Reproducible Research: Peer Assessment 1

```
require(data.table)
```

```
## Loading required package: data.table
```

```
## Warning: package 'data.table' was built under R version 3.1.3
```

```
require(dplyr)
```

```
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:data.table':
##
##     between, last
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Loading and preprocessing the data

What is mean total number of steps taken per day?

1. Calculate the total number of steps taken per day:

```
spd <- na.omit(data) %>% group_by(date) %>% summarise(n=sum(steps))
spd
```

```
## Source: local data table [53 x 2]
##
##          date     n
## 1   2012-10-02   126
## 2   2012-10-03 11352
## 3   2012-10-04 12116
## 4   2012-10-05 13294
## 5   2012-10-06 15420
## 6   2012-10-07 11015
## 7   2012-10-09 12811
## 8   2012-10-10  9900
## 9   2012-10-11 10304
## 10  2012-10-12 17382
## ..         ...   ...
```
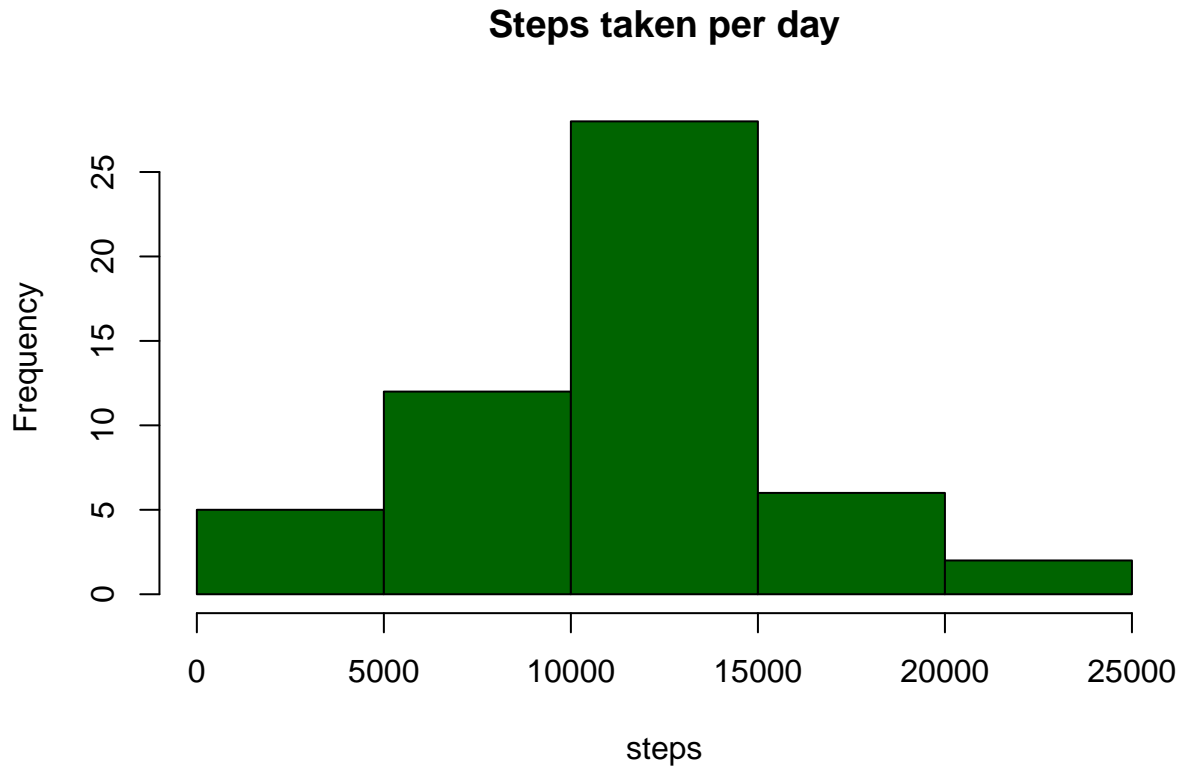
2. Make a histogram of the total number of steps taken each day

```r
hist(spd$n,main="Steps taken per day",xlab="steps",col="darkgreen")
```

Steps taken per day

3. Calculate and report the mean and median of the total number of steps taken per day

```r
mean_spd <- mean(spd$n,na.rm=TRUE)
median_spd <- median(spd$n,na.rm=TRUE)
options(scipen=1, digits=2)
```

The mean of the total number of steps taken per day is **10766.19**.

The median of the total number of steps taken per day is **10765**.

What is the average daily activity pattern?

1. Make a time series plot (i.e. type = "l") of the 5-minute interval (xaxis) and the average number of steps taken, averaged across all days (yaxis)

2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

Imputing missing values

1. Calculate and report the total number of missing values in the dataset (i.e. the total number of rows with NAs)

2. Devise a strategy for filling in all of the missing values in the dataset. The strategy does not need to be sophisticated. For example, you could use the mean/median for that day, or the mean for that 5-minute interval, etc.

3. Create a new dataset that is equal to the original dataset but with the missing data filled in.

4. 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. Do these values differ from the estimates from the first part of the assignment? What is the impact of imputing missing data on the estimates of the total daily number of steps?

Are there 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 (xaxis) and the average number of steps taken, averaged across all weekday days or weekend days (yaxis).