

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)

poll$has_children <- as.logical(poll$total_children)
poll
```

```
## # A tibble: 20,602 x 7
##   age sex marital_status self Rated_heal~ total_children self Rated_ment~
```

```
##      <dbl> <chr> <chr>          <chr>          <dbl> <chr>
## 1  52.7 Fema~ Single, never~ Excellent      1 Excellent
## 2  51.1 Male  Married          Good            5 Good
## 3  63.6 Fema~ Married          Very good       5 Good
## 4  80   Fema~ Married          Very good       1 Very good
## 5  28   Male  Living common~ Good            0 Good
## 6  63   Fema~ Married          Excellent        2 Very good
## 7  58.8 Fema~ Single, never~ Poor            2 Poor
## 8  80   Fema~ Divorced         Good            7 Very good
## 9  63.8 Fema~ Single, never~ Very good        0 Very good
## 10 25.2 Male  Single, never~ Poor            1 Poor
## # ... with 20,592 more rows, and 1 more variable: has_children <lgl>
```

```
original_poll <- poll

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

# poll <- head(poll, 1000)

poll <- cleaned_poll
```

```
nrow(poll)
```

```
## [1] 20439
```

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

```
##
## Female    Male
## 11122     9317
```

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

```
##
##           Divorced      Living common-law      Married
##           1752           2064           9424
##      Separated Single, never married      Widowed
##           639           4675           1880
```

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

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

```
model <- multinom(selfRatedMentalHealth ~ age + sex + maritalStatus +
  selfRatedHealth + hasChildren,
  data = poll)
```

```
## # weights: 75 (56 variable)
## initial value 32845.408917
## iter 10 value 23591.252534
## iter 20 value 23408.761919
## iter 30 value 23049.845274
## iter 40 value 22904.966135
## iter 50 value 22815.803566
## iter 60 value 22751.210835
## final value 22749.489605
## converged
```

```
summary(model)
```

```
## Call:
## multinom(formula = selfRatedMentalHealth ~ age + sex + maritalStatus +
##   selfRatedHealth + hasChildren, data = poll)
##
## Coefficients:
##      (Intercept)      age      sexMale maritalStatusLiving common-law
## Fair      0.9655496 -0.033433317 -0.3225311      -0.15773997
## Good      1.4709695 -0.011992596 -0.2348329      0.12043339
## Poor      1.7698564 -0.045244562 -0.2166419     -0.64404262
## Very good  0.8599432 -0.005972975 -0.2129676      0.05428992
##      maritalStatusMarried maritalStatusSeparated
## Fair      -0.29823014      0.7226140
## Good      -0.01965182      0.4335222
## Poor      -0.93029686      0.4936593
## Very good -0.04197558      0.1294732
##      maritalStatusSingle, never married maritalStatusWidowed
## Fair      0.28122033      0.18059391
## Good      0.23533960      0.23632619
## Poor      0.11148165      0.01617863
## Very good 0.07141802      0.17988087
##      selfRatedHealthExcellent selfRatedHealthFair
## Fair      -3.355777      1.5859292
## Good      -3.040435      0.1193757
## Poor      -5.151364     -0.2217835
## Very good -1.807255     -0.3743900
##      selfRatedHealthGood selfRatedHealthPoor
## Fair      -0.1586458      1.9079201
## Good      0.1641493      0.1467143
## Poor      -1.8650569      1.6118807
## Very good -0.1273074     -0.1239058
##      selfRatedHealthVery good hasChildrenTRUE
## Fair      -1.6164753      0.1034934
## Good      -1.2319557      0.1867869
## Poor      -3.5124012      0.1400445
## Very good 0.3302639      0.1342273
```

```
##
## Std. Errors:
##      (Intercept)      age      sexMale marital_statusLiving common-law
## Fair      0.4715301 0.002489899 0.06917301      0.16295948
## Good      0.3118185 0.001493053 0.04235122      0.09972757
## Poor      0.5175040 0.004639139 0.12413768      0.28918291
## Very good 0.3338051 0.001369020 0.03898844      0.09174982
##      marital_statusMarried marital_statusSeparated
## Fair      0.12384266      0.1977684
## Good      0.07790245      0.1411761
## Poor      0.20994713      0.3020930
## Very good 0.07219280      0.1365781
##      marital_statusSingle, never married marital_statusWidowed
## Fair      0.14436394      0.15754685
## Good      0.09519607      0.10029466
## Poor      0.22755558      0.25748347
## Very good 0.08867631      0.09390913
##      selfRated_healthExcellent selfRated_healthFair
## Fair      0.4575447      0.4453045
## Good      0.2952466      0.2968839
## Poor      0.5107868      0.4420480
## Very good 0.3190166      0.3249583
##      selfRated_healthGood selfRated_healthPoor
## Fair      0.4438620      0.4541042
## Good      0.2922608      0.3113751
## Poor      0.4407716      0.4441559
## Very good 0.3188949      0.3406281
##      selfRated_healthVery good has_childrenTRUE
## Fair      0.4472629      0.09363596
## Good      0.2926807      0.05899384
## Poor      0.4643647      0.16328123
## Very good 0.3179372      0.05390949
##
## Residual Deviance: 45498.98
## AIC: 45610.98
```

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

```
##      Excellent      Fair      Good      Poor Very good
## 1 0.6478455 0.014972313 0.10932262 0.0026098217 0.2252498
## 2 0.1911508 0.046258991 0.49653883 0.0062900594 0.2597614
## 3 0.2368721 0.012127783 0.16582617 0.0010588348 0.5841151
## 4 0.2602378 0.007700366 0.14965564 0.0005539027 0.5818523
## 5 0.1556600 0.084620285 0.50907776 0.0168599571 0.2337820
## 6 0.7054501 0.006472567 0.08153008 0.0006291839 0.2059181
```

```
input <- data.frame(selfRated_health = c("Excellent"), age = c(21.5), sex = c("Male"), marital_status = c("Single, never married"),
predict(model, newdata = input, "probs")
```

```
##      Excellent      Fair      Good      Poor      Very good
## 0.661652268 0.028343157 0.106478230 0.007654654 0.195871692
```

Results

By using the Canadian General Social survey collected in 2017 about Family, we were able to access over 20000 responses. Of the respondents, almost 50 percent of females rated themselves as ‘Very Good’ with respect to mental health, and over 45 % of male’s reported ‘Very Good’. The response rates between males and females were very similar overall, with slightly more females rating themselves good rather than excellent in comparison to the male respondents. Similarly, with respect to physical health by sex, the response distribution was almost identical comparing males to females. Similarly the age distribution is practically identical, with more Females responding than males. As with before, our sex to marital status distribution (Figure N) are very similar to each other, although more Males reported to be MArried and NEver married while significantly more females reported to be widowed than males did. Most people rated their own mental health as Very Good, and very few people rated their own mental health as poor. However, much less people rated their own mental health as poor if they were Married or living in a common-law relationship, and divorcees and separated had much more rating themselves as poor mental health (Figure N). Most interestingly, there seems to be the largest variance in response of self rated mental health with respect to self rated health. 70% of respondents who rated their physical health as Excellent also rated their mental health as excellent, with the trend staying consistent for almost all the ratings. About half of the people who self-assessed their physical health as Poor listed their mental health as either Good or Fair and only less than 20% listed their mental as Poor. Overall our results so fairly even and expected distributions between various demographics and there does not seem to be strong correlation between various variables we used.

As explained in the previous section, we modeled Canadians 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 as mentioned previously. An individual marking themselves as excellent physical health decreases their log chances of being in poor mental health by 5. Variables such as age and number of kids has limited effect on a persons mental health rating.

Discussion

Mental health is an increasingly popular topic, and there are many myths and facts regarding such a topic. Mental health is known to be a complicated issue that about half the population experience by the time theyre 40 (CMHA, 2013). By analyzing our results, we look through multiple variables that we as young adults viewed to be instrumental to our own mental health, and see how that effects the general Canadian public’s impression on their own mental health. With this in mind, we knew that the stigma of mental health would skew our results but certain relationships (or lack thereof) was still interesting to see. It is important to know that this discussion focuses on how a person reports their mental health, not an individuals actual mental health, and that these variables are not full representations of a persons self rating of their mental health rather just a glimpse into possible correlations.

There is a common saying “money can’t buy happiness”, and depending on who you ask, people have varying ideas of whether this is true or not. Due to known issues with people reporting their own income, we attempted to use age and gender as indicators of income (Elkins, 2019) in an attempt to possibly get a more accurate result. A study by McBride (2010) showed that peoples view on happiness is related to income, to a certain degree, and as such we expected to see people increasing in age to also increase the likelihood

of reporting their mental health is good or excellent, and also males reporting at a better self rating in comparison to females. As shown in Figure N, these 2 variables have limited roll on any individuals self assessment of mental health. Interestingly, being male lowers the log likelihood of an individual reporting any result other than excellent and there are multiple reasons why this trend would be present. First and foremost, the well discussed wage gap across sex in Canada is a possibility, however more likely is the stigma towards mental health. Previous experience at a non-profit organization dedicated towards tackling stigma against certain individuals experiencing mental health issues showed that males are much less likely to report experiencing mental health issues. Due to the stigma that males appear “weak” should they experience mental health issues, males are much more likely to suppress symptoms and asking for help, often convincing themselves they are fine even when dealing with extreme mental health issues. However age has almost no role towards an individuals self rating towards their mental health. This didn’t track with our expectations of older individuals reporting better mental health.

Another factor we thought we be interesting to see was a presence of family, more specifically, whether marital status and number of children would have effect on an individuals rating of their own mental health. The presence of an attached partner (married or common-law) decreases an individuals likelihood of reporting having poor or fair mental health and mixed results regarding good or very good mental health. Individuals who reported themselves to be single and never married has a slightly higher chance of also reporting their mental else as Fair rather than excellent. Interestingly, individuals who has children has a slightly higher chance of reporting all levels of mental health instead of excellent, despite practically every anecdotal evidence in existence suggesting otherwise. It is evident that the precence of a partner and children are not clear indicators of an individuals report of their mental health.

The last variable we decided to use in our model was self assessed physical health. There was a strong correlation between a persons report of their own physical health relative to their rating of their own mental health. Individuals extraordinarily less likely to report having poor mental health should they report themselves as having good physical health, and individuals reporting poor physical health. This is common knowledge to many, and even the ancient Roman’s has a proverb translating to “healthy mind in a healthy body”. The study on benefits of excercise on mental health have been published as early as 1985 and new studies continue to be released even now. Studies such as the one published by Mikkelsen et al. (2017) indicate that consistent excersice and a healthy body does indeed improve a multitude of mental health issues such as anxiety, depression and stress. This survey’s results does match with expecations from the studies, however it is noted that the response distribution for physical health is very similar to the distribution for mental health so both the responses maybe subject to response bias.

Overall, these conclusions potentially help contribute to finding what can cause an individual to view their mental health in a certain way. Many values that are deemed to be important to a strong mental health such as family and reaching certain goals as people age, or being married or having children did not significantly affect our models prediction on mental health. The causes of mental health issues are random and unpredictable, and this survey further shows that. The only variable that was a strong predictor of a person’s rating of their own mental health was also subject to response bias and public stigma (physical health). Being married did increase ones chances of reporting positively about one’s own mental health, but that is likely due to having a support group rather than the relationship itself, meaning a poor relationship can do harm as well and most other variables were inconclusive at best. This survey showed that hitting certain bench marks (earn a certain income, get married, get a certain grade etc) while rewarding would not contribute to ones mental health on the long run. Further more, the disproportionate amount of individuals reported their mental health as very good or excellent can mean one of two things. Optimistically, it means Canadians are doing great, both physically and mentally; realistically, it suggests that despite recent improvements to advocate for mental health support, the stigmas and lack of willingness to admit that an individual needs mental health support is still lacking. If people are not willing to answer honestly in an anynomous survey, it becomes hard to imagine the same individual seeking help personally. Our low accuracy in prediction further enforces that mental health issues are unpredictable, and unlike many physical diseases, mental health issues are not finite and can be hard to diagnose if the suffering individual refuses to acknowledge the need for help.

Weaknesses and future work

Our survey had many weaknesses, such as the fact that we chose to analyze a variable that is very hard to predict, prone to bias and relative to the respondents experience towards mental health. Another large weakness was the data excluded youths aged 15 and younger. This effects our results significantly because youths mental illness is a prominent issue as half of all mental health issues begin by the age of 15 and the lack of this data does not allow us to represent mental health thoroughly. This was a weakness we were aware of and knew it would affect our outcome. However, we were still curious to see how the responses would go and determine if the results can lead to insights. We also were not able to spend time to include all the variables we wanted. Variables such as income, detailed family situations and living situations were things we planned on adding to our model but due to time and covariance concerns we were not able to integrate them into our model. This would be part of our future plans in order to build a more encompassing model. Furthermore, we hope to gain less bias information on an individuals mental health by utilizing mental health surveys dedicated to measuring an individuals mental health with less bias.

<https://cmha.ca/fast-facts-about-mental-illness>

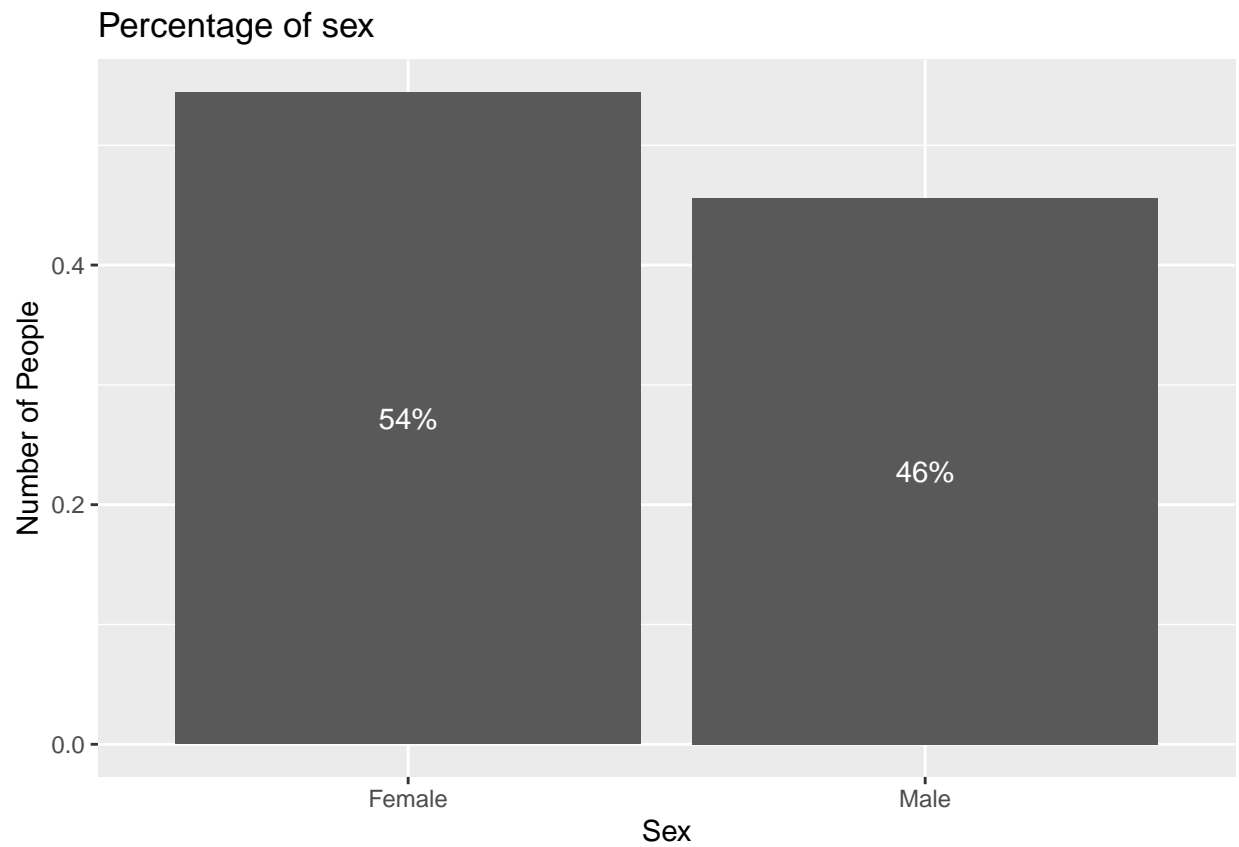
<https://www.cnn.com/2019/04/02/heres-how-much-men-and-women-earn-at-every-age.html>

<https://www.sciencedirect.com/science/article/abs/pii/S0167268110000399> <https://www.sciencedirect.com/science/article/pii/S0378512217308563>

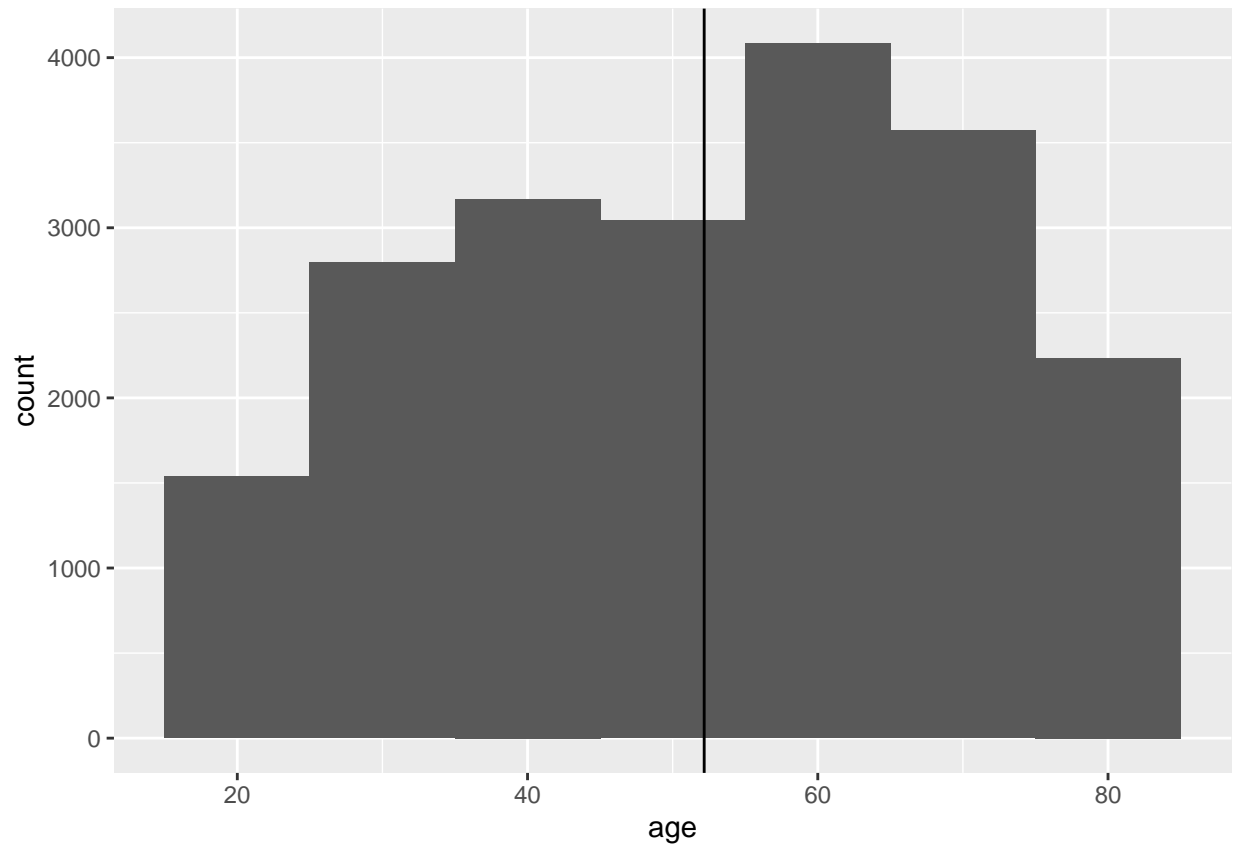
<https://www.camh.ca/en/camh-news-and-stories/transforming-mental-health-for-children-and-youth>

```
#total number of respondents measured
total_survey = nrow(poll)

#
bar_sex <- data.frame(table(poll$sex)) %>%
  rename(
    Sex = Var1,
    Total = Freq
  )
# graph people to sex
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")
```

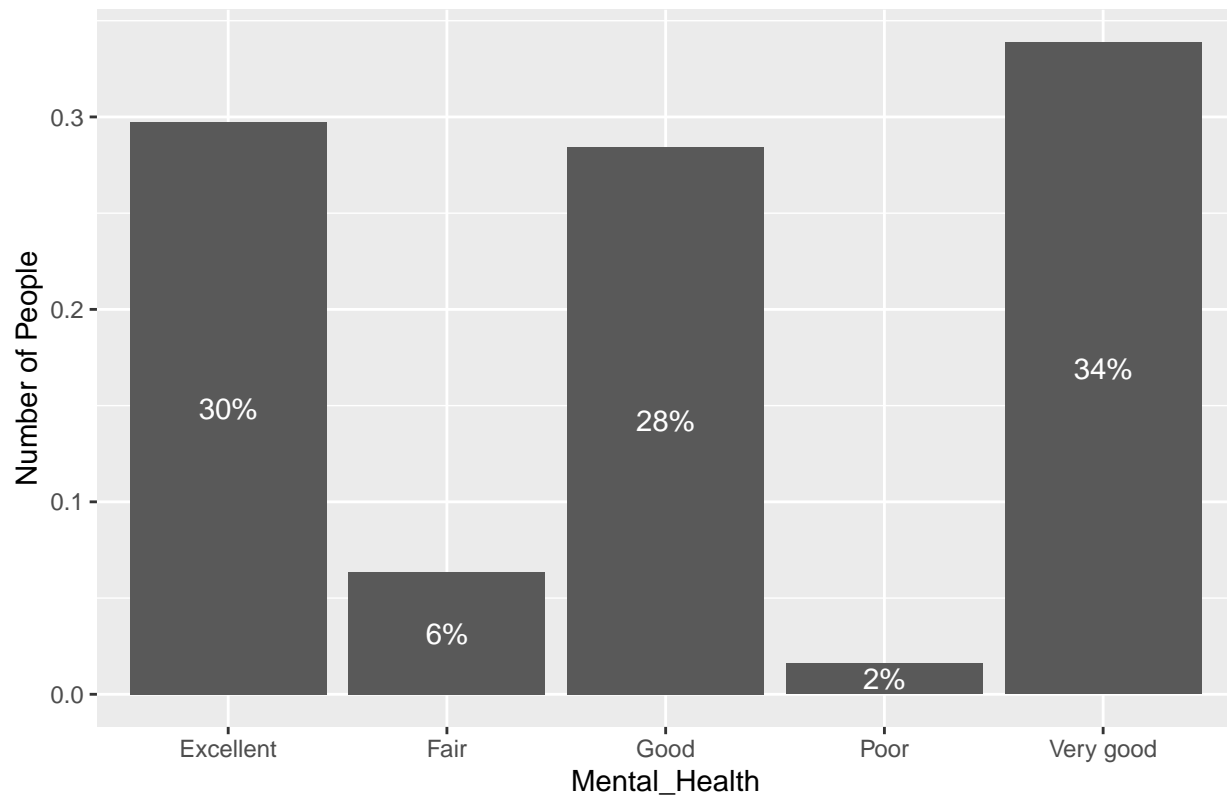


```
ggplot(poll, aes(x=age)) + geom_histogram(position="identity", binwidth = 10) +  
  geom_vline(xintercept = mean(poll$age))
```

```
total_survey = nrow(poll)
# filter and group results by self rated mental health
bar_sex <- data.frame(table(poll$self_rated_mental_health)) %>%
  rename(
    Mental_Health = Var1,
    Total = Freq
  )
# Distributions of people's self ratings of mental health
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



```
grouped_bar_chart <- function(data, xlab, ylab, title){
  ggplot(data,
    aes(y=grouped, x=value, fill=category, label=category))+
  geom_bar(stat="identity", position="dodge", width=0.8) +
  scale_x_continuous(ylab) +
  scale_y_discrete(xlab) +
  ggtitle(title)
}
```

```
# graph age vs distribution of self rated mental health

#data parsing into format I want for chart
poll<- original_poll
grouped <- poll %>%
  count(sex, self Rated mental health) %>%
  rename(
    count=n
  )
#total number of votes
total_vote <- aggregate(grouped$count,
  by=list(sex=grouped$sex), FUN=sum)

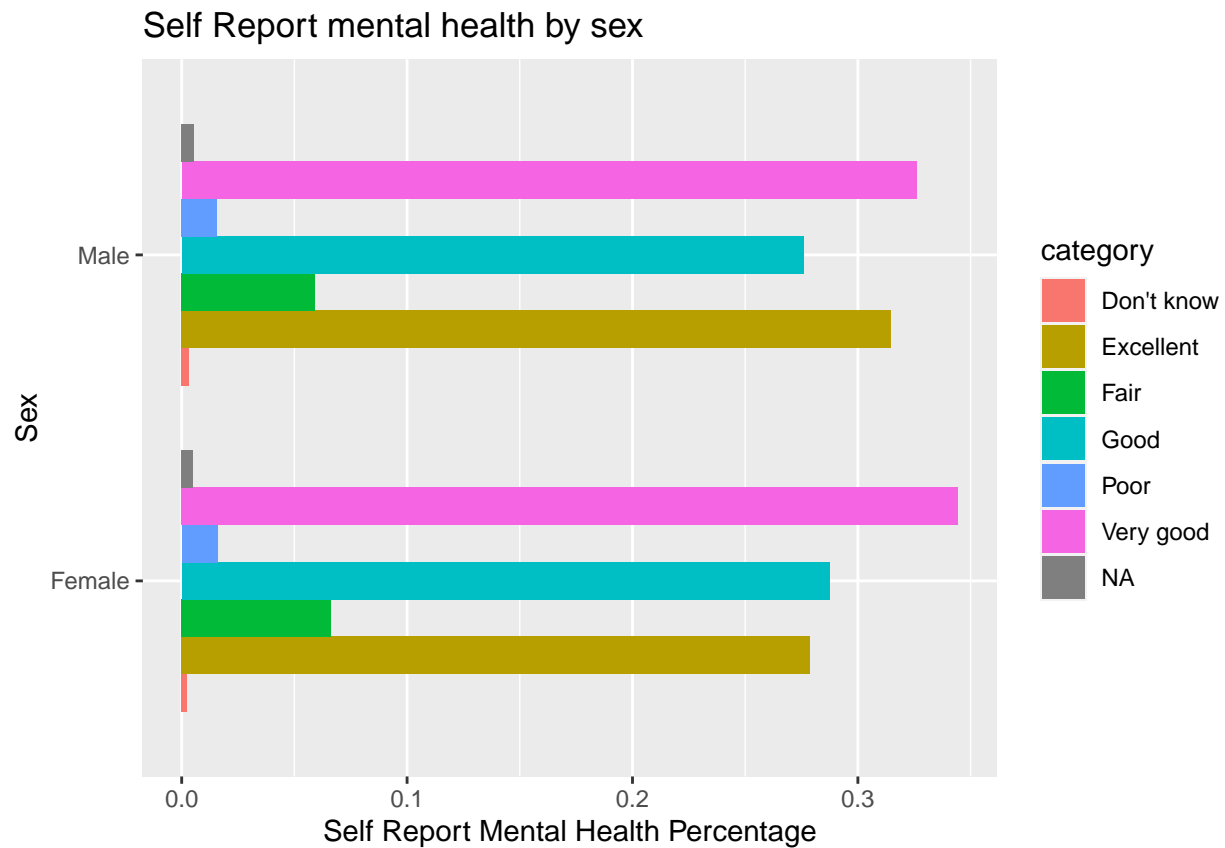
#process data into desired graphing format
grouped<-left_join(grouped, total_vote, by="sex") %>%
  rename(total=x) %>%
  mutate(value=count/total) %>%
```

```

rename(grouped=sex, category=self Rated mental health)

grouped_bar_chart(grouped, "Sex", "Self Report Mental Health Percentage",
                  "Self Report mental health by sex")

```



```

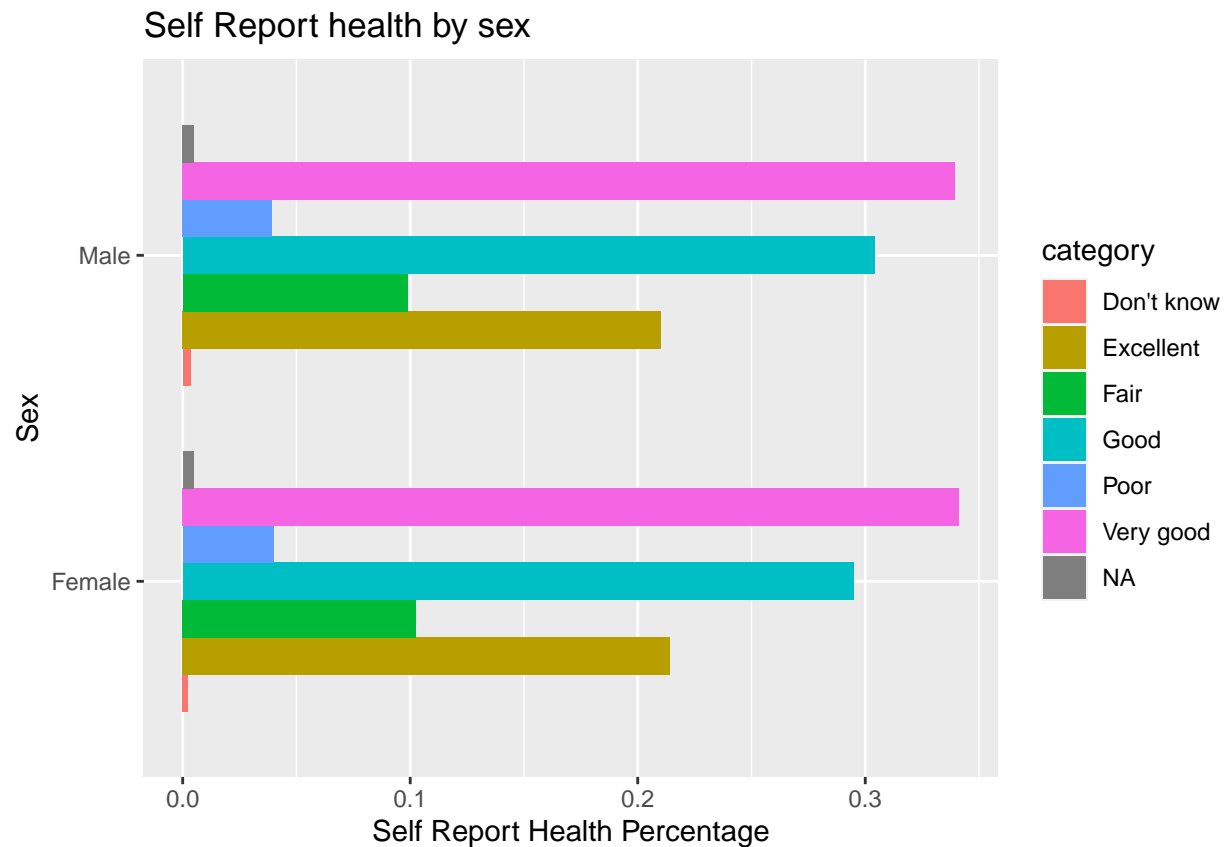
# graph age vs distribution of self rated mental health
poll<- original_poll

#data parsing into format I want for chart
grouped <- poll %>%
  count(sex, self Rated health) %>%
  rename(
    count=n
  )

#process data into desired graphing format
grouped<-left_join(grouped, total_vote, by="sex") %>%
  rename(total=x) %>%
  mutate(value=count/total) %>%
  rename(grouped=sex, category=self Rated health)

grouped_bar_chart(grouped, "Sex", "Self Report Health Percentage",
                  "Self Report health by sex")

```

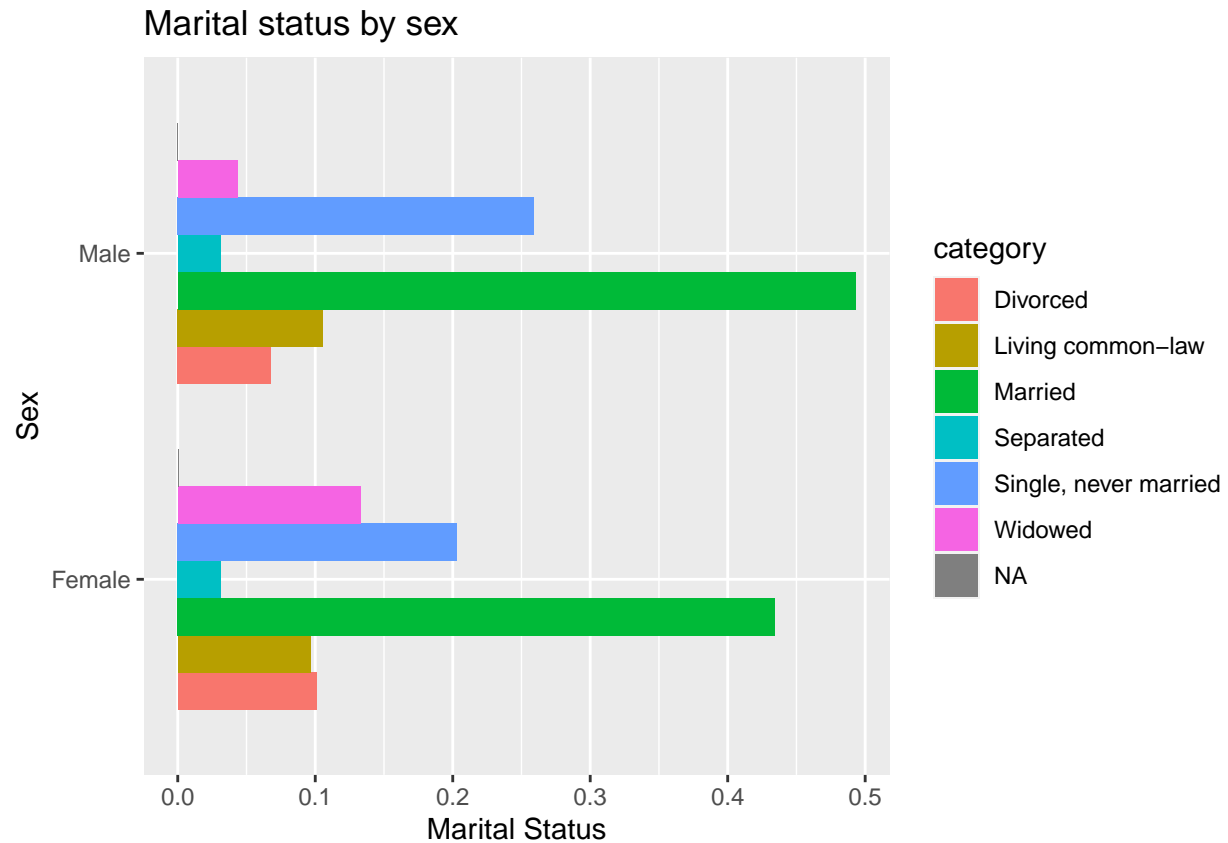


```
# graph age vs distribution of self rated mental health
poll<- original_poll

#data parsing into format I want for chart
grouped <- poll %>%
  count(sex, marital_status) %>%
  rename(
    count=n
  )

#process data into desired graphing format
grouped<-left_join(grouped, total_vote, by="sex") %>%
  rename(total=x) %>%
  mutate(value=count/total) %>%
  rename(grouped=sex, category=marital_status)

grouped_bar_chart(grouped, "Sex", "Marital Status",
  "Marital status by sex")
```

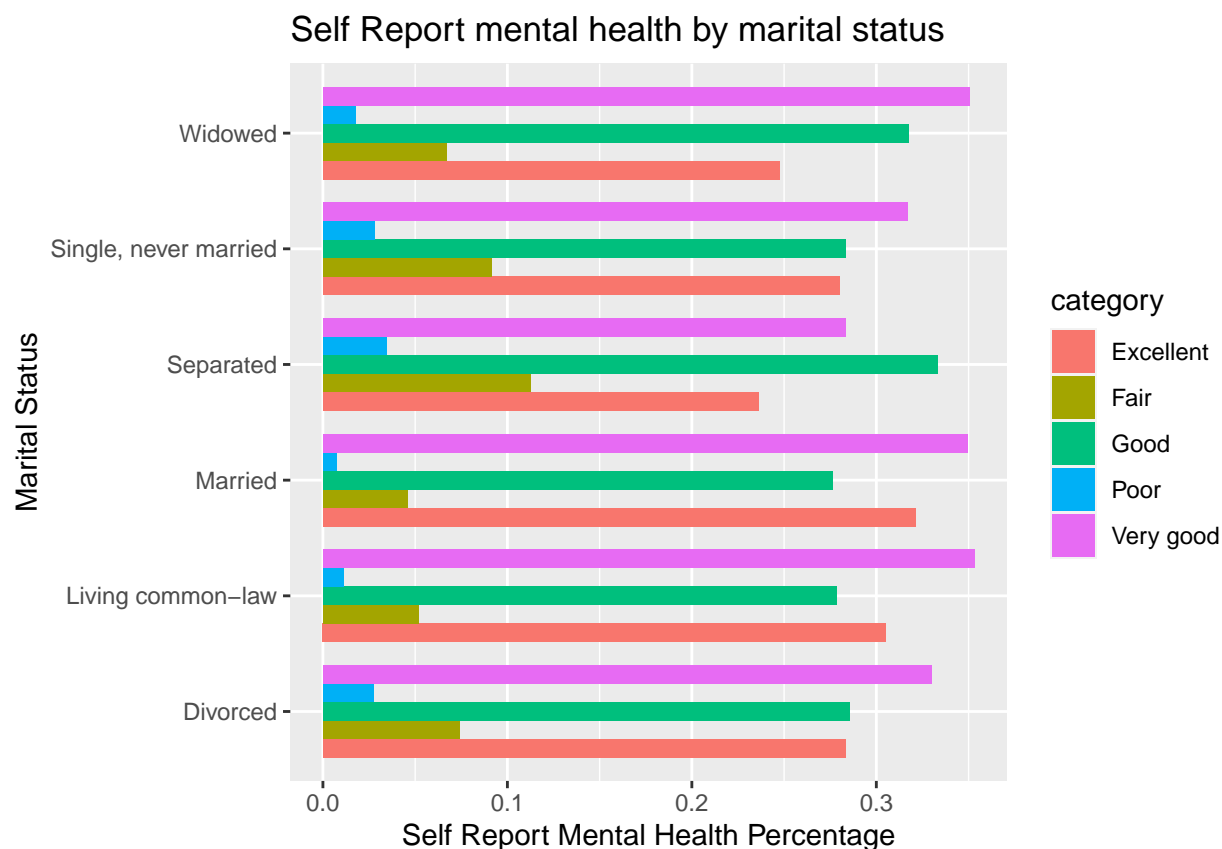


```
# graph age vs distribution of self rated mental health
#data parsing into format I want for chart
poll<- cleaned_poll[complete.cases(cleaned_poll$marital_status),]
grouped <- poll %>%
  count(marital_status, self_rated_mental_health) %>%
  rename(
    count=n
  )

#total number of votes
total_vote <- aggregate(grouped$count,
                        by=list(marital_status=grouped$marital_status), FUN=sum)

#process data into desired graphing format
grouped<-left_join(grouped, total_vote, by="marital_status") %>%
  rename(total=x) %>%
  mutate(value=count/total) %>%
  rename(grouped=marital_status, category=self_rated_mental_health)

grouped_bar_chart(grouped, "Marital Status",
                  "Self Report Mental Health Percentage",
                  "Self Report mental health by marital status"
                )
```



```
poll<- original_poll[complete.cases(original_poll$marital_status),]
```

```
grouped <- poll
# grouped <- poll %>%
#   count(self Rated health, self Rated mental health) %>%
#   rename(
#     count=n
#   )
grouped
```

```
## # A tibble: 20,595 x 7
##   age sex marital_status self Rated heal~ total_children self Rated ment~
##   <dbl> <chr> <chr> <chr> <dbl> <chr>
## 1 52.7 Fema~ Single, never~ Excellent 1 Excellent
## 2 51.1 Male Married Good 5 Good
## 3 63.6 Fema~ Married Very good 5 Good
## 4 80 Fema~ Married Very good 1 Very good
## 5 28 Male Living common~ Good 0 Good
## 6 63 Fema~ Married Excellent 2 Very good
## 7 58.8 Fema~ Single, never~ Poor 2 Poor
## 8 80 Fema~ Divorced Good 7 Very good
## 9 63.8 Fema~ Single, never~ Very good 0 Very good
## 10 25.2 Male Single, never~ Poor 1 Poor
## # ... with 20,585 more rows, and 1 more variable: has_children <lgl>
```

```

#grouped <- cor(grouped$count, use = "everything")

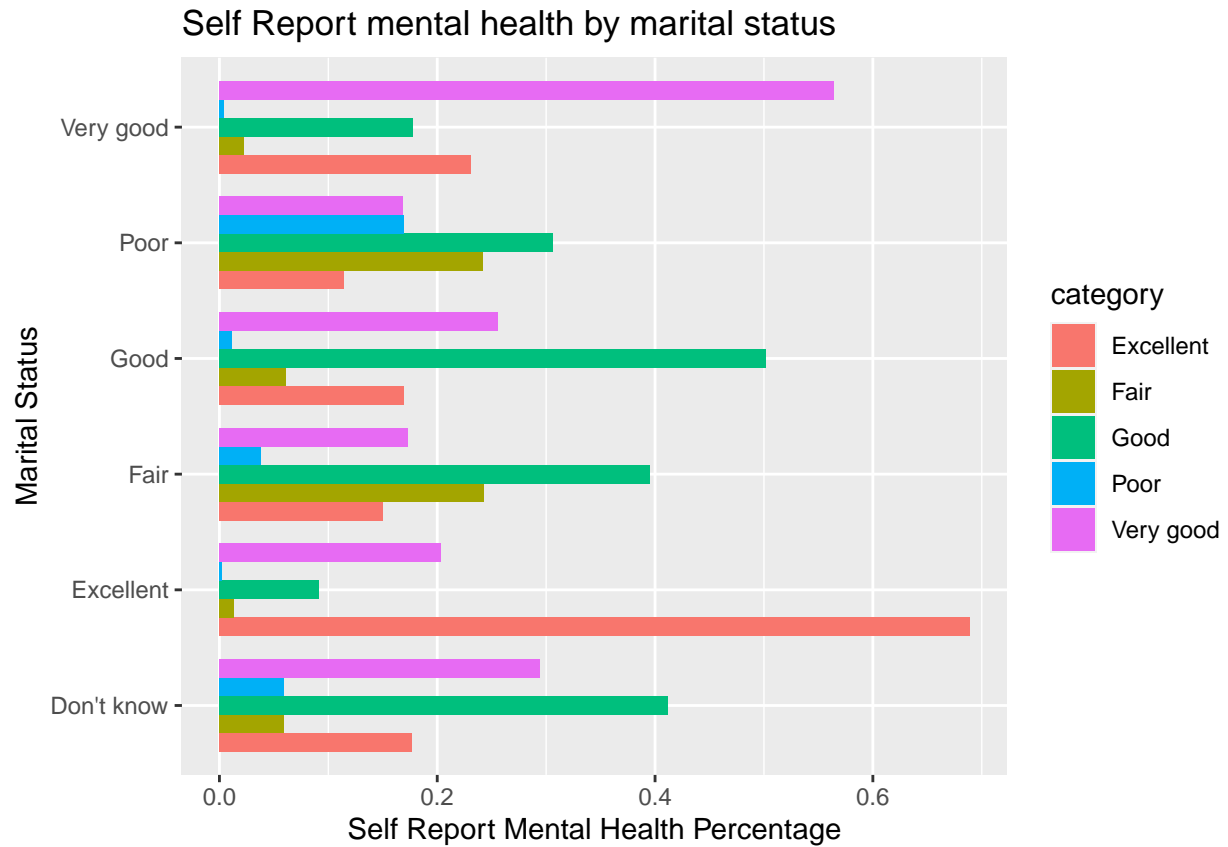
#ggcorrplot(grouped)

# graph age vs distribution of self rated mental health
#data parsing into format I want for chart
poll<- cleaned_poll[complete.cases(cleaned_poll$self_rated_health),]
grouped <- poll %>%
  count(self_rated_health, self_rated_mental_health) %>%
  rename(
    count=n
  )
#total number of votes
total_vote <- aggregate(grouped$count,
                        by=list(self_rated_health=grouped$self_rated_health), FUN=sum)

#process data into desired graphing format
grouped<-left_join(grouped, total_vote, by="self_rated_health") %>%
  rename(total=x) %>%
  mutate(value=count/total) %>%
  rename(grouped=self_rated_health, category=self_rated_mental_health)

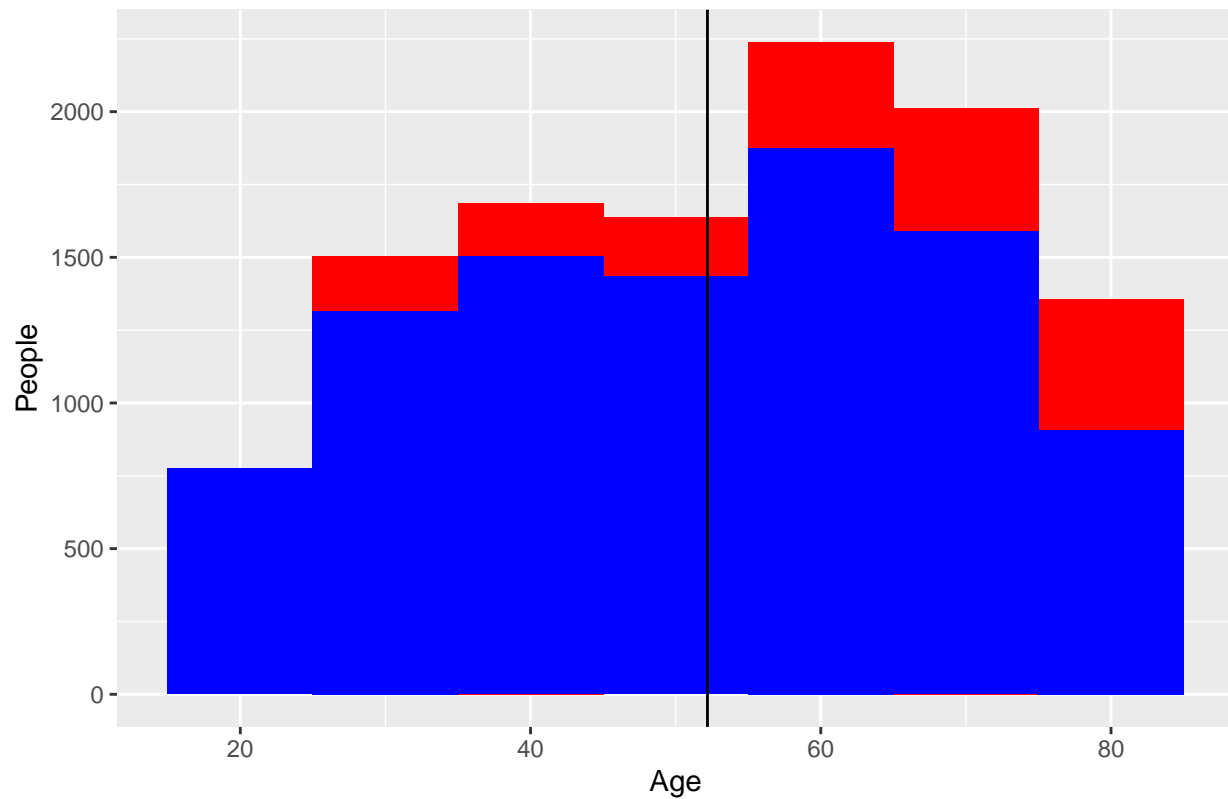
grouped_bar_chart(grouped, "Marital Status",
                  "Self Report Mental Health Percentage",
                  "Self Report mental health by marital status"
                  )

```



```
# graph age vs distribution of self rated mental health
poll<- original_poll
ggplot(poll, aes(x=age)) +
  geom_histogram(data = subset(poll, sex == 'Female'), position="identity",
    binwidth = 10, fill="red") +
  geom_histogram(data = subset(poll, sex == 'Male'), position="identity",
    binwidth = 10, fill="blue") +
  geom_vline(xintercept = mean(poll$age)) +
  xlab("Age") +
  ylab("People")+
  ggtitle("Age split by Sex")
```


Age split by Sex



poll

```
## # A tibble: 20,602 x 7
##   age sex marital_status self Rated heal~ total_children self Rated ment~
##   <dbl> <chr> <chr> <chr> <dbl> <chr>
## 1 52.7 Fema~ Single, never~ Excellent 1 Excellent
## 2 51.1 Male Married Good 5 Good
## 3 63.6 Fema~ Married Very good 5 Good
## 4 80 Fema~ Married Very good 1 Very good
## 5 28 Male Living common~ Good 0 Good
## 6 63 Fema~ Married Excellent 2 Very good
## 7 58.8 Fema~ Single, never~ Poor 2 Poor
## 8 80 Fema~ Divorced Good 7 Very good
## 9 63.8 Fema~ Single, never~ Very good 0 Very good
## 10 25.2 Male Single, never~ Poor 1 Poor
## # ... with 20,592 more rows, and 1 more variable: has_children <lgl>
```