Welcome to AP Statistics! I am so looking forward to having you all in class and meeting some of you for the first time. The following packet contains the summer assignment students need to complete by the first day of school. We will use some time in each of the first few classes to answer questions as we move forward with the course. This packet will be collected, graded and worth points towards your first semester grade. You will also be tested on this material as you will see it throughout the course. Note that this is the **most elementary** of topics in Statistics and the foundation of Descriptive Statistics. You can assess after this packet if AP Statistics is right for you.

This packet contains a section of notes and 35 exercises. Please complete these thoroughly and neatly if you want to receive full credit. Any questions that ask for explanations should be answered with complete sentences as that is what is expected on the AP exam. All graphs should be properly labeled and scaled.

A graphing calculator is needed to complete some problems. TI-83 Plus or TI-84 Plus (the CE is my favorite but not necessary). If you do not have your own, I can let you borrow one for the summer.

If you have questions over the summer while working on the packet, there are several ways to receive assistance.

These are not my videos or notes, but this instructor does a fairly good job of explaining things.
https://sites.google.com/a/cusd.kahoks.org/mrg/ap-stats-lecture-videos/chapter-1
You can also search YouTube, and Kahn Academy for more information. I want you to dig a little ☺

You can also contact me over the summer.
Email – c.smith@smmusd.org

RECOMMENDED READING:

I also recommend you read *"How to Lie with Statistics"* **by Darrell Huff (ISBN: 978-0-393-31072-6)**

It will really benefit you before we begin the course.

Again, I look forward to having you in class and I will see you in August. Please don't hesitate to contact me at the above email.

Happy Summer! Ms. Smith ☺

Statistics is the science of data. This first chapter focuses on how to display the data in graphs and use basic forms of analysis. We will start with some vocabulary:

Individuals are the objects described by a set of data. We want to use the data from small groups (***sample***) in order to draw conclusions about a large group (***population***). Individuals may be people, animals, or things.

In an ***observational study***, we observe individuals and measure variable of interest but do not attempt to influence the responses.

In an ***experiment***, we deliberately do something to individuals in order to observe their response.

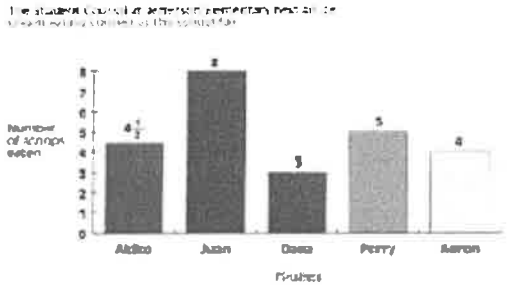
A ***variable*** is any characteristic of an individual. A variable can take different values for different individuals.

A ***categorical variable*** places an individual into one of several groups or categories. (*Example* – gender, eye color, grade in school, etc.) These are not numerical (except for grade in school, which can be thought of as freshman or sophomore instead of 9th, 10thgrade).

A ***quantitative variable*** takes numerical values for which arithmetic operations such as adding and averaging make sense.

The ***distribution*** of a variable tells us what values the variable takes and how often it takes these values.

Bar graphs and ***pie charts*** are reserved only for categorical data.

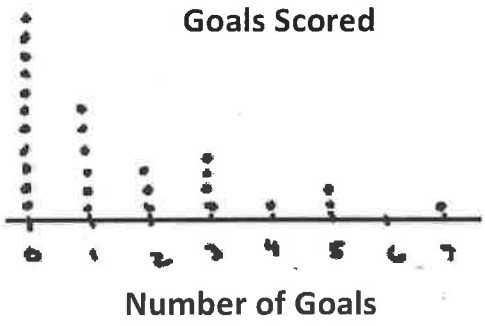


↑ notice, bars do not touch (x-axis is categories)

A ***dotplot*** is another type of graph, but it is used to display quantitative data. Each dot represents one value from the data set.

Example – The number of goals scored by each team in the first round of the California Southern Section Division V high school soccer playoffs is shown below:

5	0	1	0	7	2	1	0	4	0	3	0	2	0
3	1	5	0	3	0	1	0	1	0	2	0	3	1



Constructing Dotplots:

- Label your axis and title your graph. Draw a horizontal line and label it with the variable.
- Scale the axis based on the values of the variable.
- Mark a dot above the number on the horizontal axis corresponding to each data value.

Overall Pattern of a Distribution

To describe the overall pattern of a distribution:

- Give the center and the spread
- See if the distribution has a simple shape that you can describe in a few words

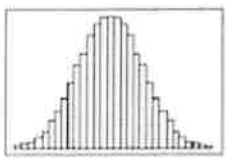
An ***outlier*** in any graph is an individual observation that falls outside the overall pattern

A distribution is ***symmetric*** if the right and left sides of the histogram are approximately mirror images of each other.

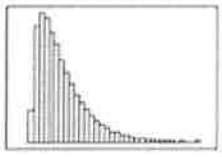
A distribution is ***skewed to the right*** if the right side of the histogram extends much farther out than the left side (Tail is to the right)

A distribution is ***skewed to the left*** if the left side of the histogram extends much farther out than the right side (Tail is to the left)

Skewed to the Left



Symmetric Bell shaped

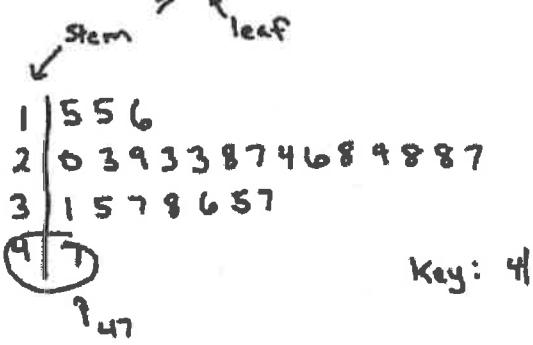


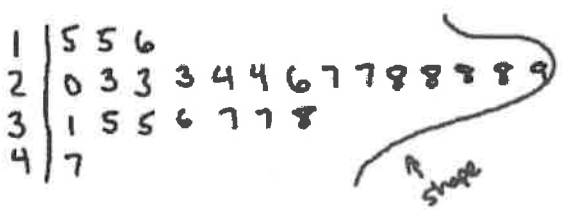
Skewed to the Right

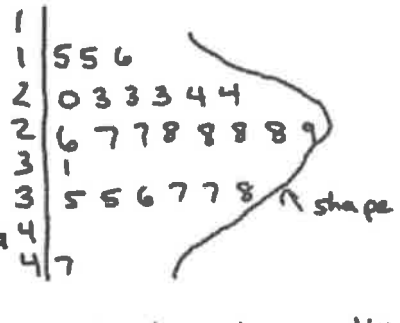
Constructing Stemplots:

- Separate each data value into a stem and a leaf. The leaf must be one digit. (So the value 45 would have a stem of 4 and a leaf of 5)
- Write the stems vertically in increasing order from the top to bottom.
- Go through the data writing each leaf to the right of its stem.
- Rewrite the graph in order and add a key.

ex) Create a stemplot for the number of mg of caffeine in several different brands of soda: 20 15 23 29 23 15 23 31 28 35 37 27
24 26 47 28 24 28 28 16 38 36 35 37 27







When there are too many values in one line you can split-stems. The first half gets values ending in (0-4) and the second half gets (5-9)

Measures of Center

The ***mean*** (average) is *not* a ***resistant measure***. This means that it cannot resist the influence of extreme observations (outliers). The notation for mean is $\bar{x}$, pronounced "x-bar".

$$mean(average) \rightarrow \bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

x_i *represents each observation;* n *represents the number of observations*

The ***median*** (M) is the midpoint of the distribution. M is the number such that half of the observations are smaller and half are larger. To find the median, arrange all observations in order of size, from smallest to largest.

1	3	4	5	9
$M = 4$				

1	3	4	5	7	9
$M = 4.5$ *If there is an even number of values, the median is the average of the 2 middle numbers.*					

Measures of Spread

The ***range*** is a measure of spread calculated by finding the difference between the largest and smallest observations. ($range = max. - min.$)

Other measures of spread include ***interquartile range*** and ***standard deviation***.

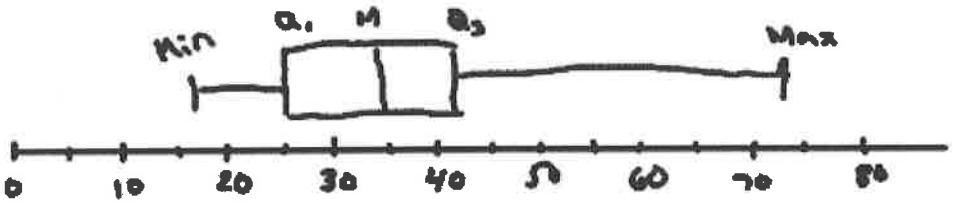
The ***first quartile*** (Q_1) marks the first quarter of the data (half of the first half). The ***third quartile*** (Q_3) marks the third quarter of the data (half of the second half).

The ***interquartile range*** (IQR) is the difference between the first and third quartiles. ($IQR = Q_3 - Q_1$)

			quarter				half				quarter				
16	19	24	25	25	33	33	34	34	37	37	40	42	46	49	73
			$Q_1 = 25$				$M = 34$				$Q_3 = 41$				
			$range = 73 - 16 = 57$						$IQR = 41 - 25 = 16$						

The ***five-number summary*** of a data set consists of the smallest observation (min.), the first quartile (Q_1), the median (M), the third quartile (Q_3), and the largest observation (max.), written in order from smallest to largest.

This leads to a graph called a ***boxplot***. It plots the values of the five-number summary as follows:



The ***variance*** (s^2) of a set of observations is the average of the squares of the deviations of the observations from their mean. In symbols, the variance of n observations $x_1, x_2, x_3, \ldots, x_n$ is:

$$s^2 = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

The ***standard deviation*** (s) is the square root of the variance (s^2):

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

The standard deviation measures spread by looking at how far the observations are from the mean.

All of these calculations can be found on your calculator. To enter data, push STAT then EDIT.

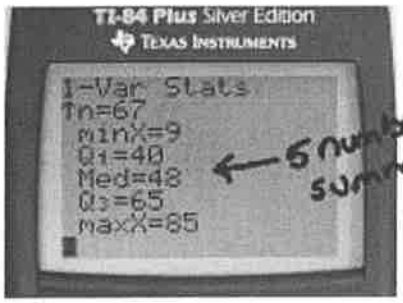
Once you have typed in your data into one of the lists, push STAT again, go over to CALC, and select 1-VAR-STATS.

This will print out the mean and the standard deviation (Sx). If you scroll down you will see the quartiles and the median as well.

mean

standard deviation

down arrow to see more



5 number summary

Practice Problems

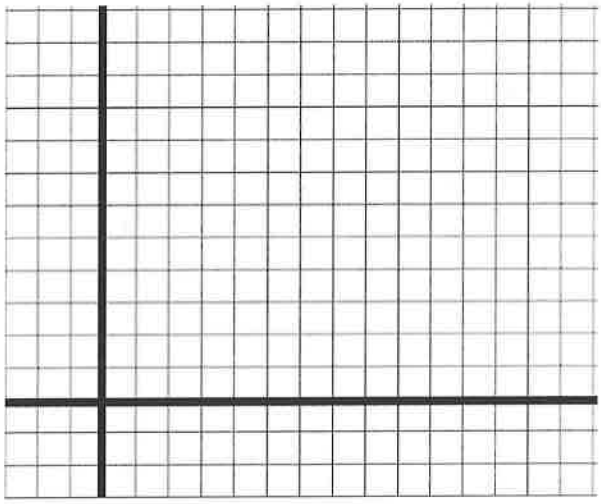
1) Data from a medical study contain values of many variables for each of the people who were the subjects of the study. Which of the following variables are categorical and which are quantitative:

a) Gender (female or male)	
b) Age (years)	
c) Race	
d) Smoker (yes or no)	
e) Systolic Blood Pressure	
f) Level of calcium in the blood	

2) Here are data on the percent of females among people earning doctorates in 1994 in several fields of study:

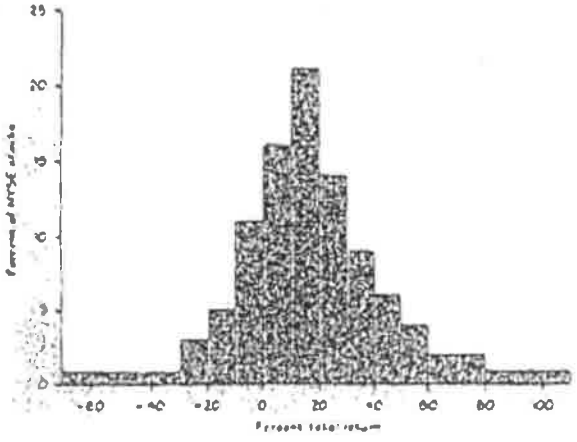
Computer Science – 15.4%	Life Sciences – 40.7%	Education – 60.8%
Physical Sciences – 21.7%	Engineering – 11.1%	Psychology – 62.2%

a) Present this data in a well-labeled bar graph.



b) Would it also be correct to use a pie chart to display these data? If so, construct a pie chart. If not, explain why not.

3) The total return on a stock is the change in its market price plus any dividend payments made. Total return is usually expresses as a percent of the beginning price. Figure 1.10 is a histogram of the distribution of total returns for all 1528 stocks listed on the New York Stock Exchange in one year. It is a histogram of the percents in each class rather than a histogram of counts:



a) Describe the overall shape of the distribution of total returns.

b) What is the approximate center of the distribution?

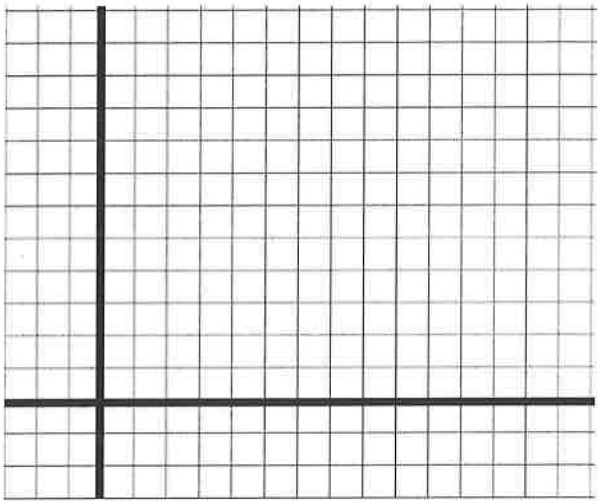
c) Approximately what were the smallest and largest total returns?

d) A return less than zero means that an owner of the stock lost money. About what percent of all stocks lost money?

4) Here are data on the rate of deaths from cancer (deaths per 100,000 people) in the United States over the 50-year period from 1945 to 1995:

Year:	1945	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995
Deaths:	134.0	139.8	146.5	149.2	153.5	162.8	169.7	183.9	193.3	203.2	204.7

a) Construct a line graph (time plot) for these data.



b) Describe what you see in a few sentences.

c) Do these data suggest that we have made progress in treating cancer? Explain.

5) Joey's first 14 quiz grades in a marking period were:

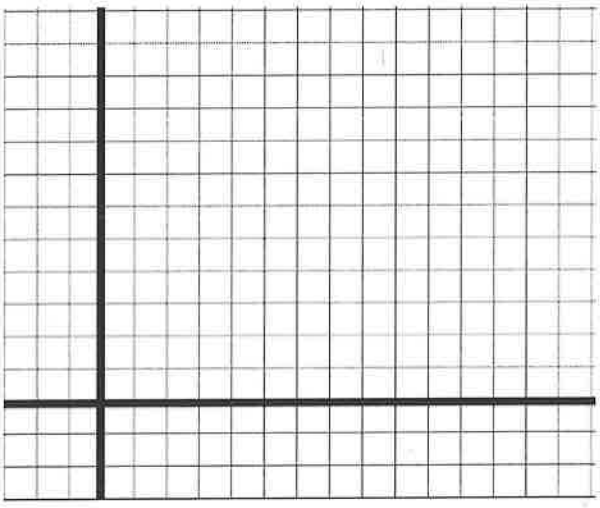
86	84	91	75	78	80	74	87	76	96	82	90	98	93

a) Calculate the mean. Check using "1-var stats" on your calculator.

$\bar{x} =$

b) Suppose Joey had an unexcused absence for the 15th quiz and received a 0. Determine his final quiz average. What property of the mean does this situation illustrate? Write a sentence about the effect of the zero on Joey's quiz average.

c) What kind of plot would best show Joey's distribution of grades? Assume an 8-point grading scale (A: 93-100, B: 85-92, C: 77-84, D: 70-76). Make an appropriate plot and justify your choice.



6) Suppose a major league baseball team's mean yearly salary for a player is $1.2 million and that the team has 25 players on its active roster. What is the team's annual payroll for players? If you knew only the median salary, would you be able to answer the question? Why or why not?

7) The distribution of individual incomes in the US is strongly skewed to the right. In 1997, the mean and median incomes of the top 1% of Americans were $330,000 and $675,000. Which of these numbers is the mean and which is the median? Explain.

8) The level of various substances in the blood influences our health. Here are the measurements of the level of phosphate in the blood of a patient made on 6 consecutive visits to a clinic:

5.6	5.2	4.6	4.9	5.7	6.4

a) Calculate the mean.

b) Use the calculator to obtain $\bar{x}$ and s.

9) New York Yankee Roger Maris held the single-season home run record from 1961 until 1998. Here are Maris' home run counts for his 10 years in the American League:

15	28	16	39	61	33	23	26	8	13

a) Maris' mean number of home runs is $\bar{x} = 26.2$. Calculate the standard deviation by hand.

b) Use your calculator to find the mean and standard deviation for the 9 observations that remain when you leave out the outlier. How does the outlier affect the values of $\bar{x}$ and s. Is s a resistant measure of spread?

10) This is a standard deviation contest. You must choose four numbers from the whole numbers 0 to 10, with repeats allowed.

a) Choose four numbers that have the smallest possible standard deviation.

b) Choose four numbers that have the largest possible standard deviation.

c) Is more than one choice possible for either (a) or (b)? Explain.

11) Maria measures the length of 5 cockroaches that she finds at school (in inches):

1.4	2.2	1.1	1.6	1.2

a) Find the mean and standard deviation of Maria's measurements

b) Maria's science teacher is furious to discover that she has measured the lengths in inches rather than centimeters. There are 2.54 cm in 1 inch. She gives Maria two minutes to report the mean and standard deviation in centimeters. Maria succeeded. Please state her answers.

12) Athletes like Cathy Freeman, Rulon Gardner, Ian Thorpe, and Jenny Thompson captured public attention by winning gold medals in the 2000 Summer Olympic Games in Sydney, Australia. Table 1.2 displays the total number of gold medals won by several countries in the 2000 Summer Olympics.

TABLE 1.2 Gold medals won by selected countries in the 2000 Summer Olympics

Country	Gold medals	Country	Gold medals
Sri Lanka	0	Netherlands	12
Qatar	0	India	0
Vietnam	0	Georgia	0
Great Britain	28	Kyrgyzstan	0
Norway	10	Costa Rica	0
Romania	26	Brazil	0
Switzerland	9	Uzbekistan	1
Armenia	0	Thailand	1
Kuwait	0	Denmark	2
Bahamas	1	Latvia	1
Kenya	2	Czech Republic	2
Trinidad and Tobago	0	Hungary	8
Greece	13	Sweden	4
Mozambique	1	Uruguay	0
Kazakhstan	3	United States	39

Source: BBC Olympics Web site.

Make a dotplot to display these data. Describe the distribution of number of gold medals won.

13) A marketing consultant observed 50 consecutive shoppers at a supermarket. One variable of interest was how much each shopper spent in the store. Here are the data (in dollars), arranged in increasing order:

3.11	8.88	9.26	10.81	12.69	13.78	15.23	15.62	17.00	17.39
18.36	18.43	19.27	19.50	19.54	20.16	20.59	22.22	23.04	24.47
24.58	25.13	26.24	26.26	27.65	28.06	28.08	28.38	32.03	34.98
36.37	38.64	39.16	41.02	42.97	44.08	44.67	45.40	46.69	48.65
50.39	52.75	54.80	59.07	61.22	70.32	82.70	85.76	86.37	93.34

a) Round each amount to the nearest dollar. Then make a stemplot using the tens of dollars as the stem and dollars as the leaves.

b) Make another stemplot of the data by splitting stems. Which of the plots shows the shape of the distribution better?

c) Describe the shape, center, and spread of the distribution. Write a few sentences describing the amount of money spent by shoppers at this supermarket.

14) Here are the scores on the Survey of Study Habits and Attitudes (SSHA) for 18 first-year college women:

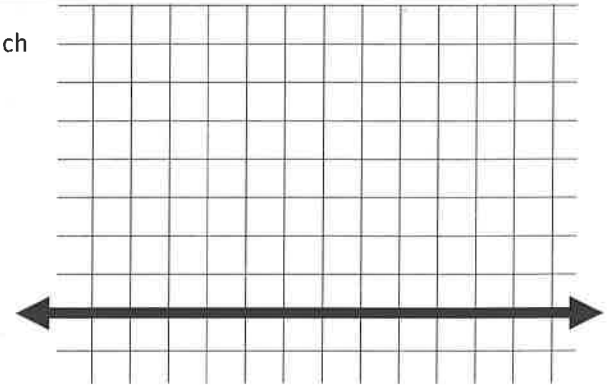
154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	128	200	148

And for 20 first-year college men:

108	140	114	91	180	115	126	92	169	146
109	132	75	88	113	151	70	115	187	104

a) Compute 5-number summary for each distribution

b) Make side-by-side boxplots to compare the distributions.



c) Write a paragraph comparing the SSHA scores for men and women.

15) Figure 1.19 displays computer output for the data on amount spent by grocery shoppers.
   a) Determine the total amount spent by the shoppers.

DataDesk

```
Summary of   spending
No Selector

Percentile  25

        Count   50
         Mean   34.7022
       Median   27.8550
       StdDev   21.6974
          Min    3.11000
          Max   93.3400
Lower 1th %tile  19.2700
Upper 1th %tile  45.4000
```

Minitab

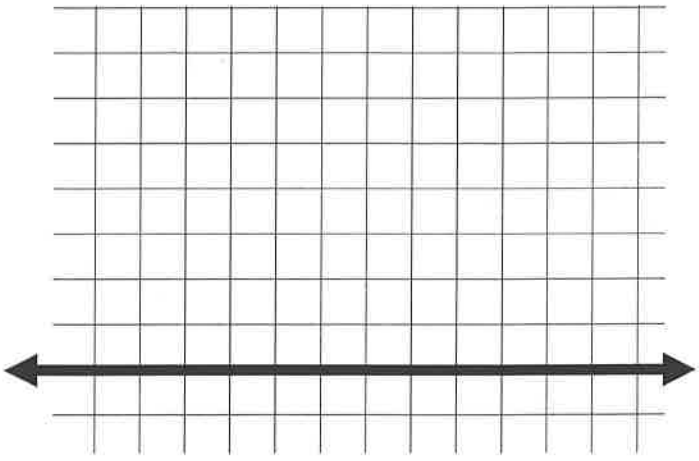
Descriptive Statistics

Variable	N	Mean	Median	TrMean	StDev	SEMean
spending	50	34.70	27.85	32.92	21.70	3.07

Variable	Min	Max	Q1	Q3
spending	3.11	93.34	19.06	45.72

FIGURE 1.19 Numerical descriptions of the unrounded shopping data from the Data Desk and Minitab software.

b) Make a boxplot from the computer output. Do you suspect there are any outliers?



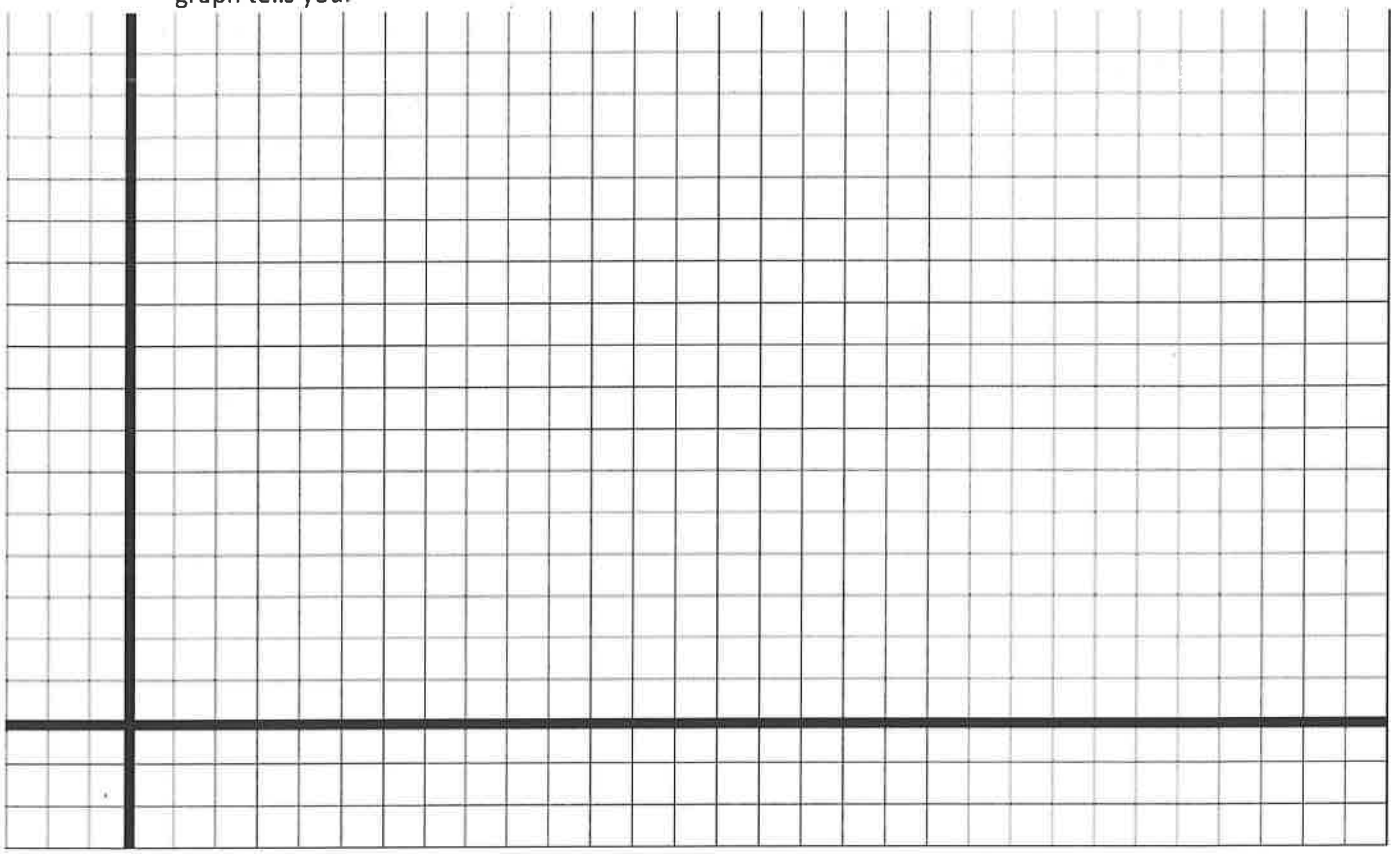
16) Has the increasing popularity of the compact disc (CD) affected sales of cassette tapes? Table 1.13 shows the number of cassettes and CDs sold from 1990 to 1999:

TABLE 1.13 Sales (in millions) of full-length cassettes and CDs, 1990–1999

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Full-length cassettes	54.7	49.8	43.6	38.0	32.1	25.1	19.3	18.2	14.8	8.0
Full-length CDs	31.1	38.9	46.5	51.1	58.4	65.0	68.4	70.2	74.8	83.2

Source: The Recording Industry Association of America, *1999 Consumer Profile*.

Make a graphical display to cassette and CD sales. Write a few sentences describing what your graph tells you.



17) The mean $\overline{x}$ and standard deviation s measure center and spread but are not a description of a distribution. Data sets with different shapes can have the same mean and standard deviation. To demonstrate this fact, use your calculator to find $\overline{x}$ and s for the following two small data sets. Then make a stemplot of each and comment on the shape of each distribution.

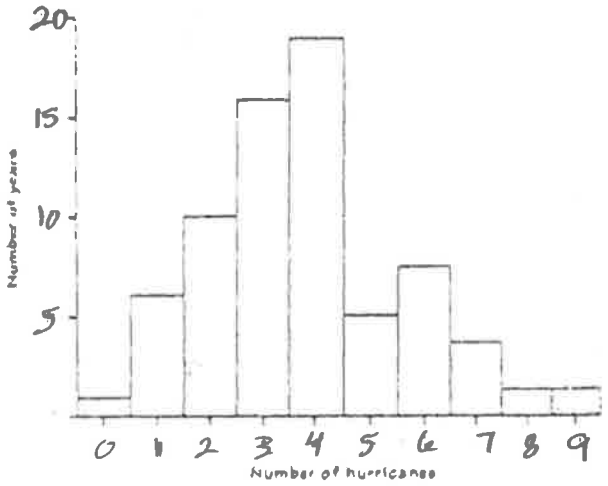
Data A:	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74
Data B:	6.58	5.76	7.71	8.84	8.47	7.04	5.25	5.56	7.91	6.89	12.50

18) Here are the percents of the popular vote won by the successful candidate in each of the presidential elections from 1948 to 2000:

Year	1948	1952	1956	1960	1964	1968	1972	1976	1980	1984	1988	1992	1996	2000
Percent	49.6	55.1	57.4	49.7	61.1	43.4	60.7	50.1	50.7	58.8	53.9	43.2	49.2	47.9

a) Make a stemplot of the winners' percents. (Round to whole numbers and use split stems.)	b) What is the median percent of the vote won by the successful candidate in presidential elections? (Work with the unrounded data.)
c) Call an election a landslide if the winner's percent falls at or above the third quarter. Find the third quartile. Which elections were landslides?	

19) The histogram in Figure 1.25 shows the number of hurricanes reaching the east coast of the United States each year over a 70-year period. Give a brief description of the overall shape of this distribution. About where does the center of the distribution lie?



20) The Survey of Study Habits and Attitudes (SSHA) is a psychological test that evaluates college students' motivation, study habits, and attitudes toward school. A private college gives the SSHA to a sample of 18 of its incoming first-year students. Their scores are below:

154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	128	200	148

a) Make a stemplot of these data. The overall shape of the distribution is irregular, as often happens when only a few observations are available. Are there any potential outliers? About where is the center of the distribution (the score with half the scores above it and half below)? What is the spread of the scores (ignoring any possible outliers)?

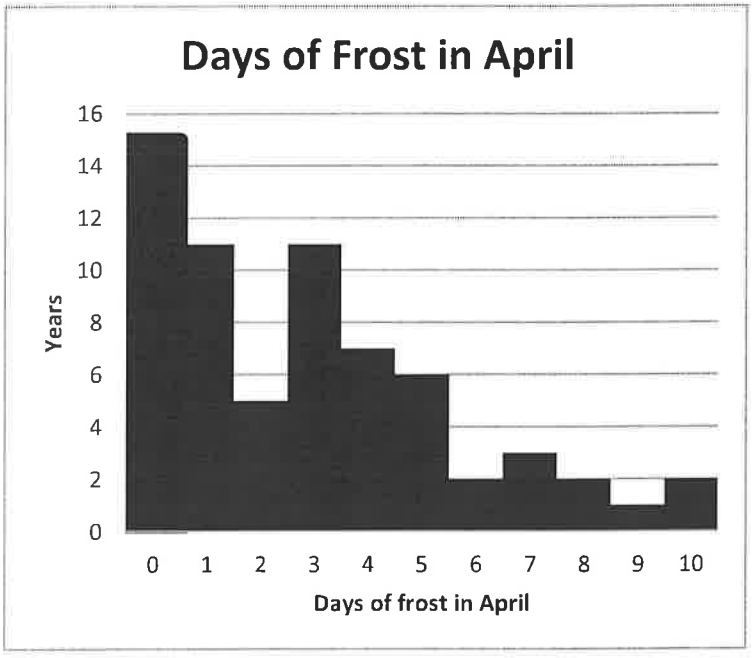
b) Find the mean score using the formula for the mean. Then enter the data into your calculator and confirm the mean using 1-Var Stats.

c) Find the median of these scores. Which is larger: the median or the mean? Explain why.

21) Is the interquartile range (IQR) a resistant measure of spread? Give an example of small data set that supports your answer.

22) This is a histogram of the number of days in the month of April on which the temperature fell below freezing at Greenwich, England. The data cover a period of 65 years.

a) Describe the shape, center, and spread of this distribution. Are there any outliers?



b) In what percent of the 65 years did the temperature fall below freezing in April?

23) The rate of return on a stock is its change in price plus any dividends paid. Rate of return is usually measured in percent of the starting value. We have data on the monthly rate of return of the stock of Wal-Mart stores for the years 1973 to 1991, the first 19 years Wal-Mart was listed on the New York Stock Exchange. There are 228 observations.

Figure 1.27 displays output from statistical software that describes the distribution of these data. The stems in the stemplot are the tens digits of the percent returns. The leaves are the ones digits. The stemplot uses split stems to give a better display. The software gives high and low outliers separately from the stemplot rather than spreading out the stemplot to include them.

```
Mean = 3.064
Standard deviation = 11.49

N = 228  Median = 3.4691
Quartiles = -2.950258, 8.4511

Decimal point is 1 place to the right of the colon.

Low: -34.04255 -31.25000 -27.06271 -26.61290

-1 : 985
-1 : 444443322222110000
-0 : 99998877766666665555
-0 : 44444444333333322222222222211111110O
 0 : 000001111111111122222233333333444444444
 0 : 55555555555555555556666666666677777778888888888999999
 1 : 000000000111111112223333444 4
 1 : 55566667889
 2 : 011334

High: 32.01920 41.80531 42.05607 57.89474 58.67769
```

FIGURE 1.27 Output from software describing the distribution of monthly returns from Wal-Mart stock.

a) Give the five-number summary for monthly returns on Wal-Mart stock.

b) Describe in words the main features of the distribution.

c) If you had $1000 worth of Wal-Mart stock at the beginning of the best month during these 19 years, how much would your stock be worth the end of the month?
If you had $1000 worth of Wal-Mart stock at the beginning of the worst month during these 19 years, how much would your stock be worth the end of the month?

d) Find the interquartile range (IQR) for the Wal-Mart data. Are there any outliers according to the 1.5*IQR criterion? Does it appear to you that the software uses this criterion in choosing which observations to report separately as outliers?

24) In 1798, the English scientist Henry Cavendish measured the density of the earth by careful work with a torsion balance. The variable recorded was the density of the earth as a multiple of the density of water. Here are Cavendish's 29 measurements.

5.50	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
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	

Present these measurements graphically in a stemplot. Discuss the shape, center, and spread of the distribution. Are there any outliers? What is your estimate of the density of the earth based on these measurements?

25) There are many ways to measure the reading ability of children. One frequently used test is the Degree of Reading Power (DRP). In a reasearch study on third graders, the DRP was administered to 44 students. Their scores are shown below:

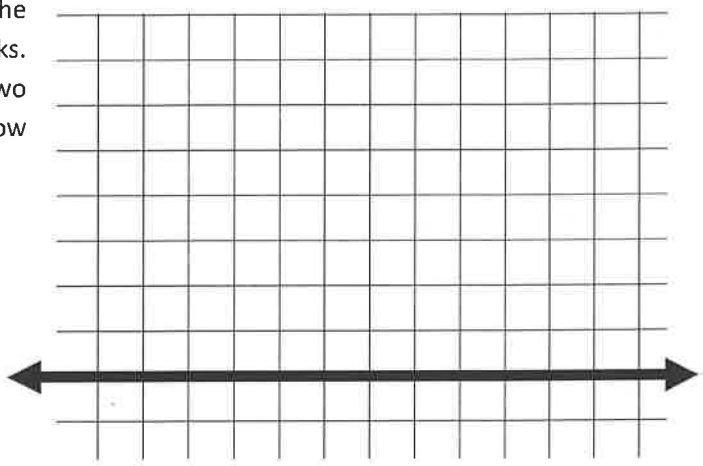
40	26	39	14	42	18	25	43	46	27	19
47	19	26	35	34	15	44	40	38	31	46
52	25	35	35	33	29	34	41	49	28	52
47	35	48	22	33	41	51	27	14	54	45

Display these data graphically. Write a paragraph describing the distribution of DRP scores.

26) Corn is an important animal food. Normal corn lacks certain amino acids, which are building block for protein. Plant scientists have developed new corn varieties that have more of these amino acids. To test a new corn as an animal food, a group of 20 one-day-old male chicks was fed a ration containing the new corn. A control group of another 20 one-day-old male chicks was fed a ration that was identical except that it contained normal corn. Here are the weight gains (in grams) after 21 days:

Normal corn				New corn			
380	321	366	356	361	447	401	375
283	349	402	462	434	403	393	426
356	410	329	399	406	318	467	407
350	384	316	272	427	420	477	392
345	455	360	431	430	339	410	326

a) Compute five-number summaries for the weight gains of the two groups of chicks. Then make boxplots to compare the two distributions. What do the data show about the effect of the new corn?



b) What are the means and standard deviations for the two groups of chicks? How much larger is the mean weight gain of chicks fed the new corn? (Answer this with percentages)

c) The weights are given in grams. There are 28.35 grams in an ounce. Use the results of part b) to compute the means and standard deviations of the weight gains measures in ounces.

27) Political party preference in the United States depends in part on the age, income, and gender of the voter. A political scientist selects a large sample of registered voters. For each voter, she records gender, age, household income, and whether they voted for the Democratic or for the Republican candidate in the last congressional election. Which of these variables are categorical and which are quantitative?

28) A study of the size of jury awards in civil cases (such as injury, product liability, and medical malpractice) in Chicago showed that the median award was about $8000. But the mean award was about $69,000. Explain how this great difference between the two measures of center can occur.

29) Here are the scores of games played in the California Division I-AAA high school basketball playoffs:

71-38	52-47	55-53	76-65	77-63	65-63	68-54	64-62
87-47	64-56	78-64	58-51	91-74	71-41	67-62	106-46

On the same day, the final scores of games in Division V-AA were:

98-45	67-44	74-60	96-54	92-72	93-46
98-67	62-37	37-36	69-44	86-66	66-58

a) Construct a back-to-back stemplot to compare the number of points scored by Division I-AAA and Division V-AA basketball teams.

b) Compare the shape, center, and spread of the two distributions. Which numerical summaries are most appropriate in this case? Why?

c) Is there a difference in "margin of victory" in Division I-AAA and Division V-AA games? Provide appropriate graphical and numerical support for your answer.

30) Which measure of center, the mean or the median, should you use in each of the following situations?

a) Middletown is considering imposing an income tax on citizens. The city government wants to know the average income of citizens so that it can estimate the total tax base.

_____

b) In a study of the standard of living of typical families in Middletown, a sociologist estimates the average family income in that city.

_____

31) A school system employs teachers at salaries between $30,000 and $60,000. The teachers' union and the school board are negotiating the form of next year's increase in the salary schedule. Suppose that every teacher is given a flat $1000 raise.

a) How much will the mean salary increase? The median salary?

b) Will a flat $1000 raise increase the spread as measured by the distance between the quartiles? Explain.

c) Will a flat $1000 raise increase the spread as measured by the standard deviation of the salaries? Explain.

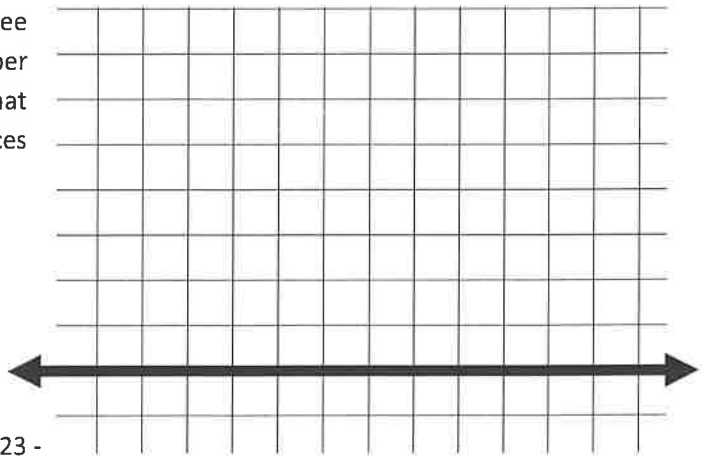
32) Suppose that the teachers in the previous exercise each receive a 5% raise. The amount of the raise will vary from $1500 to $3000, depending on present salary. Will a 5% across-the-board raise increase the spread of the distribution as measured by the distance between the quartiles? Do you think it will increase the standard deviation?

33) You are preparing to study the television viewing habits of high school students. Describe two categorical variables and to quantitative variables that you might record for each student. Give the units of measurement for the quantitative variables.

34) Different varieties of the tropical flower Heliconia are fertilized by different species of hummingbirds. Over time, the lengths of the flowers and the forms of the hummingbirds' beaks have evolved to match each other. Here are the data on the lengths in millimeters of three varieties of these flowers on the island of Dominica:

H. bihai:	47.12	46.75	46.80	47.67	47.43	46.44	46.64	48.07	48.34	48.15
	50.26	50.12	46.34	46.94	48.36					
H. caribaea red:	41.90	42.01	41.93	43.09	41.47	41.69	39.78	40.57	39.63	
	42.18	40.66	37.87	39.16	37.40	38.20	38.07	38.10	37.97	
	38.79	38.23	38.87	37.78	38.01					
H. caribaea yellow:	36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10	35.17	
	36.82	36.66	35.68	36.03	34.57	34.63				

a) Make boxplots to compare the three distributions. Report the five-number summaries along with your graphs. What are the most important differences among the three varieties of flower?



b) Find the mean and standard deviation for each variety.

c) Make a stemplot for each of flower lengths. Do the distributions appear suitable for use of mean and standard deviation summaries?

35) The following are Barry Bonds' home run counts from 1986 to 2004:

16	25	24	19	33	25	34	46	37	33
42	40	37	34	49	73	46	45	45	

a) Find the quartiles Q1 and Q3 for Barry Bonds' home run data.

b) Use the 1.5*IQR rule for identifying outliers to see in Bonds' 73 home run year is an outlier. Explain your findings.