Multiple Choice: *Select the best answer for Exercises 40–43.*

40. For which of the following would it be *inappropriate* to display the data with a single pie chart?

(a) The distribution of car colors for vehicles purchased in the last month

(b) The distribution of unemployment percentages for each of the 50 states

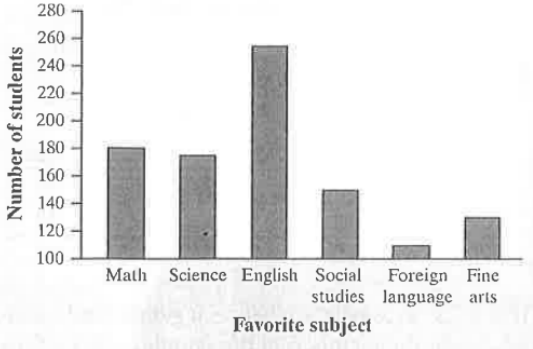
(c) The distribution of favorite sport for a sample of 30 middle school students

(d) The distribution of shoe type worn by shoppers at a local mall

(e) The distribution of presidential candidate preference for voters in a state

41. The following bar graph shows the distribution of favorite subject for a sample of 1000 students. What is the most serious problem with the graph?

d



(a) The subjects are not listed in the correct order.

(b) This distribution should be displayed with a pie chart.

(c) The vertical axis should show the percent of students.

(d) The vertical axis should start at 0 rather than 100.

(e) The foreign language bar should be broken up by language.

42. The Dallas Mavericks won the NBA championship in the 2010–2011 season. The two-way table displays the relationship between the outcome of each game in the regular season and whether the Mavericks scored at least 100 points.

d

		Points scored		
		100 or more	Fewer than 100	Total
Outcome of game	Win	43	14	57
	Loss	4	21	25
	Total	47	35	82

Which of the following is the best evidence that there is an association between the outcome of a game and whether or not the Mavericks scored at least 100 points?

(a) The Mavericks won 57 games and lost only 25 games.

(b) The Mavericks scored at least 100 points in 47 games and fewer than 100 points in only 35 games.

(c) The Mavericks won 43 games when scoring at least 100 points and only 14 games when scoring fewer than 100 points.

(d) The Mavericks won a higher proportion of games when scoring at least 100 points (43/47) than when they scored fewer than 100 points (14/35).

(e) The combination of scoring 100 or more points and winning the game occurred more often (43 times) than any other combination of outcomes.

43/ 100 High School Students

Gender

Dominant Hand	M	F	Total
R	x		90
L			90
total	40	60	(100)

No association

X =

a) 20 b) 30 c) 36 d) 45 e) Impossible to tell

- marginal relative freq
- Joint relative freq
- Conditional relative freq.

		Environmental club		
		No	Yes	Total
Snowmobile use	Never used	445	212	657
	Snowmobile renter	497	77	574
	Snowmobile owner	279	16	295
	Total	1221	305	1526

A **marginal relative frequency** gives the percent or proportion of individuals that have a specific value for one categorical variable.

Warm Up

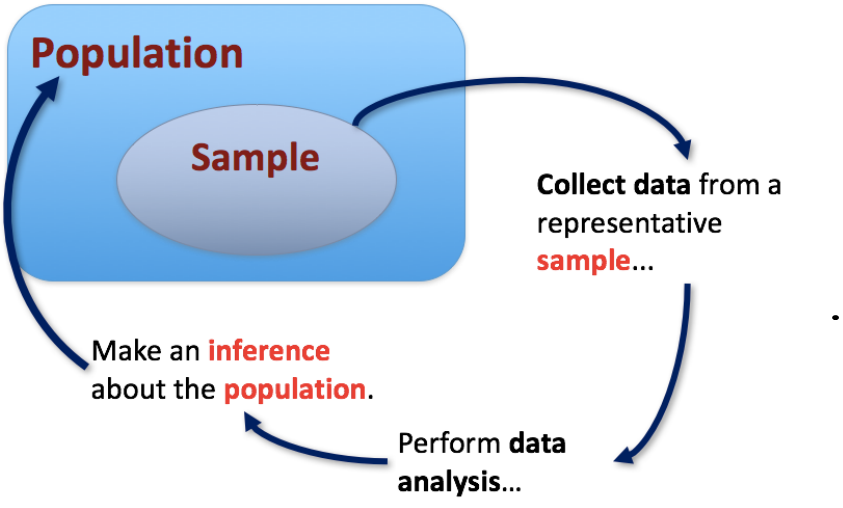
Pick Up the on-line textbook instructions

§ The Types of Quantitative Displays

Mr. C has a present for you. Are you ready?

Today: Make & Interpret & Compare Distributions using Dotplots, Stemplots, and Histograms.

Describe shape, center, outliers, and Variability of a distribution.



Dotplots

(pages 30-32)

A **dotplot** shows each data value as a dot above its location on a number line.

How to make a dotplot:

A **dotplot** shows each data value as a dot above its location on a number line.

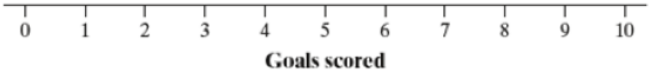
How to make a dotplot:

1) Draw a horizontal axis (a number line) and label it with the quantitative variable.

A **dotplot** shows each data value as a dot above its location on a number line.

How to make a dotplot:

1) Draw a horizontal axis (a number line) and label it with the quantitative variable.
2) Scale the axis from the minimum to the maximum value.

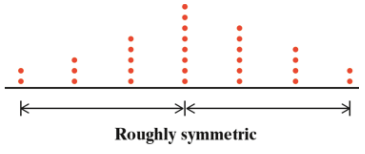


A **dotplot** shows each data value as a dot above its location on a number line.

How to make a dotplot:

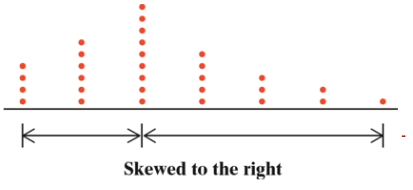
1) Draw a horizontal axis (a number line) and label it with the quantitative variable.
2) Scale the axis from the minimum to the maximum value.
3) Mark a dot above the location on the horizontal axis corresponding to each data value.

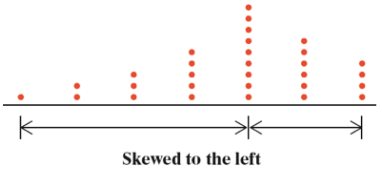
Describing Shape
(pages 32-34)



A distribution is roughly **symmetric** if the right side of the graph (containing the half of observations with the largest values) is approximately a mirror image of the left side.

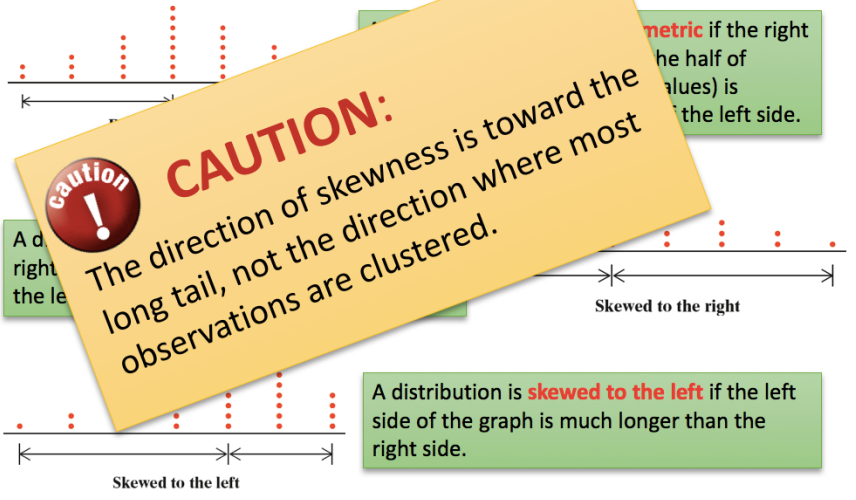
A distribution is **skewed to the right** if the right side of the graph is much longer than the left side.





A distribution is **skewed to the left** if the left side of the graph is much longer than the right side.

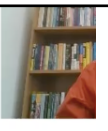
Describing Shape



metric if the right he half of alues) is the left side.

A distribution is skewed to the left if the left side of the graph is much longer than the right side.

Stemplots
(pages 37-40)



- Use for small data sets
- Include a key
- Don't omit any stems!
- Splitting stems

20	88
21	05679
22	02345566777
23	001345599
24	02
25	6

Key: 23|5 is a player with a head circumference of 23.5 inches.

20	88
21	0
21	5679
22	0234
22	5566777
23	00134
23	5599
24	02
24	
25	
25	6

These data represent the responses of 20 female AP Statistics students to the question, "How many pairs of shoes do you have?" Construct a stemplot.

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

1
2
3
4
5

Stems

These data represent the responses of 20 female AP Statistics students to the question, "How many pairs of shoes do you have?" Construct a stemplot.

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

1	1 | 93335
2	2 | 664233
3	3 | 1840
4	4 | 9
5	5 | 0701

Stems Add leaves

These data represent the responses of 20 female AP Statistics students to the question, "How many pairs of shoes do you have?" Construct a stemplot.

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

1		1	93335	1	33359
2		2	664233	2	233466
3		3	1840	3	0148
4		4	9	4	9
5		5	0701	5	0017

Stems | Add leaves | Order leaves

1
1
2
2
3
3

These data represent the responses of 20 female AP Statistics students to the question, "How many pairs of shoes do you have?" Construct a stemplot.

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

1		1	93335	1	33359
2		2	664233	2	233466
3		3	1840	3	0148
4		4	9	4	9
5		5	0701	5	0017

Stems | Add leaves | Order leaves | Add a key

Key: 4|9 represents a female student who reported having 49 pairs of shoes.

When data values are "bunched up", we can get a better picture of the distribution by **splitting stems**.
Two distributions of the same quantitative variable can be compared using a **back-to-back stemplot** with common stems.

Females

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

Males

14	7	6	5	12	38	8	7	10	10
10	11	4	5	22	7	5	10	35	7

When data values are "bunched up", we can get a better picture of the distribution by **splitting stems**.
Two distributions of the same quantitative variable can be compared using a **back-to-back stemplot** with common stems.

Females

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

Males

14	7	6	5	12	38	8	7	10	10
10	11	4	5	22	7	5	10	35	7

0
0
1
1
2
2
3
3
4
4
5
5

"split stems"

When data values are "bunched up", we can get a better picture of the distribution by **splitting stems**.
Two distributions of the same quantitative variable can be compared using a **back-to-back stemplot** with common stems.

Females

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

Males

14	7	6	5	12	38	8	7	10	10
10	11	4	5	22	7	5	10	35	7

0
0
1
1
2
2
3
3
4
4
5
5

"split stems"

Females		Males
	0	4
	0	555677778
333	1	0000124
95	1	
4332	2	2
66	2	
410	3	
8	3	58
	4	
9	4	
100	5	
7	5	

Key: 4|9 represents a student who reported having 49 pairs of shoes.

The "Resting Pulse" distribution being displayed appears to be skewed left. This is not the case.

CAUTION

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

On the AP exam students will:

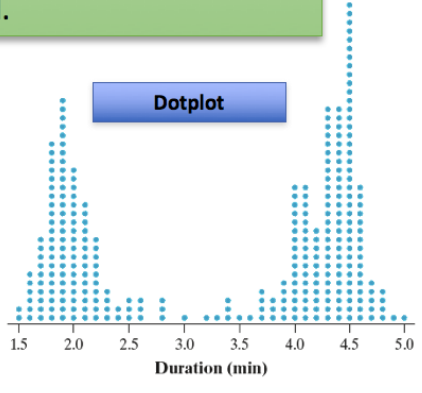
- sometimes be asked to make graphs, label them appropriately, and comment on their characteristics.
- but more often they will be asked to do some analysis based on graphs provided.

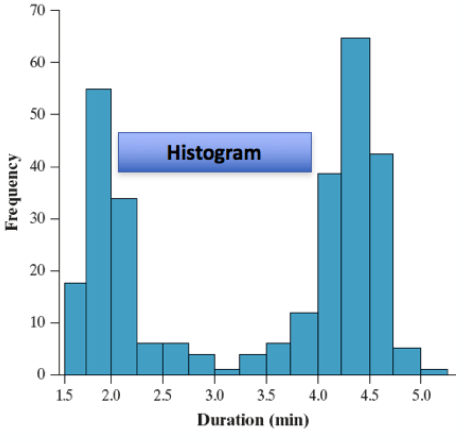
Histograms

(pages 40-44)

- Choosing intervals
- Vertical axis can show frequency or relative frequency.
- Advantage: Good for displaying large data sets
- Disadvantage: Lose sight of individual data values

A **histogram** shows each interval of values as a bar. The heights of the bars show the frequencies or relative frequencies of values in each interval.





How to make a histogram:

State	Percent	State	Percent	State	Percent
Alabama	2.8	Louisiana	2.9	Ohio	3.6
Alaska	7.0	Maine	3.2	Oklahoma	4.9
Arizona	15.1	Maryland	12.2	Oregon	9.7
Arkansas	3.8	Massachusetts	14.1	Pennsylvania	5.1
California	27.2	Michigan	5.9	Rhode Island	12.6
Colorado	10.3	Minnesota	6.6	South Carolina	4.1
Connecticut	12.9	Mississippi	1.8	South Dakota	2.2
Delaware	8.1	Missouri	3.3	Tennessee	3.9
Florida	18.9	Montana	1.9	Texas	15.9
Georgia	9.2	Nebraska	5.6	Utah	8.3
Hawaii	16.3	Nevada	19.1	Vermont	3.9
Idaho	5.6	New Hampshire	5.4	Virginia	10.1
Illinois	13.8	New Jersey	20.1	Washington	12.4
Indiana	4.2	New Mexico	10.1	West Virginia	1.2
Iowa	3.8	New York	21.6	Wisconsin	4.4
Kansas	6.3	North Carolina	6.9	Wyoming	2.7
Kentucky	2.7	North Dakota	2.1		

How to make a histogram:

1) Choose equal-width intervals that span the data.
2) Make a table that shows the frequency or relative frequency of individuals in each interval.

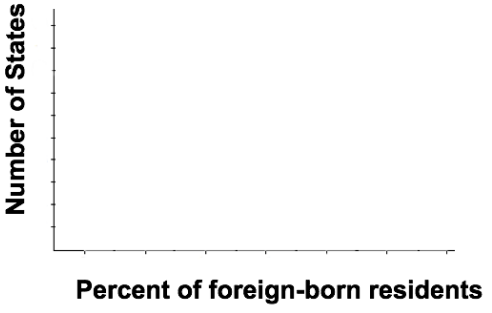
Frequency Table	
Class	**Count**
0 to <5	20
5 to <10	13
10 to <15	9
15 to <20	5
20 to <25	2
25 to <30	1
Total	50

State	Percent	State	Percent	State	Percent
Alabama	2.8	Louisiana	2.9	Ohio	3.6
Alaska	7.0	Maine	3.2	Oklahoma	4.9
Arizona	15.1	Maryland	12.2	Oregon	9.7
Arkansas	3.8	Massachusetts	14.1	Pennsylvania	5.1
California	27.2	Michigan	5.9	Rhode Island	12.6
Colorado	10.3	Minnesota	6.6	South Carolina	4.1
Connecticut	12.9	Mississippi	1.8	South Dakota	2.2
Delaware	8.1	Missouri	3.3	Tennessee	3.9
Florida	18.9	Montana	1.9	Texas	15.9
Georgia	9.2	Nebraska	5.6	Utah	8.3
Hawaii	16.3	Nevada	19.1	Vermont	3.9
Idaho	5.6	New Hampshire	5.4	Virginia	10.1
Illinois	13.8	New Jersey	20.1	Washington	12.4
Indiana	4.2	New Mexico	10.1	West Virginia	1.2
Iowa	3.8	New York	21.6	Wisconsin	4.4
Kansas	6.3	North Carolina	6.9	Wyoming	2.7
Kentucky	2.7	North Dakota	2.1		

How to make a histogram:

1) Choose equal-width intervals that span the data.
2) Make a table that shows the frequency or relative frequency of individuals in each interval.
3) Draw horizontal and vertical axes. Label the axes.

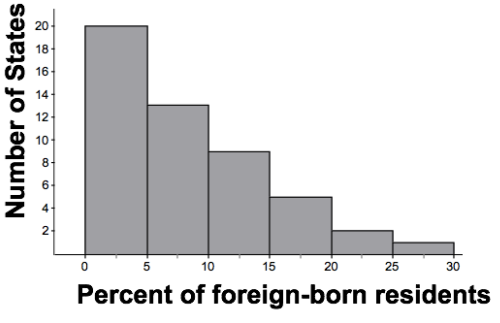
Frequency Table	
Class	**Count**
0 to <5	20
5 to <10	13
10 to <15	9
15 to <20	5
20 to <25	2
25 to <30	1
Total	50



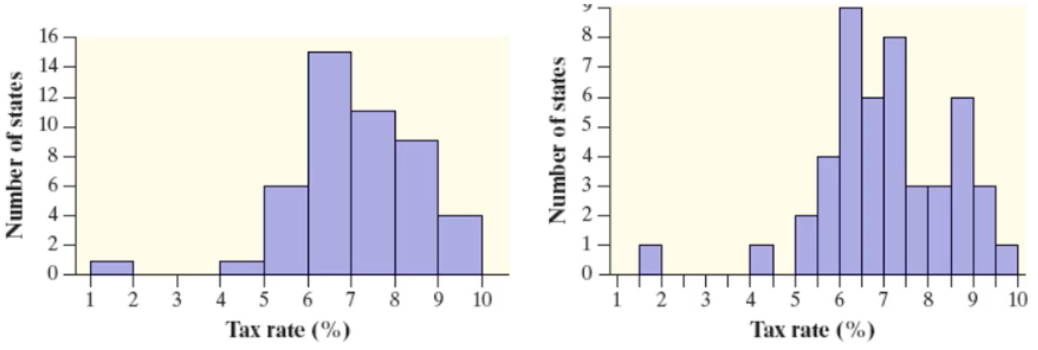
How to make a histogram:

1) Choose equal-width intervals that span the data.
2) Make a table that shows the frequency or relative frequency of individuals in each interval.
3) Draw horizontal and vertical axes. Label the axes.
4) Scale the axes.
5) Draw bars above the intervals. The bar heights correspond to the frequency or relative frequency of individuals in that interval.

Frequency Table	
Class	**Count**
0 to <5	20
5 to <10	13
10 to <15	9
15 to <20	5
20 to <25	2
25 to <30	1
Total	50



↳ Advantage to make with TI-83/84

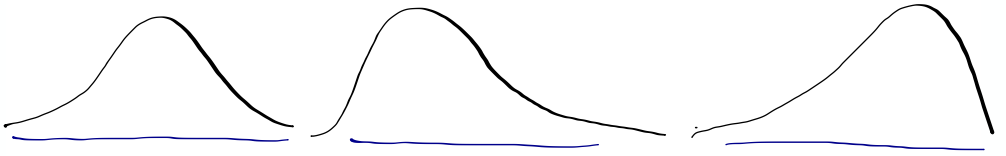


CAUTION:

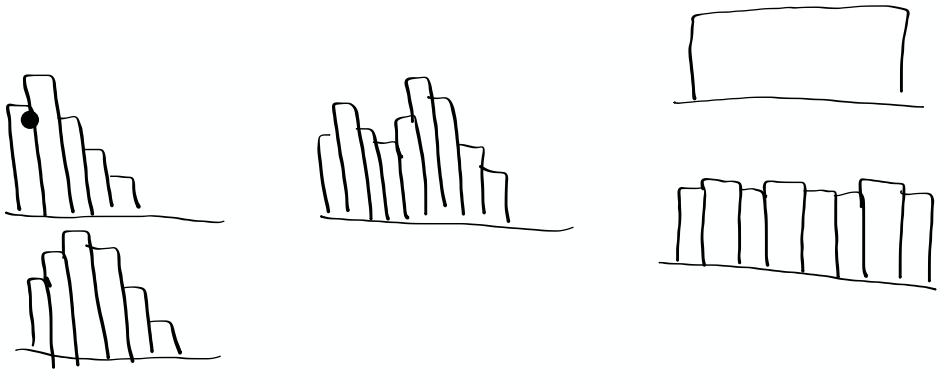
1) Don't confuse histograms and bar graphs.
2) Use percents or proportions instead of counts on the vertical axis when comparing distributions with different numbers of observations.
3) Just because a graph looks nice doesn't make it a meaningful display of data.

1. Briefly sketch a histogram to illustrate the following distribution shapes:

Symmetric | Skewed right | Skewed left



Single-peaked (Unimodal) | Double-peaked (Bimodal) | Uniform

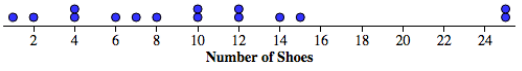


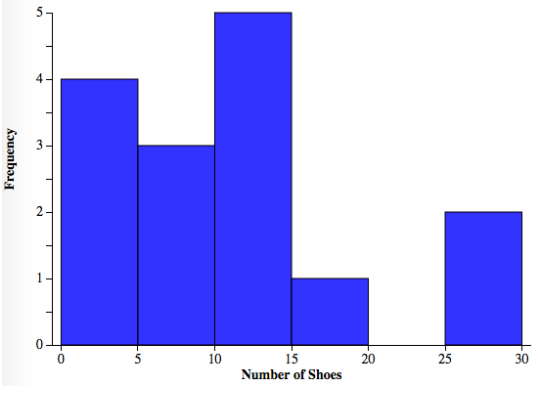
How many pairs of shoes do you own?

18	3	6					
7	20						
9	11						
8	15						
12	9						

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

Per. 1 1 10 6 14 4 25 2 4
12 25 10 7 12 15 8



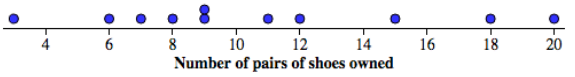


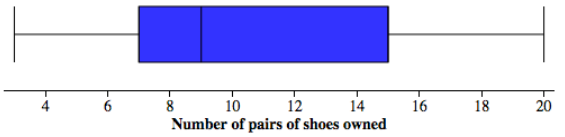
0	1 2 4 4
0	6 7 8
1	0 0 2 2 4
1	5
2	
2	5 5

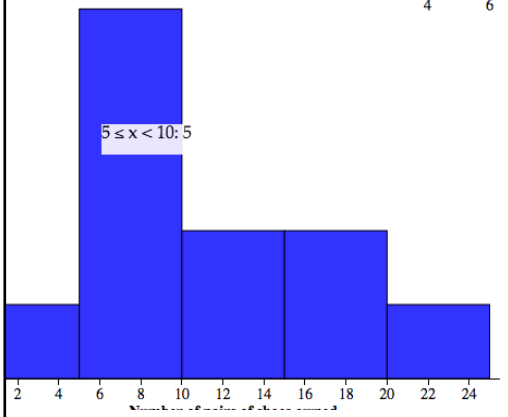
Key

1 | 2 represents 12 pairs of shoes

Per 5







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.

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.

Work together on first side up to 5

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

- Shape: slightly skewed right, unimodal
  Outliers (if any): possible outlier of 3
  Center (typical number of shoes owned) around 9
  Variability: # of pairs of shoes varies from 3 to 20

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

-ly words

roughly

approximately

unusually

extremely

slightly

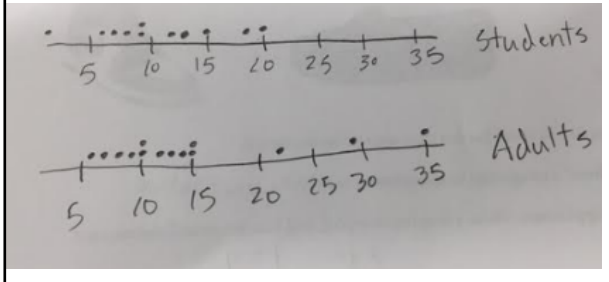
6. Mrs. Cedarlund wonders if teachers have the same number of pairs of shoes as students. He asked her colleagues to record the number of pairs that they had. The results are below.

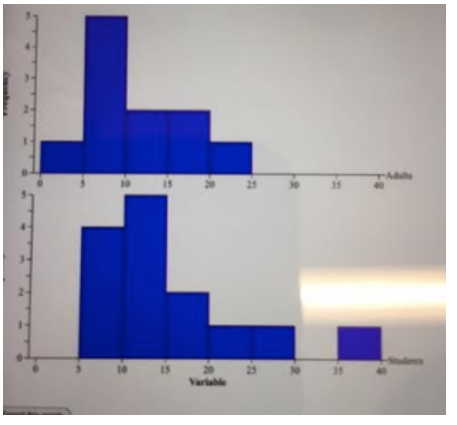
15 8 10 29 14 7 22 35 6 15 13 12 9 10

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.

Groups
6-9





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 unimodal, slightly skewed Right

Outliers Teacher: 35
Student: possibly 3

Center approximately 9 for students and approx. 12 for teachers

Variability # of pairs of shoes varies from 3-20 (students) and 6-35 (teachers)

roughly
approximately
unusually
extremely
slightly

S
O
C
V

+ context
+ comparitive language

Let's formalize

(from here on out try to be more precise with wordings on the examples)

Summary - Displaying Quantitative Data

Important Ideas:

Summary - Displaying Quantitative Data

Important Ideas:

Describing Distributions
- Use Context
- Use -ly words

Summary - Displaying Quantitative Data

Important Ideas:

Describing Distributions
- Use Context
- Use -ly words

approximately
slightly
roughly
extremely

Summary - Displaying Quantitative Data

Important Ideas:

Describing Distributions

- Use Context
- Use -ly words

approximately
slightly
roughly
extremely
⋮

Shape
Outliers
Center
Variability

S
O
C
V

Summary - Displaying Quantitative Data

Important Ideas:

Describing Distributions

- Use Context
- Use -ly words

approximately
slightly
roughly
extremely
⋮

Shape: skewed left/right, symmetric
Outliers:
Center: mean or median
Variability: How spread out the data is?
Range?

Use comparison language if comparing

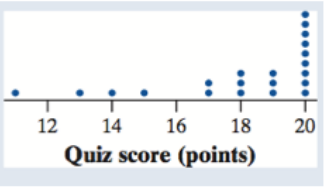
Check Your Understanding

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?

(b) Describe the shape of the distribution.

(c) Are there any potential outliers? Why?



Check Your Understanding

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?

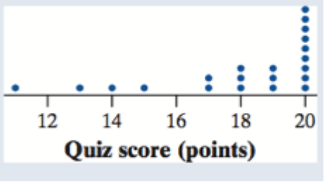
17/21 = 0.81 = 81%

(b) Describe the shape of the distribution.

Skewed left

(c) Are there any potential outliers? Why?

The score of 11 is much ~~larger~~ smaller than the rest of the data. It could be an outlier.



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.

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.

SHAPE

OUTLIERS

center

Variability

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.

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.

SHAPE

the distribution of resting pulse rates after exercise are both similarly skewed right

OUTLIERS

center

Variability

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.

Context comparing

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.

SHAPE the distribution of resting pulse rates after exercise are both similarly skewed right

OUTLIERS For resting pulse rates, 120 is a potential outlier, and for exercise, 146 is a possible outlier

Center

Variability

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.

Context comparing

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.

SHAPE the distribution of resting pulse rates after exercise are both similarly skewed right

OUTLIERS For resting pulse rates, 120 is a potential outlier, and for exercise, 146 is a possible outlier

Center The center is higher after exercise than for resting

Variability

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.

Context comparing

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.

SHAPE the distribution of resting pulse rates after excercise are both similarly skewed right

OUTLIERS For resting pulse rates, 120 is a potential outlier, and for excercise, 146 is a possible outlier

center The center is higher after excercise than for resting

Variability

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.

Context comparing

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.

SHAPE the distribution of resting pulse rates after excercise are both similarly skewed right

OUTLIERS For resting pulse rates, 120 is a potential outlier, and for excercise, 146 is a possible outlier

center The center is higher after excercise than for resting

Variability there is a high variability after excercise with a range of 60 compared to 52 bpm for resting.

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.

Context comparing

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.

SHAPE the distribution of resting pulse rates after excercise are both similarly skewed right

OUTLIERS For resting pulse rates, 120 is a potential outlier, and for excercise, 146 is a possible outlier

center The center is higher after excercise than for resting

Variability there is a high variability after excercise with a range of 60 compared to 52 bpm.

Assignment

1.2...45, 49, 51, 59, 63

and study pp.30-33