# reading the file

>

> if(!file.exists("getdata-projectfiles-UCI HAR Dataset.zip")) {

+ tmp <- tempfile()

+ download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip",tmp)

+ unzip(tmp)

+ unlink(tmp)

+ }

trying URL 'http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip'

Content type 'application/zip' length 53559 bytes (52 KB)

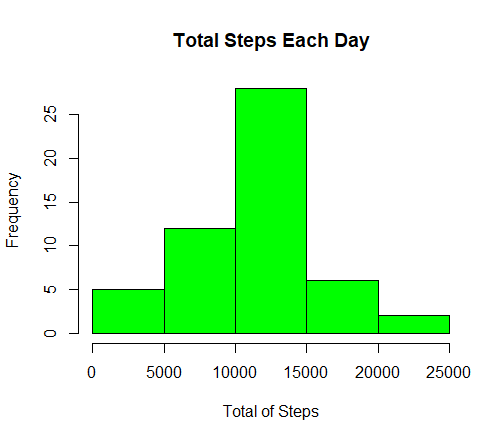
downloaded 52 KB

df <- read.csv("activity.csv")

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

> sum\_steps <- aggregate(steps ~ date, df, sum)

> hist(sum\_steps$steps, main = paste("Total Steps Each Day"), col="green", xlab="Total of Steps")



### **Calculate and report the mean and median of the total number of steps taken per day.**

> print(paste("The mean is:", mean(sum\_steps$steps), sep = " "))

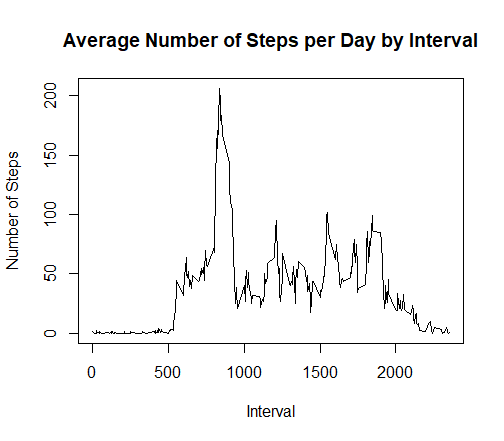
[1] "The mean is: 10766.1886792453"

> print(paste("The median is:", median(sum\_steps$steps), sep = " "))

[1] "The median is: 10765"

# What is the average daily activity pattern?

* Make 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 days (y-axis)
* Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?
* > steps\_by\_interval <- aggregate(steps ~ interval, df, mean)
* > plot(steps\_by\_interval$interval,steps\_by\_interval$steps, type="l", xlab="Interval", ylab="Number of Steps",main="Average Number of Steps per Day by Interval")



> print(paste("The 5-minute interval with the maximum number of steps is:", steps\_by\_interval[which.max(steps\_by\_interval$steps),1], sep = " "))

[1] "The 5-minute interval with the maximum number of steps is: 835"

# Imputing missing values

> print(paste("The total of missing data is: ", sum(!complete.cases(df)), sep = " "))

[1] "The total of missing data is: 2304"

**All of the missing values are filled in with mean value for that 5-minute interval and replace each missing value with the mean value of its 5-minute interval**

> stepsInterval <- aggregate(steps ~ interval, data = df, mean, na.rm = TRUE)

> interval2steps <- function(interval) {

+ stepsInterval[stepsInterval$interval == interval, ]$steps

+ }

> Filled <- df # Make a new dataset with the original data

> count = 0 # Count the number of data filled in

> for (i in 1:nrow(Filled)) {

+ if (is.na(Filled[i, ]$steps)) {

+ Filled[i, ]$steps <- interval2steps(Filled[i, ]$interval)

+ count = count + 1

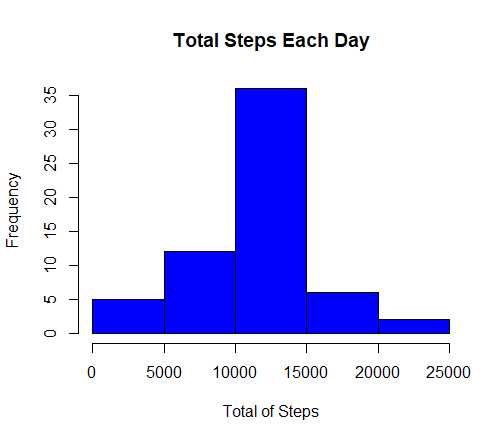
+ }

+ }

**Let’s make a histogram of the total number of steps taken each day and calculate the mean and median total number of steps.**

> totalSteps2 <- aggregate(steps ~ date, data = Filled, sum)

> hist(totalSteps2$steps, main = paste("Total Steps Each Day"), col="blue", xlab="Total of Steps")



> print(paste("This is the mean:", mean(totalSteps2$steps), sep = " "))

[1] "This is the mean: 10766.1886792453"

> print(paste("This is the median:", median(totalSteps2$steps), sep = " "))

[1] "This is the median: 10766.1886792453"

**Are there diferences in activity patterns between weekdays and weekends?**

> Filled$day = ifelse(as.POSIXlt(as.Date(Filled$date))$wday%%6 ==

+ 0, "weekend", "weekday")

> # For Sunday and Saturday : weekend, Other days : weekday

> Filled$day = factor(Filled$day, levels = c("weekday", "weekend"))

> library(lattice)

> stepsInterval2 = aggregate(steps ~ interval + day, Filled, mean)

> xyplot(steps ~ interval | factor(day), data = stepsInterval2, aspect = 1/2, type = "l")

