The fitness Movement

R CODE

#Exploratory
head(activity)
str(activity)
hist(activity$steps)

#Histogram of # of steps
hist(activity$steps)

OUTPUT

#Exploratory
3 variables in the dataset consisting of num, Factor and int variables

```
   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

'data.frame':    17568 obs. of  3 variables:
 $ steps   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ date    : Factor w/ 61 levels "2012-10-01","2012-10-02",..: 1 1
1 1 1 1 1 1 1 ...
 $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

#Histogram
Vast majority of 5 minute interval only contain 50 steps

Histogram of activity$steps
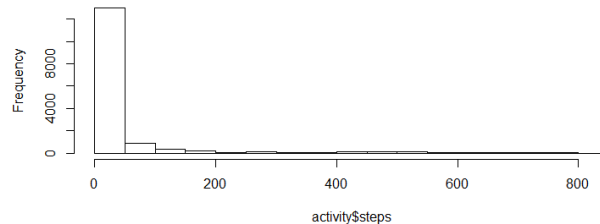
The fitness Movement

R CODE

#Mean and median number of steps taken each day
mean.default(activity$steps, trim = 0, na.rm = TRUE, activity)
median.default(activity$steps, trim = 1, na.rm = TRUE, activity)

#The 5-minute interval that, on average, contains the maximum number of steps
plot(steps ~ interval, activity)

#Code to describe and show a strategy for imputing missing data
#Histogram of the total number of steps taken each day after missing values are imputed

activity$steps[activity$steps== "NA"] <- 37.3826
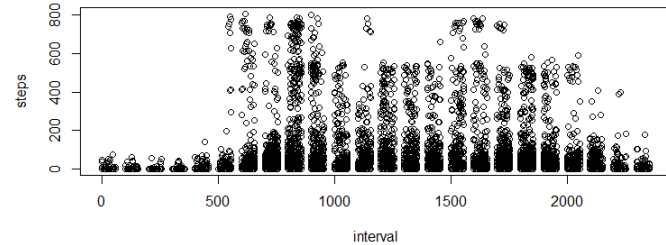plot(steps ~ date, activity)

OUTPUT
#Mean and Median
Mean = 37.3826
Median = 0

#Five minute interval
It is approximately the 900th interval based on the density of the plot

#Plot to depict total number of steps each day
2012-10-11 appears to be the most dense and containing the highest amount of steps

2