# Midterm Project

# Business Analysis of Annual Food Imports by Source Country

## Group 1

**Cameron Rangecroft, Balram Prasad, Camila Gomez Liendo, Claire Zhu**

# Table of Contents

Abstract

Food product in America is nearly as diverse as the people who consume it. With a variety of food imported to the United States, there is cause for question as to what factors lead to import differences across food groups. To better address the needs of Supermarket clientele, analysis of five major food groups was done to see long term regression analysis and predictive changes for the upcoming year. There was percent growth in import total seen for four of the five major food categories (Fruit, Dairy, Grains, Vegetables) and negative growth seen for one of the five (Meat). The regression models that were best suited to each kind of food group were linear for four food groups with one group having Second Degree Polynomial (Dairy). Based on the percent growth and adjusted R-squared values explored and analyzed, the recommendation for our client is to invest their capital in grains; as a food group, grains was seen to have the highest percent change at 9.62%, with an adjusted R-squared of 0.9539. While this was not the highest adjusted R-squared value seen throughout all five food categories, when paired with the highest percentage change we believe that it is the best food to put capital in.

<div align="center">Introduction</div>

Agriculture is a world-wide market that opens the door to a variety of business opportunities; "agricultural-processing industries" rely heavily on imports, such as fruit, vegetables, meat, dairy and grains imports. "Over the past two decades, the United States Department of Agriculture has issued roughly 100 new rules allowing specific crops to be imported from certain countries" (*Karp*). As a result, "crops that previously would have not been approved because they might introduce invasive pests and diseases were allowed in through new "systems approaches" that manage those risks by combining methods like orchard inspections, sprays and bagging of fruits" (*Karp*). Therefore, many foreign crops have been recently approved to be imported by using these new "systems approaches" thus increasing imports in the United States.

Our client is the owner of a grocery store, where he is looking to invest on imports of a specific food group. In order to advise our client, we are going analyze the total number of imports to the United States from 2002-2017, which in turn can be used to create a predictive model for the years 2018. The purpose of creating this model is to be able to predict which food group is going to increase the most. This will allow our client to invest in capital that will help grow his capabilities.

Knowing that there is an annual growth seen in the United States imports over the course of the years due to new policy changes and higher demand, we want to estimate the annual price for fruit, vegetables, meat, dairy and grains imports in 2018. We believe that the imports will increase over 2018 yet we want to see which food group will increase the most so our client can invest his capital on the one that will give him the highest returns. For the purpose of this project,

we are not going to look at cost structure, profit margins, customer preference, or domestic supply. We are only taking into consideration the total number of imports for each food group.

## Methods

### *Data Collection:*

The data we used to complete this analysis was collected by the U.S. Department of Commerce and the U.S. Census Bureau. The data provides the import values in millions of dollars for a variety of food products; we limited our selection of edible products to the following categories: Dairy, Fruits, Grains, Meat, and Vegetables. The import for these groups were compiled by year, and subset by Country of Origin.

### *Data Understanding:*

The data set is a compilation of a variety of data; the first subset within the data frame is Total U.S. Food imports in Millions of dollars, there are then continuing subsets that are separated by food type (Dairy, Fruits, Grains, Meat, Vegetables, etc.) from which we chose to look specifically at the five major food groups. Within each food group there are categories of data that break down the food group farther, we chose to look at the "Total food and preparations," which is the complete total of the imports for each food type by country.

### *Data Preparation:*

Due to the fact that we were limiting the original data set to a much smaller data set, we created a new data frame that contained specifically the data we were interested in analyzing. There were no missing data within the original set which allowed us to choose our data without issue. The measures we chose not to use were specifically left out to limit our data set. We did not need the complete source data as it was fairly extensive. By limiting our data to a smaller subset, we were able to focus our study to a specific group and apply that to our clients needs.

### *Modeling:*

We decided to do a time series model to predict food imports for the 2018 calendar year; to do this, we began by graphing each food category import by year and country to see what the general trends would be for each specific food group. Using R, we plotted Imports (Millions $) vs. Year, and subset by country. We continued by creating regression models, we used a linear regression model for Grains, Fruits, Vegetables, and Meat, and a second degree polynomial model for Dairy.

We used regressions as our modelling and predicting methods. The type of regressions for each food group was determined by experimenting with different methods and the method that gave the highest adjusted R-squared and made the most reasonable predictions in economic terms (for example, the predicted imports should not be negative) was chosen to produce the final model.

As a result, we used the second degree polynomial regression for Dairy and linear regression for the rest of the food groups.

The choice of linear regression method for 4 food groups was justified since our data set met the following assumptions: 1) we only dealt with one variable "Year," so there was no collinearity between variables; from the residual plot for each food group model (Figure 1-4 in Appendix), we verified 2) constant variance, 3) normality of residuals, and 4) no autocorrelation.

<center>Results</center>

***Evaluation:***
After running the different regressions, we were able to obtain our prediction equations and our adjusted r-squared and p-values for each food group, which can be found in Tables 1 and 2.

*Table 1: Results from linear regression by group of food*

| Food Group | Intercept | Coefficient (year since 2002) | Adjusted R-Squared | P-Value |
|---|---|---|---|---|
| Fruit | 4351.9 | 877.9 | 0.9843 | 2.995e-14 |
| Grains | 3170.9 | 545.9 | 0.9539 | 5.829e-11 |
| Vegetables | 4416.9 | 552.2 | 0.9897 | 1.609e-15 |
| Meat | 2857.21 | 279.22 | 0.6404 | 0.0001196 |

Intercept is the imports number from 2002 while the coefficient is the annual change.

*Table 2: Results from second degree polynomial model for Dairy*

| Food Group | Adjusted R-Squared | P-Value |
|---|---|---|
| Dairy | 0.8172 | 6.299e-06 |

We can observe that all of our p-values are less than alpha 0.05 and our adjusted r-squared values are relatively high. Additionally, from the results shown above we can construct our prediction equations shown below:

$$\widehat{Fruit} = 4351.90 + 877.90 * (\text{Year since 2002})$$
$$\widehat{Grains} = 3170.90 + 545.90 * (\text{Year since 2002})$$
$$\widehat{Vegetables} = 4416.90 + 552.2 * (\text{Year since 2002})$$
$$\widehat{Meat} = 2857.21 + 279.22 * (\text{Year since 2002})$$
$$\widehat{Dairy} = 1114.07 + 57.15 * (\text{Year since 2002}) - 0.39 * (\text{Year since 2002})^2$$

In order to obtain our predicted value for the year 2018, we plugged in the year into our equation. Note that 2018 is year 16 in our model since our first year was 2002. Furthermore, we calculated the percent change from 2017 to 2018 by using the value of total imports for 2017 from our data set.

*Table 3: Projected Value for 2018 and Percent Change From 2017-2018*

| Food Group | Prediction for 2018 | Percent Change from 2017-2018 |
| --- | --- | --- |
| Fruits | 18398.3 | 0.08% |
| Grains | 11905.3 | 9.62% |
| Vegetables | 13252.1 | 4.00% |
| Meat | 7324.73 | -0.02% |
| Dairy | 1928.63 | 5.55% |

From the results above we can observe four food groups (Fruits, Grains, Vegetables and Dairy) have positive growth and one food group (Meat) has negative percent growth. We were expecting for all food groups to increase, yet there may be some factors affecting the import of meat in the past year such as but not limited to trade wars, climate change, customer preference, etc.

## Discussion

***Interpretation and Recommendation:***
After analyzing each food groups regression model and distinguishing based on percent growth and adjusted R-squared, our recommendation is that our client should invest capital toward grains; our reasoning for this is that grains has the highest percent growth with a reasonable high adjusted R-squared. While the adjusted R-squared for grains is not the highest, it is still at a value of 0.95, which means that 95% of the variance in the change in year can be explained by the change in imports for grains. While other models produce a higher adjusted R-squared value, those models do not yield as much of a considerable gain in percent growth; because of this, our recommendation is to invest the capital in grains to increase our capacity for grains. Fruit and Vegetables have the highest adjusted R-squared, but their percent growth is considerably lower than that of grains, thus we would not recommend pursuing capital in those food groups primarily. Similarly we would highly discourage investing capital in meat as meat imports as seen to have a negative percent growth in imports, and as such do not project a demanding or growing market.

Furthermore, we believe that increasing capital investment in grains is practical and easy to implement for our client. Fruits, vegetables, dairy products, and meat all need heavy investment into their storage method in order to keep them fresh. In contrast, grains need relatively less investment into their storge and therefore enjoy a relatively higher profit margin.

*Limitations:*

Essentially, our model only used previous imports data from each food group to project future changes. Realistically, there are many other factors that would influence US imports. For example, political environment, such as a trade war between US and its importing countries, might alter the imports numbers significantly. Other factors include the economic situations in the US as well as its importing countries (which will influence the currency exchange rate and the buying power at each side), climate change, and cost structure. One way to improve the predicting power of our models is to include some of the factors above as explanatory variables.

*Potential Improvements:*

Another potential statistical technique that could be used moving forward to gain more accurate understanding of food imports by food group would be to analyze each food group based on subset of country of origin. In doing this we would be able to use different regression models and base them more specifically on the trends seen on a country-to-country basis rather than using Total Food Import to the US as a baseline. Based on our current recommendation for our client to put capital into the grains market, an improvement that could be made to their business to better assist our target goals would be to closely analyze purchasing patterns in their own store to better apply future recommendations as well. We see that there is a linear trend in four of the five food groups; as years progress we expect to see this trend continue, the main factor to focus on will be percent change between years and giving our client the ability to utilize this information accordingly will better prepare the supermarket on a yearly basis.

## Works Cited

1. *U.S. Department of Commerce, U.S. Census Bureau. (n.d.). Summary data on annual food imports, values and volume by food category and source country, 1999-2017. https://www.ers.usda.gov/data-products/us-food-imports.aspx. (accessed Oct 14)*
2. Karp, D. (2018, March 13). Most of America's Fruit Is Now Imported. Is That a Bad Thing? https://www.nytimes.com/2018/03/13/dining/fruit-vegetables-imports.html (accessed Oct 14)
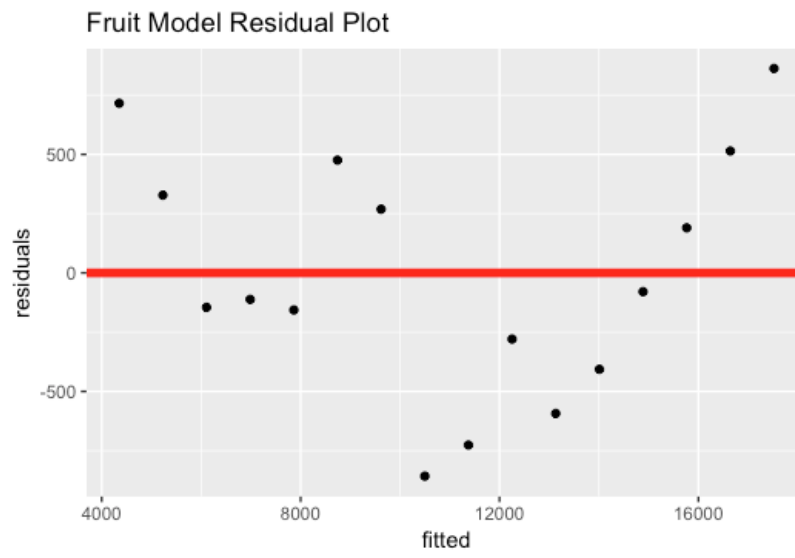
*Figure 1.* Fruit Model Residual Plot
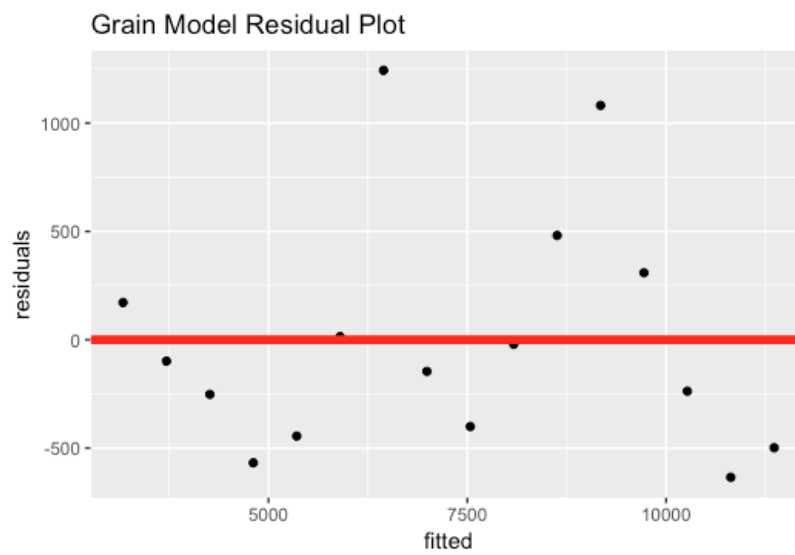


*Figure 2.* Grain Model Residual Plot

*Figure 3.* Meat Model Residual Plot
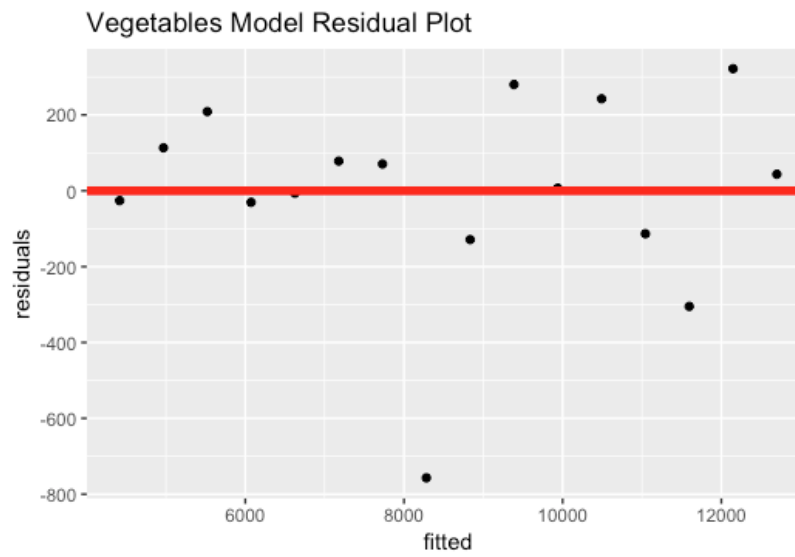


*Figure 4.* Vegetables Model Residual Plot
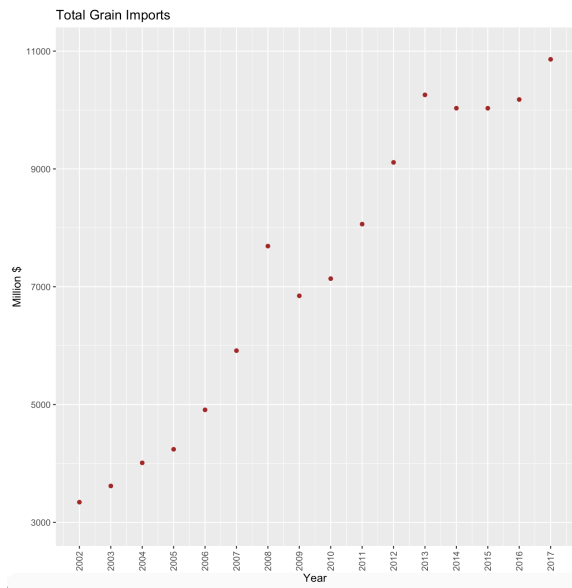
*Figure 5:* Scatter Plot of the Total Grain Imports in Millions of Dollars by Year
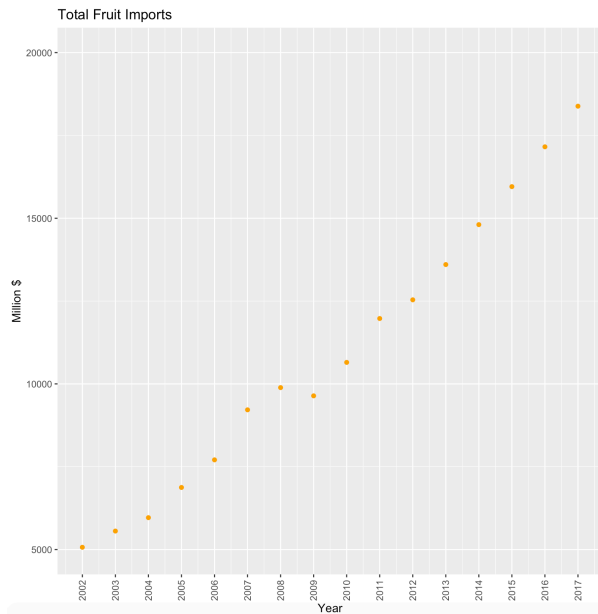


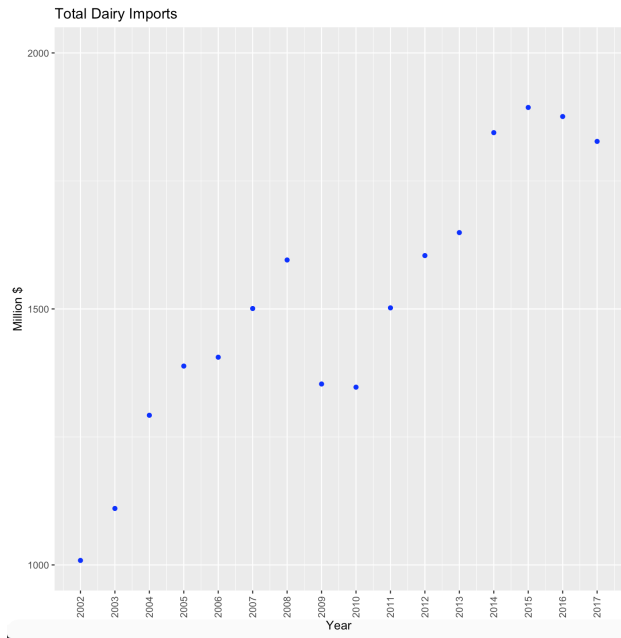*Figure 6:* Scatter Plot of the Total Fruit Imports in Millions of Dollars by Year

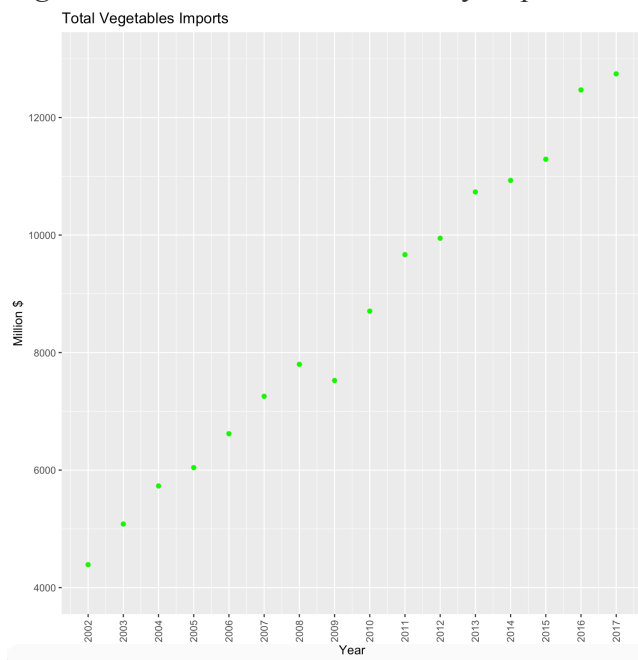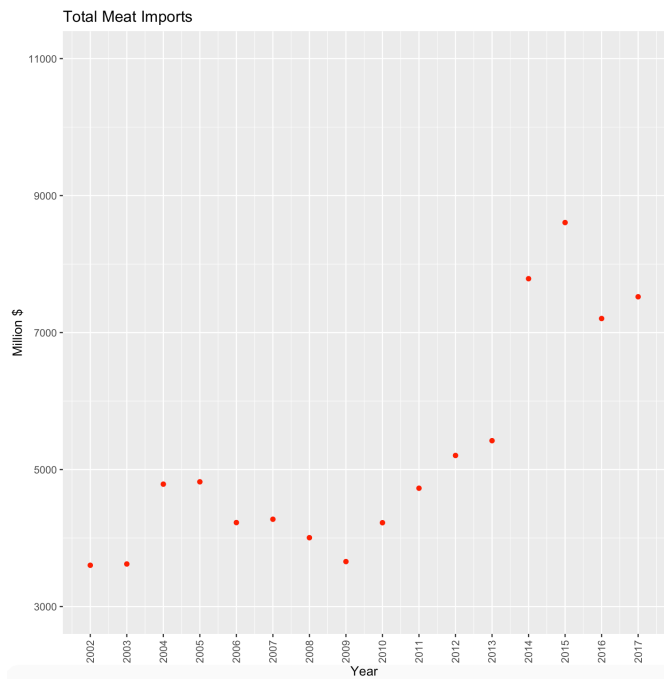*Figure 7:* Scatter Plot of Total Dairy Imports in Millions of Dollars by Year



*Figure 8:* Scatter Plot of Total Vegetable Imports in Millions of Dollars by Year

*Figure 9:* Scatter Plot of the Total Meat Imports in Millions of Dollars by Year