

Statistical Visualization I — Stat 5550, Section MW1

Fall 2020 (2 Credits)

Instructor: Dr. Jürgen Symanzik
Office: Virtual
Phone: 435-797-0696
e-mail: symanzik@math.usu.edu
Web: <http://www.math.usu.edu/~symanzik/>
http://www.math.usu.edu/~symanzik/teaching/2020_stat5550/stat5550.html

Office Hours: Monday (M), 9:00am – 10:00am; Wednesday (W) & Friday (F), 8:00am – 9:00am; and by appointment. Office hours have to be scheduled via the Zoom menu in Canvas (or via e-mail) and will be held virtually via Zoom.

Classes & Rooms:

Tuesday (T) & Thursday (R) 9:00am – 10:15am, T 9/15 – R 11/19, 2020: Virtual Meetings via Zoom.

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 lecture periods 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	9/1 No class	9/3: No class
2	9/8: No class	9/10: No class
3	9/15: Lecture 01	9/17: Lecture 02
4	9/22: Lecture 03	9/24: Lecture 04
5	9/29: Lecture 05	10/1: Lecture 06
6	10/6: Lecture 07	10/8: Lecture 08
7	10/13: Lecture 09	10/15: Lecture 10
8	10/20: Lecture 11	10/22: Lecture 12
9	10/27: Lecture 13	10/29: Lecture 14
10	11/3: Lecture 15	11/5: Lecture 16
11	11/10: Lecture 17	11/12: Lecture 18
12	11/17: Lecture 19	11/19: Lecture 20
13	11/24: Backup	11/26: No class
14	12/1: Backup	12/3: No class
15	12/8: No class	12/10: No class

Note: “No class” means guaranteed no class that day. I have marked a few days as “Backup”, e.g., in case we miss lectures because of network problems, power failures, etc. on my side. But, hopefully, this won’t happen. If nothing goes wrong, our tentative last lecture date will be on R 11/19/20.

If a Zoom meeting does not start on time, please wait for 10min so I can try to get it started differently. Similarly, if we get disconnected during a Zoom meeting, also wait for 10min and watch for announcements in Canvas and/or via e-mail. In case you lose the connection on your side, please try to reconnect as quickly as possible.

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. L^AT_EX 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”, 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.

Course Format and Lecture Attendance Points:

The course will be offered in a blended web broadcast format. See <https://www.usu.edu/ais/scheduling/deliverymethods> for requirements on your side.

Under this setup, we will basically use a flipped classroom approach. You will have to watch recordings of past classroom-based lectures for this course by yourself and you have to work through the lecture notes and R code also by yourself. The scheduled lecture periods will be used to discuss your questions related to the lecture recordings and other course materials and for help with the next homework assignment. I also plan to summarize the most important parts of each lecture at the start of each lecture period. Lecture periods will differ in lengths, depending on your questions. Some may be as short as 10min, while others may take up the entire 75min.

You will be awarded up to three lecture attendance points (LAPs) for each lecture, i.e., up to 60 LAPs in total. You will be asked early, in the middle, and towards the end of each lecture to type a short confirmation into the Chat box in Zoom so that I can see who was present at that time. LAPs will contribute to 10% of your course grade. You will obtain 10 points for 90% or more (54 to 60) of all possible LAPs, 9 points for 80% up to 90% (48 to 53) of all possible LAPs, 8 points for 70% up to 80% (42 to 47) of all possible LAPs, 7 points for 60% up to 70% (36 to 41) of all possible LAPs, 6 points for 50% up to 60% (30 to 35) of all possible LAPs, and 0 points for less than 50% (0 to 29) of all possible LAPs.

The lecture periods will be recorded. If you really cannot participate at a lecture, at least watch the recording before the next lecture. In case of an excused absence, e.g., for medical reasons, family emergencies or funerals, court appointments, university-approved travel, etc., please provide some supporting information and your LAP score will be adjusted according to the number of lecture periods you could attend. Private reasons such as travel, most family events (such as weddings), etc. do not count as an excused absence.

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 90% of your course grade. The value of each HW assignment will be roughly proportional to its importance and the amount of work involved. Your final course grade will be determined as the weighted average of your LAPs and the sum of your points in all HW assignments.

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. The USU library holds several of these books or provides online access. 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 most recent version of R, i.e., 4.0.2, on your own computer so we can exchange code. Also install RStudio (<https://www.rstudio.com/>) as a front end to R and MiKTeX (<https://miktex.org/>) that will allow us to combine code and results from R into text documents.

Courtesy:

One of the intents of the blended web broadcast format is to make our courses somewhat more personal again, compared to the pure online format during the second half of the Spring 2020 semester. For this reason, please activate your webcams during all of our Zoom meetings so we each can see each other. However, to avoid interference with the audio, please mute your microphone unless you want to speak yourself. Let me know that you want to speak by indicating so in the Chat box in Zoom or by unmuting your microphone, but wait until the previous speaker has ended. When you are done speaking, please mute your microphone again.

Please be aware of the *Code of Policies and Procedures for Students at Utah State University* (<https://studentconduct.usu.edu/studentcode/>) and follow the code accordingly. Also be aware of the USU *Coronavirus/Covid-19* regulations and requirements

(<https://www.usu.edu/covid-19/>). The latter ones may change with little advance notification. Useful resources for students can be found at <https://www.usu.edu/ready/>. Also, as a reminder, all of our lecture periods, including audio, video, and chats, will be recorded and might be used as evidence in case of any student code violations during these lecture periods, in particular those listed in Article 5 of the student code (<https://studentconduct.usu.edu/studentcode/article5>).

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.

Note:

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