Reproducible Research Course Project 1

Melrose Huang

February 18, 2017

Load packages

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

Import data

```
df<-read.csv("C:/Users/mhuang/My Documents/activity.csv")
glimpse(df)
```

```
## Observations: 17,568
## Variables: 3
## $ steps    (int) NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N...
## $ date     (fctr) 2012-10-01, 2012-10-01, 2012-10-01, 2012-10-01, 2012...
## $ interval (int) 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 100, 10...
```

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

```
aggregate(df$steps, by=list(grp=df$date), FUN=sum, na.rm=TRUE) #Total number of steps per day
```

```
##           grp     x
## 1  2012-10-01     0
## 2  2012-10-02   126
## 3  2012-10-03 11352
## 4  2012-10-04 12116
## 5  2012-10-05 13294
## 6  2012-10-06 15420
## 7  2012-10-07 11015
## 8  2012-10-08     0
## 9  2012-10-09 12811
## 10 2012-10-10  9900
## 11 2012-10-11 10304
## 12 2012-10-12 17382
## 13 2012-10-13 12426
## 14 2012-10-14 15098
## 15 2012-10-15 10139
## 16 2012-10-16 15084
## 17 2012-10-17 13452
## 18 2012-10-18 10056
```

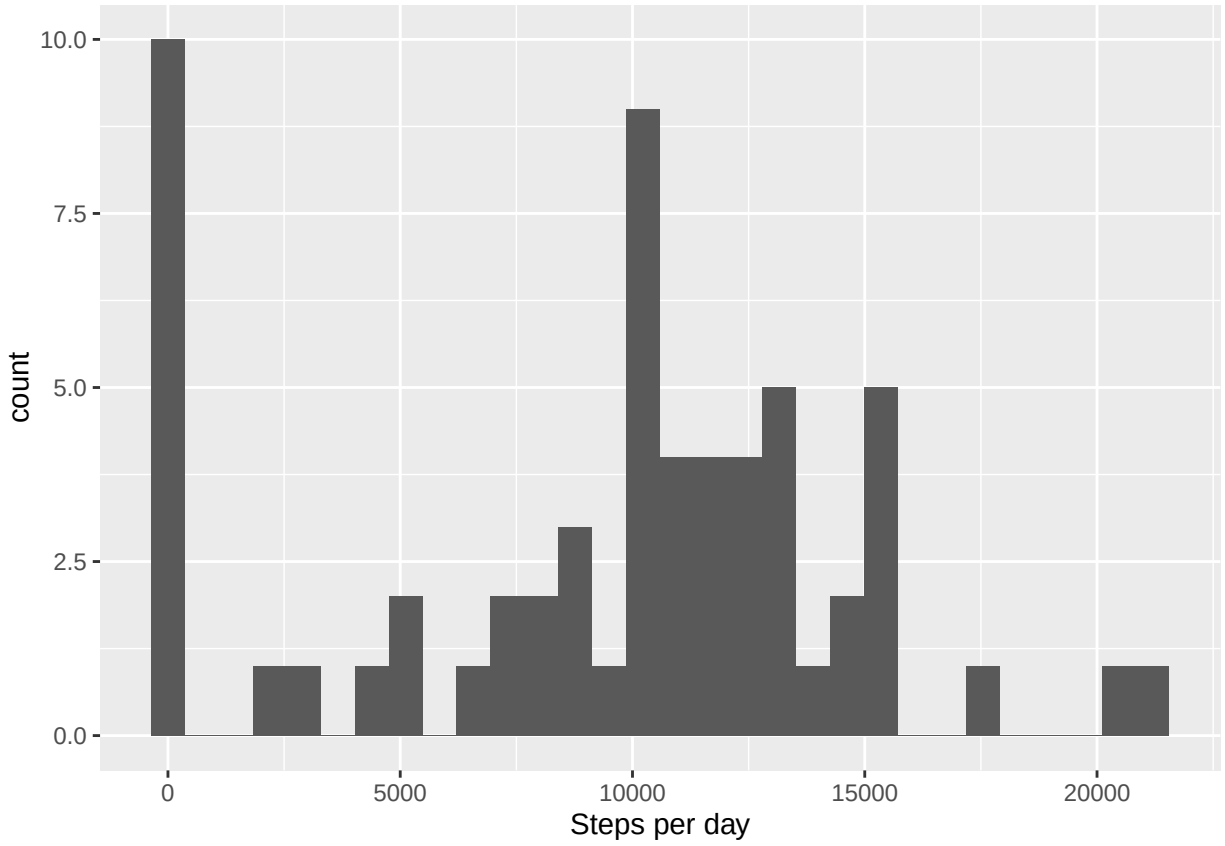
```
## 19 2012-10-19 11829
## 20 2012-10-20 10395
## 21 2012-10-21  8821
## 22 2012-10-22 13460
## 23 2012-10-23  8918
## 24 2012-10-24  8355
## 25 2012-10-25  2492
## 26 2012-10-26  6778
## 27 2012-10-27 10119
## 28 2012-10-28 11458
## 29 2012-10-29  5018
## 30 2012-10-30  9819
## 31 2012-10-31 15414
## 32 2012-11-01     0
## 33 2012-11-02 10600
## 34 2012-11-03 10571
## 35 2012-11-04     0
## 36 2012-11-05 10439
## 37 2012-11-06  8334
## 38 2012-11-07 12883
## 39 2012-11-08  3219
## 40 2012-11-09     0
## 41 2012-11-10     0
## 42 2012-11-11 12608
## 43 2012-11-12 10765
## 44 2012-11-13  7336
## 45 2012-11-14     0
## 46 2012-11-15    41
## 47 2012-11-16  5441
## 48 2012-11-17 14339
## 49 2012-11-18 15110
## 50 2012-11-19  8841
## 51 2012-11-20  4472
## 52 2012-11-21 12787
## 53 2012-11-22 20427
## 54 2012-11-23 21194
## 55 2012-11-24 14478
## 56 2012-11-25 11834
## 57 2012-11-26 11162
## 58 2012-11-27 13646
## 59 2012-11-28 10183
## 60 2012-11-29  7047
## 61 2012-11-30     0
```

```
perday<-as.data.frame(aggregate(df$steps, by=list(grp=df$date), FUN=sum, na.rm=T)) #Create data frame o
glimpse(perday)
```

```
## Observations: 61
## Variables: 2
## $ grp (fctr) 2012-10-01, 2012-10-02, 2012-10-03, 2012-10-04, 2012-10-0...
## $ x   (int) 0, 126, 11352, 12116, 13294, 15420, 11015, 0, 12811, 9900,...
```

```
qplot(x, data=perday, geom="histogram", xlab = "Steps per day") #Create histogram representing total nu
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
mean(df$steps, na.rm=TRUE) #Mean number of steps taken per day
```

```
## [1] 37.3826
```

```
median(df$steps,na.rm=TRUE) #Median number of steps taken per day
```

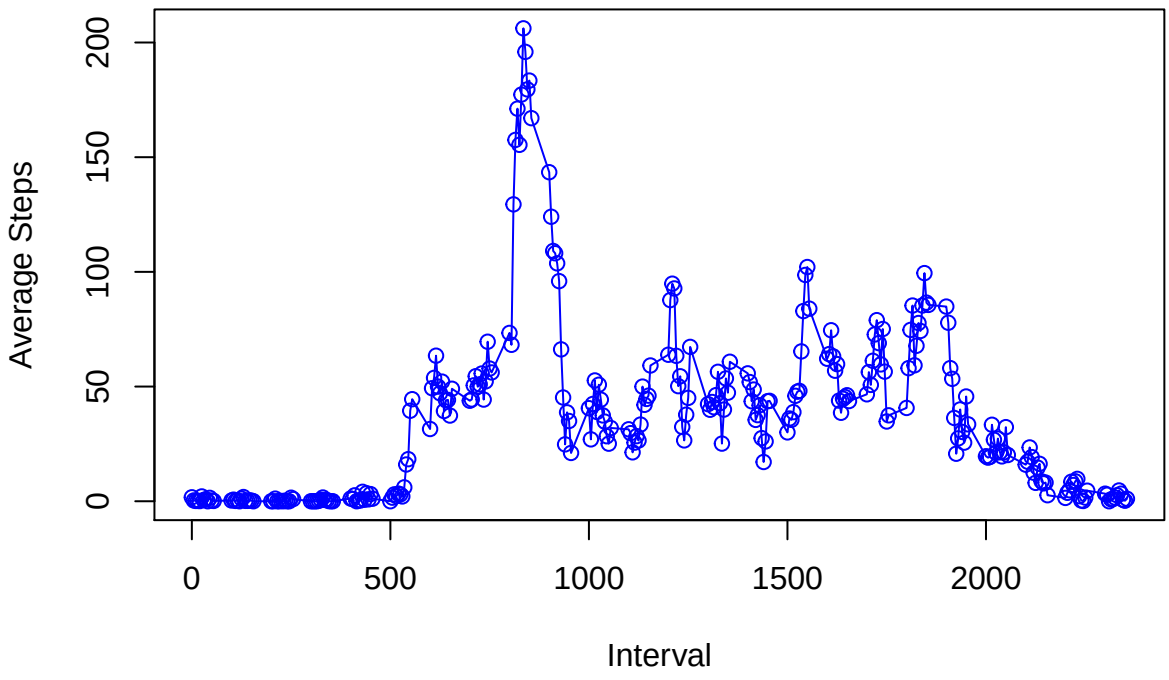
```
## [1] 0
```

What is the average daily activity pattern?

```
fivemin <-as.data.frame(aggregate(df$steps, by=list(grp=df$interval), FUN=mean, na.rm=T)) #Create data
glimpse(fivemin)
```

```
## Observations: 288
## Variables: 2
## $ grp (int) 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 100, 105, 11...
## $ x   (dbl) 1.7169811, 0.3396226, 0.1320755, 0.1509434, 0.0754717, 2.0...
```

```
plot(fivemin$grp, fivemin$x, xlab="Interval",
    ylab="Average Steps", type="o", col="blue")
```




```
#Time series plot of 5-minute interval (x-axis) and average number of steps taken averaged across all d
max_interval<-filter(fivemin, x==max(fivemin$x))
glimpse(max_interval) #Interval with maximum number of average steps is 835.
```

```
## Observations: 1
## Variables: 2
## $ grp (int) 835
## $ x   (dbl) 206.1698
```

Impute missing values

```
summary(df) #Total number of (rows with) NAs = 2304
```

```
##      steps                date          interval
##  Min.   :  0.00   2012-10-01:  288   Min.   :   0.0
##  1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8
##  Median :  0.00   2012-10-03:  288   Median :1177.5
##  Mean   : 37.38   2012-10-04:  288   Mean   :1177.5
##  3rd Qu.: 12.00   2012-10-05:  288   3rd Qu.:1766.2
##  Max.   :806.00   2012-10-06:  288   Max.   :2355.0
##  NA's   :2304     (Other)   :15840
```

```
df1<-df #Create new data frame, df1, that is a copy of the original, df
df1$steps<-ifelse(is.na(df$steps), median(df$steps, na.rm=T), df$steps) #Replace NA steps value with me
summary(df1) #Check imputation
```

```
##      steps                date          interval
##  Min.   :  0.00   2012-10-01:  288   Min.   :   0.0
```

```
##  1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8
##  Median :  0.00   2012-10-03:  288   Median :1177.5
##  Mean   : 32.48   2012-10-04:  288   Mean   :1177.5
##  3rd Qu.:  0.00   2012-10-05:  288   3rd Qu.:1766.2
##  Max.   :806.00   2012-10-06:  288   Max.   :2355.0
##                   (Other)   :15840
```

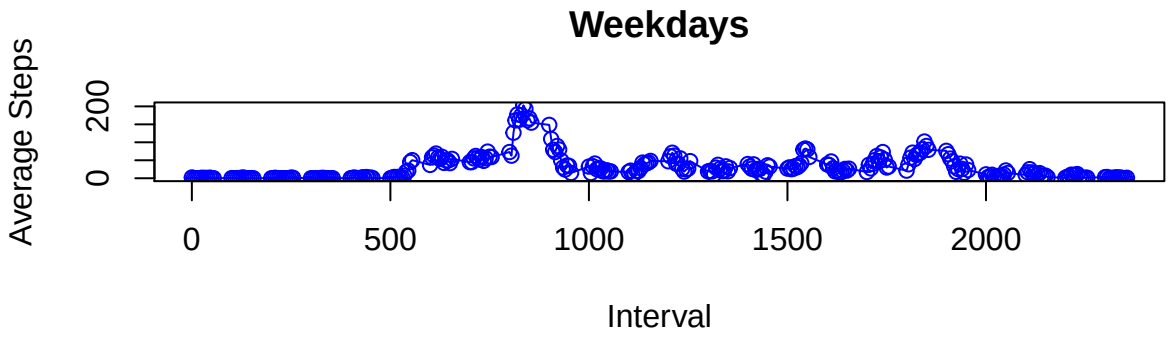
Are there any differences in activity patterns between weekdays and weekends?

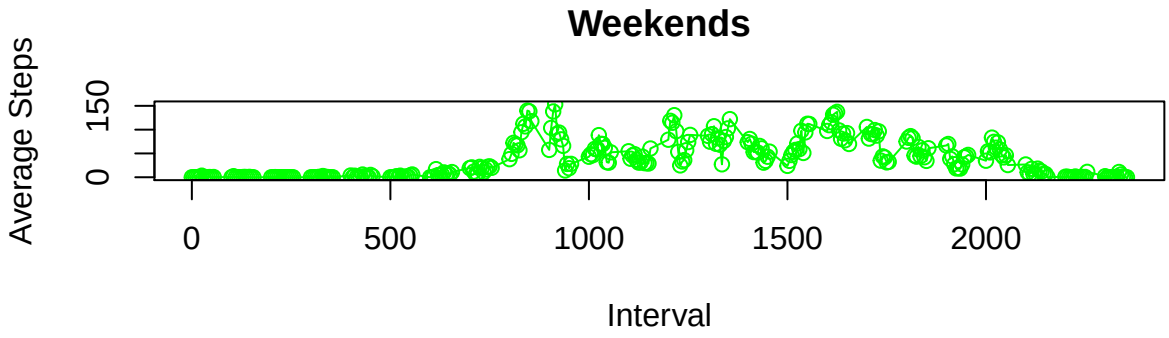
```
df1$weekday<- weekdays(as.Date(df1$date))
df1$day_type<-as.factor(ifelse((df1$weekday=="Saturday"|df1$weekday=="Sunday"), "weekend", "weekday")) 
addmargins(xtabs(~weekday+day_type, df1, na.action=na.pass, exclude=NULL)) #Check new variable by runni
```

```
##            day_type
## weekday     weekday weekend   Sum
##   Friday       2592       0  2592
##   Monday       2592       0  2592
##   Saturday        0    2304  2304
##   Sunday          0    2304  2304
##   Thursday     2592       0  2592
##   Tuesday      2592       0  2592
##   Wednesday    2592       0  2592
##   Sum         12960    4608 17568
```

```
weekdays<-filter(df1, day_type=="weekday")
weekends<-filter(df1, day_type=="weekend")
fivemin_weekday<-as.data.frame(aggregate(weekdays$steps, by=list(grp=weekdays$interval), FUN=mean, na.r
fivemin_weekend<-as.data.frame(aggregate(weekends$steps, by=list(grp=weekends$interval), FUN=mean, na.r

par(mfrow=c(2,1))
plot(fivemin_weekday$grp, fivemin_weekday$x, xlab="Interval",
    ylab="Average Steps", type="o", col="blue")
      title(main="Weekdays")
plot(fivemin_weekend$grp, fivemin_weekend$x, xlab="Interval",
    ylab="Average Steps", type="o", col="green")
      title(main="Weekends")
```






```
#Make panel plot with time series plot of 5-minute interval and average # steps across all weekday days
```