Fitbit Participant Data Analysis

Sam Valenzuela

2022-11-22

Setting up my environment

```
install.packages('tidyverse')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.2'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Importing the dailyActivity_merged and sleepDay_merged csv files

```
daily_activity <- read.csv("/cloud/project/Capstone/dailyActivity_merged.csv")
sleep_activity <- read.csv("/cloud/project/Capstone/sleepDay_merged.csv")
```

Outer join the csv files on the mutual primary key

```
merged_activity <- merge(x = daily_activity, y = sleep_activity,
                         by = "UniqueID", all = TRUE)
```

A table only on days where participants didn't wear their device

```
daily_activity_deviceless <- filter(daily_activity,
                                    daily_activity$TotalSteps == 0)
```

Variables to count the total number of participant days recorded, where participants wore their devices, where participants wore their device to sleep and where participants did not wear their device.

```
total_customer_days_observed <- sum(merged_activity$TotalSteps >= 0)
daily_activity_count <- sum(merged_activity$TotalSteps > 0)
```

```
sleep_activity_count <- sum(sleep_activity$UniqueID > 0)
participant_days_without_steps <- sum(merged_activity$TotalSteps == 0)
```

Data frame that compares the values for days participants wore their devices and days participants wore their devices to sleep

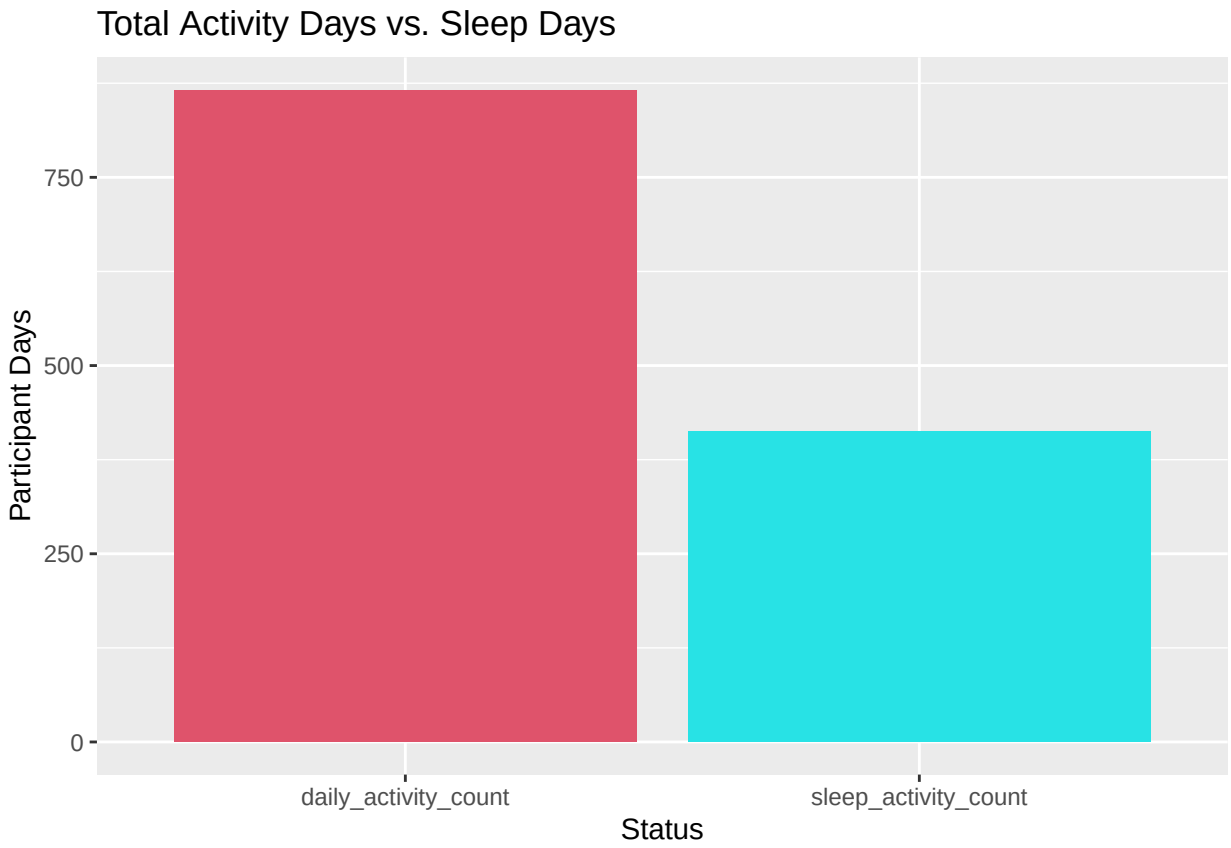
```
act_vs_sleep <- data.frame(status = c("daily_activity_count",
                                      "sleep_activity_count"),
                           participant_days = c(daily_activity_count,
                                      sleep_activity_count))
```

Data frame that compares the values for days participants wore their devices and days participants didn't wear their devices

```
days_with_and_without_steps <- data.frame(day_type = c("days_with_steps",
                                      "days_without_steps"),
                           number_of_participant_days =
                             c(daily_activity_count,
                               participant_days_without_steps))
```

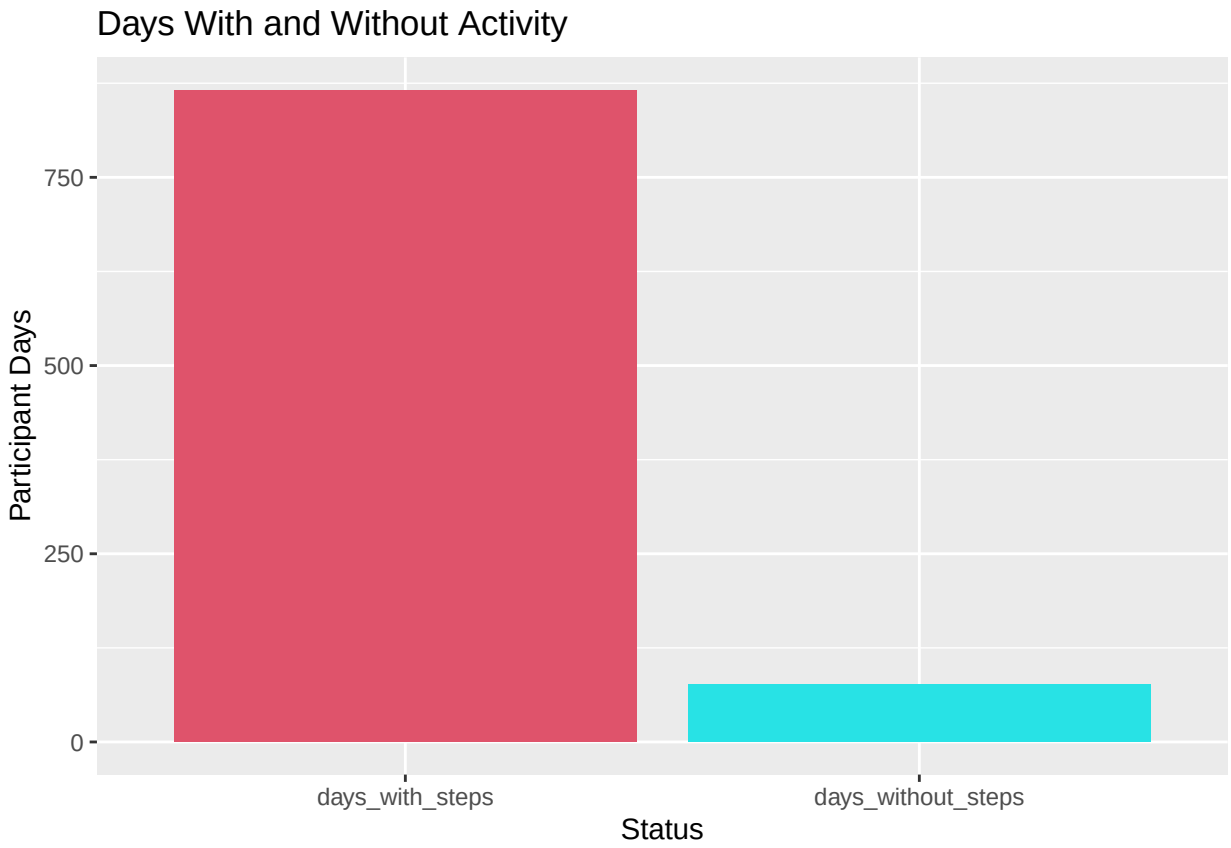
Bar graph that plots days participants wore their devices and days participants wore their devices to sleep

```
ggplot(data = act_vs_sleep, aes(x = status, y = act_vs_sleep$participant_days)) +
  geom_bar(stat = "identity", fill = act_vs_sleep$participant_days) +
  labs(title = "Total Activity Days vs. Sleep Days") +
  ylab("Participant Days") +
  xlab("Status")
```



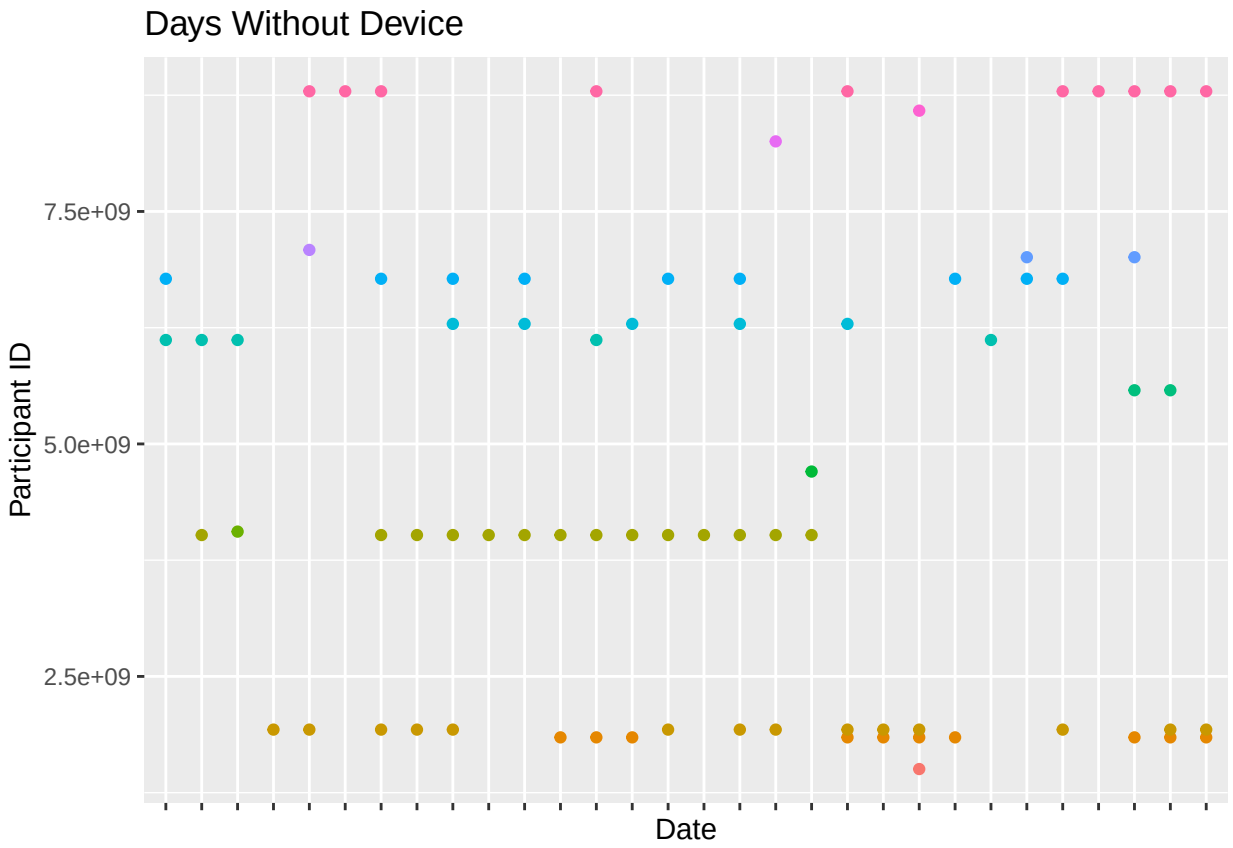
Bar graph that plots days participants wore their devices and days participants didn't wear their devices

```
ggplot(data = days_with_and_without_steps,
       aes(x = day_type, y = days_with_and_without_steps$number_of_participant_days)) +
  geom_bar(stat = "identity", fill = days_with_and_without_steps$number_of_participant_days) +
  labs(title = "Days With and Without Activity") +
  ylab("Participant Days") +
  xlab("Status")
```



Point plot that shows days where certain Ids did not wear their device

```
ggplot(data = daily_activity_deviceless) +
  geom_point(
    aes(x = ActivityDate, y = Id,
    color = as.factor(Id))) +
  labs(title = "Days Without Device") +
  theme(legend.position = "none", axis.text.x = element_blank()) +
  ylab("Participant ID") +
  xlab("Date")
```


Days Without Device
Participant ID
7.5e+09
5.0e+09
2.5e+09
Date