Reproducible Research: Peer Assessment 1

Jennyfer Combariza

December 5, 2016

## Loading and preprocessing the data

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

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

library(plyr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:plyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
  
total.steps <- tapply(data$steps, data$date, FUN = sum, na.rm = TRUE)  
  
#Mean  
mean(total.steps)

## [1] 9354.23

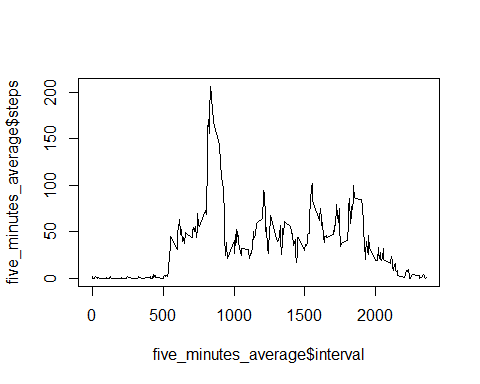
#Median  
median(total.steps)

## [1] 10395

## What is the average daily activity pattern?

We make a time series plot of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis) and we plot the result.

five\_minutes\_average <- aggregate(steps~interval, data=data, FUN=mean, na.rm=TRUE)  
plot(x = five\_minutes\_average$interval, y = five\_minutes\_average$steps, type = "l")



png("average.png", width=750)  
plot(x = five\_minutes\_average$interval, y = five\_minutes\_average$steps, type = "l")   
dev.off()

## png   
## 2

max\_steps <- max(five\_minutes\_average$steps)  
for (i in 1:288)   
{  
 if (five\_minutes\_average$steps[i] == max\_steps)  
 five\_minute\_interval\_at\_max\_steps <- five\_minutes\_average$interval[i]  
}  
five\_minute\_interval\_at\_max\_steps

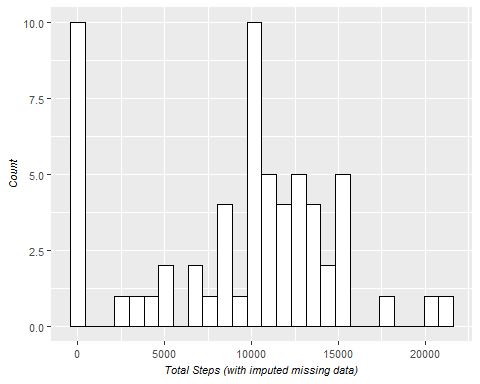
## [1] 835

## Imputing missing values

sum(!complete.cases(data))

## [1] 2304

fillfunc <- function(step, interval) {  
 ifelse(is.na(step), averages[averages$interval == interval, ]$steps, step)  
}  
  
data\_fill <- data  
tot\_steps\_day <- aggregate(steps ~ date, data=data\_fill, FUN=sum,  
 na.rm=TRUE, na.action=NULL)  
  
granularity = diff(range(tot\_steps\_day$steps)) / 25  
ggplot(data=tot\_steps\_day, aes(x=tot\_steps\_day$steps)) +  
 geom\_histogram(binwidth=granularity, col="black", fill="white") +  
 labs(x="Total Steps (with imputed missing data)", y="Count") +  
 theme(axis.text=element\_text(size=8),  
 axis.title=element\_text(size=8, face="italic"))



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

We start by creating a new factor variable in the dataset with two levels -- weekday and weekend indicating whether a given date is a weekday or weekend day.

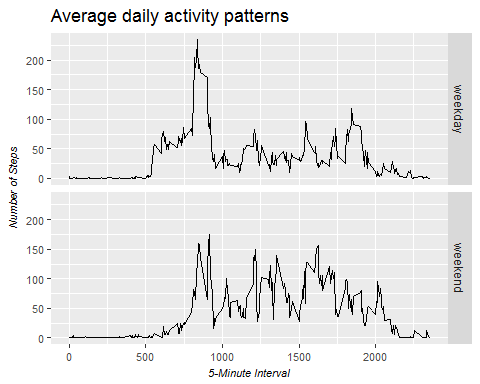
switch(Sys.info()[[ 'sysname' ]],  
 Windows = { lctime <- "English" }, { lctime <- "C" })  
Sys.setlocale("LC\_TIME", lctime)

## [1] "English\_United States.1252"

data\_fill$date <- as.Date(data\_fill$date, "%Y-%m-%d")  
daytype <- function(day) {  
 ifelse(weekdays(day, abbreviate=FALSE) %in% c("Saturday", "Sunday"),  
 "weekend", "weekday")  
}  
data\_fill$daytype <- mapply(daytype, data\_fill$date)

And we end this document by plotting the two resulting datasets:

averages <- aggregate(steps ~ interval + daytype, data=data\_fill, FUN=mean)  
ggplot(data=averages, aes(x=interval, y=steps)) +  
 geom\_line() +  
 facet\_grid(daytype ~ .) +  
 ggtitle("Average daily activity patterns") +  
 xlab("5-Minute Interval") +  
 ylab("Number of Steps") +  
 theme(axis.text=element\_text(size=8),  
 axis.title=element\_text(size=8, face="italic"))



Answer

Yes, there are differences in activity patterns between weekdays and weekends.