Reproducible Research-Peer Assessment 1

Albion Dervishi

May 16, 2015

This is project from the Reproducible Research course in Coursera Data Science specialization track. Using the file"activity" and analyse it as based in the project purpose.

Loading and preprocessing the data

```
Data<-read.csv("~/Desktop/Reproducible Research/activity.csv", header=T)
Data<-data.frame(Data)
Data$date <- as.Date(Data$date)
library(reshape2);library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
##
## Attaching package: 'ggplot2'
##
## The following object is masked _by_ '.GlobalEnv':
##
##     mpg
```

Mean total number of steps taken per day Next figure shows-histogram of the total number of steps taken each day

```
stepsd <- aggregate(steps ~ date, Data, sum)
ggplot(stepsd, aes(x=date , y=steps , fill=steps)) + geom_bar(stat = "identity", position="identity", c
```
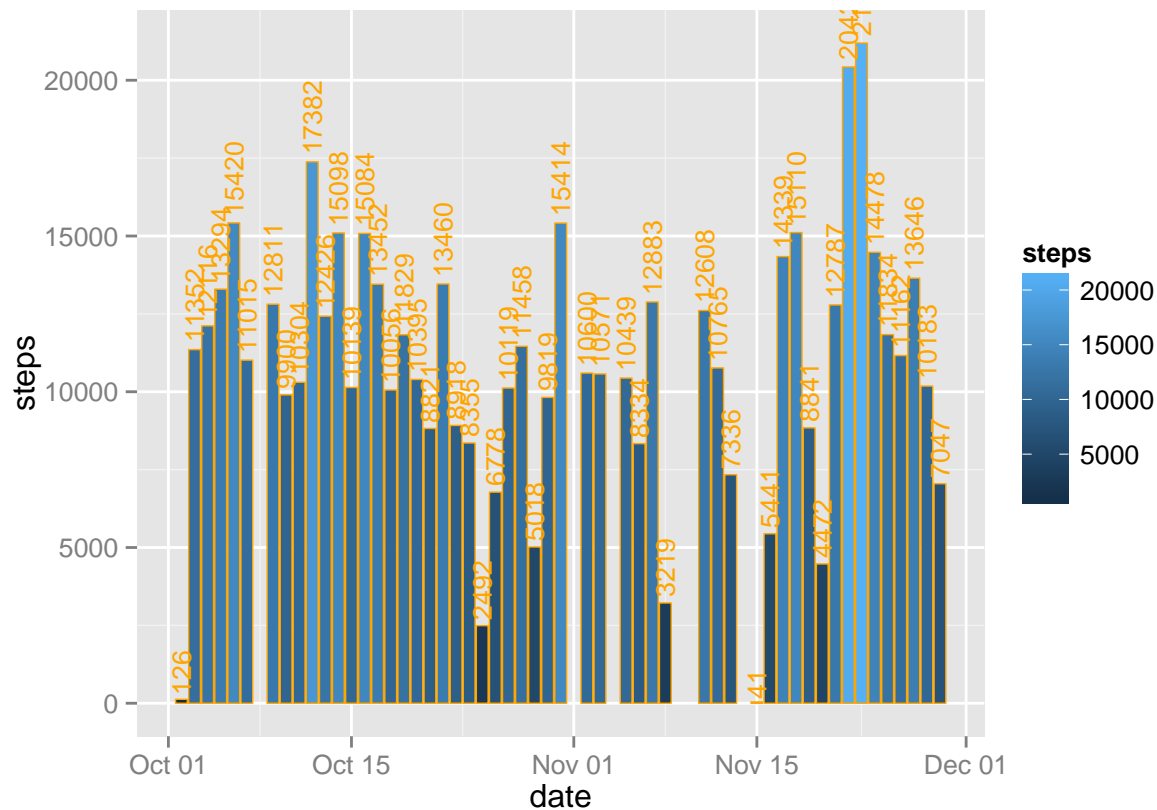
Calculate and report the mean and median of the total number of steps taken per day

```
mean_steps<-paste("Mean Steps per Day =", mean(stepsd$steps, na.rm=TRUE))
mean_steps
```

```
## [1] "Mean Steps per Day = 10766.1886792453"
```

```
median_steps<-paste("Median Steps per Day =", median(stepsd$steps, na.rm=TRUE))
median_steps
```

```
## [1] "Median Steps per Day = 10765"
```

What is the average daily activity pattern? Next figure-shows-average daily activity pattern

```
stepsi <- aggregate(steps ~ interval, Data, mean)
ggplot(stepsi, aes(x=interval , y=steps )) +geom_area(fill="blue", alpha=.2) + geom_line() + xlim(-2, 25
```
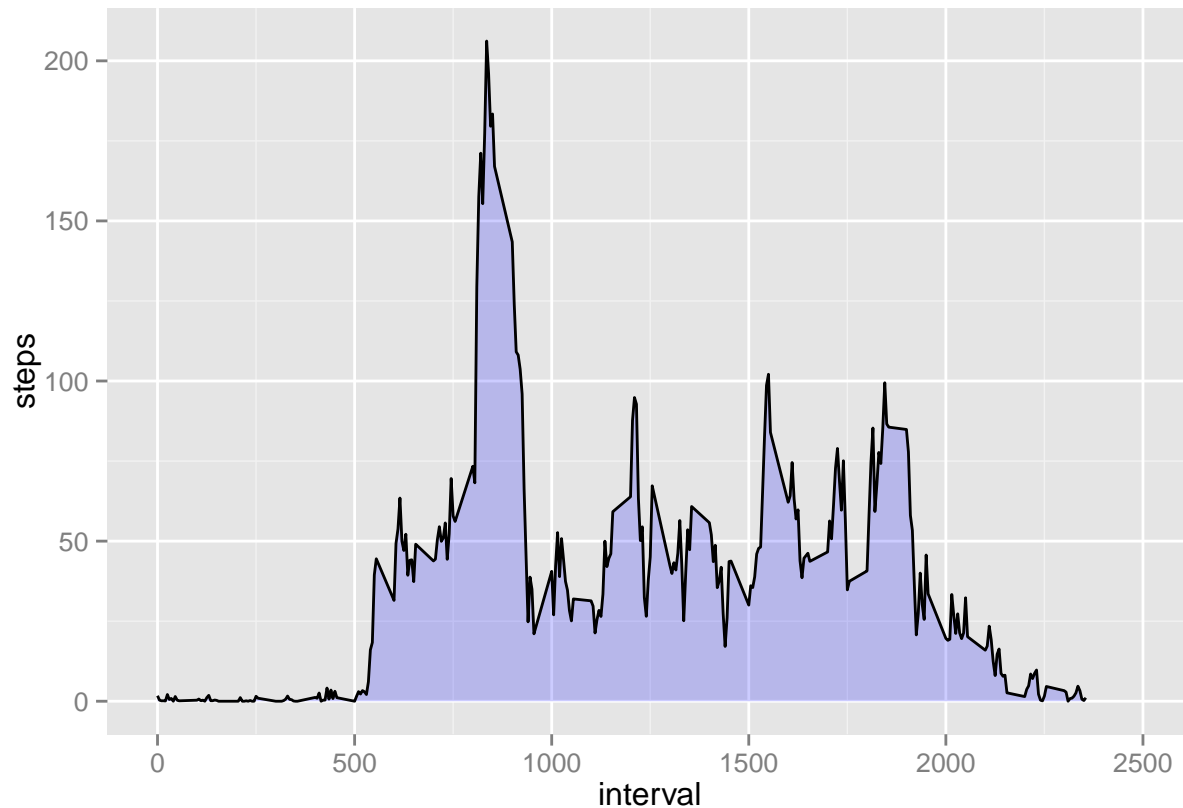
Maximum number of steps in 5-minute interval for average across all the days in the data.

```
max<-subset(stepsi, steps==max(stepsi$steps))
max
```

```
##     interval    steps
## 104      835 206.1698
```

Imputing missing values

```
incomp <- sum (is.na(Data$steps))
incomp
```

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

Devise a strategy for filling in all of the missing values in the dataset.

```
meansteps<-mean(stepsi$steps)
Data$steps[is.na(Data$steps)] <- meansteps
```

Create a new dataset that is equal to the original dataset but with the missing data filled in. Next figure-shows filled and non filled "NA"" in the project

```
data_filled <- transform(Data, (is.na(Data$steps)==meansteps))
data_filled[as.character(data_filled$date) == "2012-10-01", 1] <- 0
stepsdi<- aggregate(steps ~ date, data_filled, sum)
```

```r
hist(stepsdi$steps, main = paste("Number of Steps Each Day"), col="green", xlab="Number of Steps")
hist(stepsd$steps, main = paste(" Number of Steps Each Day"), col="yellow", xlab="Number of Steps", add=
legend("topright", c("Filled", "Non-Filled"), col=c("green", "yellow"), lwd=8)
```

Number of Steps Each Day
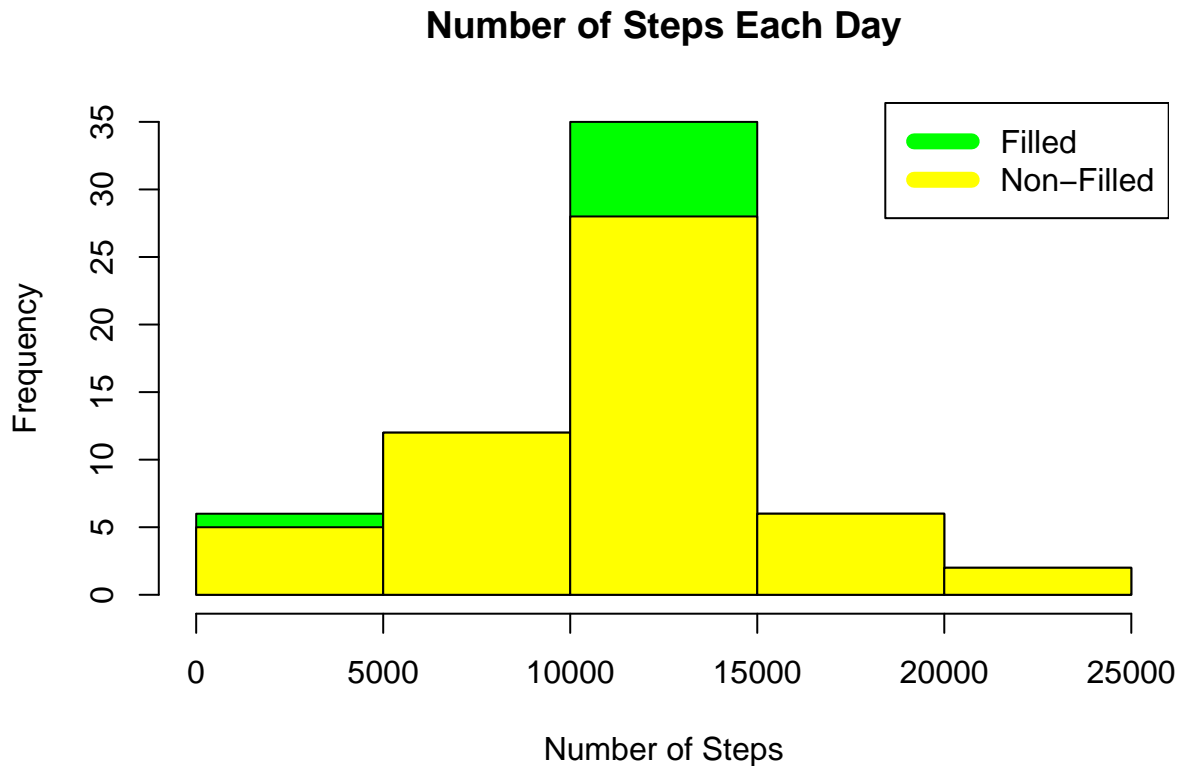
Are there differences in activity patterns between weekdays and weekends? Next figure-shows activity patterns between weekdays and weekends

```r
weekday_weekend <- function(date) {
  day <- weekdays(date)
  if (day %in% c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"))
    return("weekday")
  else if (day %in% c("Saturday", "Sunday"))
    return("weekend")
  else
    stop("invalid date")
}
Data$date <- as.Date(Data$date)
Data$day <- sapply(Data$date, FUN=weekday_weekend)
Data$day<- factor(Data$day)
Mean_week <- aggregate(steps ~ interval + day, data=Data, mean)


ggplot(Mean_week, aes(interval, steps, fill=factor(day))) + geom_line()+ geom_area(colour="black", size=
  scale_fill_brewer(palette="Blues", breaks=rev(levels(Mean_week$day)))
```