

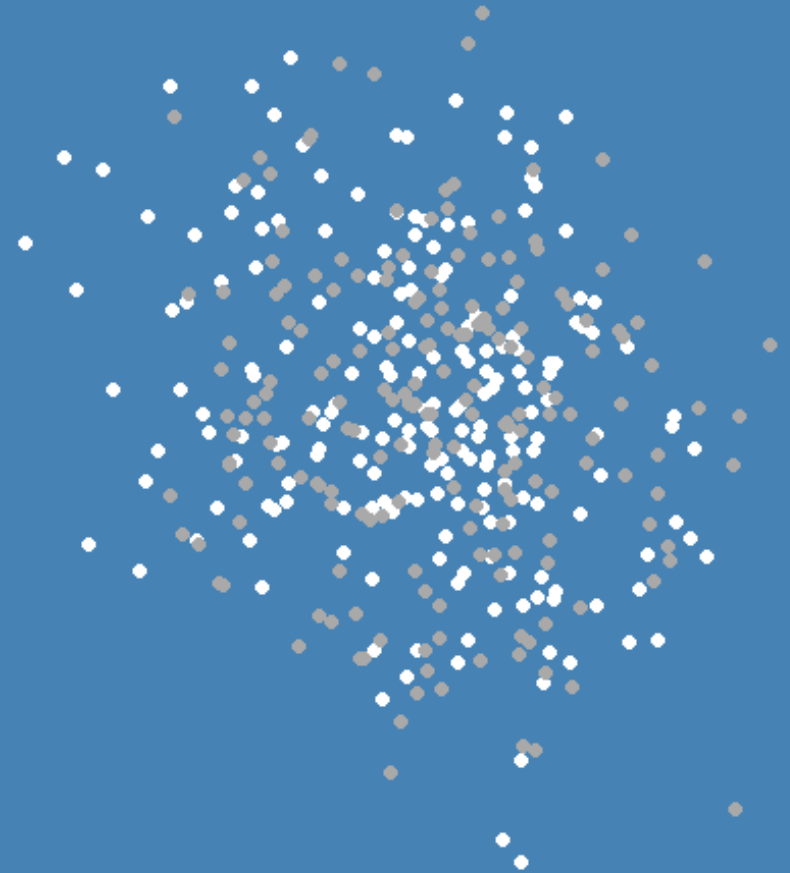
Introduction to R

2.7 Merging Data Frames

Lion Behrens, M.Sc.



University of Mannheim
Chair of Social Data Science and Methodology
Chair of Quantitative Methods in the Social
Sciences



Defining the Issue

Let's say you are collecting data on your own, and you compile [observational data from various sources](#).

One dataset might be on countries' [gdp per capita](#).

```
gdp_data <- read.csv2("dat/gdp_data.csv")
```

```
print(gdp_data)
```

```
##           country gdp_capita
## 1 United States    59.939
## 2          China     8.612
## 3          Japan    38.214
## 4          Germany  44.680
## 5           India    1.980
```

Defining the Issue

Let's say you are collecting data on your own, and you compile [observational data from various sources](#).

The other dataset might be on countries' [freedom house scores](#).

```
fh_data <- read.csv2("dat/fh_data.csv")
```

```
print(fh_data)
```

```
##           country fh
## 1 United States 83
## 2           China  9
## 3           Japan 96
## 4          Germany 94
## 5           India 66
```

Merging Data Frames

How to merge these two data frames?

```
print(gdp_data)
```

```
##           country gdp_capita
## 1 United States    59.939
## 2           China     8.612
## 3           Japan    38.214
## 4           Germany  44.680
## 5           India     1.980
```

```
print(fh_data)
```

```
##           country fh
## 1 United States  83
## 2           China   9
## 3           Japan  96
## 4           Germany 94
## 5           India  66
```

Merging Data Frames

Simply use the `cbind()` function.

```
data <- as.data.frame(  
  cbind(gdp_data, fh_data)  
)
```

```
print(data)
```

```
##           country gdp_capita           country fh  
## 1 United States    59.939 United States 83  
## 2           China     8.612           China  9  
## 3           Japan    38.214           Japan 96  
## 4          Germany    44.680          Germany 94  
## 5           India     1.980           India 66
```

Merging Data Frames

Simply use the `cbind()` function.

```
data <- as.data.frame(  
  cbind(gdp_data, fh_data[, -1])  
)
```

```
print(data)
```

```
##           country gdp_capita fh_data[, -1]  
## 1 United States    59.939         83  
## 2          China     8.612          9  
## 3          Japan    38.214         96  
## 4          Germany  44.680         94  
## 5           India     1.980         66
```

```
colnames(data)[3] <- "fh"
```

Merging Data Frames

```
print(data)
```

```
##           country gdp_capita fh
## 1 United States    59.939 83
## 2         China     8.612  9
## 3         Japan    38.214 96
## 4       Germany    44.680 94
## 5         India     1.980 66
```

Merging Data Frames - Different Orders

```
gdp_data <- read.csv2("dat/gdp_data_alt.csv")
```

How to merge these two data frames?

```
print(gdp_data)
```

```
##           country gdp_capita
## 1           China      8.612
## 2 United States    59.939
## 3           Japan    38.214
## 4           Germany  44.680
## 5           India    1.980
```

```
print(fh_data)
```

```
##           country fh
## 1 United States 83
## 2           China 9
## 3           Japan 96
## 4           Germany 94
## 5           India 66
```


Merging Data Frames - Different Orders

Use the `merge()` function.

- The variable `country` becomes the `key` that relates observations to each other

```
data <- as.data.frame(  
  merge(gdp_data, fh_data, by = "country")  
)
```

```
print(data)
```

```
##      country gdp_capita fh  
## 1      China      8.612   9  
## 2    Germany     44.680  94  
## 3      India      1.980  66  
## 4      Japan     38.214  96  
## 5 United States  59.939  83
```

References

Parts of this course are inspired by the following resources:

- Wickham, Hadley and Garrett Grolemund, 2017. *R for Data Science - Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly.
- Bahnsen, Oke and Guido Ropers, 2022. *Introduction to R for Quantitative Social Science*. Course held as part of the GESIS Workshop Series.
- Breuer, Johannes and Stefan Jünger, 2021. *Introduction to R for Data Analysis*. Course held as part of the GESIS Summer School in Survey Methodology.
- Teaching material developed by Verena Kunz, David Weyrauch, Oliver Rittmann and Viktoriia Semenova.