

INTRO DATENANALYSE MIT R - DATENIMPORT

Jan-Philipp Kolb

07 Juni, 2019

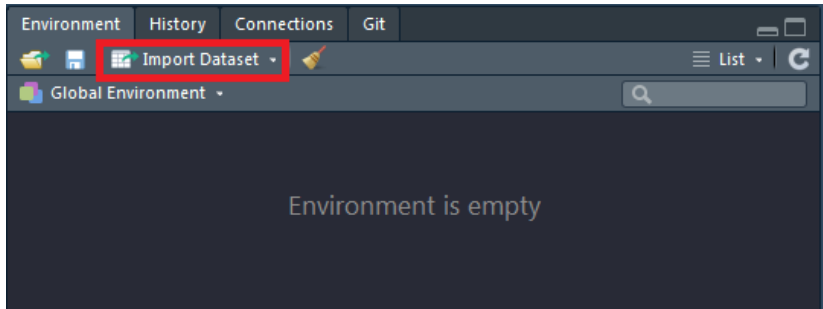
DATENIMPORT



DATEN MIT RSTUDIO IMPORTIEREN

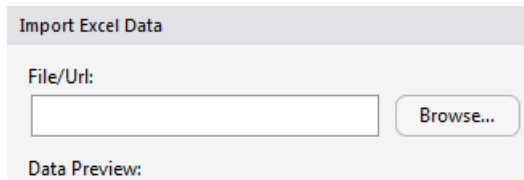
RSTUDIO FUNKTIONALITÄT UM DATEN ZU IMPORTIEREN

- Environment - Import Dataset - Filetyp auswählen



WO FINDET MAN DIE DATEN?

BROWSE BUTTON IN RSTUDIO



Import Excel Data

File/Url:

Browse...

Data Preview:

CODE VORSCHAU IN RSTUDIO



Code Preview:

```
library(readxl)
ee_recode_questionnaire_coded <- read_excel("data/ee_recode_questionnaire_coded.xls")
View(ee_recode_questionnaire_coded)
```

Import Cancel

CSV DATEN IMPORTIEREN

- ▶ `read.csv` ist ein Befehl, der im Basispaket verfügbar ist.
- ▶ Excel-Daten können als `.csv` in Excel gespeichert werden.
- ▶ Dann kann `read.csv()` zum Einlesen der Daten verwendet werden.
- ▶ Für Deutsche Daten benötigt man eventuell `read.csv2()` wegen der Komma-Trennung.

```
dat <- read.csv("../data/ZA5666_v1-0-0.csv")
```

Wenn es Deutsche Daten sind:

```
datd <- read.csv2("../data/ZA5666_v1-0-0.csv")
```

EXCEL-DATENSATZ IMPORTIEREN - MIT XLSX.

PAKET XLSX

- ▶ Titel: Read, Write, Format Excel 2007 and Excel 97/2000/XP/2003 Files
- ▶ Autoren: Adrian A. Dragulescu, Cole Arendt

```
install.packages("xlsx")
```

```
library("xlsx")  
ab_xlsx <- read.xlsx("../data/ab.xlsx",1)
```

- ▶ Das Paket xlsx benötigt Java - wenn das nicht verfügbar ist, verwenden Sie den Befehl `read_excel` aus dem Paket `readxl`.

DAS PAKET READXL

```
install.packages("readxl")
```

- ▶ **readxl hat keine externen Abhängigkeiten**
- ▶ readxl unterstützt sowohl das alte .xls Format als auch das moderne xml-basierte .xlsx Format.

```
library(readxl)  
ab <- read_excel("../data/ab.xlsx")  
head(ab)
```

SPSS DATEIEN EINLESEN

Dateien können auch direkt aus dem Internet geladen werden:

```
link<- "http://www.statistik.at/web_de/static/  
mz_2013_sds_-_datensatz_080469.sav"
```

```
?read.spss
```

```
Dat <- read.spss(link,to.data.frame=T)
```


IMPORTIEREN VON STATA DATEIEN

- ▶ Mit `read.dta13` können Stata-Dateien ab Version 13 (und höher) importiert werden.

```
library(readstata13)
dstat<-read.dta13("../data/ZA5666_v1-0-0_Stata14.dta")
```

IMPORT VON STATA DATEIEN - ÄLTERE VERSIONEN

```
library(foreign)
dst12 <- read.dta("../data/ZA5666_v1-0-0_Stata12.dta")
```

- ▶ Einführung in den Import mit R (**is.R**)

DIE BIBLIOTHEK READSTATA13

```
readstata13 {readstata13}
```

R Documentation

Import Stata Data Files

Description

Function to read the Stata file format into a data.frame.

Note

If you catch a bug, please do not sue us, we do not have any money.

Author(s)

Marvin Garbuszus jan.garbuszus@ruhr-uni-bochum.de

Sebastian Jeworutzki sebastian.jeworutzki@ruhr-uni-bochum.de

See Also

[read.dta](#) and [memisc](#) for dta files from Stata Versions < 13

Die Bibliothek rio

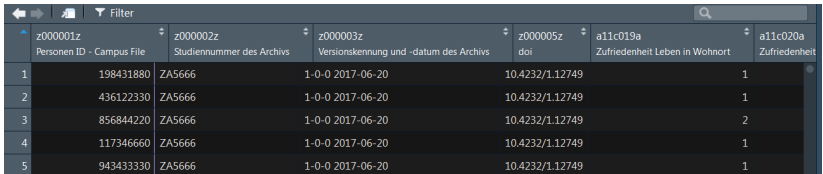
```
install.packages("rio")
```

```
library("rio")  
x <- import("../data/ZA5666_v1-0-0.csv")  
y <- import("../data/ZA5666_v1-0-0_Stata12.dta")  
z <- import("../data/ZA5666_v1-0-0_Stata14.dta")
```

- **rio: Ein Schweizer Offiziersmesser für Data I/O**

SICH EINEN ERSTEN ÜBERBLICK VERSCHAFFEN

View(datf)



The screenshot shows a data viewer interface with a table. The table has 6 columns and 5 rows of data. The columns are labeled as follows:

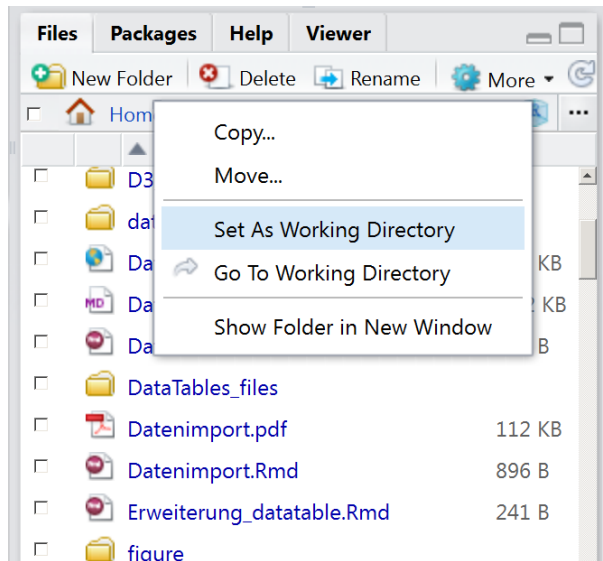
- z000001z: Personen ID - Campus File
- z000002z: Studiennummer des Archivs
- z000003z: Versionskennung und -datum des Archivs
- z000005z: doi
- a11c019a: Zufriedenheit Leben in Wohnort
- a11c020a: Zufriedenheit

The data rows are as follows:

	z000001z Personen ID - Campus File	z000002z Studiennummer des Archivs	z000003z Versionskennung und -datum des Archivs	z000005z doi	a11c019a Zufriedenheit Leben in Wohnort	a11c020a Zufriedenheit
1	198431880	ZA5666	1-0-0 2017-06-20	10.4232/1.12749		1
2	436122330	ZA5666	1-0-0 2017-06-20	10.4232/1.12749		1
3	856844220	ZA5666	1-0-0 2017-06-20	10.4232/1.12749		2
4	117346660	ZA5666	1-0-0 2017-06-20	10.4232/1.12749		1
5	943433330	ZA5666	1-0-0 2017-06-20	10.4232/1.12749		1

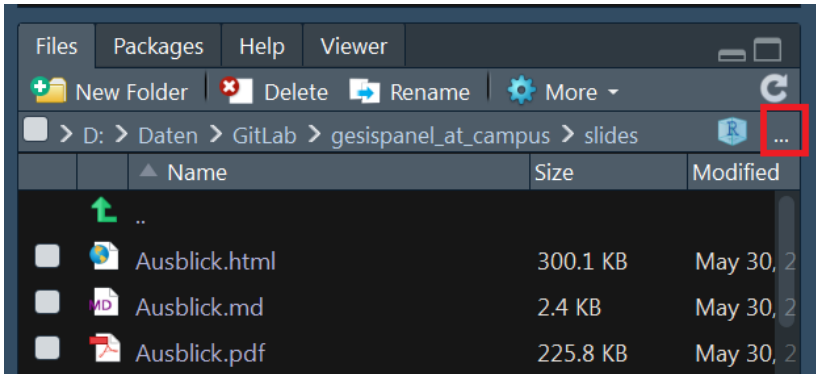
- Das gleiche kann man mit RStudio erreichen, wenn man auf das Datensatzsymbol im Umgebungsmenü klickt.

DAS ARBEITSVERZEICHNIS



...

- ▶ Wenn sich die Daten auf einem anderen Laufwerk in Windows befinden



INGEBAUTE DATENSÄTZE

- ▶ Häufig wird ein Beispieldatensatz zur Verfügung gestellt, um die Funktionalität eines Pakets zu zeigen.
- ▶ Diese Datensätze können mit dem Befehl `data` geladen werden.

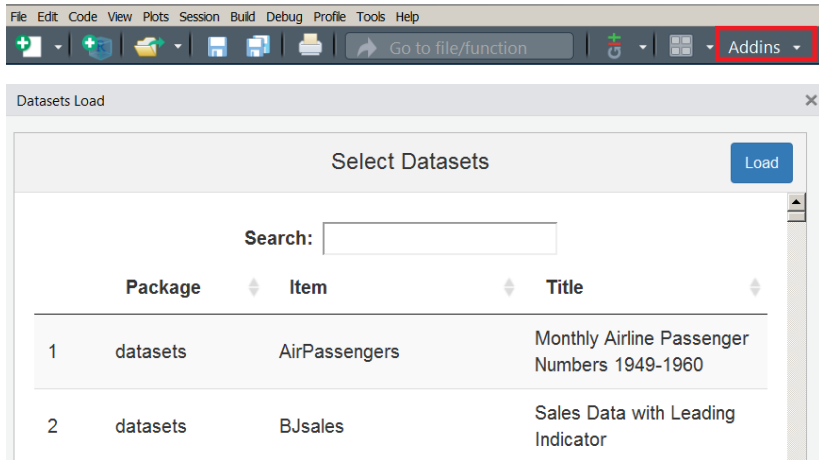
```
data(iris)
```

- ▶ Es gibt auch ein **RStudio-Add-In**, das hilft, einen Datensatz zu finden.

```
install.packages("datasets.load")
```

EXKURS RSTUDIO ADDINS

- Oben rechts befindet sich ein Button Addins



Daten einfügen

► RStudio Addin um Daten einzufügen

```
devtools::install_github("lbusett/insert_table")
```

Insert Table Add-In

Cancel Select output format and edit the Table if you wish so Done

Select Table Name Select Output Format

my_tbl None

Edit Table or cut and paste from spreadsheet

* The first row will be used as column names.
* Right click to add more lines or columns

☒ Use first row as column names. (If unchecked, 'Col_1', 'Col_2', etc. are used)

1	a	v		c
2				
3				
4				

ÜBUNG - IMPORTIEREN VON DATEN

- ▶ Importiere die Daten des österreichischen Mikrozensus und verschaffe Dir einen ersten Überblick über die Daten.