library(ggplot2)

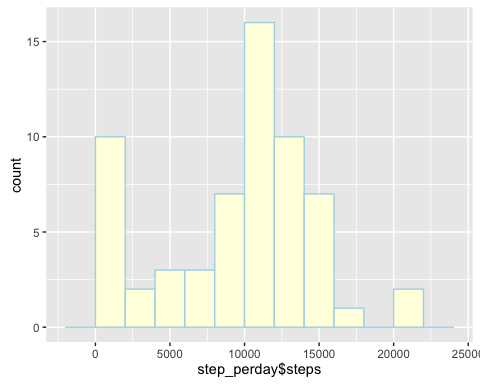
1.Code for reading in the dataset and/or processing the data

unzip("activity.zip")  
df <- read.csv("activity.csv", header = T, sep = ',', stringsAsFactors = FALSE)  
a <- summary(df)  
print(a, type="html")

## 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

step\_perday <- aggregate(df$steps, by = list(df$date),sum, na.rm = T)  
colnames(step\_perday) <- c("date","steps")  
b <- ggplot(data = step\_perday, aes(x=step\_perday$steps)) + geom\_histogram(binwidth = 2000, fill = 'lightyellow', col = 'lightblue')   
print(b,type="html")

  
 3. Mean and median number of steps taken each day

datamean <- mean(step\_perday$steps)  
datamedian <- median(step\_perday$steps)  
print(datamean, type = "html")

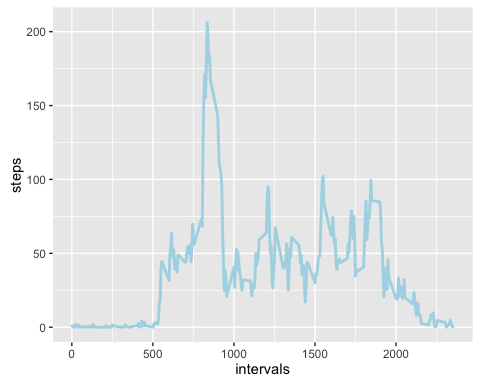
## [1] 9354.23

print(datamedian, type="html")

## [1] 10395

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

step\_int <- aggregate(df$steps, by=list(df$interval), mean, na.rm = T)  
colnames(step\_int) <- c("intervals","steps")  
ggplot(data = step\_int, aes(x=intervals, y=steps)) + geom\_line(col = 'lightblue', size =1)

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

step\_int[which.max(step\_int$steps),]

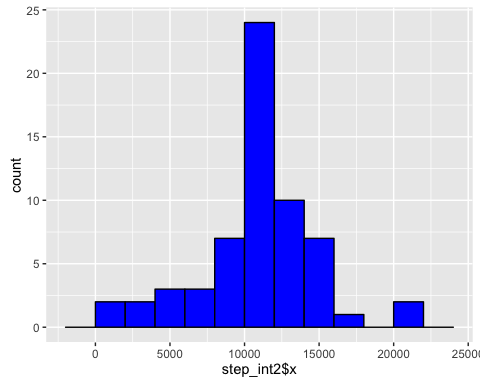
## intervals steps  
## 104 835 206.1698

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

df2 = df  
for(i in 1:nrow(df2)){  
 if(is.na(df2$steps[i])){  
 int = df2$interval[i]  
 sub = subset(step\_int, intervals == int)  
 df2$steps[i] = sub$steps  
 }  
}

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

step\_int2 <- aggregate(df2$steps, by = list(df2$date), sum)  
ggplot(data = step\_int2, aes(x=step\_int2$x)) + geom\_histogram(binwidth = 2000, fill = 'blue', col = 'black')

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

df2$date = as.Date(df2$date)  
df2$day = factor(format(df2$date, "%A"))  
levels(df2$day) = list(weekday = c("Monday", "Tuesday","Wednesday", "Thursday", "Friday"), weekend = c("Saturday", "Sunday"))  
  
we = subset(df2, df2$day == 'weekend')  
step\_int\_we = aggregate(we$steps, by=list(we$interval), mean)  
step\_int\_we$day = 'weekend'  
  
wd = subset(df2, df2$day == 'weekday')  
step\_int\_wd = aggregate(wd$steps, by=list(wd$interval), mean)  
step\_int\_wd$day = 'weekday'  
  
step\_int\_2 = rbind(step\_int\_we, step\_int\_wd)  
  
colnames(step\_int\_2) <- c('interval', 'steps', 'day')  
  
ggplot(data = step\_int\_2, aes(x = interval, y = steps)) + facet\_grid(day~.) + geom\_line(col = 'steelblue', size = 1) + labs(x = 'Interval(5 minuts)', y = 'Average numeber of steps', title ='Time series plot(new)')

