

title: "bellabeatproject" author: "Amady" date: "5/24/2021"

output: html_document

Install and load packages

```
```\r} install.packages("rmarkdown") install.packages("tidyverse") install.packages("skimr") install.packages("janitor") library(tidyverse) library(skimr) library(janitor) library(dplyr) library(lubridate) list.files(path = "../input/fitbit/Fitabase Data 4.12.16-5.12.16") getwd()
```

```
take a close look to the dataset
```\r}
View(dailyActivity_merged)
head(dailyActivity_merged)
colnames(dailyActivity_merged)
colnames(dailyCalories_merged)
colnames(dailyIntensities_merged)
colnames(dailySteps_merged)
colnames(heartrate_seconds_merged)
colnames(hourlyCalories_merged)
colnames(hourlyIntensities_merged)
colnames(hourlySteps_merged)
colnames(minuteCaloriesNarrow_merged)
colnames(minuteCaloriesWide_merged)
colnames(minuteIntensitiesNarrow_merged)
colnames(minuteMETsNarrow_merged)
colnames(minuteSleep_merged)
colnames(minuteStepsNarrow_merged)
colnames(minuteStepsWide_merged)
colnames(sleepDay_merged)
colnames(weightLogInfo_merged)
```

replacing the na values on the column Fat on weighlog per 0

```
```\r} weightLogInfo_merged[is.na(weightLogInfo_merged)] = 0
```

```
Understanding some summary statistics
How many unique participants are there in each dataframe? It looks like there may be more participants in the daily activity da
```\r}
n_distinct(dailyActivity_merged$Id)
n_distinct(sleepDay_merged$Id)
```

How many observations are there in each frame?

```
```\r} nrow(dailyActivity_merged) nrow(sleepDay_merged)
```

```
What are some quick summary statistics we would want to know about each data frame?

#####For the daily activity dataframe:
```\r}
dailyActivity_merged %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes) %>%
  summary()
```

For the sleep dataframe:

```
```\r} sleepDay_merged %>%
select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>% summary()
```

```
For the Calory dataframe
```{r}
dailyCalories_merged %>%
  select(Calories,
         ActivityDay) %>%
  summary()
```

For the weightlog dataframe

```
```{r} weightLogInfo_merged %>%
select(WeightKg, WeightPounds, Fat) %>% summary()
```

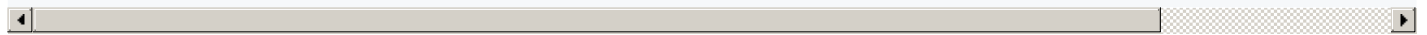
```
for the heartrate daraframe
```{r}
heartrate_seconds_merged %>%
  select(Value) %>%
  summary()
```

Plotting a few explorations

What's the relationship between steps taken in a day and sedentary minutes? How could this help inform the customer segments that we can market to? E.g. position this more as a way to get started in walking more? Or to measure steps that you're already taking?

```
```{r} ggplot(data=dailyActivity_merged, aes(x=TotalSteps, y=SedentaryMinutes, color=TotalSteps)) + geom_point() ggsave("totalsteps_sedentary.png")
```

```
What's the relationship between minutes asleep and time in bed? You might expect it to be almost completely linear - are there
```{r}
ggplot(data=sleepDay_merged, aes(x=TotalMinutesAsleep, y=TotalTimeInBed, color=TotalMinutesAsleep)) + geom_point()
ggsave("minuteasleep_timeinbed.png")
```



Merging these two datasets

```
```{r} combined_data <- merge(sleepDay_merged, dailyActivity_merged, by="Id") n_distinct(combined_data$Id) View(combined_data) q View(weightLogInfo_merged)
```

#### ## Including Plots

You can also embed plots, for example:

```
```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.