PA1_template.Rmd

abhibhaku

April 22, 2018

```
knitr::opts_chunk$set(echo = TRUE)

## setting working direcctory and reading file into R

setwd("D:/ISB Co 2018/Coursera/Data Science Specialization/Reproducible Research/Week 2/Project
 1")

actv1 <- read.csv("activity.csv")
str(actv1)
```

```
## '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 ...
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.4.3
```

```
##
## 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(zoo)
```

```
## Warning: package 'zoo' was built under R version 3.4.4
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
actv1 <- mutate(actv1, date = as.Date(as.character(date), "%Y-%m-%d"))
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.3
```

```
is.regular(actv1$date)
```

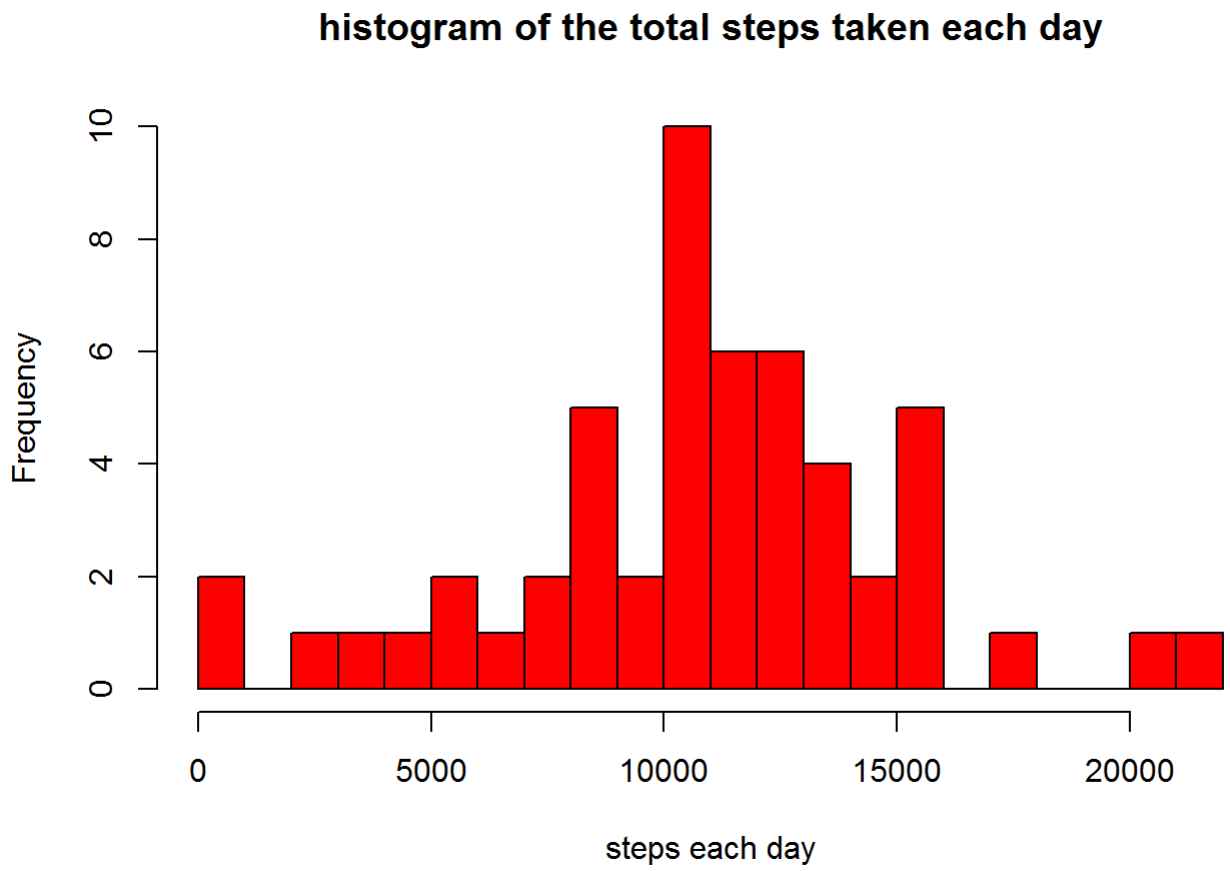
```
## [1] TRUE
```

```
unique(actv1$date)
```

```
##  [1] "2012-10-01" "2012-10-02" "2012-10-03" "2012-10-04" "2012-10-05"
##  [6] "2012-10-06" "2012-10-07" "2012-10-08" "2012-10-09" "2012-10-10"
## [11] "2012-10-11" "2012-10-12" "2012-10-13" "2012-10-14" "2012-10-15"
## [16] "2012-10-16" "2012-10-17" "2012-10-18" "2012-10-19" "2012-10-20"
## [21] "2012-10-21" "2012-10-22" "2012-10-23" "2012-10-24" "2012-10-25"
## [26] "2012-10-26" "2012-10-27" "2012-10-28" "2012-10-29" "2012-10-30"
## [31] "2012-10-31" "2012-11-01" "2012-11-02" "2012-11-03" "2012-11-04"
## [36] "2012-11-05" "2012-11-06" "2012-11-07" "2012-11-08" "2012-11-09"
## [41] "2012-11-10" "2012-11-11" "2012-11-12" "2012-11-13" "2012-11-14"
## [46] "2012-11-15" "2012-11-16" "2012-11-17" "2012-11-18" "2012-11-19"
## [51] "2012-11-20" "2012-11-21" "2012-11-22" "2012-11-23" "2012-11-24"
## [56] "2012-11-25" "2012-11-26" "2012-11-27" "2012-11-28" "2012-11-29"
## [61] "2012-11-30"
```

```
## histogram of total steps taken each day

t_steps1 <- aggregate(steps~date,data=actv1,sum,na.rm = TRUE)
names(t_steps1) <- c("date","sum of steps")
hist(t_steps1$`sum of steps`, main = "histogram of the total steps taken each day", xlab = "step
s each day", breaks = 20, col = "red")
```




```
## mean & median of total steps taken each day

mean(t_steps1$`sum of steps`)
```

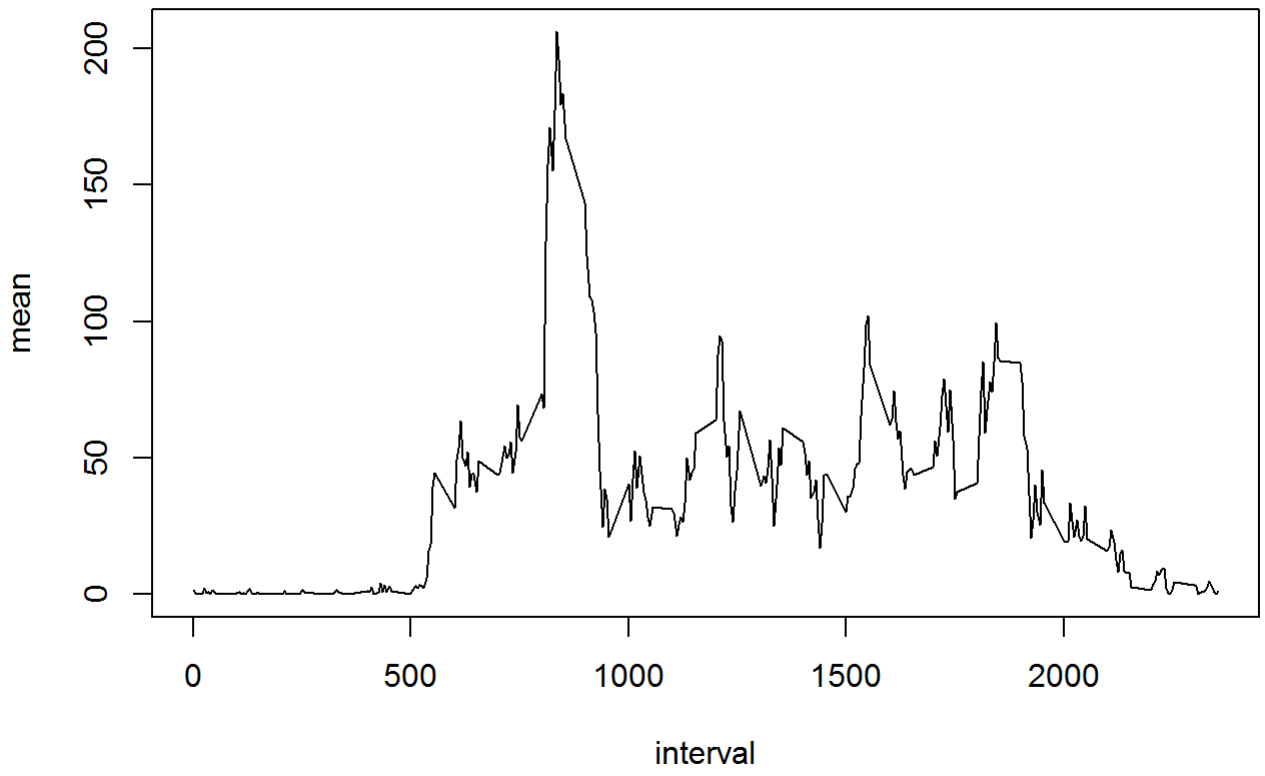
```
## [1] 10766.19
```

```
median(t_steps1$`sum of steps`)
```

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

```
## time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number o
f steps taken, averaged across all days (y-axis)

i_steps1 <- aggregate(steps~interval,data=actv1,mean,na.rm = TRUE)
names(i_steps1) <- c("interval", "mean")
plot(i_steps1,type="l")
```




```
## 5-minute interval, on average across all the days in the dataset, that contains the maximum n
umber of steps

i_steps1$interval[which.max(i_steps1$mean)]
```

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

```
## number of missing values or NA in the dataset

sum(is.na(actv1))
```

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

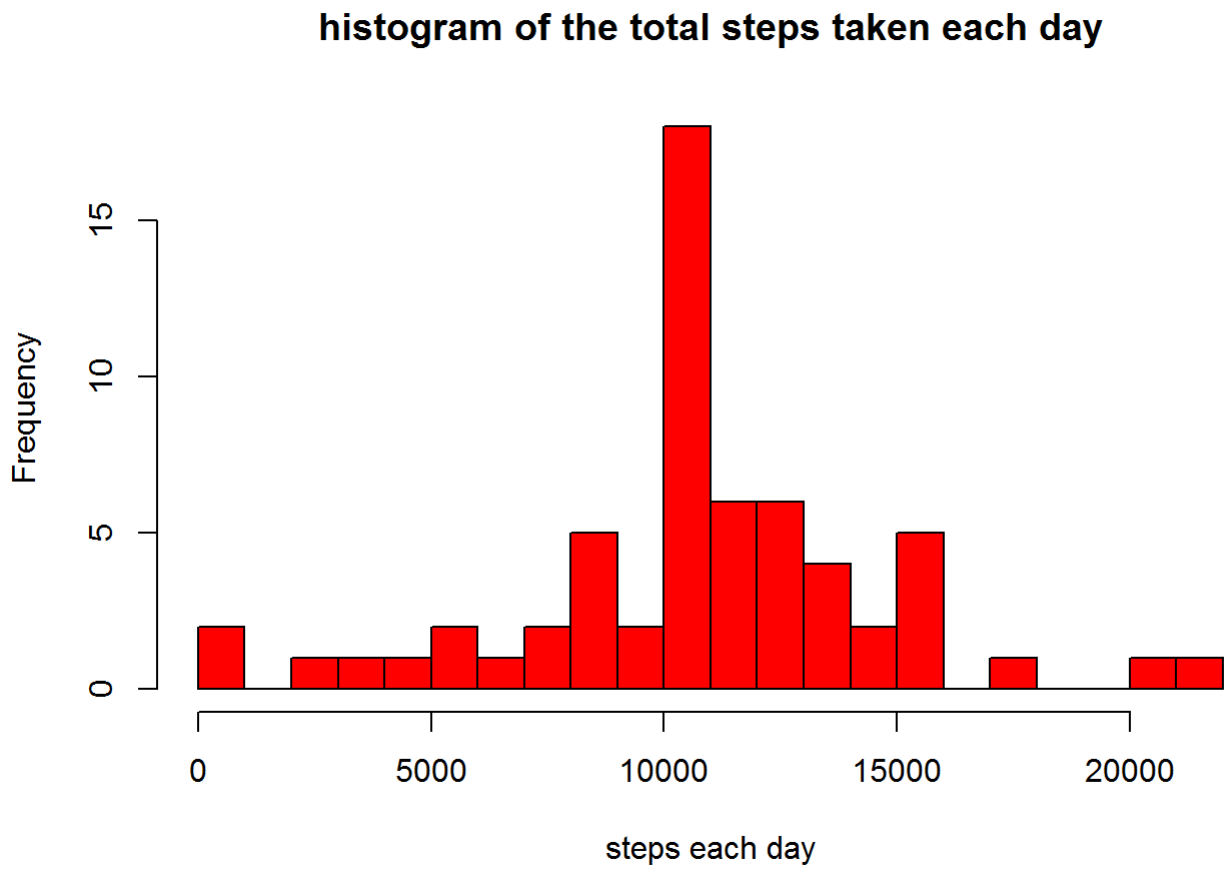
```
## for missing values or with NA in dataset, plan to replace them with the respective mean of th
at 5-min interval

## creating the new dataset with missing values replaced

actv2 <- actv1
actv2$steps[is.na(actv2$steps)] <- mean(na.omit(actv2$steps))
actv2$date <- as.Date(actv2$date,format = "%Y-%m-%d")

## histogram of total steps taken each day

t_steps2 <- aggregate(steps~date,data=actv2,sum,na.rm = TRUE)
names(t_steps2) <- c("date","sum of steps")
hist(t_steps2$`sum of steps`, main = "histogram of the total steps taken each day", xlab = "step
s each day", breaks = 20, col = "red")
```




```
## mean & median of total steps taken each day

mean(t_steps2$`sum of steps`)
```

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

```
median(t_steps2$`sum of steps`)
```

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

```
## difference in activity patterns between weekdays & weekends

library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 3.4.3
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

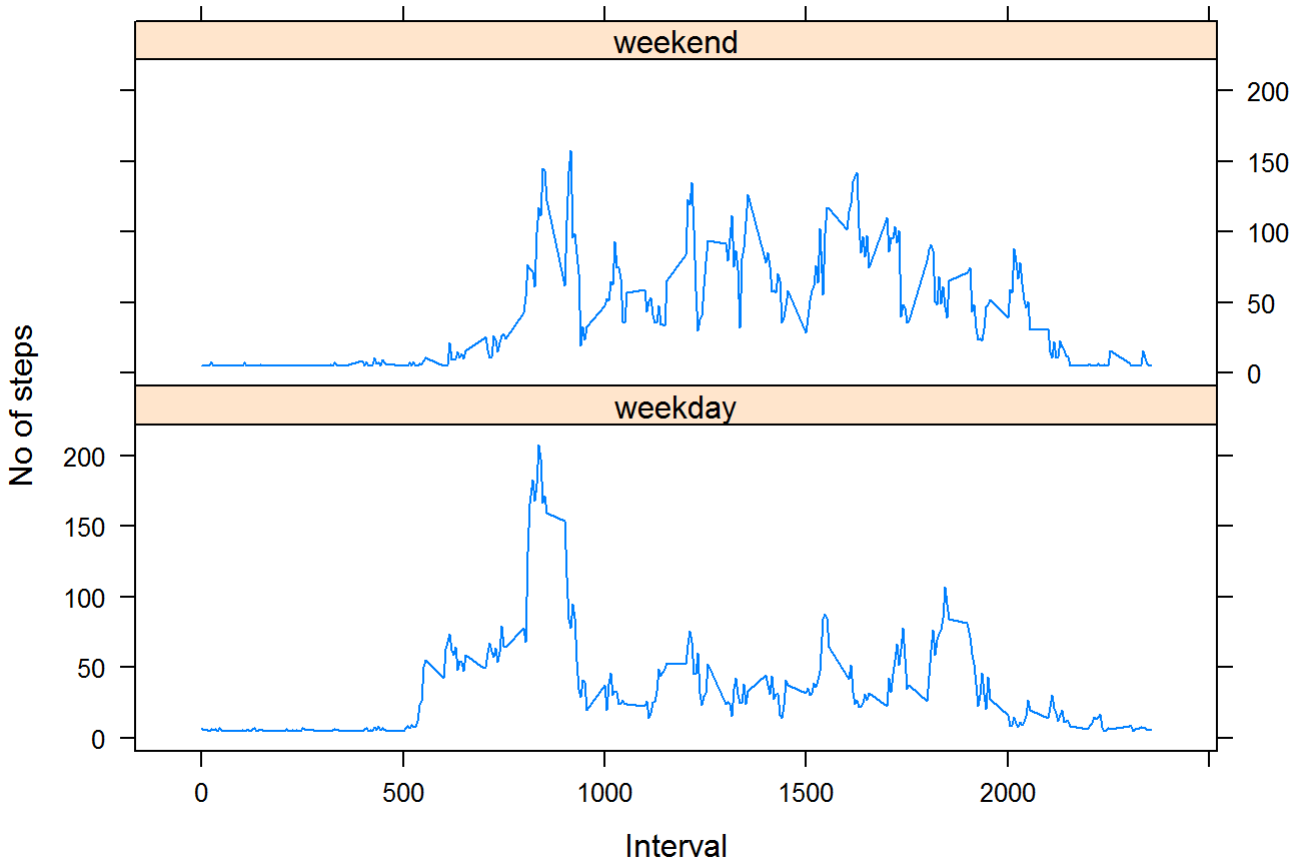
```
# write a fucntion to check if date corresponnds to weekday or weekend

whatday <- function(d){
  c <- weekdays(d)
  ifelse (c=="Saturday"|c=="Sunday","weekend","weekday")
}

e <- sapply(actv2$date,whatday)
actv2$wkday <- as.factor(e)

# panel plot containing a time series plot (i.e. type = "l") of the 5-minute interval (x-axis) a
nd the average number of steps taken, averaged across all weekday days or weekend days (y-axis)

library(lattice)
actv_wk <- aggregate(steps ~ wkday+interval, data=actv2, FUN=mean)
xyplot(steps ~ interval | factor(wkday),layout = c(1, 2), xlab="Interval", ylab="No of steps", t
ype="l", lty=1, data=actv_wk)
```


weekend
weekday
No of steps
Interval
0
50
100
150
200
0
500
1000
1500
2000