Activity\_Steps

Victor Campos

20/1/2022

It is now possible to collect a large amount of data about personal movement using activity monitoring devices such as a Fitbit, Nike Fuelband, or Jawbone Up. These type of devices are part of the “quantified self” movement – a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. But these data remain under-utilized both because the raw data are hard to obtain and there is a lack of statistical methods and software for processing and interpreting the data.

This assignment makes use of data from a personal activity monitoring device. This device collects data at 5 minute intervals through out the day. The data consists of two months of data from an anonymous individual collected during the months of October and November, 2012 and include the number of steps taken in 5 minute intervals each day.

The data for this assignment can be downloaded from the course web site:

Dataset: Activity monitoring data [52K]

The variables included in this dataset are:

steps: Number of steps taking in a 5-minute interval (missing values are coded as NA)

date: The date on which the measurement was taken in YYYY-MM-DD format

interval: Identifier for the 5-minute interval in which measurement was taken

The dataset is stored in a comma-separated-value (CSV) file and there are a total of 17,568 observations in this dataset.

Full submission

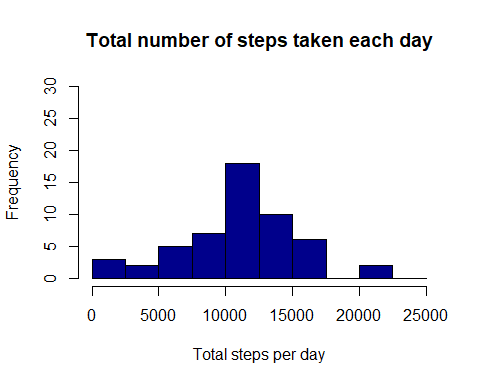
1. Code for reading in the data set and/or processing the data

library(knitr)

## Warning: package 'knitr' was built under R version 4.1.1

setwd("C:/Project Files/CURSOS/DataScience-Hopkins/ReproducibleResearch\_Week 2\_Course Project\_1")

activitydata <- read.csv("C:/Project Files/CURSOS/DataScience-Hopkins/ReproducibleResearch\_Week 2\_Course Project\_1/activity.csv")  
StepsPerDay <- aggregate(steps~date, activitydata, sum)  
hist(StepsPerDay$steps, col = "darkblue", xlab = "Total steps per day", ylim = c(0,30), main = "Total number of steps taken each day", breaks = seq(0,25000,by=2500))

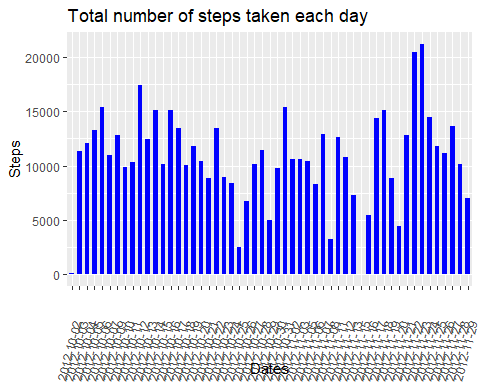


summary(activitydata)

## steps date interval   
## Min. : 0.00 Length:17568 Min. : 0.0   
## 1st Qu.: 0.00 Class :character 1st Qu.: 588.8   
## Median : 0.00 Mode :character Median :1177.5   
## Mean : 37.38 Mean :1177.5   
## 3rd Qu.: 12.00 3rd Qu.:1766.2   
## Max. :806.00 Max. :2355.0   
## NA's :2304

1. Histogram of the total number of steps taken each day

library(ggplot2)  
library(scales)  
ggplot(StepsPerDay, aes(x=date, y=steps)) + geom\_bar(stat = "identity", width=.6, fill="blue") + ylab("Steps") + xlab("Dates") + labs(title = "Total number of steps taken each day") + theme(axis.text.x = element\_text(angle=75, vjust=0.1))



1. Mean and median number of steps taken each day

rmean <- mean(StepsPerDay$steps)  
rmean

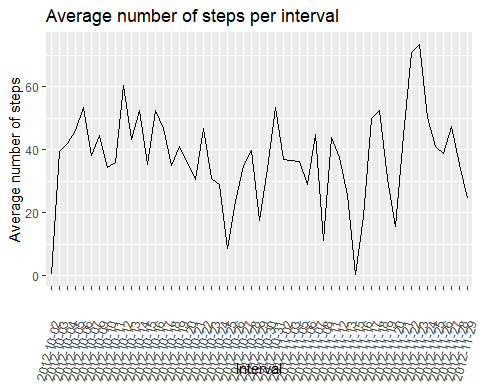
## [1] 10766.19

rmedian <- median(StepsPerDay$steps)  
rmedian

## [1] 10765

1. Time series plot of the average number of steps taken

AverageStepsPerDay <- aggregate(steps~date, activitydata, mean , na.rm=TRUE)  
names(AverageStepsPerDay)<-c("interval", "mean")  
ggplot(AverageStepsPerDay, aes(x=interval, y = mean , group=1)) + geom\_line() + ylab("Average number of steps") + xlab("Interval") + labs(title = "Average number of steps per interval") + theme(axis.text.x = element\_text(angle=75, vjust=0.1))



1. The 5-minute interval that, on average, contains the maximum number of steps

max\_interval <- activitydata[which.max(activitydata$steps),]$interval  
max\_interval

## [1] 615

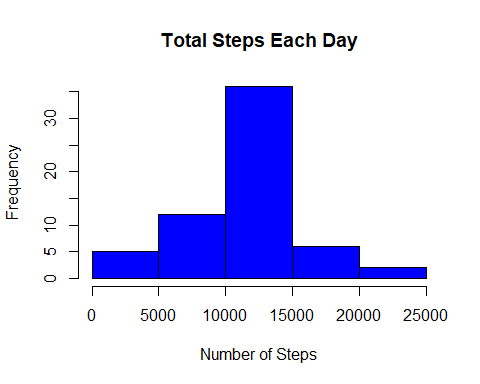
1. Code to describe and show a strategy for imputing missing data

TotalNAs <- sum(!complete.cases(activitydata))  
TotalNAs

## [1] 2304

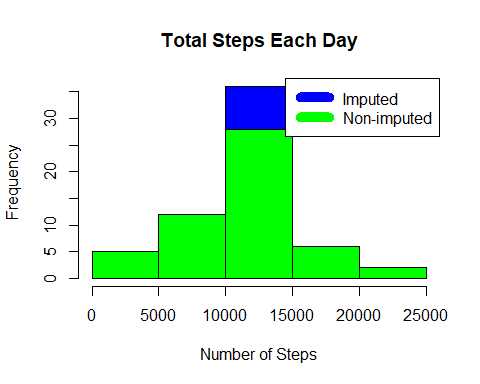
1. Histogram of the total number of steps taken each day after missing values are imputed

StepsAverage <- aggregate(steps ~ interval, data = activitydata, FUN = mean)  
fillNA <- numeric()  
for (i in 1:nrow(activitydata)) {  
 obs <- activitydata[i, ]  
 if (is.na(obs$steps)) {  
 steps <- subset(StepsAverage, interval == obs$interval)$steps  
 } else {  
 steps <- obs$steps  
 }  
 fillNA <- c(fillNA, steps)  
}  
  
new\_dataset\_activity <- activitydata  
new\_dataset\_activity$steps <- fillNA  
  
TotalSteps <- aggregate(steps ~ date, data = new\_dataset\_activity, sum, na.rm = TRUE)  
hist(TotalSteps$steps, main = paste("Total Steps Each Day"), col="blue", xlab="Number of Steps")



7.1 Create Histogram to show difference.

hist(TotalSteps$steps, main = paste("Total Steps Each Day"), col="blue", xlab="Number of Steps")  
hist(StepsPerDay$steps, main = paste("Total Steps Each Day"), col="green", xlab="Number of Steps", add=T)  
legend("topright", c("Imputed", "Non-imputed"), col=c("blue", "green"), lwd=10)



7.2 Mean and Median of the total number of steps taken per day

cmean <- mean(TotalSteps$steps)  
cmean

## [1] 10766.19

cmedian <- median(TotalSteps$steps)  
cmedian

## [1] 10766.19

1. Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

weekdays <- c("lunes", "martes", "miercoles", "jueves", "viernes")  
  
new\_dataset\_activity$dayoftheweek = as.factor(ifelse(is.element(weekdays(as.Date(new\_dataset\_activity$date)),weekdays), "Weekday", "Weekend"))  
TotalStepsU <- aggregate(steps ~ interval + dayoftheweek, new\_dataset\_activity, mean)  
  
library(lattice)  
xyplot(TotalStepsU$steps ~ TotalStepsU$interval|TotalStepsU$dayoftheweek, main="Average Steps per Day by Interval",xlab="Interval", ylab="Steps",layout=c(1,2), type="l")

