

RRproject1

lm

2023-04-14

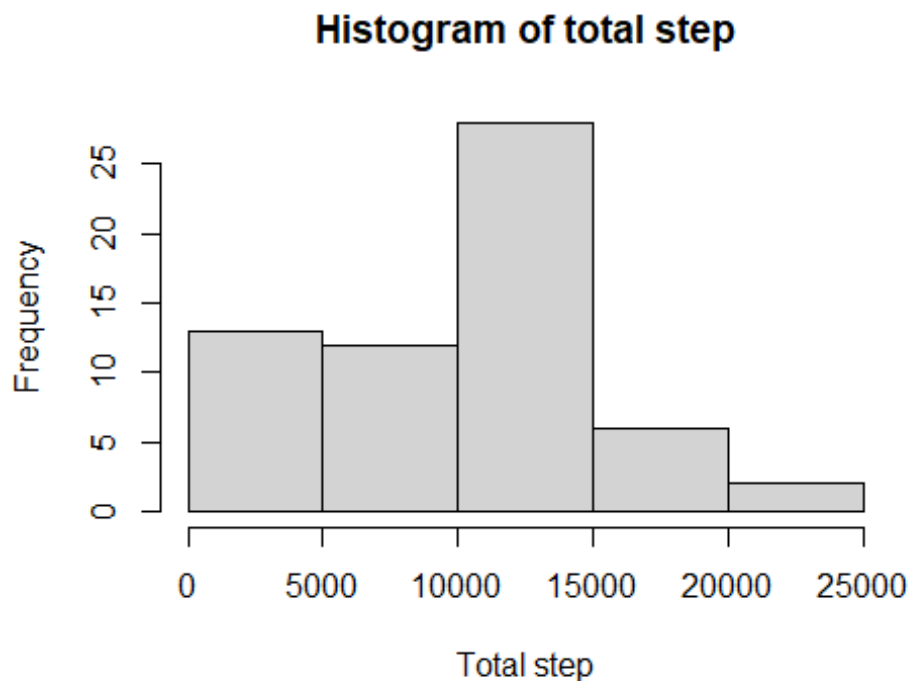
Loading and preprocessing the data

```
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
download.file(fileUrl, "repdata_data_activity.zip")
unzip("repdata_data_activity.zip")
data <- read.csv(unzip("repdata_data_activity.zip"))
```

What is mean total number of steps taken per day?

The total number of steps taken per day was calculated and its histogram was plotted. The mean and median of the total number of steps taken per day calculated and reported.

```
stepsperday <- tapply(data$steps, data$date, sum, na.rm = TRUE)
hist(stepsperday, xlab = "Total step", main="Histogram of total step")
```



```
print(mean (stepsperday))
```

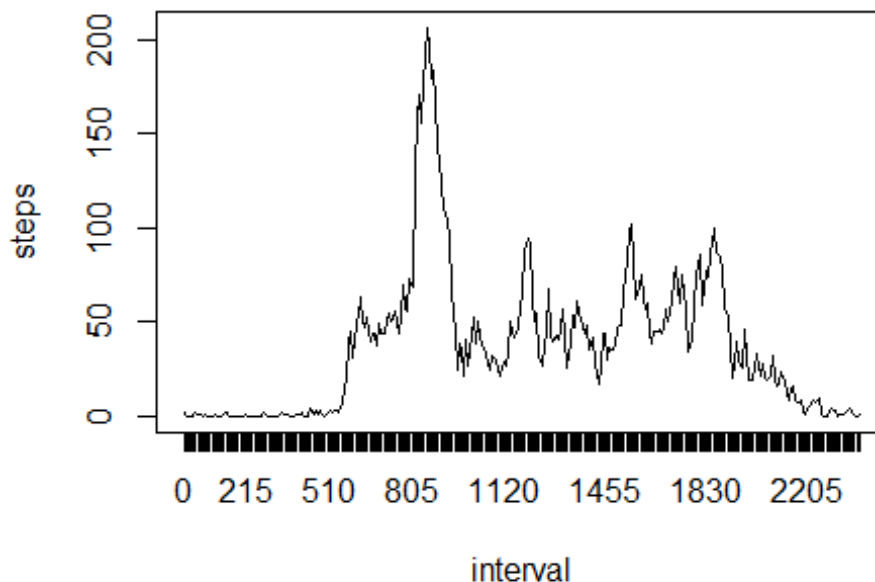
```
## [1] 9354.23
```

```
print(median (stepsperday))  
## [1] 10395
```

What is the average daily activity pattern?

a time series plot of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)

```
xnames <- names(tapply(data$steps, data$interval, mean, na.rm = TRUE))  
plot(tapply(data$steps, data$interval, mean, na.rm = TRUE), xaxt="n", type = "l",  
      xlab= "interval", ylab = "steps")  
axis(1, at=1:length(xnames), labels=xnames)
```



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
agrdata <- aggregate(steps~interval, data, mean, na.rm = TRUE)  
max <- agrdata[which.max(agrdata$steps),]$interval  
max  
## [1] 835
```

##Imputing missing values the total number of missing values in the dataset (i.e. the total number of rows with NAs calculated).

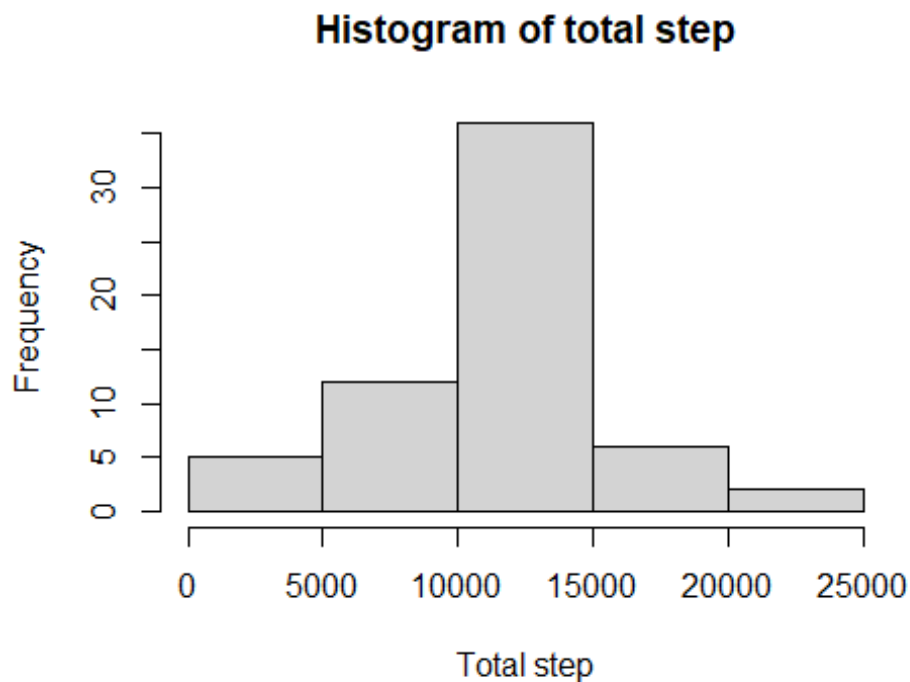
A new dataset that is equal to the original dataset but with the missing data filled in was created.

```
missingValue <- sum(is.na(data$steps))
data2 <- data
for(i in 1:ncol(data)) {
  data2[, i][is.na(data2[, i])] <- mean(data2[, i], na.rm = TRUE)
}

## Warning in mean.default(data2[, i], na.rm = TRUE): argument is not numeric
or
## logical: returning NA
```

A histogram of the total number of steps taken each day and the mean and median total number of steps taken per day calculated. Imputing missing data on the estimates of the total daily number of steps create a normal distribution according to the below histogram and make the mean and median closer to each other.

```
stepsperday2 <- tapply(data2$steps, data2$date, sum, na.rm = TRUE)
hist(stepsperday2, xlab = "Total step", main="Histogram of total step")
```



```
print(mean (stepsperday2))

## [1] 10766.19

print(median (stepsperday2))

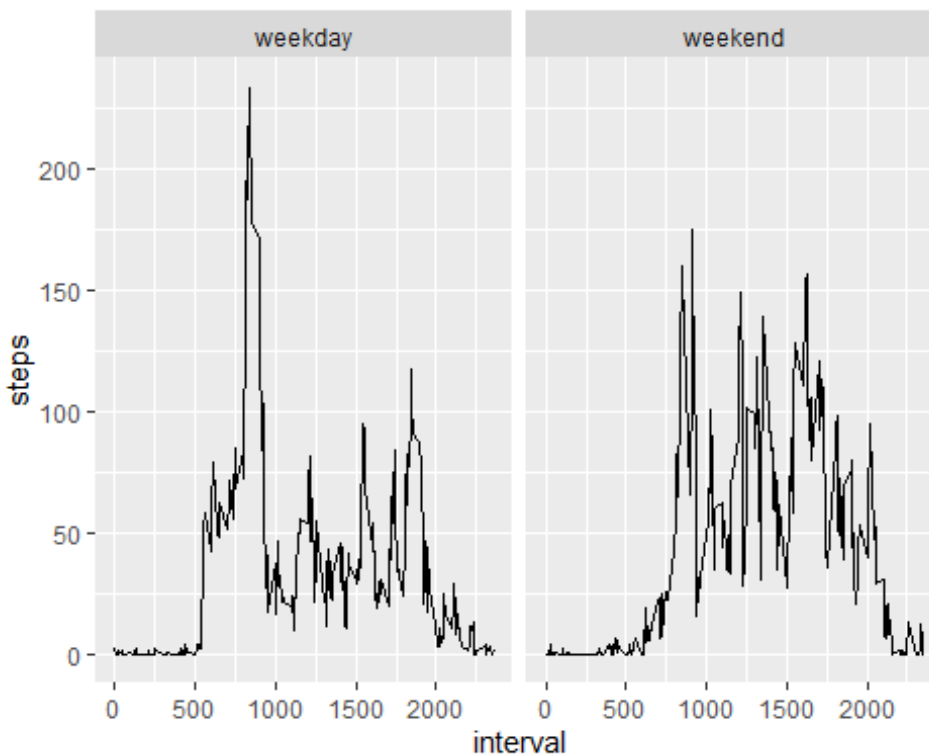
## [1] 10766.19

data$date <- as.Date(strptime(data$date, format="%Y-%m-%d"))
weekdays <- weekdays(data$date)
```

```

data2 <- cbind(data, weekdays)
for (i in 1:nrow(data2)) {
  if (data2$weekdays[[i]] %in% c("Saturday", "Sunday")) {
    data2$weekdays[[i]] <- "weekend"
  }
  else{
    data2$weekdays[[i]] <- "weekday"
  }
}
stepsByDay <- aggregate(data2$steps ~ data2$interval + data2$weekdays, data2,
mean)
names(stepsByDay) <- c("interval", "weekdays", "steps")
library(ggplot2)
par(mfrow=c(1,2))
ggplot(stepsByDay, aes(interval, steps))+
  facet_grid(~weekdays)+
  geom_line()

```



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