Factors Impacting Mental Health of Canadians during the COVID-19 Pandemic*

Anna Li

27 April 2022

Abstract

Mental health is a prevalent issue that the COVID-19 pandemic has impacted. Many studies have been done in the past regarding mental health and its relating factors. Yet, there are few studies that look at these identified relationships under the context of COVID-19, which has caused drastic changes in lifestyle and society. This paper investigates Canadians during the pandemic, utilizing the Canadian Perspective Series Survey 6 created by Statistics Canada to analyze the trends in alcohol consumption, stress, demographic factors and their relationship to mental health. Results reveal that stress and increased alcohol consumption negatively impact mental health while being employed, male, and older leads to a positive impact on mental health.

Keywords: mental health, stress, alcohol, demographic factors, COVID-19

Contents

*Code and data are available at https://github.com/annadlli/canada-mentalhealth.git.

1 Introduction

Mental health is an area of health that has become increasingly important as people suffer the consequences of the COVID-19 pandemic. The importance of a person's emotional and mental well-being can be seen in the everyday actions of a person throughout their life. Their mental state constantly affects their thoughts, actions, and quality of life as well. With the drastic changes to lifestyle that COVID-19 has brought upon the world along with its deadly consequences, many people suffered mental health, with anxiety and depression increasing by 25% globally "Covid-19 Pandemic Triggers 25" (n.d.). As the physical health situation for COVID-19 is gradually improving with the development of vaccines and health measures, it is important to also focus on the area of mental health as well.

This paper looks at what mental health was like due to COVID-19, particularly factors that can influence one's mental health. From this, it provides insight for effective actions that the government might enact in policy, as well as the areas and demographics to target for the most impact. With this newfound knowledge of the current trends and prevalent factors, then the Canadian government can best address the various issues that have arisen with the COVID-19 virus.

This paper fills the gap of the few works available in recent years regarding mental health and its related factors to a multitude of factors. Most studies were done in the field focus on mental health and its direct relationship with the pandemic, as well as mental health's relationship with either stress or alcohol. This paper looks at some of the identified key factors pre-pandemic and how their relationship with mental health comes to play altogether, as well as how this relationship may have changed due to the impact of COVID-19.

This paper focuses on 3 factors that impact mental health: alcohol, stress, and demographic factors. This was chosen to provide an overview of the different areas that relate to mental health: behavioural, mental state of mind, and social and biological characteristics. Literature has revealed that pre-pandemic, alcohol and other substance usage seems to be correlated with a higher risk of having harm as well as mental health disorders Young (2011). While a large proportion of Canadians do consume alcohol in moderate proportions in the past, it seems that the pandemic has caused alcohol consumption behaviour to change as alcohol-related hospital visits increased by 50% during the pandemic Thomas (2012) & Vaneer (2022). Different genders, age groups, and employment statuses can also impact the life of a person, including the way that they view things. This, consequently, impacts their mental health status as well. Stress itself is known to be a factor that is commonly associated with mental disorders such as depression and anxiety. Whether these relationships still exist or have changed due to the drastic effects of COVID-19 is something that this paper would like to determine.

The paper utilizes a combination of graphs, summary tables, and a linear regression model to investigate mental health and its relating factors. The main result is an investigation into the relationship mental health has with different factors, providing insight into what main factors can influence mental health. Results reveal that changes in alcohol consumption behaviour, sex, age, stress, and employment status are all significant factors that can influence a respondent's proclaimed mental health level. Therefore, to maintain high mental health, there are both factors that one can change and others that are quite inherent to nature. The factors more inherent to nature help alert which part of the population should pay extra attention to maintaining their mental health. If the respondent is female, unemployed, and of young age, they should be more aware of their mental health and spend more effort in maintaining a good healthy mindset. The behavioural factors help demonstrate ways that one could try to improve their mental health, such as by decreasing alcohol consumption, trying to get a job, and finding ways to lower stress.

The remainder of the paper is split into five sections. Section 2 explains the data source and collection methodology, potential bias and ethics issues, as well as the selected data's characteristics. The data cleaning process is briefly mentioned as well. Section 3 discusses the methods used to produce the wanted results of the change in frequency of alcohol consumption and the relationship between alcohol consumption and mental health. Section 4 presents the findings from applying the model. Section 5 discusses what the findings in results mean, the limitations in the findings, as well as next steps to enhance and expand on the findings of the paper. Section A includes testing assumptions for regression, additional figures, as well as a datasheet to shed some more insight into what was utilized in the paper.

2 Data

This report was created using the R statistical programming language R Core Team (2020). Analysis and figures were created with the packages ggplot2 Wickham (2016), knitr Xie (2014), kableExtra, Zhu (2021), gtsummary, Sjoberg et al. (2021), patchwork Pedersen (2020), and tidyverse Wickham et al. (2019). Package car Fox and Weisberg (2019) was utilized in the appendix to check the regression model. Cleaning utilized Wickham and Bryan (2019) to transform the raw dataset.

The dataset was obtained from Statistics Canada, with variables relabeled and cleaned according to the provided documentation that accompanied the dataset to eliminate skipped and not stated answers.

2.1 Data Source and Collection

The paper utilizes 2987 observations from the Canadian Perspectives Survey Series: Substance Use and Stigma During the Pandemic survey. The dataset was obtained from Statistics Canada, available as a public use microdata file.

The Canadian Perspectives Survey Services was a series of surveys conducted to better understand the current social issues Canadians were facing. Participants were chosen through probability sampling and were allowed to choose whether they wanted to participate in the survey. Each survey spanned the period of one month for its online data collection. To ensure that the sample represents Canada, Statistics Canada adds weights to the various categories to best simulate the overall situation of Canada. Specifics regarding the dataset can be seen in the appendix, with the datasheet A.1. The codebook guide to the original dataset, along with additional information regarding weights and other potential information of interest, is available when downloading the public microdata file from Statistics Canada.

2.2 Ethics and Bias

Due to the way data was collected, sampling and non-response errors both existed in the survey. This was due to the voluntary participation as well as the options given to the participants to skip questions. While this was mitigated by the utilization of weights for different categories, a study was done by Statistics Canada to verify the bias that existed in the survey after adjustment. There was still some bias that remained, specifically underrepresentation of people working in agricultural occupations as well as an overrepresentation of those with a post-secondary certification, with children, born in Canada, and working in nature and applied science occupations Government of Canada (2022). This should be taken into consideration when interpreting the results of the paper.

Another potential issue is non-sampling errors. Non-sampling errors can be due to participants not responding or giving inaccurate responses to questions. It is possible by how the survey was phrased that misunderstanding could have occurred, leading to a "wrong" answer that is not representative of what the respondent actually undertook. Moreover, it is possible that different interpretations of the various options, such as the difference between "fair" and "good," could also lead to inaccurate results. Moreover, the variable sex contains only the choices of "male" or "female." It is possible that those of nonbinary gender would have difficulty accurately representing themselves for this question.

Finally, an issue regarding the survey is its nature, which is the investigation of substance usage. Substance usage is something that is typically viewed in a negative light by society. It is possible that respondents would prefer to underreport their current situation, such as the amount of alcohol they consume. This could potentially introduce another source of error to the data.

2.3 Data Characteristics

9 variables were used to conduct the analysis within the report. The variables covered different aspects that could have a relationship with mental health: behaviour, mental, and demographic. The model focuses on 6 factors that potentially impact mental health: alcohol consumption, change in alcohol consumption behaviour, stress level, sex, age, and employment status.

Alcohol consumption was measured in terms of the self-reported results of how many drinks a respondent would have in one sitting if they were to drink. Change in alcohol consumption determines whether the respondent has displayed different alcohol consumption levels this month in comparison to their past, with five different categories: one to four drinks, five to nine drinks, ten to nineteen drinks, twenty or more, or other situations. Other situations include those who were able to skip this question during the survey for a valid reason, such as never drinking alcohol. Sex, employment status, and stress level were all made into binary variables. Stress level was measured in two broad categories, with respondents either having no to little stress or experiencing a significant amount of stress due to the pandemic. Respondents were also split into the two categories of either male or female and unemployed or employed. Age was split into seven groups: 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75 years and older.

Other variables utilized include the change in mental health and change in stress. These variables ask, in comparison to the time before the COVID-19 pandemic, how have mental health and stress levels changed? Change in stress level is measured on a scale of one to five, with one being no stress and five being extremely stressful. Change in mental health was measured similarly on a scale of one to five, with one being much worse and five being much better. Perceived mental health, the desired dependent variable, is also measured on a scale of one to five, with one being poor and five being excellent.

To arrive at the final dataset, the additional variable number of drinks was utilized in order to keep respondents who do not drink alcohol. In comparison to the original dataset, respondents who did not answer the questions coded to be "not stated" responses were eliminated from the final dataset.

2.3.1 Alcohol: General Population Consumes a bit of Alcohol, with Drinking Habits being Consistent

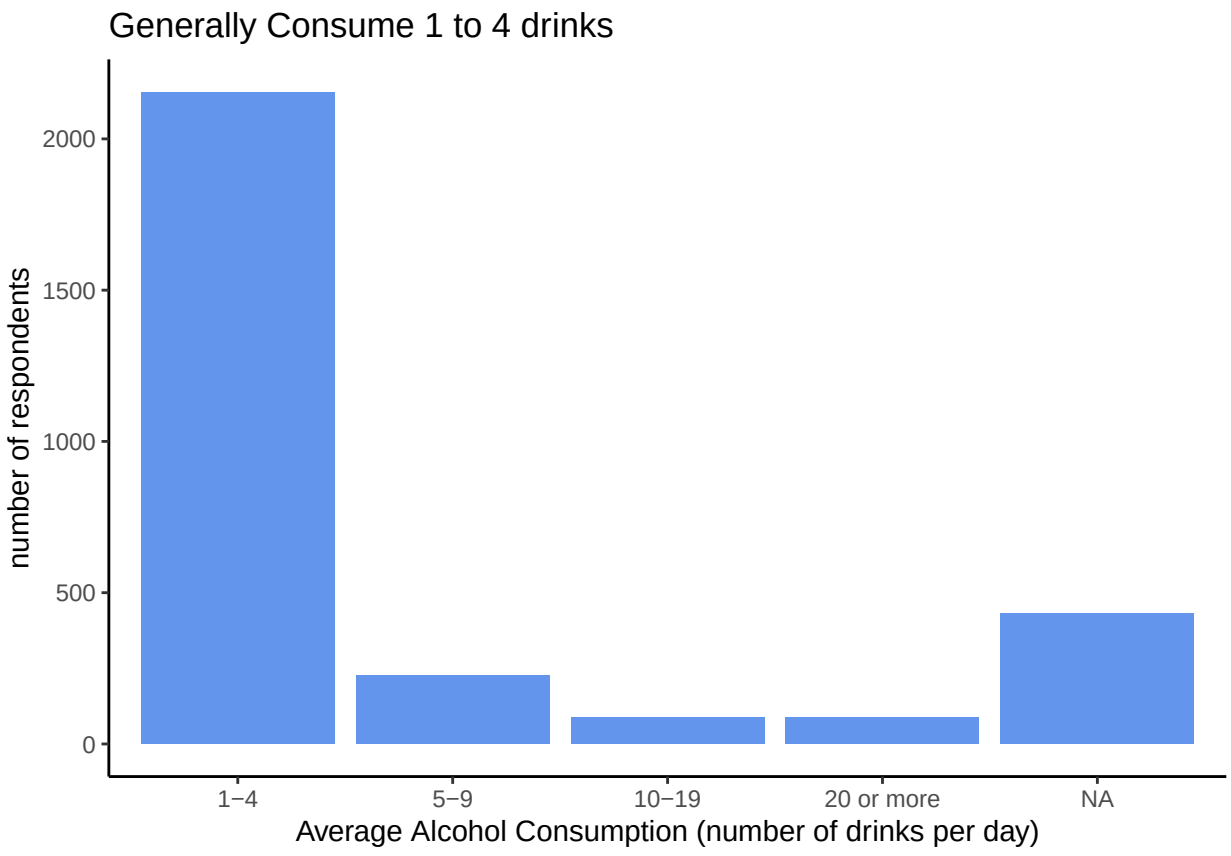

Figure 1: Alcohol consumption levels are generally low.

Figure 1 is created through ggplot(Wickham (2016)). The bar plot describes the respondent's average alcohol consumption for the last month. The majority of the respondents drank 1 to 4 drinks a day as their average alcohol consumption level for whenever they drank. There is also a significant amount of people for whom this question was inapplicable, such as not drinking in the last month.

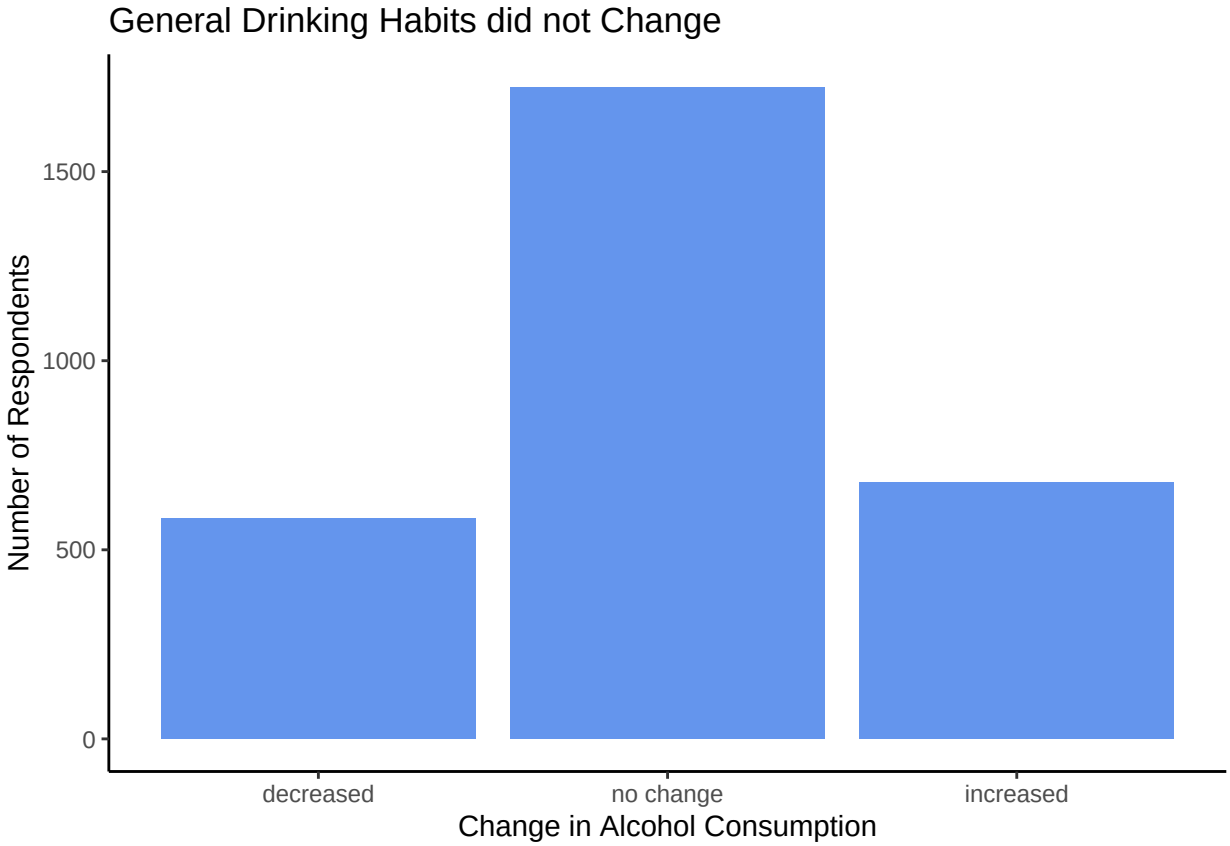

Figure 2: Alcohol consumption change due to COVID-19 is relatively equal for increase and decrease.

Figure 2 is created through ggplot(Wickham (2016)). The barplot demonstrates how drinking habits have changed due to the pandemic. The majority of the respondents experienced no change in their alcohol consumption, with a relatively equal proportion of respondents who answered that they both decreased or increased their alcohol consumption levels.

2.3.2 Mental health: Respondents Feel Good, with the Pandemic Causing Only a Slight Decrease in Mental Health

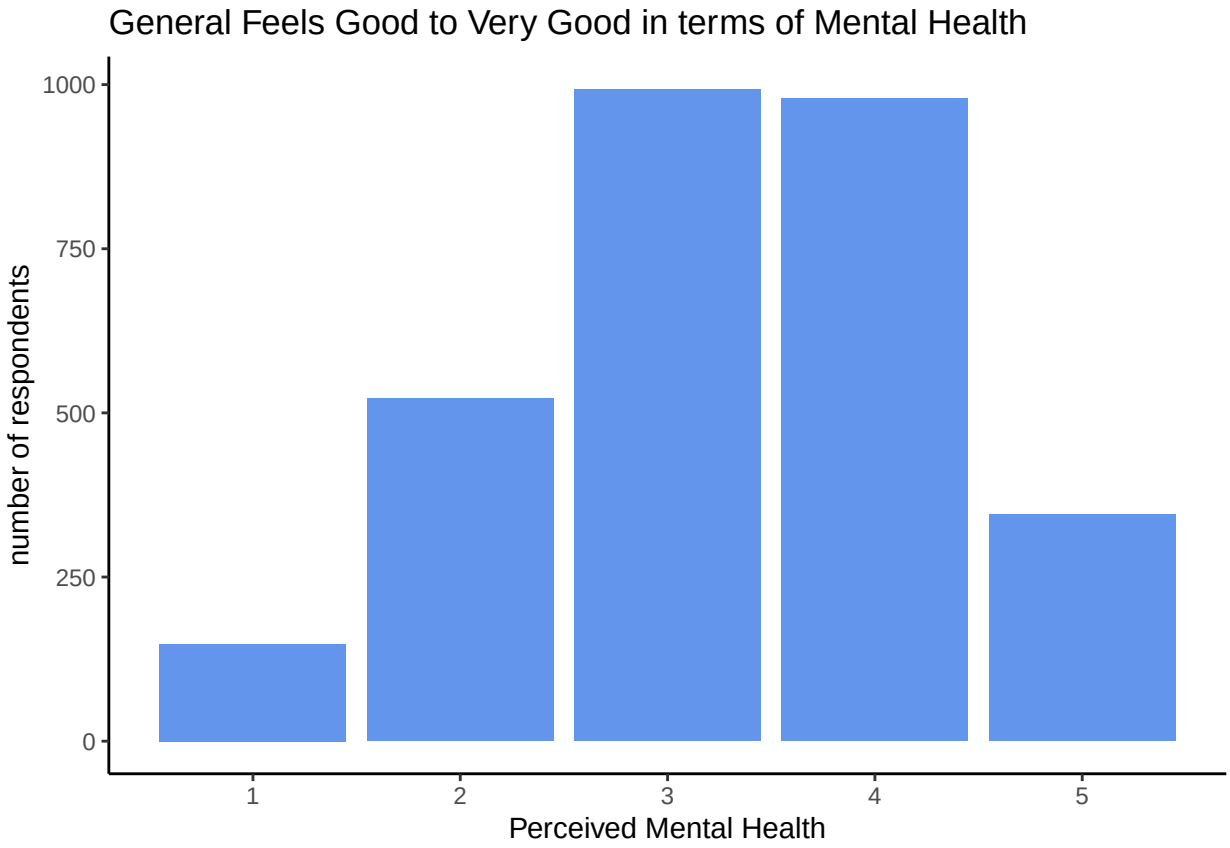

Figure 3: Mental health distribution shows more positive responses.

Figure 3 is created through ggplot(Wickham (2016)). The barplot demonstrates the distribution of the desired results, perceived mental health. 1 represents feeling terrible, and 5 feeling the best. The majority of the respondents gave answers of 3 or higher, feeling at least good about their mental health.

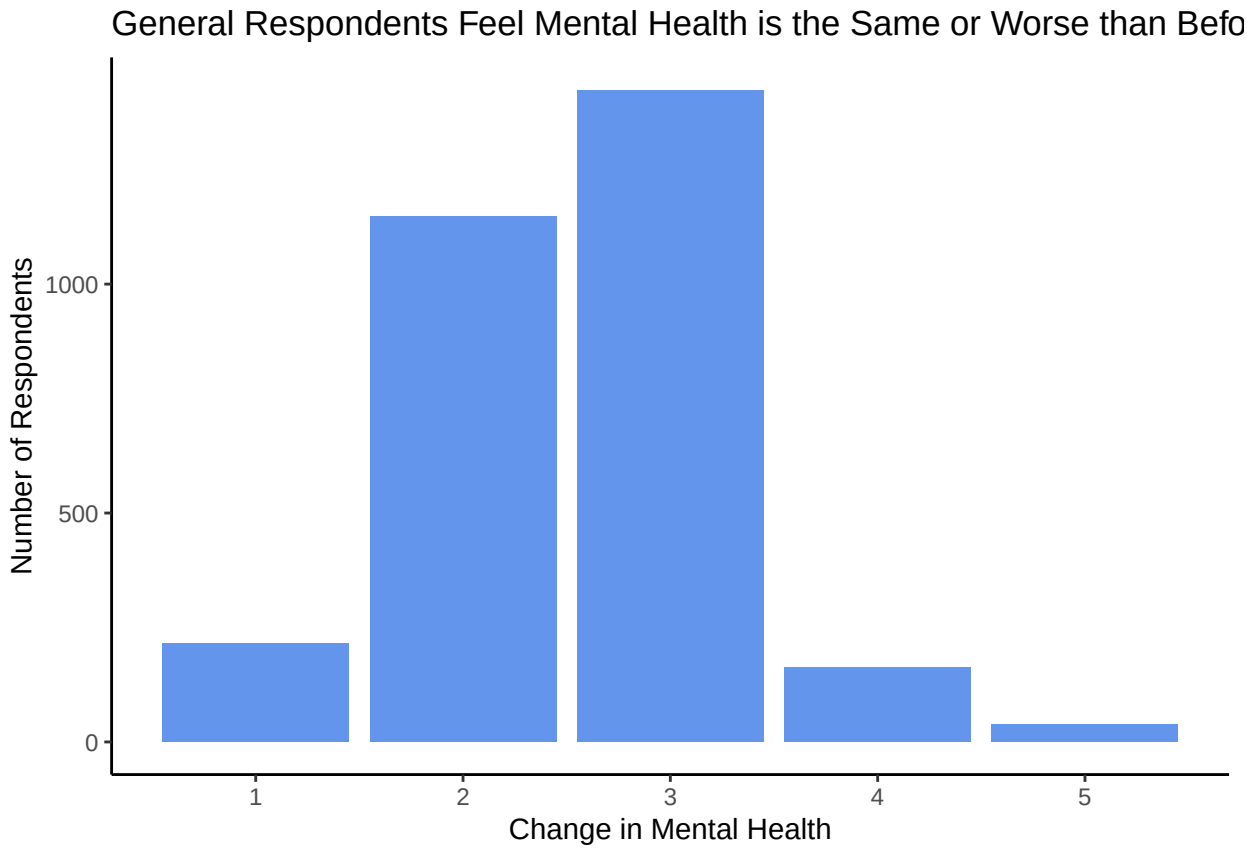

Figure 4: Change in mental health due to COVID-19, if occurred, are for worse.

Figure 4 is created through ggplot(Wickham (2016)). This barplot looks at how respondents feel that their mental health has changed due to the COVID-19 pandemic. The majority of the respondents answered 3 or below, feeling that the pandemic has caused them to have the same or worse mental health.

2.3.3 Stress: COVID-19 Caused a Slight Increase in Stress, but not by Significant Amounts

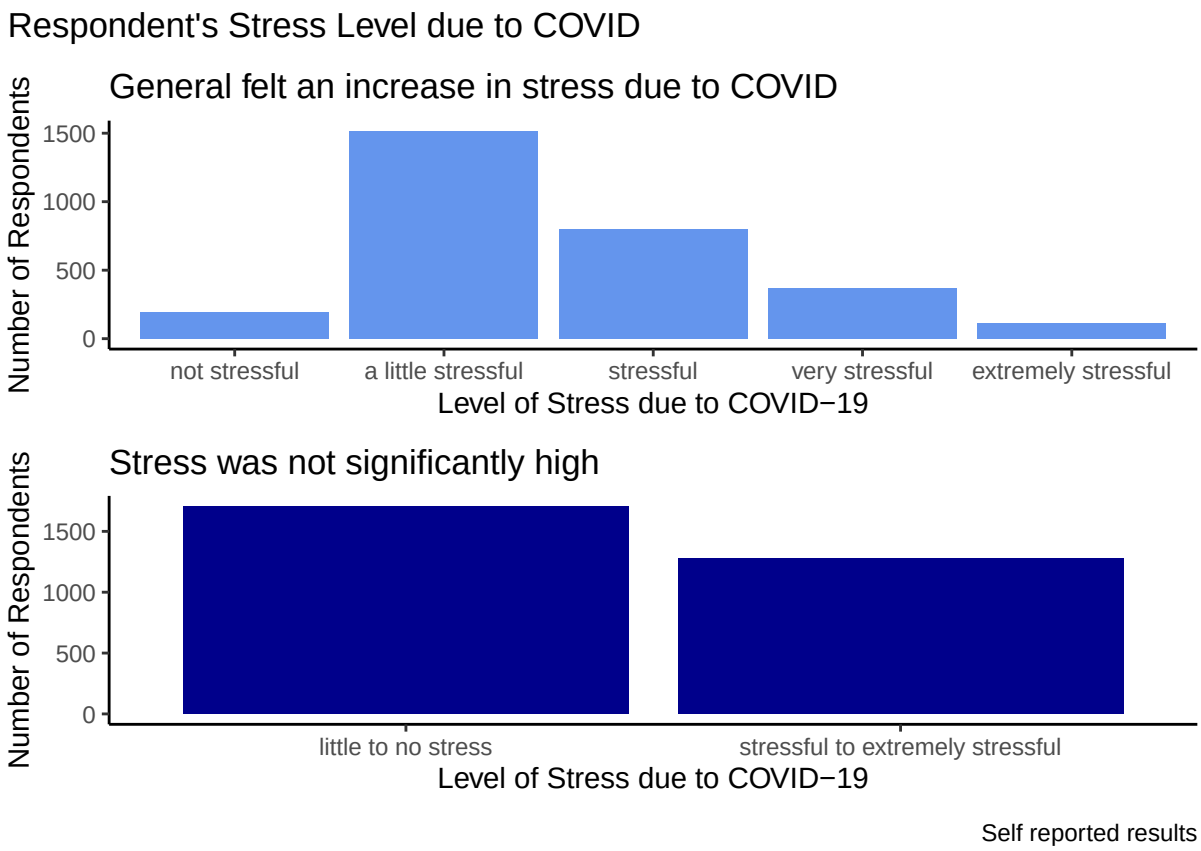

Figure 5: Stress due to COVID has increased.

Figure 5 is created through ggplot(Wickham (2016)) and patchwork (Pedersen (2020)). The 2 bar plots describe respondents' stress levels due to COVID-19. The majority of the participants feel at least a little bit of stress due to the pandemic. However, when the different levels were generalized to two categories, there was a slightly larger proportion of respondents who felt no to little stress in comparison to those who experienced great stress.

2.3.4 Summary Proportion Table: Different Demographic Proportions for Different Age Groups

Table 1: Data set: Proportion of Respondents by Age, Sex, and Employment Status

	Male	Female	Unemployed	Employed
15-24 years old	0.3261	0.6739	0.3261	0.6739
25-34 years old	0.4416	0.5584	0.1247	0.8753
35-44 years old	0.4751	0.5249	0.1113	0.8887
45-54 years old	0.4618	0.5382	0.0984	0.9016
55-64 years old	0.4846	0.5154	0.4146	0.5854
65-74 years old	0.5000	0.5000	0.7781	0.2219
75 years and older	0.5236	0.4764	0.9372	0.0628

Table 1 was created using kable in knitr Xie (2014) and kableExtra package Zhu (2021). It reveals the proportions of the results in Table 3, the raw observations which can be found in the appendix. There are significantly different trends for the different age groups. Gender is relatively evenly split for all age groups except for 15-24, where females have a larger proportion of 70-30. 25-54 years old age groups show similar trends for employment, where around 90 percent of the respondents are employed. 15-24 years old have around 70 percent employed, which makes sense considering that at these younger ages, a large portion of the respondents would have been in school full-time. Similarly, those 65 and older have a significantly lower proportion of people employed as well (22 percent and 6 percent respectively for the 2 age groups of 65-74 and 75 years older). This also is reasonable considering the general age of retirement is 65.

3 Model

Using results explored in the data characteristics section as well as the literature review, variables were narrowed to ones that related significantly to mental health. To come to the results, two different approaches were used. First, visual bar plots were utilized to help provide insight into how the change in one variable seems to impact perceived mental health. Then, linear regression was used to specifically look at the relationships that the variables had with the desired mental health.

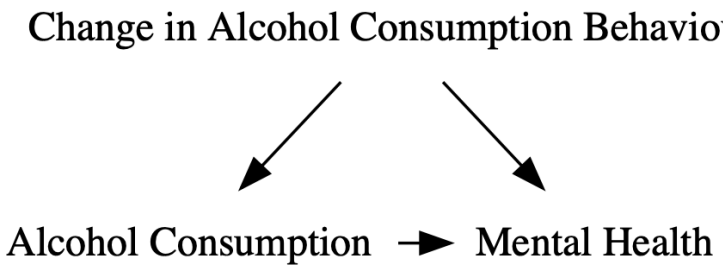

Figure 6: Diagram of mediator

Figure 6 is created by using DiagrammeR, Iannone (2022) reveals the interrelationship between the independent variables stress and alcohol, as well as the dependent variable of mental health. In order to account for the change in alcohol consumption behaviour's mediator's effect in the linear model, it is not controlled. This leads the model to include both terms for alcohol consumption and change in alcohol consumption behaviour.

To come to the model, power transformation, utilizing the package car, Fox and Weisberg (2019) was used to see if the response variable needed to be transformed. Then, the model was checked to see if for linearity assumptions, including utilizing the VIF (variance inflation factor) to check for outliers for all variables. For the checking of linear assumptions, please refer to the appendix Section A.2.2. The utilized model is as follows.

Perceived Mental Health $= \beta_0 + \beta_1 \text{alcohol} + \beta_2 \text{change in alcohol consumption} + \beta_3 \text{stress} + \beta_4 \text{age} + \beta_5 \text{sex} + \beta_5 \text{employment}$

4 Results

4.1 Alcohol and Mental Health

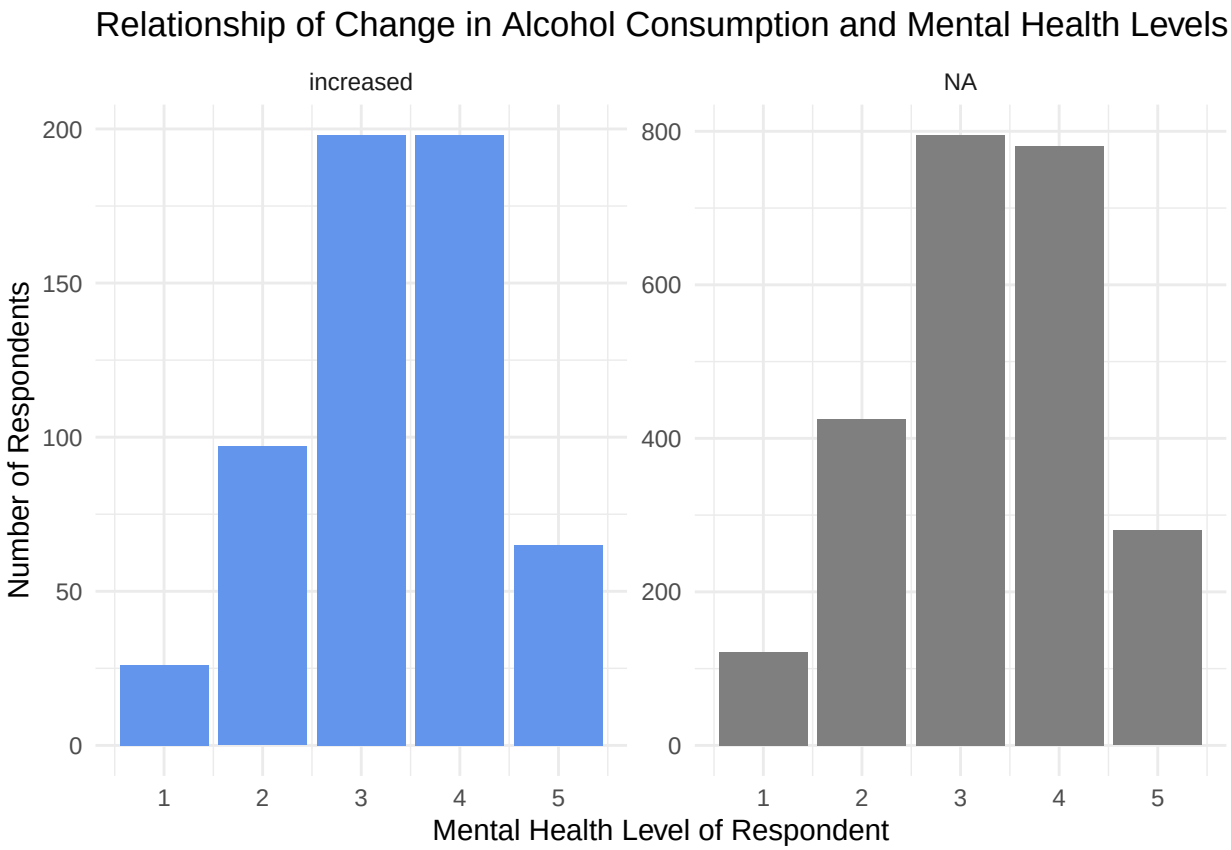

Figure 7: Overall similar distribution, with those who decrease alcohol consumption being generally higher scoring in terms of mental health.

Figure 7 compares the different mental health level distributions for the different possibilities of alcohol consumption behaviour. Those who decreased alcohol consumption have the highest proportion of respondents that is left-skewed, representing the positive mental health level of "good" and "very good."Those that increased alcohol consumption have a more right-skewed distribution, representing the mental levels of "fair" and "poor" more. For respondents who did not change their alcohol consumption habits, the general distribution is also more towards the positive mental health levels of "good and"very good."

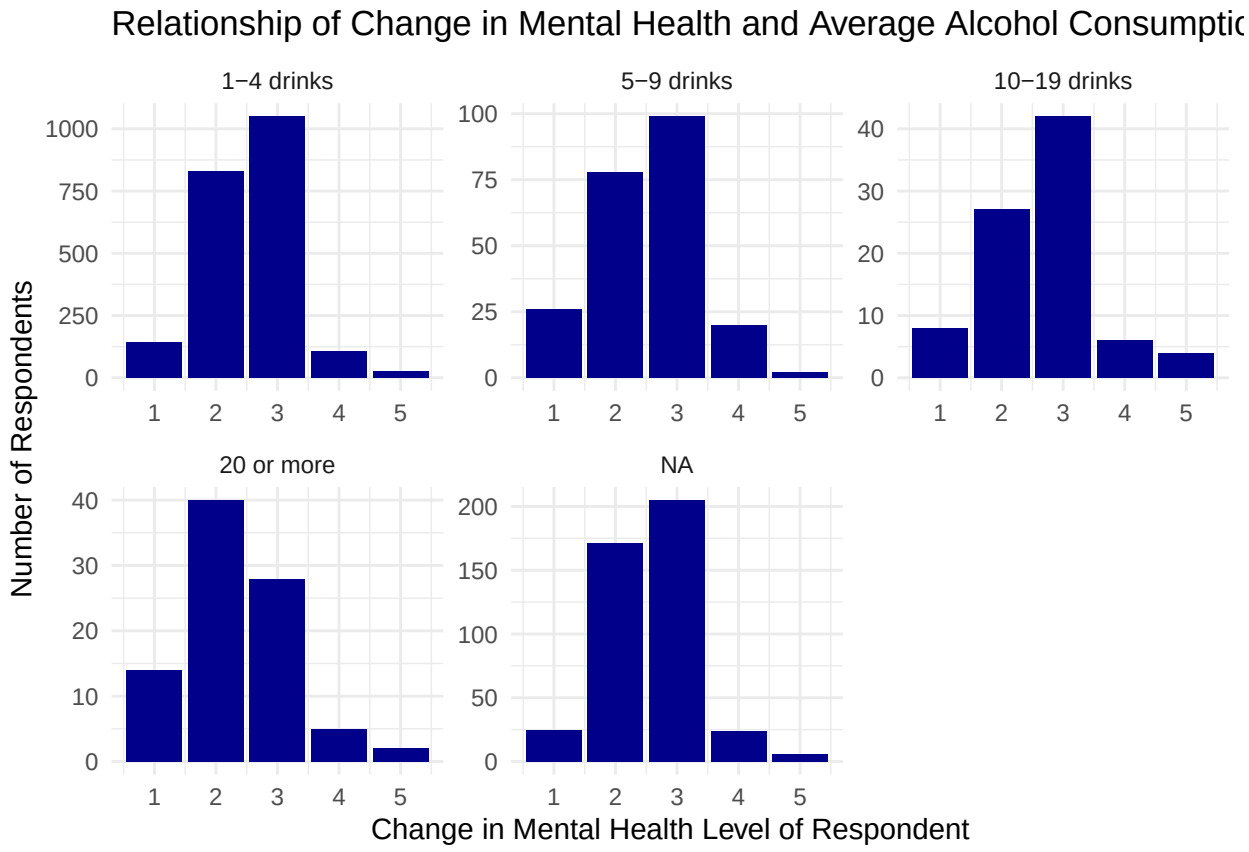

Figure 8: Those who drink more seem to have a more negative change in mental health.

Figure 8 compares how mental health has changed for respondents that have different amounts of alcohol consumption per sitting. In general, regardless of how many drinks a respondent would consume when they choose to drink alcohol, the most common response is that mental health is about the same before and after the pandemic. While there is no evident difference in distribution for the different levels of alcohol consumption, those who consume 20 or more drinks in one day when they are drinking show the highest proportion of respondents who were negatively impacted by the pandemic in terms of mental health.

4.1.1 Change in stress and Alcohol Consumption

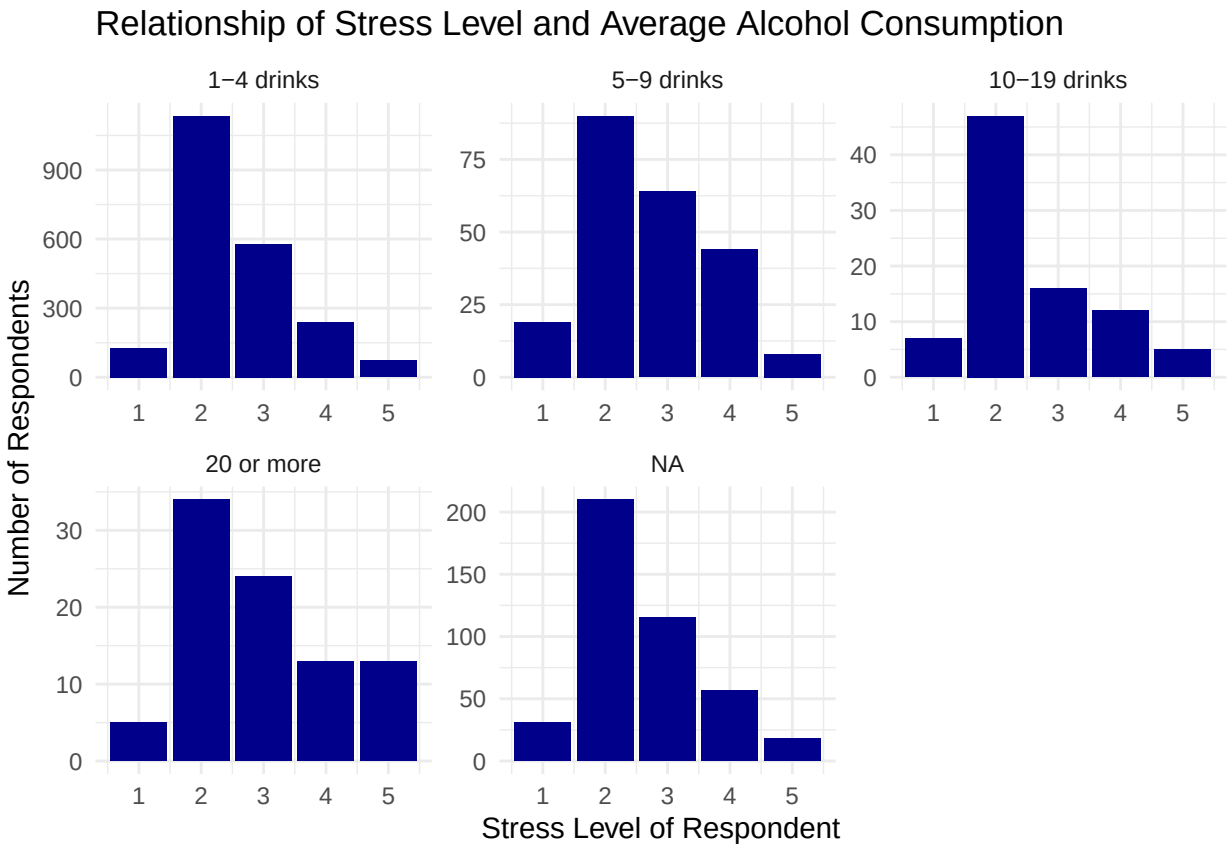

Figure 9: Those who drink more seem to have a more proportion of people who are more stressed.

Figure 9 compares the different levels of alcohol consumption with the stress levels reported by respondents. In general, for all levels of alcohol consumption, it was common for respondents to feel a little bit stressed from the pandemic. Those that drank 20 or more drinks in one go were most likely to feel very and extremely stressed due to COVID.

4.2 Stress and Mental Health

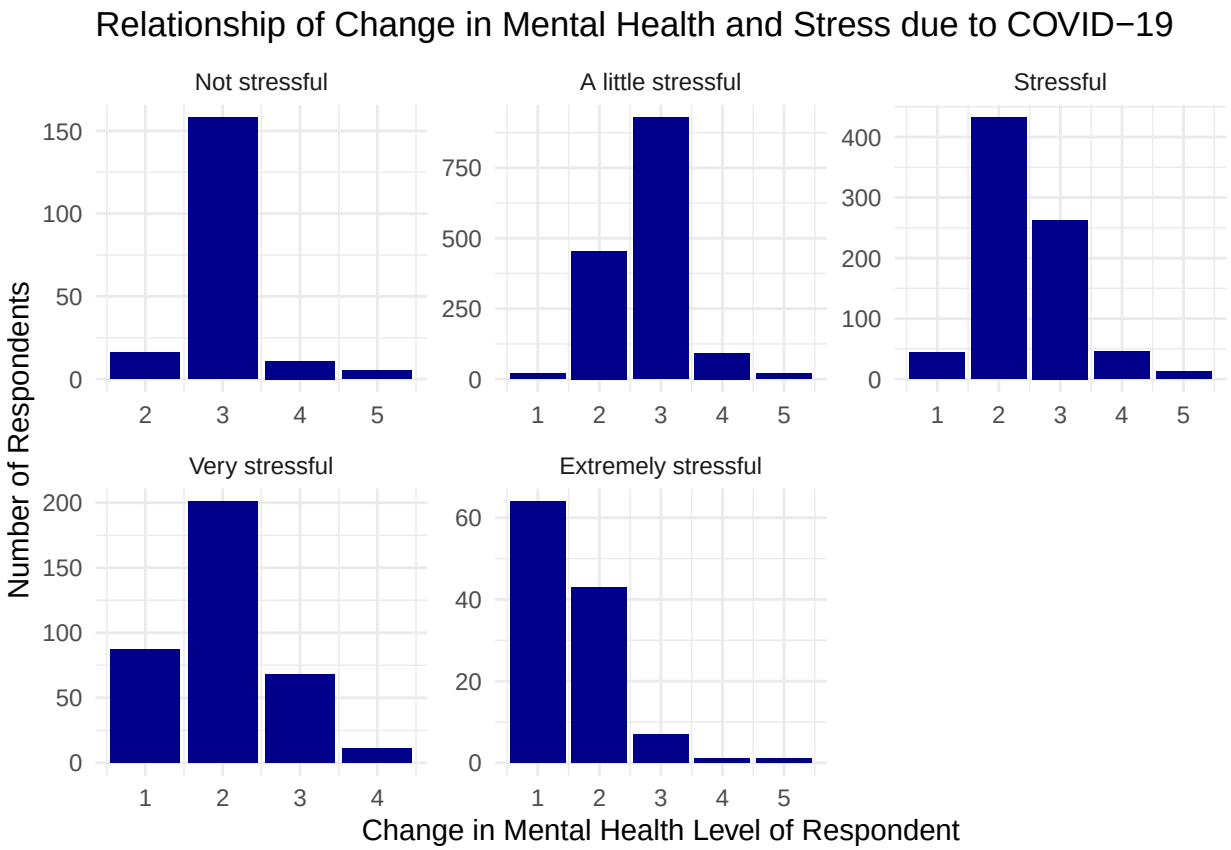

Figure 10: Those who felt stressful and above levels of stress answered that COVID had a more negative impact on their mental health.

Figure 10 compares the different levels of stress levels to the changes in mental health levels due to COVID-19 of the respondents. Those that felt extremely stressful and very stressful had a higher proportion of respondents who stated that their mental health worsened a lot due to the pandemic. Those who felt no stress or little stress generally felt that their mental health stayed the same, if not a bit worse than before. Overall, there are minimal respondents who felt that their mental health improved after the pandemic.

Table 2: Linear Regression Results

Characteristic	**Beta**	**95% CI**	**p-value**
alcohol	0.00	0.00, 0.00	<0.001
change_alcohol	-0.19	-0.25, -0.14	<0.001
stress	-0.85	-0.92, -0.78	<0.001
sex	-0.05	-0.11, 0.02	0.2
age_group	0.12	0.10, 0.14	<0.001
employment	0.16	0.08, 0.24	<0.001

4.3 Regression Model Results

Table 2 was created using gtsummary Sjoberg et al. (2021). Stress, sex, and change in alcohol consumption are negatively correlated with mental health. In other words, increasing alcohol consumption and stress levels lead to a decrease in mental health. Changing from a male respondent to a female respondent also leads to a decrease in mental health. Both stress and change in alcohol consumption are very significant. Alcohol consumption levels themselves had seemingly no effect on the perceived mental health responses. Age and employment status are positively correlated with mental health, both being very significant as well. In other words, increasing the age of the respondent and changing their employment status from unemployed to employed has a positive effect on mental health.

5 Discussion

5.1 Alcohol and Mental Health

The regression model results find the change in alcohol consumption behaviour to be negatively related to mental health. Decreasing alcohol consumption is associated with better mental health, and increased alcohol consumption leads to worse mental health. This corresponds to the idea of alcohol being used as a coping mechanism or escape from reality, with alcohol abuse being a type of mental disorder. Those who have a better state of mind would find less the need to use such a coping mechanism, leading to a decrease in alcohol consumption. It is worth noting that the habits of alcohol consumption behaviour of most respondents stayed the same despite the pandemic. When considering that most respondents consume alcohol in moderate amounts, 1-4 drinks, a decrease in alcohol consumption would mean consuming no alcohol. This, while possible, is more of a lifestyle change than just a decrease. Such drastic lifestyle changes would be much harder to make for a person. The idea of decreasing alcohol consumption leading to better mental health would be more applicable to those who consume higher amounts of alcohol, such as 10-19 drinks or even 20 drinks.

However, the amount of alcohol consumed is found to have no impact on mental health in the regression model. This is slightly different from the findings from observing figure 8. Those who consume much alcohol seem to be negatively impacted by the COVID-19 pandemic more in mental health. However, most respondents find their mental health to be the same as pre-pandemic, regardless of the level of drinking. This corresponds to the model's findings, where the amount of alcohol consumed did not affect mental health. Altogether, it would seem that changes made to alcohol consumption behaviour rather than the level of alcohol usage itself impact mental health more.

5.2 Stress and Mental Health

More stressed individuals also answered that their mental health was more negatively impacted by COVID-19. This relationship was identified in both the figures and the linear regression model utilized in the paper. Stress itself is considered a negative factor for mental health and is tied with other mental health issues such as depression and anxiety. The confirmation of the relationship and how the change in one affects the other confirms that the model is providing similar results to expectations.

When looking specifically at the overall distribution in figure 10, it seems that the relationship comes into effect when there are higher levels of stress, such as "stressful," very stressful," and "extremely stressful," where respondents answer that they experienced deteriorating mental health. For lower stress levels, the overall response is that mental health does not change due to COVID-19. This is potentially a reason for the significant result found in the model, where the variable for stress was used as a binary variable. The variable grouped no and little stress, as well as the higher level of stress, together into two categories. If the stress variable was taken into consideration with the 5 different categories instead of two, the results could be different. Nevertheless, as the five levels themselves are quite subjective, the grouping together of variables can actually provide more clarity as to stress' overall effect on mental health. As the paper provides a brief investigation of various factors, it is sufficient to satisfy the original intent.

People must understand the potential harm of stress. It is common in the current society for stress to be normalized and expected. While it is not harmful in small amounts, large amounts of stress are bad for mental health. Stress, moreover, is something that can come from multiple sources. It may not seem to be a lot initially, yet the build-up can lead someone to be extremely stressed and suffer mentally. Empathy and understanding should be increased, especially in the times of COVID-19, where stress is coming from more additional sources. While a large portion of the respondents felt that COVID-19 had caused them little to no stress, many respondents felt that COVID-19 led to a lot more stress. It is important to find a common understanding of how the same situation can affect people differently, leading to different responses. Moreover, people adjust at different rates. The government, schools, and companies should consider the varying circumstances that COVID-19 has caused people when creating policies, assignments, and benefits. In response, the public should understand that while they may not relate to or feel that certain policies are effective, it could be that they are targeting another subset of the population who are in need of such measures.

5.3 Demographic Characteristics and Mental Health

The regression model finds that gender is negatively correlated with mental health while employment status and age are positively correlated. Employment and age are very significant, while sex is less significant. Respondents who are women, unemployed, and of a young age are likely to have lower mental health in comparison to those who are men, employed, and of older age. This is reinforced by literature done in the past. In society, men face the masculinity stereotype of not showing any weakness. Even though this survey was conducted in a voluntary format, online without any others for peer pressure, it is possible that male respondents felt societal pressure to answer in a certain way that reflects "masculinity," such as more positive answers for mental health. Masculinity can be associated with the likelihood of some mental health disorders, such as aggression. However, the typical disorders associated with mental health, such as depression and anxiety, are more commonly associated with women Rosenfield and Mouzon (2013). It is highly likely that when respondents were asked about their mental health, they thought more of these internal aspects of mental health, leading to the negative relationship found in the model.

The relationship between mental health and gender could be potentially affected by age. Kiely, Brady, and Byles (2019) found that older women were more likely to experience common mental disorders, but the differences in gender narrow in younger age groups. This could be due to how society is trying to redefine gender roles and breaking stereotypes that often cause stressors that lead to poor mental health. Age overall shows a positive relationship with mental health from the model. It corresponds with the idea that there is less volatility and uncertainty regarding self-identity as one becomes older. With this better understanding of life, this stability likely causes a better state of mind, leading to better mental health. This is supported by literature findings, Hopman et al. (2009) where there is better mental health in older age groups and young adults suffer from depression and anxiety the most Weitzman (2004).

Regarding the findings for employment status, it corresponds to what is found in Modini et al. (2016), where good work and the role of work enhanced mental well-being and helped facilitate recovery from mental illness. Being employed provides both financial security as well as a social circle and a systemic daily schedule, reducing uncertainty which can cause much uneasiness. This effect is magnified by the COVID-19 pandemic, where people's daily lives were disrupted, and there existed a ton of uncertainty. Being employed would help combat many of the difficulties that COVID-19 brought, such as the feeling of isolation and instability.

However, the applicability of the positive impact of employment status could vary for the different age groups of respondents. Those who are young, such as 15 to 18 years old, and those of retirement age, such as 75, may not experience as much of a positive impact on mental health as those in middle age. This ties into the idea of when a person is typically supposed to start working in society. For those who start working at a young age, the time of schooling can be tied to other pressuring factors in life that may impact their mental health, such as financial burden. Those who are still working at ages beyond retirement face a similar situation. They are expected to be enjoying a peaceful time of rest but are still, for some reason, continuing to work. This is not deemed to be a normal circumstance by society, which could potentially lead to some burdens on mental health for being viewed as "odd," in addition to whatever circumstance the person may be facing at the time.

5.4 Weaknesses and Limitations

There are a few limitations that must be considered when considering the paper's results. While the model utilized was a linear regression model, not all assumptions were fully met. Overall, the data seems to fit a linear distribution, yet there is room for improvement. Trying to transform the response variables was unhelpful in solving the situation, as the results did not improve much after using powerTransform to change the dependent variables. Therefore, for the sake of interpretation and meaningful results, I decided to go forth with the linear regression.

The variable alcohol had a coefficient of 0 in the linear regression model, meaning it did not affect mental health. This contrasts with some of the literature in the field, where there is a correlation between alcohol usage and a higher risk of mental disorders Young (2011). The evidence of the results could be due to the effect of a mediator, such as the change in alcohol consumption. This is represented by the figure 6.

Moreover, despite the original dataset having weights that led to a pretty accurate representation of the different groups, cleaning was done to the original dataset to eliminate non-response observations. This could lead groups to be disproportionately represented within the used dataset for the analysis. Finally, the data was taken from the respondents over one month in 2021. For this study, I assume that it is enough to provide a glimpse into how COVID-19 has impacted the lives of Canadians. However, with the different waves of COVID-19 and the different lockdown procedures, likely, the experiences of people and how they would respond to the question would be different in 2020 compared to 2021. Therefore, the findings of this paper should not be considered representative of the entire pandemic. Instead, it should be thought of as providing insight and exploring potential relationships between factors and mental health during the pandemic.

5.5 Next Steps

Several studies could be done to address the weaknesses and limitations found in this paper. First, a follow-up study with different models could investigate the mediating effect of change in alcohol consumption. This would help determine whether the change in alcohol consumption was actually a mediator and how much of a mediating effect it played on the relationship between alcohol and mental health. Second, a small study could be done to investigate the bias of the cleaned dataset. This would help provide insight into the applicability of the findings as well as whether there is an underrepresentation of overrepresentation.

An extension could also be done to the current study. One possible option is to conduct a similar survey as the Canadian Perspectives Survey Series 6:Substance Use and Stigma during the Pandemic for the more recent years of 2021 and 2022. This would enable a time series analysis to see how the situation has changed over the pandemic, as well as allow for findings that could be more generally applicable to the pandemic experience. Another option could be a deeper investigation into the identified factors related to mental health. Stress can be broken down into different types and sources. Employment status could also be investigated to see which related factors are the most significant in influencing mental health. Related factors could include employment status, income level, and salary level. This could help narrow down the focus of areas that the Canadian government should target in their policies.

A Appendix

A.1 Datasheet

Format and questions taken from Gebru et al. (2021).

Motivation

1.*For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.*

The original dataset was created to quickly understand social issues while reducing the cost of collecting data. It focuses on the impact COVID-19 had on Canadians, specifically about the use of substances. The cleaned dataset was created to focus specifically on alcohol and its relationship to COVID-19 and Canadians.

2.*Who created the dataset (for example, which team, research group) and on behalf of which entity (for example, company, institution, organization)?*

The original dataset was created by Statistics Canada. The cleaned dataset was created by Anna Li.

3.*Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.*

The original dataset was funded by Statistics Canada.

4.*Any other comments?*

Composition

1.*What do the instances that comprise the dataset represent (for example, documents, photos, people, countries)? Are there multiple types of instances (for example, movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.*

The instances represent people's responses.

2.*How many instances are there in total (of each type, if appropriate)?*

There are 3941 instances in the original dataset.

3.*Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set? If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (for example, geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (for example, to cover a more diverse range of instances, because instances were withheld or unavailable).*

The dataset is a sample of Canada. It is representative of a larger dataset, using probability sampling and stratification. Weights helped adjust for nonparticipation and potential sources of error.

4.*What data does each instance consist of? "Raw" data (for example, unprocessed text or images) or features? In either case, please provide a description.*

Each instance consists of 180 variables (raw data), and 1 feature (the person who provided the response to the survey).

5.*Is there a label or target associated with each instance? If so, please provide a description.*

Each instance is given an ID by row, which helps identify each instance's individual responses to the various questions. Each response is also correlated with a column label, identifying what the response is for.

6.*Is any information missing from individual instances? If so, please provide a description, explaining why this information is missing (for example, because it was unavailable). This does not include intentionally removed information, but might include, for example, redacted text.*

Information missing has already been cleaned by Statistics Canada in the creation of the dataset. However, choices to not respond to certain questions are included, such as "not stated" and "valid skip." This is due to

the voluntary nature of the survey, which allowed for participants to freely choose whether they wanted to answer the questions asked.

7.*Are relationships between individual instances made explicit (for example, users' movie ratings, social network links)? If so, please describe how these relationships are made explicit.*

There are no evident relationships between instances, as instances were ensured to be anonymous.

8.*Are there recommended data splits (for example, training, development/validation, testing)? If so, please provide a description of these splits, explaining the rationale behind them.*

There are no recommended data splits.

9.*Are there any errors, sources of noise, or redundancies in the dataset? If so, please provide a description.*

There is the potential error of nonresponse error and coverage error. While the targeted response rate was 60%, the actual accumulative response rate was 14%, according to Statistics Canada. This however has been accounted for when using weights to adjust the responses' proportionate meaningfulness. As those above 65 typically do not use the internet, it is possible that the dataset does not fully achieve the desired target representation.

10.*Is the dataset self-contained, or does it link to or otherwise rely on external resources (for example, websites, tweets, other datasets)? If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (that is, including the external resources as they existed at the time the dataset was created); c) are there any restrictions (for example, licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.*

The dataset is self-contained.

11.*Does the dataset contain data that might be considered confidential (for example, data that is protected by legal privilege or by doctor-patient confidentiality, data that includes the content of individuals' non-public communications)? If so, please provide a description.*

There is no confidential data, and the dataset is publicly available.

12.*Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please describe why.*

Yes. The dataset focuses on substance abuse, such as alcohol, drugs, and smoking. These are typically more "taboo" subjects in society that can cause anxiety when discussing them, especially variables that measure how frequent a person may be a substance abuser.

13.*Does the dataset identify any sub-populations (for example, by age, gender)? If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.*

The dataset comprises of Canadians. Demographic characteristics are included, making it possible to identify subpopulations if desired. However, the dataset does not explicitly identify any sub-populations.

14.*Is it possible to identify individuals (that is, one or more natural persons), either directly or indirectly (that is, in combination with other data) from the dataset? If so, please describe how.*

It is not possible to identify individuals in any way.

15.*Does the dataset contain data that might be considered sensitive in any way (for example, data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)? If so, please provide a description.*

Sensitive columns include but are not limited to: employment status, mental health, and alcohol consumption amount.

Collection process

1. *How was the data associated with each instance acquired? Was the data directly observable (for example, raw text, movie ratings), reported by subjects (for example, survey responses), or indirectly inferred/derived from other data (for example, part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.*

The data associated was reported by subjects through an online survey.

2. *What mechanisms or procedures were used to collect the data (for example, hardware apparatuses or sensors, manual human curation, software programs, software APIs)? How were these mechanisms or procedures validated?*

Computer application and software was used to collect the data. The computer program was tested extensively by Statistics Canada.

3. *If the dataset is a sample from a larger set, what was the sampling strategy (for example, deterministic, probabilistic with specific sampling probabilities)?*

Probability sampling with specifically determined weights.

4. *Who was involved in the data collection process (for example, students, crowdworkers, contractors) and how were they compensated (for example, how much were crowdworkers paid)?*

Statistics Canada issued the survey to a sub-sample of Labour Force Survey respondents. No additional compensation was mentioned.

5. *Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (for example, recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.*

The data was collected from 2020 to 2021, with collection lasting approximately a week.

6. *Were any ethical review processes conducted (for example, by an institutional review board)? If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.*

Ethical review processes were not conducted, or if they were conducted, it was not mentioned in the documentation.

7. *Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (for example, websites)?*

I obtained the data from Statistics Canada's website.

8. *Were the individuals in question notified about the data collection? If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.*

The individuals voluntarily chose to participate in the survey collection. No notice is available.

9. *Did the individuals in question consent to the collection and use of their data? If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.*

The individuals consented to the collection and use of their data when they voluntary chose to participate. The exact language to which consent was granted is not available.

10. *If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses? If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).*

A mechanism to revoke consent was not provided. However, during the collection process the respondents could opt to not respond to the questions.

11.*Has an analysis of the potential impact of the dataset and its use on data subjects (for example, a data protection impact analysis) been conducted? If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.*

Such an analysis has not been conducted yet.

12.*Any other comments?*

The purpose of the dataset was to provide a quick glance at the changing circumstances of Canadians due to COVID-19. It is a part of a series of surveys which provide such perspectives.

Preprocessing/cleaning/labeling

1.*Was any preprocessing/cleaning/labeling of the data done (for example, discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)? If so, please provide a description. If not, you may skip the remaining questions in this section.*

The data was originally obtained from Statistics Canada. Cleaning was done to relabel the variables with meaningful names.

2.*Was the "raw" data saved in addition to the preprocessed/cleaned/labeled data (for example, to support unanticipated future uses)? If so, please provide a link or other access point to the "raw" data.*

The raw data obtained was saved as inputs/data/raw_data.csv.

3.*Is the software that was used to preprocess/clean/label the data available? If so, please provide a link or other access point.*

R was used to clean and label the data. R is a free software available at https://www.R-project.org/

4.*Any other comments?*

Documentation by Statistics Canada is included which discusses briefly regarding their cleaning process to arrive at the initial raw dataset.

Uses

1.*Has the dataset been used for any tasks already? If so, please provide a description.*

The dataset has not been used for other tasks yet.

2.*Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.*

https://github.com/annadlli/canada-alcohol.git

3.*What (other) tasks could the dataset be used for?*

The dataset could also be used to examine the state of mental health or the drinking statistics for Canada during COVID-19.

4.*Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses? For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (for example, stereotyping, quality of service issues) or other risks or harms (for example, legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?*

The cleaning process is done on the basis of Statistic Canada's original collection process. Weights implemented account for the potential issues mentioned. The labelling is applicable to only the original dataset and may not be transferrable to other datasets.

5.*Are there tasks for which the dataset should not be used? If so, please provide a description.*

It would be inappropriate to use the dataset for any purposes other than looking at alcohol and its impact on mental health during COVID-19 for Canada.

Distribution

1.*Will the dataset be distributed to third parties outside of the entity (for example, company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.*

No, the dataset is openly available and is currently being used for personal usage.

2.*How will the dataset be distributed (for example, tarball on website, API, GitHub)? Does the dataset have a digital object identifier (DOI)?*

The dataset will be distributed using Github and does not have a DOI.

3.*When will the dataset be distributed?*

The dataset will be distributed in April 2022. Specifically, the final version of the dataset will be available on April 27, 2022.

4.*Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)? If so, please describe this license and/ or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.*

The dataset will be released under the MIT license. The license and details are available on Github.

5.*Have any third parties imposed IP-based or other restrictions on the data associated with the instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.*

There are no restrictions.

6.*Do any export controls or other regulatory restrictions apply to the dataset or to individual instances? If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.*

No such controls or restrictions are applicable.

7.*Any other comments?*

Maintenance

1.*Who will be supporting/hosting/maintaining the dataset?*

Anna Li

2.*How can the owner/curator/manager of the dataset be contacted (for example, email address)?*

Communication is available through github and email.

3.*Is there an erratum? If so, please provide a link or other access point.*

There is no erratum available currently.

4.*Will the dataset be updated (for example, to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (for example, mailing list, GitHub)?*

There are no plans of updating the dataset after April 27, 2022. If any updates were to happen, they would be done through Github and visible from the commit history.

5.*If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (for example, were the individuals in question told that their data would be retained for a fixed period of time and then deleted)? If so, please describe these limits and explain how they will be enforced.*

The dataset was made via survey findings accumulated by Statistics Canada. There are no applicable limits as the people took part voluntarily. Moreover, personal identifiable information was eliminated.

6.*Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.*

There will be no older versions hosted. Dataset consumers can check the obsolescence of the dataset by looking at the Github commit history to see when changes have been made.

7.*If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so? If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.*

There is no way for others to build on to the dataset currently.

A.2 Additional Figures

A.2.1 Raw Observations Distribution Table

Table 3: Data set: Number of Respondents by Age, Sex, and Employment Status

	Male	Female	Unemployed	Employed
15-24 years old	30	62	30	62
25-34 years old	170	215	48	337
35-44 years old	239	264	56	447
45-54 years old	230	268	49	449
55-64 years old	346	368	296	418
65-74 years old	302	302	470	134
75 years and older	100	91	179	12

A.2.2 Linear Assumptions Checking

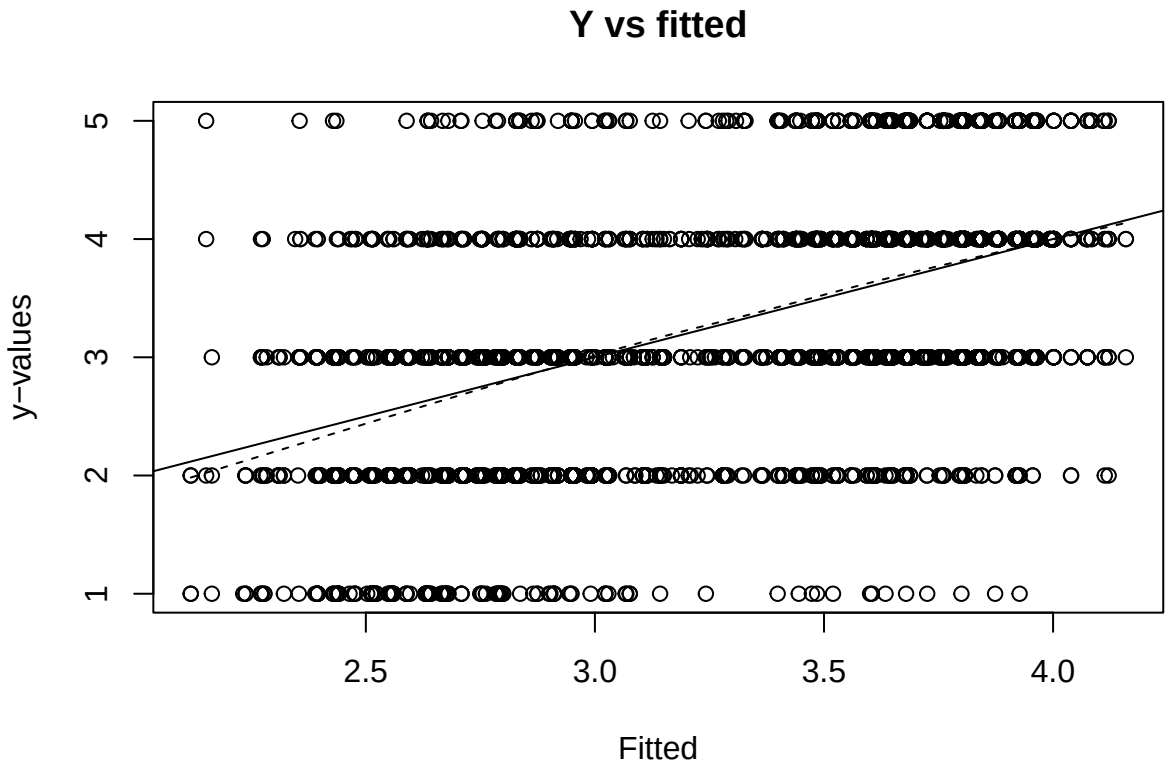

Figure 11: Testing linearity assumptions

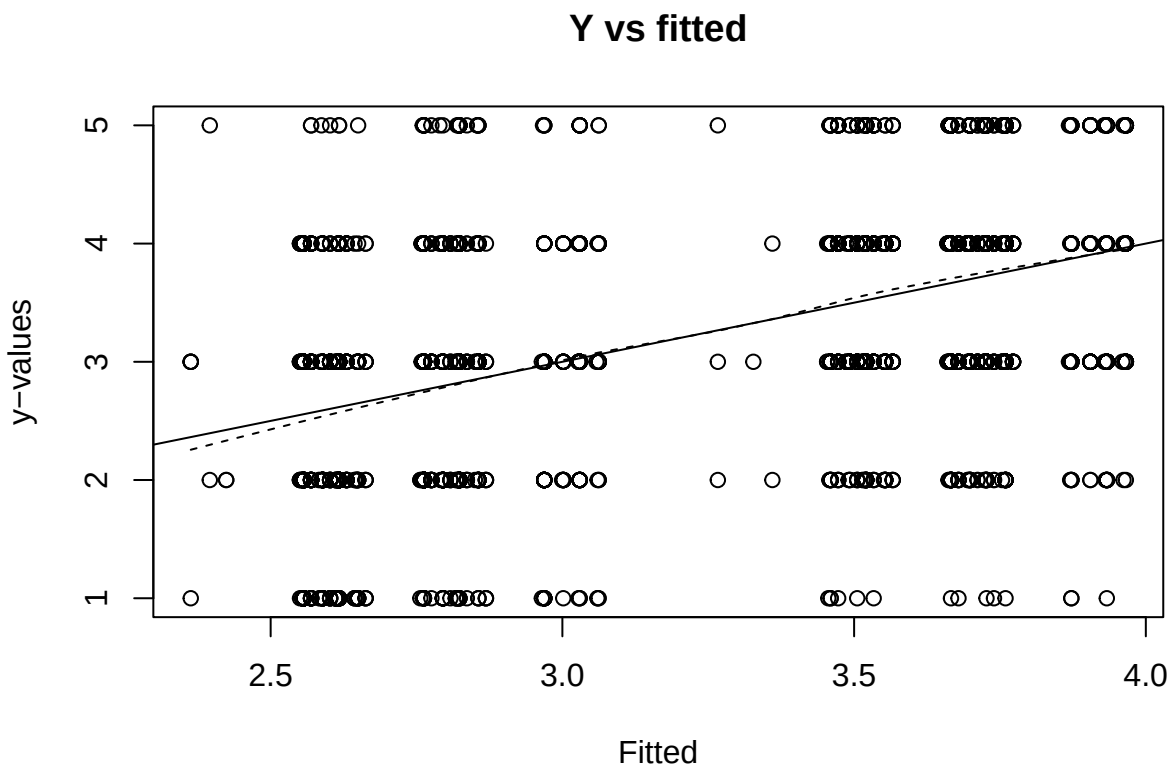

Figure 12: Testing reduced models

As the graph is significantly less linear looking, the reduced model is not utilized.

References

"Covid-19 Pandemic Triggers 25." n.d. *World Health Organization*. World Health Organization. https://www.who.int/news/item/02-03-2022-covid-19-pandemic-triggers-25-increase-in-prevalence-of-anxiety-and-depression-worldwide.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64 (12): 86–92.

Government of Canada, Statistics Canada. 2022. "Canadian Perspectives Survey Series 6: Substance Use and Stigma During the Pandemic Public Use Microdata File." *Government of Canada, Statistics Canada*. https://www150.statcan.gc.ca/n1/en/catalogue/45250012.

Hopman, W. M., M. B. Harrison, H. Coo, E. Friedberg, M. Buchanan, and E. G. VanDenKerkhof. 2009. "Associations Between Chronic Disease, Age and Physical and Mental Health Status." *Chronic Diseases in Canada* 29 (3): 108–17. https://doi.org/10.24095/hpcdp.29.3.03.

Iannone, Richard. 2022. *DiagrammeR: Graph/Network Visualization*. https://CRAN.R-project.org/package=DiagrammeR.

Kiely, Kim M., Brooke Brady, and Julie Byles. 2019. "Gender, Mental Health and Ageing." *Maturitas* 129: 76–84. https://doi.org/10.1016/j.maturitas.2019.09.004.

Modini, Matthew, Sadhbh Joyce, Arnstein Mykletun, Helen Christensen, Richard A Bryant, Philip B Mitchell, and Samuel B Harvey. 2016. "The Mental Health Benefits of Employment: Results of a Systematic Meta-Review." *Australasian Psychiatry* 24 (4): 331–36. https://doi.org/10.1177/1039856215618523.

Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rosenfield, Sarah, and Dawne Mouzon. 2013. "Gender and Mental Health." In *Handbook of the Sociology of Mental Health*, edited by Carol S. Aneshensel, Jo C. Phelan, and Alex Bierman, 277–96. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-4276-5_14.

Sjoberg, Daniel D., Karissa Whiting, Michael Curry, Jessica A. Lavery, and Joseph Larmarange. 2021. "Reproducible Summary Tables with the Gtsummary Package." *The R Journal* 13: 570–80. https://doi.org/10.32614/RJ-2021-053.

Thomas, Gerald. 2012. *Levels and Patterns of Alcohol Use in Canada*. Canadian Centre on Substance Abuse.

Vaneer, Pamela. 2022. "Alcohol-Related Hospital Visits up 50." *Kawartha 411*. https://www.kawartha411.ca/2022/04/05/hospital-visits-related-to-alcohol-use-up-50-locally-since-pandemic-began/.

Weitzman, Elissa R. 2004. "Poor Mental Health, Depression, and Associations with Alcohol Consumption, Harm, and Abuse in a National Sample of Young Adults in College." *Journal of Nervous &Amp; Mental Disease* 192 (4): 269–77. https://doi.org/10.1097/01.nmd.0000120885.17362.94.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Jennifer Bryan. 2019. *Readxl: Read Excel Files*. https://CRAN.R-project.org/package=readxl.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Young, Matthew M. 2011. *Cross-Canada Report on Student Alcohol and Drug Use: Technical Report*. Canadian Centre on Substance Abuse.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.