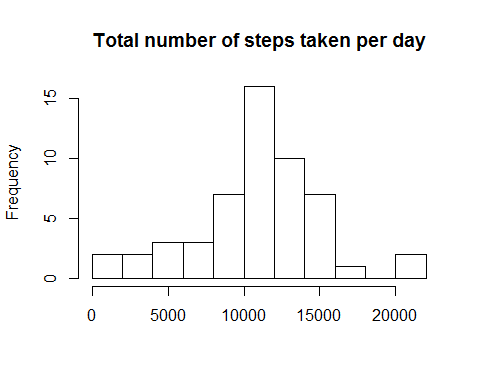
# Reproducible Researh: Course Assignment 1

## 1.Read and pre-process data

setwd("C:/Users/Peter/Documents/Coursera/JH\_DS/Reproducable Research/Week2/repdata\_data\_activity")  
activity <- read.csv("activity.csv", sep = ",")  
activity\_rm<-activity[which(!is.na(activity$steps)),]

## 2.Histogram of the total number of steps taken each day

perday<-tapply(activity\_rm$steps, activity\_rm$date, sum)   
hist(perday,10, main = "Total number of steps taken per day", xlab = "")



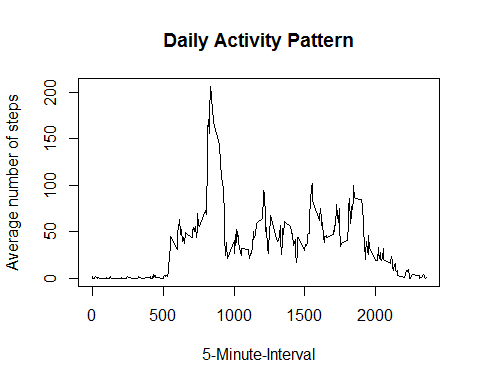
## 3.Mean and median number of steps taken each day

summary(perday)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 41 8841 10760 10770 13290 21190 8

## 4.Time series plot of the average number of steps taken

dailyact<-tapply(activity\_rm$steps, activity\_rm$interval, mean)   
plot(y = dailyact, x = names(dailyact), type = "l", xlab = "5-Minute-Interval", main = "Daily Activity Pattern", ylab = "Average number of steps")



## 5.The 5-minute interval that, on average, contains the maximum number of steps

dailyact[dailyact==max(dailyact)]

## 835   
## 206.1698

## 6.Code to describe and show a strategy for imputing missing data

act\_new <- activity   
act\_new[which(is.na(act\_new$steps)),1]<- dailyact[as.character(act\_new[which(is.na(act\_new$steps)),3])]

## 7.Histogram of the total number of steps taken each day after missing values are imputed

perday\_new<-tapply(act\_new$steps, act\_new$date, sum)   
  
hist(perday\_new,10, main = "Total number of steps taken per day (missing values imputed)", xlab = "Steps", ylim =c(0, 25))   
abline(v = median(perday\_new), col = 4, lwd = 4)

