Department of Economics — Intro to Causal Inference in Econometrics
Yale University — GEC
Prof. Edward Vytlacil — June 2023

R Basics

1 What is R?

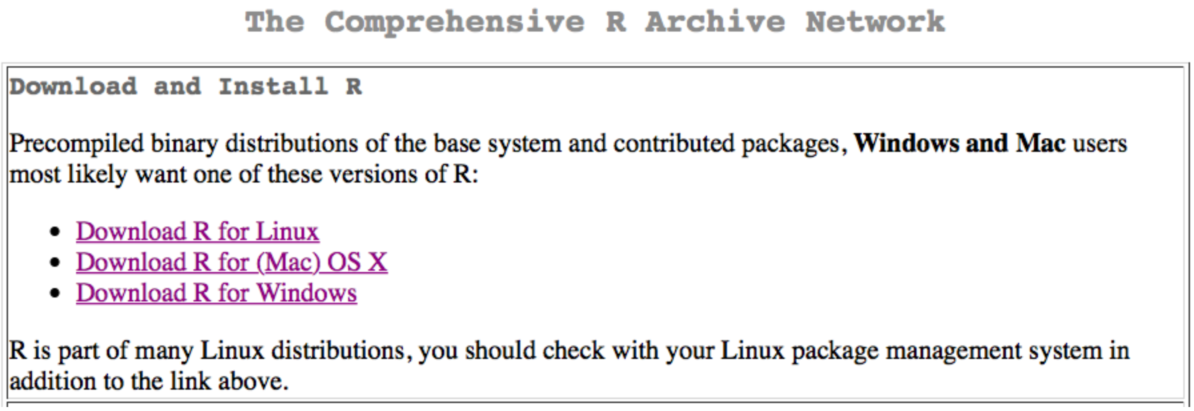
R is an open source implementation of the object-oriented programming language S. It is widely used by applied statisticians and its libraries implement a wide variety of statistical and graphical techniques with applications to a range of disciplines, such as the agricultural and biological sciences, genetics, neuroscience and economics.

2 Why should you learn R?

Like any other language, learning R cannot be done overnight. However, its intuitive syntax makes learning it fairly easy. Moreover, because of its flexibility and usefulness in tackling almost any type of data analysis problem, its powerful visualization abilities as well as the extensive and constantly expanding collection of libraries and resources, it is used in both, academia and industry. At the same time, R is a great tool for learning econometrics and completely free for any use.

3 Obtaining R

R can be downloaded from `https://cran.r-project.org`. If you are using a Windows or a Mac device, you probably want to download the precompiled binaries listed on the landing page.



The Comprehensive R Archive Network

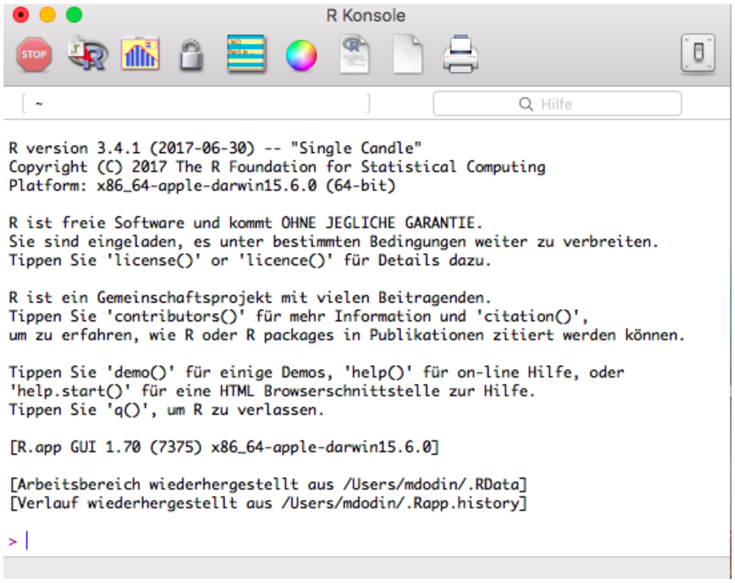
Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- Download R for Linux
- Download R for (Mac) OS X
- Download R for Windows

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

After running the R installation package and executing the program, you should see a command line interface that looks similar to the picture below.



```
R version 3.4.1 (2017-06-30) -- "Single Candle"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)

R ist freie Software und kommt OHNE JEGLICHE GARANTIE.
Sie sind eingeladen, es unter bestimmten Bedingungen weiter zu verbreiten.
Tippen Sie 'license()' or 'licence()' für Details dazu.

R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.

Tippen Sie 'demo()' für einige Demos, 'help()' für on-line Hilfe, oder
'help.start()' für eine HTML Browserschnittstelle zur Hilfe.
Tippen Sie 'q()', um R zu verlassen.

[R.app GUI 1.70 (7375) x86_64-apple-darwin15.6.0]

[Arbeitsbereich wiederhergestellt aus /Users/mdodin/.RData]
[Verlauf wiederhergestellt aus /Users/mdodin/.Rapp.history]

> |
```

3.1 Libraries

Libraries (or packages) are collections of R functions, code and data sets. Since R as a programming language is not just designed to be an econometrics program, many of the functions that we are interested in are not included in the standard set of packages. Therefore, we will sometimes ask you to install packages related to certain topics covered in class or needed for problem sets.

There are different ways to download and install packages. The most direct way is to issue a command in R. For example, if you want to install the popular graphing package `ggplot2`, type

```
> install.packages("ggplot2",dep=TRUE)
```

and select a CRAN mirror of your choice. The `dep` parameter tells R to automatically also download those packages that our package of choice depends on and we recommend that you always set it to true.

Alternatively, if you want to use the GUI, you can install packages by clicking *install package* under the *package* menu. Note that in general you must have administrator privileges to your device to install the program and packages. If you are having problems installing R and/or packages that you can not solve, please get in touch with your teaching assistant. You can always see the full list of installed packages using the `library()` command. Finally, in order to use the functions that are stored in a library, we have to load it. This is also done using the `library()` command. For example, in order to be able to use the functions of the `ggplot2` package, type

```
> library(ggplot2)
```

3.2 Help and References

The `help()` function and the `?` operator in R provide access to the documentation pages for R functions, data sets, and packages. You can also search the documentation for a word or a phrase using the `help.search()` command. The built in help functions are especially useful to understand what arguments a specific function requires. For example the commands

```
>?lm
>help(lm)
>help.search("linear model")
```

lead to the documentation of the linear model function that we will use extensively.

However, many times a simple search online will yield the best results. As you will notice, most of the questions that you will encounter have at some point been asked and answered on one of the many R discussion boards such as `https://stackoverflow.com`.

There are also many excellent and free R references available online. For example, these notes are based in parts on the book *Econometrics in R* by G. Farnsworth that is available for free. If your time permits and you want to dig deeper, there are also more programming oriented references such as *An Introduction to R* by W. N. Venables, D. M. Smith and the R Core Team. However, for the sake of this lecture, I recommend learning by trial and error, as it is the most time efficient approach and sufficient for the type of coding problems that we will consider.

4 Writing Scripts in R

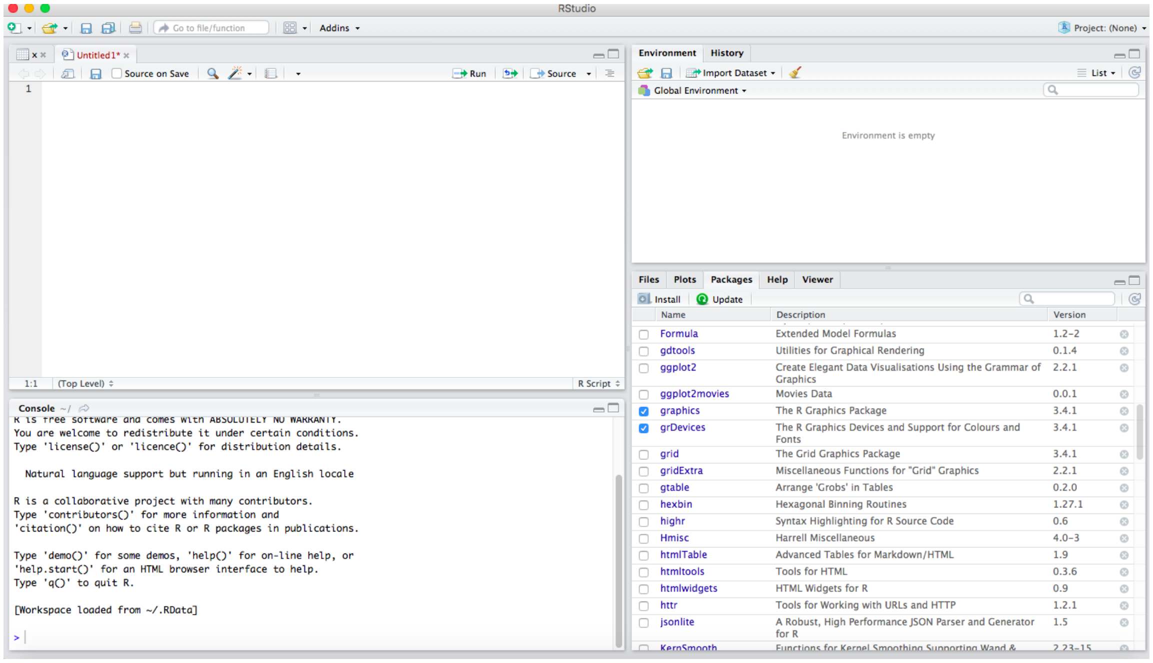
In principle, we can directly interact with R through its command prompt. For example, the command `2+2` yields the output `4`.

```
> 2+2
[1] 4
```

However, for any project that is longer than a few lines, it is recommended that you write your code in script files. Script files are simple text files that contain commands and comments. They offer several advantages over using the R console. Most importantly, they allow you to organize your work (and thoughts) in a clear fashion. Moreover, it is possible to execute your script or parts of it directy from the script file. To open a new script file click *File* and choose the option *New Script.* R script files generally have the extension ".R" and can in principle be edited using any text editor, with different editors having different advantages. In the next subsection we will give a brief recommendation regarding the editor and the GUI but you are of course free to choose whatever works best for you.

4.1 Editors and GUIs

Code written for R can be run on many computational platforms with or without a graphical user interface. However, while R has a command line interface by default, there are several graphical front-ends available. A relatively new and popular GUI is offered by RStudio, a free and open-source integrated development environment (IDE) for R. It contains many useful features that we think make using and learning R easier and it is therefor recommended. You can download RStudio (Desktop) for free from **here**. After running the installation package and executing RStudio you should see a interface that looks similar to the picture below.



```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> 
```

	Name	Description	Version
□	Formula	Extended Model Formulas	1.2-2
□	gdtools	Utilities for Graphical Rendering	0.1.4
□	ggplot2	Create Elegant Data Visualisations Using the Grammar of Graphics	2.2.1
□	ggplot2movies	Movies Data	0.0.1
☑	graphics	The R Graphics Package	3.4.1
☑	grDevices	The R Graphics Devices and Support for Colours and Fonts	3.4.1
□	grid	The Grid Graphics Package	3.4.1
□	gridExtra	Miscellaneous Functions for "Grid" Graphics	2.2.1
□	gtable	Arrange 'Grobs' in Tables	0.2.0
□	hexbin	Hexagonal Binning Routines	1.27.1
□	highr	Syntax Highlighting for R Source Code	0.6
□	Hmisc	Harrell Miscellaneous	4.0-3
□	htmlTable	Advanced Tables for Markdown/HTML	1.9
□	htmltools	Tools for HTML	0.3.6
□	htmlwidgets	HTML Widgets for R	0.9
□	httr	Tools for Working with URLs and HTTP	1.2.1
□	jsonlite	A Robust, High Performance JSON Parser and Generator for R	1.5
□	KernSmooth	Functions for Kernel Smoothing Supporting Wand &	2.23-15