

Codebooks für OpenPetition Datensätze

und ein Überblick über Ordner-/Dateienstruktur ‘data’

Contents

1	Überblick Ordner-/Dateienstruktur	2
2	Codebook 4_liste_in_zeichnung.csv	4
	url	4
	title	4
	from	4
	to	4
	region	4
	status	4
	target_support	4
	category	4
	perc_reached	4
	supporters_total	5
	supporters_for_quorum	5
	petition_text	5
	stat_url	5
	date	6
	sig_total	6
	sig_quorum	6
	id	6
	date	7
	new_sig_total	7
	new_sig_quorum	7
	id	7

Dies ist ein Codebook für die OpenPetition Daten. Zudem eine kurze Erläuterung zur Ordnerstruktur/Dateienstruktur.

1 Überblick Ordner-/Dateienstruktur

- Daten und die R Files im Unterordner “data”.
- Datenerhebung erfolgte grob gesagt in fünf Schritten -> Prefix gibt Aufschluss über Position der Datei in Workflow
 - Prefix 0 -> Master R File, in dem alle anderen R Files gesourct werden
 - csv files mit gleichem Prefix wie R File sind “Output” des R Files
 - anderer Output können html files sein, die gedownloadet wurden (s.u.)
- **relevant für Analyse:** “4_liste_in_zeichnung_clean.csv”-Datensatz in Verbindung mit den zwei Datensätzen im “statistik_und_karten”-Unterordner.
- Alphabetisch geordnete Codebooks für diese drei Datensätze findet ihr in Sektion 2, 3 und 4.

Schritt	.R	Output Files	Beschreibung
0	0_data_master	/	Master R File htmls zu Petition- slistenseiten Metadaten für jede Petition (name, url)
1	1_scrape_liste		
		html files in Unterordner listen 1_liste_in_zeichnung.csv	
2	2_get_html_files	html files in Unterordner html_files 2_liste_in_zeichnung_withid.csv	htmls zu jeder Petition fügt id Variable hinzu
3	3_scrape_liste	3_liste_in_zeichnung_scraped.csv	Daten zu jeder Petition
4	4_clean_data	4_liste_in_zeichnung_clean.csv	säubert Daten (teilweise)
5	5_get_statistik_und_karten_data	in Unterordner statistik_und_karten: 5_ganzer_zeitraum.csv 5_letzte_2_wochen.csv	für jede Petition: Anzahl Unterschriften ganzer Zeitraum (Graph 1) Anzahl Unterschriften letzte zwei Wochen

2 Codebook 4_liste_in_zeichnung.csv

Hinweis: Wenn ihr nicht wollt, dass die character-Variablen in factors umgewandelt werden, benutzt die Option “stringsAsFactors = FALSE” bei read.csv(). ## id {-} - petitionsspezifische id Variable - zum Mergen von Statistik-und-Karten-Datensätzen zu 4_liste_in_zeichnung.csv und umgekehrt

url

- URL zur Seite der Petition

title

- Titel der Petition

from

- Petitionssteller

to

- an wen sich Petition richtet

region

- Region der Petition
- mögliche Werte auch: “Deutschland”, “Schweiz”, “Österreich”

status

- Status der Petition
- für alle “in Zeichnung”, da wir nur aktuelle Petitionen gescrapt haben

target_support

- für Petitionen mit Quorum: angestrebtes Quorum
- für Petitionen ohne Quorum (d.h. nur mit Sammlungsziel): angestrebtes Sammlungsziel

category

- Kategorie der Petition
- NA: keine Kategorie angegeben

perc_reached

- erreichte Prozente von target_support

supporters__total

- Unterstützer gesamt

supporters__for__quorum

- Unterstützer, die für das Quorum zählen
- NA für Petitionen mit Sammelziel

petition__text

- Petitionstext

stat__url

- URL zur Statistik & Karten Seite

Codebook 5_ganzer_zeitraum.csv {#sec51} Dieser Datensatz enthält für jede Petition und jeden Tag die absolute Anzahl an Unterschriften (total und für-Quorum-zählend) seit Petitionsstart.

date

- Datum

sig_total

- Anzahl alle Unterstützer zu Datum

sig_quorum

- Anzahl aller Unterstützer, die für Quorum zählen, zu Datum

id

- petitionsspezifische ID Variable

Codebook 5_letzte_2_wochen.csv {#sec52} Im Unterschied zu 5_ganzer_zeitraum.csv, sind in diesem Datensatz für jede Petition und jeden Tag in den letzten zwei Wochen die **neu hinzugekommenen** Unterschriften (total und für-Quorum-zählend) enthalten.

date

- Datum

new_sig_total

- Anzahl neue Unterstützer zu Datum

new_sig_quorum

- Anzahl neuer Unterstützer, die für Quorum zählen, zu Datum

id

- petitionsspezifische ID Variable