

Daten verbinden

Jan-Philipp Kolb

22 Februar 2017

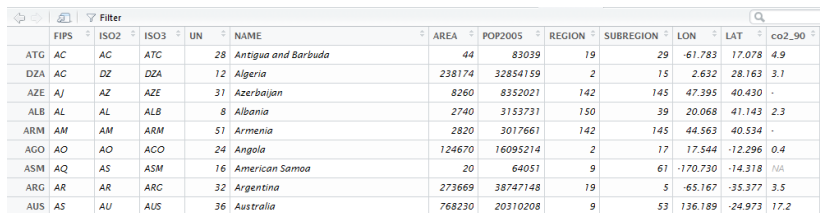
Die wrld_simpl Daten

```
library("maptools")  
data("wrld_simpl")
```

Datensatz in die Daten-Registerkarte von Rstudio schreiben

```
df_ws <- data.frame(wrld_simpl@data)
```

Jetzt können wir den Rstudio Daten-Browser verwenden



The screenshot shows the Rstudio Data Browser interface. At the top, there is a search bar and a 'Filter' button. Below this is a table with 12 columns: FIPS, ISO2, ISO3, UN, NAME, AREA, POP2005, REGION, SUBREGION, LON, LAT, and co2_90. The table contains 10 rows of data for various countries, including Antigua and Barbuda, Algeria, Azerbaijan, Albania, Armenia, Angola, American Samoa, Argentina, and Australia. Each cell in the table has a small downward arrow icon next to it, indicating that the data is interactive and can be filtered or sorted.

	FIPS	ISO2	ISO3	UN	NAME	AREA	POP2005	REGION	SUBREGION	LON	LAT	co2_90
ATG	AC	AC	ATC	28	Antigua and Barbuda	44	83039	19	29	-61.783	17.078	4.9
DZA	AC	DZ	DZA	12	Algeria	238174	32854159	2	15	2.632	28.163	3.1
AZE	AJ	AZ	AZE	31	Azerbaijan	8260	8352021	142	145	47.395	40.430	.
ALB	AL	AL	ALB	8	Albania	2740	3153731	150	39	20.068	41.143	2.3
ARM	AM	AM	ARM	51	Armenia	2820	3017661	142	145	44.563	40.534	.
AGO	AO	AO	ACO	24	Angola	124670	16095214	2	17	17.544	-12.296	0.4
ASM	AQ	AS	ASM	16	American Samoa	20	64051	9	61	-170.730	-14.318	NA
ARG	AR	AR	ARC	32	Argentina	273669	38747148	19	5	-65.167	-35.377	3.5
AUS	AS	AU	AUS	36	Australia	768230	20310208	9	53	136.189	-24.973	17.2

Figure 1: Rstudio Daten Browser

Der wrld_simpl Datensatz

Holen Sie sich die Ländernamen:

```
CNames <- wrld_simpl@data$NAME  
head(CNames)
```

```
## [1] Antigua and Barbuda Algeria Azerbaijan  
## [4] Albania Armenia Angola  
## 246 Levels: Aaland Islands Afghanistan Albania Algeria
```

```
CNames <- as.character(CNames)  
head(CNames)
```

```
## [1] "Antigua and Barbuda" "Algeria" "Azerbaijan"  
## [4] "Albania" "Armenia" "Angola"
```

Substring eines Zeichen-Vektors

```
CNames1 <- substr(CNames,1,1)  
head(CNames1)
```

```
## [1] "A" "A" "A" "A" "A" "A"
```

```
CNames2 <- substr(CNames,1,2)  
head(CNames2)
```

```
## [1] "An" "Al" "Az" "Al" "Ar" "An"
```

Auswahl vornehmen

```
CNames[CNames2=="An"]
```

```
## [1] "Antigua and Barbuda" "Angola" "Anguilla"  
## [4] "Andorra" "Antarctica"
```

CO2 Emissionen

```
link <-  
"https://raw.githubusercontent.com/Japhilko/  
GeoData/master/data/CO2emissions.csv"  
co2 <- read.csv(link)
```

X	V1	V2	V3	V4	V5	V6	V7
1	1.	Qatar	25.2	36.7	54.3	60.9	58.7
2	2.	Trinidad and Tobago	13.9	17.1	17.0	13.5	15.8
3	3.	Netherlands Antilles	32.6	26.9	22.6	35.0	34.3
4	4.	Kuwait	19.0	5.1	10.0	16.9	20.8

Wir müssen Länder in diesem Datensatz und Ländernamen in wrld_simpl-Datensatz zusammenbringen

Vektoren zum Matching

Wie bringt man zwei Vektoren zusammen:

```
A <- c(1,2,3,4)
B <- c(4,3)
match(A,B)
```

```
## [1] NA NA 2 1
```

```
match(B,A)
```

```
## [1] 4 3
```


Vektoren zum Matching

```
D <- c(1,3,5,6,7)
E <- c("A",1,98,4)
match(D,E)
```

```
## [1]  2 NA NA NA NA
```

- Matching mit Ländernamen, um eine Karte mit CO2 Indikatoren zu produzieren

Matching

```
ind <- match(wrld_simpl@data$NAME,co2$V2)
ind
```

##	[1]	80	97	73	141	137	142	NA	84	11	8	77	69	6
##	[18]	NA	172	179	126	62	NA	14	181	171	NA	215	213	5
##	[35]	182	214	189	145	130	199	116	165	104	35	45	166	13
##	[52]	56	22	200	151	201	49	26	96	31	133	NA	NA	10
##	[69]	178	120	29	38	NA	41	154	202	125	192	150	79	7
##	[86]	53	NA	105	39	92	101	184	152	NA	185	NA	4	1
##	[103]	90	193	61	NA	NA	203	85	98	20	NA	216	138	11
##	[120]	93	52	205	204	21	115	207	7	102	36	71	NA	1
##	[137]	111	NA	170	NA	NA	NA	NA	NA	NA	NA	NA	NA	M
##	[154]	28	32	206	18	87	160	51	164	140	157	46	123	7
##	[171]	94	NA	161	NA	23	208	13	82	48	42	NA	114	17
##	[188]	54	121	188	76	NA	74	2	99	153	NA	135	196	M
##	[205]	NA	210	43	59	12	198	128	89	127	68	91	NA	M
##	[222]	155	197	163	132	86	3	NA	195	NA	33	134	24	M

Struktur der Daten

```
co2vec<-co2$V3  
str(co2vec)
```

```
## Factor w/ 90 levels "", "-", "0", "0.1", ...: 48 28 57 37 47
```

```
co2vec<-as.character(co2vec)  
str(co2vec)
```

```
## chr [1:219] "25.2" "13.9" "32.6" "19.0" "25.0" "29.4"
```

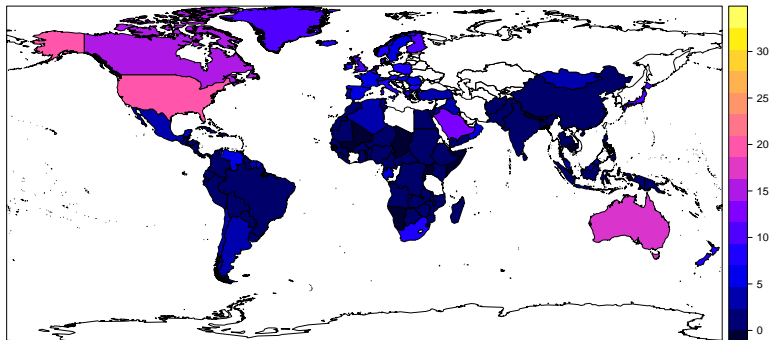
```
co2vec<-as.numeric(co2vec)  
str(co2vec)
```

```
## num [1:219] 25.2 13.9 32.6 19 25 29.4 29.1 24.1 25.9 18
```

Daten anspielen

```
wrld_simpl@data$co2_90 <- co2vec[ind]
```

```
library(sp)  
spplot(wrld_simpl, "co2_90")
```



Zusätzliche Länder matchen

```
ind2 <- match(co2$V2, wrld_simpl@data$NAME)
fehlt <- co2$V2[is.na(ind2)]
fehlt
```

```
## [1] Brunei
## [2] United Arab Emirates
## [3] Falkland Islands
## [4] South Korea
## [5] Taiwan[4] [5]
## [6] Libya
## [7] European Union
## [8] Iran
## [9] Macedonia
## [10] World
## [11] Réunion
## [12] Syria
## [13] North Korea
```

Matching mit agrep

```
Namen_ws <- as.character(wrld_simpl@data$NAME)
Namen_co2 <- as.character(co2$Country)
for (i in 1:length(ind)){
  if(is.na(ind[i])){
    ind4 <- agrep(Namen_ws[i],Namen_co2)
    if(length(ind4)==1){
      ind[i] <- ind4
    }
  }
}
```

Daten anspielen

```
wrld_simpl@data$co2_91 <- co2vec[ind]
```

```
spplot(wrld_simpl,"co2_91")
```

