Processing Google Takeout Fitbit Data

Genevieve Roberts

2025-08-25

Load Packages and Setup Functions and Constants

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

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

#add some dates of interest to highlight
dates_of_interest_start = "2025-07-29"
dates_of_interest_end = "2025-07-31"
```

Heart Rate Variablility

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

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

pander(sample_n(hrv_data, 5))
```

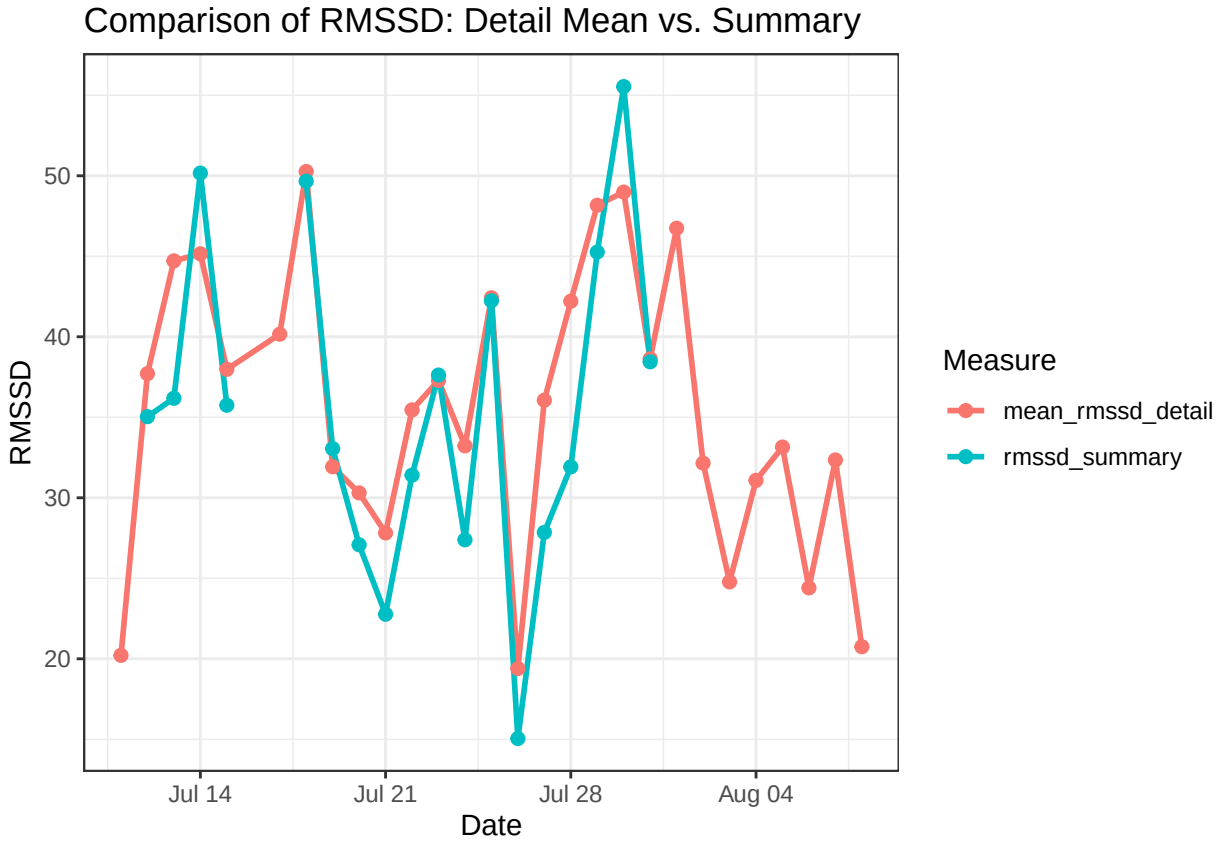
Table 1: Table continues below

timestamp_detail	rmssd_detail	coverage	low_frequency	high_frequency
2025-07-18 23:25:00	51.22	0.913	1476	704.5
2025-07-21 02:00:00	18.52	1	656.4	88.25
2025-08-07 06:35:00	28.13	0.971	970.2	164.4
2025-07-25 04:25:00	19.95	0.993	741.9	83.19
2025-08-03 06:00:00	38.12	1.003	773.9	432.1

file_date	timestamp_summary	rmssd_summary	nremhr	entropy
2025-07-18	2025-07-18	49.66	70.04	2.793
2025-07-21	2025-07-21	22.77	78.3	2.661
2025-08-07	NA	NA	NA	NA
2025-07-25	2025-07-25	42.24	70.28	2.695
2025-08-03	NA	NA	NA	NA

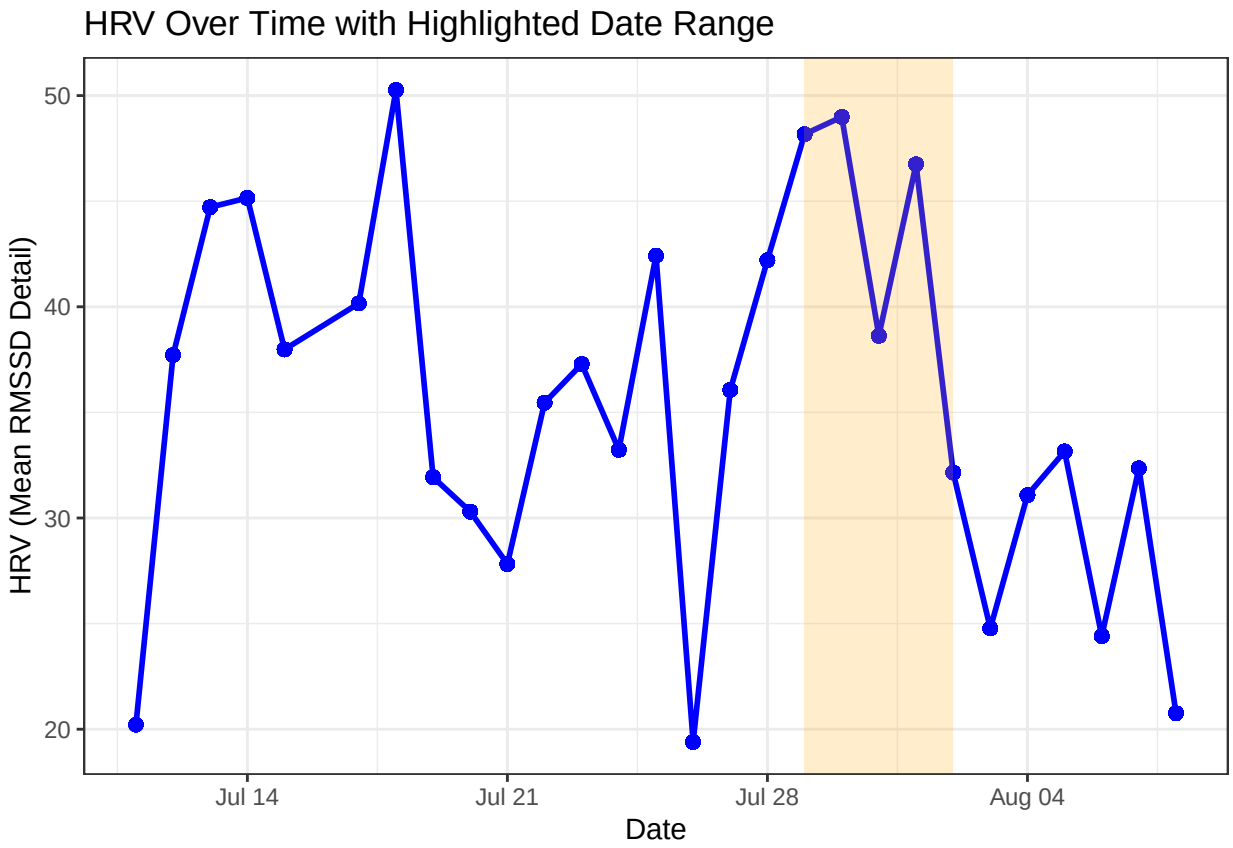
Here, I want to know if the `rmssd_summary` from the summary HRV files is simply the mean across all of the "detail" datapoints for that day. The plot below suggests they are not the same, but they are closely related.

```
check_if_mean_equals_summary <- hrv_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()
```



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

```
# 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
hrv_data <- hrv_data %>%
  group_by(file_date) %>%
  mutate(mean_rmssd_detail = mean(rmssd_detail)) %>%
  ungroup()

ggplot(hrv_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
HRV (Mean RMSSD Detail)
50
40
30
20
Jul 14
Jul 21
Jul 28
Aug 04
Date

Add Resting Heart Rate Data

```
rhr_data <- load_fitbit_resting_hr(start_date = start_date,
                                   end_date = end_date,
                                   root_dir = data_path)

# Combine by day (default)
combined <- combine_fitbit_data(hrv_data, rhr_data)
pander(sample_n(combined, 5))
```

Table 3: Table continues below

file_date	timestamp_detail	rmssd_detail	coverage	low_frequency
2025-08-01	2025-08-01 05:40:00	66.41	0.964	3073
2025-07-31	2025-07-31 08:20:00	25.63	1.003	694.9
2025-07-29	2025-07-29 02:00:00	31.3	1.003	2249
2025-07-25	2025-07-25 02:20:00	47.13	1.003	1198
2025-07-31	2025-07-31 06:05:00	43.1	0.891	3987

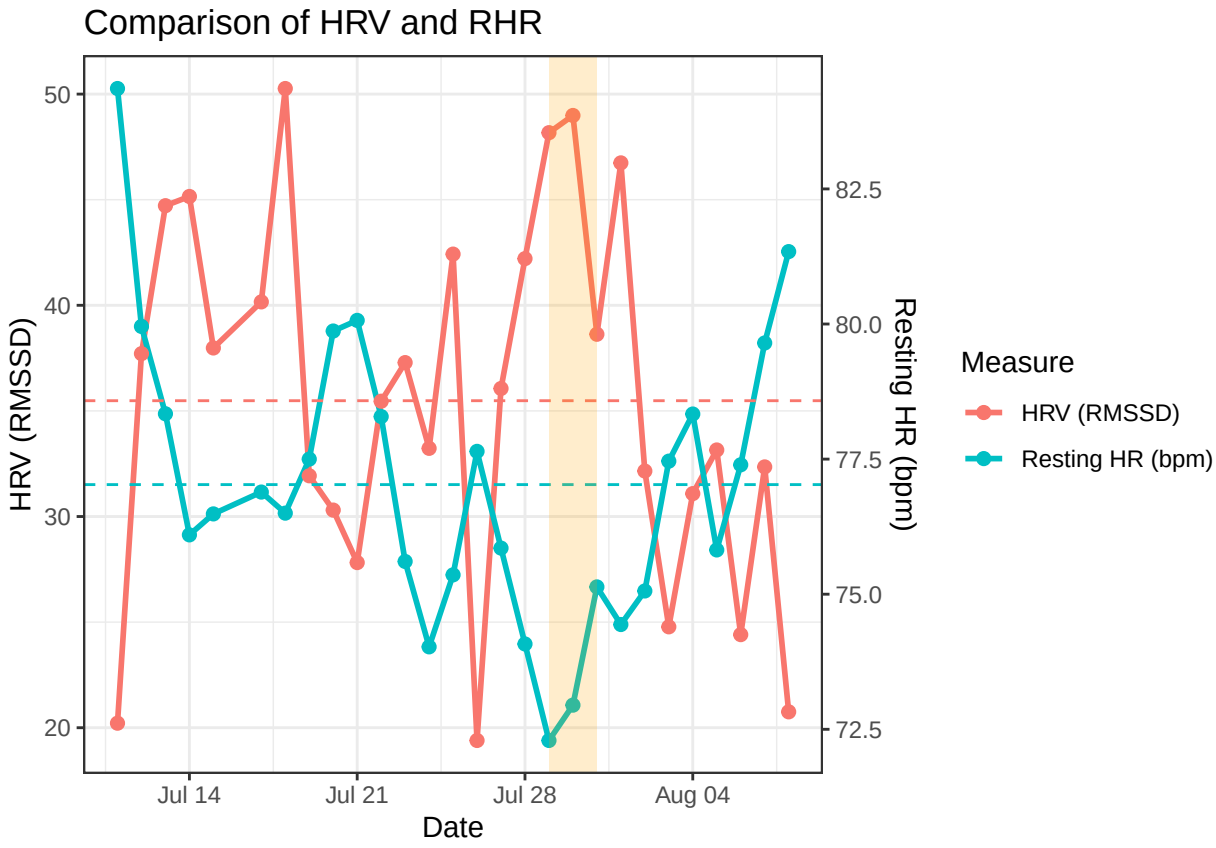
Table 4: Table continues below

high_frequency	timestamp_summary	rmssd_summary	nremhr	entropy
1189	NA	NA	NA	NA
130.1	2025-07-31	38.45	68.44	2.616
271.5	2025-07-29	45.26	64.95	2.941
697.8	2025-07-25	42.24	70.28	2.695
677.7	2025-07-31	38.45	68.44	2.616

mean_rmssd_detail	resting_hr
46.74	74.44
38.63	75.14
48.17	72.29
42.42	75.36
38.63	75.14

The plot below compares heart rate variability to resting heart rate. You can see there is a rough inverse correlation.

```
plot_dual_axis(
  data = combined,
  col1 = rmssd_detail,
  col2 = resting_hr,
  label1 = "HRV (RMSSD)",
  label2 = "Resting HR (bpm)",
  title = "Comparison of HRV and RHR",
  highlight_start = dates_of_interest_start,
  highlight_end = dates_of_interest_end
)
```



Daily Readiness Score

```
# Load the daily readiness score and combine it with the other data
daily_ready <- load_fitbit_daily_readiness(start_date = start_date,
                                           end_date = end_date,
                                           root_dir = data_path)

# Combine by day
combined <- combine_fitbit_data(combined, daily_ready)
```

```
plot_dual_axis(
  data = combined,
  col1 = daily_readiness_score,
  col2 = resting_hr,
  label1 = "Daily Readiness Score \n(Higher is Better)",
  label2 = "Resting HR \n(bpm)",
  title = "Comparison of Daily Readiness Score\nand Resting Heartrate",
  highlight_start = dates_of_interest_start,
  highlight_end = dates_of_interest_end
)
```