

NAME: Answer Key

**Directions:** Be sure to show all of your work. An answer alone will not receive any credit. You must show a formula or how you arrived at your answer. Partial credit will be given on all problems.

<sup>24</sup> 1. True/False questions.

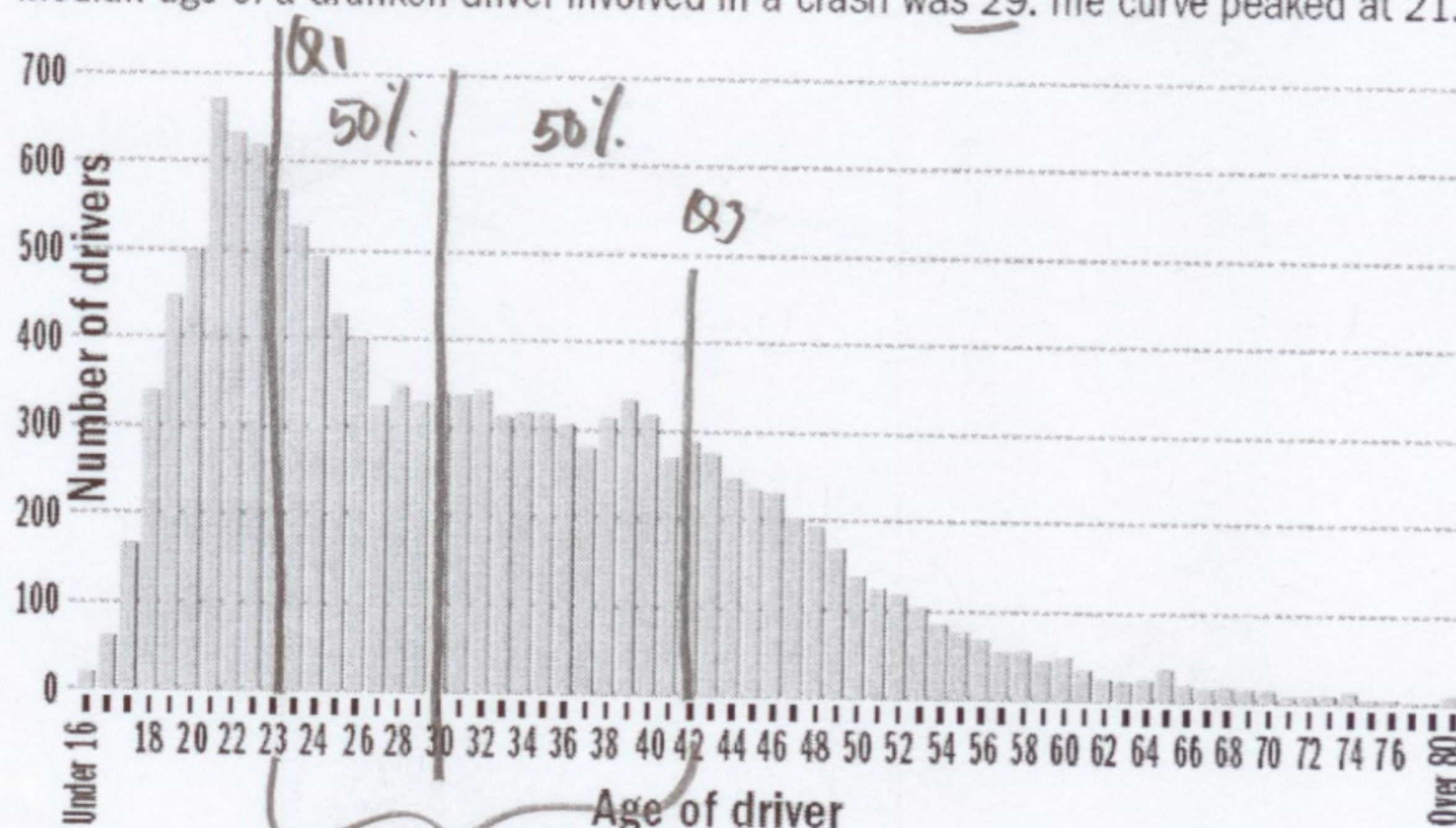
- ☐ T ☐ F <sup>3</sup>a. The Median and Trimmed Average are both robust measures of location.  
☐ T ☐ F <sup>3</sup>b. A population is only a set of numeric measurements.  
☐ T ☐ F <sup>3</sup>c. To be useful about making conclusions about the population, a sample should be representative of the population.  
☐ T ☐ F <sup>3</sup>d. A symmetric distribution would never have outliers.  
☐ T ☐ F <sup>3</sup>e. The range and IQR are both robust measures of spread.  
☐ T ☐ F <sup>3</sup>f. Normal distributions are always symmetric.  
☐ T ☐ F <sup>3</sup>g. Symmetric distributions are always Normal.  
☐ T ☐ F <sup>3</sup>h. Sample spaces are a list of possible outcomes from an experiment.

<sup>15</sup> 2. Use the figure below about drunk-driving crashes from the Cincinnati *Enquirer* to answer the True/False questions below.

- ☐ T ☐ F <sup>3</sup>a. This distribution is skewed left.  
☐ T ☐ F <sup>3</sup>b. The variance would be larger than the IQR.  
☐ T ☐ F <sup>3</sup>c. The average would be greater than the median.  
☐ T ☐ F <sup>3</sup>d. The IQR is approximately 50.  
☐ T ☐ F <sup>3</sup>e. Because of the outliers, the IQR would be a better measure of spread than the SD.

### Young and reckless

Median age of a drunken driver involved in a crash was 29. The curve peaked at 21.



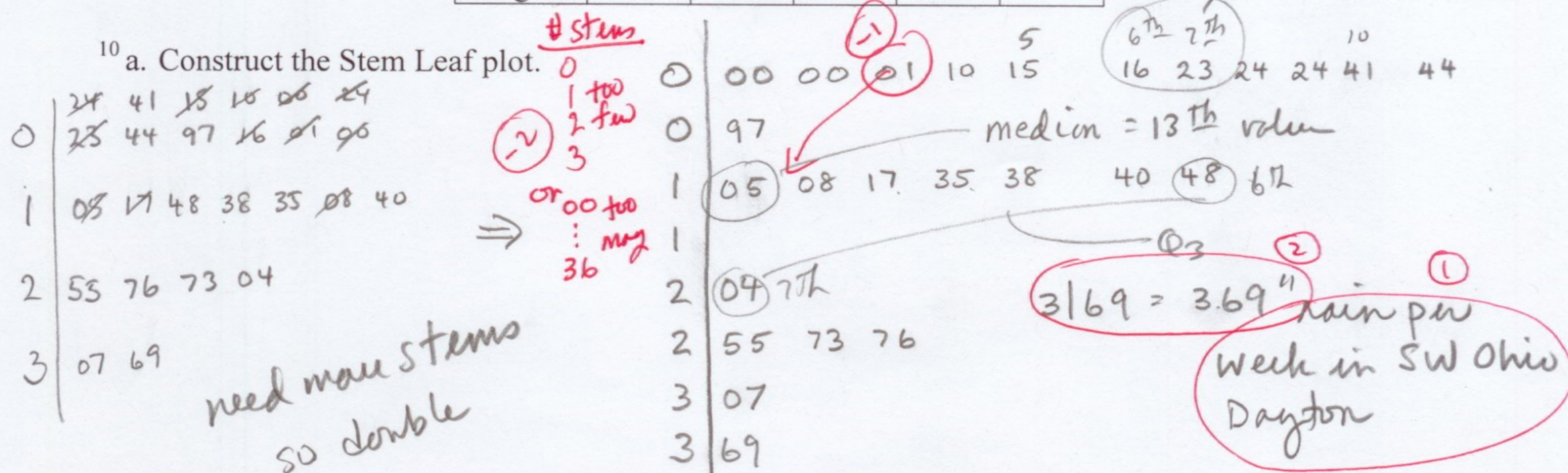
<sup>5</sup> 3. For the weekly rainfall data in SW Ohio seen in a later problem, the variance of weekly total rainfalls is approximately 1.2030. For an observation to be declared an outlier, how far away from the average would it need to be?

$sd = \sqrt{var} = \sqrt{1.2030} = 1.0968$   
 using Z-score  $\Rightarrow |Z| \geq 3 \text{ or } 4 \Rightarrow \frac{|obs - avg|}{sd} \geq 3 \text{ or } 4 \Rightarrow |obs - avg| \geq 3 \text{ or } 4 \times sd$   
 or  $3 \text{ or } 4 \times \sqrt{1.2030}$  or  $3 \text{ or } 4 \times 1.0968$   
 so  $|distance| > 3.2904 \text{ or } 4.3875$



- <sup>35</sup> 4. Weekly rainfalls in SW Ohio have varied quite a bit this year. Below are the weekly (Sunday thru Saturday) rainfall totals for the Dayton, Ohio NOAA site; use this information to answer the questions that follow.

March	2.55	0.23	2.76	1.05	
April	0.44	0.97	0.16	1.17	
May	1.48	3.07	0.24	0.41	
June	3.69	1.38	0.15	1.35	
July	1.08	2.73	0.10	2.04	0.00
August	1.40	0.01	0.00	0.24	



5# Round = {0.05, .20, 1.1, 1.75, 3.7} 5# truncated = {0.0, 0.15, 1.0, 1.7, 3.6}

- <sup>10</sup> b. Obtain the Five Number Summary, **BY HAND AND SHOW YOUR WORK!!!!** max = 3.69 min = 0.00

Med = 13th value since n=25 odd so median = 1.05

Q<sub>1</sub> = median of 12 values below Median so avg of 6th + 7th = .16 + .23 / 2 = 0.1950

Q<sub>3</sub> = " of " above median so avg of 1.48 + 2.04 / 2 = 1.76 so {0.00, 0.1950, 1.05, 1.76, 3.69}

- <sup>5</sup> c. What **TWO** facts about your 5# summary clearly illustrate this distribution is not normal?

Since normals are symmetric any indication of skew => not normal.

Since skew => non symmetry of distances btm 5# summary.

So distance btm max -> med ≠ min -> med, max -> Q<sub>3</sub> ≠ min -> Q<sub>1</sub>; Q<sub>3</sub> -> med ≠ Q<sub>1</sub> -> med

① 2.64 ≠ 1.05 ② 1.93 ≠ 0.195 ③ 0.71 ≠ 0.8550

- <sup>5</sup> d. To summarize the center of this distribution of rainfall amounts, would you use the average or median? Why? BRIEFLY!!!!

B/c not symmetric distrib skewed, median better measure of center

- <sup>5</sup> e. Using the upper fence, are there any outliers on the upper end of the distribution?

$$\text{Upper fence} = Q_3 + 1.5(IQR) = Q_3 + 1.5(Q_3 - Q_1) = 1.76 + 1.5(1.76 - 0.195) = 1.76 + 1.5(1.5650) = 1.76 + 2.3475 = 4.1075$$

no outliers since max is within upper fence.



<sup>21</sup> 5. Below are the boxplots of the weights of samples of different species of fish. Use this information and circle the most appropriate answer from the possibilities given below each statement.

<sup>3</sup>a. The Parkki species distribution is one of the \_\_\_\_\_ distributions:

more Normal less Normal more symmetric less symmetric cannot determine

*can't tell Normal from Boxplots!*

<sup>3</sup>b. \_\_\_\_\_ of the distributions are severely skewed to the left. *that only one could be is smelt but since the wgt > 0 can't be skewed left!*

None Some All cannot determine

<sup>3</sup>c. The trimmed avg for Roach would be \_\_\_\_\_ the trimmed avg of the Parkki species.

much greater than greater than about the same as smaller than much smaller than cannot determine

*b/c both very symmetric w/ similar medians*

<sup>3</sup>d. Approximately \_\_\_\_\_ of the Perch weigh less than all Pike.

100% 75% 50% 25% 0% none of these

*defn of median of Perch + end of whisker of Pike.*

<sup>3</sup>e. The standard deviation for Roach would be \_\_\_\_\_ the standard deviation for Bream.

smaller than about the same as larger than cannot determine

*b/c range + IQR much smaller than range + IQR of Bream so Avg squared deviation smaller!*

<sup>3</sup>f. There are six times as many Bream fish as Whitefish. The standard deviation for Bream would be \_\_\_\_\_ the standard deviation for Whitefish.

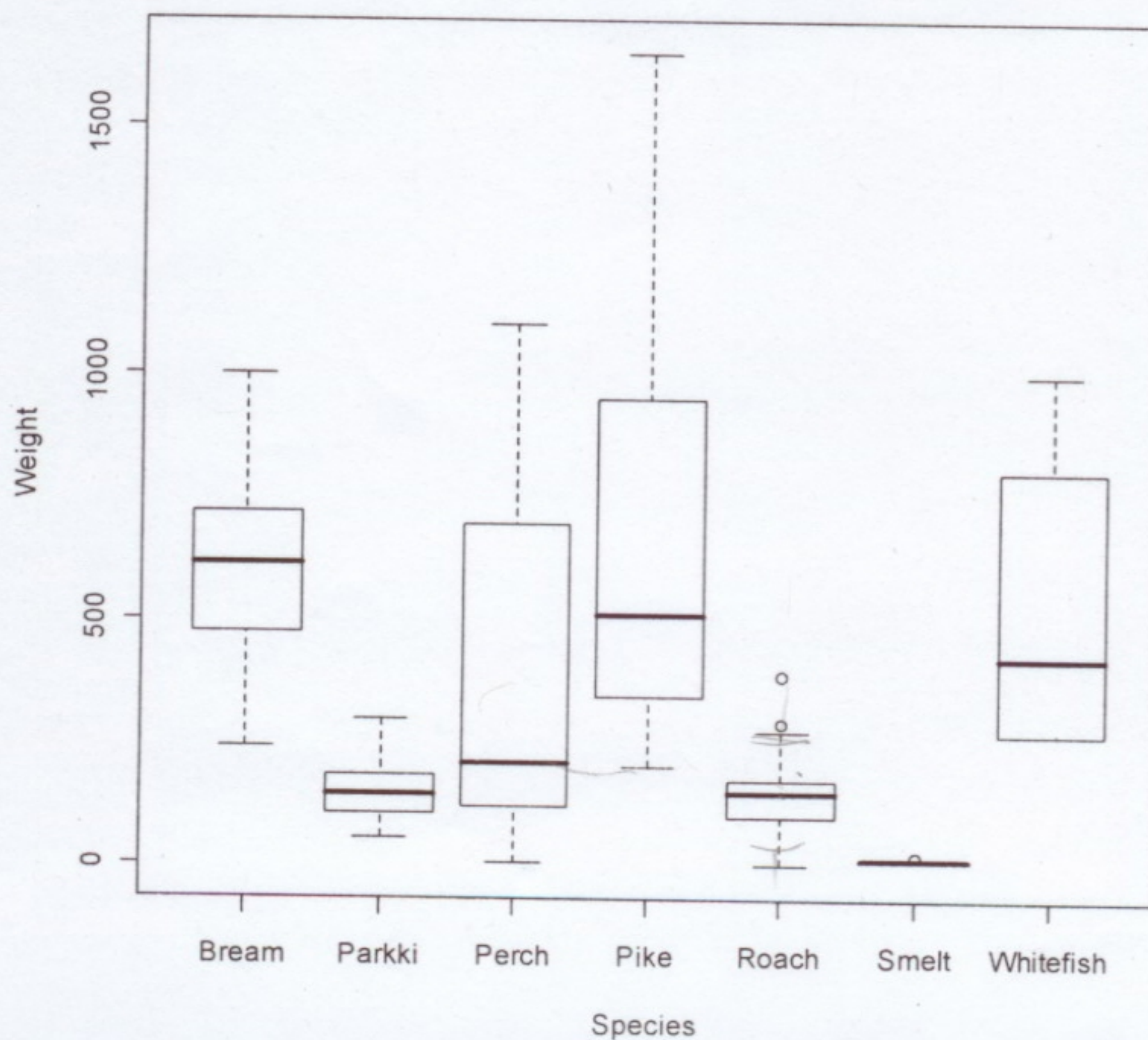
smaller than about the same as larger than cannot determine

*not important! ranges are about same but middle 50% of Bream closer to center than middle 50% of Whitefish.*

<sup>3</sup>g. When comparing measures of spread of the species' weights, EVERY measure of spread for the Smelt distribution would be the \_\_\_\_\_.

smallest about the same as largest cannot determine

*Smelt data is the most tightly grouped of all the data sets.*





```
1 | 2: represents 12
```

```
leaf unit: 1
```

```
n: 55
```

```
5 | 123
```

```
5 |
```

```
6 |
```

```
6 | 667788
```

```
7 | 024
```

```
7 | 5899
```

```
8 | 123444
```

```
8 | 5556667788999
```

```
9 | 00001233344
```

```
9 | 566667779
```

```
> summary(STA301Exam1$Exam1)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
51.00	78.50	86.00	83.49	92.50	99.00

```
> sd(STA301Exam1$Exam1)
```

```
[1] 11.93171
```