

Datenimport

Jan-Philipp Kolb

12 Juni 2017

Datenimport



Figure 1:

Dateiformate in R

- ▶ Von R werden quelloffene, nicht-proprietäre Formate bevorzugt
- ▶ Es können aber auch Formate von anderen Statistik Software Paketen eingelesen werden
- ▶ R-user speichern Objekte gerne in sog. Workspaces ab
- ▶ Auch hier jedoch gilt: (fast) alles andere ist möglich

Formate - base package

R unterstützt von Haus aus schon einige wichtige Formate:

- ▶ CSV (Comma Separated Values): `read.csv()`
- ▶ FWF (Fixed With Format): `read.fwf()`
- ▶ Tab-getrennte Werte: `read.delim()`

Datenimport leicht gemacht mit Rstudio

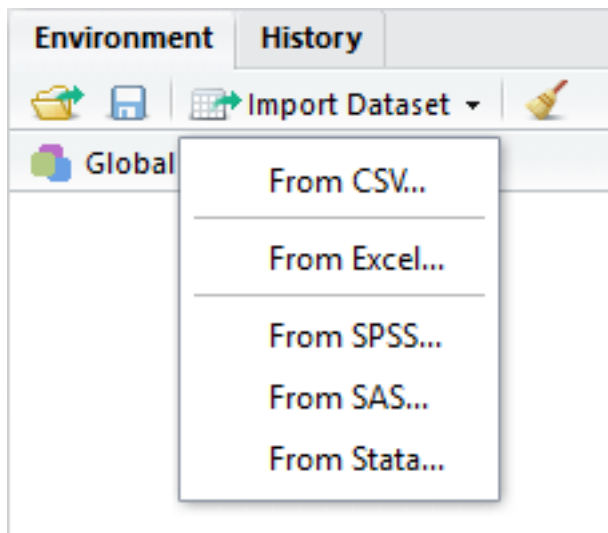


Figure 2: Import Button

Der Arbeitsspeicher

So findet man heraus, in welchem Verzeichnis man sich gerade befindet

```
getwd()
```

Und ändert dann den Pfad mit `setwd()`

```
setwd("C:/")
```

Man erzeugt ein Objekt in dem man den Pfad abspeichert:

```
main.path <- "C:/" # Beispiel für Windows  
main.path <- "/users/Name/" # Beispiel für Mac  
main.path <- "/home/user/" # Beispiel für Linux
```

Bei Windows ist es wichtig Slashes anstelle von Backslashes zu verwenden.

Alternative - Arbeitsspeicher

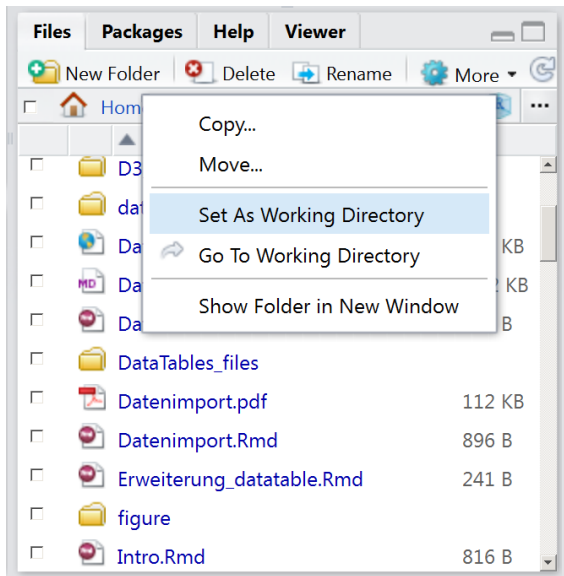


Figure 3:

Import von Excel-Daten

- ▶ `library(foreign)` ist für den Import von fremden Datenformaten nötig
- ▶ Wenn Excel-Daten vorliegen - als .csv abspeichern
- ▶ Dann kann `read.csv()` genutzt werden um die Daten einzulesen.
- ▶ Bei Deutschen Daten kann es sein, dass man `read.csv2()` wegen der Komma-Separierung braucht.

```
library(foreign)  
?read.csv  
?read.csv2
```


CSV Dateien einlesen

Zunächst muss das Arbeitsverzeichnis gesetzt werden, in dem sich die Daten befinden:

```
Dat <- read.csv("schuldaten_export.csv")
```

Wenn es sich um Deutsche Daten handelt:

```
Dat <- read.csv2("schuldaten_export.csv")
```

CSV aus dem Web einladen

► Datensatz:

<https://data.montgomerycountymd.gov/api/views/6rqk-pdub/rows.csv?accessType=DOWNLOAD>

► Datenimport mit Rstudio

Import Text Data

File/Url
<https://data.montgomerycountymd.gov/api/views/6rqk-pdub/rows.csv?accessType=DOWNLOAD> Update

Data Preview

Full Name (character) *	Gender (character) *	Current Annual Salary (character) *	2015 Gross Pay Received (character) *	2015 Overtime Pay (character) *	Department (character) *	Department Name (character) *	Division (character) *	Assignment Category (character) *	Position Title (character) *	Underfilled Job Title (character)
Ambas, Pam J.	F	\$68676.16	\$72356.79	N/A	POL	Department of Police	MSE Information Management and Technology Division	Falltime-Regular	Office Services Coordinator	N/A
Aaron, David J.	M	\$90908.09	\$101857.00	\$4640.99	POL	Department of Police	158 Major Crimes Division Negative Section	Falltime-Regular	Master Police Officer	N/A
Aaron, Marisha M.	F	\$104136.06	\$103019.73	N/A	HMS	Department of Health and Human Services	Adult Protective and Case Management Services	Falltime-Regular	Social Worker IV	N/A
Ababio, Godfred A.	M	\$50697.79	\$54181.46	\$6445.15	COR	Correction and Rehabilitation	PRSS Facility and Security	Falltime-Regular	Resident Supervisor II	N/A
Ababa, Esayas	M	\$92931.00	\$92468.35	N/A	HCA	Department of Housing and Community Affairs	Single Family Housing Program	Falltime-Regular	Planning Specialist III	N/A
Abdennasser, Drew B.	M	\$67715.00	\$81392.40	\$10027.11	POL	Department of Police	PSS 6th District Special Assignment Team	Falltime-Regular	Police Officer III	N/A
Abdennasser, Marwan A.	M	\$62298.30	\$59063.27	N/A	HMS	Department of Health and Human Services	Head Start	Falltime-Regular	Administrative Specialist II	N/A
Abdul Chani, Haseebah J.	F	\$45828.92	\$46793.23	\$6.38	POL	Department of Police	PSS Traffic Division Automated Traffic Enforcement S...	Falltime-Regular	Police Aide	N/A
Abduljabbar, Saqad	M	\$61040.57	\$66861.98	\$6589.81	DCS	Department of General Services	Facilities Maintenance	Falltime-Regular	Electrician I	N/A
Abdur-Rahman, Mikael A.	M	\$56404.96	\$71943.08	\$15342.84	DOT	Department of Transportation	Transit Silver Spring Ride On	Falltime-Regular	Bus Operator	N/A
Abdie, Huzefa	F	\$151585.60	\$168845.06	N/A	HMS	Department of Health and Human Services	STD and HIV Services	Parttime-Regular	Medical Doctor (IV) - Physician	N/A
Abdie, Zelmanus S.	M	\$44825.99	\$51693.47	\$2440.75	DOT	Department of Transportation	Transit Nicholson Ride On	Falltime-Regular	Bus Operator	N/A
Abdulin, Amrozeah	M	\$39062.00	\$450.00	N/A	DOT	Department of Transportation	Transportation Management	Falltime-Regular	Traffic Management Technician II	Traffic Management Technician
Abdine, Sherry R.	F	\$93436.50	\$90833.10	N/A	HMS	Department of Health and Human Services	Adult Protective and Case Management Services	Falltime-Regular	Social Worker III	N/A
Abrera, Joseph M.	M	\$117811.00	\$115796.22	N/A	OTS	Department of Technology Services	EASD - ERP Applications Support	Falltime-Regular	Senior Information Technology Specialist	N/A
Abu-Jomas, Rania F.	F	\$53009.99	\$46850.16	N/A	LIB	Department of Public Libraries	Oleary Library	Falltime-Regular	Library Assistant I	N/A
Abumano, Ayan Z.	M	\$17268.00	\$14665.53	N/A	LIB	Department of Public Libraries	Silver Spring Library	Parttime-Regular	Library Desk Assistant	N/A
Abus, Lynda B.	F	\$40429.58	\$41746.34	\$2500.33	DOT	Department of Transportation	Transit Catonsville Ride On	Falltime-Regular	Bus Operator	N/A
Abuwan, Maral	F	\$20925.31	\$9978.52	\$45.28	POL	Department of Police	PSS Traffic Division School Safety Section	Parttime-Regular	Crossing Guard	N/A
Aburayes, Nadia L.	F	\$16602.01	\$15592.92	N/A	HMS	Department of Health and Human Services	Community Support Network for People with Disabilities	Parttime-Regular	Office Clerk	N/A

Previewing first 50 entries.

Import Options

Name: rows ☒ First Row as Names ☒ Trim Spaces ☒ Open Data Viewer

Delimiter: Comma ☒ Quotes: Default ☒ Comment: Default ☒ Local: NA: Default

Code Preview

```
library(readr)
rows <- read_csv("https://data.montgomerycountymd.gov/api/views/6rqk-pdub/rows.csv?accessType=DOWNLOAD")
view(rows)
```

Import Cancel

Das Paket readr

```
install.packages("readr")
```

```
library(readr)
```

- ▶ [readr auf dem Rstudio Blogg](#)

Import von Excel-Daten

- ▶ `library(readr)` ist für den Import von fremden Datenformaten hilfreich
- ▶ Wenn Excel-Daten vorliegen - als .csv abspeichern

```
url <- "https://raw.githubusercontent.com/Japhilko/  
GeoData/master/2015/data/whcSites.csv"
```

```
whcSites <- read.csv(url)
```

Der Beispieldatensatz

```
head(data.frame(whcSites$name_en,whcSites$category))
```

```
##                               whcSites.name_en
## 1 Cultural Landscape and Archaeological Remains of the
## 2                               Minaret and Archaeological
## 3                               Historic Centres of Berat and
## 4
## 5                               Al Qal'a o
## 6
##   whcSites.category
## 1           Cultural
## 2           Cultural
## 3           Cultural
## 4           Cultural
## 5           Cultural
## 6           Cultural
```

Das Paket haven

Import and Export 'SPSS', 'Stata' and 'SAS' Files

```
install.packages("haven")
```

```
library(haven)
```

- ▶ Das R-Paket haven auf dem Rstudio Blogg

SPSS Dateien einlesen

- ▶ Zunächst muss wieder der Pfad zum Arbeitsverzeichnis angegeben werden.
- ▶ SPSS-Dateien können auch direkt aus dem Internet geladen werden:

```
library(haven)
mtcars <- read_sav(
  "https://github.com/Japhilko/RInterfaces/raw/master/
  data/mtcars.sav")
```

foreign kann ebenfalls zum Import genutzt werden

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

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


stata Dateien einlesen

```
library(haven)
oecd <- read_dta("https://github.com/Japhilko/IntroR/
                raw/master/2017/data/oecd.dta")
```

- ▶ Einführung in Import mit R (is.R)

Das Paket readstata13

```
# install.packages("readstata13")
```

```
library(readstata13)  
?read.dta13
```

Das Paket rio

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

```
library("rio")  
x <- import("mtcars.csv")  
y <- import("mtcars.rds")  
z <- import("mtcars.dta")
```

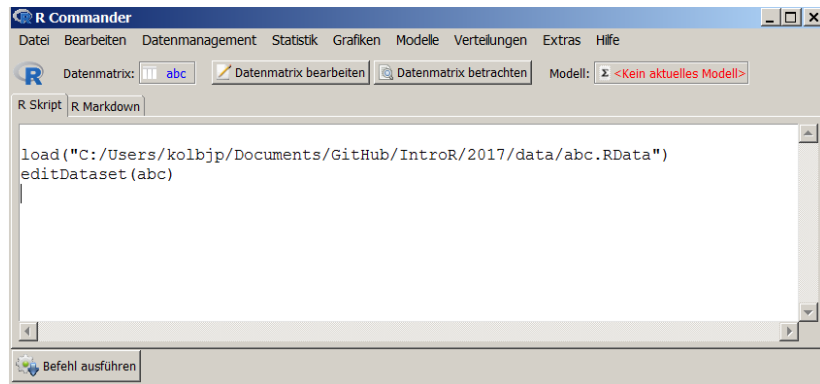
- rio: A Swiss-Army Knife for Data I/O

Weitere Alternative Rcmdr

```
install.packages("Rcmdr")
```

- Funktioniert auch mit Rstudio

```
library(Rcmdr)
```



Die Gesis Panel Daten



Figure 6:

- ▶ Gesis Panel Campus File
- ▶ Codebuch für die Gesis Panel Daten im DBK

Aufgabe

- ▶ Gehen Sie auf meine Github Seite

https:

`//github.com/Japhilko/RSocialScience/tree/master/data`

- ▶ Importieren Sie den Datensatz `GPanel.dta`