Compare The Ocular Behaviour of Novices and Visualisation Experts

MSc Data Science CSC8639 – Interim Report

Author – Harvey Yuan

Supervisor – Alma Cantu

Introduction

Data visualisations are quickly becoming increasingly prevalent in more and more industries [1] and has traditionally played a pivotal role in financial reporting [2], ranging from annual financial summaries to real time transactional volumes. The utilisation of data visualisations in financial reporting facilitates a clear and concise comprehension of intricate financial information. Complex datasets are transformed into visually intuitive representations, enabling users to discern patterns, relationships, and trends with ease. This clarity ensures that stakeholders, ranging from executives to investors, can readily grasp the essential message and gain valuable insights from the data.

The objective of this study is to analyse the eye gaze data of participants while they view three static images. One of the images is a standard income statement displaying fictitious data, while the other two images present the same data but in the form of visualisations.

The main goal of this research is to examine how participants eye movements and gaze patterns vary when exposed to different formats of presenting income statement data. By comparing their regions of interests and engagement across the three individual images, insights can be gained regarding the effectiveness of visualisation techniques in conveying financial information.

With the main objective of this study being to analyse and compare the collected data, the study also aims to establish a set of guidelines for financial reporting, specifically focusing on income statements. These guidelines are intended to improve the understanding and retention of information for users reading the information.

Aim and Objectives

Aim

The primary objective of this study is to examine and compare the eye gaze patterns between novices and experts when viewing financial data presented through data visualisations. The research aims to generate guidelines for the creation of effective data visuals. To accomplish this, the study will analyse the eye gaze data of participants as they view three distinct data visualisations: one in the form of a standard table for an income statement and two others that present the same data in a visualised format.

The research methodology involves the use of a Tobii X2-30 eye tracker to gather precise eye gaze data during the participants observation of the visuals. The Tobii eye tracker operates at a sampling rate of 30 Hz, allowing for the capture of 30 eye data points per second [3]. This high sampling rate ensures accurate tracking of participants eye movements, providing valuable insights into their visual attention and gaze behaviour.

By comparing the eye gaze patterns of novices and experts when exposed to different data visualisations of financial information, the study aims to uncover variations in their visual engagement, regions of interest and information processing. This analysis will contribute to understanding how individuals with varying levels of financial expertise perceive and interpret data visualisations. A study published in 2021 [4] which studied the visual scanning strategies of pilots determined that pilots with higher levels of expertise would spend less time on each fixation but return to those regions of interests more frequently. This suggests that the experts are more efficient at perceiving the information provided to them.

With this in mind, the hypothesis of this study is that individuals with high levels of expertise both in financial data and data visualisations will be able to understand and obtain data from a traditional

income statement at a faster rate, whereas participants with less experience will take longer to obtain this information. Displaying the same information as a data visual, the participants with less experience with financial data will hopefully improve their experience with financial data.

In time, the study intends to derive practical guidelines for creating effective data visuals based on the analysis of eye gaze data between participants of varying expertise. These guidelines will serve as a resource for users requiring creating financial data visualisations, assisting them in creating visualisations that facilitate better comprehension and engagement amongst consumers.

Objectives

This study will consist of three main objectives:

- Creating data visuals
    - Ensuring data appropriate data is used to display the three types of visuals.
    - This should allow for appropriate comparisons.
- Collecting and analysing eye gaze data from participants
    - This consists of displaying the separate visuals to the participants and using the Tobii eye tracker to gather this data.
- Draw conclusions from analysis and gaze data

If the study progresses as intended, an added objective is to develop guidelines for creating impactful data visualisations specifically tailored to income statements.

Overview of Progress

Initially, the study had a broad scope with no specific field of focus. However, due to the increasing adherence to financial and accounting standards, the decision was made to narrow the investigation to financial data. Specifically, the study will delve into the visualisation of income statements as a focal point.

Preliminary research has been conducted in the field of financial data visualisation, focusing on gathering existing studies and their findings. As the study advances, additional research will be undertaken to enhance our understanding in this domain.

The study will aim to follow the guidelines and recommendations set out by the Nielsen Norman Group in their publication on "How to Conduct Eyetracking Studies" [5]. Specifically, the focus will be on sections 2 and 7 of the report, namely "Eyetracking Method Tips" and "Beginning the Session and Calibrating the User's Eyes" respectively. These sections provide valuable insights and best practices for establishing an optimal environment during an eye tracking study. By implementing these guidelines, it is anticipated that a consistent and replicable working environment will be created, leading to reliable and consistent outcomes.

The study will use the Tobii X2-30 eye tracker as the chosen equipment, ensuring consistent and reliable results throughout the research. To control the eye tracker, a modified version of the python set up code provided in the SDK setup page [6] will be used. The environment used to write the code is using python version 3.8 which allows for the "tobii_research" library to be downloaded and installed.

This code currently activates the tracker, collects data for a period of 30 seconds, and then deactivates it. Updates made to the code will include starting the tracker when a specific key is pressed on the keyboard and allowing it to be deactivated once the experiment is completed. Additionally, the existing code contains functionality to save the gathered data to a CSV file. The collected data will be pushed

to a GitHub repository, this allows for the raw data to be consolidated and prepared easily once the study moves into the analysis phase.

This experiment will involve participants from diverse backgrounds, ranging from university students and individuals currently employed in the field of finance. Their expertise levels in finance will be documented to enable comparisons between participants with varying levels of experience. Additionally, we will gather high-level demographic information, such as gender and ethnicity, to ensure the participant sample represents a wide range of demographics and reflects the broader population. This data collection approach will contribute to a more robust analysis more representative of the general public.

The data within the visualisations that will be shown to the participants in this study is fictitious data based on real world examples. While the data is fictitious, the data aims to mirror the type of information that the participants with finance experience encounters.

Furthermore, this approach will assist with the analysis by allowing a comparison of gaze patterns between participants without prior finance experience and a baseline group of individuals with finance expertise. On top of this, the data gathered can be comparted to the information presented in the other two visualisations which will allow for more comprehensive insights which will allow for more effective guidelines.

Project Plan

In the project plan, this study has a start date of 19th of May 2023 and is anticipated to conclude on the 21st of August 2023.

The selected start date of this project corresponds with the initial deliverable, which surrounded the completion of an ethical approval. It also corresponds with when the decision was made to narrow the focus of the investigation to the visualisation of income statements.

The research phase and the design data visualisation phase are intentionally overlapping as the idea is to continue to conduct additional research throughout the process of designing and building the data visualisations. This will allow for more granular approach, where new knowledge and insight is gained can go directly in the build of the visualisations.

Similarly, the research phase is designed to overlap with the data gathering phase. This overlap is driven by the same rationale applied to the research phase, allowing for continuous exploration and investigation during the data collection process.

Finally, the data analysis phase will allow for the evaluation of the data gathered during the previous phase. The plan is to extract meaningful insight from the data and using this insight, create guidelines which will allow for future users to create data visualisations which will covey the information in a way that will appeal to a larger audience, not just people with expertise in financial data.

References

[1] N. Singer, "When the Data Struts Its Stuff," The New York Times, 2 04 2011. [Online]. Available: https://www.nytimes.com/2011/04/03/business/03stream.html. [Accessed 06 2023].

[2] J. Rowan, "Data Visualization: Why a Picture Can Be Worth a Thousand Clicks," The Wall Street Journal/Deloitte, 17 11 2017. [Online]. Available: https://deloitte.wsj.com/articles/data-visualization-why-a-picture-can-be-worth-a-thousand-clicks-1510894925. [Accessed 06 2023].

[3] Tobii AB, "Specification of Gaze Accuracy and Gaze Precision, Tobii X2-30 Eye Tracker," 2013.

[4] V. P. M. C. Christophe Lounis , "Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study," 18 2 2021. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0247061#ack. [Accessed 5 2023].

[5] J. N. Kara Pernice, "How to Conduct Eyetracking Studies," 2009. [Online]. Available: https://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/. [Accessed 06 2023].

[6] Tobii AB, "Python - Getting started," Tobii AB. [Online]. [Accessed 05 2023].
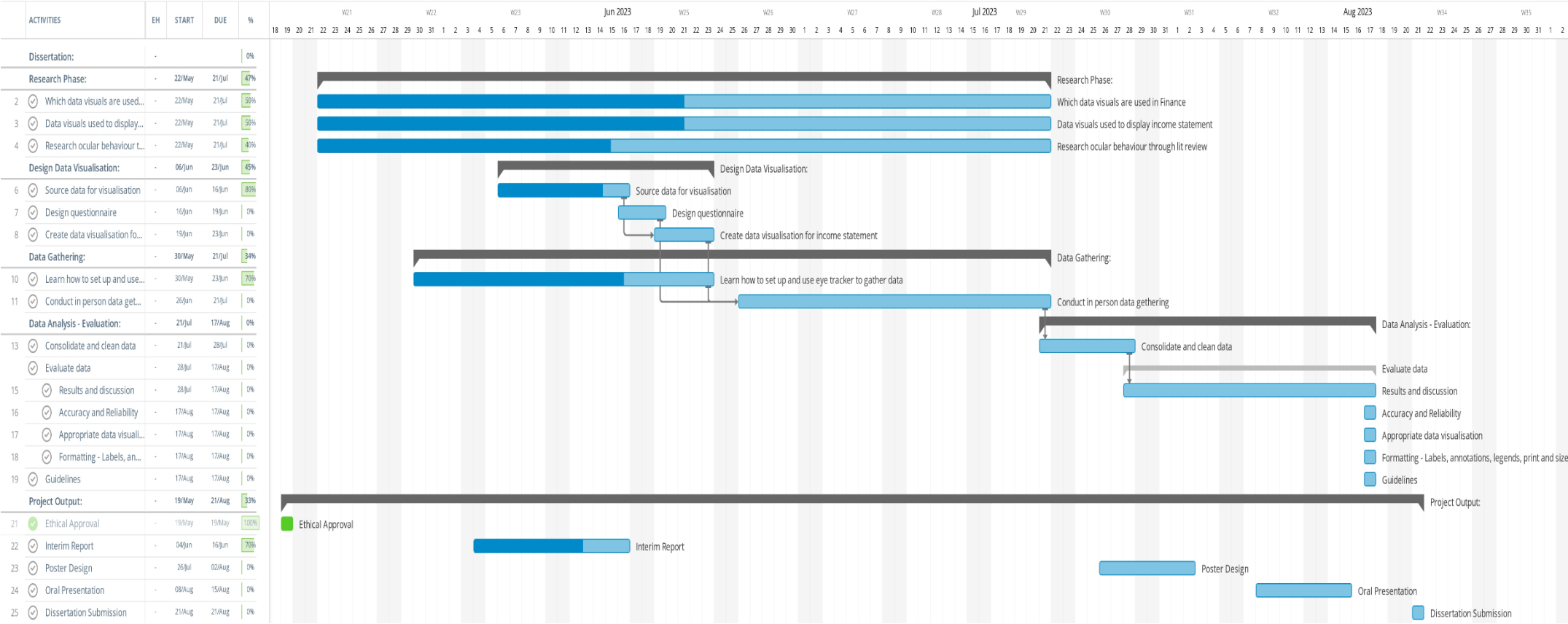
Figure 1 Project Plan

TEMPLATE FOR A DATA MANAGEMENT PLAN

The following **template** should be used to develop a Data Management Plan (DMP) to accompany a research proposal. The notes (*in italics*) provide further context and guidance for its completion. Where substantial data is generated from the research, the DMP will be more in depth and therefore likely to be 2 or 3 pages long [(3 pages maximum length - See MRC Je-S Help and Guidance for DMP)](#) for low impact studies generating small amounts of data, DMPs will be short ie less than half a page.

If you opt NOT to use the template the topics listed in the template MUST be addressed.

0. Proposal name
Compare the ocular behaviour of novices and visualisation experts

1. Description of the data
1.1   Type of study
Study the ocular behaviour between novices and experts for financial documents, specifically income statements from financial reports.
1.2   Types of data
The primary data that will be handled in this study is quantitative data, specifically collected during the participant's examination of a data visual using eye tracking technology.
Prior to the study, participants will complete a survey to record their expertise in data visualisation and financial data. Additionally, a post-study survey will be conducted to assess the usefulness of data visualisation for understanding income statements.
1.3   Format and scale of the data
The eye gaze data collected during the experiment will consist of x and y coordinates for each eye, recorded as tuples. This means that for every position the participant looks at, there will be four data points in total: two tuples for each eye. The Tobii X2-30 eye tracker used in this study has a sampling rate of 30 Hz, meaning that 30 data points will be collected every second. As the experiment duration is limited to a maximum of 1 minute, a maximum of 1800 data points will be generated for each run.

2. Data collection / generation
2.1   Methodologies for data collection / generation
The data for this study will be collected via 2 means, survey collection from the participants and the eye tracking data. To control the Tobii eye tracker, a Python script running on Python version 3.8 will be used. The script will automatically terminate the eye tracking data collection once the designated time limit has been reached.
The data will be stored on a CSV and the raw data will be moved to an R environment which will then be processed using R.
With regards to the survey, no identifiable information will be asked as part of this survey.
2.2   Data quality and standards The quality of data will be controlled by the same python script and eye tracker being used as part of this study. This will ensure the only variations will be from the physical environment such as lighting environment and participant. The eye tracker and the display will be set up using the same laptop to ensure the same device is used throughout.

3. Data management, documentation and curation

3.1 Managing, storing and curating data.

The eye gaze data will first be collected on the working laptop and then pushed and stored on a GitHub repository. When the data needs to be analysed, an R environment will be created and turned into a data frame meaning the raw data will remain untouched.

The survey data will be collected on the day of the experiment and will be saved in the repository. Once the data has been extracted from within, the surveys will be removed.

3.2 Metadata standards and data documentation

The methodology to gathering gaze data, code and hardware will be defined in the main report. This will allow for the procedure to be recreated. Data visualisations and tables will also be part of the report which will allow for comparisons in any future work.

3.3 Data preservation strategy and standards

Once the study has concluded, all raw data will remain on the GitHub repository. This will allow others to use this data and share their findings with the data.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

n/a

4.2 Main risks to data security

Confidentiality is not a concern since no identifiable information will be captured or stored as part of the study.

5. Data sharing and access

Identify any data repository (-ies) that are, or will be, entrusted with storing, curating and/or sharing data from your study, where they exist for particular disciplinary domains or data types. [Information on repositories is available here](#).

5.1 Suitability for sharing

The data is suitable for sharing as it contains onto eye gaze data. Survey data captured as part of this study will not be released, only summarised at a high level.

5.2 Discovery by potential users of the research data

If the outcome of this study is successful, an outcome of this will be to provide guidelines on creating data visualisations for income statements. This will allows for future users to search by financial data visualisation which the report and code will be freely available on GitHub. The sharing of gaze data will not be shared although the code, visualisation of gaze data and high level analysis of the participants will be shared.

5.3 Governance of access

As the data will be anonymous and not contain any identifiable information in it, the raw data will be freely available from the GitHub repository. Should the data need to be removed or if the repository need to be made private for any reason, the principal investigator and owner of the GitHub repository will be able to restrict access.

5.4 The study team's exclusive use of the data

The raw data will be made openly accessible through a dedicated GitHub repository, fostering a collaborative environment for potential users to engage in further analysis and study. This will

be made available from the start of the study which will encourage an environment for potential users to engage in further analysis and study.

5.5   Restrictions or delays to sharing, with planned actions to limit such restrictions

The study with each participant will only take place once they have consented to the use of the data to be collected and analysed. The participant will also need to consent for the raw eye gaze data to be stored on a GitHub repository which will allow other users to clone and download. Should the user agree for the data to be collected and analysed but not for the data to be stored publicly, their data will be removed from the GitHub repository.

5.6   Regulation of responsibilities of users

Indicate whether external users are (will be) bound by data sharing agreements, setting out their main responsibilities (please see page 13 section 7, titled Data-sharing agreements of the PDF file generated by selecting either of two links above).

Participants will be required to provide informed consent through consent forms. Only data from participants who have explicitly granted permission for their data to be stored on the GitHub repository will be retained and made publicly available. Consequently, external users accessing the data on the repository will not be bound by data sharing agreements, as the data provided to them will be anonymized and devoid of any identifiable information. This ensures that external users will not be in violation of the Data-sharing agreements, specifically Section 7 R13, as they will not have access to any personally identifiable information through the available data.

6.   Responsibilities

Apart from the PI, who is responsible at your organisation/within your consortia for:

- study-wide data management – Principal Investigator
- metadata creation – Principal Investigator
- data security – Principal Investigator
- quality assurance of data – Principal Investigator

7. Relevant institutional, departmental or study policies on data sharing and data security

Please complete, where such policies are (i) relevant to your study, and (ii) are in the public domain, e.g. accessible through the internet.

Policy	URL or Reference
Data Management Policy & Procedures	https://www.ncl.ac.uk/media/wwwnclacuk/research/files/ResearchDataManagementPolicy.pdf
Data Security Policy	https://www.ncl.ac.uk/data.protection/ https://www.ncl.ac.uk/data.protection/dataprotectionpolicy/
Data Sharing Policy	https://www.ncl.ac.uk/library/academics-and-researchers/research/rdm/sharing/#:~:text=The%20Research%20Data%20Service%20aims,between%20data%20users%20and%20creators.
Institutional Information Policy	

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details

Harvey Yuan