For a given data series

A) Find 5 Number Summary and

B) Draw Box Plot for a data set

Week 3

## Aniket Vaishnav

2017BTEIT00062

# Introduction

**Box Plot** is also known as **Whisker plot**. It provides a summary of 5 different quantities which are :

1. Minimum
2. Q1
3. Median (also known as Q2)
4. Q3
5. Maximum

IQR ( InterQuartile Range ) = Q3 - Q1

Upper = Q1 - 1.5 \* IQR

Lower = Q3 + 1.5 \* IQR

Box Plot are useful when

1. Handles Large Data Easily
2. Exact Values Not Retained
3. A clear summary
4. Displays outliers

# Code

Written in python can be found at [GitHub](https://github.com/OBrutus/DataMining/blob/master/3.BoxPlot/main.py)

import pandas as pd

import matplotlib.pyplot as plt

def display\_prompt(data):

i = -1

for key in data.keys():

i += 1

if i==0:

print('key','\t', 'Attribute')

continue

print(i,'\t', key)

def gui\_plot(data):

fig = plt.figure(figsize =(10, 7))

# Creating plot

plt.boxplot(data)

# show plot

plt.show()

def get\_quartile(X):

N = len(X)

Q = None

if N%2 == 0:

Q = (X[N//2] + X[N//2-1]) / 2

else:

Q = X[N//2] / 2

return Q

if \_\_name\_\_ == '\_\_main\_\_':

data = pd.read\_csv(open('dataset.csv'))

display\_prompt(data)

key = 'Chhattisgarh'

# key = input('choose a attribue :\t ')

# print(data[key])

x=list(data[key])

X=sorted(x)

# print((x))

N = len(x)

Q1 = None

Q2 = None

Q3 = None

MIN = None

MAX = None

## for Q2

Q2 = get\_quartile(X)

## for Q1

Q1 = get\_quartile(X[:N//2])

## for Q3

if N%2 == 0:

Q3 = get\_quartile(X[N//2:])

else:

Q3 = get\_quartile(X[N//2+1:])

IQR = Q3 - Q1

MIN = Q1 - 1.5\*IQR

MAX = Q3 + 1.5\*IQR

print('''

N = {}

IQR = {}

MIN = {}

Q1 = {}

Q2 = {}

Q3 = {}

MAX = {}

'''.format(N, IQR, MIN, Q1, Q2, Q3, MAX) )

gui\_plot(data[key])

# Dependencies

Python 3+

Pandas

Matplotlib

Run via python3 main.py

# Screenshots

# 

# References

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>

<https://en.wikipedia.org/wiki/Box_plot>