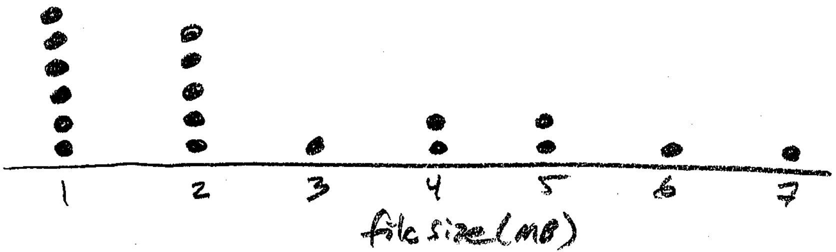
AP Statistics Ch 4 - Extra Practice **Name:** Solutions **Per:** ____

How much disk space does your music use? Here are the files sizes (in megabytes) for 18 randomly selected files on Tim's mp3 player:

1.1	1.3	1.3	1.6	1.9	1.9	2.1	2.2	2.4
2.5	2.7	3.0	4.4	4.7	5.0	5.6	6.2	7.5

#1. Make a dotplot of these data.



#2. Find the mean, standard deviation, and 5-number summary (min, Q1, median, Q3, max).

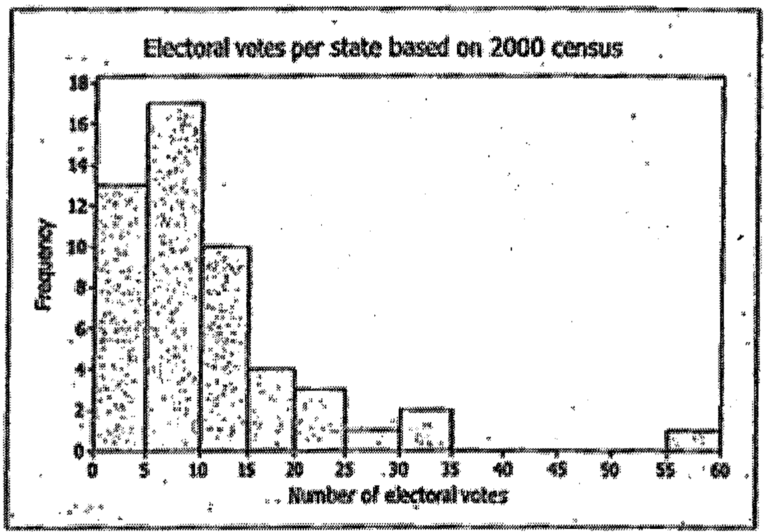
1 var stats : $\bar{x} = 3.19$ $s = 1.90$ min 1.1 Q1 1.9 med 2.45 Q3 4.7 max 7.5 (IQR = 4.7 − 1.9 = 2.8)

#3. Describe the overall pattern of the distribution and any possible outliers.

The file sizes are skewed right with a median of 2.45 MB and an IQR of 2.8 MB. There are no outliers.

(UF = 4.7 + 1.5(2.8) = 8.9)

#4. The histogram shows the distribution of electoral votes for the 50 United States and the District of Columbia. Describe the shape, center, and spread of the distribution.



The electoral vote data is skewed right with a median of 7.5 votes and an IQR of 10 votes.

The upperfence is at 27.5, so technically, all the data in the 30-35 and 55-60 ranges are outliers numerically, however only the data in the 55-60 bin appears to be an actual outlier.

L1	L2
2.5	13
7.5	17
12.5	10
17.5	4
22.5	3
27.5	1
32.5	2
55.5	1

1 var stats L1, L2

$\bar{x} = 11.19$
$s = 9.91$
$Q1 = 2.5$
med = 7.5
$Q3 = 12.5$

(IQR = 12.5 − 2.5 = 10)

(UF = 12.5 + 1.5(10) = 27.5)

Of the 50 species of oaks in the United States, 28 grow on the Atlantic coast and 11 grow in California. We are interested in the distribution of acorn volumes among oak species. Here are back-to-back stemplots on the volumes of acorns (in cubic centimeters) for these 39 oak species:

Volume of Acorns (cubic centimeters)

1varstats

$\bar{x} = 2.73$

$s = 2.73$

Qi = 1.1

med = 1.7

Q3 = 3.5

(IQR = 3.5 − 1.1 = 2.4)

(UF = 3.5 + 1.5(2.7) = 7.1)

Atlantic coast		California
998643	0	4
88864211111	1	06
50	2	06
6640	3	
8	4	1
	5	59
8	6	0
	7	1
1	8	
1	9	
5	10	
	11	
	12	
	13	
	14	
	15	
	16	
	17	1

Key: 2|6 = 2.6 cm³

1varstats

$\bar{x} = 4.846$

$s = 4.657$

Q1 = 1.6

med = 4.1

Q3 = 6

(IQR = 6 − 1.6 = 4.4)

(UF = 6 + 1.5(4.4) = 12.6)

#5. Use the stemplots to compare the distribution of acorn sizes between Atlantic Coast and California oak species.

The Atlantic Coast acorn distribution is skewed right, but the California distribution is more symmetrical and may be bimodal.

The median of the Atlantic distribution is 1.7 cm³ compared to a median of 4.1 cm³ for California (although there may actually be 2 subgroups with centers around 1.5 cm² and 5.0 cm².) In general, it appears that acorns are larger in California, on average.

The California distribution has significantly more variability with an IQR of 4.4 cm² compared to 2.4 cm² for Atlantic.

Both distributions contain samples in the outlier region — the 8.1, 9.1, and 10.5 values for Atlantic and the 17.1 value for California are all outliers.

3. **Thinking about shape.** Would you expect distributions of these variables to be uniform, unimodal, or bimodal? Symmetric or skewed? Explain why.

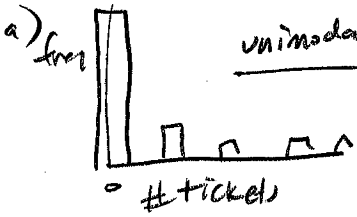
a) The number of speeding tickets each student in the senior class of a college has ever had.

b) Players' scores (number of strokes) at the U.S. Open golf tournament in a given year.

c) Weights of female babies born in a particular hospital over the course of a year.

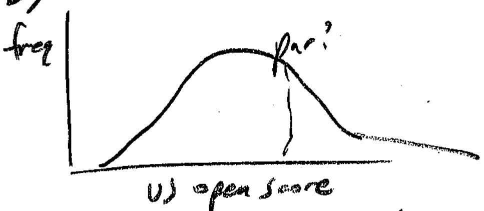
d) The length of the average hair on the heads of students in a large class.

a)



unimodal, skewed right

b)



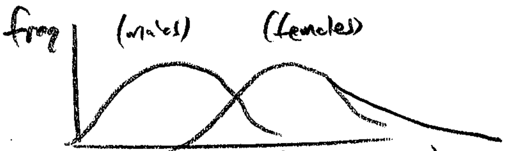
unimodal, symmetric (or maybe skewed right?)

c)

freq
wt female babies

unimodal, symmetric (probably "Normal")

d)



bimodal (by gender)
female may be skewed right

12. **The Great One.** During his 20 seasons in the NHL, Wayne Gretzky scored 50% more points than anyone who ever played professional hockey. He accomplished this amazing feat while playing in 280 fewer games than Gordie Howe, the previous record holder. Here are the number of games Gretzky played during each season:

79, 80, 80, 80, 74, 80, 80, 79, 64, 78, 73, 78, 74, 45, 81, 48, 80, 82, 82, 70

a) Create a stem-and-leaf display for these data using split stems.

b) Describe the shape of the distribution.

c) Describe the center and spread of this distribution.

d) What unusual feature do you see? What might explain this?

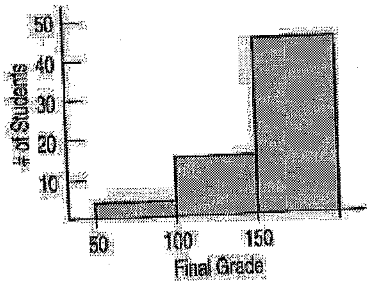
a)
```
8 |
8 | 00000 1022
7 | 9988
7 | 4340
6 |
6 | 4
5 |
5 |
4 | 58
4 |
```

b) skewed left

c) Center 79 games (median)
Spread Q1 = 73, Q3 = 80
IQR = 80 - 73 = 7 games

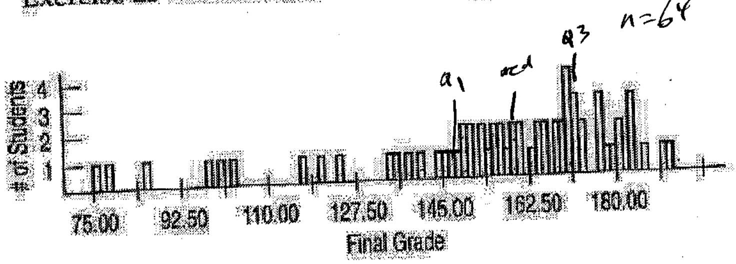
d) 3 Seasons are outliers:
64, 45, 48
- injured?
- contract issues?

25. Final grades. A professor (of something other than Statistics!) distributed the following histogram to show the distribution of grades on his 200-point final exam. Comment on the display.



The bins are really wide (too wide) so it is difficult to see shape details.
But it appears skewed left (higher grades are more common)

27. Final grades revisited. After receiving many complaints about his final grade histogram from students currently taking a Statistics course, the professor from Exercise 25 distributed the following revised histogram.



a) Comment on this display.
b) Describe the distribution of grades.

This is better, but maybe a little too narrow for bin widths. Almost too spread out to where every data value is in its own bin.

Center is 165–170
Most grades are in the 148–185 range
The distribution is skewed left and the grades around 75 are likely low outliers.
(IQR ≈ 172 − 150 = 22)
LF = Q1 − 1.5 IQR
= 150 − 1.5(22) = 117
(the bottom 6 data values are all outliers — below the lower fence)

38. Drunk driving. Accidents involving drunk drivers account for about 40% of all deaths on the nation's highways. The table tracks the number of alcohol-related fatalities for 20 years.

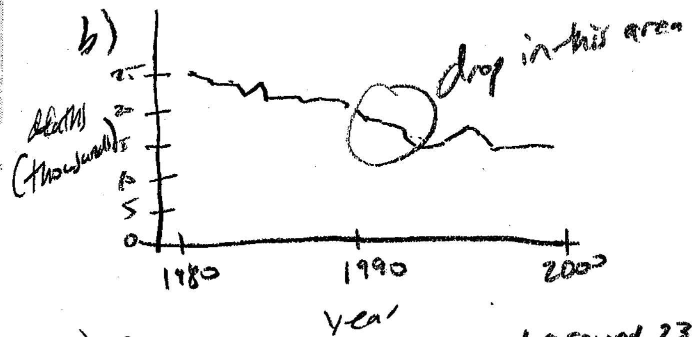
Year	Deaths (thousands)	Year	Deaths (thousands)
1982	25.2	1992	17.9
1983	23.6	1993	17.5
1984	23.8	1994	16.6
1985	22.7	1995	17.2
1986	24.0	1996	17.2
1987	23.6	1997	16.5
1988	23.6	1998	16.0
1989	22.4	1999	16.0
1990	22.0	2000	16.7
1991	19.9	2001	16.7

a) Create a stem-and-leaf display or a histogram of these data.
b) Create a timeplot.
c) Using features apparent in the stem-and-leaf display (or histogram) and the timeplot, write a few sentences about deaths caused by drunk driving.

a)

25	2
24	0
23	6866
22	740
21	
20	
19	9
18	
17	9522
16	650077

b)



c) Before 1990, deaths grouped around 23000 per year, but this dropped around 1990, in 1990's, deaths grouped around 17000 per year

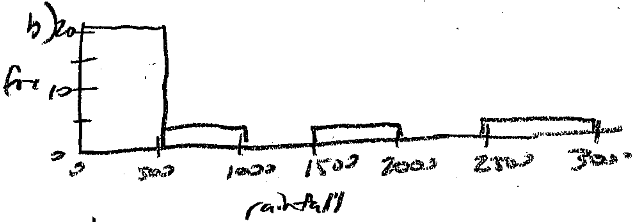
42. Rainmakers. The table lists the amount of rainfall (in acre-feet) from 26 clouds seeded with silver iodide.

log			log
3.44	2745	200	2.30
3,22	1697	198	2,29
3,22	1656	129	2,11
2,99	978	119	2,08
2,85	703	118	2,07
2,69	489	115	2,06
2,63	430	92	1,96
2,52	334	40	1,60
2,48	302	32	1,50
2,44	274	31	1,49
2,44	274	17	1.23
2,40	255	7	0,84
2,38	242	4	0,60

a) Why is "acre-feet" a good way to measure the amount of precipitation produced by cloud seeding?
b) Plot these data, and describe the distribution.
c) Create a re-expression of these data that produces a more advantageous distribution.
d) Explain what your re-expressed scale means.

a) data values give reasonable numbers

b)

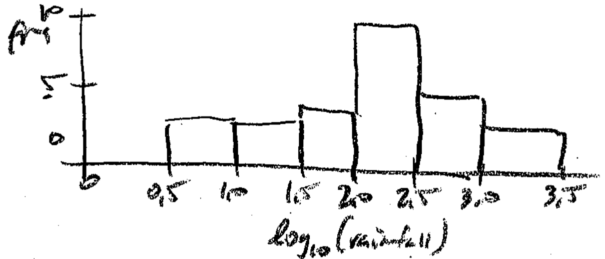


highly skewed right

c) take $\log_{10}$ of rainfall values

(suggestion: try taking the $\log_{10}$ of the rainfall values)

d)



Values are now the $\log_{10}$(rainfall)
(logarithms — double = 10 times more rainfall)
This distribution is symmetrical
with center at 2,25;
$\log_{10}(2,25) = \text{rain}$
So $\text{rain}_{\text{center}} = 10^{2,25} = 177$ acre-ft