

Reproducible Research: Peer Assessment 1

pitambar Gautam: template after RDPeng

2020/10/9

Loading and preprocessing the data

```
activity <- read.table (file ="activity.csv",header = TRUE, sep =",")  
head(activity)
```

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

```
tail(activity)
```

```
##      steps      date interval  
## 17563    NA 2012-11-30     2330  
## 17564    NA 2012-11-30     2335  
## 17565    NA 2012-11-30     2340  
## 17566    NA 2012-11-30     2345  
## 17567    NA 2012-11-30     2350  
## 17568    NA 2012-11-30     2355
```

```
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
```

Histogram of Daily Steps data

```
## install.packages("ggplot2")  
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```

daily_steps <- aggregate(steps ~ date, activity, sum)
head(daily_steps)

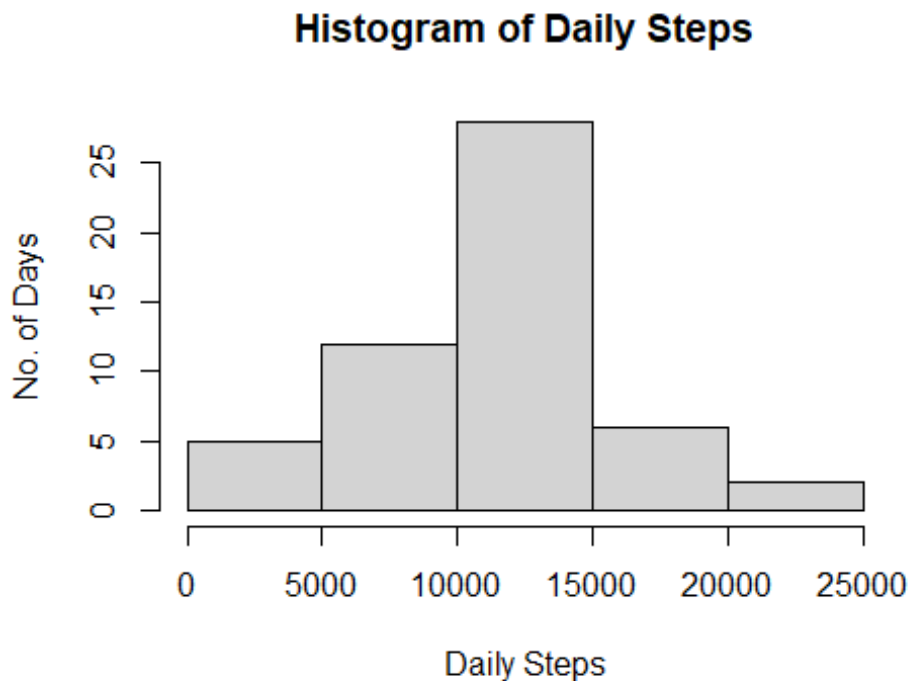
##           date steps
## 1 2012-10-02   126
## 2 2012-10-03 11352
## 3 2012-10-04 12116
## 4 2012-10-05 13294
## 5 2012-10-06 15420
## 6 2012-10-07 11015

str(daily_steps)

## 'data.frame':   53 obs. of  2 variables:
##  $ date : chr  "2012-10-02" "2012-10-03" "2012-10-04" "2012-10-05" ...
##  $ steps: int  126 11352 12116 13294 15420 11015 12811 9900 10304 1738
## 2 ...

hist_ds<- hist(daily_steps$steps, main = "Histogram of Daily Steps", xlab
= "Daily Steps", ylab = "No. of Days")

```



```

## Summary of Daily Steps data
summary(daily_steps$steps)

```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	41	8841	10765	10766	13294	21194

Average daily activity pattern:

Mean and Median number of steps taken per day

```
mean_daily_steps <- mean(daily_steps$steps)
mean_daily_steps

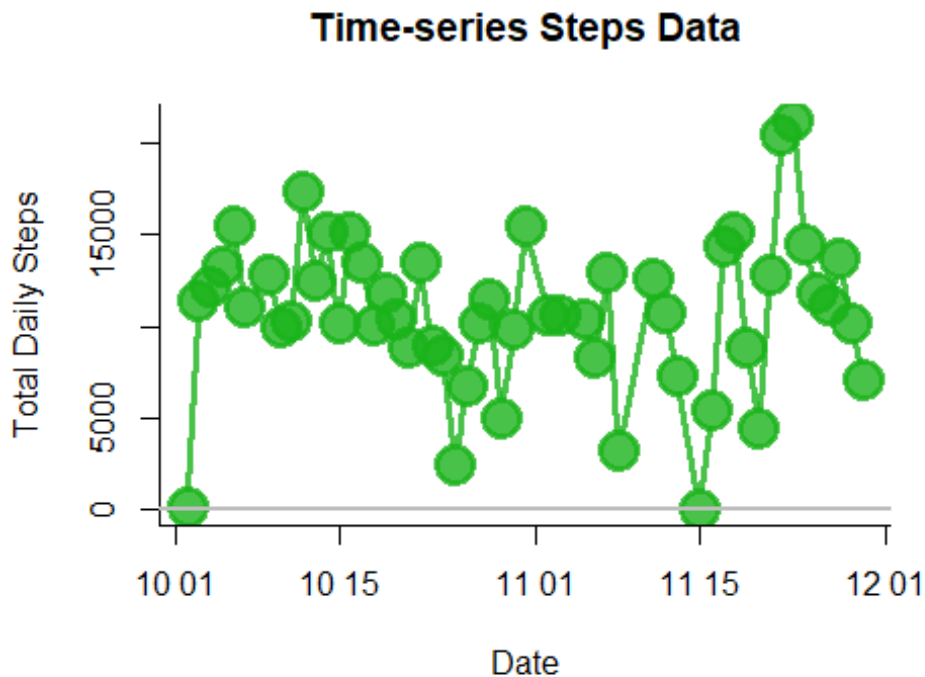
## [1] 10766.19

median_daily_steps <- median(daily_steps$steps)
median_daily_steps

## [1] 10765
```

Time-series of total steps taken per day

```
daily_steps$date <- as.Date(daily_steps$date)
daily_steps <- daily_steps[order(daily_steps$date) , ]
plot(daily_steps$steps~daily_steps$date, type="b", lwd=3, col=rgb(0.1,0.7,
0.1,0.8), main = "Time-series Steps Data",ylab="Total Daily Steps", xlab=
"Date" , bty="l" , pch=20 , cex=4) + abline(h=seq(0,100,10), col="grey",
lwd=0.8)
```



```
## integer(0)
```

5-minute time interval total steps Time-series data

```
interval_steps <- aggregate(steps ~ interval, activity, mean)
interval_steps <- interval_steps[order(interval_steps$steps, decreasing =
  TRUE) , ]
head(interval_steps)
```

```
##      interval      steps
## 104         835 206.1698
## 105         840 195.9245
## 107         850 183.3962
## 106         845 179.5660
## 103         830 177.3019
## 101         820 171.1509
```

```
interval_maxsteps <- interval_steps$interval[1]
interval_maxsteps
```

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

```
activity["interval_meansteps"]<- NA
activity$interval_meansteps = interval_steps$steps
# activity <- activity[order(activity$interval_meansteps, decreasing = TR
  UE) , ]
head(activity)
```

```
##  steps      date interval interval_meansteps
## 1    NA 2012-10-01      0          206.1698
## 2    NA 2012-10-01      5          195.9245
## 3    NA 2012-10-01     10          183.3962
## 4    NA 2012-10-01     15          179.5660
## 5    NA 2012-10-01     20          177.3019
## 6    NA 2012-10-01     25          171.1509
```

Imputing missing values (replace by interval mean)

```
# No. of missing values from summary above = 2304
activity["filled_steps"]<- NA # an empty (na) column added
activity$filled_steps = activity$steps # content replaced by steps column data
# code to replace by overall mean steps:
# activity$filled_steps = ifelse(is.na(activity$filled_steps), ave(activity$steps,
  (+contd..))
# FUN = function (x) mean (x, na.rm =TRUE)), activity$filled_steps)
activity$filled_steps = ifelse(is.na(activity$filled_steps), activity$interval_meansteps, activity$filled_steps)
head(activity)
```

	steps	date	interval	interval_meansteps	filled_steps
## 1	NA	2012-10-01	0	206.1698	206.1698
## 2	NA	2012-10-01	5	195.9245	195.9245
## 3	NA	2012-10-01	10	183.3962	183.3962
## 4	NA	2012-10-01	15	179.5660	179.5660
## 5	NA	2012-10-01	20	177.3019	177.3019
## 6	NA	2012-10-01	25	171.1509	171.1509

Average daily activity pattern (after imputing data):

```
## Histogram od Daily Steps (revised) data
## install.packages("ggplot2")
require(ggplot2)
daily_steps_filled <- aggregate(filled_steps ~ date, activity, sum)
head(daily_steps_filled)

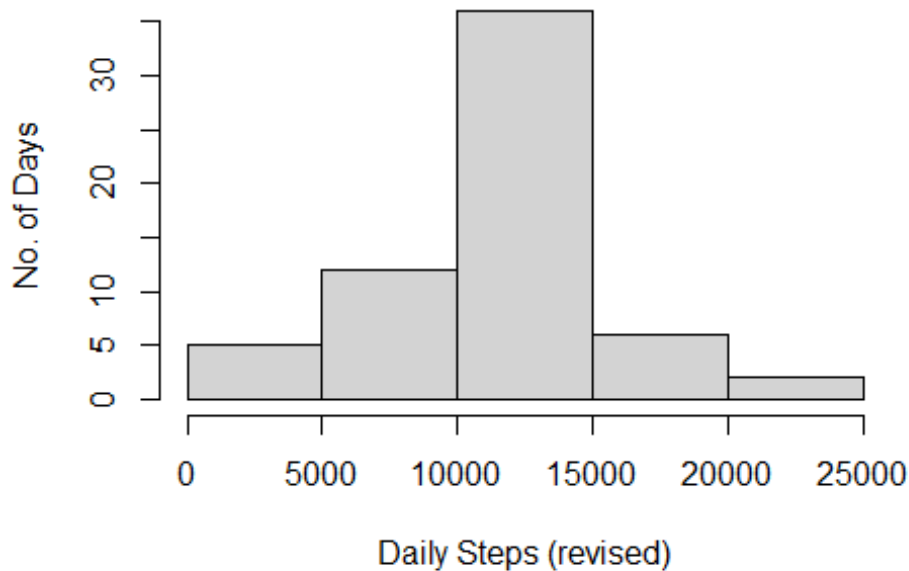
##           date filled_steps
## 1 2012-10-01      10766.19
## 2 2012-10-02         126.00
## 3 2012-10-03      11352.00
## 4 2012-10-04      12116.00
## 5 2012-10-05      13294.00
## 6 2012-10-06      15420.00

str(daily_steps_filled)

## 'data.frame':   61 obs. of  2 variables:
##  $ date          : chr  "2012-10-01" "2012-10-02" "2012-10-03" "2012-10-04" ...
##  $ filled_steps: num  10766 126 11352 12116 13294 ...

hist_dsfilled<- hist(daily_steps_filled$filled_steps, main = "Histogram of Daily Steps (revised)", xlab = "Daily Steps (revised)", ylab = "No. of Days")
```

Histogram of Daily Steps (revised)



Summary

of Daily Steps Data:

```
summary(daily_steps_filled$filled_steps)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       41   9819   10766   10766   12811   21194
```

Mean and Median number of steps taken per day:

```
mean_daily_steps_filled <- mean(daily_steps_filled$filled_steps)
mean_daily_steps_filled
```

```
## [1] 10766.19
```

```
median_daily_steps_filled <- median(daily_steps_filled$filled_steps)
median_daily_steps_filled
```

```
## [1] 10766.19
```

Differences in activity patterns between weekdays and weekends:

```
activity$date <- as.Date(activity$date) #conversion to date format for con
version with weekdays to week's day
activity$day_class <- ifelse(as.POSIXlt(activity$date)$wday %in% c(0,6),
  'weekend', 'weekday')
head(activity)
```

```
##   steps      date interval interval_meansteps filled_steps day_class
## 1    NA 2012-10-01         0          206.1698      206.1698  weekday
## 2    NA 2012-10-01         5          195.9245      195.9245  weekday
## 3    NA 2012-10-01        10          183.3962      183.3962  weekday
## 4    NA 2012-10-01        15          179.5660      179.5660  weekday
## 5    NA 2012-10-01        20          177.3019      177.3019  weekday
## 6    NA 2012-10-01        25          171.1509      171.1509  weekday
```

```
interval_steps_filled <- aggregate(filled_steps ~ interval + day_class, d
ata = activity, mean)
head(interval_steps_filled)
```

```
##   interval day_class filled_steps
## 1         0  weekday      29.51153
## 2         5  weekday      26.52327
## 3        10  weekday      24.60839
## 4        15  weekday      24.11992
## 5        20  weekday      23.72914
## 6        25  weekday      24.13124
```

```
ggplot(interval_steps_filled, aes(interval, filled_steps)) +
  geom_line() +
  facet_grid(day_class ~ .) + xlab("interval (each 5 minutes)") + ylab
("Average Number of Steps")
```

