



Dr. Lars Pelke

Universität Heidelberg, Institut für Politische Wissenschaft

# Praxiskurs Datenanalyse und Replikation

Sitzung 10 Replikationscodes veröffentlichen



# Leitfragen und Lernziele

- Reproduktionsmaterialien zur eigenen Replikationsstudie erstellen
  - Wie sollte die Readme-Datei aussehen?
  - Überprüfen der eignen Ordnerstruktur
  - Rechtliches zur Datensatzpublikation
- Übersicht Repositorien zur Archivierung von Replikationsmaterialien
- Übersicht *research-transparency requirements* bei politikwissenschaftlichen Zeitschriften

# Eigene Replikationsmaterialien erstellen



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

## Wrap-Up: Auf was achten?

Plenumsgespräch



# Eigene Replikationsmaterialien erstellen

## Wrap-Up: Auf was achten?

- Ordnerstruktur klar und nach den erarbeiteten Vorgaben?
- Informative readme-Datei
- Gut-dokumentierter und nutzerfreundlicher Code
- Benennung von Software und Packages, die notwendig zur Replikation sind
- Daten zugänglich?

“authors must begin their work with replication in mind” (Alvarez and Heuberger 2021, 1)



-> Wir haben bei der Replikation begonnen, hatten daher stets die replication unserer Ergebnisse im Blick (hoffentlich ;))



# Eigene Replikationsmaterialien erstellen

## Readme-Datei erstellen (Leitlinien)

- Wird auf der obersten Ordner Ebene gespeichert
- Muss alle Informationen enthalten damit Nutzer:innen den Code zum Laufen bringen
  - Liste aller Order/Unterordner und data files
  - Hardware Spezifikationen die genutzt worden sind
  - Software, welche genutzt worden ist
  - Liste aller Code Files, die notwendig zur Reproduktion der Originalresultate sind
  - Die geschätzte *running time* für jedes File

```
README.txt - Editor
Datei Bearbeiten Format Ansicht Hilfe
Title: Replication File for "Conceptualizing and Measuring Autocratization Episodes"
Journal: Swiss Political Science Review
Author: Lars Pelke, Aurel Croissant
Correspondence: Lars Pelke, Heidelberg University (lars.pelke@ipw.uni-heidelberg.de)
Date: January 10th, 2021

Required Software: R and RStudio (3.6.1 or Above).
Additional required software packages and libraries are specified in the replication code.

Overview: These files replicate all analyses in Lars Pelke and Aurel Croissant, "Conceptualizing and Measuring
To replicate all analyses and figures, download all data files to a folder entitled "~/data", create a second

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Replication Code %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

01_data_wrangling
Replicates the data wrangling process and figures in the research note. In addition, it compiles data need for

02_data_preparation_shiny_web_app
Replicates the data preparation for the Shiny Web Application (Online Interactive Supplementary Appendix)

03_WebApp
Replicates the Shiny Web Applicatio. The Shiny Web Application runs will with Google Chrome Browser

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Data Files (relative path "~/data") %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Replication data for research note:
-vdem_10 folder: Varieties of Democracy Version 10, Coppegde et al. (includes Polity IV and Freedom House data
-subdatasets are created when using 01_data_wrangling

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Output Files (relative path "~/outputs") %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

Store outputs, such as Figures
|
```

Leitlinien nach Alvarez and Heuberger 2021

Beispiel aus Pelke und Croissant (2021)



# Eigene Replikationsmaterialien erstellen

## Eigenes Readme-Datei erstellen

**Erstellen Sie Ihr jeweilige Readme Datei.**

Das sollte nicht mehr als 20min in Anspruch nehmen. Orientieren Sie sich an Alvarez und Heuberger (2021) und an der Originalstudie, die Sie repliziert haben.



# Eigene Replikationsmaterialien erstellen

## Rechtliches zu Replikationsdateien

- Publikation von Sekundärdatensätzen kann unter Umständen urheberrechtlich problematisch sein:
  - Unproblematisch wenn Sekundärdatensätze Open Access lizenziert oder Erlaubnis zur Publikation erteilt ist (an entsprechende Referenzen denken).
  - Ferner problematisch, wenn Daten zur Identifikation von Personen enthalten sind (es darf keine Identifikation möglich sein, bspw. CAVE georeferierte Surveydaten)
  - Auf jeden Fall müssen alle Replikationsfiles enthalten sein, damit ggf. urheberrechtlich geschützte Datensätze von Nutzer:innen erstellt werden könne (meine Benennung häufig: *01\_data\_wrangling*), wenn Nutzer:innen diese geschützten Daten selbst heruntergeladen haben.

# Übersicht data repositories



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

## Harvard Dataverse

- Langzeitarchivierung mit DOI
- Keine Versionskontrolle
- von vielen Journal unterstützt



## Open Science Framework

- Langzeitarchivierung mit DOI
- Versionskontrolle
- Auch Upload von Preprints



## Github /Git

- Langzeitarchivierung
- Versionskontrolle
- Kommt aus der Softwareentwicklung
- Signalling-Effekt





# Upload der Replikationsfiles in der HeiBox/Dropbox



UNIVERSITÄT  
HEIDELBERG  
ZUKUNFT  
SEIT 1386

Laden Sie Ihre Replikationsfiles inklusive der Readme-Dateien, klarer Benennung und sauberer Ordnerstruktur in unsere Kursinterne Dropbox/HeiBox hoch (als .zip-Datei).

Aufgabe für nächste Woche: Jede:r von Ihnen lässt mindestens eine Code Ihrer Kolleg:innen durchlaufen und schaut ob alles funktioniert. (siehe Liste *Materialien/list\_replication\_peers.docx*)

# Übersicht Zeitschriften mit research-transparency requirements



## Partner:innen-Aufgabe

- Zu zweit suchen Sie für mindestens fünf Journals aus Ihrem Fachgebiet nach *research-transparency requirements* und dokumentieren die Anforderungen. Überlegen Sie sich ggf. eine Klassifikation der Anforderung. Mögliche Punkte
  - Materialien müssen zugänglich sein
  - Materialien werden über das Journal zugänglich gemacht
  - Skripte und Daten werden durch Journal oder Beauftragten repliziert



# Vorbereitung nächste Sitzung (Woche 11)

- Bereiten Sie die Liste der Journals mit den jeweiligen *replication transparency policies* vor
- Überlegen Sie sich für Ihre Replikationsstudie mindestens zwei Titelvorschläge und kondensieren Sie Ihr Hauptargument!