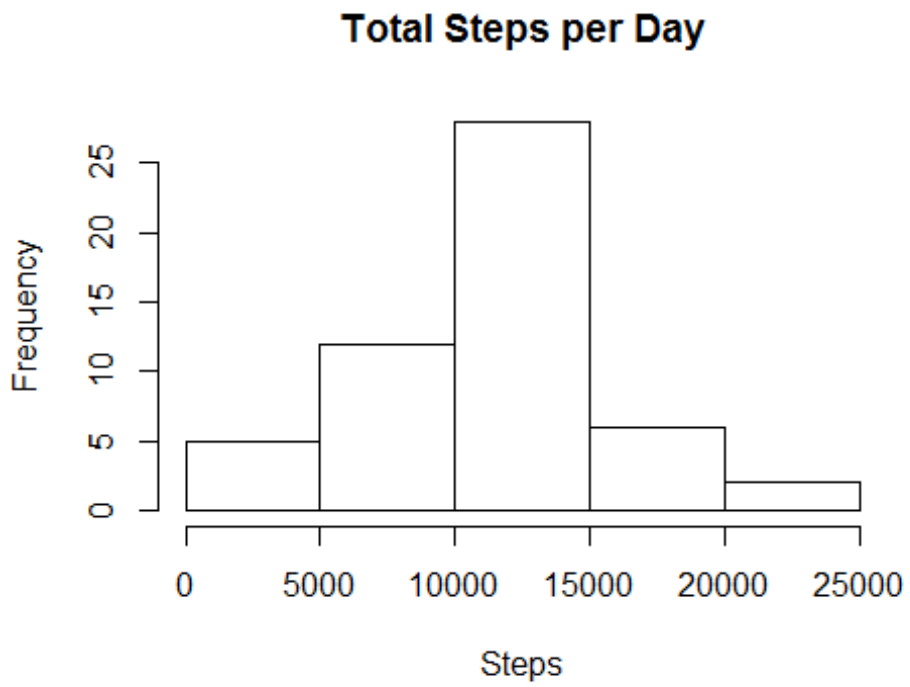
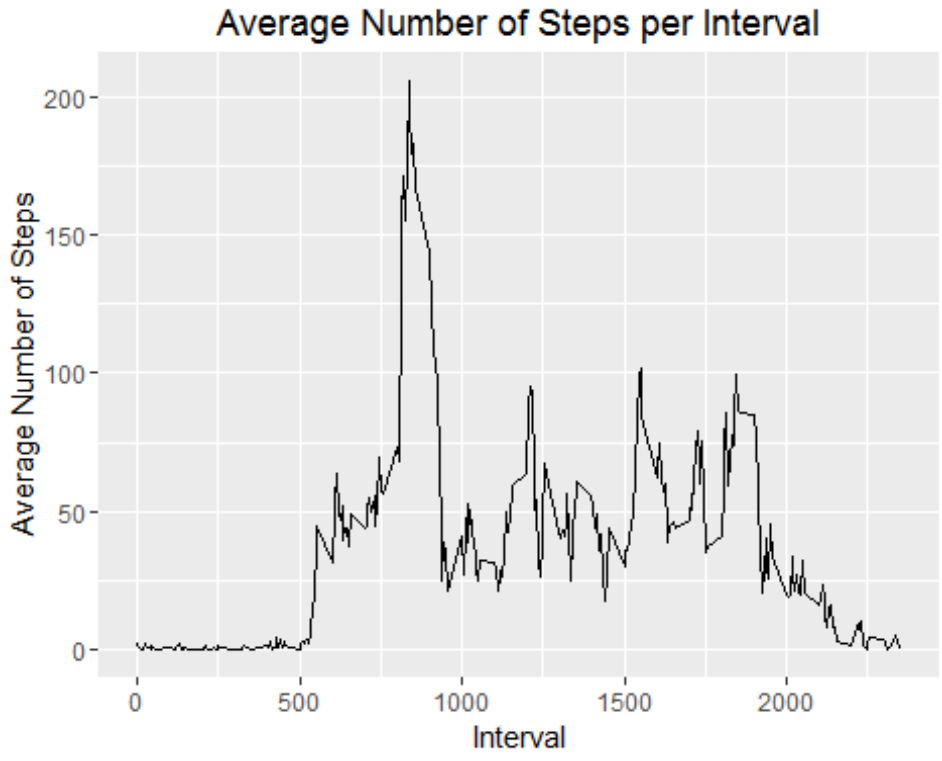
```r
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.3

library(plyr)

## Warning: package 'plyr' was built under R version 3.2.3

setwd("Q:/datasciencecoursera/ReproducibleResearch/CourseProj1")

activity <- read.csv("activity.csv")

activity$day <- weekdays(as.Date(activity$date))
activity$DateTime<- as.POSIXct(activity$date, format="%Y-%m-%d")

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

##pulling data without nas
clean <- activity[!is.na(activity$steps),]

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

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



```
## Mean of Steps
as.integer(mean(sumTable$Steps))

## [1] 10766

## Median of Steps
as.integer(median(sumTable$Steps))

## [1] 10765

## 2) Average Daily Activity

library(plyr)
library(ggplot2)
##pulling data without nas
clean <- activity[!is.na(activity$steps),]

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

##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()+xlab("Interval")+ylab("Average Number of
Steps")+ggtitle("Average Number of Steps per Interval")
```




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

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

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

```r
## 3) Imputing missing values

##Number of NAs in original data set
nrow(activity[is.na(activity$steps),])
```

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

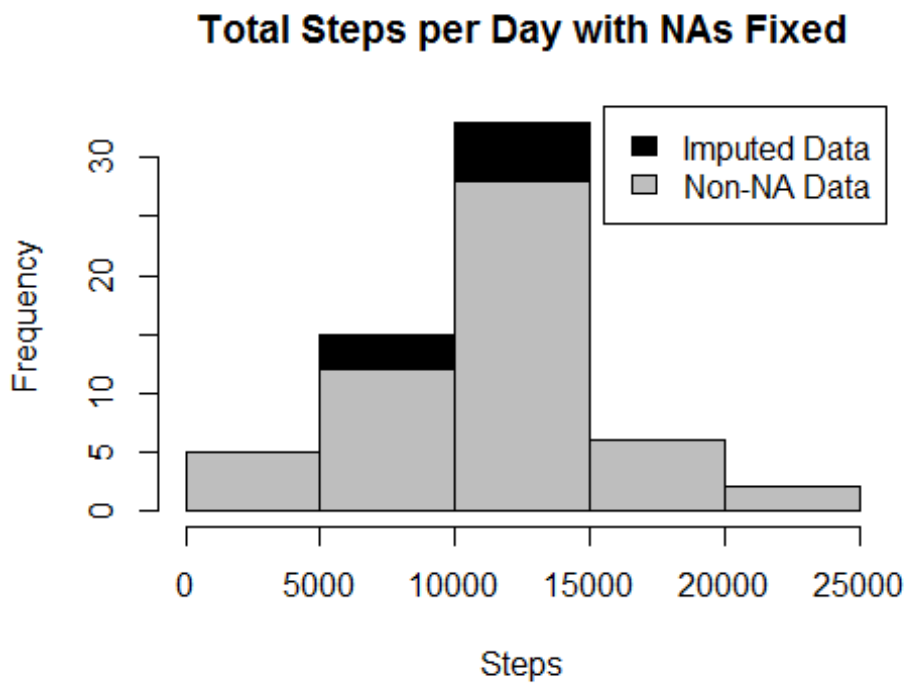
```r
## Create the average number of steps per weekday and interval
avgTable <- ddply(clean, .(interval, day), summarize, Avg = mean(steps))

## Create dataset with all NAs for substitution
nadata<- activity[is.na(activity$steps),]
## Merge NA data with average weekday interval for substitution
newdata<-merge(nadata, avgTable, by=c("interval", "day"))

## Reorder the new substituded data in the same format as clean data set
newdata2<- newdata[,c(6,4,1,2,5)]
colnames(newdata2)<- c("steps", "date", "interval", "day", "DateTime")

##Merge the NA averages and non NA data together
mergeData <- rbind(clean, newdata2)

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

## Mean of Steps with NA data taken care of
as.integer(mean(sumTable2$Steps))
```

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

```r
## Median of Steps with NA data taken care of
as.integer(median(sumTable2$Steps))
```

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

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




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

## Create new category based on the days of the week
mergeData$DayCategory <- ifelse(mergeData$day %in% c("Saturday", "Sunday"),
"Weekend", "Weekday")

library(lattice)

## Warning: package 'lattice' was built under R version 3.2.3

## Summarize data by interval and type of day
intervalTable2 <- ddply(mergeData, .(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