Data Representation and Visualisation Using Graph and Table.

COMP 1810 –Data and Web Analytics

Presentation Using Graphics and Tables.

In other to understand the information in your data, graphs and tables are used as the visual presentation.

Tables are obvious and we have been using them in excel or any other format for data.

The challenge is which type of graphs should be used that will provide the required information and make more sense in presenting the insight in the data set.

The rule of thumb is that the type of graph to used depends on the type of data you want to represent Eg, Categorical, Numeric, Discrete and Continuous data types.

Tables

This is a typical table all aspects of it must be identifies with title, column headings and values in appropriate number of decimal.

Make sure that your table is as simple and clear as possible

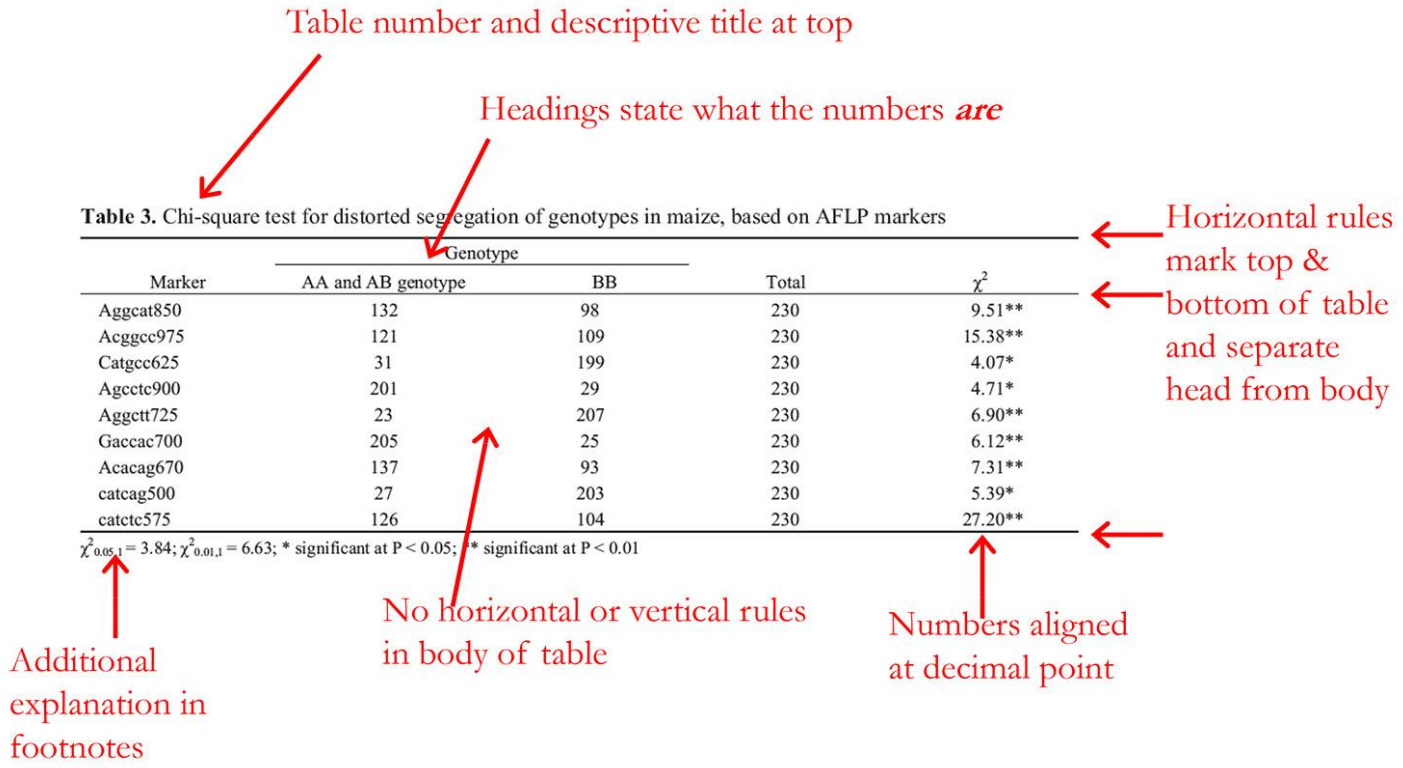


Table 3. Chi-square test for distorted segregation of genotypes in maize, based on AFLP markers

Marker	Genotype: AA and AB genotype	Genotype: BB	Total	χ^2
Aggcat850	132	98	230	9.51**
Acggcc975	121	109	230	15.38**
Catgcc625	31	199	230	4.07*
Agcctc900	201	29	230	4.71*
Aggctt725	23	207	230	6.90**
Gaccac700	205	25	230	6.12**
Acacag670	137	93	230	7.31**
catcag500	27	203	230	5.39*
catctc575	126	104	230	27.20**

$\chi^2_{0.05,1} = 3.84$; $\chi^2_{0.01,1} = 6.63$; * significant at $P < 0.05$; ** significant at $P < 0.01$

Graphs

All graph must have these basic properties, all graph must at least have SALT

- S = Scale
- A = Axes
- L = Labels
- T = Title

Is your Graph Salty
S = Scale •Horizontal and vertical scale selected •Independent variable in the horizontal axis •Dependent variable in the vertical axis
A = Axes •Each axis is shown clearly •Each axis scale increment evenly •Each axis scale may or may not start from zero
L = Labels •Each axis identifies the data and units clearly
T = Title •The title should be written clearly usually at the top •The title should identify the purpose of the graph

Common graphs for data presentations

These are the most common graph used in data analysis,

1. Scatterplot
2. Bar Chart
3. Histogram
4. Density plot
5. Pie Chart
6. Correlation Matrix
7. Box Plot
8. Line Chart

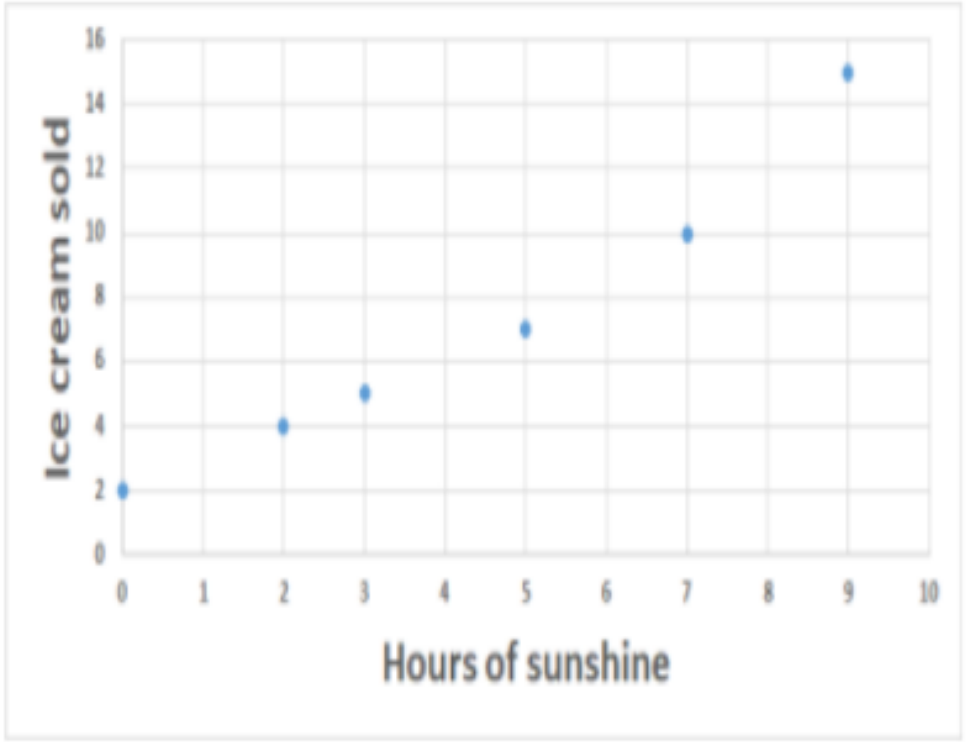
Scatterplot

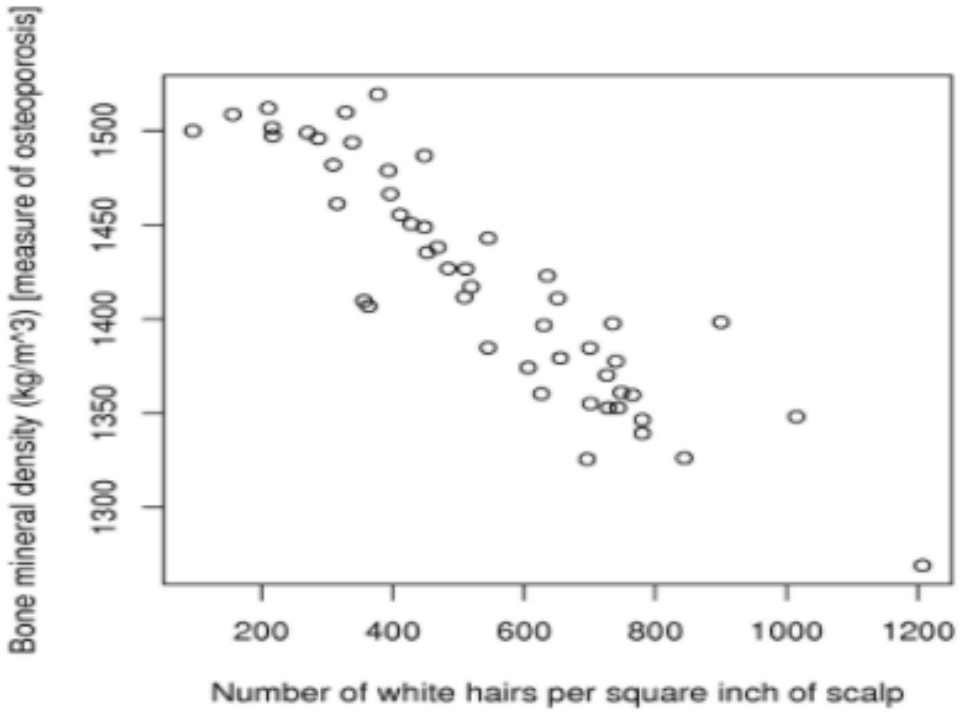
This is used for numeric data type, mostly to show relationship between two or more variable that are numeric

For example

"x" Hours of Sunshine	"y" Ice Creams Sold
0	2
2	4
3	5
5	7
7	10
9	15

Scatterplot





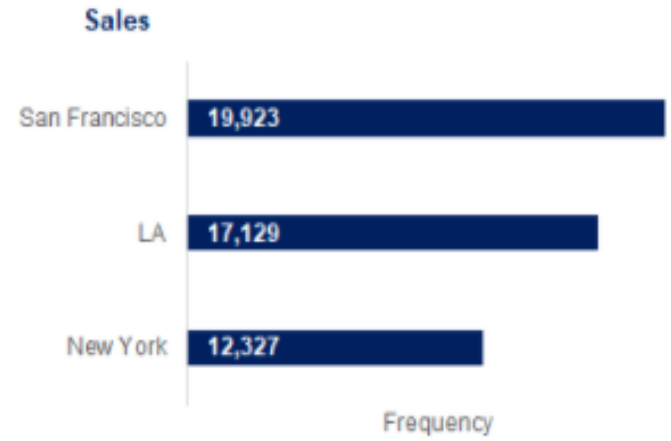
The two scattered plot shows both positive and negative relationship between the variables. The scattered plot would be reviewed in each topic when necessary

Bar Chart

A **bar chart** or **bar graph** is a chart or graph used to represent categorical data

Inform of a rectangular bar with heights of bar proportional to the size of values that they represent. It could be plotted vertically or horizontally. A vertical bar is called a **column chart**.

	Frequency
New York	12,327
LA	17,129
San Francisco	19,923
Total	49,379

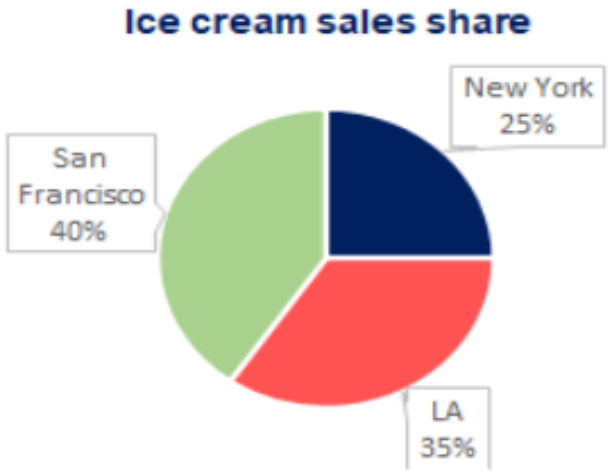


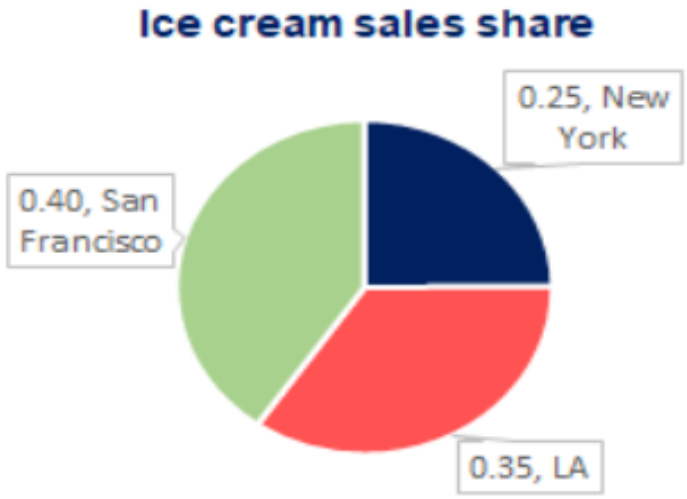
Bar Chart

Pie Chart

Solution:

	Frequency	Relative frequency	or.. relative frequency
New York	12,327	25%	0.25
LA	17,129	35%	0.35
San Francisco	19,923	40%	0.40
Total	49,379	100%	1.00



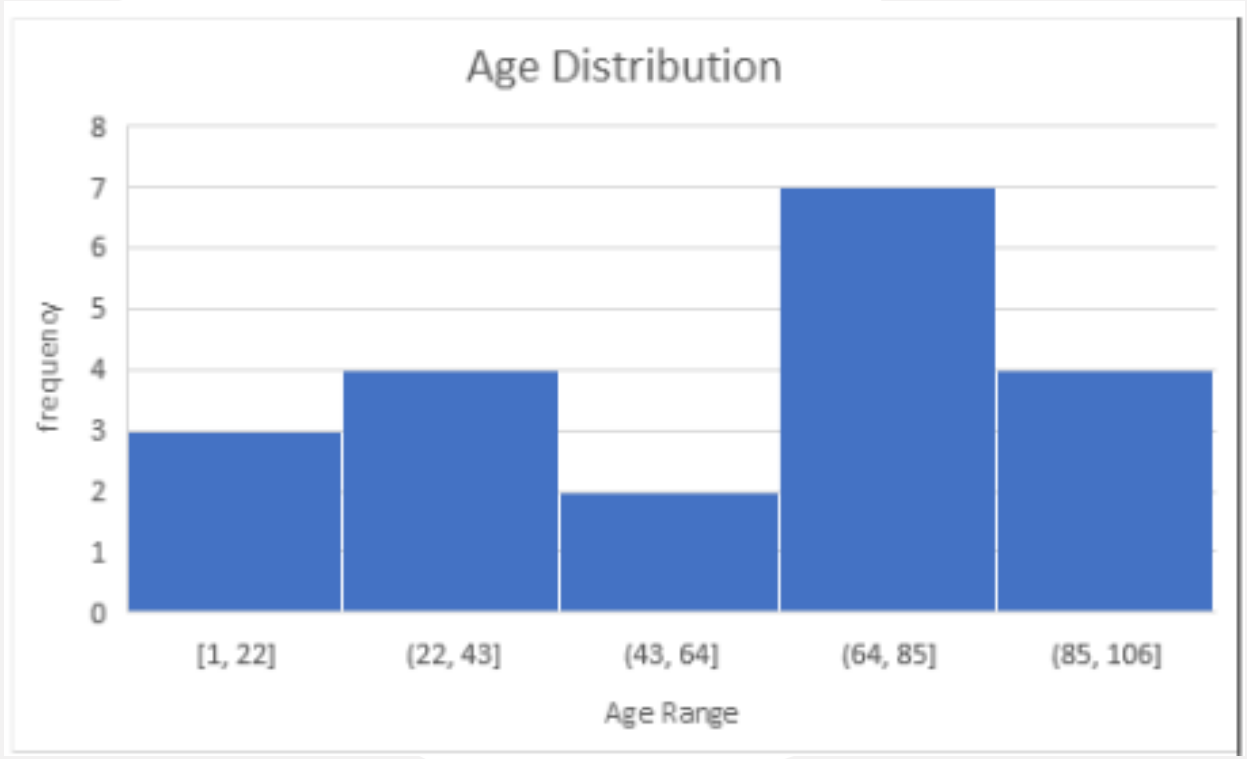


Histogram

Histogram look similar to Bar chart, but histogram is for continuous data, hence no space between the charts.

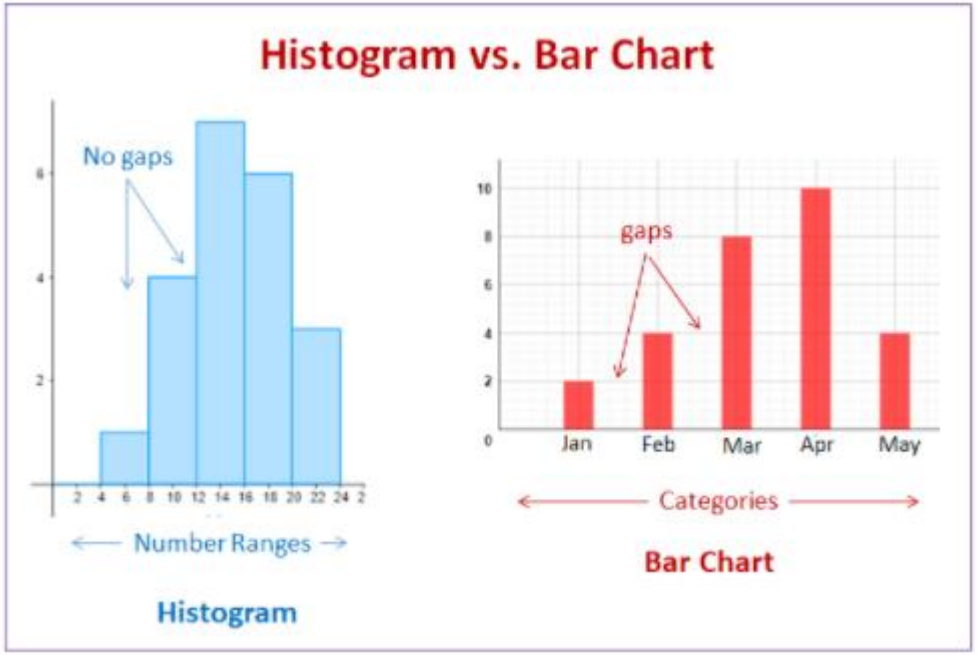
Histogram is used to show the distribution of numerical data, the horizontal is the range of values that has been bin (bucket) and the vertical is the frequency

Age
1
9
22
24
32
33
42
44
57
66
70
73
75
76
79
82
87
89
95
100



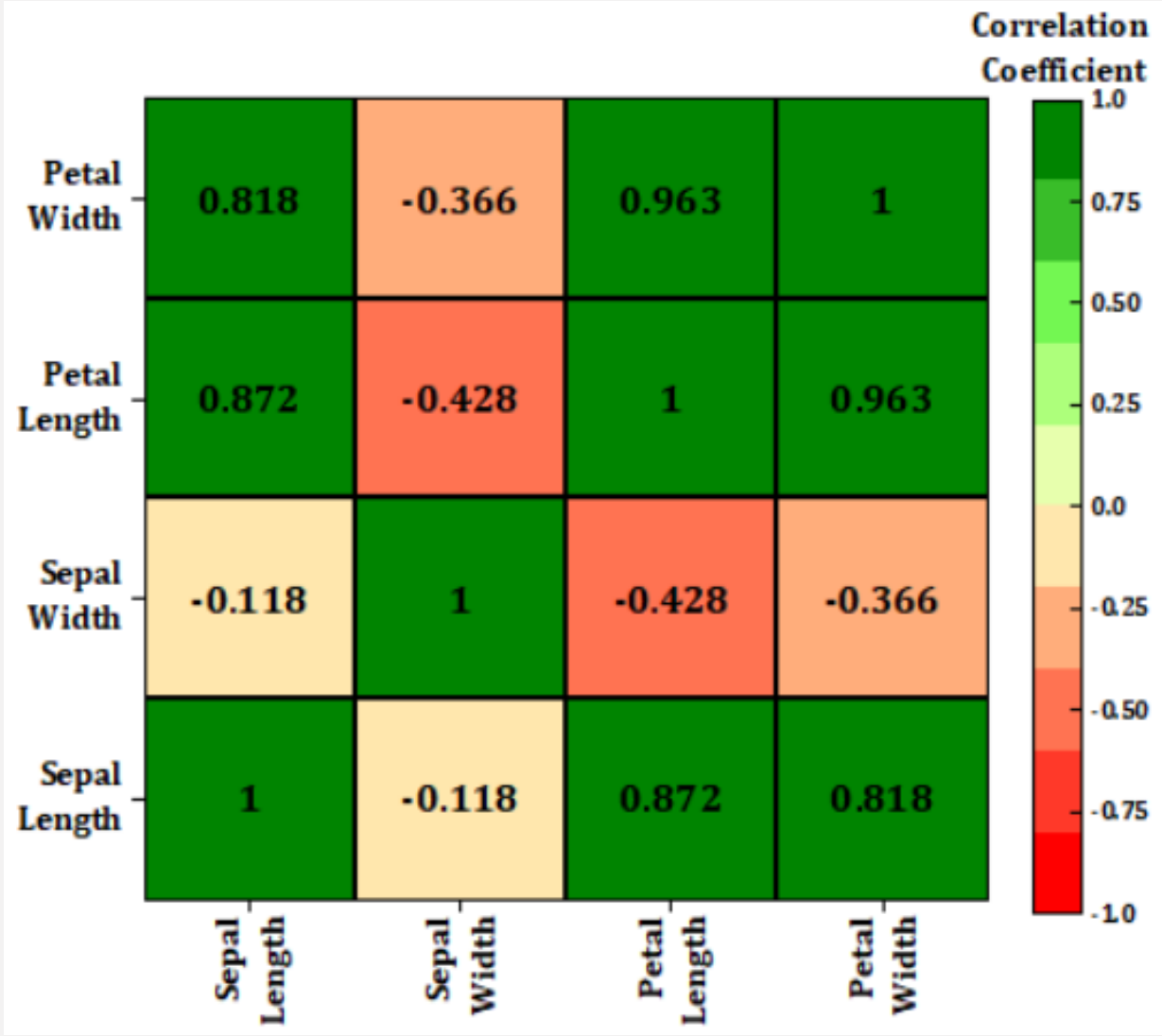
Histogram


Histogram of Monthly Salary
Frequency
200
150
100
50
0
$800
$900
$1000
$1100
$1200
$1300
$1400
$1500
$1600
$1700
$1800
Gross monthly salary

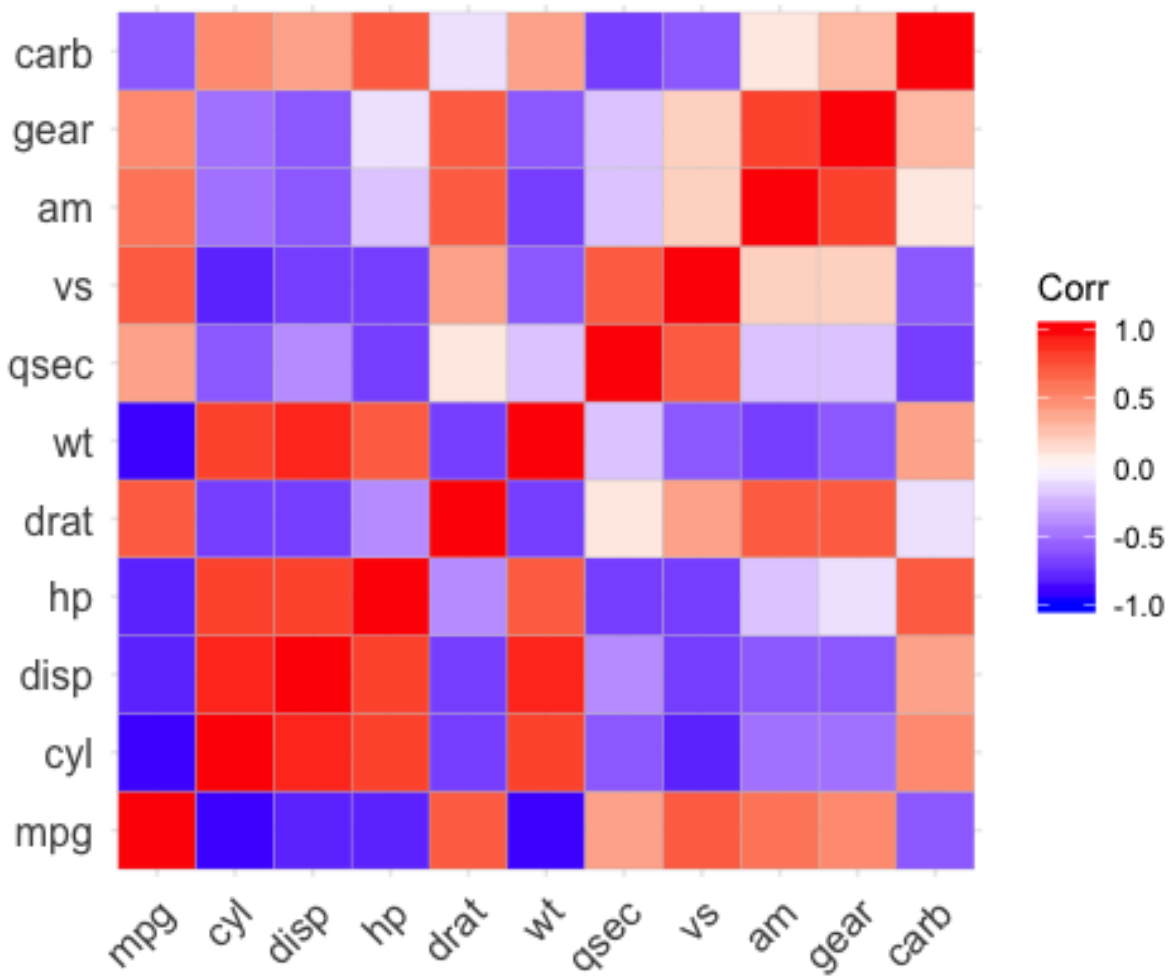

Histogram vs. Bar Chart
No gaps
Number Ranges
Histogram
gaps
Jan
Feb
Mar
Apr
May
Categories
Bar Chart

Correlation Matrix

Is used to visualized the relationship between variables in two-dimension color, the intensity of the two colors shows how strong the relationships between the variables. (it demonstrates correlation very well).

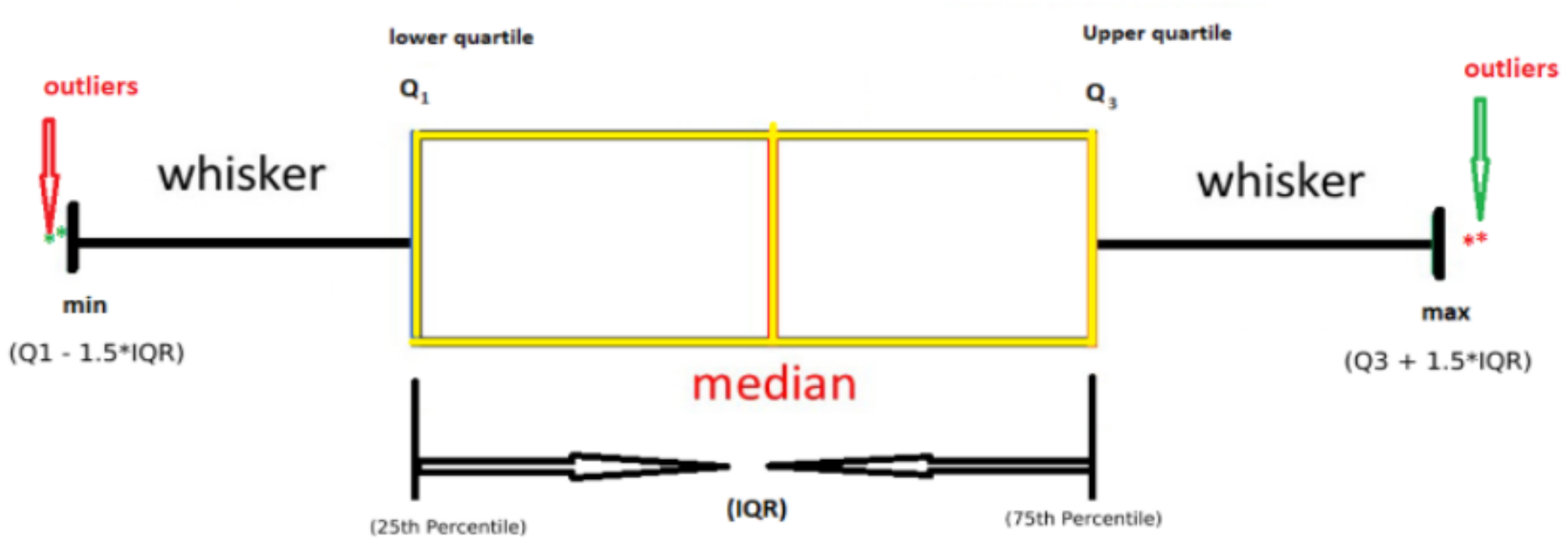


Correlation Matrix



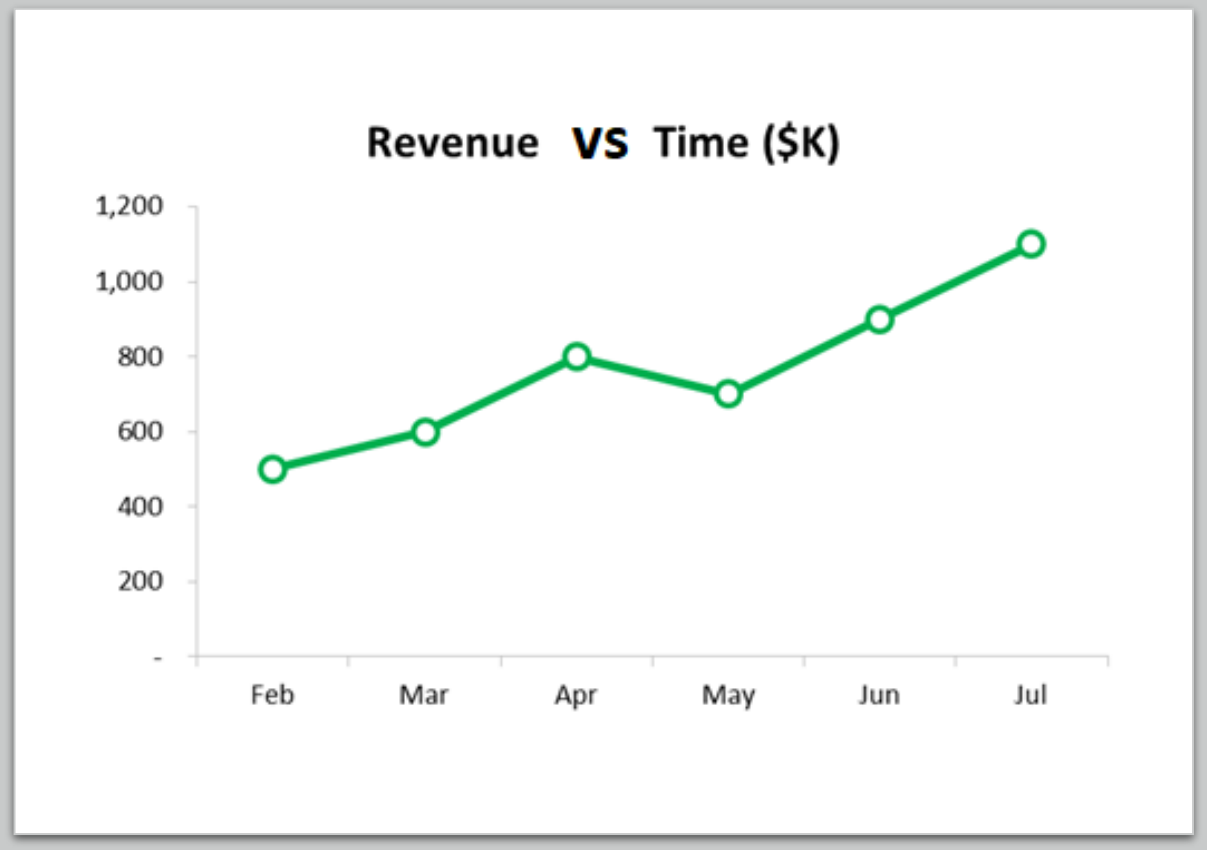
Box Plot

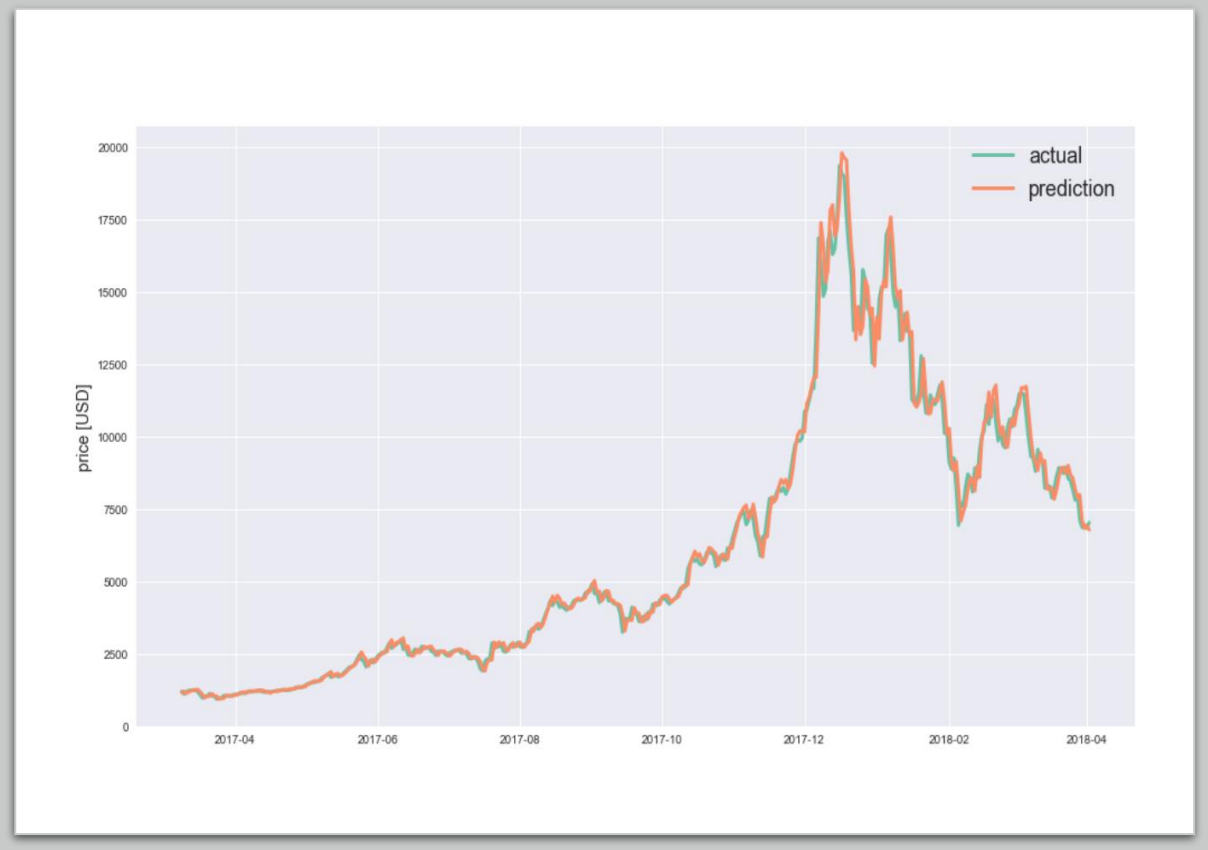
This chart is used to show the summary of a dataset; it shows the minimum, maximum, lower quartile (25 Percentile) , median, Upper quartile (75 Percentile). It also provides values that are most likely Outliers (below minimum and above maximum).



Line charts

A line chart or line graph is used to show series of data points and all the points are connected by line segments. Mostly data point that varies with times like seconds, mins, hours, days week, years, quarterly are mostly displayed using line charts.





End