DATA VISUALIZATION FOR THE NEW AGE: AN EXPERIMENTAL STUDY EXPLORING THE MERITS OF HUMAN PERCEPTION ON VISUALIZATION TECHNIQUES FOR IMPROVED STATISTICAL COMMUNICATION

Isabella Benabaye, Dominic Dayta, Patricia Rose Donato, and Sean Oliver Escalante
School of Statistics
University of the Philippines, Diliman
pbdonato@up.edu.ph

Conventional data visualization techniques give little consideration for how human perception affects one's comprehension of statistical charts. Motivated to address this inadequacy, the present paper seeks to study the improvement in statistical communication when visualization techniques are integrated with the Gestalt principles from cognitive psychology. Students from a local high school are made to take a timed chart comprehension exam where their test scores represent their level of understanding--to be compared based on the type of chart used ('conventional' versus 'Gestalt-modified')--hypothesizing that the latter is more effective in delivering its intended message to the student. This study aims to provide an evidence-based analysis towards finding more effective and efficient data visualization techniques in an age of data overload.

INTRODUCTION

We are living in the age of big data. Our increasing production and reliance on data poses a challenge not only to statistics as a growing field, but also to how we teach and learn statistics (Tishkovskaya & Lancaster, 2012). Calling for a new breed of data presentation for both educators and learners of the field, data today must not only be presented in a clear manner, but also in a way that makes it strong enough to be obvious and evocative (Alexandre & Tavares, 2012).

Despite the use of quantitative charts beginning even prior to the 18th century, it experienced little innovation in effectiveness and efficiency until late in the 20th century. In fact, through time, only basic chart types (e.g., line charts, bar charts, and pie charts) have become familiar to many (Few, 2017). While sufficient documentation on these standard statistical charts are in agreement about their proper use (e.g., preventing data misinterpretation through proper scaling and labelling), no single set of guidelines has been agreed upon as the gold standard. In the rapidly growing field of statistics, recent studies have been suggesting that traditional representation approaches are becoming insufficient and inefficient (Pandey, et. al., 2014; Alexandre & Tavares, 2010). Improving these standard charts, however, is not merely a development borne of evolving sensibilities in aesthetics. Just as computational tools for data analysis are based on theoretical knowledge, proper visual representation must be more than just being arbitrarily aesthetic and self-explanatory (Ware, 2004).

A way to improve standard charts is inspired by Gestalt perception theories. It is grounded on the principle that the whole speaks more than its parts. That is, it is only when one is subjected to seeing the totality can their brain translate it into a concept (Guberman, 2012). By using Gestalt principles in designing data visualization tools, it promotes better comprehension by linking many types of elements found in the chart, instead of focusing on the conception of the chart's content itself (Andersson, 1986). For example, *proximity* and shared visual properties (e.g. color *similarity*) of a data series name and its corresponding chart element indicate they appear as one group. Another intuitive Gestalt principle is *figure and ground*, which guides the reader to what chart element they should see first by graying out the remainder. *Enclosure*, a less common technique, can change the pattern the reader perceives by superimposing a certain shape on the data points. Out of the eight Gestalt principles, the aforementioned principles are those that have been cited as relevant in the enhancing the standard statistical charts (Kobourov, et. al., 2015; Meeks, 2015).

In this paper, the 'standard' visualization is based on the minimum guidelines for clarity and accuracy of data presentation as suggested in *Elementary Statistics* (Almeda, Capistrano, & Sarte, 2010). This study considers going beyond these standards by revising them into 'enhanced' visualizations, redesigned in a manner that takes into account the dynamics of human perception and cognitive processing of images -- a key preoccupation of the Gestalt perception theory.

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

Berinato (2016) provides detailed prompts for revising charts from ‘standard’ to ‘enhanced’, and is used as a primary reference for developing the instrument utilized for the experiment.

With scarce literature to work with, like most recent experimental investigations on data visualization, the study becomes a pioneering exploration: this study aims to fill the gap between psychological principles and statistical data analysis with insights gathered from performing a controlled experiment, onet that explores whether the integration of human perception into the design of data visualization tools has merits that make it superior to standard practices.

METHODOLOGY

An experiment was performed on ninth grade students from a private school in San Juan, Metro Manila, Philippines. A total of 18 participating students were randomized into two equal groups each taking a chart reading exam that made use of either the ‘standard’ or the ‘enhanced’ visualizations. As the students are non-native speakers of English, a reading comprehension exam was administered prior to the chart reading exam to minimize unwanted mistakes due to poor comprehension of the exam’s questions. A posttest, presenting both the enhanced and standard versions of each chart in the exam, was also given to the students asking which version would have been more helpful in answering the exam. While the posttest is not included in the design itself, it was administered to know more about the students’ preference regardless of their performance. In totality, the experiment followed a randomized single-treatment design with covariate.

To cover as much ground in terms of the students' ability to interpret charts, the exam covered three of four chart groups defined by Andrew V. Abela and discussed in Berinato (2016): *composition* (e.g. pie and stacked bar charts), *comparison* (e.g. radar and line charts), and *relationship* (e.g. scatter and bubble charts). In creating the enhanced charts, certain redesigns involved only aesthetic revisions, such as when a standard line chart was reinforced with relevant Gestalt principles (Figure 1). On the other hand, some involved a complete revamping of the chart, such as in addition to the integration of Gestalt principles, a stacked bar chart on a two-year period of time was transformed into a slope chart (Figure 2).

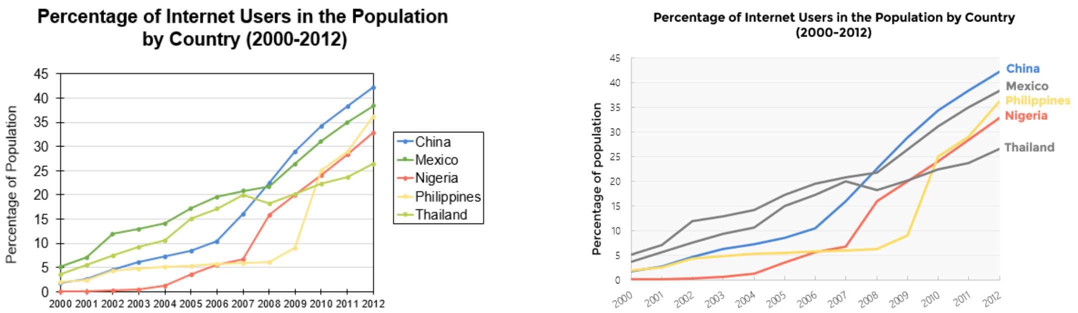

Figure 1. Aesthetic revisions on a multiple line chart from standard (left) to enhanced (right)

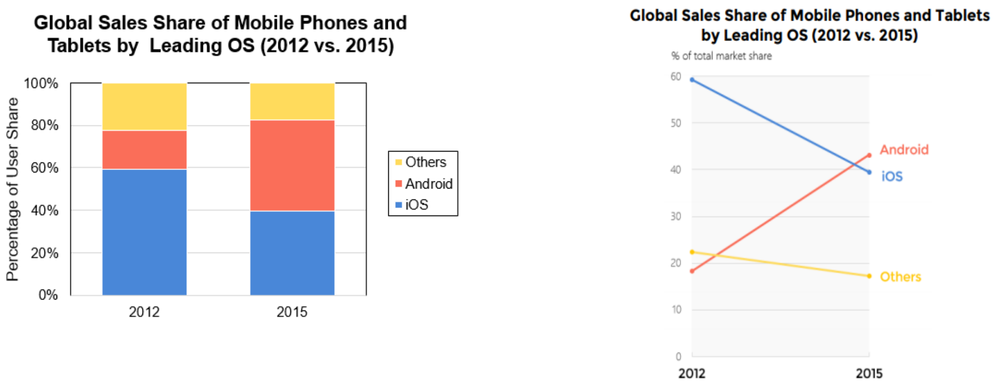

Figure 2. Completely revamping a stacked chart (standard, left) to a slope chart (enhanced, right)

The chart reading exam has a total of 12 charts, with four charts coming from each of the selected chart groups: composition, comparison, and relationship. Both the standard and enhanced chart exams contained exactly the same items, with the same questions, on the same layout, except that the enhanced charts are redesigned versions of those in the standard exam. In redesigning the charts, half of them made use of aesthetic revisions, while half were completely revamped.

The exam included 12 main items, each with its own chart from which the student had to deduce the answers to three questions. In summation, the test was 36 questions in length, to be taken without a time limit. Nevertheless, the students were encouraged to answer with both speed and accuracy. The time they took to answer the exam was used to scale their raw exam scores into a final score:

$$\text{Final Score} = (\text{\% Correct})*(1-W) + (1 - \text{Speed Percentile})*W$$

where *W* is the weight of speed. Equivalently, *1-W* becomes the weight of accuracy (as measured by percentage of correct answers in the exam). The formula, when applied, results into a combined grade ranging from 0 to 100, representing the student's speed and accuracy when taking the test. Multiple possible values of *W* were considered, beginning with a combined grade that is composed purely of 100% accuracy, then gradually increasing the weight of speed towards a combined grade that takes into account 50% accuracy and 50% speed. Such a technique facilitates an exploratory analysis of how divergent the scores become between students using the enhanced charts versus those using the standard charts when speed becomes more or less considered.

DESCRIPTIVE STATISTICS

The two chart types appear to have different effects on the respondent's level of comprehension. The boxplots below reveal similar median raw scores of 28 points, but with greater variability in the scores of students that used the standard charts, (they also have the lowest score at 18 points). They completed the exam slightly slower, averaging two minutes longer (26.11 minutes) than those who used the enhanced charts (24.22 minutes). While the time taken to finish the exam slightly varies more for those who were given the enhanced charts, the group's performance was actually not only better, but was found to be much more consistent.

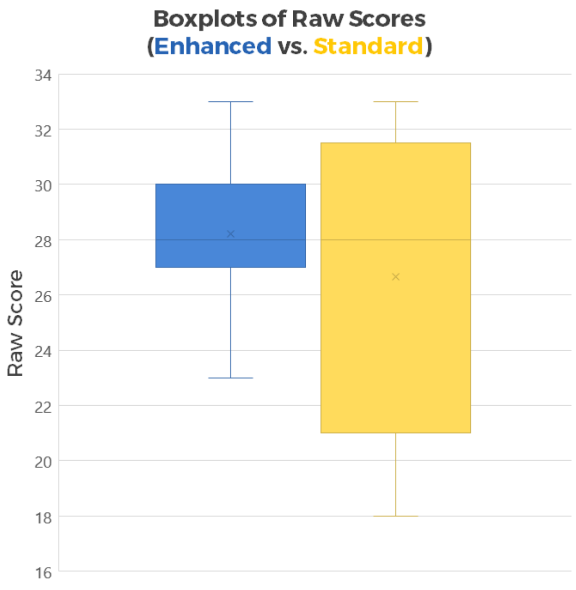

Figure 3. Comparing the two groups' raw scores

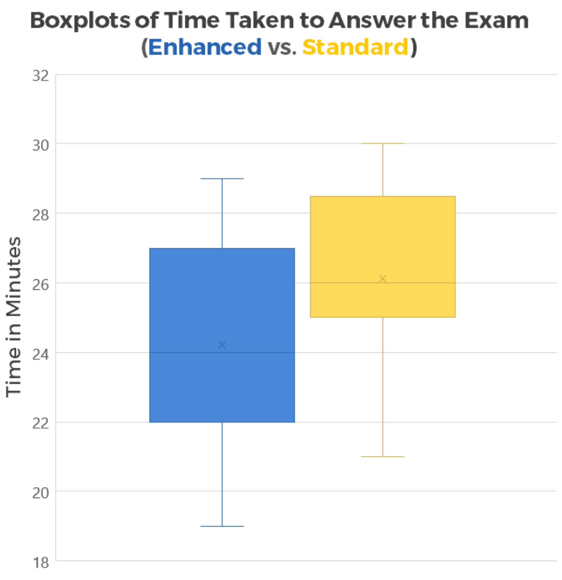

Figure 4. Comparing the time it took the two groups to complete the exam

Examining the students' preferences for the visualization type, the table below shows the percentage of times the enhanced chart was preferred and the percentage of correct answers for each item, wherein the answers were grouped by whether the charts used were standard or enhanced.

Table 1. Differences in preference and performance for each exam item based on the charts used

Charts Used	Item Number											
	1	2	3	4	5	6	7	8	9	10	11	12
	Percentage of times the enhanced chart was preferred over the standard version											
Enhanced	22%	33%	67%	78%	67%	89%	22%	22%	78%	44%	78%	67%
Standard	44%	33%	44%	67%	22%	22%	33%	44%	78%	67%	89%	89%
	Percentage of correct answers when enhanced or standard chart was used											
Enhanced	81%	81%	85%	56%	96%	89%	100%	74%	85%	74%	41%	78%
Standard	70%	70%	85%	67%	81%	89%	85%	81%	59%	74%	48%	78%

Note: Enhanced charts for Items 1 to 6 are complete revamps of their standard versions, while those for Items 7 to 12 are simply aesthetic revisions of their standard versions.

Both groups often prefer the enhanced versions when their enhancements only involve aesthetic revisions, specifically when applied to scatter plots (items 9 and 10) and bubble charts (items 11 and 12). The enhancements for these charts primarily used the Gestalt principle of enclosure, which despite being the least common technique, was seen by the students as helpful in finding patterns. The students' performance on the same items, however, do not differ that much, for only in Item 9 did the students who were given the enhanced charts outperform those given the standard charts.

For the items that used completely revamped charts, the standard group strongly prefers the charts they were originally assigned to, which is not a surprise given that these revamped versions may have been completely new to the students. In addition, items where both groups clearly disfavor the enhanced charts are Items 1 and 2, and Items 7 and 8, whose standard versions, interestingly, are pie charts and line charts, respectively. This preference for the standard versions is probably due to the students' strong familiarity with such charts. Looking at the percentage of correct answers for these same items however, it is evident that the performance is better for those who were given the enhanced charts. In fact, for the items where the enhanced group prefers the standard charts, their percentages of correct answers are generally higher than those of the standard group.

MODEL

Figure 5 shows how the average performance of students taking each exam type diverges as the importance of speed is increased in computing the final score. The leftmost side of the horizontal axis represents the scoring method where speed is given 0% weight (i.e. the final grade comes purely from the students' accuracy in answering the exam), while the rightmost side is where speed is given 100% weight. For every scoring method, the LS means was computed from an ANCOVA model discussed in Montgomery (2013), with the following form:

$$(\text{Final Score})_i = \beta_0 + \beta_1(\text{Exam Type})_i + \beta_2(\text{Pretest Score})_i + \epsilon_i$$

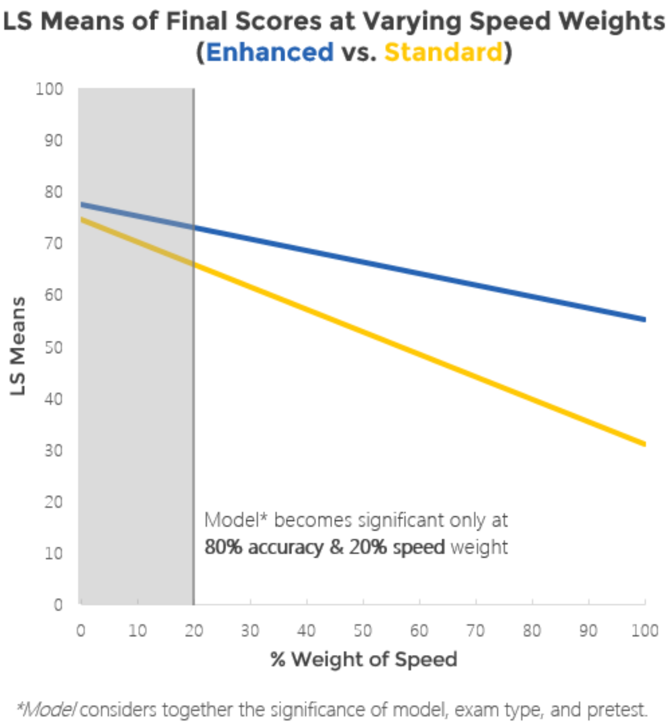

Figure 5. Performance of standard and enhanced groups using different scoring methods

When the students' performance is measured purely based on accuracy, the average scores between the two groups of students exhibit no significant difference. However, as speed gets factored in with greater weight into their final score, the performances diverge. At a threshold of 20% weight for speed and 80% weight for accuracy, the chart type begins to significantly affect the students' performance on the test, with those given the enhanced charts performing better. This shows that accuracy alone fails to distinguish the merits of using Gestalt-modified visualizations, as the existing standard techniques taught in statistics textbooks are already designed for presenting data accurately and with no distortion. It is when speed is considered that this new breed of visualization claim their merits: making accurate data insights faster, more immediate.

With the scoring weights of 80% for accuracy and 20% for speed, the table below shows the results of the ANCOVA model considered for this study.

Table 3. Proposed Analysis of Covariance Model

Source of Variation	df	Sum of Squares	Mean Square	F-statistic	P-value
Model	2	497.89	248.94	4.43	0.0308**
Error	15	842.85	56.19		
Corrected Total	17	1340.74			
Exam Type (Standard or Enhanced)	1	210.78	210.78	3.75	0.0718*
Pretest	1	199.12	199.12	3.54	0.0793*

** $p < 0.10$, **$p < 0.05$. R^2 =39.67%. Diagnostic checking reveals no violation of assumptions.*

It can be seen that not only the exam type was significant, but also the pretest results. This shows how different levels of reading comprehension could induce heterogeneity among the students' performance, and is indeed an important factor to account for in a chart comprehension exam. With speed and accuracy taken into account, using such perception principles indeed leads to an overall better comprehension of quantitative data, supporting the intention of this new breed of visualization to -- as Edward Tufte put it -- above all else, show the data.

CONCLUSION AND RECOMMENDATION

Despite maintaining a preference for the standard statistical charts, likely due to familiarity, the students' comprehension of the data presented to them was made generally more efficient when they used the enhanced visualizations, as measured by their exam performance. These results strengthen the grounds of specialists in the emerging field of data visualization who seek to marry statistics and psychology in pursuit of a scientific approach to making clearer, more efficient, and therefore more useful charts for communicating quantitative insights. To further explore the effects of such improvements on visualization, researchers are recommended to investigate, using a psychological approach, the specific skills that comprise chart reading ability, along with alternative methods for analysis such as the cognitive diagnosis models and item response theory models.

REFERENCES

Alexandre, D. S., & Tavares, J. M. (2010). Introduction of human perception in visualization. *International Journal of Imaging, 4*(10), 60-70.

Almeda, J. V., Capistrano, T. G., & Sarte, G.M.F. (2010). *Elementary Statistics*. Diliman, Quezon City: University of the Philippines Press.

Andersson, B. (1986). The experiential gestalt of causation: a common core to pupils' preconceptions in science. *European Journal of Science Education*, *8*(2), 155-171.

Berinato, S. (2016). *Good charts: the HBR guide to making smarter, more persuasive data visualizations*. Boston, MA: Harvard Business Review Press.

Gestalt Principles for Data Visualization. (2015, March). Retrieved November 27, 2017, from https://emeeks.github.io/gestaltdataviz/section1.html

Few. S. (2007, January 10). *Data visualization: Past, present, and future*[PDF]. Cognos Innovation Center for Performance Management. Retrieved November 15, 2017.

Guberman S (2015). On Gestalt theory principles. *Gestalt Theory*, *37*(1), 25–44.

Kobourov, S. G., Mchedlidze, T., & Vonessen, L. (2015). Gestalt Principles in Graph Drawing. *Lecture Notes in Computer Science Graph Drawing and Network Visualization,* 558-560.

Meeks, E. (2015, March). Gestalt Principles for Data Visualization. Retrieved November 27, 2017, from https://emeeks.github.io/gestaltdataviz/section1.html

Montgomery, D. C. (2013). *Design and analysis of experiments*. John Wiley & Sons: Hoboken, New Jersey.

Pandey, A.V., Manivannan, A., Nov, O., Satterthwaite, M. L., & Bertini, E. (2014). The persuasive power of data visualization. *New York University Public Law and Legal Theory Working Papers.* 474.

Tishkovskaya, S. & Lancaster, G. (2012) Statistical education in the 21st century: a review of challenges, teaching innovations and strategies for reform, *Journal of Statistics Education, 20*(2).

Tufte, E. R. (1983). *The visual display of quantitative information*. Graphic Press: Cheshire, Connecticut.

Ware, C. (2004). *Information visualization: perception for design.* San Francisco: Morgan Kaufmann Publisher.