Reproducible Research: Peer Assessment 1

# Load required packages

#setwd("C:/Users/Hadrien/Dropbox/Coursera Data Science/reproducibleResearch/assignment1/RepData\_PeerAssessment1-master")  
getwd()

## [1] "C:/Users/Hadrien/Dropbox/Coursera Data Science/reproducibleResearch/assignment1/RepData\_PeerAssessment1-master"

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.4

library(plyr)  
library(mice) #used to impute missing values

## Warning: package 'mice' was built under R version 3.2.4

## Loading required package: Rcpp  
## mice 2.25 2015-11-09

library(chron)  
library(cowplot)#needed to graph multiple plots

## Warning: package 'cowplot' was built under R version 3.2.4

##   
## Attaching package: 'cowplot'  
##   
## The following object is masked from 'package:ggplot2':  
##   
## ggsave

## Loading and preprocessing the data

df <- read.csv('Data/activity.csv') #Load Data  
df2 <- df[!is.na(df$steps), ] #remove NAs  
df2\_summarized <- ddply(df2, c("date"), summarize, sumSteps = sum(steps)) #summarize

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

#plot results  
hist <- ggplot(data=df2\_summarized, mapping = aes(x=date, y=sumSteps)) +  
 geom\_bar(stat="identity", fill = 'blue') +  
 theme(axis.text.x = element\_text(angle = 90)) + #rotate x axis labels  
 ggtitle('Steps Per Day') +  
 theme(plot.title = element\_text(face='bold', size=16))  
  
png('figures/totalStepsPerDay.png', width=1000)  
hist  
dev.off()

## png   
## 2

#Calculate mean & median  
mean(df2\_summarized$sumSteps, na.rm = TRUE)

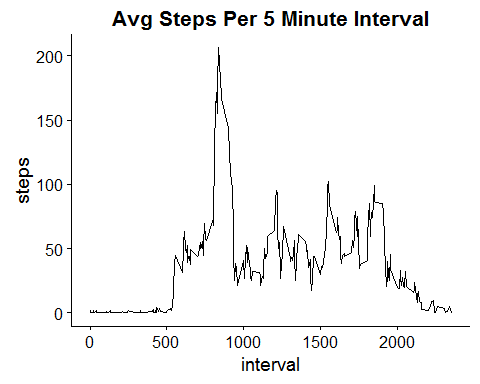
## [1] 10766.19

median(df2\_summarized$sumSteps, na.rm = TRUE)

## [1] 10765

## What is the average daily activity pattern?

ggplot(data=df2, aes(interval, steps)) +  
 stat\_summary(fun.y = mean, geom = 'line') +  
 ggtitle('Avg Steps Per 5 Minute Interval') +  
 theme(plot.title = element\_text(face='bold', size=16))



avg\_steps\_by\_interval <- ddply(df2, c("interval"), summarize, avgSteps = sum(steps)) #summarize  
  
#Print busiest time  
busiestInterval <- (avg\_steps\_by\_interval$interval[avg\_steps\_by\_interval$avgSteps == max(avg\_steps\_by\_interval$avgSteps, na.rm = T)])

## Imputing missing values

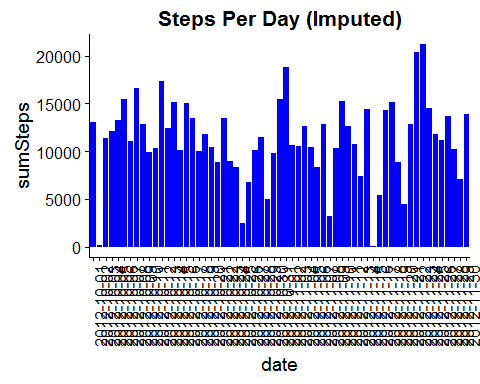
sum(is.na(df$steps)) #calculate sum of NAs

## [1] 2304

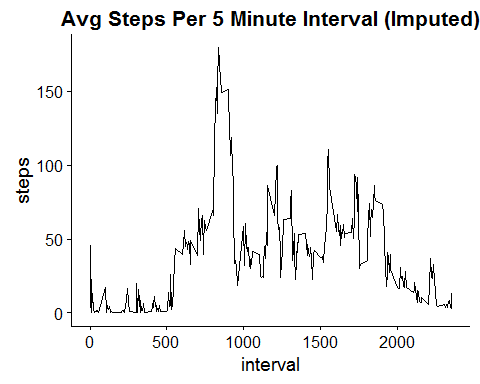
tempData <- mice(df, seed = 500) #impute missing values

##   
## iter imp variable  
## 1 1 steps  
## 1 2 steps  
## 1 3 steps  
## 1 4 steps  
## 1 5 steps  
## 2 1 steps  
## 2 2 steps  
## 2 3 steps  
## 2 4 steps  
## 2 5 steps  
## 3 1 steps  
## 3 2 steps  
## 3 3 steps  
## 3 4 steps  
## 3 5 steps  
## 4 1 steps  
## 4 2 steps  
## 4 3 steps  
## 4 4 steps  
## 4 5 steps  
## 5 1 steps  
## 5 2 steps  
## 5 3 steps  
## 5 4 steps  
## 5 5 steps

df3 <- complete(tempData, 1)  
df3\_summarized <- ddply(df3, c("date"), summarize, sumSteps = sum(steps)) #summarize  
  
#Plot new data set  
ggplot(data=df3\_summarized, mapping = aes(x=date, y=sumSteps)) +  
 geom\_bar(stat="identity", fill = 'blue') +  
 theme(axis.text.x = element\_text(angle = 90)) + #rotate x axis labels  
 ggtitle('Steps Per Day (Imputed)') +  
 theme(plot.title = element\_text(face='bold', size=16))



ggplot(data=df3, aes(interval, steps)) +  
 stat\_summary(fun.y = mean, geom = 'line') +  
 ggtitle('Avg Steps Per 5 Minute Interval (Imputed)') +  
 theme(plot.title = element\_text(face='bold', size=16))



#Calculate mean & median  
mean(df3\_summarized$sumSteps, na.rm = TRUE)

## [1] 11237.28

median(df3\_summarized$sumSteps, na.rm = TRUE)

## [1] 11352

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

df3$weekend <- is.weekend(df3$date)  
for (i in 1:nrow(df3)) {  
  
 if (df3$weekend[i] == TRUE) {  
 df3$dayType[i] <- 'weekend'  
 } else {  
 df3$dayType[i] <- 'weekday'  
 }  
}  
  
plot\_weekend <- ggplot(data=subset(df3, dayType == 'weekend'), aes(interval, steps)) +  
 stat\_summary(fun.y = mean, geom = 'line') +  
 ggtitle('Weekend Avg Steps Per 5 Minute Interval (Imputed)') +  
 theme(plot.title = element\_text(face='bold', size=16))   
  
plot\_weekday <- ggplot(data=subset(df3, dayType == 'weekday'), aes(interval, steps)) +  
 stat\_summary(fun.y = mean, geom = 'line') +  
 ggtitle('Weekday Avg Steps Per 5 Minute Interval (Imputed)') +  
 theme(plot.title = element\_text(face='bold', size=16))   
  
#Save plot to figures folder  
png('figures/final\_figure.png')  
plot\_grid(plot\_weekday, plot\_weekend, ncol = 1, nrow=2)  
dev.off()

## png   
## 2

#Print plot  
plot\_grid(plot\_weekday, plot\_weekend, ncol = 1, nrow=2)

