Bellabeat Visualizations

Alyssa McCormick

2025-03-07

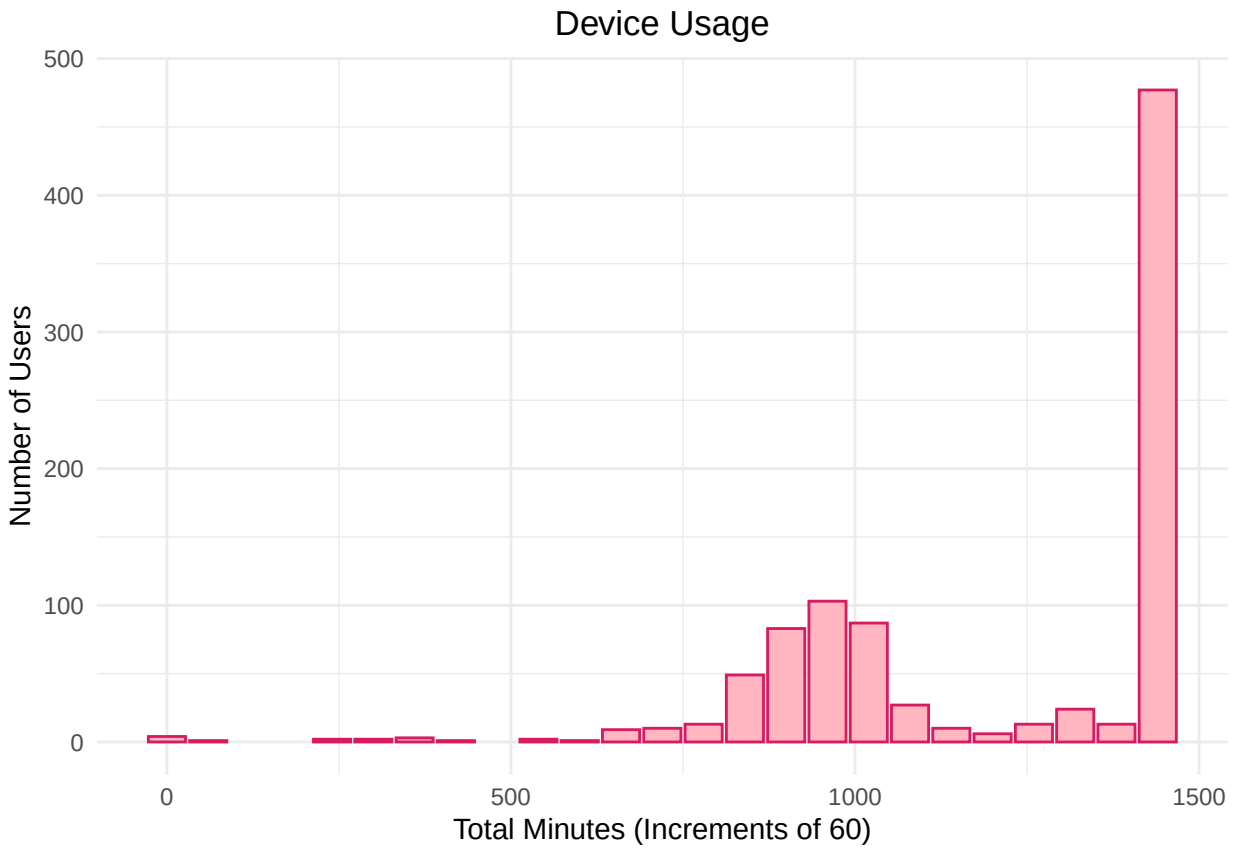
```
# Installed libraries
library(readr)
library(ggplot2)
```

User Device Usage

This histogram illustrates the distribution of the **total minutes users wear their FitBit devices per day**. The x-axis represents **total minutes worn**, grouped in increments of 60 minutes, and the y-axis represents the **number of users in each time range**.

This distribution highlights how many users wear their FitBit for the **full 1440 minutes (24 hours)** versus those who remove their device at various times throughout the day. This insight helps identify user engagement levels and can inform Bellabeat's marketing strategies by understanding typical device usage patterns.

```
# Device Usage Histogram
ggplot(data = deviceUsage, mapping = aes(x = MinuteRange, y = NumUsers)) +
  geom_col(fill = "#FFB6C1", color="#D81B60") +
  labs(title = "Device Usage", x="Total Minutes (Increments of 60)", y="Number of Users") +
  theme_minimal() +
  theme(plot.title = element_text(hjust=0.5))
```



Correlation Between Sleep Duration and Step Count

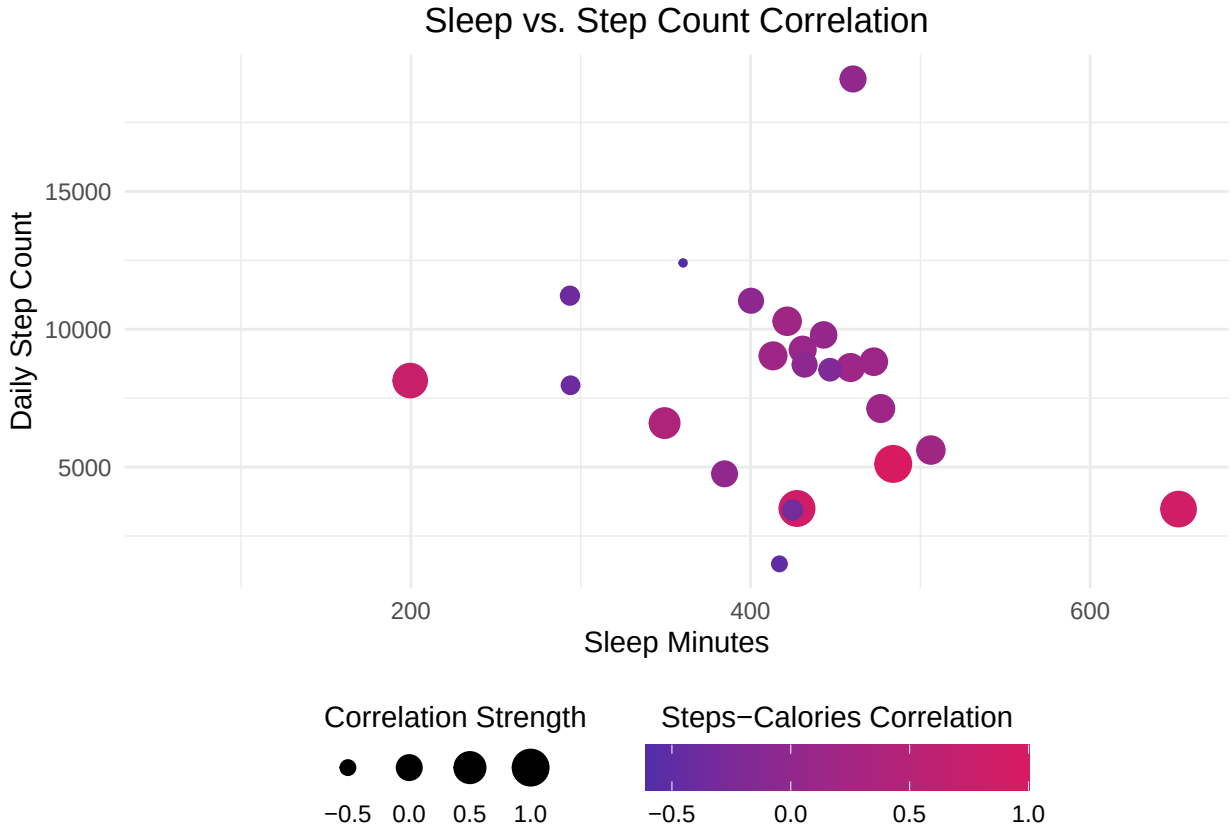
The scatter plot displays the **relationship between Average Daily Steps and Average Calories Burned**, with the strength of the correlation between these 2 variables represented through the color and size of the points. Each point represents one individual, and the x-axis shows the **average sleep duration** and the y-axis shows the **average daily steps taken**.

The color and size of each point correspond to the strength of correlation between steps and calories. The points with a **stronger correlation** appear in **green** and are larger in size, while the points with a **weaker correlation** appear **red** and are smaller in size. This helps highlight individuals that have a stronger or weaker relationship between their sleep habits and daily steps.

This visualization helps identify patterns in user behavior, showing how consistent or inconsistent the link is between sleep and activity. The correlation values indicate that some users exhibit a **strong positive correlation**, meaning more sleep is associated with more steps, while other users show a **negative or weak correlation**, suggesting a lot of variability in individual sleep and activity habits.

```
# Sleep vs. Steps Correlation scatter plot
ggplot(data = sleepStepsCorr, aes(x = AvgSleepMinutes, y = AvgDailySteps)) +
  geom_point(aes(color = SleepStepsCorrelation, size = SleepStepsCorrelation)) +
  # gradient coloring
  scale_color_gradient(low = "#512DA8", high = "#D81B60") +
  labs(title = "Sleep vs. Step Count Correlation",
       x = "Sleep Minutes",
       y = "Daily Step Count",
       color = "Sleep-Steps Correlation",
```

```
       size = "Correlation Strength") +
theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5), legend.position = "bottom") +
  # Formatting the legend
  guides(color = guide_colorbar(title = "Steps-Calories Correlation",
                                title.position = "top",
                                title.hjust = 0.5,
                                barwidth = 10),
         size = guide_legend(title = "Correlation Strength",
                             title.position = "top",
                             label.position = "bottom"))
```



Correlation Between Step Count and Calories Burned

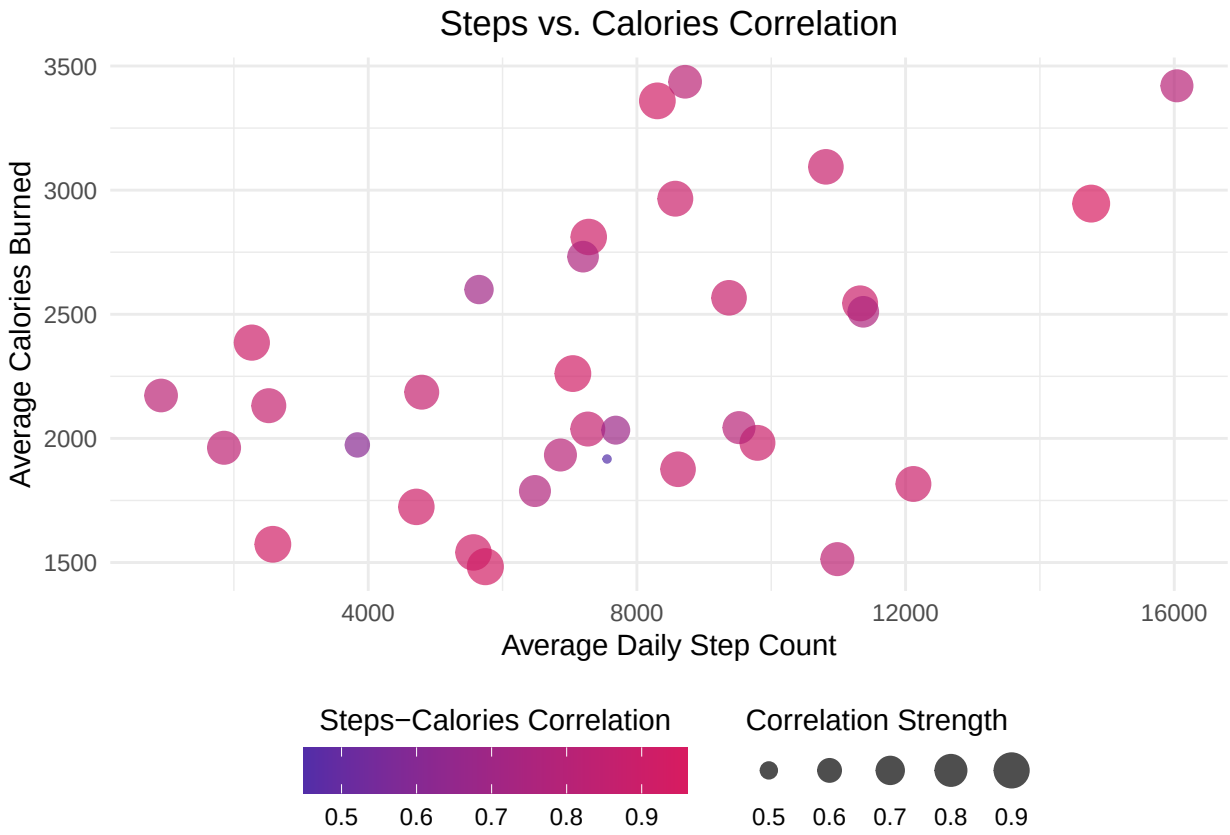
This scatter plot visualizes the **relationship between Average Daily Steps and Average Calories Burned** for different individuals, with an additional emphasis on the strength of the correlation between these two variables. Each point represents one individual, and the x-axis shows the **average daily steps** and the y-axis shows the **average calories burned**.

The color and size of the points is determined by the Steps-Calories Correlation, with points that have a **stronger correlation** between steps and calories appearing in **green** and larger in size, while those with a **weaker correlation** appearing in **red** and smaller in size. This helps highlight individuals with a stronger or weaker relationship between their daily steps and calorie expenditure.

The correlation values show a **strong positive relationship**, with a high degree of consistency in users'

step count and calories burned. Most users exhibit a positive correlation, suggesting that as daily steps increase, so does the number of calories burned. The visual highlights this **direct relationship**, with some slight variations across individual users. This visualization aids in understanding user behaviors and may help inform targeted strategies for improving fitness or device recommendations.

```
# Steps vs. Cals scatter plot
ggplot(data=stepsCalsCorr, aes(x=AvgDailySteps, y=AvgCalories)) +
  geom_point(aes(color=StepsCalsCorrelation, size=StepsCalsCorrelation),
             alpha=0.7) +
  # graident coloring
  scale_color_gradient(low="#512DA8", high="#D81B60") +
  labs(title="Steps vs. Calories Correlation",
       x="Average Daily Step Count",
       y="Average Calories Burned",
       color="Steps-Calories Correlation",
       size="Correlation Strength") +
  theme_minimal() +
  theme(plot.title = element_text(hjust=0.5), legend.position = "bottom") +
  # Formatting the legend
  guides(color = guide_colorbar(title = "Steps-Calories Correlation",
                                title.position = "top",
                                title.hjust = 0.5,
                                barwidth = 10),
         size = guide_legend(title = "Correlation Strength",
                             title.position = "top",
                             label.position = "bottom"))
```

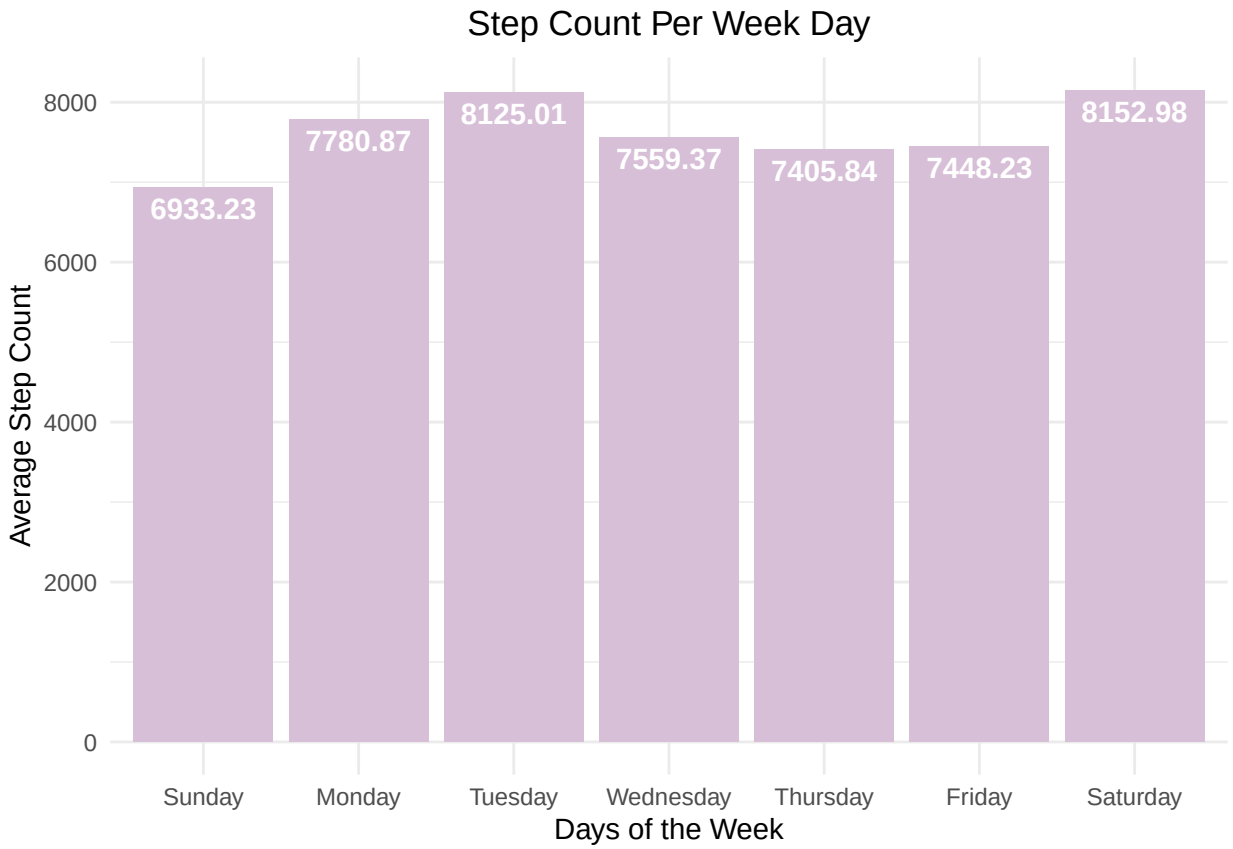


Average Step Count Per Week Day

This bar chart displays the **average daily step count for each day of the week** among FitBit users. The x-axis represents the **day on the week** (Sunday through Saturday), while the y-axis represents the **average number of steps taken per day** by all users. Each bar is labeled with the precise average step count for clarity.

Tuesday and Saturday show the **highest average step counts**, suggesting users tend to be more active on those days while Sunday has the **lowest average step counts**, indicating a possible trend of reduced physical activity on that day. This analysis can help identify weekly activity trends which can inform Bellabeat's marketing strategy by targeting users on less active days or emphasizing features that encourage daily movement.

```
# Bar Chart for average steps per week day
ggplot(data=weekdayStepCount,
       # Ensure days of the week are displayed in order
       aes(x=factor(DayOfWeek, levels=c("Sunday", "Monday", "Tuesday",     "Wednesday", "Thursday", "Fri
           y=AvgSteps)) +
  geom_bar(stat="identity", fill="#D8BFD8") +
  geom_text(aes(label=AvgSteps), vjust=1.5, color="#FFFFFF", size=4, fontface="bold") +
  labs(title="Step Count Per Week Day",
       x="Days of the Week",
       y="Average Step Count") +
  theme_minimal() +
  theme(plot.title = element_text(hjust=0.5))
```



Average Step Count Per Each Hour

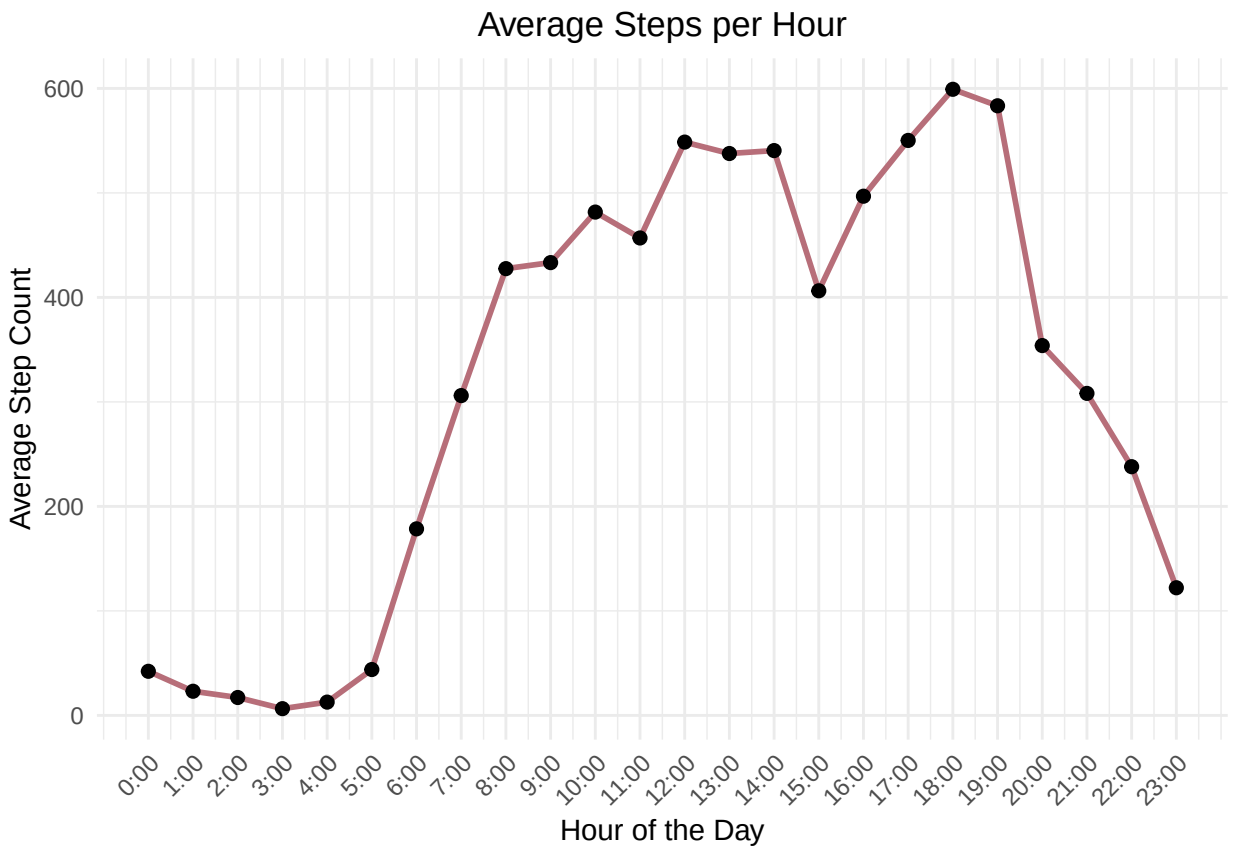
This line chart visualizes the **average number of steps taken per hour throughout the day** among all FitBit users. The x-axis displays **the hour** in 24-hour format, labeled in hourly increments, while the y-axis represents that **average number of steps** taken by all users for each hour of the day.

The visualization finds that the highest step counts occur in the evening with the **highest average steps** at 6:00 pm, 7:00 pm, and 5:00 pm. Activity levels are **consistently at their lowest** in the late night and early morning hours. This line chart highlights distinct daily movement patterns, with activity peaking in the afternoon and evening, and declining sharply overnight, reflecting common lifestyle and work routines.

```
# Format 'Hour' as a number to correctly display line graph
hourlySteps$Hour <- as.numeric(hourlySteps$Hour)

# Line chart for average steps per hour
ggplot(data=hourlySteps, mapping=aes(x=Hour, y=AvgSteps)) +
  geom_line(color="#B76E79", linewidth=1) +
  geom_point(color="#000000", size=2) +
  # Label hours in the correct format
  scale_x_continuous(
    breaks=seq(0, 23, 1),
    labels=paste0(seq(0, 23, 1), ":00")
    ) +
```

```
  labs(title="Average Steps per Hour",
       x="Hour of the Day",
       y="Average Step Count") +
  theme_minimal() +
  # Rotate x-axis labels so they are all visible
  theme(axis.text.x = element_text(angle=45, hjust=1), plot.title = element_text(hjust=0.5))
```



Average Minutes Per Activity Segment

This bar chart displays the **average total minutes spent in physical activity** for users categorized into 3 different activity segments: **Highly Active, Moderately Active,** and **Low Activity.** The x-axis represents the **3 classifications** based on each individual users' average daily intensity. The y-axis represents the **average minutes** ALL users within that class spend in activity per day.

This visualization shows that users who spend more time being sedentary or doing low intensity activity also do **less activity in general**, while users who often do high intensity workouts also spend **more time overall being active**. This chart provides insights into how users categorized by overall activity level differ in their total active minutes per day, helping identify trends in engagement with physical activity and market towards users with their specific activity behaviors in mind.

```
ggplot(data=activitySegments,
       mapping=aes(x=factor(ActivitySegment, levels=c("Low Activity", "Moderately Active", "Highly Acti
                   y=AvgMinutes)) +
  geom_bar(stat="identity", fill="#FF6F61") +
```

```
labs(title="Average Minute per Activity Segments",
     x="Activity Segment",
     y="Average Active Minutes") +
theme_minimal() +
theme(plot.title = element_text(hjust=0.5))
```