COVID-19 South African and Global Media Outlet Twitter Analysis

Information and Knowledge in Organisations 771

Data Analysis Assignment 2 Report

Johanna Engelhard (22098038), Gina Lamprecht (21773513) *

12 July 2021

* https://github.com/ginalamp/ISM_Data_Analysis

Contents

1 Overview

Since the first detection of COVID-19 in 2019, the world has experienced a global pandemic that has impacted all lives. It has also impacted media and journalism – hardly a day goes by without another COVID-19 headline, whether it's about the pro's and con's of a home-office environment, development of a vaccine, or a new mutation that has sprung up and threatens to lock down the world once more. In an effort to analyse the spread of the Coronavirus in South Africa over the past six months, this paper will explore the responses by various media agencies in South Africa.

This analysis comes in the form of an analysis of social media posts on Twitter made by a selection of South African news agencies. In addition to this, in order to further understand the South African conversations surrounding the COVID-19 pandemic, this paper will also be evaluating the global response to the pandemic and compare that to the South African response. In the following paper, we describe data collection and wrangling, followed by an analysis and overall discussion of the findings. This includes a general analysis, the sentiment of what is reported, and the topics and sentiment of the topics that are being reported both in South Africa and globally.

1.1 Data collection

In order to get a better understanding of the topics and sentiment surrounding these topics over the past six months, we made the design decision to use tweet data between the beginning of January and the end of June in 2021. In order to go beyond the minimum functionality of the Twitter API's 3200 most recent tweets, we used *TwitterScraper* [11], a scraper optimised to get tweets in a given time frame we found on GitHub. This scraper did not retrieve all the tweets between January and the end of June, but we decided that the $\approx$ 6000 tweets per media outlet spread out over six months would give more insight than the 3200 most recent tweets.

After we scraped the Tweet IDs using the *TwitterScraper*, we then used the Twitter API through *rtweet* [3] to collect the tweet data on each of these tweets.

The chosen South African media agencies are the *Daily Maverick*, *SABC News*, *Eyewitness News*, and *News24*. The *New York Times*, *the Economist*, *BBC (World)*, and the *Associated Press* were identified as additional relevant sources of data that might provide further insight into the Twitter conversations in South Africa surrounding the COVID-19, by allowing a comparison

between the South African response to the global response. These specific media outlets were chosen for the reasons outlined below.

South Africa:

1. *Daily Maverick* - independent, unbiased, large digital publisher [12].
2. *SABC News* - one of the largest multimedia (written, radio, and TV) news outlets in South Africa [5].
3. *Eyewitness News* - large digital publisher [12].
4. *News24* - largest digital publisher in South Africa [12].

Global:

1. *New York Times* - top global news source with a huge readership, largest followship on Twitter, and one of the largest news outlets worldwide [9] [10].
2. *The Economist* - top global news source that focuses on politics and business, and blends left and right views [7].
3. *BBC World* - top global news source that is unbiased, very popular, and not geographically bound [7].
4. *Associated Press* - top global news source that is unbiased, very popular, and not geographically bound [7].

1.2 Data Wrangling

Once the data was collected, the next step was to clean the data. In order to do this we started with only selecting the columns of each `.csv` file that we would need for our analysis. Next, in order to extract tweets related to COVID-19, we used an **R** function (`grepl`) to get tweets containing the words *covid*, *corona*, and *pandemic*. After extracting Corona-related tweets, we cleaned the tweets through removing elements such as hyperlinks, english and custom stopwords, mentions, and non-alphabetical characters. Section-specific data wrangling steps are discussed at the beginning of each section.

All code that produced the results in this report can be found under `assignment2` in our GitHub repository:
`https://github.com/ginalamp/ISM_Data_Analysis`

2 Graph Discussion Outline

The analysis that we performed can be grouped into three main sections: numerical analysis, overall sentiment analysis, and topic modelling with topic-specific sentiment analysis.

In the sections that follow we will discuss each analysis using the following structure:

1. Briefly explaining the method of calculation
2. Describing the result and key takeaways
3. Presenting the result (table/figure)

3 Numerical analysis

In this section we performed two main forms of numerical analysis: general and tweet frequency. The general section looks into top-level information of media outlet follower, tweet, and retweet counts and compares the South African data with the global data. The tweet frequency analysis compares the tweet amount and frequency of overall tweets per media outlet, and tweets related to COVID-19 specifically per outlet.

3.1 General Numerical Analysis

The general numerical analysis results are computed by calculating the number of followers, tweets, and retweets of each media outlet using the `followers_count`, `statuses_count`, and `retweet_count` columns provided by the data extracted through *rtweet*.

As seen in the figures below, one can observe that both in South Africa and globally the amount of followers that a media agency has, has little influence on the amount of tweets that the outlet posts. All media agencies, in South Africa and globally, posted between $\approx 5 - 7.5k$ tweets in the last six months. Furthermore, the proportion of Corona-related tweets in comparison to the overall proportion of tweets varied between media agencies in South Africa, ranging between $2\% - 13\%$, with *Eyewitness News* having the least proportion of tweets related to the pandemic and *News24* having the largest proportion. Retweets have an even smaller proportion with a maximum of *Daily Maverick's* $\approx 3\%$. Globally, media outlets' posts related to the pandemic comprise between $12\% - 16\%$ of their overall tweets.

In South Africa, tweets about COVID-19 are more likely to be retweeted than the overall tweets, where for all media outlets the ratio between their overall tweets and retweets is less than the ratio between the Corona-related tweets and retweets. Conversely, the exact opposite is true globally: people are on average less likely to retweet Covid-related tweets than tweets overall. In addition to this, one sees that in South Africa the more followers an outlet has - the more likely their tweets are to be retweeted. This is not true globally as *The Associated Press*, the media outlet with the least amount of followers ($\approx$ 15M) and the second least amount of tweets ($\approx$ 5.1K), has the most amount of retweets ($\approx$ 1.33M) over the past six months overall.

Overall, this data shows that South African media outlets has a lot less activity than globally even though all the media agencies tweeted roughly the same amount over the past six months. The selected South African media agencies' amount of conversation relating to COVID-19 is a lot more varied than globally. Furthermore, one can deduce that the followers of the South African media outlets are most interested in content relating to COVID-19, as they are more likely to retweet posts relating to it. This is different when comparing it to the global conversation, where it seems that the media outlets post more about COVID-19, but the main point of interest of their followers are not necessarily there.

South African Numerical Analysis

Media outlet	Followers	Tweets	Retweets	Covid Tweets	Covid Retweets
Daily Maverick	481,169	5,530	18,042	555	2,049
Eyewitness News	1,296,367	7,119	59,027	202	1,056
News24	4,025,308	7,482	102,905	993	14,497
SABC News	1,968,720	7,108	79,148	811	6,489

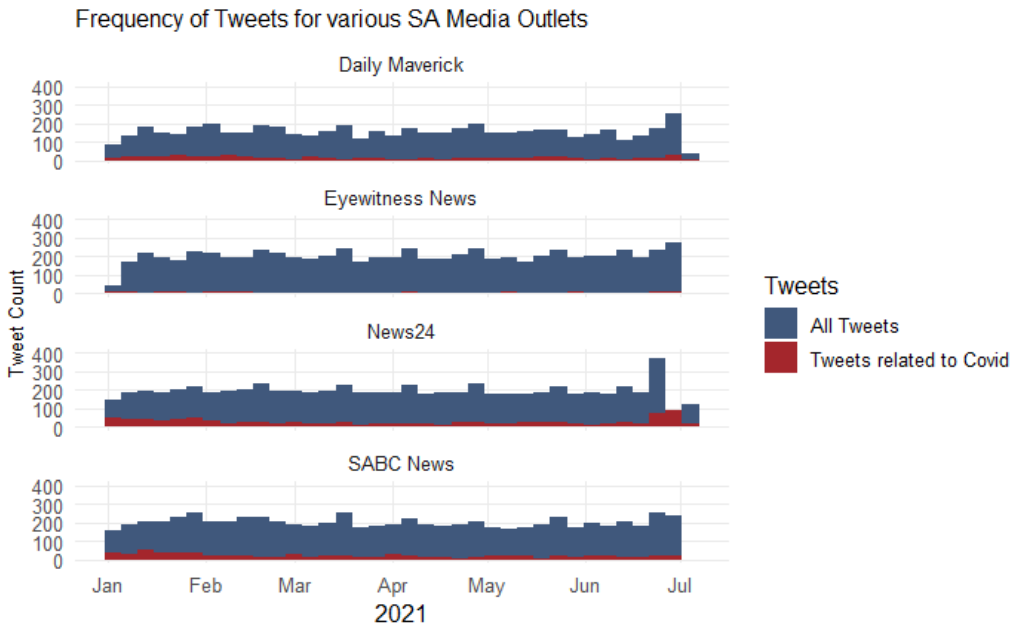
Global General Numerical Analysis

Media outlet	Followers	Tweets	Retweets	Covid Tweets	Covid Retweets
The New York Times	50,028,974	6,059	1,317,794	999	220,688
The Associated Press	15,189,986	5,194	1,337,521	838	125,588
BBC News (World)	32,225,177	4,901	824,250	607	91,536
The Economist	25,645,706	5,808	148,538	741	17,540

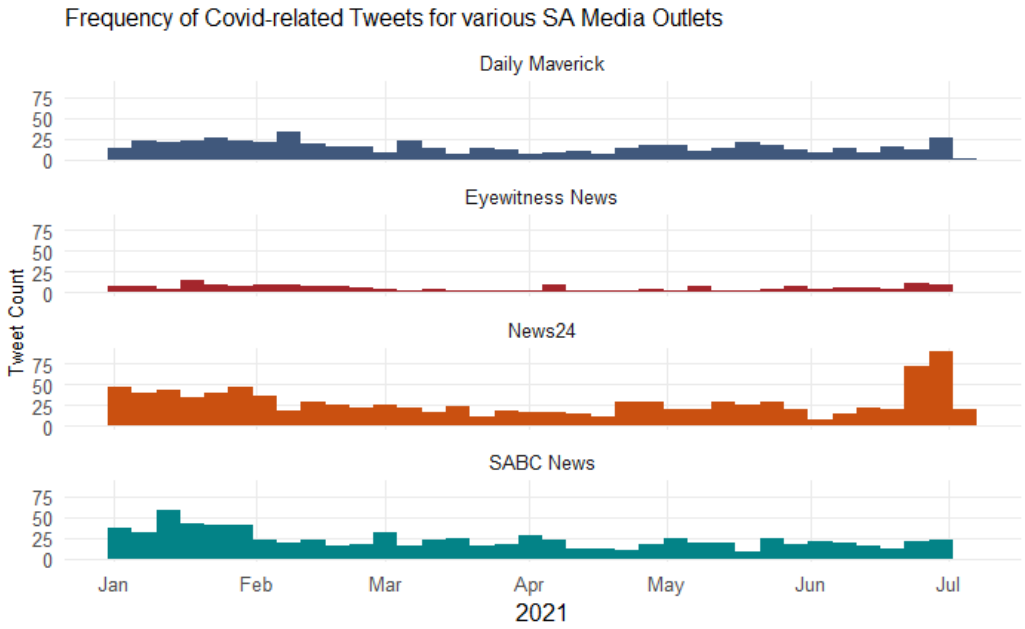
3.2 Tweet frequency

To visualise tweet frequency over time, tweets were grouped into 36 bins over the six months, or six bins per month. The resulting graphs show the number of tweets posted by each outlet for every five days over the six months.

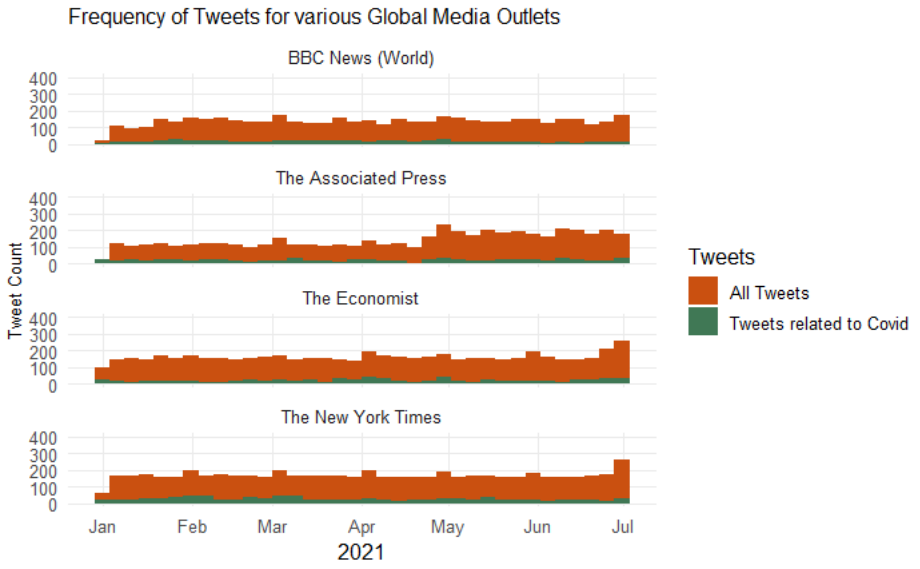
The below graph visualises the tweet frequency per chosen South African media outlet over the last six months. Although the tweet frequency fluctuates, all of the outlets tweet consistently over time at around 200 tweets per five days, with the *Daily Maverick* tweeting slightly less at around 150 tweets per five days. A general upward trend in tweet frequency appears to exist in June for the *Daily Maverick* and *Eyewitness News*.



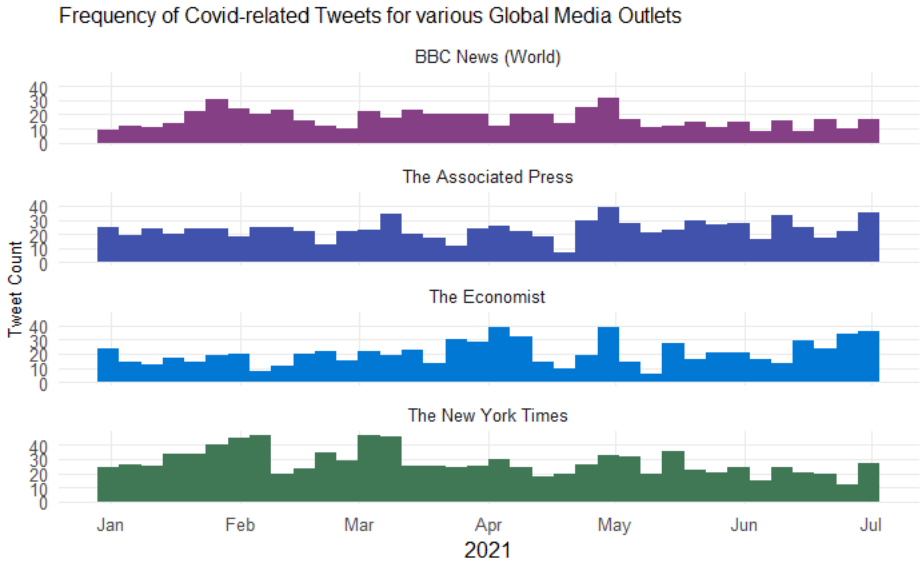
While the total tweet frequency fluctuates around a consistent average in South Africa, the number of Covid-related tweets per media agency visualised below decreased from January to April, after which it increased slightly. Following some dips, the number of tweets is again on an upward trend towards the end of June. *News24* and *SABC News* tweet most about COVID-19, and *Eyewitness News* tweets significantly less about COVID-19 compared to the other South African news outlets. This indicates that there is not necessarily a correlation between the number of overall tweets and the number of Covid-related tweets per organisation.



The below graph visualises the tweet frequency per chosen global news organisation over the last six months. Their tweet frequency fluctuates around a consistent average at around 150 tweets per five days, with the *New York Times* tweeting most consistently and the most frequently at just below 200 tweets per five days, with a spike in tweets at the beginning of each month. Toward the end of April, the *Associated Press* increases its average tweet frequency to just below 200 tweets per five days as well. The tweet frequencies of the other three news agencies show a general upward trend in June.



In contrast to the *overall* tweet frequency of global media organisations, the frequency of Covid-related tweets visualised below seems to increase and decrease in waves. The highest number of tweets per five days seem to mostly occur around the end and beginning of the months. Towards the end of June, the frequency of Covid-related tweets appears to be increasing again. There is no significant difference in the amount of Covid-related tweets posted by the various global new outlets over time.



From the above, it becomes apparent that the South African media agencies generally tweet more frequently than the global media agencies. The number of Covid-related tweets for South African and global organisations usually ranges anywhere between 10 to 50 tweets per five days. All of the organisations tweet relatively consistently, whereas the amount of Covid-related tweets posted fluctuates more significantly. It seems that, besides tweeting more frequently, South African news outlets follow a similar tweeting pattern to the global news outlets.

4 Sentiment analysis

This paper uses the `vader` package [8] to calculate the sentiment analysis score. The compound value is used to calculate the results, where we generally use the mean of the compound values in a given analysis. For interpretation, all values greater than zeroes will be considered a positive sentiment and all values less than zero would be considered as a negative sentiment [4]. In this section we discuss the total sentiment in South Africa and globally, the sentiment differences overall and for the top retweeted and favourited tweets per media agency, and how the sentiment per media outlet changes over the past six months.

4.1 Overall sentiment of all media outlets

The total overall sentiment on COVID-19 is determined by calculating the mean of all the media outlets' compound values of the extracted Covid-related tweets.

According to the *Vader* sentiment analysis as both the South African and global sentiment being negative, this means that on average both have a negative sentiment toward COVID-19. In addition to this, the global media outlets are in total almost twice as negative as the South African media outlets.

As such, we conclude that the selected global media outlets have been as a whole more negative about COVID-19 than the South African media outlets during the past six months.

$$\begin{aligned} South\ African: \ Total\ Overall\ Sentiment &\approx -0.1880734 \\ Global: \ Total\ Overall\ Sentiment &\approx -0.3565928 \end{aligned}$$

4.2 Overall sentiment per media outlet

The sentiment media outlet overall, for the top 10 retweeted tweets, and for the top 10 favourited tweets is determined by calculating the mean of each media outlet's compound value of the extracted Covid-related tweets. The top tweets are determined by filtering the compound values in descending order and extracting the top 10 tweets with the highest compound value.

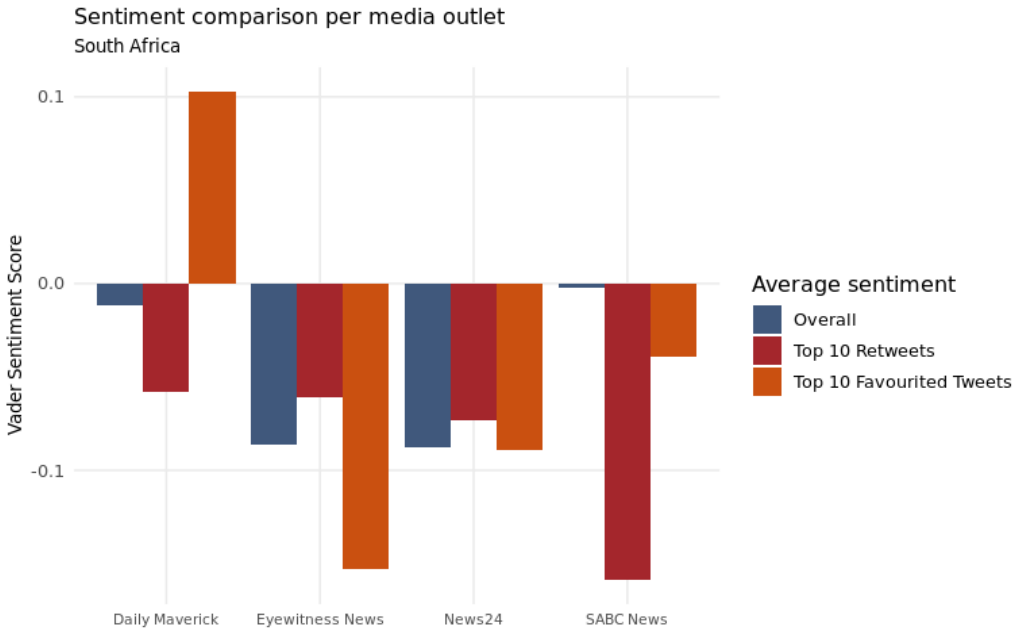
In the figures below, we see that overall both in South Africa and globally the sentiment on COVID-19 of all of the media outlets is negative. The South African media outlets are generally less negative than the Global outlets, with its most negative outlets (*Eye Witness News* and *News 24*) being

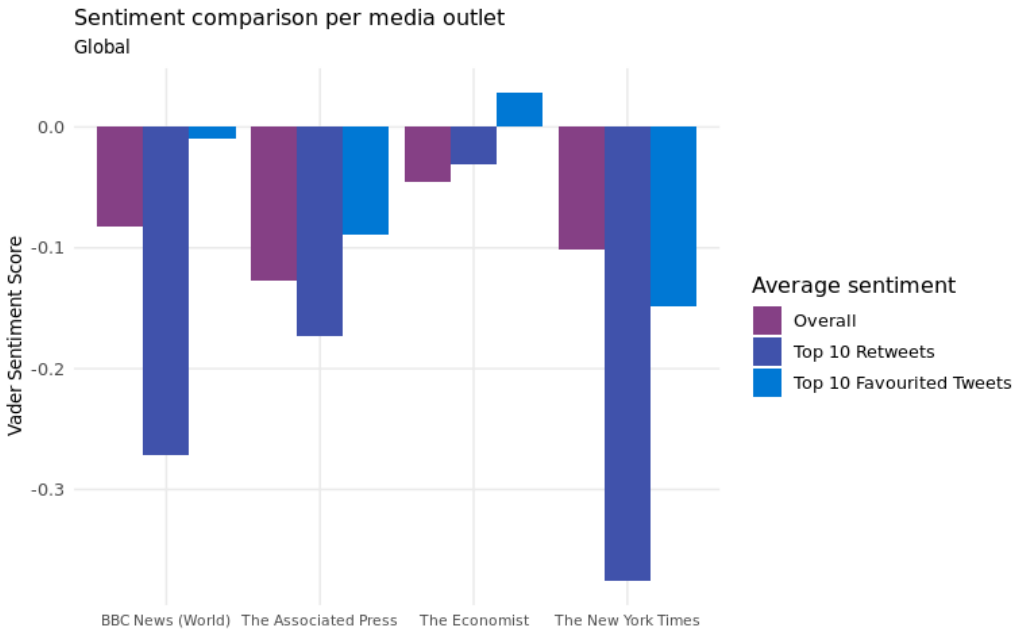
less negative than all but one of the global outlets (*The Economist*) overall. Furthermore, the *The Daily Maverick* and *SABC News* is at least seven times less negative than the other South African media outlets, which is a large discrepancy in comparison to the Global outlets where the sentiment is mostly in the same order (with the largest difference in sentiment being triple in the case of *The Economist* and *The Associated Press*).

In the South African media outlets, half of the retweeted tweets are more negative than the overall tweets per media outlet, whereas for the most favourited tweets only *The Daily Maverick* has a more positive sentiment than the overall tweets per media outlet. This seems like an outlier, as the sentiment of *The Daily Maverick's* top favourited tweets has a significantly more positive sentiment (≈ 0.1) than any other tweets graphed below in South Africa and globally.

A different trend can be seen globally where all but one media outlet's top retweet sentiment is notably more negative than the overall sentiment and only one media outlet's top favourited sentiment is more negative than the overall sentiment.

As a whole this is consistent with the total overall sentiment analysis with global media outlets being more negative than South African media outlets. Furthermore, one can conclude that the South African media outlets are more polarised on their sentiment toward COVID-19 than the Global media outlets are.





4.3 Sentiment per media outlet over time

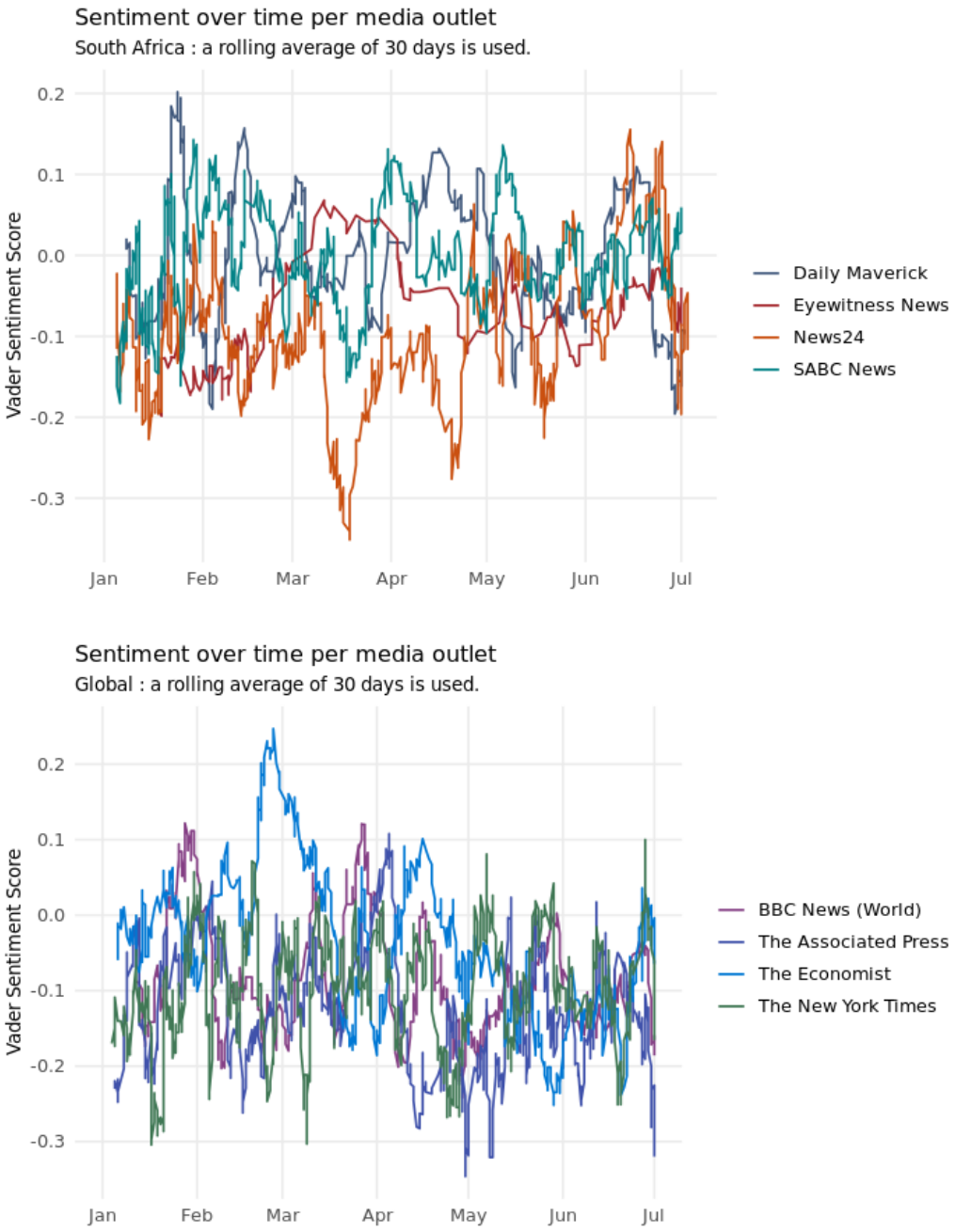
The rolling average is calculated over 30 days using the compound value of the extracted Covid-related tweets. In **R**, this entails applying the `roll_mean()` method from the `RcppRoll` package to the *Vader* sentiment calculated on each media outlet's tweets.

In the figures below, we see that the sentiment in both the South African and global media outlets are quite volatile between the months of January and June. The South African positive sentiment has a upper bound of ≈ 0.2. On the other hand, the negative sentiment generally has a lower bound of ≈ -0.2, with *News 24* significantly deviating from this in the end of March, and slightly deviating in the middle of April and beginning of May. As a whole, the South African sentiment stays between these bounds with a slight positive trend in some media outlets over the six months.

Globally, the upper bound of the positive sentiment is generally closer to ≈ 0.1 (with only one significant fluctuation above this in March by *The Economist*. On the other hand, the lower bound is generally closer to -0.3. As a whole, the global sentiment has an increase in negative sentiment after April.

This is still in line with the overall total and per media outlet sentiment, as global media outlets seem to be more negative than South African media outlets about COVID-19. In this analysis, it seems like the global media outlets has a trend of becoming more negative in the latter half of this anal-

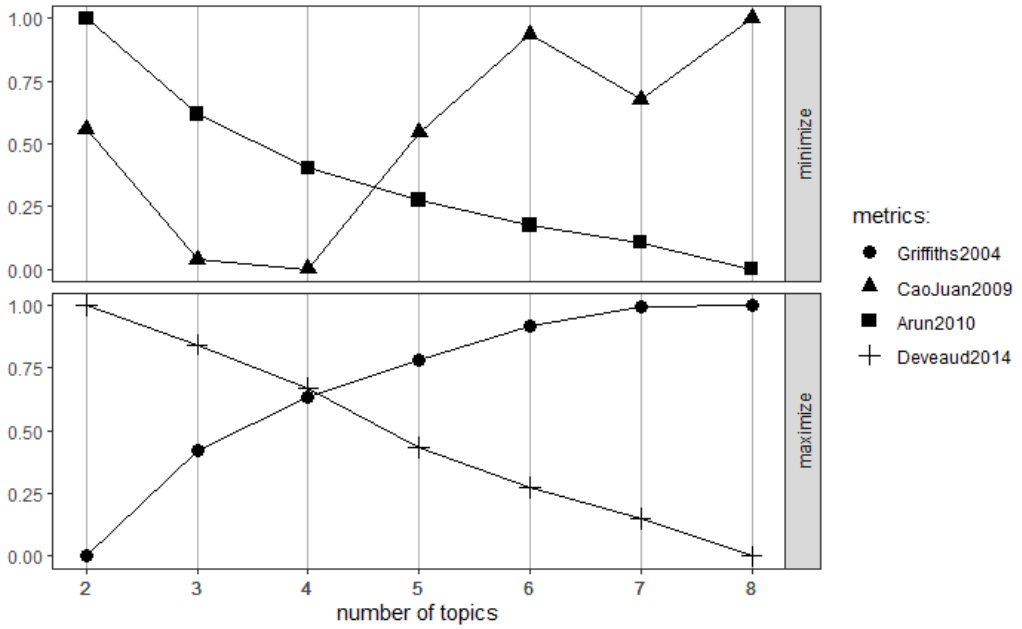
ysis, with the South African media outlets being more stable and with some outlets trending toward a less negative sentiment.



5 Topic modelling and sentiment per topic

The `topicmodels` package [1] is used to perform LDA topic modeling in this project and the number of topics to model is chosen based on the result of *LDA tuning* performed with the `ldatuning` package [6]. During topic modelling, a *beta* value is allocated to each term, describing the likelihood of that term appearing in a given topic. The terms with the highest *beta* values are therefore the most relevant terms in that topic. In order to prevent the topics from being overpowered by words that are frequently used and applicable to all topics contained in covid-related tweets, appropriate data cleaning was undertaken. Besides standard English stopwords from the `tidytext` package [2], words including *COVID*, *COVID19*, *corona*, *COVID-19*, *pandemic*, *sabcnews*, *maverick*, *south*, and *africa*, were removed from the set of terms. The top ten most relevant terms are displayed to describe the respective topics.

Using `ldatuning` to determine topic amounts:
The number of topics in each of the following topic modelling graphs is based on *LDA tuning*. The following example illustrates the process of choosing the best number of topics.



When choosing the number of topics, the aim is to minimise the two functions in the top graph, whilst maximising those in the bottom graph at the same time. In the bottom graph, one function has a positive gradient, while the other has a negative gradient. This means that when trying to maximise

both of these functions we will have to look at where they intersect. In this instance, the functions intersect where there are four topics. A similar logic applies to the top graph, where the intersection of the functions is between four and five topics. However, as one of the functions has an extremum at four the best number of topics to model in this example is therefore four.

We used this approach in our analysis when determining the amount of topics for each of the topic modelling graphs that follow. In our code we created a global variable `TUNING` (defaulted to `FALSE`) which can be set to `TRUE` if the reader wishes to run the code with the *LDA tuning* analysis.

Sentiment analysis on identified topics:
After analysing the topics for the tweets, we analyse the sentiments on the given topics. After the graphs are shown for the topics, we will have a look at the average sentiment both in South Africa and globally for the: overall topics, topics per month, and topics over time (January versus June). This paper will not be analysing the sentiment for the topics per media outlet over time.

In this analysis, the same amount of topics as in the topic modelling are used (which is determined through the initial *LDA tuning*). Furthermore, it is important to note that this analysis is done on individual terms associated with the specific topics, as unnesting the tweets into individual terms is necessary in order to apply the *LDA* topic modelling. As such, the *Vader* sentiment analysis applied on the topics could be different than it could have been if the sentiment analysis were to be applied to entire tweets associated to the topics.

5.1 Overall topics of all media outlets

The topics discussed by the South African and global news outlets over the last six months is determined by performing topic modelling over all their tweets on Corona-related themes over that time frame.

In the figures below, we see that it seems that vaccines have been on everyone's mind both in South Africa, as well as globally. Besides that, the reporting by the South African and global media outlets over the last six months each broadly surrounds five topics, which could be summarised as follows:

South African Overall Topics

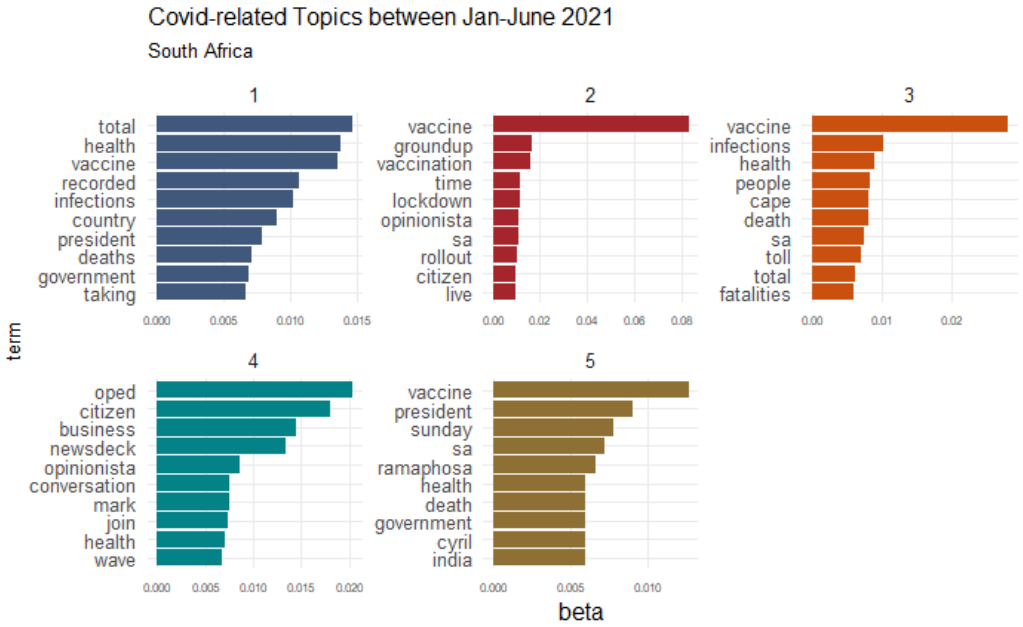
1. Total recorded infections
2. Lockdown and time surrounding vaccinations

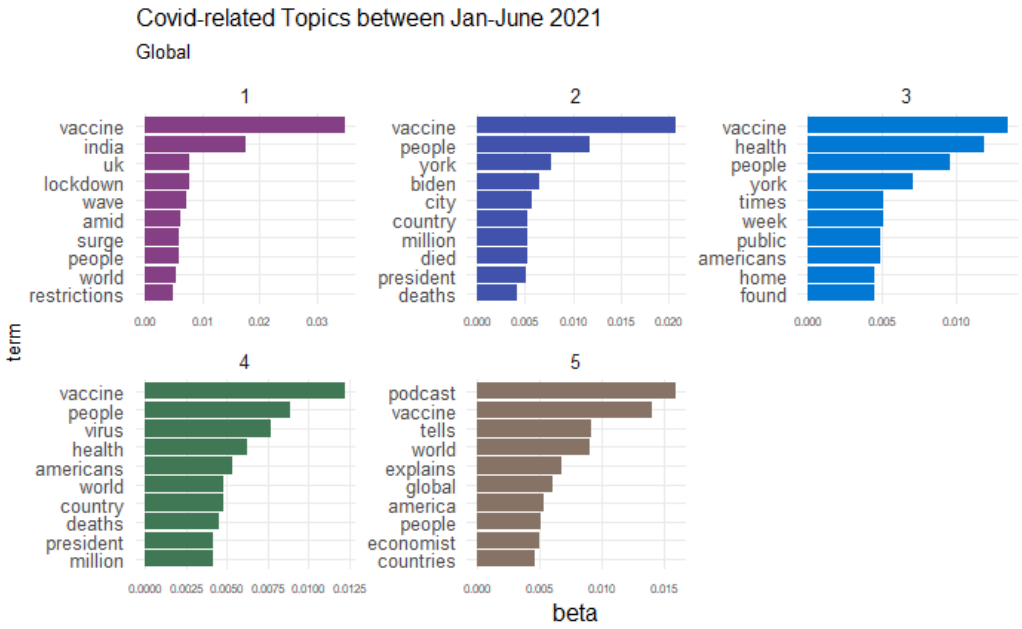
3. Death toll and health facilities
4. Articles expressing personal viewpoints
5. Response by President Cyril Ramaphosa/the South African government

Global Overall Topics

1. Indian variant, consequential 3rd wave, lockdowns, and restrictions
2. Death toll in New York and President Biden's response
3. Public health
4. America, its people and its president
5. Podcast and information explanations

Although the overarching topics do not correspond entirely, certain trends can be found in both South African and global reporting. For example, one might conclude that how presidents are responding, lockdowns, vaccinations and the general impact of COVID-19 on people's lives, are important all over the world.

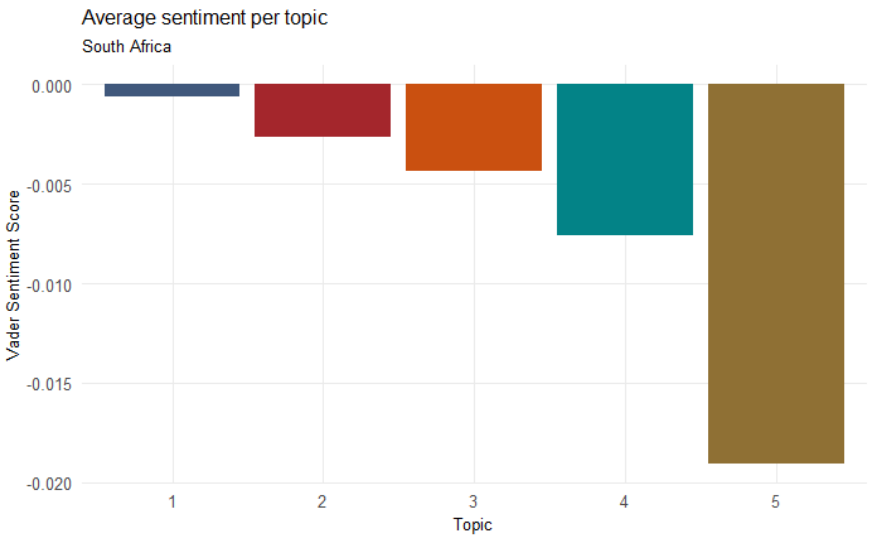




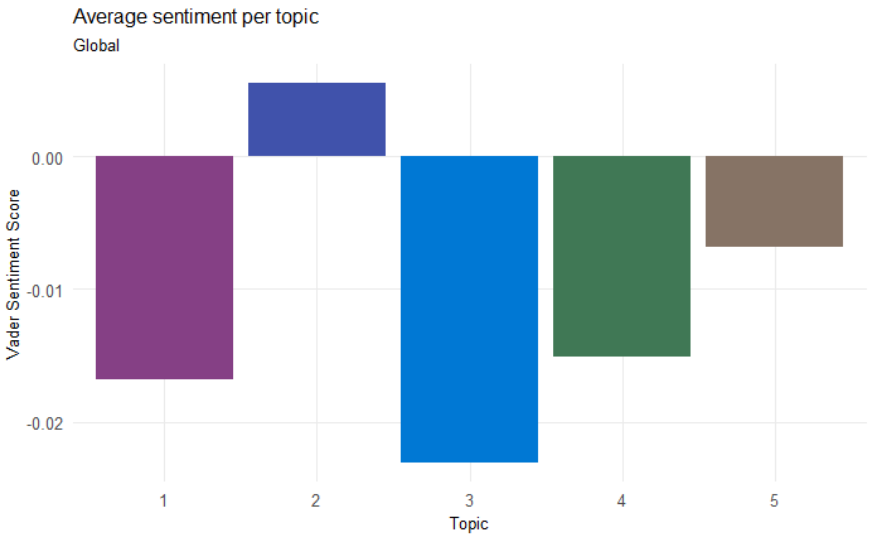
5.1.1 Average sentiment per topic

The average sentiment per topic is calculated through applying *Vader* sentiment analysis upon each of the topics.

In the graph below, we see that in South Africa all these topics have a negative sentiment where sentiment decreases from topic 1 (the total recorded infections) to topic 5 (the response by President Cyril Ramaphosa and the South African government). The difference between the sentiment is significant, with the sentiment decreasing gradually from topic 1 to 4, and there being a significant negative sentiment for the 5^{th} topic. The response by the South African president and government has more than a ten-factor difference than the first topic and more than a two-factor difference than the next most negative topic about articles expressing personal viewpoints.



Both in South Africa and globally the sentiment about topics surrounding their leaders in government is significantly negative. In South Africa this is the most negative sentiment by a significant amount, whereas globally there seems to be a more negative sentiment toward public health and the Indian variant, consequential third wave, lockdowns, and restrictions. Furthermore, in America there seems to be a positive sentiment toward President Biden's response in relation to the New York death toll. Thus, it could be that the significant negative response in South Africa toward government and presidential response is due to another reason than just it being associated with COVID-19. For example, it could be that there is generally a more negative sentiment toward government in South Africa than there is globally.

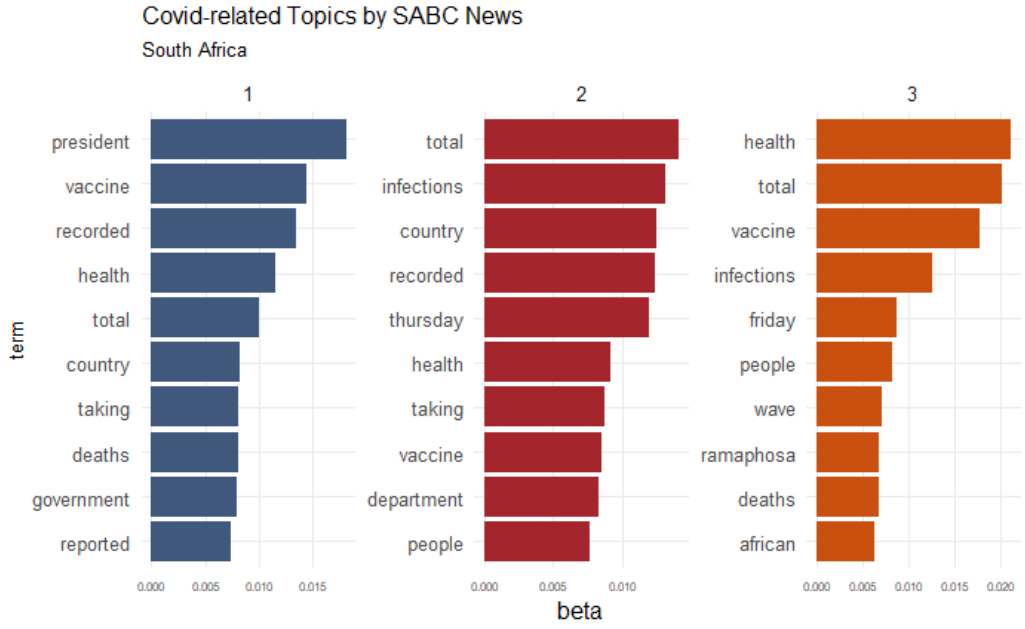


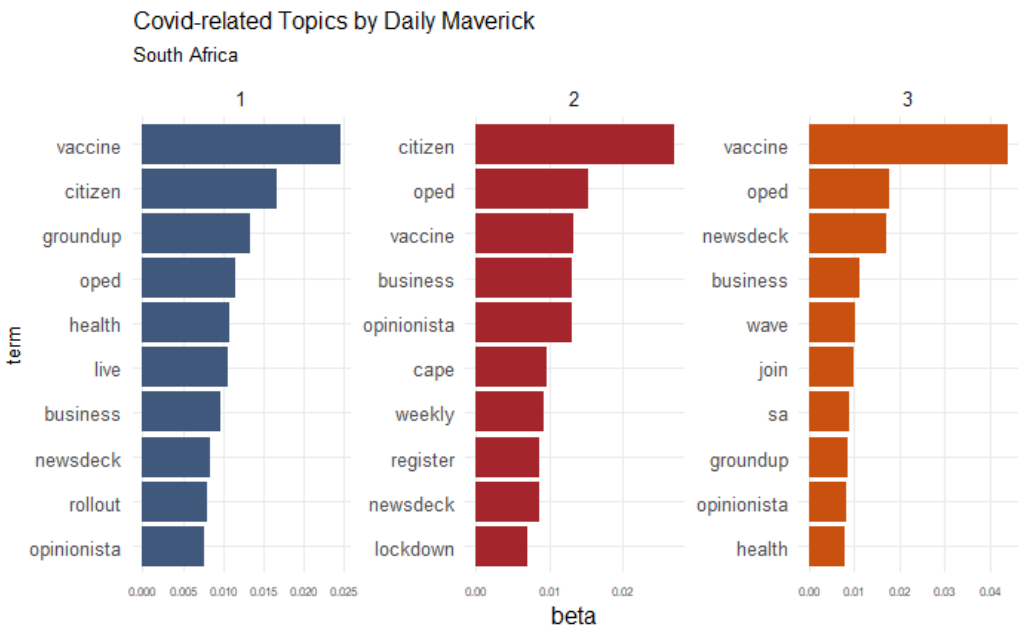
5.2 Overall topics per media outlet

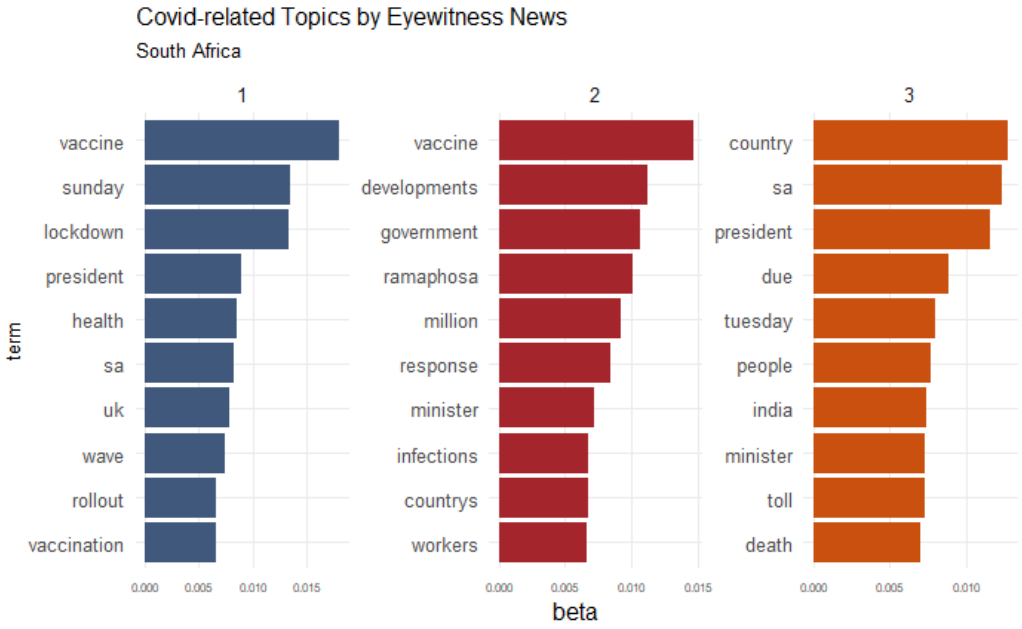
To have a closer look at how the individual media outlets contribute to the overall topics, their respective topics over the last six months extracted from the Corona-related tweets are considered.

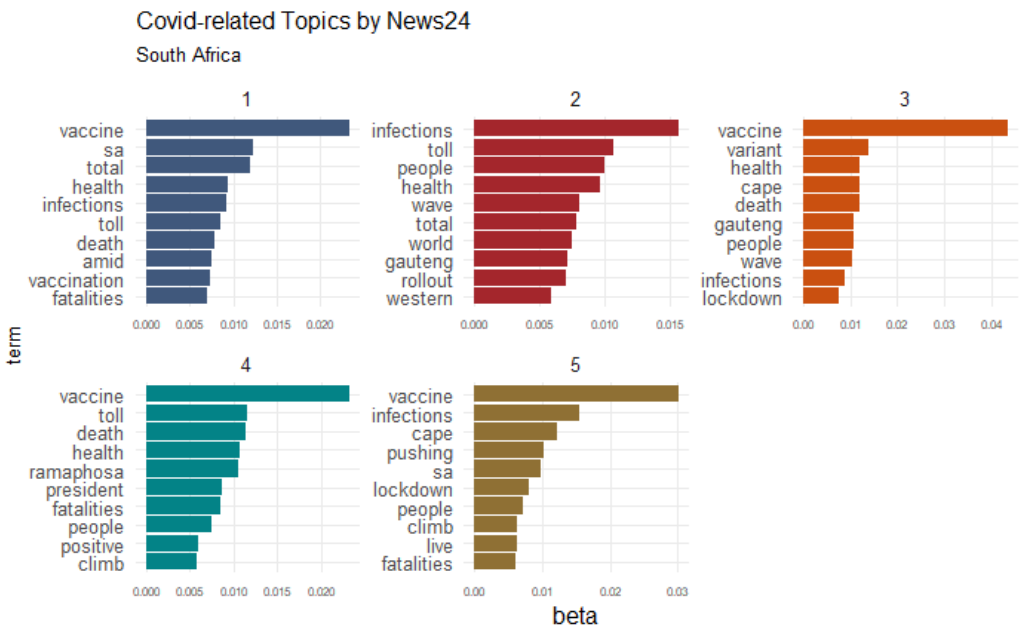
In the figures below, we see that from a South African point it seems that most of the overall topics discussed above are reported on by all of the various media outlets.

In terms of outlet-specific topics, *SABC News's* topics tend to revolve around the overall topics of vaccination, infection rates, health and death tolls, governmental and President Ramaphosa's response, and waves of infection. Furthermore, from the fact that in each of *The Daily Maverick's* topics the terms "oped" and "opinionista" is present, we can deduce that this media outlet could have more subjective reporting. In addition to this, *The Daily Maverick's* topics all also contain the "business" term, which could indicate that their reporting is more focused on business reporting, or the effects of COVID-19 on businesses. *Eyewitness News* focuses on similar topics to *SABC News*, but it seems that they have a more international focus with the inclusion of different countries in their topics. Interestingly, the articles expressing personal viewpoints (indicated by *opinionista* and *oped*, which has been reduced from *OP-ED*) originate from the *Daily Maverick*, and *News24* appears to be the only news outlet with topics relating to Gauteng, which is the current COVID-19 hotspot in South Africa.

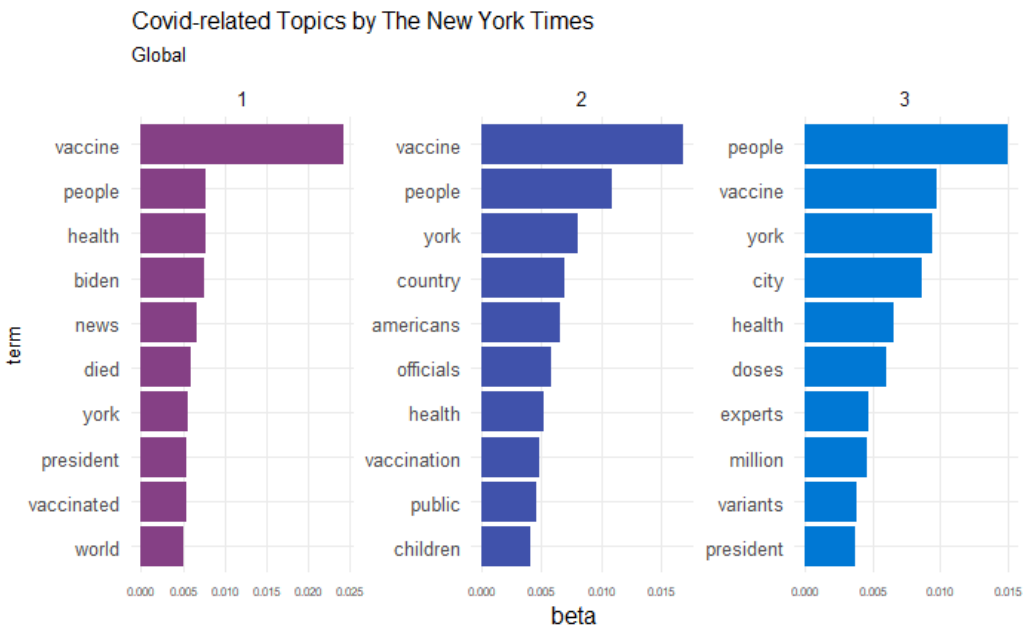


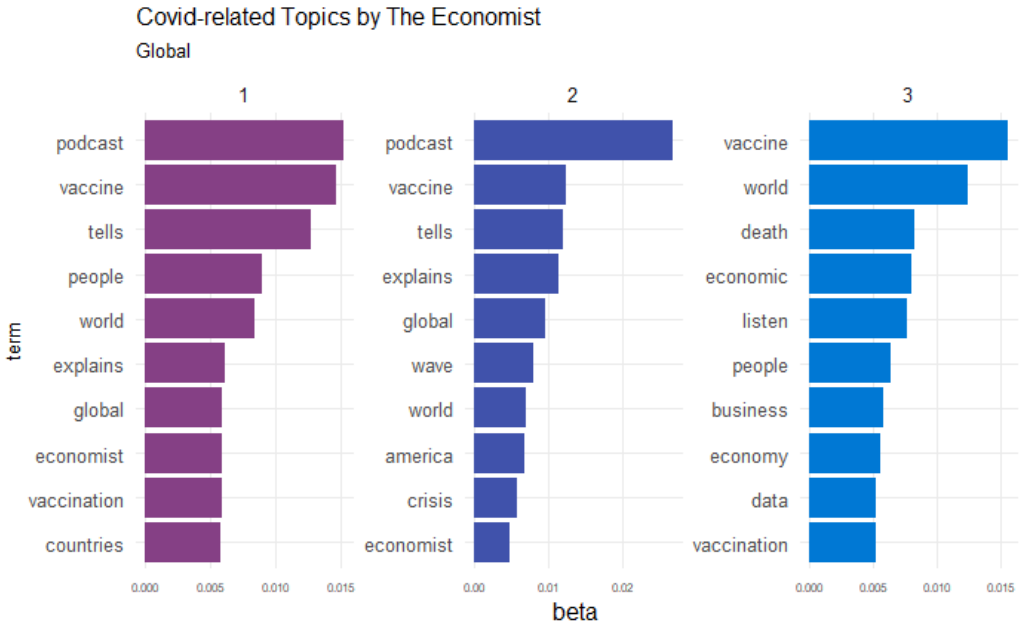

Covid-related Topics by Daily Maverick
South Africa
1
2
3
vaccine
citizen
groundup
oped
health
live
business
newsdeck
rollout
opinionista
citizen
oped
vaccine
business
opinionista
cape
weekly
register
newsdeck
lockdown
vaccine
oped
newsdeck
business
wave
join
sa
groundup
opinionista
health
term
0.000 0.005 0.010 0.015 0.020 0.025
0.00 0.01 0.02
0.00 0.01 0.02 0.03 0.04
beta

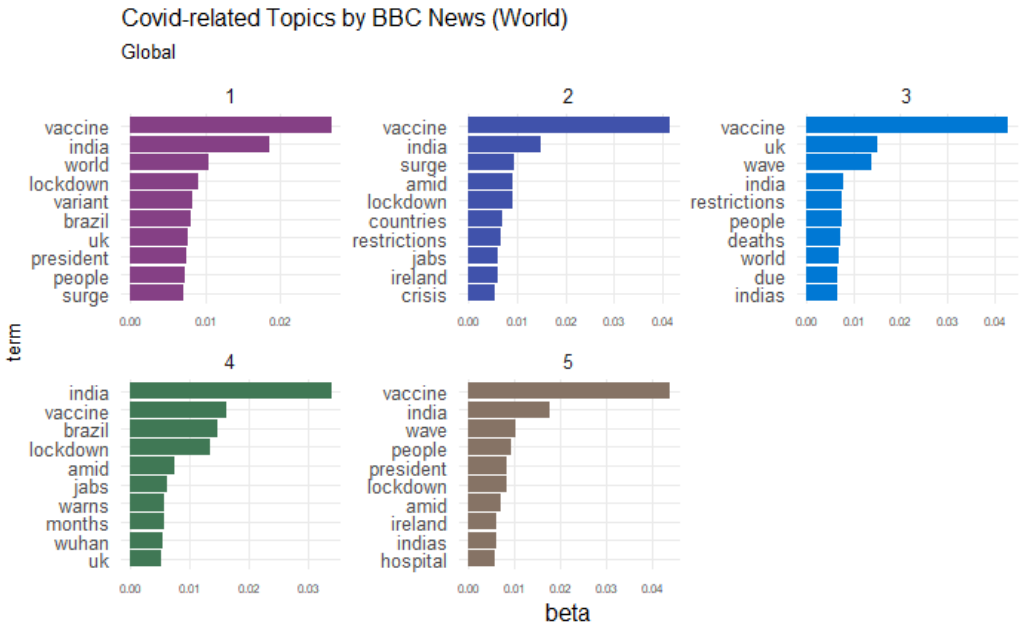

Covid-related Topics by Eyewitness News
South Africa
1
2
3
vaccine
sunday
lockdown
president
health
sa
uk
wave
rollout
vaccination
vaccine
developments
government
ramaphosa
million
response
minister
infections
countrys
workers
country
sa
president
due
tuesday
people
india
minister
toll
death
term
0.000 0.005 0.010 0.015
0.000 0.005 0.010 0.015
0.000 0.005 0.010
beta

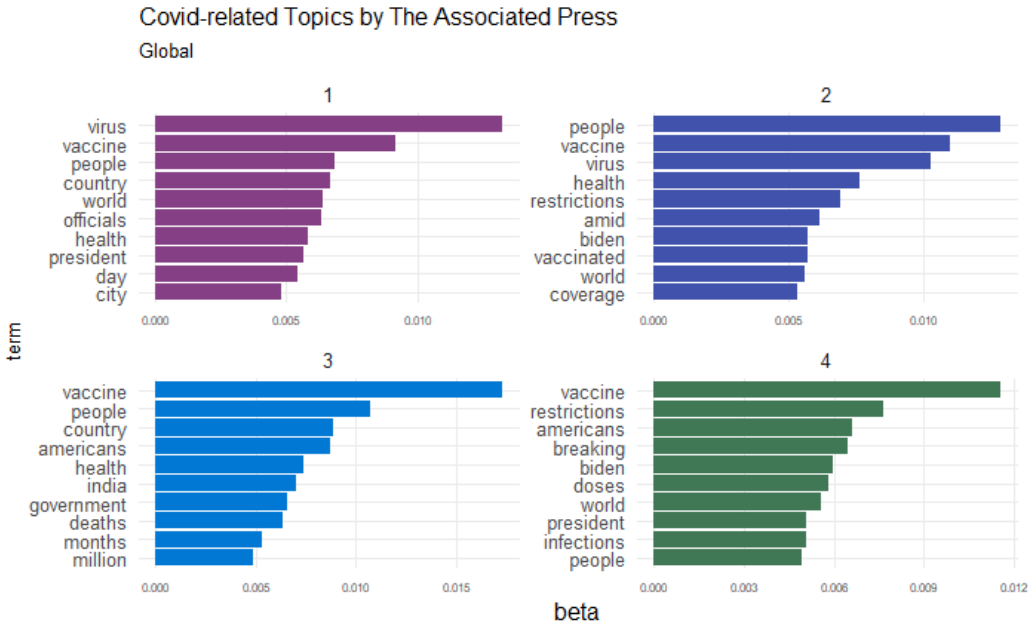


In the figures below, we see that from a global point it seems that although most overall topics are reported on by most of the selected global news outlets, certain topics receive more coverage by certain media outlets. For instance, the *New York Times* and the *Associated Press* both focus on America and use keywords such as "vaccine" and "people" together. The *Economist* includes economy and business aspects, and runs a podcast that explains and discusses topics surrounding COVID-19. The Indian variant and consequential surge of infections and following restrictions and lockdowns is a topic predominantly covered by *BBC (World)*.


Covid-related Topics by The New York Times
Global
1
2
3
vaccine
people
health
biden
news
died
york
president
vaccinated
world
vaccine
people
york
country
americans
officials
health
vaccination
public
children
people
vaccine
york
city
health
doses
experts
million
variants
president
term
beta


Covid-related Topics by The Economist
Global
1
2
3
podcast
vaccine
tells
people
world
explains
global
economist
vaccination
countries
podcast
vaccine
tells
explains
global
wave
world
america
crisis
economist
vaccine
world
death
economic
listen
people
business
economy
data
vaccination
term
beta





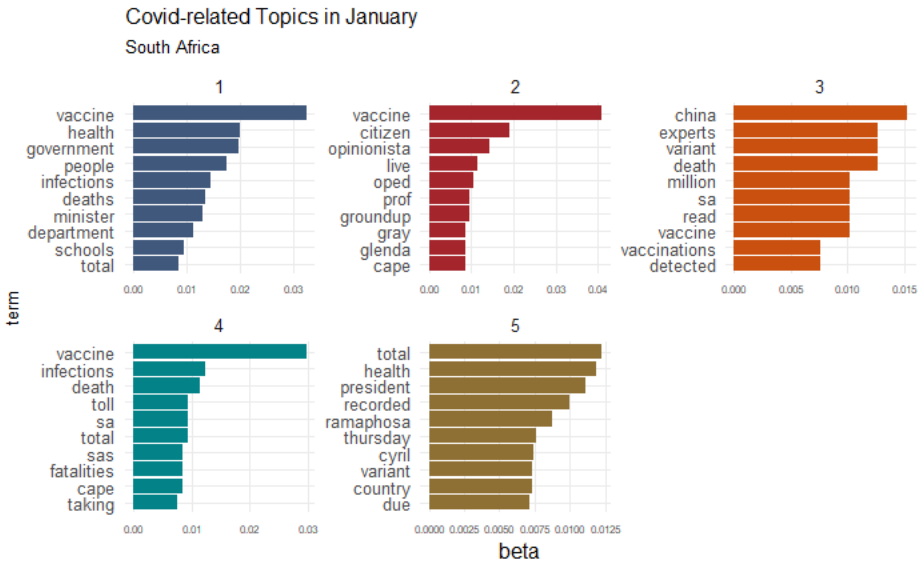
Looking at the overall topics of the last six months relating to COVID-19 covered by the selected South African and global media agencies gives an impression of the general conversations and topics of concern regarding the virus. The topics covered by the individual media outlets show where the focus of each media outlet lies.

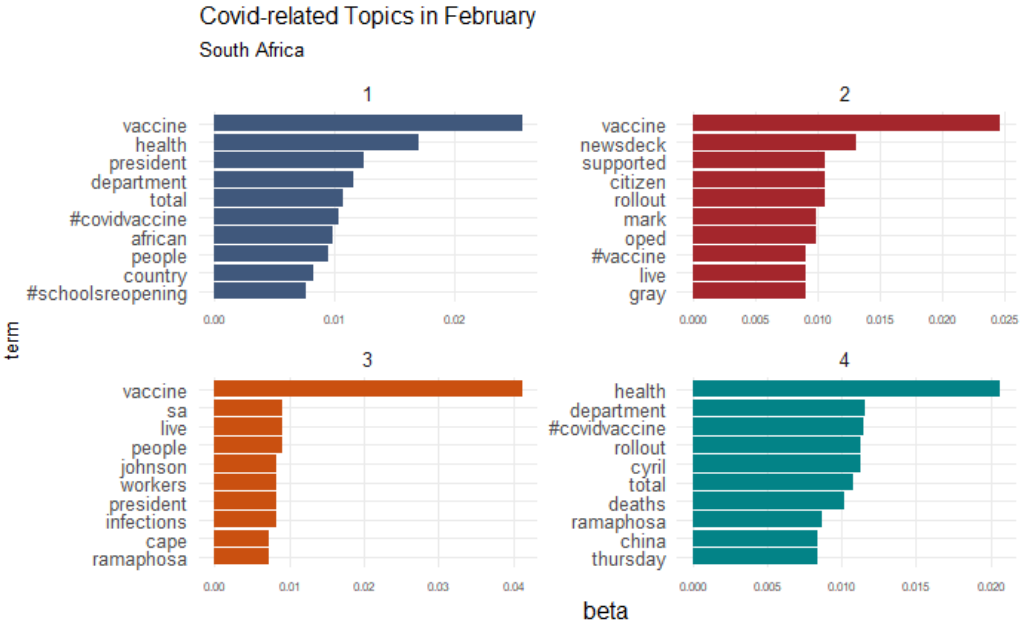
5.3 Topics of all media outlets over time

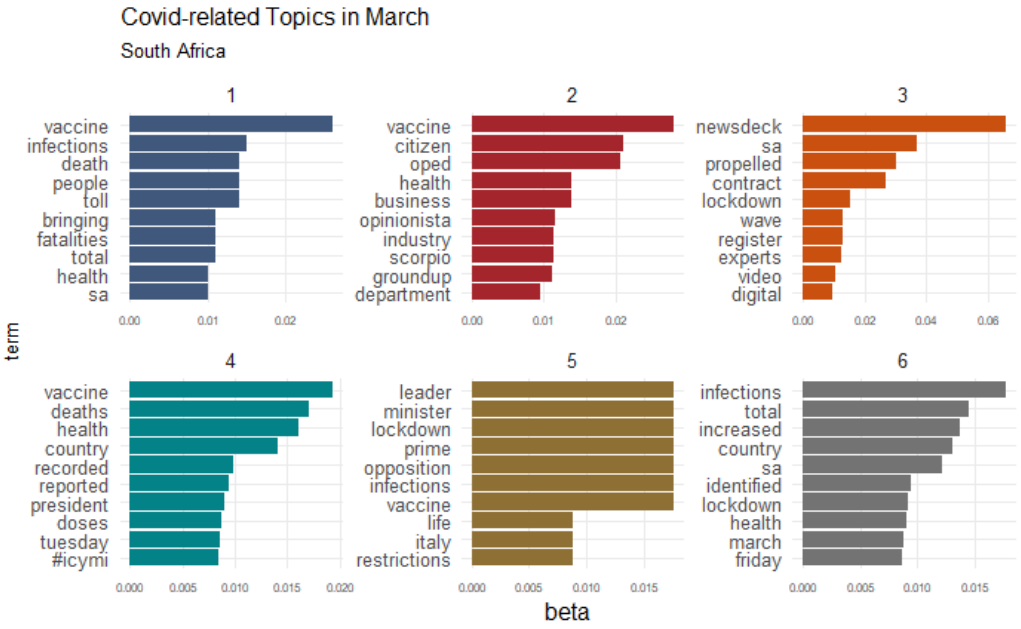
In order to gain further insight into how the spread of COVID-19 in South Africa progressed over the last six months, we determined the topics reported on by the South African and global media outlets for each month for the Corona-related tweets.

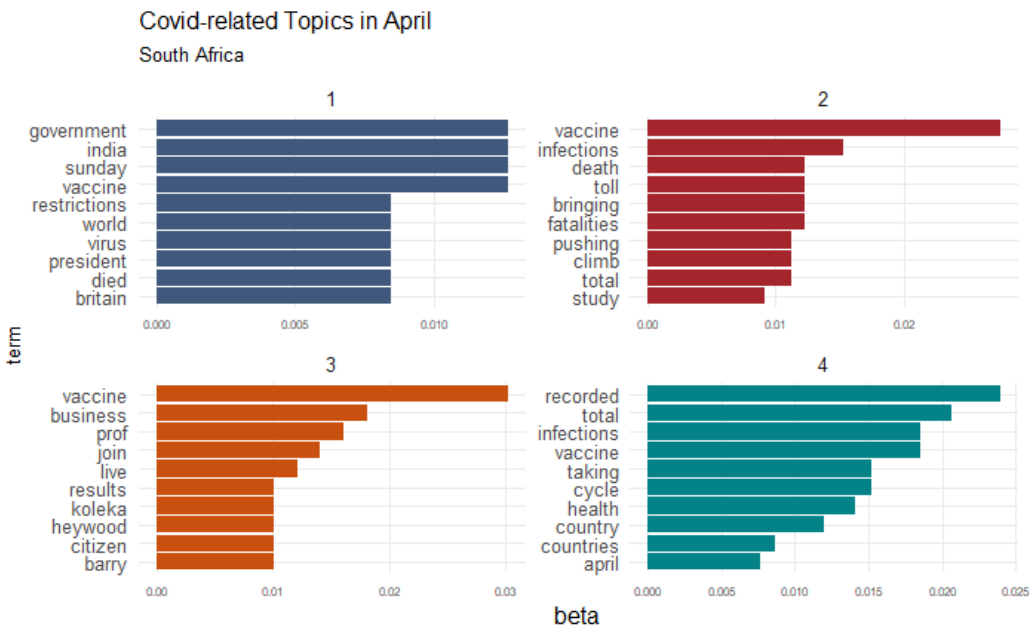
5.3.1 South African topics over time

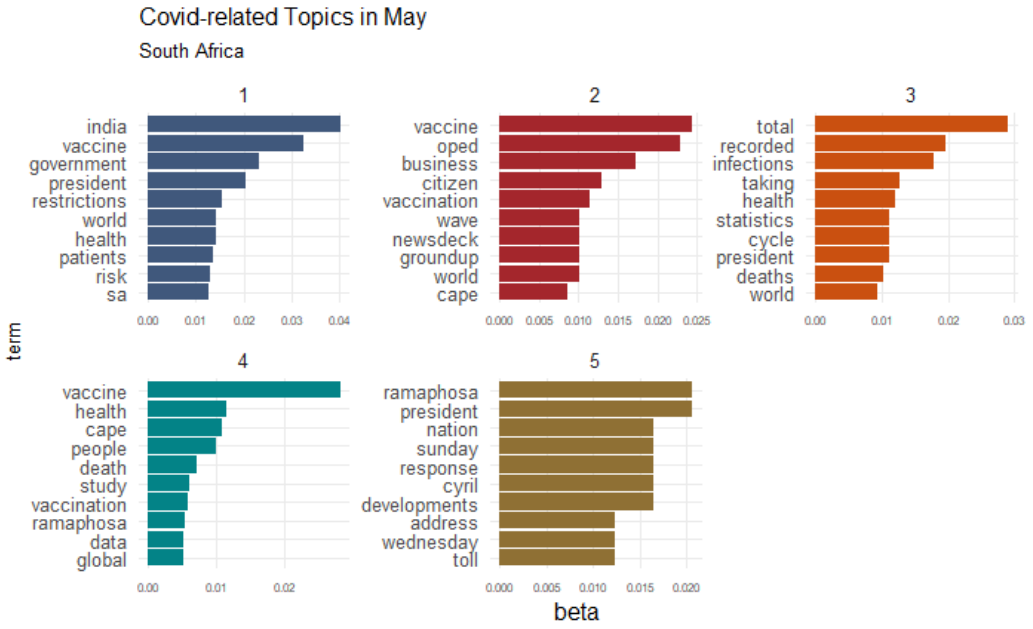
In the figures below, we see that from a South African point it seems that total recorded infections and deaths, vaccination, President Cyril Ramaphosa and other leaders and ministers of the SA government, and personal viewpoints or opinions are recurring topics in South African news throughout the first half of 2021. Additional noteworthy topics include the conversation in January and February surrounding the reopening of schools for the academic year in 2021, with the Minister of Health and Prof Glenda Gray at the centre of the discussion. In February, the first phase of the vaccination roll-out was started in South Africa using the Johnson & Johnson vaccine. Increasing infections in March indicating another wave of COVID-19 and mentions of India in April are followed by this being a central topic in May. In May these topics are central along with the presidential address where President Cyril Ramaphosa addressed the nation on Sunday 30 May 2021 in response to the developments surrounding the COVID-19. In June, infections rise as the Indian variant spreads, especially in Gauteng. Towards the end of the month, SA moves into an adjusted level 4 lockdown.

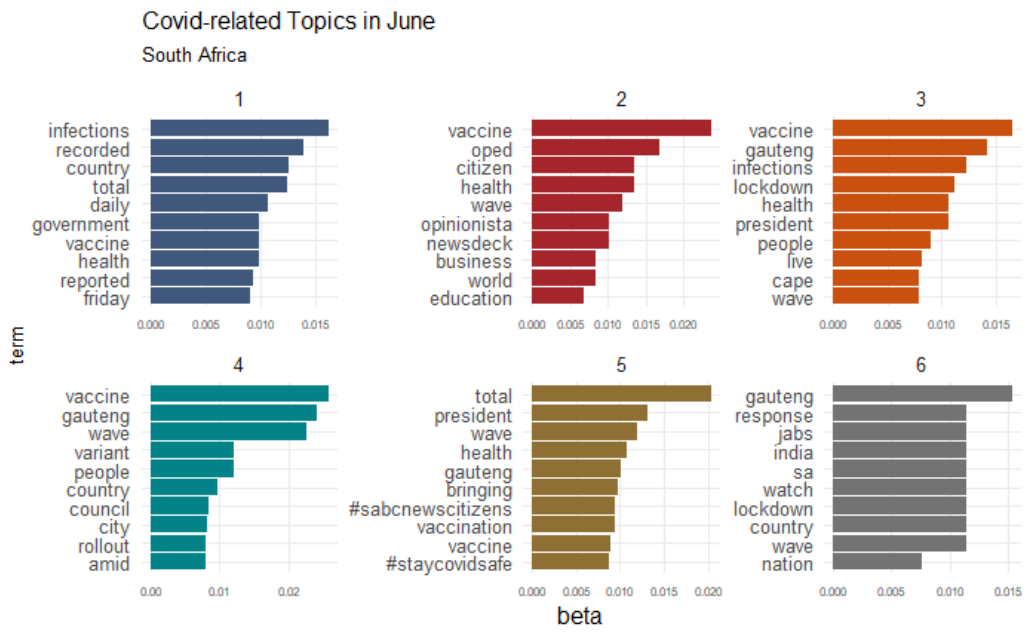



Covid-related Topics in February
South Africa
1
vaccine
health
president
department
total
#covidvaccine
african
people
country
#schoolsreopening
2
vaccine
newsdeck
supported
citizen
rollout
mark
oped
#vaccine
live
gray
3
vaccine
sa
live
people
johnson
workers
president
infections
cape
ramaphosa
4
health
department
#covidvaccine
rollout
cyril
total
deaths
ramaphosa
china
thursday
term
beta


Covid-related Topics in March
South Africa
1
vaccine
infections
death
people
toll
bringing
fatalities
total
health
sa
2
vaccine
citizen
oped
health
business
opinionista
industry
scorpio
groundup
department
3
newsdeck
sa
propelled
contract
lockdown
wave
register
experts
video
digital
4
vaccine
deaths
health
country
recorded
reported
president
doses
tuesday
#icymi
5
leader
minister
lockdown
prime
opposition
infections
vaccine
life
italy
restrictions
6
infections
total
increased
country
sa
identified
lockdown
health
march
friday
term
beta


Covid-related Topics in April
South Africa
1
government
india
sunday
vaccine
restrictions
world
virus
president
died
britain
0.000
0.005
0.010
2
vaccine
infections
death
toll
bringing
fatalities
pushing
climb
total
study
0.00
0.01
0.02
3
vaccine
business
prof
join
live
results
koleka
heywood
citizen
barry
0.00
0.01
0.02
0.03
4
recorded
total
infections
vaccine
taking
cycle
health
country
countries
april
0.000
0.005
0.010
0.015
0.020
0.025
term
beta

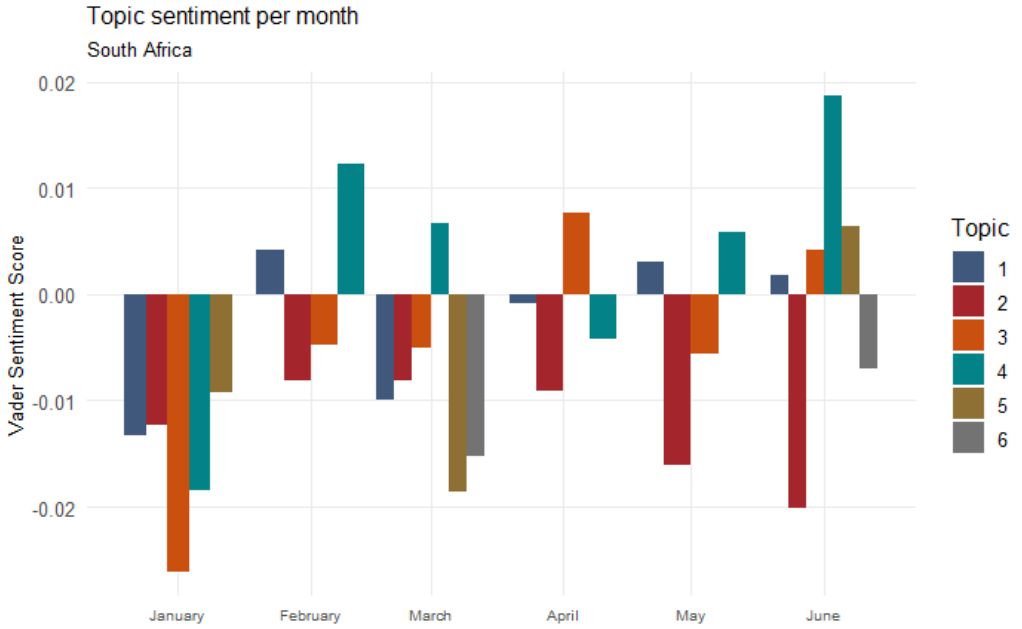

Covid-related Topics in May
South Africa
1
india
vaccine
government
president
restrictions
world
health
patients
risk
sa
0.00
0.01
0.02
0.03
0.04
2
vaccine
oped
business
citizen
vaccination
wave
newsdeck
groundup
world
cape
0.000
0.005
0.010
0.015
0.020
0.025
3
total
recorded
infections
taking
health
statistics
cycle
president
deaths
world
0.00
0.01
0.02
0.03
4
vaccine
health
cape
people
death
study
vaccination
ramaphosa
data
global
0.00
0.01
0.02
5
ramaphosa
president
nation
sunday
response
cyril
developments
address
wednesday
toll
0.000
0.005
0.010
0.015
0.020
term
beta



5.3.2 Average topic sentiment per month (SA)

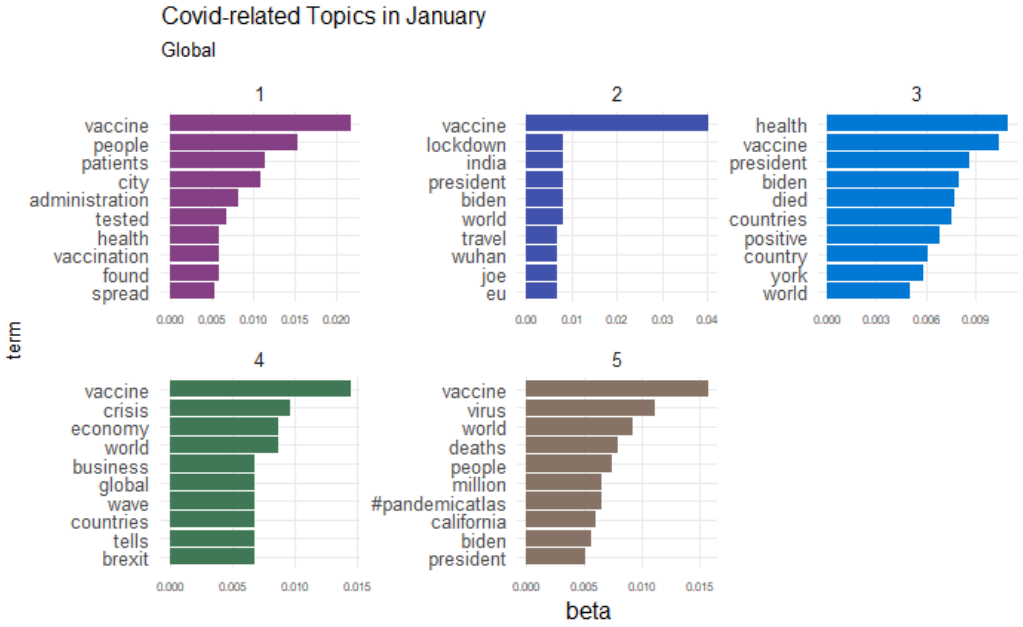
The topic sentiment is calculated for the South African topics identified in the monthly topics identified above.

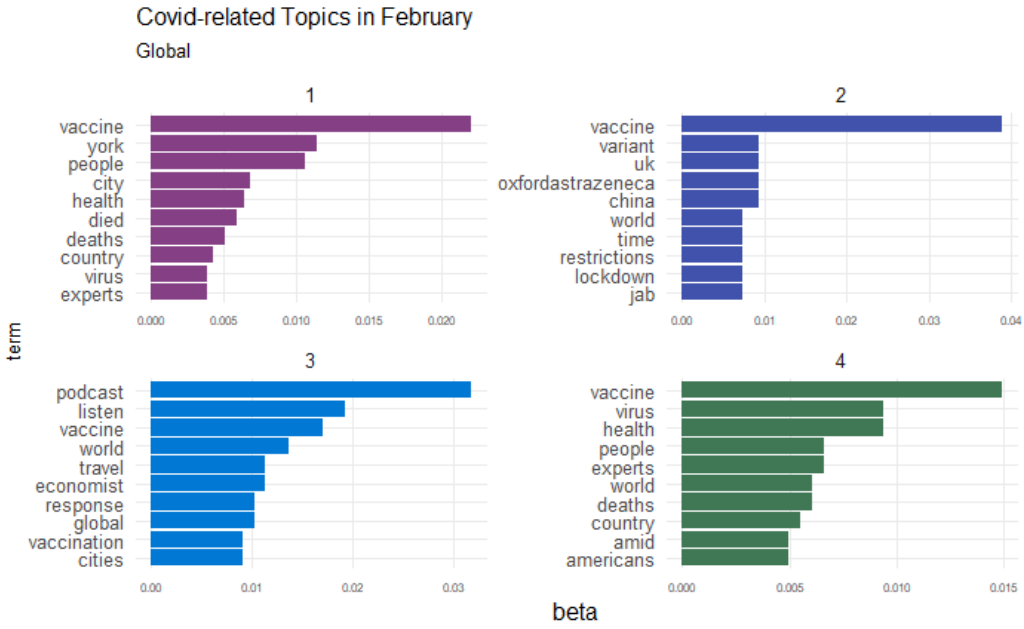
We see a trend that the topic sentiment over time becomes more positive as time goes on. Initially in January the topics that were present all had negative sentiment, but by June all but two of the topics has a positive sentiment. The most negative topic in January revolves around themes of China, virus detection, variants, and death tolls. February and May has a 50% positive rate per topic, with March and April having more negative sentiment with only one topic in each month having a positive sentiment. Furthermore, June seems like a particularly positive month, with most topics having positive sentiment. Even so, there is a significant negative topic is topic 2, which revolves around opinion pieces surrounding vaccination, health, the wave, business, education, and the world. Similarly, there is a significantly positive topic surrounding a theme of vaccination roll-out amidst the thirds wave in Gauteng.

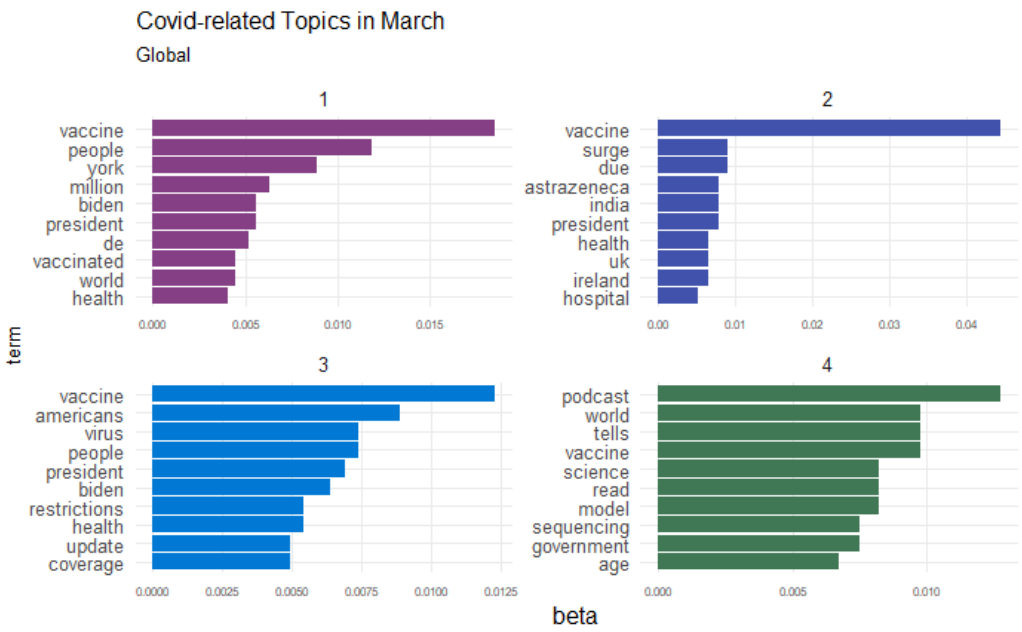


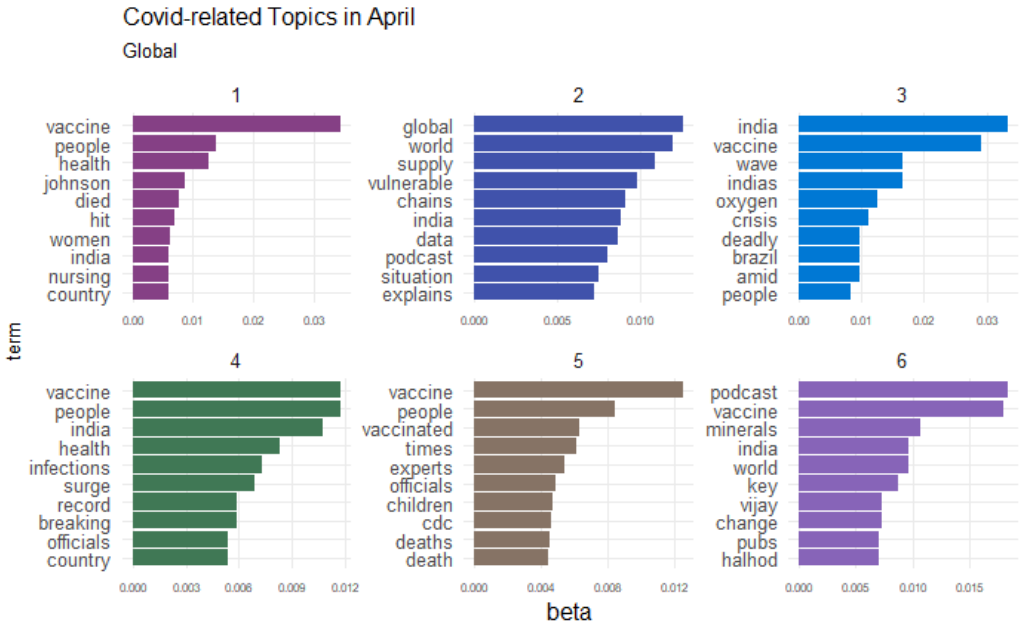
5.3.3 Global topics over time

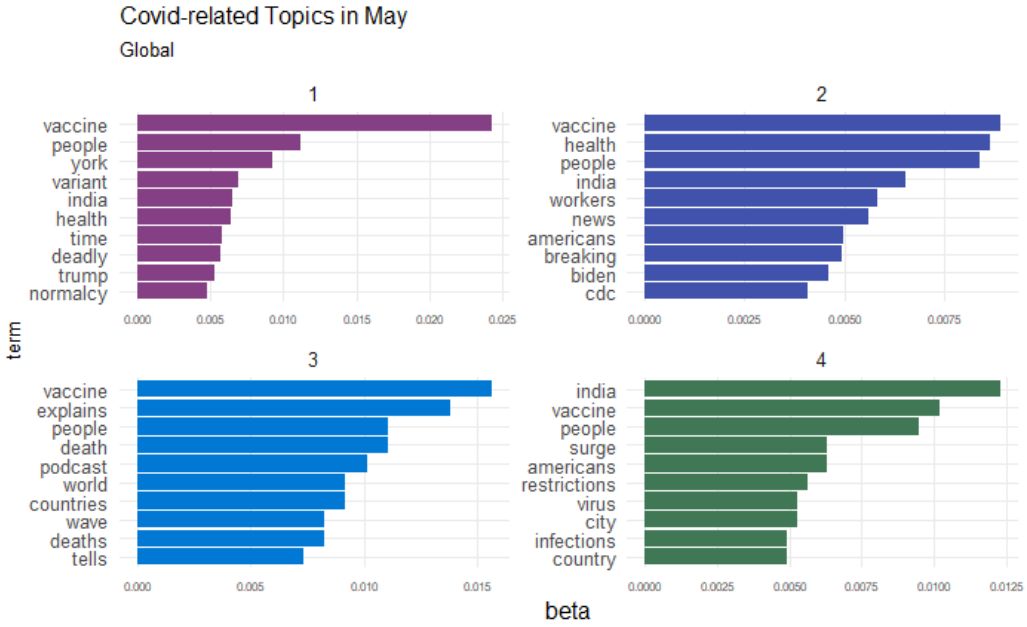
In the figures below, we see that in global news reporting vaccinating people seems to be the biggest concern. The world, President Biden, the *The Economist's* podcast explaining different aspects relating to COVID-19, deaths, and lockdowns are further topics that appear throughout the last six months. In January, economic crises is another topic, and India is already in the headlines, presumably because of the variant. There is a conversation in February to March about the vaccine that is produced by the British-Swedish biotechnology company AstraZeneca in collaboration with Oxford University. In April, as the vaccine developed by Johnson & Johnson makes headlines, the most recent wave of COVID-19 cases floods countries all over the world as the Indian variant spreads. The world faces vulnerable supply chains and an oxygen crisis where countries such as Brazil run out of oxygen to treat COVID-19 patients. Infections continue to surge throughout April and May.

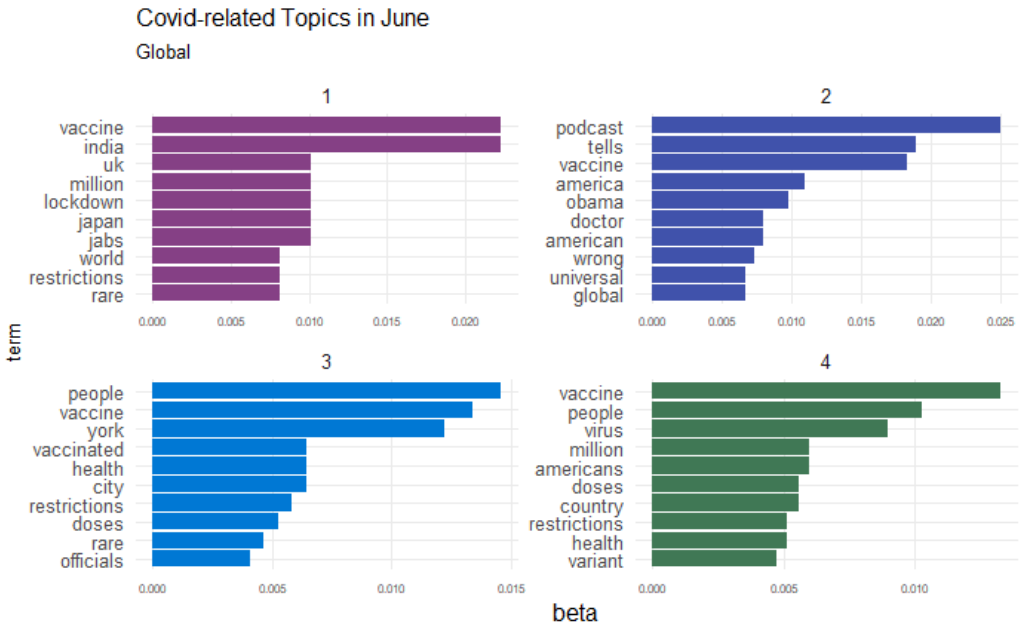

Covid-related Topics in January
Global
1
vaccine
people
patients
city
administration
tested
health
vaccination
found
spread
0.000 0.005 0.010 0.015 0.020
2
vaccine
lockdown
india
president
biden
world
travel
wuhan
joe
eu
0.00 0.01 0.02 0.03 0.04
3
health
vaccine
president
biden
died
countries
positive
country
york
world
0.000 0.003 0.006 0.009
4
vaccine
crisis
economy
world
business
global
wave
countries
tells
brexit
0.000 0.005 0.010 0.015
5
vaccine
virus
world
deaths
people
million
#pandemicatlas
california
biden
president
0.000 0.005 0.010 0.015
term
beta


Covid-related Topics in February
Global
1
vaccine
york
people
city
health
died
deaths
country
virus
experts
0.000 0.005 0.010 0.015 0.020
2
vaccine
variant
uk
oxfordastrazeneca
china
world
time
restrictions
lockdown
jab
0.00 0.01 0.02 0.03 0.04
3
podcast
listen
vaccine
world
travel
economist
response
global
vaccination
cities
0.00 0.01 0.02 0.03
4
vaccine
virus
health
people
experts
world
deaths
country
amid
americans
0.000 0.005 0.010 0.015
term
beta


Covid-related Topics in March
Global
1
vaccine
people
york
million
biden
president
de
vaccinated
world
health
0.000
0.005
0.010
0.015
2
vaccine
surge
due
astrazeneca
india
president
health
uk
ireland
hospital
0.00
0.01
0.02
0.03
0.04
3
vaccine
americans
virus
people
president
biden
restrictions
health
update
coverage
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
4
podcast
world
tells
vaccine
science
read
model
sequencing
government
age
0.000
0.005
0.010
term
beta


Covid-related Topics in April
Global
1
vaccine
people
health
johnson
died
hit
women
india
nursing
country
0.00
0.01
0.02
0.03
2
global
world
supply
vulnerable
chains
india
data
podcast
situation
explains
0.000
0.005
0.010
3
india
vaccine
wave
indias
oxygen
crisis
deadly
brazil
amid
people
0.00
0.01
0.02
0.03
4
vaccine
people
india
health
infections
surge
record
breaking
officials
country
0.000
0.003
0.006
0.009
0.012
5
vaccine
people
vaccinated
times
experts
officials
children
cdc
deaths
death
0.000
0.004
0.008
0.012
6
podcast
vaccine
minerals
india
world
key
vijay
change
pubs
halhod
0.000
0.005
0.010
0.015
term
beta



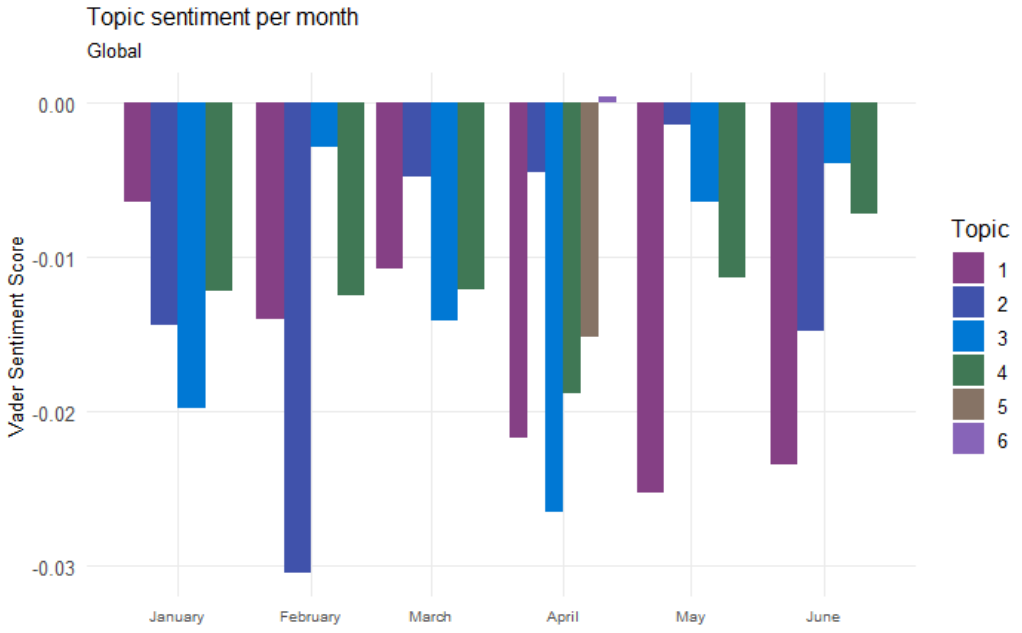


5.3.4 Average topic sentiment per month (Global)

The topic sentiment is calculated for the global topics identified in the monthly topics identified above.

We see a trend that the topic sentiment over time being that all but one of the topics having a consistent negative sentiment. There seems to be

an overall improvement of sentiment in March, but generally in cases where there is a month with a topic with a low negative topic sentiment, it gets countered by another topic with a high negative topic sentiment. The most negative topic sentiment is in February for topic 2, which revolves around the vaccine, variants, countries throughout the world, and lockdown restrictions. The most positive topic (which is hardly above a neutral 0) is in April for topic 6, which revolves around themes in podcasts, vaccines, minerals, India, and pubs.



5.3.5 Key outcomes

Besides the recurring themes of vaccines, infections, deaths, lockdowns, and the Indian variant - which seems to hit South Africa a month later than globally - topics covered by the South African news outlets do not overlap with those of the global news outlets. As expected, the South African media outlets cover news relevant to South Africa, whereas the topics covered by global media outlets are of international interest.

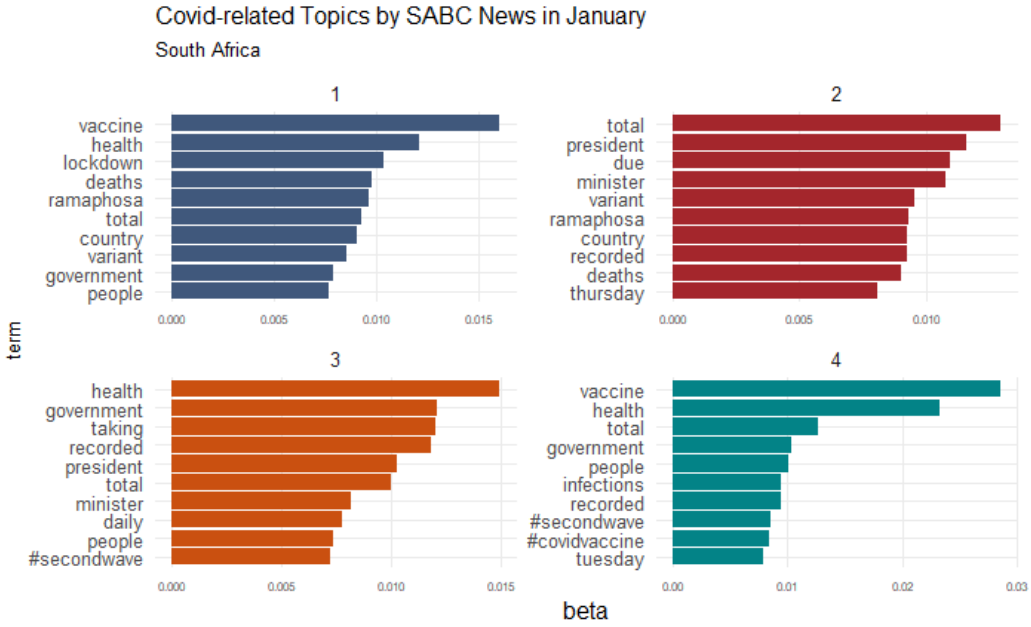
Furthermore, in the sentiment analysis per topic over the past six months we see a similar trend to that of the overall sentiment in Section 4.3 where the sentiment on the topics discussed by the selected South African media outlets are less negative than the global topics, with the South African sentiment becoming more positive as time goes on.

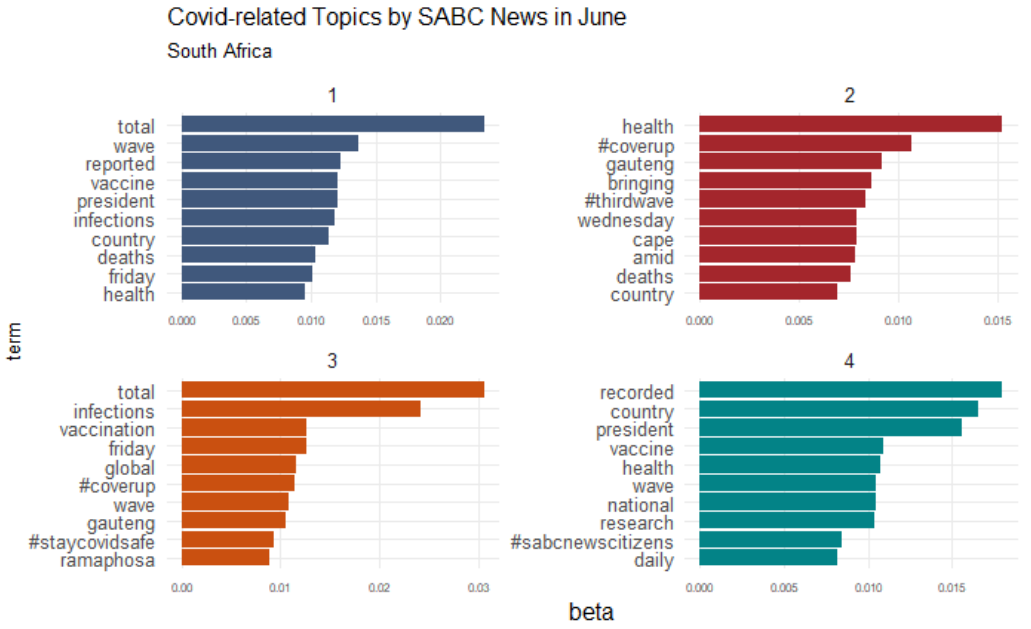
5.4 Topics per media outlet over time: January vs June

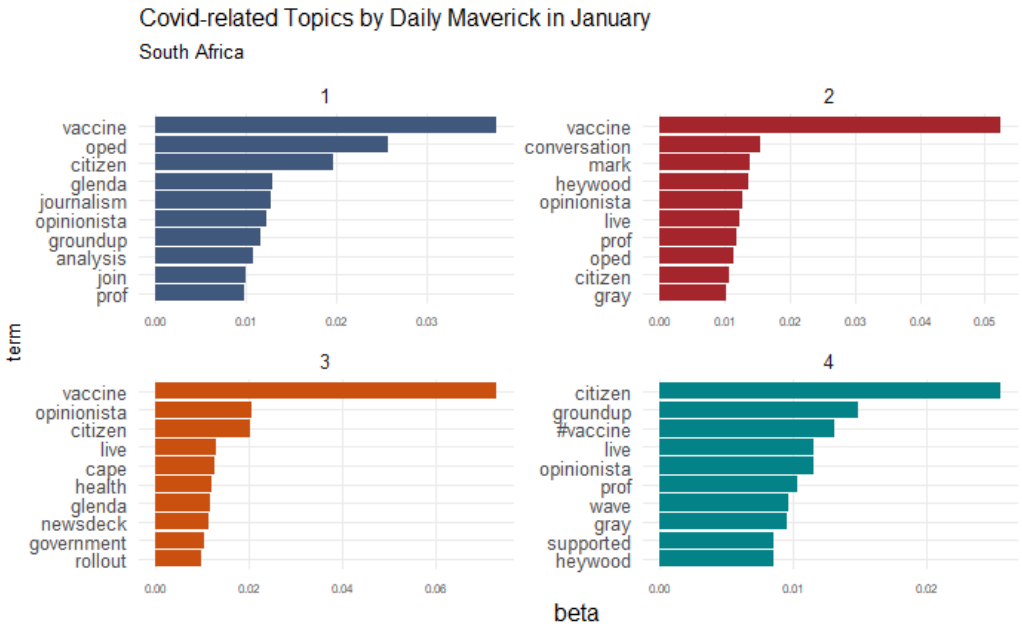
As news on COVID-19 varies from organisation to organisation, and changes over time, the question arises whether there are significant changes in how individual organisations report over time. In the following section, the Corona-related topics for each organisation are considered for the months January and June to explore whether there are any noteworthy changes in reporting on COVID-19 from the start to the end of the first six months of 2021.

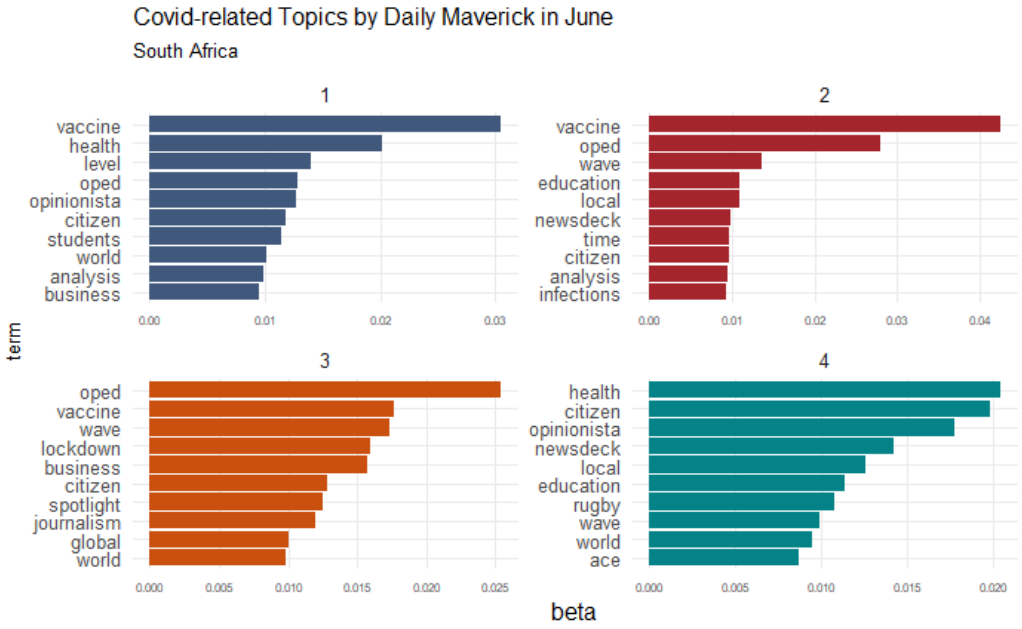
5.4.1 South African topics per media outlet over time

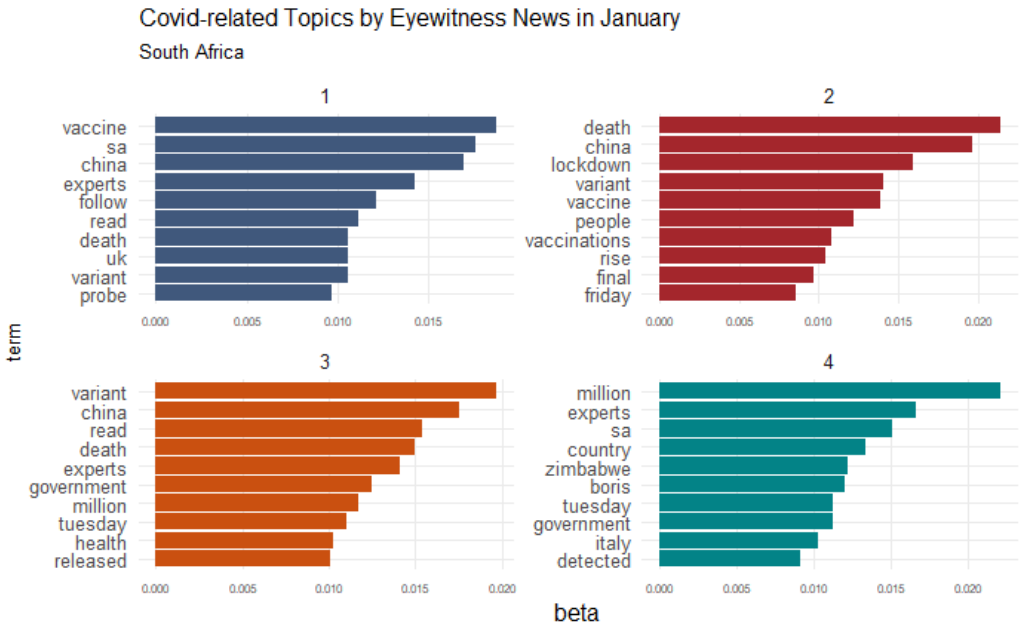
The below graphs show how reporting by the South African news outlets differ from January to June, but also how they differ from each other. Overall, the topics they report on correspond to the overall outlet topics as described in Section 5.2. For example, *SABC News* reports on the second and third waves and President Ramaphosa, and "opinionista" is again specific to the *Daily Maverick*, for both January and June, indicating that the *Daily Maverick* continuously publishes articles expressing personal opinions. *Eyewitness News* seems to report slightly differently to other organisations by reporting more specifically on various countries and areas, such as Zimbabwe, Italy and China in January, and Thailand, India and Gauteng in June. While the type of topics covered by these media organisations differ between the organisations, they remain similar from January to June. *News24* on the other hand shows more change, as topics covered in January seem to revolve around infections, fatalities and death toll, and more about the delta variant, Gauteng, and a lockdown in June.

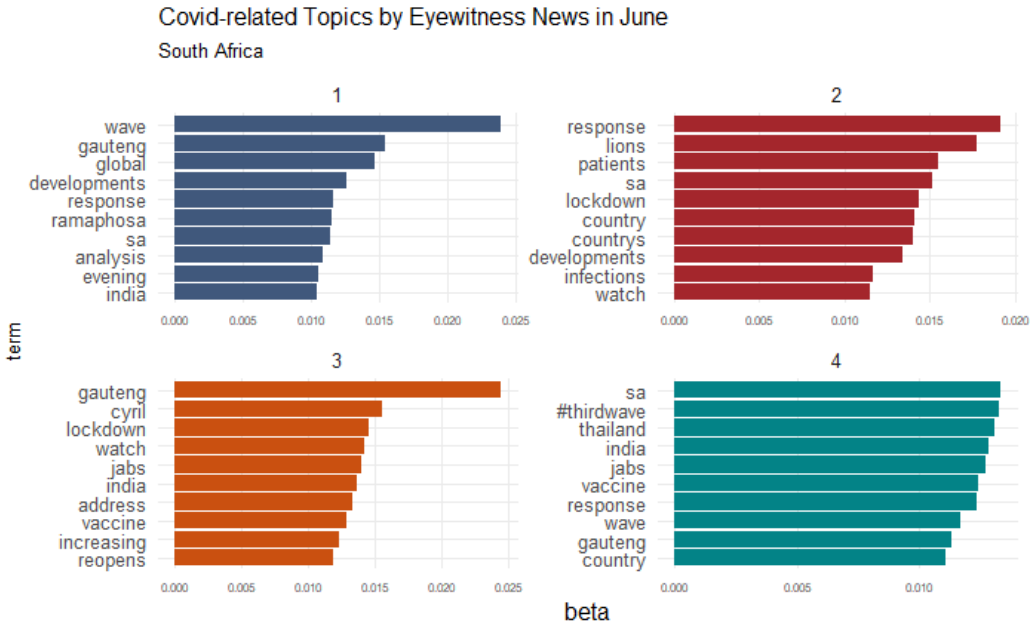

Covid-related Topics by SABC News in January
South Africa
1
vaccine
health
lockdown
deaths
ramaphosa
total
country
variant
government
people
2
total
president
due
minister
variant
ramaphosa
country
recorded
deaths
thursday
3
health
government
taking
recorded
president
total
minister
daily
people
#secondwave
4
vaccine
health
total
government
people
infections
recorded
#secondwave
#covidvaccine
tuesday
term
beta

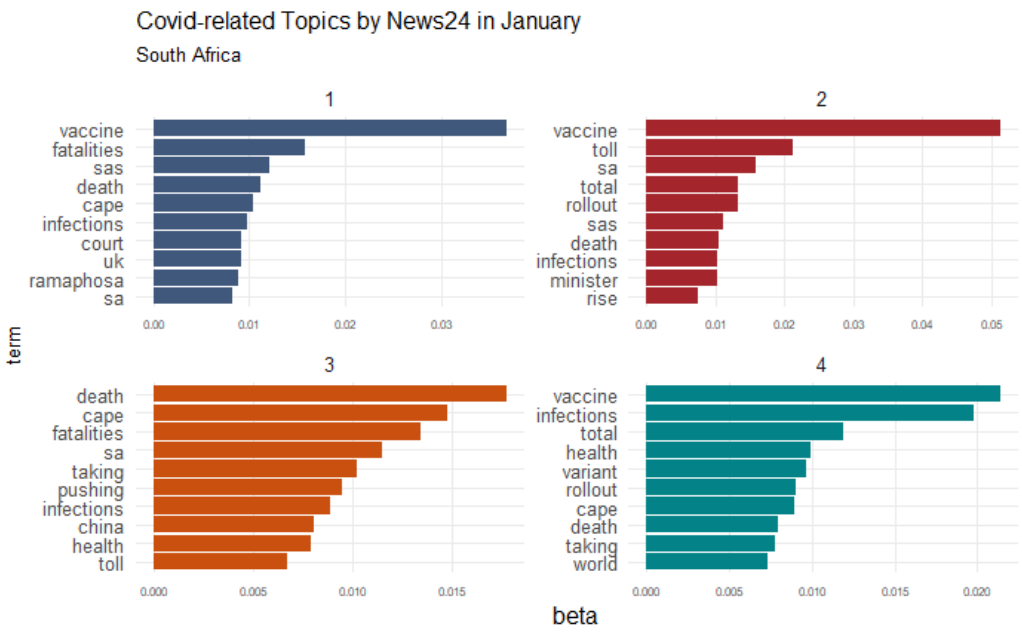

Covid-related Topics by SABC News in June
South Africa
1
total
wave
reported
vaccine
president
infections
country
deaths
friday
health
2
health
#coverup
gauteng
bringing
#thirdwave
wednesday
cape
amid
deaths
country
3
total
infections
vaccination
friday
global
#coverup
wave
gauteng
#staycovidsafe
ramaphosa
4
recorded
country
president
vaccine
health
wave
national
research
#sabcnewscitizens
daily
term
beta

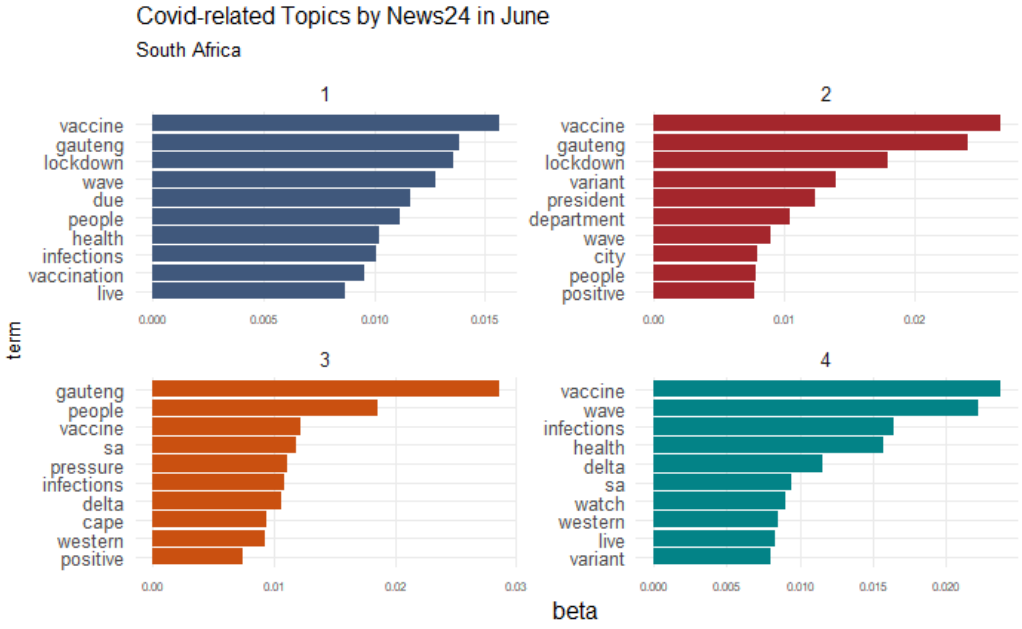

Covid-related Topics by Daily Maverick in January
South Africa
1
vaccine
oped
citizen
glenda
journalism
opinionista
groundup
analysis
join
prof
2
vaccine
conversation
mark
heywood
opinionista
live
prof
oped
citizen
gray
3
vaccine
opinionista
citizen
live
cape
health
glenda
newsdeck
government
rollout
4
citizen
groundup
#vaccine
live
opinionista
prof
wave
gray
supported
heywood
term
beta


Covid-related Topics by Daily Maverick in June
South Africa
1
vaccine
health
level
oped
opinionista
citizen
students
world
analysis
business
2
vaccine
oped
wave
education
local
newsdeck
time
citizen
analysis
infections
3
oped
vaccine
wave
lockdown
business
citizen
spotlight
journalism
global
world
4
health
citizen
opinionista
newsdeck
local
education
rugby
wave
world
ace
term
beta


Covid-related Topics by Eyewitness News in January
South Africa
1
vaccine
sa
china
experts
follow
read
death
uk
variant
probe
2
death
china
lockdown
variant
vaccine
people
vaccinations
rise
final
friday
3
variant
china
read
death
experts
government
million
tuesday
health
released
4
million
experts
sa
country
zimbabwe
boris
tuesday
government
italy
detected
0.000
0.005
0.010
0.015
0.020
term
beta


Covid-related Topics by Eyewitness News in June
South Africa
1
wave
gauteng
global
developments
response
ramaphosa
sa
analysis
evening
india
2
response
lions
patients
sa
lockdown
country
countrys
developments
infections
watch
3
gauteng
cyril
lockdown
watch
jabs
india
address
vaccine
increasing
reopens
4
sa
#thirdwave
thailand
india
jabs
vaccine
response
wave
gauteng
country
0.000
0.005
0.010
0.015
0.020
0.025
term
beta





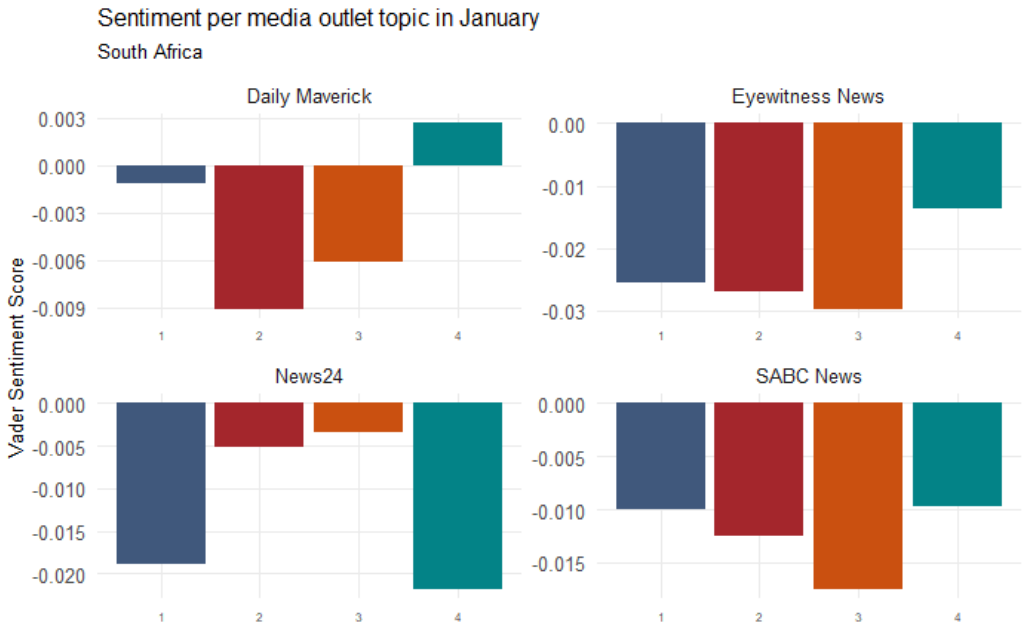
5.4.2 Average topic sentiment per outlet over time (SA)

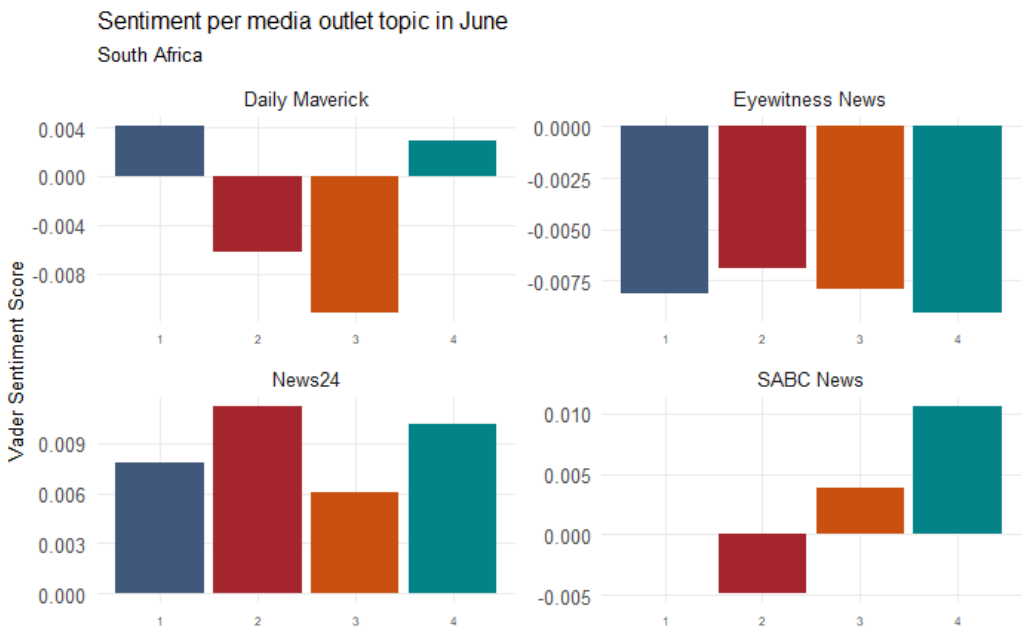
The topic sentiment is calculated for the topics in January and June identified in the monthly topics identified above.

From the graphs below there is generally a trend in increased positivity in topic sentiment per media outlet over between January and June. *The*

Daily Maveric has little change in the (negative) topic sentiment that they cover, whereas *Eyewitness News* becomes increasingly negative (by a factor of ≈ 10). Furthermore, there is a significant change in sentiment for *SABC News* and especially *News24*. *News24's* topic sentiment changes from an all negative sentiment across topics in January to a significantly positive sentiment across topics in June. As discussed above, the topics that *News24* cover in January versus June is also significantly different. *SABC News* has a similar, less drastic, change from a significantly negative sentiment on the topics covered in January to a mostly positive or neutral sentiment in June.

The most negative topic is *Eyewitness News'* topic 3 in January, which revolves around the variant, china, death tolls and experts' and governments' responses. The most positive topic covered by the media outlets overall is *News24's* topic 2 in June, which revolves around the vaccine, Gauteng, positive cases, and the lockdown. As discussed in the Limitations section, this could be an inaccurate sentiment analysis due to the fact that it identifies the term "positive". In the context of COVID-19, "positive" is likely used in conjunction with "cases", which is in fact negative. Due to the nature of the unnesting in the *LDA tuning*, the *Vader* sentiment analysis is applied to individual terms associated to topics instead of entire tweets associated to topics.

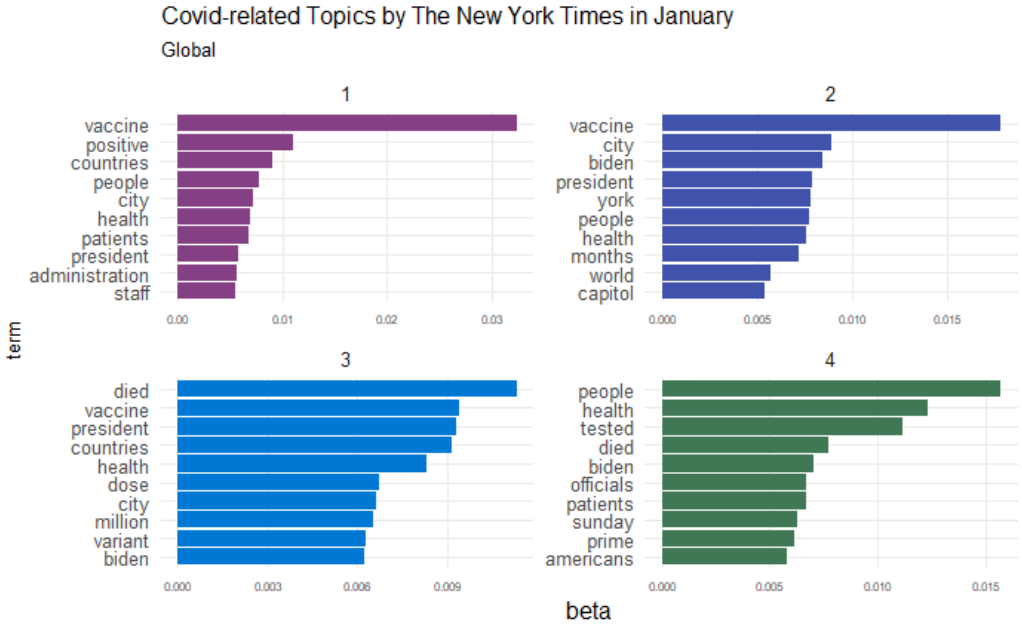


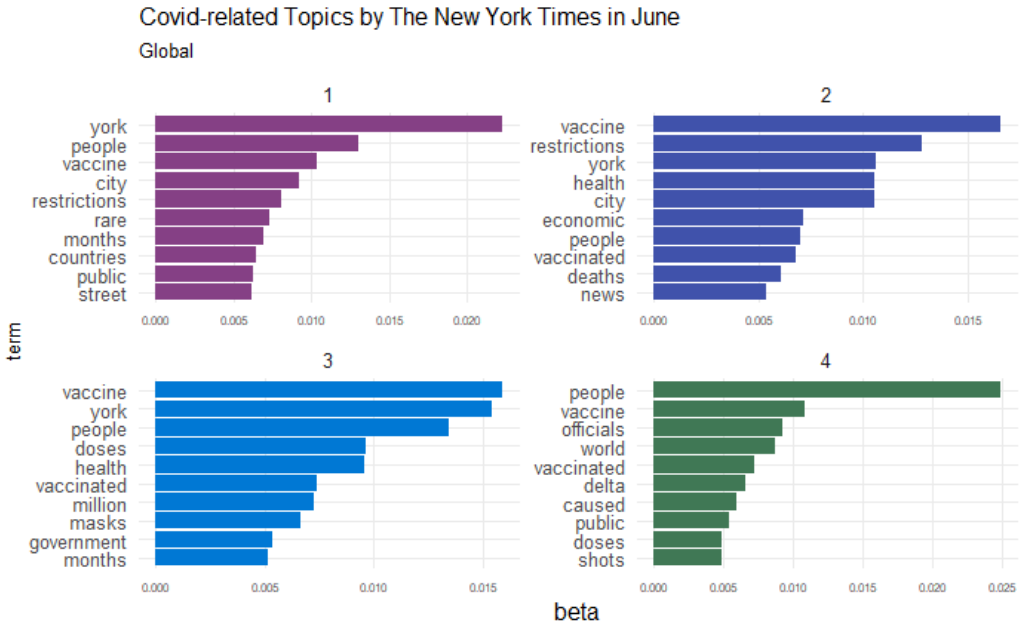


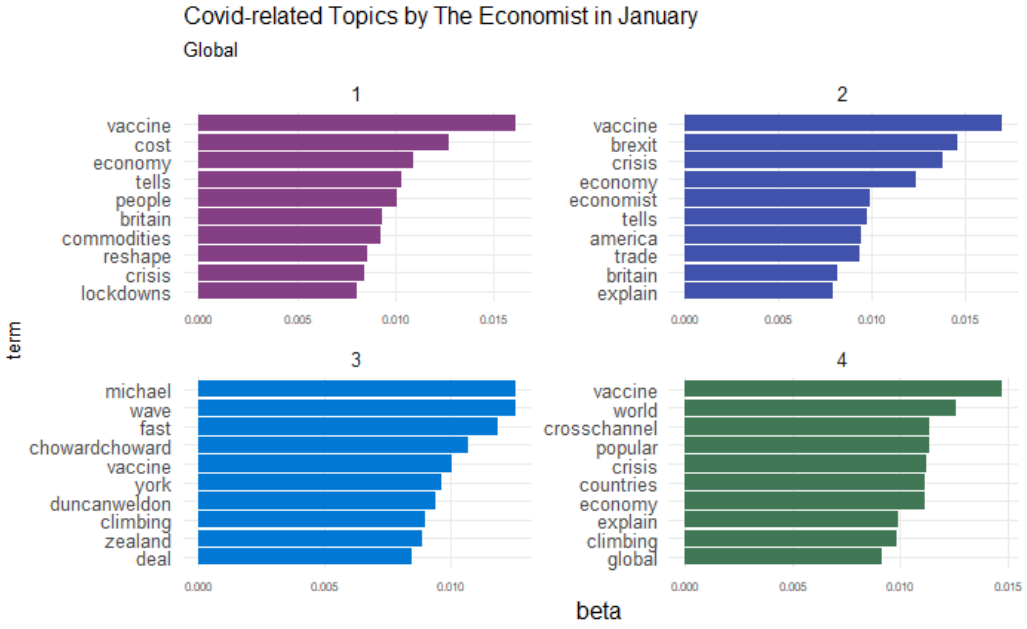
5.4.3 Global topics per media outlet over time

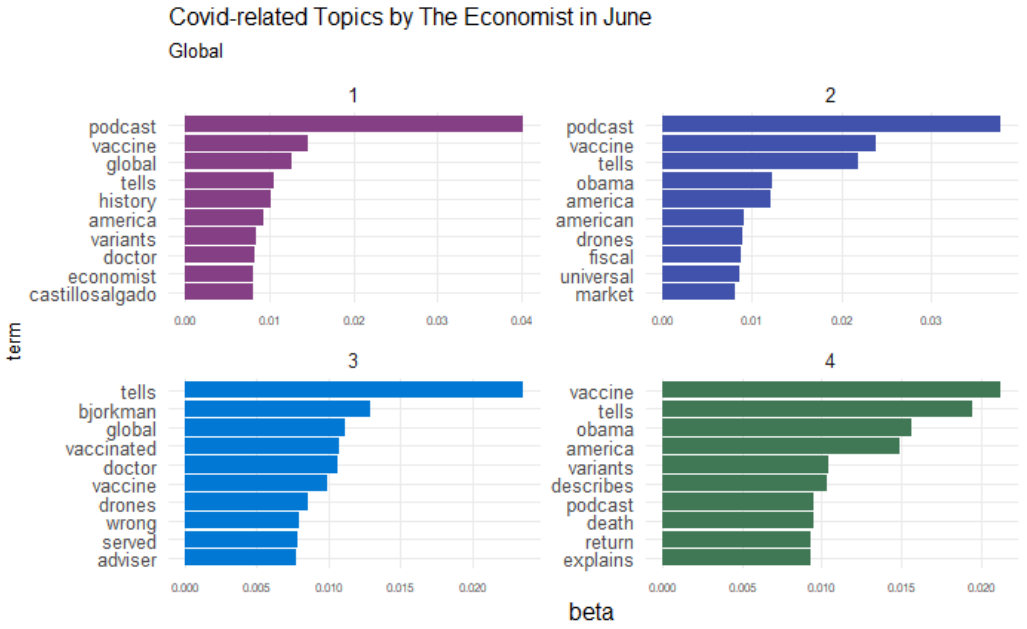
Similar to the above for South Africa, the below graphs show how news reporting by global media organisations changes over time, and how they vary from each other.

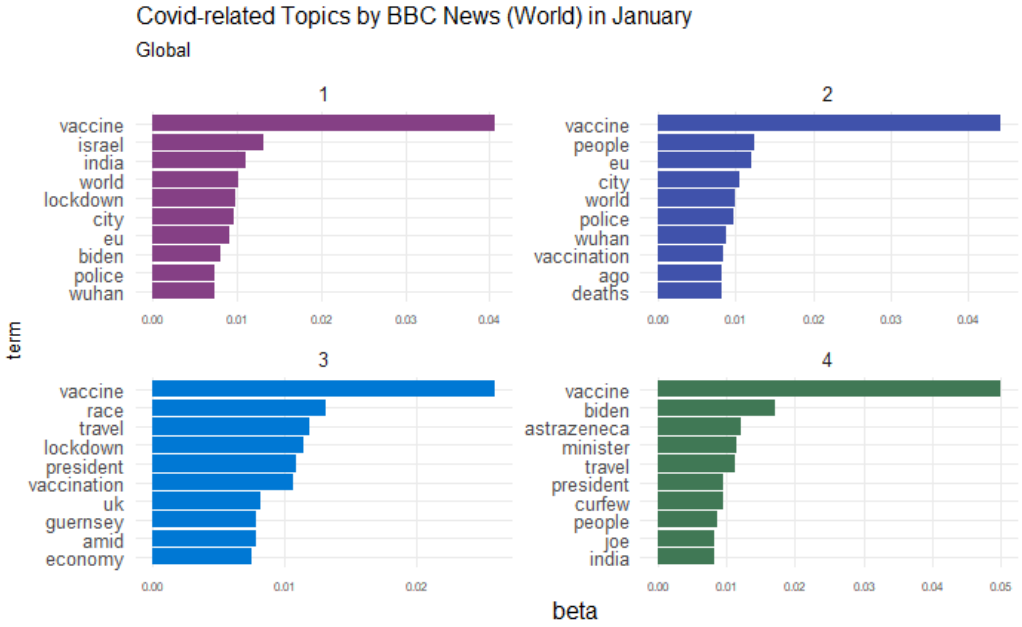
The type of reporting by the *New York Times* and the *Associated Press* appear to correspond to their overall topics as analysed in Section 5.2 throughout the months, with topics revolving around the vaccine, people, and American officials in both January and June. In March, the *Economist* started "The Jab", a podcast reporting on the race to produce an effective COVID-19 vaccination. This could explain why the podcast features heavily in June, whereas *The Economist's* reporting in January focuses on economy, lockdowns, and crises around the world. The overall topics from Section 5.2 of *BBC (World)* are reflected in both January and June. However, it is worth mentioning that *BBC (World)* includes the upcoming 2021 Olympics scheduled to start towards the end of July in Tokyo, Japan, as one of its main news topics. The *Associated Press* is the only other news outlet of all the global or South African organisations where this topic also makes an appearance.

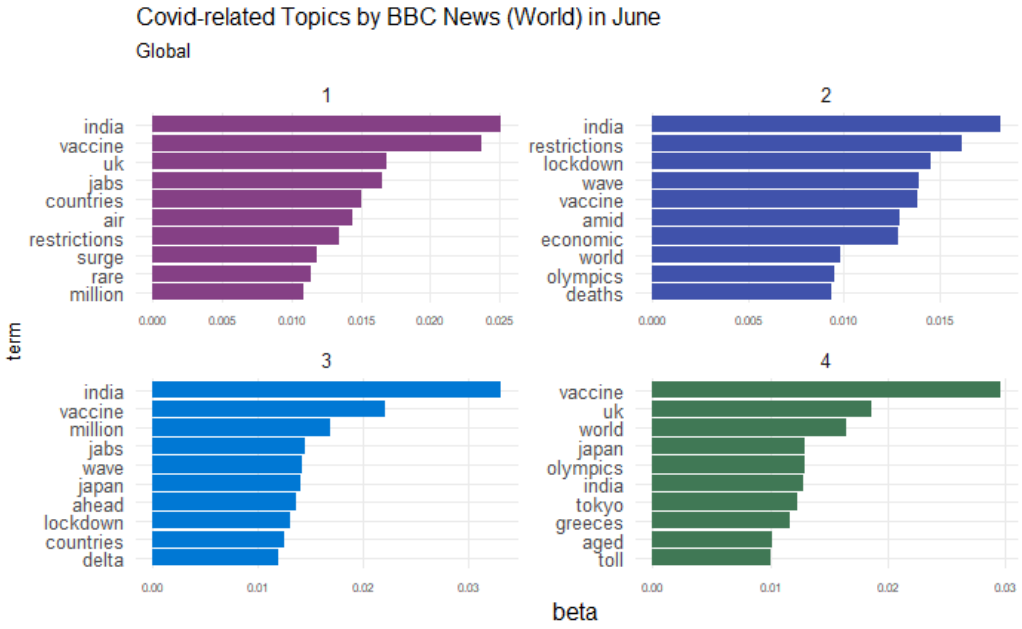

Covid-related Topics by The New York Times in January
Global
1
vaccine
positive
countries
people
city
health
patients
president
administration
staff
0.00
0.01
0.02
0.03
2
vaccine
city
biden
president
york
people
health
months
world
capitol
0.000
0.005
0.010
0.015
3
died
vaccine
president
countries
health
dose
city
million
variant
biden
0.000
0.003
0.006
0.009
4
people
health
tested
died
biden
officials
patients
sunday
prime
americans
0.000
0.005
0.010
0.015
term
beta

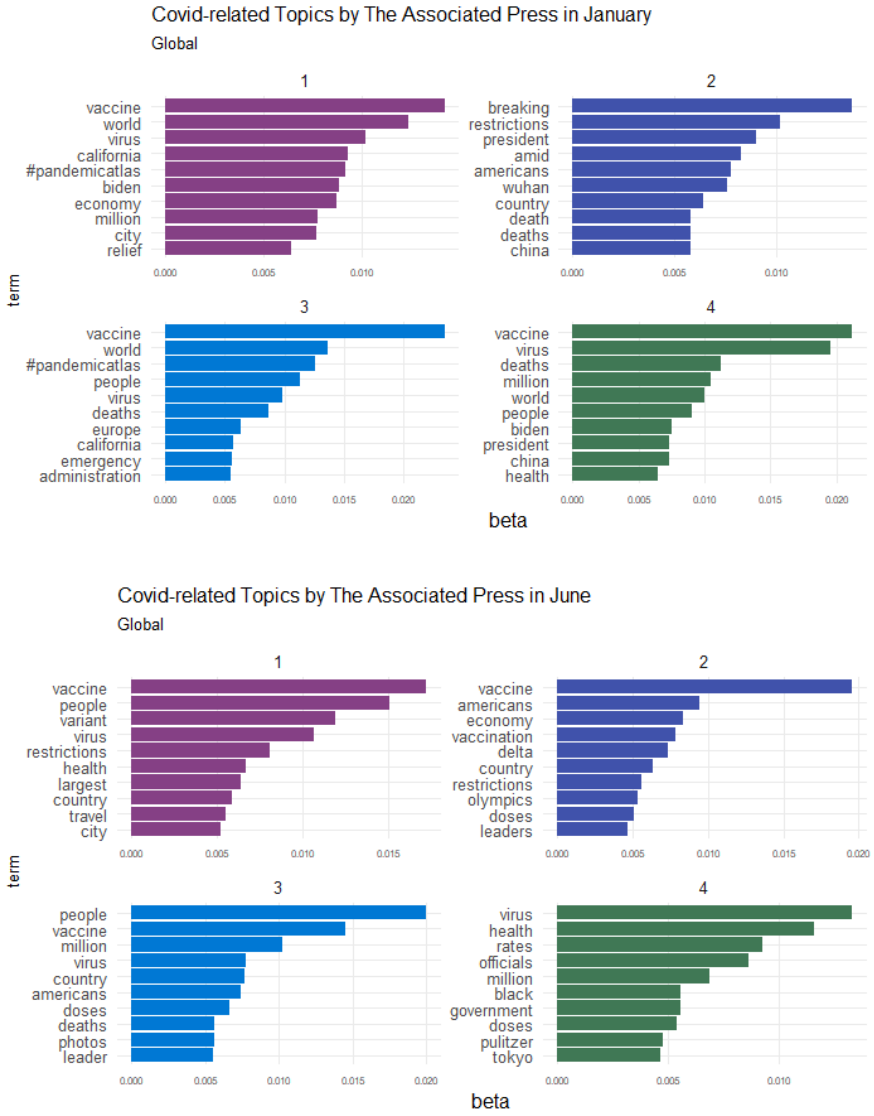

Covid-related Topics by The New York Times in June
Global
1
york
people
vaccine
city
restrictions
rare
months
countries
public
street
0.000
0.005
0.010
0.015
0.020
2
vaccine
restrictions
york
health
city
economic
people
vaccinated
deaths
news
0.000
0.005
0.010
0.015
3
vaccine
york
people
doses
health
vaccinated
million
masks
government
months
0.000
0.005
0.010
0.015
4
people
vaccine
officials
world
vaccinated
delta
caused
public
doses
shots
0.000
0.005
0.010
0.015
0.020
0.025
term
beta


Covid-related Topics by The Economist in January
Global
1
vaccine
cost
economy
tells
people
britain
commodities
reshape
crisis
lockdowns
2
vaccine
brexit
crisis
economy
economist
tells
america
trade
britain
explain
3
michael
wave
fast
chowardchoward
vaccine
york
duncanweldon
climbing
zealand
deal
4
vaccine
world
crosschannel
popular
crisis
countries
economy
explain
climbing
global
0.000
0.005
0.010
0.015
term
beta


Covid-related Topics by The Economist in June
Global
1
podcast
vaccine
global
tells
history
america
variants
doctor
economist
castillosalgado
2
podcast
vaccine
tells
obama
america
american
drones
fiscal
universal
market
3
tells
bjorkman
global
vaccinated
doctor
vaccine
drones
wrong
served
adviser
4
vaccine
tells
obama
america
variants
describes
podcast
death
return
explains
0.00
0.01
0.02
0.03
0.04
0.000
0.005
0.010
0.015
0.020
term
beta


Covid-related Topics by BBC News (World) in January
Global
1
vaccine
israel
india
world
lockdown
city
eu
biden
police
wuhan
2
vaccine
people
eu
city
world
police
wuhan
vaccination
ago
deaths
3
vaccine
race
travel
lockdown
president
vaccination
uk
guernsey
amid
economy
4
vaccine
biden
astrazeneca
minister
travel
president
curfew
people
joe
india
term
beta


Covid-related Topics by BBC News (World) in June
Global
1
india
vaccine
uk
jabs
countries
air
restrictions
surge
rare
million
2
india
restrictions
lockdown
wave
vaccine
amid
economic
world
olympics
deaths
3
india
vaccine
million
jabs
wave
japan
ahead
lockdown
countries
delta
4
vaccine
uk
world
japan
olympics
india
tokyo
greeces
aged
toll
term
beta



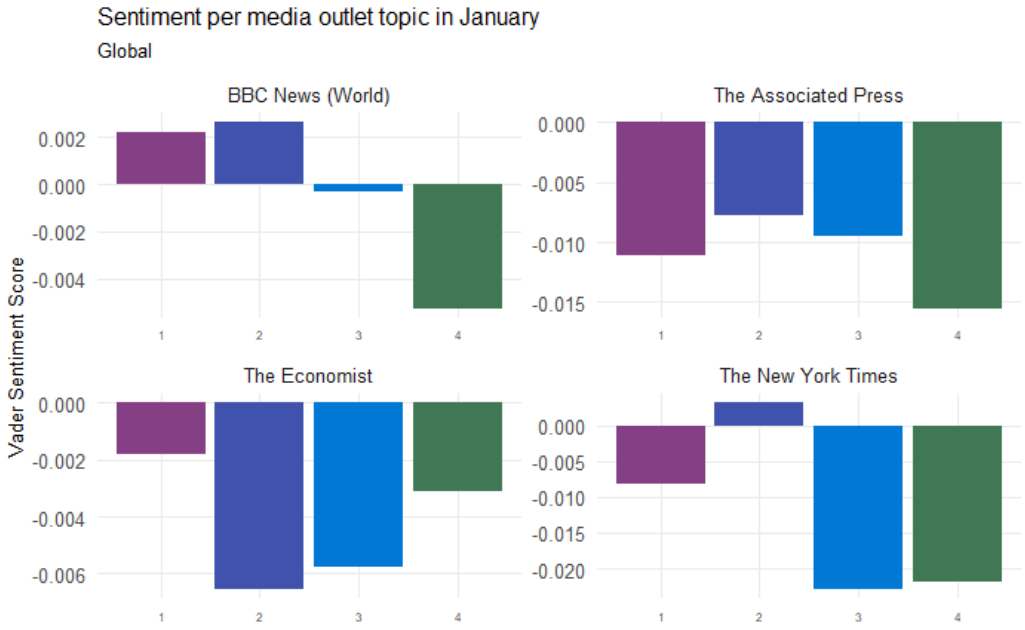
5.4.4 Average topic sentiment per outlet over time (Global)

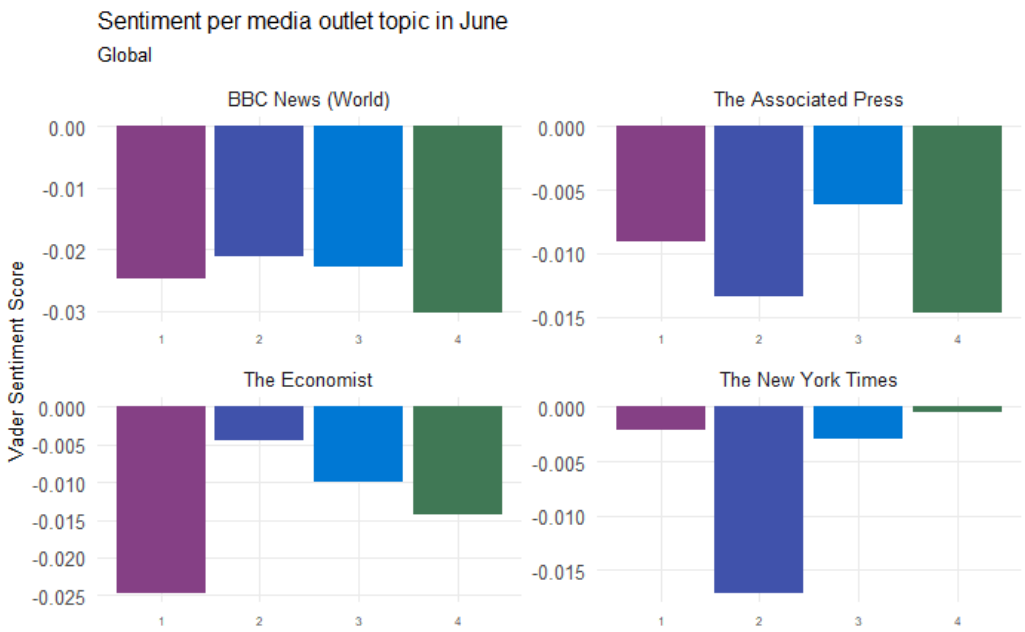
The topic sentiment is calculated for the topics in January and June identified in the monthly topics identified above.

From the graphs below there is generally a trend in change in topic sentiment per media outlet between January and June. While *The Associated Press* generally has the same sentiment toward the topics covered in January versus June, all other outlets have changed their sentiments quite significantly. BBC (World) went from sort of positive to mostly negative by more than a factor of ≈ 10, and *The Economist* and *The New York Times* went

from an overall mostly negative discussion in January to a sentiment with some topics being more negative and some being less negative in June.

The most negative topic is *BBC (World)'s* topic 4 in June, which revolves around the Olympics in Tokyo, and the different countries at play. Negative sentiment could be due to people being particularly worried what the effect of COVID-19 on this international event will be. The most positive topic is *BBC (World)'s* topic 2 in January, which revolves around vaccination, the European Union, and police. As the most positive and most negative topic is produced by the same news outlet, this shows that news outlets can change their tune over time.





5.4.5 Key outcomes

From the above, one can conclude that the various media agencies have different focal areas and sentiment within the main topics they cover, and that while most remain similar over time, some show significant change from January to June - such as *News24*, *BBC World*, and the *Economist*. Overall, it seems that the global news outlets show more change and variety between each other and over time, compared to the South African organisations. This could possibly indicate that the type of reporting and the main concerns regarding the spread of the COVID-19 in South Africa have remained relatively consistent throughout the development of the pandemic in the first six months of 2021.

6 Overall analysis

In the above sections we discussed how we discussed many results in detail. In this section we will summarise the key findings in each section.

6.1 Numerical analysis

In the numerical analysis we see that South African media outlets have a lot less activity than globally even though all the media agencies tweeted roughly the same amount over the past six months. The selected South African media agencies' amount of conversation regarding COVID-19 is a lot more varied than globally. Furthermore, one can deduce that the followers of the South African media outlets are most interested in content relating to COVID-19, as they are more likely to retweet posts relating to it. This is different when comparing it to the global conversation, where it seems like the media outlets post more about COVID-19, but the main point of interest of their followers are not necessarily there.

Furthermore, when analysing tweet frequency it becomes apparent that the South African media agencies generally tweet more frequently than the global media agencies. The number of Covid-related tweets for South African and global organisations usually ranges anywhere between 10 to 50 tweets per five days. All of the organisations tweet relatively consistently, whereas the amount of Covid-related tweets posted fluctuates more significantly. It seems that, besides tweeting more frequently, South African news outlets follow a similar tweeting pattern to the global news outlets.

6.2 Sentiment analysis

Through the overall sentiment analysis we see that the selected global media outlets are as a whole more negative about COVID-19 than the South African media outlets during the past six months. We also see that this is the case across media outlets and that the South African media outlets are more polarised on their sentiment toward COVID-19 than the global media outlets are. We also see that the global media outlets has a trend of becoming more negative in the latter half of this analysis, with the South African media outlets being more stable and with some outlets trending toward a less negative sentiment.

6.3 Topic modelling and sentiment per topic

Through topic modelling we see that although the overarching topics do not correspond entirely, certain trends can be found in both South African and global reporting. For example, one might conclude that how presidents are responding, lockdowns, vaccinations and the general impact of COVID-19 on people's lives, are important all over the world.

Both in South Africa and globally the sentiment about topics surrounding their leaders in government is significantly negative. In South Africa this is the most negative sentiment by a significant amount, whereas globally there seems to be a more negative sentiment toward public health and the Indian variant, consequential third wave, lockdowns, and restrictions. Furthermore, in America there seems to be a positive sentiment toward President Biden's response in relation to the New York death toll.

Looking at the overall topics of the last six months relating to COVID-19 covered by the selected South African and global media agencies gives an impression of the general conversations and topics of concern regarding the virus. The topics covered by the individual media outlets show where the focus of each media outlet lies.

Besides the recurring themes of vaccines, infections, deaths, lockdowns, and the Indian variant - which seems to hit South Africa a month later than globally - topics covered by the South African news outlets do not overlap with those of the global news outlets. As expected, the South African media outlets cover news relevant to South Africa, whereas the topics covered by global media outlets are of international interest.

Furthermore, in the sentiment analysis per topic over the past six months we see a similar trend to that of the overall sentiment in Section 4.3 where the sentiment on the topics discussed by the selected South African media outlets are less negative than the global topics, with the South African sentiment becoming more positive as time goes on.

One can conclude that the various media agencies have different focal areas and sentiment within the main topics they cover, and that while most remain similar over time, some show significant change from January to June - such as *News24*, *BBC World*, and the *Economist*. Overall, it seems that the global news outlets show more change and variety between each other and over time, compared to the South African organisations. This could possibly indicate that the type of reporting and the main concerns regarding the spread of the COVID-19 in South Africa have remained relatively consistent throughout the development of the pandemic in the first six months of 2021.

7 Limitations

The analysis above has four main limitations:

1. The keywords “south” and “africa” were removed in cleaning the global tweets. If we had more time, we would include this, as we initially removed it from the South African tweets as it was overpowering the topics in our topic modelling. We noticed too late that these key words should not have been removed from the global data, as these keywords could potentially have given different insights.

2. For the sentiment analysis on the topics, the sentiment analysis algorithm is run on individual words instead of entire sentences that the term is extracted from. This is due to the nature of the way that the topics are extracted using the *LDA* algorithm, where the tweets were unnested into terms. An example where this might be prevalent is where individual words on its own might have a positive sentiment, but in the context of the sentence it might not be a true positive sentence. For example, ”positive” (cases) and ”increase” (in deaths/cases) could increase the overall positivity of a topic.

3. This paper did not analyse the sentiment for the topics per media outlet over time.

4. An important note is that if the topic modelling code is run on Windows, it gives a different result than if the code is run on a Linux distribution even though the seed is set to a constant value of 1234.

8 Conclusion

In this paper, we analysed social media posts on Twitter made by a selection of South African and global news agencies. A description of data collection and wrangling was followed by an analysis and overall discussion of the findings. This included a general analysis, sentiment analysis, topic modelling and an analysis of the sentiment of the topics that are being reported both in South Africa and globally. This examination of the responses to the Coronavirus by various media agencies in South Africa over the last six months and the comparison of their response to that of global agencies provides an analysis of the spread of COVID-19 in South Africa in the first half of 2021.

References

[1] Bettina Grün et al. *Package 'topicmodels'*. URL: https://cran.r-project.org/web/packages/topicmodels/topicmodels.pdf. (accessed: 09.07.2021).

[2] Gabriela De Queiroz et al. *Package 'tidytext'*. URL: https://cran.r-project.org/web/packages/tidytext/tidytext.pdf. (accessed: 09.07.2021).

[3] Michael W. Kearney et al. *rtweet*. URL: https://cran.microsoft.com/web/packages/rtweet/rtweet.pdf. (accessed: 01.07.2021).

[4] Aditya Beri. *Sentiment Analysis using Vader*. URL: https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664. (accessed: 7.07.2021).

[5] SABC News. *SABC News*. URL: https://www.sabcnews.com/sabcnews/. (accessed: 06.07.2021).

[6] Murzintcev Nikita and Nathan Chaney. *Package 'ldatuning'*. URL: https://cran.r-project.org/web/packages/ldatuning/ldatuning.pdf. (accessed: 09.07.2021).

[7] Gavin Phillips. *Top 4 Unbiased Independent World News Sources*. URL: https://www.makeuseof.com/top-unbiased-news-sources/. (accessed: 06.07.2021).

[8] Katherine Roehrick. *Package 'vader'*. URL: https://cran.r-project.org/web/packages/vader/vader.pdf. (accessed: 7.07.2021).

[9] Shobhit Seth. *The World's Top 10 News Media Companies*. URL: https://www.investopedia.com/stock-analysis/021815/worlds-top-ten-news-companies-nws-gci-trco-nyt.aspx. (accessed: 09.07.2021).

[10] Unknown. *World's 10 Most Powerful News Outlets*. URL: https://www.mentaldaily.com/article/2019/11/worlds-most-powerful-news-outlets. (accessed: 09.07.2021).

[11] Matthew Wolff. *TwitterScraper*. URL: https://github.com/MatthewWolff/TwitterScraper. (accessed: 01.07.2021).

[12] Staff Writer. *Record readership for South African online news publications – Top websites and publishers revealed*. URL: https://mybroadband.co.za/news/internet/354497-record-readership-for-south-african-online-news-publications-top-websites-and-publishers-revealed.html. (accessed: 03.07.2021).