Reproducible Research: Peer Assessment 1

## Loading and preprocessing the data

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

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

library(data.table)

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

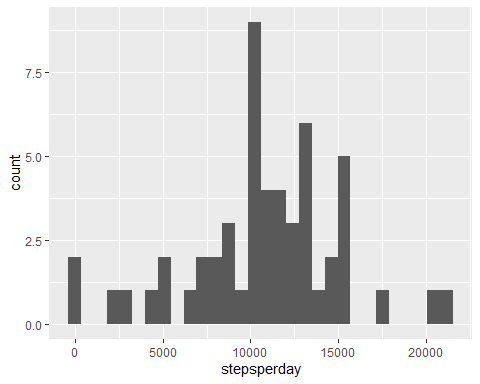
activity<-data.table::fread("./activity.csv")  
  
spd<- aggregate(activity$steps,   
 list(activity$date),  
 FUN=sum)  
  
colnames(spd)<- c("date","stepsperday")

## What is mean total number of steps taken per day?

h <- ggplot(spd, aes(stepsperday))  
h+geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 8 rows containing non-finite values (stat\_bin).



#mode  
cat(paste("\nmean is : ", mean(spd$stepsperday, na.rm = TRUE)))

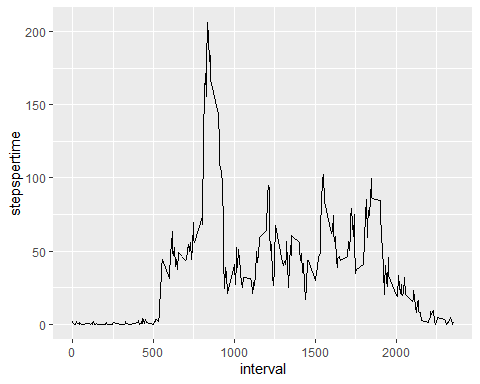
##   
## mean is : 10766.1886792453

#median  
cat(paste("\nmedian is : ", median(spd$stepsperday, na.rm = TRUE)))

##   
## median is : 10765

## What is the average daily activity pattern?

spt<- aggregate(steps~interval,  
 data=activity,  
 FUN=mean,  
 na.acton=na.omit)  
  
colnames(spt)<- c("interval","stepspertime")  
  
t <- ggplot(spt, aes(interval, stepspertime))  
t+geom\_line()



spt %>%  
select(interval, stepspertime) %>%  
filter(stepspertime==max(spt$stepspertime))

## interval stepspertime  
## 1 835 206.1698

## Imputing missing values

print(paste("total NA is : ", activity[is.na(steps), .N ]))

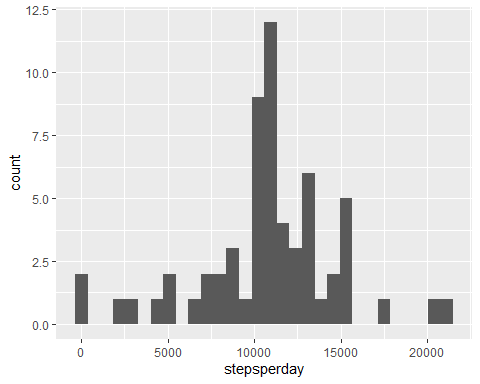
## [1] "total NA is : 2304"

activity[is.na(steps),"steps"]<-activity[,c(lapply(.SD, mean, na.rm = TRUE,)), .SDcols = c("steps")]

## Warning in `[<-.data.table`(`\*tmp\*`, is.na(steps), "steps", value =  
## structure(list(: 37.382600 (type 'double') at RHS position 1 truncated  
## (precision lost) when assigning to type 'integer' (column 1 named 'steps')

#data.table::fwrite(activity,"./activity-fillna-median.csv")  
  
fillactivity<-copy(activity)  
  
tspd<- aggregate(activity$steps,   
 list(activity$date),  
 FUN=sum)  
  
colnames(tspd)<- c("date","stepsperday")  
  
th <- ggplot(tspd, aes(stepsperday))  
th+geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



cat(paste("\nmean is : ", mean(tspd$stepsperday)))

##   
## mean is : 10751.737704918

cat(paste("\nmedian is : ", median(tspd$stepsperday)))

##   
## median is : 10656

## Are there differences in activity patterns between weekdays and weekends?

fillactivity$wkday<-weekdays(fillactivity$date)  
  
fillactivity$wkORwkend<-ifelse(fillactivity$wkday=="Saturday"|fillactivity$wkday=="Sunday","weekend","weekday")  
  
fspt<- aggregate(steps~interval+wkORwkend,  
 data=fillactivity,  
 FUN=mean)  
  
colnames(fspt)<- c("interval","wkORwkend","stepspertime")  
  
ft <- ggplot(fspt, aes(interval, stepspertime))  
ft+geom\_line()+facet\_wrap(wkORwkend ~ .)

