Processing Google Takeout Fitbit Data

Genevieve Roberts

2025-08-25

Load Packages and Setup Functions and Constants

```r
library(here)
library(tidyverse)
```

```r
source(here::here("fitbit/takeout_fitbit_processing_functions.R"))
```

```r
#current path
data_path <- here::here("fitbit/sample_fitbit_takeout_data/9Aug25_groberts_fitbit_takeout/Fitbit")
print(data_path)
```

```
## [1] "/Users/gen-omix/Documents/umass/VIGOR-surveys/fitbit/sample_fitbit_takeout_data/9Aug25_groberts_
```

Explore some FitBit data

```
#define some constants for the nb
start_date="2025-07-07"
end_date="2025-08-09"
```

Heart Rate Variablility

```
# Combined detailed + summary
data <- load_fitbit_hrv(start_date = start_date,
                        end_date = end_date,
                        root_dir= data_path,
                        summary_only = FALSE)

# Only summary data
summary_only <- load_fitbit_hrv(start_date = start_date,
                                end_date = end_date,
                                root_dir = data_path,
                                summary_only = TRUE)

pander(sample_n(data, 5))
```
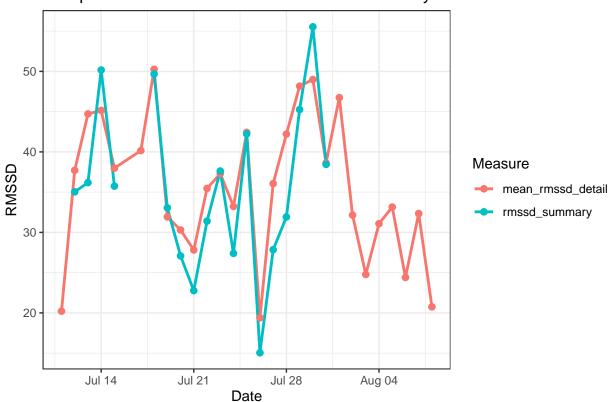
Table 1: Table continues below

timestamp_detail	rmssd_detail	coverage	low_frequency	high_frequency
2025-07-14 04:55:00	45.49	0.935	573.7	623.9
2025-08-03 03:55:00	18.49	0.952	218.7	75.9
2025-07-31 07:40:00	49.2	0.962	1083	556.6
2025-07-26 01:35:00	22.73	1.002	258.8	234.4
2025-08-02 01:10:00	35.34	1.002	888.6	335

file_date	timestamp_summary	rmssd_summary	nremhr	entropy
2025-07-14	2025-07-14	50.17	67	2.844
2025-08-03	NA	NA	NA	NA
2025-07-31	2025-07-31	38.45	68.44	2.616
2025-07-26	2025-07-26	15.05	87.78	2.098
2025-08-02	NA	NA	NA	NA

Here, I want to know if the `rmssd_summary` from the summary HRV files is simply the mean across all of the "detail" datapoints for that day. The plot below suggests they are not the same, but they are closely related.

```r
check_if_mean_equals_summary <- data %>%
  group_by(file_date) %>%
  summarize(
    mean_rmssd_detail = mean(rmssd_detail, na.rm = TRUE),
    rmssd_summary = first(rmssd_summary)  # summary has one value per date
  ) %>%
  ungroup()

# Prepare data in long format for plotting
mean_detail_compare_plot_df <- check_if_mean_equals_summary %>%
  pivot_longer(cols = c(mean_rmssd_detail, rmssd_summary),
               names_to = "Type",
               values_to = "RMSSD")

ggplot(mean_detail_compare_plot_df, aes(x = file_date, y = RMSSD, color = Type)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  labs(
    title = "Comparison of RMSSD: Detail Mean vs. Summary",
    x = "Date",
    y = "RMSSD",
    color = "Measure"
  ) +
  theme_bw()
```


Comparison of RMSSD: Detail Mean vs. Summary

The plot below shows how we can highlight a period of interest

```r
# Example dates to highlight
highlight_start <- as.Date("2025-07-29")
highlight_end <- as.Date("2025-08-02")

#add the mean rmssd from the detailed HRV information
data <- data %>%
  group_by(file_date) %>%
  mutate(mean_rmssd_detail = mean(rmssd_detail)) %>%
  ungroup()

ggplot(data, aes(x = file_date, y = mean_rmssd_detail)) +
  geom_line(color = "blue", size = 1) +
  geom_point(color = "blue", size = 2) +

  # Highlight date range with a transparent rectangle
  annotate(
    "rect",
    xmin = highlight_start, xmax = highlight_end,
    ymin = -Inf, ymax = Inf,
    alpha = 0.2, fill = "orange"
  ) +

  labs(
    title = "HRV Over Time with Highlighted Date Range",
    x = "Date",
    y = "HRV (Mean RMSSD Detail)"
  ) +

  theme_bw()
```

HRV Over Time with Highlighted Date Range