

Analyzing Healthy Diet Accessibility & Affordability on a Global Scale

Summary

For our project, we're taking a look at the cost and affordability of healthy food around the world and migration, fertility, and growth rates, as well as the life expectancy of these countries. We aim to identify which countries have experienced the highest and lowest levels of food insecurity and if the cost of a healthy diet may contribute to food insecurity. We also seek to explore if food insecurity may play a factor in a country's net migration, fertility rates, and life expectancy. Our findings revealed that the countries with the highest rates of food insecurity were Burundi, Liberia, Madagascar, and the Democratic Republic of the Congo. On the other hand, Switzerland, Iceland, Azerbaijan, and the United Arab Emirates faced the lowest levels of food insecurity. Additionally, a positive correlation, though not particularly strong, was observed between the cost of a healthy diet and the percentage of a country's population unable to afford such a diet, suggesting that higher costs may contribute to increased food insecurity. Finally, we found that the countries which had a higher share of their population unable to afford a healthy diet had higher fertility and growth rates, lower immigration rates, and a decreased life expectancy.

Team

My contributions: data preparation, background, aims, datasets, methods, and discussion sections, made the tables and scatterplots, contributed to the variable summaries and regression analysis, and did the formatting.

Chen Guo: figure captions, correlation heatmap, kickstarted the data cleaning and analysis.

Carly Greutert: data preparation, regression analysis, discussion, and datasets portions and chose the project title.

Takeen Shamloo: summary and the variable summaries.

Introduction

Background

This project is centered around the cost and affordability of a healthy diet across the world and the relationship between this affordability and population estimators (e.g. migration and fertility rates, life expectancy). The area of study this project is in dialogue with is public health, nutrition, and economics. The data mainly includes information about food costs and the affordability of food in countries around the world. We also pull from another dataset that contains global population data that may be useful in understanding how food affordability may affect a country's population. The motivation for collecting and analyzing food cost and affordability data across different countries is so we can gain insight into the affordability and accessibility of healthy diets in different contexts, as well as globally, and identify any disparities in food access across the different countries. We can also see how increased unaffordability may impact the life expectancy as well as migration, fertility, and growth rates of the countries.

Aims

The first aim of this project was to identify which countries experience the highest and lowest levels of food insecurity and how the cost of food may play a role in food insecurity. To accomplish this, we identified the countries that had the highest and lowest percentages of their population that were unable to afford a healthy diet. Then we plotted the percentage of the population that was unable to afford a healthy diet against the cost of a healthy diet for all the countries to see if there we could identify a relationship between the two and if so, what?. We found that Burundi, Liberia, Madagascar, and the Democratic Republic of the Congo had the highest share of their

population unable to afford a healthy diet and that Switzerland, Iceland, Azerbaijan, and the United Arab Emirates had the lowest. We also found that there was a positive (albeit, not very strong) correlation between the cost of a healthy diet and the percentage of a country's population that was unable to afford a healthy diet.

The second aim of this project was to see how food insecurity may play a factor in a country's net migration, fertility, and growth rates, as well as life expectancy. Our approach was to graph the percentage of the countries' population that was unable to afford a healthy diet against the net migration rate, fertility rate, growth rate, and life expectancy of the countries to identify if there was a relationship between them visually and if so, what? We found that the countries with a higher share of their population unable to afford a healthy diet also had higher fertility rates and lower immigration rates and life expectancy. We also made a correlation heatmap to reaffirm the relationships we identified for both approaches.

Materials and methods

Datasets

Two data sets were used. The first data set contains the cost and the affordability of a healthy diet, as well as the percentage of the population and the number (in millions) of people who can't afford a healthy diet for countries around the world in 2017. The second data set describes attributes of various countries, including population size, life expectancy, annual growth rates, fertility rates, and migration rates in the year 2017.

The healthy diet data was sourced from the [World Bank Group's Food Prices for Nutrition DataHub](#) which uses food item availability and prices from the International Comparison Program (ICP), combined with food composition data and nutritional requirements from a wide range of sources—including national dietary guidelines—and provides data for most of the countries around the world. The data was collected as part of the Food Prices for Nutrition project which was established to provide governments and development agencies with accurate and updated metrics to inform agricultural and food systems interventions. The census data is sourced from the [United States Census Bureau](#) and includes data for 227 countries and areas. For the sake of merging, we only used data for the countries that are also included in the food dataset. We also only analyzed countries that had data for all the variables we wanted to analyze. In the end, we had a data set containing information for 186 out of 195 countries in the world.

The data on healthy diet was collected using observed consumer prices and household expenditures to provide an operational measure of people's access to locally available foods in the proportions needed for health. The initial results are based on national average prices in 2017, expressed in terms of purchasing power parity dollars per day. Foods included as part of a healthy diet were based on nutritional standards set by dietary guidelines, with sufficient diversity and quantity within and between food groups to achieve nutrient adequacy and protect against diet-related diseases. To measure access and affordability, the cost of meeting food-based dietary guidelines is computed using the least expensive foods available in each category at each place and time to meet the needs of a representative adult requiring 2330 kcal per day. The affordability of each diet is measured by comparing diet costs to the food component of the international poverty line. The census data was collected using hundreds of data sources around the world, including censuses, surveys, administrative records, and vital statistics. National statistical offices of other countries are one of the primary resources. The scope of inference for this data is limited to the countries and regions included in the ICP data set. However, the data can still provide valuable insights into the affordability and accessibility of healthy food options across a range of contexts.

Descriptions for the observed variables we selected for use in our analysis and a glimpse at our data set are provided in Tables 1 and 2 (respectively) below.

Table 1. Variable Summaries

Variable name	Description	Type	Units of measurement
Country Name	The name of the country for which the data is attributed to	object	None
CoHD	Cost of the least expensive locally-available foods to meet requirements for food-based dietary guidelines, in current PPP dollar/person/day, for a representative person within energy balance at 2330 kcal/day	numeric	PPP (Per dollar Per person Per day)
COHD_pov	The ratio of the cost of a healthy diet to the 99 cent food poverty line (52% of the international poverty line of 1.90/day in 2011 PPP dollars)	numeric	PPP/\$0.99 food poverty line
CoHD_headcount	The share of the population whose food budget is below the cost of a healthy diet. The food budget is defined as 52% of household income, based on the average share of income that households in low-income countries spend on food.	numeric	percentage
CoHD_unafford_n	The number of people whose food budget is below the cost of a healthy diet. The food budget is defined as 52% of household income, based on the average share of income that households in low-income countries spend on food.	numeric	count (in millions)
Population	The number of people residing in the country	numeric	count(n)
Annual Growth Rate %	The percentage increase or decrease of a country's population in a year.	numeric	percentage
Total Fertility Rate	The total number of children that would be born to each woman if she were to live to the end of her child-bearing years and give birth to children in alignment with the prevailing age-specific fertility rates	numeric	count(n)
Life Expectancy at Birth, Both Sexes	The number of years a person (male or female) born in this country is expected to live.	numeric	years
Net Migration Rate	The rate in which people are migrating in or out of the country. Calculated by taking the number of immigrants minus the number of emigrants over a period, divided by the population.	numeric	percentage

Table 2. Preview of Data Set

	Country Name	CoHD	CoHD_pov	CoHD_headcount	CoHD_unafford_n	Population	Annual Growth Rate %	Total Fertility Rate	Life Expectancy at Birth, Both Sexes	Net Migration Rate
0	Albania	3.952	3.992	37.8	1.1	3055278.0	0.305	1.5117	78.83	-3.27
1	Algeria	3.763	3.801	35.2	14.6	40952312.0	1.697	2.7042	76.96	-0.89
2	Angola	4.327	4.371	92.9	27.7	29310737.0	3.523	6.1609	60.20	0.22
5	Argentina	3.341	3.375	11.0	4.8	44292731.0	0.912	2.2640	77.26	-0.10
6	Armenia	3.096	3.127	40.9	1.2	3045086.0	-0.213	1.6413	74.86	-5.66

Methods

We first merged the two data sets, handled any missing values, and familiarized ourselves with the variables to identify an efficient approach to achieving our aims. Our first approach to analyzing the data was to sort the countries based on the share of their population that could not afford a healthy diet and look at the tail and head of the data set to identify which countries experience the highest and lowest levels of food insecurity. Our next approach was to visualize the relationship between the affordability of a healthy diet and the cost of a healthy diet by making a scatterplot of the percentage of a population that couldn't afford a healthy diet v.s the cost of a healthy diet for all the countries in our data set. A regression line was added to the scatterplot to make identifying a relationship easier. Another approach was to visualize the relationship between food insecurity and migration and fertility rates, as well as life expectancy. We used the same approach as our last method and made three scatterplots with regression lines of the percentage of a population that could not afford a healthy diet against the countries' net migration rate, total fertility rate, and life expectancy. We then made a correlation heatmap to reaffirm these relationships and provide a compact and convenient way of seeing the correlations of each variable. Lastly, we conducted a regression analysis to predict a country's percentage of the population that cannot afford a healthy diet and plotted the Fitted Model Over a Scatter Plot of Country Attributes to see the relationship between the affordability of a healthy diet and growth rate.

Results

Tables 3 and 4, shown below, are tables of the head and tail (respectively) of our data set sorted by the share of the population that could not afford a healthy diet. The countries with the highest rates of food insecurity were Burundi, Liberia, Madagascar, and the Democratic Republic of the Congo. The countries with the lowest rates of food insecurity were Switzerland, Iceland, Azerbaijan, and the United Arab Emirates.

Table 3. Countries with the lowest percentage of their population unable to afford a healthy diet

Country Name	CoHD	CoHD_pov	CoHD_headcount	CoHD_unafford_n	Population	Annual Growth Rate %	Total Fertility Rate	Life Expectancy at Birth, Both Sexes	Net Migration Rate
Switzerland	2.523	2.548	0.0	0.0	8231519.0	0.683	1.5567	82.58	4.68
Iceland	2.213	2.235	0.0	0.0	339755.0	1.127	2.0000	83.08	3.96
Azerbaijan	2.348	2.372	0.0	0.0	9961850.0	0.874	1.8940	72.76	0.00
United Arab Emirates	2.755	2.783	0.0	0.0	9543192.0	1.071	1.6906	78.59	1.10
Slovenia	2.798	2.826	0.1	0.0	2101369.0	0.039	1.5765	81.05	0.76

Table 4. Countries with the highest percentage of their population unable to afford a healthy diet

Country Name	CoHD	CoHD_pov	CoHD_headcount	CoHD_unafford_n	Population	Annual Growth Rate %	Total Fertility Rate	Life Expectancy at Birth, Both Sexes	Net Migration Rate
Malawi	2.724	2.752	95.5	16.9	18365567.0	2.643	4.0220	70.97	-0.29
Congo, Dem. Rep.	2.921	2.951	96.4	78.5	92453955.0	3.255	5.9620	59.54	-0.58
Madagascar	2.987	3.017	97.1	24.8	25009416.0	2.494	4.0320	66.60	0.00
Liberia	4.018	4.059	97.1	4.6	4688786.0	2.498	5.0560	63.28	-5.65
Burundi	2.988	3.018	97.5	10.6	10951667.0	2.342	5.3811	65.41	-6.61

Figure 1, shown below, is a scatterplot that shows the relationship between the affordability and cost of a healthy diet in the countries. From it, we can see a weak positive correlation between the percentage of the countries' population unable to afford a healthy diet and the cost of a healthy diet. This indicates that countries, where the cost of a healthy diet is higher, may be associated with higher rates of food insecurity.

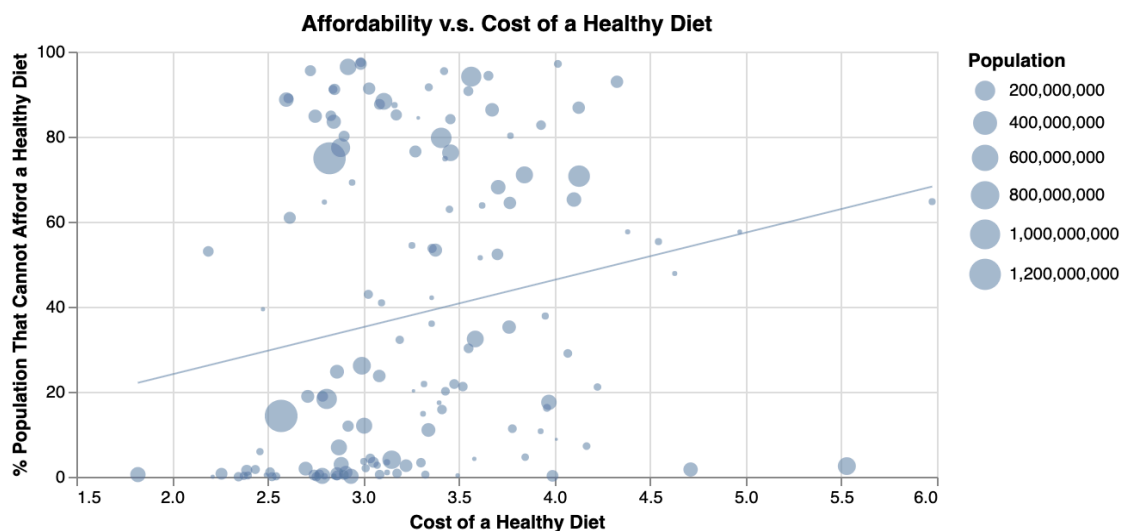


Figure 1. Linear regression relation between affordability & cost of a healthy diet

Figure 2, shown below, is a scatterplot that shows the relationship between the affordability of a healthy diet and migration rates in the countries. From the figure, we can see a weak negative correlation between the percentage of the population unable to afford a healthy diet and the net migration rates of the countries. This indicates that countries with higher rates of food insecurity may be associated with higher rates of emigration and lower rates of immigration.

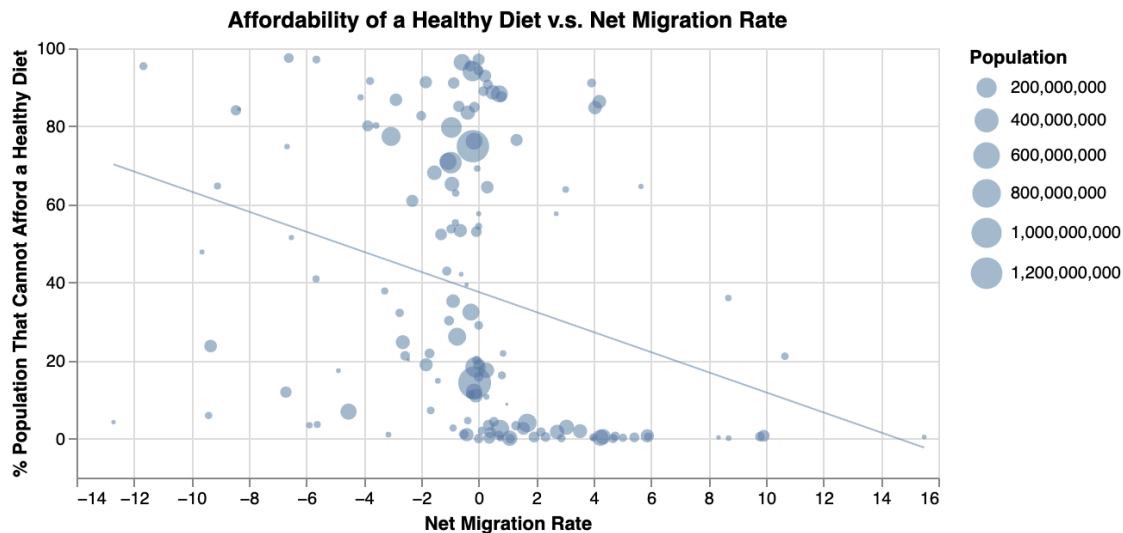


Figure 2. Linear regression relation between the affordability of a healthy diet & net migration rates

Figure 3, shown below, is a scatterplot that shows the relationship between the affordability of a healthy diet and fertility rates in the countries. From the figure, we can see a moderate positive correlation between the percentage of the population unable to afford a healthy diet and the fertility rates of the countries. This indicates that countries with higher rates of food insecurity are associated with higher rates of fertility.

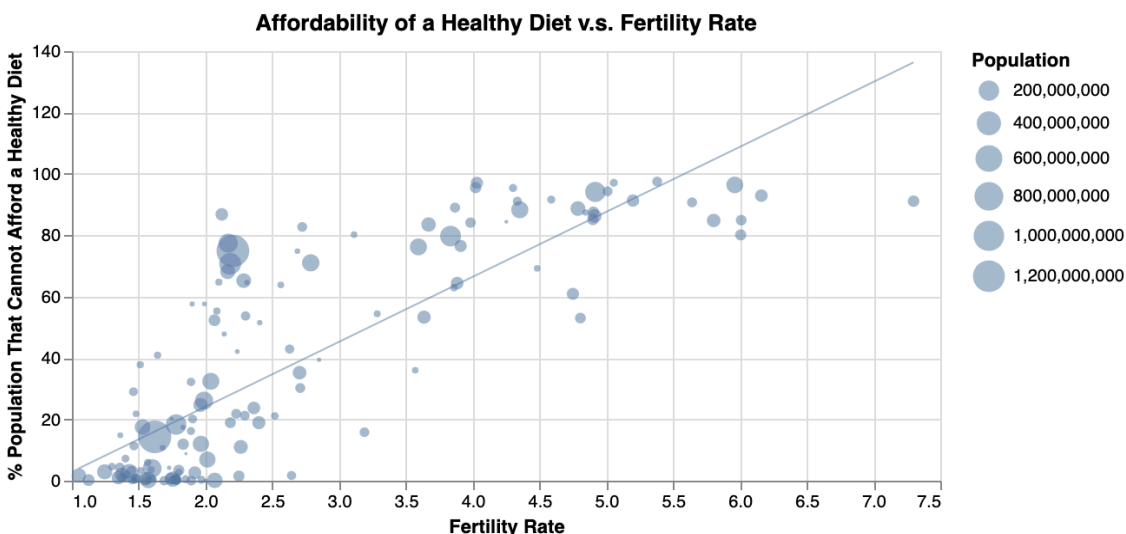


Figure 3. Linear regression relation between the affordability of a healthy diet & fertility rates

Figure 4, shown below, is a scatterplot that shows the relationship between the affordability of a healthy diet and life expectancy in the countries. From the figure, we can see a moderate negative correlation between the percentage of the population unable to afford a healthy diet and the fertility rates of the countries. This indicates that countries with higher rates of food insecurity are associated with lower life expectancy.

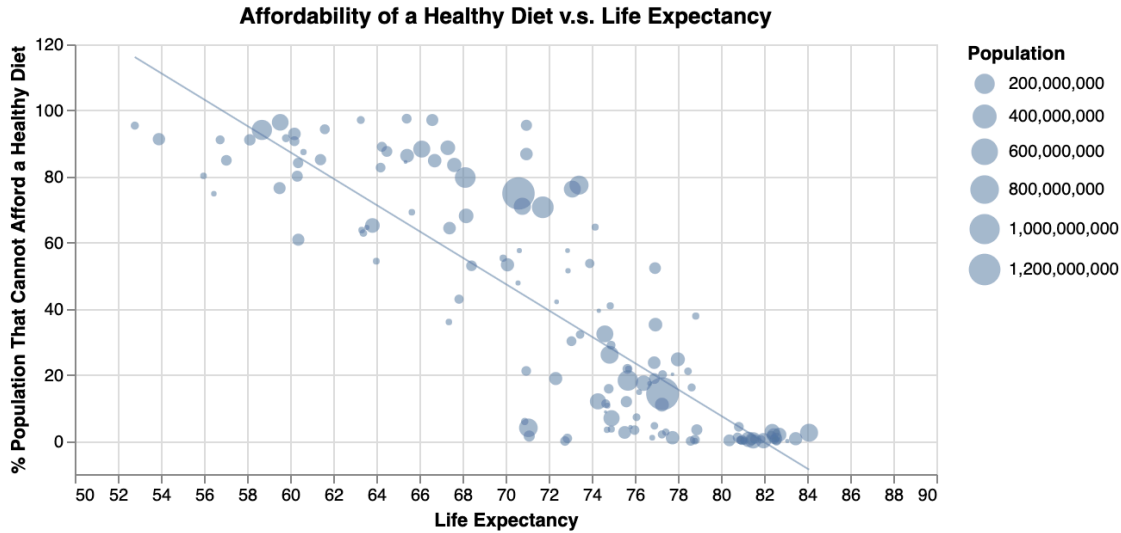


Figure 4. Linear regression relation between the affordability of a healthy diet & life expectancy

Figure 5, shown below, is a correlation heatmap. In it, we see that the proportion of a population that is unable to afford a healthy diet (CoHD_headcount) has a negative correlation with life expectancy (-0.86) and net migration rate (-0.31) and a positive correlation with fertility rate (0.82) and cost of a healthy diet (0.20). This is in line with our findings from the scatterplots above.

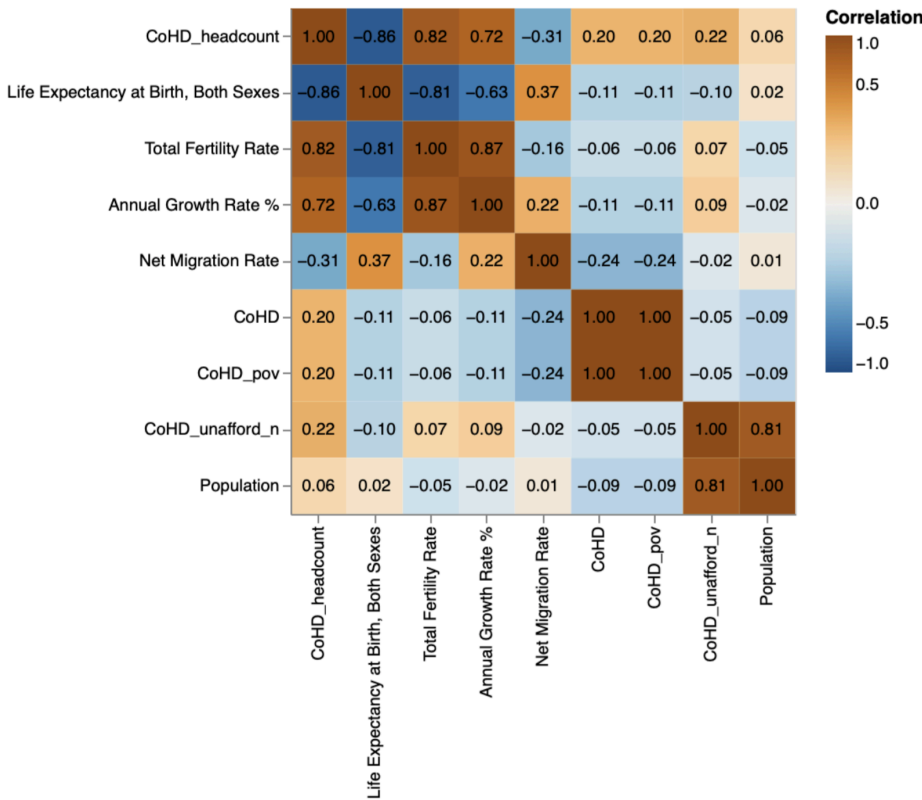


Figure 5. A heatmap of all correlation variables

Regression Analysis

To further our understanding of what attributes/features of a country are related to the proportion of their population that cannot afford a healthy diet, we employed regression analysis to predict our response variable, 'CoHD_headcount'. Figure 6 below shows the coefficient estimates and standard errors for each variable predicting 'CoHD_headcount'.

	coefficient estimate	standard error
CoHD	6.571156e+03	4.445828e+03
CoHD_pov	-6.492187e+03	4.401145e+03
CoHD_unafford_n	8.000420e-02	2.335659e-02
Population	-1.113229e-08	1.277365e-08
Annual Growth Rate %	1.341378e+01	3.195588e+00
Total Fertility Rate	8.313761e+00	2.425519e+00
Life Expectancy at Birth, Both Sexes	-6.299839e-01	8.904908e-02
Net Migration Rate	-1.994583e+00	4.529656e-01

Figure 6. Variable coefficient estimates and standard errors for predicting 'CoHD_headcount'

It appears that the cost of a healthy diet and the ratio of the cost of a healthy diet to the poverty line were the most important attributes in determining the percentage of the population that cannot afford a healthy diet. Interestingly, annual growth rate percentage and total fertility were important factors as well. Furthermore, our model did a very good job of prediction as it has an R-squared value of about 0.85.

Figure 7 below is a plot of the fitted model over a scatter plot showing the relationship between the affordability of a healthy diet and various country attributes.

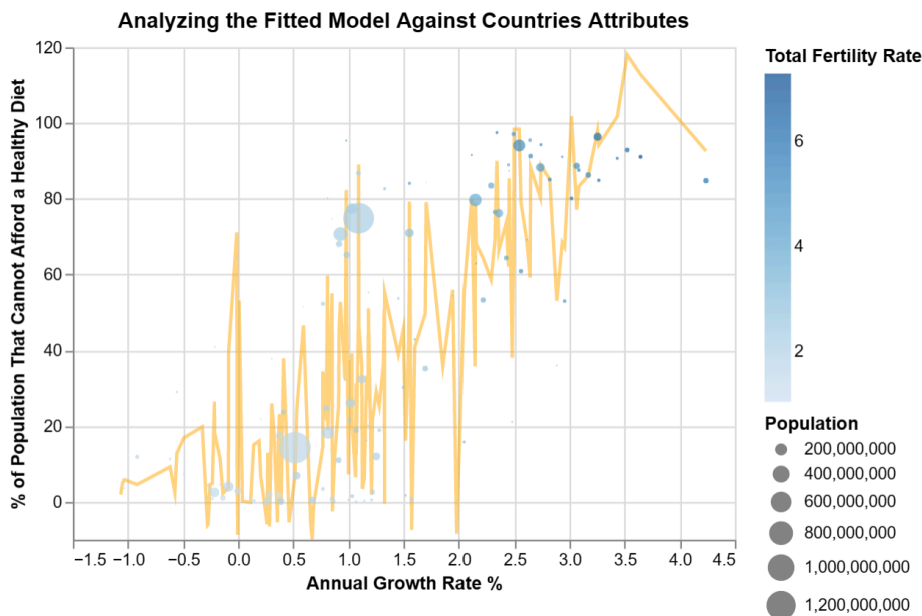


Figure 7. Plotting the Fitted Model Over a Scatter Plot of Country Attributes

From figure 7, we see that countries with higher populations and with higher annual growth rates tend to have higher rates of unaffordability of a healthy diet. Also, higher rates of fertility are shown for countries with a higher annual growth rate.

Discussion

Our analysis revealed Burundi, Liberia, Madagascar, and the Democratic Republic of the Congo as having the highest share of their population unable to afford a healthy diet and Switzerland, Iceland, Azerbaijan, and the United Arab Emirates as having the lowest share. It also highlighted a positive (albeit weak) correlation between the cost of a healthy diet and the percentage of a country's population that was unable to afford a healthy diet. This indicates that countries, where the cost of a healthy diet is higher, may be associated with higher rates of food insecurity. This is particularly interesting because though poverty rates seem to be the widely accepted factor driving food insecurity around the world, that may only be part of the bigger picture. Our analysis leads us to believe that high food prices are a part of that bigger picture. Our belief was validated again through regression analysis where we saw the cost of a healthy diet and the ratio of the cost of a healthy diet to the poverty line were the most important factors (of the variables in our data set) for predicting a country's percentage of the population that cannot afford a healthy diet. Our analysis also revealed that countries with a higher percentage of their population unable to afford a healthy diet also had higher fertility rates and annual growth rates and lower immigration rates and life expectancy. We speculate that the lower immigration rates and life expectancy are due to people not wanting to move to countries with decreased food affordability and that the lowered affordability is likely contributing to higher rates of starvation, which in turn brings down life expectancy.

It is important to note that our interpretations come with a few caveats. Further exploration is needed since the data we used was strictly from the year 2017, so seeing if these trends are consistent in other years is necessary to determine the validity of these relationships. Also, the data was not inclusive of all the world's countries, but just the countries that data was available for, and thus our findings may not hold true for every country. One topic we would have liked to explore was the affordability of healthy food with other factors such as infant mortality rates to see if there is a relationship. Also, it would have been interesting to categorize the countries based on region to see if the attributes we analyzed have any identifiable regional trends.