

VELLORE INSTITUTE OF TECHNOLOGY, VELLORE

FALL SEMESTER 2020-21

DATA VISUALIZATION

FACULTY: Dr. SIVANESAN

S.

SLOT: L13 + L14



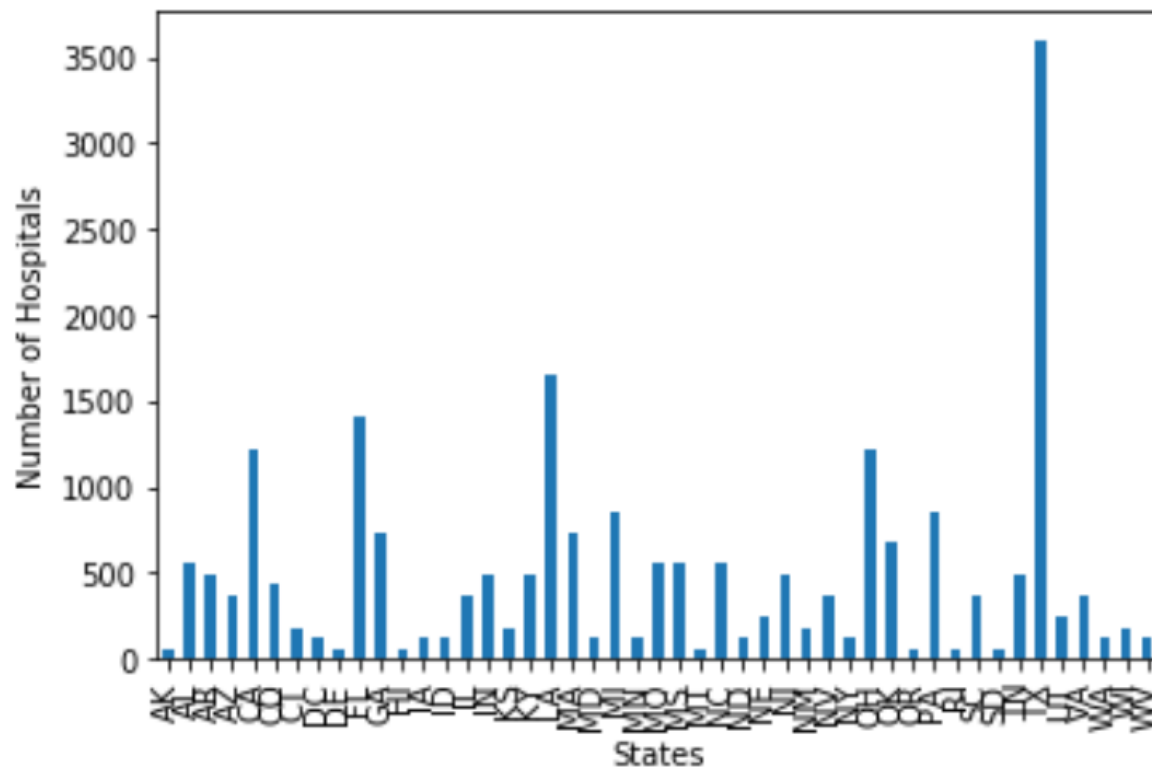
LAB ASSIGNMENT 2

NAME : ANKIT CHAWLA
 REGISTRATION NO. : 18BCE0830

CREATING GRAPHS

The following section contains all the graphs created along with their code.

- Bar Chart:



(Number of Hospitals VS States)

Code for the above graph:

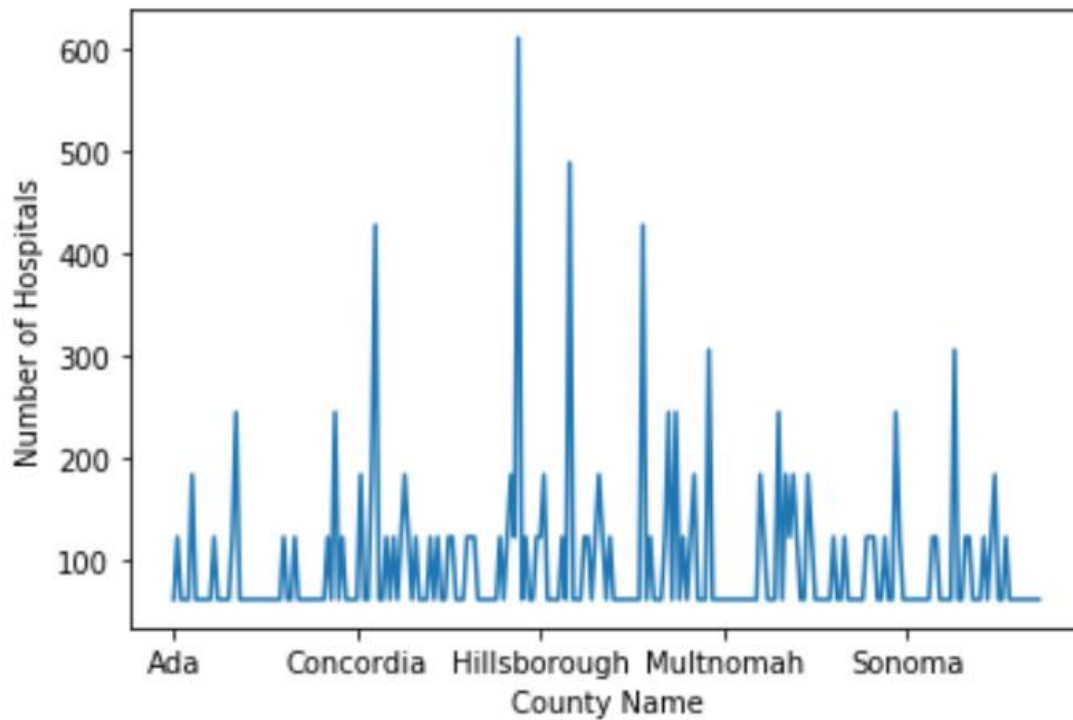
```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import sklearn

'''18BCE0830 Ankit Chawla'''

Out[1]: '18BCE0830 Ankit Chawla'

In [17]: df = pd.read_csv(r'D:\data\Long term\Long-Term Care Hospital - Provider data.csv')
ax = df.groupby("State").count().plot.bar(x= "Number of Hospitals" ,y = "States")
ax.set_ylabel("Number of Hospitals")
ax.set_xlabel("States")
ax
```

- Line Chart:



(Number of Hospitals VS County Name)

Code for the graph above:

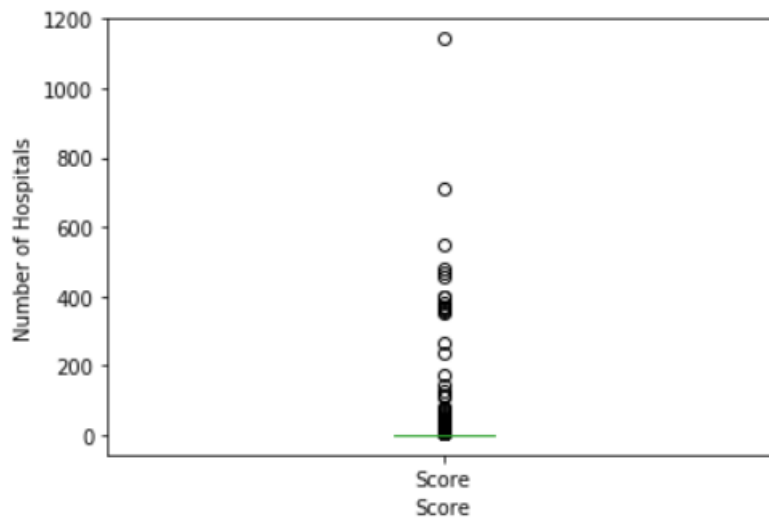
```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import sklearn

'''18BCE0830 Ankit Chawla'''

Out[1]: '18BCE0830 Ankit Chawla'

In [32]: df = pd.read_csv(r'D:\D_data\Long term\Long-Term Care Hospital - Provider data.csv')
ax = df.groupby("County Name")["County Name"].count().plot.line()
ax.set_ylabel("Number of Hospitals")
ax.set_xlabel("County Name")
ax
```

- Box Plot:



(Number of Hospitals VS Score)

Code for the above graph:

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import sklearn

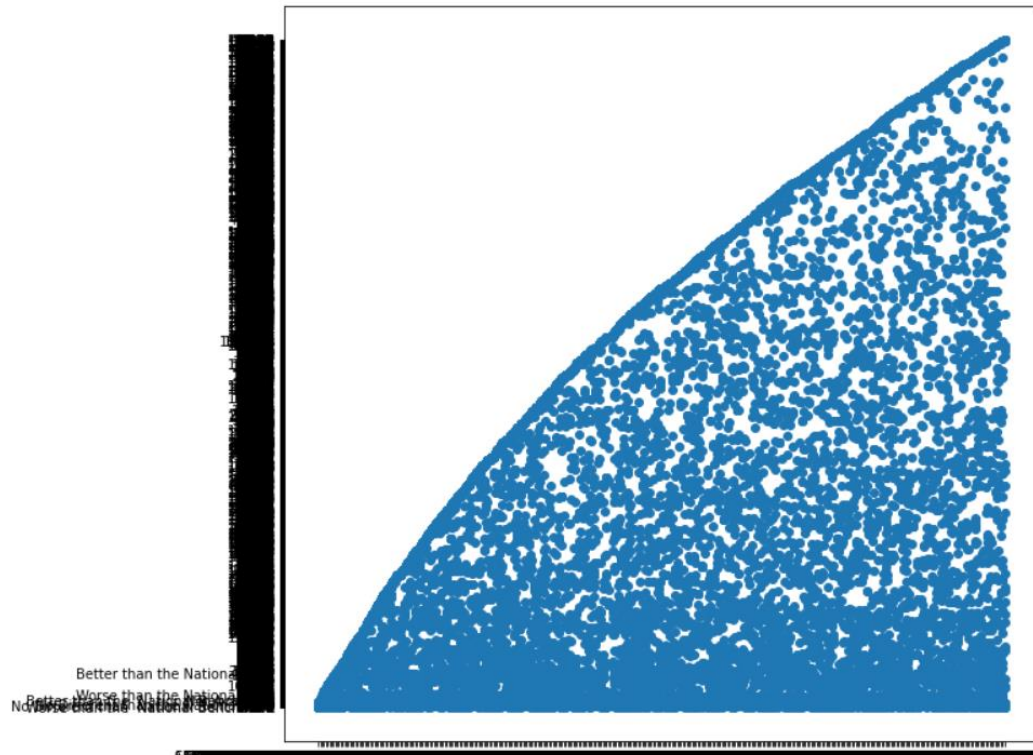
'''18BCE0830 Ankit Chawla'''

Out[1]: '18BCE0830 Ankit Chawla'

In [42]: df = pd.read_csv(r'D:\D_data\Long term\Long-Term Care Hospital - Provider data.csv')
ax = df.groupby("Score").count().plot.box()
ax.set_ylabel("Number of Hospitals")
ax.set_xlabel("Score")
ax
```

- **Scatter Plot:**

✎ <matplotlib.collections.PathCollection at 0x7f215c5d0550>



Code for the graph:

```
[2] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import sklearn
import seaborn as sns

'''18BCE0830 Ankit Chawla'''

/usr/local/lib/python3.6/dist-packages/statsmodels/tools/_testing.py:19: FutureWarning: pandas.util.testing is deprecated. Use the functions in the public API at pandas.testing instead
import pandas.util.testing as tm
'18BCE0830 Ankit Chawla'
```

```
[3] from google.colab import files
data_to_load = files.upload()
```

Choose Files

- Long-Term C...er data.csv
- Long-Term Care Hospital - Provider data.csv(application/vnd.ms-excel) - 4350767 bytes, last modified: 5/7/2020 - 100% done

Saving Long-Term Care Hospital - Provider data.csv to Long-Term Care Hospital - Provider data (2).csv

```
[4] import io
df = pd.read_csv(io.BytesIO(data_to_load['Long-Term Care Hospital - Provider data.csv']))
```

```
[5] from pandas import DataFrame
name = df["Facility Name"]
score = df["Score"]
df2 = DataFrame(index=score, columns=name)
```

```
colors = (0,0,0)
area = np.pi*3
plt.figure(figsize=(10, 10))
plt.scatter(name, score)
```

• Heat Map:

Code:

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import sklearn
import seaborn as sns

'''18BCE0830 Ankit Chawla'''

Out[1]: '18BCE0830 Ankit Chawla'
```

```
In [2]: df = pd.read_csv(r'D:\D_data\Long term\Long-Term Care Hospital - Provider data.csv')
```

```
In [3]: from pandas import DataFrame
```

```
In [6]: df.shape
```

```
Out[6]: (22570, 15)
```

```
In [10]: name = df["Facility Name"]
score = df["Score"]
```

```
In [4]: from pandas import DataFrame
name = df["Facility Name"]
score = df["Score"]
df2 = DataFrame(index=name, columns=score)
```

```
In [ ]: sns.heatmap(df2, annot = True)
```

```
In [5]: df.describe()
```

ANALYZING THE RESULTS

The following section contains the analysis of the results shown above.

1. **Effectiveness:** The effectiveness of the following graphs as analyzed by me is written down below.
 - a. **Bar Chart:** The Bar chart is a very effective and simple visualization method in my opinion, by making the Bar chart of the Number of Hospitals VS States makes it easy to understand which state has the maximum and the minimum number of Hospitals and after looking at that information we can better understand which states are in more need of new hospitals on a priority basis and which States can function just fine without any new hospitals made.
 - b. **Line Chart:** The Line chart is less effective than the Bar chart in my opinion so representing Number of Hospitals VS States would not have been very effective when plotting it on a Line Chart. On the other hand, the chart did seem to be very effective when plotting the Number of Hospitals by County, in such case when we know the details of one State, we can see which parts of the state are functioning well in terms of creating new hospitals. This way we can see which states have a trend of having less hospitals and which have a trend of having more hospitals from the principle city or the capital.
 - c. **Box Plot:** The Box plot seems to be very effective in finding outliers in the database and as a result they have been used to analyze the scores of the hospitals. We can easily make out which hospitals are performing better than the others and what is the average score of the hospitals.

- d. **Scatter Plot:** The scatter plot made was very effective in showing the no linear pattern that exists in the score value. It makes it easier to make out the maximum and the minimum values in the scores. The observation is made straightforward by the plot.
- e. **Heat Map:** The heatmap provides good visual paths so that the reader can see the numeric values properly. It is effective in providing the visual representation of numeric data in a clearer way.

2. **Completeness:** The following section analyses the completeness of the graphs.

- a. **Bar Plot:** The Bar plot seems to show the complete picture of the quantity it was meant to represent to some degree. We can clearly see what the values are and how they compare with each other. It should be pointed out that even after being complete to some degree, it fails to expose key assumptions, causes and patterns.
- b. **Line Plot:** The Line plot was meant to show trends in the data above and it show the complete information based on the data based on which we can understand the trend In the data.
- c. **Box Plot:** The Box plot shows less complete information as it can show the outliers but it cannot show the trends in the data.
- d. **Scatter Plot and Heat Map:** The Scatter plot does portray the complete information we need from it but does not give the exact extent of the correlation the heatmap provides complete visual representation.

3. **Visibility:** The visibility of information of the data is pretty good in all the graphs and they all can show information in a clear way in case the number of data points is low and they have distinct values.
4. **Accuracy:** The Data represented in the following graphs is accurate and after meticulous analysis we can understand that they are representing the correct information, accurately.
5. **Precision:** The graphs represent exactly what we wanted them to represent and hence it is very precise.
6. **Constraints:**
 - a. **Bar Plot:** The Bar plot, as pointed out earlier does fail to expose key assumptions, impacts and patterns and as a result we can understand that it can be manipulated to show and give a false impression.
 - b. **Line Plot:** A wide range of data, as was in this dataset, is challenging to be represented.
 - c. **Box Plot:** The most notable constraint of the Box plot is that it is as such that the mean of the data is difficult to locate. Even the outliers maybe exceptional cases only or may show something more than that. The box plot can be confusing for some people.
 - d. **Scatter plot and Heat Map:** Scatter plot cannot give the entire extent of correlation and due to human visual limitations, some people can have a problem interpreting heatmaps.