Reproducible Research: Peer Assessment 1

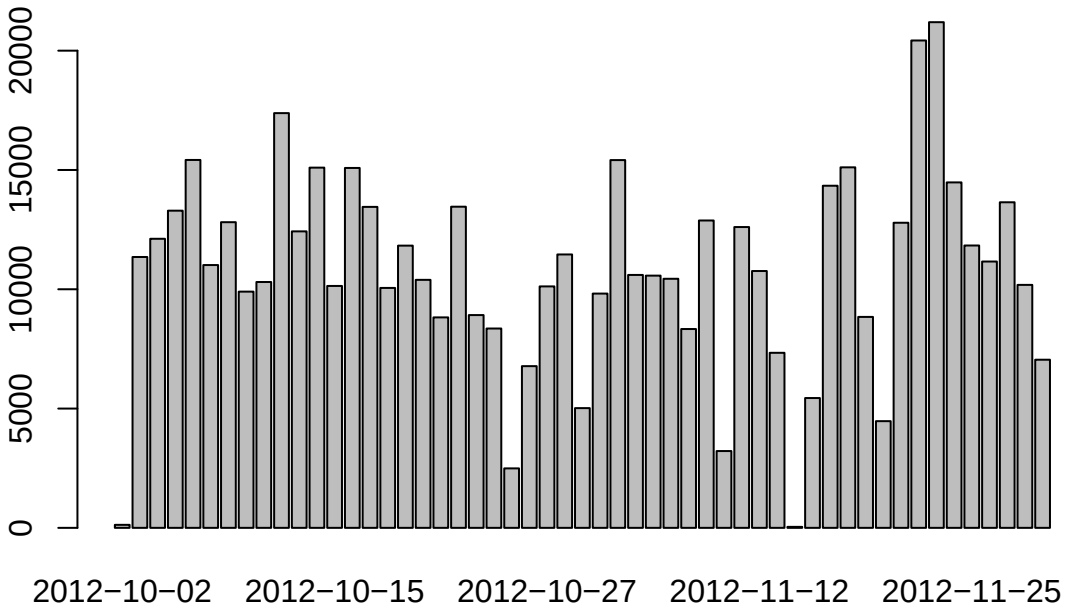
Loading and preprocessing the data

```
library(base)
step_data <- read.csv("activity.csv")
```

What is mean total number of steps taken per day?

A histogram of the number of steps per day is given below, and using the following code:

```
steps_5min <- step_data$steps[
    !is.na(step_data$steps)
]
day_factor <- as.factor(
    step_data$date[
        !is.na(step_data$steps)
    ]
)
steps_per_day <- tapply(steps_5min, day_factor, sum)
barplot(steps_per_day[!is.na(steps_per_day)])
```



The mean of the total number of steps per day is given here:

```
mean(steps_per_day[!is.na(steps_per_day)])
```

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

The median is given below:

```
median(steps_per_day[!is.na(steps_per_day)])
```

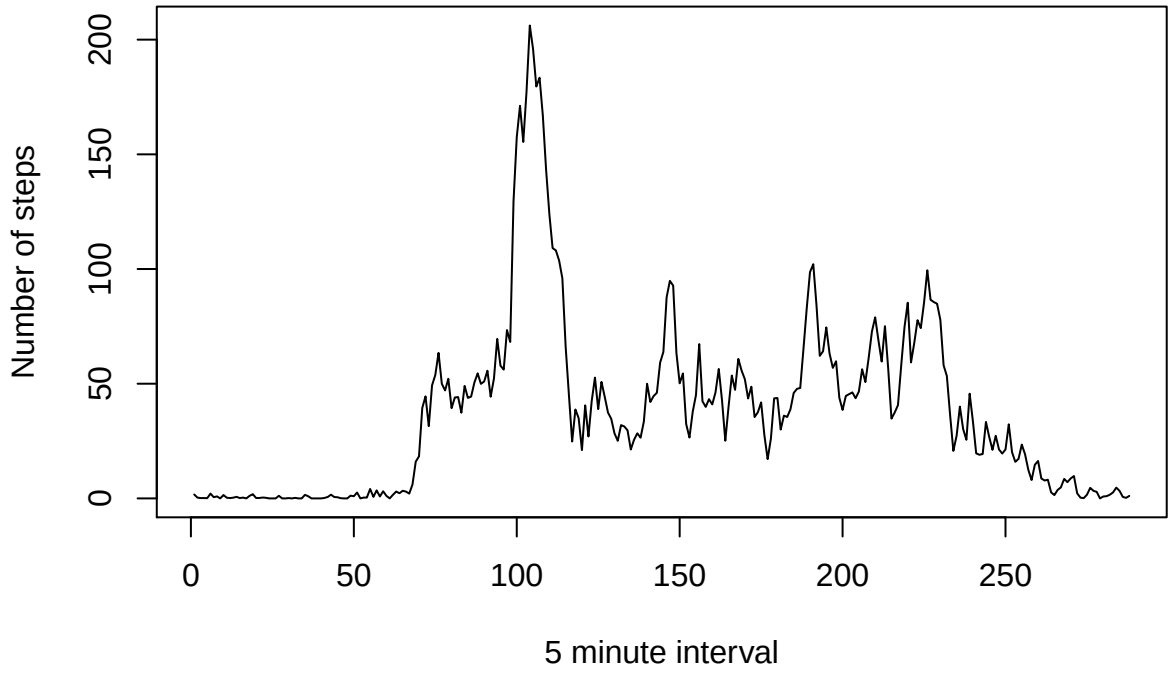
```
## [1] 10765
```

What is the average daily activity pattern?

The following is the source code and plot for the number of steps per five minute interval averaged over all days:

```
intrvl_factor <- as.factor(
    step_data$interval[
        !is.na(step_data$steps)
    ]
)
steps_per_int <- tapply(steps_5min, intrvl_factor, mean)

plot(steps_per_int,
     type="l",
     ylab="Number of steps",
     xlab="5 minute interval"
   )
```



The maximum number of steps is located at 5 minute interval label given below:

```
which.max(steps_per_int)[[1]]
```

```
## [1] 104
```

Imputing missing values

The number of missing values is given by:

```
sum(is.na(step_data$steps))
```

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

All missing data will be imputed by substitution, at the appropriate interval value, of the number of steps per interval averaged over all days with the following code:

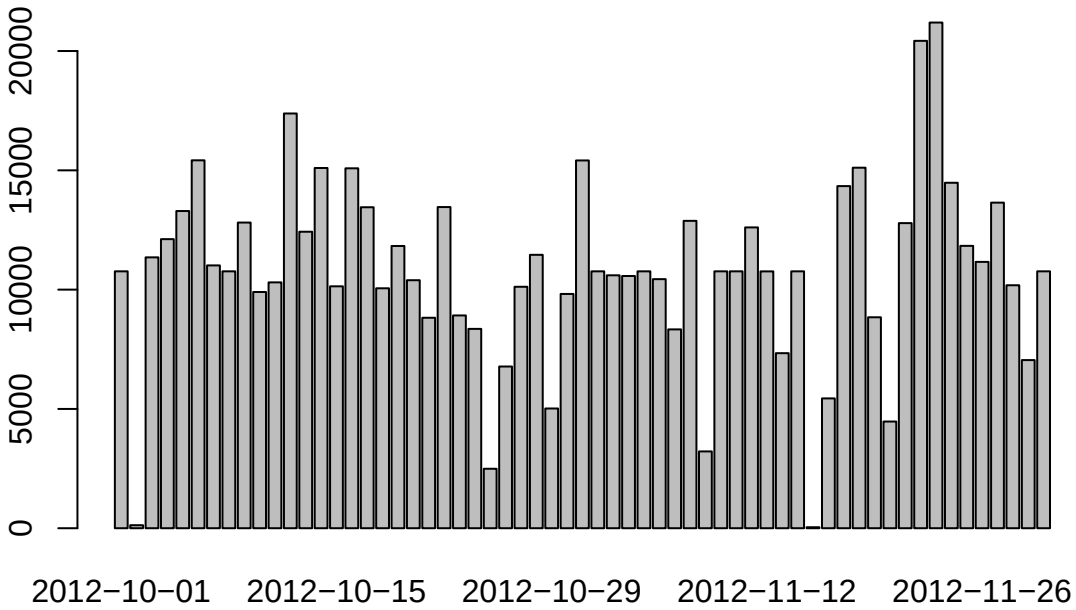
```
step_data_noNA <- step_data
for(i in seq(1,length(steps_per_int))) {
    step_data_noNA[is.na(step_data$steps)
                &
                as.character(step_data_noNA$interval)
                    == row.names(as.matrix(steps_per_int))[i]
               ,1] <- as.matrix(steps_per_int)[i]
}

sum(is.na(step_data_noNA$steps))
```

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

A histogram of the number of steps per day is given below, and using the following code:

```
day_factor_noNA <- as.factor(step_data_noNA$date)
steps_per_day_noNA <- tapply(step_data_noNA$steps, day_factor_noNA, sum)
barplot(steps_per_day_noNA)
```



The mean of the total number of steps per day is given here:

```
mean(steps_per_day_noNA)
```

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

The median is given below:

```
median(steps_per_day_noNA)
```

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

Are there differences in activity patterns between weekdays and weekends?