

DSAA 5024 Data Exploration and Visualization

Assignment 1

- Posting date: 20 Feb. 2022
 - Due date: 11:59 PM (Beijing time) on Monday, 26 Sept. 2022.
-

Data transformation

index.html presents a scatterplot of 10,000 random data points. Try to increase the number of data points to 1 million, 10 million, and even 100 million. Answer the following questions.

1. What problems do you encounter when the number of points increases? (30 points)
2. Which data transformation method would you choose to address the problems? Name two data transformation methods. (30 points)
3. What are the benefits and drawbacks of each transformation method? (40 points)
4. Implement one of the data transformation method and submit the code. (20 bonus points)

It's also OK if you do the assignment in Tableau.

TURNIN instructions:

Submit a report called *A1_student ID_name.pdf*, with all questions answered.

If you want to take the bonus points, put your code in *index.html* (or some *twb* file if you use Tableau), zip your code and report in a single file, named as *A1_studnet ID_name.zip*.