The effect of socio-demographic and attitudinal predictors on mask-wearing over Canada during coronavirus pandemic, together with a spatial analysis: female and older people are more likely to wear a mask

Yingshi Wang

21/03/2021

Abstract

Since year 2020, COVID-19 affects different people in different ways. Most infected people will develop mild to moderate illness and recover without hospitalization. Wearing a mask is one of the significant action to prevent being infected, and it is also a heated policy issue during COVID-19 pandemic. We evaluate the situation of wearing of a face mask in Canada. And mainly focusing on two research questions: who is most likely to wear a mask and where people wearing ask is the most common. An generalized linear mixed model was used to solve for the first research question by combining the information of socio-demographic and attitudinal information. On the other hand, we use spatial model attempting to find out where people wearing mask is the most common. We found out the answers after modeling. Female and people with older age are more likely to wearing a mask, people have higher confidence and trust in media and government are more likely to wearing a mask. We also found out people living in eastern area, such as Toronto, are more likely to wearing a mask during coronavirus pandemic, and people living in western area in Canada,such as Vancouver,are less likely to wear a mask. Since there are unequal number of respondents among provinces were collected, the research could be improve by increasing the sample size in each province.

1. Introducion

This research is promoted by the project collaborator and is conducted under the instruction of STA490 in University of Toronto. The report aims at analyzing and exploring the effectiveness of factors on mask uptake. As we know, the virus that causes COVID-19 can be spread by such things as coughing, sneezing, or even speaking at a close range. Using a mask is significant to limit the spread of COVID-19 as we return to our usual activities. We are interested in investigating two research questions: who wearing a mask is the most common and where mask wearing is the most common.Our data given by Policy, Elections, and Representation Lab(Pearl), is one of the largest survey data sets of Canadians' attitudes and behaviors around COVID-19 with 43272 observations. The data consists of 18 discrete, weekly surveys("wave") from Mar 25th to Aug 31st. I used 27 variables from the original dataset, including response id, wave number, age, gender, province, FSA, education level, etc. In this report, two models were used to implement the objectives, the Generalized linear mixed model and the Spatial model. The report contains the following sections: a method section describing the implementation steps of both data cleaning and model building, a 'result & discussion' session discussing the main findings of model results, a conclusion section summarizing the implications of the entire statistical analysis and commenting some limitations of this research.

2. Method

The methodologies used during this project are generalized linear mixed model and spatial model. A short introduction to these two models are presented below. In addition, we constructed a factor analysis to finish the process of variable selection, however, due to the limitation of time and computing power, we can't guarantee that the selected factors are the most suitable ones for our model.

Generalized linear mixed models (GLMMs) are an extension to both linear mixed models and generalized linear models. It combines the ideas of generalized linear models with random effects modeling overdispersion and correlation. The response variable is also allowing to follow a non-normal distribution. GLMMs cover a wide range of models, from simple linear regression to complex multilevel models for non-normal longitudinal data.

Most models assume that observations are obtained independently of each other. However, observations in most cases are correlated. Distance and location between observations could be one of the sources of correlation. For instance, the location of trees in a forest may follow a spatial smooth pattern based on soil nutrients, moisture content, etc. The tree data would have a spatial nature because the observations showing a higher degree of spatial autocorrelation. In this case, spatial model could take into account the spatial autocorrelation so that the general trend from the purely spatial random variation could be separated.

Before looking at the specific procedure of modeling, let's take a glimpse of our data on Table 1 and there are some sample survey questions shown below. In addition, the `Method` section is discussed in multiple sub-sections which would be easier for you to follow.

Table 1: A Glimpse of Data

	respid	wave	age	Q4	Q5	FSA	Q7	Q73	Q10_12	Q12	Q140	Q22	Q24	Q33	Q34
4971	1655768686	3	50	1	5	J4J	1	1	NA	4	2	1	1	1	2
4972	1607721823	3	41	1	6	L5B	1	2	NA	3	2	3	5	3	5
4973	1483422459	3	62	2	6	N3C	1	1	1	3	2	1	1	1	1
4974	1483975267	3	39	2	5	G0K	1	2	NA	2	2	1	2	4	5
4975	1684339476	3	31	2	10	V3R	2	1	NA	3	2	3	3	3	5
4976	1482904270	3	41	2	6	L6Y	1	1	NA	3	2	4	5	2	5

Q5 - Which province do you currently live in?

Q7 - How concerned are you about the coronavirus pandemic?

Q73 - How serious of a threat do you think the coronavirus (COVID-19) is to Canadians?

Q10 - Over the past week, which of the following actions have you taken as a result of the coronavirus pandemic? Please select all that apply.

Q140 - Has someone in your household been tested for the coronavirus (COVID-19)?

Q22 - How frequently did you watch, read, or listen to news about politics over the past week?

2.1 Generalized linear mixed model

Research question 1: Who is most likely to report using the mask

In the stage of data analysis, I first addressed the first research question. As our data is a multilevel dataset and respondents are nested within provinces, which in turn are nested within regions. So I considered province as our the random effect of the generalized linear mixed model because respondents are nested with provinces, and I also believe that the severity of Covid-19 spread is different among provinces so that number of people wearing mask would also differ. Therefore, respondents would be in the first level of the model and provinces would be in the second level of our model. The following equation is our model equation.

$$logit(\pi_i) = \beta_i X_i + U_i + \epsilon_i$$
$$where\ i = 1, ..., 10$$
$$U_i \sim N(0, \tau_u)$$
$$\epsilon_i \sim i.i.d\ N(0, \Sigma)$$

Noted that n_j represents the number of respondents in province j, U_i is the random effect of province, the i^{th} province.

There are also some concerns for this model. We noticed that few individuals had done the survey questions twice and these repeated measurements could possibly have a side effect since generalized linear mixed model requiring independent observations. Furthermore, some respondents may know each other, for example, they may be neighbors or friends, so our observations are not independent with each other regardless repeated measurements.

2.2 Spatial model

Where is masking wearing most common?

2.2.1 Data preparation

In this section, we investigate in the second research question and examined mask uptake situation over entire Canada in depth and plotted the observed the count of people wearing a mask in each province, expected rate of people wearing a mask for each province. Spatial Areal Unit Modeling with Conditional Autoregressive Priors is our analysis tool because we having a binary response variable and this model allowing us to group individuals into discrete areas. Moreover, We eliminated those socio-demographic predictors in order to keep the simplicity of the model. The model was constructed by first creating an binary adjacency 13*13 matrix W, where 13 is the number of provinces in Canada so that we have 13 areas. A binary specification for W matrix is based on geographical contiguity, and $W_{kj} = 1$ if areal units (S_k, S_j) sharing a common border, and is zero otherwise(W matrix is shown in Appendix). Next we implement spatial modeling using `S.CARmultilevel()` function.

2.2.2 Map

I imported map of Canada from `Statistics Canada`, and use it along with the above constructed data frame to create a `SpatialPolygonsDataFrame` which allowing us to make maps of the variables in data frame. I also calculated Standardized Mask Ratio(SMR) to estimate the situation of make wearing, where

$$SMR\ in\ province\ i = SMR_i = \frac{Y_i}{E_i} = \frac{num\ of\ observed\ cases\ in\ province\ i}{num\ of\ expected\ cases\ in\ province\ i}$$

If $SMR_i = 1$, number of observed cases in $province_i$ = number of expected cases in $province_i$.

If $SMR_i > 1$, number of observed cases in $province_i$ > number of expected cases in $province_i$, which means people in $province_i$ having relative less risk exposed to COVID-19,regardingless other factors.

If $SMR_i < 1$, number of observed cases in $province_i$ < number of expected cases in $province_i$, which means people in $province_i$ having relative higher risk exposed to COVID-19,regardingless other factors.

And finally we can visualize the observed and expected mask-wearing coutns, the SMRs in an interactive map which is constructed in `leaflet` package.

3. Result and discussion

3.1 GLMM

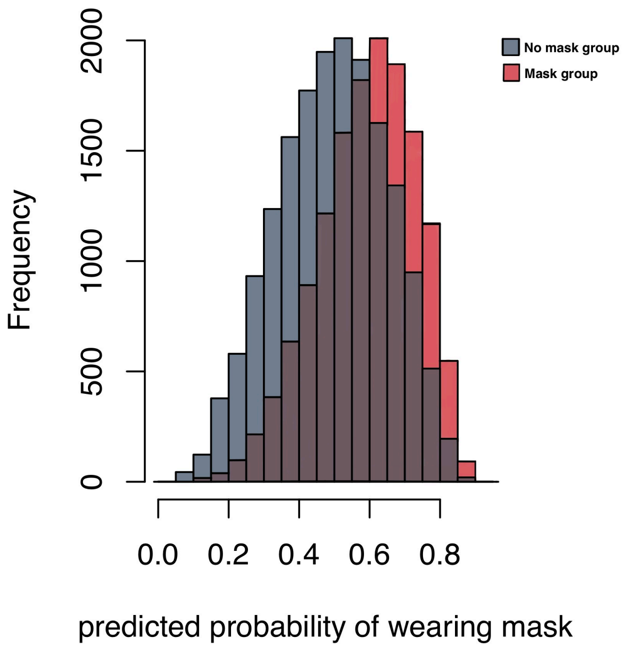
Table 3: GLMM model summary

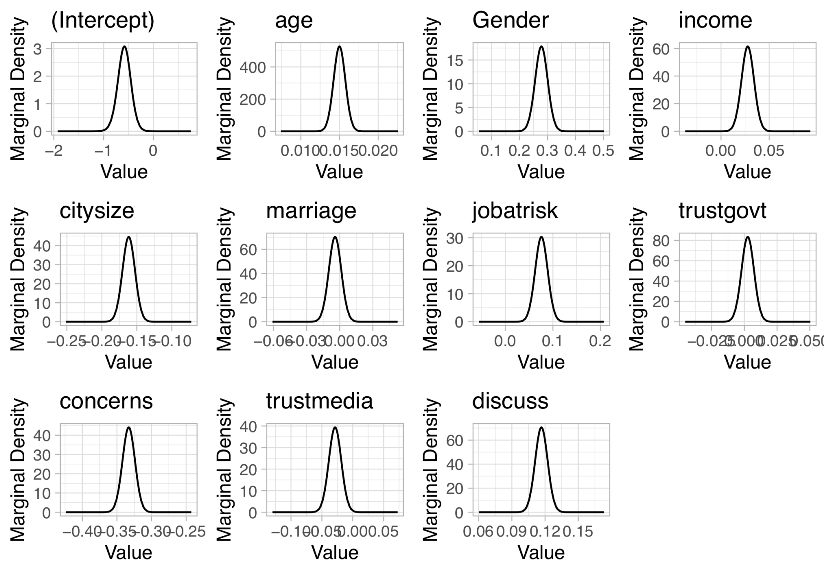
	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	0.557	1.143	0.426	0.557	0.722	0.558	1
age	1.015	1.001	1.014	1.015	1.017	1.015	1
Gender	1.321	1.023	1.264	1.321	1.380	1.321	1
income	1.028	1.007	1.015	1.028	1.041	1.028	1
citysize	0.851	1.009	0.836	0.851	0.866	0.851	1
marriage	0.996	1.006	0.985	0.996	1.007	0.996	1
jobatrisk	1.079	1.013	1.051	1.079	1.107	1.079	1
trustgovt	1.003	1.005	0.993	1.003	1.012	1.003	1
concerns	0.717	1.009	0.704	0.717	0.730	0.717	1
trustmedia	0.972	1.010	0.953	0.972	0.991	0.972	1
discuss	1.124	1.006	1.111	1.124	1.136	1.124	1

In table 3, we had an exponentiated model result and found that most variables positively affect the odds of an individual being reported to wear a mask. The generalized linear mixed model exhibits the following results. Firstly, it is not surprising that the odds that a respondent wearing a mask increase by 1.5% for each additional year of age, because older people are at higher risk for more severe cases of COVID-19. For a female, the odds of wearing a mask are 1.321 times as large as the odds for a male wearing a mask, even though males may be a risk factor for more severe outcomes of COVID-19. Perhaps wearing a mask is viewed as a sign of fragility or weakness among some men in Canada. Results also indicate the odds that a respondent wearing a mask decrease by 14.9% for each level decrease in city size, due to a greater resistance to mask-wearing may be concentrated outside of urban areas. It is also not surprising that a respondent working at a higher risk of exposing to COVID-19, the odds of wearing a mask are 1.079 times as large as the odds for a respondent working in a safer environment, for example, doctors and nurses would wear a mask every day. Furthermore, I combined 3 scores to create a new score to reflect if people's trust and confidence in the Canadian government would impact the behavior of they wearing a mask, they were a score of trust in the federal government, provincial government, and local government respectively. It is very surprising that for an individual having relatively little confidence in the Canadian government to contain the coronavirus pandemic, the odds of wearing a mask are 1.003 times as large as the other people wearing a mask. However, it is only 0.3% larger and its 95% credible interval $(0.993, 1003)$ including 1, so that we could say it is not sufficient enough to make a conclusion. Lastly, we could see that for individuals who discuss the coronavirus pandemic with friends, family, and acquaintances, the odds of wearing a mask increase by 12.4% for each level of increase in discussion frequency, because they may be more concerned about the coronavirus pandemic.

The model studies suggest that promoting social norms for widespread use of masks could be an effective intervention to increase mask usage, because this would be helpful to eliminate the stereotype of wearing a mask so that the public can properly understand why wearing a mask is so important during coronavirus pandemic. On the other hand, this is called normative conformity, people will follow the cues of the other people they like and respect. When we have one people wearing a mask, his/her relatives, friends, neighbors will possibly change their attitudes toward mask wearing.

Here we have the distribution of the predicted probability of wearing mask, and I divided individuals into two groups: wearing a mask group and not wearing a mask group. It is interesting that the group of people not wearing a mask has a normal distribution, and group of people wearing a mask having a left-tailed distribution. However, this model prediction is also consistent with our expectation that each individual has a different preference on wearing a mask during Covid-19 pandemic, and respondents who choose to wear a mask having a higher predicted probability of wearing a mask, and respondents who don't having a lower predicted probability of wearing a mask.





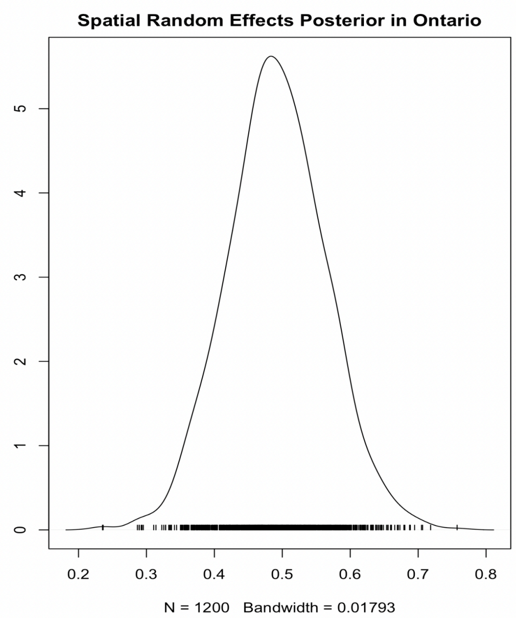
	mean	sd	q0.025	q0.5	q0.975	mode
SD for province	0.3024336	0.0780162	0.1881809	0.2893352	0.4928701	0.2670584

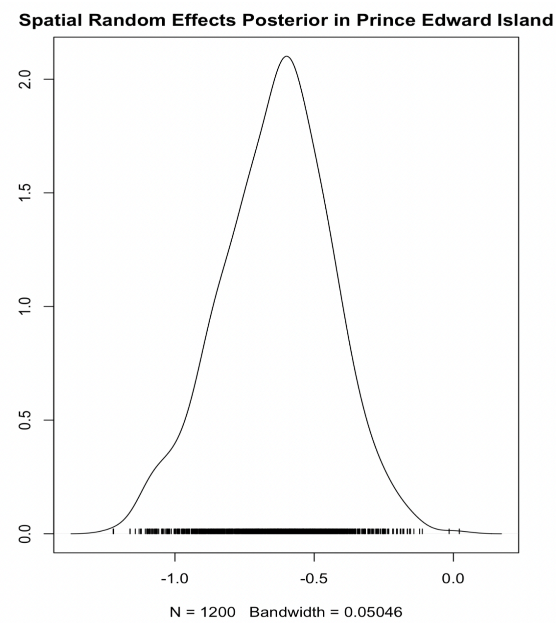
Based on the above posterior distributions, we see that all distribution concentrated at a single point and there is not so much variance covering zero, therefore, our variables are statistically significant. And we say our model is doing a reasonably good job on modeling reality. Since the problem of logistics regression is that it is not like the simple linear regression, for SLR, we have assumption, the observations are normally distributed around certain mean. And we can plot the residual plot or Normal QQ plot to check assumption. In logistics regression, all we really getting is parameters maximized some likelihood, we don't have the same normality assumption as fitting a standard regression. Instead of doing a model diagnostic, we want to see if our model is doing a reasonably good job of modeling reality and we can check it by looking at the posterior distributions.

3.2 Spatial modelling

	Median	2.5%	97.5%	n.sample	n.effective	Geweke.diag
(Intercept)	0.0110	-0.1681	0.2195	1200	415.3	-0.1
trustgovt	-0.0303	-0.0421	-0.0185	1200	583.0	0.7
citysize	-0.0966	-0.1199	-0.0740	1200	638.3	0.5
trustmedia	-0.0736	-0.1003	-0.0467	1200	465.1	-0.7
Q79	0.0941	0.0799	0.1080	1200	837.8	0.4
tau2	0.5461	0.2162	1.6879	1200	1003.2	-0.7
rho	1.0000	1.0000	1.0000	NA	NA	NA

The median value of respondents' confidence scores in government and media indicating that respondents who have less trust in government(federal, provincial, local government) and social media are more 3% and 7% less likely to wear a mask respectively, this may because they are less likely believe the suggestion from government. And we also have estimate 0.0948 for Q79 which represent the education level(from low level to high level). People who are highly educated are 9% more likely to wear a mask, and this is reasonable and probably because they are more believe in any scientific reports and news from social media or newspaper. Moreover, we have all of our convergence diagostic proposed by Geweke (Geweke.diag), which takes the form of a Z-score, within the range (-1.96,1.96), a good size of this model.



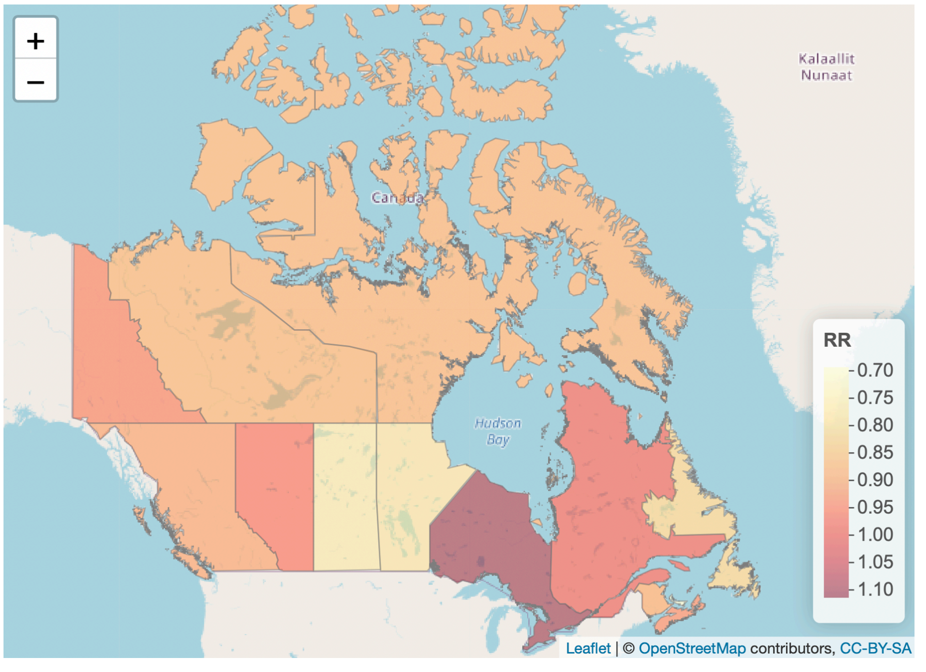


Next, we found that from the spatial random effects plots that most areas have a positive distribution, but area 6 has a relatively negative distribution. This aligns with what we would expect because Ontario has a relatively more population so that it has a higher posterior value, and Prince Edward Island is relatively sparsely populated so that there are fewer people wearing a mask on data. Moreover, it could also because the epidemic in Prince Edward Island is not as severe as it in Ontario.

Table 6: ith area Spatial Random Effect Posterior

	Spatial Random Effect Posterior
Nova Scotia	0.1954348
Prince Edward Island	-0.6438734
New Brunswick	0.0984090
Quebec	0.1576758
Ontario	0.4830500
Manitoba	-0.3002604
Saskatchewan	-0.2966068
Alberta	0.0990695
British Columbia	-0.0049307
Northwest Territories	-0.0244240
Yukon	0.6180660
Newfoundland and Labrador	-0.1707607
Nunavut	-0.2108491

In table 6, we have the matrix of spatial random effects and each entry is the posterior mean of the spatial random effect associated with a given area. An area with a greater size of population tend to have a positive posterior mean, such as Quebec and Ontario, and area like Newfoundland and Labrador & Nunavut having a negative posterior mean. Until 2020 September 1st, Ontario and Quebec both had more than 50,000 total confirmed cases, in particular, Quebec had been reporting nearly twice as many cases as Ontario, although it had a smaller population(8.5 million versus 14.5 million). Furthermore, Ontario and Quebec have consistently outperformed most other provinces in terms of finance during the last four years. Therefore, due to a greater amount of confirmed cases and business areas that were well alerted during the coronavirus pandemic, Ontario and Quebec had greater posterior values than other provinces with 0.483 and 0.157 respectively.



From the map, we observe that people in eastern Canada having relative more people wearing a mask and northern Canada has a more wearing mask. Ontario and Quebec having the most amount of people wearing a mask during the past weeks, and this aligns with what we expect that they were the center of the outbreak since COVID-19 found in Canada. And the other areas with less observed cases may because they are less population and less confirmed cases during last few weeks.

4. Conclusion

We investigated who wearing a mask and where people wearing a mask is the most common by conducted a generalized linear mixed model and spatial model. For those who wear a mask most commonly, we utilized socio-demographic and attitudinal predictors. The generalized linear mixed model reported that females and elder people, and people who have more trust in government and social media are more likely to wear a mask. For where people wearing a mask the most common, we utilized the spatial model and the model results indicated that people living in Eastern Canada are more likely to wear a mask during the coronavirus pandemic. On the contrary, people living in western Canada(e.g. Vancouver) are less likely to wear a mask, which government needs to pay more attention to those areas. Furthermore, Ontario and Quebec have the most people wearing a mask compared with the other provinces until Aug 31st, 2020, resulting from greater COVID-19 confirmed cases and more financial districts which had been well alerted. In conclusion, our models suggest that promoting mask use could be one of the key interventions governments, communities, businesses and other organizations can implement to control the spread of COVID-19. In particular, we could promote mask use more in the smaller provinces with relatively higher COVID cases.

4.1 Limitation

1. In the Generalized linear mixed model, there are repeated measurements in data so that it is very hard for us to include a random effect(e.g random slope) since we believe that there could be different levels among each respondent. For example, we could have the date as the random slope for our model at the province level, this means that the time effect on mask-wearing is different for different values of the province. Due to the limitation of repeated measurements, it is very hard for us to implement this step.

2. In the spatial model, SMRs may be misleading and unreliable when there is a small population or rare diseases. Our model can not incorporate covariates and borrow information from neighboring areas to obtain smoothed relative risks. For example, we only have 5 or 6 observations in Nunavut, the spatial model couldn't collect enough information from its neighboring areas in our current dataset due to the lack of information.

3. Another limitation of this report is the technical equipment because our computer does not have enough RAM to run a more complex model and sometimes we can't run the model on full data. As mentioned before, the limitation of computation power, we couldn't guarantee that the selected variables are the best suitable for the models. 4. Since there are small data counts in some provinces, the data got from the spatial model might not be informative. We are having a small count of the population in three territories(e.g. approximately 10 respondents from each of them), therefore the model results are not representative.

4. Since there are small data counts in some provinces, the data got from the spatial model might not be informative. We are having a small count of the population in three territories(e.g. approximately 10 respondents from each of them), therefore the model results are not representative.

4.2 Further improvement

1. Increasing the number of covariates to investigate in the GLMM model so that we could analyze more characteristics of a group of people wearing a mask.

2. Developing the computing power so that we could run the model in a more powerful analysis tool, though these tools are usually needed to be paid.

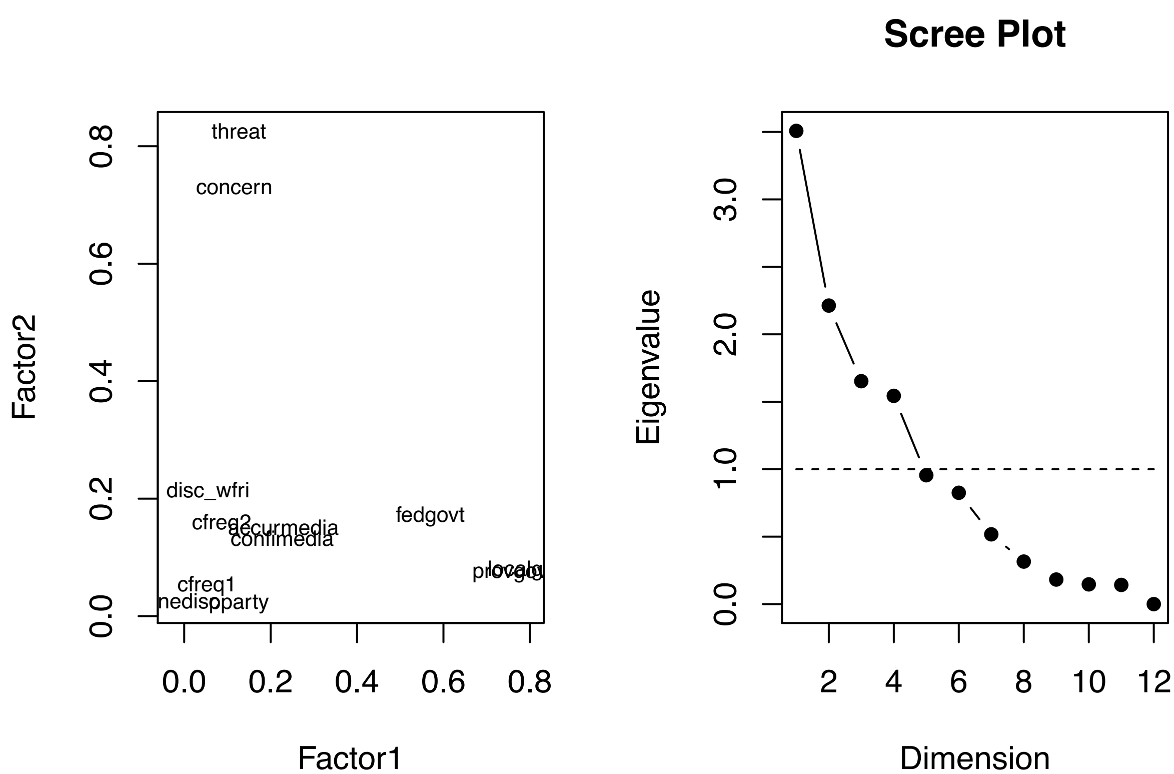
3. Giving out more survey questions to the public in three Canadian territories and collecting more data so that the information could be representative for those territories.

Appendixes

Factor analysis results

```
##
## Loadings:
##            Factor1 Factor2 Factor3 Factor4 Factor5
## concern     0.116   0.728   0.133   0.195   0.126
## threat      0.127   0.826   0.127
## disc_wfri           0.214           0.614
## onlinedisc                          0.529  -0.147
## pparty      0.127                   0.275
## cfreq1                                      0.409
## cfreq2              0.157   0.179   0.509   0.478
## confimedia  0.227   0.133   0.780   0.157   0.199
## accurmedia  0.230   0.151   0.658   0.156   0.126
## localgovt   0.799           0.138
## provgovt    0.761           0.112           0.145
## fedgovt     0.570   0.169   0.296   0.137
##
##                Factor1 Factor2 Factor3 Factor4 Factor5
## SS loadings      1.706   1.367   1.251   1.117   0.517
## Proportion Var   0.142   0.114   0.104   0.093   0.043
## Cumulative Var   0.142   0.256   0.360   0.453   0.497
```

Scree plot of factor analysis



Neighbour matrix W for spatial model

```
##       [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13]
##  [1,]    0    1    1    0    0    0    0    0    0     0     0     0     0
##  [2,]    1    0    1    0    0    0    0    0    0     0     0     0     0
##  [3,]    1    1    0    1    0    0    0    0    0     0     0     0     0
##  [4,]    0    0    1    0    1    0    0    0    0     0     0     1     0
##  [5,]    0    0    0    1    0    1    0    0    0     0     0     0     0
##  [6,]    0    0    0    0    1    0    1    0    0     0     0     0     1
##  [7,]    0    0    0    0    0    1    0    1    0     1     0     0     0
##  [8,]    0    0    0    0    0    0    1    0    1     1     0     0     0
##  [9,]    0    0    0    0    0    0    0    1    0     1     1     0     0
## [10,]    0    0    0    0    0    0    1    1    1     0     1     0     1
## [11,]    0    0    0    0    0    0    0    0    1     1     0     0     0
## [12,]    0    0    0    1    0    0    0    0    0     0     0     0     0
## [13,]    0    0    0    0    0    1    0    0    0     1     0     0     0
```

Glossary

“Attitudinal-like” variables:

- Q25 What is your primary source for news on the coronavirus pandemic?
- Q26 How often have you read, listened to, or watched news related to the coronavirus pandemic over the past week?
- Q27 In general, how much trust and confidence do you have in the mass media when it comes to reporting the news fully, accurately, and fairly?
- Q28 How accurate, do you think, is the news posted online by news organizations?
- Q7 How concerned are you about the coronavirus pandemic?
- Q73 How serious of a threat do you think the coronavirus (COVID-19) is to Canadians?
- Q33 Over the past week, how often did you discuss the coronavirus pandemic with friends, family, and acquaintances?
- Q34 Over the past week, how often did you have online discussion about the coronavirus pandemic?
- Q55 In federal politics, do you usually think of yourself as a(n)...
- Q65 To what extent do you approve or disapprove of your local government’s handling of the coronavirus pandemic so far?
- Q64 To what extent do you approve or disapprove of your provincial government’s handling of the coronavirus pandemic so far?
- Q63 To what extent do you approve or disapprove of the federal government’s handling of the coronavirus pandemic so far?