



Einführung in die Statistik

Praktische Übung – Jürgen Hermes – IDH – SoSe 2023

Ziele dieses Kurses

- Vermittlung der Motivation für quantitative Datenanalyse
- Verständnis der zentralen Begriffe und Konzepte der statistischen Auswertung
- Erwerb der wichtigsten Grundlagen der Programmiersprache R
- Umgang mit der Entwicklungsumgebung RStudio

Eingesetzte Mittel

- Vorstellung der wichtigsten Konzepte im Präsenzunterricht
- Für R: Verweis auf Online-Kurs beim HPI (zum Selbststudium)
- Bearbeitung kleinerer Hausaufgaben (Abgabe über ILIAS)
- Zwei Teil-Testate: Eines zu R und eines zu statistischen Grundlagen

Die Programmiersprache R

- Freie, vollständige Programmiersprache, die meist für statistische Berechnungen und Visualisierungen eingesetzt wird.
- Publiziert 1993, aktuelle Version 4.1.3
- Benutzerfreundliche integrierte Entwicklungsumgebung RStudio
- Besonderheiten: Für statistische Berechnungen optimierte Datenstrukturen und Funktionen, besonderes Potential in der einfachen Grafikerzeugung.
- “To understand computations in R, two slogans are helpful: Everything that exists is an object. Everything that happens is a function call.” (John Chambers)
- Extrem große Nutzer:innen-Community, die viele vorgefertigte, nutzbare Pakete zur Verfügung stellt.

Beispiel für ein R-Statistik-Projekt mit Aufbereitung, Exploration, Korrelation von Daten

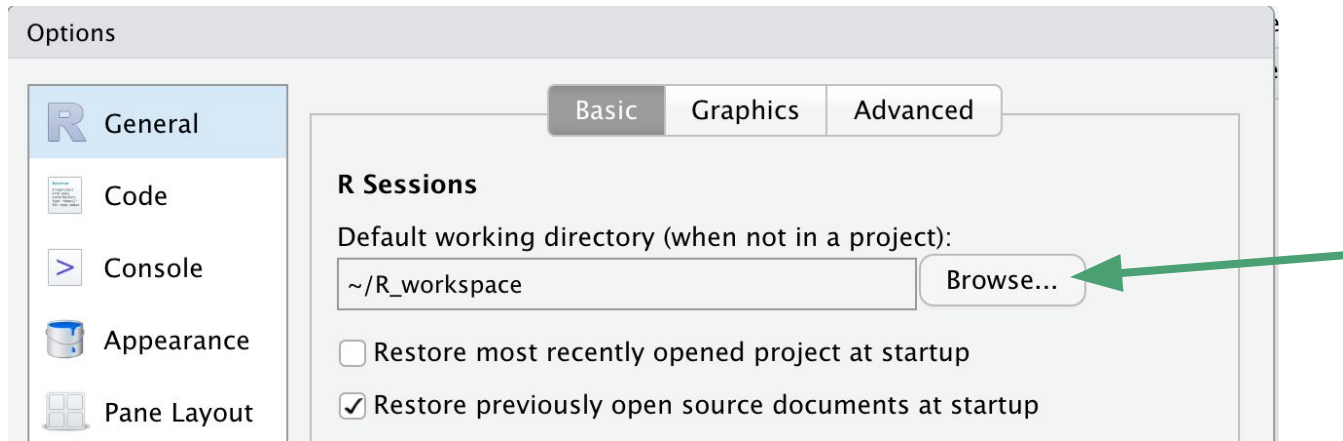
- Projekt R_Covid_Tools – GitHub: https://github.com/hermesj/R_Covid_Tools
- Entstanden im Rahmen einer Übung für unsere MA-Programme zu Thema “Datenanalyse mit R”
- Ziel war die Aufbereitung von RKI-Rohdaten zur Pandemie zur Exploration unterschiedlicher Phasen der Pandemie und zur Untersuchung der Korrelation zwischen Impfquoten, Hospitalisierungen und Todesfällen.

Hausaufgabe

- Installieren Sie sich R in einer Version ab 4.0 auf Ihrem Rechner
→ <https://cran.r-project.org/>
- Installieren Sie sich die Entwicklungsumgebung RStudio auf Ihrem Rechner
→ <https://www.rstudio.com/products/rstudio/download/>
- (optional, machen wir auch in der ersten R-Stunde gemeinsam) Richten Sie RStudio schon einmal so ein, wie es auf den nächsten drei Folien zu sehen ist.
- (optional) Treten Sie dem Online-Kurs “Programmieren mit R für Einsteiger” bei und bearbeiten Sie die Materialien von Woche 0.
→ <https://open.hpi.de/courses/programmieren-r2022/>

Einrichtung RStudio I: Default-Workspace

- Tools → Global Options → General → R Sessions → Default working directory → Browse → Ordner aussuchen.



- Dann landet man bei jedem RStudio-Start genau dort

Einrichtung RStudio II: Reproduzierbarkeit sicherstellen

- Tools → Global Options → General → Workspace
 - Haken entfernen “Restore .RData into workspace at startup”
 - “Save workspace to .RData on exit” auf “Never”

Workspace

Restore .RData into workspace at startup

Save workspace to .RData on exit: Never ▾

Einrichtung RStudio III: UTF-8-Default-Encoding setzen

- File → Reopen with Encoding... → UTF-8 auswählen → “Set as default encoding” auswählen

