Exploring the BRFSS data

## Setup

### Load packages

library(ggplot2)  
library(dplyr)  
library(corrplot)

### Load data

load("brfss2013.RData")

## Part 1: Data

The Behavioral Risk Factor Surveillance System (BRFSS) is the system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services.

# Sample Collection

The samples were collected by land-line telephonic surveys and cell-phones telephonic surveys. In conducting the BRFSS landline telephone survey, interviewers collect data from a randomly selected adult in a household. In conducting the cellular telephone version of the BRFSS questionnaire, interviewers collect data from an adult who participates by using a cellular telephone and resides in a private residence or college housing. Disproportionate stratified sampling was used for land-line telephones. And the cellular phone sample is randomly generated from the combination od area code and prefix.

# Inference

If we look at the sample collection methods, we would conclude that this is an observational study. It uses stratified sampling method based on random digit dialing methods.

# Generabizability

Using this type of random digit dialing and Disproportionate stratified sampling, we can canclude that this study can be generalized to the population of all non-institutionalized adults of 18 years and above

# Causality

As this is an observational study, so we can not be sure about causality. We can only be sure about associations.

## Part 2: Research questions

**Research quesion 1:**

How does the sleeping time (duration) affect the general health. In other words, is there any association between variable genhlth: General Health and sleptim1: How Much Time Do You Sleep. This question is of an interest because generally, good sleeping habits might be associated with general good health.

**Research quesion 2:**

Generally, people with reasonable consumption of vegetables and fruits per day are healthier. It would be interesting to know that what is the association between vegetables and fruits comsumption with the general health. In other words, association between frutsum variable and vegesum variable with the variable genhlth.

**Research quesion 3:**

Is there any association between bmi5cat: Computed Body Mass Index catefory and the consumption of vegetables and fruits

It would be interesting to know about how the BMI category is affected by vegetable and fruits intake.

## Part 3: Exploratory data analysis

**Research quesion 1 :**

The mean and STD of sleeping hours for different general health categories are computed. The following code shows the summary statistics of sleeping hours for four general health categories.

df1 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Poor') %>% filter(sleptim1 !='NA') %>% summarise(mean\_poorhealth\_sleep = mean(sleptim1), sd\_poorhealth\_sleep = sd(sleptim1), n = n())  
df1

## mean\_poorhealth\_sleep sd\_poorhealth\_sleep n  
## 1 6.737152 2.391026 26639

df2 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Good') %>% filter(sleptim1 !='NA') %>% summarise(mean\_goodhealth\_sleep = mean(sleptim1), sd\_goodhealth\_sleep = sd(sleptim1), n = n())  
df2

## mean\_goodhealth\_sleep sd\_goodhealth\_sleep n  
## 1 7.038402 1.442974 148299

df3 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Very good') %>% filter(sleptim1 !='NA') %>% summarise(mean\_vgoodhealth\_sleep = mean(sleptim1), sd\_vgoodhealth\_sleep = sd(sleptim1), n = n())  
df3

## mean\_vgoodhealth\_sleep sd\_vgoodhealth\_sleep n  
## 1 7.103533 1.206879 157833

df4 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Excellent') %>% filter(sleptim1 !='NA') %>% summarise(mean\_excellenthealth\_sleep = mean(sleptim1), sd\_excellenthealth\_sleep = sd(sleptim1), n = n())  
df4

## mean\_excellenthealth\_sleep sd\_excellenthealth\_sleep n  
## 1 7.18968 1.214332 84822

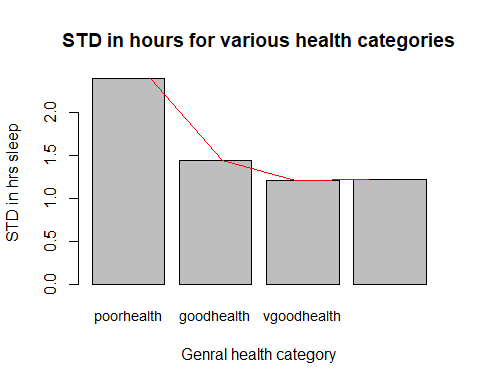
mean\_sleep\_vector <- c(df1$mean\_poorhealth\_sleep,df2$mean\_goodhealth\_sleep,df3$mean\_vgoodhealth\_sleep,df4$mean\_excellenthealth\_sleep)  
mean\_sleep\_vector

## [1] 6.737152 7.038402 7.103533 7.189680

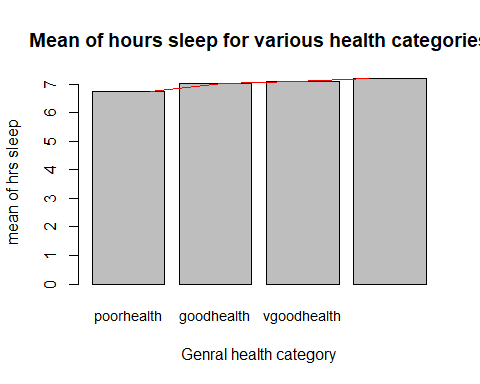
sd\_sleep\_vector <- c(df1$sd\_poorhealth\_sleep,df2$sd\_goodhealth\_sleep,df3$sd\_vgoodhealth\_sleep,df4$sd\_excellenthealth\_sleep)  
sd\_sleep\_vector

## [1] 2.391026 1.442974 1.206879 1.214332

names <- c("poorhealth", "goodhealth", "vgoodhealth", "excellenthealth")  
barplot(sd\_sleep\_vector, names.arg=names, cex.names=.9, ylab = 'STD in hrs sleep', main='STD in hours for various health categories', xlab="Genral health category")   
lines(sd\_sleep\_vector, col='Red')



barplot(mean\_sleep\_vector, names.arg=names, cex.names=.9, ylab = 'mean of hrs sleep', main='Mean of hours sleep for various health categories', xlab="Genral health category")   
lines(mean\_sleep\_vector, col='Red')



For excellent, good and very good health category, STD for sleeping hours is relatively smaller than the poor general health category. Similarly, the mean of sleeping hours for excellent, good and very good health category is more than the mean sleeping hours of poor general health category people.

From the summary statistics and the plots, it is shown the generally poor health is associated with less sleep. Though, this difference is not huge, yet general health is slightly associated with the sleep.

People with around 7 hours per day sleep are more healtheir than the people with less than 7 hours per day sleep.

**Research quesion 2:**

Below, we have computed the mean consumption of fruits and vegetables for all four general health categories.

df5 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Poor') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_poorhealth\_veg = mean(X\_vegesum))  
df5

## mean\_poorhealth\_veg  
## 1 162.3957

df6 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Good') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_goodhealth\_veg = mean(X\_vegesum))  
df6

## mean\_goodhealth\_veg  
## 1 179.8734

df7 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Very good') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_vgoodhealth\_veg = mean(X\_vegesum))  
df7

## mean\_vgoodhealth\_veg  
## 1 197.7161

df8 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Excellent') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_excellenthealth\_veg = mean(X\_vegesum))  
df8

## mean\_excellenthealth\_veg  
## 1 221.3981

df9 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Poor') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_poorhealth\_fruit = mean(X\_frutsum))  
df9

## mean\_poorhealth\_fruit  
## 1 122.9033

df10 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Good') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_goodhealth\_fruit = mean(X\_frutsum))  
df10

## mean\_goodhealth\_fruit  
## 1 132.0273

df11 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Very good') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_vgoodhealth\_fruit = mean(X\_frutsum))  
df11

## mean\_vgoodhealth\_fruit  
## 1 143.8004

df12 <-brfss2013 %>% filter(genhlth != "NA") %>% filter(genhlth == 'Excellent') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_excellenthealth\_fruit = mean(X\_frutsum))  
df12

## mean\_excellenthealth\_fruit  
## 1 161.9623

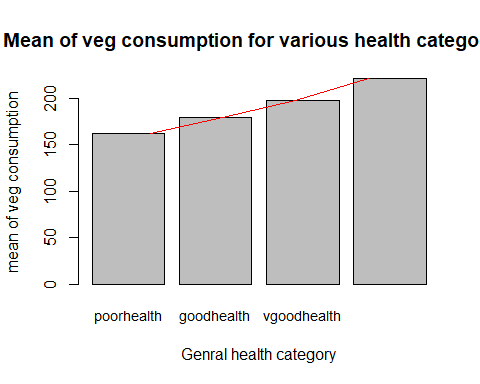
It is shown that people who consumes more vegetable and fruits are generally healthier.

We also plot the vegetable and fruit mean consumption for various general helath categories

mean\_veg\_vector <- c(df5$mean\_poorhealth\_veg,df6$mean\_goodhealth\_veg,df7$mean\_vgoodhealth\_veg,df8$mean\_excellenthealth\_veg)  
mean\_veg\_vector

## [1] 162.3957 179.8734 197.7161 221.3981

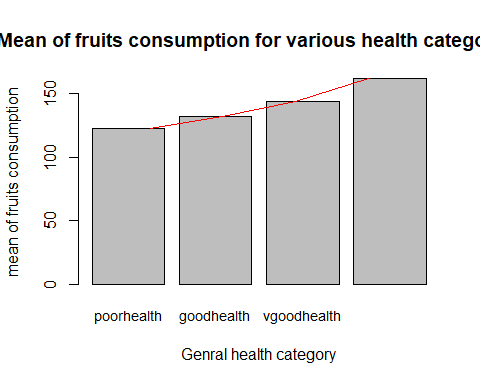
names <- c("poorhealth", "goodhealth", "vgoodhealth", "excellenthealth")  
barplot(mean\_veg\_vector, names.arg=names, cex.names=.9, ylab = 'mean of veg consumption', main='Mean of veg consumption for various health categories', xlab="Genral health category")   
lines(mean\_veg\_vector, col='Red')



mean\_fruit\_vector <- c(df9$mean\_poorhealth\_fruit,df10$mean\_goodhealth\_fruit,df11$mean\_vgoodhealth\_fruit,df12$mean\_excellenthealth\_fruit)  
mean\_fruit\_vector

## [1] 122.9033 132.0273 143.8004 161.9623

names <- c("poorhealth", "goodhealth", "vgoodhealth", "excellenthealth")  
barplot(mean\_fruit\_vector, names.arg=names, cex.names=.9, ylab = 'mean of fruits consumption', main='Mean of fruits consumption for various health categories', xlab="Genral health category")   
lines(mean\_fruit\_vector, col='Red')

 In both the above plots, we see a clear association of better health and more consumption of fruits and vegetables.

**Research quesion 3:**

Here we provide the mean veg and fruits consumption for various BMI categories.

df13 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Underweight') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_Underweight\_veg = mean(X\_vegesum))  
df13

## mean\_Underweight\_veg  
## 1 193.1601

df14 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Normal weight') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_NormalWeight\_veg = mean(X\_vegesum))  
df14

## mean\_NormalWeight\_veg  
## 1 202.8943

df15 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Overweight') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_Overweight\_veg = mean(X\_vegesum))  
df15

## mean\_Overweight\_veg  
## 1 187.7619

df16 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Obese') %>% filter(X\_vegesum !='NA') %>% summarise(mean\_Obese\_veg = mean(X\_vegesum))  
df16

## mean\_Obese\_veg  
## 1 177.8585

df17 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Underweight') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_Underweight\_fruit = mean(X\_frutsum))  
df17

## mean\_Underweight\_fruit  
## 1 142.9423

df18 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Normal weight') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_NormalWeight\_fruit = mean(X\_frutsum))  
df18

## mean\_NormalWeight\_fruit  
## 1 149.9006

df19 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Overweight') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_Overweight\_fruit = mean(X\_frutsum))  
df19

## mean\_Overweight\_fruit  
## 1 138.3816

df20 <-brfss2013 %>% filter(X\_bmi5cat != "NA") %>% filter(X\_bmi5cat == 'Obese') %>% filter(X\_frutsum !='NA') %>% summarise(mean\_Obese\_fruit = mean(X\_frutsum))  
df20

## mean\_Obese\_fruit  
## 1 127.9404

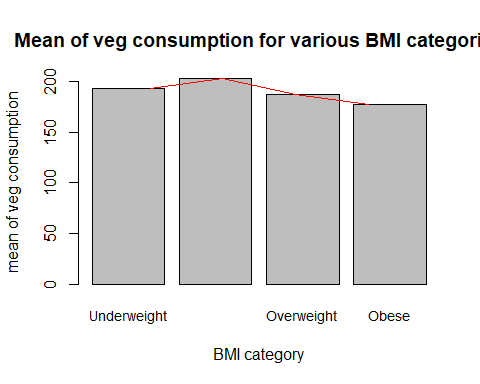
The mean fuits and vegetable consumption for normal weight people is higher.

Now we provide the plotting to see any pattern (if there is).

mean\_veg\_vector\_bmi <- c(df13$mean\_Underweight\_veg,df14$mean\_NormalWeight\_veg,df15$mean\_Overweight\_veg,df16$mean\_Obese\_veg)  
mean\_veg\_vector\_bmi

## [1] 193.1601 202.8943 187.7619 177.8585

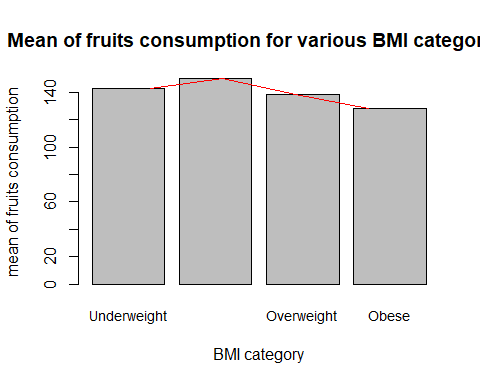
names <- c("Underweight", "NormalWeight", "Overweight", "Obese")  
barplot(mean\_veg\_vector\_bmi, names.arg=names, cex.names=.9, ylab = 'mean of veg consumption', main='Mean of veg consumption for various BMI categories', xlab="BMI category")   
lines(mean\_veg\_vector\_bmi, col='Red')



mean\_fruit\_vector\_bmi <- c(df17$mean\_Underweight\_fruit,df18$mean\_NormalWeight\_fruit,df19$mean\_Overweight\_fruit,df20$mean\_Obese\_fruit)  
mean\_fruit\_vector\_bmi

## [1] 142.9423 149.9006 138.3816 127.9404

names <- c("Underweight", "NormalWeight", "Overweight", "Obese")  
barplot(mean\_fruit\_vector\_bmi, names.arg=names, cex.names=.9, ylab = 'mean of fruits consumption', main='Mean of fruits consumption for various BMI categories', xlab="BMI category")   
lines(mean\_fruit\_vector\_bmi, col='Red')



It can be seen that normal weight BMI category people consume relatively more fruits and vegetables. On the average, obese people consume less vegetables and fruits. This relation of fruit and vegetable consumption with the BMI category is associational.