

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

rstatsZH - Data Science mit R

DINAcon 2021 - Lars Schöbitz - lse.de

2021-10-29

# Warum?

# Warum? (Bisher nur Outline)

## STAT

- Kompetenzen im Umgang mit R und git/GitHub ausbauen um Kollaboration effizienter zu machen
- Analytische Kompetenzen ausbauen um bessere datenbasierte Entscheidungen zu treffen
- Verständnis von strukturierten öffentlichen Daten (OGD) erhöhen um Anzahl der Publikationen voranzutreiben

## Kursteilnehmer\_innen

- Weg von Excel, SPSS und STAT -> Hin zu modernen Open Source Alternativen
- Datenvisualisierungen
- Effizienteres Arbeiten und Zugang zum Programmieren bekommen

Was?



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

# Was? - Lernziele

Am Ende des Kurses können Teilnehmer\_innen

1. (Daten)projekte reproduzierbar publizieren
2. Kollaborativ mit Git und GitHub arbeiten
3. Das Konzept von Tidy Data verstehen und anwenden
4. Explorative Datenanalysen durchführen
5. Sich selbst und Anderen helfen mehr über die grosse Welt der Möglichkeiten für Data Science mit R zu lernen

Wie?

# Wie? - Werkzeugkiste

## 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? - Kursüberblick (vielleicht weglassen)

Sieben Kurseinheiten über sieben Wochen

- **Unterricht:** Wöchentlich 3 Stunden
- **Übungen:** Übungen während der Kurseinheit
- **Hausaufgaben:** Einzureichen bis zum nächsten Kurstag
- **Persönliches Tutoring:** Sprechstunden ausserhalb der Kurstage
- **Abschluss Projekt (freiwillig):**
  - Eigene Arbeitsatmosphäre aufbauen
  - Arbeit mit eigenen Daten
  - Bericht erfassen
  - 5 Minuten Präsentation
- **Zertifikat:** Bestätigung über Teilnahme

# Wer?

## Welche Antwort beschreibt deine Erfahrung mit dem Programmieren im Allgemeinen?

49 responses



## Welche Antwort beschreibt deine Erfahrung mit dem Programmieren mit R?



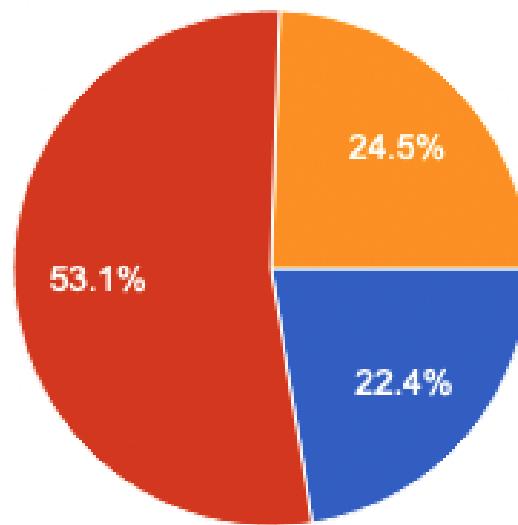
49 responses



Welche Antwort beschreibt am besten, wie leicht du ein Programm in einer beliebigen Sprache schreiben könntest, um die größte Zahl in einer Liste zu finden.



49 responses



- Ich wüsste nicht, wo ich anfangen sollte.
- Ich könnte mich mit Trial und Error und vielen Web-Suchen durchkämpfen.
- Ich könnte es schnell und mit wenig oder gar keiner zusätzlichen Hilfe erledigen.

## Text Analyse auf Frage:

"Was möchtest du nach diesem Kurs wissen oder machen können, dass du momentan nicht weisst oder nicht machen kannst?"

- Was sind die Top 5 Begriffe die genannt werden?
- Welche zwei Worte werden am häufigsten zusammen genannt?

Wieviel?

# Wieviel wurde programmiert?

Ich habe Nutzungsdaten aus der RStudio Cloud, aber momentan nur als Dashboard. Versuche an die Rohdaten zu kommen und folgende Indikatoren zu zeigen:

- Anzahl Stunden pro Teilnehmer\_in und (mean, sd) welche zusätzlich zu den 3 Kurstunden pro Woche aufgewendet wurden
- Zeitreihe der Summe Stunden pro Woche über alle Kurse

Falls ich nicht an Daten komme, dann berechne ich ein paar Kennzahlen selbst.

# Evaluation / Zufriedenheit

## Evaluation - wöchentlich

- Zeige Ergebnis einer Frage zum Inhalt der Kurseinheit
- **Plot wird mit R reproduziert**



Danke

Für die Aufmerksamkeit!

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

**Folien als PDF:**

**Folien als HTML:**

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

**Email:** Lars@Lse.de

**Twitter:** @larnsce