Reproducible Research - Assignment 1

Arun Panangatt

14 May 2015

Overarching settings

```
echo = TRUE ### to make all code visible
library(ggplot2)
```

Loading data and pre-processing

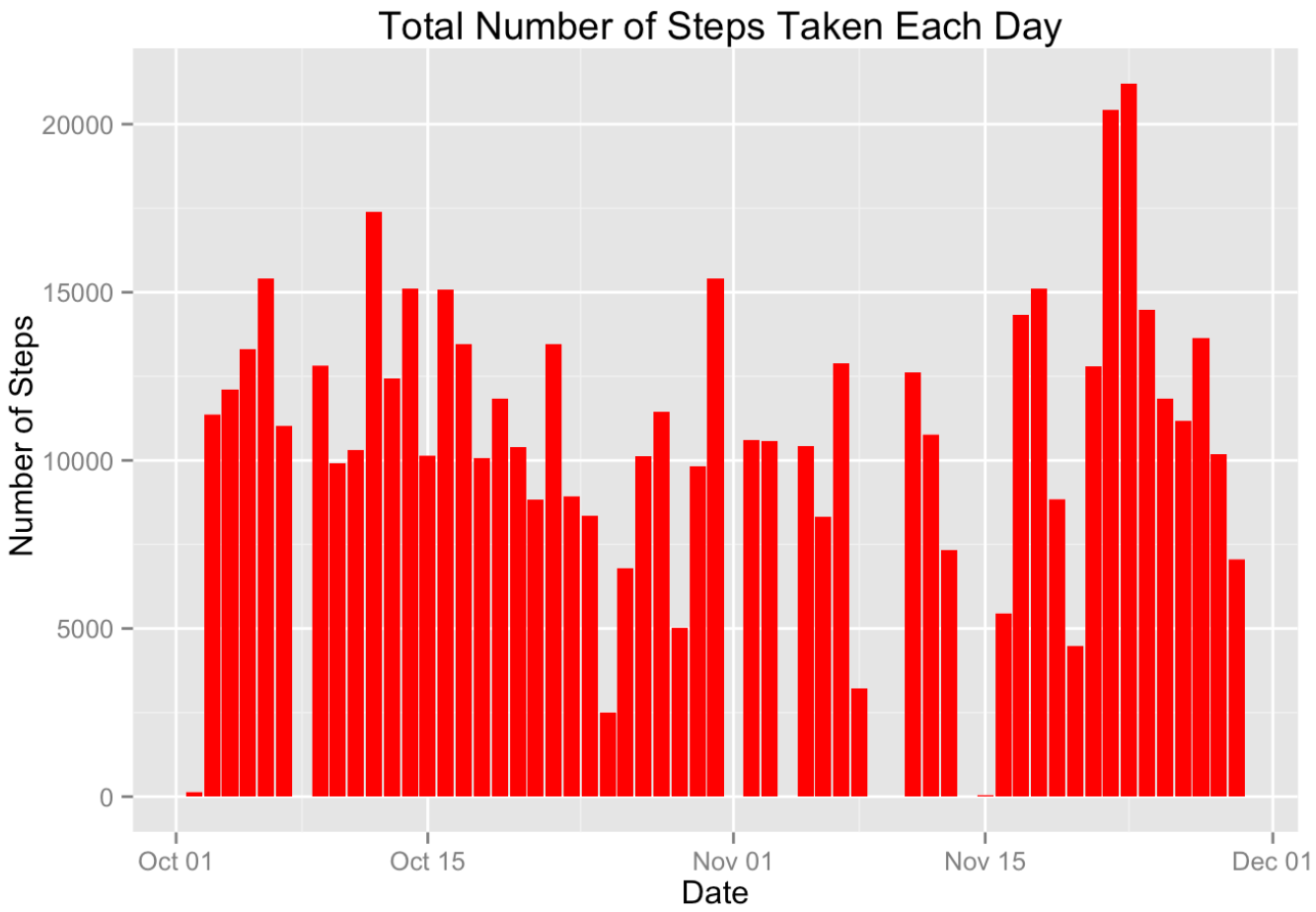
```
data <- read.table('activity.csv',sep = ",",header = TRUE, na.strings ="NA",
                   colClasses = c('integer','Date','factor'))
```

To clear rows with NA

```
new.data <- na.omit(data)
```

Plot of total number of steps taken per day

```
total.steps <- tapply(new.data$steps, new.data$date, FUN = sum)
plot1 <- ggplot(new.data, aes(date, steps)) + geom_bar(stat = "identity",fil
l='red',binwidth = .5) +
        labs(title = " Total Number of Steps Taken Each Day",x = "Date", y = "
Number of Steps")
print(plot1)
```



Mean and Median of the total numberof steps taken per day

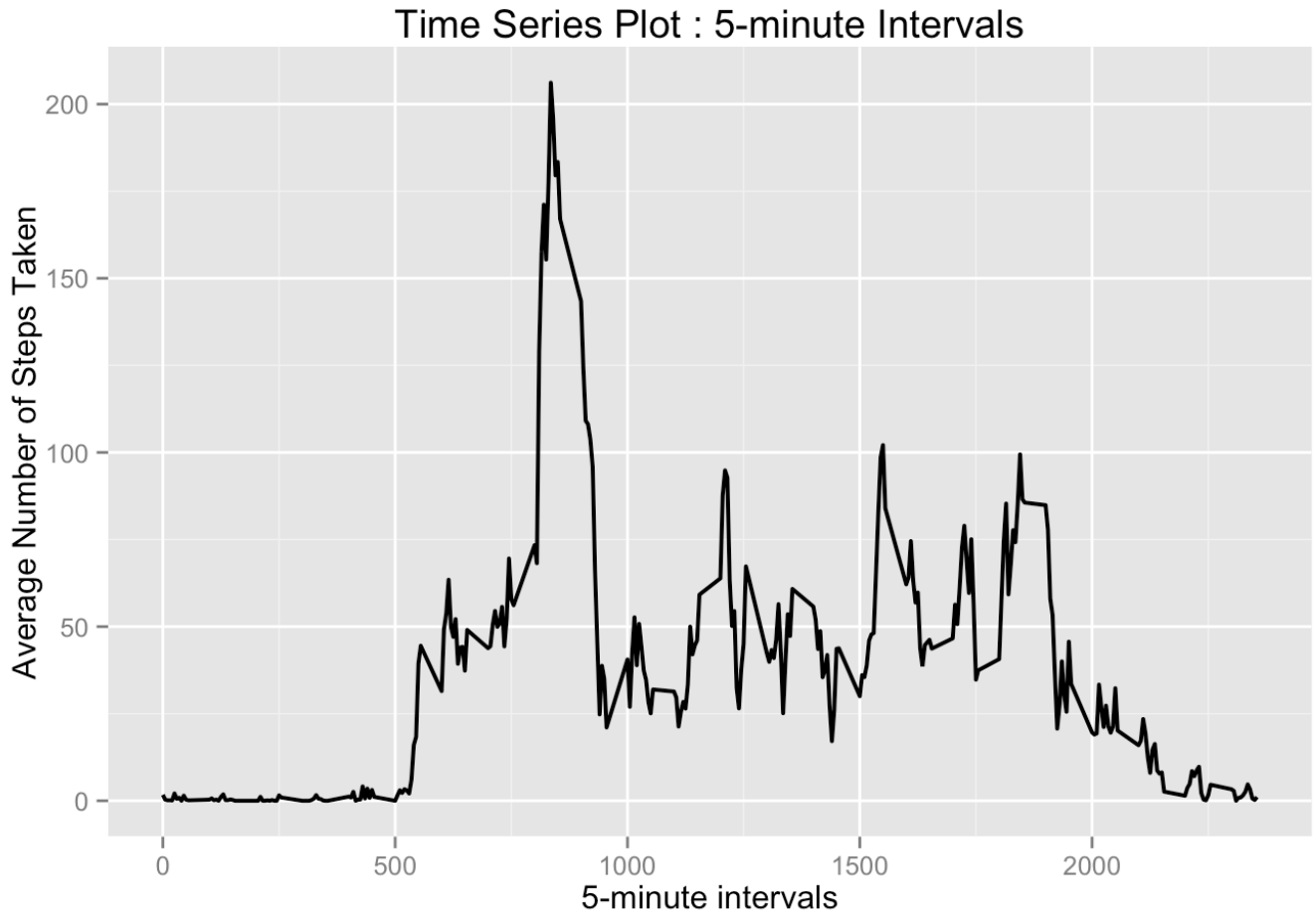
```
summary(total.steps)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      41    8841   10760   10770   13290   21190
```

Average daily actitivy pattern plot

```
averages <- aggregate(new.data$steps, list(interval = as.numeric(as.character(n
ew.data$interval))), FUN = "mean")
names(averages)[2] <- "Avg.Steps"

plot2 <- ggplot(averages, aes(interval, Avg.Steps)) + geom_line(color = "blac
k", size = 0.7) + labs(title = "Time Series Plot : 5-minute Intervals", x = "5-
minute intervals", y = "Average Number of Steps Taken")
print(plot2)
```



5 minute interval with the highest average number of steps

```
averages[averages$Avg.Steps == max(averages$Avg.Steps),]
```

```
##     interval Avg.Steps
## 104      835  206.1698
```

Interval named 835 with 206 steps at index position 104 is the highest.

Imputing missing values

Total number of incomplete cases

```
sum(!complete.cases(data))
```

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

Imputing missing values with mean for 5 minute interval

```
impData <- data
for (i in 1:nrow(impData)) {
    if (is.na(impData$steps[i])) {
        impData$steps[i] <- averages[which(impData$interval[i] == averages$inte
rval), ]$Avg.Steps
    }
}
```

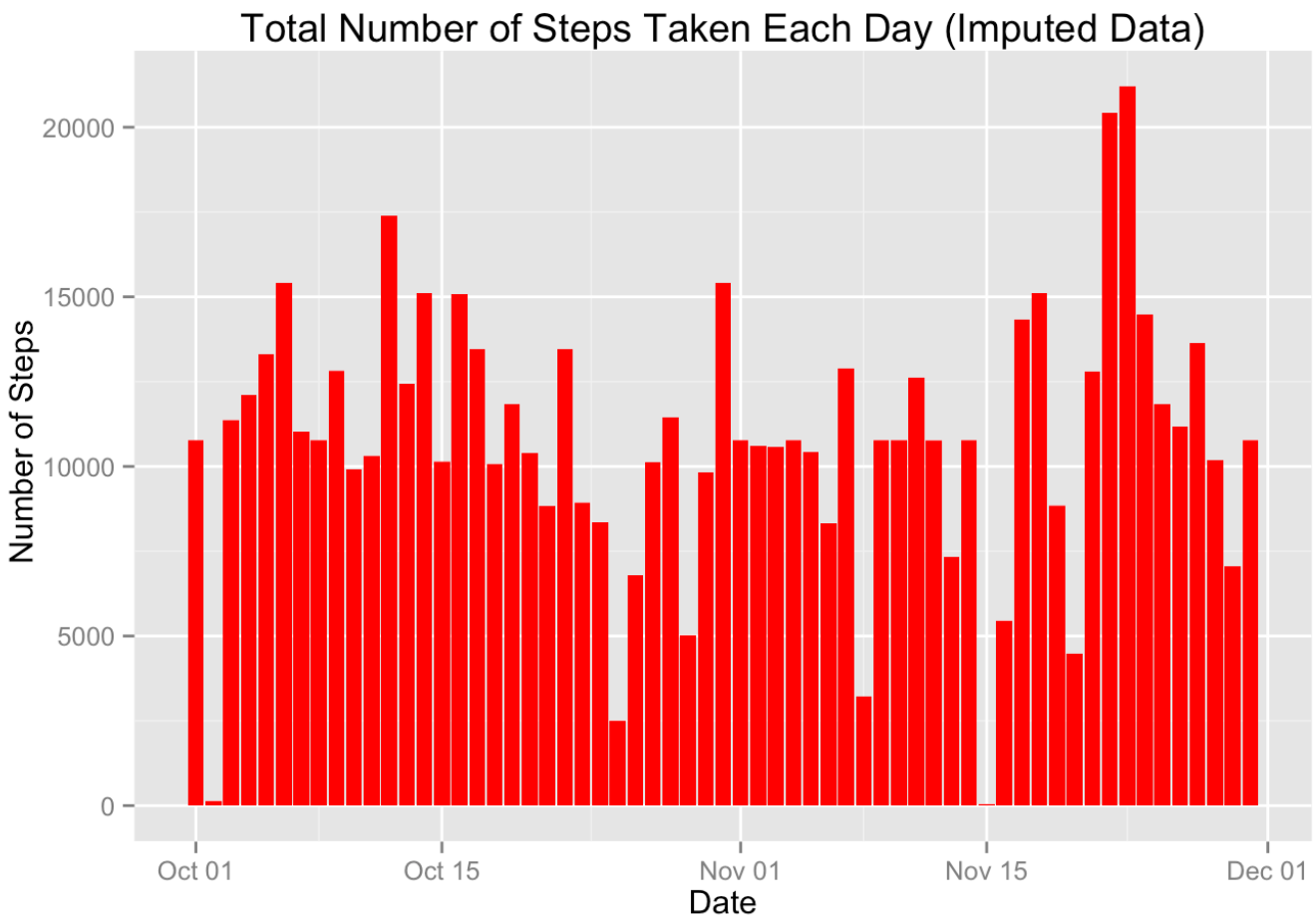
Verification of missing values having been filled in

```
sum(!complete.cases(impData))
```

```
## [1] 0
```

Histogram with imputed data

```
plot3 <- ggplot(impData, aes(date, steps)) + geom_bar(stat = "identity", fil
l="red",binwidth = .5) +
    labs(title = "Total Number of Steps Taken Each Day (Imputed Data)",x = "Dat
e", y = " Number of Steps")
print(plot3)
```



Calculation of mean and median of imputed data

```
totalsteps.impute <- tapply(impData$steps, impData$date, FUN = sum)
summary(totalsteps.impute)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      41    9819   10770   10770   12810   21190
```

It can be noted that the Median has increased in the imputed data , but the Mean remanins the same

Differences in activity patterns between Weekdays and Weekends

Creation of new variable indicating whether a given date is Weekday or Weekend

```
impData$weekdays <- factor(format(impData$date, "%A"))
levels(impData$weekdays)
```

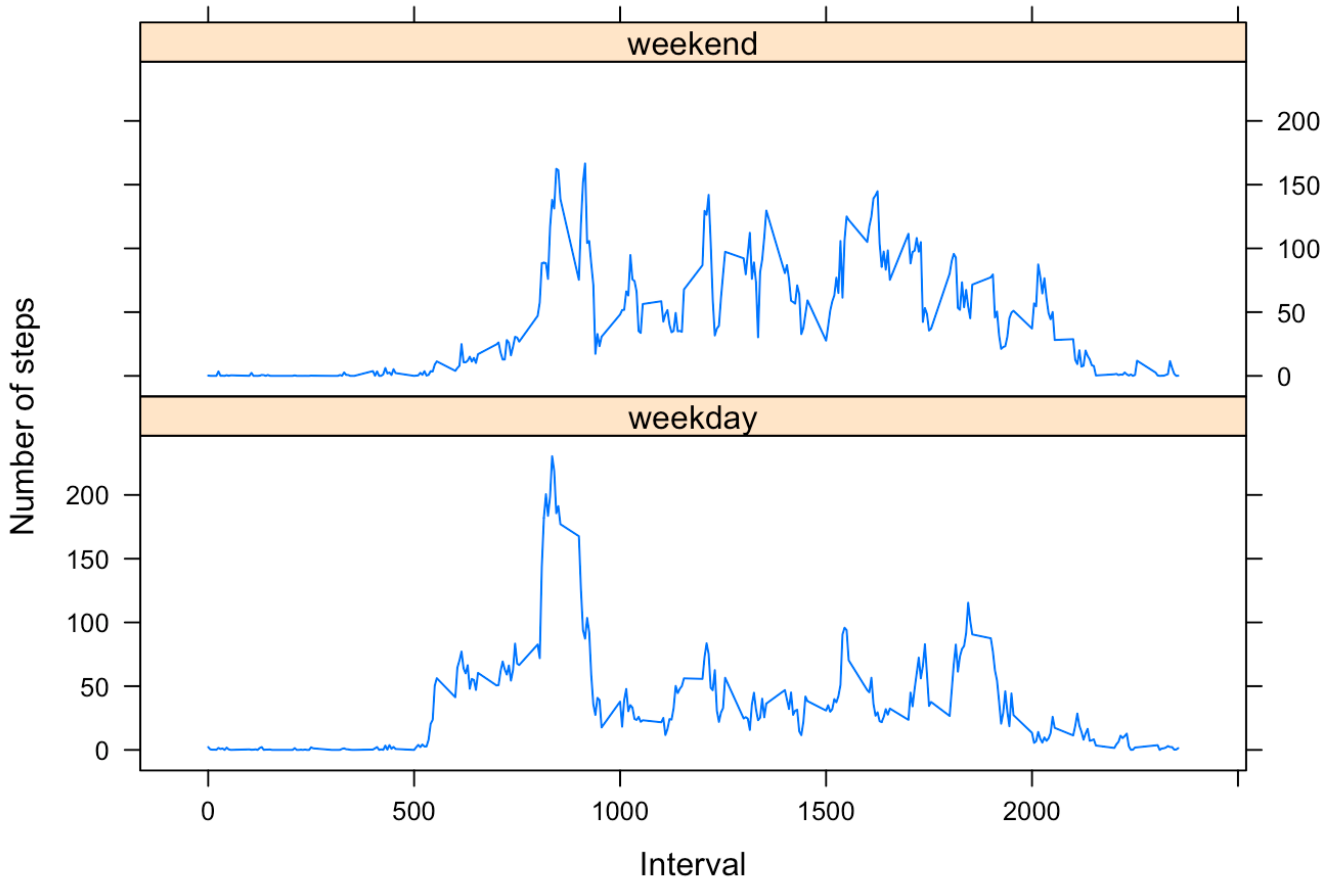
```
## [1] "Friday"    "Monday"    "Saturday"  "Sunday"    "Thursday"  "Tuesday"
## [7] "Wednesday"
```

```
levels(impData$weekdays) <- list(weekday = c("Monday", "Tuesday",
                                             "Wednesday",
                                             "Thursday", "Friday"),
                                 weekend = c("Saturday", "Sunday"))
levels(impData$weekdays)
```

```
## [1] "weekday" "weekend"
```

Graph containing time series plot (i.e. type = “l”) of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis)

```
new.averages <- aggregate(impData$steps,
                          list(interval = as.numeric(as.character(impData$interva
l)),
                               weekdays = impData$weekdays),
                          FUN = "mean")
names(new.averages)[3] <- "meanOfSteps"
library(lattice)
plot4 <- xyplot(new.averages$meanOfSteps ~ new.averages$interval | new.average
s$weekdays,
       layout = c(1, 2), type = "l",
       xlab = "Interval", ylab = "Number of steps")
print(plot4)
```



Difference between Weekdays and Weekends

```
table(impData$weekdays)
```

```
##
## weekday weekend
##   12960    4608
```