

gesis

Leibniz Institute
for the Social Sciences



Social-Media und Text-Mining mit R

Veronika Batzdorfer

Teil III

*Twitter Demo & Exkurs: Crawling Social-Web-Data,
29.09.2022*

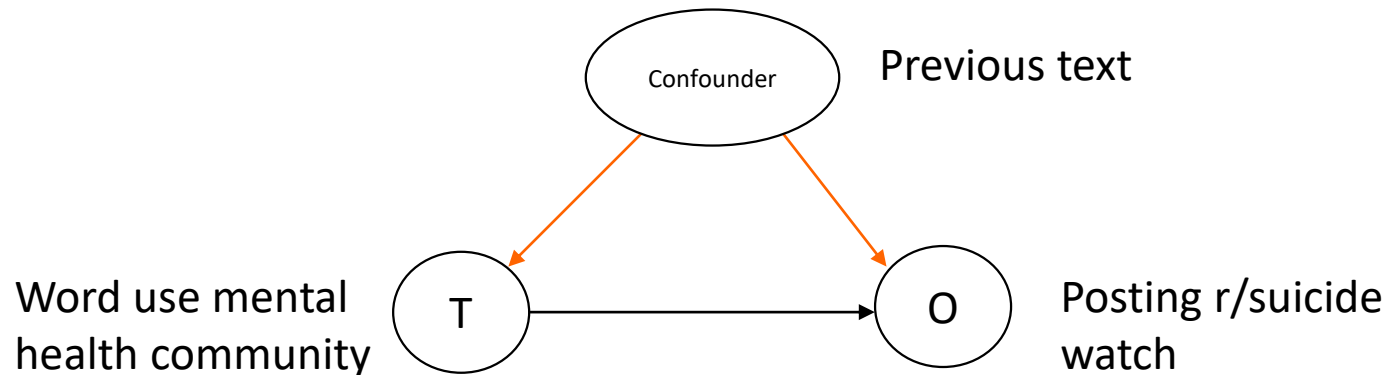
Schedule

Uhrzeit	Inhalt
09:00 - 10:30	Konzepte & Herausforderungen bei der Analyse von Social-Web-Daten (Twitter-API)
10:30 - 11:00	<i>Kaffeepause</i>
11:30 - 12:30	Getting Started mit Twitterdaten: (i) Sampling, (ii) Pre-Processing & (iii) Grundlagen der Textanalyse (Häufigkeiten, Co-Occurences, Netzwerke)
12:30 - 13:30	<i>Mittagspause</i>
* 13:30 - 15:00	Twitter Demo & Exkurs Crawling Social-Web-Data
15:00 - 15:30	<i>Kaffeepause</i>
15:30 - 17:00	Ausblick: Fortgeschrittene NLP-Techniken (z.B. Topic Modelling) & Social-Web-Data-Collection; Bias und Ethik im NLP

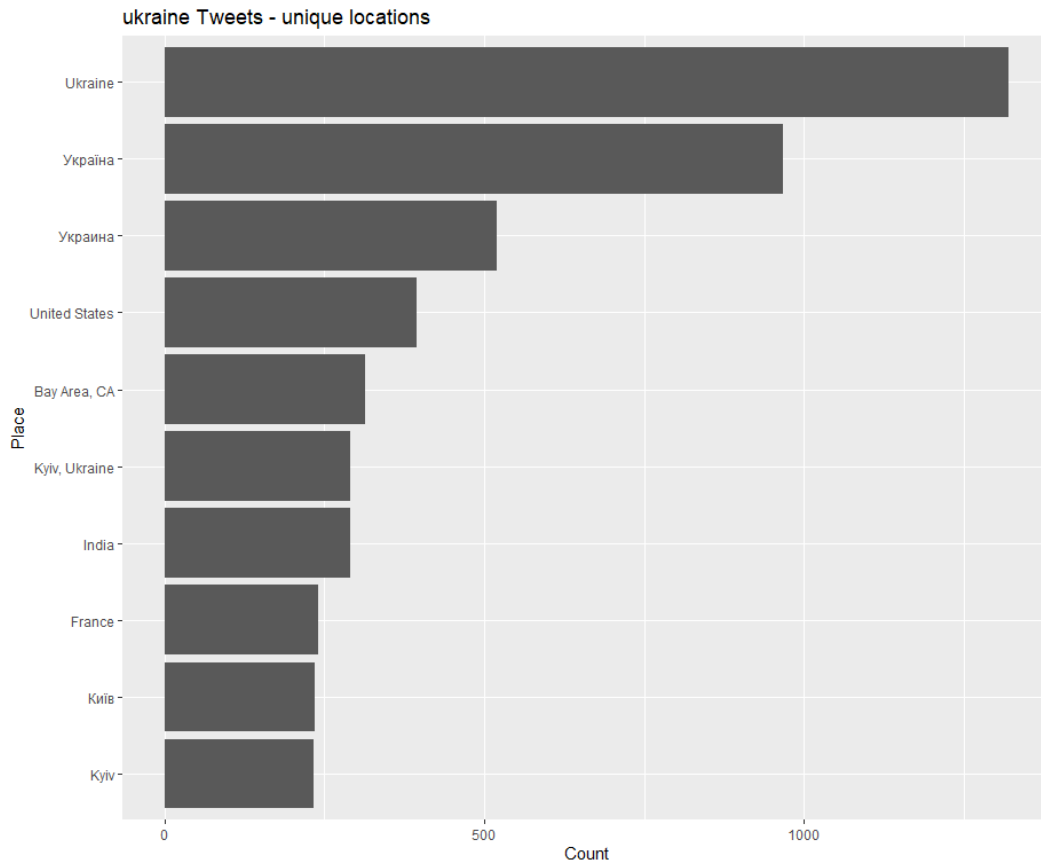
Exploration: Twitter-Demo

■ Kaggle Ukraine Data Set: 07/29/2022 (221, 482 tweets)

- Z.B. manuelle Bot-Detection (automatisiert [tweetbotornot2](#))
- Variieren von Pre-processing Entscheidungen (z.B. Stop words, Zeitfenster)
- Wie ist die Verteilung der User-Geolokation (location)?
- Was sind mögliche [Confounder, Collider](#)?



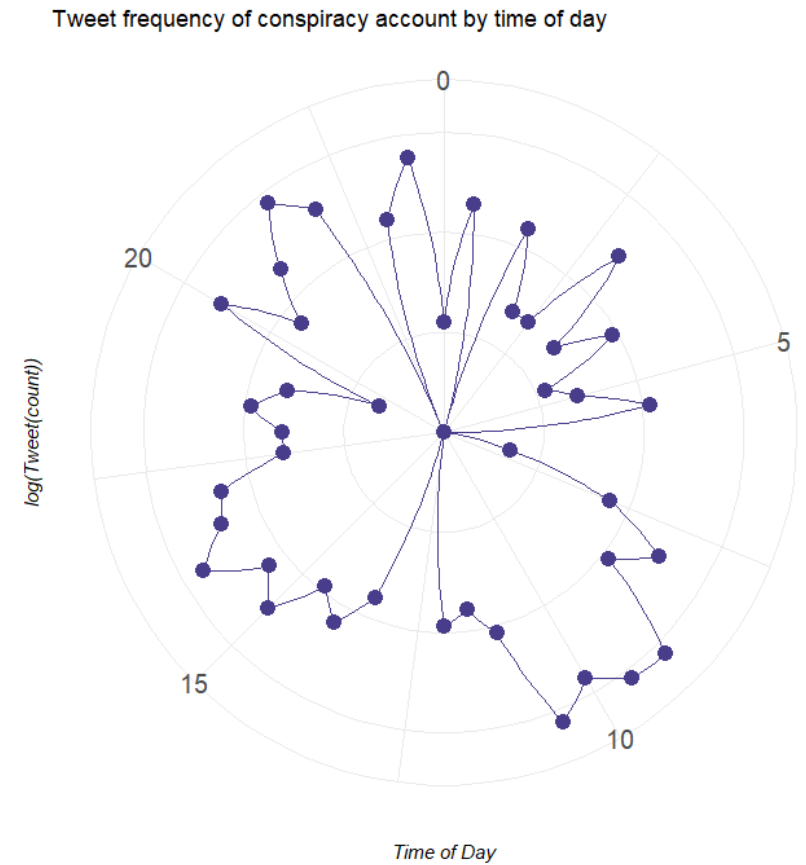
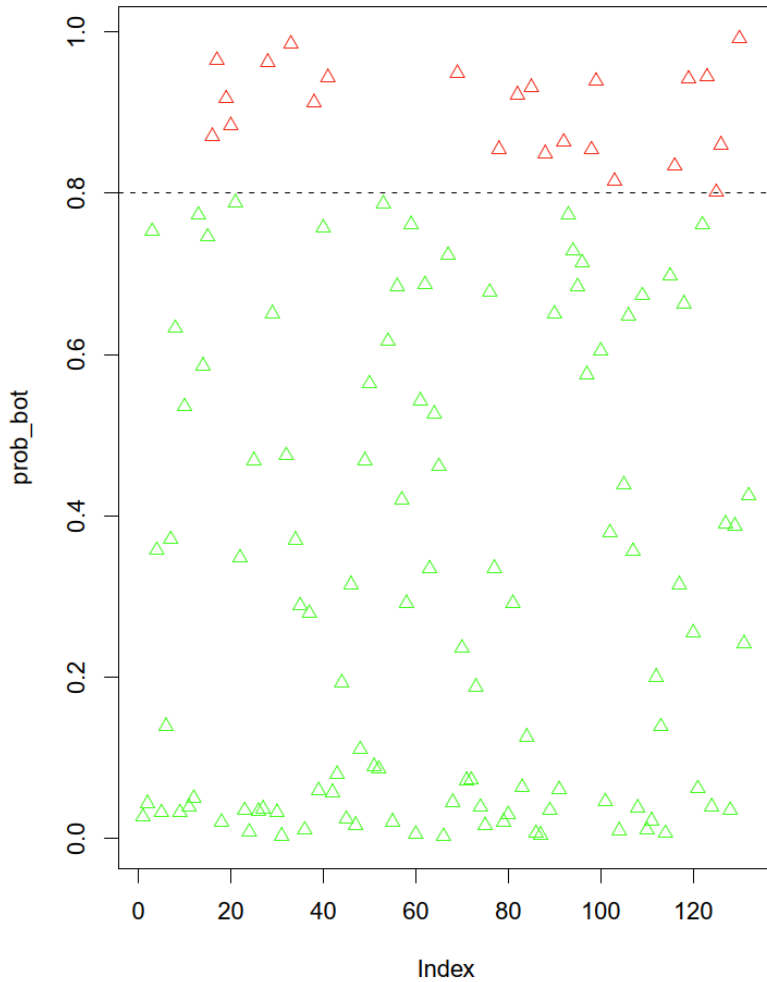
Exploration: Twitter-Demo



```
tweets %>%
  count(location, sort = TRUE) %>%
  mutate(location_var =
    reorder(location, n)) %>%
  na.omit() %>%
  top_n(10) %>%
```

```
ggplot(aes(x = location_var, y = n)) +
  geom_col() +
  coord_flip() +
  labs(x = "Place",
       y = "Count",
       title = "ukraine Tweets - unique
locations ")
```

Bots und Conspiracy-Accounts



Crawling ...

Youtube Daten und Untertitel

(vgl. auch [Kohne, Breuer, & Mohseni, 2022](#))

- ? Welche Forschungsfragen könnten damit beantwortet werden?