Data Analytics & Visualization with Python
Homework 2

Due date: June 26 2022 at 23.59 CET

Please send your solutions to ahmetcan.kabak55@gmail.com

Remarks:
Write the code yourself. ***Cheating is strictly forbidden***.
For each problem write your code in the function format and give the names of the functions as problem numbers, for example for the solution of problem1:

```
def problem1(input):
  return something
```

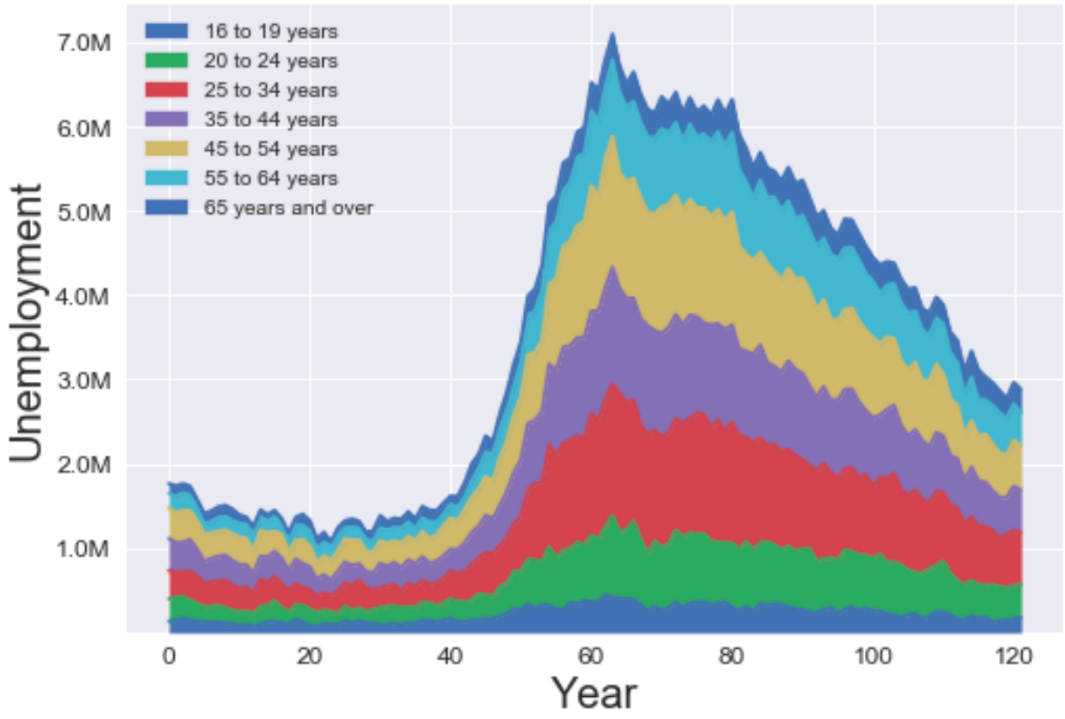
Put the codes for all problems into one file (jupter notebook file) and name that file using your student username in the following format: Ahmetcan_kabak_homework2.ipynb. The notebook file should definitely contain the outputs of the functions, if applicable. Sample solution file (sample_solution.ipynb) is given to you to show how to organize your solutions.
Also write your name the jupyter notebook as well.

Give as much as documentation for your script using comments.

Problem 1 (20 Points)

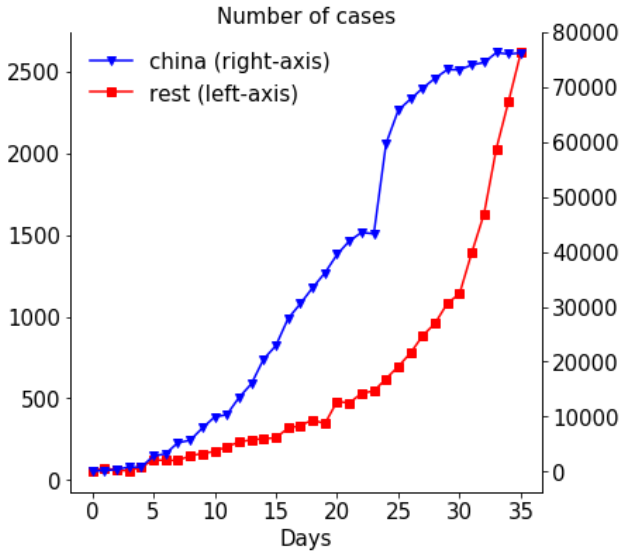
You are given " monthly_unemployment_by_ages.csv" file which contains Unemployment data. Plot stacked area plot of unemployment splitted by age as a function of year. Set x and y-axis label fontsize as 20.

```
problem1("monthly_unemployment_by_ages.csv")
```



Problem 2 (20 Points)

You are given " corona_china_vs_rest.csv" file which contains the number of corona virus cases for china and the rest of the world. Using this data, plot the figure given below. Also save this figure as a png file.

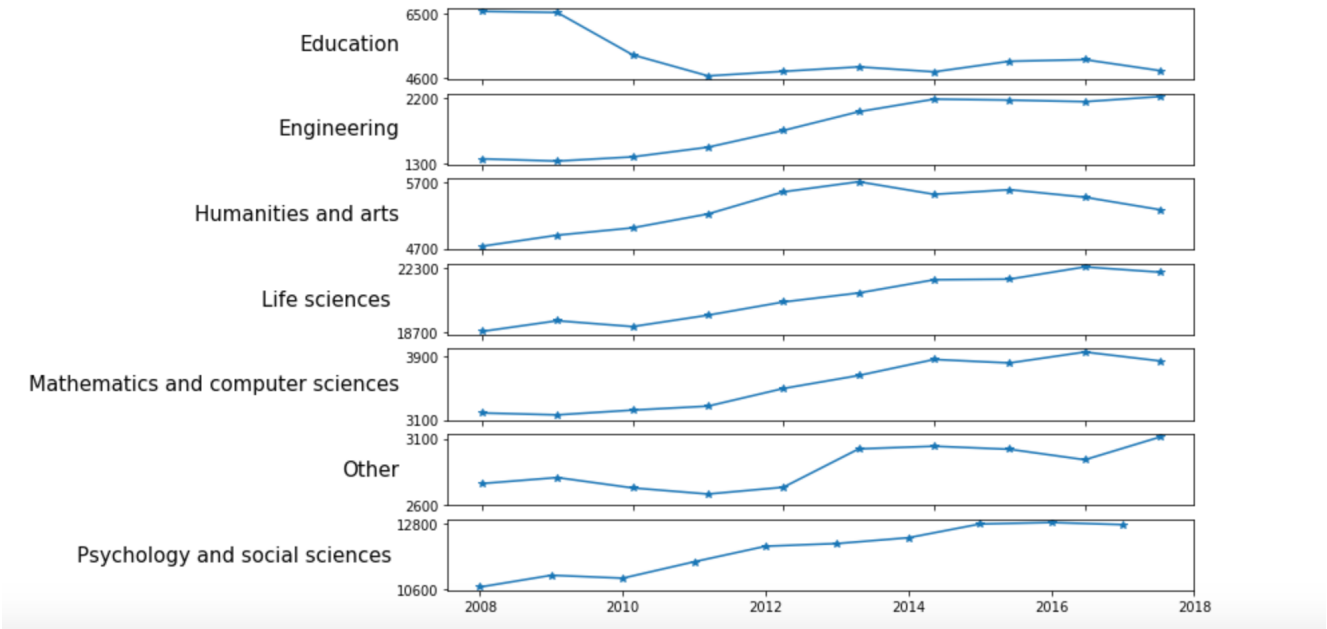


Problem 3 (20 Points)

You are given “phd_by_field.csv" file which contains the number of PhDs awarded in US by year. Use this file as an input and produce below figure for yearly total number of PhDs for each broad field. For the figure apply the following details:

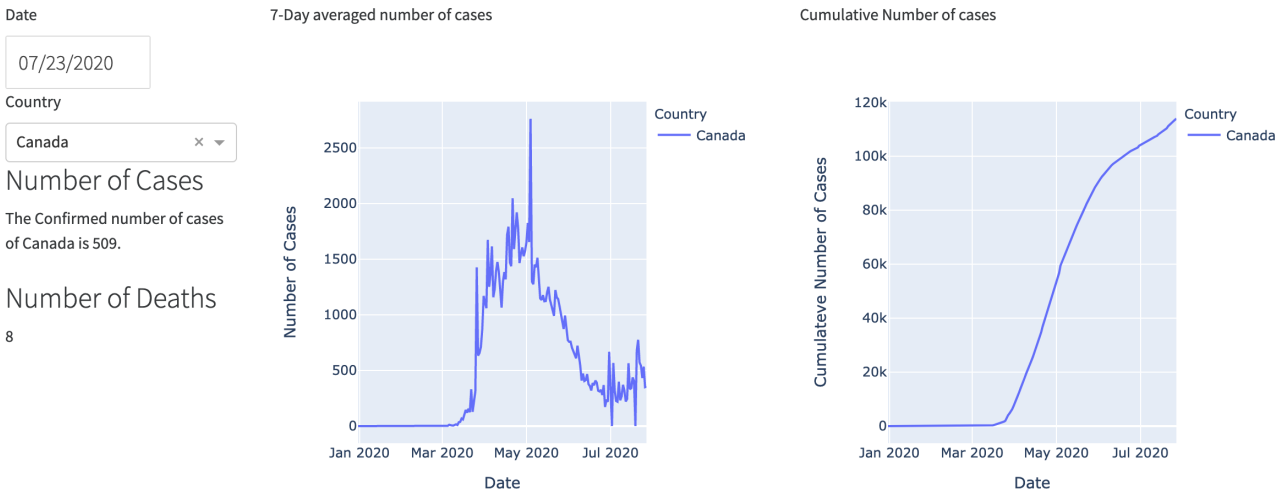
- Figure size should be 10x8 inches
- But y-labels as the field name
- For y-ticks of each subplot put only minimum and maximum values. Truncate these numbers to nearest hundred like 2245 to 2200, or 3980 to 3900

```
problem3("phd_by_field.csv")
```



Problem 4 (20 Points)

You are given "covid19_ecdc_28_07_20.csv" file containing statistics for Covid-19. Use this data as the input and produce the following dashboard using plotly-dash.
The user should be able to choose date and country. There objects in dash core components to select dates.



Just a hint you can use the following code to calculate 7-day average of number of cases.

```
# first read data into a dataframe
import pandas as pd
data=pd.read_csv("covid19_ecdc_28_07_20.csv")
# create datetime objects for the "dateRep" column
data.dateRep=pd.to_datetime(data.dateRep,format="%d/%m/%Y")
# reshape the data: Show countries in rows, dates in columns and values should be cases
confirmed_by_country=data.pivot(index="countriesAndTerritories",columns="dateRep",values="cases")
# calculate rolling average of cases for 7-day windows
confirmed_7day_mean=confirmed_by_country.rolling(7,axis=1,min_periods=1).mean()
```

```
confirmed_7day_mean.head()
```

dateRep	2019-12-31	2020-01-01	2020-01-02	2020-01-03	2020-01-04	2020-01-05	2020-01-06	2020-01-07	2020-01-08	2020-0
countriesAndTerritories										
Afghanistan	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Albania	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Algeria	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
Andorra	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
Angola	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Problem 5 (20 Points)

In this problem you are going to produce animated plot for the results of indian general election. You are given whitespace delimited “indian_elections.dat” file which contains percentages of seats won by two parties over the years.

Write a code that uses “indian_elections.dat” file and produce animated line plot for these results. You are given a video file “animation.mp4” that shows how your output should look like. Also a snapshot of the video is given below. Note that for this problem don't write your answer as a function. There is a problem calling animatefunction inside a function. There is workaround for this issue, but you don't need to spend time on that. Please follow the instructions provided below.

- For the y-axis ticklabels use percentage sign
- Animation should be repeated; you don’t have to write an outputfile just working on jupter notebook file is fine.
- Remove right and top spines
- At each frame write the percentages of seats for each party.
- Put year information as the title. The year should be updated in each frame.
- Put grid lines as dashed style