PA1\_template

Steven Sheridan

April 4, 2019

# Loading and Processing the data

Read data

activity<-read.csv(file='activity.csv')

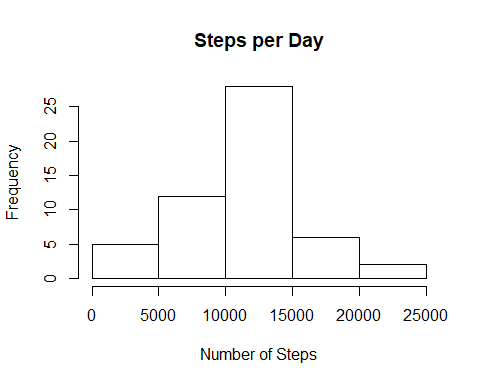
# What is the mean total number of steps taken per day?

Calculate steps per day

stepsperday<-aggregate(steps~date,activity,sum)

Create steps per day histogram

hist(stepsperday$steps,xlab='Number of Steps',ylab='Frequency',main='Steps per Day')



Mean and median of steps per day

summary(stepsperday$steps)

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

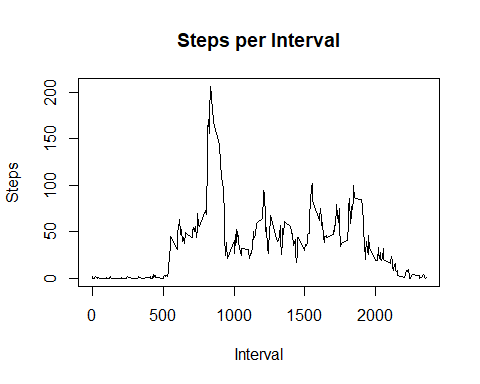
# What is the average daily activity pattern?

Calculate steps per interval

stepsperinterval<-aggregate(steps~interval,activity,mean)

Plot steps per interval

plot(stepsperinterval,type='l',xlab="Interval", ylab="Steps",main="Steps per Interval")



Most steps in a single interval

max(stepsperinterval$steps)

## [1] 206.1698

Interval with the most steps

stepsperinterval[stepsperinterval$steps==max(stepsperinterval$steps),1]

## [1] 835

# Imputing missing values

Number of missing values

nrow(activity[is.na(activity$steps),])

## [1] 2304

Impute interval mean steps

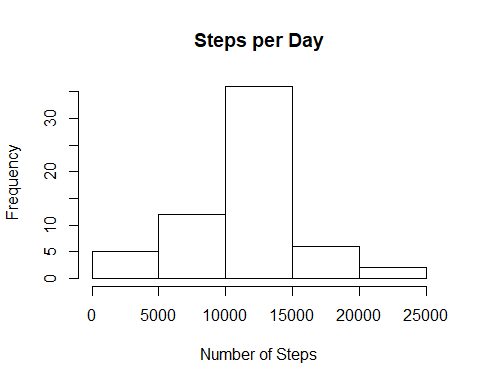
impute<-activity  
nas<-is.na(impute$steps)  
impute$steps[nas]<-stepsperinterval$steps

Calculate new steps per day

newstepsperday<-aggregate(steps~date,impute,sum)

Create new imputed steps per day histogram

hist(newstepsperday$steps,xlab='Number of Steps',ylab='Frequency',main='Steps per Day')



New imputed mean and median of steps per day

summary(newstepsperday$steps)

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

The imputed mean and median are nearly identical to the originals.

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

Create weekend/weekday categories

impute$day <- ifelse(weekdays(as.Date(impute$date)) == "Saturday" |   
 weekdays(as.Date(impute$date)) == "Sunday", "Weekend", "Weekday")

Calculate new steps per interval by weekend/weekday

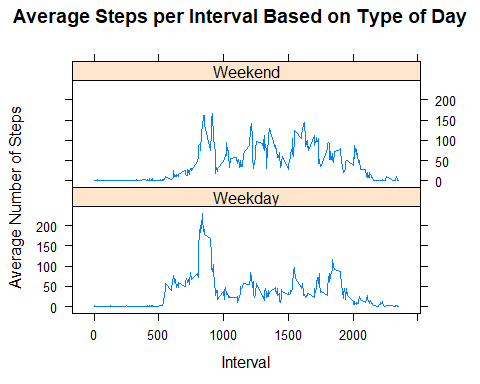
library(plyr)

## Warning: package 'plyr' was built under R version 3.5.3

newstepsperinterval<-ddply(impute,.(interval,day),summarize,avg=mean(steps))

Plot steps per interval by weekend/weekday

library(lattice)  
xyplot(avg~interval|day, data=newstepsperinterval, type="l", layout = c(1,2),  
 main="Average Steps per Interval Based on Type of Day",   
 ylab="Average Number of Steps", xlab="Interval")



Yes, it looks like people tend to start walking earlier during weekdays and there is a sharp spike at around the 800th interval. On weekends the average number of steps per interval seems to be higher during the day and afternoon.