Reproducible\_CourseraHwk1.R

trosati

Sat Mar 5 21:24:07 2016

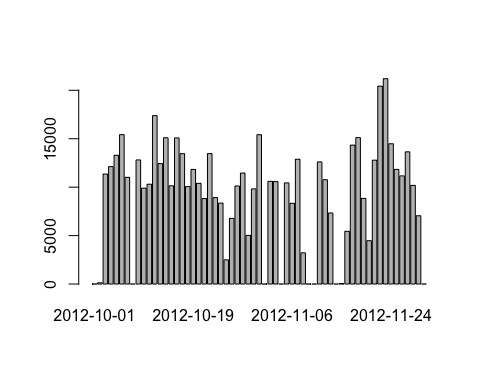
# Get the data from my working directory  
rep <- read.csv("activity.csv")  
  
# make date column Dates instead of Factor  
rep$date <- as.character(rep$date)  
rep$dates <- as.Date(rep$date, "%Y-%m-%d")  
  
# remove NAs  
data<-na.omit(rep)  
  
# Histogram of the total number of steps taken each day  
library(ggplot2)  
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(knitr)  
stepsByDay <- tapply(rep$steps, rep$date, sum, na.rm=TRUE)  
barplot(stepsByDay)



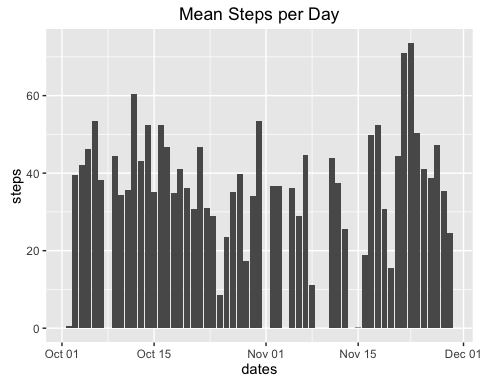
#Mean and Median number of steps taken each day  
mean(stepsByDay)

## [1] 9354.23

# answer = 9354.23  
median(stepsByDay)

## [1] 10395

# answer = 10395  
  
#Time series plot of the average number of steps taken  
meanplot <-ggplot(data = data, aes(dates, steps)) +  
 stat\_summary(fun.y = mean, geom = "bar")  
meanplot+ggtitle("Mean Steps per Day")



#The 5-minute interval that, on average, contains the maximum number of steps  
stepsByInterval <- tapply(data$steps, data$interval, mean, na.rm=TRUE)  
summary(stepsByInterval)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 2.486 34.110 37.380 52.830 206.200

sort(stepsByInterval)[length(stepsByInterval)]

## 835   
## 206.1698

#answer is interval 835  
  
#Code to describe and show a strategy for imputing missing data  
library(zoo)

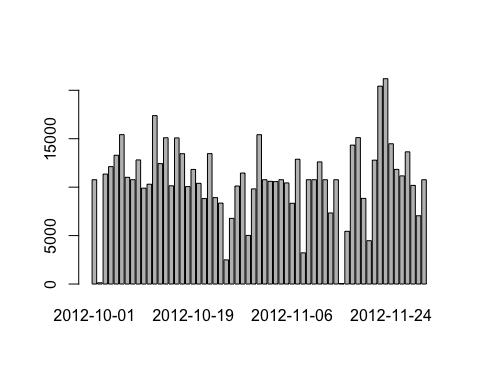
##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

rep$steps<-na.aggregate(rep$steps)  
summary(rep)

## steps date interval dates   
## Min. : 0.00 Length:17568 Min. : 0.0 Min. :2012-10-01   
## 1st Qu.: 0.00 Class :character 1st Qu.: 588.8 1st Qu.:2012-10-16   
## Median : 0.00 Mode :character Median :1177.5 Median :2012-10-31   
## Mean : 37.38 Mean :1177.5 Mean :2012-10-31   
## 3rd Qu.: 37.38 3rd Qu.:1766.2 3rd Qu.:2012-11-15   
## Max. :806.00 Max. :2355.0 Max. :2012-11-30

#Histogram of the total number of steps taken each day after missing values are imputed  
stepsByDayNA <- tapply(rep$steps, rep$date, sum, na.rm=TRUE)  
barplot(stepsByDayNA)



#Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends  
rep$week <- weekdays(rep$dates)  
rep <- mutate(rep, weekpart = ifelse(week == "Saturday" | week=="Sunday", "weekend", "weekday"))  
str(rep)

## 'data.frame': 17568 obs. of 6 variables:  
## $ steps : num 37.4 37.4 37.4 37.4 37.4 ...  
## $ date : chr "2012-10-01" "2012-10-01" "2012-10-01" "2012-10-01" ...  
## $ interval: int 0 5 10 15 20 25 30 35 40 45 ...  
## $ dates : Date, format: "2012-10-01" "2012-10-01" ...  
## $ week : chr "Monday" "Monday" "Monday" "Monday" ...  
## $ weekpart: chr "weekday" "weekday" "weekday" "weekday" ...

rep$weekpart <- as.factor(rep$weekpart)  
  
byDay<-aggregate(steps ~ interval+weekpart, data = rep, mean)  
head(byDay)

## interval weekpart steps  
## 1 0 weekday 7.006569  
## 2 5 weekday 5.384347  
## 3 10 weekday 5.139902  
## 4 15 weekday 5.162124  
## 5 20 weekday 5.073235  
## 6 25 weekday 6.295458

ggplot(byDay, aes(interval, steps)) +   
 geom\_line() +   
 facet\_grid(weekpart ~ .) +  
 xlab("5-minute Interval") +   
 ylab("Mean Steps")

