PA1\_template.R

TCTKASAPOGULLARI

Mon May 18 01:31:16 2015

setwd("D:/Data Scientist/Reproducible Reasearch")  
  
library(knitr)

## Warning: package 'knitr' was built under R version 3.1.3

opts\_chunk$set(echo = TRUE, results = 'hold')  
  
library(ggplot2) # we shall use ggplot2 for plotting figures

## Warning: package 'ggplot2' was built under R version 3.1.3

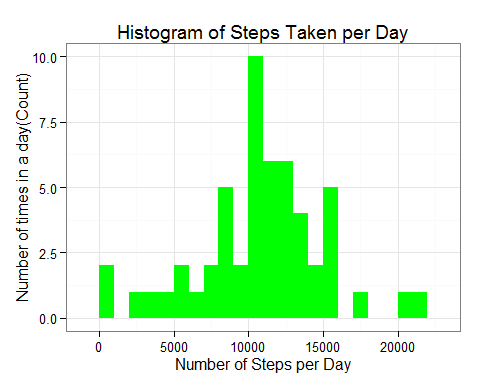
rdata <- read.csv('activity.csv', header = TRUE, sep = ",",  
 colClasses=c("numeric", "character", "numeric"))  
  
rdata$date <- as.Date(rdata$date, format = "%Y-%m-%d")  
rdata$interval <- as.factor(rdata$interval)  
  
str(rdata)

## 'data.frame': 17568 obs. of 3 variables:  
## $ steps : num NA NA NA NA NA NA NA NA NA NA ...  
## $ date : Date, format: "2012-10-01" "2012-10-01" ...  
## $ interval: Factor w/ 288 levels "0","5","10","15",..: 1 2 3 4 5 6 7 8 9 10 ...

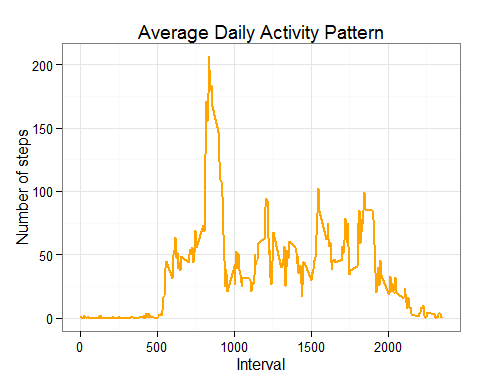
steps\_per\_day <- aggregate(steps ~ date, rdata, sum)  
colnames(steps\_per\_day) <- c("date","steps")  
head(steps\_per\_day)

## date 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

ggplot(steps\_per\_day, aes(x = steps)) +   
 geom\_histogram(fill = "green", binwidth = 1000) +   
 labs(title="Histogram of Steps Taken per Day",   
 x = "Number of Steps per Day", y = "Number of times in a day(Count)") + theme\_bw()



steps\_mean <- mean(steps\_per\_day$steps, na.rm=TRUE)  
steps\_median <- median(steps\_per\_day$steps, na.rm=TRUE)  
  
steps\_per\_interval <- aggregate(rdata$steps,   
 by = list(interval = rdata$interval),  
 FUN=mean, na.rm=TRUE)  
#convert to integers  
##this helps in plotting  
steps\_per\_interval$interval <-   
 as.integer(levels(steps\_per\_interval$interval)[steps\_per\_interval$interval])  
colnames(steps\_per\_interval) <- c("interval", "steps")  
  
ggplot(steps\_per\_interval, aes(x=interval, y=steps)) +   
 geom\_line(color="orange", size=1) +   
 labs(title="Average Daily Activity Pattern", x="Interval", y="Number of steps") +   
 theme\_bw()



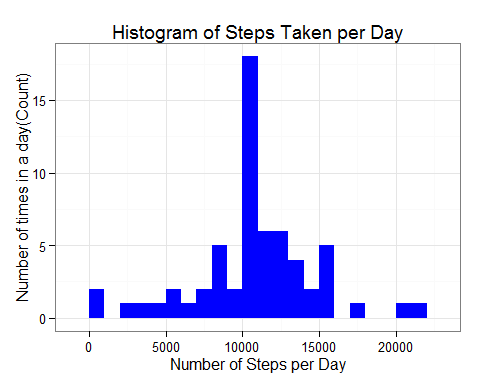
max\_interval <- steps\_per\_interval[which.max(   
 steps\_per\_interval$steps),]  
  
missing\_vals <- sum(is.na(rdata$steps))  
  
  
na\_fill <- function(data, pervalue) {  
 na\_index <- which(is.na(data$steps))  
 na\_replace <- unlist(lapply(na\_index, FUN=function(idx){  
 interval = data[idx,]$interval  
 pervalue[pervalue$interval == interval,]$steps  
 }))  
 fill\_steps <- data$steps  
 fill\_steps[na\_index] <- na\_replace  
 fill\_steps  
}  
  
rdata\_fill <- data.frame(   
 steps = na\_fill(rdata, steps\_per\_interval),   
 date = rdata$date,   
 interval = rdata$interval)  
str(rdata\_fill)

## 'data.frame': 17568 obs. of 3 variables:  
## $ steps : num 1.717 0.3396 0.1321 0.1509 0.0755 ...  
## $ date : Date, format: "2012-10-01" "2012-10-01" ...  
## $ interval: Factor w/ 288 levels "0","5","10","15",..: 1 2 3 4 5 6 7 8 9 10 ...

sum(is.na(rdata\_fill$steps))

## [1] 0

fill\_steps\_per\_day <- aggregate(steps ~ date, rdata\_fill, sum)  
colnames(fill\_steps\_per\_day) <- c("date","steps")  
  
##plotting the histogram  
ggplot(fill\_steps\_per\_day, aes(x = steps)) +   
 geom\_histogram(fill = "blue", binwidth = 1000) +   
 labs(title="Histogram of Steps Taken per Day",   
 x = "Number of Steps per Day", y = "Number of times in a day(Count)") + theme\_bw()



steps\_mean\_fill <- mean(fill\_steps\_per\_day$steps, na.rm=TRUE)  
steps\_median\_fill <- median(fill\_steps\_per\_day$steps, na.rm=TRUE)  
  
weekdays\_steps <- function(data) {  
 weekdays\_steps <- aggregate(data$steps, by=list(interval = data$interval),  
 FUN=mean, na.rm=T)  
 # convert to integers for plotting  
 weekdays\_steps$interval <-   
 as.integer(levels(weekdays\_steps$interval)[weekdays\_steps$interval])  
 colnames(weekdays\_steps) <- c("interval", "steps")  
 weekdays\_steps  
}  
  
data\_by\_weekdays <- function(data) {  
 data$weekday <-   
 as.factor(weekdays(data$date)) # weekdays  
 weekend\_data <- subset(data, weekday %in% c("Cumartesi","Pazar"))  
 weekday\_data <- subset(data, !weekday %in% c("Cumartesi","Pazar"))  
   
 weekend\_steps <- weekdays\_steps(weekend\_data)  
 weekday\_steps <- weekdays\_steps(weekday\_data)  
   
 weekend\_steps$dayofweek <- rep("weekend", nrow(weekend\_steps))  
 weekday\_steps$dayofweek <- rep("weekday", nrow(weekday\_steps))  
   
 data\_by\_weekdays <- rbind(weekend\_steps, weekday\_steps)  
 data\_by\_weekdays$dayofweek <- as.factor(data\_by\_weekdays$dayofweek)  
 data\_by\_weekdays  
}  
  
data\_weekdays <- data\_by\_weekdays(rdata\_fill)  
  
  
ggplot(data\_weekdays, aes(x=interval, y=steps)) +   
 geom\_line(color="violet") +   
 facet\_wrap(~ dayofweek, nrow=2, ncol=1) +  
 labs(x="Interval", y="Number of steps") +  
 theme\_bw()

