Exploring the BRFSS data

Setup

Load packages

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.1
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.1
```

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 BRFSS-2013 was a periodic survey, which selected people from the United States population, including all states and other regions belonging to the territory. The sample selection was given in two ways:

Stratified sample in most states and regions. Simple-sized sample in some states and regions.

The interviews all took place via telephone or mobile through a structured form, conducted by trained researchers with the aid of specific research methodology (CATI), **by randomly selecting phone / cell numbers from each region / province** from participants who met the eligibility criteria.

Given the sample size, the applied randomness and the selection in several strata, it is possible to state that with such data one can infer **generalization** in the results. In other words, the results of such a study have an **approximation** with the United States population.

On the other hand, it **is not possible** to establish a **causal relationship**, since this research is a survey and no methodologies were applied in different subgroups (**experimental vs. control**) in order to investigate a causal relationship. Nevertheless, it is possible to make

associations (correlations) between the study variables.

Part 2: Research questions

Research quesion 1:

> *Are there differences or similarities in access to health coverage between ethnicities/races in the state of Alabamas and the District of Columbia?*

The purpose of this research question is to verify the impact of race / ethnicity on access to health insurance between two states. The state of Alabamas was randomly selected for comparison with the District of Columbia, one of the most African American states in the United States in the year this survey was released (2013).

Variables involved: Alabama State `alabama` , District of Columbia `columbia` , Race/ethnicity `X_race` , Healthcare coverage `hlthpln1` .

Research quesion 2:

> *Do male veterans report high rates of depression?*

The purpose of this research question is to examine whether male war veterans have a high frequency of self-reported depression, as this is a common sense disorder associated with war veterans.

Variables involved: Veterans `veterans` , Male `male` , Depression disorder `addepev2`

Research quesion 3:

> *There are the differences or similarities in marital status in the United States male and female population (are there very different or similar prevalence)?*

This research question aims to investigate through data, the frequency and proportion of the most prevalent marital status in the United States, reported between the sexes and to look for similarities and differences.

Variables involved: Male `male` , Female `female` , Marital status `marital`

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:

To investigate this problem, it was filtered the variables from each of the selected states `columbia` and `alabama`, assigning them to a specific new dataframe.

```
alabama<-(filter(brfss2013,X_state == "Alabama"))
columbia<-(filter(brfss2013,X_state == "District of Columbia"))
```

Then, a frequency table was generated, containing the race vectors and the vector asking if the interviewee has any access to any health coverage.

```
alabamahealth<-table(alabama$X_race, alabama$hlthpln1)
columbiahealth<-table(columbia$X_race, columbia$hlthpln1)
```

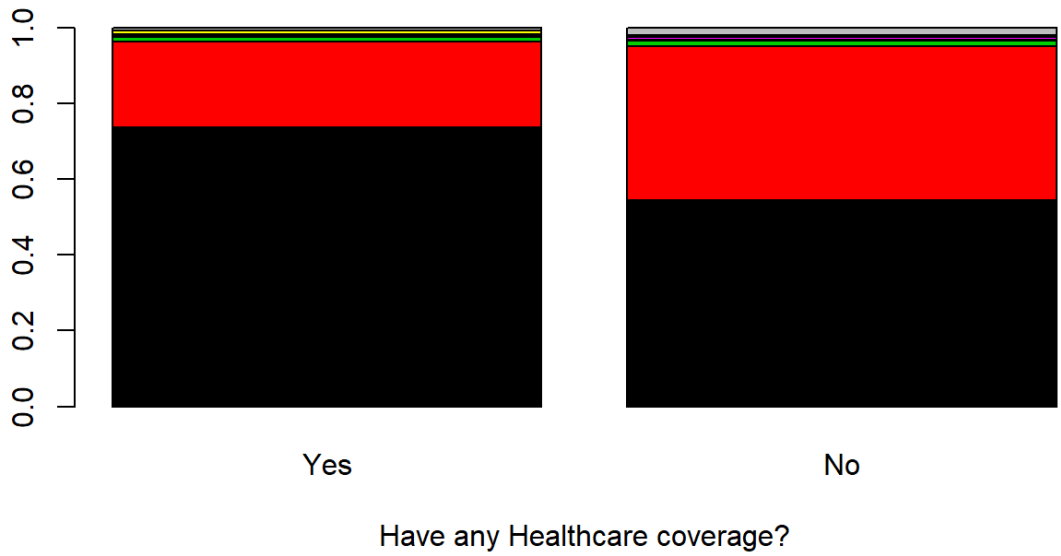
Finally, the frequency values are transformed in proportion, so that they are prepared for the creation of a barplot.

```
alabamahealth2<-prop.table(alabamahealth, 2)
columbiaahealth2<-prop.table(columbiahealth, 2)
```

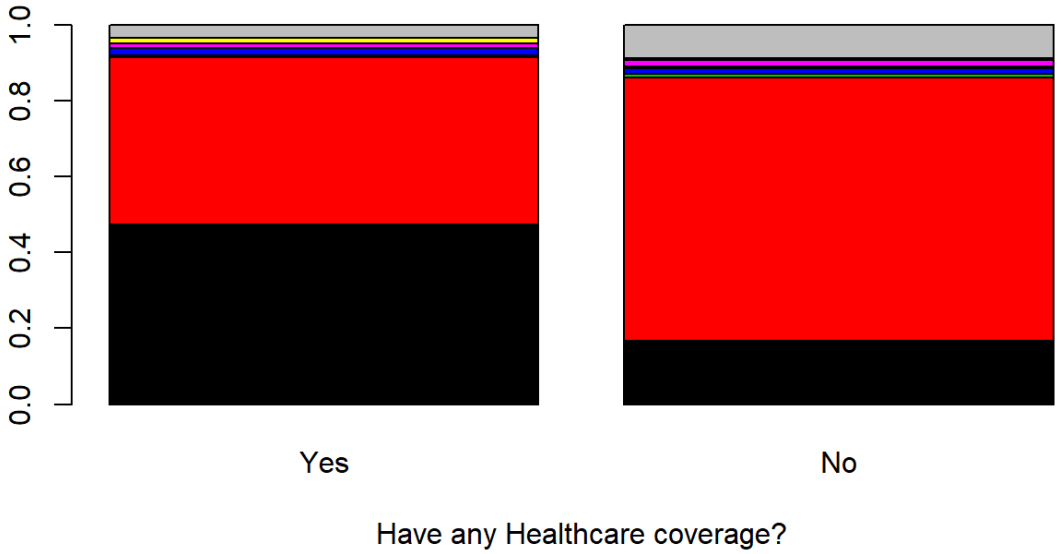
The barplot is created by assigning subtitles and attributed race specific colors in the subtitles of the barplot.

Color Insertion

```
barplot(alabamahealth2, xlab="Have any Healthcare coverage?",col=1:8)
```

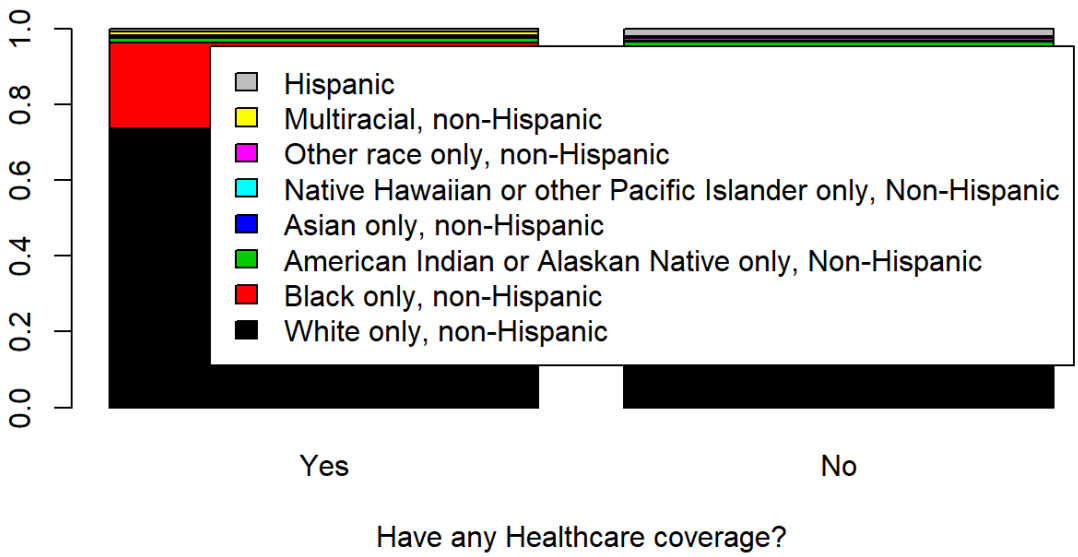


```
barplot(columbiaahealth2, xlab="Have any Healthcare coverage?",col=1:8)
```

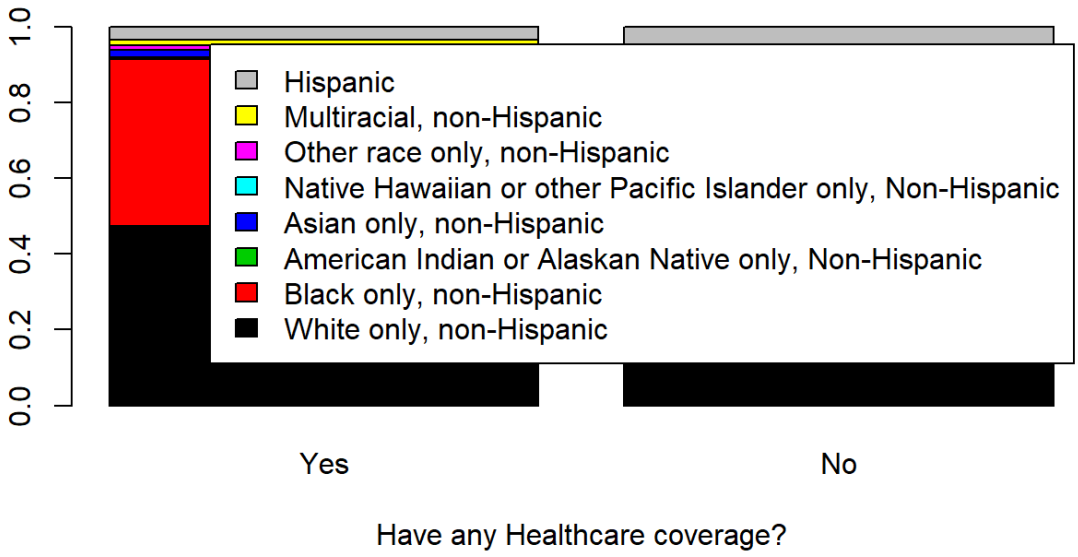


Subtitles Insertion

```
barplot(alabamahealth2,legend = rownames(alabamahealth2), xlab="Have any Hea
        lthcare coverage?",col=1:8)
```



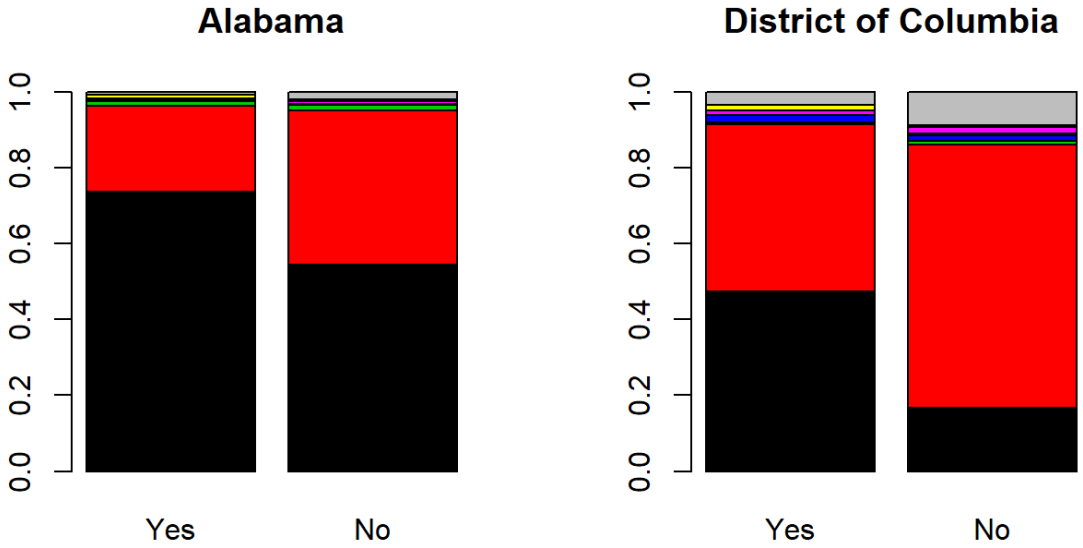

```
barplot(columbiaahealth2,legend = rownames(columbiaahealth2), xlab="Have any
        Healthcare coverage?",col=1:8)
```



After performing the process for each state (`alabama` and `columbia`), the barplots of each state are inserted in a single view.

```
par(mfrow = c(1:2))

barplot(alabamahealth2, main = "Alabama", col=1:8)
barplot(columbiaahealth2, main = "District of Columbia", col=1:8)
```



Its possible to see some interesting results. White people in both states have a higher rate in the variable access to health coverage (`hlthpln1`) and in the situation that they do not, it is

much less than black people. Thus showing strong evidence for the greater **unfeasibility** of health coverage for the public of non-white ethnic groups, and especially for the black population. A `summary` with sample-specific details is displayed in the summary below.

```
columbiahealth
```

```
##
##                                                                 Yes   No
##   White only, non-Hispanic                                     2172   36
##   Black only, non-Hispanic                                     2024  151
##   American Indian or Alaskan Native only, Non-Hispanic           18    2
##   Asian only, non-Hispanic                                       86    3
##   Native Hawaiian or other Pacific Islander only, Non-Hispanic    6    1
##   Other race only, non-Hispanic                                  47    4
##   Multiracial, non-Hispanic                                      75    1
##   Hispanic                                                      154   19
```

```
alabamahealth
```

```
##
##                                                                 Yes   No
##   White only, non-Hispanic                                     4224  377
##   Black only, non-Hispanic                                     1298  281
##   American Indian or Alaskan Native only, Non-Hispanic           67   11
##   Asian only, non-Hispanic                                       15    1
##   Native Hawaiian or other Pacific Islander only, Non-Hispanic    2    0
##   Other race only, non-Hispanic                                  28    5
##   Multiracial, non-Hispanic                                      47    3
##   Hispanic                                                       46   14
```

Research quesion 2:

To investigate this problem, it was first filtered out the variables of interest (`veterans` and `male`) by assigning them to a new dataframe (`veteransmale`).

```
veterans<-(filter(brfss2013,veteran3 == "Yes"))
veteransmale<-(filter(veterans,sex == "Male"))
```

Then, a frequency table containing the vectors of interest was generated. That was also transformed in proportion to facilitate interpretation of the analysis.

```
table(veteransmale$addepev2)
```

```
##
##   Yes    No
##  8391 47254
```

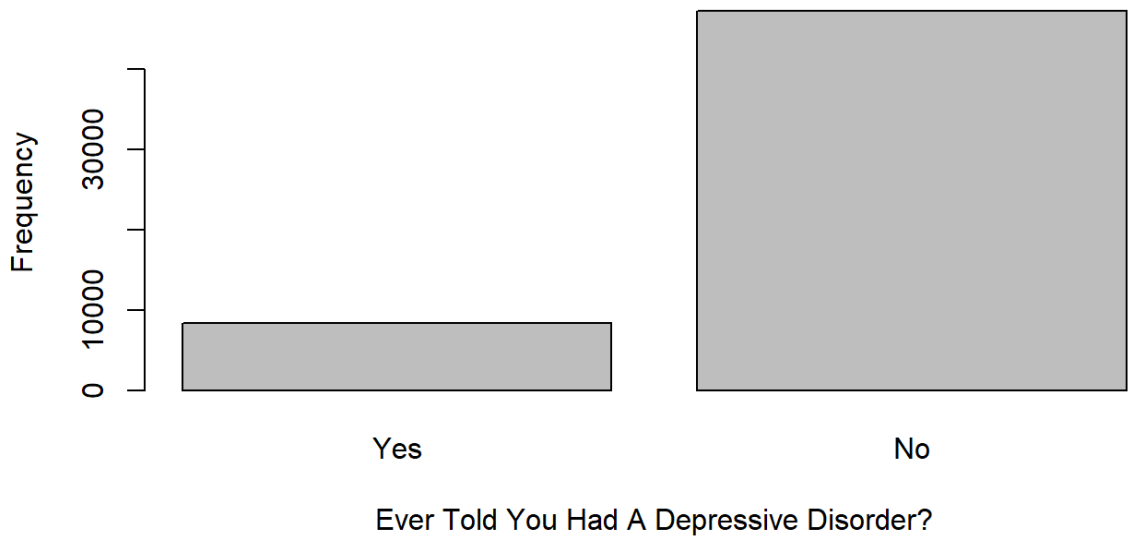
```
prop.table(table(veteransmale$addepev2))
```

```
##
##       Yes        No
## 0.1507952 0.8492048
```

The results indicate that 15% of respondents (`prop.table`) in this category indicate self-reported depressive disorder, in contrast approximately 85% claim not to have.

To facilitate the reader's visualization, it was decided to turn these results into a bar chart.

```
barplot(xtabs(~veteransmale$addepev2),xlab = "Ever Told You Had A Depressive
        Disorder?", ylab = "Frequency")
```



Compared to the overall United States average (showed below), the frequency of self-reported depression of male veterans is lower (overall US average: 19%) but slightly higher compared to the overall male population (average: 14%).

```
prop.table(table(brfss2013$addepev2))
```

```
##
##       Yes        No
## 0.1956726 0.8043274
```

```
male<-(filter(brfss2013,sex == "Male"))
prop.table(table(male$addepev2))
```

```
##
##       Yes        No
## 0.1421308 0.8578692
```

Research question 3:

To investigate this problem, it was first filtered out the variables of interest (male and female separately) by attributing to distinct dataframes (`male` - already did in the Question 2- and `female`).

```
male<-(filter(brfss2013,sex == "Male"))
female<-(filter(brfss2013,sex == "Female"))
```

To summarise the results, a frequency table was generated, containing the vectors of interest with the main variable separately (in this case, `marital`). Finally, the frequency values were also transformed in proportion to better understand the analysis performed.

The Female Marital status results:

```
table(female$marital)
```

```
##
##                       Married                      Divorced
##                        139268                         44140
##                       Widowed                     Separated
##                         52998                          6810
##                 Never married A member of an unmarried couple
##                         37845                          7302
```

```
prop.table(table(female$marital))
```

```
##
##                       Married                      Divorced
##                    0.48296071                    0.15307096
##                       Widowed                     Separated
##                    0.18378918                    0.02361607
##                 Never married A member of an unmarried couple
##                    0.13124083                    0.02532225
```

The Male Marital status results:

```
table(male$marital)
```

```
##
##                       Married                      Divorced
##                        114060                         26236
##                       Widowed                     Separated
##                         12747                          3852
##                 Never married A member of an unmarried couple
##                         37225                          5871
```

```
prop.table(table(male$marital))
```

```
##
##                       Married                      Divorced
##                    0.57032566                    0.13118590
##                       Widowed                     Separated
##                    0.06373787                    0.01926087
##                 Never married A member of an unmarried couple
##                    0.18613338                    0.02935632
```

The overall results indicate similarity responses between the sexes in general but one exceptional variable draws particular attention: Widowed people. Widowed women (18%) are 3 times larger in proportion compared to widowed men (6%) pointing out this major difference.