
hw-01: Exploring Data

This is a pair assignment

Purpose:

The purpose of this assignment is to practice the early stages of visualization design, and exploratory data analysis.

Tasks:

Start a document for this assignment. Anything you are asked to turn in for the tasks below should be recorded in this document. Responses for each task should start on a new page of the document.

1. Preliminary Design Steps (30 points)

For this portion of your assignment, you will work with VAST Challenge data. The VAST Challenge is a visual analytics design challenge offered each year at the IEEE VIS conference. This page provides background and an overview of the VAST Challenge from 2021: <https://vast-challenge.github.io/2021/description.html>.

Navigate to the link above and read about the VAST Challenge from 2021. You will notice that the VAST Challenge is broken into 3 Mini-Challenges. Read about each Mini-Challenge and think about designing a visualization tool to help solve the problem described in the challenge. For each Mini-Challenge please do the following:

- a) Identify the **domain situation** your visualization tool will support.
 - Who is the target user for the tool? Describe their area of expertise and the experience with visualization and analytics you expect them to have.
 - What is the overarching purpose of the visualization tool?
- b) Download the data for the Mini-Challenge. Assess the **data** using a tool of your choice (excel, python, Google sheets, etc.).
 - What is the quality of the data? Are there many missing values or unusable variables? Explain your answer.
 - What is the source of the data?
 - Are there any biases or ethical issues associated with the data? If so, describe them.
- c) Given your domain situation and data, identify one specific **task** your visualization tool will support. Describe the task.
- d) Name one **visual encoding** you could incorporate into your tool to support the tasks and data described above.

2. Exploratory Data Analysis (20 points)

For this portion of your assignment, you will perform an exploratory data analysis on the class using data collected from the anonymous part of the hw-00 survey.

Start by downloading the de-identified-class-data from the ~~class Google Drive~~ course website (under Homework > hw-01-data, or in the class schedule). **Edited 09.13.2022 with correct data location.**

You may use a tool of your choice to perform your exploratory analysis. Options include (but are not limited to): Python, Excel, GoogleSheets, Tableau. Use whatever is most comfortable to you. As long as you answer the questions below fully we have no preference on the tool or language you use.

Using the tool of your choice, explore your data and learn about the class. Be sure to include the following in your analysis:

- a) Are there any data quality issues? If so, what are they?
- b) Correct any data quality issues you identify as you see fit. Record how you corrected the issues in enough detail for someone to replicate what you did.
- c) Generate visualizations to answer the following research questions. For each, include a screenshot of the visualization and the answer to the question in your document.
 - i. How many first, second, third, fourth, and fifth year students are in the class?
 - ii. Which programming language is most familiar to the class? Which is least familiar?
 - iii. Are more people familiar with bar charts or bubble charts?
 - iv. Is there a relationship between temperature outside, and date and time?
- d) Come up with three additional research questions. Record each in your document. Generate visualizations to answer each of your research questions. For each, include a screenshot of the visualization and the answer to the question in your document.

3. Reflection (5 points)

Answer each of the following questions with your partner unless instructed otherwise.

- a) What is your biggest takeaway message from each of the tasks for this assignment?
- b) Did both of you put your best effort into this assignment? If not, what obstacle got in your way and how can you avoid it in future assignments?
- c) (Each partner individually) Name one contribution your partner made to this assignment that you are proud of.

Submission:

Save your document as a PDF and submit it on Gradescope. **Be sure to assign questions to pages appropriately when you submit.**