Haley Mathews & Jennifer Russell
David Mimno
INFO 2950: Introduction to Data Science
May 14, 2019

What are the Significant Factors Affecting Average Red Meat Consumption in the United States?

Part 1: Exposition

The purpose of this analysis was to determine the most significant factors affecting the average red meat consumption (in tons per capita) in the United States from 1990 to 2017. Dieting trends, U.S. financial health, the population's opinion on climate change, and health factors were considered. It is crucial to recognize the limited impact of our analysis, as meat consumption data was only gathered from years 1990 - 2017, and some factors tested had even smaller amounts of data. However, the data is still viable for sufficiently analyzing America's relationship with red meat and considering the impact of different factors on the average American diet. Throughout the article, when referring to meat consumption, meat is defined as the average per capita intake of beef, lamb, and pork in tons. Poultry is not included.

After running tests to determine how significant the findings were, it was determined that GDP, inflation, government support for agriculture, vegetarian popularity, and caring "not at all" about climate change proved sufficiently significant in impacting meat consumption in the United States. Below, the five significant factors will be discussed. More specifically, the three most significant factors affecting average meat consumption are:

1. Inflation Rates
2. Vegetarian Diet Popularity
3. Caring "Not at All" About Climate Change

GDP

GDP proved to have a relatively strong positive correlation with average red meat consumption. When plotting GDP by year, GDP steadily rises until the year 2008, where GDP falls drastically for the next few years. In that time, average meat consumption also decreased. This GDP decrease can be attributed to the financial crash of 2008. At this time, Americans were in financial distress and thus cut back on luxury goods, such as eating meat.

Inflation Rates

Similarly correlated to the financial health of the United States is the inflation rate, which in relation to the other significant factors was the most impactful. Inflation is a measure of the rate the average price of goods raises over time. When the inflation rate rises, the price of food and other commodities is affected and people have "less" money to purchase more luxury items like meat.

Agricultural Support (TSE)

Agricultural Support (TSE) has a positive relationship with meat consumption. However, TSE was pretty insignificant in the presence of the other significant variables. TSE is the amount of money transferred from taxpayers to the agriculture sector from government policies supporting agriculture. Agricultural support directly affects how much money farmers have, so when farmers have less monetary resources there is an overall reduction in meat production. Since the subsidies the agricultural industry receive are dependent on economic factors such as GDP, there is a lot of overlapping significance between these two factors.

Vegetarian Diet Popularity

Vegetarian popularity was the second most significant factor impacting average meat consumption. When a person adheres to a vegetarian diet, they stop eating meat. As the popularity of the vegetarian diet increases, the average red meat consumption decreased rather significantly. Potentially, this could be correlated to GDP and inflation rates because as people need to cut out meat as a luxury good, they may just start a vegetarian diet.

Caring "Not at All" About Climate Change

Strong preferences for not caring at all about climate change was the last significant factor found in the analysis. When polled about their concerns on climate change, the larger percentage of people that answered "Not at All", the less the red meat consumption per capita people consume. Over time, more people have become more extreme in their opinions on climate change, leading to higher percentages of people either caring a great deal or not at all.

Part 2: Scientific Methods

Recurring Methods Utilized:

A linear regression model was fit to each individual set of data. Then, descriptive statistics, including the coefficient, the coefficient of determination, and correlation coefficient were obtained from the model. To obtain the correlation coefficient, a covariance matrix was created from the meat consumption data and the chosen variable. The Pearson coefficient was obtained from this matrix. Below, the specific statistical terms' meanings are defined:

- Coefficient: indicates whether the correlation between variables is positive or negative

- Coefficient of Determination: the proportion of the variance in the dependent y-variable that can be predicted from the independent x-variable (in this case, average red meat consumption)
- Correlation Coefficient: a measure of the strength of the linear relationship between two variables, ranging from -1 to +1

Next, random linear models were generated and compared to the collected data's model. A significance test was run on each data set to determine if the factor had a significant impact on average meat consumption.

To determine significance, a two-sided hypothesis test was conducted using a permutation test for each data factor. The null hypothesis was that the model was random, while the alternative hypothesis was that the model was significant. This was tested by comparing the steepness of the actual model's slope to the slopes of 1000 randomly generated models which were based off of a random ordering of data pairs. If the alpha level was less than 0.05, then the factor was considered significant enough to have impact average meat consumption. Alpha level is the probability of rejecting the null hypothesis when the null hypothesis was actually true. Since the test is two-sided, the test statistic was divided by two to determine the approximate percentage of random models with steeper slopes in the bottom or top 2.5% of the distribution.*

Beneath each factor's analysis, the results of the permutation test is shown. The first plot in the upper right corner is the real data which is displayed in red, while the other plots are the random data. When the linear model created from the random data has a more gradual slope, then the slope line is plotted in blue. If the slope is steeper, the slope line is plotted in red. The purpose of the permutation test is to determine the number of times a random linear model has a steeper slope than the actual model, which would signify that the factor is not significant in predicting beef trends.

Finally, once the significant factors affecting red meat consumption were chosen, a linear regression model was created with the five significant factors. The resulting output shows the coefficients of each variable:

GDP: -0.01752904
TSE: 0.00259907
Inflation Rate: -0.10969436
Vegetarian Popularity: -0.10355212
Caring "Not at All" About Climate Change: -0.0565222

*Permutation test statistics presented below may be slightly different then the notebook, as they rely on randomly generated models that differ each time it is run

Dieting Trends

Vegetarian

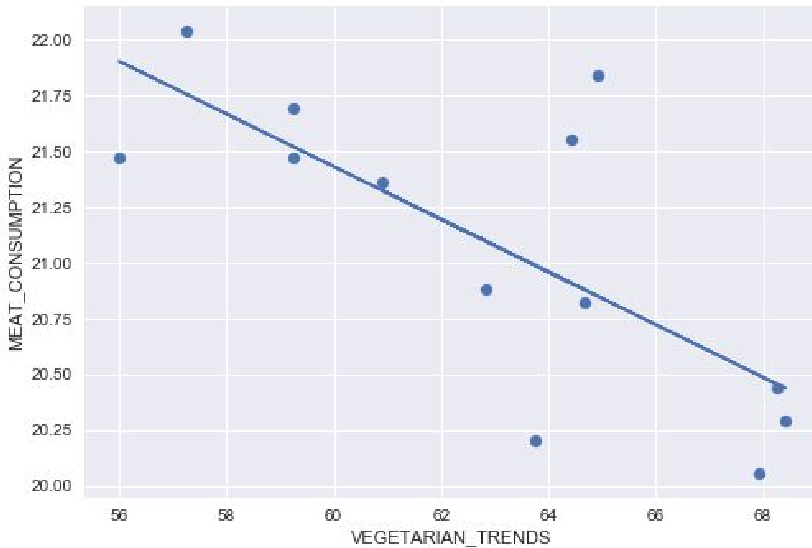

Figure 1. Vegetarian Popularity vs. Meat Consumption

After vegetarian popularity was scatter plotted against average red meat consumption for visualization purposes, a linear model was then fit to the data to determine the descriptive statistics. The coefficient was negative, showing that as vegetarianism gets more popular, average meat consumption decreases. The coefficient of determination was 0.527, showing that 52.7% of variation in average meat consumption can be predicted from variation in vegetarian popularity. Finally, the correlation coefficient was -0.726, indicated a strong negative relationship between vegetarian popularity and average meat consumption.

Generating the permutation test resulted in a test statistic of 0.001. Since 0.001 is less than 0.025, the linear model is very significant. The results of the significance test display that Google searches for vegetarian trends has a significant and non-random correlation with meat consumption trends in the U.S.

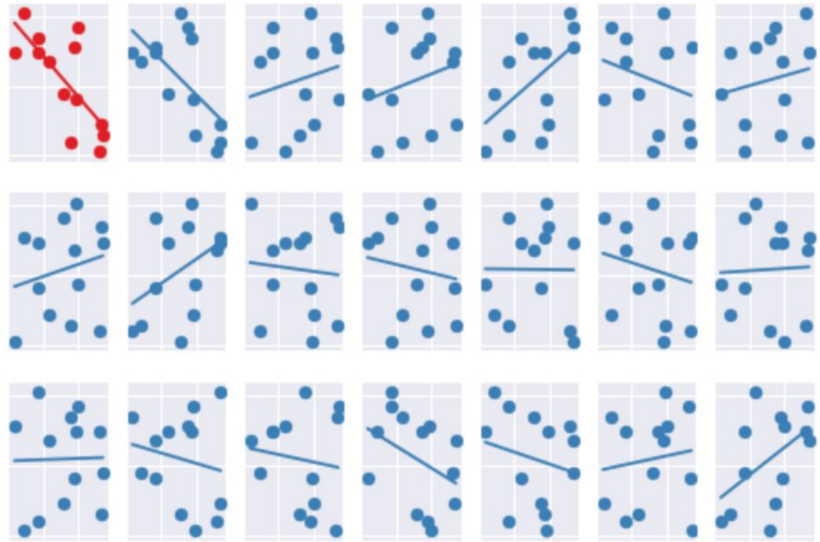

Figure 2: Vegetarian Popularity Permutation Test

Vegan

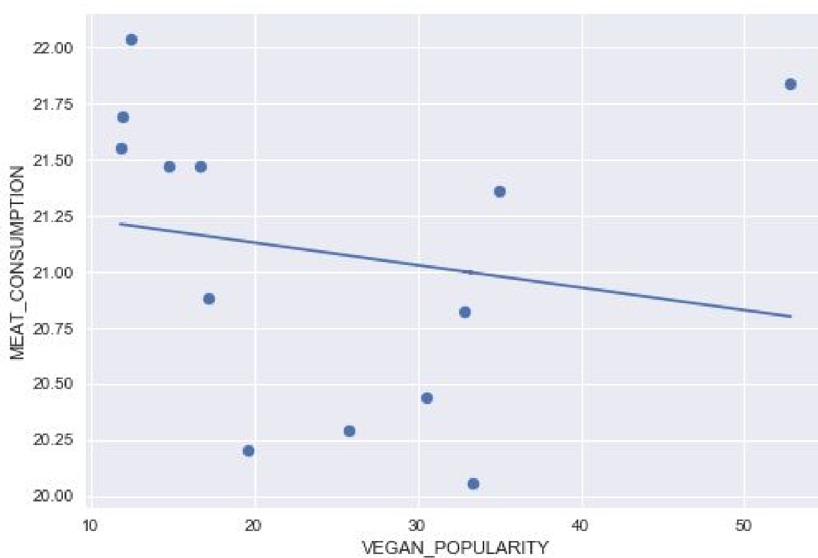

Figure 3: Vegan Popularity vs. Meat Consumption

After a scatter plot was created to visualize vegan popularity in comparison to meat consumption, a linear regression model was fit to this data to find descriptive statistics. First, the coefficient was negative, which indicates that as the popularity of the vegan diet increases, average meat consumption decreases. The coefficient of determination was 0.033, showing that only 3.3% of change in meat consumption can be explained by change in vegan popularity. Finally, the correlation coefficient was -0.183, proving a very weak negative correlation between vegan popularity and average red meat consumption.

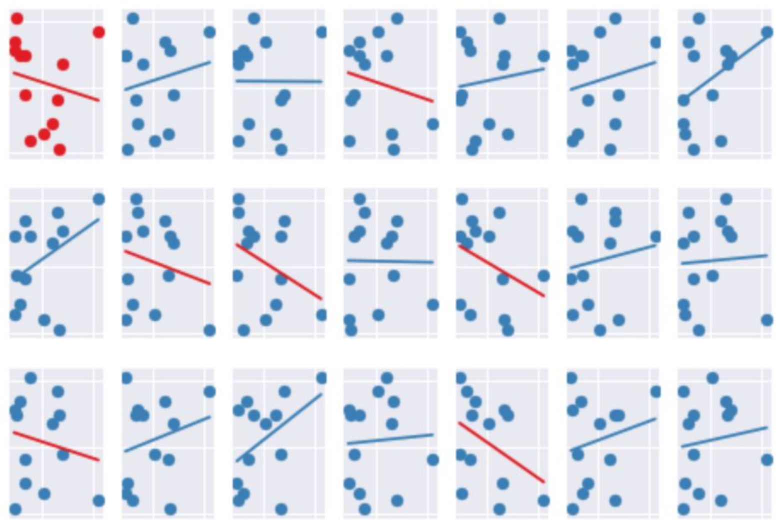

Figure 4: Vegan Popularity Permutation Test

Keto

A scatter plot was generated to visualize the relationship between the popularity of the keto diet and average meat consumption. Most of the data shows a popularity of 0.0 for keto searches, indicating that the keto data will have very limited reliability in predicting meat consumption.

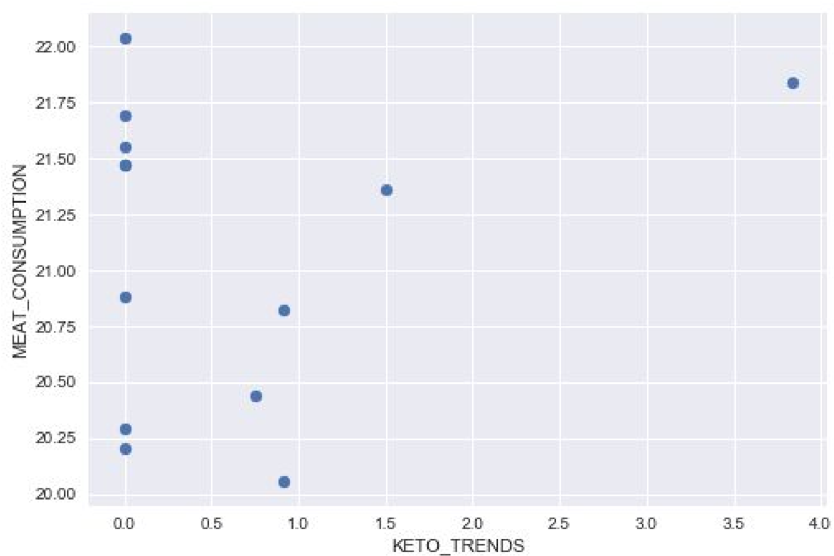

Figure 5: Keto Popularity vs. Meat Consumption

A linear regression model was fit to this data to determine some descriptive statistics. The coefficient was positive, indicating a positive relationship between the popularity of the keto diet and average meat consumption. As keto popularity increases, so does the average meat consumption. Next, the coefficient of determination was 0.337, showing that only 3.37% of variation in average meat consumption can be attributed to change in keto popularity. Finally, the correlation coefficient was 0.184, indicating a very weak positive correlation between vegan popularity and meat consumption.

A permutation test resulted in a test statistic of 0.18, far above the alpha cutoff of 0.025. This indicates any correlation between the keto diet's popularity and average red meat consumption is not significant enough to be considered.

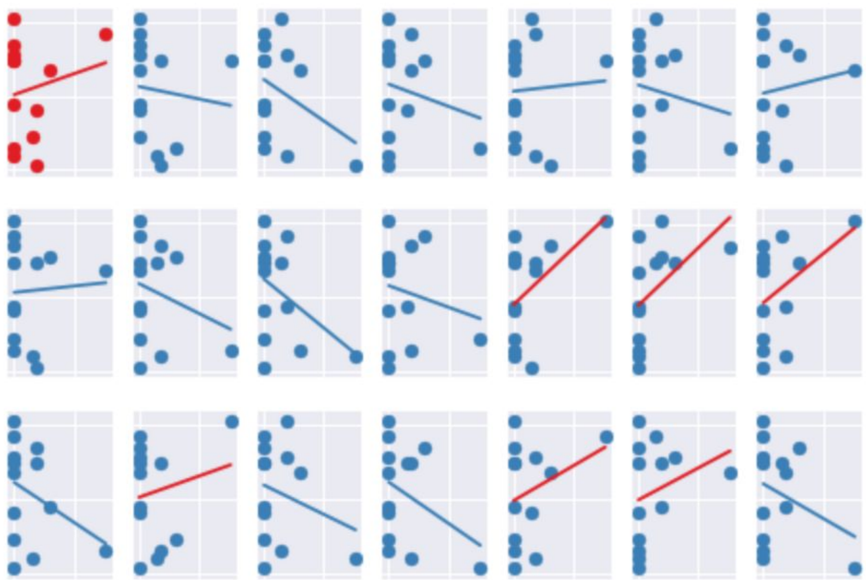

Figure 6: Keto Diet Popularity Permutation Test

Financial Health Data

GDP

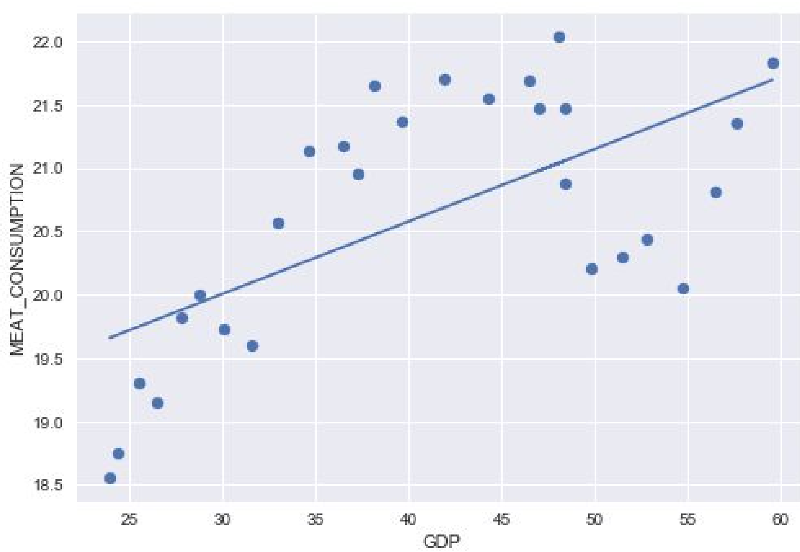

Figure 7: GDP vs Meat Consumption

A scatter plot was first generated to visualize the meat consumption compared to GDP. Then, a linear regression model was fit to the data, and descriptive statistics were found from this model. The coefficient was positive (0.057), indicating a positive correlation between red meat consumption and GDP. This means that as GDP increases, so does the average meat consumption. The coefficient of determination was 0.412, indicating that 41.2% of change in meat consumption can be explained from change in GDP data. Finally, the correlation coefficient was 0.612, indicating a relatively strong positive correlation between the GDP and average red meat consumption.

A permutation test resulted in a test statistic of 0.0, which is far below the alpha cutoff of 0.025. This indicates that GDP is a significant factor affecting meat consumption in the U.S.

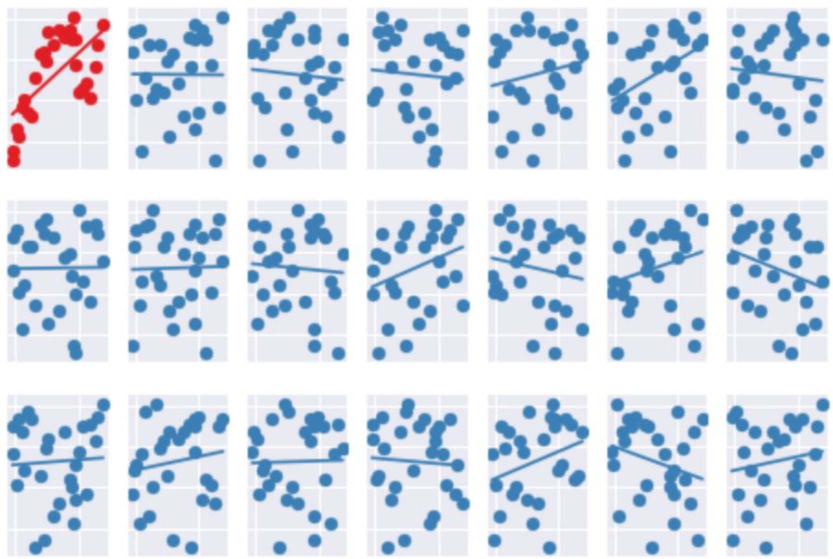

Figure 8: GDP Permutation Test

Inflation/Price Stability

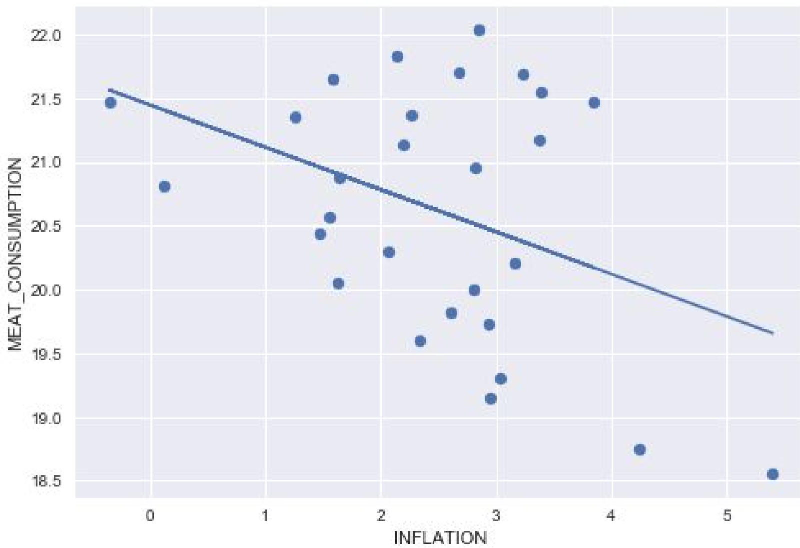

Figure 9: Inflation vs Meat Consumption

A scatter plot was first created to visualize inflation in comparison to average red meat consumption. After a linear regression model was fit to the data, descriptive statistics were found from this model. The coefficient was negative, indicating a negative correlation between inflation and red meat consumption. As inflation increases, meat consumption decreases. The coefficient of determination was 0.155, showing that about 15.5% of variation in red meat consumption can be elucidated from inflation variation. Finally, the correlation coefficient was -0.393, indicating a relatively strong negative relationship between inflation rates and average red meat consumption.

The permutation test resulted in a test statistic of 0.015. Since this is below the alpha cutoff of 0.025, it can be determined that the inflation rate has a significant effect on average red meat consumption in the United States.

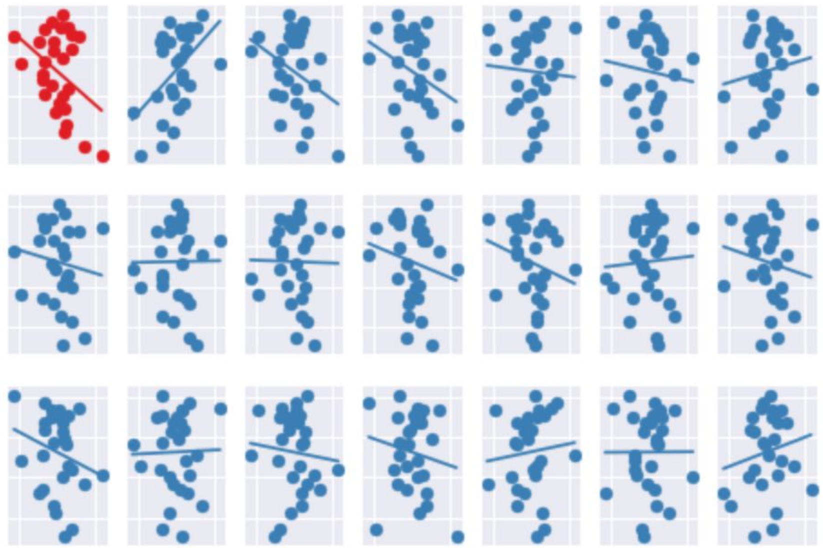

Figure 10: Inflation Rate Permutation Test

Unemployment Rate

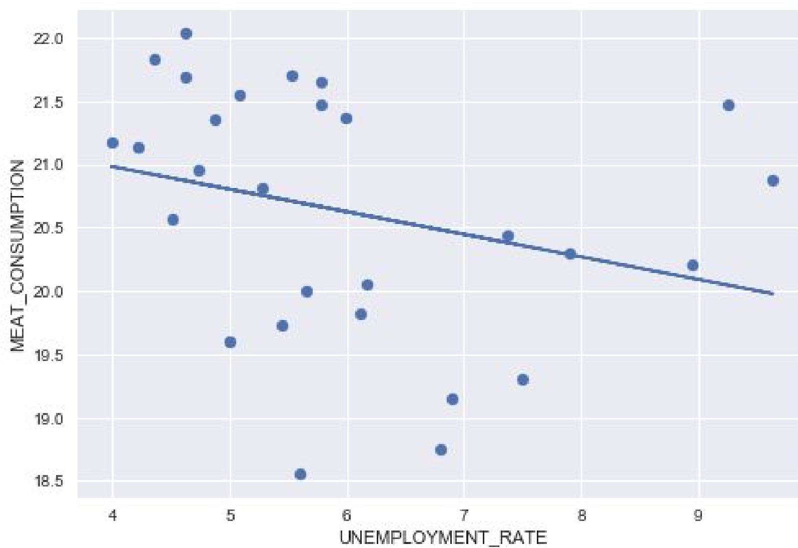

Figure 11: Unemployment Rate vs Meat Consumption

A scatter plot was generated to visualize the unemployment rate against average red meat consumption. A linear regression model was then fit to the data to determine descriptive statistics. The coefficient was found to be negative, indicating a negative relationship between unemployment rate and average red med consumption. Therefore, as the unemployment rate goes down, average meat consumption increases. The coefficient of determination was 0.077, showing that only 7.7% of change in average meat consumption is explained by unemployment rates. Finally, the correlation coefficient was -0.29, indicating a relatively weak negative correlation between unemployment rate and average red meat consumption.

A permutation test on the unemployment rate's effect on average red meat consumption resulted in a test statistic of 0.08. This is above the cutoff of 0.025, thus it can be determined that the unemployment rate does not have a significant enough effect on average red meat consumption in the United States to be considered.

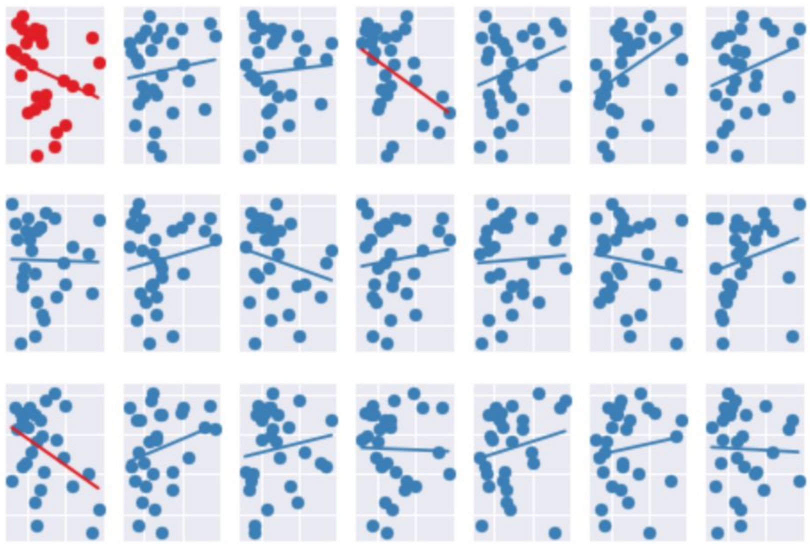

Figure 12: Unemployment Rate Permutation Test

Agricultural Support (TSE)

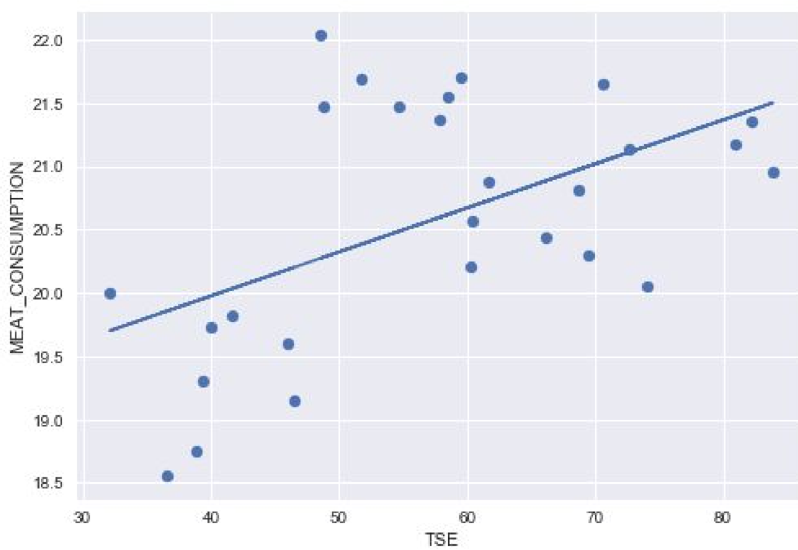

Figure 13: Agriculture Support vs Meat Consumption

To display the relationship between the total support estimate and meat consumption per capita, a scatterplot of the two variables was created. Then, a linear model was made to fit the TSE data against meat consumption. This resulted in a coefficient of about -0.332, meaning that as TSE increases by 1 billion U.S. dollars, red meat consumption decreases by 0.332 metric tons per capita. The correlation coefficient was found to be 0.527, showing a moderate positive relationship between TSE and meat consumption. The coefficient of determination of 0.278 means that 27.8% of variation in meat consumption can be explained by variation in TSE.

Finally, the TSE data was tested for significance by means of a permutation test. This resulted in a test statistic of 0.0025, revealing that only 0.25% of randomized linear models would have steeper positive slopes and 0.25% would have steeper negative slopes. 0.0025 is less than 0.025, showing that TSE has more significance than being random data.

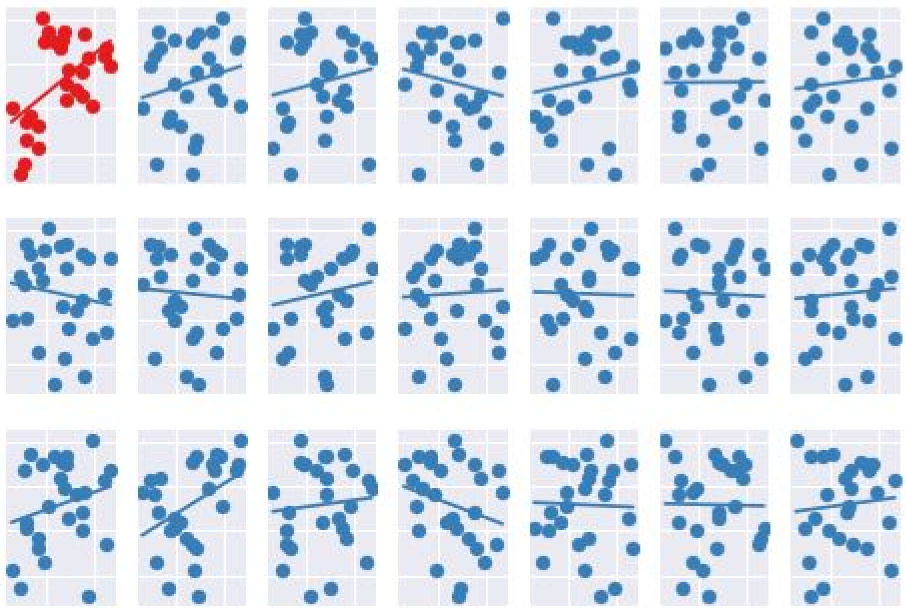

Figure 14: TSE Permutation Test

Population's Concern on Climate Change

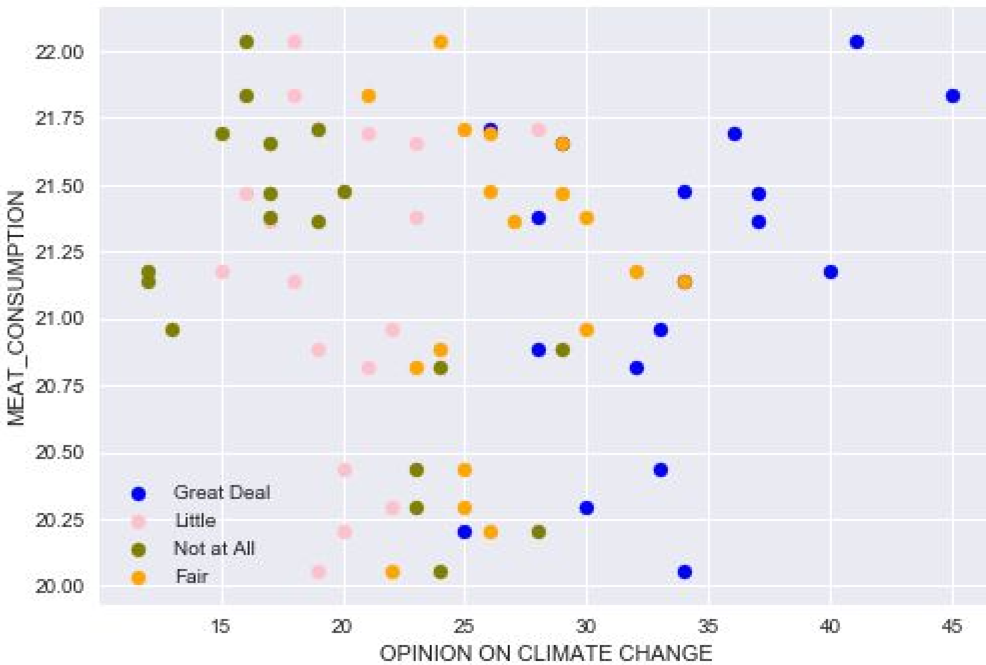

Figure 15: Opinion on Climate Change and Meat Consumption

Great Deal

A scatter plot comparing the percentage of people who cared a "great deal" about the environment was generated to visualize the relationship between the two variables. Then, a linear regression model was fitted to the data. The coefficient for the model was 0.048, which indicates that as "great deal" increases by 1%, meat consumption increases by 0.048 metric tons. The correlation coefficient was 0.433 which indicates a weak-to-moderate positive relationship between the two variables. The coefficient of determination was 0.187, meaning that 18.7% of change in meat consumption can be explained by the change in "great deal."

A permutation test was run on the data regarding people concerned a "great deal" about climate change. The test returned a test statistic of 0.0345, that while small, was not considered significant as it was larger than our alpha level of 0.025. Therefore, the data was not considered to be significant.

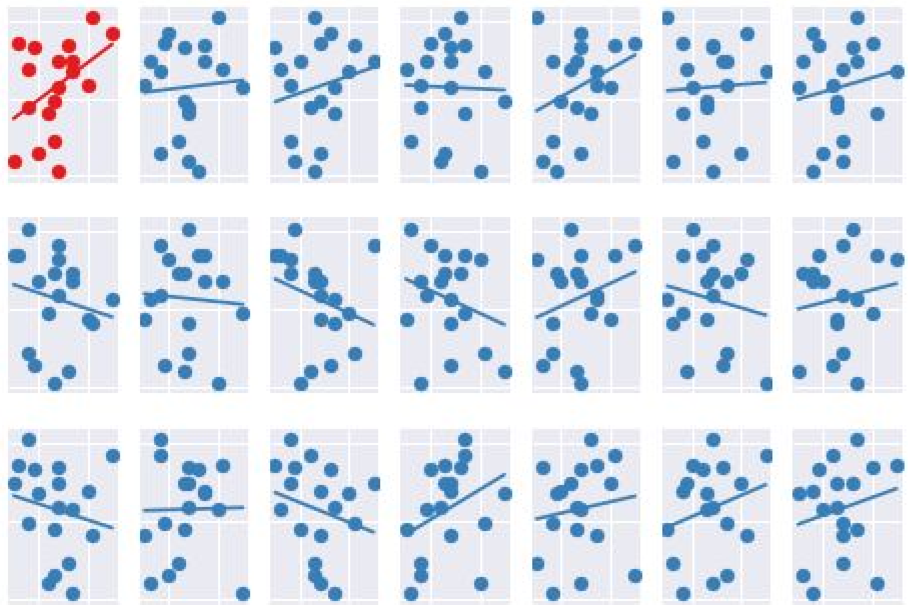

Figure 16: “Great Deal” Permutation Test

Fair

Figure 15 refers to the scatterplot created for all data of opinions on climate change. Then, a linear model for data on the “fair” response was created to assess the relationship between “fair” and meat consumption. This model had a coefficient of 0.022, indicating that for every 1% increase in “fair”, meat consumption rises by 0.022 metric tons. The correlation coefficient was 0.13, displaying that there was a very weak relationship between the variables. The coefficient of determination was 0.018, so little variation in meat consumption can be accounted for by variation in “fair.”

A t-test using the permutation method was then run on the data, resulting in a test statistic of 0.294, which is far greater than the alpha level of 0.025 meaning that “fair” was not significantly statistic.

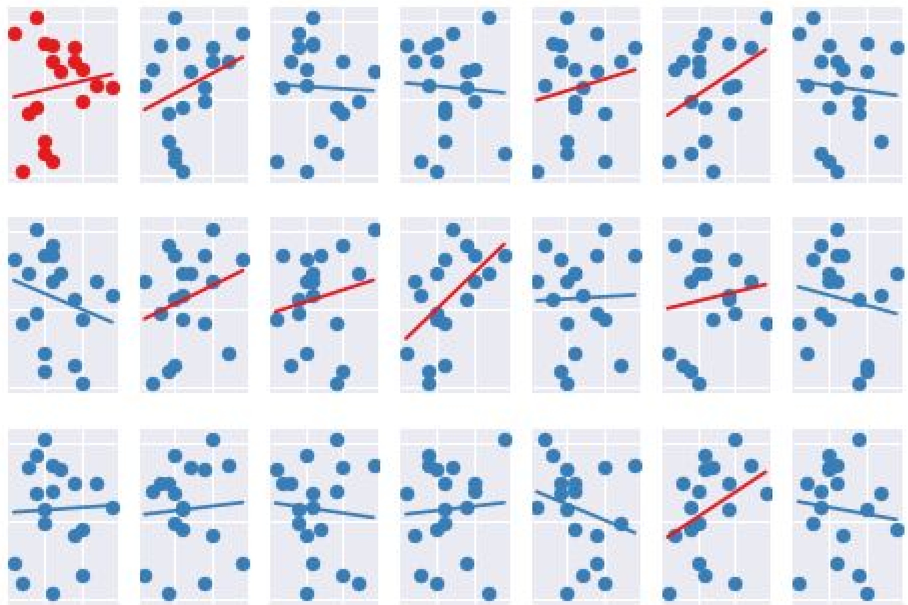

Figure 17: “Fair” Permutation Test

Little

The scatterplot created for the relationship between meat consumption and opinion “Little” can be shown in Figure 15. A linear model was created to further analyze this relationship. The coefficient was 0.006, meaning as “little” increased by 1%, meat consumption increased by 0.006 metric tons. The correlation coefficient was 0.03, showing no relationship between the variables. The correlation coefficient was 0.001, further showing a negligible relationship between variables.

The test statistic of the permutation test done for the “little” variable was 0.4535 which is much larger than the set alpha level, revealing that caring about the climate change a “little” has no significant effect on meat consumption.

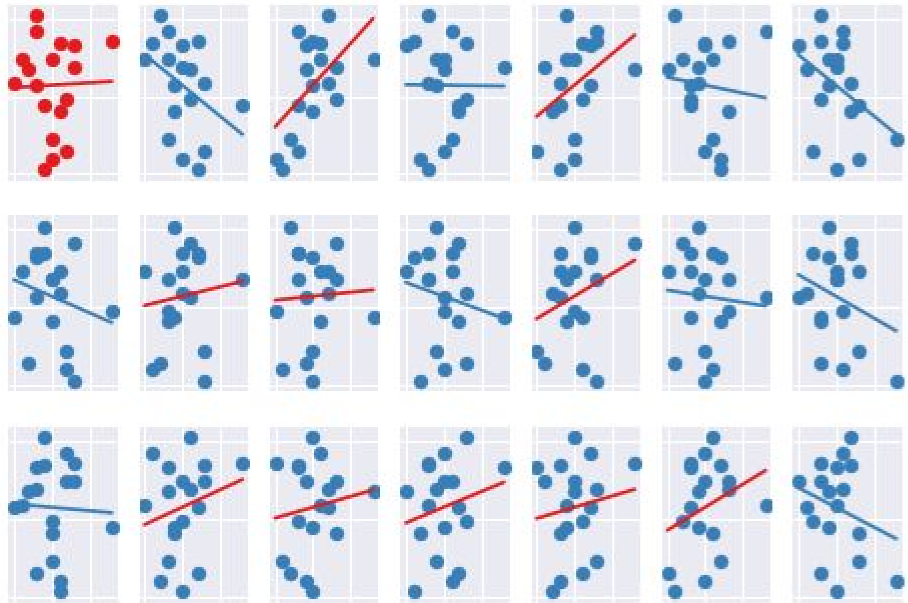

Figure 18: “Little” Permutation Test

Not at All

The opinion “Not at All” is plotted against meat consumption per capita in Figure 15. A linear model was created to determine the relationship between people not caring about the environment at all and meat consumption. As “Not at All” goes up by 1%, meat consumption decreases by 0.07 metric tons, resulting from the model’s coefficient of 0.07. The correlation coefficient is -0.6257, a moderately strong negative correlation between the two variables. The correlation of determination is 0.3915, interpreted as 39.15% of variation in meat consumption can be explained by variation in “Not at All”.

To determine the significance of this relationship, a permutation test was run on the data. The test produced a test statistic of 0.0035, which was significantly lower than the alpha level of 0.05, Therefore, the variable “Not at All” is significant in determining meat consumption per capita.

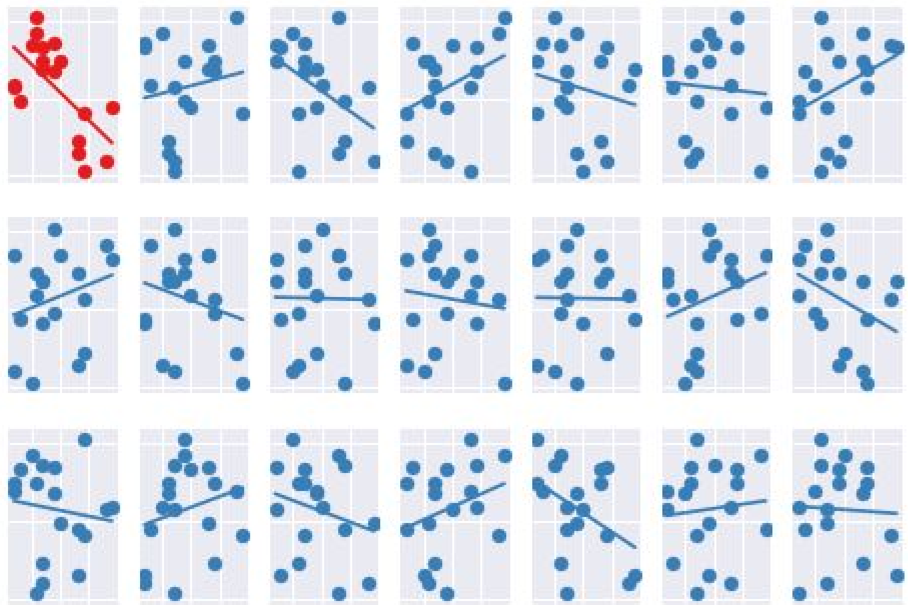

Figure 19: “Not at All” Permutation Test

Health Data

Obesity Rate

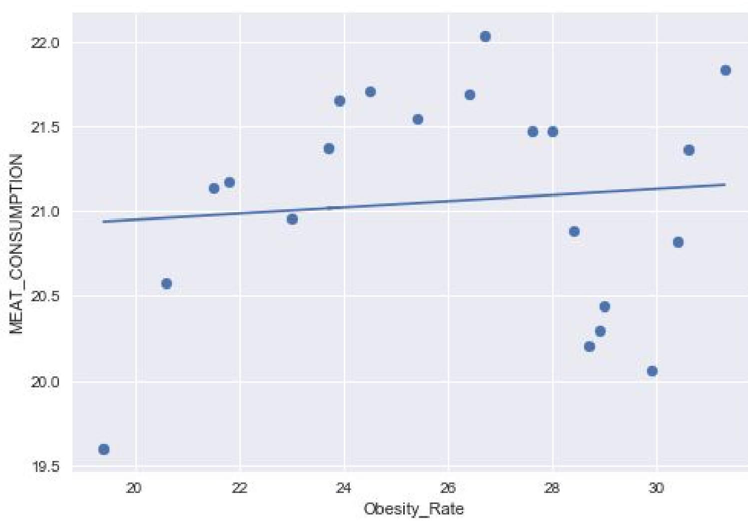

Figure 20: Obesity Rate vs Meat Consumption

A scatterplot was created to visualize the relationship between obesity rates in the U.S. and meat consumption. Then, a linear model was constructed to further analyze the relationship. The resulting model had a coefficient of 0.018, meaning that as obesity rates increased by 1%, meat consumption per capita increased by 0.018 metric tons. The correlation coefficient was 0.098, showing that there was an extremely weak relationship between the variables. The coefficient of determination was 0.009, meaning approximately 0% of meat consumption variance can be determined by obesity rates.

Afterwards, a permutation test was run to determine the significance of obesity rates on determining meat consumption per capita. This resulted in a test statistic of 0.3335, which is insignificant since it is much larger than the alpha level. It can then be determined that obesity rates are not a significant factor for meat consumption per capita.

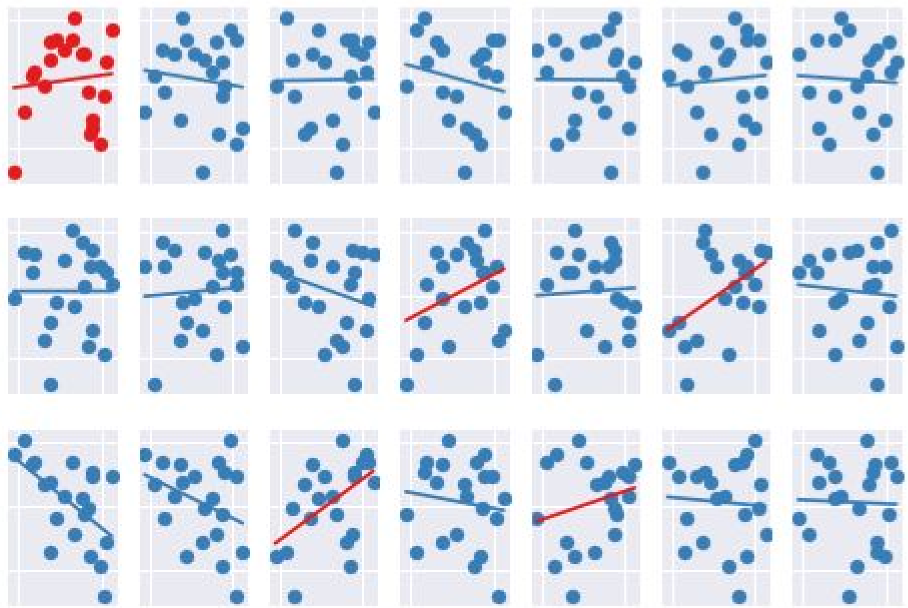

Figure 21: Obesity Rates Permutation Test