Peer Assignment - R Markdown

Abhishek Kaul

Tuesday, March 10, 2015

# Part 1 - Loading and preprocessing the data

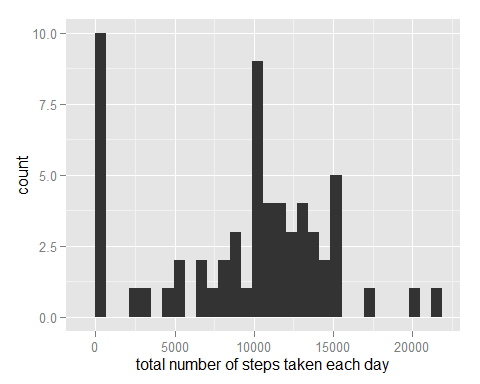
##unzip file  
unzip(zipfile="activity.zip")  
  
## Read the activity.csv file  
data <- read.csv("activity.csv")  
  
## print top rows  
head(data)

## 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

# Part 2 : What is mean total number of steps taken per day?

library(ggplot2)  
total.steps <- tapply(data$steps, data$date, FUN=sum, na.rm=TRUE)  
qplot(total.steps, xlab="total number of steps taken each day")

## stat\_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.



mean(total.steps, na.rm=TRUE)

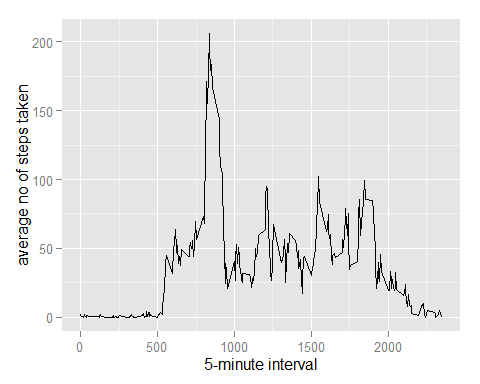
## [1] 9354.23

median(total.steps, na.rm=TRUE)

## [1] 10395

# Part 3 : What is the average daily activity pattern?

library(ggplot2)  
avg <- aggregate(x=list(steps=data$steps), by=list(interval=data$interval),FUN=mean, na.rm=TRUE)  
ggplot(data=avg, aes(x=interval, y=steps)) + geom\_line() + xlab("5-minute interval") + ylab("average no of steps taken")



## On average across all the days in the dataset, the 5-minute interval contains the maximum number of steps?

avg[which.max(avg$steps),]

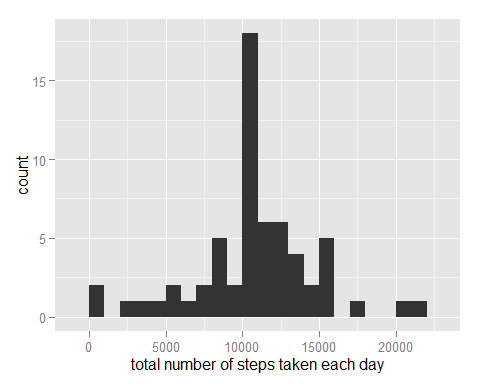
## interval steps  
## 104 835 206.1698

# Part 4: Imputing missing values

missing <- is.na(data$steps)  
# How many missing  
table(missing)

## missing  
## FALSE TRUE   
## 15264 2304

# Replace each missing value with the mean value of its 5-minute interval  
fill.value <- function(steps, interval) {  
 filled <- NA  
 if (!is.na(steps))  
 filled <- c(steps)  
 else  
 filled <- (avg[avg$interval==interval, "steps"])  
 return(filled)  
}  
filled.data <- data  
filled.data$steps <- mapply(fill.value, filled.data$steps, filled.data$interval)  
  
total.steps <- tapply(filled.data$steps, filled.data$date, FUN=sum)  
qplot(total.steps, binwidth=1000, xlab="total number of steps taken each day")



mean(total.steps)

## [1] 10766.19

median(total.steps)

## [1] 10766.19

# Part 5 : Are there differences in activity patterns between weekdays and weekends?

weekday.or.weekend <- function(date) {  
 day <- weekdays(date)  
 if (day %in% c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"))  
 return("weekday")  
 else if (day %in% c("Saturday", "Sunday"))  
 return("weekend")  
 else  
 stop("invalid date")  
}  
filled.data$date <- as.Date(filled.data$date)  
filled.data$day <- sapply(filled.data$date, FUN=weekday.or.weekend)  
  
averages <- aggregate(steps ~ interval + day, data=filled.data, mean)  
ggplot(averages, aes(interval, steps)) + geom\_line() + facet\_grid(day ~ .) +  
 xlab("5-minute interval") + ylab("Number of steps")

