library(ggplot2)

library(tidyverse)

# 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)

# Logistic regression model with all variables

bike.logit <- glm(casual\_high ~ factor(season) + factor(yr) + factor(hr) +

holiday + workingday + weathersit + atemp + temp +

hum + windspeed,

data = train, family = binomial(link = "logit"))

summary(bike.logit)

# Reduced logistic regression model

bike.logit2 <- glm(casual\_high ~ factor(season) + factor(yr) + factor(hr) +

holiday + workingday + atemp,

data = train, family = binomial(link = "logit"))

summary(bike.logit2)