HOMEWORK 3 Due: Feb.3

1. A set of data are put in numerical order, and a statistic is calculated that divides the data set into two equal parts with one part below it and the other part above. Which of the following statistics was computed?

   a. mean
   b. interquartile range
   c. range
   d. median

2. The following stemplot represents the yearly percentage increases in a college's comprehensive fees over the past 22 years. (3 | 8 means that one year had a percentage increase of 3.8%.)

3	8
4	
5	2 8
6	2 7 7 7 9
7	2 4 7 8 9
8	1
9	1 5 7 9
10	1
11	
12	6
13	6
14	
15	
16	6

a. If one were asked to give the value of the center of this data, is there a single obvious correct answer? Explain.

No, since the distribution is not close to symmetric. We could only say that the data is centered somewhere in or near the 70's.

b. Find both the median and mean yearly percentage increase.

Since there are 22 values, the median will split the data into two halves containing the lowest 11 values and the highest 11 values. Thus the median will be the midpoint of the 11th and 12th values: M = (77+78)/2 = 77.5. Thus, the median is 7.75%.
The mean $\overline{X}$ = (38 + 52 + 58 + … + 166) ÷ 22 = 1854 ÷ 22 = 84.2. That is, 8.42%

c. Which is larger, the median or the mean? Is this what you would have expected based on the stemplot? Explain.

The mean is larger. Yes, we would have expected this since the distribution is right skewed.

d. Find the range of the data. Make sure to give a single number for your answer.

16.6% – 3.8 %= 12.8%

e. Find the first and third quartiles, and the interquartile range.

Q1 is the median of the lower 11 values, so it is the 6th smallest value. Counting from the top of the stemplot down we find this value to be 6.7%. Q3 is the median of the higher

11 values, so it is the 6th largest value. Counting from the bottom of the stemplot up we find this value to be 9.7%. Finally, the IQR = Q3 – Q1 = 9.7 – 6.7 = 3.0%.

f. Fill in the blanks below to give an appropriate interpretation of the IQR:

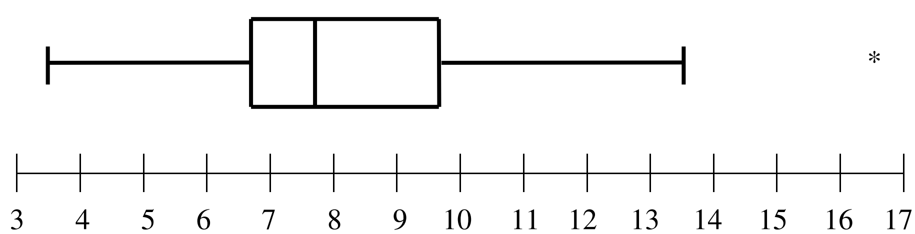
In about ___half__ of the past 22 years, the percentage increase in the college's comprehensive fees was __between__ __6.7%___ and __9.7%___.

g. Determine if the data contains any outliers as judged by the 1.5(IQR) criterion.

Q1 – 1.5(IQR) = 6.7 – 1.5(3.0) = 2.2. There are no values below this point.

Q3 + 1.5(IQR) = 9.7 + 1.5(3.0) = 14.2. The value 16.6 is above this, so according to the 1.5(IQR) criterion it is an outlier, the only one in the data set.

h. Construct a boxplot of the data. Make sure to show any outliers with asterisks.



3. Would it be more desirable for variability to be high or low for each of the following cases? Explain your decision.

   a. Age of trees in a national forest
   High; a healthy forest has trees of all ages.

   b. Diameter of new tires coming off one production line
   Low; the manufacturer would like each tire to be nearly identical in size.

   c. Weight of a box of cereal
   Low; the manufacturer would like each box to be nearly identical in weight.

   d. Prices of cars available at a used car lot
   High, cars in better shape should have higher prices than very old cars in bad shapes.

5. For what kinds of variables are side-by-side boxplots be appropriate?

   a. categorical only
   b. quantitative only
   c. one categorical and one quantitative
   d. varies according to situation

6. For what kinds of variables are a histogram be appropriate?

   a. categorical only
   b. quantitative only
   c. one categorical and one quantitative
   d. varies according to situation

7. Why is the median considered to be a RESISTANT measure of center?

   Because outliers hardly affect it (sometimes not at all).

7. Pretend you are constructing a histogram for describing the distribution of salaries for all employed people in California.

   a. What is on the Y-axis? Explain.

      The frequencies; that is, the numbers of people in each salary category.

   b. What is on the X-axis? Explain.

      A scale that covers the entire range of salaries in the data set.

   c. What would be the probable shape of the salary distribution? Explain why.

      Skewed to the right. There would be a substantial number of individuals at or near the lowest salary level, but some people make a tremendously high salary.

8. In 1798 the English scientist Henry Cavendish measured the density of the earth. Since it was difficult at that time to measure this density accurately, he repeated his work 29 times. Here are the results (the data represent the ratio of the earth's weight to an equivalent volume of water):

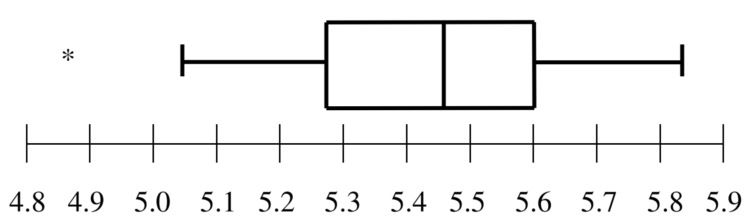
5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65	
5.51	5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85		

a. Make a stem and leaf plot of the values.

4.8	**8**
4.9	
5.0	**7**
5.1	**0**
5.2	**6 7 9 9**
5.3	**0 4 4 6 9**
5.4	**2 4 6 7**
5.5	**0 1 5 7 8**
5.6	**1 2 3 5 8**
5.7	**5 9**
5.8	**5**

b. Determine the five-number summary for the data. Use three decimal places in your answers. Then, construct a boxplot, showing outliers if any. You must show all calculations needed to determine if there are outliers.

Since (N+1)/2 = (29+1)/2 = 15,the median M is the 15th smallest value, and Q1 is the median of the 14 data values that are smaller than M and Q3 is the median of the 14 data values that are larger than M. So Q1 is the average of the 7th and 8th smallest values and Q3 is the average of the 7th and 8th largest values. So we have:
Min = 4.88 Q1 = (5.29+5.30)/2 = 5.295 M = 5.46 Q3 = (5.61+5.62)/2 = 5.615
Max = 5.85



c. Write a COMPLETE description of the distribution (include appropriate measures of center and of variability in your description).

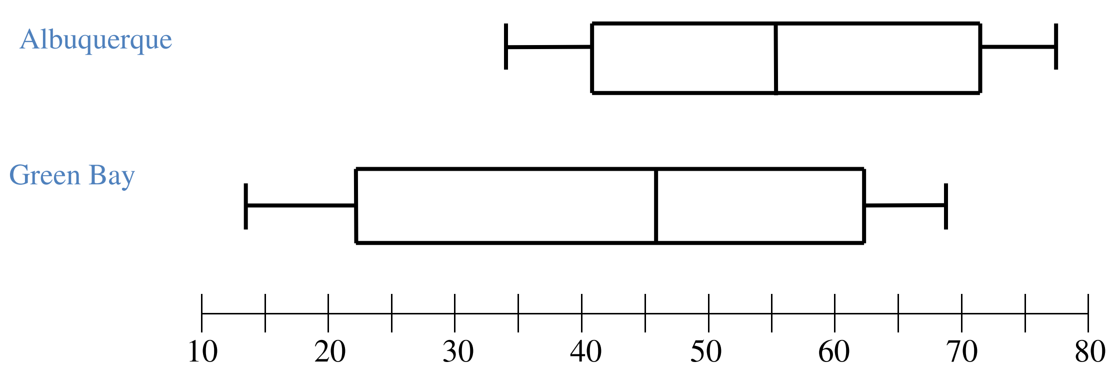
All but one of the measured densities is between 5 and 5.85, with a median density of 5.46. There is one marginal outlier at 4.88. Aside from the outlier the distribution is quite symmetric. The IQR is 0.32 (5.615 – 5.295), so half of the measured densities were within 0.32 of each other.

9. The table below shows the average monthly temperatures for Albuquerque, New Mexico and Green Bay, Wisconsin.

	Albuquerque	Green Bay
Jan.	34	14
Feb.	40	18
Mar.	46	30
Apr.	55	44
May.	64	55
Jun.	74	64
Jul.	78	69
Aug.	75	67
Sep.	68	59
Oct.	57	48
Nov.	44	34
Dec.	35	20

a. Make side-by-side boxplots of the two cities' temperature distributions.

The five-number summary for Albuquerque:
Min: 34 Q1: 41 Median: 56 Q3: 72.5 Max: 78
The five-number summary for Green Bay:
Min: 14 Q1: 22.5 Median: 46 Q3: 62.75 Max: 69



b. Summarize the key differences and similarities in the temperature distributions of the two cities, using statistical terminology.

The boxplots show clearly that Albuquerque is a much warmer city than Green Bay, with a median temperature 10 degrees higher in Albuquerque than in Green Bay. There is a greater range of temperatures in Green Bay than in Albuquerque—although Green Bay is about 10 degrees cooler than Albuquerque in the warmer months, it is about 20 degrees colder than Albuquerque in the colder months. The temperature distributions in both cities are roughly symmetrical with no outliers.