The Behavioral Risk Factor Surveillance System (BRFSS) is a project designed to measure behavioral risk factors for non-institutionalized adult population (18 years of age and older) residing in the US. Through, monthly Telephone interviews BRFSS collects data regarding chronic diseases, injuries, and preventable infectious diseases. In addition to landline telephone survey, BRFSS is also conducting cellular telephone based surveys. Few Factors assessed in these surveys include tobacco use, HIV/AIDS, hypertension awareness etc.

BRFSS is an observational study as it draws inferences from a sample to a population where the independent variable (health practices and risk behaviors) are not under the control of the researcher. Since 2011, BRFSS is conducting surveys in 2 samples landline telephone and cellular telephone based. In landline Telephone based survey only one adult is randomly selected from the household among stratas(i.e. states- according to survey) .Whereas, in cellular based survey, a person is randomly selected and he/she should reside in a private residence or college housing among stratas(i.e. states according to survey). The survey is a Stratified Sampling and the data also has a N/A row to classify the landline surveyed adults to cellular surveyed adults.

## GENERALIZABILTY

Generalizability by definition involves the extent to which the results of a study can be applied beyond the sample to the larger population.BRFSS used random Sampling and It can be generalized to all 50 states in USA and US virgin islands with non-institutionalized adult population of 18 years of age and older having a Telephone Number.

## CAUSALITY

Association does not mean causation. As BRFSS is an observational study to measure the health risk factors among individuals. Establishing causation by ruling out the possibility of any lurking variable, or by assuming that individuals differ only with respect to the values of the explanatory variable is difficult. Due to the lurking variables, it is not preferable to establish causation in an observational study.

**Is there an association between Used Prescribed Preventative Asthma Med(asthmed3) and Depression (qlmentl2) in Past 30 Days ?**

brfss2013\_Question1<-brfss2013%>%filter(!is.na(qlmentl2),qlmentl2!="NA",!is.na(asthmed3))%>%mutate(depMore=ifelse(qlmentl2=="0","Notdepressed","Depressed"))

ggplot(brfss2013\_Question1, aes(x = asthmed3, fill = depMore)) + geom\_bar(position = "fill")+ scale\_fill\_discrete(name = "Depression") + xlab("Number of Days Asthma Medication Taken(In Days)") + ylab("Proportion")

brfss2013\_Question1 %>% group\_by(asthmed3) %>% summarise(prop\_depre = sum( depMore== "Depressed") / n())

------------------xxxxxxxxxxxxxxxxxx------------------xxxxxxxxxxxxxxxxxx------------

 Is there any correlation between Drink Sugar-Sweetened Drinks(**ssbfrut2)** and Asthma Symptoms(asymptom) for males and females (sex variable)?

x

brfss2013%>%filter(!is.na(ssbfrut2),!is.na(asymptom))%>%group\_by(ssbfrut2)%>%summarise(count=n())

y

brfss2013%>%filter(!is.na(asymptom))%>%group\_by(asymptom)%>%summarise(count=n())

brfss2013\_Question3<-brfss2013%>%filter(!is.na(ssbfrut2),!is.na(asymptom))%>%mutate(sweetDrink = ifelse(ssbfrut2<200,ssbfrut2\*30,ifelse(ssbfrut2<300,ssbfrut2\*4,ssbfrut2)))%>%group\_by(asymptom,sweetDrink)%>%summarise(count=n())

ggplot(data = brfss2013\_Question3, aes(x = sweetDrink, y = count, group = asymptom, colour = asymptom))+geom\_line()+geom\_point()

ggplot(data = brfss2013\_Question3, aes(x = sweetDrink, y = count, group = asymptom, colour = asymptom))+ geom\_point() + geom\_smooth(method = lm, se = FALSE)

General Health([genhlth](file:///C:\\Users\\jaideep\\Documents\\Coursera-Project\\brfss_codebook.html" \l "#genhlth)) and How Much Time Do You Sleep([sleptim1](file:///C:\Users\jaideep\Documents\Coursera-Project\brfss_codebook.html##sleptim1)☺) for males and females

brfss2013\_Question2<-brfss2013%>%filter(!is.na(sleptim1),!is.na(genhlth),genhlth=="Excellent"|genhlth=="Poor"|genhlth=="Fair",!is.na(sex))%>%group\_by(genhlth,sleptim1,sex)%>%summarise(sum=n())

ggplot(data = brfss2013\_Question2, aes(x = sleptim1, y = sum, group = genhlth, fill = genhlth))+ geom\_bar(stat = "identity", width = 0.5, position = "dodge")+ facet\_grid(.~sex)+ theme\_bw()+ theme(axis.text.x = element\_text(angle = 90))

Irregular daily routines

How Many Hours Per Week Did You Work and How Often Feel Restless Past 30 Days for males and females

brfss2013\_Question4 <-brfss2013 %>% filter(X\_drnkdy4 != "NA",!is.na(X\_drnkdy4),!is.na(sex)) %>% filter(scntlwk1 != "NA")%>%mutate(drinkWeek=X\_drnkdy4\*7)%>%group\_by(scntlwk1,drinkWeek,sex)%>%summarise(count=n())

ggplot(brfss2013\_Question4, aes(x =scntlwk1 , y = log(drinkWeek+1), colour = sex)) + geom\_point(shape = 19, alpha = 1/4) + geom\_smooth(method = lm, se = FALSE) + scale\_colour\_discrete(name = "Sex")

---

title: "Exploring the BRFSS data"

output:

html\_document:

fig\_height: 4

highlight: pygments

theme: spacelab

---

## Setup

### Load packages

```{r load-packages, message = FALSE}

library(ggplot2)

library(dplyr)

```

### 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.

```{r load-data}

load("brfss2013.RData")

```

\* \* \*

## Part 1: Data

The Behavioral Risk Factor Surveillance System (BRFSS) is a project designed to measure behavioral risk factors for non-institutionalized adult population (18 years of age and older) residing in the US. Through, monthly Telephone interviews BRFSS collects data regarding chronic diseases, injuries, and preventable infectious diseases. In addition to landline telephone survey, BRFSS is also conducting cellular telephone based surveys. Few Factors assessed in these surveys include tobacco use, HIV/AIDS, hypertension awareness etc.

BRFSS is an observational study as it draws inferences from a sample to a population where the independent variable (health practices and risk behaviors) are not under the control of the researcher. Since 2011, BRFSS is conducting surveys in 2 samples landline telephone and cellular telephone based. In landline Telephone based survey only one adult is randomly selected from the household among stratas(i.e. states- according to survey) .Whereas, in cellular based survey, a person is randomly selected and he/she should reside in a private residence or college housing among stratas(i.e. states according to survey). The survey is a Stratified Sampling and the data also has a N/A row to classify the landline surveyed adults to cellular surveyed adults.

GENERALIZABILTY

Generalizability by definition involves the extent to which the results of a study can be applied beyond the sample to the larger population.BRFSS used random Sampling and It can be generalized to all 50 states in USA and US virgin islands with non-institutionalized adult population of 18 years of age and older having a Telephone Number.

CAUSALITY

Association does not mean causation. As BRFSS is an observational study to measure the health risk factors among individuals. Establishing causation by ruling out the possibility of any lurking variable, or by assuming that individuals differ only with respect to the values of the explanatory variable is difficult. Due to the lurking variables, it is not preferable to establish causation in an observational study.

\* \* \*

## Part 2: Research questions

\*\*Research quesion 1:\*\*

\*\*Research quesion 2:\*\*

\*\*Research quesion 3:\*\*

\* \* \*

## 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:\*\*

```{r}

brfss2013\_Question1<-brfss2013%>%filter(!is.na(qlmentl2),qlmentl2!="NA",!is.na(asthmed3))%>%mutate(depMore=ifelse(qlmentl2=="0","Notdepressed","Depressed"))%>%group\_by(asthmed3,depMore)%>%summarise(count=n())

ggplot(data = brfss2013\_Question1, aes(x = asthmed3, y = count, group = depMore, fill = depMore))+ geom\_bar(stat = "identity", width = 0.5, position = "dodge")+ theme\_bw()+ theme(axis.text.x = element\_text(angle = 90))+ scale\_fill\_discrete(name = "Depression") + xlab("Number of Days Asthma Medication Taken(In Days)") + ylab("Count")

```

\*\*Research quesion 2:\*\*

```{r}

brfss2013\_Question4 <-brfss2013 %>% filter(X\_drnkdy4 != "NA",!is.na(X\_drnkdy4),!is.na(sex)) %>% filter(scntlwk1 != "NA")%>%mutate(drinkWeek=X\_drnkdy4\*7)%>%group\_by(scntlwk1,drinkWeek,sex)%>%summarise(count=n())

ggplot(brfss2013\_Question4, aes(x =scntlwk1 , y = log(drinkWeek+1), colour = sex)) + geom\_point(shape = 19, alpha = 1/4) + geom\_smooth(method = lm, se = FALSE) + scale\_colour\_discrete(name = "Sex")+ xlab("Number of Hours Worked in per Week") + ylab("Alcohol Beverages Consumption per Week(log)")

```

\*\*Research quesion 3:\*\*

```{r}

brfss2013\_Question2<-brfss2013%>%filter(!is.na(sleptim1),!is.na(genhlth),genhlth=="Excellent"|genhlth=="Poor"|genhlth=="Fair",!is.na(sex))%>%group\_by(genhlth,sleptim1,sex)%>%summarise(sum=n())

ggplot(data = brfss2013\_Question2, aes(x = sleptim1, y = sum, group = genhlth, fill = genhlth))+ geom\_bar(stat = "identity", width = 0.5, position = "dodge")+ facet\_grid(.~sex)+ theme\_bw()+ theme(axis.text.x = element\_text(angle = 90))+ scale\_fill\_discrete(name = "GeneralHealth") + xlab("Sleep Time(in hrs)") + ylab("Count")

```