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 (fa)?**
2. **Does sugar or saturated fa 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 fa consumption?**
4. **Finally, what specific food items are associated with high amounts of sugar or saturated fa?**

About the data I used a total of four data sets for my project. The first two are answers to a 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 one two different days, with day one answers being one table and day two answers being the other table. Not all respondents were recorded in both days. Observations, or rows, in the food survey data are separated into individual food or drink items and also includes estimates on

how much of each item was consumed, as well as energy and nutrient estimates for each item. Participants were asked additional questions about their consumption, such as what time the item was consumed, what meal it was a part of, whether the meal was eaten at home, etc. The next data set I used contains general demographic information about each of the participants, such as age, gender, ethnicity, etc. 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.

Methods

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.
