



CORRELAID

GOOD CAUSES. BETTER EFFECTS.

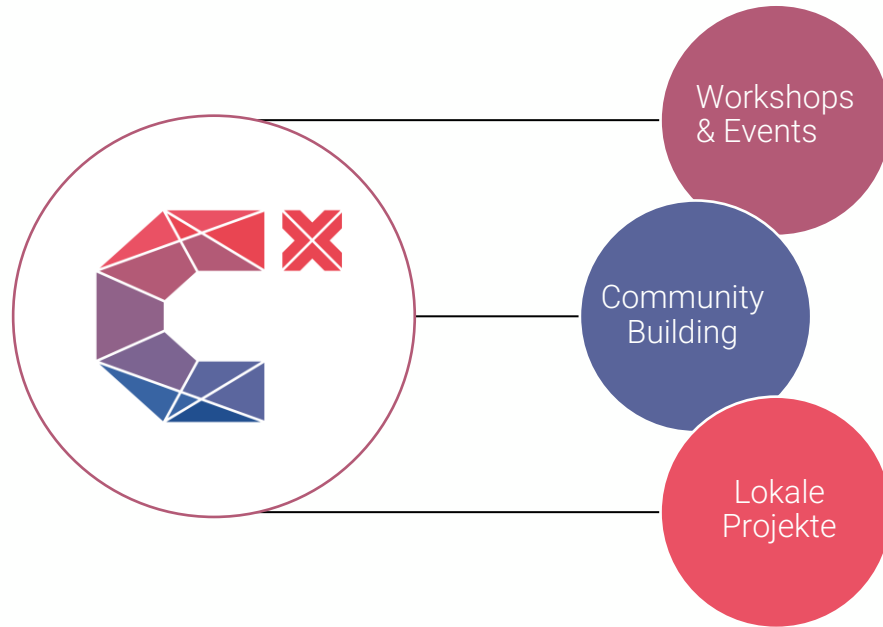
Datenbereinigung mit R

Agenda

- Was ist CorrelAid?
- Was ist CorrelAid in Konstanz?
- Datenbereinigung
 - Wofür brauchen wir das eigentlich?
 - Tidy Data
- Das Tidyverse



CorrelAid & CorrelAidXKonstanz



CorrelAid

- Netzwerk von über 1000 Datenanalyst:innen
- Studierende und Berufstätige
- Data4Good-Projekte

Local Chapter Konstanz

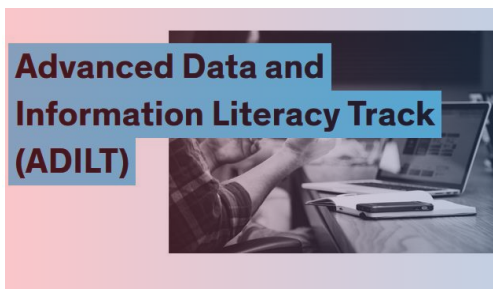
- Gründungsort!
- Workshops
- Lokale Projekte mit NPOs etc.

Unser Local Chapter



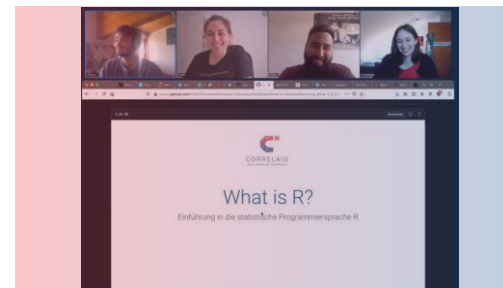
LOKALE PROJEKTE

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



LOKALE KOOPERATIONEN

In Kooperationen mit der Stadt Konstanz, dem ADILT oder auch dem SWR werden Packages geschrieben, Vorlesungen gehalten und Hackathons veranstaltet.



WORKSHOPS

An der Uni gibt es regelmäßig Workshops (bisher) zur verschiedenen Bereichen in R.

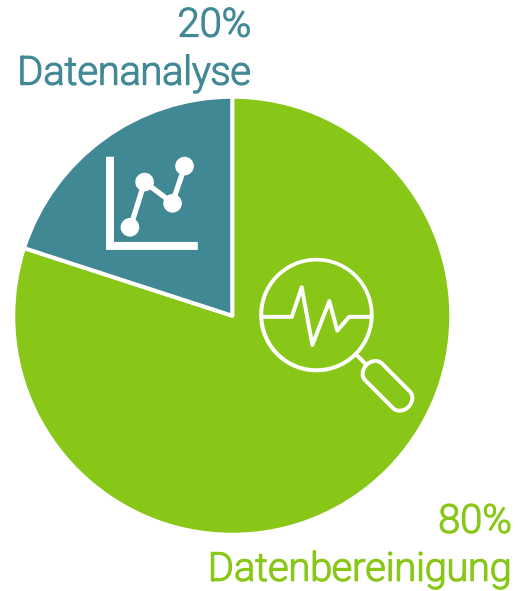
Datenbereinigung

Verbesserung der
Datenqualität

Auffinden von Fehlern in
den Daten

Beheben von
Widersprüchen in den
Daten

Umgang mit fehlenden
Daten



Das Tidyverse



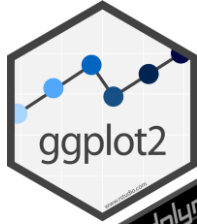
Sammlung von Packages, die aufeinander abgestimmt sind und derselben Logik folgen

Nachschlagewerk: „R for Data Science “
<https://r4ds.had.co.nz/>

Tidyverse-Philosophie:

- Verwende existierende Datenstrukturen.
- Baue einfache Funktionen mit der Pipe.
- Benutze funktionale Programmierung.
- Entwickle für Menschen.

Das Tidyverse – die Packages



Datenvisualisierung



Datenmanipulation



Datenbereinigung



Datenimport



Funktionale Programmierung



Datensätze



Zeichenfolgen



Faktoren

Das Tidyverse – die Pipe



<https://upload.wikimedia.org/wikipedia/en/b/b9/MagrittePipe.jpg>

Kombination von mehreren Operationen im Code

Vermeidung unnötiger Komplexität durch Klammern, bessere Lesbarkeit des Codes

```
package::function2(package::function1(data))
```

```
data %>%  
  package::function1() %>%  
  package::function2()
```


Das Tidyverse – die *dplyr*-Verben

filter

Auswahl von Zeilen basierend auf ihren Werten

select

Auswahl von Spalten basierend auf ihren Bezeichnungen

mutate

Erstellung neuer Variablen mit Hilfe von existierenden Variablen

arrange

Sortierung der Zeilen

group_by

Ausführen von Operationen für definierte Gruppen

summarise

Zusammenfassen von verschiedenen Werten





CORRELAID

KONSTANZ

Contact us at:



konstanz@correlaid.org



[@CorrelaidxKN](https://twitter.com/CorrelaidxKN)

or join our Slack-Channel:



[#lc-konstanz](https://join.slack.com/join/shared_invite/zt-1000000000-1000000000-1000000000)