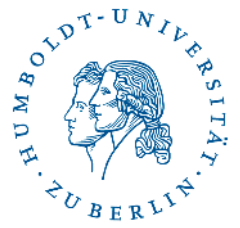


Business Analytics & Data Science



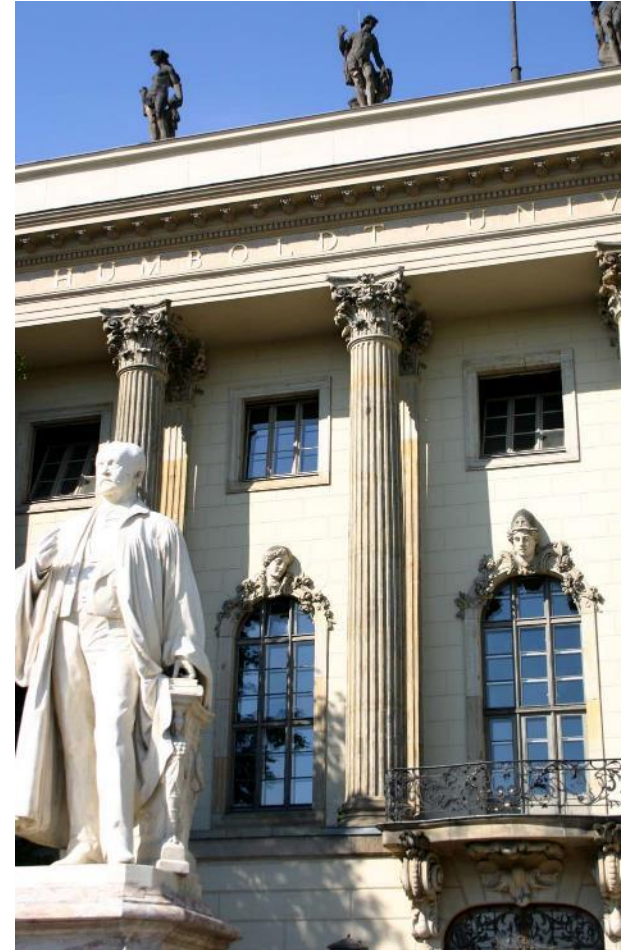
Winter Semester 2018 / 2019

Tutorial 1: A Primer in R Programming

Stefan Lessmann

Johannes Haupt

Alisa Kolesnikova

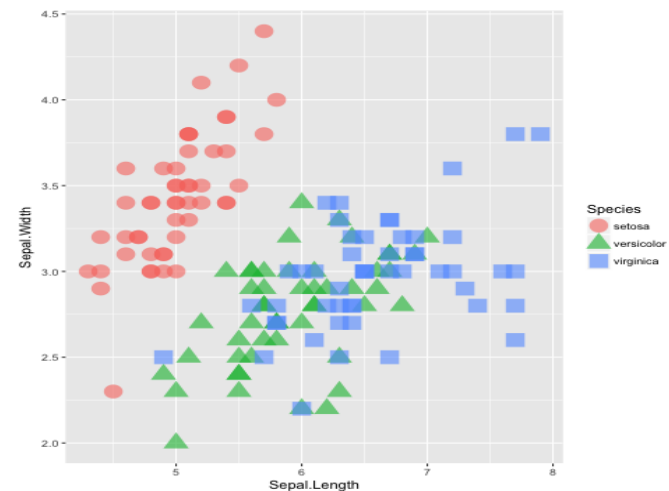


Agenda

- systems and tools for Data Science
- R Studio
- data types
- basic functions
- exercises

What R is About

- programming language for statistical computing
- most popular tool for predictive analytics (together with Python)
- open Source (<http://cran.r-project.org/>)
- best used within **R Studio** environment
- alternatively: Anaconda



R Studio



The screenshot displays the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Tools, and Help. The toolbar below the menu contains icons for file operations and a search bar. The main editor pane shows a script with R code snippets and their descriptions. The Environment pane on the right is empty. The Packages pane at the bottom right lists installed and available packages. The Console pane at the bottom left shows the command prompt.

Editor Pane:

```
Demos in package 'base':  
  
error.catching      More examples on catching and handling errors  
is.things           Explore some properties of R objects and is.FOO()  
functions. Not  
  
recursion           for newbies!  
scoping             Using recursion for adaptive integration  
                    An illustration of lexical scoping.  
  
Demos in package 'graphics':  
  
Hershey             Tables of the characters in the Hershey vector  
fonts  
Japanese            Tables of the Japanese characters in the Hershey  
vector fonts  
graphics            A show of some of R's graphics capabilities  
image              The image-like graphics builtins of R  
persp               Extended persp() examples  
plotmath            Examples of the use of mathematics annotation  
  
Demos in package 'grDevices':  
  
colors              A show of R's predefined colors()
```

Environment Pane:

Environment is empty

Packages Pane:

Name	Description	Version
System Library		
boot	Bootstrap Functions (originally by Angelo Canty for S)	1.3-11
checkpoint	Install Packages from Snapshots on the Checkpoint Server for Reproducibility	0.3.3
class	Functions for Classification	7.3-10
cluster	Cluster Analysis Extended Rousseeuw et al.	1.15.2
codetools	Code Analysis Tools for R	0.2-8
compiler	The R Compiler Package	3.1.1
datasets	The R Datasets Package	3.1.1
digest	Create cryptographic hash digests of R objects	0.6.4
evaluate	Parsing and evaluation tools that provide more details than the default.	0.5.5
foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...	0.8-61
formatR	Format R Code Automatically	1.0
graphics	The R Graphics Package	3.1.1
grDevices	The R Graphics Devices and Support for Colours and Fonts	3.1.1
grid	The Grid Graphics Package	3.1.1

Console Pane:

```
>  
>
```

R Studio



Read/write R scripts/functions stored on your computer

Workspace, shows all variables available in the current R session

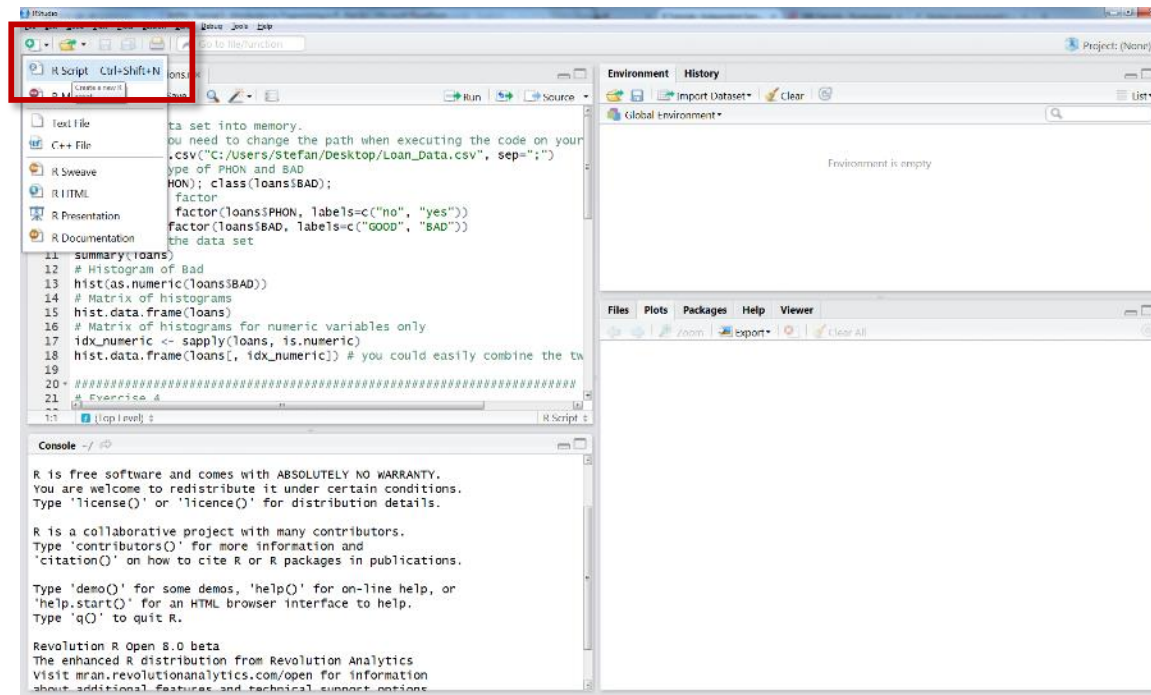
Console, to write and execute code directly

File and package explorer, plot and help window

Name	Description	Version
<input type="checkbox"/> boot	Bootstrap Functions (originally by Angelo Canty for S)	1.3-11
<input type="checkbox"/> checkpoint	Install Packages from Snapshots on the Checkpoint Server for Reproducibility	0.3.3
<input type="checkbox"/> class	Functions for Classification	7.3-10
<input type="checkbox"/> cluster	Cluster Analysis Extended Rousseeuw et al.	1.15.2
<input type="checkbox"/> codetools	Code Analysis Tools for R	0.2-8
<input type="checkbox"/> compiler	The R Compiler Package	3.1.1
<input type="checkbox"/> datasets	The R Datasets Package	3.1.1
<input type="checkbox"/> digest	Create cryptographic hash digests of R objects	0.6.4
<input type="checkbox"/> evaluate	Parsing and evaluation tools that provide more details than the default	0.5.5
<input type="checkbox"/> foreign		
<input type="checkbox"/> formatR		
<input type="checkbox"/> graphics		
<input type="checkbox"/> grDevices		
<input type="checkbox"/> grid		

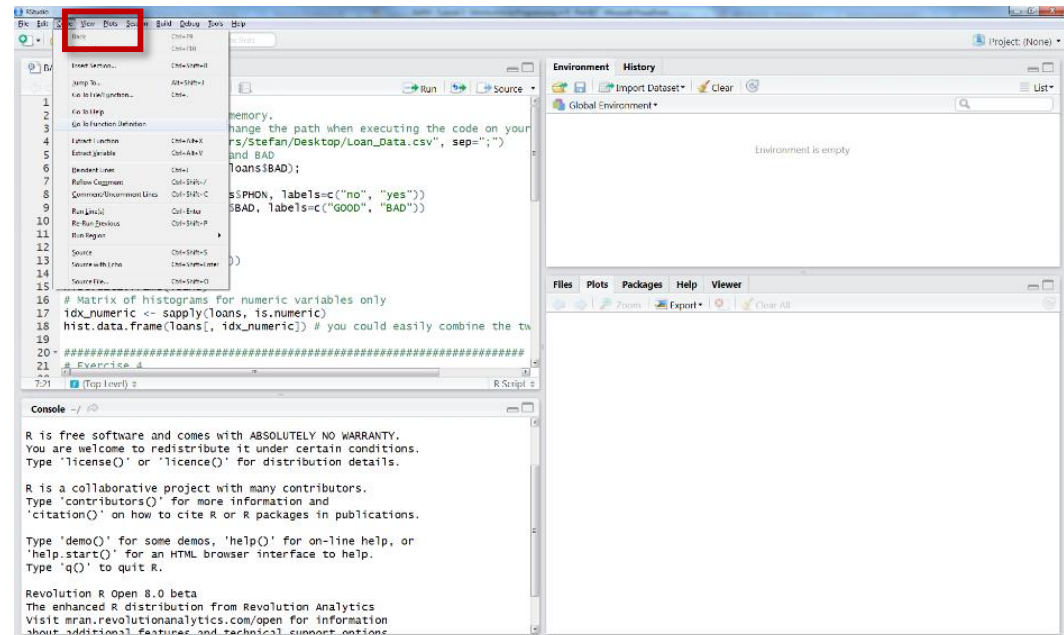
Writing R Scripts

You can write and execute R code using the **console window**. Alternatively, you can create **script files** and put your code there



Executing the written code

- run one line or part of one line
- run a block of R statements
- run the whole script



Shortcuts & Tricks

- **TAB** for auto-completion of functions and variable names
- **up/down arrow** to select past commands
- make code foldable by adding four or more hashtags or dashes after a comment, e.g. **# Section One -----**
- several shortcuts to run code:
 - ▶ **Ctrl + Enter**: Run current line
 - ▶ **Ctrl + Alt + B**: Run code up to and including current line
 - ▶ **Ctrl + Alt + R**: Run current script

Other Useful Resources

■ Learning R

- <https://swirlstats.com>
- <https://www.datacamp.com/courses/introduction-to-r>
- <http://data.princeton.edu/R/default.html>
- <https://www.coursera.org/course/rprog>

■ GOOGLE

When code doesn't work

- R help
- Google
- Experiment
- Ask during tutorial session
- Ask colleagues
- Ask on the moodle forum
- Put question on stackoverflow.

- Abstract:
 - What general problem do you need to solve?
 - Ex: "R regression outlier analysis",
"R Im missing values"
- Specific:
 - What function interests you?
 - Ex: "R glmnet"
 - What error do you get?
 - Ex: "Error in if (NULL == 1) { : **argument is of length zero**"