Data Visualization

matplotlib

Below are some of the important plots used for visualizing the data. Data Visualization is one of the core skills in Data Science.

I used Matplotlib, Seaborn & Pandas in Python.

1. Bar Chart
2. Stacked Bar Chart
3. Line Chart
4. Histogram
   4.1 Categorical
   4.2 Continuous
5. Pie Chart
6. Donut Chart
7. Scatter Plot
8. Pair Plot
9. Box Plot

Initially do – pip install and import the packages in python -

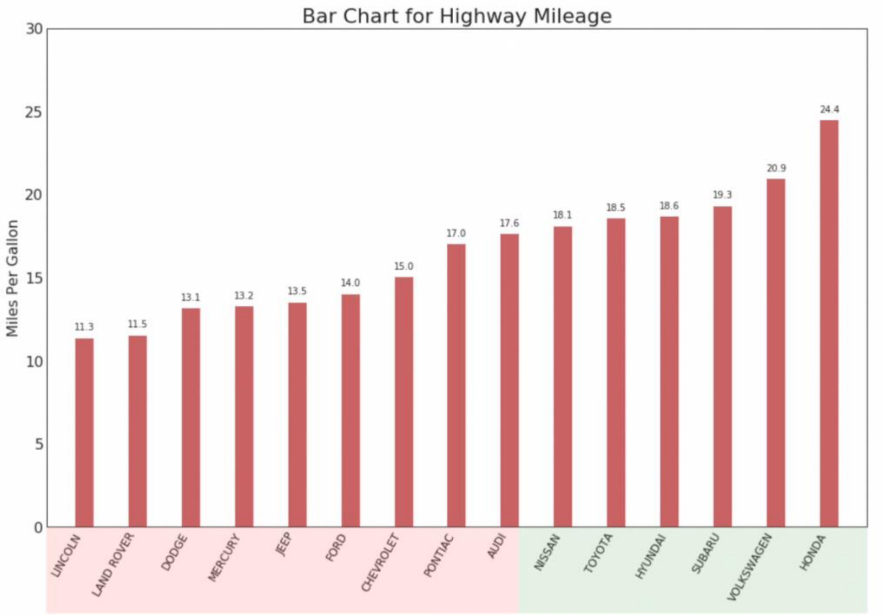
```
pip install matplotlib
```

```
import pandas as pd
```

1. **Bar Chart**:

Bar chart conveys the rank order of the items effectively.

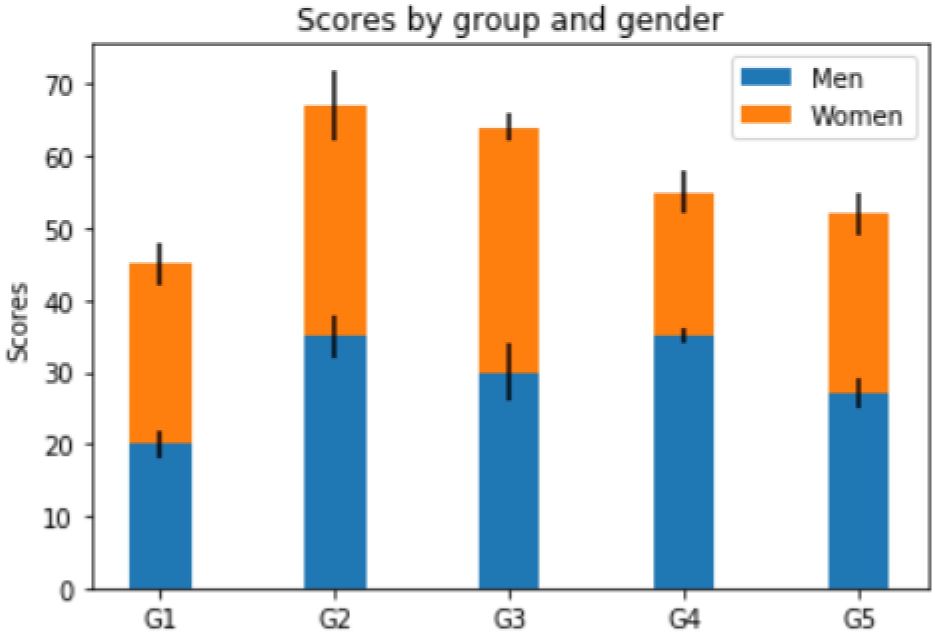
```
ax = plt.subplots(figsize=(16,10), facecolor='white', dpi= 80)

ax.vlines(x=df.index, ymin=0, ymax=df.cty, color='firebrick', alpha=0.7, linewidth=20)
```



2. **Stacked Bar Chart**:

Stacked Bar Chart is multiple datasets on top of each other in order to show how the larger category is divided into the smaller categories and their relations to the total amount

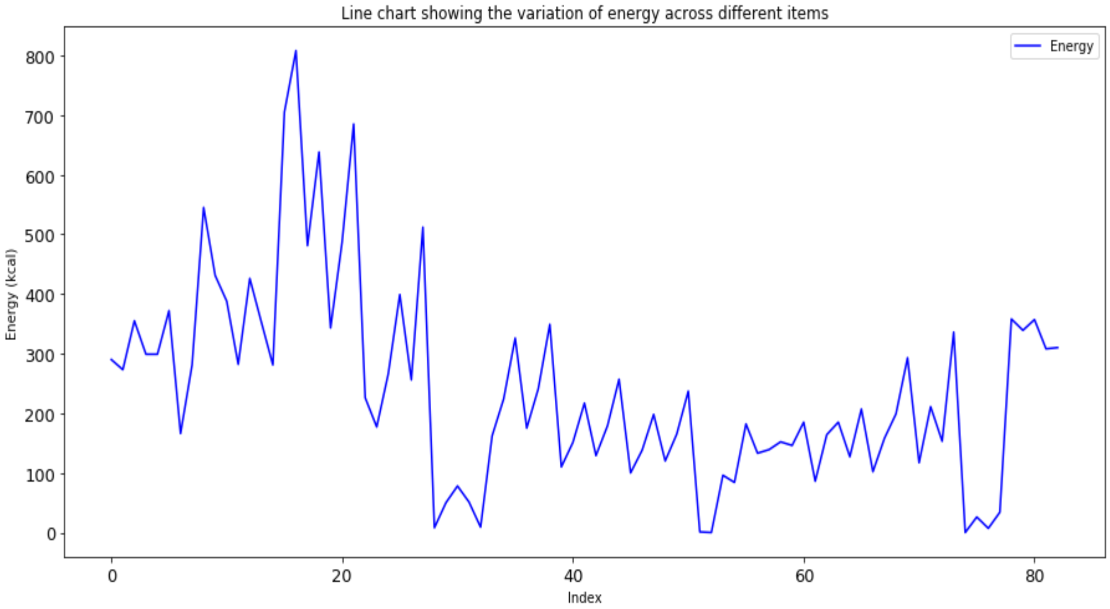
```
fig,        ax = plt.subplots()

          ax.set_ylabel('Scores')

          ax.set_title('Scores by group and gender')
```


Scores by group and gender
Scores
0
10
20
30
40
50
60
70
G1
G2
G3
G4
G5
Men
Women

3. Line Chart:

Line chart can pass over any number of individual values, making it the tool of first choice for distributions with many unique values or categories

```
df_eng[['Energy']].plot(kind='line', figsize=(15,7), color='blue', fontsize=13)
```

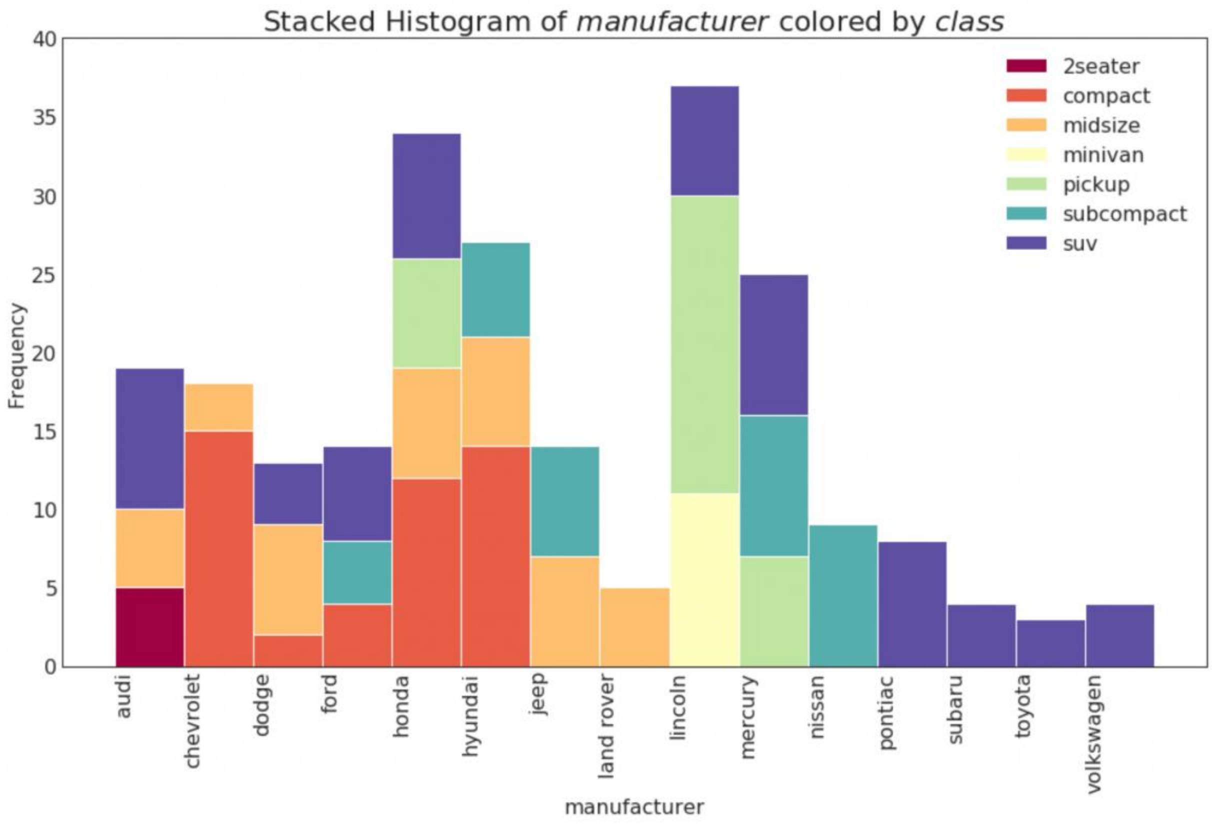


4. Histogram:

4.1 Histogram for Categorical Variable

Histogram of a categorical variable shows the frequency distribution of a that variable

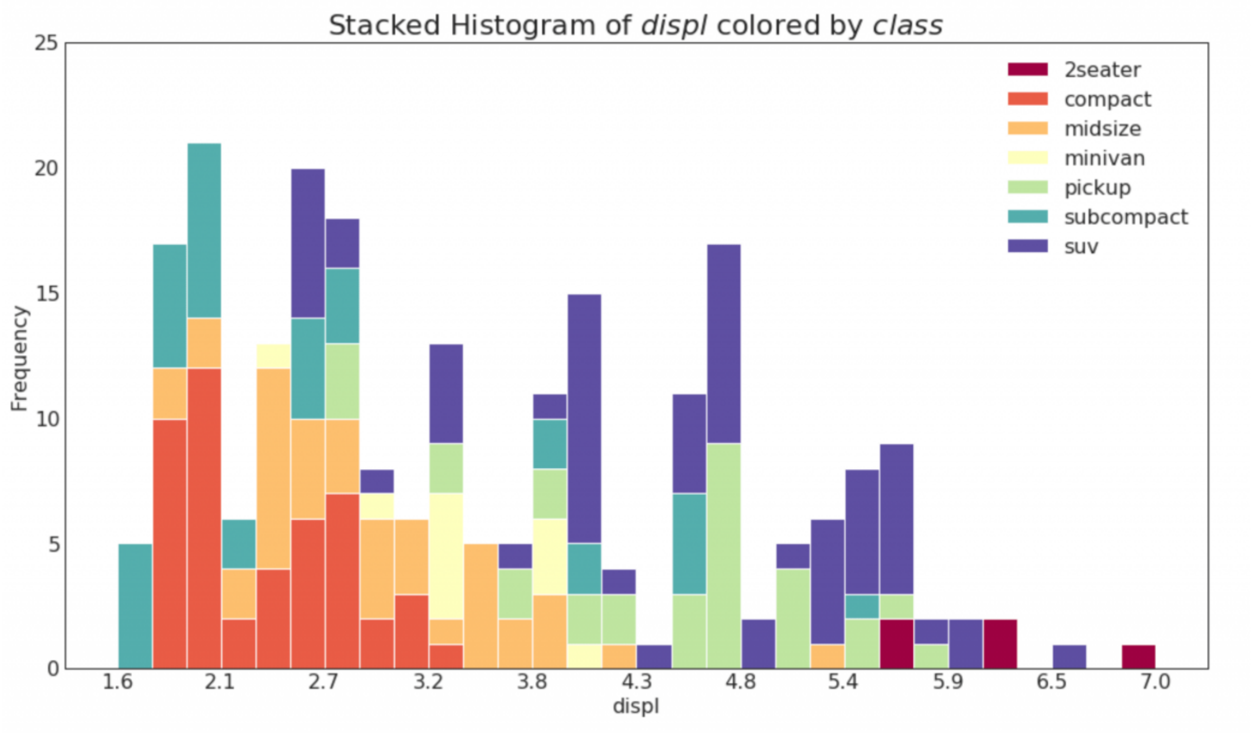
```
plt.figure(figsize=(16,9), dpi= 80)

colors = [plt.cm.Spectral(i/float(len(vals)-1)) for i in range(len(vals))]

n, bins, patches = plt.hist(vals, df[x_var].unique().__len__(), stacked=True, density=False,

color=colors[:len(vals)          ])
```



4.2 Histogram for Continuous Variable

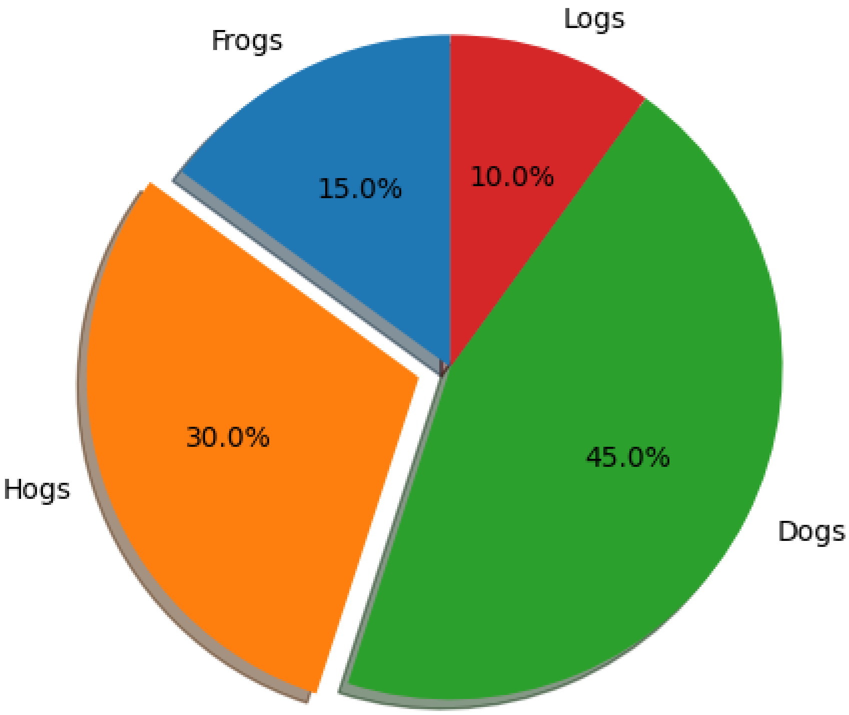
```
plt.figure(figsize=(16,9), dpi= 80)


colors = [plt.cm.Spectral(i/float(len(vals)-1)) for i in range(len(vals))]
```

```
n, bins, patches = plt.hist(vals, df[x_var].unique().__len__(), stacked=True, density=False,
color=colors[:len(vals)
```



5. Pie Chart:

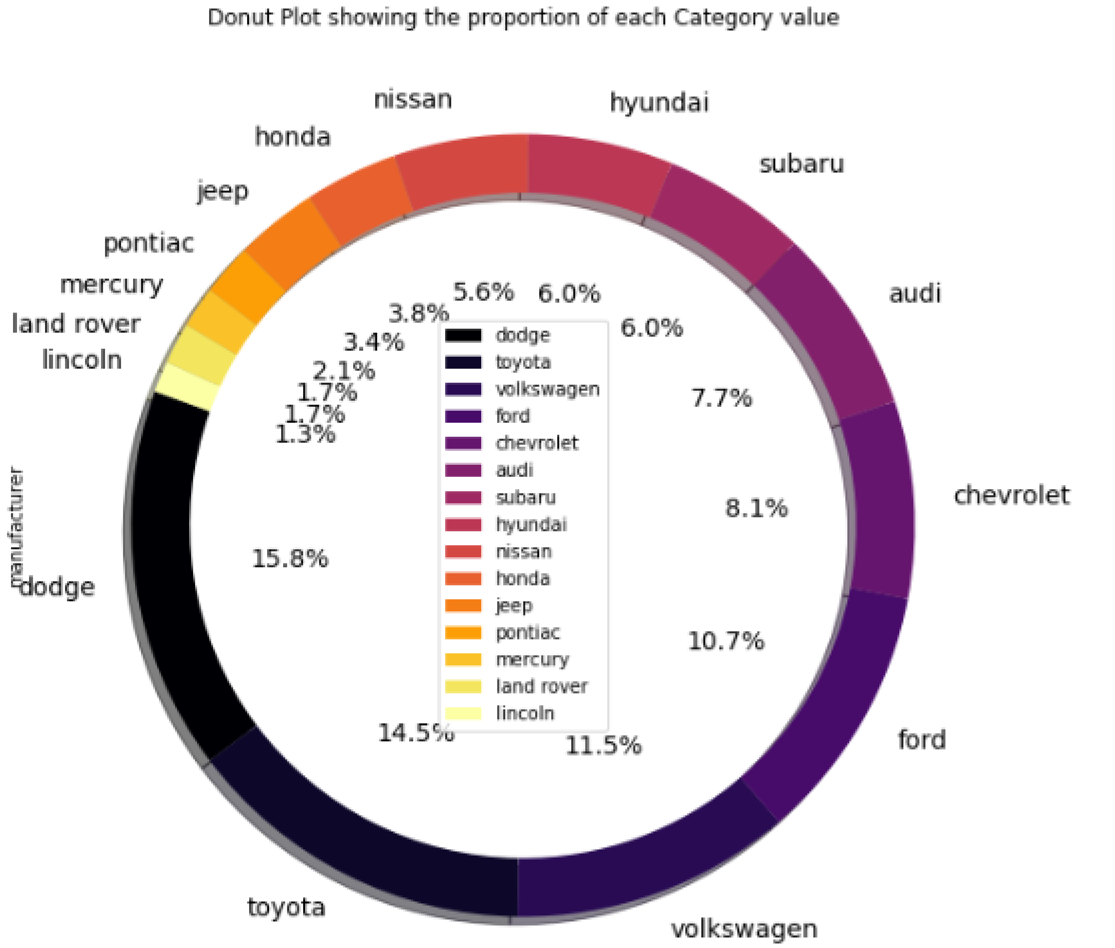
```
fig1, ax1 = plt.subplots()

ax1.pie(sizes, explode=explode, labels=labels, autopct='%1.1f%%',

        shadow=True, startangle=90)
```



6. Donut Chart:

Donut charts provide a better data intensity ratio to standard pie charts since the blank center can be used to display additional, related data.

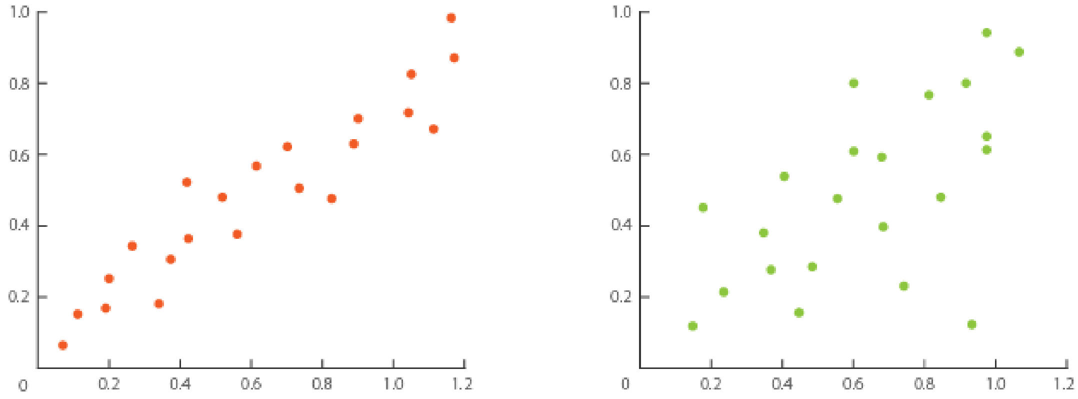
```
df_raw['manufacturer'].value_counts().plot(kind='pie', fontsize=14, autopct='%3.1f%%',
wedgeprops=dict(width=0.15), shadow=True, startangle=160, figsize=(10,10), cmap='inferno',
legend=True)
```



7. Scatter Plot:

Scatter plot is a classic and fundamental plot used to study the relationship between two variables. It’s the simplest bivariate plot is the scatter plot.

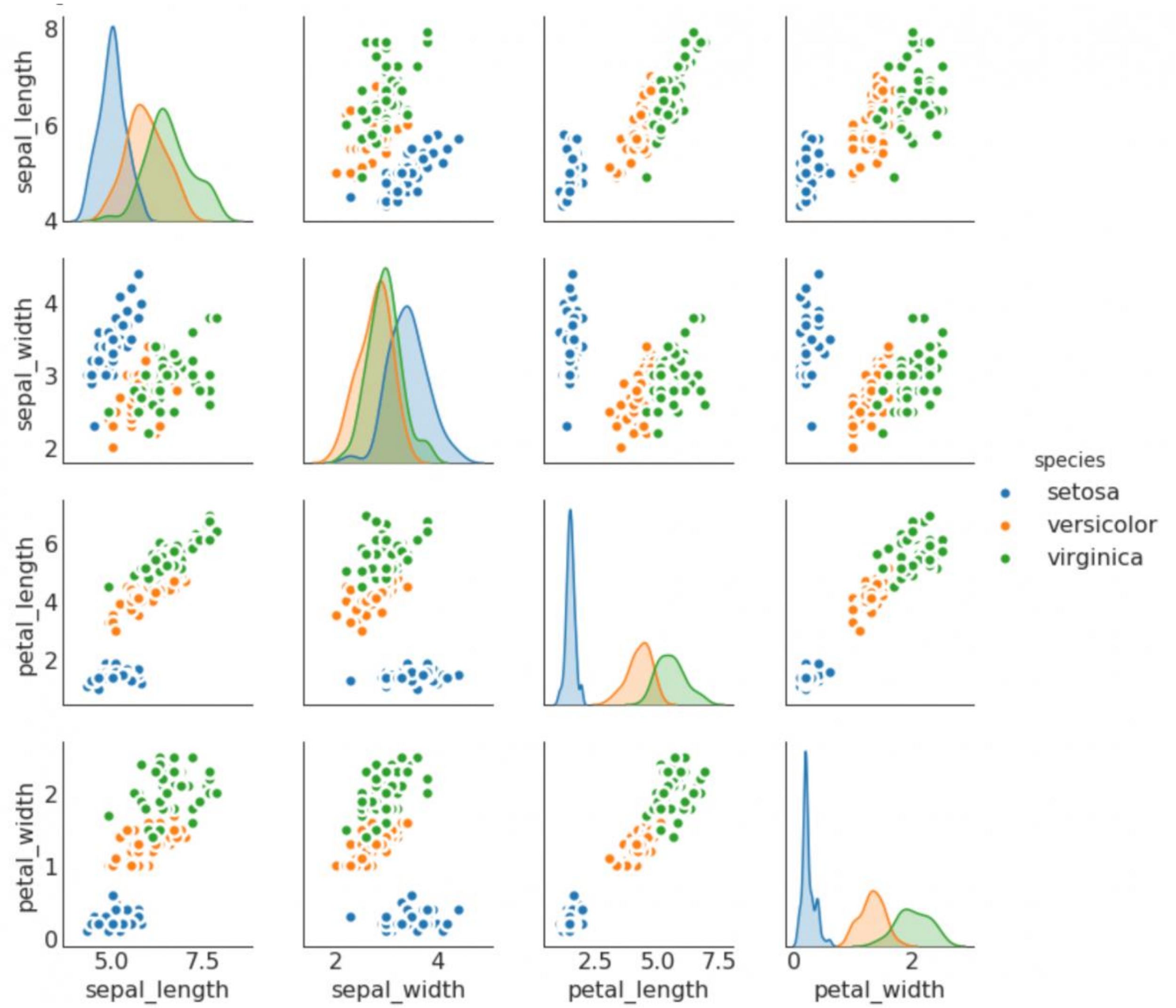
```
df.plot.scatter()
```



8. Pair Plot:

Pairwise plot is a favorite in exploratory analysis to understand the relationship between all possible pairs of numeric variables.

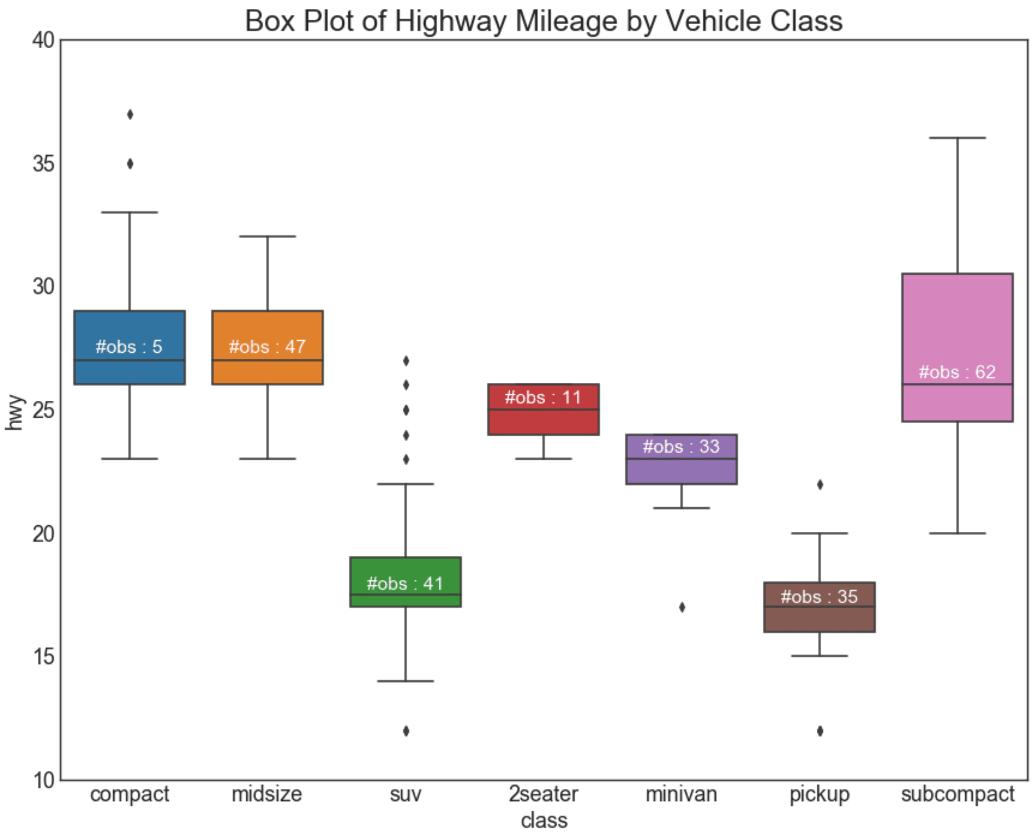
```
sns.pairplot(df, kind="reg", hue="species")
```



9. Box Plot:

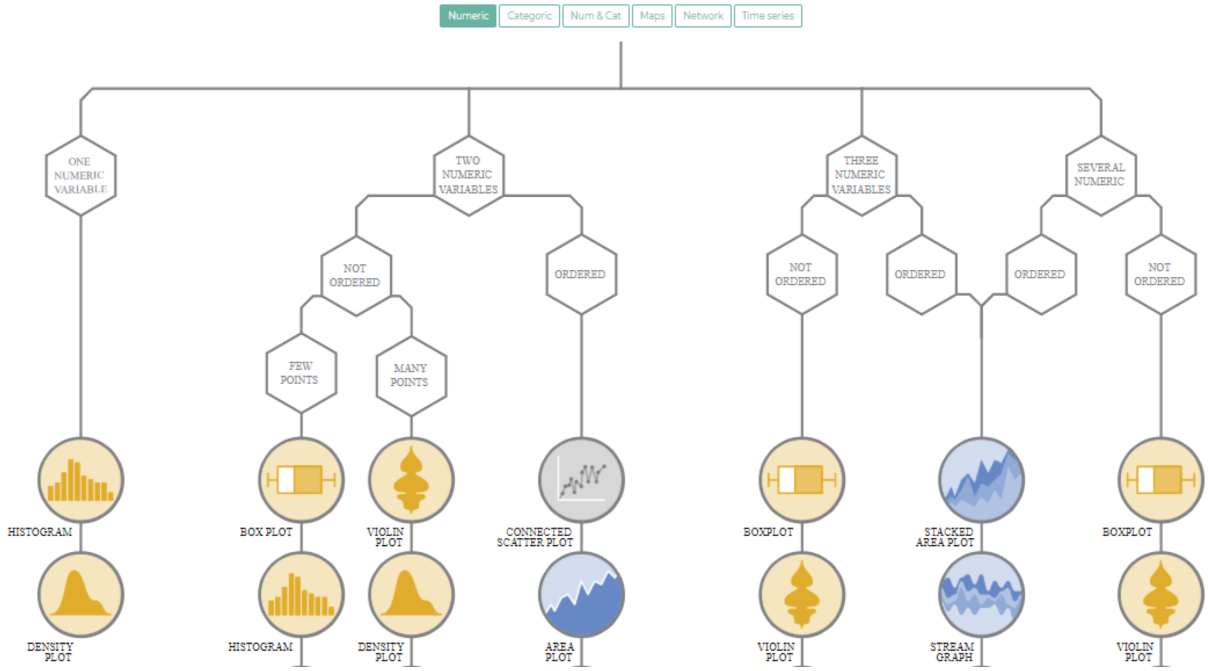
Box plots are a great way to visualize the distribution, keeping the median, 25th 75th quartiles and the outliers in mind.

```
sns.boxplot(x='class', y='hwy', data=df, notch=False)
```

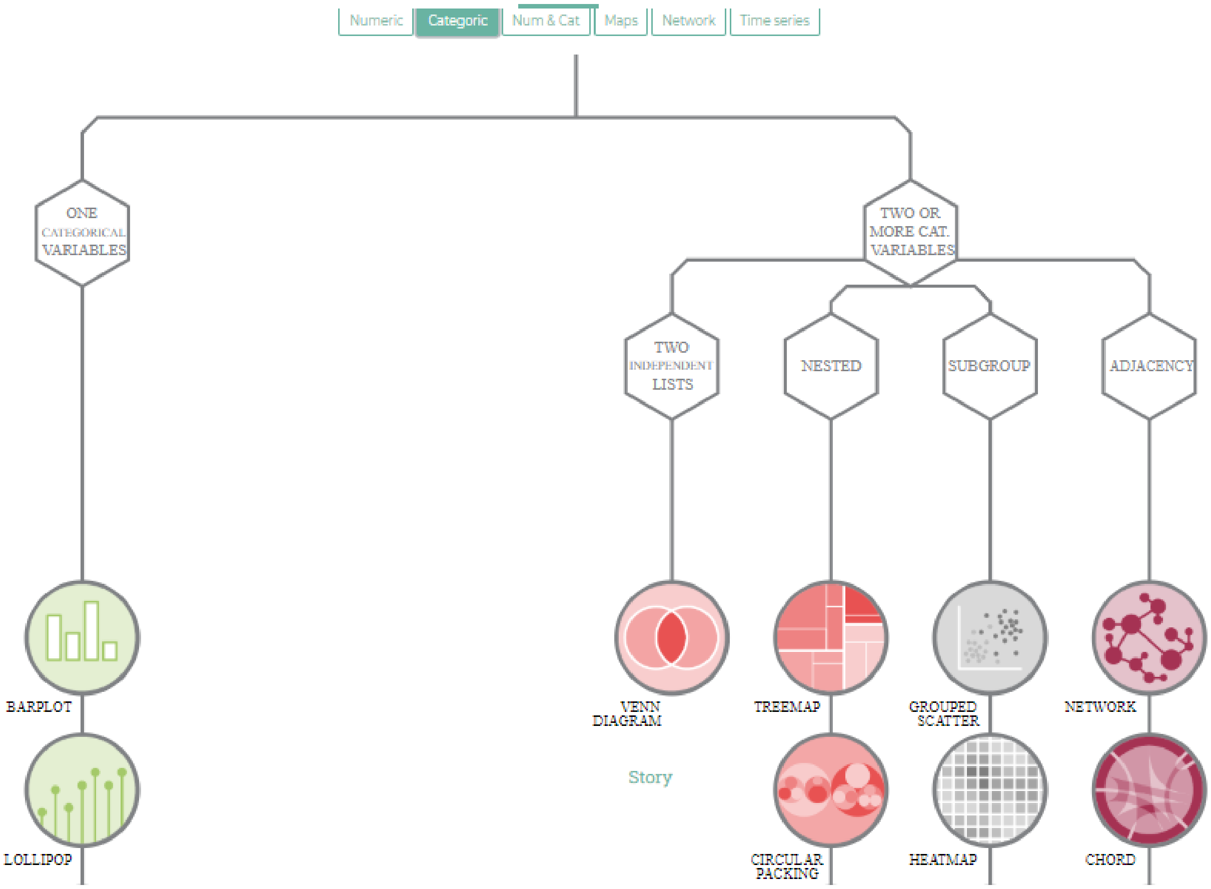


Below are different kinds of plots which used for Numerical and Categorial data:

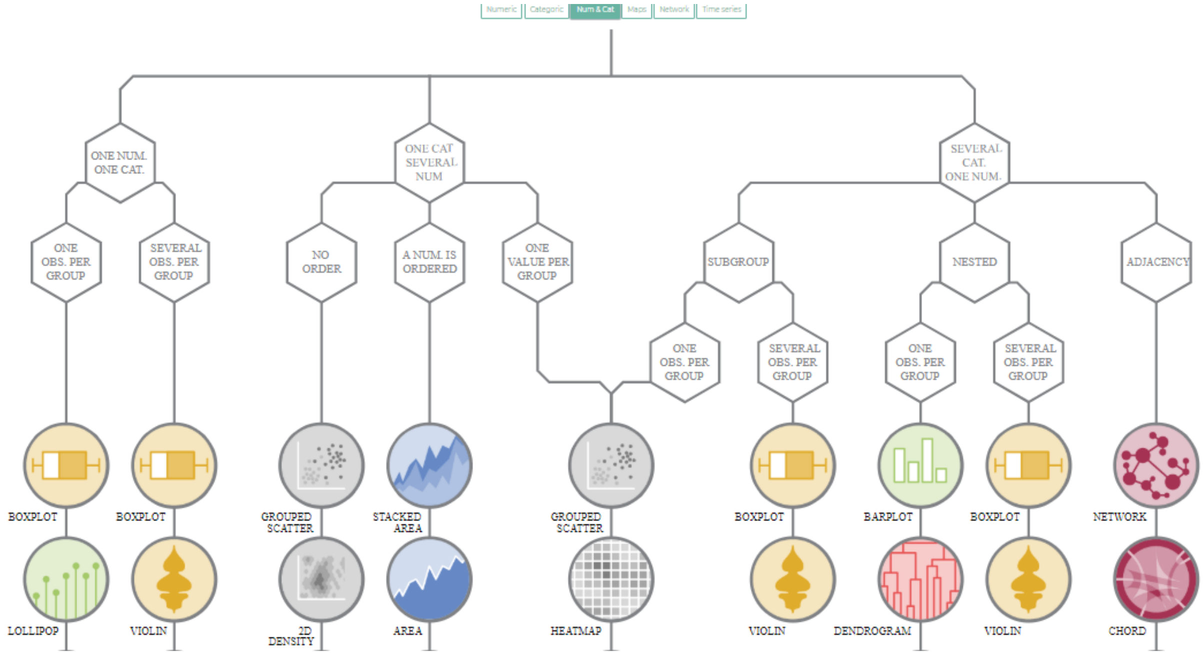
Numerical:



Categorial:


Numeric
Categoric
Num & Cat
Maps
Network
Time series
ONE CATEGORICAL VARIABLES
TWO OR MORE CAT. VARIABLES
TWO INDEPENDENT LISTS
NESTED
SUBGROUP
ADJACENCY
BARPLOT
LOLLIPOP
VENN DIAGRAM
Story
TREEMAP
CIRCULAR PACKING
GROUPED SCATTER
HEATMAP
NETWORK
CHORD

Numerical and Categorial:


Numeric
Categoric
Num & Cat
Maps
Network
Time series
ONE NUM. ONE CAT.
ONE CAT SEVERAL NUM
SEVERAL CAT. ONE NUM.
ONE OBS. PER GROUP
SEVERAL OBS. PER GROUP
NO ORDER
A NUM. IS ORDERED
ONE VALUE PER GROUP
SUBGROUP
NESTED
ADJACENCY
ONE OBS. PER GROUP
SEVERAL OBS. PER GROUP
ONE OBS. PER GROUP
SEVERAL OBS. PER GROUP
BOXPLOT
BOXPLOT
GROUPED SCATTER
STACKED AREA
GROUPED SCATTER
BOXPLOT
BARPLOT
BOXPLOT
NETWORK
LOLLIPOP
VIOLIN
2D DENSITY
AREA
HEATMAP
VIOLIN
DENDROGRAM
VIOLIN
CHORD