Exploring the BRFSS data

Setup

Load packages

```
library(knitr)
library(pander)
library(ggplot2)
library(colorspace)
library(tidyr)
library(dplyr)
```

Load data

```
load("brfss2013.RData")
brfss2013 = filter(brfss2013, X_state == 'Texas')

allprep <- function(...) {
    study <- brfss2013 %>%
        select(...)

    return(study[complete.cases(study),])
}
```

Part 1: Data

The BRFSS is an observational study targeting the non-institutionalized adult population of US citizens aged 18 and older. The study asks respondents different questions about their health and other information in the months prior to their contact. The study contacts respondents through phone call, and eliminates respondents who reside at a college or if they use a phone that is not classified as a landline. Once conditions are met, a random adult is selected from the household. Random sampling is used to select the phone numbers and to select which adult to question.

This method of collecting data contains its issues. First of all, it requires the use of telephones, limiting the scope of the data frame to only those who have a landline telephone. Furthermore, a respondent may be neglectful towards answering the phone, thus introducing non response bias.

As the survey was just an observational survey, and non a population under experimental control, there is no causation.

Part 2: Research questions

Research question 1:

Does a correlation exist between the hours slept each night and reported days with poor physical health or poor mental health in the last 30 days?

Variables: sleptim (hours slept), physhlth (days reported poor physical health), menthlth (days reported poor mental health)

Interest: This question is of importance as it can see if good sleeping habits can lead to having better mental or physical health.

Research question 2:

Does having a higher level of education increase ones income level?

Variables: educa (education level) income2 (income level)

Interest: There is plenty of debate about whether a college education is worth it or not. This also covers how important is too finish highschool/go back for a GED.

Research question 3:

Is there a relation ship between the time spent exercising and the amount of days binge drinking (5 drinks in a day for male and 4 drinks in a day for female) or frequency of smoking?

Variables: exerhmm1 (Minutes Or Hours Walking, Running, Jogging, Or Swimming), drnk3ge (Binge Drinking), smokday2 (frequency of days now smoking)

Interest: This question is important as exercising is an important part of staying healthy and if there exists a correlating between lack of exercise and consumption of alcohol or tobacco, then it may influence someone to potentially stop drinking or smoking.

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 quesion 1:

```
# CONVERT DATA TO NUMERICS

nSleep = as.numeric(brfss2013$sleptim1)
nPhys = as.numeric(brfss2013$physhlth)
nMntl = as.numeric(brfss2013$menthlth)

#GET STATS
stats = summarize(brfss2013, "sleptim1", mean(sleptim1, na.rm=T), sd(sleptim1, na.rm=T))
colnames(stats) = c("Variable", "mean", "sd")

physStats = summarize(brfss2013, "physhlth", mean(physhlth, na.rm=T), sd(physhlth, na.rm=T))
colnames(physStats) = c("Variable", "mean", "sd")
stats = rbind(stats, physStats)

mentStats = summarize(brfss2013, "menthlth", mean(menthlth, na.rm=T), sd(menthlth, na.rm=T))
colnames(mentStats) = c("Variable", "mean", "sd")
stats = rbind(stats, mentStats)
```

```
dat1 <- allprep( sleptim1, physhlth, menthlth)
sum1 <- dat1 %>%
    group_by(sleptim1, physhlth) %>%
    summarize(Sum = n()) %>%
    spread(physhlth, Sum) %>%
    mutate()
```

```
## `summarise()` regrouping output by 'sleptim1' (override with `.groups` argument)
```

```
colnames(sum1)[1] <- "Hours slept"
pandoc.table(sum1,caption="Summary Statistics for hours slept and days physical health is not go
         od", justify = "center")
```

```
##
## ------------------------------------------------------------------------
##  Hours slept    0      1     2     3    4    5    6    7    8    9    10   11
## ------------- ------ ----- ----- ----- ---- ---- ---- ---- ---- ---- ---- ----
##        1         1     NA    NA    NA    NA   NA   NA   NA   NA   NA   NA   NA
##
##        2         5     NA     1    NA    1    NA   NA   NA   1    NA   NA   NA
##
##        3         17    NA     3     1    NA   NA   NA   2    NA   NA   NA   NA
##
##        4        106    11    14    11    4    7    1    4    1    NA   6    NA
##
##        5        358    18    31    25    16   14   3    8    4    1    15   2
##
##        6        1302   86    122   62    38   87   15   51   6    NA   52   NA
##
##        7        2080   120   171   93    58   56   8    52   12   1    41   3
##
##        8        2224   89    137   102   41   82   15   62   7    NA   56   NA
##
##        9        353    10    27    14    5    18   4    17   1    2    10   NA
##
##       10        159     6    15     8    9    8    1    6    NA   1    6    1
##
##       11         11    NA    NA    NA    NA   NA   NA   NA   NA   NA   NA   NA
##
##       12         45     1     2     2    3    3    NA   2    NA   NA   1    NA
##
##       13         3     NA    NA    NA    NA   NA   NA   NA   NA   NA   NA   NA
##
##       14         4      1     1     1    2    NA   NA   NA   NA   NA   2    NA
##
##       15         3     NA     1    NA    NA   NA   NA   NA   NA   NA   NA   NA
##
##       16         2     NA    NA     1    NA   1    NA   NA   NA   NA   NA   NA
##
##       17         1     NA    NA    NA    NA   NA   NA   NA   NA   NA   NA   NA
##
##       18         NA    NA    NA    NA    NA   NA   NA   NA   NA   NA   NA   NA
##
##       24         NA    NA    NA    NA    NA   NA   NA   NA   NA   NA   NA   NA
## ------------------------------------------------------------------------
##
## Table: Summary Statistics for hours slept and days physical health is not good (continued bel
ow)
##
##
## ------------------------------------------------------------------------------------------
##  12   13   14   15   16   17   18   20   21   22   23   24   25   26   27   28   29   30
## ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- -----
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   1    2    NA   NA   NA   NA   NA   NA   NA   NA   NA   1    NA   NA   1     6
```

```
##
##  NA   NA   1    1    NA   NA   NA   1    NA   NA   NA   NA   NA   NA   NA   NA   1    26
##
##  3    1    2    9    NA   NA   NA   6    1    NA   NA   NA   2    NA   NA   NA   1    53
##
##  2    NA   10   34   1    1    1    17   6    NA   NA   NA   5    1    NA   NA   NA   88
##
##  3    2    33   42   3    1    NA   22   9    NA   1    1    10   NA   NA   2    NA   163
##
##  3    NA   25   40   2    1    1    22   6    NA   1    1    10   NA   NA   3    2    131
##
##  4    NA   41   53   3    NA   NA   25   6    1    1    NA   9    NA   2    2    5    199
##
##  NA   1    4    13   NA   NA   1    8    NA   NA   NA   NA   NA   NA   NA   NA   NA   36
##
##  NA   1    3    9    NA   NA   NA   3    NA   NA   NA   NA   1    NA   1    NA   NA   32
##
##  1    NA   NA   NA   NA   NA   NA   1    NA   NA   NA   NA   NA   NA   NA   NA   NA    3
##
##  NA   NA   1    3    NA   NA   NA   1    NA   NA   NA   NA   1    NA   NA   NA   NA   19
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  1    NA   NA   NA   NA   NA   NA   1    NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   1    NA   NA   NA   1    NA   NA   NA   NA   NA   NA   NA   NA   NA    2
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    2
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
## ------------------------------------------------------------------------------------------
```

It initially looks like people sleeping 6,7,or 8 hours tend to have better physical health, until it reaches 30 days of bad health and the highest numbers are for 6,7,and 8 hours slept.

```
sum1 <- dat1 %>%
    group_by(sleptim1, menthlth) %>%
    summarize(Sum = n()) %>%
    spread(menthlth, Sum) %>%
    mutate()
```

```
## `summarise()` regrouping output by 'sleptim1' (override with `.groups` argument)
```

```
colnames(sum1)[1] <- "Hours slept"
pandoc.table(sum1,caption="Summary Statistics for hours slept and days physical health is not go
         od", justify = "center")
```

```
##
## ------------------------------------------------------------------------
##  Hours slept    0     1     2    3    4    5    6    7    8    9    10   11
## ------------- ------ ---- ----- ---- ---- ---- ---- ---- ---- ---- ---- ----
##        1         1     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##        2         9     NA    1    NA   NA   1    NA   NA   NA   NA   NA   NA
##
##        3         20    NA    3    NA   NA   2    NA   2    NA   NA   3    NA
##
##        4        113    9     4    12   7    8    NA   8    3    NA   8    NA
##
##        5        395    16   22    25   8    20   4    13   2    NA   24   1
##
##        6        1401   46   92    76   36   77   9    39   8    1    48   NA
##
##        7        2269   85   129   81   39   73   9    25   6    1    42   1
##
##        8        2554   69   103   67   33   66   3    43   4    NA   41   1
##
##        9        406    11   21    17   7    9    NA   7    3    NA   9    NA
##
##       10        201    8     8    4    4    5    NA   3    NA   NA   5    NA
##
##       11         10    1     1    1    1    1    NA   NA   NA   NA   NA   NA
##
##       12         49    1     3    7    NA   2    NA   1    NA   NA   1    NA
##
##       13         2     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##       14         10    NA   NA    NA   NA   1    NA   NA   NA   NA   NA   NA
##
##       15         5     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##       16         4     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##       17         1     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##       18         1     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##       24         NA    NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
## ------------------------------------------------------------------------
##
## Table: Summary Statistics for hours slept and days physical health is not good (continued bel
ow)
##
##
## ------------------------------------------------------------------------------------------
##  12   13   14   15   16   17   18   19   20   21   22   23   24   25   27   28   29   30
## ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- -----
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   1    NA   NA   NA   NA   2    NA   NA   NA   NA   NA   NA   2    1     2
```

```
##
##  NA   NA   1    1    NA   NA   NA   1    NA   NA   NA   NA   NA   NA   NA   1    NA   19
##
##  NA   1    5    13   NA   NA   NA   NA   10   NA   1    NA   NA   3    NA   1    1    36
##
##  6    NA   8    22   NA   NA   NA   1    21   4    NA   NA   1    4    NA   1    2    61
##
##  6    2    18   60   NA   1    1    NA   40   2    1    NA   NA   8    NA   1    NA   140
##
##  5    NA   15   38   1    NA   NA   NA   27   1    NA   1    NA   8    NA   1    1    85
##
##  2    NA   11   43   NA   NA   NA   NA   23   2    NA   NA   NA   5    2    1    1    92
##
##  2    NA   3    7    NA   NA   NA   NA   4    NA   NA   NA   NA   2    NA   NA   NA   16
##
##  1    NA   1    6    1    NA   NA   NA   3    NA   2    NA   NA   1    NA   1    NA   16
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   3    NA   NA   NA   NA   3    1    NA   NA   NA   NA   NA   NA   1    12
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   1    NA    1
##
##  1    NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    2
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    3
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
## ------------------------------------------------------------------------------------------
```

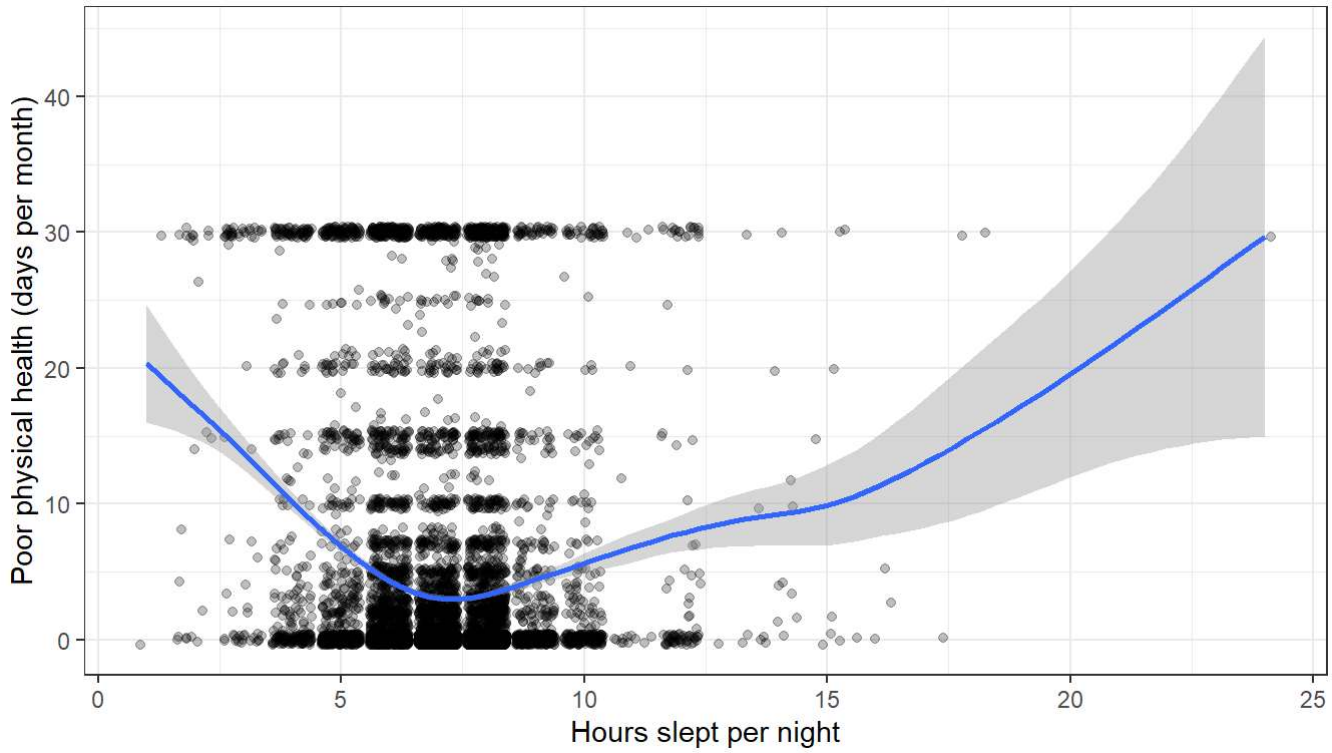
Once again It initially looks like people sleeping 6,7,or 8 hours have better mental health, but the majority of people who have 30 days with bad mental health also sleep 6,7,or 8 hours. What these tables really prove is that the majority of people sleep 6,7,or 8 hours, these tables do not give an answer to the question.

```
# PLOT PHYSICAL
ggplot(brfss2013, aes(x=sleptim1, y=physhlth)) +
 theme_bw() +
  geom_jitter(alpha=0.25) +
   geom_smooth() +
    labs(x="Hours slept per night", y="Poor physical health (days per month)")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 594 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 594 rows containing missing values (geom_point).
```




```
cor(nSleep, nPhys, use="complete.obs")
```

```
## [1] -0.07839386
```

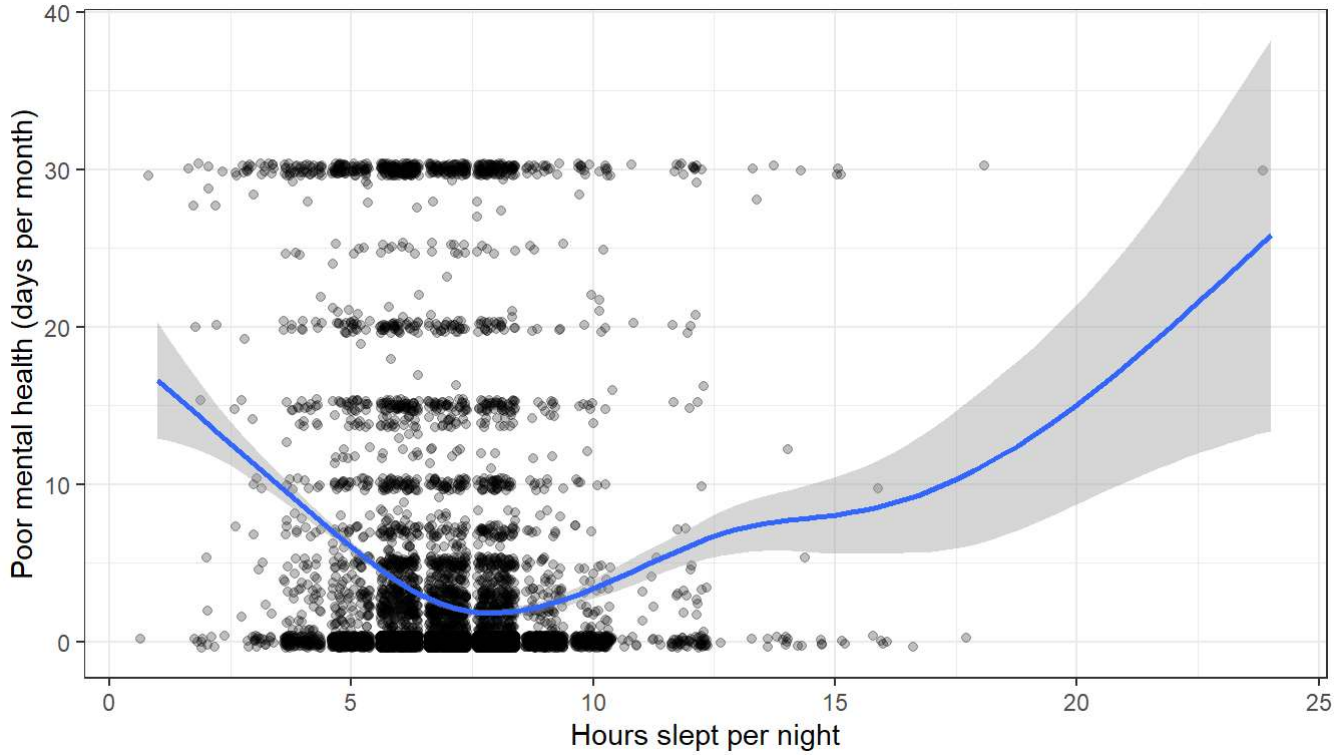
There does not appear to be any clear correlation, as the resulting coefficient is not significant.

```
# PLOT MENTAL
ggplot(brfss2013, aes(x=sleptim1, y=menthlth)) +
 theme_bw() +
  geom_jitter(alpha=0.25) +
   geom_smooth() +
    labs(x="Hours slept per night", y="Poor mental health (days per month)")
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 514 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 514 rows containing missing values (geom_point).
```




```
cor(nSleep, nMntl, use="complete.obs")
```

```
## [1] -0.1270602
```

Once again, there does not appear to be any clear correlation, as the resulting coefficient is not significant.

Summary

Initial Question: Does a correlation exist between the hours slept each night and reported days with poor physical health or poor mental health in the last 30 days?

Narrative from the Exploratory Analysis: There is no linear relationship between hours slept and the two variables. However there is a significant dip in the lowess regression line, implying that people on both extremes of the spectrum might have higher number of days with poor physical health or poor mental health.

Research quesion 2:

```
dat2 <- allprep( educa, income2)
dat2$educa <- plyr::revalue(dat2$educa, c("Never attended school or only kindergarten" = "None/K
        inder Only",
                                          "Grades 1 through 8 (Elementary)" = "1st-8th",
                                          "Grades 9 though 11 (Some high school)" = "9th - 11th"
        ,
                                          "Grade 12 or GED (High school graduate)" = "12th or GE
        D",
                                          "College 1 year to 3 years (Some college or technical
         school)" = "Some college/Tech School",
                                          "College 4 years or more (College graduate)" = "Colleg
        e Graduate"))


sum2 <- dat2 %>%
    group_by(income2, educa) %>%
    summarize(Sum = n()) %>%
    spread(educa, Sum) %>%
    mutate()
```

```
## `summarise()` regrouping output by 'income2' (override with `.groups` argument)
```

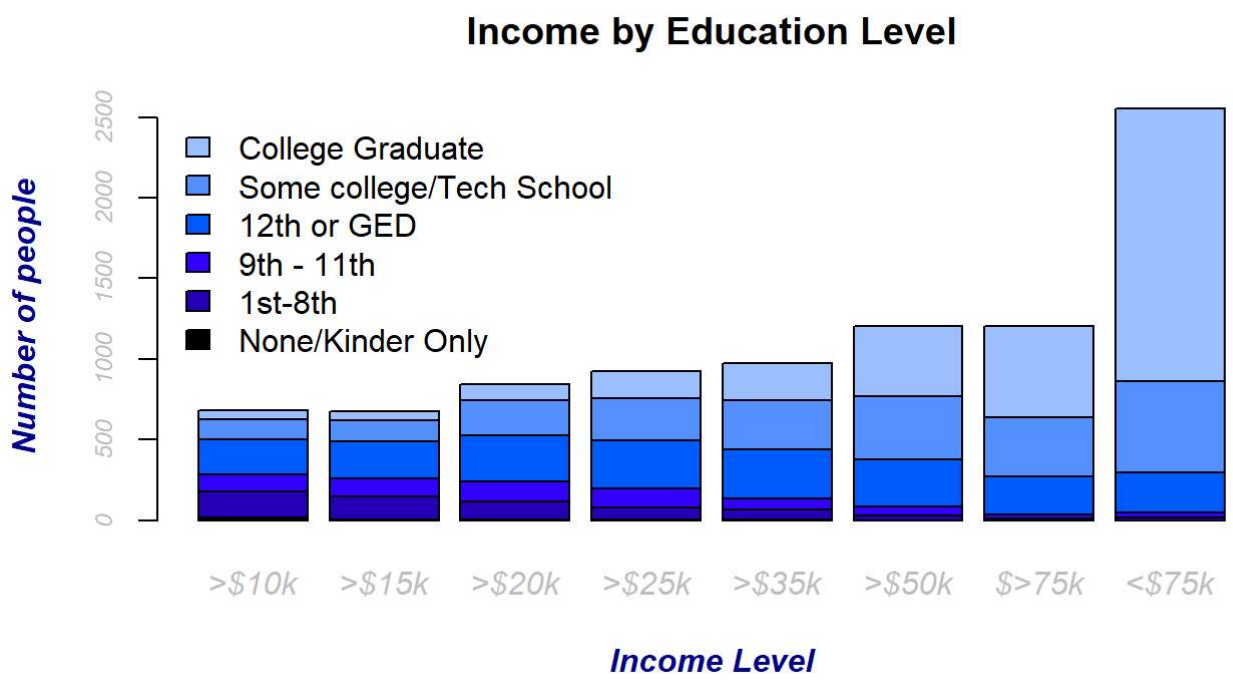
```
colnames(sum2)[1] <- "Income"
pandoc.table(sum1,caption="Summary Statistics for income by education level", justify = "center"
        )
```

```
##
## ------------------------------------------------------------------------
##  Hours slept    0     1     2    3    4    5    6    7    8    9    10   11
## ------------- ------ ---- ----- ---- ---- ---- ---- ---- ---- ---- ---- ----
##       1         1     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##       2         9     NA    1    NA   NA   1    NA   NA   NA   NA   NA   NA
##
##       3         20    NA    3    NA   NA   2    NA   2    NA   NA   3    NA
##
##       4        113    9     4    12   7    8    NA   8    3    NA   8    NA
##
##       5        395    16   22    25   8    20   4    13   2    NA   24   1
##
##       6        1401   46   92    76   36   77   9    39   8    1    48   NA
##
##       7        2269   85   129   81   39   73   9    25   6    1    42   1
##
##       8        2554   69   103   67   33   66   3    43   4    NA   41   1
##
##       9        406    11   21    17   7    9    NA   7    3    NA   9    NA
##
##      10        201    8     8    4    4    5    NA   3    NA   NA   5    NA
##
##      11         10    1     1    1    1    1    NA   NA   NA   NA   NA   NA
##
##      12         49    1     3    7    NA   2    NA   1    NA   NA   1    NA
##
##      13         2     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##      14         10    NA   NA    NA   NA   1    NA   NA   NA   NA   NA   NA
##
##      15         5     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##      16         4     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##      17         1     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##      18         1     NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##      24         NA    NA   NA    NA   NA   NA   NA   NA   NA   NA   NA   NA
## ------------------------------------------------------------------------
##
## Table: Summary Statistics for income by education level (continued below)
##
##
## ------------------------------------------------------------------------------------------------
##  12   13   14   15   16   17   18   19   20   21   22   23   24   25   27   28   29   30
## ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- ---- -----
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   1    NA   NA   NA   NA   2    NA   NA   NA   NA   NA   NA   2    1     2
##
```

```
##  NA   NA   1    1    NA   NA   NA   1    NA   NA   NA   NA   NA   NA   NA   1    NA   19
##
##  NA   1    5    13   NA   NA   NA   NA   10   NA   1    NA   NA   3    NA   1    1    36
##
##  6    NA   8    22   NA   NA   NA   1    21   4    NA   NA   1    4    NA   1    2    61
##
##  6    2    18   60   NA   1    1    NA   40   2    1    NA   NA   8    NA   1    NA   140
##
##  5    NA   15   38   1    NA   NA   NA   27   1    NA   1    NA   8    NA   1    1    85
##
##  2    NA   11   43   NA   NA   NA   NA   23   2    NA   NA   NA   5    2    1    1    92
##
##  2    NA   3    7    NA   NA   NA   NA   4    NA   NA   NA   NA   2    NA   NA   NA   16
##
##  1    NA   1    6    1    NA   NA   NA   3    NA   2    NA   NA   1    NA   1    NA   16
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   3    NA   NA   NA   NA   3    1    NA   NA   NA   NA   NA   NA   1    12
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   1    NA    1
##
##  1    NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    2
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    3
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
##
##  NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA    1
## ----------------------------------------------------------------------------------------
```

This summary table shows the association in the data between Income and Education level. THe table shows that as income increases there is also an increase in education level, particularly in college education level.

```
counts <- table(dat2$educa, dat2$income2)
barplot(counts,
        main="Income by Education Level",
        ylab = "Number of people",
        xlab = "Income Level", col=c("black","#2500B6", "#3300FF", "#005BFF","#5491FF", "#9BBFF
         F"),
        font.axis = "3",
        col.axis = "gray",
        cex.axis = ".75",
        font.lab = "4",
        col.lab = "dark blue",
        names.arg=c(">$10k",">$15k",">$20k",">$25k",">$35k", ">$50k", "$>75k", "<$75k"),
        legend = TRUE, args.legend=(list(bty="n",x="topleft", ncol =1)))
```



This stacked bar plot visualizes the summary table and shows the amount of people within each income level based on their education level. It can be clearly seen from the graph that college graduates (indicated by the lightest blue) tend to earn more than $75,000.

Summary

Initial Question: Does having a better education potentially increase ones income level?

Narrative from the Exploratory Analysis: Yes, College Graduates make up the majority of $75k or more group, and over half of College Graduates make $75k or more. In fact, the number of college graduates increases each the Income Level increases.The data also shows that as Income Level increases, the amount of people with an education level lower than 12th or GED decreases.

Research quesion 3:

```
nSmok = as.numeric(brfss2013$smokday2)
nDrnk = as.numeric(brfss2013$drnk3ge5)
dat3 <- allprep(drnk3ge5,exerhmm1, smokday2)%>%
  filter(exerhmm1 < 601 & exerhmm1 > 61)
```

*NOTE: Removing data where exercise is greater than 600 as there are very few people who exercised this often, as well as removed people who did not exercise at least an hour. This way graphs become easier to read since it allows us to focus on the bigger pieces of data.

```
sum3 <- dat3 %>%
    group_by(drnk3ge5, exerhmm1) %>%
    summarize(Sum = n()) %>%
    spread(exerhmm1, Sum) %>%
    mutate()
```

```
## `summarise()` regrouping output by 'drnk3ge5' (override with `.groups` argument)
```

```
colnames(sum3)[1] <- "Days Binge Drinked"
pandoc.table(sum3,caption="Summary Statistics for days binge drinked and hours excericed in the
          last month.")
```

```
##
## ------------------------------------------------------------------------
##  Days Binge Drinked   100   110   115   120   125   130   145   200   230   250
## -------------------- ----- ----- ----- ----- ----- ----- ----- ----- ----- -----
##           0            175   NA     8     1     1    38     3    42     8     1
##
##           1            26    NA    NA    NA    NA     7     1     5    NA    NA
##
##           2            19     1     2    NA    NA    10     1     7    NA    NA
##
##           3             7    NA    NA     1    NA     2    NA     3    NA    NA
##
##           4            10    NA    NA    NA    NA     3    NA     1    NA    NA
##
##           5             9    NA    NA     1    NA     1    NA    NA    NA    NA
##
##           6             2    NA    NA    NA    NA     2    NA     1    NA    NA
##
##           8            NA    NA    NA    NA    NA     2    NA     2    NA    NA
##
##           9            NA    NA    NA    NA    NA    NA    NA     1    NA    NA
##
##           10            4    NA    NA    NA    NA     2    NA     3    NA    NA
##
##           12            4    NA    NA    NA    NA     2    NA     1    NA    NA
##
##           13            1    NA    NA    NA    NA    NA    NA    NA    NA    NA
##
##           15            2    NA    NA    NA    NA    NA    NA    NA    NA    NA
##
##           20           NA    NA    NA    NA    NA    NA    NA     1    NA    NA
##
##           23           NA    NA    NA    NA    NA     1    NA    NA    NA    NA
##
##           26           NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
##
##           30            3    NA    NA    NA    NA    NA    NA     2    NA    NA
## ------------------------------------------------------------------------
##
## Table: Summary Statistics for days binge drinked and hours excericed in the last month. (cont
inued below)
##
##
## ----------------------------------------
##  300   330   400   402   430   500   600
## ----- ----- ----- ----- ----- ----- -----
##  27     7    31     1     4     7     3
##
##   2    NA     4    NA     2     2    NA
##
##   2    NA     9    NA     1    NA     1
##
##   1     1     2    NA    NA    NA    NA
```

```
##
##   1    NA     3    NA    NA    NA    NA
##
##   2    NA     1    NA    NA     1    NA
##
##   1    NA    NA    NA    NA    NA    NA
##
##   2    NA    NA    NA    NA    NA    NA
##
##  NA    NA    NA    NA    NA    NA    NA
##
##   1    NA    NA    NA    NA    NA    NA
##
##  NA    NA    NA    NA    NA    NA    NA
##
##  NA    NA    NA    NA    NA    NA    NA
##
##   1     1    NA    NA    NA    NA    NA
##
##  NA    NA    NA    NA    NA    NA    NA
##
##  NA    NA    NA    NA    NA    NA    NA
##
##  NA    NA     1    NA    NA    NA    NA
##
##   1    NA     2    NA    NA     1    NA
## ----------------------------------------
```

This summary table shows the relationship between days binge drinked and hours exercised, we can see that a lot of the people who are exercising are also not binge drinking.

```
sum4 <- dat3 %>%
    group_by(smokday2, exerhmm1) %>%
    summarize(Sum = n()) %>%
    spread(exerhmm1, Sum) %>%
    mutate()
```

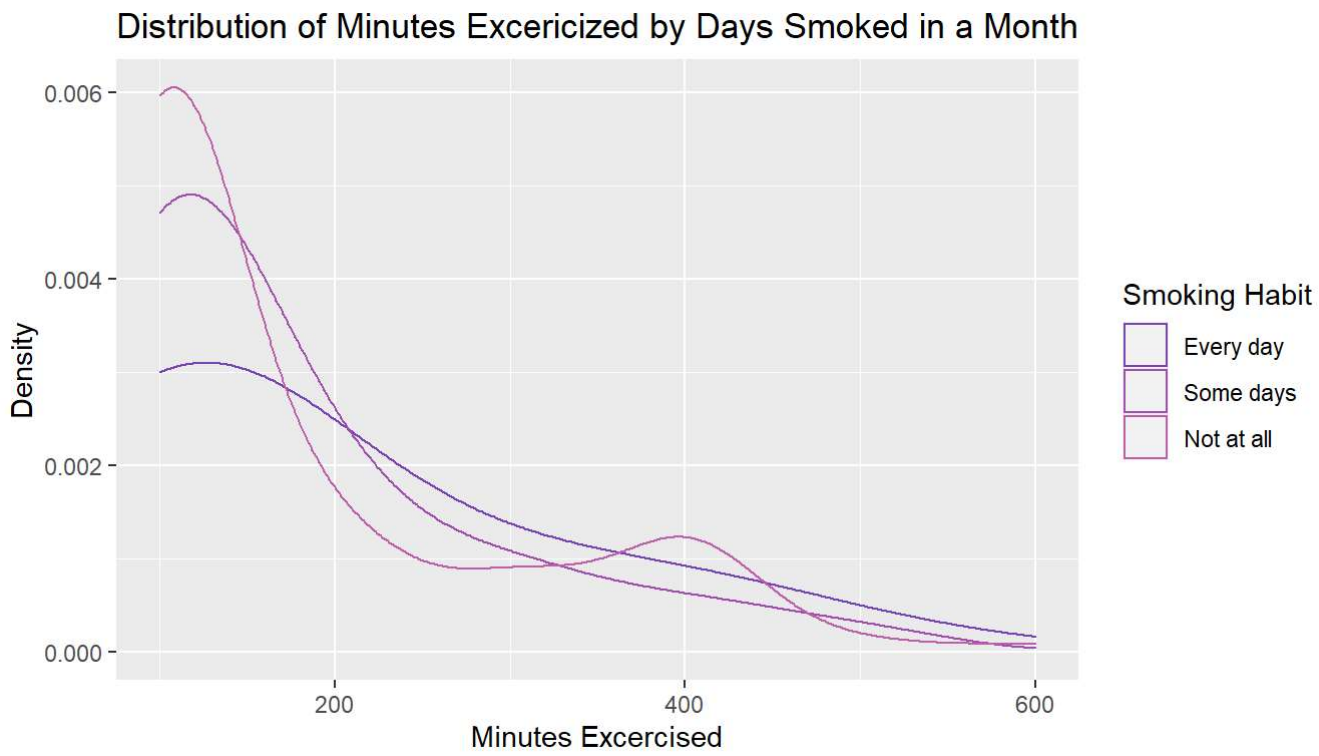
```
## `summarise()` regrouping output by 'smokday2' (override with `.groups` argument)
```

```
colnames(sum4)[1] <- "Frequency of days now smoking"
pandoc.table(sum4,caption="Summary Statistics for smoking habits and hours excericed in the last
        month")
```

```
##
## ------------------------------------------------------------------------
##  Frequency of days now smoking   100   110   115   120   125   130   145   200
## ------------------------------- ----- ----- ----- ----- ----- ----- ----- -----
##            Every day             55    NA     1     1    NA     8    NA    20
##
##            Some days             39    NA     4    NA     1    19    NA    15
##
##           Not at all             168    1     5     2    NA    43     5    34
## ------------------------------------------------------------------------
##
## Table: Summary Statistics for smoking habits and hours excericed in the last month (continued
below)
##
##
## ------------------------------------------------------
##  230   250   300   330   400   402   430   500   600
## ----- ----- ----- ----- ----- ----- ----- ----- -----
##   1    NA    10     2    11    NA     2     4     1
##
##   2    NA     9     1     5    NA     1     3    NA
##
##   5     1    22     6    37     1     4     4     3
## ------------------------------------------------------
```

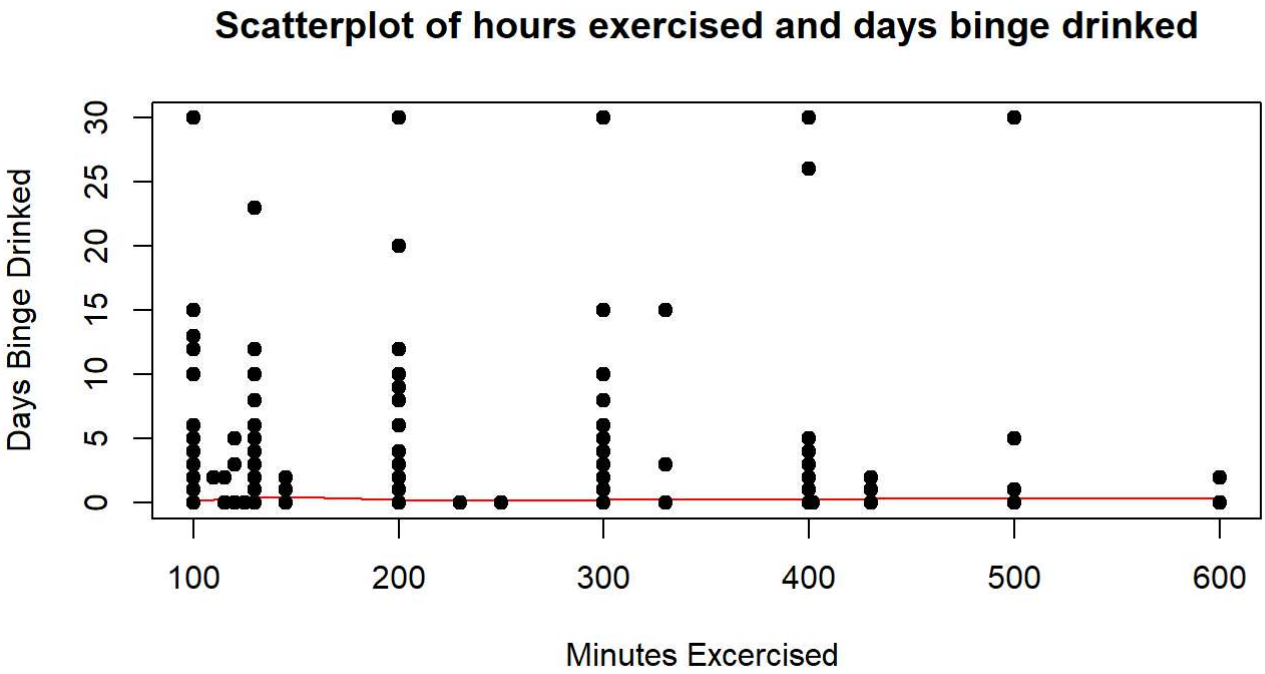
This table shows the frequency of days smoked and its relation to hours exercised. Not much can be immediately noticed from this table other than the fact that the majority of people excercising are excersing for 100 mins, however this doesn't matter for the question being asked.

```
ggplot(dat3, aes(x = exerhmm1, color = smokday2)) +
    geom_density(adjust = 2) +
    scale_color_manual(values = rev(heat_hcl(6, h = c(20,-80), l = c(75,40), c =c(40,80), power
          = 1))) +
    labs(title = "Distribution of Minutes Excericized by Days Smoked in a Month", y = "Density",
         x = "Minutes Excercised", col = "Smoking Habit") +
    theme(plot.title = element_text(hjust = 0.5))
```



We can see from the graph that while people who don't smoke at all have a higher density for 100 minutes, they end up having the lowest density for 250 minutes, and rise back up at 400 minutes.

```
plot(dat3$exerhmm1, dat3$drnk3ge5, main="Scatterplot of hours exercised and days binge drinked",
   xlab="Minutes Excercised ", ylab="Days Binge Drinked", pch=19, lines(lowess(dat3$exerhmm1, da
         t3$drnk3ge5), col = "red" ))
```



This scatterplot shows that people who binge drink for more than 15 days a month tend to exercise much less. The majority of exercisers tend to be 10 days or under, and even more for 5 days or fewer. However no direct correlation can be seen.

Summary

Initial Question: Is there a relation ship between minutes exercised and binge drinking and frequency of smoking?

Narrative from the Exploratory Analysis: No, there is no clear correlation for smoking habits and exercise or drinking habits and minutes of exercise.However we can see that people who binge drink more than 15 days a month tend to exercise much less often than those that binge drink less than 15 days.There is not a correlation for smoking and exercise either, and the reason for non smokers having a higher amount of exercises for the earlier minutes is due to the fact the majority of data consists of people who don't smoke. If there was a connection, then they would retain the highest density throughout the chart, but instead they fell off and on.