

Peer Assessment 1

loading the dataset

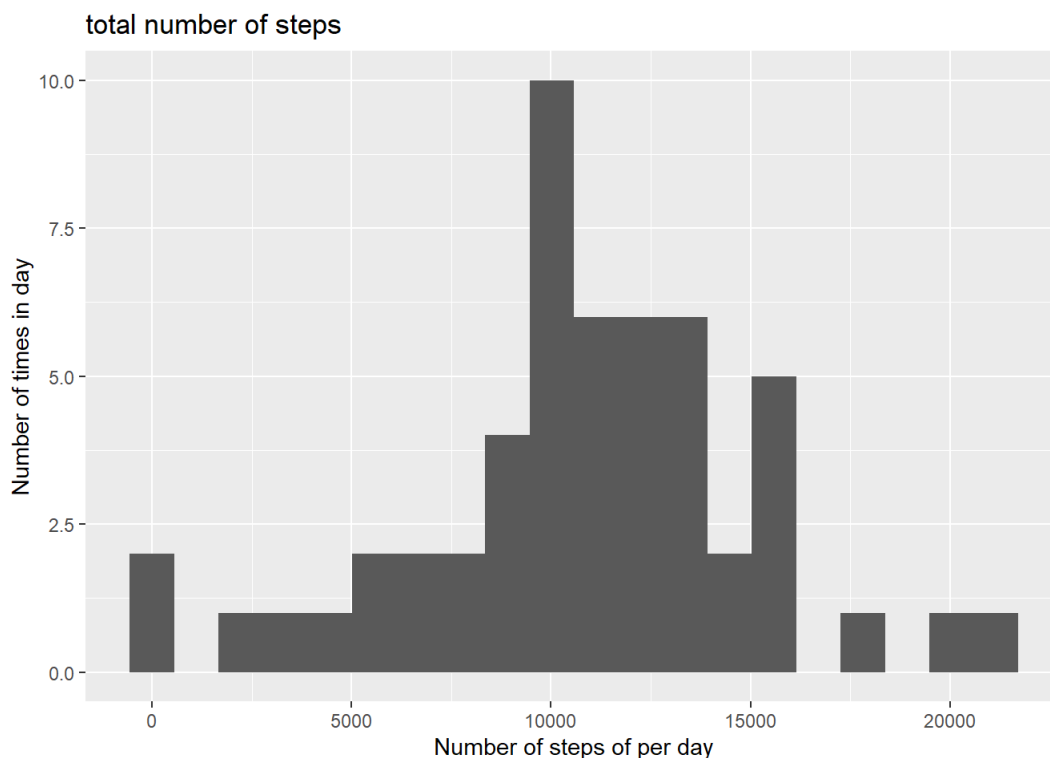
the number of step according to the date

make histogram that the numbers steps taken each day

```
step_date <- aggregate(steps~date, data = data1, FUN = sum, na.rm=FALSE)
str(step_date)
```

```
## 'data.frame':   53 obs. of  2 variables:
## $ date : Factor w/ 61 levels "2012-10-01","2012-10-02",...: 2 3 4 5 6 7 9 10 11 12 ...
## $ steps: int  126 11352 12116 13294 15420 11015 12811 9900 10304 17382 ...
```

```
ggplot(step_date, aes(x=steps))+
  geom_histogram(bins = 20)+
  labs(title = "total number of steps",
       x = "Number of steps of per day",
       y = "Number of times in day")
```



calculate the mean and median number

```
## [1] 10766.19
```

```
##      date      steps
## 2012-10-02: 1   Min.   : 41
## 2012-10-03: 1   1st Qu.: 8841
## 2012-10-04: 1   Median :10765
## 2012-10-05: 1   Mean    :10766
## 2012-10-06: 1   3rd Qu.:13294
## 2012-10-07: 1   Max.    :21194
## (Other)      :47
```

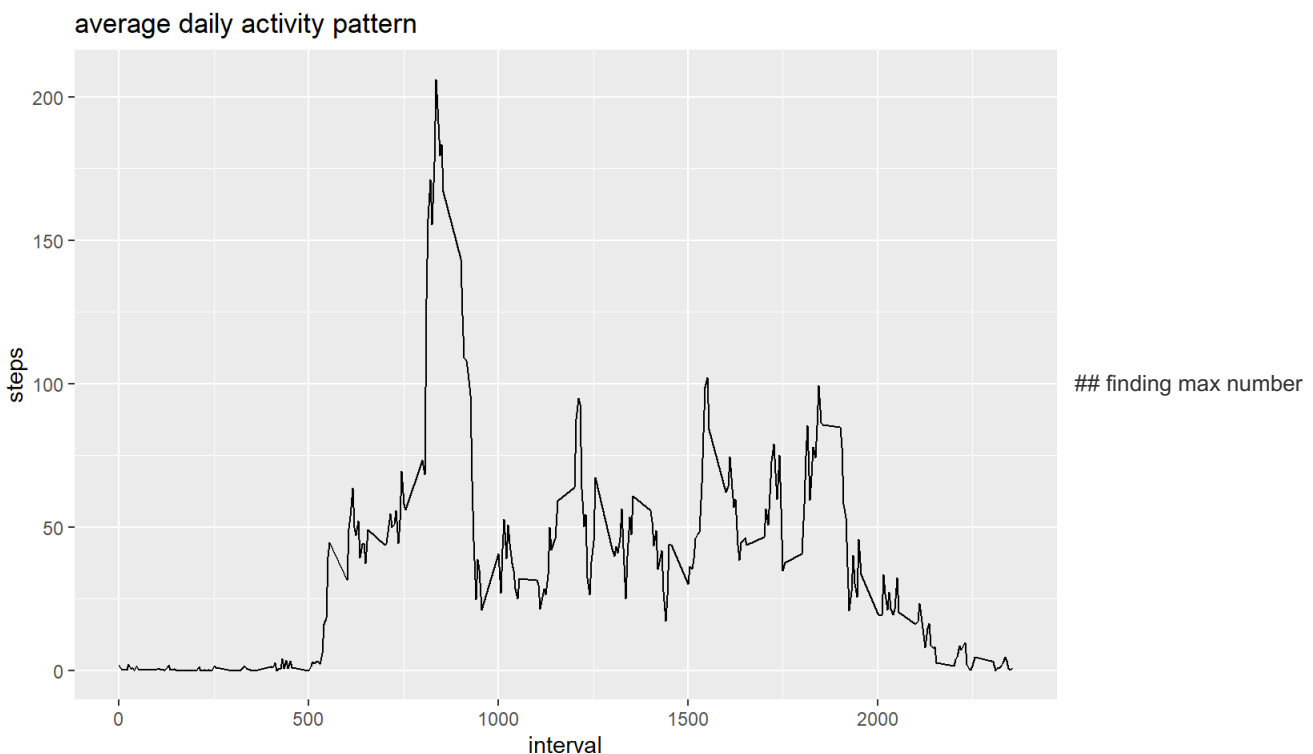
intervals that average of number of steps taken

plot in time series

```
interval_step <- aggregate(steps ~ interval, data1, mean)
head(interval_step)
```

```
##   interval    steps
## 1         0 1.7169811
## 2         5 0.3396226
## 3        10 0.1320755
## 4        15 0.1509434
## 5        20 0.0754717
## 6        25 2.0943396
```

```
ggplot(interval_step, aes(x =interval , y =steps))+
  geom_line()+
  labs(title = "average daily activity pattern",
       x ="interval", y= "steps")
```



```
max_steps <- which.max(interval_step$steps)

interval_step[max_steps,]
```

```
##   interval    steps
## 104       835 206.1698
```

imputing missing value

```
sum(is.na(data1))
```

```
## [1] 2304
```

```
data_filled <- data1
means <- mean(data1$steps, na.rm =TRUE)
data_filled$steps[is.na(data_filled$steps)] <- means

str(data_filled)
```

```
## 'data.frame': 17568 obs. of 3 variables:
## $ steps : num 37.4 37.4 37.4 37.4 37.4 ...
## $ date : Factor w/ 61 levels "2012-10-01","2012-10-02",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ interval: int 0 5 10 15 20 25 30 35 40 45 ...
```

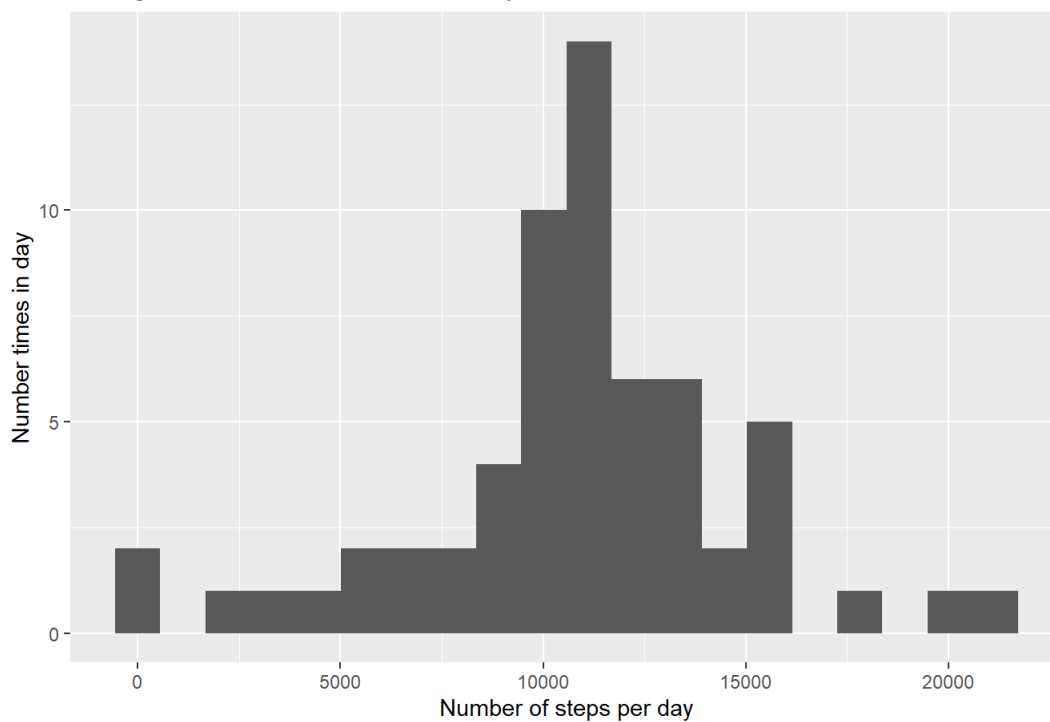
plot the dataset when missing value filled in

```
steps_datafill <- aggregate(steps~date, data = data_filled, FUN = sum)
head(steps_datafill)
```

```
##      date      steps
## 1 2012-10-01 10766.19
## 2 2012-10-02 126.00
## 3 2012-10-03 11352.00
## 4 2012-10-04 12116.00
## 5 2012-10-05 13294.00
## 6 2012-10-06 15420.00
```

```
ggplot(steps_datafill, aes(x=steps))+
  geom_histogram(bins = 20)+
  labs(title = "Histogram of the total number of steps",
       x = "Number of steps per day", y = "Number times in day")
```

Histogram of the total number of steps



```
summary(step_date)
```

```
##      date      steps
## 2012-10-02: 1   Min.   : 41
## 2012-10-03: 1   1st Qu.: 8841
## 2012-10-04: 1   Median :10765
## 2012-10-05: 1   Mean    :10766
## 2012-10-06: 1   3rd Qu.:13294
## 2012-10-07: 1   Max.    :21194
## (Other)      :47
```

```
summary(steps_datafill)
```

```
##           date           steps
## 2012-10-01: 1   Min.      : 41
## 2012-10-02: 1   1st Qu.: 9819
## 2012-10-03: 1   Median :10766
## 2012-10-04: 1   Mean      :10766
## 2012-10-05: 1   3rd Qu.:12811
## 2012-10-06: 1   Max.      :21194
## (Other)       :55
```

check differences in activity patterns between weekdays and weekends and plot it

```
dataNew <- data_filled
dataNew["type_of_day"] <- weekdays(as.Date(dataNew$date))
dataNew$type_of_day[dataNew$type_of_day %in% c('Saturday','Sunday')] <- 'weekend'
dataNew$type_of_day[dataNew$type_of_day != 'weekend'] <- 'weekday'

dataNew$type_of_day <- as.factor(dataNew$type_of_day)
dataNew_SI <- aggregate(steps~interval + type_of_day, dataNew, mean)
head(dataNew_SI)
```

```
##   interval type_of_day  steps
## 1         0     weekday 7.006569
## 2         5     weekday 5.384347
## 3        10     weekday 5.139902
## 4        15     weekday 5.162124
## 5        20     weekday 5.073235
## 6        25     weekday 6.295458
```

```
ggplot(dataNew_SI, aes(x = interval , y =steps))+
  geom_line()+
  labs(title = "Average daily steps",
       x = "interval",
       y = "Total number of steps")+
  facet_wrap(~type_of_day,ncol = 1, nrow =2)
```

