Lesson 1.2: How many pairs of shoes do you own?

1. How many pairs of shoes do you own? Record your answer on the board.

32 10 8
10 19 7
12 7

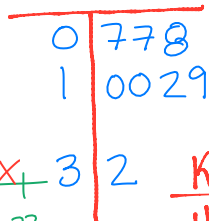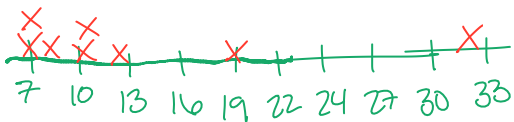2. Is "Number of pairs of shoes" a categorical or (quantitative) variable?

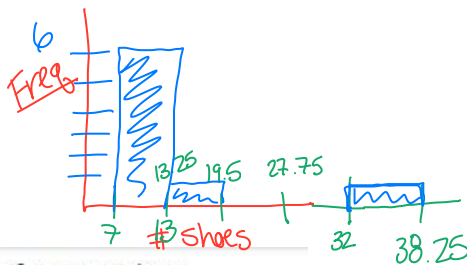3. Enter the data at www.stapplet.com. Make a dotplot, stemplot, and histogram and sketch each below.

dot plot

stem/leaf

$$
\begin{array}{c|c}
0 & 7\ 7\ 8 \\
1 & 0\ 0\ 2\ 9 \\
3 & 2 \\
\end{array}
$$

Key
1|0 means 10

Calculator histogram

6
Freq

13.25 19.5 27.75

7 #shoes 32 38.25

dotplot x-axis: 7 10 13 16 19 22 24 27 30 33

4. Describe the distribution of the number of pairs of shoes for your class.

Shape: Skewed Right, more data on left side

Outliers: appears to be 19 and 32 → most definitely 32 pull mean up

Center: median – 10 → resistant to outliers

Variability: data relatively close together until the outlier of 32 is put in

5. Which of the three types of display do you prefer? Why?

Stem/leaf more organized, easier to read

Histo – too much drawing → be sure scale equal

32 10 10 19 7 7 8 12 **L1**

6. Mrs. Gallas wonders if teachers have the same number of pairs of shoes as students. She asked her colleagues to record the number of pairs that they had. The results are below.

L2 15 8 10 29 14 7 22 35 6 15 13 12 9 10 *teachers*

7. Enter this data at stapplet.com. Be sure to make 2 groups (students and teachers).

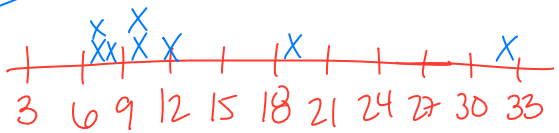**8.** Make dotplots, a side-by-side stemplot, and then histograms. Copy one of these graphs below.

Dot plot   Stem/leaf   histogram

stds | | teachers

stds

x x
xx x x   x   x
+ + + + + + + + + + + +
3 6 9 12 15 18 21 24 27 30 33

stds		teachers
	0	
	1	
	2	
	3	

teachers xxxxx xxxx   x   x   x
         x

9. Compare the distributions of the number of pairs of shoes for students versus teachers. Be sure to address shape, outliers, center, and variability.

Shape
Both teachers and stds appear to be skewed Right

Outliers
For teachers, the outliers appear to be 22, 29 and 35.
For students, the outliers appear to be 19 and 32.
Both sections have a gap, showing obvious outliers.

Center
Since both R skewed we use the median to represent the center. The median appears to be 10 for students and appears to be higher for teachers @ 12.5 shoes.

Variability
Both dist. appear to be relatively close together until we bring in the outliers, meaning that students and teachers have a relatively similar amount of shoes.

Important Ideas

Describing Distributions – we will use ly words...you MUST use CONTEXT!
- Shape — *skew / appnox normal*
- Outliers — *gap— must talk about.*
- Center — *med/mean*
- Variability — *spaling*

Don't forget your **SOCS**!

Check

1. The dotplot displays the scores of 21 statistics students on a 20-point quiz.

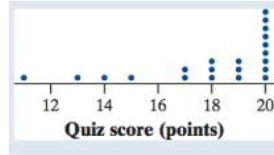(a) What percent of students scored higher than 16 points?

$$17/21 = 0.81 \times 100 = 81\%$$

(b) Describe the shape of the distribution.

left skew

(c) Are there any potential outliers? Why?

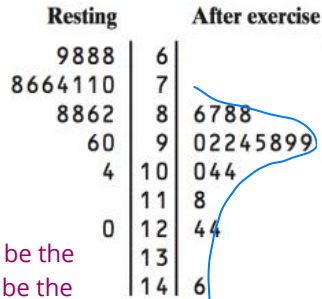*Score of 11 is much lower than the rest of the data.*

2. Here is a back-to-back stemplot of 19 middle school students' resting pulse rates and their pulse rates after 5 minutes of running.

Write a few sentences comparing the distributions of resting and after-exercise pulse rates.

Shape: The distribution of resting and after exercise are both similarly skewed right.

Resting		After exercise
9888	6	
8664110	7	
8862	8	6788
60	9	02245899
4	10	044
	11	8
0	12	44
	13	
	14	6

Key: 8|2 is a student whose pulse rate is 82 beats per minute.

Outliers: For resting heart rate, 120 appears to be the outlier. After exercise, 146 appears to be the outlier. They both have large gaps between them and the rest of their data points.

Center: The center for after exercise is larger than the center for resting. We would use the median because they have outliers and skew.

Variability: The after exercise group has a larger range, meaning the numbers are more spread out. The range for after is about 60 and for resting is 52.

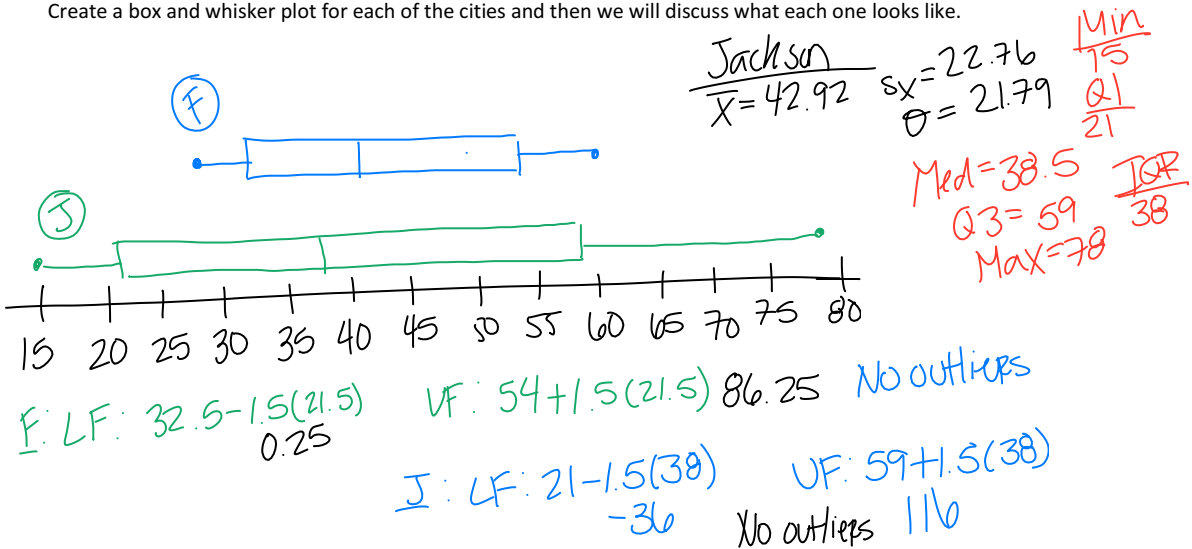Given the following set of data, find the summary stats.

Average Monthly High Temperature (F°) For Two U.S. Cities

	Franklin	Jackson
Jan.	29	15
Feb.	30	20
Mar.	35	22
Apr.	40	30
May	42	45
June	58	58
July	60	78
Aug.	59	77
Sept.	50	60
Oct.	42	58
Nov.	38	32
Dec.	30	20

430 × 220

Franklin

$\bar{X} = 42.75$

$S_x = 11.49$ → sample std. dev

$\sigma = 11.001$ → pop std. dev

IQR
21.5

Min = 29   Q1 = 32.5   Med = 41
Q3 = 54   Max = 60

List out the 5 number summary, the mean, S_x and σ, the IQR and see if you can find any outliers.

Create a box and whisker plot for each of the cities and then we will discuss what each one looks like.

(F)

(J)

15 20 25 30 35 40 45 50 55 60 65 70 75 80

Jackson
$\bar{X} = 42.92$ $S_x = 22.76$
$\sigma = 21.79$

Min
15
Q1
21
Med = 38.5
Q3 = 59
Max = 78
IQR
38

F: LF: 32.5 − 1.5(21.5)
0.25

VF: 54 + 1.5(21.5) 86.25 No outliers

J: LF: 21 − 1.5(38)
−36   No outliers

UF: 59 + 1.5(38)
116

The first quartile (Q1) of a dataset is 12 and the third quartile (Q3) is 18. What is the largest value above Q3 in the dataset that would not be a potential outlier?

Q1   Q3
12   18

UF: 18 + 1.5(6)
27 → largest value
Any value above 27

IQR = 18 − 12 = 6

Comparing the two box and whisker plots from above

Franklin appears to be approximately symmetric while Jackson appears to be slightly skewed to the right. There are no outliers in either city. The spread of the data seems to be more variable in Jackson since the IQR is larger while the spread of data in Franklin is smaller, with a range of 31 compared to a range of 63. The median for Jackson is less than the median for Franklin. Slightly more than 25% of the lower temperatures for Jackson are less than the minimum temperature for Franklin.