Introduction

Data visualisation is a method of conveying data in an easily digestible manner through graphics. It is an important aspect of data presentation and allows key information to be quickly identified by the observer. Very many subject areas rely on such visualisations to relay messages that may get lost or have less impact when presented as written word or raw numbers. The main objective of visualisation is to create figures that display the data accurately in an aesthetic manner, giving non-misleading messages in a format that is pleasing to the eye. A good visualisation strikes a balance between aesthetics and information, where the aesthetic features are designed such that they 'enhance the message of the visualisation' [Wilke, 2019]. An incorrect balance of aesthetics to information can lead to figures that are misleading, confusing, or unengaging. Wilke discusses the way in which, for example, a research scientist with limited design experience may produce a visualisation displaying the data in an informative manner, but fail to draw immediate attention to the desired message, and on the other hand, someone with a main interest in the aesthetic design of a visualisation could create a figure that is very pleasing to the eye, but create a misleading visualisation in the process.

This literature review will discuss a range of publications discussing various aspects of data visualisation with a focus on how poor or uninformed visualisation design can produce misleading figures, as well as how such visualisations may be abused to deliberately deceive the observer. Starting with publications discussing general good visualisation practice, the discussion will then lead on to look at studies investigating the implementation of different visualisation practices, from which inspiration will be drawn to design the study for this paper.

Good Visualisation Practice

The book *'Fundamentals of Data Visualization'* is renowned as *'an excellent reference about producing and understanding static figures, figures'* [see Bebeau, 2019] and described as being *'suitable to be used as a reference manual'* [see Hwang, 2020]. Thus, this book provides a good basis to understanding the principles behind data visualisation, and how to create effective, informative and aesthetic figures.

Studies in Visualisation

There is a large amount of research and literature surrounding the topic of misleading visualisations, looking into how various techniques can either deliberately or unintentionally deceive an observer in the data's message. 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 study and literature review will focus on.

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 Hlawatsch et al. [2013] 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. Motulsky [2009] additionally discusses the use of a logarithmic axis in bar plots, explaining how it is impossible for a zero value to be displayed on this axis, and thus the bar start points are arbitrary and produce an inaccurate representation of the bar height with relation to the true value. To quote the paper, *"Don't create bar graphs using a logarithmic axis if your goal is to honestly show the data"*. We see that the logarithmic scale makes the percieved difference appear smaller than in the control.

As well as scaling, another aspect of visualisation design that could potentially mislead the observer is bar width and aspect ratios. When adding a visualisation into a publication, re-sizing the visualisation to fit a specific gap may include altering the aspect ratio, in turn affecting the length to width ratio of the bars in a bar plot. As explored by Steven Few in a 2016 article for the *'Visual Business Intelligence Newsletter'* [Few, 2016], altering this ratio can affect viewer perception in the way of a narrower and taller image distorting bars to appear longer, and vice versa, meaning that perceived differences between bar heights may be affected. Part 2 of the survey was based around investigating this idea, alongside how the reading of exact values is affected.

An article from the University of Stuttgart [Huynh, 2017] gives an overview of many types of bar chart, including stacked and grouped bars. The author remarks that grouped bar charts may make the comparison of bars in the same category more difficult, while the stacked bar chart sacrifices ease of comparison of values in the bars for increased spacial efficiency. A 2018 work from the journal of *'Visual Informatics'* [Indratmo et al., 2018] also provides a discussion on the use of various forms of stacked and grouped bar charts and their efficacy. The paper notes how a classical stacked bar chart can be useful for overall comparisons as the height of the bar represents the value of the item, with the different attributes depicted as a segmentation of this single bar into different colours. When discussing grouped bar charts it is mentioned that stacked bar charts may be less useful when performing attribute comparisons, in other words comparisons between different categories on the same bar, as a result of the bar segments being non-aligned. This results in comparison taking the form of length judgment as opposed to position judgment. Cleveland and McGill in their 1984 article in the *'Journal of the American Statistical Association'* [**?**] discuss how judgments based on length are likely to be less accurate than those based on position. A grouped bar chart is a way to allow for easy comparison between individual categories, but is discussed to be less effective in overall comparison.

Bibliography

Christy M. Bebeau. "graphicacy for numeracy: Review of fundamentals of data visualization: A primer on making informative and compelling figures by claus o. wilke (2019).". *Numeracy 12*, 2, 2019. doi: https://doi.org/10.5038/1936-4660.12.2.18.

Steven Few. Bar widths and the spaces in between, 2016. URL http://perceptualedge.com/articles/visual_business_intelligence/bar_widths.pdf.

M. Hlawatsch, F. Sadlo, M. Burch, and D. Weiskopf. Scale-stack bar charts. *Computer Graphics Forum*, 32(3pt2):181–190, 2013. doi: https://doi.org/10.1111/cgf.12105. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12105.

Hai Dang Huynh. Two-dimensional bar charts. 2017. doi: http://dx.doi.org/10.18419/opus-9496. URL https://elib.uni-stuttgart.de/handle/11682/9513.

JP Hwang. Create impactful data visualizations with these books. Online, towards data science, 2020. URL https://towardsdatascience.com/create-impactful-data-visualizations-with-these-books-ca9fbfecfde5.

Indratmo, Lee Howorko, Joyce Maria Boedianto, and Ben Daniel. The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. *Visual Informatics*, 2(3):155–165, 2018. ISSN 2468-502X. doi: https://doi.org/10.1016/j.visinf.2018.09.002. URL https://www.sciencedirect.com/science/article/pii/S2468502X18300287.

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.

Harvey J. Motulsky. The use and abuse of logarithmic axes. Online, GraphPad Software, inc., 2009. URL https://web.archive.org/web/20101123050530/http://graphpad.com/faq/file/1487logaxes.pdf.

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

Claus O. Wilke. *Fundamentals of Data Visualization*. O'Reilly Media, Inc., 2019. ISBN 9781492031086. URL https://clauswilke.com/dataviz/.

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.