AP STATISTICS
1I HW

NAME:

INTRODUCTION Exercises

1. Protecting wood How can we help wood surfaces resist weathering, especially when restoring historic wooden buildings? In a study of this question, researchers prepared wooden panels and then exposed them to the weather. Here are some of the variables recorded: type of wood (yellow poplar, pine, cedar); type of water repellent (solvent-based, water-based); paint thickness (millimeters); paint color (white, gray, light blue); weathering time (months). Identify each variable as categorical or quantitative.

Categorical: type of wood, type water repellent, paint color

Quantitative: paint thickness, weathering time

Multiple choice: *Select the best answer.*

Exercises 7 and 8 refer to the following setting. At the Census Bureau Web site, you can view detailed data collected by the American Community Survey. The table below includes data for 10 people chosen at random from the more than one million people in households contacted by the survey. "School" gives the highest level of education completed.

Weight (lb)	Age (yr)	Travel to work (min)	School	Gender	Income last year ($)
187	66	0	Ninth grade	1	24,000
158	66	n/a	High school grad	2	0
176	54	10	Assoc. degree	2	11,900
339	37	10	Assoc. degree	1	6,000
91	27	10	Some college	2	30,000
155	18	n/a	High school grad	2	0
213	38	15	Master's degree	2	125,000
194	40	0	High school grad	1	800
221	18	20	High school grad	1	2,500
193	11	n/a	Fifth grade	1	0

Notice Categorical may be labeled as numbers
Gendered labeled 1(m) and 0(F) could be Quantitative to GET statistic % male.

3. A class survey Here is a small part of the data set that describes the students in an AP Statistics class. The data come from anonymous responses to a questionnaire filled out on the first day of class. pg 3

Gender	Hand	Height (in)	Homework time (min)	Favorite music	Pocket change (cents)
F	L	65	200	Hip-hop	50
M	L	72	30	Country	35
M	R	62	95	Rock	35
F	L	64	120	Alternative	0
M	R	63	220	Hip-hop	0
F	R	58	60	Alternative	76
F	R	67	150	Rock	215

(a) What individuals does this data set describe?

(b) Clearly identify each of the variables. Which are quantitative? In what units are they measured?

(a) Individuals: AP Statistics students that completed survey

(b) Variables:

Categorical
1. gender
2. left/right handed
3. favorite music

QUANTITATIVE
1. height (in)
2. time spent on HW (min)
3. pocket change (¢'s)

7. The individuals in this data set are
(a) households.
(b) people.
(c) adults.
(d) 120 variables.
(e) columns.

8. This data set contains
(a) 7 variables, 2 of which are categorical.
(b) 7 variables, 1 of which is categorical.
(c) 6 variables, 2 of which are categorical.
(d) 6 variables, 1 of which is categorical.
(e) None of these.

SECTION 1.1 Exercises

11 Birth days Births are not evenly distributed across the days of the week. Here are the average numbers of babies born on each day of the week in the United States in a recent year:[10]

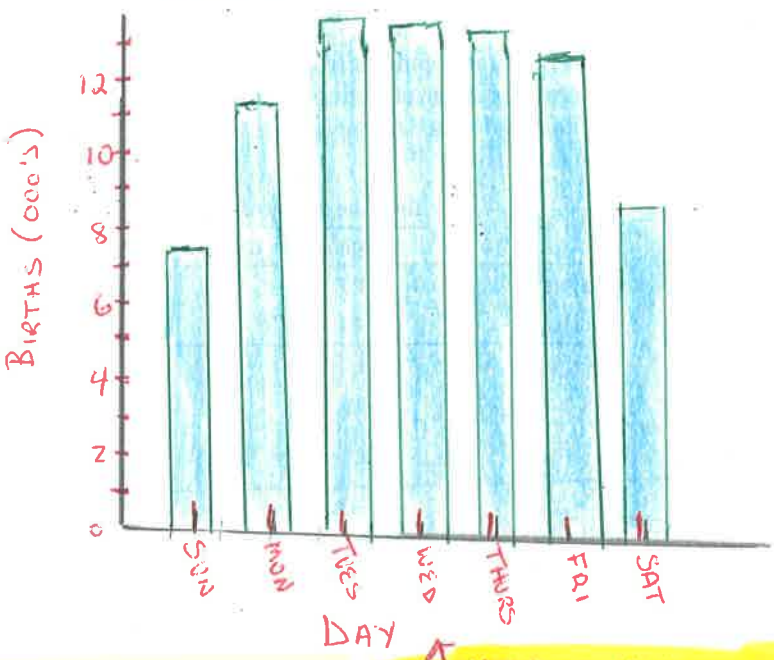
Day	Births
Sunday	7,374
Monday	11,704
Tuesday	13,169
Wednesday	13,038
Thursday	13,013
Friday	12,664
Saturday	8,459

MIN - 7,374
MAX - 13,038

(a) Present these data in a well-labeled bar graph. Would it also be correct to make a pie chart?

YES. BECAUSE ALL DAYS OF WEEK ARE INCLUDED. WOULD NEED TO CALCULATE %'s

BIRTHS BY DAY OF WEEK



INTERPRET GRAPH
WHY FEWER BIRTHS ON WEEKENDS?

NOTICE: Categorical Data. Labels in center of bar. Bars are NOT connected

15 Buying music online Young people are more likely than older folk to buy music online. Here are the percents of people in several age groups who bought music online in 2006:[14]

pg 10

Age group	Bought music online
12 to 17 years	24%
18 to 24 years	21%
25 to 34 years	20%
35 to 44 years	16%
45 to 54 years	10%
55 to 64 years	3%
65 years and over	1%

You can NOT use a pie chart. This % represents the % of 12-17 years that bought online.

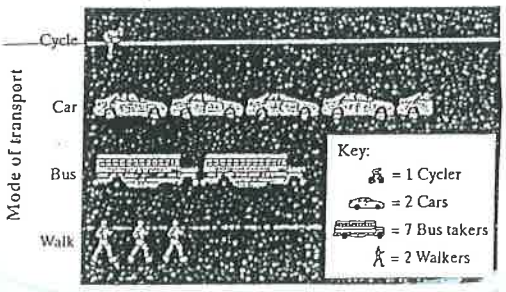
PIE CHARTS MUST BE % OF A Whole. IN THIS EXAMPLE, THE %'s would have to be buying online by each age group and total to 100%.

(a) Explain why it is *not* correct to use a pie chart to display these data.

17 Going to school Students in a high school statistics class were given data about the primary method of transportation to school for a group of 30 students. They produced the pictograph shown.

pg 11

(a) How is this graph misleading?



THE PICTURE SHOULD BE PROPORTIONAL TO THE NUMBER OF STUDENTS THEY REPRESENT

19 Attitudes toward recycled products Recycling is supposed to save resources. Some people think recycled products are lower in quality than other products, a fact that makes recycling less practical. People who actually use a recycled product may have different opinions from those who don't use it. Here are data on attitudes toward coffee filters made of recycled paper among people who do and don't buy these filters:[16]

pg 13

(a) How many people does this table describe? How many of these were buyers of coffee filters made of recycled paper? 133 people; 36 bought recycle filters.

(b) Give the marginal distribution of opinion about the quality of recycled filters. What percent think the quality of the recycled product is the same or higher than the quality of other filters?

	Think the quality of the recycled product is:			TOTAL
	Higher	The same	Lower	
Buyers	20	7	9	36
Nonbuyers	29	25	43	97
TOTAL	49	32	52	133

60.90% responded that the quality of recycled coffee filters was "the same" or "higher".

(b) Marginal distribution of QUALITY OF FILTERS → 49/133 32/133 52/133 → 36.84% 24.06% 39.10% = 100%

21 Attitudes toward recycled products Exercise 19 gives data on the opinions of people who have and have not bought coffee filters made from recycled paper. To see the relationship between opinion and experience with the product, find the conditional distributions of opinion (the response variable) for buyers and nonbuyers. What do you conclude?

pg 15

Below are 2 sets of conditional distributions one for buyers and the other for non buyers. Now you can compare opinions of QUALITY between buyers and non buyers.

* CREATE A TABLE TO DISPLAY THIS INFORMATION

	HIGHER	SAME	LOWER	TOTAL
BUYERS	55.56%	19.44%	25.00%	100%
NON BUYERS	29.90%	25.77%	44.33%	100%

Buyers are much more likely to consider recycled filters as higher QUALITY, THOUGH 25% of BUYERS STILL THINK THEY ARE LOWER QUALITY.

Multiple choice: *Select the best answer.*

Exercises 27 to 32 refer to the following setting. The National Survey of Adolescent Health interviewed several thousand teens (grades 7 to 12). One question asked was "What do you think are the chances you will be married in the next ten years?" Here is a two-way table of the responses by gender:[18]

	Female	Male	Total
Almost no chance	119	103	222
Some chance, but probably not	150	171	321
A 50-50 chance	447	512	959
A good chance	735	710	1445
Almost certain	1174	756	1930
TOTAL	2625	2252	4877

27. The percent of females among the respondents was
(a) 2625. (c) about 46%. (e) None of these.
(b) 4877. (d) about 54%. 2625/4877 = 53.82%

28. Your percent from the previous exercise is part of
(a) the marginal distribution of females.
(b) the marginal distribution of gender.
(c) the marginal distribution of opinion about marriage.
(d) the conditional distribution of gender among adolescents with a given opinion.
(e) the conditional distribution of opinion among adolescents of a given gender.

29. What percent of females thought that they were almost certain to be married in the next ten years? denominator is females "given female"
(a) About 16% (c) About 40% (e) About 61%
(b) About 24% (d) About 45% 1174/2625 = 44.72%

30. Your percent from the previous exercise is part of
(a) the marginal distribution of gender.
(b) the marginal distribution of opinion about marriage.
(c) the conditional distribution of gender among adolescents with a given opinion.
(d) the conditional distribution of opinion among adolescents of a given gender.
(e) the conditional distribution of "Almost certain" among females.

denominator is opinion
Given opinion

31. What percent of those who thought they were almost certain to be married were female?

(a) About 16% (c) About 40% (e) About 61%

(b) About 24% (d) About 45%

1174/1930 = 60.83%

32. Your percent from the previous exercise is part of

(a) the marginal distribution of gender.

(b) the marginal distribution of opinion about marriage.

(c) the conditional distribution of gender among adolescents with a given opinion.

(d) the conditional distribution of opinion among adolescents of a given gender.

(e) the conditional distribution of females among those who said "Almost certain."

---

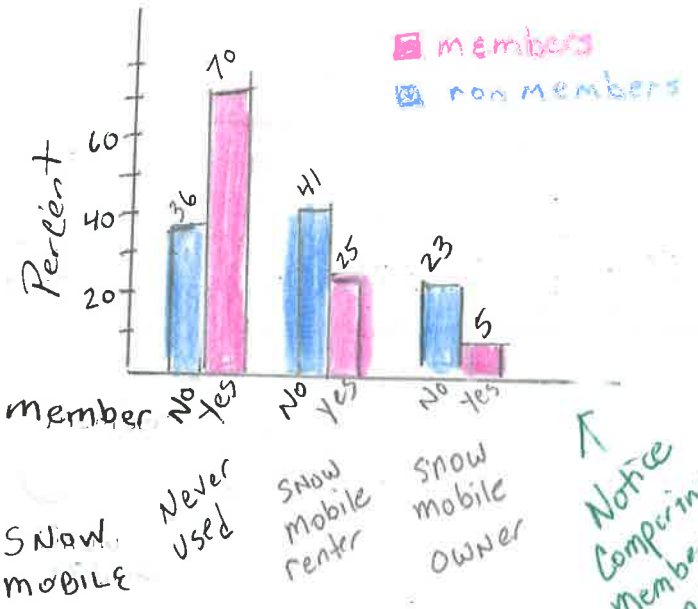
25. Snowmobiles in the park Yellowstone National Park surveyed a random sample of 1526 winter visitors to the park. They asked each person whether they owned, rented, or had never used a snowmobile. Respondents were also asked whether they belonged to an environmental organization (like the Sierra Club). The two-way table summarizes the survey responses.

pg 18

STEP 4

	Environmental Clubs		
	No	Yes	Total
Never used	445 (36%)	212 (70%)	657
Snowmobile renter	497 (41%)	77 (25%)	574
Snowmobile owner	279 (23%)	16 (5%)	295
Total	1221 (100%)	305 (100%)	1526

Do these data provide convincing evidence of an association between environmental club membership and snowmobile use for the population of visitors to Yellowstone National Park? Follow the four-step process.



Notice comparing members and non members and space between snowmobile groups

STATE: From the Population of visitors to Yellowstone National Park, What is the relationship between membership in an environmental club and use of snowmobiles..

PLAN: To see if their is a relationship, we will look at Conditional distributions for members and then non members

DO:
- Calculate Conditional distributions
- Create side-by-side bar graphs (see graph)

CONCLUDE:

Members of environmental clubs are much more likely to have never owned a snow mobile (70%) compared to Non members (36%). Members are less likely to rent or own snow mobiles compares to non members of environmental clubs.

Acronym "Cuss and BS"

Center
Unusual
Shape
Spread

Be Specific

SECTION 1.2 Exercises

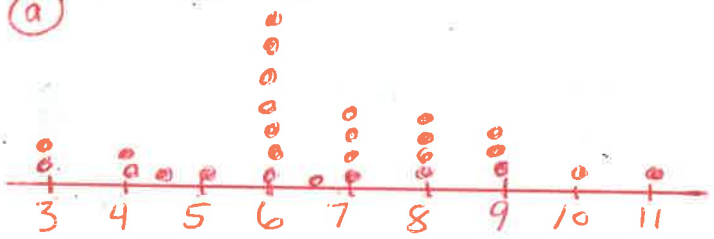
37. Feeling sleepy? Students in a college statistics class responded to a survey designed by their teacher. One of the survey questions was "How much sleep did you get last night?" Here are the data (in hours):

9	6	8	6	8	8	6	6.5	6	7	9	4	3	4
5	6	11	6	3	6	6	10	7	8	4.5	9	7	7

MINIMUM - 3
MAXIMUM - 11

(a) Make a dotplot to display the data.

(a)



(b) Describe the overall pattern of the distribution and any deviations from that pattern.

Remember Shape, Center, Spread
Outliers in context.

The shape is roughly symmetric with a center around 6 hrs. The spread has a range of 8 hours with a minimum of 3 hrs and a maximum of 11 hrs. There do not appear to be any outliers

Notice the highlighted hedging words. because Statistics are NOT an exact science.

45. Where do the young live? Below is a stemplot of the percent of residents aged 25 to 34 in each of the 50 states. As in the stemplot for older residents (page 35), the stems are whole percents, and the leaves are tenths of a percent. This time, each stem has been split in two, with values having leaves 0 through 4 placed on one stem, and values ending in 5 through 9 placed on another stem.

STEM	LEAF
11	44
11	66778
12	0134
12	666778888
13	0000001111444
13	7788999
14	0044
14	567
15	11
15	
16	0

Key: 12|1 means that 12.1% of that state's residents are aged 25 to 34.

KEY 11|4 = 11.4%

NOTICE:
- STEM PLOTS MUST have a key.
  * 2 TYPES OF KEYS:
    1. Key: STEM - ones digit, LEAF - tenths digit
    2. KEY: 11|4 = 11.4%
- LEAF ARE IN NUMERICAL ORDER
- widths of numbers are EQUAL

(a) Why did we split stems?

(a) we split to get a better understanding of the shape. IF we did not split, we would only have 6 bars.

(b) Utah has the highest percent of residents aged 25 to 34. What is that percent? Why do you think Utah has an unusually high percent of residents in this age group?

(b) Utah has the highest % at 16.0% 25-34, possibly due to Mormon church

(c) Describe the shape, center, and spread of the distribution, ignoring Utah.

(c) The shape is roughly symmetric with the center around 13% and the spread of about 3.5% (MIN 11.4% and max 15.1)

49 Do women study more than men? We asked the students in a large first-year college class how many minutes they studied on a typical weeknight. Here are the responses of random samples of 30 women and 30 men from the class:

min

Women				
180	120	180	360	240
120	180	120	240	170
150	120	180	180	150
200	150	180	150	180
120	60	120	180	180
90	240	180	115	120

n = 30

MIN - 60 MAX - 360

Men				
90	120	30	90	200
90	45	30	120	75
150	120	60	240	300
240	60	120	60	30
30	230	120	95	150
0	200	120	120	180

n = 30

MIN - 0 MAX - 300

(a) Examine the data. Why are you not surprised that most responses are multiples of 10 minutes? Are there any responses you consider suspicious?

(b) Make a back-to-back stemplot to compare the two samples. Does it appear that women study more than men (or at least claim that they do)? Justify your answer.

(a) Most people estimate # min studying in 10 min intervals. Notice, responses are in multiples of 30 and 60 EQUIVALENT TO 1/2 hr and hrs. The maximum values of 360 min (6 hrs) and 300 min (5 hrs) seem to be exagerating

minutes studying on typical weeknight

Women		Men
	0	033334
96	0	66679999
22222221	1	22222222
8888887555	1	558
4440	2	00344
	2	
	3	0
6	3	

MUST HAVE A KEY

Key:

STEM = HUNDREDS
LEAF = TENS

OR

KEY 2|0 = 200

NOTICE: Responses ending with a 5 were truncated

EXAMPLE: 115

1|1 Could round

Conclusion:

It does appear women study more since the center for women was around 180 min, while the center for men was around 120 minutes.

SHOPPING SPREE (ROUNDED TO $TENS)

(48) (A)

0	3 99
1	1 3 4 5 6 7 7 8 8 9
2	0 0 0 1 2 3 4 5 5 6 6 8 8 8 8
3	2 5 6 9 9
4	1 3 4 5 5 7 9
5	0 3 5 9
6	1
7	0
8	3 6 6
9	3

YOU MOST ALWAYS GIVE A KEY!

KEY 0|3 = $3

(B)

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

0-4

5-9

SPLIT STEM

KEY: 0|3 = $3

NOTE, THE STEM PLOT WITH THE SPLIT STEMS SHOW THE SKEWNESS, GAPS, AND OUTLIERS FOR CLEARLY.

(C) THE DISTRIBUTION OF THE AMOUNT OF MONEY SPENT BY SHOPPERS AT THIS GROCERY STORE IS CENTERED AROUND $28 (BASED ON THE MEDIAN) WITH A SPREAD OF ABOUT $90 (RANGE). THE SHAPE OF THE DISTRIBUTION IS SKEWED RIGHT. WITH GAPS BETWEEN $62 to $69 and $71 to $82 AND SOME OUTLIERS ON THE HIGH END $86 and $93.

Be specific → BY SHOPPERS AT THIS

Center → $28 (BASED ON THE MEDIAN)

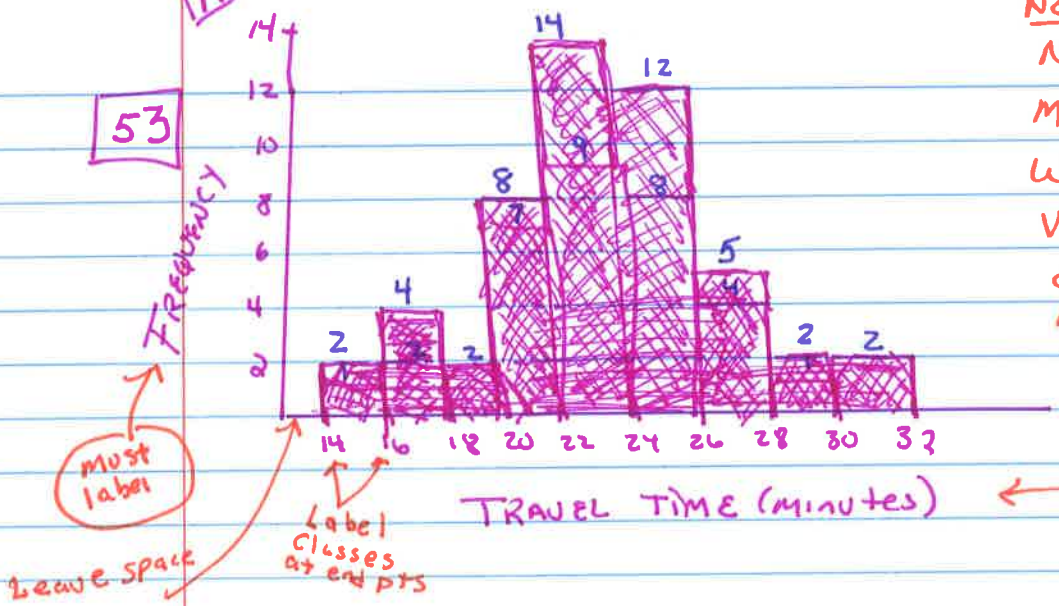
Spread → A SPREAD OF ABOUT $90 (RANGE).

Shape → THE SHAPE OF THE DISTRIBUTION IS SKEWED RIGHT.

Unusual → WITH GAPS BETWEEN $62 to $69 and $71 to $82 AND SOME OUTLIERS ON THE HIGH END $86 and $93.

A

53



NOTE: GRAPHS DO NOT NEED TO BE ARTIST MASTER PIECES. IN STATS, WE GRAPH TO GET A VISUAL UNDERSTANDING OF THE DATA. LABELS ARE IMPORTANT.

Make sure to get to know your calculator. You can use it on all tests. It's a GREAT tool.

STEPS:

① Put the data in a LIST (STAT) EDIT L1

② (ZOOM) (STAT:9) to see the graph

③ set the (WINDOW) to easily graph

the problem ask you to setup like this:

Xmin = 14
Xmax = 32
Xscl = 2

Ymin = 0
Ymax = 14
Yscl = 2

these may vary

③ USE (TRACE) TO GET FREQUENCY COUNTS FOR EACH BAR (You do not need to include these on your graph, but they can be helpful interpreting your data.

④ GET STATISTICS (STAT) (CALC) 1: 1-Var Stats (L1)

$\bar{x} = 23.1$ $S_x = 3.57$ $n = 51$ Min = 15.5 Med = 23.4 Max = 30.9

B

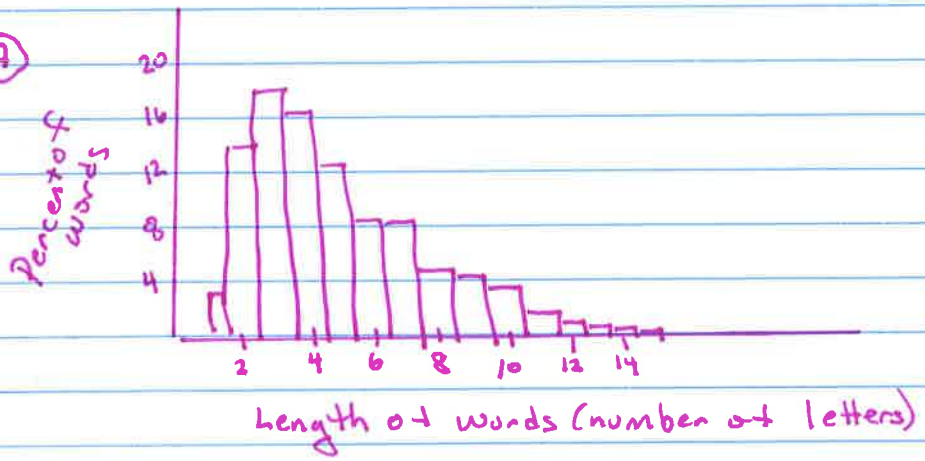
ALTHOUGH THE GRAPH IS SOMEWHAT IRREGULAR, THE DISTRIBUTION IS ROUGHLY SYMMETRIC

FROM THE GRAPH THE CENTER IS ABOUT 23 MIN

THE SPREAD MEASURED BY THE RANGE IS ABOUT 15 MIN

THERE DO NOT APPEAR TO BE ANY OUTLIERS.

57 (A)



The distribution of word length is unimodal and skewed to the right. The center is about 4 letters per word. with a spread of 14 letters (range) and there are no gaps or outliers

(B) The distribution of words in Popular Mechanics and Shakespeare have similar shapes, skewed right and unimodal. The spread for Shakespeare is less (about 13 letters). Both have centers about 4 letters per word.

Shakespeare has more short words (with 2, 3, 4 letter words) while popular mechanics has more long words

Multiple choice *Select the best answer for Exercises 69 to 74.*

69. Here are the amounts of money (cents) in coins carried by 10 students in a statistics class: 50, 35, 0, 97, 76, 0, 0, 87, 23, 65. To make a stemplot of these data, you would use stems

(a) 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
(b) 0, 2, 3, 5, 6, 7, 8, 9.
(c) 0, 3, 5, 6, 7.
(d) 00, 10, 20, 30, 40, 50, 60, 70, 80, 90.
(e) None of these.

DO NOT SKIP CATEGORIES

70. One of the following 12 scores was omitted from the stemplot below:

84 76 92 92 88 96 68 80 92 88 76 96

6	8
7	66
8	0488
9	2266

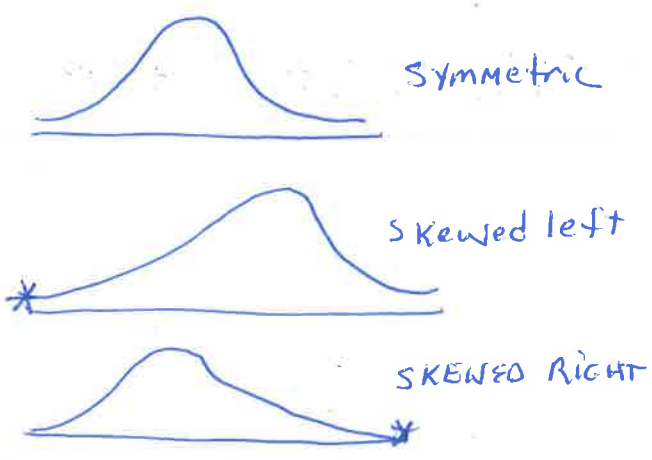
The missing number is

(a) 76. (b) 88. (c) 90. (d) 92. (e) 96.

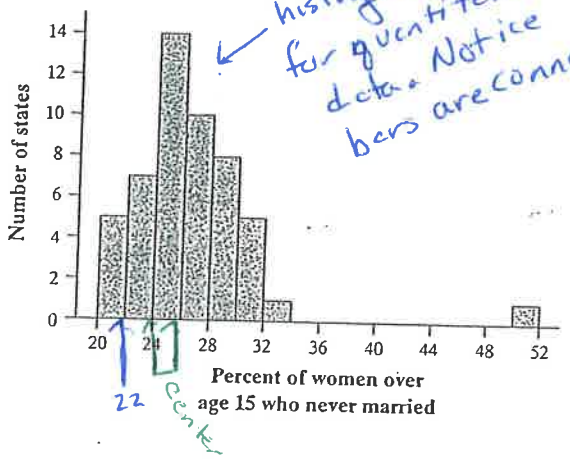
71. You look at real estate ads for houses in Naples, Florida. There are many houses ranging from $200,000 to $500,000 in price. The few houses on the water, however, have prices up to $15 million. The distribution of house prices will be

(a) skewed to the left.
(b) roughly symmetric.
(c) skewed to the right.
(d) unimodal.
(e) too high.

$200 $500 $15 million



Exercises 72 to 74 refer to the following setting. The histogram below shows the distribution of the percents of women aged 15 and over who have never married in each of the 50 states and the District of Columbia.



histogram for quantitative data. Notice bars are connected

72. The leftmost bar in the histogram covers percents of never-married women ranging from about

(a) 20% to 24%.
(b) 20% to 22%.
(c) 0% to 20%.
(d) 0% to 5%.
(e) None of these.

73. The center of this distribution is in the interval

(a) 22% to 24%.
(b) 24% to 26%.
(c) 26% to 28%.
(d) 28% to 30%.
(e) 36% to 38%.

74. In about what percent of states have at least 30% of women aged 15 and over never married?

(a) 4% (b) 7% (c) 10% (d) 14% (e) 32%

The histogram is # of states

30-32% — 5
32-34% — 1
50-52% — 1

7/50 = 14%

SECTION 1.3 Exercises

79. Quiz grades Joey's first 14 quiz grades in a marking period were

pg 51

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

Calculate the mean. Show your work. Interpret your result in context.

$$\text{Mean} = \bar{x} = \frac{\Sigma x_i}{n} = \frac{1190}{14} = 85$$

The mean of Joey's first 14 quiz's is 85%. IN CONTEXT: IF JOEY HAD SCORED THE SAME NUMBER OF POINTS OF THE 1ST 14 QUIZES, THEN HE WOULD HAVE SCORED AN 85% ON EACH QUIZ. (The mean is the balancing point. The "fair share")

89. Quiz grades Refer to Exercise 79.

pg 57, 58

(a) Find and interpret the interquartile range (*IQR*).

(b) Determine whether there are any outliers. Show your work.

MIN-74
Q1-78
Q2-85
Q3-91
MAX-98

n=14
$\bar{x}$=85
Σx=1190

(c) TI84 - ENTER DATA IN L1 AND CHECK BOX PLOT. IS YOUR HAND PLOT DIFFERENT FROM THE CALC? EXPLAIN

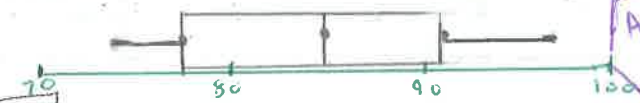
(a) IQR = Q3(91) - Q1(78) = 13

The middle 50% of the data has a 13 point spread

(b) $Q_1 - 1.5\,IQR = 78 - 1.5(13) = 58.5$
$Q_3 + 1.5\,IQR = 91 + 1.5(13) = 110.5$
There are NO OUTLIERS

ALWAYS USE 1ST BOX PLOT BC IT SHOWS OUTLIERS *

(c)



83. Incomes of college grads According to the Census Bureau, the mean and median 2008 income of people at least 25 years old who had a bachelor's degree but no higher degree were $48,097 and $60,954. Which of these numbers is the mean and which is the median? Explain your reasoning.

MEAN - $60,954
MEDIAN - $48,097
THE DISTRIBUTION IS LIKELY SKEWED TO THE RIGHT BECAUSE A FEW PEOPLE WHO HAVE VERY LARGE INCOMES ARE PULLING THE MEAN TOWARDS THE HIGHER INCOMES.

81. Quiz grades Refer to Exercise 79.

pg 53

(a) Find the median by hand. Show your work. Interpret your result in context.

(b) Suppose Joey has an unexcused absence for the 15th quiz, and he receives a score of zero. Recalculate the mean and the median. What property of measures of center does this illustrate?

(a) EASY WAY TO FIND MEDIAN IS TO DO A QUICK STEM LEAF GRAPH

7	4568
8	0246 7
9	0189 3

Median = $\frac{84+86}{2} = 85$

The median is 85%. That means half the scores are below 85 and half are above 85

(b) 15TH QUIZ = 0%

$\bar{x} = \frac{1190}{15} = 79.33\%$

Median = 84%

Notice • the mean went from 85% to 79%.
• the median went from 85% to 84%.

RESISTENCE is demonstrated.

AN OUTLIER HAS A LARGE IMPACT ON THE MEAN, WHILE THE MEDIAN IS LESS INFLUENCED BY AN OUTLIER (the median is a resistent measure)

91. **Don't call me** In a September 28, 2008, article titled "Letting Our Fingers Do the Talking," the *New York Times* reported that Americans now send more text messages than they make phone calls. According to a study by Nielsen Mobile, "Teenagers ages 13 to 17 are by far the most prolific texters, sending or receiving 1,742 messages a month." Mr. Williams, a high school statistics teacher, was skeptical about the claims in the article. So he collected data from his first-period statistics class on the number of text messages and calls they had sent or received in the past 24 hours. Here are the texting data:

0	7	1	29	25	8	5	1	25	98	9	0	26
8	118	72	0	92	52	14	3	3	44	5	42	

(a) Make a boxplot of these data by hand. Be sure to check for outliers.

(b) Do these data support the claim in the article about the number of texts sent by teens? Justify your answer with appropriate evidence.

① ENTER DATA INTO L1

② STAT 1-VAR STATS

$\bar{x} = 27.48$
$n = 25$

MIN = 0
Q1 = 3
MED = 9
Q3 = 43
MAX = 118

③ DRAW BOX PLOT. CHECK W/CALC:

2ND STAT PLOT PLOT 1> ON
Type: L1 Freq 1
ZOOM 9: ZOOM STAT

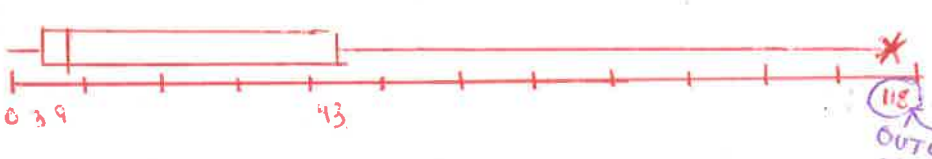
④ OUTLIERS

IQR = Q3 − Q1 = 43 − 3 = 40

Q3 + 1.5 IQR = 43 + 1.5(40) = 103

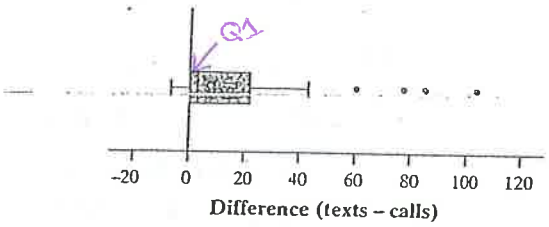
*SINCE 118 IS GREATER THAN 103, IT IS AN OUTLIER, NOTED IN BOX PLOT WITH AN *

(A)



(B) The article claims 1742 monthly messages which works out to be about 58 msg/day. That seems very high since half this class sent fewer than 10 msgs AND ONLY 4 OUT OF THE 25 students sent more than 58 messages

93. **Texts or calls?** Refer to Exercise 91. A boxplot of the difference (texts – calls) in the number of texts and calls for each student is shown below.



"0" means the same number of calls and texts were made
"+number" – more texts sent than calls
"−number" – more calls than texts sent

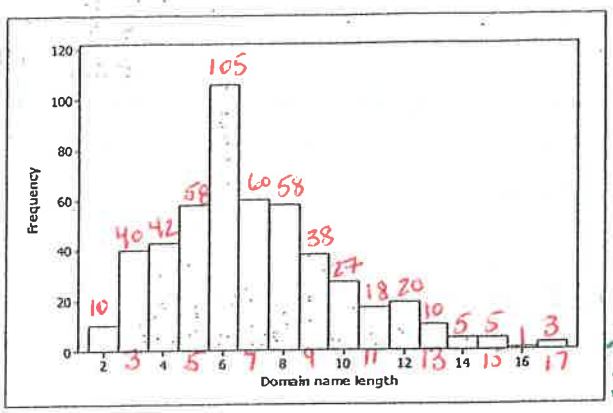
(a) Do these data support the claim in the article about texting versus calling? Justify your answer with appropriate evidence.

(b) Can we draw any conclusion about the preferences of all students in the school based on the data from Mr. Williams's statistics class? Why or why not?

(a) This box plot does support the articles claim that teenagers prefer texting over phone calls. This is based on 75% of the students shown in the box plot made more texts than calls.

(b) We can not draw conclusions since this was not a random sample and there may be similarities in this group that could not be generalized to all teenagers. (ie AP students are higher educated)

87. **Domain names** When it comes to Internet domain names, is shorter better? According to one ranking of Web sites in 2008, the top 8 sites (by number of "hits") were yahoo.com, google.com, youtube.com, live.com, msn.com, myspace.com, wikipedia.org, and facebook.com. These familiar sites certainly have short domain names. The histogram below shows the domain name lengths (in number of letters in the name, not including the extensions .com and .org) for the 500 most popular Web sites.



(A)

Length	# SITES	Total length L*S	Cumulative # SITES
2	10	20	10
3	40	120	50
4	42	168	92
5	58	290	150
6	105	630	255
7	60	420	
8	58	464	
9	38	342	
10	27	270	
11	18	198	
12	20	240	
13	10	130	
14	5	70	
15	5	75	
16	1	16	
17	3	51	
Total	500	3504	

Median is the average of the 250th & 251st VALUES

Median is 6

Tip: Use Calc

L1 L2 L3

> STAT > Calc > 1-VAR L2 → $\Sigma x = 500$

> Stat > Calc > 1-Var L3 → $\Sigma x = 3504$

MEAN

$\bar{x} = \frac{3504}{500}$

$\bar{x} = 7.008$

(a) Estimate the mean and median of the distribution. Explain your method clearly.

(b) If you wanted to argue that shorter domain names were more popular, which measure of center would you choose—the mean or the median? Justify your answer. Since the median (6) is less than the mean (7), we would use the median to argue that shorter names are more popular. The mean is skewed to the right with a few large names

97. **Phosphate levels** The level of various substances in the blood influences our health. Here are measurements of the level of phosphate in the blood of a patient, in milligrams of phosphate per deciliter of blood, made on 6 consecutive visits to a clinic: 5.6, 5.2, 4.6, 4.9, 5.7, 6.4. A graph of only 6 observations gives little information, so we proceed to compute the mean and standard deviation.

(a) Find the standard deviation from its definition. That is, find the deviations of each observation from the mean, square the deviations, then obtain the variance and the standard deviation.

(b) Interpret the value of s_x you obtained in (a).

(A)

Phosphate Level (mg/dl) x_i	Deviations $x_i - \bar{x}$	$(x_i - \bar{x})^2$
5.6	.2	.04
5.2	−.2	.04
4.6	−.8	.64
4.9	−.5	.25
5.7	.3	.09
6.4	1	1
	$\Sigma = 0$	$\Sigma = 2.06$

$\bar{x} = 5.4$

$s_x = .6419$

L2 = L1 − 5.4

L3 = L2²

$s^2 = \frac{\Sigma (x_i - \bar{x})^2}{n-1}$

$s^2 = \frac{2.06}{6-1} = .412$

$s = .6419$ mg/dl

(B) The typical phosphate level is on average .6419 mg/dl different from the mean level of 5.4 mg/dl

103. **SD contest** 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 in either (a) or (b)? Explain.

(a) 4 same number (1, 1, 1, 1)

(b) 0, 0, 10, 10 (2 of lowest and 2 of high)

(c) For (a) any group of the same 4 numbers results in $S_x = 0$

For (b) we want the largest deviation (0, 10). The mean = 5 so all deviations would be $(5)^2$.

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

pg 66

Women

154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	129	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	

Do these data support the belief that women have better study habits and attitudes toward learning than men? (Note that high scores indicate good study habits and attitudes toward learning.) Follow the four-step process.

STATE DOES THE DATA INDICATE THAT WOMEN BETTER STUDY HABITS AND ATTITUDES TOWARDS LEARNING THAN MEN?

PLAN CREATE PARALLEL BOX PLOTS FOR MEN AND WOMEN

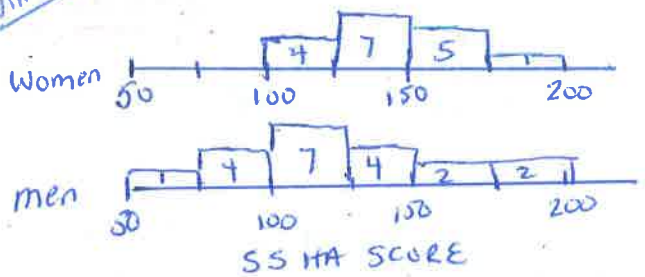
DO ENTER DATA L1 = Women L2 = men

NUMERICAL SUMMARIES

VARIABLE	N	$\bar{x}$	S_x	MIN	Q1	MED	Q3	MAX
WOMEN	18	141	26.4	101	126	138.5	154	200
MEN	20	121	32.9	70	98	114.5	143	187

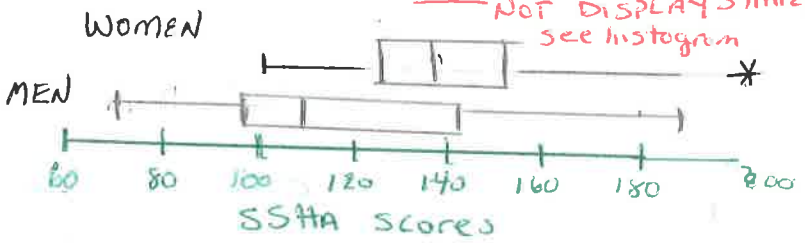
STATPLOT WINDOW – XMIN=50 XMAX=200 XSCL=25 Use trace to get frequencies

HISTOGRAMS



① BOX PLOTS MUST USE the same scale
② GRAPHS MUST show scales and be labeled

NOTE: Box plots do NOT display shape, see histogram



Conclude: It appears that Women have higher SS HA Scores than men. The median is higher for women than men (138.5 vs 114.5). The women are higher on all 5 numeric summaries. Women have a smaller standard deviation so there is less variability in their scores. Both men + women's scores appear to be symmetric with men's scores more spread out.

Conclusion Tips:
① When Comparing 2 Groups remember "CUSS and BS": You should describe center, spread, shape and unusual values and be specific. ② Compare at least 3 in the CUSS.

Multiple choice: *Select the best answer for Exercises 107 to 110.*

107 If a distribution is skewed to the right with no outliers,
(a) mean < median.
(b) mean ≈ median.
(c) mean = median.
(d) mean > median.
(e) We can't tell without examining the data.

Tip: The mean gets pulled to the outliers

108 You have data on the weights in grams of 5 baby pythons. The mean weight is 31.8 and the standard deviation of the weights is 2.39. The correct units for the standard deviation are
(a) no units—it's just a number.
(b) grams.
(c) grams squared.
(d) pythons.
(e) pythons squared.

$\bar{x} = 31.8g$
$S = 2.39g$

109 Which of the following is least affected if an extreme high outlier is added to your data?
(a) Median
(b) Mean
(c) Standard deviation
(d) Range
(e) Maximum

110 What are all the values that a standard deviation s_x can possibly take?
(a) $s_x \geq 0$
(b) $s_x > 0$
(c) $0 \leq s_x \leq 1$
(d) $-1 \leq s_x \leq 1$
(e) Any number