

Datenimport

Jan-Philipp Kolb

8 Mai 2017

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

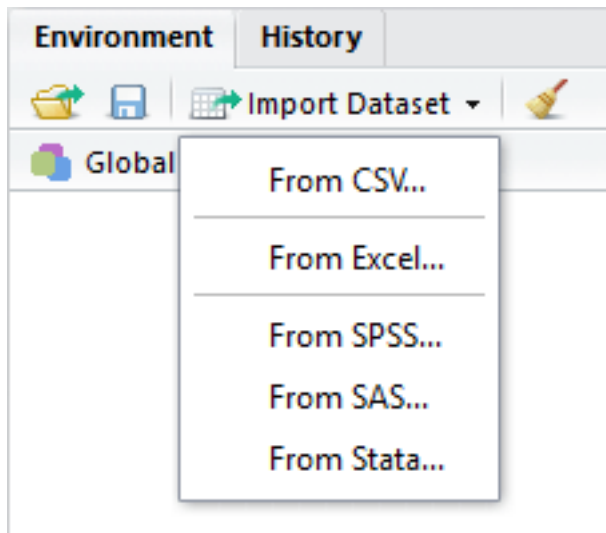


Figure 1: Import Button

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

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) *
Aarbus, Pam J.	F	\$68878.16	\$72336.79	NA	POL	Department of Police	MSB Information Management and Technology Division	Fulltime-Regular	Office Services Coordinator	NA
Aaron, David J.	M	\$96908.09	\$101857.00	\$4640.99	POL	Department of Police	ISB Major Crimes Division Fugitive Section	Fulltime-Regular	Master Police Officer	NA
Aaron, Marsha M.	F	\$104196.06	\$103019.73	NA	HHS	Department of Health and Human Services	Adult Protective and Case Management Services	Fulltime-Regular	Social Worker IV	NA
Alabio, Coffred A.	M	\$50697.79	\$54181.46	\$4445.15	COR	Correction and Rehabilitation	PRRS Facility and Security	Fulltime-Regular	Resident Supervisor II	NA
Alabio, Essayas	M	\$92931.00	\$93468.35	NA	HCA	Department of Housing and Community Affairs	Single Family Housing Program	Fulltime-Regular	Planning Specialist III	NA
Abdmonem, Marwan M.	M	\$67715.00	\$81392.40	\$10027.11	POL	Department of Police	FSB 6th District Special Assignment Team	Fulltime-Regular	Police Officer III	NA
AbdulChani, Haseenah J.	F	\$45828.92	\$46783.23	\$6.38	POL	Department of Police	Head Start	Fulltime-Regular	Administrative Specialist II	NA
Abduljabbar, Saeed	M	\$61040.57	\$66861.98	\$6569.81	DCS	Department of General Services	FSB Traffic Division Automated Traffic Enforcement S...	Fulltime-Regular	Police Aide	NA
Abdur-Rahman, Mikael A.	M	\$56404.96	\$71943.08	\$15342.84	DOT	Department of Transportation	Facilities Maintenance	Fulltime-Regular	Electrician I	NA
Abelie, Hivuth	F	\$151585.60	\$164945.06	NA	HHS	Department of Health and Human Services	STD and HIV Services	Fulltime-Regular	Bus Operator	NA
Abelie, Zekarias S.	M	\$44825.99	\$51693.47	\$5240.75	DOT	Department of Transportation	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abdulin, Amirreza	M	\$39062.00	\$450.00	NA	DOT	Department of Transportation	Transit Silver Spring Ride On	Fulltime-Regular	Bus Operator	NA
Abdulin, Sherry R.	F	\$93436.50	\$90823.10	NA	HHS	Department of Health and Human Services	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abena, Yoseph M.	M	\$117811.00	\$115786.22	NA	DTS	Department of Technology Services	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abi Jamaa, Rania F.	F	\$53009.99	\$46850.36	NA	LIB	Department of Public Libraries	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abijomas, Ryan Z.	M	\$71288.00	\$14665.53	NA	LIB	Department of Public Libraries	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abitu, Lydia B.	F	\$40429.58	\$41746.34	\$3500.33	DOT	Department of Transportation	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abikarim, Maral	F	\$20925.51	\$9976.52	\$45.28	POL	Department of Police	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA
Abouraya, Nadia L.	F	\$16602.01	\$15592.92	NA	HHS	Department of Health and Human Services	Transit Silver Spring Ride On	Fulltime-Regular	Medical Doctor III - Physician	NA

Previewing first 50 entries

Der Arbeitsspeicher

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

```
getwd()
```

So kann man das Arbeitsverzeichnis ändern:

Man erzeugt ein Objekt in dem man den Pfad abspeichert:

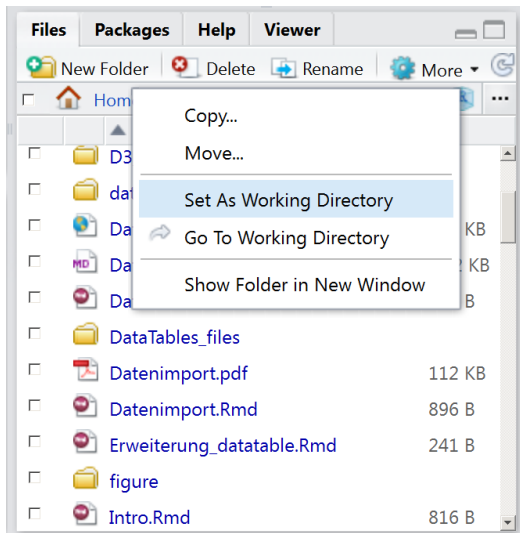
```
main.path <- "C:/" # Beispiel für Windows  
main.path <- "/users/Name/" # Beispiel für Mac  
main.path <- "/home/user/" # Beispiel für Linux
```

Und ändert dann den Pfad mit `setwd()`

```
setwd(main.path)
```

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

Alternative - Arbeitsspeicher



Das Paket readr

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

```
library(readr)
```

- readr auf dem Rstudio Blogg



readr 0.2.0

October 28, 2015 in [Packages](#), [tidyverse](#)

readr 0.2.0 is now available on CRAN. readr makes it easy to read many types of tabular

Import von Excel-Daten

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

```
library(readr)
rows <- read_csv("https://data.montgomerycountymd.gov/api/view
```

.csv-Daten aus dem Web importieren - zweites Beispiel

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

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

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

```
##                                     whcSit  
## 1 Cultural Landscape and Archaeological Remains of the Bami  
## 2                                     Minaret and Archaeological Rema  
## 3                                     Historic Centres of Berat and G  
## 4  
## 5                                     Al Qal'a of B  
## 6                                     M  
##   whcSites.category  
## 1 Cultural
```

Das Paket haven

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

```
library(haven)
```

- Das R-Paket haven auf dem Rstudio Blogg



haven 1.0.0

October 4, 2016 in [Packages](#), [tidyverse](#)

I'm pleased to announce the release of haven. Haven is designed to facilitate the transfer

SPSS Dateien einlesen

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

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

```
library(haven)
```

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

stata Dateien einlesen

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

- Quick-R - Import Data
- Datenimport bei R-bloggers
- Importing Data into R
- Mapping von Arbeitslosendaten in den USA
- Das Paket readr