Reproducible Research Course Project 1

Data

The data for this assignment was downloaded from the course web site:

Dataset: Activity monitoring data [52K] The variables included in this dataset are:

steps: Number of steps taking in a 5-minute interval (missing values are coded as NA)

date: The date on which the measurement was taken in YYYY-MM-DD format

interval: Identifier for the 5-minute interval in which measurement was taken

The dataset is stored in a comma-separated-value (CSV) file and there are a total of 17,568 observations in this dataset.

Loading and preprocessing the data

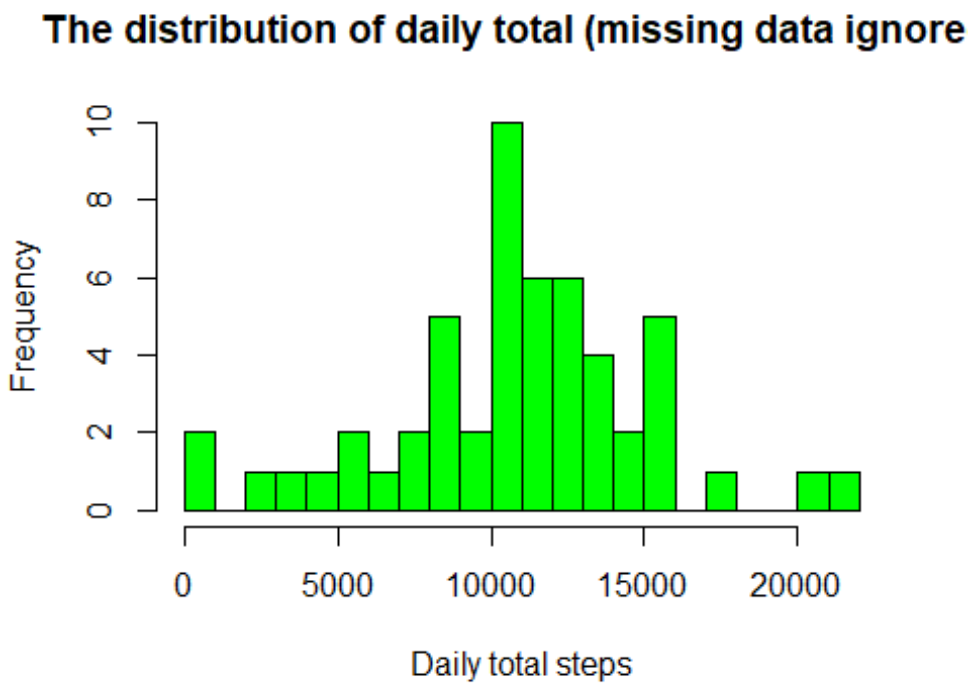
Download, unzip and load data into data frame data.

```
knitr::opts_chunk$set(
  fig.path = "images/"
)
cls = c("integer", "character", "integer")
Data <- read.csv("activity.csv", head=TRUE, colClasses=cls, na.strings="NA")
head(Data)
```

```
##   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
```

What is mean total number of steps taken per day?

```
Data$date <- as.Date(Data$date)
Data_ign <- subset(Data, !is.na(Data$steps))
dailysum <- tapply(Data_ign$steps, Data_ign$date, sum, na.rm=TRUE,
simplify=T)
dailysum <- dailysum[!is.na(dailysum)]

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

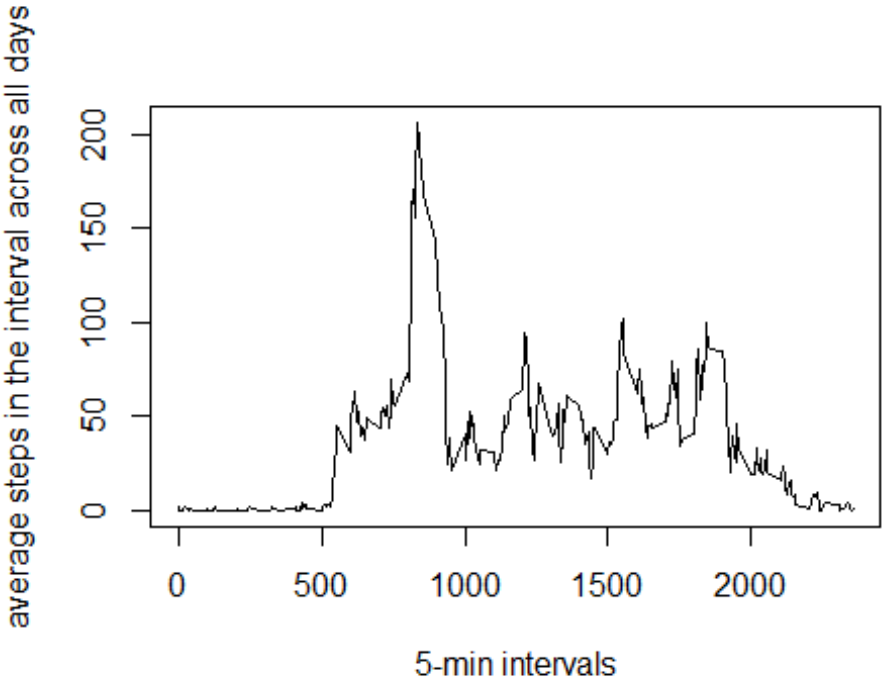


Mean and median of total number of steps taken per day

```
mean(dailysum)

## [1] 10766.19

median(dailysum)

## [1] 10765
```

What is the average daily activity pattern?

```
knitr::opts_chunk$set(
  fig.path = "images/"
)
int_avg <- tapply(Data_ign$steps, Data_ign$interval, mean, na.rm=TRUE,
simplify=T)
Data_ia <- data.frame(interval=as.integer(names(int_avg)), avg=int_avg)

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

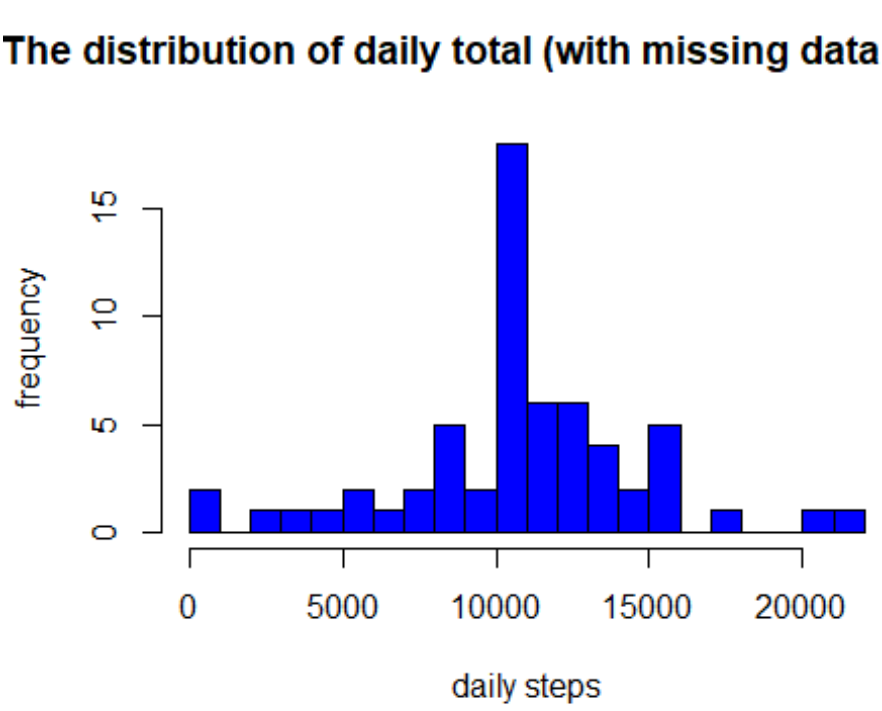



```
max_steps <- max(Data_ia$avg)
Data_ia[Data_ia$avg == max_steps, ]

##     interval      avg
## 835      835 206.1698
```

Imputing missing values

```
sum(is.na(Data$steps))

## [1] 2304

Data_impute <- Data
ndx <- is.na(Data_impute$steps)
int_avg <- tapply(Data_ign$steps, Data_ign$interval, mean, na.rm=TRUE,
simplify=T)
Data_impute$steps[ndx] <- int_avg[as.character(Data_impute$interval[ndx])]

knitr::opts_chunk$set(
  fig.path = "images/"
)
new_dailysum <- tapply(Data_impute$steps, Data_impute$date, sum, na.rm=TRUE,
simplify=T)

hist(x=new_dailysum,
     col="blue",
     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?

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

wx <- sapply(Data_impute$date, is_weekday)
Data_impute$wk <- as.factor(wx)
head(Data_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

knitr::opts_chunk$set(
  fig.path = "images/"
)
wk_Data <- aggregate(steps ~ wk+interval, data=Data_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_Data)
```