

Data Visualization

Segment 1 - Creating standard plots (line, bar, pie)

! pip install Seaborn # To Install Seaborn Package

```
import numpy as np
```

```
from numpy.random import randn
```

```
import pandas as pd
```

```
from pandas import Series, DataFrame
```

```
*****
```

```
import matplotlib.pyplot as plt
```

```
from matplotlib import rcParams
```

```
import seaborn as sb
```

```
*****
```

```
%matplotlib inline
```

```
rcParams['figure.figsize'] = 5, 4
```

```
sb.set_style('whitegrid')
```

```
*****
```

Creating a line chart from a list object

Plotting a line chart in matplotlib

```
x = range(1,10)
```

```
y = [1,2,3,4,0,4,3,2,1]
```

```
plt.plot(x, y)
```

```
*****
```

Plotting a line chart from a Pandas object

```
address = 'mtcars.csv'
```

```
cars = pd.read_csv(address)
```

```
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
```

```
mpg = cars['mpg']
```

```
mpg.plot()
```

```
*****
```

```
df = cars[['cyl', 'wt', 'mpg']]
```

```
df.plot()
```

```
*****
```

Creating bar charts

Creating a bar chart from a list

```
plt.bar(x, y)
```

```
*****
```

Creating bar charts from Pandas objects

```
mpg.plot(kind='bar')
```

```
*****
```

Plotting Horizontal Bars

```
mpg.plot(kind='barh')
```

```
*****
```

Creating a pie chart

```
x = [1,2,3,4,0.5]
```

```
plt.pie(x)
```

```
plt.show()
```

```
*****
```

Saving a plot

```
plt.savefig('pie_chart.jpeg')
```

```
plt.show()
```

```
*****
```

To See Present working Directory

```
%pwd
```

```
*****
```

```
*****
```

Segment 2 - Defining elements of a plot

```
import numpy as np
```

```
from numpy.random import randn
```

```
import pandas as pd
```

```
from pandas import Series, DataFrame
```

```
*****
```

```
import matplotlib.pyplot as plt
```

```
from matplotlib import rcParams
```

```
*****
```

```
%matplotlib inline
```

```
rcParams['figure.figsize'] = 5, 4
```

```
*****
```

Defining axes, ticks, and grids

```
x = range(1,10)
```

```
y = [1,2,3,4,0,4,3,2,1]
```

```
fig = plt.figure()
```

```
ax = fig.add_axes([.1, .1, 1, 1])
```

```
ax.plot(x,y)
```

```
*****
```

```
fig = plt.figure()
```

```
ax = fig.add_axes([.1, .1, 1, 1])
```

```
ax.set_xlim([1,9])
```

```
ax.set_ylim([0,5])
```

```
ax.set_xticks([0,1,2,4,5,6,8,9,10])
```

```
ax.set_yticks([0,1,2,3,4,5])
```

```

ax.plot(x,y)
*****

fig = plt.figure()
ax = fig.add_axes([.1, .1, 1, 1])
ax.set_xlim([1,9])
ax.set_ylim([0,5])
ax.grid()
ax.plot(x, y)
*****

```

Generating multiple plots in one figure with subplots

```

fig = plt.figure()
fig, (ax1, ax2) = plt.subplots(1,2)
ax1.plot(x)
ax2.plot(x,y)
*****
*****

```

Segment 3 - Plot formatting

```

import numpy as np
import pandas as pd
from pandas import Series, DataFrame
import matplotlib.pyplot as plt
from pylab import rcParams
import seaborn as sb
*****

%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
*****

```

Defining plot color

BAR

```

x = range(1, 10)
y = [1,2,3,4,0.5,4,3,2,1]
plt.bar(x, y)
*****

wide = [0.5, 0.5, 0.5, 0.9, 0.9, 0.5, 0.5, 0.5]
color = ['salmon']
plt.bar(x, y, width=wide, color=color, align='center')
*****

```

LINE

```

address = 'mtcars.csv'
cars = pd.read_csv(address)

```

```
cars.columns = ['car_names','mpg','cyl','disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
df = cars[['cyl', 'mpg', 'wt']]
df.plot()
*****

color_theme = ['darkgray', 'lightsalmon', 'powderblue']
df.plot(color=color_theme)
*****
```

PIE

```
z = [1,2,3,4,0.5]
plt.pie(z)
plt.show()
*****

color_theme = ['#A9A9A9', '#FFA07A', '#B0E0E6', '#FFE4C4', '#BDB76B']
plt.pie(z, colors = color_theme)
plt.show()
*****
```

Customizing line styles

```
x1 = range(0,10)
y1 = [10, 9, 8, 7, 6, 5, 4, 3, 2, 1]
plt.plot(x, y)
plt.plot(x1,y1)
*****

plt.plot(x, y, ls = 'steps', lw=5)
plt.plot(x1,y1, ls='--', lw=10)
*****
```

Setting plot markers

```
plt.plot(x, y, marker = '1', mew=20)
plt.plot(x1,y1, marker = '+', mew=15)
*****
*****
```

Segment 4 - Creating labels and annotations

```
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
import matplotlib.pyplot as plt
from pylab import rcParams
import seaborn as sb
%matplotlib inline
rcParams['figure.figsize'] = 8,4
sb.set_style('whitegrid')
```

Labeling plot features

The functional method

```
x = range(1,10)
y = [1,2,3,4,0.5,4,3,2,1]
plt.bar(x,y)
plt.xlabel('your x-axis label')
plt.ylabel('your y-axis label')
```

```
z = [1, 2, 3, 4, 0.5]
veh_type = ['bicycle', 'motorbike', 'car', 'van', 'stroller']
plt.pie(z, labels= veh_type)
plt.show()
```

The object-oriented method

```
address = 'mtcars.csv'
cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
```

```
mpg = cars.mpg
fig = plt.figure()
ax = fig.add_axes([.1, .1, 1, 1])
mpg.plot()
ax.set_xticks(range(32))
ax.set_xticklabels(cars.car_names, rotation=60, fontsize='medium')
ax.set_title('Miles per Gallon of Cars in mtcars')
ax.set_xlabel('car names')
ax.set_ylabel('miles/gal')
```

Adding a legend to your plot

The functional method

```
plt.pie(z)
plt.legend(veh_type, loc='best')
plt.show()
```

The object-oriented method

```
fig = plt.figure()
ax = fig.add_axes([.1,.1,1,1])
mpg.plot()
ax.set_xticks(range(32))
ax.set_xticklabels(cars.car_names, rotation=60, fontsize='medium')
ax.set_title('Miles per Gallon of Cars in mtcars')
```

```

ax.set_xlabel('car names')
ax.set_ylabel('miles/gal')
ax.legend(loc='best')
*****

```

Annotating your plot

```

print(mpg.max())
*****

fig = plt.figure()
ax = fig.add_axes([.1,.1,1,1])
mpg.plot()
ax.set_title('Miles per Gallon of Cars in mtcars')
ax.set_ylabel('miles/gal')
ax.set_ylim([0,45])
ax.annotate('Toyota Corolla', xy=(19,33.9), xytext = (21,35), arrowprops=dict(facecolor='black',
shrink=0.05))
*****

```

Segment 5 - Creating visualizations from time series data

```

import numpy as np
from numpy.random import randn
import pandas as pd
from pandas import Series, DataFrame
import matplotlib.pyplot as plt
from pylab import rcParams
import seaborn as sb
%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
*****

```

The simplest time series plot

```

address = 'Superstore-Sales.csv'
df = pd.read_csv(address, index_col='Order Date', parse_dates=True)
df.head()
*****

df['Order Quantity'].plot()
*****

df2 = df.sample(n=100, random_state=25, axis=0)
plt.xlabel('Order Date')
plt.ylabel('Order Quantity')
plt.title('Superstore Sales')
df2['Order Quantity'].plot()
*****

```

Segment 6 - Constructing histograms, box plots, and scatter plots

```
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
from pandas.tools.plotting import scatter_matrix
import matplotlib.pyplot as plt
from pylab import rcParams
import seaborn as sb
*****

%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
*****
```

Eyeballing dataset distributions with histograms

```
address = 'mtcars.csv'
cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am', 'gear', 'carb']
cars.index = cars.car_names
mpg = cars['mpg']
mpg.plot(kind='hist')
*****

plt.hist(mpg)
plt.plot()
*****

sb.distplot(mpg)
*****
```

Seeing scatterplots in action

```
cars.plot(kind='scatter', x='hp', y='mpg', c='darkgray', s=150)
*****

sb.regplot(x='hp', y='mpg', data=cars, scatter=True)
*****
```

Generating a scatter plot matrix

```
sb.pairplot(cars)
*****

cars_df = pd.DataFrame((cars.ix[:,(1,3,4,6)].values), columns = ['mpg', 'disp', 'hp', 'wt'])
cars_target = cars.ix[:,9].values
target_names = [0, 1]
cars_df['group'] = pd.Series(cars_target, dtype="category")
sb.pairplot(cars_df, hue='group', palette='hls')
*****
```

Building boxplots

```
cars.boxplot(column='mpg', by='am')
```

```
cars.boxplot(column='wt', by='am')
```

```
*****
```

```
sb.boxplot(x='am', y='mpg', data=cars, palette='hls')
```

```
*****
```