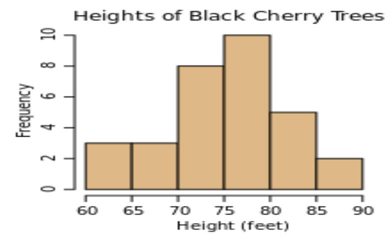
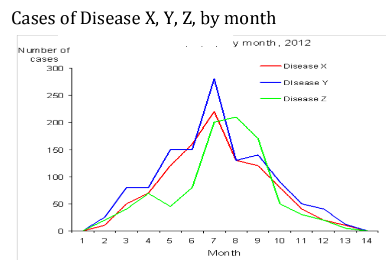
2.2 Histograms, Frequency Polygons, and Time Series Graphs

	Used for	Examples
A ***histogram*** is a bar graph that displays the data of a frequency distribution by using vertical bars to represent the frequencies.	• visually displaying the "shape" of the data, allowing the reader to determine some of the dominant characteristics of the data at a glance.	
A ***frequency polygon*** is a line graph of a frequency distribution with midpoints of the class intervals marked on the x-axis and the frequencies on the y-axis; data points are plotted and connected with straight lines.	• comparing the dominant characteristics of two sets of data simultaneously	
An ***time series graph*** is a line graph for which the horizontal axis is labeled with dates or time periods and the vertical axis	• tracking the growth or decay of a single category over an extended period of time	

Frequency Distribution

Suppose a manager of a sporting goods chain wants to reorganize it's shoe department offering a new line of running shoes that are guaranteed for 1,000 miles of wear. A researcher for the company must first collect data to determine the likely demand for such a shoe. A survey is given randomly to 500 runners asking how many miles they run per week; 20 of the data values are listed below.

The researcher would use all 500 data values, but we will use only 20 to keep our data analysis manageable while learning new graphing techniques.

Number of miles run per week of 20 runners

7	12	15	15	16	16	22	23	23	24
26	26	29	30	30	31	31	32	40	40

Let's first construct a frequency distribution with 7 classes.

$$\frac{\text{high-low}}{\#\ \text{of classes}} = \frac{40-7}{7} = \frac{33}{7} = 4.714 \approx 5$$

Number of miles (Class limits)	Number of miles (Class boundaries)	Tally	Frequency	Cumulative frequency

Histogram

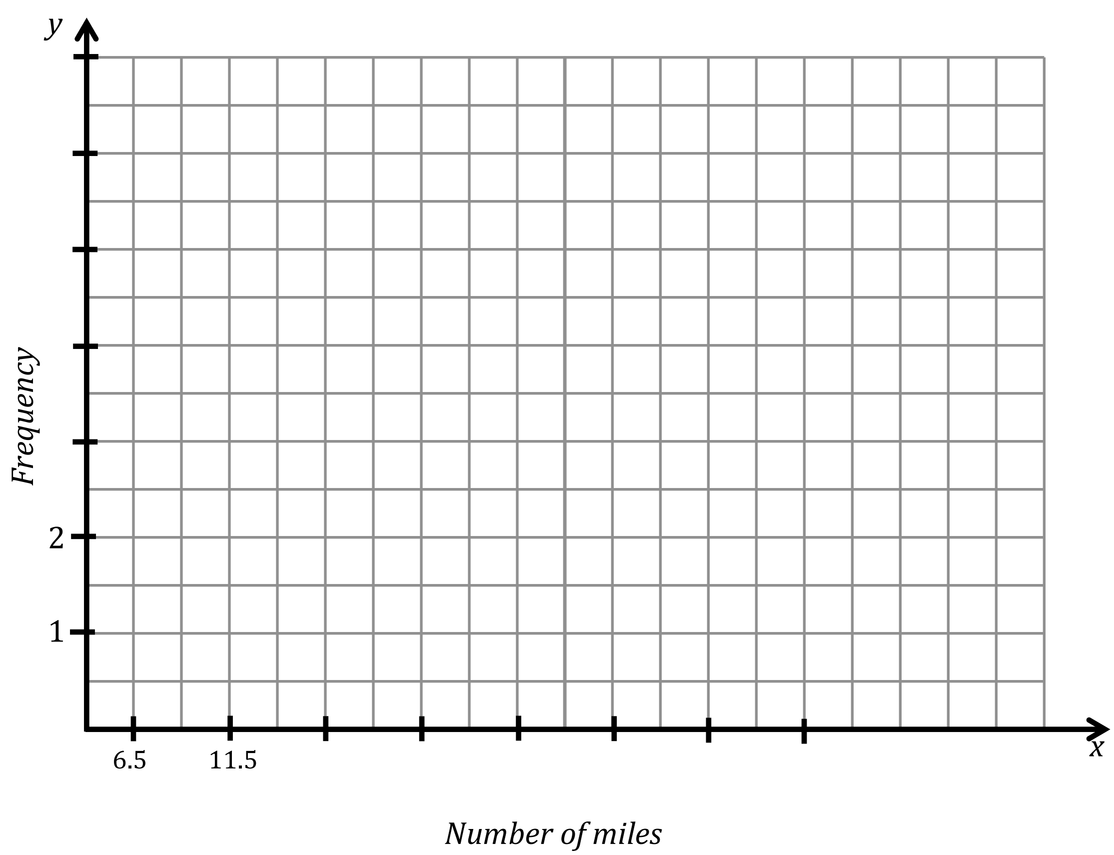
With a frequency distribution, the researcher can analyze the data for key characteristics. However, it is awkward to use in a presentation to an audience, such as a group of board members, not all of which are familiar with statistical analysis. Instead, the researcher can present the data visually in the form of a histogram.

To construct a histogram:

- Mark the class boundaries of the distribution along the horizontal axis.
- Number the vertical axis to accommodate the highest frequency in the distribution.
- Draw a bar over each class interval, with height equal to the frequency for that class.

Note that histograms designed for general public viewing usually use the lower class limits as horizontal axis demarcations for ease of understanding.

The histogram for the frequency distribution of the data from above has been started for you below:



Frequency Polygon

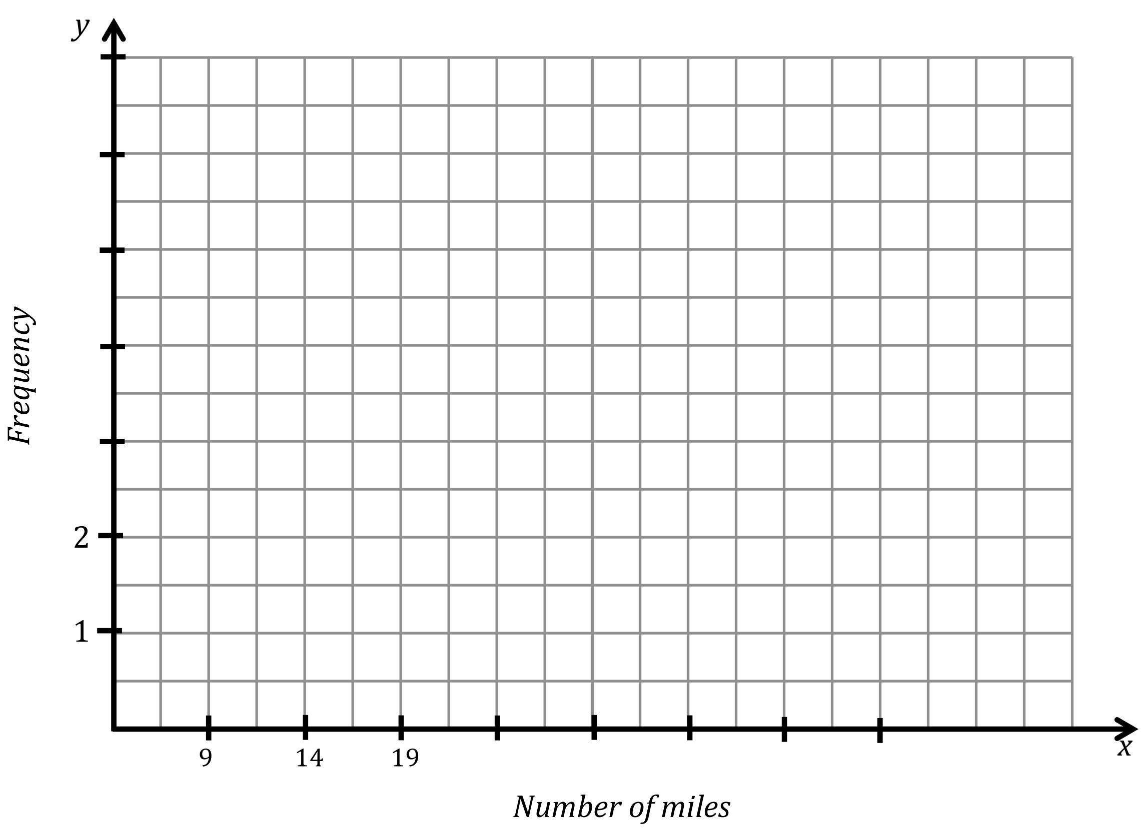
Suppose the researcher had conducted a similar survey of 500 randomly selected runners 10 years prior and wanted to compare the results of the 2 sets of data to see if the running distances are increasing, decreasing, or neither. The best graph to use to compare 2 frequency distributions is called a frequency polygon. 20 of the data values from 10 years prior are listed below:

7	8	8	12	13	15	15	16	18	20
21	21	24	25	25	26	31	32	40	40

Complete the frequency distribution for the 2nd set of data below:

Number of miles (Class limtis)	Number of miles (Class boundaries)	Tally	Frequency	Cumulative frequency
7 – 11	6.5 – 11.5			
12 – 16				

To construct a frequency polygon, mark the midpoints of the class intervals along the horizontal axis. Plot the data points (class midpoint, frequency) and connect the points with straight lines. Plot the data points for both data sets using different colors for each below.



Time Series Graph

Sometimes we want to track the growth or decay of a single category over a period time, such as rainfall, for which a time series graph is the line graph best suited for this purpose.

Consider the following climate data for Red Bluff, Redding, Weaverville, and Burney:

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Red Bluff 2014	0.06	6.41	3.17	0.29	0.61	0.00	0.01	0.00	2.44	1.20	2.87	9.76
Red Bluff 2015	0.04	0.79	0.71	0.80	0.30	0.01	0.03	0.00	0.20	0.22	1.15	4.11
Red Bluff 2016	8.80	0.41	6.21	1.28	0.51	1.46	0.00	0.00	0.00	3.48	3.26	3.37
Red Bluff 2017	6.80	5.51	2.10	2.27	0.49	0.47	0.00	0.13	0.10	0.66	3.01	0.01
Red Bluff 2018	5.02	0.03	2.95	3.18	1.22	0.10	0.00					
Red Bluff normal	*5.02*	*4.20*	*3.20*	*1.56*	*1.22*	*0.47*	*0.11*	*0.13*	*0.50*	*1.21*	*2.85*	*4.16*
Redding 2014	0.42	8.08	5.45	0.17	0.15	0.00	0.09	0.12	3.23	2.15	3.71	10.39
Redding 2015	0.26	3.38	1.06	1.18	0.34	0.57	0.00	0.00	0.55	0.27	1.38	8.21
Redding 2016	12.68	0.91	10.37	3.25	0.97	2.46	T	0.00	0.00	7.78	4.92	6.03
Redding 2017	9.92	8.25	2.38	2.93	0.05	0.10	T	0.10	0.61	0.33	5.23	0.11
Redding 2018	4.85	0.35	3.66	4.47	0.70	T	0.00	0.00				
Redding normal	*6.06*	*4.45*	*4.38*	*2.08*	*1.27*	*0.56*	*0.17*	*0.46*	*0.91*	*2.24*	*5.21*	*5.51*
Weaverville 2014	0.99	6.22	8.50	0.42	0.31	0.02	0.43	0.00	1.73	3.35	3.62	9.30
Weaverville 2015	1.14	7.66	1.43	0.75	0.74	0.62	1.05	0.04	0.58	0.66	1.35	12.40
Weaverville 2016	13.28	1.27	10.17	1.50	0.77	1.13	0.00	0.00	0.01	9.61	5.08	7.47
Weaverville 2017	12.79	10.23	5.63	4.84	0.19	1.23	0.00	0.03	0.97	1.19	6.19	0.59
Weaverville 2018	5.03	0.84	5.27	4.55	1.93	0.14	0.17					
Weaverville normal	*2.89*	*3.33*	*3.36*	*3.04*	*3.25*	*3.40*	*3.33*	*3.44*	*3.24*	*2.14*	*2.93*	*2.75*
Burney 2014	0.51	4.14	5.34	0.89	0.23	0.13	0.01	0.45	0.72	1.80	2.70	9.73
Burney 2015	0.20	5.12	0.97	1.21	1.64	0.53	0.73	0.01	0.24	1.08	1.22	4.68
Burney 2016	4.60	1.05	4.01	0.41	2.84	0.00	0.00	0.00	0.00	2.46	2.18	4.07
Burney 2017	6.87	6.27	1.57	2.91	0.06	0.62	0.00	0.02	1.20	0.46	4.41	0.26
Burney 2018	2.46	0.07	4.25	3.98	1.40	0.03	0.21					
Burney normal	*4.56*	*4.31*	*3.88*	*2.27*	*1.75*	*0.73*	*0.28*	*0.31*	*0.49*	*1.70*	*3.24*	*5.07*

Data source: http://www.cnrfc.noaa.gov/monthly_precip_2016.php
and http://weather.gov/sacramento

Use the data for both the 2017 and normal precipitation of your city to create overlaying time series graphs adhering to the following rules:

1. Label the horizontal axis by marking points for each time increment.
2. Plot each point according to the data given.
3. Connect the points with straight lines.