

The Visual Display of Quantitative Information (Second Edition). *By Edward R. Tufte.* Graphics Press, Cheshire, CT, 2001. 197pp., hardcover. ISBN 0-9613921-4-2.

It was with great anticipation that I awaited the review copy of the second edition of Tufte's "The Visual Display of Quantitative Information". The first edition is viewed as the best of Tufte's several books, at least amongst the data visualization community, so, the prospect of a revised first edition was exciting. But for all practical uses the second edition is the same as the first. There are several re-generated images, a little more color than the first edition, and the errors found between the first printing and this have been corrected. The second edition is dedicated to the memory of John W. Tukey, an original researcher and major innovator in the Statistics community in the second half of the 20th century. This dedication grounds Tufte's book in Tukey's rich legacy, with the influence of data visualization research acknowledged.

Tufte's books are indisputably the icons of today's information visualization research. His two other books are titled "Envisioning Information" and "Visual Explanations". They are tangibly accessible books which reach out to the community at large: business managers, newspaper editors, web designers, architects, academics. Some of the appeal of the books stems from the wealth of examples that Tufte has amassed from across the globe, and meticulously cataloged. These examples are used to introduce and explain his self-constructed guidelines for generating graphical displays of information.

"The Visual Display of Quantitative Information" is the first of Tufte's books, and the most quantitative of all of his books. The material in the book arises from the area of data visualization: "using points, lines, a coordinate system, numbers, symbols, words, shading, and color" to "display measured quantities". His book nicely documents the history of this pursuit. But it quickly shifts from examples of data visualization to cartographic examples, and further to visualization of diverse types of information such as train schedules. It lays the groundwork for the shift of focus in his later books to more community-drawn examples - taken from the news media, signs and advertisements - which provide content more consistent with a coffee table book than a scientific discourse. This is actually a strength of Tufte's work: a disregard for boundaries between sometimes disparate areas of visualization research and success in making academic pursuits accessible to the public at large. We will expand on these points later in the review, to place Tufte's books in the context of the active fields of research on visualization. But first we will describe the content of "The Visual Display of Information" in more detail.

What's in the book?

This book is divided into two parts: (I) graphical practice and the (II) theory of data graphics. In the first part, using examples he lays out the Tufte principles for producing data graphics. The first chapter contains excellent graphics, the second chapter contains awful graphics, and the third chapter seeks to determine the processes that lead to poor graphical displays. In the second part, the Tufte

principle tenets are described in detail in separate chapters: careful use of data ink, discarding chartjunk, use of small multiples and aesthetics.

What are the Tufte principles for good graphical displays of data? On the first page of the first chapter, this is what Tufte has to say:

“show the data
induce the viewer to think about the data...
avoid distorting what the data have to say
present many numbers in a small space
make large data sets coherent
reveal the data at several levels of detail
serve a reasonably clear purpose
be closely integrated with the statistical and verbal descriptions of data.”

His first example graph demonstrates the need for graphics in data analysis: four plots of data exhibiting dramatically different patterns yet the data have identical numerical statistics. Statisticians should take note of this example. A well-constructed graph can reveal much more about the data than numerical statistics.

In chapter 1 of part II of the book Tufte unifies several of the above points into the concept of “data-ink”, the non-erasable core of a graphic. He introduces the data-ink ratio, calculated from dividing the amount of ink used to draw data by the total amount of ink used to print the graphic. To be a good graphic this ratio should be large, close to 1: most ink should be used to represent data.

The chapter 2 of part II is devoted to the concept of “chart-junk”, graphical decoration which distracts the viewer from the data. Some awful (and comical) examples are included. He proffers some reasons why chart-junk exists, early in the book, in chapter 2:

“... was the assumption that data graphics were mainly devices for showing the obvious to the ignorant. It is hard to imagine any doctrine more likely to stifle intellectual progress in a field. The assumption led down two fruitless paths in the graphically barren years from 1930 to 1970: First, that graphics had to be ‘alive’, ‘communicatively dynamic’, overdecorated and exaggerated (otherwise all the dullards in the audience would fall asleep in the face of these boring statistics)....”

Chapters 2 and 4 of part II of the book discuss multi-functioning graphical elements and small multiples. Because the data being presented is quite often multi-dimensional in nature, and the paper that it is being drawn on is 2-dimensional, a data graphic needs to extend the viewer’s mind beyond the page. There are several ways to do this: multi-resolution, such as a stem-and-leaf

plot, which presents an overview of the shape of the data distribution with the precision of the data numbers themselves; small multiples, where small frames like movie frames, quite often generated by subsetting data by conditioning on one or two measured quantities, are arranged beside each other.

The last chapter focuses on aesthetics. The major point is: “Graphical elegance is often found in simplicity of design and complexity of data.” A poor graphic has this tenet backwards.

How does this book relate to other visualization work?

Intuition vs Experimentation: “The Visual Display of Quantitative Information” can be viewed in statistical terms as an observational study. Tufte has studied the historical development of data graphics. He has collected and cataloged many, many examples of data graphics. Tufte is like the classical botanist or zoologist who meticulously collects, records and preserves specimens. From his collection, though, he has synthesized a list of principles which can be viewed as hypotheses for good graphical displays. The weakness is these hypotheses have not been tested experimentally. Perhaps this weakness has contributed to the religious fervor with which Tufte is viewed by the general community. Without scientific experimentation, the ultimate judgment about the excellence or lack of excellence in a graphic remains the domain of the guru. Tufte is the guru of data graphics, whether he likes this or not. He will remain there until there is a body of work confirming or negating his principles. This is not to say that Tufte is not scientific, quite the contrary, what he has done is a very important service to the community, and his examples and sources are well-documented. To gain more power from Tufte’s cataloging takes experimentation, carefully designed experiments to test human perception of data graphics. The array of examples that Tufte provides is dizzying. The catalog needs to be de-constructed into simpler components that extend broadly across graphical components.

There is a history of experimental studies of graphics. Perhaps the most famous example is the work of Bertin (1983). Cleveland (1994) uses the last quarter of his book to describe his experiments in human visual perception for data graphics. Brewer (1999) focuses on the use of color, and appropriate choices of color combinations. Wilkinson (1999) formalizes visual thinking into a grammar of graphics.

Information visualization, data visualization, cartography and scientific visualization: Tufte’s books are most naturally classified as contributions to the information visualization field. This is a relatively recently recognized field of research, although from Tufte’s examples its an area that pre-dates data visualization. Data visualization perhaps is contained as a subset of this field: It focuses on information that can be represented in a table or list. Information visualization is broader, sometimes as simple as visualizing the data in the context from which it arises, but more generally it attempts to display information that is not yet processed or coded into data. Numerous books have emerged in the past 3 years on information visualization (Card, MacKinley & Schneiderman 1999, Spence 2000, Ware 2000). Most of this literature does not

adequately describe multivariate data visualization. Cartography is much older than either field, and has functions and utilities that transcend beyond purely spatial data visualization. Scientific visualization has arisen from the computer graphics revolution, with the purpose to render 3D, or further 3D+time, in as realistic a manner as possible. Perhaps it is thanks to this field that we have stunning computer-generated scenes that go beyond reality in today's movies and computer games.

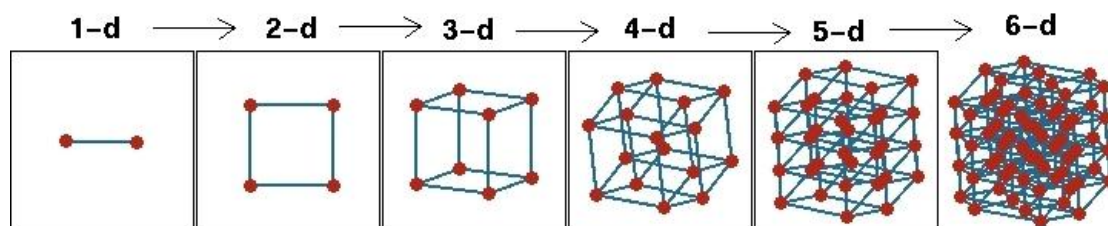
Presentation vs Exploration: Visualization research has been energized by the advances in computer technology. Along with information and data visualization the fields of cartography, scientific visualization, and computer graphics have grown and expanded into each other's domains. Cartographers are now integrating statistical data graphics with maps, and statisticians are integrating maps with statistical graphics displays. With the computing technology has come the power to push and pull, rotate and label graphical elements in real-time, interactively on the computer screen. These transient and fleeting data displays can truly be dynamic, and come alive. Make no mistake, though, that these advances are not to amuse the dullard, they allow data analysts to become explorers, facilitating exploratory data analysis and data mining, with rapid re-drawing, and fast querying of data. (See (Cleveland & McGill 1988, Buja & Tukey 1991, Wegman & Carr 1993) for some early examples.) A discourse on these tools is absent from Tufte's books. His books deal primarily with presentation graphics, once the defining features of the data are found, how do we present it accurately and informatively. We are yet to have a comprehensive catalog of interactive and dynamic graphics, but there are some taxonomic beginnings (Buja, Cook & Swayne 1996, Wilhelm 2000)).

Summary.

Tufte's books are truly gems. They are beautifully and thoughtfully executed. The introduction to the second edition of "The Visual Display of Quantitative Information" describes some of the sacrifices that Tufte endured to bring the books to print. This book remains my favorite of all his publications.

Multidimensional Spaces Can Be Visualized.

To reiterate one message of the book: Data visualization involves going beyond the 2-dimensional page to draw the viewer's mind into multi-dimensional spaces. Thus we finish with an original example (not taken from Tufte) involving a simple, and very elegant geometric shape: the cube. The figure below illustrates the evolution of a 1-dimensional cube (a line) to a 6-dimensional cube. A single 2-dimensional projection of the cube is shown in each frame. The simple cube underlies many applications in statistics: for example, experimental design, coin flipping experiments, and studying contingency tables. Notice (1) how the cube grows a dimension by doubling itself, (2) each dimension expands the physical limits of the data within the plot region, to the point of bursting out of the assigned space by 6-d, and (3) that the points concentrate in the center as dimension increases - roughly illustrating an effect known as the central limit theorem in statistics.



References

- Bertin, J. (1983), *Semiology of Graphics: Diagrams, Networks, Maps*, University of Wisconsin Press, Madison, WI.
- Brewer, C. A. (1999), Color Use Guidelines for Data Representation, in 'Proceedings of the Section on Statistical Graphics', American Statistical Association, Baltimore, pp. 55-60, See also <http://www.personal.psu.edu/faculty/c/a/cab38/>.
- Buja, A., Cook, D. & Swayne, D. (1996), 'Interactive High-Dimensional Data Visualization', *Journal of Computational and Graphical Statistics* **5**(1), 78-99. See also www.research.att.com/~andreas/xgobi/heidel/.
- Buja, A. & Tukey, P., eds (1991), *Computing and Graphics in Statistics*, Springer-Verlag, New York, NY.
- Card, S. K., MacKinley, J. D. & Schneiderman, B. (1999), *Readings in Information Visualization: Using Vision to Think*, Morgan Kaufmann Publishers, Inc., San Francisco, CA.
- Cleveland, W. S. (1994), *The Elements of Graphing Data (revised edition)*, CRC Press, New York, NY.
- Cleveland, W. S. & McGill, M. E., eds (1988), *Dynamic Graphics for Statistics*, Wadsworth, Monterey, CA.
- Spence, R. (2000), *Information Visualization*, Addison Wesley, Harlow, England.
- Ware, C. (2000), *Information Visualization: Perception for Design*, Morgan Kaufmann, San Mateo, CA.
- Wegman, E. J. & Carr, D. B. (1993), Statistical Graphics and Visualization, in C. R. Rao, ed., 'Handbook of Statistics, Vol. 9', Elsevier Science Publishers, Amsterdam, pp. 857-958.
- Wilhelm, A. F. (2000), 'Interactive Statistical Graphics: The Paradigm of Linked Views.', Habilitationsschrift, Universität Augsburg.
- Wilkinson, L., ed. (1999), *The Grammar of Graphics*, Springer, New York, NY.