Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

Data were loaded using the data.table framework into data table act using the fast fread command.

```r
library (data.table)
```

```
## Warning: package 'data.table' was built under R version 3.2.3
```

```r
library (ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.3
```

```r
# file download
setwd("C:/Coursera/Reproducible_Research/RepData_PeerAssessment1")

url <- "https://github.com/kdconway/RepData_PeerAssessment1/raw/master/activity.zip"
if (!file.exists ("./data")) {
    dir.create ("./data")
    download.file (url, "./data/activity.zip")
    unzip ("./data/activity.zip", exdir = "./data")
}

#read in activity data to act dataframe
act <- fread ("./data/activity.csv")
```

Time variables were converted from minimal string format by padding to 4 digits and inserting ":" into the between the hour and minute digits. This was most easily accomplished using the stringr package.

```r
library (stringr)



# deal with times
# add leading zeroes to time until length = 4, insert ":" in middle
act [, interval := str_pad (interval, 4, side="left", pad="0")]
act [, time := as.ITime (gsub ('^([0-9]{2})([0-9]{2})$', '\\1:\\2', interval))]

#create POSIXct dt variable for compound date,time object
act [, date := as.IDate (date)]
act [, dt := as.POSIXct (date, time)]
```
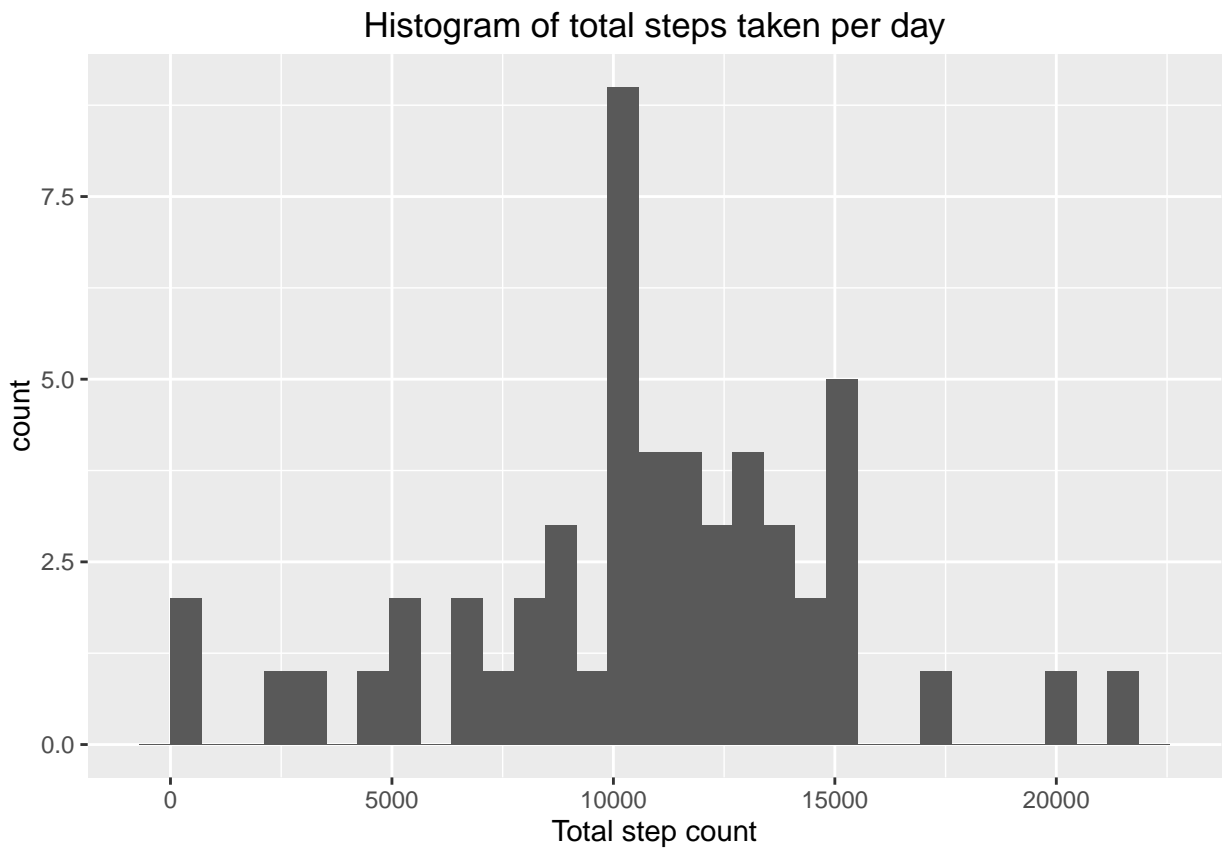
What is mean total number of steps taken per day?

```r
steps.day <- act [, sum(steps), by=date]

qplot (V1, data=steps.day, main = "Histogram of total steps taken per day",
       xlab = "Total step count")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 8 rows containing non-finite values (stat_bin).
```

Histogram of total steps taken per day


```
steps.mean <- steps.day [, mean(V1, na.rm=TRUE)]
steps.median <- steps.day [, median(V1, na.rm=TRUE)]
```

The mean number of steps taken is r steps.mean.

What is the average daily activity pattern?

Imputing missing values

Are there differences in activity patterns between weekdays and weekends?