Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

This step is completely straightforward. I unzip the existing file in the archive (no need to check in the csv), and read the frame, which already has column headers.

```
unzip("activity.zip")
activityData<-read.csv("activity.csv")
dim(activityData)
```

```
## [1] 17568     3
```

```
head(activityData,n=1)
```

```
##   steps       date interval
## 1    NA 2012-10-01        0
```

```
library(plyr)
library(ggplot2)
```

What is mean total number of steps taken per day?

```
totals<-ddply(activityData,.(date),summarise,steps=sum(steps,na.rm=F))
meanSteps<-mean(totals$steps,na.rm=T)
meanSteps
```

```
## [1] 10766.19
```

What is the average daily activity pattern?

We collect the mean number steps by interval over all dates given (removing the NA row/column combos), and we get:

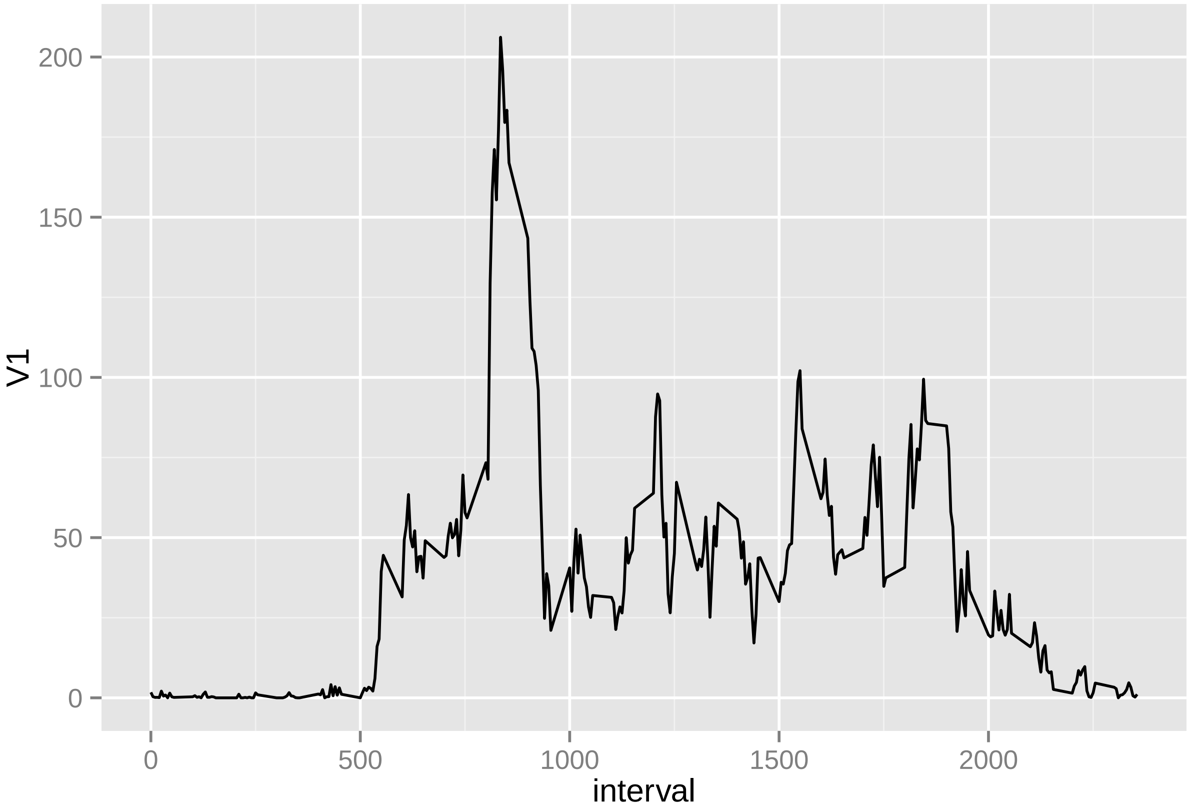
```
intervalTotals<-ddply(activityData,c("interval"),.fun=function(x) mean(x$steps,na.rm=T))
dim(intervalTotals)
```

```
## [1] 288   2
```

```
head(intervalTotals)
```

```
##   interval        V1
## 1        0 1.7169811
## 2        5 0.3396226
## 3       10 0.1320755
## 4       15 0.1509434
## 5       20 0.0754717
## 6       25 2.0943396
```

```
q<-ggplot(data=intervalTotals,aes(x=interval,y=V1))+geom_line()
q
```



Almost no steps are taken in the intervals from 0 to 500, and the number of steps peaks at around interval 800.

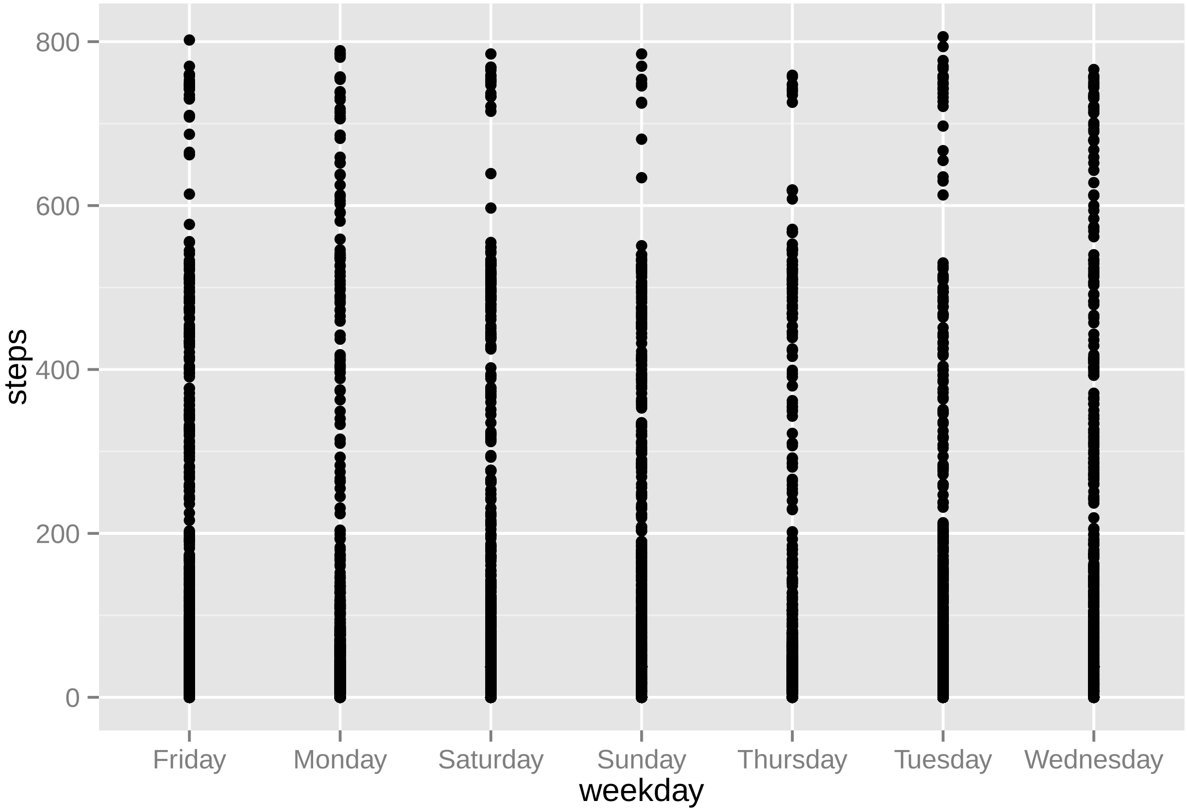
Imputing missing values using mean for day and interval

```
imputeRow<-function(date,interval){
  mean(activityData$steps[activityData$interval==interval&activityData$date==date],na.rm = T)
}

activityDataFilled<-transform(activityData,steps=ifelse(is.na(steps),imputeRow(date,interval),steps))

head(activityDataFilled)
```

```
##     steps       date interval
## 1 37.3826 2012-10-01        0
## 2 37.3826 2012-10-01        5
## 3 37.3826 2012-10-01       10
## 4 37.3826 2012-10-01       15
## 5 37.3826 2012-10-01       20
## 6 37.3826 2012-10-01       25
```

Are there differences in activity patterns between weekdays and weekends?

```
activityDataFilled<-transform(activityDataFilled,weekday=weekdays(as.Date(date)))
weekdaySummary<-ddply(activityDataFilled,.(weekday),total=mean(steps))
q<-ggplot(data=activityDataFilled,aes(x=weekday,y=steps))+geom_point()
q
```



The averages are the same, but activity seems more bi-modal on weekends, with many people taking fewer steps and a few taking far more steps, and few in between.