

Syllabus

Introduction into R for social scientists

By dr. Elze G. Ufkes and Benjamin Ziepert
Version: 15 November 2018

1. An introduction into R

For many years the most commonly used platform for data analyses in psychology was SPSS. It was first released in 1968 as the Statistical Package for the Social Sciences (SPSS). Back then SPSS was a program running on mainframe computers using punched cards for input (if you ever long for romantic stories about data-analyses you should ask one of the older professors about the time they still had to bike over to the university's data-center for running their analysis). Meanwhile the program developed to a highly versatile statistical tool, forming the standard for many social scientists. In 2009 IBM acquired SPSS for almost 2 billion dollars, and since then further developed and commercialized this tool (Wikipedia, 2017).

Besides the advantages of SPSS for data management and statistical analysis (e.g., the versatility and ease of use of the user interface) the program also has a number of important disadvantages that can be summarized in four points:

- First, although versatile SPSS has limits in the types of analysis it can perform. For more advanced, but increasingly commonly used analyses, such as Structural Equation Modelling (SEM), multi-level models for analyzing nested data, or analyses crossing the border between qualitative and quantitative analyses (such as content analysis) researchers often have to turn to alternative programs.
- Second, partially because the strong focus on user experience in SPSS the assumptions and methods used for specific analyses are often implicit (and even unknown) to the user.
- Third, the latest insights in data-analyses are only slowly incorporated because the development is in the hands of a limited number of developers.
- And fourth, because SPSS is a commercial package, each update means that universities or researchers need to invest in purchasing the license for a new version of SPSS.

Partially because of these limitations the statistical package R has become more popular over the last year. R is an open source statistical program ideal for each step of data science: from data manipulation, visualization to (advanced) analyses.

The general aim of this introductory course is to teach the basics of R, as well as getting some hands-on experience with more advanced data-analyses. That is, let you experience the reason why you would start using R in the first place.

The course is partially based on three in-class sessions as well as two online self-study assignments. It starts with a session introducing the basics of R and getting everybody started on the platform we will use for the self-study assignments. In the second session, we will discuss the self-study assignments, discuss some good-practices while using R, and practice common statistical analysis. In the third and final session, we will discover the true advantage of using R by harvesting, visualizing, and analyzing social media messages.

2. Overview of the course

Teacher

Benjamin Ziepert
research@b.ziepert.de

Time investment

Max. 16 hours

Requirements

- Laptop running Windows or OS.X with [R](#), [R Studio](#), [R Tools](#) (windows only), and the latest operating system updates installed. The laptop should boot within 60 seconds and open files within 10 seconds. See chapter 3 for instructions on how to install the R programs.
- Basic IT knowledge.

Most of you will have more than enough IT knowledge for the course and beyond. However, if you struggle with PCs, check if you are able perform the actions below and otherwise familiarize yourself with them: downloading a file from a website, installing a program, unzipping a file, creating a folder, renaming a folder, moving files to a folder, copy and paste text, perform a right click, save a file, etc.

Aims

After following this course students are able to:

1. List the benefits and challenges of using R as an alternative software package for data analysis in the social sciences
2. Execute basic functions of R, including:
 - a. Installing and operating R and R-studio
 - b. Installing and operating R-libraries
 - c. Importing data from other statistical software packages such as SPSS and EXCEL
 - d. Creating new variables such as vectors, lists and data frames in R
3. Given a structured data-set and hypotheses, conduct common statistical analyses for master thesis research, such as correlation, t-test, regression, ANOVA and process analysis using R.
4. Given step-by-step instructions and a r-script, perform content-analysis of Twitter media, including:
 - a. Harvesting twitter messages based on keywords
 - b. Performing sentiment-analyses
 - c. Creating visualizations of most frequent used words

Passing requirements

- Attendance of all sessions
- Complete DataCamp assignments with at least 7000 XP (max. possible 7250)
- Complete R script assignment with statistical analysis
- Complete Twitter analysis and present results

Sessions (6 hours)

Session 1

November 22nd, 8.45 – 10.30 (90 minutes)

- Preparation:
 - o Read assigned literature (see *Reading*)
 - o Bring laptops with R, RStudio, & Rtools (windows only) installed (see Chapter 3)
- Topics:
 - o Introduction into R
 - o Introduction *DataCamp* for self-study
 - o R setup
- In-class:
 - o Download the session files <https://benjaminziepert.com/teaching>
- Self-study Assignment
 - o Voluntary: to start working on DataCamp assignments in-class

Session 2

November 20th, 10.45 – 12.30 (90 minutes)

- Preparation:
 - o Complete all self-study assignments in Data-camp (see online self-study assignments)
 - o Prepare at least one question for Q&A
 - o Bring laptops with R, RStudio, & Rtools (windows only) installed (see Chapter 3)
- Topics
 - o Q&A pertaining self-study assignment
 - o Practice common statistical analyses
 - o Good practices when using R
- In-class
 - o Download the hand-out and files from <https://benjaminziepert.com/teaching>
- Assignment
 - o R script with statistical analysis

Session 3

November 20th, 13.45 – 17.30 (180 minutes)

- Topic
 - o In class R-project: analyzing twitter media using R
- In class
 - o Download the hand-out and data-files <https://benjaminziepert.com/teaching>
- Assignment
 - o Twitter analysis and presentation

Online self-study assignments (6 hours)

Data camp courses:

- Assignment 1: Introduction to R (4 hours)
- Assignment 2: Importing data from statistical software packages (2 hours)

Reading (4 hours)

- Baker, M. (2017). Scientific computing: Code alert. *Nature*, 541(7638), 563–565. doi:10.1038/nj7638-563a
- Culpepper, S. A., & Aguinis, H. (2010). R is for Revolution. *Organizational Research Methods*, 14(4), 735–740. doi:10.1177/1094428109355485

3. Installing R, RStudio, and Rtools

During—and hopefully after—the course you will use R and RStudio. R is the basic statistical program but has a very poor interface. Therefore, you—just as most social scientist working with R—never will have to open R. Instead we will use RStudio to write the script and inspect the outcomes of our analyses. Note that you have to install both programs because RStudio is an add-on that relies on R to work.

Mac Users

To Install R

1. Open an internet browser and go to www.r-project.org.
2. Click the "download R" link in the middle of the page under "Getting Started."
3. Select a CRAN location (a mirror site) and click the corresponding link.
4. Click on the "Download R for (Mac) OS X" link at the top of the page.
5. Click on the file containing the latest version of R under "Files."
6. Save the .pkg file, double-click it to open, and follow the installation instructions.
7. Now that R is installed, you need to download and install RStudio.

To Install RStudio

1. Go to www.rstudio.com and click on the "Download RStudio" button.
2. Click on "Download RStudio Desktop."
3. Click on the version recommended for your system, or the latest Mac version, save the .dmg file on your computer, double-click it to open, and then drag and drop it to your applications folder.

Rtools

As a Mac user you won't need to install Rtools. A similar software is already installed.

Windows Users

To Install R:

1. Open an internet browser and go to www.r-project.org.
2. Click the "download R" link in the middle of the page under "Getting Started."
3. Select a CRAN location (a mirror site) and click the corresponding link.
4. Click on the "Download R for Windows" link at the top of the page.
5. Click on the "install R for the first time" link at the top of the page.
6. Click "Download R for Windows" and save the executable file somewhere on your computer. Run the .exe file and follow the installation instructions.
7. Now that R is installed, you need to download and install RStudio.

To Install RStudio

1. Go to www.rstudio.com and click on the "Download RStudio" button.
2. Click on "Download RStudio Desktop."
3. Click on the version recommended for your system, or the latest Windows version, and save the executable file. Run the .exe file and follow the installation instructions.

To Install Rtools:

1. Open an internet browser and go to <https://cran.r-project.org/bin/windows/Rtools/>.
2. Click in the download table on the latest “Rtoolsxx.exe” (e.g. Rtools35.exe).
3. Save the executable file somewhere on your computer. Run the .exe file and follow the installation instructions.