Food Survey Investigation

Flemming Wu

Contents

Introduction

Insulin resistance and diabetes is a growing health issue for Americans. When foods with a high glycemic index (causing a rapid rise in blood sugar) are consumed, the pancreas must pump insulin to move sugar from the blood back into the cells. Over time, if these foods are consumed on a consistent basis, cells stop responding to insulin and the normal blood sugar level rises. This leads to weight gain, as excess blood sugar is sent to be stored as body fat, and sets the stage for prediabetes and type 2 diabetes.

While there are many other factors outside of diet that influence the development of insulin resistance and diabetes such as lifestyle, environmental factors, and family history, in this project, I will be investigating factors affecting our food choices using the NHANES (National Health and Nutrition Examination Survey) data. More specifically, I examined the data that was collected in What We Eat in America (WWEIA), the dietary interview component of the NHANES.

I also acknowledge that people's dietary requirements vary due to a variety of factors, but according to the CDC and other sources, people should generally be wary of continued consumption of foods high in added sugar and saturated fats. Therefore, in this project, I will use the data to investigate the following questions that I have asked:

1. What time and/or day of the week do people generally eat foods high in sugar or saturated fatty acids?
2. Does sugar or saturated fatty acid consumption vary by age, ethnicity, or gender?
3. Does the source of food or whether the meal was eaten at home have an effect on sugar or saturated fatty acid consumption?
4. What specific food items in the dataset are associated with high amounts of sugar or saturated fatty acids?

Methods

About the data

I used a total of four datasets for my project. The first two datasets contain answers to the food survey questionnaire, in which the respondents were asked to recall all food and drink they consumed in a 24 hour period. These questions were asked on two different days, with day 1 answers being one table and day 2 answers being the other table. Day 1 is collected in person in a mobile examination center and contains information from 12,634 respondents with a complete intake, including 242 individuals that were breastfed. Day 2 answers are collected over the phone 3 to 10 days following Day 1, and contains information from 10,830 respondents with a complete intake, including 203 individuals that were breastfed. This implies that not all respondents were recorded in both days. Observations, or rows, in the food survey data are separated into individual food or drink items, meaning each record is one food/drink item (identified by a USDA food code) consumed by a participant. Each record also identifies other information about the food/drink items such as the intake day of week, the time and name of the eating occasion for the food, whether food was eaten in combination with other foods (for example, milk added to cereal), amount of food consumed, and amounts of food energy and 64 nutrients/food components provided by the food amount.

The next data set I used contains general demographic information about each of the participants, such as age, gender, ethnicity, pregnancy status, language used for interview, among others. The last data set I used contains descriptions of food information. Since the food items in the food survey questionnaires were encoded as numbers, I used this table to cross-reference the food code numbers with descriptions of the food or drink items.

Data Preparation

The data provided on the website were in SAS Transport File Format, so I used the `haven` package to read in the data directly from the http link. Once I read in the data into R, I noticed that the column names were encoded with names that weren't intuitive such as "WTDRD1PP", but the data sets also contained column labels which explained the meanings of the column names. I did some text processing on the labels, such as removing non-alphanumeric characters and removing spaces, and then set these as the column names to make downstream work easier. I then noticed that all of the categorical variables in the data were encoded with numbers, such as a 1 for yes or a 2 for no. To fix this, I went through the data set documentation and updated the categorical observations with their actual character values. I then added a column to each of the food survey data tables to keep track of which day the answers were from and then concatenated the data from day 1 and day 2. Lastly, I merged all of the data into one data table, using the respondent id numbers and food code numbers as the common keys.

Results

Summary Tables

To view the summary tables, please refer to the corresponding on the website: click here. In the interest of keeping this report succinct, I will only be providing the interpretations here.

Table 1 | *Summary of the average sugar and saturated fatty acid consumption by the day of the week in which they were consumed.* According to Table 1, average sugar and saturated fatty acid consumption appears to be highest on Friday, Saturday, and Sunday. Tuesday is the day with the lowest average sugar and saturated fatty acid consumption.

Table 2 | *Summary of hourly average sugar and saturated fatty acid consumption.* The number 0 corresponds to all items consumed between 12:00 AM and 12:59 PM, 1 corresponds to all items consumed between 1:00 AM and 1:59 AM, etc. It appears that average sugar and saturated fat consumption is highest between the hours of 8:00 PM and 3:00 AM.

Table 3 | *Summary of average sugar and saturated fatty acid consumption by eating occasion.* Eating occasions that are more considered to be more formal meals, such as dinner, lunch, almuerzo (breakfast), desayano (breakfast), supper, etc. generally involve less consumption of sugar than do informal eating occasions such as snacks (including bocadillo, botana, merienda, entre comida). The reverse is true for saturated fats consumption, as the average grams consumed for these are slightly higher in more formal eating occasions.
Note: the data I used (2017-2020) did not provide English translations for Spanish meal names, but they were provided in the 2003-2004 NHANES data documentation, which is what I used for reference in this analysis:

Spanish	English
Desayano	Breakfast
Almuerzo	Breakfast
Comida	Lunch
Merienda	Snack
Cena	Dinner
Entre comida	Snack
Botana	Snack
Bocadillo	Snack
Tentempie	Snack
Bebida	Drink

Table 4 | *Summary of average sugar and saturated fatty acid consumption by age.* Consumption increases from birth until the ages of 14-18, and then decreases after age 18. Age group 14-18 sees the highest average sugar and saturated fatty acid consumption.

Table 5 | *Summary of average sugar and saturated fatty acid consumption by gender.* Males consume higher amounts of sugar and saturated fatty acids on average, which makes sense considering that generally males have larger bodies than females, requiring them to consume more calories on average.

Table 6 | *Summary of average sugar and saturated fatty acid consumption by race.* The groups with the highest average sugar and saturated fat consumption are "other race including multiracial" and "non-hispanic black", and they are closely followed by "non-hispanic white". The "non hispanic asian" group consumes less sugar and saturated fats on average than other groups.

Table 7 | *Summary of average sugar and saturated fatty acid consumption by source of food.* High average sugar consumption can be seen to come from food sources such as vending machines and convenience stores, likely due to the sale of sugar-sweetened beverages. High saturated fatty acid consumption can be seen to come from food sources such as fish (likely due to naturally occurring omega fatty acids in fish) as well as fast food restaurants and recreational facilities.

Table 8 | *Summary of average sugar and saturated fatty acid consumption by whether meal eaten at home.* Foods that were not eaten at home are generally slightly higher in sugars and saturated fatty acids content on average, however, the difference is quite small.

Figures

To interact with the plots, please refer to the corresponding page on the website: click here. Supplementary figures and their interpretations can be found on this page: click here

Figure 1

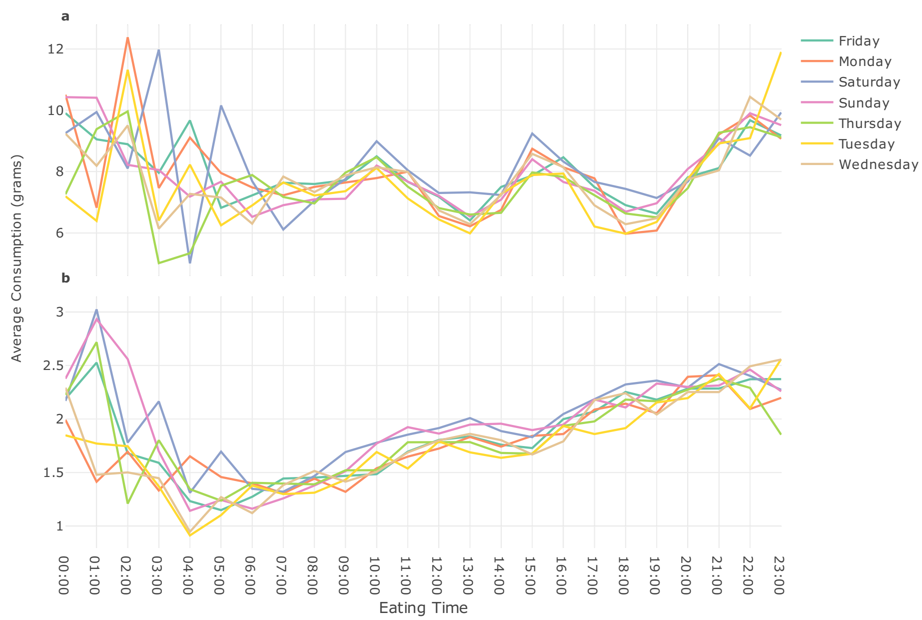

Figure 1 | *(a) Hourly average sugar consumption grouped by day of the week.* Sugar consumption starts off lower at midnight on Tuesday and Thursday, and are higher for the other days of the week. Then, the lines seem to converge at around 7:00 AM, and from there there are spikes in sugar consumption at around 10:00 AM, 3:30 PM, and 9:00 - 11:00 PM. *(b) Hourly average saturated fatty acid consumption grouped by day of the week.* Average saturated fatty acid consumption varies less by day of the week with the lines converging at around 5:00 AM and steadily increasing for the next 21 hours.

Figure 2

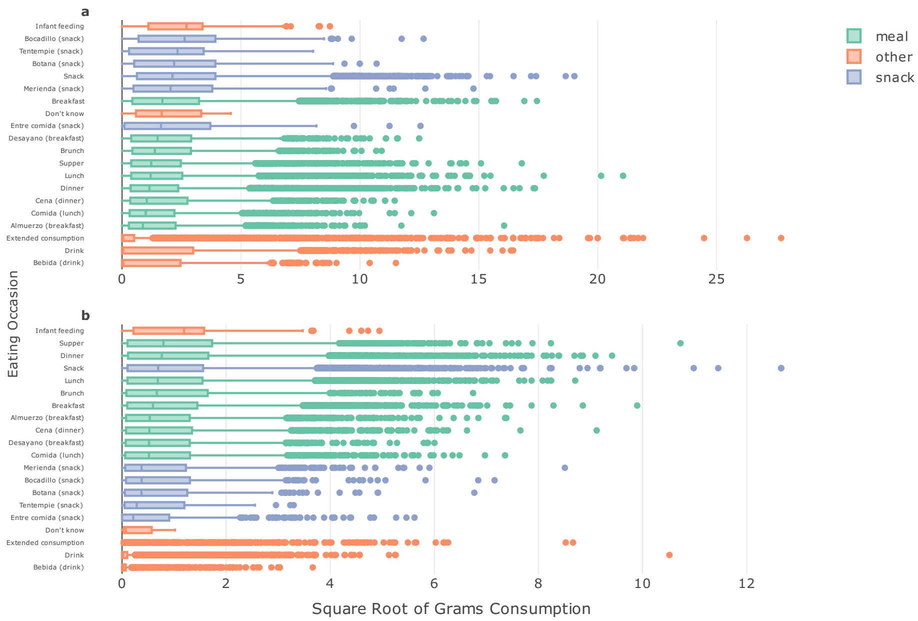

Figure 2 | *(a) Distribution of sugar consumption (square root transformed) for each eating occasion.* Eating occasions at different times of the day, such as breakfast, lunch, and dinner do not see much difference in sugar. However, the distribution of sugar consumption is lower in formal meals (such as breakfast, lunch or dinner) than in snacking occasions. *(b) Distribution of saturated fatty acid consumption (square root transformed) for each eating occasion.* Eating occasions at different times of the day, such as breakfast, lunch, and dinner do not see much difference in saturated fatty acid consumption either. While snacking occasions typically involve much higher average sugar consumption. The reverse trend can be seen for average saturated fatty acid consumption, as snacks have smaller distributions of saturated fatty acid consumption than meals.

Figure 3

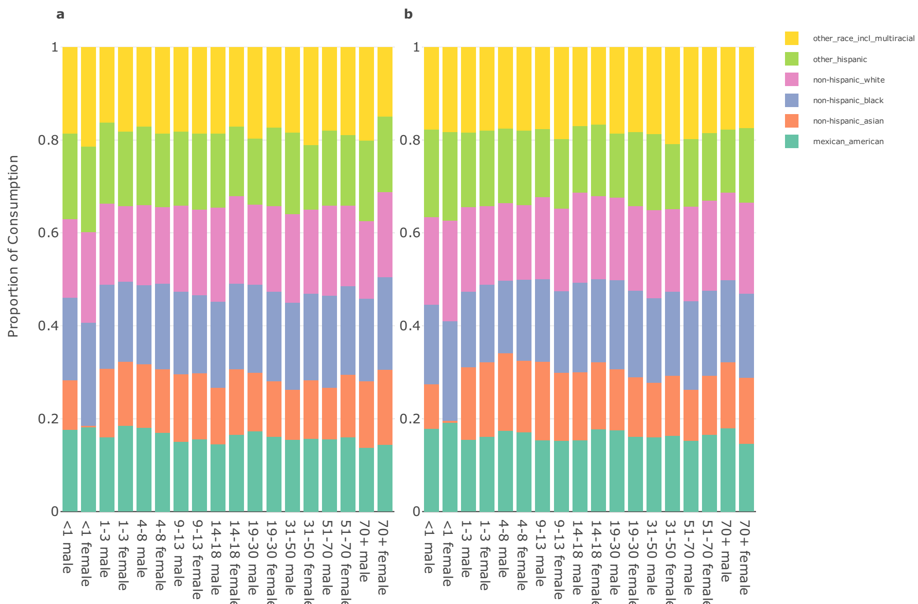

Figure 3 | *(a) Proportion of total sugar consumption within each age and gender category by race. (b) Proportion of total saturated fatty acid consumption within each age and gender category by race.* The figure above plots the proportion of sugar or saturated fatty acid consumption within each ethnicity category, separated by age and gender. The proportion of consumption appears to be relatively similar across all ethnicity groups, with each group having about 17% of the consumption. The non-hispanic asian group seems to be slightly less than others, however, with an average of about 13% of the proportion.

Figure 4

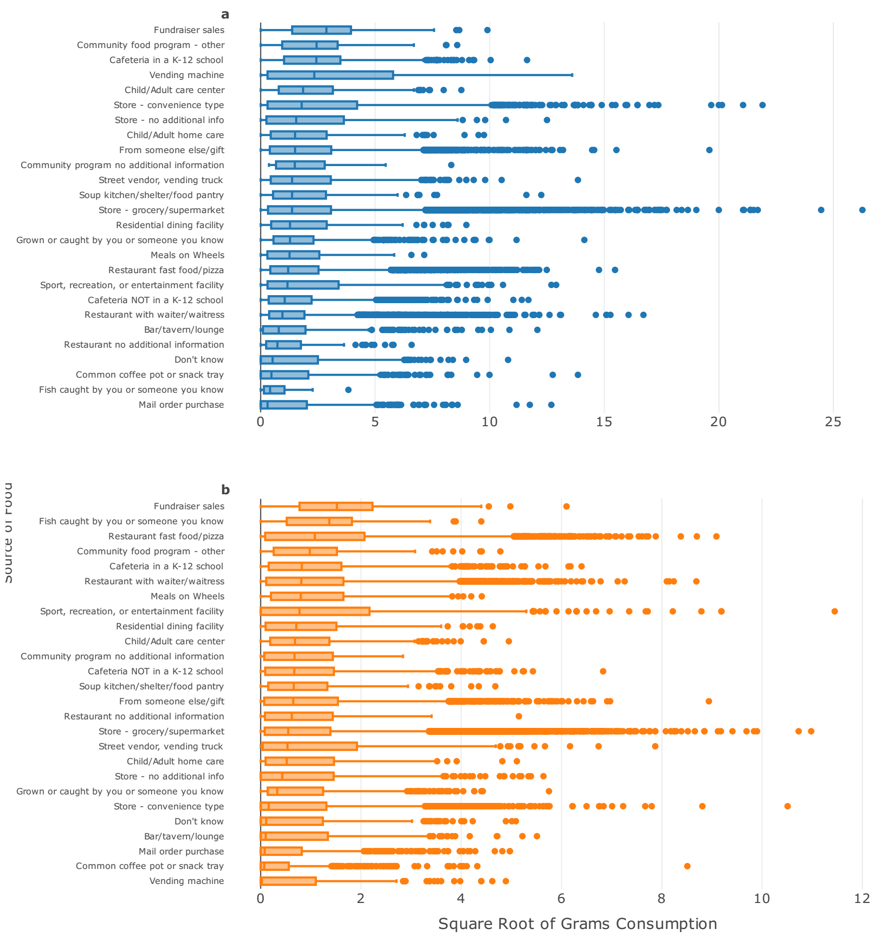

Figure 4 | *(a) Distribution of sugar consumption (square root transformed) for each source of food. (b) Distribution of saturated fatty acid consumption (square root transformed) for each source of food.* It is interesting to note that the median for both sugar and saturated fatty acid distributions is highest from food obtained from fundraiser sales. School cafeterias also are prominent sources of high sugar and saturated fatty acid consumption. Additionally, the highest consumption of sugar and saturated fatty acid consumption comes from the "Store - grocery/supermarket" category, as this category has the most large outliers. Moreover, the food category "Fish caught by you or someone you know" has a high distribution in saturated fatty acids, likely due to fish containing high amounts of omega fatty acids.

Conclusion and Summary

Out of all of the variables that I have investigated in this project, I have found that eating occasion, source of food, time of consumption, and age have stronger relationships with sugar and saturated fatty acid consumption than the other variables. Consumption of foods high in sugar or saturated fatty acids does not vary significantly by day of the week, but sugar consumption is highest between 8:00 AM and 4:00 AM, and saturated fatty acid consumption is lowest at 6:00 AM, and steadily increases throughout the day. (Figure 1) Furthermore, people in this data set consume higher amounts of sugar during snacking occasions and higher amounts of saturated fatty acids during meals rather than snacks. (Figure 2) Additionally, the sugar and saturated fatty acid content is fairly consistent between the foods eaten by people in the different ethnicity categories. However, there was a trend seen when looking at sugar and saturated fatty acid content by age. It seemed that the consumption of these ingredients increased from birth, and peaks at people ages 14-18, then starts to decline. (Figure 3) Moreover, foods that came from fundraising sales, community food programs, and K-12 school cafeterias had the highest distribution of foods high in sugars and saturated fatty acids. (Figure 4) Lastly, sugar-sweetened beverages seems to be the most common source of high intake of sugar, whereas the common foods associated with high saturated fatty acids include pizza, ice cream, milkshakes, and burgers. (Supplementary Figure 2)

Further Directions

In further analyses it would be interesting to see if any of the variables that I have suggested to have an association with sugar or saturated fatty acid consumption are correlated with each other. From there, it would also be interesting to work with more personalized information about the participants, such as primary place of residence, height, weight, income, access to food, etc.

Sources

- https://www.cdc.gov/diabetes/index.html
- https://www.ars.usda.gov/northeast-area/beltsville-md-bhnrc/beltsville-human-nutrition-research-center/food-surveys-research-group/docs/wweia-documentation-and-data-sets/