An Overview of Key Principles of Effective Data Visualization

Samyukta Rongala

Abstract: Data presentation presents information in standard graphical works like charts, graphs, and maps. Thus, visualization enables one to look at large amounts of data and identify something that is understandable and can be quickly processed. This paper explores the key rules of layout and briefly reflects on the most critical aspects of gateway communication: less is more, be consistent and be Contextually Aware. It raises an important question of how color, shape and the type of charts used affect the visualization. This paper also asks and answers questions related to the audience, the roles of design in reducing the cognitive load needed to identify patterns, and the values of interactivity in encouraging enhanced focus. This work gives a good starting point for some things to prioritize before designing data visualizations that enable users to gain insights from data.

1. Introduction

Data visualization is an essential activity that turns data collected in various forms into graphical figures. What a user can decipher when they look at a graphical image is a cognitive system that interprets and lets the user see helpful information and insights [25]**.** Whether in business, teaching, or research, data presentation is critical to convey simple data in a meaningful story. Visual information is recognized 60 thousand more quickly by the human brain than text; therefore, naturalization is among the most efficient instruments for sharing information and insight.

John Medina of Brain Rules Work **[10],** people get a 90% comprehension rate in graphics illustrations as compared to written word messages. It is shown that when information is presented in the text with some images, it is retained more often. Data visualization research reveals that decision-making can occur five times quicker and be far more precise than business intelligence based on raw data or textual reports. **[24]** has conducted research that shows that data visualization improves decisions by revealing patterns, trends, and outliers at a breakneck pace. If users are encouraged to 'dig in' to the data themselves, it can be up to 40% more effective in engaging its audience.

Given the massive amount of data characteristic of the digital era of big data, data visualization has proven to be the most crucial element of information management. For example, the volume of unstructured data ranges from 60% to 90% overall, and visualization brings insights into the ambiguity of big complex data **[3].** Another advantage of the DV is its ability to identify trends and odd ones in big datasets. The determining factor is that good visualizations can turn abstract data into something tangible that people can use to make decisions and alter their daily lives. Figure 1 below shows a Data Visualization Process, from raw data to meaningful insights.
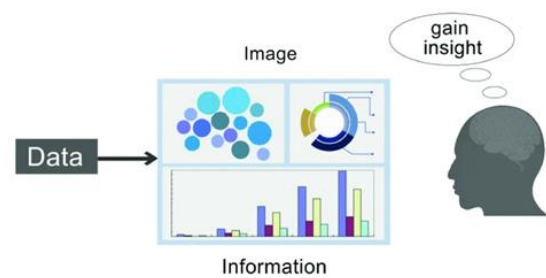

Fig 1: Data Visualization Process. [25].

2. Principals of Data Visualization

There are some rules you need to follow if you desire to have better visualizations. Regarding data dashboards, charts, graphs, or maps, work productivity increases when options of standard guidelines are applied. Data presentations include details and present information in a proper, meaningful and convincing format. It is in this paper that the basic concepts that shape successful data visualization will be addressed **[11].**

2.1. Know your Audience / Purpose of the Visualization

Knowing your audience is critical to data visualization because it enables you to convey a clear message. As with almost all internet marketing strategies, visualization is not a globally applicable remedy, and the decision of whether or not it will bring beneficial results depends on understanding who is being bombarded with the ads. The amplitude of familiarity with both the data and the specific topic can also differ significantly depending on the group of recipients and whether we are talking about entrepreneurs, analysts, or ordinary information users. Hence, each of these groups needs different types of visualization.

Beginner Audience: It is asserted that simple and uncluttered visualizations with little frills and straightforward language are suitable for those with little to no background in data representation.

Intermediate Audience: Graphical utilities, including dynamic views with selection and subset controls, are helpful for users possessing relatively good knowledge about the data [3].

Expert Audience: Despite their relatively small size and highly complicated structure, complex visualizations with descriptions and intricate statistical features are useful for analysts and data scientists to examine further [11].

Tailored Approach: The complexity of the visual portrayal should correspond to the audience's knowledge while being informative and engaging—basic graphics for non-advanced users and complicated graphics for highly advanced users [11].

2.1.1 Purpose of the Visualization

The role of the purpose of visualization is predominant in defining how the data should be presented. The purpose of the visualization can alter the kind of chart more profoundly, the complexity of the information displayed, and the interactivity ratio. The goal of the visualization has to be clearly stated to understand whether the task at hand corresponds to the expectations of the intended audience. There are several purposes for visualization, decision-making, education or broader audiences and context [5].

2.1.2 Customizing for Your Audience

In practice, it becomes essential to pay attention to the specifics of designing in line with the experience of consumers and the peculiarities of a given group. Personalizing the visualization is to make the audience interested in what is actually useful and comprehensible.

There is a need to understand the audience. Paying attention to the type of audience, its purpose, and the need for detailed information, you are free to create the best kinds of visualizations. If one makes it look complicated for others who may not understand it, it scares the experts, and on the other side, if one makes it too simple for the experts, it feels like the experts are being humored [11]. It is a common pitfall when designing visualizations; the people creating them expect the viewers to understand the data the same way as they do, which is a curse of knowledge. It can be very challenging to describe what a beginning audience will consider relevant, sometimes presenting difficulties for mentors and other pundits [17]. To address this, the use of visualizations must consider the end view of the audience to help convey the data correctly, regardless of the audience's sophistication.

2.2 Clarity & Simplicity

For data to be visualized accurately, several things should be avoided to help ensure that the visual data helps to find the messages accessible to users. These features include grid lines, three-dimensional views, and confusing texts that may hide the intended message behind the visual representation of data (Cairo, 2016). Every visualization that is made should aim to convey a single idea or point of comparison in essence at most; otherwise, the audience will be overburdened. In line with recommendations in the literature on design research and the best practices in visualizations, titles of such visualizations are probably to be noticed by views. Referring to the title also serves as a function of putting into perspective the audience's comprehension of figures. For this reason, the message must come first, with the text as an additional element to the visuals. More text is needed to include the goal of presenting the data clearly and unencumbered [17].

Several strategies can be applied to create clear and simple data visualizations: heavy lines on significant features, horizontal lines, low-order line types for different trends, and headings for substantial items and variables. Figure 2 below shows an example of Clear vs. Cluttered Visualizations used to show clarity and simplicity.

Fig 2: Examples of Clear vs. Cluttered Visualizations: To show clarity and simplicity. [1]
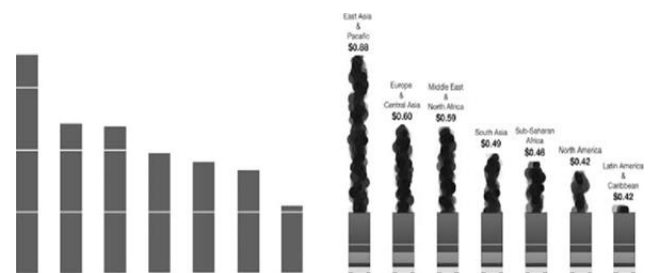

Fig 3: Examples of minimalist and embellished charts. [8].

The ideology behind minimalism is the absence of all nonessential elements in the data representation, called chartjunk. Famous designers like Edward Tufte believe that having a strong data-ink ratio is crucial, meaning the information should be as straightforward as possible. Some of these aspects include removing too many grid lines, ornamented fonts, and even three-dimensional appearances since they only clutter up the message being passed by the visualization.

In other words, the goal of minimalism in data visualization is to get the ultimate measure of the 'signal-to-noise ratio', the former being the essential message that is being conveyed in the form of data and the latter being characterized by all the other parts of the data set that are not informed by the objective of the visualization (Islam & Jin, 2019). To this end, visualizations must be clean and free of anything that does not help visualize the data. It also involves removing the unwanted gridlines, a tendency towards ornate typography faces and three-dimensional figures that skew the data and thus the perception [3]. Figure 3 below shows an example of minimalist and embellished charts, which lacks clarity.

2.3 Chart Type

Identifying the chart type to use dramatically determines the comprehension of information being passed from the data. Thus, the kind of chart you settle on depends on the data type you are working with, the specific message you want to relay to your viewers, and, most importantly, their expectations [12]. Below are some of the most common chart types, along with guidelines on when to use them:

2.3.1 Bar Chart

Bar charts are another preferred chart type utilized to illustrate categorical data or shifts in discrete data. They can show differences in the size, frequency, or extent of certain characteristics in two groups or more.
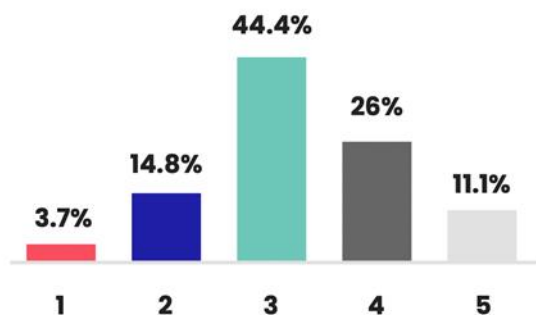

Fig 4: Example of a Bar Chart. [4].

Figure 4 above shows an example of a Bar Chart. A bar chart is suitable when presenting data relating to two or more categories in the set. For instance, they are appropriately used to analyze sales performance within specific geographical areas or the number of products sold in various classes. The length of each bar is proportional to the data value, so it is convenient to compare the sizes of the values based on their categories. Such graphs are most suitable when presenting discrete data that fall under categories that do not contain the other data type as a subset or an element. This can be illustrated by presenting the number of students in specific grade levels or customers' preferences regarding particular products.

Therefore, bar charts can be used better when comparing categories and illustrating part-to-whole relationships, ranking data, and change over time [13]. They are a fundamental and effective instrument for displaying categoric information and making direct contrasts.

2.3.2 Line Chart

It is primarily suitable for representing time series data or continuous data sets, as seen in the example below. According to [4], it clearly demonstrates how data transforms and recognizes trends, correlations, and variations.
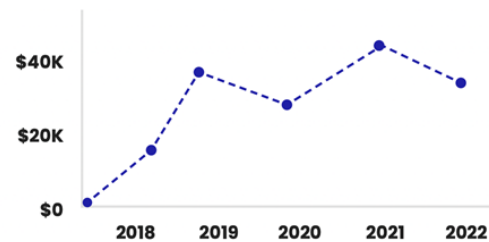

Fig 5: Example of a Line Chart. [4]

Figure 5 above shows an example of a Line Chart. A line chart works best when presenting the change in data in equal intervals and is best suited when presenting features. For instance, they can represent stock prices for the past, monthly sales revenue or website traffic. The linked data points help identify more data points moving upward or downward. Line charts are also ideal for quantitative data on the continuous line of the axis, that is, data which can be graduated along line scale such as temperature, speed, and distance [19]. For example, a line can indicate the transition of plants' growth from one season to another or in temperature from morning to evening.

2.3.3 Scatter Plot

Scatter plots are applied if there is one dependent and one independent continuous variable whose importance is ranking [16]. They can uncover relationships, tendencies, and anomalies, which makes them good for the EDA.
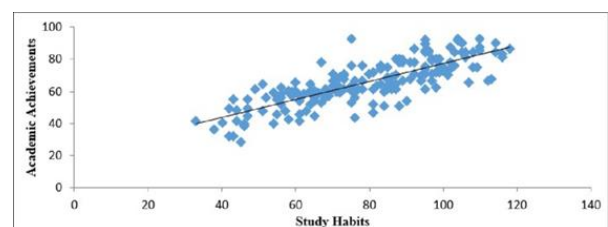

Fig 6: Example of a Scatter Plot. [16].

In particular, scatter plots are used to examine some properties of the connection between two quantitative parameters. For example, they can be used to find out the correlation between the amount of time spent studying and exam performance or the amount spent on advertising and sales results. They are points that depict observations made

on two variables placed on two dimensions of a two-dimensional axis to discover associations and trends.

In particular, a scatter plot helps determine dataset anomalies–observations beyond any conceivable expectation [14]. For example, using the scatter plot, one can learn the extent of the high and low sales or income, which is slightly far from the rest of the other values. Hence, when the scatter plot is analyzed, we get to understand the pairwise relationships of the two quantitative variables, such as the kind of relationship, the degree of the relationship, and the pattern of the relationship, as well as the opportunity to identify outliers. This kind of model is especially valuable in pattern recognition and data mining.

2.3.4 Histogram

Histograms are helpful, mainly for prevailing continuous data. They show the proportion of observations included in given sectors referred to as "bins" to uncover the density of data. Figure 7 below shows an example of a Histogram.
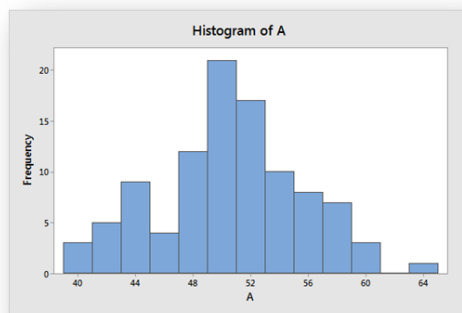

Fig 7: Example of a Histogram. [7].

Figure 6 above shows an example of a Scatter Plot of study habits against Academic achievements. In particular, scatter plots are used to examine some properties of the connection between two quantitative parameters. For example, you can find out how many people of what age are in a country or how many students in a class scored what grade. Histograms also aid in the identification of whether data is symmetrical or skewed, has multiple peaks or has outliers [22]. A histogram can present the distribution of data points within a specific interval. For example, they can demonstrate areas where customers spend more money or how often complaints appear during the specified time intervals.

Hence, histograms are most valuable when analyzing continuous distribution, frequency and spread data. They are uniquely effective at revealing the arrangement of the data as well as skewness and outliers of density

2.4 Color

Hue and saturation are relevant in data communication since they buffer the visual interface, highlight relevant information, and point viewers to relevant information.

When applied correctly, the use of color can make otherwise dense data much easier to understand. It makes it easier to pass meaning in a shorter time since it is less demanding on the audience's cognition and makes it easier for them to interpret figures of pattern and trend. Figure 8 below shows a Hue Density plot Distribution as an example to demonstrate how color is used in data visualization
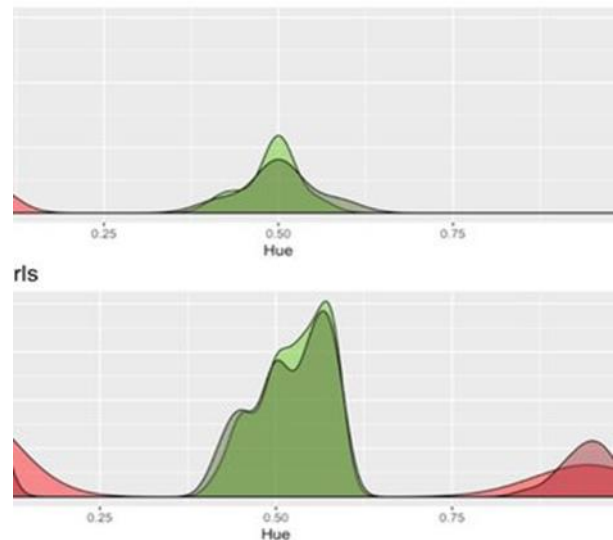

Fig 8: Hue Density plot Distribution. [21].

Color contrasts can be used in categorization, differentiation of categories, and clarification of the data or the lack of them to enhance the interpretability of the created visualizations [15]. If adequately applied, color makes the viewing experience more stimulating, and the information received more understandable. Designers must always grasp the psychological and contextual effects of color for the visualization to be informative of the encoded message in the data. Color should be used to enhance and emphasize the data further instead of acting as a visual nuisance. Contrasting and sharp colors should be applied moderately and more frequently when specific trends or values may be focused on. Intensive use of vibrant colors and high contrast may be inhibitive and overload the viewer's perception, thus resulting in an incorrect interpretation or total disregard for the vital information.

Concisely, color is an effective means of data visualization that can help further understand the intricacies presented in numbers and emphasize key phenomena in the given data set. When choosing colors, keeping the concepts coherent and using color to help highlight the data can help create good and easy-to-understand visualizations. Further, one needs to think about colorblind people better to understand how they see the designed data. This means that tools used to make choices concerning color enable the achievement of color accessibility so that visualizations are not distorted in a way that could confuse or obscure their meaning to people with color deficits.

2.5 Consistency/Story Telling

2.5.1 Consistency

Repetition and narrative are the key components when creating successful data visualizations. Both ensure clarity and increase the audience's interest so that understanding a particular matter is easy. These elements are crucial for Visual Clarity, Cognitive Ease, Professionalism and Trust, and Branding and Identity.

Copying standard equipment, including colors, fonts, and the type of charts used, is crucial because it facilitates the audience's association of patterns and correlation in the data. When the elements mentioned above are inconstant or move around erratically, they inconvenience the viewer and complicate their avenue of concentrating on the significant highlights or insights given [9]. However, for consistency, the viewer begins to understand the type of presentation being used, gaining faster understanding. Figure 9 below shows a dashboard for Visualizing Performance ensuring clarity and consistency.
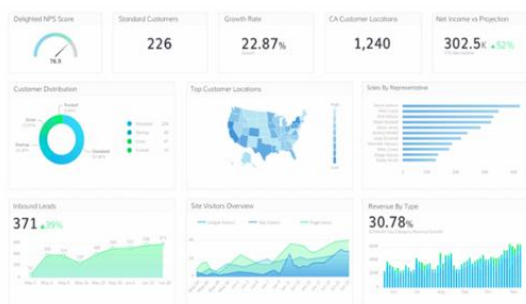

Fig 9: Dashboards for Visualizing Performance. [(2)]

Creating a set design that is employed throughout a range of visualizations creates a professional look. Coherent or related visual material will help avoid confusion regarding what is being presented and will not compromise the audience's confidence in that data. Conversely, follow-through, across charts and visuals, indicates to the viewer that the data has been presented with some level of planning, thereby developing the viewers' trust in the information sharing. The coherence in every aspect of the overall appearance is a significant aspect vital to maintaining the organization's brand image in a business setting. The consistency of the colors, fonts, and logos helps keep the visualization of the company's visual identity guidelines. This improves the aesthetic value and aligns all communication with the firm's articulated brand.

2.5.2 Storytelling

Results given without a framework can seem quite dull and naked at first observation. Storytelling enables one to create a context or a story around it, which really makes key points pop out of numerous figures and rates. This way, the information structure will be meaningful and present a more precise set of insights the viewer can follow—making sense and being relevant.

Human beings are inclined to stories, which is why there are so many narrative-based visualizations there. A storytelling type of visualization is much better as the viewer gets bombarded with many numbers, and it is easy to lose interest and forget what one has seen in the visualization (2023). It highlighted that several examples exist to show the conversion of static graphics into interesting and engaging interactions using the right story. One thing people underestimate a lot is the ability to put complexity into a single and focused narrative. When working with the data, highlighting only the necessary aspects is more straightforward when you already have a story in mind. This makes it easy for the information offered in the visualization to be understood; thus, the message is not hidden in excessive and nonessential details.

The principle of structural consistency and story, Yasemin Tanell and Meggin A. Kearney, tells how it is possible to construct the data's framework and narrative accurately and comprehensively. Consistency makes the visualization less confusing; it makes it look more credible and less confusing to the reader; storytelling makes the data more engaging, relevant, and thus easier to comprehend. When all these components are integrated, they not only facilitate the usability of the data but also promote action and provoke decisions. A well-thought-out, coherent and integrated story makes the info-graphics strong instruments of analysis and persuasion.

3. Conclusion

In Conclusion, data visualization is a powerful means of connecting data and insights into decision-making, learning, or communication outcomes. We will focus on using data visualization tools to present the results because their strength is converting complicated data sets into easy-to-understand infotainment and meaningful graphics. Keeping audience, understanding, simplicity, and constancy in mind makes data visualization meaningful and powerful. Understanding the user audience is central to deploying suitable visualizations relative to their background. Beginners appreciate simple setups; intermediate users need engagement, and power users are satisfied with detail. Simplicity and clean design make a call for the essentials and eliminate contortions by unimportant details. Analyzing the choice of charts is vital because each type is used for a specific intention, starting with trends in line charts to distributions in histograms. The combination of colours increases understanding of the text and, at the same time, creates an easily distinguishable contrast for viewers with colour vision disorders. Repetition in design induces reliability and professionalism since it coins the feel of a standard brand image, and storytelling helps viewers comprehend and deliberate on numbers. Altogether, these

principles are tools that 'sell' the message tucked behind visualizations, convincing an audience to act desirably and constructively. For this reason, making data images is seen as a helpful tool which, when done purposefully, is an effective way of sharing information in societies with a massive demand for data.

Author contributions

Samyukta Rongala: Conceptualization, Methodology, Field study Writing-Original draft preparation, Software, Visualization, Writing-Reviewing and Editing.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] S. A. Scientist Data, "Essential Principles for Effective Data Visualization," *The Deep Hub*, Mar. 28, 2024. https://medium.com/thedeephub/essential-principles-for-effective-data-visualization-a13f05d22c39

[2] AMA, "Marketing Dashboards Market to See Huge Growth Google, IBM, Microsoft, Oracle," *openPR.com*, Jan.28,2023. https://www.openpr.com/news/2902852/marketing-dashboards-market-to-see-huge-growth-google-ibm

[3] K. Brush and E. Burns, "What is data visualization and why is it important?," *SearchBusinessAnalytics*,2020. https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization

[4] S. Donnelly, "16 of the best financial charts and graphs," *Finance Alliance*, Oct. 19, 2023. https://www.financealliance.io/financial-charts-and-graphs/

[5] S. Donnelly, "16 of the best financial charts and graphs," *Finance Alliance*, Oct. 19, 2023. https://www.financealliance.io/financial-charts-and-graphs/

[6] K. Eberhard, "The Effects of Visualization on Judgment and decision-making: a Systematic Literature Review," *Management Review Quarterly*, vol. 73, Aug. 2021, Available: https://link.springer.com/article/10.1007/s11301-021-00235-8

[7] J. Frost, "Using Histograms to Understand Your Data - Statistics By Jim," *Statistics By Jim*, Feb. 11, 2019. https://statisticsbyjim.com/basics/histograms/

[8] P. Gough, X. Ho, K. M. Dunn, and T. Bednarz, "Art and Chartjunk," Aug. 2014, doi: https://doi.org/10.1145/2636240.2636852.

[9] M. Islam and S. Jin, "An Overview of Data Visualization," *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, pp. 1–7, Nov. 2019, doi: https://doi.org/10.1109/icisct47635.2019.9012031.

[10] R. Mattison, "Data Storytelling: How to Tell a Great Story with Data," *ThoughtSpot*, Mar. 14, 2023. https://www.thoughtspot.com/data-trends/best-practices/data-storytelling

[11] S. R. Midway, "Principles of Effective Data Visualization," *Patterns*, vol. 1, no. 9, p. 100141, Dec. 2020, doi: https://doi.org/10.1016/j.patter.2020.100141.

[12] Mokkup.ai, "Data Visualization — How to Pick the Right Chart Type?," *Medium*, Sep. 20, 2024. https://medium.com/@mokkup/data-visualization-how-to-pick-the-right-chart-type-f6f65824928a

[13] "LibGuides: Data Visualization: Comparison," *Morgan.edu*, 2024. https://library.morgan.edu/dataviz/cf

[14] Q. V. Nguyen, N. Miller, D. Arness, W. Huang, M. L. Huang, and S. Simoff, "Evaluation on interactive visualization data with scatterplots," *Visual Informatics*, vol. 4, no. 4, pp. 1–10, Dec. 2020, doi: https://doi.org/10.1016/j.visinf.2020.09.004.

[15] N. Desktop, "The Importance of Color Theory in Data Visualization with Tableau," *Nobledesktop.com*, Sep. 25, 2024. https://www.nobledesktop.com/learn/tableau/the-importance-of-color-theory-in-data-visualization-with-tableau

[16] J. H. Park *et al.*, "The principles of presenting statistical results: Figures," *Korean Journal of Anesthesiology*, vol. 75, no. 2, Jan. 2022, doi: https://doi.org/10.4097/kja.21508.

[17] X. Qin, Y. Luo, N. Tang, and G. Li, "Making data visualization more efficient and effective: A survey," *The VLDB Journal*, vol. 29, no. 1, pp. 93–117, Nov. 2020, Available: https://link.springer.com/article/10.1007/s00778-019-00588-3

[18] J. Ryan and H. Herleman, "A Big Data Platform for Workforce Analytics," *Routledge eBooks*, pp. 33–56, Nov. 2015, doi: https://doi.org/10.4324/9781315780504-8.

[19] S. C. Government of Canada, "5.5 Line chart," *www150.statcan.gc.ca*, Sep. 02, 2021. https://www150.statcan.gc.ca/n1/edu/power-pouvoir/ch9/line-lineaire/5214824-eng.htm

[20] C. A. Steed, "Interactive data visualization," *Elsevier eBooks*, pp. 185–211, Nov. 2024, doi: https://doi.org/10.1016/b978-0-443-13878-2.00002-3.

[21] P. Stenger, Jérémie Vidal-Dupiol, C. Reisser, S. Planes, and C. Ky, "Colour plasticity in the shells and pearls of animal graft model Pinctada margaritifera assessed by HSV colour quantification," *Scientific Reports*, vol. 9, no. 1, May 2019, doi: https://doi.org/10.1038/s41598-019-43777-4.

[22] S. Turney, "Skewness | Definition, Examples & Formula," *Scribbr*, May 10, 2022. https://www.scribbr.com/statistics/skewness/

[23] Y. Wang, H. Huang, C. Rudin, and Y. Shaposhnik, "Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization," *Journal of Machine Learning Research*, vol. 22, no. 201, pp. 1–73, 2021, Available: https://www.jmlr.org/papers/v22/20-1061.html

[24] Q. Zhang, "The Impact of Interactive Data Visualization on Decision-Making in Business Intelligence," *Advances in economics, management and political sciences*, vol. 87, no. 1, pp. 166–171, Jun. 2024, doi: https://doi.org/10.54254/2754-1169/87/20241056.

[25] Q. Li, *Embodying Data*. Springer Nature, 2020.