Reproducible Research: Peer Assessment 1

## Prepare the R global envirnoment.

library(knitr)  
library(data.table)  
library(ggplot2)  
opts\_chunk$set(echo = TRUE, results = 'hold')

## Loading and preprocessing the data

### Load the required data

The following statement is used to load the data using read.csv().

Note: It is assumed that the file activity.csv is in the current working directory.

data <- read.csv("activity.csv", header = TRUE, sep = ",", colClasses=c("numeric", "character", "numeric"))

### Preprocessing of data (Cleaning) and previewing data

we will first convert the date to date class and interval to factor calss

data$date <- as.Date(data$date, format = "%Y-%m-%d")  
data$interval <- as.factor(data$interval)  
str(data)

## 'data.frame': 17568 obs. of 3 variables:  
## $ steps : num 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","5","10","15",..: 1 2 3 4 5 6 7 8 9 10 ...

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

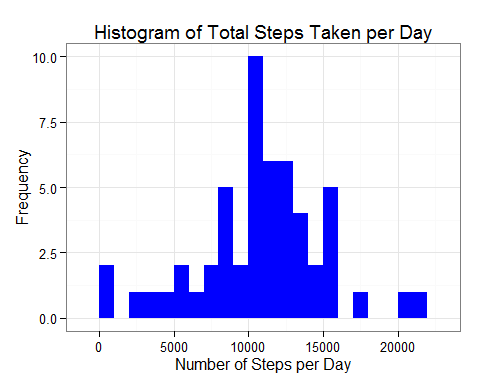
Now we will calculate the total number of steps per day

Total\_steps\_day <- aggregate(steps ~ date, data, sum)  
colnames(Total\_steps\_day) <- c("date","steps")  
head(Total\_steps\_day)

## date steps  
## 1 2012-10-02 126  
## 2 2012-10-03 11352  
## 3 2012-10-04 12116  
## 4 2012-10-05 13294  
## 5 2012-10-06 15420  
## 6 2012-10-07 11015

1). Now plot a histogram of the total steps taken per day

ggplot(Total\_steps\_day, aes(x = steps)) +   
 geom\_histogram(fill = "blue", binwidth = 1000) +   
 labs(title="Histogram of Total Steps Taken per Day",   
 x = "Number of Steps per Day", y = "Frequency") + theme\_bw()



2). Now we will calculate the mean number of steps taked per day

mean <- mean(Total\_steps\_day$steps, na.rm=TRUE)

The mean number of steps taken per day is mean

## What is the average daily activity pattern?

## Imputing missing values

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