MSDS670_CGREEN_WEEK5_ASSIGNMENT

October 10, 2021

0.1 MSDS670 - Week5 - MATPLOTLIB Visualizations - Chris Green

0.1.1 Importing Packages and Initial Data

Note: Much of the data cleaning and prep was done in Excel before importing to Jupyter

[1]:
```
conda install -c conda-forge pandas pandas-profiling matplotlib openpyxl -y
```

```
Collecting package metadata (current_repodata.json): done
Solving environment: done

# All requested packages already installed.


Note: you may need to restart the kernel to use updated packages.
```

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import phik
import seaborn as sns
```

[3]:
```
df = pd.read_excel('50 richest people 2021.xlsx', header=0, usecols = "A:G")
```

[4]:
```
df.info()
```

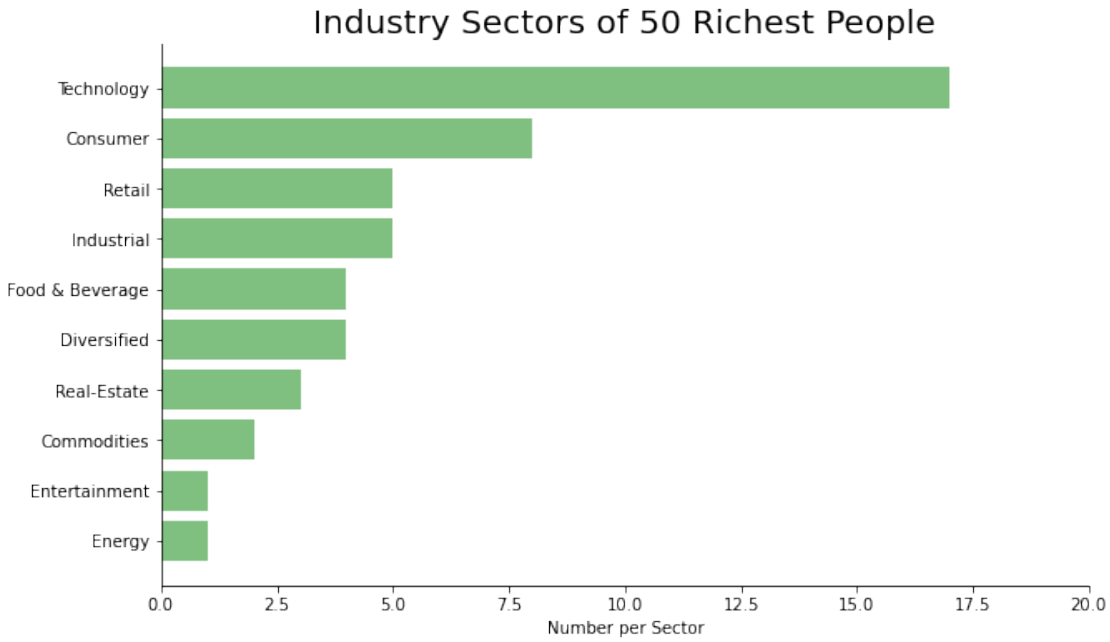
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 50 entries, 0 to 49
Data columns (total 6 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Rank                         50 non-null     int64
 1   Name                         50 non-null     object
 2   Total Net Worth              50 non-null     float64
 3   $ YTD Change (in Billions)   50 non-null     float64
 4   Country                      50 non-null     object
 5   Industry                     50 non-null     object
dtypes: float64(2), int64(1), object(3)
memory usage: 2.5+ KB
```

0.1.2 Horizontal Bar Showing the Industry Sectors for 50 Richest

[5]:
```
df_industry = pd.read_excel('50 Richest Industry Count.xlsx', header=0, usecols␣
↪= "A:B")
```

[6]:
```
df_industry_sorted= df_industry.sort_values('Count',ascending=True)
```

[7]:
```
fig, ax= plt.subplots(figsize=(10,6))
plt.figure(figsize=(10,6))
ax.barh('Industry', 'Count',data=df_industry_sorted, color='green', alpha=0.5)
ax.set(xlim=[0, 20], xlabel='Number per Sector', ylabel='')
ax.set_title('Industry Sectors of 50 Richest People', loc='center', fontsize=20)
right_side = ax.spines["right"]
top_side = ax.spines["top"]
top_side.set_visible(False)
right_side.set_visible(False)
plt.savefig("Industry_barplot.png")
```



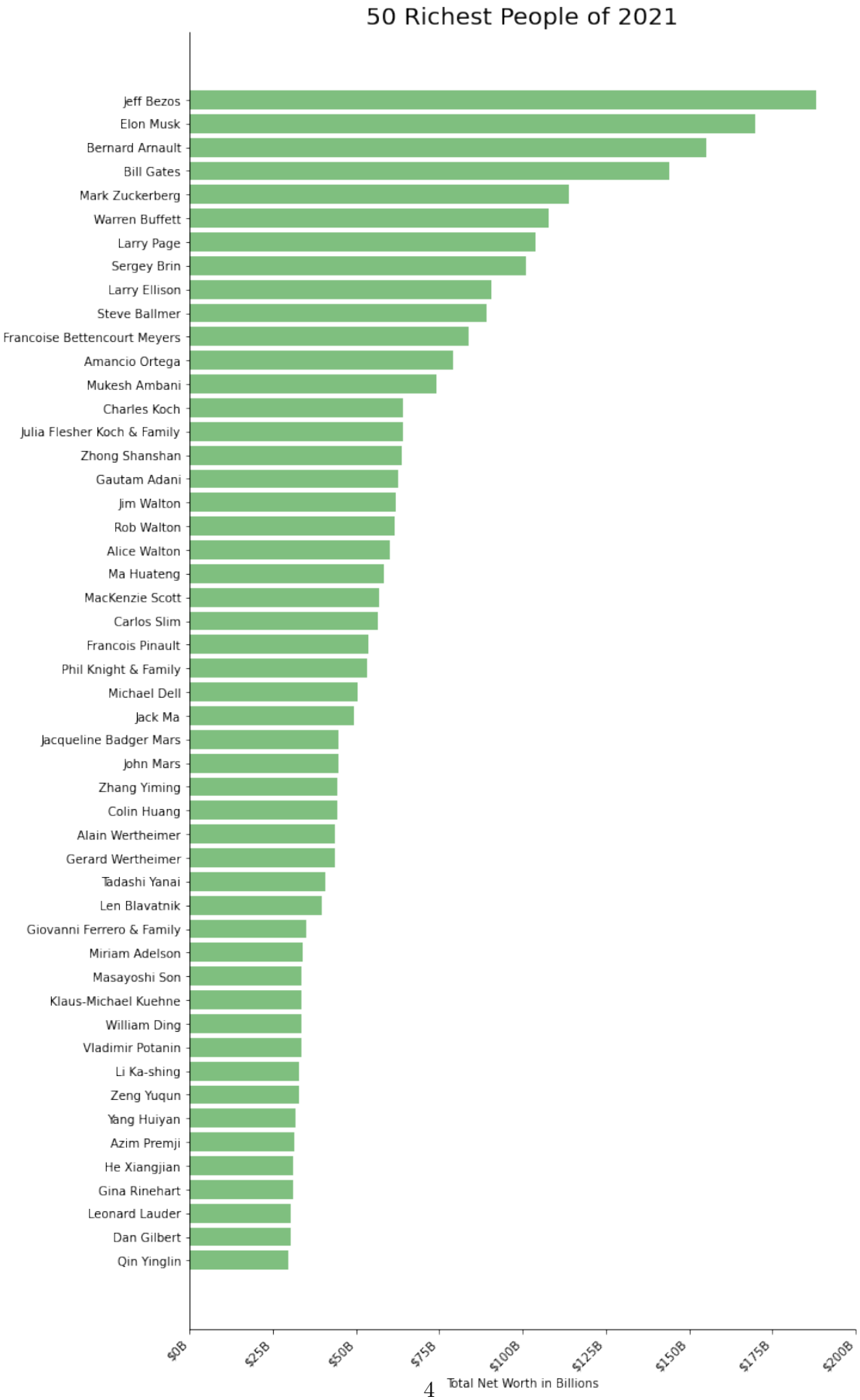
<Figure size 720x432 with 0 Axes>

0.1.3 Horizontal Bar Showing 50 Richest

[8]:
```
df_sorted= df.sort_values('Total Net Worth',ascending=True)
```

[9]:
```
fig, ax = plt.subplots(figsize=(10, 20))
ax.barh('Name', 'Total Net Worth',data=df_sorted, color='green', alpha=0.5)
```

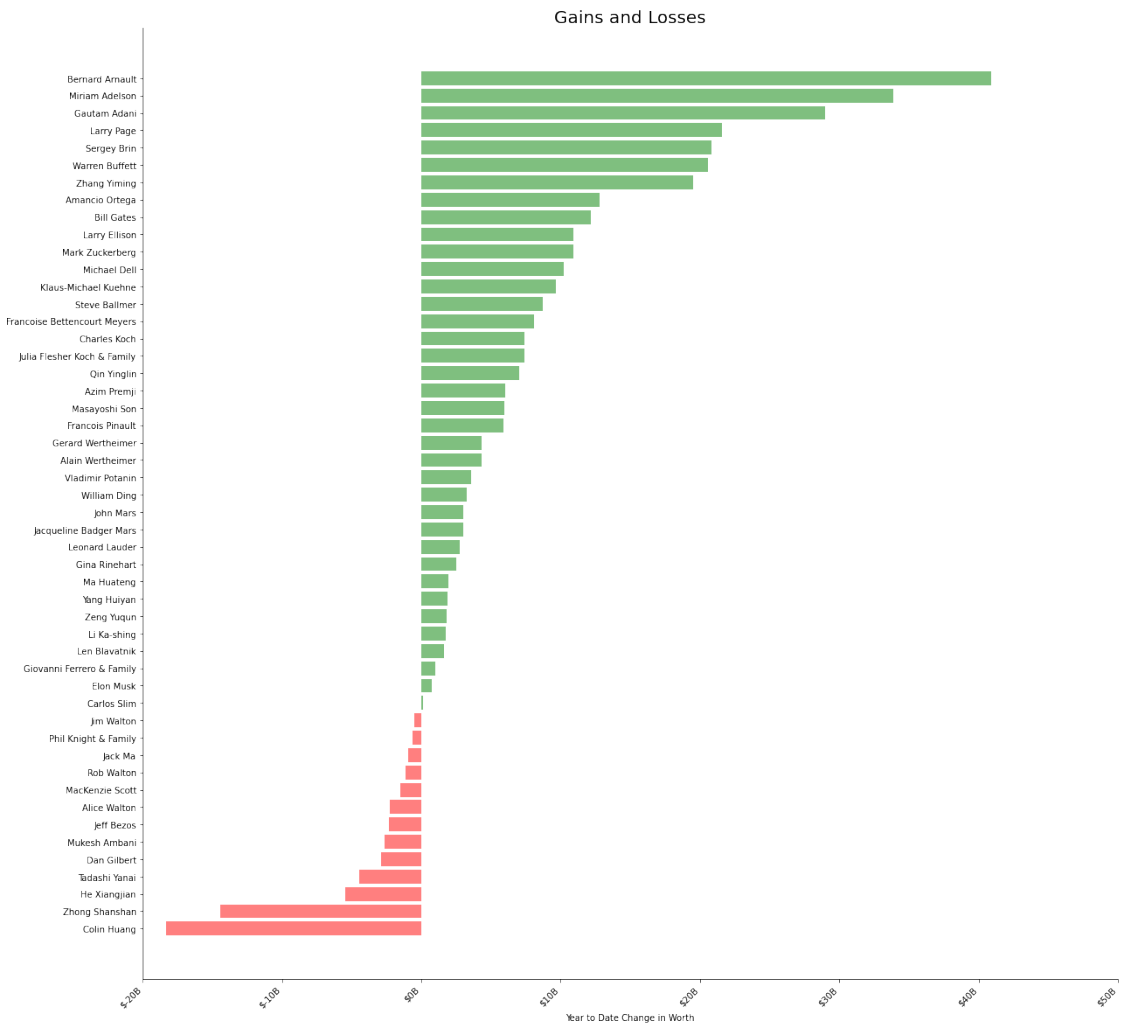
```
labels = ax.get_xticklabels()
plt.setp(labels,rotation=45, horizontalalignment='right')
ax.set(xlim=[0, 200], xlabel='Total Net Worth in Billions', ylabel='')
ax.set_title('50 Richest People of 2021', loc='center', fontsize=20)
ax.xaxis.set_major_formatter('${x:1.0f}B')
right_side = ax.spines["right"]
top_side = ax.spines["top"]
top_side.set_visible(False)
right_side.set_visible(False)
plt.savefig("50_barplot.png")
```


50 Richest People of 2021
Jeff Bezos
Elon Musk
Bernard Arnault
Bill Gates
Mark Zuckerberg
Warren Buffett
Larry Page
Sergey Brin
Larry Ellison
Steve Ballmer
Francoise Bettencourt Meyers
Amancio Ortega
Mukesh Ambani
Charles Koch
Julia Flesher Koch & Family
Zhong Shanshan
Gautam Adani
Jim Walton
Rob Walton
Alice Walton
Ma Huateng
MacKenzie Scott
Carlos Slim
Francois Pinault
Phil Knight & Family
Michael Dell
Jack Ma
Jacqueline Badger Mars
John Mars
Zhang Yiming
Colin Huang
Alain Wertheimer
Gerard Wertheimer
Tadashi Yanai
Len Blavatnik
Giovanni Ferrero & Family
Miriam Adelson
Masayoshi Son
Klaus-Michael Kuehne
William Ding
Vladimir Potanin
Li Ka-shing
Zeng Yuqun
Yang Huiyan
Azim Premji
He Xiangjian
Gina Rinehart
Leonard Lauder
Dan Gilbert
Qin Yinglin
$0B
$25B
$50B
$75B
$100B
$125B
$150B
$175B
$200B
Total Net Worth in Billions

0.1.4 Horizontal Bar Delineating YTD Gains and Losses By Bar Direction and Color

[10]:
```
winloss= df.sort_values('$ YTD Change (in Billions)',ascending=True)
```

[11]:
```
import numpy as np
x=winloss['$ YTD Change (in Billions)']
fig, ax = plt.subplots(figsize=(20, 20))
my_color = np.where(x<0, 'red', 'green')
ax.barh(winloss['Name'],winloss['$ YTD Change (in Billions)'],color=my_color,␣
↪alpha=0.5)
labels = ax.get_xticklabels()
plt.setp(labels,rotation=45, horizontalalignment='right')
ax.set(xlim=[-20, 50], xlabel='Year to Date Change in Worth', ylabel='')
ax.set_title('Gains and Losses', loc='center', fontsize=20)
ax.xaxis.set_major_formatter('${x:1.0f}B')
right_side = ax.spines["right"]
top_side = ax.spines["top"]
top_side.set_visible(False)
right_side.set_visible(False)
plt.savefig("WinLoss_barplot.png")
```



0.1.5 Scatter Plot Showing Distribution of Gains and Losses by Net Worth for 500 Richest

[12]:
```
from mpl_toolkits.axes_grid1 import make_axes_locatable

df500 = pd.read_excel('500 richest people 2021.xlsx', header=0, usecols = "A:G")


fig, ax = plt.subplots(figsize=(7, 7))
x = df500['Total Net Worth']
y = df500['$ YTD Change (in Billions)']
c = df500['$ YTD Change (in Billions)']

scat = ax.scatter(x,y, c=c, cmap = 'RdYlGn', alpha=0.5)
```

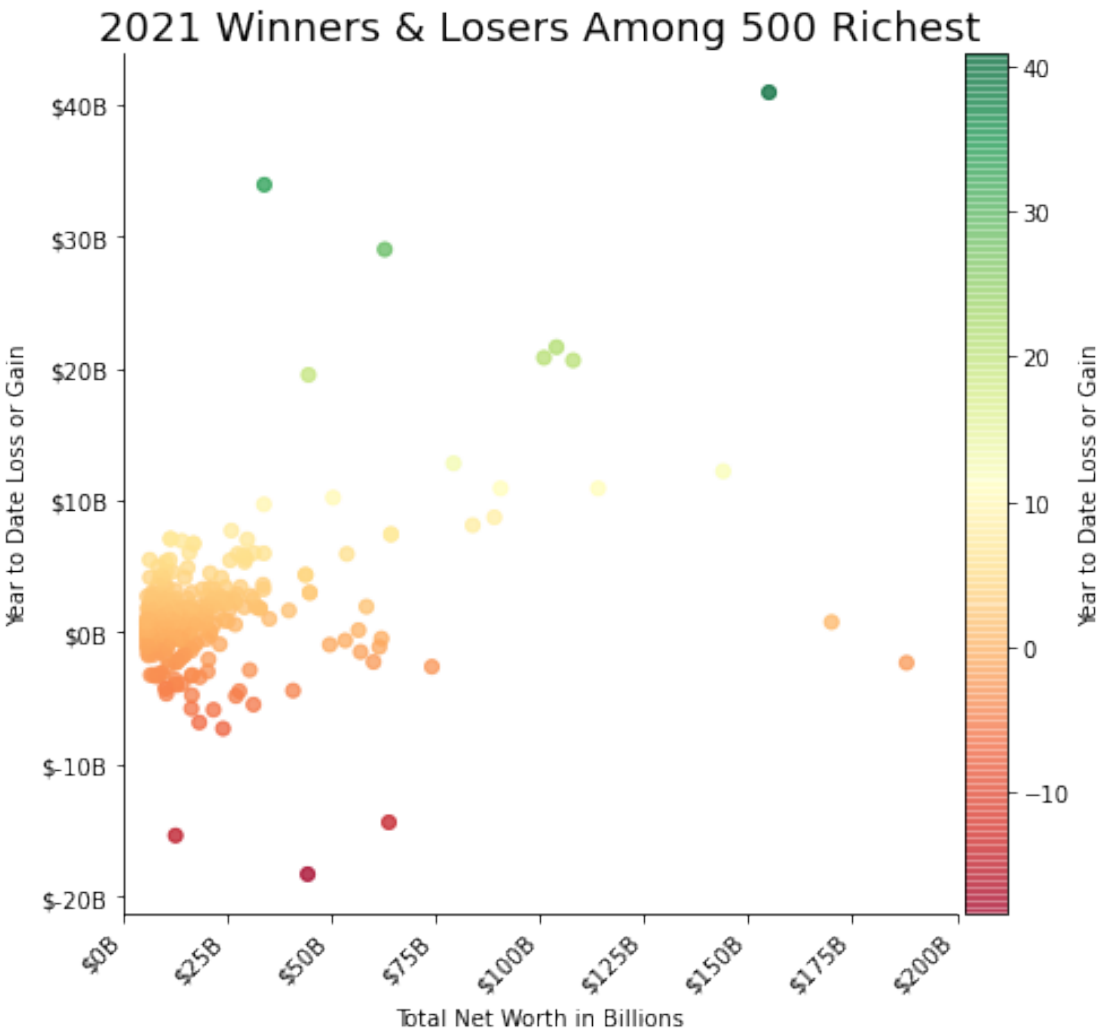
```
ax.scatter(x,y, c=c, cmap = 'RdYlGn', alpha=0.5)

labels = ax.get_xticklabels()
plt.setp(labels,rotation=45, horizontalalignment='right')
ax.set(xlim=[0, 200], xlabel='Total Net Worth in Billions', ylabel='Year to␣
 ↪Date Loss or Gain')
ax.set_title('2021 Winners & Losers Among 500 Richest', loc='center',␣
 ↪fontsize=18)
ax.xaxis.set_major_formatter('${x:1.0f}B')
ax.yaxis.set_major_formatter('${x:1.0f}B')

right_side = ax.spines["right"]
top_side = ax.spines["top"]
top_side.set_visible(False)
right_side.set_visible(False)

divider = make_axes_locatable(ax)
cax = divider.append_axes("right", size="5%", pad=0.05)
plt.colorbar(scat, cax=cax, label='Year to Date Loss or Gain')

plt.savefig("500_scarplot.png")
```


2021 Winners & Losers Among 500 Richest
Year to Date Loss or Gain
$40B
$30B
$20B
$10B
$0B
$-10B
$-20B
$0B
$25B
$50B
$75B
$100B
$125B
$150B
$175B
$200B
Total Net Worth in Billions
40
30
20
10
0
−10
Year to Date Loss or Gain

[ ]: