Problem Statement:
Consider the basic diet optimization problem. Let there be n foods from which you can construct your diet. Each food can supply you with m types of vitamins, minerals, calories, dietary fiber, and possibly other essentials. Each food item also has an associated cost.
(a) Research the data necessary to consider an linear model of a minimum cost diet that also satisfies USDA recommendations on nutrition. Consider n >= 50 foods and m >=12 nutritional requirements.
(b) Model the minimum cost diet problem as an LP, IP or MIP (as appropriate) in a standard matrix form appropriate for your software.
(c) Solve your initial model and discuss the results.
(d) Consider and implement two model improvements and discuss the results.

Research:

First we need to establish nutritional requirements, for both macro and micro nutrients, as recommended by the USDA. According to the USDA the amount of calories needed per day varies per person depending "on a number of factors, including the person's age, gender, height, weight and level of physical activities" (Dept. of Agriculture, 13).

I will be constructing this diet for myself and will use appendix 6 (page 78) of *Nutrition and Your Health: Dietary Guidelines for Americans* to approximate the total amount of calories I need to consume daily. According to appendix 6, I should be consuming 2,800 calories a day for an active 21-25 year old male. Instead of shooting for an exact amount of calories per day, I will formulate my diet to fit within a range of 2,600 and 3,000 calories, this better suits my life style since there are periods of inactivity mixed with days of running upwards of 6 hours.

Sources of caloric intake can be subdivided into three categories, known as macronutrients, and they are as follows; carbohydrates, fats (or lipids) and proteins. Though the USDA does recommend splitting caloric intake between these three categories of macronutrients, it is not required. When it comes to gaining, losing or maintaining weight, the total number of calories matter much more so than the source of such calories[1] (pg 13). For this reason, I will initially only concern myself with the total number of calories when formulating my diet. After I ensure the functionality of my linear program (or integer program) I will than concern myself with the source of my calories.

For micronutrients I will use appendix 5 (pg 76) of *Nutrition and Your Health: Dietary Guidelines for Americans* to determine the daily requirements my diet needs to satisfy. According to the USDA, some micronutrients to increase consumption of are potassium, dietary fiber, vitamin D and calcium, while it is advised to decrease consumption of sodium[1] (pg 40 - 42). The reasons may not be so obvious as to why certain micronutrients should enter one's diet more regularly while others should dissipate. This issue will not be covered but may be explored further in the *Nutrition and Your Health: Dietary Guidelines for Americans* book, pages 40-42.

Using appendix 5 and 6 I have determined the requirements that my daily diet needs to satisfy, the results are as follows:

	Calories	Calcium(mg)	Magnesium(mg)	Phosphorus(mg)	Potassium(mg)	Sodium(mg)
Requireme	2600-	1000	400	700	4700	2300

nts	3000					

	Vit. A (IU)	Vit. D (IU)	Vit. E (IU)	Vit. C (mg)	Thiamin (mg)	Riboflavin (mg)	Niacin (mg)	Vit. K (mcg)
Requirements*	3000	300	23	90	1.2	1.3	16	120

*Notice that in appendix 5 Vitamin A, D and E are given in mcg, mcg and mg respectively, though I have them listed in IU measurements. I used a conversion calculator (http://www.robert-forbes.com/public_html/index.php?option=com_content&view=article&id=61&Itemid=90) to convert mg/mcg to IU due to the fact that most sources list their nutritional values in IU. USDA suggests getting a minimum of 600 UI vitamin D a day, but they also state that majority of a person's daily vitamin D is created biologically when exposed to sunlight, therefore I will require my diet to provide half of my daily vitamin D.

I have finished the research of what I will be requiring in my diet, now all I need to do is research various food items that are plausible to eat. What I mean by the plausible criteria is the fact that I am forming this diet for myself, therefore I do not want to consider foods that I am not willing to eat. I am concerned about two main properties of food items, their cost and the amount of various nutrients a food item contains per some serving size.

I decided to concentrate on base foods that can be used in cooking and food preparation instead of refined/packaged foods and I also wanted to compose a selection of healthy foods. I was able to find a very detailed website (www.whfoods.com/foodstoc.php) on healthy food items and their corresponding nutritional value depending on the preparation method. Secondly I need an associated cost for each food item, the USDA provides a database of various foods and their average costs. The database can be found here:
http://www.cnpp.usda.gov/USDAFoodPlansCostofFood.htm

Table One summarizes the price and nutritional values of 50 different food items as discovered from the USDA database on food costs and the whfoods website on nutritional components. This table will be explained more thoroughly but for now notice that the bottom row of the table contains the requirements I have listed for my diet.

Modeling My Diet:

I wish to determine the cheapest diet that satisfies the daily nutritional requirements as recommended by the USDA.

Since the USDA list all of their requirements for daily consumption, I want my model to determine which foods and in what quantities I should be consuming per day. I will model my diet problem as a continuous linear program and not an integer program. There are various reason I decided to do this way; for one, my model will determine the amount of food to consume, not purchase, in order to satisfy nutritional requirements. Second, nutritional value within food is continuous, meaning if one eats only half a serving for a given food item they will receive half of the nutritional value listed for a serving of the specified food instead of none at all. Therefore consumption, as determined by my model, will be on the continuous real spectrum. Now this may work for consumption but most foods need to be purchased in prepackaged arrangements, meaning most food items need to be purchased in incremental amounts or integer amounts based on prepackaged sizes. We will see how to address this in a moment.

Since nutritional values for an item of food is measured in some arbitrary unit of measure, defined as a serving size, my model will then determine which foods I need to consume and in what quantities per defined serving size. One problem, our definition of serving size, right now we do not have one, I mean we know of the concept of a serving size but each food item has a different unit of serving size. In order for my model to function I need to set a universal serving size and if you look at Table One you will notice the column titled 'units' and the following values are all 100, this is my arbitrary, universal serving size, 100 grams. There are two main reasons why I choose to use 100 grams as my serving size; one, the USDA's database on food prices is given in dollars per 100 grams. Second, 100 grams is a nice arbitrarily small value, you will see why this is significant in a moment.

To recap once more, my model will specify which foods and in what quantities I should be consuming a day in order to satisfy the USDA's recommendations on nutritional requirements. This is fine in the standpoint of consumption but what about purchasing? This not as a big of a problem as imagined, all that is required is to purchase a package that contains enough substance to satisfy our model. Since the model will generate answers in terms of our defined serving size (100 grams) then all we need to do is ensure to buy packages that at least contain the amount of 100 gram servings that our model states we need to eat. This is where having a small serving size will help for it will be more concise and nicely divide prepackaged foods into sets of servings. Example, if our model states that I need to eat 6.81 beans a day, what it is really saying is that I need to eat 6.81 servings of beans or 681 grams. Grams is a nice, small concise unit of measurement that can easily be converted into ounces, pounds or other units of measurements. If a store only sell beans by the pound, roughly 454 grams, we would need to buy at least two pounds of beans to cover our daily requirements of beans. Now we could just buy two pounds of beans to cover our daily requirements or we could purchase the least common multiple of the amount required per day with the amount stored in one package, essential buying 3 pounds of beans will provide two days worth of eating. In the end it does not matter how much food we buy at one time as long as I consume the recommended servings a day (and I do not buy so much food at once that it spoils), then my price of consumption will remain constant per time and should closely resemble the cost as predicted by my model, which will be in terms of cost per day. Essentially the amount purchased at one time does not matter and will be determined by where an individual shops and how often the individual likes to shop.

The Model:

Min $Z = c^T x$
$Ax >= b$
$x >= 0$ and x is real

Where Z is the total cost a day in order to sustain a specific diet. c^T is a row vector consisting of the cost, in dollars per 100 gram serving, for each food item in our list of plausible foods. Matrix A is our matrix of constraints that must satisfy the b vector. Each row in the A matrix represents an equation in which its output must satisfy the inequality constraint against the corresponding value from the b vector. The b vector contains the values our diet must satisfy for nutritional needs; the amount of calcium, magnesium, vitamin A, Vitamin D and etc. The x vector consist of the amounts for each food item that should be consumed in 100's of grams. So in our table, x_1 is the amount of cooked lentils to be consumed in 100's grams.

Notice that in our Table One, the Price column is our c vector and therefore our row vector, c^T, it the transpose of this column and is printed directly below Table One. Next notice that the sum of the values in each column of Table One for x_1 through x_{50} must be equal to or be greater than the value listed in the requirements row for the specified column. This is the Ax >= b portion of our model where the requirements row is the transpose of our b vector and the values in the columns between calories and vit K is the transpose of our matrix A. Therefore all we need to do is transpose the values of columns calories through vitamin K to get our matrix A. For my matrix A, I decided to write the value for each vitamin as a decimal percentage of my daily requirement for that vitamin as specified by the requirements row of Table One. Therefore my b vector consist of all 1's for all vitamin requirements. The final A matrix and b vector can be seen printed below Table One. Notice that Table One consists of one column specifying the range requirements of my caloric intake but in my matrix A I need two rows to represent the range requirement of caloric intake; one for the lower bound of 2,600 and one for the upper bound of 3000. A summary of my constraints follows:

constraint one - the lower bound of caloric intake:

$\sum_{i=1}^{n} K_i x_i \geq 2600$ where i is from 1 to n foods in the list and K_i is the amount of calories per 100 grams of food x_i.

constraint two - the upper bound of caloric intake:

$\sum_{i=1}^{n} K_i x_i \leq 3000$

The remaining constraints for vitamins and minerals follows this form:

$\sum_{i=1}^{n} V_i x_i \geq 1$ where V_i represents the percentage of required daily value of a specified vitamin that food x_i contains per 100 gram serving.

Results:

I now have all the required information to solve my linear program, I will do so by using octave. The input I use for octave may be seen in the table1.m file. The results as determined by octave are as follows:

Octave suggests consuming on a daily basis:
$11.7753x_3$ or 1177.53 grams (2.6 pounds) of Pinto Beans.
$7.50188x_5$ or 750.19 grams (roughly 26 fluid ounces or 3 and one third cups) of 2% milk.
$0.0531767x_{38}$ or 5.32 grams of raw bell peppers.
$0.0294059x_{40}$ or 2.94 grams of raw carrots.
$0.0416042x_{44}$ or 4.16 grams of raw kale.
$3.56945x_{46}$ or 356.94 grams (roughly 3/4 of a pound) of baked potatoes with skin.

Octave states that this diet will cost me $1.59 a day or $12 a week or $45 a month, which is substantially lower (about $129 lower a month) than the average amount of money spent on food by the average 19 year old during the month of February 2011 as indicated by the USDA[4].

This is expected since we are forming the cheapest possible diet. All of the above values produced by octave do in fact satisfy all of the requirements and constraints I have set forth for my diet. Besides the need to eat an overwhelming amount of beans a day, this is a viable diet.

Issues:

One of the largest issues with this model is the fact that I am using data provided by the USDA for the average costs of food that is a couple of years old and outdated. This means that the data may not represent the actual price of the specified food in the current market. Now I did cross reference the price of the foods selected as my optimal diet with the actual prices observed in grocery stores and found that the price indicated by the USDA fits with observational data. Example: the data provided by the USDA suggest that milk costs around $2.60 a gallon, which fits within the margin of observational data obtained from WINCO, Walmart, Safeway and Dissmores.

Improvements:

There is plenty of room for improvement within our model. For one, we can add many more foods items to select from and two we can add more constraints/criteria for our diet to meet. Now that I have the model working, I am no longer satisfied with fulfilling my caloric intake and now concerned with the source of my calories.

Modification One - Diverse Caloric Intake:

All constraints from the original model remain but I will now add new constraints for my caloric intake. As recommended by the USDA, I want 45%-65% of my calories to come from carbs, 10%-35% coming from protein and 10%-20% coming from fat. In Table Two I have added 3 columns for carbohydrates, fat and protein per 100 gram serving of a specified food, all in grams. To convert to calories, 1 gram of carbohydrates accounts for 4 calories, 1 gram of protein accounts for 4 calories and 1 gram of fat accounts for 9 calories[1]. The following is what I have established for my caloric intake requirements for a 2,800 calorie diet:

	Grams	calories
Carbohydrates	315 - 455	1260 - 1820
Protein	70 - 245	280 - 980
fat	31.1 - 62.2	280 - 560

For my model I will refer to the constraints in caloric form for my matrix A and vector b. As one can see, Table Two currently represents the new constraints in gram form. To convert to caloric form for Matrix A and vector b, multiply the values in the columns carbohydrates, fat and protein by their corresponding conversion factor, 4, 9 and 4 respectively. Since each of these constraints is a range then we need to treat them like we originally treated our caloric constraint, by adding two constraints, one for the lower bound and one for the upper bound. The new constraints are summarized as follows:

	Lower bound	Upper bound
Carbohydrates	$\sum_{i=1}^{n} 4C_i x_i \geq 1260$	$\sum_{i=1}^{n} 4C_i x_i \leq 1820$

Fat	$\sum_{i=1}^{n} 9F_i x_i \geq 280$	$\sum_{i=1}^{n} 9F_i x_i \leq 560$
Protein	$\sum_{i=1}^{n} 4P_i x_i \geq 280$	$\sum_{i=1}^{n} 4P_i x_i \leq 980$

Carbohydrates: 4 is the conversion factor for grams of carbohydrates to calories, C_i is the amount of carbohydrates in grams per 100 grams of food x_i. The two other constraints can be read similarly.

Originally, I expressed my total number of calories as $\sum_{i=1}^{n} K_i x_i$ but now I can represent the total number of calories as $\sum_{i=1}^{n}(4C_i + 4P_i + 9F_i)x_i$ and therefore can represent my range constraint on the total number of calories as $2600 \leq \sum_{i=1}^{n}(4C_i + 4P_i + 9F_i)x_i \leq 3000$ and I do so in my new A matrix. The final A matrix and b vector can be seen directly below table 2.

Results for Modification One:

The input for octave can be seen in table2.m and the output is as follows:

$11.7672x_3$ or 1176.72 grams of Pinto Beans a day
$7.50188x_5$ or 750.19 grams of 2% milk a day.
$0.390967x_{28}$ or 39.1 grams of raw peanuts a day.
$0.0535429x_{38}$ or 5.35 grams of raw bell peppers a day.
$0.029318x_{40}$ or 2.93 grams of raw carrots a day.
$0.0416585x_{44}$ or 4.17 grams of Kale a day
$3.56526x_{46}$ or 356.53 grams of baked potatoes with skin a day.

This differs little from our original optimal diet, in fact the only new food that enters my diet is raw peanuts. Also the expected cost per day increases by less than a cent, our expected cost remains \$1.59 a day. Using the above values all of our constraints are still met and this is a valid diet.

Modification Two - Diverse Food Groups:

This modification will actually be built upon my last modification, the diversifying of caloric intake. Currently our optimal diet is pretty bland and does not implement variety. I am going to add three new constraints in order to add variety to my diet. I want my diet to consist of at least 300 grams of fruit, 500 grams of vegetables and 200 grams of meat a day. Luckily my data is already in order of categories so my new constraints are as follows:

Fruit : $\sum_{i=10}^{19} x_i \geq 3$

Vegetables: $\sum_{i=37}^{50} x_i \geq 5$

Meat: $\sum_{i=24}^{26} x_i \geq 2$

These new constraints can be seen in the last 3 columns of Table Three. The final A matrix and b vector can be seen directly below Table Three.

Results for Second Modification:

The input used for Octave may be seen in the table3.m file.

My optimal daily diet:
8.10351x_3 or 810.3 grams (1.78 pounds) of cooked Pinto Beans.
7.10128x_5 or 710.13 grams (25 fluid ounces) of 2% milk.
3x_{13} or 300 grams - roughly 2 Bananas.
2x_{26} or 200 grams of Turkey.
0.123568x_{27} or 12.36 grams of raw sunflower seeds.
0.0321267x_{40} or 3.21 grams of carrots.
0.0506821x_{44} or 5.07 grams of Kale.
4.91719x_{46} or 491.72 grams (1.09 pounds) of baked potatoes with skin.
The price is now $2.89 a day, which is still much lower than the average amount of money spent on food for a 19 year old male during the month of February of 2011 according to the USDA[4]

This is a feasible diet and does satisfy all daily requirements set forth by the USDA and in the end I will consume 2,618.11 calories a day.

Further Improvements:

Though the last diet produced by our model is feasible it may not be too reasonable, eating 2 pounds of cooked Pinto Beans a day sounds like a lot and so does consuming a pound of Potatoes a day. Both of these issues may be solved by setting more upper bound constraints for any one food item.

Modification Two did add a little variety to our diet but it did a poor job doing so. The model was able to provide us with 500 grams of vegetables but it did so primarily using Potatoes. What if we wanted 500 grams of vegetables a day spread over 6 different vegetable items. We could do this in a variety of way; for one, we could add upper bound to all vegetables so that the model will be forced to choose 6 different vegetables in order to obtain a 500 gram daily value. Or we could even solve another LP problem maximizing the number of vegetables used under a certain number of constraints.

Also what if I wanted to consume a certain amount of seafood per week and not necessarily consume seafood daily. We could divide the desired amount by 7 in order to obtain a daily amount and add that amount as a constraint to our model as is but it may be more advantageous for us to convert our model in terms of nutrients required per week. For one, we can simply add any weekly desired consumption of foods such as seafood or certain types of meats. Secondly, our model will generate the amount of food needed to be consumed in a week and therefore provide a better representation of how much food to buy at one time to last longer than a day.

Another issue with our model is with vitamins. There is no risk of over consuming water soluble vitamins but there is a health risk associated with over consuming fat soluble vitamins. In reality we would want our model to cap the amount of any fat soluble vitamin consumed a day. Also if this model were to be used by someone with allergies than the food list would need to be modified to reflect those allergies. Finally, one of the easiest ways to improve our model is simply add more food and nutritional value constraints to our table; we could include other vitamins, minerals and even more abstract nutritional values such as GI ratings or the amount of antioxidants a food contains.

1. *Nutrition and Your Health: Dietary Guidelines for Americans.* [Washington, DC]: U.S. Dept. of Agriculture, 2010. Web. 23 Mar. 2011. <http://www.mypyramid.gov/guidelines/PolicyDoc.pdf>
2. "USDA Food Plans: Cost of Food." *USDA Center for Nutrition Policy and Promotion*. Web. 14 Apr. 2011. <http://www.cnpp.usda.gov/USDAFoodPlansCostofFood.htm>.
3. "WHFoods." *The World's Healthiest Foods*. Web. 14 Apr. 2011. <http://www.whfoods.com/foodstoc.php>.
4. "USDA Food Plans: Cost of Food Cost of Food at Home." *USDA Center for Nutrition Policy and Promotion*. Web. 16 Apr. 2011. <http://www.cnpp.usda.gov/USDAFoodCost-Home.htm>.