

Introduction to Data Visualization

SURV 752

1 credit/2 ECTS

Spring 2020, sec. 1

Instructor

Richard Traunmüller, rtraunmu@mail.uni-mannheim.de

Video lecture by Richard Traunmüller

Short Course Description

Data visualization is one of the most powerful tools to explore, understand and communicate patterns in quantitative information. At the same time, good data visualization is a surprisingly difficult task and demands three quite different skills: substantive knowledge, statistical skill, and artistic sense. The course is intended to introduce participants to key principles of analytic design and useful visualization techniques for the exploration and presentation of univariate and multivariate data. This course is highly applied in nature and emphasizes the practical aspects of data visualization in the social sciences. Students will learn how to evaluate data visualizations based on principles of analytic design, how to construct compelling visualizations using the free statistics software R, and how to explore and present their data with visual methods.

Course and Learning Objectives

By the end of the course, students will...

- know how to evaluate and criticize data visualizations based on principles of analytic design
- be in the position to explore and present their data with visual methods
- understand which graphical formats are useful for which types of data and questions
- know how to construct compelling visualizations using the free statistics software R

Prerequisites

Course prerequisites are a **basic understanding of statistics and bivariate linear regression**. Some experience in the use of a statistical software package would help, preferable basic data management tasks in R (loading and merging data, generating

and recoding variables, etc.) I will provide detailed code examples to get students up to pace.

Class Structure and Course Concept

This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video recorded lectures and reading the required literature for each unit and then “attending” mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor. Just like in an on-site course, homework will be assigned and graded and there will be a final exam at the end of the course.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), in-class-activities (watching videos, participating in online meetings), and follow-up activities (working on assignments and exams) – as in an on-site course. As a rule of thumb you can expect to spend approximately 3h/week on in-class-activities and 9 hours per week on out-of-class activities (preparing for class, readings, assignments, projects, studying for quizzes and exams). Therefore, the workload in all courses will be approximately 12h/week. This is a 1-credit course that runs for 4 weeks. Please note that the actual workload will depend on your personal

Mandatory Weekly Online Meetings

Tuesday, 12:00 PM EDT/6:00 PM CEST, starting April 7, 2020

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on <https://www.elms.umd.edu/>. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students are required to post questions or comments about the materials covered in the videos or readings of the week in the forum before the meetings (deadline for posting questions is Monday, 12:00 PM EDT/6:00 PM CEST).

Students have the opportunity to use the Zoom meeting room set up for this course to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are encouraged to post the times that they will be using the room to the course website forum to avoid scheduling conflicts. Students are not required to use Zoom and can of course use other online meeting platforms such as Google Hangout or Skype.

Grading

Grading will be based on:

- Participation at online meetings (10%)
- Required Course Assignments (40%)
- Final Visualization Project (50%)

Dates of when assignment will be due are indicated in the syllabus. Extensions will be granted sparingly and are at the instructor's discretion.

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between students and the instructor during the weekly online meetings. Therefore we encourage all students in this course to use a web camera and a headset. Decent quality headsets and web cams are available for less than \$20 each. We ask students to refrain from using built-in web cams and speakers on their desktops or laptops. We know from our experience in previous online courses that this will reduce the quality of video and audio transmission and therefore will decrease the overall learning experience for all students in the course. In addition, we suggest that students use a wire connection (LAN), if available, when connecting to the online meetings. Wireless connections (WLAN) are usually less stable and might be dropped.

Long Course Description

This course is intended to provide students with a thorough introduction to the best practice of modern data visualization from a social science perspective.

The course is highly applied in nature and emphasizes the practical aspects of data visualization in the social sciences. To illustrate the concepts and methods, examples and data from the social sciences will be used throughout the course. Next to conceptual discussions, the course will spend some time on how to produce data visualizations using the free statistical programming language R. Course participants will get hands-on advice on producing modern visualizations for their practical problems.

After an introduction to data visualization showing new and classic examples, the course will discuss data visualization as a methodology for social science data analysis and exploration rather than simply turning data into visual objects. Data visualization is about solving problems with data, where visualization is the means to an overarching goal. The course distinguishes between high-level goals

(exploration vs. presentation) and low-level goals (making specific comparisons and revealing specific patterns). Understanding data visualization as a methodology also implies that, instead of focusing on single graphics and formats, one should think about how they are used in the larger context of a data analysis. This immediately leads to considerations of how to use and combine multiple graphs either of different subsets of the data or different formats of the same subset.

Next, the course will introduce the basic fundamentals of graphical perception – how humans see and process visual stimuli. We will take a closer look at how to best achieve the low-level goals of making specific comparisons and finding specific patterns in the data. Certain graphical formats are generally superior to others. Understanding the workings of graphical perception suggests specific design principles that improve the detectability of patterns in the data and decrease the cognitive load in processing them. We will apply this knowledge in a discussion of the relative merits of familiar formats such as bar, dot, and line charts.

Comparisons are the heart of any analysis of quantitative data. The course will give an overview of the graphical formats and visual techniques that optimally support some of the most fundamental data analytic tasks: comparing before and after, comparing subgroups, comparing to a standard, comparing to a larger context, etc. Students will get to know the slope graph, spark lines and the bullet graph as less well known but highly effective formats for making visual comparisons. Importantly, this session will present one of the most powerful methods of data visualization: the small multiple design. We will stress the importance of arrangement, sorting, and visual reference elements in enabling effective comparisons.

Social science data analysis is fundamentally about relations between variables. This includes how a variable changes over time as a special case. The course discusses scatter plot variants for the effective display of bivariate relationships. Considerable time will be spent on the problem of over-plotting, how to deal with it (e.g., through the use of jittering or alpha blending) and how to enhance scatterplots with additional plot elements. One important enhancement that will receive a detailed treatment is the addition of both parametric and non-parametric scatter-plot smoothers that reveal general trends in data. We will close with a discussion of how the aspect-ratio of a graphic affects the perceived strength of a correlation or time trend.

Readings

Primary readings will be from the following volume:

Few, Stephen (2012). *Show Me the Numbers. Designing Tables and Graphs to Enlighten.* (Second Edition). Analytics Press.

Additional required and recommended readings will be made available on the course website: <https://www.elms.umd.edu/>.

Academic Conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at

<https://www.president.umd.edu/sites/president.umd.edu/files/documents/policies/III-100A.pdf> (University of Maryland) and

<https://www.uni-mannheim.de/en/research/good-research-practice/> (University of Mannheim).

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work or taking any exams in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously and may result in oral or written reprimands, a lower or failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Accessibility & Disability Service (ADS) office to register in person for services. Please call the office to set up an appointment to register with an ADS counselor. Contact the ADS office at 301.314.7682; <https://www.counseling.umd.edu/ads/>.

Students at the University of Mannheim should contact the Commissioner and Counsellor for Disabled Students and Students with Chronic Illnesses at http://www.uni-mannheim.de/studienbueros/english/counselling/disabled_persons_and_persons_with_chronic_illnesses/.

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course (in addition to the standard university evaluation survey). Participation is entirely voluntary and highly appreciated.

Class Schedule

Please note that assignments and dates are subject to change. Information (e.g., articles and assignments) posted to the course website supersedes the information noted here.

Unit 1: Data Visualization as a Methodology

Video lecture: available Tuesday, March 31

Online meeting: Tuesday, April 7, 12:00 PM EDT/6:00 PM CEST

- Exploring and Describing the Datasaurus Dozen Dataset using R

Assignment/Quiz 1 due: Wednesday, April 8, 12:00 PM EDT/6:00 PM CEST

Required Readings:

- Few, S. (2012). Introduction. (Ch. 1: p. 1-13)
- Few, S. (2012). Differing Roles of Tables and Graphs. (Ch. 3: p. 39-51)

Interested students might find the following additional recommended texts helpful in preparing for the course:

- Tufte, E. (2006). Principles of Analytic Design. In: Tufte, E.: *Beautiful Evidence*. Graphics Press. (Ch. 5).
- Gelman, A & Unwin, A. (2013). Infovis and Statistical Graphs: Different Goals, Different Looks (with Discussion). *Journal of Computational and Graphical Statistics* 22: 2-28

Unit 2: Principles of Analytic Design

Video lecture: available Tuesday, April 7

Online meeting: Tuesday, April 14, 12:00 PM EDT/6:00 PM CEST

- Comparing simple plot formats using polling data using R
- OR
- Graphical redesign of line graphs using R

Assignment/Quiz 2 due: Wednesday, April 15, 12:00 PM EDT/6:00 PM CEST

Required Readings:

- Few, S. (2012). Visual Perception and Graphical Communication. (Ch 5: 61-86)
- Few, S. (2012). Fundamental Variations of Graphs. (Ch 6: 87-135)
- Few, S. (2012). General Design for Communication. (Ch 7: 141-154)

Interested students might find the following additional recommended texts helpful in preparing for the course:

- Few, S. (2012). Silly Graphs that are Best Forsaken. (Ch 12: 271-285)
- Heer, J. & Bostock, M. (2010). Crowdsourcing Human Perception: Using Mechanical Turk to Assess Visualization Design. *ACM Human Factors in Computing Systems (CHI)* 2010.

Unit 3: Making Visual Comparisons

Video lecture: available Tuesday, April 14

Online meeting: Tuesday, April 21, 12:00 PM EDT/6:00 PM CEST

- Exploring and making visual comparisons using the football data set and R

Assignment/Quiz 3 due: Wednesday, April 22, 12:00 PM EDT/6:00 PM CEST

Required Readings:

- Few, S. (2012). General Graph Design. (Ch. 9: 191-203)
- Few, S. (2012). Component Level Graph Design. (Ch. 10: 205-255)
- Few, S. (2012). Displaying Many Variables at Once. (Ch. 11: 257-270)

Interested students might find the following additional recommended texts helpful in preparing for the course:

- Tufte, E. (2001). Data Density and Small Multiples. In: Tufte, E.: *The Visual Display of Quantitative Information*. Graphics Press. Ch. 8.

Unit 4: Visualizing Relations Between Variables

Video lecture: available Tuesday, April 21

Online meeting: Tuesday, April 28, 12:00 PM EDT/6:00 PM CEST

- Visualizing the Gapminder data set using R

Assignment/Quiz 4 due: Wednesday, April 29, 12:00 PM EDT/6:00 PM CEST

Required Readings:

- Few, S. (2009). Correlation Analysis. In: Few, S.: *Now You See It*. Analytics Press. (Ch. 11: 245-279).

Interested students might find the following additional recommended books helpful in preparing for the course:

- Jacoby, W. G. (2000). Loess: a nonparametric tool for depicting relationships between variables. *Electoral Studies* 19: 577-613.

Final Visualization Project

Due: Tuesday, May 5, 12:00 PM EDT/6:00 PM CEST

Note: Student access to the course website will be revoked four weeks after the final exam.