a3_results

James Bao

10/18/2020

```
# load the csv, can be downloaded via utoronto
poll <- as_tibble(data.frame(read_csv("gss_cleaned.csv")))
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   caseid = col_double(),
##   age = col_double(),
##   age_first_child = col_double(),
##   age_youngest_child_under_6 = col_double(),
##   total_children = col_double(),
##   age_start_relationship = col_double(),
##   age_at_first_marriage = col_double(),
##   age_at_first_birth = col_double(),
##   distance_between_houses = col_double(),
##   age_youngest_child_returned_work = col_double(),
##   feelings_life = col_double(),
##   hh_size = col_double(),
##   number_total_children_intention = col_double(),
##   number_marriages = col_double(),
##   fin_supp_child_supp = col_double(),
##   fin_supp_child_exp = col_double(),
##   fin_supp_lump = col_double(),
##   fin_supp_other = col_double(),
##   is_male = col_double(),
##   main_activity = col_logical()
##   # ... with 1 more columns
## )
```

```
## See spec(...) for full column specifications.
```

```
# choose pertinent variables
poll <- poll %>% select(age, sex, marital_status,self_rated_health,
                        total_children, self_rated_mental_health)

# clean up the data
poll<-poll[!grepl("Don't know", poll$self_rated_mental_health),]

# poll <- head(poll, 1000)
```

```r
nrow(poll)
```

```
## [1] 20545
```

```r
# table(poll$age)
table(poll$sex)
```

```
##
## Female   Male
##  11177   9368
```

```r
table(poll$marital_status)
```

```
##
##              Divorced     Living common-law               Married
##                  1759                  2073                  9481
##             Separated Single, never married               Widowed
##                   641                  4698                  1887
```

```r
table(poll$self_rated_mental_health)
```

```
##
## Excellent      Fair      Good      Poor Very good
##      6080      1296      5813       326      6924
```

```r
model <- multinom(self_rated_mental_health ~ age + sex + marital_status +self_rated_health,
                  data = poll)
```

```
## # weights:  70 (52 variable)
## initial  value 32875.988237
## iter  10 value 23611.931535
## iter  20 value 23441.438461
## iter  30 value 23049.583322
## iter  40 value 22914.075187
## iter  50 value 22822.824806
## iter  60 value 22775.879430
## final  value 22775.875150
## converged
```

```r
summary(model)
```

```
## Call:
## multinom(formula = self_rated_mental_health ~ age + sex + marital_status +
##     self_rated_health, data = poll)
##
## Coefficients:
##           (Intercept)          age    sexMale marital_statusLiving common-law
## Fair        1.0217425 -0.032798399 -0.3224026                     -0.16949105
## Good        1.5874823 -0.011040378 -0.2400994                      0.09958731
```

```
## Poor        1.8561011 -0.044664300 -0.2137392                     -0.62177910
## Very good   0.9413021 -0.005288015 -0.2162900                      0.04000497
##           marital_statusMarried marital_statusSeparated
## Fair                -0.29251444               0.7299606
## Good                -0.01207268               0.4443427
## Poor                -0.92293929               0.5025271
## Very good           -0.03471334               0.1393385
##           marital_statusSingle, never married marital_statusWidowed
## Fair                               0.224221668            0.17670603
## Good                               0.130575216            0.22994484
## Poor                               0.033859427            0.01458883
## Very good                         -0.002477955            0.17862358
##           self_rated_healthExcellent self_rated_healthFair
## Fair                       -3.360131             1.5800364
## Good                       -3.050227             0.1066989
## Poor                       -5.162473            -0.2307758
## Very good                  -1.813127            -0.3809237
##           self_rated_healthGood self_rated_healthPoor
## Fair                 -0.1643287             1.9008484
## Good                  0.1545703             0.1368938
## Poor                 -1.8749204             1.6135608
## Very good            -0.1348050            -0.1311144
##           self_rated_healthVery good
## Fair                       -1.6225589
## Good                       -1.2432679
## Poor                       -3.5250590
## Very good                  0.3213029
##
## Std. Errors:
##           (Intercept)         age    sexMale marital_statusLiving common-law
## Fair        0.4695617 0.002429496 0.06897229                      0.16272558
## Good        0.3101351 0.001459802 0.04227380                      0.09946604
## Poor        0.5122282 0.004525478 0.12350622                      0.28565061
## Very good   0.3325459 0.001338815 0.03893398                      0.09150798
##           marital_statusMarried marital_statusSeparated
## Fair                 0.12358331               0.1975624
## Good                 0.07773531               0.1410396
## Poor                 0.20958816               0.3020509
## Very good            0.07207928               0.1364599
##           marital_statusSingle, never married marital_statusWidowed
## Fair                                0.13684546            0.15750158
## Good                                0.08986249            0.10021966
## Poor                                0.21528134            0.25746406
## Very good                           0.08379589            0.09385189
##           self_rated_healthExcellent self_rated_healthFair
## Fair                       0.4574703             0.4452440
## Good                       0.2951180             0.2967677
## Poor                       0.5107378             0.4420195
## Very good                  0.3189361             0.3248795
##           self_rated_healthGood self_rated_healthPoor
## Fair                  0.4437937             0.4540376
## Good                  0.2921365             0.3112635
## Poor                  0.4407436             0.4439970
## Very good             0.3188215             0.3405639
```

```
##           self_rated_healthVery good
## Fair                       0.4472054
## Good                       0.2925541
## Poor                       0.4643552
## Very good                  0.3178609
##
## Residual Deviance: 45551.75
## AIC: 45655.75
```

```
head(fitted(model))
```

```
##   Excellent        Fair       Good         Poor Very good
## 1 0.6719425 0.014404347 0.09910454 0.0024202042 0.2121284
## 2 0.1951443 0.046536541 0.49248128 0.0062707489 0.2595671
## 3 0.2389554 0.012145612 0.16505036 0.0010447006 0.5828039
## 4 0.2601140 0.007720795 0.14990947 0.0005466639 0.5817091
## 5 0.1437139 0.082680958 0.52334153 0.0175120007 0.2327516
## 6 0.7074104 0.006452120 0.08074126 0.0006178245 0.2047784
```

```
input <- data.frame(self_rated_health = c("Excellent"), age = c(21.5), sex = c("Male"), marital_status =
predict(model, newdata = input, "probs")
```

```
##   Excellent        Fair        Good        Poor   Very good
## 0.658524299 0.028453356 0.107809344 0.007717323 0.197495678
```

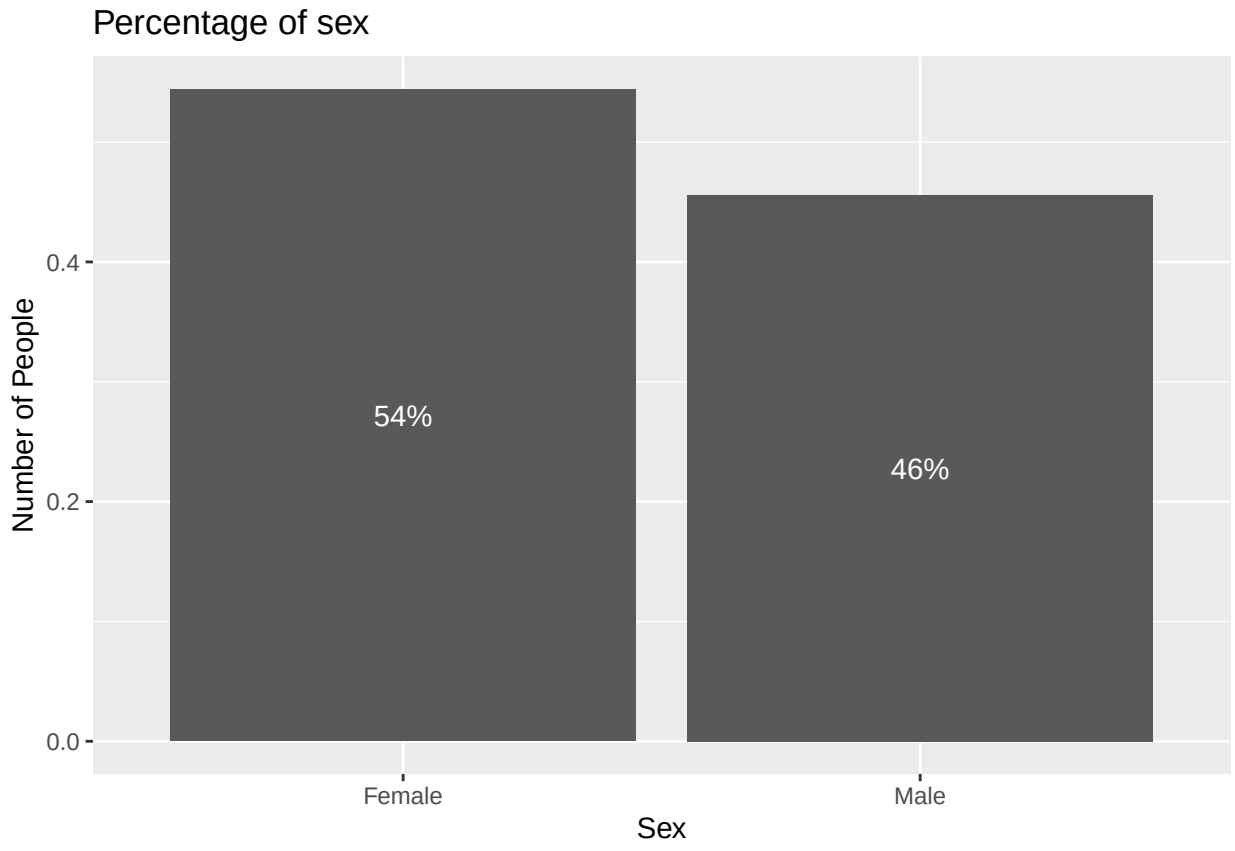
Results

By using the Canadian General Social survey collected in 2017 about Family, we were able to access over 20000 responses. Out of the respondents, about 54 percent of them were female and 46 percent was male (Figure N). Most of the respondents were in their 60's and 70's, and groups in the 30's to 50's making the second most common demographic (Figure N). These primary analysis are fairly consistent with the Canadian demographics, with female response being slightly higher and the age distribution being roughly 10 years older than the Canadian demographic (average Canadian age is 41, average respondent age was 52). This older demographic makes sense as the younger group (<18) are not eligible to respond to surveys there for are not represented. Of the 20,000 respondents, only 54 people did not self report their rating of their own mental health, and therefore these people are removed from our polls as outliers. Furthermore, according to figure N, the responses are heavily skewed towards the positive responses, with 30% of respondents replying with and 'Excellent' rating and 38% replying 'Very good'. 28% rated their own mental health as 'Fair' with the remaining 8 percent split 6 to 2 with regards to 'Fair' and 'Poor' respectively. These results indicate that either a large amount of Canadians feel very good about themselves mental health wise, or people who tend to feel poorly towards their mental health did not respond. As the "I don't know" responses were extremely low, it is possible that individuals with low self reported mental health did not respond to the entire survey. However, we still felt it was interesting to try and find more details in regards to what affects Canadian's self report of their own mental health with respect to different home and personal variables.

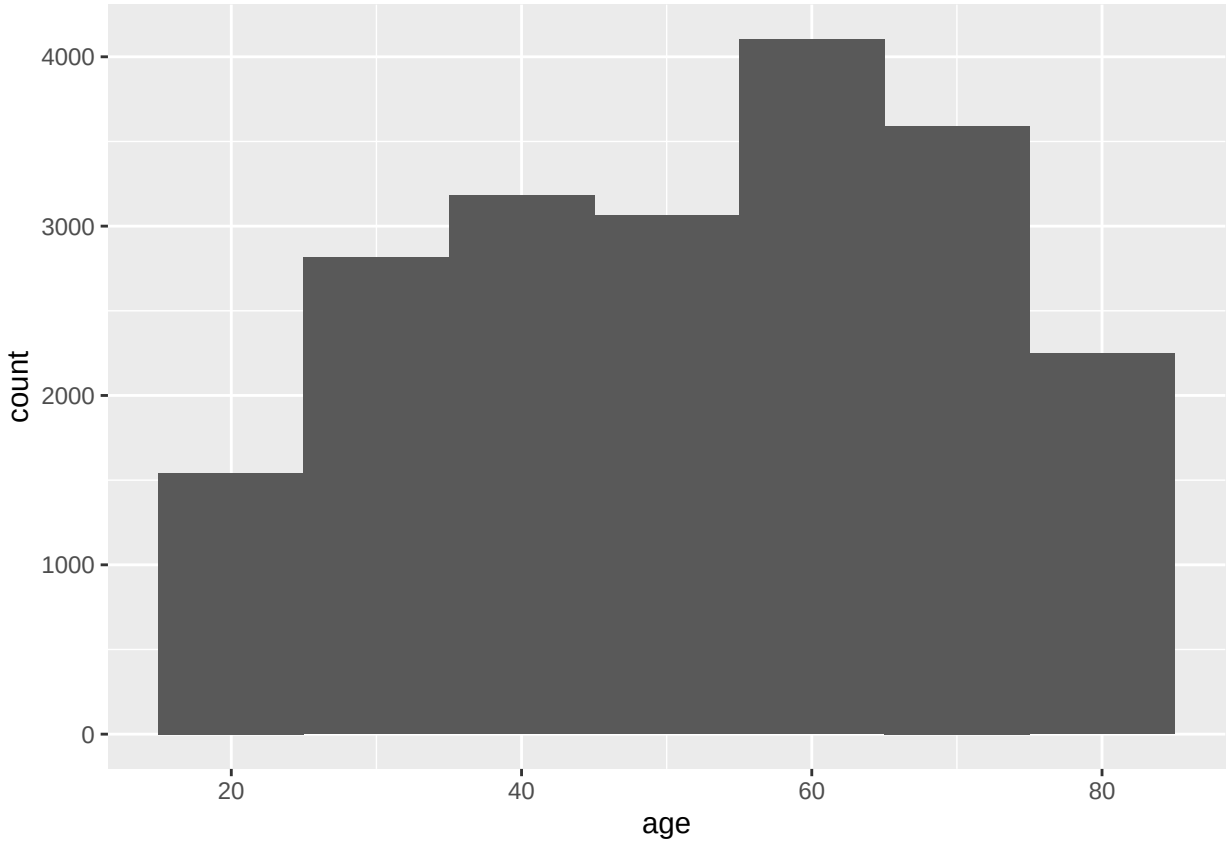
As explained in the previous section, we modeled the a persons self reported mental health against several variables such as age, sex, marital status, and self rated physical health in order to gain more insight on this topic. By analyzing the summary statistic of our model shown in figure N, we can see the various results in relation to our baseline which was chosen to be self_reported_mental_health = "Excellent". By analyzing the summary table, we can notice certain variables and their affect on a persons self rating of their mental health. For example, having a married marital status decreases your log odds having an poor self reported mental health compared to an excellent self assessment by 0.92. However, being married has a much smaller

affect on the odds of having fair mental health self assessment and close to no affect on Good and Very good self reported mental health in comparison to an excellent self assessment. More noticeably, the strongest correlation is found between self rated physical health and self rated mental health. An individual marking themselves as excellent physical health decreases their log chances of being in poor mental health by 5. Variables such as age as almost no affect on a persons mental health rating.

```
total_survey = nrow(poll)
bar_sex <- data.frame(table(poll$sex)) %>%
  rename(
    Sex = Var1,
    Total = Freq
  )
ggplot(bar_sex, aes(x = Sex, y = Total/total_survey)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0(round(Total/total_survey*100), "%")),
            color="white",
            position = position_stack(vjust = 0.5)) +
  ggtitle("Percentage of sex") +
  scale_y_continuous("Number of People")
```



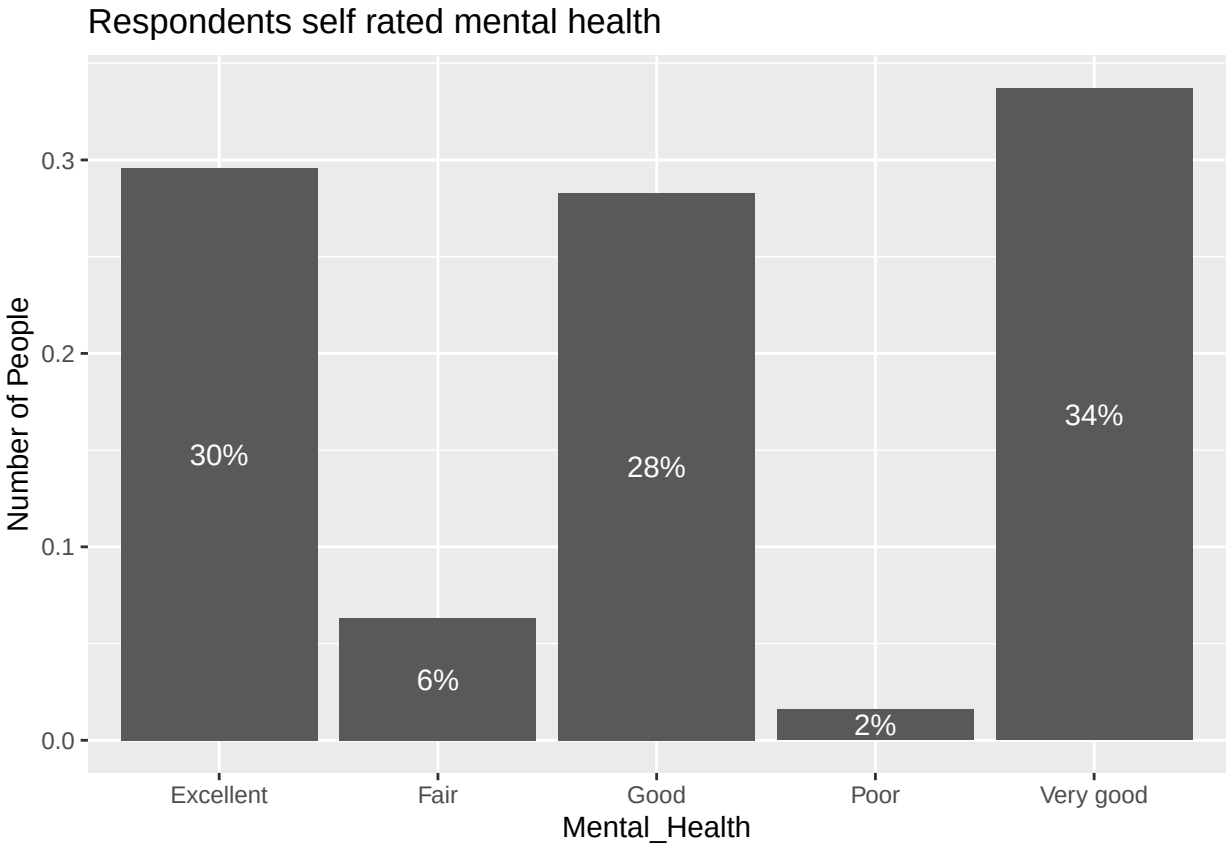

```
ggplot(poll, aes(x=age)) + geom_histogram(position="identity", binwidth = 10)
```




```
mean(poll$age)
```

```
## [1] 52.17379
```

```
total_survey = nrow(poll)
bar_sex <- data.frame(table(poll$self_rated_mental_health)) %>%
  rename(
    Mental_Health = Var1,
    Total = Freq
  )
ggplot(bar_sex, aes(x = Mental_Health, y = Total/total_survey)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0(round(Total/total_survey*100), "%")),
            color="white",
            position = position_stack(vjust = 0.5)) +
  ggtitle("Respondents self rated mental health") +
  scale_y_continuous("Number of People")
```


Respondents self rated mental health
Number of People
0.3
0.2
0.1
0.0
30%
6%
28%
2%
34%
Excellent
Fair
Good
Poor
Very good
Mental_Health