Data Visualisation Assignment
Behavioral Research: Statistical Methods,
Spring 2023

1 Instructions

Please read the following instructions carefully before attempting.

1. The dataset for all questions are available at this link. The required sheet name is mentioned in each question. Download the dataset and load in your code accordingly.

2. Make sure to label all plots correctly. Describe and comment with clear reasoning for each question.

3. If you are coding in R, make sure to submit only the pdf file after knitting. File name format should be <Roll Number>_Visualisation_Assignment.pdf

4. If you are using any other language/software, submit the appropriate file following the same nomenclature.

2 Statistical Deception

Sheet 1 (Statistical Deception) contains data for 4 groups, x_1, x_2, x_3, x_4. The data has been specifically crafted so that it is misleading in one of visualisation methods. Identify that method and plot it giving reasons as to why it is so.

Now, plot the best visualisation for the given data. Justify your reasons for the same.

3 Personality and Motion

Sheet 2 (Movement Personality Results) contains the Joint Importance values for personality traits like Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism.

Researchers analyzed the movement patterns of individuals and attempted to predict personality scores. They employed machine learning models wherein 12 joint-movement vectors were used to predict personality scores.

The researchers now want to visualize the results to showcase how each of these joints contributes to predicting each of the five personality scores.

Pick two ways to visualize this data and justify why one would be superior.

4   Data Plotting Adventure

You will be given different scenarios in the following subtasks, and you have to develop a statistical visualization for the given data that best suits the given data.

Label the plot correctly and justify your reasons for the same. Also write 2-3 lines about what one can infer from the statistical visualization that you provide.

4.1   Subtask 1: Sinking Ship

A ship hit an accident, and the following details represent the statistics of people who survived:

- 118 people from the 1st Class, who were males, died.
- 62 people from the 1st Class, who were males, survived.
- 4 people from the 1st Class, who were females, died.
- 141 people from the 1st Class, who were females, survived.
- 154 people from the 2nd Class, who were males, died.
- 25 people from the 2nd Class, who were males, survived.
- 13 people from the 2nd Class, who were females, died.
- 93 people from the 2nd Class, who were females, survived.
- 422 people from the 3rd Class, who were males, died.
- 88 people from the 3rd Class, who were males, survived.
- 106 people from the 3rd Class, who were females, died.
- 90 people from the 3rd Class, who were females, survived.
- 670 people from the Crew, who were males, died.
- 192 people from the Crew, who were males, survived.
- 3 people from the Crew, who were females, died.
- 20 people from the Crew, who were females, survived.

4.2   Subtask 2: Petal Prediction

Sheet 3 (Petal Prediction) contains a part of the IRIS dataset.

Given this dataset, extract the rows pertaining to Sepal Length and Flower Species and plot visualizations to understand the relationship between them. Mention what inferences you can make from your plot and justify the reasoning behind your choice.

4.3   Subtask 3: Spotify Wrapped

Sheet 4 (Spotify Wrapped) contains a part of the dataset which tracks global daily streams on the music streaming service Spotify. We focus on five popular songs from 2017 and 2018.

Given this dataset, plot visualizations to understand how the stream data changes over time. Mention what inferences you can make from your plot and justify the reasoning behind your choice.

5   Need For Speed: Heatmap

Sheet 5 (Need For Speed) contains data pertaining to automobiles enlisting features such as physical make of the car, horsepower, mileage, price etc.

Plot a heatmap for different types of correlations between the features present in the dataset and write about the inferences you can make from it. Also justify as to why one is better than the other.