##---

##title: "Reproducible Research: Peer Assessment 1"

##output:

## html\_document:

## keep\_md: true

##---

## Loading and preprocessing the data

#load data

activity <- NULL

activity <- read.csv("activity.csv", header = T, sep = ",")

#process data into suitable format for analysis

activity$timeanddate <- as.POSIXct(

with(

activity,

paste(

date,

paste(interval %/% 100, interval %% 100, sep=":"))

),

format="%Y-%m-%d %H:%M",tz="")

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

#libraries needed for plots

library(ggplot2)

library(scales)

library(grid)

library(gridExtra)

#calculating data for histogram

stepperday<- setNames(

aggregate(

steps~as.Date(date),

activity,

sum,

na.rm = TRUE),

c("date","steps")

)

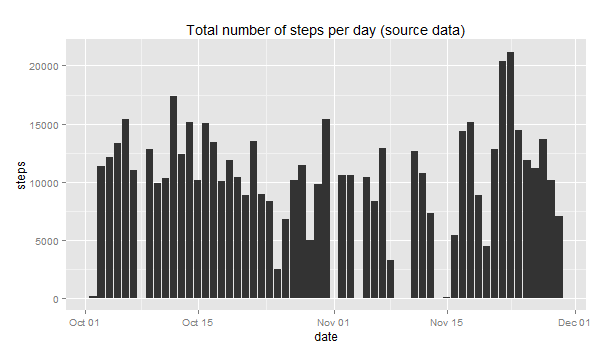
#plotting histogram

hist1 <- ggplot(stepperday,aes(x=date,y=steps)) +

geom\_bar(stat="identity") +

ggtitle("Total number of steps per day (source data)")

print(hist1)



#find mean and media

meanandmed <- c(mean = mean(stepperday$steps),median = median(stepperday$steps))

print(meanandmed)

mean median

10766.19 10765.00

## What is the average daily activity pattern?

#time series plot

avg <- aggregate(steps~interval,activity,mean,na.rm = TRUE)

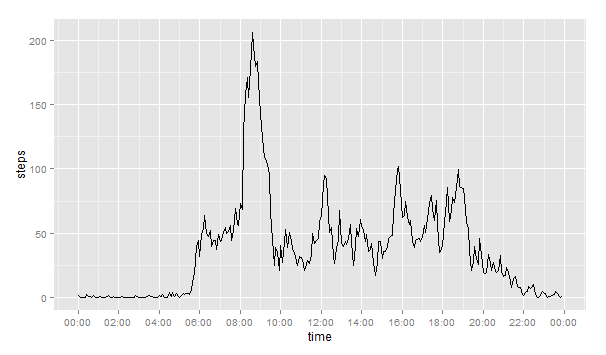
avg$time <- as.POSIXct(with(avg,paste(interval %/% 100, interval %% 100, sep=":")),format="%H:%M")

plot <- ggplot(avg,aes(x=time,y=steps)) +

geom\_line() +

scale\_x\_datetime(breaks = date\_breaks("2 hour"),labels = date\_format("%H:%M"))

print(plot)



#max number of steps

with(avg,avg[steps == max(steps),])

interval steps time

104 835 206.1698 2014-12-15 08:35:00

## Imputing missing values

#number of missing values

NAvalue <- aggregate(cnt~date,cbind(activity[is.na(activity$steps),],cnt=c(1)),sum,na.rm = FALSE)

NAvalue$dow <- weekdays(as.Date(NAvalue$date),abbreviate=TRUE)

print(NAvalue[,c(1,3,2)])

date dow cnt

1 2012-10-01 Mon 288

2 2012-10-08 Mon 288

3 2012-11-01 Thu 288

4 2012-11-04 Sun 288

5 2012-11-09 Fri 288

6 2012-11-10 Sat 288

7 2012-11-14 Wed 288

8 2012-11-30 Fri 288

#filling in NA values

unique(NAvalue$dow)

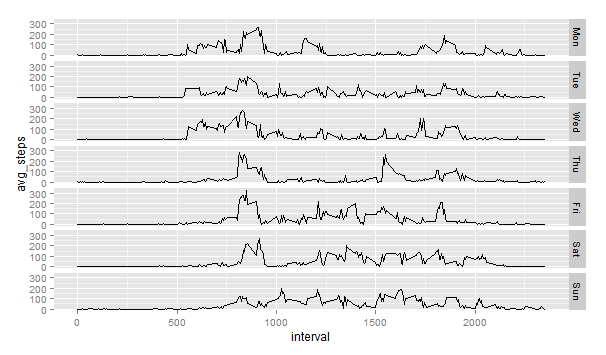
"Mon" "Thu" "Sun" "Fri" "Sat" "Wed"

newset <- aggregate(steps~interval+weekdays(timeanddate,abbreviate=TRUE),activity,FUN=mean,na.rm=TRUE)

colnames(newset) <- c("interval","dow","avg\_steps")

newset$dow <- factor(newset$dow,levels = c("Mon","Tue","Wed","Thu","Fri","Sat","Sun"))

ggplot(newset,aes(x=interval,y=avg\_steps)) + geom\_line() + facet\_grid("dow ~ .")



#full data set

activity$dow <- weekdays(activity$timeanddate,abbreviate=TRUE)

merged <- merge(activity,newset,by=c("dow","interval"),all.x = TRUE)

merged <- merged[with(merged,order(date,interval)),]

merged$fixed\_steps <- ifelse(is.na(merged$steps),merged$avg\_steps,merged$steps)

#histogram of total steps each day, mean and median

# calculating data set for histogram

NEWstepperday <- setNames(

aggregate(

fixed\_steps~as.Date(date),

merged,

sum,

na.rm = TRUE),

c("date","steps")

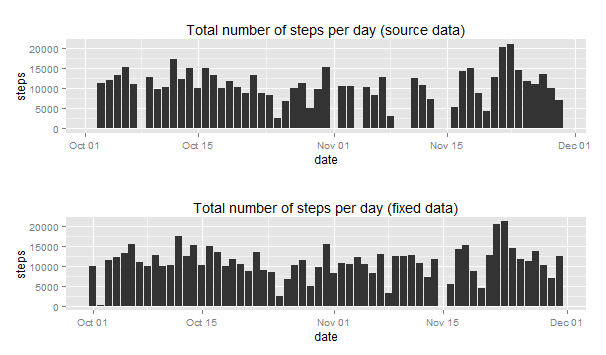
)

# plotting histogram

hist2 <- ggplot(NEWstepperday,aes(x=date,y=steps)) +

geom\_bar(stat="identity") +

ggtitle("Total number of steps per day (fixed data)")



# combining with previous

grid.arrange(hist1, hist2, nrow=2)

NEWmeanandmed <- c(mean = mean(NEWstepperday$steps),median = median(NEWstepperday$steps))

difference <- rbind(source = meanandmed, fixed = NEWmeanandmed, delta = NEWmeanandmed-meanandmed)

print(difference)

mean median

source 10766.18868 10765

fixed 10821.20960 11015

delta 55.02092 250

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

#create factor vairable

week\_diff <- aggregate(

steps~dow+interval, # group steps by weekend/weekday and interval to find average steps

with(

activity,

data.frame(

dow = factor(

ifelse(

weekdays(as.Date(date)) %in% c("Sunday","Saturday"),

"weekend", # if sunday or saturday

"weekday" # else

)

),

interval,

steps

)

),

FUN = mean,

rm.na = TRUE

)

#time series

ggplot(week\_diff,aes(x=interval,y=steps)) + geom\_line() + facet\_grid("dow ~ .")

