library(ggplot2)

library(tidyverse)

library(gmodels)

library(vcdExtra)

library(DescTools)

# Read in the bike data

bike <- read.csv('https://raw.githubusercontent.com/IAA-Faculty/statistical\_foundations/master/bike.csv')

# Split into training and test datasets

set.seed(123)

bike <- bike %>% mutate(id = row\_number())

train <- bike %>% sample\_frac(0.7)

test <- anti\_join(bike, train, by = 'id')

# Create the casual\_high variable

train$casual\_high <- train$casual >= train$registered

table(train$casual\_high)

# Chi-square test and cross-tabulation table between casual\_high and season

CrossTable(train$season, train$casual\_high)

chisq.test(table(train$season, train$casual\_high))

CMHtest(table(train$season, train$casual\_high))$table[1,]

# Chi-square test and cross-tabulation table between casual\_high and holiday

CrossTable(train$holiday, train$casual\_high)

CMHtest(table(train$holiday, train$casual\_high))$table[1,]

OddsRatio(table(train$holiday, train$casual\_high))