

# Einführung in die Datenanalyse mit R: Orga und Einführung in R und RStudio

---

**Marco Wähler**



Gebäude 37.03.03.14



marco.waehner@hhu.de



Sprechstunde nach Vereinbarung



NRW-FORSCHUNGSKOLLEG  
**ONLINE-PARTIZIPATION**

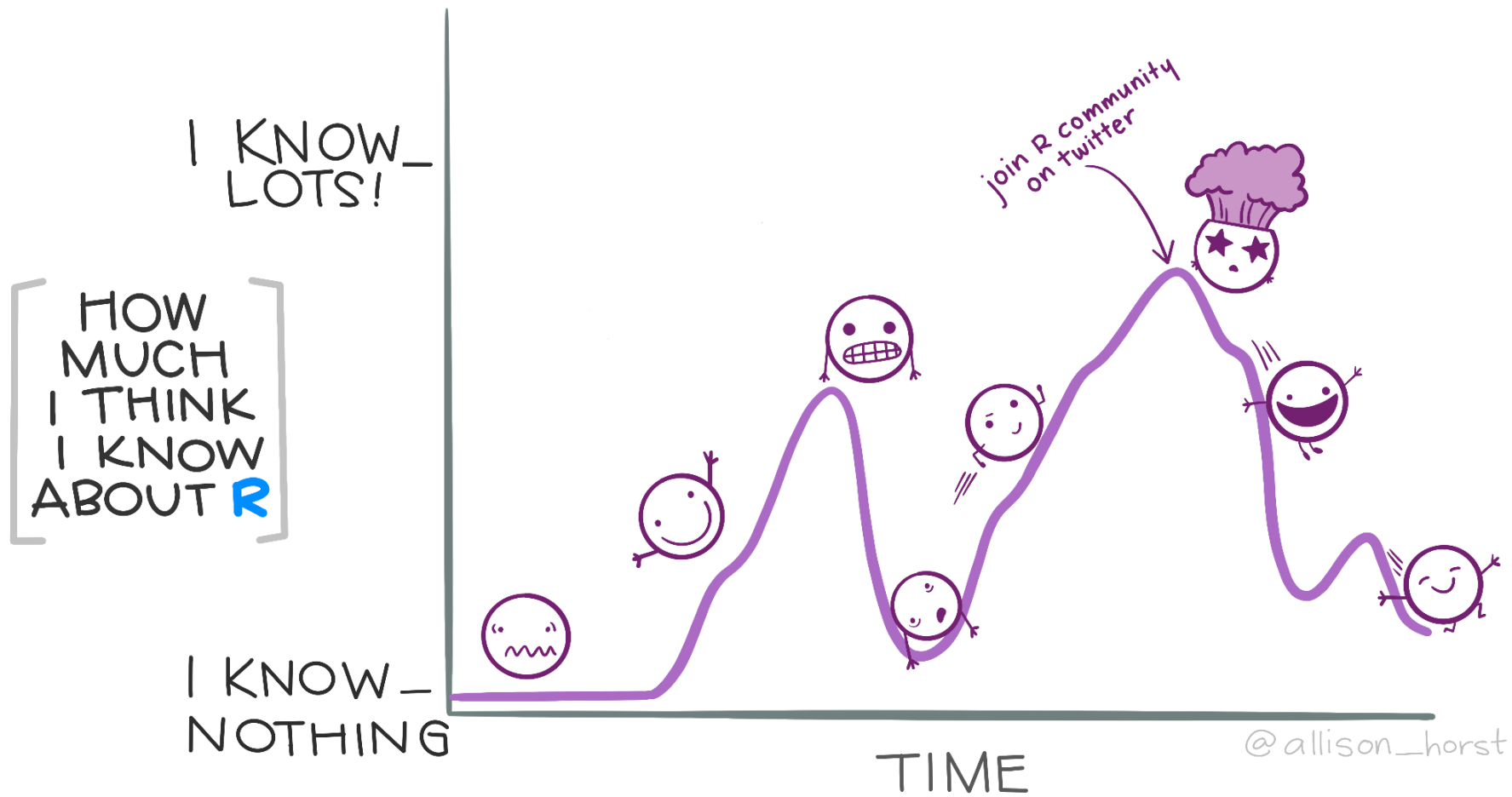
**hhu** Heinrich Heine  
Universität  
Düsseldorf



RESEARCH  
FOR THE  
DIGITAL AGE

- Veranstaltungsziele
  - Grundlegendes Verständnis der Programmiersprache R, z.B. Skripte, Rmarkdown, Packages und Funktionen
  - Umgang mit Ressourcen
  - Organisation der Datenanalyse, z.B. Projects
  - Selbstständige Bearbeitung einer empirischen Fragestellung, d.h. Datenimport, -aufbereitung, -visualisierung und -analyse





- Hinweise zur Durchführung einer hybriden Veranstaltung
  - Bitte stummschalten und gerne die Kamera anschalten
  - Rückfragen in den Chat oder direkt melden
    - Erst werden Rückfragen aus dem Seminarraum, dann aus dem Webex-Raum bearbeitet
    - Fragen aus dem Chat werden vorgelesen
    - Fragen aus dem Seminarraum für den Online-Raum wiederholen
    - Häufig ist es hilfreich Fehlermeldungen in den Chat zu kopieren
  - Ich teile stets meinen Bildschirm (RStudio, PPP etc.)
  - Bitte Bescheid sagen, wenn ich zu viel rumlaufe
  - Wenn Rückfragen oder Probleme nicht unmittelbar gelöst werden können, besteht die Möglichkeit in der Pause in den Breakout Room zu wechseln
    - Häufig ist es sinnvoll kurz den Bildschirm zu teilen

Die Inhalte basieren auf (acknowledgement):

- R – Ein Einführungsskript (Masch et al., 2021): <https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=58006>
- [R for Data Science](#) (Hadley Wickham)
- Introduction to R for Data Analysis (Johannes Breuer & Stefan Juenger, GESIS Summer School in Survey Methodology 2021): <https://github.com/jobreu/r-intro-gegis-2021>

## Kursmaterialien

- Die Materialien zum Kurs sind auf GitHub verfügbar:  
<https://github.com/marcohu/einfuehrung-in-R-hhu>
  - Unter Code -> Download ZIP können alle Files heruntergeladen werden
- Die Daten sind im Sciebo-Ordner (siehe Mail)

# **Veranstaltungsüberblick und -ablauf**

- Thematische Blöcke a 90 Minuten
- Jeder Block besteht aus einer kurzen Einführung, Beispielen in R und eigener Bearbeitung
- Hands-On: kurze Aufgaben zur eigenen Bearbeitung
  - Lösung wird am Ende kurz durchgegangen
  - Komplette Lösung im Skript
- Für einen Beteiligungsnachweis (Anrechnung nur im WP) muss ein Übungsblatt bearbeitet werden
  - Beinhaltet unterschiedliche Aufgaben, die thematisch das Seminar abdecken
  - Abgabe wird bewertet (Cut-Off: 50%)
  - Mehr zum Übungsblatt am Samstag, 9. April

**Erwartung vs.  
Befürchtung**



**Hello World!**

- R ist eine Statistik- und Programmiersprache
  - Version 1.0.0 im Jahr 2000 veröffentlicht ([Überblick zur Entwicklung und Philosophy von R von Roger D. Peng](#))
  - Entwicklung durch [R Core Group](#)
  - Frei („free as freedom“) und Open-Source



- RStudio ist eine Entwicklungsumgebung und Benutzeroberfläche (engl. Integrated Development Environment oder kurz: IDE)
  - RStudio wird seit etwa 2011 entwickelt
  - Als Desktop- und als Server-Version verfügbar (Desktop empfohlen!)
  - For free but not Open-Source ([free as free beer](#))
  - Es gibt zahlreiche Alternativen, z.B. [Jupyter Notebooks](#) für den Kurs wird allerdings RStudio empfohlen



```
R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (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.

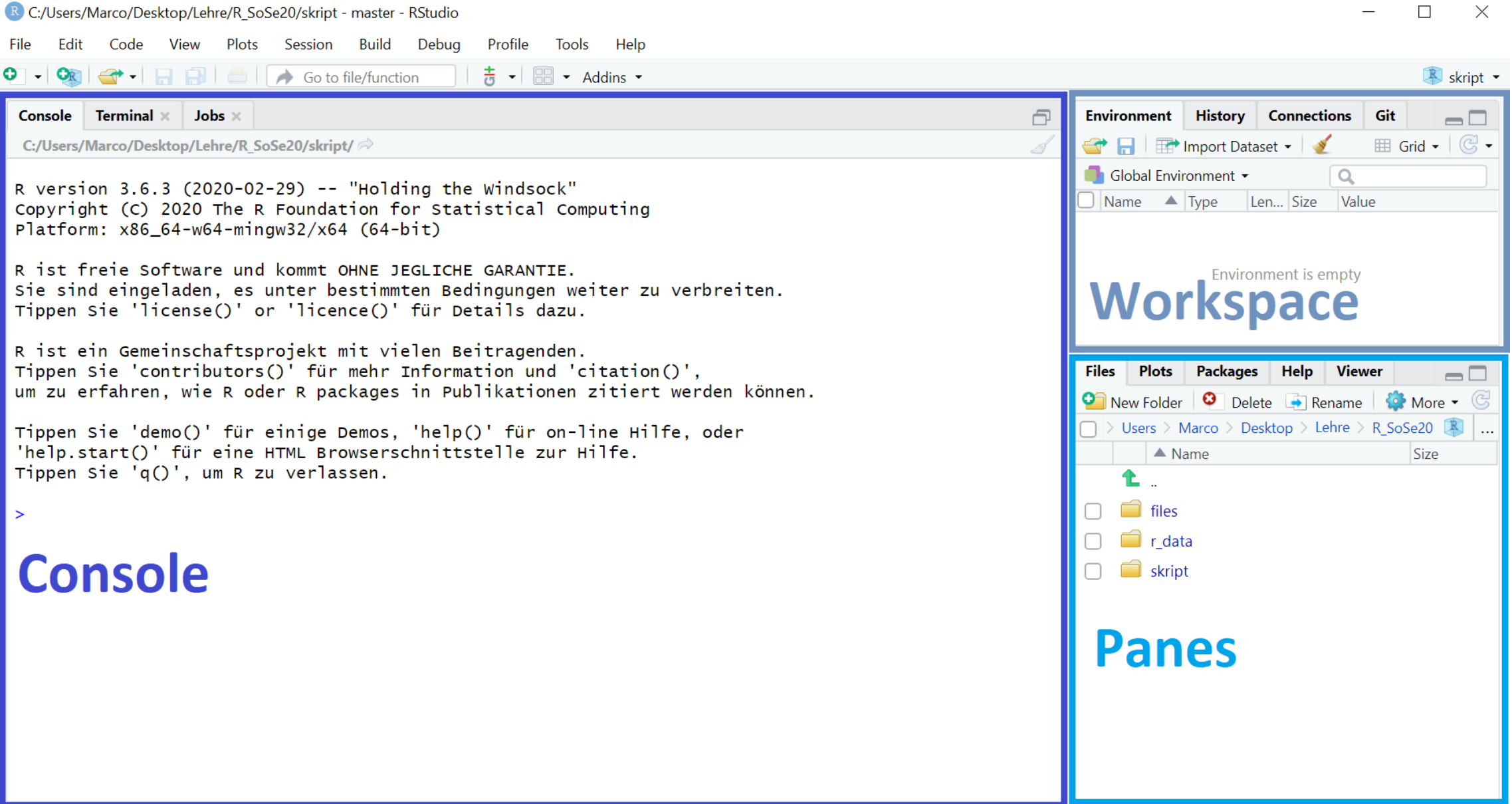
R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.

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-GUI

R Console



Untitled1 x

1

1:1 (Top Level) R Script

**Console** **Terminal** **Jobs**

C:/Users/Marco/Desktop/Lehre/R\_SoSe20/skript/

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.

> |

skript

**Environment** **History** **Connections** **Git**

Import Dataset Grid

Global Environment

Name	Type	Len...	Size	Value
Environment is empty				

**Files** **Plots** **Packages** **Help** **Viewer**

New Folder Delete Rename More

> Users > Marco > Desktop > Lehre > R\_SoSe20

Name	Size
..	
files	
r_data	
skript	

14

- Code wird allerdings nicht in der Console, sondern in Skripten (File -> new File -> R Script) geschrieben
- Haben die Endung „.R“ oder „.r“
- Cursor muss in die Zeile, die ausgeführt werden soll. Code wird anschließend mit einem Klick auf „run“ oder mit dem Shortcut „Strg“ + „Enter“ ausgeführt
- Kommentare werden mit einem „#“ eingeleitet

- 5 bis 10 Minuten, um sich mit der Umgebung vertraut zu machen
  - Rstudio öffnen
  - Skript öffnen
- R als Taschenrechner nutzen
  - $5 + 5$
  - $5 * 5$
  - $5 * 5 + 10$
  - $a <- 5 * 5$
  - $a + 8$



# **R in a nutshell: Objekte, Funktionen und Packages (Kapitel 1.4.2 & 1.4.3)**

- In R wird alles in Objekten gespeichert (Variablen, Daten, Funktionen etc.)
- Objekte haben verschiedene Klassen, Typen und Formate
- Objekte erhalten einen Wert durch Zuweisung
  - Dafür wird das „assignment“-Zeichen verwendet <-
  - „Kleiner-als“ + „minus“
  - Shortcut: „alt“ + „minus“

- Auch Funktionen werden in Objekten gespeichert
- Funktionen führen automatisiert Anweisungen/Aufgaben aus
  - Z.B. Mittelwerte mit `mean(x)`
  - Z.B. Tabellen mit `table(x)`
  - „?“ Funktion für Hilfe
- Auf Funktionsname folgte eine Klammer
  - In die Klammer kommt die Variable gefolgt von weiteren Argumenten
  - Z.B. `mean(x, na.rm = TRUE)`

- R bietet einen Kern an Funktionen (base R), die über Packages erweitert werden
  - Smartphone -> Apps
- Packages werden i.d.R. über Cran ([The Comprehensive R Archive Network](https://cran.r-project.org/)) installiert
  - Ca. 18.900 Packages werden auf Cran gelistet (Stand März 2022)
  - Qualitätskontrolle und Standards u.a. Dokumentation zur Funktionsweise
- Packages müssen einmalig installiert und für jede Session aktiviert werden
  - `install.packages(„package“)` und `library(package)`
  - Deinstallation über `remove.packages(„package“)`
  - Update aller Packages über `update.packages()`
- `sessionInfo()` zeigt aktivierte Packages

- Kernfunktion(en) von R (keine weitere Installation notwendig)
  - Z.B. `mean()` oder `table()`
- Verantwortung im „R Core Team“
  - Wird eher selten und sehr konservative erweitert/verändert

**Praxis:**

**R als Taschenrechner**

**Hands-On:**

**Webshop (Lösung)**

Vielen Dank!

**Marco Wähner**



Gebäude 37.03.03.14



marco.waehner@hhu.de



Sprechstunde nach Vereinbarung