library(dplyr)

library(ggplot2)

library(lubridate)

# Adjust the file path to your data file

data <- read.csv("C:/Users/Owner/Documents/DataScience/Reproducible Research/Scripts/activity.csv")

# Calculate the total number of steps taken each day

total\_steps\_per\_day <- data %>%

group\_by(date) %>%

summarise(total\_steps = sum(steps, na.rm = TRUE))

# Create a histogram of the total steps per day

ggplot(total\_steps\_per\_day, aes(x = total\_steps)) +

geom\_histogram(binwidth = 1000, fill = "blue", color = "black") +

labs(title = "Histogram of Total Steps per Day", x = "Total Steps", y = "Frequency")

# Calculate the mean and median number of steps taken per day

mean\_steps\_per\_day <- mean(total\_steps\_per\_day$total\_steps)

median\_steps\_per\_day <- median(total\_steps\_per\_day$total\_steps)

# Create a time series plot of the average number of steps taken

average\_steps\_per\_interval <- data %>%

group\_by(interval) %>%

summarise(avg\_steps = mean(steps, na.rm = TRUE))

ggplot(average\_steps\_per\_interval, aes(x = interval, y = avg\_steps)) +

geom\_line(color = "blue") +

labs(title = "Time Series Plot of Average Steps per Interval", x = "Interval", y = "Average Steps")

# Determine the 5-minute interval with the maximum number of steps

max\_steps\_interval <- average\_steps\_per\_interval$interval[which.max(average\_steps\_per\_interval$avg\_steps)]

# Impute missing data by replacing it with the average steps for the corresponding interval

data\_imputed <- data

for (i in 1:nrow(data\_imputed)) {

if (is.na(data\_imputed$steps[i])) {

interval <- data\_imputed$interval[i]

data\_imputed$steps[i] <- average\_steps\_per\_interval$avg\_steps[average\_steps\_per\_interval$interval == interval]

}

}

# Create a histogram of the total steps per day after imputation

total\_steps\_per\_day\_imputed <- data\_imputed %>%

group\_by(date) %>%

summarise(total\_steps = sum(steps))

ggplot(total\_steps\_per\_day\_imputed, aes(x = total\_steps)) +

geom\_histogram(binwidth = 1000, fill = "blue", color = "black") +

labs(title = "Histogram of Total Steps per Day (After Imputation)", x = "Total Steps", y = "Frequency")

# Create a panel plot comparing the average steps per interval across weekdays and weekends

data\_imputed$date <- as.Date(data\_imputed$date)

data\_imputed$day\_type <- ifelse(weekdays(data\_imputed$date) %in% c("Saturday", "Sunday"), "weekend", "weekday")

average\_steps\_per\_interval\_day\_type <- data\_imputed %>%

group\_by(interval, day\_type) %>%

summarise(avg\_steps = mean(steps))

ggplot(average\_steps\_per\_interval\_day\_type, aes(x = interval, y = avg\_steps)) +

geom\_line() +

facet\_grid(day\_type ~ .) +

labs(title = "Average Steps per Interval: Weekday vs Weekend", x = "Interval", y = "Average Steps")

C:\Users\Owner\Documents\RProjects\PeerAssesment1\activity.csv

C:\Users\Owner\Documents\RProjects\PeerAssesment1\PeerAssment1.Rproj

C:\Users\Owner\Documents\RProjects\PeerAssesment1\PA1\_template.R

C:\Users\Owner\Documents\RProjects\PeerAssesment1\PA1\_template.Rmd

C:\Users\Owner\Documents\RProjects\PeerAssesment1\PA1\_template.html

cd C:\Users\Owner\Documents\RProjects\PeerAssesment1 # Change to the project directory

# Initialize the Git repository if it hasn't been done already

git init

# Add the files to the staging area

git add activity.csv PeerAssessment1.Rproj PA1\_template.R PA1\_template.Rmd PA1\_template.html

# Commit the changes

git commit -m "Initial commit"

# Force push the files to the GitHub repository

git remote add origin https://github.com/JNEID123/RepData\_PeerAssessment1.git

git push -u origin master –force

Also:

git init

git add .

git commit -m "Initial commit"

git remote add origin <https://github.com/JNEID123/RepData_PeerAssessment1.git>

git push -u origin master --force