Assignment 1

TD

Sunday, February 08, 2015

Loading and reprocessing data

```r
setwd("E:\\Coursera\\Reproducible Research\\assignment1")
data<-read.csv(file="activity.csv", header = TRUE, sep = ",")

data$date<-as.Date(data$date)
dataComplete<-na.omit(data)
```

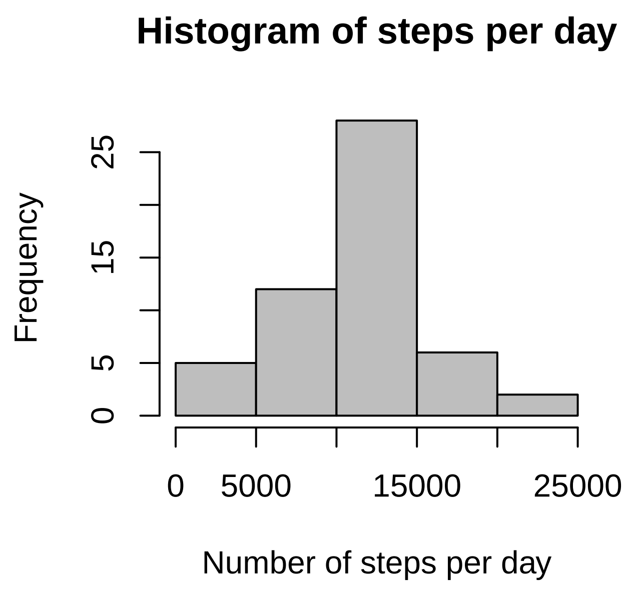
Mean total number of steps taken per day

For this part of the assignment, we ignore the missing values in the dataset.

1. Calculate the total number of steps taken per day
2. Make a histogram of the total number of steps taken each day
3. Calculate and report the mean and median of the total number of steps taken per day

```r
steps_per_day<-aggregate( steps~date,data=dataComplete, sum)
```

```r
plot<-hist(steps_per_day$steps, main = "Histogram of steps per day",
        xlab = "Number of steps per day", ylab = "Frequency",
        col = "grey", breaks = 7)
```



```r
mean(steps_per_day$steps)
```

```
## [1] 10766.19
```

```r
median(steps_per_day$steps)
```

```
## [1] 10765
```

The average daily activity pattern

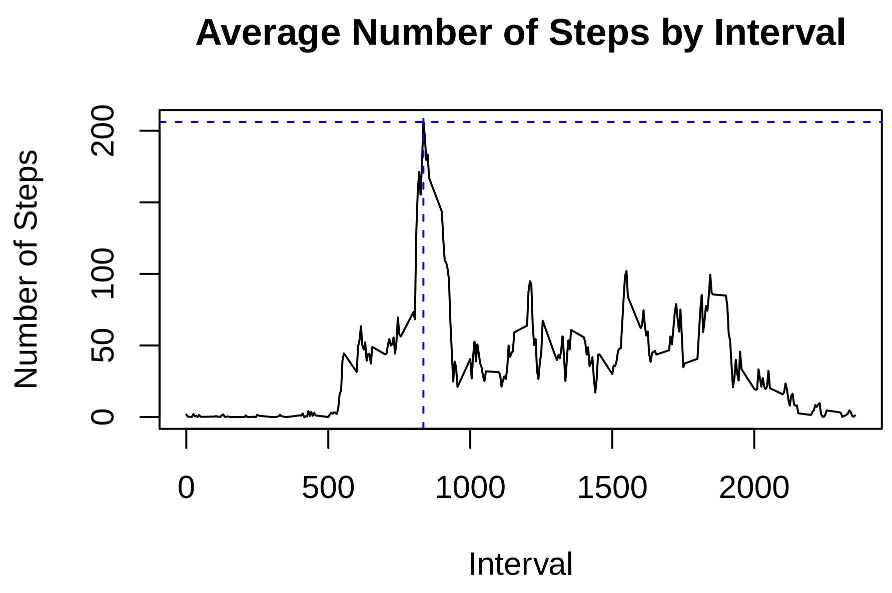
1. Make a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)
2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```r
steps_by_interval <- aggregate(steps ~ interval, dataComplete, mean)

max<-steps_by_interval[which.max(steps_by_interval$steps),]
max
```

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

```r
plot(steps_by_interval$interval,steps_by_interval$steps, type="l",
        main = "Average Number of Steps by Interval",
        xlab = "Interval", ylab="Number of Steps")
abline(v = max$interval, h = max$steps, untf = TRUE, col="blue", lty=2)
```



Imputing missing value

1. Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)
2. Here, we use the mean for that 5-minute interval for imputing the missing values in the corresponding interval
3. Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day; And comprare to the first part of the assignment

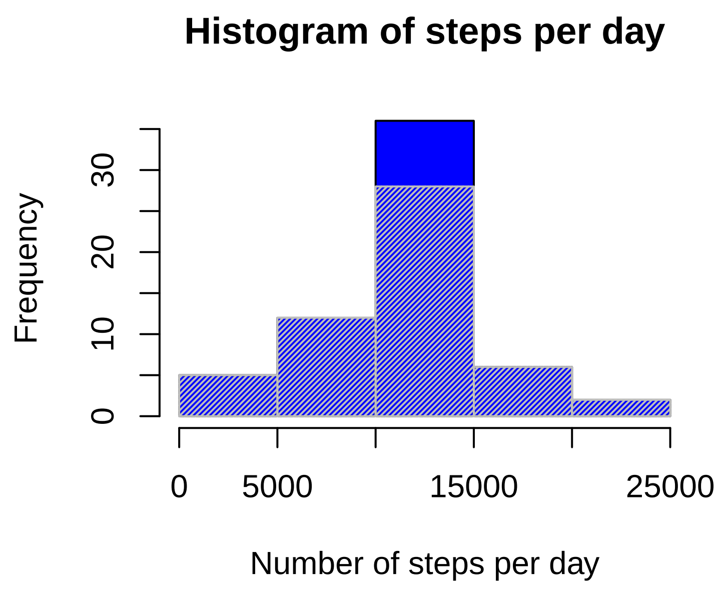
```
sapply(data, function(x) sum(is.na(x)))
```

```
##    steps     date interval 
##     2304        0        0
```

```
for (i in 1:nrow(data)){
        if (is.na(data$steps[i])){
            interval_val <- data$interval[i]
            row_id <- which(steps_by_interval$interval == interval_val)
            steps_val <- steps_by_interval$steps[row_id]
            data$steps[i] <- steps_val
        }
}

steps_per_day_IM<-aggregate(steps~date, data=data, sum)
```

```
# histogram after imputing
hist(steps_per_day_IM$steps , breaks = 7, main = "Histogram of steps per day",
        xlab = "Number of steps per day", ylab = "Frequency", col = "blue")
# histogram before imputing (part1)
hist(steps_per_day$steps, main = paste("Total Steps Each Day"),
        col="grey",  density=50, xlab="Number of Steps", add=T)
```



```r
mean(steps_per_day_IM$steps)
```

```
## [1] 10766.19
```

```r
median(steps_per_day_IM$steps)
```

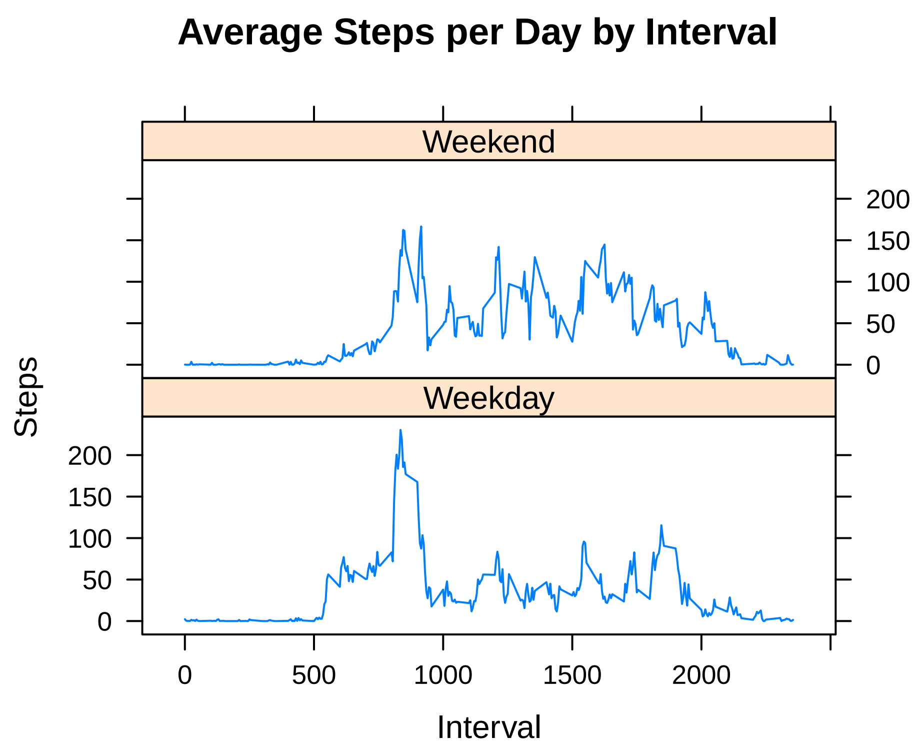
```
## [1] 10766.19
```

Differences in activity patterns between weekdays and weekends

1. Create a new factor variable in the dataset with two levels - "weekday" and "weekend" indicating whether a given date is a weekday or weekend day.
2. Make a panel plot containing a 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).

```r
weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")
data$day = as.factor(ifelse(is.element(weekdays(as.Date(data$date)),weekdays),
                            "Weekday", "Weekend"))
steps_IM <- aggregate(steps ~ interval + day, data, mean)
```

```r
library(lattice)
xyplot(steps_IM$steps ~ steps_IM$interval|steps_IM$day,
        main="Average Steps per Day by Interval",
        xlab="Interval", ylab="Steps",layout=c(1,2), type="l")
```


Average Steps per Day by Interval
Weekend
Weekday
Steps
Interval