The Pandemic within COVID-19: Assessing Misinformation Susceptibility

Project Proposal - Group 6

Kyle Rogers
Reiko Matsuda-Dunn
Ryan Karasopoulos
Son Pham

https://github.com/summeryriddles/geopolymeric-tribbles

Description

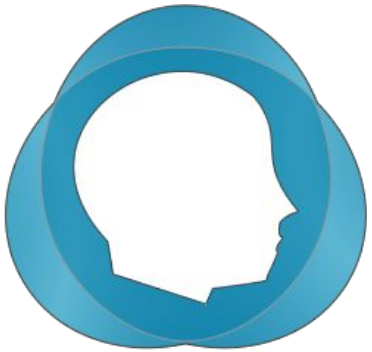
University of Colorado Boulder

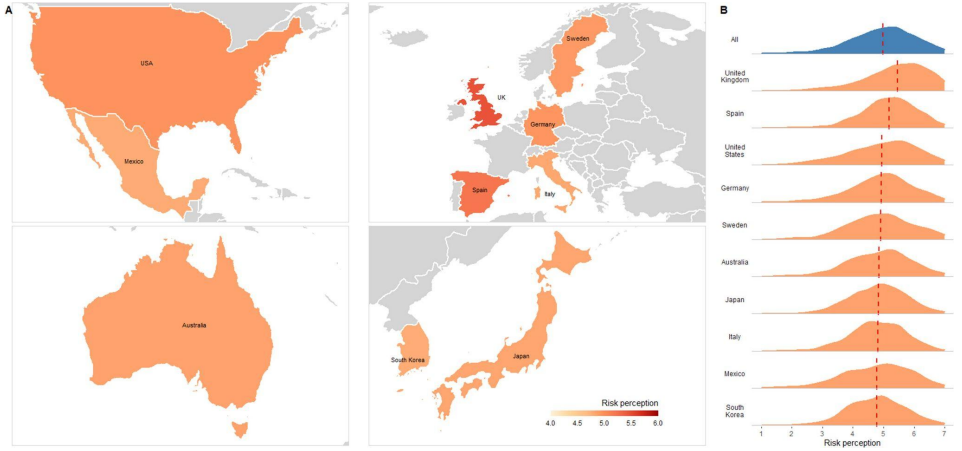
- This project focuses on individual susceptibility to misinformation regarding COVID-19. Possible predictors include:
  - How information was communicated
  - Reactions of governments to the pandemic
  - Attitudes and risk perceptions people had towards the virus
  - Trust in science and/or journalists
  - Political ideology
- Major questions to answer include:
  - How digital communications influenced people’s interpretation of the news
  - Did susceptibility to misinformation predict the responses to the new laws and mandates
  - Beliefs and concerns about COVID-19 versus other world issues
  - The similarity and trends among the different countries

Prior Work

University of Colorado Boulder

- Many COVID datasets and studies
  - This project will focus more on effects outside of cases/deaths/positive test rates
- Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research.
  - Effect of catastrophic event (COVID-19) on global collaboration (esp. US/China)
- COVID 19 Ethics and Research
  - Ethically conducting clinical trials
  - Addressing future COVID19 ethical concerns
- Effects of COVID-19 on mental health
  - Impact of Event scale for psychological effects
    - Commonly used to diagnose PTSD
    - Classifying the responses under categories Intrusion, Avoidance, and Hyperarousal
- Documentation of survey results (our data set)
  - Susceptibility to misinformation (October 2020)
  - Risk perceptions as correlation to survey results (April 2020)





Datasets

University of Colorado Boulder

- Risk perception of COVID-19/Coronavirus:
  - Survey of ~10,000 individuals from 12 different countries.
    - 104 different questions
      - Demographic data (7 attributes)
      - Questions regarding perceptions of COVID-19 risks, preparedness, information sources, trust in society, political views (90 attributes)
        - Answers are on a scale of 1-7
      - Probability math questions (4 attributes)
  - Fifteen .csv files of survey data from 12 different countries (multiple files for the US and UK).
  - One additional file of longitudinal data from the UK
- Where to find:
  - The data is made available by the Center for Open Science
  - https://osf.io/vhnk7/
- Each member individually downloaded the data

	EndDate	DemGen	quota_age	Residency	GenSocTr
0	EndDate	DemGen	quota_age	Residency	GenSocTrust
1	End Date (GMT)	What is your gender?	What is your age?	List of Countries	Generally speaking, would
2	nan	What is your gender?	nan	In which country do you currently live?	Generally speaking, would
3	nan	1 = female, 2 = male, 3 = other, 4 = prefer n...	1 = 18-24, 2 = 25-34, 3 = 35-44, 4 = 45-54, 5 = 55-64, 6 = 65+		1 = Can't be too careful to
4	2020-03-28T19:15:47Z	2	5	SE	3
5	2020-03-28T19:16:46Z	2	4	SE	4
6	2020-03-28T19:19:49Z	1	1	SE	5
7	2020-03-28T19:23:04Z	2	3	SE	5
8	2020-03-28T19:24:15Z	2	2	SE	5
9	2020-03-28T19:25:46Z	2	2	SE	7

Proposed Work

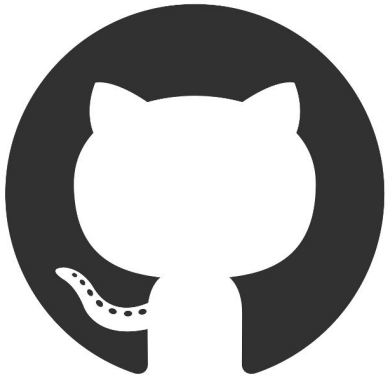
University of Colorado Boulder

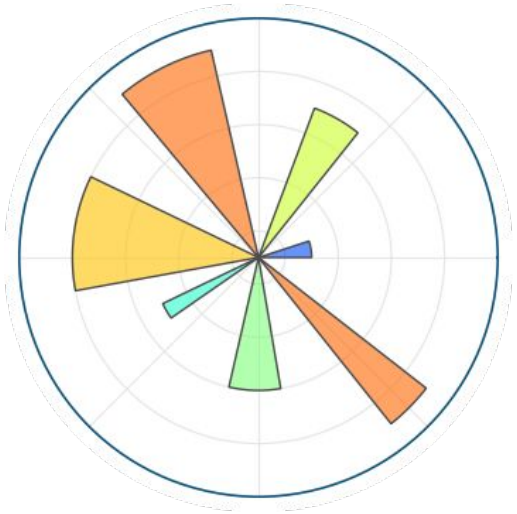
- Data cleaning
  - The data are all read into Python DataFrames as strings
    - str type is appropriate for some nominal data such as ‘Residency’
    - Numeric data will have to be converted to integers or floats where appropriate.
  - Look for participants who just clicked through the survey (e.g. all ‘1’s or all ‘7’s)
- Data preprocessing
  - Because the data is available in several different files, some attributes will have to be renamed for consistency.
  - The data was collected from a voluntarily survey;
  - There are null values in the dataset
    - Need to ensure consistent treatment of ‘null’ between csv files
    - Make decisions on how to treat each of these null values
      - Is an omitted survey answer data?
- Data integration
  - Each of the csv files will have to be appended into one dataset.
    - Consistent naming dealt with in preprocessing will facilitate this.
  - Column headers of some csv files are being read inconsistently
    - i.e., In some files, column header is the attribute name, such as “Trustingroups_1”, the next row is the exact question asked on the survey, “'How much do you trust each of the following? - People in your family'”, the next row is null, and the third row and beyond contain data. In other files, this order is different or there is no null row.

University of Colorado Boulder

List of tools

- Python, including various libraries
  - NumPy - computations, array manipulation
  - Pandas - data manipulation
  - MatPlotLib - plotting and visualization
  - Tkinter - user interface
- Git/Github - group repository
- Overleaf - Latex integration and group access
- Google drive - group presentation, spreadsheet planner access





University of Colorado Boulder

Evaluation of Data

- Comparison of individual categories in survey results
  - Correlation, outlier analysis, null response interpretation/integration
  - Cluster analysis based on country, education level, survey answers, etc.
- Result visualization
  - Histograms for each category, clustered categories
  - Scatter plots to map clusters
- Will need to avoid repeating prior work
  - Risk perception study heavily detailed data collected, used multiple linear regression and ANOVA to produce results
  - Original misinformation study only worked with 5 countries, more direct reporting of results (minimal correlation work)
    - This project will include all 12 countries
    - Clustering had not been explored by the original authors
- Analyzing longitudinal results
  - Examine longitudinal dataset
  - Stretch Goal: Investigate known population responses by country
    - Potentially look at additional data