Chapter 2 - Representation of data

Data can be represented in many different ways, including:

1. Cumulative frequency graphs (such as the one shown above)

2. Stem and leaf diagrams

For example:

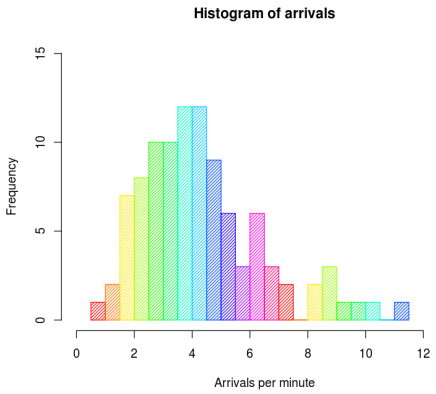
2.3, 2.5, 2.5, 2.7, 2.8 3.2, 3.6, 3.6, 4.5, 5.0

And here is the stem-and-leaf plot:

Stem	Leaf
2	3 5 5 7 8
3	2 6 6
4	5
5	0

Stem "2" Leaf "3" means **2.3**

- In this case, each leaf is a decimal
- It is OK to repeat a leaf value
- 5.0 has a leaf of "0"



3. Histograms

Histograms are used to present continuous data. They show the rough location and general shape of the data and how spread out it is. The area of each bar corresponds to the frequency of each class.

To calculate the height of each bar (frequency density):

Area of the bar = k x frequency

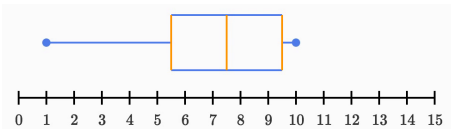
When k = 1:

Frequency density = frequency/class width

Joining the top of each bar in a histogram forms a frequency polygon.

4. Box and whisker diagrams (Boxplots)

A boxplot shows the quartiles, maximum, and minimum values and any others.



1. Collect and organize your data.
2. Calculate the median
   The line in the box is the median
3. Calculate each of the quartiles.
4. Create your plot line and draw a line for each of your quartiles.
   The orange lines on either side of the box are the lower and upper quartiles
5. Create a box connecting the quartiles.
6. Find the interquartile range.
7. Find the new upper and lower limits.

Comparing data

When comparing data, you comment on:

1) A measure of location
2) A measure of spread

You can use the mean and standard deviation or the median and interquartile range.

- The median should not be used with standard deviation and similarly, the mean should not be used with interquartile range