

Quantitative Methods in International Relations

David Weyrauch

Spring 2020

E-mail: dweyrauc@mail.uni-mannheim.de

Office Hours: Thursdays, 14:00-15:00

Office: A 351, A5 6

Web: TBD

Class Hours: Mondays, 12:00-13:30

Class Room: C -108, A5 6

Course Description

This course will teach students how to address social science questions in the fields of international relations and European integration by analyzing quantitative data in R. The course will introduce students to R, a free and versatile software environment for statistical computing and graphics. Students will learn about data management, basic principles for statistical inference, techniques for dealing with continuous, binary and count variables and data visualization.

Course Objectives

After this course, students will be able to preprocess the data required to answer their own research questions, identify the correct statistical model for different types of data and research questions and to justify their decision. They are able to correctly specify and implement such models in R and to describe and visualize the results as well as to evaluate and interpret them in light of the research question. The goal is to provide students with the foundation to perform their own analyses (e.g. for their BA theses) by transferring the acquired skills to their research interests and to become critical consumers of statistical claims made in the news, policy reports and academic research.

Organizational Information and Course Policies

General: The course is taught in English. Each session is divided into a lecture and a lab part. The goal of the lecture part is to walk you through the material necessary for understanding what we will apply hands-on in the lab part. You will be awarded 6 ECTS credits for the successful completion of this course.

Course readings: We do not use a single textbook for this course. Selected readings for each session will be made available on the ILIAS course page. We will mainly work with the following books:

Healy, K. (2019). *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press.

Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press.

Monogan, J. E. (2015). *Political Analysis Using R*. Cham: Springer.

Wickham, H. and G. Grolemund (2017). *R for Data Science*. Sebastopol: O'Reilly.

Attendance and participation: The learning process of this class is based on in-class participation, practice and discussion of the course contents and problems. Therefore, regular attendance and preparation of the course materials is highly recommended. This includes coming to class on time. We however do not formally check your attendance.

Laptop and phone policy: Since class takes place in the computer lab with all required software being installed, you do not need to bring your own laptop. If you wish to work with your own laptop, feel free to bring it to class. This has the advantage that you can set up your own computer, e.g. for the homeworks and the take-home exam. We will explain and try to help with installing the required software throughout the course.

Please avoid distracting yourself and others through electronic devices.

Datacamp: As part of this course, you will get a six-months free premium access to [DataCamp](#) which we will use as an online learning platform. You will mainly use it to get started with the basics in R. However, it is also a great opportunity for practicing some of the material we cover in this course and for checking out some of DataCamp's other courses. We will recommend suitable DataCamp courses for some of the sessions throughout the semester.

Grading policy: The examination regulations state that course assessment consists of two components: coursework (“Studienleistungen”) and examination (“Prüfungsleistungen”). Coursework is graded as pass/fail and is defined as preliminary required (“Vorleistung”) before you can be admitted to the examination. Therefore, you must pass all coursework successfully to be able to participate in the final examination. Nonaccomplishment or failure of successful completion lead to an exclusion from the final examination and consequently to failure of this seminar. The examination requirement will be graded on a scale from 1,0 (sehr gut) to 5,0 (nicht ausreichend) and will be the grade for the entire course.

Coursework:

- Completion of [“Introduction to R” on DataCamp](#) (pass/fail)
 - This DataCamp course is a good way to familiarize yourself with the very basics of R that we will need to get started with our course. To complete the DataCamp course, sign up on DataCamp via the invitation link that will be sent to your university email address and complete the course by **February 17, 2020**.
- Regular exercises on *Swirl* (pass/fail)
 - *Swirl* is a programme that allows you to revise each week's course materials interactively, at your own pace and directly in the R console. Weekly repetition exercises on *Swirl* (starting from Week 3) have to be completed by the beginning of the following session. Successful completion of 9 out of 10 exercises is required to pass this coursework. More on that in class Week 3.

Examination:

- Take-home exam (graded): Towards the end of the semester, you will have to complete a take-home exam. The take-home exam will consist of a data analysis project, for which we will hand out a data set and a problem description. The take-home exam will be handed out on June 1, 2020 and the write-up consisting of a PDF report and the RMarkdown file that has produced it are due on ILIAS by **June 7, 2020**. The hard copy of the PDF report has to be handed in by **June 10, 2020**. Failure to upload your take-home exam to ILIAS on time will result in failing the course; failure to submit the hard copy on time will result in a one grade point (1.0) deduction from your final grade.

Late work policy: Late submissions will not be accepted. If you think you will not be able to meet a deadline, we expect you to discuss these issues with us in good time *before*, not after the deadline.

Cheating and plagiarism: We expect you to complete all coursework and examination on your own. Cheating and plagiarising can lead to failing the course or expulsion from the University of Mannheim. All your submissions will be checked for cheating and plagiarism.

Software: We will work with R, a free and open-source statistical programming language. R is not only well suited for performing state-of-the-art quantitative analyses but is also great for generating really nice and professional data visualizations. You can download the R version suited for your operating system for free at <https://www.r-project.org/>.

We also recommend that you use RStudio, which is a graphical user interface for R. With a constantly growing number of features, it is the preferred choice for working with R – especially for beginners. Just like R, RStudio is free and open-source and you can download it [here](#).

To make things easier for you, in the labs we will use [RStudio Cloud](#), so you do not need to install anything on your computers.

Note that if you want to knit .Rmd files to PDFs on your own computers, you will additionally have to install a [L^AT_EX](#) distribution (e.g. MiKTeX for Windows or MacTeX for Mac OS).

Course Schedule

We expect you to prepare for class by working through the chapter(s) listed under “Required readings” for each session. The material listed under “Optional readings” and “Further material” can help you to gain a deeper understanding of the contents we discuss.

Week 01 (Feb 10): Course Organization and Introduction to R I

In this session, we will discuss the course schedule and other organizational issues. We will also take a first peek at R, RStudio and RStudio Cloud and will learn how to work with R Markdown documents.

Optional readings:

- Wickham, H. and G. Grolemund (2017). *R for Data Science*. Sebastopol: O’Reilly. Chapters [27](#), [29](#) & [30](#).
- Xie, Y., J. J. Allaire and G. Grolemund (2018). *R Markdown: The Definitive Guide*. Boca Raton: Chaoman and Hall/CRC. Available [here](#).

Further material:

- [R Markdown Cheat Sheet](#)
- [R Markdown Reference Guide](#)

Week 02 (Feb 17): Introduction to R II

We will start working with R, learning about and practicing its logic and some basics: objects, vectors, data frames and functions. We will also talk about good coding practices.

Optional readings:

- Wickham, H. and G. Grolemund (2017). *R for Data Science*. Sebastopol: O’Reilly. Chapters [4](#), [8](#), [19](#) & [20](#).

Further material:

- [Tidyverse Style Guide \(short\)](#) & [Tidyverse Style Guide \(long\)](#)
- [Introduction to R on DataCamp](#)

Week 03 (Feb 24): Data Wrangling

Data preparation is often the most tedious task in your programming workflow. In this session, we will learn how to import different data formats into R, how to combine data sets and how to manipulate data, e.g. by creating new variables or recoding existing ones.

Required readings:

- Wickham, H., and G. Grolemund (2017). *R for Data Science*. Sebastopol: O’Reilly. Chapter [5](#).

Optional readings:

- Wickham, H. and G. Grolemund (2017). *R for Data Science*. Sebastopol: O’Reilly. Chapter [18](#).

Further material:

- [Data Import Cheat Sheet](#)
- [Data Transformation with dplyr Cheat Sheet](#)

Week 04 (Mar 02): Descriptive Analyses

We will discuss statistics that allow us to describe the distribution of data and what they can tell us about the data.

Required readings:

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapter 2.6.

Optional readings:

- Agresti, A. (2018). *Statistical Methods for the Social Sciences*. London: Harlow. Chapter 3.
- Monogan, J. E. (2015). *Political Analysis Using R*. Cham: Springer. Chapters 4.1 & 4.2.

Week 05 (Mar 09): Data Visualization

We will get to know ggplot2, a powerful package to plot your results in R.

Required readings:

- Wickham, H. and G. Grolemund (2017). *R for Data Science*. Sebastopol: O'Reilly. Chapters 3 & 7.

Further material:

- Wilkinson, L. (2012). *The Grammar of Graphics*. New York: Springer.
- [Data Visualization with ggplot2 Cheat Sheet](#)
- [Data Visualization with ggplot \(Part 1\) on DataCamp](#)
- [Data Visualization with ggplot \(Part 2\) on DataCamp](#)

Week 06 (Mar 16): Uncertainty I

In this session we will look at fundamental concepts of uncertainty and inferential statistics and what underlying statistical theory we can use to get an idea about the quality of our statistical analyses.

Required readings:

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapters 6.4 & 7.1.

Optional readings:

- Monogan, J. E. (2015). *Political Analysis Using R*. Cham: Springer. Chapter 5.1.

Further material:

- [Foundations of Probability in R on DataCamp](#)
- [Foundations of Inference in R \(Chapter 1\) on DataCamp](#)

Week 07 (Mar 23): Uncertainty II

Building on the previous session, in this session we will get to know standard errors, confidence intervals and hypothesis testing, all of which are techniques to test and evaluate the performance of our statistical measurements.

Required readings:

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapters 7.1.2 & 7.1.3.
- Monogan, J. E. (2015). *Political Analysis Using R*. Cham: Springer. Chapter 5.1.

Optional readings:

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapter 7.2.

Further material:

- [Foundations of Inference in R \(Chapters 2–4\) on DataCamp](#)

Week 08 (Mar 30): Linear Regression I

We will discuss the fundamental problem of causality and how regression analysis can help us to overcome it. We begin with modelling continuous outcome variables, repeating the basics of linear regression and focussing on how to interpret the results in meaningful ways.

Required readings:

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapters 4.2 & 4.3.2.

Optional readings:

- Fox, J. (2008). *Applied Regression Analysis and Generalized Linear Models*. Los Angeles: Sage. Chapter 5.

Further material:

- [Modeling with Data in the Tidyverse on DataCamp](#)

Week 09 (Apr 06): Easter Break

no class

Week 10 (Apr 13): Easter Break

no class

Week 11 (Apr 20): Linear Regression II

In this session, we will learn how to model conditional hypotheses by using interaction terms.

Required readings:

- Brambor, T., W. Clark and M. Golder (2006). "Understanding Interaction Models: Improving Empirical Analyses". *Political Analysis* 14(1), 63-83.

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapter 4.3.3.

Further material:

- [Multiple and Logistic Regression in R \(Chapters 1, 2 & 3\) on DataCamp](#)

Week 12 (Apr 27): Linear Regression III

We will return to the discussion of how to study causality and take a peek at two regression designs for causal inference with observational data: difference-in-difference and regression discontinuity designs.

Required readings:

- Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press. Chapters 2.5.3 & 4.3.4.

Optional readings:

- Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press. Chapters 5 & 6.

Week 13 (May 04): Logistic Regression

We use logistic regression for binary or categorical dependent variables. In this session we will learn when it is appropriate to use logistic regression, how it works theoretically, how we implement it in R and how we interpret the results.

Required readings:

- Monogan, J. E. (2015). *Political Analysis Using R*. Cham: Springer. Chapters 7.1 & 7.2.

Optional readings:

- Fox, J. and S. Weisberg (2014). *An R Companion to Applied Regression*. 2nd Edition. Los Angeles: Sage. Chapters 5.1 - 5.3.
- Hobolt, S. B. (2016). "The Brexit vote: a divided nation, a divided continent". *Journal of European Public Policy* 23(9), 1259-1277.

Further material:

- [Multiple and Logistic Regression in R \(Chapter 4\) on DataCamp](#)

Week 14 (May 11): Count Models

We will learn how to model count data and how to communicate the results.

Required readings:

- Monogan, J. E. (2015). *Political Analysis Using R*. Cham: Springer. Chapter 7.3.
- Slapin, J. B. and S.-O. Proksch (2010). "Look who's talking: Parliamentary debate in the European Union". *European Union Politics* 11(3), 333-357.

Optional readings:

- Fox, J. (2008). *Applied Regression Analysis*. Los Angeles: Sage. Chapter 15.

Week 15 (May 18): TBA

no class

Week 16 (May 25): Open Office Hour

On this date we will have an open office hour.