PROYECTO\_GOOGLE\_2023

Juan Mario Moreno Chaparro

2023-01-15

## Cargamos las librerias

Notes: las librerias siguientes se pueden instalar con “install.packages”

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.2

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.2

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.1 ✔ stringr 1.5.0   
## ✔ readr 2.1.3 ✔ forcats 0.5.2   
## ✔ purrr 0.3.5

## Warning: package 'tibble' was built under R version 4.2.2

## Warning: package 'tidyr' was built under R version 4.2.2

## Warning: package 'readr' was built under R version 4.2.2

## Warning: package 'purrr' was built under R version 4.2.2

## Warning: package 'dplyr' was built under R version 4.2.2

## Warning: package 'stringr' was built under R version 4.2.2

## Warning: package 'forcats' was built under R version 4.2.2

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library(rmarkdown)

## Warning: package 'rmarkdown' was built under R version 4.2.2

library(skimr)

## Warning: package 'skimr' was built under R version 4.2.2

library(dplyr)  
library(janitor) #funciones para la limpieza de datos

## Warning: package 'janitor' was built under R version 4.2.2

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library("here") #Este paquete facilita la consulta de los archivos

## Warning: package 'here' was built under R version 4.2.2

## here() starts at C:/Users/moren/OneDrive/Documents/Google\_certifid

library(readr)

## Datos para analizar

para poder cargar un documentos cvs usamos la siguiente función de R

dailyActivity\_merged <- read\_csv("C:/Users/moren/OneDrive/Escritorio/Fitabase Data 4.12.16-5.12.16/dailyActivity\_merged.csv")

## Rows: 940 Columns: 15  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): ActivityDate  
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...  
##   
## 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.

View(dailyActivity\_merged)  
  
  
Proyecto\_LEFT\_JOIN\_VARIAS\_TABLAS <- read\_csv("C:/Users/moren/OneDrive/Escritorio/Fitabase Data 4.12.16-5.12.16/BIG\_QUERY\_CONSULTAS/Proyecto\_LEFT\_JOIN\_VARIAS\_TABLAS.csv")

## Rows: 943 Columns: 20  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): SleepHour  
## dbl (18): Id, Calories, TotalSteps, TotalDistance, TrackerDistance, LoggedA...  
## date (1): ActivityDate  
##   
## 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.

View(Proyecto\_LEFT\_JOIN\_VARIAS\_TABLAS)  
  
USUARIOS\_LEFT\_JOIN <- read\_csv("C:/Users/moren/OneDrive/Escritorio/Fitabase Data 4.12.16-5.12.16/BIG\_QUERY\_CONSULTAS/USUARIOS\_LEFT\_JOIN.csv")

## Rows: 943 Columns: 21  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (2): Usuario, SleepHour  
## dbl (18): Id, Calories, TotalSteps, TotalDistance, TrackerDistance, LoggedA...  
## date (1): ActivityDate  
##   
## 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.

View(USUARIOS\_LEFT\_JOIN)

## Reporte de datos

Usamos las siguientes funciones para que nos de un resumen de los datos que estamos usando.

skim\_without\_charts(dailyActivity\_merged) #resumen detallado de los datos

Data summary

|  |  |
| --- | --- |
| Name | dailyActivity\_merged |
| Number of rows | 940 |
| Number of columns | 15 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 1 |
| numeric | 14 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ActivityDate | 0 | 1 | 8 | 9 | 0 | 31 | 0 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Id | 0 | 1 | 4.855407e+09 | 2.424805e+09 | 1503960366 | 2.320127e+09 | 4.445115e+09 | 6.962181e+09 | 8.877689e+09 |
| TotalSteps | 0 | 1 | 7.637910e+03 | 5.087150e+03 | 0 | 3.789750e+03 | 7.405500e+03 | 1.072700e+04 | 3.601900e+04 |
| TotalDistance | 0 | 1 | 5.490000e+00 | 3.920000e+00 | 0 | 2.620000e+00 | 5.240000e+00 | 7.710000e+00 | 2.803000e+01 |
| TrackerDistance | 0 | 1 | 5.480000e+00 | 3.910000e+00 | 0 | 2.620000e+00 | 5.240000e+00 | 7.710000e+00 | 2.803000e+01 |
| LoggedActivitiesDistance | 0 | 1 | 1.100000e-01 | 6.200000e-01 | 0 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 4.940000e+00 |
| VeryActiveDistance | 0 | 1 | 1.500000e+00 | 2.660000e+00 | 0 | 0.000000e+00 | 2.100000e-01 | 2.050000e+00 | 2.192000e+01 |
| ModeratelyActiveDistance | 0 | 1 | 5.700000e-01 | 8.800000e-01 | 0 | 0.000000e+00 | 2.400000e-01 | 8.000000e-01 | 6.480000e+00 |
| LightActiveDistance | 0 | 1 | 3.340000e+00 | 2.040000e+00 | 0 | 1.950000e+00 | 3.360000e+00 | 4.780000e+00 | 1.071000e+01 |
| SedentaryActiveDistance | 0 | 1 | 0.000000e+00 | 1.000000e-02 | 0 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.100000e-01 |
| VeryActiveMinutes | 0 | 1 | 2.116000e+01 | 3.284000e+01 | 0 | 0.000000e+00 | 4.000000e+00 | 3.200000e+01 | 2.100000e+02 |
| FairlyActiveMinutes | 0 | 1 | 1.356000e+01 | 1.999000e+01 | 0 | 0.000000e+00 | 6.000000e+00 | 1.900000e+01 | 1.430000e+02 |
| LightlyActiveMinutes | 0 | 1 | 1.928100e+02 | 1.091700e+02 | 0 | 1.270000e+02 | 1.990000e+02 | 2.640000e+02 | 5.180000e+02 |
| SedentaryMinutes | 0 | 1 | 9.912100e+02 | 3.012700e+02 | 0 | 7.297500e+02 | 1.057500e+03 | 1.229500e+03 | 1.440000e+03 |
| Calories | 0 | 1 | 2.303610e+03 | 7.181700e+02 | 0 | 1.828500e+03 | 2.134000e+03 | 2.793250e+03 | 4.900000e+03 |

glimpse(dailyActivity\_merged) #resumen de las columnas

## Rows: 940  
## Columns: 15  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 150396036…  
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/…  
## $ TotalSteps <dbl> 13162, 10735, 10460, 9762, 12669, 9705, 13019…  
## $ TotalDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8…  
## $ TrackerDistance <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8…  
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ VeryActiveDistance <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5…  
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3…  
## $ LightActiveDistance <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0…  
## $ SedentaryActiveDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ VeryActiveMinutes <dbl> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4…  
## $ FairlyActiveMinutes <dbl> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21…  
## $ LightlyActiveMinutes <dbl> 328, 217, 181, 209, 221, 164, 233, 264, 205, …  
## $ SedentaryMinutes <dbl> 728, 776, 1218, 726, 773, 539, 1149, 775, 818…  
## $ Calories <dbl> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203…

head(dailyActivity\_merged)

## # A tibble: 6 × 15  
## Id Activ…¹ Total…² Total…³ Track…⁴ Logge…⁵ VeryA…⁶ Moder…⁷ Light…⁸ Seden…⁹  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1.50e9 4/12/2… 13162 8.5 8.5 0 1.88 0.550 6.06 0  
## 2 1.50e9 4/13/2… 10735 6.97 6.97 0 1.57 0.690 4.71 0  
## 3 1.50e9 4/14/2… 10460 6.74 6.74 0 2.44 0.400 3.91 0  
## 4 1.50e9 4/15/2… 9762 6.28 6.28 0 2.14 1.26 2.83 0  
## 5 1.50e9 4/16/2… 12669 8.16 8.16 0 2.71 0.410 5.04 0  
## 6 1.50e9 4/17/2… 9705 6.48 6.48 0 3.19 0.780 2.51 0  
## # … with 5 more variables: VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,  
## # LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>, and  
## # abbreviated variable names ¹​ActivityDate, ²​TotalSteps, ³​TotalDistance,  
## # ⁴​TrackerDistance, ⁵​LoggedActivitiesDistance, ⁶​VeryActiveDistance,  
## # ⁷​ModeratelyActiveDistance, ⁸​LightActiveDistance, ⁹​SedentaryActiveDistance

skim\_without\_charts(USUARIOS\_LEFT\_JOIN)

Data summary

|  |  |
| --- | --- |
| Name | USUARIOS\_LEFT\_JOIN |
| Number of rows | 943 |
| Number of columns | 21 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 2 |
| Date | 1 |
| numeric | 18 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Usuario | 0 | 1.00 | 9 | 10 | 0 | 33 | 0 |
| SleepHour | 530 | 0.44 | 8 | 14 | 0 | 2 | 0 |

**Variable type: Date**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| ActivityDate | 0 | 1 | 2016-04-12 | 2016-05-12 | 2016-04-26 | 31 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Id | 0 | 1.00 | 4.858486e+09 | 2.423712e+09 | 1503960366 | 2.320127e+09 | 4.445115e+09 | 6.962181e+09 | 8.877689e+09 |
| Calories | 0 | 1.00 | 2.307510e+03 | 7.208200e+02 | 0 | 1.829500e+03 | 2.140000e+03 | 2.796500e+03 | 4.900000e+03 |
| TotalSteps | 0 | 1.00 | 7.652190e+03 | 5.086530e+03 | 0 | 3.795000e+03 | 7.439000e+03 | 1.073400e+04 | 3.601900e+04 |
| TotalDistance | 0 | 1.00 | 5.500000e+00 | 3.930000e+00 | 0 | 2.620000e+00 | 5.260000e+00 | 7.720000e+00 | 2.803000e+01 |
| TrackerDistance | 0 | 1.00 | 5.490000e+00 | 3.910000e+00 | 0 | 2.620000e+00 | 5.260000e+00 | 7.710000e+00 | 2.803000e+01 |
| LoggedActivitiesDistance | 0 | 1.00 | 1.100000e-01 | 6.200000e-01 | 0 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 4.940000e+00 |
| StepTotal | 0 | 1.00 | 7.652190e+03 | 5.086530e+03 | 0 | 3.795000e+03 | 7.439000e+03 | 1.073400e+04 | 3.601900e+04 |
| TotalSleepRecords | 530 | 0.44 | 1.120000e+00 | 3.500000e-01 | 1 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 | 3.000000e+00 |
| TotalMinutesAsleep | 530 | 0.44 | 4.194700e+02 | 1.183400e+02 | 58 | 3.610000e+02 | 4.330000e+02 | 4.900000e+02 | 7.960000e+02 |
| TotalTimeInBed | 530 | 0.44 | 4.586400e+02 | 1.271000e+02 | 61 | 4.030000e+02 | 4.630000e+02 | 5.260000e+02 | 9.610000e+02 |
| SedentaryMinutes | 0 | 1.00 | 9.903500e+02 | 3.012600e+02 | 0 | 7.290000e+02 | 1.057000e+03 | 1.229000e+03 | 1.440000e+03 |
| LightlyActiveMinutes | 0 | 1.00 | 1.930300e+02 | 1.093100e+02 | 0 | 1.270000e+02 | 1.990000e+02 | 2.640000e+02 | 5.180000e+02 |
| FairlyActiveMinutes | 0 | 1.00 | 1.363000e+01 | 2.000000e+01 | 0 | 0.000000e+00 | 7.000000e+00 | 1.900000e+01 | 1.430000e+02 |
| VeryActiveMinutes | 0 | 1.00 | 2.124000e+01 | 3.295000e+01 | 0 | 0.000000e+00 | 4.000000e+00 | 3.200000e+01 | 2.100000e+02 |
| SedentaryActiveDistance | 0 | 1.00 | 0.000000e+00 | 1.000000e-02 | 0 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.100000e-01 |
| LightActiveDistance | 0 | 1.00 | 3.350000e+00 | 2.050000e+00 | 0 | 1.950000e+00 | 3.380000e+00 | 4.790000e+00 | 1.071000e+01 |
| ModeratelyActiveDistance | 0 | 1.00 | 5.700000e-01 | 8.800000e-01 | 0 | 0.000000e+00 | 2.400000e-01 | 8.100000e-01 | 6.480000e+00 |
| VeryActiveDistance | 0 | 1.00 | 1.500000e+00 | 2.660000e+00 | 0 | 0.000000e+00 | 2.200000e-01 | 2.060000e+00 | 2.192000e+01 |

glimpse(USUARIOS\_LEFT\_JOIN)

## Rows: 943  
## Columns: 21  
## $ Id <dbl> 1624580081, 1644430081, 1644430081, 164443008…  
## $ Usuario <chr> "Usuario 1", "Usuario 2", "Usuario 2", "Usuar…  
## $ Calories <dbl> 2690, 3226, 3300, 3108, 3846, 3324, 2897, 270…  
## $ ActivityDate <date> 2016-05-01, 2016-04-14, 2016-04-19, 2016-04-…  
## $ TotalSteps <dbl> 36019, 11037, 11256, 9405, 18213, 12850, 1511…  
## $ TotalDistance <dbl> 28.03, 8.02, 8.18, 6.84, 13.24, 9.34, 10.67, …  
## $ TrackerDistance <dbl> 28.03, 8.02, 8.18, 6.84, 13.24, 9.34, 10.67, …  
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ StepTotal <dbl> 36019, 11037, 11256, 9405, 18213, 12850, 1511…  
## $ SleepHour <chr> NA, NA, NA, NA, "00:00:00", NA, NA, NA, "00:0…  
## $ TotalSleepRecords <dbl> NA, NA, NA, NA, 1, NA, NA, NA, 1, NA, NA, 1, …  
## $ TotalMinutesAsleep <dbl> NA, NA, NA, NA, 124, NA, NA, NA, 445, NA, NA,…  
## $ TotalTimeInBed <dbl> NA, NA, NA, NA, 142, NA, NA, NA, 489, NA, NA,…  
## $ SedentaryMinutes <dbl> 1020, 1125, 1099, 1157, 816, 1115, 1053, 1061…  
## $ LightlyActiveMinutes <dbl> 171, 252, 278, 227, 402, 221, 276, 297, 206, …  
## $ FairlyActiveMinutes <dbl> 63, 58, 58, 53, 71, 94, 63, 47, 48, 72, 43, 8…  
## $ VeryActiveMinutes <dbl> 186, 5, 5, 3, 9, 10, 48, 35, 1, 66, 11, 31, 1…  
## $ SedentaryActiveDistance <dbl> 0.02, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.0…  
## $ LightActiveDistance <dbl> 1.91, 5.10, 5.30, 4.31, 9.46, 4.54, 5.40, 5.6…  
## $ ModeratelyActiveDistance <dbl> 4.19, 2.56, 2.53, 2.32, 3.14, 4.09, 1.93, 1.6…  
## $ VeryActiveDistance <dbl> 21.92, 0.36, 0.36, 0.20, 0.63, 0.72, 3.34, 2.…

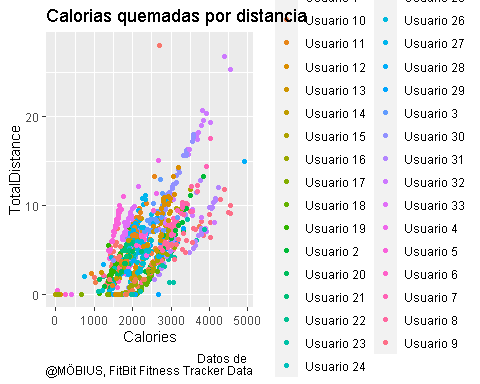
head(USUARIOS\_LEFT\_JOIN)

## # A tibble: 6 × 21  
## Id Usuario Calor…¹ Activity…² Total…³ Total…⁴ Track…⁵ Logge…⁶ StepT…⁷  
## <dbl> <chr> <dbl> <date> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1624580081 Usuario… 2690 2016-05-01 36019 28.0 28.0 0 36019  
## 2 1644430081 Usuario… 3226 2016-04-14 11037 8.02 8.02 0 11037  
## 3 1644430081 Usuario… 3300 2016-04-19 11256 8.18 8.18 0 11256  
## 4 1644430081 Usuario… 3108 2016-04-28 9405 6.84 6.84 0 9405  
## 5 1644430081 Usuario… 3846 2016-04-30 18213 13.2 13.2 0 18213  
## 6 1644430081 Usuario… 3324 2016-05-03 12850 9.34 9.34 0 12850  
## # … with 12 more variables: SleepHour <chr>, TotalSleepRecords <dbl>,  
## # TotalMinutesAsleep <dbl>, TotalTimeInBed <dbl>, SedentaryMinutes <dbl>,  
## # LightlyActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,  
## # VeryActiveMinutes <dbl>, SedentaryActiveDistance <dbl>,  
## # LightActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,  
## # VeryActiveDistance <dbl>, and abbreviated variable names ¹​Calories,  
## # ²​ActivityDate, ³​TotalSteps, ⁴​TotalDistance, ⁵​TrackerDistance, …

## Gráficas

En primera tenemos una grafica de puntos que relaciona la distancia con las calorias

ggplot(data = USUARIOS\_LEFT\_JOIN) + geom\_point((mapping =   
 aes(x = Calories,  
 y = TotalDistance, color=Usuario)))+  
 labs(title="Calorias quemadas por distancia",caption= "Datos de   
 @MÖBIUS, FitBit Fitness Tracker Data")

 ### Después tenemos Los pasos y el rastreador de ellos, esto para saber si el dispositivo concuerda con los pasos del usario

ggplot(data = USUARIOS\_LEFT\_JOIN) + geom\_smooth((mapping =   
 aes(x= TrackerDistance,   
 y=TotalSteps, color=Usuario)))+   
 geom\_jitter((mapping =  
   
 aes(x= TrackerDistance, y=TotalSteps, color=Usuario)))+   
 labs(title="Total de pasos vs DIstancia del rastreador",caption= "Datos de   
 @MÖBIUS, FitBit Fitness Tracker Data")

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : span too small. fewer data values than degrees of freedom.

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at -0.02235

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 4.0524

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 2.3788

## Warning in sqrt(sum.squares/one.delta): Se han producido NaNs

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : span too small. fewer  
## data values than degrees of freedom.

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## -0.02235

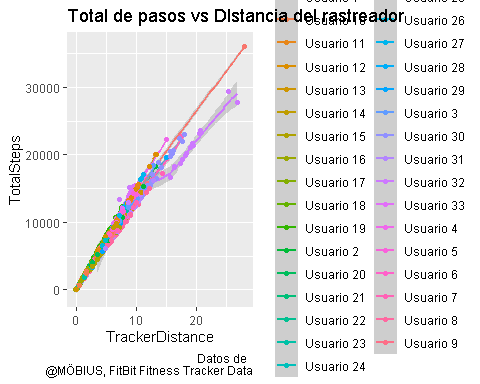
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 4.0524

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 0

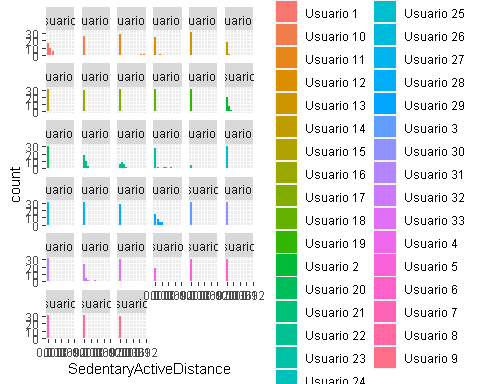
## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other near  
## singularities as well. 2.3788

## Warning in stats::qt(level/2 + 0.5, pred$df): NaNs produced

## Warning in max(ids, na.rm = TRUE): ningun argumento finito para max; retornando  
## -Inf

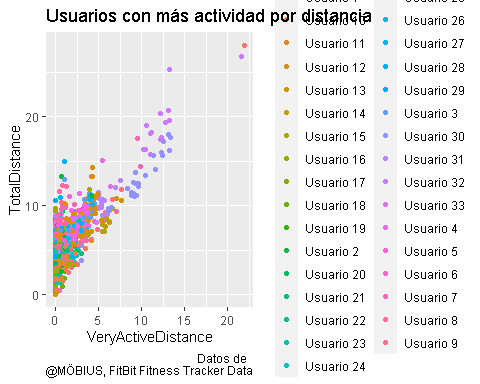
 ###Sedentarismo Niveles de Sedentarismo por usario.

ggplot(data = USUARIOS\_LEFT\_JOIN)+  
 geom\_bar(mapping=aes(x=SedentaryActiveDistance, fill=Usuario))+  
 facet\_wrap(~Usuario)



Tenemos usuarios con más actividad donde se ve que entre más distancia más actividad

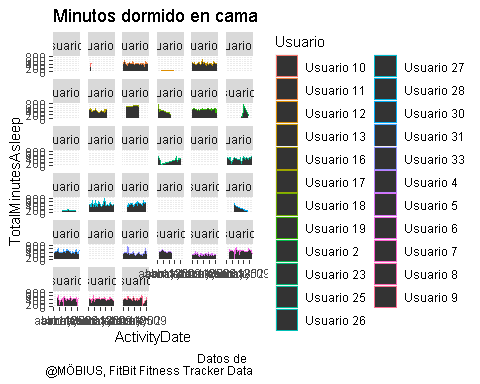
ggplot(data = USUARIOS\_LEFT\_JOIN) + geom\_point((mapping =   
 aes(x = VeryActiveDistance,  
 y = TotalDistance, color=Usuario)))+  
 labs(title="Usuarios con más actividad por distancia",caption= "Datos de   
 @MÖBIUS, FitBit Fitness Tracker Data")



Las horas de sueños vistas en minutos por usuario

ggplot(data = USUARIOS\_LEFT\_JOIN) + geom\_area((mapping=  
 aes(x = ActivityDate,   
 y=TotalMinutesAsleep, color=Usuario))) +   
 facet\_wrap(~Usuario)+  
 labs(title="Minutos dormido en cama",caption= "Datos de   
 @MÖBIUS, FitBit Fitness Tracker Data")

## Warning: Removed 530 rows containing non-finite values (`stat\_align()`).



##Estadísticas

Tenemos que las estadísticas de los datos nos da como resultado algunas respuestas sobre los usuarios. Por ejemplo podemos ver que la media de cada usuario “Mean” de calorias y la de Pasos totales.  
La desviación Estándar “sd” que nos la dispersión de los datos, tenemos en cuenta a los usuarios y sus disperción en pasos totales y calorias. La correlación es la que tan relacionados están los datos, entre más cerca de 1 más se relacionan.

USUARIOS\_LEFT\_JOIN %>%   
 group\_by(Usuario) %>%   
 summarise(mean(Calories), sd(Calories), mean(TotalSteps), sd(TotalSteps),   
 cor(Calories, TotalSteps))

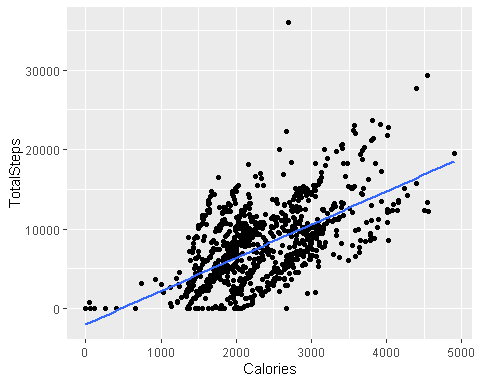
## # A tibble: 33 × 6  
## Usuario `mean(Calories)` `sd(Calories)` `mean(TotalSteps)` sd(To…¹ cor(C…²  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Usuario 1 1483. 257. 5744. 6177. 0.931  
## 2 Usuario 10 2132. 484. 2520. 3028. 0.859  
## 3 Usuario 11 1982. 296. 9795. 3942. 0.888  
## 4 Usuario 12 2544 629. 11323. 5306. 0.888  
## 5 Usuario 13 2566. 436. 9372. 3857. 0.884  
## 6 Usuario 14 1788 467. 6482. 3141. 0.765  
## 7 Usuario 15 2732. 571. 7199. 3402. 0.764  
## 8 Usuario 16 1962. 545. 1854. 2327. 0.827  
## 9 Usuario 17 1573. 308. 2580. 2713. 0.917  
## 10 Usuario 18 2173. 221. 916. 1205. 0.822  
## # … with 23 more rows, and abbreviated variable names ¹​`sd(TotalSteps)`,  
## # ²​`cor(Calories, TotalSteps)`

##Gráficas

Podemos ver en la siguiente grafica, como se ve los datos estadísticos de forma gráfica.

ggplot(USUARIOS\_LEFT\_JOIN, aes(Calories, TotalSteps)) +  
 geom\_point() + geom\_smooth(method = lm, se=FALSE)

## `geom\_smooth()` using formula = 'y ~ x'

 Tenemos ahora las estadisticas de dos variables que por su nombre puede que tengan relación. “Sendentarismo” y Poca “Avtividad” Pero el resultado es otro, su relación es negativa, esto quiere decir que mientras una variable crece la otra decrece.

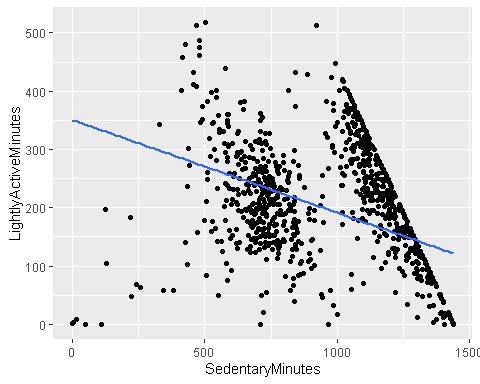
USUARIOS\_LEFT\_JOIN %>%   
 group\_by(Usuario) %>%   
 summarise(mean(SedentaryMinutes), sd(SedentaryMinutes), mean(LightlyActiveMinutes),   
 sd(LightlyActiveMinutes),   
 cor(SedentaryMinutes, LightlyActiveMinutes))

## # A tibble: 33 × 6  
## Usuario `mean(SedentaryMinutes)` sd(SedentaryMin…¹ mean(…² sd(Li…³ cor(S…⁴  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Usuario 1 1258. 94.0 153. 40.9 -0.327   
## 2 Usuario 10 1299. 221. 40.2 49.9 -0.561   
## 3 Usuario 11 662. 125. 246. 67.0 0.0533  
## 4 Usuario 12 1055. 218. 281. 106. 0.102   
## 5 Usuario 13 850. 264. 144. 59.9 -0.498   
## 6 Usuario 14 1287. 64.6 117. 55.9 -0.943   
## 7 Usuario 15 1267. 111. 138. 86.9 -0.953   
## 8 Usuario 16 1060. 371. 91.8 107. -0.620   
## 9 Usuario 17 1207. 316. 115. 123. -0.455   
## 10 Usuario 18 1317. 187. 38.6 50.6 -0.398   
## # … with 23 more rows, and abbreviated variable names ¹​`sd(SedentaryMinutes)`,  
## # ²​`mean(LightlyActiveMinutes)`, ³​`sd(LightlyActiveMinutes)`,  
## # ⁴​`cor(SedentaryMinutes, LightlyActiveMinutes)`

Gráfico

ggplot(USUARIOS\_LEFT\_JOIN, aes(SedentaryMinutes, LightlyActiveMinutes)) +  
 geom\_point() + geom\_smooth(method = lm, se=FALSE)

## `geom\_smooth()` using formula = 'y ~ x'

 Tenemos ahora la media de mucha distancia activa y Sedentarismo activo por Usuario.

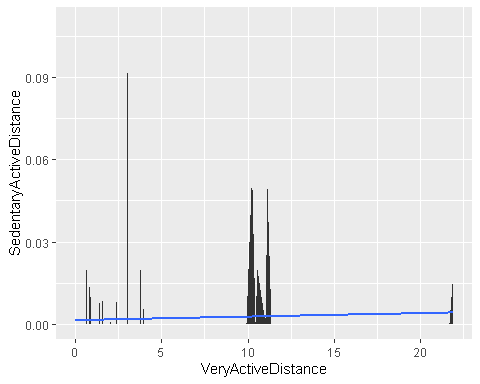
USUARIOS\_LEFT\_JOIN %>%   
 group\_by(Usuario) %>%   
 summarise(mean(VeryActiveDistance), mean(SedentaryActiveDistance))

## # A tibble: 33 × 3  
## Usuario `mean(VeryActiveDistance)` `mean(SedentaryActiveDistance)`  
## <chr> <dbl> <dbl>  
## 1 Usuario 1 0.939 0.00613   
## 2 Usuario 10 0.709 0   
## 3 Usuario 11 1.62 0.00677   
## 4 Usuario 12 2.41 0.000769  
## 5 Usuario 13 2.78 0   
## 6 Usuario 14 2.21 0.000526  
## 7 Usuario 15 0.798 0   
## 8 Usuario 16 0.0248 0   
## 9 Usuario 17 0.00839 0   
## 10 Usuario 18 0.0958 0   
## # … with 23 more rows

Gráfico

ggplot(USUARIOS\_LEFT\_JOIN, aes(VeryActiveDistance, SedentaryActiveDistance)) +  
 geom\_area() + geom\_smooth(method = lm, se=FALSE)

## `geom\_smooth()` using formula = 'y ~ x'



#Conclusiones

Como Conclusión la empresa bellabeat podria mejorar la calidad de sus productos, con publicidad enfocada a correr, hacer ejercicio, etc. La gráfica de calorias, indica que no muchos usarios no usan debidamente los productos para su cuidado.

A lo que lleva que el Sedentarismo sea más alto en los usuarios.

Veremos más a fondo las gráficas con Tableu para llegar a está conclusión.

Gracias! Seguire mejorando.