title: “Exploring the BRFSS data” output: html_document: fig_height: 4 highlight: pygments theme: spacelab —

Setup

Load packages

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.6.2

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.6.2
```

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 purpose of this document is to provide information for completion of the data analysis project for week 5 of the “Inferential Statistics” course as part of “Statistics using R” specialization by Duke University offered on Coursera.

The GSS dataset consists of observational study of societal change and growing complexity of the American Society. This dataset is very extensive and have around 491775 rows of 330 variables. The variables includes are also very diverse and provides lot of information about the dataset.

Generalizability: Due to data gathering techniques used for collecting the data from the participants of the survey, random sampling techinque can be used and the results can be generalized to the entire population of US

Causality: It was an observational study and no random assignment into groups was done. So, it wont be possible to make any casual inferences

---

Part 2: Research questions

Research question 1: Is there any correlation between the consumption of alcohol (average consumption) and anxiety level for past 30 days and does gender plays any vital role.

Research quesion 2: Is there any correlation between the emlployment status and havimg any health care coverage?

Research quesion 3: Is there any correlation between number of children at home and a person has exercised in last 30 days? Is it same for both genders?

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:

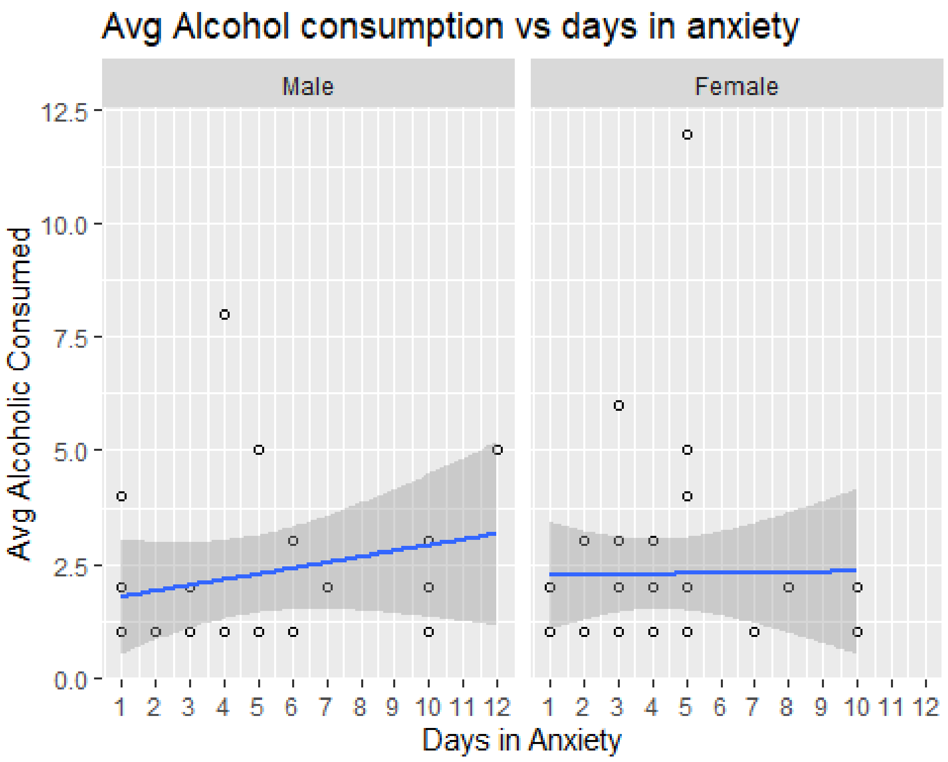
```
data1 <- select(brfss2013, qlstres2 , sex , avedrnk2) %>%
  filter(!is.na(qlstres2), !is.na(sex), avedrnk2 <= 12 )

# Plot relevant variables

ggplot(data = data1, aes(x = qlstres2, y = avedrnk2 ))+
  geom_point(shape=1) +
  geom_smooth(method=lm)    +
  scale_x_continuous(limits = c(1,12), breaks = 1:12) +
  facet_grid(. ~  sex) +
  ggtitle("Avg Alcohol consumption vs days in anxiety") +
  xlab ("Days in Anxiety") +
  ylab("Avg Alcoholic Consumed")
```

```
## Warning: Removed 101 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 101 rows containing missing values (geom_point).
```

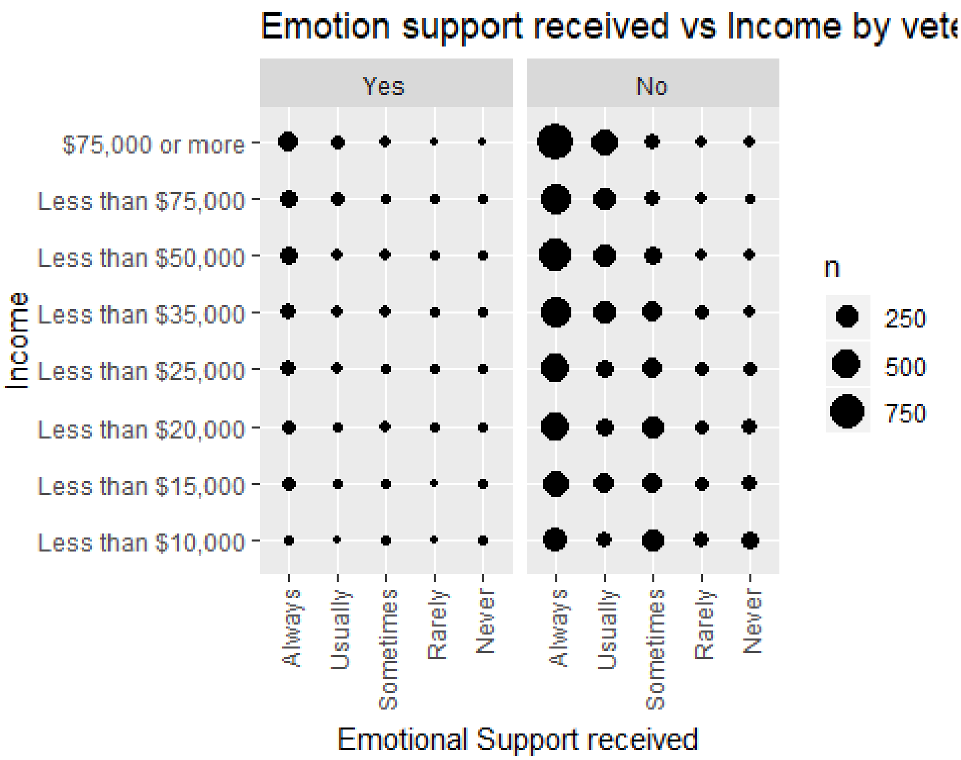


Based on the plots generated it seems to have positive linear relationship in case of men i.e. the average consumption of per glass of alcohol increases with the increase in number of days in anxiety while its almost constant in case of women and dont show any significant increase. Also the average number of days women are feeling anxious is less than those felt by men.

Research quesion 2:

```
 data2 <- select(brfss2013, emtsuprt , veteran3, income2) %>%
  filter(!is.na(emtsuprt), !is.na(veteran3), !is.na(income2))


ggplot(data = data2, aes(x = emtsuprt , y = income2 ))+
  geom_count () +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
  facet_grid(. ~  veteran3) +
  ggtitle("Emotion support received vs Income by veteran status") +
  xlab("Emotional Support received") +
  ylab ("Income")
```

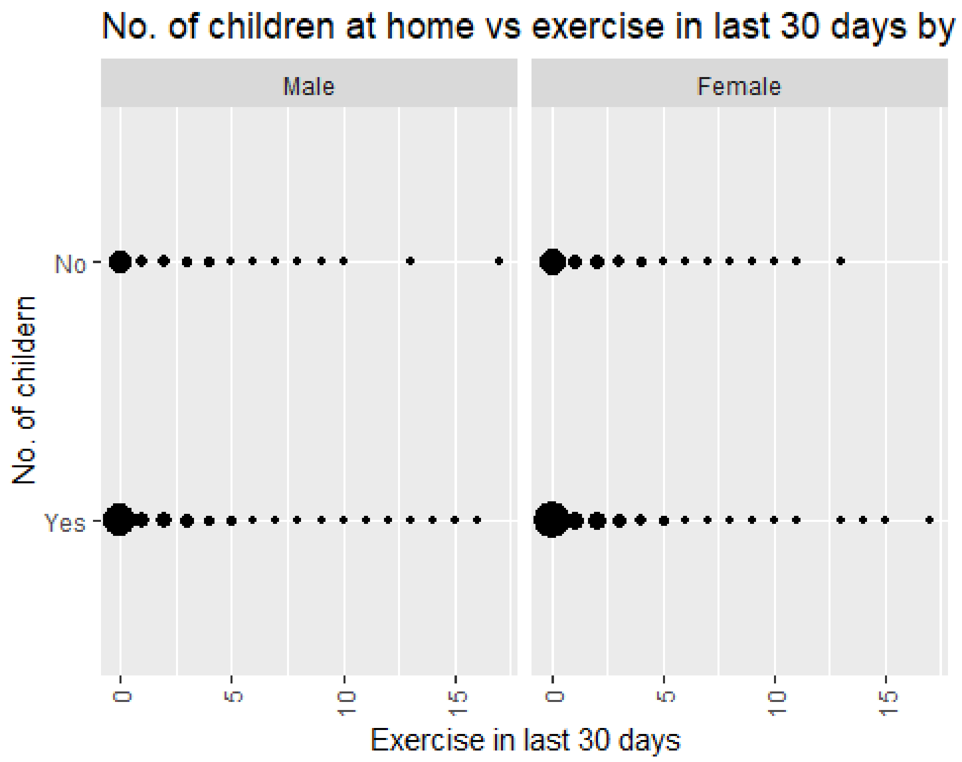


Looks like majority of population is having some or the other healthcare coverage. Out of all the people having coverage retired and employed for wages are the top ones while students, persons who are out of work are the least. So type of employment also plays an important role.

Research quesion 3:

```
data3 <- select(brfss2013, children , exerany2, sex) %>%
filter(!is.na(children), !is.na(exerany2), !is.na(sex))

ggplot(data = data3, aes(x = children , y = exerany2 ))+
geom_count () +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+
theme(legend.position = "none")+
facet_grid(. ~  sex) +
ggtitle("No. of children at home vs exercise in last 30 days by gender") +
xlab("Exercise in last 30 days") +
ylab ("No. of childern")
```



# It seems that the adults in the househould having no children do exercise but having 1 or more child does impact. Although based on the plot it seems that female with no kids at home do exercise slightly more frequently but the difference is not that much and doing exercise is more of less gender neutral