

# How to Promote Physical Activity in the Gym?

By: Johnson Vo

10/22/2024

## 1. Introduction

As individuals look for ways to improve their physical health, many will often focus on going to the gym as a starting point. Regardless of experience level, going to the gym develop or enhance one's physical well-being. However, going to the gym is not as straight-forward as it may seem at a glance.

Although some may find it to be intimidating due to the amount of equipment, it just means that there is something for everyone. This intimidation ultimately discourages people to maintain their membership for a very short amount of time. But, what can we do to address these issues?

This begins with taking a look at the “Gym Members Exercise Dataset” by Vala Khorasani.

Source: Gym Members Exercise Dataset

```
# Load the tidyverse package
library('tidyverse')

# Import the dataset
gym <-
  read_csv(
    "C:/Users/Johns/Desktop/Data Files/CSV Files/gym_members_exercise_tracking.csv"
  )
```

```
# Structure of the dataset
str(gym)
```

```
## spc_tbl_ [973 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Age : num [1:973] 56 46 32 25 38 56 36 40 28 28 ...
## $ Gender : chr [1:973] "Male" "Female" "Female" "Male" ...
## $ Weight (kg) : num [1:973] 88.3 74.9 68.1 53.2 46.1 ...
## $ Height (m) : num [1:973] 1.71 1.53 1.66 1.7 1.79 1.68 1.72 1.51 1.94 1.84 ...
## $ Max_BPM : num [1:973] 180 179 167 190 188 168 174 189 185 169 ...
## $ Avg_BPM : num [1:973] 157 151 122 164 158 156 169 141 127 136 ...
## $ Resting_BPM : num [1:973] 60 66 54 56 68 74 73 64 52 64 ...
## $ Session_Duration (hours) : num [1:973] 1.69 1.3 1.11 0.59 0.64 1.59 1.49 1.27 1.03 1.08 ...
## $ Calories_Burned : num [1:973] 1313 883 677 532 556 ...
## $ Workout_Type : chr [1:973] "Yoga" "HIIT" "Cardio" "Strength" ...
## $ Fat_Percentage : num [1:973] 12.6 33.9 33.4 28.8 29.2 15.5 21.3 30.6 28.9 29.7 ...
## $ Water_Intake (liters) : num [1:973] 3.5 2.1 2.3 2.1 2.8 2.7 2.3 1.9 2.6 2.7 ...
## $ Workout_Frequency (days/week) : num [1:973] 4 4 4 3 3 5 3 3 4 3 ...
## $ Experience_Level : num [1:973] 3 2 2 1 1 3 2 2 2 1 ...
## $ BMI : num [1:973] 30.2 32 24.7 18.4 14.4 ...
```

```
## - attr(*, "spec")=
## .. cols(
## ..   Age = col_double(),
## ..   Gender = col_character(),
## ..   'Weight (kg)' = col_double(),
## ..   'Height (m)' = col_double(),
## ..   Max_BPM = col_double(),
## ..   Avg_BPM = col_double(),
## ..   Resting_BPM = col_double(),
## ..   'Session_Duration (hours)' = col_double(),
## ..   Calories_Burned = col_double(),
## ..   Workout_Type = col_character(),
## ..   Fat_Percentage = col_double(),
## ..   'Water_Intake (liters)' = col_double(),
## ..   'Workout_Frequency (days/week)' = col_double(),
## ..   Experience_Level = col_double(),
## ..   BMI = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

## 2. How experienced are the majority of gym members?

The purpose behind this observation is to understand the common workout program gym members prefer based on their experience level. By visualizing the data, gyms can determine a demographic for advertising their workout programs. In addition, it will encourage more activity and popularity at the gym.

- Beginners (Experience\_Level = 1) - Refers to members with little to no knowledge about exercising, however understand the health benefits of increased physical activity and are look for a way to start.
- Intermediate (Experience\_Level = 2) - Refers to members with some knowledge about the gym. This includes members who are relatively healthy. In addition, many will have an idea of their goals at the gym.
- Advanced (Experience\_Level = 3) - Refers to members who have developed an understanding of their body and know what to do to achieve their goals. In addition, most of these members tend to have greater strength and/or endurance than the average person.

```
# Creates a new column in the data frame that classifies the experience level
# into more recognizable terms
```

```
gym_levels <- gym %>%
  mutate(experience = case_when(
    Experience_Level == 1 ~ "Beginner",
    Experience_Level == 2 ~ "Intermediate",
    Experience_Level == 3 ~ "Advanced"
  ))
```

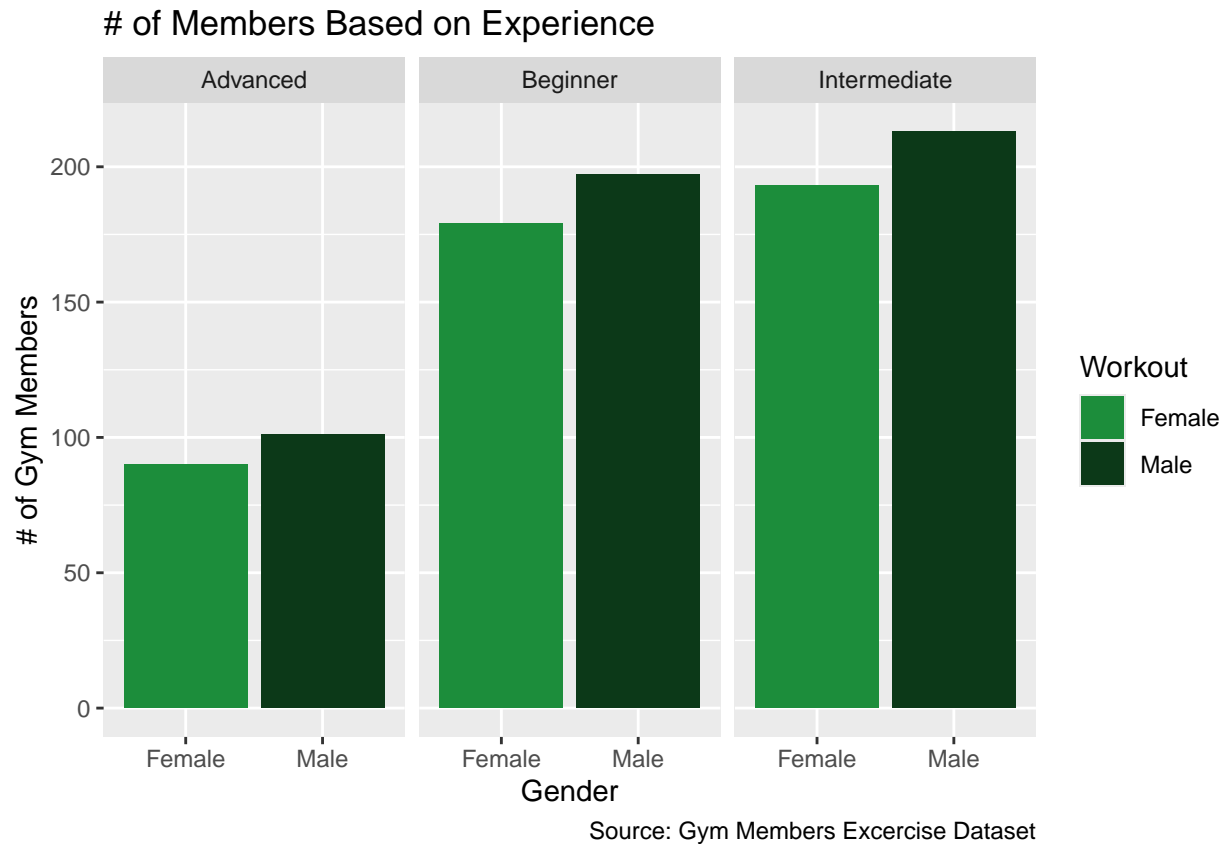
```
# Graphs the the type of workout most gym members decide to do based on
# experience level.
```

```
ggplot(data = gym_levels) +
  geom_bar(mapping = aes(x = Gender, fill = Gender)) +
  scale_fill_manual(values = c('#1c8d3b', '#0c3918'))+
```

```

facet_wrap(~experience) +
labs(title = "# of Members Based on Experience",
      caption = "Source: Gym Members Exercise Dataset") +
ylab("# of Gym Members") +
guides(fill = guide_legend(title = "Workout"))

```



### 3. What gender makes up the majority of the dataset?

At a glance, many individuals view the gym as less inclusive of women and feel uncomfortable with this negative perception. However, is that perception really true?

Using `count()`, the values that are 0 are women, while 1 are men, `n` represents the # of males/females, and `%` represents the proportion of men/women at the gym.

By using `ggplot2`, the data can be visualized.

```

# Calculates the amount of males and females in the gym
sex <- gym %>%
  count(Gender = case_when(
    Gender == "Male" ~ "Male",
    Gender == "Female" ~ "Female"
  )) %>%
  mutate(`%` = round(n/nrow(gym), 4) * 100) %>%
  mutate(raw = round(n/nrow(gym), 4))

```

```
# Prints out the data frame
sex
```

```
## # A tibble: 2 x 4
##   Gender      n    '%'   raw
##   <chr> <int> <dbl> <dbl>
## 1 Female   462  47.5  0.475
## 2 Male    511  52.5  0.525
```

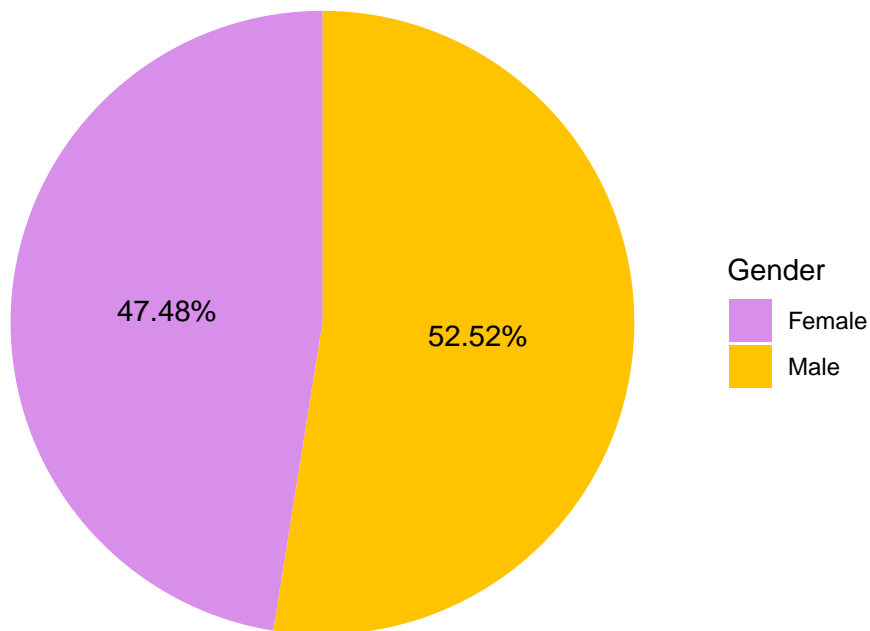
```
# Rearranges values to in dataframe
```

```
sex <- arrange(sex, desc(n))
```

```
# Visualizes the amount of men and women that attend the gym
```

```
ggplot(data = sex) +
  geom_col(mapping = aes(x="", y = raw, fill = Gender)) +
  geom_text(aes(x="", y=raw, label = paste(`%`, '%', sep = "")),
            position = position_stack(vjust = 0.5)) +
  coord_polar("y", start = 0) +
  theme_void() +
  scale_fill_manual(values = c('#d88fea', '#FFC300')) +
  labs(title = "Percentage of Men and Women at the Gym",
       caption = "Source: Gym Members Exercise Dataset")
```

Percentage of Men and Women at the Gym

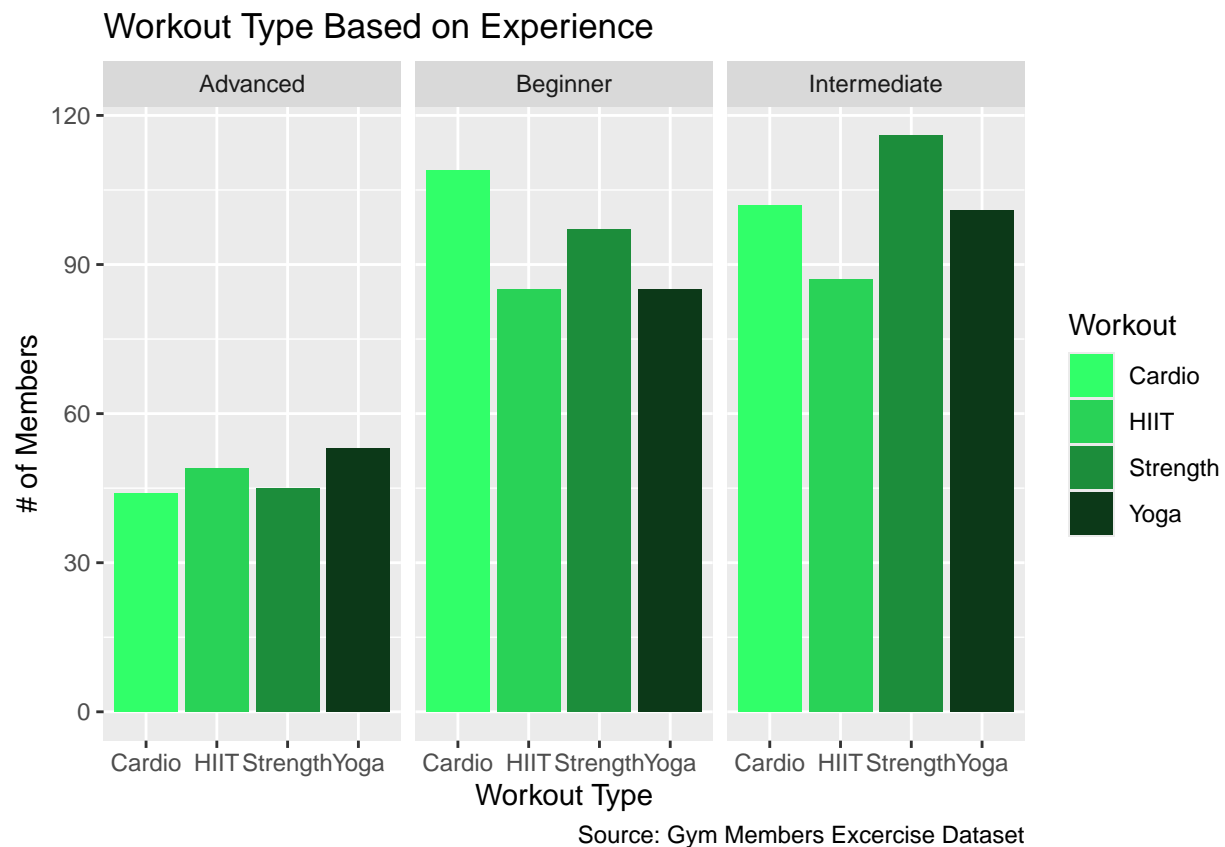


Source: Gym Members Exercise Dataset

#### 4. What is the most common workout plan for members based on experience level?

By filtering out the data set, one can see the most common type of workout plan that members gravitate to based on experience.

```
ggplot(data = gym_levels) +  
  geom_bar(mapping = aes(x = Workout_Type, fill = Workout_Type)) +  
  scale_fill_manual(values = c('#31ff69', '#29d257', '#1c8d3b', '#0c3918')) +  
  facet_grid(~experience) +  
  labs(title = "Workout Type Based on Experience",  
       caption = "Source: Gym Members Exercise Dataset")+  
  ylab("# of Members") +  
  xlab("Workout Type") +  
  guides(fill = guide_legend(title = "Workout"))
```



#### 5. Age Group in each Experience Level

By using a box whisker plot, we can see the distribution of members based on their age and workout experience. The purpose behind this plot is to see the general age range within each level of experience.

```
# Generating box-whisker plot
```

```
ggplot(data = gym_levels) +
```

```
geom_boxplot(aes(x = experience, y = `Session_Duration (hours)`,
                 fill = Gender)) +
facet_wrap(~Gender) +
guides(fill = "none") +
scale_fill_manual(values = c('#d88fea', '#FFC300')) +
labs(title = "Median Age Based on Experience and Sex",
      caption = "Source: Gym Members Exercise Dataset") +
ylab("Session Duration (hrs)") +
xlab("Experience")
```



## 6. Calories Burned by Time Spent at the Gym

Burning off calories is valuable to one's physical health because once it has been burned off, the body will eventually use its fat reserves as a source of fuel to power the bodily functions. This in turn results in a decrease in body fat.

By using `ggplot2` and `cor()`, allows us to plot the relationship between the session time at the gym to amount of calories burned.

```
# Identifying the correlation between session duration and calories burned
round(cor(gym_levels$`Session_Duration (hours)`, gym_levels$Calories_Burned,
         use = "complete.obs"), 3)
```

```
## [1] 0.908
```

```

# Initializing the values linear regression equation
activity_reg <- lm(gym_levels$Calories_Burned ~
  gym_levels$`Session_Duration (hours)` )
activity_coeff <- coefficients(activity_reg)

activity_intercept <- activity_coeff[1]
activity_slope <- activity_coeff[2]

# Scatter plot with linear regression line
ggplot(data = gym_levels) +
  geom_jitter(mapping = aes(x = `Session_Duration (hours)`,
    y = Calories_Burned, color = Gender)) +
  scale_color_manual(values = c("#bb80b0", "#c8d855")) +
  geom_abline(intercept = activity_intercept, slope = activity_slope)+
  labs(title = "Calories Burned by Time Spent",
    caption = "Source: Gym Members Exercise Dataset") +
  ylab("Calories Burned") +
  xlab("Session Duration (Hrs)")

```



## 7. Conclusion

Using the bar chart, most gym members mainly consist **Beginners** and **Intermediate** in terms of workout experience. In addition, it tells us that majority of gym members have little to no idea on what to do or what their goals may be. This is important because analyzing **Median Age Based on Experience and Sex**

suggests that men and women who are **Beginners** or **Intermediate** spend less time at the gym. To add on, the graph **Calories Burned by Time Spent at the Gym** suggests that there is a positive correlation between the amount of calories burned and time spend at the gym. This means that the amount of calories burn increase as session duration at the gym increases as well.

By taking a look at these observations, gym owners can offer classes based on **Workout Type** to encourage more time spent at the gym. The reason being is that most **Beginners** appear to higher amount of interest in **Cardio**, while **Intermediate** members prefer more **Strength** training. This tells gym owners what members will generally gravitate towards in terms of what to do at the gym. This will offer insight on how gym owners will advertise their workout programs to gym members, and could ultimately increase session duration for **Beginner** and **Intermediate** members.

Although the plots do not tell us that what variables will have a direct affect on gym activity, it instead tells us why the data appears to be what it is. This is valuable as it allows gym owners to develop marketing strategies to use to advertise. Assuming that marketing strategies are successful, it could lead to an increase in physical activity and potentially increase one's physical capabilities in the long term.

```
# According to the graph: "# of Members Based on Experience"

# Q1. What is the most common experience level among the members?
Answer_1 <- "Beginner and Intermediate Members"

# According to the graph: `Percentage of Men and Women at the Gym`

# Q2. Is there an equal amount of Men and Women at the gym?
Answer_2 <- "There is more men than women in the gym based on the data.
However, there is a small difference"

# According to the graph: `Workout Type Based on Experience`

# Q3. What is the most common workout plan for members based on
# experience level?
Answer_3 <- "For Advanced members, most do Yoga. Intermediate does Strength.
Beginners do cardio."

#According to the graph: `Median Age Based on Experience and Sex`

# Q4. Which experience level generally spends the least amount of time at the
# gym?
Answer_4 <- "Beginners tend to spend the least amount of time at the gym."
```