A brief study on data visualization techniques and its implementation using tableau software

[1] JIPSA KURIAN, [2] SONIA BHALLA,[3] SHIKHA TUTEJA
[1]*DEPARTMENT OF MATHEMATICS*, [2]*UNIVERSITY INSTITUTE OF SCIENCE*, [3]*ECE*
Chandigarh University,NH-95,Gharuan,Mohali,Punjab, (India),140413
[1]jipsakurian99@gmail.com , [2]sonia.e8843@cumail.in, [3]shikhatuteja85@gmail.com shikha.e12701@cumail.in

***Abstract*—Rather than data on unprocessed numbers, facts, and figures, everyone today is interested in well-formatted and fully depicted portraits. As a result, data visualization techniques are expanding into the commercial sphere. As it is simple for both technical and non-technical people to understand, it not only raises the bar for the business sector but also promotes numerous inventions and analyses. Currently, data visualization is at a high height, from commercial to notable scientific advances. This essay will emphasize the value of data analysis and the application of some practical and cutting-edge technologies that will raise the bar for all types of representations. Using a well-formatted superstore dataset as an example, Tableau software will be used to demonstrate its graphical application.**

KEYWORDS: Data visualization, tableau, charts, graphs, tables, bar charts, histogram, line charts.

I. INTRODUCTION

Data analysis has been discussed in a variety of ways by various authors. It is the representation of any type of data or information using some common images or graphics, such as infographics, graphics charts, and animations, in order for any type of reader and audience to understand [13], [14]. This makes it easier to analyze complex data relationships and to derive a variety of insights from the data, making it possible to foresee hidden trends in the data. As in the big data era, numerous elements like volume, veracity, diversity, and many more must be taken into account and are nearly difficult to assess using outdated methods and technologies. Even for expert users and analysts, it can be quite difficult to understand the underlying patterns in huge data. Therefore, visualization is a crucial component of any analysis. [14] There are several general purposes for data visualization.

These main goals are listed below:

- Idea illustration
- Visual discovery
- Idea generation

A. Idea illustration:

Data visualization based on ideas a method or procedure, for example, can be better understood when it is shown. In order to facilitate communication between the appropriate parties on certain tasks, it can also be used to illustrate organisational structures or procedures. It is frequently employed in educational settings such as tutorials, certification programmes, and centres of excellence. Abstraction is used to represent and to more fully comprehend the information management process within an enterprise's data flow through data modeling, enabling developers, data architects, business analysts, and others to have a better understanding of the links in a data repository.

B. Visual exploration:

Visual discovery and routine data visualisation are becoming more and more firmly linked with data teams. While data analysts, data scientists, and other professionals can identify trends and patterns in a dataset with the help of visual discovery, regular data visualisation supports the following narrative when a new insight has really been found.

C. Idea generation:

Data visualization is widely used to promote group brainstorming. During group brainstorming at the start of a project, they often used to encourage the gathering of various points of view and pay attention to the challenges that the group shares. Even though these visualizations are typically rough around the edges, they help set the project's foundation and ensure that the team is in agreement about the problem that they are working to resolve for key stakeholders. Well, there are several techniques to perform data visualization. We can find various proofs that visualization has been practiced since the 17th of century. With the running of several time periods and eras, people leveraged data analysis for multi-purposes and multi-disciplines such as finances, health sectors, logistics, medicines, geographic locations, and many more. There are thousands of data visualization techniques. Among them, the most common techniques will be discussed below:

eISSN 1303-5150

1) ***Tables:****:* This representation includes rows and columns which are commonly used in the representation and comparison of different dependent and independent variables. This technique aids in portraying data in a well-organized and simpler way. The image shown below is the table:

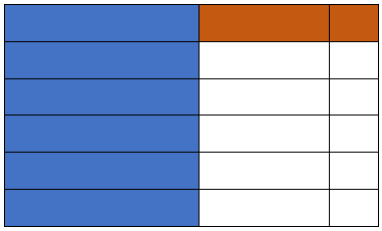

Fig. 1. Representation of the table

Here, blue represents columns while the brown one represents row

2) ***Area chart and line chart:****:* These graphics, which are commonly used in predictive analytics, by plotting a set of data points through time, you may show how one or more quantities have changed. Area charts stack parameters on top of one another, connect data points by line segments and display these changes using lines in line graphs, whereas and employ colour to distinguish between variables. Pictures given below are the examples of bar chart and area chart respectively:

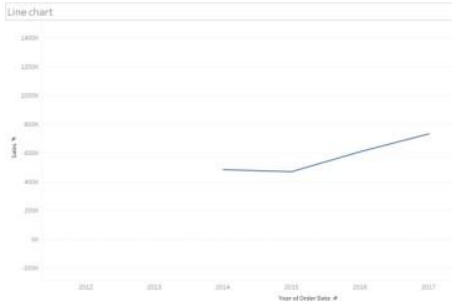

Fig. 2. Line chart

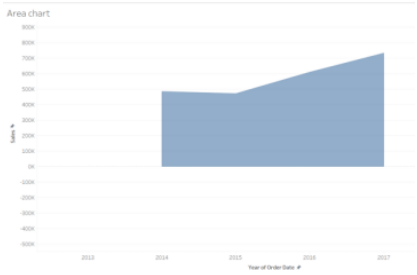

Fig. 3. Area chart

3) ***Histograms:****:* The volume of information falling inside a particular range is represented by this graph, which depicts a bar chart showing how the data are distributed (without spaces between the bars). With the help of this visual, a user can detect outliers with relative ease in a given dataset. The portal representation of the histogram is shown below:

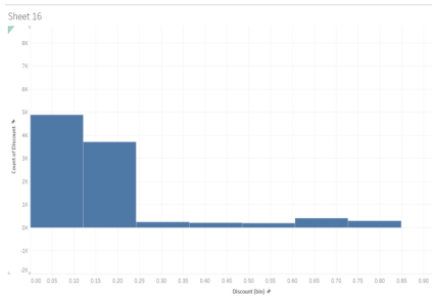

Fig. 4. Histograms

4) ***Scatter plots:****:* These illustrations are frequently used during regression data analysis because they are useful in revealing the connection between two elements. However, these can occasionally be confused with bubble charts, which use the x-axis, y-axis, and bubble size to represent three different variables. The picture below represents a scatter plot: All of the pictures explained above introduce data visualization

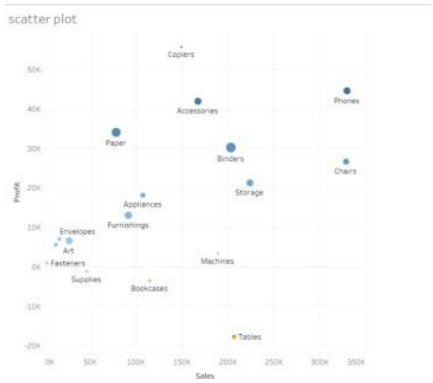

Fig. 5. Scatter plots

in a brief. In the paper ahead, several works done ahead on data visualizations will be discussed and a case study will be done relying on a particular dataset.

II. CASE STUDY AND ANALYSIS OF SEVERAL WORKS DONE ON DATA VISUALIZATION

A. Tableau:

Tableau is a fantastic business intelligence and data visualization application for reporting and analyzing enormous data sets. Tableau was established in the United States in 2003, and Salesforce acquired it in June 2019. In order to visualize and analyse data to support business decision-making, it helps users create a variety various charts, mappings, dashboards, and storylines. [1], [6]

eISSN 1303-5150

B. Domain dataset:

For the successful completion of review and research on this topic, I have chosen an eminent dataset known as 'Sample Superstore' which consists of several categorical, geographical, date, measures, and as well as parameter values which are almost efficient to perform a review on visualization techniques. Using this dataset, different visual representations and charts will be drawn via the help of Tableau public software. Some techniques that will be implemented are bar charts, pie charts, line charts, histograms, tree maps, bullet charts, and tables. Data given below represents data information about the dataset that will be going to be used. Below the table is the representation of the dataset which almost consists of around 10000 tuples of data. Generally, we cannot analyze this massive dataset with numerical representation and even if we try to understand it in numerical format, it will consume a massive amount of time to perform simple analysis. Thus, pictorial representation plays a significant role, and data visualization technique comes into existence. The portrait above is taken as a

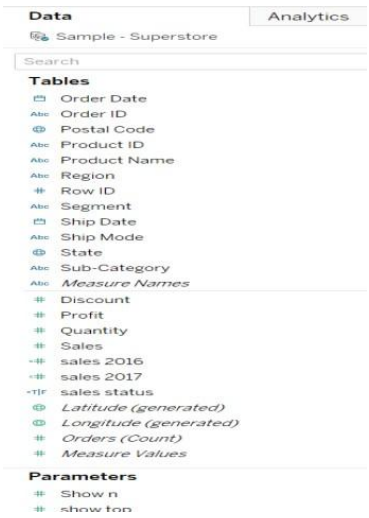

Fig. 6. Tableau UI

screenshot from a tableau which consists of several data columns. Those data are categorized as:

1) ***Dimension:****:* In this type, data consist of a category i.e., categorical type of data that gives information about the category that the user is engaged with. These kinds of values could provide a cluster or group that gives an information about the group in which the data falls under. Here, two different data that holds different property such as date which holds date type of data which is one of the best features in tableau from which different timelines about the acquisition of data can be done and significantly helps in time series analysis in a different type of tools and geographical data can be seen that gives and information of particular state.

2) ***Measures:****:* Hold types of data that can be calculated. For instance, columns likewise Sales, Quantity, Discount, etc. are measure values.

3) ***Bar Chart:****:* This is a widely used type of visualization technique which is responsible for summarizing a set of categorical data on the basis of either continuous or discrete values. In maximum functionalities, continuous data are converted to categorical values using auto-binning features that mostly exist in any kind of tools and techniques of data visualization. In a bar chart, we can perform different functionalities like splitting columns into another categorical column or data, or to split it into a same column but separated by colors. Here, the implementation of the bar chart is given below: In the above depicted graph, the x-axis is taken from

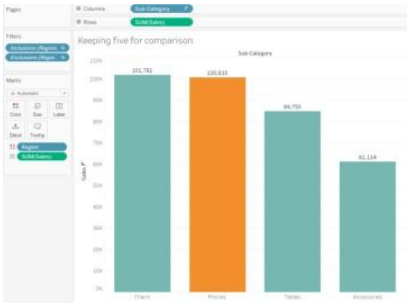

Fig. 7. Implemented bar chart

category type i.e., Sub-Category, and sales are taken from measures value. Further enhancement in this technique is done where different color is shown sub-category wise and text is taken from the measure which shows the data in a text format which clearly helps to visualize the number of sales made. From it, we can analyze the sales status based on a category of product that the Superstore is selling [12].

4) ***Pie chart:****:* It's a circular representation of data which is also called a pie shape. It is one of the graphs that are most frequently used to display data because it uses the characteristics of spheres, circles, and angular data to reflect actual information. A pie chart has a circular shape, with the pie representing the entire set of data and the slices representing its many components, which are recorded separately. This picture is mostly used depict percentiles as it gives information about

eISSN 1303-5150

as per the area of proportions covered. Given below shows an implementation of a pie chart. [8]

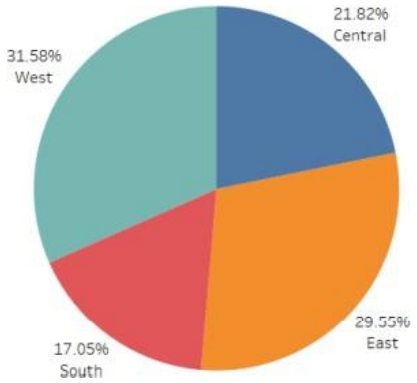

Fig. 8. Implemented pie chart

Here, the given chart indicates that the western region has maximum sales values which are in a percentile of 31.58% and the region south has minimum percentiles of sales i.e., 17.05%. From this descriptive analysis, a business analyst can plan for setting his business accordingly making a proper plan as per the demand of their products in different regions.

5) ***Line chart:****:* It is a type of graphic data representation in which points are continuously connected by lines. Depending on the data under investigation, the line might be straight or might be curved. The easiest way to depict quantitative data between two variables is using a line, which can be straight or curved. Line charts use lines to display information. Since it offers a clear depiction of the data, line charts are used by traders, investors, and financial officers to show the high and low within market for a specified value. Talking about the line charts on the tableau, it is used on the dataset which consists of date type. Implementation of a line chart is given below: In

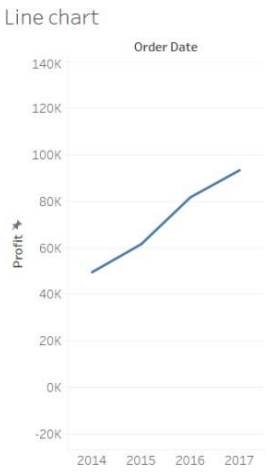

Fig. 9. Implemented line chart

the picture above, a simple line chart is implemented which consists of the date in the x axis and the profit on the y-axis. This chart helps to explain the trend of profit year-wise. It indicates that the company is earning profit in an increasing amount day by day as the line is inclined almost in a straight form.

6) ***Histograms:****:* A histogram is a visual representation of data points grouped into user-specified intervals. By gathering several data points and organising them into useful ranges or bins, the histogram, which resembles a bar graph in appearance, condenses a trend line into an understandable picture. It consists of two axes where the y-axis generally depicts the percentage or count of occurrences in data for each column. Columns are used in the visualization of patterns of data distribution. [4]

7) ***Tree Maps:****:* For the display of hierarchical data, there are three maps. They are constructed by a collection of nested rectangles whose sizes are proportionate to the associated data value. A wide rectangle is used to represent a branch of a data tree, and smaller rectangles are used to represent the size of the individual nodes within that branch. [4] On data panels, tree maps are frequently seen. To give visual variation to a busy interface, designers frequently select them. Tree maps, however, provide a number of challenges to quick comprehension as they are complicated image. This is the fundamental prerequisite for any data shown on a dashboard. The picture depicted above is an implementation of treemaps

Fig. 10. Implemented tree map

using a tableau which can be implemented by using one dimension and one measures value in a tableau. Here an advanced form of a tree map is drawn where size shows the figure of sales state-wise which indicates that California is the state with the highest sales value followed by New York with figure 457,688 whereas North Dakota has minimum sales with a figure of 920. Moreover, the colors that are shown above represent amount of profit state-wise. The red color signifies states that are bearing losses whereas the green box shows earning a profit. Here, the transparency of colors changes as per the amount of profit or loss that happened in states.

8) ***Bullet charts:****:* In a bullet graph, a bar is tagged with additional encodings to represent progress toward a goal or performance in comparison to a reference line. With extra visual elements added to each bar to add more detail, each bar concentrates the user's attention on a single measurement.

eISSN 1303-5150

Meters and gauges, which predominated early dashboards and reports, are replaced by the Stephen Few-designed bullet graph. It offers more information in a more condensed area, making a small dashboard the perfect fit. A bar that represents the featured metric appears in a bullet graph. With its vibrant color and prominent line, compared to the rest of the graph, this bar sticks out. It stands at the center of the graph. An axis with a quantitative scale has a reference line that is perpendicular to the bar and symbolizes a goal or other significant threshold. Whenever the main bar passes the reference line, the target has been attained or the circumstance has changed [15]. The chart shown below depicts the implementation of the bullet chart on the tableau Here, a given bullet chart is used

Fig. 11. Implemented bullet chart

to compare the status of sales between the years 2016 and 2017. The bar represents the year 2017 whereas a small black colored line represents sales made in the year 2016. Color is used to represent profit status in the year 2017. In this chart, the maximum sales seems to be increased in 2017 in each and every subcategory except a few of them which seemed to be decreasing. It helps analysts and sales representatives to analyze the reason for decrease of sales in particular categories and further prescriptive measures that need to be taken to come over with followed trends by data.

9) ***Tables:*** Using a table form view, a tabular visualization enables you to present data from such a metric collection. If you didn't select the alternative type, the visualization type that is first utilized when choosing data is a table, ordinarily referred to as a list box or a data table. The visual representation of the table is given below: The table depicted above represents a table with measure names at the top and the reason as an index of a column followed by a category. And, it also has two different measures depicted i.e., profit and discount. All in all, from the above we can easily analyze the discount given and profit happened in several regions. In above, the central region has faced a loss of $ 2871 in an item i.e., Furniture but in all other categories, everything is in profit.

Region	Category	Disco..	Profit
Central	Furniture	143	-2,871
	Office Supplies	359	8,880
	Technology	56	33,697
East	Furniture	93	3,046
	Office Supplies	245	41,015
	Technology	77	47,462
South	Furniture	40	6,771
	Office Supplies	167	19,986
	Technology	32	19,992
West	Furniture	93	11,505
	Office Supplies	177	52,610
	Technology	80	44,304

Fig. 12. Implemented table

10) ***Story in tableau:*** A story in Tableau is a series of visualizations used in concert to convey information. You can use stories to make a strong argument, give context, illustrate how decisions affect results, or just tell a data narrative. The portrait given below depicts a portrait of a story in tableau. Because a story is a sheet, you can create, identify, and manage stories in the same ways that you do with worksheets and dashboards. A story is also a group of sheets that have been organized in a certain order. Story points are the individual sheets that make up a story. In a portrait given below two visual representations are used to make which operates sequentially so that it will aid users in easily understand and visualizing data. [1]

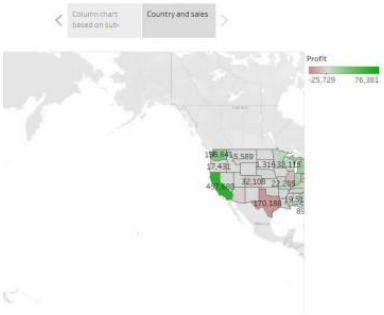

Fig. 13. Story tableau

11) ***Dashboard in tableau:*** It is a collection of various views that allows you to compare numerous pieces of data at once. Instead of navigating to different worksheets, you can construct a dashboard that shows all the images at once, for instance, if you have a group of pictures that you evaluate every day. Dashboards can be accessed from tabs there at bottom of a workbook, just like worksheets. A sheet's data is related to any dashboards that contain it, so when you edit a sheet, those dashboards also change, and the reverse. Both the sheets and the dashboards are updated with the most recent

eISSN 1303-5150

information from data sources. [16]. The image shown below represents an interactive dashboard made using four different sheets.

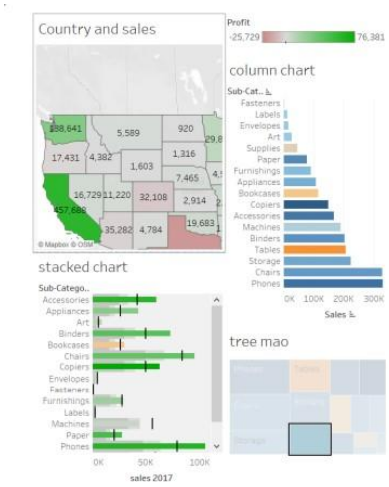

Fig. 14. Dashboard in tableau

III. APPLICATIONS:

The majority of visualization designs are intended to improve cognition and support decisionmaking. How well a data visualization will be used must be taken into consideration when developing and creating a prototype. Data visualization requires picking and rethinking the figures that serve as the basis for the portrayal, going beyond just portraying the numbers [17]. Data visualization is a crucial area of computer science that has a variety of application areas. To examine individual datasets in many applications, a number of application-specific methods have been created for several medical and scientific domains.

A. Public Health:

For public health monitoring to be successful, data analysis and presentation must be clear and understandable. To support their job, health researchers require practical and clever tools. Cloud-based medical data visualizations must prioritize security [7]. Today, you can find various graphical representations if you open any medical or health publication.

B. Renewal Power:

Finding the best approach requires calculating energy usage in relation to production. [9]

C. Environmental Science:

Environmental management needs visualization as they are obliged to make decisions based on incredibly complex facts. Applications for visualisation are now being used in applied environmental research. [5] It is preferable to have access to various programs for displaying findings.

D. Fraud Detection:

During the preliminary stages of a fraud investigation, data visualisation is essential. Fraud investigators might employ data visualisation as a preventative detection technique to identify trends that indicate fraudulent behaviour. [3]

E. Library-Decision Making:

In conjunction with data visualisation software, librarians can more easily assemble and demonstrate data compiled from several sources. It offers students the means to communicate ideas in an engaging manner [11]. The display of library data emphasizes purchasing choices, upcoming library requirements, and objectives. Due to the fact that they are de facto professionals in the domain, librarians can assist students, teachers, and researchers in visualizing their data. [2]

IV. CHALLENGES

Animations is utilized for interactive study of time-varying data. By imitating the storytelling approach, it depicts temporal events. Users' capacities for using data visualization and coming to judgments under time pressure vary [18]. Because of the vast amount of data they contain, large, time-varying data sets provide a significant problem for data visualization. Users may be able to react proactively to challenges that develop thanks to real-time data visualization. A method for creating The benefits of a data visualization technique are difficult to measure. The existence of numerous visualization algorithms and related software is due to this. The majority of these programs do not make use of the direct manipulation and multi-touch engagement features of the new devices. Big data poses a unique set of challenges for the development of visualizations, both for structured and unstructured data. This is because we have to consider the speed, volume, and variety of the data [10]. A new set of issues regarding effectiveness, operability, and amount of differentiation provide difficulties for large data visualization and analysis. Setting a sizable replicated data set takes time and effort. Additionally, selecting the ideal picture for use can be challenging. Talking about the challenges that are faced during the implementation of different techniques in tableau, it becomes complicated to interpret which techniques have to be used based on varied conditions. Several data sets cannot be represented by using these normal charts. Hence, we have to dive into other advanced data visualizations that will overall justify the representation of data.

V. CONCLUSIONS

Data visualization is the practice of effectively and clearly displaying data in a graphical or pictorial form. For the analysis and interpretation of big and complicated data, it has become a potent and extensively used tool. It is now a quick and simple method of disseminating ideas in a format that is accessible to all. It must effectively, precisely, and transparently convey complicated ideas. These advantages have made it possible for data visualization to be helpful in sever domains. Tableau is the much-advanced AI-paced technique that fully justifies data of any kind and number of dimensionalities. It

eISSN 1303-5150

is equipped with thousands of features and data visualization technique that provides a facility for digging into a core part of any kind of data set and give exploration of core insights

REFERENCES

[1] Steven Batt, Tara Grealis, Oskar Harmon, and Paul Tomolonis. Learning tableau: A data visualization tool. *The Journal of Economic Education*, 51(3-4):317–328, 2020.

[2] Tara J Brigham. Feast for the eyes: an introduction to data visualization. *Medical reference services quarterly*, 35(2):215–223, 2016.

[3]William N Dilla and Robyn L Raschke. Data visualization for fraud detection: Practice implications and a call for future research. *International Journal of Accounting Information Systems*, 16:1–22, 2015.

[4] Clarence Goh. Data dashboarding in accounting using tableau. *Available at SSRN 4141810*, 2022.

[5] Sam Grainger, Feng Mao, and Wouter Buytaert. Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software*, 85:299–318, 2016.

[6] Jamie Hoelscher and Amanda Mortimer. Using tableau to visualize data and drive decision-making. *Journal of Accounting Education*, 44:49–59, 2018.

[7] Tumaini Kilimba, Gideon Nimako, and Kobus Herbst. Data everywhere: an integrated longitudinal data visualization platform for health and demographic surveillance sites. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 551–552, 2015.

[8] Robert Kosara. The impact of distribution and chart type on part-to-whole comparisons. In *EuroVis (Short Papers)*, pages 7–11, 2019.

[9]Omesh Kumar and Abhishek Goyal. Visualization: a novel approach for big data analytics. In *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*, pages 121–124. IEEE, 2016.

[10] Xinxiao Li, Akira Kuroda, Hidenori Matsuzaki, and Nobuyasu Nakajima. Advanced aggregate computation for large data visualization. In *2015 IEEE 5th Symposium on Large Data Analysis and Visualization (LDAV)*, pages 137–138. IEEE, 2015.

[11] Sarah Anne Murphy. Data visualization and rapid analytics: Applying tableau desktop to support library decision-making. *Journal of Web Librarianship*, 7(4):465–476, 2013.

[12] Daniel G Murray. *Tableau your data!: fast and easy visual analysis with tableau software*. John Wiley & Sons, 2013.

[13] Frits H Post, Gregory M Nielson, and Georges-Pierre Bonneau. Data visualization: the state of the art (vol. 1). *Massachusetts: Kluwer Academic Publishers*, 2003.

[14] Matthew Sadiku, Adebowale E Shadare, Sarhan M Musa, Cajetan M Akujuobi, and Roy Perry. Data visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12):11–16, 2016.

[15] Donabel Santos. *Tableau 10 Business Intelligence Cookbook*. Packt Publishing Ltd, 2016.

[16] Eko Aji Saputra and Dewi Agushinta. The implementation of business intelligence on smart sales dashboard using tableau (study case: Pt. derma konsep estetika).

[17] Joanna Wolfe. Teaching students to focus on the data in data visualization. *Journal of Business and Technical Communication*, 29(3):344–359, 2015.

[18] Li Yu, Aidong Lu, William Ribarsky, and Wei Chen. Automatic animation for time-varying data visualization. In *Computer graphics forum*, volume 29, pages 2271–2280. Wiley Online Library, 2010.