

# Google Certificate Case Study

Alejandro Hernandez

2024-02-17

## Bellabeat case

### Escenario:

- Trabajo en una empresa llamada *Bellabeat*: Una empresa dedicada a los dispositivos inteligentes enfocados en la salud, mi trabajo es como junior data analyst.
- Se me pide analizar datos sobre el uso de los dispositivos inteligentes enfocados en la salud para conocer tendencias de uso de los usuarios.
- Los personas involucradas son:
  - Urška Sršen: Cofundador y Chief Creative Officer
  - Sando Mur: Matemático y cofundador; miembro clave del equipo ejecutivo de Bellabeat
  - Equipo de analytics y,
  - Equipo de marketing de Bellabeat

### Fase 1: Ask

**Business task:** Conocer las tendencias del mercado de los dispositivos inteligentes en la salud para aplicarlas en mejorar la experiencia de los clientes de Bellabeat y a orientar la estrategia de marketing.

### Stakeholders:

- Urška Sršen: Cofundador y Chief Creative Officer.
- Sando Mur: Matemático y cofundador; miembro clave del equipo ejecutivo de Bellabeat.
- Equipo de analytics y de marketing de Bellabeat.

### Fase 2: Prepare

Este análisis se realizó en Rstudio. Se ocuparon los siguiente packages:

- tidyverse
- markdown
- janitor
- lubridate

```
library(tidyverse)
library(rmarkdown)
library(janitor)
library(lubridate)
```

Se ocupó el dataset FitBit Fitness Tracker Data obtenido de Kaggle Link a la página, los cuales fueron publicados por el usuario de Kaggle: MÖBIUS. De igual forma, comparte la liga a la fuente original de los datos Link a la página.

Cuenta con 18 documentos en formato .csv.

```
daily_activity <- read_csv("C:\\Users\\magno\\OneDrive\\Escritorio\\Curso Análisis de datos\\Modulo 8\\sleep_day_merged <- read_csv("C:\\Users\\magno\\OneDrive\\Escritorio\\Curso Análisis de datos\\Modulo 8\\
```

### Fase 3: Process

Los pasos que se siguieron para realizar la limpieza de datos fueron los siguientes:

- Cargar los packages a usar.
- Cargar los datos con la función `read_csv()`.
- Se exploró los datos con la función `head()` y `str()`.

```
head(daily_activity)
```

```
## # A tibble: 6 x 15
##       Id ActivityDate TotalSteps TotalDistance TrackerDistance
##       <dbl> <chr>          <dbl>          <dbl>          <dbl>
## 1 1503960366 4/12/2016      13162          8.5            8.5
## 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
## # i 10 more variables: LoggedActivitiesDistance <dbl>,
## #   VeryActiveDistance <dbl>, ModeratelyActiveDistance <dbl>,
## #   LightActiveDistance <dbl>, SedentaryActiveDistance <dbl>,
## #   VeryActiveMinutes <dbl>, FairlyActiveMinutes <dbl>,
## #   LightlyActiveMinutes <dbl>, SedentaryMinutes <dbl>, Calories <dbl>
```

```
str(daily_activity)
```

```
## spc_tbl_ [940 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Id : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ ActivityDate : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
## $ TotalSteps : num [1:940] 13162 10735 10460 9762 12669 ...
## $ TotalDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
## $ TrackerDistance : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
## $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveDistance : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
## $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
## $ LightActiveDistance : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
## $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
## $ VeryActiveMinutes : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
## $ FairlyActiveMinutes : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
## $ LightlyActiveMinutes : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
## $ SedentaryMinutes : num [1:940] 728 776 1218 726 773 ...
## $ Calories : num [1:940] 1985 1797 1776 1745 1863 ...
```

```
## - attr(*, "spec")=
## .. cols(
## ..   Id = col_double(),
## ..   ActivityDate = col_character(),
## ..   TotalSteps = col_double(),
## ..   TotalDistance = col_double(),
## ..   TrackerDistance = col_double(),
## ..   LoggedActivitiesDistance = col_double(),
## ..   VeryActiveDistance = col_double(),
## ..   ModeratelyActiveDistance = col_double(),
## ..   LightActiveDistance = col_double(),
## ..   SedentaryActiveDistance = col_double(),
## ..   VeryActiveMinutes = col_double(),
## ..   FairlyActiveMinutes = col_double(),
## ..   LightlyActiveMinutes = col_double(),
## ..   SedentaryMinutes = col_double(),
## ..   Calories = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
head(sleep_day_merged)
```

```
## # A tibble: 6 x 5
##       Id SleepDay      TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##       <dbl> <chr>                <dbl>                <dbl>          <dbl>
## 1 1503960366 4/12/2016 12:0~           1                 327            346
## 2 1503960366 4/13/2016 12:0~           2                 384            407
## 3 1503960366 4/15/2016 12:0~           1                 412            442
## 4 1503960366 4/16/2016 12:0~           2                 340            367
## 5 1503960366 4/17/2016 12:0~           1                 700            712
## 6 1503960366 4/19/2016 12:0~           1                 304            320
```

```
str(sleep_day_merged)
```

```
## spc_tbl_ [413 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Id : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
## $ SleepDay : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:00 AM" ...
## $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 ...
## $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
## $ TotalTimeInBed : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
## - attr(*, "spec")=
## .. cols(
## ..   Id = col_double(),
## ..   SleepDay = col_character(),
## ..   TotalSleepRecords = col_double(),
## ..   TotalMinutesAsleep = col_double(),
## ..   TotalTimeInBed = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

- Se uso la función summary() para buscar errores de falta de datos en las columnas numéricas.

```
summary(daily_activity)
```

```
##           Id           ActivityDate       TotalSteps   TotalDistance
## Min.      :1.504e+09   Length:940       Min.       :    0   Min.       : 0.000
## 1st Qu.:2.320e+09   Class :character   1st Qu.: 3790   1st Qu.: 2.620
## Median :4.445e+09   Mode  :character   Median : 7406   Median : 5.245
## Mean      :4.855e+09                Mean      : 7638   Mean      : 5.490
## 3rd Qu.:6.962e+09                3rd Qu.:10727   3rd Qu.: 7.713
## Max.      :8.878e+09                Max.      :36019   Max.      :28.030
## TrackerDistance   LoggedActivitiesDistance   VeryActiveDistance
## Min.       : 0.000   Min.       :0.0000   Min.       : 0.000
## 1st Qu.: 2.620   1st Qu.:0.0000   1st Qu.: 0.000
## Median : 5.245   Median :0.0000   Median : 0.210
## Mean      : 5.475   Mean      :0.1082   Mean      : 1.503
## 3rd Qu.: 7.710   3rd Qu.:0.0000   3rd Qu.: 2.053
## Max.      :28.030   Max.      :4.9421   Max.      :21.920
## ModeratelyActiveDistance   LightActiveDistance   SedentaryActiveDistance
## Min.       :0.0000   Min.       : 0.000   Min.       :0.000000
## 1st Qu.:0.0000   1st Qu.: 1.945   1st Qu.:0.000000
## Median :0.2400   Median : 3.365   Median :0.000000
## Mean      :0.5675   Mean      : 3.341   Mean      :0.001606
## 3rd Qu.:0.8000   3rd Qu.: 4.782   3rd Qu.:0.000000
## Max.      :6.4800   Max.      :10.710   Max.      :0.110000
## VeryActiveMinutes   FairlyActiveMinutes   LightlyActiveMinutes   SedentaryMinutes
## Min.       : 0.00   Min.       : 0.00   Min.       : 0.0   Min.       : 0.0
## 1st Qu.: 0.00   1st Qu.: 0.00   1st Qu.:127.0   1st Qu.: 729.8
## Median : 4.00   Median : 6.00   Median :199.0   Median :1057.5
## Mean      :21.16   Mean      :13.56   Mean      :192.8   Mean      : 991.2
## 3rd Qu.:32.00   3rd Qu.:19.00   3rd Qu.:264.0   3rd Qu.:1229.5
## Max.      :210.00   Max.      :143.00   Max.      :518.0   Max.      :1440.0
##           Calories
## Min.       : 0
## 1st Qu.:1828
## Median :2134
## Mean      :2304
## 3rd Qu.:2793
## Max.      :4900
```

```
summary(sleep_day_merged)
```

```
##           Id           SleepDay       TotalSleepRecords   TotalMinutesAsleep
## Min.      :1.504e+09   Length:413       Min.       :1.000   Min.       : 58.0
## 1st Qu.:3.977e+09   Class :character   1st Qu.:1.000   1st Qu.:361.0
## Median :4.703e+09   Mode  :character   Median :1.000   Median :433.0
## Mean      :5.001e+09                Mean      :1.119   Mean      :419.5
## 3rd Qu.:6.962e+09                3rd Qu.:1.000   3rd Qu.:490.0
## Max.      :8.792e+09                Max.      :3.000   Max.      :796.0
## TotalTimeInBed
## Min.       : 61.0
## 1st Qu.:403.0
## Median :463.0
## Mean      :458.6
```

```
## 3rd Qu.:526.0
## Max.    :961.0
```

- Se cambio los nombres de las columnas a un formato más estandarizado respetando las buenas prácticas generales en R.

```
daily_activity <- clean_names(daily_activity)
sleep_day_merged <- clean_names(sleep_day_merged)
```

- Se utilizó la función `mdy()` para cambiar el tipo de datos la columna `activity_date` del dataframe `daily_activity` de “Character” a “Date”.
- Se utilizó la función `mdy_hms()` para cambiar la columna `sleep_date` de “Character” a “Date/Time” a “Date”.

```
daily_activity$activity_date <- mdy(daily_activity$activity_date)
sleep_day_merged$sleep_day <- mdy_hms(sleep_day_merged$sleep_day)
sleep_day_merged$sleep_day <- as.Date(sleep_day_merged$sleep_day)
```

- Se utilizó la función `get_dupes()` para buscar duplicados con los “id” y “activity\_date” ya que se tiene que tener un registro por persona (id) en cada fecha (activity\_date).
- De igual forma, se ocupó la función `get_dupes()` para buscar duplicados en el data frame `sleep_day_merged`, utilizando las columnas `id` y `sleep_day`.

```
get_dupes(daily_activity, id, activity_date)
```

```
## No duplicate combinations found of: id, activity_date
```

```
## # A tibble: 0 x 16
## # i 16 variables: id <dbl>, activity_date <date>, dupe_count <int>,
## #   total_steps <dbl>, total_distance <dbl>, tracker_distance <dbl>,
## #   logged_activities_distance <dbl>, very_active_distance <dbl>,
## #   moderately_active_distance <dbl>, light_active_distance <dbl>,
## #   sedentary_active_distance <dbl>, very_active_minutes <dbl>,
## #   fairly_active_minutes <dbl>, lightly_active_minutes <dbl>,
## #   sedentary_minutes <dbl>, calories <dbl>
```

```
get_dupes(sleep_day_merged, id, sleep_day)
```

```
## # A tibble: 6 x 6
##       id sleep_day dupe_count total_sleep_records total_minutes_asleep
##   <dbl> <date>      <int>          <dbl>              <dbl>
## 1 4388161847 2016-05-05         2             1                471
## 2 4388161847 2016-05-05         2             1                471
## 3 4702921684 2016-05-07         2             1                520
## 4 4702921684 2016-05-07         2             1                520
## 5 8378563200 2016-04-25         2             1                388
## 6 8378563200 2016-04-25         2             1                388
## # i 1 more variable: total_time_in_bed <dbl>
```

- Se eliminaron los registros duplicados encontrados con la función `distinct()`.

```
sleep_day_merged <- sleep_day_merged %>%
  distinct()
```

- Finalmente, se realizó otra exploración de datos para confirmar que se haya realizado todo correctamente y no haya algún paso más por realizar.

#### Fase 4: Analyze

**Análisis de los datos de la actividad diaria** Los datos que se analizaron de la actividad diaria de las personas son: las calorías diarias quemadas, los minutos que pasan sedentarios, los minutos en una actividad física ligera, moderada e intensa (daily\_activity).

**Actividad física intensa** Primero se hizo una correlación entre los minutos que las personas pasaban en una actividad física intensa y las calorías gastadas.

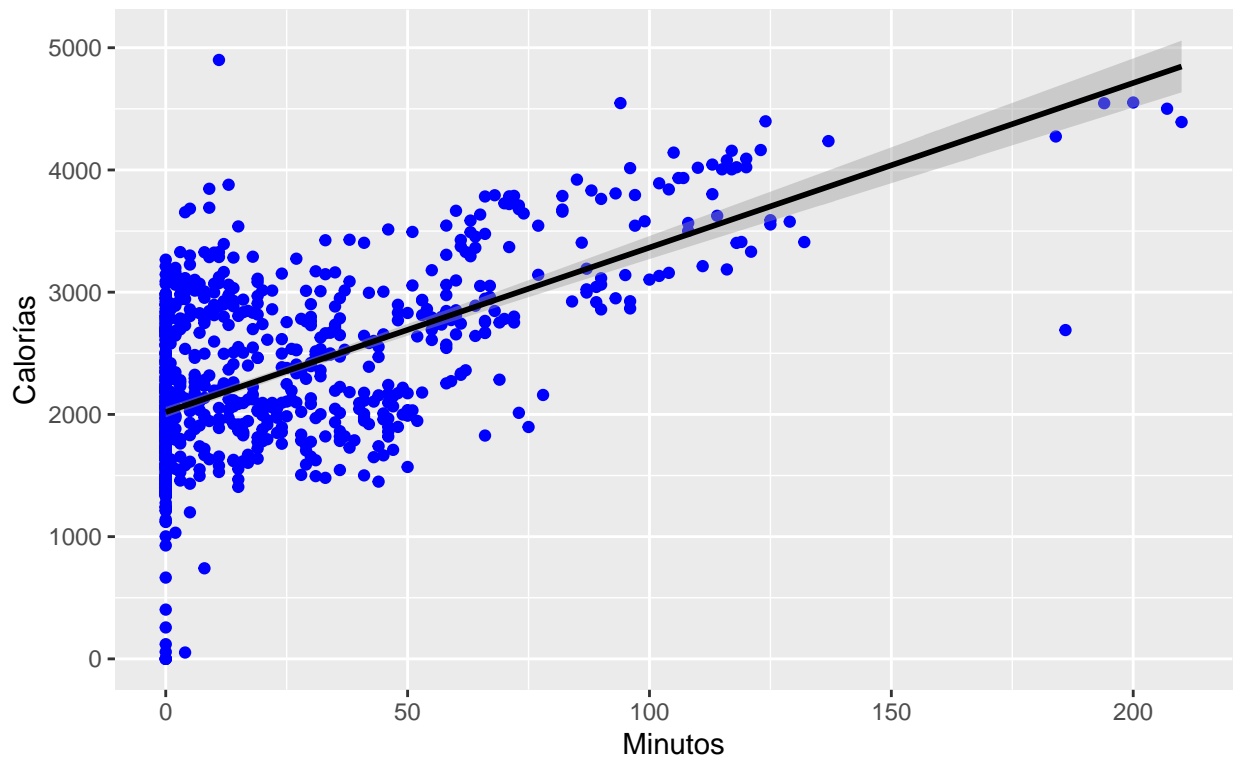
De estos datos podemos saber qué:

- La media de la actividad física intensa es de 21.16 minutos diarios.
- El valor mínimo es de 0 y el máximo es de 210 minutos.
- El 50% de los datos se encuentran entre 0 y 32 minutos.
- La mediana de los datos es de 4 minutos.
- La correlación entre los minutos de actividad física intensa y las calorías es positiva, con un valor de .615.

```
ggplot(data = daily_activity)+
  geom_point(mapping = aes(x=very_active_minutes, y=calories), color="blue")+
  geom_smooth(mapping = aes(x=very_active_minutes, y=calories), method=lm, color="black")+
  labs(title="Relación de calorías en función de los minutos en una actividad física intensa", x="Minutos", y="Calorías")

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

## Relación de calorías en función de los minutos en una actividad física inte



La línea negra representa correlación entre las variables, que es igual a 0.615838268270338

```
cor(x=daily_activity$very_active_minutes, y=daily_activity$calories)
```

```
## [1] 0.6158383
```

```
summary(daily_activity$very_active_minutes)
```

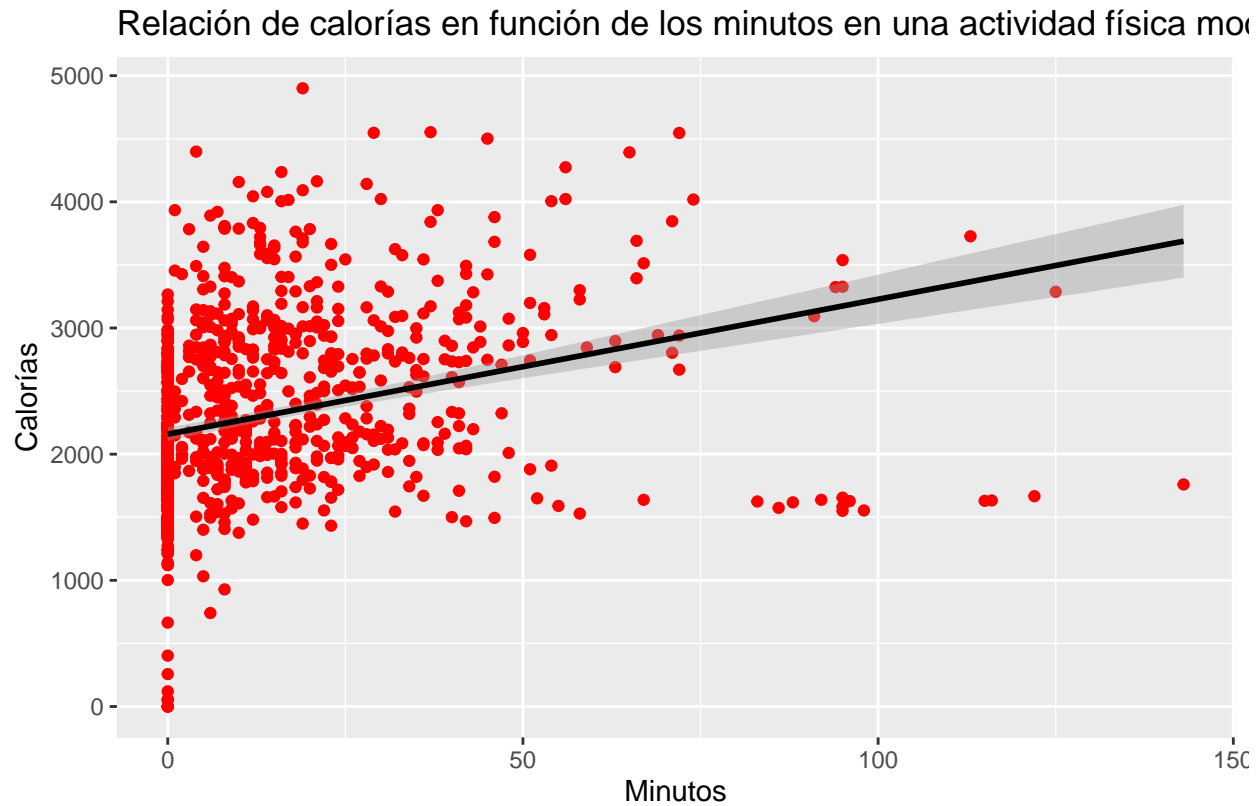
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   0.00    4.00   21.16   32.00   210.00
```

**Actividad física moderada** Ahora se analizan los datos de la actividad física moderada. Sobre estos datos podemos decir que:

- La media de la actividad física moderada es de 13.56 minutos diarios.
- El valor mínimo es de 0 y el valor máximo es de 143 minutos diarios.
- El 50% de los datos se encuentran entre 0 y 19 minutos.
- La mediana de los datos es de 6 minutos.
- La correlación entre los minutos de actividad física moderada y las calorías es positiva, con un valor de .297.

```
ggplot(data = daily_activity)+
  geom_point(mapping = aes(x=fairly_active_minutes, y=calories), color="red")+
  geom_smooth(mapping = aes(x=fairly_active_minutes, y=calories),method=lm, color="black")+
  labs(title="Relación de calorías en función de los minutos en una actividad física moderada", x="Minu
```

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



```
cor(x=daily_activity$fairly_active_minutes, y=daily_activity$calories)
```

```
## [1] 0.2976235
```

```
summary(daily_activity$fairly_active_minutes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   0.00    6.00   13.56   19.00   143.00
```

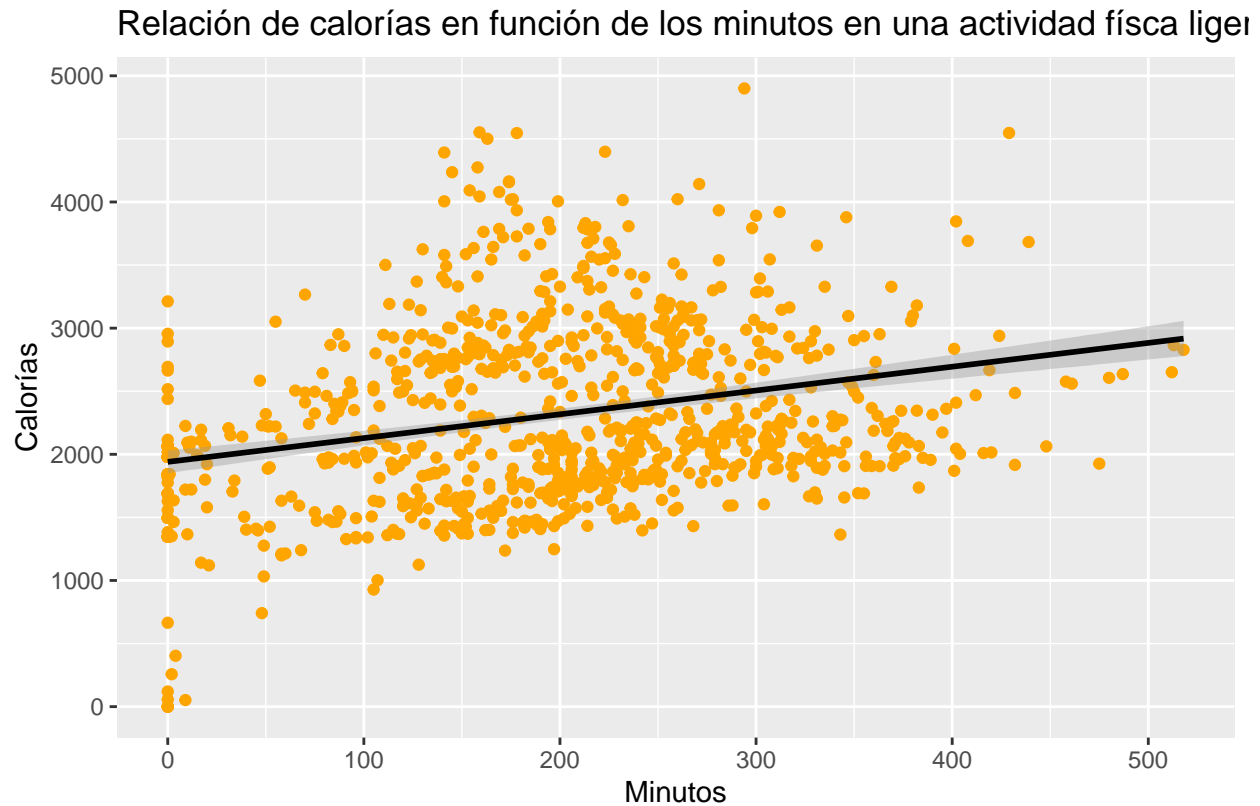
**Actividad física ligera** Por su parte, la descripción sobre los datos de la actividad física ligera es la siguiente:

- La media de la actividad física ligera es de 192.8 minutos diarios.
- El valor mínimo es de 0 y el valor máximo es de 518 minutos diarios.
- El 50% de los datos se encuentran entre los 127 y 264 minutos diarios.
- La mediana de los datos es de 199 minutos.
- La correlación entre los minutos de actividad física ligera y las calorías es positiva, con un valor de .286.



```
ggplot(data = daily_activity)+
  geom_point(mapping = aes(x=lightly_active_minutes, y=calories), color="orange")+
  geom_smooth(mapping = aes(x=lightly_active_minutes, y=calories),method=lm, color="black")+
  labs(title="Relación de calorías en función de los minutos en una actividad física ligera", x="Minutos")

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



```
cor(x=daily_activity$lightly_active_minutes, y=daily_activity$calories)
```

```
## [1] 0.2867175
```

```
summary(daily_activity$lightly_active_minutes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   127.0   199.0   192.8   264.0   518.0
```

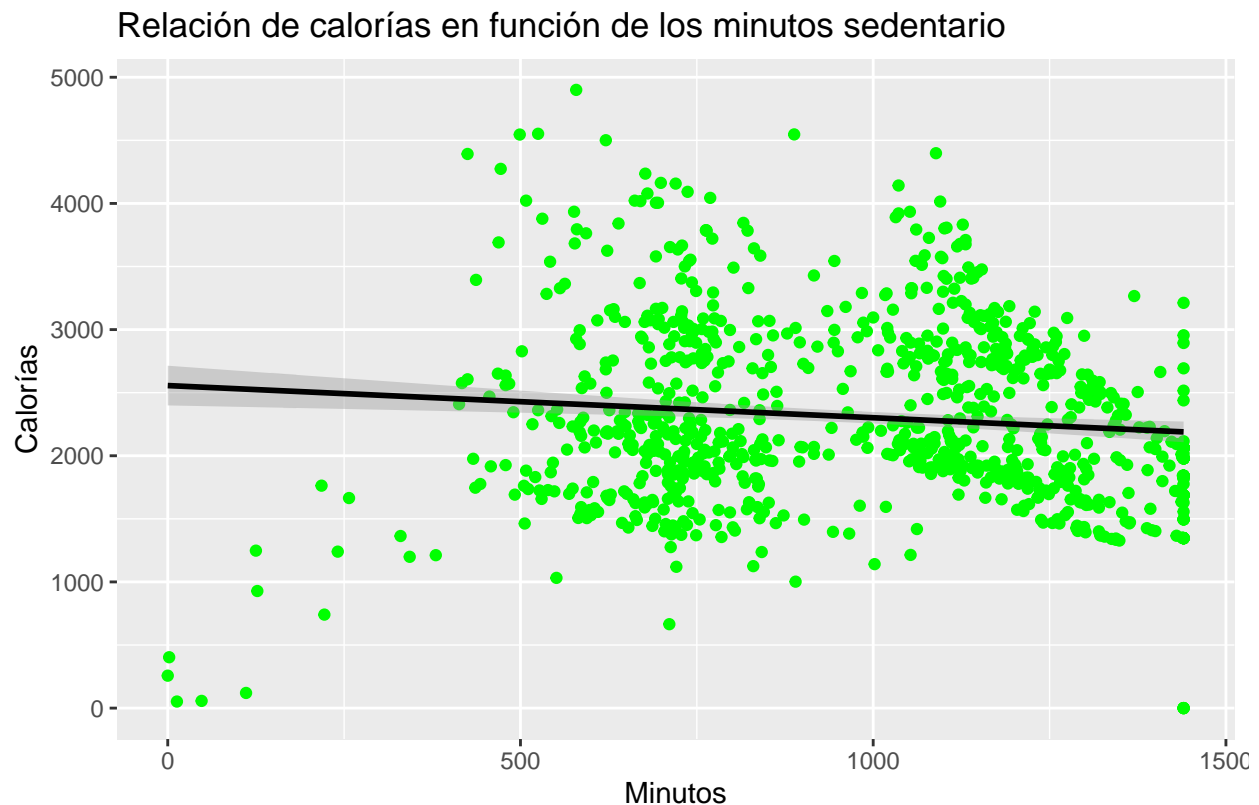
Finalmente, se describen los datos acerca de los minutos diarios de actividad sedentaria:

- La media de la actividad sedentaria es de 1057.5 minutos diarios.
- El valor mínimo es de 0 y el valor máximo es de 1440 minutos diarios.
- El 50% de los datos se encuentran entre los 729.8 y los 1229.5 minutos diarios.
- La mediana de los datos es de 1057.5 minutos.
- La correlación entre los minutos de actividad física sedentaria y las calorías es negativa, con un valor de -.106.

```
ggplot(data = daily_activity)+
  geom_point(mapping = aes(x=sedentary_minutes, y=calories), color="green")+
  geom_smooth(mapping = aes(x=sedentary_minutes, y=calories),method=lm, color="black")+
  labs(title="Relación de calorías en función de los minutos sedentario", x="Minutos",y="Calorías", cap
```

## Actividad física sedentaria

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



```
cor(x=daily_activity$sedentary_minutes, y=daily_activity$calories)
```

```
## [1] -0.106973
```

```
summary(daily_activity$sedentary_minutes)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   729.8   1057.5   991.2  1229.5   1440.0
```

Esos fueron los resultados de los datos analizados de la base de actividad diaria.

**Análisis de los datos del sueño** El análisis a los datos referentes al sueño (sleep\_day\_merged) se realizó de la siguiente forma:

**Horas de sueño** Se creó una columna con las horas totales de sueño, a partir de los minutos totales de sueño dividiéndolos entre 60. De esta forma se puede analizar los datos de una forma más práctica.

```
sleep_day_merged <- mutate(sleep_day_merged, total_hours_asleep=round(total_minutes_asleep/60, 2), nsma)

summary(sleep_day_merged$total_hours_asleep)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.970   6.020   7.210   6.987   8.170  13.270
```

De estos datos se puede obtener lo siguiente:

- La media de horas de sueño en la muestra es de 6.987 hrs.
- El 50% de los registros se encuentra entre las ~6 y las ~8 horas.
- El mínimo de horas fue menor a una hora.
- Y el máximo de horas fue de poco más de 13 horas.

**Minutos en conciliar el sueño** Por otro lado, también se creó una nueva columna llamada “time\_to\_fall\_asleep”, restando del tiempo total en cama, el tiempo total que durmieron, teniendo como sobrante, el tiempo que tardaron en quedarse dormidos.

```
sleep_day_merged <- mutate(sleep_day_merged, time_to_fall_asleep=total_time_in_bed - total_minutes_asleep)

summary(sleep_day_merged$time_to_fall_asleep, sleep_day_merged$total_minutes_asleep)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   17.00   25.50   39.31   40.00   371.00
```

De estos datos, se puede decir que:

- El promedio del tiempo para quedarse dormido es de 39.31 minutos.
- El mínimo es de 0 minutos y el máximo es de 371 minutos.
- El 50% de los datos se encuentran entre 17 y 40 minutos.
- La mediana de los datos es de 25.50 minutos.

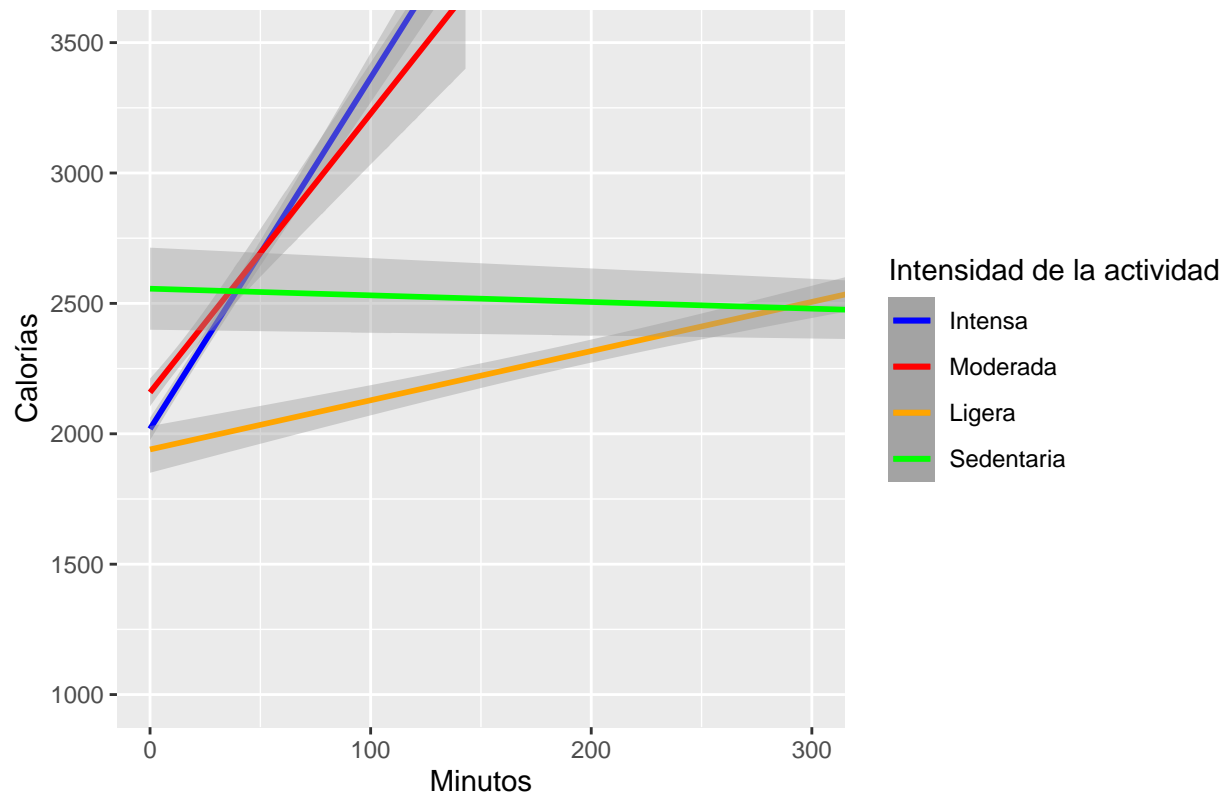
## Fase 5: Share

**Datos de la actividad diaria** De los resultados del análisis de los datos de las actividades diarias podemos concluir que:

```
ggplot(data = daily_activity)+
  geom_smooth(mapping = aes(x=very_active_minutes, y=calories, color="Intensa"),method=lm)+
  geom_smooth(mapping = aes(x=fairly_active_minutes, y=calories, color="Moderada"),method=lm)+
  geom_smooth(mapping = aes(x=lightly_active_minutes, y=calories, color="Ligera"),method=lm)+
  geom_smooth(mapping = aes(x=sedentary_minutes, y=calories, color="Sedentaria"),method=lm)+
  labs(title="Correlación de calorías en función de los minutos en distintas actividades físicas", x="M")
  coord_cartesian(xlim = c(0, 300),ylim = c(1000, 3500))+
  scale_color_manual(name="Intensidad de la actividad",
                     breaks=c("Intensa","Moderada","Ligera","Sedentaria"),
                     values=c("Intensa"="blue","Moderada"="red","Ligera"="orange","Sedentaria"="green"))
```

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

### Correlación de calorías en función de los minutos en distintas actividades



**Tendencias sobre las diferentes actividades físicas encontradas:** La correlación con la actividad intensa es la más fuerte, teniendo un valor de .61 podemos hablar de una relación fuerte entre las calorías gastadas y los minutos que las personas realizan actividades con esta intensidad. También podemos observar, que entre menor es la intensidad del ejercicio, menor es la correlación tiene con las calorías gastadas. Siguiendo con la actividad moderada y su correlación de .29, teniendo la relación, pero está siendo débil. Finalmente, la actividad ligera es la que menor correlación tiene, siendo su valor de .28.

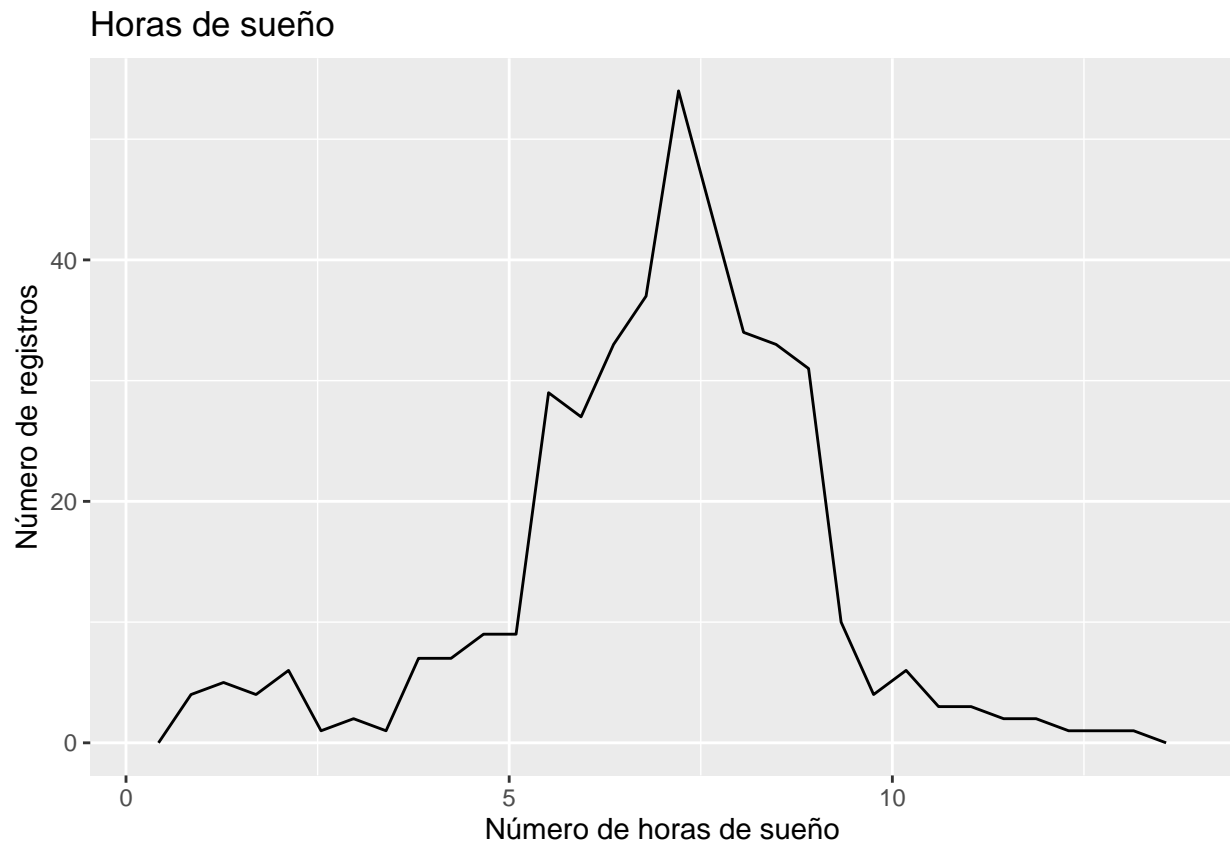
Esto lo podemos interpretar como que una parte importante de las calorías gastadas, es que dentro de la rutina incluyas actividad física intensa, no habría que malinterpretar conque mientras más actividad física intensa hagas más calorías en total gastaras, ya que las calorías totales gastadas en un día también incluyen los otros dos tipos de actividades físicas, y si realizamos actividad intensa y omitimos las otras dos, podría ser contraproducente para nuestros objetivos, sin embargo, para afirmar esto, se tendría que contar con más datos y enfocar el estudio hacia esta relación. Finalmente, que entre más intensa sea la actividad, se tiende a que se realice menos número de veces, esto lo indica la mediana, que va disminuyendo conforme la actividad se vuelve más intensa.

Por su parte, la actividad sedentaria tiene su propia correlación, siendo negativa con un valor de .10, nos indica una relación inversamente proporcional débil, esto quiere decir que entre más actividad física sedentaria realices menos calorías gastarás y viceversa, lo cual hace sentido con la realidad. Sumando las dos partes, podríamos concluir que entre más intenso sea la actividad total del día, más calorías se gastarán, no solamente importa que no tengas momentos sedentarios, si no que realmente hagas actividades físicas enfocadas en ser más intensas y por lo tanto, gastar más calorías.

**Datos del sueño** De los datos del sueño, se puede decir lo siguiente sobre las horas del sueño:

```
ggplot(data=sleep_day_merged) +  
  geom_freqpoly(mapping = aes(x=total_hours_asleep))+  
  labs(title = "Horas de sueño", x="Número de horas de sueño",y="Número de registros")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



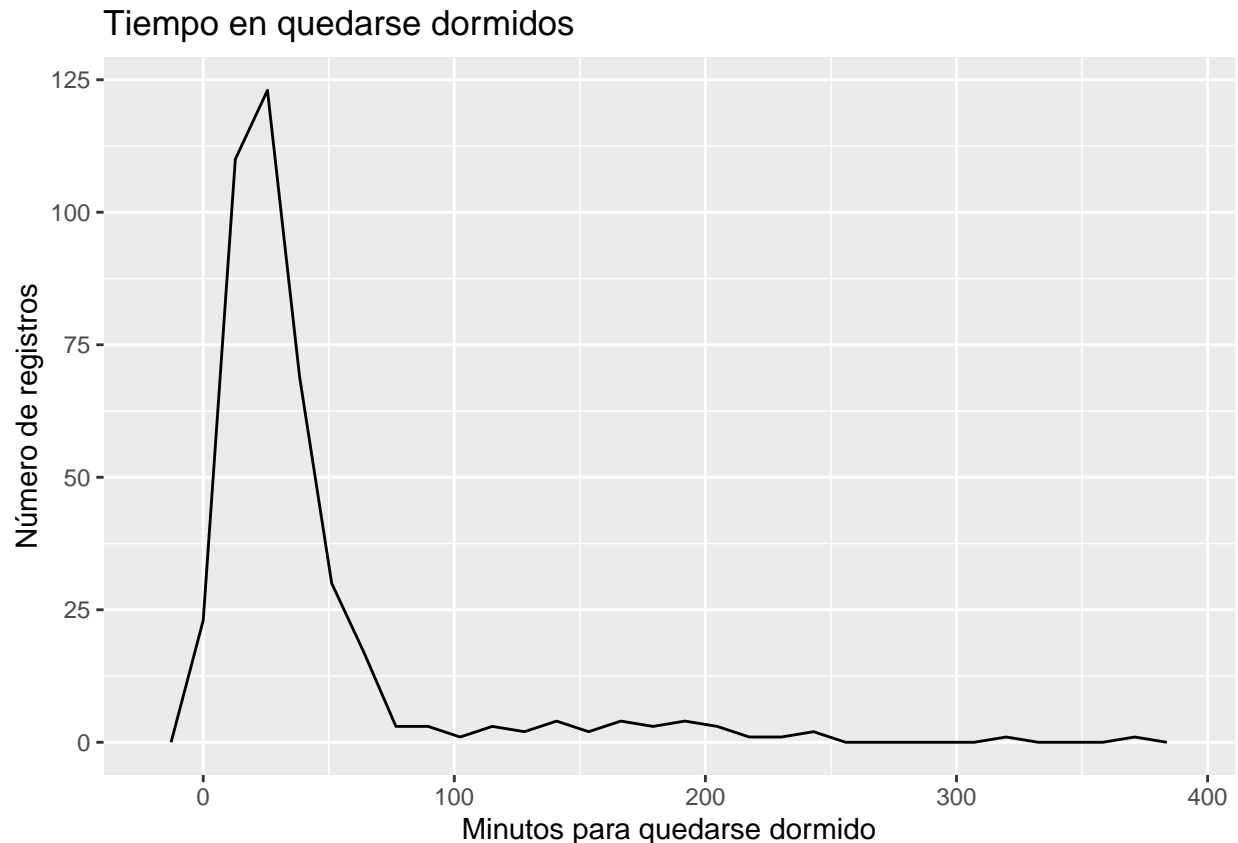
```
summary(sleep_day_merged$total_hours_asleep)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##    0.970  6.020   7.210   6.987  8.170  13.270
```

De los datos de la muestra la media de horas de sueño en la muestra es de 6.987. Si bien, cada persona tiene necesidad fisiológicas distintas, se conoce actualmente que los ciclos del sueño se componen por hora y media, esto quiere decir, que las horas en las cuales se cumplen los ciclos serían a la 6ta y a la 7ma y media hora, por lo cual, la mayoría de persona despierta sin cumplir sus ciclos del sueño.

Por su parte, los datos del tiempo en que las personas tardan en quedar dormidos se pueden interpretar de la siguiente forma:

```
ggplot(data=sleep_day_merged) +  
  geom_freqpoly(mapping = aes(x=time_to_fall_asleep))+  
  labs(title = "Tiempo en quedarse dormidos", x="Minutos para quedarse dormido",y="Número de registros")
```



```
summary(sleep_day_merged$time_to_fall_asleep)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00  17.00   25.50   39.31  40.00   371.00
```

La media de tiempo que las personas de la muestra tardan en dormir es casi de 40 minutos, lo que deja en claro, que hay una dificultad para conciliar el sueño. Además, que la mayoría de personas tomen más de 15 minutos para dormir, aunado a que los datos más alejados de la media sean muy por encima de esta, nos dice que este problema para conciliar el sueño es un problema generalizado.

**Tendencias generales del sueño encontradas:** La población muestra problemas en la conciliación del sueño y en el tiempo que pueden dormir. Esto puede deberse a una mala higiene del sueño provocada por múltiples factores, entre ellos, el trabajo, el estrés, los malos hábitos en los horarios al dormir, etc. Y pueden tener relación entre sí, ya que el tiempo que se pierde en conciliar el sueño promedio es casi la mitad de un ciclo del sueño, por lo que si se redujera el tiempo necesario para quedarse dormido, podría ocuparse en dormir.

## Fase 6: Act

Las 4 tendencias encontradas en este análisis fueron:

- Se encontró la tendencia que entre más tiempo realicen las personas actividad física intensa, más calorías totales gastarán en un día

- Sí bien, hay una relación inversa entre las calorías gastadas por las personas y el tiempo que pasan siendo sedentarios, es una relación muy débil, por lo que importa más lo que sí hacen en actividades físicas intensas y no lo que pasan siendo sedentario.
- Existe una tendencia a no cumplir con los ciclos de sueño, ya que las personas duermen menos de 7 horas y media, pero más de 6 horas.
- Finalmente, hay un problema generalizado para conciliar el sueño, ya que las personas pueden tardar ~40 minutos en poder conciliar el sueño.

**Las mejoras enfocadas en los usuarios encontradas en el análisis son:**

- Mejorar la detección de tiempo estando en la cama, ya el tiempo, según algunos registros, es de 0 minutos, lo cual puede pasar, pero es poco probable.
- Incentivar a los usuarios a realizar más tiempo actividades físicas intensas, ya que son las más relacionadas a las calorías gastadas pero las menos realizadas en número de veces, bien puede ser a través de distintos tipos de recompensas, por ejemplo, notificaciones felicitándolos por haber llegado a “X” tiempo de actividad intensa o con un sonido de fiesta al llegar.
- Mostrarle al usuario la importancia de mantener sus datos actualizados, ya que se pensaban ocupar los datos de peso de la muestra, pero dado que solo 8 usuarios los tienen y a que en varios casos son como mucho un par de registros, se descartó utilizarlos.

**Las recomendaciones del enfoque de la campaña de marketing generados por este análisis son:**

- Mostrar casos en los cuáles, los usuarios, gracias a mantener un tracking y cumplir con ciertos objetivos, pudieron llegar a ciertas metas, ya sea de peso, pasos dados, distancia recorrida y esto cómo los hizo sentir.
- Informar sobre cómo tener una buena higiene del sueño ayuda a mejorar la salud y en la vida diaria en general, y mejorar los hábitos poco a poco mediante el seguimiento de ciertas métricas.