

Exploring the BRFSS data

Setup

Load packages

```
library(tidyverse)
library(plotly)
library(gt)
```

Load data

Make sure your data and R Markdown files are in the same directory. When loaded your data file will be called `brfss2013` . Delete this note when before you submit your work.

```
load("brfss2013.RData")
```

Part 1: Data

The Behavioral Risk Factor Surveillance System (BRFSS) randomly samples non-institutionalized adults from the US population, including US territories. The data is collected from telephone based surveys where interviewers survey a randomly selected adult in a household. Interviewers ask participants in the survey questions pertaining to their health status, health care access, quality of life, sleep, weight, disease status, alcohol consumption, tobacco use, diet, exercise, income, race, and behavioral risk factors like seat belt use.

Part 2: Research questions

Research question 1:

Is there a generational difference among drivers wearing seat belts? The first seat belt law was introduced in 1968 which required vehicle manufactures to include them in the vehicles they made. Most states did not introduce mandates for drivers until the mid to late 80’s.

Why I want to know: for as long as I can remember my dad has not worn a seat belt. I’ve seen him stick cardboard in the belt clip to keep his truck from beeping. He still refuses to wear his seat belt.
My hypothesis is that older generations are more resistant to wearing a seat belt because they had to adjust to the regulations, while younger generations may have normalized the use.

Research question 2:

How much time should I spend exercising and sleeping in order to maximize my chances of being healthy?
I want to know what I can do to improve my chances of being healthy.

Why I want to know: I want to know how much time I should exercise to improve my general health.

Research question 3:

Do veterans get the emotional support they need?
Why I want to know: Veteran suicide is a troubling trend. I suspect that veterans do not receive the emotional support they need, especially when returning to civilian life after being deployed. I want to help raise awareness about this issue.

Part 3: Exploratory data analysis

NOTE: Insert code chunks as needed by clicking on the “Insert a new code chunk” button (green button with orange arrow) above. Make sure that your code is visible in the project you submit. Delete this note when before you submit your work.

Research question 1:

To answer this question we need to define generations and create a new variable named generation According to wikipedia the generations are defined as follows:
Baby Boomers: 1946 - 1964
Generation X: 1965 - 1980
Millenials: 1981 - 1996

Since the dataset we’re working with is data from 2013 we’ll focus on these three categories.

In 2013: baby boomers: 49 - 67 Gen X: 33 - 48 Millenials: 17 - 32

Since the age levels dont’ correspond exactly to these definitions we’ll have to recategorize them to the following levels: Millenails: 18 - 29 Gen X: 30 - 49 Baby Boomers: 50 or older

```
question_1 <- brfss2013 %>%
  select(X_ageg5yr, seatbelt) %>%
  mutate(generation = case_when(
    X_ageg5yr == "Age 18 to 24" ~ "Millenial",
    X_ageg5yr == "Age 24 to 29" ~ "Millenial",
    X_ageg5yr == "Age 30 to 34" ~ "Gen X",
    X_ageg5yr == "Age 35 to 39" ~ "Gen X",
    X_ageg5yr == "Age 40 to 44" ~ "Gen X",
    X_ageg5yr == "Age 45 to 49" ~ "Gen X",
    X_ageg5yr == "Age 50 to 54" ~ "Boomer",
    X_ageg5yr == "Age 55 to 59" ~ "Boomer",
    X_ageg5yr == "Age 60 to 64" ~ "Boomer",
    X_ageg5yr == "Age 65 to 69" ~ "Boomer",
    X_ageg5yr == "Age 70 to 74" ~ "Boomer",
    X_ageg5yr == "Age 75 to 79" ~ "Boomer",
    X_ageg5yr == "Age 80 or older" ~ "Boomer",
  )) %>% select(-X_ageg5yr) %>%
  arrange(generation, seatbelt) %>%
  filter(seatbelt != "Never drive or ride in a car")

# Split these into seperate datasets
boomer_drivers <- question_1 %>% filter(generation == "Boomer")
genx_drivers <- question_1 %>% filter(generation == "Gen X")
millenial_drivers <- question_1 %>% filter(generation == "Millenial")
print(question_1$seatbelt %>% levels())
```

```
## [1] "Always"           "Nearly always"
## [3] "Sometimes"        "Seldom"
## [5] "Never"            "Never drive or ride in a car"
```

```
# Now we need to calculate the percentage of each group of drivers based on seatbelt usage
boomer_always <- nrow(boomer_drivers %>% filter(seatbelt == "Always")) / nrow(boomer_drivers)
boomer_near_always <- nrow(boomer_drivers %>% filter(seatbelt == "Nearly always")) / nrow(boomer_drivers)
boomer_sometimes <- nrow(boomer_drivers %>% filter(seatbelt == "Sometimes")) / nrow(boomer_drivers)
boomer_seldom <-nrow(boomer_drivers %>% filter(seatbelt == "Seldom")) / nrow(boomer_drivers)
boomer_never <- nrow(boomer_drivers %>% filter(seatbelt == "Never")) / nrow(boomer_drivers) # Hey Dad!

genx_always <- nrow(genx_drivers %>% filter(seatbelt == "Always")) / nrow(genx_drivers)
genx_near_always <- nrow(genx_drivers %>% filter(seatbelt == "Nearly always")) / nrow(genx_drivers)
genx_sometimes <- nrow(genx_drivers %>% filter(seatbelt == "Sometimes")) / nrow(genx_drivers)
genx_seldom <- nrow(genx_drivers %>% filter(seatbelt == "Seldom")) / nrow(genx_drivers)
genx_never <- nrow(genx_drivers %>% filter(seatbelt == "Never")) / nrow(genx_drivers)

millenial_always <- nrow(millenial_drivers %>% filter(seatbelt == "Always")) / nrow(millenial_drivers)
millenial_near_always <- nrow(millenial_drivers %>% filter(seatbelt == "Nearly always")) / nrow(millenial_drivers)
millenial_sometimes <- nrow(millenial_drivers %>% filter(seatbelt == "Sometimes")) / nrow(millenial_drivers)
millenial_seldom <- nrow(millenial_drivers %>% filter(seatbelt == "Seldom")) / nrow(millenial_drivers)
millenial_never <- nrow(millenial_drivers %>% filter(seatbelt == "Never")) / nrow(millenial_drivers)

boomer_stats <- c(boomer_always, boomer_near_always, boomer_sometimes, boomer_seldom, boomer_never)
genx_stats <- c(genx_always, genx_near_always, genx_sometimes, genx_seldom, genx_never)
millenial_stats <- c(millenial_always, millenial_near_always, millenial_sometimes, millenial_seldom, millenial_never)

freq_use <- c("Always", "Nearly Always", "Sometimes", "Seldom", "Never")

q1_stats <- data.frame(freq_use, boomer_stats, genx_stats, millenial_stats)
names(q1_stats) <- c("Seatbelt", "Boomers", "GenX", "Millenials")

q1_stats %>% gt(rowname_col = "Seatbelt") %>%
  tab_header(
    title = "Summary of Seatbelt compliance",
    subtitle = "Data source: Behavioral Risk Factor Surveillance System (2013)"
  ) %>%
  opt_align_table_header(align = "left") %>%
  fmt_percent(columns = everything())
```

Summary of Seatbelt compliance

Data source: Behavioral Risk Factor Surveillance System (2013)

	Boomers	GenX	Millenials
Always	88.00%	85.22%	77.20%
Nearly Always	6.77%	8.33%	12.67%
Sometimes	2.65%	3.28%	5.41%

Summary of Seatbelt compliance

Data source: Behavioral Risk Factor Surveillance System (2013)

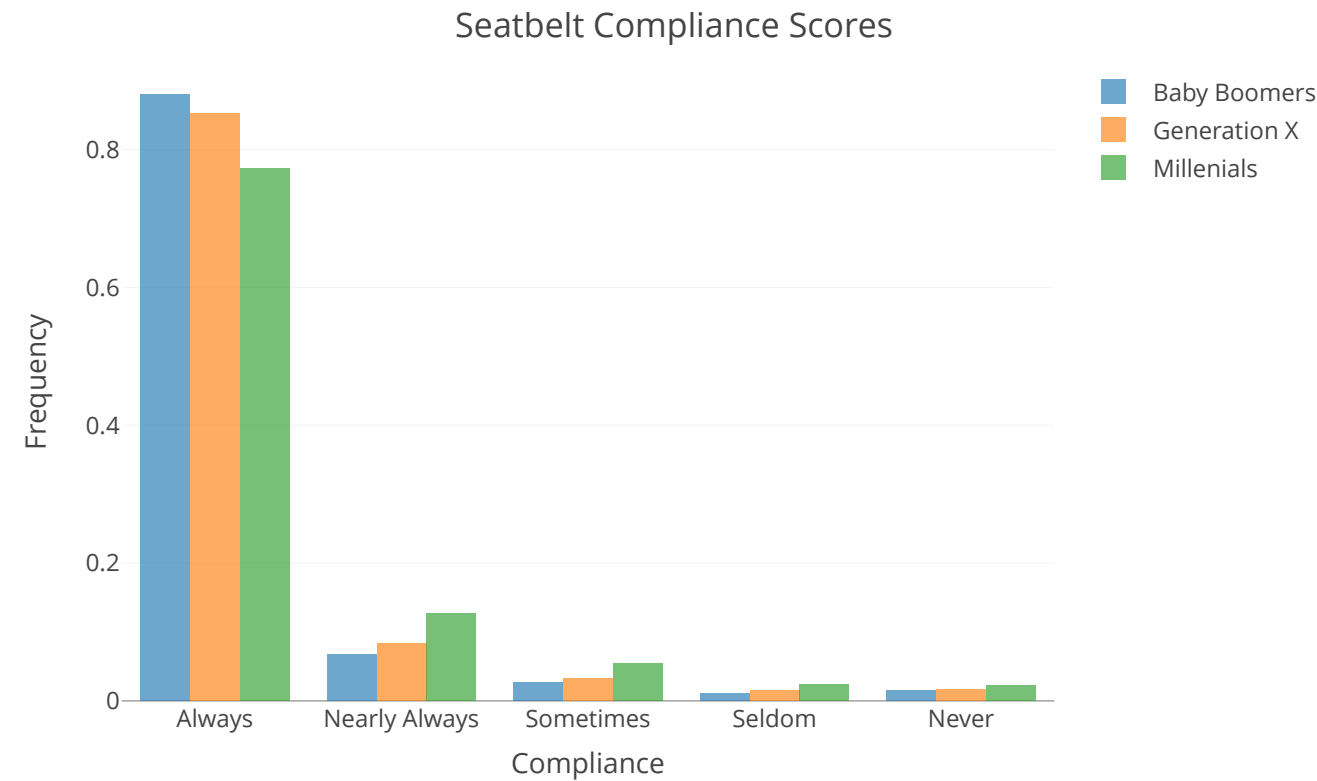
Seldom	1.05%	1.46%	2.44%
Never	1.54%	1.70%	2.28%

These results contradict my hypothesis.

88% of baby boomers always wear their seat belts, while only 77% of millenials always wear their seat belts. Only 1.5% of baby boomers never wear their seat belts, while 2.3% of millenials never wear their seat belts. It seems that the older generations are actually more seatbelt compliant than the younger generation - not what I expected! This can likely be attributed to more experienced drivers having first hand knowledge of the dangers associated with driving. Also, I learned that my dad is an outlier in the 1.5% of baby boomers that never wear a seatbelt. I'll have to talk with him about that! We can see that the vast majority of people across all generations either always or nearly always wear seatbelts.

Visualizing the results

```
fig <- plot_ly(q1_stats, x =~ Seatbelt, y =~ Boomers, type='bar', name="Baby Boomers", opacity=0.65)
fig <- fig %>% add_trace( y =~ GenX, type='bar', name="Generation X", opacity=0.65)
fig <- fig %>% add_trace( y =~ Millenials, type='bar', name="Millenials", opacity=0.65)
fig <- fig %>% layout(
  title="Seatbelt Compliance Scores",
  xaxis=list(title="Compliance",
             categoryorder="array",
             categoryarray=c("Always", "Nearly Always", "Sometimes", "Seldom", "Never")),
  yaxis=list(title="Frequency")
)
fig
```



This bar chart illustrates the relationship between generations and seatbelt compliance. Overall Older generations are more likely to always wear a seatbelt, and younger generations are less likely to always wear a seatbelt and more likely to never wear one.

Research qusion 2:

How does exercise and sleep coorelate with overall health?

```
q2 <- brfss2013 %>%
  select(genhlth, sleptim1, exerhmm1, exeroft1) %>%
  drop_na()

q2_summary <- q2 %>%
  group_by(genhlth) %>%
  summarize(
    avg_exer_time = mean(exerhmm1, trim=.1),
    avg_sleep = mean(sleptim1, trim=.1))

q2_summary %>% gt(rowname_col = "genhlth") %>%
  tab_header(
    title = "Summary of Average Exercise and Sleep Time",
    subtitle = "Data source: Behavioral Risk Factor Surveillance System (2013)"
  ) %>%
  opt_align_table_header(align = "left") %>%
  cols_label(
    avg_exer_time = md("Average Exercise Time,<br>minutes"),
    avg_sleep = md("Average Sleep Time,<br>hours")
  ) %>%
  cols_width(
    avg_exer_time ~ px(210),
    avg_sleep ~ px(210)
  ) %>%
  fmt_number( columns = everything())
```

Summary of Average Exercise and Sleep Time

Data source: Behavioral Risk Factor Surveillance System (2013)

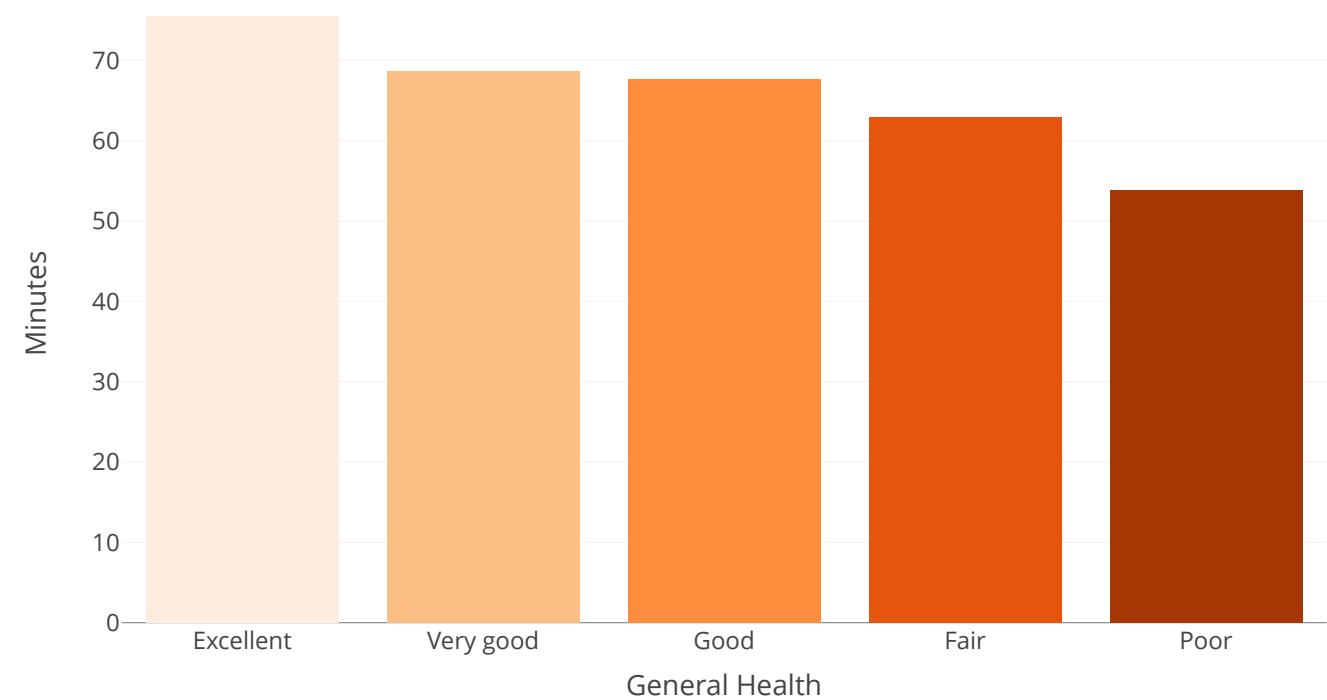
	Average Exercise Time, minutes	Average Sleep Time, hours
Excellent	75.49	7.21
Very good	68.69	7.11
Good	67.70	7.04
Fair	62.96	6.85
Poor	53.89	6.59

We can see that there is a positive relationship between both time spent exercizing and sleep on general health. In general people who are in excellent health exercise for an average of 75 minutes, and sleep an average of 7.2 hours.

Visualize the results

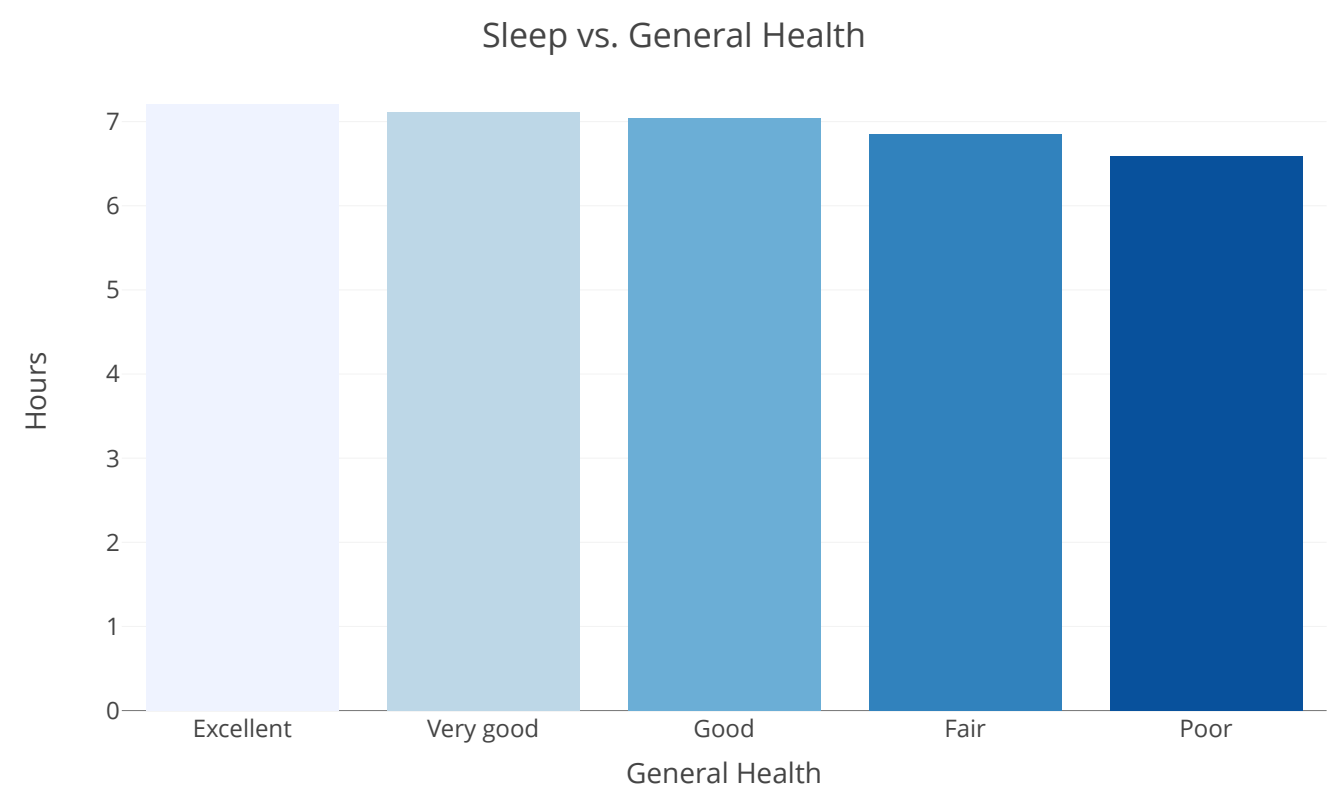
```
fig <- plot_ly(
  q2_summary,
  x =~genhlth, y=~avg_exer_time,
  type="bar", name="Average Exercise Time",
  color=~genhlth, colors="Oranges",
  showlegend=FALSE) %>%
  layout(
    title="Exercise Time vs. General Health",
    xaxis=list(title="General Health"),
    yaxis=list(title="Minutes")
  )
fig
```

Exercise Time vs. General Health



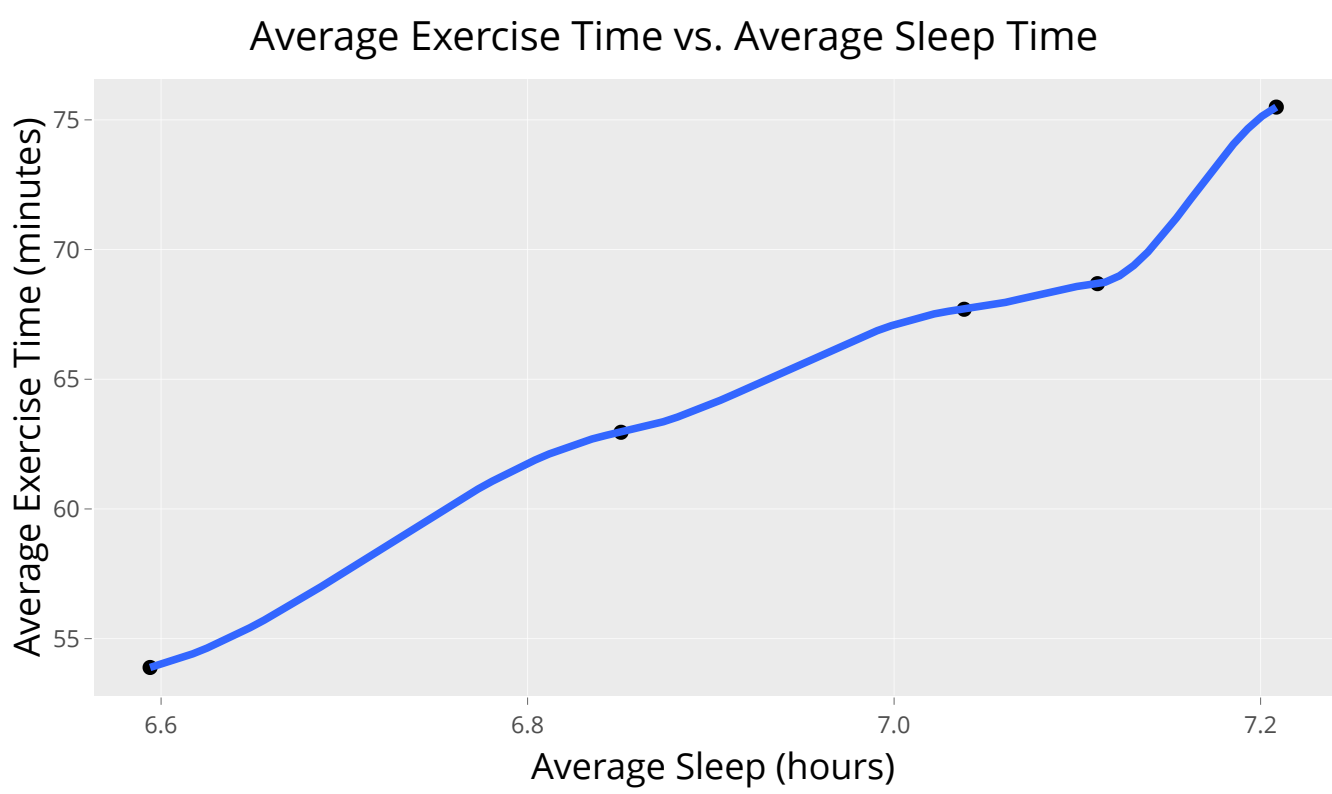
People who exercise just a little bit more greatly improve their chances of being healthy. 75 minutes seems to be the right amount. Normally I go to the gym for one hour. So by spending just an additional 15 minutes working out I can significantly improve my general health.

```
fig <- plot_ly(
  q2_summary,
  x =~genhlth, y=~avg_sleep,
  type="bar", name="Average Sleep Time",
  color=~genhlth, colors="Blues",
  showlegend=FALSE) %>%
  layout(
    title="Sleep vs. General Health",
    xaxis=list(title="General Health"),
    yaxis=list(title="Hours")
  )
fig
```



It’s no suprise that people who are in better general health get more sleep than those who are in poorer health. Getting atleast 7 hours of sleep seems very important to general health.

```
fig <- ggplot(data = q2_summary, aes(x = avg_sleep, y = avg_exer_time)) +
  geom_point() +
  geom_smooth()
fig <- ggplotly(fig) %>%
  layout(
    title="Average Exercise Time vs. Average Sleep Time",
    xaxis=list(title="Average Sleep (hours)"),
    yaxis=list(title="Average Exercise Time (minutes)")
  )
fig
```



Finally, we can see that there is a positive coorelation between exercize time and sleep. People who exercize more get more sleep. Since both of these variables have a positive effect on general health it’s important to note that exercize has the added benefit of getting better sleep, which in turn has a positive relationship on general health.

Research question 3:

Do veterans get the emotional support they need?

```
question_3 <- brfss2013 %>%
  select(veteran3, emtsuprt, lsatisfy) %>%
  filter(veteran3 == "Yes") %>%
  drop_na()

# Let's map these to numbers to gauntify the results
question_3 <- question_3 %>%
  mutate(
    support = case_when(
      emtsuprt == "Always" ~ 5L,
      emtsuprt == "Usually" ~ 4L,
      emtsuprt == "Sometimes" ~ 3L,
      emtsuprt == "Rarely" ~ 2L,
      emtsuprt == "Never" ~ 1L,
    ),
    life_satisfaction = case_when(
      lsatisfy == "Very satisfied" ~ 4L,
      lsatisfy == "Satisfied" ~ 3L,
      lsatisfy == "Dissatisfied" ~ 2L,
      lsatisfy == "Very dissatisfied" ~ 1L,
    )
  )
q3_summary <- question_3 %>%
  group_by(emtsuprt) %>%
  summarise(
    avg_support = mean(support),
    avg_satisfaction = mean(life_satisfaction)
  )

q3_summary %>% gt(rowname_col = "emtsuprt") %>%
  tab_header(
    title = "Summary of Average Emotional Support And Life Satisfaction",
    subtitle = md("Data source: Behavioral Risk Factor Surveillance System (2013)<br>Veterans")
  ) %>%
  opt_align_table_header(align = "left") %>%
  cols_label(avg_support = "Average Support",
             avg_satisfaction = "Average Satisfaction") %>%
  fmt_number(columns = everything())
```

Summary of Average Emotional Support And Life Satisfaction

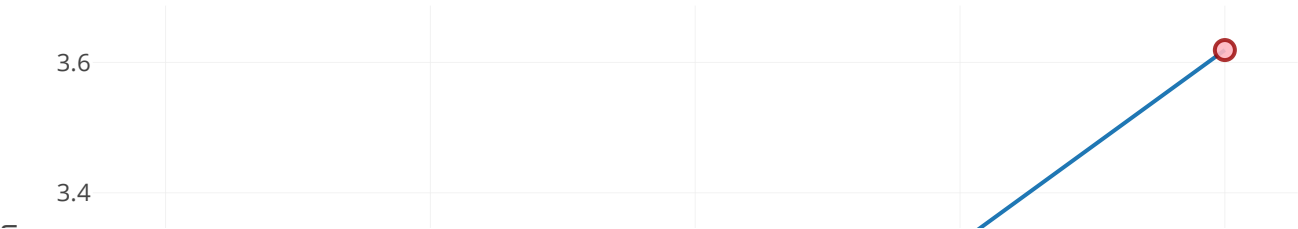
Data source: Behavioral Risk Factor Surveillance System (2013)
Veterans

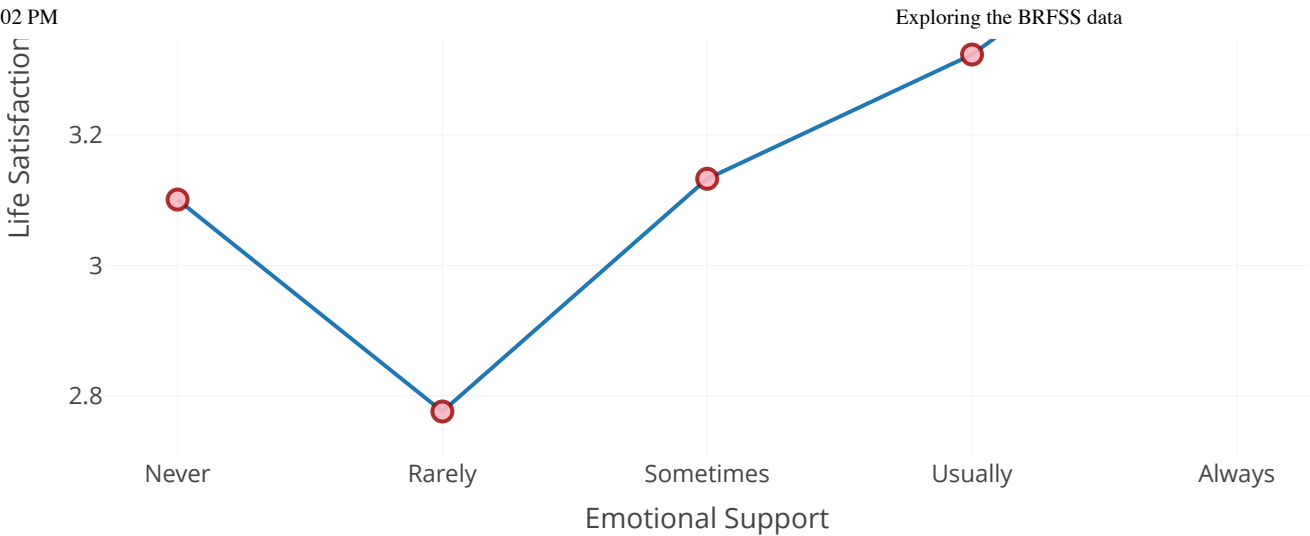
	Average Support	Average Satisfaction
Always	5.00	3.62
Usually	4.00	3.32
Sometimes	3.00	3.13
Rarely	2.00	2.78
Never	1.00	3.10

Visualizing this trend

```
xorder = q3_summary$emtsuprt %>% levels() %>% rev()
fig <- plot_ly(q3_summary, x = ~emtsuprt, y = ~avg_satisfaction, type='scatter', mode='lines+markers',
  marker = list(size = 10,
    color = 'rgba(255, 182, 193, .9)',
    line = list(color = 'rgba(152, 0, 0, .8)',
      width = 2)))
fig <- fig %>%
  layout(
    title="Life Satisfaction vs. Emotional Support Among Veterans",
    xaxis=list(title="Emotional Support",
      categoryorder="array",
      categoryarray=xorder),
    yaxis=list(title="Life Satisfaction")
  )
fig
```

Life Satisfaction vs. Emotional Support Among Veterans





There is an overall positive relationship between life satisfaction and emotional support. It’s interesting to note that people who rarely get emotional support actually score lower than people who never get emotional support, which seems to suggest that rarely getting emotional support is no better than never getting emotional support. This clearly shows that veterans who receive the emotional support they need are more satisfied than those who don’t. If you know a veteran let them know that you are there for them!

I want to compare this with the general population (non-veterans) to see if there is a similar dip with those who rarely receive emotional support.

```
gen_population <- brfss2013 %>%
  select(veteran3, emtsuprt, lsatisfy) %>%
  filter(veteran3 == "No") %>%
  drop_na()

gen_population <- gen_population %>%
  mutate(
    support = case_when(
      emtsuprt == "Always" ~ 5L,
      emtsuprt == "Usually" ~ 4L,
      emtsuprt == "Sometimes" ~ 3L,
      emtsuprt == "Rarely" ~ 2L,
      emtsuprt == "Never" ~ 1L,
    ),
    life_satisfaction = case_when(
      lsatisfy == "Very satisfied" ~ 4L,
      lsatisfy == "Satisfied" ~ 3L,
      lsatisfy == "Dissatisfied" ~ 2L,
      lsatisfy == "Very dissatisfied" ~ 1L,
    )
  )

gen_pop_summary <- gen_population %>%
  group_by(emtsuprt) %>%
  summarise(
    avg_support = mean(support),
    avg_satisfaction = mean(life_satisfaction)
  )

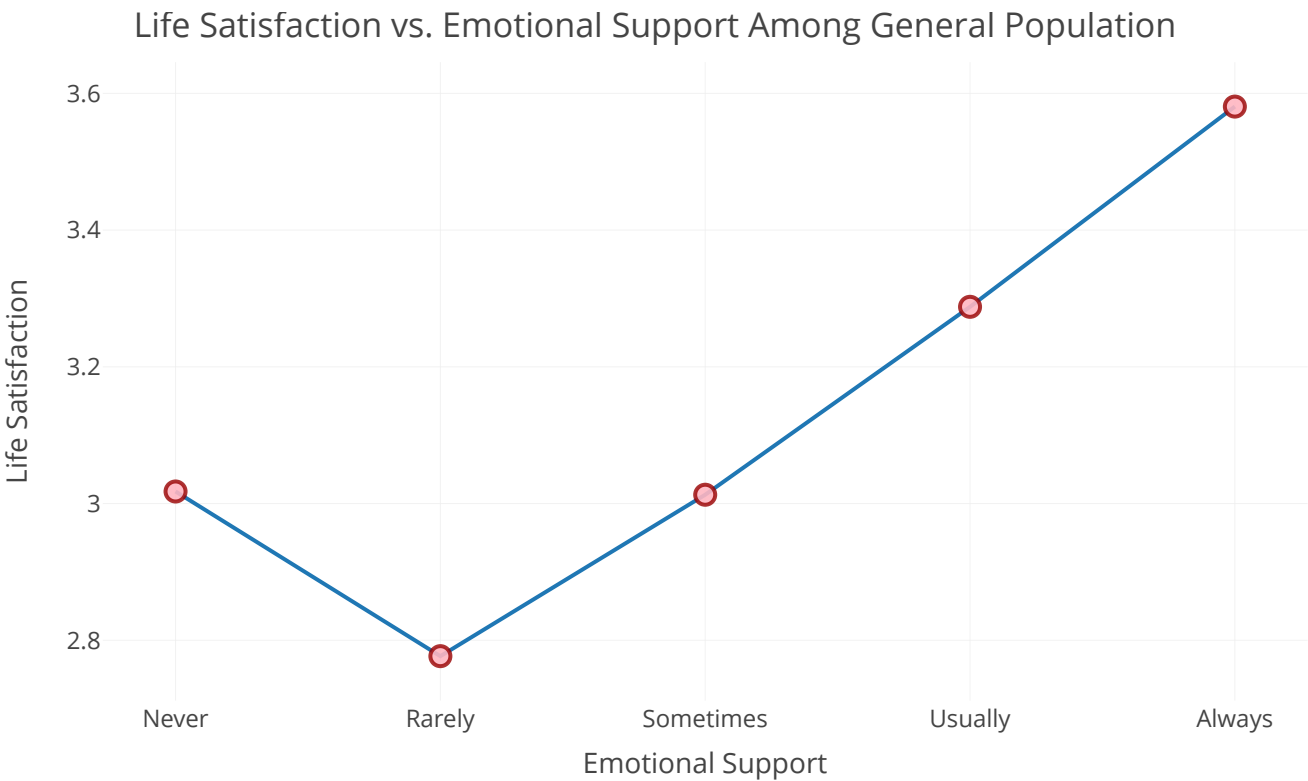
gen_pop_summary %>% gt(rowname_col = "emtsuprt") %>%
  tab_header(
    title = "Summary of Average Emotional Support and Life Satisfaction",
    subtitle = md("Data source: Behavioral Risk Factor Surveillance System (2013)<br>General Population")
  ) %>%
  opt_align_table_header(aligned = "left") %>%
  cols_label(
    avg_support = "Average Support",
    avg_satisfaction = "Average Life Satisfaction"
  ) %>%
  cols_width(avg_support ~ px(210), avg_satisfaction ~ px(210)) %>%
  fmt_number(columns = "avg_satisfaction")
```

Summary of Average Emotional Support and Life Satisfaction

Data source: Behavioral Risk Factor Surveillance System (2013)
General Population

	Average Support	Average Life Satisfaction
Always	5	3.58
Usually	4	3.29
Sometimes	3	3.01
Rarely	2	2.78
Never	1	3.02

```
xorder = gen_pop_summary$emtsuprt %>% levels() %>% rev()
fig <- plot_ly(gen_pop_summary, x = ~emtsuprt, y = ~avg_satisfaction, type='scatter', mode='lines+markers',
              marker = list(size = 10,
                            color = 'rgba(255, 182, 193, .9)',
                            line = list(color = 'rgba(152, 0, 0, .8)',
                                       width = 2)))
xorder = c("Never", "Rarely", "Sometimes", "Usually", "Always")
fig <- fig %>%
  layout(
    title="Life Satisfaction vs. Emotional Support Among General Population",
    xaxis=list( title="Emotional Support",
                categoryorder="array",
                categoryarray=xorder),
    yaxis=list(title="Life Satisfaction")
  )
fig
```



Interestingly there is a similar dip! There is still a clear overall positive trend but for some reason people who rarely receive emotional support are less satisfied in life than those who never receive emotional support. Maybe people who never receive emotional support tell themselves and others that everything is fine as a coping strategy. Maybe they mask their feelings with drugs and alcohol. But why do people who rarely receive emotional support consistently score lower? I beleive that only rarely receiving emotional hightens the awareness for the pyschological need, awareness of the lack thereof, and awareness of emotional pain. There is also a strong possibility that people who rarely provide emotional support aren’t very good at it. Imagine having a physical injury and a doctor that haphazardly addresses that injury. You might feel you were better off treating the injury yourself. This may explain why there is a higher life satisfaction score to the left and right of rarely.

Let’s get back to our veterans and see how well society is taking care of them.

```
nvets <- nrow(question_3)
vet_always <- nrow(question_3 %>% filter(emtsuprt == "Always")) / nvets
vet_usually <- nrow(question_3 %>% filter(emtsuprt == "Usually")) / nvets
vet_sometimes <- nrow(question_3 %>% filter(emtsuprt == "Sometimes")) / nvets
vet_rarely <- nrow(question_3 %>% filter(emtsuprt == "Rarely")) / nvets
vet_never <- nrow(question_3 %>% filter(emtsuprt == "Never")) / nvets

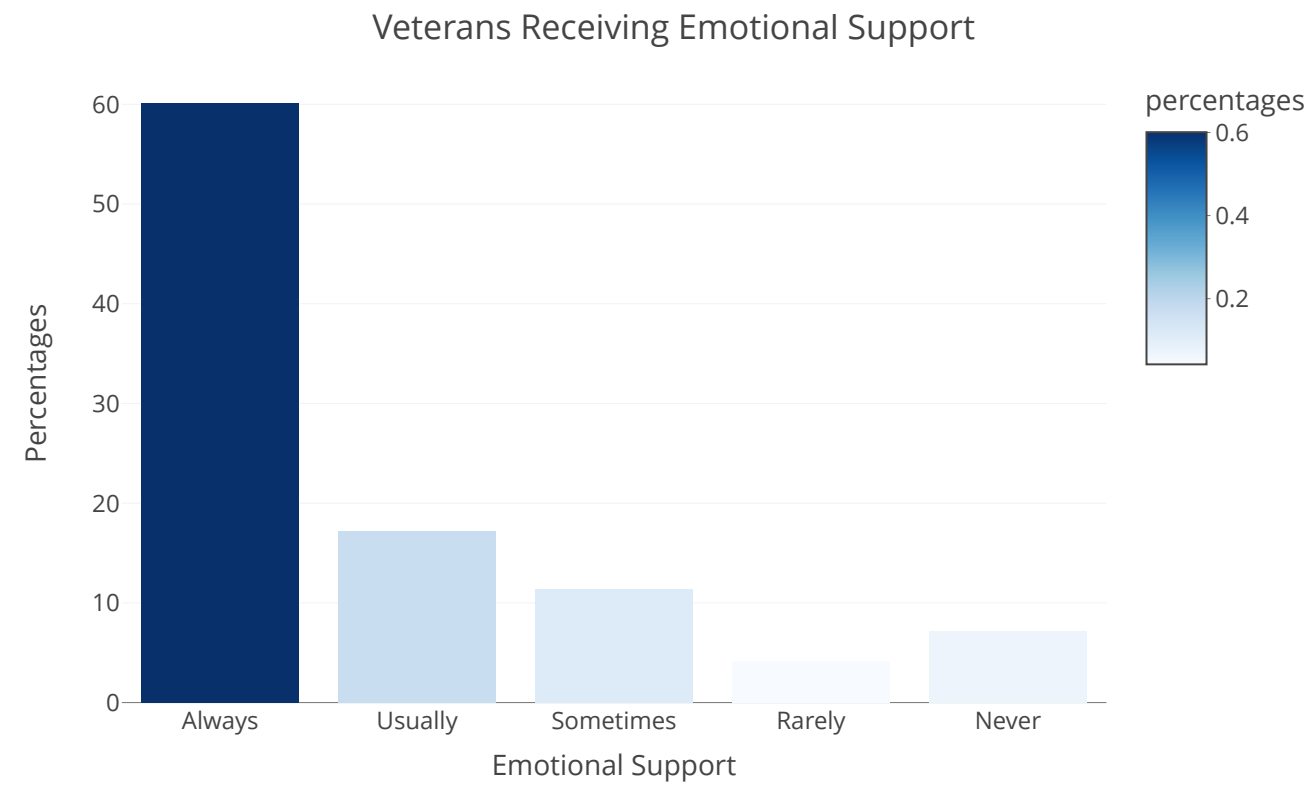
percentages <- c(vet_always, vet_usually, vet_sometimes, vet_rarely, vet_never)
emtsprt <- question_3$emtsuprt %>% levels()
vet_stats <- data.frame(emtsprt, percentages)

vet_stats %>% gt(rowname_col = "emtsprt") %>%
  tab_header(title = "Veterans Receiving Emotional Support") %>%
  opt_align_table_header(align = "left") %>%
  fmt_percent(columns = "percentages")
```

Veterans Receiving Emotional Support	
	percentages
Always	60.07%
Usually	17.18%
Sometimes	11.41%
Rarely	4.19%
Never	7.15%

Only 77% of veterans always or usually receive emotional support, while more than 11% of all veterans rarely or never receive emotional support. This is obviously an issue that needs to be addressed. If you know a veteran ask them how they're doing! Let them know that you are there for them if they need to talk. Sadly there is little to no support from the military for veterans returning home from deployment and they are expected to adjust to civilian life on their own.

```
xorder = vet_stats$emtsprt
fig <- plot_ly(vet_stats, x = ~emtsprt, y = ~percentages*100, type='bar', color = ~percentages, colors='Blues')
fig <- fig %>%
  layout(
    title="Veterans Receiving Emotional Support",
    xaxis=list(title="Emotional Support",
              categoryorder="array",
              categoryarray=xorder),
    yaxis=list(title="Percentages")
  )
fig
```



It's sad to see that only 60% of veterans always receive the emotional support they need, especially when we can also see the positive difference it makes on a persons outlook on life. This has been an interesting dataset to analyze and hopefully subsequent years show improvements in these numbers.