# EDA on Ride Sharing Dataset

A white rectangular box with black text

AI-generated content may be incorrect.

R code:  
# Load necessary libraries

if (!require(readxl)) install.packages('readxl', dependencies=TRUE)

if (!require(dplyr)) install.packages('dplyr', dependencies=TRUE)

if (!require(ggplot2)) install.packages('ggplot2', dependencies=TRUE)

if (!require(lubridate)) install.packages('lubridate', dependencies=TRUE)

library(readxl)

library(dplyr)

library(ggplot2)

library(lubridate)

# Load the dataset

ride\_data <- read\_excel("C:/Users/Msi/Desktop/R folder/Group Work/EDA Ride Sharing/Ride Sharing Dataset.xlsx")

# Display the first few rows of the dataset

print(head(ride\_data))

# Summary of the dataset

print(summary(ride\_data))

# Check for missing values

print(sum(is.na(ride\_data)))

# Convert 'Request Time' to a proper datetime format

ride\_data$`Request Time` <- as.POSIXct(ride\_data$`Request Time`, origin="1970-01-01")

# Extract additional time-based features

ride\_data$Hour <- hour(ride\_data$`Request Time`)

ride\_data$Day <- wday(ride\_data$`Request Time`, label=TRUE)

ride\_data$Month <- month(ride\_data$`Request Time`, label=TRUE)

# Distribution of rides by hour

p1 <- ggplot(ride\_data, aes(x=Hour)) +

  geom\_histogram(binwidth=1, fill="blue", color="black") +

  labs(title="Distribution of Rides by Hour", x="Hour of the Day", y="Number of Rides")

print(p1)

# Distribution of rides by day of the week

p2 <- ggplot(ride\_data, aes(x=Day)) +

  geom\_bar(fill="orange", color="black") +

  labs(title="Distribution of Rides by Day of the Week", x="Day of the Week", y="Number of Rides")

print(p2)

# Distribution of rides by month

p3 <- ggplot(ride\_data, aes(x=Month)) +

  geom\_bar(fill="green", color="black") +

  labs(title="Distribution of Rides by Month", x="Month", y="Number of Rides")

print(p3)

# Average fare amount by vehicle type

p4 <- ride\_data %>%

  group\_by(`Vehicle Type`) %>%

  summarise(Average\_Fare = mean(`Fare Amount (in $)`)) %>%

  ggplot(aes(x=`Vehicle Type`, y=Average\_Fare)) +

  geom\_bar(stat="identity", fill="purple", color="black") +

  labs(title="Average Fare Amount by Vehicle Type", x="Vehicle Type", y="Average Fare Amount ($)")

print(p4)

# Relationship between ride distance and fare amount

p5 <- ggplot(ride\_data, aes(x=`Ride Distance (in miles)`, y=`Fare Amount (in $)`)) +

  geom\_point(color="red") +

  geom\_smooth(method="lm", color="blue") +

  labs(title="Relationship Between Ride Distance and Fare Amount", x="Ride Distance (miles)", y="Fare Amount ($)")

print(p5)

# User ratings distribution

p6 <- ggplot(ride\_data, aes(x=`User Rating`)) +

  geom\_histogram(binwidth=0.5, fill="yellow", color="black") +

  labs(title="Distribution of User Ratings", x="User Rating", y="Count")

print(p6)

# Payment method preferences

p7 <- ride\_data %>%

  group\_by(`Payment Method`) %>%

  summarise(Count = n()) %>%

  ggplot(aes(x=`Payment Method`, y=Count)) +

  geom\_bar(stat="identity", fill="cyan", color="black") +

  labs(title="Payment Method Preferences", x="Payment Method", y="Count")

print(p7)

# Traffic conditions impact on ride distance

p8 <- ride\_data %>%

  group\_by(`Traffic Condition`) %>%

  summarise(Average\_Distance = mean(`Ride Distance (in miles)`)) %>%

  ggplot(aes(x=`Traffic Condition`, y=Average\_Distance)) +

  geom\_bar(stat="identity", fill="pink", color="black") +

  labs(title="Impact of Traffic Conditions on Ride Distance", x="Traffic Condition", y="Average Ride Distance (miles)")

print(p8)

# Peak hours analysis

p9 <- ride\_data %>%

  group\_by(`Peak Hours`) %>%

  summarise(Count = n()) %>%

  ggplot(aes(x=`Peak Hours`, y=Count)) +

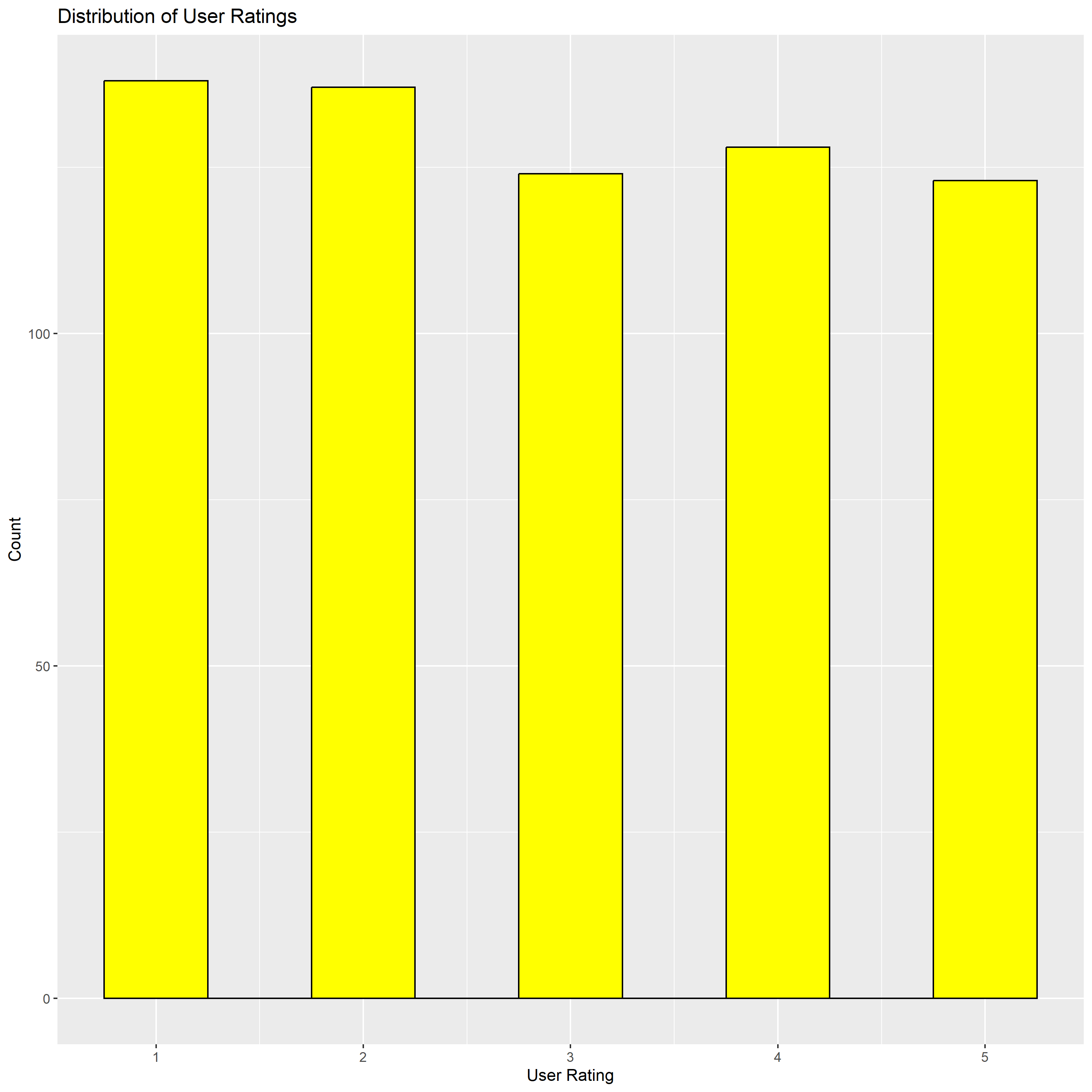
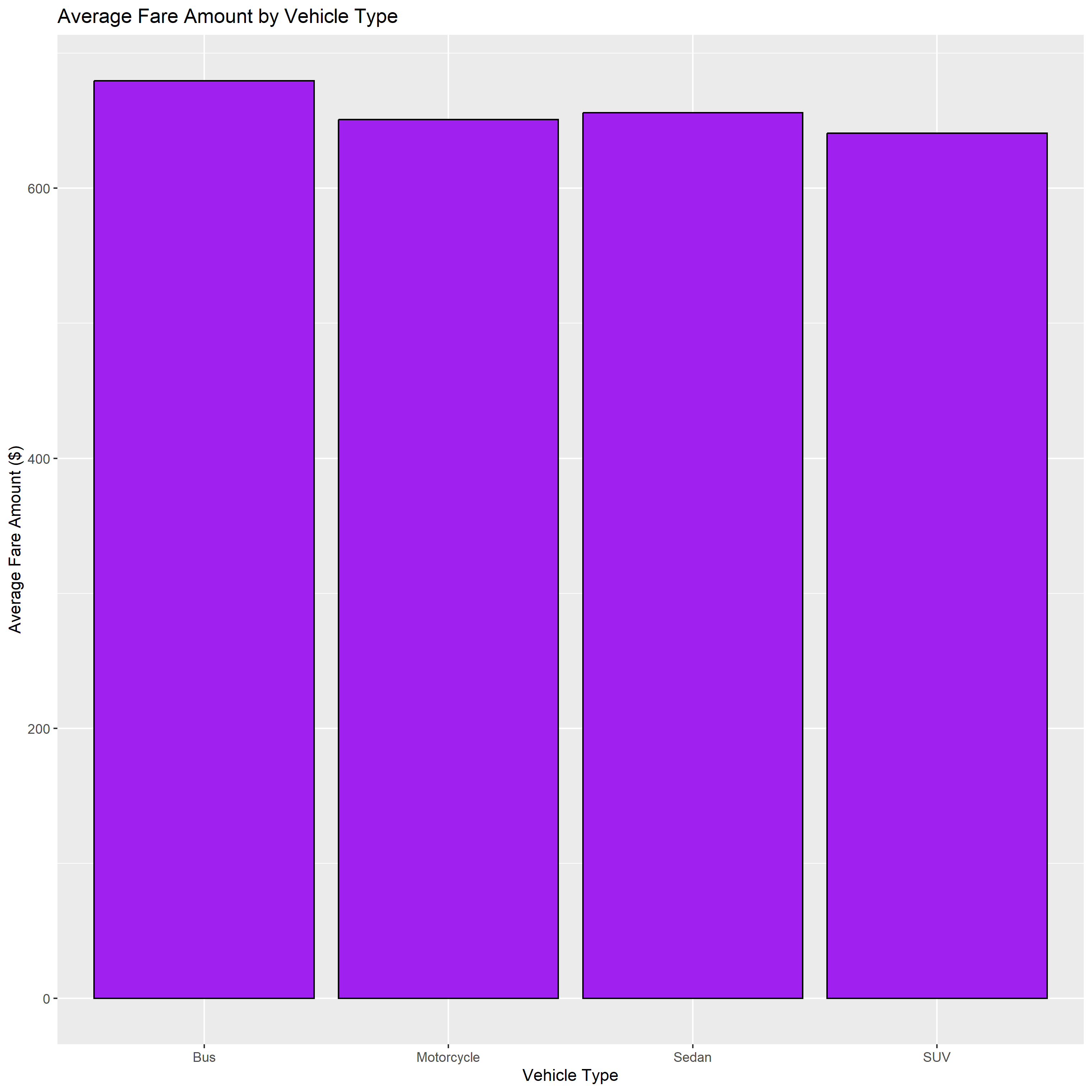
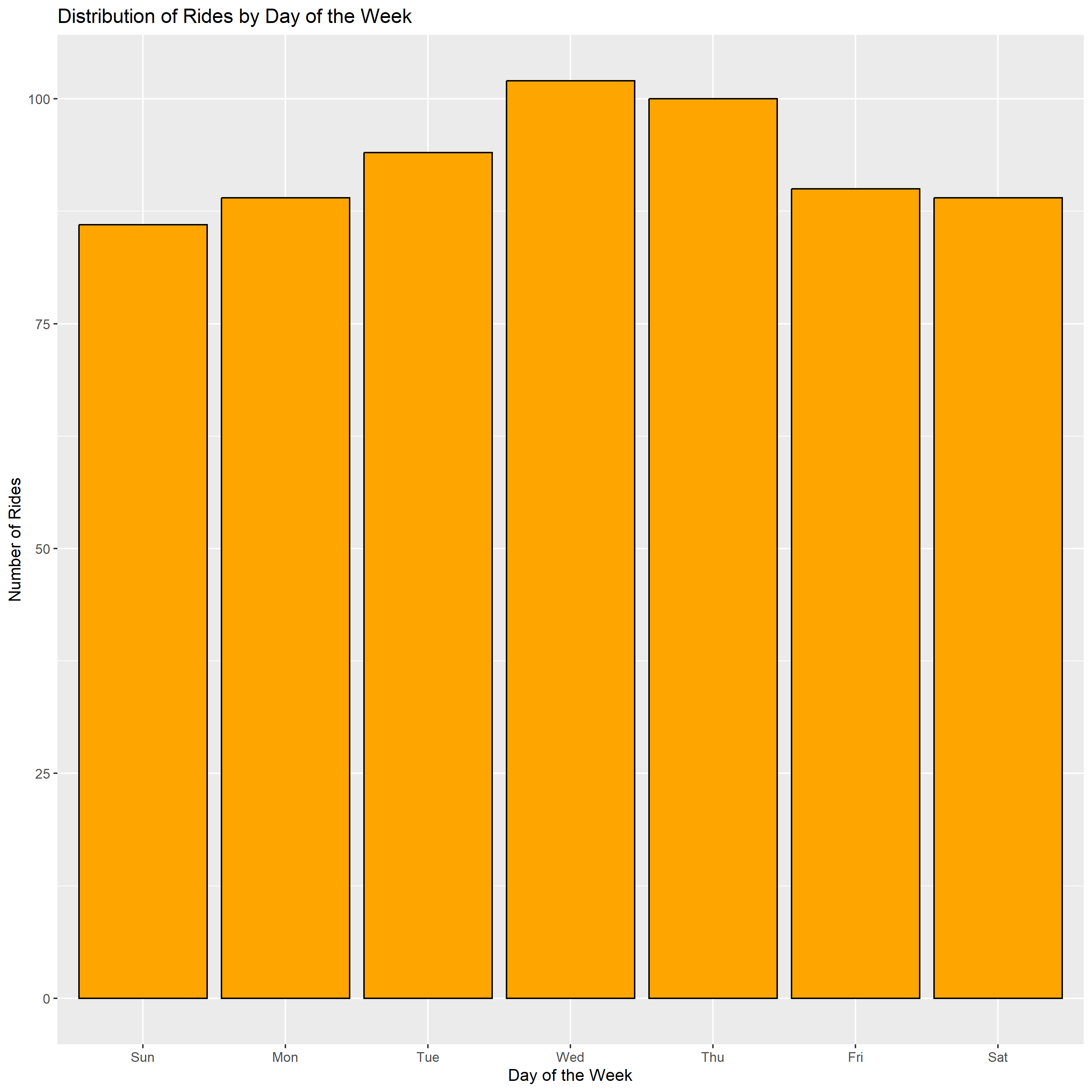
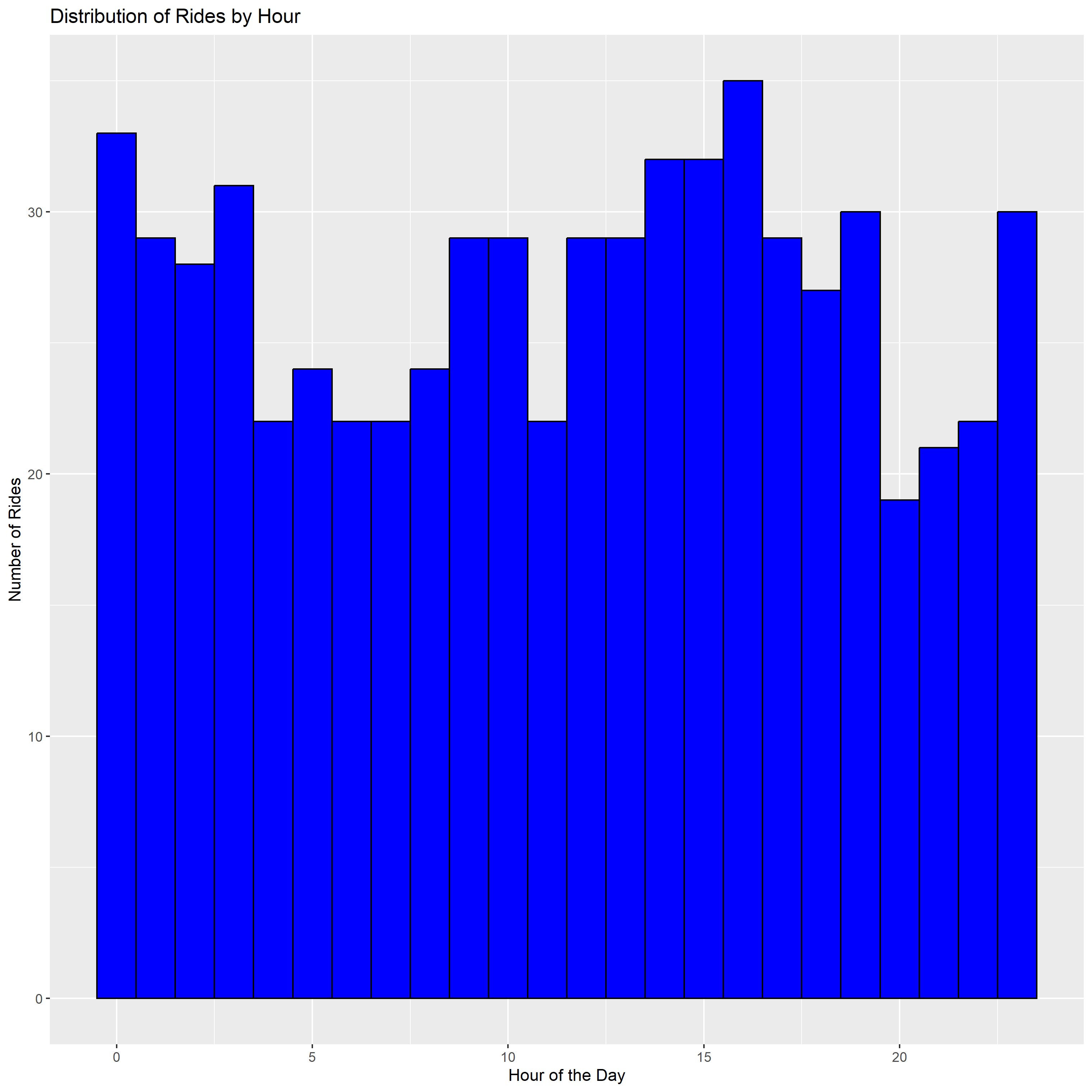
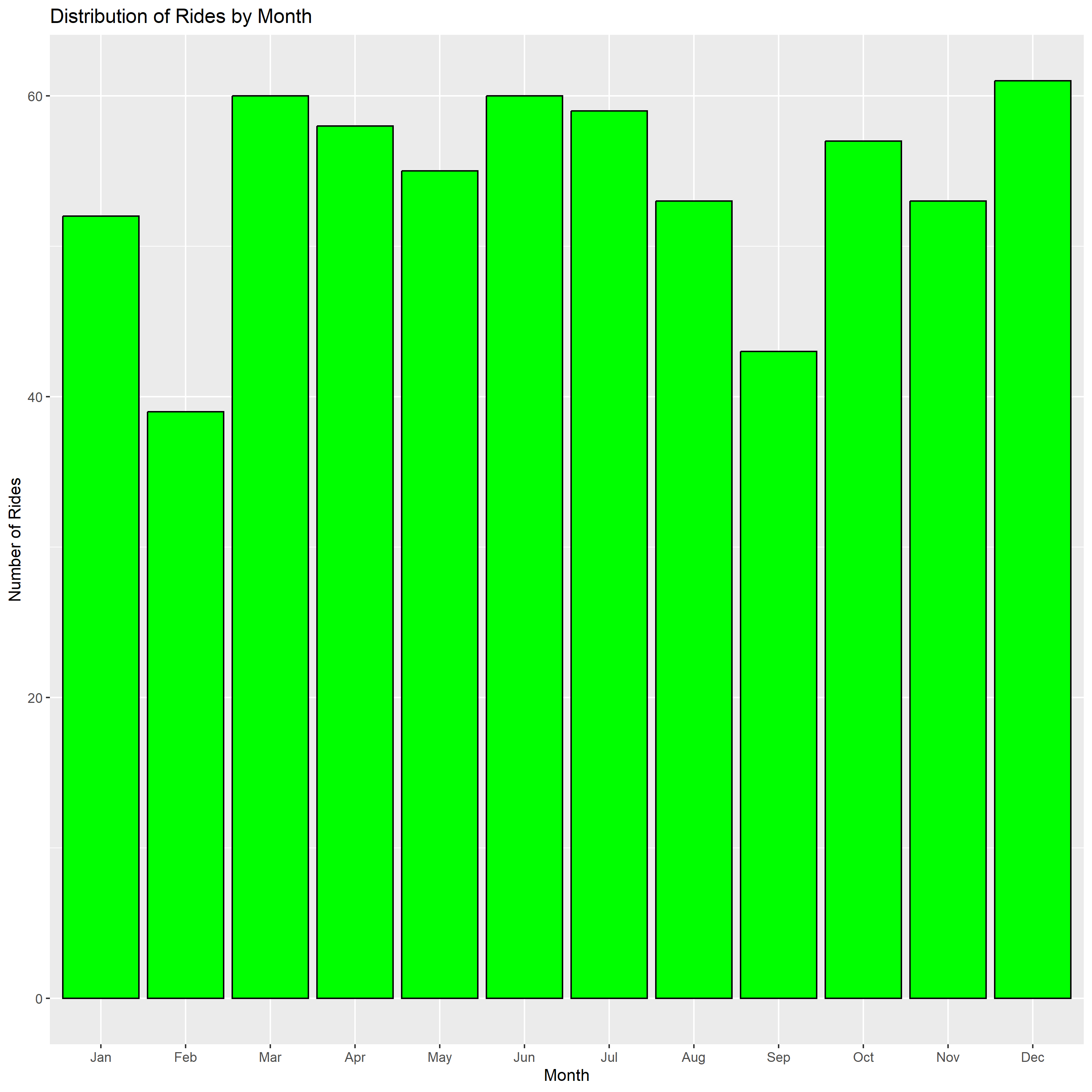
  geom\_bar(stat="identity", fill="brown", color="black") +

  labs(title="Ride Count During Peak Hours", x="Peak Hours", y="Number of Rides")

print(p9)

A red rectangular object with white lines

AI-generated content may be incorrect.



A graph with a red line

AI-generated content may be incorrect.A graph of blue rectangular bars

AI-generated content may be incorrect.