PA1\_template

## Loading and preprocessing the data

##### Loading Data #####  
library(data.table)  
library(ggplot2)  
#unzip("repdata%2Fdata%2Factivity.zip",exdir = "Data")  
##### Reading Data #####  
DT <- data.table::fread(input = "repdata\_data\_activity/activity.csv")

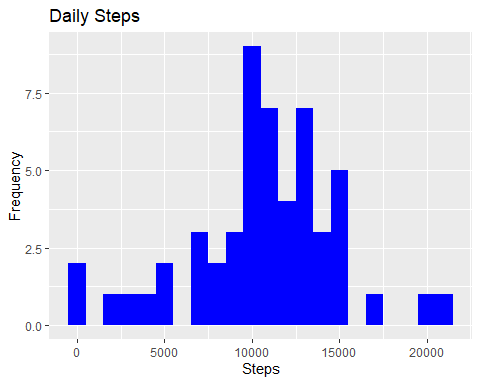
## What is mean total number of steps taken per day?

##### Total Steps #####  
Total\_Steps <- DT[, c(lapply(.SD, sum, na.rm = FALSE)), .SDcols = c("steps"), by = .(date)]   
head(Total\_Steps, 10)

## date steps  
## 1: 2012-10-01 NA  
## 2: 2012-10-02 126  
## 3: 2012-10-03 11352  
## 4: 2012-10-04 12116  
## 5: 2012-10-05 13294  
## 6: 2012-10-06 15420  
## 7: 2012-10-07 11015  
## 8: 2012-10-08 NA  
## 9: 2012-10-09 12811  
## 10: 2012-10-10 9900

ggplot(Total\_Steps, aes(x = steps)) +  
 geom\_histogram(fill = "blue", binwidth = 1000) +  
 labs(title = "Daily Steps", x = "Steps", y = "Frequency")

## Warning: Removed 8 rows containing non-finite values (stat\_bin).

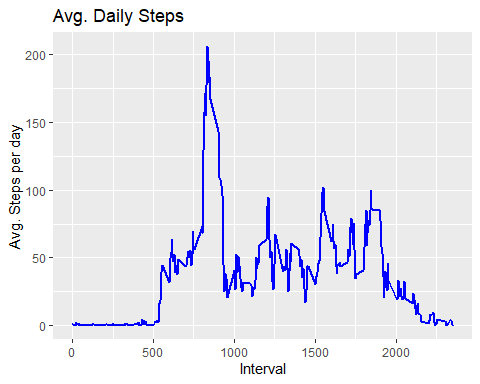


Total\_Steps[, .(Mean\_Steps = mean(steps, na.rm = TRUE), Median\_Steps = median(steps, na.rm = TRUE))]

## Mean\_Steps Median\_Steps  
## 1: 10766.19 10765

## What is the average daily activity pattern?

##### Avg daily activity pattern #####  
IntervalDT <- DT[, c(lapply(.SD, mean, na.rm = TRUE)), .SDcols = c("steps"), by = .(interval)]   
ggplot(IntervalDT, aes(x = interval , y = steps)) + geom\_line(color="blue", size=1) + labs(title = "Avg. Daily Steps", x = "Interval", y = "Avg. Steps per day")



IntervalDT[steps == max(steps), .(max\_interval = interval)]

## max\_interval  
## 1: 835

## Imputing missing values

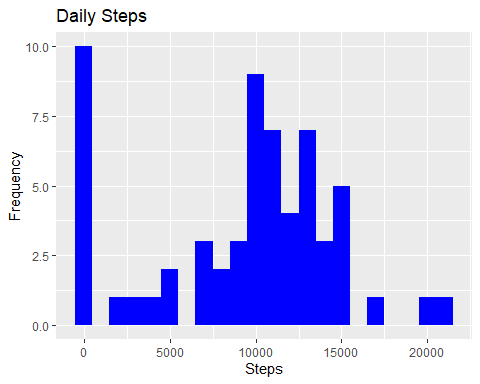
##### Missing Values #####  
DT[is.na(steps), .N ]

## [1] 2304

DT[is.na(steps), "steps"] <- DT[, c(lapply(.SD, median, na.rm = TRUE)), .SDcols = c("steps")]  
data.table::fwrite(x = DT, file = "Data/Replaced\_NA\_Data.csv", quote = FALSE)  
# total number of steps taken per day  
Total\_Steps <- DT[, c(lapply(.SD, sum)), .SDcols = c("steps"), by = .(date)]   
# mean and median total number of steps taken per day  
Total\_Steps[, .(Mean\_Steps = mean(steps), Median\_Steps = median(steps))]

## Mean\_Steps Median\_Steps  
## 1: 9354.23 10395

ggplot(Total\_Steps, aes(x = steps)) + geom\_histogram(fill = "blue", binwidth = 1000) + labs(title = "Daily Steps", x = "Steps", y = "Frequency")



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

##### Differences in activity patterns #####  
activityDT <- data.table::fread(input = "repdata\_data\_activity/activity.csv")  
activityDT[, date := as.POSIXct(date, format = "%Y-%m-%d")]  
activityDT[, `Day of Week`:= weekdays(x = date)]  
activityDT[grepl(pattern = "Monday|Tuesday|Wednesday|Thursday|Friday", x = `Day of Week`), "weekday or weekend"] <- "weekday"  
activityDT[grepl(pattern = "Saturday|Sunday", x = `Day of Week`), "weekday or weekend"] <- "weekend"  
activityDT[, `weekday or weekend` := as.factor(`weekday or weekend`)]  
head(activityDT, 10)

## steps date interval Day of Week weekday or weekend  
## 1: NA 2012-10-01 0 Monday weekday  
## 2: NA 2012-10-01 5 Monday weekday  
## 3: NA 2012-10-01 10 Monday weekday  
## 4: NA 2012-10-01 15 Monday weekday  
## 5: NA 2012-10-01 20 Monday weekday  
## 6: NA 2012-10-01 25 Monday weekday  
## 7: NA 2012-10-01 30 Monday weekday  
## 8: NA 2012-10-01 35 Monday weekday  
## 9: NA 2012-10-01 40 Monday weekday  
## 10: NA 2012-10-01 45 Monday weekday

activityDT[is.na(steps), "steps"] <- activityDT[, c(lapply(.SD, median, na.rm = TRUE)), .SDcols = c("steps")]  
IntervalDT <- activityDT[, c(lapply(.SD, mean, na.rm = TRUE)), .SDcols = c("steps"), by = .(interval, `weekday or weekend`)]   
ggplot(IntervalDT , aes(x = interval , y = steps, color=`weekday or weekend`)) + geom\_line() + labs(title = "Avg. Daily Steps by Weektype", x = "Interval", y = "No. of Steps") + facet\_wrap(~`weekday or weekend` , ncol = 1, nrow=2)

