Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

This section describes the process to unzip dataset file and read it into a dataframe called "**activity**." The first two lines of code which have been commented out shows the steps to download the dataset if unavailable in the project folder.

```
# file.URL <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
# download.file(file.URL, file.dir, method = "auto")


file.name <- "activity.csv"
file.dir <- "./activity.zip"
unzip(file.dir)
activity <- read.csv(file.name, na.strings = "NA")
file.remove(file.name)
```

Next, we will check the data and perform any pre-processing as required.

```
# check dataframe
str(activity)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : chr  "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
# check for missing values
colSums(is.na(activity))
```

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

```
# convert date from character to date format
activity$date <- as.Date(activity$date, format="%Y-%m-%d")
head(activity)
```

```
##   steps       date interval
## 1    NA 2012-10-01        0
## 2    NA 2012-10-01        5
## 3    NA 2012-10-01       10
## 4    NA 2012-10-01       15
## 5    NA 2012-10-01       20
## 6    NA 2012-10-01       25
```

```r
str(activity)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : Date, format: "2012-10-01" "2012-10-01" ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

What is mean total number of steps taken per day?

What is the average daily activity pattern?

Imputing missing values

Are there differences in activity patterns between weekdays and weekends?