Reproducible Research: Peer Assessment 1

## Loading and preprocessing the data

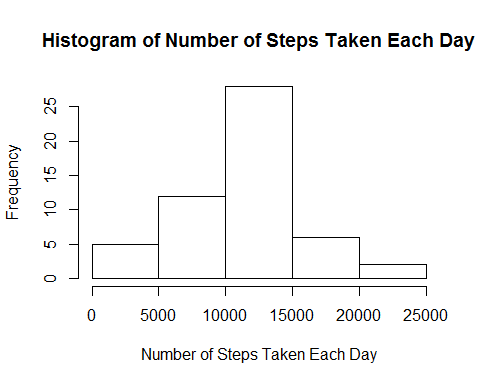
Assumption: The data is in a zipped file called *activity.zip*, as is found in the original repository by Roger Peng.

df <- read.csv(unzip("activity.zip"))  
df\_complete <- df[complete.cases(df), ] # This dataframe does not have NA observations.

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

The below is a histogram showing the number of steps taken each day.

df\_agg\_date <- aggregate(steps ~ date, data = df, FUN = sum)  
hist(df\_agg\_date$steps, xlab = "Number of Steps Taken Each Day", main = "Histogram of Number of Steps Taken Each Day")



The *mean* total number of steps taken per day is given by

mean(df\_agg\_date$steps)

## [1] 10766

The *median* total number of steps taken per day is given by

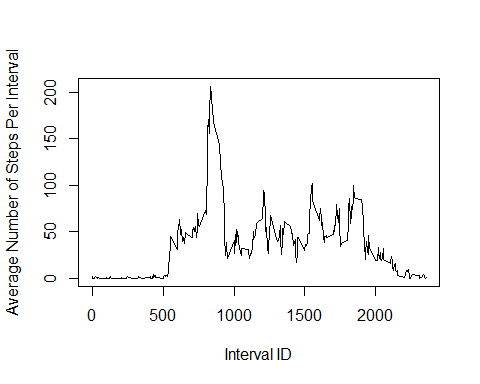
median(df\_agg\_date$steps)

## [1] 10765

## What is the average daily activity pattern?

The below time series plot gives an idea of the average daily activity pattern of this person.

# Get the average steps per interval ID. Missing values (NA) are ignored, ie: the observation with the missing value is omitted entirely.  
df\_agg\_interval <- aggregate(steps ~ interval, data = df, FUN = mean, na.action = na.omit)   
  
plot(df\_agg\_interval$steps ~ df\_agg\_interval$interval , type = "l", xlab = "Interval ID", ylab = "Average Number of Steps Per Interval")



On average across all the days in the dataset, the 5-minute interval which contains the maximum number of steps is: **835**

## Imputing missing values

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