201588070, 201636200, 201587988, 201599713

COMPARING THE EFFECTIVENESS OF SCATTERPLOT AND HEATMAP

1 INTRODUCTION

1.1 Aim

The goal of this study is to compare the effectiveness of scatterplot and heatmap in supporting extremum recognition tasks. Extremum recognition is the ability to identify maximum and minimum values quickly and accurately, which is a crucial skill in data interpretation to help extracting insights from complex datasets. This experiment seeks to assess which of these two visualisations better facilitates quick and accurate identification of extremum.

1.2 Design Experiment

Participants in this study were asked to interact with both scatterplot and heatmap, identifying maximum and minimum values across a range of datasets and task conditions. By using randomly generated datasets of intermediate difficulty, we aimed to minimise any potential advantage based on participant experience, creating a balanced testing environment that could yield generalisable findings.

1.3 Task

The tasks were designed to mimic low-level analytical activities commonly encountered in everyday data interpretation, focusing specifically on extremum detection. All visualisations were standardised in terms of size, font, and colour scheme to ensure consistency across tasks. Task order was randomised to prevent participants from inferring patterns, thus improving the reliability of responses. Through careful design adjustments, such as the inclusion of labels and appropriate colour schemes, we try to enhance clarity and usability in both scatterplots and heatmaps.

1.4 Duration

The experiment was structured to be completed within 20 minutes per participant. This time frame included a brief training session to acquaint participants with the visualisation types and tasks, followed by the main experimental tasks.

2 METHOD

2.1 Participants

Ten volunteers from the University of Leeds participated in the study. They received an information sheet and gave verbal consent before starting. Participation was voluntary, with no compensation provided.

To ensure anonymity and adhere to ethical guidelines, no personal or demographic data (e.g., age, gender, educational background) were collected. All participants completed 20 trials involving both heatmaps and scatterplots, with their responses and response times recorded for analysis.

2.2 Materials

The application used during participant assessments was created using Python and rendered to the web with HTML. Libraries imported include Dash and Plotly, both of which allowed the rendering of interactive plots onto the web, Pandas, which was used to read data from our tasks' Excel file, and NumPy, which helped structure indicated data. Some other libraries include time, random, csv, and os. These would help in the timekeeping of participants' engagement, creating a time interval between questions, generation of questions, and parsing through data files.

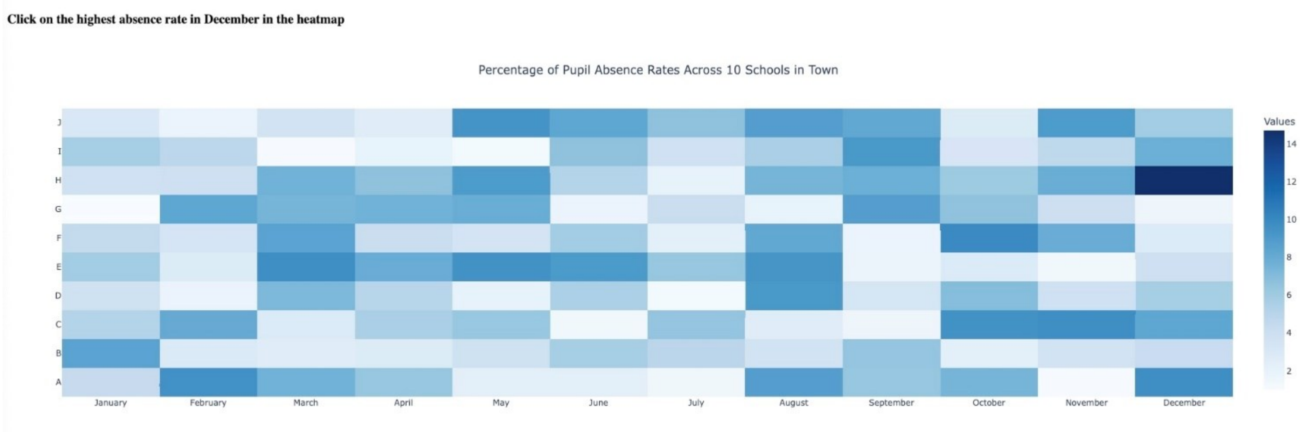

Figure 1: Example of the heatmap visualisation interface displaying the same dataset of pupil absence rates. The heatmap uses colour intensity to represent absence rates, with darker blues indicating higher values.

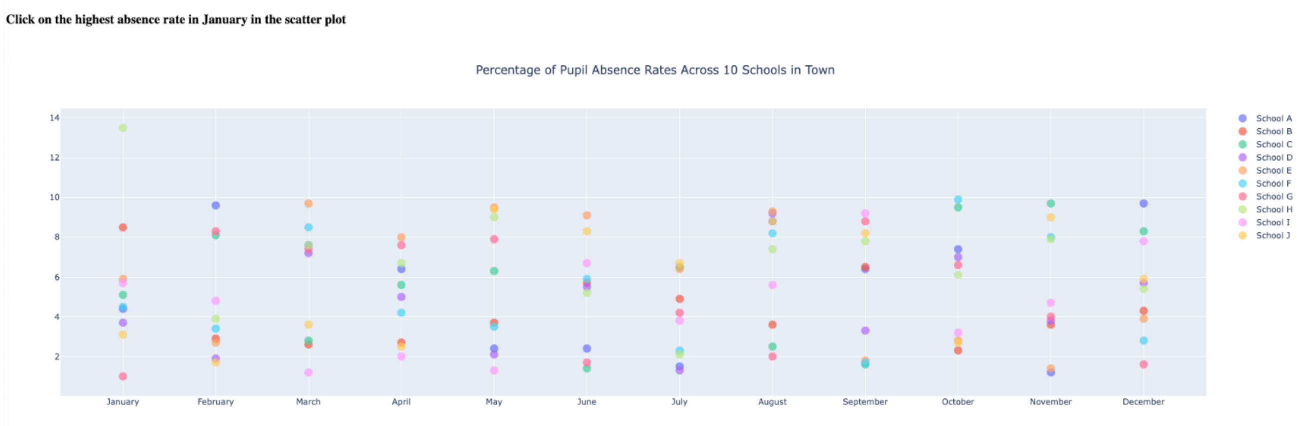

Figure 2: Example of the scatterplot visualisation interface used in the experiment, showing pupil absence rates across 10 schools over 12 months.

An Excel file was created with 10 separate sheets which each represent a different set of data and its pairing question. Tables of months by schools were populated with randomly generated data of percentages of student absences per sheet. Questions created used the template "Click on the {extreme_value} absence rate in {month} in the {chart_type}". The *extreme_value* depended on whether it was the first set of 5 sheets, for highest value, or the second set, for lowest value. The *month* would then be related to the corresponding *extreme_value* in that dataset sheet. Finally, the *chart_type* which could either be scatterplot or heatmap.

To initialise the program, this software reads the Excel and chooses one of the sheets to create a plot and present its question to the participant. As soon as it loads, a timer is started to record the time taken by a participant to complete the total of 20 questions. By using Dash and Plotly, they can directly click on each data marker or cell according to the instruction given. The software compares whether clicked data equates to the actual answer and adds to the correct answer count. A one second time interval leaves the screen blank in between each question after a data is clicked to reset the chart. By the end of the test, an "Experiment Complete" page is displayed along with data of average time taken, time taken to complete heatmap questions and time taken to complete scatterplot questions, as well as the correct answers of each chart type and total correct answers.

The results of each trial are appended into a CSV file with no identifying data to promote anonymity. Other precautions were taken as well to ensure the software worked properly and produced a constant fair environment.

We ensured that each dataset is only chosen two times and both scatterplot and heatmap are created for them. To promote fairness, one dataset will not be shown consecutively in both chart forms to avoid familiarity of questions.

2.3 Procedure

Participants underwent a structured 20-minute experiment to compare the effectiveness of heatmaps and scatterplots in visualising pupil absence rates across ten schools over a school year. The experiment took place in a quiet room to minimize distractions.

2.3.1 Briefing and Instructions

Participants were welcomed, read an information sheet, and received a standardized briefing using slides, which covered:

- **Overview:** Explanation of the study's goal to compare visualisation methods for interpreting pupil absence rates.
- **Task Description:** Participants would view heatmaps or scatterplots and identify the school with the highest or lowest absence rate for a given month.
- **Interaction Instructions:**
    - **Heatmaps:** Click on the cell intersecting the specified month and school.
    - **Scatterplots:** Click on the data point representing the chosen school's absence rate.
- **Emphasis on Speed and Accuracy:** Participants were advised to respond quickly and accurately, as both metrics were recorded.
- **Software Overview:** Screenshots of the Dash Plotly interface highlighted key features, such as submitting responses and proceeding to the next trial.
- **Questions:** Participants could ask questions before starting.

2.3.2 Experimental Trials

Each participant completed **20 trials** (10 with heatmaps and 10 with scatterplots), with the order randomised to prevent order effects. Each trial displayed a chart showing absence rates for all ten schools over 12 months. Participants were prompted to identify the school with the highest or lowest absence rate for a specified month and click on their selection. The software recorded the accuracy and response time. A 1-second blank screen between trials helped reset attention.

2.3.3 Post-Experiment Debriefing

After completing all trials, participants were reminded that their data would remain anonymous and were once again thanked for their valuable contribution to the research.

2.3.4 Controls and Considerations

- **Standardisation:** Uniform instructions and equipment were used.
- **Environment:** The room conditions were consistent for all participants.
- **Technical Support:** One team member was present for technical assistance, minimising interaction to avoid influencing performance.
- **Data Security:** Data were securely stored and accessed only by our team for analysis.

3 RESULTS

The results were analysed using **Analysis of Variance (ANOVA)** by performing t-test and 95% confidence intervals for both response time and accuracy when using scatterplot and heatmap.

3.1 ANOVA on response time

Null hypothesis (H_0): There is no significant difference in response times between heatmaps and scatterplots visualisations.

Alternative hypothesis (H_1): There is a significant difference in response time between heatmaps and scatterplots visualisations.

	Mean	N	Std. Deviation	95% Confidence Interval	
				Lower	Upper
Heatmap	**8.428**	**10**	**3.866143**	**5.662328**	**11.19367**
Scatterplot	**7.579**	**10**	**2.888158**	**5.512937**	**9.645063**

Table 1: Paired sample statistics for response time comparison between heatmap and scatterplot visualisations.

	Paired difference				t	df	p-value
	Mean	Std. Deviation	95% Confidence Interval				
			Lower	Upper			
Heatmap and Scatterplot	**0.849**	**0.977985**	**0.149391**	**1.548609**	**1.043002**	**9**	**0.324152**

Table 2: Paired sample tests for response time comparison between heatmap and scatterplot visualisations.

A paired sample t-test was performed to compare the response time when using heatmap and scatterplot visualisation to identify extremum value. There is a difference in response time between heatmap (mean = 8.428, std dev = 3.866143) and scatterplot (mean = 7.579, std dev = 2.888158); t(9) = 1.043002, p = 0.324152.

The p-value is much greater than conventional significance level of 0.05, indicating that we fail to reject the null hypothesis for response time. There is not enough evidence to conclude a statistically significant difference in responses time between the visualisations. The observed difference in response time is likely due to random variation rather than a meaningful effect of visualisation type.

The analysis yielded a mean difference of 0.849 where scatterplot has faster response times than heatmaps. The 95% confidence interval in the heatmap [5.662, 11.194] and scatterplot [5.513, 9.645] conditions overlap, suggesting no strong evidence of a difference between them. The confidence interval for the difference in means [0.149, 1.549] suggests a small difference with heatmap possibly being slightly slower. However, this difference is not statistically significant according to the t-test. We conclude that there is insufficient evidence to claim that there is a difference in response time of using scatterplot or heatmap in detecting extremum value.

3. 2 ANOVA on accuracy

Null hypothesis (H_0): There is no significant difference in accuracy between heatmaps and scatterplots visualisations.

Alternative hypothesis (H_1): There is a significant difference in accuracy between heatmaps and scatterplots visualisations.

	Mean	N	Std. Deviation	95% Confidence Interval	
				Lower	Upper
Heatmap	**97**	**10**	**6.749486**	**92.17171**	**100**
Scatterplot	**94**	**10**	**15.77621**	**82.71438**	**100**

Table 3: Paired sample statistics for accuracy comparison between heatmap and scatterplot visualisations.

	Paired difference				t	df	p-value
	Mean	Std. Deviation	95% Confidence Interval		t	df	p-value
			Lower	Upper			
Heatmap and Scatterplot	**3**	**9.02673**	**-3.45733**	**9.457332**	**0.557086**	**9**	**0.591051**

Table 4: Paired sample tests for accuracy comparison between heatmap and scatterplot visualisations.

A paired sample t-test was performed to compare the accuracy when using heatmap and scatterplot visualisation to identify extremum value. There is a difference in accuracy between heatmap (mean = 97, std dev = 6.749486) and scatterplot (mean = 94, std dev = 15.77621); $t(9) = 0.557086$, $p = 0.591051$.

The p-value is much greater than conventional significance level of 0.05, indicating that we fail to reject the null hypothesis for response time. There is no statistically significant difference between using heatmap and scatterplot in terms of accuracy. The observed difference in response time is likely due to random variation rather than a meaningful effect of visualisation type.

The analysis yielded a mean difference of 3% where heatmap has higher accuracy than scatterplot. The 95% confidence interval in heatmap is [92.172%, 100%] and scatterplot is [82.714%, 100%]. The upper interval is adjusted because accuracy cannot exceed 100%. Although the mean accuracy for heatmap is higher, the confidence interval for the paired difference is [-3.45733%, 9.457322%] which includes zero that is a null value. There is no statistically meaningful difference between the two visualisations. We conclude that there is no true effect to accuracy of using heatmap or scatterplot in detecting extremum value.

3. 3 Discussion

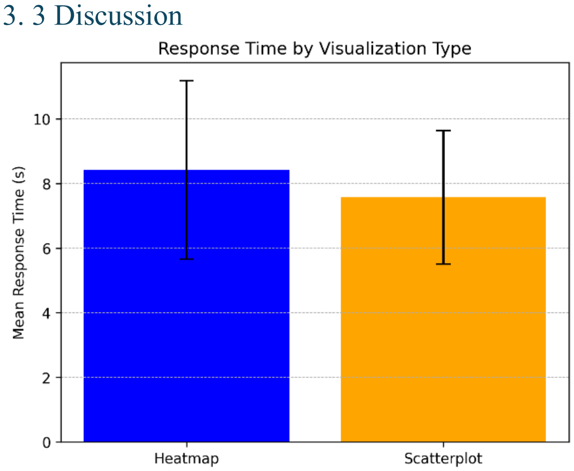

Figure 3: Mean response time comparison between heatmap and scatterplot visualisations with 95% Confidence Interval error bars.

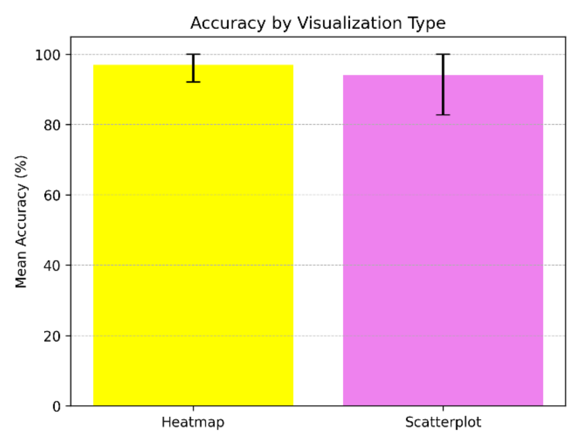

Figure 4: Mean accuracy comparison between heatmap and scatterplot visualisations with 95% Confidence Interval error bars.

While the statistical analysis found no significant differences between heatmap and scatterplot, qualitative observation of the result data suggests some variation in individual preferences and effectiveness across participants. Several participants achieved perfect accuracy with heatmap but had errors with the scatterplot. Conversely, some participants performed better with scatterplot, indicating that both visualisations have unique strengths that may align differently with each individual. This suggests that individual differences in preference and familiarity may influence performance.

These observations lead to an inference that visualisation effectiveness may depend on perceptual tendencies or prior experience with similar visual formats. Participants who found heatmap more intuitive may have benefitted from its continuous colour gradient which make extremum values stand out more. Conversely, those who performed better with scatterplot might find point-based data more straightforward for locating discrete extremum identification.

4 CONCLUSION

The study aimed to compare the effectiveness of heatmap and scatterplot visualisations in terms of response time and accuracy for identifying extremum values. Statistical analysis including paired sample t-tests and confidence intervals, found no significant differences between the two visualisation types for either response time or accuracy. Although slight mean differences were observed, they were small and not statistically significant. Observational insights reveal that individual performance may vary. The result of this experiment suggests that both visualisations are similarly effective for this task and any observed performance differences are likely due to random variation rather than a true effect. We conclude that neither heatmap nor scatterplot has a definitive advantage in supporting accurate or rapid identification of extremum values while user-centred preference should be considered to optimise performance outcomes.

5 APPENDIX

5.1 Participant information sheet

Students: Muhamad Ihtisyam Bin Shahrulnizam, Nadhilah Binti Mohd Nazri, Nurrin Shabihah Binti Lukman, Shasha Nabila Binti Al Malek Faisal

Address: School of Computer Science, University of Leeds

Purpose of the Study:
We are students at the University of Leeds conducting research on how different types of data visualisations affect the interpretation of information. This study aims to compare heatmaps and scatterplots in terms of how effectively they help users identify specific data points, specifically pupil absence rates across schools over a school year.

What Will Happen During The Experiment:
If you volunteer for this experiment, you will perform a series of tasks involving data visualisations of pupil absence rates in 10 schools over a school year. In each trial:

- You will be presented with a chart (either a heatmap or a scatterplot) displaying absence rates for all schools across all months.
- A prompt will ask you to identify the school with the highest absence rate in a specified month.
- Your task is to click on the part of the visualisation that corresponds to your chosen school for that month.

The experiment consists of **20 trials** (10 with heatmaps and 10 with scatterplots) and will take approximately **20 minutes** to complete. There will be a brief 1-second blank screen between each trial to reset your attention.

Data Collection and Confidentiality:

- **Responses Recorded:** Your selections (correct or incorrect) and the time it takes for you to respond will be recorded.
- **Anonymity:** No personal data will be collected. You will remain completely anonymous throughout the study.
- **Use of Data:** The data collected will be used solely for academic purposes in our coursework report. At no point will your identity be disclosed, and no one will be able to identify you from the reported results.

Voluntary Participation and Withdrawal:

- **Voluntary Participation:** Your participation in this experiment is entirely voluntary.
- **Right to Withdraw:** You are free to withdraw from the experiment at any time without any consequences or explanation.
- **Data Deletion**: If you choose to withdraw, any data collected from you up to that point will be deleted and not used in the study.

Consent:
By proceeding with the experiment, you acknowledge that you have read and understood this information sheet and consent to participate in the study.

Questions and Contact Information:
Please let us know if you have any questions or would like to discuss anything with us.

Thank you for considering participating in this experiment. Your contribution is highly valuable to our research.