The Pandemic within COVID-19: Assessing Misinformation Susceptibility

Kyle Rogers
Reiko Matsuda-Dunn
Ryan Karasopoulos
Son Pham

Project Overview

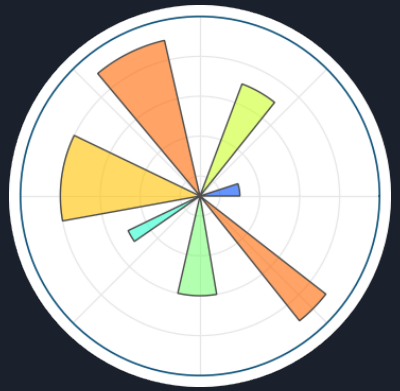
- Data from late March - mid May 2020 study on susceptibility to coronavirus misinformation.
- Survey data includes:
  - demographic information
  - political affiliation
  - personal views on coronavirus
  - degree of preparedness for coronavirus
  - media sourcing
  - trust in institutions and community

Questions We Sought to Answer

- How do digital communications influence people's interpretation of the news?
- Are there trends in political beliefs and in susceptibility to misinformation?
- Beliefs and concerns about COVID-19 versus other world issues.
- The similarity and trends among the different countries.

List of Tools

- Python, including various libraries
  NumPy, Pandas, MatPlotLib, Plotly, Tkinter, scikit Learn
- Git/Github - group repository
- Overleaf - group Latex integration
- Google Drive - group presentation, spreadsheet planner access



Data Preparation

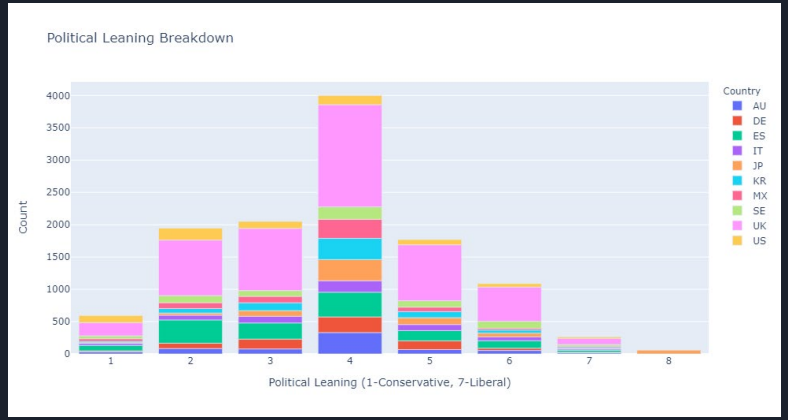
- Data originally in 15 separate files based on country:
  - Aligned and appended to one another.
- Data converted to ‘floats’ where appropriate.
- Null values were left, and dealt with on a case by case basis.
- 12,744 objects were contained in the final dataset.

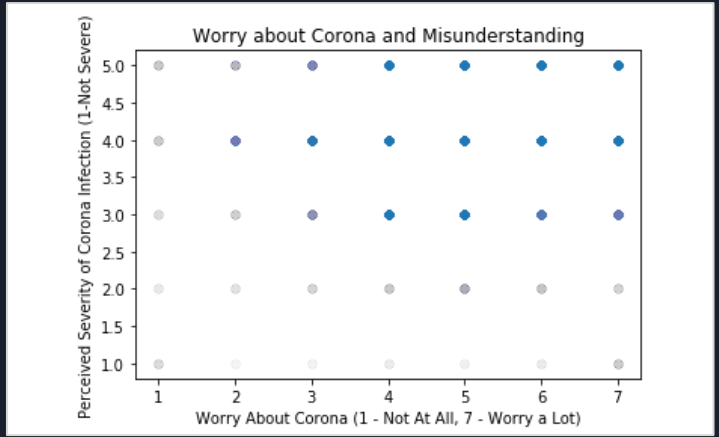
Methods Applied

- Exploratory Statistical Analysis
- Apriori Analysis
- Classification
  - Decision Trees
  - Naive Bayesian
- Clustering
  - k-Means
  - DBSCAN

Knowledge Gained - Preliminary Results

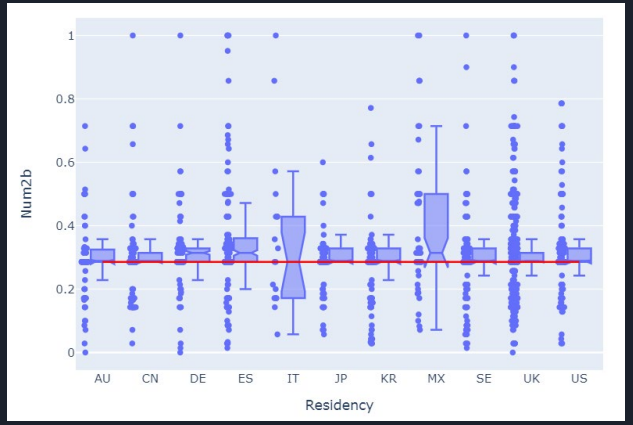
- The UK is vastly overrepresented in the data compared to other countries
- Political Affiliations followed an approximate bell curve, skewed slightly towards “Liberal”
- Relatively Low Correlations Between Question Classes
  - Most Correlation Coefficients were between 0 and 0.2
- “Severity of Infection” most closely correlated with “Worry About Coronavirus” (~0.45 CC)

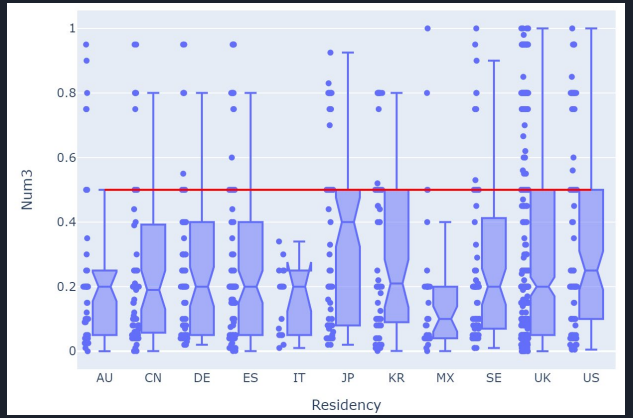




Knowledge Gained - Exploratory Statistics

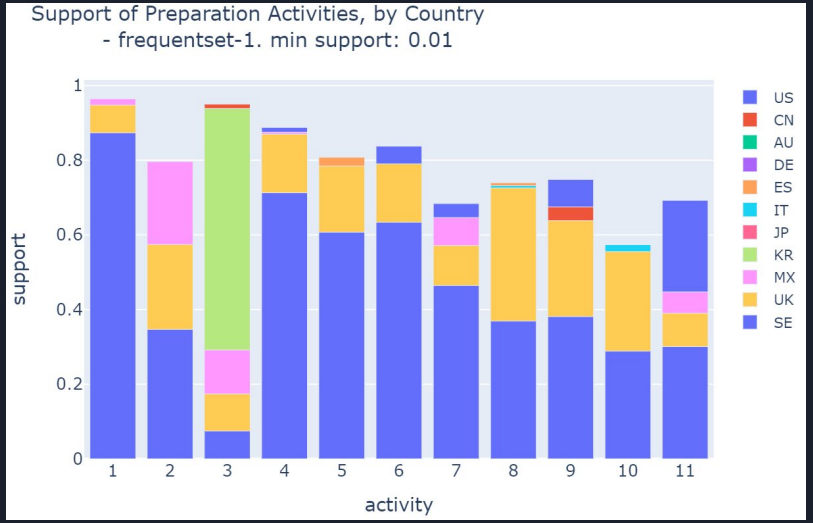
- Probability questions for participants to gauge knowledge in math
- Attributes:
  - Num1, Num2a, Num2b, Num3
- Pattern
  - MX, SE, IT has wide quartile ranges
  - Higher deviations from the truth





Knowledge Gained - Apriori Analysis

- Visualize the level of support each country has for each activity shown
- Frequent-1 Itemsets shows pattern among maskwearers.
  - US, UK, SE have low mask support
  - Same countries remain top with cases per million population



Activity	Description
1	washing hands more often
2	using alcohol-based hand sanitizer more often
3	wearing a face mask
4	avoiding social events
5	avoiding public transport
6	eating out less
7	touching your face less
8	shopping for groceries less
9	cooking at home more
10	staying home from work
11	purchasing extra supplies

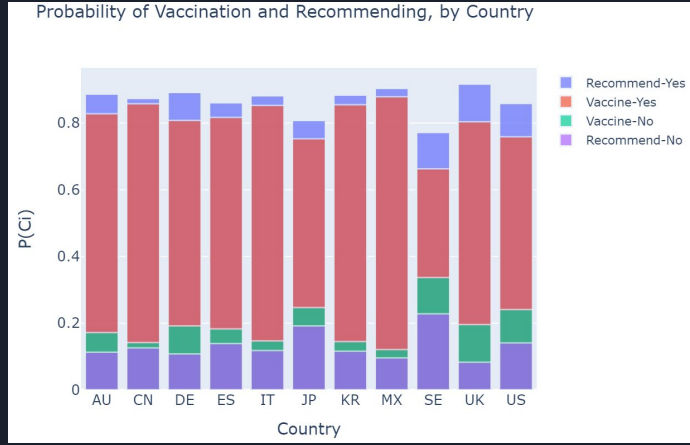
Knowledge Gained - Bayesian Classification

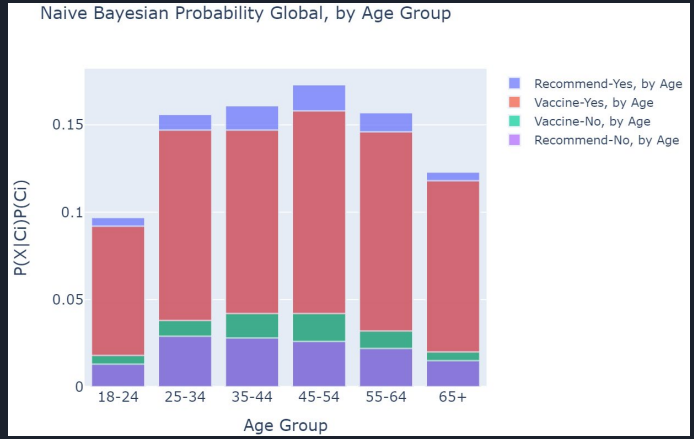
- Build a classification model to predict which attributes contribute to the susceptibility to misinformation
- Gain table shows top attributes that contain the highest information gain are trust-related attributes

attribute	gain
PosterstrustQ1	0.0535
WHOtrustQ1	0.0454
workplacetrustQ1	0.0367
SocialmediatrustQ1	0.0326
FinitePool_2	0.0236
FriendstrustQ1	0.0221
Govrestrict_3	-0.0072
Trustingroups_9	-0.0073
Politics	-0.0076
FinitePool_3	-0.0081
Personal_6	-0.0082
Friends_6	-0.0094

Knowledge Gained - Bayesian Classification

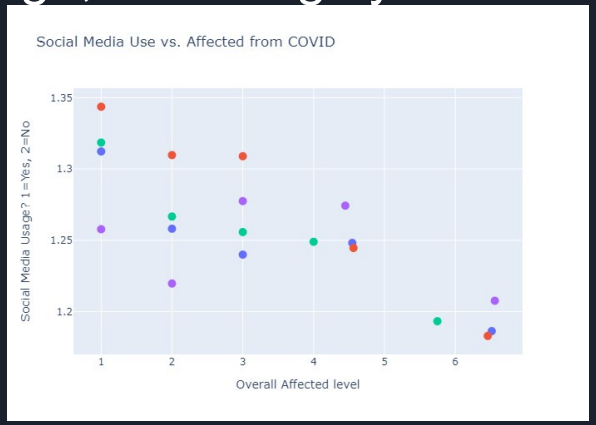
- Our model used Vaccine1, Vaccine2, and CanadaQ1 as potential class labels
- CanadaQ1 was split into 2 groups:
  - 1,2,3 = Not Serious
  - 4,5 = Serious
- Highest probability of acceptance
  - Age group 45-54
- Lowest probability
  - Age group 18-24

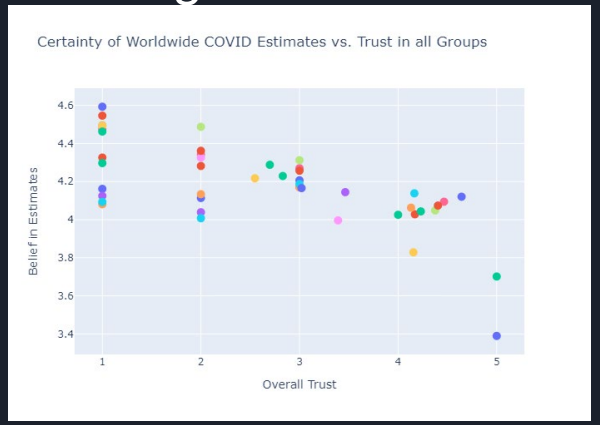




Knowledge Gained - K-means clustering

- Trust in general (13) & Trust in Gov. (4) higher for more WHO exp.
  - Less WHO exp.-> Lower trust in non-gov officials
- More social media use-> more affected
- All believe certainty of worldwide case/death/spread estimates
- Worry from COVID High, those highly affected even higher





Knowledge Gained - DBSCAN

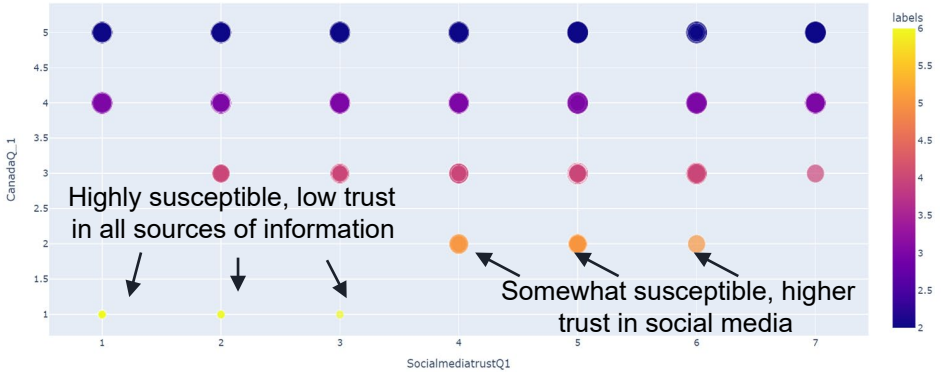
Most susceptible to misinformation:

- Low trust in all sources of information

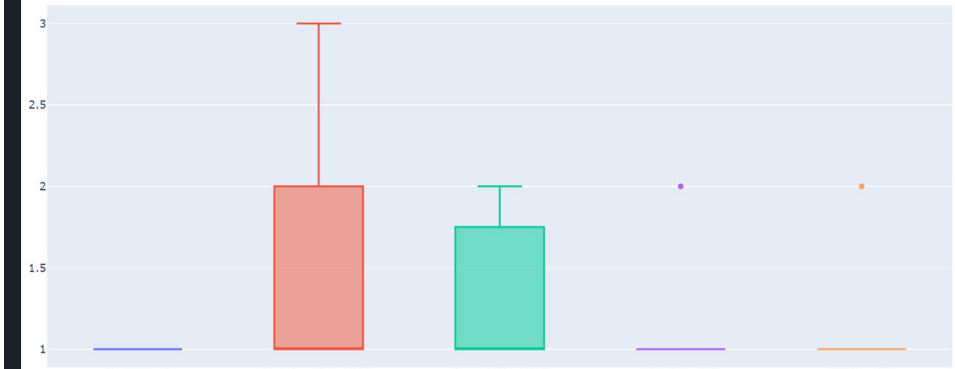
Somewhat susceptible to misinformation:

- Slightly higher trust in social media.

Clustering results based on trust in various information sources



Distribution of trust in information sources of those most susceptible to misinformation



Knowledge Gained - DBSCAN

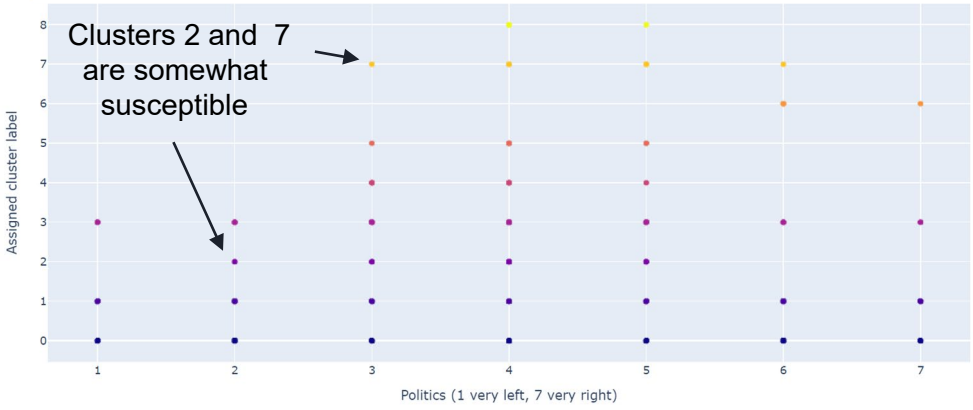
Most susceptible to misinformation:

- Politically diverse group

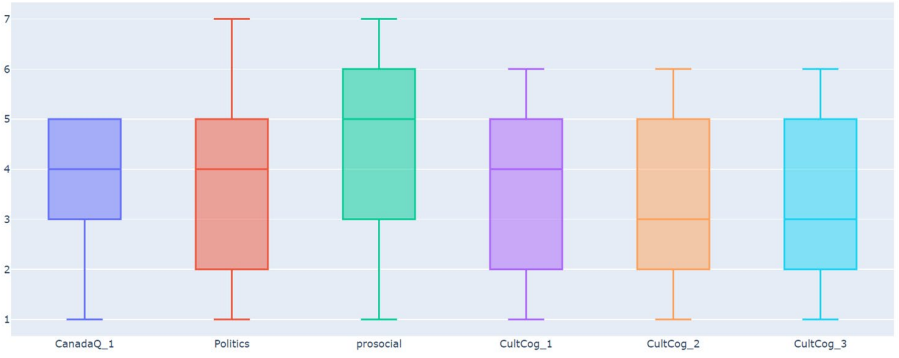
Somewhat susceptible to misinformation:

- A range of political opinions:
  - One small cluster (N=17) :
    - Right-wing, but believe in greater government control
  - Others are centrists across are variety of measures

Clustering results based on political attributes



Political Attributes of those most susceptible to misinformation



Future Work and Applications

Repeat:

- Survey to provide insight on changes in responses.

More nuanced approach:

- Group that is susceptible to misinformation is very small (3.8% of the dataset)
- Diversity in that group makes it resistant to classification methods.
- Difficult to create an accurate predictive model to determine who is susceptible to misinformation from the current data

Resources

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

- Visualization: https://pharsalus.herokuapp.com/