Monitoring data activity

Carlos-Melgarejo

6/1/2021

Monitoring data assignment

1. Loading and preprocessing the data

```
unzip(zipfile = "activity.zip")
```

```
## Warning in unzip(zipfile = "activity.zip"): error 1 al extraer del archivo zip
```

```
data=read.csv("activity.csv")
data$date=as.Date(as.character(data$date))
summary(data)
```

```
##      steps             date               interval     
##  Min.   :  0.00   Min.   :2012-10-01   Min.   :   0.0  
##  1st Qu.:  0.00   1st Qu.:2012-10-16   1st Qu.: 588.8  
##  Median :  0.00   Median :2012-10-31   Median :1177.5  
##  Mean   : 37.38   Mean   :2012-10-31   Mean   :1177.5  
##  3rd Qu.: 12.00   3rd Qu.:2012-11-15   3rd Qu.:1766.2  
##  Max.   :806.00   Max.   :2012-11-30   Max.   :2355.0  
##  NA's   :2304
```

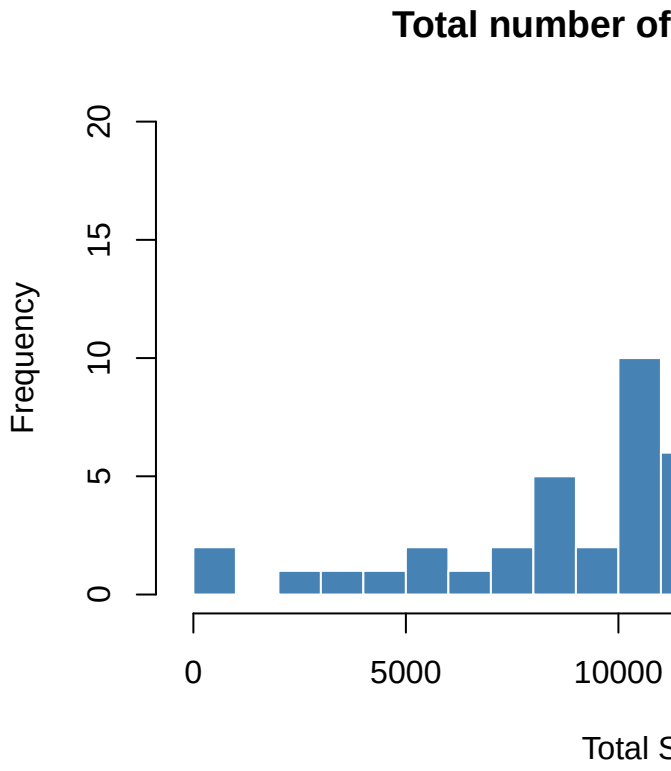
2. What is mean total number of steps taken per day?

```
byStep=aggregate(data$steps,by=list(data$date),sum)
names(byStep)=c("Date","TotalSteps")
head(byStep)
```

a. Total steps taken on each day

```
##         Date TotalSteps
## 1 2012-10-01         NA
## 2 2012-10-02        126
## 3 2012-10-03      11352
## 4 2012-10-04      12116
## 5 2012-10-05      13294
## 6 2012-10-06      15420
```

```
hist(byStep$TotalSteps,breaks = 25,col = "steelblue",border = "white",
     main = "Total number of Steps per Day", xlab = "Total Steps",ylim = c(0,20))
```



b. Histogram of the total number of steps taken each day

```
mean=round(mean(byStep$TotalSteps,na.rm=TRUE),2)
median=round(median(byStep$TotalSteps,na.rm = TRUE),2)
```

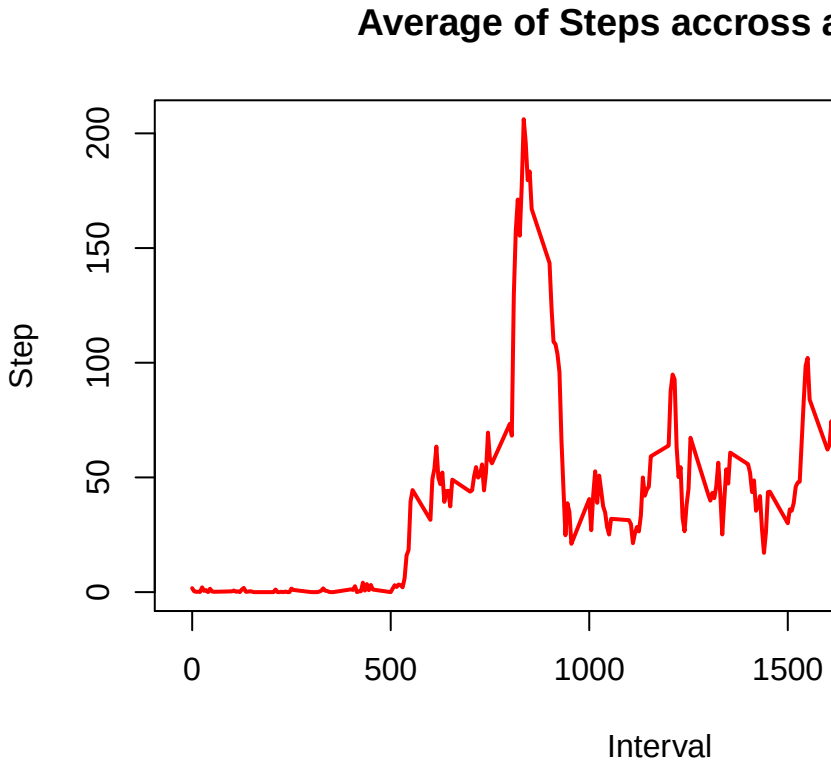
c. Mean and median of the total number of steps taken per day Mean of the total number of steps taken per day is: 1.076619×10^4 Median of the total number of steps taken per day is: 1.0765×10^4

3. What is the average daily activity pattern?

```
byInterval=aggregate(na.omit(data)$steps,by=list(na.omit(data)$interval),mean)
names(byInterval)=c("Interval","Mean.Steps")

plot(x=byInterval$Interval,y=byInterval$Mean.Steps,type = "l",col =10,lwd=2,main ="Average of Steps acc
```

a. Time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of



steps taken, averaged across all days (y-axis)

```
byInterval[byInterval$Mean.Steps==max(byInterval$Mean.Steps),]
```

b. The 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps

```
##     Interval Mean.Steps
## 104      835   206.1698
```

4. Imputing missing values

```
sapply(X=data,FUN = function(x) sum(is.na(x)))
```

a. Total number of missing values in the dataset (i.e. the total number of rows with NAs)

```
##    steps     date interval
##     2304        0        0
```

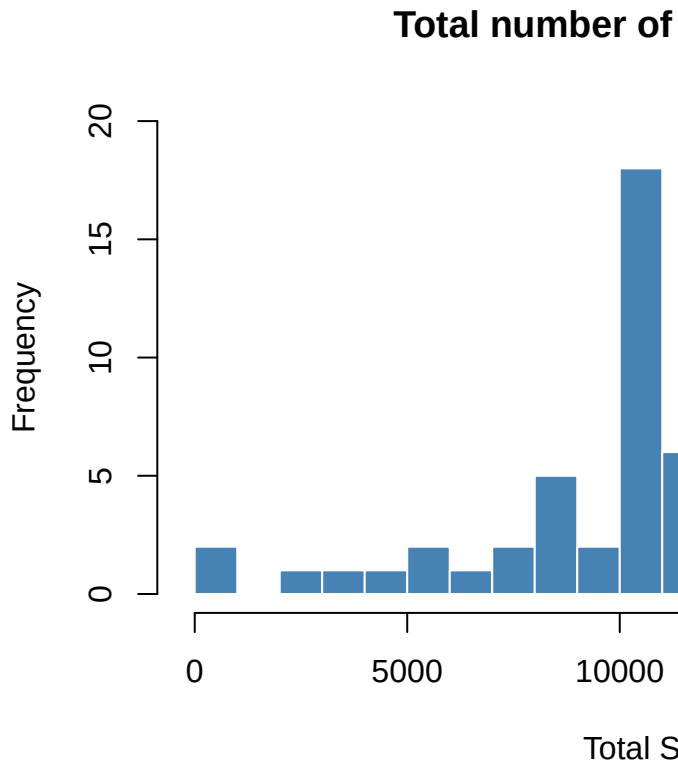
b. All the NAs will be replaced by the mean of steps taken on its interval

```
 complete.data=data
for (i in c(which(is.na(data$steps)))) {
  complete.data$steps[i]=byInterval[which(byInterval$Interval==data$interval[i]),2]
}
 sapply(X=complete.data,FUN = function(x) sum(is.na(x)))
```

c. New dataset that is equal to the original dataset but with the missing data filled in.

```
##    steps     date interval
##        0        0        0
```

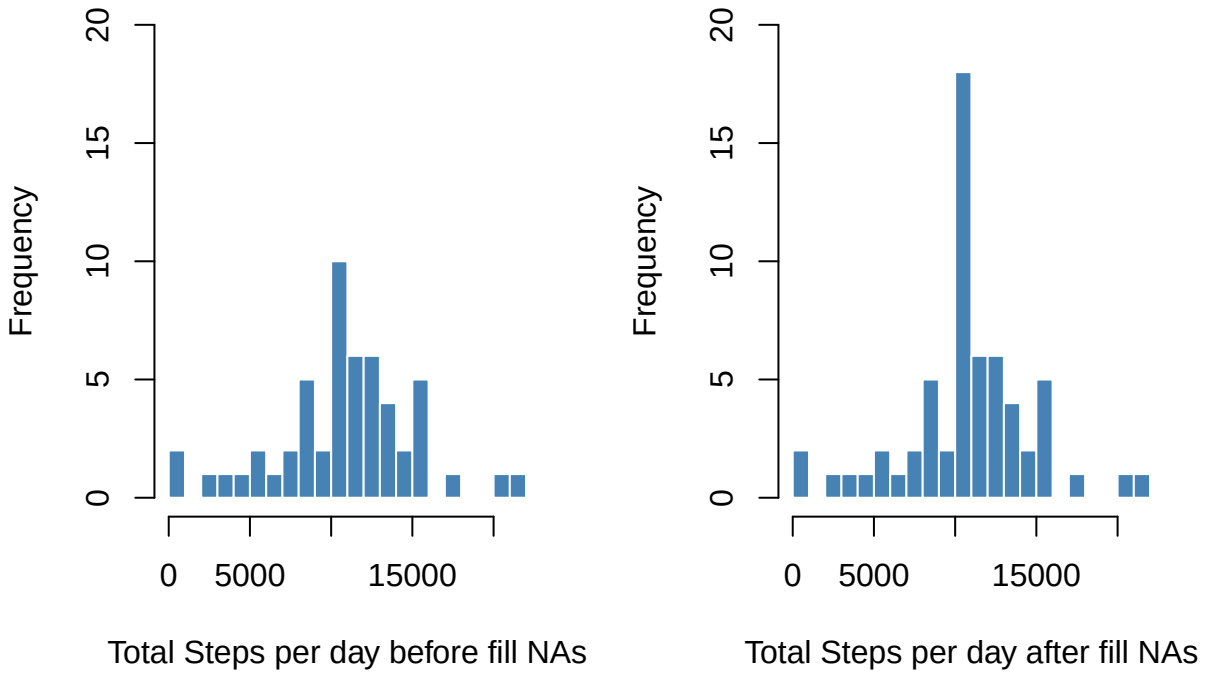
```
byStep.new=aggregate(complete.data$steps,by=list(complete.data$date),sum)
names(byStep.new)=c("Date","TotalSteps")
hist(byStep.new$TotalSteps,breaks = 25,col = "steelblue",border = "white",
     main = "Total number of Steps per Day", xlab = "Total Steps",
     ylim = c(0,20))
```



d. Histogram of the total number of steps taken each day

```r
#Compare between histogram
newmean=round(mean(byStep.new$TotalSteps),2)
newmedian=round(median(byStep.new$TotalSteps),2)
par(mfrow=c(1,2))
hist(byStep$TotalSteps,breaks = 25,col = "steelblue",border = "white",
     main=NULL,xlab = "Total Steps per day before fill NAs",ylim = c(0,20))
hist(byStep.new$TotalSteps,breaks = 25,col = "steelblue",border = "white",
     main=NULL,xlab = "Total Steps per day after fill NAs",
     ylim = c(0,20))
```

e. Any differences in mean and median between original data and data with NAs replaced?



New Mean of the total number of steps taken per day is: 1.076619×10^4 New median of the total number of steps taken per day is: 1.076619×10^4 The mean between those two data is the same which is 1.076619×10^4. However, the median rise from 1.0765×10^4 to 1.076619×10^4.

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

```r
new.data=complete.data
new.data$day.of.week=ifelse(weekdays(as.Date(new.data$date)) %in% c("Saturday","Sunday"),
                            "weekend","weekday")
new.data=aggregate(new.data$steps,by=list(new.data$interval,new.data$day.of.week),mean)
```

```
names(new.data)=c("Interval","day.of.week","Mean.Steps")
head(new.data)
```

a. New factor variable in the dataset with two levels - "weekday" and "weekend" indicating whether a given date is a weekday or weekend day

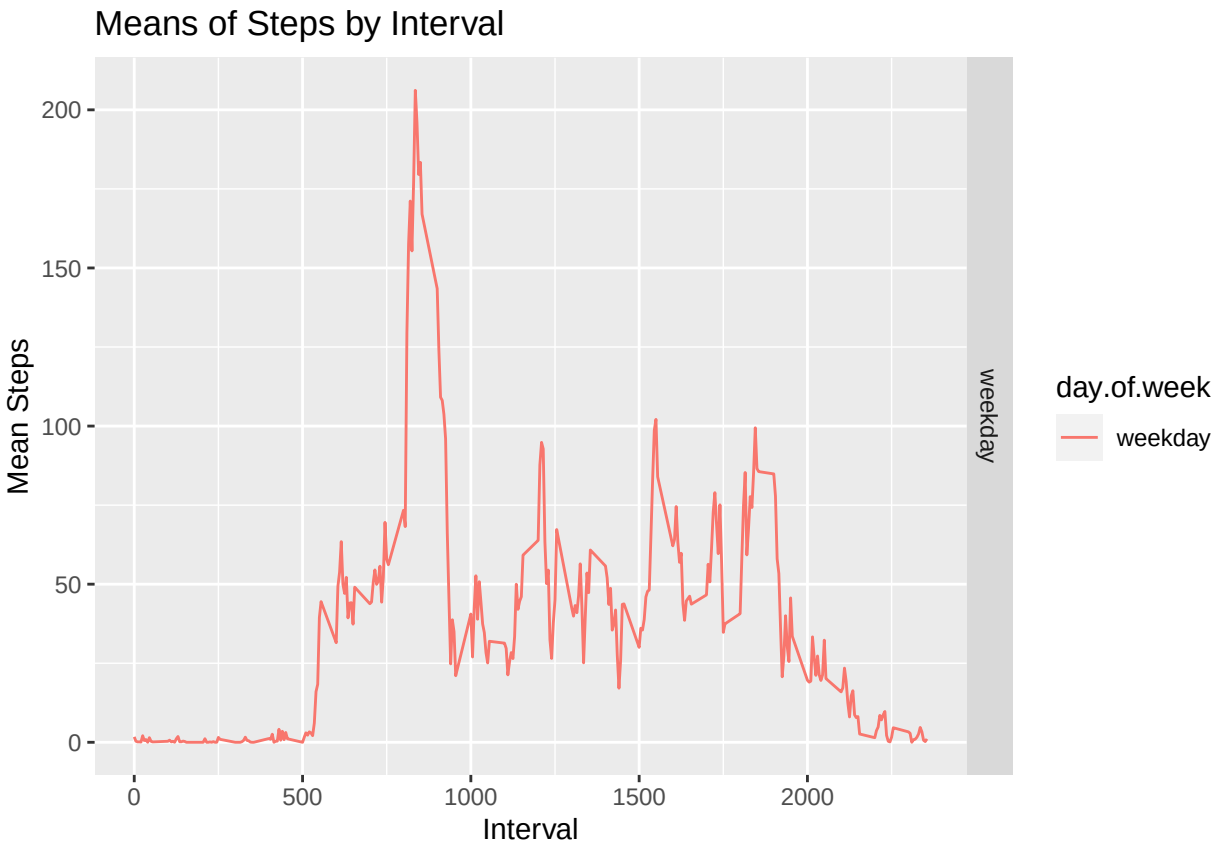
```
##   Interval day.of.week Mean.Steps
## 1        0     weekday  1.7169811
## 2        5     weekday  0.3396226
## 3       10     weekday  0.1320755
## 4       15     weekday  0.1509434
## 5       20     weekday  0.0754717
## 6       25     weekday  2.0943396
```

```
library(ggplot2)
```

b. Panel plot containing a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis)

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
ggplot(new.data,aes(x=Interval,y=Mean.Steps,color=day.of.week))+
  geom_line()+
  facet_grid(day.of.week~.)+
  labs(title="Means of Steps by Interval", x="Interval",y="Mean Steps")
```


Means of Steps by Interval
Mean Steps
200
150
100
50
0
0
500
1000
1500
2000
Interval
weekday
day.of.week
weekday