Homework 05-STUDENTS

December 4, 2021

USC Marshall School of Business

DSO 545- Homework 5

Fall 2020

0.1 Kuochang Lan

0.2 Assignment Learning Objectives

- Customize seaborn plots in Python
- Deal with date variables in Python
- Wrangle data in Python

0.3 Dataset

The dataset consists of 3 variables:

Variables	Explanation
date	date of the poll
approval	President's Trump approval rating
disapproval	President's Trump disapproval rating

0.4 Questions

1. Create an EXACT copy of the following graph that shows Persident's Trump daily approval ratings.

- Each of the red horizontal lines represent the annual average approval ratings in 2017, 2018, 2019, and 2020 respectively
- The line chart is grey in color with alpha = 0.3
- The figure size is 10 x 5

```
[1]: import numpy as np
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     df = pd.read_csv('TrumpRatings.csv',parse_dates = ['date'])
```

```
[8]: df['year'] = df['date'].dt.year
     df.groupby('year').agg({"approve":"mean"})
```

```
[8]:        approve
     year
     2017   39.130631
     2018   40.855546
     2019   41.092979
     2020   41.903386
```

```
[2]: plt.figure(figsize = (10,5))
     plt.gca().spines['top'].set_visible(False)
     plt.gca().spines['right'].set_visible(False)
     sns.lineplot(data = df,
                 x = 'date',
                 y = 'approve',
                 alpha = 0.3,
                 color = 'grey')
     plt.title("President Trump's Approval Rating",
             loc = 'center',
             fontsize = 12,
             fontweight = 'normal',
             color = 'black',
             pad = 20)
     plt.ylabel("approve", color = 'black',fontsize = 9)
     tickpositions = [34,36,38,40,42,44,46,48]
     ticklabels = ['34%','36%','38%','40%','42%','44%','46%','48%']
     plt.yticks(tickpositions,ticklabels)
     avg_2017 = 39.130631
     avg_2018 = 40.855546
     avg_2019 = 41.092979
     avg_2020 = 41.903386
     plt.hlines(y = avg_2017,xmin = pd.to_datetime('2017-01-22'),xmax = pd.
      →to_datetime('2017-12-31'),color = 'red')
     plt.hlines(y = avg_2018,xmin = pd.to_datetime('2018-01-01'),xmax = pd.
      →to_datetime('2018-12-31'), color = 'red')
     plt.hlines(y = avg_2019,xmin = pd.to_datetime('2019-01-01'),xmax = pd.
      →to_datetime('2019-12-31'), color = 'red')
     plt.hlines(y = avg_2020,xmin = pd.to_datetime('2020-01-01'),xmax = pd.
      →to_datetime('2020-09-22'),color = 'red')
```

```
[2]: <matplotlib.collections.LineCollection at 0x7fc9b27851f0>
```

President Trump's Approval Rating
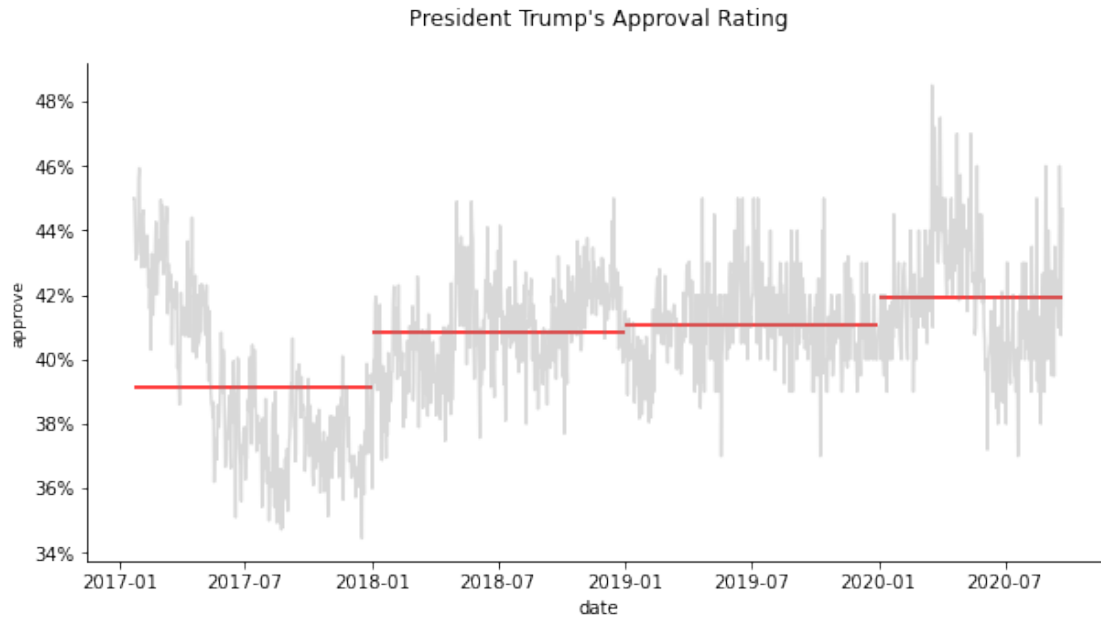
2. Create an EXACT copy of the following graph that shows the average monthy difference between Persident's Trump approval and disapproval ratings (Disapproval% - Approval%) since he took office in 2017.

- The span area is green colored with alpha = 0.3
- The 2017, 2018, and 2019 lines are grey with alpha = 0.3
- The 2020 line is red
- The annotation text fontsize for the years is 10
- The figure size is 10 x 5

```
[12]: df['diff'] = df['disapprove']-df['approve']
      df['month'] = df['date'].dt.month
```

```
[13]: plt.figure(figsize=(10,5))

      diff2017 = df.loc[df['year'] == 2017]
      diff2018 = df.loc[df['year'] == 2018]
      diff2019 = df.loc[df['year'] == 2019]
      diff2020 = df.loc[df['year'] == 2020]

      sns.lineplot(data=diff2017,x = 'month',y = 'diff',color = 'grey',alpha = 0.3,ci
       ↪= False)
      sns.lineplot(data=diff2018,x = 'month',y= 'diff',color = 'grey',alpha = 0.3,ci
       ↪= False)
      sns.lineplot(data=diff2019,x = 'month',y= 'diff',color = 'grey',alpha = 0.3,ci
       ↪= False)
```

```python
sns.lineplot(data=diff2020,x = 'month',y= 'diff',color = 'red',alpha = 1,ci =
 →False)


plt.title("Ratings Diff: Disapproval(%) - Approval(%)",
loc= 'center',
fontsize = 12,
fontweight = 'normal',
pad = 10,
color = 'black')


plt.gca().spines['top'].set_visible(False)
plt.gca().spines['right'].set_visible(False)


tickpositions = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]
ticklabels = ['1', '2', '3', '4', '5', '6', '7', '8', '9', '10',
 →"Election's\nMonth", '12']

tickpositionsv = [0,5,10,15,20]
ticklabelsv = ['0%', '5%', '10%', '15%', '20%']

plt.xticks(tickpositions, ticklabels)
plt.yticks(tickpositionsv, ticklabelsv)

plt.xlabel(" ")
plt.ylabel("Ratings Diff", color= 'black', fontsize = 12)

plt.axvspan(10.5, 11.5, color = 'green', alpha = 0.3)

plt.annotate('2017',(12, 19),color='grey',fontsize=12)
plt.annotate('2019',(12, 11),color='grey',fontsize=12)
plt.annotate('2018',(12, 10),color='grey',fontsize=12)
plt.annotate('2020',(9, 10),color='red',fontsize=12)
```
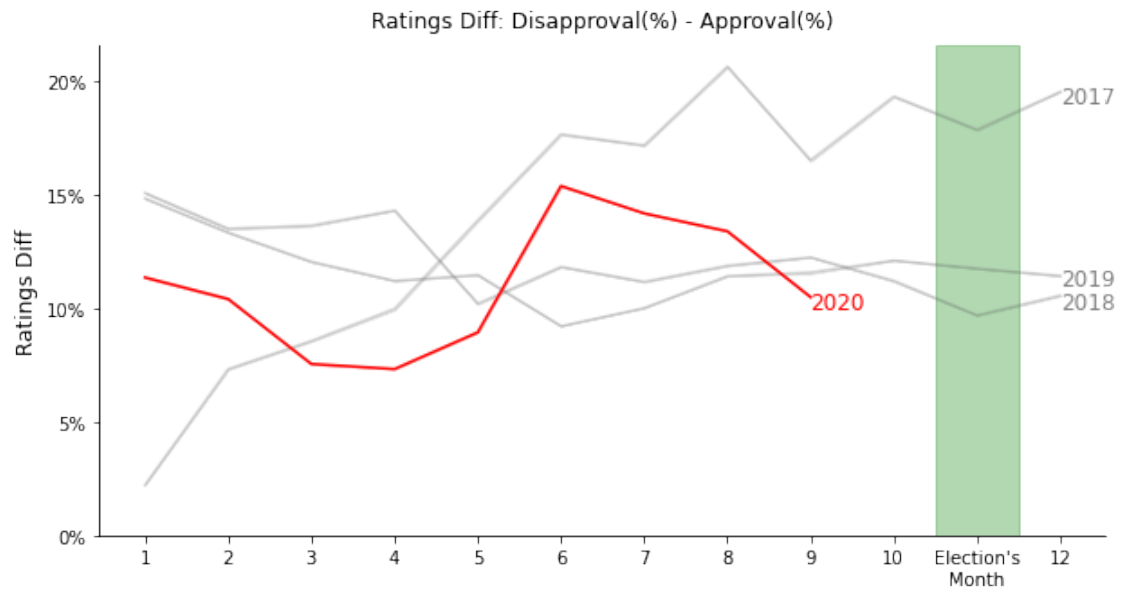
```
[13]: Text(9, 10, '2020')
```

Ratings Diff: Disapproval(%) - Approval(%)

[ ]: