Caitlin Sisilli

Class Assignment 7

Question 1)

Data visualization is important in data analysis because showing raw data can be useful to visualize the data that is there to gain information from the data.

Question 2)

The three graphs to describe a distribution of samples is: A Bar graph, A line graph, and a Histogram.

Question 3)

a)

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data=pd.read_table('timeseries_stockprice.csv',sep=',')

print(data)
```

Output:

```
  Date    Amazon  Google  Facebook
0    5/1/18   927.800  901.94   151.740
1    5/2/18   946.645  909.62   153.340
2    5/3/18   946.000  914.86   153.600
```

```
3    5/4/18   944.750  926.07   150.170
4    5/7/18   940.520  933.54   151.450
5    5/8/18   940.950  926.12   150.710
6    5/9/18   952.800  936.95   151.490
7   5/10/18   953.500  931.98   150.230
8   5/11/18   945.110  925.32   150.310
9   5/14/18   954.500  931.53   150.400
10  5/15/18   958.730  932.95   150.170
11  5/16/18   961.000  940.00   150.110
12  5/17/18   954.700  935.67   148.000
13  5/18/18   944.800  921.00   144.720
14  5/21/18   962.840  931.47   148.445
15  5/22/18   964.000  935.00   148.080
16  5/23/18   975.020  947.92   148.520
17  5/24/18   976.000  952.98   148.510
18  5/25/18   984.850  957.33   150.300
19  5/29/18   995.000  969.70   152.230
20  5/30/18   996.510  970.31   151.970
21  5/31/18  1000.000  975.02   152.700
```

b)

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data=pd.read_table('timeseries_stockprice.csv',sep=',')

print(data.info())
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22 entries, 0 to 21
Data columns (total 4 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Date      22 non-null     object
 1   Amazon    22 non-null     float64
 2   Google    22 non-null     float64
 3   Facebook  22 non-null     float64
dtypes: float64(3), object(1)
memory usage: 832.0+ bytes
```

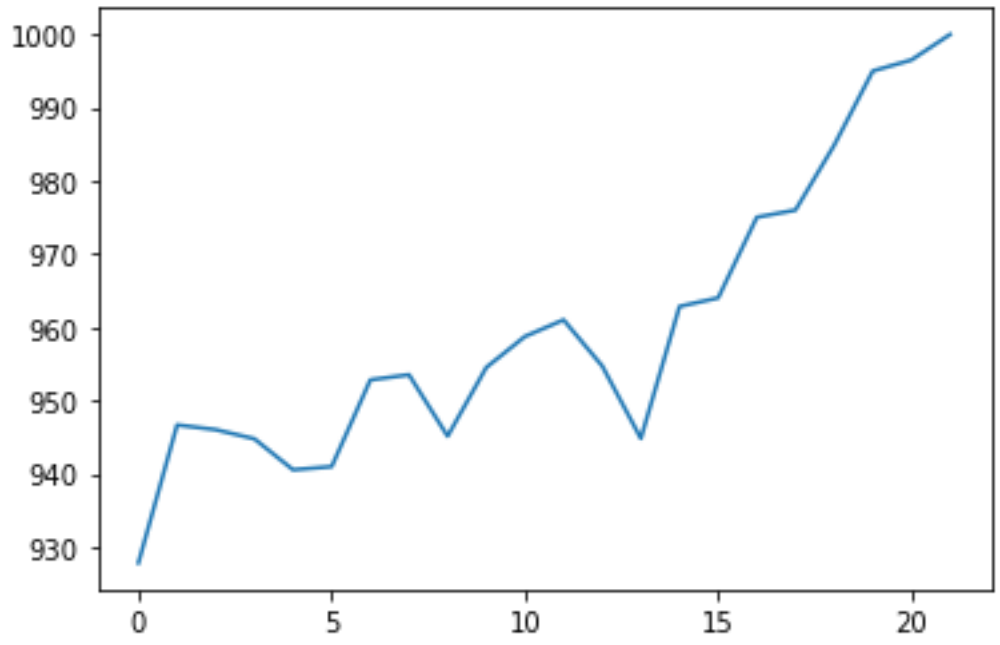
```
None
```

c)

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data=pd.read_table('timeseries_stockprice.csv',sep=',')

print(data['Amazon'].plot())
```

Output:

```
AxesSubplot(0.125,0.125;0.775x0.755)
```



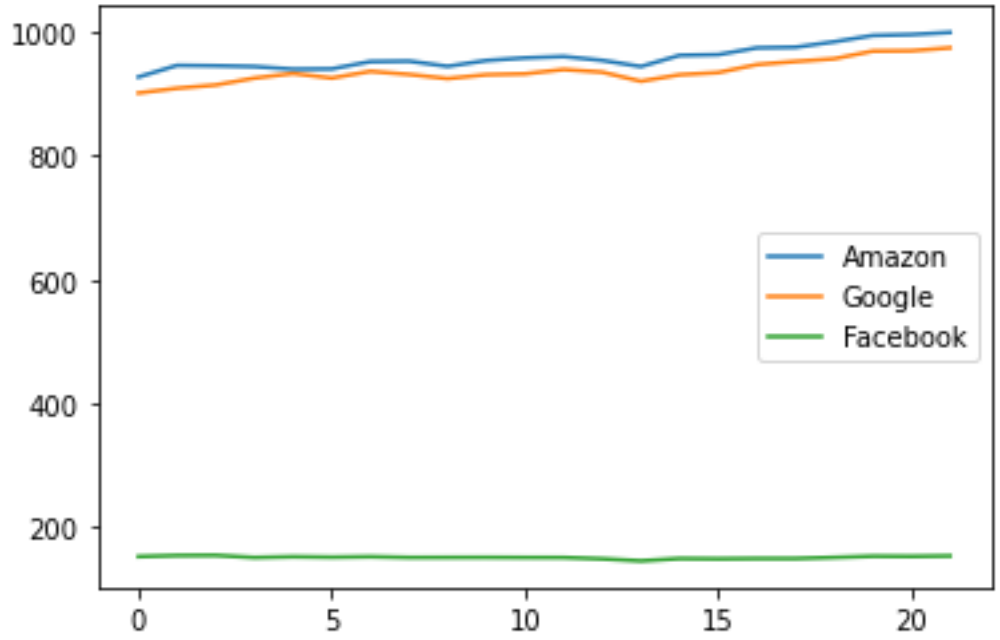
d)

```
import numpy as np
```

```
import pandas as pd

import matplotlib.pyplot as plt

data=pd.read_table('timeseries_stockprice.csv',sep=',')

print(data.plot())
```

Output:

```
AxesSubplot(0.125,0.125;0.775x0.755)
```



Question 4)

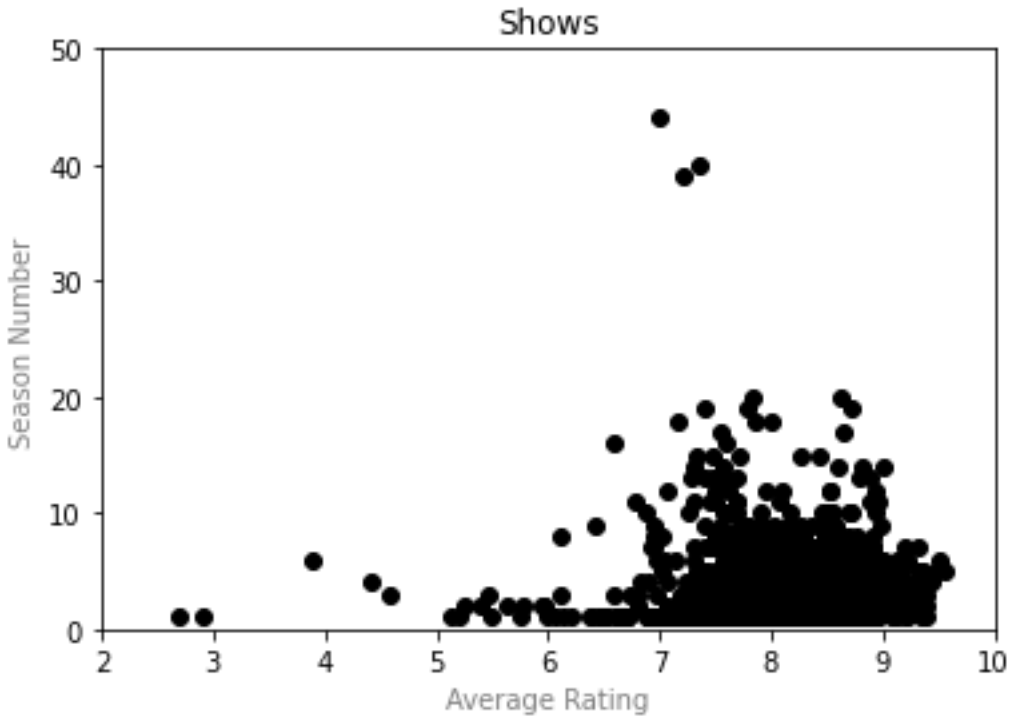
a)

```
import numpy as np

import pandas as pd
```

```
import matplotlib.pyplot as plt

data=pd.read_table('tv.csv',sep=',')

print(data)

y=data['seasonNumber']

x=data['av_rating']

plt.axis([2,10,0,50])

plt.title('Shows')

plt.xlabel('Average Rating',color='gray')

plt.ylabel('Season Number',color='gray')

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

Output:

```
[<matplotlib.lines.Line2D at 0x28e0910a370>]
```



b)

```
import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

data=pd.read_table('tv.csv',sep=',')

df=data[data['seasonNumber'] <=7]

y=df['seasonNumber']

x=df['av_rating']

plt.axis([2,10,0,10])

plt.title('Shows')
```

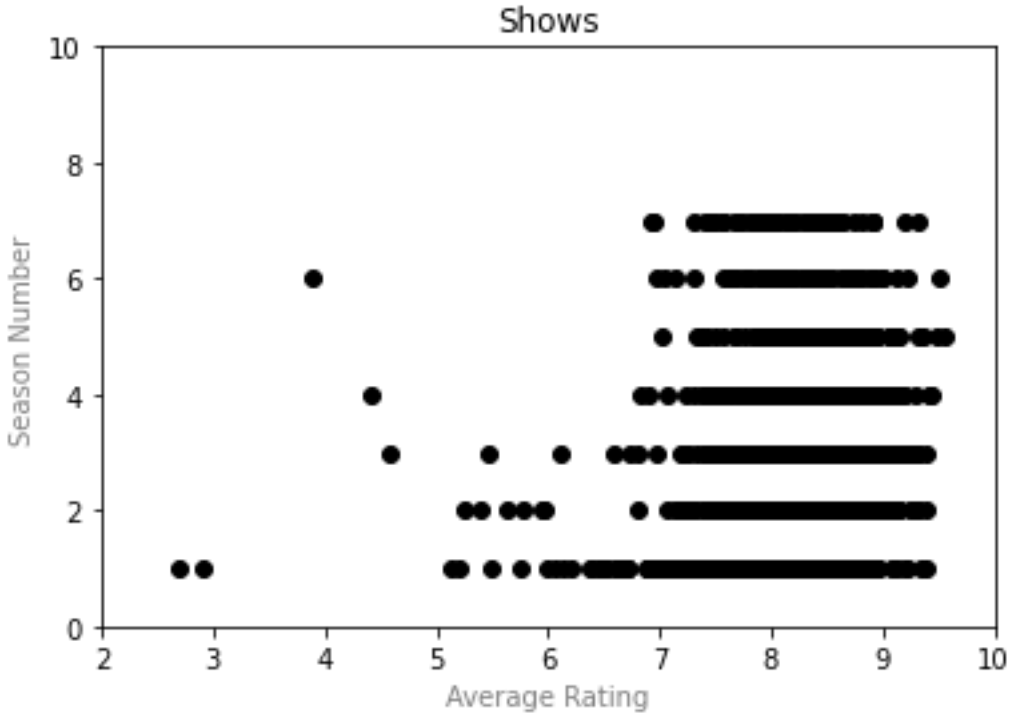
plt.xlabel('Average Rating',color='gray')

plt.ylabel('Season Number',color='gray')

plt.plot(x,y,'ko')

Output:

```
[<matplotlib.lines.Line2D at 0x28e0aeb3c10>]
```



c) There is a difference between both graphs where the last graph is more spread out with the data since the season from seven and less was removed to show more of the information.

The first one is more squished since the graph is all in one part.