Reproducible_Research_Assingment_1

Paul M.

2022-10-31

Loading and preprocessing the data

```
data <- read.csv('~/activity.csv')
```

##What is mean total number of steps taken per day? #For this part of the assignment, you can ignore the missing values in the dataset.

```
daysteps <- tapply(data$steps, data$date, sum, na.rm=TRUE)
summary(daysteps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    6778   10395    9354   12811   21194
```

```
qplot(daysteps, xlab='Steps Per Day', ylab='Frequency', binwidth=500)
```

1. Make a histogram of the total number of steps taken each day
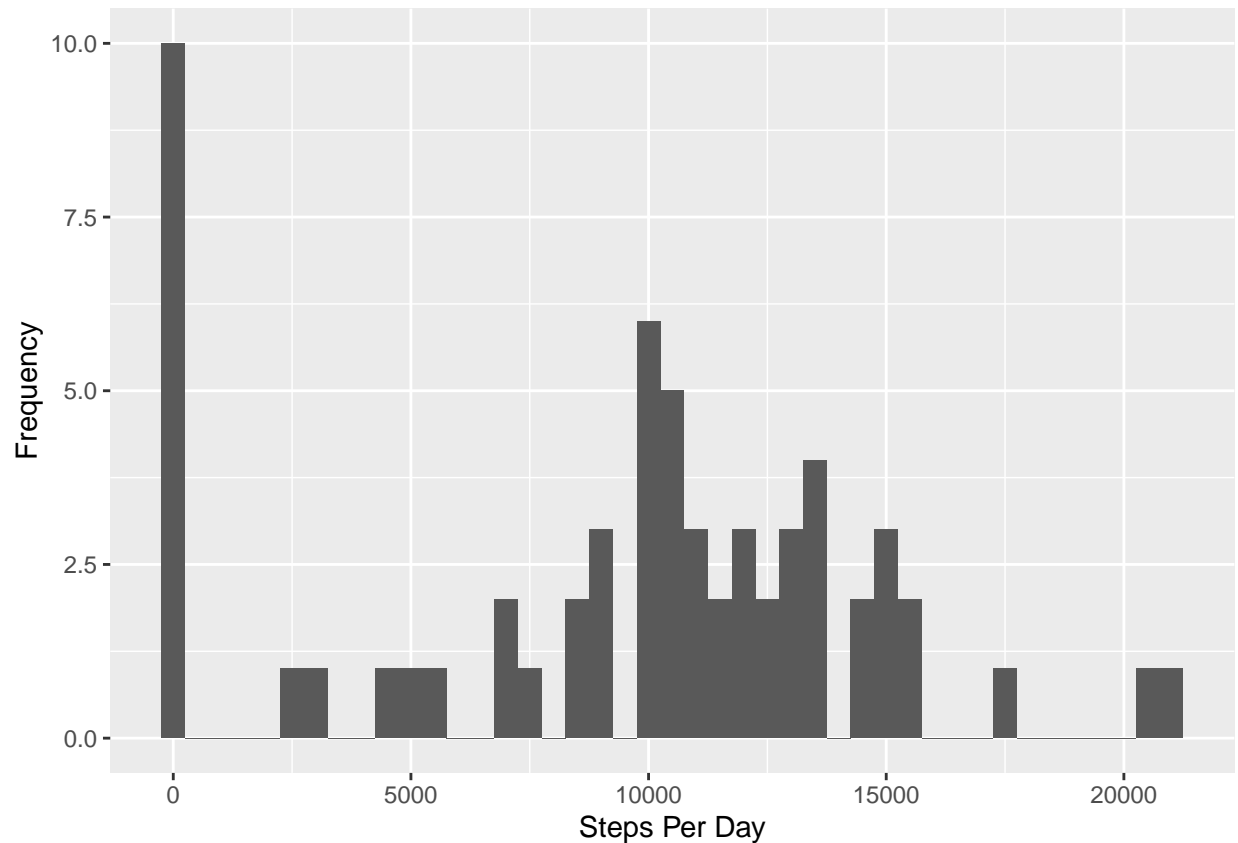
```
daystepsMean <- mean(daysteps)
daystepsMedian <- median(daysteps)
```

2. Calculate and report the mean and median total number of steps taken per day

- Mean: 9354.2295082
- Median: 10395

What is the average daily activity pattern?

```
averagep<- aggregate(x=list(meanSteps=data$steps), by=list(interval=data$interval), FUN=mean, na.rm=TRU
summary(averagep)
```

```
##     interval        meanSteps
## Min.   :   0.0   Min.   :  0.000
## 1st Qu.: 588.8   1st Qu.:  2.486
## Median :1177.5   Median : 34.113
## Mean   :1177.5   Mean   : 37.383
## 3rd Qu.:1766.2   3rd Qu.: 52.835
## Max.   :2355.0   Max.   :206.170
```

```
ggplot(data=averagep, aes(x=interval, y=meanSteps)) +
    geom_line() +
    xlab("5 minute interval") +
    ylab("average steps")
```

1. Make a time series plot
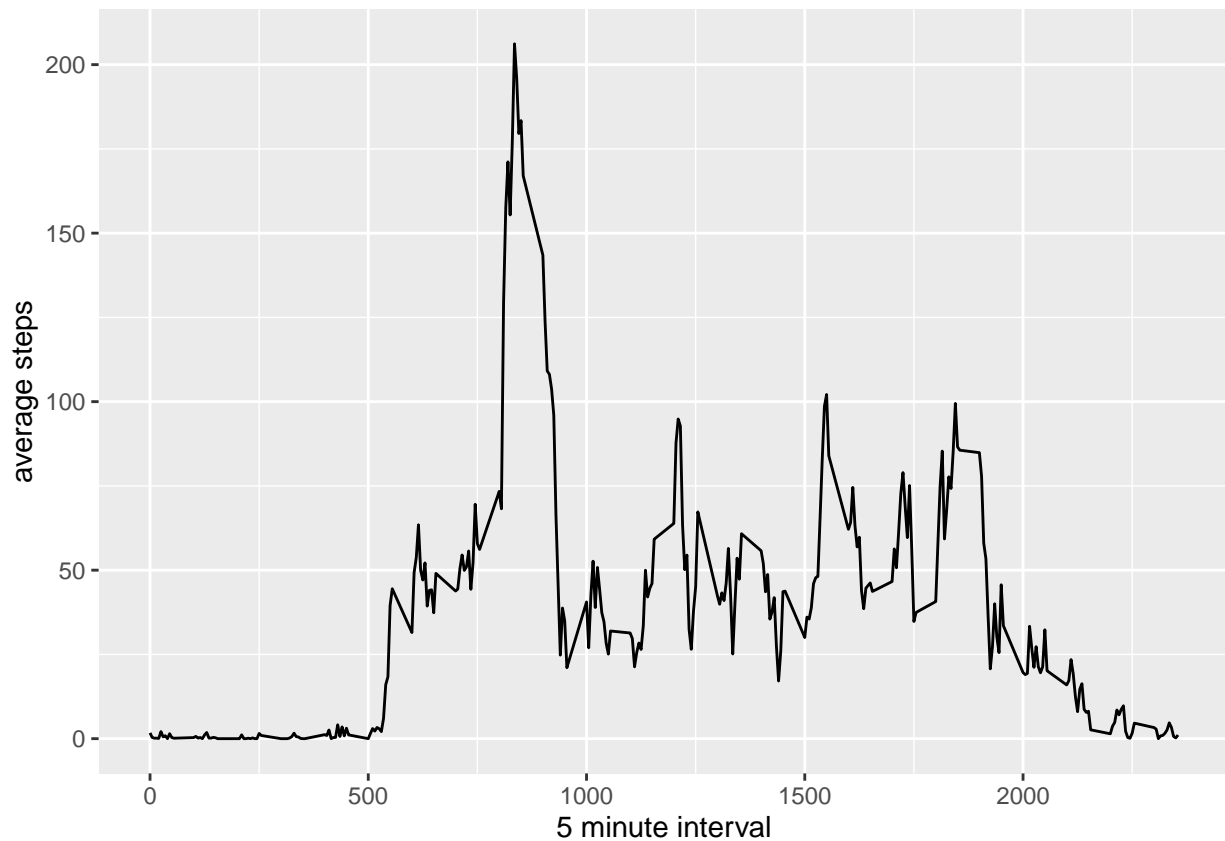

```
maxsteps <- which.max(averagep$meanSteps)
timemax <-  gsub("([0-9]{1,2})([0-9]{2})", "\\1:\\2", averagep[maxsteps,'interval'])
```

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

- Most Steps at: 8:35

_____

Imputing missing values

```
null <- length(which(is.na(data$steps)))
```

1. Calculate and report the total number of missing values in the dataset

- Number of missing values: 2304

2. Devise a strategy for filling in all of the missing values in the dataset.

```
DataImputed <- data
DataImputed$steps <- impute(data$steps, fun=mean)
```

3. Create a new dataset that is equal to the original dataset but with the missing data filled in.

```
stepsImputed <- tapply(DataImputed$steps, DataImputed$date, sum)
qplot(stepsImputed, xlab='steps per day (Imputed)', ylab='Frequency', binwidth=500)
```

4. Make a histogram of the total number of steps taken each day
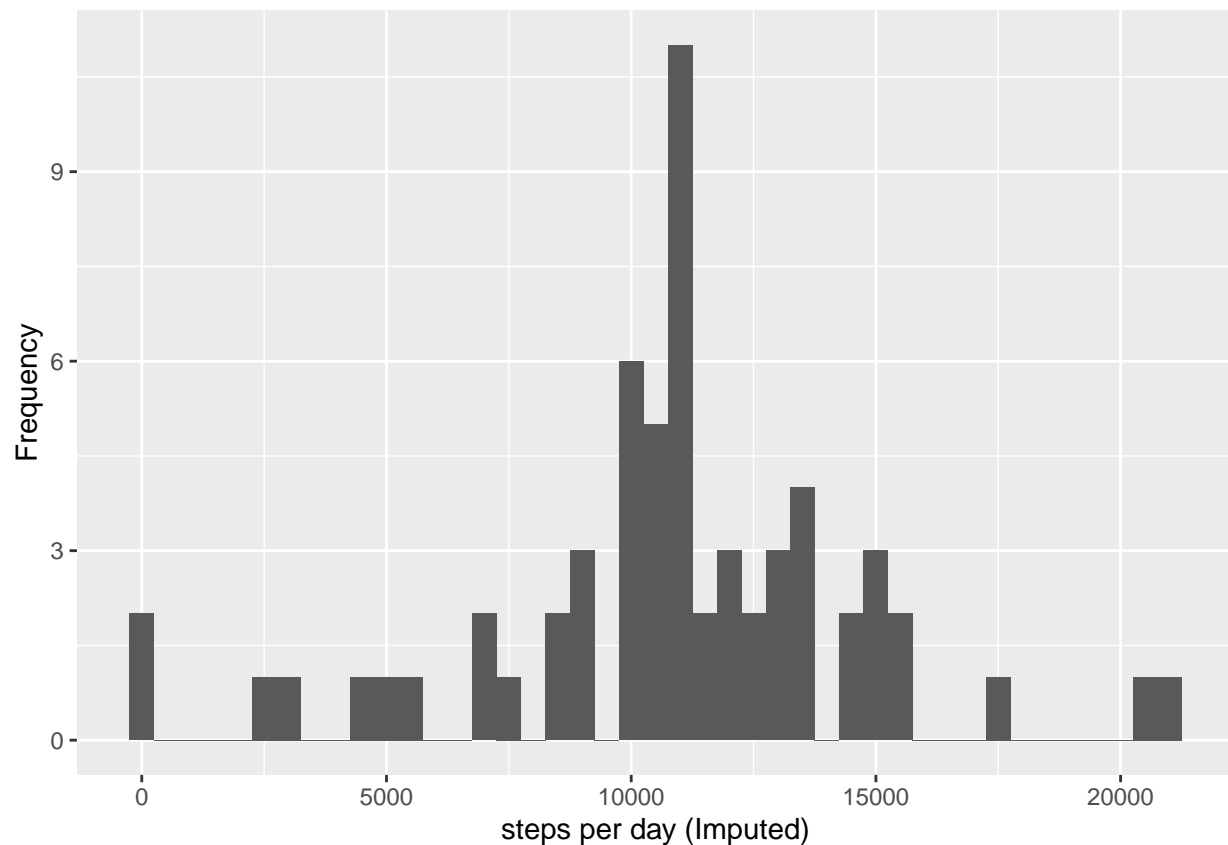
```
MeanImputed <- mean(stepsImputed)
MedianImputed <- median(stepsImputed)
```

Calculate and report the mean and median total number of steps taken per day.

- Mean (Imputed): 1.0766189×10^4
- Median (Imputed): 1.0766189×10^4

Are there differences in activity patterns between weekdays and weekends?

```
DataImputed$dateType <-  ifelse(as.POSIXlt(DataImputed$date)$wday %in% c(0,6), 'weekend', 'weekday')
```

1. Create a new factor variable in the dataset with two levels – "weekday" and "weekend" indicating whether a given date is a weekday or weekend day.

```
avgDataImputed <- aggregate(steps ~ interval + dateType, data=DataImputed, mean)
ggplot(avgDataImputed, aes(interval, steps)) +
    geom_line() +
    facet_grid(dateType ~ .) +
    xlab("5 minute interval") +
    ylab("avarage steps")
```

2. Make a panel plot containing a time series plot