

**PhD Course: Introduction to R**

January 16 - January 18, 2017

Department of Political Science, University of Copenhagen

Sebastian Barfort ([s.barfort@lse.ac.uk](mailto:s.barfort@lse.ac.uk)).

**Objective:** The aim of this course is to introduce and prepare participants to do quantitative academic work within the social sciences. As increasing emphasis in academics is being placed on the skills needed to effectively gather, handle, model, and analyze data as well as present results to a range of audiences, this course will provide participants with important tools for doing independent research. Furthermore, the skills taught in this course are also in high demand outside academia. Participants will be introduced to a modern workflow for doing effective quantitative research using the R programming language. R is a free (based on open source principles) statistical software package, which is supported by a very large international research community. The program can be used to do all types of statistical analyses.

**Content:** A tentative outline of the course is below:

- *Day 1:* Introduction to R, basics (reading/writing data, data cleaning, merging, types of data), functions, getting help when things don't work, data visualization.
- *Day 2:* Linear and binary choice models, panel data, interactions, standard errors, methods for causal inference, data visualization.
- *Day 3:* Analyzing text data, web scraping, statistical learning, communicating results.

My plan is to tailor the third day to the specific needs of the course participants. Therefore, if you're enrolled and have a preference for some subset of topics for the third day (or other topics not listed), please send me an email.

**Format and Evaluation:** The course will be hands-on and will primarily be practical instruction using R. Focus will be on giving participants hands-on experience doing actual analysis, and finish the course with a series of small scripts that can be used in the future. Practical exercises using real-world research examples from economics and political science will be assigned for each session.

**Preparation:** Please make sure you install R and RStudio before the first day of class. R is a free software environment for statistical computing and graphics. It works on all platforms. If you want to watch a step-by-step tutorial on how to install R for Mac or Windows, you can watch these videos

- [Install R for Windows](#)
- [Installing R on the Mac](#)

I recommend you use an [IDE](#) to interact with R. IDEs integrate text editing and syntax highlighting, and highly recommend you download and use [RStudio](#). It's free and modern, and if you're new to R this will make it much easier to get started. There's a nice video of how to install RStudio [here](#)

- [Installing RStudio](#)

Once installed, try to open up RStudio and type the following in the console (lines that start with `##` are the console's output) - don't worry if the commands make no sense to you right now.

```
x = 5
y = 2
x + y
## [1] 7
```

```
print("R is running on my computer")
## [1] "R is running on my computer"
```

**Important:** If you for some reason can't get R and RStudio installed before we meet then send me an email before the first class. However, I expect that you've tried as hard as you can using all the resources available online before emailing me.

**References:** Below are references for each day of the course. It's up to you to decide whether to read these before the course, they're provided primarily if you want to catch up on some details after the course. I will mention supplementary texts in my slides in case you want to go further with the topics covered in the course.

### Day 1

- Leeper, Thomas. 2016. “Really Introductory Introduction to R”. Page 1-17.
- Golemund, Garrett and Hadley Wickham. 2016. “R for Data Science”. The book is to be published by O’Reilly in late 2016/early 2017, but it’s free online. *Chapters 3-5, 11-15, 19, 21.*
- Schwabish, Jonathan A. 2014. “An Economist’s Guide to Visualizing Data”. *Journal of Economic Perspectives*, 28(1): 209-34.
- Healy, Kieran and James Moody. 2014. “Data Visualization in Sociology”. *Annual Review of Sociology*, 40:105–128.

### Day 2

- Golemund, Garrett and Hadley Wickham. 2016. “R for Data Science”. *Chapters 23-25, 28.*
- Croissant, Yves, and Giovanni Millo. 2008. “Panel data econometrics in R: The plm package.” *Journal of Statistical Software* 27.2: 1-43.
- Esarey, Justin, and Andrew Menger. 2015. “Practical and Effective Approaches to Dealing with Clustered Data.” Department of Political Science, Rice University.

### Day 3

- Wickham, Hadley. 2014. “rvest: Easy web scraping with R”. RStudio Blog.
- Shiab, Nael. 2015. “On the Ethics of Web Scraping and Data Journalism”. Global Investigative Journalism Network.
- The Economist. 2016. “What APIs are.”
- Imai, Kosuke. 2016. *A First Course in Quantitative Social Science. Chapter 5, Section 1-2.*
- Grimmer, Justin, and Brandon M. Stewart. 2013. “Text as data: The promise and pitfalls of automatic content analysis methods for political texts.” *Political Analysis*, 21.3: 267-297.
- Silge, Julia and David Robinson. 2016. *Tidy Text Mining with R. Chapters 2, 6, 7.*
- Varian. Hal. 2014. “Big Data: New Tricks for Econometrics”. *Journal of Economic Perspectives*, 28.2: 3-27.
- Kleinberg, Jon, et al. “Prediction policy problems.” *American Economic Review*, 105.5 (2015): 491-495.

- Grolemond, Garrett and Hadley Wickham. 2016. “R for Data Science”. *Chapters 27, 29*.

We will work with data from these papers

- Jones, Benjamin F., and Benjamin A. Olken. 2005. “Do Leaders Matter? National Leadership and Growth Since World War II.” *The Quarterly Journal of Economics*, 120.3: 835-864.
- Andersen, Asger L., David D. Lassen, and Lasse H. W. Nielsen. 2012. “Late budgets.” *American Economic Journal: Economic Policy*, 4.4: 1-40.
- Gentzkow, Matthew. 2006. “Television and voter turnout.” *The Quarterly Journal of Economics*. 931-972.
- Meyersson, Erik. 2014. “Islamic rule and the empowerment of the poor and pious.” *Econometrica*, 82.1: 229-269.
- Ananat, Elizabeth Oltmans. 2011. “The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality.” *American Economic Journal: Applied Economics*, 3.2: 34-66.