PA1_template

This my first Rmarkdown.file

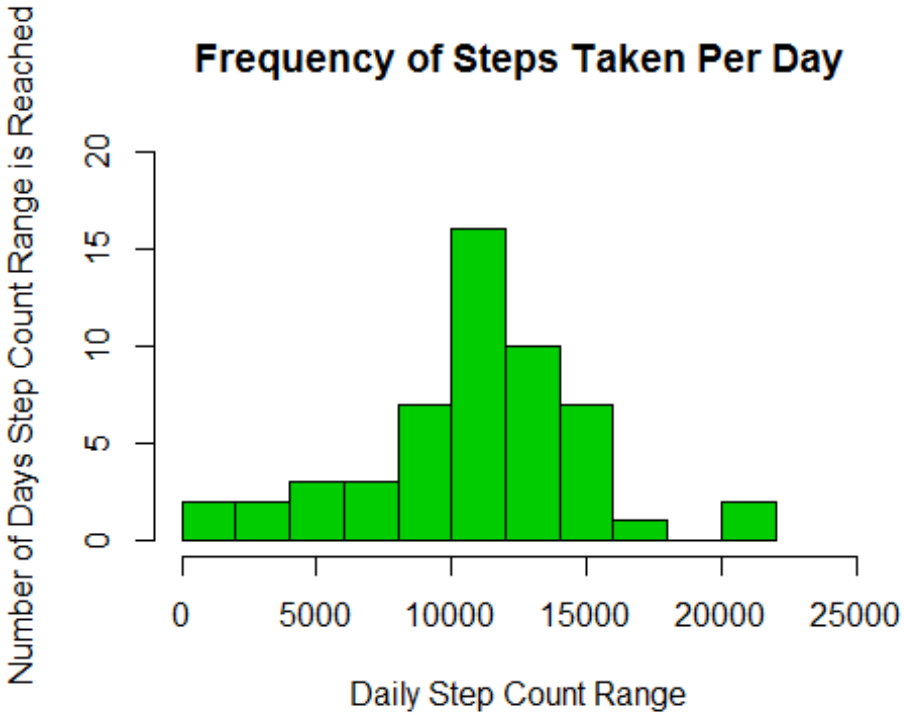
Loading and preprocessing the data

```
activity <- read.csv("C:/activity.csv", colClasses = c("numeric", "c
haracter", "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", co
l = "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"
```

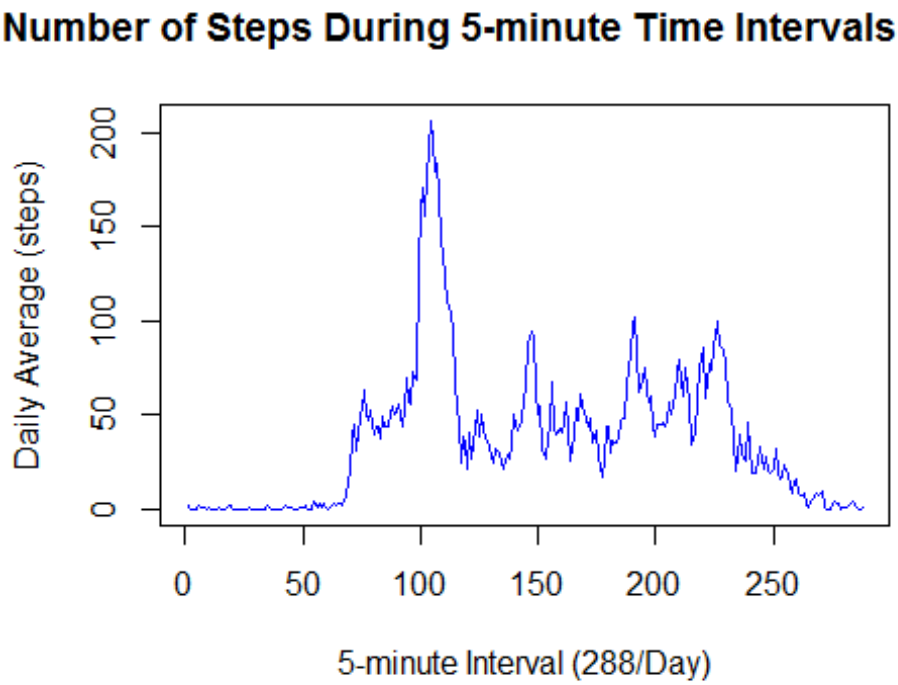
```
am_pm_end <- "AM"

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-minu
te 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
```

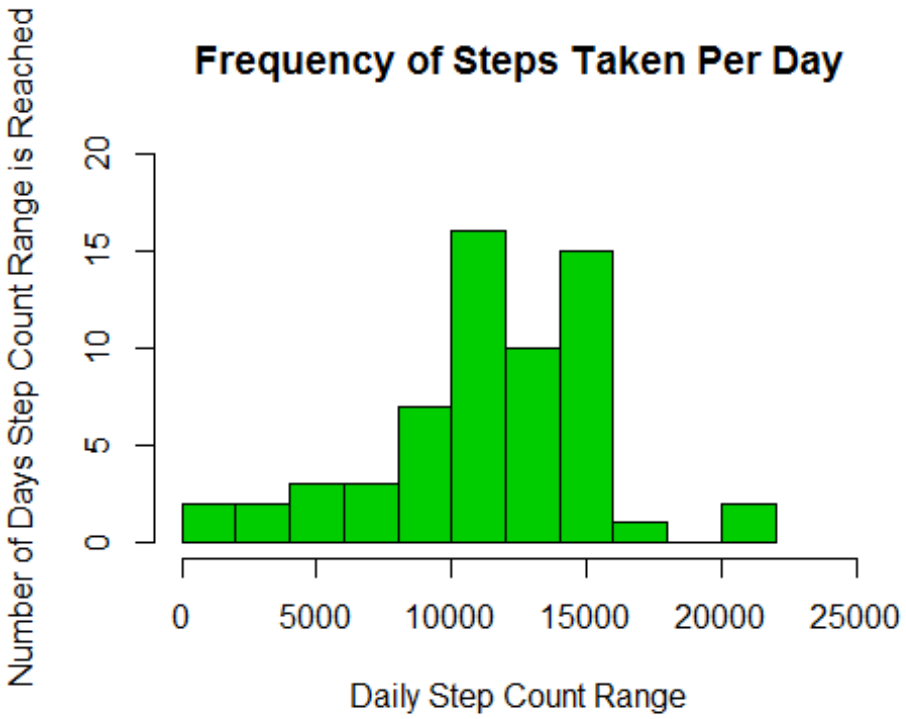
```
## [1] 2304
```

```
suppressMessages(require(reshape))
```

```
## Warning in library(package, lib.loc = lib.loc, character.only = T
RUE,
## 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 = ro
wMeans(a, na.rm = TRUE))
r2 <- reshape(r, direction = "long", varying=list(names(r)[2:length
(names(r))]),
                                v.names=c("steps"), timevar="date", i
dvar=c("interval"),
                                times=names(r)[2:length(names(r))], n
ew.row.names=1:dim(activity)[1])

steps_per_day2 <- aggregate(steps ~ date, data = r2, sum, na.rm = TR
UE)
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
```

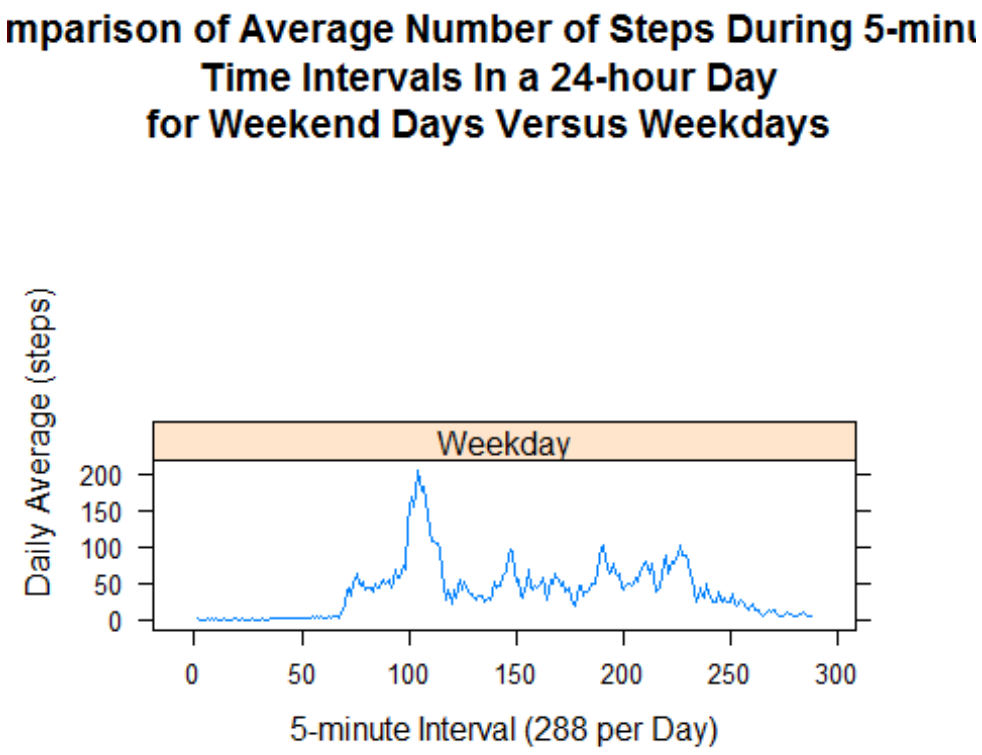
```
## [1] 11458
```

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", lay
out = c(1, 2),xlab = "5-minute Interval (288 per Day)", ylab = "Dail
y Average (steps)")
update(splot,main="Comparison of Average Number of Steps During 5-mi
nute\nTime Intervals In a 24-hour Day\nfor Weekend Days Versus Weekd
ays")
```



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.