

# 607 Project 2 Food

Lu Beyer

## Overview

This dataset looks at food choices and preferences among 126 college students at Mercyhurst University. Upon first glance at this data, there are a lot o variables that can be looked at, but I will be focusing on gender and the variance of how both documented genders answered questions within this survey.

```
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.4.4      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

x <- read.csv("food_coded.csv")
```

I'm interested to see if either gender seems to perceive themselves are more overweight or more fit, and if that perception shows any correlation between how rewarding life feels, the students GPA, the frequency of exercise, etc., so I wanted to narrow down the data to more easily look at these variables.

```
y <- x %>%
  select(GPA, grade_level, Gender, healthy_feeling, self_perception_weight, life_rewarding, income, emp

      # comfort_food_reasons_coded, cook, cuisine, diet_current_coded, eating_out, fruit_day, veggi
```

I didn't want to rely on the codebook in order to read my data, so I decided to mutate my data so it is more easily read.

```
z <- y %>%
  # mutate(weight = as.numeric(gsub("[0-9]", "", weight))) %>%
  mutate(GPA = as.numeric(gsub("[^0-9.]", "", GPA))) %>%
  mutate(grade_level = as.character(gsub("1", "Freshman", grade_level))) %>%
  mutate(grade_level = as.character(gsub("2", "Sophomore", grade_level))) %>%
  mutate(grade_level = as.character(gsub("3", "Junior", grade_level))) %>%
  mutate(grade_level = as.character(gsub("4", "Senior", grade_level))) %>%
  mutate(Gender = as.character(gsub("1", "Female", Gender))) %>%
```

```

mutate(Gender = as.character(gsub("2", "Male", Gender))) %>%
mutate(self_perception_weight = as.numeric(gsub("[^0-9]", "", self_perception_weight))) %>%
mutate(self_perception_weight = as.character(gsub("6", "i dont think myself in these terms", self_perception_weight))) %>%
mutate(self_perception_weight = as.character(gsub("5", "overweight", self_perception_weight))) %>%
mutate(self_perception_weight = as.character(gsub("4", "slightly overweight", self_perception_weight))) %>%
mutate(self_perception_weight = as.character(gsub("3", "just right", self_perception_weight))) %>%
mutate(self_perception_weight = as.character(gsub("2", "very fit", self_perception_weight))) %>%
mutate(self_perception_weight = as.character(gsub("1", "slim", self_perception_weight))) %>%
mutate(income = as.character(gsub("^1$", "less than $15,000", income))) %>%
mutate(income = as.character(gsub("^2$", "$15,001 to $30,000", income))) %>%
mutate(income = as.character(gsub("^3$", "$30,001 to $50,000", income))) %>%
mutate(income = as.character(gsub("^4$", "$50,001 to $70,000", income))) %>%
mutate(income = as.character(gsub("^5$", "$70,001 to $100,000", income))) %>%
mutate(income = as.character(gsub("^6$", "higher than $100,000", income))) %>%
mutate(life_rewarding = as.numeric(gsub("[^0-9]", "", life_rewarding))) %>%
mutate(employment = as.numeric(gsub("[^0-9]", "", employment))) %>%
mutate(employment = as.character(gsub("1", "yes full time", employment))) %>%
mutate(employment = as.character(gsub("2", "yes part time", employment))) %>%
mutate(employment = as.character(gsub("3", "no", employment))) %>%
mutate(employment = as.character(gsub("4", "other", employment))) %>%
mutate(exercise = as.numeric(gsub("[^0-9]", "", exercise))) %>%
mutate(exercise = as.character(gsub("1", "Everyday", exercise))) %>%
mutate(exercise = as.character(gsub("2", "Twice or three times per week", exercise))) %>%
mutate(exercise = as.character(gsub("3", "Once a week", exercise))) %>%
mutate(exercise = as.character(gsub("4", "Sometimes", exercise))) %>%
mutate(exercise = as.character(gsub("5", "Never", exercise))) %>%
mutate(life_rewarding_bin = cut(life_rewarding, breaks = c(0,2.5,5,7.5,10), labels = c("0-2.5", "2.5-5", "5-7.5", "7.5-10"))) %>%
mutate(GPA_bin = cut(GPA, breaks = c(2, 2.5, 3, 3.5, 4), labels = c("2-2.5", "2.5-3", "3-3.5", "3.5-4"))) %>%

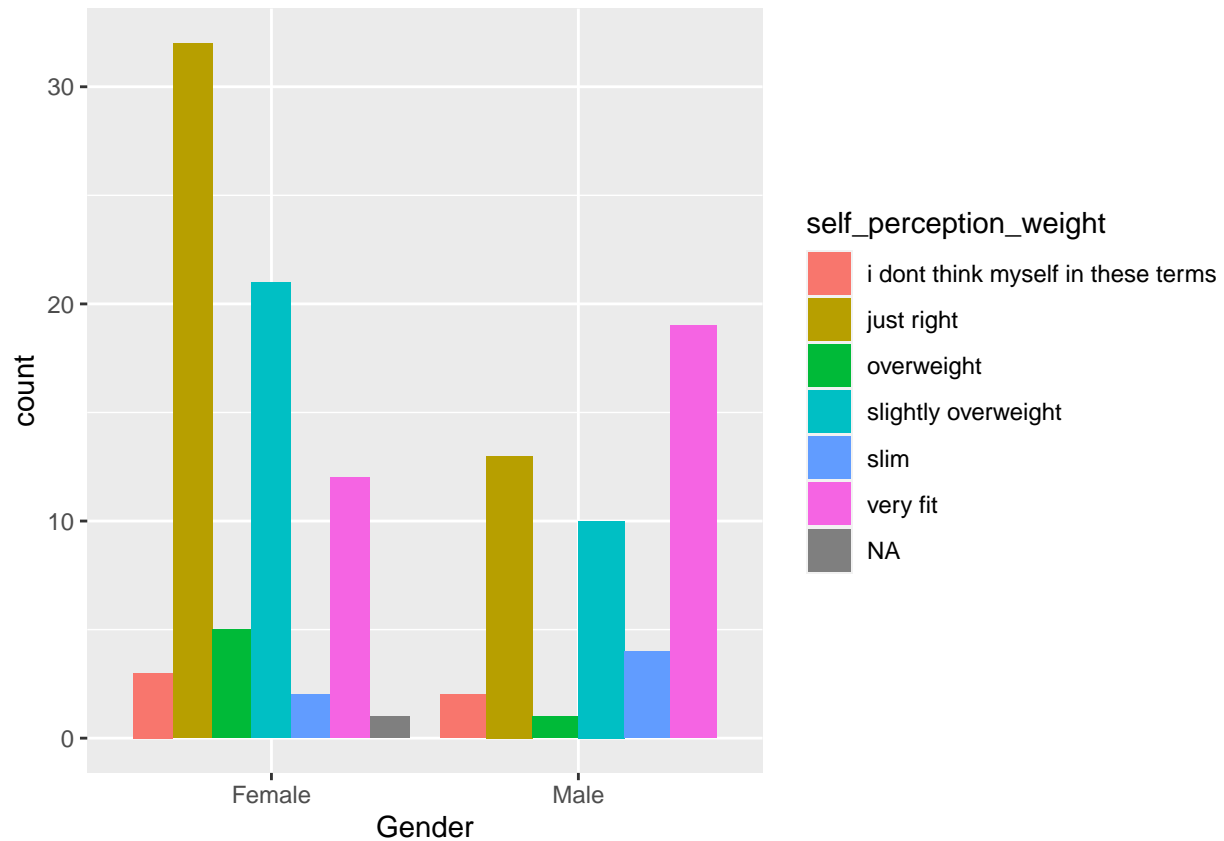
```

I wanted to create a visualization to show the self perception of weight between males and females to more easily see how that data is split.

```

ggplot(data = z, aes(x = Gender, fill = self_perception_weight)) +
  geom_bar(position = "dodge")

```

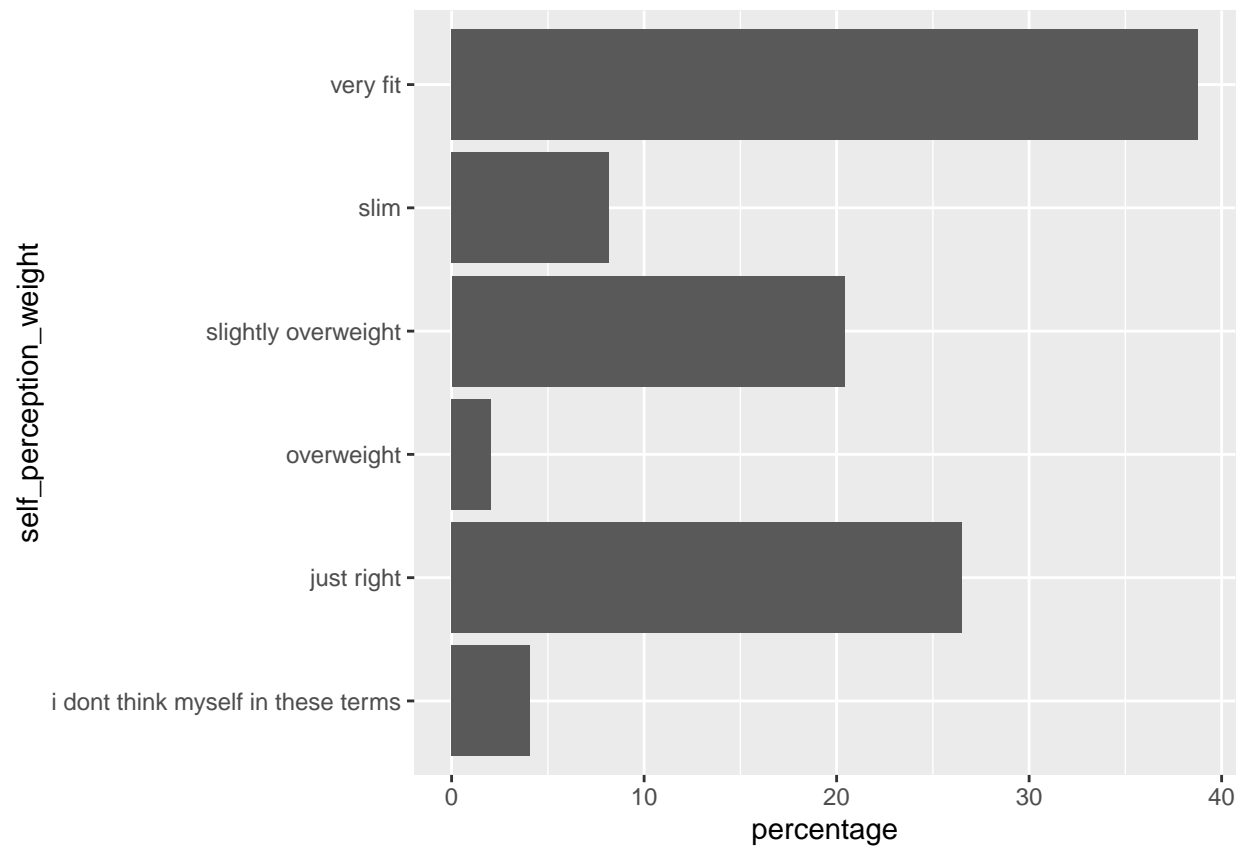


Seeing this data side by side, I wanted to see how the percentages of each category by gender to see how those numbers differed.

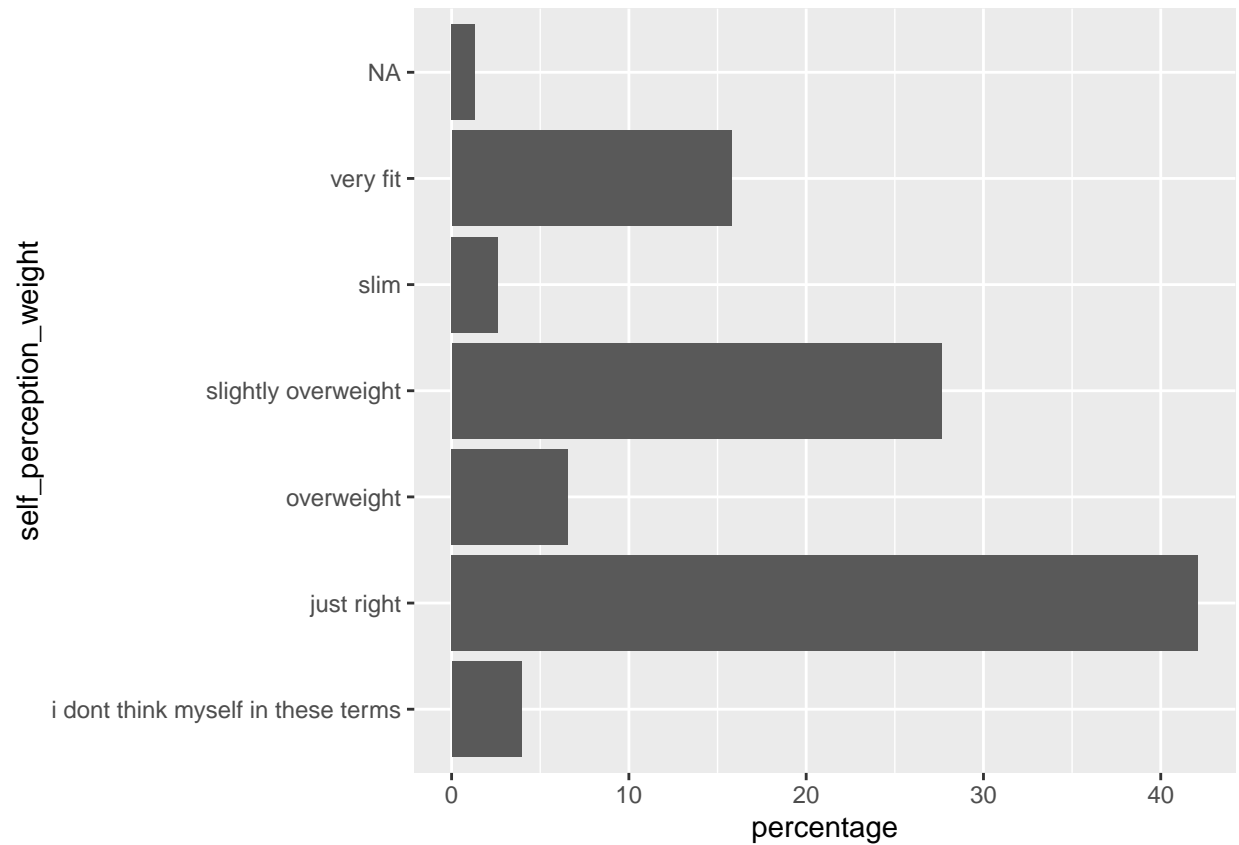
```
men <- z %>%
  filter(Gender == "Male") %>%
  group_by(self_perception_weight) %>%
  summarise(count = n()) %>%
  mutate(percentage = (count / sum(count))*100)
```

```
women <- z %>%
  filter(Gender == "Female") %>%
  group_by(self_perception_weight) %>%
  summarise(count = n()) %>%
  mutate(percentage = (count / sum(count))*100)
```

```
ggplot(data = men, aes(x = percentage , y = self_perception_weight)) +
  geom_bar(stat = "identity")
```

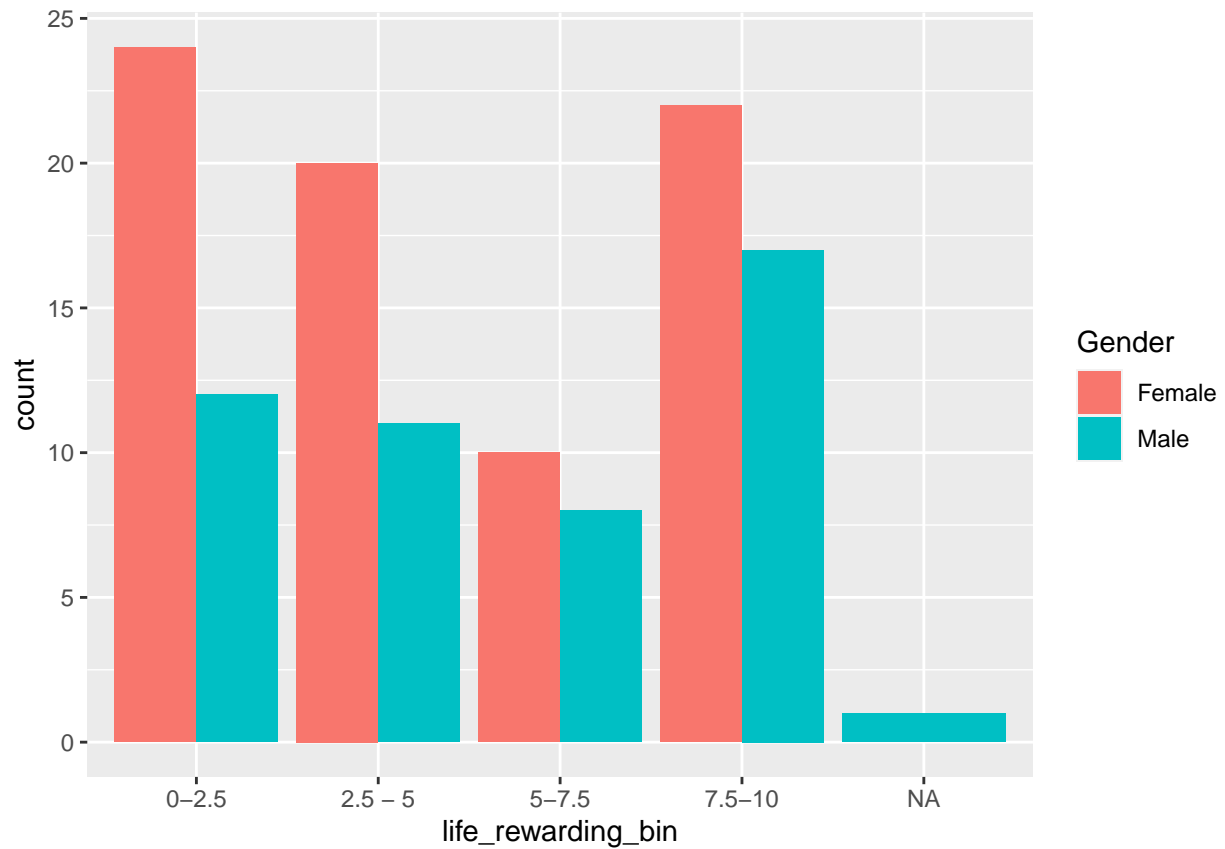


```
ggplot(data = women, aes(x = percentage , y = self_perception_weight)) +  
  geom_bar(stat = "identity")
```



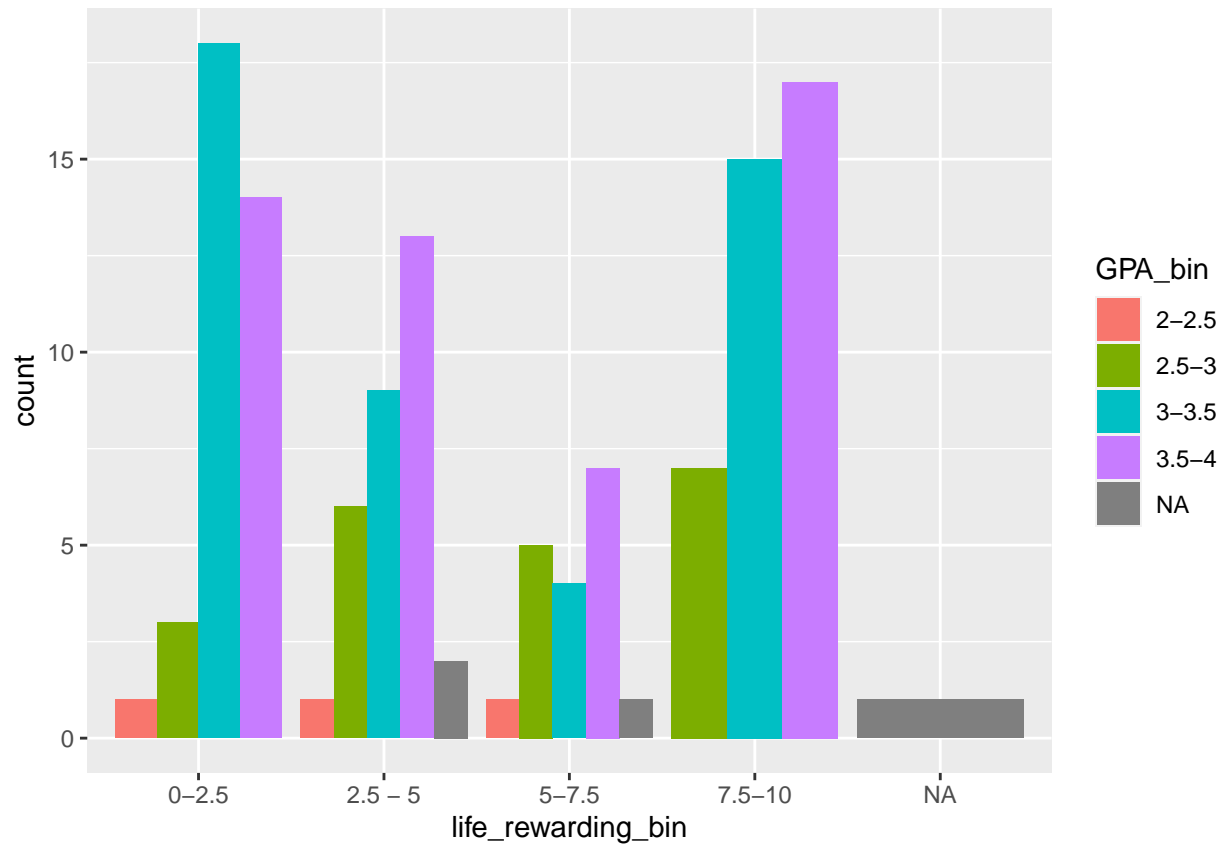
I also wanted to see how males and females viewed how rewarding their lives are.

```
ggplot(data = z, aes(x = life_rewarding_bin, fill = Gender)) +  
  geom_bar(position = "dodge")
```



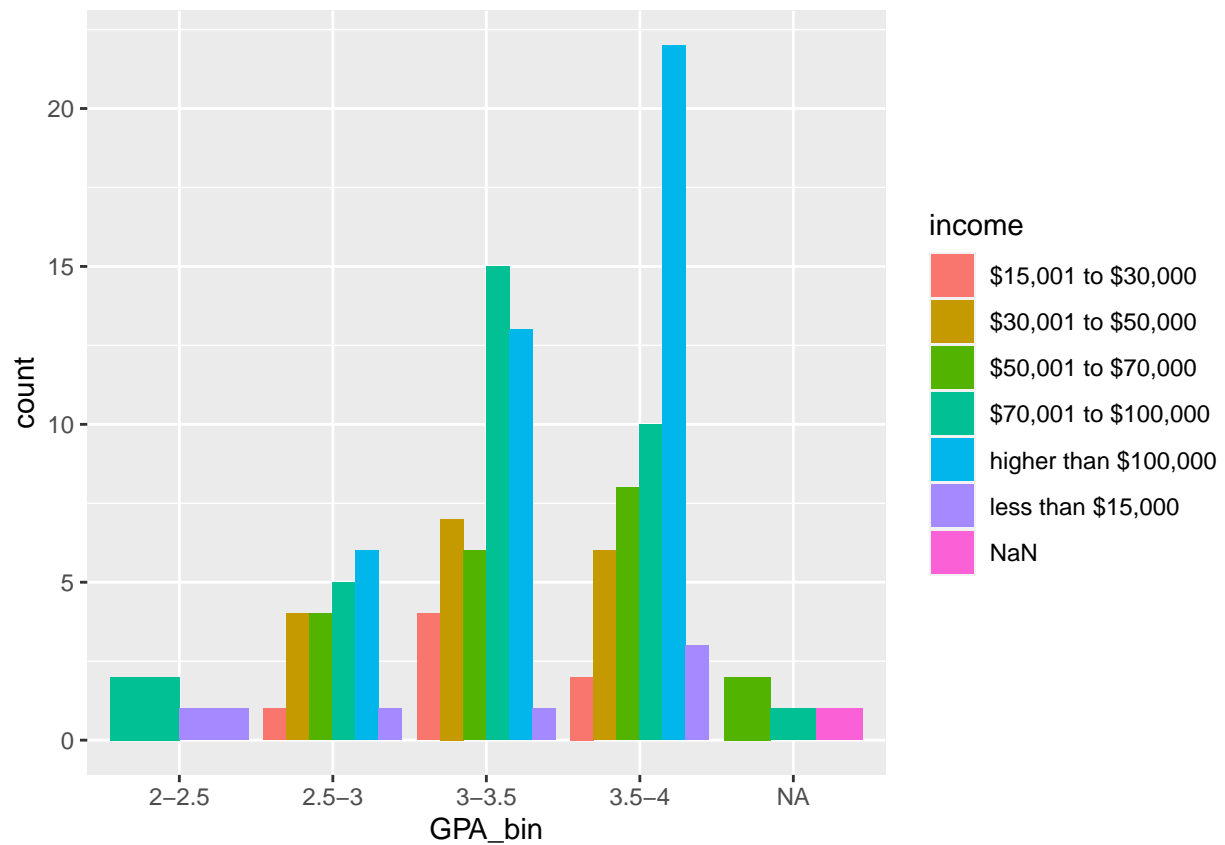
I then wanted to see how a students GPA compares to how rewarding they say their life feels. I oranized their GPA and their ratings for how rewarding their lives are into bins to consolidate the data.

```
ggplot(data = z, aes(x = life_rewarding_bin, fill = GPA_bin)) +  
  geom_bar(position = "dodge")
```



I created visuals to compare reported income and GPA of the surveyed students.

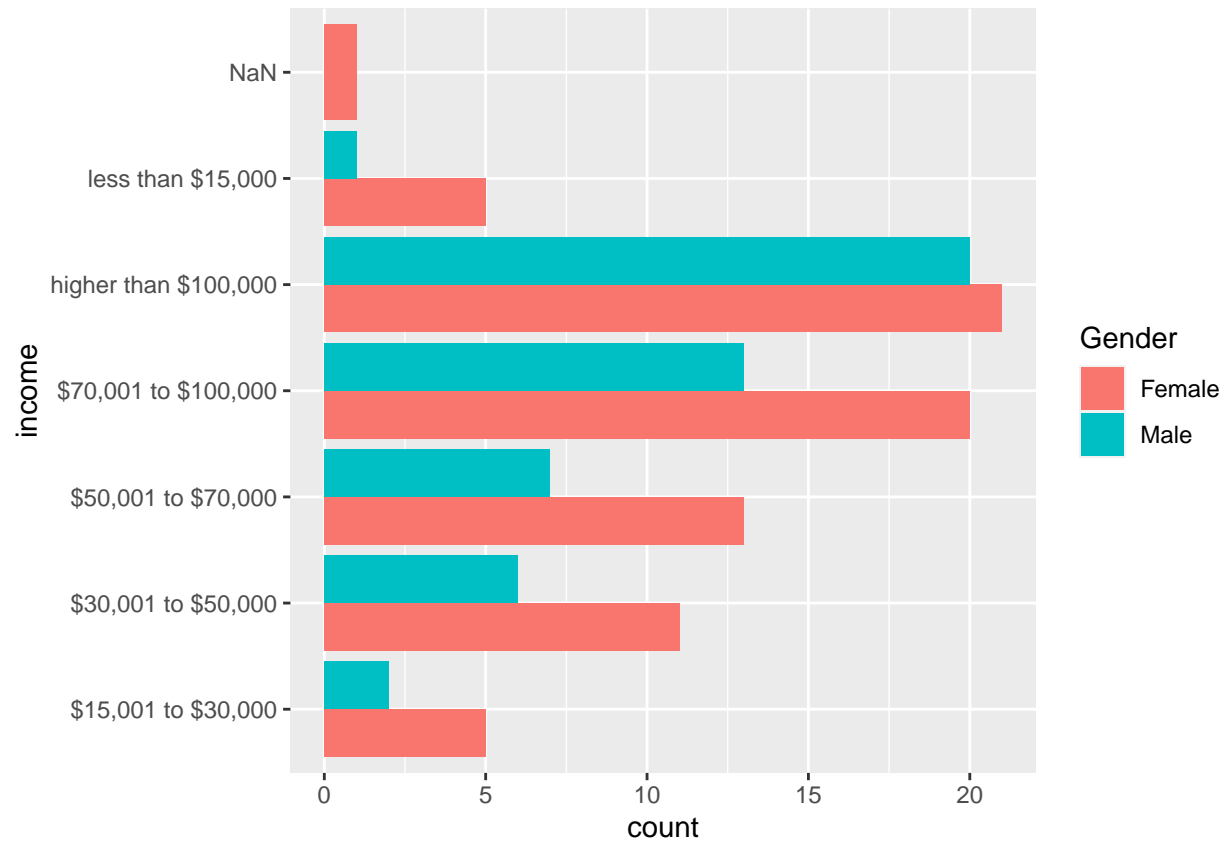
```
ggplot(data = z, aes(x = GPA_bin, fill = income)) +  
  geom_bar(position = "dodge")
```



I wanted to compare the reported income against the gender of the surveyed students.

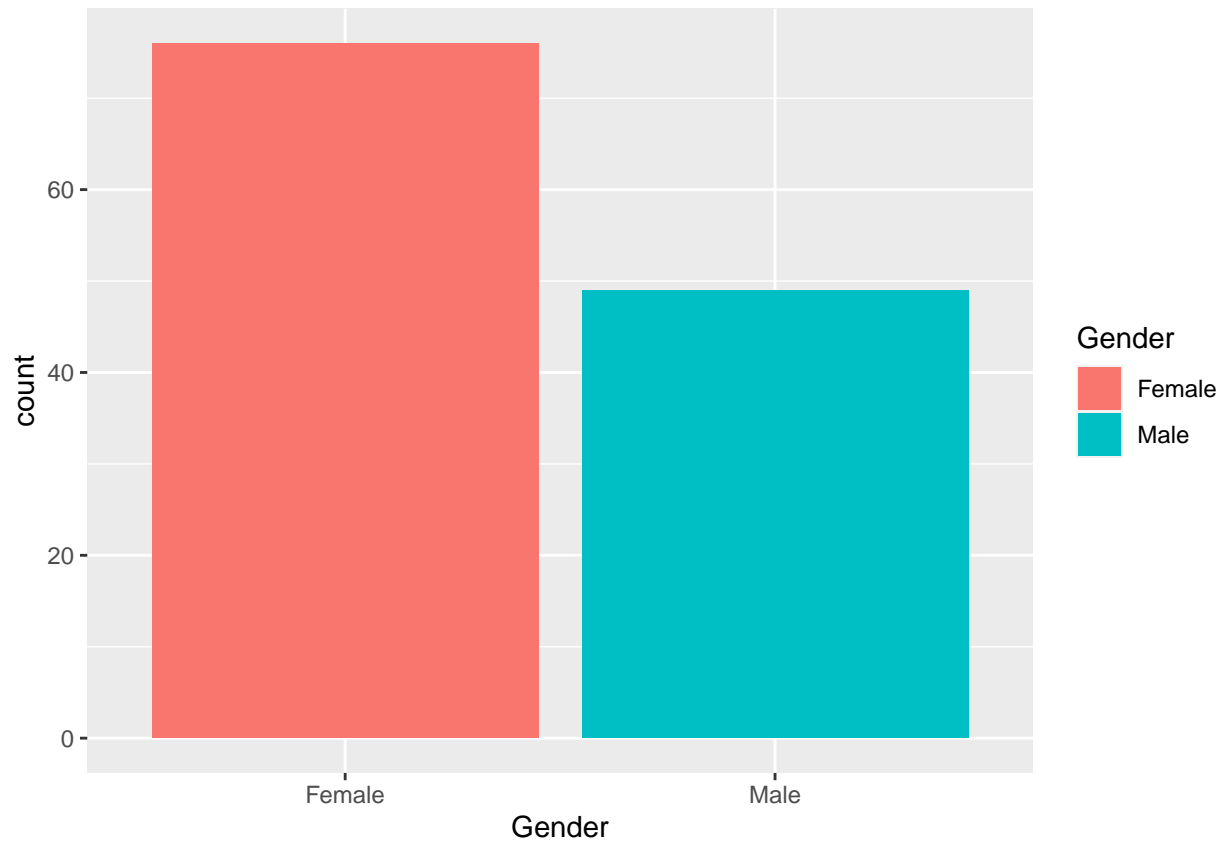
```
ggplot(data = z, aes(x = income, fill = Gender)) +
  geom_bar(position = "dodge") + coord_flip()
```





I felt that, in general, the counts for females were higher than males so I wanted to see how many males vs females were surveyed.

```
ggplot(data = z, aes(x = Gender, fill = Gender)) +  
  geom_bar()
```

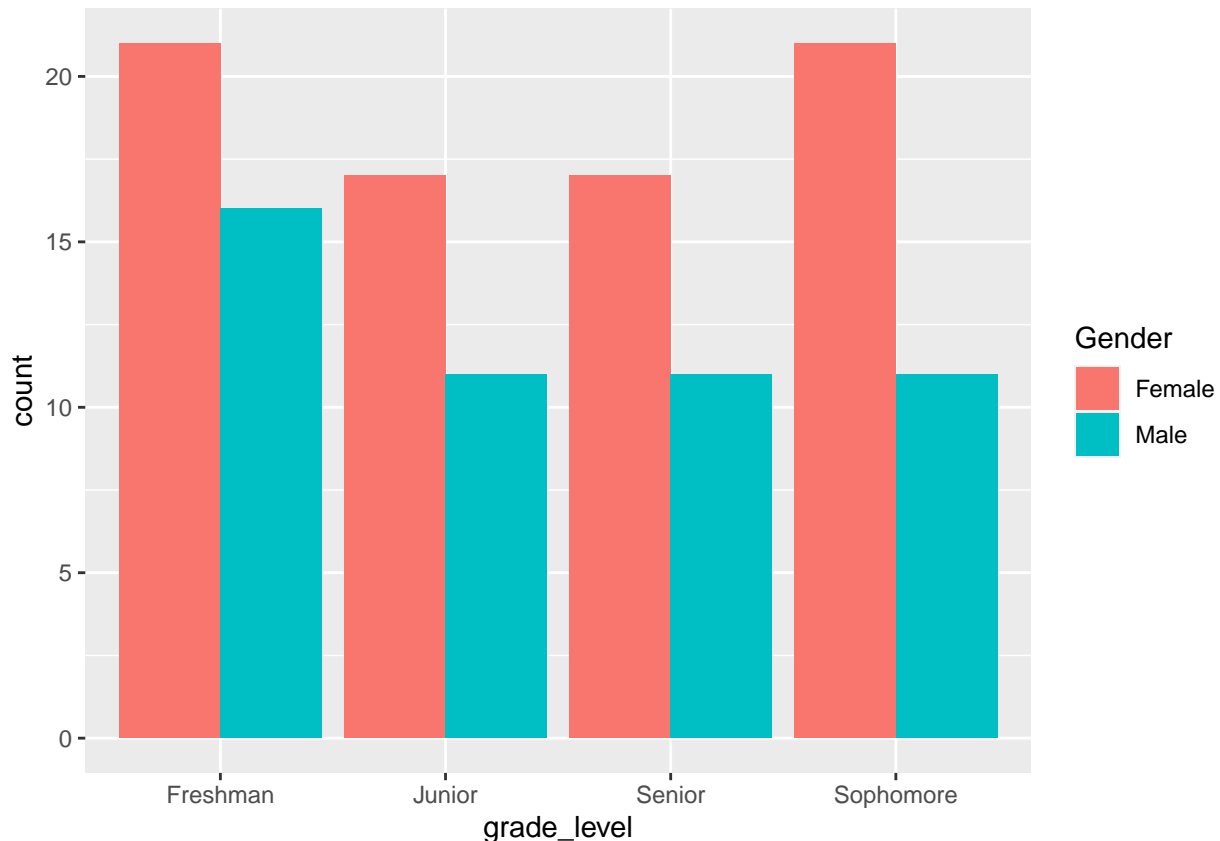


```
table(z$Gender)
```

```
##  
## Female    Male  
##      76     49
```

Finally, I wanted to look at the counts of students surveyed across their grade level.

```
ggplot(data = z, aes(x = grade_level, fill = Gender)) +  
  geom_bar(position = "dodge")
```



#### # Conclusion

Initially, I thought the data would be very clear to analyze based on gender, but because the gender ratio skews 60/40 in favor of women, all of the data generally had higher numbers for women. It was harder to determine trends based on gender because we would assume that women would generally have higher counts in any given category. Something I found interesting while exploring the data, it was much more common for men to perceive themselves as very fit than women were with 38.7% of men viewing themselves as very fit and 15.7% of women. On the other hand, 34.1% of women and 22.8% of men perceive themselves as overweight or slightly overweight. 46% of women surveyed perceived themselves as “just right” compared to 26% of men. This was the most interesting comparison of genders in the data for me. The variables are somewhat difficult to quantify because it’s more emotional and subjective—is it a more positive view to be “just right” vs “very fit”, who among those who answered “overweight” views that as negative vs positive? I thought this would give me some information on how the students surveyed perceived themselves, but once I actually went through the data and thought about it, I didn’t really have much more clarity.

Comparing the students perception of how rewarding their life was and their GPA, my assumption was that the more rewarding the students life felt, the higher their GPA would be. I assumed students who felt their lives were more rewarding would have more capacity for school and likely be able to do better. That largely wasn’t the case; students who felt their lives were the least rewarding based on the given scale had the highest count of students with a GPA of 3-3.5 and the second highest count of students with a GPA of 3.5-4. Students with the highest rating of life fulfillment had no counts of students with a GPA of 2-2.5, but the highest count of students with a 2.5-3. None of the data I chose to analyze supported any of my initial thoughts before organizing this data, but after organizing, I didn’t really gain any clarity either.

The most shocking part of this data is the amount of college students claiming they make more than \$100,000. This did make me somewhat skeptical about the data set as a whole, but maybe college students make more money than I was aware of.