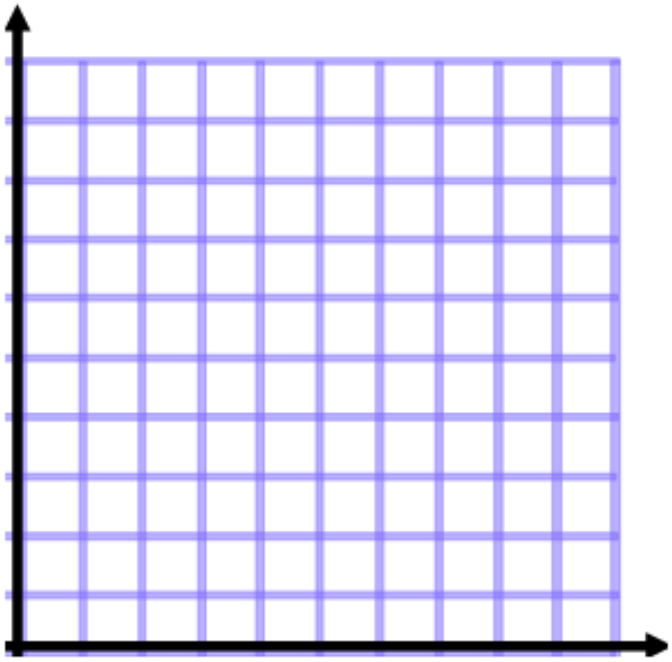




Section 2: Scatterplots

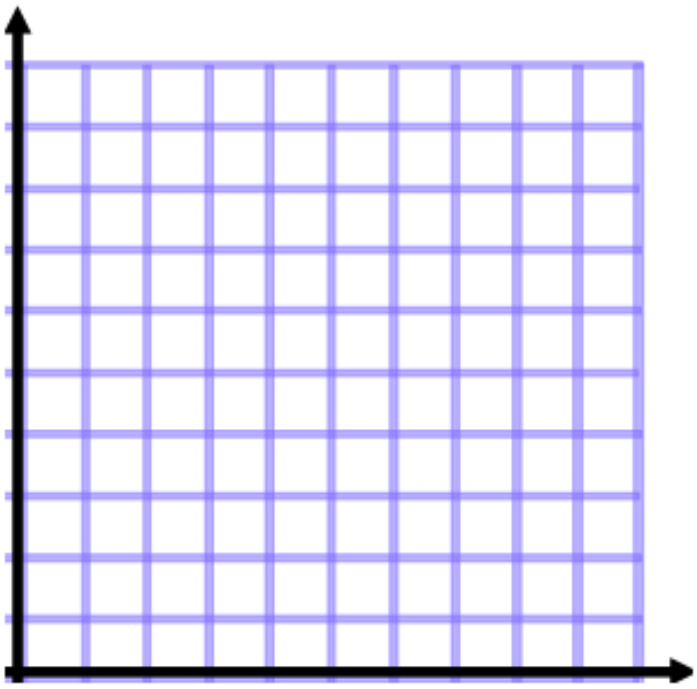
1. Make a scatterplot of the following data:

Hours Running	Calories Burned
3	400
2	350
1.5	290
4	550
1	210
2.5	380
3	420
2	330
4	490
5	600



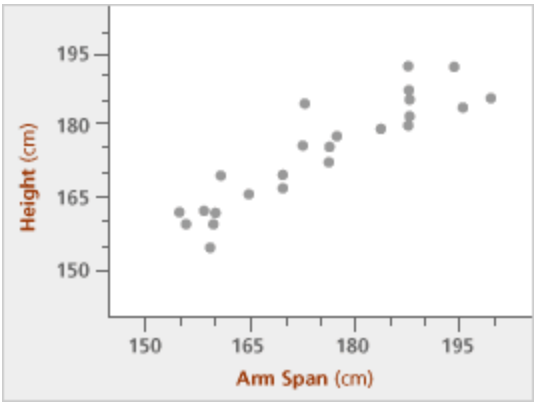
2. Make a scatterplot of the following data **use a break in scale on both axis:**

Score on Alg. Regents	Score on US Regents
82	80
89	61
72	75
90	88
71	80
67	73
88	91
80	75



3. Draw a line of best fit on both graphs. Which graph shows a stronger correlation? Explain:

Section 3: Best Fit Line and correlation.



1.

a. Examine the scatterplot to the right. Draw in a line of best fit. What kind of correlation do you see?

b. If someone's arm span is 187.5 cm, approximately how tall are they?

c. If a person is 150 cm tall, approximately what is their arm span?

d. If you were to approximate the R value as 1, .5, 0, -.5 or -1 which is the best choice? Why?

2. You are studying a data set on the amount of time students spent on their cell phones during class and test scores. The data set has strong negative correlation. What does that mean? What happens to a student's test score as the time they spend on their cell phone during class increases?

3. Describe the difference between an R value of -.8 and +.8.

Section 4: Measures of Central Tendency

1. Define the following mathematical terms in non-mathematical words:
   a. Mean:
   b. Median:
   c. Mode:
   d. Range:

2. Find the 4 MCTs of the following data set:

{22, 29, 22, 21, 33, 34, 35, 19, 19, 22, 28, 32, 28, 39, 32, 29, 17, 22, 29}

3. Find the MCT for the following data set:

1.3, 0.6, 1.8, 2.2. 2.3, 0.9, 1.1, 1.0, 1.8, 2.0, 0.7

Section 5: Shifts and Outliers

1. If the data set from number two in section 4 were to incur a shift of plus 5, how would the MCT be effected?

2. A data set has mean = X and range = Y. If the data set incurs a shift of minus 10, what will the new mean and ranges be?

3. If we were to add the number 4 to the data set in number 2 in section 4, what would happen to the MCT ? Write a sentence for each MCT.

4. The following data set are the scores that Jessie earned on her English quizes. Find the MCT and identify any outliers in the data set. How did these outliers effect the MCT.

89, 94, 68, 92, 90, 88, 96, 83

Section 6: 5 Number Summary

1. What are the 5 numbers in the 5 number summary. What do they mean?

2. Find the 5 number summary for the following data set:

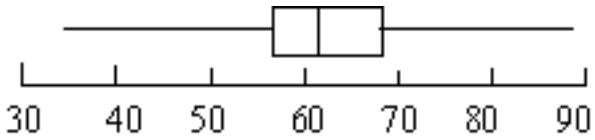
   9, 7, 2, 12, 8, 14, 11, 10, 5, 6, 14

3. If a data set has a maximum value of 44, which of the following is a possible value for the third quartile?
   a. 44
   b. 45
   c. 54
   d. 24

4. If a data set has a first quartile value of 23 and a maximum of 28, what is a possible value for the thrid quartile of the data set?

5. What is the interquarterile range IQR for the data set in number 2?

Section 7: Box and Whisker Plots



1. Study the scatterplot below of the number of calories in 20 different health bars and answer the questions below.

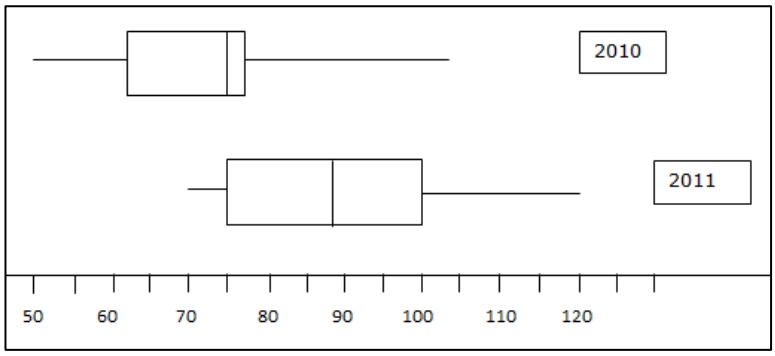
   a. Approximately _____% of the bars have less than 68 calories.
   b. Approximately _____ bars have more than 33 calories.
   c. The _____ number of caloires in the bars is 89.
   d. The range of the number of caloris for the data set is about _____.

2. Construct a box and whisker plot for the following data set, state the 5#S.

100, 183, 117, 142, 134, 104, 174, 138, 122, 179

3. The comparison box and whisker plot shows the scores for the chemistry regents in 2010 and 2011.

a. Did students do better in 2010 or 2011?



b. What percentage of students scored higher than 75 in 2010? _____

What percentage of studetns scored higher than 75 in 2011?_____

c. Which year has a higher IQR? _____ Which year has a larger range? _____

Section 8: Univariate/Bivariate and Quanitative/Qualitative

Identify if the following data is univariate or bivariate and if it is Quantitative or Qualitiative.

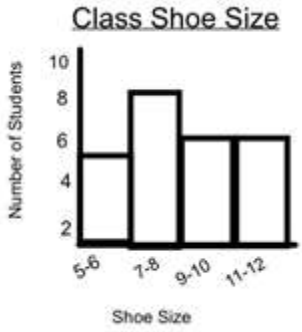
1. Sally took a survey of all of the people in her english class. She asked them how much money they spent on lunch that day.

2.

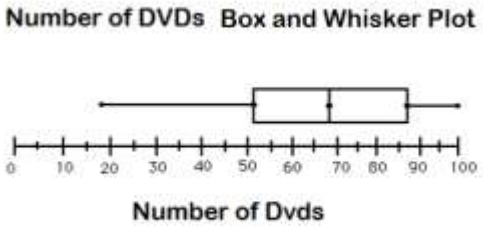
Fav. Color	Frequency
Red	21
Blue	28
Green	37
Purple	25
Pink	18

3. Miss P did a survey about how far away students live from school and how long it takes them to get here in the morning.

4.



5.



Section 9: Bias

1. You are doing a survey about animal rights. Identify the following as Biased or Unbiased and a survey method or survey question:
    a. Asking every other person who comes out of a library._____
    b. Asking " Do you think animal rights are imporant?" ._____
    c. Asking every other person who comes out a pets store. ._____
    d. Asking " don't you think baby animals are the cutest?!" ._____

2. You are doing a survey about peer tutoring.
    a. Give an example of an unbiased survey method.

    b. Give an example of a biased survey method.

    c. Give an example of an unbiased survey question

    d. Give an example of a biased survey question

Section 10: Data Analysis with a Calculator

1. Consider the regents data:
   a. Enter the data into your calculator and run a linear regression.
   b. What is the equation of the regression line:
      _____
   c. What is the corellation coefficient?_____
   d. What does the r-value say about the line of best fit?
   e. Use your regression line to determine the predicted US score for someone who scored a 75 on the Algebr regents_____
   f. Use your regression line to predict the algebra score for someone who scored an 79 on the US exam:_____

Score on Alg. Regents	Score on US Regents
82	80
89	61
72	75
90	88
71	80
67	73
88	91
80	75

2. The following data table shows the number of bacteria in a petri dish after a certain number of hours.

Hours	1	2	3	4	5
Bacteria	10	12	17	21	28

   a. Use your calculator to find an **exponential** model that represents the bacteria growth:_____

   b. At this growth rate, how many bacteria will there be after one full day?

3. The following list is the number of text messages a group of 20 students send in a day.

   { 18, 22, 93, 50, 52, 38, 21, 8, 29, 35, 53, 47, 21, 30, 51, 38, 29, 40, 50, 27}

   a. Enter the data into your calculator and run the 1-var stats
      Mean:_____ Med:_____ Mode:_____ Range:_____

      Min:_____ Q1:_____ Med:_____ Q3:_____ Max:_____

   b. Is the Mean or the Median a better representation of the data set? Explain why: