Alcohol Abuse and Mental/Physical Health in the United States

Jane Condon

2024-03-18

Loading packages

```
library(haven)
library(sf)
library(ggplot2)
library(sjPlot)
library(sjmisc)
library(sjlabelled)
library(plyr)
library(tidyverse)
library(formatR)
```

Importing data from the CDC's BRFSS - 2018

```
zip_tf <- tempfile()

zip_url <- "https://www.cdc.gov/brfss/annual_data/2018/files/LLCP2018XPT.zip"

download.file(zip_url, zip_tf, mode = "wb")

brfss_tbl <- read_xpt(zip_tf)

brfss_df <- data.frame(brfss_tbl)

names(brfss_df) <- tolower(names(brfss_df))

brfss_df[, "one"] <- 1
```

Data Manipulation and Cleaning

Selecting only the variable names that we want to keep to make the data more manageable

```
variables <- c("x_state", "maxdrnks", "alcday5", "avedrnk2",
    "drnk3ge5", "menthlth", "poorhlth", "addepev2", "genhlth",
    "physhlth", "x_ageg5yr", "sex1", "one", "x_llcpwt", "x_rfbing5")
brfss_df <- brfss_df[variables]
```

Adding state abbreviation column

```
fips <- read.csv("fips.csv")

colnames(brfss_df)[colnames(brfss_df) == "x_state"] = "fips"

brfss_df <- merge(fips[, c("fips", "st", "stname")], brfss_df,
    by = "fips", all.x = F, all.y = T)
```

Subsetting the data

Binge Drinking Prevalence

```
binge_prev <- brfss_df[(brfss_df$x_rfbing5 == 1 | brfss_df$x_rfbing5 ==
    2) & !is.na(brfss_df$x_rfbing5), ]
```

Binge Drinking Intensity

```
# Dataframe excluding NA and 'not sure' or 'refused'
# responses from the 'max drinks' column
max_drinks <- brfss_df[(brfss_df$maxdrnks %in% 1:76) & !is.na(brfss_df$maxdrnks),
    ]
```

Binge Drinking Frequency

```
# Dataframe excluding NA and 'not sure' or 'refused'
# responses from the binge drinking column
binge_freq <- brfss_df[(brfss_df$drnk3ge5 %in% 1:88) & !is.na(brfss_df$drnk3ge5),
    ]

# Replacing '88' values with '0', since a response of '88'
# means 'None'
binge_freq$drnk3ge5[binge_freq$drnk3ge5 == "88"] <- 0
```

Average Alcohol Consumption

```
# Dataframe excluding NA and 'not sure' or 'refused'
# responses from the 'average number of drinks' consumed
# column
avg_drinks <- brfss_df[(brfss_df$avedrnk2 %in% 1:76) & !is.na(brfss_df$avedrnk2),
    ]
```

Average Drinking Frequency

```
# Dataframe excluding NA and 'not sure' or 'refused'
# responses from the 'number of days per week/month where
# you consumed an alcoholic beverage' column
alc_days <- brfss_df[(brfss_df$alcday5 %in% 101:107 | brfss_df$alcday5 %in%
    201:230 | brfss_df$alcday5 == 888) & !is.na(brfss_df$alcday5),
    ]

# Replacing '888' values with '0', since a response of
# '888' means 'No drinks in the past 30 days'
alc_days$alcday5[alc_days$alcday5 == "888"] <- 0

# Remove the '2' from 201-230 values, since 2 _ _ = days
# per month and dividing it by 7 to get 'days per week'

alc_days$alcday5 <- ifelse(alc_days$alcday5 %in% 201:209, gsub("20",
    "", alc_days$alcday5), alc_days$alcday5)

alc_days$alcday5 <- ifelse(alc_days$alcday5 %in% 210:230, gsub("2",
    "", alc_days$alcday5), alc_days$alcday5)

alc_days$alcday5 <- ifelse(alc_days$alcday5 %in% 1:30, round((as.numeric(alc_days$alcday5))/7,
    2), alc_days$alcday5)

# Remove the '10' from 101-107 values, since 1 _ _ = days
# per week

alc_days$alcday5 <- ifelse(alc_days$alcday5 %in% 101:107, gsub("10",
    "", alc_days$alcday5), alc_days$alcday5)
```

Exploratory Data Analysis and Visualizations

Binge Drinking Prevalence

Calculating the proportion of binge drinkers in each state

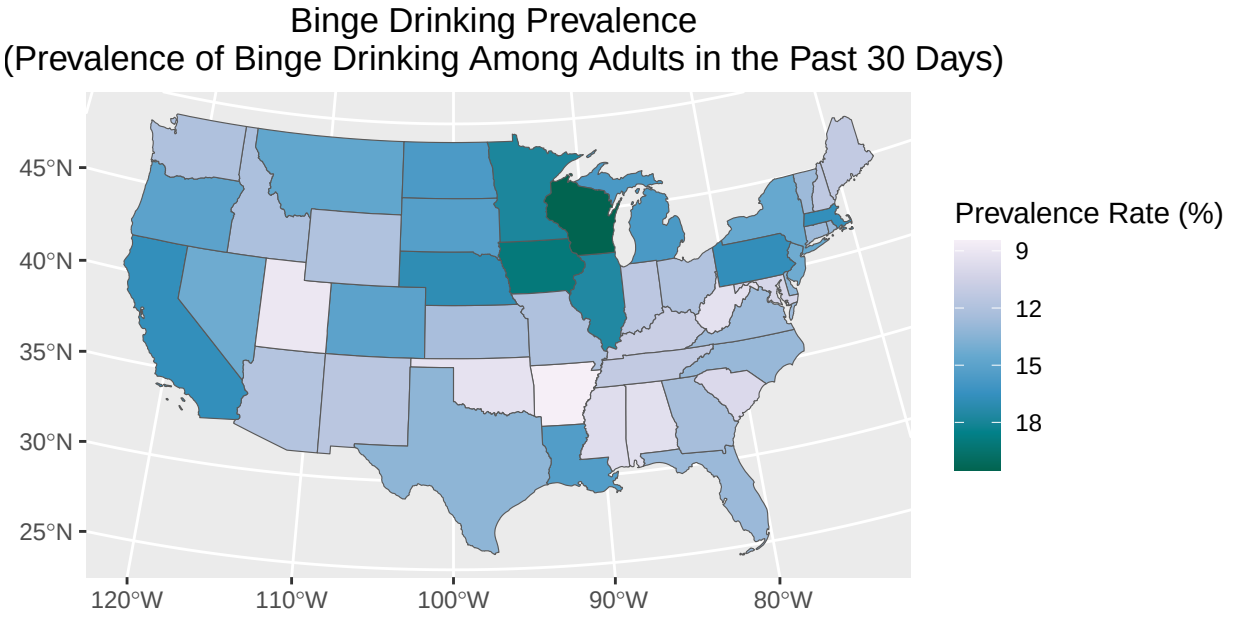
```
yesNo <- data.frame(Yes = rowSums(binge_prev["x_rfbing5"] ==
    2), No = rowSums(binge_prev["x_rfbing5"] == 1))
binge <- aggregate(yesNo, binge_prev["stname"], sum)
binge$percent <- (binge$Yes)/(binge$Yes + binge$No) * 100
binge$percent <- round(binge$percent, 2)
```

Creating a map to display the proportion of binge drinkers in each state

```
states <- st_read("2015-2019-acs-states.geojson")
```

```
## Reading layer ‘2015-2019-acs-states’ from data source
##   ‘C:\Users\jane9\OneDrive\Documents\March Analytics Project\2015-2019-acs-states.geojson’
##   using driver ‘GeoJSON’
## Simple feature collection with 52 features and 93 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -179 ymin: 17.91377 xmax: -65.22157 ymax: 71.35256
## Geodetic CRS:  WGS 84
```

```
states <- states[!(states$ST %in% c("PR", "AK", "HI")), ]

binge <- binge[!(binge$stname %in% c("Puerto Rico ", "Alaska",
    "Hawaii")), ]

states <- st_transform(states, 6580)

states_prevalence <- merge(states, binge[, c("stname", "percent")],
    by.x = "Name", by.y = "stname")

map <- ggplot() + geom_sf(data = states_prevalence, aes(fill = binge$percent)) +
    scale_fill_distiller(palette = "PuBuGn", trans = "reverse") +
    labs(title = "Binge Drinking Prevalence \n (Prevalence of Binge Drinking Among Adults in the Past 3
        fill = "Prevalence Rate (%)") + theme(plot.title = element_text(hjust = 0.5))

print(map)
```



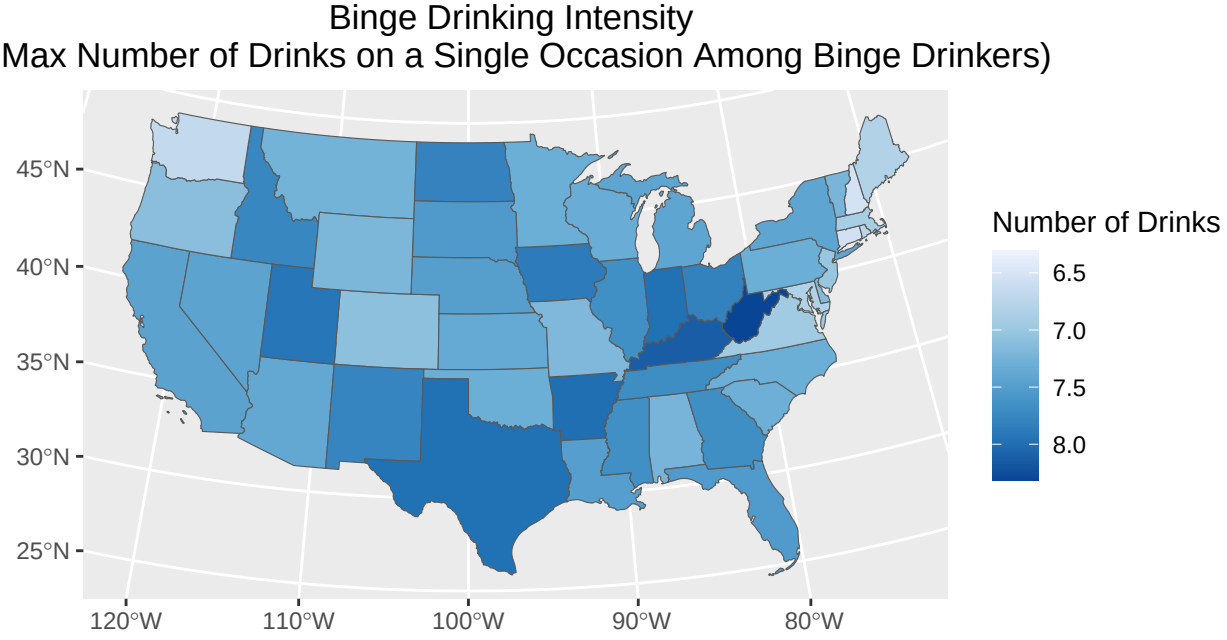
Binge Drinking Intensity

Looking at state means and quartiles of 'most drinks consumed on a single occasion in the past 30 days' among those who engaged in binge drinking in the past 30 days

```
# Creating a subset of the 'max drinks' dataframe, only
# including those who have engaged in binge drinking in the
# past 30 days (>= 4 drinks for women or >= 5 drinks for
# men)
max_drinks_binge <- max_drinks[(max_drinks$maxdrnks >= 5 & max_drinks$sex1 ==
    1) | (max_drinks$maxdrnks >= 4 & max_drinks$sex1 == 2), ]

# Finding state means and quantiles

intensity_state_means <- aggregate(max_drinks_binge$maxdrnks,
    list(max_drinks_binge$stname), FUN = mean)  #Mean
intensity_state_quantiles <- aggregate(max_drinks_binge$maxdrnks,
    list(max_drinks_binge$stname), FUN = quantile)  #Quantiles
intensity_state_75 <- data.frame(intensity_state_quantiles$Group.1,
    intensity_state_quantiles$x[, "75%"])  #75th percentile
```

Creating a map to display the state means for binge drinking intensity

```
intensity_state_means <- intensity_state_means[!(intensity_state_means$Group.1 %in%
    c("Puerto Rico ", "Alaska", "Hawaii")), ]

states_bngdrnk <- merge(states, intensity_state_means[, c("Group.1",
    "x")], by.x = "Name", by.y = "Group.1")

map <- ggplot() + geom_sf(data = states_bngdrnk, aes(fill = x)) +
    scale_fill_distiller(palette = "Blues", trans = "reverse") +
    labs(title = "Binge Drinking Intensity \n (Max Number of Drinks on a Single Occasion Among Binge Dri
        fill = "Number of Drinks") + theme(plot.title = element_text(hjust = 0.5))

print(map)
```



Binge Drinking Frequency

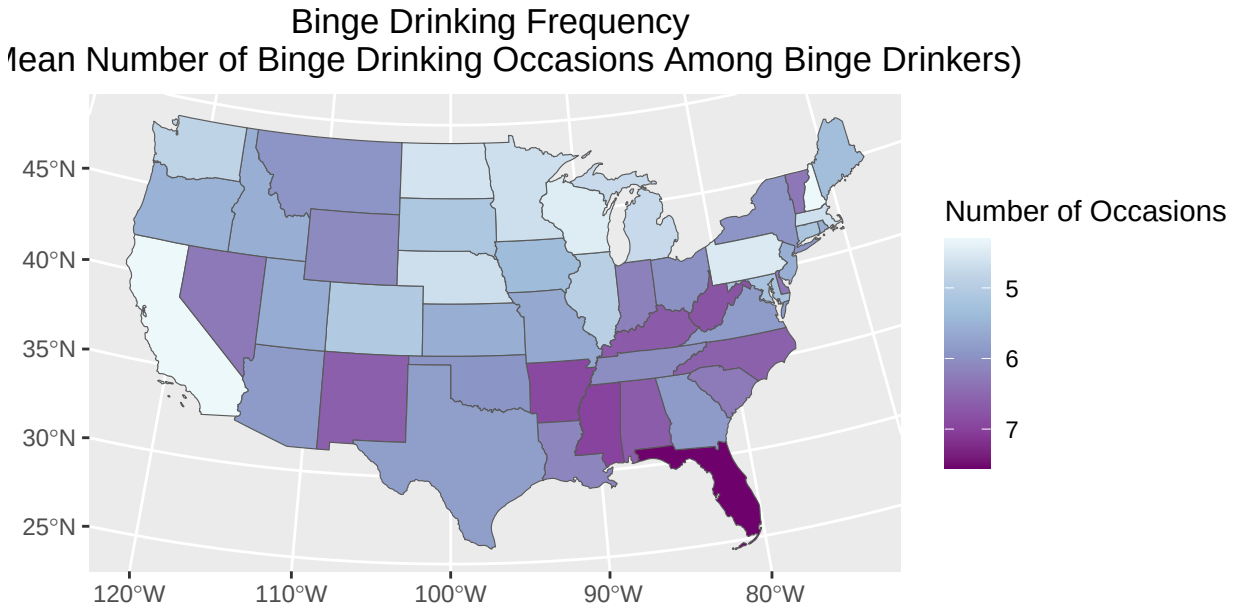
Calculating state means for binge drinking frequency among those who engage in binge drinking

```
# Creating a subset of the data, which only includes those
# who have engaged in binge drinking in the past 30 days
```

```
binge_freq <- binge_freq[(binge_freq$maxdrnks %in% 1:76) & !is.na(binge_freq$maxdrnks),
    ]
binge_freq <- binge_freq[(binge_freq$maxdrnks >= 5 & binge_freq$sex1 ==
    1) | (binge_freq$maxdrnks >= 4 & binge_freq$sex1 == 2), ]

# Finding state means and quantiles

frequency_state_means <- aggregate(binge_freq$drnk3ge5, list(binge_freq$stname),
    FUN = mean)  #Mean
frequency_state_quantiles <- aggregate(binge_freq$drnk3ge5, list(binge_freq$stname),
    FUN = quantile)  #Quantiles
frequency_state_75 <- data.frame(frequency_state_quantiles$Group.1,
    intensity_state_quantiles$x[, "75%"])  #75th percentile
```

Creating a map displaying binge drinking frequency

```
frequency_state_means <- frequency_state_means[!(frequency_state_means$Group.1 %in%
    c("Puerto Rico ", "Alaska", "Hawaii")), ]

states_freq <- merge(states, frequency_state_means[, c("Group.1",
    "x")], by.x = "Name", by.y = "Group.1")

map <- ggplot() + geom_sf(data = states_freq, aes(fill = x)) +
    scale_fill_distiller(palette = "BuPu", trans = "reverse") +
    labs(title = "Binge Drinking Frequency \n (Mean Number of Binge Drinking Occasions Among Binge Drink
        fill = "Number of Occasions") + theme(plot.title = element_text(hjust = 0.5))

print(map)
```



Depressive Disorders and Mental/Physical Health Condition vs Binge Drinking

```
# Creating a subset of the data, excluding NA or 'not
# sure'/'refused' values for binge drinking, depressive
# disorder, physical health, and general health

depression_prev <- binge_prev[(binge_prev$addepev2 == 1 | binge_prev$addepev2 ==
    2) & !is.na(binge_prev$addepev2), ]
health_prev <- depression_prev[(depression_prev$genhlth %in%
    1:5) & !is.na(depression_prev$genhlth) & (depression_prev$menthlth %in%
    1:30 | depression_prev$menthlth == 88) & !is.na(depression_prev$menthlth) &
    (depression_prev$physhlth %in% 1:30 | depression_prev$physhlth ==
        88) & !is.na(depression_prev$physhlth), ]
health_prev$menthlth <- as.numeric(gsub("88", "0", health_prev$menthlth))  # Changing '88' to '0', sinc
health_prev$physhlth <- as.numeric(gsub("88", "0", health_prev$physhlth))  # Changing '88' to '0', sinc
```

Constructing a Logistic Regression Model to Predict Whether a Survey Respondent is a Binge Drinker Based on Their Health Condition

Preparing the Data for Logistic Regression

```r
health_prev$x_rfbing5 <- as.numeric(gsub("1", "0", health_prev$x_rfbing5))  # No = 0
health_prev$x_rfbing5 <- as.numeric(gsub("2", "1", health_prev$x_rfbing5))  # Yes = 1

health_prev$addepev2 <- as.numeric(gsub("2", "0", health_prev$addepev2))  # No = 0, Yes = 1

health_prev <- health_prev[c("x_rfbing5", "addepev2", "menthlth",
    "physhlth", "genhlth")]


col_names <- c("x_rfbing5", "addepev2")

health_prev[, col_names] <- lapply(health_prev[, col_names],
    factor)
```

Logistic Regression Model: Binge Drinker ~ Depressive Disorder + Mental Health + Physical Health + General Health

```r
model <- glm(x_rfbing5 ~ addepev2 + menthlth + physhlth + genhlth,
    family = binomial(link = "logit"), data = health_prev)
summary(model)
```

```
##
## Call:
## glm(formula = x_rfbing5 ~ addepev2 + menthlth + physhlth + genhlth,
##     family = binomial(link = "logit"), data = health_prev)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.5511005  0.0124565 -124.52   <2e-16 ***
## addepev21    0.0213137  0.0132365    1.61    0.107
## menthlth     0.0259187  0.0006286   41.23   <2e-16 ***
## physhlth    -0.0222096  0.0007225  -30.74   <2e-16 ***
## genhlth     -0.1272917  0.0052016  -24.47   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 315952  on 396886  degrees of freedom
## Residual deviance: 312393  on 396882  degrees of freedom
## AIC: 312403
##
## Number of Fisher Scoring iterations: 5
```

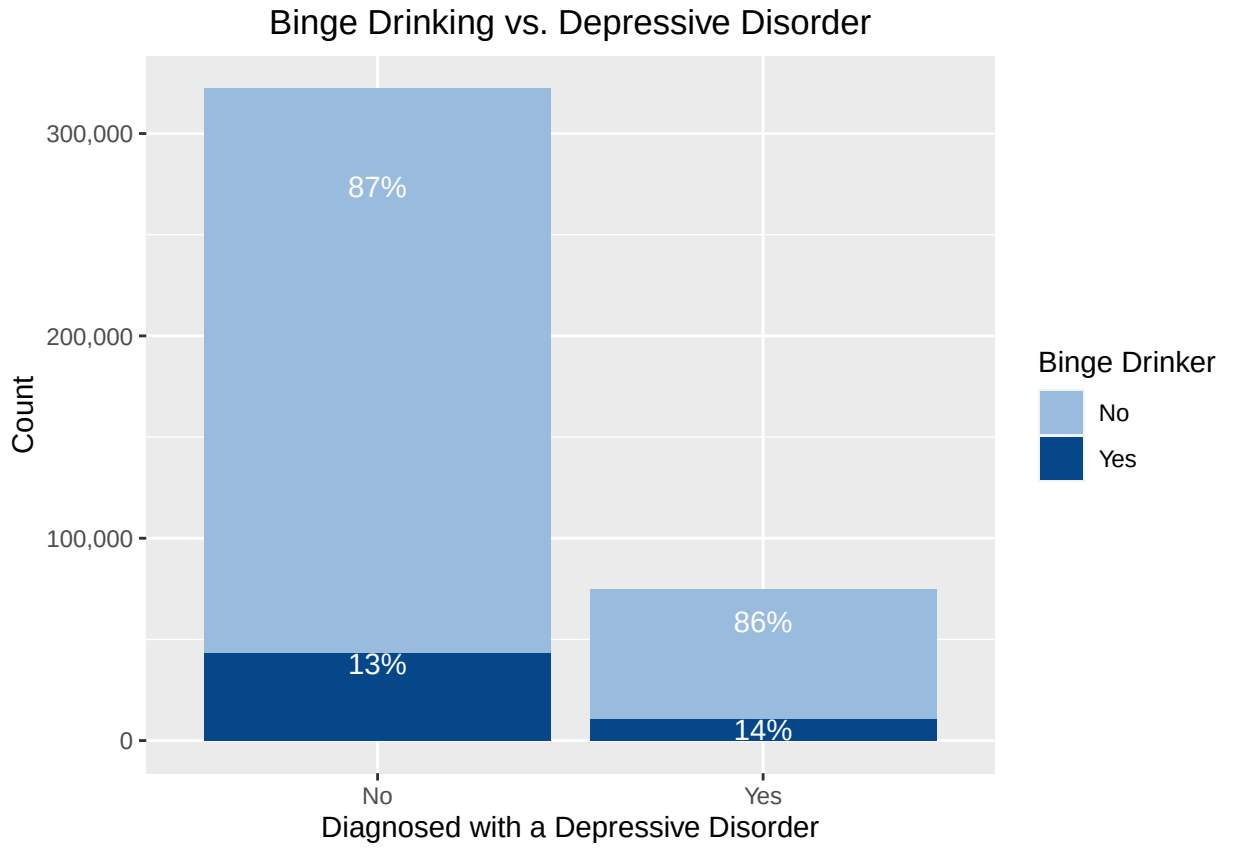
```r
# Displaying results of logistic regression as a table
tab_model(model, dv.labels = "Binge Drinker (No = 0, Yes = 1)",
    pred.labels = c("(Intercept)", "Depressive Disorder", "Mental Health",
        "Physical Health", "General Health"), show.est = TRUE,
    show.stat = TRUE)
```

```
## Profiled confidence intervals may take longer time to compute.
```

```
##   Use 'ci_method="wald"' for faster computation of CIs.
```

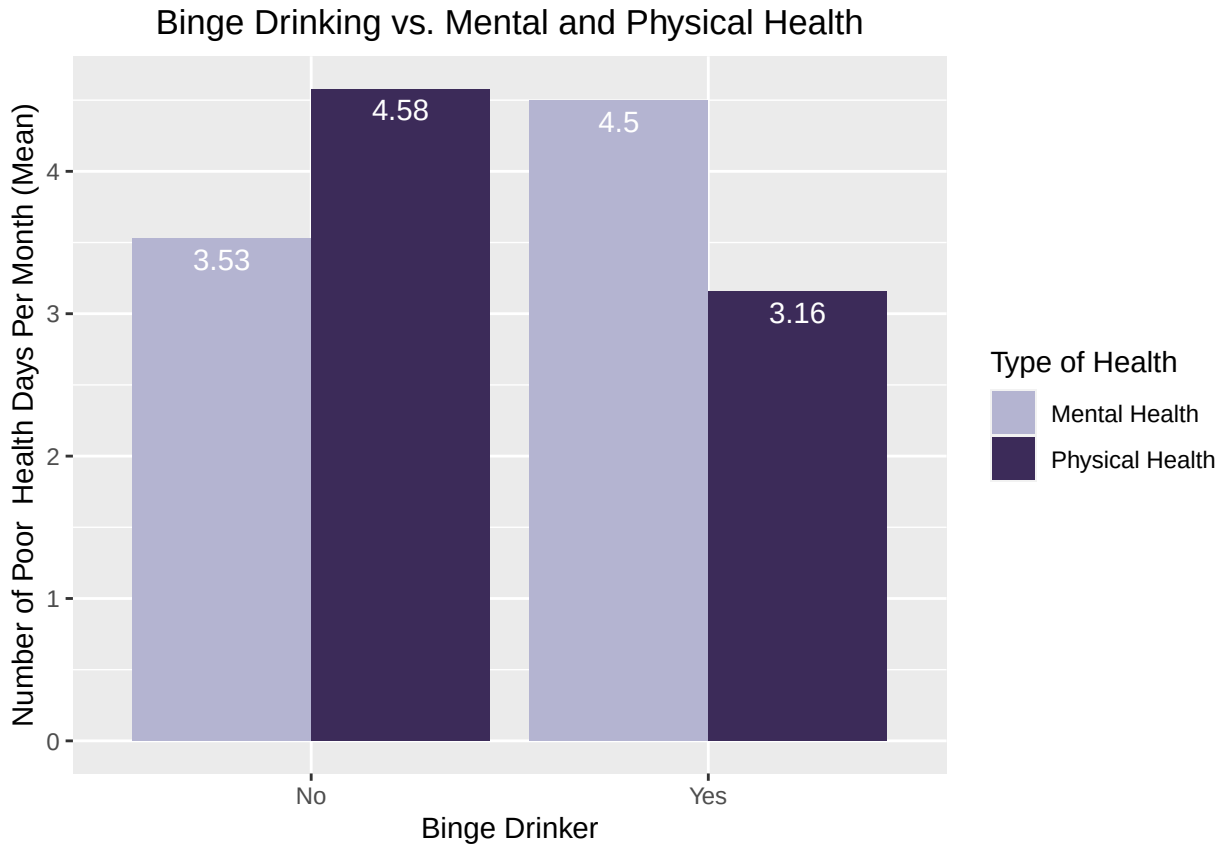
Stacked Bar Plot Showing Binge Drinking vs. Deppressive Disorder

```
dat <- data.frame(table(health_prev$addepev2, health_prev$x_rfbing5))
names(dat) <- c("addepev2", "x_rfbing5", "count")
dat <- ddply(dat, .(addepev2), transform, percent = count/sum(count) *
    100)
dat$label = paste0(sprintf("%.0f", dat$percent), "%")

ggplot(data = dat, aes(x = addepev2, y = count, fill = x_rfbing5)) +
    geom_bar(stat = "identity") + scale_x_discrete(labels = c("No",
    "Yes")) + scale_fill_manual(values = c("#99bbde", "#06478a"),
    labels = c("No", "Yes")) + labs(title = "Binge Drinking vs. Depressive Disorder",
    fill = "Binge Drinker", x = "Diagnosed with a Depressive Disorder",
    y = "Count") + geom_text(aes(label = label), hjust = 0.5,
    vjust = 1, size = 4, color = "white") + theme(plot.title = element_text(hjust = 0.5)) +
    scale_y_continuous(labels = scales::comma_format())
```



Binge Drinking vs. Poor Mental and Physical Health

```
dat2 <- health_prev[c("x_rfbing5", "menthlth", "physhlth")]
dat2 <- aggregate(. ~ x_rfbing5, dat2, mean)

dat2$rfbing5 <- row.names(dat2)
dat2_long <- gather(dat2, key = var, value = value, menthlth,
    physhlth)

ggplot(dat2_long, aes(x = x_rfbing5, y = value, fill = var)) +
    geom_bar(stat = "identity", position = "dodge") + scale_x_discrete(labels = c("No",
    "Yes")) + scale_fill_manual(values = c("#b4b4d1", "#3c2b59"),
    labels = c("Mental Health", "Physical Health")) + labs(title = "Binge Drinking vs. Mental and Physi
    fill = "Type of Health", x = "Binge Drinker", y = "Number of Poor  Health Days Per Month (Mean)") +
    geom_text(aes(label = round(value, 2)), vjust = 1.5, position = position_dodge(0.9),
        size = 4, color = "white") + theme(plot.title = element_text(hjust = 0.5))
```



Binge Drinking vs. General Health

```
dat3 <- aggregate(genhlth ~ x_rfbing5, health_prev, median)
ggplot(data = dat3, aes(x = x_rfbing5, y = genhlth, fill = x_rfbing5)) +
    geom_bar(stat = "identity", width = 0.5) + scale_x_discrete(labels = c("No",
    "Yes")) + scale_fill_manual(values = c("#a7d4bf", "#3a6953"),
    labels = c("No", "Yes")) + labs(title = "Binge Drinking vs. General Health",
    fill = "Binge Drinker", x = "Binge Drinker", y = "Median General Health Rating (1-5)") +
```

```
    geom_text(aes(label = round(genhlth, 2)), hjust = 0.5, vjust = 4,
        size = 4, color = "white") + theme(plot.title = element_text(hjust = 0.5))
```