Coursework 1: Evaluation exercise

Prof. Roy Ruddle

Alejandra Velasco Zarate - 201891211
Fernanda Alanis Montfort - 201899603
Louie Dove - 201497236
Corey Knight - 201497161

Leeds, Yorkshire 18th Nov

1 Introduction

Our experiment aims to systematically investigate the effectiveness of heatmaps and scatterplots in visualising pupil absences across a hypothetical dataset encompassing ten schools over an academic year. By assessing participants accuracy, response time, and the ability to interpret the data, our research seeks to determine which method is able to portray the information about pupil absences with a greater impact.

A within-participants design has been selected, allowing each participant to experience both methods of visualisation and reducing variability between personal differences (e.g. reaction time). The experiment consists of a total of 20 multiple-choice questions, of which 10 are associated with scatter plots and the other 10 with heat maps, both types of visualisations will be presented in a black-and-white scale to eliminate any potential negative consequences related to colour perception, thereby accommodating individuals who may have difficulty distinguishing colours. The questions are interleaved so that one question on scatter plots is followed by another on heat maps, and so on. Additionally, there will be a 3-seconds white screen displayed between each question. This structure is implemented in order to mitigate the effect of participant fatigue during the test while also preventing bias to the results and keeping the user engaged. No personal data will be collected from participants and participants will not receive any type of compensation for their collaboration in the experiment. This task will take no longer than 35 minutes, although participants have the full right and freedom to withdraw from the experiment at any time if they experience discomfort or prefer not to continue.

2 Methodology

The method we employed to carry out this experiment consisted of the following:

2.1 Participants

We gathered 10 participants, all university students, who provided verbal informed consent to participate in the experiment, after being thoroughly briefed on the purpose of the experiment, the task they would perform, and the fact that there would not be any kind of compensation. We also assured them that no personal information would be stored and that their anonymity would be preserved throughout the experiment, analysis and reporting stages. To ensure a controlled experiment, all relevant information and instructions provided to participants are duly documented in a 'Participation Information Sheet'. This document ensures that participants clearly understand the objectives, procedures, conditions of participation, as well as their rights and responsibilities during the conduct of the experiment.

2.2 Materials

For the development of the experiment's software, Python programming language was used, known for its versatility and wide support community. Among the main libraries used were Streamlit, Matplotlib, Pandas and NumPy, as well as some auxiliary libraries such as random, time and path, which allowed the implementation of additional functions necessary for the operation of the program. Pandas was used for data manipulation and analysis, thanks to its ability to work with efficient data structures and perform complex operations in a simple way. NumPy facilitated the

handling of arrays and mathematical operations, while Matplotlib took care of the graphical visualisation, generating high quality static graphics. The user interface was built with Streamlit, which made it possible to create an interactive web application with an easy-to-use graphical interface, without the need for advanced front-end development skills. These libraries ensure the proper functioning of the experiment software and its implementation.

The experiment software begins with a start page that provides instructions to the user, explaining that they will answer questions about school absences presented in either a scatterplot or heatmap format. A "Start Experiment" button is displayed on this page, and when clicked, the experiment begins, initialising the session state and setting up the first question. Once the experiment starts, the software generates random data for school absences across several months, using NumPy to create random values and Pandas to store and manipulate the data. For each question, the program alternates between generating a scatterplot or heatmap, depending on whether the question number is odd or even. The question is dynamically created based on the data, asking participants to identify the school with the highest or lowest absences for a specific month, or the month with the highest or lowest absences for a particular school. Participants are presented with the visualisation and must select an answer from multiple choices. The software tracks the time taken for the participant to answer, checks if the response is correct, and stores the results (including the time taken and the correct/incorrect answer) in a CSV file for later analysis. After each question, the participant is shown a brief white screen for 3 seconds before moving on to the next question. Once all 20 questions are answered, an end page is displayed, informing the participant that the experiment is complete and thanking them for their participation. The results are saved, and the experiment ends, with the user unable to return to previous questions.

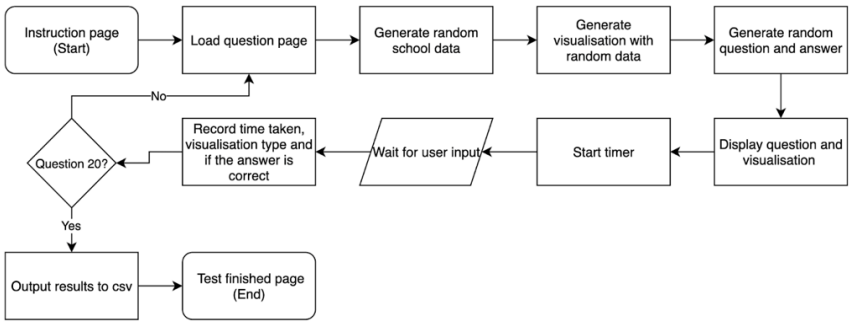


Figure 1: Flowchart of the Visualisation Evaluation Experiment

2.3 Procedure

Each participant followed a number of structured steps for the experiment, ranging from briefing to final debriefing. The following procedure outlines the participant's experience:

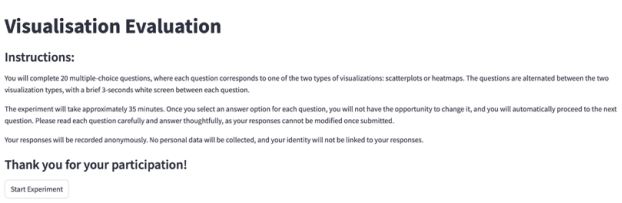
1. *Initial briefing.* Participants were welcomed and provided with the Participation Information Sheet, as shown in the Appendix 5. The participant read the information and then gave verbal informed consent to participate in the study.

2. *Start of the experiment.* Participants had been forwarded to the first question of the experiment by clicking on the "Start Experiment" button (Fig. 2a).

3. *Trial structure.* Each participant completed 20 trials, alternating between scatterplot and heatmap visualisations. The software dynamically created each question based on the random data. For each trial, participants were asked to read the question, view the visualisation and select an answer (Fig. 3). Following this, the participants saw a short white screen for 3 seconds before being forwarded to the next question.

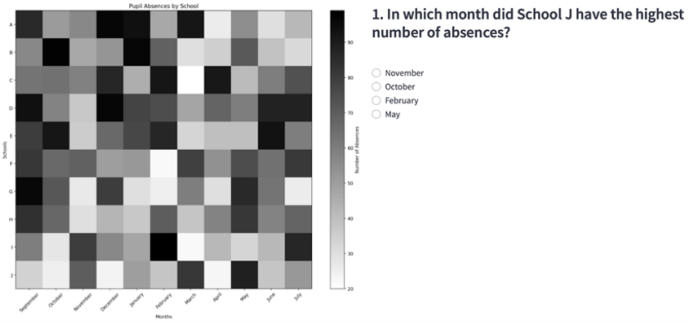
4. *Recording results.* After each question was answered, the software recorded the participant's ID, the question's number, the visualization type, time taken to give the answer and whether the answer was correct, data which was saved for future analysis.

5. *Completion of the experiment.* After completing all 20 questions, an End Page appeared, informing participants that the experiment was complete (Fig. 2b). The results were saved automatically to a CSV file.

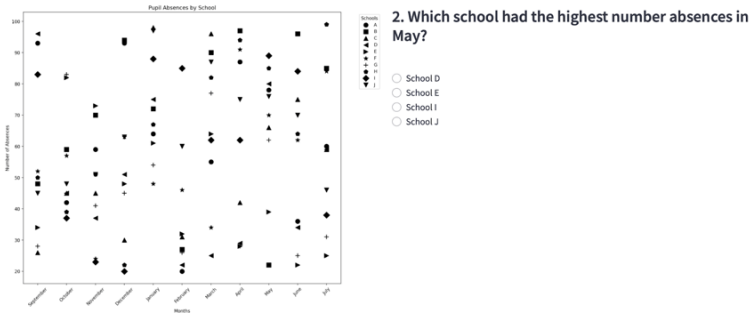


(a) Start page (b) End page

Figure 2: Experiment software interface start and end



(a) Software interface displaying heatmap visualisation



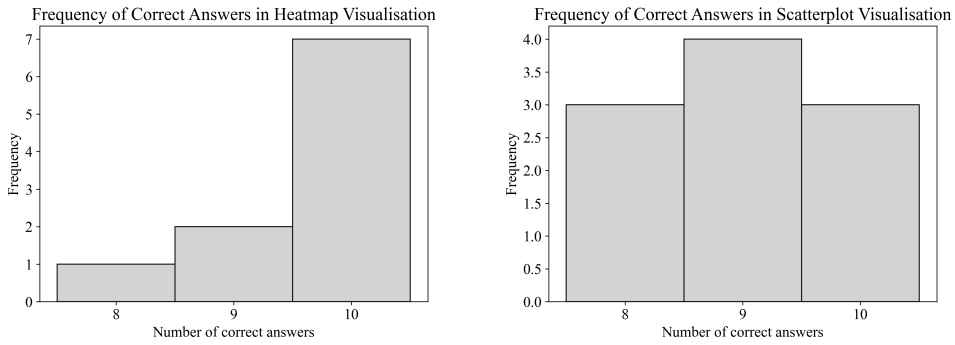
(b) Software interface displaying scatterplot visualisation

Figure 3: Experiment software interface during trials.

3 Results

The results were analysed using the pandas library in Python. As part of the data preparation for hypothesis testing and confidence interval estimation, the data was grouped by participant. This enabled the calculation of the percentage of correct responses for each visualisation type (heatmap and scatterplot), as well as the total time taken to answer the 10 questions related to each visualisation type. Separate columns were then created to represent the percentage of correct answers and the total time for each visualisation type for each participant.

Histograms were generated for the correct response columns to examine the distribution of correct answers per participant. This allowed for an analysis of the variability in participant performance for each type of visualisation (Fig. 4). A clear difference is observed in the distribution of correct responses between the two visualisation types, heatmap and scatterplot. In Figure 4a, which corresponds to the heatmap, most participants (7 out of 10) answered all questions correctly, with only one participant achieving 8 correct answers and two achieving 9. In contrast, in Figure 4b, which shows the scatterplot results, the most frequent number of correct responses was 9, with more participants achieving 8 correct answers and fewer achieving all correct answers. Comparing both histograms, it is evident that the frequency of correct responses is higher for the heatmap than for the scatterplot.



(a) Heatmap visualisation (b) Scatterplot visualisation

Figure 4: Frequency of correct answers by visualisation type

The visualisation of the frequency of correct responses paved the way for statistical tests, such as the Student's t-test, which in this case was used to determine whether the difference in the percentage of correct answers between the two visualisation types was statistically significant. The following hypotheses were defined based on the percentage of correct answers for each visualisation type:

- H_0: There is no significant difference in the percentage of correct answers between the two visualisation types.
- H_a: There is a significant difference in the percentage of correct answers between the two visualisation types.

The test yielded a t-statistic of $t = 1.62$ with a p-value of $p = 0.1405$. Given a 95% confidence level ($\alpha = 0.05$), the null hypothesis is not rejected, as $p > \alpha$. Therefore, we conclude that there

is no statistically significant difference between the two visualisation types. Despite the distinct distribution of correct answers for each visualisation, the statistical test indicates that this difference is not as substantial as it may initially appear. This result may be attributed to the fact that participants are able to answer the questions correctly if they allocate sufficient time to them.

To further investigate this, a second t-test was conducted, this time comparing the average time taken to answer the 10 questions related to each visualisation type. The test resulted in a t-statistic of $t = 3.77$, with a one-tailed p-value of $p_1 = 0.0044$ and a two-tailed p-value of $p_2 = 0.0022$. Since $p_1 < \alpha$, the null hypothesis is rejected, indicating a statistically significant difference in the average response time between the two visualisation types. For the two-tailed t-test, which assesses the direction of this difference, the following hypotheses were proposed:

- H_0: The average response time for the heatmap visualisation is greater than or equal to the average response time for the scatterplot visualisation.
- H_a: The average response time for the heatmap visualisation is less than the average response time for the scatterplot visualisation.

Since $p_2 < \alpha$, the null hypothesis is rejected, indicating that the average response time for the scatterplot visualisation is greater than that for the heatmap visualisation.

To complement these findings, Figure 5 displays the confidence intervals for the average percentage of correct responses for both visualisation types. As shown in the figure, with 95% confidence, participants' correct response rates for the heatmap visualisation will fall between $(91, 100)$, while for the scatterplot visualisation, the correct response rates will range between $(84.16, 95.84)$, although the confidence intervals largely overlap between the heatmap and scatterplot, suggesting no significant difference in accuracy.

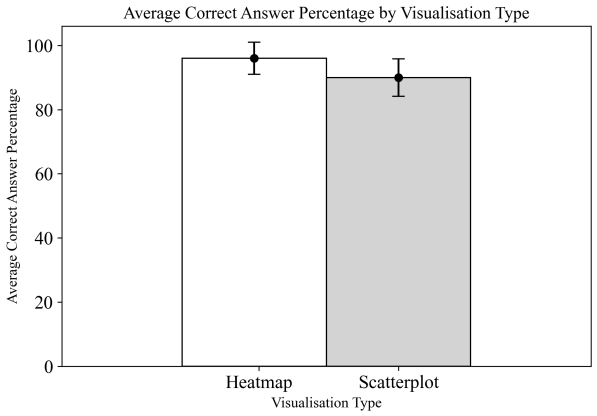


Figure 5: Average of correct answers percentage by visualisation type

In contrast, Figure 6 shows the confidence intervals for the average response time. With 95% confidence, the time taken to answer the 10 questions related to the heatmap visualisation is expected to fall between $(100.89, 151.73)$ seconds, while for the scatterplot visualisation, the time will range between $(139.86, 270.54)$ seconds. The fact that the confidence intervals for response

time do not overlap further supports the conclusion that the scatterplot visualisation requires more time to process than the heatmap.

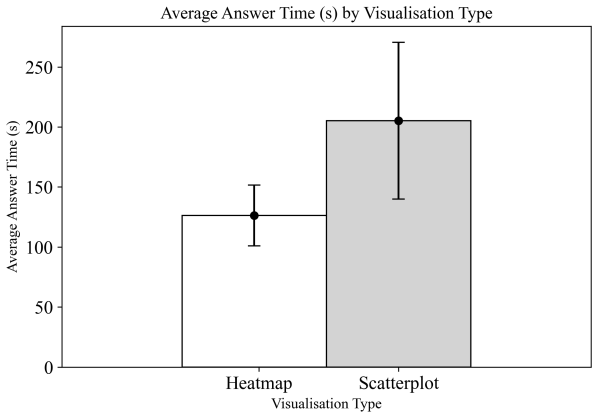


Figure 6: Average answer time by visualisation type

Thus, while there was no significant difference in the average percentage of correct responses between the two visualisations, a significant difference was found in the average time it took participants to comprehend the scatterplot compared to the heatmap. This suggests that although participants provided correct answers for both types of visualisation, the higher average response time for the scatterplot visualisation may indicate that participants needed more time to process and interpret this visualisation compared to the heatmap.

4 Conclusions

In this experiment, we explored the effectiveness of heatmaps and scatterplots in visualising pupil absences across a hypothetical dataset for ten schools over an academic year. By analysing participants' accuracy and response times, we aimed to assess which visualisation method provided a clearer, more impactful representation of the data.

The results show that, although the heatmap visualisation led to a higher percentage of correct answers, the difference in accuracy between the two methods was not statistically significant. Participants were generally more accurate when responding to questions based on the heatmap, with most answering all questions correctly, whereas the scatterplot resulted in more variation in responses. However, a significant difference was found in the average response time between the two visualisations. Participants required significantly more time to interpret the scatterplot compared to the heatmap. The scatterplot's increased processing time suggests that while it provided similar accuracy in answers, it was more complex to interpret and required more careful analysis.

Overall, the findings suggest that both visualisation methods can be effective in conveying information about pupil absences, but the heatmap is more efficient in terms of response time. The scatterplot, although equally accurate, demands more time for participants to comprehend. These results imply that heatmaps may be more suitable for scenarios where quick interpretation is important.

5 Appendix

Information Sheet: Visualization Experiment

Student: <Name>
Address: School of Computing, University of Leeds

I am a student at the University of Leeds, conducting research on how to design effective visualizations for descriptive statistics. This experiment aims to evaluate the clarity and effectiveness of two visualization types (scatterplots and heatmaps) in representing student absences in a hypothetical dataset covering 10 schools over an academic year. Your participation will help us assess which visualization method better supports accurate and quick data interpretation. If you volunteer for the experiment, you will complete 20 multiple-choice questions, where each question corresponds to one of the two types of visualizations: scatterplots or heatmaps. The questions are alternated between the two visualization types, with a brief 3-seconds white screen between each question and are designed to measure how effectively each type of visualization conveys information on student absences.

The experiment will take approximately 35 minutes. Once you select an answer option for each question, you will not have the opportunity to change it, and you will automatically proceed to the next question. Please read each question carefully and answer thoughtfully, as your responses cannot be modified once submitted. Your responses will be recorded anonymously. No personal data will be collected, and your identity will not be linked to your responses. The results may be reported in student research work, but you will not be identifiable in any way. Please note that you will not receive any financial compensation or other form of reward for participating in this study.

By participating in this experiment, you are giving your informed consent to be part of this research. However, you are free to withdraw at any time without any negative consequences or impact on your participation.

If you have any questions about the experiment or wish to discuss any aspect of the study, please feel free to ask me. Thank you for considering participation in this research. Your contribution is highly valuable for advancing our understanding of effective data visualization techniques.