

# Be-Healthy

Pooja Rathee

2022-04-08

**First we install required packages and go through it**

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.6      v dplyr  1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1

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

library(ggplot2)
```

**Read csv file**

```
daily_calories <- read_csv("data/dailyCalories_merged.csv")

## Rows: 940 Columns: 3
## -- Column specification -----
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, Calories
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

weight_log <- read.csv("data/weightLogInfo_merged.csv")
```

**Take a look at the daily calories data**

```
head(daily_calories)
```

```
## # A tibble: 6 x 3
##       Id ActivityDay Calories
##   <dbl> <chr>      <dbl>
## 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
```

Identify all the columns in daily\_\_activity

```
colnames(daily_calories)
```

```
## [1] "Id"          "ActivityDay" "Calories"
```

Take a look at the weight\_log

```
head(weight_log)
```

```
##       Id              Date WeightKg WeightPounds Fat  BMI
## 1 1503960366 5/2/2016 11:59:59 PM    52.6    115.9631 22 22.65
## 2 1503960366 5/3/2016 11:59:59 PM    52.6    115.9631 NA 22.65
## 3 1927972279 4/13/2016 1:08:52 AM   133.5    294.3171 NA 47.54
## 4 2873212765 4/21/2016 11:59:59 PM    56.7    125.0021 NA 21.45
## 5 2873212765 5/12/2016 11:59:59 PM    57.3    126.3249 NA 21.69
## 6 4319703577 4/17/2016 11:59:59 PM    72.4    159.6147 25 27.45
##   IsManualReport      LogId
## 1             True 1.462234e+12
## 2             True 1.462320e+12
## 3            False 1.460510e+12
## 4             True 1.461283e+12
## 5             True 1.463098e+12
## 6             True 1.460938e+12
```

Identify all the columns in weight log

```
colnames(weight_log)
```

```
## [1] "Id"          "Date"          "WeightKg"      "WeightPounds"
## [5] "Fat"         "BMI"          "IsManualReport" "LogId"
```

## Understanding some summary statistics

```
n_distinct(daily_calories$Id)
```

```
## [1] 33
```

```
n_distinct(weight_log$Id)
```

```
## [1] 8
```

## Calculate the observations

```
nrow(daily_calories)
```

```
## [1] 940
```

```
nrow(weight_log)
```

```
## [1] 67
```

## summary statistics of daily calories

```
daily_calories %>%  
  select(Id,  
          ActivityDay,  
          Calories) %>%  
  summary()
```

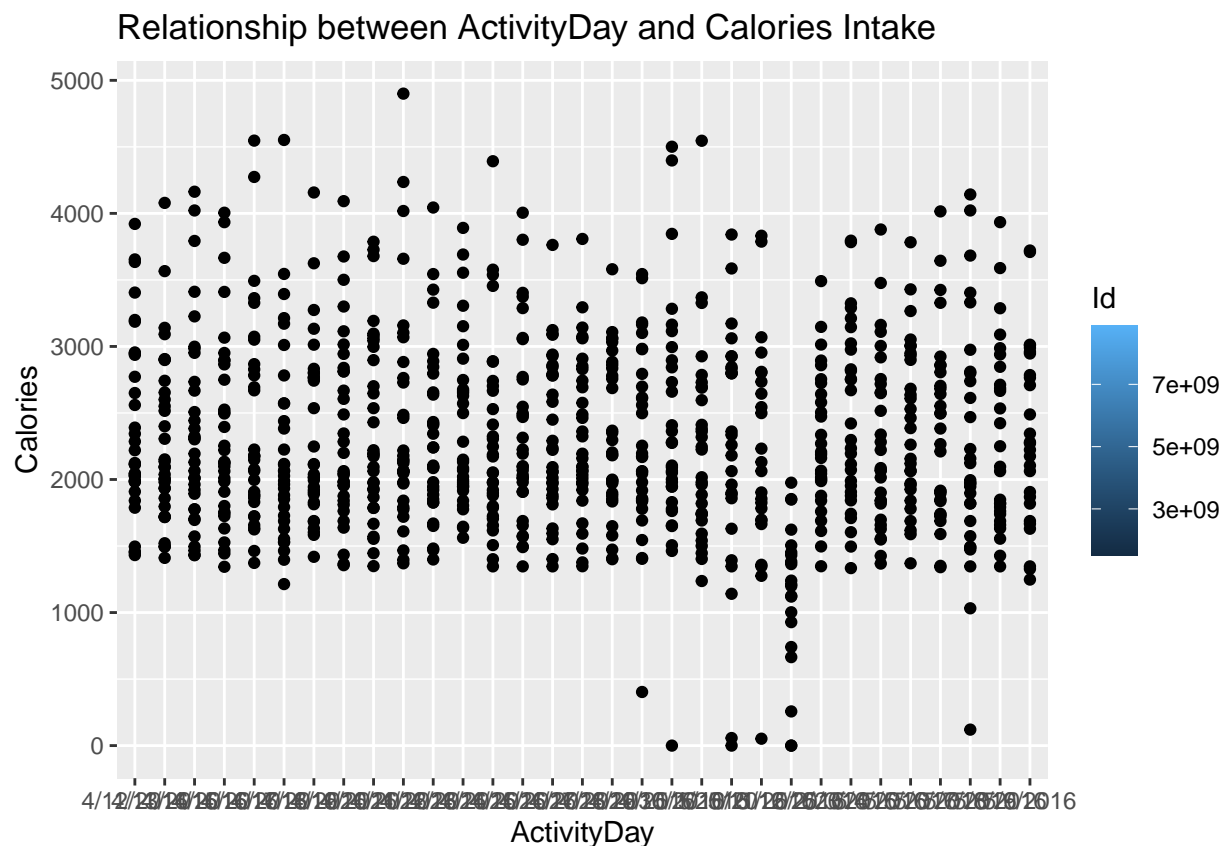
```
##           Id           ActivityDay           Calories  
##  Min.      :1.504e+09  Length:940      Min.       :  0  
## 1st Qu.:2.320e+09    Class :character 1st Qu.:1828  
## Median :4.445e+09    Mode  :character Median :2134  
## Mean   :4.855e+09                      Mean   :2304  
## 3rd Qu.:6.962e+09                      3rd Qu.:2793  
## Max.   :8.878e+09                      Max.   :4900
```

## For the weight\_log dataframe

```
weight_log %>%  
  select(Id,  
          Date,  
          WeightKg, BMI, Fat) %>%  
  summary()
```

```
##           Id           Date           WeightKg           BMI
## Min.      :1.504e+09   Length:67       Min.      : 52.60   Min.      :21.45
## 1st Qu.:6.962e+09   Class :character   1st Qu.: 61.40   1st Qu.:23.96
## Median :6.962e+09   Mode  :character   Median : 62.50   Median :24.39
## Mean      :7.009e+09                Mean      : 72.04   Mean      :25.19
## 3rd Qu.:8.878e+09                3rd Qu.: 85.05   3rd Qu.:25.56
## Max.      :8.878e+09                Max.      :133.50   Max.      :47.54
##
##           Fat
## Min.      :22.00
## 1st Qu.:22.75
## Median :23.50
## Mean      :23.50
## 3rd Qu.:24.25
## Max.      :25.00
## NA's      :65
```

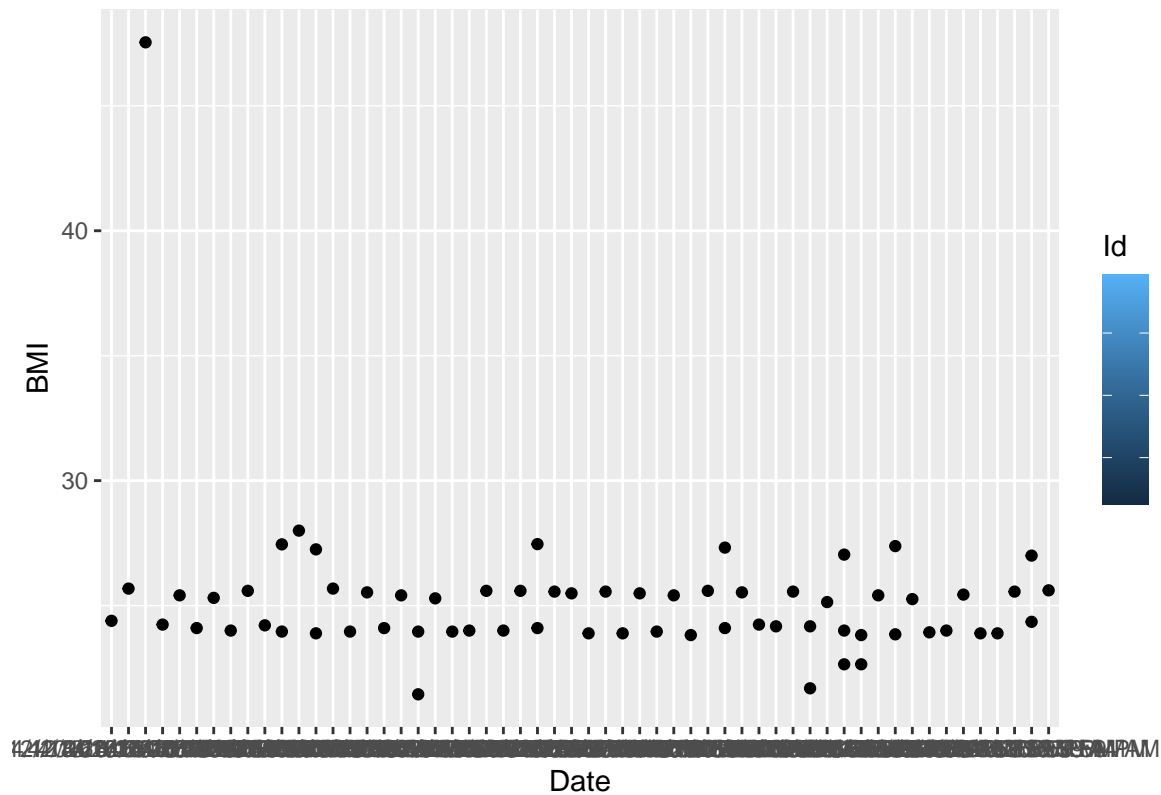
```
ggplot(data=daily_calories, aes(x=ActivityDay, y=Calories, fill=Id)) + geom_point()+
  labs(title = "Relationship between ActivityDay and Calories Intake")
```



```
# plotting the graph for weight_log
```

```
ggplot(data=weight_log, aes(x=Date, y=BMI, fill=Id)) + geom_point() +
  labs(title = "Relationship between Date and BMI rate as per Id's")
```

Relationship between Date and BMI rate as per Id's

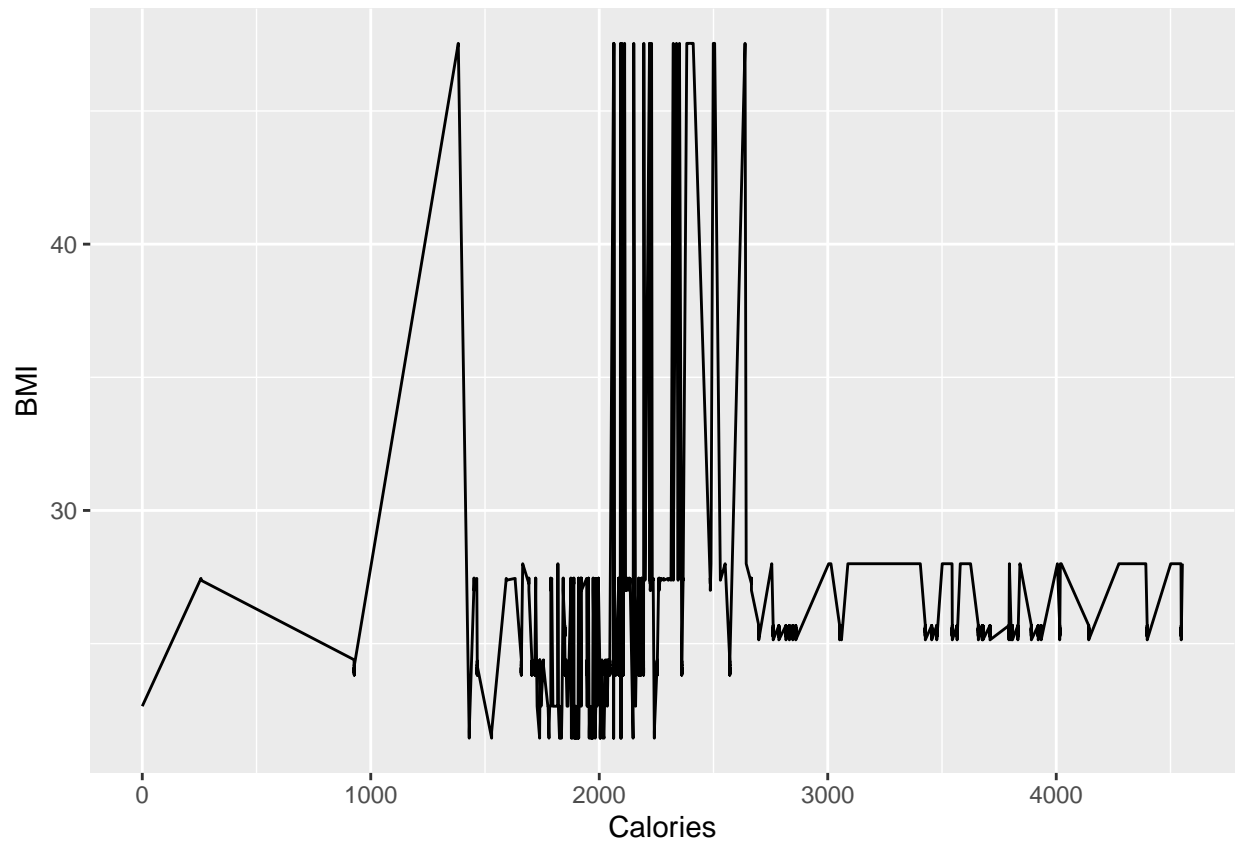


What could these trends tell you about how to help market this product? Or areas where you might want to explore further?

## Merging these two datasets together

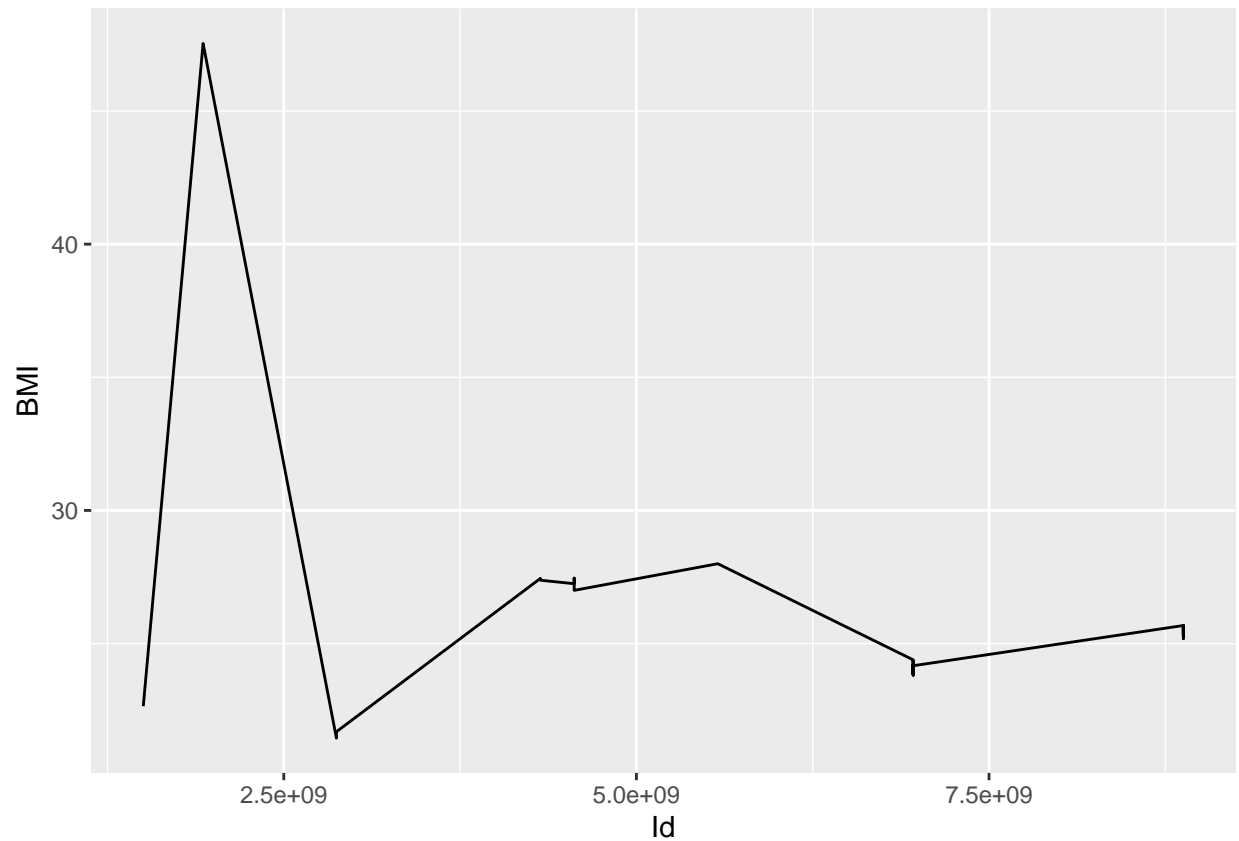
How Calories intake and BMI rate influences health?

```
combined_data <- merge(weight_log, daily_calories, by="Id")
ggplot(data=combined_data, aes(x=Calories, y=BMI)) +
  geom_line()
```



How many participants are there in data

```
ggplot(data=combined_data,aes(x=Id, y=BMI)) +  
  geom_line()
```



```
n_distinct(combined_data$Id)
```

```
## [1] 8
```

There were more participant Ids in the daily calories dataset that have been filtered out using merge.