GYM Data Analysis

Arnav, Sam, Quinn, and Santiago

2024-11-13

Question 1

For lab credit, submit at least one table or figure from (some of?) the data that you plan to use for the final project. Describe the data in the figure or the values in the table.

Choosing your research question and data is a very important step. Try to choose something that you care about and that you can use to practice and further develop your data science skills.

```
library(tidyverse)
library(tidymodels)
library(ggplot2)
library(dplyr)
```

With this data set we are considering taking the approach of "being" a personal trainer who is analyzing this data set in order to best determine what types of lessons to offer for what specific person.

```
gym_members_data <- read_csv("gym_members_exercise_tracking.csv")
```

```
Rows: 973 Columns: 15
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (2): Gender, Workout_Type
dbl (13): Age, Weight (kg), Height (m), Max_BPM, Avg_BPM, Resting_BPM, Sessi...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
mega_gym_data <- read_csv("megaGymDataset.csv")
```

```
New names:
Rows: 2918 Columns: 9
-- Column specification
-------------------------------------------------------- Delimiter: "," chr
(7): Title, Desc, Type, BodyPart, Equipment, Level, RatingDesc dbl (2): ...1,
Rating
i Use `spec()` to retrieve the full column specification for this data. i
Specify the column types or set `show_col_types = FALSE` to quiet this message.
* `` -> `...1`
```

```r
head(gym_members_data)
```

```
# A tibble: 6 x 15
    Age Gender `Weight (kg)` `Height (m)` Max_BPM Avg_BPM Resting_BPM
  <dbl> <chr>          <dbl>        <dbl>   <dbl>   <dbl>       <dbl>
1    56 Male            88.3         1.71     180     157          60
2    46 Female          74.9         1.53     179     151          66
3    32 Female          68.1         1.66     167     122          54
4    25 Male            53.2         1.7      190     164          56
5    38 Male            46.1         1.79     188     158          68
6    56 Female          58           1.68     168     156          74
# i 8 more variables: `Session_Duration (hours)` <dbl>, Calories_Burned <dbl>,
#   Workout_Type <chr>, Fat_Percentage <dbl>, `Water_Intake (liters)` <dbl>,
#   `Workout_Frequency (days/week)` <dbl>, Experience_Level <dbl>, BMI <dbl>
```

```r
head(mega_gym_data)
```

```
# A tibble: 6 x 9
   ...1 Title            Desc  Type  BodyPart Equipment Level Rating RatingDesc
  <dbl> <chr>            <chr> <chr> <chr>    <chr>     <chr>  <dbl> <chr>
1     0 Partner plank ba~ "The~ Stre~ Abdomin~ Bands     Inte~      0 <NA>
2     1 Banded crunch is~ "The~ Stre~ Abdomin~ Bands     Inte~     NA <NA>
3     2 FYR Banded Plank~ "The~ Stre~ Abdomin~ Bands     Inte~     NA <NA>
4     3 Banded crunch     "The~ Stre~ Abdomin~ Bands     Inte~     NA <NA>
5     4 Crunch            "The~ Stre~ Abdomin~ Bands     Inte~     NA <NA>
6     5 Decline band pre~ "The~ Stre~ Abdomin~ Bands     Inte~     NA <NA>
```

```r
type_mega_counts <- mega_gym_data %>%
  count(Type) %>%
  arrange(desc(n))

# View the results
type_mega_counts
```

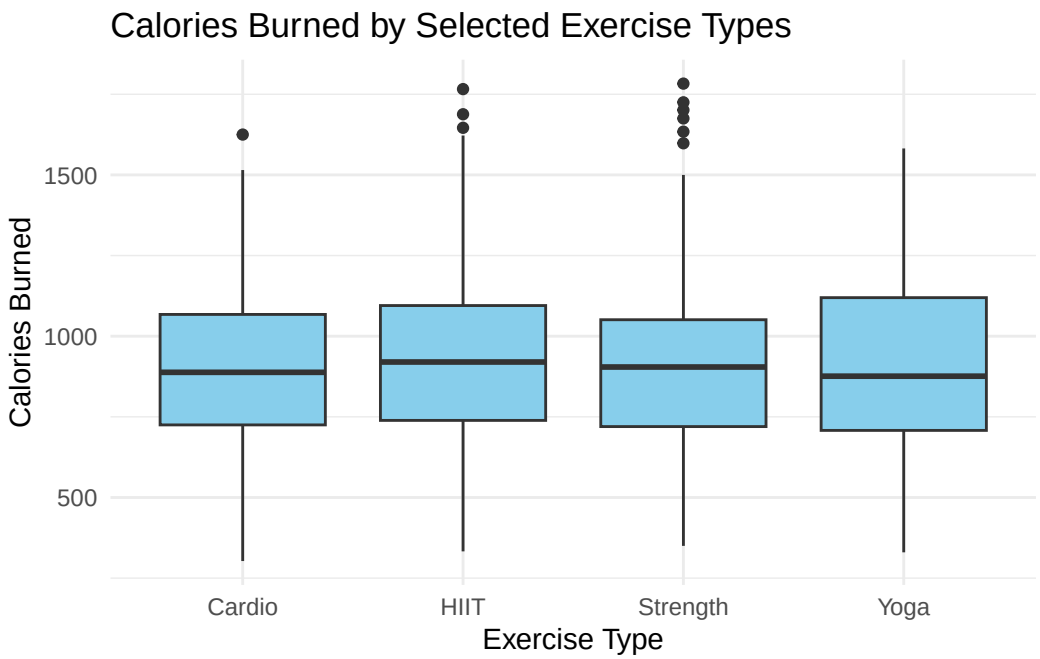
```
# A tibble: 7 x 2
  Type                     n
  <chr>                <int>
1 Strength              2545
2 Stretching             147
3 Plyometrics             97
4 Powerlifting            37
5 Cardio                  35
6 Olympic Weightlifting   35
7 Strongman               22
```

```r
type_gym_counts <- gym_members_data %>%
  count(Workout_Type) %>%
  arrange(desc(n))

# View the results
type_gym_counts
```

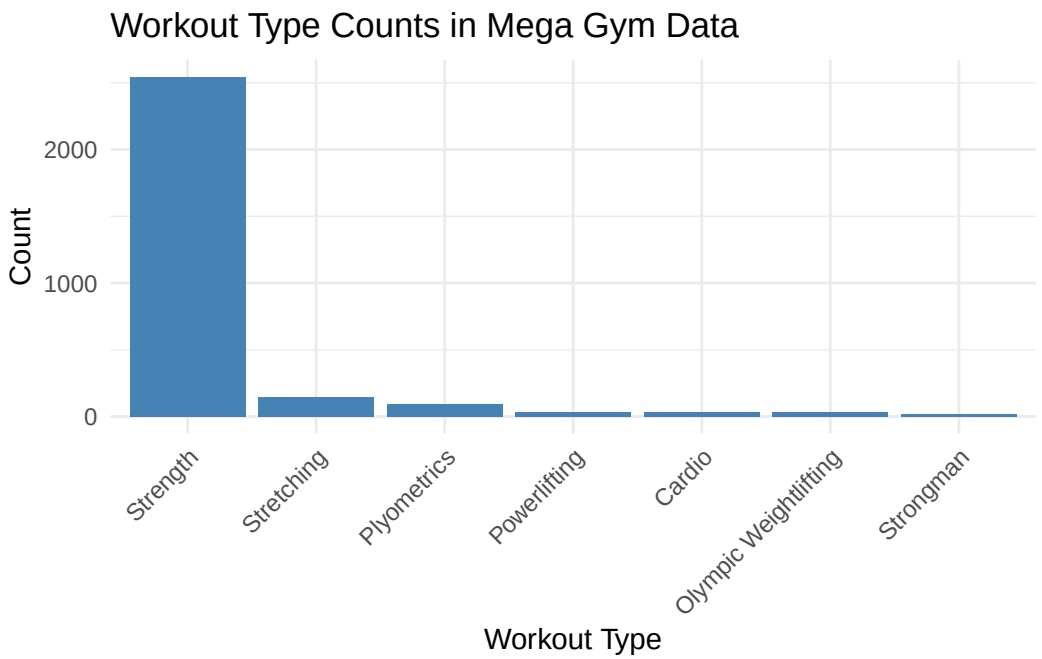
```
# A tibble: 4 x 2
  Workout_Type     n
  <chr>        <int>
1 Strength       258
2 Cardio         255
3 Yoga           239
4 HIIT           221
```

```r
# Boxplot of Calories Burned by the specified exercise types
ggplot(gym_members_data, aes(x = Workout_Type, y = Calories_Burned)) +
  geom_boxplot(fill = "skyblue") +
  labs(title = "Calories Burned by Selected Exercise Types", x = "Exercise Type", y = "Calori
  theme_minimal()
```

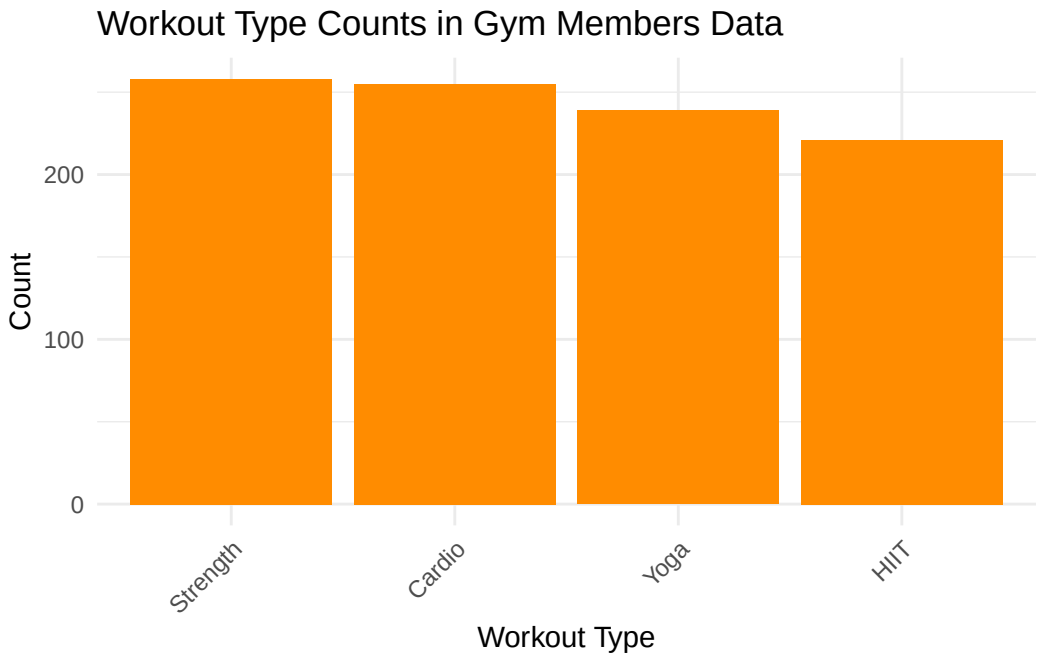


in these two figures above we had each data set be counted by the type of workout. What we can see is that for Mega Counts strength workouts are a lot more common with 2,545 followed by stretching with only 147. This story continues for the Gym Counts as strength is the most common one but not by nearly as large as a margin as we see that the count is 258 and is followed by Cardio with 255. The two graphics below help display this better.

```
ggplot(type_mega_counts, aes(x = reorder(Type, -n), y = n)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Workout Type Counts in Mega Gym Data",
       x = "Workout Type",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Plot for type_gym_counts
ggplot(type_gym_counts, aes(x = reorder(Workout_Type, -n), y = n)) +
  geom_bar(stat = "identity", fill = "darkorange") +
  labs(title = "Workout Type Counts in Gym Members Data",
       x = "Workout Type",
       y = "Count") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Research Question

How does the type of exercise and experience level influence calorie burn and heart rate metrics across different age groups and genders, and is there a correlation with the exercise ratings provided by gym members?

How does the type of exercise and experience level influence calorie burn and heart rate metrics across different age groups and genders, and is there a correlation with the exercise ratings provided by gym members?