

# 1. Definição do problema.

---

Quais são algumas tendências no uso de dispositivos inteligentes?

A partir dos Dados da FitBit Fitness Tracker Data

## Salving PDF

---

```
# https://bookdown.org/yihui/rmarkdown/pdf-document.html
# Press F1 or Ctrl+Shift+P
# Type export and select below
rmarkdown::pandoc_available()
library(knitr)
pandoc("projectFinal/finalProject.Rmd", format = "pdf")
```

```
install.packages("tinytex")
tinytex::install_tinytex()
```

## Library may be used

---

```
install.packages("*")
```

```
library(janitor) # janitor has simple little tools for examining and cleaning
dirty data.
library(arrow)
library(tidyverse)
library(naniar)
library(ggsci)
library(skimr) # Skim a data frame, getting useful summary statistics
library(lubridate) # Lubridate provides tools that make it easier to parse and
manipulate dates.
library(ggpubr) # library and require load and attach add-on packages. stat_cor
```

## reading files csv

---

```
# sleep_day_file <- read_csv("projectFinal/fitDatabase/sleepDay_merged.csv")
# daily_activity_file <-
```

```
read_csv("projectFinal/fitDatabase/dailyActivity_merged.csv")
# daily_intensities_file <-
read_csv("projectFinal/fitDatabase/dailyIntensities_merged.csv")
# hourly_intensities_file <-
read_csv("projectFinal/fitDatabase/hourlyIntensities_merged.csv")
# hourly_calories_file <-
read_csv("projectFinal/fitDatabase/hourlyCalories_merged.csv")
```

## Converting csv files to parquet to gain performance

---

```
# write_parquet(sleep_day_file,
"projectFinal/projectDataParquet/sleepDay.parquet")
# write_parquet(daily_activity_file,
"projectFinal/projectDataParquet/dailyActivity.parquet")
# write_parquet(daily_intensities_file,
"projectFinal/projectDataParquet/dailyIntensities.parquet")
# write_parquet(hourly_intensities_file,
"projectFinal/projectDataParquet/hourly_intensities.parquet")
# write_parquet(hourly_calories_file,
"projectFinal/projectDataParquet/hourly_calories.parquet")
```

## Reading Parquet files

---

```
sleep_day_file <- read_parquet("projectFinal/projectDataParquet/sleepDay.parquet")
daily_activity_file <-
read_parquet("projectFinal/projectDataParquet/dailyActivity.parquet")
daily_intensities_file <-
read_parquet("projectFinal/projectDataParquet/dailyIntensities.parquet")
hourly_intensities_file <-
read_parquet("projectFinal/projectDataParquet/hourly_intensities.parquet")
hourly_calories_file <-
read_parquet("projectFinal/projectDataParquet/hourly_calories.parquet")
```

## Undestaining data

---

```
skim_without_charts(sleep_day_file)
skim_without_charts(daily_activity_file)
skim_without_charts(daily_intensities_file)
skim_without_charts(hourly_intensities_file)
skim_without_charts(hourly_calories_file)
```

```
sleep_day_file  
daily_activity_file  
daily_intensities_file  
hourly_intensities_file  
hourly_calories_file
```

```
glimpse(sleep_day_file)  
glimpse(daily_activity_file)  
glimpse(daily_intensities_file)  
glimpse(hourly_intensities_file)  
glimpse(hourly_calories_file)
```

## Number of unique users

```
count(distinct(sleep_day_file, Id))  
count(distinct(daily_activity_file, Id))  
count(distinct(daily_intensities_file, Id))  
count(distinct(hourly_intensities_file, Id))  
count(distinct(hourly_calories_file, Id))
```

## Data Cleaning

---

```
anyDuplicated(sleep_day_file)  
anyDuplicated(daily_activity_file)  
anyDuplicated(daily_intensities_file)  
anyDuplicated(hourly_intensities_file)  
anyDuplicated(hourly_calories_file)
```

## Dropping NA and duplicates

```
sleep_day_file <- sleep_day_file %>%  
  distinct() %>%  
  drop_na()  
  
anyDuplicated(sleep_day_file)
```

## Cleaning names to the format used in the classes

```
sleep_day_file <- clean_names(sleep_day_file)
daily_activity_file <- clean_names(daily_activity_file)
daily_intensities_file <- clean_names(daily_intensities_file)
hourly_intensities_file <- clean_names(hourly_intensities_file)
hourly_calories_file <- clean_names(hourly_calories_file)
```

## Tranforming the Dates

```
glimpse(sleep_day_file)
glimpse(daily_activity_file)
glimpse(daily_intensities_file)
glimpse(hourly_intensities_file)
glimpse(hourly_calories_file)
```

```
sleep_day_file$sleep_day <- mdy_hms(sleep_day_file$sleep_day)
daily_activity_file$activity_date <- mdy(daily_activity_file$activity_date)
daily_intensities_file$activity_day <- mdy(daily_intensities_file$activity_day)
hourly_intensities_file$activity_hour <-
mdy_hms(hourly_intensities_file$activity_hour)
hourly_calories_file$activity_hour <- mdy_hms(hourly_calories_file$activity_hour)
```

```
sleep_day_file
daily_activity_file
daily_intensities_file
hourly_intensities_file
hourly_calories_file
```

## Analizing

---

```
glimpse(sleep_day_file)
glimpse(daily_activity_file)
glimpse(daily_intensities_file)
glimpse(hourly_intensities_file)
glimpse(hourly_calories_file)
```

### Device use throughout the day

```
daily_activity_file$total_time =
rowSums(daily_activity_file[c("very_active_minutes", "fairly_active_minutes",
```

```
"lightly_active_minutes", "sedentary_minutes"]])
glimpse(daily_activity_file)
```

```
daily_activity_file |>
  group_by(id) |>
  summarise(daily_usage_hour = mean(total_time / 60)) |>
  ggplot(aes(x = daily_usage_hour)) +
  geom_histogram(
    # mapping = aes(x = daily_usage_hour),
    color = "black", fill = "#008b3a"
  ) +
  scale_color_igv() +
  scale_fill_igv() +
  theme_grey() +
  scale_x_continuous(breaks = c(1:24)) +
  labs(
    title = "Average Daily Device Usage Time",
    subtitle = "",
    x = "Daily Use Time(hours)",
    y = "Count"
  )
```

## Calories vs Steps

```
daily_activity_file |>
  ggplot() +
  (mapping = aes(x = total_steps, y = calories)) +
  geom_jitter() +
  geom_smooth() +
  stat_cor(method = "pearson", label.x = 20000, label.y = 4800) +
  # MODERATE CORRELATION
  # https://www.google.com/imgres?
imgurl=https%3A%2F%2Fsphweb.bumc.bu.edu%2Fotlt%2FMPH-Modules%2FPH717-
QuantCore%2FPH717-Module9-Correlation-
Regression%2FCorrelation%2520Coefficient%2520examples.png&imgrefurl=https%3A%2F%2F
sphweb.bumc.bu.edu%2Fotlt%2FMPH-Modules%2FPH717-QuantCore%2FPH717-Module9-
Correlation-Regression%2FPH717-Module9-Correlation-
Regression4.html&tbnid=12LiAxIyZkRAFM&vet=12ahUKEwiQ3sK-
x_f8AhXBuZUCHdWqCR0QMygDegUIARDeAQ..i&docid=odIvFkLwH6fSkM&w=800&h=448&q=correlati
on%20Coefficients&ved=2ahUKEwiQ3sK-x_f8AhXBuZUCHdWqCR0QMygDegUIARDeAQ
  # https://www.scribbr.com/statistics/pearson-correlation-
coefficient/#:~:text=The%20Pearson%20correlation%20coefficient%20(r,the%20relation
ship%20between%20two%20variables.
  scale_color_igv() +
  scale_fill_igv() +
  theme_grey() +
  labs(
    title = "Daily steps vs. calories",
    subtitle = "Pearson correlation coefficient (r)",
```

```
x = "Daily Steps",  
y = "Calories"  
)
```

The scale of Pearson's Correlation Coefficient  $R = 0.59$  (Moderate Correlation), Between 0 and 1 Positive correlation With a ModerateCorrelation may be other factors; faster steps...

```
hourly_intensities_file$day <- format(hourly_intensities_file$activity_hour,  
format = "%Y %m %d")  
hourly_intensities_file$calories <- cbind(hourly_calories_file$calories)  
glimpse(hourly_intensities_file)
```

```
hourly_intensities_file |>  
  group_by(day) |>  
  summarise(total_int = total_intensity, total_cal = calories) %>%  
  ggplot() +  
  (mapping = aes(x = total_int, y = total_cal)) +  
  geom_jitter() +  
  geom_smooth() +  
  stat_cor(method = "pearson", label.x = 50, label.y = 750) +  
  labs(  
    title = "Hourly intensity vs. calories",  
    x = "Hourly Intensity",  
    y = "Calories"  
  )
```

Now we can notice a Very High Correlation; 0.9. When one variable changes, the other variable changes in the same direction.