

# BRFSS2023\_ANALYSIS.R

abdullahsiddiqui

2025-09-18

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(readr)
library(ggplot2)
```

*# 1. Load dataset*

```
brfss <- read_xpt("~/Downloads/LLCP2023.XPT ")
```

*# 2. Select only needed variables*

```
brfss_small <- brfss %>%
  select(MENTHLTH,      # Poor mental health days
         EXERANY2,      # Exercise in past 30 days
         SMOKDAY2,      # Smoking status
         ALCDAY4,       # Alcohol use frequency
         SEXVAR,        # Sex of respondent
         EDUCA,         # Education level
         INCOME3,       # Income category
         `_AGEG5YR`)    # Age group (5-year intervals)
```

*# 3. Clean & recode variables*

```
# Mental health: valid = 0-30; 88 = "None" + 0; 77/99/etc + NA
brfss_small <- brfss_small %>%
  mutate(MENTHLTH = case_when(
    MENTHLTH %in% 0:30 ~ MENTHLTH,
```

```

    MENTHLTH == 88 ~ 0,
    TRUE ~ NA_real_
  ))

# Exercise: 1=Yes, 2=No; 7/9=NA
brfss_small <- brfss_small %>%
  mutate(EXERANY2 = ifelse(EXERANY2 %in% c(7, 9), NA, EXERANY2))

# Smoking: 1=Every day, 2=Some days, 3=Not at all; 7/9=NA
brfss_small <- brfss_small %>%
  mutate(SMOKDAY2 = ifelse(SMOKDAY2 %in% c(7, 9), NA, SMOKDAY2))

# Alcohol: 888=No drinks + 0; 777/999=NA
brfss_small <- brfss_small %>%
  mutate(ALCDAY4 = case_when(
    ALCDAY4 == 888 ~ 0,
    ALCDAY4 %in% c(777, 999) ~ NA_real_,
    TRUE ~ as.numeric(ALCDAY4)
  ))

# Education: 9=NA
brfss_small <- brfss_small %>%
  mutate(EDUCA = ifelse(EDUCA == 9, NA, EDUCA))

# Income: 77/99=NA
brfss_small <- brfss_small %>%
  mutate(INCOME3 = ifelse(INCOME3 %in% c(77, 99), NA, INCOME3))

```

```

# 4. Save cleaned dataset

```

```

write_csv(brfss_small, "BRFSS2023_subset_clean.csv")

```

```

# 5. Descriptive statistics

```

```

cat("\n--- Cleaned Descriptive Statistics ---\n")

```

```

##
## --- Cleaned Descriptive Statistics ---

```

```

# Average poor mental health days
avg_mental_health <- mean(brfss_small$MENTHLTH, na.rm = TRUE)
cat("Average Poor Mental Health Days:", avg_mental_health, "\n")

```

```

## Average Poor Mental Health Days: 4.358383

```

```

# Exercise
cat("\nExercise Frequency (%):\n")

```

```

##
## Exercise Frequency (%):

```

```
print(round(prop.table(table(brfss_small$EXERANY2, useNA = "no"))) * 100, 1))
```

```
##  
##      1      2  
## 75.3 24.7
```

```
# Smoking  
cat("\nSmoking Status (%):\n")
```

```
##  
## Smoking Status (%):
```

```
print(round(prop.table(table(brfss_small$SMOKDAY2, useNA = "no"))) * 100, 1))
```

```
##  
##      1      2      3  
## 20.1  8.5 71.5
```

```
# Alcohol (non-drinkers vs others)  
cat("\nAlcohol Use:\n")
```

```
##  
## Alcohol Use:
```

```
cat("Non-drinkers (%):", round(mean(brfss_small$ALCDAY4 == 0, na.rm = TRUE) * 100, 1), "\n")
```

```
## Non-drinkers (%): 47.3
```

```
# Education  
cat("\nEducation Levels (%):\n")
```

```
##  
## Education Levels (%):
```

```
print(round(prop.table(table(brfss_small$EDUCA, useNA = "no"))) * 100, 1))
```

```
##  
##      1      2      3      4      5      6  
## 0.2  1.9  3.7 24.7 26.5 42.9
```

```
# Income  
cat("\nIncome Categories (%):\n")
```

```
##  
## Income Categories (%):
```

```
print(round(prop.table(table(brfss_small$INCOME3, useNA = "no")) * 100, 1))
```

```
##
##      1      2      3      4      5      6      7      8      9     10     11
##  2.7   2.9   3.7   5.3  11.1  13.7  16.7  14.2  15.1   7.0   7.7
```

## # 6. Plots

```
# Histogram of poor mental health days
p1 <- ggplot(brfss_small, aes(x = MENTHLTH)) +
  geom_histogram(binwidth = 1, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Poor Mental Health Days",
       x = "Days in past 30 where mental health not good", y = "Count")

ggsave("hist_mentalhealth.png", p1, width = 7, height = 5)
```

```
## Warning: Removed 8108 rows containing non-finite outside the scale range
## ('stat_bin()').
```

```
# Bar chart of exercise
p2 <- ggplot(brfss_small, aes(x = factor(EXERANY2))) +
  geom_bar(fill = "lightgreen", color = "black") +
  labs(title = "Exercise in Past 30 Days",
       x = "Exercise (1=Yes, 2=No)", y = "Count")

ggsave("bar_exercise.png", p2, width = 7, height = 5)
```

```
# Boxplot of poor mental health days by sex
p3 <- ggplot(brfss_small, aes(x = factor(SEXVAR), y = MENTHLTH)) +
  geom_boxplot(fill = "lightpink") +
  labs(title = "Poor Mental Health Days by Sex",
       x = "Sex (1=Male, 2=Female)", y = "Days")

ggsave("boxplot_mentalhealth_sex.png", p3, width = 7, height = 5)
```

```
## Warning: Removed 8108 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```