

1 BACKGROUND

2 OBJECTIVES

3 METHODOLOGY - SELECT THE DATA

4 FINDINGS

4.1 Findings - Observations from scrutinizing the data

4.2 Findings - missing values

4.3 Findings - Recoding

4.4 Findings - Visualizing variables and explanation

4.4.1 Demographic profiles of the respondents to the survey wave 6th

4.4.2 Investigating the difference between the drinking and not drinking groups

4.4.3 The association of *Binge drinking* and stress level in respondents

5 CONCLUSION

6 REFERENCE

Code ▾

Examining the Impact of COVID-19 on Mental Health and Substance Use among Canadians

Assignment 1 | R markdown

Author: Le Thanh Nguyen

- MSc eHealth program
- DeGroote School of Business, McMaster University
- Github page: <https://github.com/thanhnguyen3849> (<https://github.com/thanhnguyen3849>)

30 January 2023

NB: This document is my original work. No code nor text has been copied from any other student

Code

1 BACKGROUND

In a series of national surveys, CAMH and research technology company Delvinia led research to understand the mental health and substance use impacts of COVID-19, and to track changes as the pandemic unfolded [1]. The following interactive dashboard represents selected findings based on online surveys conducted over the following intervals:

- **Wave 1:** May 8 to 12, 2020, with 1,005 respondents
- **Wave 2:** May 29 to June 1, 2020, with 1,002 respondents
- **Wave 3:** June 19 to 23, 2020 with 1,005 respondents
- **Wave 4:** July 10 to 14, 2020 with 1,003 respondents
- **Wave 5:** September 18 to 22, 2020, with 1,003 respondents
- **Wave 6:** November 27 to December 1, 2020 with 1,003 respondents
- **Wave 7:** March 19 to 23, 2021 with 1,000 respondents
- **Wave 8:** July 9 to 13 2021, with 1,001 respondents
- **Wave 9:** January 7 to 11, 2022, with 1,004 respondents

All survey participants were English-speaking Canadians ages 18 and older [1].

Below are some key findings for this particular study, wave 6.

Data source: CAMH Studies Mental Health over Coronavirus Pandemic (<https://www.delvinia.com/camh-coronavirus-mental-health/>)

Code

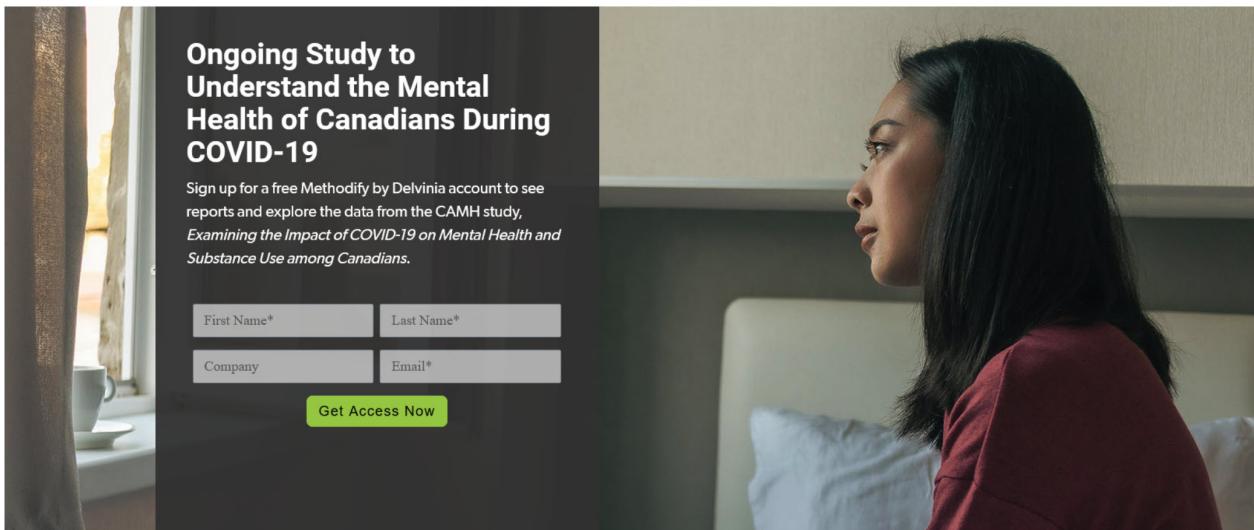


Figure 1.1: The Delvinia Website

2 OBJECTIVES

In this report, I employ the different packages in R and various functions to generate firstly the initial analysis of the data, and second the subsequent steps to prepare data for analysis. I use the dataset provided by the 6th wave of the CAMH study to generate insights regarding:

- The dataset itself
- The study's research results

3 METHODOLOGY - SELECT THE DATA

In this report, I use RStudio to scrutinize the data and generate descriptive analyses using relevant packages. Some of the packages that I use and their functions are:

1. **dplyr**: A powerful and efficient data manipulation package that provides a set of tools for filtering, grouping, and summarizing data. It is particularly useful for working with large datasets and is widely used in data wrangling tasks.
2. **ggplot2**: A data visualization package that provides a flexible and powerful way to create a wide range of static and interactive plots. It is widely used for creating high-quality data visualizations and is particularly useful for exploratory data analysis.
3. **pacman**: An R package management tool, which provides a convenient way to install and manage R packages, similar to the package manager on Linux and Mac OS. It is a wrapper around the basic functionality provided by the "install.packages()" function and provides a more user-friendly interface for installing, updating, and removing packages.

3.1 Import the dataset and the datamap

The data and the datamap for CAMH study wave 6th were imported.

[Code](#)

[Code](#)

| | |
|----------|--------------|
| respid, | respid |
| status, | status |
| L, | Language |
| hWave, | wave tracker |
| Consent, | Consent |

Table 3.1: The datemap of the survey

| Variable.ID | Question.Label |
|-------------|--|
| S1, | In which province or territory do you currently live? |
| hRegion, | In which province or territory do you currently live? |
| S2, | To which of the following age groups do you belong? (detailed) |

[Code](#)

3.2 Trim and filter only important variables

A brief overview shows that there are **1003 observations of 66 variables** in this dataframe. I will limit my work to the following **17 variables** only:

- hAge (Age_group),
- Gender (male, female, other),
- Q4_1, Q4_3, Q4_5,
- Q5 (switch coding 1->4, 2->3, 3->2, 4->1),
- Q8x1 - Q8x3,
- Q15,
- Q17,
- Q20x1, Q20x2, Q20x3,
- Q25 Education
- Q26 Marital status
- Q27 Ethnicity/Race

[Code](#)

Table 3.2: New dataframe of only 17 variables of interest

| Age_group | Gender | Q4_1_me_covid_test_pos | Q4_3_me_covid_symptoms_no_test | Q4_5_me_covid_test_neg | Q5_fin_worry | Q8x1_nervous_frq |
|-----------|--------|------------------------|--------------------------------|------------------------|--------------|------------------|
| 1 | 2 | 1 | | 1 | 1 | 2 |
| 3 | 1 | 1 | | 1 | 1 | 2 |
| 1 | 1 | 1 | | 1 | 1 | 2 |
| 2 | 2 | 1 | | 1 | 1 | 2 |
| 3 | 2 | 1 | | 1 | 1 | 1 |
| 3 | 2 | 1 | | 1 | 2 | 2 |

[Code](#)

I switched coding for Q5 - "How worried are you about the impact of COVID-19 on your personal financial situation?" as per instruction (switch coding 1->4, 2->3, 3->2, 4->1).

Table 3.3: New dataframe of switched codes

| Age_group | Gender | Q4_1_me_covid_test_pos | Q4_3_me_covid_symptoms_no_test | Q4_5_me_covid_test_neg | Q5_fin_worry | Q8x1_nervous_frq |
|-----------|--------|------------------------|--------------------------------|------------------------|--------------|------------------|
| 1 | 2 | 1 | | 1 | 1 | 2 |
| 3 | 1 | 1 | | 1 | 1 | 2 |
| 1 | 1 | 1 | | 1 | 1 | 2 |
| 2 | 2 | 1 | | 1 | 1 | 2 |
| 3 | 2 | 1 | | 1 | 1 | 1 |
| 3 | 2 | 1 | | 1 | 3 | 2 |

3.3 Renaming the variables

I also rename the variable in this new dataframe to enhance clarity and comprehensiveness.

[Code](#)

Table 3.4: New dataframe of selected variables renamed

| age_group | gender | covid_positive | covid_no_test | covid_test_negative | financial_worry | nervous_frq | not_stop_worry_frq | worry_diff_frq | alcohol_ |
|-----------|--------|----------------|---------------|---------------------|-----------------|-------------|--------------------|----------------|----------|
| 1 | 2 | 1 | 1 | 1 | 4 | 2 | 2 | 2 | |
| 3 | 1 | 1 | 1 | 1 | 4 | 2 | 2 | 3 | |
| 1 | 1 | 1 | 1 | 1 | 4 | 2 | 2 | 2 | |
| 2 | 2 | 1 | 1 | 1 | 3 | 2 | 2 | 2 | |
| 3 | 2 | 1 | 1 | 1 | 3 | 1 | 1 | 4 | |
| 3 | 2 | 1 | 1 | 1 | 3 | 2 | 2 | 2 | |

4 FINDINGS

4.1 Findings - Observations from scrutinizing the data

Now the data is prepared, I carry on to examine the data, identify any missing values, or any errors in the collected data.

[Code](#)

Table 4.1: Summary statistic of the dataframe

| | vars | n | mean | sd | median | trimmed | mad | min | max | range | skew | kurtosis | se |
|---------------------|------|------|-----------|------------|--------|----------|--------|-----|-----|-------|------------|-------------|-----------|
| age_group | 1 | 1003 | 1.914257 | 0.8302094 | 2 | 1.892902 | 1.4826 | 1 | 3 | 2 | 0.1607858 | -1.5327394 | 0.0262142 |
| gender | 2 | 1003 | 1.517448 | 0.5156674 | 2 | 1.511831 | 0.0000 | 1 | 3 | 2 | 0.1048699 | -1.5731583 | 0.0162824 |
| covid_positive | 3 | 1003 | 1.006979 | 0.0832903 | 1 | 1.000000 | 0.0000 | 1 | 2 | 1 | 11.8268145 | 138.0111426 | 0.0026299 |
| covid_no_test | 4 | 1003 | 1.027916 | 0.1648151 | 1 | 1.000000 | 0.0000 | 1 | 2 | 1 | 5.7229353 | 30.7826824 | 0.0052041 |
| covid_test_negative | 5 | 1003 | 1.175474 | 0.3805614 | 1 | 1.094645 | 0.0000 | 1 | 2 | 1 | 1.7038136 | 0.9038850 | 0.0120164 |
| financial_worry | 6 | 1003 | 2.713858 | 0.9040149 | 3 | 2.767123 | 1.4826 | 1 | 4 | 3 | -0.2571560 | -0.7102377 | 0.0285447 |
| nervous_frq | 7 | 1003 | 2.000997 | 0.9864344 | 2 | 1.876712 | 1.4826 | 1 | 4 | 3 | 0.7770422 | -0.3969858 | 0.0311471 |
| not_stop_worry_frq | 8 | 1003 | 1.818544 | 0.9607959 | 2 | 1.657534 | 1.4826 | 1 | 4 | 3 | 0.9954127 | -0.0365771 | 0.0303376 |
| worry_diff_frq | 9 | 1003 | 1.988036 | 0.9643632 | 2 | 1.860523 | 1.4826 | 1 | 4 | 3 | 0.7442076 | -0.3956563 | 0.0304502 |
| alcohol_drink | 10 | 1003 | 10.163509 | 27.0889564 | 1 | 2.099626 | 1.4826 | 0 | 99 | 99 | 2.9444639 | 6.7534390 | 0.8553460 |
| alcohol_frq | 11 | 1003 | 4.184447 | 10.5602008 | 3 | 2.938979 | 0.0000 | 1 | 96 | 95 | 8.5153101 | 71.0667061 | 0.3334431 |
| depressed_frq | 12 | 1003 | 1.813559 | 0.9520100 | 2 | 1.667497 | 1.4826 | 1 | 4 | 3 | 0.9330074 | -0.1906952 | 0.0300601 |
| lonely_frq | 13 | 1003 | 1.826520 | 0.9908839 | 2 | 1.667497 | 1.4826 | 1 | 4 | 3 | 0.9235850 | -0.3350125 | 0.0312876 |
| hopeful_frq | 14 | 1003 | 2.429711 | 1.0264927 | 2 | 2.412204 | 1.4826 | 1 | 4 | 3 | 0.0427304 | -1.1416828 | 0.0324120 |
| education | 15 | 1003 | 5.257228 | 10.0462819 | 5 | 4.353674 | 0.0000 | 1 | 96 | 95 | 8.8104130 | 76.6374240 | 0.3172159 |
| marital | 16 | 1003 | 4.076770 | 11.8837134 | 1 | 2.427148 | 0.0000 | 1 | 96 | 95 | 7.3743845 | 54.1124869 | 0.3752336 |
| race | 17 | 1003 | 8.980060 | 14.7451268 | 8 | 7.069738 | 0.0000 | 1 | 97 | 96 | 5.5261981 | 29.8387681 | 0.4655840 |

[Code](#)

Table 4.2: The structure of the missing values

| Variable | q | qNA | pNA | qZero | pZero | qBlank | pBlank | qlnf | plnf | qDistinct | type | anomalous_percent |
|---------------------|------|-----|-----|-------|-------|--------|--------|------|------|-----------|---------|-------------------|
| alcohol_drink | 1003 | 0 | • | 347 | 34.6% | 0 | • | 0 | • | 10 | Integer | 34.6% |
| covid_positive | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 2 | Integer | • |
| covid_no_test | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 2 | Integer | • |
| covid_test_negative | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 2 | Integer | • |
| age_group | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 3 | Integer | • |
| gender | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 3 | Integer | • |
| financial_worry | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 4 | Numeric | • |
| nervous_frq | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 4 | Integer | • |
| not_stop_worry_frq | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 4 | Integer | • |
| worry_diff_frq | 1003 | 0 | • | 0 | • | 0 | • | 0 | • | 4 | Integer | • |

| Variable | q | qNA | pNA | qZero | pZero | qBlank | pBlank | qInf | pInf | qDistinct | type | anomalous_percent |
|---------------|------|-----|-----|-------|-------|--------|--------|------|------|-----------|---------|-------------------|
| depressed_frq | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 4 | Integer | * |
| lonely_frq | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 4 | Integer | * |
| hopeful_frq | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 4 | Integer | * |
| alcohol_frq | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 6 | Integer | * |
| education | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 6 | Integer | * |
| marital | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 7 | Integer | * |
| race | 1003 | 0 | * | 0 | * | 0 | * | 0 | * | 12 | Integer | * |

Code

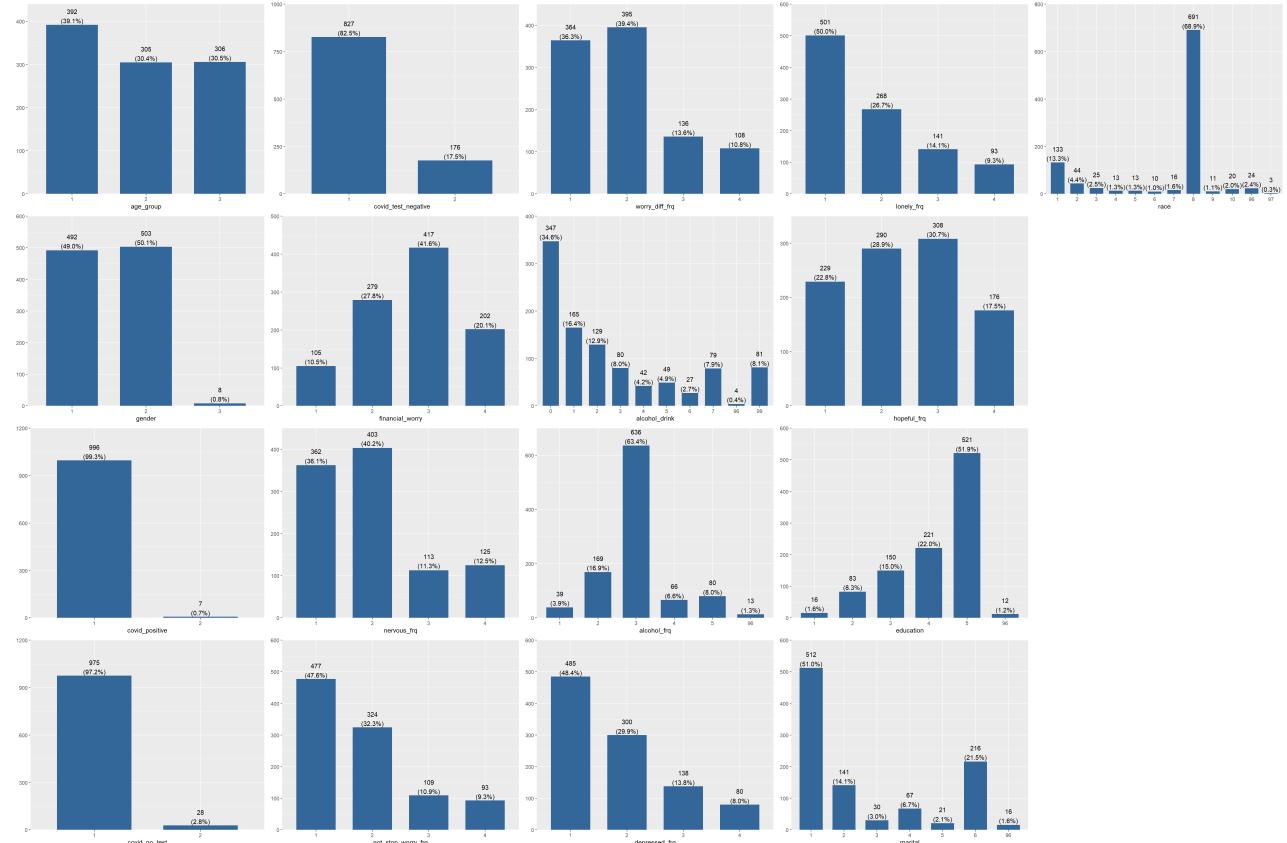


Figure 4.1: Plot frequencies of the variables in the dataset

Observations on scrutinizing the data:

- There is no N/A value in the dataframe of interest
- For Q15_alcohol, "0" which can be an anomalous value in other dataset, for this particular case, the value is used to distinguish the "non_alcohol" drinkers from "alcohol" drinkers but did not consume alcohol within the 7 days of the survey.
- For some variables, existing extreme values like "95", "96", and "99" which skew the distribution and affecting the mean values (Such as Q15 and Q17 regarding the alcohol consumption).
- For other variables in this dataframe, they are **categorical**, which means the value is limited and usually based on a particular finite group. However, for "Q15- During the PAST 7 DAYS, on how many days did you drink ALCOHOL?", there are both **continuous values** (The number of days people drinking) and the **categorical values** ("I do not drink alcohol" and "Prefer not to answer").
- Therefore, for question "Q15- During the PAST 7 DAYS, on how many days did you drink ALCOHOL?", there are three groups of respondents:
 1. People who drink alcohol (Corresponding to answer from 0 - 7)
 2. People who do NOT drink alcohol (Corresponding to answer 99 - "I do not drink alcohol")
 3. People who choose not to answer the question (Corresponding to answer 96 - Prefer not to answer")
- So I will break Q15 into 3 categories accordingly. And within the first category, I will maintain the numeric values for further analysis. The third category, I identify it as **not applicable or n/a** and thus, will proceed to remove them.
- The rest of the data was fine

4.2 Findings - missing values

4.2.1 Remove n/a values

I change value "96" from Q15 into **n/a** and remove those **n/a** entries from the dataframe. This is only to demonstrate how to remove n/a from a dataset. Otherwise can use the subset() function to remove entries associated with the "96" value.

The new data frame from now will be referred to as "camh_clean". Now there is only **999 observations**.

Code

Table 4.3: Removing n/a values

| val | label | frq | raw.prc | valid.prc | cum.prc |
|-----|--------|-----|---------|-----------|---------|
| 0 | <none> | 347 | 34.73 | 34.73 | 34.73 |
| 1 | <none> | 165 | 16.52 | 16.52 | 51.25 |
| 2 | <none> | 129 | 12.91 | 12.91 | 64.16 |
| 3 | <none> | 80 | 8.01 | 8.01 | 72.17 |
| 4 | <none> | 42 | 4.20 | 4.20 | 76.38 |
| 5 | <none> | 49 | 4.90 | 4.90 | 81.28 |
| 6 | <none> | 27 | 2.70 | 2.70 | 83.98 |
| 7 | <none> | 79 | 7.91 | 7.91 | 91.89 |
| 99 | <none> | 81 | 8.11 | 8.11 | 100.00 |
| NA | NA | 0 | 0.00 | NA | NA |

4.2.2 Remove unnecessary data

For some questions, there are the "Preferred not to answer" option, which coded by "96"; as well as "Not sure" coded by "97", and they occupy less than 5% of the total observations. Therefore, I will remove observations that have a value of "96" and "97" to make the data more relevant.

Now there is only **954 observations**

Code

Table 4.4: Removing unnecessary data

| val | label | frq | raw.prc | valid.prc | cum.prc |
|-----|--------|-----|---------|-----------|---------|
| 1 | <none> | 129 | 13.52 | 13.52 | 13.52 |
| 2 | <none> | 42 | 4.40 | 4.40 | 17.92 |
| 3 | <none> | 24 | 2.52 | 2.52 | 20.44 |
| 4 | <none> | 12 | 1.26 | 1.26 | 21.70 |
| 5 | <none> | 13 | 1.36 | 1.36 | 23.06 |
| 6 | <none> | 9 | 0.94 | 0.94 | 24.00 |
| 7 | <none> | 16 | 1.68 | 1.68 | 25.68 |
| 8 | <none> | 679 | 71.17 | 71.17 | 96.86 |
| 9 | <none> | 11 | 1.15 | 1.15 | 98.01 |
| 10 | <none> | 19 | 1.99 | 1.99 | 100.00 |
| NA | NA | 0 | 0.00 | NA | NA |

4.3 Findings - Recoding

Data in the “camh new” dataframe will be recoded, following the instructions in the new datamap, “20-3082-DATAMAP.xlsx”.

4.3.1 Recode Demographic variables

Demographic variables will be decoded directly (without creating the factor forms, as we do not need their numeric values for calculation) such as:

- Age groups
 - Genders
 - Covid status (Variable 3 -5)
 - Education level
 - Marital status
 - Race/ Ethnicity

Code

Table 4.5: Recode Demographic variables

age group gender covid positive covid no test covid test negative financial worry nervous freq not stop worry freq worry diff freq alcohol

| age_group | gender | covid_positive | covid_no_test | covid_test_negative | financial_worry | nervous_frq | not_stop_worry_frq | worry_diff_frq | alcohol_ |
|-----------|--------|----------------|---------------|---------------------|-----------------|-------------|--------------------|----------------|----------|
| 18-39 | Female | No | No | No | 4 | 2 | 2 | 2 | |
| 60+ | Male | No | No | No | 4 | 2 | 2 | 3 | |
| 18-39 | Male | No | No | No | 4 | 2 | 2 | 2 | |

4.3.2 Recode Non-Demographic variables

Non-demographic variables will be decoded by creating a factor forms, namely:

- Financial worry
- Nervous frequency
- Not stop worry frequency
- Worry different things frequency
- Alcohol consumption frequency
- Depressed frequency
- Lonely frequency
- Hopeful frequency

[Code](#)

Table 4.6: Recode Non-Demographic variables

| age_group | gender | covid_positive | covid_no_test | covid_test_negative | financial_worry | nervous_frq | not_stop_worry_frq | worry_diff_frq | alcohol_ |
|-----------|--------|----------------|---------------|---------------------|-----------------|-------------|--------------------|----------------|----------|
| 18-39 | Female | No | No | No | 4 | 2 | 2 | 2 | |
| 60+ | Male | No | No | No | 4 | 2 | 2 | 3 | |
| 18-39 | Male | No | No | No | 4 | 2 | 2 | 2 | |
| 60+ | Female | No | No | No | 3 | 1 | 1 | 4 | |
| 60+ | Female | No | No | No | 2 | 2 | 2 | 2 | |

4.3.3 Recode Q15 into 2 categories “Drinking” and “No drinking”

Values in question 15 will be re-coded into 2 different categories:

1. **People who drink alcohol** (Corresponding to answer from 0 - 7)
2. **People who do NOT drink alcohol** (Corresponding to answer 99 - “I do not drink alcohol”)

[Code](#)

[1] 0

[Code](#)

Table 4.7: Clean data set ready for analysis

| age_group | gender | covid_positive | covid_no_test | covid_test_negative | education | marital | race | financial_worry_factored | nervous_frq_factored | not_stop_worry_factored | worry |
|-----------|--------|----------------|---------------|---------------------|----------------------------|---------------|-------|--------------------------|----------------------|-------------------------|-------|
| 18-39 | Female | No | No | No | University diploma/degree | Never married | White | Very worried | Several days | Several days | Seve |
| 60+ | Male | No | No | No | College diploma/degree | Married | White | Very worried | Several days | Several days | Over |
| 18-39 | Male | No | No | No | Post-high school education | Married | White | Very worried | Several days | Several days | Seve |
| 60+ | Female | No | No | No | Completed high school | Married | White | Some what worried | Not at all | Not at all | Near |
| 60+ | Female | No | No | No | Post-high school education | Widowed | White | Some what worried | Several days | Several days | Seve |

| age_group | gender | covid_positive | covid_no_test | covid_test_negative | education | marital | race | financial_worry_factored | nervous_frq_factored | not_stop_worry_factored | wor |
|-----------|--------|----------------|---------------|---------------------|---------------------------|---------|-------|--------------------------|----------------------|-------------------------|------|
| 18-39 | Female | No | No | Yes | University diploma/degree | Married | White | Very worried | Several days | Several days | Over |

4.4 Findings - Visualizing variables and explanation

4.4.1 Demographic profiles of the respondents to the survey wave 6th

[Code](#)

Table 4.8: Number of respondents infected with COVID-19

| | Infected with COVID | | Male |
|-------|---------------------|------|------|
| | Female | Male | |
| 18-39 | 2 | | 2 |
| 40-59 | 0 | | 1 |
| 60+ | 0 | | 1 |

[Code](#)

Table 4.9: Number of respondents not infected with COVID-19

| | No COVID | | |
|-------|----------|------|-------|
| | Female | Male | Other |
| 18-39 | 192 | 169 | 4 |
| 40-59 | 145 | 142 | 1 |
| 60+ | 146 | 148 | 1 |

[Code](#)

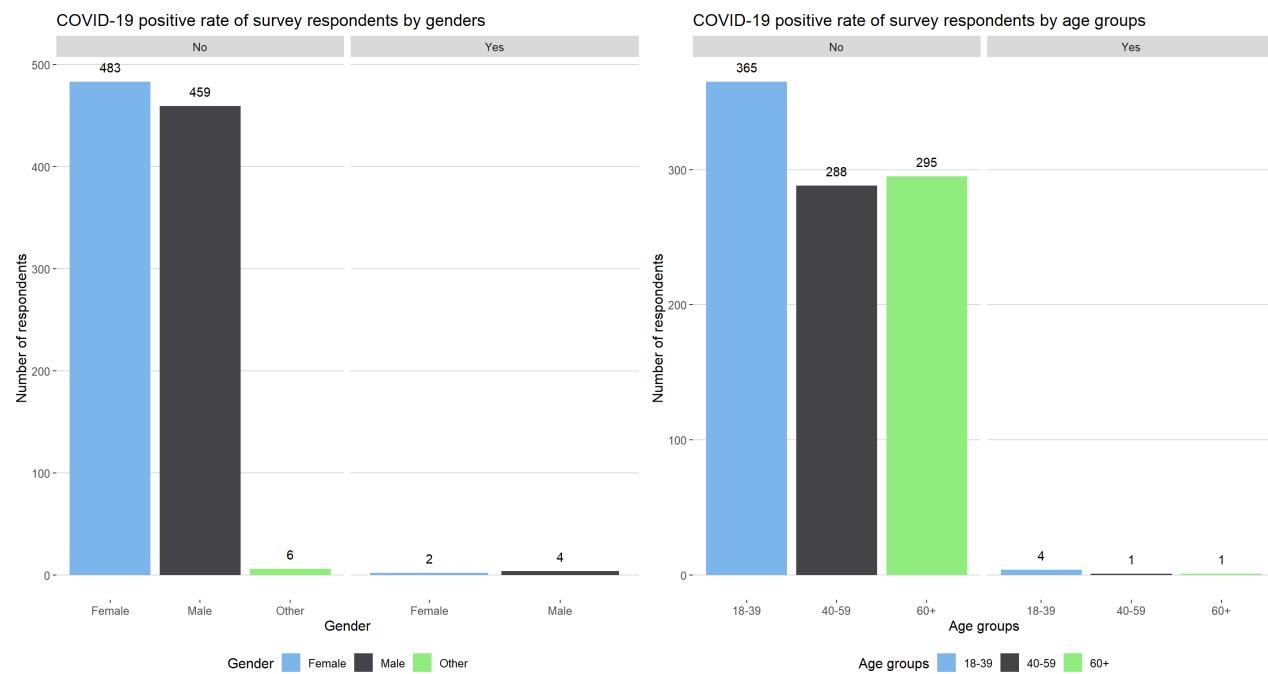


Figure 4.2: Graph of COVID-19 status of survey respondents

From Table 4.8, Table 4.9, and Figure 4.2 above, we can see that the majority of people took part in this survey were not COVID-19, be it asymptomatic or have been tested. There are people with confirmed COVID-19 infection, but their number is too small to be accounted for, at **6 people over the total of 954 responses**.

In terms of **genders**, there is an equal number of males and females taking part in this survey, with the latter group (483) is a slightly more than the former one (459). The third group of "Other" gender made up a humble statistic of 6 people.

In terms of **age groups**, the youngest group of respondents who are from 18 to 39 years old was significantly larger than the other two groups, with a gap as wide as 100 respondents, making up approximately 20% - 25% of the total people in each group. This may be owing to the fact that young adults are more technical savvy than their older counterparts, thus they have a greater tendency to engage in online surveys and studies similar to the CAMH research. However, it still requires further analysis to determine whether or not this difference carries any significance.

4.4.2 Investigating the difference between the drinking and not drinking groups

- Investigating into the population who drink from the responses, we can see that the majority of people who participated in this survey drink alcohol.
- Nevertheless, it is interesting to see that the number of males and females in each drinking and not drinking group is not significantly different.
- Among people who drink, the average days on drinking in the past 7 days before taking the survey is **$mean(x)=1.99$** .
- The majority of people reported that they do not drink more or less within the past 7 days, but a considerable number of people recalled to drink a slightly more during the same period.
- Females are observed to have the tendency to be nervous and worried more than their male counterparts.

[Code](#)

[Code](#)

Table 4.10: Respondents who drink

| Drinkers | | |
|----------|--------|-------|
| gender | people | pct_a |
| Female | 448 | 0.508 |
| Male | 429 | 0.486 |
| Other | 5 | 0.006 |

[Code](#)

Table 4.11: Respondents who do not drink

| Non drinkers | | |
|--------------|--------|-------|
| gender | people | pct_b |
| Female | 37 | 0.514 |
| Male | 34 | 0.472 |
| Other | 1 | 0.014 |

[Code](#)

[Code](#)

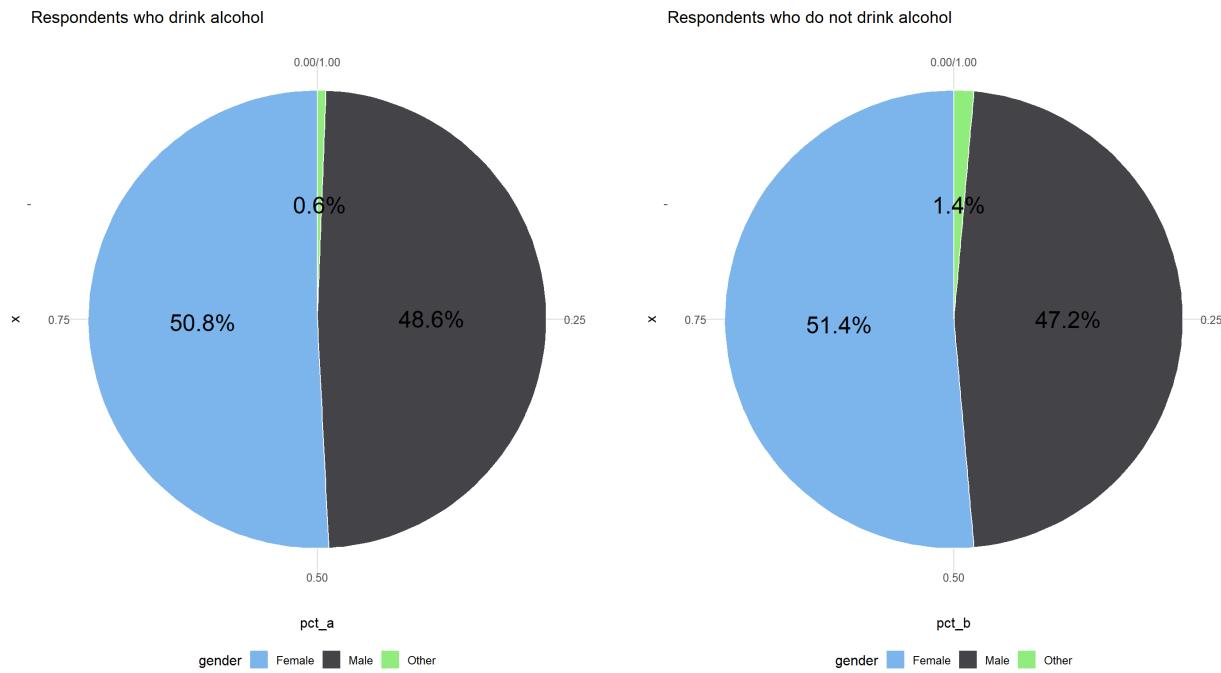


Figure 4.3: Graph of drinking status of survey respondents

[Code](#)

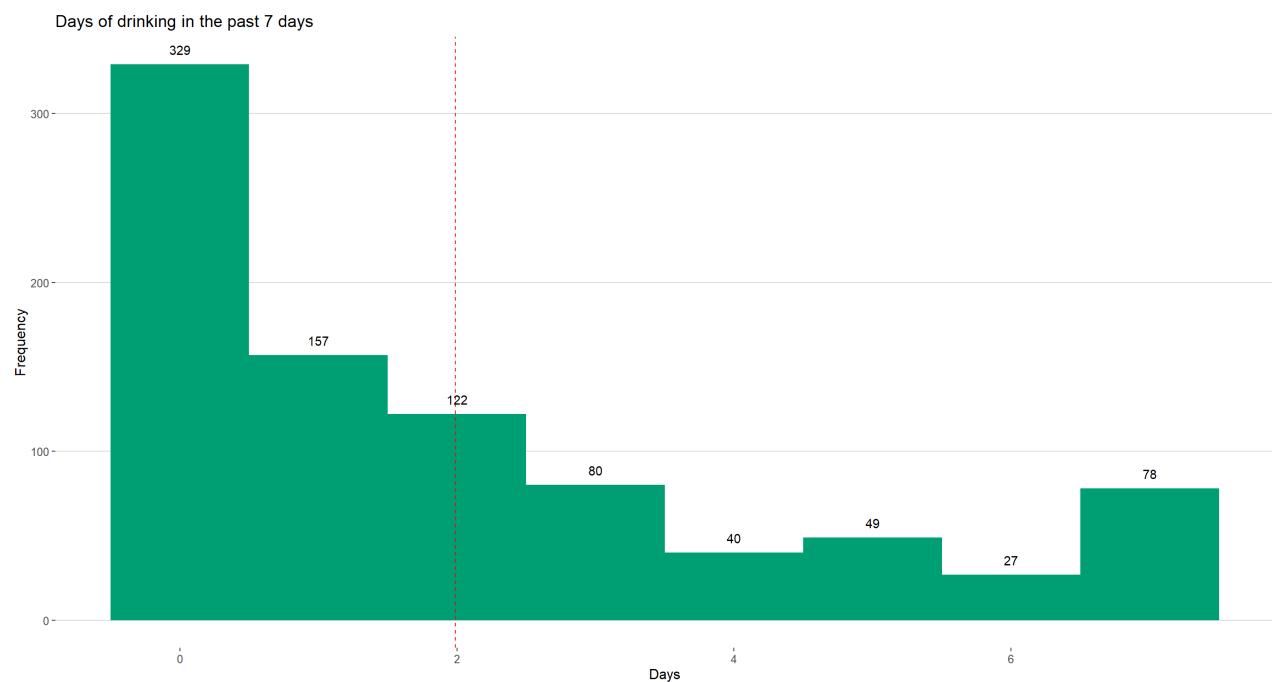


Figure 4.4: Graph of drinking status of survey respondents

[Code](#)

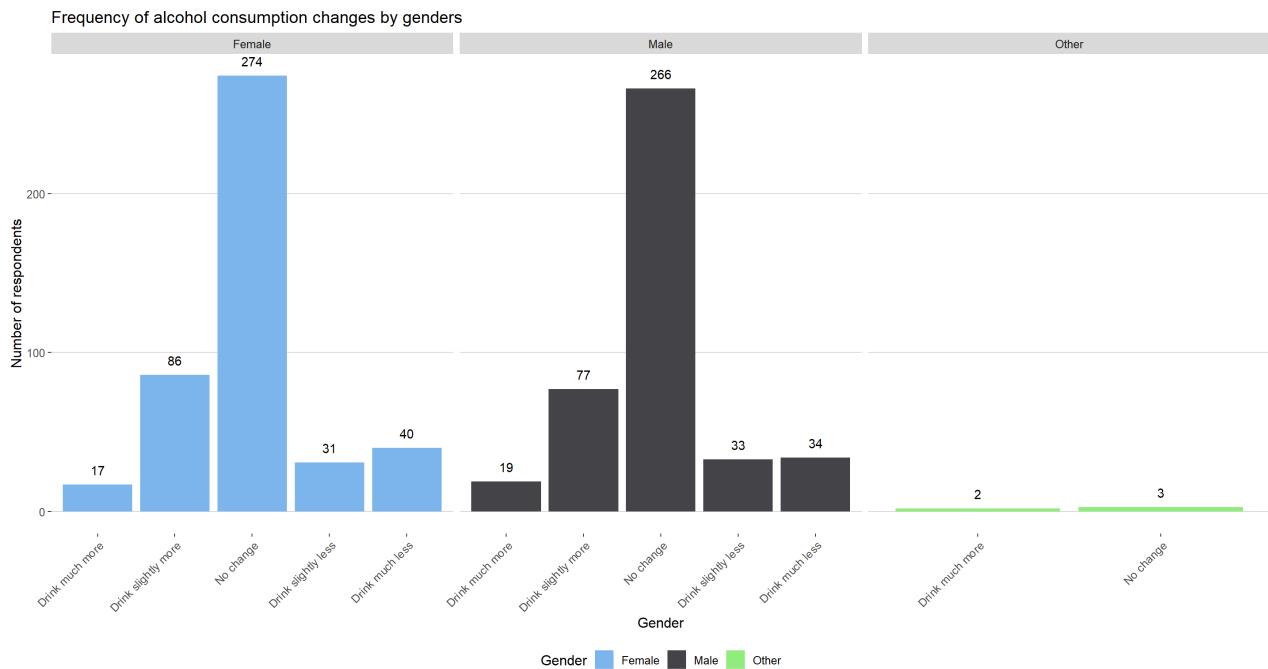


Figure 4.5: Graph of drinking status of survey respondents

4.4.3 The association of *Binge drinking* and stress level in respondents

- A closer look into people who did not drink during the past 7 days and who drank every day regarding their stress frequency
- Overall, there is no significant difference between the two groups from observing the graphs
- Nonetheless, the number of people who experienced nervousness and worries for several days in the binge-drinking group accounted for the largest group.
- In contrast, for people who did not drink, they appeared to do not experience any episodes of anxiety
- From the graphs, people who binge_drinking might experience longer periods of stress than people who did not drink alcohol. However, more investigation is still needed to make the association of *Binge drinking* and stress level in respondents clearer

[Code](#)

Table 4.12: The structure of the number of days people drink

| val | label | frq | raw.prc | valid.prc | cum.prc |
|-----|--------|-----|---------|-----------|---------|
| 0 | <none> | 329 | 37.30 | 37.30 | 37.30 |
| 1 | <none> | 157 | 17.80 | 17.80 | 55.10 |
| 2 | <none> | 122 | 13.83 | 13.83 | 68.93 |
| 3 | <none> | 80 | 9.07 | 9.07 | 78.00 |
| 4 | <none> | 40 | 4.54 | 4.54 | 82.54 |
| 5 | <none> | 49 | 5.56 | 5.56 | 88.10 |
| 6 | <none> | 27 | 3.06 | 3.06 | 91.16 |
| 7 | <none> | 78 | 8.84 | 8.84 | 100.00 |
| NA | NA | 0 | 0.00 | NA | NA |

[Code](#)

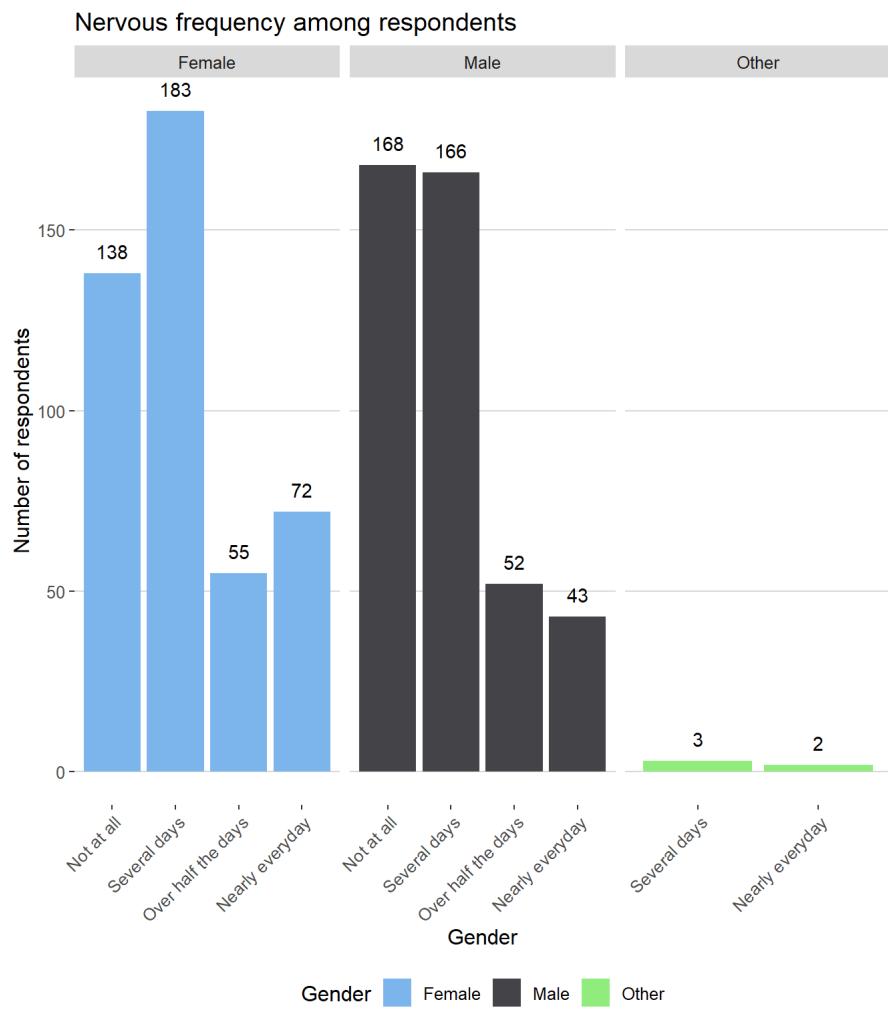


Figure 4.6: Stress frequency among respondents

[Code](#)

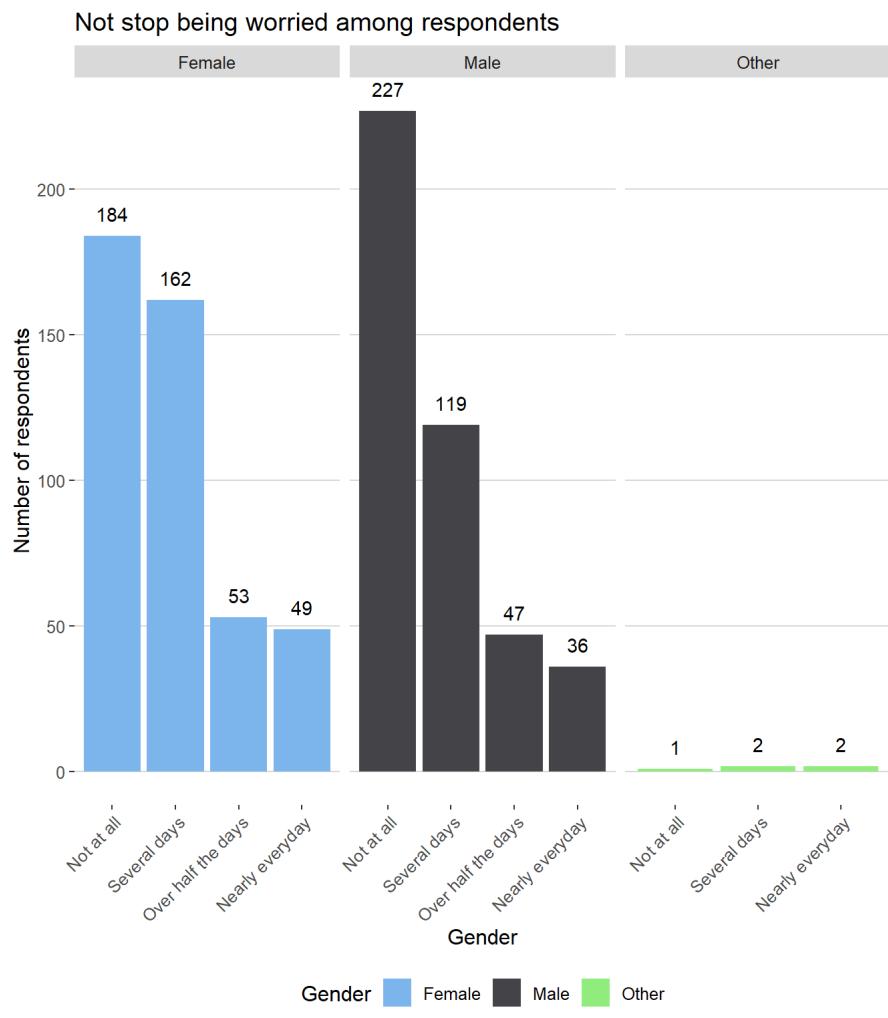


Figure 4.7: Stress frequency among respondents

Worries about different things among respondents

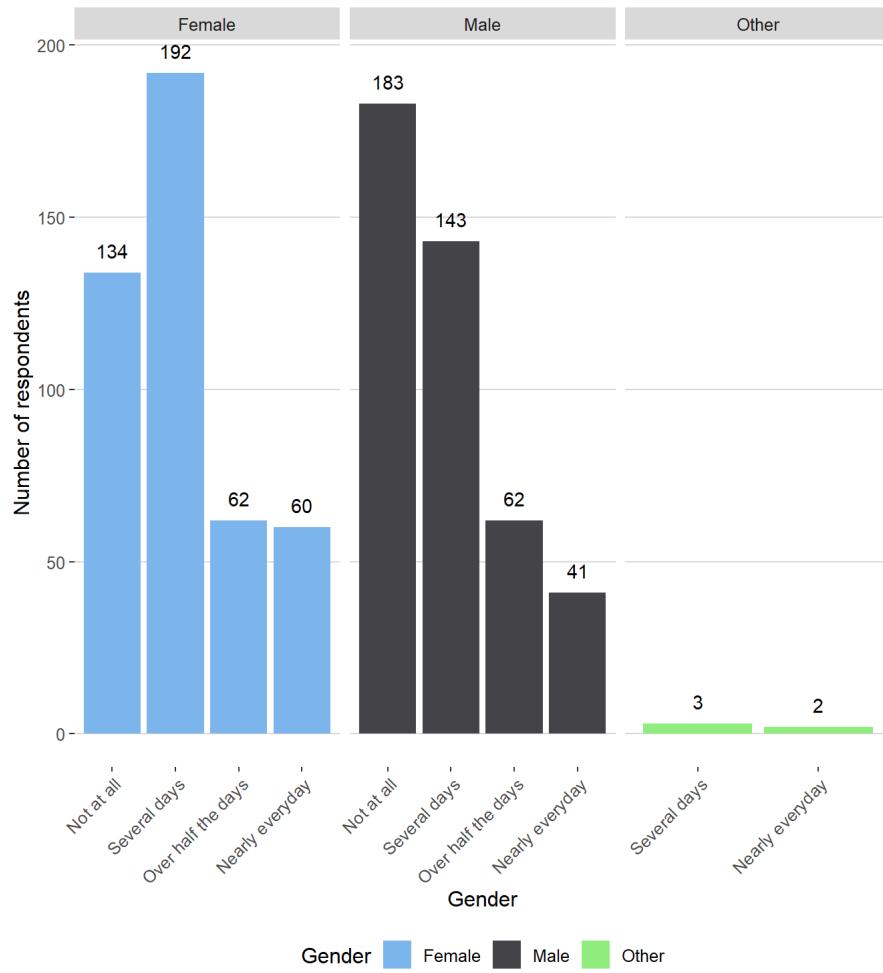


Figure 4.8: Stress frequency among respondents

Code

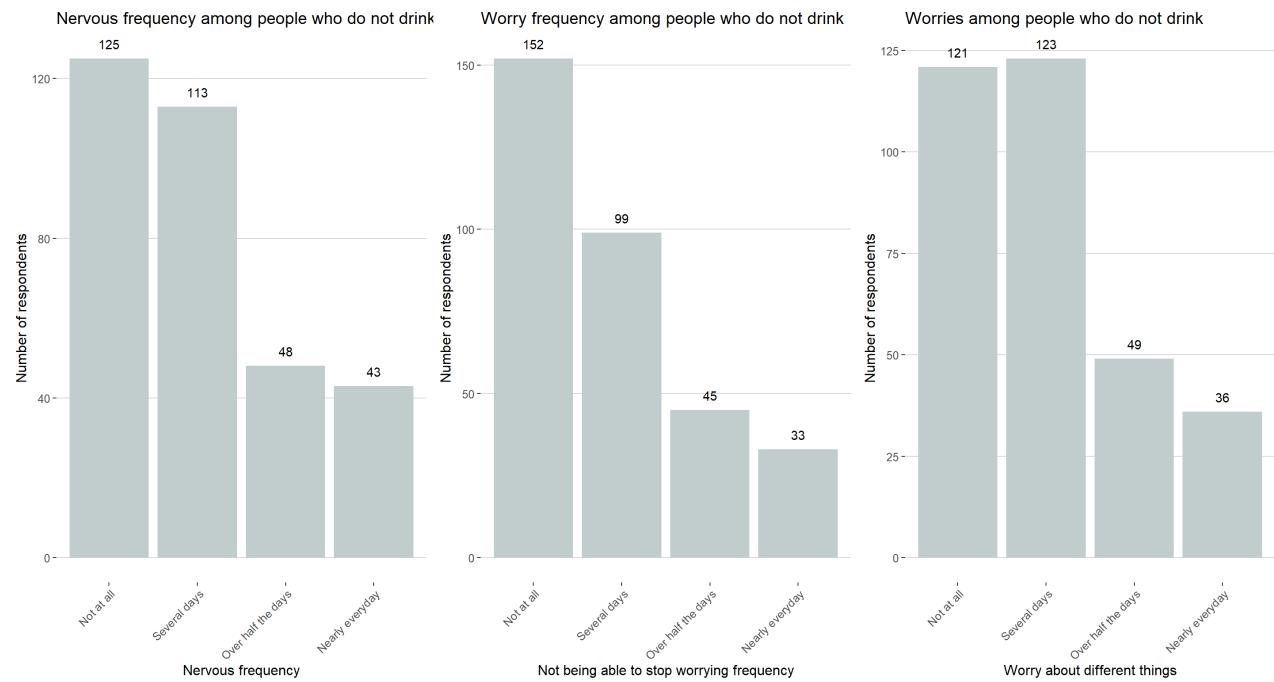


Figure 4.9: Stress levels between people who did not drink and who binged drinking

Code

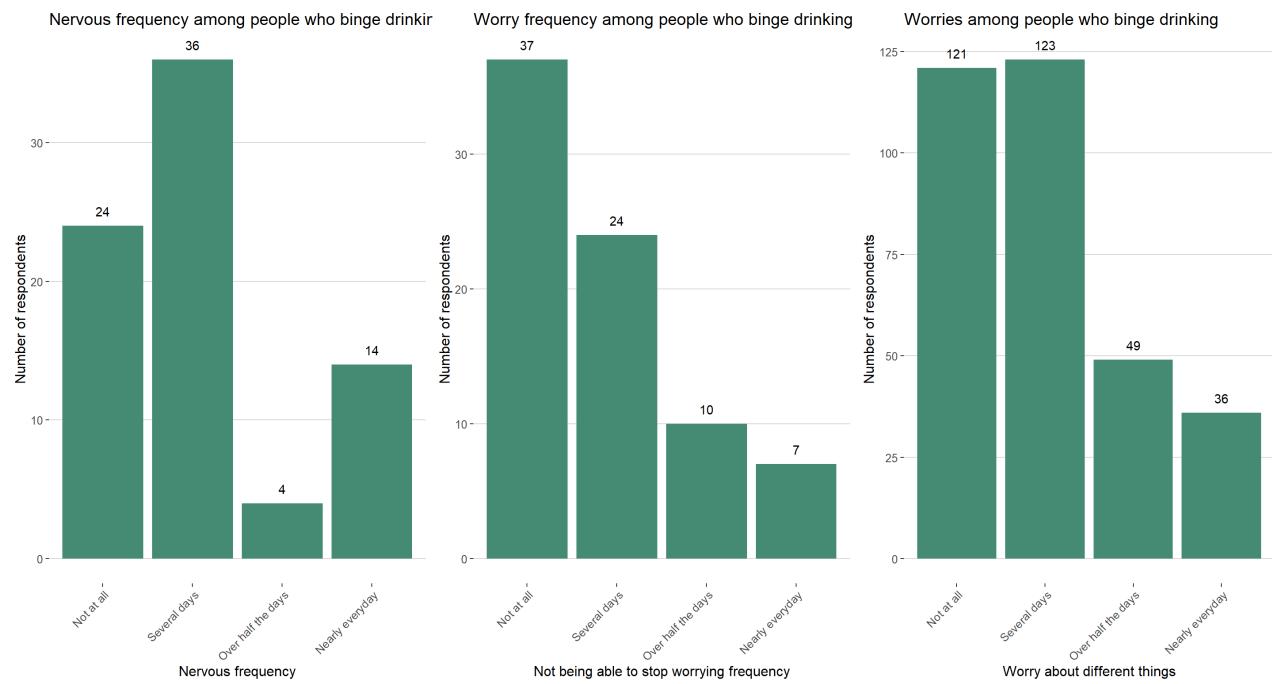


Figure 4.10: Stress levels between people who did not drink and who binged drinking

5 CONCLUSION

In conclusion, I would like to summarize my process for this analytic report:

1. I studied the data, the datamap, and filtered the 17 variables of interest
2. I evaluated n/a and anomalous values, weighing their impact against the value of the data set and subsequent analysis. For example, I maintained the value of 99 as "Not drinking" and used it for analysis. On the other hand, I removed values corresponding to option "Preferred not to answer" as they would not contribute to insights
3. I recoded the values using different methods - the factor form and the direct recoding - with respect to their nature, demographic or non-demographic values
4. I used different approaches - tables and graphs - to generate insights from the data set

5. Summary of findings:

- There are only 6 cases of COVID-19 positive out of 954 observations, making around 0.6% of the population studied
- There are equal numbers of male and female respondents
- The larger group of respondents is the youngest group of people from 18 to 39 years old
- Respondents who drink, reported to drink an average of 2 days during the past 7 days before taking the survey
- Females showed considerable tendencies to experience nervousness and worries more frequent than males. However, the relationship between this mental issues and COVID-19 need more investigation
- There is no significant difference between people who did not drink during the past 7 days and people who binged drinking the same period of time
- People who binged drinking might experience more prolonged nervous episodes but more investigation required before coming to a definitive conclusion

6 REFERENCE

[1] COVID-19 national survey dashboard. (n.d.). CAMH. <https://www.camh.ca/en/health-info/mental-health-and-covid-19/covid-19-national-survey> (<https://www.camh.ca/en/health-info/mental-health-and-covid-19/covid-19-national-survey>)