chapter{Data collection}

Background on survey design

As explained by Wiley-Interscience [2004], a survey is a means of obtaining quantitative information regarding opinions and experiences of the respondents in order to explore the views of the target population as a whole. In this book, a survey is noted as a "systematic" method of collecting data, where the author states that the word "systematic" is deliberately used in order to separate surveys from other methods of information collection. "systematic" is defined by the Collins English Dictionary as something that *"is done according to a fixed plan, in a thorough and efficient way"* [Collins], and this reflects the manner in which surveys are created in accordance with a given system, where methods for distribution, implementation and analysis are defined under a pre-determined structure. The survey will be delivered to potential respondents in the target population, who will then be asked to complete a series of standardised questions, or questions for which the question ordering and wording is identical for every respondent, unless different formats are to be used to research purposes. It is once again discussed by Wiley-Interscience [2004] that standardised questioning was not always the norm; most interviewers would more likely have a list of objectives, and each interviewer would formulate and word questions based around these. It was discovered that question wording can have a drastic effect on respondents' answers.

Whether or not the survey is 'thorough' and 'efficient' depends heavily on the survey structure and design. Designing an effective, systematic survey involves balancing efficiency with completeness, creating a survey that can obtain as much information as possible whilst not boring or fatiguing participants, which can lead to non-response and measurement errors due to participants skipping questions or selecting answers at random. A well-designed systematic survey has the capacity to yield large amounts of both qualitative and quantitative information regarding the research topic while minimising these errors.

There exist a variety of methods for delivering a survey, such as self-completed questionnaires and interviewer-administered interviews. Depending on the aims of the study, there will be advantages and disadvantages to each method. There may also be times when a combined approach is helpful in gathering the necessary information.

The first method of surveying, a questionnaire, may consist of either physical paper forms that are mailed or handed out to people within the target population, or in an online format. As discussed by Brace [2004], this form of surveying constitutes a method of indirect communication between the respondent and researcher, in effect a non-verbal conversation in which the respondent is replying to the researcher's questions. The non-face-to-face aspect of this method can be beneficial in terms of anonymity; an anonymous respondent is more likely to be honest in their answers than a respondent for whom the identity is known. As a result, an anonymous questionnaire can mitigate errors that may be caused by respondents fearing judgment of their answers. It is also possible to administer a large number of these questionnaires in a short period of time since they are self-administered, and thus constraints such as the number of interviewers or time taken to administer the survey has less effect on the amount of information obtained.

There are, however negatives to this questionnaire method. In his book, Brace discusses the way in which question wording must be

very carefully thought about when using this method of indirect conversation, for reasons such as there being no way to correct participant misunderstanding of questions. Additionally, the fact that the researcher and participant never come into contact may allow the researcher to write questions without considering the human nature of the participants; it is easy to become absorbed in attempting to gather information and fall into forgetting that long-winded or complicated questions may bore or confuse respondents, leading to poorer quality responses. Similarly including too many questions in the questionnaire may lead to response errors for the same reasons. It is then crucial to be as clear and concise as possible in question wording, leaving little room for interpretation. This type of survey is also a very static medium; it does not allow for much expansion on participants' answers, with reasoning behind answers unknown unless specifically requested, which again could add to respondent fatigue and affect quality of response.

We can attempt to implement some dynamic discussion into a questionnaire in the form of 'open-ended questions', mentioned above as specifically requesting reasoning behind answers. A questionnaire is composed of two types of questions; closed-ended questions, for which the respondent selects their answer from a given set of potential responses, and open-ended questions, in which the participants are able to write their answers in a free-form format. Closed-ended questions are very good for obtaining quantitative data that may be easily categorised and counted, which is useful for gathering empirical evidence in order to form objective conclusions regarding the sample population.

Open-ended questions are generally used where more expansion may be required in addition to the closed-form answer, or if using a closed-form question would limit the answer range. The Leibniz Institute for the Social Sciences [Züll, 2016] provides guidance on open-ended questions, in which the occasions for using open-ended questions are outlined as:

- "knowledge measurement"; with with multiple choice, respondents would have a chance of guessing the correct answer, and thus this would be a sub-optimal way to measure raw knowledge

- "Unknown range of possible answers"; multiple choice may be limiting for certain questions, and may cause the researcher to miss important information

- "Avoidance of excessively long lists of response options"; if there is a known range of answers, but this range is very large, it may overwhelm respondents to see all of these as options

- "Avoidance of directive questions"; certain questions may have options based on the researcher's own opinions, and thus have the potential to direct the participant in a certain direction, and may not reflect the participants' true views. This links to "unknown range of answers" in that the researcher may incorrectly assume the potential range of answers and thus the given options may not cover the respondents' true opinions.

- "Cognitive pretesting", which covers instances such as ensuring the question was understood correctly.

To summarise, open-ended questions are useful when either there is not enough information to set a standardised range of potential responses or if more information is needed after a closed-ended response.

A method of surveying that is, by design, more

dynamic is an interview. An interview may be structured, semi-structured or structured and each of these have a different set of features that distinguish them from one another. Structured interviews, as by the name, are rigid in nature and comprise of a vocal conversation in which the interviewer has a specific set of questions from which the discussion does not deviate. The slightly less rigid semi-structured interview is similar, but slight deviation from the plan is allowed in order to explore new avenues and ideas that might not be found with a structured interview, but the interviewer will still have a set of specific questions for which to obtain responses. For the most flexible of the three, the unstructured interview, the interviewer will tend to follow a loose plan of what they wish to explore rather than a strict question schedule, with the discussion led by the respondent's answers.

Phone calls and other forms of interview-based survey allow the interviewer to form a personal connection with the survey participant, which can be especially helpful for a company's image if the interviewer is particularly professional or charismatic. Additionally, while the interviewer will still be limited to asking the pre-set questions, the format of such a survey can be considered semi-structured and with much more room for interpretation. This can lend itself to gaining additional insights that may not have otherwise been gathered from a more closed-form paper or online survey. Additionally, the more open format can negate any error as a result of participants misinterpreting questions due to the interviewer's ability to immediately clarify on any misunderstandings. This type of survey also provides an instant response, which is beneficial if there is only a short time frame available in which to gather information.

However, there are also shortfalls to an interview-based survey method. For instance,

although a charismatic interviewer can positively impact the image of whoever is conducting the survey, this could also lead to biases, such as the respondent answering in a way they feel will please the interviewer. Additionally, the image of the organisation could potentially be tainted if the interviewer appears rude or unprofessional, alongside potentially providing bias in the opposite direction. As well as this, telephone surveys are likely to be interpreted as a telemarketing scheme, and thus potentially have a negative impact on the number of willing respondents. The reduced anonymity of this type of survey may also create bias in the way of participants avoiding making statements that could be deemed socially unacceptable, or that they feel they may be judged for, and therefore may not provide answers accurate to their true line of thought.

The UK Household Longitudinal Study [lon] is an ongoing study and an example of implementation of a combined use of the above mentioned surveying methods. Initially, in 'wave 1' of the study, a sample of 40,000 households in the UK were selected to be surveyed on a yearly basis. The survey involves all members of each selected household, overall comprising of around 100,000 individuals, and asks them a range of questions regarding areas such as family life, income, employment and health. The study consists of a self-administered youth paper questionnaire given to respondents ages 10-15, and an interview for those aged 16 and up. This split in age demographic allows some questions to be omitted from the youth survey, such as those about income and employment, and some to be added such as about pocket money habits and 'future intentions', as the website states. Giving the youth respondents a paper questionnaire may help obtain more useful or relevant answers, as the respondent may be more comfortable with this than being interviewed by an adult. The youth questionnaire is also shorter, which could perhaps just

be a result of many questions not being relevant to this demographic, or it could be a conscious decision, but either way this with help to ensure the young respondent doesn't lose interest and potentially incur bias in their answers due to either rushing to finish the survey or not paying attention. The adult survey also includes a section specific to 16-21 year olds. The surveys contain a standardised set of core questions asked each year alongside a set asked every other year. The reasoning behind this is given to be that this study has a very large scope, asking about many aspects of each respondents' life, and so it becomes inefficient and counterproductive to include all questions every year since, as mentioned previously, the longer a survey is, the more likely a respondent is to get bored or mentally fatigued. The fact that the adult survey is administered in an interview also means that there may be limits on the amount of time the survey can take, as interviewers may have to get through a certain number of respondents in a day, additionally to the interviewer potentially also becoming fatigued. If the interviewer is fatigued, their tone and how they hold themselves may change, and potentially cause a subconscious bias in how the respondent answers the questions.

Specific goals of survey tool for this study

While visualisations can be a very useful tool for understanding data, they also have the potential to be highly misleading. This section of the study will explore how modifying certain aesthetic features of visualisations can impact perception and interpretation of data, and how these modifications can be exploited in order to mislead the observer. Misleading visualisations may be created in an effort to deliberately influence the viewers' perceptions,

or accidentally as a result of poor practice and knowledge surrounding data visualisation. In either case, visualisations have the ability to communicate different messages and stories depending on how they present the data to the observer. There is a large amount of research and literature surrounding this topic, both in terms of providing frameworks for good visualisation practice as well as looking into how various techniques are used to deceive viewers. Results from some of these papers will be replicated, as well as used to form hypotheses which this survey will investigate.

A large amount of the literature exploring misleading tactics in data visualisation focuses mainly on bar plots and line plots for categorical and time series data, and so this is what the survey will focus on. The specific aim of the survey is to test whether altering y-axis scaling, bar width, bar grouping method and colouring will have an impact on single data value interpretation and subjective interpretation of differences in data values.

Survey Design

The survey design will be inspired by a series of papers, all of which investigate how different aesthetic and design choices have the potential to mislead the observer or alter perception.

The 2020 paper "The Deceptive Potential of Common Design Tactics Used in Data Visualizations" [Lauer and O'Brien, 2020], as the title suggests, explores how using different design tactics may mislead the person seeing the visualisation. Similarly to "An Empirical Study of Data Visualisation", the Claire and O'Brian paper uses a survey to explore how deceptive visualisation techniques can be employed as well as their impact on perception of the data. The survey discussed in this paper presents the participant with four plots; a bar

plot, a line plot, a pie chart and a bubble plot. Additionally to changing aesthetic features of the plots themselves, the study investigates the use of exaggerated, leading titles, for example one control plot has the title " Home Sales Show Increase From 2015 - 2016", which is altered to"Huge Increase in Home Sales From 2015 – 2016¡". The control plots consist of using a y-axis scaling beginning at 0 for the bar and line plots, a standard pie chart, and a bubble plot with proportionally sized bubbles, all alongside the non-exaggerated titles. The altered plots involve truncating the y-scale for the bar and line plots, making the pie chart in 3D, and arbitrarily altering the sizes of the bubbles on the bubble plot. The altered plots are referred to as the"deceptive" plots. The survey used sets of plots as crossed between deceptive aesthetics and deceptive titles; two had control aesthetics, one with the control title and one for the exaggerated title, and two had deceptive aesthetics with one having the exaggerated titling.

With regard to truncated axes, Claire and O'Brian asked participants to subjectively judge the difference between two data points using a 6 point scale ranging from "a little" to "a lot". For both the bar plot and line plot it was found the the use of a truncated scale increases the perceived difference between the data points. The use of a truncated scale is also discussed by Yang et al. [2021], whereby 5 empirical studies were performed in order to assess the effect of altering the scale in this way. The first of the 5 studies once again assessed how large the difference between data points is perceived to be in the truncated plot as compared to a control, again using a subjective scale from "Not at all different" to "Extremely different" on a 7 point scale. This scale differed, however, in the way that a midpoint label of "Moderately different" was provided. The 7 point scale may be preferable to the 6 point scale as the 7 point has a defined

midpoint at 4, whereas the 6 point does not. This study once again concludes that the differences in data points tended to be perceived as larger than for the control plot. Alongside these studies, a 2014 blog post [Parikh, 2014] discusses axis truncation and its effect on perceived data point difference for bar plots alongside other aesthetic features. The first example shows how truncating the y-axis of a bar plot can over-exaggerate differences in the heights of the bars, perhaps leading to incorrect observations regarding comparisons of values within the data.

The paper **?** performs a similar study, but instead investigates the use of 'stack-scale', or 'stacked' bar charts and logarithmic scaling. The aim of the study was to explore whether stack-scale bar charts are an effective way to visualise large value data, which is less relevant to our study since we have relatively low-valued data compared to the paper, but nevertheless provides a framework for exploring the use of logarithmic scaling and stacked bars in a respondent study. Participants were shown three plots; a control with a linear scale, a bar plot using a stack-scale, and one with logarithmic scaling. The questions asked determined how the different scaling affected accuracy in reading individual values, interpreting differences in values and determining which time-step exhibits the largest difference in values.

Following this, questions will be included in part 1 the survey with the aim of gauging whether altering the y-scale to be truncated or logarithmic has an effect on user perception of difference in data point values, for both bar and line plots. These will ask the respondents to gauge both individual values and differences in values, with the former providing an open answer box in which the may type their answer to allow for maximum freedom and obtain their true observation, unimpeded by

the bias of having a specific set of numbers to pick from when their true observation may lie outside this range. The question for gauging difference perception will follow Lauer and O'Brien [2020] and Yang et al. [2021] in using a numbered scale from not much difference up to a large difference. The Yang et al. [2021] method of a 7-point scale will be employed here.

Additionally, stacked bar charts will be investigated, showing a comparison between using the stacking method as opposed to a grouped bar plot.

Although the content of the surveys for this study is not likely to be controversial or highly personal, anonymity is still important as the participants could otherwise potentially feel pressure to give a 'correct' answer, given the mathematical nature of the questions. Anonymity here means that this pressure is potentially reduced and thus the relevant measurement bias may be mitigated.

The decision to focus on the two plot types was made in part to follow the existing literature, but also to ensure the survey is not too long. As discussed in [CITE: https://ojs.ub.uni-konstanz.de/srm/article/view/7145], too long a survey can result in higher measurement error due to factors such as waning interest or mental fatigue of respondents, resulting in careless responding and non-response. This is also further explored in [CITE: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=3059&context=etd_all]. While [CITE: https://ojs.ub.uni-konstanz.de/srm/article/view/7145] does conclude that a 'split survey' design, where each respondent is only asked to answer a selection of questions from the whole set, is effective in reducing error while gathering large amount of information, this will not be employed here. The reasoning for this is that there will already

be a large set of different surveys being sent, and creating further splits could potentially lead to small sample sizes and thus inconclusive results. Additionally to this, the paper investigates how placement of questions in the survey can affect responses, concluding that questions asked later in the survey are more susceptible to bias.

%%%% TODO:

% Perform meta-analysis of different sources to find optimal number of survey questions? % "20 minute rule" -> Don't go beyond 20 mins % Work out how long each question will take... Guessing 1-2 mins per question => ~10-20 questions? % Long surveys can increase measurement error %

%%%% Potential sources about survey length %%%% % https://journals.sagepub.com/doi/pdf/10.2501/IJMR-2017-039 -> Study about ideal length of survey in terms of time % https://ojs.ub.uni-konstanz.de/srm/article/view/7145 -> Study about reducing measurement error due to survey length % https://corescholar.libraries.wright.edu/etd_all/1918/ -> About 'careless responding' and 'insufficient effort responding' % https://journals.sagepub.com/doi/abs/10.1177/089443930101900202

Two types of questionaire will be written for this study. The first type will present a series of fairly basic visualisations accompanied by questions created to guage whether each presented visualisation has an impact on how the data is interpreted by the survey participant.

The set will consist of three separate surveys, which will be identical up to the visualisation package used. Particularly, one will contain visualisations made with R's ggplot2, the next with matplotlib from Python, and the last with the JavaScript library D3. These sur-

veys will be distributed to the general public by sharing links on social media platforms such as Facebook. The reasoning behind creating three separate surveys in a variety of languages is to ascertain whether the language used influences the interpretation.

%TODO explain the second type of survey further The second type will be a single survey comparing the implementation of the different languages and will be distributed to a group specialising in visual analytics for pharmaceutical research. This second survey aims to explore opinions on the coding languages themselves, in terms of features such as readability, reproducibility and ease of implementation.

%TODO explain that I'm using questionaires and why

In the set of three surveys we aim to answer questions such as:

- In which form is the given information of interest most accurately interpreted by the viewer?

- What factors of a plot can bias interpretation?

- Does the visualisation tool used have an impact on interpretation?

- Does the tool used have an impact on opinions regarding aesthetic features?

% Idea: Since I have had to extract things from ninja warrior data (lots of data), could show data and then a plot to show how % visualisations are useful to find things you can't immediately see in the data itself. % Idea: Axes in time series plots - Use short and tall axes and ask which varies most? (Maybe use HR data from http://ecg.mit.edu/time-series/) % Idea: Distance between bars in barplot - further apart harder to interpret? %

Idea: Does uning a logarithmic scale impact interpretation?

The survey

Potentially will use google sheets and post all three, and ask people to complete one of them. %TODO Specify further

Demographic Questions

The questions below are used to

- Please enter your age

- If you are a university student or past university graduate please specify your area of study. (Drop down box: Science, Technology, Engineering, Maths, Arts, Social Sciences, Humanities, Business, N/A, Other (please specify))

- How strongly do you agree with each of the following statements? (Linear scale with 1 - 5, 1=strongly disagree, 5=strongly agree)

-

- I have good spatial awareness skills

-

- I have good observational skills

-

- I have good numerical skills

- Are you colourblind? (Checkbox: Yes, No, Prefer not to answer)

- Do you have any disorders that may affect visual processing? (this could be a general visual processing disorder or dyslexia,

dyscalculia etc) ((Checkbox: Yes, No, Prefer not to answer))

Bar Plot Questions

American Ninja Warrior

The following bar charts present information regarding how many times 4 obstacles were used over the course of 10 seasons of the TV show 'American Ninja Warrior'. Please note that the answers to this section are entirely subjective, and that there are no 'correct' answers.

The first three questions refer to this bar chart, bar chart A.

Figure

%Note: Deliberately design 'bad' as well as good plots?

%%%% Misleading aspects of plots: %%%%

% Axis Scales % - Sources: % - https://www.nature.com/articles/s41559-018-0610-7 % - https://web.archive.org/web/20101123050530/http://graphpad.com/faq/file/1487logaxes.pdf % - https://dl.acm.org/doi/fullHtml/10.1145/3231772 % - https://journals.sagepub.com/doi/pdf/10.1177/1050651920958392 % - https://journals.sagepub.com/doi/pdf/10.1177/1050651920958392 % -

%%%%

% TODO Write about choosing more 'basic' questions to cut down on number of questions as well as appealing to audience

Conclusion

Bibliography

Understanding society - the uk household longitudinal study. URL https://www.understandingsociety.ac.uk/.

Ian Brace. *Questionnaire Design: How To Plan, Structure And Write Survey Material for Effective Market Research*. 01 2004.

Collins. Systematic. In *Collins.com dictionary*. URL https://www.collinsdictionary.com/dictionary/english/systematic.

Claire Lauer and Shaun O'Brien. The deceptive potential of common design tactics used in data visualizations. In *Proceedings of the 38th ACM International Conference on Design of Communication*, SIGDOC '20, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450375252. doi: 10.1145/3380851.3416762. URL https://doi.org/10.1145/3380851.3416762.

Ravi Parikh. How to lie with data visualization, Apr 2014. URL https://heap.io/blog/data-stories/how-to-lie-with-data-visualization.

Wiley-Interscience. *Survey Methodology (Wiley Series in Survey Methodology)*. 01 2004.

Brenda W. Yang, Camila Vargas Restrepo, Matthew L. Stanley, and Elizabeth J. Marsh. Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition*, 2021. ISSN 2211-3681. doi: https://doi.org/10.1016/j.jarmac.2020.10.002. URL https://www.sciencedirect.com/science/article/pii/S2211368120300978.

C. Züll. Open-ended questions. *GESIS Survey Guidelines.*, 2016. doi: doi:10.15465/gesis-sg_en_002.