

# Project1\_ReproducibleResearch.R

*anaco*

*Sun Aug 21 12:03:59 2016*

```
# Load libraries
```

```
library(RCurl)
```

```
## Loading required package: bitops
```

```
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.2.5
```

```
##  
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':  
##  
##     date
```

```
library(plyr)
```

```
##  
## Attaching package: 'plyr'
```

```
## The following object is masked from 'package:lubridate':  
##  
##     here
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```

library(lattice)

# Section #1: Loading and Processing the Data

#Get Data from Website and unzip file in your current working directory

data <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip'
download.file(data, 'repdata-data-activity.zip')
unzip('repdata-data-activity.zip')
activity <- read.csv("activity.csv", stringsAsFactors=FALSE)

#Change date field and clean the data
activity$day <- weekdays(as.Date(activity$date))
activity$DateTime<- as.POSIXct(activity$date, format="%Y-%m-%d")
clean <- activity[!is.na(activity$steps),]

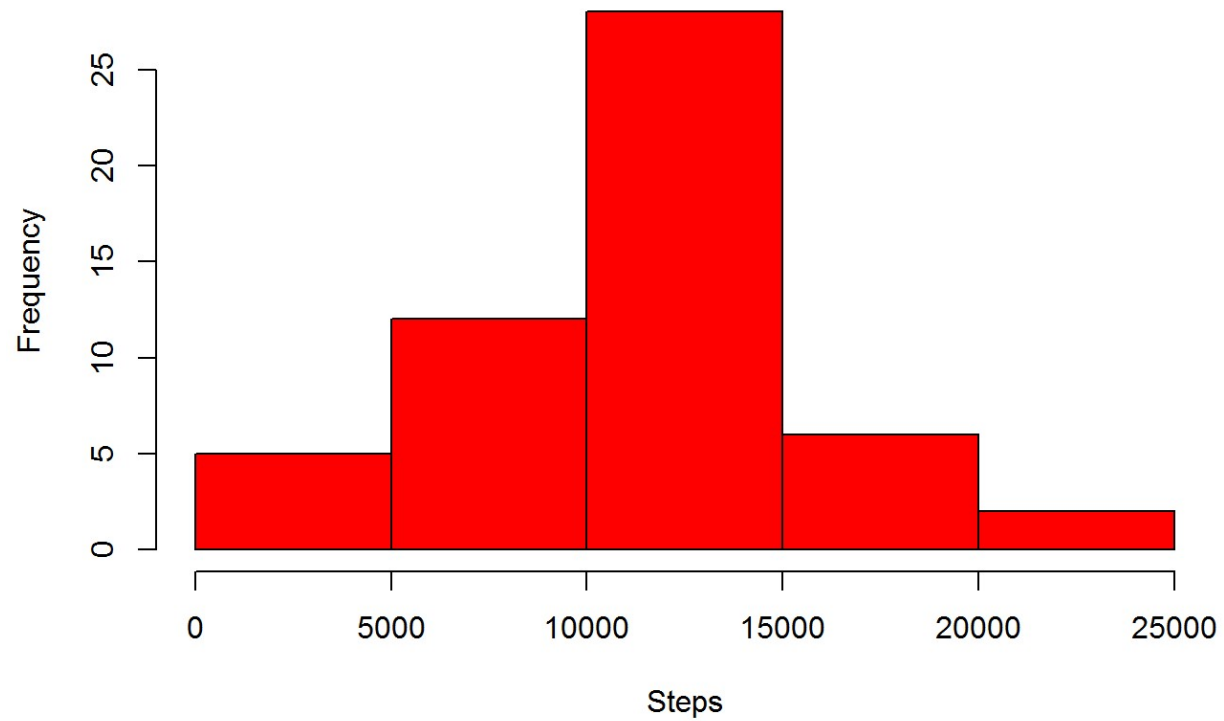
# summary of total steps per date
sumTable <- aggregate(activity$steps ~ activity$date, FUN=sum)
colnames(sumTable)<- c("Date", "Steps")

## Section # 2. What is mean total number of steps taken per day?

# Creating the histogram of total steps per day
hist(sumTable$Steps, breaks=5, xlab="Steps", col = "red", main = "Total Steps p
er Day")

```

## Total Steps per Day



```
# Calculate the mean and the median of the steps per day  
as.integer(mean(sumTable$Steps))
```

```
## [1] 10766
```

```
as.integer(median(sumTable$Steps))
```

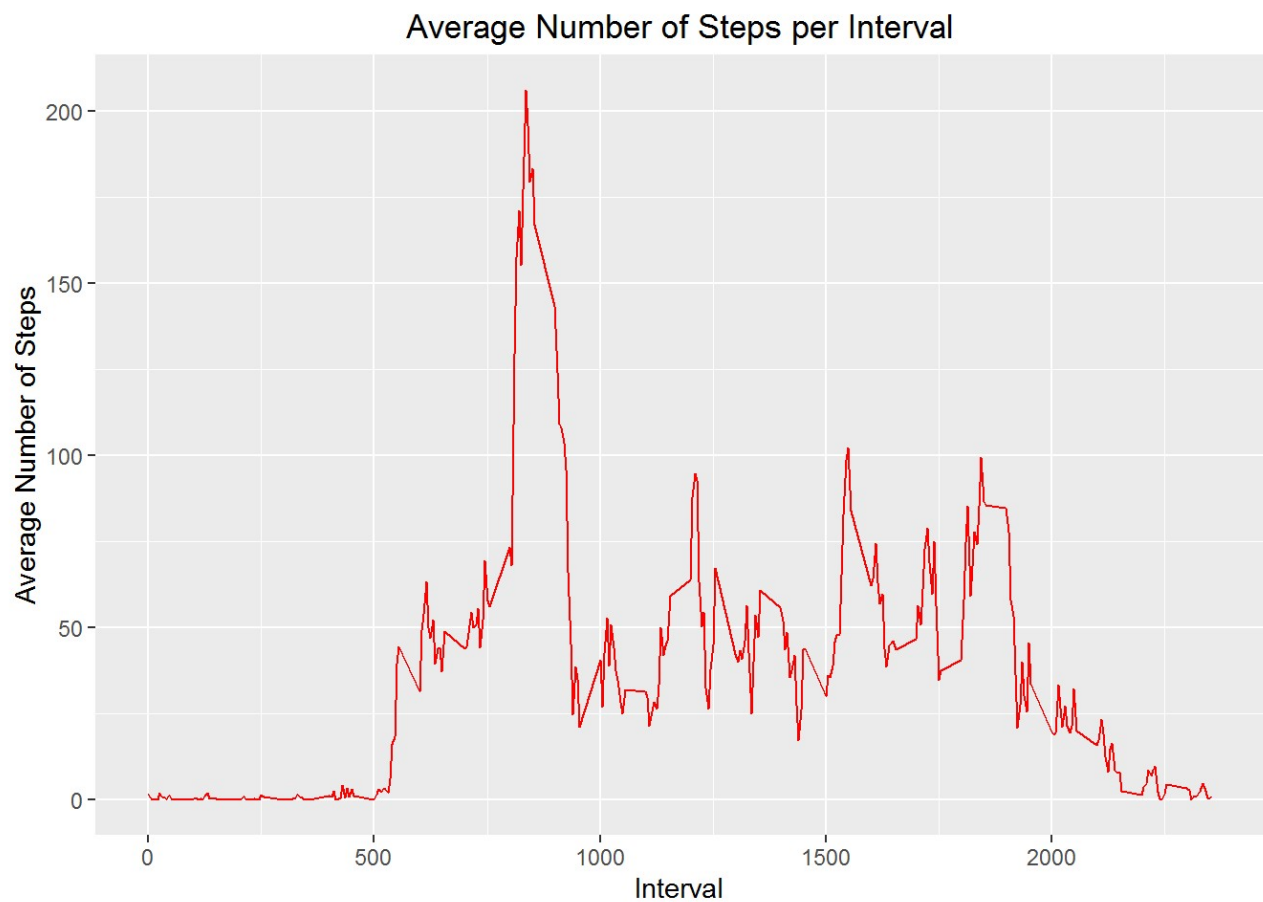
```
## [1] 10765
```

```
## Section # 3. What is the average daily activity pattern?

# create average number of steps per interval
intervalTable <- ddply(clean, .(interval), summarize, Avg = mean(steps))

# create max steps by interval
maxSteps <- max(intervalTable$Avg)

# Create line plot of average number of steps per interval
p <- ggplot(intervalTable, aes(x=interval, y=Avg), xlab = "Interval", ylab="Average Number of Steps")
p + geom_line(colour = 'red') + xlab("Interval") + ylab("Average Number of Steps") +
  ggtitle("Average Number of Steps per Interval")
```



```
# Which interval contains the maximum average number of steps
intervalTable[intervalTable$Avg==maxSteps,1]
```

```
## [1] 835
```

```
# Section # 4. Input Missing Values
```

```
# Process and calculate the number of NAs in original data set  
nrow(activity[is.na(activity$steps),])
```

```
## [1] 2304
```

```
# Create table with the average number of steps per weekday and interval  
avgTable <- ddply(clean, .(interval, day), summarize, Avg = mean(steps))
```

```
# Create a new dataset with all NAs for substitution  
NAdata<- activity[is.na(activity$steps),]
```

```
# Merge NA data with average weekday interval for substitution  
newdataset<-merge(NAdata, avgTable, by=c("interval", "day"))
```

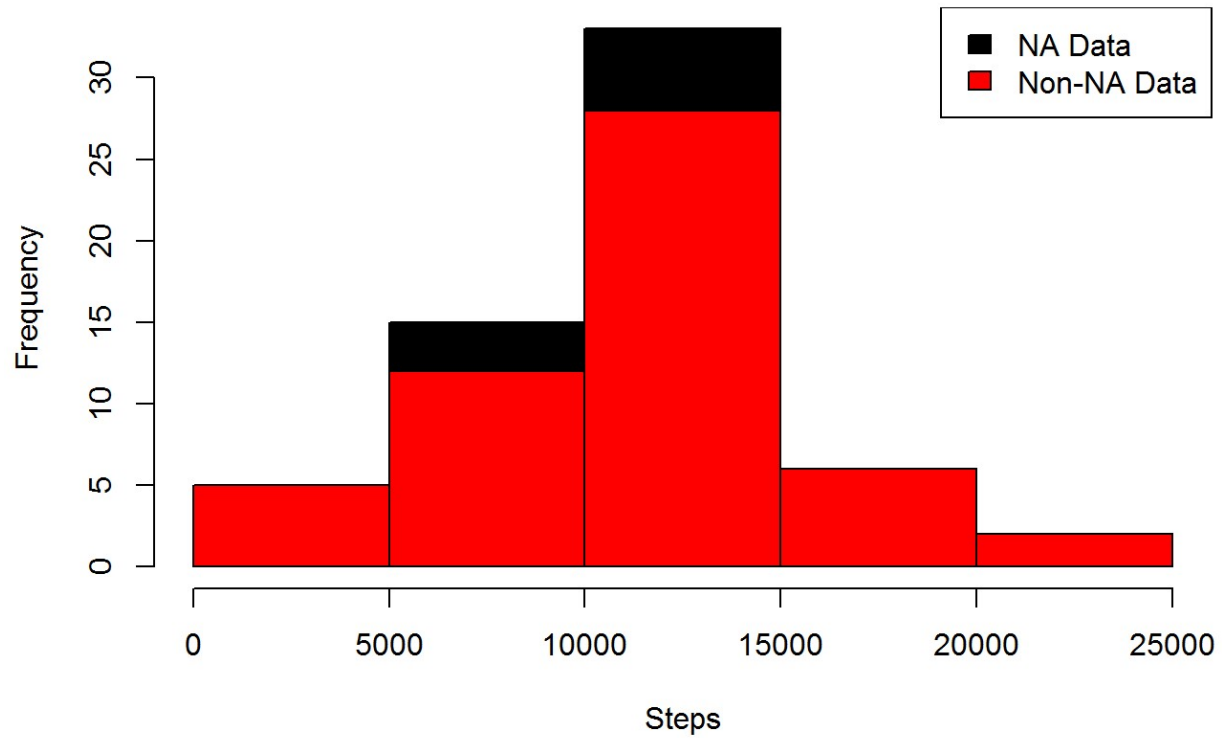
```
# Reorder the new substituted data in the same format as clean data set  
finalnewdata<- newdataset[,c(6,4,1,2,5)]  
colnames(finalnewdata)<- c("steps", "date", "interval", "day", "DateTime")
```

```
# Merge the NA averages and non NA data together  
mergedData <- rbind(clean, finalnewdata)
```

```
# Create sum of steps per date to compare with step 1  
sumTable2 <- aggregate(mergedData$steps ~ mergedData$date, FUN=sum)  
colnames(sumTable2)<- c("Date", "Steps")
```

```
## Creating the histogram of total steps per day, categorized by data set to show impact  
hist(sumTable2$Steps, breaks=5, xlab="Steps", main = "Total Steps per Day with NA and Non NA", col="Black")  
hist(sumTable$Steps, breaks=5, xlab="Steps", main = "Total Steps per Day with NA and Non NA", col="Red", add=T)  
legend("topright", c("NA Data", "Non-NA Data"), fill=c("black", "red") )
```

## Total Steps per Day with NA and Non NA



```
# Calculate the mean and the median with new data set  
as.integer(mean(sumTable2$Steps))
```

```
## [1] 10821
```

```
as.integer(median(sumTable2$Steps))
```

```
## [1] 11015
```

```
# Section # 5. Are there differences in activity patterns between weekdays and weekends?
```

```
## Create new category based on the days of the week
```

```
mergedData$DayCategory <- ifelse(mergedData$day %in% c("Saturday", "Sunday"),  
  "Weekend", "Weekday")
```

```
## Summarize data by interval and type of day
```

```
intervalTable2 <- ddply(mergedData, .(interval, DayCategory), summarize, Avg =  
  mean(steps))
```

```
## Plot data in a panel plot
```

```
xyplot(Avg~interval|DayCategory, data=intervalTable2, type="l", layout = c(1,  
  2),  
  main="Average Steps per Interval Based on Type of Day",  
  ylab="Average Number of Steps", xlab="Interval")
```

## Average Steps per Interval Based on Type of Day

