Observation on Physical Health according to 2013 BRFSS data

Code

PART 1: BRFSS Data

Behavioral Risk Factor and Surveillance System (BRFSS (https://www.cdc.gov/brfss/about/about_brfss.htm)) is a research controlled by the Centers for Disease Control (CDC). Initial point-in-time state surveys were conducted in 29 states from 1981–1983. In 1984, the Centres for Disease Control and Prevention (CDC) established the Behavioural Risk Factor Surveillance System (BRFSS), and 15 states participated in monthly data collection.

Currently data are collected monthly in all 50 states, the District of Columbia, American Samoa, Palau, Puerto Rico, the U.S. Virgin Islands, and Guam.

After the collection of data in a random based order and after a certain period all the data are sorted and based on their priority, they are made available to the general public. This study conducted is an observational study focused on the resident population of the United States.

General Bias Factors

This study adopts a complex strategy as to represent the entire U.S Population. Though it should be taken into account the Bias created by this sampling, some Examples are:

- Dishonest answering about the questions asked to the participants as it may result in being quite sensitive to answer correctly.
- Interview's held over the telephone tend to increase the Non-Responsiveness rate, as the participants may or may not be available to take calls or not intend to accept calls from highly unknown sources.
- The participants may or may not be interested in answering the questions and this would go on creating casual answering with False or wrong answers.

PART 2: Research Questions

> **Research quesion 1.**
> It is observed that deprivation of sleep affects the physical health of a person so, How does the amount of sleep interfere with the physical health of a person? A person who is intoxicated with either some drinks or smoking is more likely to be impacted, with their sleeping patterns being uneven. So accordingly, Does smoking, drinking and deprivation of sleep, deteriorates the health of a person, but keeping a good sleeping habit and continuing to drink and smoke doesn't affect the physical health that much?

Research quesion 2.
According to the BRFSS dataset of 2013, 15.58% persons are smokers and 12.6% persons are drinkers(Obtained by calculating the percentile of the difference between smokers and drinkers to non-smokers and non-drinkers). Drinking habbit is judged by Harmful drinkers (https://www.who.int/health-topics/alcohol#tab=tab_2) who drink more than two doses of alchol per day. Judging on this data, can it be highlighted with the fact that Physical health is direcly related to smoking and drinking. Alcohol has considerable toxic effects on the digestive- and cardiovascular systems. Alcoholic beverages are classified as carcinogenic by the International Agency for Research on Cancer and increase the risk of several cancer types. Alcohol as an immunosuppressant increases the risk of communicable diseases, including tuberculosis and HIV.

Research quesion 3.
Following the 2013 BRFSS dataset, can it be deduced that the income of an individual is related to his/her Physical health? Assuming, Smoking and drinking causes the physical health of a person to deteriorate. Smoking and drinking rate of a person prevails more in the high aspect of income level, i.e. Is it true that people with bad or deteriorating physical health varies between different income levels?

PART 3: Exploratory Data Analysis (EDA)

Load Packages

Hide

```
library(statsr)      # Companion packages for statistics in R
library(dplyr)       # Used for making data manipulation smooth and easy
library(ggplot2)     # Visulization Package
library(tidyr)       # Cmpliment od dyplr package
library(stringr)     # To deal with String Variables
library(gridExtra)   # To arrange multiple grid based plots on a single page
library(scales)      # Customise and control the appearance of axis
```

Load DataSet

Hide

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

Data Sorting for Question 1

For Question 1 he Variable `sleptim1` represents the sleeping time of individuals in hours. The sleeping time of a normal adult is about 7 to 9 hours per day.The National Sleep Foundation (https://www.sleepfoundation.org/press-release/national-sleep-foundation-recommends-new-sleep-times) recommends 7 to 9 hours of sleep per night in order to stay healthy. For this, a categorical variable `sleep.time` is assigned with three different levels in order to identify if the Amount of hours spent sleeping is less, adequate or high.

- 1 to 6 hours: Less
- 7 to 9 hours: Adequate
- 10 to 24 hours: High

Sleep plays an important role in your physical health. For example, sleep is involved in healing and repair of your heart and blood vessels. Ongoing sleep deficiency is linked to an increased risk of heart disease, kidney disease, high blood pressure, diabetes, and stroke.

Data Sorting for Question 2

In order to anayze data for Question 2,

- People who smoke will be categorized into a variable called `smoker`
- People who drink harmfully according to the WHO database will be grouped into a varible called `drinker` All other factors are discarded from this part for the time being.

Data Sorting for Question 3

The `income2` varible from the databse will be used which displays different income levels of the people.

Steps To Prepapre Data For Data Analysis

In order to prepare our data the following steps are required to be followed;

- Load the Packages & Data File.
- Filter out unwanted & Not Available(NA) values in `physhlth` .
- Transform `X_smoker 3` into a binary variable called `smoker` , identifying who Smokes and who does not Smoke.
- Transform `avedrnk2` into a binary variable called `drinker` , identifying who has a drinking behaviour and who doesn't. As because drinking affects sleep and Physical health.
- Create a categorical variable to classify `sleptim1` into 3 different sleeping behaviours, based on the number of hours slept per day.
- Finally, select the group of variables required for this analysis.

Hide

```
vtemp <- brfss2013 %>%
  filter(physhlth <= 30) %>%
  mutate(physical.illness = as.numeric(physhlth > 0),
         illness.level = cut(physhlth, c(-1, 0, 3, 10, 25, 30),
                             c("no disorder", "low", "moderate", "severe", "extreme")),
         smoker = if_else(X_smoker3 == "Former smoker" | X_smoker3 == "Never smoked",
  0 , 1),
         drinker = as.numeric(avedrnk2 > 2),
         sleep.time = if_else(sleptim1 < 7, "Less", if_else(sleptim1 > 9, "High", "Ad
  equate")),
         sleep.time = factor(sleep.time, levels = c("Less", "Adequate", "High"))) %>%
  select(physhlth,physical.illness,illness.level,# 3 base variables
         sleptim1, sleep.time,# 2 variables for question 1
         smoker, drinker,  # 2 variables for question 2
         income2, X_incomg) # 2 variables for question 3
```

Analysis

Research Questin 1

How does sleep impact the physical health of a person?
People with habbits of smoking and drinking and sleep deprivation are likely to prohibit physical disorder than people who smoke or drink but sleep properly.

Sleep plays an important role in your physical health. For example, sleep is involved in healing and repair of your heart and blood vessels. Ongoing sleep deficiency is linked to an increased risk of heart disease, kidney disease, high blood pressure, diabetes, and stroke.

Hide

```
mean1.sleep <- mean(vtemp$sleptim1, na.rm = TRUE)# mean sleep value = 7.049
```

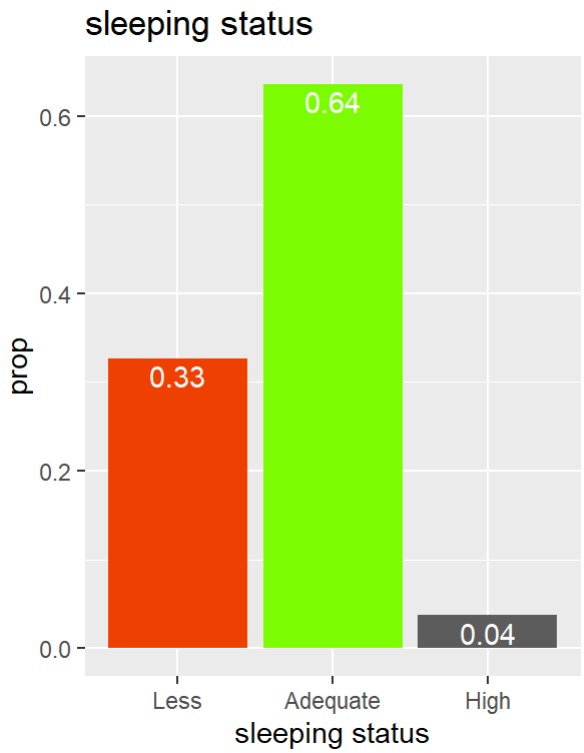
The aveage sleeping time of people in the U.S. is 7.05 hours per day but according to the National sleep foundation (https://www.sleepfoundation.org/press-release/national-sleep-foundation-recommends-new-sleep-times) the recommended time of sleep per night is 7 to 9 hours.

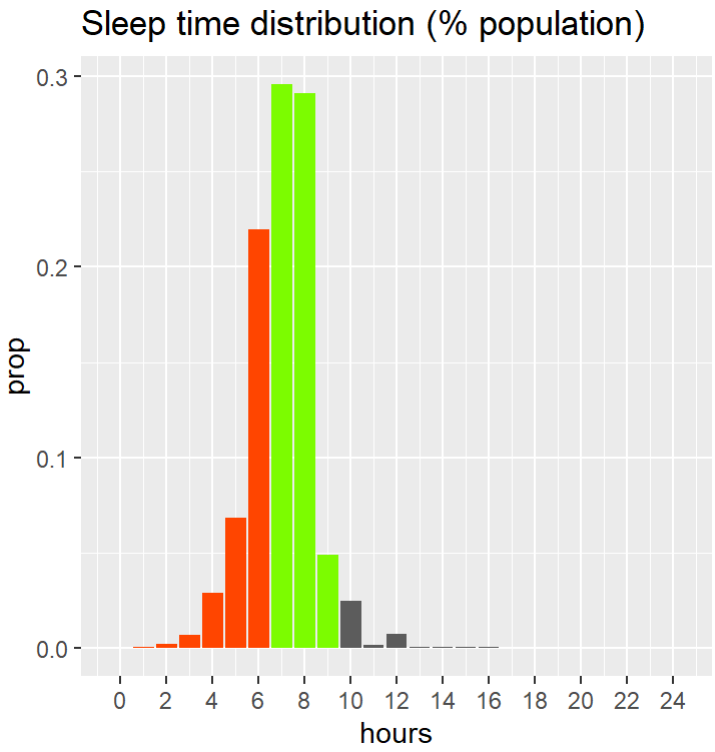
Hide

```
v1 <- vtemp %>%
  filter(!is.na(sleep.time)) %>%
  count(sleep.time) %>%
  mutate(prop = prop.table(n)) %>%
  ggplot(aes(x = sleep.time, y = prop)) +
  geom_bar(stat = "identity", fill = c("orangered2", "lawngreen", "grey36")) +
  geom_text(aes(label = round(prop, 2)), vjust = 1.3, color = "white") +
  labs(title = "sleeping status", x = "sleeping status")

v2 <- vtemp %>%
  filter(!is.na(sleptim1)) %>%
  ggplot(aes(x = sleptim1, y = ..prop..)) +
  geom_bar(fill = c(rep("orangered", 7), rep("lawngreen", 3), rep("grey36", 15))) +
  scale_x_continuous(breaks = seq(0, 24, 2)) +
  labs(title = "Sleep time distribution (% population)", x = "hours")

grid.arrange(grobs = list(v1, v2), ncol = 2, widths = c(.45, .55))
```





Grouping all these together:

Hide

```
vtemp %>%
  filter (!is.na(sleep.time)) %>%
  group_by(sleep.time) %>%
  summarise(avg.hour= round(mean(sleptim1),1))
```

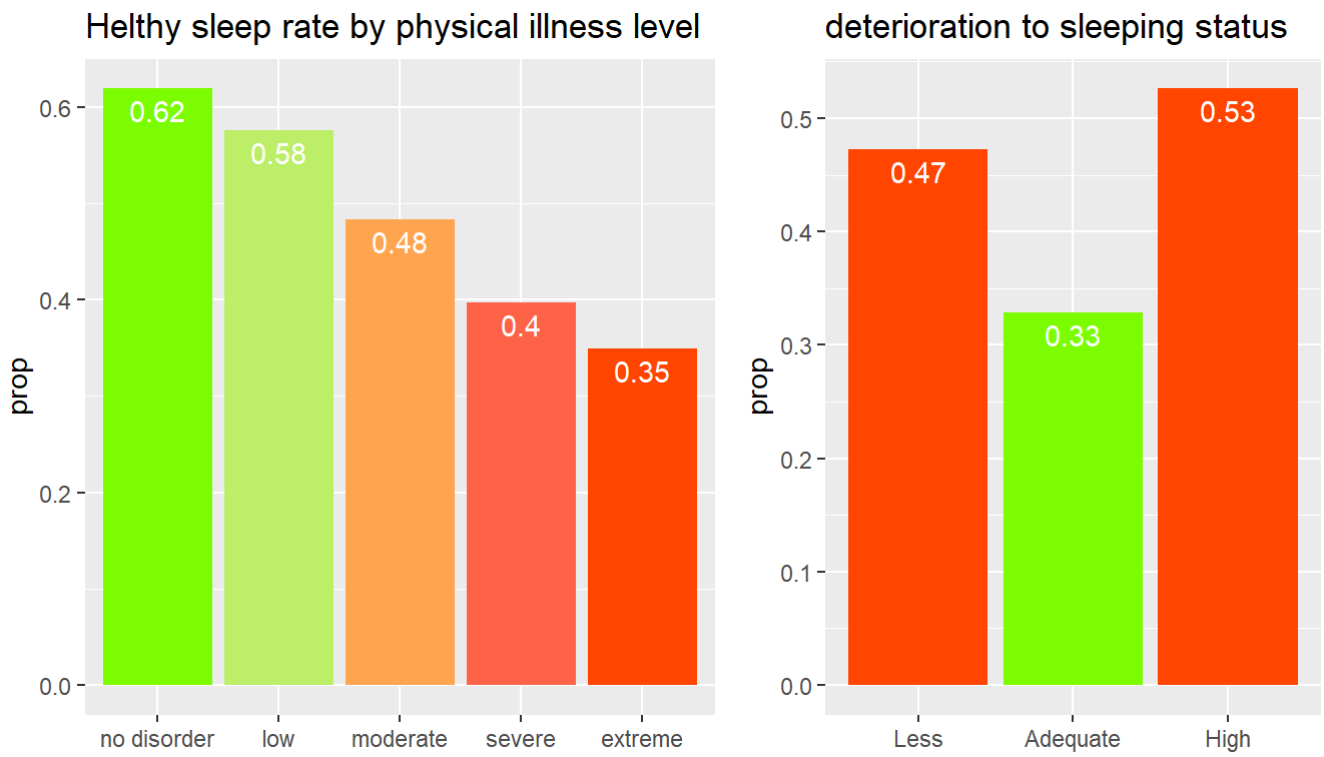
```
## # A tibble: 3 x 2
##   sleep.time avg.hour
##   <fct>         <dbl>
## 1 Less            5.5
## 2 Adequate        7.6
## 3 High           11
```

According to the dataset obtained 64% of people sleep for an average of 7.6 hours, 33% of people sleep for an average of 5.5 hours and 4% of people sleep for an average of 11 hours each day.

Hide

```
v <- list(
  vtemp %>%
    filter(!is.na(sleep.time), smoker == 1 | drinker == 1) %>%
    group_by(illness.level) %>%
    summarise(prop = mean(sleep.time == "Adequate")) %>%
    ggplot(aes(x = illness.level, y = prop)) +
    geom_bar(stat = "identity", fill = c("lawngreen","darkolivegreen2","tan1", "tomat
  o", "orangered")) +
    geom_text(aes(label = round(prop, 2)), vjust = 1.5, color = "white") +
    labs(title = "Helthy sleep rate by physical illness level", x = NULL),
  vtemp%>%
    filter(!is.na(sleep.time), smoker == 1 | drinker == 1) %>%
    group_by(sleep.time) %>%
    summarise(prop = mean(physical.illness))  %>%
    ggplot(aes(x = sleep.time, y = prop)) +
    geom_bar(stat = "identity", fill = c("orangered", "lawngreen", "orangered")) +
    geom_text(aes(label = round(prop, 2)), vjust = 1.5, color = "white") +
    labs(title = "deterioration to sleeping status", x = NULL)
  )

grid.arrange(grobs = v, ncol = 2, widths = c(.55, .45))
```



Graph showing sleeping habbits of different people. ##### **Conclusions**
The anaysis done above depicts well defined insights on the relationships between sleep, tobacco & alcohol consumption & Physical disorder.

- It can be proved that good sleep per day educes the risks of having a bad physical health.
- Physical health is directly linked with smoking and alcohol consumption.

Research Questin 2

Is Physical Health related to smoking and drinking? According to the BRFSS dataset of 2013, 15.58% persons are smokers and 12.6% persons are drinkers. If the dataset is observed then it can be seen that the smoking rates of people in the US in 2013 were 16% and the harmful drinking rate is 24%(Rounded off

to the nearest value for the sake of eliminating confusions.)# Original data is used while computing.

Creating the variables required and assigning the required values to them.

Hide

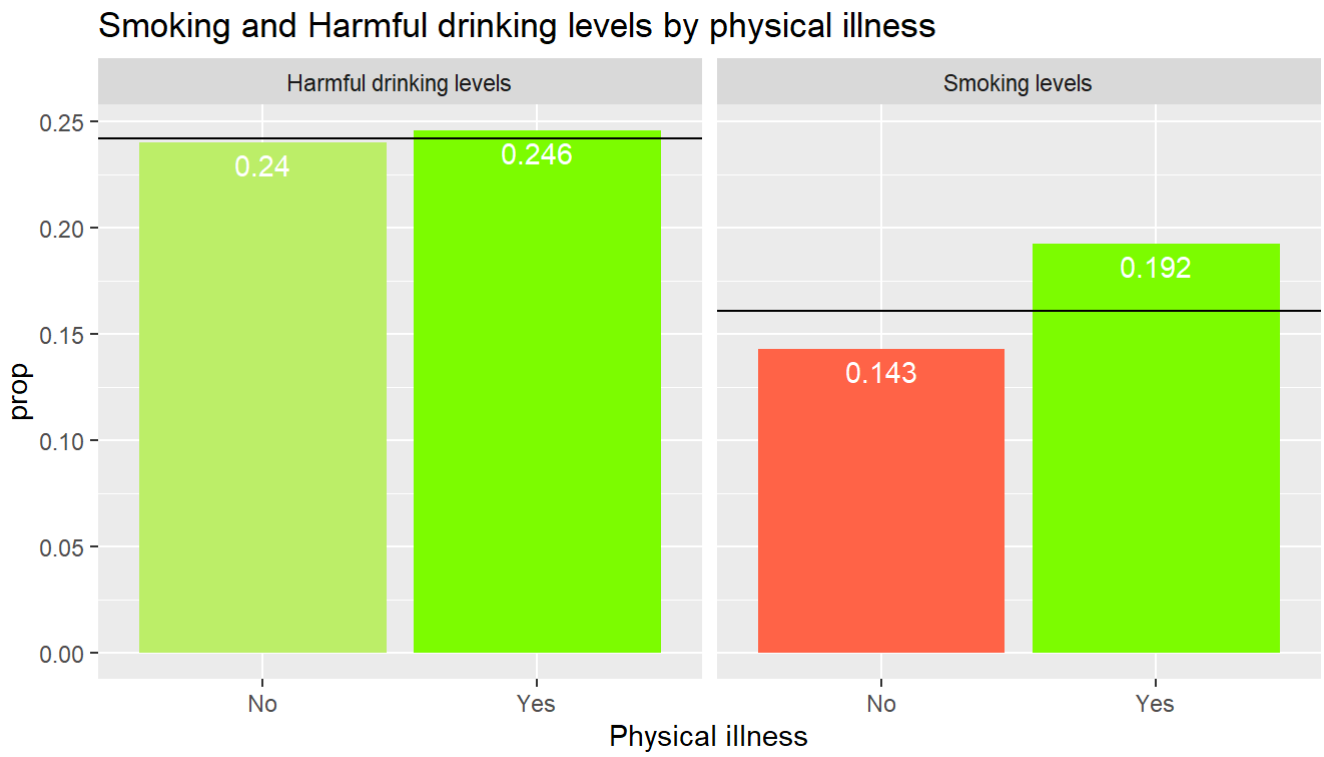
```
level.physical <- mean(vtemp$physhlth > 0)
level.smoker <- mean(vtemp$smoker, na.rm = TRUE)
level.drinker <- mean(vtemp$drinker, na.rm = TRUE)
```

Creating the required graphical plots to show if smoking and drinking rates is related to Physically Disturbed health.

Hide

```
physical.illness<- vtemp %>%
  group_by(physical.illness = ifelse(physhlth > 0, "Yes", "No")) %>%
  summarise("Harmful drinking levels" = mean(drinker, na.rm = TRUE),
            "Smoking levels" = mean(smoker, na.rm = TRUE))

physical.illness%>%
  gather(key, prop, - physical.illness) %>%
  ggplot(aes(x = physical.illness, y = prop)) +
  geom_bar(stat = "identity", fill = c("darkolivegreen2", "lawngreen","tomato", "lawn
 green")) +
  geom_text(aes(label = round(prop, 3)), vjust = 1.5, color = "white") +
  geom_hline(aes(yintercept = c(rep(level.drinker, 2), rep(level.smoker, 2)))) +
  labs(title = "Smoking and Harmful drinking levels by physical illness ",
       x = "Physical illness") +
  facet_grid(~key)
```



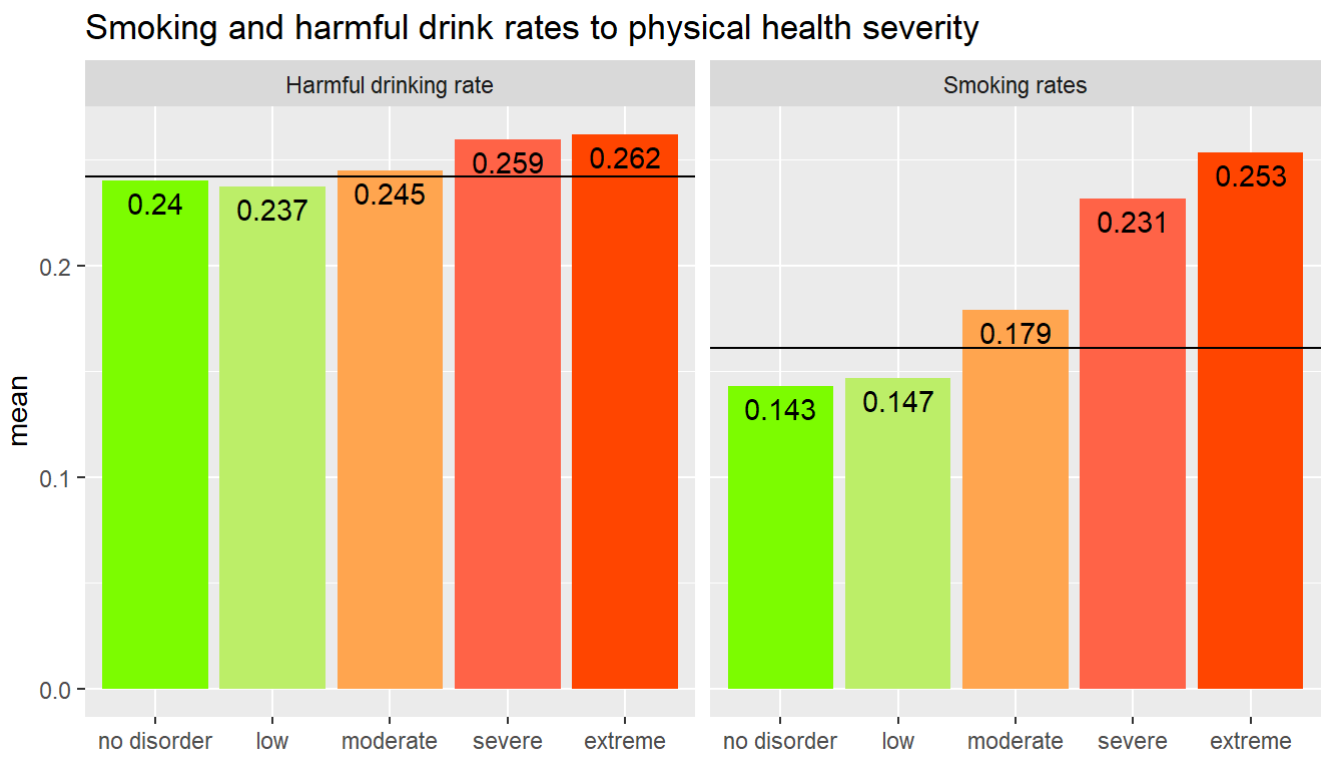
The graphical plots obtained above depicts that the persons with no physical illness has very little or negigible effect on drinking rates compared to the persons who has a physical illness. On the otherhand it is observed that a person who smokes has a considerable effect on the illness level of physical health whereas the person who desent smoke remains normal to the plot.

Plots to show evidential difference between smoking/drinking levels and physical health:

Hide

```
physical.illness.drugs <- vtemp %>%
  group_by(illness.level) %>%
  summarise("Harmful drinking rate" = mean(drinker, na.rm = TRUE),
            "Smoking rates" = mean(smoker, na.rm = TRUE))

physical.illness.drugs %>%
  gather(key, mean, -illness.level) %>%
  ggplot(aes(x =illness.level, y = mean)) +
  geom_bar(stat = "identity",fill = c("lawngreen","darkolivegreen2","tan1", "tomato",
  "orangered","lawngreen","darkolivegreen2","tan1", "tomato", "orangered")) +
  geom_text(aes(label = round(mean, 3)), vjust = 1.5, color = "black") +
  geom_hline(aes(yintercept = c(rep(level.drinker, 5), rep(level.smoker, 5)))) +
  labs(title = "Smoking and harmful drink rates to physical health severity", x = NUL
  L) +
  facet_grid(~key)
```



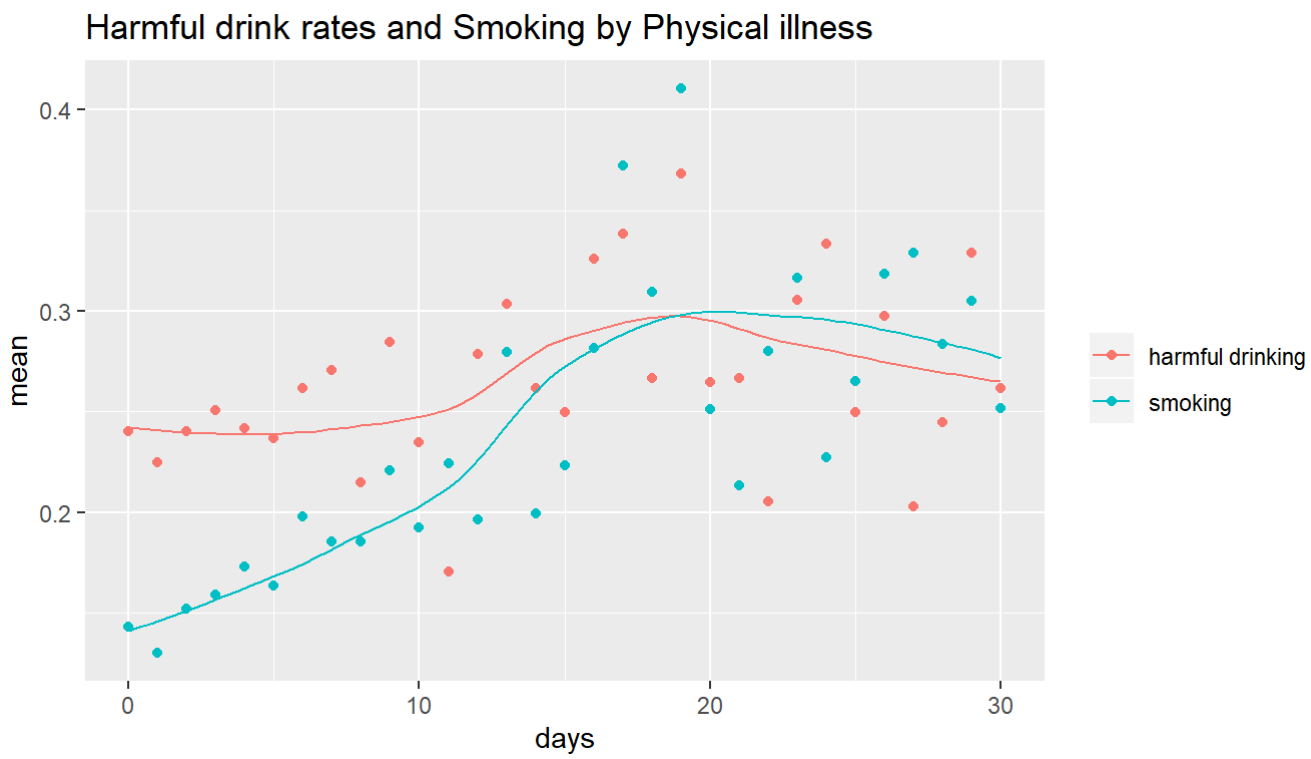
The plots given above depicts that harmful drinking levels and smoking varies with the physical illness duration. Among the people who were physically healthy 14.3% were smokers and 24% drank harmfully (i.e. 2 alcohol duses per day). Whereas in the bar of extreme physical illness, it can be seen that 26.2% of people were harmful drinkes and 25.3% of people were smokers.

Scaterplot for Comparing harmful drinking rates with physical illness

Hide

```
illness.level.drugs <- vtemp%>%
  select(physhlth, smoker, drinker) %>%
  gather(key, value, -physhlth) %>%
  group_by(physhlth, key) %>%
  summarise(mean = mean(value, na.rm = TRUE))

ggplot(illness.level.drugs, aes(x = physhlth, y = mean, color = key)) +
  geom_point() +
  stat_smooth(method = "loess", se = FALSE, size = .5) +
  scale_color_discrete(NULL, label = c("harmful drinking", "smoking")) +
  labs(title = "Harmful drink rates and Smoking by Physical illness", x = "days")
```



Conclusions

According to the data obtained above from the graph plots it can be said that:

- smoking and drinking affects meantal health.
- Alcohol consumption may not be of much complaint in the lower levels of illness but at the higher levels it can be seen that the alcohol consumption has increased considerably.
- Good physical health van be maintained if smoking and drinking rates can be reduced.
- After around 19 - 20 days of bad physical health drinking rates decreases a bit, whereas the smoking rates increases till 20th and then the ratre of decreasing is slow.

Research Questin 3

Is the income of an individual related to his/her Physical health? Following the 2013 BRFSS dataset,Smoking and drinking rate of a person prevails more in the high aspect of income level, i.e. Is it true that people with bad or deteriorating physical health varies between different income levels?

Studying the plots that we get we then would be able to deduce the answers to these questions.

Hide

```
vtemp %>%
  filter(!is.na(income2)) %>%
  group_by(income2) %>%
  summarise(smoking = sd(smoker, na.rm = TRUE),
            "Harmful drink" = sd(drinker, na.rm = TRUE))
```

```
## # A tibble: 8 x 3
##   income2           smoking `Harmful drink`
##   <fct>               <dbl>           <dbl>
## 1 Less than $10,000   0.461           0.491
## 2 Less than $15,000   0.445           0.477
## 3 Less than $20,000   0.429           0.473
## 4 Less than $25,000   0.410           0.456
## 5 Less than $35,000   0.388           0.445
## 6 Less than $50,000   0.367           0.433
## 7 Less than $75,000   0.337           0.423
## 8 $75,000 or more     0.278           0.400
```

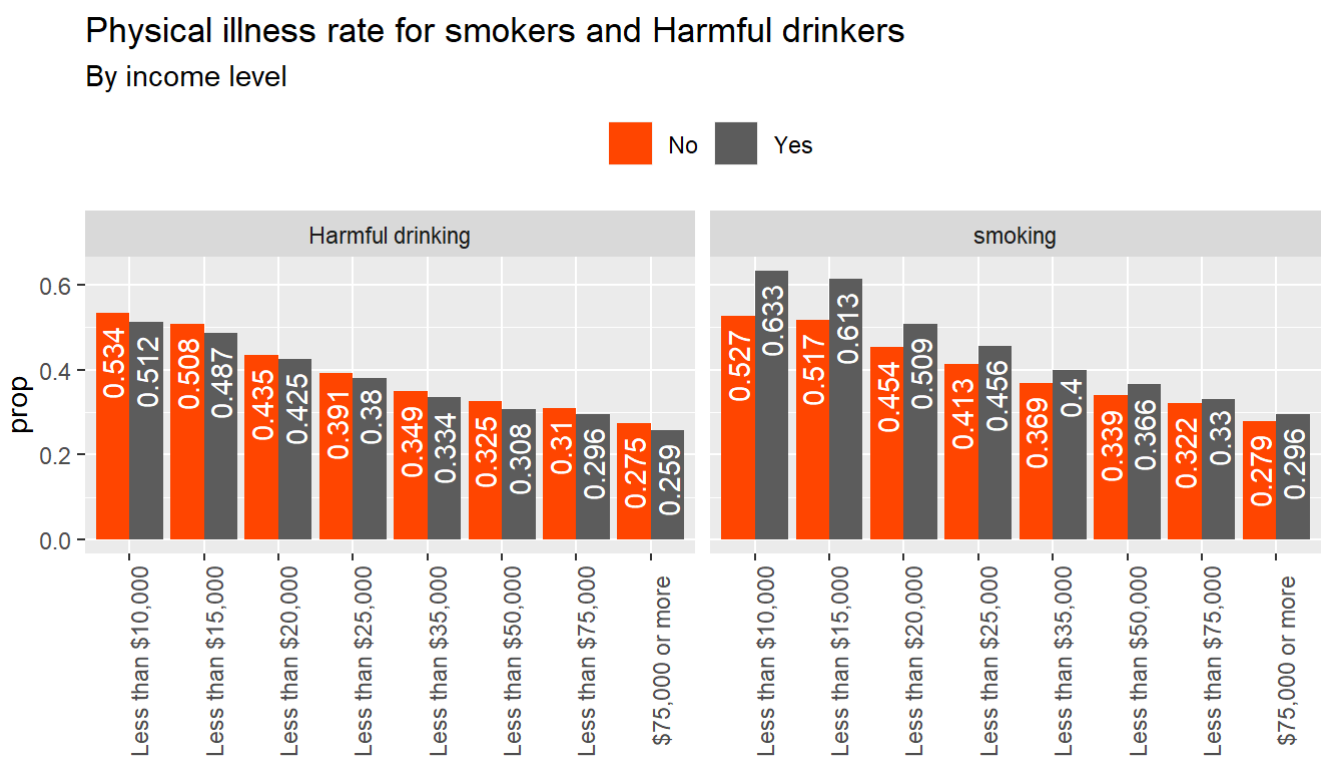
The next graoh is shown in order to compare physical illness rate with harmful drinking and smoking according to the income levels.

Hide

```
drinkers.income <- vtemp %>%
  filter(!is.na(drinker), !is.na(income2)) %>%
  group_by(income2, act = drinker) %>%
  summarise(physical.level = mean(physical.illness)) %>%
  mutate(behavior = "Harmful drinking")

smokers.income <- vtemp %>%
  filter(!is.na(smoker), !is.na(income2)) %>%
  group_by(income2, act = smoker) %>%
  summarise(physical.level = mean(physical.illness)) %>%
  mutate(behavior = "smoking")

rbind(drinkers.income, smokers.income) %>%
  ggplot(aes(x = income2, y = physical.level, fill = ifelse(act == 1, "Yes", "No")))
  +
  geom_bar(stat = "identity", position = position_dodge(width = .9)) +
  geom_text(aes(label = round(physical.level, 3)), hjust = 1.2, angle = 90,
            color = "white", position = position_dodge(width = .9)) +
  scale_fill_manual(values = c("orangered", "grey36"), NULL) +
  labs(title = "Physical illness rate for smokers and Harmful drinkers",
       subtitle = "By income level", x = NULL, y = "prop") +
  facet_grid(~behavior) +
  theme(legend.position = "top",
        axis.text.x = element_text(angle = 90))
```



We can see that harmful drinkers and smokers decrease as the income level of the individuals increase.

Conclusion

- As the per capita income decreases the severity of the physical illness increases.
- Variablity for Harmful drinking and smoking rates decreases in the higher income levels.
- Physical disorder and income is negatively associated.

After all these we can assume that the data provided by BRFSS depicts the fact that people who sleep healthy leads a healthy physical life.