

Computerwerkstatt: Umgang mit Daten in R (- Studio)

Dipl.-Psych. Jörg-Henrik Heine

14.11.2012

1 Allgemeines und Vorbereitung

Angedacht ist zunächst eine knappe Einführung in das graphical user Interface (GUI) R-Studio für R. Bei Interesse wird die Funktionalität von R-Studio PDF und HTML Dokumente zu erstellen kurz in groben Zügen vorgestellt. Ein weiterer Hauptpunkt soll in der Präsentation und Übung der Organisation von Daten in R sein. Dazu wird anhand eines Beispieldatensatzes (NEO-PI-R) der komplette Workflow in R, vom einlesen div. Rohdatenformate bis zu Organisation von Meta-Informationen wie Itemwortlaut, Polung, etc., demonstriert. Bei Interesse, können Teilnehmer auch eigene Datensätze mitbringen welche dann entsprechend verarbeitet werden.

Um möglichst wenig Zeit mit dem Download von Software und von noch fehlenden Paketen zu verbrauchen, habe ich im folgenden Abschnitt eine Liste mit der zu installierenden Software aufgestellt - jeweils für MAC und Windows User (vgl. Downloadlinks in Abschnitt 2.1 und 2.2).

2 Software Voraussetzungen MAC und WIN

Mindestvoraussetzung ist eine aktuelle R base Installation und die aktuellste Version von R-Studio. Ferner werden zur Demonstration der Funktionalität von R-Studio einige Erweiterungspakete benötigt. Diese Erweiterungspakete sind als R Paketinstallations-Befehle in Abschnitt 2.3 angegeben und teilweise eher optional, mit zwei Ausnahmen, nämlich dem Paket **plyr** und dem Paket **xlsx**, die beide in direktem Zusammenhang mit der Thematik *Umgang mit Daten in R (-Studio)* stehen. Zur Installation der Pakete können die entsprechenden Befehle in Abschnitt 2.3 einfach aus diesem PDF-Dokument herauskopiert werden und in ein R-Syntaxdokument eingefügt und dann in R ausgeführt werden. Die Installation dieser Erweiterungspakete kann aber (bei bestehender Internetverbindung) auch während des Workshops nachgeholt werden.

Die Installation eines L^AT_EX- Systems und die Installation des Paketes **knitr** stehen im Zusammenhang mit den erweiterten Möglichkeiten von R-Studio, PDF-Dokumente aus L^AT_EX- Dateien zu kompilieren, in denen der R-Code zur

Berechnung von Koeffizienten und/oder zur Erstellung von Grafiken bereits integriert ist. Es wäre von meiner Seite her noch offen ob, und wenn ja wie ausführlich, diese Thematik überhaupt besprochen werden soll. Allerdings stellt es eine elegante Möglichkeit dar, während der Computerwerkstatt sozusagen “mitzuschreiben” und am Ende entweder ein PDF oder ein HTML- Dokument zu haben, in dem alles dokumentiert ist. Nachdem es sich bei der \LaTeX - Installation um einen größeren Download handelt, empfiehlt es sich diese Installation bereits vor dem Workshop auf dem eigenen Laptop vorzunehmen - auch wenn die Thematik, je nach Interesse der Teilnehmer, ggf. nur am Rand besprochen wird. Abgesehen davon schadet eine aktuelle \LaTeX - Installation auf dem Computer auch nicht.

2.1 Software Installation für MAC OSX Nutzer (Version 10.8.2):

- R - Installation (ca. 64MB):
<http://cran.r-project.org/bin/macosx/R-2.15.2.pkg>
- R-Studio Installation (29.9 MB):
<http://download1.rstudio.org/RStudio-0.96.331.dmg>
- \LaTeX - Installation (ca. 2.16 GB):
<http://mirror.ctan.org/systems/mac/mactex/MacTeX.pkg>
- ggf. noch die Extrapakete installieren unter:
<http://mirror.ctan.org/systems/mac/mactex/MacTeXtras.zip>

2.2 Software Installation für Windows Nutzer (Win-XP, Win-7):

- R - Installation (47,1 MB):
<http://cran.r-project.org/>
- R-Studio Installation (32.1 MB):
<http://www.rstudio.com/ide/download/desktop>
- \LaTeX - Installation (1,3 GB):
<http://www.latex-project.org/ftp.html>
- Hinweise zu \LaTeX : Herunterladen und entpacken der Datei und im entstandenen Ordner

\backslash ProTeXt

die Datei Setup.exe ausführen, in der sich öffnenden Dialogbox ggf. die Anleitung öffnen oder gleich MiKTeX installieren und dann TeXstudio installieren (*bitte genau in dieser Reihenfolge*).

2.3 Für MAC OSX und Windows Nutzer: Installations Befehle für R - Pakete (*die teilweise nur eventuell zum Einsatz kommen könnten*)

```
install.packages("xlsx",dependencies=T,repos="http://cran.at.r-project.org")
# Read, write, format Excel 2007 and Excel 97/2000/XP/2003 files

install.packages("plyr",dependencies=T,repos="http://cran.at.r-project.org")
# Tools for splitting, applying and combining data

install.packages("knitr",dependencies=T,repos="http://cran.at.r-project.org")
# für die PDF / Latex und HTML Funktionalität von RStudio

install.packages("mixRasch",dependencies=T,repos="http://cran.at.r-project.org")
# Mixture Rasch Models with JMLE

install.packages("cfa",dependencies=T,repos="http://cran.at.r-project.org")
# Analysis of configuration frequencies (CFA)

install.packages("poLCA",dependencies=T,repos="http://cran.at.r-project.org")
# Polytomous variable Latent Class Analysis

install.packages("e1071",dependencies=T,repos="http://cran.at.r-project.org")
# Misc Functions of the Department of Statistics (e1071), TU Wien

install.packages("PP",dependencies=T,repos="http://cran.at.r-project.org")
# Person Parameter estimation

install.packages("CircStats",dependencies=T,repos="http://cran.at.r-project.org")
# Circular Statistics, from "Topics in circular Statistics" (2001)

install.packages("circular",dependencies=T,repos="http://cran.at.r-project.org")
# Circular Statistics, from "Topics in circular Statistics" (2001)

install.packages("MplusAutomation",dependencies=T,repos="http://cran.at.r-project.org")
# Automating Mplus Model Estimation and Interpretation
```

3 Inhalte und Themen der Computerwerkstatt

Vorstellung der Benutzeroberfläche R-Studio

- Ggf. Nachholen der Installation unter MAC OSX oder Windows
- RStudio und \LaTeX bzw. HTML Dokumente zu Dokumentation des eigenen R-Codes nutzen.
- Organisation und Navigation innerhalb von (umfangreichem) R-Code;
- Was sind R-Codesections, Erstellen von R-Codesections, ...
- Ein- und Ausblenden von R-Code, Springen innerhalb von R-Code.

Umgang mit Daten in R

Vorbereiten und Einlesen von Datensätzen

- Einlesen von einfachen Text-Datenformaten: csv, csv2, table, ...
- Einlesen von .xlsx Excel-Tabellen mit mehreren Tabellenblättern
- Direktes einlesen von SPSS Daten ...

Metainformation zu Datensätzen in R organisieren

- Itemreihenfolge, Zuordnung zu Dimensionen,
- 'Value Labels' und 'Variable Labels' in R ?

Kurzer Überblick zu Objektklassen in R

- Integer, Vektor, Matrix, Array, ...
- Listen, Dataframes, ...

Organisation von Daten in Matrizen in R

- Recodieren,
- Indizieren, Auswahl von Teilstichproben,
- Umsortieren Zeilen und Spaltenweise,
- Variablenamen ändern, Fallbezeichnungen ändern

Einfache eigene Funktionen

- Ganz einfache (Hilfs-) Funktionen
- Anwendung von Funktionen auf Datensätze
- lapply, sapply, mapply
- Funktionsgruppen und do call Aufrufe