project_1.R

bwr590

2024-01-22

```
## 1. load the data into Rstudie, already did 'Extract all' in windows libary

library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.2.3
```

```
activity <- read.csv("C:/Users/bwr590/Downloads/repdata_data_activity/activity.csv")
View(activity)

summary (activity)
```

```
##      steps            date              interval     
##  Min.   :  0.00   Length:17568       Min.   :   0.0  
##  1st Qu.:  0.00   Class :character   1st Qu.: 588.8  
##  Median :  0.00   Mode  :character   Median :1177.5  
##  Mean   : 37.38                      Mean   :1177.5  
##  3rd Qu.: 12.00                      3rd Qu.:1766.2  
##  Max.   :806.00                      Max.   :2355.0  
##  NA's   :2304
```

```
##2.  what is mean total number of steps taken per day?

Steps_a_day <- tapply(activity$steps, activity$date, sum, na.rm=FALSE)
head (Steps_a_day)
```

```
## 2012-10-01 2012-10-02 2012-10-03 2012-10-04 2012-10-05 2012-10-06 
##         NA        126      11352      12116      13294      15420
```

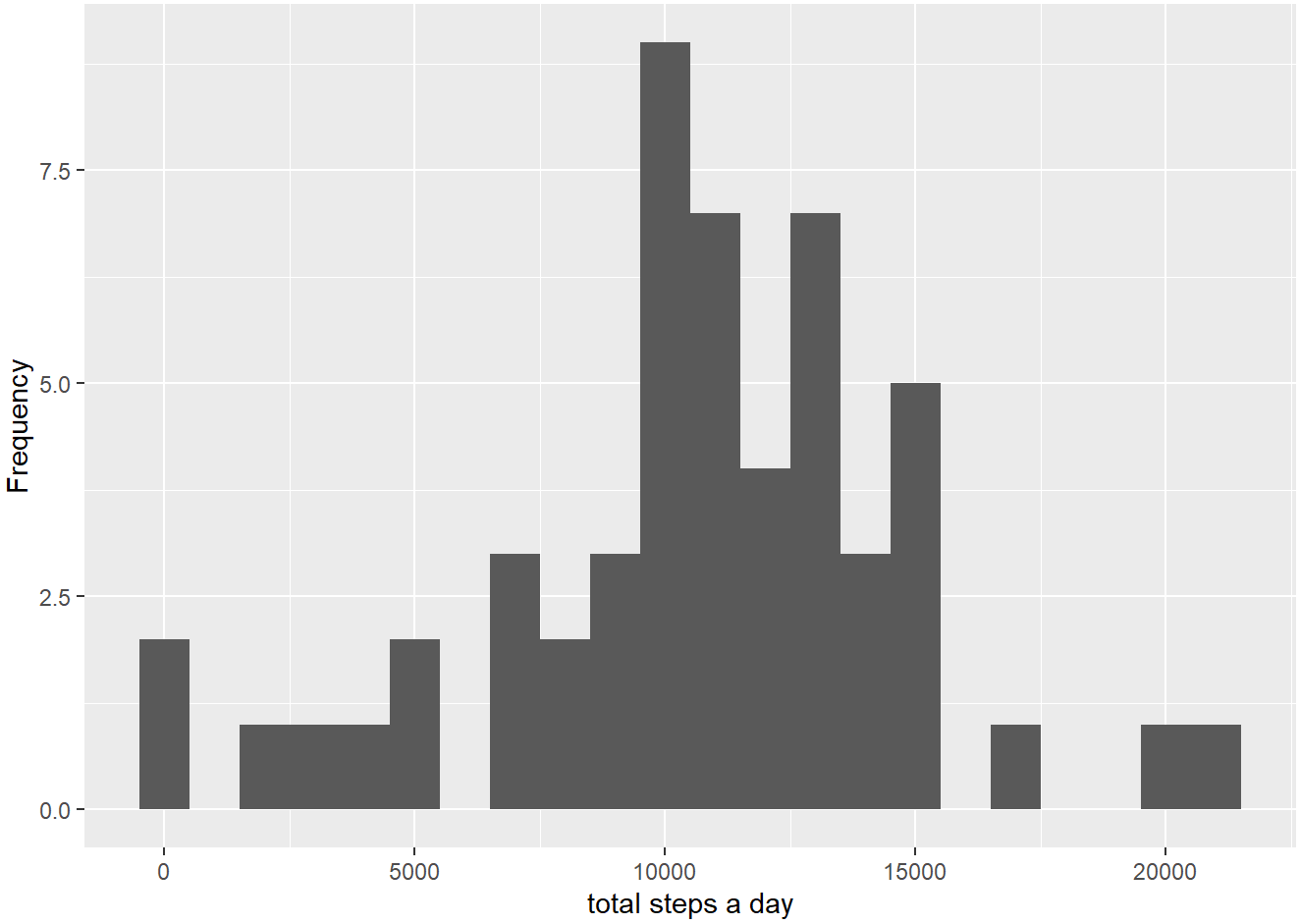
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

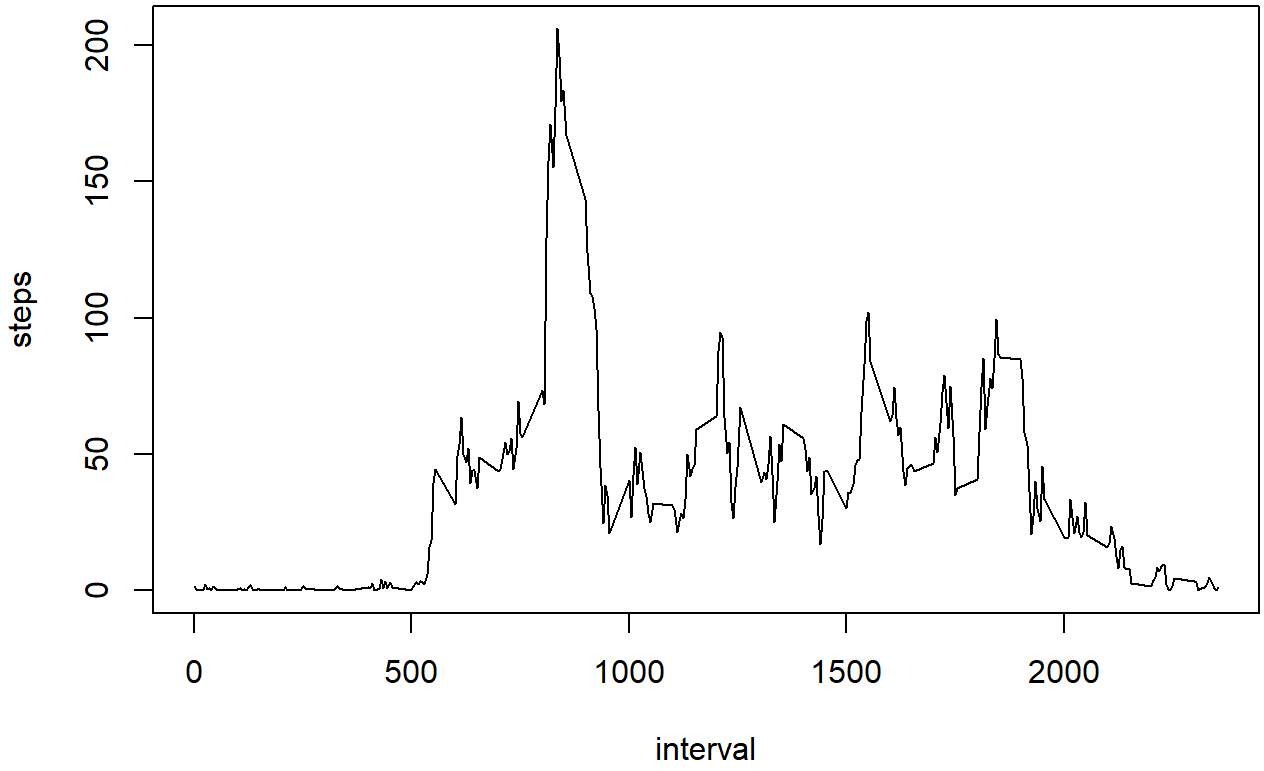
```
qplot (Steps_a_day, xlab= 'total steps a day', ylab= 'Frequency',binwidth = 1000)
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## Warning: Removed 8 rows containing non-finite values (`stat_bin()`).
```




```
mean_steps_a_day <- mean (Steps_a_day, na.rm = TRUE)

median_Steps_a_day <- median (Steps_a_day, na.rm = TRUE)

##3. What is the average daily activity pattern?

Steps_interval<-aggregate(steps~interval,data=activity,mean,na.rm=TRUE)
plot(steps~interval,data=Steps_interval,type="l")
```




```
max(Steps_interval$steps)
```

```
## [1] 206.1698
```

```
Steps_interval[which.max(Steps_interval$steps),]$interval
```

```
## [1] 835
```

```
##4. Imputing missing values

missing_values <- sum(is.na(activity), na.rm = TRUE)
cat("Total number of missing values in the dataset:", missing_values, "\n")
```

```
## Total number of missing values in the dataset: 2304
```

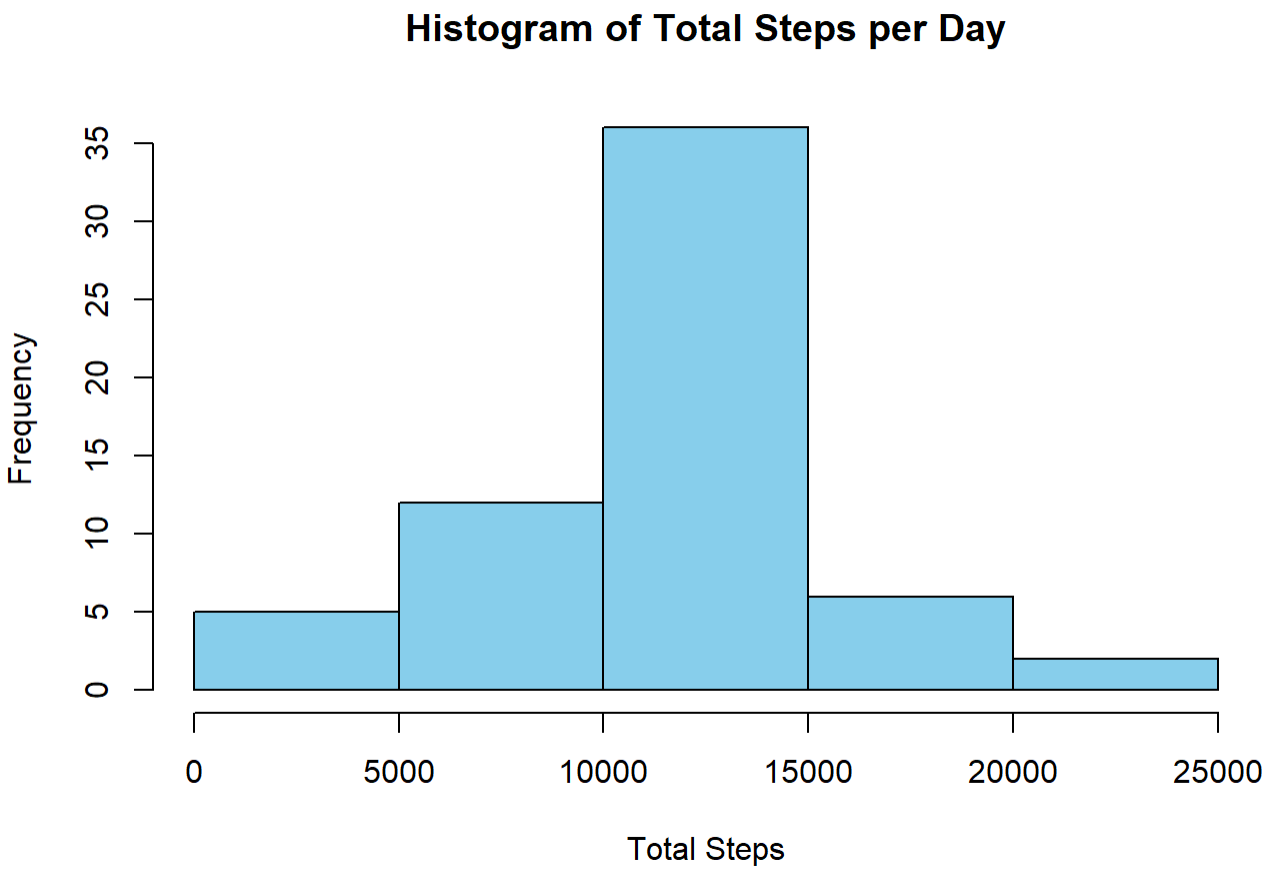
```
activity_filled <- apply(activity, 2, function(x) ifelse(is.na(x), mean(x, na.rm = TRUE), x))
```

```
## Warning in mean.default(x, na.rm = TRUE): argument is not numeric or logical:
## returning NA
```

```
summary(activity_filled)
```

```
##     steps               date               interval
##  Length:17568       Length:17568       Length:17568
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
```

```
activity_filled <- transform(activity, steps = ifelse(is.na(steps), mean(steps, na.rm = TRU
E), steps))

total_steps_per_day <- aggregate(steps ~ date, activity_filled, sum)

hist(total_steps_per_day$steps, main = "Histogram of Total Steps per Day", xlab = "Total Step
s", col = "skyblue", border = "black")
```




```
mean_steps <- mean(total_steps_per_day$steps)
median_steps <- median(total_steps_per_day$steps)

cat("Mean total number of steps per day:", mean_steps, "\n")
```

```
## Mean total number of steps per day: 10766.19
```

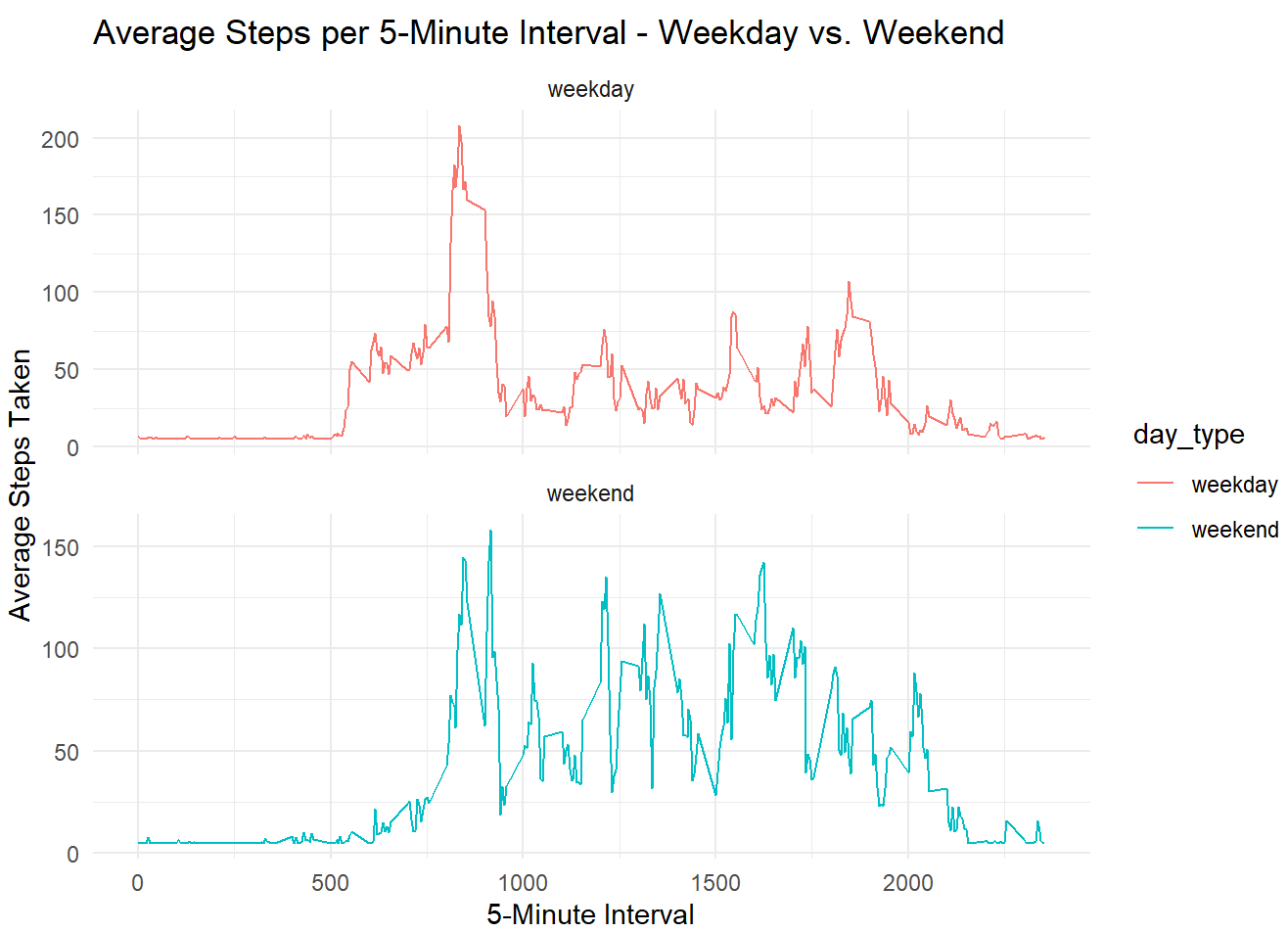
```
cat("Median total number of steps per day:", median_steps, "\n")
```

```
## Median total number of steps per day: 10766.19
```

```
##no impact difference after imputing missing data on the estimates of the daily number of st
eps. This is because I filled the missing data with the mean.

##5. Are there differences in activity patterns between weekdays and weekends?

activity_filled$date <- as.Date(activity_filled$date)
activity_filled$day_type <- ifelse(weekdays(activity_filled$date) %in% c("Monday", "Tuesday",
"Wednesday", "Thursday", "Friday"), "weekday", "weekend")
activity_filled$day_type <- factor(activity_filled$day_type, levels = c("weekday", "weeken
d"))
head(activity_filled)
```

```
##     steps       date interval day_type
## 1 37.3826 2012-10-01        0  weekday
## 2 37.3826 2012-10-01        5  weekday
## 3 37.3826 2012-10-01       10  weekday
## 4 37.3826 2012-10-01       15  weekday
## 5 37.3826 2012-10-01       20  weekday
## 6 37.3826 2012-10-01       25  weekday
```

```
Steps_per_interval <- aggregate(x=list(meanSteps=activity_filled$steps), by=list(interval=act
ivity_filled$interval, day_type=activity_filled$day_type), FUN=mean, na.rm=TRUE)

# Create a panel plot
ggplot(data=Steps_per_interval, aes(x=interval, y=meanSteps)) +
  geom_line(aes(color=day_type)) +
  facet_wrap(~day_type, scales="free_y", ncol=1) +
  xlab("5-Minute Interval") +
  ylab("Average Steps Taken") +
  ggtitle("Average Steps per 5-Minute Interval - Weekday vs. Weekend") +
  theme_minimal()
```


Average Steps per 5-Minute Interval - Weekday vs. Weekend
weekday
weekend
Average Steps Taken
5-Minute Interval
0
50
100
150
200
500
1000
1500
2000
day_type
weekday
weekend