



# Die datengetriebene Verwaltung

Daten- und Statistikkompetenz für den  
Kanton Zürich

DINAcon 2021

Andrea Schnell, Christian Ruiz, Thomas Knecht, Lars Schöbitz



Kanton Zürich

# Statistisches Amt

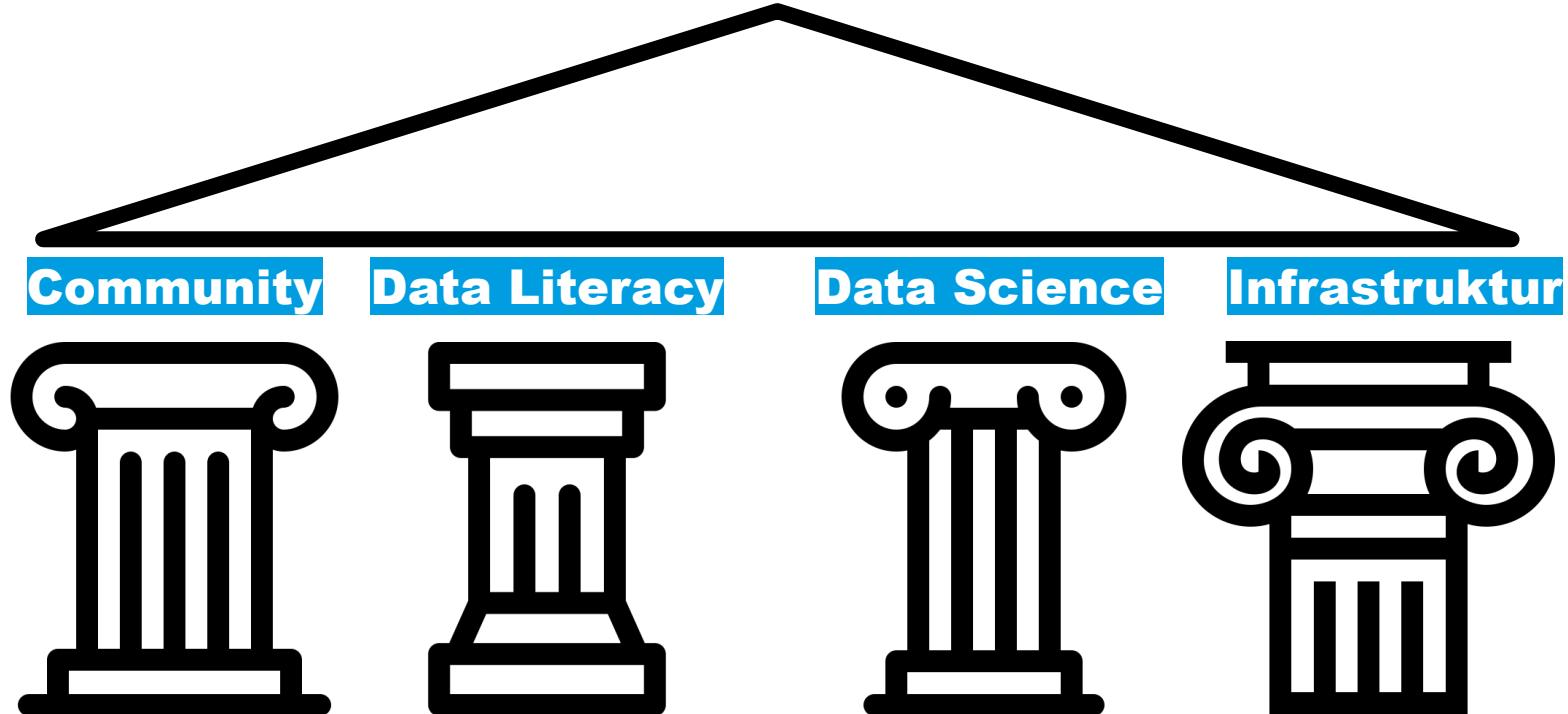
Daten...

- **Erhebung**
- **Management**
- **Analyse**
- **Diffusion**



<https://unsplash.com/@anniespratt>

# Die datengetriebene Verwaltung



A photograph of a concert stage with red and pink lights, and silhouettes of raised hands in the foreground.

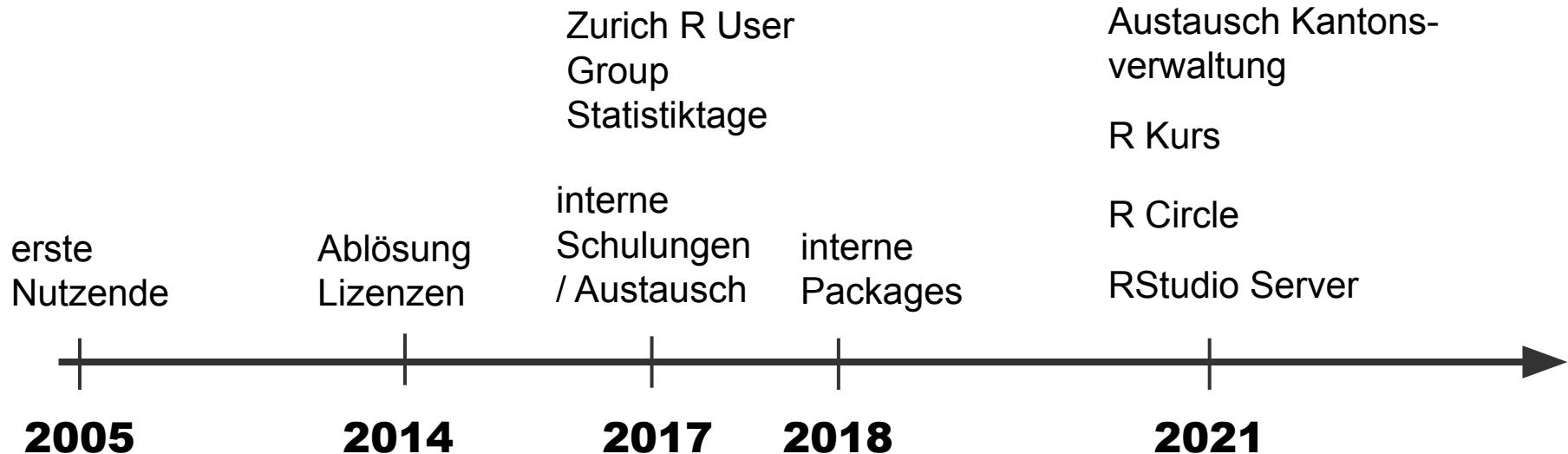
COMMUNITY

on Zürich

Community



# R im Statistischen Amt





**“The R community is one of R’s  
best features”**

Quelle: David Smith, <https://blog.revolutionanalytics.com/2017/06/r-community.html>



## STAT-intern

- Austausch ca. 4x jährlich
- interne Packages
- Anleitungen
- Infrastruktur
- Gefahr: “Einwegkommunikation”
- #rstats\_support
- ~ 15 Teilnehmer:innen





## Kantonsverwaltung

- “Kennenlernen” über Amtsgrenzen hinweg
- ca. 4x jährlich
- Vorstellen eigener R-Lösungen/ Packages/ Applikationen
- Gefahr: “Expert:innen-Gruppe”
- `#r_stat`
- ~ 25 Teilnehmer:innen



Kanton Zürich

# Teilnehmer:innen 2021



The word "stat" is written in large, bold, dark blue letters. Above the letter "t", there is a smaller "ji". To the left of the "stat" logo, the letters "stva" are written vertically above "bi-bildungsplanung" and "awel energie". Below the "stat" logo, the letters "stat ji" are written vertically above "strassenverkehrsamt ds", "gesundheitsdirektion", and "statistisches amt ji". To the right of the "stat" logo, the letters "stat dji" are written vertically above "awel" and "statistisches amt awa". At the bottom, the word "verwaltungsgericht" is written in large, bold, dark blue letters.

bi-bildungsplanung  
awel energie

stva

bildungsdirektion  
stat ji

strassenverkehrsamt ds  
awel baudirektion

ji

stat

gesundheitsdirektion  
statistisches amt ji

stat dji

awel

statistisches amt awa

verwaltungsgericht

Zürich R User Group



# R Communities

- **Zurich R User Group**  
The use of R in public administration
- **Schweizer Statistiktage**  
R in der Öffentlichen Statistik
- **adminR**  
R in Swiss Official Statistics



## Fazit & Ausblick

- Vom R User zum Contributor zur Community
- Community entsteht nicht von selbst
- Challenge R User zu Contributors zu machen
- Etablieren weiterer Communities im Kanton



# Data Literacy

**“Datenwerkstatt”**

- CO2 vs CO19
- Kurse





Kanton Zürich

# CO<sub>2</sub> versus CO<sub>19</sub>

## Do-It-Yourself CO<sub>2</sub>-Ampel



Quelle: STAT, CC0



Kanton Zürich

# CO<sub>2</sub> versus CO<sub>19</sub>

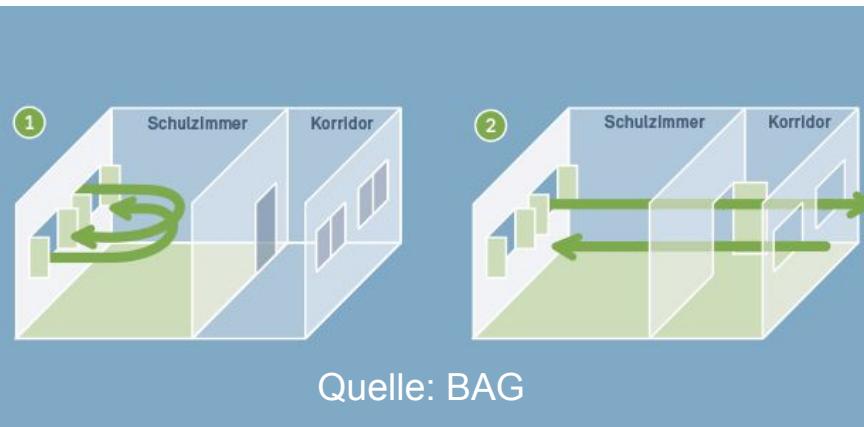


<https://unsplash.com/@kwook>

**CO<sub>2</sub>-Wert als indirekten  
Indikator für mögliche  
Virenlast im Schulraum**



# CO2 versus CO19



**CO2-Wert mit direkten  
Einfluss auf Konzentration**

**Pre-Covid BAG-Kampagne**



Kanton Zürich

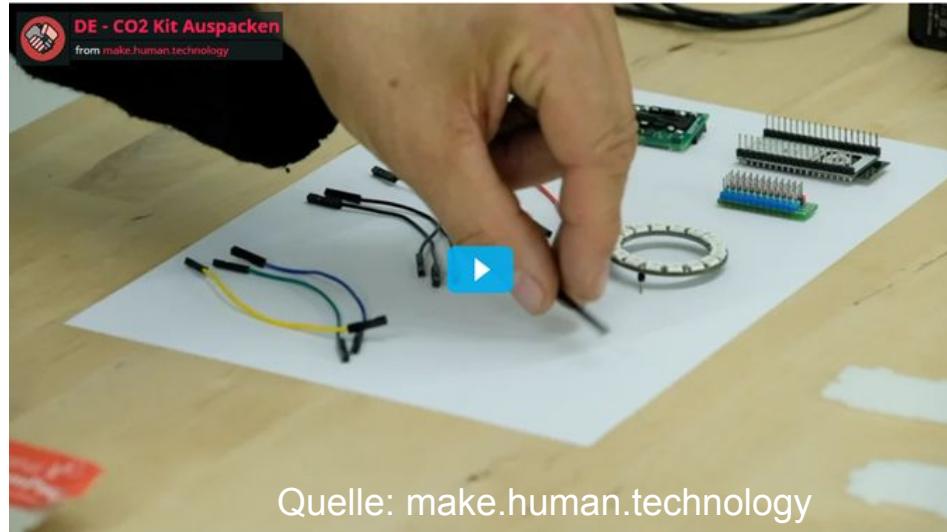
# CO2 versus CO19

## Chance

### Nicht nur Elektronik:

- **Datenkompetenz**
- **Programmierung**
- **Statistik**

### Teilhabe; Interesse



Quelle: make.human.technology



### Die Wissenspyramide anhand der CO2-Ampel verstehen

Benötigte Materialien: CO2-Ampel

Mit den Schülern und Schülerinnen kann die Wissenspyramide anhand der CO2-Ampel besprochen werden.

#### Daten:

Das Gerät misst Werte. Obwohl die Ampel uns in Farben sagt, wie es mit dem CO2 im Raum steht, so misst sie in Wirklichkeit reine Werte: 300, 500, 700, 1000. Das sind Daten, mit denen wir ohne Kontext aber nichts anfangen können.

#### Information:

Was bedeuten diese Werte? Diese Zahlen stehen für Mikrogramm CO2 pro Kilogramm Luft-Volumen. Das ist die Bedeutung dieser Rohdaten und so erschliesst sich uns ein Teil der Information.

#### Wissen:

Doch auch hier fehlt noch der Vergleich: was bedeutet nun 300, 1000 oder 2000 ppm (part per million) CO2 in der Luft? Wenn man draussen ist, dann liegt der Anteil bei ca. 400 ppm (Anmerkung für Lehrpersonen: parts per Million ist zwar eine verwirrende aber die offizielle Bezeichnung dieser Grösse). Dieser kann zwar unterschiedlich sein, je nachdem, ob man beispielsweise in einem Stadtzentrum oder in der Peripherie ist, der Wert 400ppm ist aber ein brauchbarer Durchschnitt.

#### Weisheit:

Nun hat das BAG auch Grenzwerte für die Innenluft angegeben. Man sollte ver-

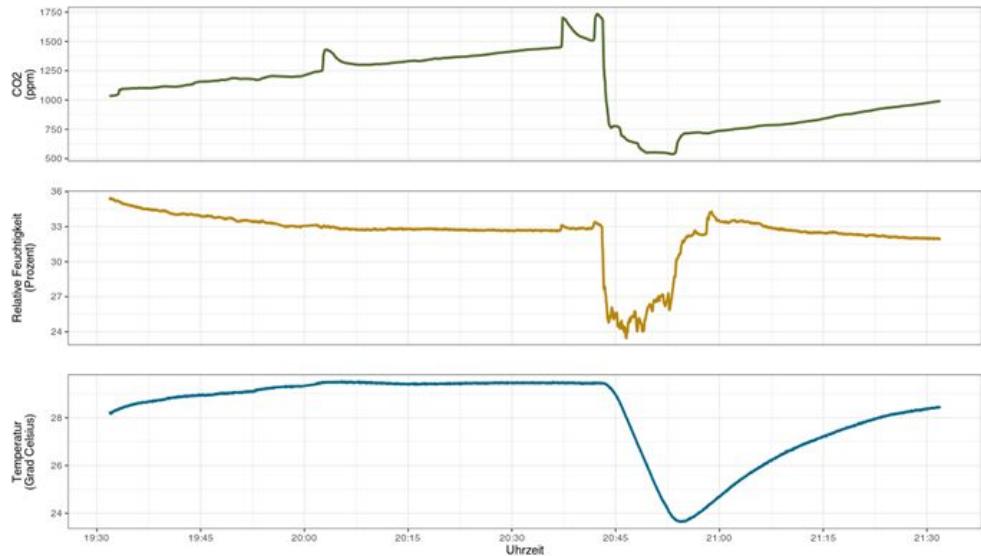
# CO2 versus CO19

**Zielpublikum:  
Lehrpersonen**

Quelle: STAT, CC0



# CO<sub>2</sub> versus CO<sub>19</sub>



Quelle: STAT, CC0



Kanton Zürich

# CO<sub>2</sub> versus CO<sub>19</sub>





Kanton Zürich

# Weiterbildungen und Schulungen



**Data  
Science  
(R und git)**



**Metadaten und  
Open Data**  
**(DCAT und Formale Leitlinien)**

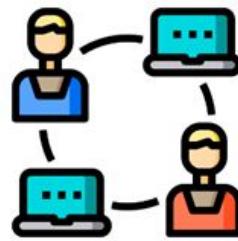


<https://unsplash.com/@tegan>

**Weitere  
Ideen**

# Aktuelle Ideen

Barrierefreiheit für Visualisierungen

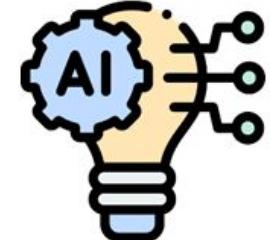
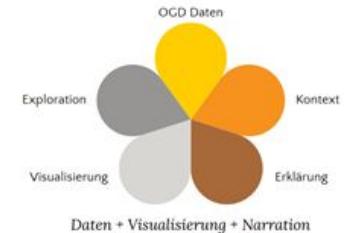


Visualisierungen (D3)

Infrastruktur für  
Datenanalysen aufbauen



Machine Learning



# Die datengetriebene Verwaltung: Daten- und Statistikkompetenz für den Kanton Zürich

rstatsZH - Data Science mit R

Lars Schöbitz | DINACon 2021 | 2021-10-29  
[lse.de/slides/dinacon2021](https://lse.de/slides/dinacon2021)

# Hoi! 🙌

Ich bin Lars  
@larnsce



Warum wird's gemacht?

# Warum wird's gemacht?

## Statistisches Amt

- Effiziente Kollaboration durch Kompetenzaufbau
- Transparente datenbasierte Entscheidungen in der Verwaltung
- Erhöhtes Verständnis für strukturierte, öffentliche Behördendaten (OGD)

## Kursteilnehmer\_innen

- Weg von Excel, SPSS und STATA -> Hin zu modernen Open Source Alternativen
- Datenvisualisierungen !!!
- Zugang zum Programmieren bekommen
- Repetitive Arbeiten effizienter lösen

# Was wird abgedeckt?



Artwork from @juliesquid for @openscapes (illustrated by @allison\_horst).

Wie wird's gemacht?

# Wie wird's gemacht?

## Kursarbeit

- Zoom
- Slack
- <https://rstatszh.github.io/website/>

## Data Science

### Programmieren

- R
- RStudio (Cloud)
- R Tidyverse Packages
- R Markdown

### Versionskontrolle und Kollaboration

- Git
- GitHub

# Wie wird's gemacht?

- **Unterricht:**
  - 7 Wochen á 3 Stunden
- **Lehrmethoden:**
  - Live Coding
  - Pair Programming
- **Hausaufgaben:**
  - Material zum Üben und Vertiefen

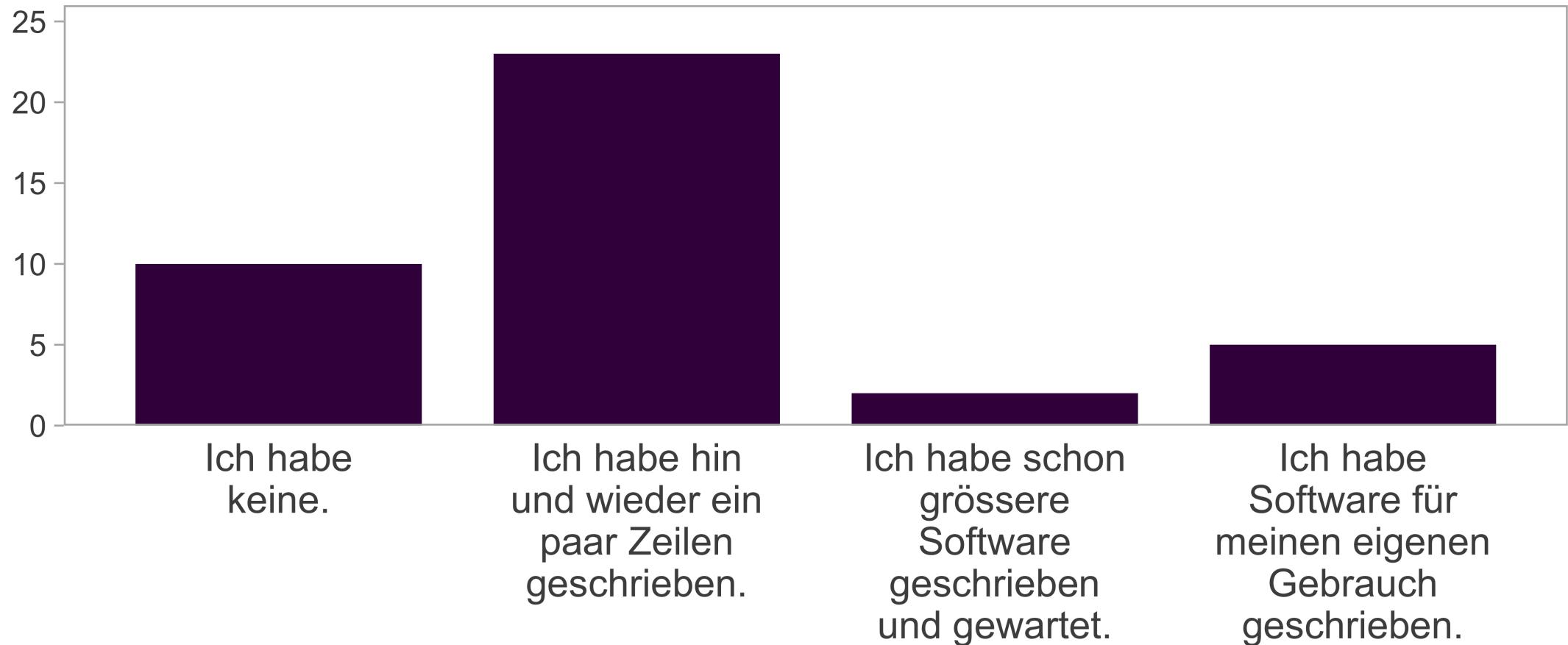
# Wie wird's gemacht?

- **Unterricht:**
  - 7 Wochen á 3 Stunden
- **Lehrmethoden:**
  - Live Coding
  - Pair Programming
- **Hausaufgaben:**
  - Material zum Üben und Vertiefen
- **Persönliches Tutoring:**
  - Verfügbar auf Slack/Zoom
- **Abschluss Projekt:**
  - Arbeitsumgebung aufbauen
  - Mit eigenen Daten arbeiten
  - Das Gelernte anwenden
  - Gezielte Unterstützung bekommen
- **Zertifikat:**
  - Bestätigung über Teilnahme

Wer ist dabei?

# Anzahl Teilnehmer\_innen mit Erfahrung im...

...Programmieren im Allgemeinen



# Anzahl Teilnehmer\_innen mit Erfahrung im...

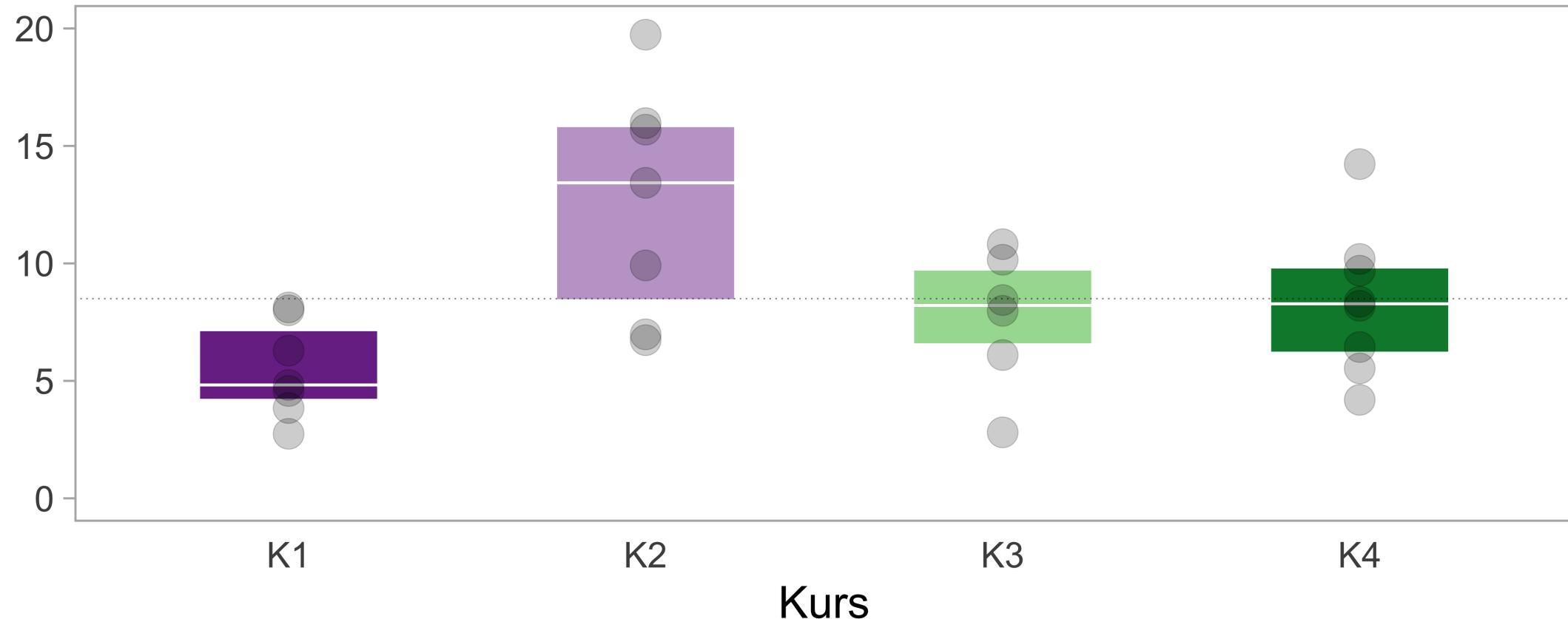
...Programmieren mit R



# Wieviel wurde programmiert?

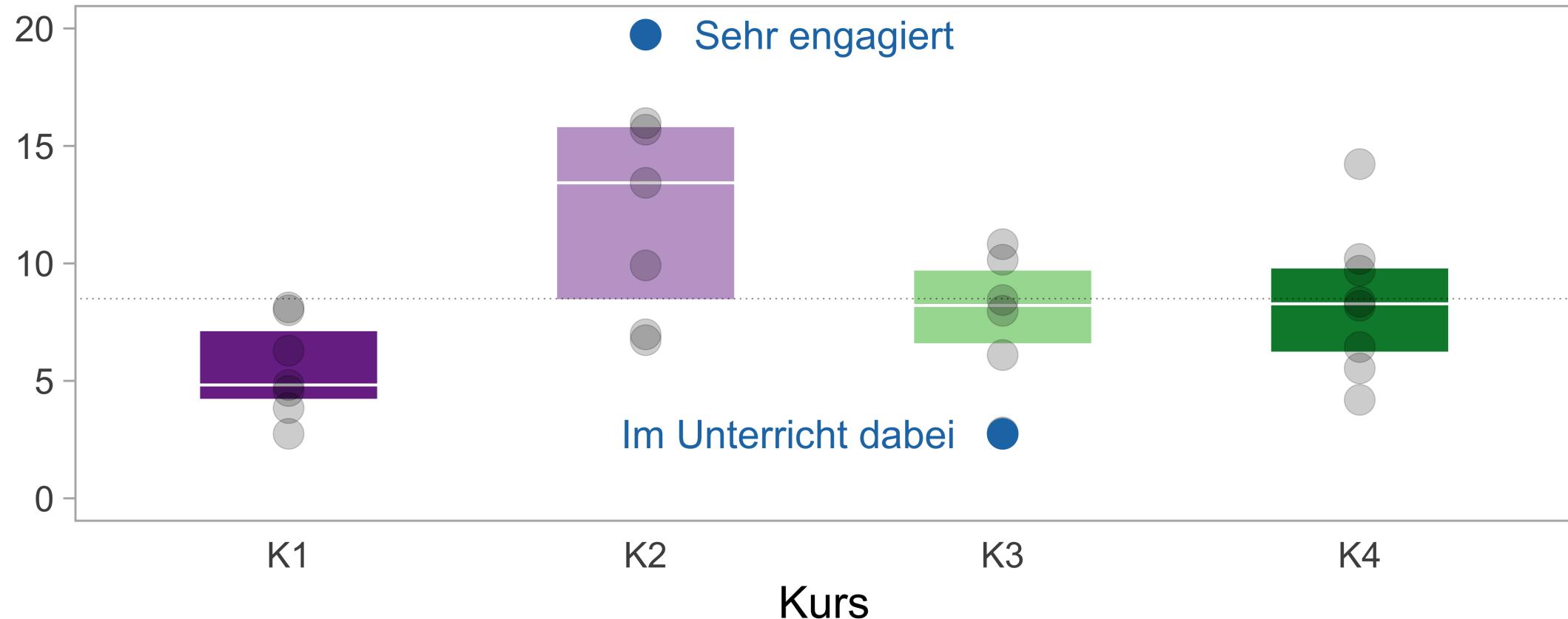
# Arbeitsstunden in der RStudio Cloud

Im Durchschnitt arbeitet jede\_r Teilnehmer\_in 8.5 Stunden/Woche mit R



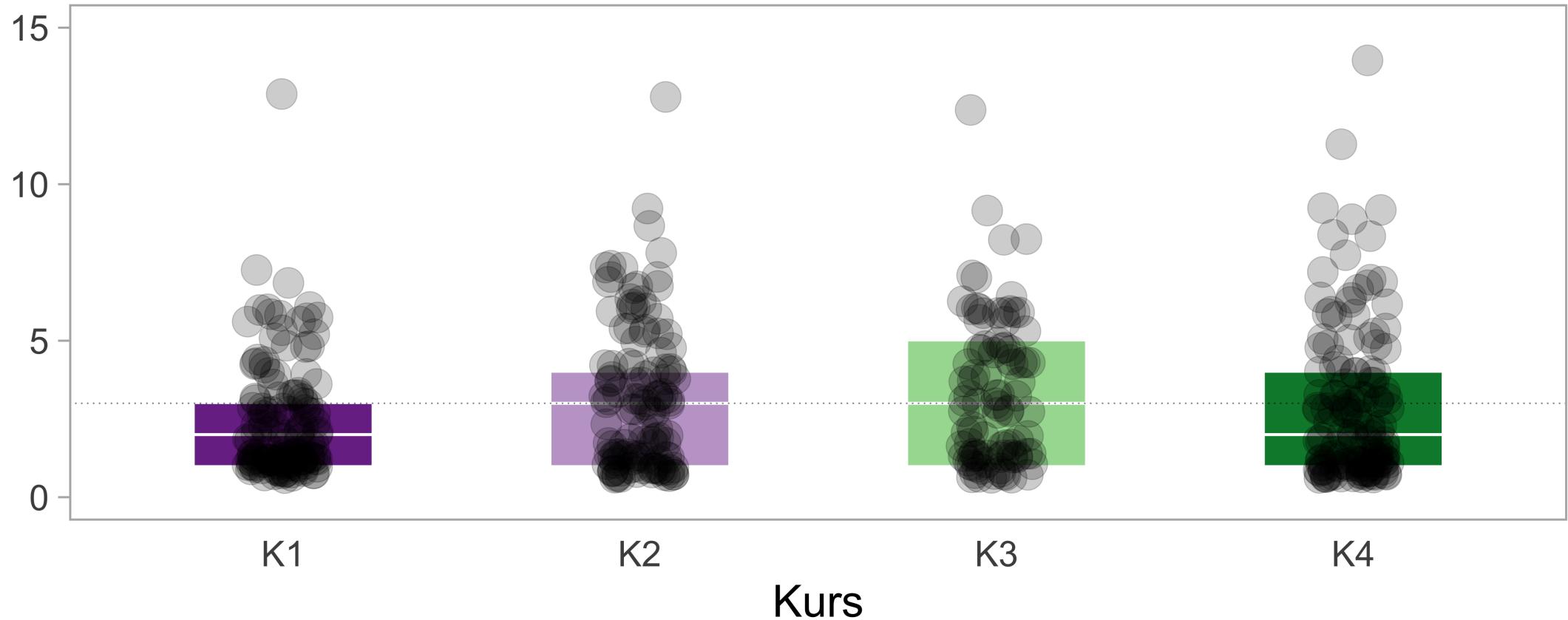
# Arbeitsstunden in der RStudio Cloud

Im Durchschnitt arbeitet jede\_r Teilnehmer\_in 8.5 Stunden/Woche mit R



# Anzahl Commits auf GitHub Repos (Übungen)

In 508 Repos wurden im Schnitt 3 Commits pro Repo gemacht



Was braucht's noch?

# Was braucht's noch?

- Praxisnahe Case Studies
- Nachschlagewerk für Funktionen und Konzepte
- Vertiefung im Arbeiten mit git und GitHub
- Community Einbindung der Teilnehmer\_innen





Danke

Für die Aufmerksamkeit!

Für die R packages `{xaringan}` und `{xaringanthemer}`

**Folien:** [lse.de/slides/dinacon2021](https://lse.de/slides/dinacon2021) (Download Folien als PDF Download)

**Code und Daten für Folien auf GitHub:**

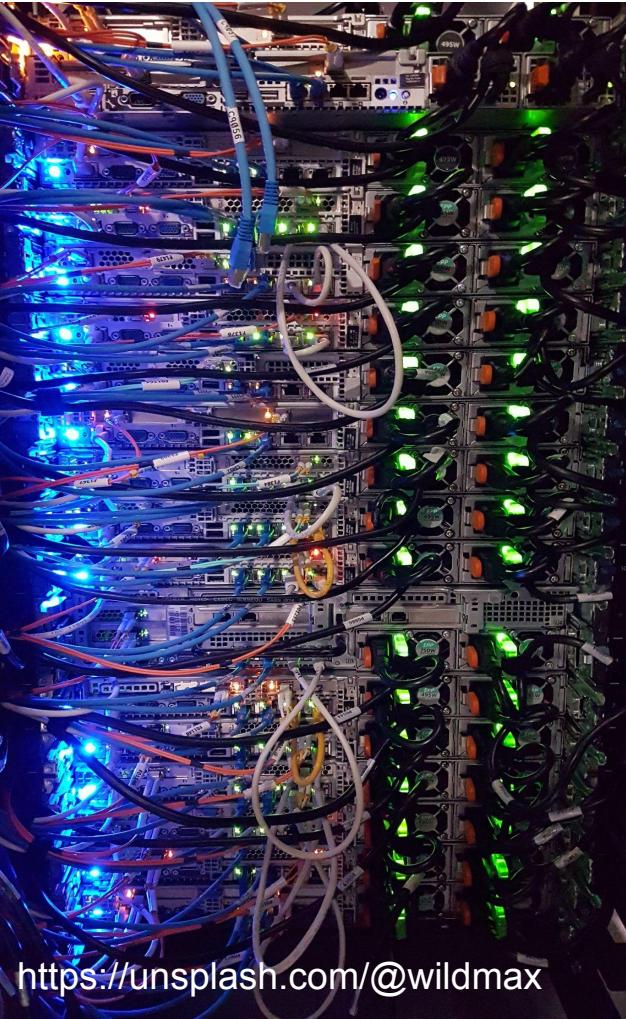
<https://github.com/larnsce/DINAcon2021>

**Kurswebsite:** <https://rstatszh.github.io/website/posts/2021-02-17-ueberblick/>

**Email:** [Lars@Lse.de](mailto:Lars@Lse.de)

**Twitter:** [@larnsce](https://twitter.com/larnsce)

**License:** CC BY 4.0



Kanton Zürich

# Infrastruktur

Dream  
Big



Kanton Zürich



**R ist eines der  
meistgenutzten  
Programmier-  
sprachen für Data  
Science im  
Kanton....**



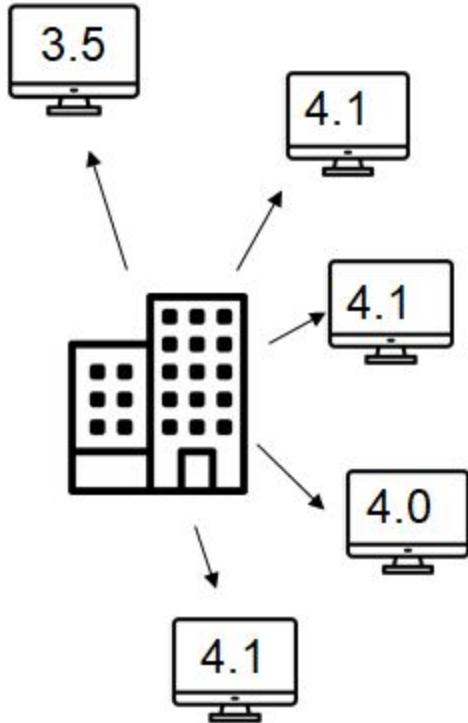
# Konkrete Ziele

- Alle die wollen sollen R einfach und sofort nutzen können
- Einheitliche R-Versionen und stabile Umgebung für einfache Zusammenarbeit
- Regelmässige Updates um mit der Zeit zu gehen (Security / Innovation etc.)
- IDLE zur Verfügung stellen, welche eine einfache Data Science ermöglicht (im Kanton ist RStudio am Verbreitesten)

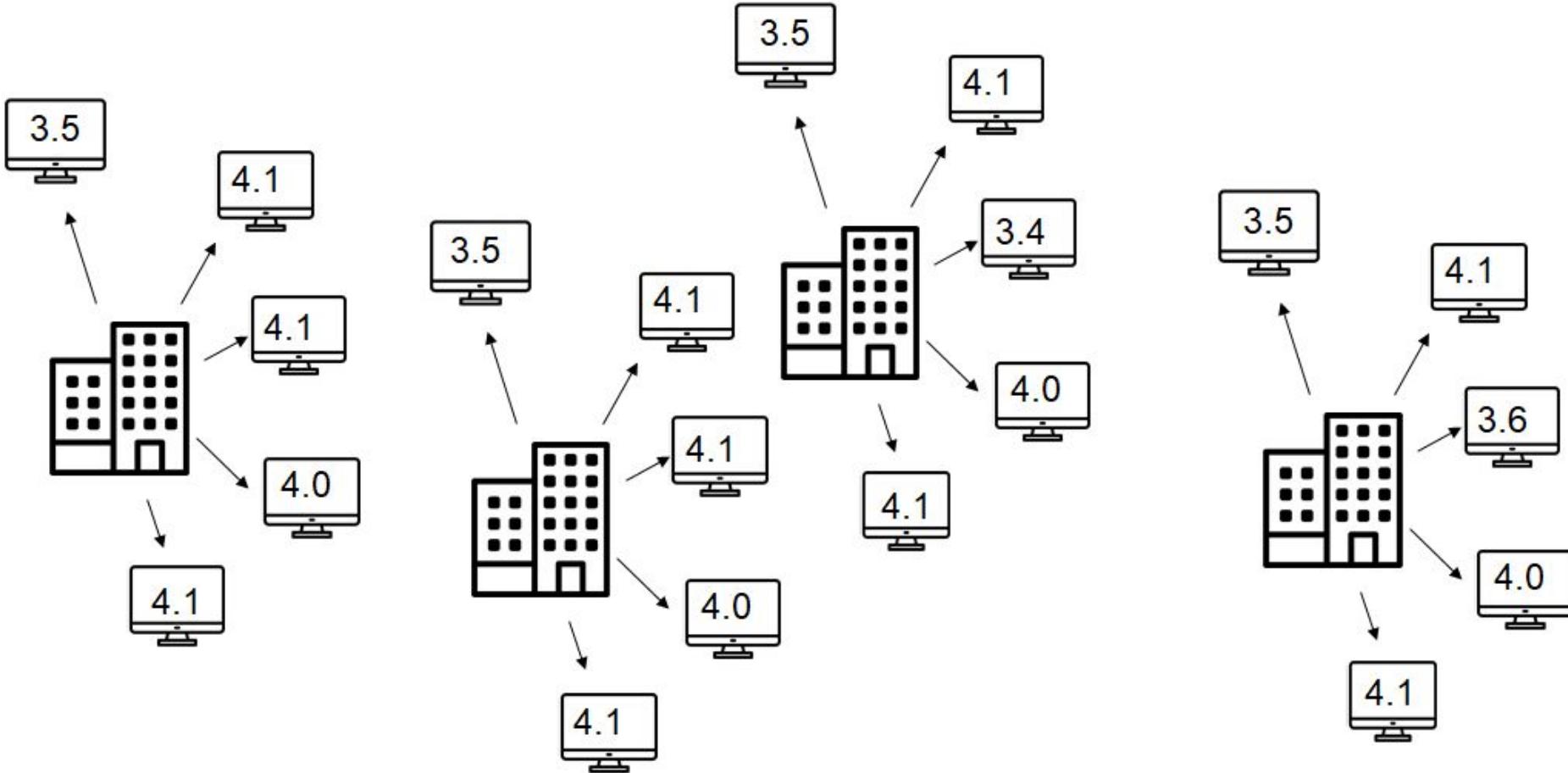


# R und RStudio als Fachapplikation

- Installation und Updates auf Anfrage
- Bei IT oft nur wenig Wissen vorhanden bezüglich R
- Unterhalt liegt bei den Usern (einrichten von proxys, DB-Verbindungen usw.)



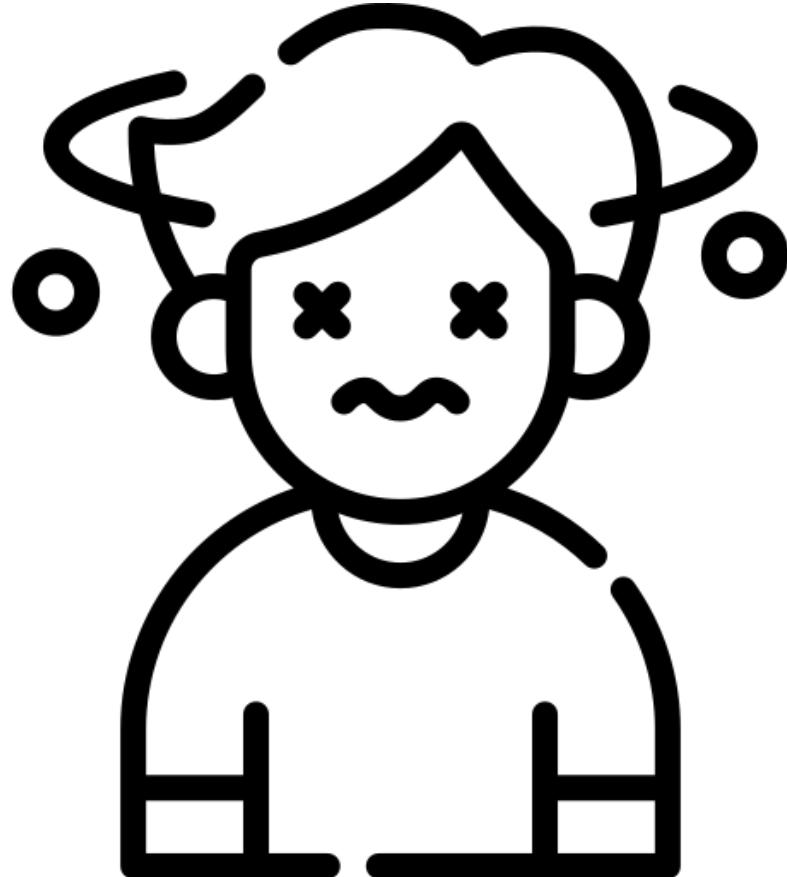
**Was bedeutet  
das...**



Icons made by Freepik and xnimrodx from flaticon



Kanton Zürich

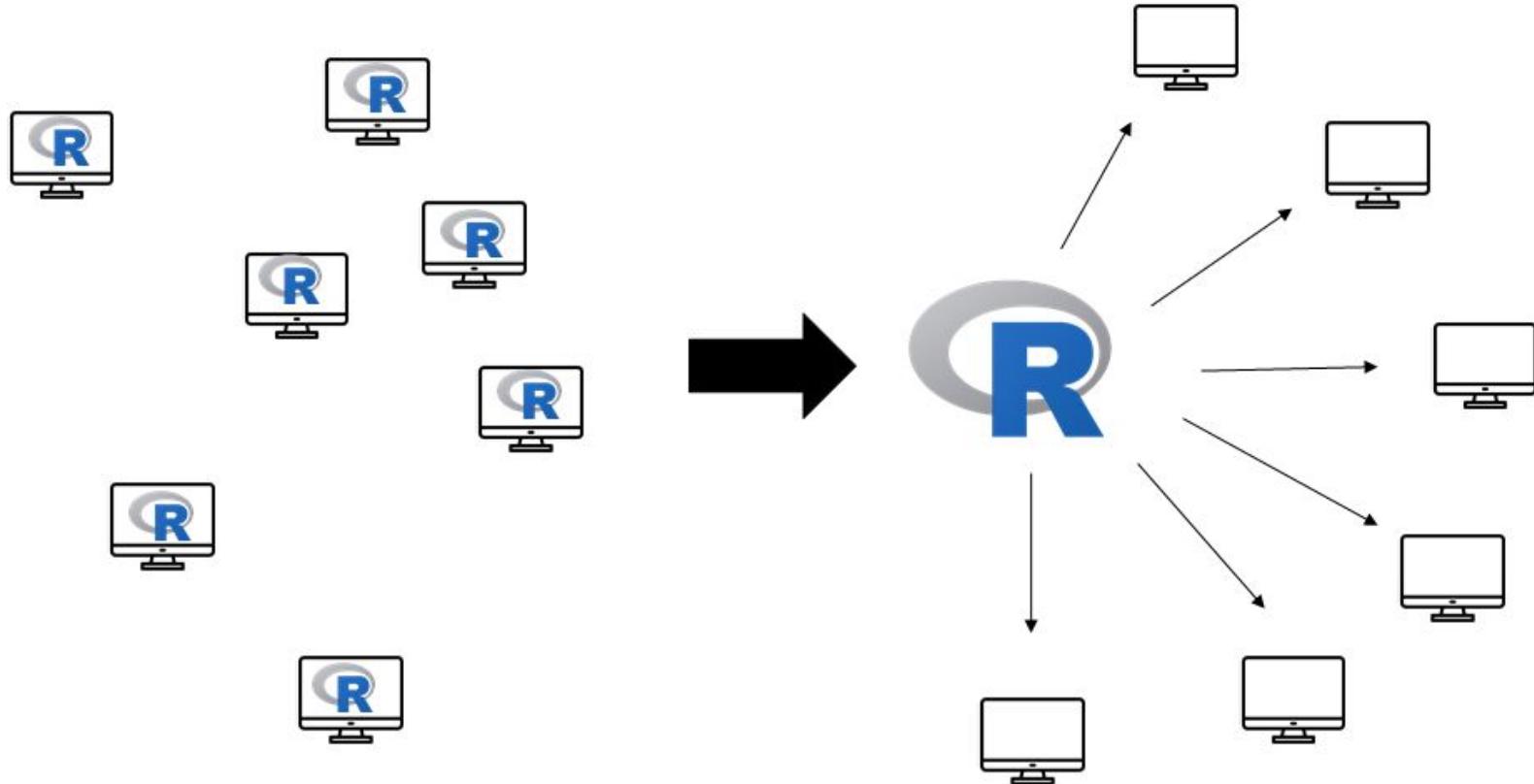


**Gar nicht so  
einfach da den  
Überblick zu  
kriegen...**



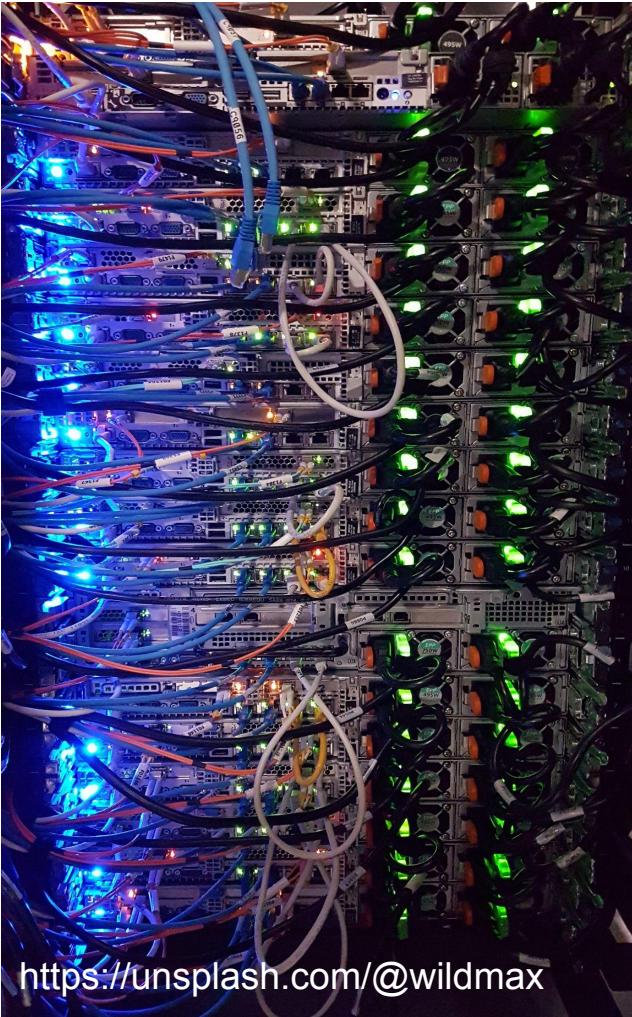
# Deshalb...

- Schritt für Schritt die R-Infrastruktur zu verbessern versuchen
- Bewusstsein dafür wecken, dass R ein zentrales Instrument ist (sowohl in Amtsleitungen als auch seitens IT)
- Inkrementelles Vorgehen: beim Statistischen Amt anfangen und skalieren





- RStudio-Server ermöglicht eine Zentrale Verwaltung der Installation
- Externe Fachleute kümmern sich um Updates und den Unterhalt





# Erste Erkenntnisse

- Wenn der Server nicht geht, dann geht nichts mehr...
- komplexe Einbindung in die IT
- Der Server ist nicht dynamisch skalierbar in unserer Installation
- Wenn man den Usern viel RAM gibt, dann wird die auch gebraucht



# Das heisst nun....

... Ideal wäre eine Server basierte Lösung

- kantonsweite Userverwaltung
- Einfach Skalierbar
- Backup-Systeme benötigt
- Einbindung von unterschiedlichen File-Systemen und DB's



# Der Weg dahin ist noch weit deshalb....

- Versuchen wir die Vereinheitlichung der Client Installation nebenbei voranzutreiben
- Direkt beeinflussen können wir es in unserer Direktion
- Indirekt können wir über den R-Austausch unsere Erfahrungen mit den anderen Direktionen teilen