Statistical Visualization I — Stat 5810, Section 003, Fall 2018

Instructor: Dr. Jürgen Symanzik
Office: AnSc 313
Phone: 797–0696
FAX: 797–1822
e–mail: symanzik@math.usu.edu
Web: `http://www.math.usu.edu/~symanzik/`
`http://www.math.usu.edu/~symanzik/teaching/2018_stat5810_003_fall/stat5810_003.html`

Office Hours: Tuesday (T) & Thursday (H) 1:30pm – 2:30pm and by appointment.

Classes & Rooms:
TH 9:00am – 10:15am, T 9/11 – H 11/29, 2018 (tentatively): AnSc 320.
Please visit the course Web page listed above and/or Canvas frequently for lecture notes, data sets, graphical examples, R code, etc. — in particular if you miss class for any reason.
Detailed Class Schedule:
For a 2–credit course, we need 20 lectures/lecture days (in contrast to 29 or 30 lectures/lecture days for a 3–credit course). Those days are marked as "Lecture 01" to "Lecture 20" in the overview below:

Week	Tuesday	Thursday
1	8/28 No class	8/30: No class
2	9/4: No class	9/6: No class
3	9/11: Lecture 01	9/13: Lecture 02
4	9/18: Lecture 03	9/20: Lecture 04
5	9/25: Lecture 05	9/27: Lecture 06
6	10/2: Lecture 07	10/4: Lecture 08
7	10/9: Lecture 09	10/11: Lecture 10
8	10/16: Lecture 11	10/18: Lecture 12
9	10/23: Lecture 13	10/25: Lecture 14
10	10/30: Lecture 15	11/1: Lecture 16
11	11/6: No class	11/8: No class
12	11/13: Lecture 17	11/15: Lecture 18
13	11/20: No class	11/22: No class
14	11/27: Lecture 19	11/29: Lecture 20
15	12/4: Backup	12/6: Backup

Note: "No class" means guaranteed no class that day. I have marked the last week as "Backup", e.g., in case we miss lectures because I am sick or have to travel. But hopefully, this won't happen. If nothing goes wrong, our tentative last lecture date will be on H 11/29/2018.

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).

The course will address three main questions:

1. Why statistical graphics (and which ones to draw)?
2. How to construct statistical graphics in R?
3. How to distinguish between **good** and **bad** statistical graphics?

This course is **not** an introduction into a single R graphics package. Rather, a variety of R graphics packages will be used, such as baseR, `ggplot2`, `lattice`, etc.

The course will be broken down largely by the dimension of the available data, starting with categorical data. One– and two–dimensional quantitative data sets require and allow far different methods than those of more than three dimensions. Towards the end of this course, we will deal with presentation graphics, including a discussion of tools and principles that lead to a clear communication and those that serve only to confuse or mislead.

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 principles will lead to a successful analysis and presentation.

Prerequisites:

I expect basic knowledge of R as taught in the "Introduction to R" course. Moreover, you should be familiar with a tool such as R Markdown, knitr, or sweave that allows you to combine text, R code, graphics, and numerical results in high-quality documents. LaTeX is a plus but is not formally required.

IDEA Center Learning Objectives:

Objective 1) Gaining factual knowledge (terminology, classifications, methods, trends).

Objective 2) Learning fundamental principles, generalizations, or theories.

Objective 3) Learning to apply course material (to improve thinking, problem solving, and decisions).

Topics: (subject to change)

1. Introduction.
2. Basic Graph Construction and Refinement.
3. Graphs for Categorical Data.
4. Graphs for Univariate Data.
5. Graphs for Bivariate Data.
6. Good and Bad Graphs.

If time permits, we will start with the discussion of some of the topics from the follow–up course "Statistical Visualization II", offered in the spring semester, such as graphs for trivariate data, history of graphics, color and cognition, statistical maps, graphs for "hypervariate" (high–dimensional) data, interactive and dynamic graphics, and web–based graphics.

Homework 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 (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. For individual assignments, you will be allowed to discuss general approaches to questions on the assignments with other students, but each student must write and submit their own code and comments. Any students caught sharing code will fail the class.

Unless otherwise stated on the assignment sheet, all homework assignments have to be submitted electronically via Canvas.

The following deductions will be applied to late homework submissions: 1 min – 24 hours late: 10% off; > 24 hours – 48 hours late: 25% off; > 48 hours – 72 hours late: 50% off. Homeworks won't be accepted later than 72 hours (i.e., 3 days) after the submission deadline.

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

Textbooks:

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

Unwin, Antony (2015) *Graphical Data Analysis with R*, Boca Raton, FL: CRC Press/Taylor & Francis.

Wickham, Hadley (2009) *ggplot2 — Elegant Graphics for Data Analysis*, New York, NY: Springer.

Every student should have access to each of these 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 who purchases which book(s). If you plan to work in the area of statistical visualization for your MS or PhD degree, you should consider to purchase these books for an ongoing use beyond this course.

Software:

We will primarily be using R (`http://cran.r-project.org/`), a free software environment for statistical computing and graphics. Please install the current version of R, i.e., 3.5.1, on your own computer so we can exchange code. Also install RStudio (`https://www.rstudio.com/`) as a front end to R.

Courtesy:

Please turn off cell phones and similar devices 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. If you have to miss a lecture, there will be a recording of the lecture available in Canvas Panopto (if the technology doesn't fail).

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 (DRC), 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, or Braille.

Note:

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