The Food Access Module to the PSID

User Guide and Codebook

2018

Codebook Authors:

Sandra Tang, Ph.D. and Stephanie McCracken, M.P.H.

Population, Neurodevelopment, and Genetics Program
Survey Research Center, Institute for Social Research
The University of Michigan, Ann Arbor, MI

Acknowledgements:

The Clark R. Smith Family Foundation generously provided funding to support the development of this data module. Gratitude is extended to the research analysts who prepared the geospatial data: Stephanie McCracken and Yajing Zhou; the team of research assistants who helped validate the food establishment data: Lauren Ely, Weixuan He, Dana LaBuda, Nejra Malanovic, Megan Mitchell, Paige Porter, Swetha Reddi, Bethany Rookus, Aanya Salot, Kendall Sidnam, and Emily Yerington; and the individuals who provided instrumental advice and guidance at various stages of the project: Natalie Colabianchi, Pamela Davis-Kean, Linda Eggenberger, Noura Insolera, Kari Moore, Fabian Pfeffer, Nicholas Prieur, and Nicole Scholtz.

Suggested APA citations:

FAM User Guide:

Tang, S. & McCracken, S. (2018). *PSID Food Access Module user guide and codebook*. Retrieved date.

FAM Data:

Tang, S., McCracken, S. & Zhou, Y. (2018). *PSID Food Access Module: Release 1*. [Data file].

Suggested acknowledgement of the FAM Data:

Funding for the development of this data module was provided by the Clark R. Smith Family Foundation awarded to Principal Investigator, Sandra Tang, and Co-Investigators, Pamela Davis-Kean and Natalie Colabianchi.

Content Contact: Sandra Tang (sandtang@umich.edu)

Table of Contents

I. Overview

The Food Access Module (FAM) to the PSID is a data resource for researchers interested in examining how the physical food environment is associated with individual well-being in the United States using a nationally representative, longitudinal dataset. Created primarily from data drawn from the National Establishment Time-Series (NETS), a commercial database, FAM contains geographic indicators of various categories of food establishments across the United States from the years 2000 to 2014. These indicators, which can be linked to the residential location of PSID families at the block-group level, allow researchers to investigate how food establishment availability in the physical environment relates to the health and wellbeing of individuals and families at a population-level.

The **food establishment categories** available in this data module are:

- Club Stores
- Convenience Stores
- Ethnic Restaurants
- Fast Food Restaurants
- Fruit and Vegetable Markets
- Full Service Restaurants
- Grocery Stores
- Health Food Stores
- Meat and Fish Markets
- Specialty Food Stores
- Supercenters
- Supermarkets

The ***distance and density*** **indicators of geographic food availability** for each food establishment category in this data module are:

Euclidean (straight-line)	Road Network
0.5 mile	0.5 mile
1 mile	1 mile
5 miles	5 miles
10 miles	

In addition to data derived from the NETS database, this module includes data from the United States Department of Agriculture's (USDA) Food Access Research Atlas, which contains data from 2010 at the census-tract level, and the USDA's Food Environment Atlas, which contains data from 1999-2014 at the state-, county-, and regional-levels.

In total, there are 516 variables and 133,783 observations in FAM. The data are structured in long format so that each row contains information on one PSID family for one specific year.

Since this data module contains restricted geocode data, users should refer to PSID's website (https://simba.isr.umich.edu/restricted/RestrictedUse.aspx) for information on how to obtain access to the restricted data.

II. Procuring the Original Data

The majority of the variables created for this data module were generated from the **National Establishment Time-Series (NETS) database**, a longitudinal, commercial database created by Walls & Associates (2014) from Dun and Bradstreet archival data. The NETS database contains annual snapshots of existing establishments in the United States beginning from the year 1990. For each establishment, NETS provides information on company names, business geocodes (latitude/longitude), Standard Industrial Classification (SIC) and North American Industry Classification System (NAICS) codes, and the years a company was active (Walls & Associates, 2013).

We ordered commercial data on establishments with food-related Standard Industrial Classification (SIC) codes for the years 2000-2014 from Walls & Associates. For our order, we primarily used the 8-digit SIC codes to identify food-relevant establishments instead of the commonly used 6-digit SIC codes because SIC codes with more digits provide more detailed information on the industry of each business (Walls & Associates, 2013). In some cases, we used the 3- and 4-digit SIC codes when we wanted to include establishments under a more general category (e.g., Meat and Fish Markets, SIC3 = 542); 3- and 4-digit codes include all the 8-digit codes under that category. In total, 300 SIC codes were used to identify food-relevant establishments in the final data module. (See Appendix A for the SIC codes that we used to order establishment data from Walls & Associates for this data module.)

In addition to requesting establishment data via SIC codes, we requested data for some establishments by name. This was necessary because an establishment's SIC code can vary each year for a couple of reasons. One, the SIC code is self-selected by each establishment, and two, the SIC code is based on the company's primary line of business for that year at that specific location (https://siccode.com/en/pages/how-to-select-your-sic-code). Thus, if we relied only on food-related SIC codes, we would have missed stores that have a sizeable portion of their store space allocated to selling groceries, but that may not have self-identified with a food-related SIC code. In total, 6 of these types of establishments were requested by name (BJ's Wholesale Club, Costco, Kmart, Sams Club, Target, Walmart).

While we waited for the delivery of the NETS data, we piloted the data linking process by adding two publicly available data resources on food availability from the USDA, described below.

The **USDA's Food Access Research Atlas** contains data from 2010 at the census-tract level. The atlas provides information on low food accessibility, prevalence of low income households, vehicle access, group housing indicators, population sociodemographic and socioeconomic characteristics, and regional characteristics (ERS, 2017). Variables representing low food access or limited supermarket accessibility were measured at multiple Euclidean distances (0.50-, 1-, 10-, and 20- miles). These indicators were based on calculations made by the USDA after subdividing the United States into ½- kilometer grid cells. Distances were measured from the center of each of these cells to the closest supermarket (see Ver Ploeg et al., 2012 for more information about the methodology.) Additional information on the Food Access Research Atlas can be found on their website: https://www.ers.usda.gov/data-products/food-access-research-

atlas/

The **USDA Food Environment Atlas** contains data from 1999-2014 at state-, county-, and regional-levels. Some of the food environment data available in this atlas include store and restaurant availability, aggregate use of government assistance for food, local food providers (e.g., farmer's markets), and aggregate obesity levels in adults and children. This ERS resource represents a conglomerate of information that was derived from multiple sources, such as the U.S. Census, the Center for Disease Control, and the USDA's Agricultural Marketing Service (ERS, 2016). Additional information on the Food Environment Atlas can be found on their website: https://www.ers.usda.gov/data-products/food-environment-atlas/

The variables from these data sources that were added to FAM are listed in Section V of the user guide.

III. Methodology for Creating the Data Module

Once the NETS data were delivered, the research team cleaned, validated, and converted the time-series data into a format that could be linked to the PSID and used to analyze the built food environment. The data cleaning process was based largely on the methods used by the Jackson Heart Study (JHS) and MESA Neighborhood Study, two studies that also used Dun and Bradstreet data to examine the built physical and food environment. Below we detail each step of the cleaning and validation process that was used to create this data module.

A. Data Cleaning and Validation

Headquarters. In the first step of the data cleaning process, all establishment headquarters were removed from the database because food establishments generally do not sell food products to the public at these locations (N. Colabianchi, personal communication, July 18, 2016). A particular establishment was considered a headquarters location if the headquarters identifier (HQDuns) was listed multiple times in the NETS database and the establishment had a Duns number that matched the HQDuns number. These criteria ensured that we removed only the corporate headquarters and retained establishments that occupy a single location.

Duplicates. Some establishments were listed multiple times in the dataset due to variations in the spelling of the company name, tradename, or address. In the second step of the cleaning process, we flagged establishments with matching latitudinal and longitudinal coordinates as potential duplicates (K. Moore, personal communication, September 14, 2016), and used SAS code to standardize business addresses (e.g., changed "Rd" to "Road") to minimize variation in address spellings. Next, we used the SPEDIS command in SAS to identify establishments with exact and near identical matches in company name, trade name, and address (Moore, K., personal communication, September 14, 2016).

Many duplicates were detected using SAS code, but there were some establishment duplicates that required additional verification. Once a list of potential duplicate establishments was generated by the SPEDIS command, the establishments on the list were checked individually by a team of research assistants to determine whether duplicates flagged by the SPEDIS code were actually duplicates or different establishments. The process for verifying the geographic location of an establishment in-person is known as "ground-truthing" (Liese et al., 2007). In the past this process was accomplished through checks done on foot or in a car. Due to advancements in technology, we used online tools such as Google Street-View to check and validate the geographic location of each establishment (Rundle et al., 2011).

However, Google-Street View does not have street view for some establishments in the early 2000s. In these instances, we used a combination of internet searches via Yelp, Opentable, and other resources to determine whether an establishment was a duplicate. When it was not possible to determine whether an establishment was truly a duplicate, we left it in the database as a unique establishment.

Table 1 lists estimates of the number of establishments removed from FAM at each step of the data cleaning and validation process.

Table 1

FAM Establishment Totals from Data Cleaning

Year	Number of Establishments in Original File	Headquarter Locations Removed	Potential Duplicates Detected in SAS	Verified Duplicates Removed
2000	1,012,402	23,796	13,267	9,164
2001	1,082,038	23,741	14,796	10,345
2002	1,101,673	22,920	14,271	8,588
2003	1,100,532	22,623	12,859	9,232
2004	1,125,512	21,630	12,376	8,571
2005	1,157,746	20,665	12,756	8,921
2006	1,220,522	19,517	13,292	9,024
2007	1,249,965	19,034	13,421	9,001
2008	1,244,102	18,222	13,310	9,059
2009	1,230,286	17,373	13,574	9,360
2010	1,162,716	15,478	12,648	7,848
2011	1,161,965	14,579	12,728	8,482
2012	1,115,061	13,877	12,022	7,565
2013	1,034,228	13,222	10,624	5,415
2014	1,227,077	13,647	10,314	9,672

B. Identifying Food Categories

Once headquarters and duplicates were removed, each establishment was assigned to a food category. During this step, we also removed establishments that do not provide food directly to the general public. For example, food establishments located on military bases (i.e., military commissaries), food distribution centers, and caterers were excluded from this module.

A total of 14 different food categories were developed after extensive literature review and exploration of the NETS database. The food establishment categories in FAM are the following: grocery stores, supermarkets, convenience stores, gas stations with convenience stores, supercenters, discount clubs, fast food restaurants, full-service restaurants, ethnic restaurants, bars, fruit and vegetable markets, meat markets, health food stores, and specialty stores.

The criteria used to identify each category and the citations of studies that have used similar criteria are listed in Table 2.

Table 2
Food Establishment Category Criteria and Sources

Food Category	SIC Codes	SIC Code Description	Additional Criteria	Citations
Bars	5813	Drinking Places		• Auchincloss et al. (2012).
Club Stores	--	--	SAS SPEDIS command used to find common discount club stores, including Sam's Club, BJ's Wholesale Club, and Costco.	• Martz et al. (2014).
Convenience Stores	54110200	Convenience Stores	Establishments with grocery store SIC codes and 5 or fewer employees also in this category to include corner grocery stores ("bogedas")	• Auchincloss et al. (2012). • Jilcott et al. (2010). • Lee. (2012). • Moore & Diez Rouz. (2006). • Rundle et al. (2009).
	54110201	Convenience Stores, Chain		
	54110202	Convenience Stores, Independent		
	54110000	Grocery Stores		
	54110100	Supermarkets		
	54110102	Supermarkets, >100,000 ft.		
	54110103	Supermarkets, Independent		
	54110104	Supermarkets, 55,000-65,000		
	54110105	Supermarkets, 66,000-99,000		
	54119900	Grocery Stores, nec		
	54119901	Cooperative Food Stores		
	54119904	Grocery Stores, Chain		
	54119905	Grocery Stores, Independent		
Ethnic Restaurants	581201	Ethnic Food Restaurants	Any restaurant that identified as an "American Restaurant" industry was removed from this category.	n/a
Fast Food Restaurants	581202	Ice Cream Stands or Dairy Bars	Based on a list of fast food companies and tradenames. See Appendix B for Fast Food Restaurant list.	• Jilcott et al. (2010). • Lee. (2012). • Moore. (2016)* • Powell & Yanjun. (2009). • Rundle et al. (2009).
	58120300	Fast Food Restaurants & Stands		
	58120301	Box Lunch Stand		
	58120302	Carry-Out Only Restaurant		
	58120303	Chili Stand		
	58120305	Delicatessen (Eating Places)		
	58120306	Drive-In Restaurant		
	58120307	Fast-Food Restaurant, Chain		
	58120308	Fast-Food Rest., Independent		
	58120309	Food Bars		
	58120310	Grills (Eating Places)		
	58120311	Hamburger Stand		
	58120312	Hot Dog Stand		
	58120313	Sandwiches & Submarine Shop		
	58120314	Snack Bar		
	58120315	Snack Shop		
	58120600	Pizza Restaurants		
	58120601	Pizzeria, Chain		
	58120602	Pizzeria, Independent		
	58120400	Lunchrooms & Cafeterias		
	58120401	Automat (Eating Places)		
	59630206	Lunch Wagon		
	59630204	Mobile Food Service		
	59631205	Mobile Ice Cream Wagon		
Fruit/Vegetable	543	Fruit and Vegetable Markets		• Auchincloss et al. (2012). • Fleischhacker et al. (2012). • Moore et al. (2013). • Moore & Diez Rouz. (2006). • Rundle et al. (2009).
Full-Service Restaurants	5812	Restaurants	Establishments not already categorized as fast food or any other food category	• Jilcott et al. (2010). • Moore. (2016)* • Powell & Yanjun. (2009). • Rundle et al. (2009).

Table 2 (continued)
Food Establishment Category Criteria and Sources

Food Category	SIC Codes	SIC Code Description	Additional Criteria	Citations
Gas Stations with Convenience Stores	--	Gasoline Service Stations w/ Convenience Stores	Businesses with a primary SIC code containing "554" for the first three digits and a secondary SIC code with "541" were also included in this category. Establishments with SIC codes for supermarkets/grocery stores and industry listed as "Filling station, gasoline" or "Gasoline service stations."	• Fleischhacker et al. (2012). • Jilcott et al. (2010). • Liese et al. (2010). • Martz et al. (2014). • Morland et al. (2002).
Grocery Stores	54110000 54110100 54110101 54110102 54110103 54110104 54110105 54119900 54119901 54119904 54119905	Grocery Stores Supermarkets Supermarkets, Chain Supermarkets, > 100,000 Supermarkets, Independent Supermarkets, 55,000 - 65,000 Supermarkets, 66,000 - 99,000 Grocery Stores, nec Cooperative Food Stores Grocery Stores, Chain Grocery Stores, Independent	Gross < $2 million in sales and employees < 25 but > 5. If sales data is missing, only number of employees is used.	• Auchincloss et al. (2012). • Gibson. (2011). • Jilcott et al. (2010). • Lee (2012). • Liese et al. (2007). • Moore & Diez Rouz. (2006). • Rundle et al. (2009).
Health Food Stores	549901	Health and Dietetic Food Stores		• Moore. (2016).* • Rundle et al. (2009).
Meat and Fish Markets	542	Meat and Fish Markets		• Auchincloss et al. (2012). • Fleischhacker et al. (2012). • Moore & Diez Rouz. (2006). • Rundle et al. (2009).
Specialty Food Stores	544 546 5451 549900 549902 549999 58120304 58129902 59630207	Candy, Nut, Confectionary Retail Bakeries Dairy Product Stores Miscellaneous Food Stores, nec Beverage Stores Miscellaneous Food Stores, nec Coffee Shop Café Snacks, Direct Sales		• Fleischhacker et al. (2012). • Moore & Diez Rouz. (2006). • Morland et al. (2002). • Rundle et al. (2009).
Supercenters	--	--	SAS SPEDIS command was used to find common supercenters, including Walmart, K-Mart, Target, and Meijer.	• Fleischhacker et al. (2012). • Martz et al. (2014).
Supermarkets	54110000 54110100 54110101 54110102 54110103 54110104 54110105 54119900 54119901 54119904 54119905	Grocery Stores Supermarkets Supermarkets, Chain Supermarkets, > 100,000 Supermarkets, Independent Supermarkets, 55,000 - 65,000 Supermarkets, 66,000 - 99,000 Grocery Stores, nec Cooperative Food Stores Grocery Stores, Chain Grocery Stores, Independent	Gross ≥ $2 million in sales and have ≥ 25 employees. If sales data is missing, only number of employees is used.	• Auchincloss et al. (2012). • Jilcott et al (2010). • Lee (2012). • Liese et al. (2007). • Moore et al. (2013). • Moore & Diez Rouz. (2006). • Rundle et al. (2009).

Note. nec = "Not Elsewhere Classified"; *Citation refers to personal communication with Kari Moore.

Note that each establishment in this data module was assigned to one food category. The "ethnic restaurant" category is the only exception to this rule. Below we highlight the criteria used to create certain food categories.

Club stores. Our review of the literature indicated that the North American Industry Classification System (NAICS) codes were primarily used to distinguish which establishments were considered discount or warehouse club stores. The specific NAICS code referenced in the reviewed literature was 452910 (Martz, et al., 2014). Similar to the inconsistent categorization of supercenters (see below), this classification code did not identify all the club stores. For example, in the 2005 NETS data, this NAICS code identified 891 establishments as club stores, but missed approximately 145 club stores. To address this problem, we identified larger discount stores such as Sam's Club, BJ's, and Costco using the SAS SPEDIS command to find matches by company name.

Ethnic restaurants. Ethnic restaurants are a unique food category in FAM. We used the definition from *The SAGE Encyclopedia of Food Issues,* which defined ethnic restaurants as those "that serve food from countries other than the one in which they are located" (Fox, 2015). The 6-digit SIC code for ethnic restaurant (581201) contained restaurants belonging to an industry of "American Restaurants". These locations were removed so that the ethnic restaurant category was a better match to the established definition for families living in the U.S. Establishments categorized as an ethnic restaurant can also be categorized as a fast food or full-service restaurants in FAM.

Fast food restaurants. Fast food restaurants were identified using SIC codes and the SPEDIS command with company and tradenames. We found that relying on SIC codes alone did not adequately identify all fast food restaurants. Similar to methods used by other research teams, we compiled a list of fast food restaurant chains using listings assembled by Restaurants & Institutions (R&I) and Nation's Restaurant News (NRN) (K. Moore, personal communication, September 14, 2016). Our final list of fast food restaurants included establishments listed in the "Top 100 Chains Ranked by U.S. Systemwide Foodservice Sales" for 2010 and 2013 by NRN and R&I's "2009 Top 400 Restaurant Chains". These documents were available online (NRN, 2010; NRN, 2013; R&I, 2009). After compiling these restaurant names we removed establishments that provided waiter service from the final fast food list. (See Appendix B for the Fast Food Restaurant Chain List.)

Gas stations with convenience stores. We observed a significant presence of prominent gas station companies, such as "Sunoco" or "Speedway", listed under the SIC codes for the convenience store, grocery store, and supermarket categories. Thus, we created the category – Gas Stations with Convenience Stores – using a combination of primary and secondary SIC codes along with using the Industry variable available with the NETS database. As indicated by Table 2, the category Gas Station with Convenience stores includes establishments with primary or secondary SIC codes with "554" in the first three digits. Note that it is possible that not every gas station with a convenience store was included in this module because the primary SIC code for gas station was not included in our NETS database order. We created this category to assist with transparency regarding the establishments in each category, but we leave it to users to decide whether to include gas stations in their analyses of convenience stores.

Supercenters. Our original order of the NETS dataset included names of specific supercenters (BJ's Wholesale Club, Costco, Kmart, Sams Club, Target, Walmart). If we had not requested these stores by name, over 90% of these establishments would not have been identified

based on our requested SIC codes alone (D. Walls, personal communication, May 31, 2016). Further analysis of the data from our NETS order showed that stores like Walmart were primarily identified by SIC code 5311, which represents department stores. Other SIC codes used for these supercenters included 4225 (general warehousing and storage) and 5999 (miscellaneous retail store). Only a few of these supercenters possessed secondary SIC codes that helped identify these centers as a source for groceries (SIC code beginning with 5411). For these reasons, we used the SPEDIS command in SAS to identify and classify these named establishments as food supercenters.

Initially, we assumed that establishments with the word pharmacy in the tradename (e.g., "Walmart Pharmacy") should be removed on the basis that pharmacies generally are not sources of food. However, we found that some large supercenter chains (i.e., Walmart) with "pharmacy" in the tradename actually represented the entire supercenter rather than just the pharmacy because there were no other entries for that establishment in the NETS database at that location. Thus, establishments fitting these criteria were coded as supercenters.

Meijer was classified inconsistently by SIC code as well. Unfortunately, we assumed that Meijer would be classified as a supermarket, and did not request it in our order of 'named' supercenters. Since SIC codes are self-selected by each establishment and are based on the company's primary line of business that "generates the highest revenue for that company at a specific location in the past year" (https://siccode.com/en/pages/how-to-select-your-sic-code), there can be discrepancies in how each establishment self-identifies each year. Meijer stores had SIC codes that classified their establishment across a range of store types including but not limited to discount stores and general warehouses. These inconsistencies were similar to those found with other supercenter establishments mentioned previously. Because it was difficult to identify Meijer stores using the SIC codes that helped identify supermarkets, we included Meijer stores in the supercenter category.

C. Calculating Geographic Indicators: Density, Euclidean, and Road-Network Distance

Geographic indicators were calculated for all 14 food categories. The process used for calculating geographic indicators are adapted in part from the Obesity and Neighborhood Environment Database (ONEdata) (http://www.cpc.unc.edu/projects/onedata), which incorporated various geographic indicators of the food environment for Add Health respondents in study waves I and III.

The FAM links families from the PSID to food access indicators that were calculated with ArcMap Software (Version 10.3.1) using food establishment information derived from the NETS containing business information from 2000-2014 (Walls & Associates, 2016).

We created food outlet density and closest facility distance variables to characterize PSID families' residential physical food environment using 2010 U.S. Census boundaries (the most recent Census). Specifically, we used the block-group centroid (a point that is the defined

"center" of the block-group according to the U.S. Census) as a proxy for the PSID family's residential location.

Details on the projection selection and calculation of the geospatial indicators can be found in Section VI. Technical Notes on Constructing the Data Module.

Two Types of Indicators

- **Distance indicators** refer to the distance (in miles) from a PSID family's residential location to the closest food outlet.

- **Density calculations** refer to the total number of food establishments within a specified buffer surrounding a family's residential location.

Similar to the OneData data module, both distance and density variables in the FAM were extracted using Euclidean (straight-line) measures and road network analysis (Gordon-Larson, 2009). Food establishment densities were measured at ½-, 1-, 5-, and 10- mile Euclidean radii and ½-, 1-, and 5- mile road network buffers.

Figures 1 and 2 demonstrate an example of the difference between Euclidean (straight-line) and road network distance calculations using simulated locations of food establishments for a fictional residence. Figure 3 provides an example of a Euclidean buffer and Figure 4 shows an example of a road network buffer.

Methodology for extracting distance and density food environment indicators are outlined in the appendices.

Figures 1-4.
Examples of Geographic Indicators for a Fictional Residence

Figure 1
Euclidean Distance to Closest Grocery Store

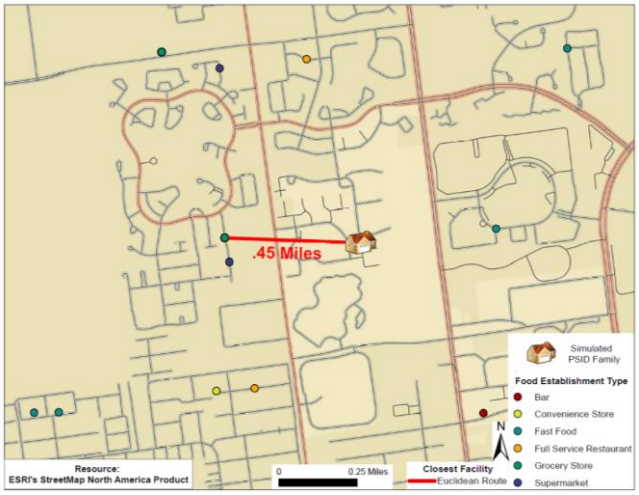


Figure 2
Road Network Distance to Closest Grocery Store

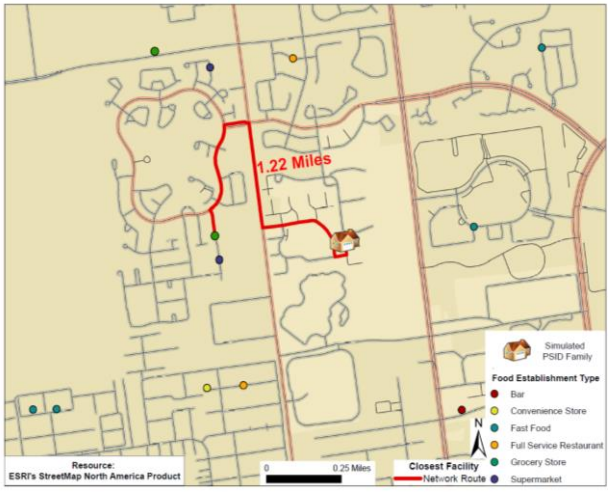


Figure 3
Number of Food Establishments within a 1-mile Euclidean Buffer

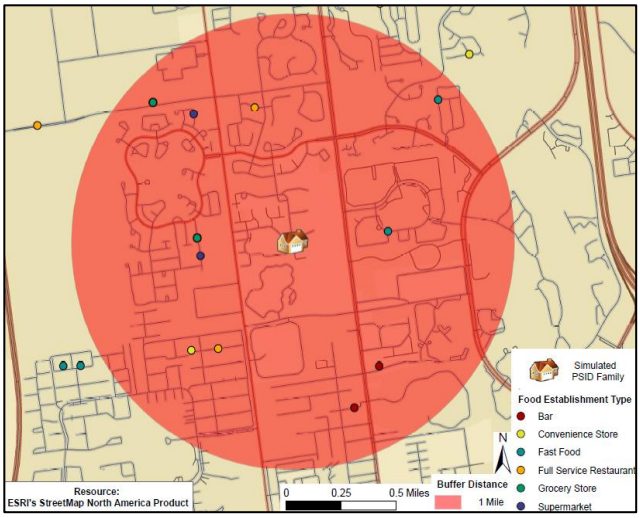
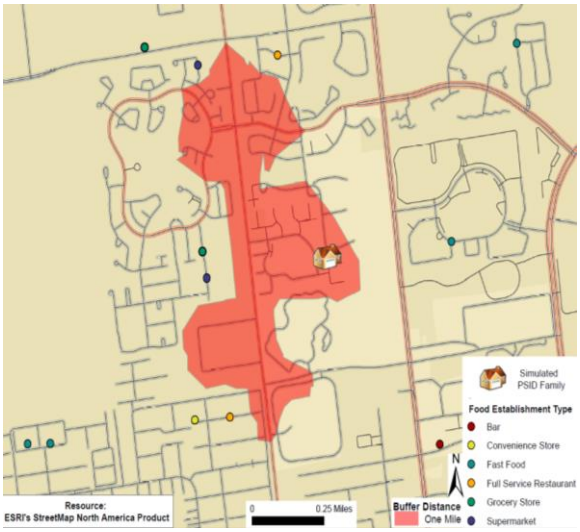


Figure 4
Number of Food Establishments within a 1-mile Road Network Buffer



Note. The house icons in the figures represent a family's residential location based on the block-group centroid.

IV. How to Use the FAM Data with PSID Survey Data

The Panel Study of Income Dynamics (PSID) is a longitudinal survey that has continuously represented families nationally since 1968 and contains a wealth of data involving various subject areas, including, but not limited to, income, child development, employment status, and wealth. The original sample of the PSID consisted of approximately 5,000 family units and by 2015 has expanded to over 9,000 family units.

The PSID is composed of a main interview, which has been conducted biennially since 1997, and multiple supplemental studies such as, the Child Development Supplement (CDS) and the Transition into Adulthood Supplement (TAS) (McGonagle, Schoeni, Sastry, & Freedman, 2012). The CDS, which focuses on eligible children within participating PSID families, provides a nationally representative, comprehensive resource for researchers interested in studying child development. Data for the Original CDS were collected in 1997, 2002, and 2007, with a new cohort beginning in 2014. Data collection for the Original TAS began in 2005 and focuses on young adults who participated in the CDS. The TAS is conducted during the same years as the main PSID interview, shortly after the main PSID collection is completed. Original TAS data are available for 2005, 2007, 2009, 2011, 2013, and 2015. For additional information on the main PSID interview and its supplemental studies, visit the PSID website: https://psidonline.isr.umich.edu.

The FAM can be linked to every year of the main PSID survey beginning in the year 2001 through 2015; to the TAS for the years 2005, 2007, 2009, 2011, and 2013; and to the CDS for the years 2002, 2007, and 2014. Instructions on how to link the FAM to the main or supplemental studies are outlined in Section IV.B. Linking FAM with the PSID and Supplemental Studies.

A. PSID Geocodes in FAM

This module contains information on the physical food environment created from the NETS data for the years 2000-2014. The food environment indicators for each year was linked to PSID families' residential location at the block-group level using 2010 U.S. Census boundary geocodes (the most recent Census) (PSID, June 2017). Although NETS data contains food establishments that exist each year, the restricted PSID data contain geolocations of participating families during the years in which the main interview was conducted. For non-PSID interview years (2000, 2002, 2004, 2006, 2008, 2010, and 2012), the block-group designations were drawn from PSID restricted block-group data from the interview year that occurred one year later. For example, the food environment indicators for the year 2000 used geolocations that represented family locations at the time of the 2001 main PSID interview.

The exception to this rule is year 2014, where food access indicators were linked to PSID families' residences in 2013 *and* 2015 to maximize flexibility for answering research questions related to the re-established Food Security items in the 2014 CDS and 2015 main PSID

interview. The rationale for using the PSID geocodes in this manner was driven by PSID's propensity to ask participants to recall events from the previous year for various questions (see Table 3).

The main PSID interview also includes relocation questions that asked if the family had moved since the last interview and the date the move occurred if the family had relocated (PSID, May 2017). FAM users may want to use information on whether participants moved during or prior to the interview year to determine which families to include in analyses for non-PSID interview years (i.e., years without a residential PSID geocode).

B. Linking FAM with the PSID and Supplemental Studies

Merging FAM data with PSID data can be accomplished by using two variables in FAM: **FAMID10**, which represents the Family Interview (ID) Number, and **PSID_year**, the main PSID interview year. These two items are necessary to properly link FAM with the PSID family files because **FAMID10** changes from year to year (PSID, 2016).

Thus, to facilitate the merging process, users should first rename the PSID Family Interview (ID) Number variable in the PSID data to FAMID10. Next, users can determine which years of the FAM to include in their analytic dataset by using the **PSID_year** and **Food_year** variables. The PSID_year variable indicates the PSID interview year from which the geocode information originated, and the **Food_year** variable indicates the year from which the food access indicators were created. For example, if a user selected PSID_year = 2007 and Food_year = 2007, then the user would have data on PSID families interviewed in the year 2007 using food indicators that were calculated using food establishment data from 2007. To merge this year 2007 subsample of FAM with the main PSID data, the user would have to use the Family Interview Number from the 2007 survey.

A similar method is used to merge FAM and TAS data because the TAS dataset already contains the Family Interview Number variable for that year. However, merging FAM with CDS data requires an additional step to ensure that the correct family information is linked to each participating child. CDS waves conducted in 2002 and 2007 do not include PSID Family Interview numbers for those specific years. Addressing this issue requires linking the children within the CDS to the PSID's cross-year individual file using the variables within the CDS, ER30001 and ER30002, which are the 1968 Family Interview Number and the Person Number, respectively. A combination of these two variables serves as a unique identifier for everyone who has participated in the PSID. After merging the children with the PSID Individual file, find the PSID wave of interest and use the Family Interview numbers from that year to merge with the FAM. The CDS survey conducted in 2014 already contains the Family Interview ID for 2013 PSID survey (PSID, 2017). For more information about the structure or the PSID, please refer to this document https://psidonline.isr.umich.edu/Guide/FileStructure.pdf (PSID, 2016).
For additional information on the CDS, including eligibility requirements for participation, please see the User Guides for both the Original CDS (collected from 1997-2007) and the Ongoing CDS (begun in 2014) located here:
https://psidonline.isr.umich.edu/Guide/documents.aspx

C. Food-related PSID Survey Questions

Questions on food accessibility, diet, and use of food stamps are available in the PSID for various interview years. Tables 3, 4, and 5 include a non-exhaustive list of the different food-related subject areas in the main PSID, the CDS, and TAS, along with the interview years that these items were examined. Table 3, which represents the main PSID, contains additional columns, including "Year referenced #1," "Year referenced #2," and "Year referenced #3." These columns were included because there are multiple interview questions within the main interview. For example, questions related to food security and food stamp usage are asked about the previous year or two years prior; a question asked during the main PSID interview in 2003 asked, "Did you (or anyone else in your family) use government food stamps at any time in 2002?" In general, food-related questions asked of participants within the CDS and TAS referred to the same year or "within the past 12 months" (PSID, 2017).

Table 3

Food-Related Survey Questions in the Main PSID Survey

Food-Related Data Availability

Panel Study of Income Dynamics, Main Interview

Food Assistance Questions

Example Question:

"Did you (or anyone else in your family) use government food stamps at any time in (year)?"

Interview Year	Year Referenced #1	Year Referenced #2	Year Referenced #3
2001	1999	2000	2001
2003	2001	2002	2003
2005	2003	2004	2005
2007	2005	2006	2007
2009	2007	2008	2009
2011	2009	2010	2011
2013	2011	2012	2013
2015	2013	2014	2015

Family Food Cost Questions:

Example Question:

"How much do you (and everyone else in your family) spend on food that you use at home in an average week?"

Interview Year	Year Referenced #1	Year Referenced #2	Year Referenced #3
2001	2001		
2003	2003		
2005	2005		
2007	2007		
2009	2009		
2011	2011		
2013	2013		
2015	2015		

Child & Adult Care Food Availability

Example Question:

"Were meals or snacks provided in any of these day care arrangements (attended last year)?"

Interview Year	Year Referenced #1	Year Referenced #2	Year Referenced #3
2001	2000		
2003	2002		
2005	2004		
2007	2006		
2009	2008		
2011	2010		
2013	2012		
2015	2014		

Free or Reduced Cost Meals

Example Question:

"During (year), did you (anyone in your family) receive free or reduced-cost meals for the elderly?"

OR

"During (year), did (NAME/any cgukd ub the family between 5 and 18 years old,) receive free or reduced-cost lunches at school?"

Interview Year	Year Referenced #1	Year Referenced #2	Year Referenced #3
2001	2000		
2003	2002		
2005	2004		
2007	2006		
2009	2008		
2011	2010		
2013	2012		
2015	2014		

Food Security Section in Section F of Questionnaire:

Example Question:

"During (previous year), did you (or other adults in your household) ever cut the size of your meals or skip meals because there wasn't enough money for food?"

Interview Year	Year Referenced #1	Year Referenced #2	Year Referenced #3
2001	2000		
2003	2002		
2015	2015		

Food Availability Questions:

Example Question:

"Which of these statements best describes the food eaten in your household in (year): (I/We) had enough to eat and the kinds of food (I/We) wanted; (I/We) had enough to eat but not always the kinds of food (I/we) wanted; sometimes (I/We) didn't have enough food to eat; or often (I/We) didn't have enough to eat?"

Interview Year	Year Referenced #1	Year Referenced #2	Year Referenced #3
2001	2000		
2003	2002		

Table 4

Food-Related Survey Questions in the CDS

Food-Related Data Availability	
Child Development Supplement	
Family Food Cost Questions: **Example Question:** *"How much did you and anyone else living with you spend on food for (CHILD) during the past 12 months"*	**Child Food Preferences** **Example Question:** *"What foods do you like to eat? (SELECT ALL THAT APPLY)"*
Interview Year	Interview Year
2002	2002
Household Rules on Child Food Choice **Example Question:** *"How much choice is (CHILD) allowed in deciding what foods (he/she) eats at breakfast and lunch? Would you say no choice, a little choice, some choice, or a great deal of choice?"*	**Household Rules on Consuming Sweets/Snacks** **Example Question:** *"Do you have rules about how much candy, sweets or other snacks (CHILD) has?"*
Interview Year	Interview Year
2002	2002
2014	2007
Child Eating Habits **Example Question:** *"What do you usually have for breakfast on a weekday morning (SELECT ALL THAT APPLY)"* *AND* *"How many days last week did you eat vegetables?"*	**Food Expenditures Spent By Child (children 12 and over)** **Example Question:** *"How much money did you spend on food last month?"*
Interview Year	Interview Year
2002	2002
2007	2007
Food Assistance **Example Question:** Duration of Activity (in seconds) in Time Diary *"Weekday: Welfare, food stamps; applying for or collecting welfare, food stamps."*	**School Lunch and Free or Reduced-Cost Meal Questions** **Example Question:** *"Are the luches full-price, reduced-price or free?"*
Interview Year	Interview Year
2002	2002
2007	2007
2014	2014
Food Security Section in Section R of Questionnaire **Example Question:** *"(I/We) worried whether (my/our) food would run out before (I/we) got money to buy more. Was that often true, sometimes true, or never true for (you/ your household) in the last 12 months?"*	**Food Preparation:** **Example Question:** *"In the past month, how often did you and (CHILD) prepare food together?"*
Interview Year	Interview Year
2014	2002
	2007

Table 5

Food-Related Survey Questions in the TAS

Food-Related Data Availability
Transition into Adulthood
Eating Habits: **Example Question:** *"How often do you eat abnormally large amounts of food within a few hours, that is eat in binges? Would you say: Never, hardly ever, less than once a month, a couple of times a month, more than once a week, or every day?"*
Interview Year
2005
2007
2009
2011
2013
2015

V. Variables in FAM

A. Variables Created from NETS (Years 2000-2014 Available)

FAM Identifier/Geographic Variables	
Variable Description	**Variable Name**
Family Interview Number Locations of food outlets nationwide derived from the National Establishment Time-Series Database (NETS), which provided the foundation to create distance and density food indicators nationwide. The Food Environment Year is the year within the NETS database used to calculate the food accessibility indicators. For example, a Food Environment year equal to "2001" represents food indicators that were calculated using businesses active throughout 2001 according to the NETS database.	FAMID10
PSID Geocode Year The geocodes used in locating PSID families in this module were only extracted during that main PSID interview years at the block-group level. This variable represents the PSID interview year used for the geocodes, which were used to conduct geographic calculations for the food access indicators.	Psid_year
Food Environment Year Locations of food outlets nationwide derived from the National Establishment Time-Series Database (NETS), which provided the foundation to create distance and density food indicators nationwide. The Food Environment Year is the year within the NETS database used to calculate the food accessibility indicators. For example, a Food Environment year equal to "2001" represents food indicators that were calculated using businesses active throughout 2001 according to the NETS database.	Food_year
State Name PSID family location at the state level.	State
FIPS Code This Federal Information Processing Standards (FIPS) code represents identifiers for U.S. counties in which PSID families are located. This variable is a concatenation of variables STATE10 and COUNTY10, which are available through the restricted PSID dataset.	FIPS
County Name PSID family location at the county level. The name of the county matches the county represented by the FIPS Code variable.	County
Census Tract Code This code represents unique identifiers for U.S. Bureau Census tracts in which PSID families are located. This variable is a concatenation of variables STATE10, COUNTY10, and TRACT10, which are available through the restricted PSID dataset.	Censustract

Road Network Distance Variables

These variables represent the distance (in miles) to the closest food outlet of each food establishment category using a road network from a PSID family locations at the block group level. A value of "99999" indicated that the ArcMap software could not find a resolution for that food category. This situation generally occurred in geographically isolated areas. See Figure 2 for a visual representation of a road network calculation.

Variable Description	Variable Name
Closest Road Network Distance -- Bar	Barn
Closest Road Network Distance -- Convenience Store	Corn
Closest Road Network Distance -- Club Store	Csrn
Closest Road Network Distance -- Ethnic Restaurant	Etrn
Closest Road Network Distance -- Fruit/Vegetable Market	Fvrn
Closest Road Network Distance -- Fast Food Restaurant	Ffrn
Closest Road Network Distance -- Full-Service Restaurant	Fsrn
Closest Road Network Distance -- Gas Station w/ Convenience Store	Garn
Closest Road Network Distance -- Grocery Store	Grrn
Closest Road Network Distance -- Health Food Store	Hern
Closest Road Network Distance -- Meat Market	Mern
Closest Road Network Distance -- Specialty Store	Sprn
Closest Road Network Distance -- Supercenter	Scrn
Closest Road Network Distance -- Supermarket	Smrn

Euclidean (Straight-Line) Distance Variables

These variables represent the straight-line distance (in miles) to the closest food outlet of each food establishment category from PSID family locations at the block group level. Geographically isolated families that did not have access to a specific food outlet category was given a value of "99999" as no solution could be calculated. See Figure 1 for a visual representation of a straight-line distance calculation.

Variable Description	Variable Name
Closest Euclidean (Straight-Line) Distance -- Bar	Bane
Closest Euclidean (Straight-Line) Distance -- Convenience Store	Cone
Closest Euclidean (Straight-Line) Distance -- Club Store	Csne
Closest Euclidean (Straight-Line) Distance -- Ethnic Restaurant	Etne
Closest Euclidean (Straight-Line) Distance -- Fruit/Vegetable Market	Fvne
Closest Euclidean (Straight-Line) Distance -- Fast Food Restaurant	Ffne
Closest Euclidean (Straight-Line) Distance -- Full-Service Restaurant	Fsne
Closest Euclidean (Straight-Line) Distance -- Gas Station w/ Convenience Store	Gane
Closest Euclidean (Straight-Line) Distance -- Grocery Store	Grne
Closest Euclidean (Straight-Line) Distance -- Health Food Store	Hene
Closest Euclidean (Straight-Line) Distance -- Meat Market	Mene
Closest Euclidean (Straight-Line) Distance -- Specialty Store	Spne
Closest Euclidean (Straight-Line) Distance -- Supercenter	Scne
Closest Euclidean (Straight-Line) Distance -- Supermarket	Smne

Road Network Density Variables	
Half-Mile Road Network Density	
The Half-Mile Road Network Density is the total count of food outlets within a specific food establishment category within a half-mile area that can be accessed through the road network. Figure 4 presents a visual representation of a road network density area.	
Variable	**Variable Name**
Half Mile Road Network Density -- Bar	Bahnd
Half Mile Road Network Density -- Convenience Store	Cohnd
Half Mile Road Network Density -- Club Store	Cshnd
Half Mile Road Network Density -- Ethnic Restaurant	Ethnd
Half Mile Road Network Density -- Fruit/Vegetable Market	Fvhnd
Half Mile Road Network Density -- Fast Food Restaurant	Ffhnd
Half Mile Road Network Density -- Full-Service Restaurant	Fshnd
Half Mile Road Network Density -- Gas Station w/ Convenience Store	Gahnd
Half Mile Road Network Density -- Grocery Store	Grhnd
Half Mile Road Network Density -- Health Food Store	Hehnd
Half Mile Road Network Density -- Meat Market	Mehnd
Half Mile Road Network Density -- Specialty Store	Sphnd
Half Mile Road Network Density -- Supercenter	Schnd
Half Mile Road Network Density -- Supermarket	Smhnd
Road Network Density Variables	
One-Mile Road Network Density	
The One-Mile Road Network Density is the total count of food outlets within a specific food establishment category within a one-mile area that can be accessed through the road network. Figure 4 presents a visual representation of a road network density area.	
Variable	**Variable Name**
One Mile Road Network Density -- Bar	Baond
One Mile Road Network Density -- Convenience Store	Coond
One Mile Road Network Density -- Club Store	Csond
One Mile Road Network Density -- Ethnic Restaurant	Etond
One Mile Road Network Density -- Fruit/Vegetable Market	Fvond
One Mile Road Network Density -- Fast Food Restaurant	Ffond
One Mile Road Network Density -- Full-Service Restaurant	Fsond
One Mile Road Network Density -- Gas Station w/ Convenience Store	Gaond
One Mile Road Network Density -- Grocery Store	Grond
One Mile Road Network Density -- Health Food Store	Heond
One Mile Road Network Density -- Meat Market	Meond
One Mile Road Network Density -- Specialty Store	Spond
One Mile Road Network Density -- Supercenter	Scond
One Mile Road Network Density -- Supermarket	Smond

Road Network Density Variables

Five-Mile Road Network Density

The Five-Mile Road Network Density is the total count of food outlets within a specific food establishment category within a five-mile area that can be accessed through the road network. Figure 4 presents a visual representation of a road network density area.

Variable	Variable Name
Five Mile Road Network Density -- Bar	Bafnd
Five Mile Road Network Density -- Convenience Store	Cofnd
Five Mile Road Network Density -- Club Store	Csfnd
Five Mile Road Network Density -- Ethnic Restaurant	Etfnd
Five Mile Road Network Density -- Fruit/Vegetable Market	Fvfnd
Five Mile Road Network Density -- Fast Food Restaurant	Fffnd
Five Mile Road Network Density -- Full-Service Restaurant	Fsfnd
Five Mile Road Network Density -- Gas Station w/ Convenience Store	Gafnd
Five Mile Road Network Density -- Grocery Store	Grfnd
Five Mile Road Network Density -- Health Food Store	Hefnd
Five Mile Road Network Density -- Meat Market	Mefnd
Five Mile Road Network Density -- Specialty Store	Spfnd
Five Mile Road Network Density -- Supercenter	Scfnd
Five Mile Road Network Density -- Supermarket	Smfnd

Euclidean (Straight-Line) Density Variables

Half-Mile Road Network Density

The Half-Mile Euclidean Density is the total count of food outlets within a specific food establishment category within a half-mile radius around a PSID family location. Figure 3 presents a visual representation of a road network density area.

Variable	Variable Name
Half Mile Euclidean Density -- Bar	Bahd
Half Mile Euclidean Density -- Convenience Store	Cohd
Half Mile Euclidean Density -- Club Store	Cshd
Half Mile Euclidean Density -- Ethnic Restaurant	Ethd
Half Mile Euclidean Density -- Fruit/Vegetable Market	Fvhd
Half Mile Euclidean Density -- Fast Food Restaurant	Ffhd
Half Mile Euclidean Density -- Full-Service Restaurant	Fshd
Half Mile Euclidean Density -- Gas Station w/ Convenience Store	Gahd
Half Mile Euclidean Density -- Grocery Store	Grhd
Half Mile Euclidean Density -- Health Food Store	Hehd
Half Mile Euclidean Density -- Meat Market	Mehd
Half Mile Euclidean Density -- Specialty Store	Sphd
Half Mile Euclidean Density -- Supercenter	Schd
Half Mile Euclidean Density -- Supermarket	Smhd

Road Network Density Variables

One-Mile Road Network Density

The One-Mile Road Network Density is the total count of food outlets within a specific food establishment category within a one-mile area that can be accessed through the road network. Figure 4 presents a visual representation of a road network density area.

Variable	Variable Name
One Mile Road Network Density -- Bar	Baond
One Mile Road Network Density -- Convenience Store	Coond
One Mile Road Network Density -- Club Store	Csond
One Mile Road Network Density -- Ethnic Restaurant	Etond
One Mile Road Network Density -- Fruit/Vegetable Market	Fvond
One Mile Road Network Density -- Fast Food Restaurant	Ffond
One Mile Road Network Density -- Full-Service Restaurant	Fsond
One Mile Road Network Density -- Gas Station w/ Convenience Store	Gaond
One Mile Road Network Density -- Grocery Store	Grond
One Mile Road Network Density -- Health Food Store	Heond
One Mile Road Network Density -- Meat Market	Meond
One Mile Road Network Density -- Specialty Store	Spond
One Mile Road Network Density -- Supercenter	Scond
One Mile Road Network Density -- Supermarket	Smond

Road Network Density Variables

Five-Mile Road Network Density

The Five-Mile Road Network Density is the total count of food outlets within a specific food establishment category within a five-mile area that can be accessed through the road network. Figure 4 presents a visual representation of a road network density area.

Variable	Variable Name
Five Mile Road Network Density -- Bar	Bafnd
Five Mile Road Network Density -- Convenience Store	Cofnd
Five Mile Road Network Density -- Club Store	Csfnd
Five Mile Road Network Density -- Ethnic Restaurant	Etfnd
Five Mile Road Network Density -- Fruit/Vegetable Market	Fvfnd
Five Mile Road Network Density -- Fast Food Restaurant	Fffnd
Five Mile Road Network Density -- Full-Service Restaurant	Fsfnd
Five Mile Road Network Density -- Gas Station w/ Convenience Store	Gafnd
Five Mile Road Network Density -- Grocery Store	Grfnd
Five Mile Road Network Density -- Health Food Store	Hefnd
Five Mile Road Network Density -- Meat Market	Mefnd
Five Mile Road Network Density -- Specialty Store	Spfnd
Five Mile Road Network Density -- Supercenter	Scfnd
Five Mile Road Network Density -- Supermarket	Smfnd

Euclidean (Straight-Line) Density Variables Half-Mile Road Network Density	
The Half-Mile Euclidean Density is the total count of food outlets within a specific food establishment category within a half-mile radius around a PSID family location. Figure 3 presents a visual representation of a road network density area.	
Variable	**Variable Name**
Half Mile Euclidean Density -- Bar	Bahd
Half Mile Euclidean Density -- Convenience Store	Cohd
Half Mile Euclidean Density -- Club Store	Cshd
Half Mile Euclidean Density -- Ethnic Restaurant	Ethd
Half Mile Euclidean Density -- Fruit/Vegetable Market	Fvhd
Half Mile Euclidean Density -- Fast Food Restaurant	Ffhd
Half Mile Euclidean Density -- Full-Service Restaurant	Fshd
Half Mile Euclidean Density -- Gas Station w/ Convenience Store	Gahd
Half Mile Euclidean Density -- Grocery Store	Grhd
Half Mile Euclidean Density -- Health Food Store	Hehd
Half Mile Euclidean Density -- Meat Market	Mehd
Half Mile Euclidean Density -- Specialty Store	Sphd
Half Mile Euclidean Density -- Supercenter	Schd
Half Mile Euclidean Density -- Supermarket	Smhd
Euclidean (Straight-Line) Density Variables One-Mile Road Network Density	
The One-Mile Euclidean Density is the total count of food outlets within a specific food establishment category within a one-mile radius around a PSID family location. Figure 3 presents a visual representation of a road network density area.	
Variable	**Variable Name**
One Mile Euclidean Density -- Bar	Baod
One Mile Euclidean Density -- Convenience Store	Cood
One Mile Euclidean Density -- Club Store	Csod
One Mile Euclidean Density -- Ethnic Restaurant	Etod
One Mile Euclidean Density -- Fruit/Vegetable Market	Fvod
One Mile Euclidean Density -- Fast Food Restaurant	Ffod
One Mile Euclidean Density -- Full-Service Restaurant	Fsod
One Mile Euclidean Density -- Gas Station w/ Convenience Store	Gaod
One Mile Euclidean Density -- Grocery Store	Grod
One Mile Euclidean Density -- Health Food Store	Heod
One Mile Euclidean Density -- Meat Market	Meod
One Mile Euclidean Density -- Specialty Store	Spod
One Mile Euclidean Density -- Supercenter	Scod
One Mile Euclidean Density -- Supermarket	Smod

Euclidean (Straight-Line) Density Variables	
Five-Mile Road Network Density	
The Five-Mile Euclidean Density is the total count of food outlets within a specific food establishment category within a five-mile radius around a PSID family location. Figure 3 presents a visual representation of a road network density area.	
Variable	**Variable Name**
Five Mile Euclidean Density -- Bar	Bafd
Five Mile Euclidean Density -- Convenience Store	Cofd
Five Mile Euclidean Density -- Club Store	Csfd
Five Mile Euclidean Density -- Ethnic Restaurant	Etfd
Five Mile Euclidean Density -- Fruit/Vegetable Market	Fvfd
Five Mile Euclidean Density -- Fast Food Restaurant	Fffd
Five Mile Euclidean Density -- Full-Service Restaurant	Fsfd
Five Mile Euclidean Density -- Gas Station w/ Convenience Store	Gafd
Five Mile Euclidean Density -- Grocery Store	Grfd
Five Mile Euclidean Density -- Health Food Store	Hefd
Five Mile Euclidean Density -- Meat Market	Mefd
Five Mile Euclidean Density -- Specialty Store	Spfd
Five Mile Euclidean Density -- Supercenter	Scfd
Five Mile Euclidean Density -- Supermarket	Smfd
Euclidean (Straight-Line) Density Variables	
Ten-Mile Road Network Density	
The Ten-Mile Euclidean Density is the total count of food outlets within a specific food establishment category within a ten-mile radius around a PSID family location. Figure 3 presents a visual representation of a road network density area.	
Variable	**Variable Name**
Ten Mile Euclidean Density -- Bar	Batd
Ten Mile Euclidean Density -- Convenience Store	Cotd
Ten Mile Euclidean Density -- Club Store	Cstd
Ten Mile Euclidean Density -- Ethnic Restaurant	Ettd
Ten Mile Euclidean Density -- Fruit/Vegetable Market	Fvtd
Ten Mile Euclidean Density -- Fast Food Restaurant	Fftd
Ten Mile Euclidean Density -- Full-Service Restaurant	Fstd
Ten Mile Euclidean Density -- Gas Station w/ Convenience Store	Gatd
Ten Mile Euclidean Density -- Grocery Store	Grtd
Ten Mile Euclidean Density -- Health Food Store	Hetd
Ten Mile Euclidean Density -- Meat Market	Metd
Ten Mile Euclidean Density -- Specialty Store	Sptd
Ten Mile Euclidean Density -- Supercenter	Sctd

B. Variables from USDA Research Atlas (Year 2010 Available)

USDA Research Atlas Variable	Variable Name
Urban Tract Indicator	Urban
Tract Population	POP2010
Total Number of Housing Units	OHU2010
High Population Living in Group Quarters Indicator	GroupQuartersFlag
Total Number of Group Quarters	NUMGQTRS
Percent Pop. Residing in Group Quarters	PCTGQTRS
Low Income and Low Access Tract (1 mile for Urban Areas and 10 Miles for Rural Areas)	LILATracts_1And10
Low Income and Low Access Tract (1/2 mile for Urban Areas and 10 Miles for Rural Areas)	LILATracts_HalfAnd10
Low Income and Low Access Tract (1 mile for Urban Areas and 20 Miles for Rural Areas)	LILATracts_1And20
Low Income and Low Access Tract Using Vehicle Access (or Low Access Tract calculated at 20 miles	LILATracts_Vehicle
Low Vehicle Access Tract	HUNVFlag
Low Income Tract	LowIncomeTracts
Percent Pop. Living Below the Federal Poverty Rate	PovertyRate
Tract Family Median Year Income	MedianFamilyIncome
Low Access Tract (1 mile for Urban Areas and 10 Miles for Rural Areas)	LA1and10
Low Access Tract (1/2 mile for Urban Areas and 10 Miles for Rural Areas)	LAhalfand10
Low Access Tract (1 mile for Urban Areas and 20 Miles for Rural Areas)	LA1and20
Low Access Tract at ½ Mile	LATracts_half
Low Access Tract at 1 Mile	LATracts1
Low Access Tract at 10 Miles	LATracts10
Low Access Tract at 20 Miles	LATracts20
Tract w/ Low Access & Low Vehicle Access*	LATractsVehicle_20
Population w/ Low Access (1 mile for Urban Areas and 10 Miles for Rural Areas)	LAPop1_10
Population w/ Low Access (1/2 mile for Urban Areas and 10 Miles for Rural Areas)	LAPop05_10
Population w/ Low Access (1 mile for Urban Areas and 20 Miles for Rural Areas)	LAPop1_20
Population w/ Low Access and Low Income (1 mile for Urban Areas and 10 Miles for Rural Areas)	LALowi1_10
Population w/ Low Access and Low Income (1/2 mile for Urban Areas and 10 Miles for Rural Areas)	LALowi05_10
Population w/ Low Access and Low Income (1 mile for Urban Areas and 20 Miles for Rural Areas)	LALowi1_20

USDA Research Atlas Variable (continued)	Variable Name
Population w/ Low Access at ½ Mile	Lapophalf
Percent Population w/ Low Access at ½ Mile	Lapophalfshare
Population w/ Low Access & Low Income at ½ Mile	Lalowihalf
Percent Population w/ Low Access & Low Income at ½ Mile	Lalowihalfshare
Children (ages 0-17) w/ Low Access at ½ Mile	Lakidshalf
Percent Children (ages 0-17) w/ Low Access at ½ Mile	Lakidshalfshare
Seniors (age >=65) w/ Low Access	Laseniorshalf
Percent Seniors (age >= 65) w/ Low Access	laseniorshalfshare
White Population w/ Low Access at 1/2 Mile	lawhitehalf
Percent of White Population w/ Low Access at 1/2 Mile	lawhitehalfshare
Black Population w/ Low Access at 1/2 Mile	lablackhalf
Percent of Black Population w/ Low Access at 1/2 Mile	lablackhalfshare
Asian Population w/ Low Access at 1/2 Mile	laasianhalf
Percent of Asian Population w/ Low Access at 1/2 Mile	laasianhalfshare
Native Hawaiian or Other Pacific Islander Population w/ Low Access at 1/2 Mile	lanhopihalf
Percent of Native Hawaiian or Other Pacific Islander Population w/ Low Access at 1/2 Mile	lanhopihalfshare
American Indian or Alaska Native Population w/ Low Access at 1/2 Mile	laaianhalf
Percent of American Indian or Alaska Native Population w/ Low Access at 1/2 Mile	laaianhalfshare
Other/Multiple Race Population w/ Low Access at 1/2 Mile	laomultirhalf
Percent of Other/Multiple Race Population w/ Low Access at 1/2 Mile	laomultirhalfshare
Hispanic or Latino Population w/ Low Access at 1/2 Mile	lahisphalf
Percent of Hispanic or Latino Population w/ Low Access at 1/2 Mile	lahisphalfshare
Population without vehicle access and Low Access at ½ mile	Lahunvhalf
Percent without Vehicle Access and Low Access at ½ Mile	Lahunvhalfshare
Population of SNAP Households w/ Low Access at 1/2 Mile	Lasnaphalf
Percent of SNAP Households w/ Low Access at 1/2 Mile	Lasnaphalfshare
Population w/ Low Access at 1 Mile	lapop1
Percent w/ Low Access at 1 Mile	lapop1share
Population with Low Income and Low Access at 1 Mile	lalowi1
Percent with Low Income and Low Access at 1 Mile	lalowi1share
Children (ages 0-17) with Low Access at 1 Mile	lakids1
Percent Children (ages 0-17) with Low Access at 1 Mile	lakids1share
Seniors (age >= 65) with Low Access at 1 Mile	laseniors1
Percent Seniors (age >= 65) with Low Access at 1 Mile	laseniors1share

USDA Research Atlas Variable (continued)	Variable Name
White Population w/ Low Access at 1 Mile	lawhite1
Percent of White Population w/ Low Access at 1 Mile	lawhite1share
Black Population w/ Low Access at 1 Mile	lablack1
Percent of Black Population w/ Low Access at 1 Mile	lablack1share
Asian Population w/ Low Access at 1 Mile	laasian1
Percent of Asian Population w/ Low Access at 1 Mile	laasian1share
Native Hawaiian or Other Pacific Islander Population w/ Low Access at 1 Mile	lanhopi1
Percent of Native Hawaiian or Other Pacific Islander Population w/ Low Access at 1 Mile	lanhopi1share
American Indian or Alaska Native Population w/ Low Access at 1 Mile	laaian1
Percent of American Indian or Alaska Native Population w/ Low Access at 1 Mile	laaian1share
Other/Multiple Race Population w/ Low Access at 1 Mile	laomultir1
Percent of Other/Multiple Race Population w/ Low Access at 1 Mile	laomultir1share
Hispanic or Latino Population w/ Low Access at 1 Mile	lahisp1
Percent of Hispanic or Latino Population w/ Low Access at 1 Mile	lahisp1share
Population without Vehicle Access and Low Access at 1 Mile	lahunv1
Percent without Vehicle Access and Low Access at 1 Mile	lahunv1share
Population of SNAP Households w/ Low Access at 1 Mile	lasnap1
Percent of SNAP Households w/ Low Access at 1 Mile	lasnap1share
Population with Low Access at 10 Miles	lapop10
Percent Population with Low Access at 10 Miles	lapop10share
Population with Low Income and Low Access at 10 Miles	lalowi10
Percent Population with Low Income and Low Access at 10 Miles	lalowi10share
Children (ages 0-17) with Low Access at 10 Miles	lakids10
Percent Children (ages 0-17) with Low Access at 10 Miles	lakids10share
Seniors (age >= 65) with Low Access at 10 Miles	laseniors10
Percent Seniors (age >= 65) with Low Access at 10 Miles	laseniors10share
White Population w/ Low Access at 10 Miles	lawhite10
Percent of White Population w/ Low Access at 10 Miles	lawhite10share
Black Population w/ Low Access at 10 Miles	lablack10
Percent of Black Population w/ Low Access at 10 Miles	lablack10share
Asian Population w/ Low Access at 10 Miles	laasian10
Percent of Asian Population w/ Low Access at 10 Miles	laasian10share
Native Hawaiian or Other Pacific Islander Population w/ Low Access at 10 Miles	lanhopi10
Percent of Native Hawaiian or Other Pacific Islander Population w/ Low Access at 10 Miles	lanhopi10share
American Indian or Alaska Native Population w/ Low Access at 10 Miles	laaian10
Percent of American Indian or Alaska Native Population w/ Low Access at 10 Miles	laaian10share

USDA Research Atlas Variable (continued)	Variable Name
Native Hawaiian or Other Pacific Islander Population w/ Low Access at 10 Miles	lanhopi10
Percent of Native Hawaiian or Other Pacific Islander Population w/ Low Access at 10 Miles	lanhopi10share
American Indian or Alaska Native Population w/ Low Access at 10 Miles	laaian10
Percent of American Indian or Alaska Native Population w/ Low Access at 10 Miles	laaian10share
Other/Multiple Race Population w/ Low Access at 10 Miles	laomultir10
Percent of Other/Multiple Race Population w/ Low Access at 10 Miles	laomultir10share
Hispanic or Latino Population w/ Low Access at 10 Miles	lahisp10
Percent of Hispanic or Latino Population w/ Low Access at 10 Miles	lahisp10share
Population without Vehicle Access and Low Access at 10 Miles	lahunv10
Percent without Vehicle Access and Low Access at 10 Miles	lahunv10share
Population of SNAP Households w/ Low Access at 10 Miles	lasnap10
Percent of SNAP Households w/ Low Access at 10 Miles	lasnap10share
Population with Low Access at 20 Miles	lapop20
Percent Population with Low Access at 20 Miles	lapop20share
Population with Low Income and Low Access at 20 Miles	lalowi20
Percent with Low Income and Low Access at 20 Miles	lalowi20share
Children (ages 0-17) with Low Access at 20 Miles	lakids20
Percent Children (ages 0-17) with Low Access at 20 Miles	lakids20share
Seniors (age >= 65) with Low Access at 20 Miles	laseniors20
Percent Seniors (age >= 65) with Low Access at 20 Miles	laseniors20share
White Population w/ Low Access at 20 Miles	lawhite20
Percent of White Population w/ Low Access at 20 Miles	lawhite20share
Black Population w/ Low Access at 20 Miles	lablack20
Percent of Black Population w/ Low Access at 20 Miles	lablack20share
Asian Population w/ Low Access at 20 Miles	laasian20
Percent of Asian Population w/ Low Access at 20 Miles	laasian20share
Native Hawaiian or Other Pacific Islander Population w/ Low Access at 20 Miles	lanhopi20
Percent of Native Hawaiian or Other Pacific Islander Population w/ Low Access at 20 Miles	lanhopi20share
American Indian or Alaska Native Population w/ Low Access at 20 Miles	laaian20
Percent of American Indian or Alaska Native Population w/ Low Access at 20 Miles	laaian20share
Other/Multiple Race Population w/ Low Access at 20 Miles	laomultir20
Percent of Other/Multiple Race Population w/ Low Access at 20 Miles	laomultir20share
Hispanic or Latino Population w/ Low Access at 20 Miles	lahisp20
Percent of Hispanic or Latino Population w/ Low Access at 20 Miles	lahisp20share

USDA Research Atlas Variable (continued)	Variable Name
Population without Vehicle Access and Low Access at 20 Miles	lahunv20
Percent without Vehicle Access and Low Access at 20 Miles	lahunv20share
Population of SNAP Households w/ Low Access at 20 Miles	lasnap20
Percent of SNAP Households w/ Low Access at 20 Miles	lasnap20share
Population w/ Low Income	TractLowi
Population of Children (ages 0-17)	TractKids
Population of Seniors (age >= 65)	TractSeniors
White Population Count	TractWhite
Black Population Count	TractBlack
Asian Population Count	TractAsian
Native Hawaiian or Other Pacific Islander Population Count	TractNHOPI
American Indian or Alaska Native Population Count	TractAIAN
Other/Multiple Race Population Count	TractOMultir
Hispanic or Latino Population Count	TractHisp
Population w/out a Vehicle	TractHUNV
Population Receiving SNAP Benefits	TractSNAP

C. Variables from USDA's Food Environment Atlas (Years 1999-2014 Available)

Food Assistance			
Supplemental Nutrition Assistance Program (SNAP)			
Variable Description	Variable Name	Scale	Year
Available Online Application	SNAP_OAPP00	State	2000
Available Online Application	SNAP_OAPP05	State	2005
Available Online Application	SNAP_OAPP10	State	2010
Face Interview Waiver	SNAP_FACEWAIVER00	State	2000
Face Interview Waiver	SNAP_FACEWAIVER05	State	2005
Face Interview Waiver	SNAP_FACEWAIVER10	State	2010
Vehicle Exclusion Criteria	SNAP_VEHEXCL00	State	2000
Vehicle Exclusion Criteria	SNAP_VEHEXCL05	State	2005
Vehicle Exclusion Criteria	SNAP_VEHEXCL10	State	2010
Broad-Based Eligibility	SNAP_BBCE00	State	2000
Broad-Based Eligibility	SNAP_BBCE05	State	2005
Broad-Based Eligibility	SNAP_BBCE10	State	2010
Available Simplified Reporting	SNAP_REPORTSIMPLE00	State	2000
Available Simplified Reporting	SNAP_REPORTSIMPLE05	State	2005
Available Simplified Reporting	SNAP_REPORTSIMPLE10	State	2010
Redemptions/Authorized Store	REDEMP_SNAPS08	County	2008
Redemptions/Authorized Store	REDEMP_SNAPS12	County	2012
% Change in Redemptions	PCH_REDEMP_SNAPS_08_12	County	08-12
Benefits Per Capita	PC_SNAPBEN08	County	2008
Benefits Per Capita	PC_SNAPBEN10	County	2010
% Eligible Pop. Participating	SNAP_PART_RATE08	County	2008
% Eligible Pop. Participating	SNAP_PART_RATE10	County	2010
% Change Benefits Per Capita	PCH_PC_SNAPBEN_08_10	County	08-10
% of Population Participating	PCT_SNAP09	County	2009
% of Population Participating	PCT_SNAP14	County	2014
% Change of Participating Pop	PCH_SNAP_09_14	County	09-14
Women, Infants, and Children (WIC)			
Variable Description	Variable Name	Scale	Year
Redemptions/Authorized Stores	REDEMP_WICS08	County	2008
Redemptions/Authorized Stores	REDEMP_WICS12	County	2012
% Change in Redemptions	PCH_REDEMP_WICS_08_12	County	08-12
Redemptions per Capita	PC_WIC_REDEMP08	County	2008
Redemptions per Capita	PC_WIC_REDEMP12	County	2012
% Change Redemptions Per Capita	PCH_PC_WIC_REDEMP_08_12	County	08-12
% of Population Participating	PCT_WIC09	State	2009
% of Population Participating	PCT_WIC14	State	2014
% Change in Participating Pop.	PCH_WIC_09_14	State	09-14
WIC Participants - 2009	WIC_Participants_FY_2009	State	2009
WIC Participants - 2011	WIC_Participants_FY_2011	State	2011
WIC Participants - 2012	WIC_Participants_FY_2012	State	2012
WIC Participants - 2013	WIC_Participants_FY_2013	State	2013
WIC Participants - 2014	WIC_Participants_FY_2014	State	2014

Food Assistance (continued)			
National School Lunch Program (NSLP)			
Variable Description	Variable Name	Scale	Year
% Eligible for Free Lunch	PCT_FREE_LUNCH06	County	2006
% Eligible for Free Lunch	PCT_FREE_LUNCH10	County	2010
% Eligible Reduced-Price Lunch	PCT_REDUCED_LUNCH06	County	2006
% Eligible Reduced-Price Lunch	PCT_REDUCED_LUNCH10	County	2010
% Population in NSLP	PCT_NSLP09	State	2009
% Population in NSLP	PCT_NSLP14	State	2014
% Change in Participants	PCH_NSLP_09_14	State	09-14
# of NSLP Partipants	National_School_Lunch_Program_pa	State	2009
# of NSLP Partipants	National_School_Lunch_Program_p0	State	2011
# of NSLP Partipants	National_School_Lunch_Program_p1	State	2012
# of NSLP Partipants	National_School_Lunch_Program_p2	State	2013
# of NSLP Partipants	National_School_Lunch_Program_p3	State	2014
School Breakfast Program (SBP)			
Variable Description	Variable Name	Scale	Year
% Population in SBP	PCT_SBP09	County	2009
% Population in SBP	PCT_SBP14	County	2014
% Change in Participants	PCH_SBP_09_14	County	09-14
# of SBP Partipants	School_Breakfast_Program_partic	State	2009
# of SBP Partipants	School_Breakfast_Program_partic0	State	2011
# of SBP Partipants	School_Breakfast_Program_partic1	State	2012
# of SBP Partipants	School_Breakfast_Program_partic2	State	2013
# of SBP Partipants	School_Breakfast_Program_partic3	State	2014
Summer Food Service Program (SFSP)			
Variable Description	Variable Name	Scale	Year
% Population in SFSP	PCT_SFSP09	County	2009
% Population in SFSP	PCT_SFSP14	County	2014
% Change in Participants	PCH_SFSP_09_14	County	09-14
# of SFSP Participants	Summer_Food_participants_FY_2009	State	2009
# of SFSP Participants	Summer_Food_participants_FY_2011	State	2011
# of SFSP Participants	Summer_Food_participants_FY_201	State	2012
# of SFSP Participants	Summer_Food_participants_FY_200	State	2013
# of SFSP Participants	Summer_Food_participants_FY_202	State	2014
Child & Adult Care			
Variable Description	Variable Name	Scale	Year
% Pop. in Child & Adult Care	PCT_CACFP09	County	2009
% Pop. in Child & Adult Care	PCT_CACFP14	County	2014
% Change in Child & Adult Care	PCH_CACFP_09_14	County	09-14
# of Child & Adult Care Participants	Child_and_Adult_Care_participants	State	2009
# of Child & Adult Care Participants	Child_and_Adult_Care_participant0	State	2011
# of Child & Adult Care Participants	Child_and_Adult_Care_participant	State	2012
# of Child & Adult Care Participants	Child_and_Adult_Care_participan0	State	2013
# of Child & Adult Care Participants	Child_and_Adult_Care_participan1	State	2014
Food Distribution Program on Indian Reservations (FDPIR)			
Variable Description	Variable Name	Scale	Year
# of FDPIR Sites	FDPIR12	County	2012

Food Insecurity Measures			
Variable Description	Variable Name	Scale	Year
% of Food Insecure Households (three-year average)	FOODINSEC_00_02	State	00-02
% of Food Insecure Households (three-year average)	FOODINSEC_07_09	State	07-09
% of Food Insecure Households (three-year average)	FOODINSEC_10_12	State	10-12
% Change in Insecure Households (from 00-02 to 10-12)	CH_FOODINSEC_02_12	State	00-12
% Change in Insecure Households (from 07-09 to 10-12)	CH_FOODINSEC_09_12	State	07-12
% Households w/ Very Low Food Security (Three-Year average)	VLFOODSEC_00_02	State	00-02
% Households w/ Very Low Food Security (Three-Year average)	VLFOODSEC_07_09	State	07-09
% Households w/ Very Low Food Security (Three-Year average)	VLFOODSEC_10_12	State	10-12
% Change in Very Low Food Security Households (from 00-02 to 10-12)	CH_VLFOODINSEC_02_12	State	00-12
% Change in Very Low Food Security Households (from 07-09 to 10-12)	CH_VLFOODINSEC_09_12	State	07-12
% Households w/ Child Food Insecurity (multi-year average)	FOODINSEC_CHILD_01_07	State	01-07
% Households w/ Child Food Insecurity (multi-year average)	FOODINSEC_CHILD_03_11	State	03-11
Food Outlet Availability			
Fast Food Restaurants			
Variable Description	Variable Name	Scale	Year
# of Outlets	FFR07	County	2007
# of Outlets	FFR12	County	2012
# of Outlets/1,000 Pop	FFRPTH07	County	2007
# of Outlets/1,000 Pop	FFRPTH12	County	2012
% Change in # of Outlets	PCH_FFR_07_12	County	07-12
% Change /1,000 Pop	PCH_FFRPTH_07_12	County	07-12
Full-Service Restaurants			
# of Outlets	FSR07	County	2007
# of Outlets	FSR12	County	2012
# of Outlets/1,000 Pop	FSRPTH07	County	2007
# of Outlets/1,000 Pop	FSRPTH12	County	2012
% Change in # of Outlets	PCH_FSR_07_12	County	07-12
% Change /1,000 Pop	PCH_FSRPTH_07_12	County	07-12
Grocery Stores			
# of Stores	GROC07	County	2007
# of Stores	GROC12	County	2012
# of Stores/1,000 Pop	GROCPTH07	County	2007
# of Stores/1,000 Pop	GROCPTH12	County	2012
% Change in # of Stores	PCH_GROC_07_12	County	07-12
% Change /1,000 Pop	PCH_GROCPTH_07_12	County	07-12

Food Outlet Availability (continued)			
Supercenters			
Variable Description	Variable Name	Scale	Year
# of Stores	SUPERC07	County	2007
# of Stores	SUPERC12	County	2012
# of Stores/1,000 Pop	SUPERCPTH07	County	2007
# of Stores/1,000 Pop	SUPERCPTH12	County	2012
% Change in # of Stores	PCH_SUPERC_07_12	County	07-12
% Change /1,000 Pop	PCH_SUPERCPTH_07_12	County	07-12
Convenience Stores			
# of Stores	CONVS07	County	2007
# of Stores	CONVS12	County	2012
# of Stores/1,000 Pop	CONVSPTH07	County	2007
# of Stores/1,000 Pop	CONVSPTH12	County	2012
% Change in # of Stores	PCH_CONVS_07_12	County	07-12
% Change /1,000 Pop	PCH_CONVS_07_12	County	07-12
Specialized Food Stores			
# of Stores	SPECS07	County	2007
# of Stores	SPECS12	County	2012
# of Stores/1,000 Pop	SPECSPTH07	County	2007
# of Stores/1,000 Pop	SPECSPTH12	County	2012
% Change in # of Stores	PCH_SPECS_07_12	County	07-12
% Change /1,000 Pop	PCH_SPECSPTH_07_12	County	07-12
SNAP-Authorized Stores			
# of Stores	SNAPS08	County	2008
# of Stores	SNAPS12	County	2012
# of Stores/1,000 Pop	SNAPSPTH08	County	2008
# of Stores/1,000 Pop	SNAPSPTH12	County	2012
% Change in # of Stores	PCH_SNAPS_08_12	County	08-12
% Change /1,000 Pop	PCH_SNAPSPTH_08_12	County	08-12
WIC-Authorized Stores			
# of Stores	WICS08	County	2008
# of Stores	WICS12	County	2012
# of Stores/1,000 Pop	WICSPTH08	County	2008
# of Stores/1,000 Pop	WICSPTH12	County	2012
% Change in # of Stores	PCH_WICS_08_12	County	08-12
% Change /1,000 Pop	PCH_WICSPTH_08_12	County	08-12
Restaurant Expenditures			
Fast Food Restaurants			
Spending Per Capita	PC_FFRSALES02	County	2002
Spending Per Capita	PC_FFRSALES07	County	2007
Full-Service Restaurants			
Spending Per Capita	PC_FSRSALES02	County	2002
Spending Per Capita	PC_FSRSALES07	County	2007

Food Prices & Taxes			
Prices			
Variable Description	Variable Name	Scale	Year
Low-Fat Milk Price/National Avg	MILK_PRICE10	County	2010
Soda Price/National Avg.	SODA_PRICE10	County	2010
Low-Fat Milk Price/Soda Price	MILK_SODA_PRICE10	County	2010
Sales Tax			
Soda Sales Tax, Retail Stores	SODATAX_STORES11	County	2011
Soda Sales Tax, Vending	SODATAX_VENDM11	County	2011
Chip Sales Tax, Retail Stores	CHIPSTAX_STORES11	County	2011
Chip Sales Tax, Vending	CHIPSTAX_VENDM11	County	2011
General Food Tax, Retail Stores	FOOD_TAX11	County	2011
Local Foods			
Direct Sales From Farms			
# of Farms w/ Direct Sales	DIRSALES_FARMS07	County	2007
% of Farms w/ Direct Sales	PCT_LOCLFARM07	County	2007
Direct Farm Sales, Percent	PCT_LOCLSALE07	County	2007
Direct Farm Sales	DIRSALES07	County	2007
Direct Farm Sales, Per Capita	PC_DIRSALES07	County	2007
Farmers' Markets			
# of Outlets	FMRKT09	County	2009
# of Outlets	FMRKT13	County	2013
% Change in Number of Outlets	PCH_FMRKT_09_13	County	09-13
# of Outlets/ 1,000 pop.	FMRKTPTH09	County	2009
# of Outlets/ 1,000 pop	FMRKTPTH13	County	2013
% Change / 1,000 pop.	PCH_FMRKTPTH_09_13	County	09-13
# of Markets Accepting SNAP	FMRKT_SNAP13	County	2013
% of Markets Accepting SNAP	PCT_FMRKT_SNAP13	County	2013
# of Markets Accepting WIC	FMRKT_WIC13	County	2013
% of Markets Accepting WIC	PCT_FMRKT_WIC13	County	2013
# of Markets Accepting WIC Cash	FMRKT_WICCASH13	County	2013
% of Markets Accepting WIC Cash	PCT_FMRKT_WICCASH13	County	2013
# of Markets Accepting SFMNP	FMRKT_SFMNP13	County	2013
% of Markets Accepting SFMNP	PCT_FMRKT_SFMNP13	County	2013
# of Markets w/ Fruits & Vege.	FMRKT_FRVEG13	County	2013
% of Markets w/ Fruits & Vege.	PCT_FRMKT_FRVEG13	County	2013
# of Markets w/ Animal Products	FMRKT_ANMLPROD13	County	2013
% of Markets w/ Animal Products	PCT_FRMKT_ANMLPROD13	County	2013
# of Markets w/ Other Products	FMRKT_OTHER13	County	2013
% of Markets w/ Other Products	PCT_FMRKT_OTHER13	County	2013

Local Foods (continued)			
Prices			
Variable Description	Variable Name	Scale	Year
Farms & Acreage for Produce			
# of Vegetable Farms	VEG_FARMS07	County	2007
Harvested Vegetable Acres	VEG_ACRES07	County	2007
Harvested Vegetable Acres / 1,000 Pop.	VEG_ACRESPTH07	County	2007
# of Farms w/ Vegetables for Fresh Markets	FRESHVEG_FARM07	County	2007
Harvested Vegetables Acres for Fresh Markets	FRESHVEG_ACRES07	County	2007
Harvested Vegetable Acres for Fresh Markets / 1,000 Pop.	FRESHVEG_ACRESPTH07	County	2007
# of Orchard Farms	ORCHARD_FARMS07	County	2007
Orchard Acres	ORCHARD_ACRES07	County	2007
Orchard Acres/ 1,000 Pop.	ORCHARD_ACRESPTH07	County	2007
# of Berry Farms	BERRY_FARMS07	County	2007
Berry Acres	BERRY_ACRESPTH07	County	2007
Berry Acres/ 1,000 Pop.	BERRY_ACRESPTH07	County	2007
# of Fresh Herb & Greenhouse Vegetable Farms	GHVEG_FARMS07	County	2007
Square Feet of Fresh Herb & Greenhouse Vegetable Farms	GHVEG_SQFT07	County	2007
Square Feet of Fresh Herb & Greenhouse Vegetable Farms / 1,000 Pop.	GHVEG_SQFTPTH07	County	2007
Other Local Food Resources			
# of Small Slaughterhouses	SLHOUSE07	County	2007
# of CSA Farms	CSA07	County	2007
# of Agritourism Operations	AGRITRSM_OPS07	County	2007
Total Agritourism Receipts	AGRITRSM_RCT07	County	2007
Farm-to-School Program	FARM_TO_SCHOOL	County	2009
Food Hub Availability	FOODHUB12	County	2012
Health			
% Obesity in Low-Income Preschoolers (ages 2-4) (two-year aggregate)	PCT_OBESE_CHILD08	County	06-08
% Obesity in Low-Income Preschoolers (ages 2-4) (two-year aggregate)	PCT_OBESE_CHILD11	County	09-11
% Change in Obesity in Low-Income Preschoolers	PCH_OBESE_CHILD_08_11	County	06-11
% of High-Schoolers who are Physically Active	PCT_HSPA09	State	2009
% of Adults Considered Obese	PCT_OBESE_ADULTS09	County	2009
% of Adults Considered Obese	PCT_OBESE_ADULTS10	County	2010
% of Adults Considered Obese	PCT_OBESE_ADULTS13	State	2013
% of Adults w/ Diabetes	PCT_DIABETES_ADULTS09	County	2009
% of Adults w/ Diabetes	PCT_DIABETES_ADULTS10	County	2010
Recreational & Fitness Facility Availability			
# of Facilities	RECFAC07	County	2007
# of Facilities	RECFAC12	County	2012
# of Facilities/ 1,000 Pop.	RECFACPTH07	County	2007
# of Facilities/ 1,000 Pop	RECFACPTH12	County	2012
% Change in Facilities	PCH_RECFAC_07_12	County	07-12
% Change / 1,000 Pop.	PCH_RECFACPTH_07_12	County	07-12
ERS Natural Amenities Index	NATAMEN	County	1999

Sociodemographics			
Variable Description	Variable Name	Scale	Year
Population Estimate - 2007	Population_Estimate__2007	County	2007
Population Estimate - 2008	Population_Estimate__2008	County	2008
Population Estimate - 2009	Population_Estimate__2009	County	2009
Population Estimate - 2011	Population_Estimate__2011	County	2011
Population Estimate - 2012	Population_Estimate__2012	County	2012
State Population Estimate - 2008	State_Population__2008	State	2008
State Population Estimate - 2009	State_Population__2009	State	2009
State Population Estimate - 2010	State_Population__2010	State	2010
State Population Estimate - 2011	State_Population__2011	State	2011
State Population Estimate - 2012	State_Population__2012	State	2012
State Population Estimate - 2013	State_Population__2013	State	2013
State Population Estimate - 2014	State_Population__2014	State	2014
2010 Census Population	_010_Census_Population	County	2010
% White	PCT_NHWHITE10	County	2010
% Black	PCT_NHBLACK10	County	2010
% Hispanic	PCT_HISP10	County	2010
% Asian	PCT_NHASIAN10	County	2010
% American Indian or Alaska Native	PCT_NHNA10	County	2010
% Hawaiian or Pacific Islander	PCT_NHPI10	County	2010
% 65 or Older	PCT_65OLDER10	County	2010
% Under 18 Years of Age	PCT_18YOUNGER10	County	2010
Median Household Income	MEDHHINC10	County	2010
% Pop. In Poverty	POVRATE10	County	2010
Counties w/ Persistent Poverty	PERPOV10	County	2010
% of Households w/ Children in Poverty	CHILDPOVRATE10	County	2010
Counties w/ Persistent Child Poverty	PERCHLDPOV10	County	2010

VI. Technical Notes on Constructing the Data Module

In this section we provide a more detailed account of each step we took to construct the FAM data module. Instructions provided to our research analysts and research assistants as well as example syntax are included at the end of this section and in the appendices. Our hope is that this information will be useful to researchers who are seeking additional information on how FAM variables were created and/or who are interested in creating a similar type of data module with their own data.

Ordering the NETS database

In 2016, NETS data for the years 2000-2013 was purchased through Donald Walls, President of Walls & Associates. Establishment selection was done primarily through choosing relevant 8-digit Standard Industrial Classification (SIC) codes. These codes are not only more descriptive than the typical 6-digit codes but have also remained consistent since 1992. In some cases we used the 3- or 4-digit SIC codes when we wanted to include all establishments under a more general category (e.g., Meat and Fish Markets, SIC3 = 542).

The initial NETS database delivery contained three large ASCII files and included 2.5 million establishments with a little over 1 million establishments still active in 2013. The NETS2013.txt (3,840,595,010 KB) file contained all of the primary information about the establishments, including company name, trade name, address, SIC codes, and many other variables. The NAICS2013.txt (225,306,114 KB) file contained the relevant NAICS codes for each establishment for every year. The third file was MOVES2013.txt (80,614,035 KB), which consisted of address changes that occurred from 1990 to 2013.

In 2017, the research team added 2014 establishment data to the module once it was available.

NETS Data Cleaning Process

After retrieving the NETS data, the ASCII files were transferred into SAS datasets to begin the data cleaning process. Because the license for using the NETS database was for two years, all cleaning processes were completed within this timeframe.

Creating Yearly NETS Databases

The original NETS file contains establishment data from 1990 to 2013. We created separate NETS files containing food establishments for each year. We used the FirstYear and LastYear variables to determine which establishments existed during a particular year. The main NETS and NAICS data files were merged using the Dunsnumber variable, which is the common variable between these files.

Using the Moves.txt file, the most relevant business relocations were joined with the yearly NETS datasets. The purpose of this step was to ensure that the appropriate coordinates were assigned to each establishment for a specific year. This process was performed on historical data

that incorporated any years prior to 2013 from the original data delivery or 2014 upon the second data delivery.

Removing Unneeded Variables from the NETS

The original NETS document contained over one million businesses with over 350 variables. Multiple fields were removed from the NETS file to reduce the amount of computing time needed for analysis or cleaning. The remaining variables included the NETS identifier variable, the Dunsnumber, company name, tradename, address, zip code, city, state, SIC codes, first year, last year, industry, latitude, longitude, level code, sales, employment size, and headquarters information. Establishments that were last active prior to 2000 were also removed from the dataset.

Deleting Headquarters Data

Similar to other researchers who have worked with NETS data, any establishment that represented a headquarters location was removed because these locations typically do not sell food products to the general public. An establishment was considered a headquarters location if the headquarters identifier (HQDuns) was listed multiple times in the database and possessed a Dunsnumber that matched the HQDuns number. The goal was to remove headquarter locations and avoid removing companies that may only have a single location that provides food services to the public (N. Colabianchi, personal communication, July 18, 2016).

Geocode Assignments for Historical NETS Data

The NETS data delivery included geocoded latitudes and longitudes for every business location. The geocoding was completed by Dun & Bradstreet using the North American Datum of 1983 (NAD83) coordinate reference system. The main NETS Database lists the most current coordinates for each business location while the supplementary NETS Moves file catalogs geocodes of businesses that have relocated. Only "significant moves" were recorded in the Moves data file, which included any relocation in which the 5-digit zipcode had changed (Walls & Associates, 2013). The level code was updated to represent a significant move. For additional information on the level code, which is a variable that represents geocode spatial accuracy, see Geocoding Accuracy of Food Establishment Locations described further down in this section.

Duplicate Detection and Removal

Initial inspection of the NETS database for each year suggested that there were no Dunsnumber duplicates, yet closer analysis indicated that there were duplicates throughout the dataset. Establishments may be listed multiple times with different spelling variations in the company name, tradename, or address. The first step taken to remove the duplicates involved determining which establishments had matching geocoded latitudes and longitudes (K. Moore, personal communication, September 14, 2016).

After determining which establishments contained matching geocodes, we created a separate dataset representing potential duplicates. The first step in duplicate extraction was using SAS code to standardize the addresses to minimize address spelling variation. However, there were still multiple variations in the way a single address was represented, and so we standardized establishments' addresses. Sample code of this process is provided in Appendix C. The SAS address standardization codes used in this project replicated syntax by Sata Hackenbruck. This syntax is available through the following link: http://www.sascommunity.org/wiki/Standardizing_Addresses_in_SAS_for_Geocoding (2017).

Once addresses were standardized, we used SAS code to parse out potential duplicates with the same coordinates and matching company name, tradename, and addresses using multiple spelling variation criteria for each of these variables. We adopted a similar methodology used by researchers who linked Dun & Bradstreet geospatial data with participant information for the MESA Neighborhood Study and the Jackson Heart Study. We used the SPEDIS function to extract both exact matches and nearly identical observations (K. Moore, personal communication, September 14, 2016). The SPEDIS command in SAS calculates the spelling distances between a query and a keyword of interest. For more information about this SAS argument, please refer to http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000245949.htm (Statistical Analysis Software (SAS), 2011).

Two establishments with SPEDIS values of ≤ 15 for both the address and company name were automatically considered matching establishments and were removed. Spelling distances of 16-45 were verified manually (K. Moore, personal communication, September 14, 2016). See Appendix C for an example of using the SPEDIS command to extract duplicates.

Ground-Truthing Procedure

Verification of duplicate establishments began by developing Excel documents that met the coordinate and SPEDIS spelling distance criteria of 16-45. These spreadsheets were formatted with each row representing two businesses that represented potential duplicates. In the first step of this process, research assistants went through these spreadsheets and either left the rows unedited to indicate that the establishments in a row were duplicates based on evident similarity between establishments, or highlighted the row to indicate that further ground-truthing was needed to confirm duplicate status.

In the second step, a team of research assistants used Google Street View to ground-truth the highlighted establishments. Businesses with geocoding at the zip-code level (the lowest spatial accuracy) were excluded from our ground-truthing process. Cases that could not be identified as duplicates were kept in the final dataset as separate establishments. In general, we were conservative in our removal of duplicates.

The ground-truthing procedure is detailed below:

1. An additional column was added to the potential duplicate spreadsheet called, "RAdup," which allowed each potential duplicate match to be assigned a numeric code that indicated whether these matches were duplicates. These numeric codes were used later in SAS to help identify and remove duplicate establishment.
2. RAs entered the address shared between two businesses into the Google search bar. Generally, one or both of the company names would appear along with the address. If the Google Street view option did not automatically appear, the address was manually pasted into Google Street View.
3. The time-scale was changed in the top-left corner of Google Street View to match the exact year of interest when the option was available. If the exact year was not an option, the "firstyr" and “lastyr" variables in the NETS data were used to inform the most appropriate years for these businesses. A "lastyr" of 2013 or 2014 indicated that the business was still open.
4. If the Street View image indicated that only one business was present at that location and no other establishment appeared in the nearby area, then the two establishments were considered to be duplicates. The Excel spreadsheet containing the potential duplicates was highlighted green and the business name was bolded to indicate this finding. A '2' was entered into the RAdup column.
5. If the location did not appear to be an actual food distributor/supplier to the general public (i.e. a corporate office, home office, or an industrial warehouse space), the text in the row was highlighted red and a 3 was entered into the RAdup column to indicate that these establishments should be removed from the final dataset.
6. If neither of the listed business names appeared in the Street View window, but another relevant food supplier was detected in close proximity, the text for that entry in the Excel file was changed to blue and a 4 was listed under RAdup, which signified that only one food outlet exists at that location. If the two businesses possessed matching SIC codes, a single entry was removed from the main NETS data file. However, if the SIC codes did not match, both establishments remained in the dataset as it could not be determined which food outlet needed to be deleted. This scenario rarely happened.
7. Two businesses were not considered duplicates if both company names were evident at the location of interest. These rows were highlighted blue and labeled a 5 in the RAdup column to indicate two distinct businesses.
8. Circumstances in which the research assistant was unable to determine whether two businesses were actual duplicates due to obstructions within Street View (e.g., a view for that location was unavailable for a particular year, the establishment was located within a building, such as a mall or other large shopping center), a 6 was entered in the RAdup column. This numeric code indicated that duplicate status could not be verified and both establishments remained in the dataset.
9. If the establishment was no longer a food establishment for a specific year, the spreadsheet row was highlighted orange and a 7 was entered in the RAdup column. These entries were removed from the final database.

Food Establishment Categories

Standard Industrial Classification (SIC) Codes in the NETS Database

Data from the NETS contained industrial classification assignments for each establishment for each year a business was active. The team primarily used Standard Industrial Classification (SIC) codes to identify food establishments. In some instances the company and tradename were used to determine the proper food category (e.g. fast food restaurants). The SIC code assigned for a specific year was often used to define an establishment except when an outlet was not allocated a SIC designation for that year. For example, if a business became active in 2009 and the geospatial coordinates were available in 2009, NETS did not provide a SIC09 for that outlet. In such cases, the SIC code for the following year was used to define that establishment. Syntax for creating Food Establishment Categories is in Appendix D (available under restricted data contract).

Creation of GIS-Based Variables

Panel Study of Income Dynamics Family Locations

PSID family geolocations were drawn from the restricted PSID data file containing numeric codes pinpointing family locations at the block-group level during the years that the main PSID interview was conducted (2001, 2003, 2005, 2007, 2009, 2011, and 2013). Census block-groups are the smallest geographic subdivisions used by the U.S. Census Bureau to extract decennial statistics. These divisions generally include 600 to 3,000 individuals and are enveloped within census tracts (U.S. Census Bureau, 2012). The information available through the PSID's restricted file was converted to point feature classes in ArcMap by merging the PSID file with the "Centers of Population by Block Group: 2010".csv file from the United States Census Bureau's website. This .csv file assigned block-group centroid coordinates based on a specific area where most residents are located (U.S. Census Bureau, 2015). This block-group file was joined with the restricted PSID geospatial information using block-group values as the common variable to link each PSID family to a specific coordinate. The 12-digit block-group values were manually created by concatenating the state, county, tract, and block-group codes.

Projection Selection

The geographic calculations in this module were calculated after re-projecting all of the geographic data to the USA Contiguous Albers Equal Area Projection for PSID families living within the continental United States. The motivation to use this specific projection was to reduce the geographic distortion. This projection is often used for maps of the continental United States because it is well-suited for geographic regions with east-west orientation that is located within the mid-latitudes. The primary benefit of this projection that is most relevant to creating FAM is its ability to preserve most distance accuracy and minimize distortion in the mid-latitudes (ESRI, 2000). PSID families located outside of the continental United States were re-projected according to the most appropriate projection with respect to location to minimize distortion.

Preparing NETS Files for ArcMap Software

The instructions below detail the steps to convert the NETS and PSID family location data files into ArcMap point features.

Instructions:

Converting The NETS & PSID Data Files into ArcMap Shapefiles

Convert the stored longitudes into negative values.

1. Ensure that the NETS database has been exported into a .csv file.
2. In ArcMap, use ArcCatelog to connect to the file folder where the NETS .csv is located.
3. Within the ArcCatelog window, right-click the .csv file and select **Create Feature Class**
4. Click on **From XY Table**.
5. Under the **X Field** drop-down box select Longitude.
6. Under the **Y Field** drop-down box select Latitude.
7. Click on **Coordinate System of Input Coordinates** box.
8. Select NAD 1983 as the XY Coordinate System.
9. Keep all other defaults and click **OK**.
10. Repeat steps 3-10 for the merged PSID/Blockgroup Centroid file.
11. Use the **Project** tool to reproject the shapefiles into the **Albers Equal Area** projection.

Geocoding Accuracy of Food Establishment Locations

The NETS dataset possessed variables representing geocoded latitudes and longitudes for each establishment along with a "LevelCode" variable indicating the geospatial accuracy of these code assignments. The table below represents the various Level Code definitions.

Code	Level Category	Definition
D	Block Face	Criteria for exact address met.
S	Street Segment	Placed on correct street with accuracy within .1 to .2 miles.
B	Block-Group	Placed on correct street with accuracy within .2 to .4 miles.
T	Census Tract	Placed on street or block according to address range (.4 to .6 mile accuracy.)
Z	ZIP Code	Establishment coded at ZIP Code Level. (Lowest level of accuracy).

These codes are essential because the geographic level that the establishment was coded depended on the years the establishment was active (Walls & Associates, 2013). This variability in geospatial accuracy, its dependence on the years the establishments were active, and potential establishment moves between1990-2014 were one of the reasons that it was necessary to create a NETS data file representing only a single year of data.

Calculating Food Outlet Densities

In this module, density was defined using various distances of both Euclidean-based and Network-based buffers. A review of the current food environment literature prompted the use of four specific distances for this project's measurements (1/2-, 1-, 5-, and 10- mile areas) to calculate the total number of food establishments from each category that exist around each PSID family unit at the block-group level.

Instructions:

Euclidean Density Calculations

1. Ensure that the extension **Spatial Analyst** is activated. To activate this extension:
    a. Select **Customize > Extensions.**
    b. Check the **Spatial Analyst** box in the **Extensions** menu.

2. Within **ArcToolbox** navigate to the **Buffer** tool by clicking **Analysis Tools > Proximity > Buffer**

3. Select the PSID blockgroup centroid shapefiles using the drop-down list under **Input Features**

4. Navigate to the desired folder or geodatabase where the file should be saved using the **Output Feature Class** section.

5. Type the desired distance under **Distance** and ensure that the proper unit (Miles) is selected.

6. Keep all of the default selections with all other drop-down menus within this tool and click **OK**.

7. Repeat steps 2-6 until all applicable buffers are created.

8. To create tables with total food outlet counts, use the **Tabulate Intersection** tool by clicking **Analysis Tools > Statistics > Tabulate Intersection** within the **ArcToolbox**.

9. Under **Input Zone Features,** select the buffer of interest. In addition, select **FAMID10** to ensure that these counts can be linked to a PSID family.

10. Select one of the food establishment shapefiles from the drop-down list under **Input Class Features.**

11. Within **Output Table**, make sure to name the .dbf table and navigate to the appropriate folder.

12. Click **OK**.

13. Repeat steps 8-12 until all the relevant counts have been calculated for each distance and food establishment of interest.

*** Note:** To make steps 8-12 more efficient, use **ModelBuilder** or **Python** code to help streamline this process.

* Sample Python Code to Calculate Euclidean Density are in Appendices E and F.

Network-Based Density Calculations

1. Ensure that the extension **Network Analyst** is activated. To activate this extension:
    a. Select **Customize > Extensions.**
    b. Check the **Network Analyst** box in the **Extensions** menu.

2. Add the relevant *Streets* Network dataset to the ArcMap document.

3. Within the **Network Analyst** toolbar in the drop-down menu, select **New Service Area**.

4. Click on the box icon next to the drop-down list. Select the tab **Network Locations** and check the box **Exclude Restricted Portions of the Network**. Then, under **Analysis Settings,** change **Impedance** to **Length (Miles)** and type in the desired distance next to **Default Breaks**. Uncheck **Use Hierarchy**. Click **OK**.

*** Note:** There are a few PSID centroids that cannot be located onto the network dataset. In these cases, it may be necessary to reload these specific centroids separately and unchecking the box next to **Non-routeable Segments** under **Restrictions** in **Analysis Settings**.

5. Right-click **Facilities** and select **Load Locations**.

6. Select the PSID block-group shapefile from the **Load From** list. Also, under **Locational Analysis Properties**, ensure to select **FAMID10** under **Name** so that these densities can be linked to a specific PSID family. Keep Default **Location Positions** and click **OK.**

7. Click the **Solve** icon within the **Network Analyst** toolbar.

8. After the network buffer polygons have been created, right-click **Polygons** and select **Export Data** to create a permanent shapefile. Once the shapefile is created, you may right-click on **Polygons**, click **Delete All.**

9. Repeat steps 4-9 until all of the density shapefiles have been created.

*** Note:** Depending on the computing capabilities at the workstation, this process may be quite time-consuming and might require the analyst to divide the PSID participant shapefiles into smaller subsets to prevent ArcMap from crashing.

10. To create tables with total food outlet counts, use the **Tabulate Intersection** tool by clicking **Analysis Tools > Statistics > Tabulate Intersection** within the **ArcToolbox**.

11. Under **Input Zone Features,** select the network buffer of interest. In addition, select **Name** to ensure that these counts can be linked to a PSID family.

12. Select one of the food establishment shapefiles from the drop-down list under **Input Class Features.**

13. Within **Output Table**, make sure to name the .dbf table and navigate to the appropriate folder.

14. Click **OK**.

15. Repeat steps 8-12 until all the relevant counts have been calculated for all food outlets.

Calculating Food Outlet Distances

Euclidean (straight-line) and network distances to the closest food establishments for all food establishment categories were computed from PSID population-centered block-group centroids. This module used two ESRI StreetMap files to calculate network distances: the StreetMap North America datasets from ESRI Data & Maps 9.3 and ESRI Data & Maps 10.1. The network file from version 9.3 was used to calculate road network distances for years 2000-2004 while the StreetMap NA file from 10.1 was utilized for any year after 2004. The decision to use these different street networks during specific years was based on the metadata of both network datasets. The StreetMap data origin and the temporal timeframe represented by these files are presented in the table below (ESRI, 2008; ESRI, 2012). During the time that this module was created, StreetMap North America from ESRI Data & Maps 10.1 was the most current file available to the research team. These two versions of Data & Maps were obtained through the University of Michigan's Spatial and Numeric Data section at the Clark library.

ESRI Data & Maps StreetMap NA Version	Road Network Temporal Timeframe	StreetMap Data Origin
9.3	2003	TeleAtlas
10.1	2007	TomTom

Instructions:

Euclidean Distance Calculations to Closest Facility

1. Ensure that the extension **Spatial Analyst** is activated. To activate this extension:
    a. Select **Customize > Extensions.**
    b. Check the **Spatial Analyst** box in the **Extensions** menu.

2. Within **ArcToolbox** navigate to the **Generate Near Table** tool by clicking **Analysis Tools > Proximity > Generate Near Table.**

3. Select the PSID blockgroup centroid shapefiles using the drop-down list under **Input Features.**

4. Only select one food outlet category within the **Near Features** section.

5. Navigate to the desired folder or geodatabase where the file should be saved using the **Output Table** section.

6. Keep default selections click **OK**.

7. Repeat steps 2-6 until near calculation tables have been created for all food categories.

* Sample Python Code to Calculate Euclidean Distances are in Appendix G.

Instructions:

Network-Based Distance Calculations

1. Ensure that the extension **Network Analyst** is activated. To activate this extension:
    a. Select **Customize > Extensions.**
    b. Check the **Network Analyst** box in the **Extensions** menu.

2. Add the relevant *Streets* Network dataset to the ArcMap document.

3. Within the **Network Analyst** toolbar in the drop-down menu, select **New OD Cost Matrix**

4. Click on the box icon next to the drop-down list. Select the tab **Network Locations** and check the box **Exclude Restricted Portions of the Network**. Then, under **Analysis Settings,** change **Impedance** to **Length (Miles)** and ensure that **Facilities to Find** is equal to 1. Keep all other default settings and click **OK**.

*** Note:** There are a few PSID centroids that cannot be located onto the network dataset. In these cases, it may be necessary to reload these specific centroids separately. When reloading these shapefiles, uncheck the box **Exclude Restricted Portions of the Network** mentioned in the previous step. Also, uncheck the box next to **Non-routeable Segments** under **Restrictions** in **Analysis Settings**.

5. Right-click **Facilities** and select **Load Locations**.

6. Select the one food category shapefile from the **Load From** list. Keep Default **Location Positions** and click **OK.**

7. Right-click **Incidents** and select **Load Locations.**

8. Select the PSID family block-group centroid shapefile from the **Load From** list. Also, under **Locational Analysis Properties**, ensure to select **FAMID10** under **Name** so that these densities can be linked to a specific PSID family. Keep Default **Location Positions** and click **OK.**

9. Click the **Solve** icon within the **Network Analyst** toolbar.

10. After all routes have been created, right-click **Routes** and click on **Open Attribute Table.**

11. Within the drop-down list at the top of the attribute table, select **Export**. Click on the folder icon next to **Output Table**, navigate to the folder you would like to save this table, and save the table as a text, .csv file. Click **Ok**.

12. Repeat steps 4-9 until all distances of interest have been calculated.

*** Note:** Depending on the computing capabilities of the computer at the workstation, this process may, not only be time-consuming, but could also cause the ArcMap software to crash. For example, in cases in which there are over 100,000 facilities within a food outlet category, such as fast food restaurants, it might be necessary to divide the PSID families into smaller subsets to allow Network Analyst to run properly.

* Sample Python Code to Calculate Network Distances are in Appendix H.

VII. References

Auchincloss, A. H. et al. (2012). Improving retrospective characterization of the food environment for a large region in the United States during a historical time period. *Health & Place, 18*(6), 1341-1347.

Child Development Supplement, Panel Study of Income Dynamics. (2008). *The Panel Study of Income Dynamics' Child Development Supplement Transition into Adulthood 2005: User Guide.* Institute for Social Research, University of Michigan, Ann Arbor, MI.

Carolina Population Center [CPC], University of North Carolina (n.d.). *OneData – Obesity & Neighborhood Database*. Retrieved November 2016, from http://www.cpc.unc.edu/projects/onedata

Economic Research Service [ERS], U.S. Department of Agriculture (USDA). (2016). *Food Environment Atlas* [Data file]. https://www.ers.usda.gov/data-products/food-environment-atlas/

ERS. (2017). *Food Access Research Atlas* [Data file]. https://www.ers.usda.gov/data-products/food-access- research-atlas/

Environmental Systems Research Institute [ESRI]. (2000). *Understanding Map Projections.* ESRI, Inc., Redlands, CA, 37.

ESRI. (2008). *ArcGIS Release 9.3*. Redlands, CA.

ESRI. (2012). *ArcGIS Release 10.1*. Redlands, CA.

Fleischhacker, S. E., et al. (2012). Evidence for validity of five secondary data sources for enumerating retail food outlets in seven American Indian communities in North Carolina. *International Journal of Behavioral Nutrition and Physical Activity, 9*(137).

Fox, K. (2015). Ethnic restaurants. In K. Albala (Ed.), *The SAGE Encyclopedia of Food Issues, 3*, 414-418. Thousand Oaks, CA: SAGE Publications Ltd. doi: 10.4135/9781483346304.n138.

Gibson, D. M. (2011). The neighborhood food environment and adult weight status: estimates from longitudinal data. *American Journal of Public Health, 101*(1), 71-78.

Gordon-Larson, L. (2009). *National Longitudinal Study of Adolescent Health: Obesity and Neighborhood Environment* (User guide). University of North Carolina at Chapel Hill, NC.

Hackenbruck, S. (n.d.). *Standardizing Addresses in SAS for Geocoding*. Retrieved September 2016 from http://www.sascommunity.org/wiki/Standardizing_Addresses_in_SAS_for_Geocoding

Jilcott, S. B. et al (2010). The association between the food environment and weight status among eastern North Carolina youth. *Public Health Nutrition, 14*(9), 1610-1617.

Lee, H. (2012). The role of local food availability in explaining obesity risk among young school-aged children. *Social Science & Medicine, 74*(8), 1193-1203.

Liese, A. et al. (2007). Food store types, availability, and cost of foods in a rural environment. *Journal of the American Dietetic Association, 107*(11), 1916-1923.

Liese, A. D. et al. (2010). Validation of 3 food outlets databases: completeness and geospatial accuracy in rural and urban food environments. *American Journal of Epidemiology, 172*(11), 1324-1333.

Martz, M. et al. (2014). Pediatric obesity and food access in Durham, North Carolina. *International Journal of Child Health and Human Development, 7*(3), 267-275.

McGonagle, K. A., Schoeni, R. F., Sastry, N., and Freedman, V. A. (2012). The Panel Study of Income Dynamics: Overview, recent innovations, and potential for life course research. *Longitudinal and Life Course Studies 2012, 3*(2), 268-284.

Moore, L. V. & Diez Rouz, A. V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *Research and Practice, 96*(2), 325-331.

Moore, K. et al (2013). Home and work neighborhood environments in relation to body mass index: the Multi-Ethnic Study of Atherosclerosis (MESA). *Journal of Epidemiology and Community Health, 67*(10), 846-853.

Morland, K. et al. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventative Medicine, 22*(1), 23-29.

Nation's Restaurant News (NRN). (2010, June 28). *Top 100 Chains Ranked by U.S. Systemwide Foodservice Sales*. Retrieved October 2016 from http://go.galegroup.com.proxy.lib.umich.edu/ps/i.do?p=ITOF&u=lom_umichanna&id=GALE|A230958006&v=2.1&it=r&sid=summon

NRN. (2013, June 24). *Top 100 U.S. Chain Systemwide Food Services Sales.* Retrieved October 2016 from http://go.galegroup.com.proxy.lib.umich.edu/ps/i.do?p=ITOF&u=lom_umichanna&id=GALE|A374101938&v=2.1&it=r&sid=summon&authCount=1

Panel Study of Income Dynamics [PSID]. (2003). *Panel Study of Income Dynamics: 2003 Public Release Family File Codebook.* Institute for Social Research, University of Michigan, Ann Arbor, MI.

PSID. (2016, May). *PSID File Structure and Merging PSID Data Files.* Institute for Social Research, University of Michigan, Ann Arbor, MI.

PSID. (2017, May). *PSID Main Interview User Manual: Release 2017.* Institute for Social Research, University of Michigan, Ann Arbor, MI. Retrieved from https://psidonline.isr.umich.edu/data/Documentation/UserGuide2015.pdf

PSID. (2017, June). *1968-2015: Geocode File – Census Tract, Block-Group and Block Codebook.* Institute for Social Research, University of Michigan, Ann Arbor, MI.

Ploeg, M. V., Breneman, V., Dutko, P., Williams, R., Snyder, S., Dicken, C., & Kaufman, P. (2012). Access to affordable and nutritious Food: Updated estimates of distance to supermarkets using 2010 data. *U.S. Department of Agriculture, Economic Research Service* (November), 15. Retrieved from http://www.ers.usda.gov/publications/err-economic-research- report/err143.aspx#.Ut7xXZGtu-k

Powell, L. & Yanjun, B. (2009). Food prices, access to food outlets and child weight. *Economics and Human Biology, 7*(1), 64-72.

Restaurants & Institutions (R&I). (2009, July 15). *R&I 2009 Top 400 Restaurant Chains.* Retrieved October 2016 from https://www.rolypoly.com/news/articles/R&I%202009%20Top%20400%20Restaurant%20Chains.pdf

Rundle et al. (2009). Neighborhood food environment and walkability predict obesity in New York City. *Environmental Health Perspectives, 117*(3). 442-447.

Rundle, A. G., Bader, M. D.M., Richards, C. A., Neckerman, K. M., & Teitler, J. O. (2011). Using Google Street View to audit neighborhood environments. *American Journal of Preventative Medicine*. *40*(1). 94-100.

Transition into Adulthood Supplement, Panel Study of Income Dynamics. (2013). *PSID Transition into Adulthood Supplement 2013: User Guide*. Institute for Social Research, University of Michigan, Ann Arbor, MI.

U.S. Census Bureau. (2012, December 06). *2010 Geographic Terms and Concepts – Block Groups*. Retrieved September 2016 from https://www.census.gov/geo/reference/gtc/gtc_bg.html

U.S. Census Bureau. (2015, March 18). *Geography Atlas – Blocks.* Retrieved September 2016 from https://www.census.gov/geo/reference/webatlas/blocks.html

Ver Ploeg, M., Breneman, P. D., Williams, R., Snyder, S., Dicken, C., & Kaufman, P. (2012, Nov). *Access to affordable and nutitious food: Updated estimates of distance to supermarkets using 2010 data* (Economic Research Report Number 143). Washington, DC: United States Department of Agriculture, Economic Research Service. Retrieved from https://www.ers.usda.gov/webdocs/publications/45032/33845_err143.pdf?v=41505

Walls & Associates. (2014). *NETS: National Establishment Time-Series Database* [Data file].

Walls & Associates. (2014). *National Establishment Time-Series (NETS) Database: 2013 Database Description.* Denver, CO.

Walls & Associates. (2016). *National Establishment Time-Series (NETS) Database: 2014 Database Description.* Denver, CO.

Appendix A. Original SIC Code Request for the NETS database

Food Establishment Category	SIC Code	Food Establishment Category	SIC Code
Groceries and Related Products	514	Coffee shop	58120304
Department Stores*	5331, 5399	Delicatessen	58120305
Grocery stores	54110000	Drive-in restaurant	58120306
Supermarkets	54110100	Fast-food restaurant, chain	58120307
Supermarkets, chain	54110101	Fast-food, independent	58120308
Supermarkets, greater than 100,000 square ft.	54110102	Food bars	58120309
Supermarkets, independent	54110103	Grills (eating places)	58120310
Supermarkets, 55,000 - 65,000 square feet	54110104	Hamburger stand	58120311
Supermarkets, 66,000 - 99,000 square feet	54110105	Hot dog stand	58120312
Convenience stores	54110200	Sandwiches and submarine shop	58120313
Convenience stores, chain	54110201	Snack bar	58120314
Convenience stores, independent	54110202	Snack shop	58120315
Grocery stores, nec	54119900	Lunchrooms and cafeterias	58120400
Cooperative food stores	54119901	Automat (eating places)	58120401
Delicatessen stores	54119902	Cafeteria	58120402
Grocery stores, chain	54119904	Luncheonette	58120403
Grocery stores, independent	54119905	Lunchroom	58120404
Meat and Fish Markets	542	Restaurant, lunch counter	58120405
Fruit and Vegetable Markets	543	Family restaurants	581205
Candy, Nut, and Confectionery Stores	544	Pizza restaurants	581206
Dairy Products Stores	545	Seafood restaurants	581207
Retail Bakeries	546	Steak and barbecue restaurants	581208
Miscellaneous food stores	549900	Eating places, nec	58129900
Health and dietetic food stores	549901	Buffet (eating places)	58129901
Beverage stores	549902	Café	58129902
Miscellaneous food stores, nec	549999	Caterers	58129903
Eating places	58120000	Chicken restaurant	58129904
Japanese restaurant	58120109	Commissary restaurant	58129905
Korean restaurant	58120110	Contract food services	58129906
Lebanese restaurant	58120111	Diner	58129907
Mexican restaurant	58120112	Dinner theater	58129908
Spanish restaurant	58120113	Health food restaurant	58129909
Sushi bar	58120114	Drinking places	5813
Thai restaurant	58120115	Food vending machines	596201
Vietnamese restaurant	58120116	Beverage vending machines	596202
Pakistani restaurant	58120117	Food services, direct sales	59630200
Ice cream and soda stands	581202	Food service, mobile, not coffee-cart	59630204
Fast food restaurants	58120300	Ice cream wagon	59630205
Box lunch stand	58120301	Lunch wagon	59630206
Carry-out only (except pizza)	58120302	Snacks, direct sales	59630207
Chili stand	58120303		

Note. We also requested by name department stores that provide food to their consumers, such as Walmart. Without this request for these specific stores, 90% of these department stores would not have been included in our final dataset with the SIC code request only (Walls, D., personal communication, May 31, 2016).

Appendix B.

Fast Food Restaurant Chain List (Alphabetical Order)

1. A&W Restaurants
2. Amigos/Kings Classic
3. Arby's
4. Arctic Circle Restaurants
5. Arthur Treacher's
6. Atlanta Bread Company
7. Au Bon Pain
8. Auntie Anne's
9. Baja Fresh
10. Bakers Square
11. Blimpies
12. Bojangles' Famous Chicken 'n Biscuits
13. Boston Market
14. Braum's
15. Burger Chef
16. Burger King
17. Burger Street
18. Burgerville
19. Captain D's Seafood Kitchen
20. Carino's Italian Grill
21. Carl's Jr.
22. Carrows
23. Charley's Grilled Subs
24. Checkers / Rally's
25. Cheeburger Cheeburger
26. Chevys Fresh Mex
27. Chicken Express
28. Chick-fil-A
29. Chipotle
30. Chronic Tacos
31. Chuck-A-Rama
32. Church's / Texas Chicken
33. Cici's Pizza
34. Cinnabon
35. Claim Jumper
36. Coco's Bakery
37. Cookout
38. Copeland's
39. Corner Bakery Café
40. Country Buffet
41. Culver's
42. Dairy Queen
43. Del Taco
44. Dixie Chili and Deli
45. Domino's
46. Druther's
47. Dunkin' Donuts
48. EatZi's
49. Eat'n Park
50. Eegee's
51. Pita Pit
52. Pizza Hut
53. Pollo Tropical
54. Popeyes Chicken & Biscuits
55. Port of Subs
56. Potbelly Sandwich Works
57. Portillo's
58. Quizno's Classic Subs
59. Qdoba Mexican Grill
60. Raising Cane's Chicken Fingers
61. Rax
62 Robeks
63. Roy Rogers Restaurants
64. Rubio's
65. Runza
66. Saladworks
67. Sarku Japan
68. Sbarro
69. Schlotzsky's
70. Seattle's Best Coffee
71. Shake Shack
72. Skyline Chili
73. Smashburger
74. Smoothie King
75. Sneaky Pete's
76. Sonic Drive-In
77. Spangles
78. Starbuck's
79. Steak 'n Shake
80. Steak Escape
81. Stir Crazy
82. Sub Station II
83. Subway
84. Swensen's
85. Swensons
86. Taco Bell
87. Taco Bueno
88. Taco John's
89. Taco Mayo
90. Taco Tico
91. Taco Time
92. Tim Horton's
93. ThunderCloud Subs
94. Twin Peaks
95. Umami Burger
96. Wendy's
97. Wetzel's Pretzels
98. Whataburger
99. White Castle
100. Wienerschnitzel
101. Zaxby's
102. Zero's Subs
103. Zippy's

Appendix C. Sample Syntax for Duplicate Retrieval

```
**************************
Creating concatenated latlong variable
**************************;
data netsdata; set netsdata;
      latlong = latitude||longitude;
      latlong = compbl(latlong);
Run;

** Ensure that the proper latitude and longitudes are assigned for the
year of interest. Remember that the NETS Moves file only records
significant moves.;

*****************************************
Creating a table that lists and counts
duplicated latitudes and longitudes
*****************************************;
PROC SQL;
      create table lat_longcount as
            select latlong, count(latlong) as count
            from netsdata
            group by latlong
            having count > 1; quit;
run;

*********************************************
Merging the PROC SQL file with the main file to
Find duplicates based on matching lat/long.
*********************************************;
PROC SORT data = lat_longcount; by latlong; run;
PROC SORT data = netsdata; by latlong; run;

data possible_dup;
            merge lat_longcount (in = a) netsdata;
            by latlong;
            if a;
run;

*********************************************
After creating a potential duplicate data file,
We also attempted to standardize addresses. The
Code below represents much of the efforts toward
This standardization but, please note, that additional
Coding may be necessary.
*********************************************;

* Because of the size of the NETS database, it is advised
To remove unnecessary variables.;



***********************************
Standardizing Addresses
```

```
************************************;

Data possible_dup; set possible_dup;
      ** Removing periods, commas, and semicolons;
      address = compress(address, '.,;');
      ** Removing additional blanks.;
run;

************************************
Standardizing All Numerical Street Addresses
************************************;

Data possible_dup; set possible_dup;
      address = tranwrd(address, "FIRST", "1ST");
      address = tranwrd(address, "SECOND", "2ND");
      address = tranwrd(address, "THIRD", "3RD");
      address = tranwrd(address, "FOURTH", "4TH");
      address = tranwrd(address, "FIFTH", "5TH");
      address = tranwrd(address, "SIXTH", "6TH");
      address = tranwrd(address, "SEVENTH", "7TH");
      address = tranwrd(address, "EIGHTH", "8TH");
      address = tranwrd(address, "NINETH", "9TH");
      address = tranwrd(address, "TENTH", "10TH");
      address = tranwrd(address, "ELEVENTH", "11TH");
      address = tranwrd(address, "TWELFTH", "12TH");
      address = tranwrd(address, "THIRTEENTH", "13TH");
      address = tranwrd(address, "FOURTEENTH", "14TH");
      address = tranwrd(address, "FIFTEENTH", "15TH");
      address = tranwrd(address, "SIXTEENTH", "16TH");
      address = tranwrd(address, "SEVENTEETH", "17TH");
      address = tranwrd(address, "EIGHTEENTH", "18TH");
      address = tranwrd(address, "NINETEENTH", "19TH");
      address = tranwrd(address, "TWENTIETH", "20TH");
run;
*** only went up to 20th ;

************************************
Parsing each word in address into separate variables
************************************;

DATA possible_dup; set possible_dup;
      word1 = scan(address, 1, ' ');
      word2 = scan(address, 2, ' ');
      word3 = scan(address, 3, ' ');
      word4 = scan(address, 4, ' ');
      word5 = scan(address, 5, ' ');
      word6 = scan(address, 6, ' ');
      word7 = scan(address, 7, ' ');
      word8 = scan(address, 8, ' ');
      word9 = scan(address, 9, ' ');
      word10 = scan(address, 10, ' ');
      word11 = scan(address, 11, ' '); run;
************************************
Some addresses were labeled “ONE HALF” or
```

```
"ONE AND A HALF" or "1/2" The aim was to standardize
These address types.;
************************************;
data possible_dup; set possible_dup;
      if word1 = "ONE" and word2 = "HALF" then do
      word1 = "1/2"; word2 = word3; word3 = word4; word4 = word5;
      end;
run;

************************************
Other address standardization attempts
************************************;
data possible_dup; set possible_dup;
      if word3 = "STREET" and word4 ne "RD" then word3 = "ST";
      else if word2 = "STREET" and word3 = " " then word2 = "ST";
run;

data possible_dup; set possible_dup;
      if word2 = "AVENUE" and word3 = " " then word2 = "AVE";
      else if word3 = "AVENUE" and word4 = " " then word3 = "AVE";
run;

data possible_dup; set possible_dup;
      if word1 = "BOULEVARD" and word2 = " " then word1 = "BLVD";
      else if word2 = "BOULEVARD" and word3 = " " then word2 = "BLVD";
run;

data possible_dup; set possible_dup;
      if word1 = "BOUL" or word1 = "BOULV" then word1 = "BLVD";
      if word2 = "BOUL" or word2 = "BOULV" then word2 = "BLVD";
      if word3 = "BOUL" or word3 = "BOULV" then word3 = "BLVD";
      if word4 = "BOUL" or word4 = "BOULV" then word4 = "BLVD";
run;

data possible_dup; set possible_dup;
      if word1 = "HIGHWAY" then word1 = "HWY";
      if word2 = "HIGHWAY" THEN word2 = "HWY";
      if word3 = "HIGHWAY" THEN word3 = "HWY";
      if word4 = "HIGHWAY" THEN word4 = "HWY";
run;

data possible_dup; set possible_dup;
      if word2 = "PARKWAY" and word3 = " " then word2 = "PKWY";
run;

Data possible_dup; set possible_dup;
      if word1 = "ROAD" then word1 = "RD";
      if word2 = "ROAD" then word2 = "RD";
      if word3 = "ROAD" then word3 = "RD";
      if word4 = "ROAD" then word4 = "RD";
      if word5 = "ROAD" then word5 = "RD";
run;

data possible_dup; set possible_dup;
```

```
      if word1 = "MLK" then word1 = "MARTIN LUTHER KING";
      if word2 = "MLK" THEN word2 = "MARTIN LUTHER KING";
      if word3 = "MLK" then word3 = "MARTIN LUTHER KING";
      if word4 = "MLK" then word4 = "MARTIN LUTHER KING";
      if word5 = "MLK" THEN word5 = "MARTIN LUTHER KING";
run;

************************************
Combining all of the words to make the new address variable.;
************************************;
data possible_dup; set possible_dup;
      true_address =
      word1||word2||word3||word4||word5||word6||word7||word8;
run;

*** Note: This code only incorporates the most common standardization
issues found within the address variable. Further cleaning steps may be
necessary and a useful resource for additional address cleaning procedures
can be found at
http://www.sascommunity.org/wiki/Standardizing Addresses in SAS for Geocod
ing;

************************************
This concludes the code pertaining to address standardization
************************************;

************************************
The following code represent the steps taken
To create potential duplicate tables within the NETS.
************************************;

PROC SORT data = possible_dup; by true_address; run;

data possible_dup;
      set possible_dup;
      by true_address;
      if first.true_address then address_id+1;
run;

PROC SQL;
      create table add as
      select address_id, true_address, company, dunsnumber, city, state,
        zip2, tradename, latlong, sic07,
        naics07, emp07, sales07, firstyear, lastyear,
        levelcode, count(true_address) as count
            from possible_dup;
            group by true_address; quit;
run;

** NOTE: The SIC07, NAICS07, EMP07, and SALES07 variables are all variable
that are specific to 2007 in the NETS. As such, users should write their
code specific to their year of interest. **
************************************
Keeping only the first address listed so
```

```
That there are no duplicates in the comparison
file
************************************;
PROC SORT data = add; by address_id; run;

data add; set add; by address_id; if first.address_id; run;

data add; set add;
      rename levelcode = levelcode_2 ;
      rename true_address = true_address2;
      rename company = company2;
      rename dunsnumber = dunsnumber2;
      rename city = city2;
      rename state = state2;
      rename zip = zip2;
      rename latlong = latlong2;
      rename tradename = tradename2;
      rename sic07 = sic07_2;
      rename naics07 = naics07_2;
      rename emp07 = emp07_2;
      rename sales07 = sales07_2;
      rename firstyear = firstyear_2;
      rename lastyear = lastyear_2;
run;

**********************************
Performing the SPEDIS function on the potential duplicates
dataset using various criteria
**********************************;

***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, addresses,
and company names.;

PROC sql;
      create table exact_matches as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and company = company2 and latlong = latlong2
      and city = city2 and state = state2 and zip = zip2; quit;
run;


***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, small spelling
disparities between addresses, and exactly matching company names.;

PROC sql;
      create table possible_match1 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) between 1 and 15
      and dunsnumber ne dunsnumber2
      and company = company2 and latlong = latlong2
```

```
      and city = city2 and state = state2 and zip = zip2; quit;
run;

***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, larger spelling
disparities between addresses, and exactly matching company names.;

PROC sql;
      create table possible_match2 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) between 16 and 45
      and dunsnumber ne dunsnumber2
      and company = company2 and latlong = latlong2
      and city = city2 and state = state2 and zip = zip2; quit;
run;

***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, small spelling
disparities between addresses, and minor spelling differences
between two company names.;

PROC sql;
      create table possible_match3 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) between 1 and 15
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 1 and 15
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, larger spelling
disparities between addresses, and minor spelling differences between
two company names.;

PROC sql;
      create table possible_match4 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) between 16 and 45
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 1 and 15
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, exactly matches between
addresses, and minor spelling differences between
two company names.;

PROC sql;
      create table possible_match5 as
      select * from possible_dup, work.add
```

```
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 1 and 15
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, minor spelling disparities
between addresses, and larger spelling differences between
two company names.;

PROC sql;
      create table possible_match6 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) between 1 and 15
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 16 and 45
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;


***Creating a table in which two establishments listed for
Comparison have exactly matching coordinates, larger spelling disparities
between addresses, and larger spelling differences between
two company names.;

PROC sql;
      create table possible_match7 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) between 16 and 45
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 16 and 45
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

***Possible_match8_2 - Possible_Match8_7 are all potential duplicates with
exactly matching addresses. Different spelling variations for company name
were assigned to each of these possible_match tables and, combined, should
represent company name spelling variations ranging from 16-45.;


PROC sql;
      create table possible_match8_2 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 16 and 20
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;
```

```
PROC sql;
      create table possible_match8_3 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 21 and 25
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

PROC sql;
      create table possible_match8_4 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 26 and 30
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

PROC sql;
      create table possible_match8_5 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 31 and 35
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

PROC sql;
      create table possible_match8_6 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 36 and 40
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;

PROC sql;
      create table possible_match8_7 as
      select * from possible_dup, work.add
      where spedis(true_address2, true_address) = 0
      and dunsnumber ne dunsnumber2
      and spedis(company, company2) between 41 and 45
      and latlong = latlong2 and city = city2 and state = state2
      and zip = zip2; quit;
run;
```

```
**********************************
Reordering the variables so that they are
Easier to interpret in the final possible duplicate
Tables.
**********************************;

%macro reorder(data1);

data &data1;
      retain dunsnumber dunsnumber2 latlong latlong2
      levelcode levelcode_2 true_address true_address2
      company company2 tradename tradename2 city city2
      state state2 zip zip2 sic07 sic07_2 firstyear
      firstyear_2 lastyear lastyear_2;
      set &data1; run;
%mend reorder;

%reorder(exact_matches); %reorder(possible_match1);
%reorder(possible_match2); %reorder(possible_match3);
%reorder(possible_match4); %reorder(possible_match5);
%reorder(possible_match6); %reorder(possible_match7);
%reorder(possible_match8_2); %reorder(possible_match8_3);
%reorder(possible_match8_4); %reorder(possible_match8_5);
%reorder(possible_match8_6); %reorder(possible_match8_7);

*** NOTE: We recommend removing many variables from these tables to make
accessing these potential duplicate matches easier to interpret. ;

**********************************
Exporting all tables into Excel files
**********************************;
*** The folder listed below is just an example folder as
The Excel file should be assigned to the proper destination folder.;

%macro export_dup(data1, data2);
proc export
      data = &data1
      dbms= xlsx
      outfile = "C:\Users\mccsteph\Documents\NETS\Duplicate Files
      \&data2.xlsx"
      replace;
Run;

%mend;

*** Prior to this step, the possible match files were saved into as a
permanent file within a Library called “NETS.”;

%export_dup(nets.exact_matches, exact_match);
%export_dup(nets.possible_match1, possible_match1);
%export_dup(nets.possible_match2, possible_match2);
%export_dup(nets.possible_match3, possible_match3);
%export_dup(nets.possible_match4, possible_match4);
%export_dup(nets.possible_match5, possible_match5);
```

```
%export_dup(nets.possible_match6, possible_match6);
%export_dup(nets.possible_match7, possible_match7);
%export_dup(nets.possible_match8_2, possible_match8_2);
%export_dup(nets.possible_match8_3, possible_match8_3);
%export_dup(nets.possible_match8_4, possible_match8_4);
%export_dup(nets.possible_match8_5, possible_match8_5);
%export_dup(nets.possible_match8_6, possible_match8_6);
%export_dup(nets.possible_match8_7, possible_match8_7);

*** Excel tables with potential duplicate matches listed side-by-side will
be saved as an Excel file within the desired destination folder.;

*** There are establishments that were geocoded at the zipcode level (i.e.
levelcode = 'Z'). These files cannot be verified using Google Earth, yet
there are establishments with Z levelcodes that were duplicated. It may be
easier to examine these separately to determine duplicate status.;
```

Appendix D. Sample Syntax for Creating Food Categories

This section contains potentially identifying information and is available in the restricted data enclave under restricted data contract.

Appendix E. Sample Python Syntax to Create Euclidean Density Feature Classes

```
arcpy.Buffer_analysis(in_features="psid_continental_project",
out_feature_class="P:/Tang-McCracken/GIS Work/2014/FAM14.gdb/hhalfbuff",
buffer_distance_or_field="0.5 Miles", line_side="FULL",
line_end_type="ROUND", dissolve_option="NONE", dissolve_field="",
method="PLANAR")

arcpy.Buffer_analysis(in_features="psid_continental_project",
out_feature_class="P:/Tang-McCracken/GIS Work/2014/FAM14.gdb/honebuff",
buffer_distance_or_field="1 Miles", line_side="FULL", line_end_type="ROUND",
dissolve_option="NONE", dissolve_field="", method="PLANAR")

arcpy.Buffer_analysis(in_features="psid_continental_project",
out_feature_class="P:/Tang-McCracken/GIS Work/2014/FAM14.gdb/hfivebuff",
buffer_distance_or_field="5 Miles", line_side="FULL", line_end_type="ROUND",
dissolve_option="NONE", dissolve_field="", method="PLANAR")

arcpy.Buffer_analysis(in_features="psid_continental_project",
out_feature_class="P:/Tang-McCracken/GIS Work/2014/FAM14.gdb/htenbuff",
buffer_distance_or_field="10 Miles", line_side="FULL", line_end_type="ROUND",
dissolve_option="NONE", dissolve_field="", method="PLANAR")
```

Appendix F. Sample Python Syntax to Calculate Outlets within Density Feature Classes

```
arcpy.TabulateIntersection_analysis(in_zone_features="halfbuff",
zone_fields="FAMID10", in_class_features="ethnic", out_table="P:/Tang-
McCracken/GIS Work/2014/FAM14_buffer.gdb/halfbuff_ethnic", class_fields="",
sum_fields="", xy_tolerance="-1 Unknown", out_units="UNKNOWN")
```

Appendix G. Sample Python Syntax to Calculate Euclidean Distances

```
arcpy.GenerateNearTable_analysis(in_features="psid_continental_project",
near_features="ethnic", out_table="P:/Tang-McCracken/GIS
Work/2014/FAM14_near.gdb/ethnic_near")
```

Appendix H. Sample Python Syntax to Calculate Road Network Distances

```
import arcpy
from arcpy import env
try:
      arcpy.CheckOutExtension("Network")
      env.workspace = "P:/Tang-McCracken/GIS Work"
      env.overwriteOutput = True
      inNetworkDataset = "Geodata/Network/streetmap_na/data/streets4"
      outNALayerName = "ODMatrix"
      impedanceAttribute = "Length"
      searchTolerance = "5000 Meters"
      inOrigin = "2014/FAM14.gdb/psid_continental_project"
      #creat a OD matrix layer, one destination to be found, 100 miles
impedance
      outNALayer = arcpy.MakeODCostMatrixLayer_na(inNetworkDataset,
outNALayerName, impedanceAttribute, "100", "1", "", "ALLOW_UTURNS", "'Non-
routeable Segments';OneWay", "USE_HIERARCHY", "", "STRAIGHT_LINES", "")
      outNALayer = outNALayer.getOutput(0)
      #load origin locations
      arcpy.AddLocations_na(outNALayer, "Origins", inOrigin, "Name FAMID10
#", "5000 Meters", "", "'SDC Edge Source' SHAPE", "MATCH_TO_CLOSEST",
"CLEAR", "NO_SNAP", "5 Meters", "INCLUDE", "'SDC Edge Source' #")
      destinationList = ["2014/FAM14.gdb/bars", "2014/FAM14.gdb/cong",
"2014/FAM14.gdb/ethnic", "2014/FAM14.gdb/ffrest", "2014/FAM14.gdb/fsrest",
"2014/FAM14.gdb/conv", "2014/FAM14.gdb/discount", "2014/FAM14.gdb/fruitveg",
"2014/FAM14.gdb/grocery", "2014/FAM14.gdb/health", "2014/FAM14.gdb/meat",
"2014/FAM14.gdb/specialty", "2014/FAM14.gdb/supercenter",
"2014/FAM14.gdb/supermarket"]
      for inDestinations in destinationList:
            #load destination locations
            arcpy.AddLocations_na(outNALayer, "Destinations", inDestinations,
"", "5000 Meters", "", "'SDC Edge Source' SHAPE", "MATCH_TO_CLOSEST",
"CLEAR", "NO_SNAP", "5 Meters", "INCLUDE", "'SDC Edge Source' #")
            arcpy.Solve_na(outNALayer, "SKIP", "TERMINATE", "5000 Meters")
            #get sublayers
            subLayers = dict((lyr.datasetName, lyr) for lyr in
arcpy.mapping.ListLayers(outNALayer)[1:])
            LinesSubLayer = subLayers["ODLines"]
            #copy features
            outLayerName = inDestinations.split("/")[2]

      arcpy.CopyFeatures_management(LinesSubLayer,"2014/FAM14_OD.gdb/"+str(ou
tLayerName)+"_OD1")
            print outLayerName + " completed successfully"
      print "Script completed successfully"
except Exception as e:
      import traceback, sys
      tb = sys.exc_info()[2]
      print "An error occured on line %i" % tb.tb_lineno
      print str(e)
```