

# Reproducible Research: Peer Assessment 1

Peter A. Edelsbrunner

March 10, 2015

# Reproducible Research: Peer assessment 1

# Introduction

It is now possible to collect a large amount of data about personal movement using activity monitoring devices such as a Fitbit, Nike Fuelband, or Jawbone Up. These type of devices are part of the “quantified self” movement – a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. But these data remain under-utilized both because the raw data are hard to obtain and there is a lack of statistical methods and software for processing and interpreting the data.

This assignment makes use of data from a personal activity monitoring device. This device collects data at 5 minute intervals through out the day. The data consists of two months of data from an anonymous individual collected during the months of October and November, 2012 and include the number of steps taken in 5 minute intervals each day.

# Data

The data for this assignment can be downloaded from the course web site:

- ▶ Dataset: Activity monitoring data [52K]

The variables included in this dataset are:

- ▶ **steps**: Number of steps taking in a 5-minute interval (missing values are coded as NA)
- ▶ **date**: The date on which the measurement was taken in YYYY-MM-DD format
- ▶ **interval**: Identifier for the 5-minute interval in which measurement was taken

The dataset is stored in a comma-separated-value (CSV) file and there are a total of 17,568 observations in this dataset.

## Assignment

This assignment will be described in multiple parts. You will need to write a report that answers the questions detailed below. Ultimately, you will need to complete the entire assignment in a **single R markdown** document that can be processed by **knitr** and be transformed into an HTML file.

Throughout your report make sure you always include the code that you used to generate the output you present. When writing code chunks in the R markdown document, always use `echo = TRUE` so that someone else will be able to read the code. **This assignment will be evaluated via peer assessment so it is essential that your peer evaluators be able to review the code for your analysis.**

For the plotting aspects of this assignment, feel free to use any plotting system in R (i.e., base, lattice, ggplot2)

Fork/clone the GitHub repository created for this assignment. You will submit this assignment by pushing your completed files into your forked repository on GitHub. The assignment submission will

Pre-work: Load packages, set nice APA-conform and neat theme for ggplot2

```
library(stats)
library(lattice)

#Set working directory
setwd("C:/Users/petere/Documents/4IV_Stud/Reproducible rese
```

## Load data

```
#Load data
```

```
ass1data <- read.csv("activity.csv", colClasses = c("integer",  
str(ass1data)
```

```
## 'data.frame':    17568 obs. of  3 variables:
```

```
## $ steps      : int  NA NA NA NA NA NA NA NA NA NA NA ...
```

```
## $ date       : Date, format: "2012-10-01" "2012-10-01" ...
```

```
## $ interval: Factor w/ 288 levels "0","10","100",...: 1 2
```

```
attach(ass1data)
```

What is the mean total number of steps taken per day?

For this part of the assignment, you can ignore the missing values in the dataset.

1. Make a histogram of the total number of steps taken each day

```
#Extract mean steps per day
```