

PROYECTO_GOOGLE_2023

Juan Mario Moreno Chaparro

2023-01-15

Cargamos las librerías

Notes: las librerías 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	4.855 407e +09	2.424 805e +09	1503 9603 66	2.320 127e +09	4.445 115e +09	6.962 181e +09	8.877 689e +09
TotalSteps	0	1	7.637 910e +03	5.087 150e +03	0	3.789 750e +03	7.405 500e +03	1.072 700e +04	3.601 900e +04
TotalDistance	0	1	5.490 000e +00	3.920 000e +00	0	2.620 000e +00	5.240 000e +00	7.710 000e +00	2.803 000e +01
TrackerDistance	0	1	5.480 000e +00	3.910 000e +00	0	2.620 000e +00	5.240 000e +00	7.710 000e +00	2.803 000e +01
LoggedActivitiesDistance	0	1	1.100 000e- 01	6.200 000e- 01	0	0.000 000e +00	0.000 000e +00	0.000 000e +00	4.940 000e +00
VeryActiveDistance	0	1	1.500 000e +00	2.660 000e +00	0	0.000 000e +00	2.100 000e- 01	2.050 000e +00	2.192 000e +01
ModeratelyActiveDistance	0	1	5.700 000e- 01	8.800 000e- 01	0	0.000 000e +00	2.400 000e- 01	8.000 000e- 01	6.480 000e +00
LightActiveDistance	0	1	3.340 000e +00	2.040 000e +00	0	1.950 000e +00	3.360 000e +00	4.780 000e +00	1.071 000e +01
SedentaryActiveDistance	0	1	0.000 000e +00	1.000 000e- 02	0	0.000 000e +00	0.000 000e +00	0.000 000e +00	1.100 000e- 01
VeryActiveMinutes	0	1	2.116 000e +01	3.284 000e +01	0	0.000 000e +00	4.000 000e +00	3.200 000e +01	2.100 000e +02

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
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, 1
50396036...
## $ ActivityDate <chr> "4/12/2016", "4/13/2016", "4/14/2016"
, "4/15/...
## $ TotalSteps <dbl> 13162, 10735, 10460, 9762, 12669, 970
5, 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, ...
## $ 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, ...
## $ VeryActiveMinutes <dbl> 25, 21, 30, 29, 36, 38, 42, 50, 28, 1
9, 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, 26
4, 205, ...
## $ SedentaryMinutes <dbl> 728, 776, 1218, 726, 773, 539, 1149,
775, 818...
```

```
## $ Calories <dbl> 1985, 1797, 1776, 1745, 1863, 1728, 1
921, 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>      <db
l>      <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>, FairlyActiveMinute
s <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
, and
## #   abbreviated variable names 1ActivityDate, 2TotalSteps, 3TotalDista
nce,
## #   4TrackerDistance, 5LoggedActivitiesDistance, 6VeryActiveDistance,
## #   7ModeratelyActiveDistance, 8LightActiveDistance, 9SedentaryActiveD
istance

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	4.858 486e +09	2.423 712e +09	1503 9603 66	2.320 127e +09	4.445 115e +09	6.962 181e +09	8.877 689e +09
Calories	0	1.00	2.307 510e +03	7.208 200e +02	0	1.829 500e +03	2.140 000e +03	2.796 500e +03	4.900 000e +03
TotalSteps	0	1.00	7.652 190e +03	5.086 530e +03	0	3.795 000e +03	7.439 000e +03	1.073 400e +04	3.601 900e +04
TotalDistance	0	1.00	5.500 000e +00	3.930 000e +00	0	2.620 000e +00	5.260 000e +00	7.720 000e +00	2.803 000e +01
TrackerDistance	0	1.00	5.490 000e +00	3.910 000e +00	0	2.620 000e +00	5.260 000e +00	7.710 000e +00	2.803 000e +01
LoggedActivitiesDistance	0	1.00	1.100 000e- 01	6.200 000e- 01	0	0.000 000e +00	0.000 000e +00	0.000 000e +00	4.940 000e +00
StepTotal	0	1.00	7.652 190e +03	5.086 530e +03	0	3.795 000e +03	7.439 000e +03	1.073 400e +04	3.601 900e +04
TotalSleepRecords	530	0.44	1.120 000e +00	3.500 000e- 01	1	1.000 000e +00	1.000 000e +00	1.000 000e +00	3.000 000e +00
TotalMinutesAsleep	530	0.44	4.194 700e +02	1.183 400e +02	58	3.610 000e +02	4.330 000e +02	4.900 000e +02	7.960 000e +02

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
TotalTimeInBed	530	0.44	4.586 400e +02	1.271 000e +02	61	4.030 000e +02	4.630 000e +02	5.260 000e +02	9.610 000e +02
SedentaryMinutes	0	1.00	9.903 500e +02	3.012 600e +02	0	7.290 000e +02	1.057 000e +03	1.229 000e +03	1.440 000e +03
LightlyActiveMinutes	0	1.00	1.930 300e +02	1.093 100e +02	0	1.270 000e +02	1.990 000e +02	2.640 000e +02	5.180 000e +02
FairlyActiveMinutes	0	1.00	1.363 000e +01	2.000 000e +01	0	0.000 000e +00	7.000 000e +00	1.900 000e +01	1.430 000e +02
VeryActiveMinutes	0	1.00	2.124 000e +01	3.295 000e +01	0	0.000 000e +00	4.000 000e +00	3.200 000e +01	2.100 000e +02
SedentaryActiveDistance	0	1.00	0.000 000e +00	1.000 000e- 02	0	0.000 000e +00	0.000 000e +00	0.000 000e +00	1.100 000e- 01
LightActiveDistance	0	1.00	3.350 000e +00	2.050 000e +00	0	1.950 000e +00	3.380 000e +00	4.790 000e +00	1.071 000e +01
ModeratelyActiveDistance	0	1.00	5.700 000e- 01	8.800 000e- 01	0	0.000 000e +00	2.400 000e- 01	8.100 000e- 01	6.480 000e +00
VeryActiveDistance	0	1.00	1.500 000e +00	2.660 000e +00	0	0.000 000e +00	2.200 000e- 01	2.060 000e +00	2.192 000e +01

```
glimpse(USUARIOS_LEFT_JOIN)
```

```
## Rows: 943
## Columns: 21
## $ Id                <dbl> 1624580081, 1644430081, 1644430081, 1
64443008...
## $ Usuario           <chr> "Usuario 1", "Usuario 2", "Usuario 2"
, "Usuar...
## $ Calories          <dbl> 2690, 3226, 3300, 3108, 3846, 3324, 2
897, 270...
## $ ActivityDate      <date> 2016-05-01, 2016-04-14, 2016-04-19,
2016-04-...
## $ TotalSteps        <dbl> 36019, 11037, 11256, 9405, 18213, 128
50, 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, 128
50, 1511...
## $ SleepHour            <chr> NA, NA, NA, NA, "00:00:00", NA, NA, N
A, "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, 10
53, 1061...
## $ LightlyActiveMinutes  <dbl> 171, 252, 278, 227, 402, 221, 276, 29
7, 206, ...
## $ FairlyActiveMinutes   <dbl> 63, 58, 58, 53, 71, 94, 63, 47, 48, 7
2, 43, 8...
## $ VeryActiveMinutes     <dbl> 186, 5, 5, 3, 9, 10, 48, 35, 1, 66, 1
1, 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>    <db
l>    <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

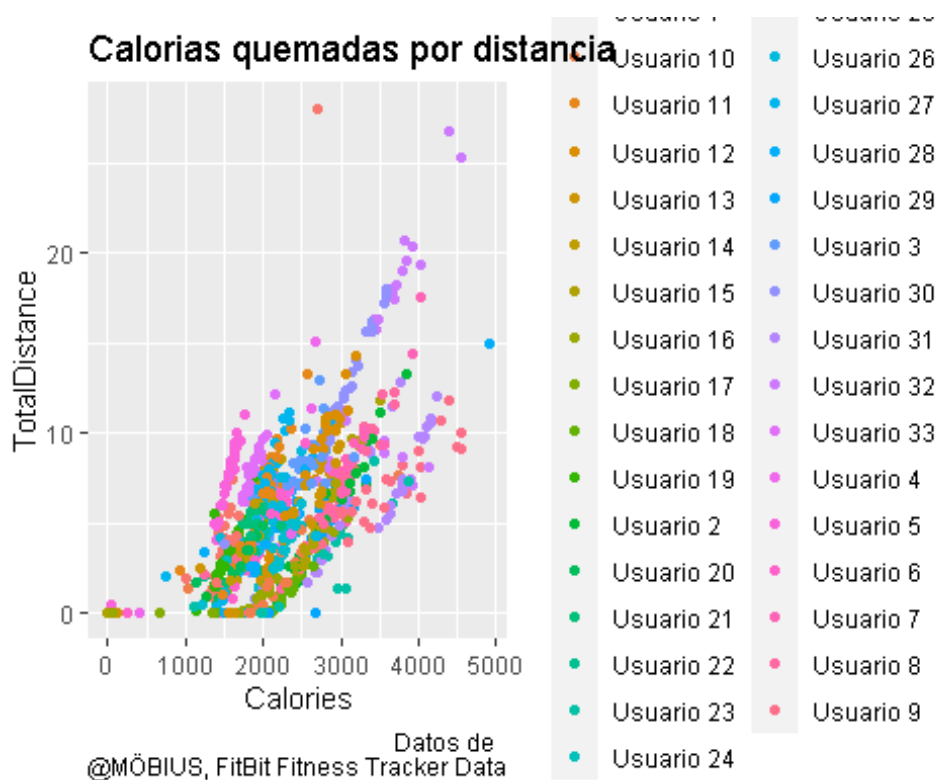
```

```
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 1Calories
## # 2ActivityDate, 3TotalSteps, 4TotalDistance, 5TrackerDistance, ...
```

Gráficas

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

```
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 usuario

```
ggplot(data = USUARIOS_LEFT_JOIN) + geom_smooth(mapping =
  aes(x= TrackerDistance
    ,
```

```

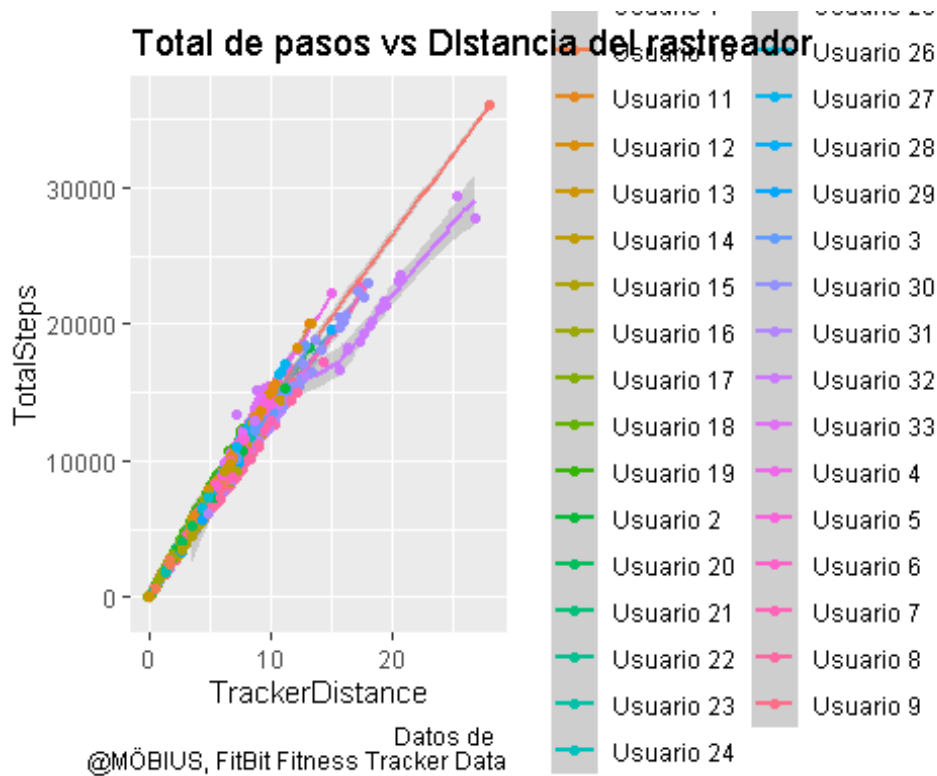
y=TotalSteps, color=
r=Usuario)))+
  geom_jitter(mapping =
    aes(x= TrackerDistance, y=TotalSteps, color=Usuario)))+
  labs(title="Total de pasos vs Distanacia del rastreador",caption= "Datos
de
  @MÖBIUS, FitBit Fitness Tracker Data")
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =
## parametric, : span too small. fewer data values than degrees of freedo
m.
## 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$x else if (is.data.frame(newdata))
## as.matrix(model.frame(delete.response(terms(object))), : neighborhood r
adius
## 4.0524
## Warning in predLoess(object$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 con
dition
## 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
## singularities as well. 2.3788

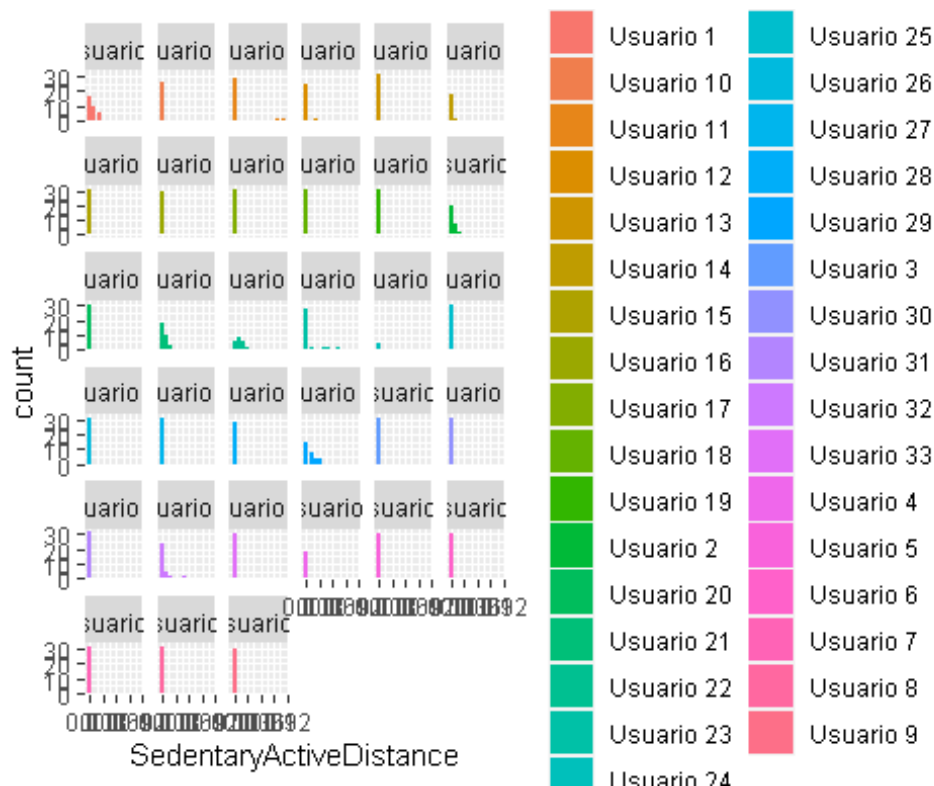
## Warning in stats::qt(level/2 + 0.5, pred$df): NaNs produced

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



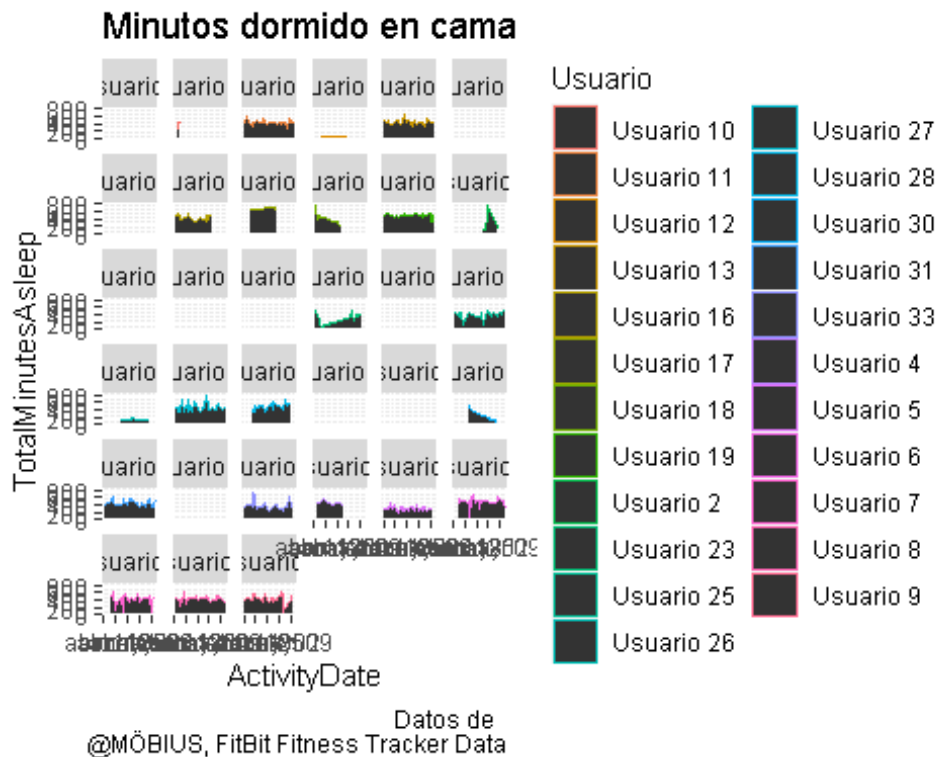
###Sedentarismo Niveles de Sedentarismo por usuario.

```
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=Userario)) +
  labs(title="Usuarios con más actividad por distancia",caption= "Datos de
    @MÖBIUS, FitBit Fitness Tracker Data")
```

##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 calorías y la de Pasos totales.

La desviación Estándar “sd” que nos la dispersión de los datos, tenemos en cuenta a los usuarios y sus dispersión en pasos totales y calorías. 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(TotalSteps)`	cor(Calories, TotalSteps)
1	Usuario 1	1483.	257.	5744.	617	0.931
2	Usuario 10	2132.	484.	2520.	302	0.859
3	Usuario 11	1982.	296.	9795.	394	0.888
4	Usuario 12	2544	629.	11323.	530	

```

6.    0.888
## 5 Usuario 13          2566.          436.          9372.    385
7.    0.884
## 6 Usuario 14          1788          467.          6482.    314
1.    0.765
## 7 Usuario 15          2732.          571.          7199.    340
2.    0.764
## 8 Usuario 16          1962.          545.          1854.    232
7.    0.827
## 9 Usuario 17          1573.          308.          2580.    271
3.    0.917
## 10 Usuario 18         2173.          221.           916.    120
5.    0.822
## # ... with 23 more rows, and abbreviated variable names 1`sd(TotalSteps)
`,`
## #    2`cor(Calories, TotalSteps)`

```

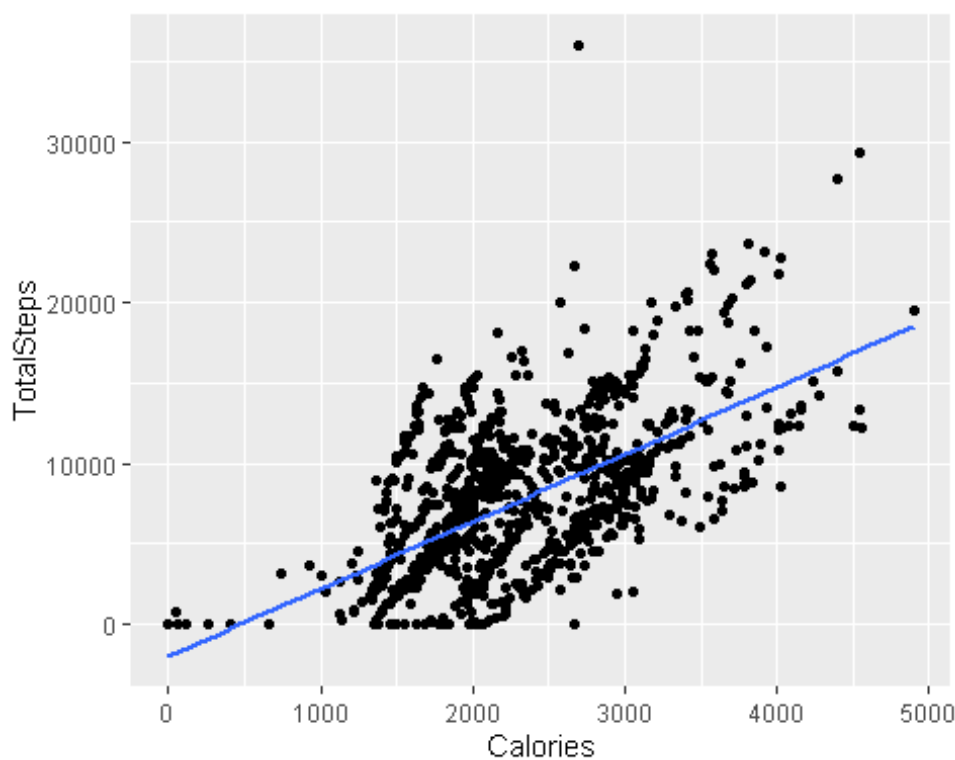
##Gráficas

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 estadísticas de dos variables que por su nombre puede que tengan relación.

“Sedentarismo” y Poca “Actividad” Pero el resultado es otro, su relación es negativa, esto quiere decir que mientras una variable crece la otra decrece.

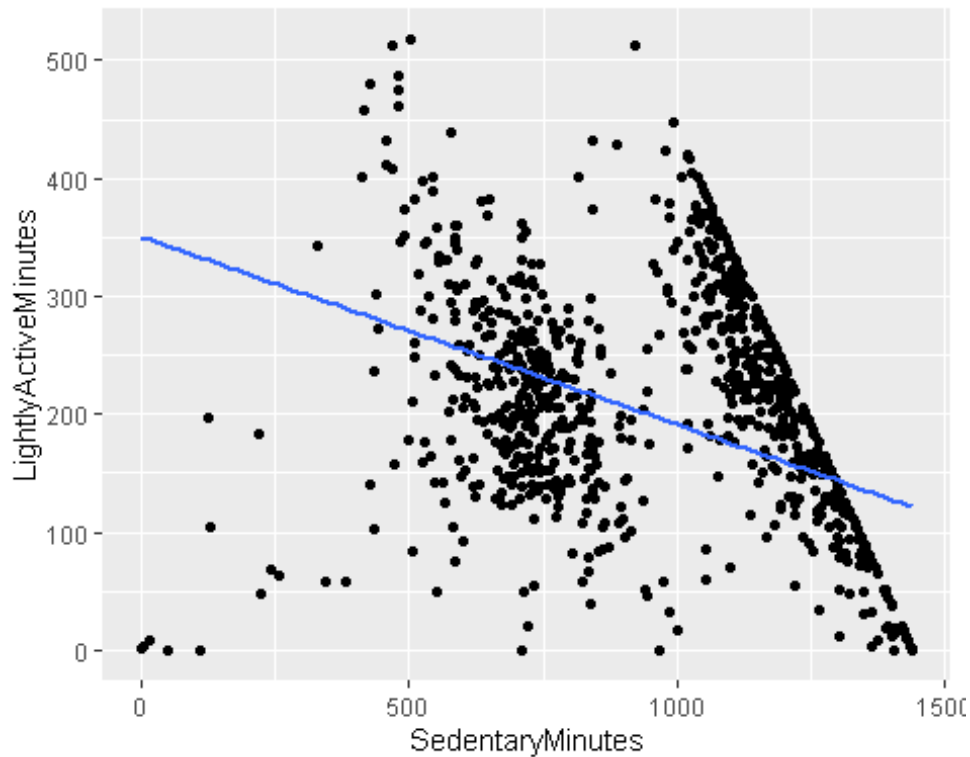
```
USUARIOS_LEFT_JOIN %>%
  group_by(Usuario) %>%
  summarise(mean(SedentaryMinutes), sd(SedentaryMinutes), mean(LightlyActiveMinutes),
            sd(LightlyActiveMinutes),
            cor(SedentaryMinutes, LightlyActiveMinutes))

## # A tibble: 33 × 6
##   Usuario `mean(SedentaryMinutes)` sd(SedentaryMin...1 mean(...2 sd(Li
...3 cor(S...4
##   <chr>                <dbl>                <dbl>    <dbl>    <db
l>    <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(SedentaryMi
nutes)`,
## # 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))
```

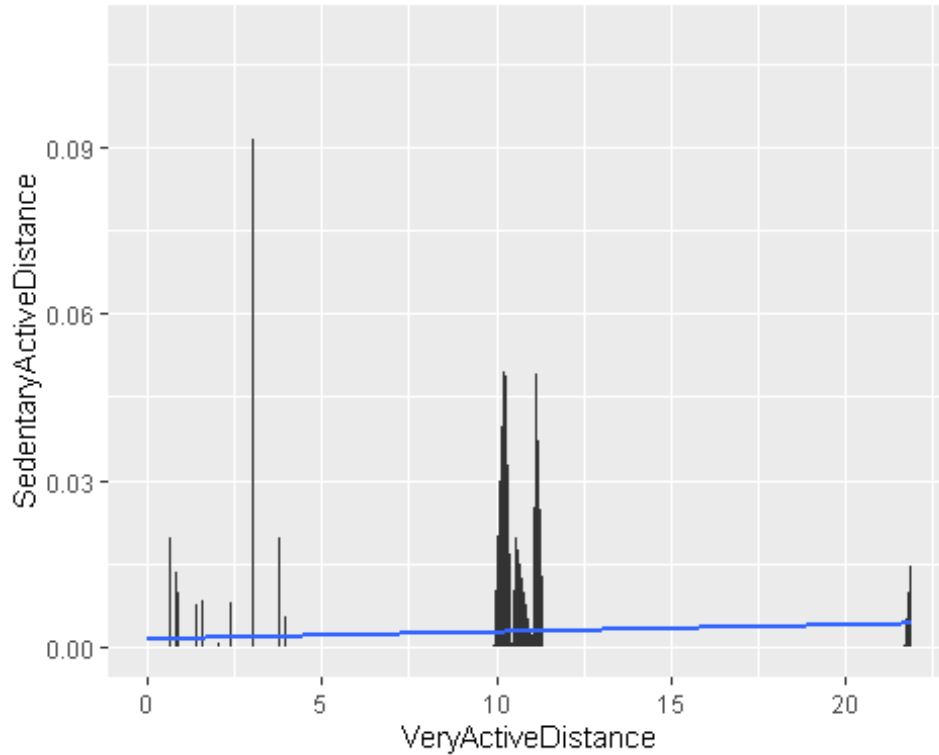
A tibble: 33 × 3

Usuario	mean(VeryActiveDistance)	mean(SedentaryActiveDistance)
1 Usuario 1	0.939	0.0061
2 Usuario 10	0.709	0
3 Usuario 11	1.62	0.0067
4 Usuario 12	2.41	0.0007
5 Usuario 13	2.78	0
6 Usuario 14	2.21	0.0005
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 calorías, indica que no muchos usuarios 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 Tableau para llegar a está conclusión.

Gracias! Seguire mejorando.