

Awareness for Climate Change amid the Spread of COVID-19

Fundamentals of Computing and Data Display

Leonie Gehrmann

2020-06-15

Contents

Introduction	1
Data	2
Results	6
Discussion	11
References	11

Introduction

“Climate change is the defining issue of our time [...]. From shifting weather patterns [...] to rising sea levels [...], the impacts of climate change are global in scope and unprecedented in scale.” -(United Nations 2019)

This issue has been brought to the general attention throughout the past 25 years, e.g. through the development of the United Nations Framework Convention on Climate Change in 1994, the introduction of legally binding emission reduction targets defined in the Kyoto Protocol in 1997, and the recent Paris Agreement striving to keep the global temperature rise under 2°C (United Nations 2019). Albeit being a relevant topic that affects us all, it could be argued that intensified awareness and action has only recently surged in the wake of the global Fridays for Future movement. These protests against the relative inaction in response to the climate crisis were first initiated by Greta Thunberg (Fridays for Future 2020b). On August 20, 2018 she started her protest outside of the Swedish parliament, which later grew into the movement #FridaysForFuture (Fridays for Future 2020b). However, even this recent heightened awareness might not be enough to initiate the necessary actions and substantially inhibit this impending crisis.

Interestingly, the world is currently facing a much more salient health crisis, the COVID-19 pandemic. This new type of coronavirus has caused a global outbreak of respiratory illness since its first official report on December 31, 2019 (WHO 2020). As of mid-June there have been more than 7 million cases worldwide, with the disease being linked to over 400.000 deaths (WHO 2020). It seems that both the news and all social conversations ultimately revolve around this topic, especially during the early stages of its spread in Germany.

The difference between the salience and general awareness of these two crises motivates the question how important topics should be positioned and communicated in order to receive adequate interest. More specifically, the interesting question arises whether journalistic professionals differ in their susceptibility to hypes or trend topics in comparison to the general population. It seems that news on COVID-19 is creating large waves and a type of self-inflating media coverage, popularly referred to as media-hype (Vasterman 2005).

On the one hand, it is to be expected that journalists write about topics that are of interest to the general population, their readership. On the other hand, it would seem that their professional training should sensitize them towards the pitfalls of highly trending topics.

This research project aims to analyze the development of both public and journalistic interest in climate change and COVID-19. Since both of these topics are fairly recent, this investigation strongly benefits from the use of Google Trends data. First, this source provides fine-grained daily data over a long period of time. Second, it allows for a rather direct comparison of general interest, in the form of regular search queries, and journalistic interest, captured by the articles written for various news outlets and assembled by Google News. Regarding public interest, one might argue that search activity does not accurately represent actual efforts taken to fight climate change, e.g. waste reduction or more efficient energy use. However, alternative data such as donations to relevant groups (e.g. Greenpeace) or the average carbon footprint of a resident of Germany is far less up-to-date due to the lag in availability of the necessary information. This argument holds even more so for interest in COVID-19 due to the very current nature of the crisis.

In the following, the data collection and aggregation is detailed before a descriptive exploration and analysis of the posed research questions follows. Finally a discussion summarizes the most important findings.

Data

This research project focuses on the interest in the two topics climate change and COVID-19 in Germany, as reflected in Google search behavior. Furthermore, data on Fridays for Future marches in Germany is used to further validate the data set. The relevant observation period begins on August 20, 2018 when Greta Thunberg started striking and ends on June 14, 2020.

Google Trends: Web Searches

First, the search terms most appropriate to capture the general interest in the two topics are determined. Generally, Google Trends search terms are case-insensitive and only report relative numbers, rather than absolute frequency. A comparison of searches for the terms *Friday for Future* and *Fridays for Future* shows that the later is used more frequently.

Similarly, a look at different German search terms related to climate change (*Klimawandel*, *Fridays for Future*, *Klimakrise*, *climate change*, *Erderwärmung*) indicates that the first two are searched for most frequently. Although they have a similar baseline number of searches, queries for *Fridays for Future* spike substantially at certain times, likely around big protests.

Finally, the popularity of various terms referring to the current pandemic (*Coronavirus*, *COVID-19*, *COVID*, *Sars-Cov-2*) is assessed. Coronavirus is used most often by far.

One issue with using Google Trends data is that search periods longer than 9 months will only report weekly data, but one of the explicit benefits of using Google searches is the opportunity of fine-grained daily information. (Medium 2018) suggest a valid approach to obtain daily data for longer time periods. First, monthly data is gathered and stored for the entire observation period, once for the two climate change-related search terms (*Fridays for Future*, *Klimawandel*) in `monthly_res_terms1` and once for the term *Coronavirus* in `monthly_res_terms2` (code not shown).

```
# monthly climate change searches
res_terms1_1 <- gtrends(c("Klimawandel", "Fridays for Future"),
                        geo = "DE", time = "2014-06-14 2020-06-14", low_search_volume = T)

# store the monthly hits for Fridays for Future and Klimawandel in a separate data frame
# (excluding unnecessary variables time and category)
# only for relevant period (starting in row 50 2018-08)
monthly_res_terms1 <-
```

```

res_terms1_1$interest_over_time[50:nrow(res_terms1_1$interest_over_time), c(1,2,3,4,6)]

# store as .tsv
write_delim(monthly_res_terms1, "monthly_res_terms1.tsv", delim = " ")

```

Next, daily searches are gathered per month separately to have a valid benchmark. Afterwards, the daily data for each month is combined into one complete dataframe per search topic `daily_res_terms1` for climate change and `daily_res_terms2` for COVID-19 (code not shown). Notice that the searches for the climate-change related terms are obtained separately to the searches for *Coronavirus*. This is due to the relative nature of the data and the different observation periods. The very high number of searches revolving around COVID-19 in the spring of 2020 would otherwise render past searches for *Fridays for Future* or *Klimawandel* too small in comparison. Hence, the comparison of interest in the two topics later on will also only be of relative nature rather than the difference in absolute levels. Also, *Coronavirus* was not searched for until January 2020.

```

# daily climate change searches
# create a vector with the months in the observation period
dates1 <- c("2018-08-20 2018-08-31", "2018-09-01 2018-09-30",
            "2018-10-01 2018-10-31", "2018-11-01 2018-11-30",
            "2018-12-01 2018-12-31", "2019-01-01 2019-01-31",
            "2019-02-01 2019-02-28", "2019-03-01 2019-03-31",
            "2019-04-01 2019-04-30", "2019-05-01 2019-05-31",
            "2019-06-01 2019-06-30", "2019-07-01 2019-07-31",
            "2019-08-01 2019-08-31", "2019-09-01 2019-09-30",
            "2019-10-01 2019-10-31", "2019-11-01 2019-11-30",
            "2019-12-01 2019-12-31", "2020-01-01 2020-01-31",
            "2020-02-01 2020-02-29", "2020-03-01 2020-03-31",
            "2020-04-01 2020-04-30", "2020-05-01 2020-05-31",
            "2020-06-01 2020-06-14")

# loop through the months to get separate results for each one
for (i in 1:length(dates1)){
  timeperiod <- dates1[i]
  res <- gtrends(c("Klimawandel", "Fridays for Future"), geo = "DE",
                 time = timeperiod, low_search_volume = T)
  assign(paste("daily_res_terms1_", i, sep = ""), res)
}

# combine all 23 results to one data frame for climate change
daily_res_terms1 <- rbind(daily_res_terms1_1$interest_over_time,
                           daily_res_terms1_2$interest_over_time,
                           daily_res_terms1_3$interest_over_time,
                           daily_res_terms1_4$interest_over_time,
                           daily_res_terms1_5$interest_over_time,
                           daily_res_terms1_6$interest_over_time,
                           daily_res_terms1_7$interest_over_time,
                           daily_res_terms1_8$interest_over_time,
                           daily_res_terms1_9$interest_over_time,
                           daily_res_terms1_10$interest_over_time,
                           daily_res_terms1_11$interest_over_time,
                           daily_res_terms1_12$interest_over_time,
                           daily_res_terms1_13$interest_over_time,
                           daily_res_terms1_14$interest_over_time,

```

```

        daily_res_terms1_15$interest_over_time,
        daily_res_terms1_16$interest_over_time,
        daily_res_terms1_17$interest_over_time,
        daily_res_terms1_18$interest_over_time,
        daily_res_terms1_19$interest_over_time,
        daily_res_terms1_20$interest_over_time,
        daily_res_terms1_21$interest_over_time,
        daily_res_terms1_22$interest_over_time,
        daily_res_terms1_23$interest_over_time)

# store as .tsv
write_delim(daily_res_terms1, "daily_res_terms1.tsv", delim = " ")

```

Finally, remember that Google Trends data is not absolute. Therefore, the daily data must be adjusted using the monthly searches over the entire observation period, thereby allowing for a cross-monthly comparison of the individual data points, as shown by (Medium 2018). For this, the two data sets per search term group are joined by matching `keyword` and `month` to create `res_terms1_final` and `res_terms2_final` (code not shown). Then, the daily search hits `hits.x` are weighted by the monthly hits `hits.y` to give the comparable variable `totalhit` over the entire observation period.

```

# adjusted daily climate change searches
# add month variable to be able to match to monthly_res_terms1
daily_res_terms1$month <-
  paste(format(daily_res_terms1$date, format = "%Y-%m"), "01", sep = "-")
# rename date in monthly_res_terms1 to match the name
monthly_res_terms1$month <- as.character(monthly_res_terms1$date)

res_terms1_final <-
  left_join(daily_res_terms1, monthly_res_terms1, by = c("keyword", "month"))

# remove unnecessary columns
res_terms1_final$category <- NULL
res_terms1_final$date.y <- NULL
res_terms1_final$geo.y <- NULL
res_terms1_final$gprop.y <- NULL

# turn hits into a numeric variable
res_terms1_final$hits.x <- as.numeric(res_terms1_final$hits.x)
res_terms1_final$hits.x[is.na(res_terms1_final$hits.x)] <- 0
res_terms1_final$hits.y <- as.numeric(res_terms1_final$hits.y)
res_terms1_final$hits.y[is.na(res_terms1_final$hits.y)] <- 0

# create adjusted hit score
res_terms1_final$totalhit <- res_terms1_final$hits.x * res_terms1_final$hits.y / 100

# store as .tsv
write_delim(res_terms1_final, "terms1_final.tsv", delim = " ")

```

Google Trends: News

The data gathered so far captures the general interest in the topics over time. Gathering the hits for Google News should provide a proxy for journalistic interest and representation of the topics in various media outlets.

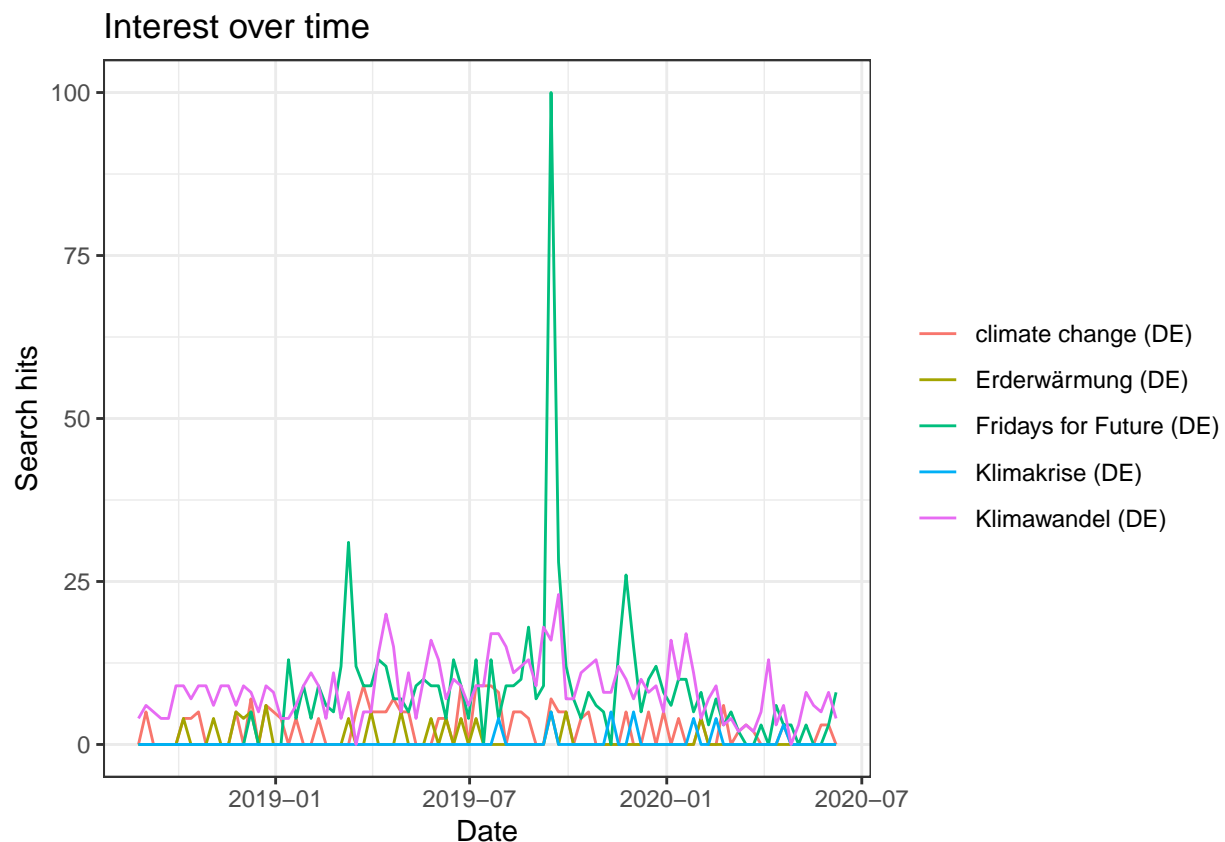
Since an equally detailed data set should be gathered, a very similar code to the case of web searches is run to create `res_news1_final` and `res_news2_final` (code not shown).

Alternatively, the number of articles published in various German national newspapers (e.g. Spiegel, Süddeutsche, FAZ) could have been scraped individually. However, this absolute number of articles with the relevant keywords is difficult to compare to the relative interest as portrayed by Google Trends. Therefore, Google News results are used in this project, providing the same fine-grained daily data as Google Trends.

Taking a look at the common terms in Google News shows a similar pattern for COVID-19 related keywords. The vast majority uses the term *Coronavirus* rather than alternatives such as *COVID-19*, *COVID*, or *Sars-Cov-2*. However, regarding climate change keywords, news articles show a slightly more balanced use of *Klimawandel* and *Fridays for Future* (except for an unordinary spike in interest for Fridays for Future during the global climate strike in September 2019), as well as more instances of the terms *Klimakrise*, *climate change* and even *Erderwärmung*.

```
# climate change
news_terms1 <- gtrends(c("Klimawandel", "Fridays for Future",
                        "Klimakrise", "climate change",
                        "Erderwärmung"), geo = "DE", gprop = "news",
                        time = "2018-08-20 2020-06-14", low_search_volume = T)

plot(news_terms1)
```



Finally, the four overall data sets `res_terms1_final`, `res_terms2_final`, `res_news1_final` and `res_news2_final` are combined to one final larger data set `res_final`, containing both general and journalistic interest as indicated by Google searches and online published articles. This data set consists of 2980 observations on 9 variables.

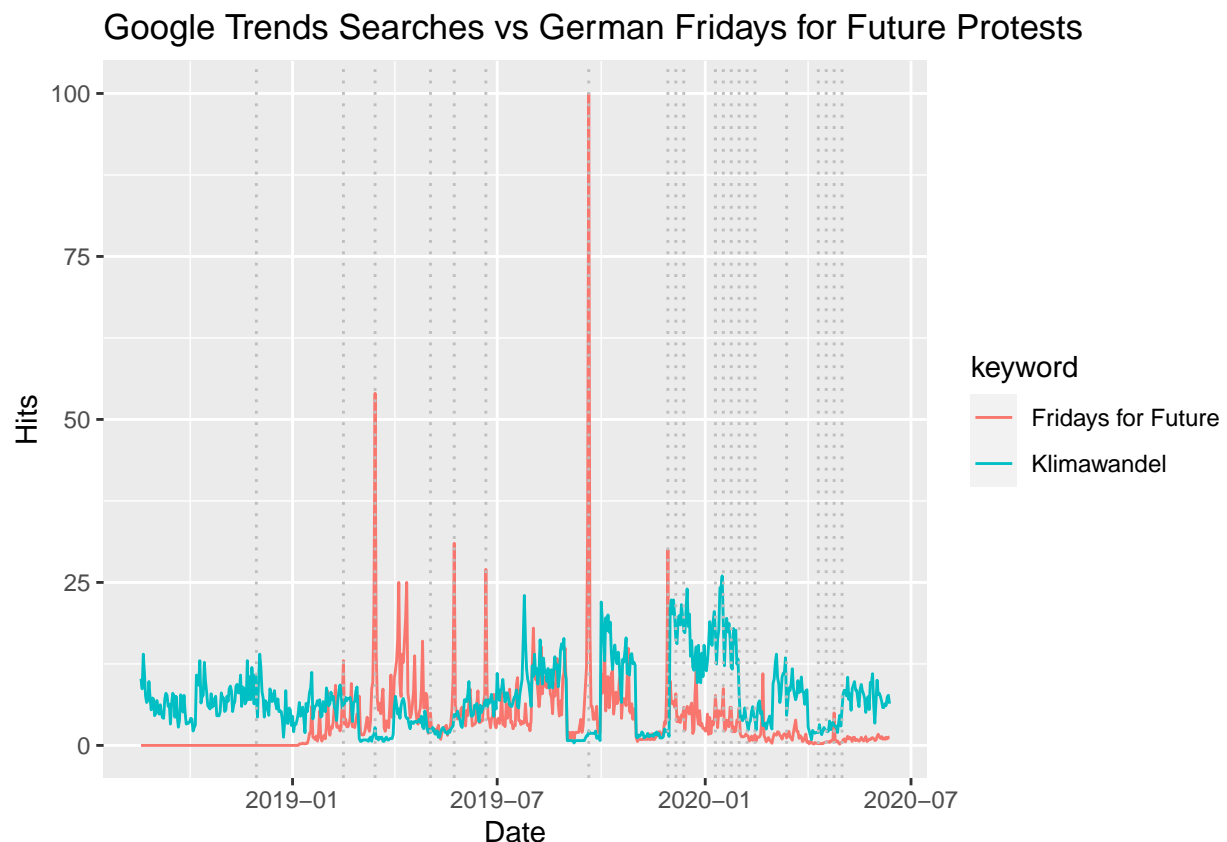
```
res_final <- rbind(res_terms1_final, res_terms2_final, res_news1_final, res_news2_final)

# store as .tsv
write_delim(res_final, "res_final.tsv", delim = " ")
```

Results

Now, having introduced the research idea and the used data sets, some visualizations are used to explore the data further and analyze the posed questions. As mentioned, the number of Fridays for Future protests FFF in Germany on a given day is used to validate the Google Trends searches. The data represents the number of strikes occurring in Germany on the given date (Fridays for Future 2020a). The graph illustrates the development of searches for the two climate-change related terms over the observation period. The vertical lines indicate dates on which Fridays for Future protests occurred in Germany. Notice that spikes in searches for Fridays for Future coincide nicely with the dates of the protests. Interestingly, searches for *Klimawandel* remain constant and relatively high at the end of 2019 and throughout the beginning of 2020, suggesting that the regular protests might have increased general awareness for the issue.

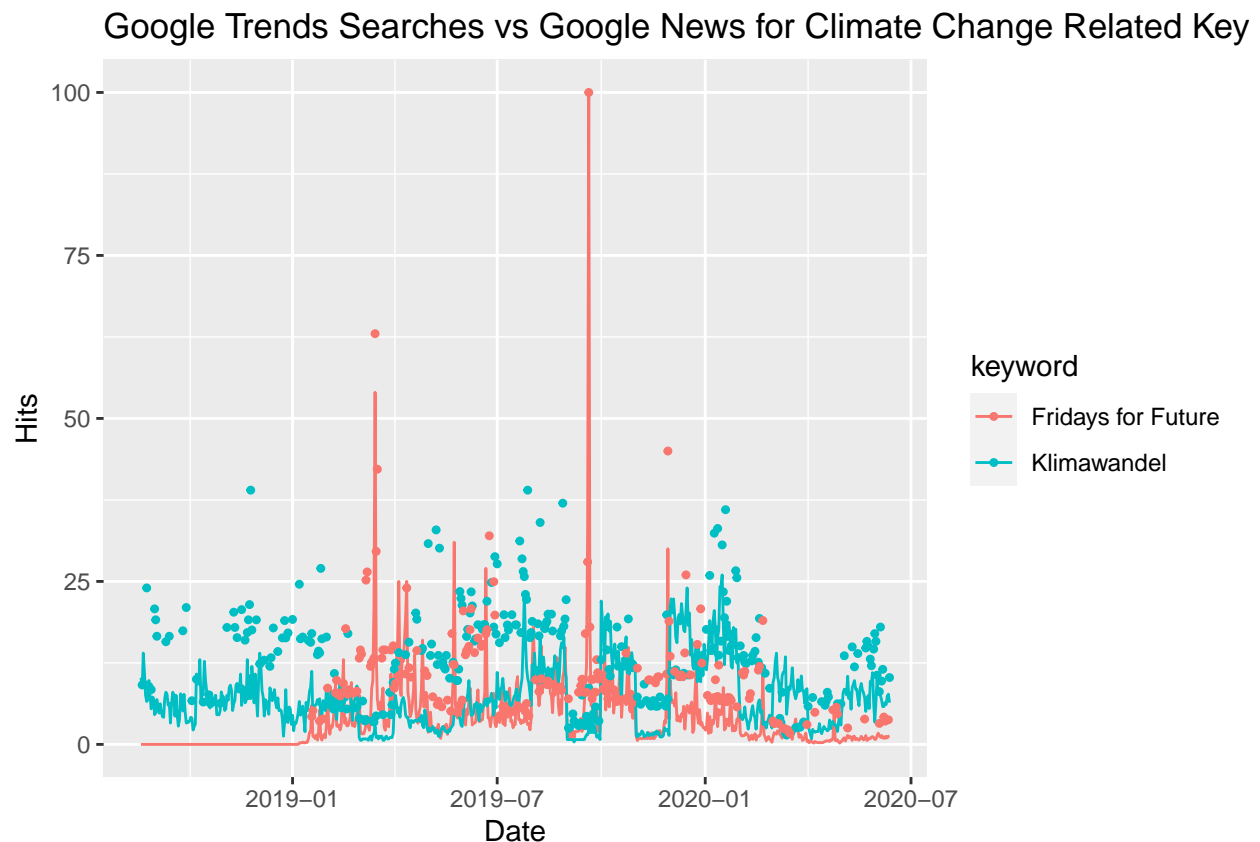
```
res_final_climate <-
  res_final[(res_final$keyword == "Klimawandel" | res_final$keyword == "Fridays for Future"),]
ggplot(data = res_final_climate[res_final_climate$gprop.x == "web",]) +
  geom_path(aes(x = date.x, y = totalhit, col = keyword)) +
  labs(x = "Date", y = "Hits",
       title = "Google Trends Searches vs German Fridays for Future Protests") +
  geom_vline(data = FFF, aes(xintercept = date), linetype = 3, color = "grey")
```



Next, a look at the development of Google searches and articles published on climate change related topics

indicate some differences between general and journalistic interest in the issue. The general interest is represented by the lines in the graph, whereas the amount of articles published on a given day is illustrated as data points. This difference in visualization is chosen to avoid a strongly fluctuating line for Google News hits that occurs since articles in general are not published every day, whereas general searches happen daily. Observe that the general and journalistic interest in the *Fridays for Future* movement seems to coincide quite well. Notice however, that journalistic interest in *Klimawandel* is generally above interest in *Fridays for Future*, a pattern that seems rather opposite for Google Trends. This suggests journalists could be able to differ between a substantial topic itself and a current related movement causing more hype better compared to the general public.

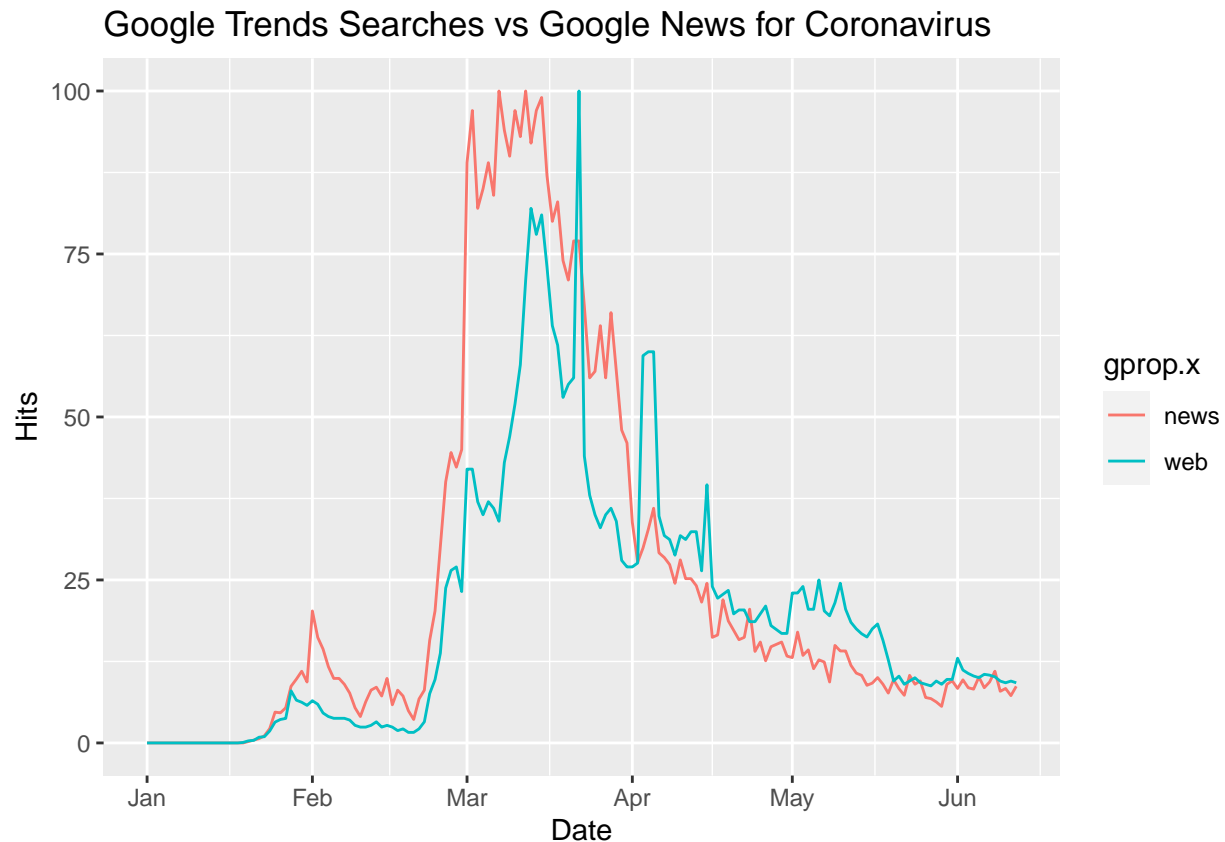
```
ggplot(data = res_final_climate) +
  geom_path(data = res_final_climate[res_final_climate$gprop.x == "web",],
    aes(x = date.x, y = totalhit, col = keyword)) +
  geom_point(data = res_final_climate[(res_final_climate$gprop.x == "news" &
    res_final_climate$totalhit > 0),],
    aes(x = date.x, y = totalhit, col = keyword), size = 0.9) +
  labs(x = "Date", y = "Hits",
    title = "Google Trends Searches vs Google News for Climate Change Related Keywords")
```



Similarly, the following graph visualizes the Google searches and articles published related to COVID-19 in 2020. For this topic, both the general and journalistic interest develop in a rather comparable manner. Two interesting things to note are the spike in articles published at the beginning of February, before the general public seems broadly aware of the virus, as well as the earlier heightened journalistic interest starting in March. The entire month sees many articles published with the daily number staying fairly constant and close below the relative maximum. Google Trends hits indicate the lag of the general population's interest which spikes at the end of March. Overall, this graph rather contradicts the previous notion that journalists

are able to resist the hype, rather supporting the idea that articles published tend to cover topics that are of current interest to the general population.

```
res_final_corona <-
  res_final[(res_final$keyword == "Coronavirus"),]
ggplot(data = res_final_corona) +
  geom_path(data = res_final_corona, aes(x = date.x, y = totalhit, col = gprop.x)) +
  labs(x = "Date", y = "Hits",
       title = "Google Trends Searches vs Google News for Coronavirus")
```



Finally, a look at the correlation between the different keywords should provide further indication of the extent to which COVID-19 has potentially crowded out the awareness for the less salient crisis. First, the pairwise correlations between the Google News and Google Trends hits for each of the three relevant keywords support the findings from the previous two graphs. While journalistic and general interest in *Fridays for Future*, and *Coronavirus* move in relative unison, the number of articles published on *Klimawandel* is not necessarily an accurate representation of the general interest in the topic, as reflected by searches.

```
# Klimawandel
cor(res_final_climate$totalhit[(res_final_climate$gprop.x == "news" &
  res_final_climate$keyword == "Klimawandel")],
  res_final_climate$totalhit[(res_final_climate$gprop.x == "web" &
  res_final_climate$keyword == "Klimawandel")])
```

```
## [1] 0.2177997
```



```
# Fridays for Future
cor(res_final_climate$totalhit[(res_final_climate$gprop.x == "news" &
  res_final_climate$keyword == "Fridays for Future")],
  res_final_climate$totalhit[(res_final_climate$gprop.x == "web" &
  res_final_climate$keyword == "Fridays for Future")])
```

```
## [1] 0.6761377
```

```
# Coronavirus
cor(res_final_corona$totalhit[res_final_corona$gprop.x == "news"],
  res_final_corona$totalhit[res_final_corona$gprop.x == "web"])
```

```
## [1] 0.8577402
```

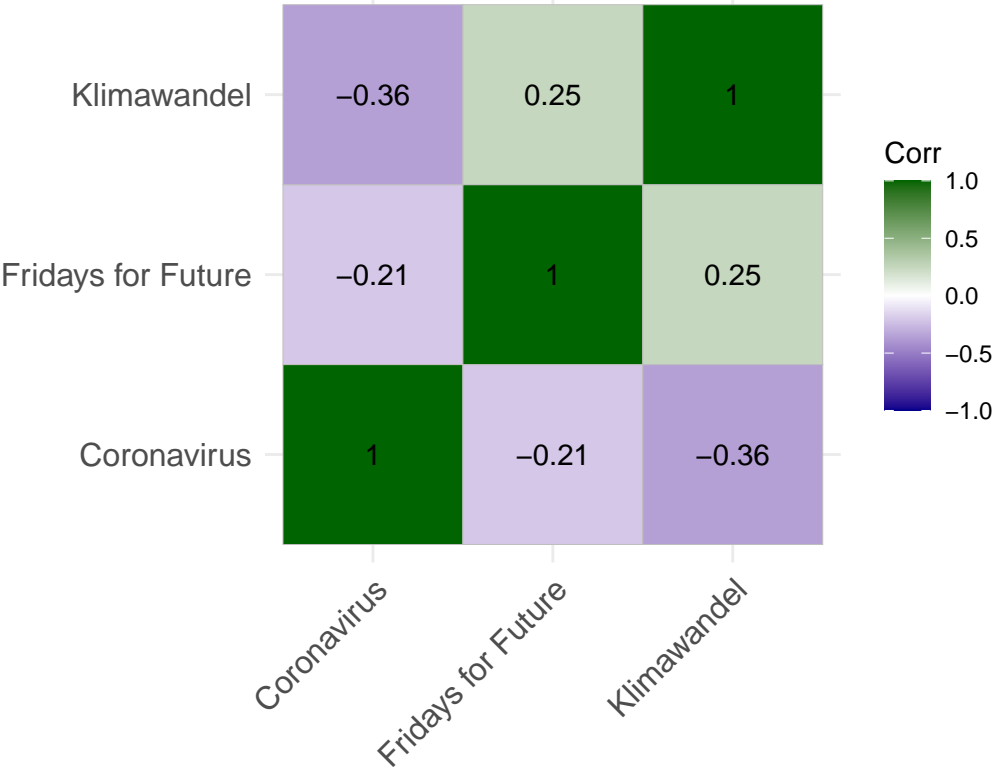
A separate analysis for each of the two groups provides insight into the extent to which COVID-19 has affected interest in climate change topics. A data set for each is created, containing the daily searches `res_final_trends` (code not shown) or articles published `res_final_news` for each of the three relevant keywords. All calculated correlations are significantly different from zero. As to be expected, interest for climate change related topics is positively correlated among both groups, while interest in such topics is adversely related to interest in COVID-19, as reflected by searches or articles published. Generally, the positive correlation for climate change topics is stronger for the general population, but overall the magnitude of the correlations is relatively low.

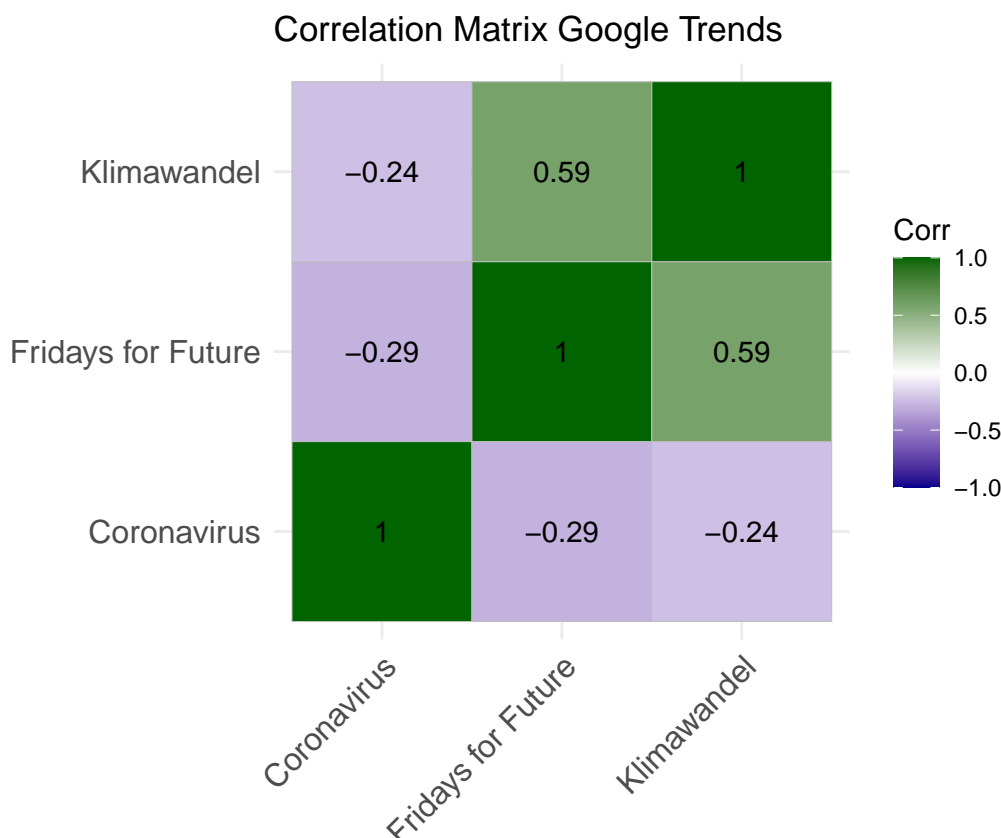
```
# vector of days in 2020
days2020 <- unique(res_final_corona$date.x)

# subset a data frame to only get Google News articles for the 3 keywords in 2020
res_final_news <-
  res_final[res_final$gprop.x == "news", c("date.x", "keyword", "totalhit")]
res_final_news <- res_final_news[(res_final_news$date.x %in% days2020),]
# reshape to wide format
res_final_news <- spread(res_final_news, key = keyword, value = totalhit)

# calculate correlation matrix
corr_news <- round(cor(res_final_news[,2:4]), 2)
# compute matrix with correlation p-values
p_news <- cor_pmat(res_final_news[,2:4])
# visualize
ggcorrplot(corr_news, colors = c("darkblue", "white", "darkgreen"), lab = TRUE,
  insig = "blank", p.mat = p_news, title = "Correlation Matrix Google News")
```

Correlation Matrix Google News





Discussion

This project has taken a closer look at the general and journalistic interest in two current crises, with one being more salient than the other. It seems that overall, journalists are not substantially better at resisting the hype and that COVID-19 is crowding out awareness of climate change. As argued in the beginning, this is likely also related to the fact that journalists generally write for the public, suggesting their articles should address relevant and current issues.

Using Google Trends and News data, a fine-grained data set containing hits for climate change and COVID-19 related keywords is built and explored. Thereby, the relative nature of the data is used to adjust reported hits and create data that is valid for the entire observation period ranging from August 2018 to June 2020.

Another advantage of Google Trends data is its geographic specificity, providing a possible extension of this project in the future. For one, the spread of COVID-19 in Germany varied across regions, something likely to be reflected in different search behavior. Furthermore, urban areas are condensed, with more people living in less space, likely a relevant factor in the extent and felt restriction of the measures enforced to ensure social distance and curb the spread of the virus. Hence, it could be expected that urban and higher populated areas experience a larger crowding-out of interest in climate change through COVID-19.¹

References

Fridays for Future. 2020a. “Strike Statistics.” <https://fridaysforfuture.org/what-we-do/strike-statistics/list-of-countries/>.

¹Link to GitHub repository: <https://github.com/lgehrm/fundamentals>

- . 2020b. “Who We Are.” <https://fridaysforfuture.org/what-we-do/who-we-are/>.
- Medium. 2018. “Google Trends: How to Acquire Daily Data for Broad Time Frames.” <https://medium.com/@bewerunge.franz/google-trends-how-to-acquire-daily-data-for-broad-time-frames-b6c6dfe200e6>.
- United Nations. 2019. “Climate Change.” <https://www.un.org/en/sections/issues-depth/climate-change/>.
- Vasterman, Peter L. M. 2005. “Media-Hype: Self-Reinforcing News Waves, Journalistic Standards and the Construction of Social Problems.” *European Journal of Communication* 20 (4): 508–30.
- WHO. 2020. “Coronavirus Disease (Covid-19): Situation Report - 145.” https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200613-covid-19-sitrep-145.pdf?sfvrsn=bb7c1dc9_2.