Week2\_projet\_report

Heiden89

February 26, 2018

## Loading and preprocessing the data

#### 1 & 2- Loading of the data

the\_data <- read.csv("./activity.csv",header=TRUE)

For the moment, no transformation is done

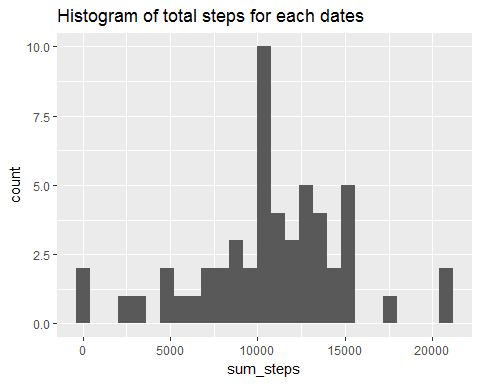
## What is mean total number of steps taken per day?

#### 1 & 2

the following graph illustrate an histogram of the total number of steps per day

grouped\_data <- the\_data %>%   
 group\_by(date) %>% summarise(sum\_steps = sum(steps,rm.na=TRUE))  
  
g <- ggplot(data=grouped\_data,aes(sum\_steps))+  
 geom\_histogram(binwidth = 800)+ggtitle("Histogram of total steps for each dates")  
 g

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



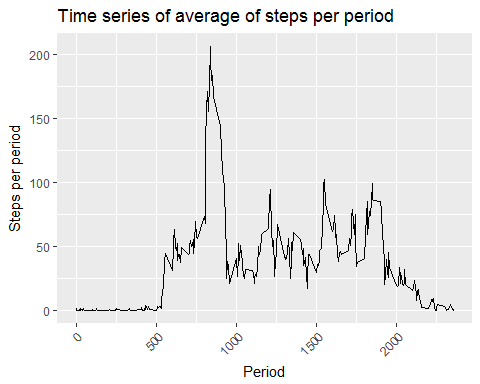
#Mean and Median  
   
steps\_summary <- summary(grouped\_data$sum\_steps)  
print(steps\_summary)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 42 8842 10766 10767 13295 21195 8

## What is the average daily activity pattern?

#### 1 & 2

by\_int <- the\_data%>% group\_by(interval) %>% summarise(the\_avr= mean(steps,na.rm=TRUE))  
  
  
  
ggplot(by\_int) + geom\_line(mapping= aes(x=interval, y= the\_avr)) +  
 xlab("Period") + ylab("Steps per period")+theme(axis.text.x = element\_text(angle = 45, hjust = 1,vjust=1))+  
 ggtitle("Time series of average of steps per period")



the\_max <- by\_int%>% filter(by\_int$the\_avr==max(by\_int$the\_avr))

The maximum average steps is 206.1698113 and is observable in the interval 835 The

## Imputing missing values

#### 1,2,3 & 4

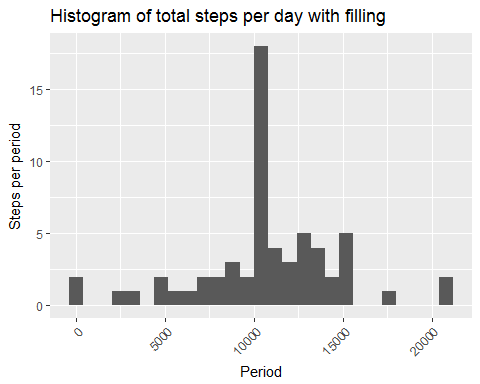
num\_missing= nrow(the\_data[is.na(the\_data$steps)==TRUE,])

There is 2304 period of 5 minutes that are Not Available in the data.frame

the\_filling= the\_data %>% merge(by\_int,by="interval") %>% mutate(with\_filled= ifelse(!is.na(steps)==TRUE,steps,the\_avr))

The strategy to fill the Na’s was to use the same interval average for the period. The new dataset is called **the\_filling**

total\_with\_filling <- the\_filling%>% group\_by(date) %>% summarise(sum\_with\_filled = sum(with\_filled))  
  
ggplot(total\_with\_filling,aes(sum\_with\_filled)) + geom\_histogram(binwidth = 800) +  
 xlab("Period") + ylab("Steps per period")+theme(axis.text.x = element\_text(angle = 45, hjust = 1,vjust=1))+  
 ggtitle("Histogram of total steps per day with filling")



The difference in mean is 0 wich is caused by the error handling in the mean function. There is no difference in the mean of the two distribution

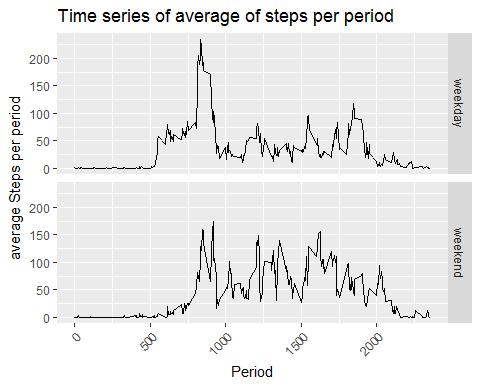
The real difference resides in the standard deviation. the difference in sd is -6.6707538 which is negative and caused by a more central distribution for the data that have been filled with the mean

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

the\_day <- factor("weekday","weekend")  
the\_filling$day <- weekdays(as.Date(the\_filling$date))  
the\_filling <- the\_filling %>% mutate(week\_timing = factor(case\_when(.$day %in% c("Monday","Tuesday","Wednesday","Thursday","Friday")~ "weekday",.$day %in% c("Saturday","Sunday")~"weekend")))  
  
print(head(the\_filling))

## interval steps date the\_avr with\_filled day week\_timing  
## 1 0 NA 2012-10-01 1.716981 1.716981 Monday weekday  
## 2 0 0 2012-11-23 1.716981 0.000000 Friday weekday  
## 3 0 0 2012-10-28 1.716981 0.000000 Sunday weekend  
## 4 0 0 2012-11-06 1.716981 0.000000 Tuesday weekday  
## 5 0 0 2012-11-24 1.716981 0.000000 Saturday weekend  
## 6 0 0 2012-11-15 1.716981 0.000000 Thursday weekday

by\_int\_filled <- the\_filling%>% group\_by(interval,week\_timing) %>% summarise(the\_avr\_filled= mean(steps,na.rm=TRUE))  
  
  
ggplot(by\_int\_filled) + geom\_line(aes(x=interval,y=the\_avr\_filled))+  
 facet\_grid(week\_timing ~ .)+  
 xlab("Period") + ylab("average Steps per period")+  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1,vjust=1))+ ggtitle("Time series of average of steps per period")



The previous Graphs shows the two pattern of average steps per period for the weekday and weekend. we can see that people tends to wake up later and stay active longer

knit2html(“PA1\_template.html”)