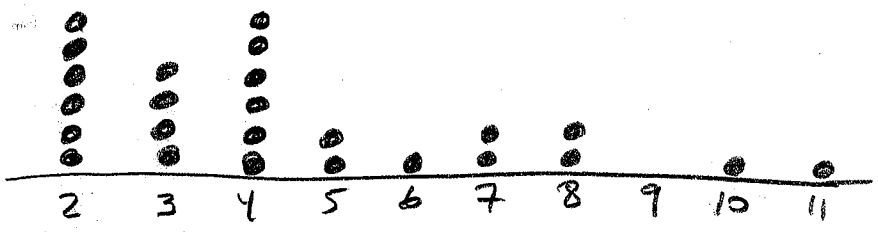
AP Statistics Ch 4 - Required Practice **Name:** Solutions **Per:** ____

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

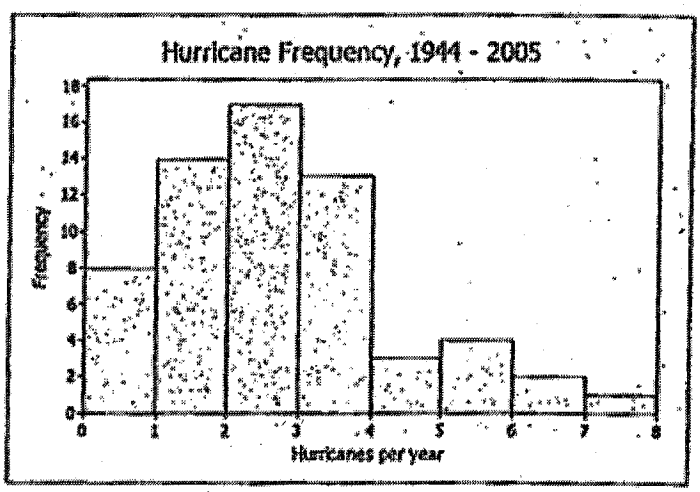
From 1-var stats: $\bar{x} = 4.16$ $s = 2.582$ min = 2 Q1 = 2.5 median = 4 Q3 = 6.5 max = 11

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

The word length is skewed right, with a median of 4 letters and an IQR of 4 letters. The points at 10 and 11 are not outliers.

$UF = Q3 + 1.5\,IQR$
$= 6.5 + 1.5(4)$
$= 12.5$

#4. The histogram 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.



The hurricane data is skewed right, with a median of 2 hurricanes, and an IQR of 2 hurricanes.

(data) L1	L2 (freq)
0	8
1	14
2	17
3	13
4	3
5	4
6	2
7	1

1 var-stats L1, L2:
$\bar{x} = 2.226$
$s = 1.624$
Q1 = 1
med = 2
Q3 = 3

IQR = 3 − 1 = 2

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	
96	1	023
86554	2	04679
774433	3	0244899
9665520	4	00444457
	5	
	6	
3	7	

Key: 1|4 = 14 home runs

1 Var Stats
$\bar{x} = 34.1$
$s = 14.04$
$n = 22$
$Q1 = 25$
med = 34
$Q3 = 45$
(IQR = 45 − 25 = 20)

1 Var Stats
$\bar{x} = 32.8$
$s = 11.18$
$n = 23$
$Q1 = 26$
med = 34
$Q3 = 44$
(IQR = 44 − 26 = 18)

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

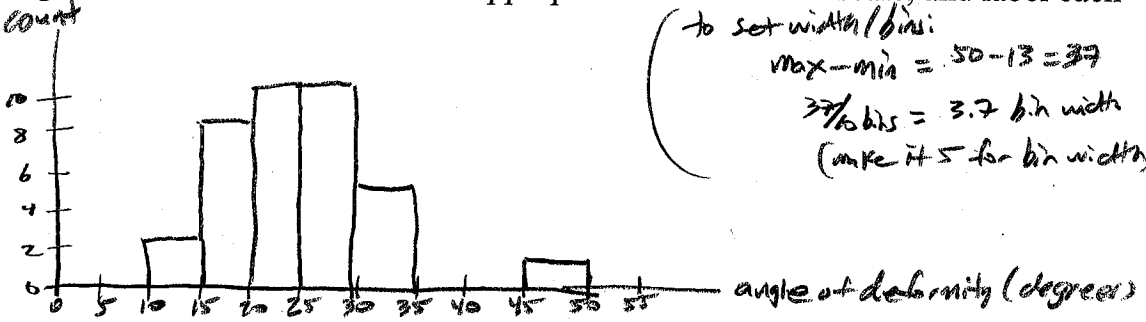
Both home run distributions are skewed left.
Both distributions are centered at a median of 34 home runs.
Bonds has slightly more variability in home runs with an IQR of 20 vs. Aaron's 18 homeruns.
There are no outliers in Aaron's data.
Bonds had one season with 73 homeruns which appears to be an outlier (although technically is not)

$UF = Q3 + 1.5(IQR)$
$= 45 + 1.5(20)$
$= 75$

Hallux abducto valgus (call it HAV) is a deformation of the big toe that is not common among young people and often requires surgery. Doctors used X-rays to measure the angle (in degrees) of deformity in 38 consecutive patients under the age of 21 who came to a medical center for surgery to correct HAV. The higher the angle measure the more severe the deformity. Here are the data.

13 14 16 16 17 18 18 20 20 20 21 21 21 21 22 23 25 25 25
25 26 26 26 26 28 28 28 30 30 30 31 32 32 32 34 38 38 50

#6. Make a histogram of these data. Choose an appropriate bin width and scale, and label each axis.



(to set width/bins:
max − min = 50 − 13 = 37
37/10 bins = 3.7 bin width
(make it 5 for bin width))

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

1 Var-Stats: $\bar{x} = 25.421$, $s = 7.475$, $n = 38$

min	Q1	Med	Q3	max
13	20	25	30	50

(IQR = 30 − 20 = 10)

#8. Write a brief discussion of the distribution of the angle of deformity among young patients needing surgery for this condition.

The distribution of angle of deformity is roughly symmetrical, with a mean of 25.4° and a standard deviation of 7.5°. The 50° data value is an outlier. (UF = 30 + 1.5(10) = 45)

Below are the resting heart rates of 26 ninth-grade biology students.

61	78	77	81	48	75	70	77	70	76	86	55	65
60	63	79	62	71	72	74	74	64	66	71	66	68

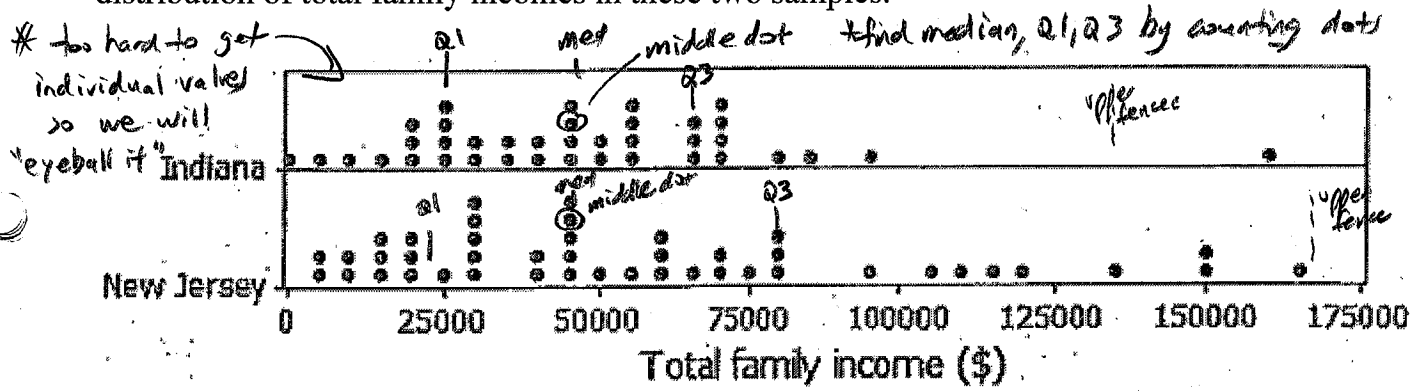
#9. Make a stemplot of these data with split stems.

← always include a key

6|5 = 65

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

#10. The dotplots below show the total family income of randomly-chosen individual from Indiana (38 individuals) and New Jersey (44 individuals). Write a few sentences comparing the distribution of total family incomes in these two samples.



(Shape) The Indiana distribution of income is roughly symmetrical, but the New Jersey distribution is skewed right.

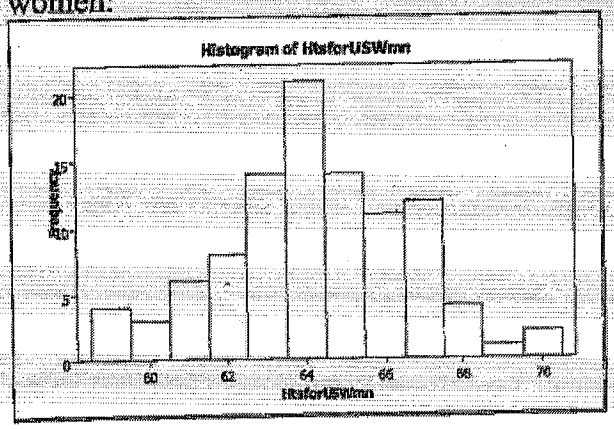
(c) Both distributions have a median of about $45,000

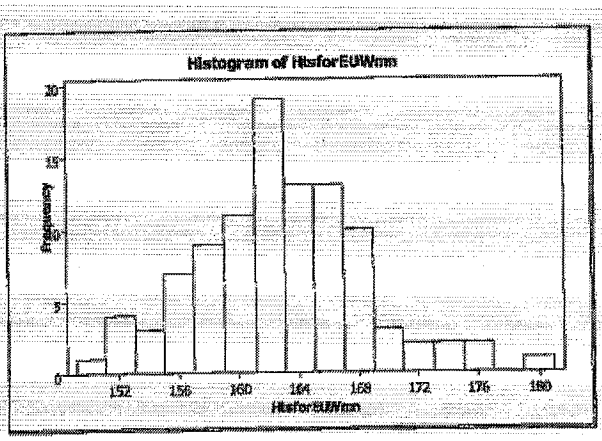
(Spread) The IQR of the New Jersey distribution is slightly larger than the Indiana distribution ($55500 vs. $40000).

(o) The Indiana distribution appears to have an outlier above the upper fence (approximated graphically). There appear to be no outliers in the New Jersey distribution.

Chapter 4 Practice Quiz

The following are histograms for the heights of 100 US women and the heights of 100 European women:





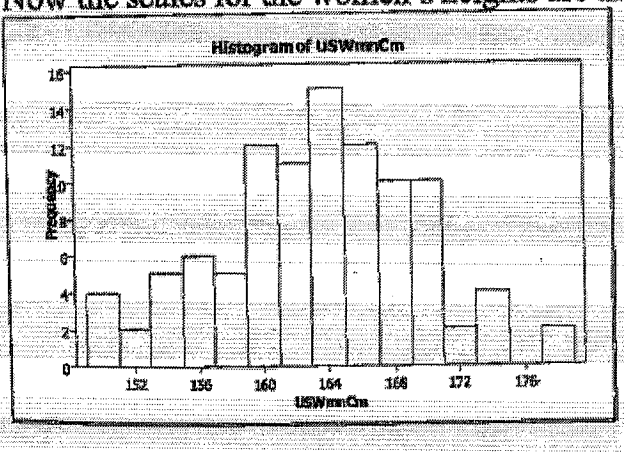
Note that the scales for the women's heights are very different, and thus it makes it hard to make a comparison between the heights of these women in the US and in Europe.

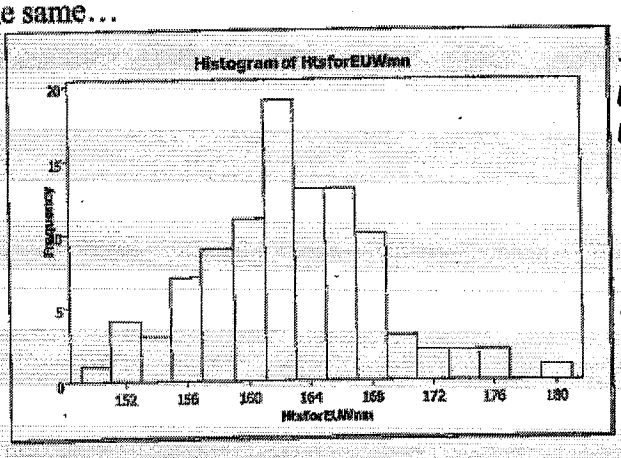
1. What might the cause of this difference in scale be?

US is in inches, European is in centimeters.

Now the scales for the women's heights are the same…

L1 L2
4
152 2
154 5
$\bar{x} = 163.3$
$s = 6.2$





L1 L2
150 1
152 4
154 3
$\bar{x} = 163.15$
$s = 5.9$

2. Compare the two distributions of the women's heights. Be sure to talk about shape, center, and spread. Both women's height distributions are roughly symmetrical with means of about 163 cm. The US women's heights are slightly more varied with standard deviation of 6.2cm compared to 5.9 cm for European women.

3. While the scales for heights the same in the second set of histograms are, there is still something that could be improved so that we could compare these two distributions better. Identify this improvement and explain why it would be better.

The frequency scales are different. This is hiding the fact that the European peak at ~162cm is much higher than the 164cm peak for U.S. It would be better to use a relative frequency (percentage) scale.