2.1: Random Samples

_____ – sample that is representative of the entire population.
        - Each member of the population has an _____ chance of being
        included in the sample.
        - Each sample of the same size is _____ likely to be selected.

_____ - of n measurements from a population
is one selected in such a manner that every sample of size n from the population
has equal probability of being selected, and every member of the population has
equal probability of being included in the sample.

Is open space around metropolitan areas important? Players of the Colorado
Lottery might think so, since some of the proceeds of the game go to fund open
space and outdoor recreational space. To play the game, you pay one dollar and
choose any six different numbers from the group of numbers 1 through 42. If your
group of six numbers matches the winning group of six numbers selected by
simple random sampling, then you are a winner of a grand prize of at least 1.5
million dollars.

(a) Is the number 25 as likely to be selected in the winning group of six numbers as
the number 5?

(b) Could all the winning numbers be even?

(c) Your friend always plays the numbers
1   2   3   4   5   6
Could she ever win?

How do we get random samples???

One way to pick a random sample is to number cards. For example, suppose we need to know how many bicycle helmets fit the given safety requirements. We want to pick a random sample of 25 bicycles out of a shipment of 200. We could number cards from 1 – 200 and select 25 of these cards at random.

Another method to selecting numbers is to use a _____.
You can make one of these for yourself by writing the digits 0 – 9 on separate cards and mixing these cards up. Draw a card and record the digit, return the card, and mix the cards up again. Draw another card; record the digit, etc…

Random Numbers on the Calculator

1. Be sure you "seed" your calculator. Â This will insure that the same random numbers do not appear on everyone's calculator. Enter a random number in the calculator such as the student ID number or telephone number. Then press [STO->] [MATH] "PRB" "1:rand".

Note: This step only needs to be completed once unless you reset the calculator!   And you will not get the same numbers!

```
TI-84 Plus Silver Edition
   TEXAS INSTRUMENTS
4568213→rand
          4568213
```

2. Specify the minimum and maximum integers possible, and how many random integers to generate.   For example, [MATH] "PRB" "5:randInt(1,10,5)" [ENTER] generates 5 numbers between 1 and 10.

Note: Hit [Enter] to get 5 more random numbers between 1 and 10.   Of course, you will not likely get the same random numbers.

```
randInt(1,10,5)
     {5 10 5 6 1}
randInt(1,10,5)
     {3 1 8 5 4}
```
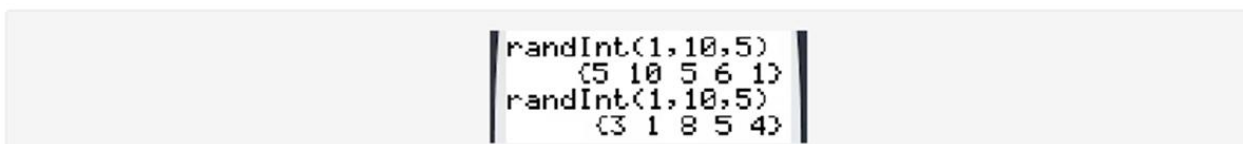
Examples: Using your calculator, pick a random sample for each of the following:

1) You want to pick a random sample of 30 cars from a population of 500 cars.

Solution: Assign each car a different number from 1 – 500. Then use your calculator.
*** If a number is repeated, skip it.

Now try these out (Explain you work!!!):
2) You want to pick a random sample of 25 trucks from a population of 300 trucks.

3) You want to pick a random sample of 40 jerseys from a population of 1000 jerseys.

4) You want to pick a random sample of 10 volleyballs from a population of 65 volleyballs.

5) Use a random-number table to get a list of 9 random numbers from 1 – 772.

6) Use a random-number table to get a list of 6 random numbers from 1 – 3256.

7) Use a random-number list to get a list of 8 random numbers from 278 – 881.

8) Use a random-number list to get a list of 4 random numbers from 32 – 70.

Another important use of random-number tables is in _____.
It is not always necessary to read the random-number tables by rows or blocks, but there are also other schemes that we can follow. Some of these include columns and diagonals. However, it is necessary to use a consistent scheme in order to pick up a random sample.

A _____ is a numerical facsimile or representation of a real-world phenomenon.

Use a random-number table or your calculator to simulate the outcomes of tossing a balanced (that is, fair) penny 10 times.

(a) How many outcomes are possible when you toss a coin once?

(b) There are several ways to assign numbers to the two outcomes. Because we assume a fair coin, we can assign an even digit to the outcome "heads" and an odd digit to the outcome "tails."

(c) What are the outcomes associated with the 10 digits?

(d) If you repeat the process, will you get the same sequence of outcomes?

2.1: Homework

1) Marcie conducted a study of the cost of breakfast cereal. She recorded the costs of several boxes of cereal. However, she neglected to take into account the number of servings in each box. Someone told her not to worry because she just had some sampling error. Comment on that advice.

2) Consider the students in your statistics class as the population and suppose they are seated in four rows of 10 students each. To select a sample, you toss a coin. If it comes up heads, you use the 20 students sitting in the first two rows as your sample. If it comes up tails, you use the 20 students sitting in the last two rows as your sample.
(a) Does every student have an equal chance of being selected for the sample? Explain.

(b) Is it possible to include students sitting in row 3 with students sitting in row 2 in your sample? Is your sample a simple random sample? Explain.

(c) Describe a process you could use to get a simple random sample of size 20 from a class of size 40.

3) Suppose you are assigned the number 1, and the other students in your statistics class call out consecutive numbers until each person in the class has his or her own number. Explain how you could get a random sample of four students from your statistics class.

(a) Explain why the first four students walking into the classroom would not necessarily form a random sample.

(b) Explain why four students coming in late would not necessarily form a random sample.

(c) Explain why four students sitting in the back row would not necessarily form a random sample.

(d) Explain why the four tallest students would not necessarily form a random sample.

4) Use a random-number table or your calculator to generate a list of eight random numbers from 1 to 976. Explain your work.

5) Use a random-number table or your calculator to simulate the outcomes of tossing a quarter 25 times. Assume that the quarter is balanced (i.e., fair).

2.2: Graphs

_____ – a visual representation of a set of data.

Suppose we are given these two scenarios:
1) Over the past couple of decades it has been reported that the following number of children were born in New York City: In 1940, there were 25,000, in 1950, there were 25,600, in 1960, there were 38,000, in 1970, there were 45,000, in 1980, there were 32,600, in 1990, there were 42,320, and in 2000, there were 13,978.

2)

| Decade	Children
1940	25,000
1950	25,600
1960	38,000
1970	45,000
1980	32,600
1990	42,320
2000	13,978

Which of the above was easier to interpret? Why???

Bar Graphs:

1) Bars can be _____ or vertical, but they should be a uniform width and uniformly spaced.

2) The length of a bar represents the _____ we wish to compare under various conditions.

3) When interpreting a bar graph be aware of _____.

4) Both axes are _____ and have scale markings.

Changing Scale

Whenever you use a change in scale in a graphic, warn the viewer by using a squiggle ∿ on the changed axis. Sometimes, if a single bar is unusually long, the bar length is compressed with a squiggle in the bar itself.

Life Expectancy
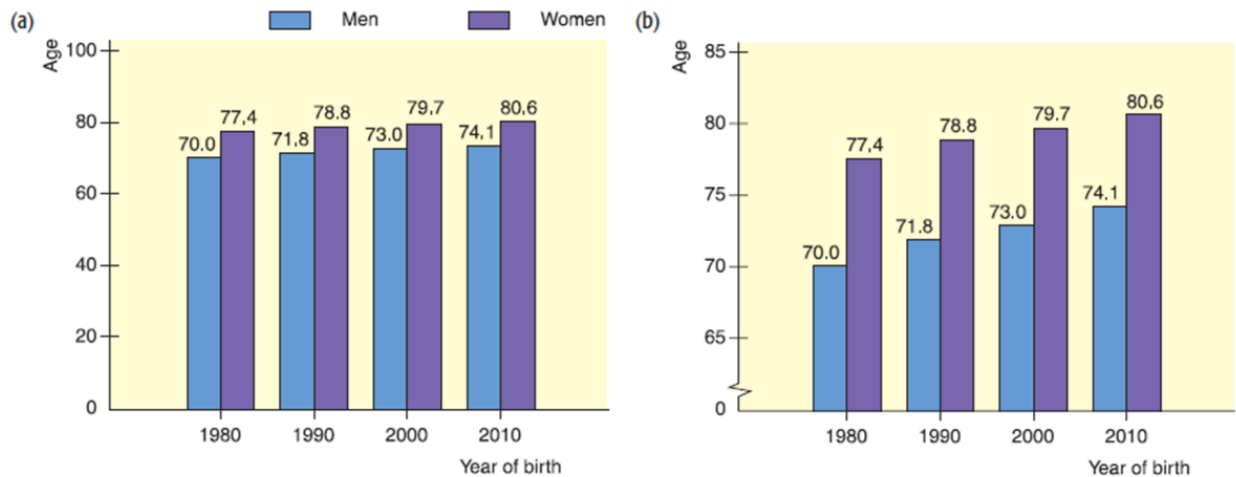
Source: U.S. Census Bureau

Recently published was a list of the World's Longest Suspension Bridges with the following: Humber (England) 1,400 meters, Verrazano-Narrows (NYC) 1,300 meters, Golden Gate (San Francisco) 1,280 meters, Mackinac Straits (Michigan) 1,150 meters, Minami Bisan-Seto (Japan) 1,100 meters, Bosporus (Turkey) 1,050 meters, and George Washington (NYC) 1,025 meters.

To Construct a Bar Graph:
First construct a table to help you organize the data

1) We now need to construct two axes, horizontal and vertical. We will label the vertical axis, length (in meters), and the horizontal axis with the bridge.
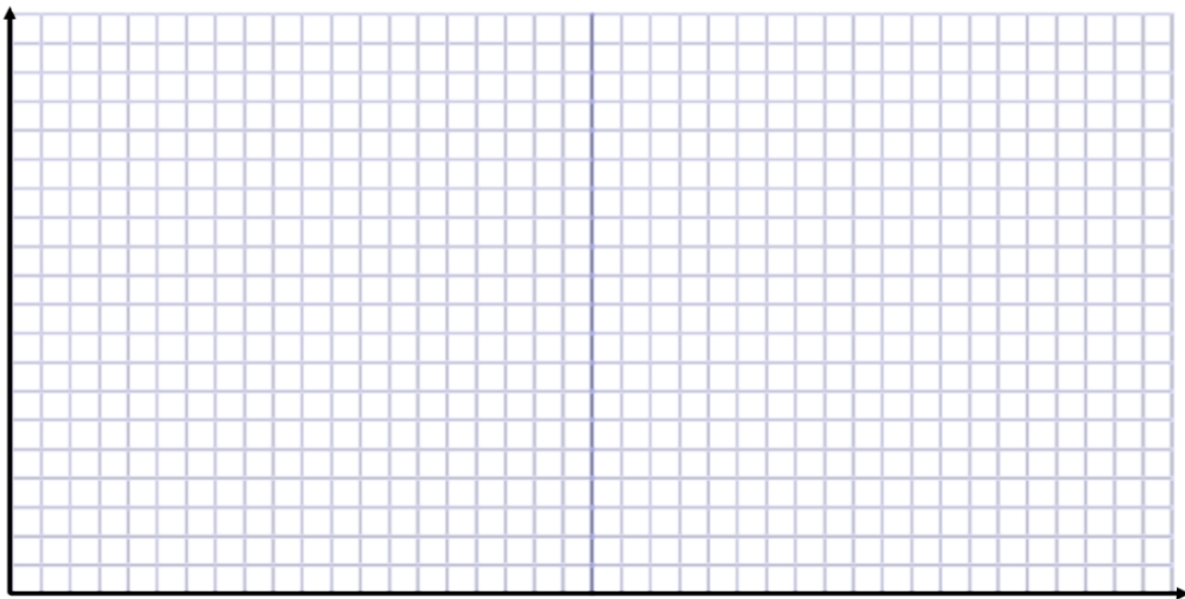
2) Include _____ markings on each axis.
3) Each bar will be _____ over its bridge, and the height of the bar will represent the length (in meters).
4) Make sure the bars have the _____ width and the spacing between all the bars is the _____.
5) At the _____ of each bar, put the actual length.

Make Bar Graphs for the Following

1) The heights of five dams are: Boulder, 730 ft., Kensico, 310 ft., Shasta, 600 ft., Grand Coulee, 450 ft., and Gatun, 120 ft. Construct a vertical bar graph to display the data.
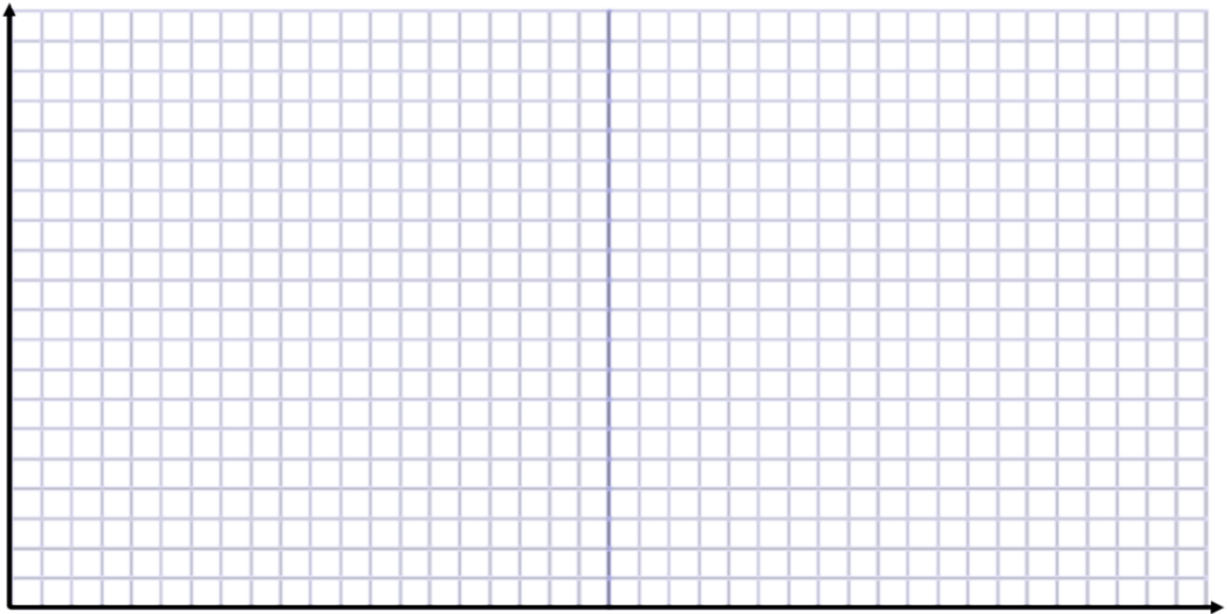
2) Sunshine Travel Agency offers a rain insurance policy on their Hawaii tours. It costs an extra $100. If you buy the optional policy and it rains during more than 15% of the days of your trip, you will be reimbursed for meals and lodging during the extra rainy days (beyond 15% and up to 5 days). You are planning a trip to Hawaii and you are debating about taking the insurance. You obtain the following rainfall information from the U.S. National Oceanic and Atmospheric Administration.

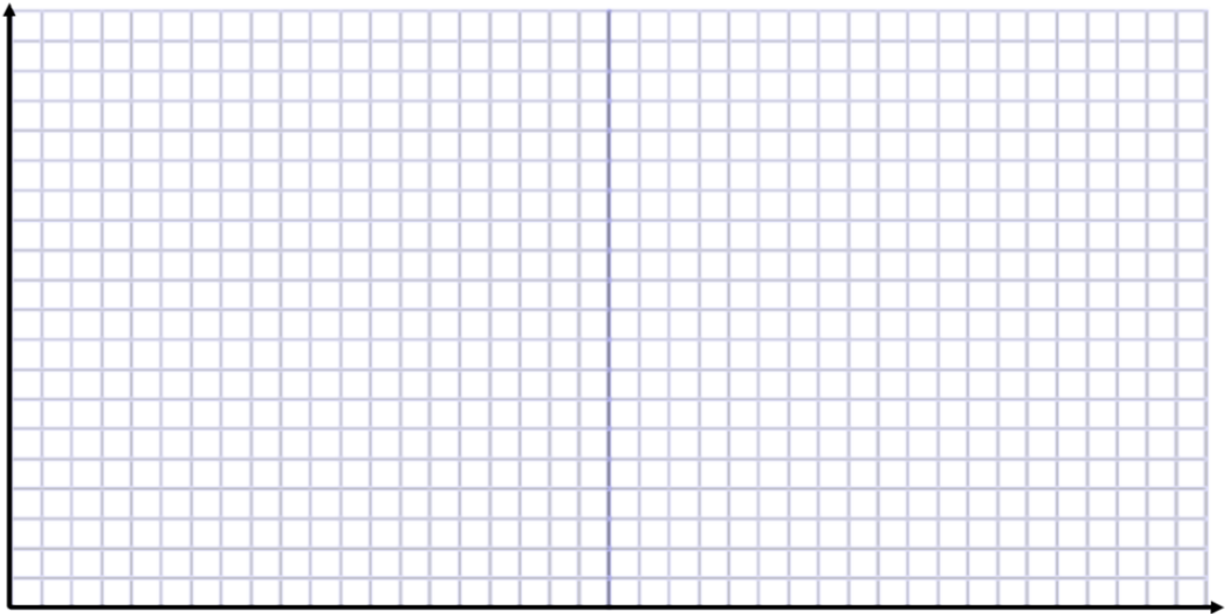| Month	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
Rainfall (inches)	4.40	2.46	3.18	1.36	0.96	0.32	0.60	0.76	0.67	1.51	2.99	3.64

a) Make a bar graph of this information with month on the horizontal axis and rainfall on the vertical axis.
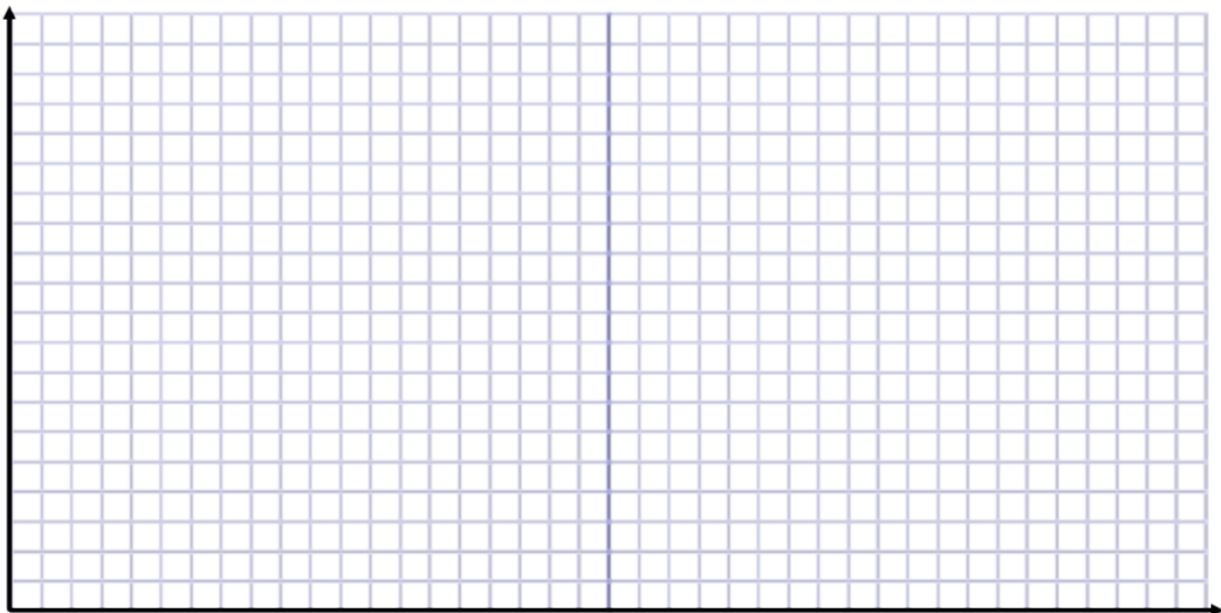
b) There is the rainy season and there is the dry season. From the graph, which 6 months would you say make up the rainy season? The dry season?

c) Without the rain insurance, which winter month (Nov., Dec., Jan., or Feb.) would be best for your trip?

3) The capacity at some National League ballparks is as follows, AT&T Park, 41,600, Shea Stadium, 57,333, Minute Maid Park, 40, 950, Petco Park, 42, 500, Chase Field, 49,033, Busch Stadium, 46,700. Construct a horizontal bar graph with the stadium on the vertical axis and the capacity on the horizontal axis.

Pareto Charts:

1) Listed as one of the most helpful charts.
2) _____ bar graphs in which the height represents frequency.
3) The bars are arranged according to _____, with the tallest bar placed on the _____.
4) Often used to organize data about causes of problems so as to highlight major causes from _____.

Suppose you want to arrive at your college class 15 minutes before your first class so that you can feel relaxed when you walk into class. An early arrival time also allows room for unexpected delays. However, you always find yourself arriving "just in time" or slightly late. What causes you to be late? One student made a list of possible causes and then kept a checklist for 2 months. On some days more than one item was checked because several events occurred that caused the student to be late. (Refer to the table below.) Make a Pareto chart showing the causes for lateness.

| Cause	Frequency
Snoozing after alarm goes off	15
Car trouble	5
Breakfast stop	13
Last-minute studying	20
Finding something to wear	8
Talking too long w/ roommate	9
Other	3

To Construct the Pareto Chart:

1) Construct a table to display the _____ of each given event. (This has been done already above!)

2) From your table, you will now have an idea of the order the events should be placed. Remember: Most frequent events are placed on the _____.
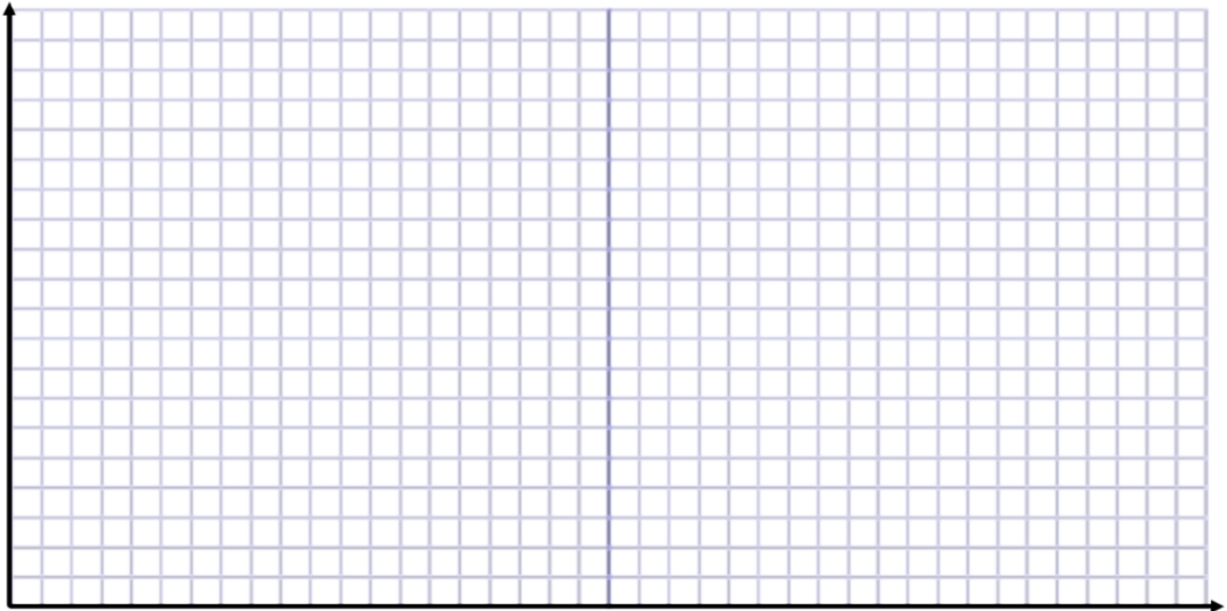
3) _____ each axis.

4) Plot a bar for each event. The height of the bar should equal the _____ from the table constructed in step 1.

1) You recently inherited $10,000 and would like to apply it to some of your outstanding bills. Here is what you owe: Home Improvement Loan Balance, $1,956; Visa Bill, $2,007; MasterCard Bill, $1,983; St. Francis Donation, $2,000; Balance of Car Loan, $1,971; and School Tuition, $2,030. Construct a Pareto Chart to show your outstanding debts.
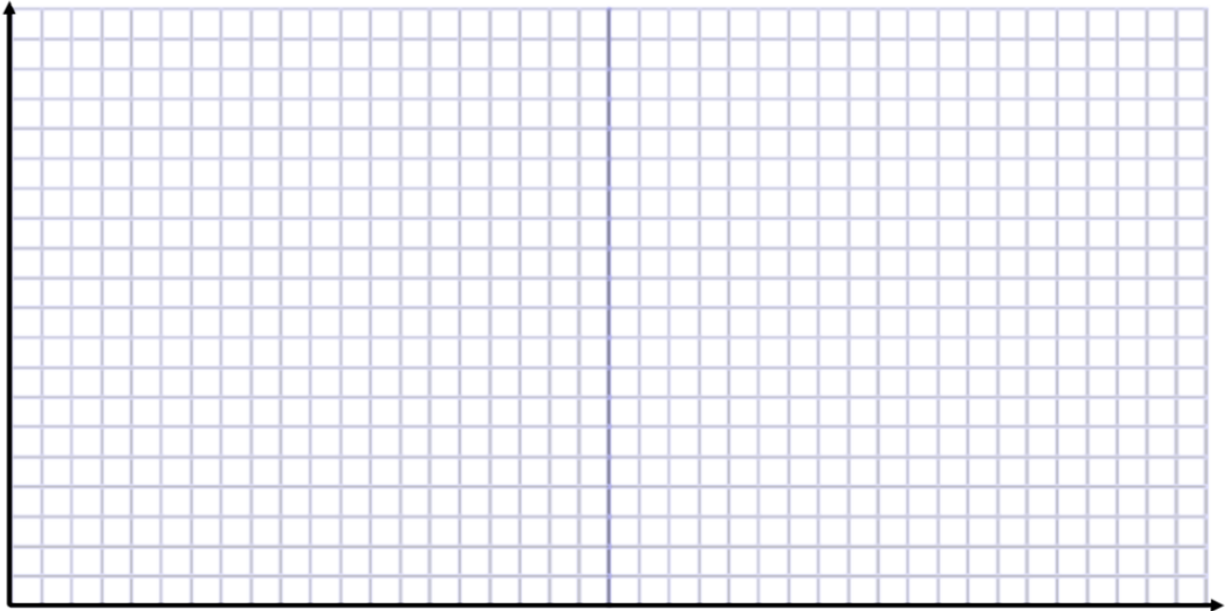
2) You have just been given a new job as the superintendent of a building in NYC. You conduct a survey of the tenants to see where you can make improvements to the overall building. The results came back as follows:

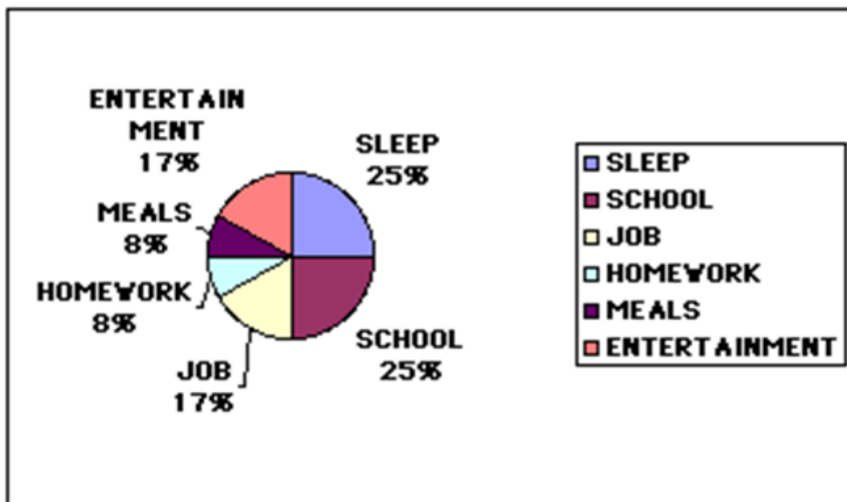| Complaint	Frequency
Loud Music	30
Insufficient Hot Water	23
Inadequate Lighting	175
Poor Quality TV Reception	10
Worn-out Furniture	1
Insufficient Storage Space	25
Cockroaches	324
Rooms Too Warm or Too Cold	300
Garbage Disposal	12

Construct a Pareto Chart to aid you in finding where the major problems are and come up with solutions on how maybe you can get rid or reduce the frequency of these problems.
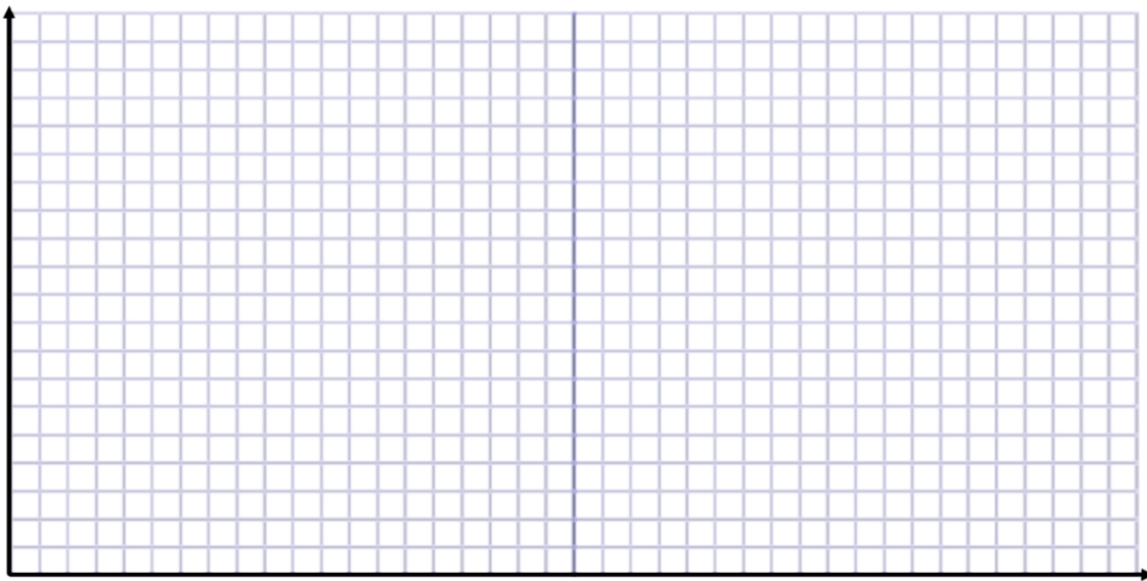
Circle Graphs (Pie Chart):

1) Safe from _____ and is especially useful for showing the division of a total quantity into its component parts.

2) _____ is represented by the entire circle.

3) Each wedge of the circle represents a component part of the total, and the wedge is labeled with corresponding _____ of the total.

Percent of Hours of a Day Spent on Activities

ENTERTAIN MENT 17%

SLEEP 25%

MEALS 8%

HOMEWORK 8%

JOB 17%

SCHOOL 25%

- SLEEP
- SCHOOL
- JOB
- HOMEWORK
- MEALS
- ENTERTAINMENT

Notre Dame does not have on-campus housing. However, there is a special housing service to help out-of-town students find approved living quarters. The current list includes 720 rental units divided up as follows:

Room, no kitchen	360
Efficiency apartment	130
One-bedroom apartment	120
Two-bedroom apartment	90
Three-bedroom apartment	20

Before we can construct a circle graph we need to make a table with multiple columns. First, we will put down the data and the frequency:

Housing	Frequency	Fractional Part	Percent	# of Degrees
Room, no kitch.	360			
Efficiency	130			
1 Bedroom	120			
2 Bedroom	90			
3 Bedroom	20			
Total	720			

Now, we have to make the actual circle graph. We should have a _____ and _____ each of the wedges that we place in our circle.
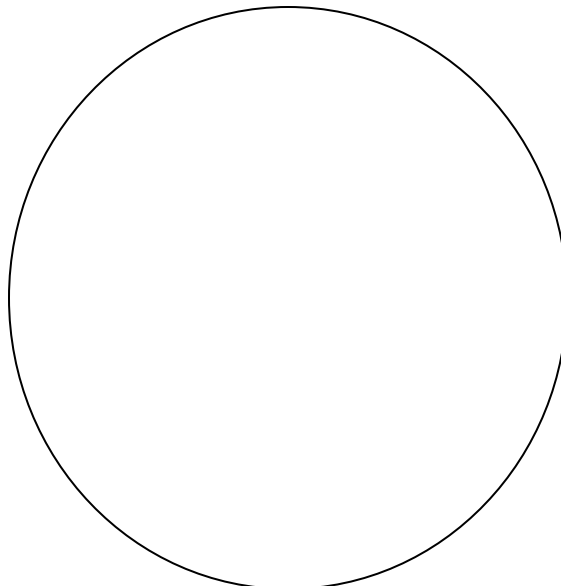
1. What type of information is being presented on this graph?

2. If the total spending is $50,000, how much money was spent on highways?

3. Approximately how many times the amount of spending on highways is spent on education?

4. Approximately what fraction of the total expenditures are spent on highways and public welfare combined?

Expenditures for State and Local Governments

Public Welfare 14%

Highways 7%

Other 45%

Education 34%
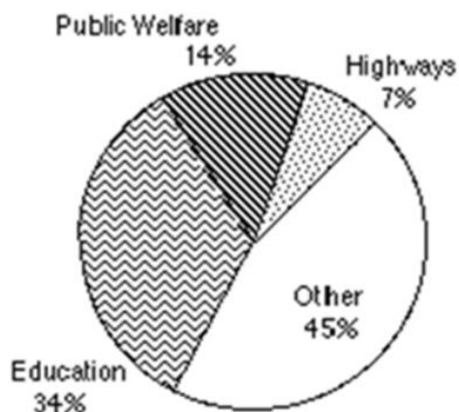
2) Use the circle graph below to answer the following questions about the types of automobiles sold at a dealership this year.

Vehicles Sold at a Dealership (Number per Type)
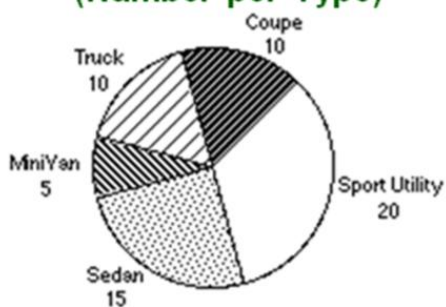
A) What percent of the sales were Sport Utility vehicles?

B) What percent of the sales were Minivans?

C) What is the ratio of the number of Sedans sold to the number of Minivans sold?

3) Given the table below, create an accurate circle graph which represents the same information.

Shares of Stock Owned by an Investor

Type of Stock	Number of Shares
Coca-Cola	8
Pepsi	10
IBM	4
Exxon	8
General Motors	20
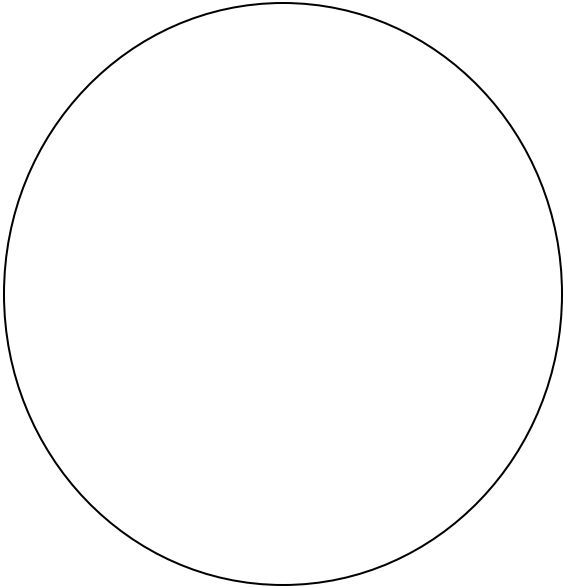
4) Given the table shown below containing information about local revenues in 1992, create an accurate circle graph which contains the same information.

Local Tax Revenues, 1992	
Source of Revenue	Total
Property Taxes	76
Sales and Excise Taxes	15
Personal and Corporate Income Taxes	6
Licenses and Other	4
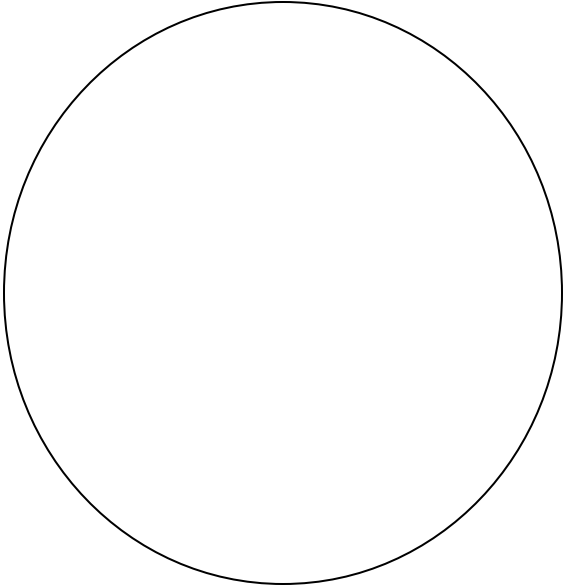
Time Plot:

1) Graph showing the data measurements in _____ order.

2) Put _____ on the horizontal scale and the variable being measured on the vertical scale.
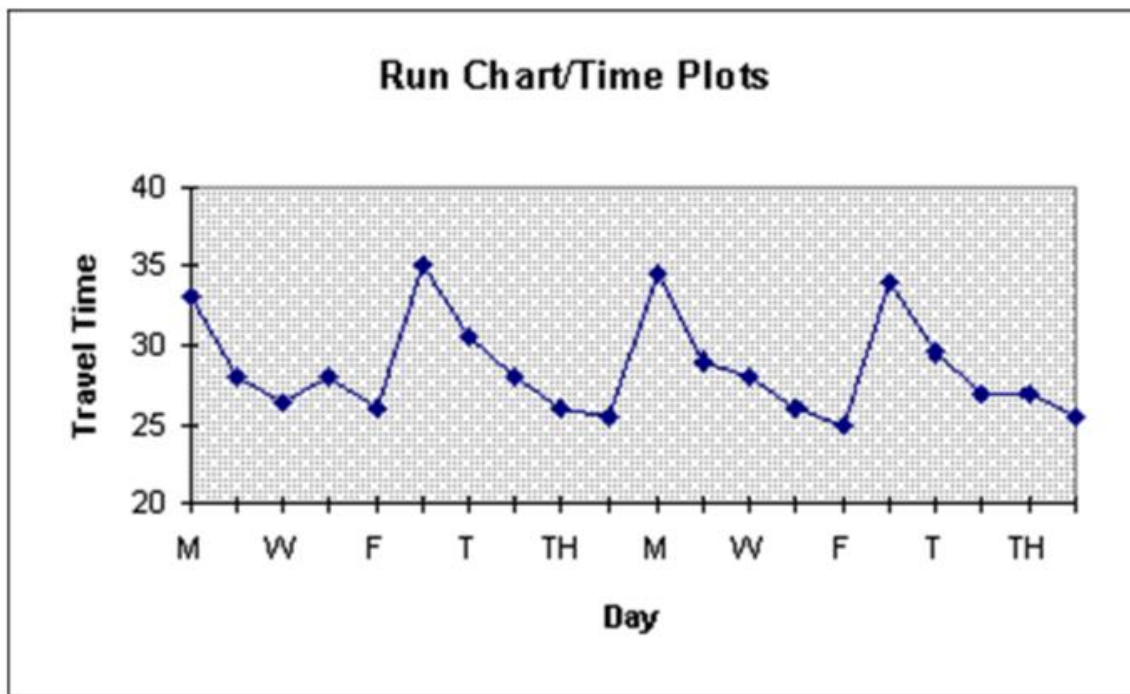
3) _____ the data points by lines.

Run Chart/Time Plots

Time Series:

1) Data sets composed of _____ measurements taken at regular intervals over time.

2) Often used in economics, finance, sociology, medicine, and any situation where we want to study or monitor a similar measure over _____.

3) A _____ can reveal some of the main features of a time series.

To Construct a Time Plot:

1) Put _____ on the horizontal scale and the variable being measured on the vertical scale.

2) _____ the data points by lines.

Example: The following chart contains the price of a bottle of soda over the past 10 years:

Year	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Cost	$1.00	$1.00	$1.10	$1.15	$1.25	$1.25	$1.30	$1.40	$1.45	$1.50
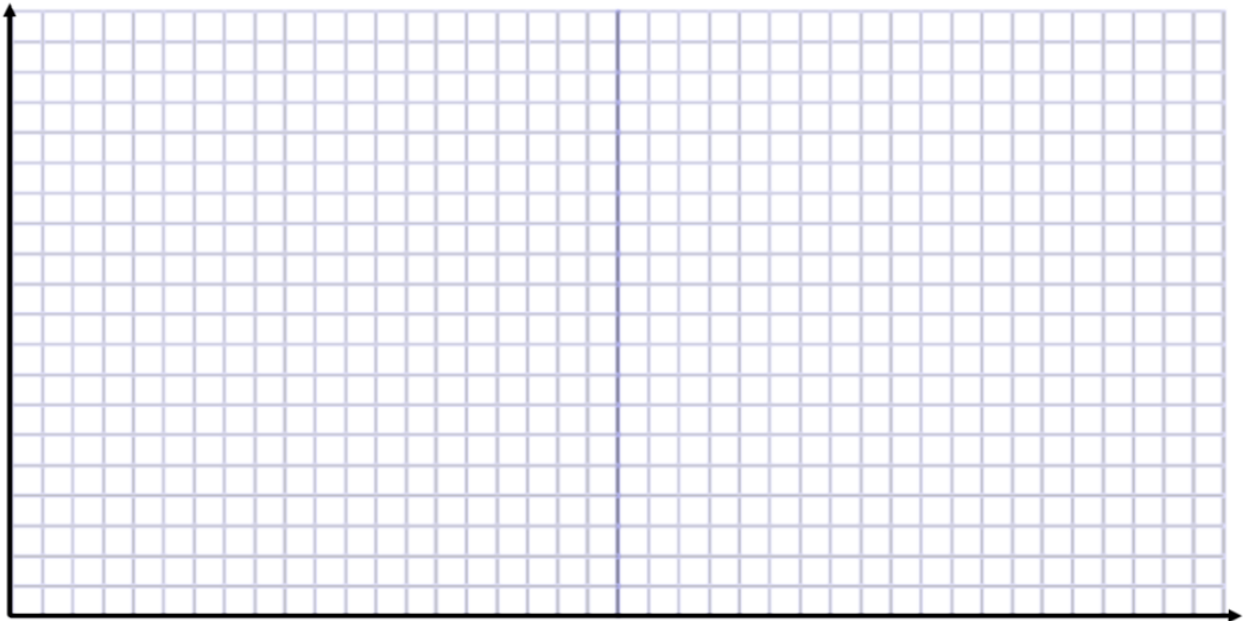
Two researchers measured the pH (a scale on which a value of 7 is neutral and values below 7 are acidic) of water collected from rain or snow over a 26 month period in Allegheny County, Pennsylvania. The first value listed is January 1984 and it proceeds in order thereafter.

4.57	5.62	4.12	5.29	4.64	4.31	4.30
4.39	4.45	5.67	4.39	4.52	4.26	4.26
4.40	5.78	4.73	4.56	5.08	4.41	4.12
	5.51	4.82	4.63	4.29	4.60	

Construct a time plot to display the data.
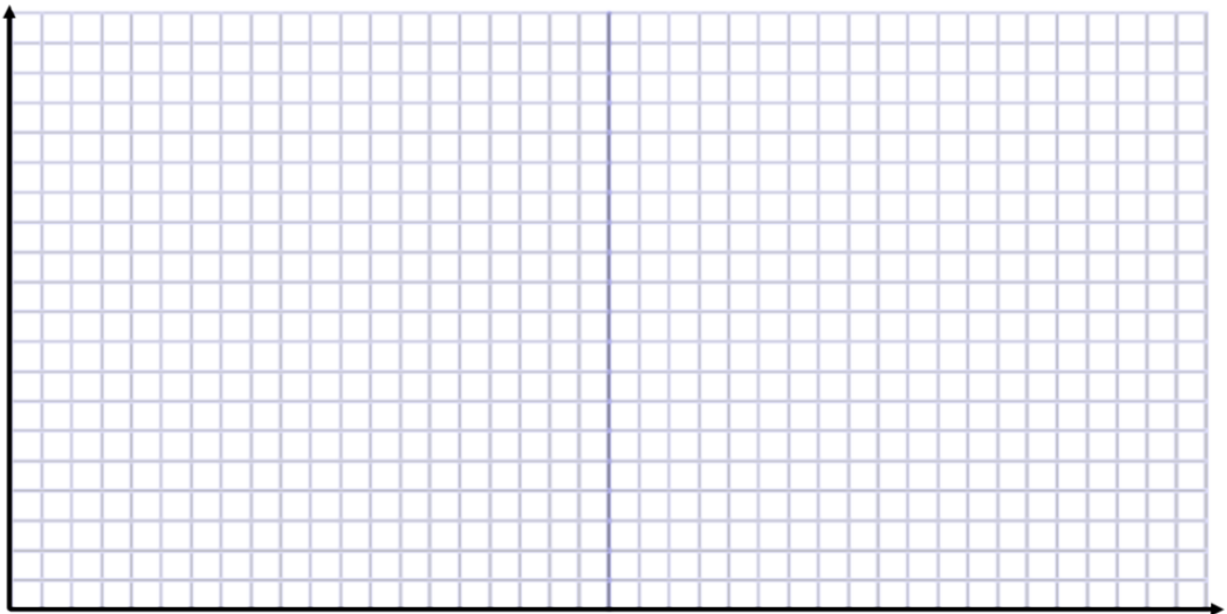
You have just moved into a new area that you are not familiar with. Your desire is to arrive at work on time, but you have just noticed over your first couple of weeks on the job that it doesn't take the same amount of time each day of the week. You decide to monitor the amount of time it takes (in minutes) to get to work over the next four weeks. Construct a time plot based on the data below.

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	33	28	26.5	28	26
Week 2	35	30.5	28	26	25.5
Week 3	34.5	29	28	26	25
Week 4	34	29.5	27	27	25.5
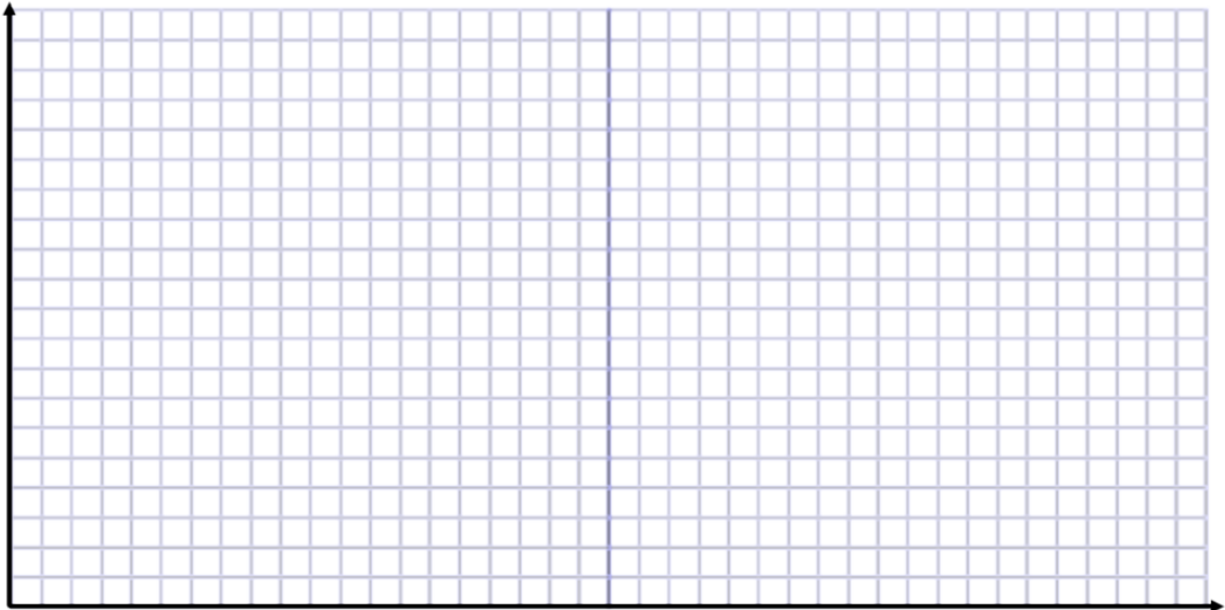
2.2: Homework

1) It's not an easy life, but it's a good life! Suppose you decide to take the summer off and sign on as a deck hand for a commercial fishing boat in Alaska that specializes in deep-water fishing for groundfish. What kind of fish can you expect to catch? One way to answer this question is to examine government reports on groundfish caught in the Gulf of Alaska. The following list indicates the types of fish caught annually in thousands of metric tons (Source: *Report on the Status of U.S. Living Marine Resources,* National Oceanic and Atmospheric Administration): flatfish, 36.3; Pacific cod, 68.6; sablefish, 16.0; Walleye pollock, 71.2; rockfish, 18.9. Make a Pareto chart showing the annual harvest for commercial fishing in the Gulf of Alaska.
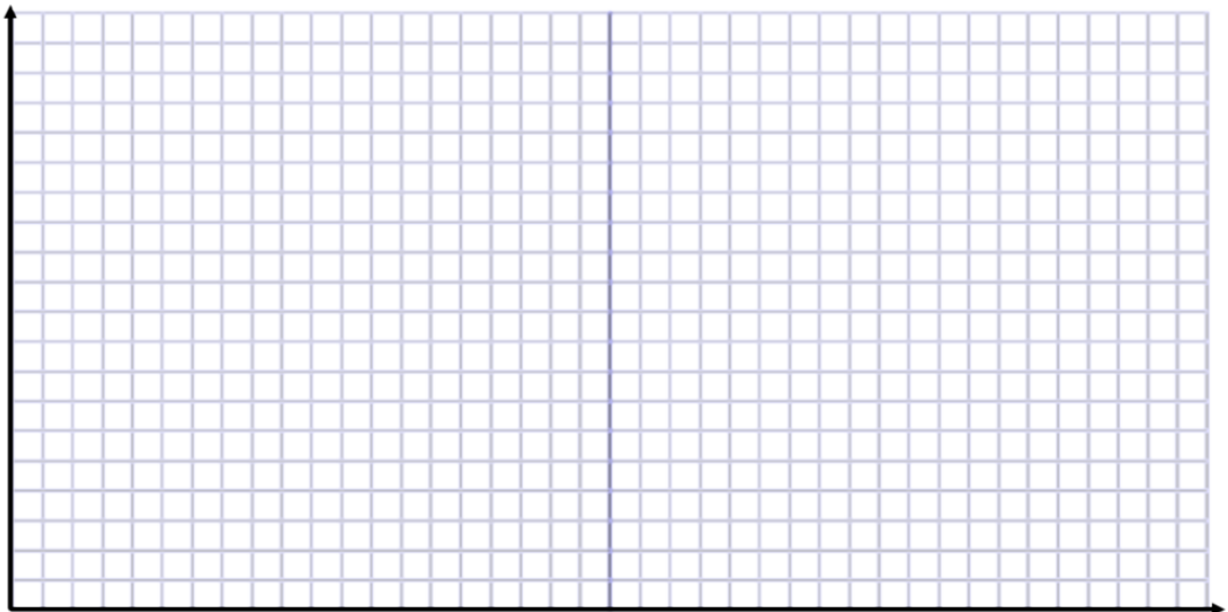
2) It is costly in both time and money to go to college. Does it pay off? According to the Bureau of Census, the answer is yes. The average annual income (in thousands of dollars) of a *household* headed by a person with the stated education level is as follows: 16.1 if ninth grade is the highest level achieved, 34.3 for high school graduates, 48.6 for those holding associate degrees, 62.1 for those with bachelor's degrees, 71.0 for those with master's degrees, and 84.1 for those with doctoral degrees. Make a bar graph showing household income for each education level.
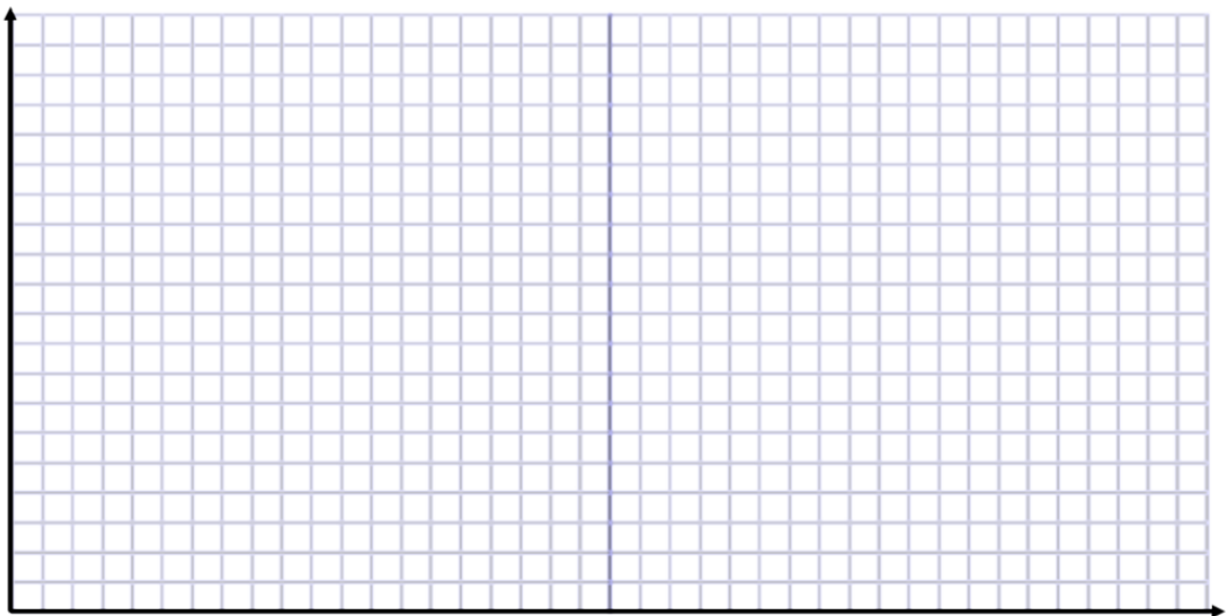
3) A survey of 1000 adults (reported in *USA Today*) uncovered some interesting housekeeping secrets. When unexpected company comes, where do we hide the mess? The survey showed that 68% of the respondents toss their mess in the closet, 23% shove things under the bed, 6% put things in the bathtub, and 3% put the mess in the freezer. Make a circle graph to display this information.
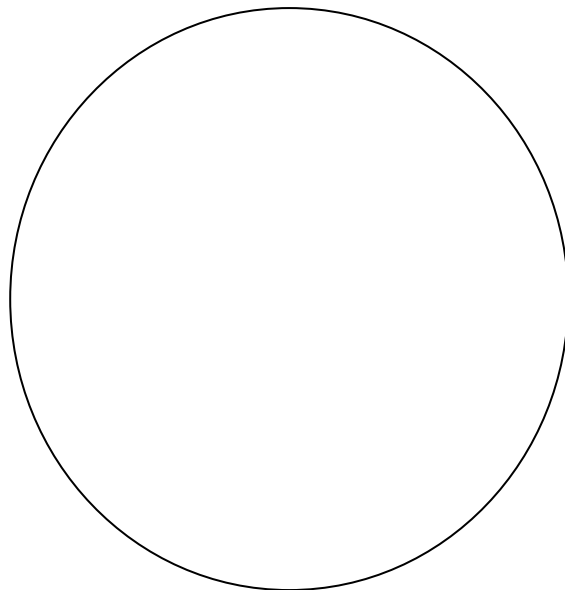
4) Pyramid Lake, Nevada, is described as the pride of the Paiute Indian Nation. It is a beautiful desert lake famous for very large trout. The elevation of the lake surface (feet above sea level) varies according to the annual flow of the Truckee River from Lake Tahoe. The U.S. Geological Survey provided the following data:

Year	Elevation	Year	Elevation	Year	Elevation
1986	3817	1992	3798	1998	3811
1987	3815	1993	3797	1999	3816
1988	3810	1994	3795	2000	3817
1989	3812	1995	3797		
1990	3808	1996	3802		
1991	3803	1997	3807		

Make a time-series graph displaying the data.
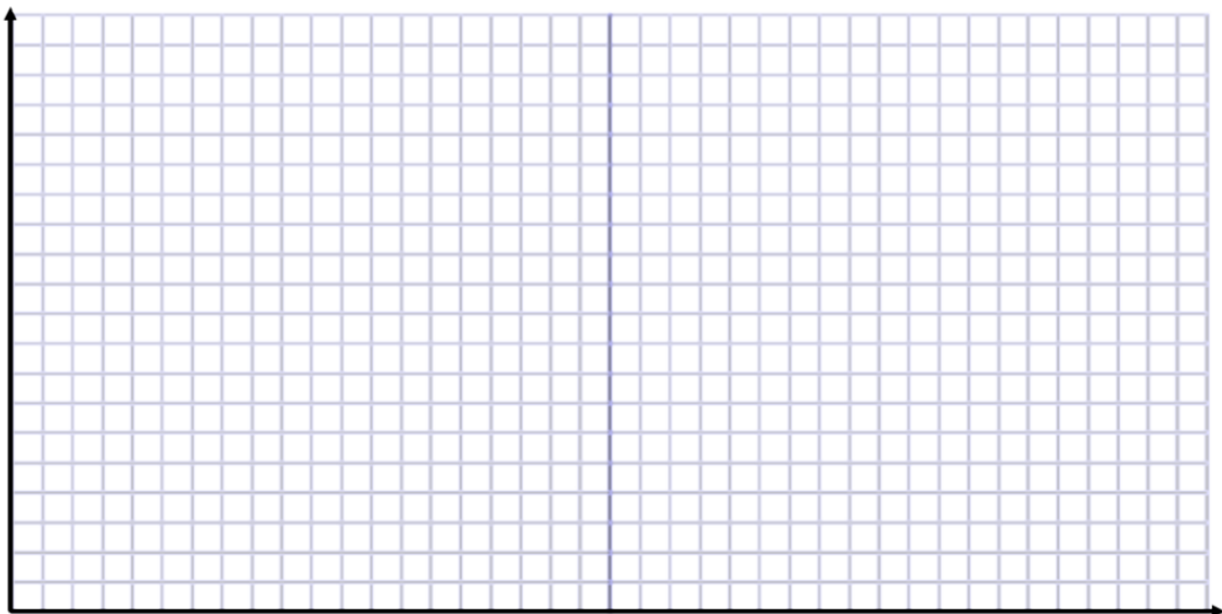
2.3: Histograms and Frequency Distributions

Histogram – a _____ where the bars always touch and the _____ of a bar represents a quantitative value.
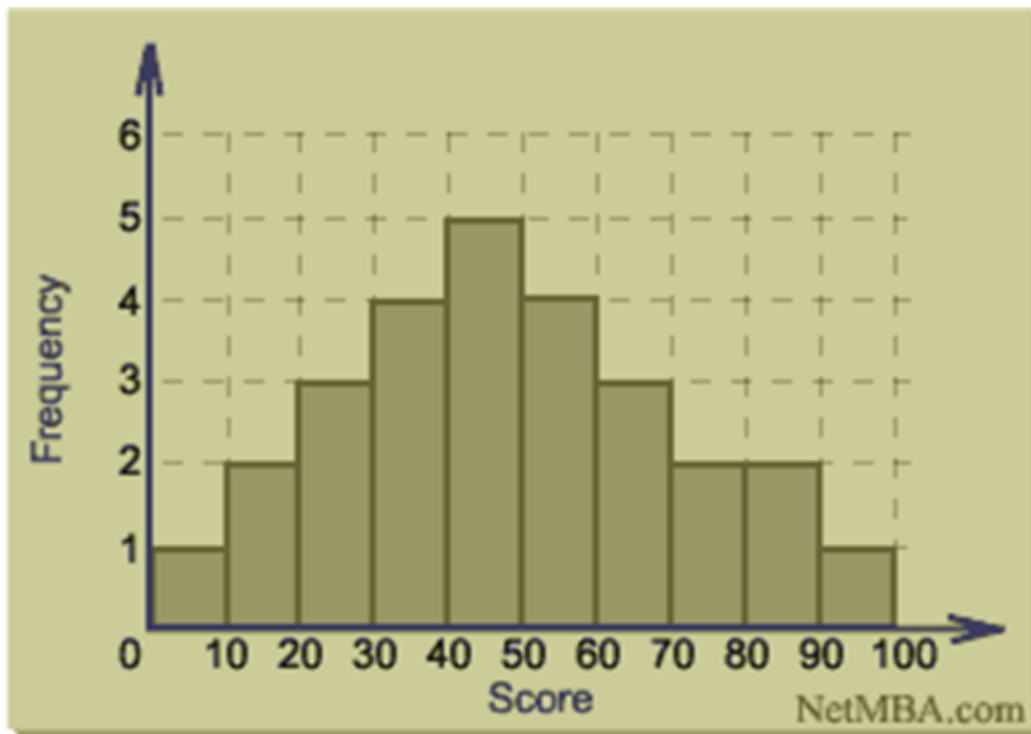
To Construct a Histogram:
1) Decide how many bars or _____ you want. (Usually 5 to 15)
2) Find a convenient _____. (Each class should have the same width; however, usually the first or last class width is a slight bit longer or shorter!)
3) Organize the data into a _____.
4) Since we want the bars to touch, put in _____.

Class Width:

*Round the value computed to the next highest whole number.
Lower Class Limit – lowest value that fits in a class.
Upper Class Limit – highest value that fits in a class.

Midpoint:

*The midpoint is often used as a representative value of the entire class.
Frequency Table – shows the limits of each class, the _____ with which the data fall in a class, and the _____.

To Construct a Frequency Table:
1) Determine the _____.
2) Create the distinct classes. We use the convention that the lower class limit of the first class is the _____ data value. Add the class width to this number to get the lower class limit of the next class.
3) _____ the data into classes. Each data value should fall exactly into one class. Total the tallies to obtain each class frequency.
4) Compute _____ (class mark) for each class.
5) Determine the _____.

Class Boundaries:

1) Assures us that the bars of the histogram _____ and that no data fall on the boundaries.

2) To calculate, take the lower class limit and _____ .5, take the upper class limit and _____ .5

* Sometimes a histogram using class boundaries looks silly. Instead, some use lower class limits as labels, with the premise that a data value falling on the class limit is included in the next higher class or the class to the right of the limit.

The following example will help us to see how easily a histogram can be constructed:

Bridget Phillips is the star of the school's basketball team. The numbers of points scored by Bridget in her last 20 games are: 36, 32, 28, 30, 33, 36, 24, 33, 29, 30, 30, 25, 34, 36, 34, 31, 36, 29, 30, and 34.

First, we need to decide how many bars we would like in our histogram. If we analyze the numbers, we see that 5 will be sufficient.

Second, we need to come up with the class width. We use the formula:

The class width tells us that we will have intervals of 3 numbers. We can now begin constructing the frequency table and filling in the information:

Frequency Table of Bridget Phillips Points Scored Over Last 20 Games Played

Interval (Class)	Tally	Frequency	Class Midpoint
35 – 37			
32 – 34			
29 – 31			
26 – 28			
23 - 25			

Sometimes, it is necessary to calculate the _____.
This is used as a representative of the entire class. To do this we use the following formula:

***Calculate the midpoint for each class and fill in the above table.

Now, since we want our bars to touch we need to compute the class boundaries for each bar. To calculate the class boundaries, we look at each interval. We take the lower class limit and subtract .5. Next, we take the upper class limit and add .5. For example, our lowest interval is 23 – 25. We take 23 and subtract .5 to get 22.5, and we take 25 and add .5 to get 25.5. These boundaries will go on the horizontal axis of our histogram.
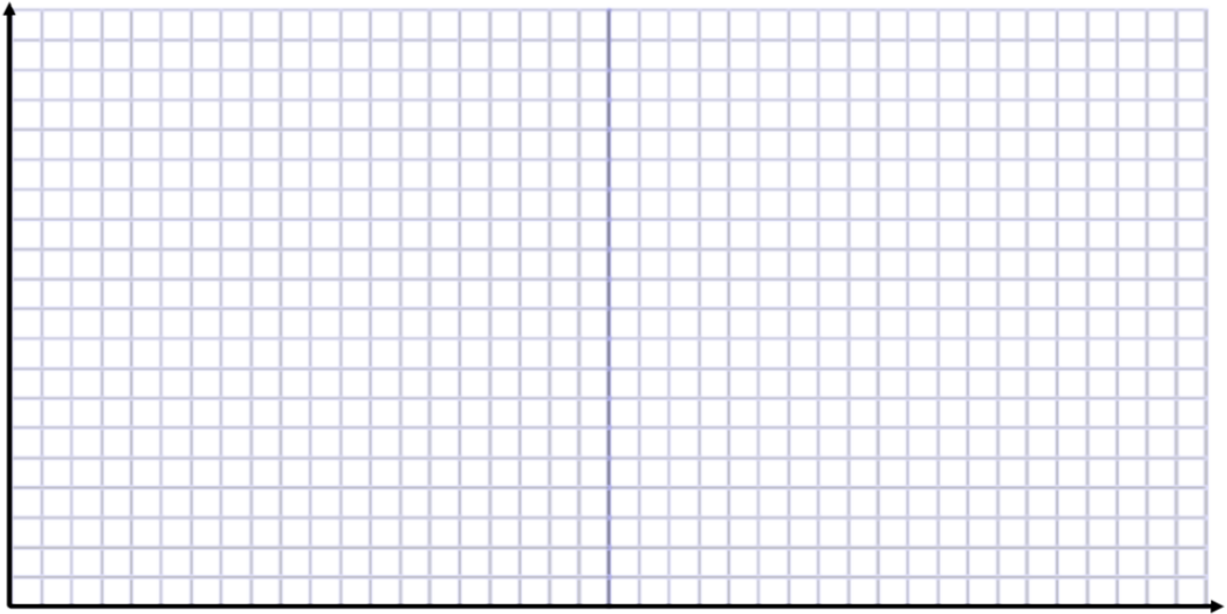
We now have everything we need to construct the histogram! We can also answer the following questions:

A) Which interval contains the greatest frequency?

B) In how many games did Bridget score 32 or more points?

C) In what percent of these 20 games did Bridget score fewer than 26 points?

Examples

1) The table below represents the number of miles per gallon (mpg) of gasoline obtained by 40 drivers of compact cars in a large city.

Interval	Frequency
16 – 19	5
20 – 23	11
24 – 27	8
28 – 31	5
32 – 35	7
36 – 39	3
40 – 43	1

Construct a frequency table and frequency histogram based on the data. Then, answer the following questions.
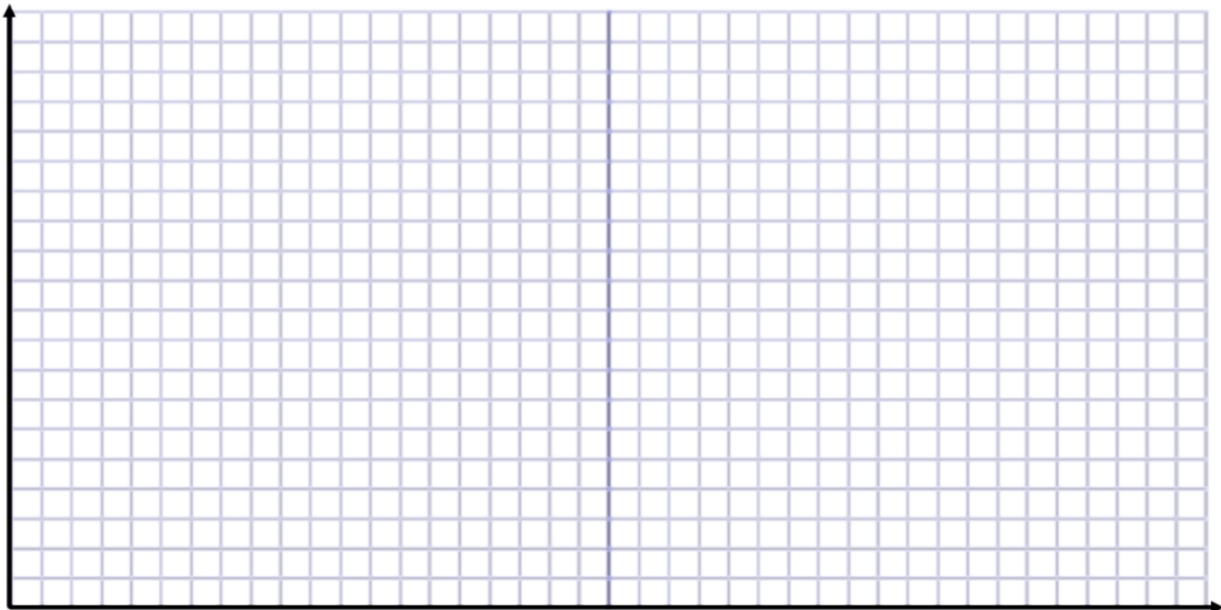
A) In what interval is the greatest frequency found?

B) What is the number (or frequency) of cars reporting mileage between 28 and 31 mpg?

C) In what interval are the fewest cars reported?

D) How many of the cars reported mileages greater than 31 mpg?

E) What percent of the cars reported mileages from 24 to 27 mpg?

Thirty students on the track team were timed in the 200-meter dash. Each student's time was recorded to the nearest tenth of a second. The times were: 29.3, 31.2, 37.6, 30.9, 26.0, 32.4, 31.8, 36.6, 35.0, 38.0, 37.0, 22.8, 35.2, 35.8, 37.7, 38.1, 34.0, 34.1, 28.8, 29.6, 26.9, 36.9, 39.6, 29.9, 30.0, 36.0, 36.0, 36.1, 38.2, and 37.8. Complete the table below and construct a frequency histogram for the given data.

Interval	Tally	Frequency	Midpoint
37.0 – 40.9			
33.0 – 36.9			
29.0 – 32.9			
25.0 – 28.9			
21.0 – 24.9			
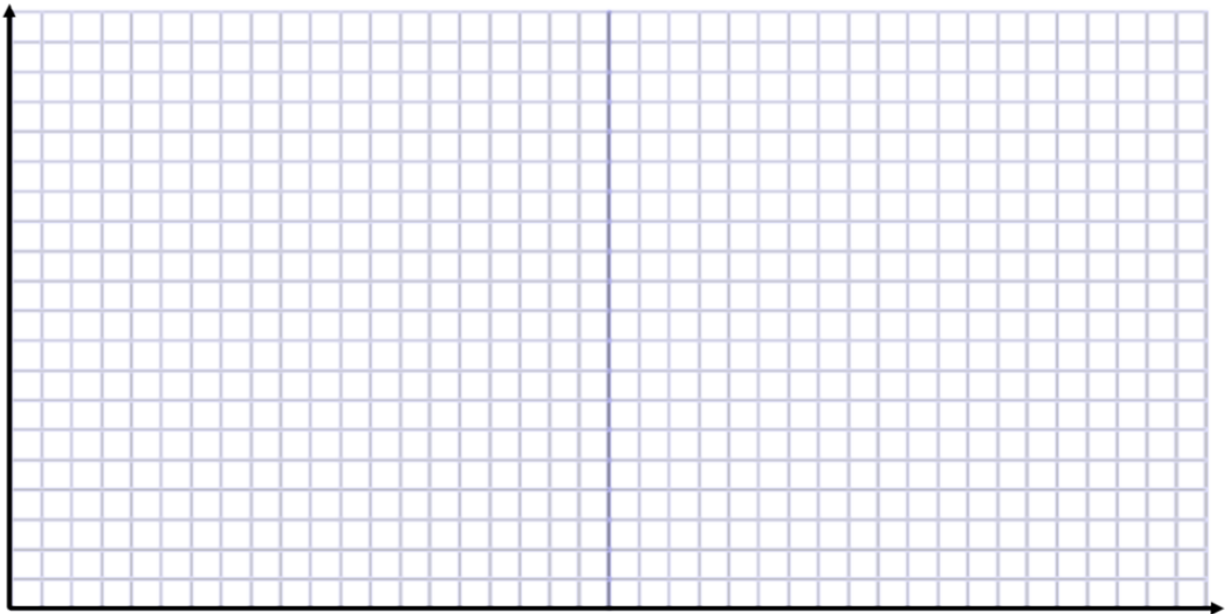
Complete the following table and construct a histogram using the data:

Interval	Frequency	Midpoint
1 – 3	24	
4 – 6	30	
7 – 9	28	
10 – 12	41	
13 – 15	19	
16 - 18	8	
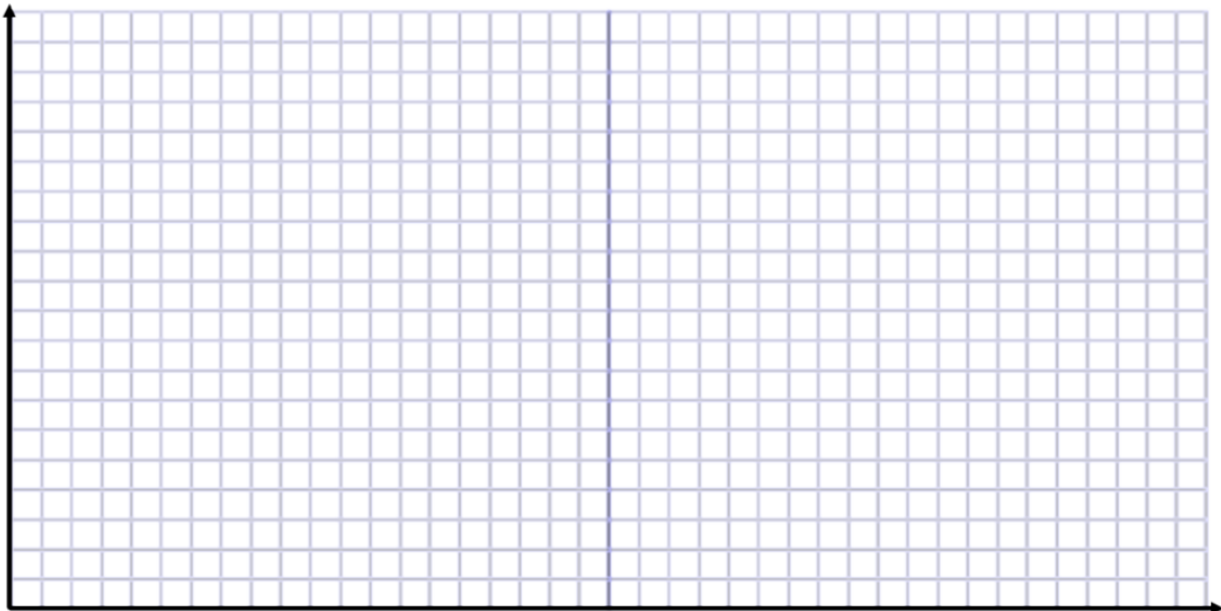
Calculator Instructions

We can also construct histograms on the calculator. Let's try the following example using the calculator:

The number of houses sold each month by two realtors, Pete and Mildred, is shown below for the last 36 months.

Pete:

4	9	5	4	9	5	2	0	8	4	8	10
6	7	13	8	2	10	0	12	8	4	6	8
6	3	3	6	7	13	9	11	3	4	0	8

Mildred:

6	9	0	1	2	3	6	3	9	8	7	5
5	3	8	1	0	5	0	4	8	9	5	7
4	1	1	6	2	3	1	8	5	3	13	5

Let's complete the frequency table and histogram for Pete.

1) First, we should figure out the class sizes using what we have learned. We know we want five classes, with our largest data value being 13 and our smallest data value being 0. Therefore, we will have a class width of 3. We should now begin constructing our frequency table.

2) Hit STAT and choose #1 EDIT. This will bring you to a screen that looks like a table.

3) Under L1 we will enter all of the individual number of houses sold for Pete, making sure we have entered 36 values.

4) Now we will hit STAT PLOT which is 2ND Y =. Select #1. Turn this ON and under TYPE: choose the figure that looks like a histogram.

5) Now we will hit ZOOM 9 for the statistics window. Unfortunately, this gives us 7 classes for our data when all we wanted was 5. This means that we will have to change our window setting. Hit WINDOW and you will see a list that looks like this and we will fill in the appropriate information:

Xmin = (lowest class boundary after rounding) = 0
Xmax = (highest class boundary after rounding) = 15
Xscl = (class width) = 3
Ymin = (lowest value we would like to see on our axis) = 0
Ymax = (highest value we would like to see on our axis) = 15
Yscl = 1
Xres = 1

6) Select GRAPH to view the histogram.
7) Select TRACE to see the frequency of each bar.
8) Copy and complete all parts of the question.

Do now: Complete the histogram and frequency table for Mildred.
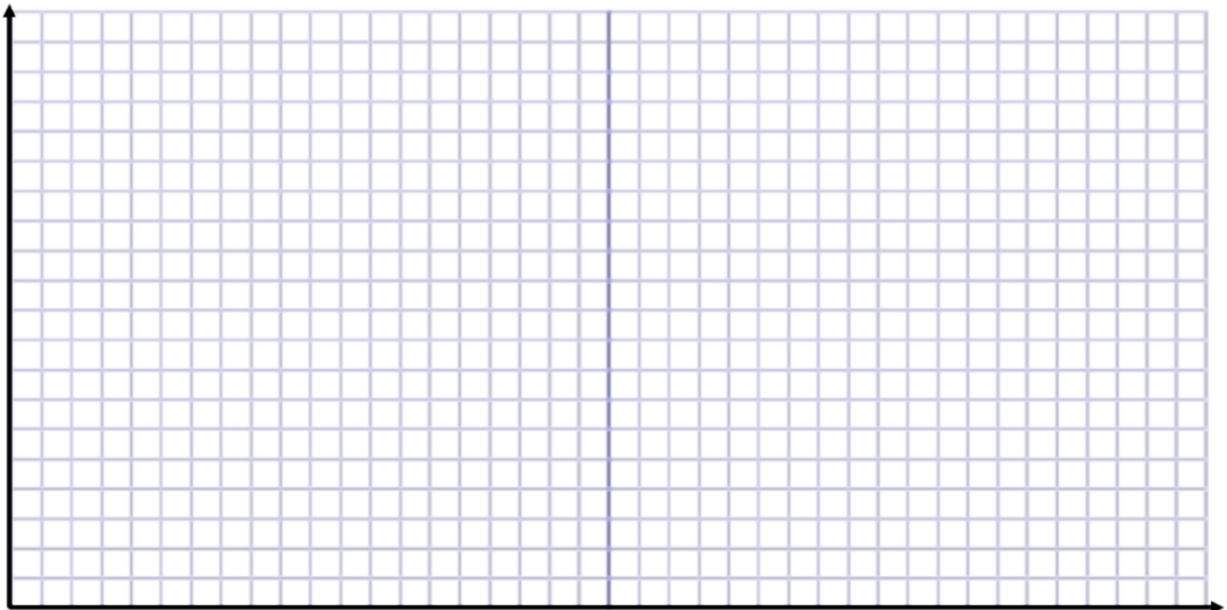
Frequency Table – if you know how to construct a frequency table, then just divide the class frequency by the total of all frequencies (sample size).

Relative Frequency

*** The total of the relative frequencies should be 1. (Unless you have rounded your answers!)

In a health care utilization study, a random sample of 25 patients having total care insurance at a designated clinic was studied. The number of times each patient went to the clinic during a one-year period was recorded. The results follow:

15	7	3	13	1	22	12	4	3	8
12	15	11	19	3	2	5	3	4	7
2	13	6	5	4					

    A) Construct a relative frequency table using five classes.
    B) Construct a relative frequency histogram using the same classes.

Solution: Since we already have an understanding of the frequency table, we should know how to come up with the 5 classes.

Health Care Utilization Study of 25 Patients

Class Boundary	Tally	Frequency	Relative Frequency
.5 – 5.5			
5.5 – 10.5			
10.5 – 15.5			
15.5 – 20.5			
20.5 – 25.5			

$$\Sigma =$$

Under the frequency column, there is a total. We will take the individual frequency for each class and divide it by the total frequency, which in this case will be 25. For the first class, we tally equals 12. We now take 12 and divide it by 25 to get .48.

***Hint: If take a _____ of the relative frequency column, you should always get approximately _____.

Now, we will construct the relative frequency histogram. The horizontal column would remain the same as if we were constructing a regular histogram. The vertical axis will be marked with relative frequencies instead of the actual frequencies.

Examples:

Home Video Incorporated rents out new releases. A random sample of 32 days gave the following information about the number of new releases rented each day at one of the local stores:

12	27	29	4	21	38	21	5
15	25	23	10	8	22	13	17
43	8	22	23	26	28	31	35
26	34	36	33	32	39	37	29

A) Construct a relative frequency table using 5 classes.

B) Construct a relative frequency histogram.
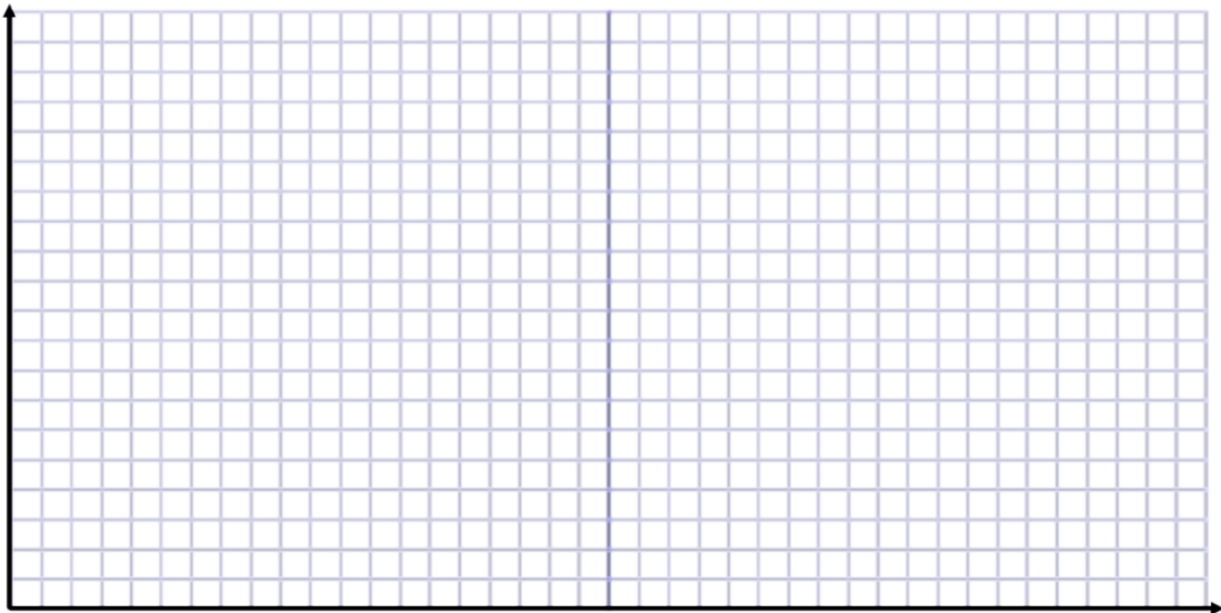
The district manager of a suburban park and recreation area wants to estimate the amount of time people spend in their local park in July. A random sample of 50 local residents gave the following information on the estimated number of hours they spent in the regional park during the month of July:

8	5	9	12	15	18	2	3	5	4
2	8	6	10	13	14	8	10	9	13
21	2	4	3	7	7	6	7	6	16
19	11	3	3	8	4	11	10	9	8
6	6	6	7	8	12	6	7	20	7

A) Construct a relative frequency table using 5 classes.

B) Construct a relative frequency histogram.

Distribution Shapes
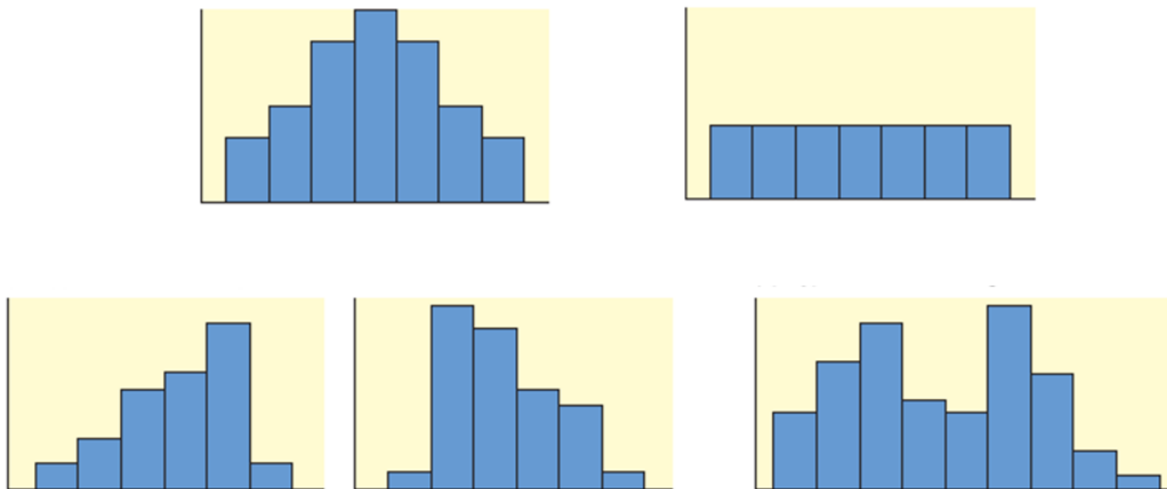
(a) **Mound-shaped symmetrical:** This term refers to a histogram in which both sides are (more or less) the same when the graph is folded vertically down the middle. Figure 2-8(a) shows a typical mound-shaped symmetrical histogram.
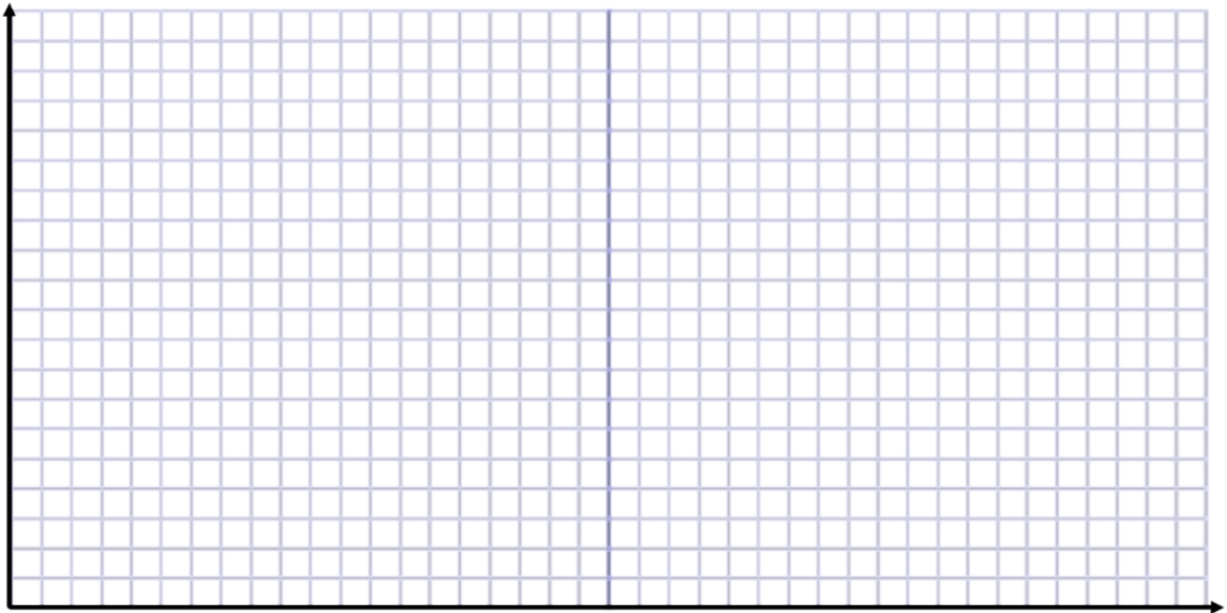
(b) **Uniform or rectangular:** These terms refer to a histogram in which every class has equal frequency. From one point of view, a uniform distribution is symmetrical with the added property that the bars are of the same height. Figure 2-8(b) illustrates a typical histogram with a uniform shape.

(c) **Skewed left or skewed right:** These terms refer to a histogram in which one tail is stretched out longer than the other. The direction of skewness is on the side of the *longer* tail. So, if the longer tail is on the left, we say the histogram is skewed to the left. Figure 2-8(c) shows a typical histogram skewed to the left and another skewed to the right.

(d) **Bimodal:** This term refers to a histogram in which the two classes with the largest frequencies are separated by at least one class. The top two frequencies of these classes may have slightly different values. This type of situation sometimes indicates that we are sampling from two different populations. Figure 2-8(d) illustrates a typical histogram with a bimodal shape.

Frequency Polygons:

_____ emphasizes the continuous _____
of the frequencies. It is made by connecting, in order, the top midpoints of the bars
in a histogram. Frequency polygons are especially useful if you wish to compare
two distributions.

To Construct a Frequency Polygon:
1) Construct a histogram. (Optional)
2) Calculate the _____ of each class and place a point on the
tops of each bar where the midpoint would be.
3) Connect, in order, the _____.
4) Extend the lines on the left and on the right so that the polygon begins and ends
with a frequency of _____.
5) Place the extensions one class width to the left and right of the first and last
midpoints, respectively.

 *** Without the histogram, simply plot the class frequency over the class
  midpoint, and then connect the points in order.

A random sample of 40 days gave the following information about the total
number of people treated each day at the Community Hospital emergency room:

40	35	42	6	13	50	60	27
8	42	53	17	25	23	24	12
26	32	28	28	31	29	30	28
21	46	22	19	20	30	31	30
36	30	40	38	30	29	31	41

a) Make a frequency table using five classes.

b) Make a histogram.
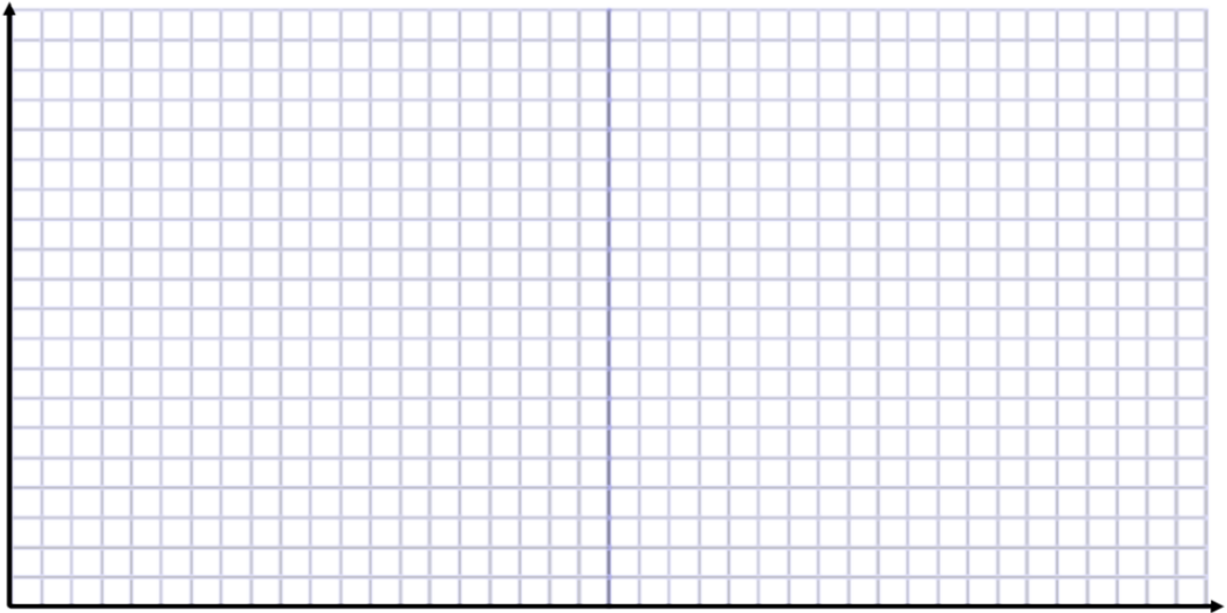
c) Make a frequency polygon.
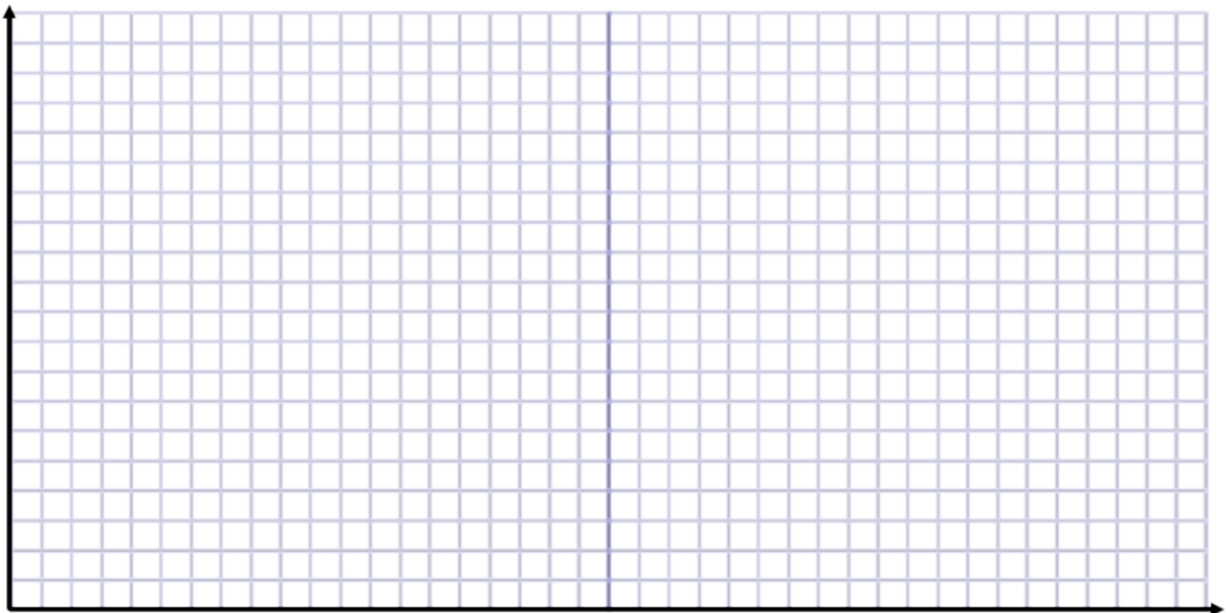
The number of hamburgers sold at McDonald's in each of the 50 states during August of last year was as follows (in units of 1,000 hamburgers):

86	70	38	100	115	135	51
72	53	65	110	136	56	25
42	60	93	23	58	96	108
131	148	150	73	156	82	71
173	200	147	68	92	15	119
110	172	183	66	65	52	97
212	63	88	93	88	95	31
44						

a) Make a frequency table using 5 classes.

b) Make a frequency polygon.
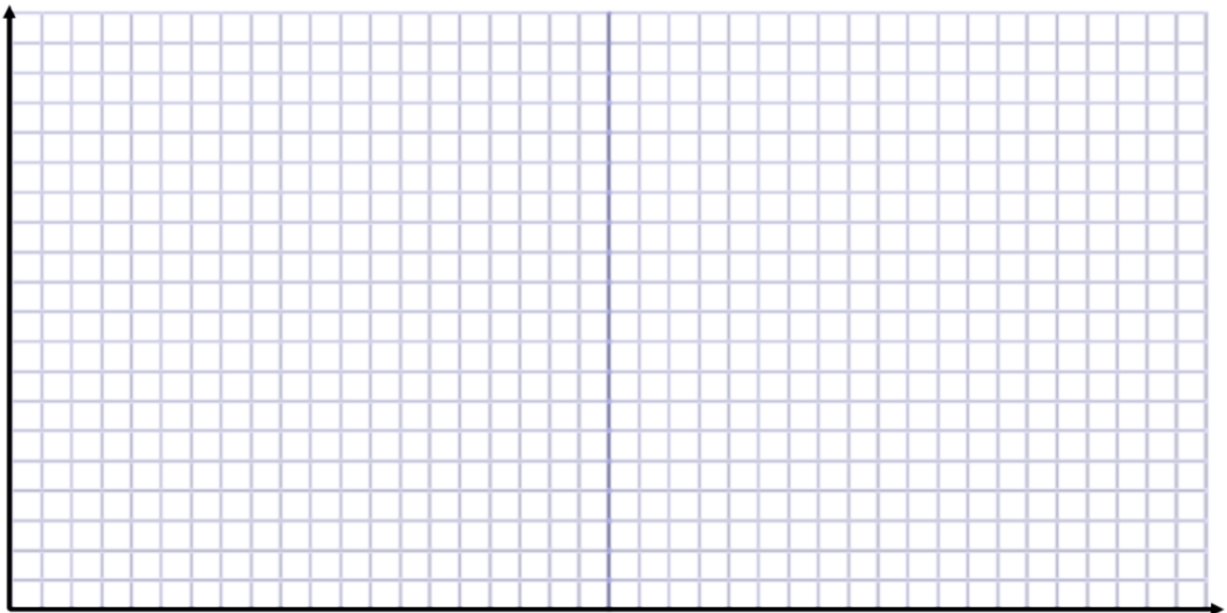
Ogives:

An _____ is a way to graph information showing _____ frequencies. To make an ogive, we first construct a _____. Your **vertical scale** will represent _____ and your **horizontal scale** will represent **upper class boundaries**. The reason that you use upper class boundaries on the horizontal scale is that we cannot be sure that we have all the data in that class until we get to the upper class boundary. Ogives always start at the left with a _____ of zero at the lower class boundary of the first class. The ogive should end on the right with the cumulative frequency equal to the sample size at the upper class boundary of the last class. Heavy dots indicate the cumulative frequency for a class and these dots are connected by straight lines.

To Construct an Ogive:

1) Construct a _____.
2) Your vertical scale represents _____ frequencies.
3) Your horizontal scale represents upper class boundaries.
4) Plot the points and _____ with straight lines.
5) The ogive should start on the left with a _____ at the lower class boundary of the first class.
6) The ogive should end on the right with a cumulative frequency equal to the _____ at the upper class boundary of the last class.

High Temperatures During the Aspen Ski Season (°F)

| Class Boundaries | | Frequency | Cumulative Frequency |
Lower	Upper		
10.5	20.5	23	23
20.5	30.5	43	66 (sum 23 + 43)
30.5	40.5	51	117 (sum 66 + 51)
40.5	50.5	27	144 (sum 117 + 27)
50.5	60.5	7	151 (sum 144 + 7)

(a) The cumulative frequency for a class is computing by adding the frequency of that class to the frequencies of previous classes.

(b) To draw the corresponding ogive, we place a dot at cumulative frequency 0 on the lower class boundary of the first class. Then we place dots over the *upper class boundaries* at the height of the cumulative class frequency for the corresponding class. Finally, we connect the dots.
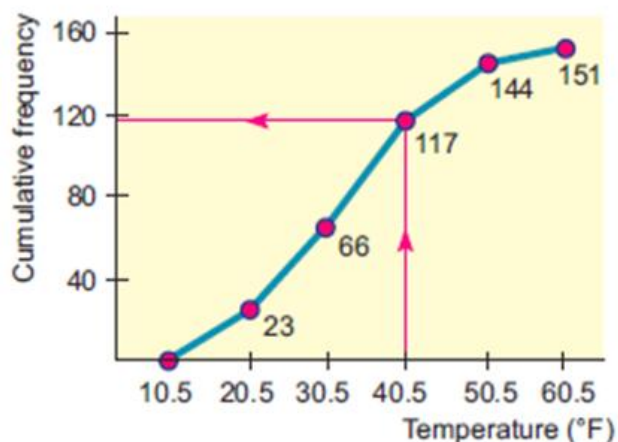
(c) Looking at the ogive, estimate the total number of days with a high temperature lower than or equal to 40°F.

Examples:

On a test consisting of 20 questions, 15 students received the following scores:

17	14	16	18	17
19	15	15	16	13
17	12	18	16	17

1) Construct a cumulative frequency table using 4 classes.

2) Construct an ogive.
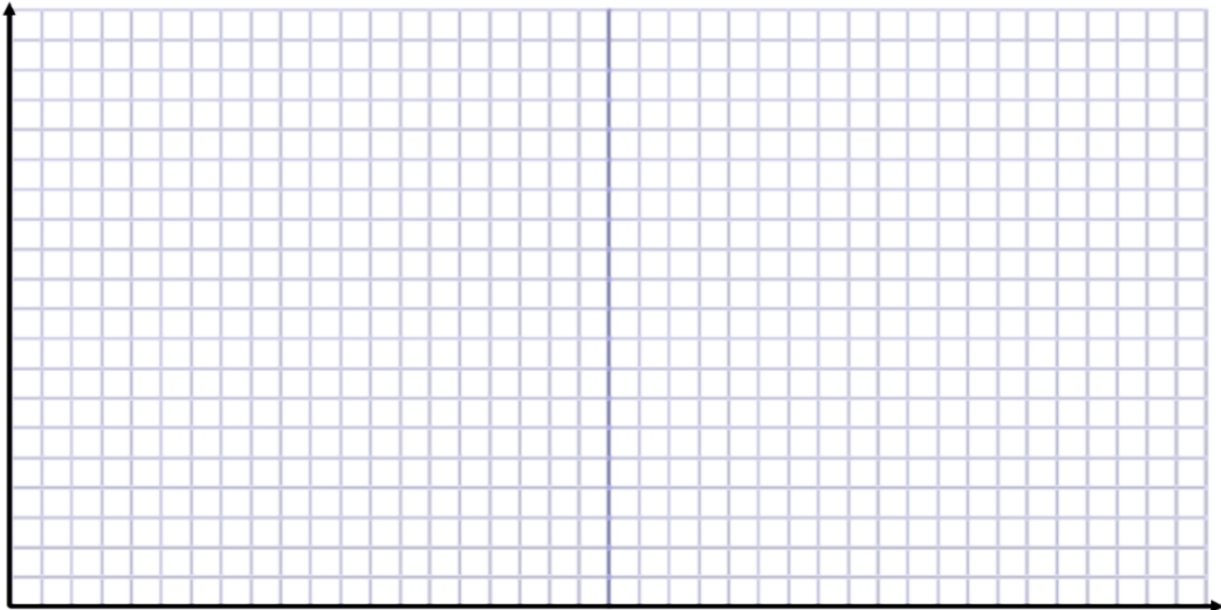
The following data consists of weights, in pounds, of 35 adults:

176	154	161	125	138	142	108
115	187	158	168	162	135	120
134	190	195	117	142	133	138
151	150	168	172	115	148	112
123	137	186	171	166	166	179

1) Construct a cumulative frequency table using 5 classes.
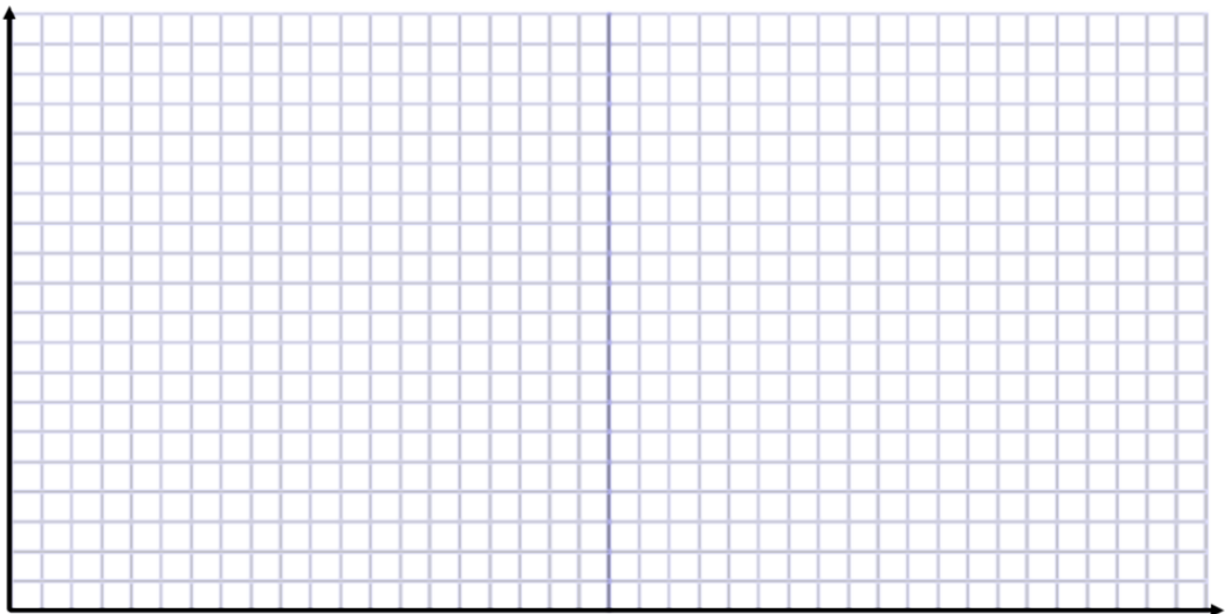
2) Construct an ogive.

Thirty students on the track team were timed in the 200-meter dash. Each student's time was recorded to the nearest tenth of a second.
The times were:

29.3	31.2	28.5	37.6	30.9	26.0
32.4	31.8	36.6	35.0	38.0	37.0
22.8	35.2	35.8	37.7	38.1	34.0
34.1	28.8	29.6	26.9	36.9	39.6
29.9	30.0	36.0	36.1	38.2	37.8

1) Construct a cumulative frequency table using 5 classes.

2) Construct an ogive.
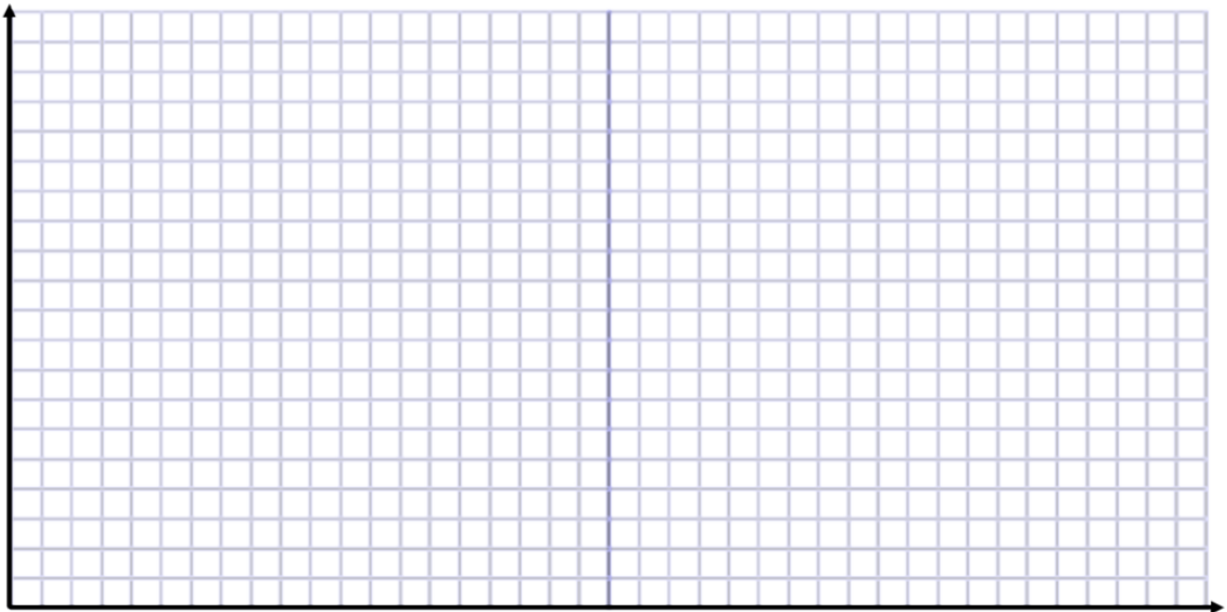
2.3: *Homework*

1) How long does it take to finish the 1161-mile Iditarod Dog Sled Race from Anchorage to Nome, Alaska? Finish times (to the nearest hour) for 57 dogsled teams are shown below.

261	271	236	244	279	296	284	299	288	288	247	256
338	360	341	333	261	266	287	296	313	311	307	307
299	303	277	283	304	305	288	290	288	289	297	299
332	330	309	328	307	328	285	291	295	298	306	315
310	318	318	320	333	321	323	324	327			

Use five classes.
(a) Find the class width.

(b) Make a cumulative frequency table showing class limits, class boundaries, midpoints, frequencies, relative frequencies, and cumulative frequencies.

(c) Draw an ogive.
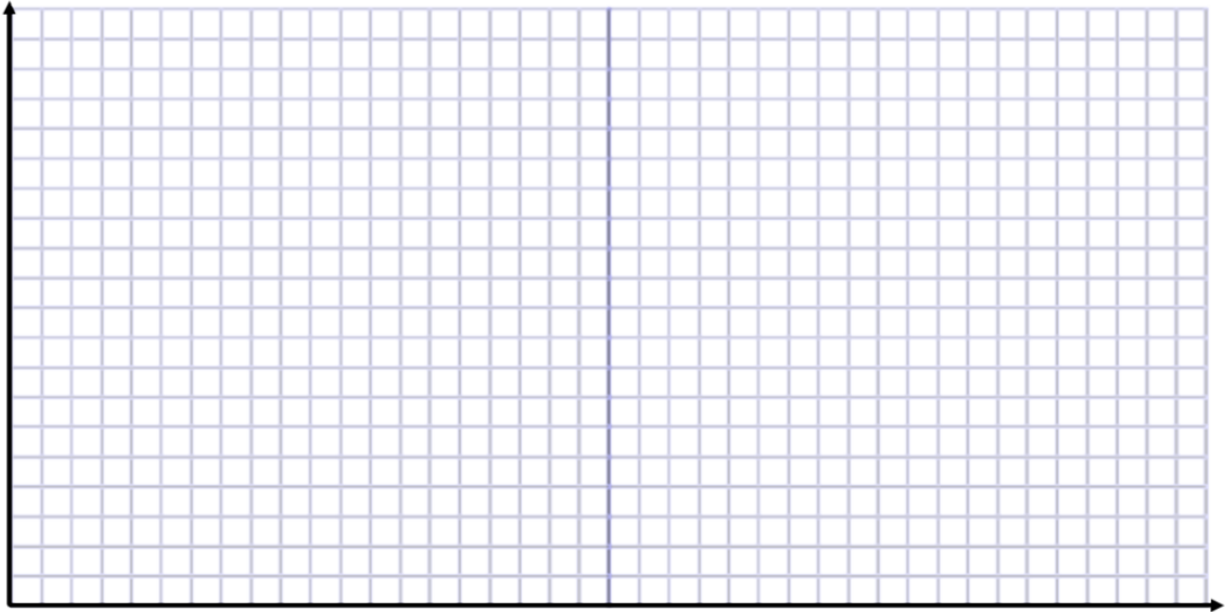
2) Certain kinds of tumors tend to recur. The following data represent the lengths of time, in months, for a tumor to recur after chemotherapy.

19	18	17	1	21	22	54	46	25	49
50	15	9	39	43	39	5	9	38	18
14	45	54	59	46	50	29	12	19	36
38	40	43	41	10	50	41	25	19	39
27	20								

Use five classes.
(a) Find the class width.

(b) Make a cumulative frequency table showing class limits, class boundaries, midpoints, frequencies, relative frequencies, and cumulative frequencies.

(c) Draw a relative frequency histogram.
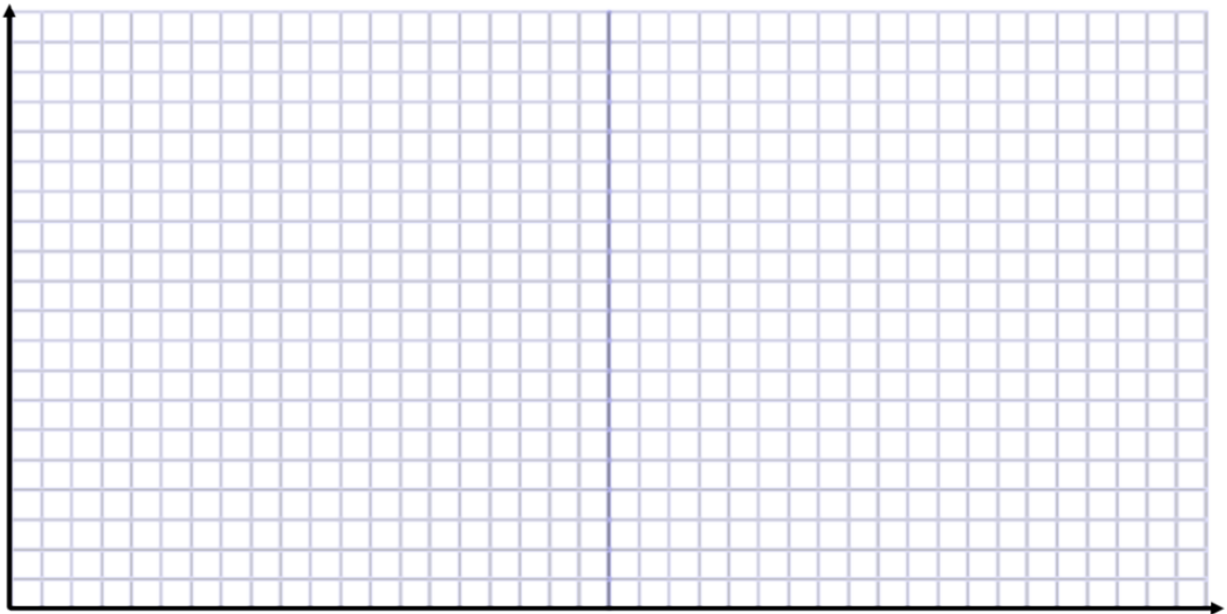
3) The following data represent highway fuel consumption in miles per gallon (mpg) for a random sample of 55 models of passenger cars.

30	27	22	25	24	25	24	15
35	35	33	52	49	10	27	18
20	23	24	25	30	24	24	24
18	20	25	27	24	32	29	27
24	27	26	25	24	28	33	30
13	13	21	28	37	35	32	33
29	31	28	28	25	29	31	

Use five classes.

(a) Find the class width.

(b) Make a cumulative frequency table showing class limits, class boundaries, midpoints, frequencies, relative frequencies, and cumulative frequencies.

(c) Draw a frequency polygon.
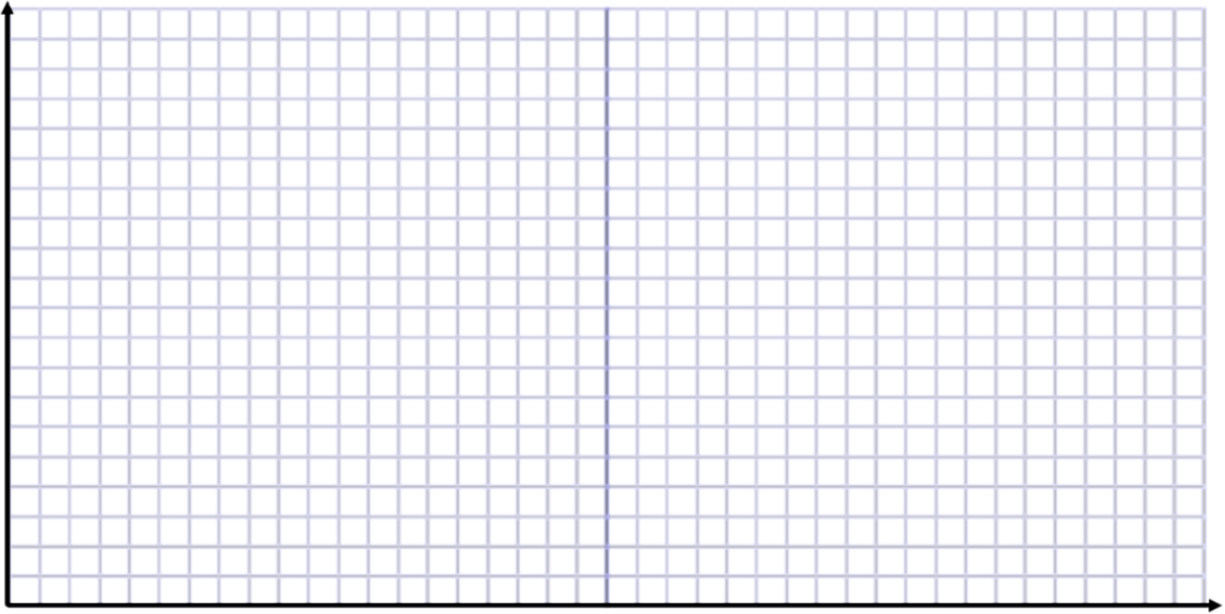
2.4 Stem-and-Leaf Displays

_____ – useful for detecting patterns and extreme data values and are designed to help us explore a data set, to ask questions we had not thought of before, or to pursue leads in many directions.

_____ – device that organizes and groups data but allows us to see many of the digits in each data value as we wish.

To Construct a Stem-and-Leaf Display:
1) Break the digits of each data value into _____ parts.
      a) Left group of digits is called a _____.
      b) The remaining group of digits is called a _____.
2) List each possible stem once on the _____ and all of its leaves in the same row on the _____.
3) Indicate the _____.
4) Give it a title.

*** The lengths of the leaves give the visual impression that a sideways histogram would be present.
*** Decimal points are omitted in the stems and leaves, but indicated in the unit designation as appropriate.

```
   3 | 2    represents 32 lb
Stem | Leaves

   0 | 0 3

   1 | 2 2 7 8 8 8 9

   2 | 1 1 1 2 6 6 7 7 8 9 9

   3 | 0 1 1 2 2 2 2 3 3 5 5 6 8 8

   4 | 1 2 3 5 7

   5 | 1
```
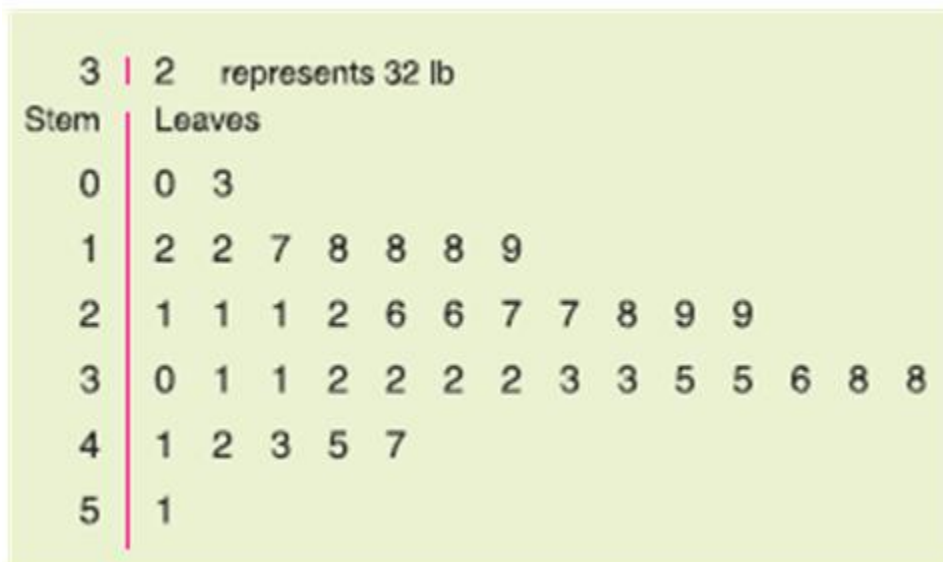
Examples:
On a recent exam, 32 students received the following grades:

56	97	97	96	95	92	90
89	81	80	78	78	76	75
74	73	89	89	86	85	85
84	83	83	72	70	70	68
67	65	62	58			

1) Construct a stem-and-leaf plot to display the data.

The following scores were recorded in an algebra class at St. Francis:

42	45	41	48	46	42	50
52	58	55	60	62	64	63
58	80	78	82	77	69	70
78	65	84	88	89	92	86
95	98	76	81			

1) Construct a stem-and-leaf plot to display the data.

Multiple Lines Per Stem:

We could also construct a stem-and-leaf display which spreads the data out even more. We call this _____. In this case we can say that leaves 0 - 4 would go with one stem and leaves 5 - 9 would go with a second stem. However, if you would like to use a multiple line stem-and-leaf, you must use an _____ for your first stem and a _____ for your second stem.

The following grades were scored on a recent exam:

100	93	71	74	85	56	62
68	70	100	99	85	77	85
48	51	79	25	86	93	88
70	100	26	67			

1) Construct a stem-and-leaf plot with two lines per stem.

The following number of slices was sold each month at Fresh Meadows (over the last 2 years):

308	309	319	298	297	294	316
328	329	334	308	300	311	319
326	324	333	337	308	305	291
290	333	319				

1) Construct a stem-and-leaf plot with two lines per stem.

Back –to-Back Stem Plot:

Many real-life applications of statistics involve _____ of two populations.

Depth (in cm) of Artifact Location

```
              5 | 2 | 0     = 25 cm at Site I and 20 cm at Site II
      Site I              Site II
              5    2   0 5 5
             5 0   3   0 0 0 0 5 5
          5 5 5 5  4   0 0 0 5
             5 0   5   0 0 5 5
        5 5 5 5 5 0  6  0 0 0 5 5 5
    5 5 5 5 5 5 0 0  7
        5 5 0 0 0  8
        5 5 0 0  9
          5 5  10
            0  11  0 0 5 5 5 5
               12  0 0 0 0 5
```
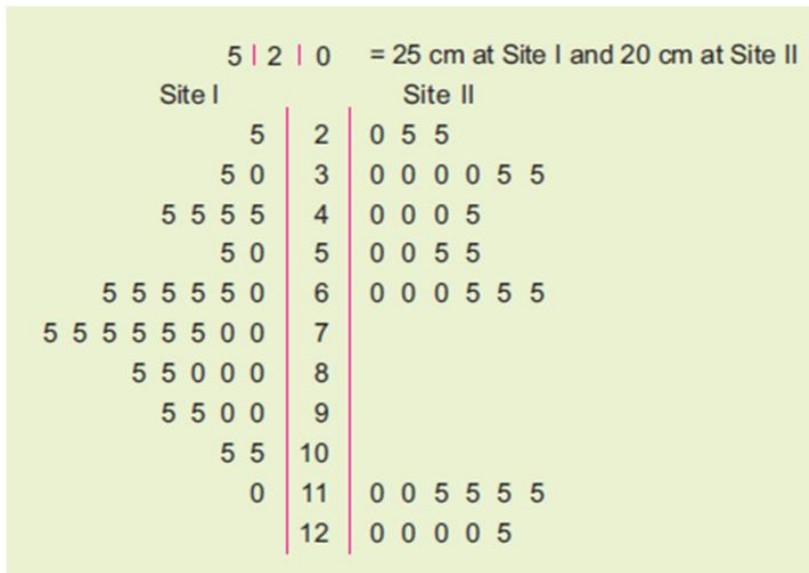
The **caloric intake** of 20 people on each of two weight loss programs are recorded as follows:

Program A:	1184	1186	1189	1188	1188
	1188	1187	1184	1184	1182
	1180	1183	1182	1187	1188
	1188	1184	1184	1187	1186

Program B:	1187	1183	1187	1182	1185
	1187	1189	1184	1187	1184
	1187	1182	1180	1182	1181
	1187	1184	1187	1186	1184

1) Construct a back-to-back stem-and-leaf plot to display the data.

In a 40 year study, **survival years** were measured for cancer patients undergoing one of two different chemotherapy treatments. The data for 25 patients on the first drug and 30 on the second were as follows:

Drug A:
5	10	17	39	29	25	20	4	8	31
21	3	12	11	19	10	4	22	17	18
13	28	11	14	21					

Drug B:
19	12	20	28	22	35	1	21	21	26
18	28	29	20	15	32	31	24	22	26
18	20	22	35	30	18	25	24	19	21

1) Construct a back-to-back stem-and-leaf plot to display the data.

2.4 Homework

1) Country Club, Bethesda, Maryland, with prizes ranging from $465,000 for first place to $5000. Par for the course is 70. The tournament consists of four rounds played on different days. The scores for each round of the 32 players who placed in the money (more than $17,000) were given on a web site.

The scores for the first round were as follows:

71	65	67	73	74	73	71	71	74	73	71
70	75	71	72	71	75	75	71	71	74	75
66	75	75	75	71	72	72	73	71	67	

The scores for the fourth round for these players were as follows:

69	69	73	74	72	72	70	71	71	70	72
73	73	72	71	71	71	69	70	71	72	73
74	72	71	68	69	70	69	71	73	74	

(a) Make a back-to-back stem-and-leaf display for the first-round scores.

(b) Compare the two distributions. How do the highest scores compare? How do the lowest scores compare?

2) How long did *real* cowboys live? One answer may be found in the book *The Last Cowboys* by Connie Brooks. This delightful book presents a thoughtful sociological study of cowboys in West Texas and Southeastern New Mexico around the year 1890. A sample of 32 cowboys gave the following years of longevity:

58	52	68	86	72	66	97	89	84	91	91
92	66	68	87	86	73	61	70	75	72	73
85	84	90	57	77	76	84	93	58	47	

(a) Make a stem-and-leaf display for these data.