

# Untitled

*Sankarshan Acharya*

*May 15, 2016*

## Loading and preprocessing the data into R

The following step loads the requisite data into R.

```
monitoring_activity <- read.csv("C:/Users/sanka/Documents/activity.csv")
```

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

The following does a graphical analysis on the data and creates a histogram with the mean and the median represented.

```

activity_array <- array(0, 61)

i = NULL
for(i in 1:61)
{
  a = 288*i - 287
  b = 288*i
  activity_array[i] <- sum(monitored_activity[a:b, 1])
}

monitored_activity$Total_Steps <- 0

i = NULL
for(i in 1:17568)
{
  if(i %% 288 != 0)
  {
    a = i %% 288
    monitored_activity$Total_Steps[i] <- activity_array[a + 1]
  }

  else
  {
    monitored_activity$Total_Steps[i] <- activity_array[i/288]
  }
}

c_activity_array = NULL
i = NULL
for(i in 1:length(activity_array))
{
  if(is.na(activity_array[i]))
  {
    c_activity_array = c(c_activity_array, 0)
  }

  else
  {
    c_activity_array = c(c_activity_array, activity_array[i])
  }
}

paste("The mean of the data is", mean(c_activity_array))

```

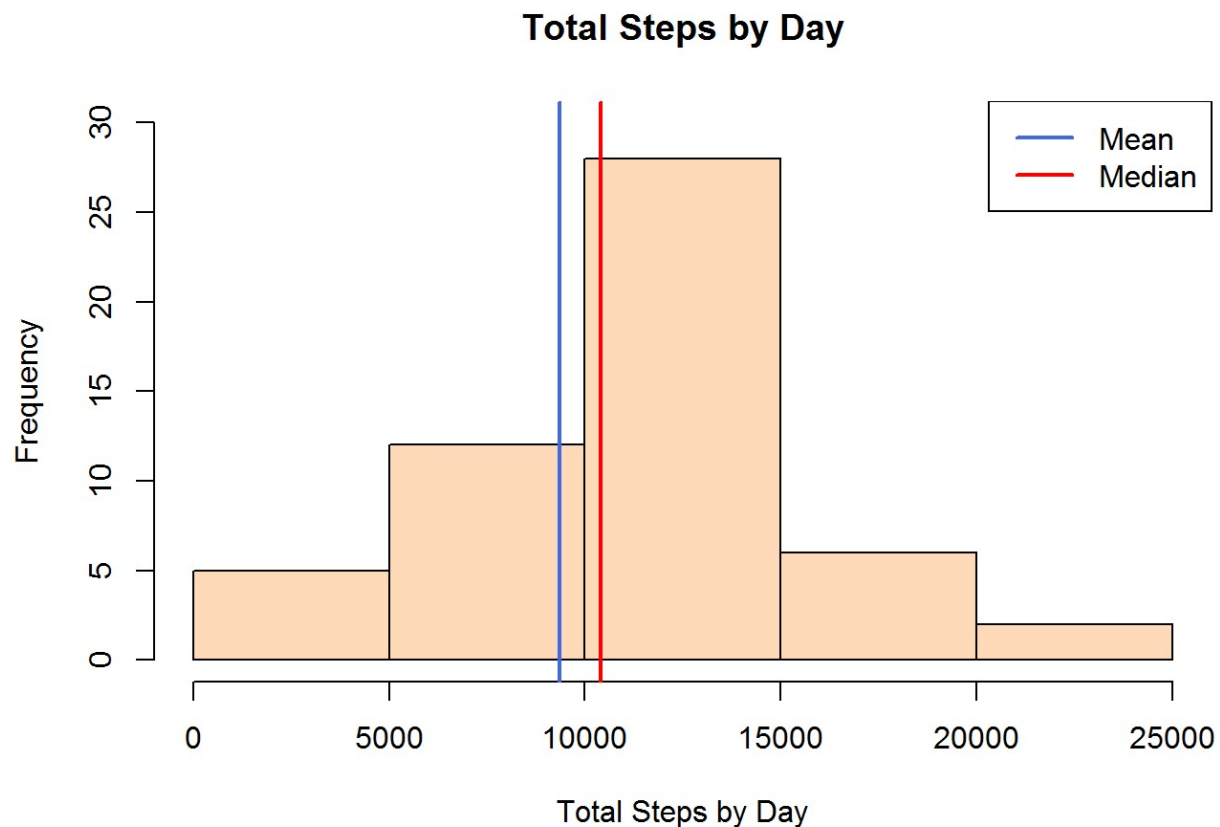
```
## [1] "The mean of the data is 9354.22950819672"
```

```
paste("The mean of the data is", median(c_activity_array))
```

```
## [1] "The mean of the data is 10395"
```

The following is the plot of the histogram with the mean and median displayed:

```
hist(activity_array, col = "peachpuff", xlab = "Total Steps by Day", main = "Total Steps by Day", ylim = c(0, 30))  
abline(v = mean(c_activity_array), col = "royalblue", lwd = 2)  
abline(v = median(c_activity_array), col = "red", lwd = 2)  
legend(x = "topright", c("Mean", "Median"), col = c("royalblue", "red"), lwd = c(2, 2))
```



## What is the average daily activity pattern?

The following code looks at each 5-minute interval data and averages them over every day in the dataset.

```

activity_array_1 <- array(0, 288)
day_time_count_array <- monitoring_activity$steps

i= NULL
for(i in 1:17568)
{
  a = is.na(day_time_count_array[i])
  if(a)
  {
    day_time_count_array[i] = 0
  }
}

i = NULL
for(i in 1:288)
{
  activity_array_1[i] <- mean(day_time_count_array[seq(i, i + 17280, by = 288)])
}

x <- monitoring_activity[1:288, 3]

a = which.max(activity_array_1)
paste("The interval containing the max interval is", x[a])

```

```
## [1] "The interval containing the max interval is 835"
```

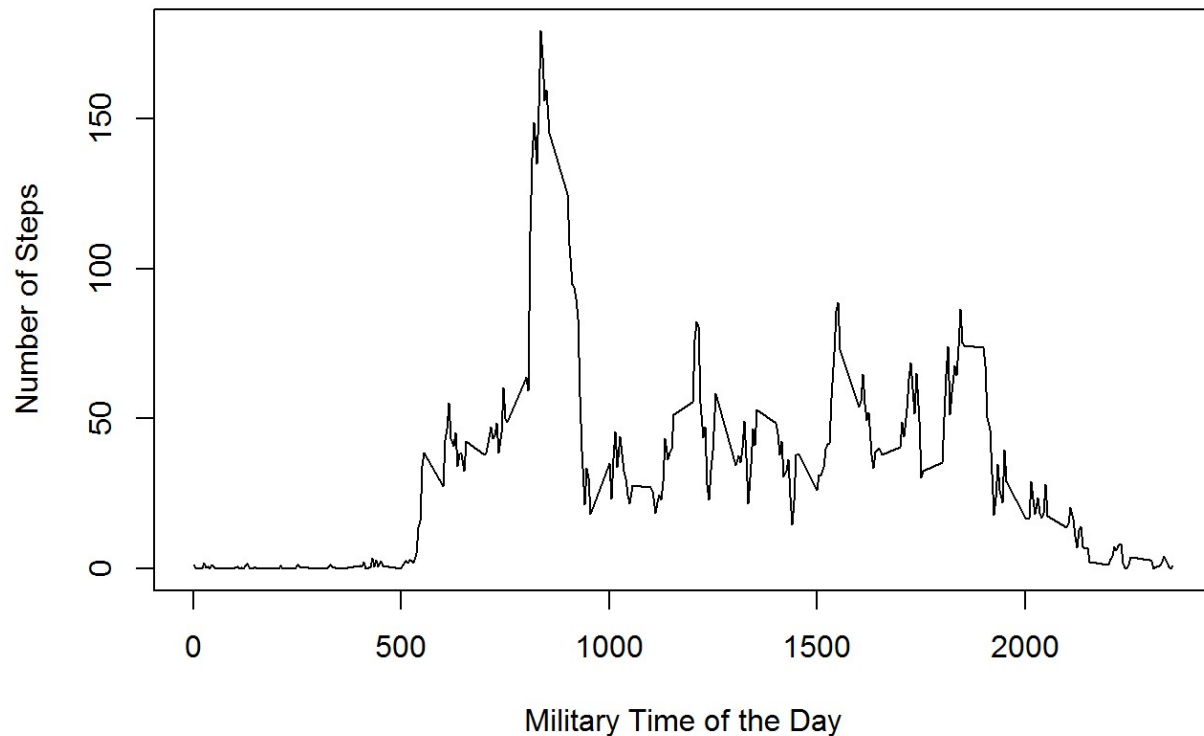
The following is the line plot of the average number of steps (averaged over all the days) given by each 5-minute interval:

```

plot(x, activity_array_1, type = "n", xlab = "Military Time of the Day", ylab = "Number of Steps", main = "Number of Steps through out the Day")
lines(x, activity_array_1, type = "l")

```

## Number of Steps through out the Day



## Imputing missing values

The following code inputs data into the NA points based on the the average assuming the values at the NA points were originally zero. More clearly, it puts the mean as the data.

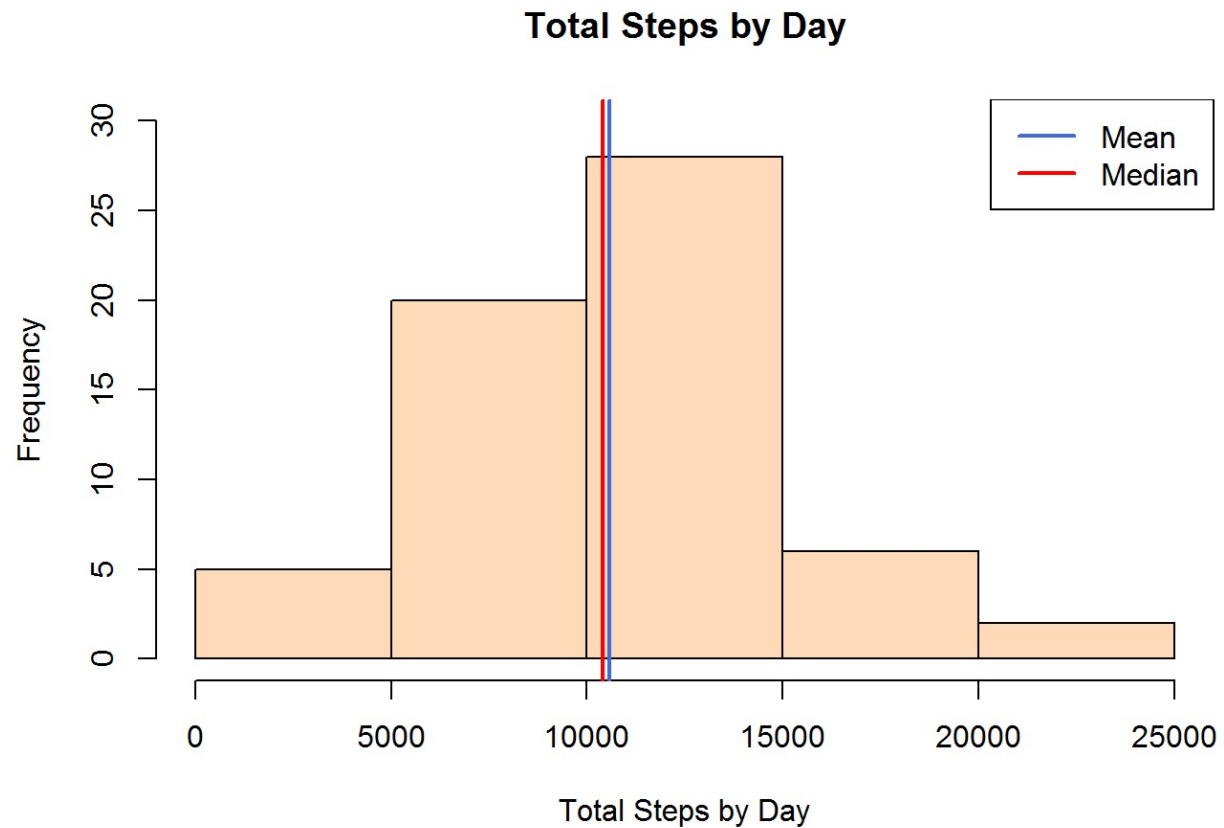
```
## [1] "The number of NA rows is 2304"
```

```
## [1] "The mean of the data is 10580.7213114754"
```

```
## [1] "The mean of the data is 10395"
```

The following code is used to create the histogram of step numbers given the fact that all the NA spots have been filled:

```
hist(activity_array_NA, col = "peachpuff", xlab = "Total Steps by Day", main =  
"Total Steps by Day", ylim = c(0, 30))  
abline(v = mean(activity_array_NA), col = "royalblue", lwd = 2)  
abline(v = median(activity_array_NA), col = "red", lwd = 2)  
legend(x = "topright", c("Mean", "Median"), col = c("royalblue", "red"), lwd =  
c(2, 2))
```



## Are there differences in activity patterns between weekdays and weekends?

The following code graphs the data based on the information given about the averages of the 5-minute intervals in the weekdays and the weekends (days).

```

monitoring_activity_non_NA$day_type <- 0
i = NULL
for(i in 1:17568)
{
  a <- weekdays(as.Date(monitoring_activity_non_NA$date[i]))

  if(a == "Saturday" || a == "Sunday")
  {
    monitoring_activity_non_NA[i, 5] <- "weekend"
  }

  else
  {
    monitoring_activity_non_NA[i, 5] <- "weekday"
  }
}

monitoring_activity_non_NA_weekday = monitoring_activity_non_NA[monitoring_acti
vity_non_NA$day_type == "weekday", ]
monitoring_activity_non_NA_weekend = monitoring_activity_non_NA[monitoring_acti
vity_non_NA$day_type == "weekend", ]

activity_array_1_NA_weekday <- array(0, 288)
activity_array_1_NA_weekend <- array(0, 288)
day_time_count_array_NA_weekday <- monitoring_activity_non_NA_weekday$steps
day_time_count_array_NA_weekend <- monitoring_activity_non_NA_weekend$steps

i = NULL
for(i in 1:288)
{
  activity_array_1_NA_weekday[i] <- mean(day_time_count_array_NA_weekday[seq
(i, i + 12672, by = 288)])
  activity_array_1_NA_weekend[i] <- mean(day_time_count_array_NA_weekend[seq
(i, i + 4320, by = 288)])
}

x <- monitoring_activity[1:288, 3]

```

The following code is used to plot the average of the 5-minute intervals over the weekdays first and then to plot the average of the 5-minute intervalse over the weekends:

```

par(mfrow=c(2,1))

plot(x, activity_array_1_NA_weekday, type = "n", xlab = "Military Time of the Day", ylab = "Number of Steps", main = "Number of Steps through out the Weekdays")
lines(x, activity_array_1_NA_weekday, type = "l")

plot(x, activity_array_1_NA_weekend, type = "n", xlab = "Military Time of the Day", ylab = "Number of Steps", main = "Number of Steps through out the Weekends")
lines(x, activity_array_1_NA_weekend, type = "l")

```

