Week 1 Assignment

John Morse

This code is based on small assignments from the course *Open Data With Python*. Used here as a refresher on using **pandas**, **numpy**, and **matplotlib**.

In [1]:
```
import numpy as np
import pandas as pd
```

```
Create some data in Python, turn it into a DataFrame.
```

In [2]:
```
df1 = pd.DataFrame({"Var1":[1, 2, 1, 2, 2],
                    "Var2":[9, 8, 7, 6, 6],
                    "Var3":[5, np.nan, 4, 3, 3]})
df1
```

Out[2]:

	Var1	Var2	Var3
0	1	9	5.0
1	2	8	NaN
2	1	7	4.0
3	2	6	3.0
4	2	6	3.0

Write the DataFrame to a .csv file.

In [4]:
```
filename = "df1.csv"
df1.to_csv(filename, index=0)
```

Import the file just created.

In [5]:
```
df2 = pd.read_csv(filename)
df2
```

Out[5]:

	Var1	Var2	Var3
0	1	9	5.0
1	2	8	NaN
2	1	7	4.0
3	2	6	3.0
4	2	6	3.0

Use a Pandas method to drop the row with the missing data.

In [6]:
```
df_no_missing = df2.dropna()
df_no_missing
```

Out[6]:

	Var1	Var2	Var3
0	1	9	5.0
2	1	7	4.0
3	2	6	3.0
4	2	6	3.0

Use a Pandas method to drop the duplicated row.

In [7]:
```
df_no_dups = df_no_missing.drop_duplicates()
df_no_dups
```

Out[7]:

	Var1	Var2	Var3
0	1	9	5.0
2	1	7	4.0
3	2	6	3.0

Import matplotlib.pyplot, then create two arrays to plot.

In [8]:
```
import matplotlib.pyplot as plt
```

In [9]:
```
x = np.array([1,2,3,4,5,6,7,8,9,10])
y = np.array([1,4,9,16,25,36,49,64,81,100])
```

Build up a plot part by part using a series of functions.

In [19]:
```
plt.plot(x, x, label = "linear", ls="solid")
plt.plot(x, y, label = "quadratic", ls="dashed")
plt.plot(x, x**3, label = "cubic", ls="dotted")
plt.xlabel("x label")
plt.ylabel("y label")
plt.title("Simple Plot")
plt.legend()
```

Out[19]:
```
<matplotlib.legend.Legend at 0x20c8b1a95d0>
```