Reproducible research project 1

Narendra Sahu

8 August 2016

# Loading and preprocessing the data

## checking the structure of variables and converting them after

activity<-read.csv("activity.csv")  
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

str(activity)

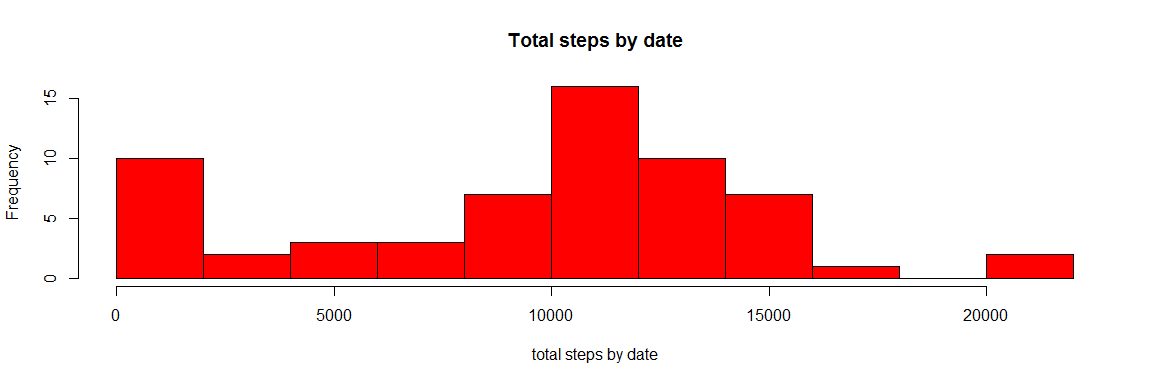
## 'data.frame': 17568 obs. of 3 variables:  
## $ steps : int NA NA NA NA NA NA NA NA NA NA ...  
## $ date : Factor w/ 61 levels "2012-10-01","2012-10-02",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ interval: int 0 5 10 15 20 25 30 35 40 45 ...

activity<-read.csv("activity.csv",colClasses = c("numeric","Date","numeric"))

# mean total number of steps taken per day

## histogram plot

totalsteps\_by\_date<-tapply(activity$steps,activity$date,sum,na.rm=T)  
hist(totalsteps\_by\_date,main = "Total steps by date",xlab="total steps by date",breaks = 10,col="red")

 ##mean and median

mean(totalsteps\_by\_date)

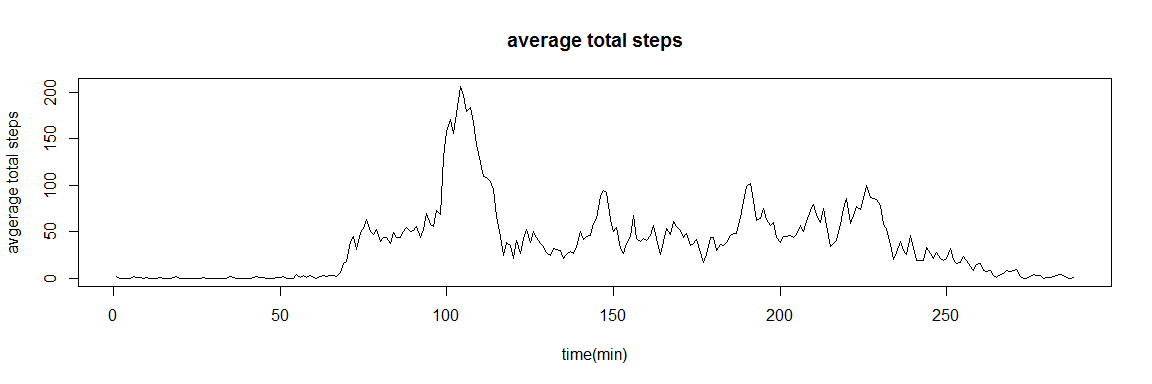
## [1] 9354.23

median(totalsteps\_by\_date)

## [1] 10395

# average daily activity pattern

interval\_mean<-tapply(activity$steps,activity$interval,mean,na.rm=T)  
plot(interval\_mean,type="l",xlab="time(min)",ylab="avgerage total steps",main="average total steps")



max\_steps<-max(interval\_mean)  
interval\_mean[max\_steps==interval\_mean]

## 835   
## 206.1698

# Imputing missing values

## number of missing value rows

sum(is.na(activity))

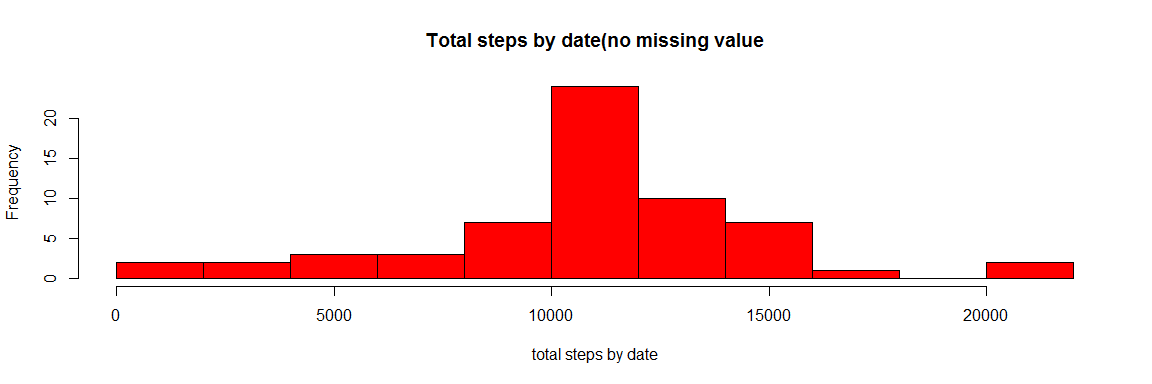
## [1] 2304

activity\_nomiss<- transform(activity, steps = ifelse(is.na(steps), round(mean(steps, na.rm=TRUE)), steps))  
sum(is.na(activity\_nomiss))

## [1] 0

## replacing missing value by means and plotting histogram

new\_steps\_by\_date<-tapply(activity\_nomiss$steps,activity\_nomiss$date,sum)  
hist(new\_steps\_by\_date,main = "Total steps by date(no missing value",xlab="total steps by date",breaks = 10,col="red")



## mean and median of total sets(no missing values)

mean(new\_steps\_by\_date)

## [1] 10751.74

median(new\_steps\_by\_date)

## [1] 10656

# differences in activity patterns between weekdays and weekends

## assigning weekdays and weekend level

activity\_nomiss$days <- weekdays(activity\_nomiss$date)  
library(plyr)  
activity\_nomiss$days <- revalue(activity\_nomiss$days,c("Monday"="weekday","Tuesday"="weekday","Wednesday"="weekday","Thursday"="weekday","Friday"="weekday"))  
activity\_nomiss$days <- revalue(activity\_nomiss$days,c("Saturday"="weekend","Sunday"="weekend"))

## panel plot containing a time series plot of the 5-minute intervaland the average number of steps taken, averaged across all weekday days or weekend days.

library(lattice)  
activity\_mean <- aggregate(steps ~ days+interval, data=activity\_nomiss , FUN=mean)  
xyplot(steps ~ interval | factor(days),data=activity\_mean,type="l")

