# Unzip and store data, clean up the raw file  
library(dplyr)

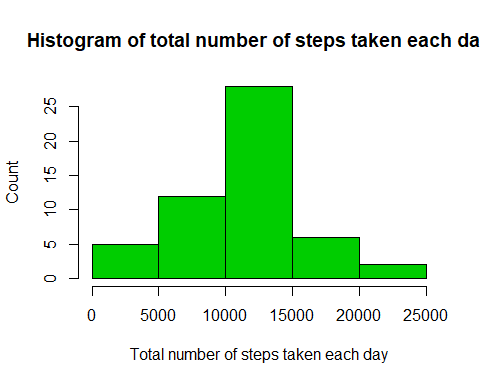
##   
## Attaching package: 'dplyr'

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

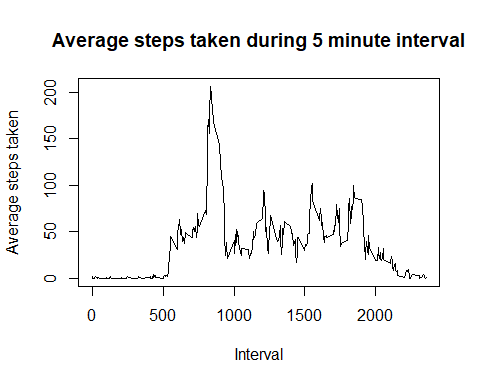
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

rawdata <- read.csv("C:\\Users\\Mahe\\Desktop\\Data Science\\Reproducible-research\\project1\\activity.csv")  
  
  
# Convert date variable from factor to date, and remove NAs  
data <- na.omit(rawdata)  
data$date <- as.Date(data$date)

stepsperday <- data %>%   
 group\_by(date) %>%   
 summarize(TotalSteps=sum(steps))  
  
# Show histogram of steps per day  
hist(stepsperday$TotalSteps,   
 xlab="Total number of steps taken each day",   
 ylab="Count",   
 main="Histogram of total number of steps taken each day",  
 col=3)

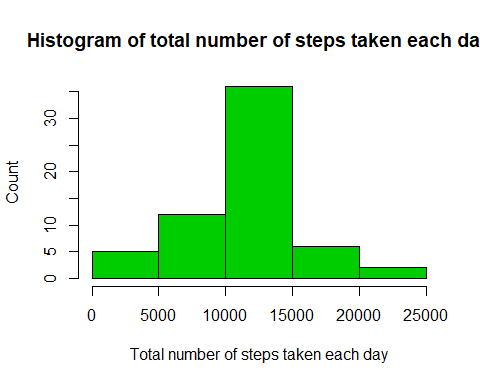


# Group data by 5 minute interval and summarize the average  
# number of steps in that interval  
fiveaverage <- data %>%  
 group\_by(interval) %>%  
 summarize(AverageSteps=mean(steps))  
  
# Make an average activity plot  
plot(fiveaverage$interval, fiveaverage$AverageSteps,   
 type="l",  
 xlab="Interval",  
 ylab="Average steps taken",  
 main="Average steps taken during 5 minute interval")



maxstepinterval <- fiveaverage$interval[which.max(fiveaverage$AverageSteps)]

missing <- sum(is.na(rawdata$steps))  
filldata <- rawdata  
for (i in 1:nrow(filldata)) {  
 if (is.na(filldata$steps[i])) {  
 # Find the index value for when the interval matches the average  
 ndx <- which(filldata$interval[i] == fiveaverage$interval)  
 # Assign the value to replace the NA  
 filldata$steps[i] <- fiveaverage[ndx,]$AverageSteps  
 }  
}  
  
# Make sure the date variable is still a date.  
filldata$date <- as.Date(filldata$date)  
# Group data by date, and summarize the sum of steps  
fillstepsperday <- filldata %>%   
 group\_by(date) %>%   
 summarize(TotalSteps=sum(steps))  
  
# Show histogram of steps per day  
hist(fillstepsperday$TotalSteps,   
 xlab="Total number of steps taken each day",   
 ylab="Count",   
 main="Histogram of total number of steps taken each day",  
 col=3)



fillmeansteps <- mean(fillstepsperday$TotalSteps)  
fillmeansteps

## [1] 10766.19

fillmediansteps <- median(fillstepsperday$TotalSteps)  
fillmediansteps

## [1] 10766.19

# Make weekday variable  
filldata$day <- weekdays(filldata$date)  
# Define all days as weekdays  
filldata$daytype <- "weekday"  
# Fix days that are saturday or sunday to be weekends  
filldata$daytype[filldata$day %in% c("Saturday", "Sunday")] <- "weekend"  
# Group data by 5 minute interval and summarize the average  
# number of steps in that interval  
dayaverage <- filldata %>%  
 group\_by(daytype, interval) %>%  
 summarize(AverageSteps=mean(steps))  
  
library(ggplot2)  
  
# Use qplot because facets easily divides the graph into two better than  
# the base graphics system.  
qplot(interval, AverageSteps, data=dayaverage,  
 type="l",  
 geom="line",  
 xlab="Interval",  
 ylab="Number of Steps (Average)",  
 main="Average steps taken Weekends vs. Weekdays",  
 facets =daytype ~ .)

## Warning: Ignoring unknown parameters: type

