Statistical Visualization I — Stat 5550, Section 001

Fall 2021 (2 Credits)

Instructor: Dr. Jürgen Symanzik
Office: AnSc 313
Phone: 435–797–0696
e–mail: symanzik@math.usu.edu
Web: `http://www.math.usu.edu/~symanzik/`
`http://www.math.usu.edu/~symanzik/teaching/2021_stat5550/stat5550.html`

Office Hours: Tuesday (T) & Thursday (H) 1:30pm – 2:20pm in AnSc 313 (please wear a face mask), Friday (F) TBA (via Zoom), and by appointment.

Classes & Rooms:
TH 9:00am – 10:15am, T 9/14 – H 11/18, 2021 (tentatively): ENGR 302 (face–to–face).

Please visit the course Web page listed above for emergency announcements, e.g., when Canvas is unavailable. Otherwise, visit Canvas frequently for lecture notes, data sets, R code, etc. — in particular if you miss our face–to–face lectures for any reason. All (additional and updated) materials, announcements, discussions, recordings, etc. from Canvas are part of the course materials. Not seeing one of these in time does not serve as an excuse for not getting point deductions for the course. Deadlines may change or *Coronavirus/Covid–19* regulations and requirements may be updated. It is your responsibility to make sure to receive all announcements in time.

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	8/31 No class	9/2: No class
2	9/7: No class	9/9: No class
3	9/14: Lecture 01	9/16: Lecture 02
4	9/21: Lecture 03	9/23: Lecture 04
5	9/28: Lecture 05	9/30: Lecture 06
6	10/5: Lecture 07	10/7: Lecture 08
7	10/12: Lecture 09	10/14: Lecture 10
8	10/19: Lecture 11	10/21: Lecture 12
9	10/26: Lecture 13	10/28: Lecture 14
10	11/2: Lecture 15	11/4: Lecture 16
11	11/9: Lecture 17	11/11: Lecture 18
12	11/16: Lecture 19	11/18: Lecture 20
13	11/23: Backup	11/25: No class
14	11/30: Backup	12/2: Backup
15	12/7: Backup	12/9: Backup

Note: "No class" means guaranteed no class that day. I have marked the dates in the three

last weeks of the semester as "Backup", e.g., in case we miss lectures because of Covid–19 restrictions, sickness, or any other reason. If nothing goes wrong, our tentative last lecture date will be on H 11/18/2021. Backup lectures (if any) will be held via Zoom (and not face–to–face). There is no need to return to Logan after Thanksgiving for a backup lecture in this course.

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:

STAT 5050 ("Introduction to R") with a C− or better. 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. Good and Bad Graphs.

The following topics will be discussed in the follow–up course "Statistical Visualization II", typically offered in the spring semester: graphs for bivariate and trivariate data, graphs for "hypervariate" (high–dimensional) data, color and cognition, statistical maps, interactive and dynamic graphics, web–based graphics, and history of graphics.

Homework Assignments:

There will be 3 HW assignments for this course, roughly one every three weeks. Each HW assignment will include a value (typically 20–100 points) that it will be scored out of. HW assignments will contribute to 100% of your course grade. The value of each HW assignment will be roughly proportional to its importance and the amount of work involved.

You will be allowed to discuss general approaches to questions on the HW assignments with other students, but each student must write and submit their own R code and comments. Any students caught sharing R code or other parts of their homework submissions will fail the class.

Unless otherwise stated on the HW assignment sheet, all homework assignments have to be submitted electronically via Canvas. **You will have 2 or 3 weeks after the last lecture to finalize and submit the last HW assignment.**

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. We will have a few worksheets for training purposes only.

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 a recent version of R, i.e., 4.0.4 or later, on your own computer so we can exchange code. Also install RStudio (`https://www.rstudio.com/`) as a front end to R and a version of LaTeX that can be used in

connection with RStudio. My personal recommendation is TeX Live (`https://www.tug.org/texlive/`).

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 Kaltura recording of the lecture available in Canvas (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 – `https://www.usu.edu/drc/`), preferably 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.

Covid–19 Requirements and Restrictions:

1. Although not formally required, please wear a face mask in all university buildings, in particular in all classrooms and all private offices. I will do the same.
2. Please continue to maintain a 6–foot distance to other students, instructors, and staff. Leave at least one (or more if possible) empty seat to the next student in the classroom. Fill in the seating chart in Canvas in the week of 9/20/2021.
3. If you aren't vaccinated against Covid–19 yet, please get vaccinated and upload your proof of vaccination to `https://aggiehealth.usu.edu`. As you may be aware, USU is finalizing plans to require students to be vaccinated for the Spring 2022 semester. According to a recent release on 9/10/2021 from the Centers for Disease Control and Prevention (CDC), fully vaccinated people have a reduced risk of infection ($5\times$), hospitalization ($> 10\times$), and death ($> 10\times$) from Covid–19 (including the Delta variant), compared to unvaccinated people (see `https://www.cdc.gov/mmwr/volumes/70/wr/mm7037e1.htm`). So, get vaccinated — the sooner the better (if you aren't vaccinated yet).
4. Stay home if you feel sick or suffering even mild symptoms of illness that could be related to Covid–19 and to get a free Covid–19 test. All lectures will be recorded via Kaltura so you can watch them in Canvas afterwards.
5. Make sure that you regularly check the guidelines for risk reduction and case management as they are updated on the main USU Covid–19 website (`https://www.usu.edu/covid-19`), and specifically the "Communications to Student" section (`https://www.usu.edu/covid-19/communications/students`).
6. Check for announcements in Canvas (and/or via e–mail), ideally each time before you plan to leave for one of our face–to–face lectures.

Note:

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