Activity Assignment

Reading in the data.frame

data0 <- na.omit(read.csv("./data/activity.csv")) ##reading in data  
data0$date <- as.Date(data0$date)  
str(data0)

## '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 ...  
## - attr(\*, "na.action")= 'omit' Named int [1:2304] 1 2 3 4 5 6 7 8 9 10 ...  
## ..- attr(\*, "names")= chr [1:2304] "1" "2" "3" "4" ...

head(data0, n=20)

## steps date interval  
## 289 0 2012-10-02 0  
## 290 0 2012-10-02 5  
## 291 0 2012-10-02 10  
## 292 0 2012-10-02 15  
## 293 0 2012-10-02 20  
## 294 0 2012-10-02 25  
## 295 0 2012-10-02 30  
## 296 0 2012-10-02 35  
## 297 0 2012-10-02 40  
## 298 0 2012-10-02 45  
## 299 0 2012-10-02 50  
## 300 0 2012-10-02 55  
## 301 0 2012-10-02 100  
## 302 0 2012-10-02 105  
## 303 0 2012-10-02 110  
## 304 0 2012-10-02 115  
## 305 0 2012-10-02 120  
## 306 0 2012-10-02 125  
## 307 0 2012-10-02 130  
## 308 0 2012-10-02 135

Total Steps per day

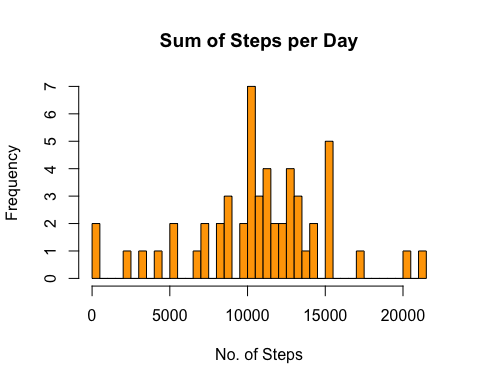
s <- split(data0, data0$date)   
 sapply(s, function(x) colSums(x[, c("steps", "interval")], na.rm = TRUE)) ##calculates total steps for eacy day.

## 2012-10-02 2012-10-03 2012-10-04 2012-10-05 2012-10-06 2012-10-07  
## steps 126 11352 12116 13294 15420 11015  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-10-09 2012-10-10 2012-10-11 2012-10-12 2012-10-13 2012-10-14  
## steps 12811 9900 10304 17382 12426 15098  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-10-15 2012-10-16 2012-10-17 2012-10-18 2012-10-19 2012-10-20  
## steps 10139 15084 13452 10056 11829 10395  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-10-21 2012-10-22 2012-10-23 2012-10-24 2012-10-25 2012-10-26  
## steps 8821 13460 8918 8355 2492 6778  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-10-27 2012-10-28 2012-10-29 2012-10-30 2012-10-31 2012-11-02  
## steps 10119 11458 5018 9819 15414 10600  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-11-03 2012-11-05 2012-11-06 2012-11-07 2012-11-08 2012-11-11  
## steps 10571 10439 8334 12883 3219 12608  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-11-12 2012-11-13 2012-11-15 2012-11-16 2012-11-17 2012-11-18  
## steps 10765 7336 41 5441 14339 15110  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-11-19 2012-11-20 2012-11-21 2012-11-22 2012-11-23 2012-11-24  
## steps 8841 4472 12787 20427 21194 14478  
## interval 339120 339120 339120 339120 339120 339120  
## 2012-11-25 2012-11-26 2012-11-27 2012-11-28 2012-11-29  
## steps 11834 11162 13646 10183 7047  
## interval 339120 339120 339120 339120 339120

##you can also use aggregate  
data1 <- aggregate(data0$steps, by = list(Steps.Date = data0$date), FUN ="sum")  
data1 <- as.data.frame(data1)  
colnames(data1) <- c("Date", "Sum\_Steps")  
head(data1)

## Date Sum\_Steps  
## 1 2012-10-02 126  
## 2 2012-10-03 11352  
## 3 2012-10-04 12116  
## 4 2012-10-05 13294  
## 5 2012-10-06 15420  
## 6 2012-10-07 11015

hist(data1$Sum\_Steps, col= "orange", breaks=36, xlab="No. of Steps", main= "Sum of Steps per Day")



mean\_step <- mean(data1$Sum\_Steps)  
mean\_step

## [1] 10766.19

median\_step <- median(data1$Sum\_Steps)  
median\_step

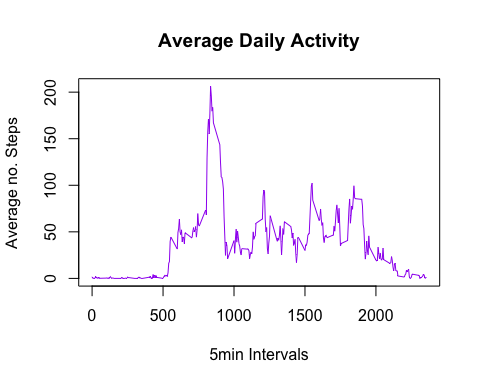
## [1] 10765

Plotting Average Steps

data2 <- aggregate(data0$steps, by = list(Interval = data0$interval), FUN ="mean")  
colnames(data2) <- c("Interval", "Average\_Steps")  
head(data2)

## Interval Average\_Steps  
## 1 0 1.7169811  
## 2 5 0.3396226  
## 3 10 0.1320755  
## 4 15 0.1509434  
## 5 20 0.0754717  
## 6 25 2.0943396

with(data2, plot(Interval, Average\_Steps, col = "purple",   
 main= "Average Daily Activity",   
 xlab= "5min Intervals",   
 ylab="Average no. Steps",   
 type= "l"))



max <- which.max(data2$Average\_Steps)  
max2 <- data2[max,1]  
max2

## [1] 835

Working with NA’s

dataNA <- read.csv("./data/activity.csv")  
length(which(is.na(dataNA))) ##number of NA's

## [1] 2304

library(Hmisc)

## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

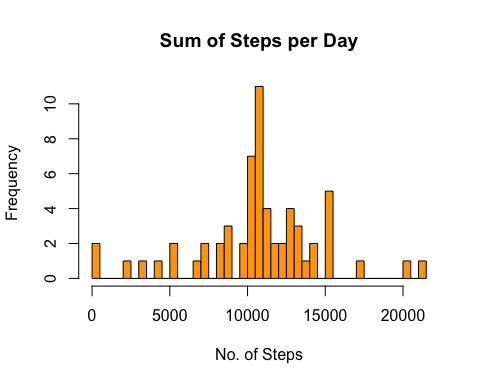
##   
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':  
##   
## format.pval, units

dataFill <- dataNA  
 dataFill$steps <- impute(dataNA$steps, fun=mean)  
  
  
data3 <- aggregate(dataFill$steps, by = list(Date = dataFill$date), FUN ="sum")  
data3 <- as.data.frame(data3)  
 colnames(data3) <- c("Date", "Sum\_Steps")  
 head(data3)

## Date Sum\_Steps  
## 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

hist(data3$Sum\_Steps, col= "orange", breaks=36, xlab="No. of Steps", main= "Sum of Steps per Day")



mean\_step <- mean(data3$Sum\_Steps)  
mean\_step

## [1] 10766.19

median\_step <- median(data3$Sum\_Steps)  
median\_step

## [1] 10766.19

Weekday and Weekend

dataFill$date <- as.Date(dataFill$date)  
dataFill$weekday <- weekdays(dataFill$date)  
dataFill$weekend <- ifelse(dataFill$weekday=="Saturday" | dataFill$weekday=="Sunday", "Weekend", "Weekday" )  
head(dataFill)

## steps date interval weekday weekend  
## 1 37.3826 2012-10-01 0 Monday Weekday  
## 2 37.3826 2012-10-01 5 Monday Weekday  
## 3 37.3826 2012-10-01 10 Monday Weekday  
## 4 37.3826 2012-10-01 15 Monday Weekday  
## 5 37.3826 2012-10-01 20 Monday Weekday  
## 6 37.3826 2012-10-01 25 Monday Weekday

library(ggplot2)  
data4 <- aggregate(dataFill$steps , by= list(dataFill$weekend, dataFill$interval), FUN = "mean") ##aggregating by interval, and weekend code.  
data4 <- as.data.frame(data4)  
 colnames(data4) <- c("Weekend", "Interval", "Mean\_Steps")  
  
ggplot(data4, aes(x=Interval, y=Mean\_Steps, color=Weekend)) +   
 geom\_line()+  
 facet\_grid(Weekend ~.) +  
 xlab("Interval") +   
 ylab("Mean Steps") +  
 ggtitle(" Average Number of Steps in Each Interval by Weekday/Weekend")

