

QUANTITATIVE METHODS

Martin Huard
Winter 2005

Organizing Data

Features of a **bar graph**.

1. Bars can be vertical or horizontal.
2. Bars are of uniform width and uniformly spaced.
3. The lengths of the bars represent values of the variable being displayed, the frequency of occurrence, or the percentage of occurrence. The same measurement scale is used for the length of each bar.
4. The graph is well annotated with title, label for each bar, and vertical scale or actual value for the length of each bar.

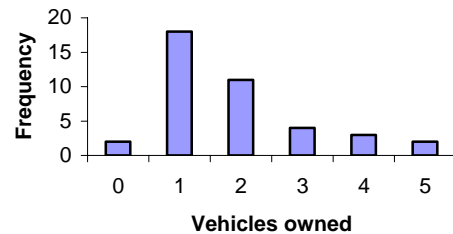
A **Pareto chart** is a bar graph in where the bars are arranged from left to right according to decreasing height. Usually used to represent nominal data.

A **circle graph** or **pie chart** may be used instead of a bar graph.

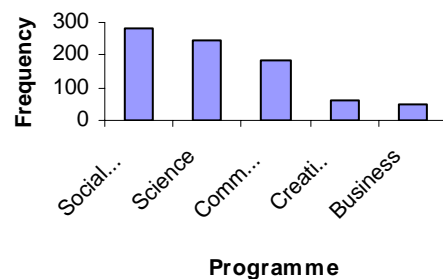
A **time plot** is a graph showing data measurements in chronological order. Time is put on the horizontal axis and the data points are connected by line segments.

Note: for any graph, provide a title, label axes and identify units of measure.

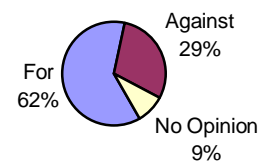
Vehicles Owned per Family for a Sample of 40 Families



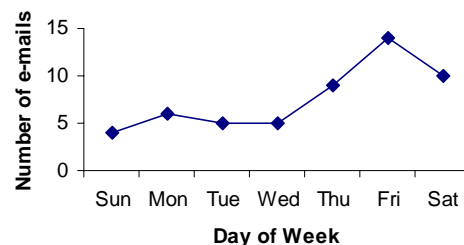
Distribution of SLC students according to programme



Canadian's opinion on gay mariages



Daily average for the number of e-mails received by students



Histograms are bar graphs in which

1. The bars have the same width and always touch.
2. The width of a bar represents a quantitative value, such as age, rather than a category.
3. The height of each bar indicates frequency.

The number of classes is usually from 5 to 15. For a sample of size n , the number of classes can be approximated with \sqrt{n} .

To make a frequency distribution table from integer data, we do the following

1. Determine the *class width*:

Compute $\frac{\text{Largest data value} - \text{Smallest data value}}{\text{Desired number of classes}}$ and round to a convenient number

2. Create the distinct classes. The *lower class limit* of the first class is usually a nice number smaller than the smallest data value. Add the class width to this number to get the lower class limit of the next class. Note that the limits should be rounded off to as many decimals as was present in the original data.
3. Tally the data into classes. Each data value should fall into exactly one class.
4. The relative frequency is calculated using $\frac{f}{n}$ where f is the class frequency and n the total sample size
5. Compute the *midpoint* (class mark) for each class.

$$\text{Midpoint} = \frac{\text{Lower class limit} + \text{Upper class limit}}{2}$$

6. Sketch the histogram.

Frequency Distribution for the Grades Obtained by the Students in a Statistics Class			
Grade (class limits)	Frequency	Relative Frequency	Midpoint
40 to 49	2	5%	44.5
50 to 59	1	3%	54.5
60 to 69	4	11%	64.5
70 to 79	9	24%	74.5
80 to 89	15	41%	84.5
90 to 99	6	16%	94.5

