Submitting Project Week 2

Oscar Alvarado

2/9/2020

## Loading and preprocessing the data

library(lubridate)

##   
## Attaching package: 'lubridate'

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

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)  
library(latexpdf)  
  
  
urldata<- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip"  
download.file(urldata, "stepsdata.zip")  
unzip("stepsdata.zip")  
stepsdata<-read.csv("activity.csv")  
head(stepsdata)

## 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

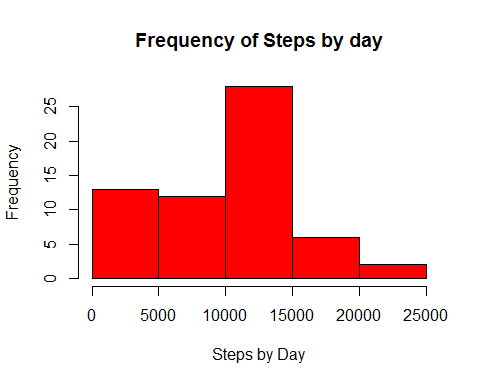
stepsdata$date<- ymd(stepsdata$date)

## Histrogram, mean and median of Steps by Day

stepsdataday<-stepsdata%>%group\_by(date)%>%summarise(steps.by.day=sum(steps, na.rm = T))

## `summarise()` ungrouping output (override with `.groups` argument)

with(stepsdataday, hist(steps.by.day, xlab="Steps by Day", main = "Frequency of Steps by day", col = "red"))



mean(stepsdataday$steps.by.day, na.rm = T)

## [1] 9354.23

median(stepsdataday$steps.by.day, na.rm = T)

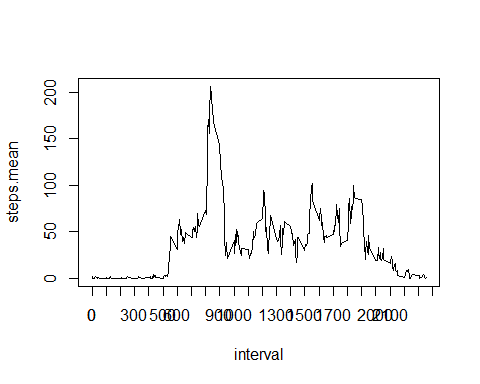
## [1] 10395

## What is the average daily activity pattern?

intervaldata<-stepsdata%>%group\_by(interval)%>%summarise(steps.mean=mean(steps, na.rm = T))

## `summarise()` ungrouping output (override with `.groups` argument)

with(intervaldata, plot(interval, steps.mean, type="l"))  
axis(side=1, at= seq(0,2500, by=100))



intervaldata[which.max(intervaldata$steps.mean),]

## # A tibble: 1 x 2  
## interval steps.mean  
## <int> <dbl>  
## 1 835 206.

## Imputing missing values

initialmean<-tapply(stepsdata$steps, stepsdata$interval, mean, na.rm=T)  
sum(!complete.cases(stepsdata))

## [1] 2304

## Replacing NAs by The median of each interval  
stepsdataNoNA<-stepsdata  
for(i in seq(from=0, to=2355, by=5)){  
 stepsdataNoNA$steps[is.na(stepsdataNoNA$steps) & stepsdataNoNA$interval==i]<-mean(stepsdataNoNA$steps[stepsdataNoNA$interval==i], na.rm = T)  
}  
sum(!complete.cases(stepsdataNoNA))

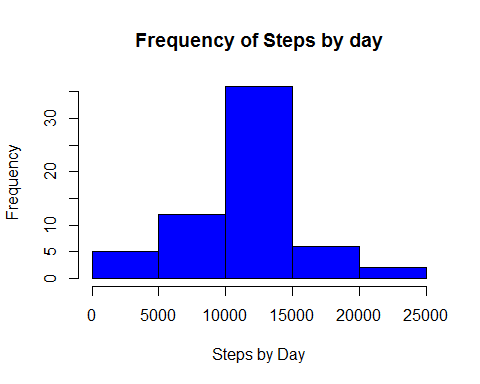
## [1] 0

### Histrogram, mean and median of Steps by Day without NAs

stepsdatadayNoNA<-stepsdataNoNA%>%group\_by(date)%>%summarise(steps.by.day=sum(steps, na.rm = T))

## `summarise()` ungrouping output (override with `.groups` argument)

with(stepsdatadayNoNA, hist(steps.by.day, xlab="Steps by Day", main = "Frequency of Steps by day", col = "blue"))



mean(stepsdatadayNoNA$steps.by.day, na.rm = T)

## [1] 10766.19

median(stepsdatadayNoNA$steps.by.day, na.rm = T)

## [1] 10766.19

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

stepsdataNoNA$weekday<- weekdays(stepsdataNoNA$date)  
stepsdataNoNA$WeekGroup<- ifelse(stepsdataNoNA$weekday%in%c("sábado", "domingo"), "weekend","weekday")  
Groups<-stepsdataNoNA%>%group\_by(WeekGroup, interval)%>%summarise(mean.steps.by.day=mean(steps))

## `summarise()` regrouping output by 'WeekGroup' (override with `.groups` argument)

ggplot(Groups, aes(interval, mean.steps.by.day))+facet\_wrap(.~WeekGroup)+geom\_line()+theme\_bw()

