
FACULTY OF
COMPUTER SCIENCE

Data Science with R - Project Proposal

# "BEHAVIOURAL & PSYCHOLOGICAL DISTRESS OF COVID-19 AND INFODEMICS"

Madhuri Sajith, Usama Ashfaq, Vishnu Jayanand, Sujith Sudhakaran, Ranjiraj Nair

May 18th, 2020


FACULTY OF
COMPUTER SCIENCE

# DATA SCIENCE WITH R - PROJECT PROPOSAL

## TEAM MEMBERS


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


# BEHAVIOURAL & PSYCHOLOGICAL DISTRESS OF COVID-19 AND INFODEMICS

### BACKGROUND AND MOTIVATION

The coronavirus COVID-19 pandemic is an unprecedented health crisis that has impacted the world to a large extent. According to WHO, mental disorders are one of the leading causes of disability worldwide, so considering that this pandemic has caused further complications to mental ailment. The stress, anxiety, depression stemming from fear, isolation, and stigma around COVID-19 affected all of us in one way or another. We could see that many people are losing their jobs and the elderly are isolated from their usual support network. The issue is that the effects of the pandemic and mental health, maybe for longer-lasting than the disease itself.

In this limelight, although the measures are taken to slow the spread of the virus, it has affected our physical activity levels, our eating behaviors, our sleep patterns, and our relationship with addictive substances, including social media. Into this last point, both our increased use of social media while stuck at home, as well as the increased exposure to disaster news past year, have amplified the negative effects of social media on our mental health. This motivates us to perform some diagnostic analysis of this pattern and portray some meaningful insights on global survey data.

### PROJECT OBJECTIVES

**Objective 1:** Here we aim to analyze and visualize the survey dataset to come up with certain descriptive and diagnostic statistics including the number of countries that participated in the survey. We look forward to analyzing the various stress levels from each country, with a focus on visualizing the distress scale with the source of distress and the level of distress. Further, we plan to visualize the coping scale which consists of the source of coping and detailing what was the personal conscious effort, to solve personal and interpersonal problems, to master, minimize or tolerate stress and conflict. We also target to depict the visualization of the level of trust such as Country's Civil service, Country's Police, Country's Healthcare system, WHO, Government's measures against COVID which also play a crucial role. We extend to diagnose on areas such as compliance and the level of agreement, the concern level, and the key factor of media from where

the respondents sought to agree to take the information from. The different relationship between perceived stress, social support, loneliness, and extroversion according to different age groups will also be set as our basis for analysis.

**Objective 2:** We propose to know the impact of COVID-19 tackling infodemics and misinformation on Twitter. This is done by extracting recent popular tweets from a specific location across different countries. It will help us describe the false information that is spread with the sole purpose of causing confusion and harm. We target to extract hashtags like #covid19, #misinformation, #fakenews, #disinformation, #, etc., to get the related posts about it and analyze how the information processing and decision-making behaviors are compromised. We perform sentimental analysis on the tweets to understand the sentiments of people which is crucial during the time of this pandemic.

**Objective 3:** Our final key interest is to perform a comparative analysis on `Infodemics`. That is to outline the interaction patterns of fake news information spreading across media. The news which is rolled out in different environments having different interaction settings and audiences are verified. We report the rumor amplification parameters for each of the social media platform due to COVID-19 to show how the misinformation is spread on different mainstream online platforms.

## DATASETS

**1. COVIDISTRESS all global survey data**
(The COVIDiSTRESS global survey is an open science collaboration, created by researchers in over 40 countries to rapidly and organically collect data on human experiences of the Coronavirus epidemic 2020.) Dataset can be downloaded here: [COVIDiSTRESS global survey network (2020, March 30). COVIDiSTRESS global survey. DOI 10.17605/OSF.IO/Z39US, Retrieved from osf.io/z39us] *https://osf.io/z39us/*

These datasets mainly focus on the stress levels, sources of stress, and trust in institutions across the EU. Furthermore, it also includes factors such as loneliness, media use, personality, social provisions, and perceived sources of psychological relief.

**2. Twitter Data**
We aim to work on the most recent dataset aggregated from Twitter using twitteR and rtweet libraries within a particular time and location.

Here `twitteR` which provides an interface and access to Twitter web API respectively, `rtweet` which acts as the client for Twitter's REST and stream APIs will be used to retrieve data.

**3. COVID19 Infodemics Observatory**
(The Role of Trust and Information during the COVID-19 Pandemic and Infodemic) Dataset can be downloaded here: [R. Gallotti, N. Castaldo, F. Valle, P. Sacco and M. De Domenico, COVID19 Infodemics Observatory (2020). DOI: 10.17605/OSF.IO/N6UPX] [Van Mulukom, V. (2021, May 15). The Role of Trust and Information during the COVID-19 Pandemic and Infodemic. https://doi.org/10.17605/OSF.IO/GFWBQ] *https://osf.io/n6upx/*, *https://osf.io/67zhg/*, *https://osf.io/rtacb/*, *https://osf.io/dh879/*, *https://osf.io/c37wq/*

These datasets comprises of summary of infodemics data collected from across countries, the world risk index, population emotional state, and news reliability.

## DESIGN OVERVIEW

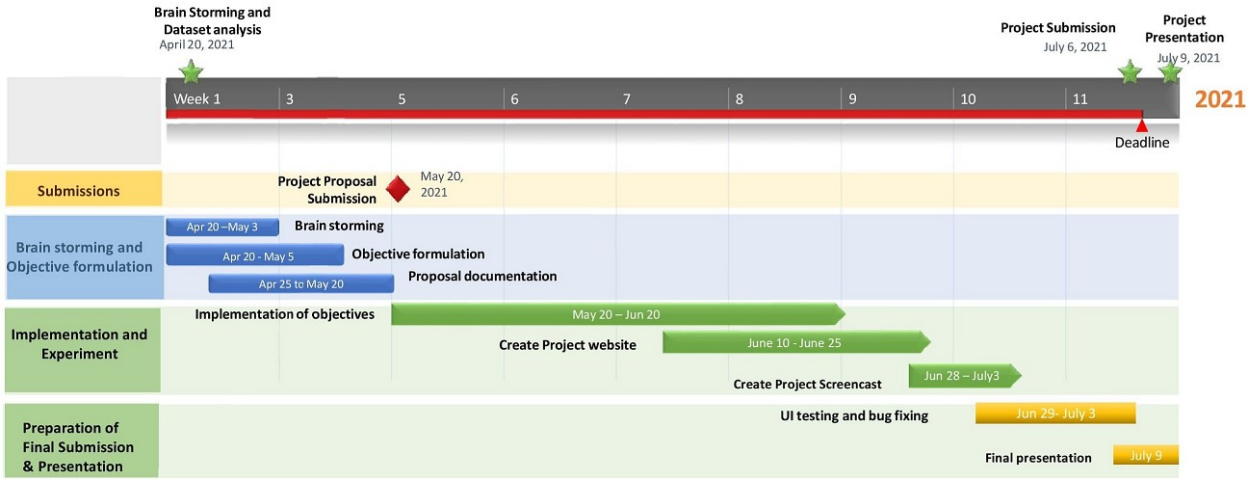
**Objective 1** We target to access the survey data using the `qualtRics` library for understanding our data and for building a first impression of the data at hand. For continuous recoded variables such as level of distress, coping, trust, concern, agreement, etc. we use frequency histograms from `ggplot2`. To see how the variables are correlated we use bivariate plots which will be used to visualize the relationship between perceived stress, social support, loneliness, and extraversion. Bar plots will be used to see the distributions of age, gender, and education level. We will use geo-visualization plots from `plotly` and `maps` to show the isolation, stress, trust, and concern scores across different regions of the world. Moreover, to visualize the country-wise stress levels, box plots will be used to do a comparative analysis with the help of quartiles. Also, with the help of `psych` package, we aim to analyze the data at multiple levels within and between-group statistics, including correlations and provide some basic descriptive statistics.

**Objective 2** Using `twitteR`/`rtweet` we carry out sentimental analysis on the retrieved tweets using respective API to analyze how citizens have been impacted all over the world and plot the trends from the specific locations and from the exact time-frame to produce word clouds.

**Objective 3** First, we calculate the Infodemic Risk Index (IRI) using basic descriptive statistics and calculate the cumulative mean. Then using `ggplot2` we depict the cumulative number of reported cases. The IRI scores from a specific time showing the trend and how significantly it changes accordingly with time will be displayed. We plot the histograms for various countries to highlight the unverified and verified scores along with the increase in the number of new cases within the same plot. Moreover, using the `map` package for a specific continent the IRI for low to high with color gradient scale will also be visualized. A comparative analysis of IRI concerning the number of confirmed COVID19 cases with the volume for each country in the continent will be visualized. A Scatter plot will be shown with some critical information like the trust in the country's government, scientists, and other government officials. We also perform a regression analysis using the `lme4` package to estimate the trust in both online and offline media which can be used as a reference for how each individual should interact and engage with these media.

## WORKLOAD DISTRIBUTION

| TASKS | RESPONSIBILITIES |
|---|---|
| Brainstorming | All team members |
| Objective Formulation | All team members |
| Implementation of Objective 1 | Ranji Raj, Usama, Sujith |
| Implementation of Objective 2 | Madhuri, Vishnu, Sujith |
| Implementation of Objective 3 | Ranji Raj, Vishnu, Usama |
| Create Project Website | Sujith, Usama, Madhuri |
| Create Project Screencast | Ranji Raj, Madhuri, Vishnu |
| Final Project Presentation | All team members |


Brain Storming and Dataset analysis
April 20, 2021
Project Submission
July 6, 2021
Project Presentation
July 9, 2021
Week 1
3
5
6
7
8
9
10
11
2021
Deadline
Submissions
Project Proposal Submission
May 20, 2021
Brain storming and Objective formulation
Apr 20 –May 3
Brain storming
Apr 20 - May 5
Objective formulation
Apr 25 to May 20
Proposal documentation
Implementation and Experiment
Implementation of objectives
May 20 – Jun 20
Create Project website
June 10 - June 25
Create Project Screencast
Jun 28 – July3
Preparation of Final Submission & Presentation
UI testing and bug fixing
Jun 29- July 3
Final presentation
July 9