Reproducible Research: Peer Assessment 1

## Loading and preprocessing the data

Data are located in a subdirectory called */data*.

We load data from the data subdirectory. All other preprocessing varies according to the question, and is performed per step.

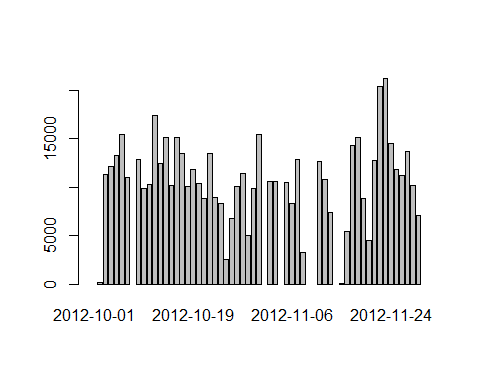
library(lattice)  
dataFile <- "Data/activity.csv"  
data <- read.csv(dataFile)

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

aggregate(data$steps ~ data$date, data, sum)

## data$date data$steps  
## 1 2012-10-02 126  
## 2 2012-10-03 11352  
## 3 2012-10-04 12116  
## 4 2012-10-05 13294  
## 5 2012-10-06 15420  
## 6 2012-10-07 11015  
## 7 2012-10-09 12811  
## 8 2012-10-10 9900  
## 9 2012-10-11 10304  
## 10 2012-10-12 17382  
## 11 2012-10-13 12426  
## 12 2012-10-14 15098  
## 13 2012-10-15 10139  
## 14 2012-10-16 15084  
## 15 2012-10-17 13452  
## 16 2012-10-18 10056  
## 17 2012-10-19 11829  
## 18 2012-10-20 10395  
## 19 2012-10-21 8821  
## 20 2012-10-22 13460  
## 21 2012-10-23 8918  
## 22 2012-10-24 8355  
## 23 2012-10-25 2492  
## 24 2012-10-26 6778  
## 25 2012-10-27 10119  
## 26 2012-10-28 11458  
## 27 2012-10-29 5018  
## 28 2012-10-30 9819  
## 29 2012-10-31 15414  
## 30 2012-11-02 10600  
## 31 2012-11-03 10571  
## 32 2012-11-05 10439  
## 33 2012-11-06 8334  
## 34 2012-11-07 12883  
## 35 2012-11-08 3219  
## 36 2012-11-11 12608  
## 37 2012-11-12 10765  
## 38 2012-11-13 7336  
## 39 2012-11-15 41  
## 40 2012-11-16 5441  
## 41 2012-11-17 14339  
## 42 2012-11-18 15110  
## 43 2012-11-19 8841  
## 44 2012-11-20 4472  
## 45 2012-11-21 12787  
## 46 2012-11-22 20427  
## 47 2012-11-23 21194  
## 48 2012-11-24 14478  
## 49 2012-11-25 11834  
## 50 2012-11-26 11162  
## 51 2012-11-27 13646  
## 52 2012-11-28 10183  
## 53 2012-11-29 7047

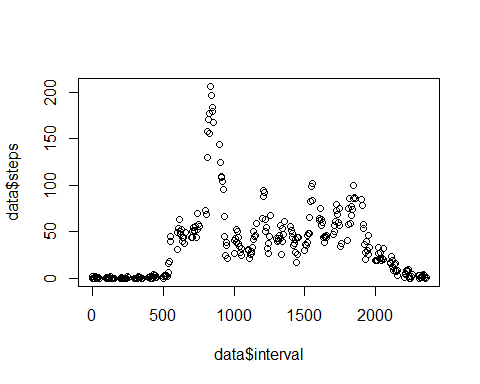
analysis1a <- tapply(data$steps, data$date, FUN=sum)  
barplot(analysis1a)



## What is the average daily activity pattern?

We will 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). Missing data are ignored.

stepsPerInterval <- aggregate(data$steps ~ data$interval, data, mean)  
plot(stepsPerInterval)



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

maxNumOfSteps <- max(stepsPerInterval["data$steps"], na.rm = TRUE)  
intervalWithMaxAvgSteps <- stepsPerInterval[which(stepsPerInterval["data$steps"] == maxNumOfSteps),1]

## Imputing missing values

We first calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)

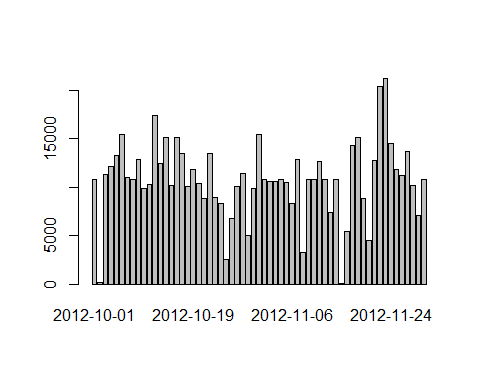
numberOfNAs <- sum(!complete.cases(data))

Next, we devise a strategy for filling in all of the missing values in the dataset. We use the mean for that 5-minute interval. We start by creating a new dataset that is equal to the original dataset but with the missing data filled in.

imputedData <- data  
for ( ndx in 1:nrow(imputedData)) if ( is.na(imputedData[ndx,1]) ) imputedData[ndx,1] <- stepsPerInterval[which(stepsPerInterval$`data$interval`==imputedData[ndx,3]),2]

Next, we make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day. Do these values differ from the estimates from the first part of the assignment? What is the impact of imputing missing data on the estimates of the total daily number of steps?

analysis3a <- tapply(imputedData$steps, imputedData$date, FUN=sum)  
barplot(analysis3a)



Finally, we report the Mean...

dayMean3a <- mean(analysis3a, na.rm = TRUE)

...and the Median

dayMedian3a <- median(analysis3a, na.rm = TRUE)

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

We will create a new factor variable in the dataset with two levels -- "weekday" and "weekend" indicating whether a given date is a weekday or weekend day.

data4 <- imputedData  
data4$date <- as.POSIXlt(data4$date)  
data4$dayType <- ifelse ( data4$date$wday %in% c(0,6),"weekend","weekday")  
data4$dayType <- as.factor(data4$dayType)

And finally present a panel plot containing a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis).

stepsPerInterval2 <- aggregate(data4$steps ~ data4$interval\*data4$dayType, data4, mean)  
xyplot(stepsPerInterval2$`data4$steps`~stepsPerInterval2$`data4$interval`|stepsPerInterval2$`data4$dayType`,  
 layout=(c(1,2)),  
 type="l",  
 xlab = "Interval",  
 ylab = "Number of Steps")

