DSAA 5024 Data Exploration & Visualization

Assignment 2

- Posting date: 28 Sept. 2022
- Due date: 11:59 PM (Beijing time) on Tuesday, 4 Oct. 2022.

Scatter plot

- Data: *full_data.csv* (latest data can be found in https://github.com/owid/covid-19-data/tree/master/public/data/jhu) describes the number of COVID-19 new, total, and weekly cases and deaths, per country/region. Each row represents daily updates, and each row contains 10 columns: *date*, *location, new_cases, new_deaths, total_cases, total_deaths, weekly_cases, weekly_deaths, bi_weekly_cases, biweekly_deaths.*

- Task:
  - Coding:

    1. Create a scatter plot (https://en.wikipedia.org/wiki/Scatter_plot) to visualize *total_cases* and *total_deaths* on the latest day as in the data. (20 points)

    2. The scatterplot shall provide basic interaction, e.g., to display country/region name and case-fatality rate upon mouse hover over a data point. (20 points)

  - Report: Answer the following questions:

    1. What is the data type of each data attribute (*date, location, new_cases*, …)? (20 points)

    2. Describe what you have learned from the data and the visualization. For instance, which country has the most number of total cases/deaths? Which country has the highest case-fatality rate? (30 points)

  - Extension of the scatterplot: On basis of the scatterplot, how would you add further information including *population* and *continent*?

You don't need to implement the code. Simply describing the visual mapping is enough. (10 points)

TURNIN instructions:

- You can code using either html and d3.js, Tableau, or any other tool. Your code should be executable, and you are required to include a README file to describe how to run your code.

- Zip your code and report in a single file, named as *A2_studnet ID_name.zip*.