

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

$IQR \text{ (InterQuartile Range)} = Q3 - Q1$

$Upper = Q1 - 1.5 * IQR$

$Lower = Q3 + 1.5 * IQR$

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](#)

```
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`

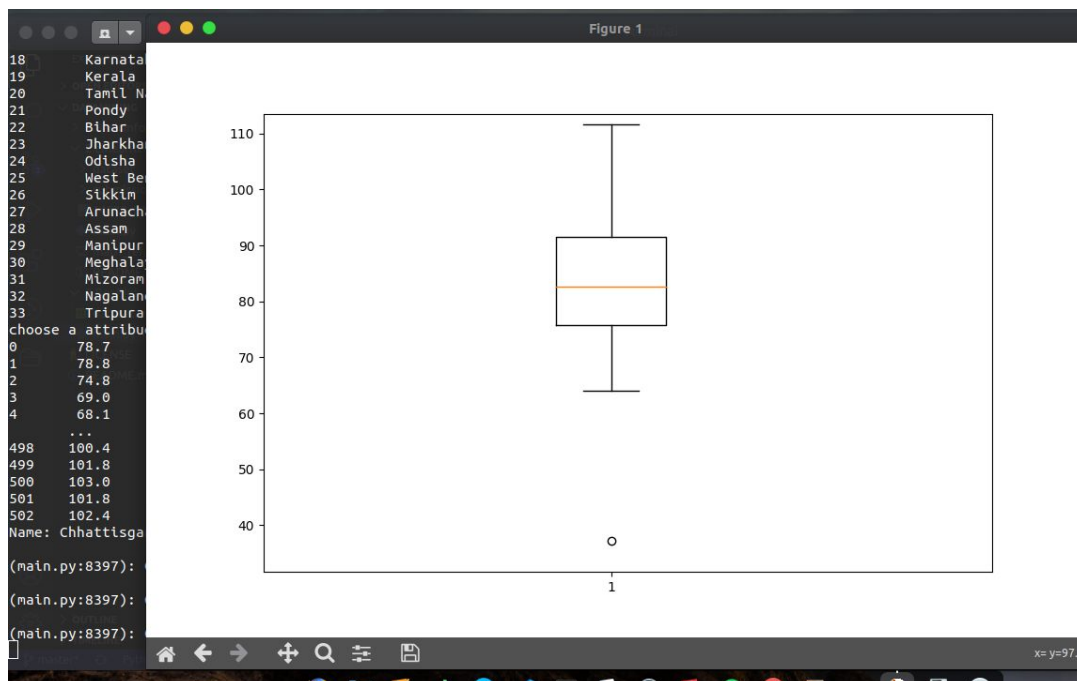
Screenshots

```
$python3 main.py
key    Attribute
1      Punjab
2      Haryana
3      Rajasthan
4      Delhi
5      UP
6      Uttarakhand
7      HP
8      J&K
9      Chandigarh
10     Chhattisgarh
11     Gujarat
12     MP
13     Maharashtra
14     Goa
15     DNH
16     Andhra Pradesh
17     Telangana
18     Karnataka
19     Kerala
20     Tamil Nadu
21     Pondy
22     Bihar
23     Jharkhand
24     Odisha
25     West Bengal
26     Sikkim
27     Arunachal Pradesh
28     Assam
29     Manipur
30     Meghalaya
31     Mizoram
```

```
Applications Places Terminal
15     DNH
16     Andhra Pradesh
17     Telangana
18     Karnataka
19     Kerala
20     Tamil Nadu
21     Pondy
22     Bihar
23     Jharkhand
24     Odisha
25     West Bengal
26     Sikkim
27     Arunachal Pradesh
28     Assam
29     Manipur
30     Meghalaya
31     Mizoram
32     Nagaland
33     Tripura

N = 503
IQR = 8.0
MIN = 25.85
Q1 = 37.85
Q2 = 41.3
Q3 = 45.85
MAX = 57.85

(main.py:3845): Gtk-WARNING **: 19:01:15.883: *
```



References

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
https://en.wikipedia.org/wiki/Box_plot