Midterm Examination: Part 2

Due date: 15 March 2022 before 10 am.

Points: 15 points for Part 2 (+5 extra points)

Do not alter this file.

Duplicate this file and move it into the Personal folder. Rename the file as Midterm_id (id is your student ID number, e.g. Midterm_Part1_6305001).

Write Python commands to answer each of the questions. For each question that requires numerical values (not list or dataframe), you also need to assign the variable e.g. ans1 to store the numerical answer for question 1. If there is more than 1 answer required, you must create more variables e.g. ans1_1, ans1_2 to store the values of the answers.

When you want to submit your file, you simply share access with me using my email pairote.sat@mahidol.edu and my TA p.pooy.pui.i@gmail.com. Do not move your file into the DS@MathMahidol team.

Data set

In the midterm exam part 1, we have already combined the datasets by adding the column 'total_population' to the dataset Covid-19. In Part 2, we will use this combined dataset to perform a statistical data analysis and present the results of the analysis.

You can download the file (US_covid19_part2.csv) from my Github page:

https://raw.githubusercontent.com/pairote-sat/SCMA470/master/US_covid19_part2.csv

```
import pandas as pd
import numpy as np

# https://stackoverflow.com/questions/20625582/how-to-deal-with-settingwithcopywarning-in-
pd.options.mode.chained_assignment = None
```

```
from plotnine import *
```

1. Enter student IDs of your group.

```
id =
```

```
#path = '/Users/Kaemyuijang/SCMA248/Data/US_covid19_part2.csv'
#final = pd.read_csv(path, parse_dates=True, index_col = 'date')

url = 'https://raw.githubusercontent.com/pairote-sat/SCMA470/master/US_covid19_part2.csv'
final = pd.read_csv(url)
```

```
final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1312449 entries, 0 to 1312448
Data columns (total 9 columns):
 #   Column            Non-Null Count    Dtype
---  ------            --------------    -----
 0   Unnamed: 0        1312449 non-null  int64
 1   date              1312449 non-null  object
 2   county            1312449 non-null  object
 3   code              1312449 non-null  object
 4   state             1312449 non-null  object
 5   fips              1312449 non-null  float64
 6   cases             1312449 non-null  int64
 7   deaths            1312449 non-null  float64
 8   total_population  1312449 non-null  int64
dtypes: float64(2), int64(3), object(4)
memory usage: 90.1+ MB
```

Instructions:

1. Read the US_covid19_part2.csv dataset from my Github page.
2. As a reference, go to the **Worldometer** website.

https://www.worldometers.info/coronavirus/country/us/

Worldometer is a reference website that provides counters and real-time statistics on various topics, including the Covid-19 pandemic statistics.

1. Write Python code to perform statistical data analysis of your own choices. The results of your Python code may include (but not be limited to) the following:

3.1. Recent coronavirus cases in the United States and the number of deaths.

3.2. The table shows the total number of cases, total number of deaths, total number of cases/1 million population, total number of deaths/1 million population, and population **classified by U.S. states**.

3.3. Graphs of total coronavirus cases in the United States in linear and logarithmic representation. **Be sure to add axis labels, title, and legend as appropriate.**

3.4. Graphs of daily New Cases in the United States. **Be sure to add axis labels, title, and legend as appropriate.**

3.5. **Extra credit (5 points):** Add the 7-day moving average to the graphs of daily new cases in the United States.

> **key aspects of data science: data retrieval and manipulation, data visualization, statistical computation, presentation and communication;**