

Software Tools for Data Analysis, Fall 2018

Data Analysis in R: An Introduction

Anatoly Arlashin

Woods College

Outline

- 1 R's basics
- 2 Importing data
- 3 Subsetting data
- 4 Graphics in R
- 5 Efficient data management
- 6 Econometrics

Downloading and installing

The two essential things are R itself and the coding environment for it called RStudio.

- www.r-project.org

- www.rstudio.com

Installation of both packages is pretty straightforward, the default options will be sufficient for all our needs.

Online resources

- [An Introduction to R](#) — a manual from R Project, the creators of R
- [A \(very\) short introduction to R](#) — a 12 page summary of the most basic features of R and RStudio
- [Cookbook for R](#) — a collection of detailed and colorful tutorials
- [R bloggers](#) — a blog platform with lots of small tutorials and Q&As
- [Quick-R](#) — yet another collection of tutorials
- [Getting Started in Data Analysis](#) — a workshop from Princeton
- [Online learning](#) — a collection of links from RStudio's authors

UI and coding basics

Tutorials on using Rstudio and some basic stuff about R:

- [Introduction to R and RStudio](#) — #1
- [Introduction to R and RStudio](#) — #2
- [Introduction to R and RStudio](#) — #3
- [Coding Style guide](#) — useful advice on how to write clean and readable code

Some more advanced all-in-one tutorials:

- [Introduction to R Seminar](#)

Importing data

The way you import data into R depends on two things:

- the existing format and structure of the data
- the desired format and structure of the data

As with almost any other aspect of R, there exists a variety of packages designed to import specific data types. Some of those packages are minimalistic, while others offer a wide selection of options.

The differences in various packages are mostly about

- data formats that are supported (*.xls, *.dta, *.csv, etc.)
- type of R object that will contain the imported data (matrix, list, dataframe, etc.)
- speed of import (only matters for large datasets)

Importing data tutorials

As with everything else, there's plenty of tutorials online on how to import various types of data into R. Below is a small selection of them:

- [Quick-R — Importing Data](#) (short)
- [R Tutorial — Data Import](#) (short)
- [IDRE UCLA — How to Input data into R](#) (short)
- DataCamp's "Data importing in R" — [Part I](#), [Part II](#) (average)
- [CRAN — R Data Import/Export](#) (long)

Data subsetting tutorials

We will follow the first tutorial in the list below in our class. Others contain shorter and less formal versions of intro to data subsetting in R.

- Advanced R
- Quick R
- Cookbook for R

Basic plotting in R

Below are a couple of short tutorials on how you can do basic data visualization using R's base function `plot()`

- Producing Simple Graphs with R

ggplot2 tutorials

ggplot2 is the most popular package for producing complex graphic visualizations in R. We will use the first two tutorials in the list below in our class, but other are very helpful too.

- [Introduction to R Graphics with ggplot2](#)
- [Top 50 ggplot2 Visualizations](#)
- [ggplot2 Cheat-Sheet from RStudio](#)
- [Data visualization using ggplot2](#)
- [3 hours to learn ggplot2](#)
- [Data Visualization with ggplot2](#)
- [R Graphics: Introduction To ggplot2](#)

Visualizing time-series data in R

Below are a couple of short tutorials on how you can do some specialized visualization for time-series data

- Exploring & Visualizing Time Series
- Time Series Graphs
- Time series in ggplot2

Data management with data.table

The `data.table` package is one of the most efficient ways to manage large datasets in R. Below are links to some of the more useful introductory materials on `data.table`:

- [Package documentation](#) with a lot of simple and useful examples
- [A collection](#) of useful links by authors, including [FAQ](#)
- [DataCamp tutorial](#) from the authors
- [Intro to the data.table Package](#)
- [Data.Table Tutorial with 50 Examples](#)
- [Advanced tips and tricks with data.table](#)

Econometrics tutorials

Below are tutorials on basic econometrics in R from several sources:

- [Applied Econometrics with R](#)
- [Data Analysis Workshop from Princeton](#)
- [Harvard Workshop](#)