

1.0 Datasets

1.1 United States Department of Agriculture

The USDA offers datasets and the new Food Data Central database (live on April 1, 2019) for the express use of “dietitians to provide specific dietary guidance; researchers to better link dietary intakes to disease measures; and policy makers to develop guidance which promotes public health.” URL: fdc.nal.usda.gov.

1.1.1 Initial Dataset: USDA Branded Food Products Database (August 2018 release)

Filename	BFPD_csv_07132018.zip
Files	BFPD_Doc.pdf Derivation_Code_Description.csv Nutrients.csv Products.csv Serving_size.csv
Source URL	https://data.nal.usda.gov/dataset/usda-branded-food-products-database/resource/c929dc84-1516-4ac7-bbb8-c0c191ca8cec
Authors	Pehrsson, Pamela R. (ORCID); Haytowitz, David B. (ORCID); McKillop, K.A.; Moore, G.; Finley, John W. (ORCID); Fukagawa, Naomi K. (ORCID)
Download Date	2019-04-02
Product Type	Database
Publisher	USDA Agricultural Research Service
Contact Name	Haytowitz, David B.
Contact Email	David.Haytowitz@ARS.USDA.GOV
Access Level	Public
Primary Article	Kretser, A., Murphy, D., & Starke-Reed, P. (2017). A partnership for public health: USDA branded food products database. <i>Journal of Food Composition and Analysis</i> , 64(1), 10-12 DOI: 10.1016/j.jfca.2017.07.019
Related Content	USDA National Nutrient Database for Standard Reference
License	U.S. Public Domain
Funding Source(s)	Agricultural Research Service
Program Code	005:037 - Department of Agriculture - Research and Education
Bureau Code	005:18 - Agricultural Research Service
Modified Date	2019-04-04 (Dataset)
Release Date	2017-11-16 (Article)

This dataset was discontinued as a standalone package and integrated as part of FoodData Central on April 1, 2019. Although we had already started the project, we abandoned this version and moved to the newer version described below.

1.1.2 Final Dataset:

Filename	FoodData_Central_branded_food_csv_2019-04-02.zip
Files	all_downloaded_table_record_counts.csv *branded_food.csv Download & API Field Descriptions-2019-04-02-14-48.xlsx food.csv *food_nutrient.csv
Source URL	https://fdc.nal.usda.gov/download-datasets.html
Download Date	2019-04-11

Filename	FoodData_Central_Supporting_Data_csv_2019-04-02.zip
Files	all_downloaded_table_record_counts.csv Download & API Field Descriptions-2019-04-02-14-48.xlsx food_attribute_type.csv food_category.csv food_nutrient_derivation.csv food_nutrient_source.csv lab_method.csv lab_method_code.csv lab_method_code.csv lab_method_nutrient.csv measure_unit.csv * nutrient.csv nutrient_incoming_name.csv retention_factor.csv wweia_food_category.csv
Source URL	https://data.nal.usda.gov/dataset/usda-branded-food-products-database/resource/c929dc84-1516-4ac7-bbb8-c0c191ca8cec
Download Date	2019-04-11

[Note: The] data in FoodData Central are updated on a varying schedule. New downloads will be made available as new data are added to FoodData Central. The filename for full downloads will include a date indicating when the download file was last compiled.

The files we used were similar to the files we began with in the initial dataset. They are highlighted with an asterisk above and described below. All are .csv files. Definitions for the attributes provided below are available in the file "Download & API Field Descriptions-2019-04-02-14-48.xlsx."

1.1.2.1 Table Attributes in FoodData_Central_branded_food_csv_2019-04-02.zip

branded_food.csv										
fd_c_id	brand_owner	gtin_upc	ingredient_s	servi_ze	servi_size_unit	household_serving_fulltext	branded_food_category	data_source	modified_date	available_date

fdc_id USDA food product code (Primary Key)

brand_owner The company that owns the brand of the referenced food product. (Note: the name of the product, for example, “Chef Boyardee Beef Ravioli, 14 oz.” is stored in a different table, “food” and is keyed to this file with the fdc_id.)

gtin_upc Global Trade Item Number, Universal Product Code. “There are a variety of different types of barcodes. However, the UPC symbol is the most recognized barcode in the United States, since it appears on almost every retail product. The UPC (Universal Product Number) symbol is the barcode representation of the GTIN-12 which consists of twelve numeric characters that uniquely identify a company’s individual product. The GTIN-12 number is part of the family of GS1 global data structures that employ 14 digits and can be encoded into various types of data carriers.”

<https://www.gtin.info/upc/>

ingredients A list of all ingredients contained in the food product, as required by regulation.

Serving_size Expressed in grams or milliliters, defines the size of a portion of the food contained in the package

Serving_size_unit Grams or milliliters.

Household_serving_fulltext The size of a portion of the food contained in the package, expressed as a common household measurement such as cups, ounces, or units (bars, chips, etc.)

Branded_food_category The category of the branded food, assigned by GDSN or Label Insight.

Data_source The source of the data for this food. GDSN (for GS1) or LI (for Label Insight).

Modified_date This date reflects when the product data was last modified by the data provider, i.e., the manufacturer.

Available_date The date the current information about the food was made available to the public via the Food Data Central service.

foodnutrient.csv										
id	fdc_id	nutrient_id	amount	data_points	derivation_id	min	max	median	footnote	min_year_acquired

This table associates the nutrients table with the branded foods table. It includes, for each branded product with an fdc_id, an analysis of each nutrient listed in the nutrients table. The information is provided in grams or milliliters per serving (as described on the Nutrition Facts Panel on the product

package). On the new Food Data Central site, this information can be requested on a per serving size basis and on 100 gram-basis, 100 ml-basis, or fluid oz-basis

Id Unique id for each transaction associating an `fdc_id` with a nutrient (Primary Key)

Fdc_id USDA food product code, as described above (foreign key to branded foods table).

Nutrient_id 4-digit code for nutrient (foreign key to nutrient table).

Amount Amount of the nutrient per 100g of food. Specified in unit defined in the nutrient table.

Data_points Number of observations on which the value is based. Column unused.

Derivation ID of the food nutrient derivation technique used to derive the value. Column unused.

Min Column unused. Minimum amount of nutrient found in any sample of the food.

Max Column unused. Maximum amount of nutrient found in any sample of the food.

Median Column unused. Median amount of nutrient food among all samples of food.

Footnote Comments on any unusual aspects of the food nutrient. Examples might include why a nutrient value is different than typically expected.. Column unused.

Min_year_acquired Minimum purchase year of all acquisitions used to derive the nutrient value. Column unused.

1.1.2.2 Files Used in FoodData_Central_Supporting_Data_csv_2019-04-02.zip

nutrient.csv				
id	name	unit_name	nutrient_nbr	rank

This is a lookup table that describes the nutrient associated with each code. There are 227 entries in this table that identify specific nutrients or combinations of nutrients. A nutrient is a chemical constituent of a food (e.g. calcium, vitamin E) officially recognized as essential to human health

Id Unique ID for each entry in the table (Primary Key)

Name the common name of the nutrient

Unit name The standard unit of measure for the nutrient (per 100g of food), e.g., g, mg, mcg.

Nutrient_nbr A unique code identifying a nutrient or food constituent. (Please note that this attribute name may differ from the attribute name used to refer to it (for example, `nutrient_id` in the brandedfood table).

Rank Attribute not defined in the FDI food attribute definitions fiile; assumed to mean relative importance of the nutrient

1.1.3 Table Modifications for DGDiet Project

For the purpose of this project the USDA files were modified to simplify the database and make attribute names consistent across all tables. Underscores were removed from attribute names.

tbrandedfood.csv				
fdcid	gtinupc	servingsize	servingsizeunit	householdsgivingfulltext

tfoodnutrient.csv			
tfnid	fdc_id	nutrient_id	amount

tnutrient.csv			
tnid	name	unit_name	nutrient_id

Note: nutrient_nbr changed to nutrientid to be consistent with tfoodnutrient.

1.2 Dollar General Dataset

The Dollar General dataset was constructed by the project team because digital data was not available from the company despite multiple requests. The manager of the local Turtle Creek store (“Adina”) requested permission to collaborate with us but was denied permission by DG management. Phone: 412-824-8733. The designated representative of the DG public relations department (Mary Catherine Colbert) at corporate headquarters did not return calls after an initial request and multiple voicemail messages. Phone: 615-855-4000.

The DGInventory data table includes the following field attributes:

dginventory.csv						
dgid	gtinupc	dgitemid	dgdescription	dgprice	totaloz	totalg

Dgid Unique ID for each entry in the table, beginning with 100.

Gtinupc As described above, Global Trade Item Number, Universal Product Code.

Dgitemid Dollar General SKU (Stock Keeping Unit).

Dgdescription Product description. For example, “Chef Boyardee Beef Ravioli, 15 oz.”

Dgprice Posted price of product, as packaged, at Dollar General website on April 14, 2019.

Totaloz The total ounces of product contained in the package.

Totalg Total grams or milliliters of product contained in the package, obtained by converting totaloz to grams.

The dataset was constructed in the following manner. An excel spreadsheet was used to compile information.

1. A list of representative stock items was selected from the Dollar General website. For each item, its name and size, the Dollar General item number (SKU), and current price were noted in the spreadsheet. No UPCs were published on the DG website.

- Items selected were judged to be “more healthful” based on product category; for example, dessert items (cookies and cakes) and certain snack foods (potato chips) were not included.
- Only a limited number of brands/packages of each food product were included. For example, only three examples of refried beans were included in the list.
- Food for children and infants was specifically excluded; babyfood and fruit juices were not added to the list.
- Only shelf-stable foods (not requiring refrigerator or freezer) were selected because not all Dollar General stores have these storage facilities. The team discovered that the USDA approves stores for its WIC program (Special Supplemental Nutrition Program for Women, Infants, and Children) only if they have the ability to offer refrigerated milk.

2. Because there was no link between the USDA Branded Food table and Dollar General information collected, UPCs for the DG items were collected. To find the UPCs, a Google search for each product name was attempted. In many cases a matching UPC could be found on one of the following sites:

<https://www.barcodelookup.com/> “Product information and images for millions of items worldwide.Total current barcode entries: 124,153,567”

<https://www.upcitemdb.com/> “With over 227 million unique UPC/EAN numbers, UPCitemdb.com is the largest UPC lookup database where you can search a broad range of UPC numbers to find related product information, images, barcodes, online shopping guide and more.”

Nevertheless, UPCs for many of the products – particularly those associated with the Dollar General brand “Clover Valley” were not found.

To fill in these gaps, the team made two trips to Dollar General to collect the missing UPCs from the product packaging. Two trips were required because some products were not available in the selected store (Turtle Creek) and were found another store (Trafford).

3. The team found that the USDA Branded Foods table could not be linked via UPCs because of the inconsistent use of UPCs both by industry and the USDA itself. For example,

Most DG Clover Valley brand products were not listed in the USDA branded foods database.

USDA fdc_ids often provided multiple analyses for the same product, e.g., same can used for two different formulations of Campbell’s Cream of Mushroom Soup

The USDA maintains multiple UPCs for the same food name and package size, e.g., 10 fdc_id Codes for “Original Kraft Macaroni & Cheese Dinner”

Therefore, UPCs were found unusable for the purposes of this project.

4. The team searched the USDA Food Data Central website for branded food products in the Dollar General inventory. For example:

- Kraft lists only 22 items with its brand name, e.g., does not include Miracle Whip (except Light).

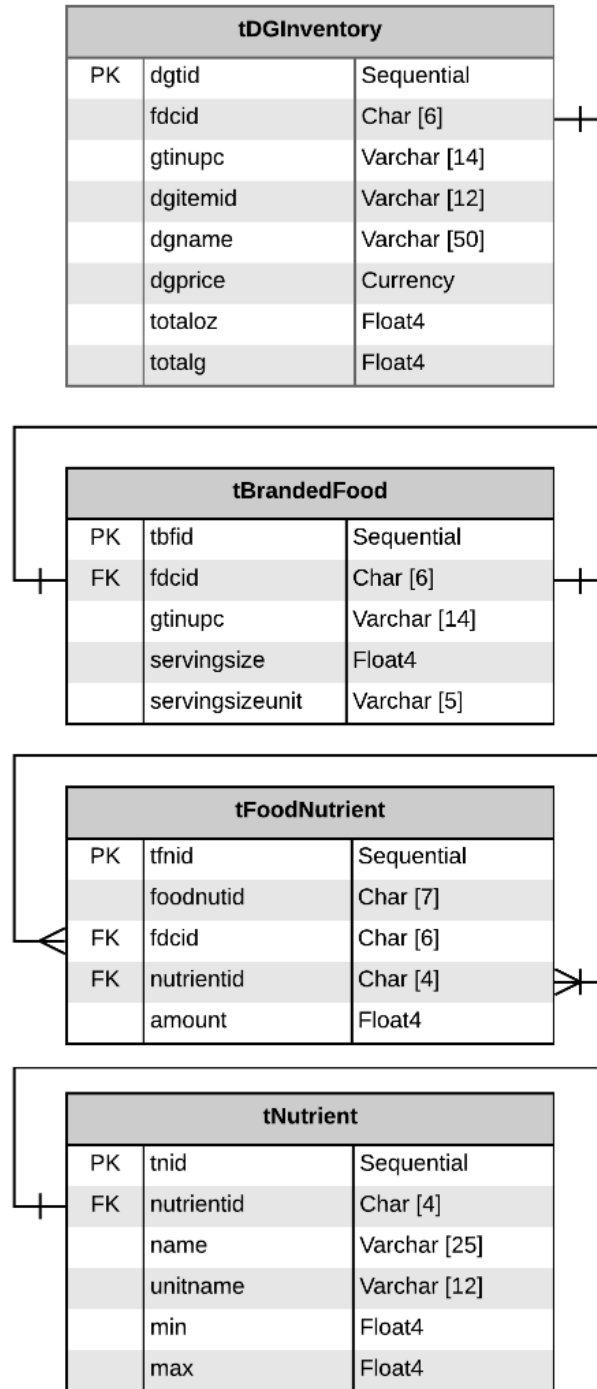
- Zatarain's lists 163 items.
- Smucker's lists only 12 items, and does not include Jif Peanut Butter or Smucker's Strawberry Jam.
- Of eight Clover Valley soups (Dollar General private label), only two are in the USDA database.

Of the 300+ products initially selected, only about 200 could be found. Because manufacturer/brand owner cooperation with the USDA is voluntary rather than mandated, many companies do not provide information. The DG Inventory table was extended to include the `fdc_id` attribute.

2.0 Database Table Structure

DGDiet Data Structure Diagram (ERD)

pshubock@ccac.edu | May 19, 2019



3.0 Query Specification

3.1 Research Question: What is the cheapest healthy diet that you can purchase at a dollar store?

To answer the research question via the Solver LP in Excel (Simplex Method) the following table of data would be required and exported from PostgreSQL as a .csv file.

Attribute	fdc_id	dg_item_id	dg_description	dg_price	totalg	serving_size	serving_unit	1008	1003	1004	1050	1087	1104-1108	1162	1089
Table	tDGInventory	tDGInventory	tDGInventory	tDGInventory	tDGInventory	tBrandedFood	tBrandedFood	tFoodNutrition	tFoodNutrition	tFoodNutrition	tFoodNutrition	tFoodNutrition	tFoodNutrition	tFoodNutrition	tFoodNutrition
								KCAL	G PROTEIN	G FAT	G CARB (by summation)	MG CALCIUM, CA	IU/UG VIT A?	MG VIT C (total ascorbic acid)	MG IRON, FE

Initially, Vitamin A will be eliminated from the model because there are five different types (retinol, beta-carotene, and three others) that are measured in different units that would need to be converted (IU/UG): nutrients 1104, 1105, 1106, 1107, and 1008. If you like we can add something else.

Table	Attributes							
tDGInventory	fdc_id	dg_item_id	dg_description	dg_price	totalg			
tBrandedFood	serving_size	serving_unit						
tFoodNutrition	1008	1003	1004	1050	1087	1104-1108	1162	1089
	KCAL	G PROTEIN	G FAT	G CARB	MG CALCIUM	IU/UG VIT A?	MG VIT C	MG IRON, FE

3.2 Constraints for LP Optimization

3.2.1 Dietary Reference Intakes (DRI) and Life Stage Group

DRI is the general term for a set of reference values used to plan and assess nutrient intakes of healthy people. These values, which vary by age and gender, include:

- Recommended Dietary Allowance (RDA): average daily level of intake sufficient to meet the nutrient requirements of nearly all (97%-98%) healthy people.
- Adequate Intake (AI): established when evidence is insufficient to develop an RDA and is set at a level assumed to ensure nutritional adequacy.
- Tolerable Upper Intake Level (UL): maximum daily intake unlikely to cause adverse health effects.

DRI are calculated by “Life Stage Group”:

Infants	Children	Males	Females	Pregnancy	Lactation
0–6 mo	1–3 years	9–13 y	9–13 y	≤ 18 y	≤ 18 y
7–12 mo	4–8 years	14–18 y	14–18 y	19-30y	19-30y
		19–30 y	19–30 y	31-50 y	31–50 y
		31-50 y	31-50 y		
		50-70 y	50-70 y		
		> 70 y	> 70 y		

For the purposes of this project, a 2000-calorie diet for an adult between the ages of 31-50 with no special dietary provisions was selected. In the future, a front-end form may be used to collect additional information to fine-tune the optimization.

3.2.2 LP Objective and Constraints

The project goal is to minimize the total cost of a diet that includes only foods purchased from Dollar General and yet meets all the constraints established for a 2000-calorie diet with maximum and minimum amounts of nutrients per current DRIs. The selected nutrients include only those shown on the Food Nutrition Facts Label. Please see 2019-05-02_Dollar_General_LP_Constraints.xlsx included in OneDrive project folder.

<https://www.fda.gov/food/nutrition-education-resources-and-materials/how-understand-and-use-nutrition-facts-label>

	Constrain	Min	Max	Unit	FDA Target (Daily Value)	Calculated Calories	Cal/g	
Calories (Energy)	Target		2000	kCal		1985		Adult Male
Total Fat	Minimize	0	65	g	65	585	9	Label
Saturated Fat	Minimize	0	20	g				
Trans Fat	Minimize	0	0					
Cholesterol	Minimize	0	300	mg				
Sodium	Minimize	0	2400	mg				
Total Carbohydrate		0		g	300	1200	4	https://www.accessdata.fda.gov/scripts/interactivenutritionfactslabel/carbohydrate.html
Dietary Fiber		0	25	g				
Sugars								
Protein				g	50	200	4	https://www.accessdata.fda.gov/scripts/interactivenutritionfactslabel/protein.html
Vitamin A	Get Enough	5000	10000	IU				https://ods.od.nih.gov/factsheets/VitaminA-Consumer/
Vitamin C	Get Enough	90	2000	mg				https://ods.od.nih.gov/factsheets/VitaminC-Consumer/
Calcium	Get Enough	1000	2500	mg				https://ods.od.nih.gov/factsheets/Calcium-Consumer/
Iron	Get Enough	8	45	mg				https://ods.od.nih.gov/factsheets/Iron-Consumer/