

Matching

Jan-Philipp Kolb

Wed Oct 14 10:22:30 2015

The wrld_simpl data set

```
library(maptools)
```

```
## Loading required package: sp  
## Checking rgeos availability: TRUE
```

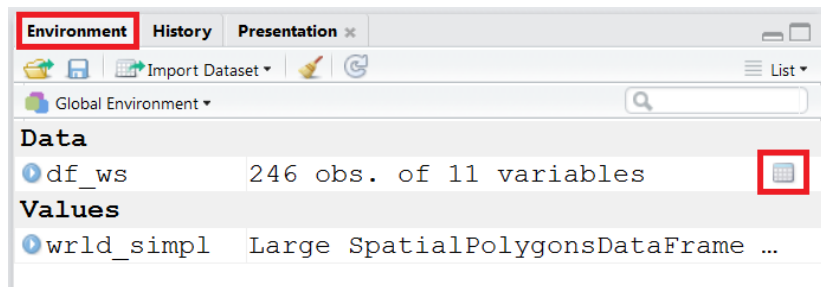
```
data("wrld_simpl")
```

Get the dataset to the environment tab of Rstudio

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

Rstudio Data Browser

Now you can use the Rstudio data browser



The screenshot shows the RStudio Environment pane with the 'Environment' tab selected. The pane displays the 'Global Environment' and lists two objects: 'df_ws' and 'wrld_simpl'. The 'df_ws' object is highlighted, and its details are shown below it. The 'Values' section shows 'wrld_simpl' as a 'Large SpatialPolygonsDataFrame ...'. A red box highlights the 'Environment' tab, and another red box highlights the data browser icon (a small grid) next to the 'df_ws' object.

Environment	
df_ws	246 obs. of 11 variables
wrld_simpl	Large SpatialPolygonsDataFrame ...

The wrld_simpl data set

Get the country names:

```
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"
```

Substrings of a Character Vector

```
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"
```

Subsetting

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

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

Why matching?

You can download a statistic on housing saving rate from Eurostat
link

```
## Loading required package: rJava
```

```
## Loading required package: xlsxjars
```

geo	X2012Q3	X2012Q4	X2013Q1
Euro area (19 countries)	9.82	11.86	11.37
EU (28 countries)	8.67	10.92	9.42
Belgium	12.52	9.33	13.99
Czech Republic	10.16	14.81	9.46
Denmark	7.04	2.56	10.15
Germany	15.06	14.77	19.45

Is it possible to match countries in this dataset and country name in
wrld_simpl dataset?

Matching vectors

How to match two vectors

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

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

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

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


Matching vectors

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

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

- ▶ Let's do the matching with country names to produce a map with economic indicators

The wrld_simpl data

We only need the European continent

```
EUR <- wrld_simpl[wrld_simpl$REGION==150,]  
plot(EUR)
```



Europe without Russia

- ▶ Russia is too big for our map:

```
EUR <- EUR[-which(EUR@data$NAME=="Russia"),]  
plot(EUR)
```



Matching Countries

```
as.character(HHsr[5:10,1])
```

```
## [1] "Denmark" "Germany" "Ireland" "Spain" "France" "Czech Republic"
```

```
EUR_Names <- as.character(EUR@data$NAME)  
head(EUR_Names)
```

```
## [1] "Albania" "Bosnia and Herzegovina"  
## [3] "Bulgaria" "Denmark"  
## [5] "Ireland" "Estonia"
```

```
ind <- match(EUR_Names, HHsr[,1])  
head(ind)
```

```
## [1] NA NA NA 5 7 NA
```

```
table(is.na(ind))
```

Add the data to the polygon data

```
EUR@data$HHsr <- as.numeric(as.character(HHsr[ind,2]))
```

```
# Plot the map:
```

```
library(sp)
```

```
spplot(EUR, "HHsr")
```

