==========

title: “IndividualAtivityAnalysis.Rmd”
author: “M Dugda”
date: “Saturday, June 04, 2016”
#output: html_document
Creating Dirctory, downloading and unzipping data

```
Url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
if (!file.exists("data")) dir.create("data")
if (!file.exists("data/data.zip")) download.file(url, destfile = "data/data.zip", mode="wb")
if (!file.exists("data/activity.csv")) unzip("data/data.zip", exdir="data")
```

Load the data and check its structure and summary

```
activity <- read.csv("data/activity.csv",quote="\"")
str(activity)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : Factor w/ 61 levels "2012-10-01","2012-10-02",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
head(activity)
```

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

```
summary(activity)
```

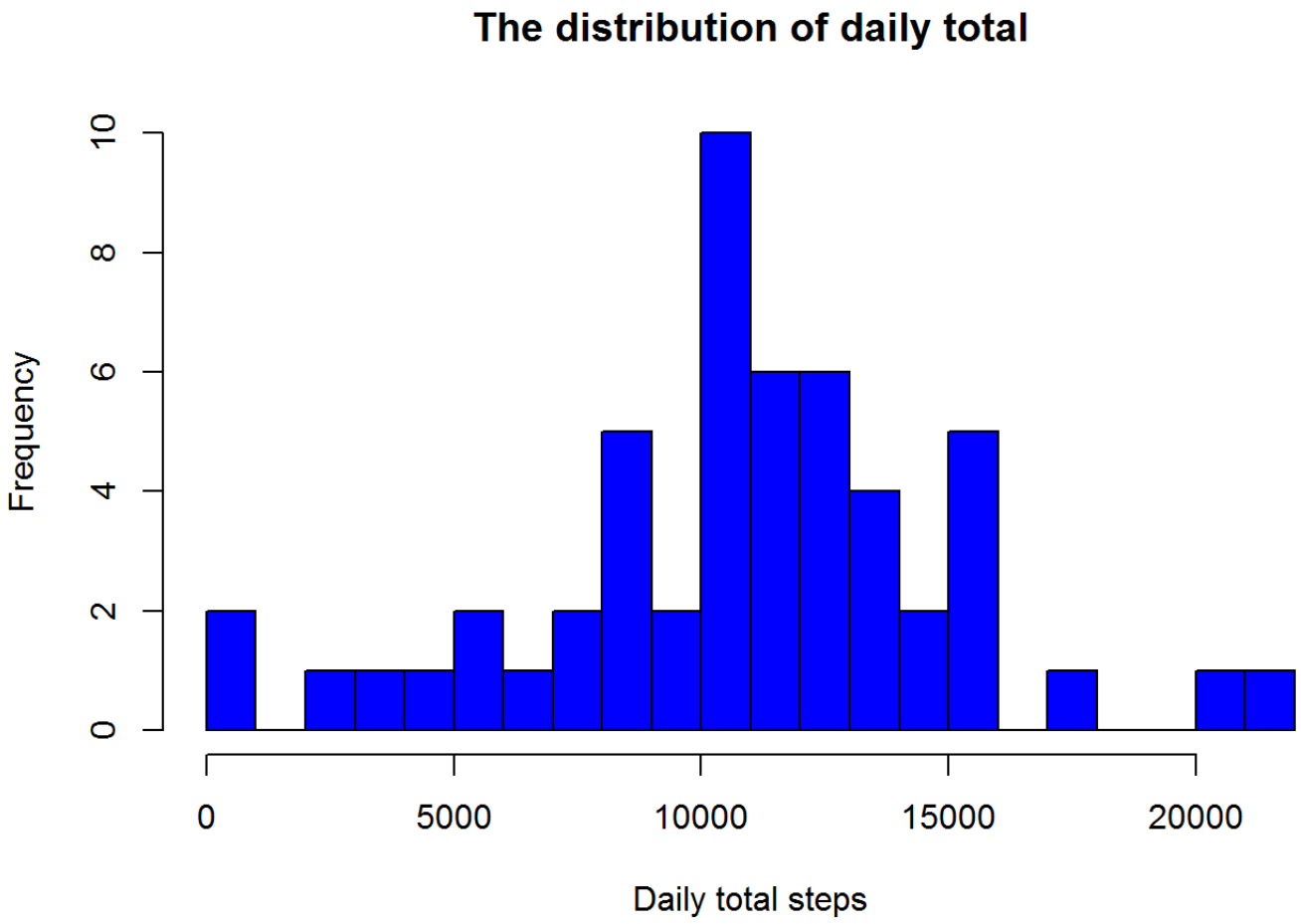
```
##      steps               date          interval     
##  Min.   :  0.00   2012-10-01:  288   Min.   :   0.0  
##  1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8  
##  Median :  0.00   2012-10-03:  288   Median :1177.5  
##  Mean   : 37.38   2012-10-04:  288   Mean   :1177.5  
##  3rd Qu.: 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
```

```
activity$date <- as.Date(activity$date)
activity_ign <- subset(activity, !is.na(activity$steps))
```

What is mean total number of steps taken per day?

```
dailysum <- tapply(activity_ign$steps, activity_ign$date, sum, na.rm=TRUE, simplify=T)
dailysum <- dailysum[!is.na(dailysum)]

hist(x=dailysum,
     col="blue",
     breaks=20,
     xlab="Daily total steps",
     ylab="Frequency",
     main="The distribution of daily total")
```




```
mean(dailysum)
```

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

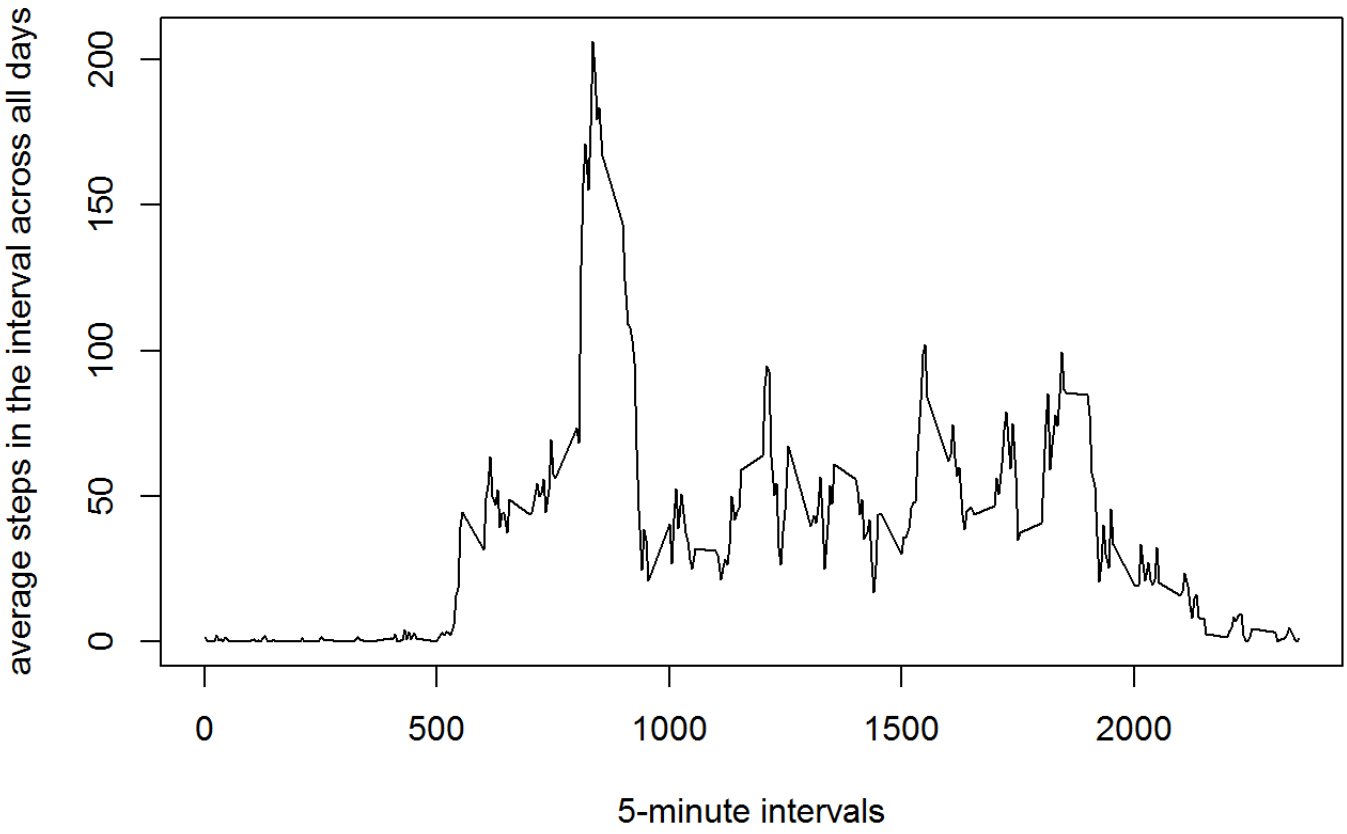
```
median(dailysum)
```

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

What is the average daily activity pattern?

```
int_avg <- tapply(activity_ign$steps, activity_ign$interval, mean, na.rm=TRUE, simplify=T)
activity_ia <- data.frame(interval=as.integer(names(int_avg)), avg=int_avg)

with(activity_ia,
     plot(interval,
          avg,
          type="l",
          xlab="5-minute intervals",
          ylab="average steps in the interval across all days"))
```

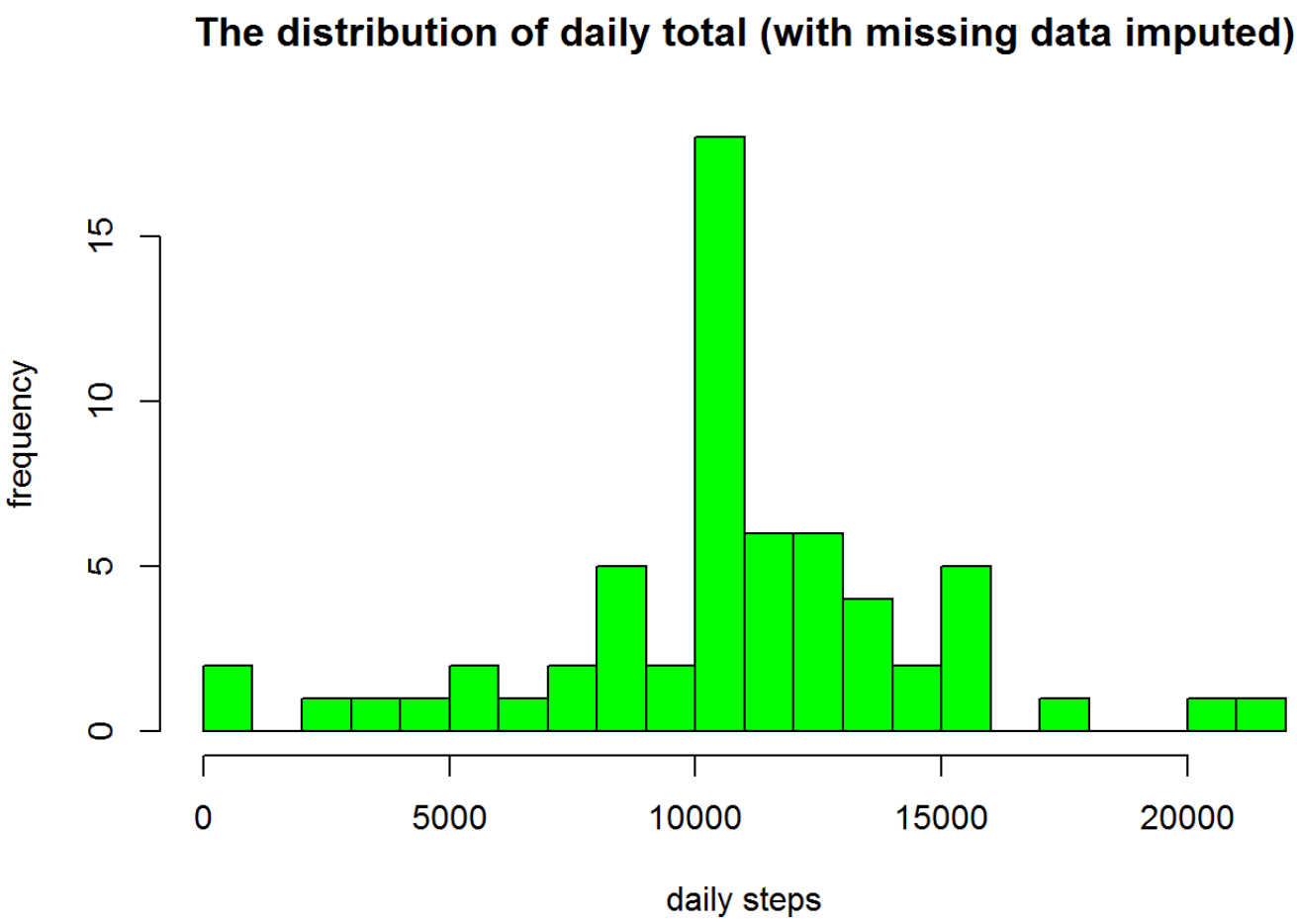


Imputing missing values

```
activity_impute <- activity
ndx <- is.na(activity_impute$steps)
int_avg <- tapply(activity_ign$steps, activity_ign$interval, mean, na.rm=TRUE, simplify=T)
activity_impute$steps[ndx] <- int_avg[as.character(activity_impute$interval[ndx])]

new_dailysum <- tapply(activity_impute$steps, activity_impute$date, sum, na.rm=TRUE, simplify=T)

hist(x=new_dailysum,
     col="green",
     breaks=20,
     xlab="daily steps",
     ylab="frequency",
     main="The distribution of daily total (with missing data imputed)")
```




```
mean(new_dailysum)
```

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

```
median(new_dailysum)
```

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

Are there differences in activity patterns between weekdays and weekends?

helper function to decide if a day is a week day or not

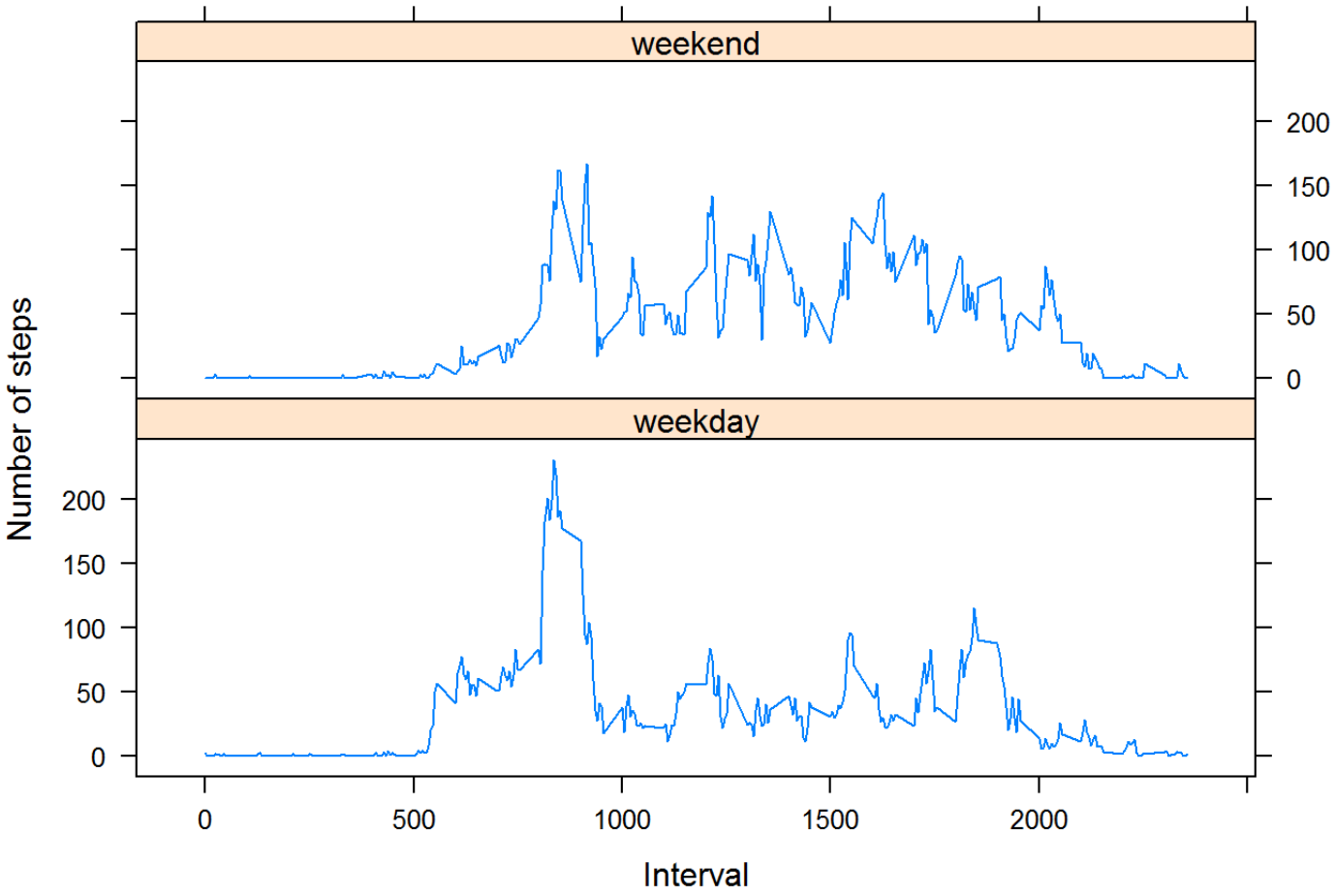
```
is_weekday <- function(d) {
    wd <- weekdays(d)
    ifelse (wd == "Saturday" | wd == "Sunday", "weekend", "weekday")
}

wx <- sapply(activity_impute$date, is_weekday)
activity_impute$wk <- as.factor(wx)
head(activity_impute)
```

```
##       steps       date interval      wk
## 1 1.7169811 2012-10-01        0 weekday
## 2 0.3396226 2012-10-01        5 weekday
## 3 0.1320755 2012-10-01       10 weekday
## 4 0.1509434 2012-10-01       15 weekday
## 5 0.0754717 2012-10-01       20 weekday
## 6 2.0943396 2012-10-01       25 weekday
```

make a panel plot containing a time series plot

```
wk_activity <- aggregate(steps ~ wk+interval, data=activity_impute, FUN=mean)
library(lattice)
xyplot(steps ~ interval | factor(wk),
       layout = c(1, 2),
       xlab="Interval",
       ylab="Number of steps",
       type="l",
       lty=1,
       data=wk_activity)
```



The END of the Rmd Code