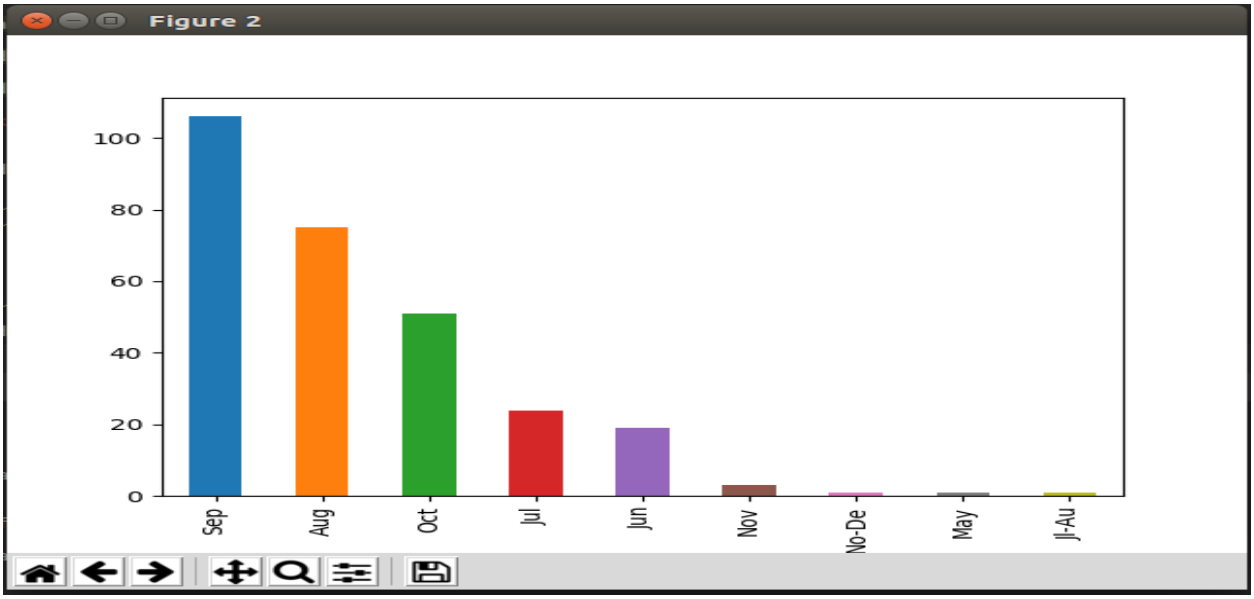
Data Analysis and Visualization on Hurricanes Data

a. A

Visualization

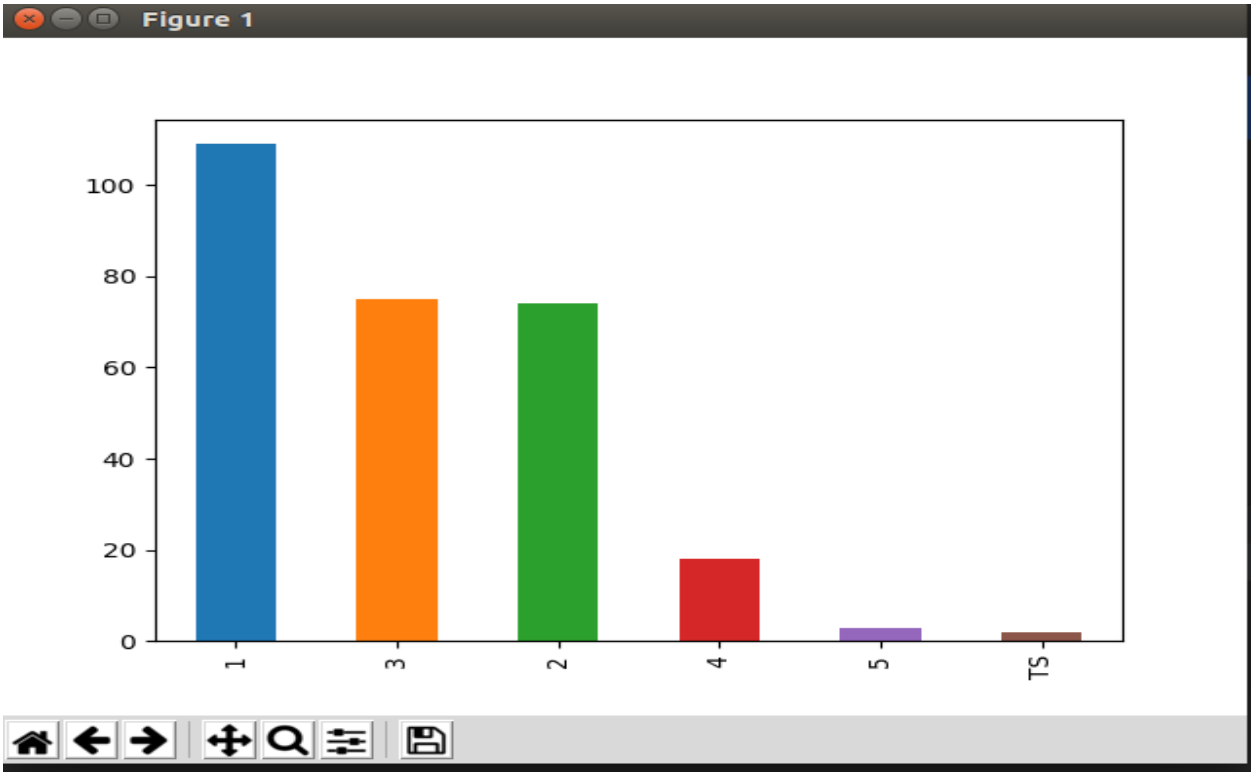
Number of Hurricanes month-wise

Months

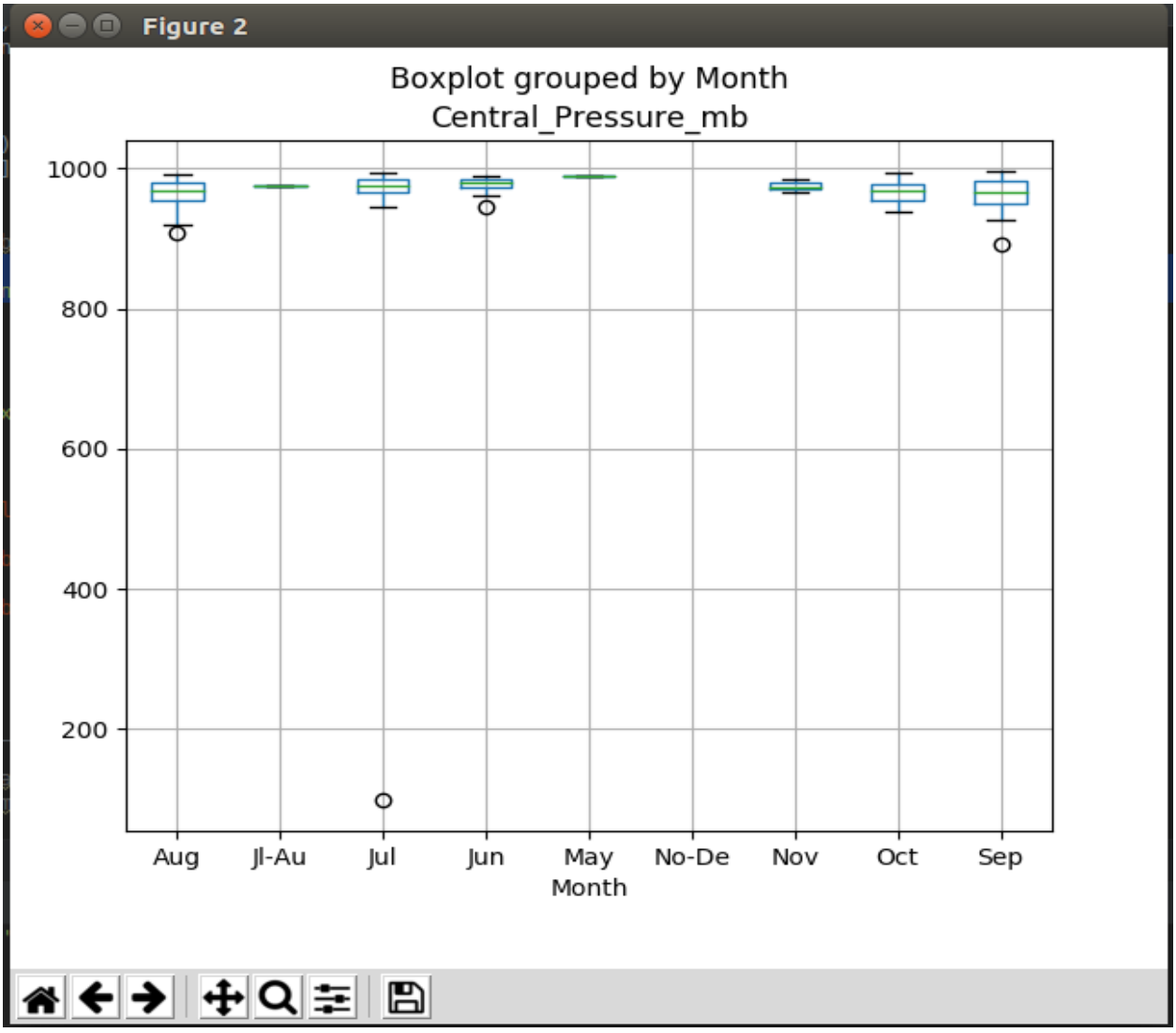


Number of Hurricanes category wise

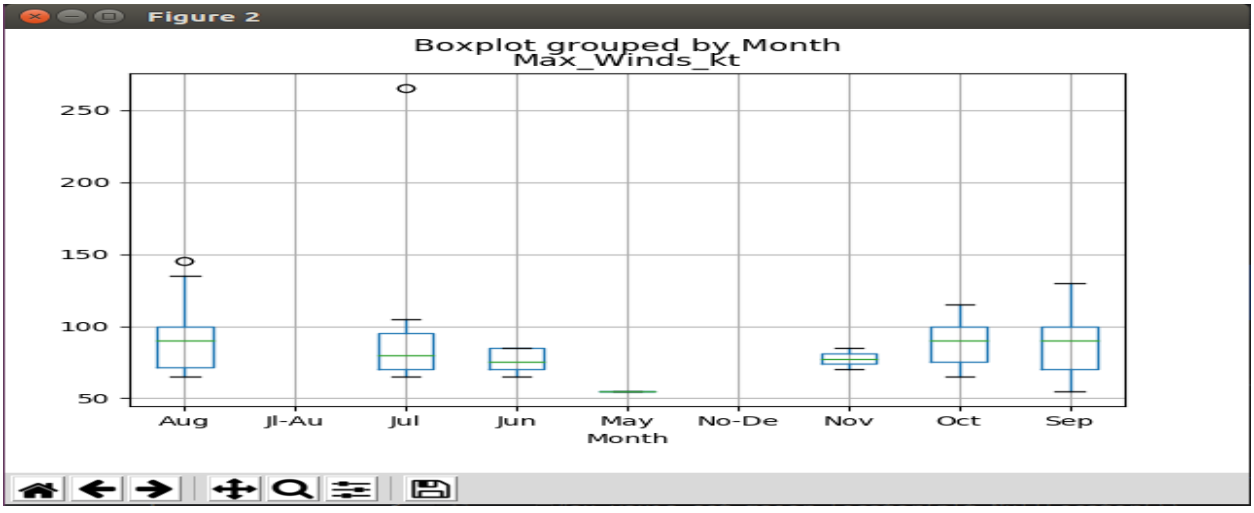
Highest Category



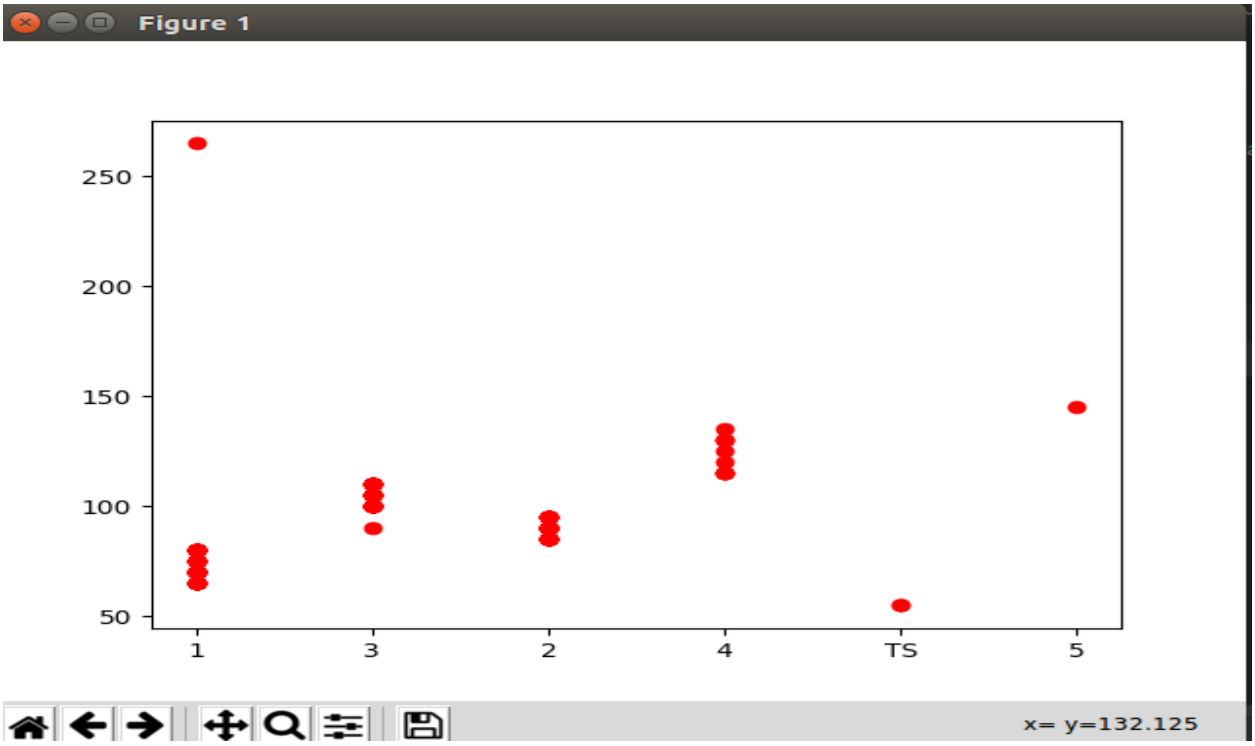
Pressure (through boxplot, understanding outliers)



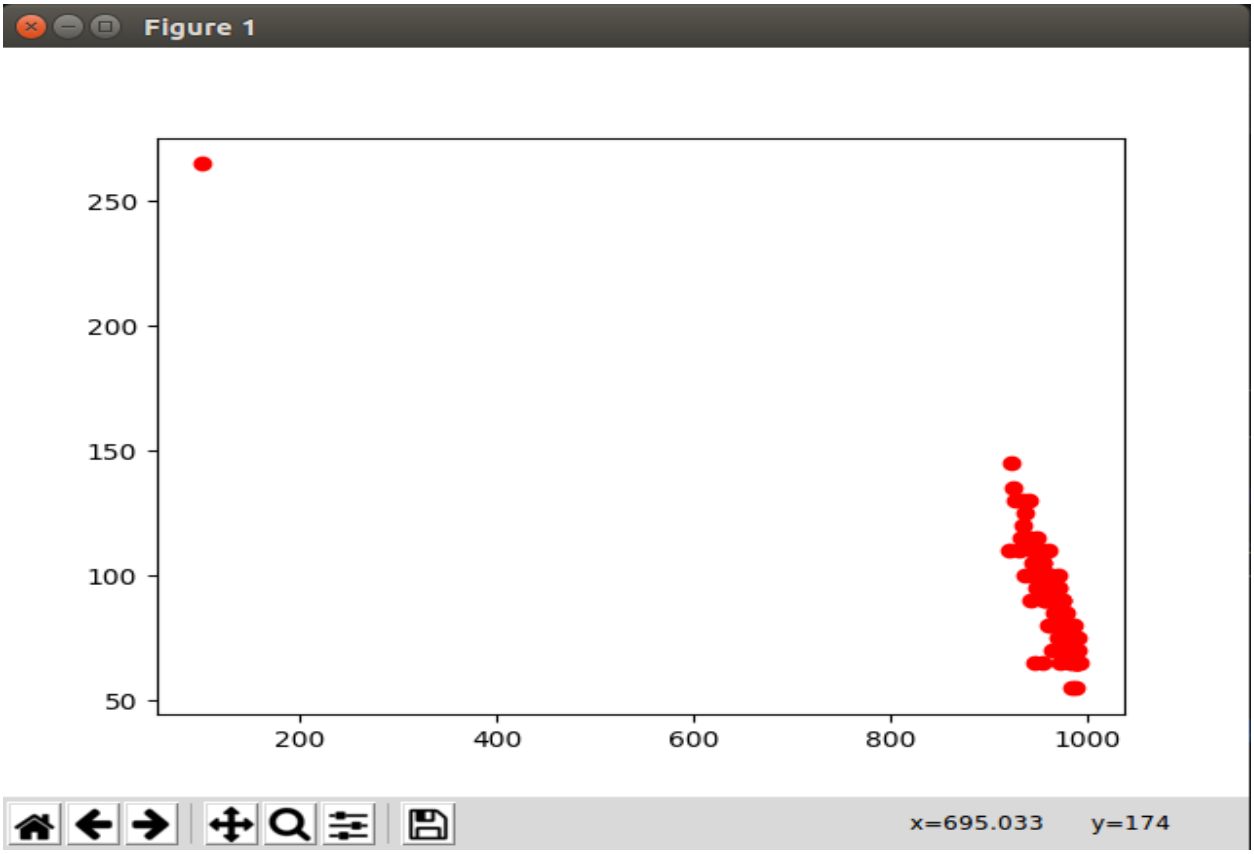
Max_Winds_Kt (through boxplot, understanding outliers)



b. In this figure Highest_Category is on the x-axis and Max_Winds_kt is on Y-axis. Max_Winds_kt is increasing with the category however, there are few exceptions and specially the major exception is where Max_Winds_kt is more than 250 and category is 1.



The Central_Pressure_mb is on the x-axis and Max_Winds_kt is on y-axis. There is gradual increase in Central_Pressure_mb as Max_Winds_kt decreases. The change is very small, and all the data lies between 900 and 1000 on the x-axis. In addition, there is one exception where Central_Pressure_mb is less then 100 and Max_Winds_kt is above 250.



c. For dealing with the missing data in part(b). I created two separate tables for the two graphs and removed the whole row where there was a null value. This was to ensure that only those null values are removed from the table which may affect that particular graph.

Code

The following code was executed in smaller chunks. Selecting part of script then right click and pressing “Execute Selection in Console.

```
import pandas as pd
import matplotlib.pyplot as plt


data = pd.read_csv("Hurricane.csv")

data.head()

#For Months
fig, ax = plt.subplots()
data['Month'].value_counts().plot(ax=ax, kind='bar')
plt.show()

#For Highest_Category
fig, ax = plt.subplots()
data['Highest_Category'].value_counts().plot(ax=ax, kind='bar')
plt.show()

#For Central_Pressure_mb
plt.figure()
data.boxplot(column='Central_Pressure_mb',by='Month')
plt.show()

#For Max_Winds_kt
plt.figure()
data.boxplot(column='Max_Winds_kt',by='Month')
plt.show()


print(data.describe(include='all').transpose())

datab1 = data.dropna(subset=['Highest_Category', 'Max_Winds_kt'])

datab2 = data.dropna(subset=['Central_Pressure_mb', 'Max_Winds_kt'])

plt.plot(x='Months')
```

```
        plt.plot(datab1.Highest_Category,datab1.Max_Winds_kt, 'ro')
        plt.show()

        plt.plot(datab2.Central_Pressure_mb,datab2.Max_Winds_kt, 'ro')
plt.show()
```