

# BRFSS2023\_correlation.R

abdullahsiddiqui

2025-09-23

```
#####  
# BRFSS2023_correlation.R  
# Week 4: Correlation analysis, scatterplots, cross-tabs  
# Author: Abdullah Siddiqui  
# Date: Oct 2, 2025  
#####
```

```
library(ggplot2)  
library(reshape2)  
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(readr)  
df <- read_csv("~/Downloads/BRFSS2023_subset_clean.csv")
```

```
## Rows: 433323 Columns: 8  
## -- Column specification -----  
## Delimiter: ","  
## db1 (8): MENTHLTH, EXERANY2, SMOKDAY2, ALCDAY4, SEXVAR, EDUCA, INCOME3, _AGE...  
##  
## i Use 'spec()' to retrieve the full column specification for this data.  
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
# Inspect dataset structure  
str(df)
```

```
## spc_tbl_ [433,323 x 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ MENTHLTH: num [1:433323] 88 88 2 88 88 3 NA 88 88 88 ...
## $ EXERANY2: num [1:433323] 2 1 1 1 1 1 2 2 2 1 ...
## $ SMOKDAY2: num [1:433323] NA NA 3 NA NA NA 3 NA NA 3 ...
## $ ALCDAY4 : num [1:433323] 0 0 0 0 202 205 0 0 0 0 ...
## $ SEXVAR : num [1:433323] 2 2 2 2 2 2 1 2 2 1 ...
## $ EDUCA : num [1:433323] 5 5 4 5 5 5 4 5 5 4 ...
## $ INCOME3 : num [1:433323] NA NA 2 NA 7 7 6 NA 6 7 ...
## $ _AGEG5YR: num [1:433323] 13 13 13 12 12 9 13 12 13 12 ...
## - attr(*, "spec")=
## .. cols(
## .. MENTHLTH = col_double(),
## .. EXERANY2 = col_double(),
## .. SMOKDAY2 = col_double(),
## .. ALCDAY4 = col_double(),
## .. SEXVAR = col_double(),
## .. EDUCA = col_double(),
## .. INCOME3 = col_double(),
## .. '_AGEG5YR' = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
if (!dir.exists("plots")) dir.create("plots")
if (!dir.exists("tables")) dir.create("tables")
```

```
# Select only numeric variables
num_vars <- df %>% select_if(is.numeric)

# Compute correlation matrix
cor_matrix <- cor(num_vars, use = "complete.obs")

# Save correlation heatmap
heatmap_data <- melt(cor_matrix)
p1 <- ggplot(heatmap_data, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "blue", high = "red", mid = "white", midpoint = 0) +
  theme_minimal() +
  labs(title = "Correlation Heatmap") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
ggsave("plots/correlation_heatmap.png", plot = p1, width = 7, height = 6)
```

```
# Income vs. Poor Mental Health Days
p2 <- ggplot(df, aes(x = INCOME3, y = MENTHLTH)) +
  geom_point(alpha = 0.2) +
  geom_smooth(method = "lm", se = FALSE, color = "blue") +
  labs(title = "Income vs. Poor Mental Health Days",
       x = "Income Category", y = "Poor Mental Health Days") +
  theme_minimal()
ggsave("plots/scatter_income_mentalhealth.png", plot = p2, width = 7, height = 5)
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 91511 rows containing non-finite outside the scale range
## ('stat_smooth()').
```

```
## Warning: Removed 91511 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
# Exercise vs. Poor Mental Health Days
p3 <- ggplot(df, aes(x = factor(EXERANY2), y = MENTHLTH)) +
  geom_boxplot(fill = "lightblue") +
  labs(title = "Exercise vs. Poor Mental Health Days",
       x = "Exercise (1=Yes, 2=No)", y = "Poor Mental Health Days") +
  theme_minimal()
ggsave("plots/scatter_exercise_mentalhealth.png", plot = p3, width = 7, height = 5)
```

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

```
# Smoking × Exercise
table_smoke_exercise <- table(df$SMOKDAY2, df$EXERANY2)
write.csv(table_smoke_exercise, "tables/crosstab_smoking_exercise.csv")

# Alcohol × Frequent Distress (binary >14 days poor MH)
df$frequent_distress <- ifelse(df$MENTHLTH > 14, 1, 0)
table_alcohol_distress <- table(df$ALCDAY4, df$frequent_distress)
write.csv(table_alcohol_distress, "tables/crosstab_alcohol_distress.csv")

#####
# End of Script
#####

file.show("plots/correlation_heatmap.png")      # opens the image in a viewer
file.show("tables/crosstab_smoking_exercise.csv") # opens CSV as text
```