# Quiz 1.1A   AP Statistics   Name:

Below is some information about films for the "Best Movie" Academy Award (Oscar) in 2009.

Name	Genre	Budget (millions of dollars)	Total # of Oscar Nominations	Running time (minutes)	MMPA Rating
Avatar	Adventure	237	9	162	PG-13
The Blind Side	Drama	29	2	128	PG-13
District 9	Action	30	4	112	R
An Education	Drama	7	3	95	PG-13
The Hurt Locker	Action	11	9	131	R
Inglourious Basterds	Drama	70	8	153	R
Precious	Drama	10	6	110	R
A Serious Man	Comedy	7	2	106	R
Up	Animated	175	5	96	PG
Up in The Air	Comedy	30	6	109	R

1. What are the individuals in this data set?

2. Identify the variables that were recorded, and indicate whether each one is categorical or quantitative.

3. Here is a pie chart for the distribution of the variable "MMPA rating." Fill in the blanks with the appropriate values of the variable.
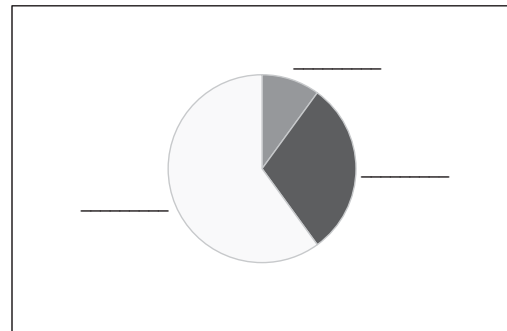
4. Below is a graph showing the total number of Oscar nominations for the four films that had PG or PG-13 ratings. What's wrong with the way the information is presented in this graph?
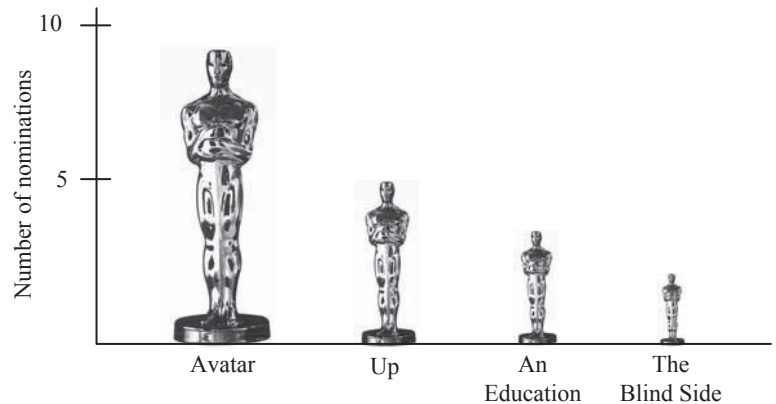
Researchers looking at the relationship between the type of college attended (public or private) and achievement gather the following data on 3265 people who graduated from college in the same year. The variable "management level" describes their job description 20 years after graduating from college.

| | | Type of College | |
		Public	Private
Management level	High	75	107
	Medium	962	794
	Low	732	595

5. Calculate the marginal distribution of management level in percents.

6. Find the conditional distribution of management level for each college type, in percents.

7. Sketch a segmented bar graph for the two conditional distributions in **6**.
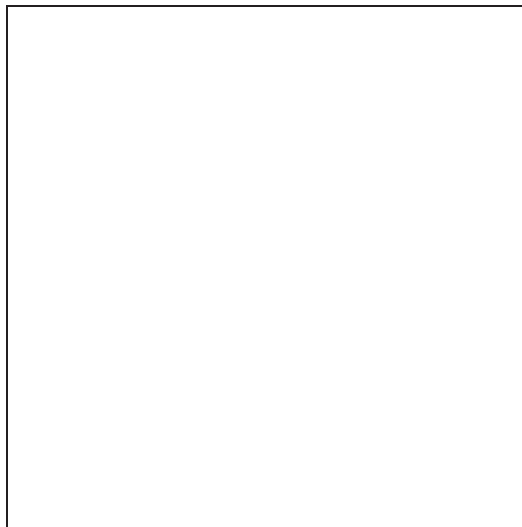
8. Write a brief description of what the information in **6.** and **7.** tell you about the relationship between these variables.

Literary scholars sometimes use the distribution of word lengths in a work as a test of authenticity. Here are the words lengths for the first 25 words on a randomly-selected page from Toni Morrison's *Song of Solomon*.

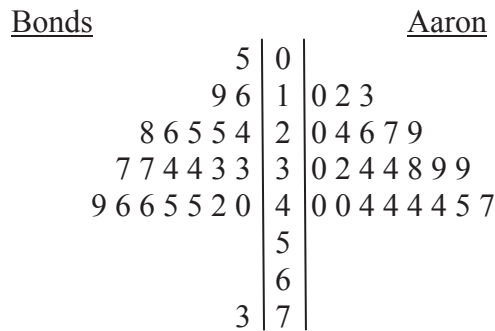| 2	3	4	10	2	11	2	8	4	3	7	2	7
5	3	6	4	4	2	5	8	2	3	4	4	

1. Make a dotplot of these data.

2. Describe the overall pattern of the distribution and any possible outliers.

3. The histogram below shows the number of major hurricanes that reached the East Coast of the United States from 1944 to 2005. Describe the shape, center, and spread of the distribution.
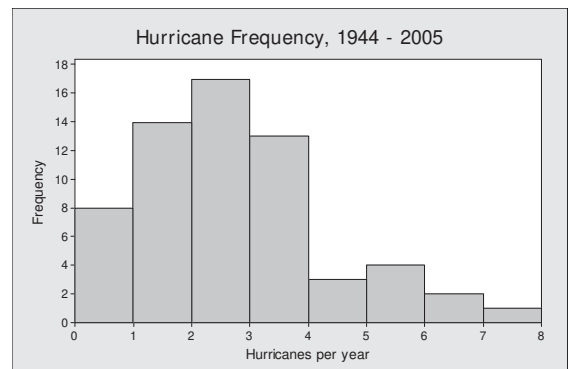
On August 7, 2007 Barry Bonds hit his 756th home run, breaking the all-time career home run record, formerly held by Hank Aaron. Does that make Bonds a better home run hitter than Aaron? Let's compare their annual home run production over their entire careers. Below is a side-by-side stemplot. (Bonds played between 1986 and 2007. Aaron played between 1954 and 1978.)

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

4. Use the plot to write a few sentences comparing Bonds and Aaron as home run hitters.

How much oil wells in a given field will ultimately produce is key information in deciding whether to drill more wells. Here are the estimated total amounts of oil recovered from 38 wells in the Devonian Richmond Dolomite area of the Michigan basin, in thousands of barrels. The data is provided in ascending order, along with a dotplot.

3	22	35	43	49	57	70	92
13	25	35	43	50	59	70	98
15	31	37	45	50	63	74	157
19	33	37	46	53	65	80	
21	35	38	48	56	66	82	

Total Oil Recovered (thousands of barrrels)
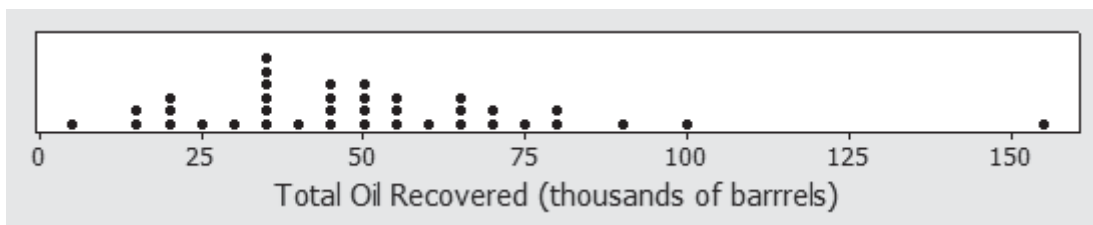
1. What measures would you use to describe the center and spread of these data? Justify your answer.

2. Find the five-number summary for these data.

3. Are there any outliers? Justify your answer.

4. Draw a boxplot of this distribution.

5. For the oil well data on the previous page, how can you tell *without doing any calculations*, that the mean of these data is larger than the median?

6. Five students reported the amount of time (in minutes) they spent studying for an AP Statistics test the night before the test. Here are the results:

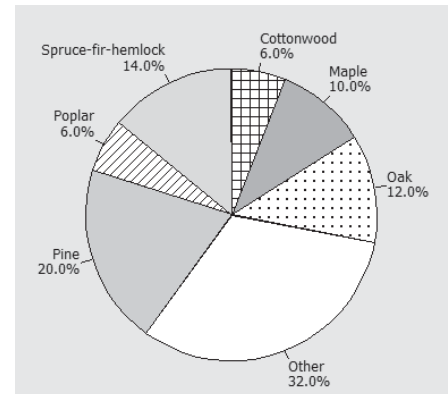$$45 \quad 50 \quad 60 \quad 65 \quad 80$$

Calculate the mean and standard deviation of study time *using the formula* for each. Show your work!

7. Below are dotplots for three small datasets: A, B, and C. Without performing any calculations, rank the standard deviations of the datasets from lowest to highest. Justify your answer.
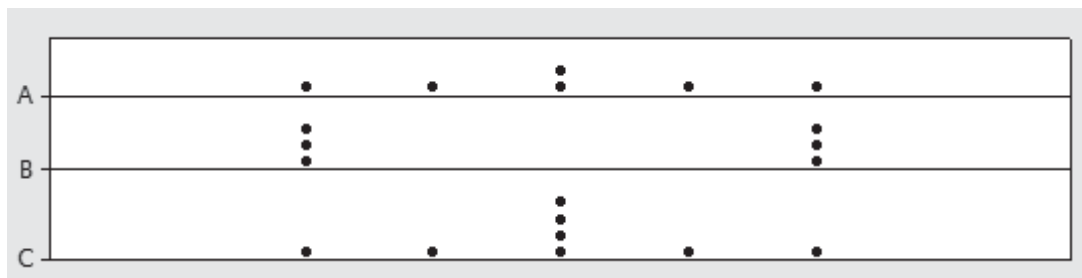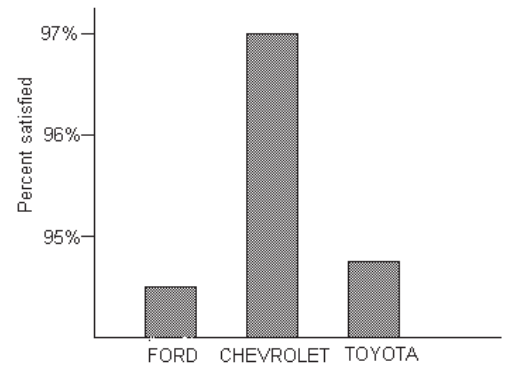
Part 1: Multiple Choice. *Circle the letter corresponding to the best answer.*

1. The pie chart at right describes the distribution of state
   tree types for the 50 states in the United States. The
   category "Other" include all trees that are the state tree
   for two or fewer states. Which of the following
   conclusions can we draw from this chart?
   (a) Some states have not designated a "state tree."
   (b) The cottonwood is the state tree for 12 states.
   (c) Taken together, oak, pine, and maple are the state
       trees for more than half the states.
   (d) There are 10 states that have designated a pine as
       their state tree.
   (e) There is no state that has designated the Eastern
       Red Cedar as its state tree.

2 The following bar graph gives the percent of owners of three brands of trucks who are
   satisfied with their truck. From this graph, we may conclude that
   (a) owners of other brands of trucks are less satisfied
       than the owners of these three brands.
   (b) Chevrolet owners are substantially more satisfied
       than Ford or Toyota owners.
   (c) there is very little difference in the satisfaction of
       owners for the three brands.
   (d) Chevrolet probably sells more trucks than Ford or
       Toyota.
   (e) a pie chart would have been a better choice for
       displaying these data.

3. Here are the IQ test scores of 10 randomly chosen fifth-grade students:
   145   139   126   122   125   130   96   110   118   118
   To make a stemplot of these scores, you would use as stems
   (a) 0 and 1.
   (b) 09, 10, 11, 12, 13, and 14.
   (c) 96, 110, 118, 122, 125, 126, 130, 139, and 145.
   (d) 0, 2, 3, 5, 6, 8, 9.
   (e) None of the above is a correct answer.

4. If a distribution is skewed to the right, which of the following is true?
   (a) The mean must be less than the median.
   (b) The mean and median must be equal.
   (c) The mean must be greater than the median.
   (d) The mean is either equal to or less than the median,
   (e) It's impossible to tell which of the above statements is true without seeing the data.

5. Rainwater was collected in water collectors at 30 different sites near an industrial complex and the amount of acidity (pH level) was measured. The data ranged from pH 2.6 to pH 6.3. The following stemplot of the data was constructed.
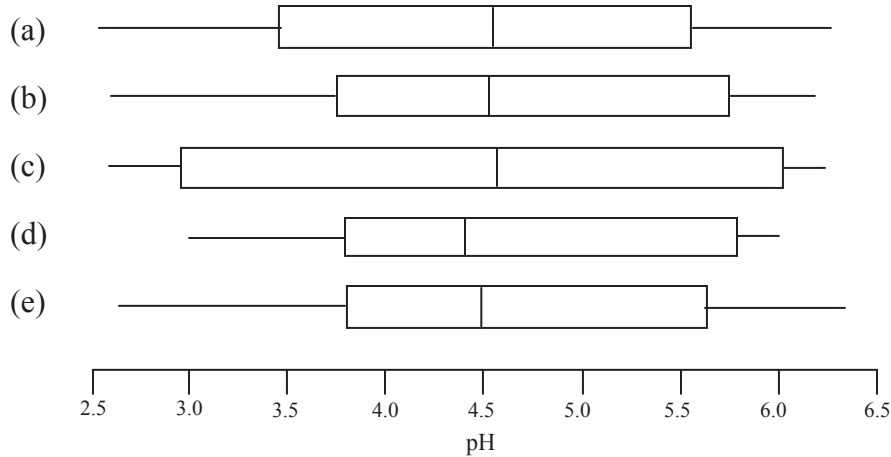
```
2 679
3 237789            Key: 3|7 = pH 3.7
4 1222446899
5 0556788
6 0233
```

Which of the following boxplots is a correct representation of the same distribution?

(a)

(b)

(c)

(d)

(e)

```
      2.5    3.0    3.5    4.0    4.5    5.0    5.5    6.0    6.5
                                  pH
```

6. A sample of 250 high school students were asked, "If you had $1000 to contribute to one kind of charitable organization, which type of organization would you choose? Below is a two-way table of responses to this question and gender.

Gender	Organization				
	Education	Environment	Health	International Aid	Other
Female	19	33	50	28	10
Male	23	29	28	17	13

Which of the following conclusions seems to be supported by the data?
(a) Most of the females who chose a health organization would have chosen an environmental organization as their second choice, had they been asked.
(b) There is no association between gender and choice of organization.
(c) The proportion of males who said they would contribute to an environmental organization was higher than the proportion of females who said they would contribute such an organization.
(d) None of the students surveyed said they would contribute to religious organizations.
(e) The marginal distribution of Organization is 140, 110.

7. A small company that prints custom t-shirts has 6 employees, one of whom is the owner and manager. Suppose the owner makes $120,000 per year and the other employees make between $40,000 and $50,000 per year. One day, the owner decides to give himself a $30,000 raise. Which of the following describes how the company's mean and median salaries would change?
(a) The mean and median would both increase by $5,000.
(b) The mean would increase by $5,000 and the median would not change.
(c) The mean would increase by $6,000 and the median would not change.
(d) The median would increase by $6,000 and the mean would not change.
(e) The mean would increase by $6,000, but we cannot determine the change in the median without more information.

8. The mean speed of vehicles in the "cars only" lanes of the New Jersey turnpike is 68 miles per hour. The mean speed of vehicles in the "any vehicle" lanes is 64 miles per hour. What must be true about the mean speed of all vehicles on the turnpike, assuming these are the only types of lanes?
(a) It could be any number between 64 and 68 miles per hour.
(b) It must be larger than the median speed.
(c) It must be larger than 66 miles per hour.
(d) It must be 66 miles per hours.
(e) We don't have enough information to draw any conclusion about the mean speed of all vehicles.

9. The mean birth weight of infants born at a certain hospital in the month of April was 128 oz. with a standard deviation of 10.2 oz. Which of the following is a correct interpretation of standard deviation?
(a) All the infants born in April weighed between 117.8 oz. and 138.2 oz.
(b) About half the infants born in April weighed between 117.8 oz. and 138.2 oz.
(c) The difference between the mean weight and the median weight of infants born in April was 10.2 oz.
(d) The distance between the weight of each infant bon in April and the mean weight was, on average, about 10.2 oz.
(e) The mean weight of infants born in subsequent months is likely to be within 10.2 oz. of the mean weight in April.

10. A medical researcher collects health data on many women in each of several countries. One of the variables measured for each woman in the study is her weight in pounds. The following list gives the five-number summary for the weights of adult women in one of the countries.

Country A: 92, 110, 120, 160, 240

About what percent of Country A women weigh between 110 and 240 pounds?
(a) 50%
(b) 65%
(c) 75%
(d) 85%
(e) 95%

Part 2: Free Response

Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

We all "know" that the body temperature of a healthy person is 98.6°F. In reality, the actual body temperature of individuals varies. Here are dotplots and summary statistics of the body temperatures for 90 healthy individuals (45 males and 45 females).
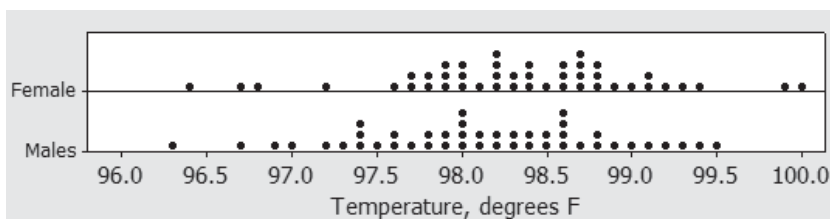

```
 Total
Variable   Count    Mean   SE Mean   StDev   Minimum       Q1   Median       Q3   Maximum
Female       45    98.344    0.111    0.745    96.400    97.900   98.400   98.800   100.000
Males        45    98.122    0.110    0.735    96.300    97.600   98.100   98.600    99.500
```

**11.** Determine if there are any outliers in each distribution. Show your work.

12. Draw parallel boxplots of these two distributions. Be sure to label the plots and provide a scale.

_____

13. Write a few sentences comparing the body temperatures of healthy males and females.

The data below is the number of unprovoked attacks by alligators on people in Florida each year for a 33-year period.

6	12	2	4	17	4	6	10	3	9	13
9	15	14	6	18	1	9	6	6	11	24
14	14	5	17	17	5	13	22	20	3	5

14. Construct a histogram for this distribution. Choose an appropriate bin width, and be sure to provide a label and scale for each axis.
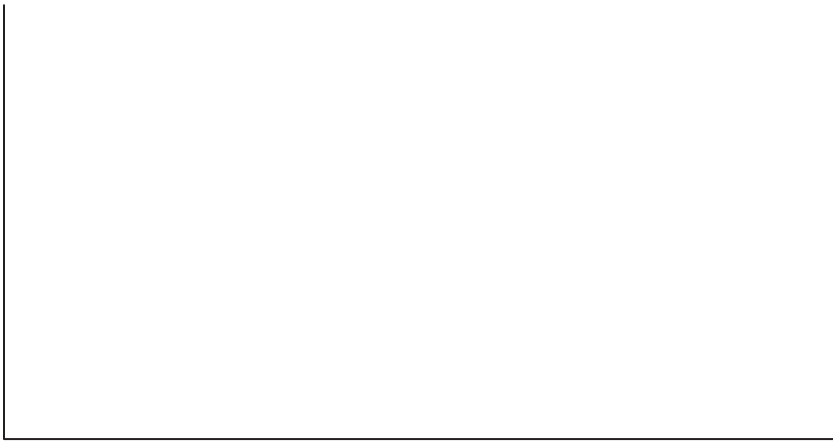
15. Based on your histogram, what numerical measures of center and spread would be best to use for this distribution? Explain your choice.

The Practice of Statistics, 4/e- Chapter 1

Do people's opinion of the alt-rock band Coastal Surface depend on their age? Here are some data that address this question.

Age	Opinion of band			Total
	Like	Dislike	Uncertain	
11-14 years	18	30	15	63
15-18 years	38	10	17	65
19-22 years	40	67	15	122
Total	96	107	47	250

16. Calculate three conditional distributions for opinion of the band, one for each age group. You may present your results in either a table or a graph.

17. Discuss the relationship between age and opinion of the band in two or three sentences.