Graphical Methods — Stat 6560, Section 001, Spring 2011

Instructor: Dr. Jürgen Symanzik
Office: Lund 325
Phone: 797–0696
FAX: 797–1822
e–mail: symanzik@math.usu.edu
WWW: http://www.math.usu.edu/~symanzik/
http://www.math.usu.edu/~symanzik/teaching/2011_stat6560/stat6560.html

Office Hours: MWF 11:00am – 12:00noon and by appointment.

Classes & Rooms:
Section 001: MWF 8:30am – 9:20am, Mo 1/10 – Fr 4/29, 2011: Geol 310.

Please note that there are no classes on Mo 1/17 (Martin Luther King Day) and on Mo 2/21 (Presidents' Day). However, Monday classes will be held on Tu 2/22 (i.e., no regular Tu classes will be held during that week). There are also no classes during Spring Break (Mo 3/7 – Fr 3/11).

Please visit the course Web page listed above frequently for lecture notes, data sets, graphical examples, R code, etc. — in particular if you miss class for any reason.

Course Objectives:

Statistical graphics and data visualization are critical elements of modern data analysis and presentation. From initial exploration of a data set to the final presentation of results to the end user, statistical graphics play a vital role in shaping our understanding of our data. Through proper use of graphics, we can make critical discoveries, and communicate them clearly. Conversely, poor use or misuse of graphics can seriously mislead (by accident or design).

In this course, we will start with presentation graphics, including discussion of both tools and principles which lead to clear communication and those which serve only to confuse or mislead. We will spend most of the semester in exploratory graphics and data analysis, including data mining. This will be broken down largely by the dimension of the applicable data. One– and two–dimensional datasets require and allow far different methods than those of more than three dimensions. Categorical and regression data call for their own specialized methods.

Even more than most aspects of statistics, graphics and visualization involve art as well as science. In most cases, there are many reasonable approaches. Only an understanding of the options available and the underlying priciples will lead to a successful analysis and presentation.

Prerequisites:

I expect basic knowledge of R and LaTeX, as taught in the USU Statistical Computing course (or from any other graduate course or from independent studies). Please refer to the USU Statistical Computing course from Fall 2010 for an introduction of R and LaTeX(http://www.math.usu.edu/~corcoran/classes/10fall6550/main.htm).

Topics: (subject to change)

1. Introduction: A Couple of Good and Bad Examples.
2. History of Statistical Graphics: Plots, People, and Events.
3. Use of Color.
4. Categorical Plots.
5. Univariate Plots.
6. Bivariate Plots.
7. Trivariate Plots.
8. "Hypervariate" (High–Dimensional) Plots.
9. Statistical Maps.
10. Interactive and Dynamic Graphics.
11. Graphics Galleries and Sources on the Web.

Assignments:

There will be a variety of assignments throughout the semester. Each assignment will include a value (typically 20–100 points) that it will be scored out of. Your final grade will be determined by the sum of your points in all assignments. Some assignments will include combinations of analysis of existing graphics, creation of your own, computer work in R, GGobi, or others, and short oral presentations. The value of each assignment will be roughly proportional to its importance and the amount of work involved.

Regular homework assignments will be done individually or in groups of 2 or 3 students. Major assignments such as seminar–like presentations or a major graphical project will be done individually or also in groups of 2 or 3 students. One major data source will be related to the Deepwater Horizon oil spill in 2010 (see `http://streaming.stat.iastate.edu/dataexpo/2011/` for more details). Full details will be provided once we know our final class size (as of 1/9/2011, there were 9 students registered so far).

There will be no regular (in–class or take–home) midterm or final exams.

Textbooks:

Carr, Daniel B., and Pickle, Linda W. (2010) *Visualizing Data Patterns with Micromaps*, Boca Raton, Florida: Chapman & Hall/CRC Press.

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

Cook, Dianne, and Swayne, Deborah F. (2007) *Interactive and Dynamic Graphics for Data Analysis — With R and GGobi*, New York, New York: Springer.

Theus, Martin, and Urbanek, Simon (2009) *Interactive Graphics for Data Analysis — Principles and Examples*, Boca Raton, Florida: Chapman & Hall/CRC Press.

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

Every student should have easy access to each of these four books, but it is not necessary that every student buys all of these books. Perhaps you can make arrangements with some of the other students in class (or your group members) who purchases which book(s).

Software:

We will primarily be using R (`http://cran.r-project.org/`), a GNU–license statistical package and clone of S–Plus. Please install the current version of R, i.e., 2.12.1, on your own computer so we can exchange code. We will also be using GGobi (`http://ggobi.org/`) for high–dimensional analysis and Mondrian (`http://rosuda.org/mondrian/`) for categorical data analysis. All tools are available on the Web for free download from the URLs listed above.

Credits:

This course uses some of the course materials provided by Dr. Mike Minnotte (formerly USU, now with the University of North Dakota) as held in the Fall 2006 semester. Additional material have been taken from other Statistical Graphics courses, such as the ones offered by Dr. Di Cook (Iowa State University: `http://www.public.iastate.edu/~dicook/`) and Dr. Dan Carr (George Mason University: `http://mason.gmu.edu/~dcarr/`). We are likely to include parts from additional Web sources that will be specified later on.

Courtesy:

Please turn off pagers and cell phones before class, and please keep conversations to a minimum during lectures. Please do not read/reply to your e–mails or browse other Web pages than the ones discussed during class.

I will not keep track if you come to class or not. However, I would highly recommend to attend all lectures.

Americans with Disabilities Act:

If a student has a disability that will likely require some accommodation by the instructor, the student must contact the instructor and document the disability through the Disability Resource Center, during the first week of the course. Any requests for special considerations relating to attendance, pedagogy, taking of examination, etc. must be discussed with and approved by the instructor. In cooperation with the Disability Resource Center, course materials can be provided in alternative formats — large print, audio, diskette or Braille.

Note:

The above schedule and procedures in this course are subject to change in the event of extenuating circumstances.

Additional References:

Beyond the required textbooks, material will be drawn from a number of sources. Some additional useful references are:

Chen, C., Härdle, W., and Unwin, A., eds., (2008), *Handbook of Data Visualization*, Berlin/Heidelberg, Springer–Verlag.

Cleveland, William S., (1994), *The Elements of Graphing Data*, Summit, NJ, Hobart Press.

Cleveland, William S. and McGill, Marylyn E., eds., (1988), *Dynamic Graphics for Statistics*, Belmont, CA, Wadsworth & Brooks/Cole.

Cook, R. Dennis and Weisberg, Sanford (1994), *An Introduction to Regression Graphics*, New York, Wiley.

Deboeck, Guido and Kohonen, Teuvo, eds., (1998), *Visual Explorations in Finance with Self-Organizing Maps*, New York, Springer.

du Toit, S. H. C., Steyn, A. G. W., and Stumpf, R. H., (1986), *Graphical Exploratory Data Analysis*, New York, Springer-Verlag.

Gentle, J. E., Härdle, W., and Mori, Y., eds., (2004), *Handbook of Computational Statistics — Concepts and Methods*, Berlin/Heidelberg, Springer–Verlag.

Harris, Robert L., (1999), *Information Graphics*, New York, Oxford University Press.

Henry, Gary T., (1995), *Graphing Data: Techniques for Display and Analysis*, Thousand Oaks, CA, SAGE Publications.

Murrell, Paul, (2005), *R Graphics*, London, Chapman & Hall.

Playfair, William, (2005), *The Commercial and Political Atlas and Statistical Breviary*, (ed. Howard Wainer and Ian Spence), New York, Cambridge University Press. (First published in 1786–1801!)

Robbins, Naomi B., (2005), *Creating More Effective Graphs*, Hobken, NJ, Wiley.

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

Tufte, Edward R., (1997), *Visual Explanations: Images and Quantities, Evidence and Narritive*, Cheshire, CT, Graphics Press.

Tukey, John W., (1977), *Exploratory Data Analysis*, Reading, MA, Addison–Wesley.

Unwin, A., Theus, M., and Hofmann, H., (2006), *Graphics of Large Datasets: Visualizing a Million*, New York, Springer.

Wainer, Howard, (1997), *Visual Revelations: Graphical Tales of Fate and Deception from Napoleon Bonaparte to Ross Perot*, New York, Springer–Verlag.

Wainer, Howard, (2005), *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton, NJ, Princeton University Press.

Wallgren, A., Wallgren, B., Persson, R., Jorner, U., and Haaland, J., (1996), *Graphing Statistics and Data: Creating Better Charts*, Newbury Park, CA, SAGE Publications.

Westphal, Christopher and Blaxton, Teresa, (1998), *Data Mining Solutions: Methods and Tools for Solving Real–World Problems*, New York, Wiley.

Wilkinson, Leland, (1999), *The Grammar of Graphics*, New York, Springer.