Relationship Between Nutrition Elements Intake and Income

Run Yu

Data Analytics

Tufts University

16 December 2019

ABSTRACT

Health disparities are pervasive among Americans. It has been studied that individuals with higher socioeconomic status have a lower rick of having the chronic diseases. Meanwhile, Diet has been known as a contributor of the chronic diseases. Dietary preferences have changed in recent decades with trends of the healthy diet in order to prevent the chronic diseases. However, a healthier diet requires higher cost since added sugar and fat are more affordable. As a result, socioeconomic disparities in diet occur. The purpose of this study is to 1) discuss the relationship between the income and the intake of five nutrition elements including sugar, fiber, fat, carbohydrate and protein, 2) examine which nutrition elements intake is affected by the income, 3) help create comprehensive strategies to reduce problematic dietary patterns among all sociodemographic groups in the United States.

1. INTRODUCTION

Socioeconomic status is frequently implicated as a contributor to health disparities in the United States. Individuals with lower incomes tend to consume lower quality diets since healthier diets which defined as higher quality diets are associated with higher diet cost. On the other hand, individuals with higher incomes are more likely to have healthy dietary intakes. They are able to spend extra money on their health and willing to keep higher quality diets. Meanwhile, diet quality plays an important role in health disparities. The goal of this study is to find out the relationship between the income and the intake of nutrition elements including protein, sugar, carbohydrate, fat and fiber by first providing an overview of the dietary patterns among individuals within different income levels. Additionally, this study will use the regression model to answer the question that which nutrition elements intakes are significantly affected by the income.

Previous researches provide evidence of relatively high cost of healthy food and well established that added sugar and fat are more affordable, this study makes the hypothesis that individuals with higher income are going to consume higher quality diets with more protein, more fiber, less sugar, less fat and lower carbohydrate. Also, this study assumes the income plays the most significant role in affecting the dietary intakes. The result of the study can provide insights of the dietary patterns among different income groups. Meanwhile, the insights can help restaurants located in different income areas understand the need of their potential customers so that they can modify their menus in order to gain more profits. Also, if the result of the regression model shows that the income does have the most significant impact on dietary intakes, the government can use the result to come out with the opportune policy to help improve dietary qualities of individuals in different income levels as much as possible.

This study uses the most recent data from National Health and Nutrition Examination Survey (NHANES) 2015-2016 to access the intake of nutrition elements in 24 hours among different income groups. The samples involved in this study are the participants who are older than 25 and younger than 50 years old since individuals outside of this year range are mostly students or elderly group who do not have representative income. The measurement used in this study is the daily nutrition goal acknowledged by the U.S. Department of Health and Human Services and the U.S. Department of Agriculture. Comparing the intakes of the participants with the standards provides the overview of different dietary intake qualities among low-income, middle-income and high-income groups. In addition, an examination of the income impact on the nutrition intakes is conducted using linear regression models. By adjusting model variables and calculating the elasticities of the variables, the relationship of income and nutrition intakes and the significance of income in affecting the intakes can thus be shown.

The rest of this paper is structured as follows. Section 2 presents a summarized review of the existing researches about the sociographic influence on the dietary intake quality. Section 3 provides detailed description and visualization about the data used in this study. Section 4 develops the model we use to estimate the impact of income on nutrition elements intake. Section 5 includes the results of this study and the discussion about the results including some limitations involved in the study. Section 6 summarizes the results and raises policy implications of the results including some future work.

2. LITERATURE REVIEW

A Aggarwal et al. (2001) used mediation analysis to indicate that the relation between the diet quality and socioeconomic position can be mediated by the diet cost. Using the result of this study, I make the hypothesis that higher quality diets have higher cost and thus cause problems for low-income groups. Additionally, Adam and Nicole (2005) observe that low-income households care more about cost and taste of the food instead of nutrition. Those family attempt to consume more energy-dense food and food with more carbohydrate, added sugar and added fats. These two studies set up the background for my study. Combining the results of two studies, my study sets the topic dietary patterns among different income groups.

Hazel et al. (2012) find that the dietary pattern varies among different sociodemographic groups. The background of the study indicates that the diet quality closely relates to the chronic diseases. The poor diet quality is the leading cause of many chronic diseases. The research aimed to find out the diet quality variances among Americans with different sociodemographic characteristics including age, sex, race, income and education level. The data used in this study is also the 24-hour intake dietary data from NHANES but from 2003 to 2004 (versus 2015 to 2016). Instead of comparing the intake data with the standard directly, the study uses Healthy Eating Index-2005 (HEI-2005) components. For each component, the intakes are compared with standards which reflect the Dietary Guidelines for Americans and the HEI-2005 scores are then been calculated by adding all components together. The HEI-2005 scores are used as the measurement of the dietary quality of participants with diverse sociodemographic characteristics. The findings of this study imply that problematic dietary patterns actually occur among different sociodemographic groups. If the individual is child or older adult, woman, Hispanic, and in higher income level, he or she will have better diet qualities. Similar with my study, the authors provide descriptive results indicating the current dietary patterns instead of prescriptive methods to modify the patterns. My study differs from this study in three ways. First, my study chooses the nutrition elements like fiber and carbohydrate instead of types of food such as vegetables and whole grain used in Hazel's study. Food type cannot provide detailed information about the dietary quality since the food preparation process is unknow. For example, the broccolini can be either boiled or fried while the fried one contains more fat and is less healthy. Second, my study focuses on the income impact on nutrition intakes while Hazel's study evaluates sociodemographic characteristics individually. Building and adjusting a model in my study help me determine whether income has significant effect or not. Additionally, children and older adults are excluded in my study since these two groups of individuals may not have meaningful and stable income.

The study which is more similar to my study is Kirkpatrick et al. (2011). The goal of their study is also to explore the dietary intakes patterns among different sociographic groups. They find that if an individual is in lower income level and is non-Hispanic black, he or she will have unhealthier dietary intake pattern. Compared with Hazel et al. (2012), Kirkpatrick et al. (2011) not only include the food groups but also the nutrition elements such as solid fats and added sugar. The study of Kirkpatrick et al. (2011) is related to my study in two ways. First, they use 24-hour intake data form NHANES but from 2001 to 2004 instead of 2015 to 2016 and compare the intake with the Dietary Guidelines for

Americans and Mypyramid standard directly in order to see the dietary performance among different groups. Second, this study also divides the income into three levels including low, middle and high levels. The method they use provides me the recommendation to divide the samples in my study. Different with my study, Kirkpatrick et al. (2011) do not build the regression model to evaluate the impact of each variable on dietary intake. My study is adding a further step to their study in order to evaluate the income effect.

3. DATA DESCRIPTION AND VISUALIZATION

3.1 NHANES background & sample creation

The data used in this study is from the National Health and Nutrition Examination Survey (NHANES), a unique survey that combines interviews and physical examinations. The interview part asks about demographic, socioeconomic, dietary, and health-related questions. The examination part includes medical, dental, physiological measurements, and laboratory. The goal of the program is to access the health and nutritional status of adults and children in the United States. The NHANES program began in the early 1960s, in this study, I use the data from 2015 to 2016 which is the most recent one.

The sample in this study is consisted of interview data in order to have detailed dietary intake information from NHANES participants. All participants are required to take two 24-hour dietary recall interviews. The first dietary recall interview is collected in-person in the Mobile Examination Center (MEC) and the second interview is collected by telephone 3 to 10 days later. First. this study focuses on the intake of sugar, fat, carbohydrate, fiber and protein, the data reflecting participants' intake of these five nutrition elements within 24 hours on the first day is used as sample. (Table 1) This study does not use the second-day date because some diet changes may happen during 3 to 10 days after the first interview so that diet habits cannot be reflected without bias.

In order to collect variables which will affect the intake of nutrition elements, this study also include participants' demographics information including gender, age, country of birth, race, education level and annual household income. (Table 2) In the study, the education level variable is changed into two dummy variables, HighSchool and College, representing whether been to high school and college or not. Additionally, this study needs age range from 25 to 50 so other participants outside of this age range are dropped from the sample.

The income variables from NHANES is a categorial variable with each categorial value representing an income range. This study takes the median value of each category and assigns back to the categorical value. Then the income becomes a continuous variable ranging from 2,500 dollars to 125,000 dollars. Pew Research Center defines "middle income" households as those with an income two-thirds to double the median income for that household size. In 2015, low-income ranges from 0 to 41,869 dollars. Middle-income ranges from 41,869 to 125,609 dollars. Using Pew Research Center standard, the income has been divided into three subgroups, low-income group, middle-income group and high-income group. Additionally, since linear regression models Model 1 and Model 2 in this study focus on the income elasticity instead of coefficient value, income will not be standardized for these two models. Model 3 cares about the coefficient value of income, income variable is thus standardized in order to have meaningful coefficient values instead of zero values.

3.2 Comparison standard

In order to see the diet quality in each income group, this study needs to compare the intake in the sample with a dietary standard. 2015-2020 Dietary Guidelines for Americans acknowledged by U.S. Department of Health and Human Services and U.S. Department of Agriculture is used in this study as the measurement of whether the intake is healthy or not. (Table 3)

3.3 Comparison method

For each income group, this study calculates the percentage of participants who have reached the least standard of the nutrition intake. This method is repeated for all of five nutrition elements including carbohydrate, sugar, fiber, fat and protein.

3.4 Sample data visualization

The sample has 2,044 observations of 12 variables. (Table 4) There are five dummy variables including Gender, CountryOfBirth, Race, HighSchool and College. The relationship between income and intake of five nutrition elements can be seen in Graph 1a, 1b, 1c, 1d, 1e. Based on the race, the sample is divided into Hispanic group and non-Hispanic group. For each group, the relationship between income and intake of five nutrition elements can be seen in Graph 2a, 2b, 2c, 2d, 2e, 3a, 3b, 3c, 3d, 3e.

4. MODELS

This section will present the regression model development in order to explore the relationship between income and intake of five nutrition elements. The dependent variables are Protein, Sugar, Carbohydrate, Fat and Fiber. The main model (Model 1) includes all variables, Age, Gender, CountryOfBirth, Income, Race, HighSchool and College, while each of them may have impact on the intake of five nutrition elements. The main model is presented as following

Model 1:

$$Y_i = \beta_0 + \beta_1 Age_j + \beta_2 Gender_j + \beta_3 CountryOfBirth_j + \beta_4 Income_j + \beta_5 Race_j + \beta_6 HighSchool_j + \beta_7 College_j + u_j$$

j = 1, …, 2044
Y = {1: Protein, 2: Sugar, 3: Carbohydrate, 4: Fat, 5: Fiber}

where j represents each participant, β_0 is the intercept and u_j is the error term. There are actually 5 sub models in the main model while each sub model is used to show the relationship between each nutrition elements with affecting variables. β_4 measures the change in the conditional expectation of nutrition intake for a marginal change in income holding all other regressors constant. Another continuous variable in the model is age. β_1 measures the change in the conditional expectation of nutrition intake for a marginal change in age holding all other regressors constant.

Actually, the education level can be affected by the income. Individuals who have higher income may have more chances to achieve higher education levels. Those individuals are able to pay the tuition and they may need to receive more education in order to fit in the high-income careers. In the next model (Model 2), the education level effect is removed from the main model. The model excluding HighSchool and College is as following

Model 2: (No Education Level Effect)

$$Y_i = \beta_0 + \beta_1 Age_j + \beta_2 Gender_j + \beta_3 CountryOfBirth_j + \beta_4 Income_j + \beta_5 Race_j + u_j$$

j = 1, …, 2044

Y = {1: Protein, 2: Sugar, 3: Carbohydrate, 4: Fat, 5: Fiber}

It is expected that there will be an increase in β_4 meaning that income have more impact on the nutrition intake. The change in the conditional expectation of nutrition intake for a marginal change in income holding all other regressors constant will be larger.

For the next step, this study considers about the race influence. Individuals from different race will have diverse cultures while they may have different income allocation preferences. For example, a Hispanic individual may want to consume more fiber when he or she has higher income while a non-Hispanic individual may begin to intake more protein when the income level becomes higher. In Model 3, variable income is intersected with variable race and is shown as following

Model 3: (No Education Level Effect, Race x Income)

$$Y_i = \beta_0 + \beta_1 Age_j + \beta_2 Gender_j + \beta_3 CountryOfBirth_j + \beta_4 Income_j + \beta_5 Race_j + \beta_6 Race_j \cdot Income_j + u_j$$

j = 1, …, 2044

Y = {1: Protein, 2: Sugar, 3: Carbohydrate, 4: Fat, 5: Fiber}

β_4 measures the change in the conditional expectation of nutrition intake of a non-Hispanic for a marginal change in income holding all other regressors constant. β_6 measures the difference of changes in the conditional expectation of nutrition intake as a Hispanic compared with a non-Hispanic for a marginal change in income holding all other regressors constant.

5. EMPIRICAL ANALYSIS

5.1 Discuss the results

5.1.1 Dietary intake patterns among income groups

After generating the proportions of individuals in each income group whose intake of nutrition element met minimum of dietary guide requirement, the results are shown in Table 5. First, among all income groups, the percentages of individuals who have reached the fiber minimum intake standard are around 12% which are very low. This indicate that intake of fiber is problematic regardless of income levels. However, the proportion of individuals who have met the fiber intake standard in high-income group is 15.13% which is higher than middle-income (12.95%) and low-income (12.92%) groups. As the income increases, proportion of individuals who have met the fiber intake standard becomes larger. The first reason here is that food with high fiber tends to be more expensive which will cause problems for low-income individuals. Also, according to the study of Adam and Nicole (2005), it is observed that low-income households care more about cost and taste of the food instead of nutrition so that they prefer food with more carbohydrate, added sugar and added fats which help full up instead of more fiber which helps digest. Furthermore, individuals have higher income may have more access to high fiber food. They may care more about healthy diets and their working or living areas may be surrounded by more healthy food restaurants.

There is a large difference between low-income group and high-income group in proportions of individuals who have met the protein intake standard. As the income increase, it shows an increasing trend of proportion of individuals who have met the protein minimum intake standard. Same as fiber, protein tends to be more expensive compared with sugar, carbohydrate and fat.

For intakes of carbohydrate and sugar, there are no increasing trends in proportions of individuals who have satisfied the minimum intake standards as the income increases. The result has been showed that individuals from all income groups are consuming lots of carbohydrate and sugar. One of the reasons will be that the prices of carbohydrate and sugar are not as high as protein or fiber and they are affordable for all income groups. Additionally, carbohydrate and sugar provide energy for individuals to maintain full which is exactly what the individuals from low-income group need.

Although there is an increasing trend of proportion of individuals who have reached the minimum fat intake standard as the income level becomes higher, there are some bias for the variable fat. Fat can be either healthy fat or unhealthy fat. The dataset used in this study does not specify which kind of fat is the participant intaking. Healthy fat costs more than unhealthy fat so that there is a probability that individuals with higher income may consume more healthy fat. Meanwhile, the individuals from low-income group may not have many access to healthy fat so they consume lots of unhealthy and cheap fat which can cause problems in their diet patterns.

Overall, the intake of fiber is the most problematic one compared with other dietary intake patterns. Also, the income does affect fiber consumption since there is a larger proportion of individuals who have met the fiber intake standard in the high-income group compared with two other income groups.

In the next section, the linear regression models are built in order to examine whether income has the most significant impact on the intakes of nutrition elements compared with other variables.

5.1.2 Regression models

This study aims to determine whether income is the most significant variable affecting the intakes of the nutrition elements compared with other variables. In order to make comparison, the study chooses age, which is also a continuous variable, to compare with income. The method used in the comparison is calculating the elasticities for income and age in each sub model. Elasticity can indicate whether the impact of an increase in x on y is large or not. It measures the percentage of change in y resulting from a one percentage increase in x on average and holding the other regressors fixed. If the absolute value of the elasticity of the variable is larger, it means that the variable is more significant.

The summary result of Model 1 is presented in Table 6. Only for the intake of Fiber column, Income is marked as significant based on its small p-value. However, Gender, CountryOfBirth, College, Age and Race are also marked significant since they all have small p-values. Also, p-values of Gender, CountryOfBirth and Race are smaller than the p-value of Income in the fiber model. In order to evaluate whether income impact is the most significant, even though Income in 4 out of 5 models is marked as insignificant due to its large p-values, the study compares the elasticity of income with the elasticity of age (Table 7). The absolute value of age elasticity is larger than the absolute value of income elasticity among models of protein, fiber, fat, sugar and carbohydrate. It has been shown that the percentage of change in nutrition intakes resulting from a one percentage increase in income on average and holding the other regressors fixed is smaller, meaning that income does not have the most significant impact on intake of all five nutrition elements.

When the education level is removed from Model 1 resulting the Model 2, Income is marked as significant in more sub models. From summary table for Model 2 (Table 8), Income is marked as a significant variable in models of Protein, Fiber, Fat and Sugar. After comparing the elasticity of Income with the elasticity of Age, the fixed effect results are shown in Table 9. Again, the absolute value of age elasticity is larger than the absolute value of income elasticity among models of protein, fiber, fat, sugar and carbohydrate. Thus, Age is more significant than Income among all sub models.

In Model 3, income is divided into income of Hispanic and income of non-Hispanic. The summary table in provided in Table 10 and Income variable is standardized in Model 3. The results show that both Income and Race•Income are still not marked as significant in some sub models such as Sugar model. The other reason for building Model 3 is to explore whether there is a culture difference in consuming nutrition intakes between Hispanic and non-Hispanic groups as the income level changes. In the Fiber model, the coefficient value for Income is 1.73, meaning that there will an increase of 1.73 gram in the conditional expectation of fiber intake of a non-Hispanic for an increase of 1 in income holding all other regressors constant. Meanwhile, the coefficient value for Race•Income is -2.76, indicating that the difference of changes in the conditional expectation of fiber intake as a Hispanic compared with a non-Hispanic for a marginal change in income holding all other regressors constant is -2.76 grams. It has been proved that if the individual is non-Hispanic, as the income level becomes higher, the individual is going to consume more fiber. Otherwise, if the individual is Hispanic, as the income level becomes higher, the individual is going to consume less fiber. Overall, income is not the most significant variable affecting the nutrition intakes. There are many factors related with income affecting the intakes. In order to understand the relationship between income and other affecting factors, this study needs to dive deeper using the ridge regression or mediation analysis to deal with the multicollinearity problem.

5.2 Limitations

First, the study uses 24-hour dietary intake data of the participates. Although the data is a good estimate of long-term daily average intake of the nutrition elements, it does not consider emergent conditions of participants on that day. Maybe some individuals eat more protein on that day due to some special occasions while they do not have lots of protein in their daily meals. Second, the multicollinearity in the linear regression models causes lots of bias of the results. In this study, income is related to most of other variables. Third, participants in the study intend to offset the low income. Some individuals who are in the low-income group may receive food stamps or other help from the government and other institutions and do not need to pay for their daily food. Even the participants are all in the same income group, their living situations are different and thus bias are created for the results. Furthermore, comparing with lowest standard for food intake in order to evaluate the dietary patterns is biased. The highest standard is ignored while it is equally significant. Some individuals have extremely unbalanced nutrition intakes while having too much of particular nutrition elements is as problematic as not meeting the lowest standard.

6. CONCLUSION

According to the descriptive part of this study which shows the dietary intake patterns among income groups, the dietary quality is different in each income group. First, all income groups have fiber intake problems since proportions of individuals who have met the minimum fiber intake standard among all income groups are small. Restaurants located in high-income areas can add more fiber into their menus in order to help their customers intake more fiber and have healthier diets. To help low-income group who receive food from the government, the government may modify the policy and provide more fiber for low-income group. Second, diet cost is also a problem for low-income group while food with higher protein, higher fiber, lower sugar, lower carbohydrate and lower fat is more expensive. Some changes in food price will help more individuals reach nutrition intake standards.

Based on the regression model results, it has been proved that income is not the most significant variable affecting the intakes of nutrition elements. Further study is needed to explore the deeper relationship between income and nutrition intakes while minimize the multicollinearity problem.

7. REFERENCE

Adam Drewnowski; Nicole Darmon, “Symposium: Modifying the Food Environment: Energy Density, Food Costs, and Portion Size”, published by American Society for Nutritional Sciences, 2005

A Aggarwal, P Monsivais, AJ Cook and A Drewnowski, “Does diet cost mediate the relation between socioeconomic position and diet quality?”, published online, 11 May 2011.

Dietary Guidelines 2015-2020, “Dietary Guidelines for Americans 2015–2020 8th Edition.”, health.gov/dietaryguidelines/2015/guidelines/, 2015

“DR1TOT_I.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DR1TOT_I.htm#SEQN.

“DR1TOT_I.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DR1TOT_I.htm#SEQN.

"DEMO_I." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, wwwn.cdc.gov/Nchs/Nhanes/2015-2016/DEMO_I.htm.

Hazel A. B. Hiza; Kellie O. Casavale; Patricia M. Guenther; Carole A. Davis, "Diet Quality of Americans Differs by Age, Sex, Race/ Ethnicity, Income, and Education Level." published by Elsevier Inc., 15 November 2012.

Sharon I. Kirkpatrick; Kevin W. Dodd; Jill Reedy; Susan M. Krebs-Smith, "Income and Race/Ethnicity Are Associated with Adherence to Food-Based Dietary Guidance among US Adults and Children.", published by Elsevier Inc. on behalf of the Academy of Nutrition and Dietetics, 22 November 2011.

U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015 – 2020 Dietary Guidelines for Americans. 8th Edition. December 2015. Available at https://health.gov/dietaryguidelines/2015/guidelines/.

8. APPENDIX

Table 1

Intake variables.

Intake Variables	
Variable	**Definition**
ID	Respondent sequence number
Protein	Total protein intake in 24 hours (g)
Sugar	Total sugar intake in 24 hours (g)
Carbohydrate	Total carbohydrate intake in 24 hours (g)
Fat	Total fat intake in 24 hours (g)
Fiber	Total fiber intake in 24 hours (g)

Table 2

Demographics variables.

Regression Model Variables	
Variable	**Definition**
Age	Age in years of the participant at the time of screening
Gender	Gender (**1**: Female **0**: Male)
Race	Race (**1**: Hispanic **0**: Non-Hispanic)
CountryOfBirth	Country of birth (**1**: United States **0**: Other)
HighSchool	Education level (**1**: At most high school degree **0**: Otherwise)
College	Education level (**1**: BA degree **0**: Otherwise)
Income	Total family income (dollars)

Table 3

USDA and HHS Dietary Standards.

USDA and HHS Dietary Standards				
	Female 25-30	**Male 25-30**	**Female 31-50**	**Male 31-50**
Protein (g)	46	56	46	56
Carbohydrate(g)	130	130	130	130
Fiber (g)	28	33.6	25.2	30.8
Fat (g)	20	20	20	20
Sugar (g)	50	60	45	55

Table 4

Variables Summary Statistics Result.

Summary Statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Protein	86.44	44.95	1.55	56.53	107.80	381.00
Sugar	109.70	73.28	0.33	58.24	142.60	533.40
Carb	259.60	122.60	9.91	176.70	320.00	1,086.00
Fat	85.73	47.82	5.12	52.79	108.70	498.60
Fiber	17.38	11.05	0	9.7	22.5	108
Gender	0.53	0.50	0	0	1	1
CountryOfBirth	0.65	0.48	0	0	1	1
Age	37.20	7.52	25	31	44	50
Income	58,892.00	39,499.00	2,500	30,000	87,500	125,000
Race	0.30	0.46	0	0	1	1
HighSchool	0.20	0.40	0	0	0	1
College	0.61	0.49	0	0	1	1

Table 5

Percentage of Reaching the Minimum Nutrition Intake Standard.

Percentage (%) of Reaching the Minimum Nutrition Intake Standard					
	Protein	**Carbohydrate**	**Fiber**	**Fat**	**Sugar**
Low Income	78.65	88.06	12.92	95.51	80.20
Middle Income	81.53	87.90	12.95	97.24	79.09
High Income	87.18	92.31	15.13	98.72	82.56

Table 6

Summary table for Model 1.

Seven Fators Affecting Nutrition Intake

	Dependent variable:				
	Protein (1)	Carb (2)	Fiber (3)	Fat (4)	Sugar (5)
Gender	-26.11***	-65.60***	-3.10***	-23.17***	-23.11***
	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***
	t = -13.59	t = -12.45	t = -6.56	t = -11.20	t = -7.19
CountryOfBirth	-1.63	1.11	-4.32***	11.49***	18.59***
	p = 0.46	p = 0.86	p = 0.00***	p = 0.0000***	p = 0.0000***
	t = -0.75	t = 0.19	t = -8.09	t = 4.92	t = 5.12
HighSchool	-1.18	-8.72	-0.97	-1.83	-6.06
	p = 0.71	p = 0.32	p = 0.22	p = 0.59	p = 0.26
	t = -0.37	t = -1.01	t = -1.25	t = -0.54	t = -1.15
College	2.22	-11.52	1.15*	3.19	-13.86***
	p = 0.44	p = 0.14	p = 0.10*	p = 0.30	p = 0.004***
	t = 0.79	t = -1.49	t = 1.66	t = 1.05	t = -2.93
Age	-0.38***	-1.08***	-0.07**	-0.28**	-0.23
	p = 0.004***	p = 0.003***	p = 0.04**	p = 0.05**	p = 0.29
	t = -2.95	t = -3.05	t = -2.07	t = -1.99	t = -1.06
Income	0.0000	0.0000	0.0000***	0.0000	-0.0000
	p = 0.17	p = 0.71	p = 0.001***	p = 0.27	p = 0.56
	t = 1.37	t = 0.37	t = 3.38	t = 1.11	t = -0.59
Race	6.34***	16.22**	2.74***	8.26***	5.82
	p = 0.01***	p = 0.02**	p = 0.0000***	p = 0.002***	p = 0.14
	t = 2.70	t = 2.52	t = 4.74	t = 3.27	t = 1.48
Constant	110.40***	336.30***	21.68***	94.93***	127.80***
	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***
	t = 19.15	t = 21.27	t = 15.27	t = 15.30	t = 13.25
Observations	2,044	2,044	2,044	2,044	2,044
Log Likelihood	-10,581.00	-12,643.00	-7,717.00	-10,731.00	-11,632.00
Akaike Inf. Crit.	21,177.00	25,301.00	15,449.00	21,478.00	23,281.00
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 7

Fixed Effects Results for Model 1.

Fixed Effects Results for Model 1		
	Income Elasticity	Age Elasticity
Protein	0.02	- 0.16
Carb	0.01	- 0.15
Fiber	0.07	- 0.14
Fat	0.02	- 0.12
Sugar	- 0.01	- 0.08

Table 8

Summary table for Model 2.

Without Educ

	Dependent variable:				
	Protein (1)	Carb (2)	Fiber (3)	Fat (4)	Sugar (5)
Gender	-25.79***	-66.38***	-2.91***	-22.70***	-24.32***
	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***
	t = -13.54	t = -12.70	t = -6.19	t = -11.06	t = -7.61
CountryOfBirth	-1.67	0.22	-4.37***	11.43***	17.85***
	p = 0.44	p = 0.98	p = 0.00***	p = 0.0000***	p = 0.0000***
	t = -0.78	t = 0.04	t = -8.22	t = 4.93	t = 4.95
Age	-0.39***	-1.01***	-0.07**	-0.30**	-0.15
	p = 0.002***	p = 0.004***	p = 0.02**	p = 0.04**	p = 0.50
	t = -3.10	t = -2.90	t = -2.34	t = -2.17	t = -0.69
Income	0.0000*	-0.0000	0.0000***	0.0000*	-0.0001*
	p = 0.06*	p = 0.94	p = 0.0000***	p = 0.08*	p = 0.09*
	t = 1.89	t = -0.08	t = 4.57	t = 1.76	t = -1.73
Race	5.79**	18.55***	2.44***	7.47***	8.78**
	p = 0.02**	p = 0.003***	p = 0.0001***	p = 0.003***	p = 0.03**
	t = 2.55	t = 2.97	t = 4.35	t = 3.05	t = 2.30
Constant	111.50***	327.40***	22.17***	96.47***	118.00***
	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***	p = 0.00***
	t = 20.98	t = 22.46	t = 16.90	t = 16.86	t = 13.25
Observations	2,044	2,044	2,044	2,044	2,044
Log Likelihood	-10,582.00	-12,644.00	-7,723.00	-10,733.00	-11,637.00
Akaike Inf. Crit.	21,175.00	25,300.00	15,457.00	21,478.00	23,286.00
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 9

Fixed Effects Results for Model 2.

Fixed Effects Results for Model 2		
	Income Elasticity	Age Elasticity
Protein	0.03	- 0.17
Carb	- 0.00	- 0.15
Fiber	0.09	- 0.16
Fat	0.03	- 0.13
Sugar	- 0.04	- 0.05

Table 10

Summary table for Model 3.

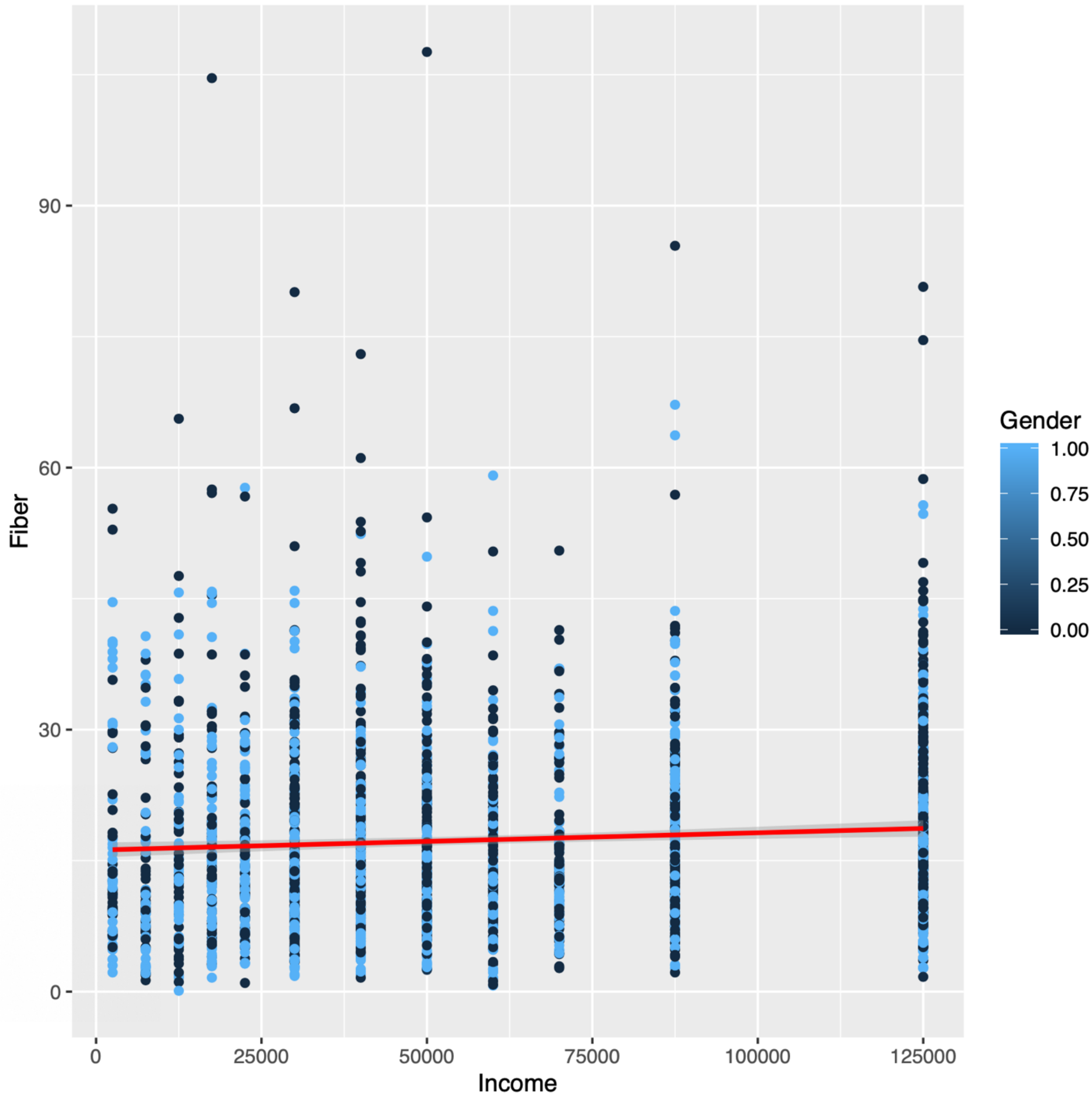
Without Educ Race x Income

	Dependent variable:				
	Protein (1)	Carb (2)	Fiber (3)	Fat (4)	Sugar (5)
Gender	-25.80*** p = 0.00*** t = -13.54	-66.43*** p = 0.00*** t = -12.73	-2.92*** p = 0.00*** t = -6.24	-22.71*** p = 0.00*** t = -11.07	-24.33*** p = 0.00*** t = -7.62
CountryOfBirth	-1.14 p = 0.61 t = -0.52	3.03 p = 0.62 t = 0.51	-3.91*** p = 0.00*** t = -7.28	11.88*** p = 0.0000*** t = 5.05	18.56*** p = 0.0000*** t = 5.06
Age	-0.40*** p = 0.002*** t = -3.14	-1.04*** p = 0.003*** t = -2.99	-0.08** p = 0.02** t = -2.52	-0.30** p = 0.03** t = -2.20	-0.16 p = 0.47 t = -0.73
Income	2.57** p = 0.03** t = 2.31	3.65 p = 0.24 t = 1.20	1.73*** p = 0.00*** t = 6.34	2.47** p = 0.04** t = 2.06	-1.84 p = 0.33 t = -0.99
Race	5.43** p = 0.02** t = 2.37	16.64*** p = 0.01*** t = 2.65	2.13*** p = 0.0002*** t = 3.79	7.16*** p = 0.004*** t = 2.91	8.29** p = 0.04** t = 2.16
Income:Race	-3.19 p = 0.18 t = -1.36	-16.91*** p = 0.01*** t = -2.63	-2.76*** p = 0.0000*** t = -4.78	-2.72 p = 0.29 t = -1.08	-4.30 p = 0.28 t = -1.09
Constant	114.00*** p = 0.00*** t = 21.71	325.70*** p = 0.00*** t = 22.65	23.58*** p = 0.00*** t = 18.29	99.00*** p = 0.00*** t = 17.51	113.50*** p = 0.00*** t = 12.90
Observations	2,044	2,044	2,044	2,044	2,044
Log Likelihood	-10,581.00	-12,640.00	-7,711.00	-10,732.00	-11,637.00
Akaike Inf. Crit.	21,175.00	25,295.00	15,436.00	21,479.00	23,287.00

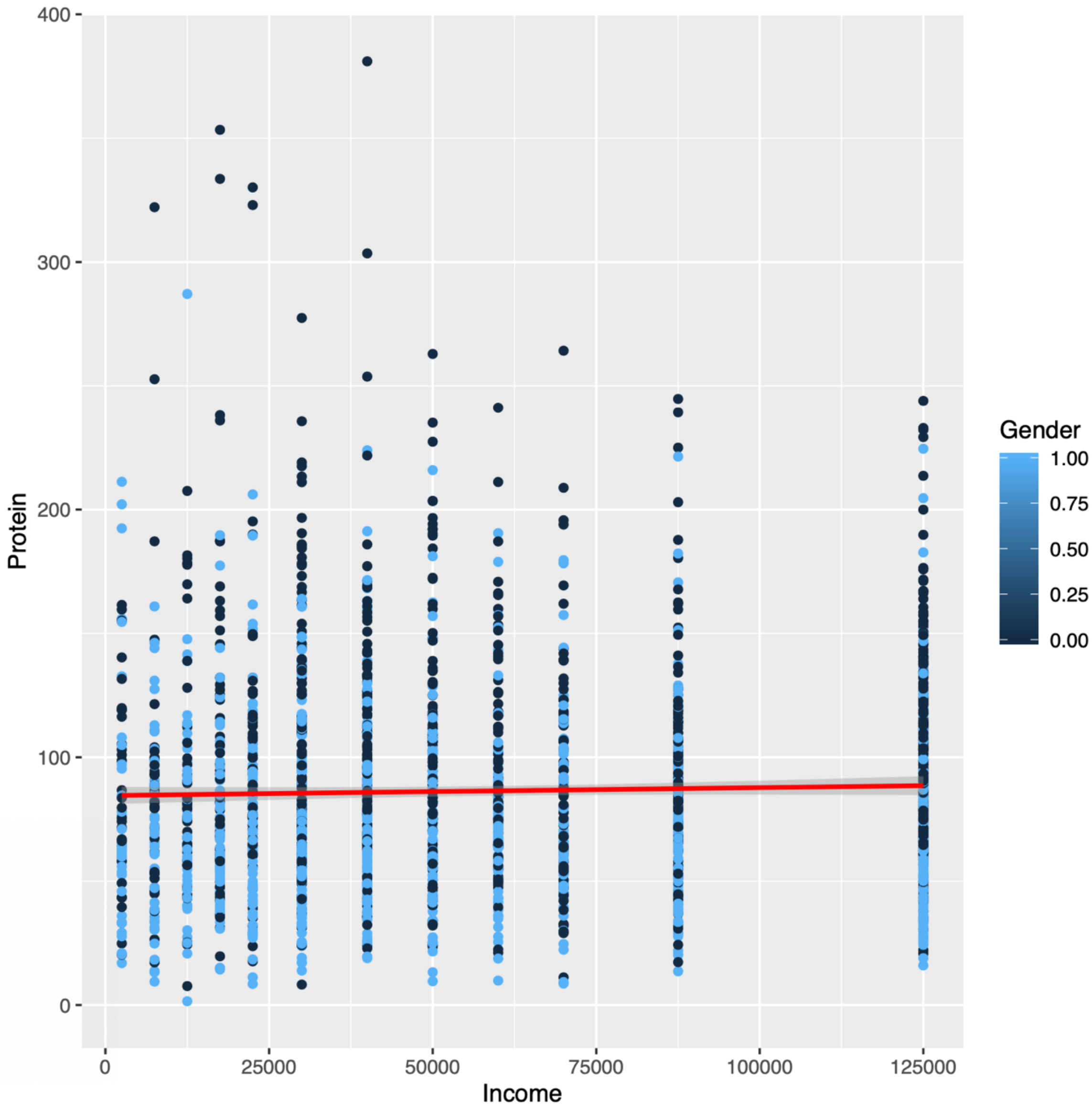
Note: *p<0.1; **p<0.05; ***p<0.01

Graph 1 Income and nutrition intake.

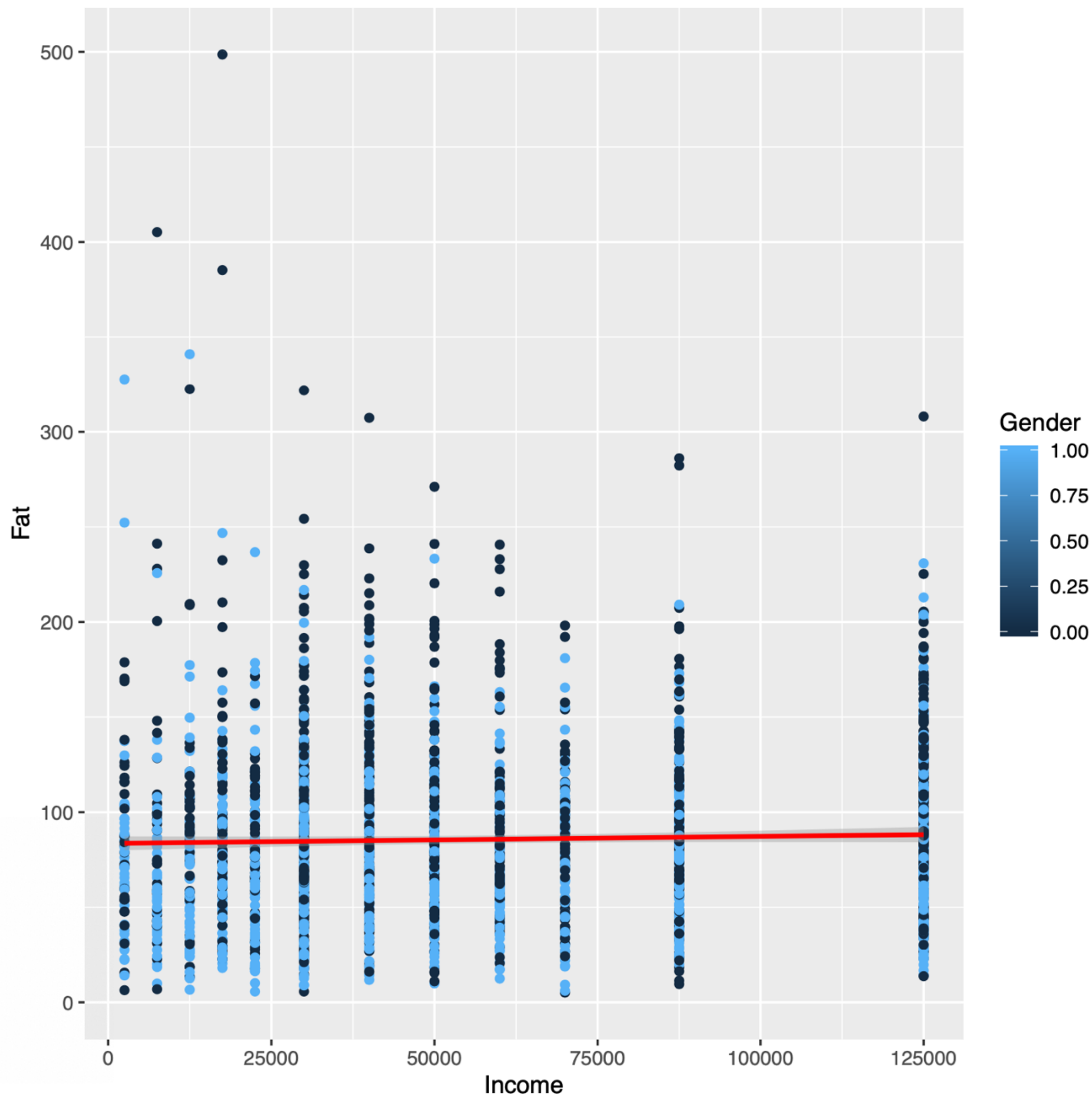
1a Income & Fiber



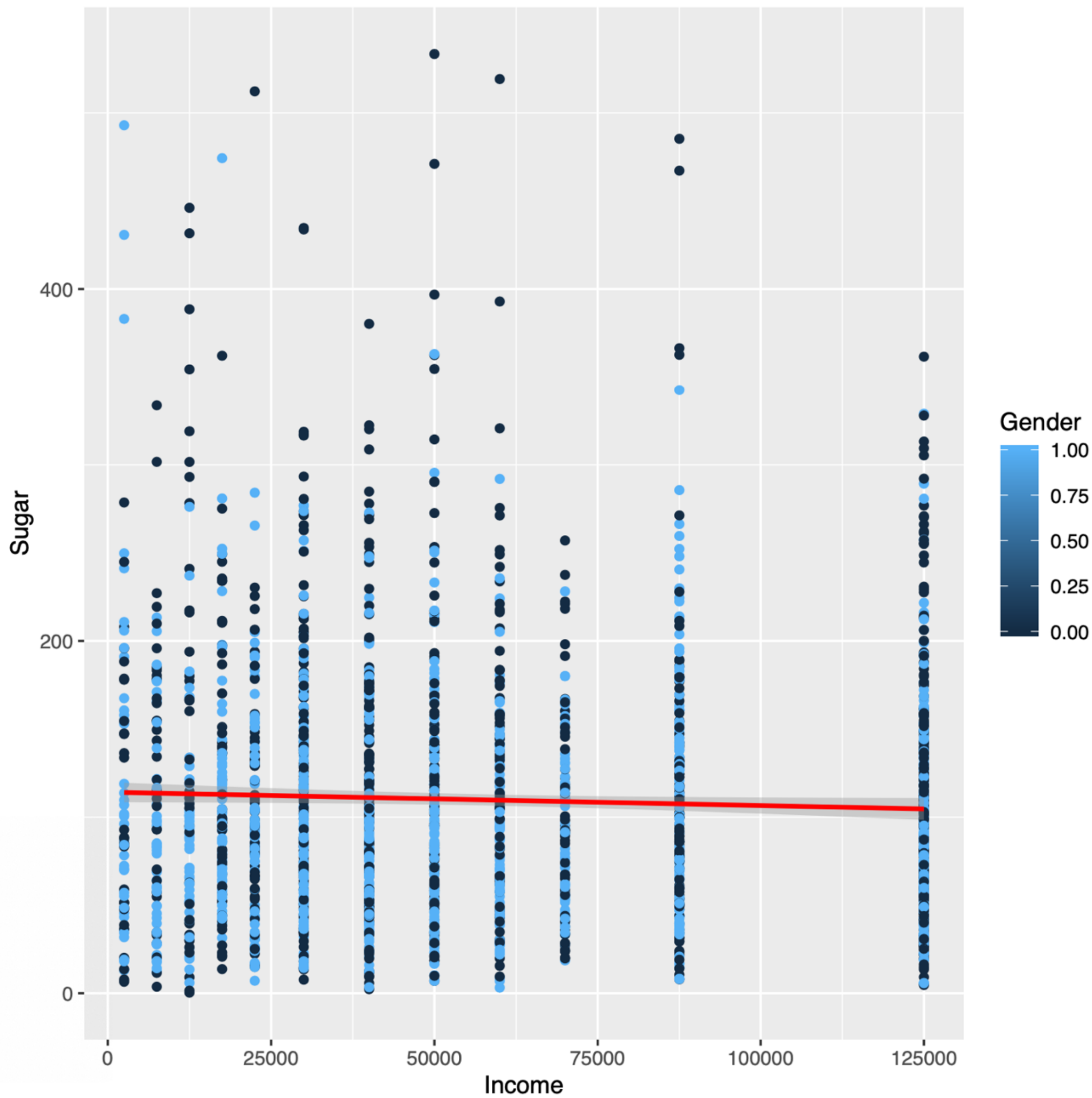
1b Income & Protein



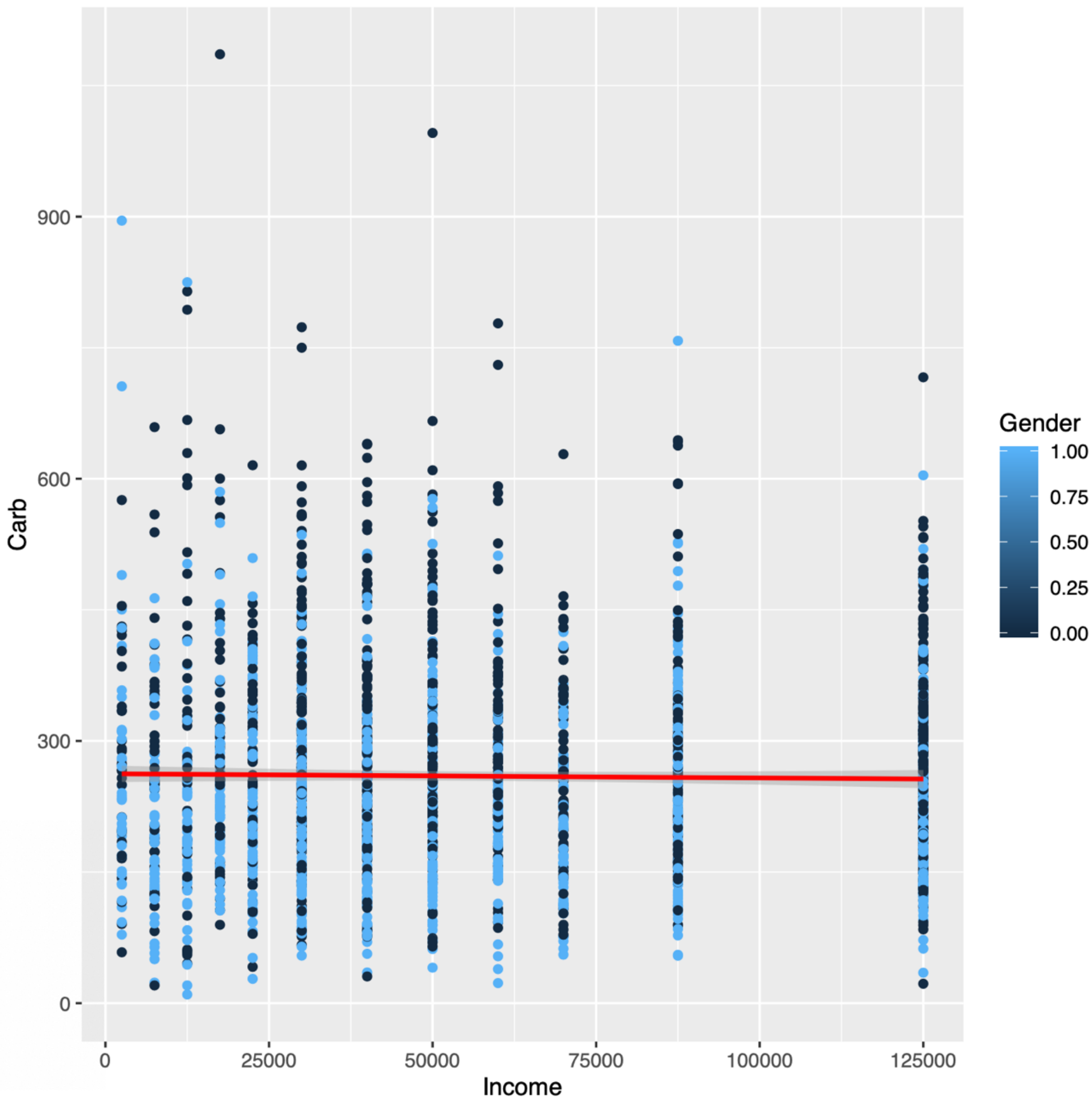
1c Income & Fat



1d Income & Sugar

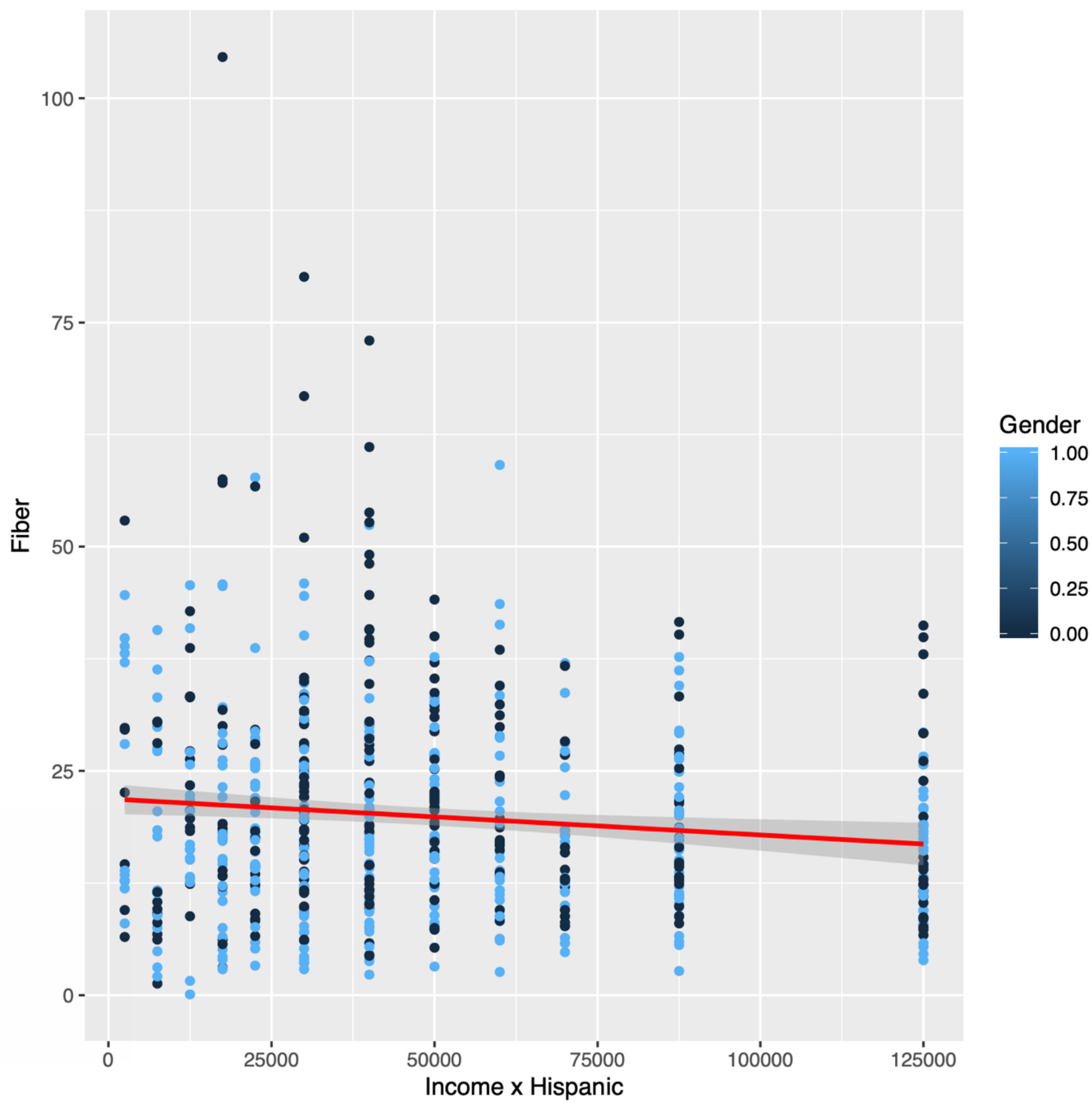


1e Income & Carbohydrate

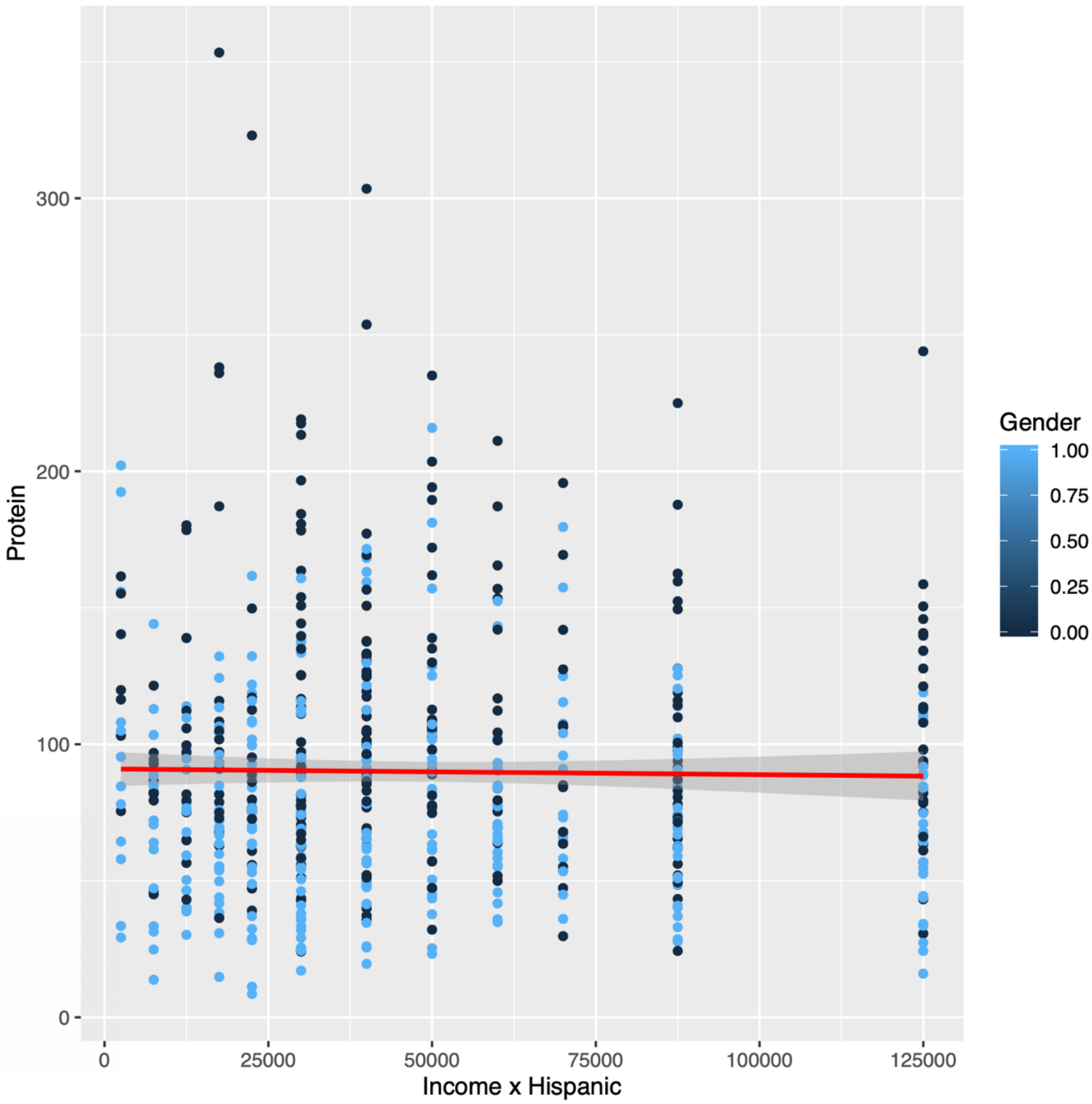


Graph 2 Income and nutrition intake (Hispanic).

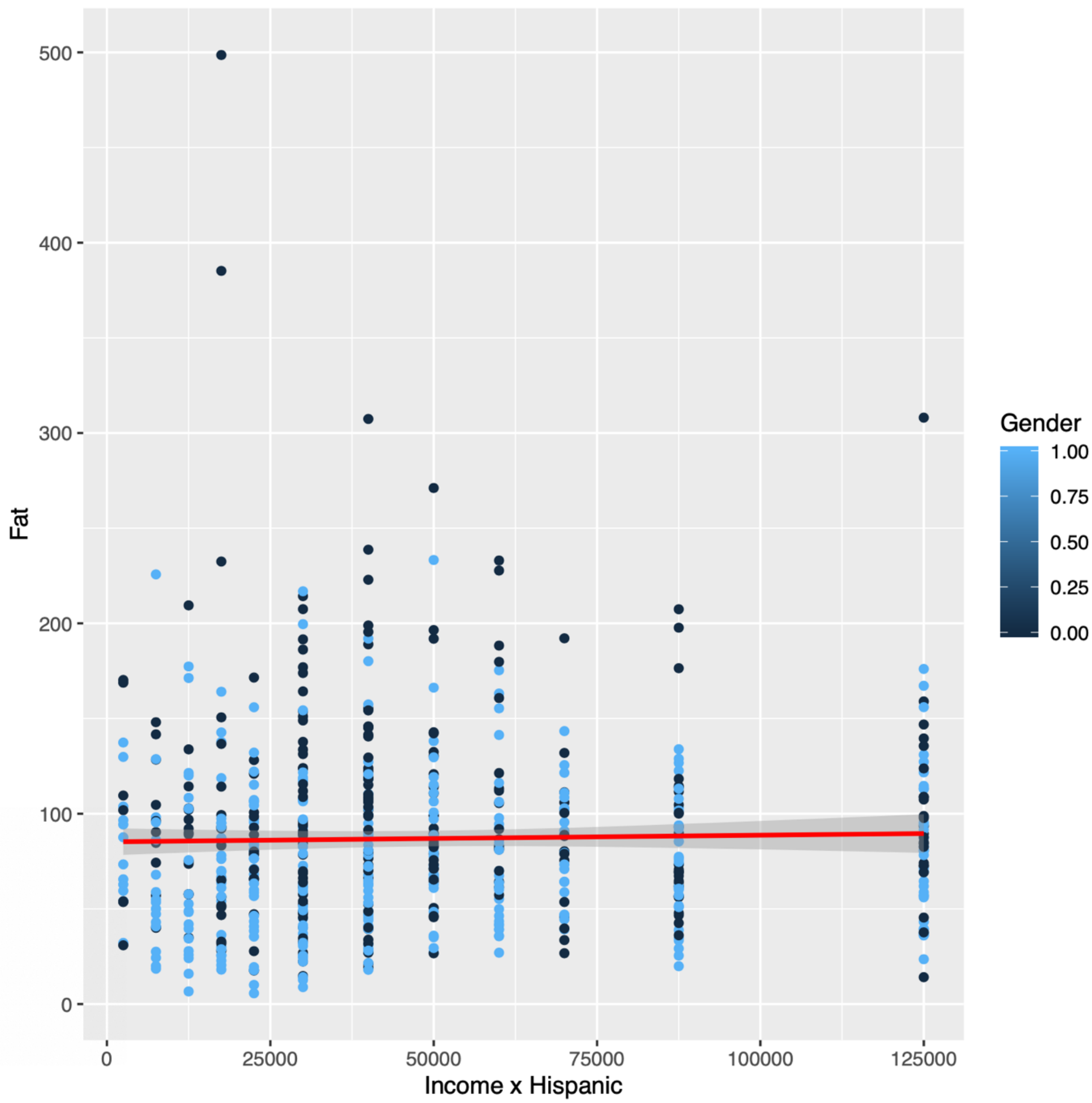
2a Income & Fiber



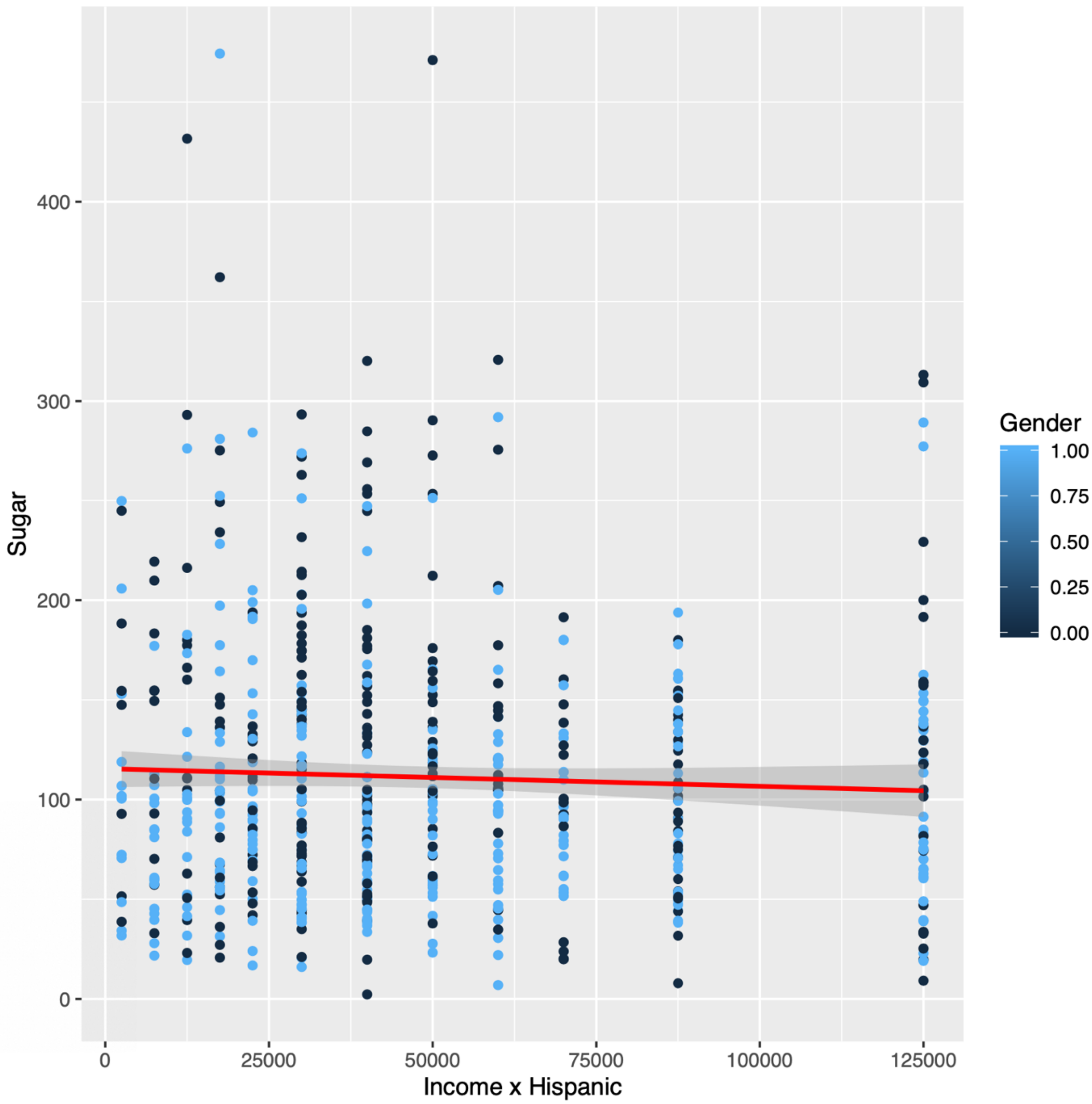
2b Income & Protein



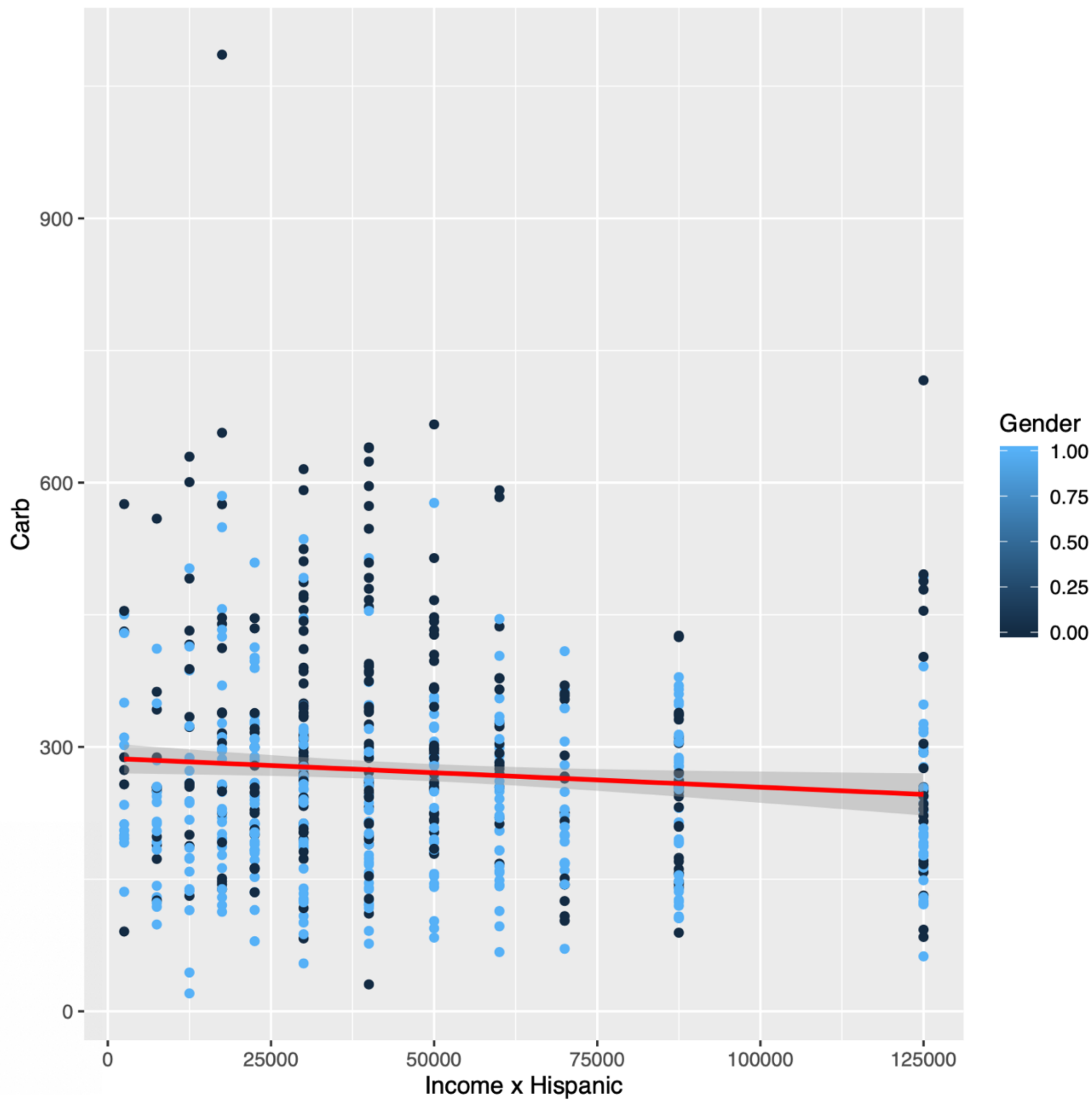
2c Income & Fat



2d Income & Sugar

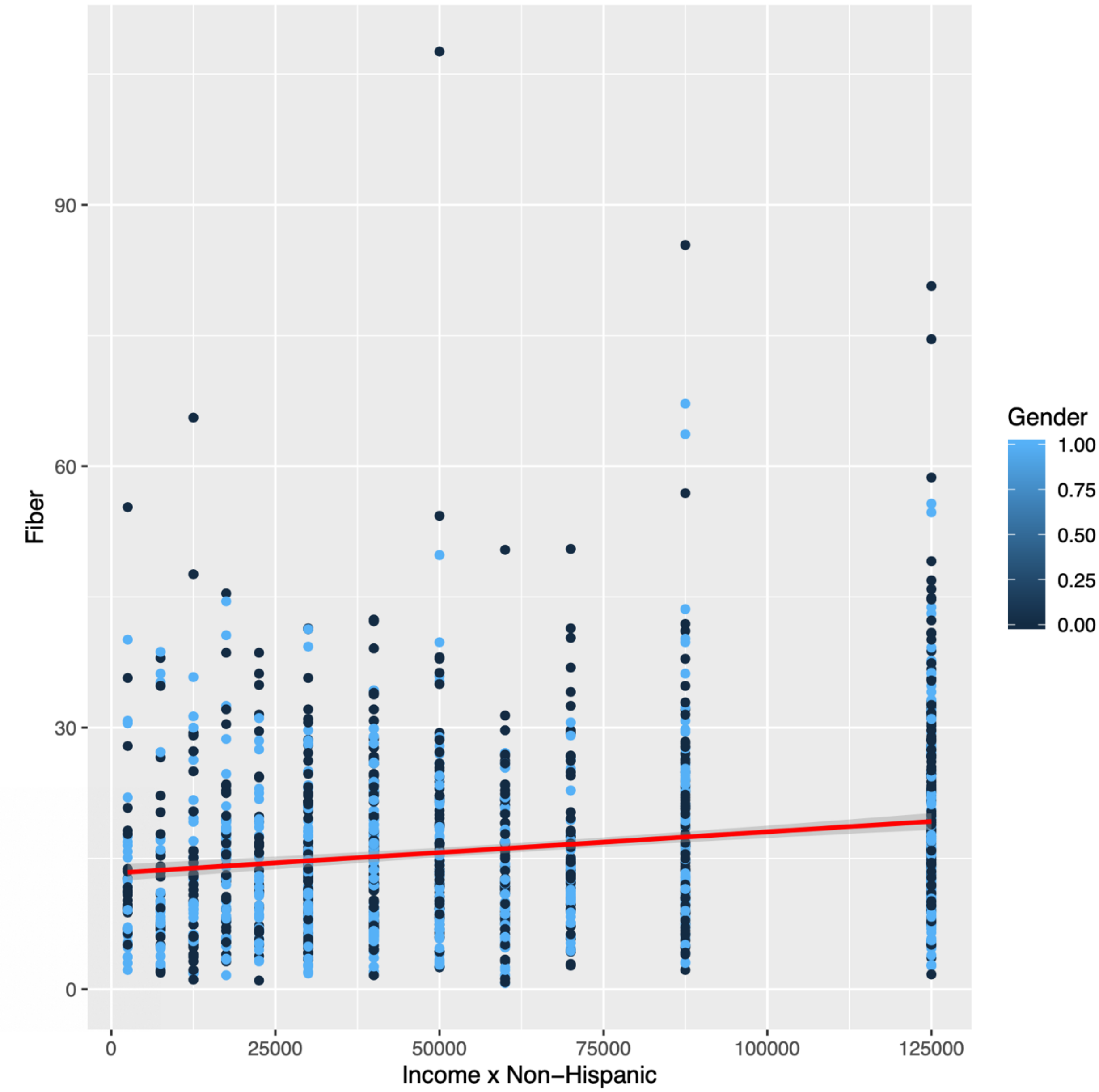


2e Income & Carbohydrate

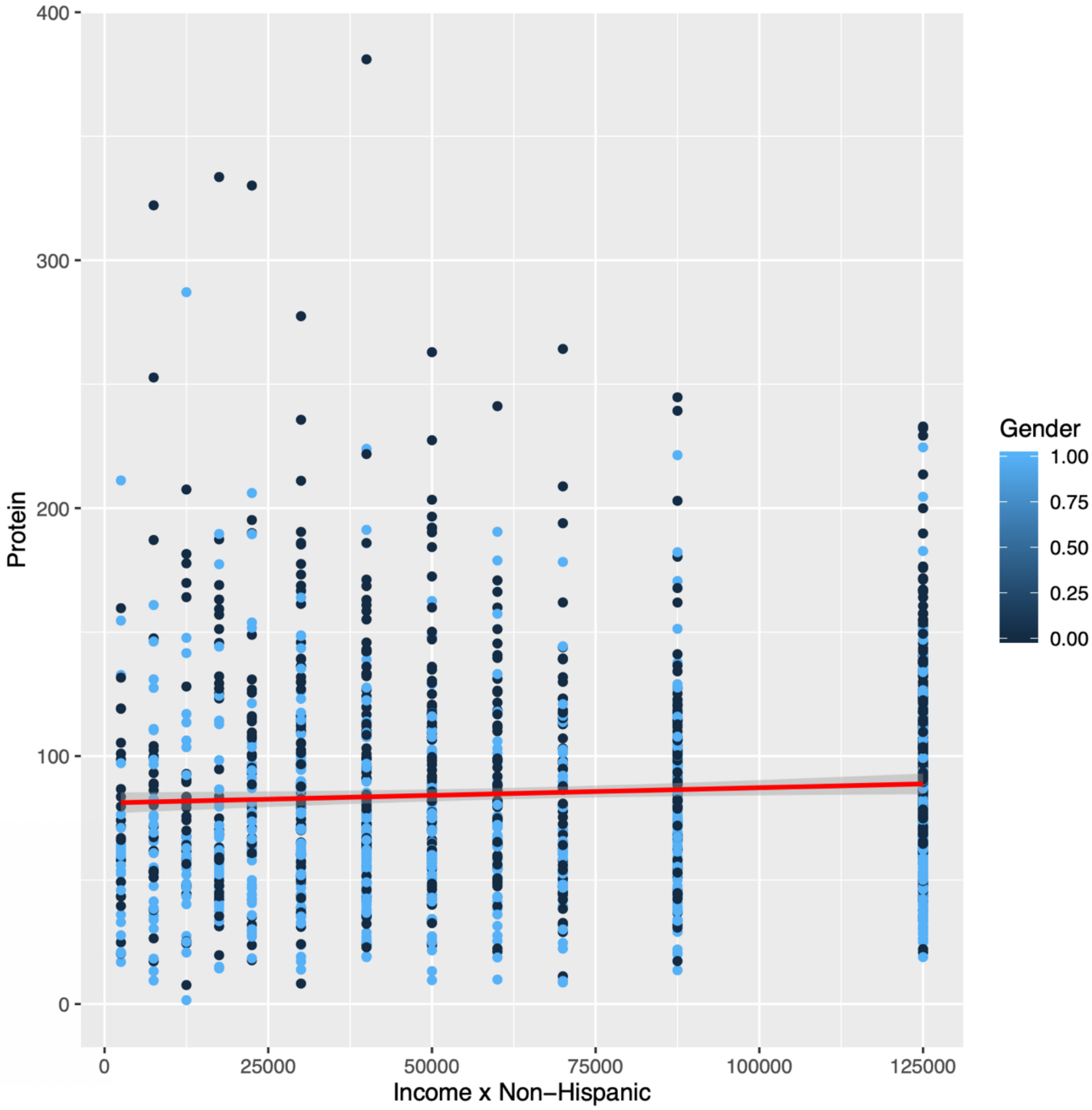


Graph 3 Income and nutrition intake (Non-Hispanic).

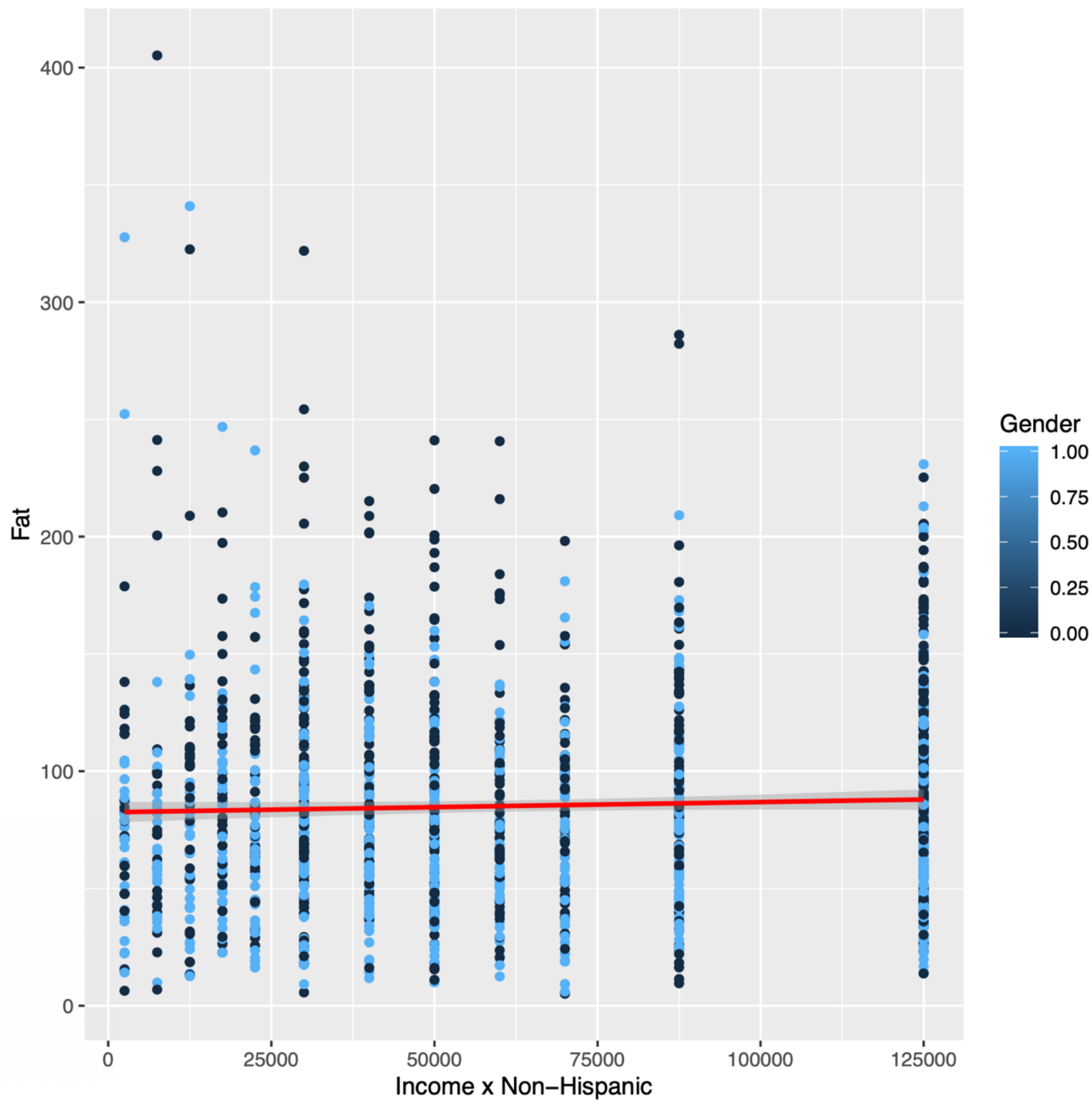
3a Income & Fiber



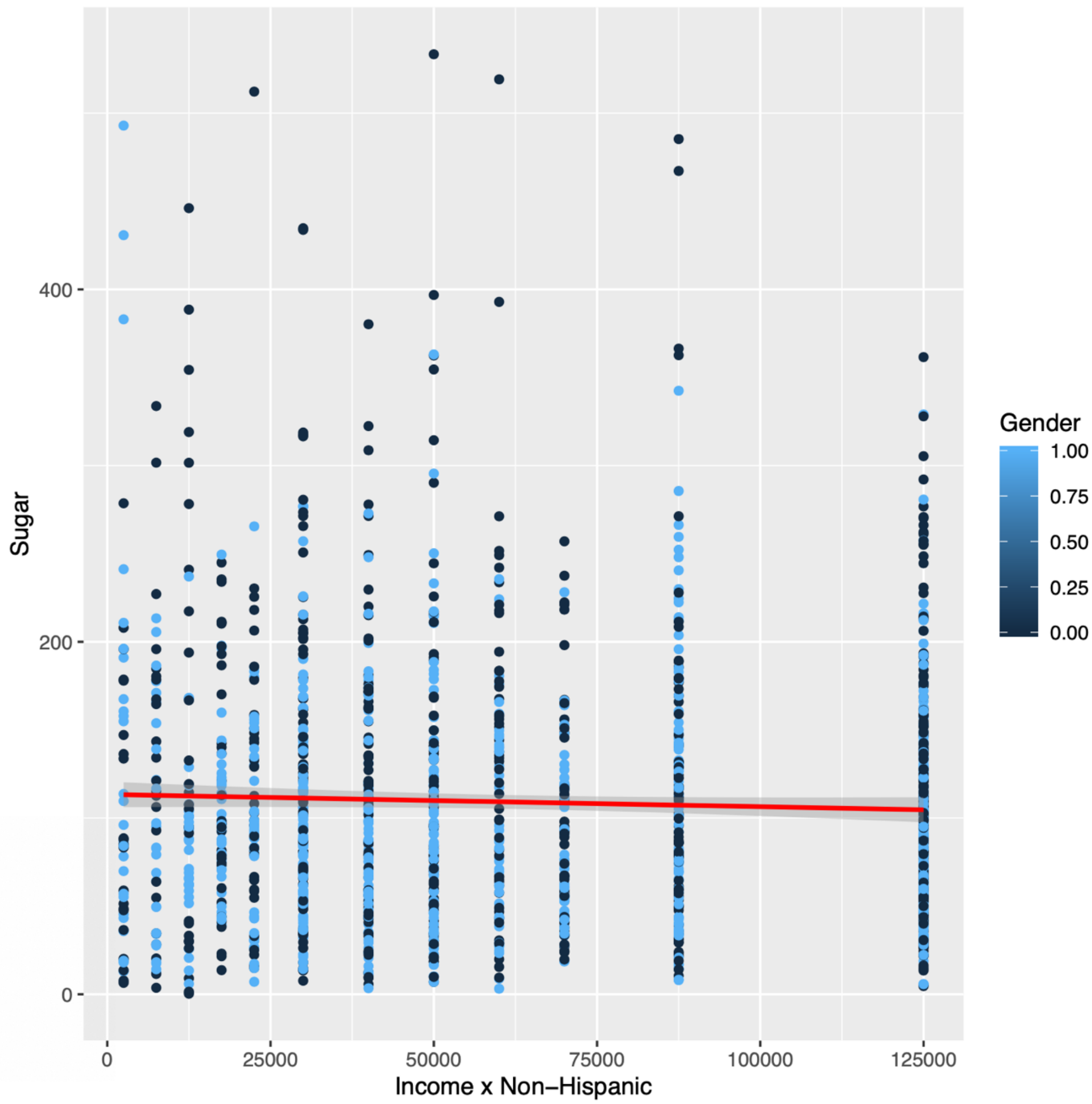
3b Income & Protein



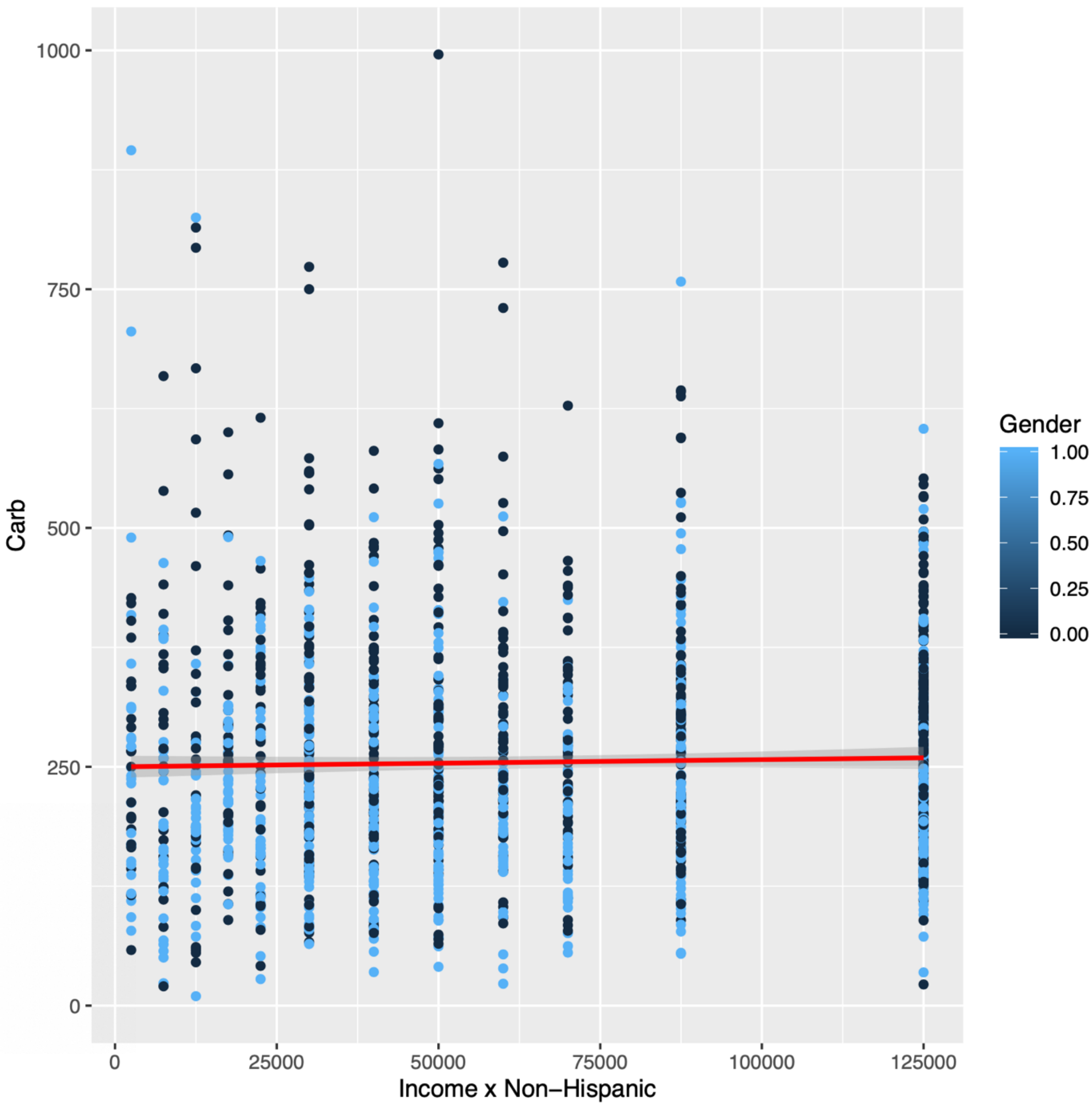
3c Income & Fat



3d Income & Sugar



3e Income & Carbohydrate



R script

```r
getwd()
library(foreign)
library(dplyr)
library(stargazer)
library(ggplot2)
library(data.table)
library(glmnet)


# Nutrition intake
diet_ind_1st <- read.xport('./Dietary/DR1IFF_I.xpt')
intake <- dplyr::select(diet_ind_1st, SEQN, DR1IPROT, DR1ISUGR, DR1ICARB, DR1ITFAT, DR1IFIBE) %>%
                 filter(!is.na(DR1IPROT))
intake <- dplyr::group_by(intake,SEQN) %>%
                 summarise_all(sum)

# Read in Demographics data
demographics_ori <- read.xport('./Demographics/DEMO_I.XPT')

# Change variable labels
demographics_ori$RIAGENDR[demographics_ori$RIAGENDR == 1] <- 0
demographics_ori$RIAGENDR[demographics_ori$RIAGENDR == 2] <- 1
demographics_ori$DMDBORN4[demographics_ori$DMDBORN4 == 2] <- 0
demographics_ori$RIDRETH3[demographics_ori$RIDRETH3 == 6] <- 5
demographics_ori$RIDRETH3[demographics_ori$RIDRETH3 == 7] <- 6

# Select the project related variables
demographics <- dplyr::select(demographics_ori, SEQN, RIAGENDR, DMDBORN4, DMDEDUC2, RIDAGEYR, INDFMIN2, RIDRETH3) %>%
                filter(!is.na(INDFMIN2)) %>%
                filter(RIDAGEYR >= 25 & RIDAGEYR <= 50)%>%
                filter(INDFMIN2 >= 1 & INDFMIN2 <= 15)%>%
                filter(INDFMIN2 != 12 & INDFMIN2 != 13)%>%
                filter(DMDEDUC2 != 7 & DMDEDUC2 != 9)

# Change variable labels
demographics$INDFMIN2[demographics$INDFMIN2 == 14] <- 11
demographics$INDFMIN2[demographics$INDFMIN2 == 15] <- 12
demographics$RIDRETH3[demographics$RIDRETH3 == 2] <- 1
demographics$RIDRETH3[demographics$RIDRETH3 == 3] <- 0
demographics$RIDRETH3[demographics$RIDRETH3 == 4] <- 0
demographics$RIDRETH3[demographics$RIDRETH3 == 5] <- 0
demographics$RIDRETH3[demographics$RIDRETH3 == 6] <- 0


# Join the intake and demographics tables
dataset <- dplyr::right_join(intake, demographics, by = 'SEQN') %>%
  filter(!is.na(DR1IPROT))
```

```r
# Rename the variables
dataset <- dataset %>%
    rename(
    ID = SEQN,
    Protein = DR1IPROT,
    Sugar = DR1ISUGR,
    Carb = DR1ICARB,
    Fat = DR1ITFAT,
    Fiber = DR1IFIBE,
    Gender = RIAGENDR,
    CountryOfBirth = DMDBORN4,
    Educ = DMDEDUC2,
    Age = RIDAGEYR,
    Income = INDFMIN2,
    Race = RIDRETH3)

dataset$HighSchool[dataset$Educ == 3] <- 1
dataset$HighSchool[dataset$Educ != 3] <- 0
dataset$College[dataset$Educ >= 4] <- 1
dataset$College[dataset$Educ < 4] <- 0

# Delete 'educ' column
dataset = subset(dataset, select = -c(Educ) )

# Change Income into a continus variable
dataset$Income[dataset$Income == 1] <- 2500
dataset$Income[dataset$Income == 2] <- 7500
dataset$Income[dataset$Income == 3] <- 12500
dataset$Income[dataset$Income == 4] <- 17500
dataset$Income[dataset$Income == 5] <- 22500
dataset$Income[dataset$Income == 6] <- 30000
dataset$Income[dataset$Income == 7] <- 40000
dataset$Income[dataset$Income == 8] <- 50000
dataset$Income[dataset$Income == 9] <- 60000
dataset$Income[dataset$Income == 10] <- 70000
dataset$Income[dataset$Income == 11] <- 87500
dataset$Income[dataset$Income == 12] <- 125000

# Export dataset
dataset <- as.data.frame(dataset)
write.csv(dataset, 'dataset.csv')

# Summary Statistics
stargazer(dataset, type = 'text', title = 'Summary Statistics',
          omit = 'ID', omit.summary.stat = "n", digits = 2,  out = 'summary_stats.txt')

# Define Income Classes
low_income <- subset(dataset, Income <= 30000)
middle_income <- subset(dataset, Income >= 40000 & Income <= 87500)
high_income <- subset(dataset, Income == 125000)
```

```r
# Divide each class into groups with different ages and genders
# Female 25-30
low_f_25 <- subset(low_income, Age <= 30)%>%
            filter(Gender == 1)
middle_f_25 <- subset(middle_income, Age <= 30)%>%
               filter(Gender == 1)
high_f_25 <- subset(high_income, Age <= 30)%>%
             filter(Gender == 1)
# Male 25-30
low_m_25 <- subset(low_income, Age <= 30)%>%
            filter(Gender == 0)
middle_m_25 <- subset(middle_income, Age <= 30)%>%
               filter(Gender == 0)
high_m_25 <- subset(high_income, Age <= 30)%>%
             filter(Gender == 0)
# Female 31-50
low_f_50 <- subset(low_income, Age >= 31)%>%
            filter(Gender == 1)
middle_f_50 <- subset(middle_income, Age >= 31)%>%
               filter(Gender == 1)
high_f_50 <- subset(high_income, Age >= 31)%>%
             filter(Gender == 1)
# Male 31-50
low_m_50 <- subset(low_income, Age >= 31)%>%
            filter(Gender == 0)
middle_m_50 <- subset(middle_income, Age >= 31)%>%
               filter(Gender == 0)
high_m_50 <- subset(high_income, Age >= 31)%>%
             filter(Gender == 0)
```

```r
# Calculate compliance rate of 5 nutrition factors
# Protein
low_protein_per <- ((length(which(low_f_25$Protein >= 46))) +
                    (length(which(low_m_25$Protein >= 56))) +
                    (length(which(low_f_50$Protein >= 46))) +
                    (length(which(low_m_50$Protein >= 56)))) /
                    length(low_income$Protein) * 100
middle_protein_per <- ((length(which(middle_f_25$Protein >= 46))) +
                       (length(which(middle_m_25$Protein >= 56))) +
                       (length(which(middle_f_50$Protein >= 46))) +
                       (length(which(middle_m_50$Protein >= 56)))) /
                       length(middle_income$Protein) * 100
high_protein_per <- ((length(which(high_f_25$Protein >= 46))) +
                     (length(which(high_m_25$Protein >= 56))) +
                     (length(which(high_f_50$Protein >= 46))) +
                     (length(which(high_m_50$Protein >= 56)))) /
                     length(high_income$Protein) * 100
# Carb
low_carb_per <- ((length(which(low_f_25$Carb >= 130))) +
                 (length(which(low_m_25$Carb >= 130))) +
                 (length(which(low_f_50$Carb >= 130))) +
                 (length(which(low_m_50$Carb >= 130)))) /
                 length(low_income$Carb) * 100
middle_carb_per <- ((length(which(middle_f_25$Carb >= 130))) +
                    (length(which(middle_m_25$Carb >= 130))) +
                    (length(which(middle_f_50$Carb >= 130))) +
                    (length(which(middle_m_50$Carb >= 130)))) /
                    length(middle_income$Carb) * 100
high_carb_per <- ((length(which(high_f_25$Carb >= 130))) +
                  (length(which(high_m_25$Carb >= 130))) +
                  (length(which(high_f_50$Carb >= 130))) +
                  (length(which(high_m_50$Carb >= 130)))) /
                  length(high_income$Carb) * 100
# Fiber
low_fiber_per <- ((length(which(low_f_25$Fiber >= 28))) +
                  (length(which(low_m_25$Fiber >= 33.6))) +
                  (length(which(low_f_50$Fiber >= 25.2))) +
                  (length(which(low_m_50$Fiber >= 30.8)))) /
                  length(low_income$Fiber) * 100
middle_fiber_per <- ((length(which(middle_f_25$Fiber >= 28))) +
                     (length(which(middle_m_25$Fiber >= 33.6))) +
                     (length(which(middle_f_50$Fiber >= 25.2))) +
                     (length(which(middle_m_50$Fiber >= 30.8)))) /
                     length(middle_income$Fiber) * 100
high_fiber_per <- ((length(which(high_f_25$Fiber >= 28))) +
                   (length(which(high_m_25$Fiber >= 33.6))) +
                   (length(which(high_f_50$Fiber >= 25.2))) +
                   (length(which(high_m_50$Fiber >= 30.8)))) /
                   length(high_income$Fiber) * 100
```

```r
# Fat
low_fat_per <- ((length(which(low_f_25$Fat >= 20))) +
                (length(which(low_m_25$Fat >= 20))) +
                (length(which(low_f_50$Fat >= 20))) +
                (length(which(low_m_50$Fat >= 20)))) /
                length(low_income$Fat) * 100
middle_fat_per <- ((length(which(middle_f_25$Fat >= 20))) +
                   (length(which(middle_m_25$Fat >= 20))) +
                   (length(which(middle_f_50$Fat >= 20))) +
                   (length(which(middle_m_50$Fat >= 20)))) /
                   length(middle_income$Fat) * 100
high_fat_per <- ((length(which(high_f_25$Fat >= 20))) +
                 (length(which(high_m_25$Fat >= 20))) +
                 (length(which(high_f_50$Fat >= 20))) +
                 (length(which(high_m_50$Fat >= 20)))) /
                 length(high_income$Fat) * 100
# Sugar
low_sugar_per <- ((length(which(low_f_25$Sugar >= 50))) +
                  (length(which(low_m_25$Sugar >= 60))) +
                  (length(which(low_f_50$Sugar >= 45))) +
                  (length(which(low_m_50$Sugar >= 55)))) /
                  length(low_income$Sugar) * 100
middle_sugar_per <- ((length(which(middle_f_25$Sugar >= 50))) +
                     (length(which(middle_m_25$Sugar >= 60))) +
                     (length(which(middle_f_50$Sugar >= 45))) +
                     (length(which(middle_m_50$Sugar >= 55)))) /
                     length(middle_income$Sugar) * 100
high_sugar_per <- ((length(which(high_f_25$Sugar >= 50))) +
                   (length(which(high_m_25$Sugar >= 60))) +
                   (length(which(high_f_50$Sugar >= 45))) +
                   (length(which(high_m_50$Sugar >= 55)))) /
                   length(high_income$Sugar) * 100

# Summary table of compliance rate
options(digits = 4)
comp_table <- data.table(IncomeClass = c('Low Income', 'Middle Income', 'High Income'),
                         Protein = c(low_protein_per, middle_protein_per, high_protein_per),
                         Carb = c(low_carb_per, middle_carb_per, high_carb_per),
                         Fiber = c(low_fiber_per, middle_fiber_per, high_fiber_per),
                         Fat = c(low_fat_per, middle_fat_per, high_fat_per),
                         Sugar = c(low_sugar_per, middle_sugar_per, high_sugar_per))
```

```
# Linear Regression Models
# (1) All 7 factors
# Protein
lm_protein <- glm(Protein ~ Gender + CountryOfBirth + HighSchool + College + Age + Income + Race, data = dataset)
# Carb
lm_carb <- glm(Carb ~ Gender + CountryOfBirth + HighSchool + College + Age + Income + Race, data = dataset)
# Fiber
lm_fiber <- glm(Fiber ~ Gender + CountryOfBirth + HighSchool + College + Age + Income + Race, data = dataset)
# Fat
lm_fat <- glm(Fat ~ Gender + CountryOfBirth + HighSchool + College + Age + Income + Race, data = dataset)
# Sugar
lm_sugar <- glm(Sugar ~ Gender + CountryOfBirth + HighSchool + College + Age + Income + Race, data = dataset)
# Summary
lm_summary <- stargazer(lm_protein, lm_carb, lm_fiber, lm_fat, lm_sugar, type = 'text',
                        title = 'Seven Fators Affecting Nutrition Intake',
                        digits = 2, report=('v*c*p*t'), out = 'all_seven_lm.txt')

# Factor Elasticity (All)
# (A 1% increase in x causes a [elas%] cahnge in y)
protein_elas_income <- as.numeric(lm_protein$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Protein))
protein_elas_age <- as.numeric(lm_protein$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Protein))
fiber_elas_income <- as.numeric(lm_fiber$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Fiber))
fiber_elas_age <- as.numeric(lm_fiber$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Fiber))
carb_elas_income <- as.numeric(lm_carb$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Carb))
carb_elas_age <- as.numeric(lm_carb$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Carb))
fat_elas_income <- as.numeric(lm_fat$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Fat))
fat_elas_age <- as.numeric(lm_fat$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Fat))
sugar_elas_income <- as.numeric(lm_sugar$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Sugar))
sugar_elas_age <- as.numeric(lm_sugar$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Sugar))


# (2) No Educ
# Examine the quadratic relationship between education and nutrition intake
# Protein
lm_protein_no_educ <- glm(Protein ~ Gender + CountryOfBirth + Age + Income + Race, data = dataset)
# Carb
lm_carb_no_educ <- glm(Carb ~ Gender + CountryOfBirth + Age + Income + Race, data = dataset)
# Fiber
lm_fiber_no_educ <- glm(Fiber ~ Gender + CountryOfBirth + Age + Income + Race, data = dataset)
# Fat
lm_fat_no_educ <- glm(Fat ~ Gender + CountryOfBirth + Age + Income + Race, data = dataset)
# Sugar
lm_sugar_no_educ <- glm(Sugar ~ Gender + CountryOfBirth + Age + Income + Race, data = dataset)
# Summary
lm_summary_no_educ <- stargazer(lm_protein_no_educ, lm_carb_no_educ, lm_fiber_no_educ,
                                lm_fat_no_educ, lm_sugar_no_educ, type = 'text',
                           title = 'Without Educ',
                           digits = 2, report=('v*c*p*t'), out = 'no_educ_lm.txt')
```

```
# Factor Elasticity (No Educ)
# (A 1% increase in x causes a [elas%] cahnge in y)
protein_elas_noeduc_income <- as.numeric(lm_protein_no_educ$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Protein))
protein_elas_noeduc_age <- as.numeric(lm_protein_no_educ$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Protein))
fiber_elas_noeduc_income <- as.numeric(lm_fiber_no_educ$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Fiber))
fiber_elas_noeduc_age <- as.numeric(lm_fiber_no_educ$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Fiber))
carb_elas_noeduc_income <- as.numeric(lm_carb_no_educ$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Carb))
carb_elas_noeduc_age <- as.numeric(lm_carb_no_educ$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Carb))
fat_elas_noeduc_income <- as.numeric(lm_fat_no_educ$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Fat))
fat_elas_noeduc_age <- as.numeric(lm_fat_no_educ$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Fat))
sugar_elas_noeduc_income <- as.numeric(lm_sugar_no_educ$coefficients['Income'] * mean(dataset$Income) / mean(dataset$Sugar))
sugar_elas_noeduc_age <- as.numeric(lm_sugar_no_educ$coefficients['Age'] * mean(dataset$Age) / mean(dataset$Sugar))


# Standardize Income
dataset$Income <- scale(dataset$Income)
# (3) No Educ Race x Income
# Examine the quadratic relationship between education and nutrition intake
# Protein
lm_protein_no_educ_ri <- glm(Protein ~ Gender + CountryOfBirth + Age + Income + Race+ Race*Income, data = dataset)
# Carb
lm_carb_no_educ_ri <- glm(Carb ~ Gender + CountryOfBirth + Age + Income + Race+ Race*Income, data = dataset)
# Fiber
lm_fiber_no_educ_ri <- glm(Fiber ~ Gender + CountryOfBirth + Age + Income + Race+ Race*Income, data = dataset)
# Fat
lm_fat_no_educ_ri <- glm(Fat ~ Gender + CountryOfBirth + Age + Income + Race+ Race*Income, data = dataset)
# Sugar
lm_sugar_no_educ_ri <- glm(Sugar ~ Gender + CountryOfBirth + Age + Income + Race+ Race*Income, data = dataset)
# Summary
lm_summary_no_educ_ri <- stargazer(lm_protein_no_educ_ri, lm_carb_no_educ_ri, lm_fiber_no_educ_ri,
                                   lm_fat_no_educ_ri, lm_sugar_no_educ_ri, type = 'text',
                                   title = 'Without Educ Race x Income',
                                   digits = 2, report=('v*c*p*t'), out = 'no_educ_ri_lm.txt')
```

```r


# Run before income is standardized in Model 3
# Visualization
ggplot(dataset, aes(Age)) + geom_histogram(bins=10)

# -----------------Income-----------------------------
# (1) Income and Fiber
dataset %>%
  ggplot(aes(Income, Fiber, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income") + ylab('Fiber')
ggsave("IncomeFiber.pdf")

# (2) Income and Protein
dataset %>%
  ggplot(aes(Income, Protein, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income") + ylab('Protein')
ggsave("IncomeProtein.pdf")

# (3) Income and Sugar
dataset %>%
  ggplot(aes(Income, Sugar, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income") + ylab('Sugar')
ggsave("IncomeSuagr.pdf")

# (4) Income and Fat
dataset %>%
  ggplot(aes(Income, Fat, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income") + ylab('Fat')
ggsave("IncomeFat.pdf")

# (5) Income and Carb
dataset %>%
  ggplot(aes(Income, Carb, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income") + ylab('Carb')
ggsave("IncomeCarb.pdf")

```

```r
# ---------------------Income x Race------------------------
# (1) Income x Race (Carb)
dataset %>%
  filter(Race == 1) %>%
  ggplot(aes(Income, Carb, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Hispanic") + ylab('Carb')
ggsave("IncomeHisCarb.pdf")

dataset %>%
  filter(Race == 0) %>%
  ggplot(aes(Income, Carb, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Non-Hispanic") + ylab('Carb')
ggsave("IncomeNonHisCarb.pdf")

# (2) Income x Race (Protein)
dataset %>%
  filter(Race == 1) %>%
  ggplot(aes(Income, Protein, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Hispanic") + ylab('Protein')
ggsave("IncomeHisProtein.pdf")

dataset %>%
  filter(Race == 0) %>%
  ggplot(aes(Income, Protein, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Non-Hispanic") + ylab('Protein')
ggsave("IncomeNonHisProtein.pdf")

# (3) Income x Race (Fiber)
dataset %>%
  filter(Race == 1) %>%
  ggplot(aes(Income, Fiber, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Hispanic") + ylab('Fiber')
ggsave("IncomeHisFiber.pdf")

dataset %>%
  filter(Race == 0) %>%
  ggplot(aes(Income, Fiber, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Non-Hispanic") + ylab('Fiber')
ggsave("IncomeNonHisFiber.pdf")
```

```
# (4) Income x Race (Fat)
dataset %>%
  filter(Race == 1) %>%
  ggplot(aes(Income, Fat, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Hispanic") + ylab('Fat')
ggsave("IncomeHisFat.pdf")

dataset %>%
  filter(Race == 0) %>%
  ggplot(aes(Income, Fat, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Non-Hispanic") + ylab('Fat')
ggsave("IncomeNonHisFat.pdf")

# (5) Income x Race (Sugar)
dataset %>%
  filter(Race == 1) %>%
  ggplot(aes(Income, Sugar, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Hispanic") + ylab('Sugar')
ggsave("IncomeHisSugar.pdf")

dataset %>%
  filter(Race == 0) %>%
  ggplot(aes(Income, Sugar, col=Gender)) +
  geom_point() +
  geom_smooth(method="glm", color = 'red')+
  xlab("Income x Non-Hispanic") + ylab('Sugar')
ggsave("IncomeNonHisSugar.pdf")
```