analysis of smart device usage

yunjing huang

September 9, 2021

Ask

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?
3. How could these trends help influence Bellabeat marketing strategy?

Prepare

For this analysis we use an open source dataset avalible on kaggle.link (https://www.kaggle.com/arashnic/fitbit). There are 18 csv files in the datasets. After carefully check the dataset we decide to use the dailyActivity_merged.csv,sleepDay_merged.csv and dailyCalories_merged.csv.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.
1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.1     v dplyr   1.0.6
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts
() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
daily_activity <- read.csv("D:/FitabaseData/dailyActivity_merged.csv")
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    4/12/2016      13162          8.50            8.50
## 2 1503960366    4/13/2016      10735          6.97            6.97
## 3 1503960366    4/14/2016      10460          6.74            6.74
## 4 1503960366    4/15/2016       9762          6.28            6.28
## 5 1503960366    4/16/2016      12669          8.16            8.16
## 6 1503960366    4/17/2016       9705          6.48            6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
## 5                5.04                       0                36
## 6                2.51                       0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

```
sleep_day <- read.csv("D:/FitabaseData/sleepDay_merged.csv")
head(sleep_day)
```

```
##           Id              SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 4/12/2016 12:00:00 AM                 1                327
## 2 1503960366 4/13/2016 12:00:00 AM                 2                384
## 3 1503960366 4/15/2016 12:00:00 AM                 1                412
## 4 1503960366 4/16/2016 12:00:00 AM                 2                340
## 5 1503960366 4/17/2016 12:00:00 AM                 1                700
## 6 1503960366 4/19/2016 12:00:00 AM                 1                304
##   TotalTimeInBed
## 1            346
## 2            407
## 3            442
## 4            367
## 5            712
## 6            320
```

```
daily_calories <- read.csv("D:/FitabaseData/dailyCalories_merged.csv")
head(daily_calories)
```

```
##           Id ActivityDay Calories
## 1 1503960366   4/12/2016     1985
## 2 1503960366   4/13/2016     1797
## 3 1503960366   4/14/2016     1776
## 4 1503960366   4/15/2016     1745
## 5 1503960366   4/16/2016     1863
## 6 1503960366   4/17/2016     1728
```

Process

in dataframe sleep_day the column sleepDay is in datatime format. we convert the datatime format to data format to be consistent with the other two dataframes.

```
sleep_day$SleepDay = as.Date(sleep_day$SleepDay, format = "%m/%d/%Y %I:%M:%S %
p")
head(sleep_day)
```

```
##           Id   SleepDay TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 2016-04-12                 1                327            346
## 2 1503960366 2016-04-13                 2                384            407
## 3 1503960366 2016-04-15                 1                412            442
## 4 1503960366 2016-04-16                 2                340            367
## 5 1503960366 2016-04-17                 1                700            712
## 6 1503960366 2016-04-19                 1                304            320
```

We rename the column sleepDay to ActivityDate in dataframe sleep_day. rename the column ActivityDay to ActivityDate. We then convert the column ActivityDate to %Y-%m-%d format. After that we convert the column ActivityDay in dataframe daily_calories to %Y-%m-%d format and rename the column ActivityDay to ActivityDate. The three dataframe has consistent data format after these operation.

```
sleep_day<-sleep_day%>% rename(ActivityDate=SleepDay)
head(sleep_day)
```

```
##           Id ActivityDate TotalSleepRecords TotalMinutesAsleep TotalTimeInBe
d
## 1 1503960366   2016-04-12                 1                327            34
6
## 2 1503960366   2016-04-13                 2                384            40
7
## 3 1503960366   2016-04-15                 1                412            44
2
## 4 1503960366   2016-04-16                 2                340            36
7
## 5 1503960366   2016-04-17                 1                700            71
2
## 6 1503960366   2016-04-19                 1                304            32
0
```

```
daily_activity$ActivityDate = as.Date(format(daily_activity$ActivityDate), "%
m/%d/%Y")
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50             8.50
## 2 1503960366   2016-04-13      10735          6.97             6.97
## 3 1503960366   2016-04-14      10460          6.74             6.74
## 4 1503960366   2016-04-15       9762          6.28             6.28
## 5 1503960366   2016-04-16      12669          8.16             8.16
## 6 1503960366   2016-04-17       9705          6.48             6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
## 5                5.04                       0                36
## 6                2.51                       0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

```
daily_calories$ActivityDay = as.Date(daily_calories$ActivityDay, format = "%m/%
d/%Y")
daily_calories<-daily_calories%>% rename(ActivityDate=ActivityDay)
head(daily_calories)
```

```
##           Id ActivityDate Calories
## 1 1503960366   2016-04-12     1985
## 2 1503960366   2016-04-13     1797
## 3 1503960366   2016-04-14     1776
## 4 1503960366   2016-04-15     1745
## 5 1503960366   2016-04-16     1863
## 6 1503960366   2016-04-17     1728
```

merge the three dataframe into a final dataframe and select only the columns we are interested.

```
final_data <- merge(daily_activity,sleep_day,daily_calories,by.x=c("Id","Activi
tyDate"),by.y=c("Id","ActivityDate"),by.z=c("Id","ActivityDate"))
head(final_data)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-15       9762          6.28            6.28
## 4 1503960366   2016-04-16      12669          8.16            8.16
## 5 1503960366   2016-04-17       9705          6.48            6.48
## 6 1503960366   2016-04-19      15506          9.88            9.88
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.14                     1.26
## 4                        0               2.71                     0.41
## 5                        0               3.19                     0.78
## 6                        0               3.53                     1.32
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                2.83                       0                29
## 4                5.04                       0                36
## 5                2.51                       0                38
## 6                5.03                       0                50
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  34                  209              726     1745
## 4                  10                  221              773     1863
## 5                  20                  164              539     1728
## 6                  31                  264              775     2035
##   TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1                 1                327            346
## 2                 2                384            407
## 3                 1                412            442
## 4                 2                340            367
## 5                 1                700            712
## 6                 1                304            320
```

```
data<-final_data%>%select(Id,ActivityDate,TotalSteps,TotalDistance,VeryActiveMi
nutes,Calories,TotalMinutesAsleep)
head(data)
```

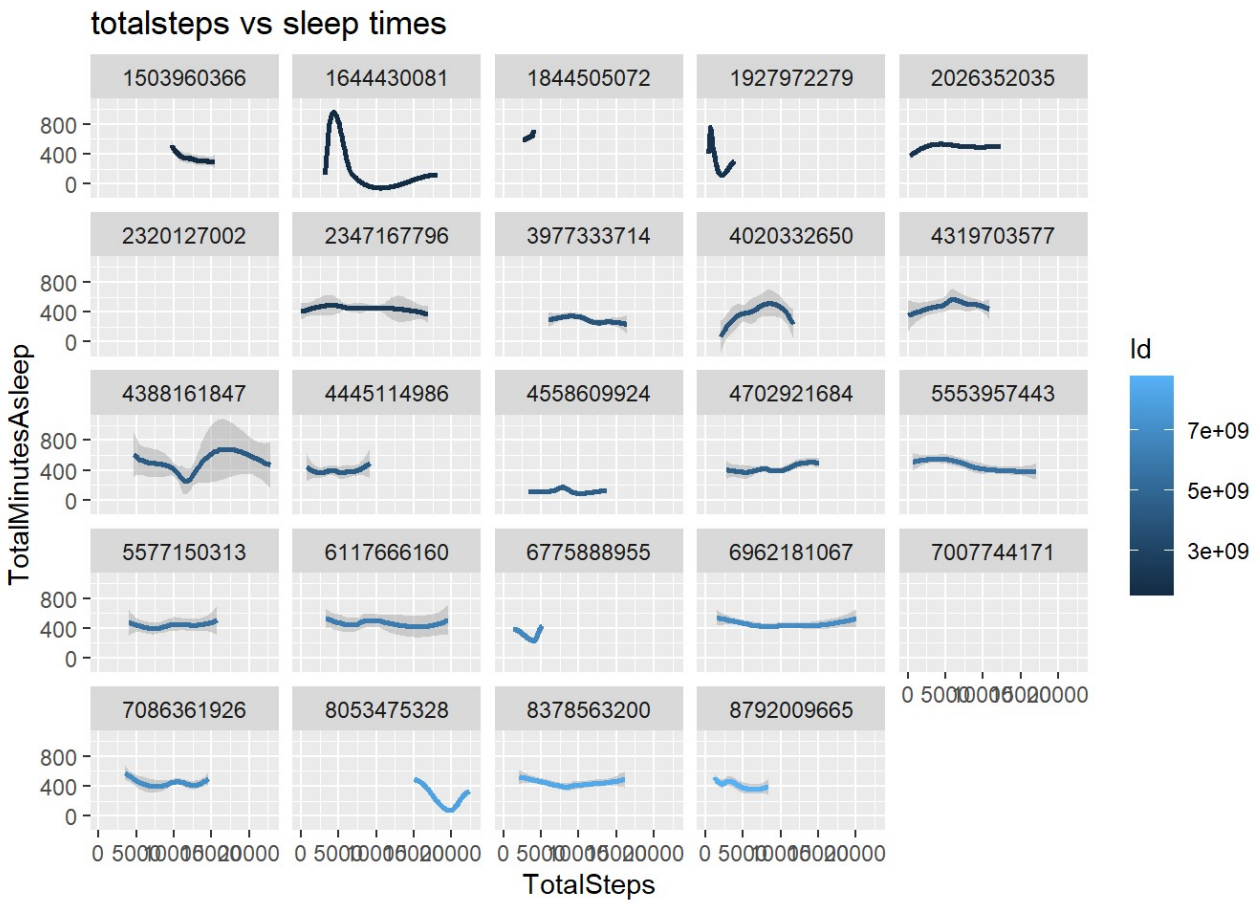
```
##           Id ActivityDate TotalSteps TotalDistance VeryActiveMinutes Calorie
s
## 1 1503960366   2016-04-12      13162          8.50                25     198
5
## 2 1503960366   2016-04-13      10735          6.97                21     179
7
## 3 1503960366   2016-04-15       9762          6.28                29     174
5
## 4 1503960366   2016-04-16      12669          8.16                36     186
3
## 5 1503960366   2016-04-17       9705          6.48                38     172
8
## 6 1503960366   2016-04-19      15506          9.88                50     203
5
##   TotalMinutesAsleep
## 1                327
## 2                384
## 3                412
## 4                340
## 5                700
## 6                304
```

Analyze

first we check the relationships between TotalSteps and TotalMinutesAsleep

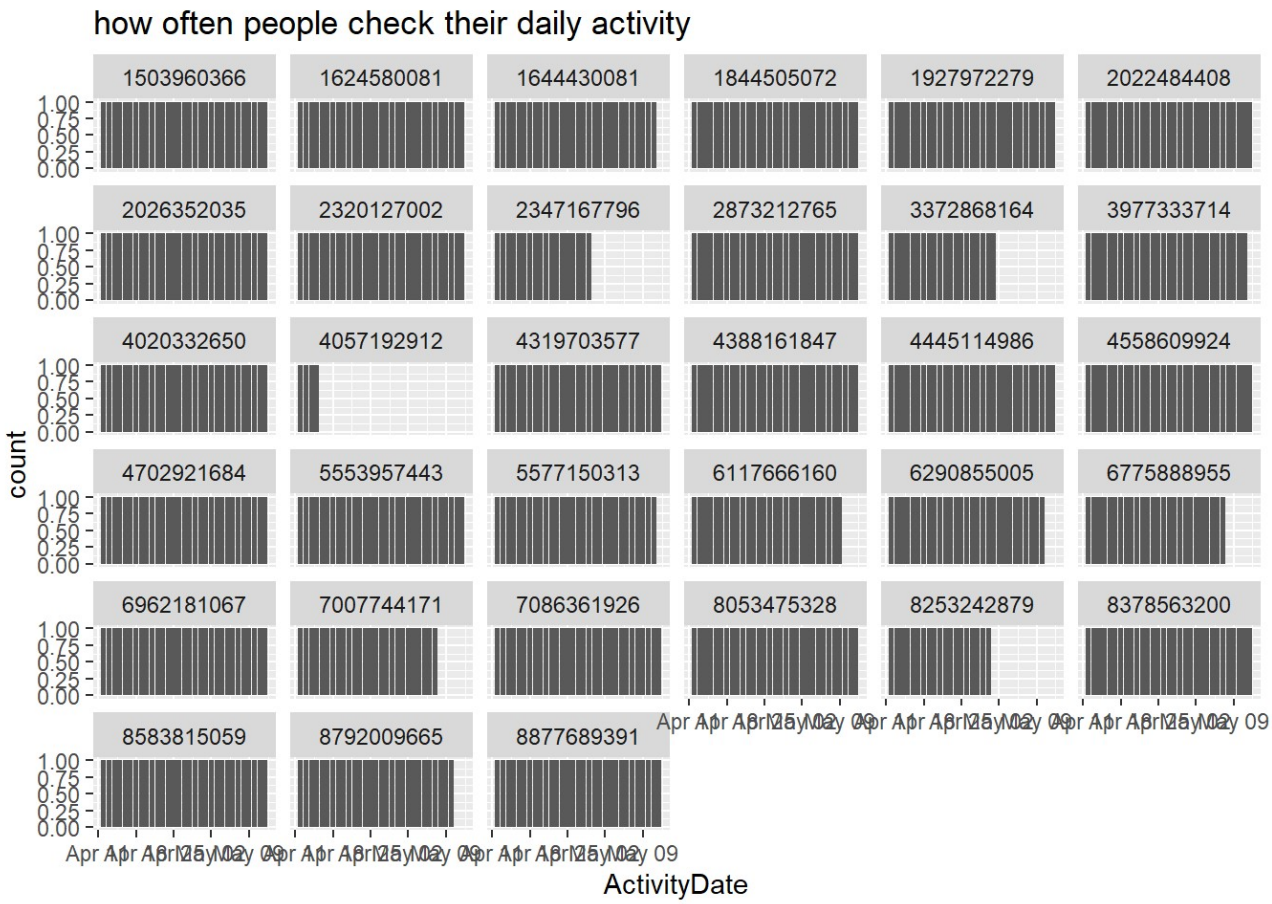
```
ggplot(data=data)+geom_smooth(mapping=aes(x=TotalSteps,y=TotalMinutesAsleep,col
or=Id))+facet_wrap(~Id)+labs(title="totalsteps vs sleep times")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

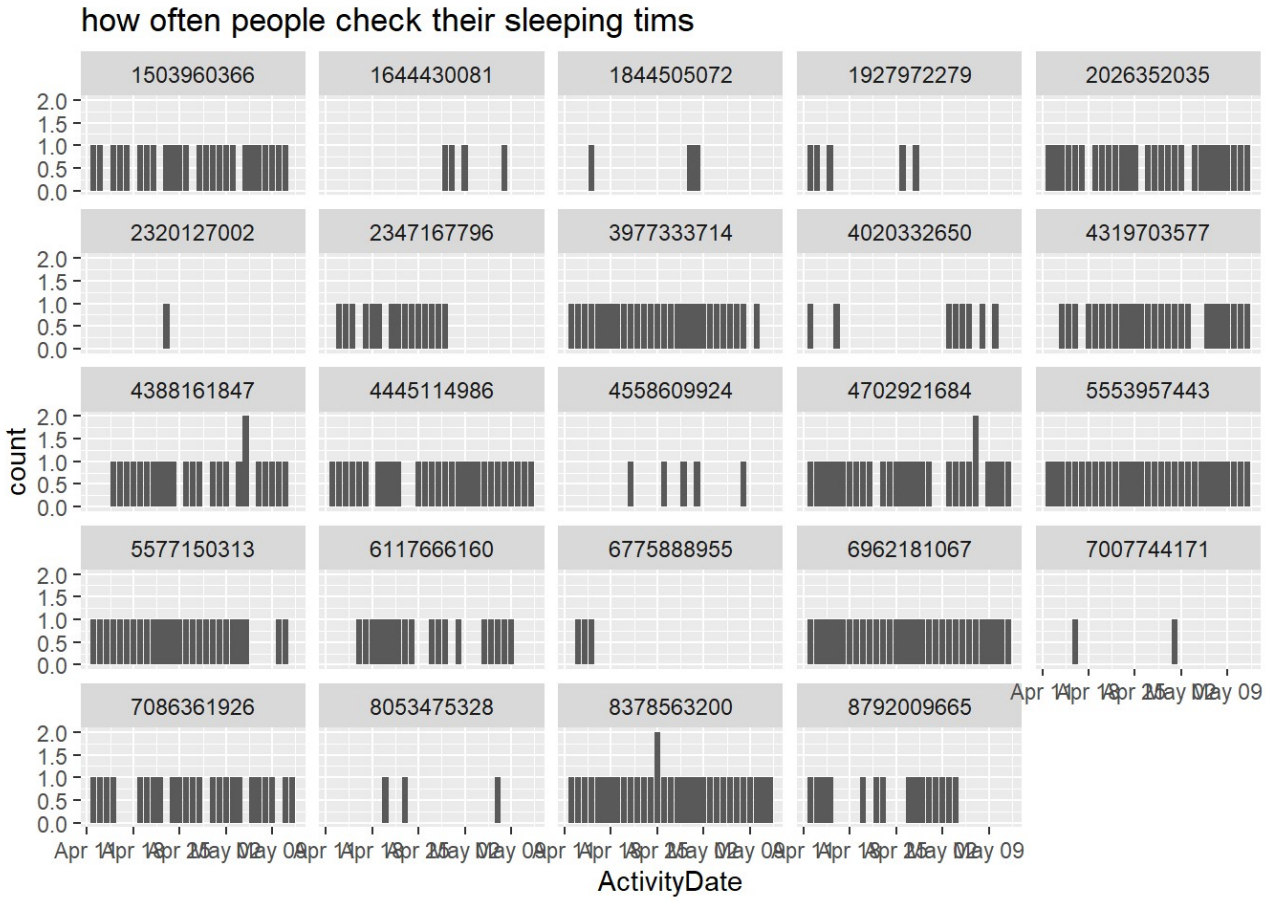


these graphs show that there are no clear relationship between totalsteps and sleeptimes. There are some peaks and valleys in the plots. Indicate apropriate execies is helpful for some people.

```
ggplot(data=daily_activity)+geom_bar(mapping=aes(x=ActivityDate))+facet_wrap(~Id)+labs(title="how often people check their daily activity")
```




```
ggplot(data=sleep_day)+geom_bar(mapping=aes(x=ActivityDate))+facet_wrap(~Id)+la
bs(title="how often people check their sleeping tims")
```

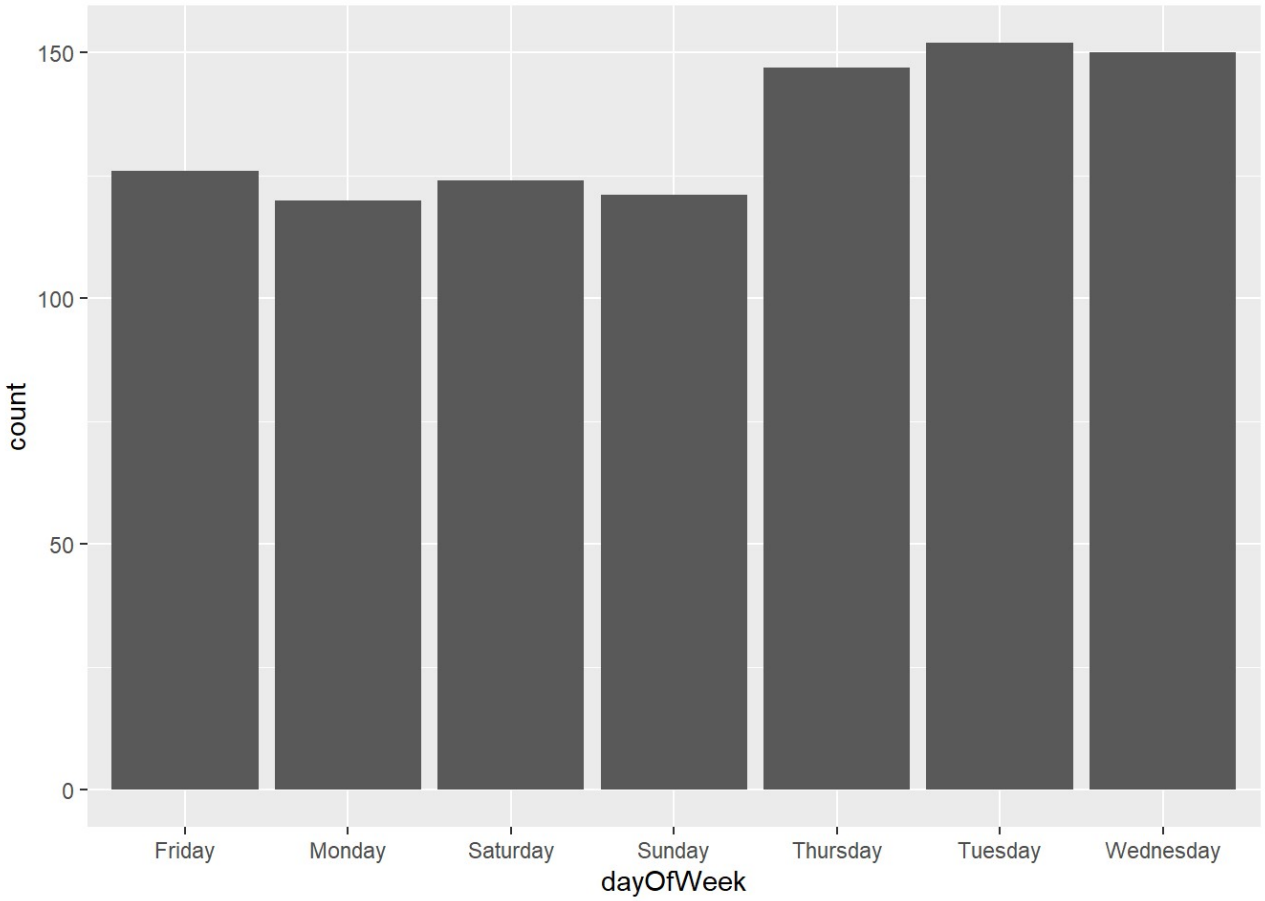


From the results we can clearly see that people are prefer the use devices to check their daily activities than to check their sleeping times as more user stop report their sleeping times than user stop report daily activities.

```
daily_activity$dayOfWeek<-weekdays(as.Date(daily_activity$ActivityDate))
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50             8.50
## 2 1503960366   2016-04-13      10735          6.97             6.97
## 3 1503960366   2016-04-14      10460          6.74             6.74
## 4 1503960366   2016-04-15       9762          6.28             6.28
## 5 1503960366   2016-04-16      12669          8.16             8.16
## 6 1503960366   2016-04-17       9705          6.48             6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
## 5                5.04                       0                36
## 6                2.51                       0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories dayOfWe
ek
## 1                  13                  328              728     1985   Tuesd
ay
## 2                  19                  217              776     1797 Wednesd
ay
## 3                  11                  181             1218     1776  Thursd
ay
## 4                  34                  209              726     1745    Frid
ay
## 5                  10                  221              773     1863  Saturd
ay
## 6                  20                  164              539     1728    Sund
ay
```

```
ggplot(data=daily_activity)+geom_bar(mapping=aes(x=dayOfWeek))
```

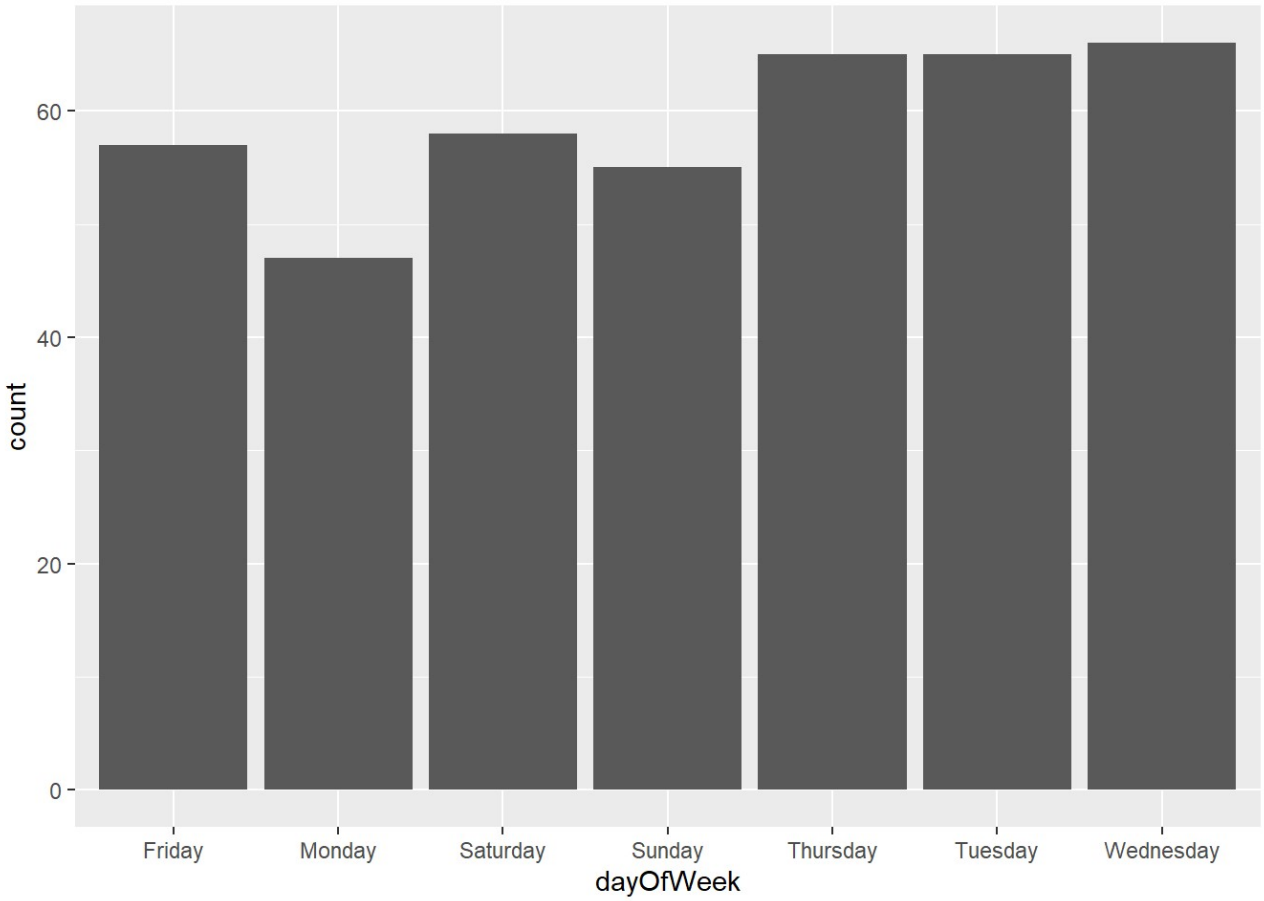


the graph shows that people are more likely to record excierse on Tuesday, Wednesday and Thursday.

```
sleep_day$dayOfWeek<-weekdays(as.Date(sleep_day$ActivityDate))
head(sleep_day)
```

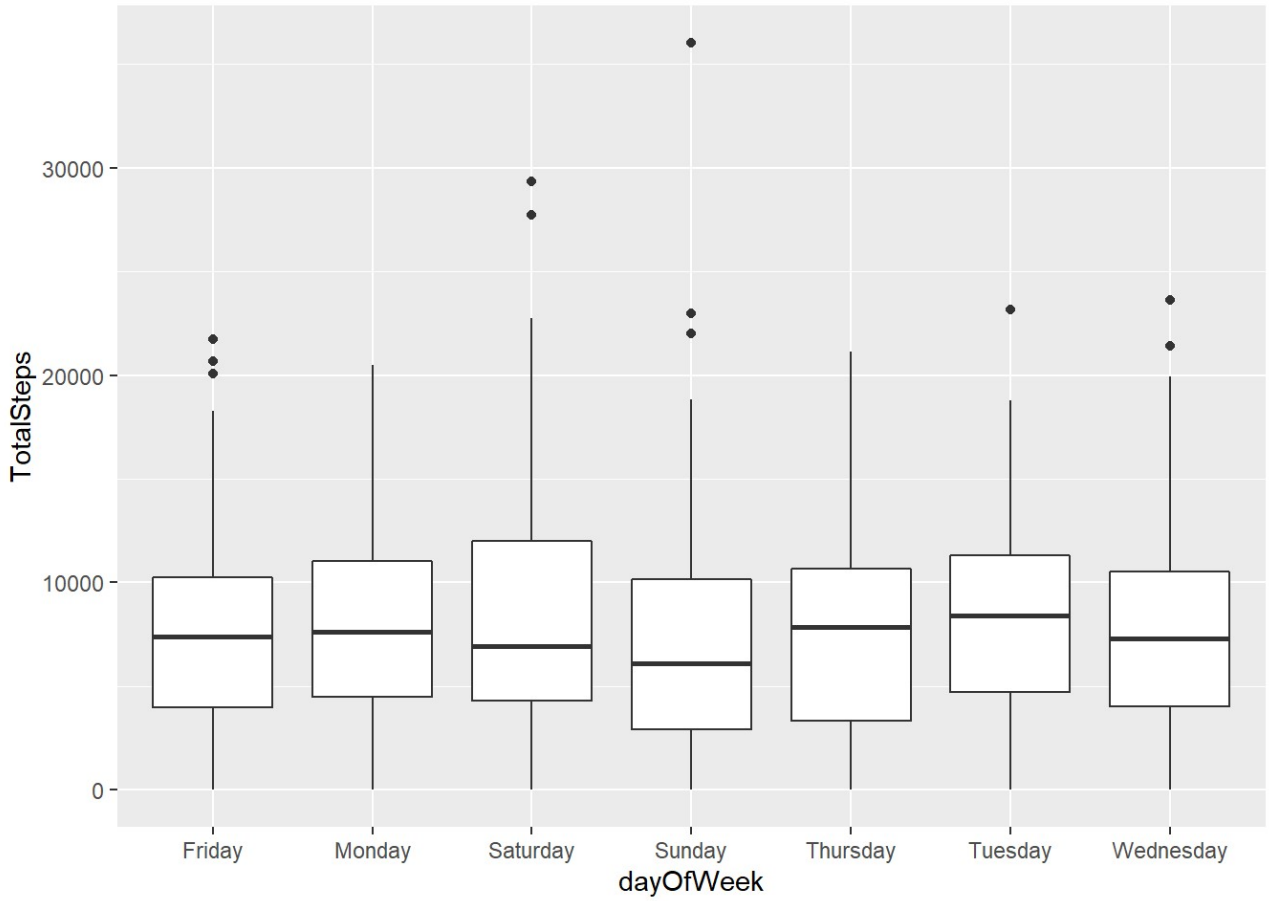
```
##           Id ActivityDate TotalSleepRecords TotalMinutesAsleep TotalTimeInBe
d
## 1 1503960366   2016-04-12                 1                327            34
6
## 2 1503960366   2016-04-13                 2                384            40
7
## 3 1503960366   2016-04-15                 1                412            44
2
## 4 1503960366   2016-04-16                 2                340            36
7
## 5 1503960366   2016-04-17                 1                700            71
2
## 6 1503960366   2016-04-19                 1                304            32
0
##   dayOfWeek
## 1   Tuesday
## 2 Wednesday
## 3    Friday
## 4  Saturday
## 5    Sunday
## 6   Tuesday
```

```
ggplot(data=sleep_day)+geom_bar(mapping=aes(x=dayOfWeek))
```

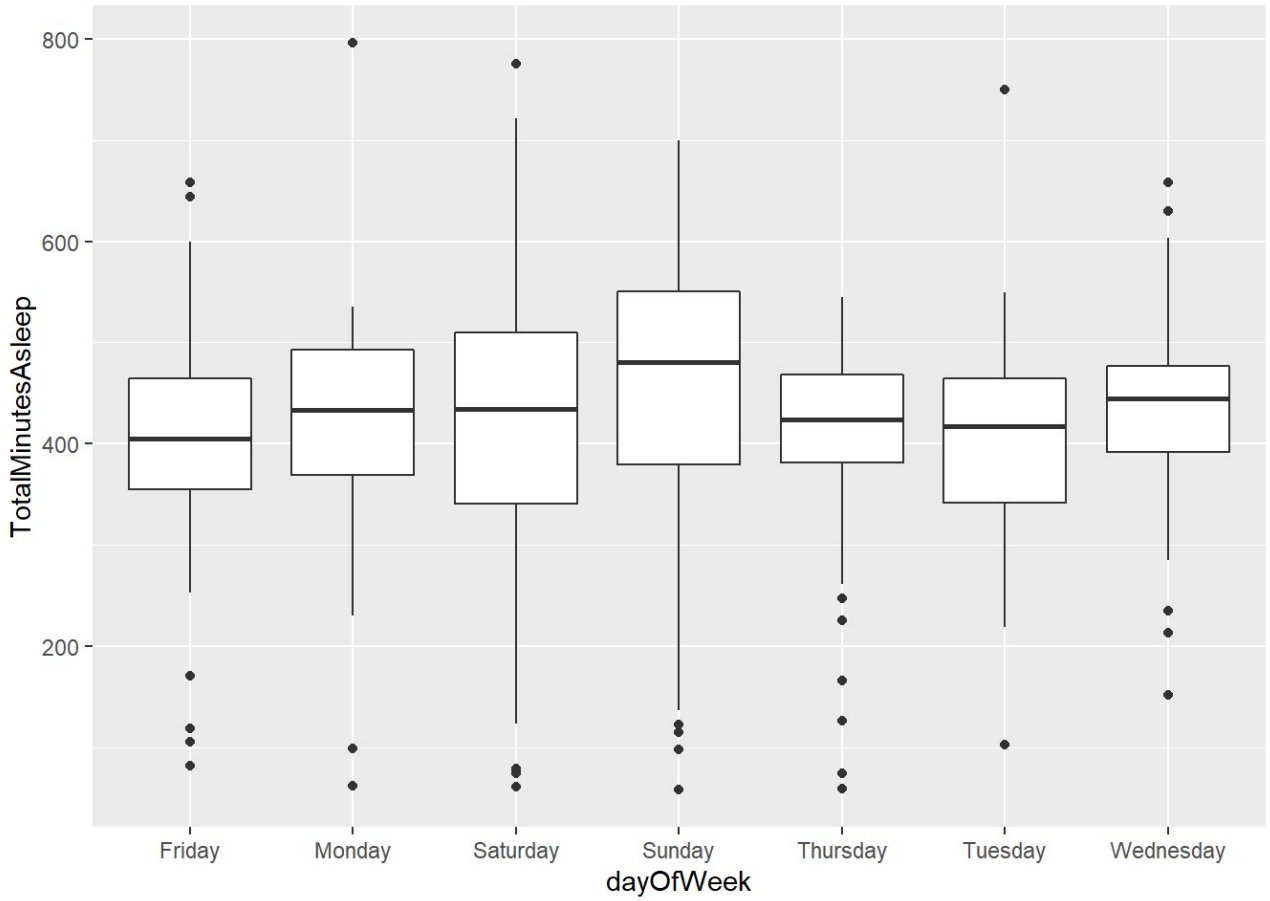


the graph shows that people are more likely to record sleeping times on Tuesday, Wednesday and Thursday.

```
ggplot(data=daily_activity)+geom_boxplot(mapping=aes(x=dayOfWeek,y=TotalSteps))
```




```
ggplot(data=sleep_day)+geom_boxplot(mapping=aes(x=dayOfWeek,y=TotalMinutesAslee
p))
```



these graphs shows that people record less activity and longer sleeping times on weekend.

Share

from the resuts above we can see

- people are more likely to quite recording their sleeping times than activities.
- people are more likely to record excierse on Tuesday, Wednesday and Thursday.

From the analysis I make below suggestions.

- Emphasize more on activity record in advertise and products design.

Thanks google for provide me this oportunity to do this analysis. Thanks Mobius for share this dataset.