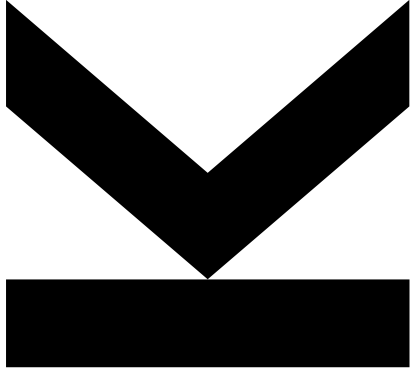


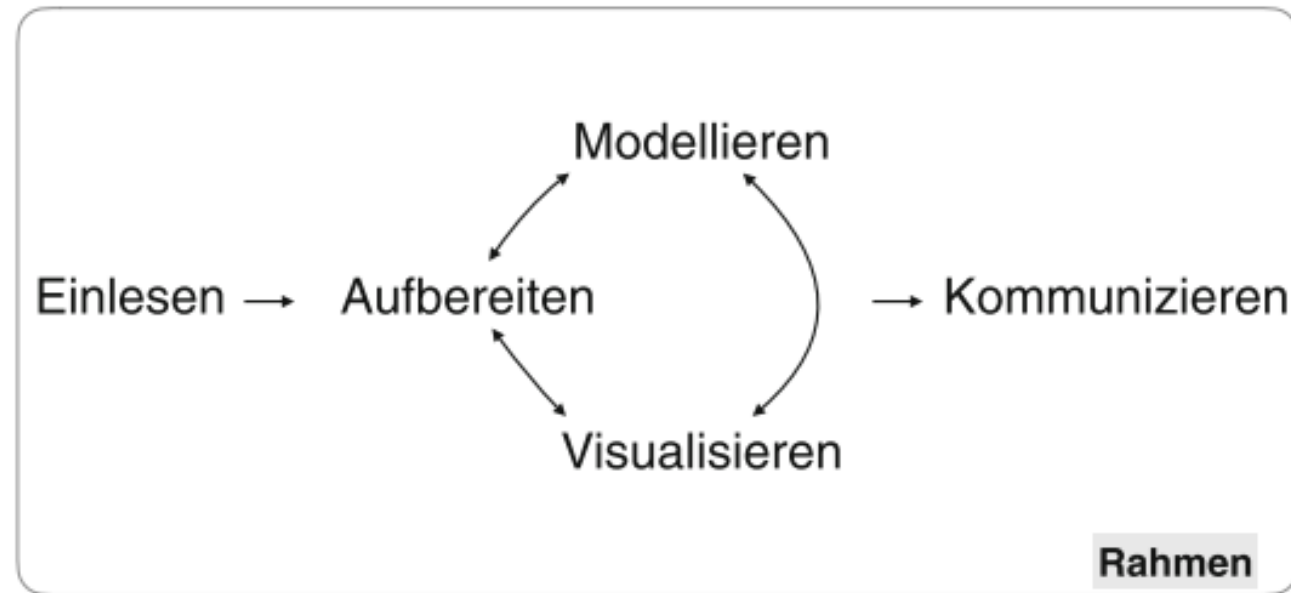
Einführung in R

#1 Was ist R & RStudio?



Tobias Wiß, Carmen Walenta & Felix Wohlgemuth

Was ist Datenanalyse?



Quelle: Sauer, S. (2018) *Moderne Datenanalyse mit R*

Daten Einlesen

Struktur von Datensätzen #1

Show **5** entries

Search:

| | COUNTRY | DATE | REFORM_NAME | CHANGE |
|---|---------|---------|--|---------------|
| 1 | Austria | 2019-01 | Papa-Monat SPÖ paternity leave | introduction |
| 2 | Austria | 2018-06 | Family Bonus Plus | introduction |
| 3 | Austria | 2016-05 | Flexible parental leave payment scheme, equal share bonus, child benefit account | recalibration |
| 4 | Austria | 2016-05 | Family Time + Partnership Bonus | introduction |
| 5 | Austria | 2015-12 | Part-time fathers & mothers | recalibration |

Showing 1 to 5 of 10 entries

Previous **1** 2 Next

Daten Einlesen

Struktur von Datensätzen #2

Show **8** entries

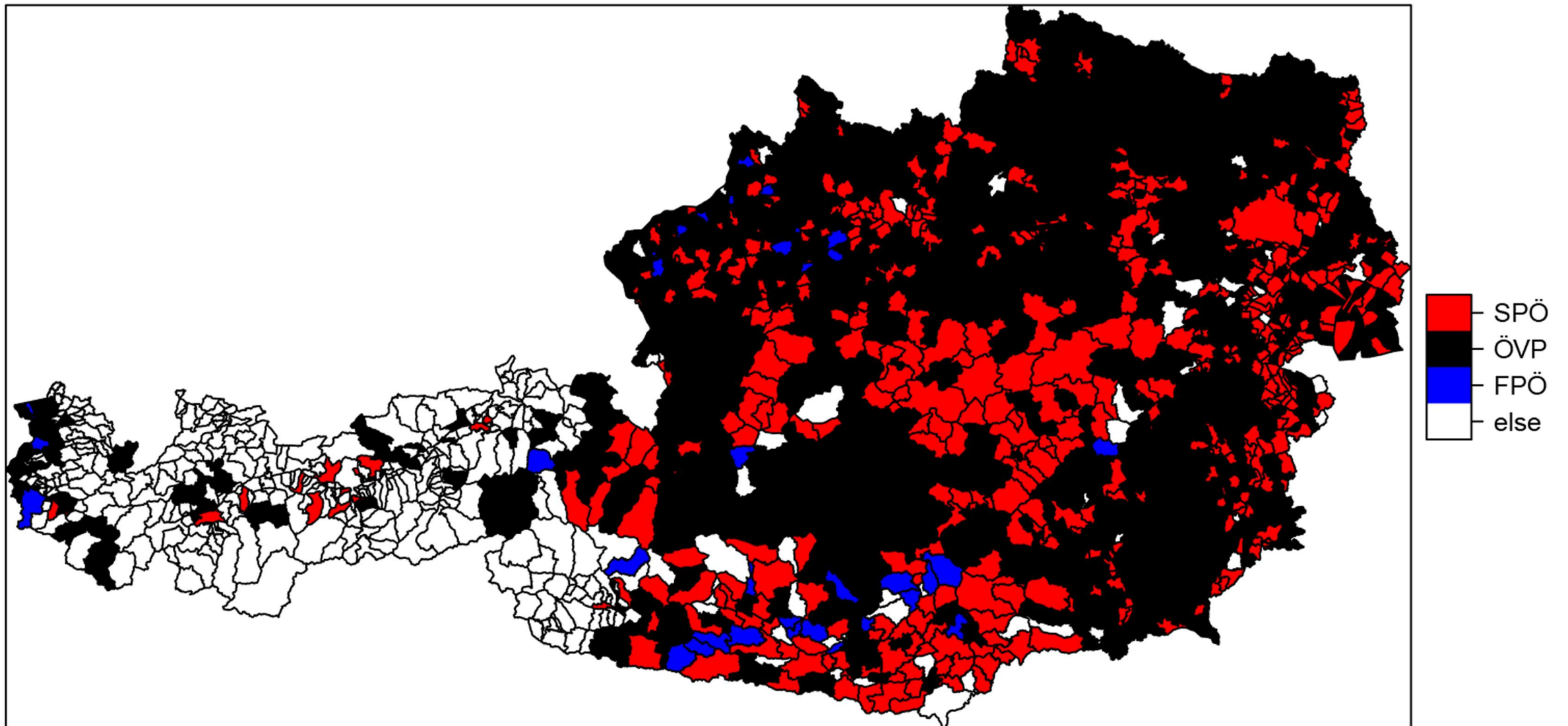
Search:

| | Country | YEAR | maternity_leave_WKS | fam_pol_exp_GDP |
|---|-------------|------|---------------------|-----------------|
| 1 | Australia | 2013 | 18 | 2.8 |
| 2 | Austria | 2013 | 60 | 2.6 |
| 3 | Belgium | 2013 | 32.3 | 3.3 |
| 4 | Bulgaria | 2013 | | |
| 5 | Canada | 2013 | 52 | 1.4 |
| 6 | Switzerland | 2013 | 14 | 2 |
| 7 | Chile | 2013 | | 1.4 |
| 8 | Cyprus | 2013 | | |

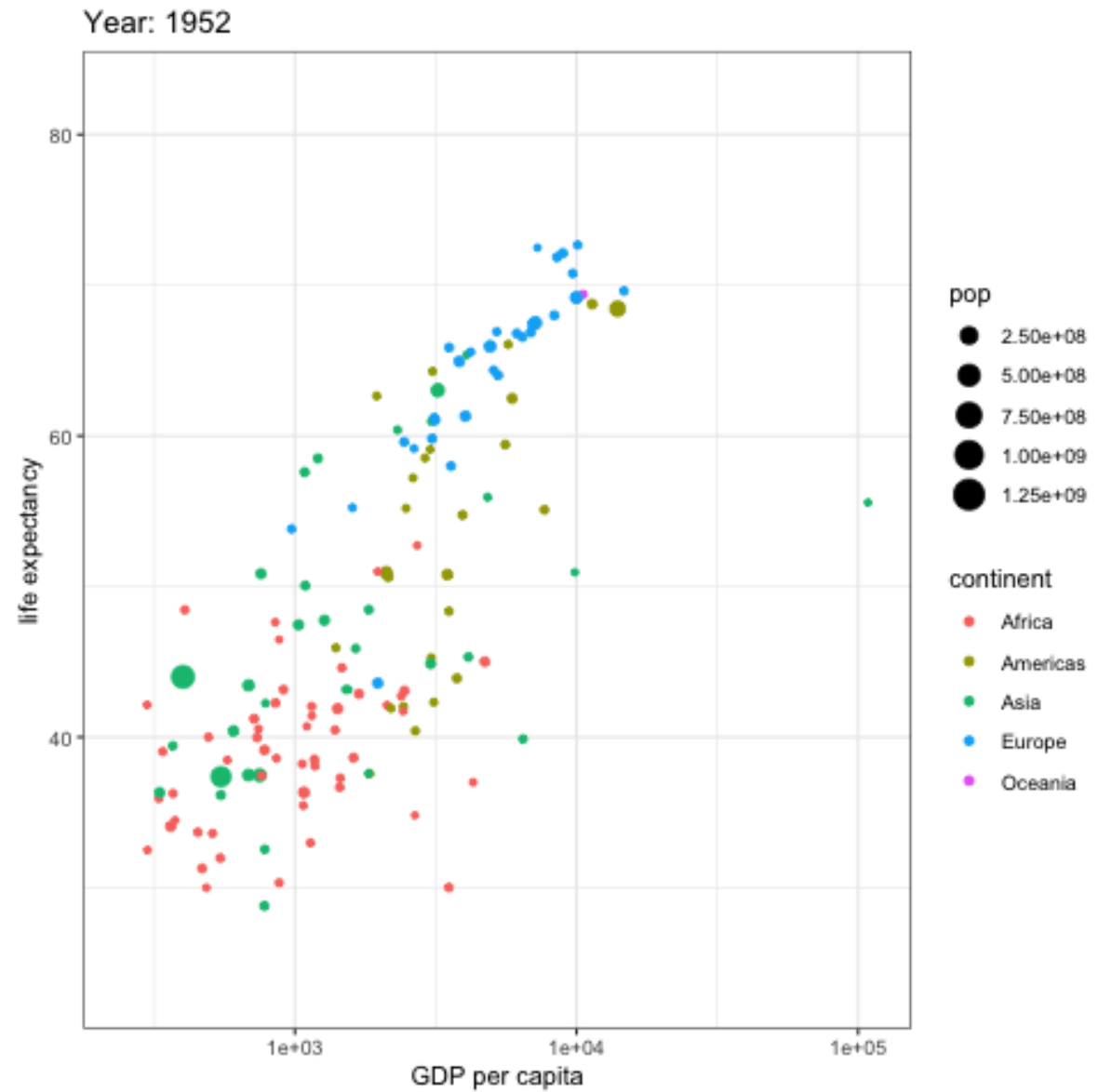
Showing 1 to 8 of 10 entries

Previous **1** 2 Next

Daten Visualisieren



Daten Visualisieren

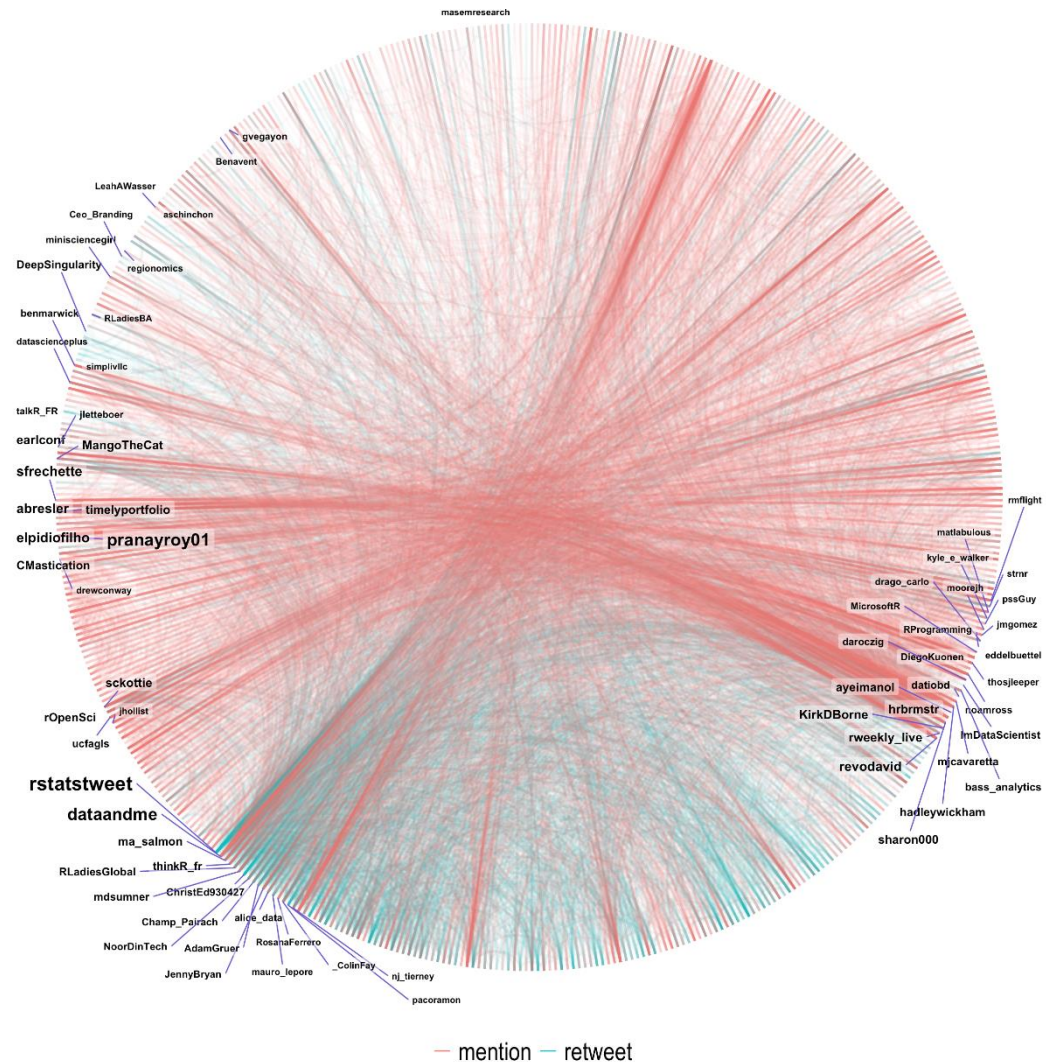


Quelle: <https://www.r-graph-gallery.com/>

Daten Visualisieren

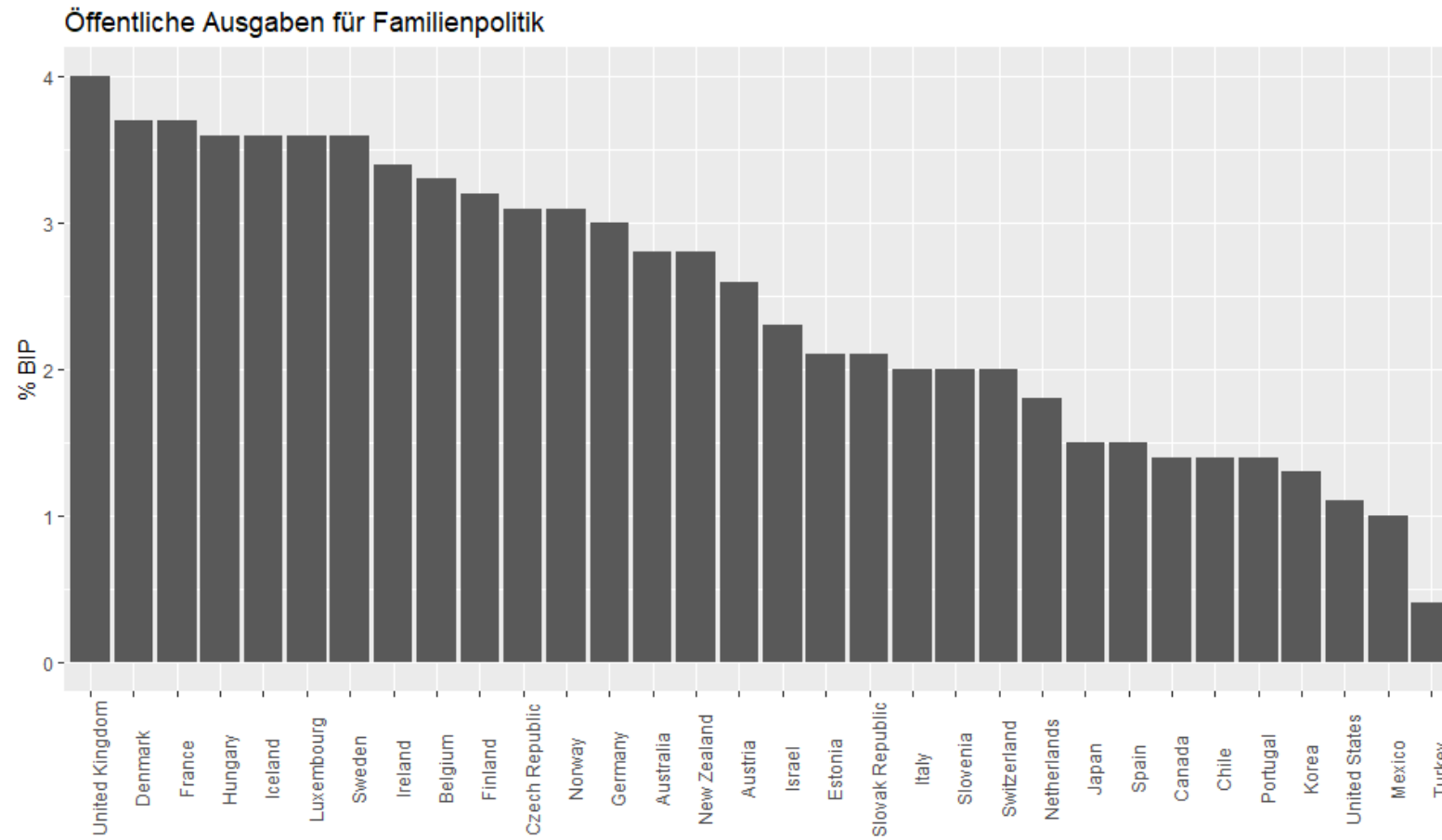
R Twitter Activity

Retweets and mentions between key members of the community



Quelle: <https://perrystephenson.me/2018/09/29/the-r-twitter-network/>

Daten Visualisieren



Daten Modellieren

- Hat die EU-Mitgliedschaft Einfluss auf die Höhe der Ausgaben für Familienpolitik?

```
summary.aov(aov(fam_pol_exp_GDP ~ EU, socx_fam))
```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) | |
|-----------|----|--------|---------|---------|---------|----|
| EU | 1 | 6.623 | 6.623 | 8.399 | 0.00695 | ** |
| Residuals | 30 | 23.657 | 0.789 | | | |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
9 observations deleted due to missingness

Ablauf des Seminars

| | Datum | Datenanalyse mit R |
|----|------------|--|
| 1 | 05.03.2020 | Einführung: Was ist R und RStudio? |
| 2 | 12.03.2020 | Erste Schritte mit RStudio |
| 3 | 19.03.2020 | Daten und Datentypen |
| 4 | 26.03.2020 | Daten finden und importieren |
| 5 | 01.04.2020 | |
| 6 | 02.04.2020 | Daten aufbereiten |
| | | |
| 7 | 23.04.2020 | Beschreibende Statistik: Einzelvariablen |
| 8 | 30.04.2020 | Visualisierungen mit ggplot2 |
| 9 | 07.05.2020 | Visuelle Darstellung von Einzelvariablen |
| 10 | 15.05.2020 | Beschreibende Statistik: mehrere Variablen |
| | | |
| 11 | 28.05.2020 | Visuelle Darstellung mehrerer Variablen |
| 12 | 04.06.2020 | Multiple lineare Regression |
| | | |
| 13 | 18.06.2020 | Logistische Regression |
| 14 | 25.06.2020 | Kommunikation von Ergebnissen |

Was ist R?



If statistics programs/languages were cars...

Image courtesy of Darren Dahly @stapsepi



Warum R?



- **Open source!**
 - Kostenlos
 - Wird durch Nutzer*innen konstant erweitert
 - Jede/r kann neuen Pakete veröffentlichen (CRAN)
- Daten so visualisieren wie man möchte
- Die ganze Datenanalyse kann mit R gemacht werden
 - Daten erstellen / einlesen (mit API direkt vom Server)
 - Daten visualisieren / modellieren
 - Interaktive Berichte schreiben mit RMarkdown / Präsentation erstellen mit xaringan
 - Auf Basis der Daten web-apps erstellen mit Shiny
- Reproduzierbarkeit & Transparenz
- Fragen & Probleme? Riesige Community die ihre Skripte frei zu Verfügung stellt

Was ist RStudio?



- R ist eine Programmiersprache
- RStudio ist eine „integrated development environment“ / Entwicklungsumgebung

R: Engine



RStudio: Dashboard



Quelle: Ismay & Kim (2020) A ModernDive into R and the Tidyverse: <https://moderndive.netlify.com/1-getting-started.html>

Installation von R & RStudio



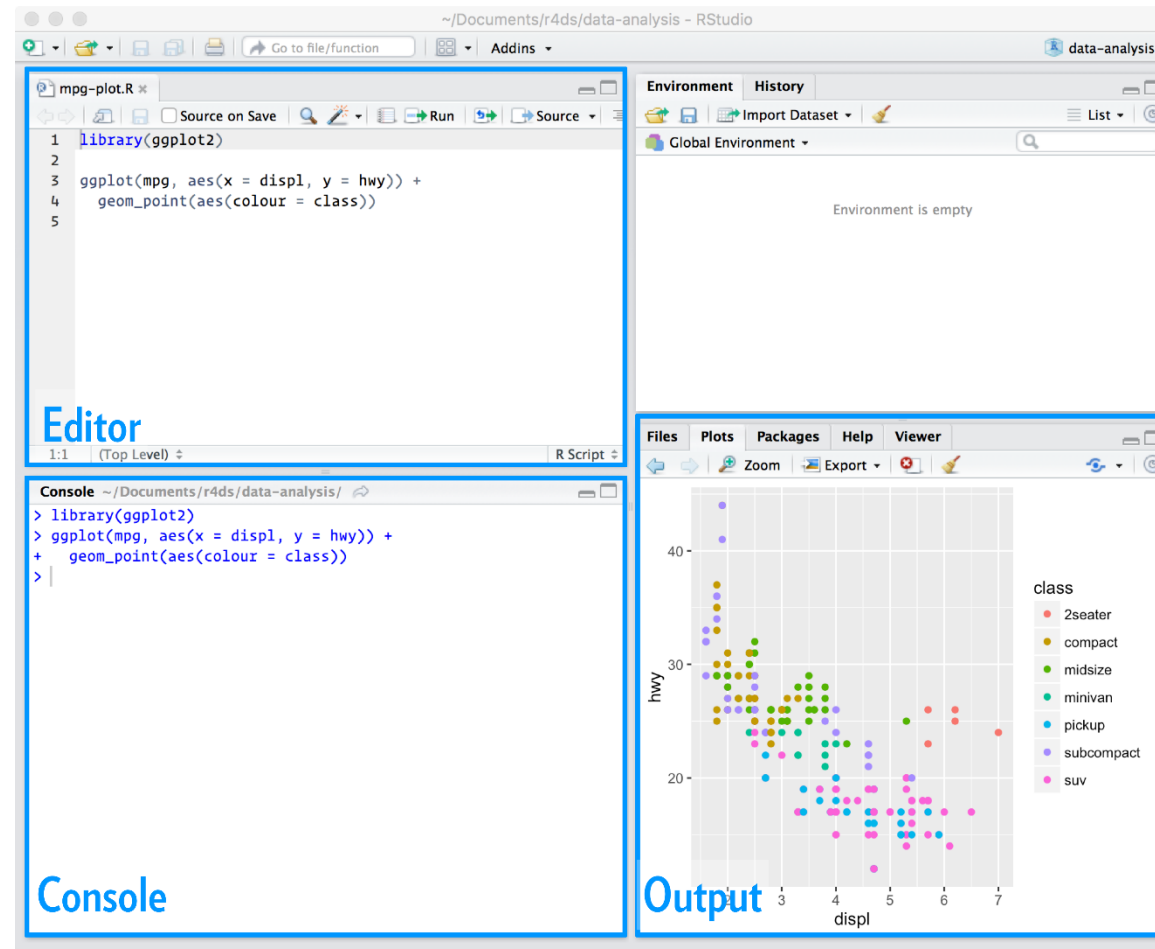
R

- <https://cloud.r-project.org/>
- Version für jeweiliges Betriebssystem auswählen
 - Download R for (Mac) OS X > R-3.6.2.pkg
 - Download R for Windows > base > Download R 3.6.3 for Windows

RStudio

- <https://rstudio.com/products/rstudio/download/#download>
- Version für jeweiliges Betriebssystem auswählen
- Tutorial zur Installation: <https://learnr-examples.shinyapps.io/ex-setup-r/>

Oberfläche von RStudio



Quelle: Golemund & Wickham (2019) R for Data Science

Erste Schritte in RStudio



- Neues R Script erstellen
 - File > New File > R Script
- Kurze Rechnung
 - Was ist $2*2$?
 - Ist $4+4$ ident mit 8?

Erste Schritte in RStudio

- Nachschlagen
 - Was ist das Paket „base“
 - Was ist die Funktion „sqrt“
- Neue Pakete installieren & laden: nycflights13 – on-time data for all flights departing NYC 2013
 - `install.packages(„nycflights13“)` # Installation
 - `library(nycflights13)` # Paket laden
 - `help(package = nycflights13)` # Dokumentation aufrufen

Aufgabe 1

- Erstellen Sie ein neues Skript in RStudio
- Führen Sie folgende Schritte per Skript aus:
 - installieren Sie das Paket "tydiverse" mit **install.packages()**
 - laden Sie das Paket "tydiverse" mit **library()**
 - Rufen Sie den help-Eintrag von "ggplot" auf mit **help()** oder ?
 - Kopieren Sie die Beschreibung von "ggplot()" als #-Kommentar in Ihr Skript
- Laden Sie das R-Skript mit dem folgender Dateinamen auf Moodle hoch:
"R_Aufgabe1_IHR_NAME.R,,

Deadline: 11.02.2020 12:00