PA1\_template

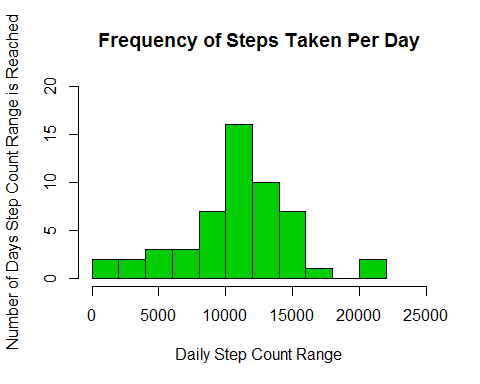
This my first Rmarkdown.file

**Loading and preprocessing the data**

activity <- read.csv("C:/activity.csv", colClasses = c("numeric", "character", "numeric"))  
suppressMessages(require(lattice))  
activity["date"] <- as.Date(activity$date, "%Y-%m-%d")

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

steps\_per\_day <- aggregate(steps ~ date, data = activity, sum, na.rm = TRUE)  
mean\_steps\_per\_day <- mean(steps\_per\_day$steps)  
median\_steps\_per\_day <- median(steps\_per\_day$steps)  
  
hist(steps\_per\_day$steps, main = "Frequency of Steps Taken Per Day",  
 xlab = "Daily Step Count Range",  
 ylab="Number of Days Step Count Range is Reached", col = "green3", xlim=c(0,25000), ylim=c(0,20), breaks=10)

 *The mean steps taken per day is 10766*

mean\_steps\_per\_day

## [1] 10766.19

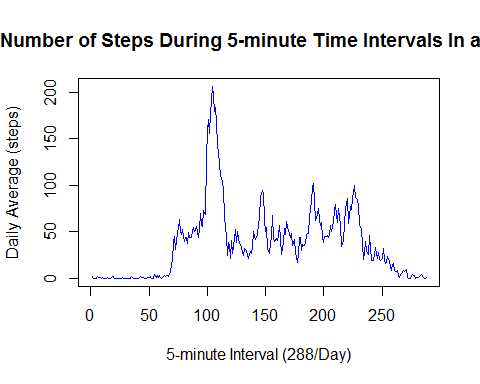
*The median steps taken per day is 10765*

median\_steps\_per\_day

## [1] 10765

**What is the average daily activity pattern?**

time\_series <- tapply(activity$steps, activity$interval %% 100 / 5 + activity$interval %/% 100 \* 12 + 1, mean, na.rm = TRUE)  
  
max\_interval <- which.max(time\_series)  
hour\_of\_day\_start <- (max\_interval \* 5) %/% 60  
minute\_of\_hour\_start <- (max\_interval \* 5) %% 60  
  
hour\_of\_day\_end <- ((max\_interval + 1) \* 5) %/% 60  
minute\_of\_hour\_end <- ((max\_interval + 1) \* 5) %% 60  
  
am\_pm\_start <- "AM"  
am\_pm\_end <- "AM"  
  
if(hour\_of\_day\_start > 12) {  
 hour\_of\_day\_start <- hour\_of\_day\_start - 12  
 am\_pm\_start <- "PM"  
}  
  
if(hour\_of\_day\_end > 12) {  
 hour\_of\_day\_end <- hour\_of\_day\_end - 12  
 am\_pm\_end <- "PM"  
}  
  
  
plot(row.names(time\_series), time\_series, type = "l", xlab = "5-minute Interval (288/Day)",   
 ylab = "Daily Average (steps)", main = "Average Number of Steps During 5-minute Time Intervals In a 24-hour Day",   
 col = "blue")



max\_interval

## 104   
## 104

**Imputing missing values**

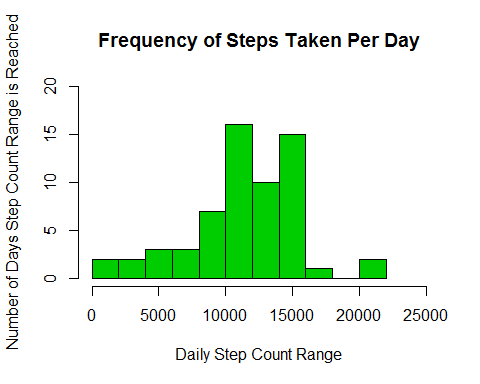
NA\_count <- sum(is.na(activity))  
NA\_count

## [1] 2304

suppressMessages(require(reshape))

## Warning in library(package, lib.loc = lib.loc, character.only = TRUE,  
## logical.return = TRUE, : there is no package called 'reshape'

suppressMessages(require(reshape2))  
a <- dcast(activity, interval ~ date, value.var="steps", fill=0)  
r <- dcast(activity, interval ~ date, value.var = "steps", fill = rowMeans(a, na.rm = TRUE))  
r2 <- reshape(r, direction = "long", varying=list(names(r)[2:length(names(r))]),  
 v.names=c("steps"), timevar="date", idvar=c("interval"),  
 times=names(r)[2:length(names(r))], new.row.names=1:dim(activity)[1])  
  
steps\_per\_day2 <- aggregate(steps ~ date, data = r2, sum, na.rm = TRUE)  
hist(steps\_per\_day2$steps, main = "Frequency of Steps Taken Per Day", xlab = "Daily Step Count Range",  
 ylab="Number of Days Step Count Range is Reached", col = "green3", xlim=c(0,25000), ylim=c(0,20), breaks=10)



mean\_steps\_per\_day <- mean(steps\_per\_day2$steps)  
median\_steps\_per\_day <- median(steps\_per\_day2$steps)  
  
mean\_steps\_per\_day

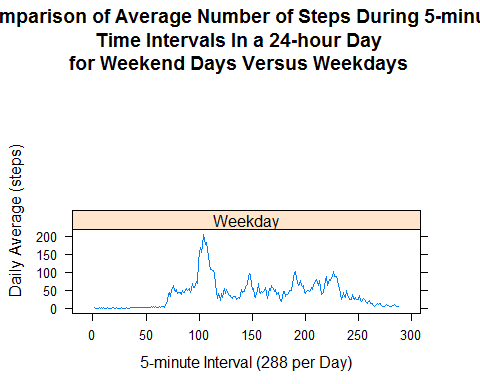
## [1] 11278.56

median\_steps\_per\_day

## [1] 11458

**re there differences in activity patterns between weekdays and weekends?**

r2["date"] <- as.Date(r2$date, "%Y-%m-%d")  
r2[(weekdays(r2$date) %in% c("Saturday", "Sunday")), "TypeOfDay"] <- "Weekend"  
r2[!(weekdays(r2$date) %in% c("Saturday", "Sunday")), "TypeOfDay"] <- "Weekday"  
  
steps <- aggregate(steps ~ interval + TypeOfDay, data = r2, mean)  
names(steps) <- c("Interval", "TypeOfDay", "Steps")  
steps$Interval <- steps$Interval %% 100 / 5 + steps$Interval %/% 100 \* 12 + 1  
  
splot <- xyplot(Steps ~ Interval | TypeOfDay, steps, type = "l", layout = c(1, 2),xlab = "5-minute Interval (288 per Day)", ylab = "Daily Average (steps)")  
update(splot,main="Comparison of Average Number of Steps During 5-minute\nTime Intervals In a 24-hour Day\nfor Weekend Days Versus Weekdays")

 Weekday steps show a large peak around 8:40AM followed by 4 smaller peaks around lunch time, afternoon break time, and supper time. Step data appears more uniform throughout weekend days and have smaller peaks.