

# Text Mining von YouTube-Kommentaren am Beispiel des Stoneman Douglas High School Shootings

---

LISA HEHNKE (UNIVERSITÄT BIELEFELD)

JOSEF HOLNBURGER (UNIVERSITÄT HAMBURG)

# Hintergrund

---

# Stoneman Douglas High School Shooting

---



Am **14. Februar 2018** starben an der Marjory Stoneman Douglas High School in Portland, Florida, 14 Schüler und drei Erwachsene bei einem Amoklauf.

Der Amoklauf wurde von dem ehemaligen Schüler Nikolas Cruz (19 Jahre) verübt.

Die Tat zog ein großes Medienecho und zahlreiche Proteste nach sich. Zur Vermeidung solcher Taten wurde u.a. die Bewaffnung von Lehrern diskutiert.

# Wie alles begann...

---

Ursprüngliches Forschungsinteresse galt den **Verschwörungstheorien**, die im Kontext des Amoklaufs verbreitet wurden.

Bereits während früheren Amokläufen wurden zahlreiche Verschwörungstheorien verbreitet.





## The Alex Jones Channel

---

*Infowars* verbreitete u.a. Verschwörungstheorien zu dem Amoklauf an der Sandy Hook Elementary School in Las Vegas 2012.

Hintergrund der Verschwörungstheorien: Die Anschläge seien inszeniert (sog. *false flags*). Die Opfer und Beteiligten seien sog. *crisis actors*.



### Full Show - Evidence Unraveling A Giant False Flag

96.817 Aufrufe • vor 2 Monaten



### Full Show - DISGUSTING: Leftists Already Using Florida

36.930 Aufrufe • vor 2 Monaten



### Dems Announce Plans To Repeal The 2nd Amendment

29.971 Aufrufe • vor 2 Monaten



### FBI Ignored Cruz Warning To Investigate Russia

11.610 Aufrufe • vor 2 Monaten



### Strange Coincidences Surround Florida High School

136.298 Aufrufe • vor 2 Monaten



### To Prevent Mass Killings Stop Drugging Our Young

35.955 Aufrufe • vor 2 Monaten



### MSM Coverage Of Florida Shooting Is Completely One-

20.035 Aufrufe • vor 2 Monaten



### Alex Jones: They're Coming After Our 2nd Amendment

22.982 Aufrufe • vor 2 Monaten

# Ursprüngliches Forschungsinteresse

---

**Arbeitshypothese:** Videos mit Verschwörungstheorien zum Amoklauf werden mit zeitlicher Verzögerung auf YouTube veröffentlicht. Sie zirkulieren dafür länger als Videos „klassischer“, d. h. berichterstattender, Medien (Bessi et al. 2014, Del Vicario et al. 2016).

**Vorgehen:** Vergleich der amoklaufbezogenen Videos von Alex Jones mit denen von ABC News, CNN, Fox News und NBC.



# Storytelling gone wrong...

---

Die Arbeitshypothese konnte mit den Daten nicht zufriedenstellend getestet werden.

Allerdings wurden große Unterschiede in der Aktivität und dem Verhalten kommentierender User zwischen den Nachrichtenkanälen beobachtet.

# Forschungsfrage

---

## Wie wird der Amoklauf auf ausgewählten YouTube-Kanälen rezipiert?

→ Explorative Auswertung der Nutzeraktivität und Rezeption  
der Videos von ABC News, Alex Jones, CNN, Fox News und  
NBC



# Konzeptioneller Rahmen

---

**Viewer engagement** als  
*„user-initiated interactions between viewers  
and the content creator, and/or other members  
of the community“*

(nach Brodie et al 2013: 107)

# Weiterführende Forschungsfragen

---

1. Wer „engagiert“ sich (wie)?
2. Was kommentieren User und auf welche Art und Weise tun sie es?

# Methodische Vorgehensweise

---

Explorative Analyse der Kommentare und Videostatistiken:

1. Userbezogene *engagement metrics* (u.a. Anzahl der Kommentare, Bewertungen, Sentiments)
2. Semantische *engagement metrics* (u.a. Lesbarkeit, Komplexität der Kommentarinhalt)

# Datenmanagement

---

# Download von YouTube-Daten mit R

YouTube-Kommentare und Videostatistiken können über das R-Paket **tuber** heruntergeladen werden.

Mit der Funktion `yt_search()` kann nach Videos, Kanälen oder Listen gesucht werden. Daten können anschließend mit `get_stats()` bzw. `get_all_comments()` heruntergeladen werden.

Wichtig: YouTube-Algorithmus beachten!

## Package ‘tuber’

March 1, 2018

Title Client for the YouTube API  
Version 0.9.5  
Maintainer Gaurav Sood <gsood07@gmail.com>  
Description Get comments posted on YouTube videos, information on how many times a video has been liked, search for videos with particular content, and much more. You can also scrape captions from a few videos. To learn more about the YouTube API, see <<https://developers.google.com/youtube/v3/>>.

URL <http://github.com/soodoku/tuber>

BugReports <http://github.com/soodoku/tuber/issues>

Depends R (>= 3.4.0)

License MIT + file LICENSE

LazyData true

Imports httr, plyr, dplyr, purrr

VignetteBuilder knitr

Suggests knitr (>= 1.11), testthat, rmarkdown, xml2, lint

RoxygenNote 6.0.1

NeedsCompilation no

Author Gaurav Sood [aut, cre],  
Kate Lyons [ctb]

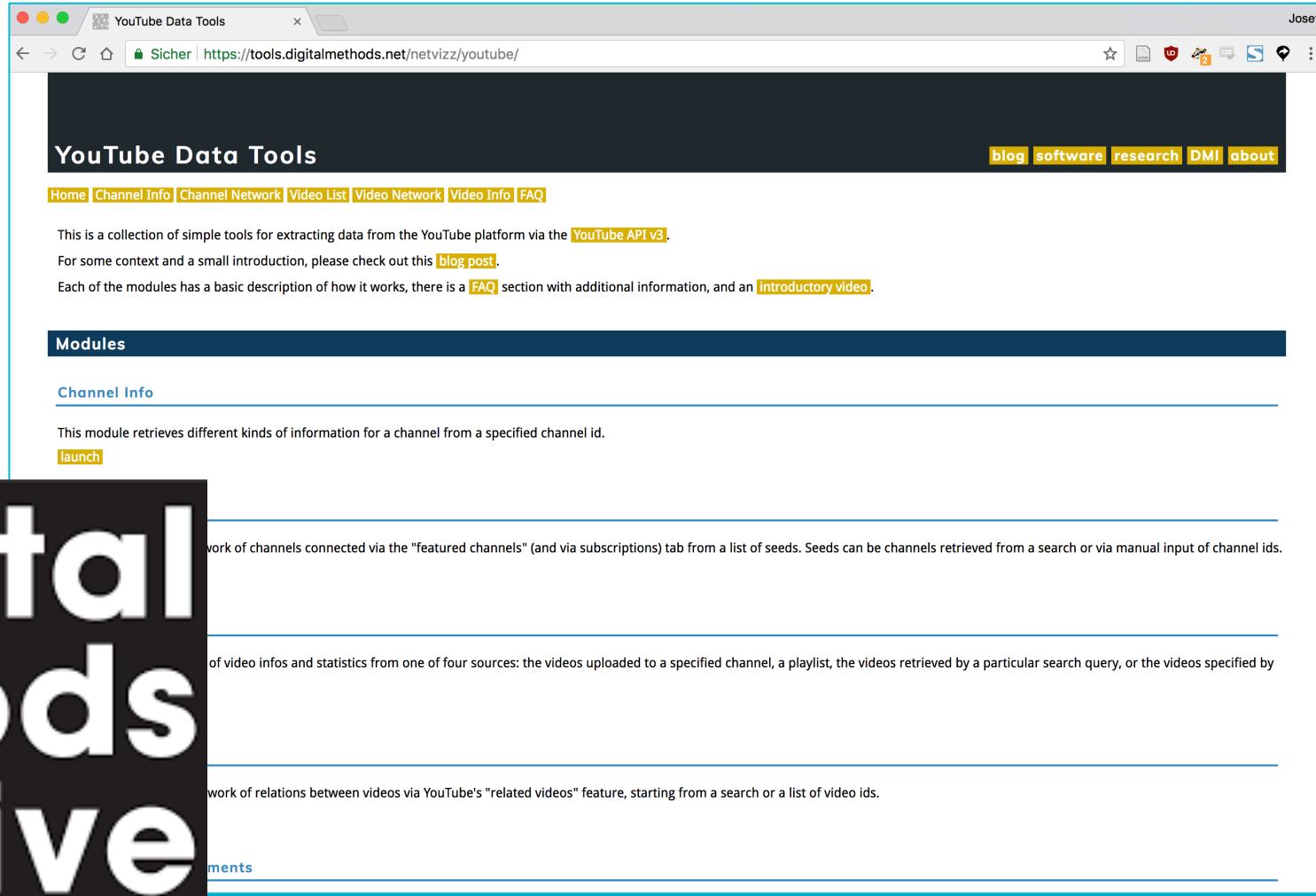
Repository CRAN

Date/Publication 2018-03-01 12:35:03 UTC

### R topics documented:

delete_captions	2
delete_channel_sections	3
delete_comments	4
delete_playlists	4
delete_playlist_items	5
delete_videos	6

# Download von YouTube-Daten mittels YouTube Data Tools



The screenshot shows a web browser window titled "YouTube Data Tools" with the URL <https://tools.digitalmethods.net/netvizz/youtube/>. The page has a dark header with the title and a navigation bar with links for blog, software, research, DMI, and about. Below the header, there's a menu with Home, Channel Info, Channel Network, Video List, Video Network, Video Info, and FAQ. A note says it's a collection of tools for extracting data via the YouTube API v3, with a link to a blog post for context. Each module has a basic description, a FAQ section, and an introductory video. The main content area is titled "Modules" and includes sections for "Channel Info", "Video List", "Video Network", "Video Info", and "Comments". Each section has a brief description and a "launch" button.

YouTube Data Tools

blog software research DMI about

Home | Channel Info | Channel Network | Video List | Video Network | Video Info | FAQ

This is a collection of simple tools for extracting data from the YouTube platform via the [YouTube API v3](#).

For some context and a small introduction, please check out this [blog post](#).

Each of the modules has a basic description of how it works, there is a [FAQ](#) section with additional information, and an [introductory video](#).

**Modules**

[Channel Info](#)

This module retrieves different kinds of information for a channel from a specified channel id.

[launch](#)

work of channels connected via the "featured channels" (and via subscriptions) tab from a list of seeds. Seeds can be channels retrieved from a search or via manual input of channel ids.

[Video List](#)

work of video infos and statistics from one of four sources: the videos uploaded to a specified channel, a playlist, the videos retrieved by a particular search query, or the videos specified by

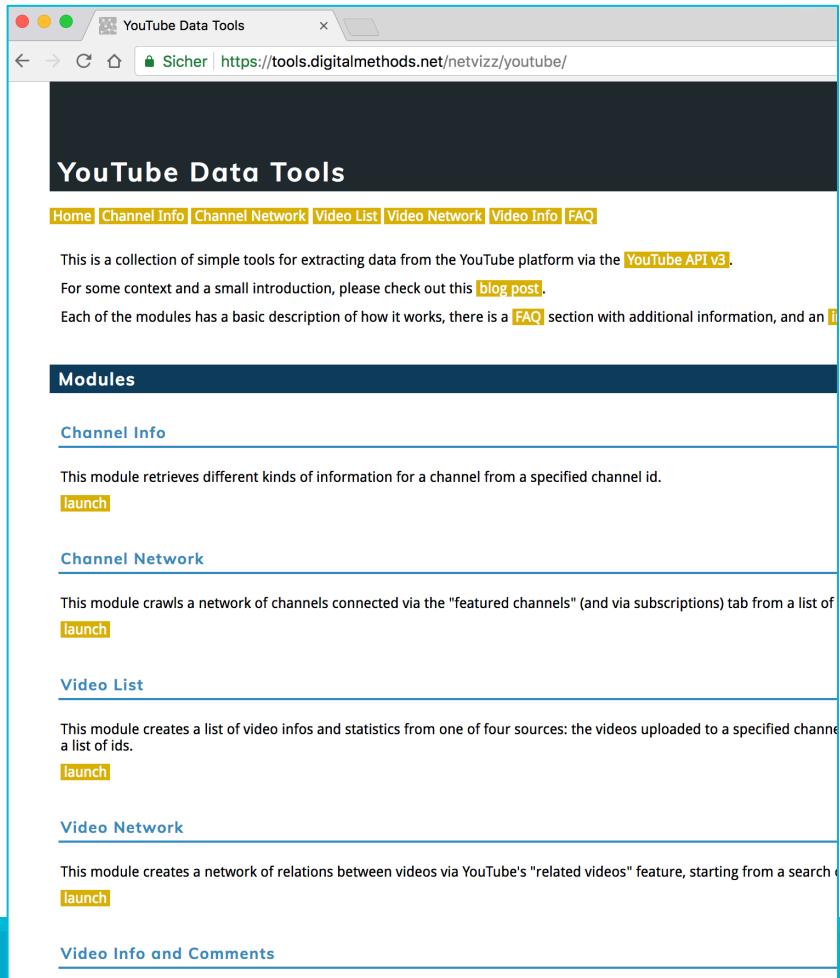
[Video Network](#)

work of relations between videos via YouTube's "related videos" feature, starting from a search or a list of video ids.

[Comments](#)



# Download von YouTube-Daten mittels YouTube Data Tools



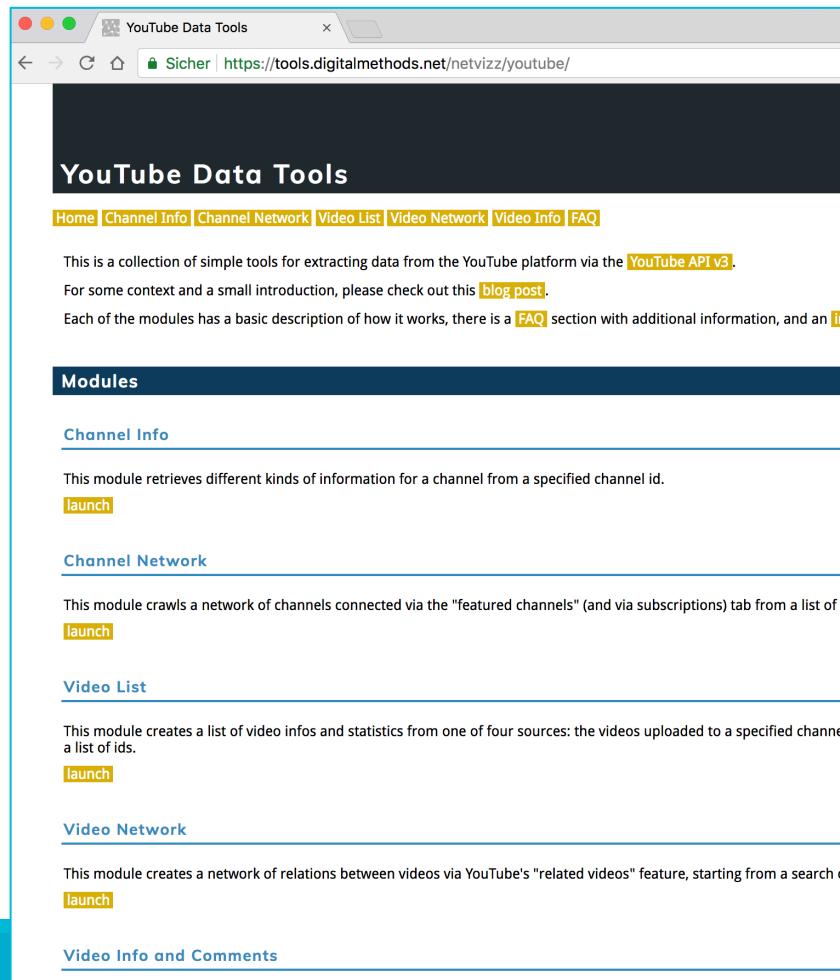
Umfangreiches und sehr nützliches Tool, Kommentare lassen sich jedoch nur für einzelne Videos herunterladen.

Downloads von Statistiken (Anzahl der Videos, Klickzahlen, Bewertung durch NutzerInnen) sind auch für Kanäle und Listen möglich.

**YouTube Data Tools** ist Open Source:  
<https://github.com/bernorieder/YouTube-Data-Tools>

# Datengewinnung

---



Das von Bernhard Rieder bereitgestellte Tool wurde für die Datengewinnung angepasst.

Es wurden Statistiken und Kommentare zu sämtlichen Videos der genannten Kanäle vom 14.02.2018 bis 28.02.2018 heruntergeladen und in einer SQL-Datenbank abgespeichert.

## Videos

<b>videoID</b>	 varchar(20)
publishedAt	datetime
channelId	varchar(255)
title	text
description	text
thumbnails.default.url	text
...	
channelTitle	text
liveBroadcastContent	tinyint
viewCount	int
likeCount	int
dislikeCount	int
favoriteCount	int
commentCount	int

## Comments

<b>videoid</b>	varchar(20)
<b>id</b>	 int
replyCount	int
likeCount	int
publishedAt	datetime
author	text
text	text
authorChannelId	varchar(255)
authorChannelUrl	text
isReply	tinyint
isReplyToChannel	varchar(255)
isReplyToName	varchar(255)

videoID	id	replyCount	likeCountComment	publishedAtComment	author	text	
1	A1LEprbmlow	UgzFkuDfjrBBvyTjkuV4AaABAq	0	0	2018-03-01 06:36:48	KOBETHEGOAT24	this notion that put
2	A1LEprbmlow	Ugw1K9Amul9eK1nkX_N4AaABAq	0	0	2018-03-01 04:27:21	Ahbobie	this fucker hasd REA
3	A1LEprbmlow	UgwugrvK0-5APhAZFo54AaABAq	0	0	2018-03-01 04:18:52	Ahbobie	fuck ugt littlye cunt
4	A1LEprbmlow	UgwZSWrtocF9g75OX1d4AaABAq	0	0	2018-02-28 19:42:40	fjolla ha	Omg and where was
5	A1LEprbmlow	UgycgU9CTrlwhbdB0i54AaABAq	0	0	2018-02-28 11:16:34	Sheila Kemunto	Why talk after the in
6	A1LEprbmlow	Ugx3jjscZHfpwF9UZ6N4AaABAq	0	0	2018-02-28 08:05:50	Maria Ingrid Bartolome	That guy was a victim
7	A1LEprbmlow	UgwU4-XVwpnRhXYS6Bx4AaABAq	0	0	2018-02-28 02:55:30	Jeff Taylor	Fake
8	A1LEprbmlow	UgzWXWB19rG0XGfiMtx4AaABAq	0	0	2018-02-28 02:35:53	BJ Martin	Yup, the dumb ass a
9	A1LEprbmlow	UgzcMD5tynNEt8LL94AaABAq	0	0	2018-02-28 02:15:15	Big Chief Trucker	the only these libera
10	A1LEprbmlow	UgzkgyTAFPk0WQwE7-R4AaABAq	0	0	2018-02-28 00:50:06	Lindell Paints	FBI and CIA are gett
11	A1LEprbmlow	UgwcDO0kj-l7AbOEnDx4AaABAq	0	0	2018-02-28 00:39:13	Anthony 2b Channel 6	Horseshit
12	A1LEprbmlow	Ugyvq1m0iVWM72umcvl4AaABAq	0	0	2018-02-28 00:05:26	KarasCyborg	The fact that the Sh
13	A1LEprbmlow	UgxD8uRtHcLU-ADjmyJ4AaABAq	0	0	2018-02-27 22:15:34	Judas	15 guns on his bed?
14	A1LEprbmlow	UgxnxGK0mSrjgLY3oQHp4AaABAq	0	0	2018-02-27 22:04:16	Lucila Morin	damn they're j
15	A1LEprbmlow	Ugx2LAkr6B5kV-ASN-B4AaABAq	0	0	2018-02-27 20:22:42	LRRPFco52	Let me get this strai
16	A1LEprbmlow	UgxOidCKg-OU687TDXR4AaABAq	0	0	2018-02-27 16:08:03	chloejohanna31	So much neglect :(
17	A1LEprbmlow	UgzAow5iLxehinUar6x4AaABAq	1	0	2018-02-27 11:24:46	richard mullins	<a href="https://w
18	A1LEprbmlow	UgzAow5iLxehinUar6x4AaABAq.8dACM4ZliNh8dTuj0...	0	0	2018-03-01 03:15:09	Will Millard	Its 1984
19	A1LEprbmlow	UgyXXRU1ynxmWfhRrcp4AaABAq	0	0	2018-02-27 11:02:30	Croxann	Seemed like the sch
20	A1LEprbmlow	UgyFRG3yB3BirzQLJpd4AaABAq	0	0	2018-02-27 10:40:41	Armilite1	It's Sad this ha
21	A1LEprbmlow	UgzblfEAcGRnjnn0PyJ4AaABAq	0	0	2018-02-27 07:06:24	J dog	hat in the the world
22	A1LEprbmlow	UgwvWDvDn88ar13Skel4AaABAq	0	0	2018-02-27 05:32:54	123toaster	OH look, it's D
23	A1LEprbmlow	Ugxk0qI0qc6nQkl4X4t4AaABAq	0	0	2018-02-27 04:18:31	Jdor D	Yeshua/Jesus is the
24	A1LEprbmlow	UqxngTa78QyxtXb6ihB4AaABAq	0	0	2018-02-27 03:06:32	Savvy Maharaj	Feel sorry for his ex



927.389 Kommentare



240.047 NutzerInnen



1.199 Videos

Zeitraum: 14.02.2018 bis 28.02.2018

# Aufbereitung von Textdaten

---

1. Kommentare um HTML-Code, URLs, Emojis, Zahlen, Satz- und Sonderzeichen bereinigen
2. Großschreibung in Kleinschreibung umwandeln
3. Stammformreduktion durchführen (sog. *stemming*)
4. Stoppwörter entfernen
5. *Tokenization*

# Aufbereitung von Textdaten

---

DAVOR

<a href="https://www.youtube.com/watch?v=A1LEprbmIw&t=0m44s">0:44</a> They have picked a very good looking crisis actor for this interview. Before buying here story, please note that the whole event as a hoax, so what she is saying is simply a piece of theatre.

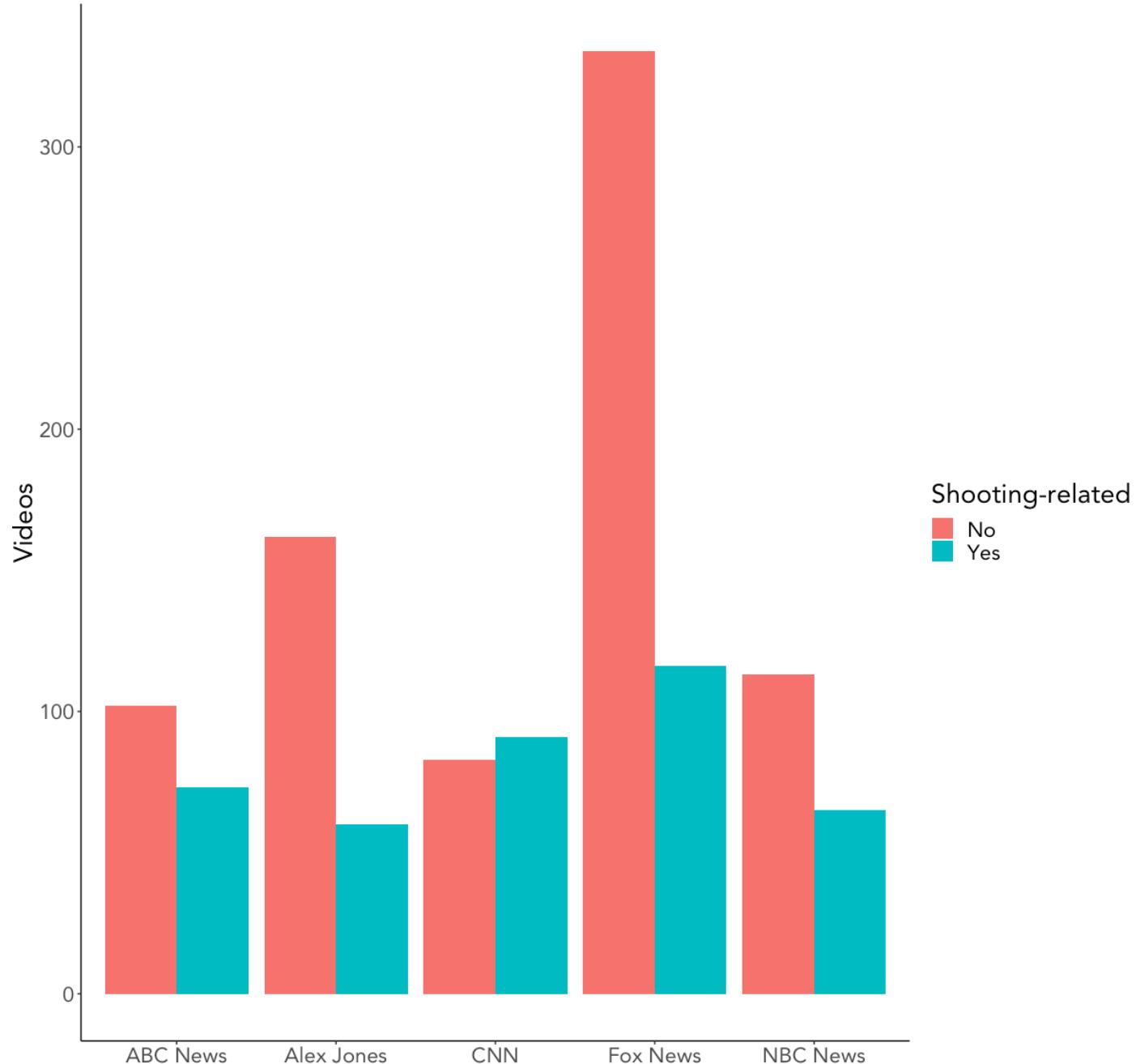
DANACH

they have picked a very good looking crisis actor for this interview before buying here story please note that the whole event as a hoax so what she is saying is simply a piece of theatre

# Erste Ergebnisse

---

Number of shooting-related vs. other videos



# YouTube-Videos zum Amoklauf

ZEITRAUM 14.02.2018 BIS  
28.02.2018

GEFILTERT DURCH  
FLORIDA|HIGH SCHOOL|  
SHOOTING|NRA  
IN TITEL ODER  
BESCHREIBUNG



624.334 Kommentare

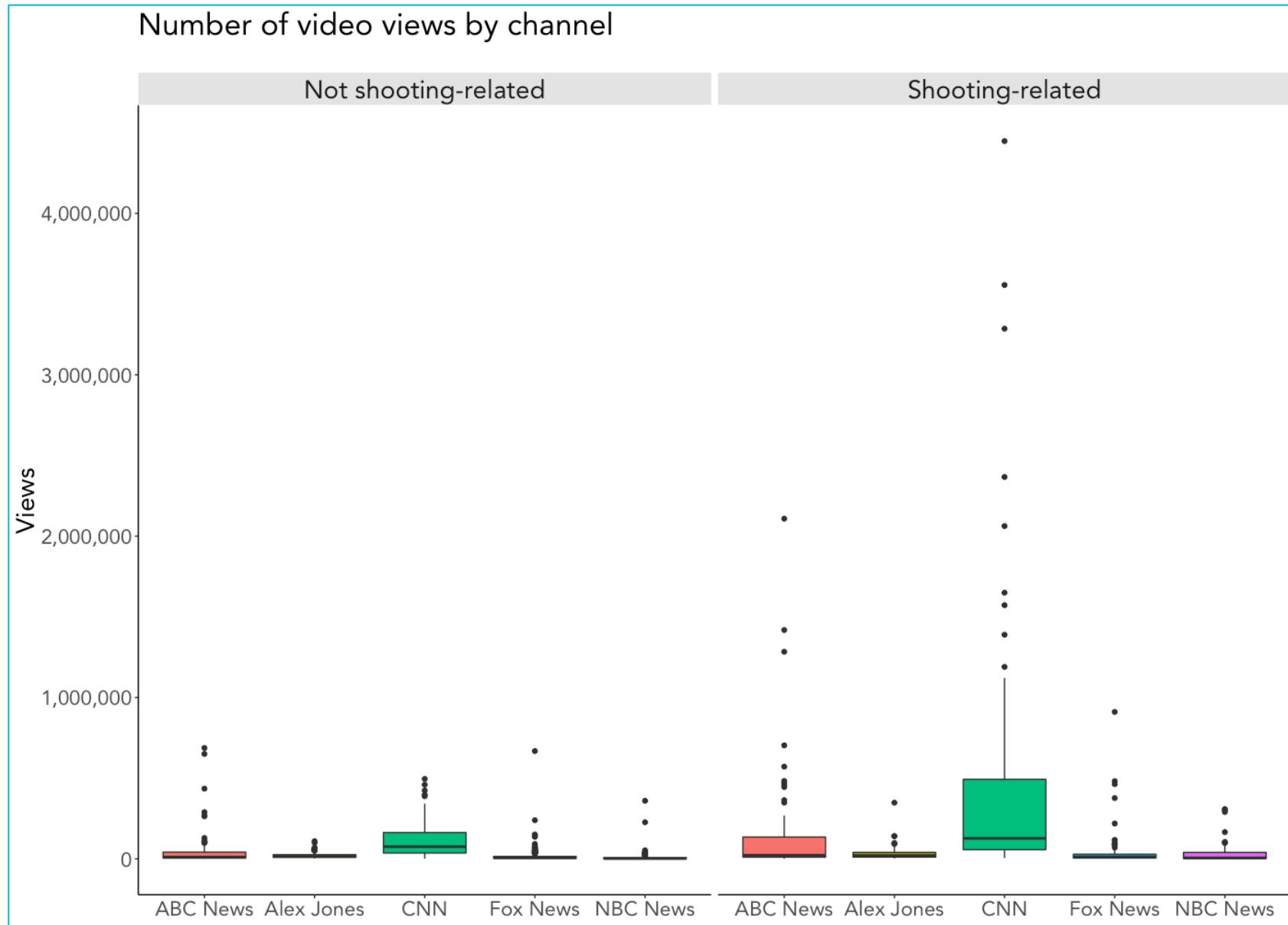


190.830 Nutzer

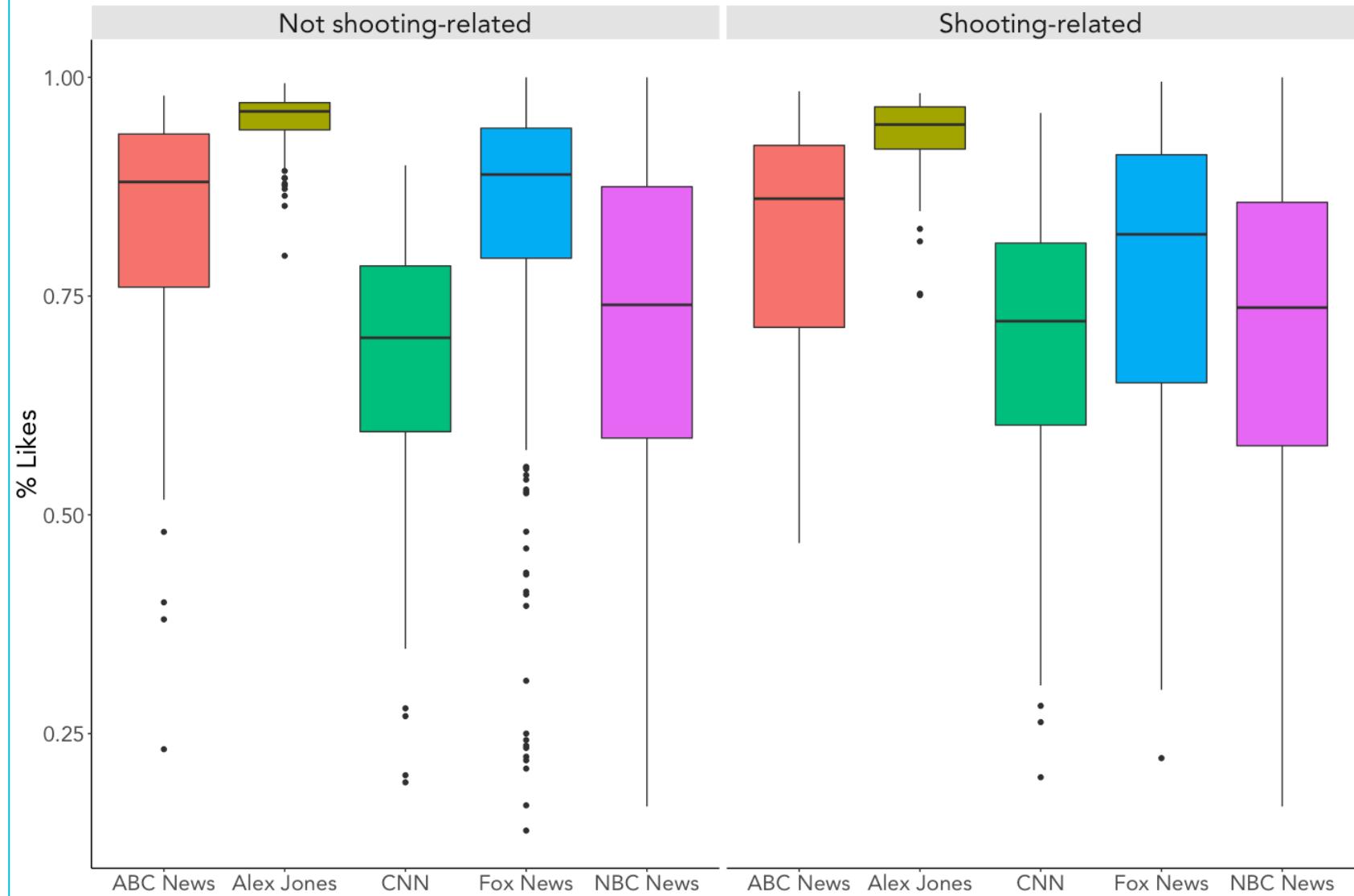


405 Videos

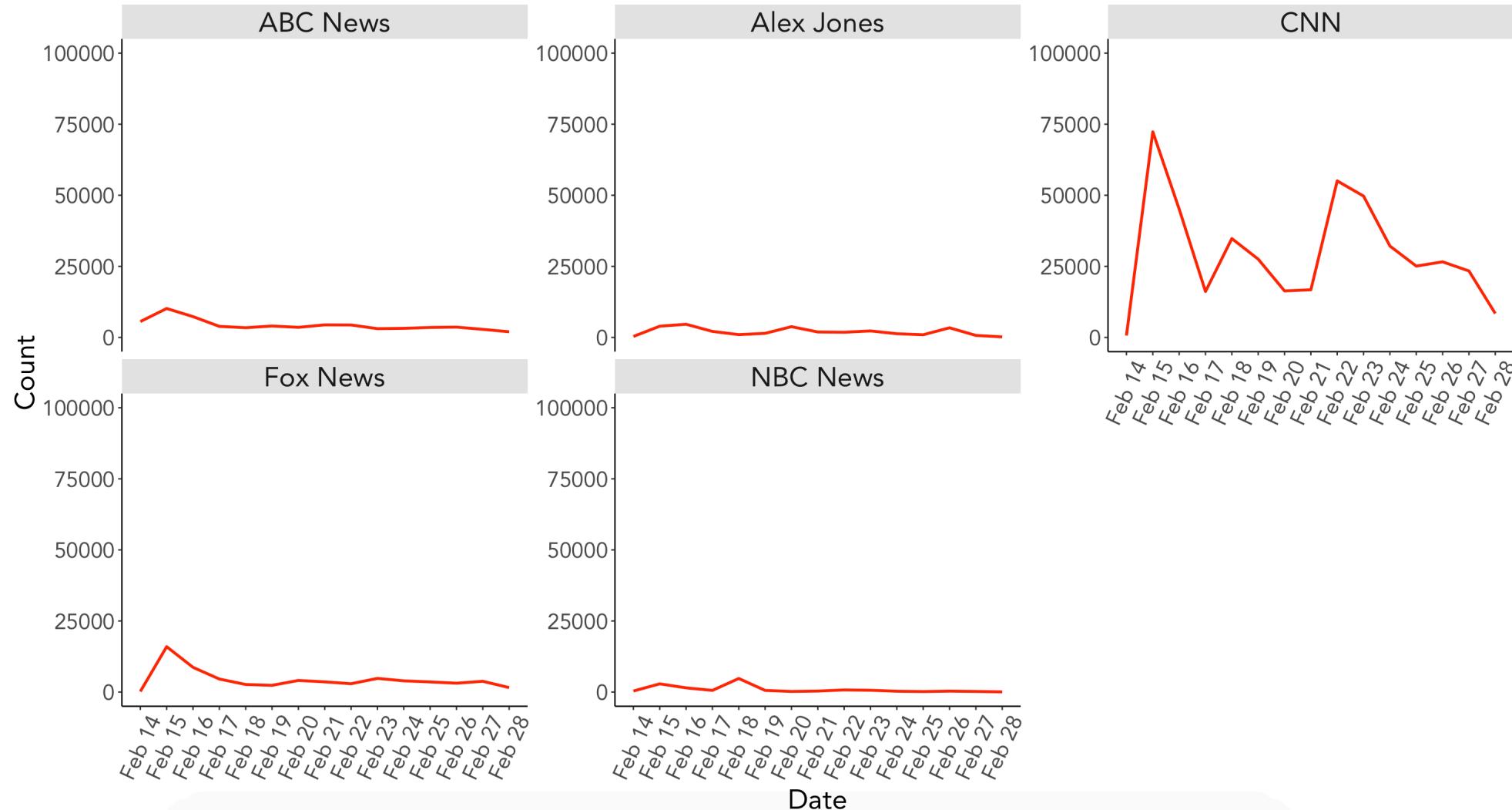
Zeitraum: 14.02.2018 bis 28.02.2018



## Percentage of likes by channel



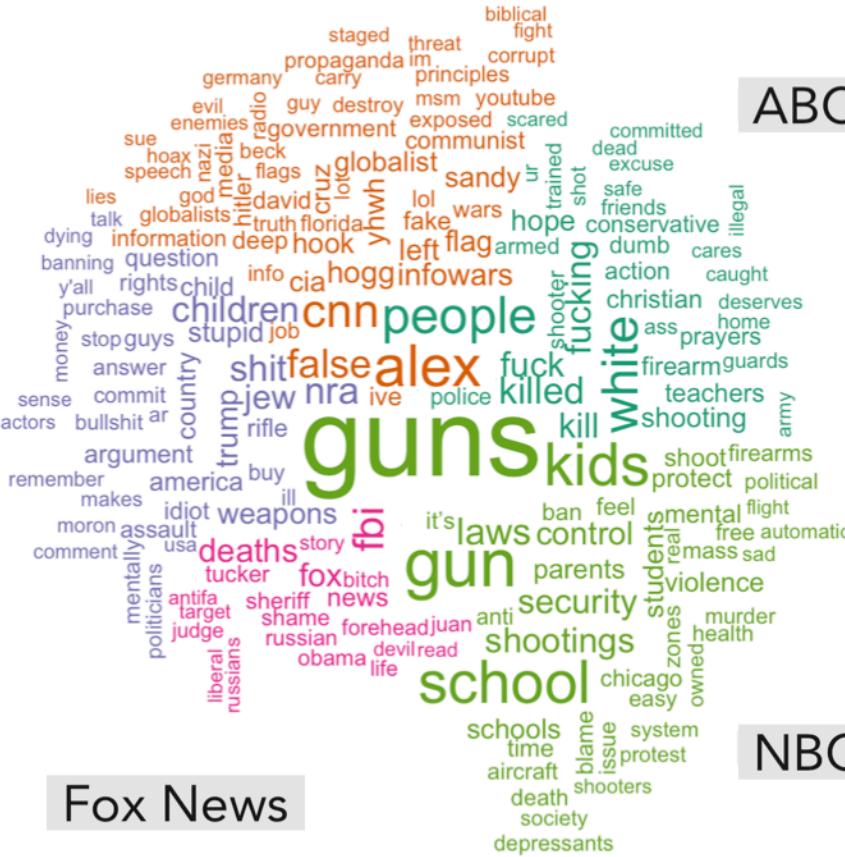
## Number of YouTube comments over time



Alex Jones

CNN

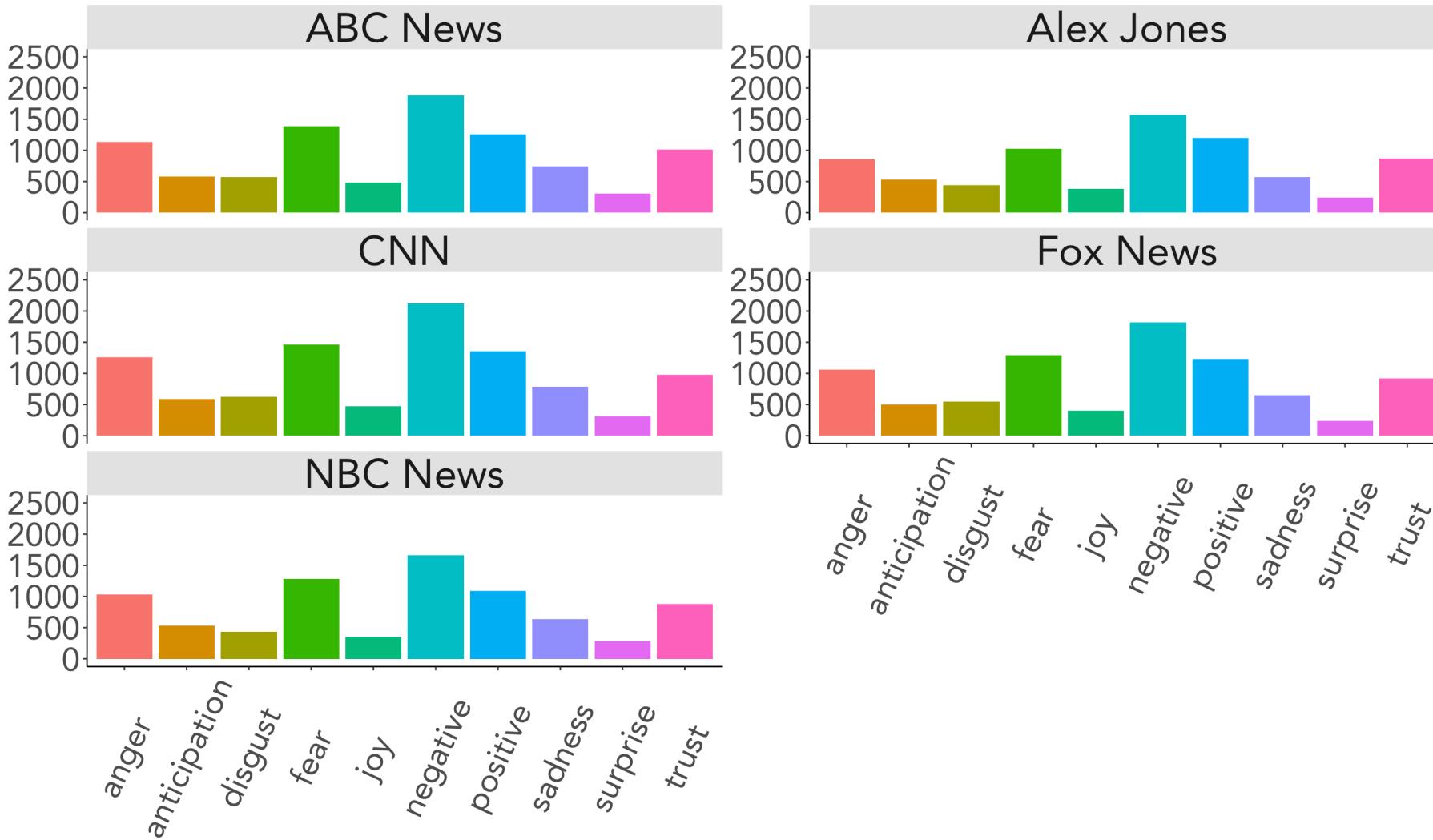
Fox News



ABC News

NBC News

## Total sentiment scores in YouTube comments (nrc)



# Diskussion

---

# YouTube-Daten

---

## PROS

- Riesiges Datenvolumen
- Daten sind ständig verfügbar und können bspw. bei aktuellen Ereignissen direkt heruntergeladen werden
- Oftmals ungefilterte Meinungen und geringe Reaktivität der User
- Geringe Kosten

## CONS

- Datenqualität? → Daten sind häufig unvollständig (Beispiel: Kommentare können gelöscht werden)
- „*Messiness*“ erfordert sorgfältige Datenaufbereitung
- Daten sind nur äußerst selten repräsentativ
- YouTube-Algorithmus bestimmt, wem welche Videos angezeigt werden
- Veränderungen an der Plattform bewirken Veränderungen im Userverhalten
- Sensible Daten



# Herausforderungen

---

1. Ohne geeignetes Forschungsdesign kaum Rückschlüsse über kausale Mechanismen, sondern nur deskriptive Aussagen möglich
2. Ethische Fragen im Umgang mit Daten der NutzerInnen
3. Zukünftige Einschränkungen beim Download von YouTube-Daten?

# Replication, Replication

---

## R SCRIPT

<https://github.com/lhehnke/youtube-data>

## SCRAPER

<https://github.com/bernorieder/YouTube-Data-Tools>

# Referenzen

---

- Bessi, A., Coletto, M., Davidescu, G. A., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2015). Science vs conspiracy: Collective narratives in the age of misinformation. *PLoS one*, 10(2), e0118093.
- Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, 66(1), 105-114.
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., & Scala, A. et al. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554-559.
- Rieder, B. (2017). *YouTube Data Tools*. Computer software. Vers. 1.08. N.p., 5 December 2017. Online: <https://tools.digitalmethods.net/netvizz/youtube/>.
- Rieder, B. et al. (2018). From ranking algorithms to ‘ranking cultures’: Investigating the modulation of visibility in YouTube search results. *Convergence*, 24(1), 50–68.
- Sood, G. (2018). *tuber: Access YouTube from R*. R package version 0.9.3.

# Danke für die Aufmerksamkeit!



@DATAPLANES



@HOLNBURGER   
HOLNBURGER 