CAMH Data set Analysis

2003549

2022-11-15

About the data

The data was collected by two groups, The Centre for Addiction an Mental Health (CAMH) and Delvinia. They released two surveys (I will be analysing the first one), the first survey was conducted on May 8th - 12th 2020 receiving responses from just over 1000 Canadians. These surveys were conducted to see whether Canadians' mental health and substance usage changed from the COVID-19 pandemic. The dataset can be found on this website, and the survey questions can be found as the first row of the .csv file.

A lot of the columns of the data frame are duplicates, they have the responses to the question as the respondents answers (e.g. 'not at all', 'somewhat', 'usually' or 'all the time') and given the numeric value as well (e.g 1,2,3 or 4).

I removed the worded rows to only give the numeric values and also removed some rows of the dataset which are not relevant to my analysis. This includes rows relating to alcohol and substance usage.

Data Cleaning

```
# data cleaning make subset of dataset with only the
# required variables
camh <- camh[, -c(2:5, 8, 10, 18, 19, 22, 24, 25, 27, 29, 31,
    33, 35, 37, 39, 41:49, 50, 52, 54, 58:64, 65, 66, 68, 70,
    72, 74)]
# change variable names
names(camh) <- c("ID", "Province", "ProvinceCode", "AgeRange",
    "Sex", "Question4.1", "Question4.2", "Question4.3", "Question4.4",
    "Question4.5", "Question4.6", "Question5", "ChangeInEmployment",
    "ChangeInWorkingHours", "Question7", "Question8.1", "Question8.2",
    "Question8.3", "Question8.4", "Question8.5", "Question8.6",
    "Question8.7", "Question20.1", "Question20.2", "Question20.3",
    "Question23", "LivingStatus", "MaritalStatus", "Ethnicity",
    "HouseholdIncome", "LivingArea")
```

```
# set variables as factors
camh$AgeRange <- as.factor(camh$AgeRange)
camh$Sex <- as.factor(camh$Sex)
camh$LivingStatus <- as.factor(camh$LivingStatus)
camh$MaritalStatus <- as.factor(camh$MaritalStatus)
camh$HouseholdIncome <- factor(camh$HouseholdIncome, levels = c("Less than 40k",
    "40-79k", "80-119k", "120k+", "Prefer not to answer"))
```

```r
camh$Ethnicity <- as.factor(camh$Ethnicity)
camh$LivingArea <- as.factor(camh$LivingArea)
camh$ChangeInEmployment <- as.factor(camh$ChangeInEmployment)
camh$ChangeInWorkingHours <- as.factor(camh$ChangeInWorkingHours)
camh$Question23 <- as.factor(camh$Question23)

# inspect the data set
str(camh)
```

```
## 'data.frame':    1005 obs. of  31 variables:
##  $ ID                  : int  94 2484 2154 1902 2997 2276 2642 6719 1749 2791 ...
##  $ Province            : chr  "Alberta" "Prince Edward Island" "Ontario" "Ontario" ...
##  $ ProvinceCode        : chr  "AB" "QC/ATL" "ON" "ON" ...
##  $ AgeRange            : Factor w/ 3 levels "18-39","40-59",..: 1 2 2 1 2 2 3 3 3 1 ...
##  $ Sex                 : Factor w/ 3 levels "Female","Male",..: 2 2 2 1 1 2 1 2 2 2 ...
##  $ Question4.1         : int  0 0 1 0 0 0 0 0 0 0 ...
##  $ Question4.2         : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Question4.3         : int  0 1 0 1 0 0 0 0 0 0 ...
##  $ Question4.4         : int  0 0 1 1 0 1 1 1 0 0 ...
##  $ Question4.5         : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ Question4.6         : int  1 1 1 0 1 0 0 0 0 0 ...
##  $ Question5           : int  2 3 3 3 3 3 4 3 1 3 ...
##  $ ChangeInEmployment  : Factor w/ 8 levels "I am not currently working, or I have been laid off/let
##  $ ChangeInWorkingHours: Factor w/ 6 levels "","Decreased a lot",..: 6 5 6 1 3 1 6 1 1 6 ...
##  $ Question7           : int  2 3 4 2 4 3 4 4 3 3 ...
##  $ Question8.1         : int  1 1 4 3 4 3 3 2 3 1 ...
##  $ Question8.2         : int  1 2 4 2 4 2 2 2 2 1 ...
##  $ Question8.3         : int  1 1 3 4 4 2 3 2 2 2 ...
##  $ Question8.4         : int  1 1 3 3 4 2 3 2 2 1 ...
##  $ Question8.5         : int  1 2 2 1 4 2 3 1 1 1 ...
##  $ Question8.6         : int  1 2 4 4 4 1 2 2 2 1 ...
##  $ Question8.7         : int  1 1 4 2 4 2 2 2 1 2 ...
##  $ Question20.1        : int  1 1 3 2 3 2 2 2 1 1 ...
##  $ Question20.2        : int  2 1 1 3 4 1 4 2 1 1 ...
##  $ Question20.3        : int  4 3 3 1 2 4 3 1 2 3 ...
##  $ Question23          : Factor w/ 8 levels "1","2","3","4",..: 2 4 3 2 2 2 1 2 2 4 ...
##  $ LivingStatus        : Factor w/ 3 levels "Live alone","Live with others",..: 2 2 2 2 2 2 1 2 2 2
##  $ MaritalStatus       : Factor w/ 7 levels "Divorced","Living with a partner",..: 3 3 3 2 3 7 1 3 3
##  $ Ethnicity           : Factor w/ 11 levels "Asian ? East (e.g., Chinese, Japanese, Korean)",..: 11
##  $ HouseholdIncome     : Factor w/ 5 levels "Less than 40k",..: 4 2 3 4 5 2 1 3 4 4 ...
##  $ LivingArea          : Factor w/ 3 levels "Rural area","Suburban area",..: 3 3 2 3 3 3 2 3 3 2 ...
```

Questions 4 relate to exposure to COVID-19 and include whether the participant is of high risk of catching COVID-19 or if relatives are of high risk/have tested positive for COVID-19. These are boolean variables given as 1 (yes) or 0 (no)

Question 5 relates to how worried individuals are about how the pandemic will affect their financial situation from Not worried at all (1) to very worried (4)

Questions 6 relates to individuals employment and whether they have been affected by the pandemic.

Question 7 relates to how worried an individual is that someone close to them will get ill with COVID-19

Questions 8 relate to symptoms of anxiety and stress where participants responded on a scale of 1 (Not at all) to 4 (Nearly everyday).

Questions 20 asks the participants whether in the past week they have felt lonely, depressed or hopeful about the future. This is scored from 1 (some of the time (less than 1 day)) up to 4 (Most or all of the time (5-7 days))

Question 23 asks how many people, including yourself live in your household.

With the Ethnicity column, there are many different options with long names. I merged some of the factors and renamed them to make it cleaner. I will group them to four levels: White, Black, Asian and minority Ethnics (BAME), Other and Prefer Not to Answer.

```
# changing the names and combining the different
# ethnicities.
levels(camh$Ethnicity) <- list(White = "White (European, North American)",
    Asian = c("Asian ? East (e.g., Chinese, Japanese, Korean)",
        "Asian ? South (e.g., Indian, Pakistani, Sri Lankan)",
        "Asian ? South East (e.g., Malaysian, Filipino, Vietnamese)",
        "Middle Eastern (e.g., Egyptian, Iranian, Lebanese)"),
    Black = "Black (Africa, Caribbean, North American)", Other = c("Latin American (e.g., Argentinean, (
        "Mixed heritage (e.g. Black ? North American & White ? North American)",
        "Indigenous (First Nations, Inuit, M,tis)", "Other"),
    `Prefer not to answer` = "Prefer not to answer")
# merge factor levels together
levels(camh$Question23) <- list(`1` = "1", `2` = "2", `3` = "3",
    `4+` = c("4", "5", "6", "7"), `Prefer not to answer` = "Prefer not to answer")
# merge factor levels together
levels(camh$MaritalStatus) <- list(`Married/Living with Partner` = c("Living with a partner",
    "Married"), `Divorced/Seperated` = c("Divorced", "Separated"),
    `Never Married/Widowed` = c("Widowed", "Never married"),
    `Prefer not to answer` = "Prefer  not to answer")
```

I will also rename and reduce the levels for Question 6.1: How have physical distancing measures due to the COVID-19 pandemic affected your employment situation?

```
# changing the names and combining the different answers
levels(camh$ChangeInEmployment) <- list(`Laid off/Let go` = "I am not currently working, or I have been
    `Working from home instead` = "I have continued working, but now I am working from home instead of n
    `Previously not working, now back in work` = "I was previously not working/ laid off/let go due to
    `Previously working from home, now outside` = "I was working from home due to the pandemic, but now
    `No change to work` = c("No change - I have continued working outside my home, as I always did",
        "No change - I have continued working from home, as I always did"),
    `Unemployed (No change)` = "No change - I was not employed prior to the pandemic (e.g., retired, st
    Other = "Other")
# combining the different answers
levels(camh$ChangeInEmployment) <- list(`Laid off/Let go` = "Laid off/Let go",
    `Working from home instead` = "Working from home instead",
    `Unemployed (No change)` = "Unemployed (No change)", `No change/Other` = c("No change to work",
        "Previously not working, now back in work", "Previously working from home, now outside",
        "Other"))
```

I will rename the answers to question 6.1 as one answer has no name and is blank

```
# changing the names and combining the different answers
levels(camh$ChangeInWorkingHours) <- c("NA", "Decreased a lot",
    "Decreased somewhat", "Increased a lot", "Increased somewhat",
    "No change")
```

To start with I will sum up the boolean scores of question 4. This new row, will be a higher number for individuals who have a high risk of being affected by COVID-10 or have had close friends exposed to COVID-19.

I will also sum up question Question 8 and question 20, as these can measure how anxious, depression and stressed each individual has been, which will be our response variable.

Before I sum up the scores, question 20.3 asks: In the PAST 7 DAYS, how often have you felt hopeful about the future?. Where 1 is 'Rarely or none of the time (less than 1 day)' and 4 is 'Most or all of the time (5-7 days)' just like the other variables. So I must invert the values of this row before summing, meaning 1 will become 4, 2 will become 3 and vice versa. As a higher score is meant to be more negative

```r
# inverting question 20.3
camh$newQuestion20.3 <- camh$Question20.3
camh <- within(camh, newQuestion20.3[Question20.3 == 1] <- 4)
camh <- within(camh, newQuestion20.3[Question20.3 == 2] <- 3)
camh <- within(camh, newQuestion20.3[Question20.3 == 3] <- 2)
camh <- within(camh, newQuestion20.3[Question20.3 == 4] <- 1)
```

```r
# summing the columns of question 4 and 8 + 20 sum of
# question 4
camh <- mutate(camh, Q4Sum = rowSums(dplyr::select(camh, Question4.1,
    Question4.2, Question4.3, Question4.4, Question4.5, Question4.6)))

# factor variable of the sum of Q4. Since very few values
# from 3 onwards we will group these together
camh$Q4factor <- factor(camh$Q4Sum)
levels(camh$Q4factor) <- list(`0` = "0", `1` = "1", `2` = "2",
    `3+` = c("3", "4", "5", "6"))


# sum of question 8
camh <- mutate(camh, Q8Sum = rowSums(dplyr::select(camh, Question8.1,
    Question8.2, Question8.3, Question8.4, Question8.5, Question8.6,
    Question8.7)))

# sum of question 20
camh <- mutate(camh, Q20Sum = rowSums(dplyr::select(camh, Question20.1,
    Question20.2, newQuestion20.3)))

# question 8 + 20
camh <- mutate(camh, TotalSum = rowSums(dplyr::select(camh, Q8Sum,
    Q20Sum)))

summary(camh$TotalSum)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   13.00   18.00   19.56   24.00   40.00
```

```r
print(table(camh$Q4Sum))
```

```
##
##   0   1   2   3   4   5   6
## 409 411 118  48  15   3   1
```

Exploratory Data Analysis

```
# create box plot of variables
kable(table(camh$Q4Sum))
```

Var1	Freq
0	409
1	411
2	118
3	48
4	15
5	3
6	1

```
ggplot(data = camh, aes(x = factor(Q4Sum), y = TotalSum)) + geom_boxplot() +
    labs(x = "Sum of question 4 scores", y = "sum of Question 8 and Question 20 scores")
```
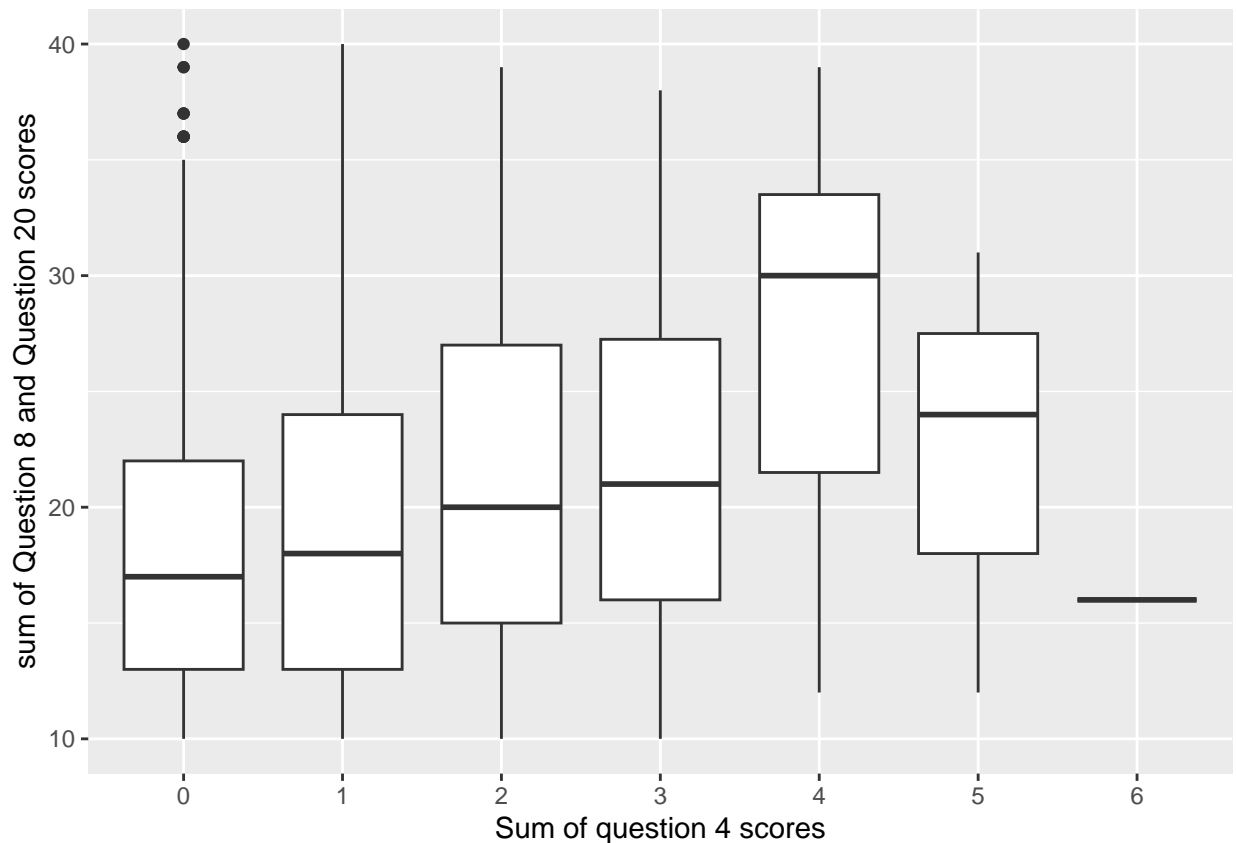

The boxplot shows a positive trend of Question 8 and 20 scores against Question 4 scores. However, the number of individuals with high scores for question 4 are extremely small. Only 2% of the data points have scores of 4 or above. Meaning this is not a reliable score.

```
# create box plot of variables
kable(summary(camh$Sex), caption = "Distribution of Sex") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 1: Distribution of Sex

	x
Female	498
Male	504
Other	3

```
ggplot(camh, aes(x = Sex, y = TotalSum)) + geom_boxplot()
```
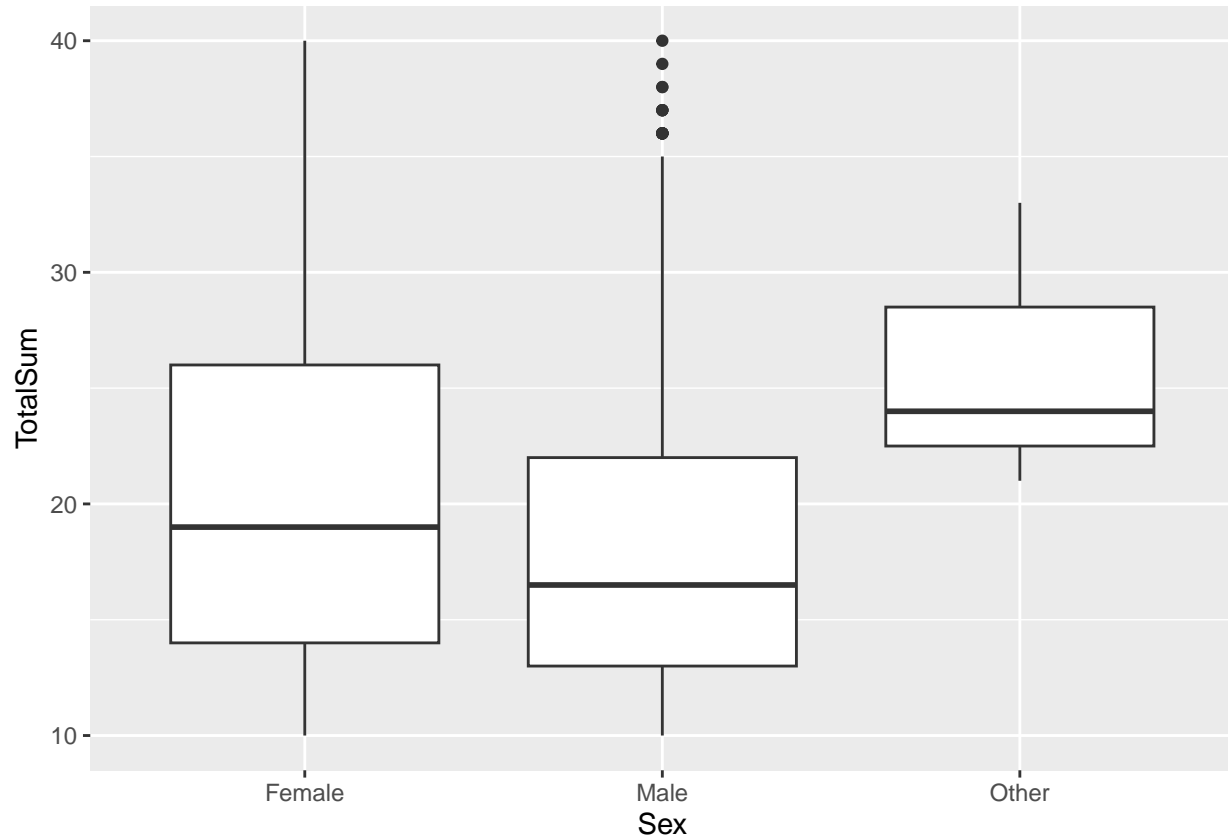

We can see that women had an overall high score compared to men, and the distribution of men and women on the survey are fairly even. There are only 3 individuals who chose other, meaning there are not enough data points to accurately compare people with other sex compared to male and female.

```
# create box plot of variables
kable(summary(camh$AgeRange), caption = "Distribution of Age Range") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 2: Distribution of Age Range

	x
18-39	394
40-59	306
60+	305

```
ggplot(camh, aes(x = AgeRange, y = TotalSum)) + geom_boxplot()
```
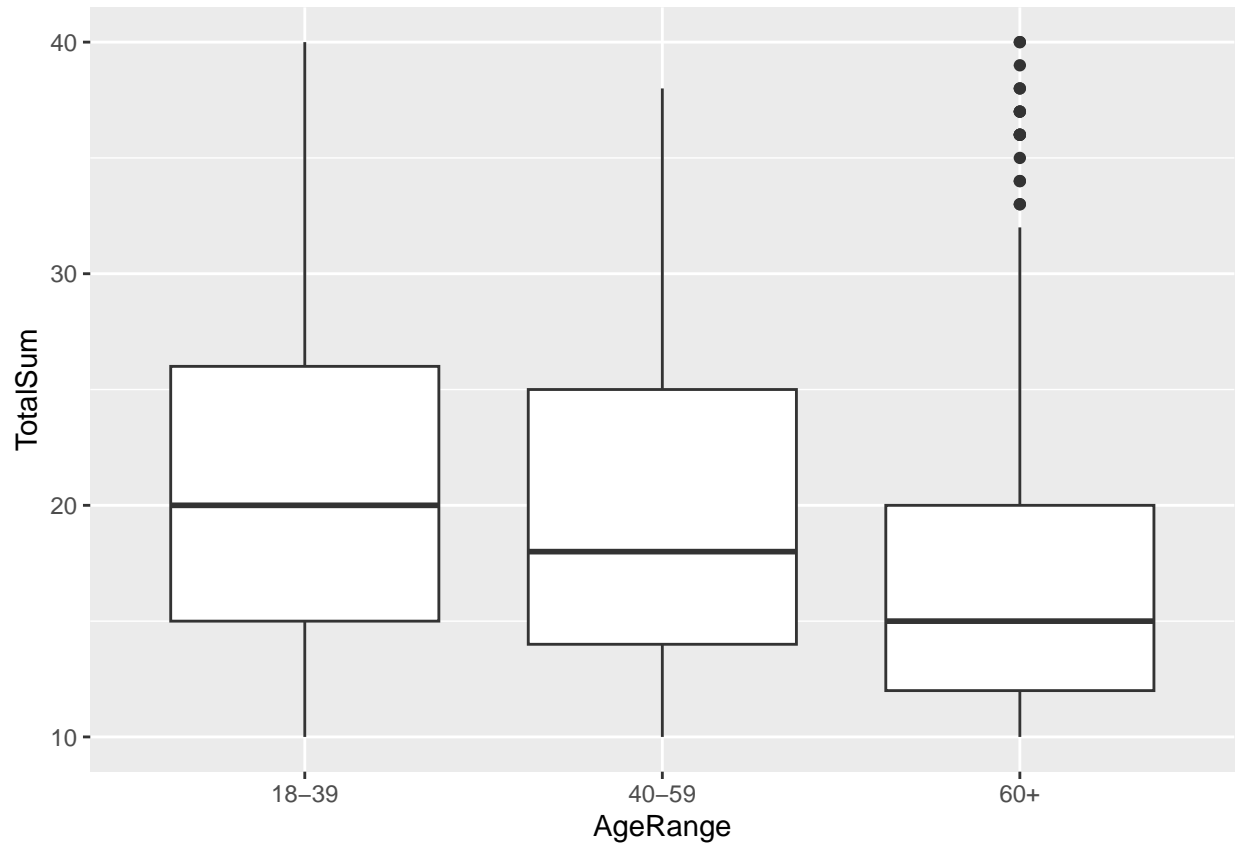

There is a decline in individuals' score as the age groups increases. This suggests evidence of younger individuals experiencing worse mental health compared to older individuals.

```
# create box plot of variables
kable(summary(camh$Ethnicity), caption = "Distribution of Ethnicity") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 3: Distribution of Ethnicity

	x
White	698
Asian	211
Black	16
Other	60
Prefer not to answer	20

```
ggplot(camh, aes(x = Ethnicity, y = TotalSum)) + geom_boxplot()
```
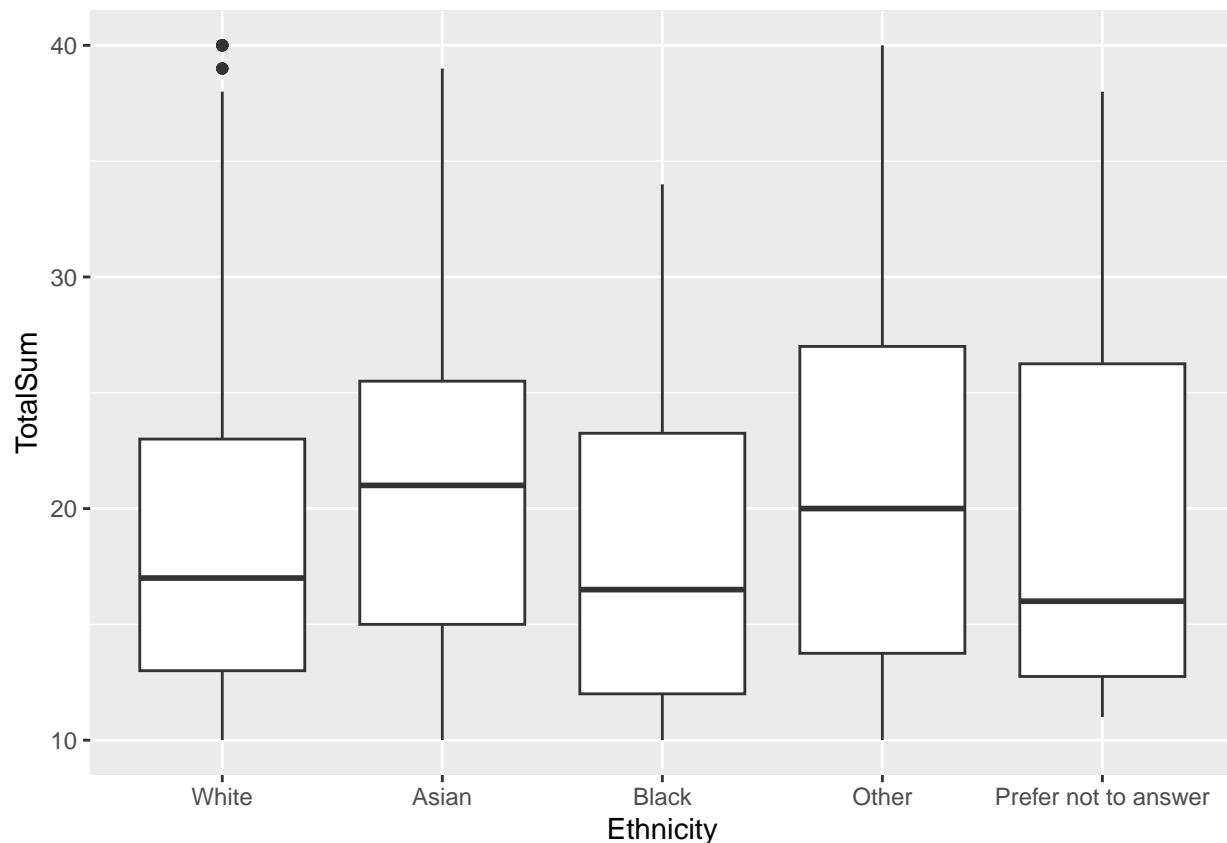
Majority of the individuals are white (69%) with the next common ethnicity group being East Asian (13%).

The box plot shows that South Asians had a higher score than all other groups followed by other Asian ethnicity groups. Black and white individuals had lower scores in comparison to the other ethnic groups.

```
# create box plot of variables
kable(summary(camh$LivingStatus), caption = "Distribution of Living Status") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 4: Distribution of Living Status

	x
Live alone	205
Live with others	797
Prefer not to answer	3

```
ggplot(camh, aes(x = LivingStatus, y = TotalSum)) + geom_boxplot()
```

8

Individuals who live with others gave slightly higher scores compared to individuals who live alone, although they are very close to each other.

```
# create box plot of variables
kable(as.array(summary(camh$Question23)), caption = "Distribution of number of people in household") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 5: Distribution of number of people in household

	x
1	205
2	406
3	176
4+	215
Prefer not to answer	3

```
ggplot(camh, aes(x = Question23, y = TotalSum)) + geom_boxplot()
```
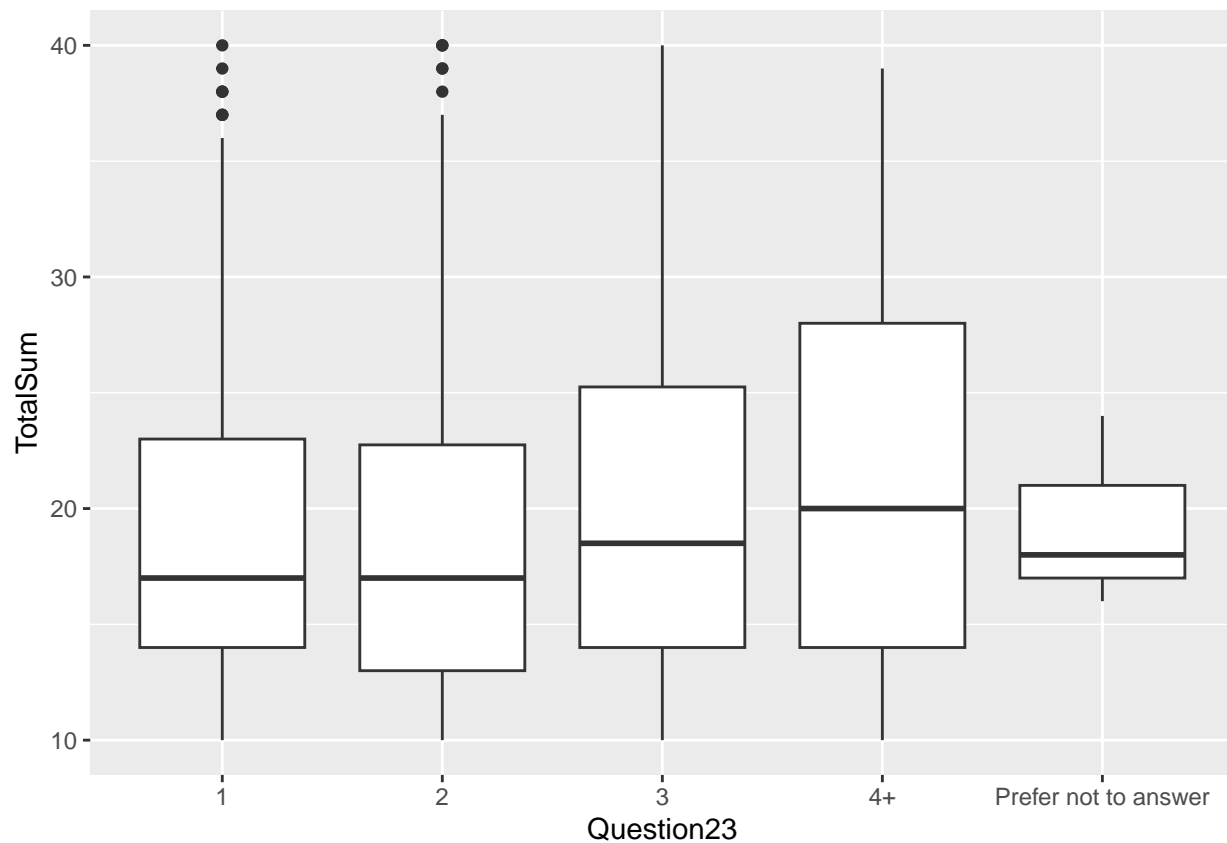
```r
# create box plot of variables
kable(summary(camh$LivingArea), caption = "Distribution of Living Area") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 6: Distribution of Living Area

	x
Rural area	158
Suburban area	382
Urban area	465

```r
ggplot(camh, aes(x = LivingArea, y = TotalSum)) + geom_boxplot()
```
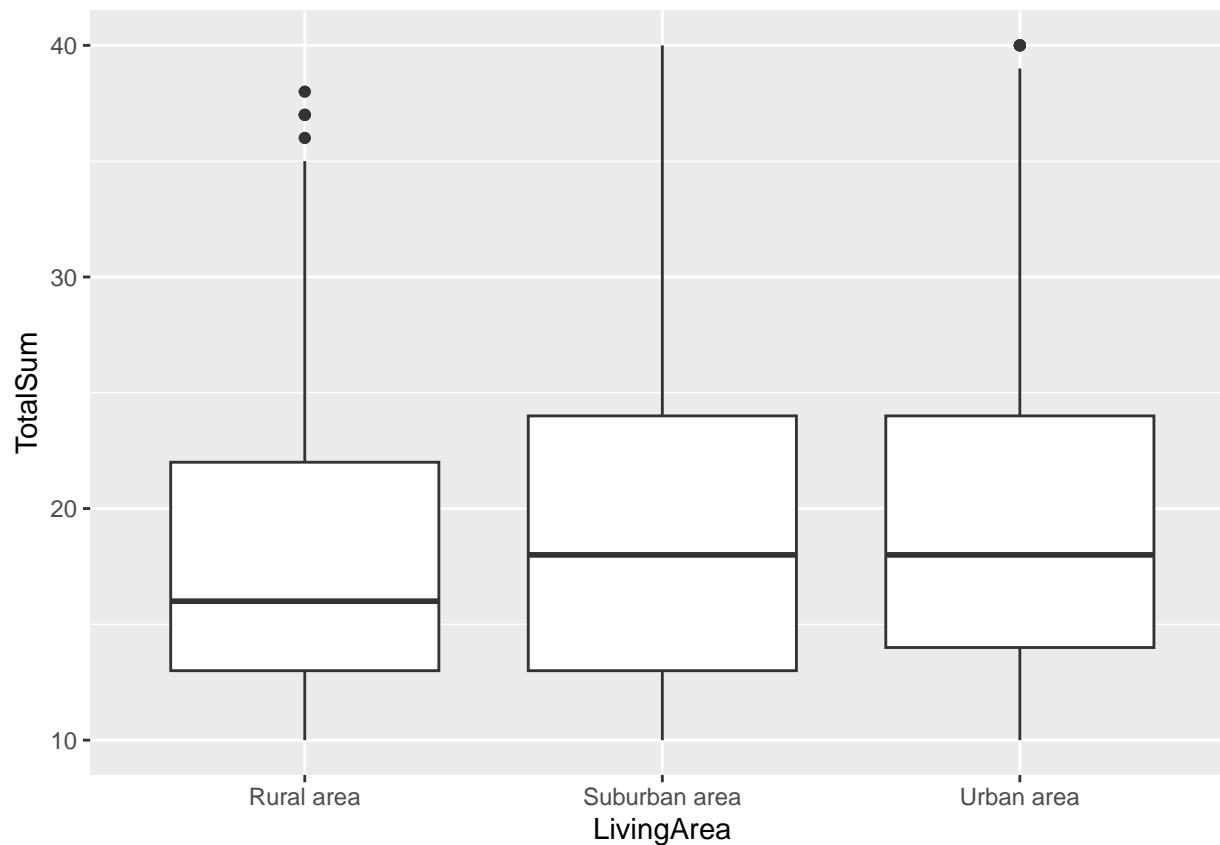
Individuals in rural areas had lower scores compared to individuals living in suburban and urban areas, with both of those being having almost identical scores.

```
# create box plot of variables
kable(summary(camh$HouseholdIncome), caption = "Distribution of Household Income") %>%
    kable_styling(latex_options = "HOLD_position")
```

Table 7: Distribution of Household Income

	x
Less than 40k	128
40-79k	268
80-119k	226
120k+	217
Prefer not to answer	166

```
ggplot(camh, aes(x = HouseholdIncome, y = TotalSum)) + geom_boxplot()
```
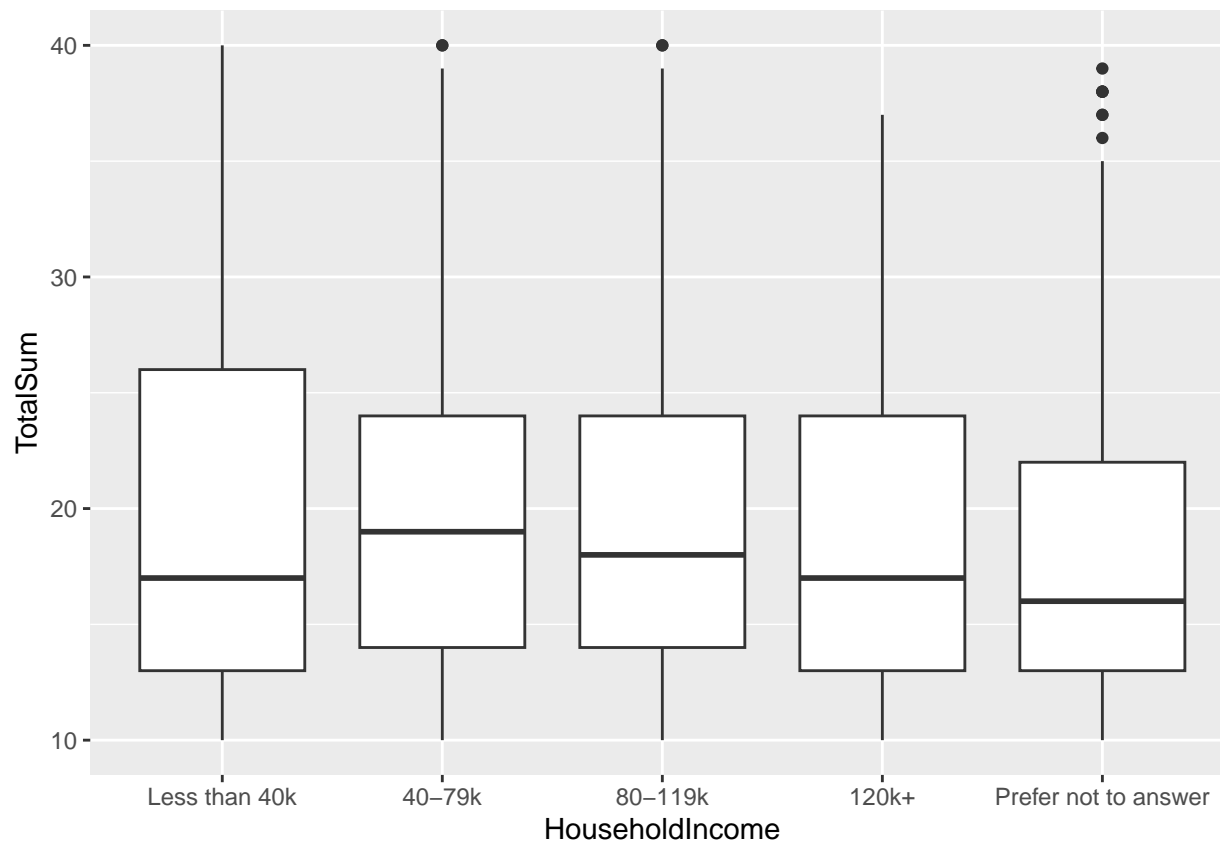
Between earning 40k - 120k+, we can see a slight decrease of scores as the income increases, however individuals individuals less than 40k have scores similar to earning 80k-120k+. Individuals who prefered not to answer had the lowest scores overall.
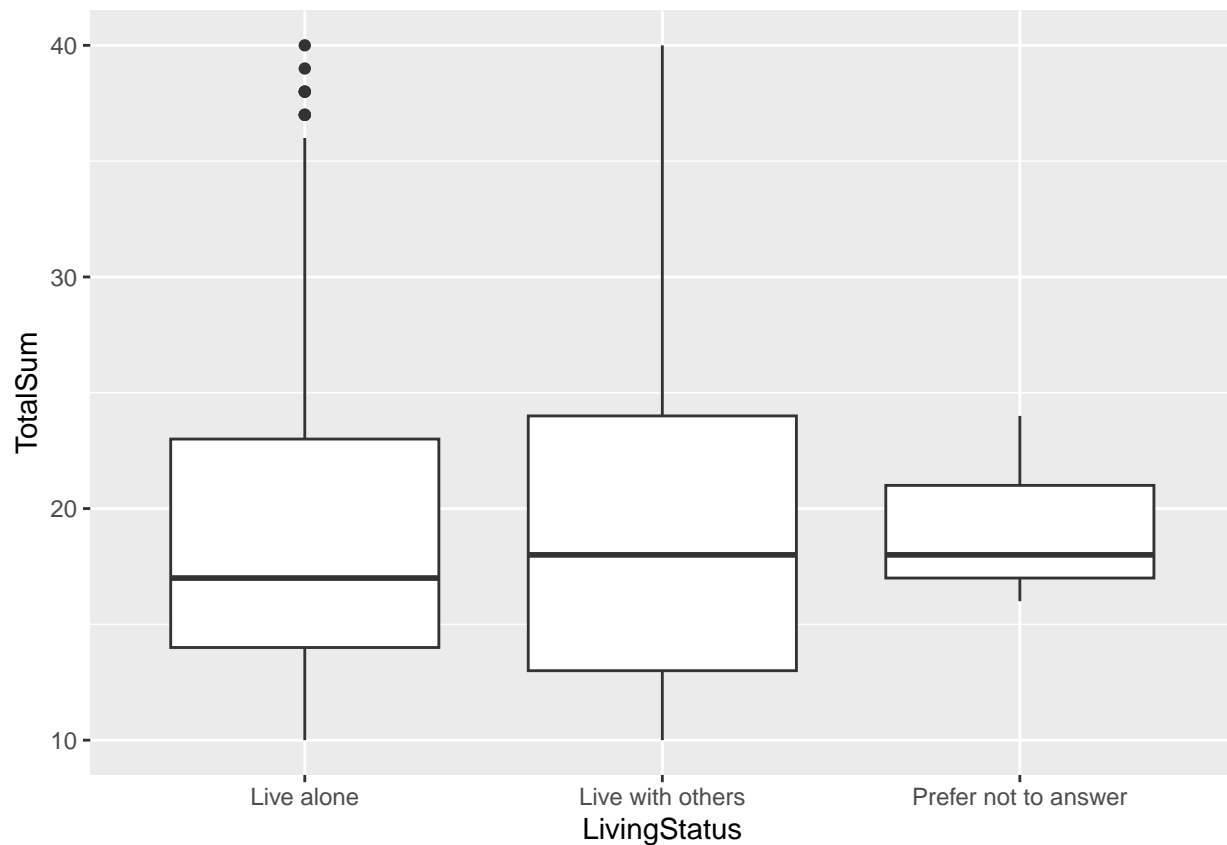
```
# create box plot of variables
kable(summary(camh$MaritalStatus), caption = "Distribution of Marital Status") %>%
    kable_styling(latex_options = "HOLD_position")
```

Table 8: Distribution of Marital Status

	x
Married/Living with Partner	613
Divorced/Seperated	88
Never Married/Widowed	291
Prefer not to answer	13

```
ggplot(camh, aes(x = MaritalStatus, y = TotalSum)) + geom_boxplot()
```
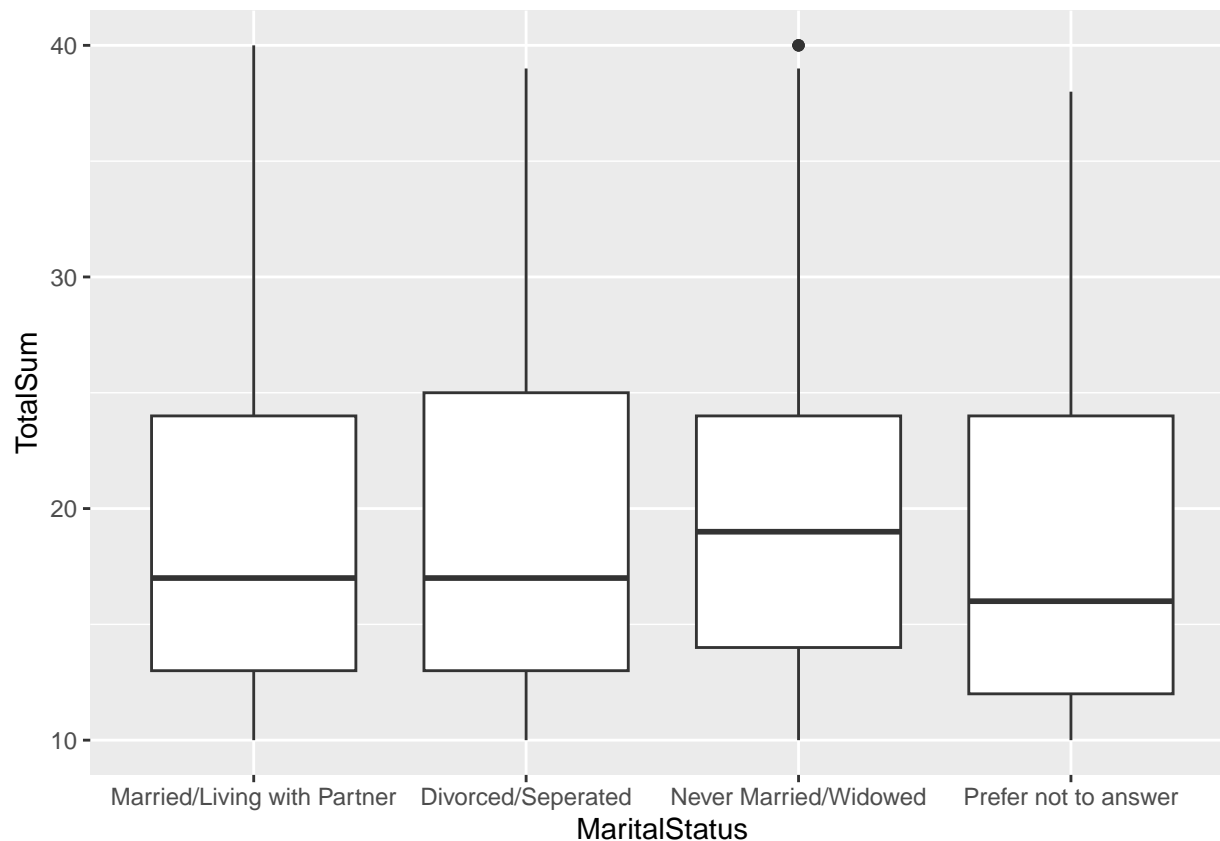
Individuals who were never married or separated had higher scores. Where as widowed, married and divorced had lower scores. One reason for widowed scores being so low is that widowed individuals are more likely to be older and as we saw previously older individuals had lower scores. Similarly to income, individuals who preferred not to answer also had lower scores.

```
# create box plot of variables
kable(summary(camh$ChangeInWorkingHours), caption = "Distribution of Question 6.2") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 9: Distribution of Question 6.2

	x
NA	455
Decreased a lot	41
Decreased somewhat	76
Increased a lot	41
Increased somewhat	75
No change	317

```
ggplot(camh, aes(x = ChangeInWorkingHours, y = TotalSum)) + geom_boxplot() +
    labs(x = "how has the number of hours you are working for pay been affected by the COVID-19 pandemi
```
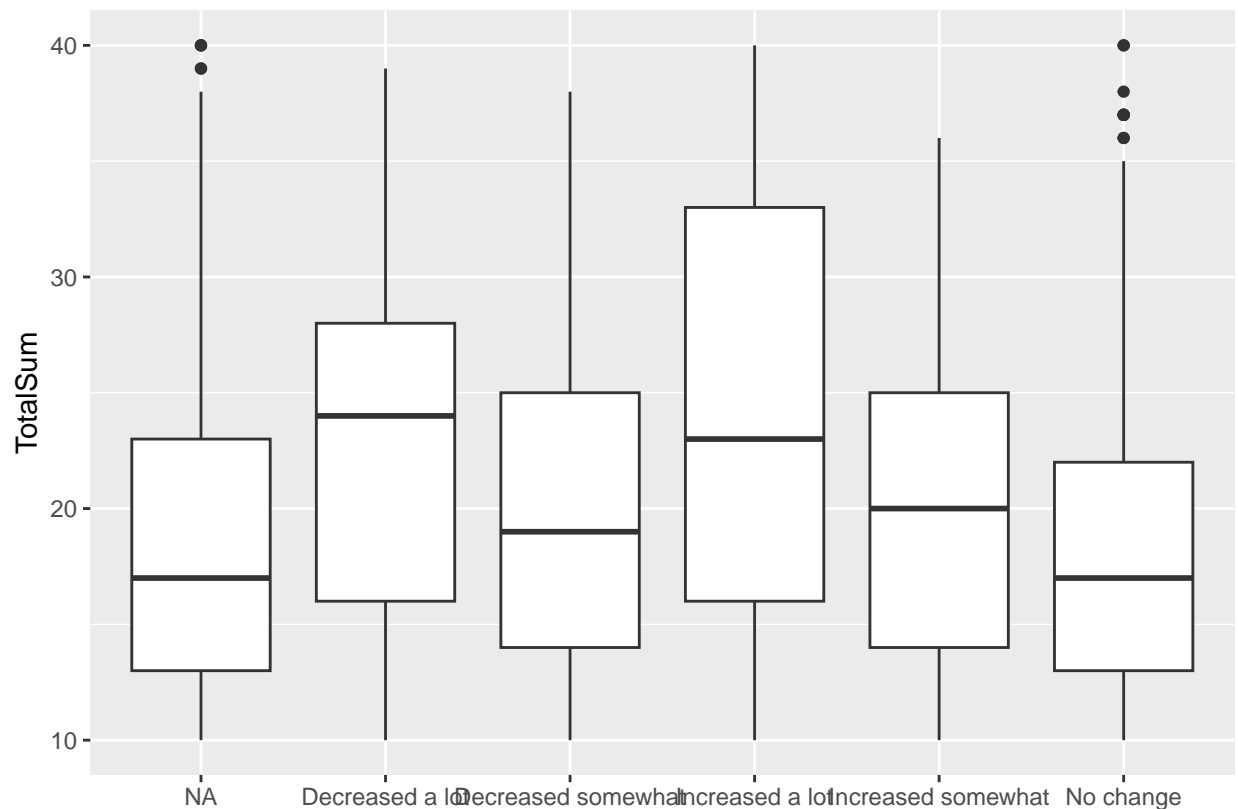
how has the number of hours you are working for pay been affected by the COVID−19 pandem

First thing to notice is that 455 of the responses are blank, most likely na. Individuals who decreased or increased working hours had higher scores, with individuals who decreased hours a lot having the highest scores. Individuals who had no change in working hours had the lowest scores.

```
# create box plot of variables
kable(summary(camh$ChangeInEmployment), caption = "Distribution of Question 6.1") %>%
    kable_classic(latex_options = "HOLD_position")
```

Table 10: Distribution of Question 6.1

	x
Laid off/Let go	172
Working from home instead	255
Unemployed (No change)	283
No change/Other	295

```
ggplot(camh, aes(x = ChangeInEmployment, y = TotalSum)) + geom_boxplot() +
    labs(x = "How have physical distancing measures due to the COVID-19 pandemic \naffected your employr
```

14
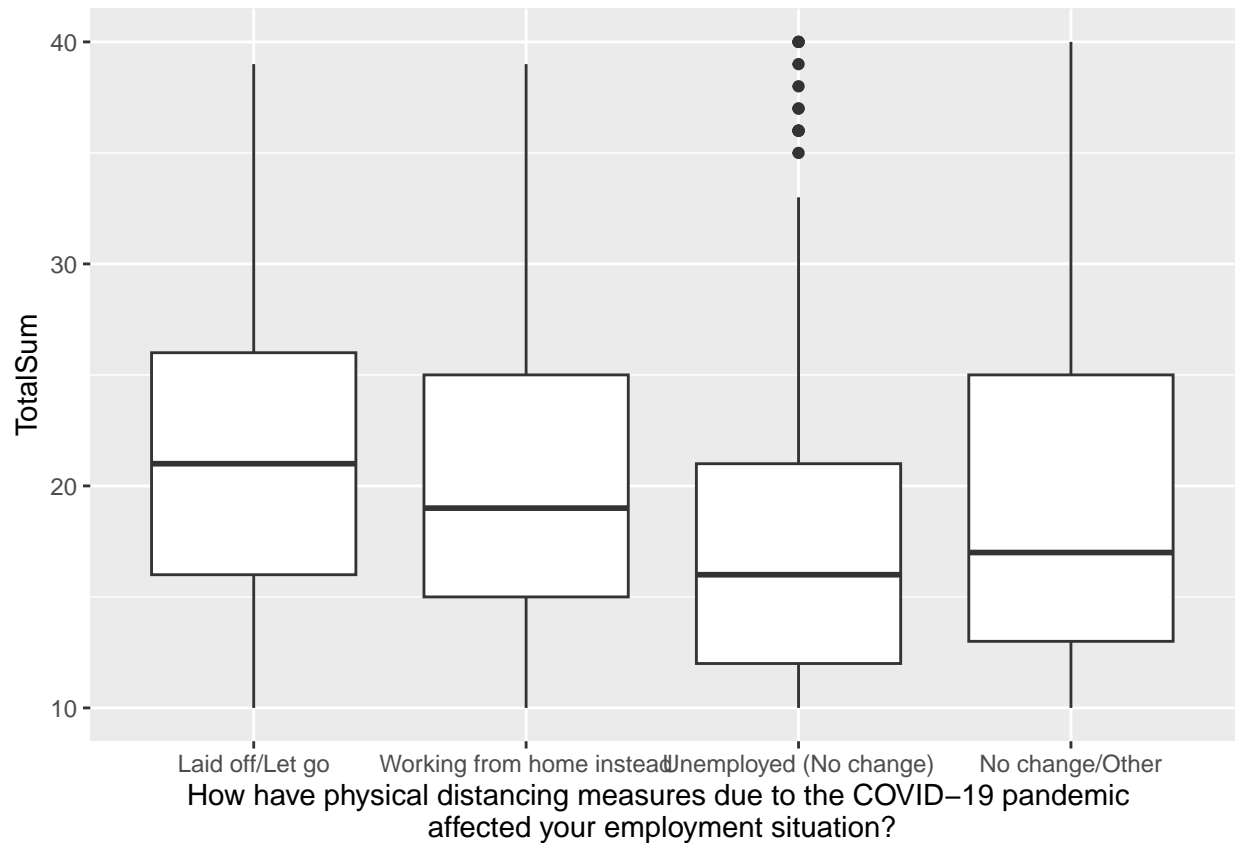
Visualisation

Summary Table:

```
# print summary table nicely
table1 %>%
    kbl() %>%
    kable_classic(latex_options = c("HOLD_position"), full_width = F) %>%
    add_indent(c(3:5, 7:9, 11:13, 15:18, 20:24, 26:28, 30:34,
        36:39, 41:46), level_of_indent = 1)  #%>% save_kable(file = 'CAMHSummaryTable.png', zoom = 1.5)
```

	Overall
n	1005
AgeRange (%)	
18-39	394 (39.2)
40-59	306 (30.4)
60+	305 (30.3)
Sex (%)	
Female	498 (49.6)
Male	504 (50.1)
Other	3 (0.3)
LivingStatus (%)	
Live alone	205 (20.4)
Live with others	797 (79.3)
Prefer not to answer	3 (0.3)
MaritalStatus (%)	
Married/Living with Partner	613 (61.0)
Divorced/Seperated	88 (8.8)
Never Married/Widowed	291 (29.0)
Prefer not to answer	13 (1.3)
HouseholdIncome (%)	
Less than 40k	128 (12.7)
40-79k	268 (26.7)
80-119k	226 (22.5)
120k+	217 (21.6)
Prefer not to answer	166 (16.5)
LivingArea (%)	
Rural area	158 (15.7)
Suburban area	382 (38.0)
Urban area	465 (46.3)
Ethnicity (%)	
White	698 (69.5)
Asian	211 (21.0)
Black	16 (1.6)
Other	60 (6.0)
Prefer not to answer	20 (2.0)
ChangeInEmployment (%)	
Laid off/Let go	172 (17.1)
Working from home instead	255 (25.4)
Unemployed (No change)	283 (28.2)
No change/Other	295 (29.4)
ChangeInWorkingHours (%)	
NA	455 (45.3)
Decreased a lot	41 (4.1)
Decreased somewhat	76 (7.6)
Increased a lot	41 (4.1)
Increased somewhat	75 (7.5)
No change	317 (31.5)
Q4Sum (mean (SD))	0.87 (0.95)
Q8Sum (mean (SD))	13.38 (5.88)
Q20Sum (mean (SD))	6.17 (2.10)

Plots

```r
# select only the Likert scale questions
newCamh = camh[16:25]
newCamh[1:10] <- lapply(newCamh[1:10], factor)
# rename the levels of the questions
levels(newCamh$Question8.1) <- list(`1 (Not at all/less \nthan 1 day)` = "1",
    `2 (Several days/\n1-2 days)` = "2", `3 (Over half the \ndays/3-4 days)` = "3",
    `4 (Nearly every \nday/5-7 days)` = "4")
# set as a likert using likert package
camh_likert = likert(newCamh)
# plot the heat map
plot(camh_likert, type = "heat", group.order = names(camh[16:25])) +
    ggtitle("Heat map showing the individual's responses about their mental \nwellbeing ")
```

Heat map showing the individual's responses about their mental wellbeing

	Mean (SD)	1 (Not at all/less than 1 day)	2 (Several days/ 1–2 days)	3 (Over half the days/3–4 days)	4 (Nearly every day/5–7 days)
Question8.7	1.83 (0.98)	48.2%	29.8%	12.6%	9.5%
Question8.6	2.01 (1.00)	37.8%	35.9%	14.1%	12.1%
Question8.5	1.73 (0.95)	54.1%	26.9%	10.8%	8.2%
Question8.4	1.99 (1.00)	38.9%	34.7%	14.7%	11.6%
Question8.3	2.01 (1.01)	37.7%	36.2%	13.6%	12.4%
Question8.2	1.80 (0.95)	49.0%	30.3%	12.2%	8.5%
Question8.1	2.01 (0.99)	36.1%	38.6%	13.1%	12.1%
Question20.3	2.42 (1.00)	21.0%	32.4%	29.8%	16.8%
Question20.2	1.83 (0.97)	48.7%	28.2%	14.6%	8.6%
Question20.1	1.77 (0.92)	50.4%	29.2%	13.8%	6.6%

Percent: 100, 75, 50, 25, 0

```r
#select only the Likert scale questions
newCamh = camh[16:25]
#set these variables as factors
newCamh[1:10] <- lapply(newCamh[1:10], factor)
#rename the levels to include the days of the week to make it clearer
levels(newCamh$Question8.1) <- list("1 (Not at all)"="1",
        "2 (Several days)"="2",
        "3 (Over half the days)"="3",
        "4 (Nearly every day)"="4")
newCamh <- newCamh %>% # rename the questions to show what symptom they ae asking about
```

17

```
  rename(
    "Nervousness/anxiousness (8.1)" = "Question8.1",
    "Unable to control worrying (8.2)" = "Question8.2",
    "Worrying about things  (8.3)" = "Question8.3",
    "Trouble relaxing (8.4)" = "Question8.4",
    "Found it hard to sit still (8.5)" = "Question8.5",
    "Feeling easily annoyed/irritable (8.6)" = "Question8.6",
    "Feeling afraid (8.7)" = "Question8.7",
    "Feeling Depressed (20.1)" = "Question20.1",
    "Feeling Lonely (20.2)" = "Question20.2",
    "Feeling Hopeful about the future (20.3)" = "Question20.3"
  )
#set as a likert using likert package
camh_likert = likert(newCamh)
plot(camh_likert, centered=FALSE, group.order = names(newCamh[1:10])) + ggtitle("How often has the foll
```


How often has the following occurred to you...

```
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)

newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
```

```
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$Q4factor))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```
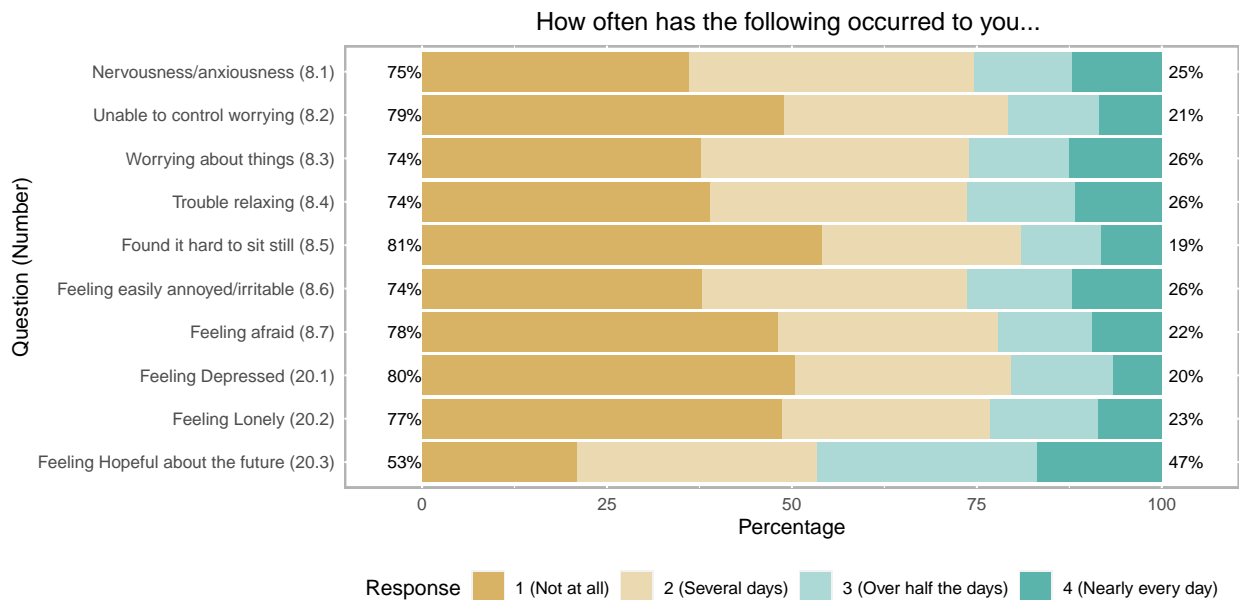

How often has the following occurred to you?

```
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)

newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
```

```r
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$AgeRange))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```

How often has the following occurred to you?
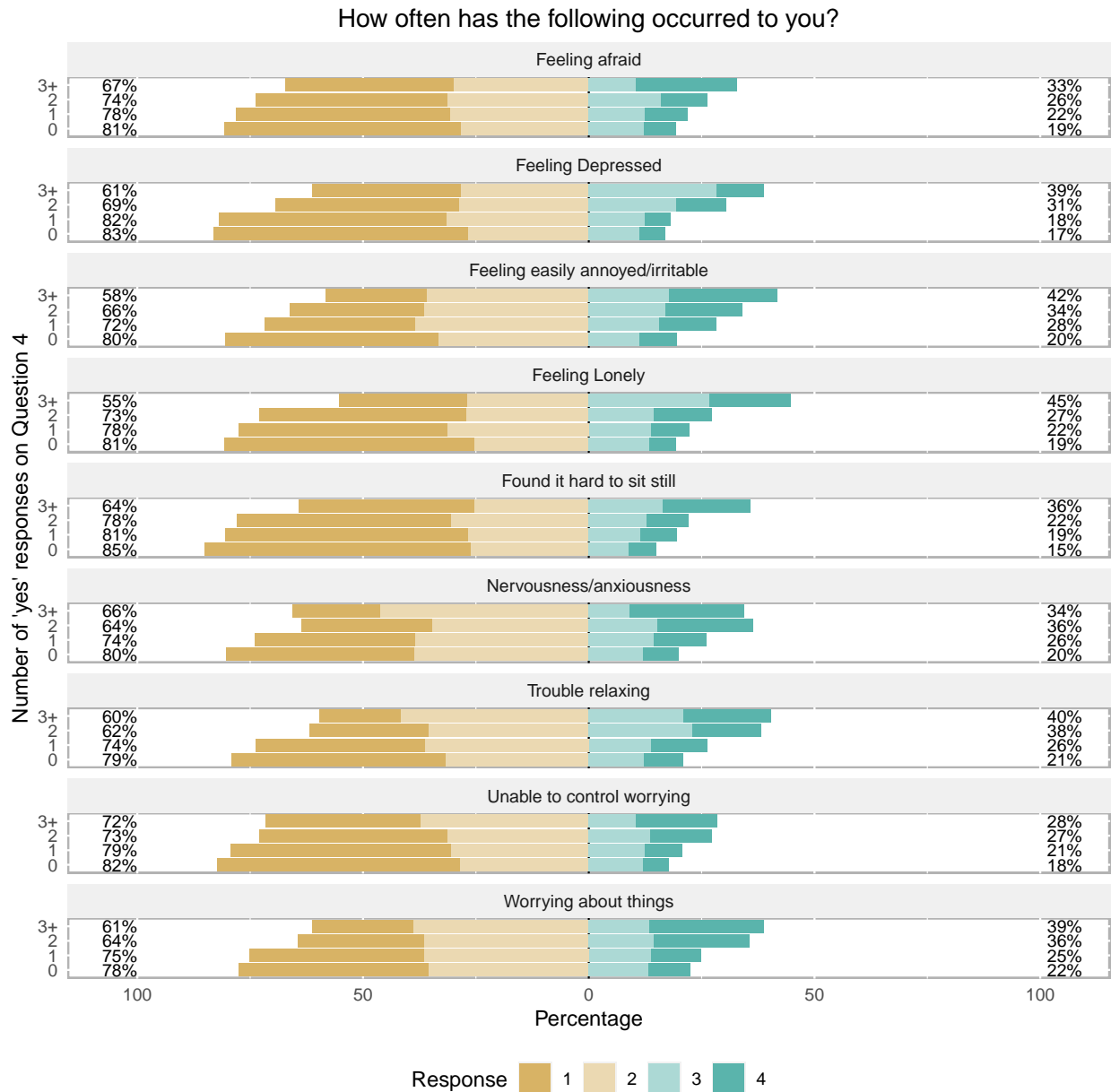


```
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)

newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
```

```
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$ChangeInEmployment))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```

How often has the following occurred to you?
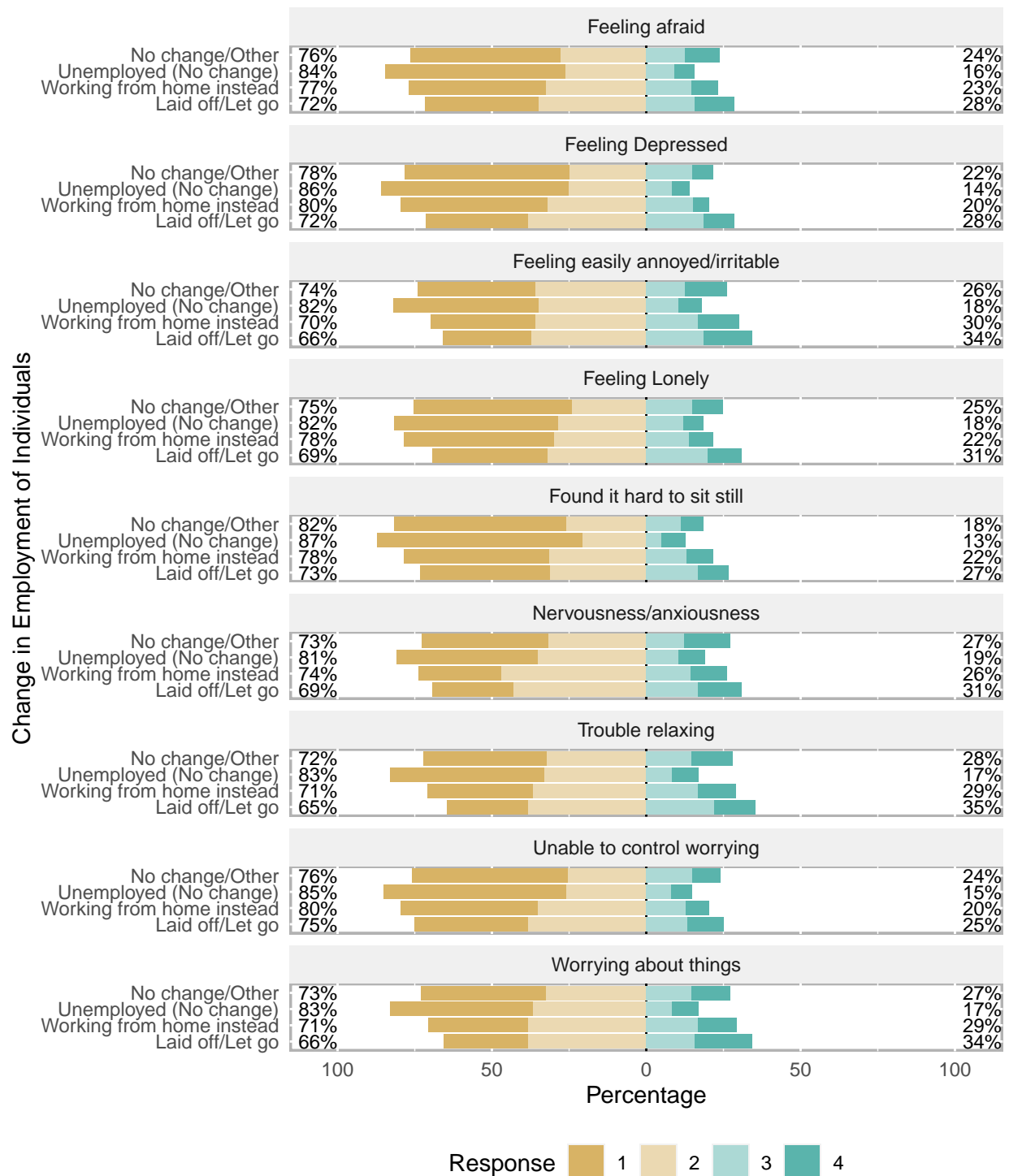
Below the three individuals who responded with Other for their sex have been removed as their sample size is too small.

```r
#select only the Likert scale questions
newCamh = camh[c(5,16:24)]
```

```r
#set these variables as factors
newCamh[1:10] <- lapply(newCamh[1:10], factor)
newCamh = newCamh[!newCamh$Sex == "Other",]
newCamh <- droplevels(newCamh)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh[-c(1)], grouping=newCamh$Sex)
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```

How often has the following occurred to you?
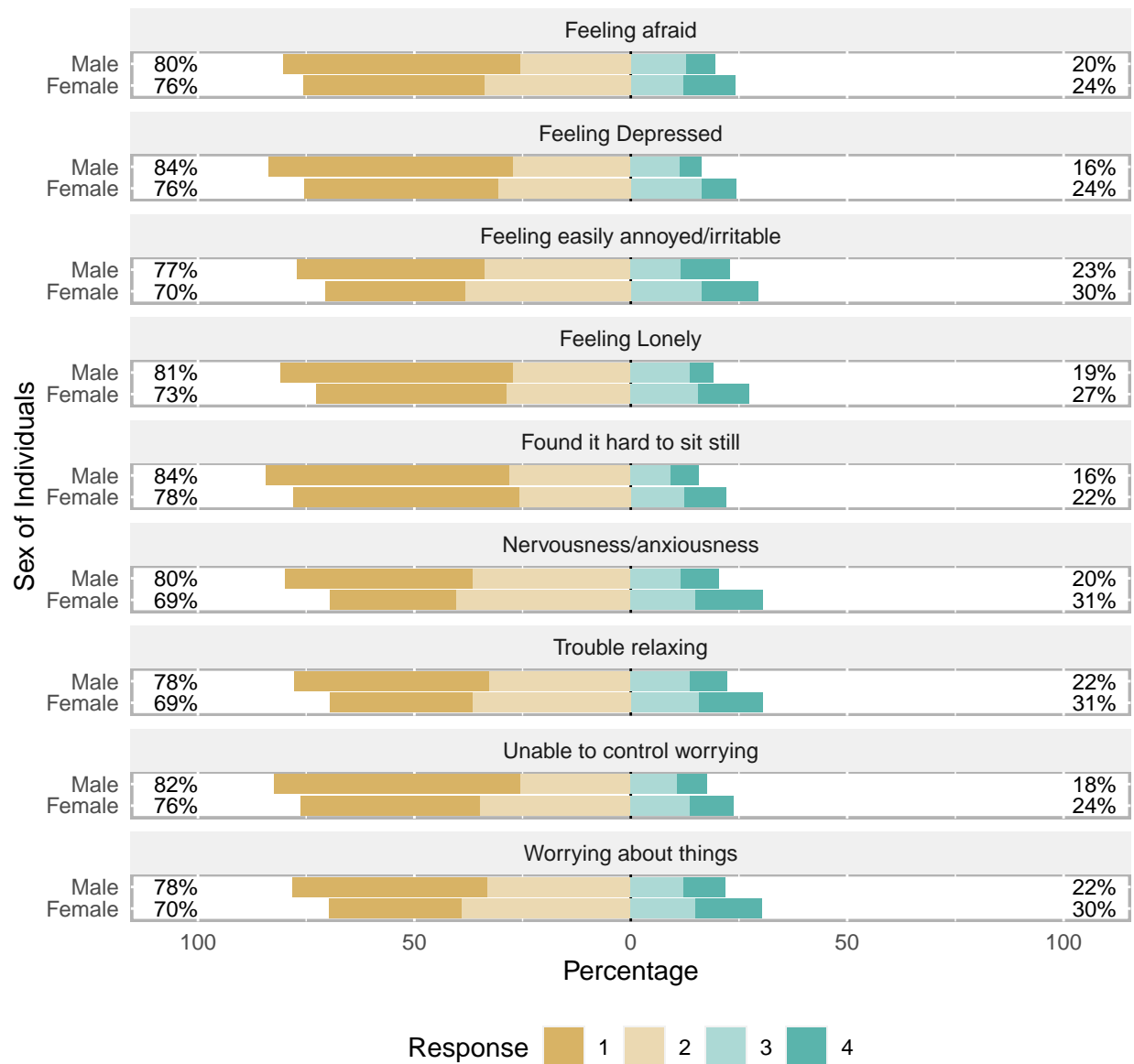

```
#select only the Likert scale questions
newCamh = camh[c(16:24, 29)]
#set these variables as factors
newCamh[1:10] <- lapply(newCamh[1:10], factor)
newCamh = newCamh[!newCamh$Ethnicity == "Prefer not to answer",]
newCamh <- droplevels(newCamh)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
```
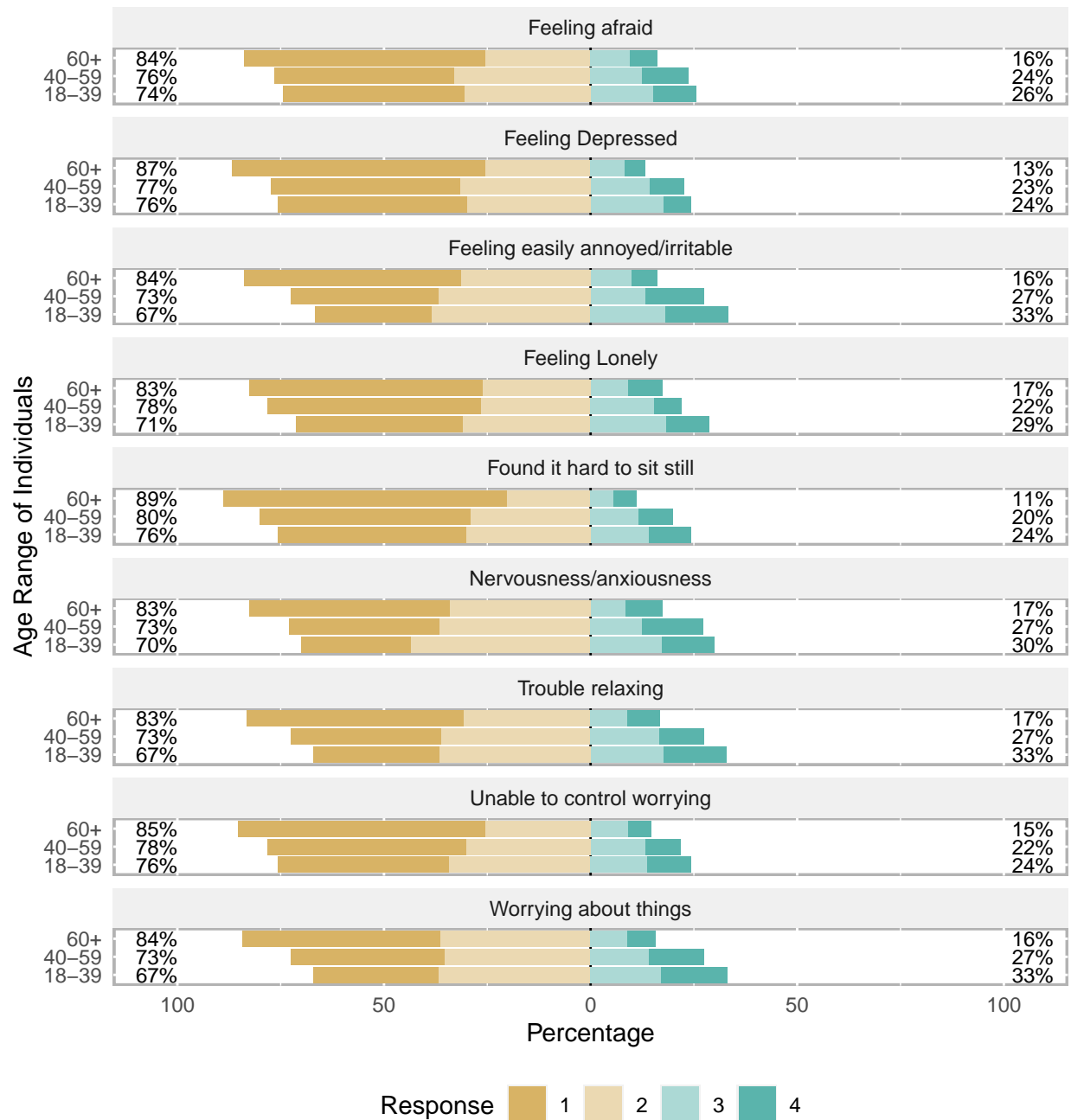
```
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh[-c(10)], grouping=factor(newCamh$Ethnicity))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```

How often has the following occurred to you?

Feeling afraid

Ethnicity	Left	Right
Other	68%	32%
Black	75%	25%
Asian	69%	31%
White	82%	18%

Feeling Depressed

Ethnicity	Left	Right
Other	68%	32%
Black	88%	12%
Asian	76%	24%
White	81%	19%

Feeling easily annoyed/irritable

Ethnicity	Left	Right
Other	67%	33%
Black	88%	12%
Asian	70%	30%
White	76%	24%

Feeling Lonely

Ethnicity	Left	Right
Other	72%	28%
Black	94%	6%
Asian	76%	24%
White	77%	23%

Found it hard to sit still

Ethnicity	Left	Right
Other	72%	28%
Black	75%	25%
Asian	81%	19%
White	82%	18%

Nervousness/anxiousness

Ethnicity	Left	Right
Other	65%	35%
Black	75%	25%
Asian	71%	29%
White	77%	23%

Trouble relaxing

Ethnicity	Left	Right
Other	68%	32%
Black	62%	38%
Asian	73%	27%
White	75%	25%

Unable to control worrying

Ethnicity	Left	Right
Other	77%	23%
Black	81%	19%
Asian	72%	28%
White	82%	18%

Worrying about things

Ethnicity	Left	Right
Other	65%	35%
Black	69%	31%
Asian	66%	34%
White	78%	22%

Ethnicity of Individuals

Percentage

Response   1   2   3   4

Below I have removed three individual's whose response as prefer not to answer

```r
#select only the Likert scale questions
newCamh = camh[c(16:24,26)]
#set these variables as factors
newCamh[1:10] <- lapply(newCamh[1:10], factor)
newCamh = newCamh[!newCamh$Question23 == "Prefer not to answer",]
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh[-c(10)], grouping=factor(newCamh$Question23))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```
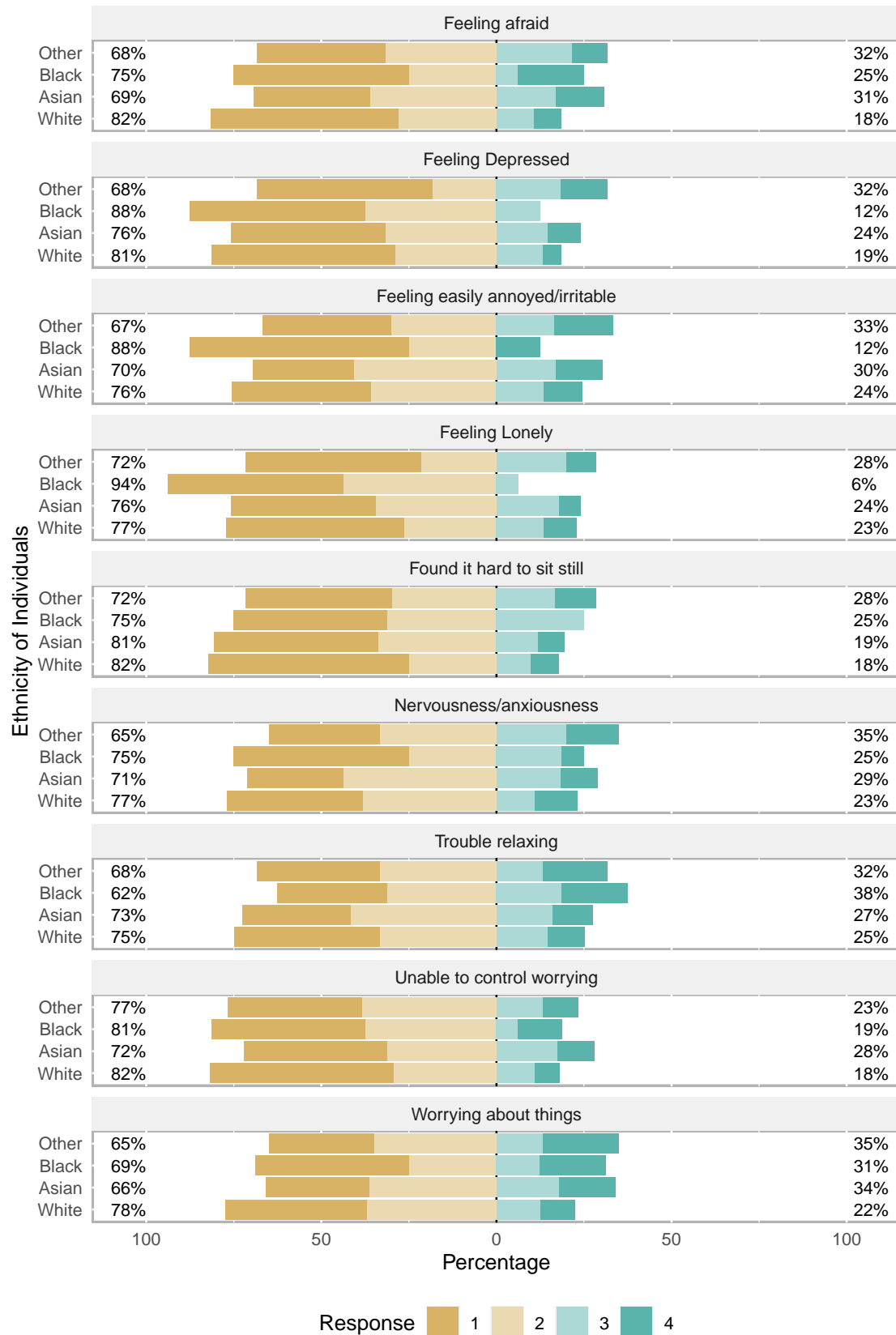
How often has the following occurred to you?

Figure showing diverging stacked bar charts by "Number of people in individual's household, including themselves" (y-axis, grouped by categories 4+, 3, 2, 1) across nine panels, with x-axis "Percentage". Response legend: 1, 2, 3, 4.

Feeling afraid

	Left %	Right %
4+	68%	32%
3	77%	23%
2	81%	19%
1	84%	16%

Feeling Depressed

	Left %	Right %
4+	70%	30%
3	76%	24%
2	85%	15%
1	80%	20%

Feeling easily annoyed/irritable

	Left %	Right %
4+	62%	38%
3	67%	33%
2	79%	21%
1	81%	19%

Feeling Lonely

	Left %	Right %
4+	72%	28%
3	76%	24%
2	85%	15%
1	66%	34%

Found it hard to sit still

	Left %	Right %
4+	78%	22%
3	81%	19%
2	81%	19%
1	83%	17%

Nervousness/anxiousness

	Left %	Right %
4+	67%	33%
3	74%	26%
2	77%	23%
1	79%	21%

Trouble relaxing

	Left %	Right %
4+	65%	35%
3	69%	31%
2	78%	22%
1	78%	22%

Unable to control worrying

	Left %	Right %
4+	73%	27%
3	80%	20%
2	81%	19%
1	82%	18%

Worrying about things

	Left %	Right %
4+	64%	36%
3	75%	25%
2	77%	23%
1	77%	23%

Response: 1 2 3 4

```r
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)
newCamh <- newCamh %>%
```

```r
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$LivingArea))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert,cex.lab=4) + ggtitle("How often has the following occurred to you?") + theme(plot.titl
```
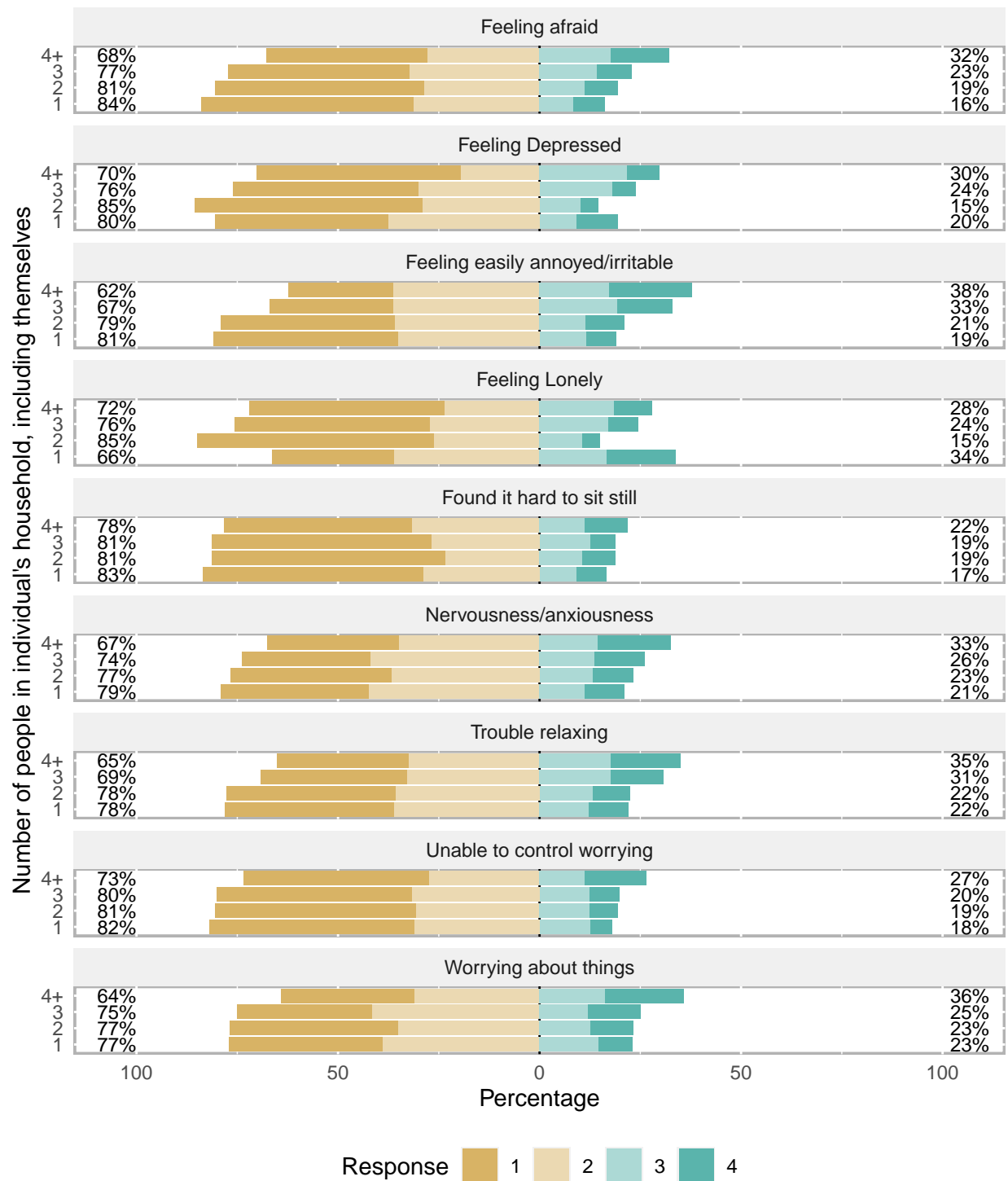
How often has the following occurred to you?


```r
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
```

```r
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$HouseholdIncome))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```
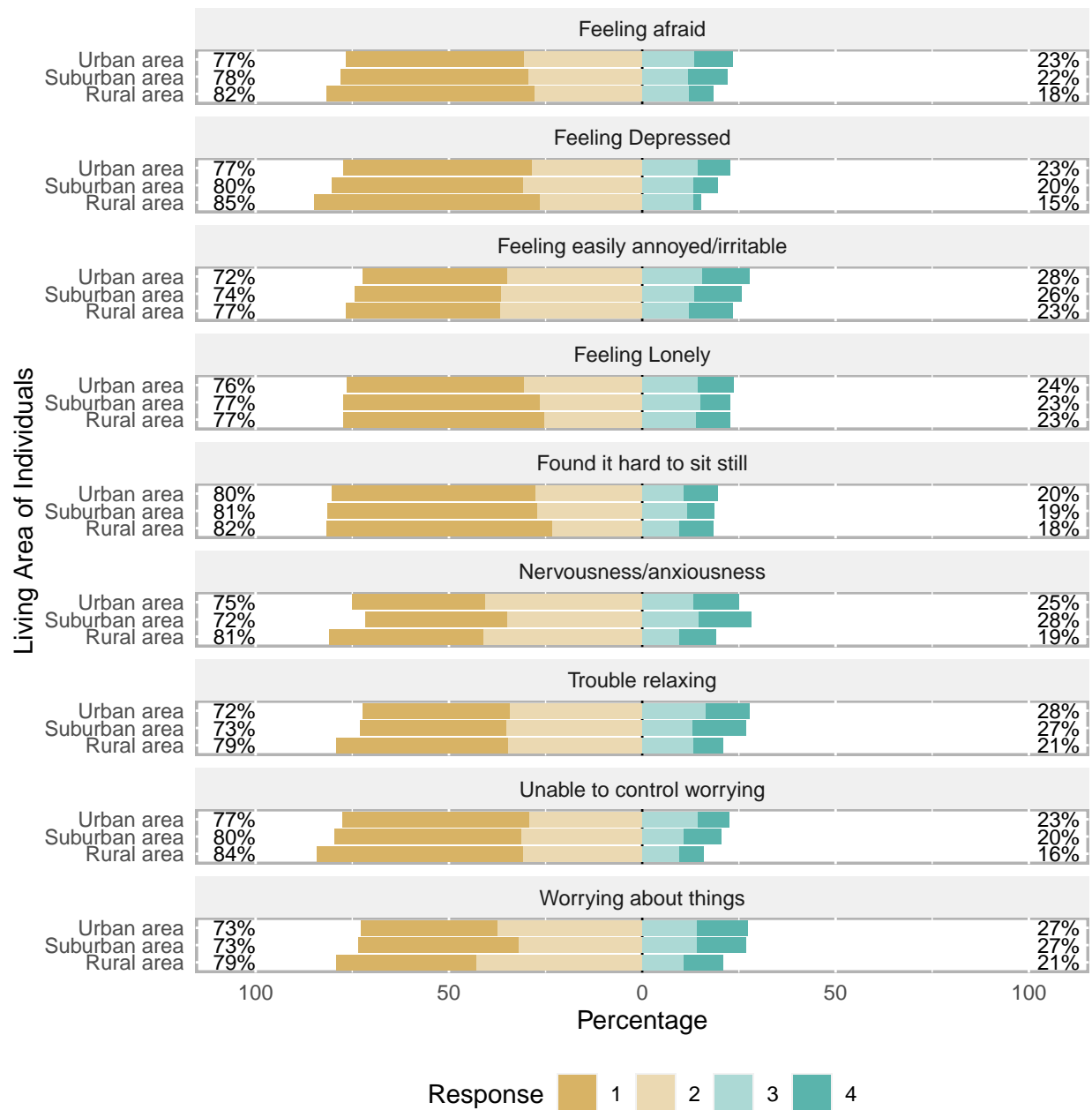
How often has the following occurred to you?

```r
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$MaritalStatus))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```

How often has the following occurred to you?

Feeling afraid

Marital Status	Left %	Right %
Prefer not to answer	85%	15%
Never Married/Widowed	77%	23%
Divorced/Seperated	83%	17%
Married/Living with Partner	77%	23%

Feeling Depressed

Marital Status	Left %	Right %
Prefer not to answer	92%	8%
Never Married/Widowed	77%	23%
Divorced/Seperated	78%	22%
Married/Living with Partner	81%	19%

Feeling easily annoyed/irritable

Marital Status	Left %	Right %
Prefer not to answer	69%	31%
Never Married/Widowed	74%	26%
Divorced/Seperated	76%	24%
Married/Living with Partner	73%	27%

Feeling Lonely

Marital Status	Left %	Right %
Prefer not to answer	77%	23%
Never Married/Widowed	66%	34%
Divorced/Seperated	65%	35%
Married/Living with Partner	84%	16%

Found it hard to sit still

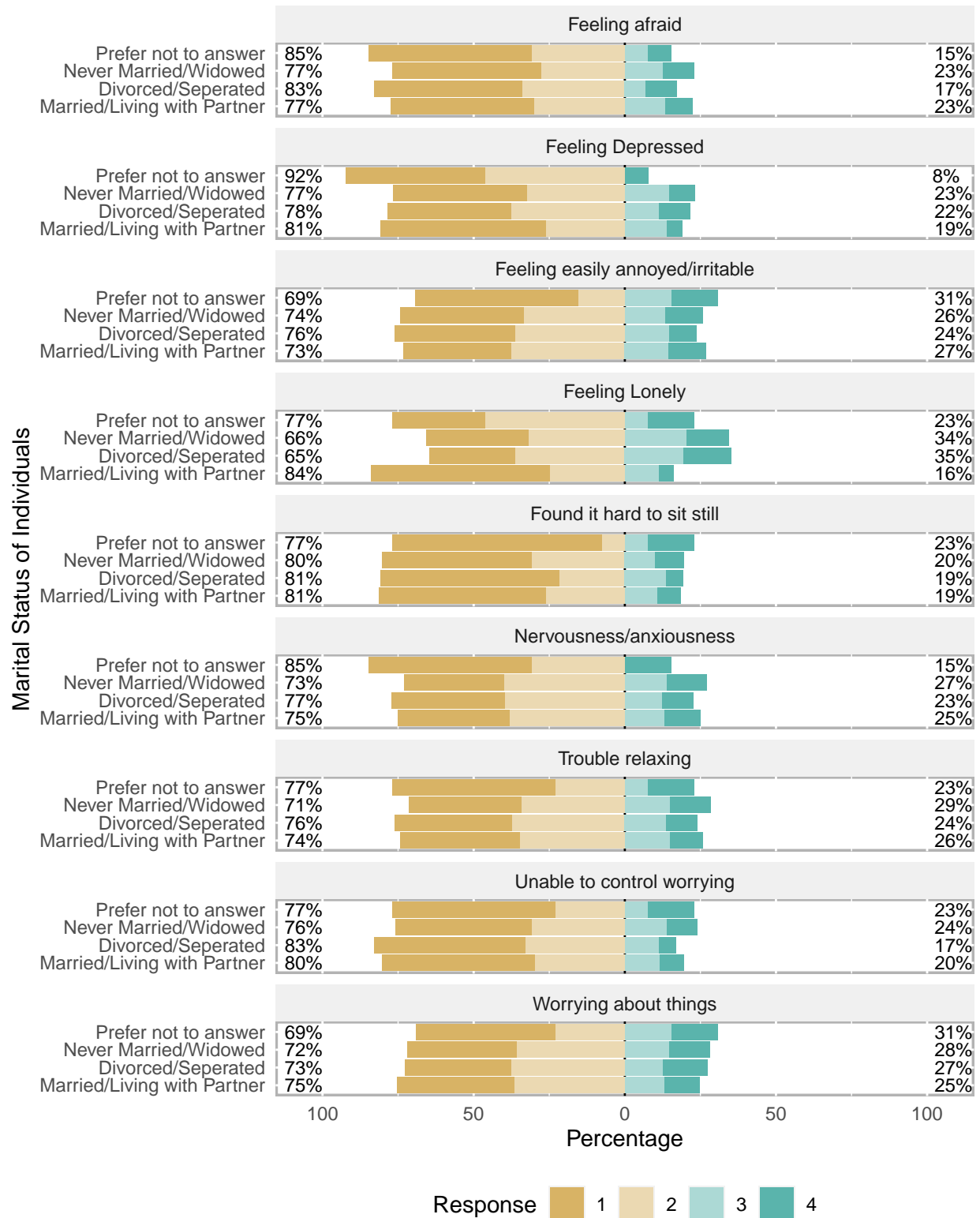| Marital Status	Left %	Right %
Prefer not to answer	77%	23%
Never Married/Widowed	80%	20%
Divorced/Seperated	81%	19%
Married/Living with Partner	81%	19%

Nervousness/anxiousness

Marital Status	Left %	Right %
Prefer not to answer	85%	15%
Never Married/Widowed	73%	27%
Divorced/Seperated	77%	23%
Married/Living with Partner	75%	25%

Trouble relaxing

Marital Status	Left %	Right %
Prefer not to answer	77%	23%
Never Married/Widowed	71%	29%
Divorced/Seperated	76%	24%
Married/Living with Partner	74%	26%

Unable to control worrying

Marital Status	Left %	Right %
Prefer not to answer	77%	23%
Never Married/Widowed	76%	24%
Divorced/Seperated	83%	17%
Married/Living with Partner	80%	20%

Worrying about things

Marital Status	Left %	Right %
Prefer not to answer	69%	31%
Never Married/Widowed	72%	28%
Divorced/Seperated	73%	27%
Married/Living with Partner	75%	25%

Percentage — Marital Status of Individuals

Response: 1 2 3 4

```r
#set this variablea as a factor variable
camh$ChangeInWorkingHours <- factor(camh$ChangeInWorkingHours, levels = c("Increased a lot", "Decreased
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$ChangeInWorkingHours))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = elemen
```
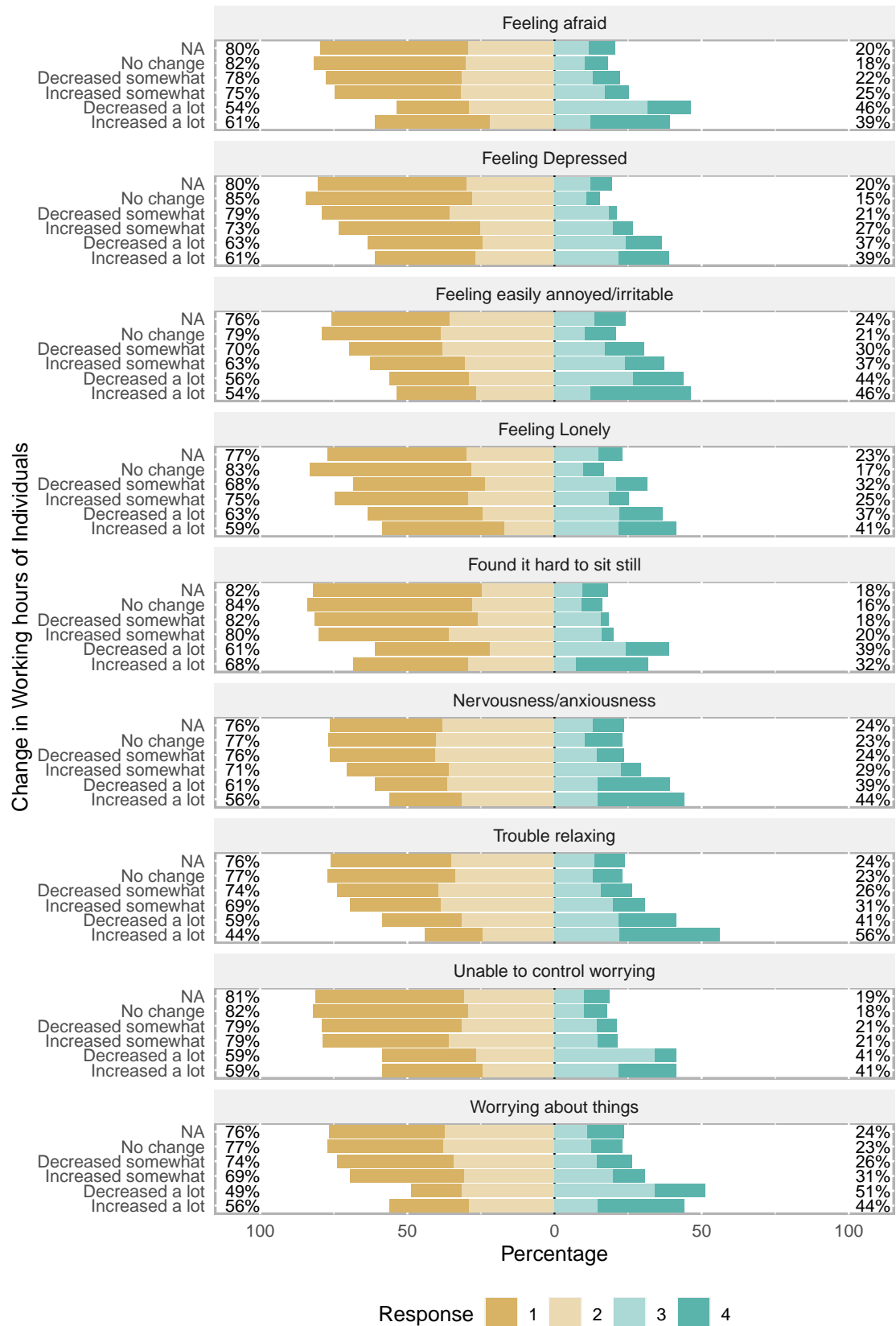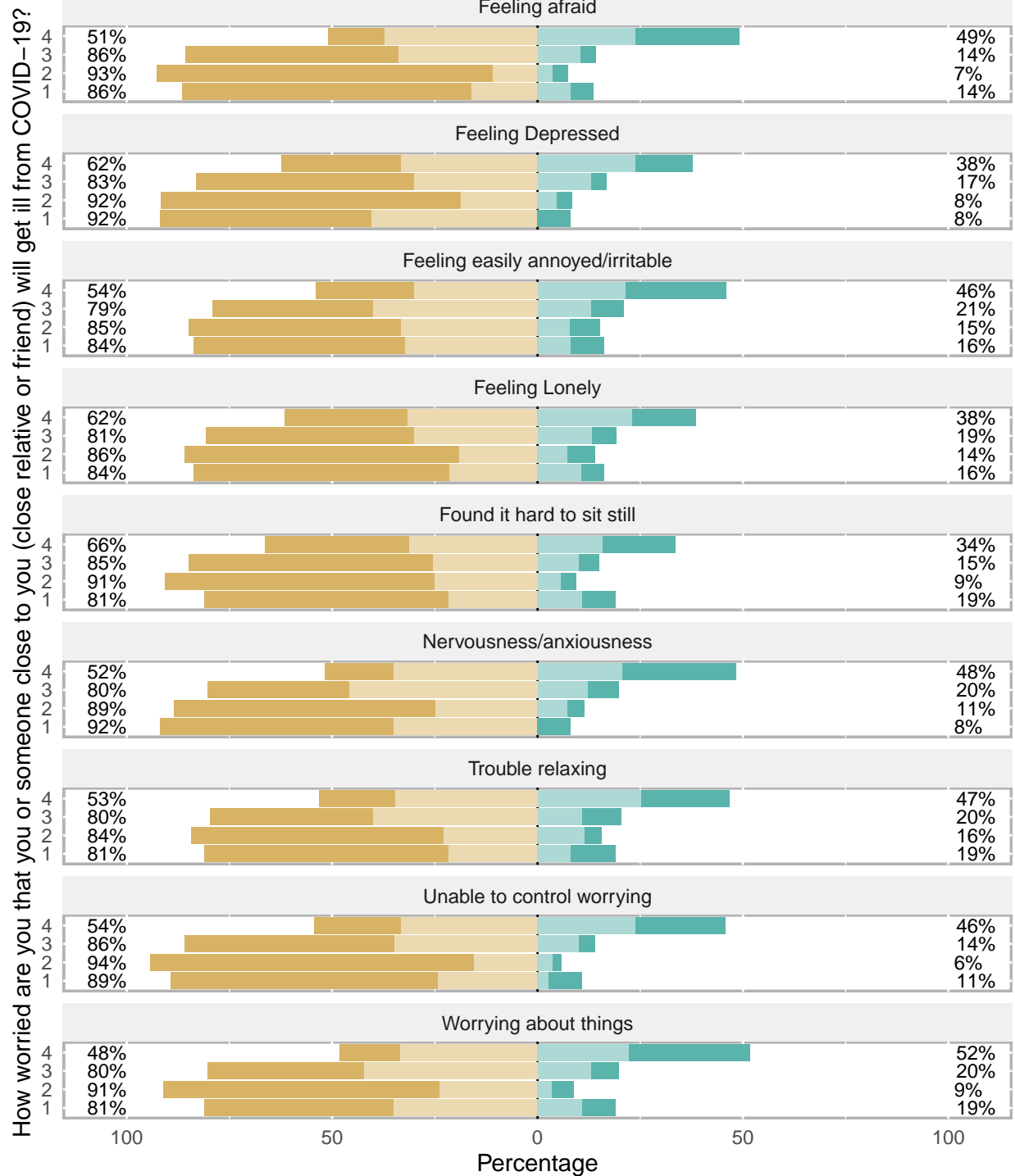
How often has the following occurred to you?

Feeling afraid

Change in Working hours	Left %	Right %
NA	80%	20%
No change	82%	18%
Decreased somewhat	78%	22%
Increased somewhat	75%	25%
Decreased a lot	54%	46%
Increased a lot	61%	39%

Feeling Depressed

Change in Working hours	Left %	Right %
NA	80%	20%
No change	85%	15%
Decreased somewhat	79%	21%
Increased somewhat	73%	27%
Decreased a lot	63%	37%
Increased a lot	61%	39%

Feeling easily annoyed/irritable

Change in Working hours	Left %	Right %
NA	76%	24%
No change	79%	21%
Decreased somewhat	70%	30%
Increased somewhat	63%	37%
Decreased a lot	56%	44%
Increased a lot	54%	46%

Feeling Lonely

Change in Working hours	Left %	Right %
NA	77%	23%
No change	83%	17%
Decreased somewhat	68%	32%
Increased somewhat	75%	25%
Decreased a lot	63%	37%
Increased a lot	59%	41%

Found it hard to sit still

Change in Working hours	Left %	Right %
NA	82%	18%
No change	84%	16%
Decreased somewhat	82%	18%
Increased somewhat	80%	20%
Decreased a lot	61%	39%
Increased a lot	68%	32%

Nervousness/anxiousness

Change in Working hours	Left %	Right %
NA	76%	24%
No change	77%	23%
Decreased somewhat	76%	24%
Increased somewhat	71%	29%
Decreased a lot	61%	39%
Increased a lot	56%	44%

Trouble relaxing

Change in Working hours	Left %	Right %
NA	76%	24%
No change	77%	23%
Decreased somewhat	74%	26%
Increased somewhat	69%	31%
Decreased a lot	59%	41%
Increased a lot	44%	56%

Unable to control worrying

Change in Working hours	Left %	Right %
NA	81%	19%
No change	82%	18%
Decreased somewhat	79%	21%
Increased somewhat	79%	21%
Decreased a lot	59%	41%
Increased a lot	59%	41%

Worrying about things

Change in Working hours	Left %	Right %
NA	76%	24%
No change	77%	23%
Decreased somewhat	74%	26%
Increased somewhat	69%	31%
Decreased a lot	49%	51%
Increased a lot	56%	44%

Change in Working hours of Individuals — Percentage

Response: 1 2 3 4

```r
#select only the Likert scale questions
newCamh = camh[16:24]
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh, grouping=factor(camh$Question7))
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```
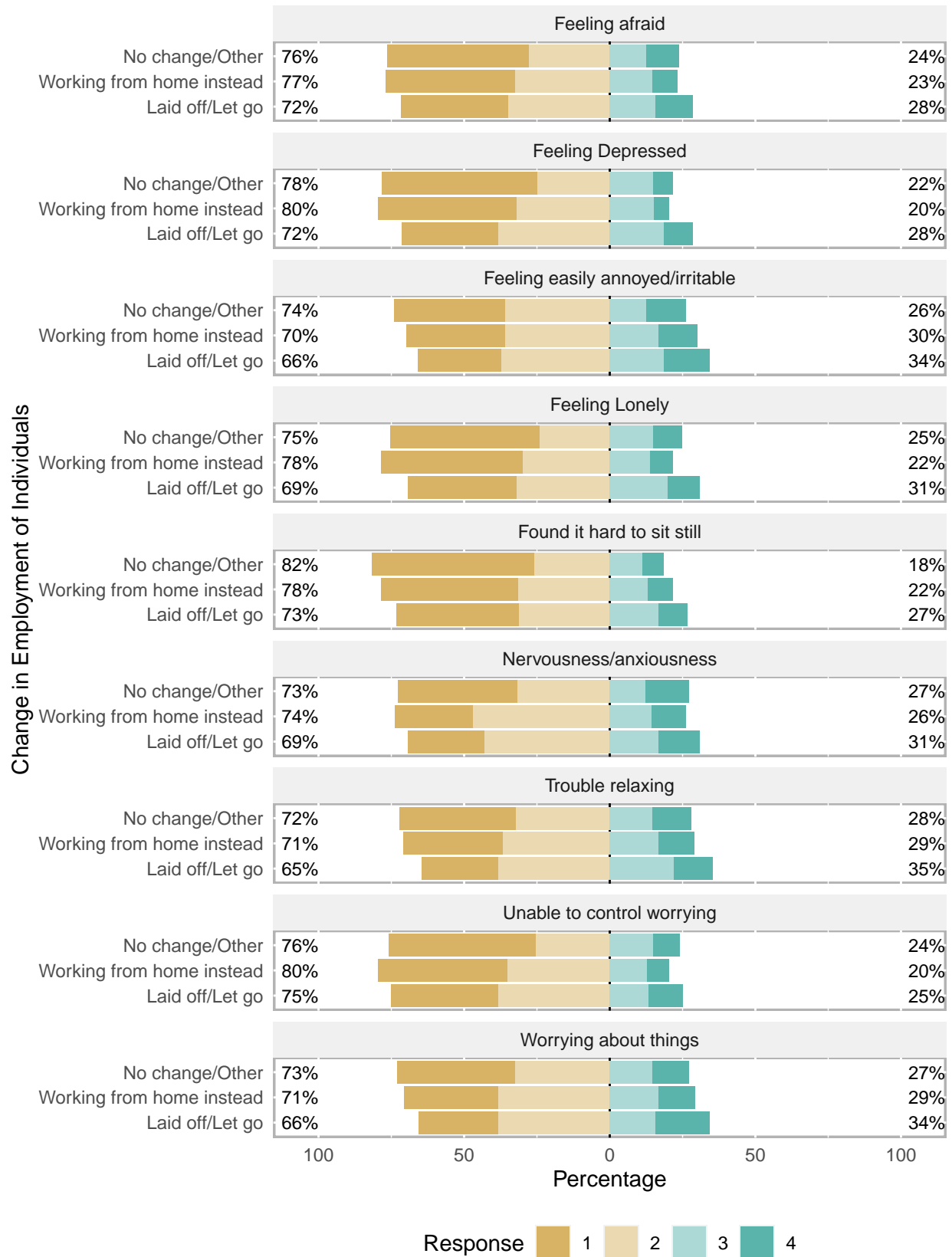
How often has the following occurred to you?

Feeling afraid

	Left	Right
4	51%	49%
3	86%	14%
2	93%	7%
1	86%	14%

Feeling Depressed

	Left	Right
4	62%	38%
3	83%	17%
2	92%	8%
1	92%	8%

Feeling easily annoyed/irritable

	Left	Right
4	54%	46%
3	79%	21%
2	85%	15%
1	84%	16%

Feeling Lonely

	Left	Right
4	62%	38%
3	81%	19%
2	86%	14%
1	84%	16%

Found it hard to sit still

	Left	Right
4	66%	34%
3	85%	15%
2	91%	9%
1	81%	19%

Nervousness/anxiousness

	Left	Right
4	52%	48%
3	80%	20%
2	89%	11%
1	92%	8%

Trouble relaxing

	Left	Right
4	53%	47%
3	80%	20%
2	84%	16%
1	81%	19%

Unable to control worrying

	Left	Right
4	54%	46%
3	86%	14%
2	94%	6%
1	89%	11%

Worrying about things

	Left	Right
4	48%	52%
3	80%	20%
2	91%	9%
1	81%	19%

How worried are you that you or someone close to you (close relative or friend) will get ill from COVID−19?

Percentage

Response: 1 2 3 4

```r
#select only the Likert scale questions
newCamh = camh[c(13,16:24)]
#remove the unemployed individuals
newCamh = newCamh[!newCamh$ChangeInEmployment == "Unemployed (No change)",]
newCamh <- droplevels(newCamh)
#set these variables as factors
newCamh[1:9] <- lapply(newCamh[1:9], factor)
newCamh <- newCamh %>%
  rename(# rename the questions to show what symptom they ae asking about
    "Nervousness/anxiousness" = "Question8.1",
    "Unable to control worrying" = "Question8.2",
    "Worrying about things" = "Question8.3",
    "Trouble relaxing" = "Question8.4",
    "Found it hard to sit still" = "Question8.5",
    "Feeling easily annoyed/irritable" = "Question8.6",
    "Feeling afraid" = "Question8.7",
    "Feeling Depressed" = "Question20.1",
    "Feeling Lonely" = "Question20.2"
  )
#set as a likert using likert package and group by the factor
camh_likert <- likert(newCamh[-c(1)], grouping=factor(newCamh$ChangeInEmployment))
```

```
## Warning in likert(newCamh[-c(1)], grouping =
## factor(newCamh$ChangeInEmployment)): items parameter contains non-factors. Will
## convert to factors
```

```r
camh_likertgs <- likert(summary = camh_likert$results, grouping = camh_likert$results[,1])
#create the plot and add the titles
plot(camh_likert) + ggtitle("How often has the following occurred to you?") + theme(plot.title = element
```

How often has the following occurred to you?

Change in Employment of Individuals

Feeling afraid
- No change/Other: 76% | 24%
- Working from home instead: 77% | 23%
- Laid off/Let go: 72% | 28%

Feeling Depressed
- No change/Other: 78% | 22%
- Working from home instead: 80% | 20%
- Laid off/Let go: 72% | 28%

Feeling easily annoyed/irritable
- No change/Other: 74% | 26%
- Working from home instead: 70% | 30%
- Laid off/Let go: 66% | 34%

Feeling Lonely
- No change/Other: 75% | 25%
- Working from home instead: 78% | 22%
- Laid off/Let go: 69% | 31%

Found it hard to sit still
- No change/Other: 82% | 18%
- Working from home instead: 78% | 22%
- Laid off/Let go: 73% | 27%

Nervousness/anxiousness
- No change/Other: 73% | 27%
- Working from home instead: 74% | 26%
- Laid off/Let go: 69% | 31%

Trouble relaxing
- No change/Other: 72% | 28%
- Working from home instead: 71% | 29%
- Laid off/Let go: 65% | 35%

Unable to control worrying
- No change/Other: 76% | 24%
- Working from home instead: 80% | 20%
- Laid off/Let go: 75% | 25%

Worrying about things
- No change/Other: 73% | 27%
- Working from home instead: 71% | 29%
- Laid off/Let go: 66% | 34%

Percentage

Response: 1, 2, 3, 4

Modelling

I will sum up the scores on question 8 and 20 as `Total Sum` to be the dependant variable.

```
# histogram of total score
hist(camh$TotalSum, xlab = "Sum of scores", main = "")
```
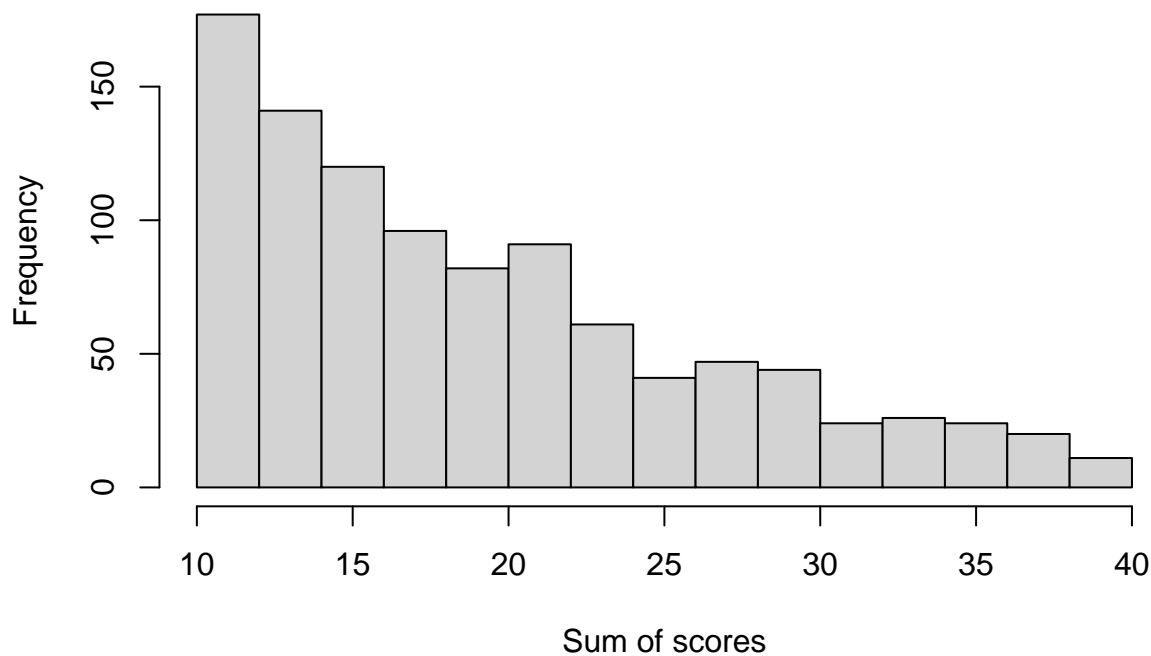

The histogram above shows that the `TotalSum` is not normally distributed, however due to the large sample size and central limit theorem we can still use parametric testing.

Question23 (How many people including yourself live in your household) and LivingStatus (Alone or with others) are going to cause multicollinearity, so I will only include question23 in the model. I will use VIF to check for any more multicollinearity in my model.

As the scale-location plot originally showed an increasing red line, I took a log transformation of the response variable to help improve the assumption of homoscedasticity.

```
# set the reference category as no change in working hours
camh$ChangeInWorkingHours <- relevel(camh$ChangeInWorkingHours,
    ref = "No change")
# create the linear model
lm1 <- lm(log(TotalSum) ~ Sex + Q4Sum + AgeRange + Ethnicity +
    Question23 + LivingArea + HouseholdIncome + MaritalStatus +
    ChangeInWorkingHours, data = camh)
# plot the diagnostic plots
plot(lm1)
```
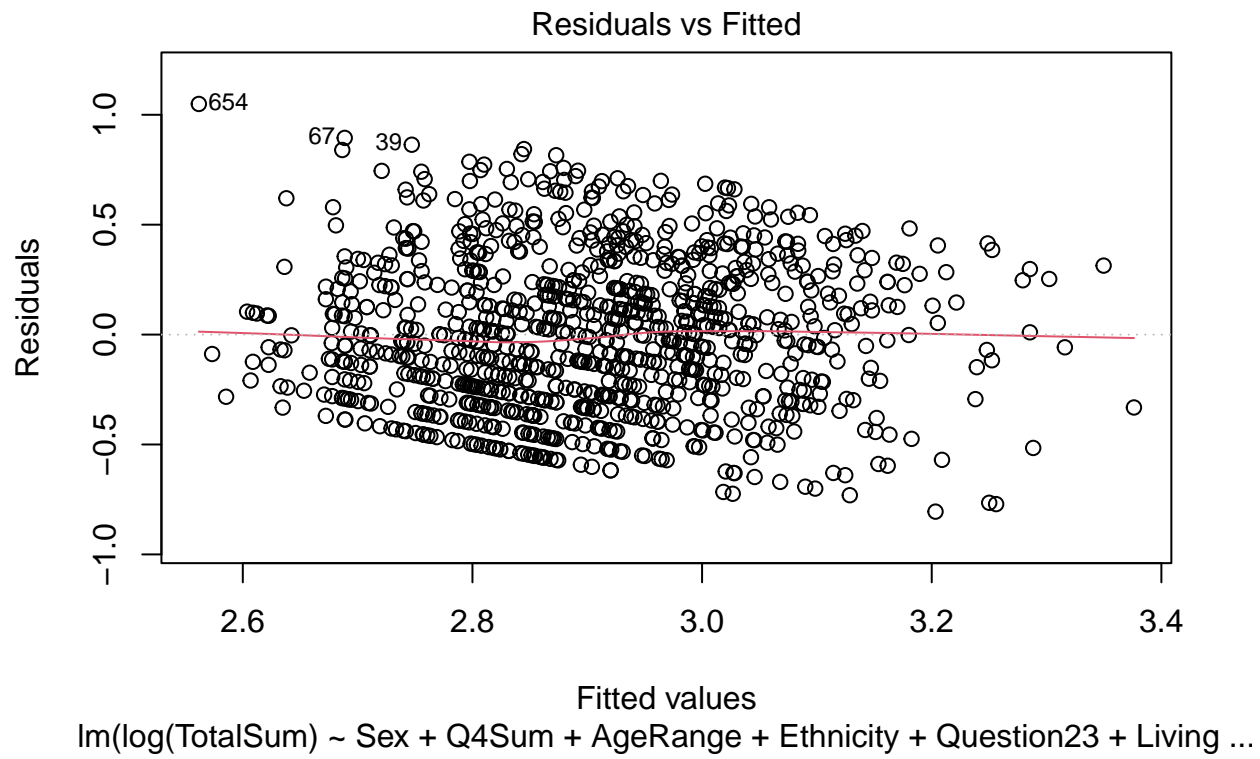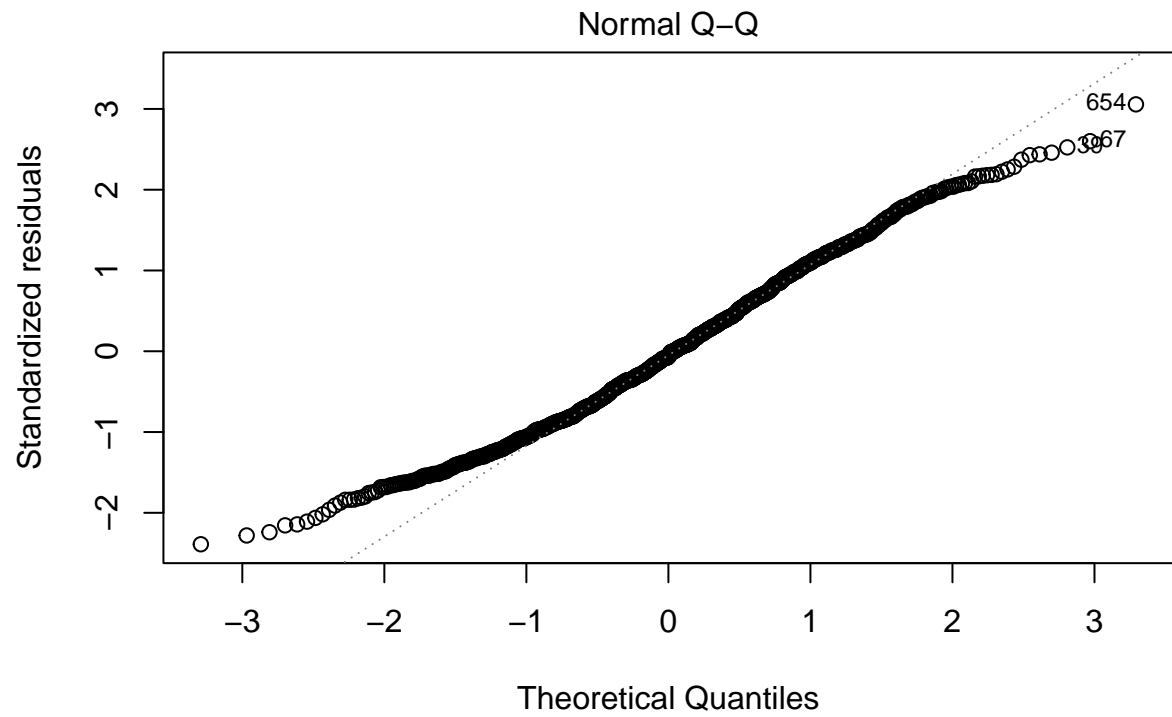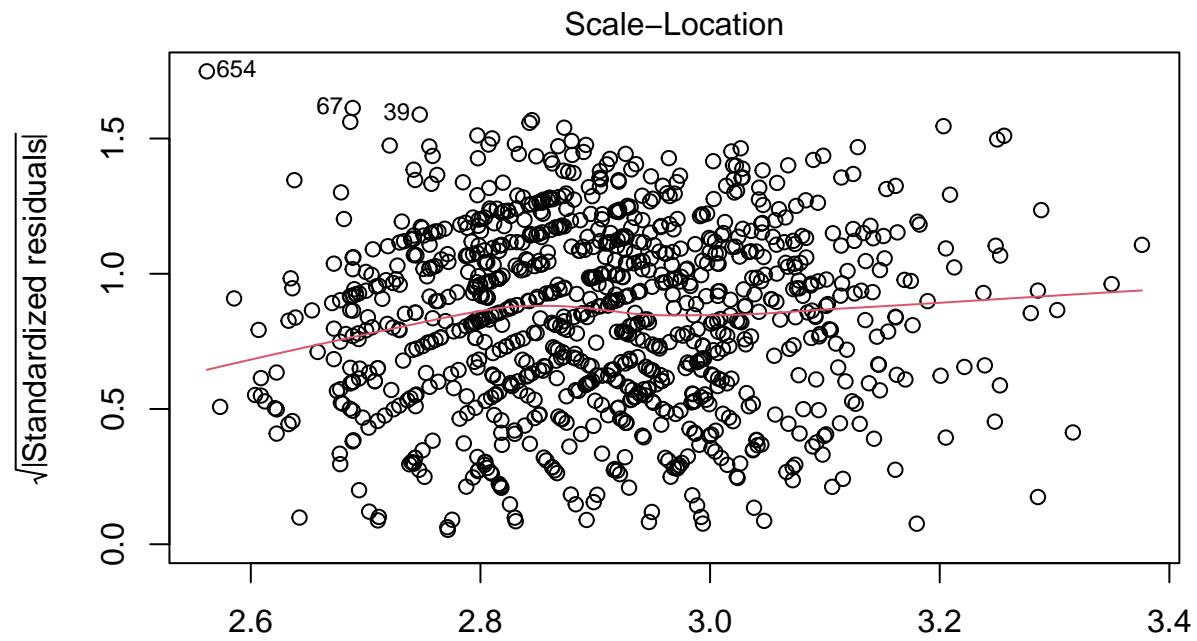
Residuals vs Fitted

Fitted values
lm(log(TotalSum) ~ Sex + Q4Sum + AgeRange + Ethnicity + Question23 + Living ...

Normal Q–Q

Theoretical Quantiles
lm(log(TotalSum) ~ Sex + Q4Sum + AgeRange + Ethnicity + Question23 + Living ...

Scale−Location

Fitted values
lm(log(TotalSum) ~ Sex + Q4Sum + AgeRange + Ethnicity + Question23 + Living ...

Residuals vs Leverage

lm(log(TotalSum) ~ Sex + Q4Sum + AgeRange + Ethnicity + Question23 + Living ...
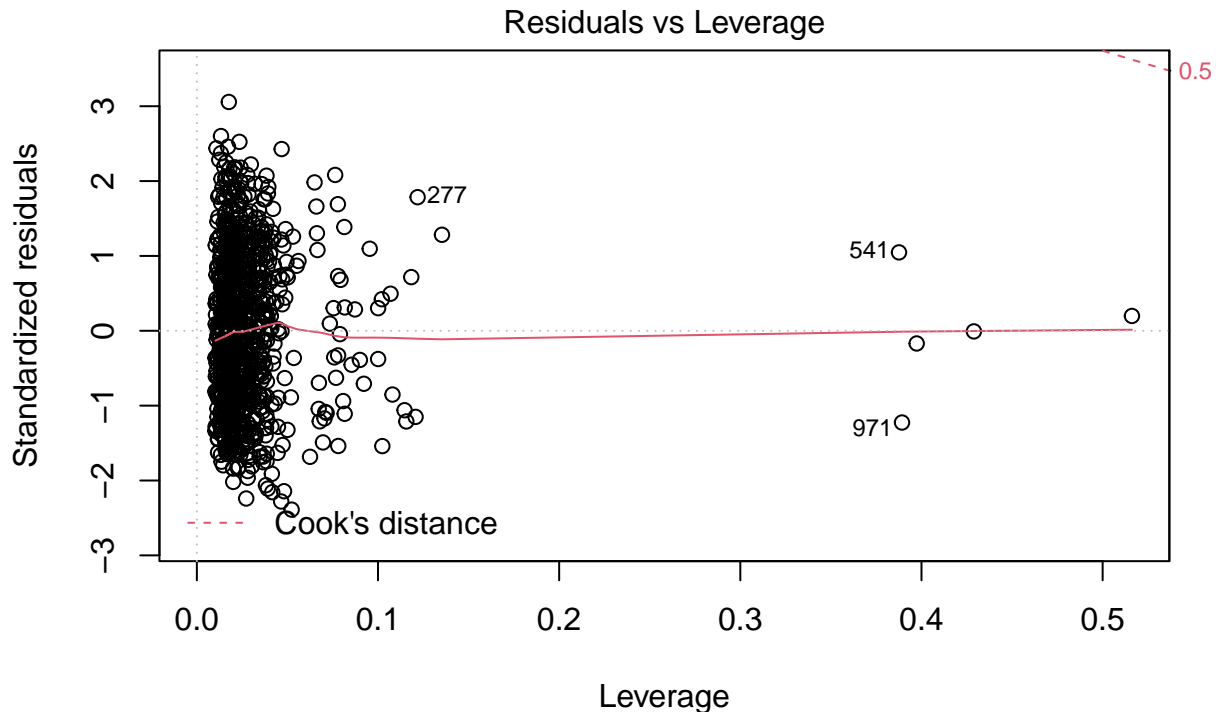
```r
# the full model
lm2 <- lm(log(TotalSum) ~ (Sex + Q4Sum + AgeRange + Ethnicity +
    Question23 + LivingArea + HouseholdIncome + MaritalStatus +
    ChangeInWorkingHours)^2, data = camh)
# perform anova
anova(lm2)
```

```
## Analysis of Variance Table
##
## Response: log(TotalSum)
##                       Df Sum Sq Mean Sq F value    Pr(>F)
## Sex                    2  3.361  1.6803 14.4863 6.766e-07 ***
## Q4Sum                  1  3.380  3.3801 29.1413 9.107e-08 ***
## AgeRange               2  6.676  3.3379 28.7776 9.339e-13 ***
## Ethnicity              4  1.257  0.3143  2.7101 0.0292091 *
## Question23             4  0.342  0.0855  0.7367 0.5670696
## LivingArea             2  0.085  0.0423  0.3648 0.6944516
## HouseholdIncome        4  1.238  0.3095  2.6684 0.0313059 *
## MaritalStatus          3  0.072  0.0241  0.2077 0.8911206
## ChangeInWorkingHours   5  2.441  0.4881  4.2083 0.0008889 ***
## Sex:Q4Sum              2  0.030  0.0150  0.1294 0.8786782
## Sex:AgeRange           3  0.315  0.1050  0.9056 0.4379067
## Sex:Ethnicity          4  0.233  0.0583  0.5023 0.7340915
## Sex:Question23         3  0.077  0.0258  0.2223 0.8809235
## Sex:LivingArea         2  0.293  0.1463  1.2611 0.2839658
## Sex:HouseholdIncome    4  0.058  0.0145  0.1247 0.9735493
```

```
## Sex:MaritalStatus                        3   0.266  0.0886  0.7639 0.5145075
## Sex:ChangeInWorkingHours                  5   0.594  0.1187  1.0234 0.4026205
## Q4Sum:AgeRange                            2   0.044  0.0220  0.1901 0.8269351
## Q4Sum:Ethnicity                           4   0.718  0.1796  1.5484 0.1863833
## Q4Sum:Question23                          4   0.276  0.0689  0.5942 0.6669236
## Q4Sum:LivingArea                          2   0.308  0.1541  1.3284 0.2655491
## Q4Sum:HouseholdIncome                     4   0.340  0.0849  0.7321 0.5701762
## Q4Sum:MaritalStatus                       3   0.268  0.0894  0.7711 0.5103693
## Q4Sum:ChangeInWorkingHours                5   0.365  0.0729  0.6287 0.6779475
## AgeRange:Ethnicity                        8   1.006  0.1258  1.0842 0.3720117
## AgeRange:Question23                       6   0.280  0.0466  0.4020 0.8779176
## AgeRange:LivingArea                       4   0.342  0.0855  0.7370 0.5669065
## AgeRange:HouseholdIncome                  8   0.974  0.1217  1.0497 0.3969278
## AgeRange:MaritalStatus                    6   0.858  0.1431  1.2334 0.2868279
## AgeRange:ChangeInWorkingHours            10   0.943  0.0943  0.8133 0.6159216
## Ethnicity:Question23                     11   0.824  0.0749  0.6455 0.7901445
## Ethnicity:LivingArea                      8   1.625  0.2031  1.7509 0.0835739 .
## Ethnicity:HouseholdIncome                13   3.027  0.2328  2.0074 0.0179032 *
## Ethnicity:MaritalStatus                  10   0.752  0.0752  0.6485 0.7723637
## Ethnicity:ChangeInWorkingHours           13   1.909  0.1468  1.2657 0.2285086
## Question23:LivingArea                     6   0.308  0.0513  0.4424 0.8505386
## Question23:HouseholdIncome               12   1.497  0.1248  1.0756 0.3777128
## Question23:MaritalStatus                  7   1.357  0.1939  1.6719 0.1126696
## Question23:ChangeInWorkingHours          15   2.775  0.1850  1.5949 0.0694955 .
## LivingArea:HouseholdIncome                8   1.087  0.1359  1.1720 0.3132928
## LivingArea:MaritalStatus                  4   0.920  0.2301  1.9835 0.0952402 .
## LivingArea:ChangeInWorkingHours          10   1.291  0.1291  1.1133 0.3490711
## HouseholdIncome:MaritalStatus             8   2.275  0.2844  2.4517 0.0127049 *
## HouseholdIncome:ChangeInWorkingHours     20   2.445  0.1222  1.0538 0.3952321
## MaritalStatus:ChangeInWorkingHours       10   1.818  0.1818  1.5676 0.1118230
## Residuals                               730  84.672  0.1160
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The anova shows us that there is a interaction between household income and marital status. This makes sense as married households have more people, hence more overall income.

Most variables have VIF close to one, except `Question23` and `MaritalStatus`. This makes sense as married/living with partner individuals are likely to live with more people (their spouse/children). However the VIF is still lower than 5 so I will keep both these parameters in my model.

Testing at a 5% significance level we can see that Sex, Q4factor, AgeRange, Ethnicity, Household income and Question6.2 (Change in working hours) all have significantly different scores depending on these variables.

Let us use a t-test to compare sex. I will remove the individuals who recorded their sex as Other as there are very few.

```
# performing t-test of sex in model
newCamh <- camh
# remove the other values from the sex column
newCamh = newCamh[!newCamh$Sex == "Other", ]
newCamh <- droplevels(newCamh)
# perform t-test
t.test(TotalSum ~ Sex, data = newCamh)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  TotalSum by Sex
## t = 4.6653, df = 990.21, p-value = 3.504e-06
## alternative hypothesis: true difference in means between group Female and group Male is not equal to
## 95 percent confidence interval:
##   1.261972 3.094359
## sample estimates:
## mean in group Female    mean in group Male
##              20.63253              18.45437
```

We have a very small p value, meaning that there is a significant difference in sexes. Specifically, men have lower scores compared to women. As men have an average score of 18.5 whereas women have an average score of 20.6.

```
# perform backward stepwise model selection
lm.min <- lm(log(TotalSum) ~ 1, data = camh)
lm.step <- step(lm1, direction = "backward", trace = 0)  # Set trace=0 to turn off printed output
```

```
# get summary of the linear model
summary(lm.step)
```

```
##
## Call:
## lm(formula = log(TotalSum) ~ Sex + Q4Sum + AgeRange + Ethnicity +
##     ChangeInWorkingHours, data = camh)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.79007 -0.26748 -0.02107  0.24379  1.00131
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                         2.89823    0.02995  96.780  < 2e-16 ***
## SexMale                            -0.10465    0.02241  -4.669 3.44e-06 ***
## SexOther                            0.20534    0.20284   1.012 0.311637
## Q4Sum                               0.05444    0.01177   4.625 4.24e-06 ***
## AgeRange40-59                      -0.04395    0.02658  -1.654 0.098522 .
## AgeRange60+                        -0.18398    0.02933  -6.272 5.33e-10 ***
## EthnicityAsian                      0.08382    0.02872   2.918 0.003603 **
## EthnicityBlack                     -0.06415    0.08787  -0.730 0.465545
## EthnicityOther                      0.04956    0.04703   1.054 0.292239
## EthnicityPrefer not to answer       0.02773    0.07993   0.347 0.728691
## ChangeInWorkingHoursIncreased a lot    0.22267    0.05773   3.857 0.000122 ***
## ChangeInWorkingHoursDecreased a lot    0.18573    0.05798   3.203 0.001402 **
## ChangeInWorkingHoursIncreased somewhat 0.05851    0.04477   1.307 0.191595
## ChangeInWorkingHoursDecreased somewhat 0.06857    0.04445   1.543 0.123222
## ChangeInWorkingHoursNA              0.07469    0.02703   2.763 0.005825 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3459 on 990 degrees of freedom
## Multiple R-squared:  0.1291, Adjusted R-squared:  0.1168
## F-statistic: 10.48 on 14 and 990 DF,  p-value: < 2.2e-16
```

```r
# create table of coeffiecients
table <- data.frame(coef(lm.step), confint(lm.step), summary(lm.step)$coefficients[,
    4], exp(coef(lm.step)), exp(confint(lm.step)))
table %>%
    round(3) %>%
    kable(col.names = c("coefficient", "2.5%", "97.5%", "p-value",
        "exp(coefficient)", "exp(2.5%)", "exp(97.5%)")) %>%
    kable_classic(latex_options = "HOLD_position", full_width = F) %>%
    add_header_above(c(` ` = 2, `95% Confidence Interval` = 2,
        ` ` = 2, `95% Confidence Interval` = 2))  # %>% save_kable(file = 'CAMHcoefficients.png', zoom
```

		95% Confidence Interval				95% Co
	coefficient	2.5%	97.5%	p-value	exp(coefficient)	exp(2.5
(Intercept)	2.898	2.839	2.957	0.000	18.142	17.1
SexMale	-0.105	-0.149	-0.061	0.000	0.901	0.8
SexOther	0.205	-0.193	0.603	0.312	1.228	0.8
Q4Sum	0.054	0.031	0.078	0.000	1.056	1.0
AgeRange40-59	-0.044	-0.096	0.008	0.099	0.957	0.9
AgeRange60+	-0.184	-0.242	-0.126	0.000	0.832	0.7
EthnicityAsian	0.084	0.027	0.140	0.004	1.087	1.0
EthnicityBlack	-0.064	-0.237	0.108	0.466	0.938	0.7
EthnicityOther	0.050	-0.043	0.142	0.292	1.051	0.9
EthnicityPrefer not to answer	0.028	-0.129	0.185	0.729	1.028	0.8
ChangeInWorkingHoursIncreased a lot	0.223	0.109	0.336	0.000	1.249	1.1
ChangeInWorkingHoursDecreased a lot	0.186	0.072	0.300	0.001	1.204	1.0
ChangeInWorkingHoursIncreased somewhat	0.059	-0.029	0.146	0.192	1.060	0.9
ChangeInWorkingHoursDecreased somewhat	0.069	-0.019	0.156	0.123	1.071	0.9
ChangeInWorkingHoursNA	0.075	0.022	0.128	0.006	1.078	1.0

```r
# diagnostic plots
par(mfrow = c(2, 2))
plot(lm.step, 1)
plot(lm.step, 2)
plot(lm.step, 3)
plot(lm.step, 4)
```

```r
par(mfrow = c(1, 1))
```
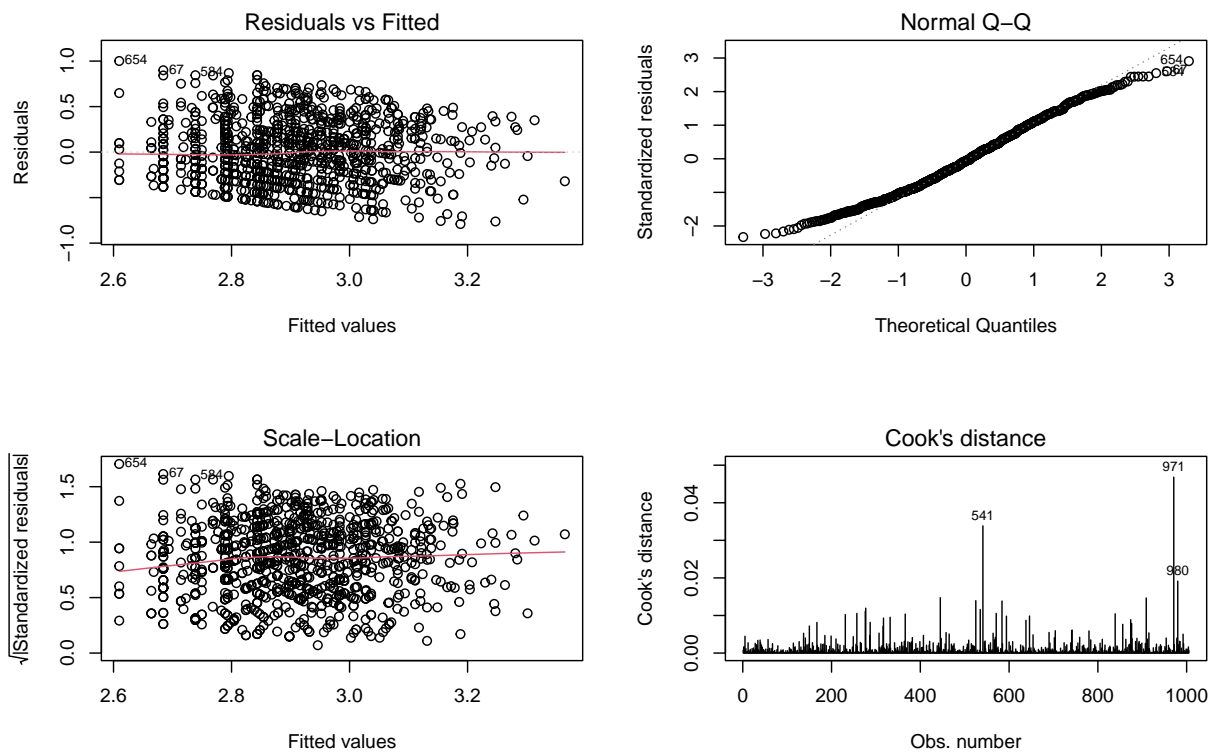
```r
# cross table of change in employment and change in working
# hours
table1 <- table(camh$ChangeInEmployment, camh$ChangeInWorkingHours)
kable(table1) %>%
    kable_classic(latex_options = "HOLD_position", full_width = F) %>%
    add_header_above(c(`Change in Employment` = 1, `Change in Working Hours` = 6))  #%>% save_kable(fil
```

Change in Employment	Change in Working Hours				
	No change	Increased a lot	Decreased a lot	Increased somewhat	Decreased somewhat
Laid off/Let go	0	0	0	0	(
Working from home instead	150	25	15	33	32
Unemployed (No change)	0	0	0	0	(
No change/Other	167	16	26	42	44

TABLES CREATED FOR FINAL REPORT:

Sensitivity to findings table from Bayes Net:

```r
sensitivity <- round(matrix(c(62.63, 2.623, 1.887, 1.361, 1.266,
    1.204, 0.8914, 0.6042, 0.3158, 0.2285, 0.09338, 100, 4.19,
    3.01, 2.17, 2.02, 1.92, 1.42, 0.965, 0.504, 0.365, 0.149,
    2.31838, 0.04527, 0.03479, 0.02283, 0.02119, 0.01568, 0.01754,
    0.01811, 0.00997, 0.00611, 0.00708, 100, 1.95, 1.5, 0.985,
```

```
        0.914, 0.677, 0.756, 0.781, 0.43, 0.264, 0.305, 0.6386338,
        0.0026238, 0.001738, 0.0014021, 0.0019867, 0.0008925, 0.0010116,
        0.0011042, 0.0005838, 0.0003579, 0.0004271), nrow = 11, dimnames = list(c("Total Score",
        "Age Range", "Change In Employment Status", "Count of Interactions with COVID-19",
        "Change in Working Hours", "Sex", "Number of People in Household",
        "Ethnicity", "Household Income", "Marital Status", "Living Area"),
        c("Variance Reduction", "Percent", "Mutual Info", "Percent",
            "Variance of Beliefs"))), 4)

kable(sensitivity, align = "r") %>%
    kable_classic(latex_options = "HOLD_position", full_width = F) %>%
    add_header_above(c(Node = 1, ` ` = 5))  #%>% save_kable(file = 'SensitivityFindings.png', zoom = 1)
```

Node		Variance Reduction	Percent	Mutual Info	Percent	Variance of Beliefs
Total Score		62.6300	100.000	2.3184	100.000	0.6386
Age Range		2.6230	4.190	0.0453	1.950	0.0026
Change In Employment Status		1.8870	3.010	0.0348	1.500	0.0017
Count of Interactions with COVID-19		1.3610	2.170	0.0228	0.985	0.0014
Change in Working Hours		1.2660	2.020	0.0212	0.914	0.0020
Sex		1.2040	1.920	0.0157	0.677	0.0009
Number of People in Household		0.8914	1.420	0.0175	0.756	0.0010
Ethnicity		0.6042	0.965	0.0181	0.781	0.0011
Household Income		0.3158	0.504	0.0100	0.430	0.0006
Marital Status		0.2285	0.365	0.0061	0.264	0.0004
Living Area		0.0934	0.149	0.0071	0.305	0.0004

Confusion Matrix from Bayesian Network Testing

```
sensitivity <- round(matrix(c(64, 61, 42, 26, 21, 44, 43, 24,
    24, 20, 32, 44, 37, 36, 35, 23, 34, 53, 53, 38, 28, 44, 63,
    51, 62), nrow = 5, dimnames = list(c("10-13", "13-16", "16-20",
    "20-26", "26-40"), c("10-13", "13-16", "16-20", "20-26",
    "26-40"))), 4)

kable(sensitivity, align = "r") %>%
    kable_classic(latex_options = "HOLD_position", full_width = F) %>%
    add_header_above(c(Actual = 1, Predicted = 5))  # %>% save_kable(file = 'ConfusionMatrix.png', zoom
```

Actual	Predicted				
	10-13	13-16	16-20	20-26	26-40
10-13	64	44	32	23	28
13-16	61	43	44	34	44
16-20	42	24	37	53	63
20-26	26	24	36	53	51
26-40	21	20	35	38	62

FOR APPENDIX

```r
camhappendix <- matrix(c("ID", "Province", "ProvinceCode", "AgeRange",
    "Sex", "Question4.1", "Question4.2", "Question4.3", "Question4.4",
    "Question4.5", "Question4.6", "Question5", "ChangeInEmployment",
    "ChangeInWorkingHours", "Question7", "Question8.1", "Question8.2",
    "Question8.3", "Question8.4", "Question8.5", "Question8.6",
    "Question8.7", "Question20.1", "Question20.2", "Question20.3",
    "Question23", "LivingStatus", "MaritalStatus", "Ethnicity",
    "HouseholdIncome", "LivingArea", "ID", "Province", "Province/region",
    "Age group", "Gender identity", "Q4_1 : I, or someone close to me, has tested positive for COVID-19
    "Q4_2 : I, or someone close to me, has had symptoms of COVID-19 but has not been tested
",
    "Q4_3 : I, or someone close to me, has been tested for COVID-19 but it was negative",
    "Q4_4 : I, or someone close to me, is elderly and/or has a health condition that increases the risk
    "Q4_5 : I have a job that exposes me to high risk of getting COVID-19 ",
    "Q4_6 : Someone close to me has a job that exposes them to high risk of getting COVID-19 ",
    "Q5 : How worried are you about the impact of COVID-19 on your personal financial situation?",
    "Q6 : How have physical distancing measures due to the COVID-19 pandemic affected your employment s
    "Q6b : On average, how has the number of hours you are working for pay been affected by the COVID-19
    "Q7 : How worried are you that you or someone close to you (close relative or friend) will get ill
    "Q8x1 : P2W frequency - Feeling nervous, anxious or on edge",
    "Q8x2 : P2W frequency - Not being able to stop or control worrying",
    "Q8x3 : P2W frequency - Worrying too much about different things",
    "Q8x4 : P2W frequency - Trouble relaxing", "Q8x5 : P2W frequency - Being so restless that it?s hard
    "Q8x6 : P2W frequency - Becoming easily annoyed or irritable",
    "Q8x7 : P2W frequency - Feeling afraid as if something awful might happen",
    "Q20x1 : In the PAST 7 DAYS, how often have you felt depressed?",
    "Q20x2 : In the PAST 7 DAYS, how often have you felt lonely?",
    "Q20x3 : In the PAST 7 DAYS, how often have you felt hopeful about the future?",
    "Q23 : Including yourself, how many people are currently living in your household?",
    "Live alone or with others?", "Q26 : What is your current marital status?",
    "Q27 : Which of the following best describes your racial or ethnic group?",
    "hIncome : Total household income received in the 2019 year before taxes?",
    "Q29 : Area you consider to be living in
"), nrow = 31)

kable(camhappendix) %>%
    kable_classic(latex_options = "HOLD_position", full_width = F,
        "striped") %>%
    add_header_above(c(`New Variable Name` = 1, `Raw Variable Name/Question` = 1)) %>%
    row_spec(1:nrow(camhappendix), extra_css = "padding: 10px")  #%>% save_kable(file = 'rawCamh.png',
```

New Variable Name	Raw Variable Name/Question
ID	ID
Province	Province
ProvinceCode	Province/region
AgeRange	Age group
Sex	Gender identity
Question4.1	Q4_1 : I, or someone close to me, has tested positive for COVID-19
Question4.2	Q4_2 : I, or someone close to me, has had symptoms of COVID-19 but has not been tested
Question4.3	Q4_3 : I, or someone close to me, has been tested for COVID-19 but it was negative
Question4.4	Q4_4 : I, or someone close to me, is elderly and/or has a health condition that increases the ris
Question4.5	Q4_5 : I have a job that exposes me to high risk of getting COVID-19
Question4.6	Q4_6 : Someone close to me has a job that exposes them to high risk of getting COVID-19
Question5	Q5 : How worried are you about the impact of COVID-19 on your personal financial situation?
ChangeInEmployment	Q6 : How have physical distancing measures due to the COVID-19 pandemic affected your emp
ChangeInWorkingHours	Q6b : On average, how has the number of hours you are working for pay been affected by the C
Question7	Q7 : How worried are you that you or someone close to you (close relative or friend) will get ill
Question8.1	Q8x1 : P2W frequency - Feeling nervous, anxious or on edge
Question8.2	Q8x2 : P2W frequency - Not being able to stop or control worrying
Question8.3	Q8x3 : P2W frequency - Worrying too much about different things
Question8.4	Q8x4 : P2W frequency - Trouble relaxing
Question8.5	Q8x5 : P2W frequency - Being so restless that it?s hard to sit still
Question8.6	Q8x6 : P2W frequency - Becoming easily annoyed or irritable
Question8.7	Q8x7 : P2W frequency - Feeling afraid as if something awful might happen
Question20.1	Q20x1 : In the PAST 7 DAYS, how often have you felt depressed?
Question20.2	Q20x2 : In the PAST 7 DAYS, how often have you felt lonely?
Question20.3	Q20x3 : In the PAST 7 DAYS, how often have you felt hopeful about the future?
Question23	Q23 : Including yourself, how many people are currently living in your household?
LivingStatus	Live alone or with others?
MaritalStatus	Q26 : What is your current marital status?
Ethnicity	Q27 : Which of the following best describes your racial or ethnic group?
HouseholdIncome	hIncome : Total household income received in the 2019 year before taxes?
LivingArea	Q29 : Area you consider to be living in