Chapter 1 § 1

Displaying Distributions with Graphs

Definitions:

Individuals – Objects described by a set of data. (i.e. people, animals, or things)

Variables – Any characteristic of an individual.

Exploratory data analysis – Statistical tools and ideas that can help examine data in order to describe their main features.

Categorical – Records which of several groups or categories an individual belongs to.

Quantitative – Takes numerical values for which it makes sense to do arithmetic operations like adding and averaging.

Distribution – of a variable tells us what values the variable takes and how often it takes these values. (The pattern of variation of a variable.)

Dotplots – another name for a line plot, which is used to graph a distribution of data.

Histograms – The most common graph of distributions with one quantitative variable

Stemplots – an additional method of graphing a distribution of quantitative data (for small data sets) using a 'steam' and a 'leaf'.

Definitions:

Overall pattern – a distribution that can be described by observing it's center, spread, and shape.

Deviations – The difference between a value in a frequency distribution and a fixed number (as the mean).

Center – The value or description of the middle of the data.

Spread – The extent of the data from the smallest to largest value.

Shape – the approximate design of a distribution as either symmetric or skewed.

Symmetric – the right and left sides of a distribution of a graph are approximately mirror images of each other.

Skewed to the right – the upper half of the observations extends much farther out than the left half.

Skewed to the left – the lower half of the observations extends much farther out than the right half.

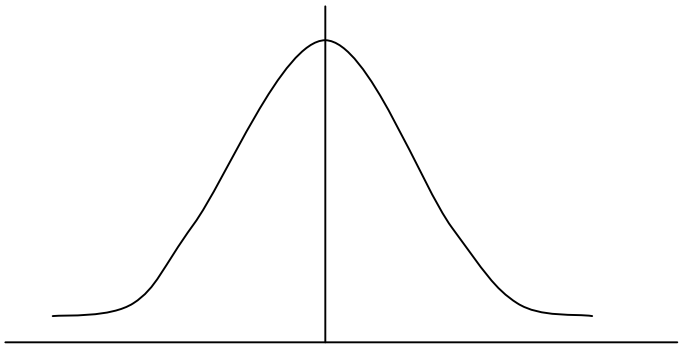
Outliers – any graph of data is an individual observation that falls outside the overall pattern of the graph.

Time plot – of a variable plots each observation against the time at which it was measured.

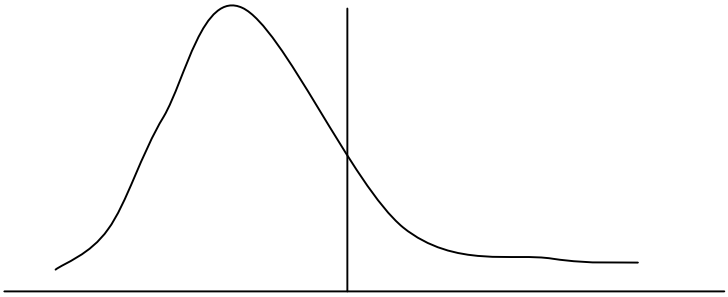
Trends – A long time upward or downward movement over time.

Categorical or Quantitative

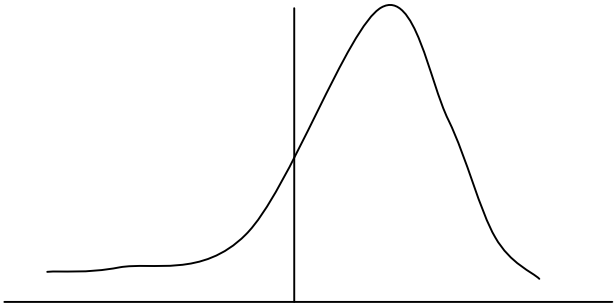
Model of Car:	Categorical
Height:	Quantitative
Shoe Size:	Quantitative
Grades:	Quantitative
Type of Dog:	Categorical
Gender :	Categorical
Pulse Rate:	Quantitative
ACT score:	Quantitative
Class Size:	Quantitative
Race:	Categorical
Type of Music:	Categorical
Radio Station:	Quantitative



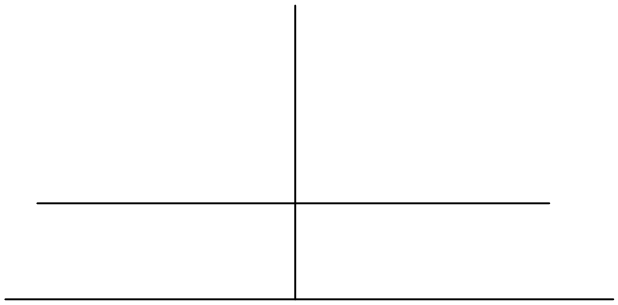
Normal Distribution



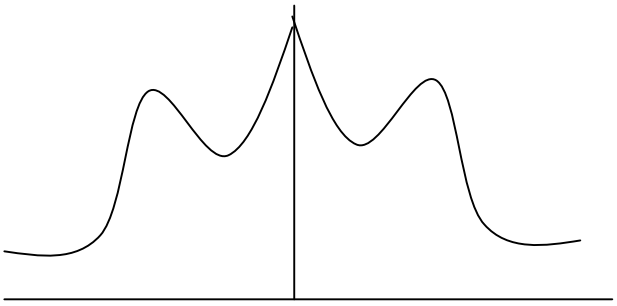
Skewed Right



Skewed Left



Uniform



Symmetric

To describe the overall pattern of a distribution:
(4 features of a histogram)

Center {mean, median, center, etc…}

Spread {range from lowest to highest}

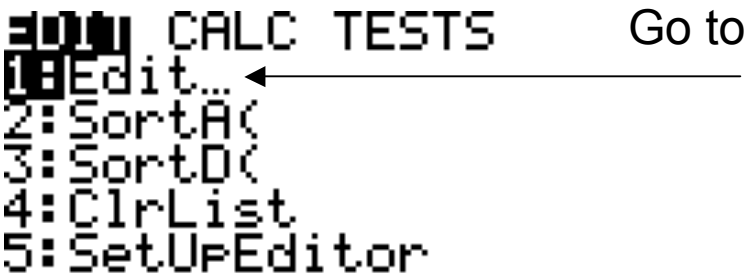
Shape {normal, skewness, etc…}

Exceptions {outliers}

Here's a way to think of the types of variables

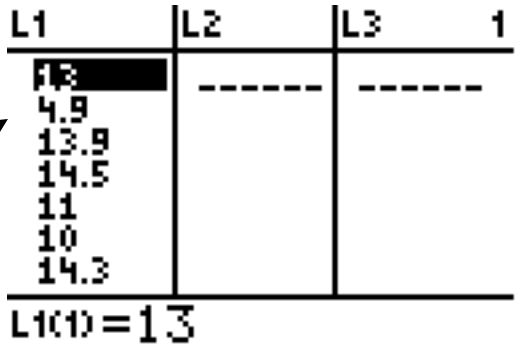
Variable	Graphical Displays	Numerical Summaries
Categorical	Pie Charts, Bar Graphs	Counts, percents, proportions
Quantitative	Dotplots, stemplots, histograms	Center: mean, median, mode; Spread: range, standard deviation, Interquartile range

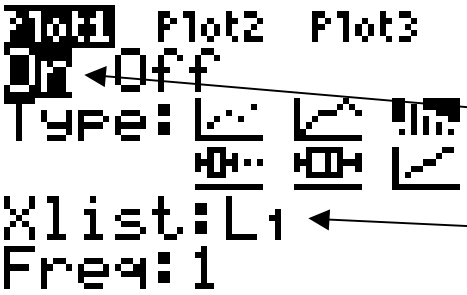
Note the data from table 1.1 on page 10 on
the percent of population 65 years old and over by state (1995)



Go to **Stat** then **Edit**

Then enter the
data under L_1





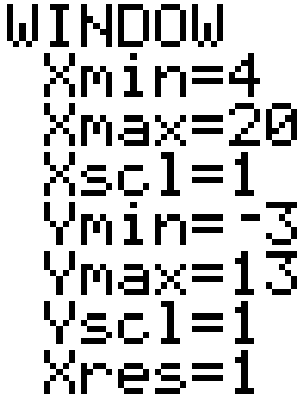
Go to **2nd Stat Plot** and press **Enter**

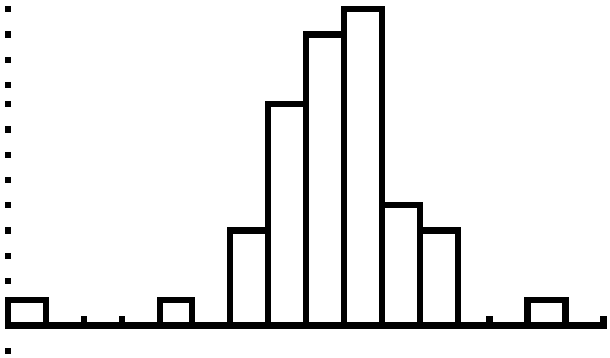
Turn the plot on

Select the Histogram

Select L_1 for the Xlist

Go to **Window**
and set the following
conditions





Then **Graph**

Let’s create a stemplot with the data from table 1.1

4	9
5	
6	
7	
8	8
9	
10	0 0 2 9
11	0 1 1 3 4 4 4 6 9
12	0 0 3 4 4 5 5 5 6 6 6 6
13	0 1 3 3 4 4 5 6 7 7 9 9 9
14	2 3 4 5 5
15	2 3 7 9
16	
17	
18	6

What comparisons and contrasts do you notice?

Describe the distribution of the data.