Analysis

2023-06-02

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.1     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.1     ✓ tibble    3.2.1
## ✓ lubridate 1.9.2     ✓ tidyr     1.3.0
## ✓ purrr     1.0.1
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflict
s to become errors
```

```
library(skimr)
library(janitor)
```

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

```
library(here)
```

```
## here() starts at G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de datos/Caso p
ractico/R
```

```
library(chron)
```

```
##
## Attaching package: 'chron'
##
## The following objects are masked from 'package:lubridate':
##
##     days, hours, minutes, seconds, years
```

```
daily_activity <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de dato
s/Caso practico/R/Boards/daily_activity_csv.csv")
```

```
sleep_day <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de datos/Cas
o practico/R/Boards/sleep_day_csv.csv")
```

Analysis

Frequency of use

The first step was to create a frequency table of device usage. This table considers the days on which each ID recorded information.

```
sleep_day_frequency <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de datos/Caso practico/R/Boards/daily_activity_frequency_csv.csv")
```

```
daily_activity_frequency <- data.frame(daily_activity %>%
                                         group_by(Id) %>%
                                         summarize(min_date = min(ActivityDate),
                                                   max_date = max(ActivityDate)),
                                       table(daily_activity$Id))
```

```
daily_activity_frequency <- daily_activity_frequency %>%
  select(-Var1)
```

```
daily_activity_frequency
```

```
##            Id   min_date   max_date Freq
## 1  1503960366 2016-04-12 2016-05-12   31
## 2  1624580081 2016-04-12 2016-05-12   31
## 3  1644430081 2016-04-12 2016-05-11   30
## 4  1844505072 2016-04-12 2016-05-12   31
## 5  1927972279 2016-04-12 2016-05-12   31
## 6  2022484408 2016-04-12 2016-05-12   31
## 7  2026352035 2016-04-12 2016-05-12   31
## 8  2320127002 2016-04-12 2016-05-12   31
## 9  2347167796 2016-04-12 2016-04-29   18
## 10 2873212765 2016-04-12 2016-05-12   31
## 11 3372868164 2016-04-12 2016-05-01   20
## 12 3977333714 2016-04-12 2016-05-11   30
## 13 4020332650 2016-04-12 2016-05-12   31
## 14 4057192912 2016-04-12 2016-04-15    4
## 15 4319703577 2016-04-12 2016-05-12   31
## 16 4388161847 2016-04-12 2016-05-12   31
## 17 4445114986 2016-04-12 2016-05-12   31
## 18 4558609924 2016-04-12 2016-05-12   31
## 19 4702921684 2016-04-12 2016-05-12   31
## 20 5553957443 2016-04-12 2016-05-12   31
## 21 5577150313 2016-04-12 2016-05-11   30
## 22 6117666160 2016-04-12 2016-05-09   28
## 23 6290855005 2016-04-12 2016-05-10   29
## 24 6775888955 2016-04-12 2016-05-07   26
## 25 6962181067 2016-04-12 2016-05-12   31
## 26 7007744171 2016-04-12 2016-05-07   26
## 27 7086361926 2016-04-12 2016-05-12   31
## 28 8053475328 2016-04-12 2016-05-12   31
## 29 8253242879 2016-04-12 2016-04-30   19
## 30 8378563200 2016-04-12 2016-05-12   31
## 31 8583815059 2016-04-12 2016-05-12   31
## 32 8792009665 2016-04-12 2016-05-10   29
## 33 8877689391 2016-04-12 2016-05-12   31
```

Frequency of use during sleep

We proceeded to create an identical table, incorporating the information we have about users' sleep.

```
sleep_day_frequency <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de datos/Caso practico/R/Boards/sleep_day_frequency_csv.csv")
```

```
str(sleep_day_frequency)
```

```
## 'data.frame':    11 obs. of  11 variables:
##  $ TotalSteps              : num  1 0.985 0.74 0.507 0.692 ...
##  $ TotalDistance           : num  0.985 1 0.795 0.471 0.662 ...
##  $ VeryActiveDistance      : num  0.74 0.795 1 0.193 0.158 ...
##  $ ModeratelyActiveDistance: num  0.507 0.471 0.193 1 0.238 ...
##  $ LightActiveDistance     : num  0.692 0.662 0.158 0.238 1 ...
##  $ SedentaryActiveDistance : num  0.0705 0.08239 0.04612 0.00579 0.0995 ...
##  $ VeryActiveMinutes       : num  0.667 0.681 0.827 0.225 0.155 ...
##  $ ModeratelyActiveMinutes : num  0.499 0.463 0.212 0.947 0.22 ...
##  $ LightlyActiveMinutes    : num  0.5696 0.5163 0.0598 0.1621 0.8857 ...
##  $ SedentaryMinutes        : num  -0.3275 -0.2881 -0.0618 -0.2214 -0.4136 ...
##  $ TotalActiviteMinutes    : num  -0.01728 0.00452 0.07262 -0.0853 -0.06921 ...
```

Avg. days of use

```
mean(daily_activity_frequency[ ,4])
```

```
## [1] 28.48485
```

```
mean(daily_activity_frequency[ ,4])/31
```

```
## [1] 0.9188661
```

On average, users use the device on 92% of the days during the study period, which is equivalent to 28 out of 31 days.

Avg. days of device usage during sleep

```
mean(sleep_day_frequency[ ,4])
```

```
## [1] 0.3129316
```

```
mean(sleep_day_frequency[ ,4])/31
```

```
## [1] 0.01009457
```

On average, the device is used during sleep for only 17 out of 31 days, which is approximately 55%.

Usage of the device by minutes of activity

Create a data frame that contains only the columns of ID, the different activity times according to their intensity, and a last column with the total sum of minutes of device usage.

```
activity_minutes <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de da
tos/Caso practico/R/Boards/activity_minutes_csv.csv")
```

```
head(activity_minutes)
```

```
##           Id VeryActiveMinutes ModeratelyActiveMinutes LightlyActiveMinutes
## 1 1503960366                25                      13                  328
## 2 1503960366                21                      19                  217
## 3 1503960366                30                      11                  181
## 4 1503960366                29                      34                  209
## 5 1503960366                36                      10                  221
## 6 1503960366                38                      20                  164
##   SedentaryMinutes TotalActiveMinutes
## 1              728               1094
## 2              776               1033
## 3             1218               1440
## 4              726                998
## 5              773               1040
## 6              539                761
```

Averages of device usage in minutes are calculated based on the intensity of the activity.

```
activity_minutes_avg <- data.frame(VeryActiveMinutes = round(mean(activity_minutes$VeryActive
Minutes)),
           ModeratelyActiveMinutes = round(mean(activity_minutes$ModeratelyActiveMinutes)),
           LightlyActiveMinutes = round(mean(activity_minutes$LightlyActiveMinutes)),
           SedentaryMinutes = round(mean(activity_minutes$SedentaryMinutes))
           )
```

```
activity_minutes_avg
```

```
##   VeryActiveMinutes ModeratelyActiveMinutes LightlyActiveMinutes
## 1                21                      14                  193
##   SedentaryMinutes
## 1              991
```

```
activity_minutes_avg_percent <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/An
alisis de datos/Caso practico/R/Boards/activity_minutes_avg_percent_csv.csv")
```

```
activity_minutes_avg_percent
```

```
##         ActivityIntensity Average
## 1       VeryActiveMinutes       2
## 2 ModeratelyActiveMinutes       1
## 3    LightlyActiveMinutes      16
## 4        SedentaryMinutes      81
```

Coincidencias

Activity Data and Sleep Data

To analyze the activity data in conjunction with the sleep data, a new data frame is created. It contains the activity data only for users who also recorded sleep activity.

Merge the columns of interest from each data frame.

```
coincidences <- merge(sleep_day ,
      daily_activity %>%
        select(-VeryActiveDistance,
               -ModeratelyActiveDistance,
               -LightActiveDistance,
               -SedentaryActiveDistance),
      by.x = c("Id", "ActivityDate"), by.y = c("Id","ActivityDate"))
```

An additional column is added to the data frame with the total sum of activity minutes.

```
coincidences <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de datos/
Caso practico/R/Boards/coincidence_csv.csv")
```

```
coincidences <- coincidences %>%
  mutate(TotalActiveMinutes =
           VeryActiveMinutes +
           ModeratelyActiveMinutes +
           LightlyActiveMinutes +
           SedentaryMinutes)
```

```
head(coincidences)
```

```
##           Id       date TotalMinutesAsleep TotalTimeInBed TotalSteps
## 1 1503960366 2016-04-12                327            346      13162
## 2 1503960366 2016-04-13                384            407      10735
## 3 1503960366 2016-04-15                412            442       9762
## 4 1503960366 2016-04-16                340            367      12669
## 5 1503960366 2016-04-17                700            712       9705
## 6 1503960366 2016-04-19                304            320      15506
##   TotalDistance VeryActiveMinutes ModeratelyActiveMinutes LightlyActiveMinutes
## 1          8.50                25                      13                  328
## 2          6.97                21                      19                  217
## 3          6.28                29                      34                  209
## 4          8.16                36                      10                  221
## 5          6.48                38                      20                  164
## 6          9.88                50                      31                  264
##   SedentaryMinutes TotalActiviteMinutes TotalActiveMinutes
## 1              728                1094               1094
## 2              776                1033               1033
## 3              726                 998                998
## 4              773                1040               1040
## 5              539                 761                761
## 6              775                1120               1120
```

Finally, a correlation coefficient table is created to understand the relationship that exists between the variables.

```
correlation <- read.csv("G:/Otros ordenadores/Mi Portátil/Gaspar Facultad/Analisis de datos/C
aso practico/R/Boards/correlation_csv.csv")
```

```
correlation <- mutate_all(correlation, as.numeric)
```

```
correlation <- cor(correlation)
```

```
(correlation)
```

```
##                          TotalSteps TotalDistance VeryActiveDistance
## TotalSteps                1.0000000     0.9967033         0.74413006
## TotalDistance             0.9967033     1.0000000         0.79174537
## VeryActiveDistance        0.7441301     0.7917454         1.00000000
## ModeratelyActiveDistance  0.5446125     0.5072264         0.20674103
## LightActiveDistance       0.7567183     0.7190983         0.19233243
## SedentaryActiveDistance  -0.3099165    -0.3075096        -0.29084925
## VeryActiveMinutes         0.7638781     0.7999989         0.96629517
## ModeratelyActiveMinutes   0.5608720     0.5258406         0.25209164
## LightlyActiveMinutes      0.6587205     0.6151313         0.07119636
## SedentaryMinutes         -0.8450779    -0.8119061        -0.39749230
## TotalActiviteMinutes     -0.7797097    -0.7498637        -0.39884488
##                          ModeratelyActiveDistance LightActiveDistance
## TotalSteps                              0.5446125          0.75671829
## TotalDistance                           0.5072264          0.71909833
## VeryActiveDistance                      0.2067410          0.19233243
## ModeratelyActiveDistance                1.0000000          0.30915062
## LightActiveDistance                     0.3091506          1.00000000
## SedentaryActiveDistance                -0.3352948         -0.08992192
## VeryActiveMinutes                       0.3130197          0.21294197
## ModeratelyActiveMinutes                 0.9950952          0.29273730
## LightlyActiveMinutes                    0.2348969          0.98352886
## SedentaryMinutes                       -0.6129181         -0.81802948
## TotalActiviteMinutes                   -0.6569536         -0.68554175
##                          SedentaryActiveDistance VeryActiveMinutes
## TotalSteps                           -0.30991650        0.76387807
## TotalDistance                        -0.30750959        0.79999895
## VeryActiveDistance                   -0.29084925        0.96629517
## ModeratelyActiveDistance             -0.33529477        0.31301975
## LightActiveDistance                  -0.08992192        0.21294197
## SedentaryActiveDistance               1.00000000       -0.31799566
## VeryActiveMinutes                    -0.31799566        1.00000000
## ModeratelyActiveMinutes              -0.36237610        0.37044015
## LightlyActiveMinutes                 -0.01925509        0.09227875
## SedentaryMinutes                      0.02721373       -0.48707446
## TotalActiviteMinutes                 -0.03480542       -0.50410262
##                          ModeratelyActiveMinutes LightlyActiveMinutes
## TotalSteps                             0.5608720           0.65872047
## TotalDistance                          0.5258406           0.61513125
## VeryActiveDistance                     0.2520916           0.07119636
## ModeratelyActiveDistance               0.9950952           0.23489691
## LightActiveDistance                    0.2927373           0.98352886
## SedentaryActiveDistance               -0.3623761          -0.01925509
## VeryActiveMinutes                      0.3704402           0.09227875
## ModeratelyActiveMinutes                1.0000000           0.21444593
## LightlyActiveMinutes                   0.2144459           1.00000000
## SedentaryMinutes                      -0.6182230          -0.77610552
## TotalActiviteMinutes                  -0.6661547          -0.63908700
##                          SedentaryMinutes TotalActiviteMinutes
## TotalSteps                   -0.84507788          -0.77970974
## TotalDistance                -0.81190606          -0.74986371
## VeryActiveDistance           -0.39749230          -0.39884488
## ModeratelyActiveDistance     -0.61291811          -0.65695364
## LightActiveDistance          -0.81802948          -0.68554175
## SedentaryActiveDistance       0.02721373          -0.03480542
```

```
## VeryActiveMinutes            -0.48707446          -0.50410262
## ModeratelyActiveMinutes      -0.61822303          -0.66615475
## LightlyActiveMinutes         -0.77610552          -0.63908700
## SedentaryMinutes              1.00000000           0.97759805
## TotalActiviteMinutes          0.97759805           1.00000000
```