



CORRELAID
GOOD CAUSES. BETTER EFFECTS.

What is R?

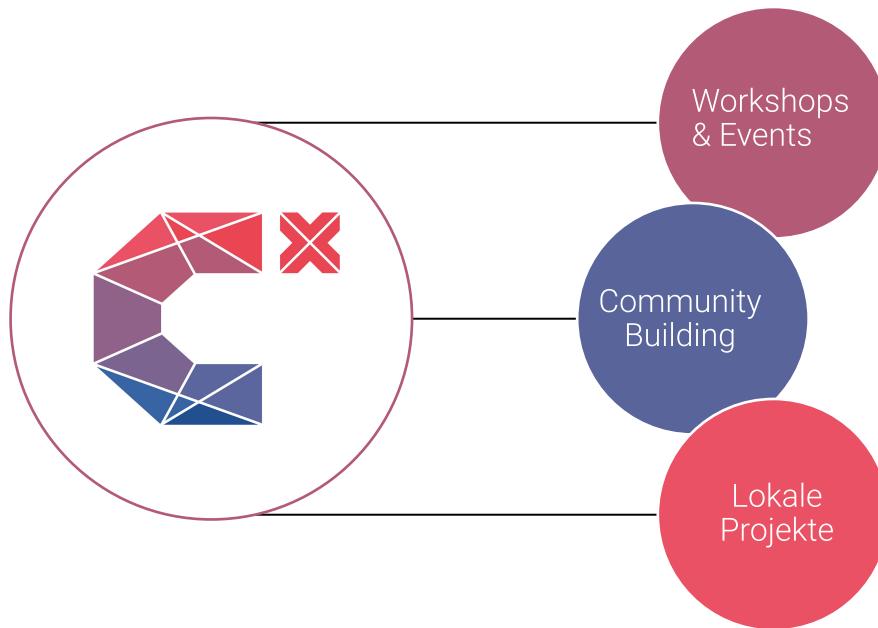
Einführung in die statistische Programmiersprache R

Outline

- Was ist CorrelAid?
- Warum ausgerechnet Konstanz?
- Was kann R eigentlich so?
- R-Studio vs R
 - Was ist eine IDE?
 - Installation
- IDE Walkthrough (customize it!)
- Unser erstes Projekt
 - Unsere ersten Befehle
 - Unser erstes Skript
 - Unsere ersten Funktionen



CorrelAid und CorrelAid X Konstanz



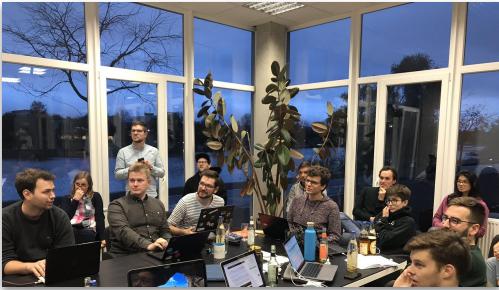
CorrelAid

- Netzwerk von über 1000 Datenanalysten
- Studenten und Berufstätige
- Data4Good-Projekte
- gegründet in Konstanz!
- mittlerweile ziemlich groß (und cool)

Local Chapter in Konstanz

- Workshops
- Hackathons (mit dem SWR)
- lokale Projekte

Was wir in Konstanz sonst so treiben



Hackathons

Gemeinsam mit dem Datenjournalismus-Team des SWR suchen wir spannende Stories in Daten.



Lokale Projekte

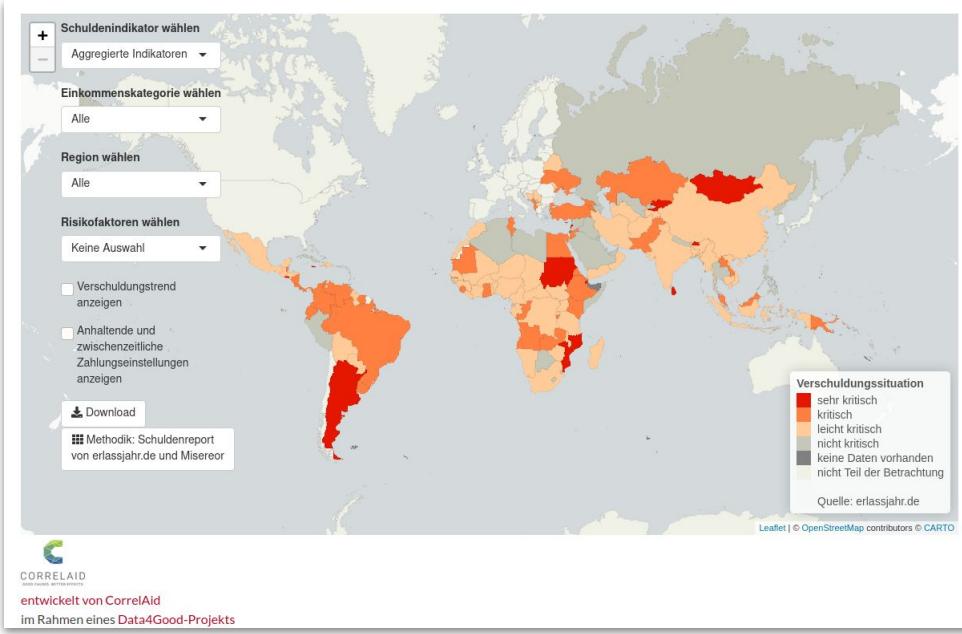
In Data4Good-Projekten unterstützen wir andere Non-Profits, das Meiste aus ihren Daten rauszuholen.



Daten-Stammtisch

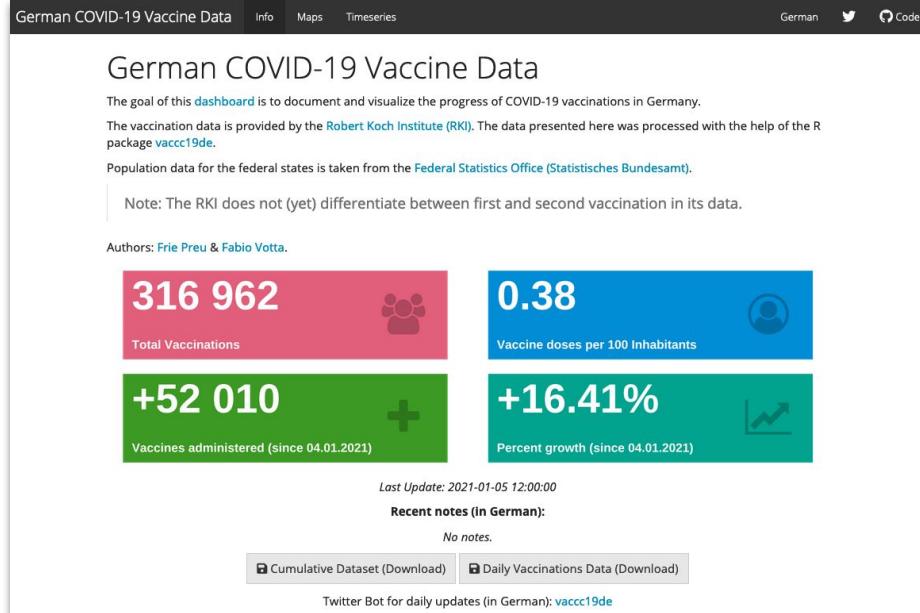
Wir treffen uns regelmäßig (wenn nicht gerade Corona), um uns auszutauschen und zu vernetzen.

R-Beispiele: Interaktive Karte für erlassjahr.de



- erlassjahr.de setzt sich für weltweite Schuldenschnitte ein
- Echtzeitberechnung von Staatsverschuldung
- Berechnung eigener Indikatoren für kritisch verschuldete Staaten
- Visualisierung mit shiny und leaflet

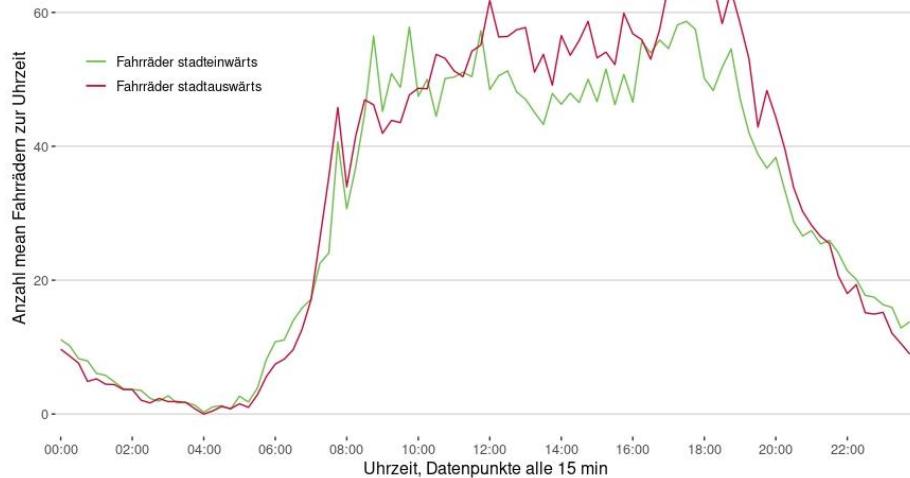
R-Beispiele: German COVID-19 Data



- https://favstats.github.io/vaccine19de_dashboard/en/index.html
- Dashboard und Visualisierung deutscher COVID-19 Impfungen
- Entwickelt von zwei CorrelAid Mitgliedern (davon eine Ex-PoLVerin! 😎)



R-Beispiele: Offene Daten der Stadt Konstanz

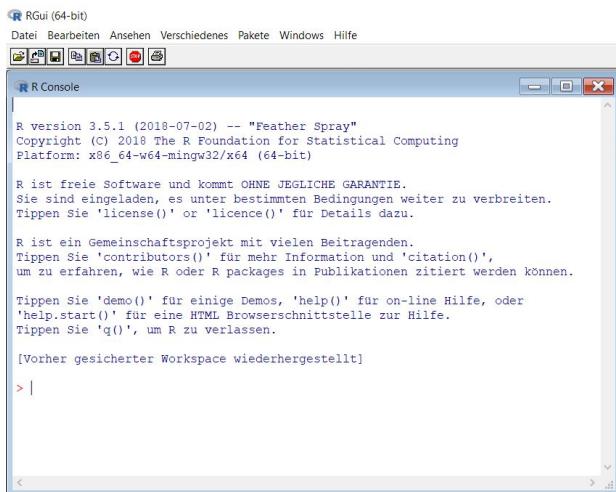


- Visualisierung von Fahrrädern die über die Fahrradbrücke fahren.
- Mean über ein Jahr
- Entwickelt von the man himself: Tilman Kerl 🔥

R vs R-Studio

Was ist R?

- die Sprache an sich
- sieht ohne R-Studio so aus:



The screenshot shows the RGui (64-bit) console window. The title bar reads "RGui (64-bit)". The menu bar includes "Datei", "Bearbeiten", "Ansehen", "Verschiedenes", "Pakete", "Windows", and "Hilfe". Below the menu is a toolbar with icons for file operations. The main window is titled "R Console". It displays the R startup message:
R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R ist freie Software und kommt OHNE JEGLICHE GARANTIE.
Sie sind eingeladen, es unter bestimmten Bedingungen weiter zu verbreiten.
Tippen Sie 'license()' or 'licence()' für Details dazu.

R ist ein Gemeinschaftsprojekt mit vielen Beitragenden.
Tippen Sie 'contributors()' für mehr Information und 'citation()',
um zu erfahren, wie R oder R packages in Publikationen zitiert werden können.

Tippen Sie 'demo()' für einige Demos, 'help()' für on-line Hilfe, oder
'help.start()' für eine HTML Browserschnittstelle zur Hilfe.
Tippen Sie 'q()', um R zu verlassen.

[Vorher gesicherter Workspace wiederhergestellt]
> |

Was ist R-Studio?

- Integrated Development Environment (IDE)
- Oberfläche, mit der wir R bedienen, Skripte schreiben, Plots und Hilfefenster anschauen
- "*Word für R*"



Let's get into it / R-Studio Walkthrough

The screenshot shows the R-Studio IDE interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. The left sidebar has tabs for Environment, History, and Connections, with Global Environment currently selected. The main workspace displays a script editor with the following R code:

```
1 - #####  
2 # CorrelAid x Konstanz #  
3 #  
4 # What is R? #  
5 #####  
6  
7 ## Laden von packages  
8 |library(memr)
```

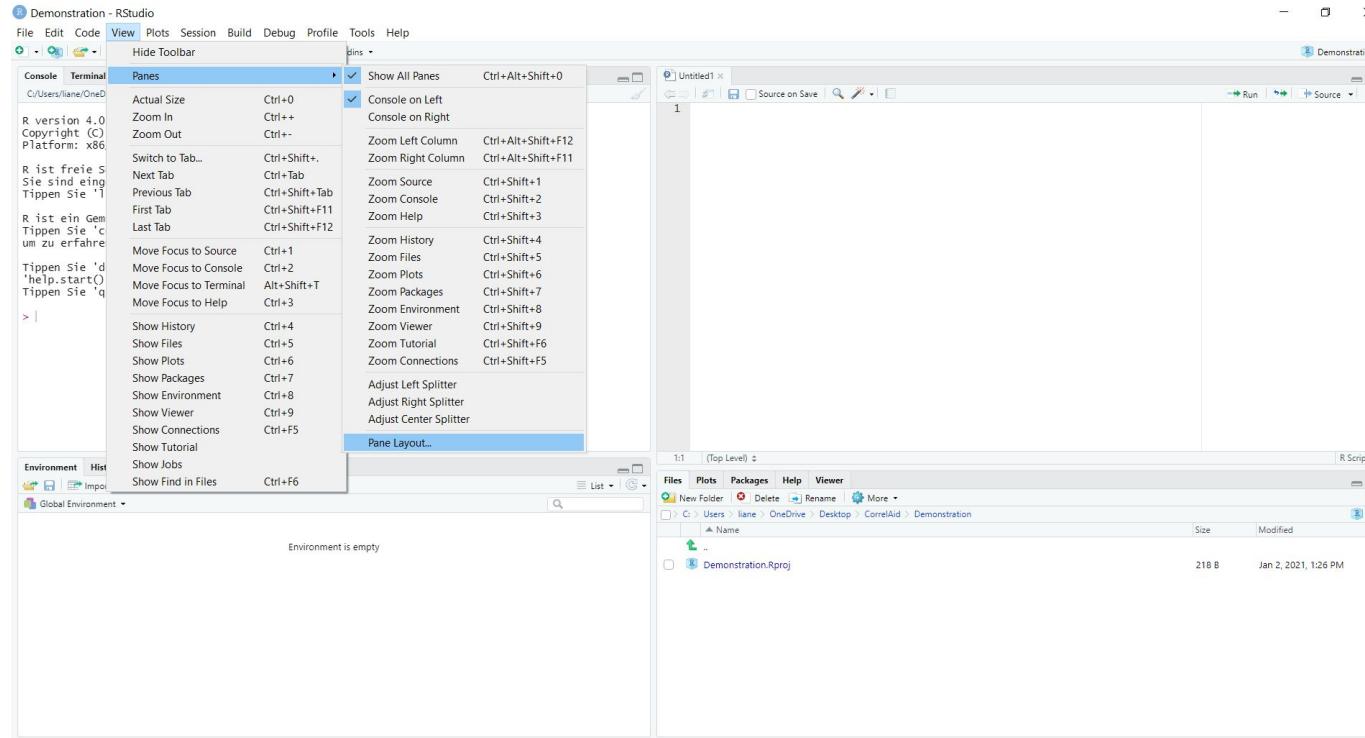
The right side of the interface features a Console window showing the R startup message and a Terminal window. Below these is a file browser titled "What-is-R" showing the project structure:

Name	Size	Modified
..		
What-is-R.Rproj	218 B	Dec 22, 2020, 10:51 PM
script.R	174 B	Dec 23, 2020, 1:52 AM
data		
plots		
.Rhistory	37 B	Dec 23, 2020, 1:52 AM

The bottom status bar indicates "Environment is empty".



R-Studio Walkthrough - Change Layout



Pfade - Wo liegen meine Dateien?

The screenshot shows a Windows File Explorer window with the following details:

Left Sidebar (Schnellzugriff):

- Häufig verwendete Ordner (8):
 - Desktop
 - Downloads
 - Dokumente
 - Bilder
 - Arbeitsblätter
 - Auswertung2020
 - neue COVID-Grafik
 - Übung
- Dropbox
- OneDrive
- Dieser PC
- 3D-Objekte
- Bilder
- Desktop
- Dokumente
- Downloads
- Musik
- Videos
- Acer (C)
- Netzwerk

Recent Files (Zuletzt verwendete Dateien 20):

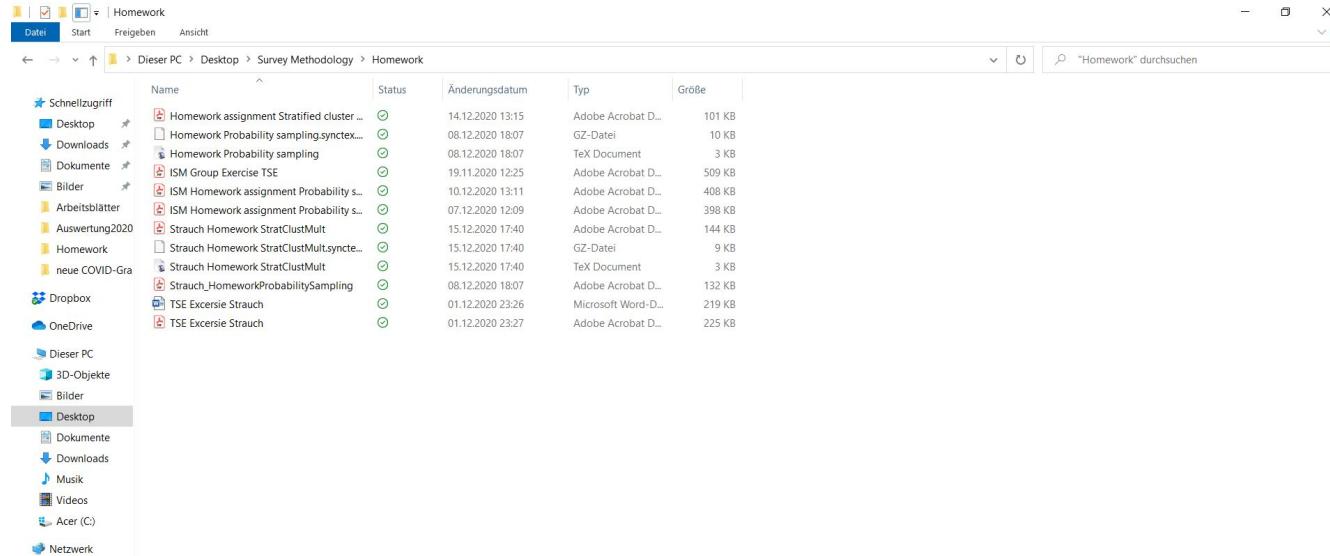
Name	Pfad
COVID-Tabsels_2020_2_Strauch	OneDrive\Desktop\Aufträge EEP\UN Paper\neue COVID-Graphik
covid-graphik	OneDrive\Desktop\Aufträge EEP\UN Paper\neue COVID-Graphik
r-primer	OneDrive\Desktop\Web Data Collection
packages	OneDrive\Desktop\Web Data Collection
Strauch Homework StratClustMult	OneDrive\Desktop\Survey Methodology\Homework
ISM Group Exercise TSE	OneDrive\Desktop\Survey Methodology\Homework
Formelsammlung	OneDrive\Desktop\Survey Methodology
05_StratClustMulti	OneDrive\Desktop\Survey Methodology
08_Algorithmitk-Dynamische_Programmierung	OneDrive\Desktop\Kdf\Vorlesung
03_1Informationsspeicherung	OneDrive\Desktop\Kdf\Vorlesung
Ergebnis2013	Liane Strauch\Documents\Studium (Archiv)\R Kurs\Material
Assignment8_Liane Strauch	Liane Strauch\Documents\Studium (Archiv)\R Kurs\Material
Task4_3	Liane Strauch\liane_strauch\assignments\PK1\assignment04
Node	Liane Strauch\pub\material\lesson00\binaryTree
Recursion	Liane Strauch\pub\lectureScript\lesson05\script
RecursionExamples	Liane Strauch\pub\lectureScript\lesson05\script
Antrag_Vertragserstellung_korrekt_IPA	Dieser PC\Downloads
Immatriculationsbescheinigung Liane Strauch WS2_21	Dieser PC\Downloads
3.Strauch.KDL07	OneDrive\Desktop\Kdf\Arbeitsblätter
3 Strauch.KDL03	OneDrive\Desktop\Kdf\Arbeitsblätter

Bottom Taskbar:

- Start button
- Search bar: Zur Suche Text hier eingeben
- Icons for File Explorer, Task View, Edge, Internet Explorer, File History, Task Scheduler, File Explorer, Mail, Microsoft Store, and File Explorer
- System tray: Battery, Network, Volume, DEU, 13:40, 02.01.2021



Pfade - Wo liegen meine Dateien?



The screenshot shows a Windows File Explorer window with the following details:

- Title Bar:** Homework
- Menu Bar:** Datei, Start, Freigeben, Ansicht
- Address Bar:** Dieser PC > Desktop > Survey Methodology > Homework
- Search Bar:** "Homework" durchsuchen
- Table:** A list of files in the 'Homework' folder.
- Left Sidebar:** Navigation pane with links like Schnellzugriff, Desktop, Downloads, Dokumente, Bilder, Arbeitsblätter, Auswertung2020, Homework, neue COVID-Gra, Dropbox, OneDrive, Dieser PC, 3D-Objekte, Bilder, Desktop, Dokumente, Downloads, Musik, Videos, Acer (C), and Netzwerk.

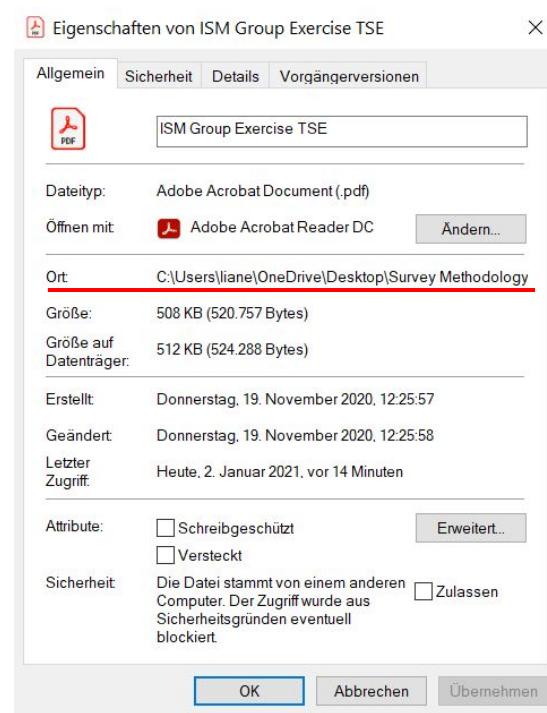
Name	Status	Änderungsdatum	Typ	Größe
Homework assignment Stratified cluster ...	✓	14.12.2020 13:15	Adobe Acrobat D...	101 KB
Homework Probability sampling.synctex...	✓	08.12.2020 18:07	GZ-Datei	10 KB
Homework Probability sampling	✓	08.12.2020 18:07	TeX Document	3 KB
ISM Group Exercise TSE	✓	19.11.2020 12:25	Adobe Acrobat D...	509 KB
ISM Homework assignment Probability s...	✓	10.12.2020 13:11	Adobe Acrobat D...	408 KB
ISM Homework assignment Probability s...	✓	07.12.2020 12:09	Adobe Acrobat D...	398 KB
Strauch Homework StratClustMult	✓	15.12.2020 17:40	Adobe Acrobat D...	144 KB
Strauch Homework StratClustMult.syncte...	✓	15.12.2020 17:40	GZ-Datei	9 KB
Strauch Homework StratClustMult	✓	15.12.2020 17:40	TeX Document	3 KB
Strauch_HomeworkProbabilitySampling	✓	08.12.2020 18:07	Adobe Acrobat D...	132 KB
TSE Excercise Strauch	✓	01.12.2020 23:26	Microsoft Word-D...	219 KB
TSE Excercise Strauch	✓	01.12.2020 23:27	Adobe Acrobat D...	225 KB



Pfade - Wo liegen meine Dateien?

Pfad einer Datei herausfinden
(Windows):

Rechtsklick auf die Datei >>
Eigenschaften



Pfade in R

- öffnet R Studio
- mit dem Befehl `getwd()` bekommt ihr den Pfad in dem ihr euch befindet zurück
- ihr könnt unter "Files" sehen auf welche Dateien R sofort zugreifen kann



Pfade in R setzen

- erstellt einen Ordner mit dem Namen: “*Testprojekt*”
- und zwei Unterordnern mit den Namen “*Daten*” und “*Ergebnisse*”
- (wenn wir schon mit Datensätzen arbeiten wäre es ganz praktisch, wenn die diese jetzt hier reinkopieren können)
- geht in R-Studio
- gibt den Befehl `setwd (“PFAD”)` mit dem Pfad eures Testprojekt-Ordners in den Klammern ein
- mit `getwd ()` könnt ihr überprüfen, ob der Pfad jetzt richtig gelegt ist
- Unter “*Files*” solltet ihr jetzt die beiden Unterordner “*Daten*” und “*Ergebnisse*” sehen können.



relative und absolute Pfade

- bisher haben wir nur **absolute Pfade** kennengelernt
 - sie fangen mit c : / an und gehen dann von Ordner zu Ordner
- Es gibt auch **relative Pfade** die im aktuellen Pfad beginnen und von dort in weitere Unterordner gehen
 - mit dem Befehl setwd ("./Ergebnisse") setzt ihr euren Pfad in den Unterordner "Ergebnisse"
 - der Punkt in setwd () ist eine Art Platzhalter für den aktuellen Pfad



R Projekte

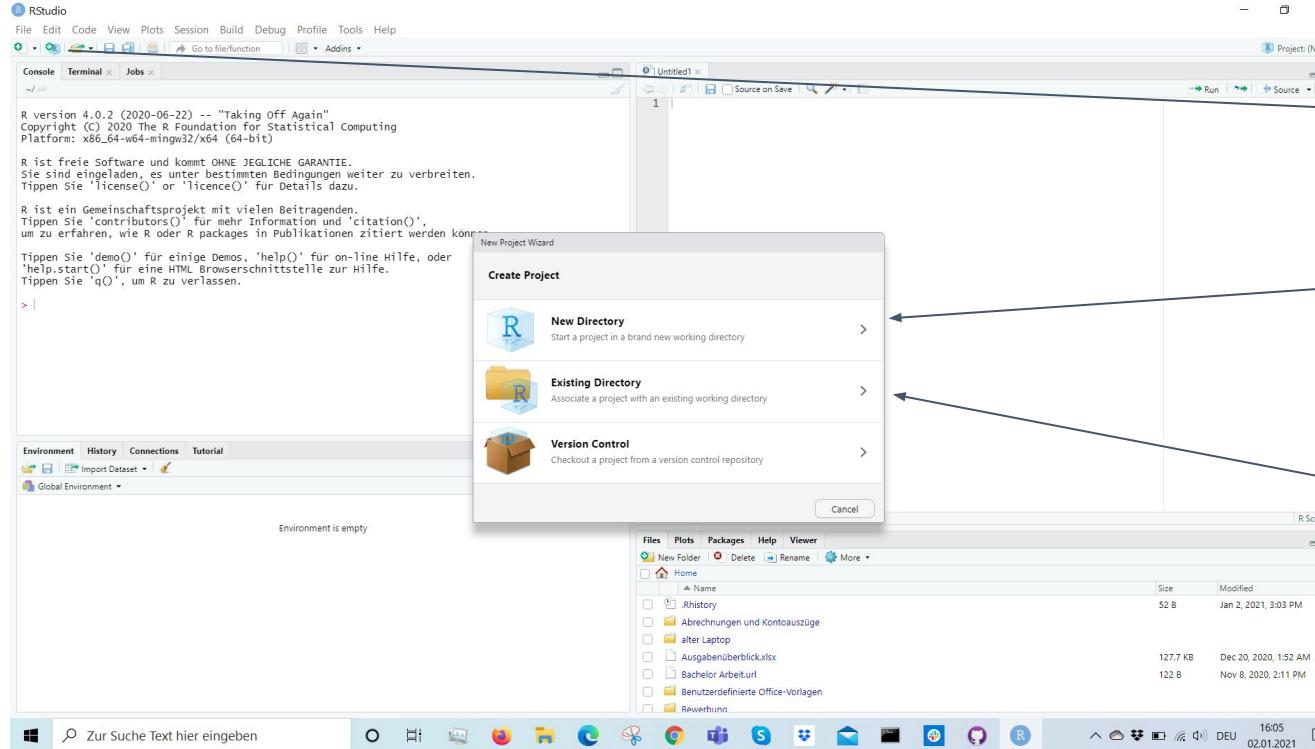
```
1 - #####  
2 # CorrelAid x Konstanz #  
3 # #  
4 # What is R? #  
5 - #####  
6  
7 ## aktuellen Pfad abfragen  
8 getwd()  
9  
10 ## Pfad ändern  
11 setwd("C:/Users/liane/OneDrive/Dokumente/GitHub/CorrelAidXKonstanz-Education/what-is-R")  
12
```

Angenommen ihr wollt den Befehl in **Zeile 11** genauso wie er hier steht auf eurem Rechner ausführen.
Auf welches Problem werdet ihr stoßen?

Der Pfad auf meinem Rechnern ist nicht der gleich wie auf eurem. Besser ist es, mit **R Projekten** zu arbeiten.



R Projekt

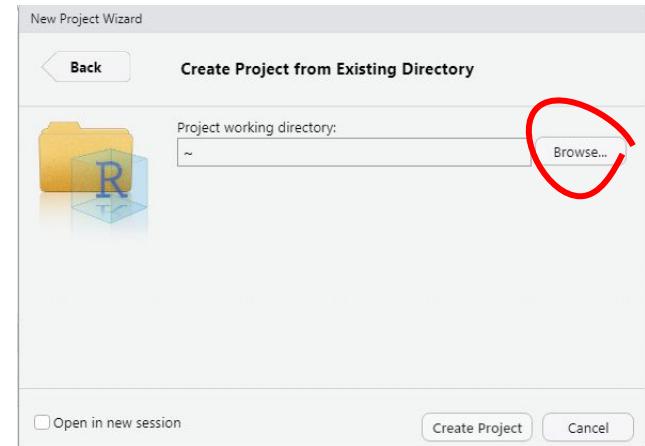


Wenn ihr noch keinen
Ordner erstellt habt

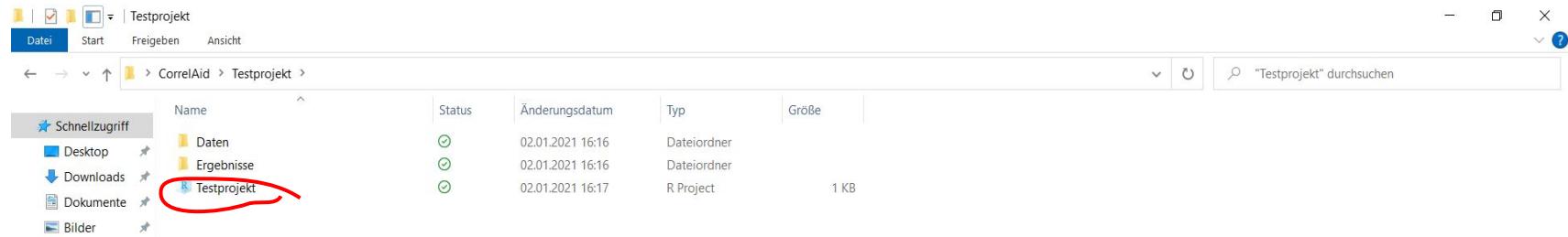
Wenn ihr bereits
einen Ordner habt

R Projekt

- da wir schon einen Ordner für unsere Analyse haben wählen wir "*Existing Directory*"
- Es öffnet sich ein neues Fenster (siehe links)
- entweder könnt ihr den Pfad per Hand eingeben (nicht empfohlen → Tippfehler)
- oder mit *Browse* in eurem PC nach dem entsprechenden Ordner suchen



R Projekt



Wenn ihr jetzt an eurem Projekt arbeiten wollt, öffnet ihr nicht mehr R Studio, sondern das R Projekt in eurem Projektordner.

Der Pfad ist sofort richtig gesetzt und ihr müsst euch darum keine weiteren Gedanken machen.







CORRELAID
GOOD CAUSES. BETTER EFFECTS.

Session 2

Jetzt geht's richtig los
fasten your seatbelts, we be touring the tidyverse 

Outline

- Warum eigentlich R-Projects?
 - relative vs. absolute Pfade
- The powerhouse of R: Packages
- Das tidyverse
- Tidyverse in action:
 - einlesen von Daten
 - cleaning
 - aufbereiten
 - mergen
 - modellieren
 - visualisieren



R-Projekte



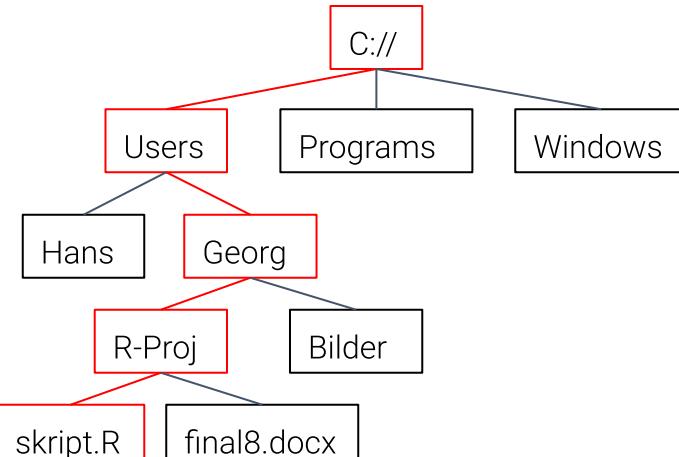
- speichern Benutzerdaten
 - bereits zuvor erstellte Objekte
 - Benutzereinstellungen
 - Oberfläche beim Schließen der letzten Sitzung
- erstellen beim Starten eine neue Sitzung
- das Verzeichnis des R-Projekts wird automatisch zum Arbeitsverzeichnis

→ genau da weitermachen, wo ihr aufgehört habt

Relative vs. absolute Pfade

Absolute Pfade:

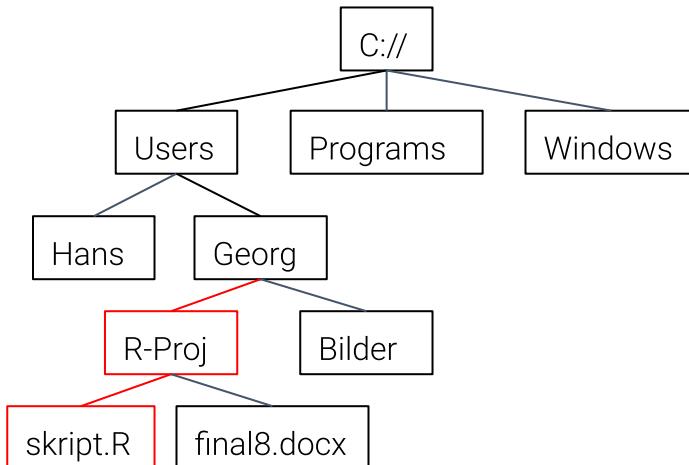
- starten von Null (c:/)



C://Users/Georg/R-Proj/skript.R

Relative Pfade:

- ausgehend vom Arbeitsverzeichnis



/R-Proj/skript.R

packages - Funktionalität erweitern

- R hat eine sehr aktive Community, die ständig neue *packages* entwickelt
- die meisten *packages* könnt ihr euch wie zusätzliche Funktionen vorstellen, die andere bereits für euch geschrieben haben
- *packages* haben eine Dokumentation, in der steht, wie man die Funktionen anwendet
- mit *readr::* könnt ihr alle Funktionen in einem package anzeigen (hier im package "readr")



Where to find help??

- `?mean()` zeigt die Dokumentation der *mean*-Funktion
- R Cheatsheets (von R-Studio herausgegeben)
- online R-Bücher
- GOOGLE!!!!!!!
- auf stackoverflow.com kann man Fragen stellen (Fragen, die euch auch Google beantwortet, werden dort aber nicht gerne gesehen.)

→ [Übersicht](#) über Ressourcen und Bücher



Ihr seid nicht allein im UniveRsum

Irgendjemand hatte vor euch ein ähnliches Problem und hatte die Courage online um Hilfe zu bitten

- Meistens geschieht das auf [Stackoverflow](#)
- Einfach euer Problem googeln und ein "r" hinzufügen



Aufgabe von Session 1



Touring the tidyverse: dplyr

- tidyverse: Sammlung von packages welche derselben Logik folgen und aufeinander abgestimmt sind
 - Erster Parameter aller Funktion bestimmt den Datensatz
 - Die darauf folgenden Parameter definieren was mit dem Datensatz gemacht wird
 - Nachschlagewerk: [R for Data Science](#)
- dplyr: Teil des tidyverse
 - erleichtert die Modifikation und Analyse von Daten



CORRELAID

KONSTANZ



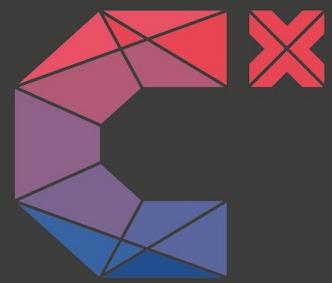
For instant updates on our work
follow us on Twitter
@CorrelAid



Receive the latest infos in our
projects via Facebook
@WeAreCorrelAid



Sign up for our mailing list on
www.correlaid.org



CORRELAID

KONSTANZ

