Data Visualization

Some Introductory Thoughts

Resources

By Dr. Edward Tufte - www.edwardtufte.com

The Visual Display of Quantitative Information	ISBN 096139210X
Envisioning Information	ISBN 0961392118
Visual Explanations	ISBN 0961392126
Visual and Statistical Thinking	
Beautiful Evidence	ISBN 0961392177
The Cognitive Style of Powerpoint	ISBN 0961392150
Visual Design of the User Interface	

By Stephen Few - www.perceptualedge.com

Show Me the Numbers	ISBN 0970601999
Information Dashboard Design	ISBN 0596100167

By Dr. Cynthia Brewer – www.colorbrewer.org

A web tool for selecting effective colour combinations

By Dr. LeRoy Bessler

Get the Best out of SAS/GRAPH® and ODS – SGF 2007 Paper 228
How to Make the "Best Choice" from the Many Ways to Create and Deliver SAS® Graphs – SGF 2007 Paper 093
Communication-Effective Use of Color for Web Pages, Graphs, Tables, Maps, Text, and Print – SUGI 29 Paper 176

Other:
http://www.pickford.abelgratis.co.uk/vislit/graphicacyPaper.doc
http://encarta.msn.com/dictionary_1861615038/graphicacy.html
http://en.wikipedia.org/wiki/Graphicacy

GHSUG September 2007

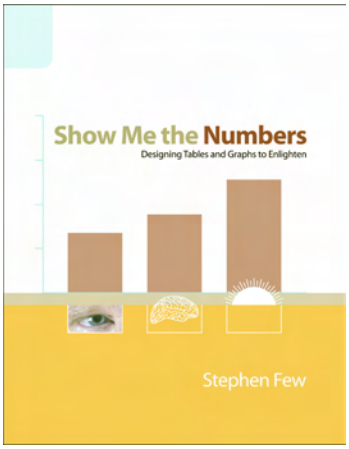
DESIGNING EFFECTIVE TABLES AND GRAPHS

Stephen Few

Most presentations of quantitative business data are poorly designed – painfully so, often to the point of misinformation. This problem, however, is rarely noticed. We use tables and graphs to communicate quantitative information: the critical numbers that measure the health, identify the opportunities, and anticipate the future of our businesses. Even the best information is useless, however, if its story is poorly told.

This problem exists because few have been trained in table and graph design for effective and efficient communication. Some key lessons for doing better include the following:

- The effective display of quantitative information involves **two fundamental challenges**: **1) selecting the right medium of display** (for example, a table or a graph, and the appropriate kind of either), and **2) designing the individual visual components of the selected medium to display the information and its message as clearly as possible**.
- **A table works best when**:
  - It is used to look up individual values
  - It is used to compare individual values
  - The values must be expressed precisely
- **A graph works best when the message is contained in the shape of the data**, such as patterns, trends, co-relationships, and exceptions to the norm.
- **Meaningful quantitative information always involves relationships**. With rare exceptions in business graphs, these relationships always boil down to one or more of the seven relationships described on the next page.
- A common problem with tables and graphs is the excessive presence of visual content that doesn't represent actual data. **Whenever quantitative information is presented, the data itself should stand out clearly, without distraction.** This involves eliminating anything that doesn't represent data, except for visual devices that support the data in a necessary way (for example, axes in a graph), in which case they should be displayed in muted fashion so as to not distract from the data itself.



Available from Analytics Press

Show Me the Numbers: Designing Tables and Graphs to Enlighten. Few, Stephen. 2004. Oakland, CA: Analytics Press.

Other Complementary References

Tufte, Edward R. 1995. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

Harris, Robert L., 1999. *Information Graphics: A Comprehensive Illustrated Reference*. New York, NY: Oxford University Press.

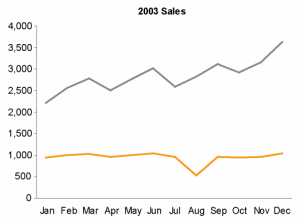
Contact: Stephen Few, Perceptual Edge, (510) 558-7400, sfew@perceptualedge.com

Seven common quantitative relationships in graphs and how to display them

Time-Series

Expresses the rise and fall of values through time.

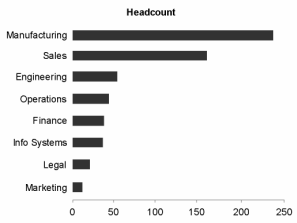
- Use lines to emphasize overall pattern.
- Use bars to emphasize individual values.
- Use points connected by lines to slightly emphasize individual values while still highlighting the overall pattern.
- Always place time on the horizontal axis.



Ranking

Expresses values in order by size.

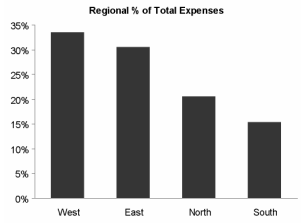
- Use bars only (horizontal or vertical).
- To highlight high values, sort in descending order.
- To highlight low values, sort in ascending order.



Part-to-Whole

Expresses the portion of each part relative to the whole.

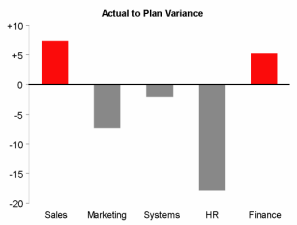
- Use bars only (horizontal or vertical).
- Use stacked bars only when you must display measures of the whole as well as the parts.



Deviation

Expresses how and the degree to which one or more things differ from another.

- Use lines to emphasize the overall pattern only when displaying deviation and time-series relationships together.
- Use points connected by lines to slightly emphasize individual data points while also highlighting the overall pattern when displaying deviation and time-series relationships together.
- Use bars to emphasize individual values, but limit to vertical bars when a time-series relationship is included.
- Always include a reference line to compare the measures of deviation against.



Distribution

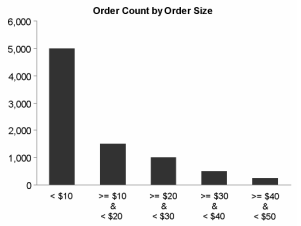
Expresses a range of values as well as the shape of the distribution across that range.

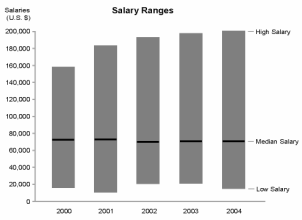
Single distribution:

- Use vertical bars to emphasize individual values
- Use lines to emphasize the overall shape.

Multiples distributions:

- Use vertical or horizontal bars (a.k.a. range bars or boxes) to encode the full range from the low value to the high value, or some meaningful portion of the range (for example, 90% of the values).
- Use points or lines together to encode measures of center (for example, the median).





Correlation

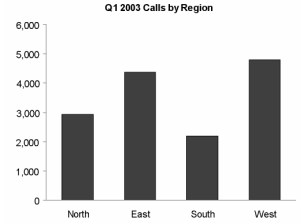
Expresses how two paired sets of values vary in relation to one another.

- Use points and a trend line in the form of a scatter plot.

Nominal Comparison

Simply expresses the comparative sizes of multiple related but discrete values in no particular order.

- Use bars only (horizontal or vertical).



Points to take home with you:

Designing Effective Graphs and Tables – from Show Me the Numbers

Principles displayed above and those espoused by Dr Edward Tufte (BS, MS, PhD, Professor Emeritus, Yale, CT) and Stephen Few (BA, MA, Lecturer at the Haas School of Business, Berkley CA)

Use the simplest style possible
Make every ink dot count – minimal use of borders, grids, backgrounds (clutter);
Let the data tell the story;
Use colour subtly – and use subtle colours;
Choose and test colours that will print AND photocopy in black and white
Make consistent use of colour and font
Consider colour deficient people when choosing colours (**red**/**green**, www.colorbrewer.org)
Use titles, footnotes and labels to clarify and define context;
Use mixed case titles, labels etc. for increased readability
Avoid acronyms and jargon
Avoid thick, heavy fonts
Will the chart stand independently of the report / presentation?
Verify wysiwyg with a hardcopy prior to distribution / presentation
A printout of your presentation slide set is NOT a report
Leave an audit trail (data source, time/space context, author, date), such that the table / chart could be reconstructed.

"A seemingly small point which deserves special mention is that one should get in the habit of making all charts, tables, reports, and problems complete in themselves. It should be possible to learn from the sheet, six months after its completion, what the problem was all about, who gathered the data and calculated it, who drew the graph, what were the important conditions at hand, what were the sample size and the units, the date, etc. Only in this way can we make sure that our records will not grow 'cold'"
I W Burr, Professor, Statistics Dept, Purdue University, Indiana from 1941 to 1974.

Tim Dietrich tim_dietrich@dofasco.ca
Bill Droogendyk bill_droogendyk@dofasco.ca

Quality Systems
Dofasco Inc

13Sep2007