

# The Art and Science of Biostatistical Display

Dominic T. Moore, Duke Clinical Research Institute, Durham, North Carolina

## Introduction

Graphs communicate. There is both an art and a science to the visual display of biostatistical data. This paper is about visual communication between the data analyst and an audience of interest. The analyst wants to tell the story of the data. The major emphasis is placed on the principles as described by Edward TUFTE in his articles, books and seminars.

This paper will provide practical guidelines for the data analyst who is looking to communicate data analysis results by means of presentation graphics. This is not about exploratory data analysis, the interaction of the data analyst and the computer, but the art and science of visual communication. The author asserts with only a hint of exaggeration, that he puts forth not one original idea, but rather restates the major principles by the who's who of graphs and data visualization, (Cleveland , Playfair, ,Tufte , Tukey, Wainer, Young). The goal is to give guidelines so that one can strive for excellence in graphics whose principles come from graphic design in art and science.

## The Purpose of Graphs

A good graph will:

- **Save the user time.**  
Trivia and extraneous matter is dispensed with - information is sifted and refined.
- **Gain user attention.**  
The display should be pleasing and exciting to the eye and generate reader motivation.
- **Show relationships and permit study of the data.**  
The user should be able to look for regularities / irregularities within the data. Study of data points and their spatial relationships to each other may reveal meanings not otherwise observed. In deciding whether to use a chart or a graph, the form which best shows the relationships should be used.
- **Suggest new ideas.**  
A good graphic might reveal connections that otherwise would not be perceived.
- **Make efficient use of information:**  
It will display fundamental relationships by way of visual conventions.

## Three things will happen with a good graph

1. It will be studied.
2. It will be accepted.
3. It will be remembered.

If you are presenting information, don't you WANT this to happen?

## Tufte's Principles for the Design of Good Graphic Display

Edward R. Tufte's The Visual Display of Quantitative Information is "*A tour de force*" according to John Tukey. Everyone who is interested in graphic presentation should have this gorgeous, important book.

TUFTE says STRIVE for GRAPHIC EXCELLENCE by stressing the following points:

- A well designed presentation of interesting data is a matter of substance, of statistics and of design.
- Complex ideas should be communicated with clarity, precision, and efficiency.
- Graphics should give the viewer the greatest number of ideas in the shortest amount of time while using the least ink in the smallest place.
- The true data must be shown.
- Most relationships, and therefore graphical displays of the relationships, are multivariate. Graphics should emphasize this important fact.

A summary of TUFTE's thoughts on graphical displays includes many important fundamentals in the field. He says that the DESIGN of a GOOD GRAPHIC DISPLAY should:

- SHOW the data and encourage the reader to think about SUBSTANCE but NOT methodology, technology or something else.
- Do NOT distort what data has to say.
- Be closely integrated with statistical and verbal descriptions of the dataset.
- Present many numbers in a small space.
- Make large data sets coherent.
- Reveal the data in many levels of detail.

- Serve a reasonably clear purpose

Description  
Exploration  
Tabulation  
Decoration

- Enforce VISUAL COMPARISONS.
- Focus on CONTENT driven design.
- Be EXPLANATORY and not just descriptive.
- Show not just a mean or simple estimate but display DISPERSION as well.

#### Tufte's rules for Graphic Integrity include:

- Display of numbers that are represented in direct proportion to the value they have in the dataset.
- Clear, detailed and thorough LABELING should be used to defeat graphical distortion and ambiguity. Definitions and explanations of the data should be written out and important events in the data should be clearly marked (i.e., the *annotate* feature in statistical graphics packages).
- Data variation and NOT design variation should be shown.
- Graphics must NOT quote or display data out of context.

TUFTE encourages those who design statistical graphics to:

- Make graphs that are REPRODUCIBLE.
- Place data in the proper context.
- Plot the same data in several ways to make sure the story you want to tell is being told.
- Be clear about your thinking.
- Be clear about your design.
- Always try to follow good design principles.
- Treat the audience with endless RESPECT.
- REMEMBER, graphs have IMPACT.

#### Guidelines for the Consultant who does Graphic Display

The experts in graphic display emphasize recurring principles.

- SHOW the data accurately.
  - Watch out for data distortions.
  - Don't insist that zero always be included.
  - Use NATURAL scales.

Choose appropriate scales when groups are being compared.

Use scale breaks only when necessary.

Don't connect values on two sides of a break.

Check and recheck for "human errors".

- SHOW comparisons as much as possible !! Good graphs INVITE comparisons. TUKEY reminds us to ask the following questions:

*Are the values small, medium or large ?*

*Is the difference or change up down or neutral ?*

*Is the difference or change small, medium or large ?*

*Are successive changes growing, shrinking or staying constant ?*

*What about change in ratio terms ?*

*What is the direction or scatter of the change ?*

*What pattern does it show ? OR is there an absence of pattern of interest itself ?*

*Values all about the same*

*No definite curvature*

*Vertical scatter does not change*

- Construct a graph that is data rich.

#### TUFTE'S Data-ink Ratio:

The amount of ink used for graphing to the total amount of ink in the graph. Zero, is low (bad) and means there is too much decoration and not much information.

TUFTE'S Data Density Index: The number of numbers plotted per square inch of display.

- Stay away from 3-D graphics, pie charts, anything that's "cutsie", and anything that smacks of default options.
- Remember the benefits of black and white (see section below).
- Don't mess with color unless you really have to. This goes with the saying:

*"Just because you CAN do something doesn't mean you SHOULD.."*

**BESSLER** points out that:

*Color does NOT improve a poor design.*

*Color MUST have a function - use it for COMMUNICATION, NOT decoration.*

*Color prioritizes information whether meant to or not.*

*The less color the better.*

**TUFTE** calls color used by computer software companies, "*A grim parody of video games.*"

HOWEVER, executives like color (and shiny objects as well). You may have to give the customer what they want.

- Benefits of BLACK and WHITE.

Compared with color, black and white hard copy is faster, cheaper, more reliable and easier to use. Additionally, simpler equipment can be used and there is no agonizing over color strategy.

Perhaps most important is that black and white graphics are more copyable than color. and GOOD graphs GET COPIED.

**Learn from others.** There is no need to reinvent the wheel.

**TUKEY** says:

*"...Learn from the classical era of statistical graphics..."*

*"...plagiarize from the great ones."*

*"...Make up a portfolio of templates and COPY them."*

**WAINER** sarcastically remarks that "*...if it has been done well in the past, try something new*"

- Look for what works for you in books, magazines and on Web pages. Review the work of chartologists including map makers, who's use of color is particularly wonderful and often studied. Most importantly, Read and look critically at journals and periodicals in YOUR AREA OF INTEREST.
- Read the LAY PRESS such as Chance, The Skeptical Inquirer, Scientific American, Byte and other various computer magazines.
- Observe and stay aware of new techniques and tools in the graphical sections of the SUGI Proceedings: I particularly recommend the works of Art Carpenter, LeRoy Bessler, Lori Griffith, and the annual Winners of Graphics Competition.
- Review books for the specific graphics packages that you use, such as SAS Institute's Books By Users (i.e., Michael Friendly's book on Graphics in SAS).
- Learn from other statistical packages such as:
  - S-Plus
  - VISTA by Forrest Young
  - Lisp-Stat
  - Stata
  - Mathematica, etc.

- Of course, review and emulate TUFTE, TUFTE, TUFTE.

- Keep a NOTEBOOK of your own programs and graphic examples from others (How QUICKLY we forget code!).

### Summary Points for the Consultant

- You should be able to free-hand sketch what you want to show OR have your client draw an approximation.
- Don't be afraid of revision.
- Edit, edit and then edit some more.
- SHOW alternatives to your client.

Try different symbol sizes, jittered data points, and different labeling schemes (annotation skills are important here).

- Ask the client to talk about what it is they're trying to show / emphasize.
- Don't ASSUME that they know what they are trying to show (or even if they know what they are talking about). Your job is to make the client look good.....even if it is only to himself.
- Don't put titles in all caps.
- Break long titles so they can be read easily.
- Avoid abbreviations and acronyms when possible.
- Annotate and label when necessary.
- Use bold, italics, underlines and change of size.
- Keep conventions consistent throughout the document.
- Strive to be explanatory with your graphs - Do they tell the story of the phenomena ?
- The excitement should come from the phenomena revealed.....not necessarily the "flash" of the graph.
- Examine the data carefully enough to know what it has to say and then let it say it with a minimum of adornment (Wainer).
- DON'T clog up a graph with a lot of chart junk
- Don't break a design "rule" unless you have a good reason. Breaking a design rule can have a tremendous impact, so make sure that impact you are making is something that you want. Poets, novelists, musicians and artists have done it with tremendously positive results, i.e.:
  - e.e. cummings with lack of capitalization and punctuation.
  - Degas - cutting off forms with a frame (the ballet dancers).

Van Gough with the liquid stars in the starry night painting.  
The Beatles starting the song "She Loves You, Yeah, Yeah" with the chorus.

## Summary

Think about graphics as communication. You want your data's message understood. Review **TUFTE's** books. They are extremely entertaining.

## Acknowledgments:

I would like to gratefully acknowledge the encouragement (prodding) of Tamara R. Fischell, Art Carpenter, and the unwavering, (semi-wavering) support of my professors at UNC. I am deeply indebted to Susan Lynn Hogan for her extensive assistance.

## Particularly Important References

Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire, Connecticut: Graphics Press, 1983)

Edward R. Tufte, *Envisioning Information* (Cheshire, Connecticut: Graphics Press, 1990)

Gary T. Henry, *Graphing Data: Technique for Display and Analysis* (Sage Publications Inc., 1995)

William S. Cleveland, *The Elements of Graphing Data* (Monterey California: Wadsworth Advanced Books and Software, 1985)

William S. Cleveland. *Visualizing Data*. (Summit, NJ: Hobart Press, 1993)

Howard Wainer, *How to Display Data Badly*. *The American Statistician*. May, 1984 Vol 8 (2). Also see Wainer's column, *Visual Revelations*, in *Chance* magazine.

John W. Tukey. *Exploratory Data Analysis*. (Phillippines: Addison-Wesley Publishing Company, 1977)

Cecil H. Meyers. *Handbook of Basic Graphs: A Modern Approach*. (Univ. of Minn., Dickensen Publications, CA, 1970).

John M. Chambers, William S. Cleveland, Beat Kleiner, and Paul Tukey. *Graphical Methods for Data Analysis*. (Beaumont, California: Wadsworth International Group and Advanced Books and Software, 1985)

A graphics oriented issue of *Statistical Sciences*. 1990, Vol 5(3). Includes articles by John Tukey, Howard Wainer, among others.

William Playfair, *The Commercial and Political Atlas*. London, 1786.

Lori Siebert, *Working With Words and Pictures*. North Light Books, Cincinnati, Ohio, 1993.

C. F. Schmid, *Handbook of Graphic Presentation*, Wiley, New York, 1979.

LeRoy Bessler. SUGI Proceedings paper, "Communicate Effectively in Color with SAS/GRAPH Software".

Michael Friendly, *SAS System for Statistical Graphics*, First Edition. (Cary NC: SAS Institute Inc., 1991).

The SAS Institute literature has a number of good articles and books, both on 'theory' and 'how to', that are excellent references.

Forrest Young's VISTA is freely available at <http://forrest.psych.unc.edu/>

Dominic T. Moore  
dmoore@bios.unc.edu