# Reproducible Research: Peer Assessment 1

*Enzo Alda*

*Sunday, December 14, 2014*

**Load libraries**

```
library(ggplot2);
library(psych);
```

```
##
## Attaching package: 'psych'
##
## The following object is masked from 'package:ggplot2':
##
##     %+%
```

**Load and preprocess data**

```
data <- read.csv("activity.csv");
times <- paste(as.character(as.integer(data$interval / 100)), as.character(data$interval %% 100), sep="
# Add proper timestamp column
data$time <- strptime(paste(data$date, times), format = "%Y-%m-%d %H:%M");
# Convert *date* field to proper Date
data$date <- as.Date(data$date);
# Add weekday column
data$weekday <- weekdays(data$date);
# Remove useless interval column
data$interval <- NULL;
```

```
# Check data types
str(data);
```

```
## 'data.frame':    17568 obs. of  4 variables:
##  $ steps  : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date   : Date, format: "2012-10-01" "2012-10-01" ...
##  $ time   : POSIXlt, format: "2012-10-01 00:00:00" "2012-10-01 00:05:00" ...
##  $ weekday: chr  "Monday" "Monday" "Monday" "Monday" ...
```

```
# show summary
summary(data);
```

```
##      steps            date                time
##  Min.   :  0.00   Min.   :2012-10-01   Min.   :2012-10-01 00:00:00
##  1st Qu.:  0.00   1st Qu.:2012-10-16   1st Qu.:2012-10-16 05:58:45
##  Median :  0.00   Median :2012-10-31   Median :2012-10-31 11:57:30
##  Mean   : 37.38   Mean   :2012-10-31   Mean   :2012-10-31 12:23:59
```

```
##   3rd Qu.: 12.00    3rd Qu.:2012-11-15    3rd Qu.:2012-11-15 17:56:15
##   Max.   :806.00    Max.   :2012-11-30    Max.   :2012-11-30 23:55:00
##   NA's   :2304
##     weekday
##   Length:17568
##   Class :character
##   Mode  :character
##
##
##
##
```

**Check mean and median for the total number of steps taken per day**

```r
# What is mean total number of steps taken per day?
# For this part of the assignment, you can ignore the missing values in the dataset.
# Make a histogram of the total number of steps taken each day
# Calculate and report the mean and median total number of steps taken per day

tapply(data$steps, data$weekday, mean, na.rm = T)
```

```
##     Friday     Monday   Saturday     Sunday   Thursday    Tuesday  Wednesday
##   42.91567   34.63492   43.52579   42.63095   28.51649   31.07485   40.94010
```

```r
tapply(data$steps, data$weekday, median, na.rm = T)
```

```
##     Friday     Monday   Saturday     Sunday   Thursday    Tuesday  Wednesday
##          0          0          0          0          0          0          0
```

```r
describeBy(data$steps, data$weekday, na.rm = T)
```

```
## group: Friday
##   vars    n  mean     sd median trimmed mad min max range skew kurtosis
## 1    1 2016 42.92 121.61      0    9.43   0   0 802   802 3.77    14.76
##     se
## 1 2.71
## ------------------------------------------------------------
## group: Monday
##   vars    n  mean     sd median trimmed mad min max range skew kurtosis
## 1    1 2016 34.63 114.85      0    6.34   0   0 789   789 4.55    21.27
##     se
## 1 2.56
## ------------------------------------------------------------
## group: Saturday
##   vars    n  mean     sd median trimmed mad min max range skew kurtosis
## 1    1 2016 43.53 120.33      0   10.89   0   0 785   785 3.74    14.39
##     se
## 1 2.68
## ------------------------------------------------------------
## group: Sunday
```

```
##    vars    n  mean     sd median trimmed mad min max range skew kurtosis
## 1     1 2016 42.63 108.94      0    12.8   0   0 785   785 3.56    13.52
##     se
## 1 2.43
## -----------------------------------------------------------
## group: Thursday
##    vars    n  mean    sd median trimmed mad min max range skew kurtosis
## 1     1 2304 28.52 99.67      0    4.17   0   0 759   759 4.71    23.03
##     se
## 1 2.08
## -----------------------------------------------------------
## group: Tuesday
##    vars    n  mean    sd median trimmed mad min max range skew kurtosis
## 1     1 2592 31.07 99.23      0    6.46   0   0 806   806 4.89    26.86
##     se
## 1 1.95
## -----------------------------------------------------------
## group: Wednesday
##    vars    n  mean     sd median trimmed mad min max range skew kurtosis
## 1     1 2304 40.94 119.82      0    9.05   0   0 766   766 4.02    16.84
##     se
## 1 2.5
```

## What is the average daily activity pattern?

```
# Make a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of
# Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of
```

## Imputing missing values

```
# Note that there are a number of days/intervals where there are missing values (coded as NA). The pres

# Calculate and report the total number of missing values in the dataset (i.e. the total number of rows

# Devise a strategy for filling in all of the missing values in the dataset. The strategy does not need

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

# Make a histogram of the total number of steps taken each day and Calculate and report the mean and me
```

## Are there differences in activity patterns between weekdays and weekends?

```
# For this part the weekdays() function may be of some help here. Use the dataset with the filled-in mi

# Create a new factor variable in the dataset with two levels - "weekday" and "weekend" indicating whet

# Make a panel plot containing a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) a
```