Good morning •

Today • → Start 1.2 (See Learning Targets on side board)

→ Small Learning Check Quiz (LCQ)

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

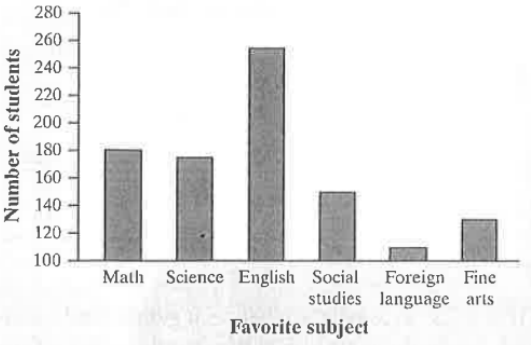
(c) The distribution of favorite sport for a sample of 30 middle school students

(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.

Section 1.2
Learning Targets

- Make and interpret dotplots, stemplots, and histograms of quantitative data.
- Identify the shape of a distribution from a graph.
- Describe the overall pattern (shape, center, and variability) of a distribution and identify any major departures from the pattern (outliers).
- Compare distributions of quantitative data using dotplots, stemplots, and histograms.

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.

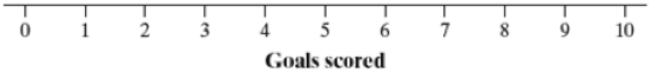
How to make a dotplot:

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.

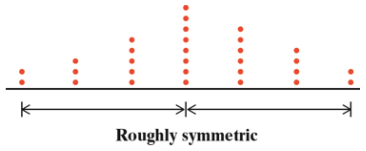
page 31

Toyota 4RUNNER DATA

ON-line Student site → TOOLS

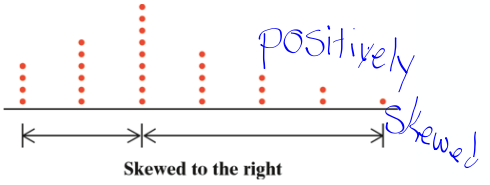
Describing Shape

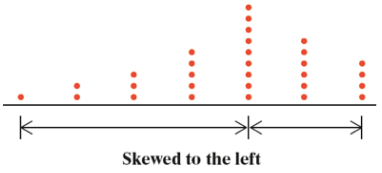
(pages 32-34)



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.

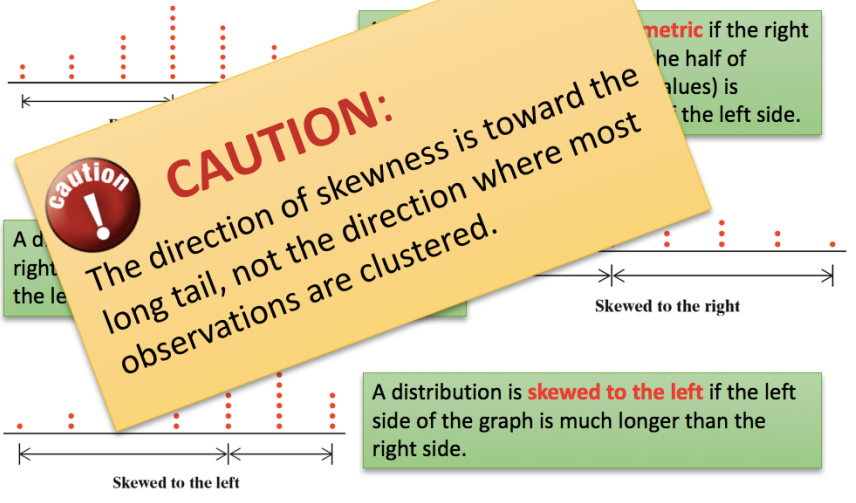




A distribution is **skewed to the left** if the left side of the graph is much longer than the right side.

Negatively Skewed

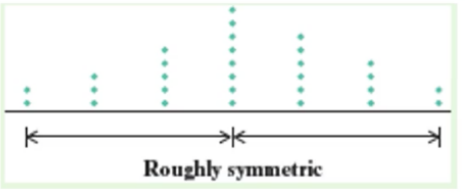
Describing Shape

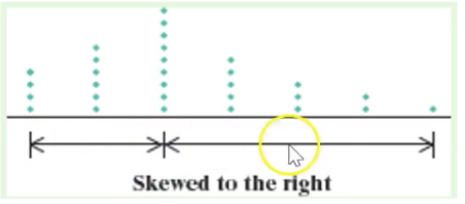


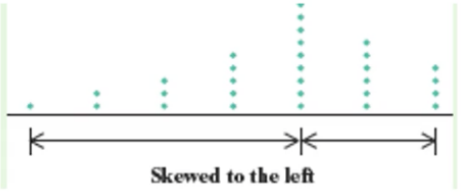
CAUTION:

The direction of skewness is toward the long tail, not the direction where most observations are clustered.

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

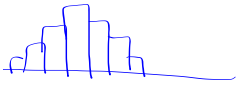




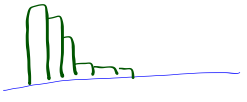


1. Briefly sketch a histogram to illustrate the following distribution shapes:

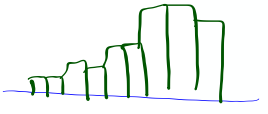
Symmetric



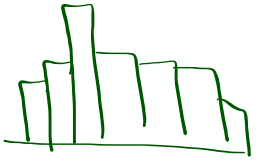
Skewed right



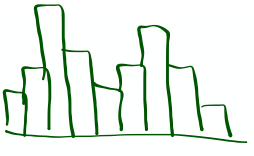
Skewed left



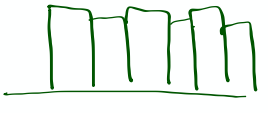
Single-peaked (~~Uni~~modal)



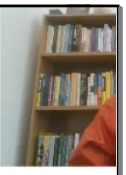
Double-peaked (Bi~~modal~~)



Uniform



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	0013
23	5599

Stemplots

A **stemplot** shows each data value separated into two parts: a stem, which consists of all but the final digit, and a leaf, the final digit. The stems are ordered from lowest to highest and arranged in a vertical column. The leaves are arranged in increasing order out from the appropriate stems.

How to make a stemplot:

1) Separate each observation into a stem, consisting of all but the final digit, and a leaf, the final digit. Write the stems in a vertical column with the smallest at the top. Draw a vertical line at the right of this column.
2) Write each leaf in the row to the right of its stem.
3) Arrange the leaves in increasing order out from the stem.
4) Provide a key that identifies the variable and explains what the stems and leaves represent.

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

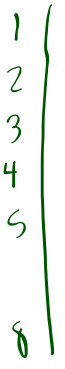
Stems Add leaves

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



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

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

Add a key

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

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.

3.
What is the most important thing to remember when making a stem plot?

✓ don't hide gaps
✓ include key

On the AP exam students will:

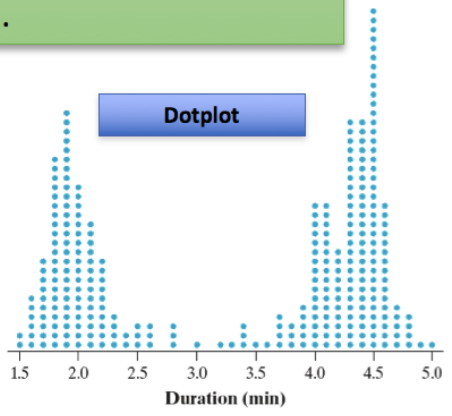
- sometimes be asked to make graphs, label them appropriately, and comment on their characteristics.

- but more often they will be asked to do some analysis based on graphs provided.

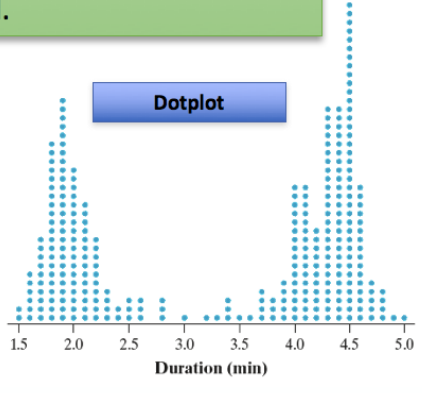
Histograms
(pages 40-44)

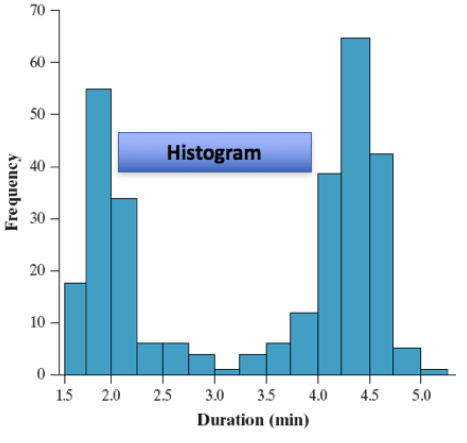
- Choosing intervals
- Vertical axis can show frequency or relative frequency.
- Advantage: Good for displaying large data sets
- Disadvantage: Lose sight of individual data values

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.



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.

Frequency Table	
Class	
0 to <5	
5 to <10	
10 to <15	
15 to <20	
20 to <25	
25 to <30	
Total	

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.

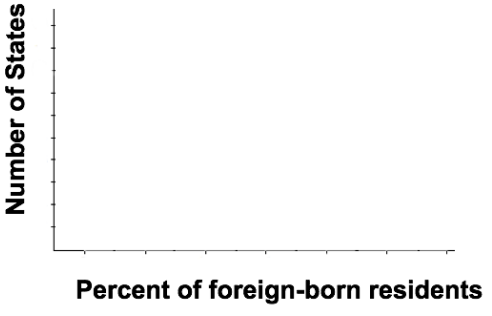
Frequency Table	
Class	**Count**
0 to <5	20
5 to <10	13
10 to <15	9
15 to <20	5
20 to <25	2
25 to <30	1
Total	50

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.

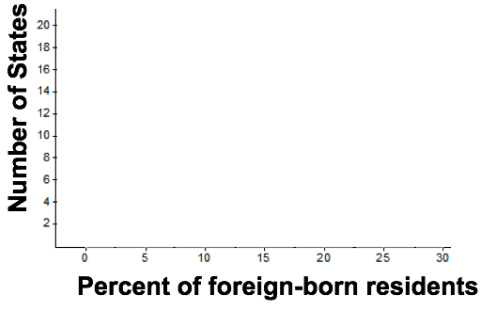
Frequency Table	
Class	**Count**
0 to <5	20
5 to <10	13
10 to <15	9
15 to <20	5
20 to <25	2
25 to <30	1
Total	50



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.

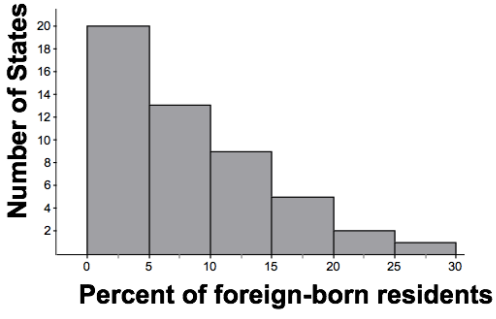
Frequency Table	
Class	**Count**
0 to <5	20
5 to <10	13
10 to <15	9
15 to <20	5
20 to <25	2
25 to <30	1
Total	50

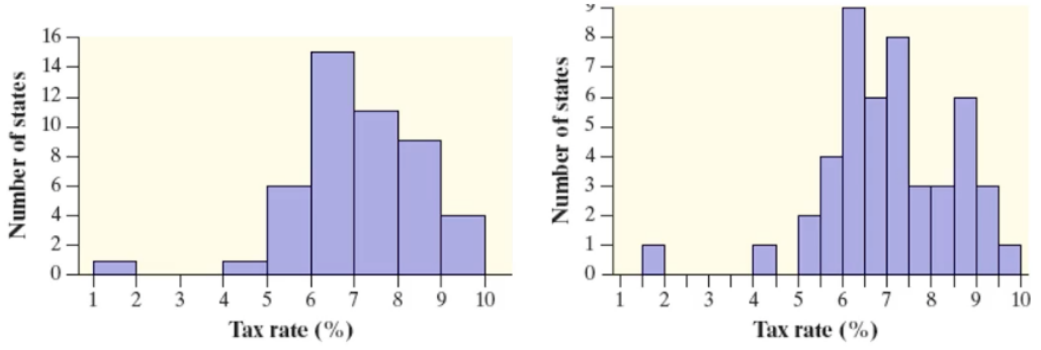


How to make a histogram:

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





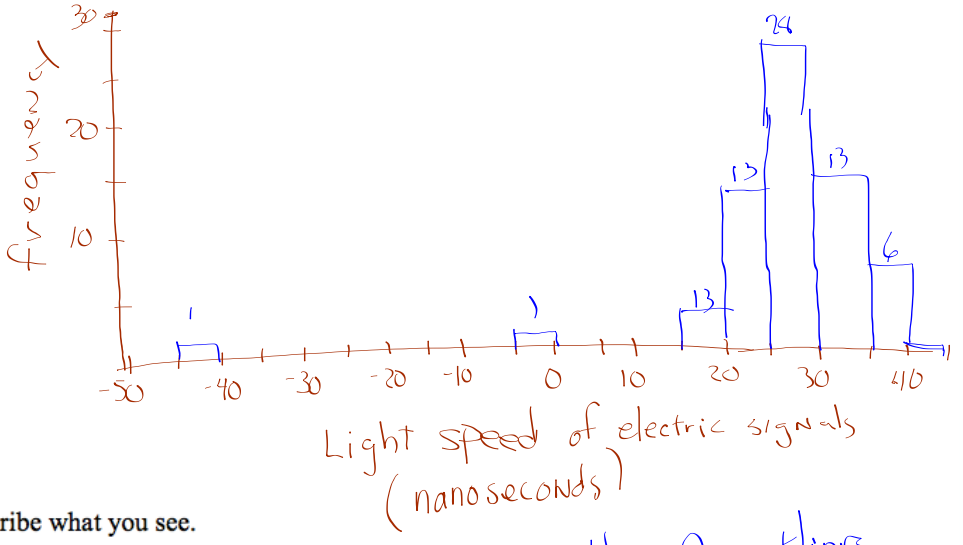
4. Example The speed of light

Light travels fast, but it is not transmitted instantly. Light takes over a second to reach us from the moon and over 10 billion years to reach us from the most distant objects in the universe. Because radio waves and radar also travel at the speed of light, an accurate value for that speed is important in communicating with astronauts and orbiting satellites. An accurate value for the speed of light is also important to computer designers because electrical signals travel at light speed. The first reasonably accurate measurements of the speed of light were made over 100 years ago by A. A. Michelson and Simon Newcomb. The table below contains 66 measurements made by Newcomb between July and September 1882.

Newcomb measured the time in seconds that a light signal took to pass from his laboratory on the Potomac River to a mirror at the base of the Washington Monument and back, a total distance of about 7400 meters. Newcomb's first measurement of the passage time of light was 0.000024828 second, or 24,828 nanoseconds. (There are 10^9 nanoseconds in a second.) The entries in the table record only the deviation from 24,800 nanoseconds.

28	26	33	24	34	-44	27	16	40	-2	29	22	24	21
25	30	23	29	31	19	24	20	36	32	36	28	25	21
28	29	37	25	28	26	30	32	36	26	30	22	36	23
27	27	28	27	31	27	26	33	26	32	32	24	39	28
24	25	32	25	29	27	28	29	16	23				

1. Construct a histogram to display these data. (What intervals/bins/classes will you use?)



2. Describe what you see.

Roughly symmetric, with 2 outliers

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.

LCQ

Assignment

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