

Providing the most cost-effective, nutritionally-sufficient, food recommendations for Americans

Andrew Freiburger

March 19, 2025

Abstract

An increasing percentage of Americans find themselves financially strained in paying for weekly groceries. This often leads to buying the most calorically dense and processed foods, since these are cheaper, over healthy and ostensibly more expensive foods. There have been attempts in literature and at the USDA to create optimizations that find healthful dietary choices while adhering to strict budgets, but these attempts have generally been academic exercises and have failed to impact the demographic that they were trying to help. Motivated by this shortcoming, I leveraged the “official” and open-source USDA data sets pertaining to food prices, food nutritional composition, and human physiological needs to design an optimization model that finds a nutritionally adequate diet for the user, explored for myself herein, at the lowest possible cost. The optimum diet offered a surprisingly diverse set of ingredients and quantities that met my nutritional needs for \$9.68 per day at current prices (January 2025).

Keywords: Food, nutrition, USDA, optimization, open-source, LP, cost-effective, foodstamps

Introduction

The Food Stamp Act of 1964, passed as part of President Lyndon Johnson’s “Great Society” legislative acts, ushered in the first continuous program of direct government assistance in the purchasing of groceries (a concept first temporarily employed during the Great Depression). The program initially covered 0.2% of the US population but now subsidizes the groceries of 12.4% of Americans, peaking at 15.1% in 2012 (1), and costs ~\$100B in just Federal expenditures (2). The impact of this program and the needs of its growing population of Americans must therefore be an increasing focus of public and private research..

Food choices at a grocery store are exceptionally difficult, despite seeming so ordinary, because there are thousands of options with varying nutritional densities and all 92 of human’s nutritional needs are only known by a tiny fraction of the population. Low-income Americans therefore often make food choices based on some compromise between taste, satiation, and price, which generally results in choosing foods with the most calories and least micronutritional value (3) and will negatively affect their well-being and perhaps the development of their children (4). Given the complexity of this situation and the severe absence of health and nutritional education, however, I believe that even the most health-conscious consumers make suboptimal dietary choices.

Background

The evident problem of finding nutritional-rich diets on strict budgets has been a topic of academic investigation for a century (5), but has received growing interest presumably because

of the advances in computational resources and object-oriented programming languages. A review from 2015 detailed efforts in this direction over the previous decade (6). Unfortunately, these studies generally only examined calories or the aggregate cost of different diet regimes, and their optimizations did not offer specific foods that meet all of the known micronutritional dietary goals.

There are several studies that have included micronutritional data in their optimization, although these studies generally examined the affects of dietary choices on nutrition or cost and did not provide targeted dietary guidance (7, 8). The only study I found that offered specific dietary advice to meet nutritional needs at the least possible cost (9) was designed specifically for the Ghanan population, and included local wild foods as a part of the diet that obviously cannot be directly used for American consumers who do not have access to these free foods.

The USDA's Thrifty Food Plan (10) is the government's attempt to provide dietary guidance through suggesting proportions of food groups (e.g. 10% leafy greens, 20% whole grains, etc) at various cost regimes. Unfortunately, this endeavor and its annual report fail to offer concise, actionable, advice for Americans earnestly looking to make the best dietary choices in the grocery aisles.

After reviewing available literature, it appears that there has not been a comprehensive optimization model constructed that provides the Western consumer with suggestions of food selection that most cost-effectively satisfy accepted nutritional guidelines. I believe that this is a huge unaddressed market and would dramatically improve the health of the people in our communities who are perhaps in the most need of aid.

Model Formulation

Data aggregation

Three data sets were acquired for my optimization model: food pricing, physiological needs, and food nutrients. I elected to source these data from the USDA, assuming that sourcing all of these data from the same government entity would create the most robust optimization and streamline connecting the disparate data. These assumptions unfortunately proved false: the pricing data was incredibly sparse (150 foods, and only fruits and vegetables), the physiological data contained less than half of the needs identified in literature, and the data sources used different namespaces so they were not interoperable. The following adjustments were therefore necessary for each of the respective data sources to correct some oversights and standardize the data sources so that they could be integrated into the optimization herein.

Pricing data

The pricing data was sourced from the Economic Research Service Division of the USDA(11), which defined only 155 food items and had extensive duplication: e.g. fresh raspberries and frozen raspberries, or fresh artichoke and canned artichoke. Frozen foods have minimal differences with fresh foods, and canned foods are much less nutritionally defined in the food composition data than fresh foods, so frozen and canned foods were omitted from consideration to create this optimization. The resulting 69 foods were the basis of the optimization, and are listed in **Table A1**.

After the optimization failed to converge with just these variables, I noticed that it was not possible to meet the fat requirements given these food options. I therefore added flaxseeds to the optimization as a low-cost, whole-food, source of fat and protein. The pricing data was sourced from Amazon (12), the yield was assumed to be 1 (since the whole seed is consumed), and the cupEquivalentSize (defined by the USDA as weight per cup of a food) derived from an analytical website (13).

Physiological data

Physiological needs for myself – Active, 28 years-old man, 5’10”, 150lb – were sourced from the USDA “DRI [daily required intake] Calculator for Healthcare Professionals” and processed into **Table 1**. While this is not the most comprehensive list of nutritional needs, containing less than half of the nutrients that are actually required by humans, it was selected because it is an “official” and open-source set of nutritional needs and covers the most important nutrients. The assumption likely made by the USDA is that by following these nutritional guidelines is that by satisfying these nutritional needs through eating whole foods, one would likely also satisfy nutritional needs for the other trace nutrients that are not captured here.

Several nutrients – Chromium, Vitamin B12, Vitamin D, Cholesterol, Iodine, Molybdenum, and Phosphorus – were contained in less than 6 foods, and were therefore omitted to prevent the optimization from either becoming infeasible (such as Vitamin D and Iodine which were in no foods) or from forcing the inclusion of the few foods that contain these nutrients (such as Molybdenum in just 3 foods). This is partly based on the belief that some of these nutrients are present in most of whole foods despite them not being explicitly measured in these data: i.e. I believe that unreported nutrient abundance in foods cannot be interpreted as a measured zero abundance of a nutrient in a food.

Table 1: The nutritional requirements that are used for the optimization, with changes noted. Notably, all of the “ND” (Not Determined) upper bounds were changed to 10,000 of the respective units, which are not included in the “Changes” column to highlight the more impactful changes to the physiological nutrient needs. All of the units were also standardized – grams for

macronutrients and milligrams for micronutrients – to facilitate the direct comparison between nutrients and easier programming.

Nutrient	Lower bound	Upper bound	Modifications
Carbohydrate	331g	478g	
Total Fiber	41g	ND	
Protein	54g	ND	The range was changed to [100,150] to meet my specific exercise needs
Fat	65g	114g	
Saturated fatty acids	"As low as possible"		Upper bound of 10% total calories (33.3g) based on recent literature
Linolenic Acid	1.6g	ND	
Linoleic Acid	17g	ND	
Dietary Cholesterol	"As low as possible"		Upper bound of 800mg, as a feasible upper limit (~4 eggs worth).
Total Water	NA	3.7L	A low of 0.37 (10% o max) was created since food might be too dry below this.
Energy	2400kcal	3200kcal	Added this term because it was NOT included in the USDA recommendations
Vitamins			
Vitamin A	900 mcg	3,000 mcg	
Vitamin C	90 mg	2,000 mg	
Vitamin D	15 mcg	100 mcg	
Vitamin B6	1.3 mg	100 mg	
Vitamin E	15 mg	1,000 mg	
Vitamin K	120 mcg	ND	
Thiamin	1.2 mg	ND	
Vitamin B12	2.4 mcg	ND	
Riboflavin	1.3 mg	ND	
Folate	400 mcg	1,000 mcg	The upper bound was changed to 3000mcg to reflect more recent literature
Niacin	16 mg	35 mg	
Choline	0.55 g	3.5 g	
Pantothenic Acid	5 mg	ND	
Biotin	30 mcg	ND	

Carotenoids	NA	ND	The range was arbitrarily defined as [0,1E5], since carotenoids are generally not nutritional
Minerals			
Calcium	1,000 mg	2,500 mg	
Chloride	2.3 g	3.6 g	
Chromium	35 mcg	ND	
Copper	900 mcg	10,000 mcg	
Fluoride	4 mg	10 mg	
Iodine	150 mcg	1,100 mcg	
Iron	8 mg	45 mg	
Magnesium	400 mg	350 mg	The upper bound here is amazingly lower than the lower bound, so the upper bound was created to be 10,000mg
Manganese	2.3 mg	11 mg	
Molybdenum	45 mcg	2,000 mcg	
Phosphorus	0.7 g	4 g	
Potassium	3,400 mg	ND	The ND here was replaced with 15,000mg because potassium is generally not a concern in excess and was too close to the lower bound at the 1E4mg default.
Selenium	55 mcg	400 mcg	
Sodium	1,500 mg	2,300 mg	
Zinc	11 mg	40 mg	

Food composition data

The nutritional composition of various foods was sourced from the 10/24 release of the USDA's "FoodData Central" collection of food data (14). The resulting zip file contained ~50 files, of both data and metadata. The metadata, in addition to documented terminology (15), were used to decypher the IDs from the data files. The resulting data files contained 18.1M nutritional data points and 1.9M foods. Given the limited scope of pricing data, the array of food options were filtered for just foods containing "raw" to match the exclusive list of fruits and vegetables

from the pricing data, which resulted in 2254 foods. These foods included raw meats as well as fruits and vegetables, but these were filtered later in the optimization code.

Model construction

The model consists of a variable for each food item – ultimately 70 including flaxseeds – that were bound $[0,5]$ to limit consolidation of the solver to just a few nutritionally dense and cheap foods. This is also designed to make the dietary suggestions more attractive for people who like to eat some diversity of foods over the course of a week.

The primary constraints in the system is

$$lb_n \leq \sum_n (q_n * f_n \forall f \in F) \leq ub_n$$

where q_n is the quantity of the nutrient n in food f_n for the F foods that contain nutrient $n \in N$. Each sum for each nutrient is bounded by the lb_n and ub_n USDA physiological requirements for each nutrient displayed in **Table 1**. Another constraint was added for volume

$$5 [cups] \leq \sum_f (vol_f * f_n) \leq 20 [cups]$$

to ensure that the volume of food is reasonable. The objective function

$$\min \sum_f (p_f * f)$$

finds the total minimal yielded-weight price p_f for 100-gram units of all food $f \in F$.

Solution

Algorithm

The optimization was solved through the Simplex algorithm via the GLPK (GNU Linear Programming Kit) solver (16), through the Optlang and OptlangHelper Python package. All variables were continuous, so the problem was strictly a linear program and not a MILP.

Results

The optimized diet was \$9.68 per day and consisted of the foods in **Table 2**. The diet was highly skewed towards pinto beans and carrots, presumably as sources of protein & vitamins

and minerals, respectively. Interestingly, only 38 grams (~¼ of a cup) of flax seeds were needed to balance the diet, and specifically meet the fatty acid requirements, from the otherwise completely whole-food vegan diet. The respective nutrient quantities of this optimized diet were visualized between lower and upper bounds in **Figure 1**, which, importantly, indicates which nutrient constraints were limiting the optimization.

Table 2: The gram amount of each food that satisfies all of the constraints for nutritional needs and volume requirements.

Food	Carrots	Pinto beans	Blueberries	Collards	Corn	Celery	Raspberries	Cucumber	Flax seeds
Amount (g)	500	500	249	197	173	155	118	75	38

The optimum ingredients from **Table 2** were passed into ChatGPT to create an actionable dietary regime that could be followed. The prompt read: “create a day of meals from the following ingredients, without adding any calories. The numbers are denominated in grams of the given food.” The meal regime proposed by ChatGPT is contained in **Figure 2**. The optimization could conceivably be run iteratively to simulate an entire week, and then the cooking instructions for the full week would be specified by ChatGPT, effectively emulating the role of a personal nutritionist.

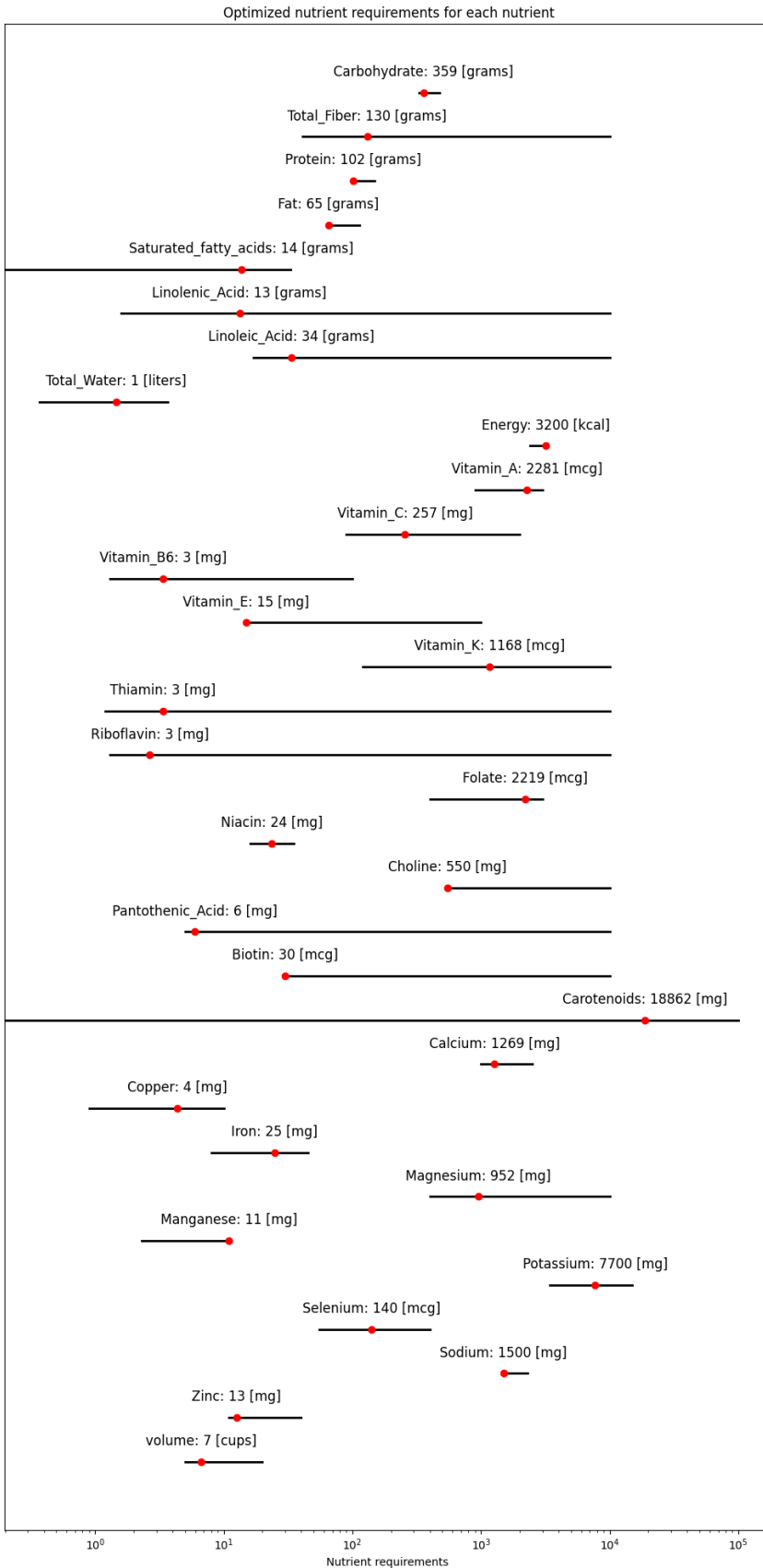


Figure 1: The nutritional requirements of each nutrient on a log-scale to normalize the distribution of requirements in the plotting space, where the lower and upper bounds are the ends of each line segment. The red dot and printed quantity above each segment represents the optimized values for each nutrient. This figure illustrates where the optimized nutrient values fall within the required ranges, and importantly indicates which nutrients are particularly limiting in the optimization.

Here's a full day of meals using only your provided ingredients, without adding extra calories:

Breakfast: Berry & Flax Smoothie

- Blueberries (125g)
- Raspberries (118g)
- Flax seeds (38g)
- Water (as needed)
- Blend everything together into a smoothie.

Lunch: Pinto Bean & Corn Salad

- Pinto beans (250g)
- Corn (173g)
- Celery (156g)
- Cucumber (75g)
- Toss together into a refreshing salad.

Snack: Carrot & Celery Sticks

- Carrots (250g)
- Celery (156g)
- Crunchy and hydrating.

Dinner: Collard Bean Wraps

- Collard greens (197g)
- Pinto beans (250g)
- Corn (173g)
- Wrap beans and corn inside the collard leaves.

Would you like any modifications or different meal structures?

Figure 2: The meal guidelines proposed by ChatGPT given the optimized diet, while not adding any calories to the diet. This could be the final output of a future tool, by giving people the most tangible suggestions that they could easily and quickly implement into their lifestyles. The Corn (173g) dietary element was also duplicated in the Lunch and Dinner suggestions, which demonstrates the need to improve the prompts and refinement of the ChatGPT dialogue to give consistent recommendations with the dietary contents

Conclusions

The USDA's data sources for dietary nutritional needs and the price and nutritional composition of foods were sufficient to parameterize an optimization model that makes the most efficient dietary suggestions to meet established dietary needs. The pricing data was unfortunately only fruits and vegetables, which bottlenecked the optimization model to only consider a subset of whole-food vegan diets and was initially an infeasible problem. This has interesting implications that some vegan diets are inherently incompatible with human needs. The addition of just 38 grams of flaxseeds, however, was sufficient to create a nutritionally sufficient diet at the price of \$9.68 per day. I believe that this framework, with better data and a proper user interface, would be a powerful tool for both financially-stressed households and for health-conscious individuals who are trying to find the most efficient means of a nutritious and healthful lifestyle.

Future Directions

The primary bottleneck in the food selection and thus completeness of the project was the USDA food pricing data, which only contained 69 fruits and vegetables. In addition, the incompatibility of this data set with the other USDA data sets further diminishes its utility. Future iterations of this model should either use a 3rd-party data set or manually create pricing data, which could be done by transcribing prices at a common grocery store such as Trader Joe's. Better pricing data would allow the full usage of the food compositional data, and thus through including new food groups – such as nuts, seeds, dairy, and eggs – would expand the optimization solution space and find a better cost-efficient nutritious diet.

Extending the model with new nutritional sources would further enhance the model, such as Oxygen Radical Antioxidant Capacity (17) values to consider the antioxidant capacity of various foods. The carbon and water footprints of foods could also be considered. These additions would broaden the scope of the optimization, and possibly lead to new insights about the optimal diet.

Another improvement of the model could be to constrain individual amino acids instead of broadly constraining total protein consumption. This improves the accuracy of nutritional needs from the model, and specifically ensures the nutritional value recommended diets for vegetarian users who are at greater risk of meeting protein requirements while missing amino acid requirements.

Ideally, there should also be optional blacklisting of foods, according to allergies or strong preferences, and whitelisting according to either advanced user specifications or nutrients that support managing certain diseases (e.g. lycopene improves the survival of lung cancer (18)). Whitelisting and blacklisting would either force a food(s) to be included or excluded, respectively, in the optimized diet. This would facilitate acceptance of the dietary recommendations and maximize positive impact on the users' health. In this same vein, the diet should also limit representation of food-groups to not overwhelm a diet with similar foods, which should further improve dietary adoption.

The final set of improvements would be to create a refined software package. First, the optimization outputs would be programmatically ported into ChatGPT and propose meals that follow the optimized dietary regime. This ensures that the dietary suggestions are both tangible and more likely to be followed. This could be even further extended to give cooking instructions for the meals. Second, the optimization would be created for a whole week since this is the timescale on which people shop for groceries and plan their meals, as well as integrate dietary variation. Thirdly, a simple user interface should be created that parameterizes the optimization and displays suggested foods & quantities with their associated suggested meals and cooking instructions. The interface would be written in HTML+CSS and be connected to the back-end Python code of this project through an abstraction such as Django. This interface would be the final end product that could actually support the target audience and allow, particularly those on food stamps, to be nutritionally satiated while being cost-conscious.

References

1. A Short History of SNAP. *USDA Food Nutr. Serv.* Available at: <https://www.fns.usda.gov/snap/history>.
2. “Supplemental Nutrition Assistance Program” (USDA Food and Nutrition Service, 2018).
3. N. Darmon, A. Drewnowski, Does social class predict diet quality? *Am. J. Clin. Nutr.* **87**, 1107–1117 (2008).
4. E. B. Tate, *et al.*, Children’s Executive Function and High-Calorie, Low-Nutrient Food Intake: Mediating Effects of Child-Perceived Adult Fast Food Intake. *Health Educ. Behav.* **42**, 163–170 (2015).
5. G. Stigler, The Cost of Subsistence. *J. Farm Econ.* **27**, 303–314.
6. N. Darmon, A. Drewnowski, Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. *Nutr. Rev.* **73**, 643–660 (2015).
7. N. Darmon, E. L. Ferguson, A. Briend, A Cost Constraint Alone Has Adverse Effects on Food Selection and Nutrient Density: An Analysis of Human Diets by Linear Programming. *J. Nutr.* **132**, 3764–3771 (2002).
8. M. F. Koenen, R. Geelen, M. Balvert, H. Fleuren, A tree-based approach to identify indispensable foods in minimum-cost food baskets. *Front. Nutr.* **11**, 1425749 (2024).
9. E.-P. Nykänen, H. Dunning, R. Aryeetey, A. Robertson, A. Parlesak, Nutritionally Optimized, Culturally Acceptable, Cost-Minimized Diets for Low Income Ghanaian Families

Using Linear Programming. *Nutrients* **10**, 461 (2018).

10. “Thrifty Food Plan, 2021” (USDA Food and Nutrition Service, 2021).
11. J. Hyman, H. Stewart, “Fruit and Vegetable Prices - Documentation” (USDA Economic Research Service).
12. Badia Organic Ground Flax Seed. Available at:
https://www.amazon.com/Badia-Organic-Flax-Ground-16-Ounce/dp/B0048IMRZ6/ref=sr_1_44_pp?crid=24Q4I8PT0RT2K&dib=eyJ2IjojMSJ9.URkA3XTShImPiJLGIOH_jobO3G5_H00sfHqrW2LAVBR3nL2Sp7IdU-QB-XD40Ti0-GpyQyoEHZpLwQXI54Tccv6BuL9LgVanXVzfO0j6-bFV04aHli74QYXia6f3nbp6OO4hTYYRA7jUzZyMGdEB6Eh4YVULWE8UE7F41tzt8tlV0EzJcUTfLxpPwu_PiLIOQ1XNtqlRfpYbK21z8aE3_UfkPZZdAN6rYWYB8HaUHftUritSWwxgcF_jKNxj_AmYGTLI_OVEOhjwG2_m6UP9LrI1dc5fu5GoQ8POcjX7bICYxwiZMflukmePW5GvYduiikgzjTg3lIFBlgd1JTBWDnV-AwQvOwJDenUvAz8wyU-ddOnRc4dw095mWz7zIHpMjA_Jwx4Vzgbbi_xpMfeEoQWMeCRsXnzR7VAMoTcDIP8f4ZfESDTGVD4-Y793aI2D.W8UcSuWGSW7_xRo_fMCeSUziOMrh1TDe0INpSw1emdg&dib_tag=se&keywords=flax%2Bseeds&qid=1742143137&prefix=%2Caps%2C126&sr=8-44&th=1.
13. “Density of Seeds, flaxseed (whole) (food).”
14. Full Download of All Data Types. FoodData Central Download Datasets.
15. FoodData Central Download Field Descriptions. (2020).
16. A. Makhorin, GLPK (GNU Linear Programming Kit). (2012).
17. ORAC Values: Antioxidant Values of Foods and Beverages. Superfoodly.
18. J. Cheng, B. Miller, E. Balbuena, A. Eroglu, Lycopene Protects against Smoking-Induced Lung Cancer by Inducing Base Excision Repair. *Antioxidants* **9**, 643 (2020).

Appendix

Table A1: A sampling of refined food prices. The full list of these food prices is listed in the associated Python Notebook from this project.

Food	RetailPrice	Yield	CupEquivalentSize	CupEquivalentPrice
Acorn squash	1.2136	0.4586	0.4519	1.1961
Apples	1.8541	0.9	0.2425	0.4996
Apricots	5.63865	0.965	0.25355	1.25615
Artichoke	2.4703	0.375	0.3858	2.5415
Asparagus	4.88715	0.76365	0.3968	2.4961
Avocados	2.6737	0.7408	0.3197	1.1538
Bananas	0.5971	0.64	0.3307	0.3085
Berries	4.2673	1	0.3307	1.4112
Black beans	1.525	2.4692	0.3858	0.2383
Blackberries	5.788	0.98	0.3252	1.9218
Blackeye peas	1.9265	2.5397	0.3858	0.2926
Blueberries	3.98925	0.975	0.3252	1.3313
Broccoli	2.781333	0.887266	0.3417	1.08523333
Brussels sprouts	2.7544	1.01	0.3417	0.93155
Butternut squash	1.2691	0.714	0.4519	0.8033
Cabbage, green	0.797	0.7788	0.3307	0.3384
Cabbage, red	1.2604	0.7791	0.3307	0.535
Cantaloupe	0.7523	0.51	0.3748	0.5528
Carrots	1.25765	0.92145	0.29765	0.405125
Cauliflower	2.79456	0.931066	0.2756	0.82146666
Celery	1.8289	0.865	0.2646	0.54075
Cherries	4.6632	0.92	0.3417	1.7321

Clementines	1.5811	0.77	0.463	0.9507
Collard	2.71965	1.0209	0.3252	0.86645
Corn	1.9585	0.7515	0.3638	1.06945
Cranberries	5.0828	1	0.1232	0.626
Cucumber	1.2473	0.85	0.2646	0.3961
Figs	7.3233	0.96	0.1653	1.2613
Grapefruit	1.4444	0.49	0.463	1.3647
Grapes	3.01495	0.98	0.248	0.68085
Great northern beans	1.7202	2.4692	0.3858	0.2688
Green beans	2.3091	0.8748	0.2866	0.7557
Green peas	1.996	0.8924	0.3527	0.7889
Green peppers	1.4789	0.82	0.2646	0.4771
Honeydew	1.1589	0.46	0.3748	0.9442
Kale	2.86605	0.97475	0.3252	0.9334
Kidney beans	1.9176	2.4692	0.3858	0.2996
Kiwi	2.6064	0.76	0.3858	1.3231
Lentils	1.839	2.4692	0.3858	0.2873
Lettuce	2.7938	0.895	0.2094	0.66115
Lettuce, iceberg	1.2512	0.95	0.2425	0.3194
Lima beans	2.4167	1.81835	0.3803	0.5986
Mango	5.88665	0.855	0.24455	1.02185
Mushrooms	3.90765	0.985	0.1543	0.61235
Mustard greens	2.4324	0.7478	0.3307	1.0756
Navy beans	1.6073	2.4692	0.3858	0.2511
Nectarines	2.3721	0.91	0.3197	0.8333

Okra	3.70475	0.8345	0.36375	1.6513
Onions	1.1062	0.9	0.3527	0.4335
Oranges	1.4624	0.68	0.4079	0.8771
Papaya	4.41145	0.81	0.23145	0.9105
Peaches	2.89215	0.98	0.3362	0.98395
Pears	1.8472	0.9	0.3638	0.7466
Pineapple	3.81515	0.755	0.25905	0.7612
Pinto beans	1.4173	2.4692	0.3858	0.2215
Plum	4.4276	0.97	0.2756	1.0818
Pomegranate	2.4672	0.56	0.3417	1.5055
Potatoes	0.8166	0.8113	0.2646	0.2663
Radish	1.8126	0.9	0.2756	0.555
Raspberries	6.9464	0.98	0.3252	2.30605
Red peppers	1.8742	0.82	0.2646	0.6047
Spinach	3.5074667	0.8486666	0.2866	1.169
Strawberries	3.15515	0.97	0.3252	1.0573
Sweet potatoes	1.1565	0.8818	0.4409	0.5782
Tomatoes	2.435	0.91	0.3748	1.0028333
Turnip greens	2.72095	0.763	0.3362	1.19535
Watermelon	0.382	0.52	0.3307	0.2429
Zucchini	1.6359	0.7695	0.3968	0.8437

All data processing, optimization design, simulation, and visualization occurred in a single Python Notebook, displayed in the following pages. Python was used because it connects to workflows developed for my research and allows this project to more easily scale into a proper tool with a user interface.