Section 11.4
Tools for Exploring Data

GRAPHS

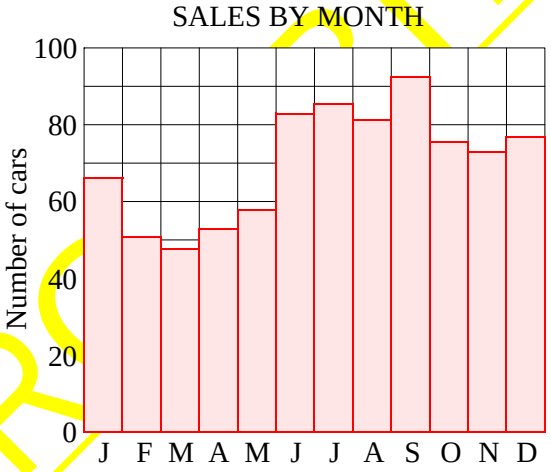
Someone once said "a picture is worth a thousand words." If the statement is true, then it is probably the reason why in statistics so many graphs are used.

A graph is a way to summarize data into an organized demonstration that could be read quickly.

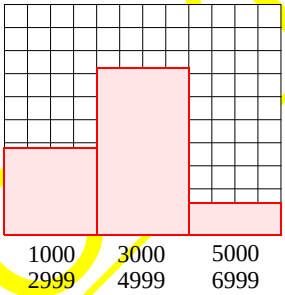
Bar Graphs and Histograms

As the name indicates, a bar graph uses bars. Bar graphs represent data based on frequency (how often.) In the case of the bar graph, bars show a specific category. The graph below shows the category "number of cars" sold by a dealership during 2006. To build a bar graph, first decide the size of the vertical scale. Because from the table the highest number of cars sold per month is close to 100, that is the limit set for the graph. The division of the scale is set at 10 for neatness. The horizontal scale was set per month because that is the way sales commissions are computed.

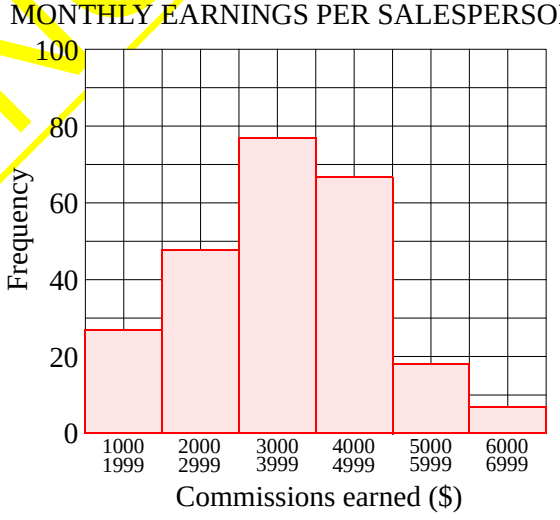
Month	number of cars
Jan.	67
Feb.	51
Mar.	48
Apr.	53
May	58
June	83
July	85
Aug.	80
Sept.	93
Oct.	76
Nov.	73
Dec.	77

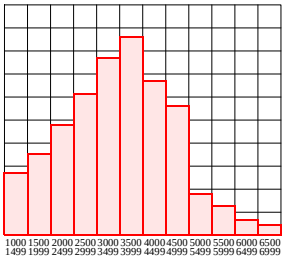


Unlike in the bar graph, in a **histogram**, each bar represents a **range** of data (see graphs below). The intervals selected are the actual commissions earned by salespersons. Notice that selecting the size of the interval is important. If the intervals are too few or too many, the graph loses meaning.



Too few intervals

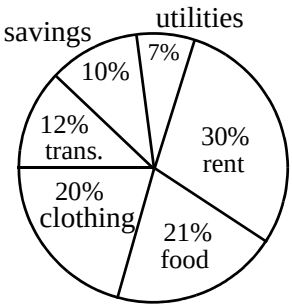


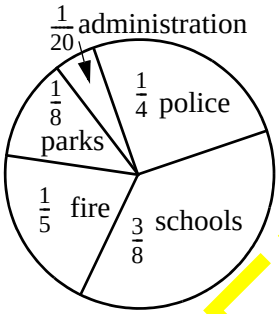


Too many intervals

Circle Graphs

Circle graphs are also called "pie charts." Although circle graphs have limitations and cannot be expanded like a bar graph, they are an excellent way to present fractions, portions, or percentage as a "piece of a pie."





The first circle graph to the right shows a home budget. Notice that all the percents add up to 100%. To build a circle graph you must recall that a circle has 360º; therefore, the piece of the pie devoted to, for example, 7%, has to be proportional to the other pieces of the pie where 100% = 360º:

$$\frac{7}{100} = \frac{angle}{360} \qquad angle = \frac{7 \times 360}{100} = 25.2^{o}$$

25.2º is the angle for 7%

The second circle graph above shows how a city spends taxes. Similarly, forming a circle graph with fractions equates a fix amount, in this case one (1), with 360º. For example, the angle of the piece of the "pie" devoted to parks will be:

$$\frac{1/8}{1} = \frac{angle}{360} \qquad angle = \frac{\frac{1}{8} \times 360}{1} = \frac{\frac{1 \times 360}{8}}{1} = \frac{\frac{360}{8}}{1} = 45^{o}$$

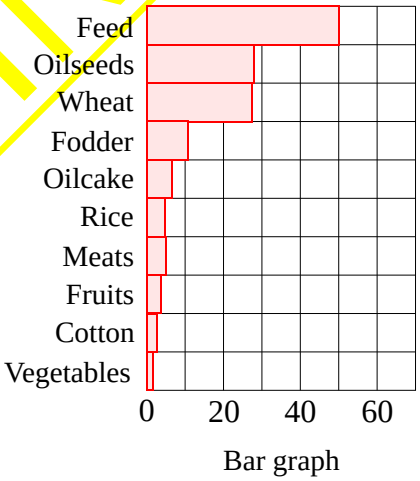
PLOTS

Plots are built from points that define a location. The location of points, in turn, may or may not develop into recognizable patterns that define the direction a set of data is taking.

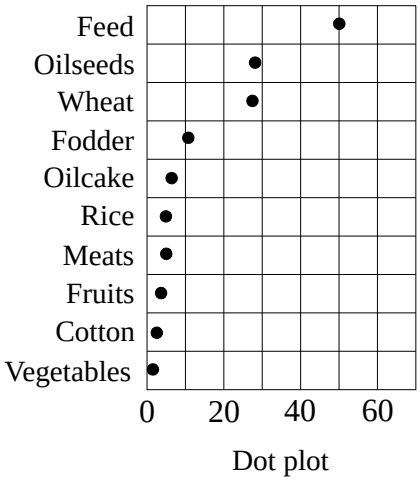
Dot plots may be used as an alternative to bar charts, and many researchers prefer them because they are cleaner and easier to read. For example, the dot plot below shows the same data as the bar graph, yet it is simpler to produce and easy to read.

commodity	millions of metric tons
Feed	50
Oilseeds	28
Wheat	27
Fodder	11
Oil cake	6
Rice	5
Meats	5
Fruits	4
Cotton	3
Vegetables	2

US AGRICULTURAL EXPORTS
PRINCIPAL COMMODITES
(2005) Millions of metric tons.



Bar graph



Dot plot

Box plots, also called "box and whiskers" plots, besides establishing location, contain data on the median, upper quartile (UQ), lower quartile (LQ), interquartile range (IQR), upper limit (UL), lower limit (LL), and outliers (far-off points).

To build a box plot:

1. Organize the data from lowest to highest.
2. Find the median (number where 50% of all data is above and 50% below, see page 180.)
3. Find the lower quartile (number where 75% of all data is above and 25% below)
4. Find the upper quartile (number where 25% of all data is above and 75% below)

Example: Build a box plot for the data: 86, 73, 55, 96, 100, 75, 73, 66, 80, 40, 120, 86, 78, 92, 67, 82. Find the interquartile range, upper and lower limits, and outliers.

Organize: 40, 55, 66, 67, 73, 73, 75, 78, 80, 82, 86, 86, 92, 96, 100, 120 (LQ, median, UQ)

Median: 79 Lower quartile: $\frac{67 + 73}{2} = 70$ Upper quartile: $\frac{86 + 92}{2} = 89$

Because there are 16 numbers (even amount), the median must be after the 8th number (half the numbers), but before the 9th number, in the middle of 78 and 80, or 79.

The lower quartile must be in the middle of the first 8 numbers (after the 4th, but before the 5th). Add the 4th and 5th number, divide by 2, and the LQ = 70.

The upper quartile must be in the middle of the last 8 numbers (after the 12th, but before the 13th). Add the 12th and 13th number, divide by 2, and the UQ = 89.

The interquartile range is the distance between UQ and LQ: IQR = UQ – LQ = 19

Outliers are those numbers that fall beyond 1.5 the distance of the IQR, or $IQR \times 1.5 = 19 \times 1.5 = 28.5$

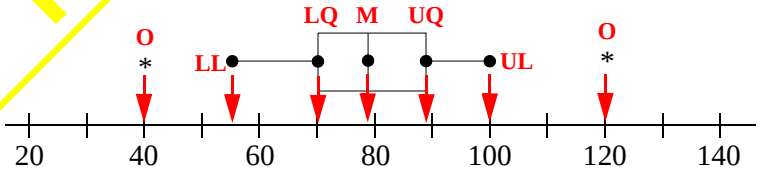
Add 28.5 to the UQ 28.5 + 89 = **117.5** and Subtract 28.5 from the LQ 70 – 28.5 = **41.5**

Any numbers below **41.5** or above **117.5** are not worth considering because they are **outliers**.

Therefore, both 40 and 120 are outliers and the UL falls to 100 and the LL moves up to 55.

Building the box:

On a scale that includes all numbers observed, draw a box with the LQ and UQ as the sides of the box, and draw a line dividing the box at the median (M). Add lines (whiskers) to the LL and UL. Mark the outliers (O) with an asterisk or a cross.



Stem-and-Leaf plots are frequency plots that are used to organize data and determine central tendency. The idea is to split every number into two sections: the left part is the **stem**, the other the **leaf**. For example, the number 123 would be split 12|3, and the number 48 4|8. A column of the stem is matched by a string representing the leaf.

Example: Make a stem-and-leaf plot of the following airplane take-off waiting times, in minutes.

5, 6, 3, 1, 16, 22, 34, 21, 1, 7, 55, 12, 33, 8,
4, 55, 34, 17, 25, 44, 48, 25, 22, 7, 14, 31, 40

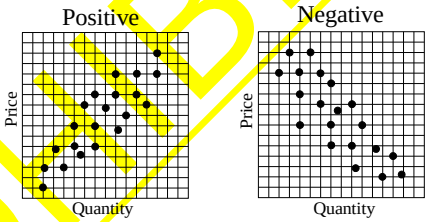
STEM	LEAF
0	1, 1, 3, 4, 5, 6, 7, 7, 8
1	2, 4, 6, 7
2	1, 2, 2, 5, 5
3	1, 3, 4, 4
4	0, 4, 8
5	5, 5

The stem will be the ten digits and the leaf the one digits. The central tendencies may be checked by looking at the distribution. In this particular case the data is "skewed" (leaning) towards the lower values, while the median is 21 minutes and there are 6 modes: Poor central tendencies.

Notice that a stem-and-leaf plot looks like a horizontal histogram: They both have intervals and frequency.

A benefit of the stem-and-leaf plot is that it displays all of the individual values within that interval, while the histogram does not.

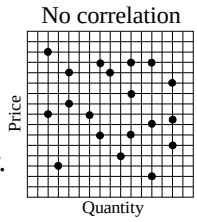
Scatter plots are used to summarize how a set of data behaves when compared to another set. Called correlation, it is a way of checking patterns of plotted points using two variables. The more the points tend to bunch up to form a distinctive group that shows direction (like a flock of birds), to the verge that they look like going somewhere, the stronger the correlation (the variables follow each other.) There are three ways of reading scatter plots:



1. If they aim towards the northeast (lower left to upper right), the correlation is positive.

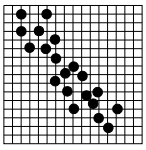
2. If they aim towards the southeast (upper left to lower right), the correlation is negative.

3. If the points are random without showing any particular aim, the correlation is very low.



Practice:

1. The weekly closing price for a certain stock is 85, 78, 77, 86, 88, 77, 81, 72, 75, 77, 85, 88, 92, 88, 90. Draw a bar graph to show the data.
2. The weight, in pounds, of some boxes is 20, 30, 35, 56, 39, 55, 57, 33, 45, 32, 38, 42, 58, 33, 42, 47, 28. Draw a histogram to exhibit the data.
3. On a monthly basis, Joe spends $854.40 for rent, $106.80 for utilities, $64.08 for gas, $149.52 for clothing and $961.20 for groceries. Draw a percent circle graph to show the expense distribution.
4. Monthly averages, in inches, of rain-fall in the Everglades from January to December are: 3, 2, 5, 10, 35, 34, 32, 40, 42, 30, 7, 3. Draw a dot plot for the data.
5. Mark is cutting pieces of wood that measure, in inches, 0.45, 0.44, 0.40, 0.41, 0.43, 0.44, 0.33, 0.46, 0.44, 0.42, 0.45, 0.59, 0.41, 0.43. Draw a box plot and find the interquartile range and outliers, if any.
6. Make a stem-and-leaf plot representing the height of the students in math class if they measure, in inches, 72, 68, 69, 74, 59, 60, 68, 67, 68, 66, 68, 69, 70, 69, 65, 64, 66, 68, 69, 67, 66, 68, 69.
7. Determine the type of correlation shown in the graph to the right, if any.



8. Draw stem-and-leaf graph using the following qualifying speeds: 199, 201, 210, 198, 200, 213, 185, 190, 193, 186, 208, 214, 202, 197, 199, 211, 196, 199, 203, 207, 203, 202.
9. The cookies Clara made have the following weights, in grams: 28, 29, 28, 26, 27, 31, 26, 27, 31, 30, 29, 24, 32, 28, 29, 27, 26, 26, 27, 29, 30, 26, 27, 28. Draw a box plot and find the interquartile range and outliers, if any.