

# Exploring the BRFSS data

## Setup

### Load packages

```
library("gridExtra")
library("ggplot2")
library("dplyr")
library("maps")
library("evaluate")
library("vcd")
library("alluvial")
```

### Load data

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

## Part 1: Data

### General information about data

The Behavioral Risk Factor Surveillance System (BRFSS) is a project between all US states (US) and participating US territories and the Centers for Disease Control and Prevention (CDC).

This telephone surveys system collects health information from the public. Polls are observational. Surveys are conducted in 50 states, as well as in several US territories, including the District of Columbia, Puerto Rico and Guam. The health characteristics assessed on the basis of the BRFSS refer to the adult population aged 18 and over, non-institutionalized adults living in the United States. They include questions about the level of the physical and psychological state of the population, about the presence of diseases such as heart disease, cancer, asthma, HIV/AIDS, depression and others, as well as about various types of activity, diet, sleep, bad habits, and others things that characterize a person's lifestyle and health.

### Causality and generability

These studies were carried out on the basis of a large set of data that were obtained through stratified random sampling. Based on this, the results can be generalized to the entire population of the United States of America. However, the data were not obtained experimentally, because this is an observational study, and therefore we cannot establish causal relationships, but can only investigate the variables for the presence of correlation.

Collecting data in the way presented in the study could lead to a number of errors. They could be related to the fact that it was not taken into account that some people could not or did not want to answer a phone call - in this regard, certain biases could appear. Also, some people could give incorrect information or not give information at all due to ignorance, for example, the fact of the presence of an illness or the fact that they did not want to mention their illnesses, psychological problems, sex life, financial situation. The information may have been incorrectly recorded by the respondent who conducted the survey. In addition, some of the records in the survey were incomplete, which could also affect the research results.

## Part 2: Research questions

A large number of mental disorders are known around the world, and there is a popular professional opinion that medications are not always the best way to deal with a disorder. Doctors say that drugs, for example, in the form of antidepressants and antipsychotics, in particular, lead to the fact that the human body loses its ability to fight the disease on its own and may not recover it. Therefore, it is recommended to start treatment with different sessions with a psychologist, and then, if necessary, introduce medications into the treatment.

Depression is a common mental disorder and is one of the leading disabilities worldwide. An estimated 264 million people worldwide suffer from depression. There are statistics that women are more susceptible to this disorder than men.

This study aims to obtain statistics on the segment of the population that has experienced depression or mental ill health in the last 30 days according to the 2013 BRFSS. These findings, coupled with those of other years of research, could help guide the decision to build additional mental health centers specializing in depressive disorders in the most sought after regions.

To do this, we need to answer three questions.

### Research question 1:

What part of the population of those who were depressed in the last 30 days needs psychological help and how often? Since a person may not be aware of the presence of depression, it is additionally worth finding out which part of the population as a whole needs psychological help and how often.

### Research question 2:

Deepening the question, let us study what part of the population among women and men experienced depression in the last 30 days, while having a corresponding diagnosis. At the same time, we are more interested in what part of people have never officially confirmed the diagnosis, which means that they most likely have no idea how to identify this disorder and where to go for help.

### Research question 3:

The third question is to find regions with the greatest concentration of people who feel mentally ill. This information is necessary to decide on the construction of specialized medical and psychological care institutions.

---

## Part 3: Exploratory data analysis

### Research question 1:

Let's designate the variables that will be used in this question:

1. misdeprd: How Often Feel Depressed Past 30 Days

What does it mean?

"During the past 30 days, about how often did you feel so depressed that nothing could cheer you up? (If necessary: all, most, some, a little, or none of the time.)"

2. emtsuprt: How Often Get Emotional Support Needed

What does it mean?

"How often do you get the social and emotional support you need? (If necessary: always, usually, sometimes, rarely, or never.)"

To avoid missing values in our data, let's put the data filtered from missing records into a new dataset called brfss2013\_quest1.

```
brfss2013_quest1 <- brfss2013 %>%  
  filter(!is.na(brfss2013$misdeprd), !is.na(brfss2013$emtsuprt))
```

Let's count the number of responses about the need for emotional assistance.

```
brfss2013_quest1 %>%  
  group_by(emtsuprt) %>%  
  summarise(count = n())
```

```
## # A tibble: 5 x 2  
##   emtsuprt   count  
## * <fct>     <int>  
## 1 Always      2770  
## 2 Usually    1029  
## 3 Sometimes    677  
## 4 Rarely      209  
## 5 Never       249
```

And let's count the number of these responses for the category of people who felt depressed over the past 30 days.

```
brfss2013_quest1 %>%  
  filter(misdeprd != 'None') %>%  
  group_by(emtsuprt) %>%  
  summarise(count = n())
```

```
## # A tibble: 5 x 2  
##   emtsuprt   count  
## * <fct>     <int>  
## 1 Always      281  
## 2 Usually    207  
## 3 Sometimes    270  
## 4 Rarely     112  
## 5 Never       92
```

We see that the total number of people in need of emotional help significantly exceeds the number of those who do not need it at all, by more than 18 times.

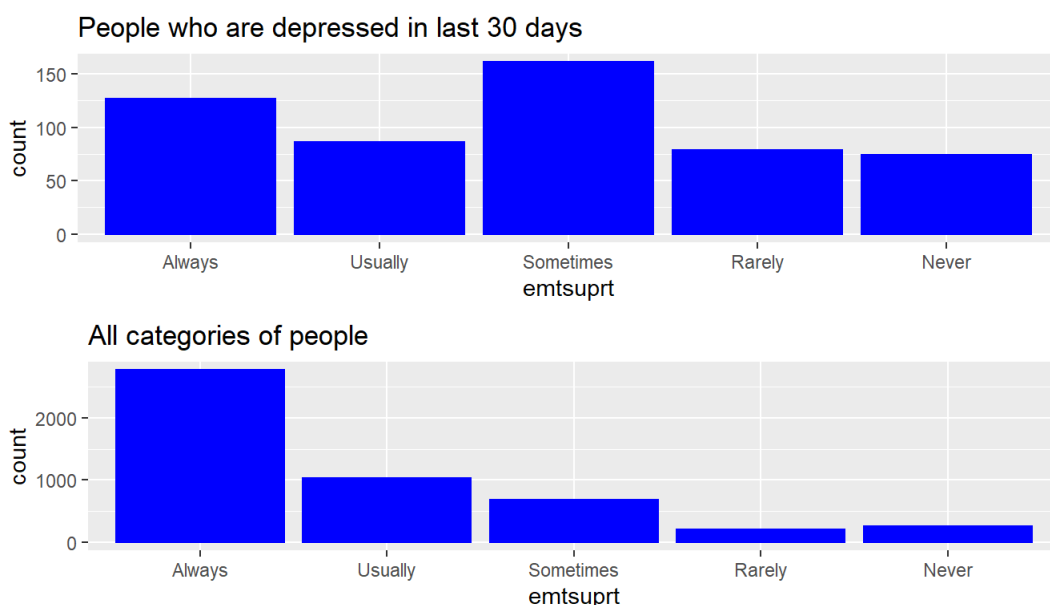
You can also notice that among people who have suffered from depression in the last 30 days, there are relatively fewer people who need help. From this we can conclude that people who have not been depressed in the last 30 days but need help most likely either experienced temporary difficulties at the time of the survey and their condition was momentary, or did not suspect that they had a disorder, but felt that they needed help. About the category of people who were depressed but did not need help, it can be concluded that most likely they already received the necessary help, that is, they knew where to get it. For the category of people who are depressed and in need of help, obviously, it can be concluded that they may receive less help due to their financial situation, lack of specialists, lack of hospital beds in public institutions and drugs and other circumstances that can be solved by construction of accessible centers for psychological and medical assistance.

For clarity, let's build graphs.

```
plot1 <- ggplot(brfss2013_quest1 %>%
  group_by(emtsuprt) %>%
  filter(misdeprd %in% c('All', 'Most', 'Some')) %>%
  summarise(count = n()), aes(emtsuprt, count)) +
  geom_col(colour = 'blue', fill = 'blue') +
  ggtitle('People who are depressed in last 30 days')

plot2 <- ggplot(brfss2013_quest1 %>%
  group_by(emtsuprt) %>%
  summarise(count = n()), aes(emtsuprt, count)) +
  geom_col(colour = 'blue', fill = 'blue') +
  ggtitle('All categories of people')

grid.arrange(plot1, plot2)
```



According to statistics, many people, regardless of the observation of depression in the last 30 days, need psychological help. Consequently, the issue of its provision is relevant.

### Research question 2:

Let's designate the variables that will be used in this question:

1. sex: Respondents Sex

What does it mean?

"Indicate sex of respondent"

2. misdeprd: How Often Feel Depressed Past 30 Days

What does it mean?

"During the past 30 days, about how often did you feel so depressed that nothing could cheer you up? (If necessary: all, most, some, a little, or none of the time.)"

3. addepev2: Ever Told You Had A Depressive Disorder

What does it mean?

"Has a doctor, nurse, or other health professional ever told you that you had any of the following? For each, tell me "Yes", "No", or you're "Not sure": (Ever told) you that you have a depressive disorder, including depression, major depression, dysthymia, or minor depression?"

To avoid missing values in our data, let's put the data filtered from missing records into a new dataset called brfss2013\_quest2.

```
brfss2013_quest2 <- brfss2013 %>%
  filter(!is.na(brfss2013$misdeprd), !is.na(brfss2013$sex), !is.na(brfss2013$addepev2))
```

Let's count the number of responses about a diagnosis associated with depressive disorder.

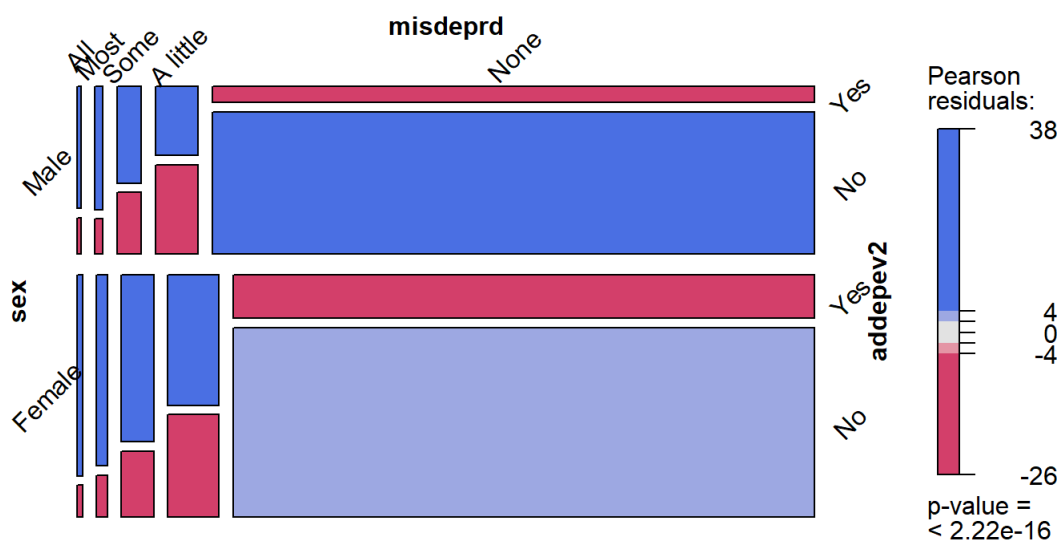
```
brfss2013_quest2 %>%
  group_by(addepev2) %>%
  summarise(count = n())
```

```
## # A tibble: 2 x 2
##   addepev2 count
## * <fct>    <int>
## 1 Yes      7715
## 2 No      28236
```

We see that 7715 people out of 35951 have ever had a depressive disorder and this is an impressive number.

Now let's deepen the research and find out what proportion of people of each gender experienced depression in the last 30 days, while having a corresponding diagnosis.

```
mosaic(~ sex + misdeprd + addepev2,
  data = brfss2013_quest2, shade = TRUE,
  labeling = labeling_border(rot_labels = c(45, 45, 45)))
```



We can see from this graph that the number of women suffering from depressive disorders exceeds the number of men, which is in line with well-known statistics. We also see that there is a category of people who do not have diagnoses of the presence of a depressive disorder, but feel on themselves its symptoms. Therefore, people in this category either have depression but do not know about it, or have another disorder, for example, similar in symptoms, but less common in everyday life, or have temporary difficulties. The main purpose of building centers is to help people with depression. But also those who do not know about their diagnosis or who cope with temporary problems will be able to get help.

### Research question 3:

Let's designate the variables that will be used in this question:

1. \_state: State Fips Code

What does it mean?

"State FIPS Code"

2. sex: Respondents Sex

What does it mean?

“Indicate sex of respondent”

### 3. menthlth: Number Of Days Mental Health Not Good

What does it mean?

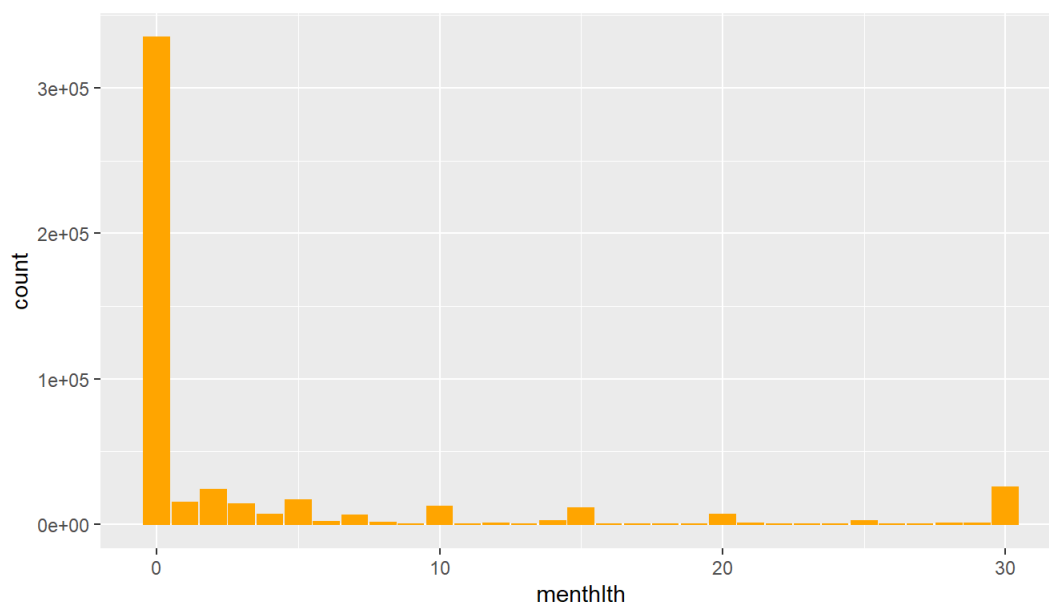
“Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?”

To avoid missing values in our data, let's put the data filtered from missing records into a new dataset called brfss2013\_quest3.

```
brfss2013_quest3 <- brfss2013 %>%  
  filter(!is.na(brfss2013$menthlth), !is.na(brfss2013$sex))
```

Let's count the number of days during which a person felt bad mentally (this includes stress, depressive states, unstable emotional states, and so on).

```
ggplot(brfss2013_quest3 %>%  
  group_by(menthlth) %>%  
  summarise(count = n()), aes(menthlth, count)) +  
  geom_col(colour = 'orange', fill = 'orange')
```



The graph shows that mostly people have not felt mentally bad over the past 30 days (this is evidenced by the clear peak of the mark at 0 days). But at the same time, we see small, but characteristic peaks at levels 10, 15, 20, 30 days, and this fact indicates the presence of people who have problems with the psychological state for a long time.

Let's divide the datasets by gender.

```
brfss2013_men <- brfss2013_quest3 %>%  
  filter(brfss2013_quest3$sex == 'Male')  
  
brfss2013_women <- brfss2013_quest3 %>%  
  filter(brfss2013_quest3$sex == 'Female')
```

And let's create the appropriate dataframes.

```
m_men <- data.frame(days_menthlth = brfss2013_men$menthlth,  
  state = tolower(brfss2013_men$X_state),  
  gender = brfss2013_men$sex)  
  
m_women <- data.frame(days_menthlth = brfss2013_women$menthlth,  
  state = tolower(brfss2013_women$X_state),  
  gender = brfss2013_women$sex)
```

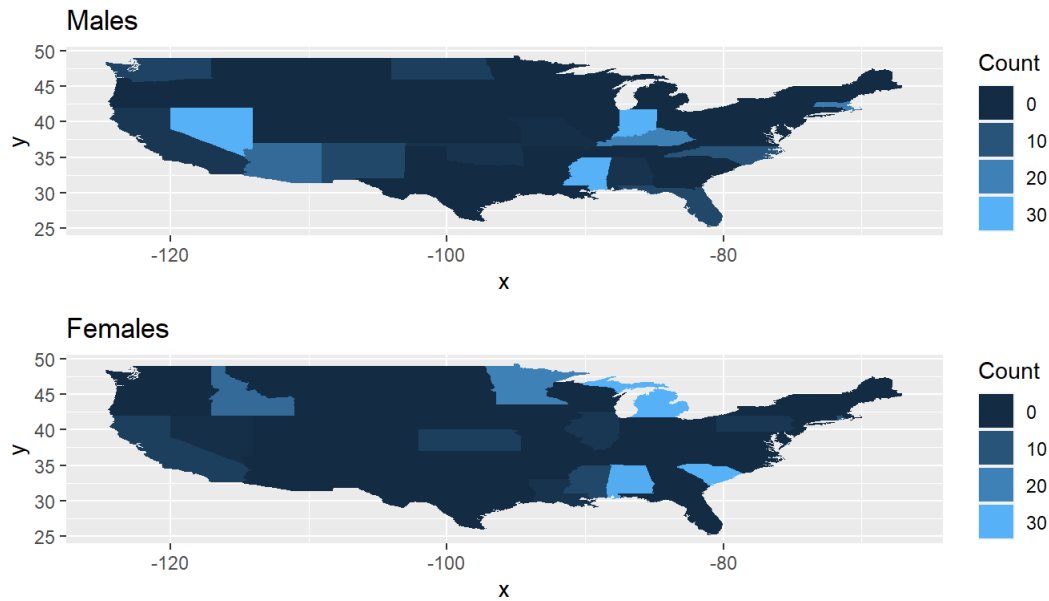
Now let's build the graphs of the distribution by regions of the number of days in which representatives of different genders felt mentally bad. We need this to select the location for the construction of psychological centers.

```
map <- map_data("state")

map_m <- ggplot(m_men, aes(fill = days_menthlth)) +
  geom_map(aes(map_id = state), map = map) +
  expand_limits(x = map$long, y = map$lat) +
  ggtitle('Males') +
  guides(fill = guide_legend(title = "Count"))

map_w <- ggplot(m_women, aes(fill = days_menthlth)) +
  geom_map(aes(map_id = state), map = map) +
  expand_limits(x = map$long, y = map$lat) +
  ggtitle('Females') +
  guides(fill = guide_legend(title = "Count"))

grid.arrange(map_m, map_w)
```



So, as we can see, women with poor mental health predominantly predominate in the east and northwest of the states, while men are predominantly in the southeast and southwest. Hence, these locations that should be taken into account when implementing a project.