

Understanding the Impact of Covid-19 on Mental Health in Canada

Logistic regression models using 2020 Mental Health Survey data

Promiti Datta

December 21, 2020

Abstract

This analysis of a 2020 survey on mental health in Canada looks at different factors that impact a person's anxiety levels during the covid-19 pandemic. We use logistic regression to model whether a person has anxiety or not, based on demographics and categorical responses to questions in the survey. In this analysis we will identify the most impactful factors contributing to anxiety to understand how mental health has changed in the pandemic, and what we can do about it.

Code supporting this analysis can be found at: <https://github.com/promitid/sta304-final-project>

Keywords: Mental Health, COVID-19, Logistic Regression, Survey Data

Introduction

Throughout one of the biggest global health crises of the past century, people around the world have experienced a lot of uncertainty about their future, whether it be around their health, financial status, or other reasons. Like many countries around the world, Canadians have spent the majority of 2020 in varying degrees of pandemic related lockdown. With people's daily lives changing so much during the coronavirus pandemic, it is important to address increasing concerns around mental health. Through large scale mental health surveys like the one in this paper, state and federal governments can understand the most vulnerable populations for poor mental health and launch targeted support programs.

The most common mental illnesses in Canada are anxiety disorders and depression, with around 20% of Canadians experiencing a mental illness in the year 2013 (Slaughter, 2020). Many surveys have concluded that mental health has worsened since the start of the pandemic, but few have done follow up iterations of the same survey to see the responses at different points in the pandemic like the survey analyzed in this paper. It is important to note that anxiety and depression are clinical conditions, and while this paper looks at mental health issues that survey participants face, all answers provided are self assessed and do not necessarily equate to a clinical diagnosis. This paper uses binary logistic regression, a common model in statistics for predicting a binary dependent variable using independent categorical and numerical variables in a linear model (Assumptions of Logistic Regression, 2020).

This paper analyzes a survey done by the Center for Addiction and Mental Health to gauge the overall mental health of Canadians at various points throughout the Covid-19 pandemic. The Methodology section goes into detail on how the data was sourced and provides relevant visualizations and descriptions of the dataset, and introduces the logistic model. Then the key takeaways are presented in the Results section and followed by a discussion on how the findings in this paper are relevant to the current world we live in.

Methodology

Data

Results of this survey are from a total of 1,003 Canadian adults. The survey was collected online on the Asking Canadians web panel, which is an online research community of participants of different demographic backgrounds who voluntarily complete surveys. The same survey was rerun five times throughout 2020: on May 8, May 29, June 19, July 10, and September 18. Each iteration had different participants, but was sampled in the same way as outlined above. The data in this survey was collected by the Centre for Addiction and Mental Health in Toronto, Canada (COVID-19 Pandemic, 2020). Further instructions for accessing the data and reproducing results can be found in the Github repository associated with this project.

The target population for this survey would be all adults in Canada, however the sampling frame was limited to only the users of the Asking Canadians website. Everyone in the sample was at least 18 years old and living in Canada, but there was likely some voluntary response bias. Unfortunately, there was a very limited amount of information on the data collection methodology provided with the dataset, so it is hard to tell from the sampling method what kind of biases there may have been. There is also no mention of what took place in situations where the participant did not respond to some questions, although there is an answer provided in every cell within the data. Even though the survey was not conducted in a very randomized way, we see in the following section that there are not any glaring disproportions of any demographic group, in relation to the whole of Canada.

The basic frequency distributions of the participants by province, age group, gender, and race are shown in Table A in the Appendix. From these charts, we see a fairly equal gender distribution and age distribution, with slightly more respondents from the 18-39 age group. Within the race and province distributions, there were a lot more white people than any other race and more people from Ontario than any other province. However, given the demographic data of Canada, this is to be expected in a random survey and therefore this does not indicate a flaw in the survey (Government of Canada, 2020).

Table 1 and 2 show the mean estimates for the percentage of the population with anxiety and with depression, using finite population correction on the survey mean. Based on the data, we can see that a much larger percentage of the population experiences frequent anxiety, making up around 1 in 4 people, whereas just under 7% experience depression. In a previous survey done not long before covid-19 emerged, 5% of Canadians self reported high levels of anxiety, and 4% reported depression (Slaughter, 2020). Since we observe a much larger difference in anxiety since the start of the pandemic, we will focus on anxiety levels of the survey participants for the rest of the analysis.

Table 1: Survey-based estimates of fraction of population with frequent anxiety

Frequent Anxiety	Mean	SE
Yes	0.253	0.0137
No	0.747	0.0137

Table 2: Survey-based estimates of fraction of population with frequent depression

Frequent Depression	Mean	SE
Yes	0.0688	0.008
No	0.931	0.008

Next we look at the baseline of Canadians who reported feeling anxious throughout the five iterations of the survey. The highest percentage was in the first iteration on May 8, and the overall anxiety trend was negative over time, until rising again in September. There are many potential reasons for the decrease, such as the relaxing of lockdown rules in many places due to lower covid-19 case numbers in the summer, or the warmer weather allowing for more time outdoors. As many scientists had predicted at the time, covid-19 case numbers did start to increase around the fall time (Neustaeter, 2020), and therefore would likely have contributed to the overall anxiety going into lockdown again.

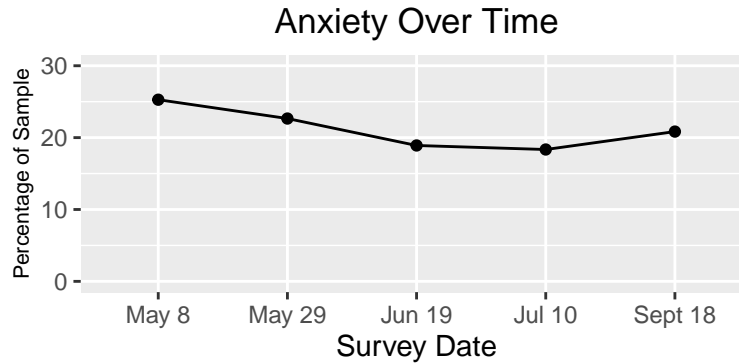


Figure 1: Anxiety of participants at different points in time

Lastly we look at the breakdown of anxiety levels for people in different demographic groups to assess impact over time. We can see from the first plot in Figure 2 that females tend to experience more anxiety than males, but the general trend of anxiety seems to be going in the same direction of the overall population. Because of the large difference between these two groups, gender would be a good factor to consider in the model.

The second plot in Figure 2 presents a more interesting breakdown of anxiety by age. Although generally people in earlier adulthood are likely to experience more anxiety (Bandelow, 2015), the anxiety levels for people above 60 follows a different trend than the general pattern of the other two age groups and the general population. We see that the oldest age group experiences a visible decrease in anxiety between July and September, while all other groups experience increase. It is worth noting that while this age group is most vulnerable to the deathly effects of covid-19 (COVID-19 and Your Health, 2020) and the number of covid-19 cases happened to be increasing around this time in September (Neustaeter, 2020), this did not impact anxiety levels of the elderly. This indicates that there are likely other reasons for anxiety besides the direct threat to one's health that the virus has. Likely causes to consider are financial instability, and the stress of changing lifestyles to follow lockdown rules again. In the next section, we will look at some potential variables of interest: Age group, Gender, Income group, Geographic region, and type of region.

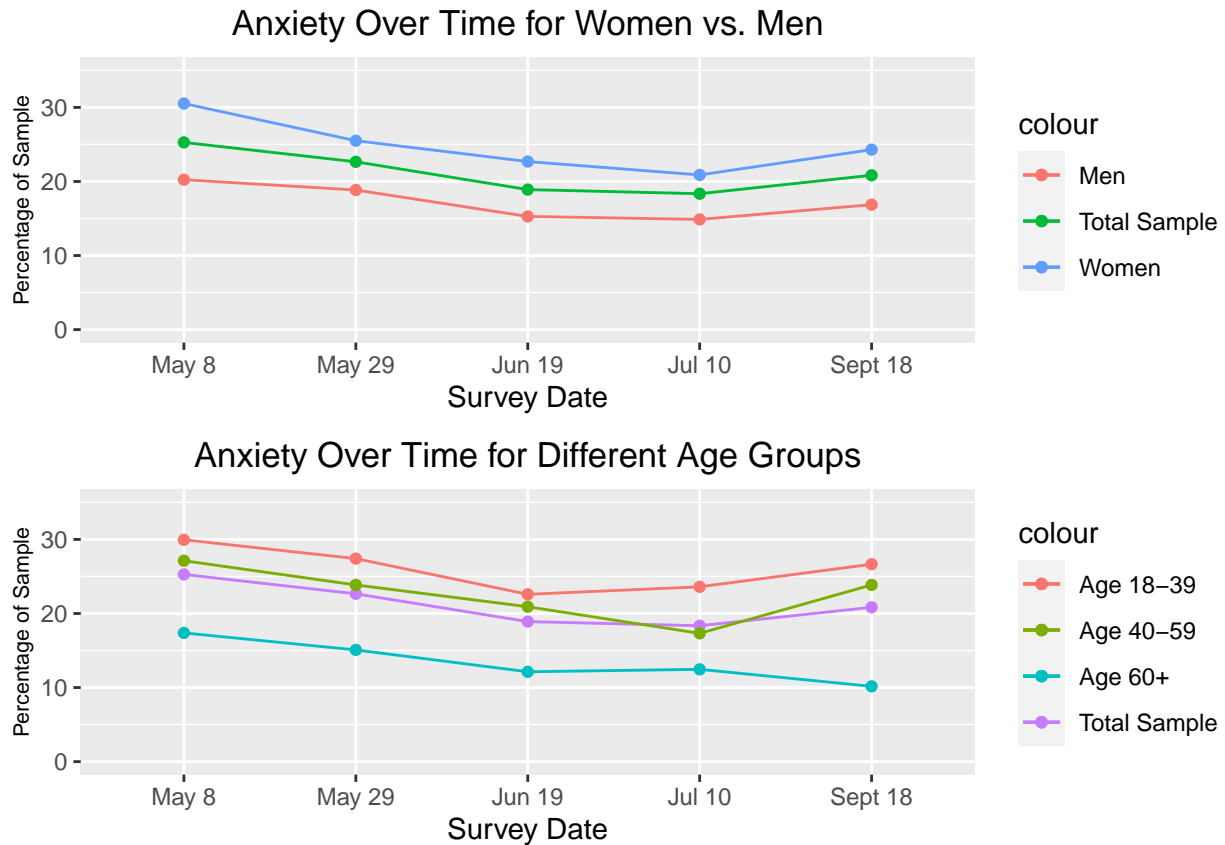


Figure 2: Anxiety of participants by Demographic Groups

Model

For this analysis, we chose a logistic regression model to understand how different factors affect whether a person experiences frequent anxiety or not. This model was chosen because the predictor variables are all categorical, whereas in a linear model they are numeric. A multilevel model could have been applied in this analysis, but was not necessary for this data since our analysis in the previous section showed that the sample was mostly representative in terms of demographics for the population of interest, which was adults in Canada.

Before fitting a logistic model, there are a few assumptions that the data needs to satisfy. First, we assume that responses are independent from person to person. Although we discussed previously that the lack of transparency on data collection may be hiding some bias, it is unlikely that this would be significant. We also assume that there is little to no multicollinearity between independent variables in our model. This will mean that the variables we select to consider in the model should not be dependent on each other. This is an important consideration because many of the questions in the survey ask users to self report on different feelings they have, for example how often they feel worried or sad, and these types of responses will likely not be independent of each other. This means that we will focus on more demographic factors like income and location, which will likely have some association, but not significant enough to invalidate the model assumptions. Lastly, the number of participants was large enough to assume all sample size requirements were satisfied for a logistic model with the number of parameters being considered (Assumptions of Logistic Regression, 2020).

We identified previously in the data section that age, gender, income, region, and type of region were likely factors that had a high impact on the outcome of the predictor variable, whether the person had anxiety or not. As mentioned in the model assumptions, we will only look at the variables for demographic questions

and not the variables for personal questions in this survey. We grouped age into 3 categories, as shown in Figure 1, spanning 20 years each. In the survey data, the age was reported in groups of 10 years each (for example 50-59), but there was also a column in the dataset categorizing age into these 3 groups. This was the best option since the impact of age is better observed with these larger groups, and the resulting model will be less complex since there are fewer age groups to account for. The same thing has been done with income, splitting it into 4 groups: less than 40k a year, 40-79k, 80k-120k, and above 120k. This creates four reasonably different income classes that we can consider to be used for the logistic regression model. Another factor we considered is region, which is grouped into 5 categories that include provinces that are located close to each other. The largest provinces, Ontario, British Columbia, and Quebec, have their own group. Lastly, the type of region will either be rural, urban, or suburban.

To simplify and limit the model to only the most important features, we will use stepwise AIC selection, which is a common method in statistics for selection of independent variables. In this algorithm, linear models are sequentially compared and evaluated using the Akaike Information Criterion (AIC), which is a value that should be minimized for the optimal model selection. In each step of the algorithm, either variables that decrease AIC are added, or variables that increase AIC are removed from the linear model (Portugués, 2020).

The equation for the logistic model after running stepwise selection is:

$$(1) \quad \log \left[\frac{P(\text{anxiety_binary} = 1)}{1 - P(\text{anxiety_binary} = 1)} \right] = \beta_1(\text{Age}_{40-59}) + \beta_2(\text{Age}_{60+}) + \beta_3(\text{MaleNot Male}) + \\ \beta_4(\text{Income}_{40-79k}) + \beta_5(\text{Income}_{80-119k}) + \\ \beta_6(\text{Income}_{\text{Less than 40k}}) + \beta_7(\text{Income}_{\text{Prefer not to answer}}) + \\ \beta_8(\text{Area}_{\text{Suburban area}}) + \beta_9(\text{Area}_{\text{Urban area}})$$

The modelling in this section was done using RStudio 1.3, using the packages **survey** (Lumley, 2020) and **tidyverse**. Packages for creating plots and tables included **ggplot2** and **jtools**. Stepwise selection was done with the **stepAIC** function in the package **MASS** (Ripley, n.d.).

Results

The exact coefficients and their p-values for the model described in Equation 1 are listed in the table in Section B of the Appendix. In Figure 3 below, the left and right plot show the distribution of coefficients in the model, with the coefficients above 1 having a positive impact on anxiety, and coefficients below one having a negative impact. The plot below shows only the variables selected by the stepwise AIC method from the initial subset of all categories for age, income, region, type of region, and gender. The stepwise selection ended up discarding every category for region, which was originally organized by groups of provinces. Only the highest of the 5 income groups was discarded from the model, with every other income group being included, signalling that it is likely an important factor. The stepwise algorithm also included the older two groups out of the 3 age groups. Two types of regions, urban and suburban, were also included with only the rural category being discarded.

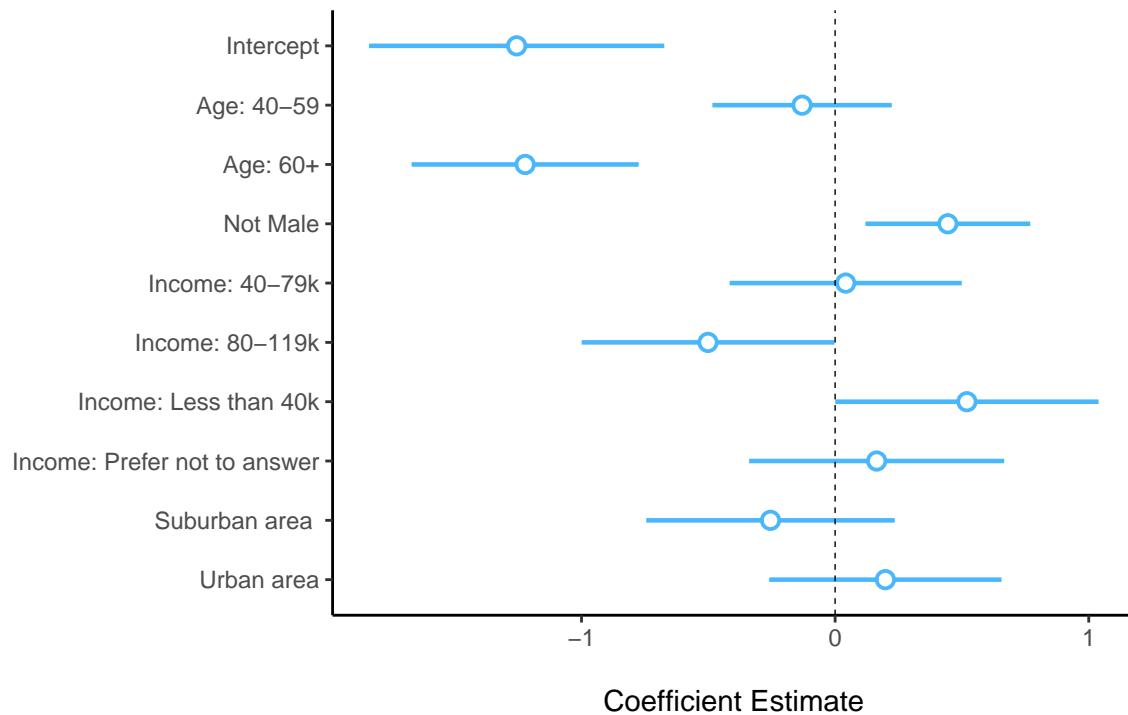


Figure 3: Coefficient Estimates for Stepwise Regression Model

The coefficients closer to the zero line in Figure 3 also had the highest p-values according to the table in Section B of the Appendix, which meant they were less significant. The most significant categories with p-values under 0.1 were age 60+, not male, income in range 80-119k, and income less than 40k. As expected, these categories had the highest magnitude positive or negative coefficients.

Looking at Figure 3, the variables that had a stronger positive impact on whether or not the person had anxiety were making under \$40,000 a year or not reporting income, living in an urban area, and being a woman or other non-male individual. The unreported income group also had a positive association with anxiety, which would likely include those who are unemployed. Conversely, the variables that had a stronger negative impact on anxiety were being over 60, making over \$80,000 a year, and living in a suburban area. From these observations, it would be reasonable to believe that lower income will contribute to more anxiety, as well as living in urban areas and being female.

Discussion

So far in this analysis we have looked at different factors that impact whether or not a person has anxiety. After looking at trends in the data and factors we believe may be impactful based on domain knowledge, we decided to focus on age group, gender, income group, geographic region, and type of region for a logistic model. We then used the stepwise AIC algorithm to further decrease the number of categories to only the most impactful. From the final model, we see that the most important factors that positively impacted anxiety were having lower income, living in an urban area, and gender, whereas being older, having higher income, and living in a suburban area had a negative impact on anxiety.

An interesting observation from our model results was that being in the highest age group (60+) had a negative impact on anxiety, with the largest negative coefficient in the model. We saw this in the Data section when the oldest consistently had the least anxiety through all 5 iterations of the survey, which was

surprising given that they are the most vulnerable population to death from covid-19 (COVID-19 and Your Health, 2020).

The three factors with the highest coefficients that would increase chances of having anxiety were having under 40k in income, identifying as not male, and living in an urban area. In the data section we mentioned that the percentage of females with anxiety is significantly more than the percentage of males with anxiety in the general population (Bandelow, 2015). There are many factors that could cause this, but the impact of covid-19 on any gender specifically is less obvious than some of the other factors. For example, living in an urban area was also a factor that contributed to anxiety, and it makes more sense since urban areas have higher densities of people, meaning the virus can spread more, and the lockdown rules are likely stricter. Lastly, having an income under 40k had the greatest positive impact on anxiety. In some parts of Canada, this salary would be at or below the poverty line (Neustaeter 2020), and financial instability is a known source of anxiety for many people. Even the next income group, 40-79k, had a positive model coefficient, meaning being in that group would make someone more likely to experience frequent anxiety. Depending on the region of Canada a person is living in, a salary of 40-79k may be equivalent to a salary of under 40k in an area with lower cost of living.

In the charts below, we look at responses from the most recent survey iteration in September to the questions, “How worried are you about the impact of COVID-19 on your personal financial situation?” and “How have physical distancing measures due to the COVID-19 pandemic affected your employment situation?” from only the people who reported having frequent anxiety.

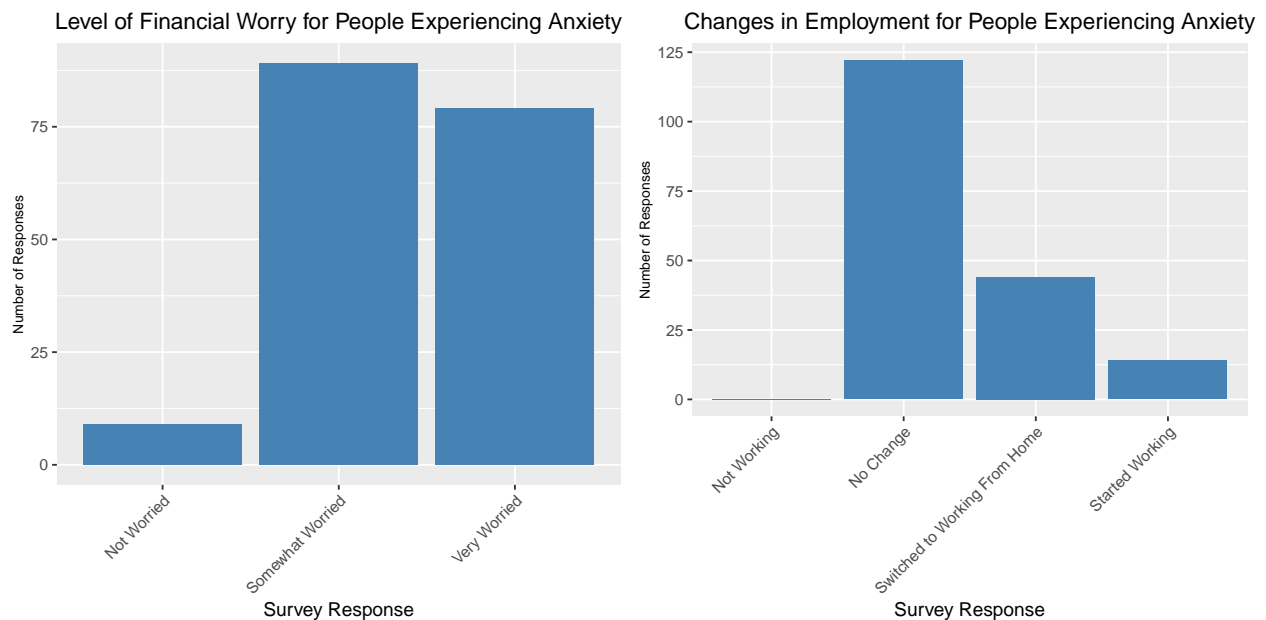


Figure 4: Frequency of Answers to Questions on Financial Status from People Experiencing Anxiety

We can see from the left frequency chart that almost everyone with anxiety is worried about their financial situation, and everyone has some job, whether they work from home or in person. This could indicate that people are likely not earning enough to feel financial secure, or that the status of their employment is unstable during the pandemic. This may also be a reason why the 60+ age group felt less anxiety, since many of the respondents would be retired or close to retiring, meaning they likely already have adequate savings. Financial instability has been a pain point of the working class for a long time, with the average Canadian having under 1000 in savings before the pandemic (MacGee, 2020). With the pandemic causing many businesses that rely on in-person interactions to lose money, many people are even more worried about their finances. While there have already been measures like the Canadian Emergency Response Benefit (MacGee, 2020) put in place to provide financial aid, these types of grants will likely not provide a sense

of financial stability to those most in need. From a mental health standpoint, it is important to support working adults and families in every way possible.

One weakness of the survey data and therefore the analysis was the lack of questions on the participant's prior mental health history. For example, if the participant was seeking mental health resources prior to the pandemic there would likely be a higher chance of them doing so during the pandemic as well. Having this data would allow us to draw a more direct conclusion on the impact of the pandemic on mental health.

Another related area of weakness is that we considered a fairly small subset of the variables in the dataset, and we could have included questions with more subjective responses. For example, when we considered financial status we only looked at the income category, but the survey also included questions on how worried a person was about their finances, and how covid-19 has changed their employment status. Similarly for health concern, we only looked at the person's age group, while we could have looked at questions asking how at risk of covid-19 they are through their jobs, or how much members of their household are at risk. These factors would have made for a more interesting analysis on the causes of anxiety. In a future analysis of this data, these types of variables may be suited for a different type of model besides a logistic regression model shown in this paper.

References

- Assumptions of Logistic Regression. (2020, June 22). Statistics Solutions. <https://www.statisticssolutions.com/assumptions-of-logistic-regression/>
- Bandelow, B., & Michaelis, S. (2015). Epidemiology of anxiety disorders in the 21st century. *Dialogues in clinical neuroscience*, 17(3), 327–335. <https://doi.org/10.31887/DCNS.2015.17.3/bbandelow>
- COVID-19 and Your Health. (2020, February 11). Centers for Disease Control and Prevention. <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html#:~:text=The%20risk%20for%20severe%20illness,older%20adults%20at%20highest%20risk.&text=need%20to%20know-,Risk%20for%20severe%20illness%20with%20COVID%2D19%20increases%20with%20age,increase%20risk%20for%20severe%20illness.>
- COVID-19 National Survey Dashboard. (2020, December 1). CAMH. <https://www.camh.ca/en/health-info/mental-health-and-covid-19/covid-19-national-survey>
- COVID-19 pandemic adversely affecting mental health of women and people with children. (2020, October 14). CAMH. <https://www.camh.ca/en/camh-news-and-stories/covid-19-pandemic-adversely-affecting-mental-health-of-women-and-people-with-children>
- Government of Canada, S. (2020, September 29). Population and demography statistics. Retrieved December 09, 2020, from https://www.statcan.gc.ca/eng/subjects-start/population_and_demography
- Lumley, T. (2020). Package “survey.” <https://cran.r-project.org/web/packages/survey/survey.pdf>
- MacGee, J. (2020, November 27). The Heterogeneous Effects of COVID-19 on Canadian Household Consumption, Debt and Savings. Bank of Canada. <https://www.bankofcanada.ca/2020/11/staff-working-paper-2020-51/>
- Neustaeter, B. (2020, September 15). Canada “very well might be” starting its second wave of COVID-19, doctor says. CTV News. <https://www.ctvnews.ca/health/coronavirus/canada-very-well-might-be-starting-its-second-wave-of-covid-19-doctor-says-1.5105105>
- Portugués, E. G. (2020). Model selection. Notes for Predictive Modelling. <https://bookdown.org/egarpor/PM-UC3M/lm-ii-modsel.html>
- Ripley, B. (n.d.). stepAIC Function. R Documentation. Retrieved December 21, 2020, from <https://www.rdocumentation.org/packages/MASS/versions/7.3-53/topics/stepAIC>
- Slaughter, G. (2020, May 1). Anxiety and depression have spiked among Canadians: survey. CTV News. <https://www.ctvnews.ca/health/coronavirus/anxiety-and-depression-have-spiked-among-canadians-survey-1.4919741>

Appendix

A: Raw distributions of predictor variables

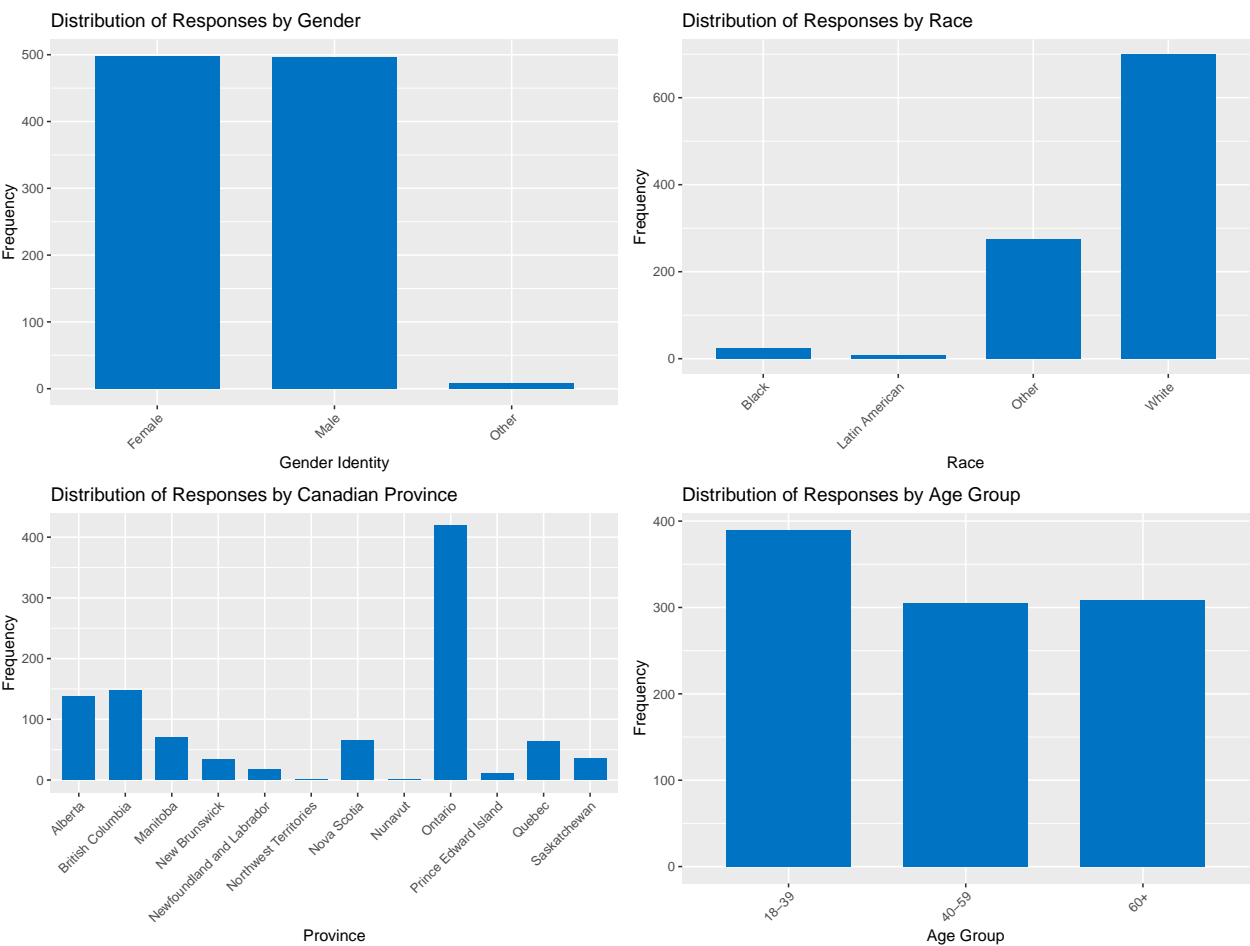


Figure 5: Category Distributions by Variable

B: Model Coefficients

Table 3: Survey-based estimates of logistic model of Factors impacting Anxiety

	Coefficients, P-values
Intercept	-1.26 *** (p-value = 0.00)
Age: 40-59	-0.13 (p-value = 0.47)
Age: 60+	-1.22 *** (p-value = 0.00)
Not Male	0.44 ** (p-value = 0.01)
Income: 40-79k	0.04 (p-value = 0.86)
Income: 80-119k	-0.50 * (p-value = 0.05)
Income: Less than 40k	0.52 * (p-value = 0.05)
Income: Prefer not to answer	0.16 (p-value = 0.52)
Suburban area	-0.25 (p-value = 0.31)
Urban area	0.20 (p-value = 0.40)
N	1003
AIC	980.97
BIC	1030.08
Pseudo R2	0.10

Standard errors are heteroskedasticity robust. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.