PA1\_Template.R

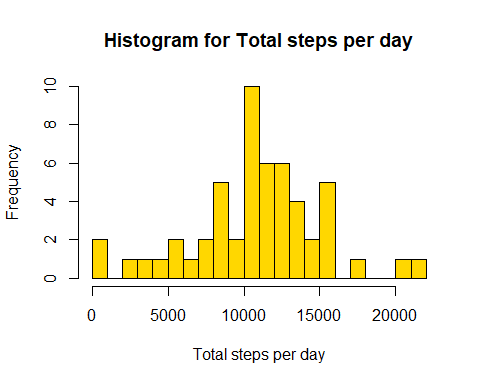
drall

2020-07-19

## RR-project 1 - Abdou Allayeh  
## Data loading  
  
setwd("~/RDataScience/Reproducible Research/Project 1")  
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip", destfile = "activity.zip", mode="wb")  
  
## unzip data and read   
  
unzip("activity.zip")  
Rawdata <- read.csv("activity.csv", header = TRUE)  
head(Rawdata)

## 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

## Calculate total number of steps per day  
  
main\_data <- na.omit(Rawdata)  
steps\_per\_day <- aggregate(main\_data$steps, by = list(Steps.Date = main\_data$date), FUN = "sum")  
  
hist(steps\_per\_day$x, col = "gold",   
 breaks = 20,  
 main = "Histogram for Total steps per day",  
 xlab = "Total steps per day")



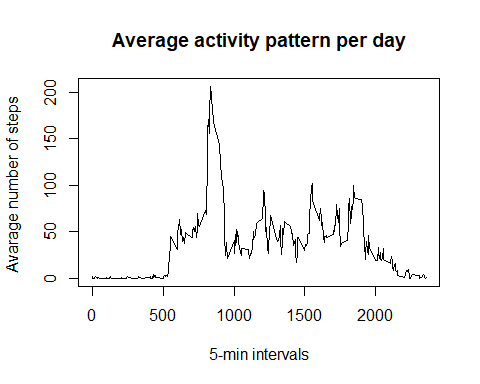
## Calculate the mean and median of all steps per day  
  
main\_mean <- mean(steps\_per\_day[,2])  
print(main\_mean)

## [1] 10766.19

main\_median <- median(steps\_per\_day[,2])  
print (main\_median)

## [1] 10765

## average activity pattern per day  
## Time Plotting for all steps per averaged of all days, along all 5-min intervals  
  
avaraged\_day <- aggregate(main\_data$steps,   
 by = list(Interval = main\_data$interval),   
 FUN = "mean")  
plot(avaraged\_day$Interval, avaraged\_day$x, type = "l",   
 main = "Average activity pattern per day",   
 ylab = "Avarage number of steps ",   
 xlab = "5-min intervals")



## define the interval with the maximum number of steps  
  
interval\_row <- which.max(avaraged\_day$x)  
max\_interval <- avaraged\_day[interval\_row,1]  
print (max\_interval)

## [1] 835

## calculate the total number of NA values  
  
NA\_number <- length(which(is.na(Rawdata$steps)))  
print (NA\_number)

## [1] 2304

## Histogram for new frequencies of all steps  
  
missingVals <- sum(is.na(data))

## Warning in is.na(data): is.na() applied to non-(list or vector) of type  
## 'closure'

library(magrittr)  
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

replacewithmean <- function(x) replace(x, is.na(x), mean(x, na.rm = TRUE))  
meandata <- Rawdata %>% group\_by(interval) %>% mutate(steps = replacewithmean(steps))  
head(meandata)

## # A tibble: 6 x 3  
## # Groups: interval [6]  
## steps date interval  
## <dbl> <chr> <int>  
## 1 1.72 2012-10-01 0  
## 2 0.340 2012-10-01 5  
## 3 0.132 2012-10-01 10  
## 4 0.151 2012-10-01 15  
## 5 0.0755 2012-10-01 20  
## 6 2.09 2012-10-01 25

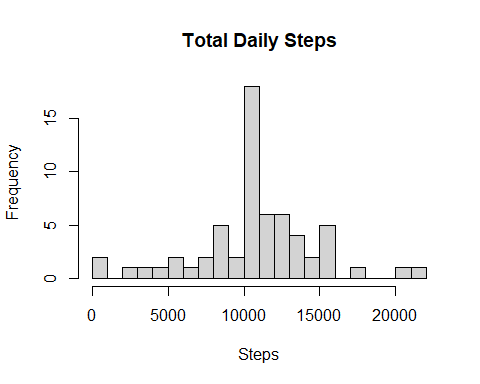
## Histogram of all steps per day   
## Calculate and report the mean and median all steps per day  
  
FullSummedDataByDay <- aggregate(meandata$steps, by=list(meandata$date), sum)  
  
names(FullSummedDataByDay)[1] ="date"  
names(FullSummedDataByDay)[2] ="totalsteps"  
head(FullSummedDataByDay,15)

## date totalsteps  
## 1 2012-10-01 10766.19  
## 2 2012-10-02 126.00  
## 3 2012-10-03 11352.00  
## 4 2012-10-04 12116.00  
## 5 2012-10-05 13294.00  
## 6 2012-10-06 15420.00  
## 7 2012-10-07 11015.00  
## 8 2012-10-08 10766.19  
## 9 2012-10-09 12811.00  
## 10 2012-10-10 9900.00  
## 11 2012-10-11 10304.00  
## 12 2012-10-12 17382.00  
## 13 2012-10-13 12426.00  
## 14 2012-10-14 15098.00  
## 15 2012-10-15 10139.00

## Summary of newest data and making histogram  
  
summary(FullSummedDataByDay)

## date totalsteps   
## Length:61 Min. : 41   
## Class :character 1st Qu.: 9819   
## Mode :character Median :10766   
## Mean :10766   
## 3rd Qu.:12811   
## Max. :21194

hist(FullSummedDataByDay$totalsteps, xlab = "Steps", ylab = "Frequency", main = "Total Daily Steps", breaks = 20)  
  
## Compare between the mean and median of Old and New data  
  
oldmean <- main\_mean  
newmean <- mean(FullSummedDataByDay$totalsteps)  
oldmedian <- main\_median  
newmedian <- median(FullSummedDataByDay$totalsteps)  
  
## differences in activity patterns between weekdays and weekends  
## Plotting for Comparison of average all steps per each interval  
  
meandata$date <- as.Date(meandata$date)  
meandata$weekday <- weekdays(meandata$date)  
meandata$weekend <- ifelse(meandata$weekday=="Saturday" | meandata$weekday=="Sunday", "Weekend", "Weekday" )  
  
library(ggplot2)



meandataweekendweekday <- aggregate(meandata$steps , by= list(meandata$weekend, meandata$interval), na.omit(mean))  
names(meandataweekendweekday) <- c("weekend", "interval", "steps")  
  
ggplot(meandataweekendweekday, aes(x=interval, y=steps, color=weekend)) + geom\_line()+  
 facet\_grid(weekend ~.) + xlab("Interval") + ylab("Mean Steps") +  
 ggtitle("Comparison of Average All Steps per each Interval")

