Statistics Corner: Data Visualization–I

Kamal Kishore¹, Vidushi Jaswal²

REALITY CHECK

Above all else show the data – Edward Tufte

The presentation of data with the help of various charts and graphs has become an almost indispensable component of scientific research. However, unlike many other subjects, data visualization is rarely taught in undergraduate and postgraduate courses. An investigator wants to communicate complex ideas of the study findings with the help of graphs. A significant challenge before the researcher is to produce quality graphs for easy and effective communication. Therefore, the researcher intends to know the answers to a few questions before finalizing the graphical display.

- What is graphical excellence?
- How to achieve graphical excellence?
- What is the role of the preattentive attributes in graphical excellence?
- Why Gestalt principles are important for attaining graphical excellence?

Keywords: Chart, Diagram, Graph, Illustration, Visualization.

Journal of Postgraduate Medicine, Education and Research (2020): 10.5005/jp-journals-10028-1412

INTRODUCTION

An eternal quest of mankind is to solve the puzzle of healthy living and disease-free life. Therefore, researchers across the globe work tirelessly to find the answers to perplexing problems faced by humanity. The investigators seeking answers carefully plan and document the studies to collect data in the hope of finding the best solutions. Despite clear evidence of results from the study, it is challenging to convince others regarding the same. The scientists bolster persuasion by building narratives around the facts. The text, tables, and charts are used to present facts. However, besides basic training in making tables and graphs, most of the researchers learn to present data by hit and trial method.

The data presentation is an art based on science. Thus, akin to good writing there are good data presentation techniques. Researchers basking in the glory of achieving convincing results perhaps forget many a time that presentation of facts is not about them, it is about others. Therefore, it is of paramount importance to carefully select data dissemination techniques. The graphs are more misunderstood as compared to tables. A significant reason for poor understanding of graphs may be either due to a lack of specialized data visualization courses during undergraduate and postgraduate training or because of its recent development or both. Therefore, it demands more care and planning to make effective graphs. The presentation of data in the form of charts became scientific and popular after William Playfair's writing in the 18th century and John Tukey's writing in the second half of the 20th century in contrast to the popular usage of tabular presentation from the 2nd century onward.

People tend to remember and recall more with the visual display.¹ Healthcare research has also shown that patients prefer and understand results better with a visual display. Therefore, it is important to understand the fundamentals of visual attention for effective communication. There is a significant rise in the popularity of charts due to either the wide availability of computers or data display software packages or both. The increasing usage

^{1,2}Department of Biostatistics, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Corresponding Author: Kamal Kishore, Department of Biostatistics, Postgraduate Institute of Medical Education and Research, Chandigarh, India, Phone: +91 9591349768, e-mail: kkishore.pgi@gmail.com

How to cite this article: Kishore K, Jaswal V. Statistics Corner: Data Visualization–I. *J Postgrad Med Edu Res* 2020;54(4):253–256.

Source of support: Nil

Conflict of interest: None

of computers and the convenience of various visualization tools further accelerated research and publication of more charts. As per estimates, 900 billion to 2 trillion graphical displays are produced annually.² However, a rapid rise in volume, velocity, and variety augmented with default software options do not generate an effective visual display for communication. Therefore, graphs need diligent preparation, more time, polishing, and visual cues for readers.

A chart is an effective tool that reveals novel and interesting relationships, trends over time, and patterns within a data. Edward Tufte defined data graphics as a visual display of measured quantities with the use of points, lines, symbols, words, shading, color, and a coordinate system.² Many people believe that the charts are secondary to tabular display. However, as demonstrated by Anscombe's quartet, visual inspection and presentation of data have inherent advantages as compared to tabular display. The graphs can be broadly segmented as exploratory and explanatory. The exploratory graphs help in the identification of various characteristics of data before analysis. Thus, exploratory graphs are primarily for the researcher to understand data and its peculiar characteristics, such as, data distributions, heteroscedasticity, outliers, and missing patterns. The explanatory graphs need to be lucid and compelling as they target readers for effective and objective communication. The explanatory graphs need to utilize

the preattentive attributes and Gestalt principles for actionable insights. Figure 1 displays the number of consultations provided by the department. Figure 1A is the default output whereas Figure 1B utilizes the preattentive attribute of color and Gestalt principle of similarity for effective communication. The default title, text, font size, gridlines, x-axis, and the gap between bars and data are also modified in Figure 1B for improving the communication.

The tabular and graphical display of data is linked with verbal and visual systems, respectively.³ The visual system has played a significant role by discerning patterns and relationships to counter threats for human survival. The human brain's perception speed varies with preattentive and attentive attributes. Attentive attributes need more focus and are slow in processing the

information as compared to preattentive attributes. Furthermore, Gestalt principles lay the foundation for understanding visual perception in human beings. Therefore, an appropriate selection of attributes and Gestalt principles can be leveraged to draw and direct readers' attention to important data in visualization. The researchers especially those involved in healthcare need to tap the potential of preattentive attributes and Gestalt principles to achieve graphical excellence for better and effective communication. In this article, we will try to understand the fundamentals of graphical excellence.

GRAPHICAL EXCELLENCE

Graphical excellence is an approach in which complex ideas are communicated to the readers with clarity, accuracy, and efficiency.² In other words, the communicator needs to convey the idea to the readers in a friendly, understandable, and time-saving manner. As the first step to graphical excellence, researchers need to have a maximum share of data-ink for displaying data. A classic example of the high data-ink ratio is an electroencephalogram (ECG) where removal of anything will compromise the information. The second vital step is to concentrate on revealing the actual data by strategically minimizing non-data components, such as, title, gridlines, texts, color, axis naming and scaling, shading, and symbols (revisit Fig. 1 for clarity). The non-data component is a significant source of chartjunk. The color plays an important role in revealing complex patterns and relationships in data. However, the bulk of software's produce default color graphs just to spice up figures rather than highlighting essential and important component of data. Figure 2 depicts the importance of color selection. Single different colors in Figure 2B receive immediate attention as compared to considerable efforts needed to identify multiple colors in Figure 2C. The color needs to be strategically and sparingly used to highlight the important characteristics of the graph.⁴ Further usage of colors needs to be consistent while also keeping in mind that approximately 10% of people are color blind.⁴

PREATTENTIVE ATTRIBUTES

It is well known that a picture is worth a thousand words and garner more attention. Despite the obvious advantages of visual displays, the majority fails to capitalize on its strengths. The usage of default software options, such as, color, saturation, symbols, gridlines, scales, and marks, rarely produces a quality graph. Effective communication desire more than the default options available in the software. The first step in preparation for compelling and lucid graphs is to understand and utilize preattentive attributes for effective communication. Figure 1B used the preattentive attribute of color for better communication. Effective visual presentations are

Figs 1A and B: Summary of frequencies of biostatistics consultations given by the Department of Biostatistics, PGIMER, Chandigarh, from 2018 to 2019: (A) Default output in Microsoft Excel®; (B) The final graph for actionable insight

Figs 2A to C: An example of strategic usage of color: (A) One color; (B) Two colors; (C) Many colors

fundamentally oriented for action due to evolutionary nature for survival.³ A typical task in a visual display that can be accomplished in <200 to 250 ms is considered as preattentive.⁵ Therefore, a significant advantage of exploiting a preattentive attribute in visual perception is that process in the brain happens at an incredible speed as compared to attentive attributes. Thus, the researcher needs to maximize the usage of preattentive attributes for making effective graphs. Figure 3, besides color, gives a list of various other preattentive attributes for readers' perusal.

GESTALT PRINCIPLES

The working memory works at an incredible speed. Moreover, when a human being processes visual information brain automatically starts finding patterns to take action. However,

a significant limitation of working memory is the amount of information it can process and retain. The capacity and speed of working memory can be enhanced with grouping cues. In 1920, the Gestalt principle of visual perception gave insights about brain functioning in terms of distinguishing patterns and organizing information for utilization. Later, it came to be known as the Gestalt principle of proximity, similarity, enclosure, closure, continuity, and correction.^{6,7} Figure 4 describes and demonstrates the Gestalt principles. The visual perception can be enhanced by understanding and application of the Gestalt principle for effective communication. It helps in reducing clutter which increases the cognitive load in processing information. Therefore, researchers need to understand and apply the Gestalt principle for improving the design of tables and graphs. The Gestalt principle of similarity was used in Figure 1B.

Fig. 3: A list of preattentive attributes for effective visual display

Principles	Visual Display Example			Explanation
	A-illustration	B-illustration	C-illustration	
Proximity				Closer objects are perceived as group
Similarity				Objects with similar attributes (shape and color) are perceived as groups
Enclosure				Objects with boundary (area marked with a line or same color) are perceived as group
Closure				Open structures are perceived as complete
Continuity				Aligned objects are perceived as groups
Connection				Connected objects are perceived as groups

Fig. 4: Gestalt principle of visual perception

CONCLUSION

The charts are essential and effective communication tools to convey results. However, the majority of researchers use default graph options. Similar to tables, default graph options do not automatically and objectively convey the message of researchers to the audience. The preattentive attributes and the Gestalt principles can be leveraged for actionable insights from the data. Therefore, the researcher's need to understand and strategically apply preattentive attributes and Gestalt principles to reduce clutter and cognitive load for better communication.

ACKNOWLEDGMENTS

We acknowledge Dr Suchet Sachdev and Dr Meenakshi Sharma for their valuable time and inputs to improve the quality of the article.

REFERENCES

1. Medina J. Brain Rules: 12 Principles for Surviving and Thriving at Work, Home and School. Pear Press.
2. Tufte ER. The Visual Display of Quantitative Information. USA: Graphics Press, 2001.
3. Few S. Show me the numbers 2012.
4. Knaflitz CN. Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons; 2015.
5. Healey CG, Enns JT. Attention and visual memory in visualization and computer graphics. *IEEE Trans Vis Comput Graph* 2012;18(7):1170–1188. DOI: 10.1109/TVCG.2011.127.
6. Peterson DJ, Berryhill ME. The gestalt principle of similarity benefits visual working memory. *Psychon Bull Rev* 2013;20(6):1–16. DOI: 10.3758/s13423-013-0460-x.
7. Schwabish J. Better presentations: A guide for scholars, researchers, and wonks. Columbia University Press; 2016.