# Reproducible Research: Peer Assessment 1

## Loading and preprocessing the data

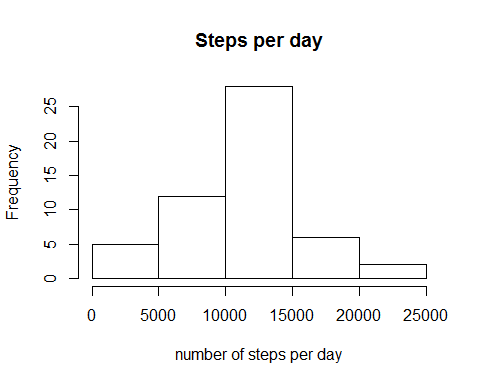
Loading the data:

setwd("Z:/Dokumente/Coursera/Reproducible Research/Assignment1/RepData\_PeerAssessment1")  
studis <- read.csv("activity.csv", header=T)

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

Rearrangin the data, aggregated by days and plot a histogram:

stepsPday <- tapply(studis$steps, studis$date, sum)  
stepsPday <- stack(stepsPday)  
hist(stepsPday$values, main="Steps per day", xlab="number of steps per day")



Statistics of the steps per day (mean, median):

mean(stepsPday$values, na.rm=T)

## [1] 10766

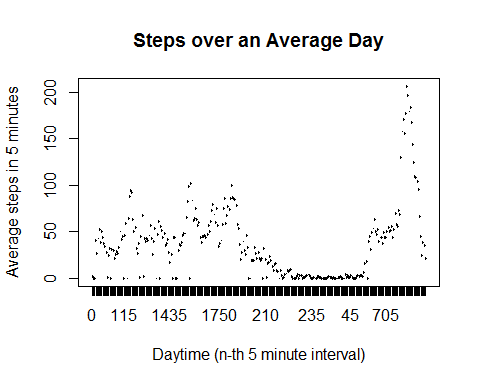
median(stepsPday$values, na.rm=T)

## [1] 10765

## What is the average daily activity pattern?

Time series plot:

timeList <- stack(tapply(studis$steps, studis$interval, mean, na.rm=T, simplify=T))  
plot(timeList$values~timeList$ind, type="l", main="Steps over an Average Day", xlab="Daytime (n-th 5 minute interval)", ylab="Average steps in 5 minutes")



Interval with maximum average steps:

timeList[which.max(timeList$values),]

## values ind  
## 104 206.2 835

## Imputing missing values

Total number of NA's:

countNA <- 0  
for (step in studis$steps) {  
 if (is.na(step)){  
 countNA <- countNA + 1  
 }  
}  
countNA

## [1] 2304

Filling out the NA's:

intBYdate <- reshape(studis, timevar="interval" , idvar="date", direction="wide")  
aveDays <- stack(lapply(intBYdate[,-1], mean, na.rm=T))  
aveDays[,2] <- substring(aveDays$ind,7)  
  
studisWoNa <- studis  
studisWoNa <- merge(studisWoNa, aveDays, by.x = "interval", by.y = "ind")  
  
studisWoNa$steps[is.na(studisWoNa$steps)] <- studisWoNa$values[is.na(studisWoNa$steps)]  
studisWoNa <- studisWoNa[,1:3]

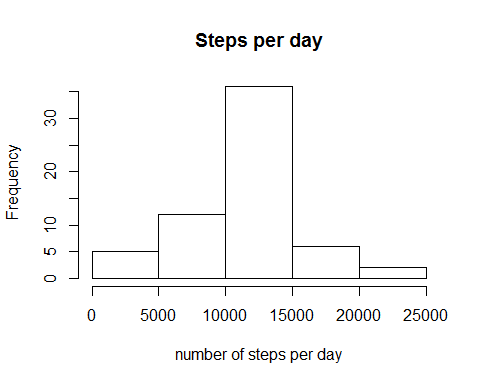
Show first lines of result witout NA's:

studisWoNa[1:20,]

## interval steps date  
## 1 0 1.717 2012-10-01  
## 2 0 0.000 2012-11-23  
## 3 0 0.000 2012-10-28  
## 4 0 0.000 2012-11-06  
## 5 0 0.000 2012-11-24  
## 6 0 0.000 2012-11-15  
## 7 0 0.000 2012-10-20  
## 8 0 0.000 2012-11-16  
## 9 0 0.000 2012-11-07  
## 10 0 0.000 2012-11-25  
## 11 0 1.717 2012-11-04  
## 12 0 0.000 2012-11-08  
## 13 0 0.000 2012-10-12  
## 14 0 0.000 2012-10-30  
## 15 0 0.000 2012-11-26  
## 16 0 47.000 2012-10-04  
## 17 0 0.000 2012-11-27  
## 18 0 0.000 2012-10-31  
## 19 0 0.000 2012-11-18  
## 20 0 0.000 2012-10-05

Histogram without the NA's:

stepsPdayWoNa <- tapply(studisWoNa$steps, studisWoNa$date, sum)  
stepsPdayWoNa <- stack(stepsPdayWoNa)  
hist(stepsPdayWoNa$values, main="Steps per day", xlab="number of steps per day")



Impact of imputing NA's:

mean(stepsPday$values, na.rm=T)- mean(stepsPdayWoNa$values)

## [1] 0

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

Creating weekday column:

daycat <- function(x){  
 if (x == "Samstag" | x == "Sonntag"){  
 "weekend"  
 }else{  
 "weekday"  
 }  
}  
studisWoNa[,"daytype"]<- stack(sapply(weekdays(as.Date(studisWoNa$date)),daycat))[,1]

Plotting:

library("lattice")  
  
timeListWeek <- aggregate(steps~interval+daytype, data = studisWoNa, mean)  
xyplot(steps ~ interval | daytype, data = timeListWeek, layout = c(1,2), type= "l")

