CSC 502 – MASTER'S PROJECT PROPOSAL SPRING 2021

Inculcating Ethics in Data Visualization Dashboards.

Under the Guidance of

Dr. Anna Baynes

Department of Computer Science

California State University, Sacramento

Proposed by

Abhinay Gupta Somisetty

301618980

TABLE OF CONTENTS

1. ABSTARCT

Any data which is visualized has a significant influence on its readers. But often it is discussed that the visual dashboards mislead or misinterpret the facts. This could lead to ethical issues. Such factors in the data visualizations have the potential of changing the audience's perception towards a data. This leads to critical ethical dilemmas which is the major concern. Zui Chen states that "Data visualization relies substantively on authors' choices of axes arrangement, such as truncation and scaling, as well as the shapes, including figure types and relative sizes [1]". Though, the deceptive and non-deceptive analysis are the two popular words among the data visualizers, it is often overlooked by the visualizers while working building the dashboards. In this project I propose a study which highlights the importance of education and avoid un-ethical/deceptive practices while building data visualization dashboards and build a dashboard to educate the visual researchers, visual developers, students and viewers.

2. MOTIVATION

An article published online [2] broadly explained how different charts or graphs mislead the viewers. The choice of chart type is essential. When using a wrong chart type, the resulting visualization can be highly ambiguous or deceiving. Suppose let us take an example of a pie chart in Fig 1. It is depicting the percentage of favourable votes to each contestant. For instance, when a reader sees the chart, they conclude that all the three contenders have the same probability of winning the election since the slices' area looks the same. However, if we carefully read the text, each slice of the pie is the same size, but the labels show differing values, which add up to 193%. Using a pie chart is wrong because each person polled was asked their opinion about each candidate. We should not expect the favorability values to add to 100%. It cannot be considered a mistake, but this violates Journalism ethics [9].
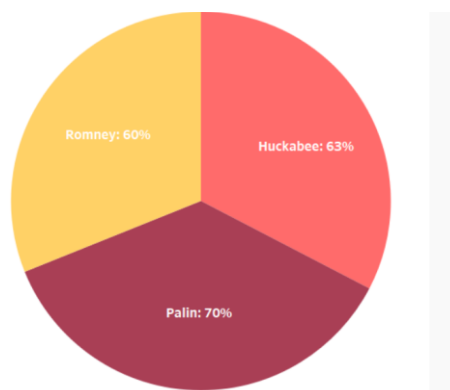
Fig 1: Pie chart misleading the data [3]

Similarly, a well know issue where visualizers try to manipulate the viewers by excluding a Zero in the bar chart graph. Below Fig 2 is one such example. This is published on German economic development agency GTAI website [3]. From the below graph it gives a larger perception that countries like Germany have an edge over countries such as Romania and France. But if we carefully observe, we could see that the chart did not start from zero, but it has cut off only 39. This is causing an illusion and misleading the viewers.
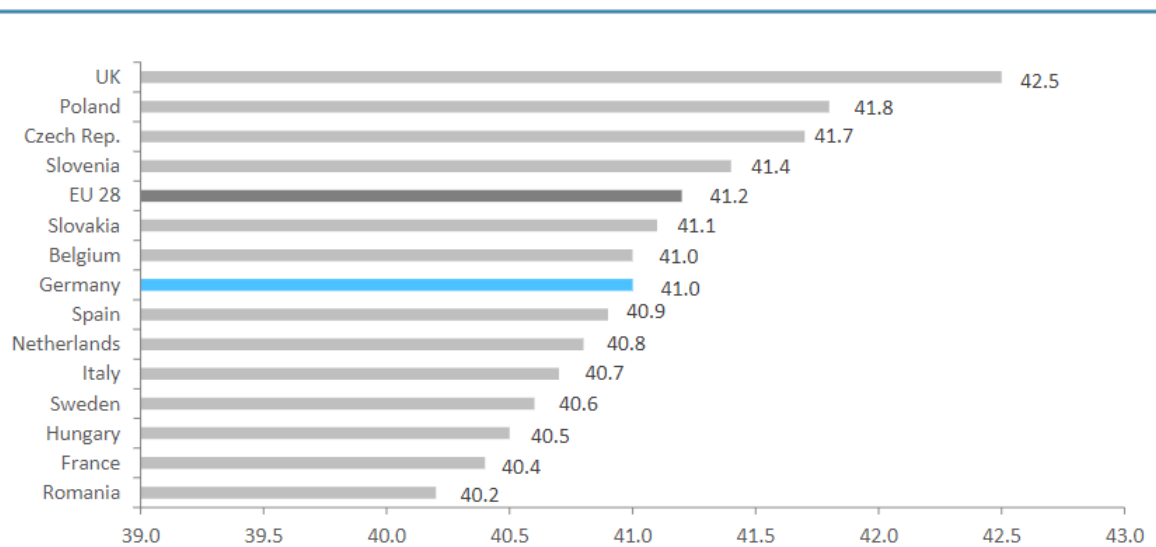

Fig 2: Representing the average number of actual weekly hours of work in main job, full-time employees, 2018.

3. OBJECTIVE

There are many scholarly articles and studies which talk about the ethical challenges in this domain. However, it is later realized that very few studies mainly focused on the interests of visualizers and viewers about the ethical challenges in data visualizations. In this project, I propose a method to handle the ethical challenges mentioned in the motivation section and educate the visualizers, the viewers about the ethical and non-ethical practices in data visualization.

In this project, I will use multiple visual encodings like bar charts, graphs and pie charts on different types of datasets such as census data, electoral poll data and other datasets that can mislead its viewers. These visualizations contain both static visuals and interactive visuals. Interactive visuals are designed using D3.js [6] and JavaScript, whereas static visuals are developed using Tableau [7].

4. METHODOLOGY

In this project, I propose a method to build multiple visuals which are deceptive and non-deceptive. Deceptive visuals can be created by excluding zero in the bar chart, using a pie chart, multiple axes on a single graph, adding 3-D exaggeration, inappropriate sizing of the charts. Research is conducted on these visuals, where I propose to survey certain groups of people by showing these visuals and asking for their feedback.

These groups are divided based on multiple factors. Considering the complexity of the project and the level of understanding, I would choose people/students who are above 18 years to be part of this survey. I tentatively plan to prepare a set of questions for the survey based on the deceptive and non-deceptive visuals we have developed. in this survey. This is one of the ways to understand the level of the deceptiveness of the data visualizations.

I propose to include some statistical data and provide an overview of the datasets or some text about the deceptive visual. This approach aims to identify if the data's unethical representations have the edge over the actual data or the textual.

Finally, create a website by analyzing the volunteers' feedback to educate the visual researchers, developers, students, and viewers about the ethical and non-ethical ways of representing the data and it's impact and severity in this domain.

5. DATASET

For the proposed idea, I would use different types of datasets. These datasets might be in the form of JSON, CSV, or text. Examples of these datasets are census dataset, demographic dataset, digital marketing trends data and election poll dataset. We can include other datasets as well to support the questions and survey process.

6. IMPLEMENTATION

Below steps are the proposed method for building the dashboard.

1. First step is to identify different types of datasets and collect the datasets which can potentially mislead the viewers.

2. This step will process the data by performing preprocessing techniques to eliminate or transform redundant, unwanted data and missing values from the data sources.

3. Study and learn different types of unethical practices to represent the data and its counterparts.

4. Now we build various ethical and misleading visuals of the obtained data. These visuals could be developed using static visuals or dynamic visuals.

5. Identify the questions to conduct the ethical vs unethical practices survey. These questions include but not limited to some of the user's personal information such as age, educational qualification and some information about the different datasets, textual representation of the data, and the different data visualizations we built using ethical and unethical practices.

6. Now identify different volunteers who are willing to participate in the survey with the consent and dividing these volunteers based on factors like age, educational background.

7. Based on the obtained feedback, analyze the severity of the impact of the un-ethical visuals and create a website including the various visual dashboards, survey research details which can be used to educate the researchers, visual developers, and viewers in this domain.

7. TECHNOLOGY STACK

- Data Pre-processing   :   Numpy, Pandas, Trifacta, MS Excel.

- Tools   :   Tableau, Pycharm, Jupyter Notebook.

- Frameworks   :   D3.js, Chart.js.

- Programming languages: Python, JavaScript, HTML5, CSS, Bootstrap.

8. TIMELINE

TASK	NO OF WEEKS
Idea and Proposal	1 week
Research on ethical challenges in Visualization	2 weeks
Collecting data	2 weeks
Preprocessing data	2 weeks
Researching Tableau/Implementation	3 weeks
Researching interactive framework	2 weeks
Developing Interactive Visual Dashboard	2 weeks
Work on the questions for the survey	3 weeks
Identify the volunteers and collective the feedback through survey	4 weeks
Building the final dashboard	3 weeks
Final Report	3 weeks

9. REFERENCES

[1] https://dvia.samizdat.co/2018/tag/r1/

[2] https://www.tessellationtech.io/top-five-ways-mislead-data-visualization/

[3] https://www.gtai.de/gtai-en/invest/business-location-germany/business-climate/motivated-and-dependable-employees-81844

[4] https://bib.dbvis.de/uploadedFiles/343.pdf

[5] Interactive Visuals - https://observablehq.com/

[6] Fan Bao and Jia Chen, "Visual framework for big data in d3.js," 2014 IEEE Workshop on Electronics, Computer and Applications, Ottawa, ON, 2014, pp. 47-50, doi: 10.1109/IWECA.2014.6845553.

[7] Getting started Tutorial – Tableau https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.htm

[8] https://learningsolutionsmag.com/articles/misleading-data-visualizations-can-confuse-deceive-learners

[9] Elizabeth Farley, Fiona Grady, Dean S Miller, Rory O'Connor, Howard Schneider, Michael Spikes, Constantia Constantinou, et al. 2014. SPJ Code of Ethics. The Power of Images (2014).

[10] Kirk, A. (2012). Data Visualization: a successful design process. Birmingham: Packt Publishing.

[11] https://www.mapbusinessonline.com/Whitepaper.aspx/Avoid-Data-Visualization-Misinformation

[12] https://cyber.harvard.edu/getinvolved/studygroups/datastorytelling_ethics

[13] https://arxiv.org/ftp/arxiv/papers/1808/1808.05686.pdf

[14]   A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite and E. Bertini, "The Persuasive Power of Data Visualization," in IEEE Transactions on Visualization and Computer Graphics, vol. 20, no. 12, pp. 2211-2220, 31 Dec. 2014, doi: 10.1109/TVCG.2014.2346419. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6876023

[15]