

# **Impact of Displaying Calorie Information on Menus**

Jay Chaudhary, Hector He, Kwan Wing Tuet, Yuesen (Sam) Zhang, Yu Ting Hung

## **Background/Introduction**

Obesity and overconsumption of high-calorie foods are prevalent issues in the United States. To address these concerns, policies and mandates have been implemented to require food manufacturers and retailers to label calorie information on their products, allowing consumers to make informed choices regarding their calorie intake. Additionally, in 2018 the FDA required food establishments with 20 or more locations to display Calorie information on their menus. Calorie information on restaurant menus is also becoming more widespread, even mandatory for all establishments in certain states, to help inform consumers and ultimately reduce the consumption of calorie-excessive foods. The purpose of this study is to determine whether the presence of calorie information on a restaurant's menu affects consumer food choices.

## **Research Question**

Does displaying calorie information on a restaurant's menu affect consumer food choices?

## **Hypothesis**

We hypothesize that including calorie information on restaurant menus will lead consumers to choose lower-calorie food options.

## **Treatment & Experimental Design**

This study employed a within-subjects experimental design to investigate the impact of calorie menu labeling on food choices at a restaurant. The participants in this study ranged in age from 18 to 73 and were recruited from the local community. A diverse group of participants was recruited in terms of gender and occupation, with a total of 110 participants recruited for this study.

The survey began by collecting basic information from each participant, including age, gender, dietary preferences or restrictions, health conditions that affected food choices, and how often they ate out per week. Participants were then presented with a restaurant menu and asked to select the food items they would like to order. After making their initial selections, a second menu was shown that displayed the calorie count for each food item. Participants were asked to make their food choices again and explain whether or not they changed their selection after seeing the calorie information.

To assess the impact of calorie labeling, participants' food choices before and after the treatment were compared. Additional information was gathered to gauge participants' attitudes towards calorie information, like whether they typically considered it when making food choices, whether they were currently on a diet, and their estimated daily calorie intake and activity levels.

The data collected from this study was organized into a table with columns for experiment subjects, including their age, gender, as well as the number of participants in total. The table also included columns for participants' food choices before and after the treatment, as well as their attitudes towards calorie information. By using a within-subjects design, individual differences in food preferences were controlled for, providing a more accurate assessment of the impact of calorie labeling on food choices.

### **Target Outcome**

We measured the calorie count of the items chosen before and after presenting the treatment menu. The control menu was shown first without calorie information, and the treatment menu was shown after with calorie information. We used a paired T-test to measure the average treatment effect (ATE) and conditional average treatment effect (CATE) based on gender, health conscientiousness, and physical activity levels. We also used regression analysis to validate our results and provide further insights.

## **Data Analysis**

### **Data Collection**

We collected data through a Qualtrics survey, which included demographic information, dietary preferences and restrictions, health conditions affecting food choices, and frequency of eating out. Each participant selected their food choices from the two menus presented, and we asked if the calorie information influenced their decision. We also gathered additional information to understand participants' attitudes toward calorie information, such as whether they typically consider it when making food choices, whether they are currently on a diet, and their estimated daily calorie intake and activity levels.

### **Data Cleaning & Preparation**

For Data Cleaning and Preparation, our initial step was to filter and extract relevant data from the raw excel table. We retained all participants' responses to the questions and the "Duration (in seconds)" column, while removing irrelevant columns such as "StartDate", "EndDate", and "ResponseId". Additionally, we excluded all participants who indicated that they did not want to order anything to prevent dilution of the treatment effect.

We encountered anomalies in the "weight" column, likely due to participants' different assumptions about weight units. To address this, we adjusted the anomalies to unit pounds. Similarly, the calorie intake was originally presented as a range, so we took the average for different ranges and used "1200" for the "< 1,600" option and "3400" for the "3000+" option.

To measure the impact of our calorie-labeling menu on food choice decision-making, we recorded participants who switched between two menus and transformed their ordered items from the two menus into actual calorie numbers. The code for this process can be found in the appendix.

For EDA, we examined people who don't exercise and conducted Paired T-test Analysis. In the T-test on the entire population, we tested whether calorie posting on the menu truly has an impact on the total calories resulting from two orders. Our test results indicate that the total calories from the two menus differ for the calorie-matter group of people but not for the calorie-does-not-matter group (health conscientiousness).

Furthermore, we conducted a T-test conditioned on gender and found that the calorie-labeling menu has a significant impact on the calories consumed by females from food items they ordered on the regular menu and the calorie-labeling menu. We also conducted T-tests on exercise and did regression analysis using covariates to increase precision. Lastly, we used a power test to ensure that our sample size is sufficient for statistical significance.

### **Within-study Design**

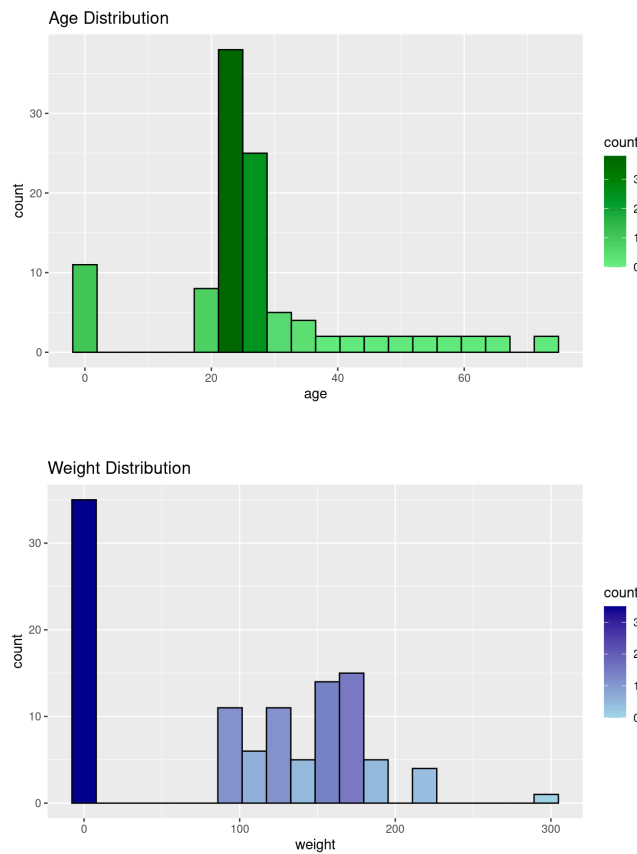
Since the study design was within-subjects, where each participant was exposed to both conditions (with and without the intervention), blocking, randomization, or clustering were not necessary. Within-subject designs aim to minimize the impact of individual differences and increase statistical power by using each participant as their own control, reducing variability and increasing sensitivity to detect treatment effects.

In a within-subject design, the order in which conditions are presented can be counterbalanced to control for order effects. Counterbalancing involves presenting the conditions in different orders across participants to ensure that any potential effects of order are distributed evenly across the groups. This approach can help increase the validity and reliability of the study's results.

Overall, within-subject designs can offer several advantages, including increased power, reduced variability, and better control of potential confounding variables. By using each participant as their own control, within-subject designs can be a useful approach for investigating treatment effects and minimizing potential sources of bias.

## Sample size

A total of 110 participants were included in the study.

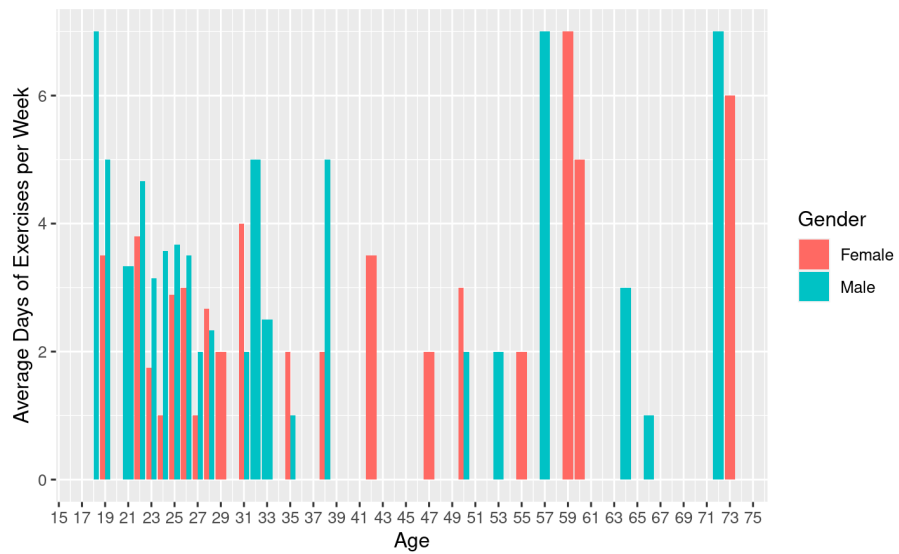
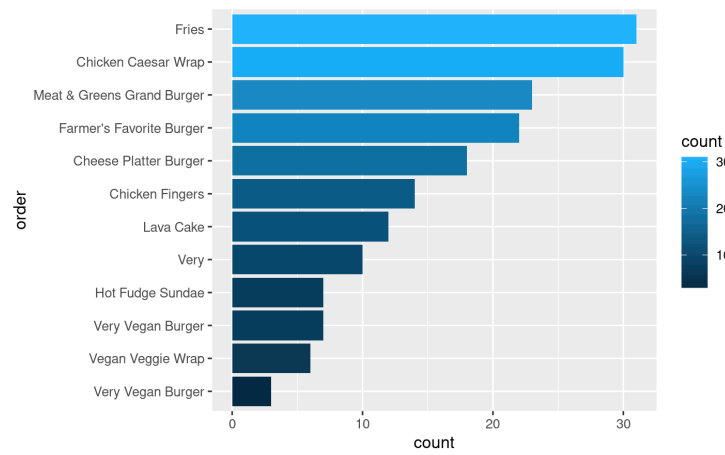
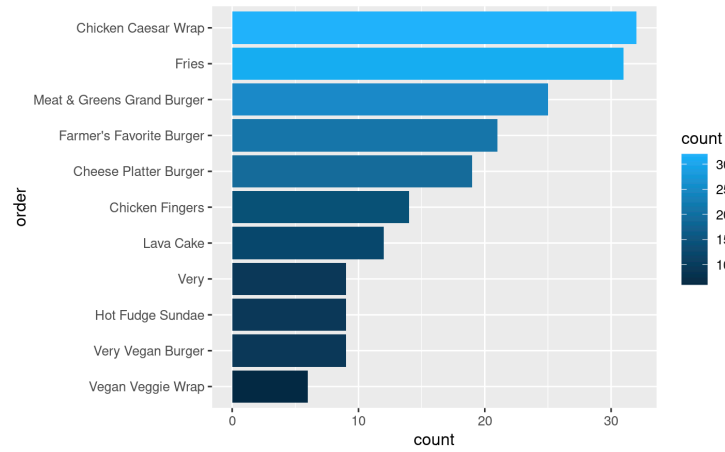


## EDA (Please see page 5)

1. What kind of foods people ordered? (p5 top, middle graphs)
2. Average number of exercise (days per week) in Age and Gender (p5 bottom graph)

## **Results**

The study found that displaying the calorie count of food items on a menu led to customers ordering items with fewer calories, indicating a significant average treatment effect of -37 calories. This suggests that providing calorie information can be an effective strategy for encouraging healthier food choices among consumers.



Moreover, the results showed that women tended to switch their order more often than men and ordered fewer calories, resulting in a C.A.T.E. of -47 calories for women. This may indicate that women are more health-conscious or more likely to be influenced by calorie information than men. Similarly, health-conscious individuals also switched their order and ordered fewer calories, leading to a C.A.T.E. of -174 calories for individuals who reported being health-conscious. This finding suggests that individuals who are more concerned about their health and nutrition are more likely to be influenced by calorie information.

## **Business Implications**

Restaurants could benefit from displaying calorie information on their menus. Lower calorie foods tend to be more cost-effective, and promoting them through menu labeling could increase their sales and profits. Restaurants could also use calorie information as a marketing tool to attract health-conscious customers and differentiate themselves from competitors.

## **Limitations & Future Scope**

There are limitations to this study that need to be addressed. Firstly, the sample size was relatively small, and the participants were from a similar demographic. This may limit the generalizability of the findings to other populations. Secondly, the menu used in this study may not be appealing to all food preferences, and the impact of calorie information may vary depending on the type of food and cuisine. Future studies could explore the impact of calorie information on various types of foods and cuisines. Additionally, future research could explore the long-term impact of calorie information on the overall health and well-being of consumers. Lastly, displaying nutritional information like protein, fat, etc. on restaurant menus could be explored.

## **Conclusion**

In conclusion, this study aimed to investigate the impact of displaying calorie information on restaurant menus on consumer food choices. The results showed that providing calorie information on a menu led to customers choosing items with fewer calories, indicating a significant treatment effect. Furthermore, the study found that women and health-conscious individuals were more likely to be influenced by calorie information than men, indicating a conditional average treatment effect based on gender and health conscientiousness.

The findings of this study have significant implications for the food industry and public health. Restaurants could use calorie information as a marketing tool to attract health-conscious customers and differentiate themselves from competitors. Lower calorie foods tend to be more cost-effective, and promoting them through menu labeling could increase their sales and

profits. At the same time, consumers can make informed choices regarding their calorie intake and reduce the risk of overconsumption of high-calorie foods.

However, the study has certain limitations that need to be considered. The sample size was relatively small and may not be representative of the broader population. Additionally, the menu used in this study may not be appealing to all food preferences, and the impact of calorie information may vary depending on the type of food and cuisine. Future research could explore the impact of calorie information on various types of foods and cuisines and the long-term impact on the overall health and well-being of consumers.

Overall, the findings of this study suggest that displaying calorie information on restaurant menus could be an effective strategy for encouraging healthier food choices among consumers.

## Appendix - Code

```
# load data  
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

```
filter, lag
```

The following objects are masked from 'package:base':

```
intersect, setdiff, setequal, union
```

```
library(readxl)  
library(data.table)
```

Attaching package: 'data.table'

The following objects are masked from 'package:dplyr':

```
between, first, last
```

```

library(fixest)
library(broom)
library(pwr)
library(ggplot2)
library(modelsummary)

survey <- read_excel('BA 830 Final Project_March 14, 2023_19.46.xlsx', sheet = 1)

# rename the column
survey <- survey %>%
  select(`Duration (in seconds)`, Q1, Q2, Q3, Q4, Q5,
        Q6, Q7, Q8, Q9, Q10, Q11, Q12, Q13, Q14, Q15, Q19) %>%
  rename(duration = `Duration (in seconds)`,
        consent = Q1, age_consent = Q2, age = Q3, gender = Q4,
        dietary_restriction = Q5, medical_cond = Q6,
        eat_out = Q7, order_1 = Q8, order_2 = Q9, reason = Q10,
        calorie_consideration = Q11, calorie_intake = Q12,
        diet = Q13, physical_active = Q14, num_exercise = Q15,
        weight = Q19)

survey <- survey[-1,]

# change data type
survey <- survey %>%
  mutate(duration = as.numeric(duration),
        age = as.numeric(age),
        eat_out = as.numeric(eat_out),
        num_exercise = as.numeric(num_exercise),
        weight = as.numeric(weight),
        gender = as.factor(gender),
        consent = as.factor(consent),
        age_consent = as.factor(age_consent),
        diet = as.factor(diet),
        physical_active = as.factor(physical_active),
        calorie_consideration = as.factor(calorie_consideration),
        medical_cond = as.factor(medical_cond),
        dietary_restriction = as.factor(dietary_restriction),
        calorie_intake = as.factor(calorie_intake))

```



```

survey <- survey %>%
  filter(consent != 'No' & age_consent != 'No') %>%
  filter_at(vars(order_1, order_2), all_vars(!is.na(.)))

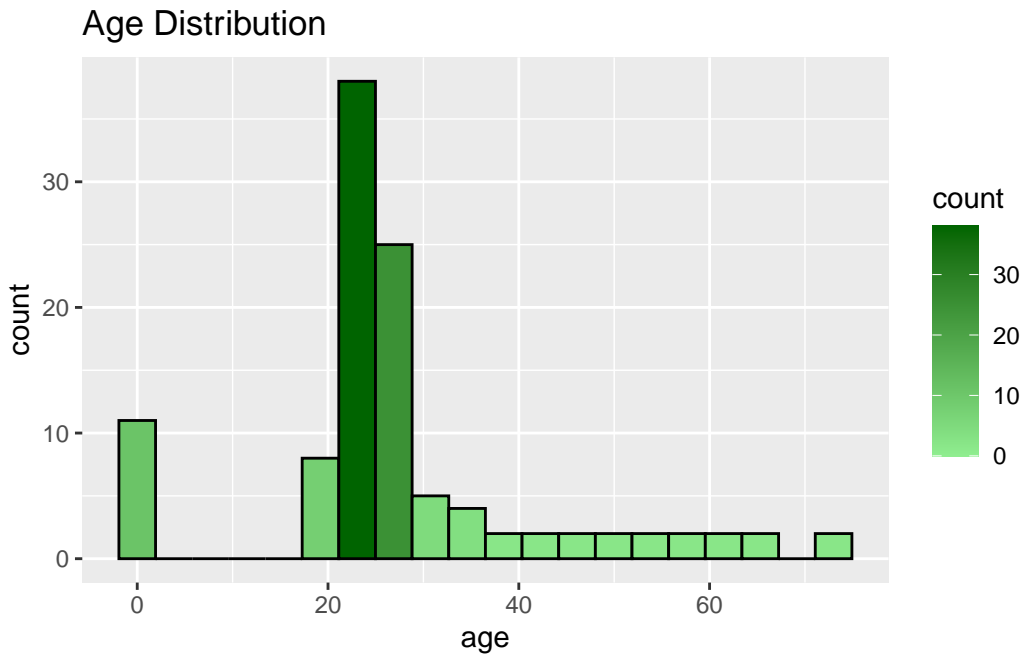
# change kg to pound
survey <- survey %>%
  mutate(weight = case_when(gender == "Male" & weight < 100 ~ weight*2.2, gender == "Female" ~ weight*2.2, TRUE ~ 0))

# change caalorie intake range into average calorie intake
survey <- survey %>%
  mutate(calorie_intake = case_when(
    calorie_intake == "2,500 - 2,999" ~ 2750,
    calorie_intake == "< 1,600" ~ 1200,
    calorie_intake == "3,000+" ~ 3400,
    calorie_intake == "2,000 - 2,499" ~ 2250,
    calorie_intake == "1,600 - 1,999" ~ 1800,
    TRUE ~ 0))

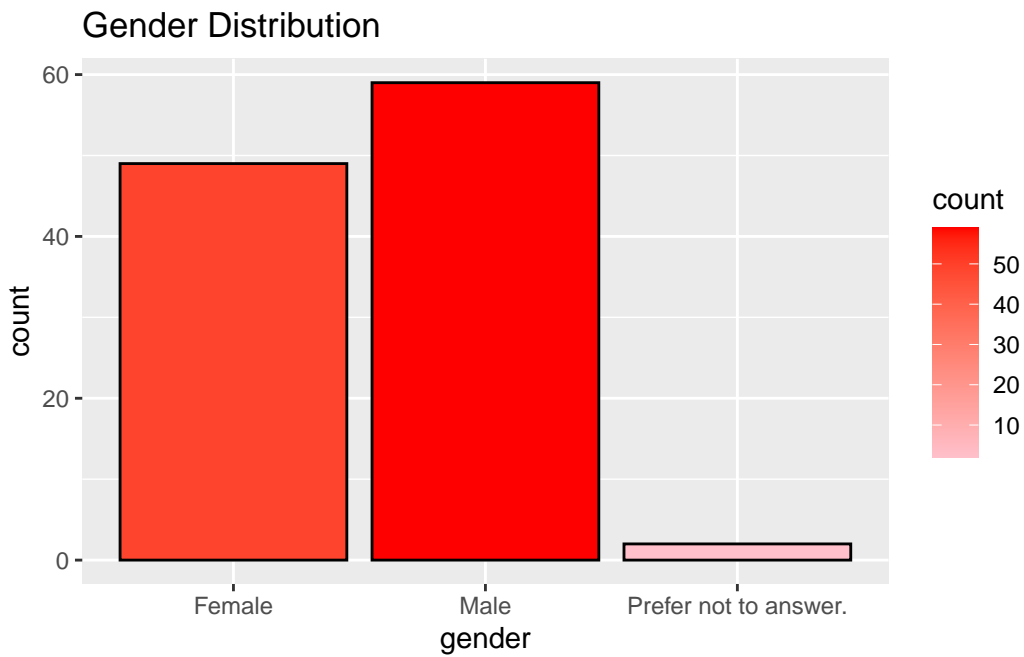
ggplot(survey, aes(x = age, fill = after_stat(count))) +
  geom_histogram(color = "black", bins = 20) +
  scale_fill_gradient(low = "lightgreen", high = 'darkgreen') +
  labs(title = 'Age Distribution')

```

Warning: Removed 1 rows containing non-finite values (`stat\_bin()`).

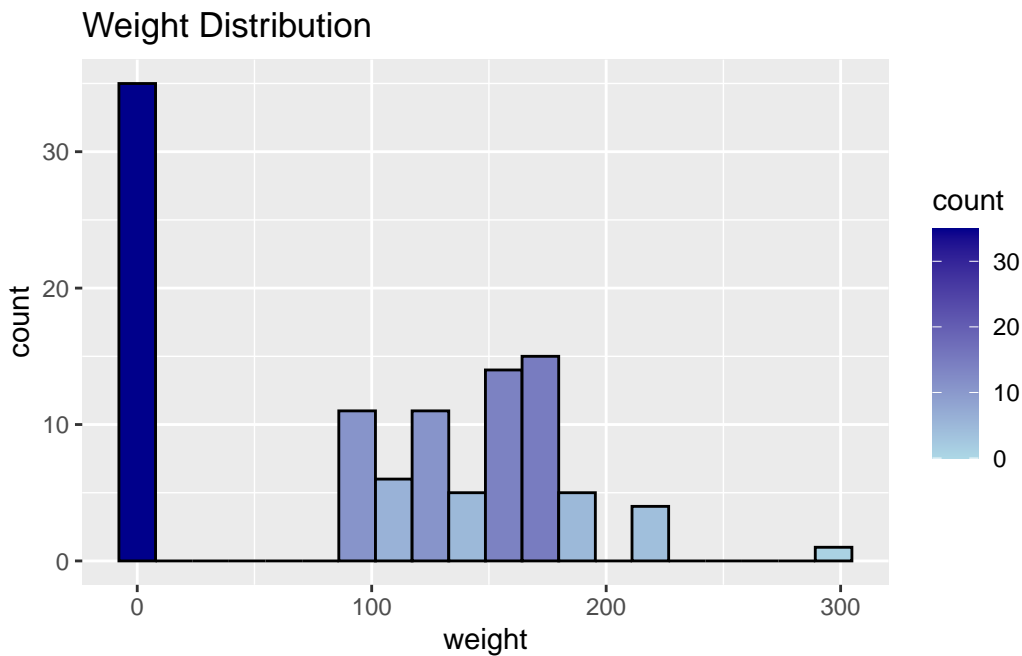


```
ggplot(survey, aes(x = gender, fill = after_stat(count))) +  
  geom_bar(color = "black") +  
  scale_fill_gradient(low = "pink", high = "red") +  
  labs(title = "Gender Distribution")
```



```
ggplot(survey, aes(x = weight, fill = after_stat(count))) +
  geom_histogram(color = 'black', bins = 20) +
  scale_fill_gradient(low = 'lightblue', high = 'darkblue') +
  labs(title = 'Weight Distribution')
```

Warning: Removed 3 rows containing non-finite values (`stat\_bin()`).



## EDA

```
calories_table <- tribble(
  ~entree, ~calorie,
  "Very Very Vegan Burger", 750,
  "Farmer's Favorite Burger", 800,
  "Cheese Platter Burger", 1020,
  "Meat & Greens Grand Burger", 970,
  "Vegan Veggie Wrap", 650,
  "Chicken Caesar Wrap", 750,
  "Fries", 350,
  "Chicken Fingers", 400,
  "Hot Fudge Sundae", 550,
  "Lava Cake", 580
```

```
)

entree_info <- calories_table$entree
calories_info <- calories_table$calorie
```

**How long does it take to complete our survey? 5.22 minutes averagely**

```
# get average duration time to complete our survey
survey %>%
  summarize(mean_duration = mean(duration, na.rm = TRUE)/60) %>%
  pull(mean_duration)
```

```
[1] 5.221061
```

**How many people change their food choice after seeing the calorie menu? 21 participants.**

```
# check if people change their food choices
survey <- survey %>%
  mutate(switch = case_when(
    stringr::word(order_2, 1, 3) == "Keep My Order" |
    gsub("\\s*\\(.*?\\)", "", order_2) == order_1 ~ 0,
    TRUE ~ 1))

sum(survey$switch == 1)
```

```
[1] 21
```

**What kind of foods people ordered?**

```
get_entrees <- function(x) {
  food <- gsub(" \\$\\d+(\\.\\d{2})?", "", x)
  food <- stringr::str_split(food, ",")[[1]]
  return(food)
}

get_entrees_from_order_2 <- function(x) {
  food <- strsplit(x, ",")[[1]]
  food <- gsub("\\(.*?\\)", "", food)
```

```

    food <- trimws(gsub("\\$.*", "", food))
    food <- unlist(strsplit(food, ", "))
    return(food)
}

find_calories <- function(z) {
  x <- entree_info
  y <- calories_info
  names(y) <- x
  entree_calories <- y[z]
  return(sum(entree_calories))
}

#survey %>%
# group_by(gender, switch) %>%
# tally(sort = TRUE, name = 'count')

food_count_1 <- function(x) {
  orders <- c()
  all_orders <- x
  for (i in x) {
    entrees <- get_entrees(i)
    for (j in entrees) {
      orders <- append(orders, j)
    }
  }
  return(orders)
}

food_count_2 <- function(x, y) { # x is order_1, y is order_2
  orders <- c()
  for (i in 1:length(x)) {
    entrees_1 <- get_entrees(x[i])
    entrees_2 <- get_entrees_from_order_2(y[i])
    if (stringr::word(entrees_2[1], 1, 1) == "Keep") {
      for (j in entrees_1) {
        orders <- append(orders, j)
      }
    } else {
      for (k in entrees_2) {
        orders <- append(orders, k)
      }
    }
  }
}

```

```

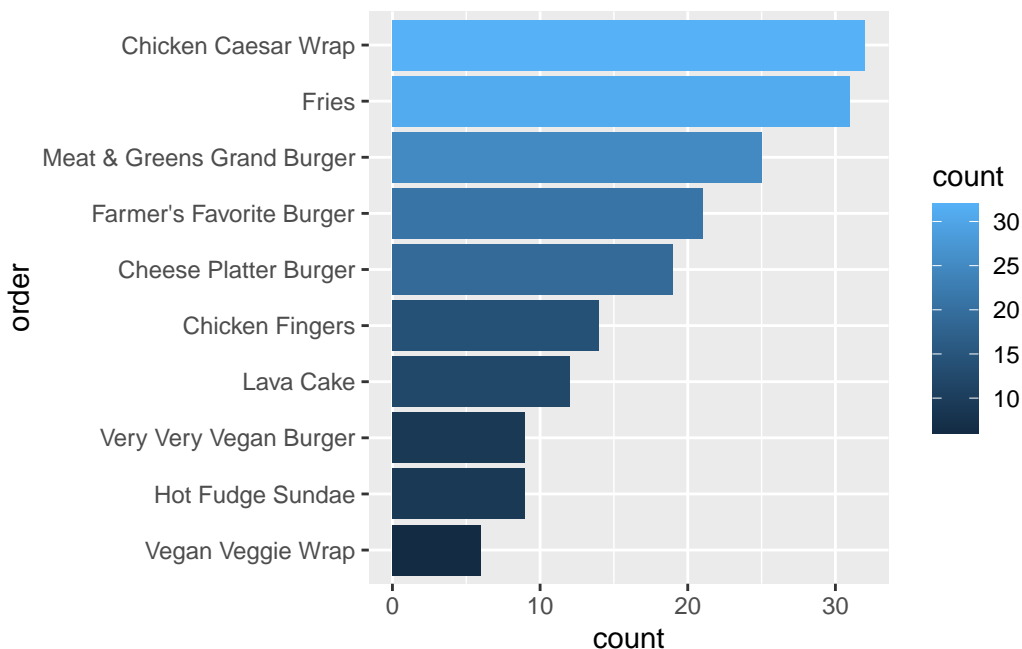
    }
  }
}
return(orders)
}

orders_1 <- as.data.frame(food_count_1(survey$order_1))
orders_2 <- as.data.frame(food_count_2(survey$order_1, survey$order_2))

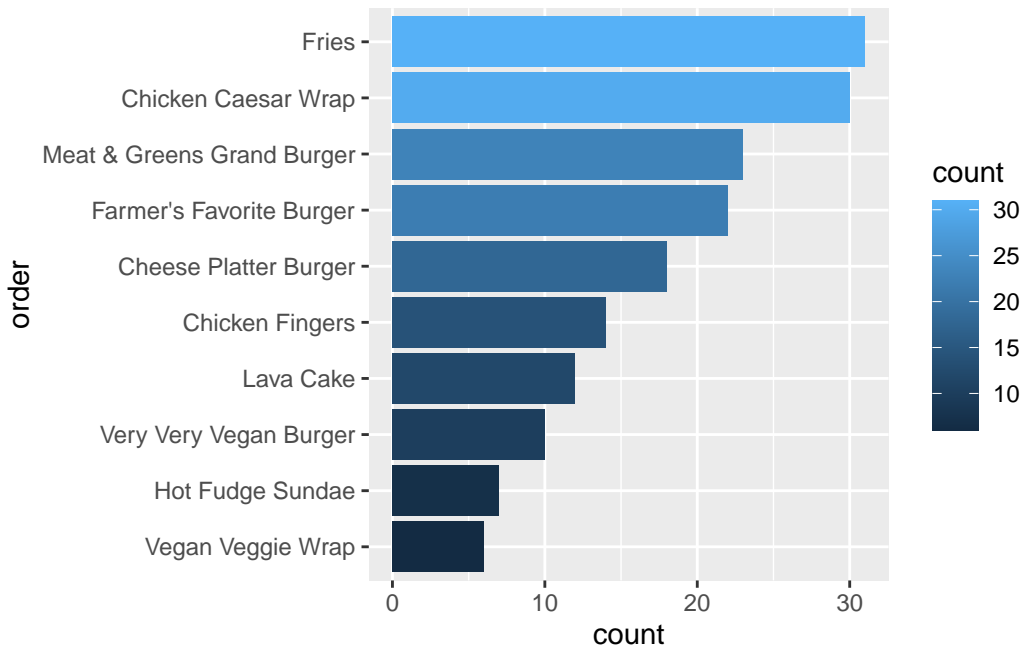
orders_1 <- orders_1 %>%
  rename(order = `food_count_1(survey$order_1)`) %>%
  filter(order != "I don't want to order anything.") %>%
  group_by(order) %>%
  tally(sort = TRUE, name = "count")
orders_2 <- orders_2 %>%
  rename(order = `food_count_2(survey$order_1, survey$order_2)`) %>%
  filter(order != "I don't want to order anything.") %>%
  group_by(order) %>%
  tally(sort = TRUE, name = "count")

orders_1$order <- factor(orders_1$order, levels = orders_1$order[order(orders_1$count, dec
ggplot(data = orders_1, aes(x = count, y = order, fill = count)) + geom_col()

```



```
orders_2$order <- factor(orders_2$order, levels = orders_2$order[order(orders_2$count, dec
ggplot(data = orders_2, aes(x = count, y = order, fill = count)) + geom_col()
```



## Exercise Level

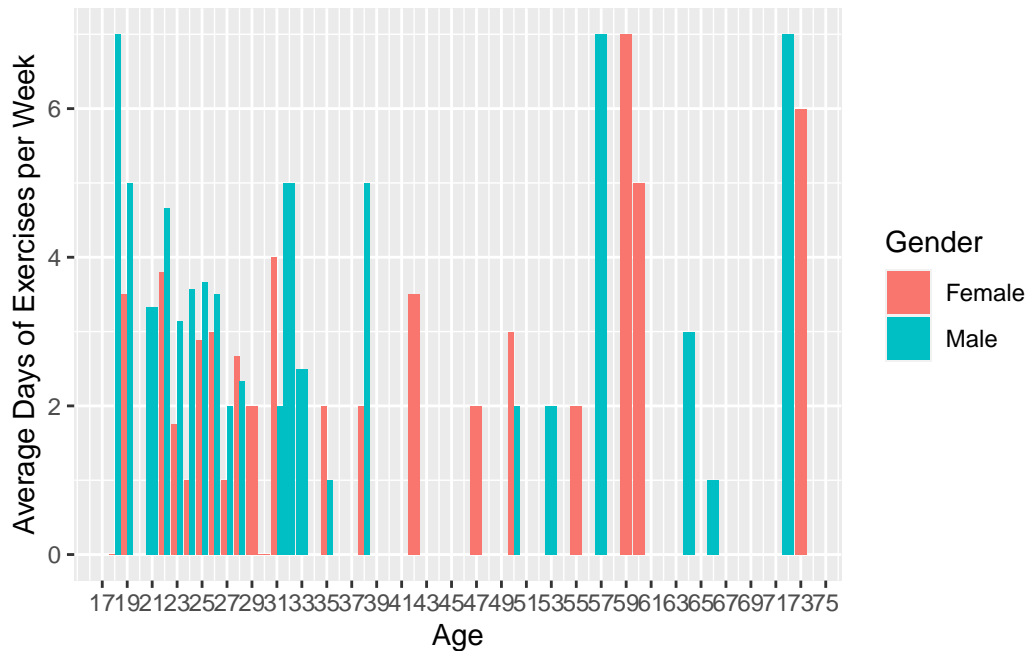
Average number of exercise (days per week) in Age and Gender

```
survey_TrueExercise <- subset(survey, num_exercise < 8 & age > 0)

survey_TrueExercise$gender_group <- ifelse(survey_TrueExercise$gender == "Male", "Male", "Female")

survey_avg <- aggregate(num_exercise ~ age + gender_group, data = survey_TrueExercise, FUN = mean)

ggplot(survey_avg, aes(x = age, y = num_exercise, fill = gender_group)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  labs(x = "Age", y = "Average Days of Exercises per Week", fill = "Gender") +
  scale_x_continuous(breaks = seq(17, 75, 2))
```



```

survey <- survey %>%
  mutate(calorie_1 = case_when(
    order_1 != "I don't want to order anything." ~
      sapply(order_1, function(x) find_calories(get_entrees(x))),
    TRUE ~ 0
  )) %>%
  mutate(calorie_2 = case_when(
    order_2 != "I don't want to order anything." & switch == 1 ~
      sapply(order_2, function(x) find_calories(get_entrees_from_order_2(x))),
    order_2 != "I don't want to order anything." & switch == 0 ~
      calorie_1,
    TRUE ~ 0
  )) %>%
  mutate(change = case_when(
    calorie_2 < calorie_1 ~ 1,
    TRUE ~ 0
  ))

calorie_aware <- survey %>%
  filter(calorie_consideration == "Yes")
calorie_unaware <- survey %>%
  filter(calorie_consideration != "Yes")

```



```

men <- survey %>%
  filter(gender == "Male")
women <- survey %>%
  filter(gender == "Female")
exercise <- survey %>%
  filter(physical_active != "No")
non_exercise <- survey %>%
  filter(physical_active == "No")

```

### T-Test on entire sample

```

t.test(survey$calorie_2, survey$calorie_1, alternative = "less",
       paired = TRUE)

```

Paired t-test

```

data:  survey$calorie_2 and survey$calorie_1
t = -1.6611, df = 109, p-value = 0.04978
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
 -Inf -0.0471196
sample estimates:
mean difference
 -36.45455

```

### T-test conditioned on being calorie aware or not

```

t.test(calorie_aware$calorie_2, calorie_aware$calorie_1, alternative = "less", paired = TRUE)

```

Paired t-test

```

data:  calorie_aware$calorie_2 and calorie_aware$calorie_1
t = -3.4584, df = 33, p-value = 0.000759
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
 -Inf -89.06327

```

```
sample estimates:
mean difference
-174.4118
```

```
t.test(calorie_unaware$calorie_2, calorie_unaware$calorie_1, alternative = "less", paired = TRUE)
```

Paired t-test

```
data: calorie_unaware$calorie_2 and calorie_unaware$calorie_1
t = 1.3539, df = 75, p-value = 0.9101
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
-Inf 56.33827
sample estimates:
mean difference
25.26316
```

```
t.test(men$calorie_2, men$calorie_1, alternative = "less", paired = TRUE)
```

Paired t-test

```
data: men$calorie_2 and men$calorie_1
t = -1.3358, df = 58, p-value = 0.09341
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
-Inf 11.11813
sample estimates:
mean difference
-44.23729
```

```
t.test(women$calorie_2, women$calorie_1, alternative = "less", paired = TRUE)
```

Paired t-test

```
data:  women$calorie_2 and women$calorie_1
t = -2.2506, df = 48, p-value = 0.01451
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
    -Inf -11.80265
sample estimates:
mean difference
    -46.32653
```

```
t.test(exercise$calorie_2, exercise$calorie_1, alternative = "less", paired = TRUE)
```

#### Paired t-test

```
data:  exercise$calorie_2 and exercise$calorie_1
t = -1.5748, df = 105, p-value = 0.05915
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
    -Inf 1.921778
sample estimates:
mean difference
    -35.75472
```

```
t.test(non_exercise$calorie_2, non_exercise$calorie_1, alternative = "less", paired = TRUE)
```

#### Paired t-test

```
data:  non_exercise$calorie_2 and non_exercise$calorie_1
t = -1, df = 3, p-value = 0.1955
alternative hypothesis: true mean difference is less than 0
95 percent confidence interval:
    -Inf 74.43499
sample estimates:
mean difference
    -55
```

```
reg_gender <- feols(data = survey, change ~ gender)
etable(reg_gender)
```

```

                                reg_gender
Dependent Var.:                change

Constant                0.1429** (0.0506)
genderMale              -0.0073 (0.0684)
genderPrefernottoanswer. 0.3571 (0.2554)

-----
S.E. type                IID
Observations            110
R2                      0.01883
Adj. R2                 0.00049
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### CATE (can get this from t-test results)

```
ATE <- mean(survey$calorie_2) - mean(survey$calorie_1)
```

### Power

```
cohen_d <- ATE/sd(survey$calorie_2)
cohen_d
```

```
[1] -0.04761407
```

```
pwr.t.test(n = length(survey$calorie_1), d = cohen_d, sig.level = .05, power = NULL)
```

Two-sample t test power calculation

```

      n = 110
      d = 0.04761407
sig.level = 0.05
```

```
power = 0.06427552  
alternative = two.sided
```

NOTE: n is number in *each* group

```
pwr.t.test(n = NULL, d = cohen_d, sig.level = .05, power = 0.7)
```

Two-sample t test power calculation

```
n = 5445.807  
d = 0.04761407  
sig.level = 0.05  
power = 0.7  
alternative = two.sided
```

NOTE: n is number in *each* group