Activities_P1

Steps

Loading and preprocessing the data

```
# download data
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"
destfile <- "repdata_data_activity.zip"
download.file(url, destfile)
# unzip data
unzip("repdata_data_activity.zip",exdir = "data")
```

```
activity <- read.csv("data/activity.csv", stringsAsFactors=FALSE)
str(activity)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : chr  "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
summary(activity)
```

```
##      steps            date              interval
##  Min.   :  0.00   Length:17568       Min.   :   0.0
##  1st Qu.:  0.00   Class :character   1st Qu.: 588.8
##  Median :  0.00   Mode  :character   Median :1177.5
##  Mean   : 37.38                      Mean   :1177.5
##  3rd Qu.: 12.00                      3rd Qu.:1766.2
##  Max.   :806.00                      Max.   :2355.0
##  NA's   :2304
```

```
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
```

```
# data set with NA rows removed
activity <- activity[which(!is.na(activity$steps)), ]
# further investigation
nlevels(as.factor(activity$steps))
```

```
## [1] 617
```

```
nlevels(as.factor(activity$date))
```

```
## [1] 53
```

```
nlevels(as.factor(activity$interval))
```

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

adjust date into POSIX

```
class(activity$date)
```

```
## [1] "character"
```

```
# turn date into
library(lubridate)
```

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

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

```
activity$date <- ymd(activity$date)
str(activity)
```

```
## 'data.frame':    15264 obs. of  3 variables:
##  $ steps   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ date    : Date, format: "2012-10-02" "2012-10-02" ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
class(activity$date)
```

```
## [1] "Date"
```

What is the average daily activity pattern?

```
require(dplyr)
```

```
## Loading required package: dplyr
```

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

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

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

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

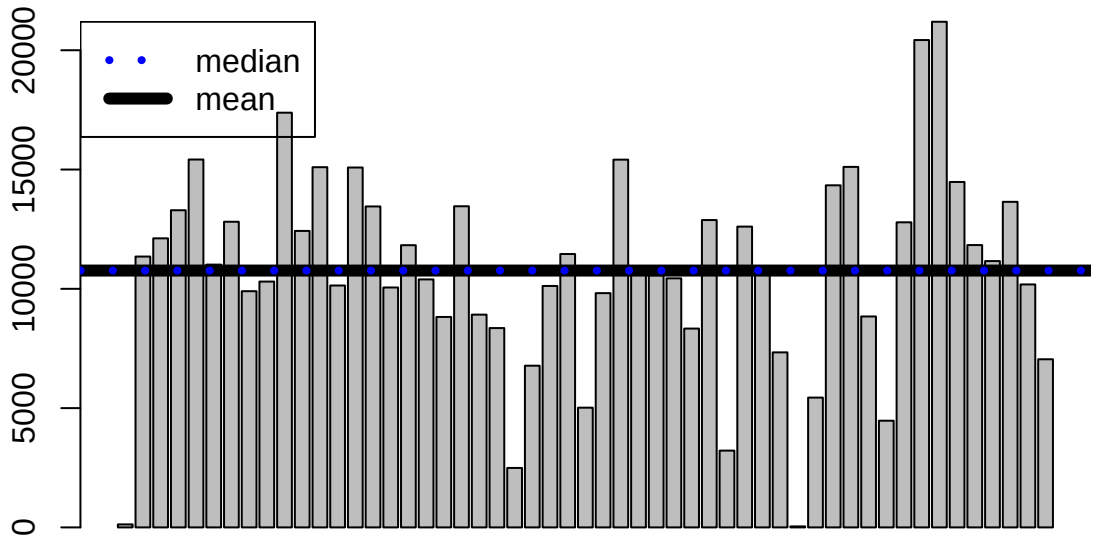
```
# piping usage:
total_day <- activity %>% group_by(date) %>%summarise(total_steps=sum(steps,na.rm=TRUE),na=mean(is.na(s
```

```
## # A tibble: 53 x 3
##    date       total_steps    na
##    <date>           <int> <dbl>
##  1 2012-10-02         126    0.
##  2 2012-10-03       11352    0.
##  3 2012-10-04       12116    0.
```

```
##  4 2012-10-05       13294    0.
##  5 2012-10-06       15420    0.
##  6 2012-10-07       11015    0.
##  7 2012-10-09       12811    0.
##  8 2012-10-10        9900    0.
##  9 2012-10-11       10304    0.
## 10 2012-10-12       17382    0.
## # ... with 43 more rows
```

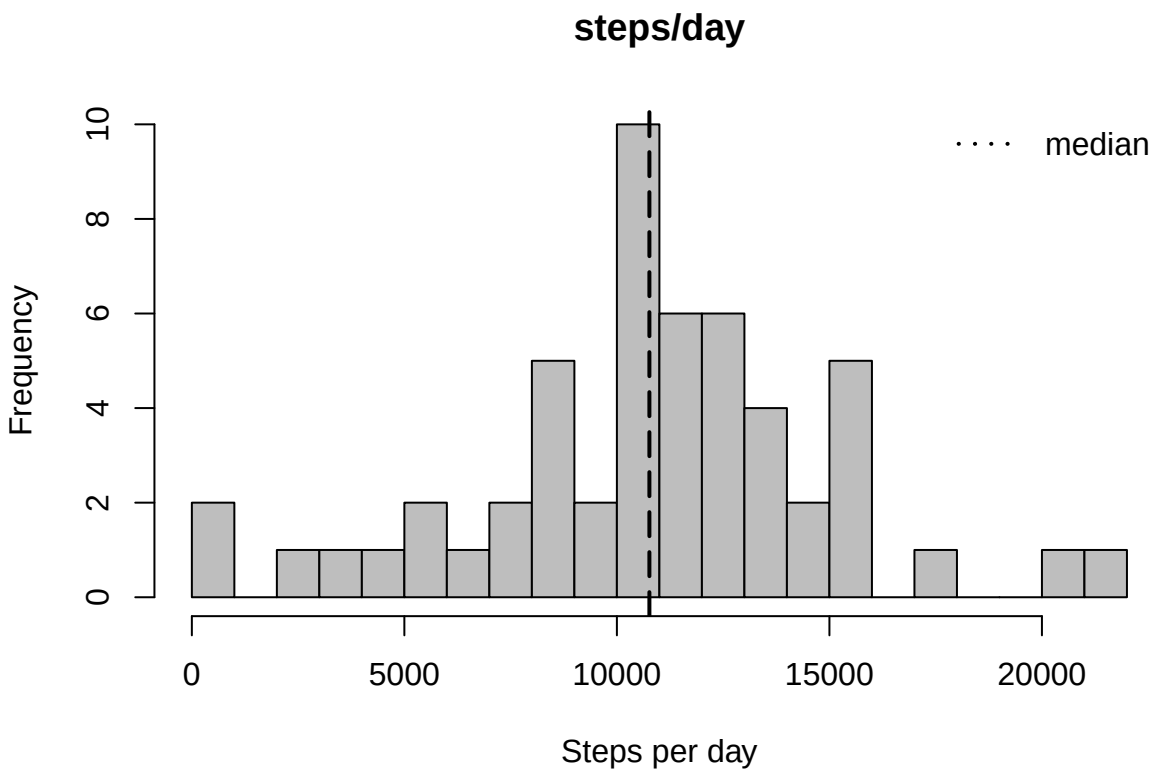
```
mean_steps <- mean(total_day$total_steps,na.rm=TRUE)
median_steps <- median(total_day$total_steps,na.rm=TRUE)
```

```
barplot(height = total_day$total_steps,col="grey")
abline(h=mean(total_day$total_steps), lwd=6, col="black")
abline(h=median(total_day$total_steps), lty=15,lwd=4, col="blue")
legend(legend=c("median","mean"),"topleft",lty=c(15,1),lwd=c(4,6), col=c("blue","black"))
```




```
#legend(legend="mean","topleft", lwd=6, col="black")
```

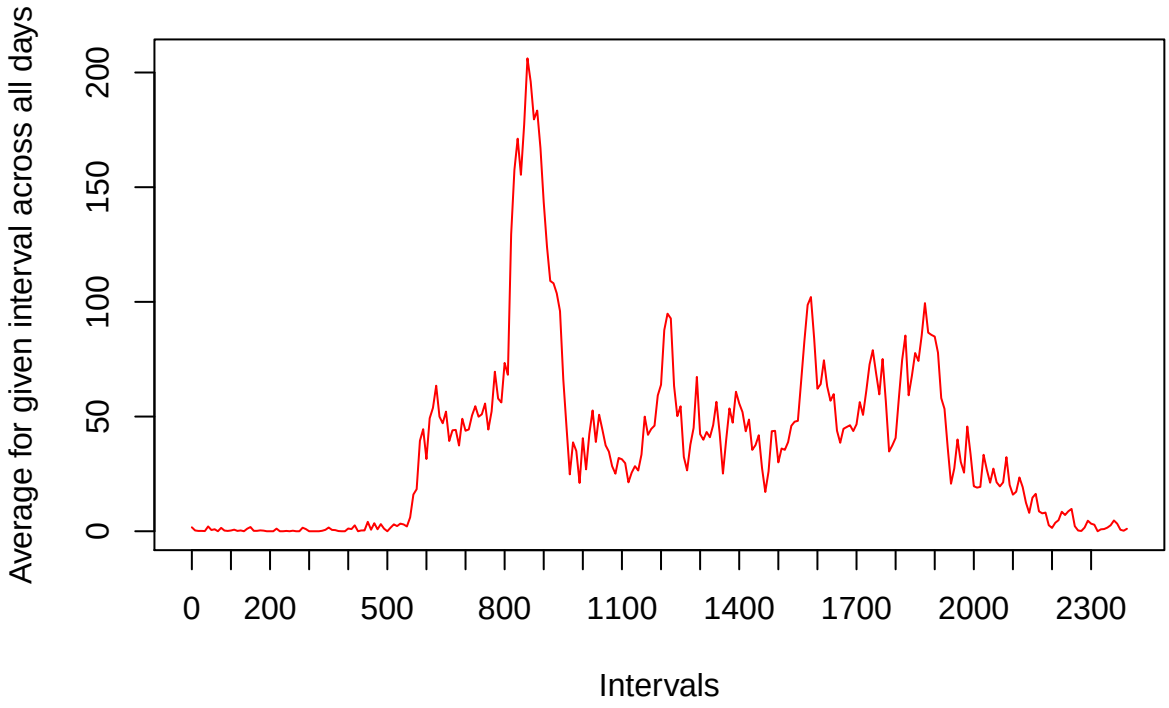
```
total_day <- filter(total_day, na < 1)
hist(total_day$total_steps,col="grey",breaks=20,main="steps/day",xlab="Steps per day")
abline(v=median(total_day$total_steps),lty=5, lwd=2, col="black")
legend(legend="median","topright",lty=3,lwd=2,bty = "n")
```




```
mean_steps <- mean(total_day$total_steps,na.rm=TRUE)
median_steps <- median(total_day$total_steps,na.rm=TRUE)
```

What is the average daily activity pattern?

```
library(dplyr,quietly = TRUE)
daily_patterns <- activity %>% group_by(interval) %>% summarise(average=mean(steps,na.rm=TRUE))
plot(x = 1:nrow(daily_patterns),y = daily_patterns$average,type = "l",
     col = "red", xaxt = "n",xlab="Intervals",
     ylab = "Average for given interval across all days")
axis(1,labels=daily_patterns$interval[seq(1,288,12)],
     at = seq_along(daily_patterns$interval)[seq(1,288,12)])
```




```
max_numb_steps_interval <- filter(daily_patterns,average==max(average))
max_numb_steps_interval
```

```
## # A tibble: 1 x 2
##   interval average
##      <int>   <dbl>
## 1      835    206.
```

Imputing missing data

```
activity2 <- split(activity, activity$interval)
activity2 <- lapply(activity2, function(x) {
        x$steps[which(is.na(x$steps))] <- mean(x$steps, na.rm = TRUE)
        return(x)
})
activity2 <- do.call("rbind", activity2)
row.names(activity2) <- NULL

activity2 <- split(activity2, activity2$date)
df <- lapply(activity2, function(x) {
        x$steps[which(is.na(x$steps))] <- mean(x$steps, na.rm = TRUE)
        return(x)
})
activity2 <- do.call("rbind", activity2)
row.names(activity2) <- NULL
```

```
head(activity2)
```

```
##   steps       date interval
## 1     0 2012-10-02        0
## 2     0 2012-10-02        5
## 3     0 2012-10-02       10
## 4     0 2012-10-02       15
## 5     0 2012-10-02       20
## 6     0 2012-10-02       25
```

Are there differences in activity patterns between weekdays and weekends?

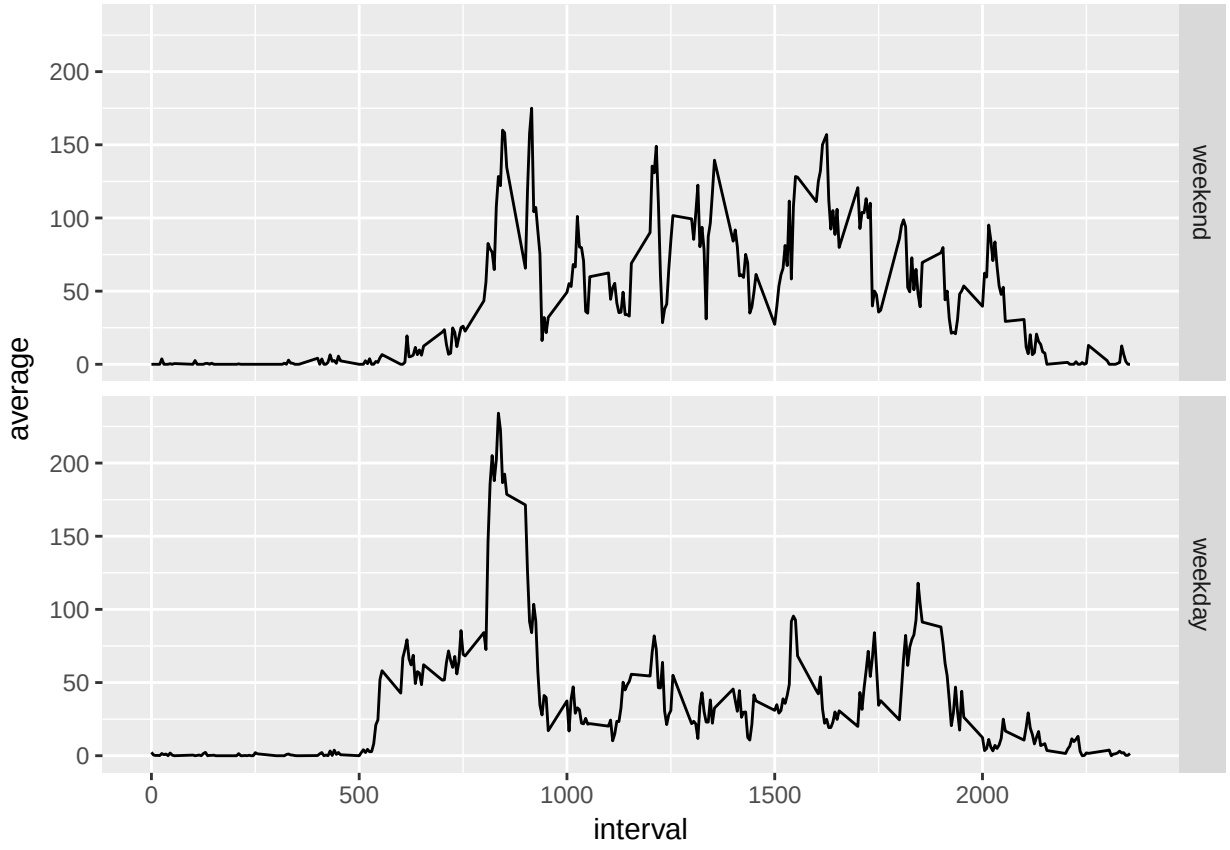
```
library(lubridate)
is_weekday <-function(date){
        if(wday(date)%in%c(1,7)) result<-"weekend"
        else
                result<-"weekday"
        result
}

activity_without_NAs <- mutate(activity_without_NAs,date=ymd(date)) %>% mutate(day=sapply(date,is_weekda

table(activity_without_NAs$day)
```

```
##
## weekday weekend
##   11232    4032
```

```
library(ggplot2)
daily_patterns <- activity_without_NAs %>% mutate(day=factor(day,levels=c("weekend","weekday")),steps_n
qplot(interval,average,data=daily_patterns,geom="line",facets=day~.)
```


average
200
150
100
50
0
weekend
weekday
0
500
1000
1500
2000
interval