



12.07.2021 - Behavioural & Psychological distress of COVID-19 and Infodemics

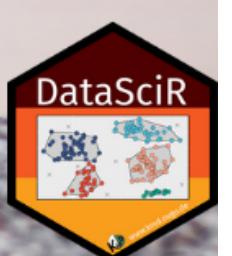
Data Science with R · Summer 2021

- Knowledge Management & Discovery Lab

<https://rpubs.com/ranjiraj9/covidistress>

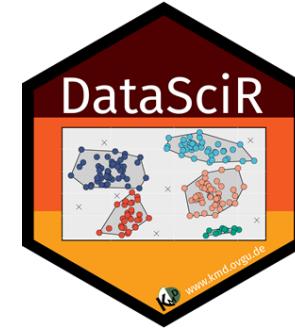


Psychological & Behavioural Distress of COVID-19 and Infodemics



Motivation & Problem statement

- COVID-19 arrival in **2020** and at its peak and in various forms now.
- Aggravated **mental disorders** and its ill-effects on a larger scale.
- Being **isolated** affects productivity and fall prone to addictive substances.
- Increased **social media** usage and **unplanned sleep** causing distress.
- **Infodemics** plays an **evil role** to all above.
- Initial questions:
 - How people are getting stressed and how they are coping with it?
 - Role of *Twitter community* during this period?
 - Whom should be trusted and whom should be not?
- Shiny app <https://covid-distress-infodemics.shinyapps.io/shinyapp/>



Team

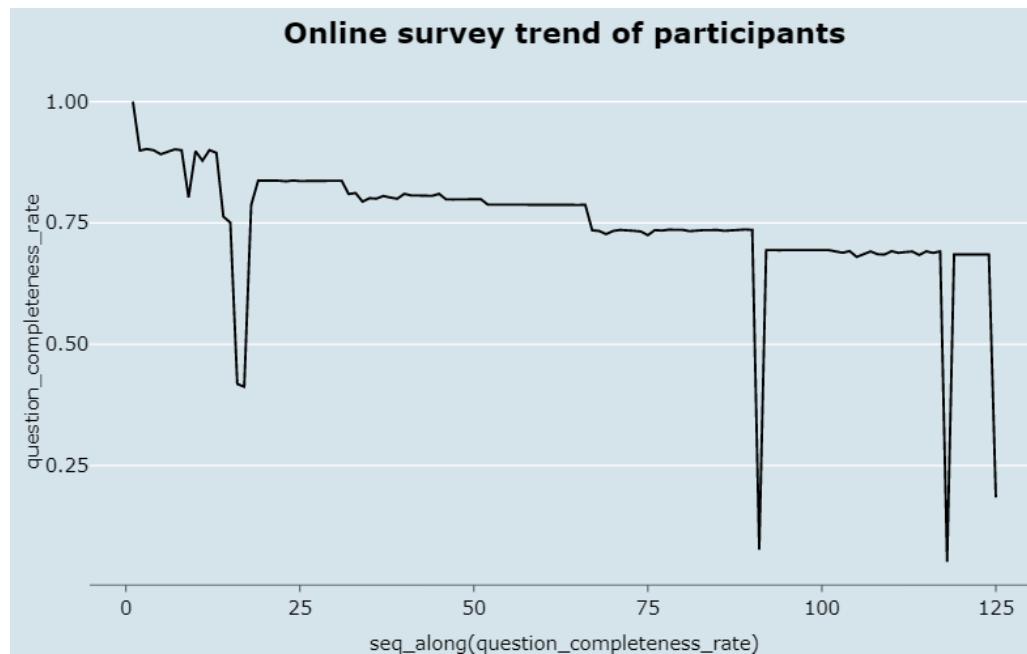
- Madhuri Sajith
- Usama Ashfaq
- Vishnu Jayanand
- Sujith Nyarakkад Sudhakaran
- Ranjiraj Rajendran Nair

Objective 1: Global distress survey

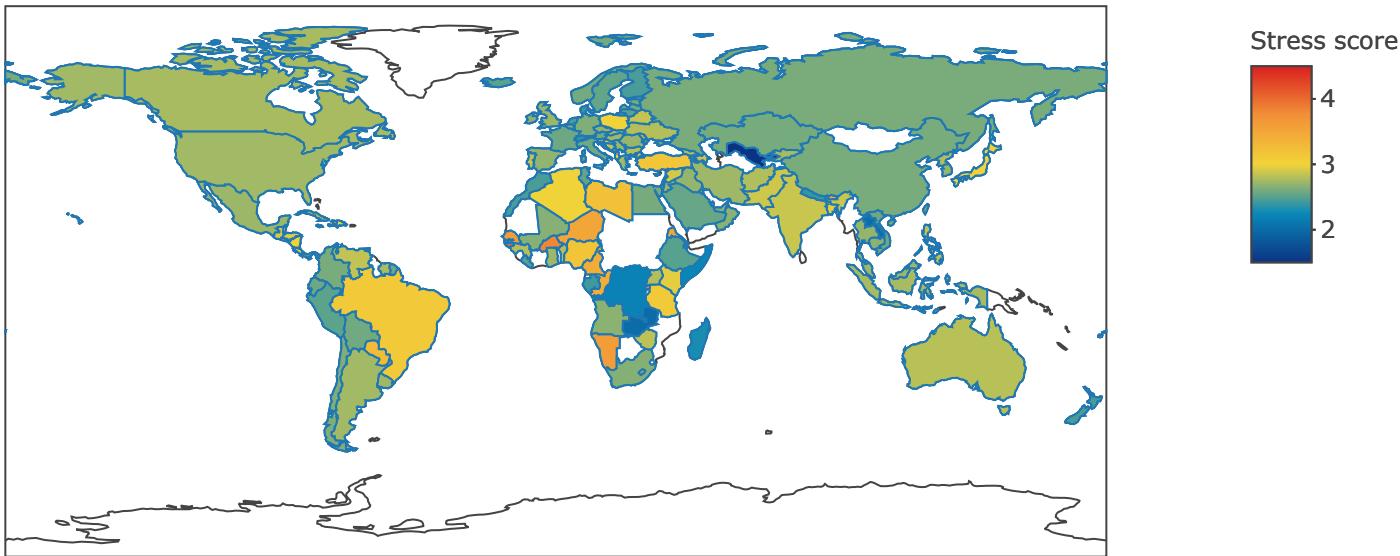


Survey Overview

- The survey was conducted with **86,751** participants. Out of which **64.32%** attended the survey and have answered every question till the end of the survey. Only **8,068** participants have answered just **one** question.
- On an average **75%** of the questions have been answered by all the participants. The average rate of completeness for each individual question is also **75%**.



World map of Corona stress



- The stress severity level is more observed in *South African* countries like **Namibia** and *West African* countries like **Senegal, Guinea-Bissau** and **Burkina Faso**.
 - This helps us to uncover less spoken parts of the world where the people are adversely affected and the media coverage is less.

Analysis by different scales...

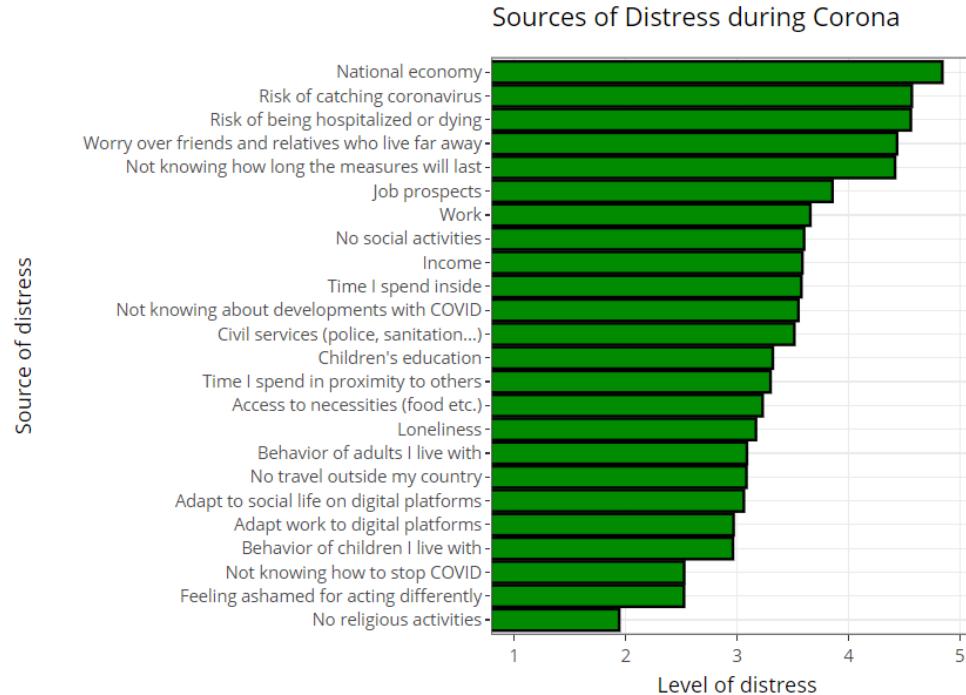
Distress

Trust

Concern

Compliance

Media



People are more stressed due to the fear of economic collapse, catching coronavirus, and the risk of being dying. While not able to perform the religious activities have very little effect on people.

Analysis by different scales...

Distress

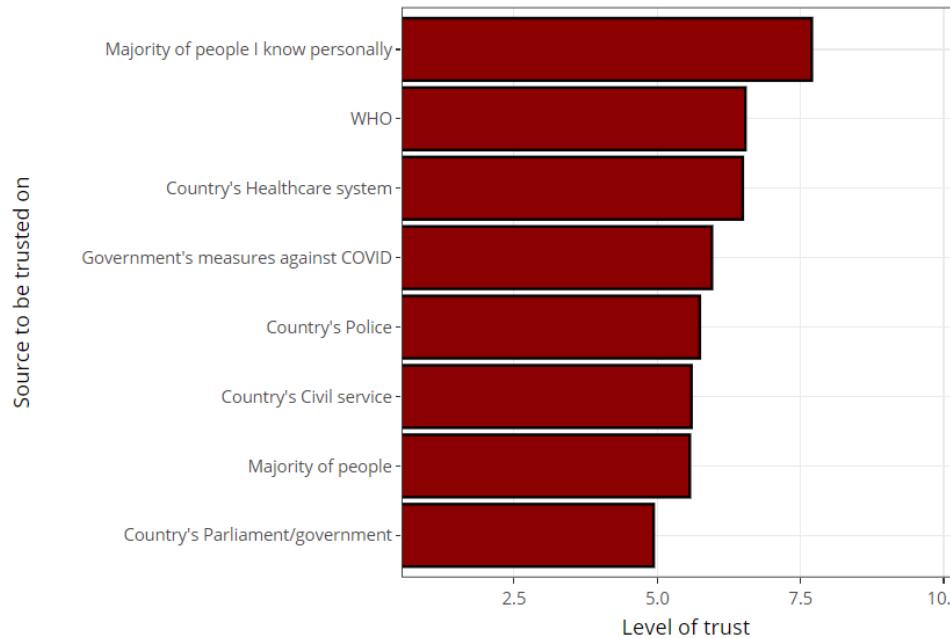
Trust

Concern

Compliance

Media

Trust level on different information sources



Clearly, people trust the information coming from majority of the people they know personally and W.H.O. An interesting finding here is that people mistrust their own country's parliament/government and turn least towards them for information related to the Corona pandemic.

Analysis by different scales...

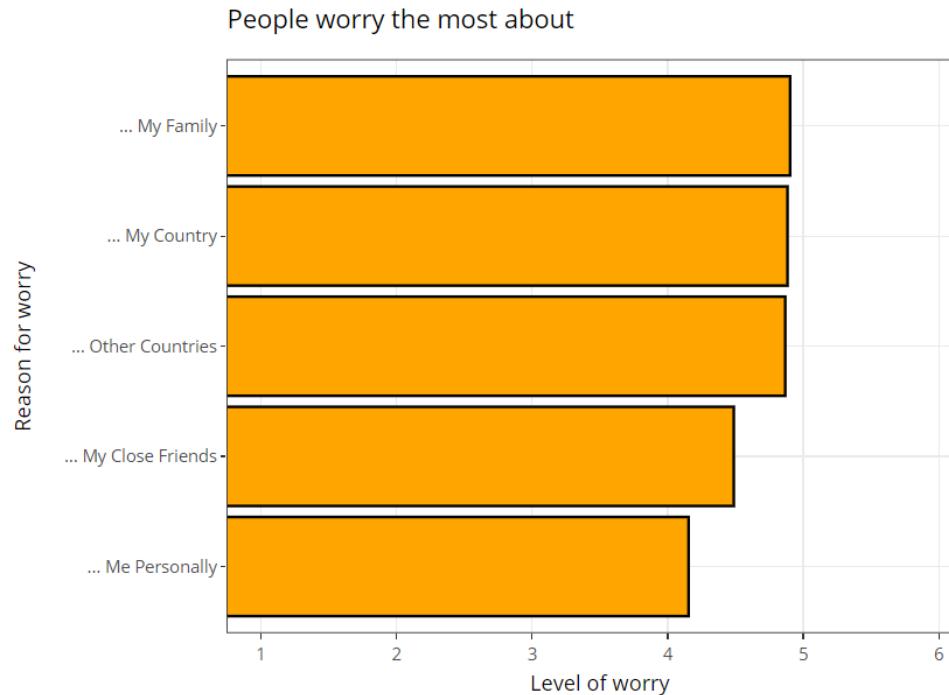
Distress

Trust

Concern

Compliance

Media



People worry most about their family and their country during the corona pandemic. Here an interesting finding is that the people care the least about themselves as compared to their country, friends, and family.

Analysis by different scales...

Distress

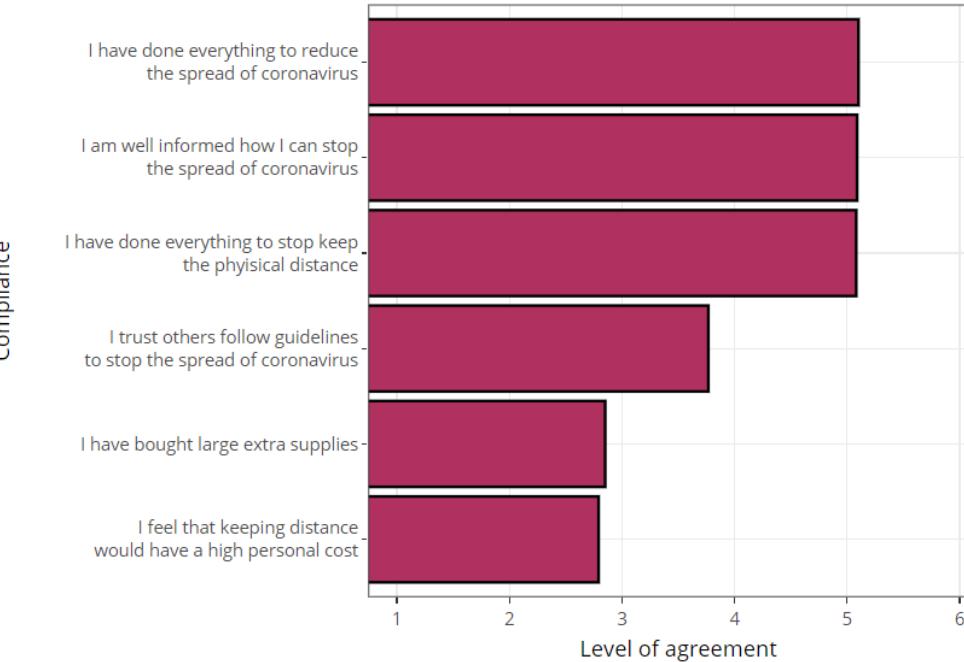
Trust

Concern

Compliance

Media

Obiedience level of instructions



People follow general health and other instructions given to them by different organizations and their countries. This data can be biased because people tend to avoid leaving a negative impression on themselves.

Analysis by different scales...

Distress

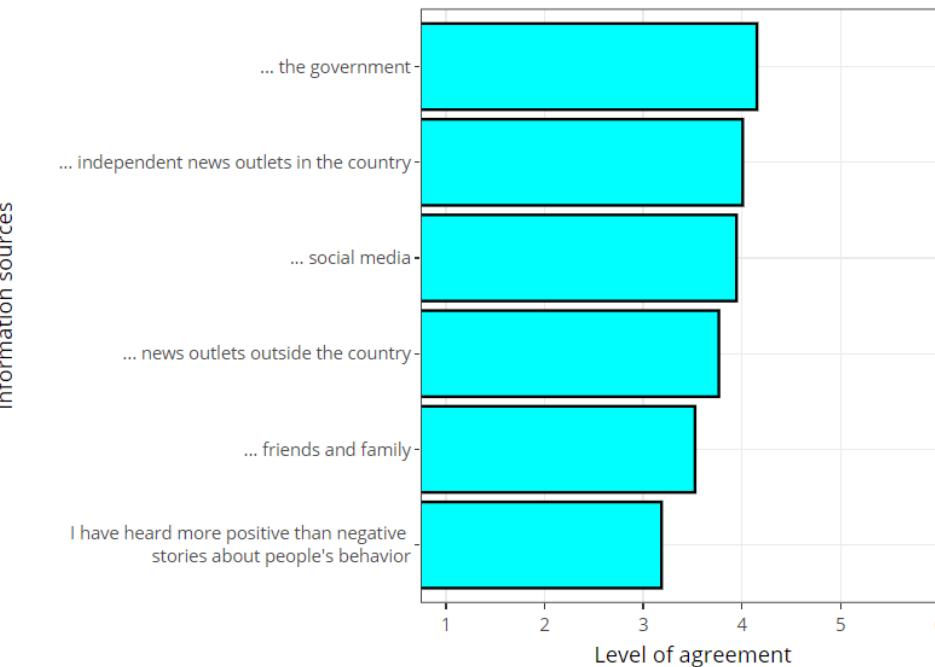
Trust

Concern

Compliance

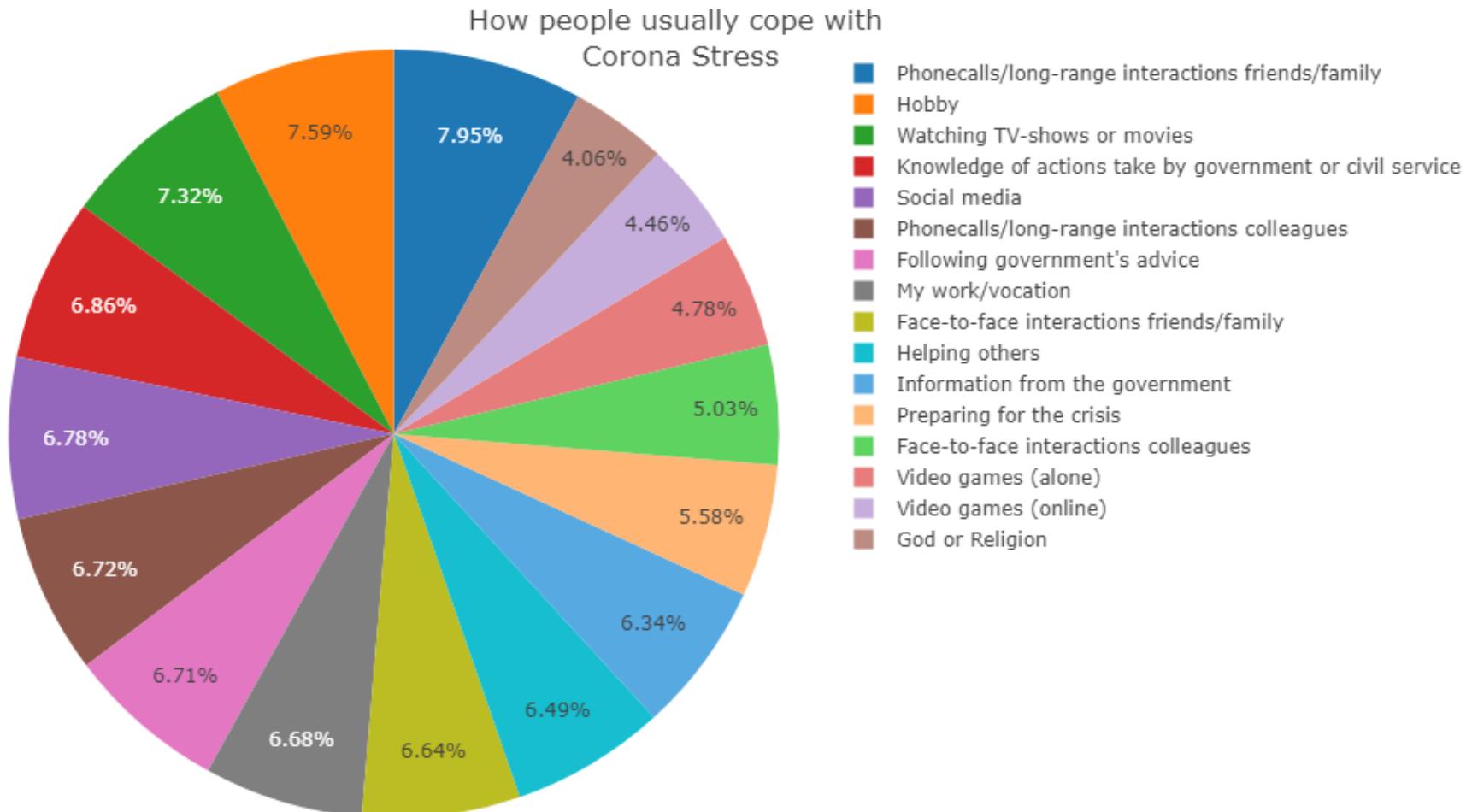
Media

Agreements on information sources



People look mostly for the information regarding corona from their government and the independent news outlets in their country.

Coping with stress



Final survey analysis

Levels of stress were moderate or lower in many countries. **Poland** and **Portugal** reported the highest levels of stress in *Europe*, and **Denmark** and **Netherlands** the lowest. Overall levels of stress remained higher in women compared to men throughout the period under consideration.

Relationship between, *Extraversion*, *Perceived Support*, *Perceived Stress*, *Loneliness* was part of our analysis and it was found that **Loneliness** is the root cause of stress leading to high risk for mental health.

We report that **Loneliness** tends to increase **Perceived Stress** in females and it is about **55.9%** more as compared to male groups.

Extraversion and **Perceived Support** shows a strong correlation of about **25.1%** among male and female groups.

Objective 2: Twitter sentimental analysis



Twitter dataset (Overview)

The Twitter analysis comprises of extracting the tweets related to tags such as `#covid19`, `#coronavirus` and performing a comparative study between the year **2020** and **2021**.

Libraries used: ***rtweet***, ***wordcloud***, ***syuzhet***

For the year 2020, we used the Twitter dataset which was collected for the purpose of research of coronavirus. The dataset can be downloaded here: <https://zenodo.org/record/3831406#.YOGah-j7Q6b/>

For the year 2021, we extracted the latest tweets for the month of June using the ***rtweet*** library.

Followed by extracting the tweets from both years, we performed certain pre-processing techniques such as removing stop words, emojis, cryptic characters, and also text conversion to lowercase to maintain semantic integrity.

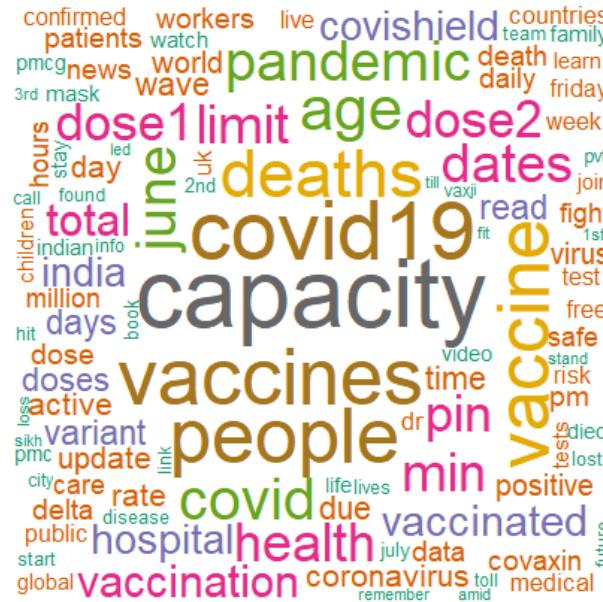
Most frequent and common words

Unigrams-21

Unigrams-20

Bigrams-21

Bigrams-20



It can be seen that in **2021** the greater prominence is given to tweets with hashtags "**capacity**" that appear more frequently in the recent time period.

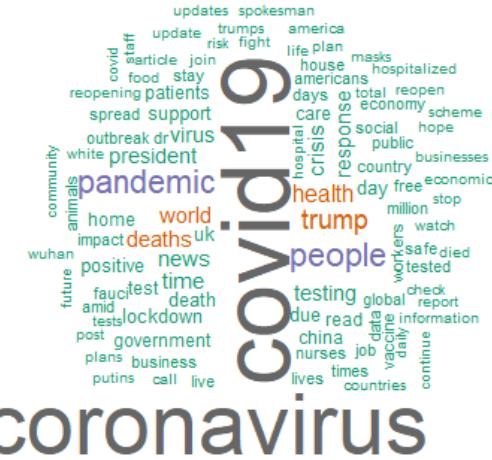
Most frequent and common words

Unigrams-21

Unigrams-20

Bigrams-21

Bigrams-20



It is pretty evident that "**coronavirus**" and "**covid19**" outweigh other tags in the year **2020** along with "**pandemic**" and "**people**" which was trending when the arrival of the pandemic was sensed which remarks the panic globally to a greater extent.

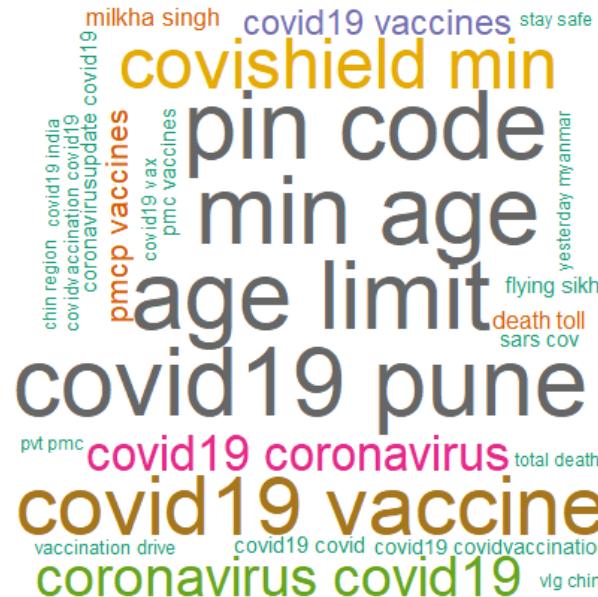
Most frequent and common words

Unigrams-21

Unigrams-20

Bigrams-21

Bigrams-20



It is evident that conjunctive words such as "**min age**", "**age limit**" are more frequent and also two-word sequences like "**covid19 vaccine**", "**covishield min**" also draw attention giving an impression that posts related to topics such as vaccination and vaccines are more common.

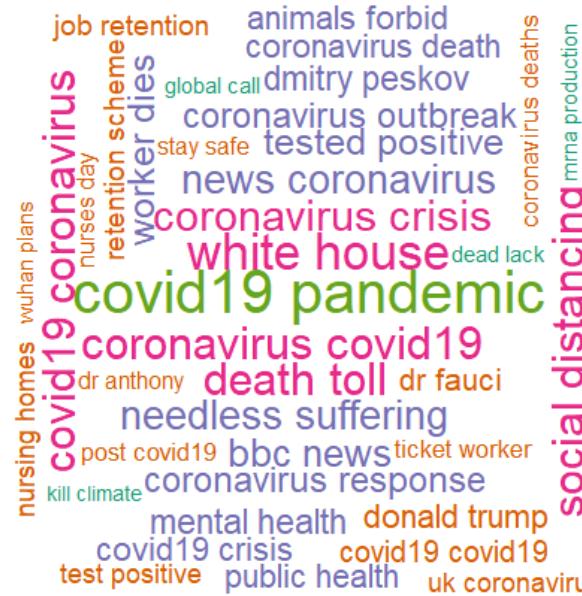
Most frequent and common words

Unigrams-21

Unigrams-20

Bigrams-21

Bigrams-20



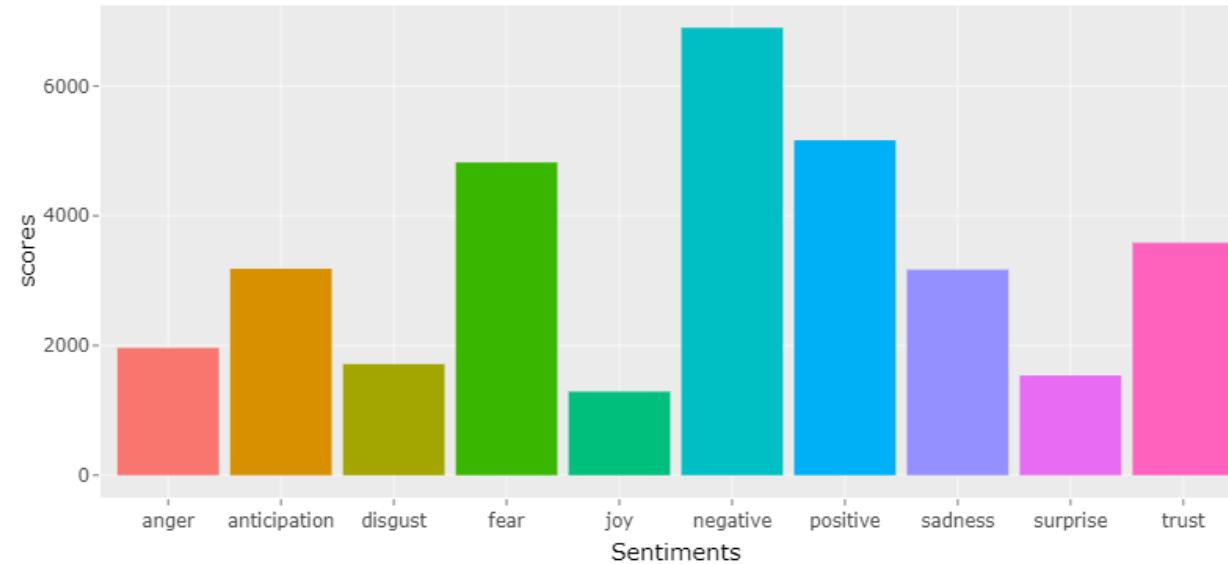
In the latter 2020 when the outbreak was sensed "wuhan coronavirus", "coronavirus outbreak" was more to be noticeable in the twitter community. Also words like **mental health** and **job retention** are also noteworthy.

NRC emotions during pandemic

2020

2021

Emotions of people behind the tweets on COVID19 in 2020 May



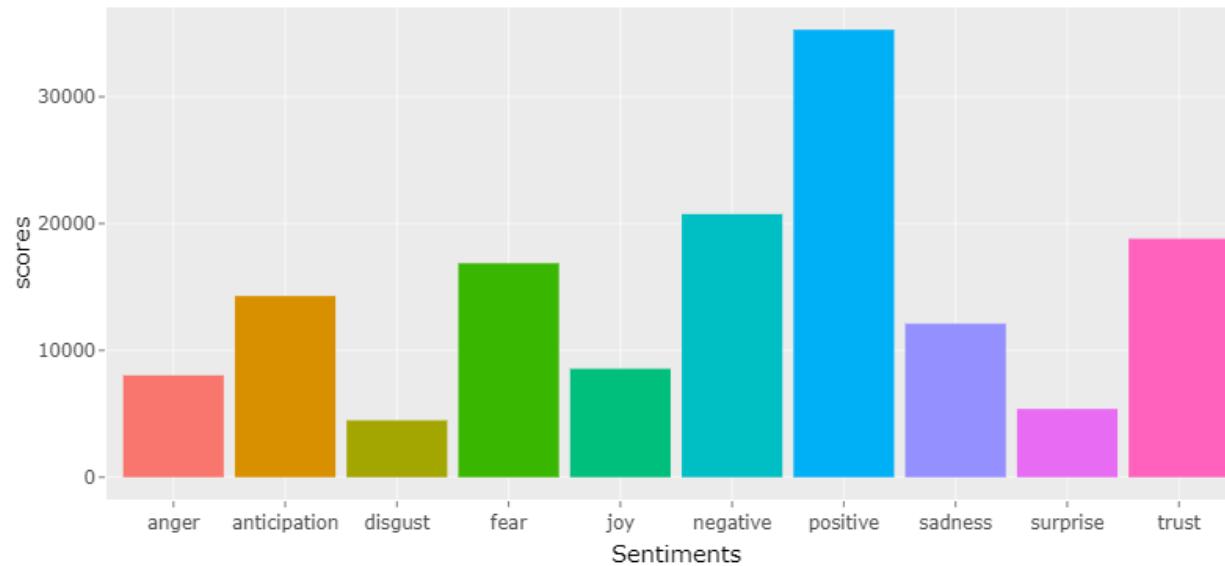
It provides an impression that in 2020, the community expressed a strong negative emotion followed by fear with respect to the amount of tweets that were circulated during the time period.

NRC emotions during pandemic

2020

2021

Emotions of people behind the tweets on COVID19 in 2021 June



We observe that after a year the volume of sentiment scores suddenly increased by 10K folds with the positive emotion being dominant.

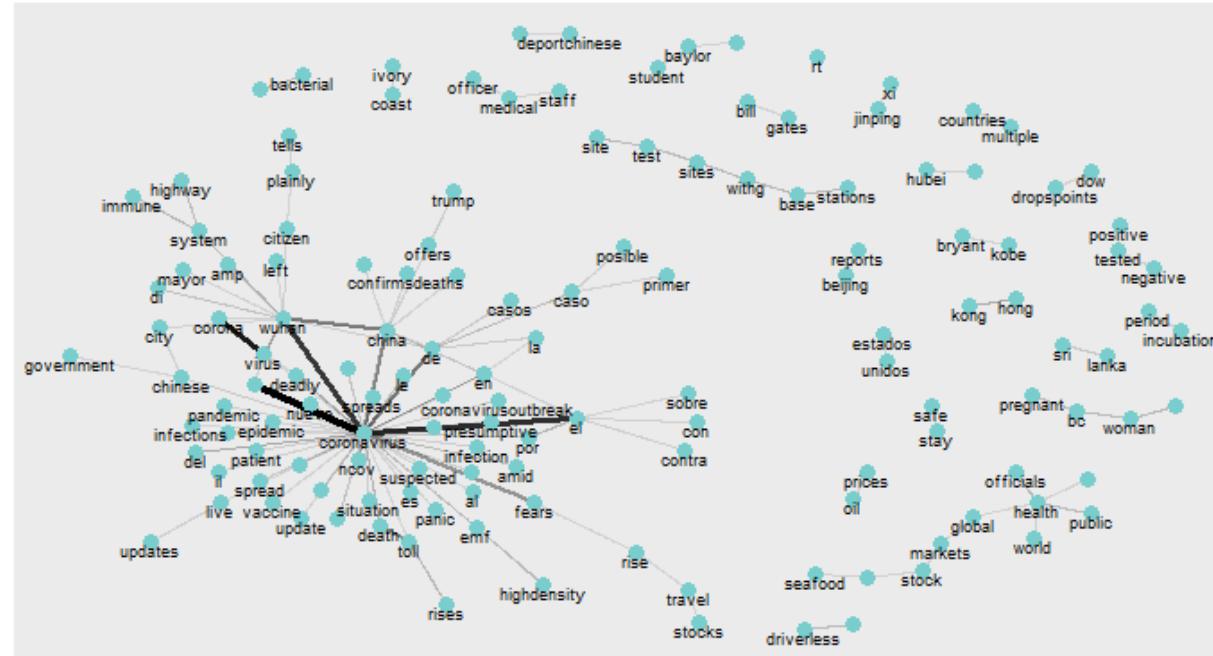
Word network for correlated tags

2020

2021

Word Network: Tweets using coronavirus hashtags

Year 2020



It is evident from the network that words like **wuhan**, **corona**, **deadly**, **spreads**, **outbreak** share very strong association with the word **coronavirus**.

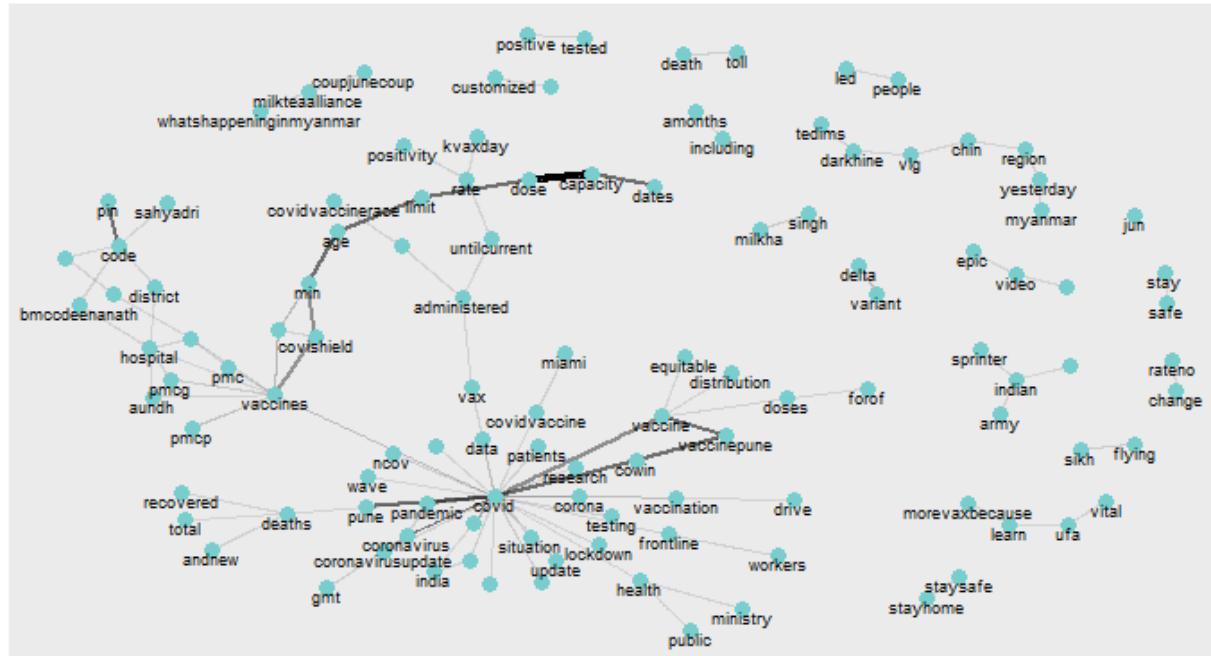
Word network for correlated tags

2020

2021

Word Network: Tweets using coronavirus hashtags

Year 2021



In 2021, **Dose** and **Capacity** have been the most commonly occurring words which tells us how important the number of doses and capacity in hospitals has been an issue throughout the world.

Final Twitter analysis

- We make use of **35,725** tweets dated from *June 17, 2021* to *June 19, 2021* and about **15,000** tweets in English from last year *June* for our analysis.
- In 2020, people were more panicked about the coronavirus outbreak. While in 2021, the people went more talking about getting vaccinated.
- Words such as **crisis**, **virus**, and **death** contributed to developing more negativity in **2020**. Whereas words like **support**, **safe** and **free** helped to keep the proportion of positive words in **2021** much higher than in **2020**.
- Overall, we can conclude that we see a more definitive pattern in **2020** that caused more distress in people than **2021**.

Objective 3: Infodemics



Infodemics Observatory data (Overview)

- Data from **COVID-19 Infodemics Observatory** <https://covid19obs.fbk.eu/#/>, comprises of summary of infodemics data collected from various countries, that indicates the world risk index, population emotional state, and news reliability.

```
##           date iso3 TWI_VOLUME EPI_CONFIRMED IRI_UNVERIFIED IRI_VERIFIED   IRI_ALL
## 7934 2020-03-06 ZWE      1561            0     0.6144987       0 0.6144987
## 7935 2020-03-07 ZWE      1197            0     0.1297782       0 0.1297782
## 7936 2020-03-08 ZWE      1685            0     0.2169017       0 0.2169017
## 7937 2020-03-09 ZWE      1686            0     0.1177609       0 0.1177609
## 7938 2020-03-10 ZWE      1445            0     0.4692687       0 0.4692687
```

- "*Infodemics Reduced*" has the `iso3` country code and the volume of the infodemic score spread across continents which we find is very vital for extracting specific insights to get an idea of how the IRI has evolved across the timeline.

```
##           date world_risk_index
## 1 2020-01-22          0.40613
```

- "*World Risk Index*" has the world risk index for each day for 2020. Shows the indices ranging from 0 (Lowest Risk) to 1 (Highest Risk), with intermediate cases.

IRI analysis by countries

Italy

USA

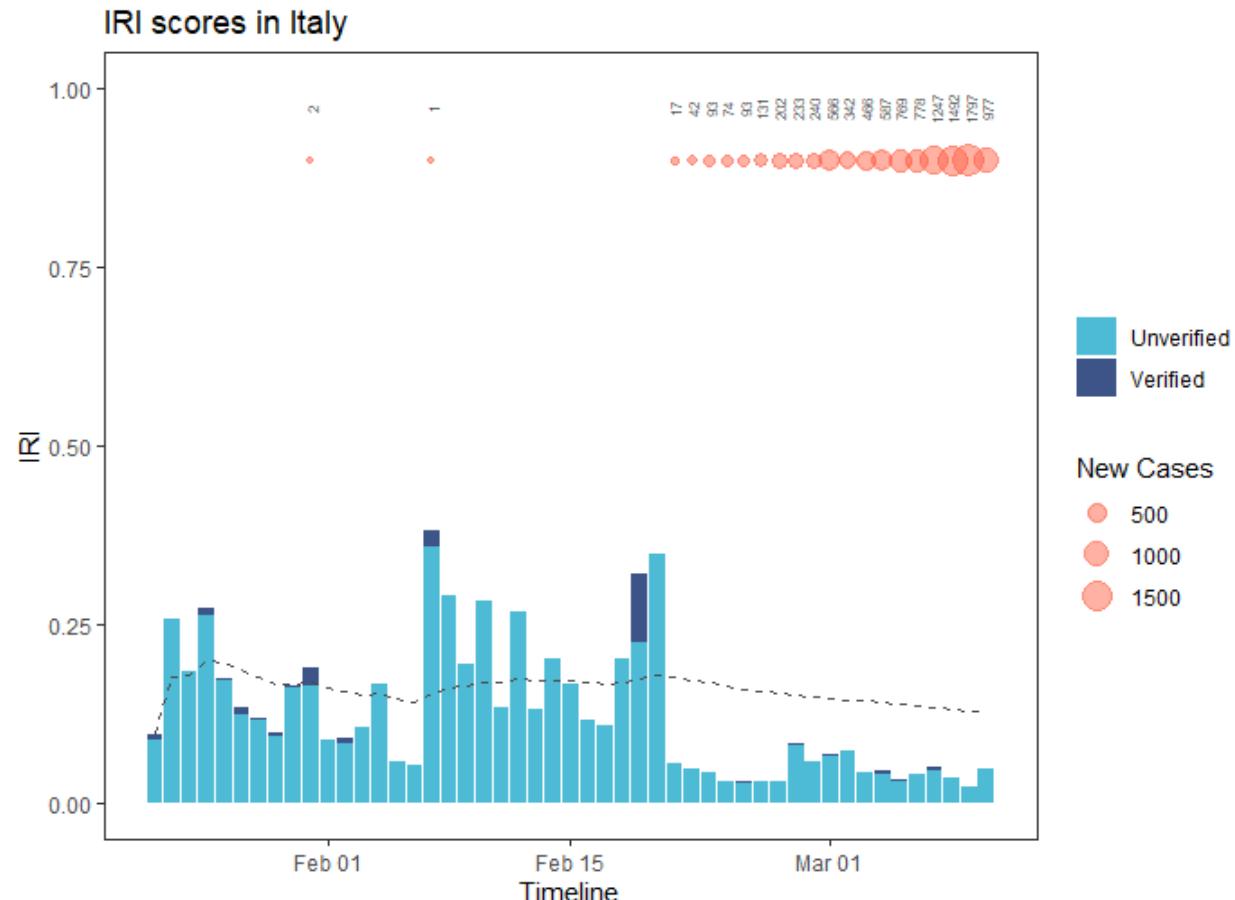
Venezuela

Russia

South Korea

Canada

Continent



IRI analysis by countries

Italy

USA

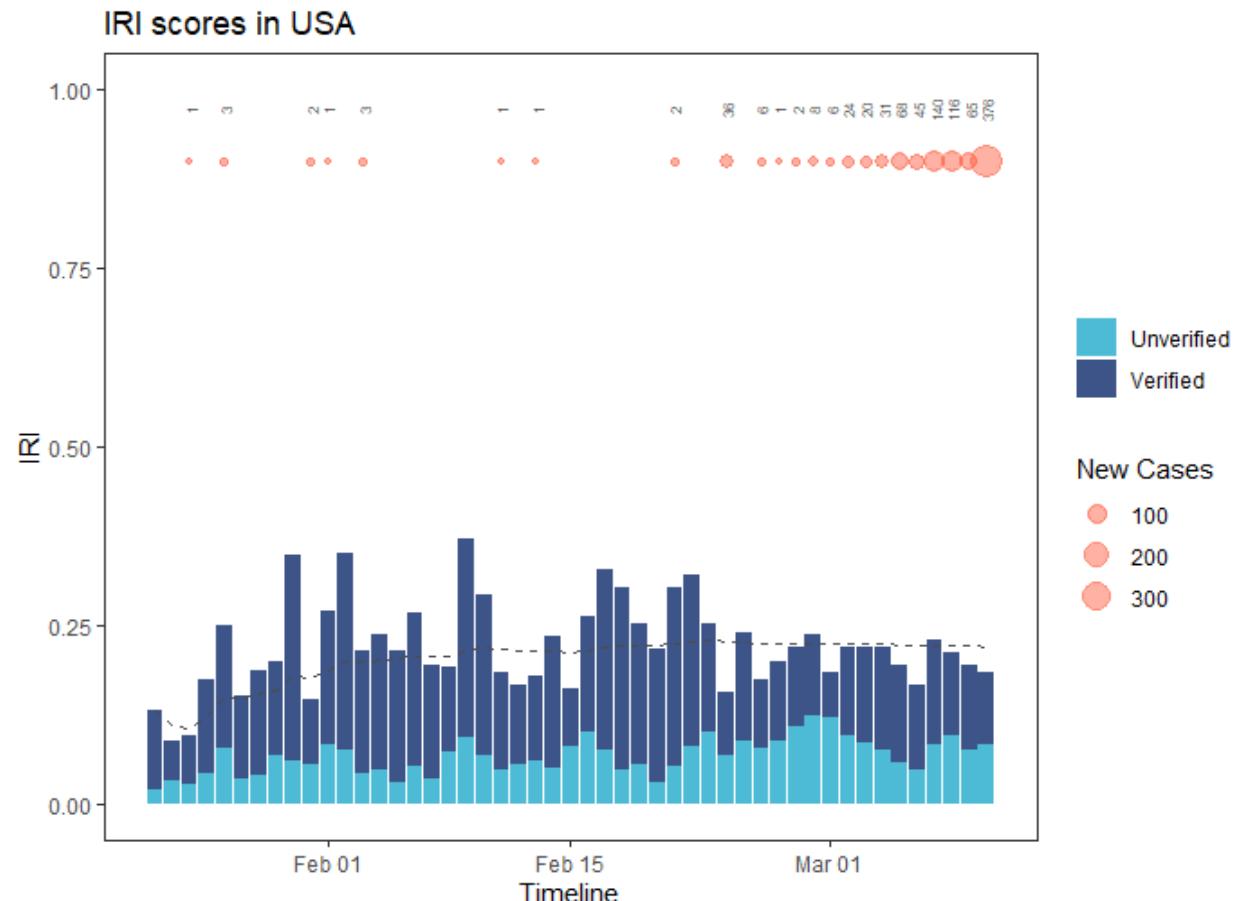
Venezuela

Russia

South Korea

Canada

Continent



IRI analysis by countries

Italy

USA

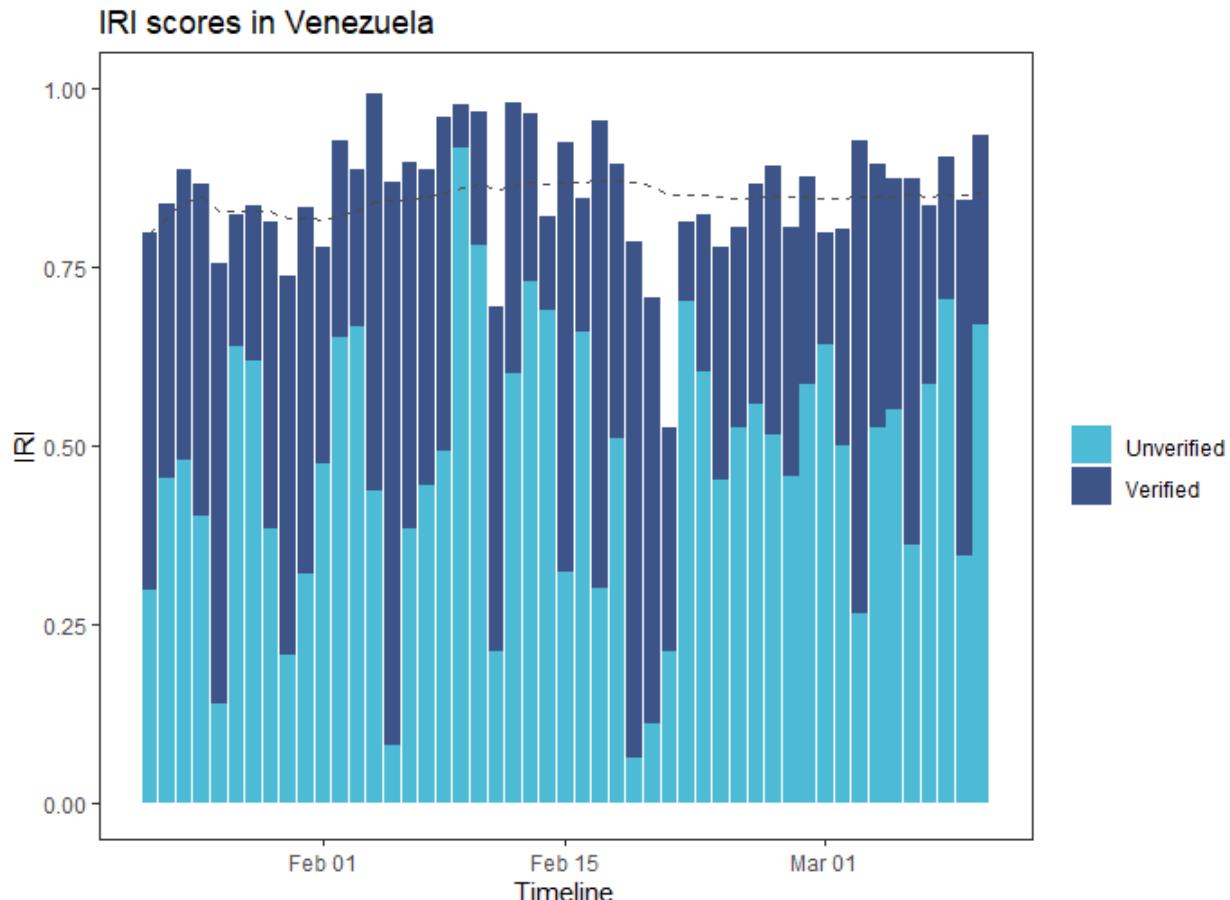
Venezuela

Russia

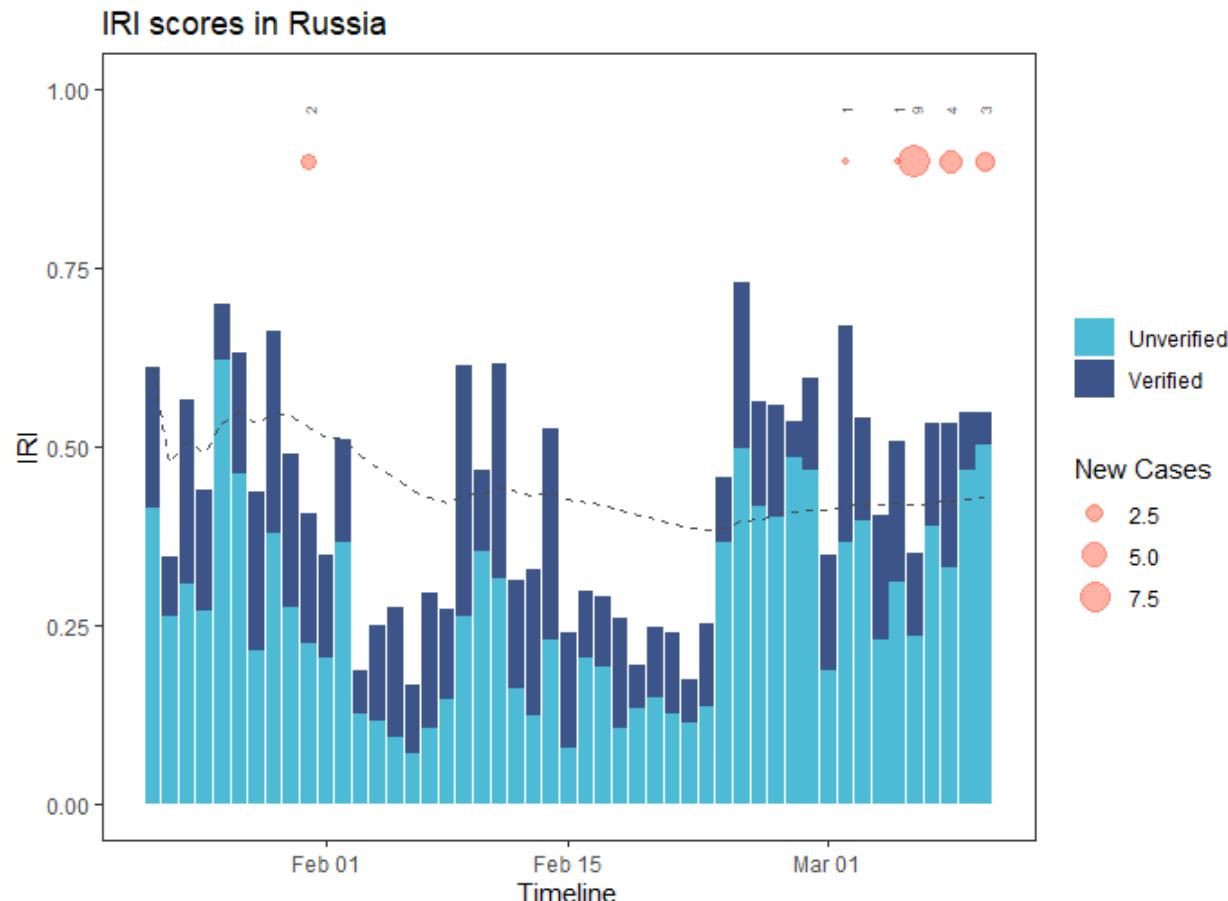
South Korea

Canada

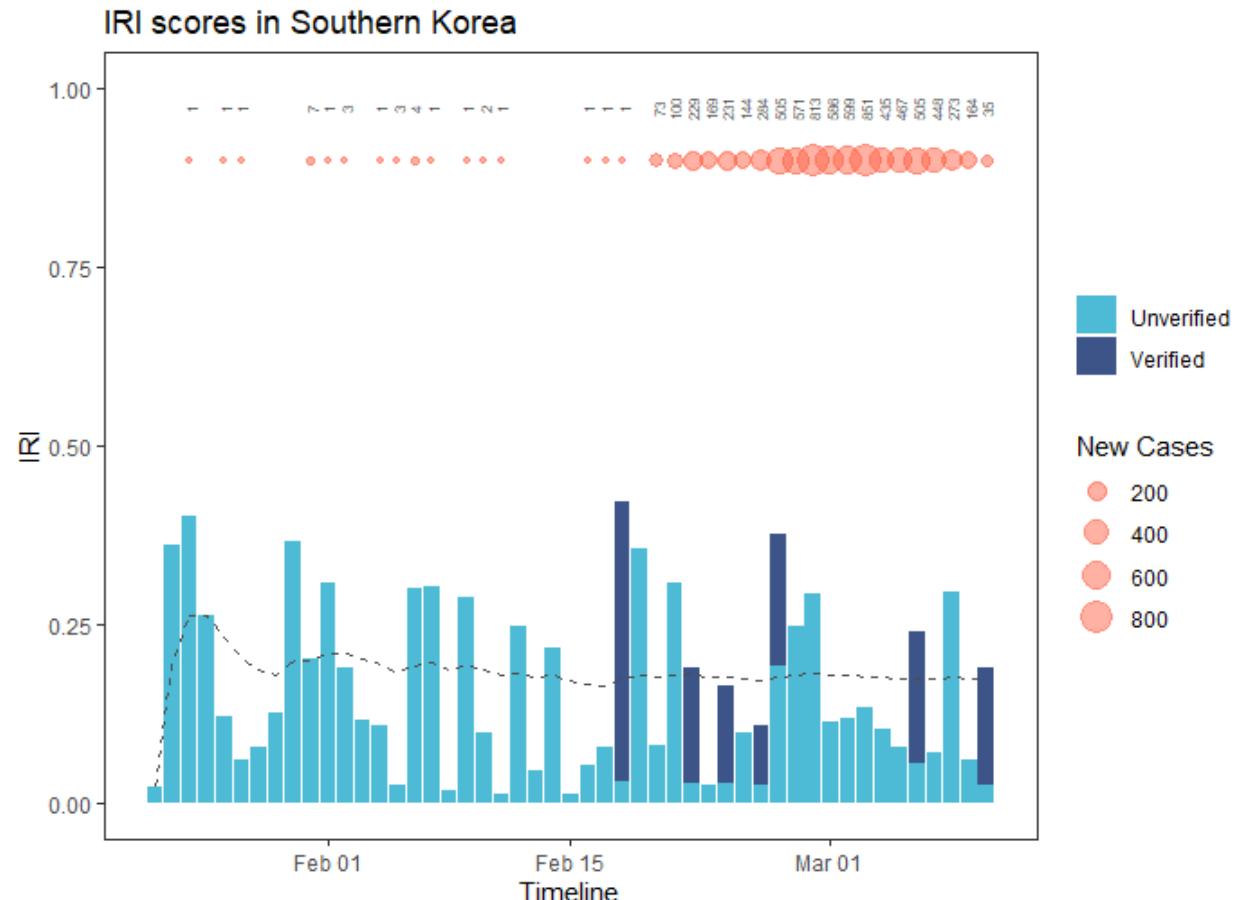
Continent



IRI analysis by countries

[Italy](#)[USA](#)[Venezuela](#)[Russia](#)[South Korea](#)[Canada](#)[Continent](#)

IRI analysis by countries

[Italy](#)[USA](#)[Venezuela](#)[Russia](#)[South Korea](#)[Canada](#)[Continent](#)

IRI analysis by countries

Italy

USA

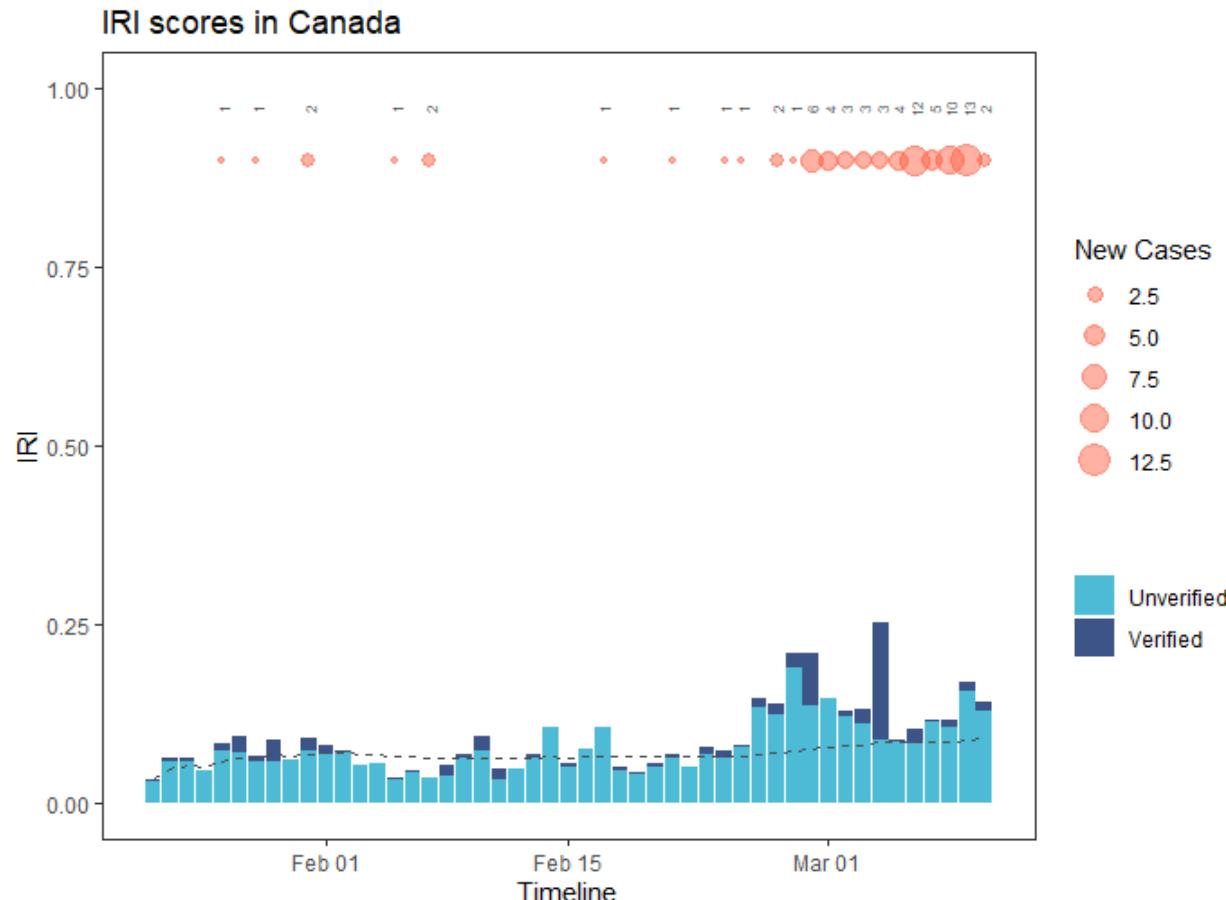
Venezuela

Russia

South Korea

Canada

Continent



IRI analysis by countries

Italy

USA

Venezuela

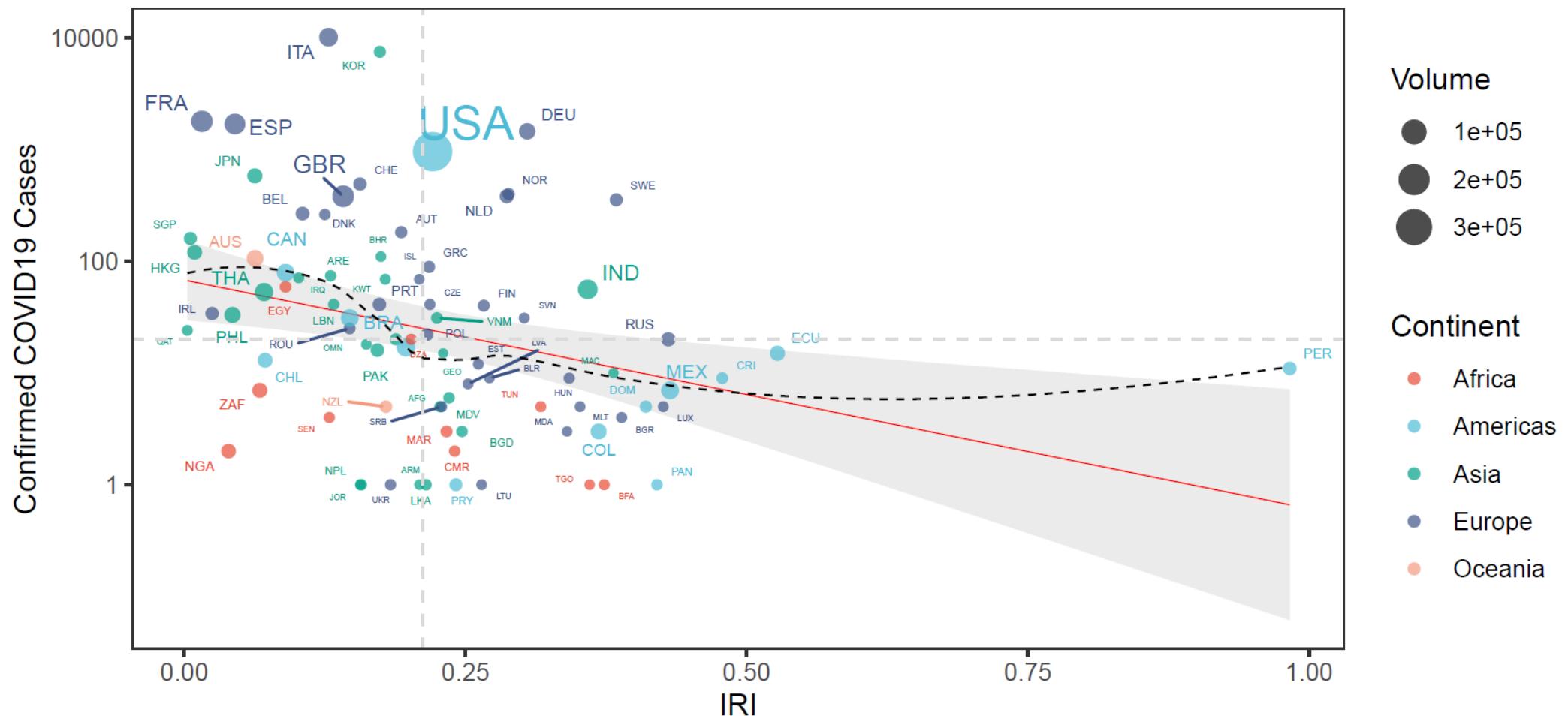
Russia

South Korea

Canada

Continent

Showing confirmed COVID-19 cases across countries and the IRI score



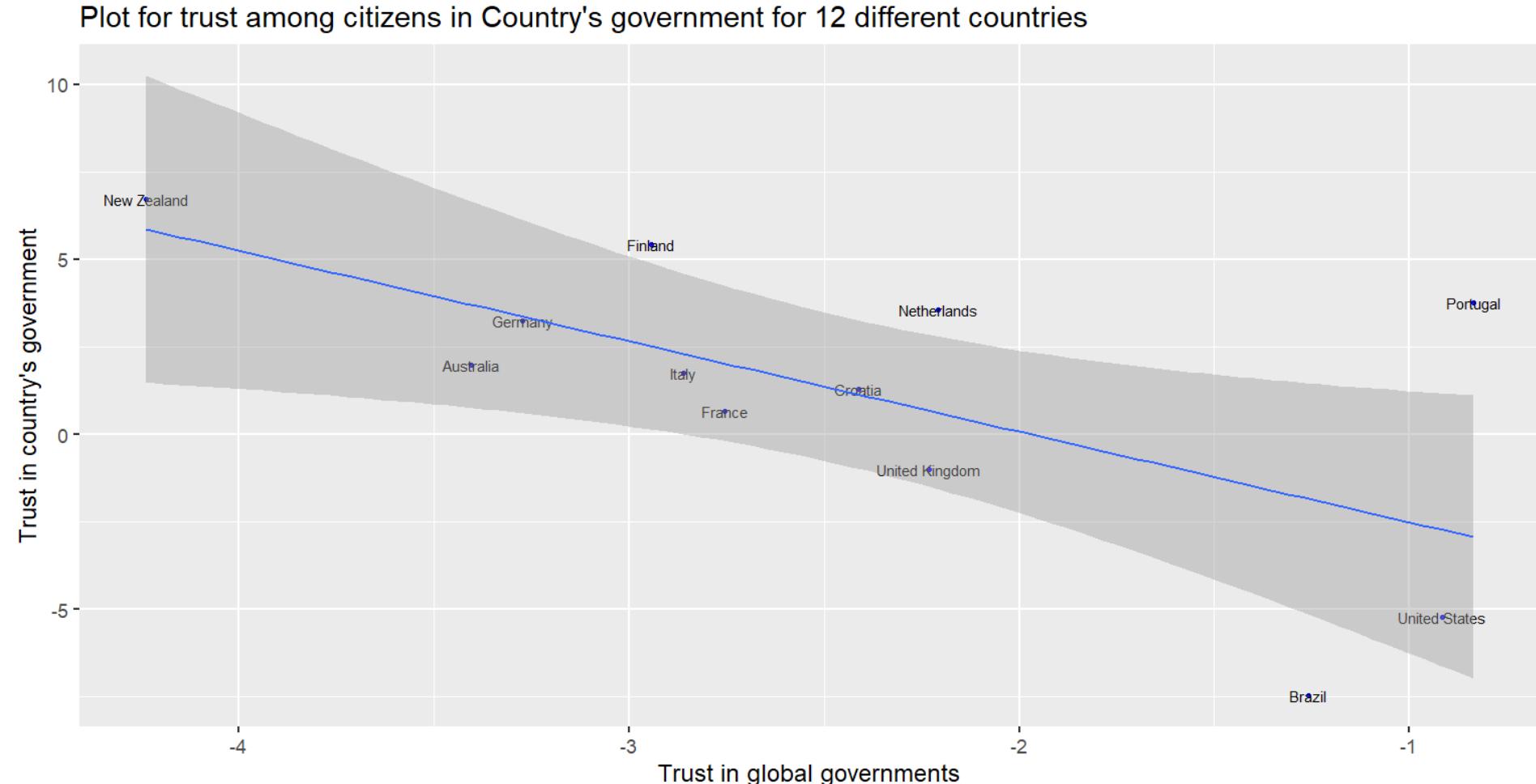
Trust scores in...



Government

Scientist

Social Media



Trust scores in...

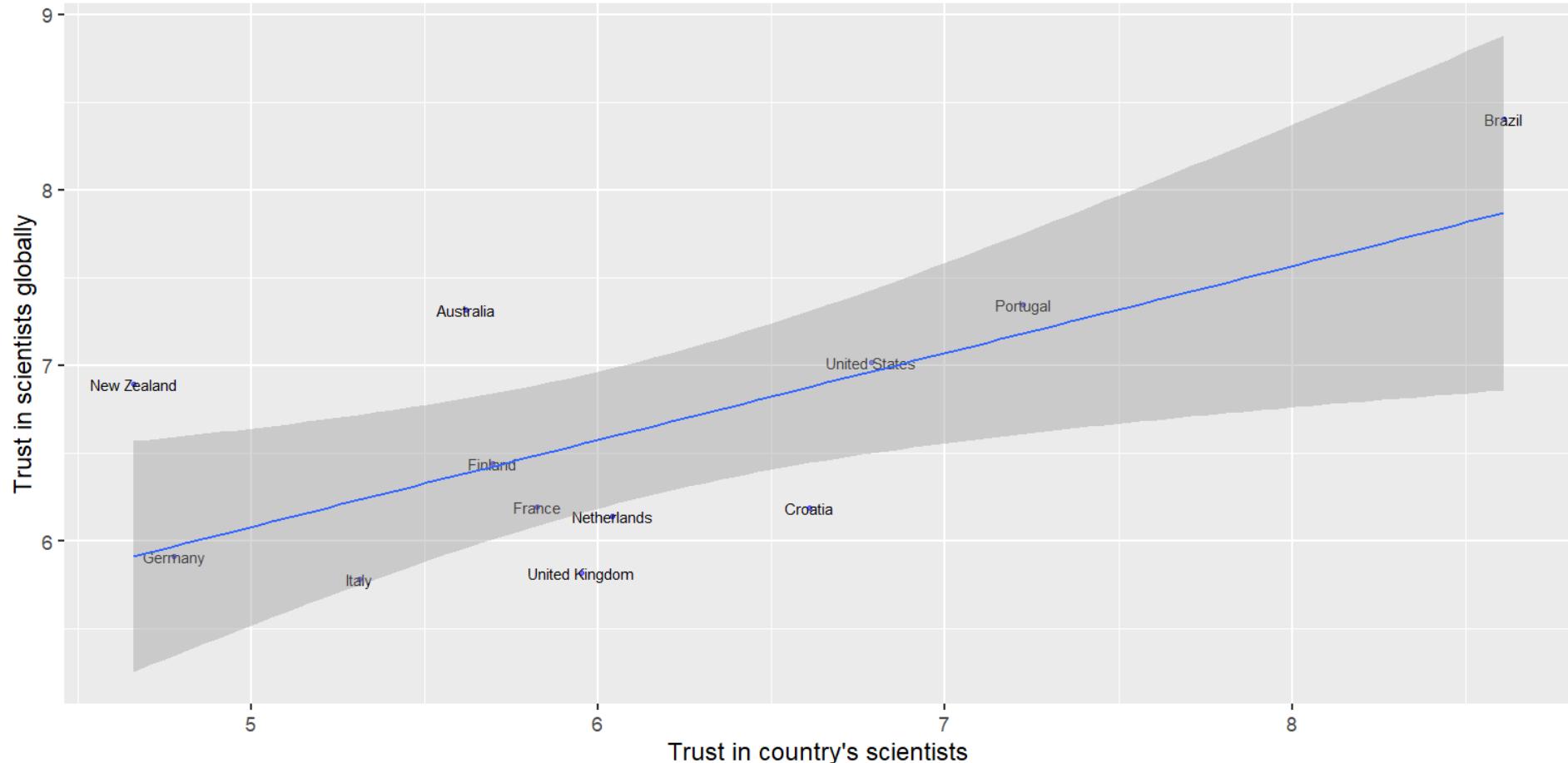


Government

Scientist

Social Media

Plot for trust among citizens in Country's scientists for 12 different countries



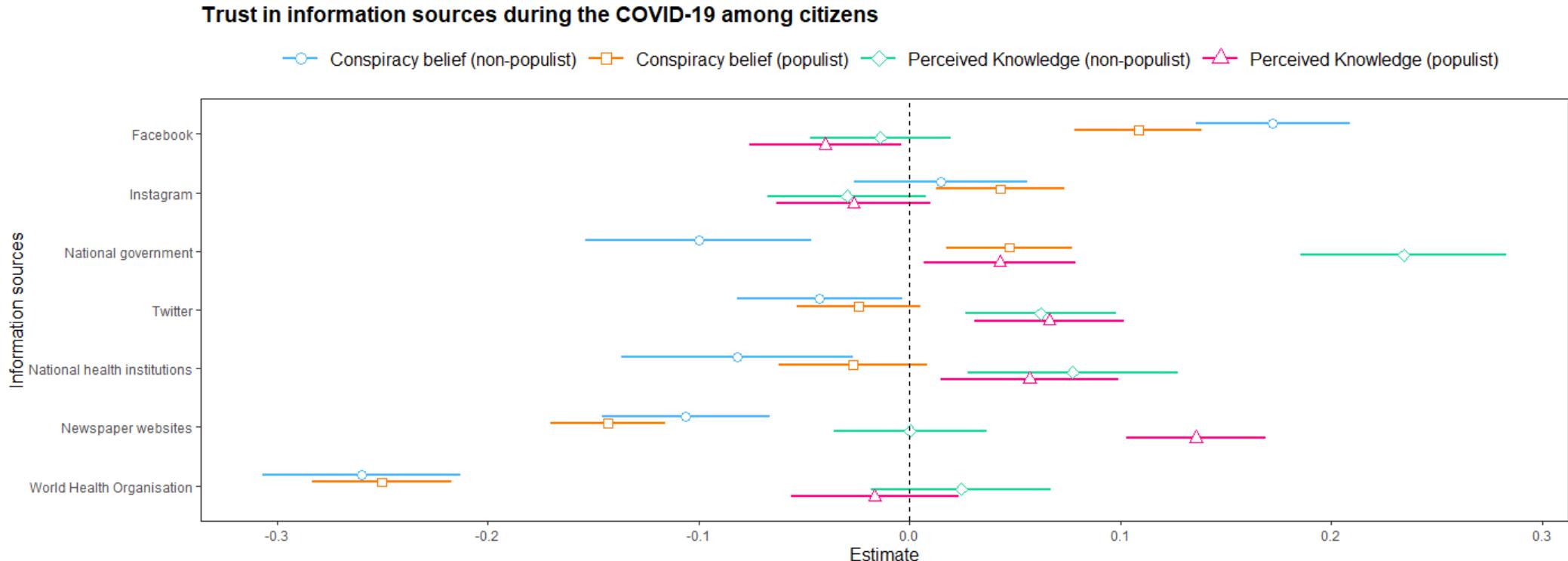
Trust scores in...



Government

Scientist

Social Media



Citizens became more conscious about COVID-19 from **National government** with an estimate score of [0.17, 0.27].

"We need to flatten this Infodemic curve"

- Just as we need to flatten the COVID-19 curve we must also tackle the infodemic curve. As with Covid-19 we must attack the curve on two fronts (suppress the contagion and increase our capacity to deal with the surge of information that is coming our way).
- From our analysis, for relationship between "virus naturally occurring or not" and "media", the true correlation is around **13%** which means media plays a key role for overhyping on this question.
- Also, "virus naturally occurring or not" and "fake news" has the least correlation with **3.6%** it is difficult to verify the trueness of this question as it is very subjective.
- We performed an analysis on how informed the citizens are feeling about the general news about COVID-19 with respect to **fake news** and **media hype**.
 - For the former a *positive* correlation is observed meaning an average number of people are convinced by fake news over social media.
 - Whereas, for the latter a *negative* correlation is observed meaning an average number of people are likely to be less informed by the media tacit.

"We need to flatten this Infodemic curve"

- Just as we need to flatten the COVID-19 curve we must also tackle the infodemic curve. As with Covid-19 we must attack the curve on two fronts (suppress the contagion and increase our capacity to deal with the surge of information that is coming our way).
- From our analysis, for relationship between "virus naturally occurring or not" and "media", the true correlation is around **13%** which means media plays a key role for overhyping on this question.
- Also, "virus naturally occurring or not" and "fake news" has the least correlation with **3.6%** it is difficult to verify the trueness of this question as it is very subjective.
- We performed an analysis on how informed the citizens are feeling about the general news about COVID-19 with respect to **fake news** and **media hype**.
 - For the former a *positive* correlation is observed meaning an average number of people are convinced by fake news over social media.
 - Whereas, for the latter a *negative* correlation is observed meaning an average number of people are likely to be less informed by the media tacit.

Fact check to alert yourself what is currently going around, Stick to trusted sources, Do not forward without checking the authenticity of messages, Increase supply of data by engaging regularly and meaningfully on the platforms that people are already using, In short: Be an informed digital citizen.

References (Slides reused from course materials)

- [1] Andreas Lieberoth, Sabrina Stoeckli, Jesper Rasmussen. 2020. "COVIDiSTRESS Global Survey Network (2020 , March 30)". <https://doi.org/10.17605/OSF.IO/Z39US>.
- [2] Cohen, Kamarck, S. 1983. "Journal of Health and Social Behavior, 24(4), 385-396. A Global Measure of Perceived Stress." <https://doi.org/10.2307/2136404>.
- [3] Matošević, Goran, and Vanja Bevanda. 2020. "Sentiment Analysis of Tweets about COVID-19 Disease During Pandemic," 1290–95. <https://doi.org/10.23919/MIPRO48935.2020.9245176>.
- [4] Mulukom, Valerie van. 2021. "The Role of Trust and Information During the COVID-19 Pandemic and Infodemic," May. <https://doi.org/10.17605/OSF.IO/GFWBQ>.
- [5] R. Gallotti, F. Valle, N. Castaldo, and M. De Domenico. 2021. "Covid19 Infodemics Observatory (2020)," May. <https://doi.org/10.17605/OSF.IO/N6UPX>.
- [6] Rul, Céline Van den. n.d. "A Guide to Mining and Analysing Tweets with R." <https://towardsdatascience.com/a-guide-to-mining-and-analysing-tweets-with-r-2f56818fdd16>.
- [7] Travaglino, Lieberoth, G. A. 2020. "How Is Covid19 Affecting Europeans' Lives?. Report of the COVIDiSTRESS Global Survey." <https://doi.org/10.13140/RG.2.2.30558.59209>.
- [8] Vijay, Tanmay, Ayan Chawla, Balan Dhanka, and Purnendu Karmakar. 2020. "Sentiment Analysis on COVID-19 Twitter Data," 1–7. <https://doi.org/10.1109/ICRAIE51050.2020.9358301>.



Thank you! Questions?