WHAT INFLUENCE CANADIAN RESIDENTS' HEALTH SELF-REPORTED

Yuhan Zhao, 1003986840

22 Dec, 2020

Abstract

With the increasing social pressure, many people feel that their health (including psychological and physical) has appeared problems. Fortunately, the 2017 General Social Survey(GSS)on families has offered a wide range of data regarding individuals' social, financial, educational background. The main focus of the study was to assess the distribution of the respondents' demographic variables and finally to assess the factors the influence ones' feeling of health. The descriptive statistics, frequency tabulation as well as the ordinal logistic regression model were used to derive the findings of the study. The following derivations may be made from the results of the regression model; Higher-income individuals reported a higher overall feeling of health.

Introduction

The study I used is the GSS 2017 data set to analyze the demographic variables and self-reported perception of overall health. The primary objectives of the General Social Survey (GSS) are the provision of information on specific social policy issues of emerging interests and gathering information related to social trends which enables monitoring changes in the living conditions and overall well-being of Canadians over time. The utility of the collected information lies greatly with program and policy areas such as childcare strategies, parental benefits, child custody, and spousal support programs as well as means-tested allocations. The data analysis procedure involves both exploratory and inferential statistics. The first part presents the summary statistics through various measures of spread and central tendencies and also frequency tabulation through cross tabs. The inferential statistics applies the use of chi-square tests and the ordinal logistic regression model to assess the association between the GSS data set variables. The data set is considered appropriate for the study since it has the appropriate sample size of over 20,000 observations which are relevant to the research objectives. The research question was as follows; Is the education level of individuals associated with their citizenship type? And what are the variables that influence individuals' feeling of self-reported health?

Data

```
gss=read.csv("gss.csv")
head(gss)
```

```
##   caseid  age age_first_child age_youngest_child_under_6 total_children
## 1      1 52.7              27                          NA              1
## 2      2 51.1              33                          NA              5
## 3      3 63.6              40                          NA              5
## 4      4 80.0              56                          NA              1
## 5      5 28.0              NA                          NA              0
## 6      6 63.0              37                          NA              2
```

```
##   age_start_relationship age_at_first_marriage age_at_first_birth
## 1                     NA                    NA               25.9
## 2                     NA                    NA                 NA
## 3                     NA                    NA               23.2
## 4                     NA                    NA               27.3
## 5                   25.3                    NA                 NA
## 6                     NA                    NA               25.8
##   distance_between_houses age_youngest_child_returned_work feelings_life    sex
## 1                      30                                NA             8 Female
## 2                      NA                                NA            10   Male
## 3                      NA                                NA             8 Female
## 4                      NA                                NA            10 Female
## 5                      NA                                NA             8   Male
## 6                      NA                                NA             9 Female
##   place_birth_canada  place_birth_father  place_birth_mother
## 1    Born in Canada      Born in Canada      Born in Canada
## 2    Born in Canada      Born in Canada      Born in Canada
## 3    Born in Canada      Born in Canada      Born in Canada
## 4    Born in Canada Born outside Canada Born outside Canada
## 5    Born in Canada      Born in Canada      Born in Canada
## 6    Born in Canada      Born in Canada      Born in Canada
##   place_birth_macro_region place_birth_province year_arrived_canada province
## 1                     <NA>               Quebec                <NA>   Quebec
## 2                     <NA>              Ontario                <NA> Manitoba
## 3                     <NA>              Ontario                <NA>  Ontario
## 4                     <NA>              Alberta                <NA>  Alberta
## 5                     <NA>               Quebec                <NA>   Quebec
## 6                     <NA>               Quebec                <NA>   Quebec
##          region                                          pop_center
## 1         Quebec              Larger urban population centres (CMA/CA)
## 2 Prairie region              Larger urban population centres (CMA/CA)
## 3        Ontario Rural areas and small population centres (non CMA/CA)
## 4 Prairie region              Larger urban population centres (CMA/CA)
## 5         Quebec              Larger urban population centres (CMA/CA)
## 6         Quebec              Larger urban population centres (CMA/CA)
##         marital_status aboriginal           vis_minority age_immigration
## 1 Single, never married         No Not a visible minority            <NA>
## 2               Married         No Not a visible minority            <NA>
## 3               Married         No Not a visible minority            <NA>
## 4               Married         No Not a visible minority            <NA>
## 5    Living common-law         No Not a visible minority            <NA>
## 6               Married         No Not a visible minority            <NA>
##   landed_immigrant citizenship_status
## 1             <NA>           By birth
## 2             <NA>           By birth
## 3             <NA>           By birth
## 4             <NA>           By birth
## 5             <NA>           By birth
## 6             <NA>           By birth
##                                                  education
## 1 High school diploma or a high school equivalency certificate
## 2                               Trade certificate or diploma
## 3                  Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
## 4 High school diploma or a high school equivalency certificate
```

```
## 5  College, CEGEP or other non-university certificate or di...
## 6 High school diploma or a high school equivalency certificate
##                                                      own_rent
## 1 Owned by you or a member of this household, even if it i...
## 2 Owned by you or a member of this household, even if it i...
## 3 Owned by you or a member of this household, even if it i...
## 4 Owned by you or a member of this household, even if it i...
## 5                       Rented, even if no cash rent is paid
## 6 Owned by you or a member of this household, even if it i...
##   living_arrangement                                hh_type hh_size
## 1              Alone Low-rise apartment (less than 5 stories)       1
## 2        Spouse only                   Single detached house       2
## 3        Spouse only                   Single detached house       2
## 4        Spouse only                                   Other       2
## 5        Spouse only Low-rise apartment (less than 5 stories)       2
## 6        Spouse only                   Single detached house       2
##   partner_birth_country partner_birth_province   partner_vis_minority
## 1               Canada                 Quebec Not a visible minority
## 2               Canada               Manitoba Not a visible minority
## 3               Canada                Ontario Not a visible minority
## 4               Canada                Alberta Not a visible minority
## 5               Canada                 Quebec Not a visible minority
## 6               Canada                 Quebec Not a visible minority
##   partner_sex                                       partner_education
## 1        <NA>                                Trade certificate or diploma
## 2        <NA>                Bachelor's degree (e.g. B.A., B.Sc., LL.B.)
## 3        <NA>  High school diploma or a high school equivalency certi...
## 4        <NA>            Less than high school diploma or its equivalent
## 5        <NA> College, CEGEP or other non-university certificate or d...
## 6        <NA> College, CEGEP or other non-university certificate or d...
##   average_hours_worked worked_last_week             partner_main_activity
## 1   30.0 to 40.0 hours              Yes Working at a paid job or business
## 2  50.1 hours and more              Yes                   Going to school
## 3                 <NA>               No                           Retired
## 4                 <NA>               No                           Retired
## 5   30.0 to 40.0 hours               No Working at a paid job or business
## 6                 <NA>               No                           Retired
##   self_rated_health self_rated_mental_health  religion_has_affiliation
## 1         Excellent                Excellent Has religious affiliation
## 2              Good                     Good                Don't know
## 3         Very good                     Good Has religious affiliation
## 4         Very good                Very good Has religious affiliation
## 5              Good                     Good Has religious affiliation
## 6         Excellent                Very good Has religious affiliation
##    regilion_importance language_home      language_knowledge
## 1   Somewhat important        French             French only
## 2           Don't know       English            English only
## 3       Very important        French Both English and French
## 4 Not at all important       English            English only
## 5 Not at all important        French Both English and French
## 6       Very important        French             French only
##           income_family  income_respondent
## 1    $25,000 to $49,999 $25,000 to $49,999
## 2    $75,000 to $99,999  Less than $25,000
```

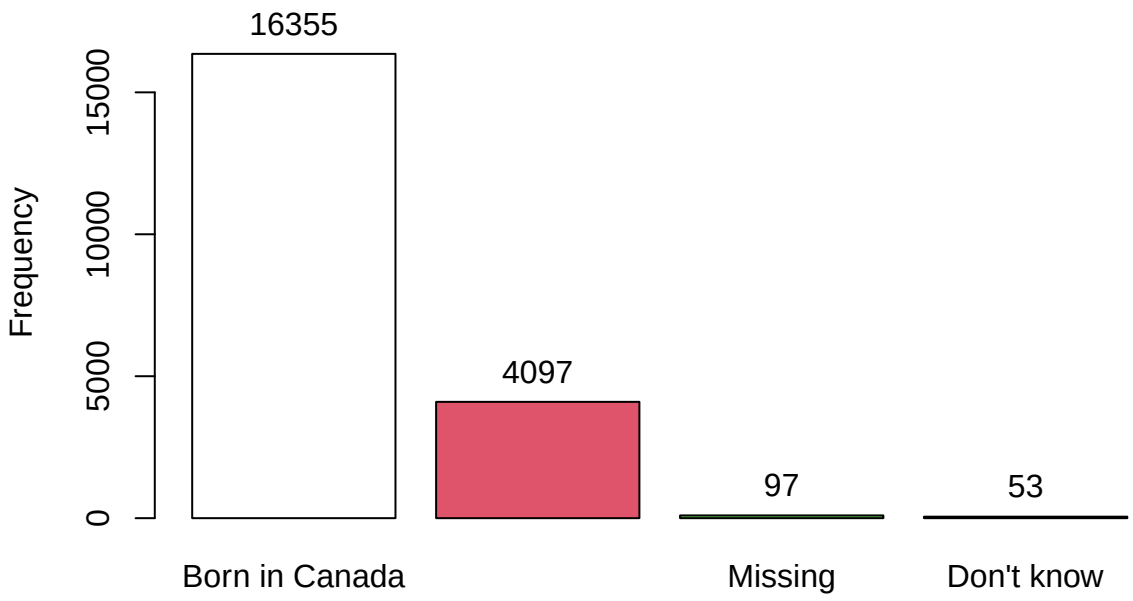
```
## 3    $75,000 to $99,999 $25,000 to $49,999
## 4 $100,000 to $ 124,999 $50,000 to $74,999
## 5    $50,000 to $74,999  Less than $25,000
## 6    $50,000 to $74,999  Less than $25,000
##                                                 occupation childcare_regular
## 1                             Sales and service occupations              <NA>
## 2 Trades, transport and equipment operators and related oc...              <NA>
## 3                                                      <NA>              <NA>
## 4                                                      <NA>              <NA>
## 5                             Sales and service occupations              <NA>
## 6                                                      <NA>              <NA>
##   childcare_type childcare_monthly_cost ever_fathered_child ever_given_birth
## 1           <NA>                   <NA>               <NA>              Yes
## 2           <NA>                   <NA>                Yes             <NA>
## 3           <NA>                   <NA>               <NA>              Yes
## 4           <NA>                   <NA>               <NA>              Yes
## 5           <NA>                   <NA>                 No             <NA>
## 6           <NA>                   <NA>               <NA>              Yes
##   number_of_current_union lives_with_partner children_in_household
## 1                    <NA>                No             No child
## 2                    <NA>                No             No child
## 3                    <NA>                No             No child
## 4                    <NA>                No             No child
## 5             Second union               Yes             No child
## 6                    <NA>                No             No child
##   number_total_children_intention has_grandchildren grandparents_still_living
## 1                              NA                No                       No
## 2                              NA               Yes                       No
## 3                              NA               Yes                       No
## 4                              NA                No                       No
## 5                               2                No                      Yes
## 6                              NA               Yes                       No
##   ever_married current_marriage_is_first number_marriages
## 1           No                      <NA>               0
## 2          Yes                       Yes               1
## 3          Yes                       Yes               1
## 4          Yes                       Yes               1
## 5           No                      <NA>               0
## 6          Yes                       Yes               1
##   religion_participation partner_location_residence full_part_time_work
## 1   Once or twice a year       In the same province               <NA>
## 2             Don't know                       <NA>               <NA>
## 3   At least once a week                       <NA>               <NA>
## 4             Not at all                       <NA>               <NA>
## 5             Not at all                       <NA>               <NA>
## 6             Not at all                       <NA>               <NA>
##   time_off_work_birth reason_no_time_off_birth returned_same_job
## 1                <NA>                    <NA>              <NA>
## 2                <NA>                    <NA>              <NA>
## 3                <NA>                    <NA>              <NA>
## 4                <NA>                    <NA>              <NA>
## 5                <NA>                    <NA>              <NA>
## 6                <NA>                    <NA>              <NA>
##   satisfied_time_children provide_or_receive_fin_supp fin_supp_child_supp
```

```
## 1                    <NA>                         <NA>                   NA
## 2                    <NA>                         <NA>                   NA
## 3                    <NA>                         <NA>                   NA
## 4                    <NA>                         <NA>                   NA
## 5                    <NA>                         <NA>                   NA
## 6                    <NA>                         <NA>                   NA
##   fin_supp_child_exp fin_supp_lump fin_supp_other fin_supp_agreement
## 1                 NA            NA             NA                <NA>
## 2                 NA            NA             NA                <NA>
## 3                 NA            NA             NA                <NA>
## 4                 NA            NA             NA                <NA>
## 5                 NA            NA             NA                <NA>
## 6                 NA            NA             NA                <NA>
##   future_children_intention is_male main_activity
## 1                      <NA>       0            NA
## 2                      <NA>       1            NA
## 3                      <NA>       0            NA
## 4                      <NA>       0            NA
## 5                      <NA>       1            NA
## 6                      <NA>       0            NA
##                                    age_diff number_total_children_known
## 1                                      <NA>                           0
## 2                 Respondent is 4 years older                         0
## 3                 Respondent is 3 years older                         0
## 4 Respondent and spouse/partner are same age                          0
## 5                                      <NA>                           1
## 6                  Respondent is 1 year older                         0
```

The data set is a General Social Survey data- related to Canadian families. The variables in the data include information related to family origins, conjugal and parental history (chronology of marriages, common-law unions, and children), children's home leaving, and other socioeconomic characteristics. The study targeted the population of all non-institutionalized persons 15 years of age or older, residing in the 10 provinces of Canada. The data was collected using a survey developed and administered questionnaire based on research and extensive liaison with key partners and data users. A pilot study was carried out with respondents in two cities, selected based on representative criteria. From the pilot, the Questions were fashioned in a manner that would ensure collecting important information required for the study with utmost efficiency. The sampling followed a cross-sectional design. The utility of the collected information lies greatly with program and policy areas such as childcare strategies, parental benefits, child custody, and spousal support programs as well as means-tested allocations. The data set had over 20,000 observations which make it appropriates for use in deriving some generalizations about the population. The study based on the GSS survey data is generalizable to the target population if the non-response bias is ignored.

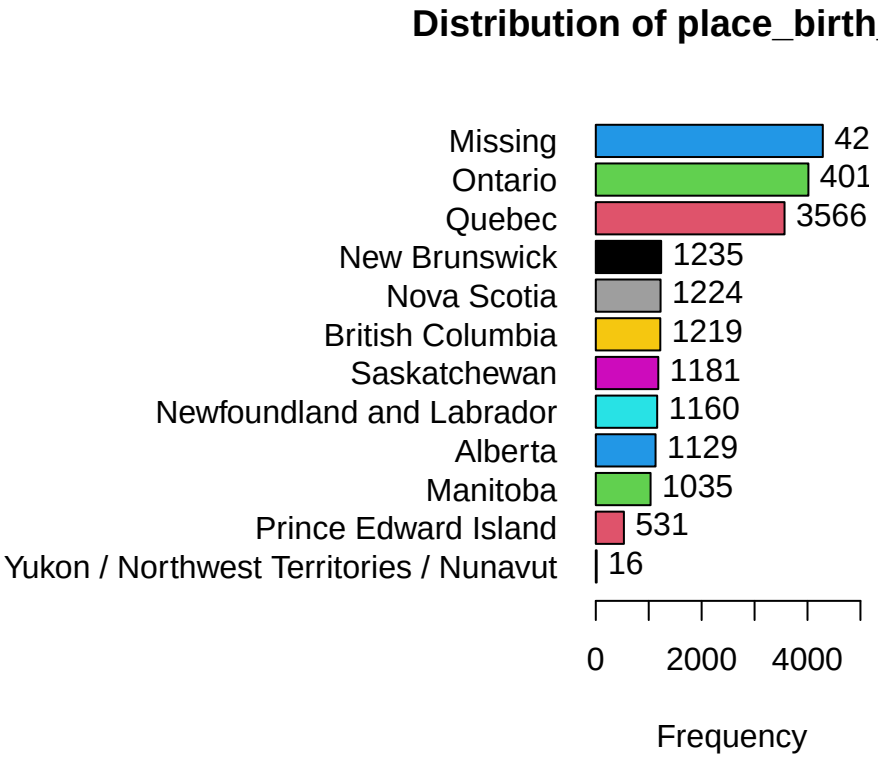
Exploratory Data Analysis

```
library(epiDisplay)
tab1(gss$place_birth_canada, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of plac
```



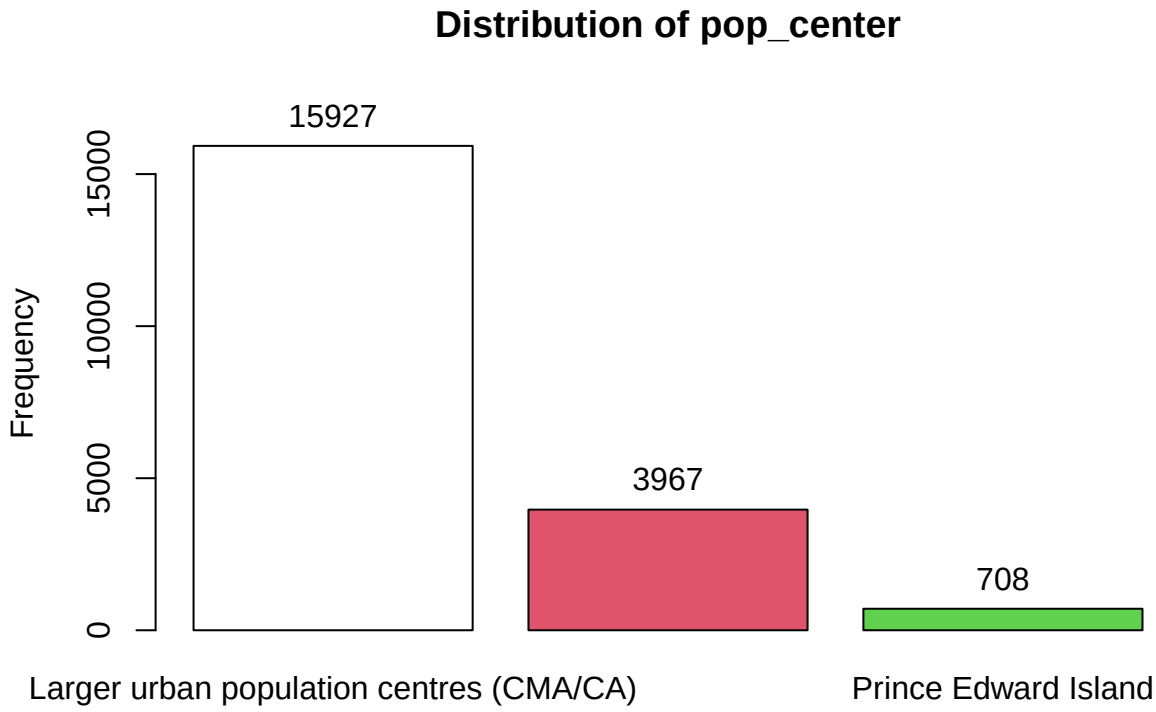

```
## gss$place_birth_canada :
##                       Frequency   %(NA+) cum.%(NA+)   %(NA-) cum.%(NA-)
## Born in Canada            16355     79.4       79.4     79.8       79.8
## Born outside Canada        4097     19.9       99.3     20.0       99.7
## <NA>                         97      0.5      100.0      0.0      100.0
## Don't know                   53      0.3       99.5      0.3      100.0
##   Total                   20602    100.0      100.0    100.0      100.0
```

```
tab1(gss$place_birth_province, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of pla
```



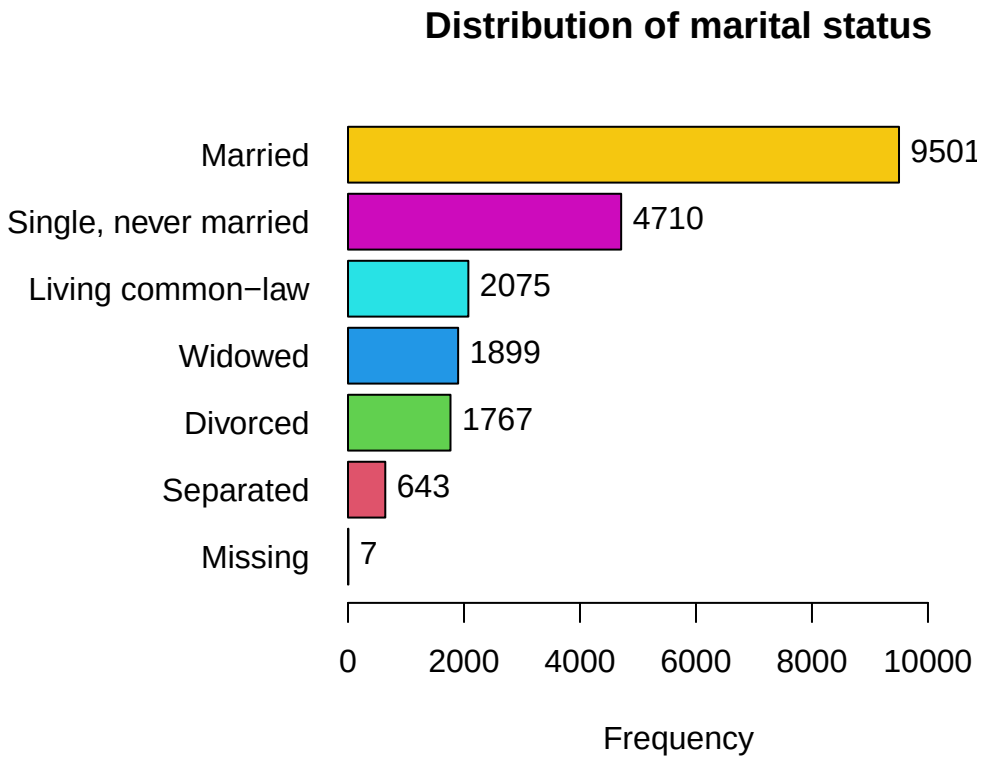

```
## gss$place_birth_province :
##                                        Frequency   %(NA+) cum.%(NA+)   %(NA-)
## <NA>                                        4289     20.8       100.0      0.0
## Ontario                                     4017     19.5        53.5     24.6
## Quebec                                      3566     17.3        73.4     21.9
## New Brunswick                               1235      6.0        22.4      7.6
## Nova Scotia                                 1224      5.9        34.0      7.5
## British Columbia                            1219      5.9        11.4      7.5
## Saskatchewan                                1181      5.7        79.1      7.2
## Newfoundland and Labrador                   1160      5.6        28.0      7.1
## Alberta                                     1129      5.5         5.5      6.9
## Manitoba                                    1035      5.0        16.4      6.3
## Prince Edward Island                         531      2.6        56.1      3.3
## Yukon / Northwest Territories / Nunavut       16      0.1        79.2      0.1
##   Total                                    20602    100.0       100.0    100.0
##                                        cum.%(NA-)
## <NA>                                        100.0
## Ontario                                      67.5
## Quebec                                       92.7
## New Brunswick                                28.3
## Nova Scotia                                  42.9
## British Columbia                             14.4
## Saskatchewan                                 99.9
## Newfoundland and Labrador                    35.4
## Alberta                                       6.9
## Manitoba                                     20.7
## Prince Edward Island                         70.8
## Yukon / Northwest Territories / Nunavut     100.0
##   Total                                     100.0
```

```r
tab1(gss$pop_center, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of pop_center")
```



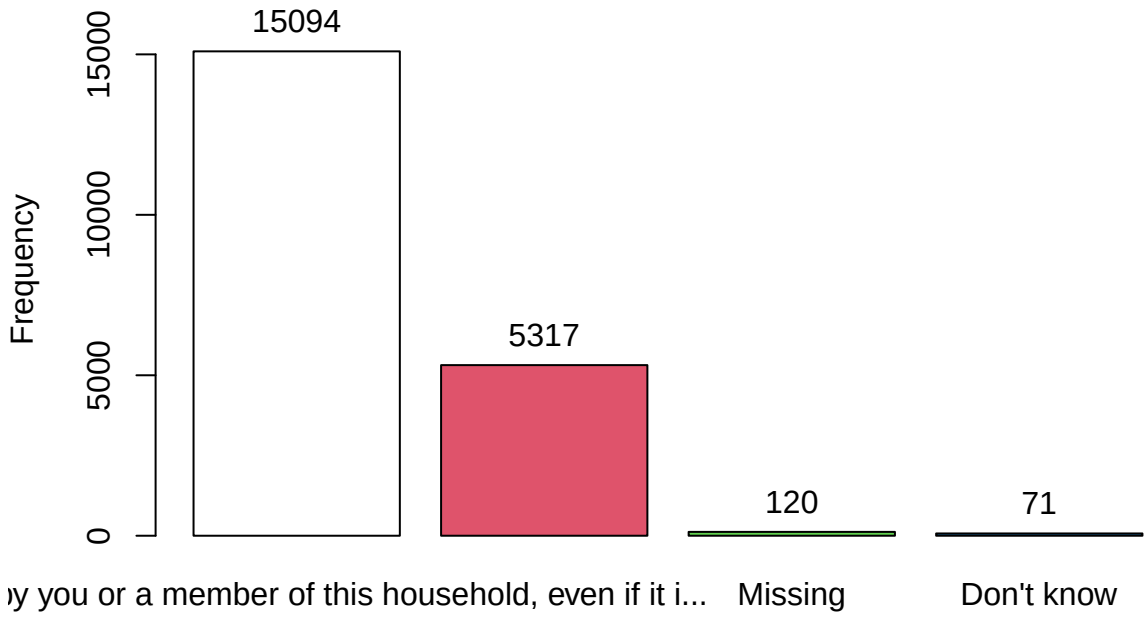

```
## gss$pop_center :
##                                                    Frequency Percent
## Larger urban population centres (CMA/CA)               15927    77.3
## Rural areas and small population centres (non CMA/CA)   3967    19.3
## Prince Edward Island                                     708     3.4
##   Total                                                20602   100.0
##                                                    Cum. percent
## Larger urban population centres (CMA/CA)                   77.3
## Rural areas and small population centres (non CMA/CA)      96.6
## Prince Edward Island                                      100.0
##   Total                                                   100.0
```

```r
tab1(gss$marital_status, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of marital s
```



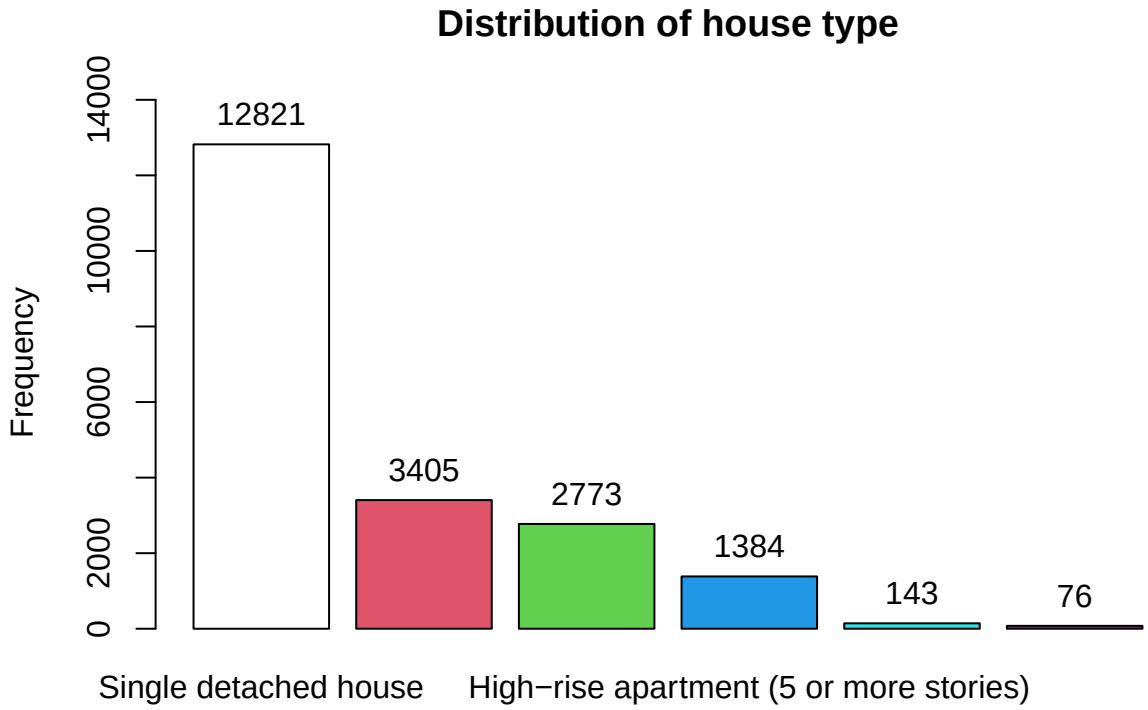

```
## gss$marital_status :
##                        Frequency   %(NA+) cum.%(NA+)   %(NA-) cum.%(NA-)
## Married                     9501     46.1       64.8     46.1       64.8
## Single, never married       4710     22.9       90.7     22.9       90.8
## Living common-law           2075     10.1       18.6     10.1       18.7
## Widowed                     1899      9.2      100.0      9.2      100.0
## Divorced                    1767      8.6        8.6      8.6        8.6
## Separated                    643      3.1       67.9      3.1       67.9
## <NA>                           7      0.0      100.0      0.0      100.0
##   Total                    20602    100.0      100.0    100.0      100.0
```

```
tab1(gss$own_rent, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of residence statu
```




```
## gss$own_rent :
##                                                        Frequency   %(NA+)
## Owned by you or a member of this household, even if it i...     15094     73.3
## Rented, even if no cash rent is paid                            5317     25.8
## <NA>                                                             120      0.6
## Don't know                                                        71      0.3
##   Total                                                        20602    100.0
##                                                        cum.%(NA+)   %(NA-)
## Owned by you or a member of this household, even if it i...       73.6     73.7
## Rented, even if no cash rent is paid                             99.4     26.0
## <NA>                                                            100.0      0.0
## Don't know                                                        0.3      0.3
##   Total                                                         100.0    100.0
##                                                        cum.%(NA-)
## Owned by you or a member of this household, even if it i...       74.0
## Rented, even if no cash rent is paid                            100.0
## <NA>                                                            100.0
## Don't know                                                        0.3
##   Total                                                         100.0
```

```
tab1(gss$hh_type, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of house type")
```




```
## gss$hh_type :
##                                          Frequency   %(NA+) cum.%(NA+)   %(NA-)
## Single detached house                        12821     62.2       99.6     62.5
## Other                                         3405     16.5       37.4     16.6
## Low-rise apartment (less than 5 stories)      2773     13.5       20.9     13.5
## High-rise apartment (5 or more stories)       1384      6.7        7.4      6.7
## Don't know                                     143      0.7        0.7      0.7
## <NA>                                            76      0.4      100.0      0.0
##   Total                                      20602    100.0      100.0    100.0
##                                          cum.%(NA-)
## Single detached house                         100.0
## Other                                          37.5
## Low-rise apartment (less than 5 stories)       20.9
## High-rise apartment (5 or more stories)         7.4
## Don't know                                      0.7
## <NA>                                          100.0
##   Total                                       100.0
```

```
tab1(gss$living_arrangement, sort.group = "decreasing", cum.percent = TRUE, main = "Distribution oflivi
```

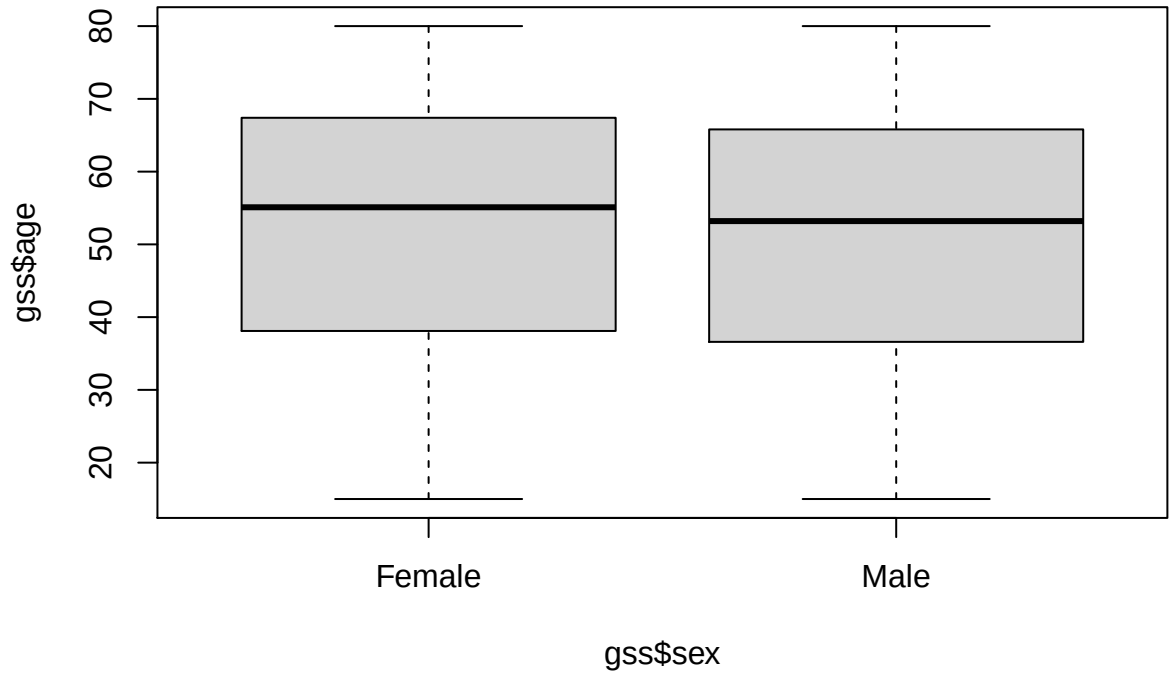

```
## gss$living_arrangement :
##                                                 Frequency Percent
## Spouse only                                          6398    31.1
## Alone                                                5815    28.2
## Spouse and single child under 25 years of age        4136    20.1
## Other living arrangement                             1134     5.5
## Living with two parents                               972     4.7
## No spouse and single child under 25 years of age      814     4.0
## Living with one parent                                511     2.5
## Spouse and single child 25 years of age or older      373     1.8
## No spouse and single child 25 years of age or older   213     1.0
## Spouse and other                                      205     1.0
## No spouse and non-single child(ren)                    20     0.1
## Spouse and non-single child(ren)                       11     0.1
##   Total                                             20602   100.0
##                                                 Cum. percent
## Spouse only                                             31.1
## Alone                                                   59.3
## Spouse and single child under 25 years of age           79.4
## Other living arrangement                                84.9
## Living with two parents                                 89.6
## No spouse and single child under 25 years of age        93.5
## Living with one parent                                  96.0
## Spouse and single child 25 years of age or older        97.8
## No spouse and single child 25 years of age or older     98.9
## Spouse and other                                        99.8
## No spouse and non-single child(ren)                     99.9
## Spouse and non-single child(ren)                       100.0
##   Total                                                100.0
```

Descriptive statistics

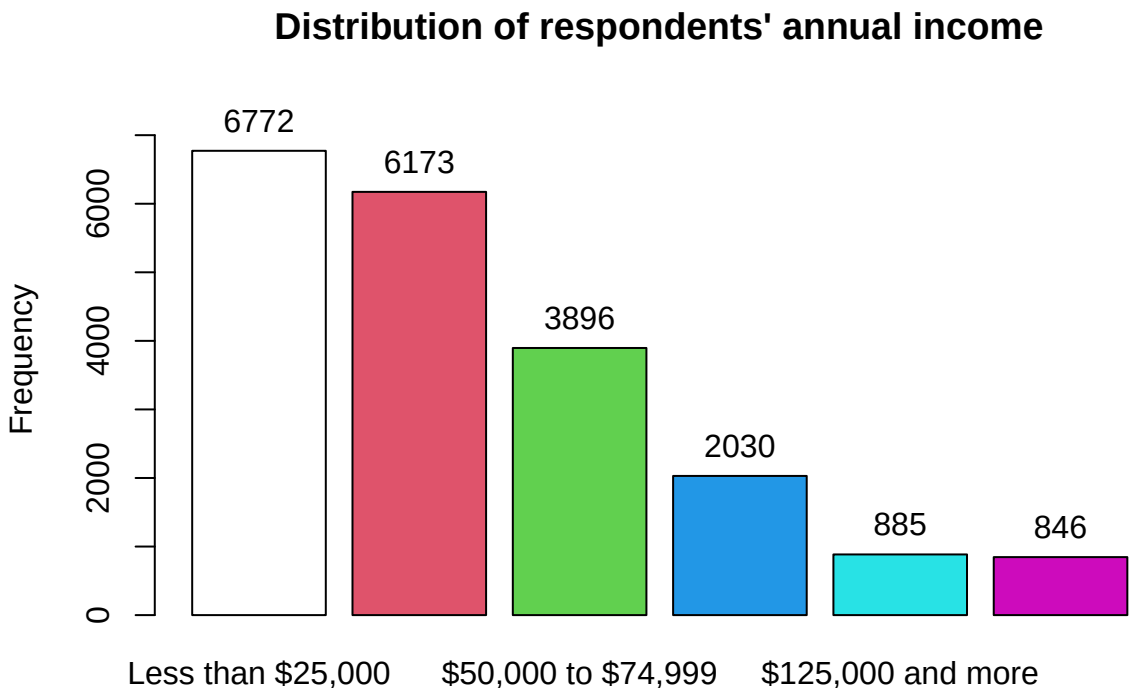
```
tapply(gss$age,gss$sex,mean)
```

```
##   Female     Male
## 52.92865 51.31056
```

```
boxplot(gss$age~gss$sex)
```

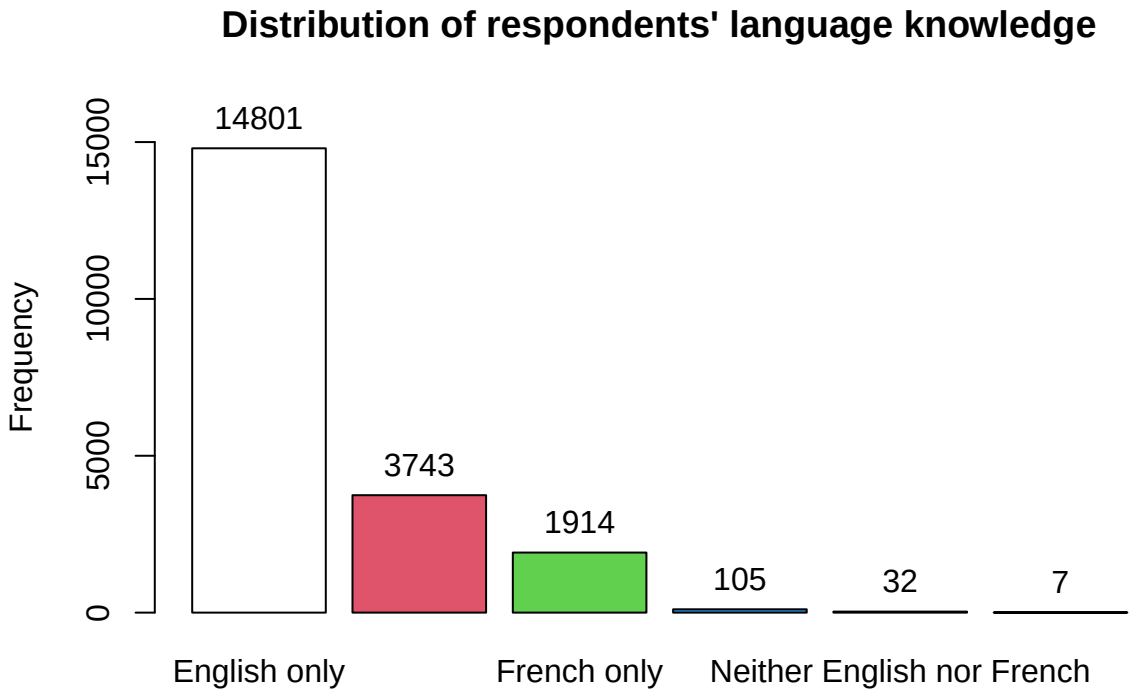



```
tab1(gss$income_respondent, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of respo
```



```
## gss$income_respondent :
##                         Frequency Percent Cum. percent
## Less than $25,000            6772    32.9         32.9
## $25,000 to $49,999           6173    30.0         62.8
## $50,000 to $74,999           3896    18.9         81.7
## $75,000 to $99,999           2030     9.9         91.6
## $125,000 and more             885     4.3         95.9
## $100,000 to $ 124,999         846     4.1        100.0
##   Total                     20602   100.0        100.0
```

```
tab1(gss$language_knowledge, sort.group = "decreasing", cum.percent = TRUE,main = "Distribution of respo
```




```
## gss$language_knowledge :
##                              Frequency   %(NA+) cum.%(NA+)   %(NA-) cum.%(NA-)
## English only                     14801    71.8       90.0    72.2       90.5
## Both English and French           3743    18.2       18.2    18.3       18.3
## French only                       1914     9.3       99.3     9.3       99.8
## <NA>                               105     0.5      100.0     0.0      100.0
## Neither English nor French          32     0.2       99.5     0.2      100.0
## Don't know                           7     0.0       18.2     0.0       18.3
##   Total                          20602   100.0      100.0   100.0      100.0
```

The distribution of respondents' income reveals that most earn less than 25000 dollars and only a few earning 125000 dollars or more annually. The individual incomes are skewed to the right. The female respondents were more that the males, at 11203 and 9399 respectively. The female had a relatively higher average age of 52.92 years compared to the males who had a mean age of 51.31.

Methodology

Inferential statistics

The inferential statistics were performed using the chi-square test for association and the ordinal logistic regression model. The choice of the ordinal logistic regression model was informed by the fact that the dependent variable, "self_reported_health" occurs in ordered levels Stewart et al (2019). The categories were as follows; "Poor", "fair", "Good", "very good" and "Excellent". The coefficients of the model presents the odds of belonging to a certain category given a level of the independent variable.

Logistic regression modelling

```
require(foreign)
require(ggplot2)
require(MASS)
require(Hmisc)
require(reshape2)
##Ordering the dependent variable
 self_rated_health= factor(gss$self_rated_health, levels = c("Poor", "Fair", "Good","Very good","Excell

## Fitting ordered logit model and store results 'mod1'
mod1 <- (polr(as.factor(self_rated_health)~income_respondent + age + marital_status+occupation+sex+incor
##### Summarizing coefficients in a table
broom::tidy(mod1)
```

```
## # A tibble: 32 x 5
##    term                                 estimate std.error statistic coef.type
##    <chr>                                   <dbl>     <dbl>     <dbl> <chr>
##  1 income_respondent$125,000 and more  -0.127      0.119      -1.06  coefficient
##  2 income_respondent$25,000 to $49,999  0.0631     0.107       0.589 coefficient
##  3 income_respondent$50,000 to $74,999  0.0819     0.101       0.808 coefficient
##  4 income_respondent$75,000 to $99,999 -0.0198     0.103      -0.191 coefficient
##  5 income_respondentLess than $25,000  -0.0306     0.121      -0.254 coefficient
##  6 age                                 -0.000386   0.00187    -0.207 coefficient
##  7 marital_statusMarried                1.60       1.30        1.23  coefficient
##  8 marital_statusSeparated              1.46       1.29        1.13  coefficient
##  9 occupationHealth occupations        -0.180      0.0943     -1.90  coefficient
## 10 occupationManagement occupations    -0.0770     0.0877     -0.878 coefficient
## # ... with 22 more rows
```

```
#(ctable <- coef(summary(mod1)))
## calculate and store p values
#p <- pnorm(abs(ctable[, "t value"]), lower.tail = FALSE) * 2

## combined table
#p_val=(ctable <- cbind(coef(summary(mod1)), "p value" = p));p_val
```

The values for the model coefficients and intercepts are displayed in the table. The results also contain the standard errors and t values which are used to calculate the p-values. All the parameters for the fitted model were significant at the 0.05 levels since the p-values were all less than 0.05. The coefficients present the relative influence of the individual variables on the overall self-reported feeling of health. Holding everything other factors constant, an increase in value of age by one unit decreases the expected value of self reported

health in log odds by 0.00287. Respondents who earned less than 25000 dollars annually had higher likelihood of feeling less healthy. Those in the middle incomes if between 25000 and 74,000 dollars had higher odds of having an excellent self reported health. The marital_status, "Living common-law" is associated with a decreases the expected value of self reported health in log odds by 1.497. The married individuals are highly likely to report a better rating for self reported health. Those individuals in occupational health, natural and applied sciences, management, art and culture, education, manufacturing and utilities as well as service occupation are likely to reporting lowering ratings of self reported health. On the flip side, those in agriculture and related production, and transport and equipment operators are likely to report higher health index. Long hours of work are associated with lower ratings of self reported health. The odds of feeling less healthy are even more among those who work for over 50 hours per week. Higher family income is found to be a predictor for a higher rating of self-reported health. individuals that live with partners are on average likely to report higher rating variable had a positive influence on self reported health in both the two levels, it was noted that individuals who reported that current marriage is first have hihger odds of reproting better feeling ofself reported health.

Results

Distribution of citizenship in relation to the education level

```
ed_contigency=table(gss$education, gss$citizenship_status)
ed_contigency
```

```
##
##                                                                 By birth
##   Bachelor's degree (e.g. B.A., B.Sc., LL.B.)                       2752
##   College, CEGEP or other non-university certificate or di...       3712
##   High school diploma or a high school equivalency certificate      4035
##   Less than high school diploma or its equivalent                   2636
##   Trade certificate or diploma                                      1319
##   University certificate or diploma below the bachelor's level      560
##   University certificate, diploma or degree above the bach...       1212
##
##                                                                 By naturalization
##   Bachelor's degree (e.g. B.A., B.Sc., LL.B.)                                691
##   College, CEGEP or other non-university certificate or di...                655
##   High school diploma or a high school equivalency certificate               591
##   Less than high school diploma or its equivalent                            290
##   Trade certificate or diploma                                               125
##   University certificate or diploma below the bachelor's level               120
##   University certificate, diploma or degree above the bach...                441
##
##                                                                 Don't know
##   Bachelor's degree (e.g. B.A., B.Sc., LL.B.)                            8
##   College, CEGEP or other non-university certificate or di...           10
##   High school diploma or a high school equivalency certificate          19
##   Less than high school diploma or its equivalent                        8
##   Trade certificate or diploma                                           3
##   University certificate or diploma below the bachelor's level          4
##   University certificate, diploma or degree above the bach...            3
```

Most of the respondents who are Bachelor's degree holders, college certificate holders and higher diploma holders are Canadian citizens by birth. The study further analyzed whether the citizenship type was associated with the education level using the chi-square test for association.

```
(ct <- chisq.test(ed_contigency))
```

```
## Warning in chisq.test(ed_contigency): Chi-squared approximation may be incorrect
```

```
##
##  Pearson's Chi-squared test
##
## data:  ed_contigency
## X-squared = 374.42, df = 12, p-value < 2.2e-16
```

The results of the test imply that an individual's level of education is in fact associated with the citizenship variable. The p-value for the test of approximately 0.00 presents evidence against the null hypothesis that education is not associated with type of citizenship.

There are certainly a wide range of factors that influence one's feeling of self reported health, among them is the marital status, income, the number of hours of work as well as the occupation. The married individuals are highly likely to report a better rating for self reported health.This may be attributed to the emotional and physical support that one may derive for a partner. Those individuals in occupational health, natural and applied sciences, management, art and culture, education, manufacturing and utilities as well as service occupation are likely to reporting lowering ratings of self reported health. On the flip side, those in agriculture and related production, and transport and equipment operators are likely to report higher health index. It is not apparent why some occupations are associated with better self reported health while others are not. Long hours of work are associated with lower ratings of self reported health. The odds of feeling less healthy are even more among those who work for over 50 hours per week. Higher family income is found to be a predictor for a higher rating of self-reported health. individuals that live with partners are on average likely to report higher rating of self reported health with increase in odds of up to 1.11 times. Even though the current marriage variable had a positive influence on self reported health in both the two levels, it was noted that individuals who reported that current marriage is first have higher odds of reporting better feeling of self reported health.

Discussion

The study used the GSS data set whose target population for 2017 survey is all non-institutionalized persons 15 years of age or older, living in the 10 provinces of Canada. The analysis presents distribution of individuals in terms demographic variables such as gender, age, education, occupation, place of birth, marital status as well as their opinions on their perceived state of health and mental health. The data also contains information house tenure type, house types and living arrangements. The results reveal that most of the individuals are married and live as spouses, most of them also own their house or at least the houses are owned by a member of their households. Ontario was the most represented province in the survey. A vast majority of the respondents have a good knowledge of English language or both French and English. The single detached houses, other housed types and low-rise houses below 5 stories were most common among the individuals in the study sample. The following derivations may be made from the results of the regression model; Respondents who earned less than 25000 dollars annually had higher likelihood of feeling less healthy. Those in the middle incomes if between 25000 and 74,000 dollars had higher odds of having an excellent self reported health. The association between income and higher health index may be explained by the fact that as income increases, one may have access to better medical coverage i.e through an insurance policy or otherwise. The individual may also afford specialized treatment and care, there is also a sense of security that may be associated with high incomes. The marital status, "Living common-law" is associated with a decreases the expected value of self reported health in log odds by 1.497. While, it might means married individuals are highly likely to report a better rating for self-reported health. This may be attributed to the emotional and physical support that one may derive from a partner. Those individuals in occupational health, natural and applied sciences, management, art and culture, education manufacturing and utilities as well as service occupation are likely to reporting lowering ratings of self-reported health. On the flip-side, those in agriculture and related products, and transport and equipment operators are likely to report higher

health index. It is not apparent why some occupations are associated with better self-reported health while others are not. Long hours of work are associated with lower ratings of self-reported health. The odds of feeling less healthy are even more among those who work for over 50 hours per week. Higher family income is found to be a predictor for a higher rating of self-reported health. individuals that live with partners are on average likely to report a higher rating of self-reported health with an increase in odds of up to 1.11 times. Even though the current marriage variable had a positive influence on self-reported health in both the two levels, it was noted that individuals who reported that current marriage is the first to have higher odds of reporting a better feeling of self-reported health.

Weaknesses

The missing values may negatively impact the model estimates obtained using the data, the biggest weakness of our research is that some factors of the dataset are missing value. In statistics, we call this non-response problem, it is a kind of non-sampling error. Therefore, it can not be sure that the information is completely valid. Besides, although this survey uses Two-Stage Cluster sampling, the sampling units are the groups of telephone numbers, which means people who do not use the telephone can not be invited to this survey. Moreover, in this survey, respondents could be interviewed in French or English, in this way, people who can not use these two languages fluently(for example, indigenous) can not join this survey either. These also caused other non-sampling errors, imperfect sampling populations, and selection biases. And since these data are based on a person's sample, sampling errors are easy to occur. In other words, sample-based estimates vary from sample to sample, and the survey we are using right now was launched three years ago(2017), so our result may not work with the next survey. The missing values may negatively impact the model estimates obtained using the data. More attention should be given to addressing this challenge of missing values. Last but not least important aspect to be noted is that the study is not causal, since it does not employ random assignments and is only observational.

Next Steps

The results of the study were not exhaustive, possible further direction may include using other classification techniques such as the random forest models and the support vector machine for determining the variables the influence one's feeling of health. In the next step, we can respond to our results with the related government to let them follow up, and the government may launch the policy to let people who have low self-reported get some economic or other non-money subsidies. For example, we can research do males or females feel happier? And people in which industry feel better? All the results would help to advance the research into analyzing the factors that influence the variation in self-reported health index among individuals from different occupations and status. Basically, if we will do the next research, I think it will be focus on why there are differences in self-reported feeling of health across different gender and jobs.

References

[1]General Social Survey - Family:https://www150.statcan.gc.ca/n1/en/surveys/4501

[2]Stewart, G., Kamata, A., Miles, R., Grandoit, E., Mandelbaum, F., Quinn, C., & Rabin, L. (2019). Predicting mental health help seeking orientations among diverse Undergraduates: An ordinal logistic regression analysis???. Journal of affective disorders, 257, 271-280.