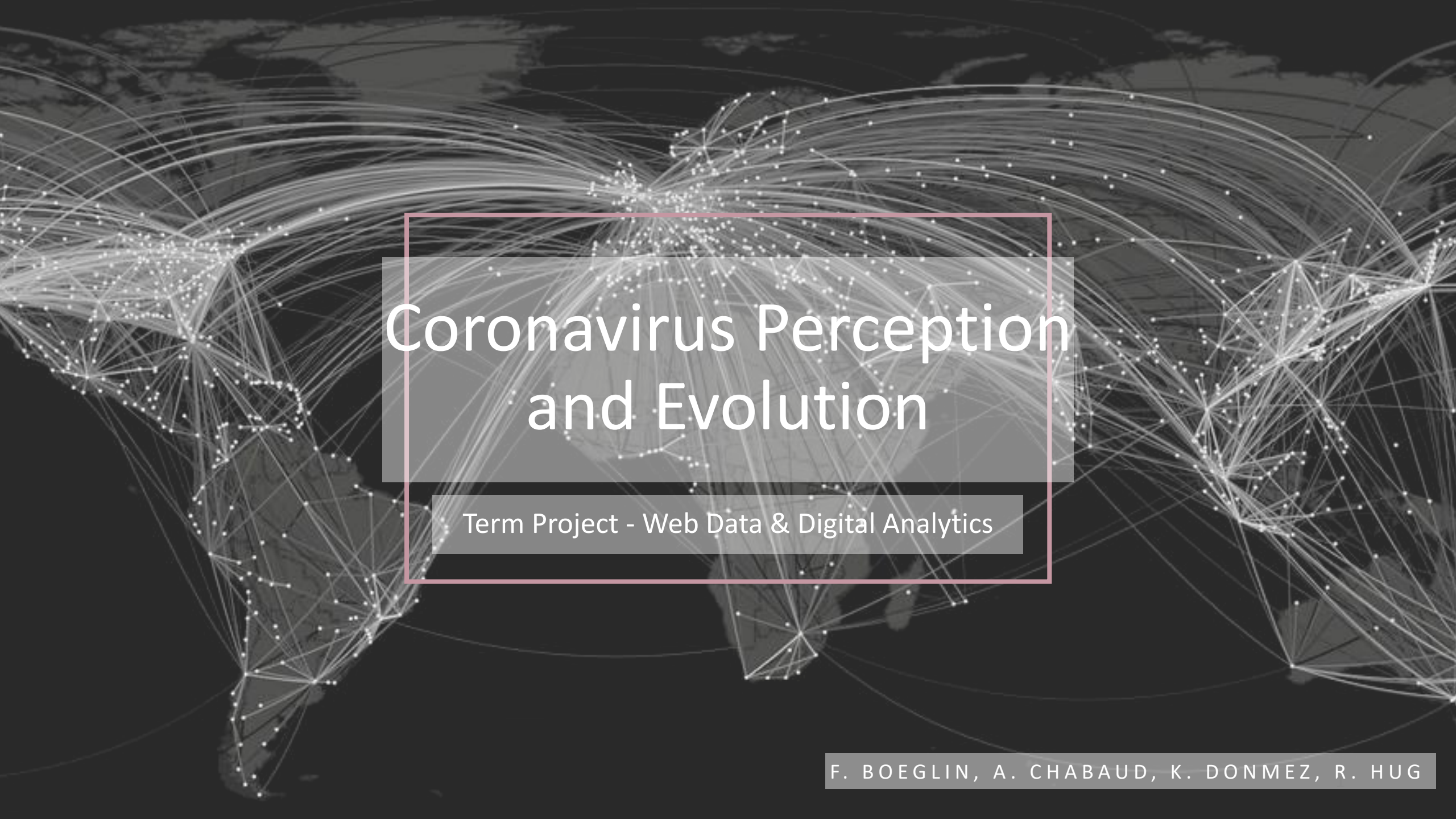


Coronavirus Perception and Evolution

Term Project - Web Data & Digital Analytics

F. BOEGLIN, A. CHABAUD, K. DONMEZ, R. HUG

Universität St.Gallen

INTRODUCTION

How can we use network analysis and text mining to analyze the perception and evolution of Coronavirus?

By using available information about the spread of the Coronavirus and the measures taken by different countries, we try to see :

- **Network analysis**: what are the similarities between the spread of the virus and the structure of worldwide air routes?
- **Text Mining**: how did the overall sentiment about this pandemic evolve over time, and what were different countries' reactions?

By combining our findings, we develop conclusions about the perception and evolution of Coronavirus between selected countries.

NETWORK ANALYSIS

THE DATA

- OpenFlights – 2012
- 59,036 routes
- 3,209 airports
- 531 airlines
- Directed network

THE ANALYSIS

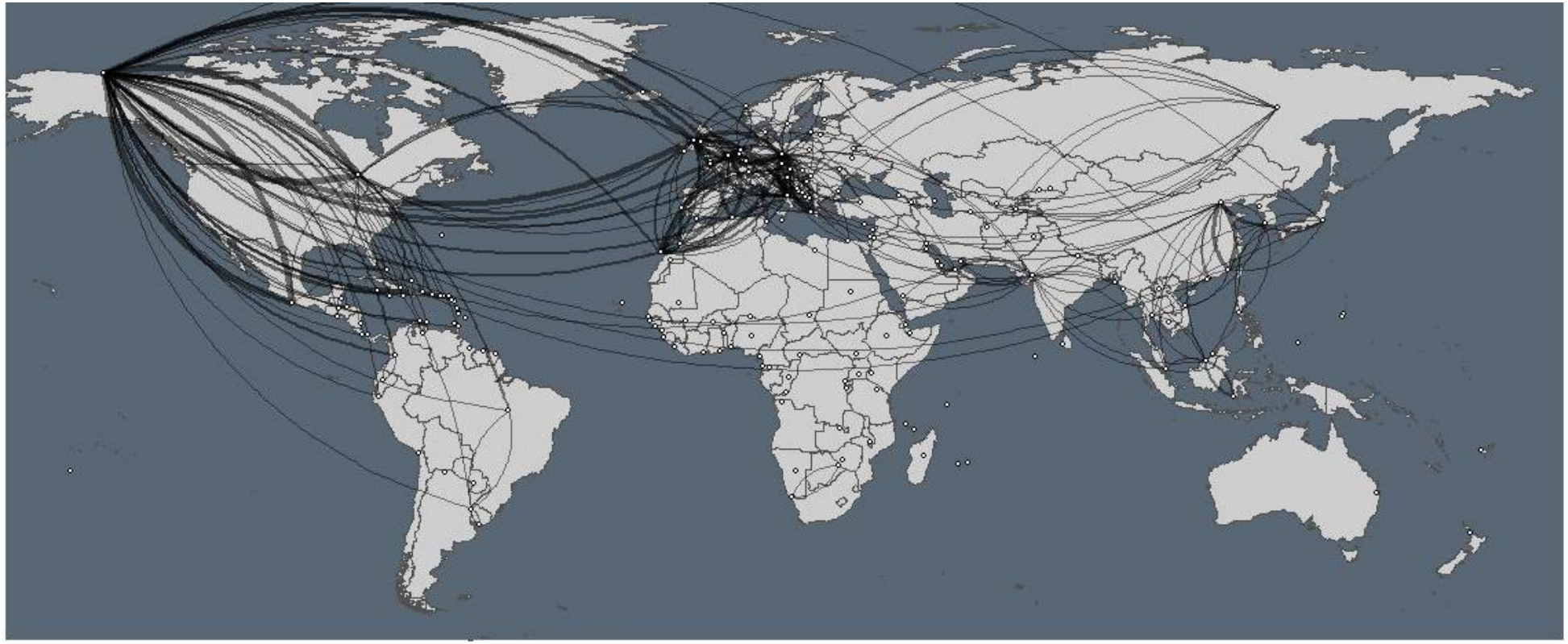
- Modeling the network
- Hubs
- Comparison with Coronavirus map

Modeling

Worldwide air routes

- **Very heavy database**: originally listing the connections between airports, by airlines; grouped by country
- **Weight of connection**: same connection from different airports
- **Representation**: only the connections with weight > 5; width of the edges represents weight

- **Observations**: Graph density is 0.214, but some areas seem more connected
  - Mediterranean basin
  - Europe-US
  - US-Central America
  - South-East Asia



Hubs

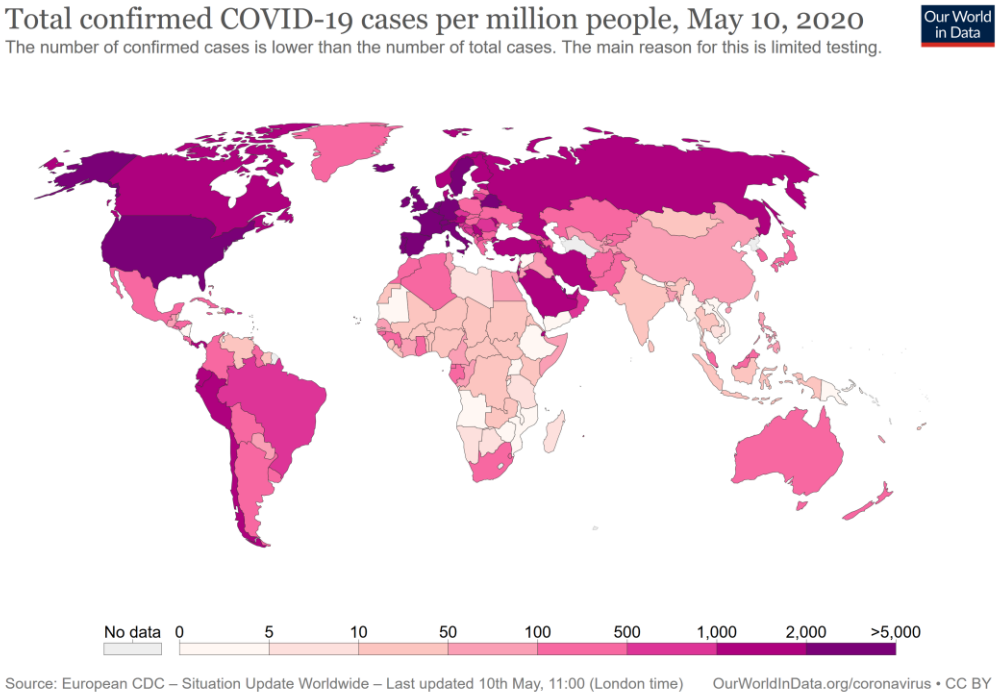
Central countries for worldwide air routes

- **In line with model visualization**: mostly countries in Mediterranean basin + United States & Mexico
- **Apparent similarity with COVID-19 cases**: 8 of the 14 most touched countries are also among the top 14 worldwide hubs

Hubs Ranking	
United States	1
Germany	0,764852
United Kingdom	0,700827
France	0,605118
Canada	0,590904
Italy	0,583538
Mexico	0,498101
Spain	0,465886
Greece	0,264316
Morocco	0,143263
Austria	0,132506
Netherlands	0,112909
Algeria	0,106361
Switzerland	0,10193

COVID-19 Cases per one million population*	
Spain	5661
Ireland	4657
Belgium	4612
United States	4133
Singapore	4072
Italy	3623
Switzerland	3502
United Kingdom	3229
France	2711
Portugal	2705
Sweden	2606
Belarus	2530
Netherlands	2488
Germany	2051

*Of countries with a population over four million; China and Iran excluded
As of 11 May 2020 10:39 GMT
Source: worldometer



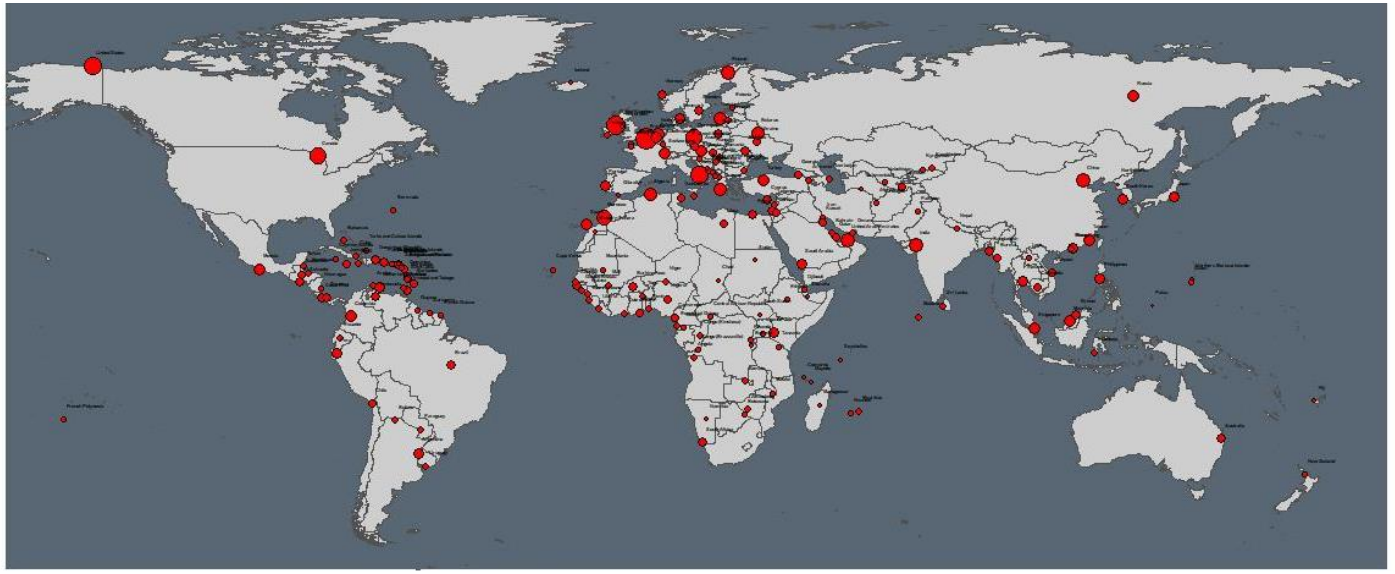
Universität St.Gallen

Map Comparison

Inferences by comparing the two maps

- **Similarities**: in general, it seems that there is some correlation between the worldwide air routes and the spread of the Coronavirus
- **Multitude of factors**: there are many factors that explain the spread, and the worldwide air routes are only one aspect
- **Limitations**: our model presents limitations, as our data is static & older vs. a dynamic & constant evolution of the pandemic

- **Countries for text mining**: choice of countries based on this network analysis
  - US: main hub and strongly touched
  - Germany: second biggest hub, but not too strongly touched
  - UK: third hub and interesting change of strategy



TEXT MINING

THE DATA

- Opinion leaders' tweets from selected countries
- US, UK, Germany
- A politician, a national influencer, a news media

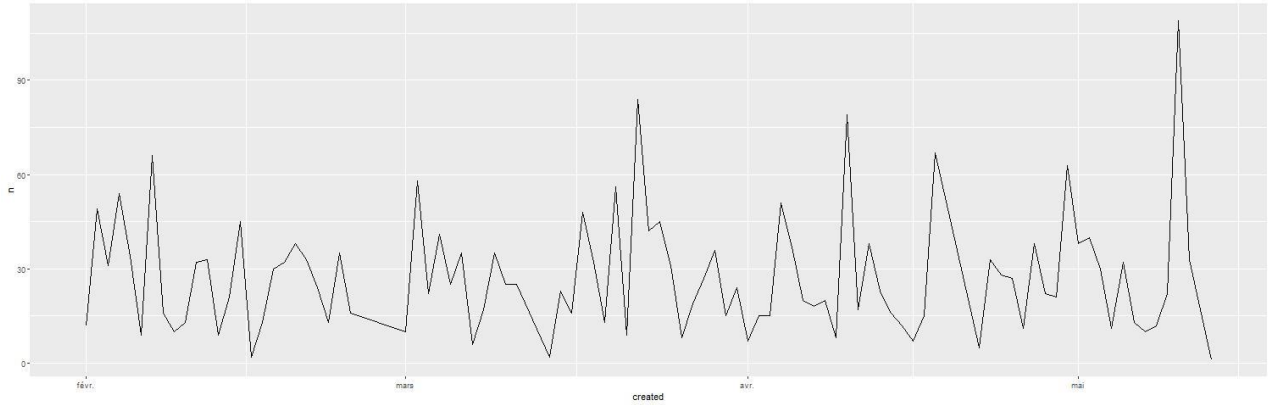
THE ANALYSIS

- Measures of the three countries
- Sentiment analysis over time
- Comparison between opinion leaders in each country
- Comparison across countries

Sentiment Analysis

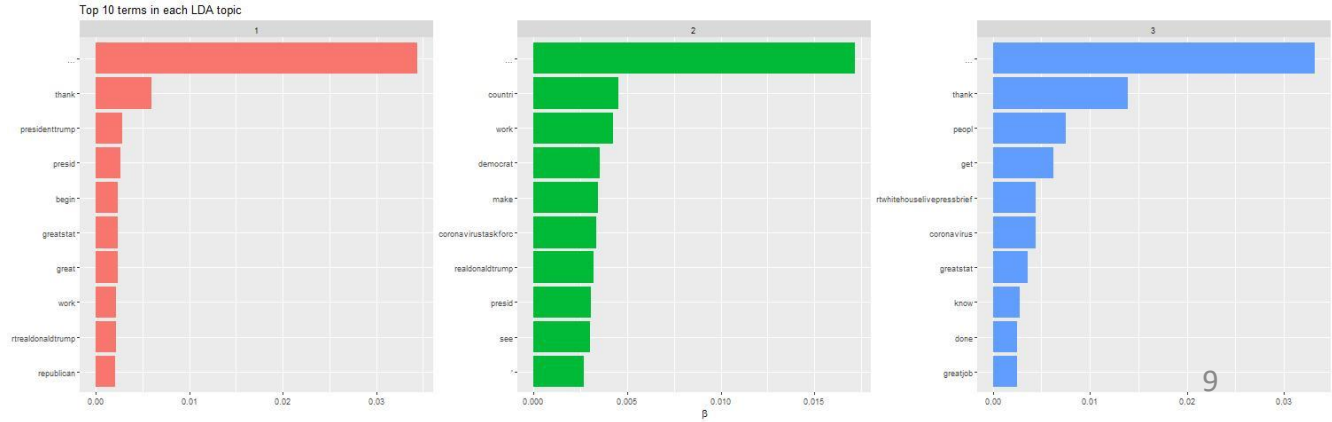
Procedure

- **Collecting the data**:
  - Limitation in tweet numbers; challenge of finding people who tweet enough but not too much and who are opinion leaders for the country
  - Taking only tweets between February 1st and May 12th
  - Sometimes taking retweets when interesting
- **Analyzing and exporting the data**:
  - Sentiment evolution over time for each person
  - General and polarized wordcloud for each person to check for main words
  - Topic analysis for each person to make sure that Coronavirus is at least one main topic
  - Aggregated sentiment evolution and polarized wordcloud for countries comparison
- **Comparing and making inferences**:
  - Keeping in mind our limitations
  - Dynamic evolution; spread of the virus inside the country and actions taken over time
- **All the graphs are in the appendix**



negative

positive



United States

Opinion Leaders' Level

OBSERVED PERSONS

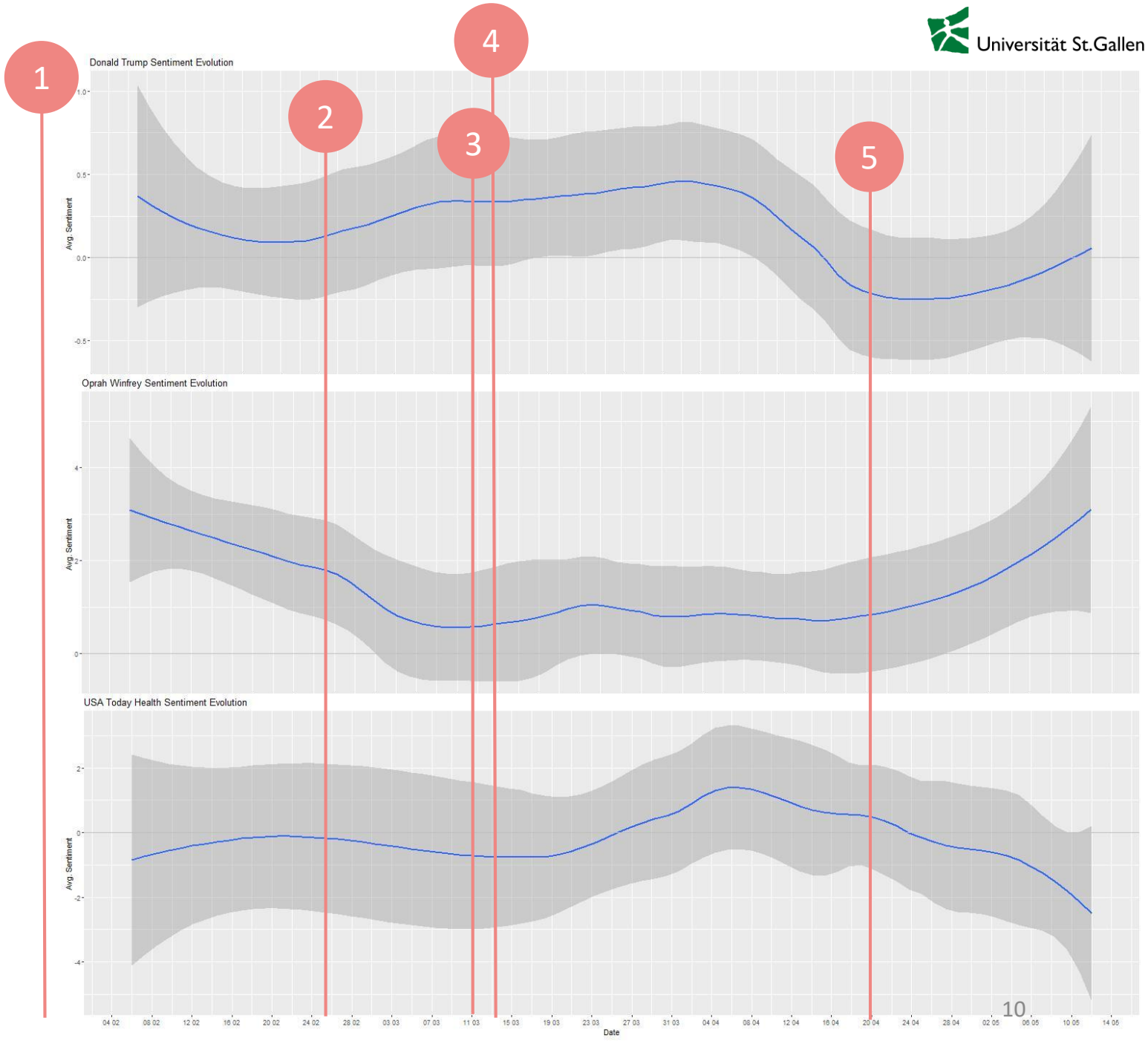
- **Trump** (with retweets) – sentiment values are close to 0
- **Oprah Winfrey** (with retweets) – highest sentiment values
- **USA Today Health** – sentiment is most of the times below 0

SENTIMENT EVOLUTION

- Sentiment evolution of Donald Trump continuously increases after end of February while Oprah Winfrey and USA Today Health sentiment evolution goes down
- Sentiment of Donald Trump and Oprah Winfrey goes up after denial of immigration; sentiment of USA Today health goes down

IMPORTANT POINTS IN TIME/POSSIBLE TURNING POINTS

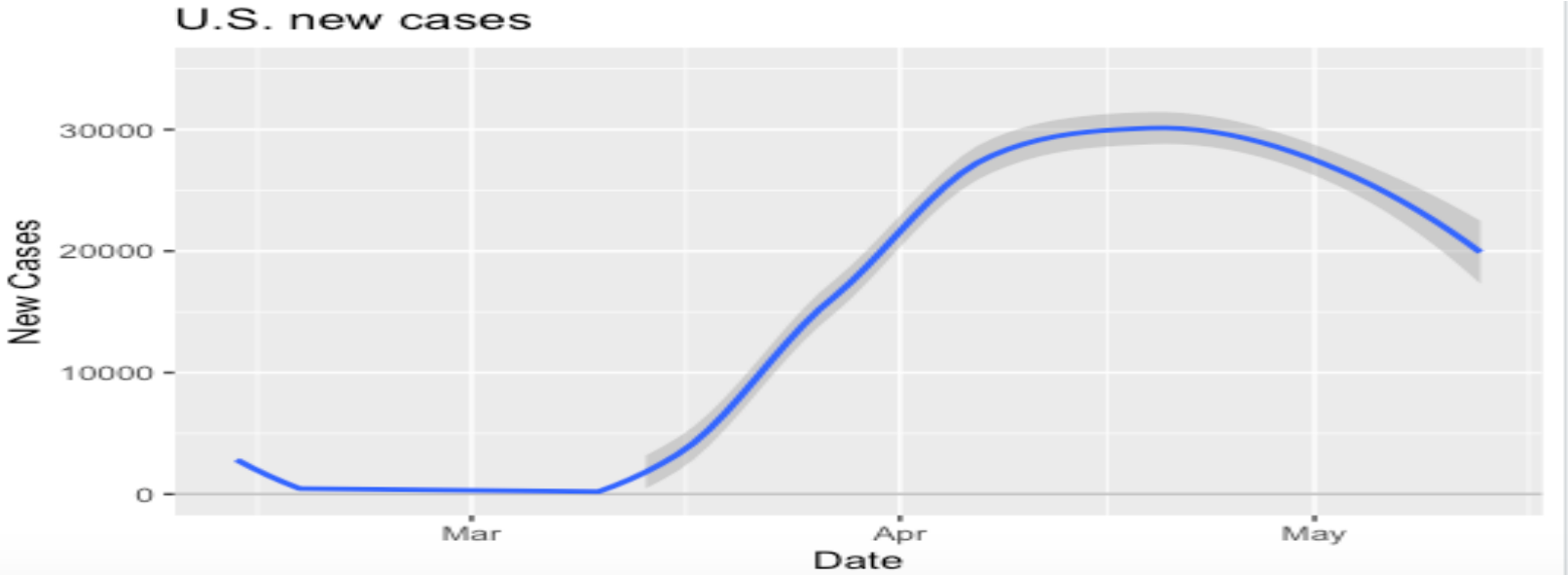
1. First COVID-19 case confirmed
2. US CDC warns that the spread to the United States is likely and that people should prepare
3. WHO declares COVID-19 as a pandemic
4. States start to cancel gatherings of more than 250 people
5. Donald Trump announces he will temporarily suspend immigration to the US for 60 days by executive order.



United States

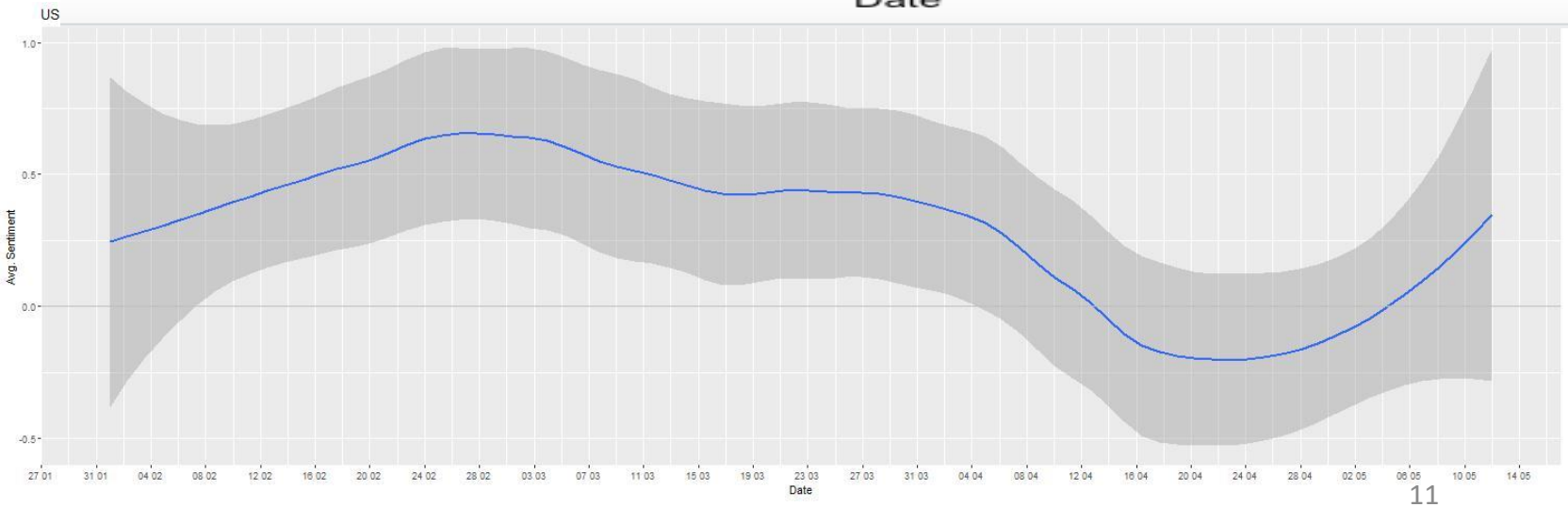
Country's Level

- New daily cases seem to have an inverse relationship with daily sentiment
- The sentiment decrease can be seen starting before the virus started growing
- Wordcloud shows "fake" (comes from Donald Trump) and "hard", as well as "strong", "protect" and "support"



negative

positive



United Kingdom

Opinion Leaders' Level

OBSERVED PERSONS

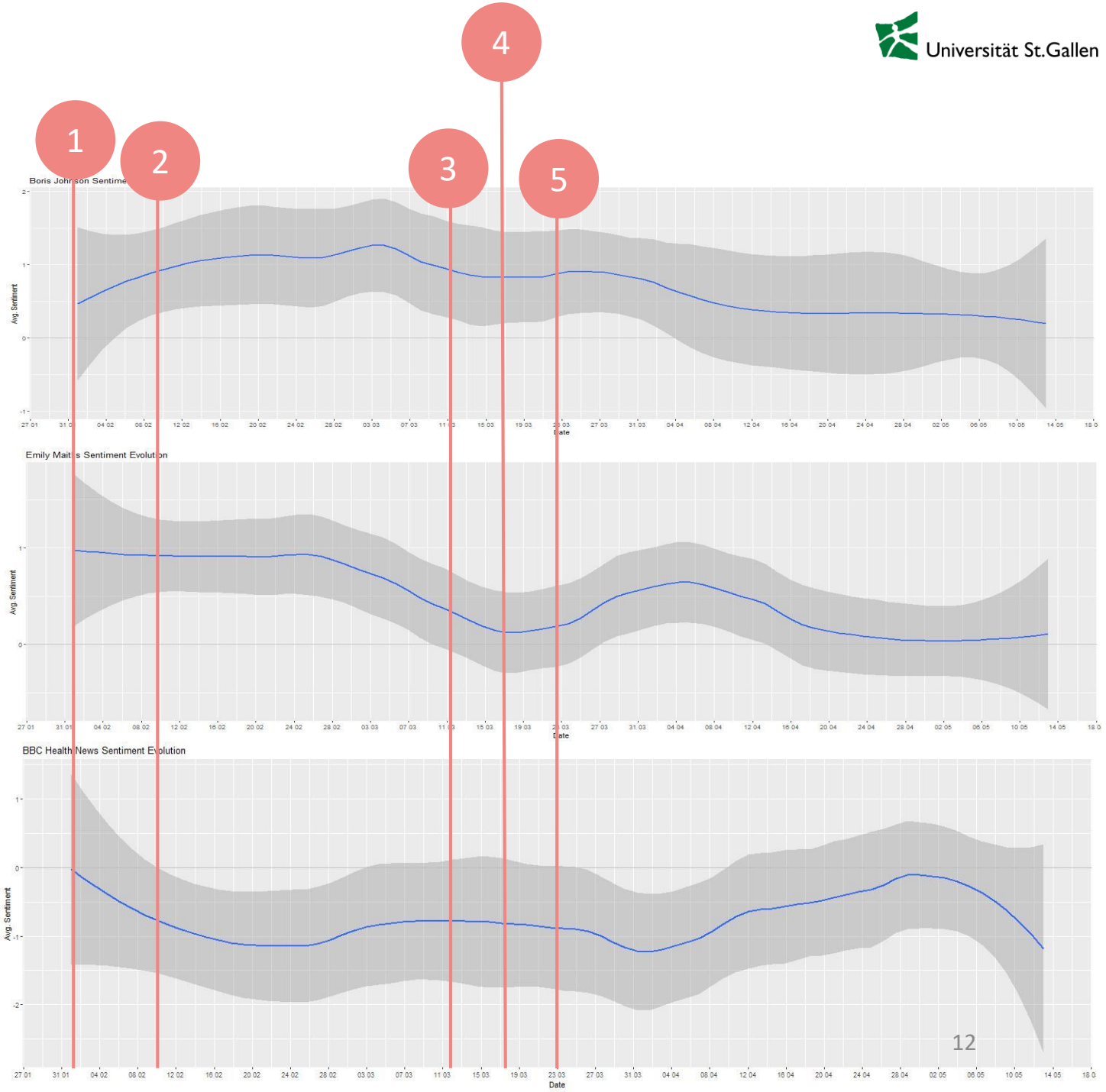
- **Boris Johnson** – sentiment goes up and then gradually decreases
- **Emily Maitlis** – compared to the others, high starting point and with time close to 0
- **BBC Health News** – always below 0

SENTIMENT EVOLUTION

- Emily Maitlis' sentiment evolution develops against the direction of the others (beginning of April)
- Beginning of April, sentiment of Boris Johnson goes gradually down
- BBC health decreases and then increases again mid March; it sinks again after May
- BBC's sentiment increases when Boris' and Emily's gradually goes down

IMPORTANT POINTS IN TIME/POSSIBLE TURNING POINTS

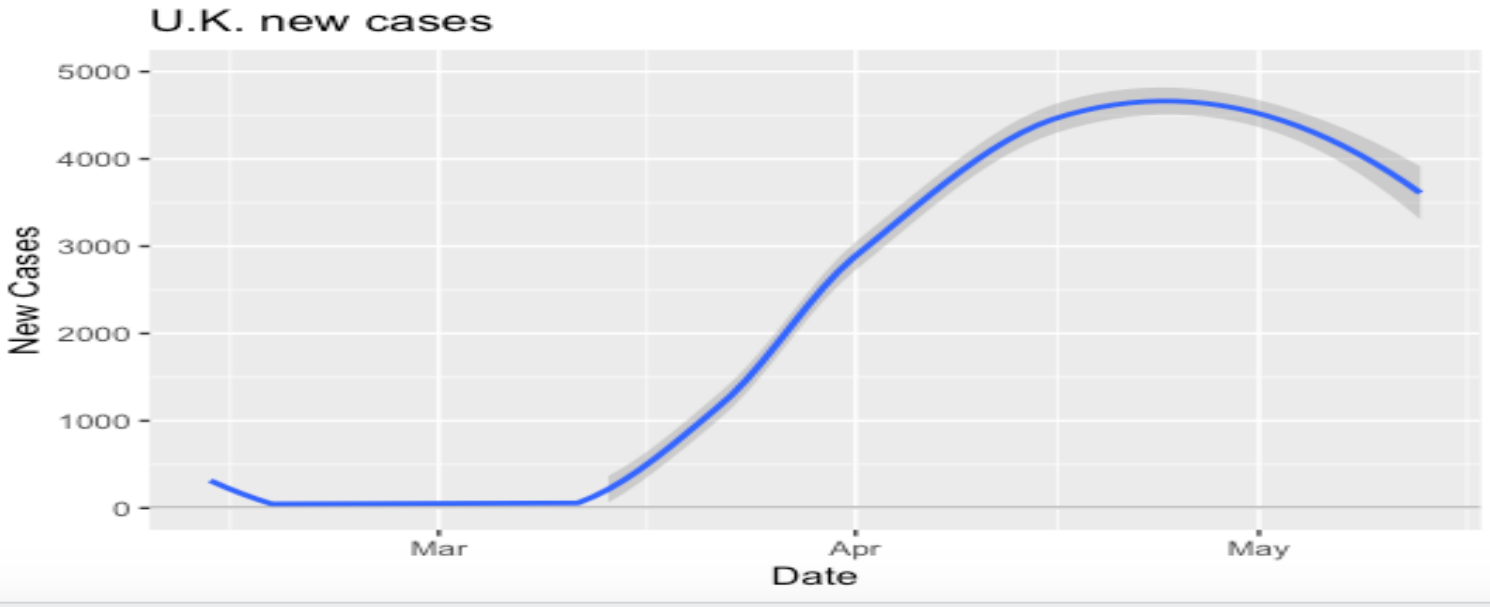
1. First COVID-19 case confirmed
2. WHO declares COVID-19 as a pandemic
3. Declaration that the coronavirus constitutes a serious and imminent threat to public health
4. UK advises citizens against all but essential international travel and strongly starts advising social distancing
5. Lockdown

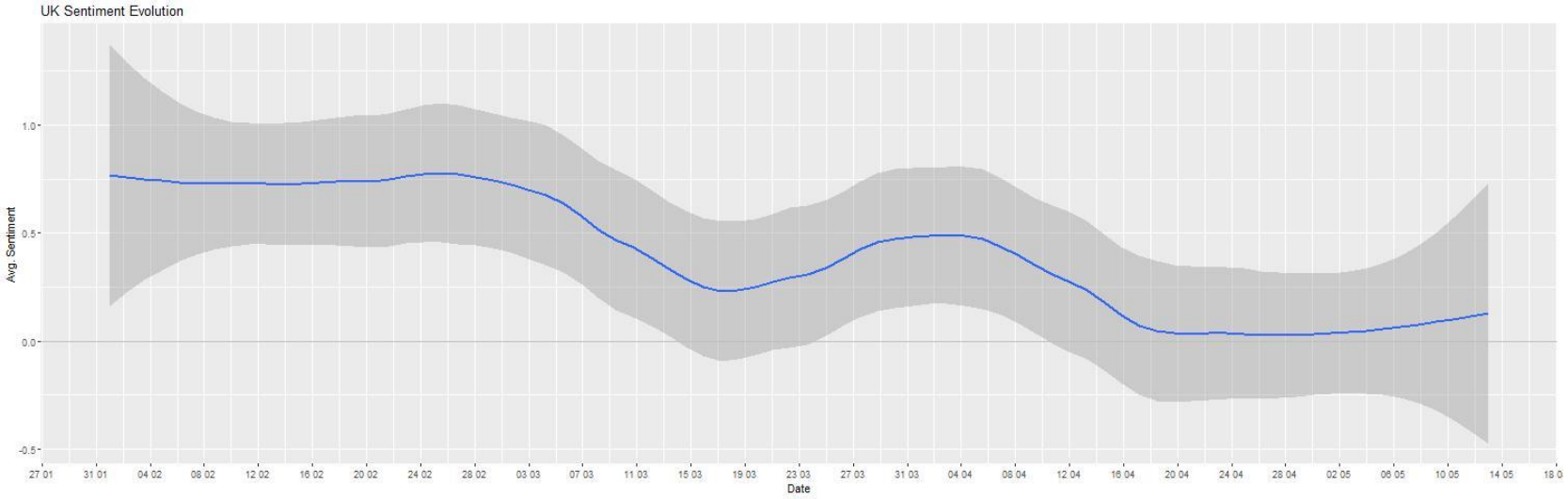


United Kingdom

Country's Level

- Interesting case due to complete change of strategy
- Interestingly, sentiment started turning positive as cases started growing; it only started going down again after change of strategy and lockdown implementation
- Wordcloud points to heavy focus on protection but also points towards current crisis and death





Germany

Opinion Leaders' Level

OBSERVED PERSONS

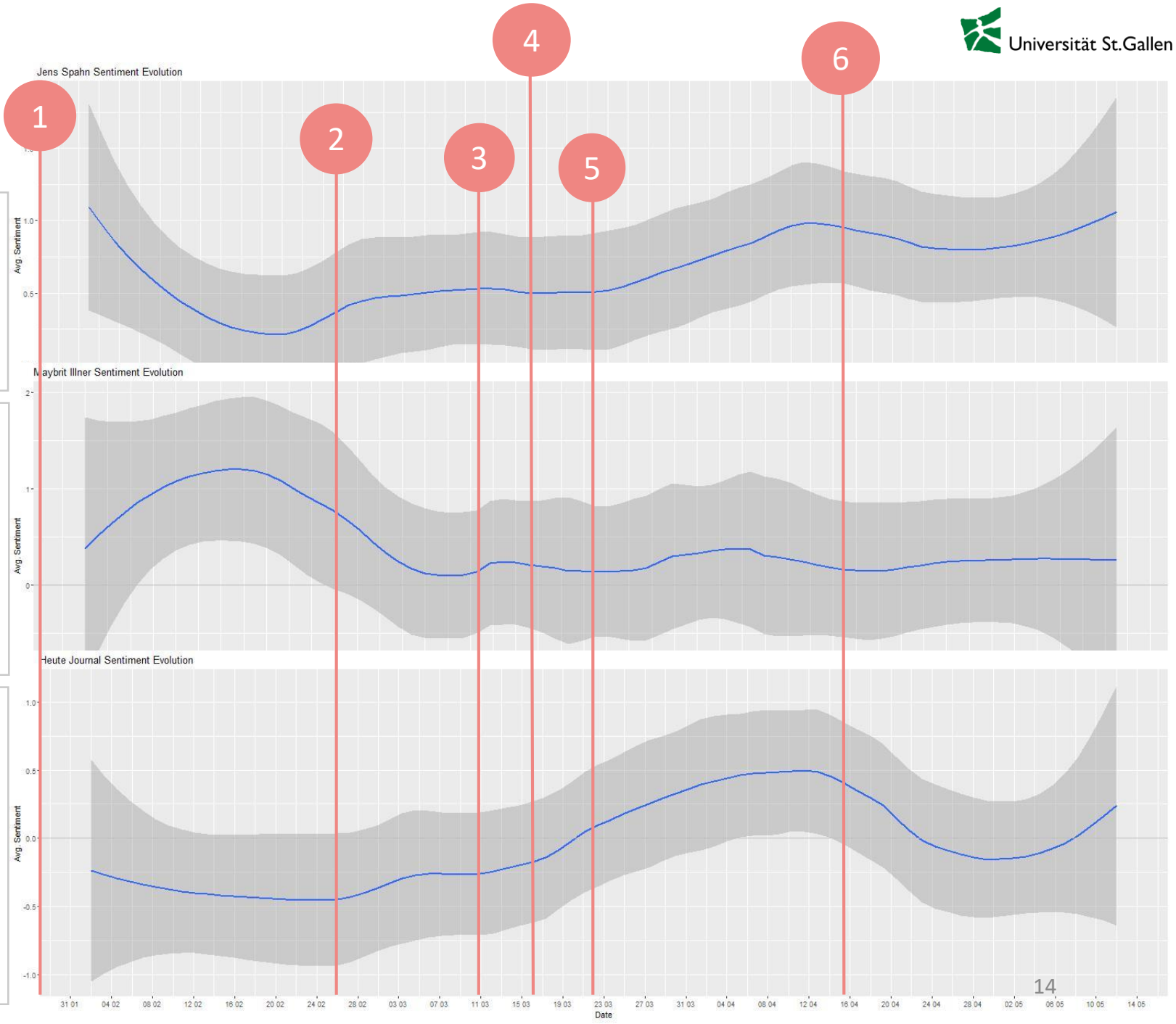
- **Jens Spahn** – low sentiment at the beginning and then gradually up
- **Maybrit Illner** – up and down at the beginning and then stays close to 0
- **Heute Journal** – stays close to 0, increase and decrease

SENTIMENT EVOLUTION

- Jens Spahn's sentiment is contrasting with Illner's and starts rising at the end of February; aligns with Heute Journal's sentiment evolution
- Maybrit Illner's sentiment peaks midst of February and then stays close to zero, without big changes
- Heute Journal's sentiment goes gradually up, unlike Maybrit's; it goes down after mid April and then rises again

IMPORTANT POINTS IN TIME/POSSIBLE TURNING POINTS

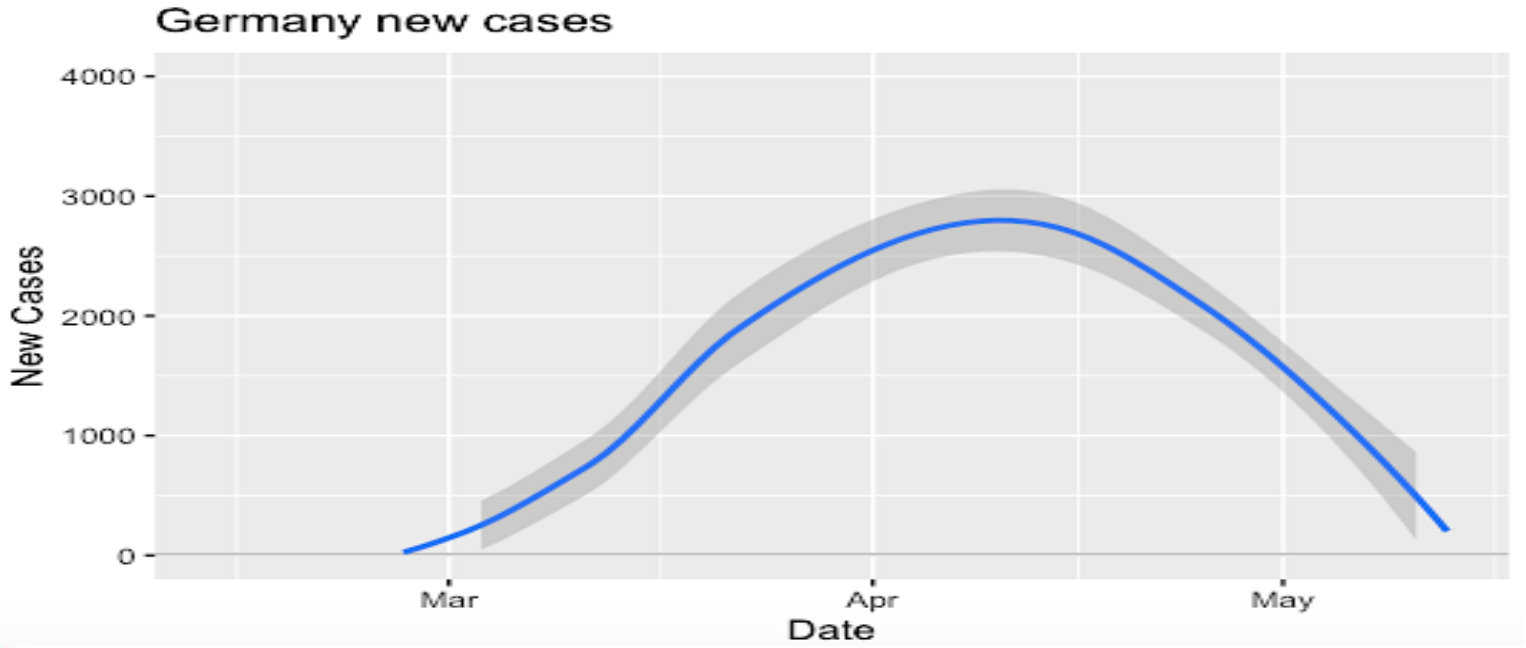
1. First COVID-19 case confirmed
2. Germanys minister of health announces that the country is at the beginning of an epidemic
3. WHO declares COVID-19 as a pandemic
4. Germany closes borders
5. "Contact ban" of more than two people
6. Germany announces plans to reopen their economy starting April 20th

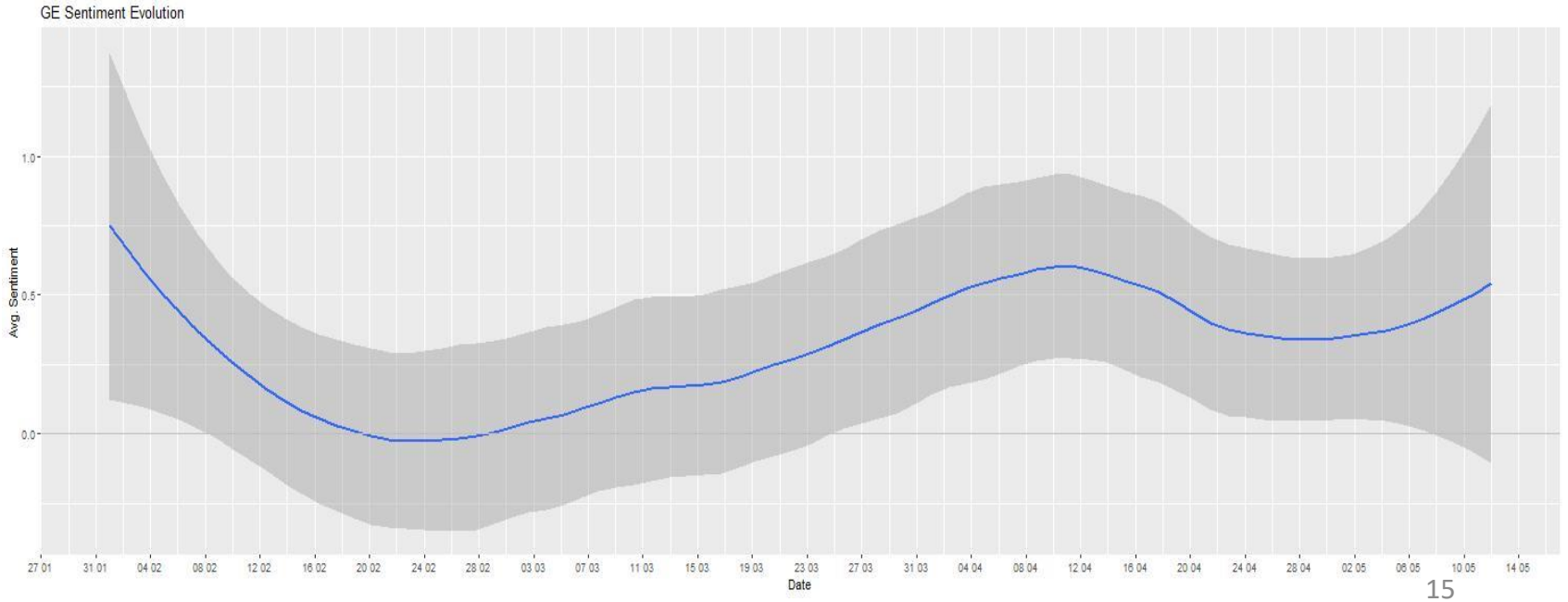


Germany

Country's Level

- There seems to be a lag between sentiment evolution and virus spread: sentiment declines before the virus starts growing, then it increases and starts declining again mid-April, until it goes back up
- The sentiment probably correlates with the measures, as people feel safer when they know that something is put in place to protect them
- Negative words are about the difficult time and the measures that are taken, positive words are about going through the situation together

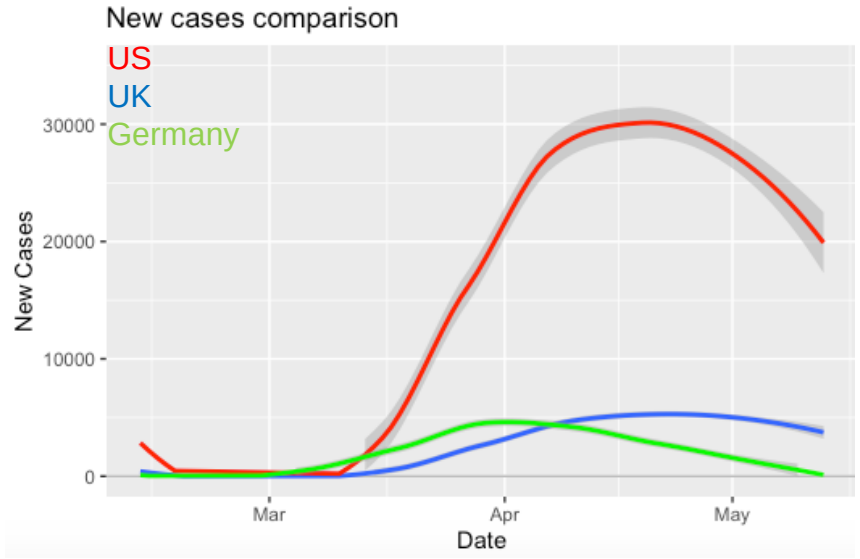


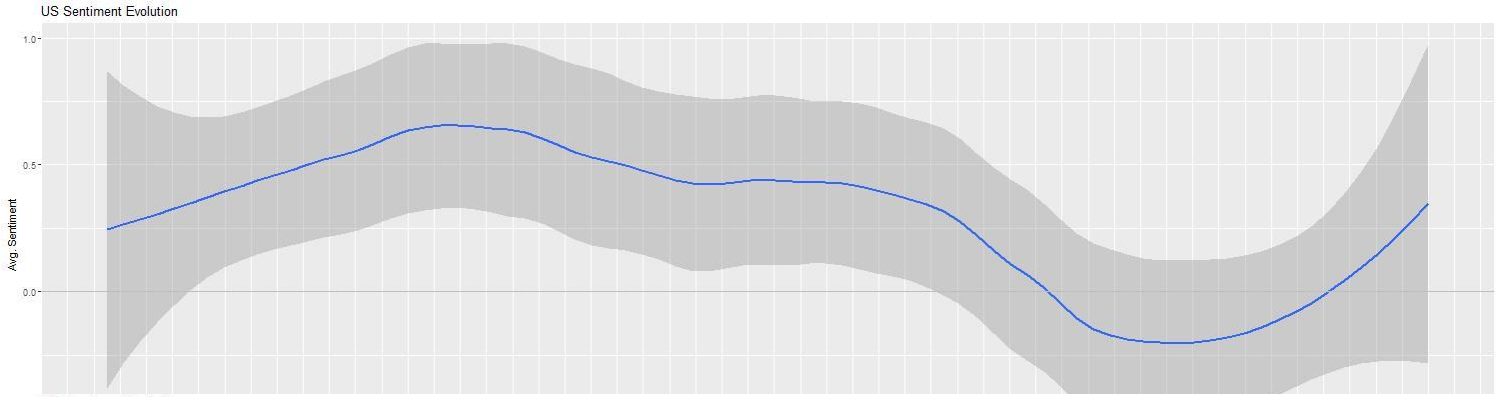


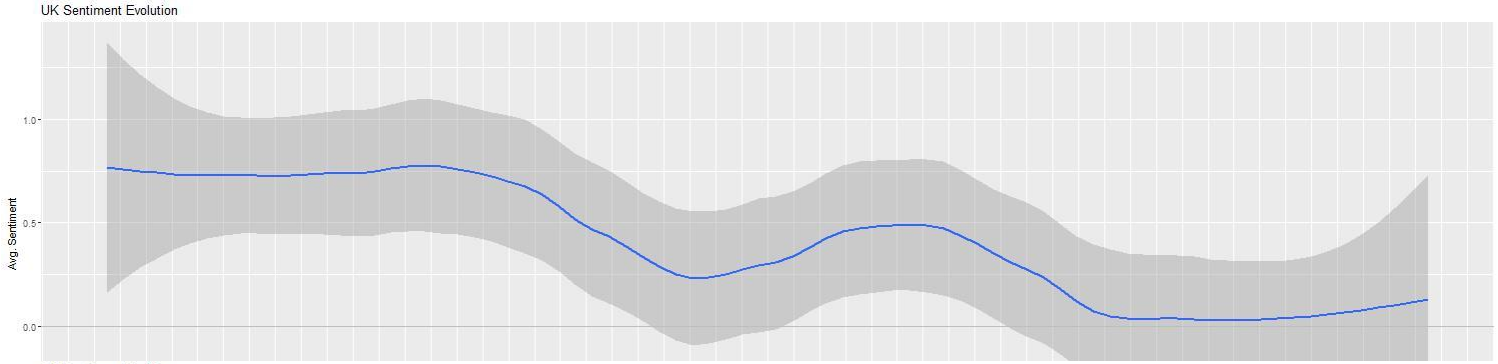
Countries' Comparison

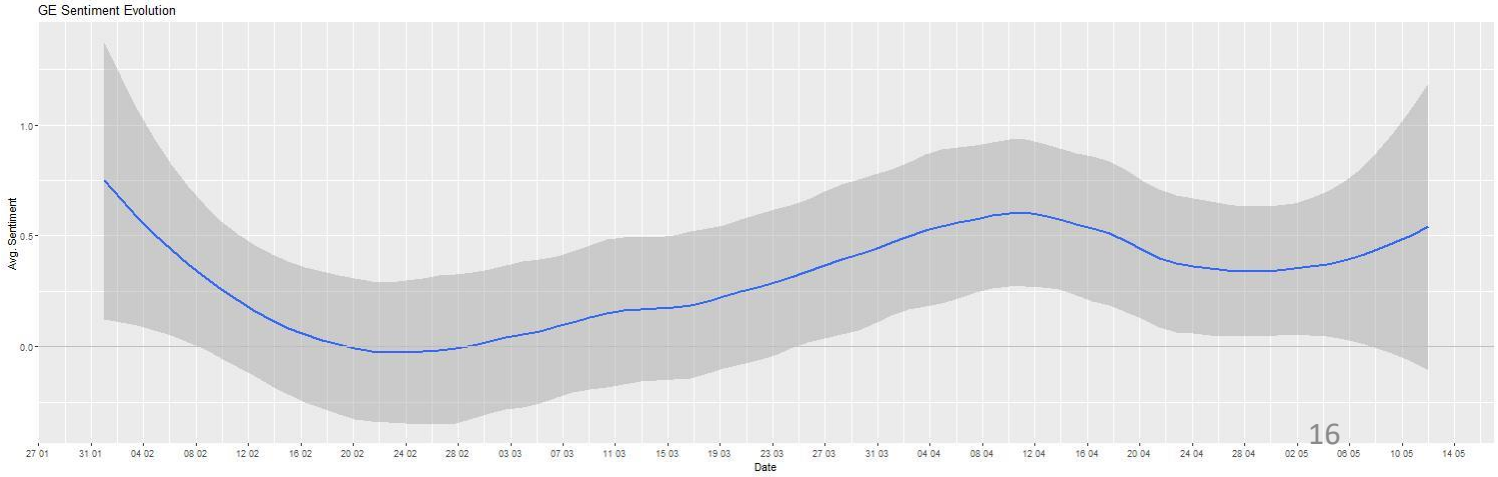
Sentiment Analysis Across Countries

- In all cases, it seems that a decrease of sentiment started before the virus actually started growing
- The sentiment curves of the US and UK seem to be similar when compared to Germany, which could be explained by their similarities in the "new cases" curves
- Germany seems to have dealt with the crisis faster and more effectively than the US and UK
- Hypothesis – Germany realized the dangers of the virus early, compared to UK and the US









FINAL NOTES

CONCLUSION & POTENTIAL EXTENSION

- Our results are mostly hypotheses, but they allow for some better understanding of the situation
- Big potential for many hypotheses testing, with a bigger and more representative sample:
  - How the sentiment evolves during a crisis (lag or not),
  - How efficient are different countries' reactions,
  - Who really influences the global sentiment during a crisis
  - …
- Comparison with past disease outbreaks could give us a better understanding of how to use sentiment as a prediction for crisis development

LIMITATIONS

- Old and static data for network analysis
- Limited data for tweets (maximum of 3,200 tweets)
- Non-representativeness of our sample and risk of bias
- Different languages, so difficulty of analysis
- Dynamic data, as the situation is not over

THANK YOU!

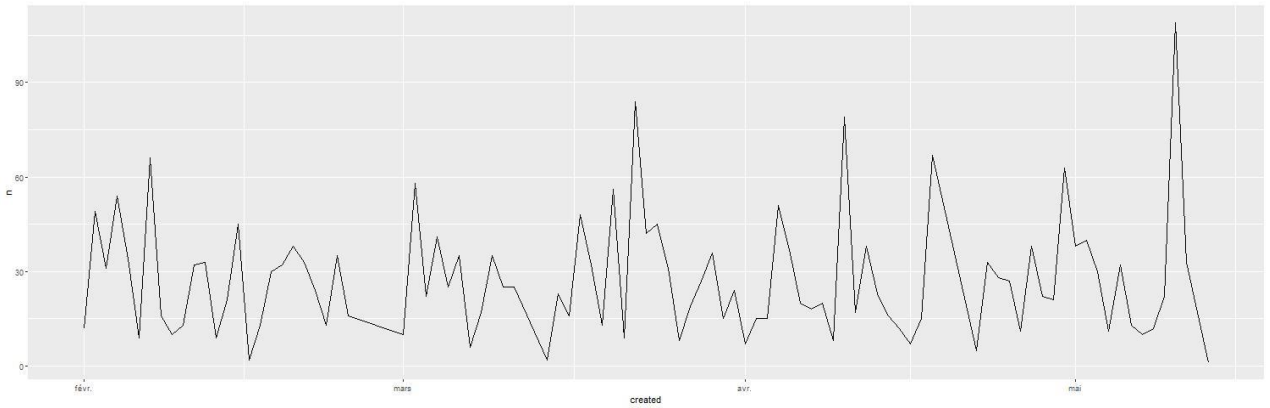
ANY QUESTION?

A P P E N D I X

Additional data

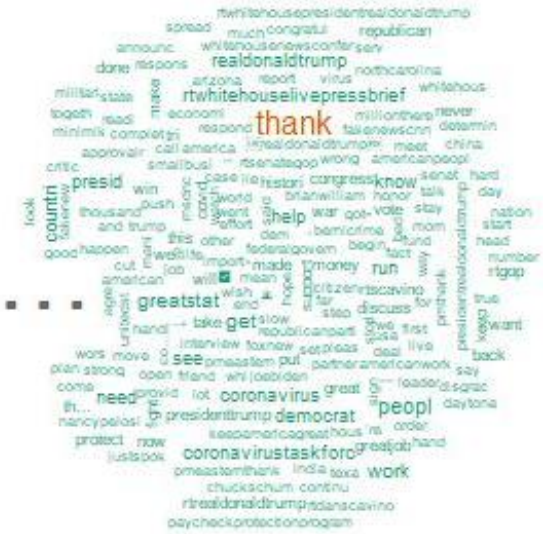
Donal Trump

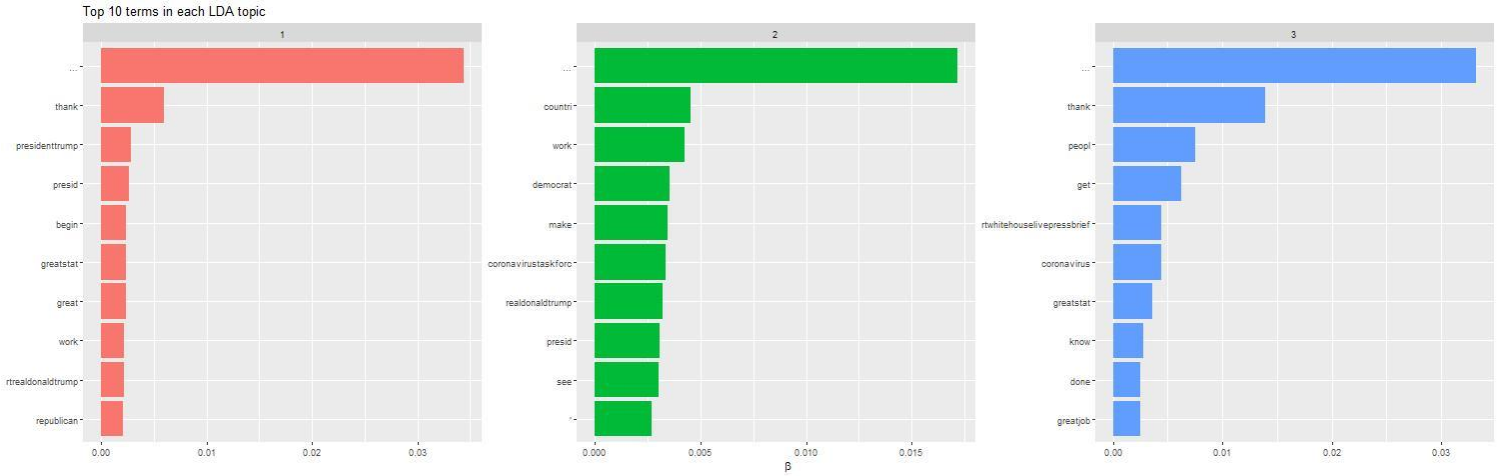
Wordclouds & Topics



negative

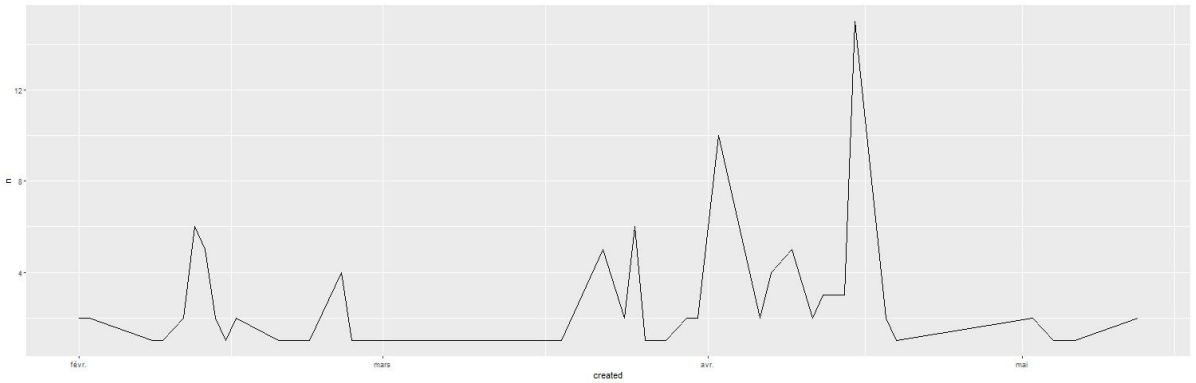
positive





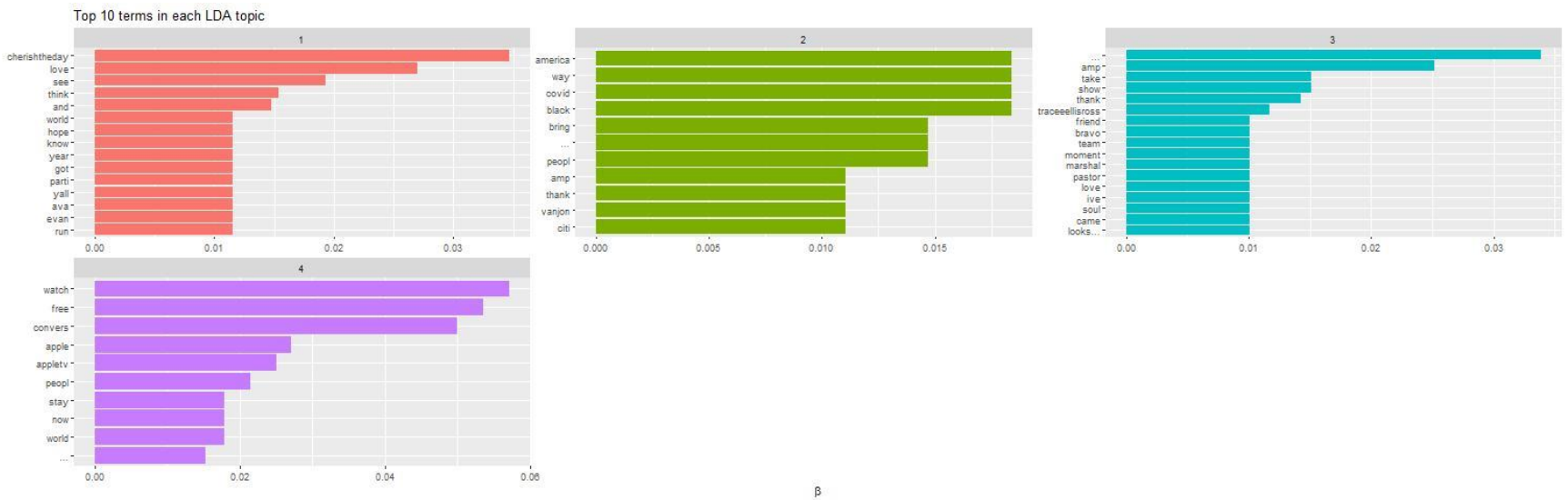
Oprah Winfrey

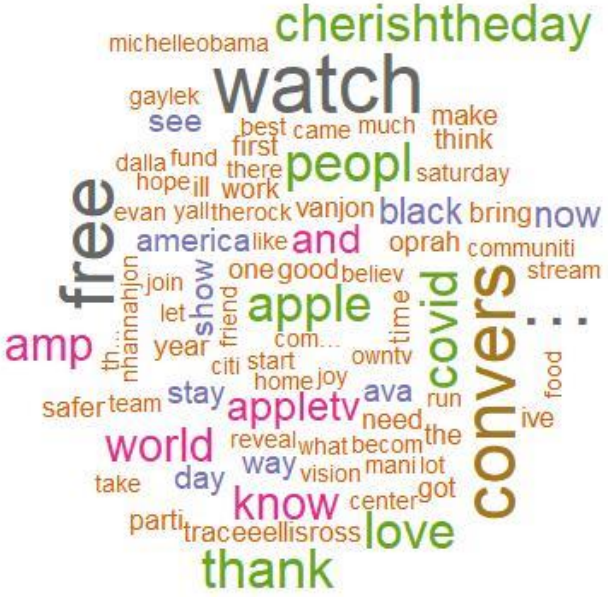
Wordclouds & Topics



negative

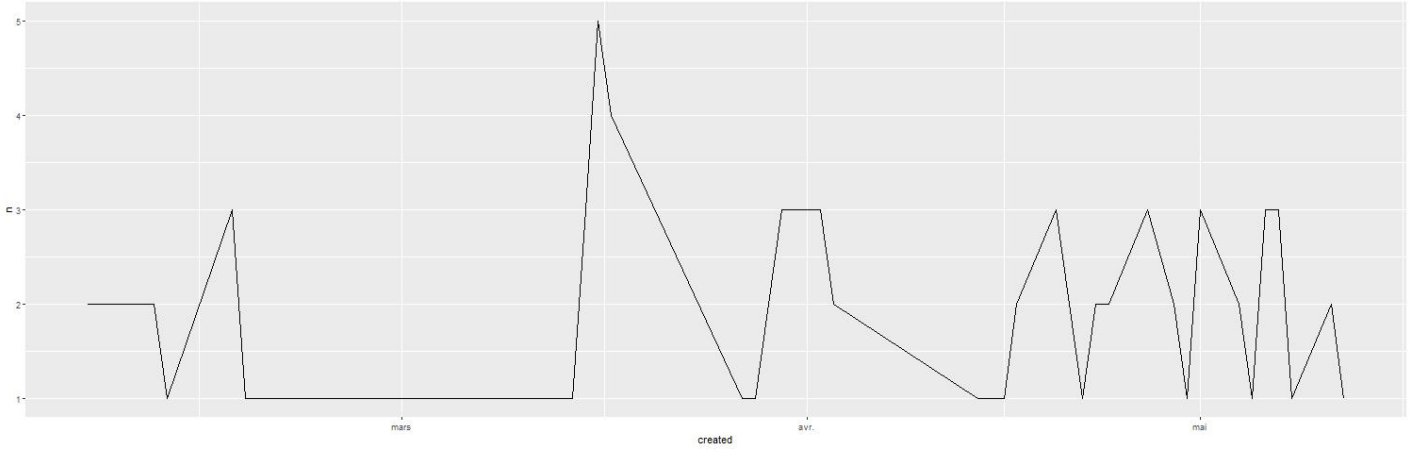
positive





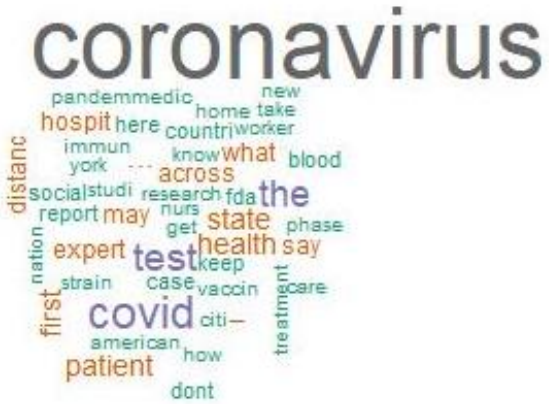
USA Today Health

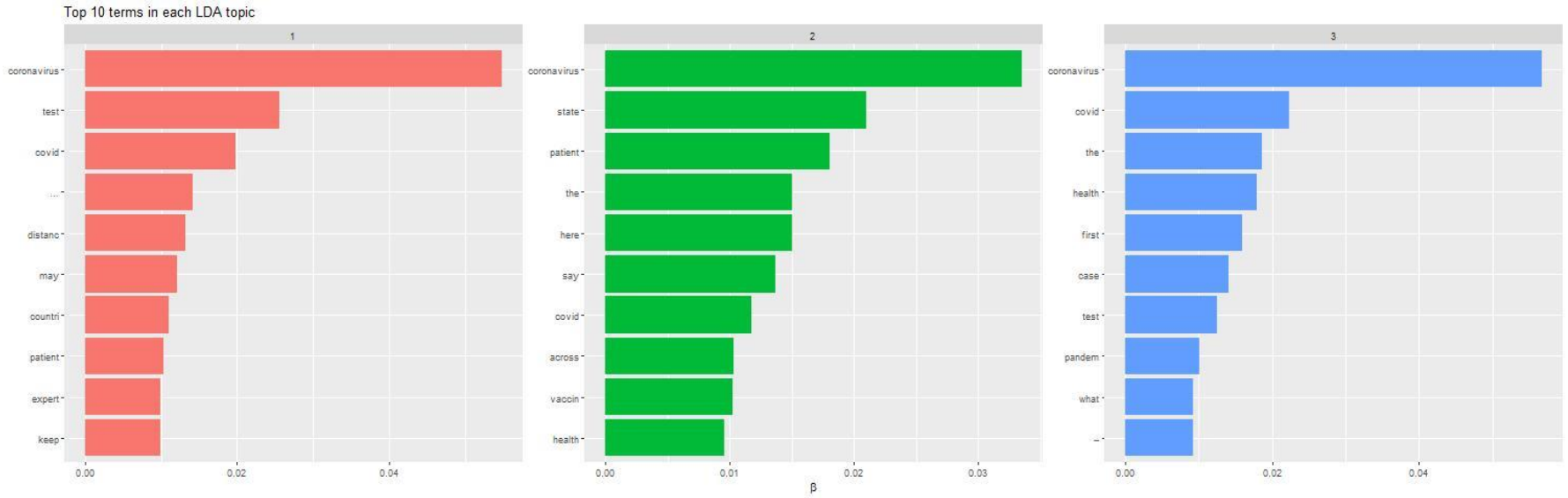
Wordclouds & Topics



negative

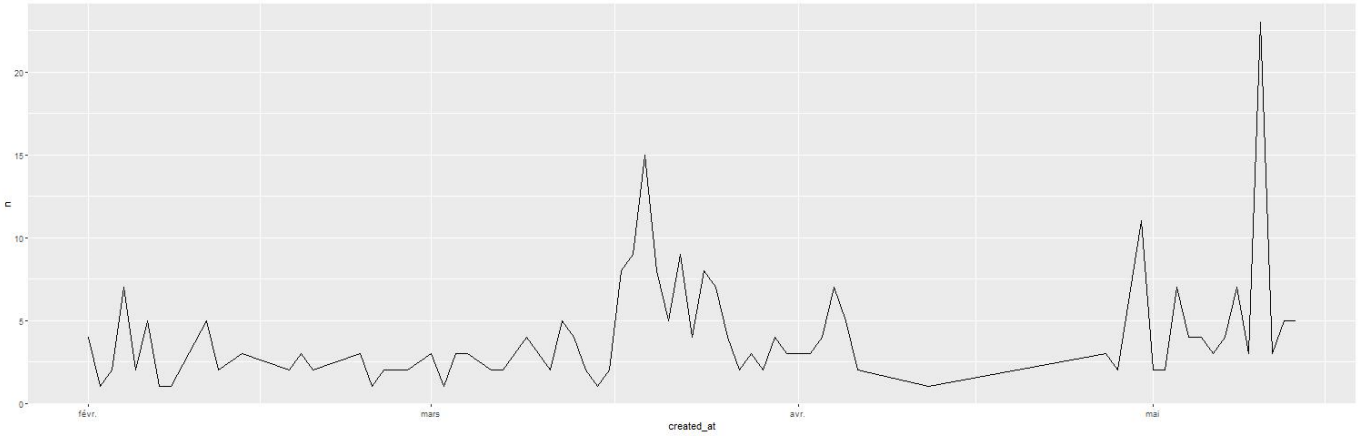
positive





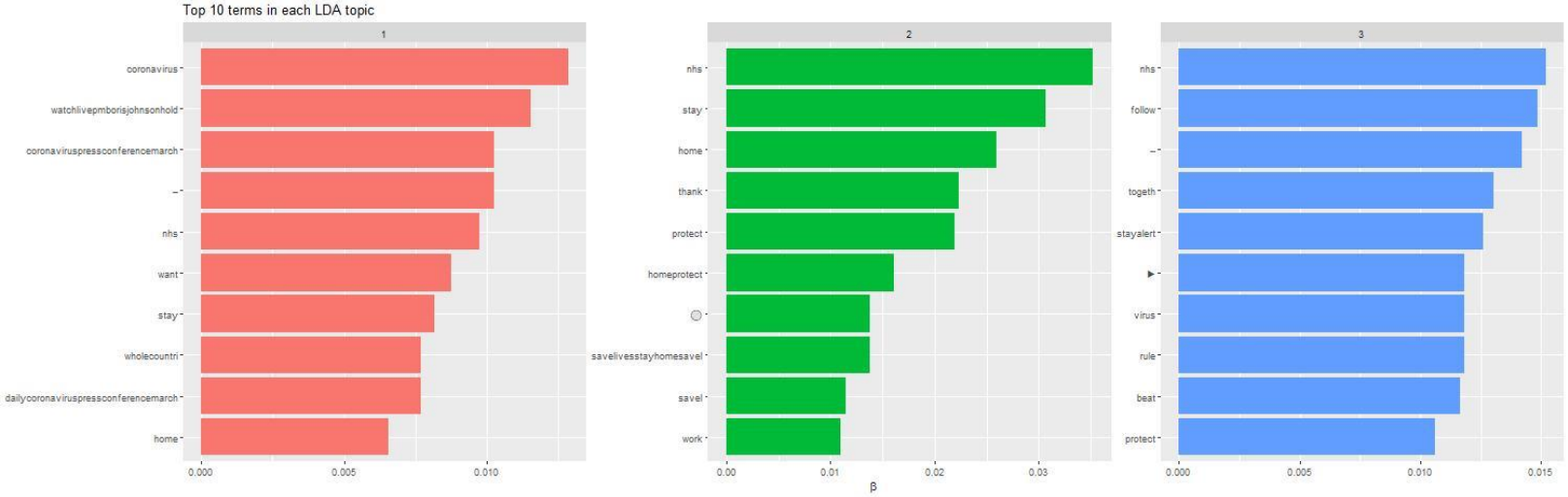
Boris Johnson

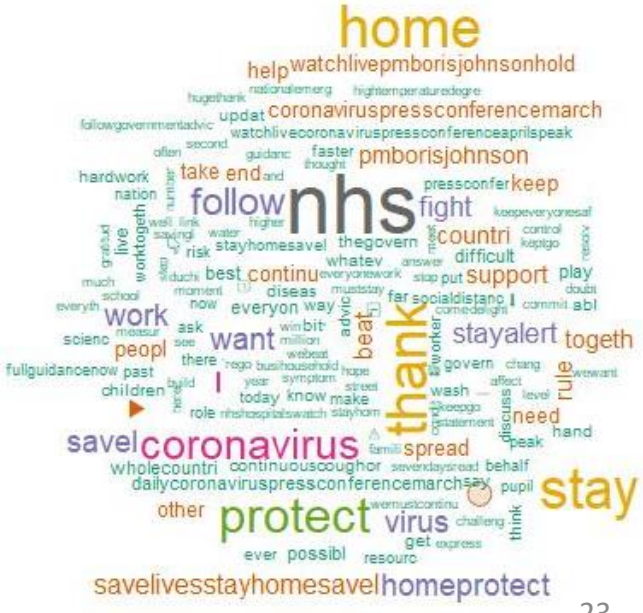
Wordcloud & Topics



negative

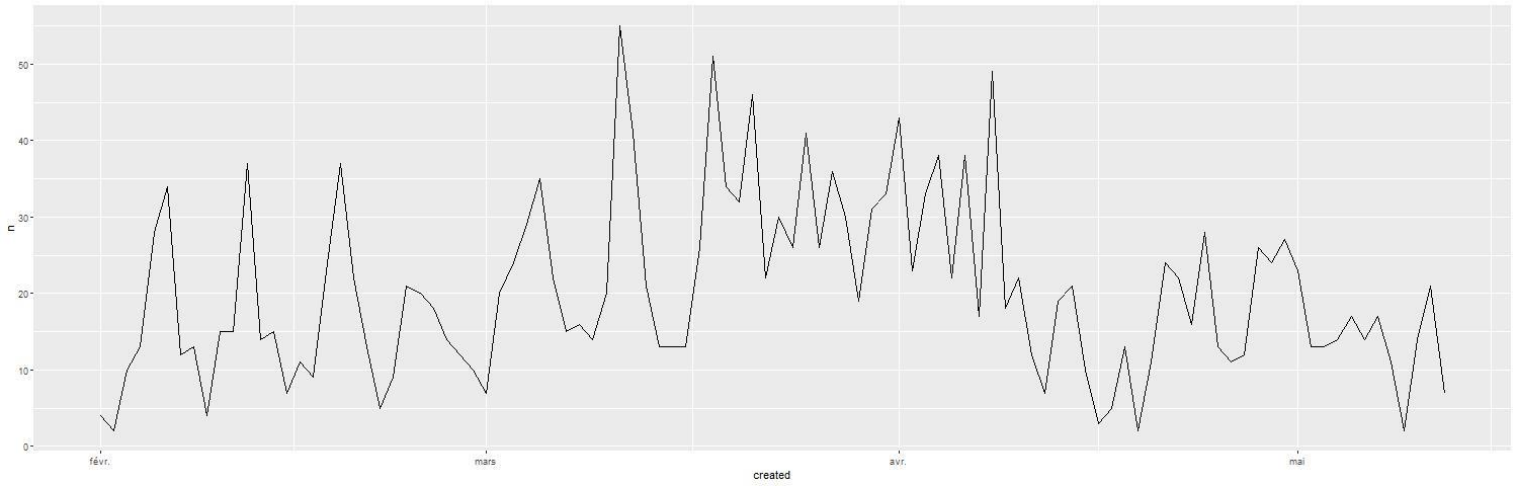
positive

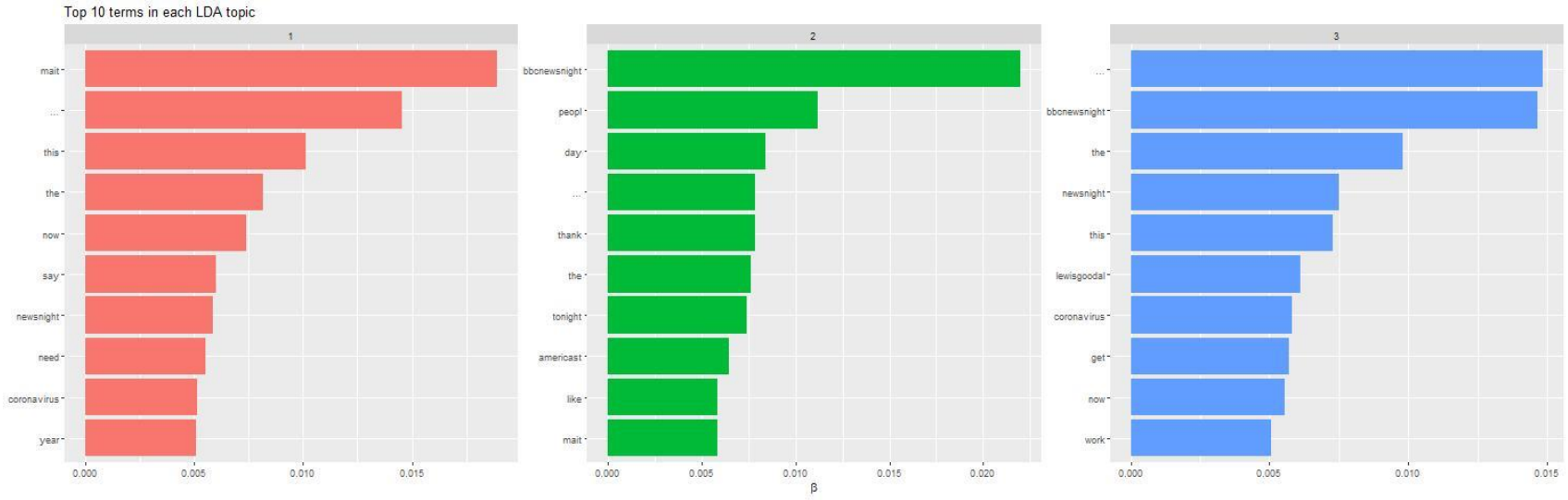


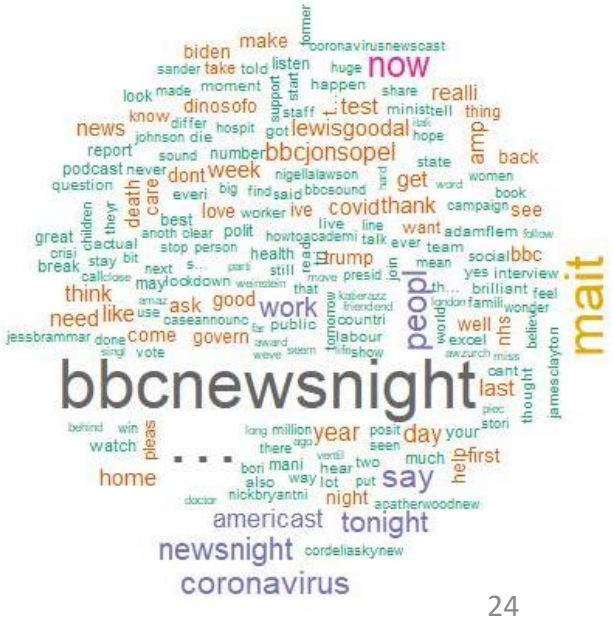


Emily Maitlis

Wordcloud & Topics

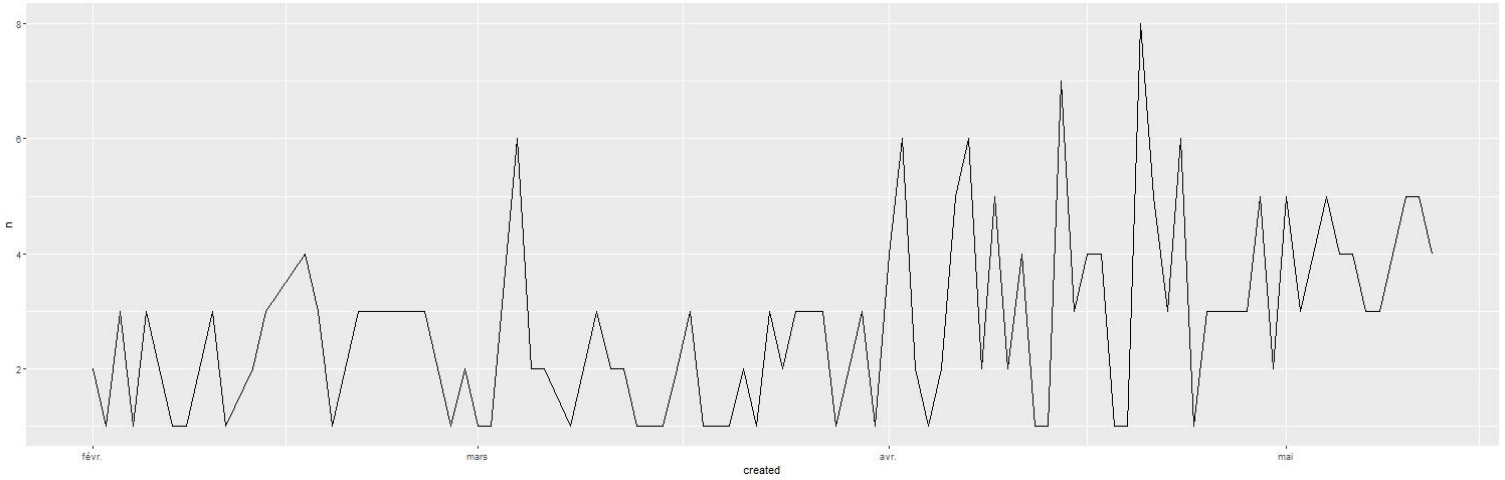


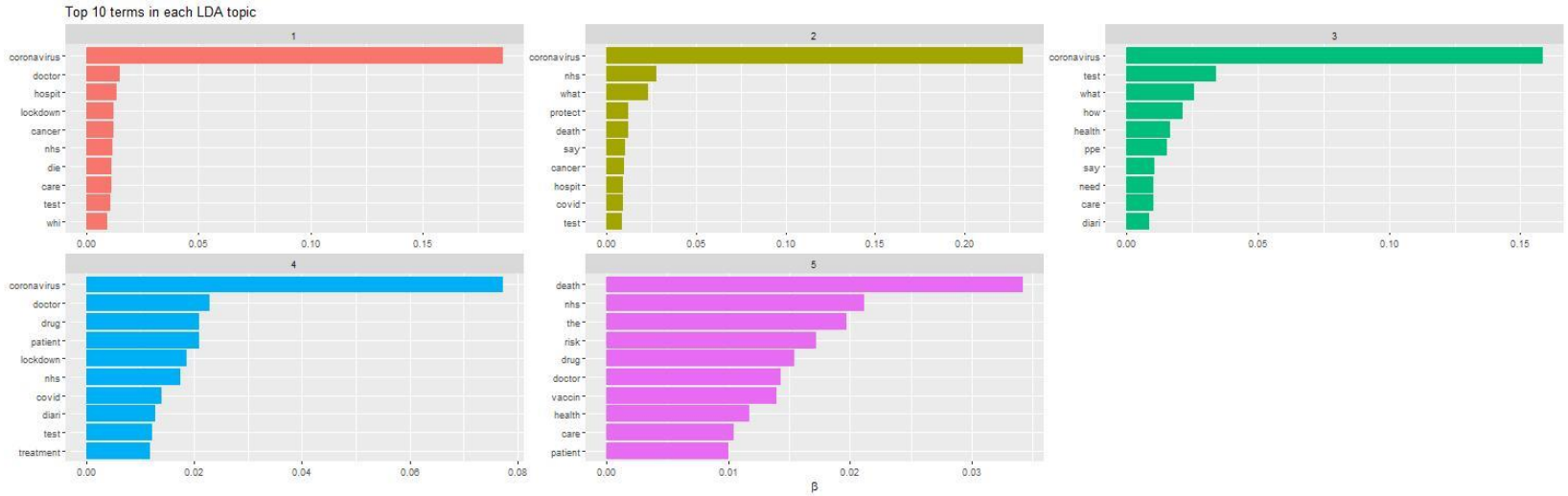




BBC Health News

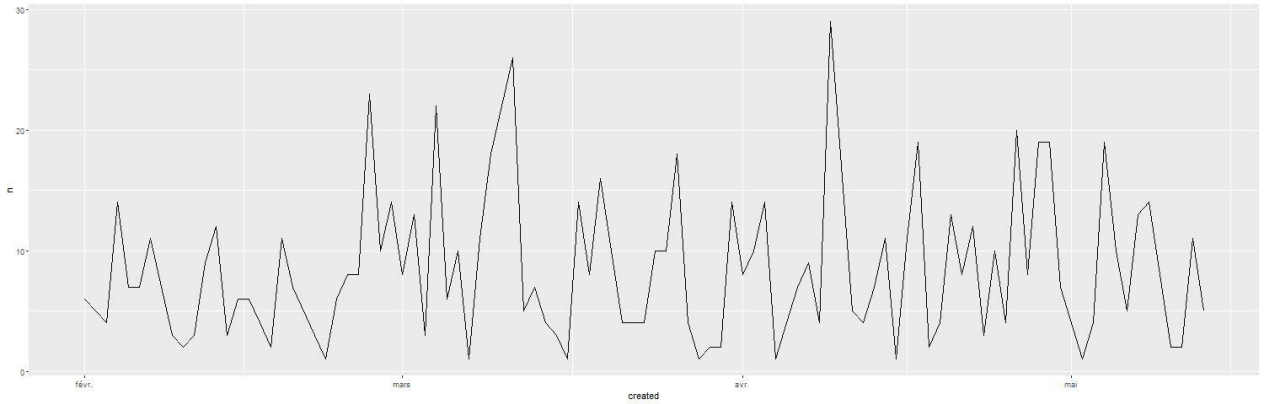
Wordcloud & Topics





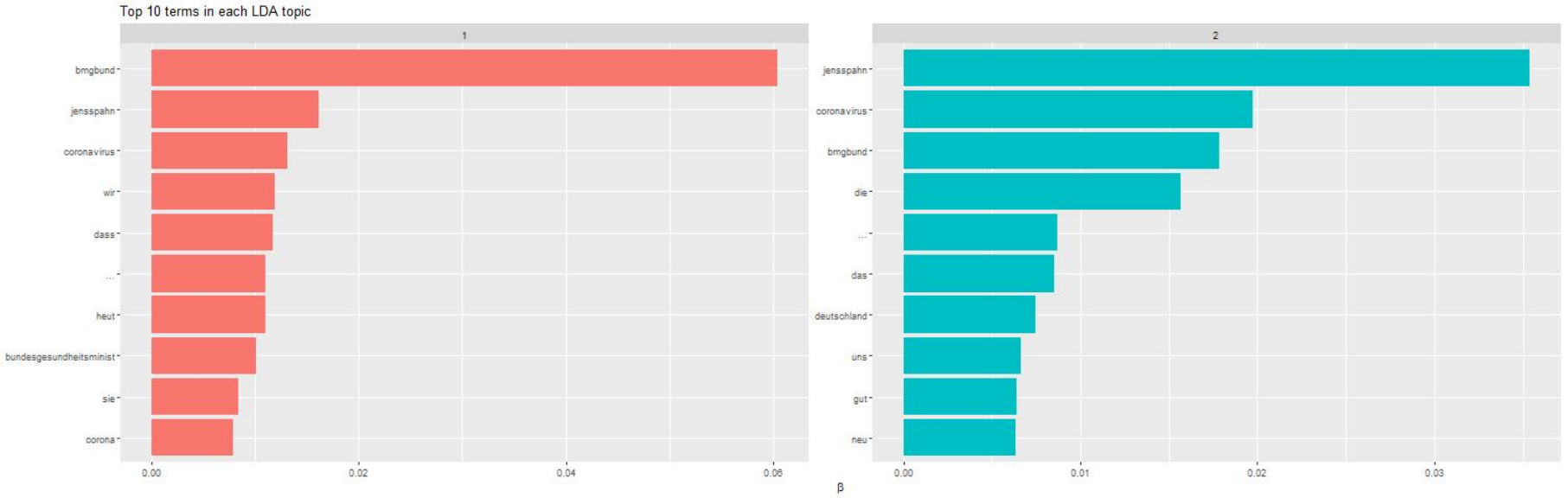
Jens Spahn

Wordcloud & Topics



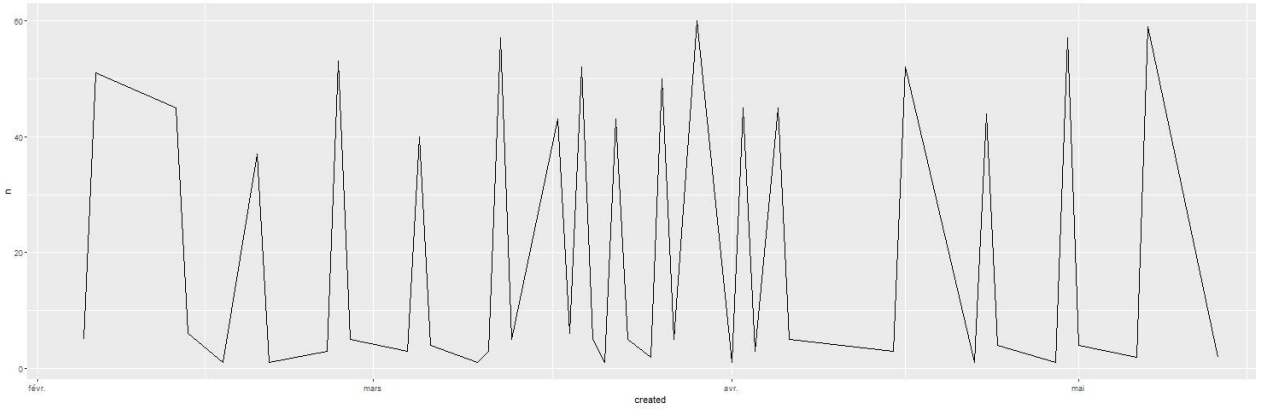
neg

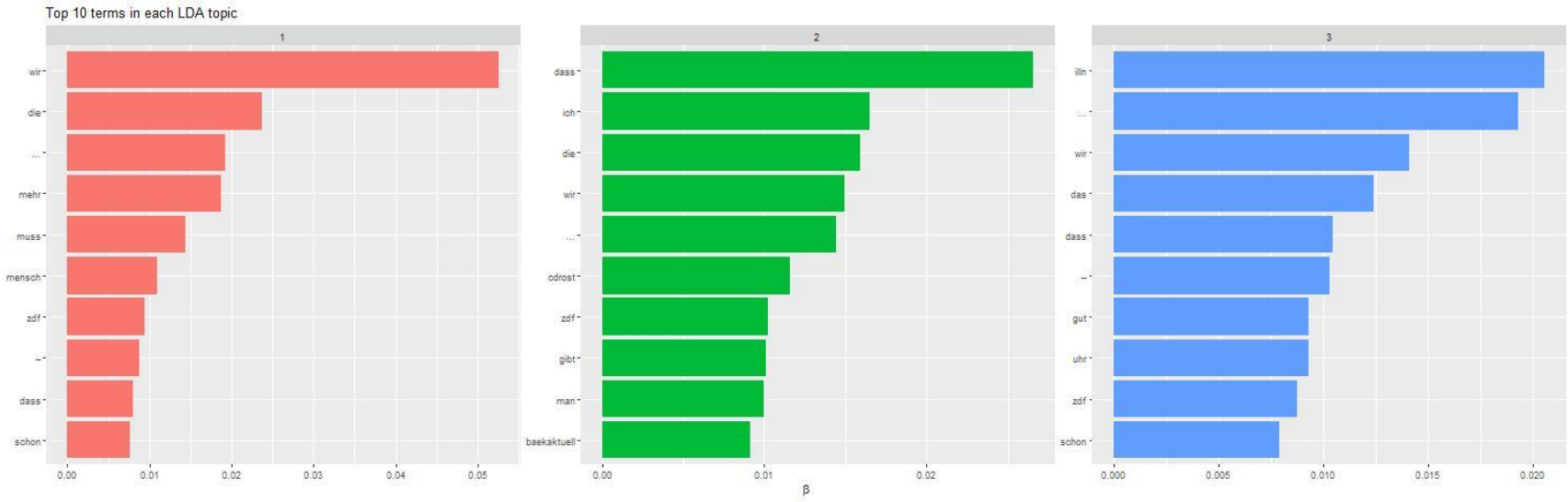
pos



Maybritt Illner

Wordcloud & Topics





Heute Journal

Wordcloud & Topics