STATS 15 Final Project: Analyzing McDonald's Nutrition

2023-06-07

Emma Morrison, Naomi Boodhoo, Tishi Avvaru, Yoonji Kim

Background Information

Main question

How do McDonald's menu items vary in affordability and nutrition?

Scope of analysis

McDonalds is one of the most populous fast food chains in the world, having locations serving cheap and tasty food in over a hundred countries. The chain prides itself on its ease of access to quick service fast food – food that isn't the most nutritious, but is made quickly and tastes good (most of the time). By no means does McDonald's claim to have the best tasting or the most healthy food. From a cursory glance at the menu, it is obvious that they do serve dishes with very high calorie counts. In this paper, we seek to answer the motivating question: How can a typical consumer get the most caloric "bang for their buck", and how unhealthy would that meal truly be?

Background on McDonalds

McDonalds is a quick service fast food chain invented over half a century ago in 1955. The chain operates on having quick and reliable service, often placing their restaurants at easy to access locations and making sure to have plenty of locations within one city or one area. Alongside their recognizable and longstanding reputation and marketing tactics, this allows the chain to gain a sort of notoriety for their constant appearances and their food products. This, in turn, helps them gain "easy customers" who are desperate for quick food, as well as reach everyday consumers who simply enjoy their cost-effective meals.

McDonalds has garnered a reputation of being unhealthy and nutritiously weak, in large part due to the high amount of calories and fats that are in their items. However, according to the McDonald's Global Happy Meals Goal's 2020 progress report, an average of 43% of Happy Meal Bundle offerings (the complete set of items that is included in the Happy meal) met the nutrition criteria across twenty markets. This criteria expected meals to be under or at 600 calories, with 10% of calories from saturated fat, 10% of calories from added sugar, and 650 milligrams of sodium. This figure has increased from years prior. However, this report does not tell the full story about McDonald's meals, as it only pertains to Happy Meals. Happy Meals are required to be bought with a healthy or healthy-adjacent side and are often smaller than regular meals. Thus, these statistics on Happy Meal nutrition do not apply broadly to the McDonald's menu.

There are no official studies pertaining to the "healthiness" of McDonald's items, thus we have included this factor into our exploration. While looking for the most cost-effective and filling item from McDonalds, we continue to explore the nutritional value of these items against the advisory intake of sodium, added sugars, etc. from the Dietary Guidelines For Americans published in collaboration with the USDA (United States Department of Agriculture) and the HHA (Human Healthy Services). This guide further published a list of guidelines for the average meal, which we have shown below. Using this breakdown of expected caloric values of meals in comparison to the meals recommended by McDonalds, as well as compared to the items we find to be the most "worth it" on the menu, we hope to create the most cost-effective meal McDonalds has to offer, and analyze that meal's nutritional value.

Data structure

The data set we use in this report, menu.csv, consists of 260 rows detailing the nutritional and caloric details of every item from the McDonald's menu. In cleaning up our data and ensuring it would fit out purposes, we manually entered prices for each of these items. For these prices we referenced the McDonalds closest to UCLA's campus: 11920 Wilshire Blvd. Any items that we were unable to find at this McDonalds, likely due to the fact that they were discontinued after this dataset was created, were eliminated from menu.csv. The final product that we used, finalmenudata.csv, consisted of 131 rows. The dataset has 25 variables being measured, of which we used 11. The variables we omitted did not pertain to our question of caloric density per price, and did not assist us in our nutritional analysis.

Explanation of variables *category (character)* - The category refers to the section under which McDonald's places an item when customers order online or in-person. There are nine categories, and thus nine classifications given to food items on the McDonald's menu.

item (character) - The item serves as a description of the food a customer can order. Items are numerous due to the small changes, such as the addition of an egg or bacon, which can constitute a whole different item. Items are classified according to the aforementioned categories.

serving_size (character) - The serving size is a measured amount of food or any other consumable item for which nutritional information is given. These serving sizes are taken straight from the McDonald's menu, and are listed in both ounces (fluid ounces included) and grams for each item. The nutritional values, such as calories and sodium, are based upon these serving sizes.

price (double) - The price refers to how much each item costs at the 11920 Wilshire Blvd that is closest to UCLA. These prices from this specific McDonalds were chosen in order to emphasize the cost-effectiveness of the question we hope to answer, we are able to explore what the most "bang for your buck" meal a UCLA student could acquire from the closest McDonalds to campus.

calories (double) - The number of total calories found within one serving size of an item.

total fat (double) - The total fat in grams found within one serving size of an item.

cholesterol (double) - The amount of cholesterol in grams found within one serving size of an item.

sodium (double) - The amount of sodium in grams found within one serving size of an item.

carbohydrates (double) - The amount of carbohydrates in grams found within one serving size of an item.

sugars (double) - The amount of sugars in grams found within one serving size of an item.

protein (double) - The amount of protein in grams found within one serving size of an item.

Explanation of created variables

serving_size_oz (double) - Due to the initial serving size variable showing the serving size in both oz and grams, as well as including the labels oz and g, the values in the column were classified as characters. This hindered our ability to measure any item or value that relied on serving size, so we were forced to truncate these values to only include the first four values of each row. This method gave us a list of serving sizes only in ounces or fluid ounces.

calorie_per_dollar (double) - In order to measure cost-effectiveness, we created a variable which divides the flat rate of calories for each item by its respective price. This variable allows us to peer deeper into the caloric "bang for your buck".

cal_density (double) - In order to survey caloric density, the amount of calories that exist within a given item per its serving size, we created a variable which divides the flat rate of calories for an item by its serving size in ounces. This allows us to have a method to measure calories per serving, rather than only relying on the flat amount of calories in each item.

Background information Overall, we want to look at how healthy and affordable McDonald's is. With the current perception and image of McDonald's as unhealthy and cheap, we want to explore how actual data correlates with this image.

As college students, we often turn to cheaper meal alternatives, such as McDonald's, to seek quick and easy meals. Despite the image that McDonald's holds as a food brand, how actually unhealthy or healthy can a typical McDonald's meal be? We wanted to explore these questions as collegiate McDonald's fanatics and see how much nutrition an average McDonald's order brings to our bodies.

McDonald's Global Happy Meal Goals 2020 Progress Report By 2019, 43% of Happy Meals met the Nutrition Criteria and more than 2.5 billion Happy Meals items sold contained the recommended food groups (protein, fruits, veggies, etc). With their goals of simplifying their ingredients and offering balanced, healthier meals, McDonald's is working towards the betterment of their brand image and quality of food.

Our data consists of the McDonald's menu with prices as well as the nutritional facts of each menu item. Through this data, we want to find how much nutritional value we are getting per dollar for each menu item and analyze what a typical meal (one entree, side, and drink) would nutritionally consist of.

Typical meal examples

Breakfast: egg mcmuffin, hash browns, premium roast coffee - 625 calories

- 22g protein
- 1150 mg sodium
- 0.5 g trans fat
- 1g added sugar

Breakfast: big breakfast w/ hotcakes, americano - 1340 calories

- 36g protein
- 2080 mg sodium
- 0.5g trans fat
- 41g added sugars

Lunch/dinner: quarter pounder w/ cheese, small fries, small sprite - 890 calories

- 33g protein
- 1400 mg sodium
- 1.5g trans fat
- 43g added sugars

Lunch/dinner: mccrispy, 4pcs chicken mcnuggets, small coke - 790 calories

- 35g protein
- 1510 mg sodium
- 0g trans fat
- 46g added sugars

```r
library(tidyverse)
library(janitor)
library(patchwork)
menu <- read_csv("finalmenudata.csv")
data(menu)
```

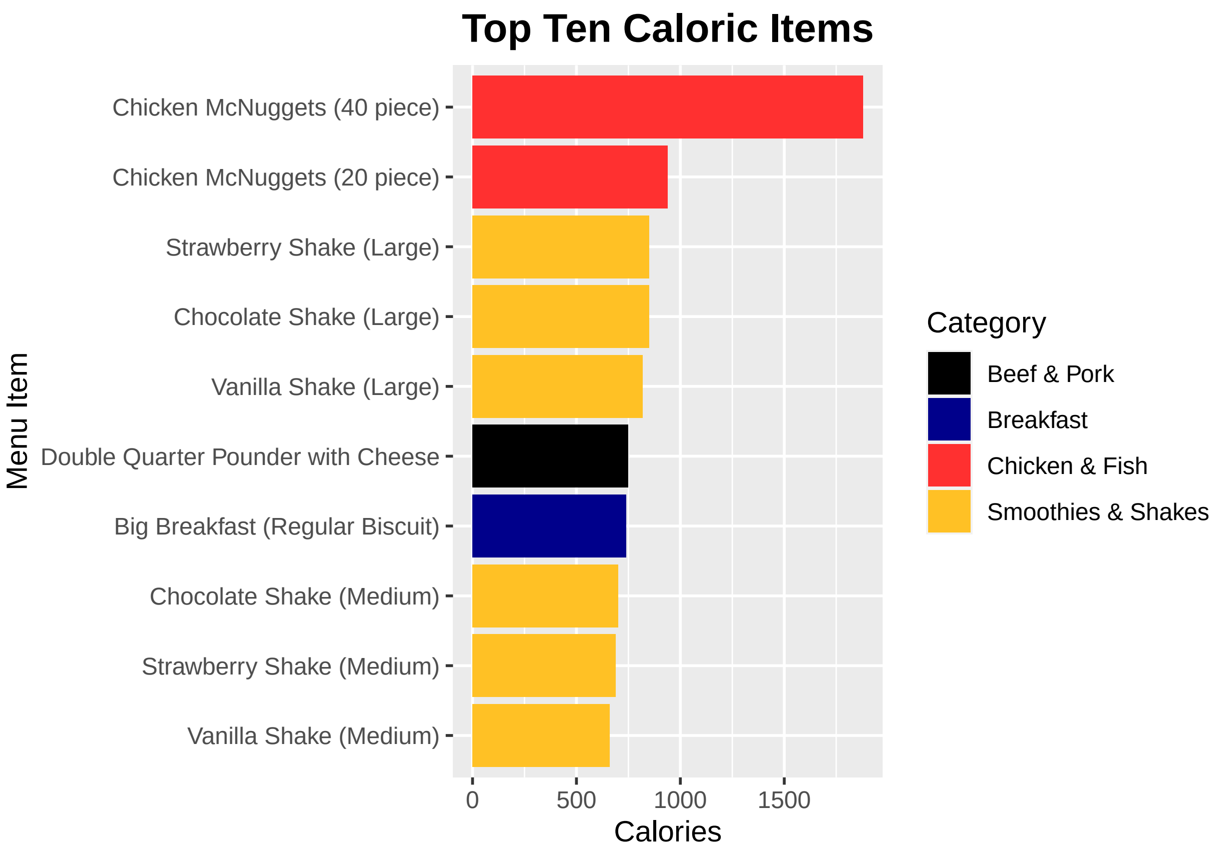
Exploratory Data Analysis

Calorie Exploration

Firstly, we decided to explore the most caloric items at McDonald's, to get a basis of what would be the items that could potentially be the most economical. We cleaned the data, and then started this exploration.

```
cleanmenu <- menu %>%
  clean_names()
```

```
cleanmenu %>%
  arrange(desc(calories)) %>%
  head(10) %>%
  ggplot(aes(x = calories, y = reorder(item, calories), fill = category)) +
  geom_col() +
  labs(x = "Calories", y = "Menu Item", title = "Top Ten Caloric Items",
       fill = "Category") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5)) +
  scale_fill_manual(values = c("black", "darkblue", "firebrick1", "goldenrod1"))
```



In looking for the most caloric options, we first found the top ten most caloric items on the menu. In doing so, we found that items from the smoothies & shakes contained the most calories. however, this may not be the most accurate representation of the most calorically dense options since the scale only accounts for a flat rate of calories, not the amount of calories per the serving size of each item. Thus, we resolved to create a new variable to measure caloric density. Furthermore, from our prior exploration, we have concluded that food has more nutritional value than beverages. still, this visualization has given us a valuable insight into most caloric items on the menu.

Furthermore, we also decided to remove the outlier of the 40 piece nugget. This is because there are also 20 piece, 10 piece, 6 piece, and 4 piece–all which have the same nutritional and price value–just multiplied.

Therefore, we determined that the 40 piece is normally ordered as a catering or shared dish, and not helpful in our exploration of single serving meals.

Calorie per Dollar and Caloric Density

After a simple exploration of the most caloric items, we determined that a better measure of value would be to ask what items had the most calories per dollar, and which were the most calorie dense per serving size.

First, for calorie per dollar–we mutated the data to create a new variable–calories over price. Then, we looked at the top ten items that have the most calories per dollar.

First, we had to change the serving size into a numeric variable, as it came in both oz and grams. We chose to use oz, as this accounts for fl oz as well with drinks.

```
serve_oz <- read.fwf(textConnection(as.character(cleanmenu$serving_size)), 4)
serve <- serve_oz[,1]
cleanmenu$serving_size_oz <- serve
```

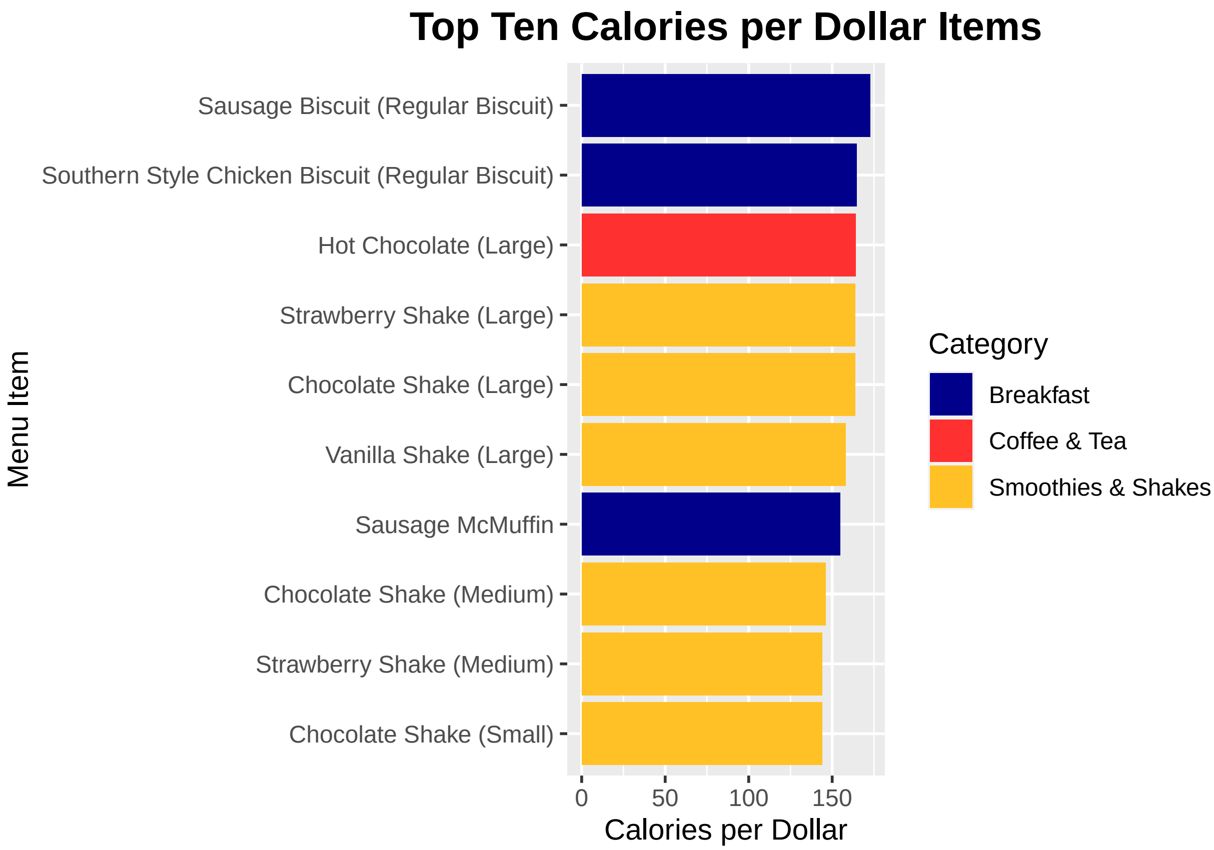
Next, we mutated the data to create our two new variables and add them to our data.

```
pricemenu <- cleanmenu %>%
  mutate(calorie_per_dollar = calories / price,
         cal_density = calories / serving_size_oz) %>%
  select("category", "item", "price", "calories", "calorie_per_dollar", "cal_density",
         "total_fat", "cholesterol", "sodium", "carbohydrates", "sugars", "protein",) %>%
  arrange(desc(calorie_per_dollar))

head(pricemenu)
```

```
## # A tibble: 6 x 12
##   category         item  price calories calorie_per_dollar cal_density total_fat
##   <chr>            <chr> <dbl>    <dbl>              <dbl>       <dbl>     <dbl>
## 1 Breakfast        Saus~  2.49      430               173.       105.         27
## 2 Breakfast        Sout~  2.49      410               165.        82          20
## 3 Coffee & Tea     Hot ~  3.29      540               164.        27          20
## 4 Smoothies & Sha~ Stra~  5.19      850               164.        38.6        24
## 5 Smoothies & Sha~ Choc~  5.19      850               164.        38.6        23
## 6 Smoothies & Sha~ Vani~  5.19      820               158.        37.3        23
## # i 5 more variables: cholesterol <dbl>, sodium <dbl>, carbohydrates <dbl>,
## #   sugars <dbl>, protein <dbl>
```

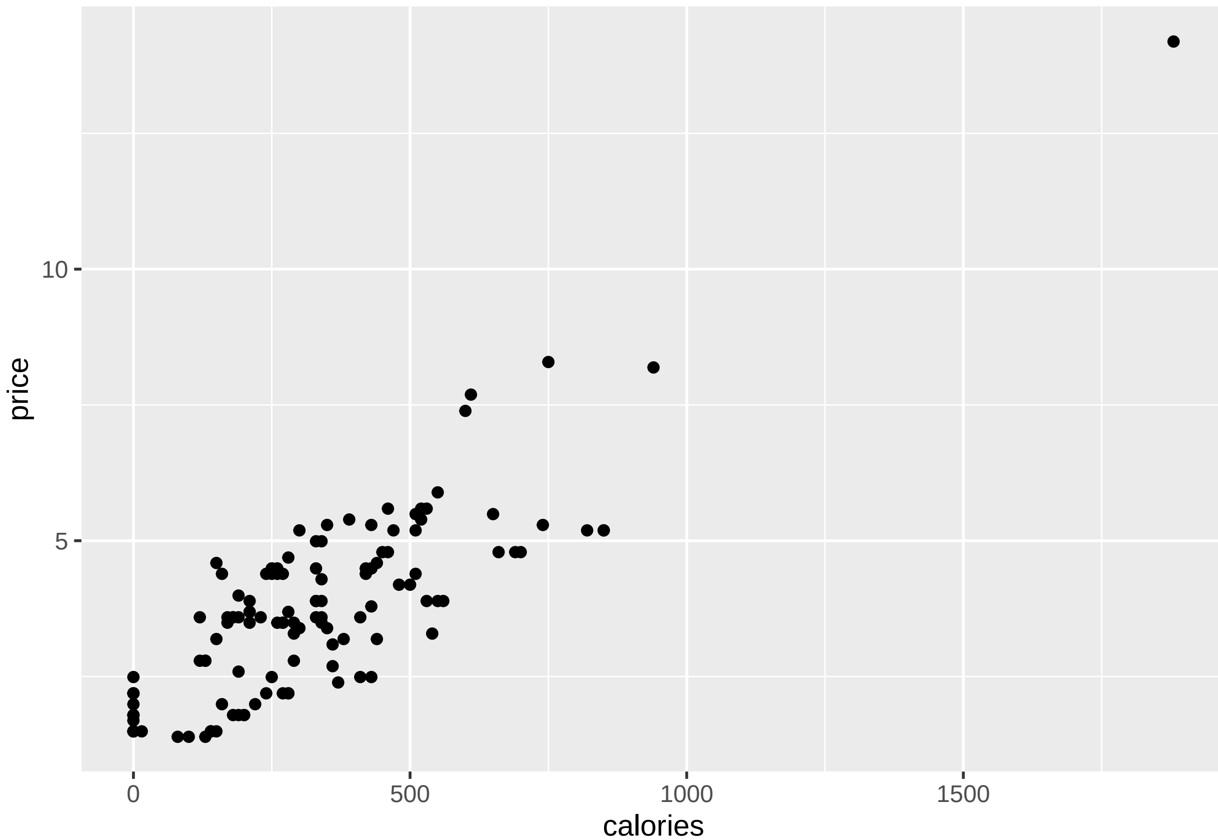
Now for some graphs with these new variables!

```
pricemenularge <- head(pricemenu, 10)

ggplot(pricemenularge, aes(x = calorie_per_dollar, y = reorder(item, calorie_per_dollar),
                           fill = category)) +
  geom_col() +
  labs(x = "Calories per Dollar", y = "Menu Item",
       title = "Top Ten Calories per Dollar Items", fill = "Category") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5)) +
  scale_fill_manual(values = c("darkblue", "firebrick1", "goldenrod1"))
```

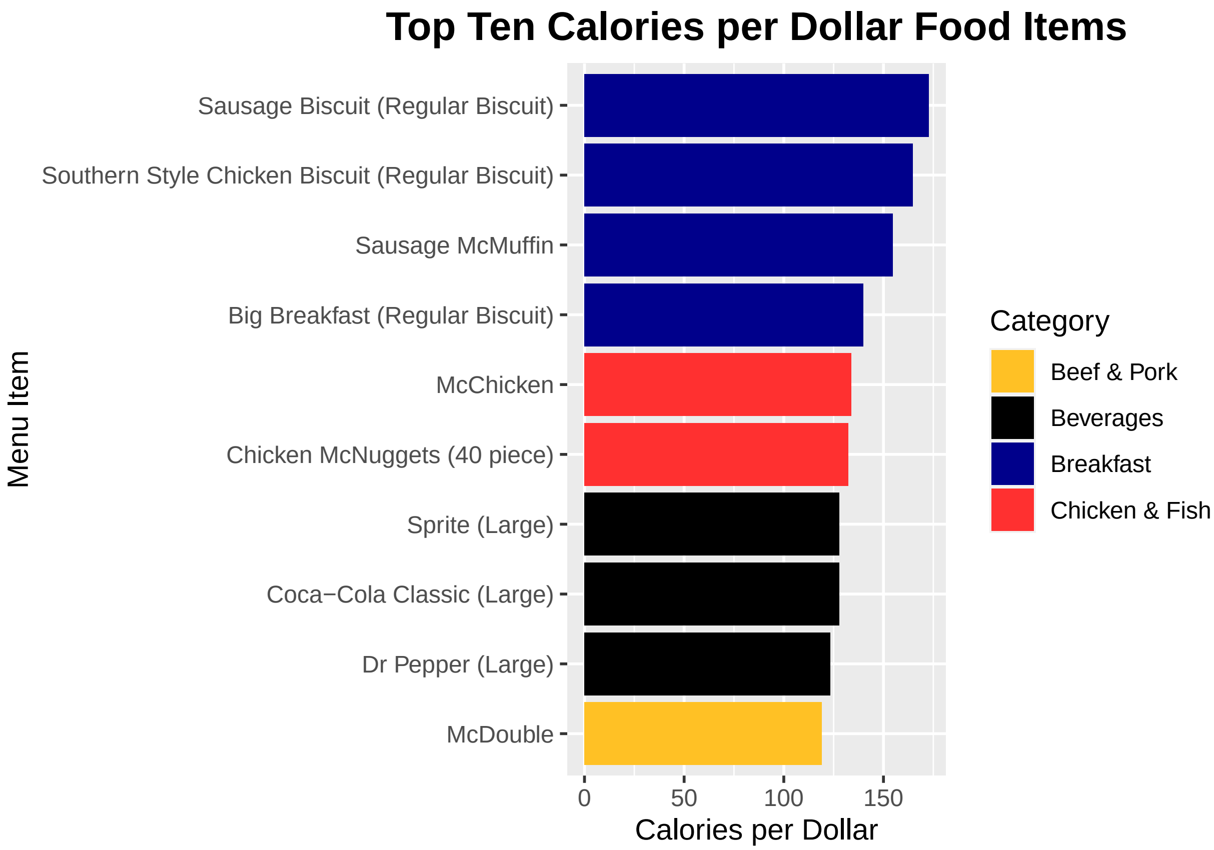


In both of these graphs, most of the items in the top ten are drinks, which are not the most nutritious, so while you may get the most calories for your dollar–we want to discover what food items would be the most bang for your buck. Therefore, we filtered for only food items.

```
ggplot(pricemenu, aes(x = calories, y = price)) +
  geom_point()
```



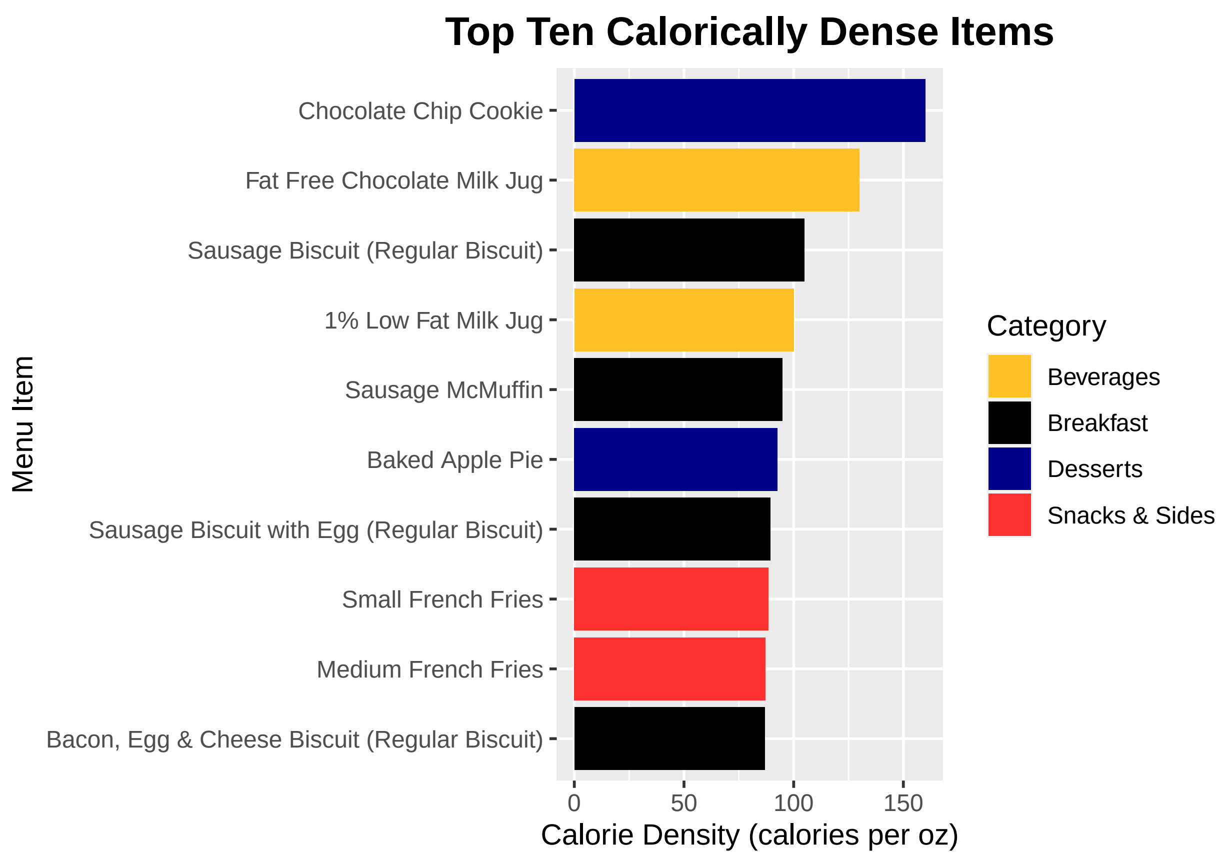

```
pricemenu %>%
  filter(category != "Smoothies & Shakes", category != "Coffee & Tea") %>%
  head(10) %>%
  ggplot(aes(x = calorie_per_dollar, y = reorder(item, calorie_per_dollar),
             fill = category)) +
  geom_col() +
  labs(x = "Calories per Dollar", y = "Menu Item",
       title = "Top Ten Calories per Dollar Food Items", fill = "Category") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5)) +
  scale_fill_manual(values = c("goldenrod1", "black", "darkblue", "firebrick1"))
```



From these graphs, breakfast appears to be the highest calorie per dollar value overall, so we chose to further investigate the nutritional value of a McDonald's breakfast.

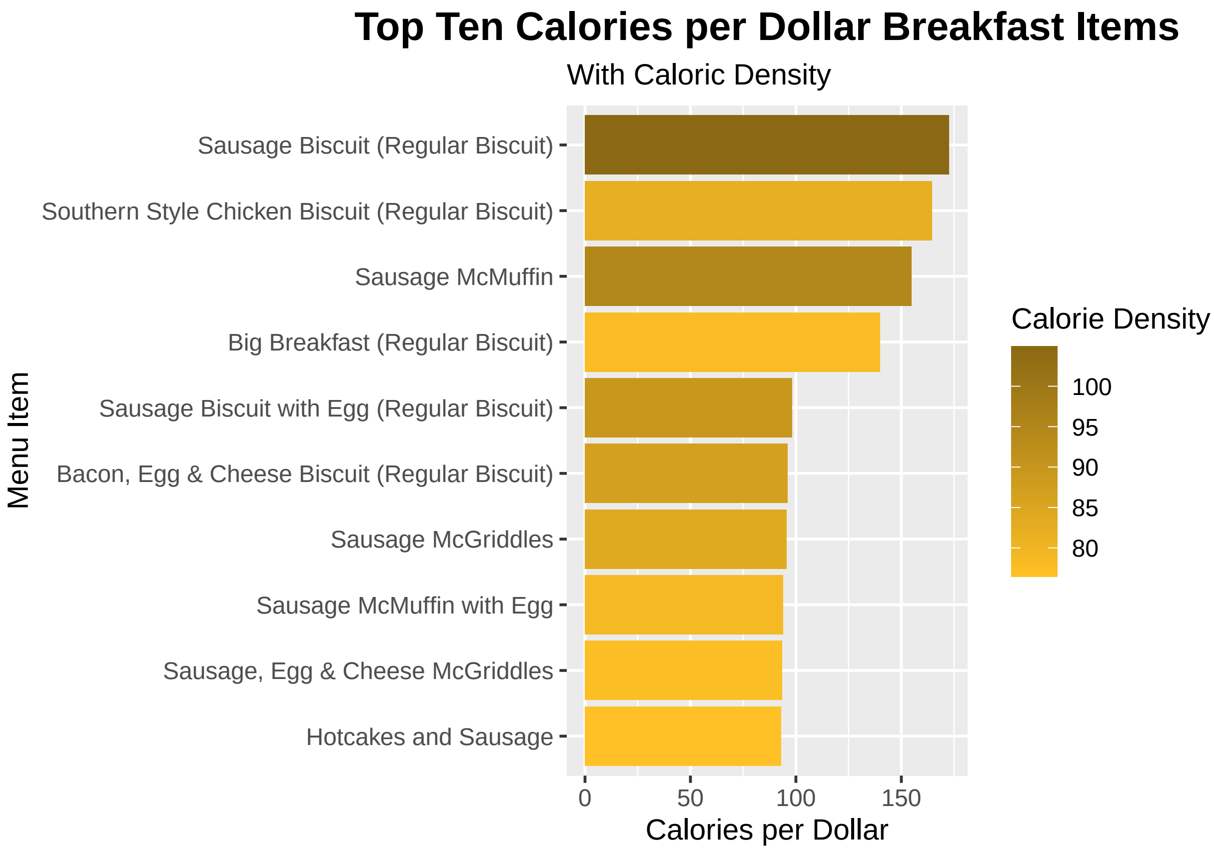
But first, we wanted to explore calorie density–to determine if those items were also the most valuable–or if while breakfast was the most calories per dollar, some other items would prove to be the most calorically dense.

```
pricemenu %>%
  arrange(desc(cal_density)) %>%
  head(10) %>%
  ggplot(aes(x = cal_density, y = reorder(item, cal_density), fill = category)) +
  geom_col() +
  labs(x = "Calorie Density (calories per oz)", y = "Menu Item", title = "Top Ten Calorically Dense Ite
       fill = "Category") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5)) +
  scale_fill_manual(values = c("goldenrod1", "black", "darkblue", "firebrick1"))
```



Once again, in this graph–breakfast further dominates the top ten–even if beverages were excluded. There are four/ten breakfast items, compared to only two each of the other categories. So we moved on to exploration of breakfast items at McDonald's.

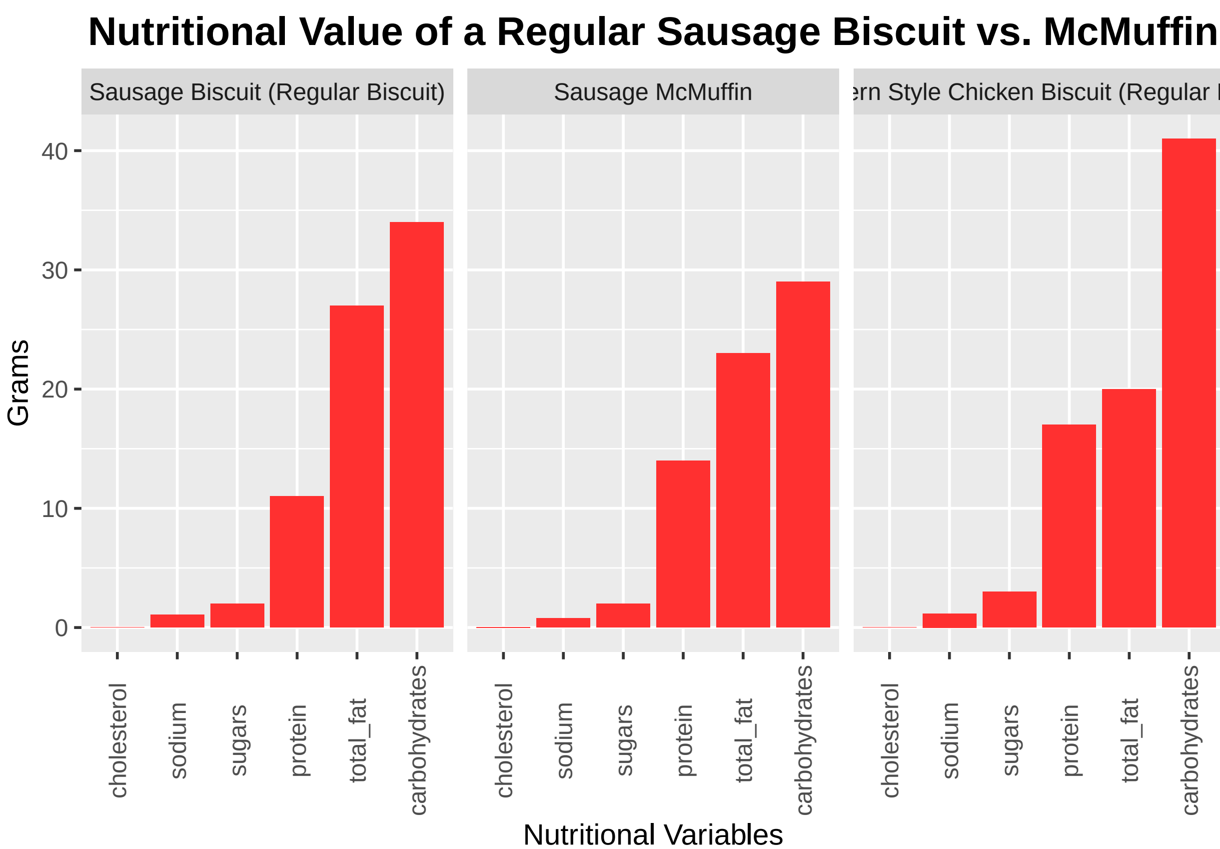
```
pricemenu %>%
  filter(category == "Breakfast") %>%
  head(10) %>%
  ggplot(aes(x = calorie_per_dollar, y = reorder(item, calorie_per_dollar))) +
  geom_col(aes(fill = cal_density)) +
  scale_fill_gradient(low = "goldenrod1", high = "goldenrod4") +
  labs(x = "Calories per Dollar", y = "Menu Item",
       title = "Top Ten Calories per Dollar Breakfast Items",
       subtitle = "With Caloric Density", fill = "Calorie Density") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5))
```



This visualization allows us to confirm that the regular sausage biscuit is the most cost effective, calorically dense, and most caloric option on the menu.

Yet, we also hoped to look into the nutrional density of each of these items. In this exploration, we standardized each of the nutritional items, this included making every item into grams, and also eliminating items like vitamins and trans fat–which were low to minimal, and not large building blocks of our diets.

```
pricemenu %>%
  filter(item %in% c("Sausage Biscuit (Regular Biscuit)",
                     "Southern Style Chicken Biscuit (Regular Biscuit)",
                     "Sausage McMuffin")) %>%
  mutate(sodium = sodium/1000) %>%
  mutate(cholesterol = cholesterol/1000) %>%
  pivot_longer(cols = total_fat:protein, names_to = "nutrition", values_to = "grams") %>%
  ggplot(aes(x = reorder(nutrition, grams), y = grams)) +
  geom_col(fill = "firebrick1") +
  facet_wrap(~item) +
  labs(x = "Nutritional Variables", y = "Grams",
       title = "Nutritional Value of a Regular Sausage Biscuit vs. McMuffin") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5)) +
  theme(axis.text.x = element_text(angle = 90, vjust = .5, hjust = 0.5))
```



We chose to compare nutritional values in order to determine if these calorically dense items are worth much nutritional value, or if they only provide value in the amount of caloric "bang for your buck" they give.

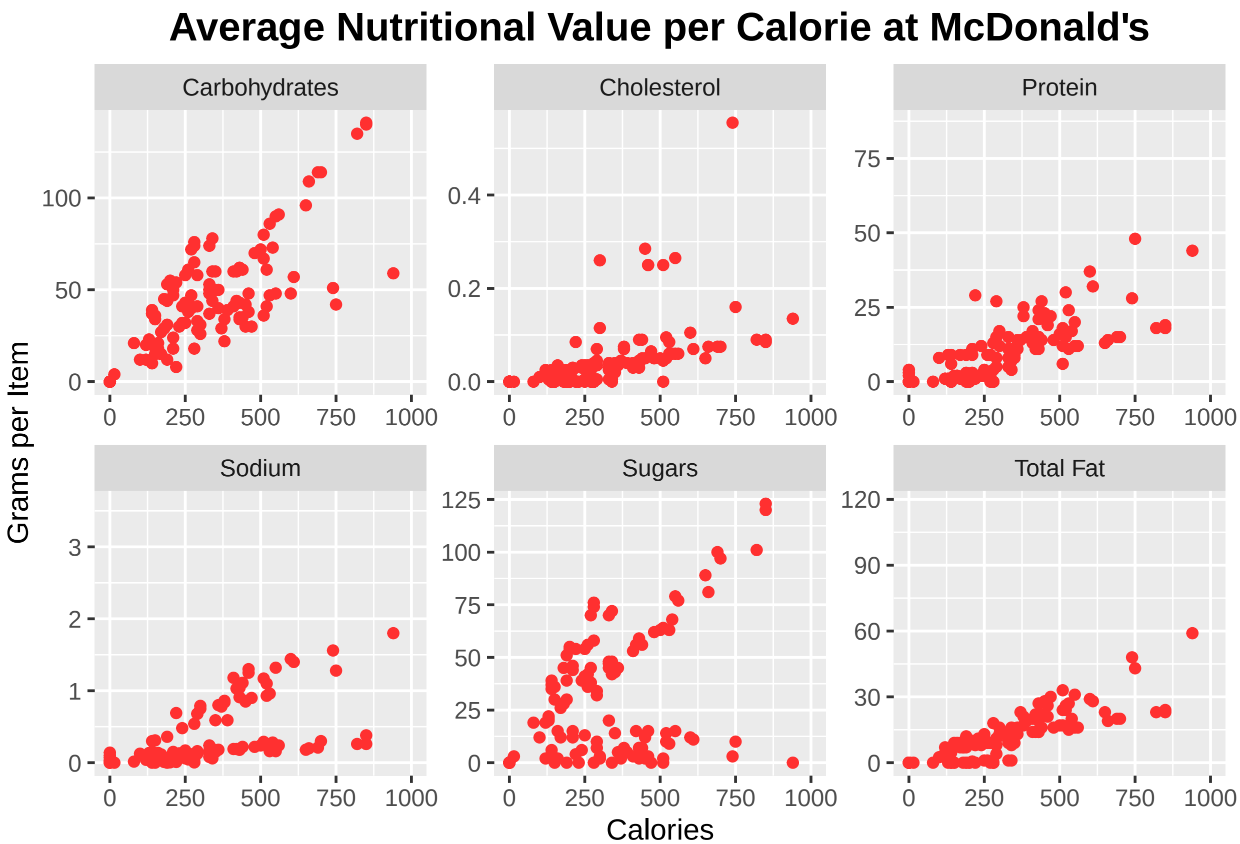
Then, we continued to explore the nutritional value of each McDonald's item.

Nutrition

We chose to display each of these facets on their own scales, as to see if there are any commonalities across different values.

```
longmenu <- cleanmenu %>%
  select("item", "calories", "total_fat", "cholesterol", "sodium", "carbohydrates",
         "sugars", "protein") %>%
  mutate(cholesterol = cholesterol/1000,
         sodium = sodium/1000)

longmenu <- setNames(longmenu, c("Item", "Calories", "Total Fat", "Cholesterol", "Sodium",
                                 "Carbohydrates", "Sugars", "Protein"))

longmenu <- pivot_longer(longmenu, cols = "Total Fat":"Protein", names_to = "nutrition",
                         values_to = "values")

ggplot (longmenu, aes(x = Calories, y = values)) +
  geom_point(color = "firebrick1") +
  facet_wrap(~nutrition, scales = "free") +
  labs(x = "Calories", y = "Grams per Item",
       title = "Average Nutritional Value per Calorie at McDonald's") +
  xlim(c(0,1000)) +
```

```
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5))
```

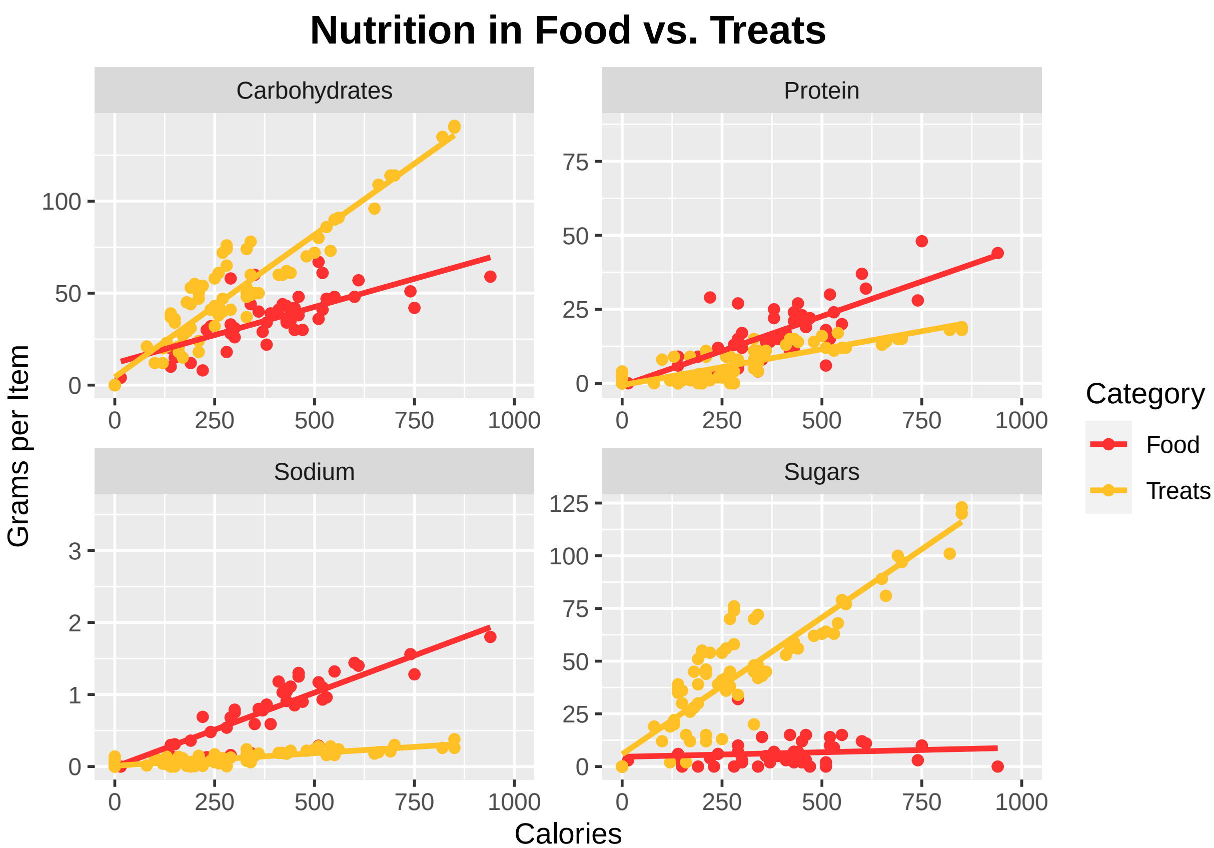


We've seen that there seems to be a trend of two lines in some of the facets, so we hoped to explore this more and determine what is causing these two distinct lines. For this test, we chose the categories sugars, sodium, carbohydrates, and protein to investigate. In this investigation, we chose to combine all drinks and all food, as we hypothesized that the two lines would be determined by these two separate categories.

Data Modeling

```
longmenu2 <- cleanmenu %>%
  select("category", "item", "calories", "sodium", "carbohydrates", "sugars",
         "protein") %>%
  mutate(sodium = sodium/1000) %>%
  setNames(c("Category", "Item", "Calories", "Sodium", "Carbohydrates", "Sugars",
             "Protein")) %>%
  pivot_longer(cols = "Sodium":"Protein", names_to = "nutrition", values_to = "values")

longmenu2[longmenu2 == "Breakfast"] <- "Food"
longmenu2[longmenu2 == "Beef & Pork"] <- "Food"
longmenu2[longmenu2 == "Chicken & Fish"] <- "Food"
longmenu2[longmenu2 == "Salads"] <- "Food"
longmenu2[longmenu2 == "Snacks & Sides"] <- "Food"
longmenu2[longmenu2 == "Beverages"] <- "Treats"
longmenu2[longmenu2 == "Coffee & Tea"] <- "Treats"
longmenu2[longmenu2 == "Smoothies & Shakes"] <- "Treats"
longmenu2[longmenu2 == "Desserts"] <- "Treats"
```

```r
ggplot(longmenu2, aes(x = Calories, y = values, color = Category)) +
  geom_point() +
  facet_wrap(~nutrition, scales = "free") +
  xlim(c(0,1000)) +
  labs(x = "Calories", y = "Grams per Item", title = "Nutrition in Food vs. Treats") +
  theme(plot.title = element_text(size = 15, face = "bold", hjust = 0.5)) +
  scale_color_manual(values = c("firebrick1", "goldenrod1")) +
  stat_smooth(method = 'lm', se = FALSE)
```



We chose to fit a linear regression model to the graph comparing the calories with grams per various nutritional measures - Carbohydrates, Protein, Sodium, and Sugar. Because we observed what looked like two separate linear trends in the data, we could model the different patterns with separate linear regression models. After color coding by the category of food, a clear pattern emerged between two classes of menu items that we labeled "Food" and "Treats". The Treats class consisted of categories that wouldn't be a sufficient or filling meal on their own, and were more sugar-based. This included the categories "Beverages," "Coffee & Tea," "Smoothies & Shakes," and "Desserts." All of the other categories of sandwiches, salads, and sides were grouped into the "Food" class.

We performed our linear regression and found that in carbohydrates and sugar, the Treats class had a steeper slope - in other words, the increase of one calorie in a Treats menu items predicted a sharper increase in grams of carbohydrates and sugar than was predicted for the Food menu items. On the other hand, when we examined sodium and protein, every one-calorie increase predicted a steeper increase in grams of sodium and protein in the Food items than in the Treats items. This pattern makes sense, given that the shakes, coffees, and beverages sold tend to be laden in sugar, while the sandwiches, salads, and sides include salty meats, which would be higher in sodium and protein.