The Co-evolution of Crop, Dietary, and Flavor diversity in the U.S

TEAM NAME: Wireless Octopus

TEAM MEMBERS: Tung-Lin Liu, John Richardson, Thomas Hanlon and Harrison Thompson

DATE: 2021-12-16

Abstract

The study investigates the relationships between crop, dietary and flavor diversity using U.S. supply (production and trade) and dietary data over time. The data suggests the decline of crop diversity in the last 20 years co-occurs with the decline of dietary and flavor diversity. Unlike low- and middle-income countries, income is a weak predictor of dietary diversity in the U.S. Alternatively, the model in this study found that flavor diversity follows the same S-curve relationship expected between income and dietary diversity in the Economics literature, therefore, provides another means to understand the complex process between food environment and human diets.

Introduction

There has been increasing attention to studying planetary and human health together. This study uses longitudinal supply (production and trade) and dietary data to investigate the relationship between crop and agricultural diversity in the U.S. over time. Crop diversity has a small but positive effect on dietary diversity through subsistence and income-generating pathways in low- and middle-income countries. The relationships are less understood in high-income countries.

Because food consists of a small portion of the budget in such countries, the increase of dietary diversity is most likely not from necessities (e.g., starch stables) but the luxurious goods (see Figure 1). These goods are often produced with intensive inputs (e.g., animal proteins) or imported (e.g., coffee) and do not contribute to domestic crop diversity. Therefore, to investigate the relationship between crop and dietary diversity, we need to example both production and trade data. 23 percent of global food production are traded between countries (D'Odorico et al., 2014)

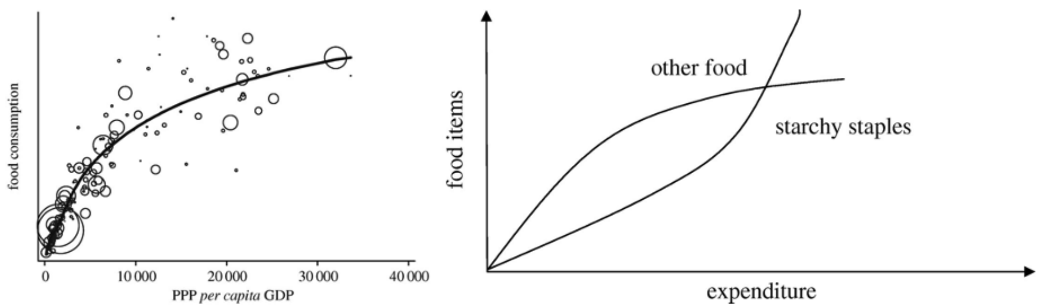

Figure 1 Engel curve of nations (left). Engel's vs Bennett's law (right) (Cirera & Masset, 2010)

Other than the level of economic development, food production and preferences are also governed by many other feedbacks such as labor availability and food quality. (See Hammond & Dubé, 2012for full discussion). Here, we are interested in how existing dietary patterns mediate the relationships between crop and dietary diversity (Béné et al., 2020). It is interesting that Ahn et al. found that Western cuisines tend to combine ingredients that share similar flavor compounds, whereas East Asian cuisines do not (2011). As these economies are highly developed, the effect on dietary patterns is likely to be at least partially independent from income. Flavor could be a key to understanding these differences.

Therefore, we hypothesize that food choice and its resulting diversity are regulated by the co-evolution of intensifying food production and exchanges across borders and cultural eating patterns. We specifically ask the following questions:

1. How have food diversity and food production and trade (suppy) changed together over time in the U.S.?
2. What are the trends of flavor diversity in the U.S?

From Engel's Law, we expect to find non-linear relationships between dietary and supply diversity in the US, and the income level of the individuals moderates the relationships. Also, we expect to observe a decline in flavor diversity over time.

Materials and Methods

We obtain U.S. food supply data from ***Global Expanded Nutrient Supply*** (GENuS) Model (Smith et al., 2016). GENuS is built upon the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT) Food Balance Sheet. GENuS disaggregates FAOSTAT commodity categories and provides better estimates of individual food components.

Currently, FAOSTAT contains production and trade data of 186 countries from 1961 to 2019. Although GENuS only has a coverage till 2011, it provides the supply of 225 commodities that are validated against historical USDA data. We chose GENuS over FAOSTAT or USDA Food Availability (Per Capita) Data System (FADS) because it provides more accurate and more categories of food the diversity metric aims to measure.

National Health and Nutrition Examination Survey (NHANES) provides cross-sectional diets and health behaviors data of ~5,000 individuals since 1971 and continuously since 1999. The US dietary data is retrieved from What We Eat in America (WWEIA) module began in 1999. WWEIA uses 24-hour dietary recall to capture portions of dishes an individual ate in a day.

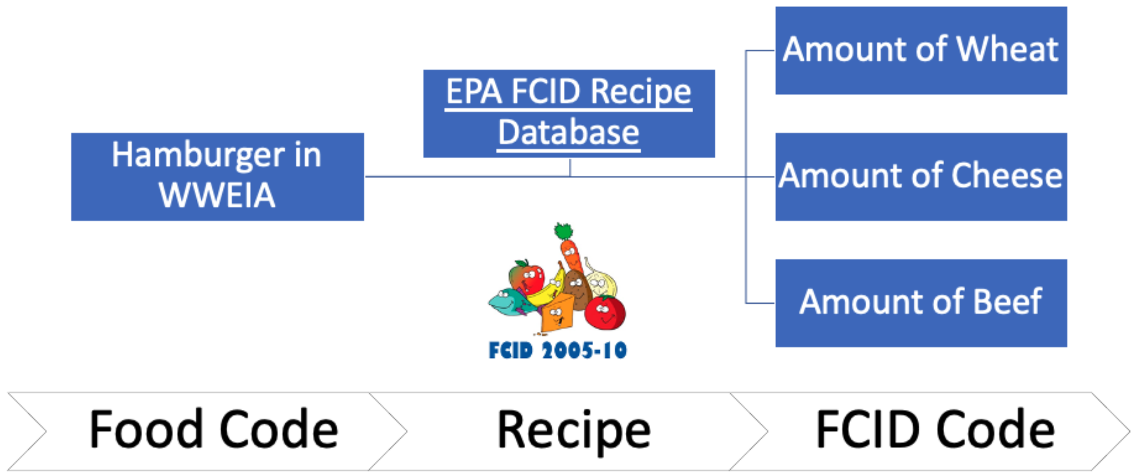

Figure 2 Converting food items in WWEIA into crop commodities using FCID database

NHANES provides conversion tables to convert the information from 24-hour recall to micro- and macronutrients. Yet, this does not fit our goal to capture dietary diversity at the ingredient level. Therefore, we follow the protocol used in Conrad et al. to get the amounts of commodities each person consumes using a separate database (2018).

Environmental Protection Agency ***Food Commodity Intake Database (FCID)*** was originally designed between 2005 and 2010 to measure the exposure of pesticide from food commodities. The database can translate 7,000+ dishes used in WWEIA to 500+ commodities (see figure 2). However, the database has stopped updating since 2010, and more dishes are being added in WWEIA every year. Therefore, there will be food that consumed in the survey post 2010 that are not converted.

At last, favor molecules data will be retrieved from FlavorDB using a scraper (Garg et al., 2018). FlavorDB contains 25,595 flavor molecules, and 2,254 of them are associated with 936 natural ingredients.

Table 1. Comparison of four databases.

	GENuS	NHANES-WWEIA	FCID	FlavorDB
Time	1961-2011	1999-2017; revision in 2002	2005-2010	NA

# of Ingredients	225	NA	500+	936

To understand the relationship between crop and dietary diversity, we first compute the count-based dietary diversity (C) using Food and Agriculture Organization (FAO) protocol. The commodities derived from 24-hour recall in WWEIA are categorized into ten categories: cereal grains, white tubers, and root foods, dark leafy greens, vitamin A-rich vegetables/tubers, vitamin A-rich fruits, other fruits and vegetables, meat and fish foods, eggs, legumes/nuts/seeds, and milk and milk products. The dietary diversity score ranges from 0 to 10. If the individuals consume all ten categories of food on the given day, they receive scores of 10 and vice versa.

Count based (C)

$$C = (n)$$

We use entropy-based diversity to measure crop diversity. Let n be the number of food commodities from 1, 2, 3, …, to w_i, and the n_i commodity has the weight w_i. The diversity metric can then be computed using Shannon-Wiener Entropy (SE), popular in Ecology. Oftentimes, the metric is normalized between 0 and 1 by dividing log(n). When SE is 0, it indicates the lowest crop diversity (no uncertainty), and when SE is 1, it indicates the highest crop diversity (high uncertainty).

Shannon-Wiener Entropy (SE)

$$SE = - \sum_{i=1}^{n} w_i \log(w_i)$$

However, when the n is large, the count- and entropy-based diversity tend to overestimate the diversity (e.g., spices of very little use). Therefore, we also use Berry-Simpson Index (BI), more common in Economics, to calculate both crop and dietary diversity. BI is less guided by probability because w_i is squared (Jekanowski & Binkley, 2000). The BI ranges from 0 to 1-1/n. BI of 0 indicates the individuals only have one food item in a day or countries produce only one crop. BI of 1-1/n suggested an equal share amount of food commodities possible. BI also gives a unified index to compare both two types of diversity.

Berry-Simpson Index (BI)

$$BI = 1 - \sum_{i=1}^{n} w_i^2$$

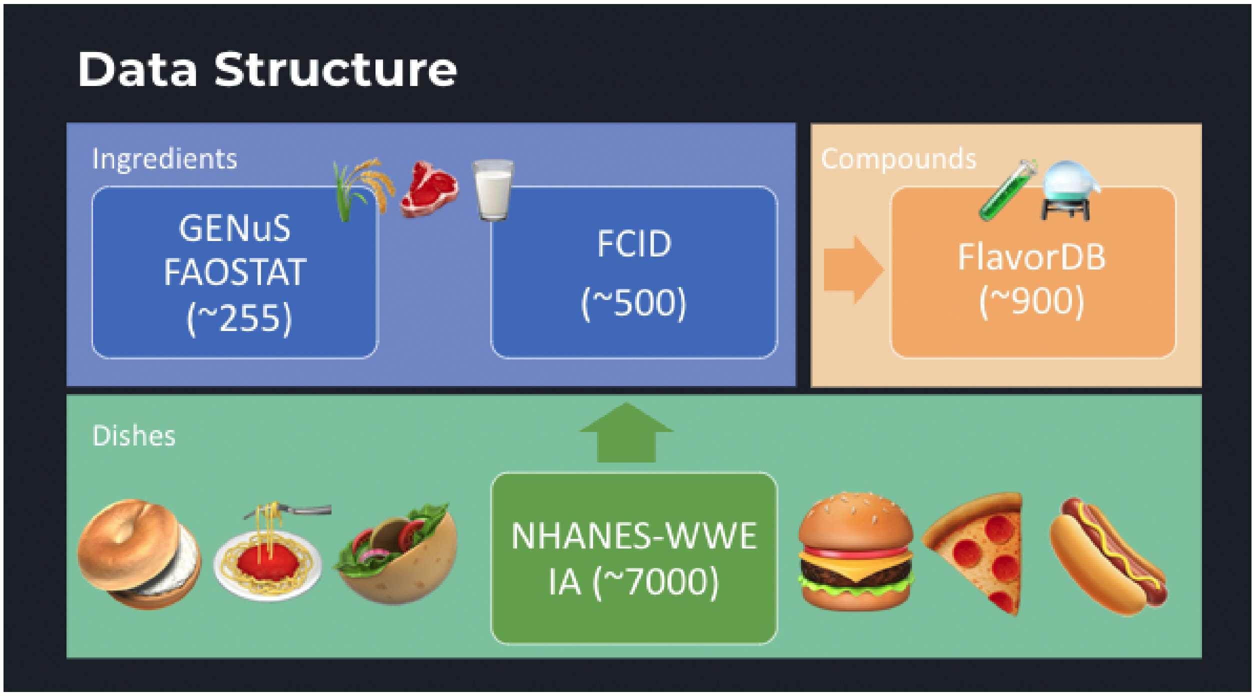

Figure 3: Datasets used and perspective of conversion flow

The data was analyzed using python. The data organization is described in Figure 3. Each dataset (GENuS, FCID, NHANES, FlavorDB) is separately loaded, combined and cleaned. Look up tables are created to link the codes used in respective databases to their names. After the initial cleaning, the portion of dishes eaten by a person in a given day in NHANES-WWEIA is converted into commodities weight using the FCID conversion table. These commodities amounts then are matched to the food in FlavorDB. For each flavor compound, we sum all the weights of the commodities containing such molecules.

US Crop diversity (CD) is estimated using Berry Index from GENuS US food supply (production and import) of each year (1961-2011). Individual Dietary Diversity (DD) is estimated using UNFAO IDDS and Shannon Entropy from the FCID commodity weights of each individual in NHANES-WWEIA. Flavor diversity (DD) is then estimated using the amount of food each molecule shares using Shannon Entropy. The first metric is then linked with the last two metrics using the crop diversity six years prior to the survey. Demographics data then is linked from other NHANES modules using the unique IDs given.

To study the changes of crop, dietary and flavor diversity over time. Independent t-tests are used to compare the average of these metrics across years. Moreover, to model the non-linear relationships between diversity and income and other control important variables (e.g., age, year of the survey), cubic regression with knots are selected.

Results

The trend of Shannon entropy and Berry Index in figure 4 shows that, US food supply has become more diverse over time over the last 50 years. However, both metrics also indicate that there has been a recent decline in crop diversity since 2000. As the figure suggests, the decline of crop diversity is a result of continuous increased production in certain foods and decline of the others, for example 747 gram of fluid milk was available per person per day in 2011 whereas only 675 gram was available in 1971. In contrast, at the highest, US beer supply reached 276 gram per gram in 1981 and since then the supply steadily declined over years. In 2011, there was only 220 gram available per capita.

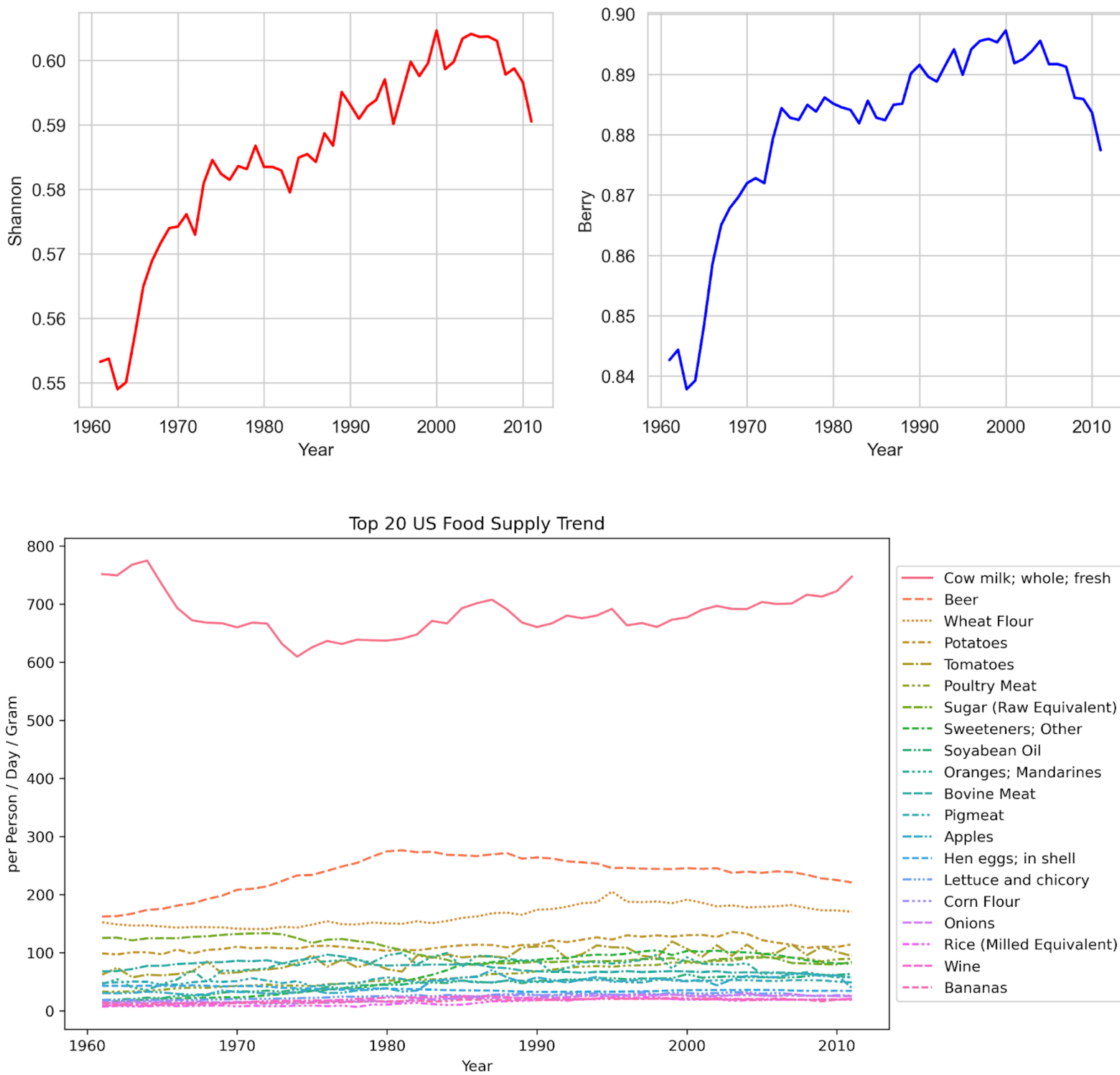

Figure 4: (upper) Top 20 individual supplied foods summed total shannon diet diversity for

each year from 1961-2011 (lower) United States Top 20 individual supplied foods (per person / day / gram) from 1961- 2011

The food supply and its resulting crop diversity is also reflected in the US diet. For the dietary data of the last ~20 years, over 80 percent of FCID commodities people consumed are concentrated in several crops – milk, water (33%) and wheat, flour (22%), corn field syrup (10%), milk, nonfat solids (7%), soybean, oil (7%), milk, fat (4%). These numbers are taken from the median of each 500 FCID commodities converted from 7000+ food items people ate between 2001 to 2018). Although people may consume different dishes, after conversion, the ingredients of such foods tend to come from major crops that are also heavily present in the US food supply.

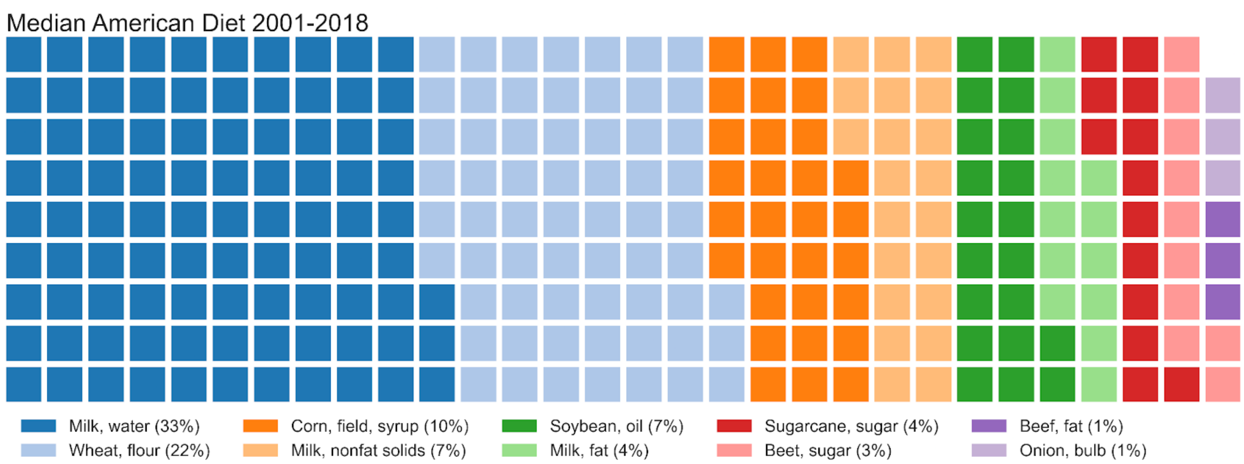

Figure 5: Waffle Chart of top 10 foods and their proportions in the Median American Diet from 2001-2018

Table 2 shows the summary statistics of the combined data used in the analysis. After removing the respondents who are 20 years or younger, we obtained 39,755 samples over 9 waves of NHANES (2001-2018). For those of whom have zero Shannon dietary diversity or UN-FAO IDDS-10 score, it indicates the person either had only one food consisting of one ingredient or the person has not eaten any food on the given day. The most significant source of missing values (n=3,217) are from the income poverty ratio. The income variable is also capped at 5.0 and does not account for anyone earned five times more than the county poverty line. This presents a major challenge in our regression analysis.

The correlation heatmap in figure 6 provides correlation coefficients of the key variables. As expected, Shannon dietary diversity is moderately correlated (r=0.59) with IDDS-10 and weakly correlated with crop diversity of six year prior (r=0.3). The negative correlation between dietary (r=-.3) and crop diversity (r=-.82) and the year of survey suggest the decline of crop and dietary diversity could occur together. At last, conforming with the nonlinear expectation, income has a negligible correlation (r=0.1) to dietary diversity. We also found an unexpected moderate correlation (r=0.57) between dietary and flavor diversity.

Table 2. Summary Statistics of key variables

	DD	IDDS-10	AGE	INCOME	FD	CD
count	39742	39755	39755	36538	39755	39755
mean	0.39	6.65	49.62	2.53	0.80	0.89
std	0.07	1.34	18.05	1.62	0.05	0.00
min	0.00	0.00	20.00	0.00	0.19	0.88
50%	0.40	7.00	49.00	2.13	0.81	0.89
max	0.59	9.00	85.00	5.00	0.73	0.90

DD: Dietary Diversity (Shannon); IDDS-10: UN-FAO Individual Dietary Diversity Score, INCOME: Income poverty ratio; FD: Flavor Diversity (Shannon); CD: Crop Diversity (Berry)

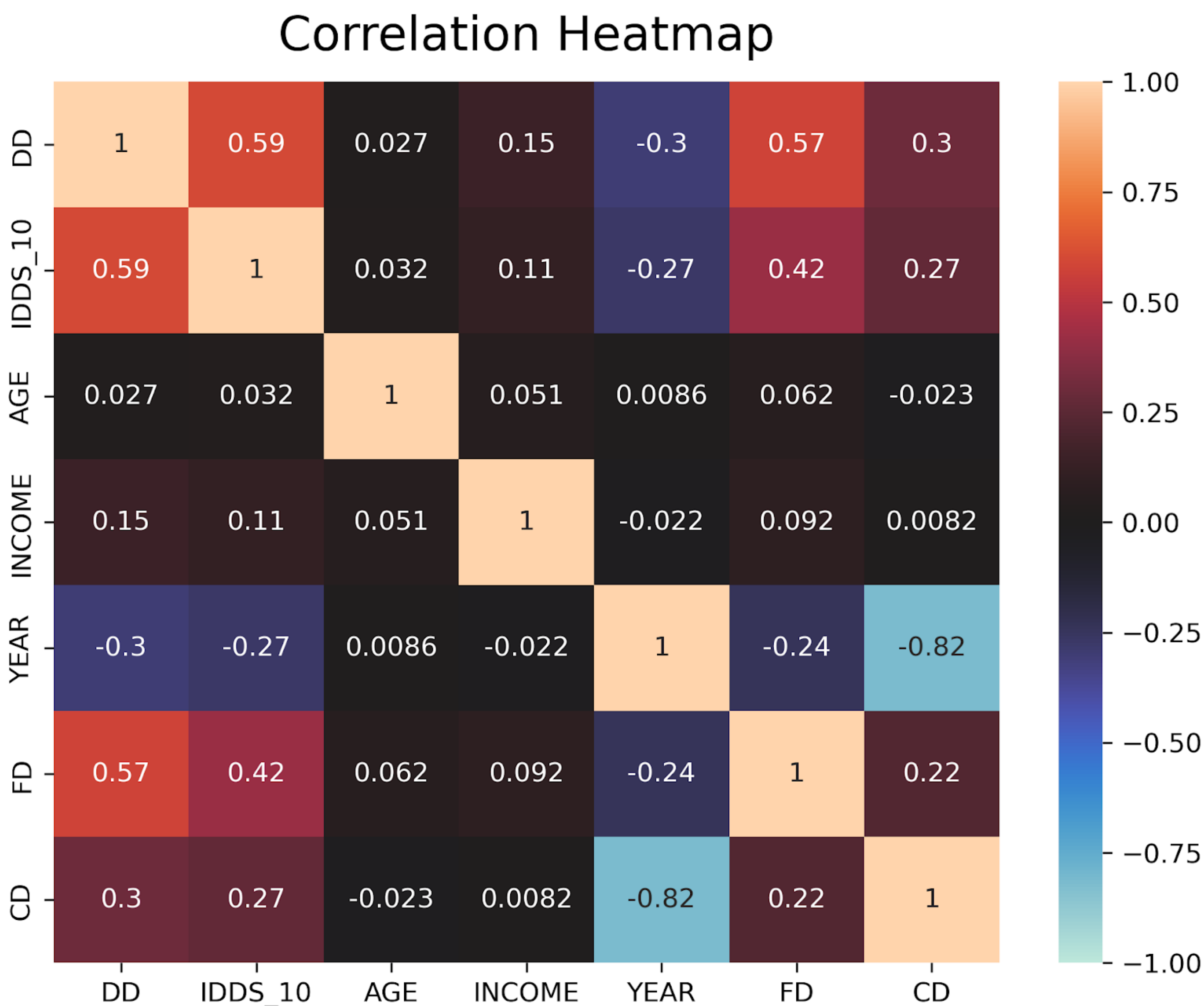

Figure 6 Correlation of key variables in the analysis

Figure 7 shows as the time progresses in the survey circle, individual dietary diversity in the sample declines over time. The peak of the distribution shifts toward low dietary diversity every two year. The changes also occur in the shape of the distribution. Every two years, the distribution also grows flatter. Colors in figure 7 highlight how the changes in dietary diversity with respect to crop diversity. Although with some fluctuations in crop diversity every two years, the trend is on the decrease.

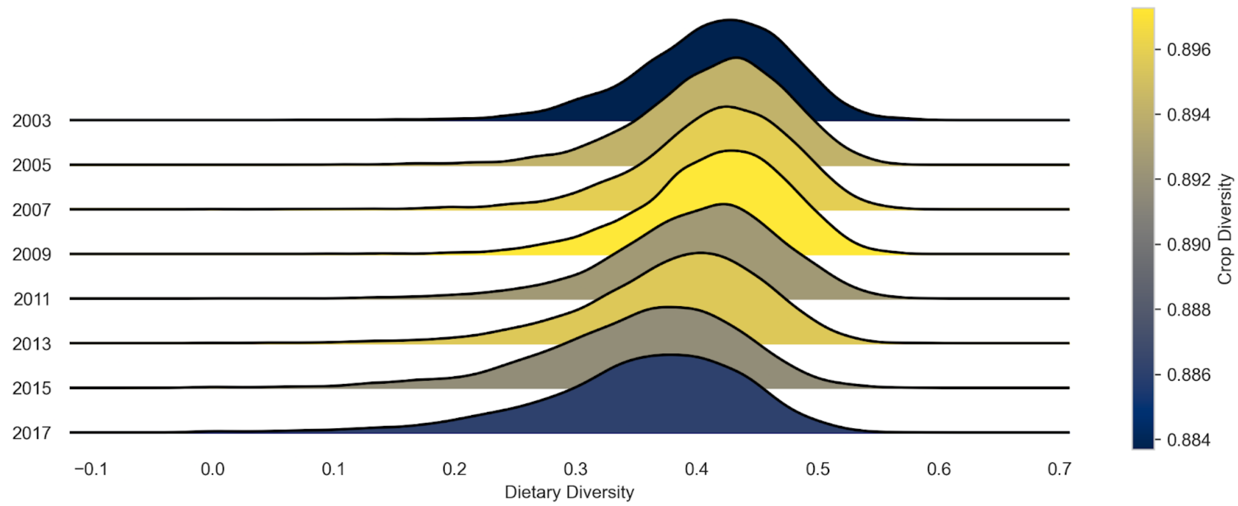

Figure 7: Decline of dietary diversity from 2003-2017, crop diversity increase and decline from 2003-2017

Similarly, the flavor diversity converted from dietary data also reveals the decline over years and the fattening of the tails in figure 8. Since the FCID conversion table used in the study stopped updating after 2010, some new food (e.g., plant-based meat) added subsequently in the NHANES-WWEIA food codes are not factored in our dietary and flavor diversity calculations. To test the effects of decline is independent from the conversion table, three independent t-tests are carried out.

The test results indicate the decline of flavor diversity occurred prior to 2010. There are significant differences in the mean of favor diversity between 2003 and 2005 (t-statistics = .1451 , p<.0001; `plotting.py`, line 61) and between 2003 and 2009 (t-statistics = .1436 , p<.0001; `plotting.py`, line 61). Although the effect of the outdated conversion table could not be ruled out, the difference between 2003 and 2017 is also highly significant (t-statistics =.3452; p<.0001; `plotting.py`, line 61). Therefore, the decline observed in both dietary and flavor diversity is likely not an artifact due to conversion.

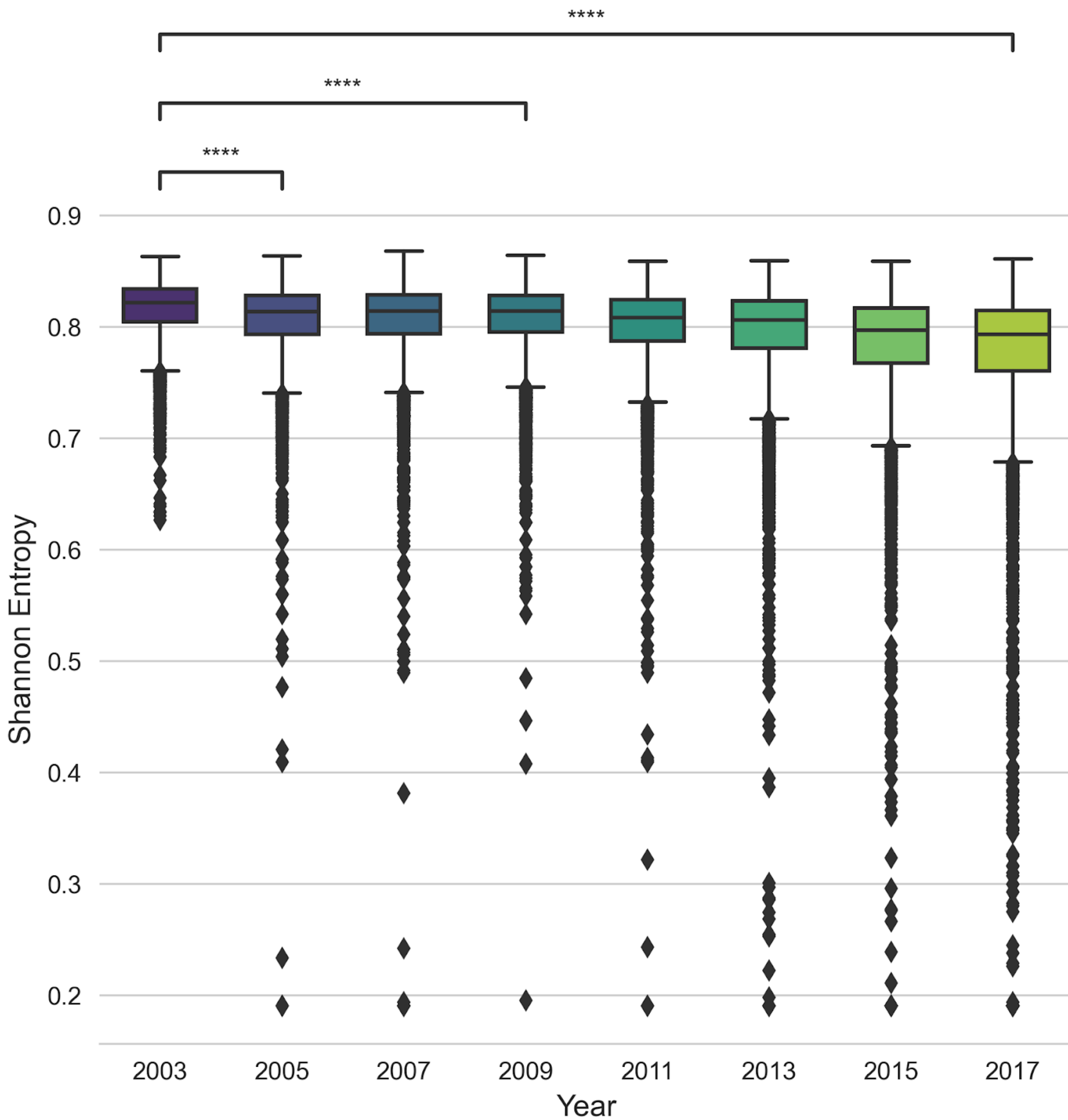

Figure 8: Shannon entropy value from 2003-2017 shown in boxplots

At last, the cubic regression model with 3 knots model demonstrated that income has a very weak effect on dietary diversity (Model 1). The effect further diminishes as the income poverty ratio continues to rise. This curve is different from the decreasing return expected in Engel curves. The model also has poor accuracy on predicting 20 percent of unseen data (RMSE = 0.0692, R2: 0.1119; `analysis.py`, lines 94-96). The fitted line in figure 9 also shows the model is not a good candidate.

Table 3.

	Model 1	Model 2
Intercept	2.6779***	-1.5974***
bs(INCOME, df=3, degree=3)[0]	0.0260***	
bs(INCOME, df=3, degree=3)[1]	0.0263***	
bs(INCOME, df=3, degree=3)[2]	0.0379***	
bs(FD, df=3, degree=3)[0]		0.4763***
bs(FD, df=3, degree=3)[1]		0.0201*
bs(FD, df=3, degree=3)[2]		0.4799***
YEAR	-0.0025***	-0.0003*
CD	3.0119***	2.4848***
INCOME		0.0038***
Adj. R-squared:	0.119	0.425

DD: Dietary Diversity (Shannon); INCOME: Income poverty ratio; FD: Flavor Diversity (Shannon); CD: Crop Diversity (Berry)

In contrast, when dietary diversity is regressed on flavor diversity using cubic regression with 3 knots (Model 2), the s-curve relationship expected in Bennets law's is observed (See Figure 9). The model has a much greater accuracy in predicting unseen data (RMSE = 0.0565, R2: 0.4108; `analysis.py`, lines 106-108). Again, the coefficients in table 3 also show that flavor diversity (FD) has a better polynomial fit than income. Overall, crop diversity (CD) has significant positive effects in both models.

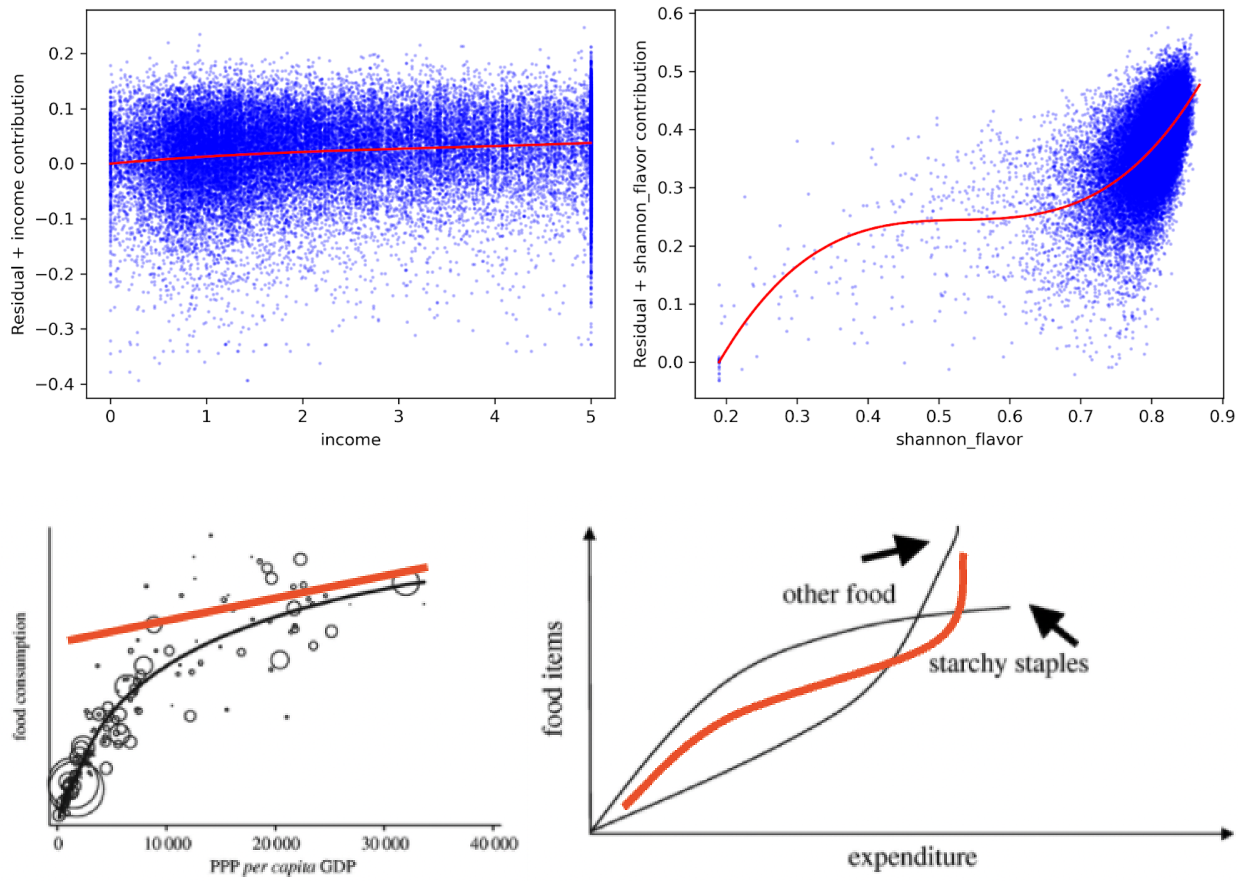

Figure 9: Polynomial Spline Regression for income + residuals vs. shannon diet diversity score and shannon flavor diversity + residuals vs. shannon diet diversity

Discussion and Conclusions

US supply and dietary data reveal the co-occurence of crop, dietary and flavor diversity declines over the last 20 years. It highlights the importance of how biodiversity of a nation's food supply can have an impact on a person's diet and possible health (de Oliveira Otto et al., 2018) and the potential of how dietary habits could impact our ability to stay within the limits of planetary boundaries and achieve the future of secured food supply at the same time (Willett et al., 2019).

The study advances the understanding of relationships between crop diversity, dietary diversity in high-income countries. At least for the US, income is not a good predictor of dietary diversity. The GDP per capita of the US is $63,543 in 2020 (Source: World Bank). It is likely the US data captures the flat tail of Engel Curve relations between income and food consumption. However, the dietary data also does show the increase in luxurious consumption as expected in Bennet's Law. As the diversity calculation did not differentiate these qualitative differences and the amount of certain luxurious food is consumed in small quantities.

One of the possible reasons for such behavior is the income is capped at '5' in NHANES dataset. Five denotes the highest income poverty ratio recorded in the survey, for example, if a county's poverty line is $20,000 any person with an income of $100,000 is not accurately reflected in the analysis. Given that the United States is a wealthy country, it is likely that given an uncapped income poverty ratio, the model could demonstrate the nonlinearity relating to dietary diversity and income.

However, our study opens up the possibility to use flavor diversity as an alternative to income to predict dietary diversity in high income countries. Flavor consists of a significant portion of the taste of food (Shepherd, 2011). Studies found cuisines from similar levels of economic development have different flavor combinations and some of these country hosts seem to have higher crop diversity (Mariani et al., 2021). Similarly, (Watson & Cooper, 2021) use the rising of specialty beer like IPA as an example to show how flavor could combat the trend of homogenous global food supply (Khoury et al., 2014) .

At last, with the limitations of each dataset imposed, the results could be further enhanced by using the latest FAOSTAT data (the latest one is 2019) instead of GENuS, a more comprehensive "recipe" database that convert all the food items recorded in NHANES into FCID commodities and methods that can model the relationship of diversity at the commodity level.

References

Béné, C., Fanzo, J., Haddad, L., Hawkes, C., Caron, P., Vermeulen, S., Herrero, M., & Oosterveer, P. (2020). Five priorities to operationalize the EAT–Lancet Commission report. *Nature Food*, *1*(8), 457–459. https://doi.org/10.1038/s43016-020-0136-4

Cirera, X., & Masset, E. (2010). Income distribution trends and future food demand. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1554), 2821–2834. https://doi.org/10.1098/rstb.2010.0164

Conrad, Z., Niles, M. T., Neher, D. A., Roy, E. D., Tichenor, N. E., & Jahns, L. (2018). Relationship between food waste, diet quality, and environmental sustainability. *PLOS ONE*, *13*(4), e0195405. https://doi.org/10.1371/journal.pone.0195405

de Oliveira Otto, M. C., Anderson, C. A. M., Dearborn, J. L., Ferranti, E. P., Mozaffarian, D., Rao, G., Wylie-Rosett, J., Lichtenstein, A. H., & null, null. (2018). Dietary Diversity: Implications for Obesity Prevention in Adult Populations: A Science Advisory From the American Heart Association. *Circulation*, *138*(11), e160–e168. https://doi.org/10.1161/CIR.0000000000000595

D'Odorico, P., Carr, J. A., Laio, F., Ridolfi, L., & Vandoni, S. (2014). Feeding humanity through global food trade. *Earth's Future*, *2*(9), 458–469. https://doi.org/10.1002/2014EF000250

Garg, N., Sethupathy, A., Tuwani, R., NK, R., Dokania, S., Iyer, A., Gupta, A., Agrawal, S., Singh, N., Shukla, S., Kathuria, K., Badhwar, R., Kanji, R., Jain, A., Kaur, A., Nagpal, R., & Bagler, G. (2018). FlavorDB: A database of flavor molecules. *Nucleic Acids Research*, *46*(D1), D1210–D1216. https://doi.org/10.1093/nar/gkx957

Hammond, R. A., & Dubé, L. (2012). A systems science perspective and transdisciplinary models for food and nutrition security. *Proceedings of the National Academy of Sciences*, *109*(31), 12356–12363. https://doi.org/10.1073/pnas.0913003109

Jekanowski, M. D., & Binkley, J. K. (2000). Food purchase diversity across U.S. markets. *Agribusiness*, *16*(4), 417–433. https://doi.org/10.1002/1520-6297(200023)16:4<417::AID-AGR3>3.0.CO;2-H

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., Rieseberg, L. H., & Struik, P. C. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences*, *111*(11), 4001–4006. https://doi.org/10.1073/pnas.1313490111

Mariani, R. O., Cadotte, M. W., Isaac, M. E., Vile, D., Violle, C., & Martin, A. R. (2021). National-scale changes in crop diversity through the Anthropocene. *Scientific Reports*, *11*(1), 20361. https://doi.org/10.1038/s41598-021-99728-5

Shepherd, G. M. (2011). *Neurogastronomy: How the Brain Creates Flavor and Why It Matters*. Columbia University Press. http://ebookcentral.proquest.com/lib/bu/detail.action?docID=909501

Smith, M. R., Micha, R., Golden, C. D., Mozaffarian, D., & Myers, S. S. (2016). Global Expanded Nutrient Supply (GENuS) Model: A New Method for Estimating the Global Dietary Supply of Nutrients. *PLOS ONE*, *11*(1), e0146976. https://doi.org/10.1371/journal.pone.0146976

Watson, D. L. B., & Cooper, C. M. (2021). Visceral geographic insight through a ‘source to senses’ approach to food flavour. *Progress in Human Geography*, *45*(1), 111–135. https://doi.org/10.1177/0309132519890913

Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J. A., Vries, W. D., Sibanda, L. M., … Murray, C. J. L. (2019). Food in the Anthropocene: The EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, *393*(10170), 447–492. https://doi.org/10.1016/S0140-6736(18)31788-4