

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_fitbit_takeout/Fitbit"
```

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 = "2025-08-02"
```

Heart Rate Variability

```
# 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
2025-07-13 23:50:00	35.04	0.989	434.5	484.6
2025-07-21 05:40:00	15.91	0.947	392.7	118.2
2025-08-07 06:55:00	21.33	1.002	1295	150
2025-07-23 03:40:00	53.89	0.918	877.9	441.3
2025-08-06 07:20:00	51.67	0.996	1873	1064

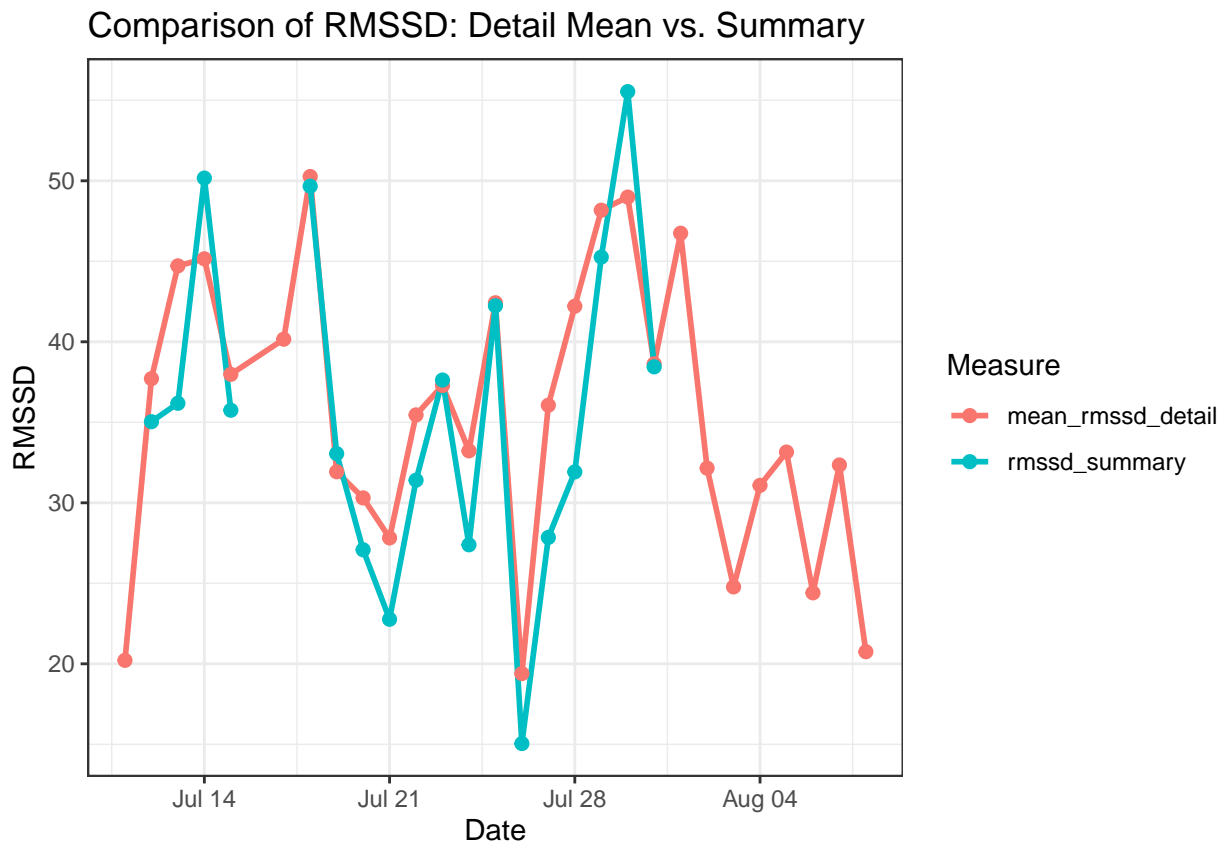
file_date	timestamp_summary	rmssd_summary	nremhr	entropy
2025-07-13	2025-07-13	36.18	73.27	2.75
2025-07-21	2025-07-21	22.77	78.3	2.661
2025-08-07	NA	NA	NA	NA
2025-07-23	2025-07-23	37.62	66.24	2.559
2025-08-06	NA	NA	NA	NA

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")

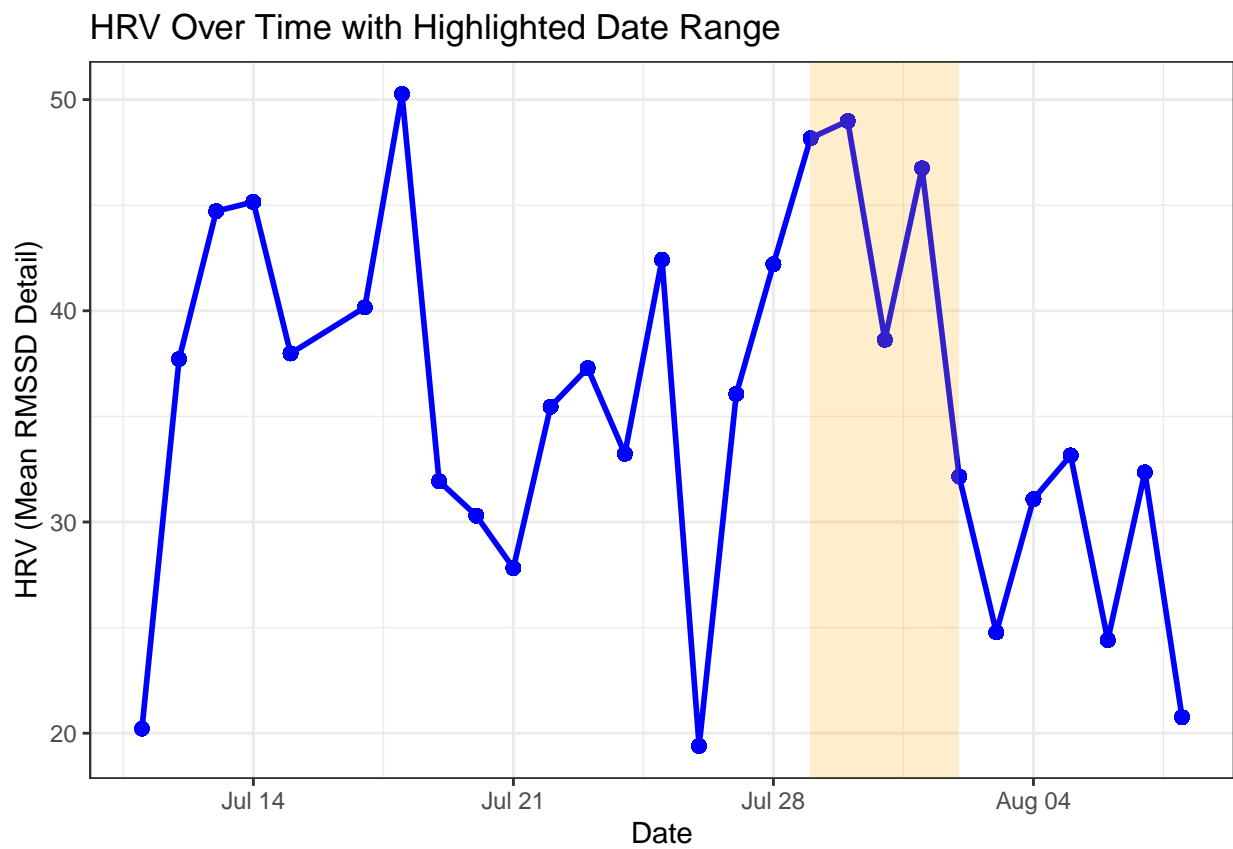
#add the mean rmssd from the detailed HRV information
hrv_data <- hrv_data %>%
  group_by(file_date) %>%
  mutate(mean_rmssd_detail = mean(rmssd_detail)) %>%
  ungroup()

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()
```



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
2025-07-18	2025-07-18 01:40:00	23.09	0.945	312.3
2025-08-02	2025-08-02 22:35:00	13.59	0.798	287.8
2025-07-19	2025-07-19 02:20:00	76.15	1.003	2067
2025-07-14	2025-07-14 21:55:00	22.72	0.919	337
2025-07-22	2025-07-22 04:45:00	20.76	1.003	455

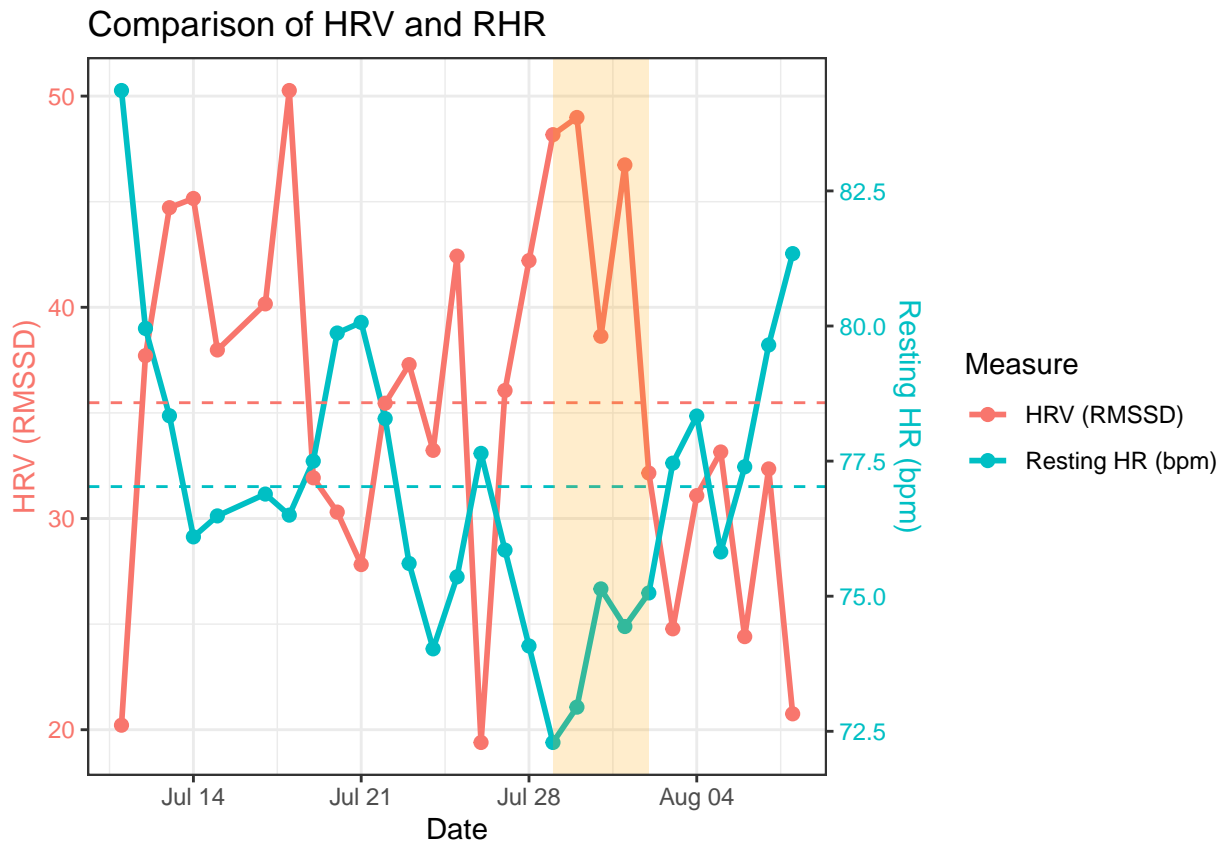
Table 4: Table continues below

high_frequency	timestamp_summary	rmssd_summary	nremhr	entropy
160	2025-07-18	49.66	70.04	2.793
53.2	NA	NA	NA	NA
1769	2025-07-19	33.05	72.56	2.614
243.9	2025-07-14	50.17	67	2.844
65.23	2025-07-22	31.4	68.38	2.606

mean_rmssd_detail	resting_hr
50.26	76.5
32.15	75.06
31.93	77.5
45.15	76.1
35.46	78.29

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)
```

The plot below shows a roughly inverse correlation between Daily Readiness Score and Resting Heartrate:

```
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
)
```

