Data Visualisation Assignment
Behavioral Research: Statistical Methods,
Spring 2024

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 have problems knitting to PDF, please knit to HTML first, then save it/print it as PDF and submit the file.
5. If you are using any other language/software, submit the appropriate file following the same nomenclature.

1 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.

2 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.

3 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.

3.1 Subtask 1: The Last of Us

You're part of a research team analyzing the survival outcomes of individuals in different types of locations during a zombie apocalypse. The goal is to understand where survival chances are highest and what the common outcomes are in each location.

- 118 people from the Safe Zone, who were males, turned into zombies.
- 62 people from the Safe Zone, who were males, survived.
- 4 people from the Safe Zone, who were females, turned into zombies.
- 141 people from the Safe Zone, who were females, survived.
- 154 people from the Contaminated City, who were males, turned into zombies.
- 25 people from the Contaminated City, who were males, survived.
- 13 people from the Contaminated City, who were females, turned into zombies.
- 93 people from the Contaminated City, who were females, survived.
- 422 people from the Rural Area, who were males, turned into zombies.
- 88 people from the Rural Area, who were males, survived.
- 106 people from the Rural Area, who were females, turned into zombies.
- 90 people from the Rural Area, who were females, survived.
- 670 people from the Isolated Island, who were males, turned into zombies.
- 192 people from the Isolated Island, who were males, survived.
- 3 people from the Isolated Island, who were females, turned into zombies.
- 20 people from the Isolated Island, who were females, survived.

3.2 Subtask 2: Glass Glimpse

Sheet 3 (Glass Glimpse) contains a part of the Glass Classification dataset.

Given this dataset, extract the rows pertaining to Glass Type and RI (Refractive Index) 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.
The *types* are numbered 1 to 7, each number corresponding to the following type:

- 1 building-windows-float-processed
- 2 building-windows-non-float-processed
- 3 vehicle-windows-float-processed
- 4 vehicle-windows-non-float-processed (none in this database)
- 5 containers
- 6 tableware
- 7 headlamps

3.3 Subtask 3: Night at the Museum

Sheet 4 (Museum Visitor) contains a part of the dataset which tracks visitors at the Los Angeles Museums. We focus on five different museums and track visitors from 2014 to 2023.

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

4 Fast and Furious: Heatmap

Sheet 5 (Fast and Furious) contains data pertaining to automobiles enlisting features such as a car's brand, miles per gallon, manufactured year 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.
HINT: Make sure you convert all types to numeric and handle missing values.