

STA304 - Fall 2021

Assignment 1

Joseph Hotz - 1005953354

The `header-includes` section in the intro lets us use the `hyperref` package when compiling this R Markdown document into a pdf. This package allows us to set up coloured hyperlinks in the document. (karpfen, 2017)

Part 1

Goal

McDonald's is one of the largest fast food companies on the planet, operating nearly 40,000 restaurants across the world (Kahar, 2020), and serving tens of millions of customers daily. The restaurant chain, alongside many other franchises in the fast food industry, are often correlated with (Hollands, 2012), and blamed for the ongoing obesity epidemics prevalent among youth in countries such as Canada, the United Kingdom and the United States.

The health effects of frequently eating food from McDonald's were famously depicted in Morgan Spurlock's 2004 documentary *Super Size Me*, in which he exclusively ate food from the chain for 30 days, and documented the deteriorating effects which this 'diet' had on his physical health. During this 30-day span, Spurlock's weight increased by over 20 pounds. On top of this massive weight gain, he experienced other health issues such as depression, chest pains, and increased blood pressure. (Spurlock, 2004) He continued to face health issues after the 30-day experiment had ended, and one of his doctors claimed that Morgan's month-long experiment may have caused some irreversible lifelong damage to his heart.

Although the results of the *Super Size Me* experiment were damning, a common criticism of the documentary is that the experiment was unrealistic (Hayasaki, 2004), as most people do not exclusively eat fast food. Spurlock has rebuked these claims by arguing that many fast food locations are in lower-income areas, and the low food prices mean that lower-income families may be reliant on these nearby fast-food franchises (Hayasaki, 2004).

A 2012 study by a graduate student at the University of Western Ontario determined that there is a statistically significant positive correlation between Canadian adults' body mass index (BMI) and the density of fast-food restaurants near their place of residence (Hollands, 2012).

Time and time again, it has been shown that fast food restaurants are significant factors driving rising obesity rates in many countries across the world, including right here in Canada, and McDonald's is one of the worst offenders.

A major issue with these fast food establishments is that the food at these restaurants are high in calories, and low in price, meaning that these restaurants (which have many locations across Canada) provide a great opportunity for their customers to consume calories cheaply, and these restaurants are often the most popular restaurants around.

The goal of this survey is to acquire information about McDonald's customers' spending habits when they are buying a meal from McDonald's during the restaurant's "lunchtime hours" between 11:00 AM and 4:00 PM (The Breakfast Hours, 2021). In particular, the information which we aim to examine is customers' average meal prices and caloric intakes when buying food from McDonald's, the reasons why they enjoy

eating at McDonald's, how often they eat from the chain, and the demographic information of McDonald's patrons.

We can use the data from this survey's responses to examine why McDonald's customers enjoy eating food from the chain, as well as examining the prices and caloric values of frequently ordered meals from McDonald's, in order to examine the 'average' McDonald's meal in terms of its health effects. We can also use the collected data to examine relationships between customers' demographic information (age and gender identity) and variables such as how often they buy food from McDonald's, and the prices and caloric content of the meals which they purchase when at McDonald's. This information can help us determine which demographic groups are particularly affected by the unhealthy options at McDonald's and the affects which frequently eating fast food can have.

As fast food consumption is correlated with obesity rates in Canada (Hollands, 2012), examining Canadians' fast food consumption rates can provide more information about how McDonald's menu options affect Canadians, and how the franchise developed such a stronghold across the world. The obesity epidemic is a public health crisis (albeit, not the largest public health crisis at the moment), and the data from this survey can help public health officials combat increasing obesity rates, by providing information on the popularity of fast food among different demographic groups, and comparing these groups' fast food intakes to the recommended daily nutrition values.

Procedure

The goal of this survey is to learn more about the eating habits of Canadians when they are at McDonald's restaurants in Canada.

The target population for this survey would be all Canadians who have eaten a meal at McDonald's in the past year, as this survey is interested in the demographics and ordering habits of Canadians at McDonald's restaurants.

Ideally, the best way to implement this survey would be to publicize this survey in McDonald's restaurant locations across Canada over a period of time, as our target population comprises of Canadians who have been to a McDonald's restaurant, and the easiest way to find Canadians who have recently eaten at McDonald's is to simply survey people who are in Canadian McDonald's locations. One possible way to publicize this survey would be to have a link to the survey posted in

The 'ideal' sampling implementation described above may lead to a non-response bias, as many possible respondents may be unwilling to do a survey about their eating habits, since people who have poorer overall health often avoid participating in health-based surveys (Turk, Heneghan, & Nunan, 2019), and this survey does not seem like it would be an exception to that trend.

A common surveying method which many fast food restaurants use to increase the number of responses to their customer satisfaction surveys and reduce non-response biases is to entice customers to participate by offering some reward for taking these surveys, such as a discount on a customer's next purchase. (TTAG Systems, 2016) Links to fill out these surveys online and information about the rewards which a customer may receive for filling out the survey are often provided on customers' receipts, and this practice helps to decrease non-response biases in these restaurants' customer satisfaction surveys.

If possible, a great way to limit non-response biases that may impact this survey would be to use this surveying technique, and have the McDonald's corporation itself assist in promoting the survey to its customers. However, restaurants such as McDonald's tend to use discounts and/or free food to incentivize their customers to complete the survey, so this would create a catch-22, as the incentive for McDonald's customers to complete a survey about unhealthy eating habits at fast food restaurants would be that they can get more fast food in the future, which is quite counterproductive.

The best way to do this survey would likely be to advertise the survey at McDonald's locations across the country, and the incentive for filling out the survey could be a chance to win a cash prize, which helps to avoid the issue of incentivizing this survey with free McDonald's food as described above, while still incentivizing customers to fill out the survey.

If we were to implement the survey by advertising the survey at McDonald's locations across the country, the frame population would be anybody who had eaten at a McDonald's location in Canada during the time period in which we were broadcasting that this survey was happening. As mentioned above, our target population is everybody who had eaten at McDonald's in the last year. Since promoting the survey for an entire year is unrealistic, the McDonald's customers who are given the opportunity to complete this survey would likely be the customers who more frequently attend McDonald's, as these customers would be more likely to order food from McDonald's during the time when this survey was being broadcasted. Ideally, we would be able to advertise that this survey was happening for a long enough period of time to allow for most people in the target population to hear about the survey and participate in it, but that is somewhat idealistic.

Since this survey would be run as an online survey where anybody who is interested in participating is able to fill out the survey, the sample population for the survey would be everyone who knows about the survey and is willing to spend time doing the survey. This survey would be based on a sample of volunteers from the frame population, but we have detailed some possible measures which we could implement to try and increase the rate of volunteers.

Showcasing the survey.

The survey which I designed can be accessed via this [Google Forms](#) link. (Grolemond, 2014) The survey contains 10 questions: 3 of these questions have numerical answers, whereas the other 7 have categorical answers.

Why do you like eating at McDonald's?

- There are low prices at McDonald's.
- There is a McDonald's location near me.
- The food at McDonald's is enjoyable.
- It is quick and efficient to get food at McDonald's.
- There is good customer service at McDonald's.
- Other:

This question on the survey uses the "Checkboxes" feature in Google Forms, which allows the survey respondents to check as many of the boxes as they would like. As shown above, the survey provides 5 'default' options which survey respondents can choose to check off, as well as a 6th checkbox labelled "Other:" which allows users to enter their own response to this question.

This question is one of the most important questions in the survey, as it is vital in answering one of the key guiding questions for our research: "Why do Canadians enjoy fast food restaurants such as McDonald's?" This question allows survey respondents to provide as many options as they would like as to why they enjoy eating at McDonald's. Even though the question is a subjective question at heart, giving respondents multiple potential options to choose from and also allowing them to fill in their own personal reasons, while also ensuring that the responses can be varied which still permitting respondents to answer truthfully.

A drawback of this question is that the 5 'default' options are somewhat narrow in scope, and these do not necessarily cover the range of reasons why people enjoy McDonald's. Although there is an "Other:" option provided as a potential response to this question, participants in this survey may not want to think too deeply about their personal answers to this question, and they may instead prefer to just choose some of the pre-written answers which they feel most strongly about. A potential improvement for this question would be to set the response type to "Short Answer" or "Long Answer" on Google Forms, as this could help avoid the potential problem where respondents only choose pre-written answers by eliminating the pre-written answers altogether.

Another potential drawback of this survey question is that the wording of the question specifically asks about "eating at McDonald's", but the 5 pre-written answers are not all related to the quality of McDonald's food itself. The question's wording detracts from the goal of this question on the survey. This survey question could be improved by rewording the question to ask why the survey participants buy food from McDonald's, without specifically making the question about "eating".

Approximately how many meals have you eaten at McDonald's in the past year?

For this question, the respondents provide a numerical input, and Google Forms ensures that this input is a “whole number” (an integer), and it does not accept non-integer inputs for this question. A benefit of this question is that the target population for this survey is everybody who has eaten at McDonald's in the past year, so if a person's answer to this question is at least 1, then we know that they are a part of our target population. Another potential benefit of this question is that the results to this question can be used to examine how frequently Canadians go to fast food restaurants such as McDonald's, and we can use this information (alongside our other survey questions) to investigate which types of people are more likely to frequent fast food establishments. We can also use responses to this question to create subgroups of McDonald's customers based on how often they eat from McDonald's compared to the average survey respondent, as there could possibly be differences in customers' attitudes depending on how often they eat at McDonald's.

A possible drawback of this question is that a “meal” is not well-defined in this question. The goal of this survey is to learn about Canadians' spending habits during the “lunchtime hours” at McDonald's. The description of the form and the goals of the survey which are displayed at the top mention that the information collected is used to determine characteristics of McDonald's customers' lunchtime (11:00 AM - 4:00 PM) meals. However, this question mentions “meals”, and it does not specify that these meals are meant to be lunchtime meals at McDonald's, so some people may have answered incorrectly, as they might have included breakfast meals in their estimated number of meals, since it was not clear that the question was about lunchtime meals only.

Another possible drawback is that it is difficult to determine how many meals you had from McDonald's over a long period of time such as a year. Unless someone knows exactly how often they eat food from McDonald's, their answer to this question may be far from the true answer, which may cause issues when analyzing the responses to this question, as the distribution of responses to this question may be significantly different than the true distribution of how often survey respondents ate at McDonald's over the past year.

When you go to McDonald's for a lunchtime meal, which of the following menu items is your “go-to” order?

For this question, a survey respondent can choose exactly one of the pre-written answers (most of the popular ‘meal’ items from the McDonald's menu), or they can fill in the “Other:” section with their own answer instead.

One of the main components of the data which we aim to investigate with our survey results is the average caloric content and price of McDonald's customers' meal choices at McDonald's. This question (alongside the following questions about ordering fries and a drink) provides a meaningful way for us to gauge the calorie and price values without having to directly ask the survey respondents for these values.

Survey respondents likely might not know how much they tend to spend when they are buying lunch at McDonald's, whereas they are more likely to know which product(s) they tend to order if they are at McDonald's, especially since a majority of the lunchtime meal options are presented in this question, so this question almost mimics the sensation of customers choosing what to order at McDonald's. Similarly, people who are filling out this questionnaire will almost surely not know the average amount of calories in their McDonald's order off-hand, and people who are filling out this survey would likely either enter a rough guess of the caloric content, or give up on finishing the survey entirely.

The greatest benefit of this survey question is that this question is much more accessible for many potential survey-takers, and it allows us to get the information which we need while also not prompting people to make rough guesses and/or quit the survey entirely. Phrasing this question in terms of the menu items at McDonald's also helps us get more accurate answers than we may have gotten otherwise, since it is much easier to answer a question about what you tend to order from McDonald's than having to find specific information about the foods which you normally order.

Similarly to the first question which was showcased, a potential drawback of this question is that the survey provides multiple options for respondents to choose from, but the list is non-exhaustive. This is likely a

smaller issue than in the previous question (Why do you like eating at McDonald's), as there are still finitely many answers to this question, but some of these answers may not have been in the pre-established list of answers. Like the first of the three showcased questions, this question would likely fare better if it was a "Short Answer" question instead of having a non-exhaustive list of options, and such a change would also have the added benefit of making the survey look shorter and less daunting.

Part 2

Data

Data Collection

The data presented in this section of the report was collected via [a survey](#) which was hosted on Google Forms. The Google Form was open and accepting responses for a somewhat short span of time; the survey became available at 2:00 AM on September 30, 2021, and closed at 11:59 PM on that same day. In total, 86 people responded to the survey, and although this is a fair number of people, it is far from a sample which is representative of the entire target population.

Joey Hotz, the lead (and only) researcher for this survey was also the person in charge of sending out links to the survey and requesting that people answered the questions on the survey in order to collect the data presented in this section. Links to the survey were posted on multiple UofT-related Discord servers such as ‘UofT Statistical Sciences’ and ‘Math Union Toronto’, each of which has around 1000 members. A link to the survey was also posted on the STA304 class Piazza page.

Outside of UofT-related channels where the link to the survey was posted, Joey also directly sent the survey link to many of his family and friends and asked them to fill out the survey, and requested that they shared these links with other people in order to broaden the ‘reach’ which the survey had.

The largest obstacle affecting our data collection process was that most of the people who were given a link to the survey from our lead researcher (Joey) were people who he knew from high school or university. Since a majority of the people in Joey’s social circles are also around his age (20 years old), the demographics of the people who filled out the survey are most likely starkly different from the overall demographics of all Canadians who have eaten at McDonald’s in the past year (our target population), and this would be a clear example of a bias in the sample caused by convenience-based sampling.

The median age of respondents in this survey is only 25 (more on the data itself later), whereas the median age in Canada is 41.1 years old (Statistics Canada, 2021). Although the target population for this sample is all Canadians who have eaten at McDonald’s in the past year, and not all Canadians in general, McDonald’s restaurants are so prevalent across the country that the median age of the target population is almost surely not as low as the median age of the people who responded to this survey.

In an ideal scenario (as described in Part 1), we would be able to send a link to the survey to a large proportion of the target population, which would help to nullify the issues that can be caused by convenience-based sampling. However, as mentioned in Part 1, under these ideal surveying circumstances, the sample population still may not match up with the frame and target populations due to issues such as non-response biases, which are especially prevalent when performing health-related surveys. (Turk, Heneghan, and Nunan, 2019).

Data Cleaning

Once the data had been collected, the next step was to clean the data. Google Forms (the platform used to host the survey) automatically records the timestamps of every survey submission, so the first thing to do was to remove this column from our data, as these timestamps are irrelevant to our survey results.

Once the Timestamp column was removed, we renamed all of the remaining columns (Schork, 2020), and gave them significantly shorter names. This was not strictly necessary for our data analysis, but it is very convenient, as the original column names were the full questions in the survey, and these were far too long and wordy to be used conveniently.

Next, we looked at the variables in our survey to ensure that there were no ‘joke answers’ or other response biases which could impact our results. The survey data only had one massive outlier; one of the survey respondents said that they had eaten approximately 410 McDonald’s meals in the last month, but only 10 over the past year. Although we are not sure if this was a joke or a strange typo, it was an issue nonetheless, so we removed this observation from the data, since this brings the validity of their other responses into question.

Of the remaining 85 survey respondents, 5 of them reported that they had not eaten a meal from McDonald's at all in the past year. As our target population for this survey is Canadians who have eaten at McDonald's in the past year, these survey responses were removed, as they were not relevant to the target population.

Next, we needed to deal with finding the caloric values and prices corresponding to the answer to the last 4 questions of the survey. In order to do this, we collected caloric values from the 'Nutrition Calculator' (McDonald's, 2021) on the McDonald's company website, and found the prices on the CostFinderCanada website (CostFinderCanada, 2021), as these prices were not available on the McDonald's company website. We stored this information in three separate csv files; one for the entrees, one for the fries, and one for the drinks.

In order to turn the answers from our survey into the numbers which we collected above, we defined functions in R (Blischak et al., 2016) which would accept a string matching a McDonald's menu item, and would return the corresponding price (before tax) or caloric value of the item, based on the McDonald's menu information collected above. (Siddiqui, 2020) In total, we defined 6 functions, as we needed a separate function to get the calorie count and price for each type of menu item (entrees, fries, and a drink).

We then applied each of these 6 newly-defined functions in a "row-wise" fashion (using the `rowwise()` function), which applies each of these functions on only one survey response at a time, allowing us to compute the calorie and price values for each individual survey response, and store these as new variables in our dataframe.

Next, after we had found the prices and calorie amounts of the individual menu items which our survey respondents would tend to order from McDonald's, we added two more columns to the data, representing the total prices and calories for our respondents' meals at McDonald's.

Unfortunately, we were not able to calculate the calorie and price values for most of the people who wrote their own answer in the "Other:" section, as a majority of these custom inputs (such as "I'm a vegetarian") did not correspond to individual entrees on the McDonald's menu which had available nutrition and price data.

Variable Information

In total, our sampled data contains 18 different variables. However, some of these variables are inherently more important to our analyses than other variables. For example, the variable storing about how McDonald's customers would choose to get their food from McDonald's is not as immediately relevant to our analyses as the overall number of calories which people consumed in a McDonald's meal. Similarly, variables such as the price of a drink which a person would tend to order are also not as useful in our analysis, but this variable was instrumental in computing the totals which we care about.

Some of our main variables of interest are:

Age was a self-reported variable in the survey, as the first question in the survey asks participants to input their age in years. The Google Forms document ensures that the answers to this question are integers.

Gender was also a self-reported variable, and the survey question allowed survey-takers to choose one of 4 gender identities, or fill in the "Other:" box if they chose to do so.

The ages and genders of our respondents are the main summary statistics we are interested in as far as the demographic information about our surveyed population. Although other information such as racial identity and annual salary could be useful, people might feel less comfortable with disclosing such information on a survey.

Ages and Gender Identities of Survey Respondents (n = 80)

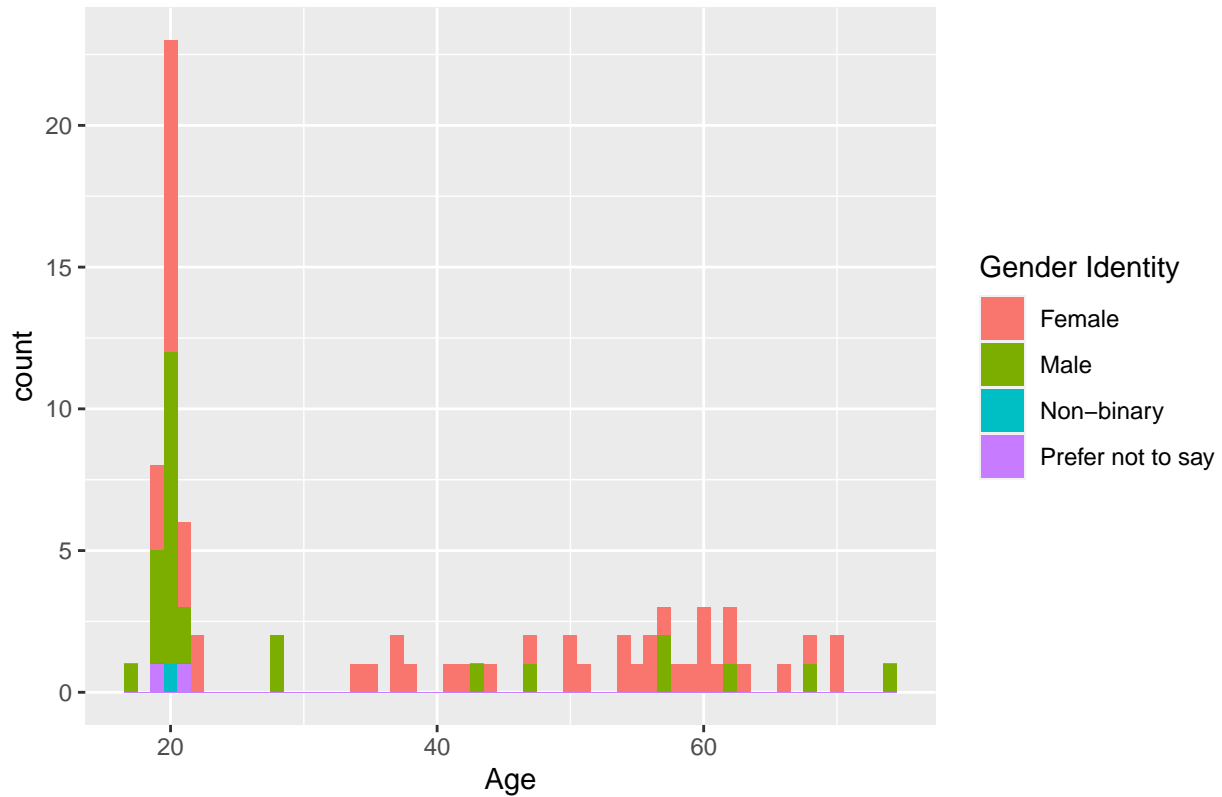


Table 1: Distribution of ages of survey respondents

	Observations	Mean	Standard Deviation	Median	Minimum	Maximum	Range
Age	80	36.5	18.628	25	17	74	57

Table 2: Distribution of ages of survey respondents, based on their gender identity

Gender Identity	Count	Minimum Age	1st Quartile	Median Age	3rd Quartile	Maximum Age	Age Range	Mean Age	Variance
Female	50	19	20.0	41.5	57.75	70	51	40.820	334.600
Male	27	17	20.0	20.0	35.50	74	57	30.333	316.923
Non-binary	1	20	20.0	20.0	20.00	20	0	20.000	NA
Prefer not to say	2	19	19.5	20.0	20.50	21	2	20.000	2.000

From the outputs above, it is clear that based on their ages alone, the people sampled in this survey are not representative of Canadians as a whole, and are likely not representative of our target population; Canadians who have eaten a meal from McDonald's in the past year, as we expect the target population to comprise of most of the Canadian population. It is visually evident that a vast majority of the surveyed population are between the ages of 15 and 25, and this likely due to an error caused by convenience sampling, as mentioned in a prior section. Additionally, there is a massive gender discrepancy in the sample population (only 33.75%

of respondents were male-identifying) which is not present in the overall demographics of Canada. (Statistics Canada, 2021)

Next, we will look at how many meals people reported that they ate from McDonald's in the past month and in the past year.

Past Month is a self-reported measure of how many times a person has eaten a meal at McDonald's in the past month. Like the "Age" variable described above, the Google Forms survey ensures that responses for this question are integers.

Past Year is a very similar measure to "Past Month", but it instead asks for how many meals the person has eaten from McDonald's over the past year. As mentioned in Part 1, this variable may be somewhat inaccurate, as respondents may find it difficult to remember buying fast food many months prior, and may be pressured to simply make an "educated guess".



Table 3: Distribution of number of McDonald's meals eaten by survey respondents in the past month and past year

	Observations	Mean	Standard Deviation	Median	Minimum	Maximum
McDonald's meals eaten in the past month	80	1.837	1.603	1	0	8
McDonald's meals eaten in the past year	80	15.537	16.084	10	1	100

The scatter plot above demonstrates the relationship between the number of McDonald's meals which a person has eaten in the past month, and the amount of McDonald's meals which they have eaten in the past year. The red line on the scatter plot represents the 'line of best fit', and it is visually clear that there is a positive correlation between the number of McDonald's meals someone ate in the last month and the last year. We will be exploring this relationship further in the "Methods" section of the report below.

Total Calories is a measure of the total number of calories which a person's average lunchtime meal at McDonald's contains, and it was computed using the caloric values which we collected from the '[Nutrition Calculator](#)' (McDonald's, 2021) on the McDonald's company website.

Number of calories in survey respondents' McDonald's lunches (n = 60)

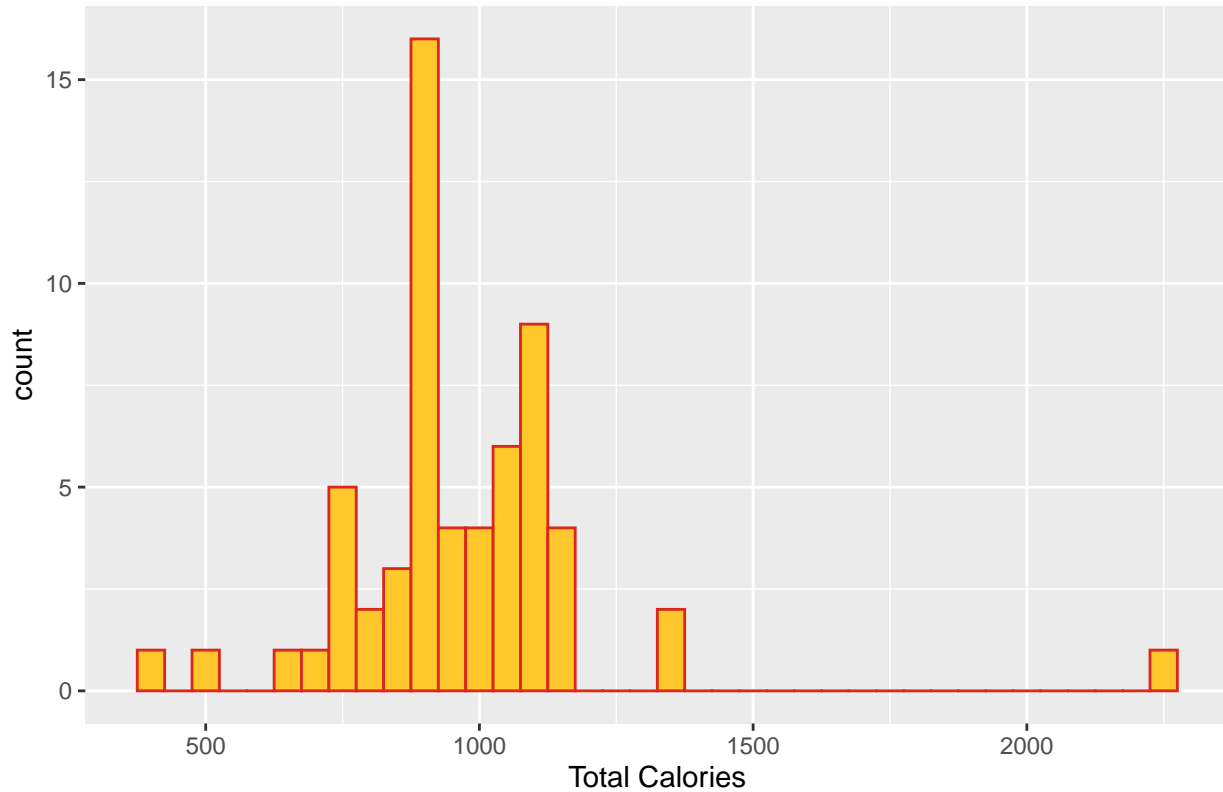


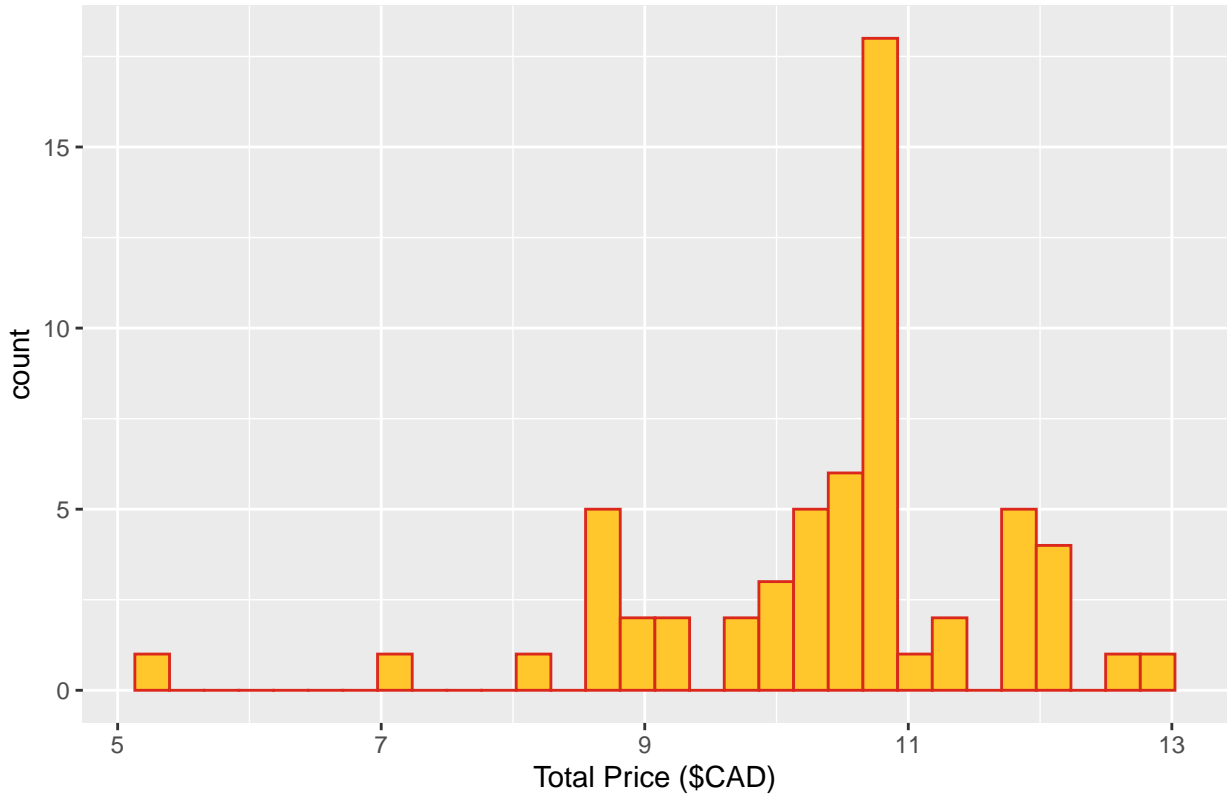
Table 4: Total caloric value of survey respondents' lunches from McDonald's.

	Observations	Mean	Standard Deviation	Median	Minimum	Maximum	Range
Calories in the meal	60	967.8	240.597	926	410	2240	1830

From the outputs above, we can see that the median number of calories eaten in a McDonald's lunch is a bit over 900 calories. There is a very large standard deviation in our results, due to the number of possible lunch options at McDonald's as well as the small sample size, but nonetheless, it is a relatively large number of calories to consume in one meal. We will be examining this variable more in the "Methods" section below.

Total Price is a measure of the total price (before tax) of a person's average lunchtime meal at McDonald's, and similarly to the "Total Calories" variable, we computed these values using the prices which were listed on the [CostFinderCanada](#) website, as the McDonald's corporate website did not display the prices of their prod-

Total cost (before tax) of survey respondents' McDonald's lunches (n = 60)



ucts.

Table 5: Total price (before sales tax) of survey respondents' lunches from McDonald's.

	Observations	Mean	Standard Deviation	Median	Minimum	Maximum	Range
Price of the meal (\$CAD)	60	10.404	1.328	10.745	5.39	13.02	7.63

Based on these results, the mean price of McDonald's lunch seems to be about \$10.40, but as we can see from the histogram above, a majority of the respondents spend between \$9 and \$12 on their McDonald's lunches, with some outliers on the fringes of the histogram which seem to be lowering the mean away from where most of the results are concentrated. McDonald's is often criticized for providing calorie-dense meals at low prices (Spurlock, 2004), and based on the histograms above, these results definitely demonstrate why those criticisms may be valid.

Astute readers may have noticed that there are only 60 values for the total number of calories and total prices of a person's meal, even though there are 80 observations in the dataset. The reason for this discrepancy is because many of the respondents put invalid answers to one of the questions about their meal choices, so we were unable to find the price information and caloric information for their meal choices, and we chose not to replace or impute the missing data.

All analysis for this report was programmed using R version 4.1.1 with the tidyverse 1.3.1 R package (Wickham et al., 2021) and the psych 2.1.9 R package (Revelle, 2021). Tables were made using the knitr 1.3.6 package. (Xie et al., 2021). The report was compiled with MiKTeX 21.6 (Schenk, 2021).

Methods

Hypothesis Testing

Most modern dietary guidelines will have different caloric intake and overall food recommendations for adult men and adult women (Fenton & Thompson, 2019), and having these differences in the recommendations makes sense, as males are typically taller than females. However, one thing which many dietary guidelines agree upon is that people should avoid eating fast food (Spurlock, 2004), as these restaurants are often considered to be less healthy options, for reasons explore in this report and in countless other works.

In our survey, we collected information about many aspects of people’s fast food choices, such as how frequently they have visited McDonald’s restaurants in the past month, and the past year.

We can use this information to compare whether men or women eat food from McDonald’s on a more regular basis. Although we had more than two gender identities as options in the survey, there were only 3 people who did not select ‘Male’ or ‘Female’ as their gender, so that group of people is not large enough to draw any meaningful conclusions about.

In order to compare the frequency in which men and women ate from McDonald’s, as well as determining if there is a statistical difference between these groups, we will perform a *hypothesis test*. Let μ_M represent the expected value of how many times a man will have McDonald’s for lunch in a month, and similarly, let μ_W represent the expected number of times a woman has McDonald’s for lunch in a month.

Our null hypothesis will be $H_0 : \mu_M - \mu_W = 0$, which we could instead write as $H_0 : \mu_M = \mu_W$. Our null hypothesis is that there is no difference between men and women with regard to how often they eat from McDonald’s in a month. This gives rise to the alternative hypothesis $H_A : \mu_M \neq \mu_W$, i.e. that there is a difference between how many times men and women eat lunch at McDonald’s in a month.

Of the 80 people who responded to the survey, 27 of them identified as male, and 50 of them identified as female. Since these are groups of different sizes (and the people in these groups are unrelated), we will perform a two-sample t-test to compare the means of these groups. As we do not know the underlying distribution of how often men and women will eat food from McDonald’s, we will not assume that the variances corresponding to these two groups are equal, as this is a strong assumption. As we have not assumed that the variances of how often men and women eat lunch at McDonald’s are equal, we will need to perform Welch’s t-test, which does not make this assumption. (Statology, 2020)

Welch’s t-test is a statistical test which allows us to compare the means of two groups of data which are not necessarily of the same size, and may not have the same variance. Welch’s t-test assumes that each group follows a Normal distribution, but the parameters of each group’s respective distribution may differ, and it computes the probability (also known as a p-value) that these underlying Normal distributions have the same mean.

We will perform our hypothesis test at a 5% level of significance, which is equivalent to a confidence level of 95%.

Confidence Intervals

Dietitians and health organizations throughout North America recommend that people eat approximately 2000 calories per day to meet their caloric needs. (Ontario Government, 2021) These organizations often recommend people to take their own needs into account, as caloric needs do not have a “one-size-fits-all” solution, so one person’s needs may vastly differ from another person’s.

Dietary guidelines from the United States recommend that a ‘typical’ adult woman should eat between 1600 and 2400 calories per day, and a typical adult man should eat between 2000 and 3000 calories per day. Assuming that an adult is eating three meals per day, these guidelines tell us that an average adult woman’s lunch should contain between 533 and 800 calories, and an average adult man’s lunch should be between 667 and 1000 calories. (Fenton & Thompson, 2019)

Recall that the sample of 60 people’s lunches shown in the previous section, the median caloric value was 926 calories, which is well above the recommended average for women, and close to the upper limit for the

average adult man’s recommended number of lunchtime calories.

Using a non-parametric bootstrap method, we will create a 95% confidence interval for the mean number of total calories in an ‘average’ McDonald’s lunchtime meal, so that we can compare this caloric intake to the recommended daily calorie averages for males as well as for females.

A non-parametric bootstrap is a method of using an already collected sample to determine a range of values for the true mean of a distribution (in this case, our distribution is the number of total calories in a McDonald’s lunch) without having to derive another sample and survey another group of people.

To perform a bootstrap, we select a random sample of the observations in our dataset and calculate the mean of that sample, and we repeat this process many times to get a sense of the distribution of different possible sample means, which helps us to determine the variance in the mean of a sample of the population. This is useful as it helps us to estimate the probabilities of encountering different sample means if we were to perform a sample of a specific size, and this can be used to create a range of values which should include the ‘true population mean’, as the mean of our surveyed sample is likely not the overall mean. (Dekking et al., 2005)

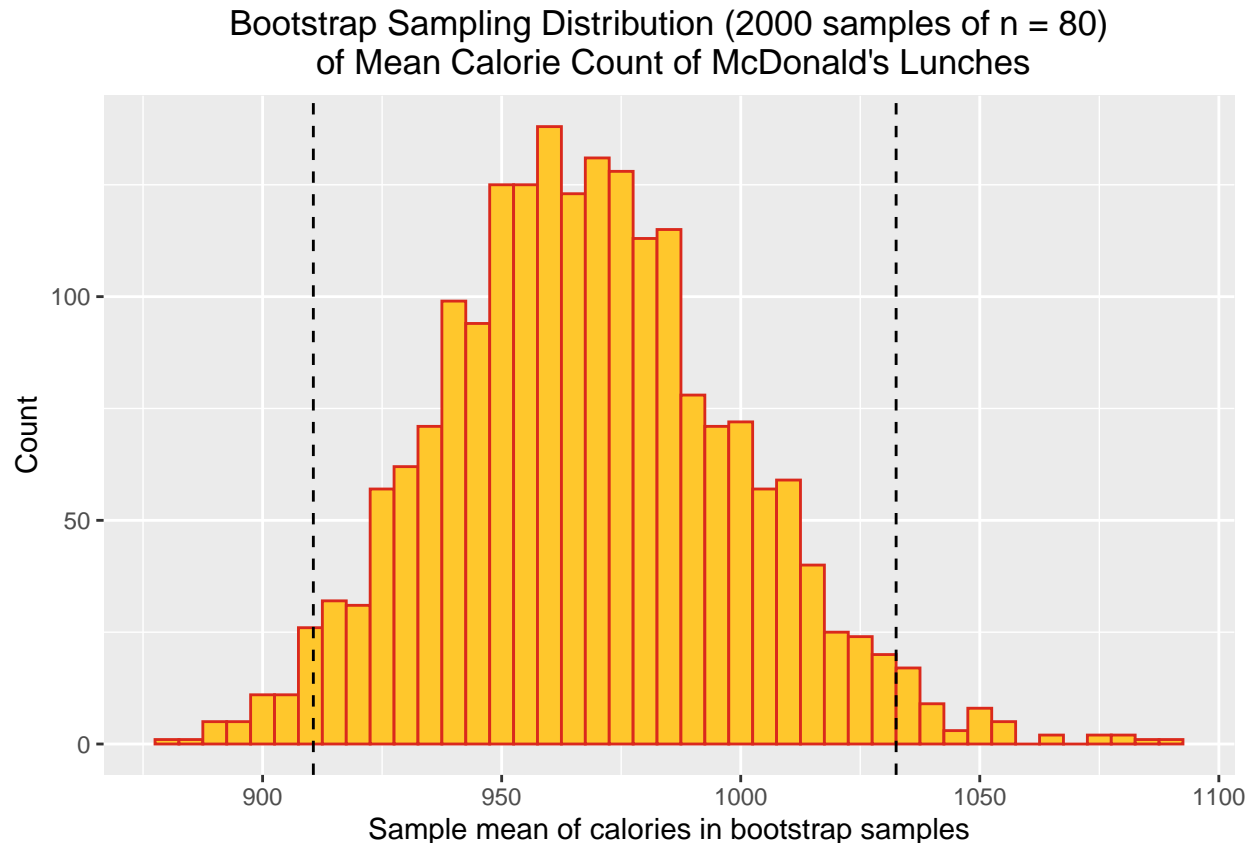
Once we have approximated the distribution of the sample means using the bootstrap method described above, we will be able to find the ‘middle’ 95% of the distribution, which we will take to be our 95% confidence interval for the mean number of total calories in a McDonald’s lunch.

Results

Recall from the “Methods” section that our null hypothesis was $H_0 : \mu_M - \mu_W = 0$, and our alternative hypothesis was $H_A : \mu_M \neq \mu_W$, where μ_M is the mean number of times a man buys lunch from McDonald’s per month, and μ_W is the mean number of times a woman buys lunch from McDonald’s per month.

In the results of the Welch’s t-test which we described in the Methods section, we found that the p-value corresponding to the null hypothesis was approximately 0.8127, which is well above the p-value of 0.05 required in order to reject the null hypothesis in favour of the alternative hypothesis.

Having a p-value which is this high means that we have virtually no evidence against the null hypothesis at all, and as such, we cannot claim that there is a statistically significant difference between how many times men and women eat lunch at McDonald’s in a month.



In the plot displayed above, the dotted black lines represent the 95% confidence interval for the mean calorie count of a lunch at McDonald's. As described in the Methods section, this range mostly overlaps with the upper range of the recommended number of calories in a meal (667 to 1000) for adult men (assuming 3 evenly-sized meals per day), and this is well above the upper limit for an adult woman's recommended lunchtime calories.

Furthermore, as we saw from the results of our hypothesis test above, there is no reason to assume that men and women visit McDonald's at a different frequency to one another, so these high-calorie means are not caused by the fact that men are the primary customers of McDonald's.

The lower bound of our 95% confidence interval is that the average lunch at McDonald's contains more than 900 calories, and most modern health guidelines (Fenton & Thompson, 2019; Ontario Government 2021) recommend a diet of well below 2700 calories per day. It is clear that the mean calorie count in a McDonald's lunch is far too high for it to be part of a healthy diet if someone even eats there somewhat regularly, as the calorie count for a meal is significantly above one-third of the recommended daily number of calories for a vast majority of the population.

Bibliography

- Blischak, J., Chen, D., Dashnow, H., & Haine, D. (2016, June). Creating Functions. Retrieved September 30, 2021, from <https://swcarpentry.github.io/r-novice-inflammation/02-func-R/>
- Campbell, N., & Finucane, F. (2018, November 26). McDonald's is a social and healthcare burden – whatever its charity PR might indicate. Retrieved September 29, 2021, from <https://theconversation.com/mcdonalds-is-a-social-and-healthcare-burden-whatever-its-charity-pr-might-indicate-105627>
- Calories on menus. (2021, September 23). Retrieved October 1, 2021, from <https://www.ontario.ca/page/calories-menus>
- Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics*. London: Springer Science + Business Media.

- Fast Food Receipt Surveys: Who Completes Them and Why. (2016, April 08). Retrieved September 29, 2021, from <http://www.ttagsystems.com/sms-marketing-blog/fast-food-receipt-surveys-who-completes-them-and-why/>
- Fenton, C., & Thompson, C., PhD. (2019, October 25). How Many Calories Should I Eat at Lunch? Retrieved from <https://www.livestrong.com/article/333407-how-many-calories-should-i-eat-at-lunch/>
- Grolemond, G. (2014, July 16). *Introduction to R Markdown*. Retrieved September 30, 2021, from https://rmarkdown.rstudio.com/articles_intro.html
- Hayasaki, E. (2004, December 1). ‘Super Size Me’ Filmmaker Offers Food for Thought. *Los Angeles Times*. Retrieved from <https://www.latimes.com/archives/la-xpm-2004-dec-01-me-class1-story.html>
- Hollands, S. (2012). *Association between the Fast-Food Environment and Obesity in Canada: A Cross-sectional Analysis* (Unpublished master’s thesis). University of Western Ontario. Retrieved September 28, 2021, from <https://ir.lib.uwo.ca>.
- Kahar, A. (2020, July 16). Top 10 Largest Fast Food Chains In The World. Retrieved from <https://www.greattopten.com/top-largest-fast-food-chains-in-the-world/>
- karpfen. (2017, June 30). How to set colour for toc links in R markdown document? Retrieved September 28, 2021, from <https://stackoverflow.com/questions/44846350/how-to-set-colour-for-toc-links-in-r-markdown-document>
- McDonald’s Menu Prices in Canada. (n.d.). Retrieved September 30, 2021, from <https://costfindercanada.com/mcdonalds-prices-in-canada/>
- Moore, S. (2017, November 4). R - Descriptive Statistics. Retrieved October 1, 2021, from <https://rpubs.com/stephenmoore56/136318>
- Nutrition Calculator. (n.d.). Retrieved September 30, 2021, from <https://www.mcdonalds.com/ca/en-ca/about-our-food/nutrition-calculator.html>
- *Population estimates on July 1st, by age and sex*. (2021, September 29). Retrieved September 29, 2021, from Statistics Canada. doi: <https://doi.org/10.25318/1710000501-eng>
- Revelle, W. (2021). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.9, <https://CRAN.R-project.org/package=psych>
- Schork, J. (2020, October 23). How to Rename a Column Name in R | 3 Examples to Change Colnames of a Data Frame. Retrieved September 29, 2021, from <https://statisticsglobe.com/rename-column-name-in-r-data-frame/>
- Siddiqui, N. (2020, July 06). How to find the index of an element in a vector in R? Retrieved October 1, 2021, from <https://www.tutorialspoint.com/how-to-find-the-index-of-an-element-in-a-vector-in-r>
- Spurlock, M. (Director). (2004). *Super Size Me* [Motion picture]. United States: Roadside Attractions, Samuel Goldwyn Films.
- Statology. (2020, October 21). How to Perform Welch’s t-Test in R. Retrieved October 1, 2021, from <https://www.statology.org/welch-t-test-in-r/>
- Tables. (n.d.). Retrieved October 1, 2021, from <https://rmarkdown.rstudio.com/lesson-7.html>
- Turk, A., Heneghan, C., & Nunan, D. (2019, November 14). Non-response bias. Retrieved September 30, 2021, from <https://catalogofbias.org/biases/non-response-bias/>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- What Time Does Mcdonalds Serve Lunch – McDonalds Lunch Hours 2021. (2021, May 9). Retrieved September 30, 2021, from <https://thebreakfasthours.com/what-time-does-mcdonalds-serve-lunch/>

Appendix

The following table is a glimpse of the data set which we used in our analyses. Aside from the removal of the `Timestamp` variable (which we had not intended to record), and the renaming of the columns, the first 10 columns (everything from `age` until `drink_choice`) is precisely the same as in the original data, and the remaining 8 columns were computed as described in the “Data Cleaning” section above.

```
## Rows: 80
## Columns: 18
## Rowwise:
## $ age                <dbl> 19, 20, 21, 20, 21, 28, 20, 20, 21, 20, 19, 20,~
## $ gender             <chr> "Female", "Male", "Male", "Male", "Female", "Ma~
## $ how_you_order      <chr> "Ordering from the drive-thru.", "Ordering from~
## $ why_you_like_mcdonalds <chr> "There are low prices at McDonald's., There is ~
## $ past_month         <dbl> 2, 1, 0, 3, 1, 1, 5, 5, 2, 0, 1, 0, 0, 1, 4, 0,~
## $ past_year          <dbl> 20, 15, 2, 30, 3, 12, 30, 15, 30, 2, 10, 10, 2,~
## $ entree             <chr> "Poutine", "Big Mac", "Double Filet-O-Fish", "F~
## $ fry_size           <chr> "Large", "Medium", "Medium", "Medium", "Large",~
## $ drink_size         <chr> "Large", "Medium", "Medium", "Medium", "I do no~
## $ drink_choice       <chr> "Sprite", "Diet Coke or Coke Zero", "NESTEA Ice~
## $ entree_calories    <dbl> NA, 570, 570, 400, NA, 480, 570, 570, 480, 530,~
## $ fry_calories       <dbl> 560, 350, 350, 350, 560, 350, 350, 350, 240, 35~
## $ drink_calories     <dbl> 280, 1, 150, 150, 0, 0, 150, 190, 3, 0, 4, 190,~
## $ entree_price       <dbl> NA, 5.69, 6.59, 5.19, NA, 5.39, 5.69, 5.69, 5.3~
## $ fry_price          <dbl> 3.29, 2.89, 2.89, 2.89, 3.29, 2.89, 2.89, 2.89,~
## $ drink_price        <dbl> 2.74, 2.24, 2.24, 2.24, 0.00, 0.00, 2.24, 2.24,~
## $ total_calories     <dbl> NA, 921, 1070, 900, NA, 830, 1070, 1110, 723, 8~
## $ total_price        <dbl> NA, 10.82, 11.72, 10.32, NA, 8.28, 10.82, 10.82~
```