Clarification questions

Please ask any question related to this assignment in the class or in the Canvas forum. To be fair to everyone, it is best to ask assignment question openly rather than in email. Do not share your code with anyone. Do not post your code online. Do not ask for help in any public discussion forum or online website. You can search over the internet and read books. Let me know if you find any error in the assignment question as soon as possible. DO NOT SHARE OR DESCRIBE YOUR OUTPUT TO ANY OTHER STUDENT. DO NOT POST YOUR VISUALIZATION ON CANVAS. You can always ask questions about the methods, definitions or approaches to solve a question. PLEASE DO NOT ASK ME 'WHETHER MY OUTPUT LOOKS CORRECT' (your outcome is only assessed after you submit your assignment).

Assignment 1

Submission Instruction: You must write your code in the given HTML file. Submit the file and a screenshot of the visualization. Clicking on the file must generate all the outputs on the browser. In some questions you are asked to write python code and generate csv files. You must submit the python file and the csv file in your submission. Do not log any unnecessary values in the console. Do not submit unnecessary files such as web scrapped data.

Q1: GOAL: Simple Plotting in D3 [Use of AI-generated code allowed], Mark : 30

Write a program that generates real numbers x and y at random in the range $-2 < x < 10$ and $-3 < y < 3$ until you find at least 1000 points that satisfy the following equation

$(2x^2 + y^2)^2 - 14x(2x^2 - 3y^2) + 50(y^2 - x^2) < 0$.

Plot these points in a scatter plot. Scale and translate the points as necessary so that all the points are displayed on the screen. The goal is to show a smooth shape. Choose an appropriate number of points and radius for each point so that the shape is clearly visible.

Hint

You may need to search for how to generate random number in javascript and then write code to create an array of points that satisfy the equation and finally use d3 to generate the plot. You may want to learn how to create object array in javascript.

If you do it right, you will get a fish shape. You may need to increase the number of points to get a smooth shape. You can also try it out with $(x^2+y^2-1)^3 - x^2 \cdot y^3 < -0.07$ or $(x^2+y^2-1)^3 - x^2 \cdot y^3 < 0$ to get a Mickey mouse or a heart.

Q2: GOAL: Data Collection and Cleaning Up [Use of AI-generated code allowed] Mark : 30

Web scraping is a popular way to collect data. The task is to collect data about top 20 cited papers of an author and create a xn6BVdoAAAAJ.csv file that writes the yearly citation count for each paper from 2014 as follows. You must edit the given python code to complete the task and submit the python file and xn6BVdoAAAAJ.csv file in your submission.

For example, here is the information of the author with userid xn6BVdoAAAAJ:

https://scholar.google.ca/citations?user=xn6BVdoAAAAJhl=en

The xn6BVdoAAAAJ.csv file that you produce should look like this. Note that the numbers may not match because by the time you are reading this, Google may update the values.

```
PaperId,Year,CiteCount
1,2014,19
1,2015,23
1,2016,26
1,2017,17
1,2018,38
1,2019,38
1,2020,23
1,2021,22
1,2022,18
1,2023,17
1,2024,9
2,2014,0
2,2015,0
2,2016,6
2,2017,23
2,2018,30
2,2019,26
...
...
```

Write a code in the HTML file to create 20 rows of circles. The ith row contains 11 circles representing the citation for 11 years [2014, 2015, ... , 2024] for the ith paper, where $1 \leq i \leq 20$. Your code must read the csv file to generate this visualization. Here is a visualization for the top two rows. The radii of the circles should be scaled based on the citation data. The circles should utilize the space allocated to them as much as possible. The circles can touch but not overlap.

Make sure that the visualization code works with any other author if we change the csv data and the csv filename. Changing the author must not require any change in the visualization code except for changing the filename.

Hint

You will need to understand the given python code and what it prints. The python code may take some time to run. You can ask questions about the given code in Canvas if something is unclear. You may need to search for how to write a line in a file in python. You may need to iterate over javascript object array and access the javascript object values.

Q3: GOAL: Use Statistics in D3 [Use of AI-generated code allowed] Mark : 40

Write a program that reads the data file msl.csv, normalizes the values between 0 to 1, and writes the first row and first column using console.log(). For each x, y coordinate, draw a square with a size length equal to 5 times the value at that coordinate. Draw the squares so that they do not overlap.

After you are done with the normalization, find the 25, 50 and 75 percentile values and log those into the console. Color the squares of your previous plot with value falling within [0-25%], (25%-50%], (50%-75%], (75%-100%] using White, Green, Blue, Red colors, respectively.

Hint

It is normal to have the same number for some percentile values.