

Food Networks

A Thesis Presented to the Faculty of the Economics Department of Cornell University in Partial
Fulfillment of the Requirements for the Degree of Bachelor of Arts in Economics with Honors

Eric Schulman

May 2017

Contents

1	Background	1
1.1	Related Work	1
1.1.1	Demographics, Policy and Nutrition	1
1.1.2	Spatial Data and Nutrition	1
1.1.3	Optimization Problem and Food	1
1.2	Simplex Algorithm	2
1.2.1	Computational Complexity	2
1.3	Minimum Cost Flow	2
1.4	Minimum Cost Cycles	2
1.5	Worked out Example	2
1.6	Primal Problem	4
1.7	Dual Problem	6
1.7.1	Characterizing Solutions	7
1.7.2	Software	7
2	Problem Specification	8
2.1	Reduction to Minimum Cost Flow	8
2.2	Dual Problem Formulation	9
2.3	Data Sources	9
2.3.1	Farms	9
2.3.2	Intermediate Processors	10
2.3.3	Stores	10
2.3.4	Edge Costs	12
3	Results	13
3.1	Onion Farms	13
3.2	Cherry Farms	13
3.3	Grape Farms	13
3.4	Cabbage Farms	13
4	Discussion	22
4.1	Robustness of Solutions	22
4.2	Comparing Different Results	22
4.2.1	Farms	25
4.2.2	Intermediate Processors	25
4.2.3	Stores	25
4.3	Checking the Uniqueness of Results	25
4.4	Transportation Problem Assumptions	28
5	Conclusion	29

Abstract

For my proposal, I ask whether towns centrally located in the network of farms, processors, and vendors enjoy lower costs and higher availability of fresh produce. To answer this question, I estimate prices for apples, grapes, cabbage, onions, and cherries based on transportation costs to each registered food retailer in New York as an optimization problem using linear programming software packages. To specify the problem, I use New York's publicly available geographic information system (GIS) data on agricultural land use, farm product dealers and food vendor permits. All of the code to used to complete this project is publicly available on GitHub.

Chapter 1

Background

1.1 Related Work

1.1.1 Demographics, Policy and Nutrition

Researchers have studied the statistical relationship between transfer programs, demographics and nutritional outcomes for a long time. Calculating these statistical relationships is straightforward for the US because the Census Bureau publishes a monthly current population survey monthly and nutrition supplement.

This literature finds reducing costs and increasing availability of nutritional foods improves health outcomes. One study evaluated the Farmers' Market Nutrition Program, a national program intended to increase consumption of fresh fruits and vegetables by providing coupons and information. Researchers found subsidizing nutritionally rich food with coupons leads to increased consumption when complemented with information [5]. The literature also found nutritional outcomes are related to systemic factors. For example, researchers analyzed nutritional outcomes in Oregon between 1999 and 2001, when Oregon had the highest average rate of hunger in the nation. The study found county-level factors like residential location and housing costs were most related to food insecurity [1].

1.1.2 Spatial Data and Nutrition

Researchers realized state-level data has its limitations. To understand systemic nutritional patterns, researchers began describing food supply in detail using spatial coordinates and mapping software. One study conducted among grocery stores across inner cities and suburbs within the Minneapolis and St. Paul metropolitan area concluded the urban poor pay marginally more for food because supermarket chains don't open in inner cities [2]. Researchers began applying the label "food deserts" to inner cities. With more data becoming available online, researchers continue describing food supply in more detail to understand systemic food insecurity. In one project, researchers leveraged Google Maps, and local online news to build a graphical representation of food availability in Bogota, Columbia [4]. Another study used online directories of grocery stores to index neighborhoods' vulnerability to food insecurity in the twin cities area [6].

1.1.3 Optimization Problem and Food

Food allocation is closely related with nutrition and often studied using a transportation problem. Transportation problems are an optimization problem designed to find the cheapest possible way of sending a certain amount of supply through a network. A typical application of this problem involves finding the best delivery route from a factory to a warehouse where the road network has some capacity and cost associated. Solving the problem involves finding a vector with an economic interpretation of optimal prices from a social planning perspective. It can be solved efficiently using the network simplex algorithm and linear programming software packages [3]. Researchers have often applied transportation problems to food allocation. One paper studies maize allocation in South Africa where rails carry maize from supply and

demand points, scattered throughout the country. Researchers minimize the total rail cost, as is standard, but add a secondary objective of distributing costs fairly among all users [7].

1.2 Simplex Algorithm

The simplex algorithm is a relatively common tool in optimization. It's like the ordinary least squares of optimization. As you can imagine, people have written textbooks about it describing its properties. Since this is an honors thesis in economics I'll explain the intuition for why it works. Essentially it finds a solution to the optimization problem that is the lowest maximum and highest minimum. These happen to coincide, so as a result you know you've found a maximum.

1.2.1 Computational Complexity

Another point to talk about is the computational complexity of my problem. Computer scientists often concern themselves with how efficient an algorithm is. However, measuring a program's efficiency by factors like how long it takes to run or how many lines of code it takes can vary with the computer and the compiler. It is also variable with the input to the program. As a result, in order to have a more universal measure of efficiency Computer scientists measure efficiency using run-time bounds. Basically, run-time bounds tell you an upper bound on the number of steps you need to solve the problem, given the input.

The simplex algorithm's run time bound is actually very bad as far as we know. It can be $O(2^n)$. Meaning that if there 2^n inputs, the number of instructions is some factor of 2^n . Proving that it has a low average run time bound is an area of active research. The general rule of thumb is that the simplex algorithm has a run time that is some polynomial of n .

1.3 Minimum Cost Flow

The minimum-cost flow problem (MCFP) is an optimization and decision problem to find the cheapest possible way of sending a certain amount of flow through a flow network. A typical application of this problem involves finding the best delivery route from a factory to a warehouse where the road network has some capacity and cost associated. The minimum cost flow problem is one of the most fundamental among all flow and circulation problems because most other such problems can be cast as a minimum cost flow problem and also that it can be solved very efficiently using the network simplex algorithm.

There can be multiple solutions to the algorithm

1.4 Minimum Cost Cycles

The dual definitely says the problem in a less straightforward way, the way to understand it is to understand that the algorithm has no minimum cost cycles. This is a condition for optimality.

In this case, a suitable set of prices are just the edge costs (1,2,1) The reduced cost of an edge (v, w) is $cp(v, w) = c(v, w) + p(v) - p(w)$. We can think of the reduced cost as the cost of buying a widget at v , shipping it to w and selling it there. Note that if $cp(v, w)$ is positive, we would therefore not ship the item from v to w . Using this definition, we can say that a price function is feasible for a residual graph if no residual edge has a negative reduced cost.

1.5 Worked out Example

I used the network simplex algorithm implemented in Gurobi, I've worked out a quick example of the minimum cost flow. Below is an example network of suppliers and demanders

We start by looking to match supply with demand and draw edges with according units of flow. This solution is not optimal.

This works out to a total cost of 14, we can see by labeling the edges with the cost they incur

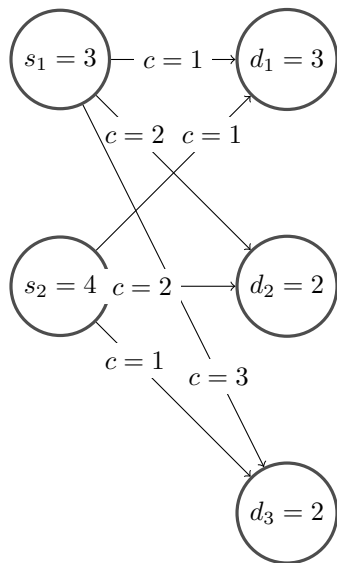


Figure 1.1: This shows something

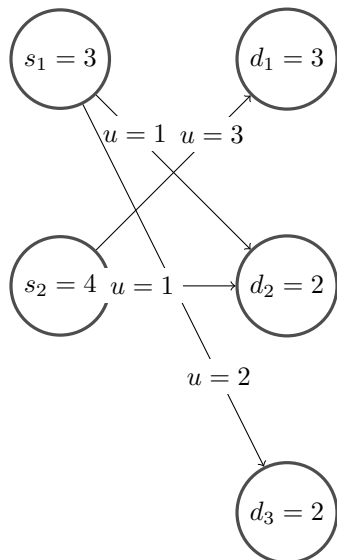


Figure 1.2: This shows something

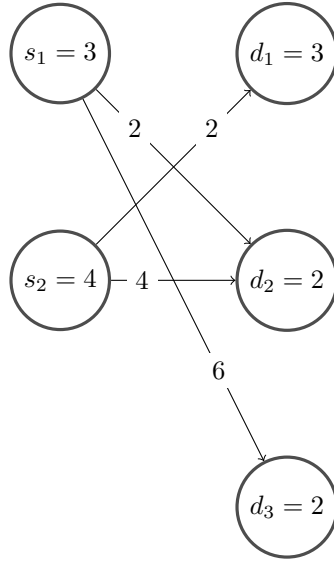


Figure 1.3: This shows something

This flow is not optimal, in order to improve it we find a reduced cost cycle, by traversing the cycle we can reduce the cost by 4 to the optimal amount.

There are no more reduced cost cycles, so we know our solution is optimal. Below is the result

Keep in mind that there are other possible solutions as 0 cost cycles exist in this simple network. Below would be one such example.

1.6 Primal Problem

As we stated before, the problem tries to send as much units of a good from a suppliers s in the set S to demanders d in the set D . However, there are costs for sending a unit of good across every route between suppliers and demand-ers. We want to minimize these costs which leads to our objective function (in the objective c is costs and u is units of the good).

$$\text{Minimize } \sum_{s,d \in \text{Routes}} c_{s,d} u_{s,d}$$

The constraints on the objective function reflects the fact demand must be satisfied. First we have that supply must be met

$$\sum_{s,d \in \text{Routes}} u_{s,d} = \text{demand at } d \text{ (for all demand-ers } s \in S)$$

The second constraint reflects the fact that no supply goes to waste. In other words, no product is just left at suppliers. In this case, supply is a negative quantity because supply leaves the supply nodes and flows to demand nodes.

$$\sum_{s,d \in \text{Routes}} u_{s,d} = \text{supply at } s \text{ (for all suppliers } d \in D)$$

Finally, we can't have negative amounts of units flow across the routes.

$$0 \leq u_{s,d}, \forall s, d \in \text{Routes}$$

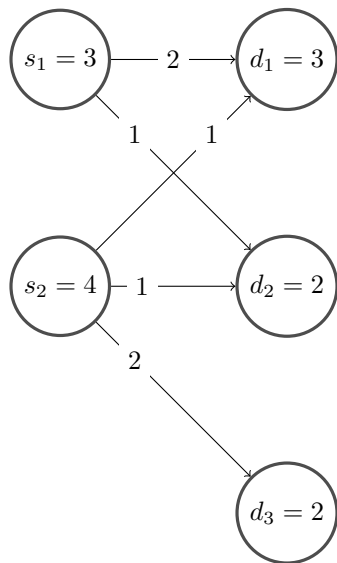


Figure 1.4: This shows something

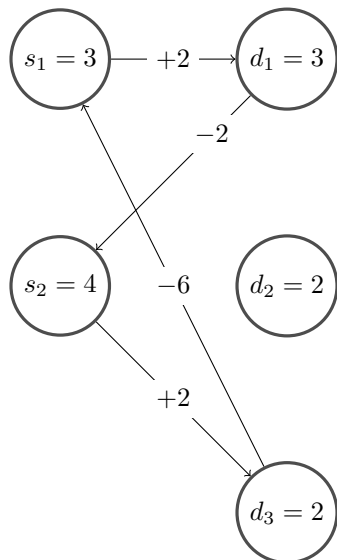


Figure 1.5: This shows something

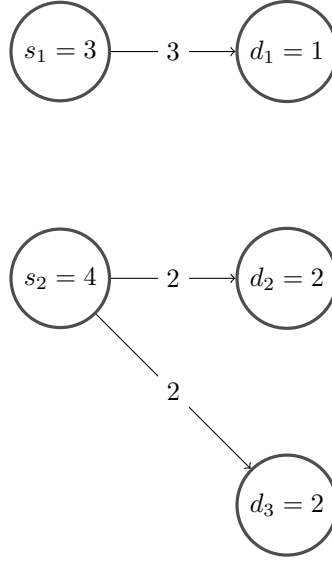


Figure 1.6: This shows something

Putting it all together we have

$$\begin{aligned}
 & \text{Minimize } \sum_{s,d \in \text{Routes}} c_{s,d} u_{s,d} \\
 & \text{subject to} \\
 & \sum_{s,d \in \text{Routes}} u_{s,d} = \text{demand at } d \text{ (for all suppliers } s \in S) \\
 & \sum_{s,d \in \text{Routes}} u_{s,d} = \text{supply at } s \text{ (for all demand-ers } d \in D) \\
 & 0 \leq u_{s,d}, \forall s, d \in \text{Routes}
 \end{aligned}$$

The objective function minimizes total costs. The first and second constraints ensure nodes either produce or consume based on their type; the third ensures supply moves forward from supply to demand.

1.7 Dual Problem

The intuition behind the dual problem comes from the fact that you are trying to get rid of augmenting cycles.

Below I state the problem formally.

The dual problem actually yields a more intuitive economic description of the system. Basically, instead of finding the amount of units sent between each pair of suppliers and demand-ers, we can look at the amount of flow going between suppliers and demanders.

$$\text{Maximize } \sum_{d \in D} (\text{demand at } d) \cdot p_d - \sum_{s \in S} (\text{supply at } s) \cdot p_s$$

Where p is a price. The dual variable y_v can be interpreted as optimal prices from a social planning perspective. The problem maximizes profits from a social planning perspective.

The constraint that arbitrage profits do not exist in the network. The price you pay at the demand node p_d , cannot exceed the price you pay at the supplier p_s plus the cost of sending the unit of good $c_{d,s}$. This has to do with the fact that the algorithm completes when the reduced costs cycles are exhausted.

$$-p_s + p_d \leq c_{s,d}, \forall s, d \in \text{Routes}$$

Putting it together we have

$$\begin{aligned} &\text{Maximize } \sum_{d \in D} (\text{demand at } d) \cdot p_d - \sum_{s \in S} (\text{supply at } s) \cdot p_s \\ &\quad \text{subject to} \\ &\quad -p_s + p_d \leq c_{s,d}, \forall s, d \in \text{Routes} \end{aligned}$$

1.7.1 Characterizing Solutions

As an economist, I would hope you are skeptical of any solution to this problem involving supply and demand found by the simplex algorithm. First of all, I am making point estimates as opposed to

The algorithm can actually return more than one solution. Going back to the previous example, You can see that the problem actually has more than one Solution.

1.7.2 Software

In order to actually solve the problem, I had to write a bit of code. Specifically, solving the optimization problem involved installing a software package that implemented the simplex algorithm. Additionally, since the results produce point estimates, a matrix with 5000 prices isn't readable. As a result, I wrote code to parse some descriptive statistics from the solution files.

I'm a big believer in open source code. As a result, I've made all the code I wrote for this project publicly available on GitHub with a detailed explanation about how to get it running. Since open source is important to me, most of the software dependencies that need to be installed in order to get my code working are open source. The one exception is the linear optimizer. In this case, I used Gurobi which is free with an academic license.

Chapter 2

Problem Specification

For each crop, New York has farms, agricultural dealers and food retailers. In the problem, farms send their goods to intermediaries and which in turn send their crop to stores purely based on transportation cost. Exports are assumed to be insignificant. Farms, intermediaries, and vendors incur transportation costs equals to the expected travel time between them. Farms can ship to any intermediary associated with the same crop and intermediaries can ship to any store. Intermediaries can scale as much as they desire. Farms produce agricultural products proportional to their size. Stores satisfy demand proportional to the market value of their square footage.

The I chose to look at the classification XX

I chose to group by census tracts, these are roughly municipal level.

You can see the amount of sq footage allotted to each tract

here you can see the stores, the color indicates their relative size. I've put the census tracts behind them below is the census tracts with the median household income.

2.1 Reduction to Minimum Cost Flow

A common thing to do in computer science is to change one complicated problem you may not know how to solve into a simpler one you can solve. When you prove the two problems are equivalent you have "reduced" the more complex problem to the simpler one. In my case, I want to show that the problem I've specified above can be specified as a linear program which I can solve with the simplex algorithm.

A transportation problem is a natural framework for New York's agricultural supply network. Supply nodes are farms and demand nodes are stores. The transportation problem assumes that supply and demand balance. Since New York exports agricultural products, I add a special demand node with demand equal to New York's net exports for each crop. Farms send goods to this node at no cost. This is equivalent to my original problem which also ignored exports.

$$\begin{aligned} & \text{Minimize} \quad \sum_{f,p \in \text{Routes}} c_{f,p} u_{f,p} + \sum_{p,s \in \text{Routes}} c_{p,s} u_{p,s} \\ & \text{subject to} \\ & \quad \sum_{f,p \in \text{Routes}} u_{f,p} = \text{supply at } f \text{ (for all farms } f \in F) \\ & \quad \sum_{f,p \in \text{Routes}} u_{f,p} + \sum_{p,s \in \text{Routes}} u_{p,s} = 0 \text{ (for all processors } p \in P) \\ & \quad \sum_{p,s \in \text{Routes}} u_{p,s} = \text{demand at } s \text{ (for all stores } s \in S) \\ & \quad 0 \leq u_{s,d}, \forall s, d \in \text{Routes} \end{aligned}$$

All we need to do is

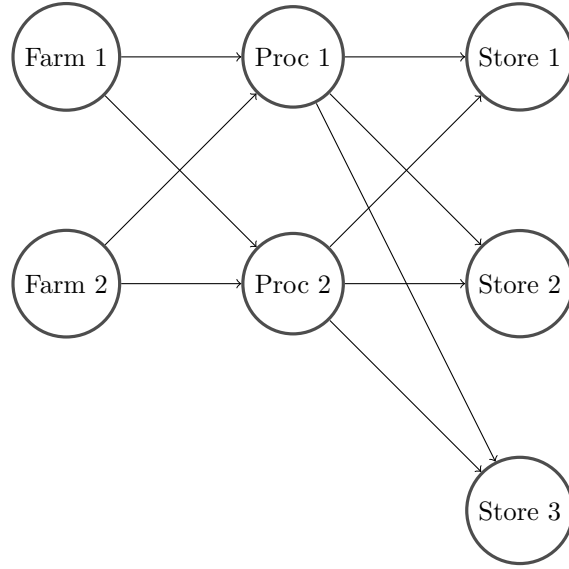


Figure 2.1: This shows something

2.2 Dual Problem Formulation

Of course, I will be solving the dual problem, not the primal to find prices. The dual problem I've formulated is

$$\begin{aligned} & \text{Maximize } \sum_{s \in S} (\text{demand at } s) \cdot p_s - \sum_{f \in F} (\text{supply at } f) \cdot p_f \\ & \text{subject to} \end{aligned}$$

We need two sets of constraints, one for the edges between the farms and processors, and another for the edges between the processors and stores. They are stated below.

$$\begin{aligned} -p_f + p_p &\leq c_{f,p}, \forall f, p \in \text{Routes} \\ -p_p + p_s &\leq c_{p,s}, \forall p, s \in \text{Routes} \end{aligned}$$

2.3 Data Sources

2.3.1 Farms

My data on acreage and location comes from a GeoTIFF image created by satellite pictures from the National Agricultural Statistical Service. The file forms a grid. Each square represents a 900 square meter plot labeled according to the crop growing. I will convert continuous stretches of pixels into polygons representing farms and assign locations to farms based on each polygon's centroid. I normalize the size of farms based on their size in pixels as a share of each crop's output. For some crops, the pixel accuracy can be as low as 50 percent. I am looking into ways to improve this. Landsat represents the world's longest continuously acquired collection of space-based moderate-resolution land remote sensing data. Four decades of imagery provides a unique resource for those who work in agriculture, geology, forestry, regional planning, education, mapping, and global change research. Landsat images are also invaluable for emergency response and disaster relief. The USDA, NASS Cropland Data Layer (CDL) is a raster, geo-referenced, crop-specific land cover data layer. The 2010 CDL has a ground resolution of 30 meters. The CDL is produced using satellite imagery from

Type	Band	Correct Pixels	Accuracy	Omission Error
Onions	49	900	87.04	12.96
Cherries	66	45	22.50	77.50
Grapes	69	2254	66.16	33.84
Cabbage	243	1067	54.97	45.03

Table 2.1: Another table caption

the Landsat 5 TM sensor, Landsat 7 ETM+ sensor, and the Indian Remote Sensing RESOURCESAT-1 (IRS-P6) Advanced Wide Field Sensor (AWiFS) collected during the current growing season. Some CDL states used additional satellite imagery and ancillary inputs to supplement and improve the classification.

The Cropland Data Layer (CDL) is produced using agricultural training data from the Farm Service Agency (FSA) Common Land Unit (CLU) Program and non-agricultural training data from the United States Geological Survey (USGS) National Land Cover Dataset 2001 (NLCD 2001)

2.3.2 Intermediate Processors

The connection between produce and nutrition isn't the only reason I focus my analysis on produce. Produce isn't drastically altered during intermediate processing. According to New York States' records on agricultural supplier permits there are roughly 300 licensed agricultural product dealers. New York State requires farm product vendors who buy more than 10,000 dollars in farm products to register for a license annually. By registering they disclose what crops they sell and their location. I use this information to include intermediate nodes in the transportation problem for their corresponding crop. There are actually a few from out of state, I've dropped these observations

2.3.3 Stores

New York State keeps a ledger detailing the locations and sizes of farmers markets and stores. There are roughly 700 farmers markets held in the state. This data includes GPS coordinates. The food retail establishment data includes 30,000 registered food vendors. For my analysis, I exclude certain classifications of stores like bakers and meat markets. The data includes labels for filtering purposes. I use stores' addresses to determine their GPS coordinates using MapQuest. In my model, stores' share of total consumption equals the cost of their square footage as share of the total. The data on property values comes from the Census's 2010 American Community Survey.

The store data set has roughly 20,000 different stores with the classification involved with selling produce.

This map show, all of the stores mapped. They've been overlayed to the census tract so you can get a sense of how many of them there are

This map shows the distribution of incomes through out the census tracts.

So as not to have a ridiculous number of edges between the stores and the intermediaries, I've actually aggregated the number of edges by Census district. The reason for using this method to aggregate stores is because I am multiplying square footage of the stores by the property to more appropriately weight the value of the square footage (i.e. a store in new york city might be the same size as a store upstate, but it you might expect it to sell significantly more produce even though they are the same size. This is because the population density is much tighter in new york and the consumers might have more income to spend on produce in New York). I'm essentially assuming that when the stores choose how big they are, they value the properties based on how much they can how much one city versus anther translates into sales and they choose an optimal size appropriately. This is by no means meant to be an exact estimate and specifying a problem where store owners choose optimal square footage to meet demand would be a lot more complicated, it's only supposed to be better than a naive estimate of area by its self.

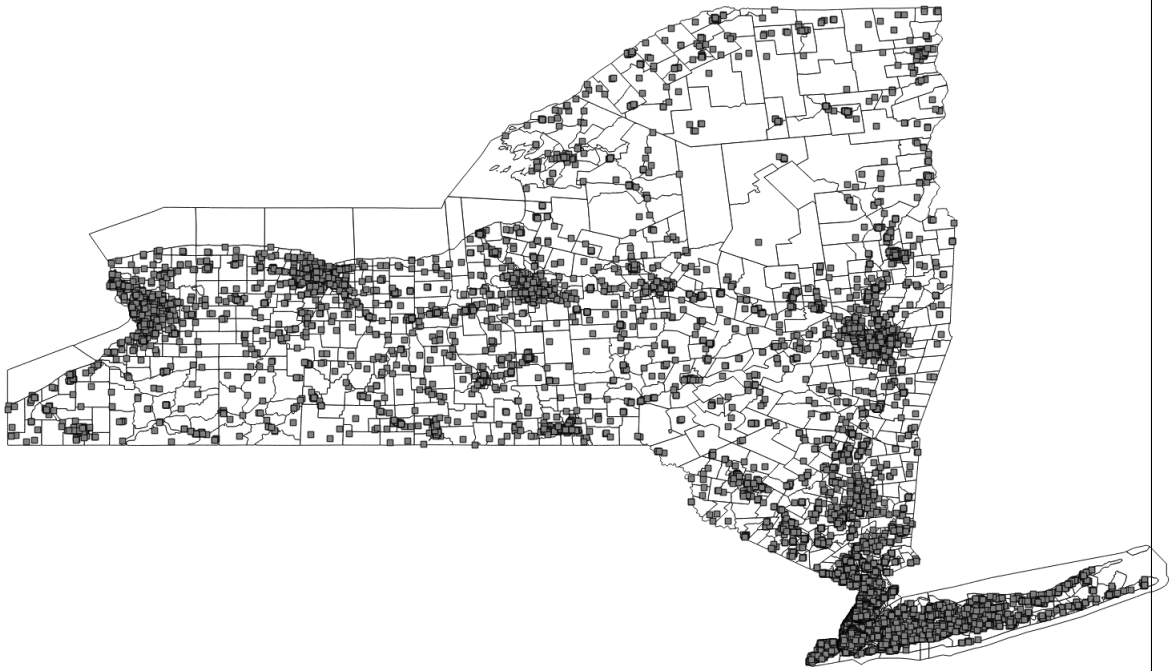


Figure 2.2: This shows something

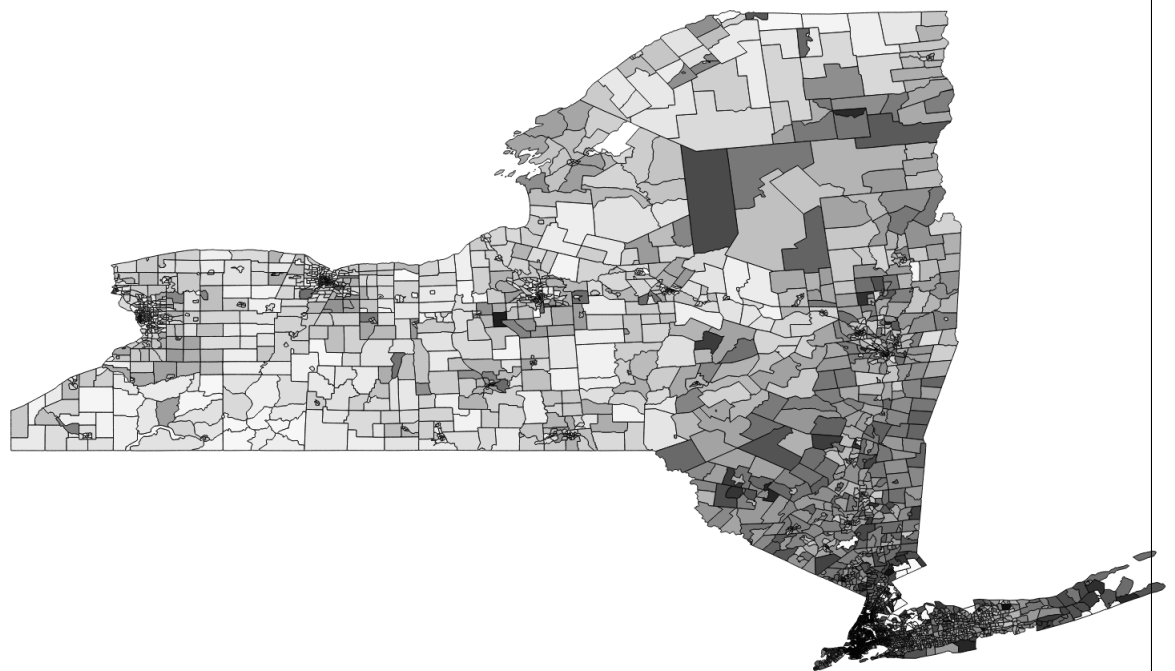


Figure 2.3: This shows something

Band	Average	Minimum	Maximum	Variance
15001.412749004	0	855900	372946.551970531	45.03

Table 2.2: Another table caption

Band	Average	Minimum	Maximum	Variance
49	14999.9461276334	51	35096.1	72442100.7350631
66	16938.8480737609	226.2	38904.2	107387638.261423
69	19756.949971294	21.4	42957.3	132807227.789542
243	16628.4401627168	64.5	39008.3	113068066.485

Table 2.3: Another table caption

2.3.4 Edge Costs

Basically, I set up a local instance of Open source routing machine and used OSRM. The prices are measured in seconds. The software loaded in a map and then it acts as a server that does routing. I let it run on my computer, connect over the local network and it responds.

In applied mathematics, the method of contraction hierarchies is a technique to speed up shortest-path routing by first creating precomputed "contracted" versions of the connection graph. It can be regarded as a special case of "highway-node routing".

Contraction hierarchies can be used to generate shortest-path routes much more efficiently than Dijkstra's algorithm or previous highway-node routing approaches,[1] and is used in many advanced routing techniques. It is publicly available in open source software to calculate routes from one place to another.

The speeds are based on expected times on each road.

Band	Average	Minimum	Maximum	Variance
49	15573.1896664454	12.1	43438.7	104114168.314326
66	16217.1197402983	12.1	43438.7	108504668.974705
69	16861.092352043	4.6	43438.7	101000630.46157
243	16205.5756165641	3.2	43438.7	108158531.635644

Table 2.4: Another table caption

Chapter 3

Results

3.1 Onion Farms

3.2 Cherry Farms

3.3 Grape Farms

3.4 Cabbage Farms

Type	Average	Variance	Deviation
Farms	2.382222222222	399.223332163	19.9805738697
Processors	2516.51147541	10468640.2301	3235.52781321
Store	5073.48738715	9636831.91217	3104.32471114

Table 3.1: Another table caption

Type	Max Price	Max County	Min Price	Min County
Farms	209.6	Orange County	0	Jefferson County
Processors	17243.3	Clinton County	0	Ontario County
Store	17798.3	Clinton County	171.4	Orleans County

Table 3.2: Another table caption



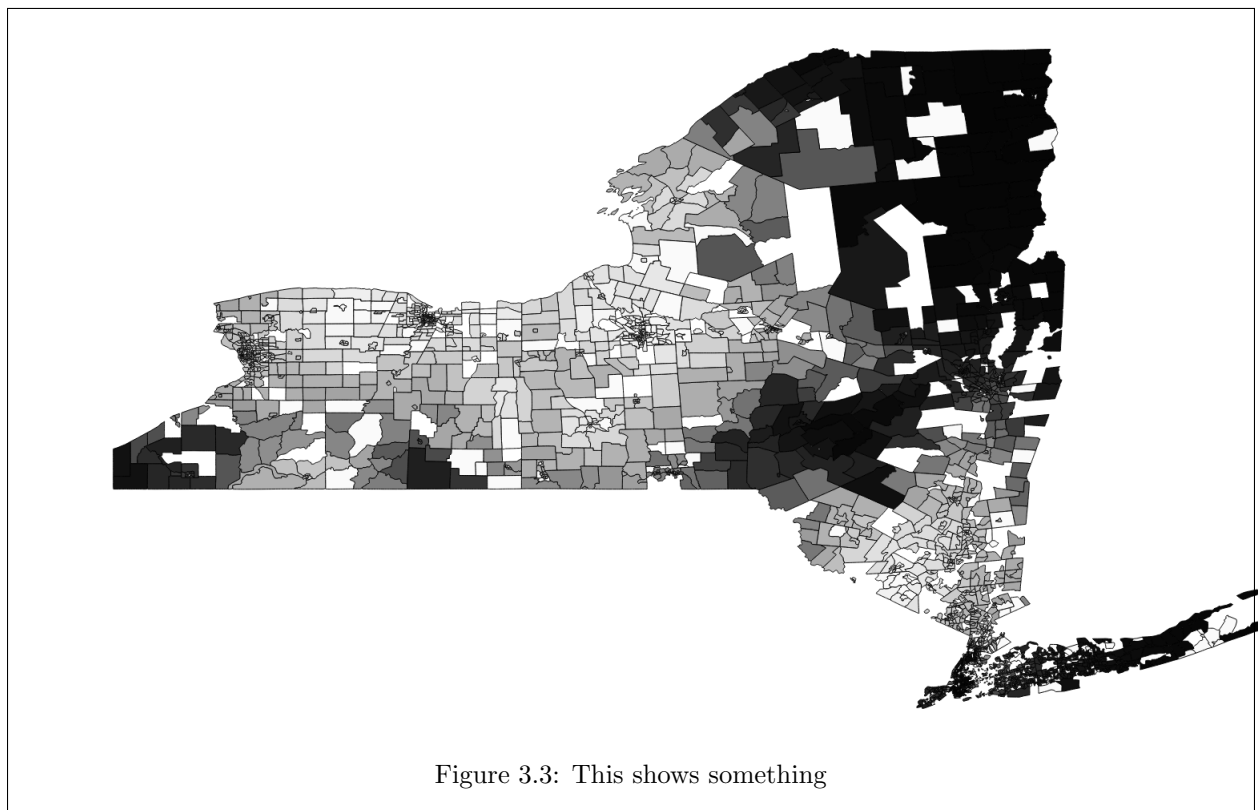
Figure 3.1: This shows something

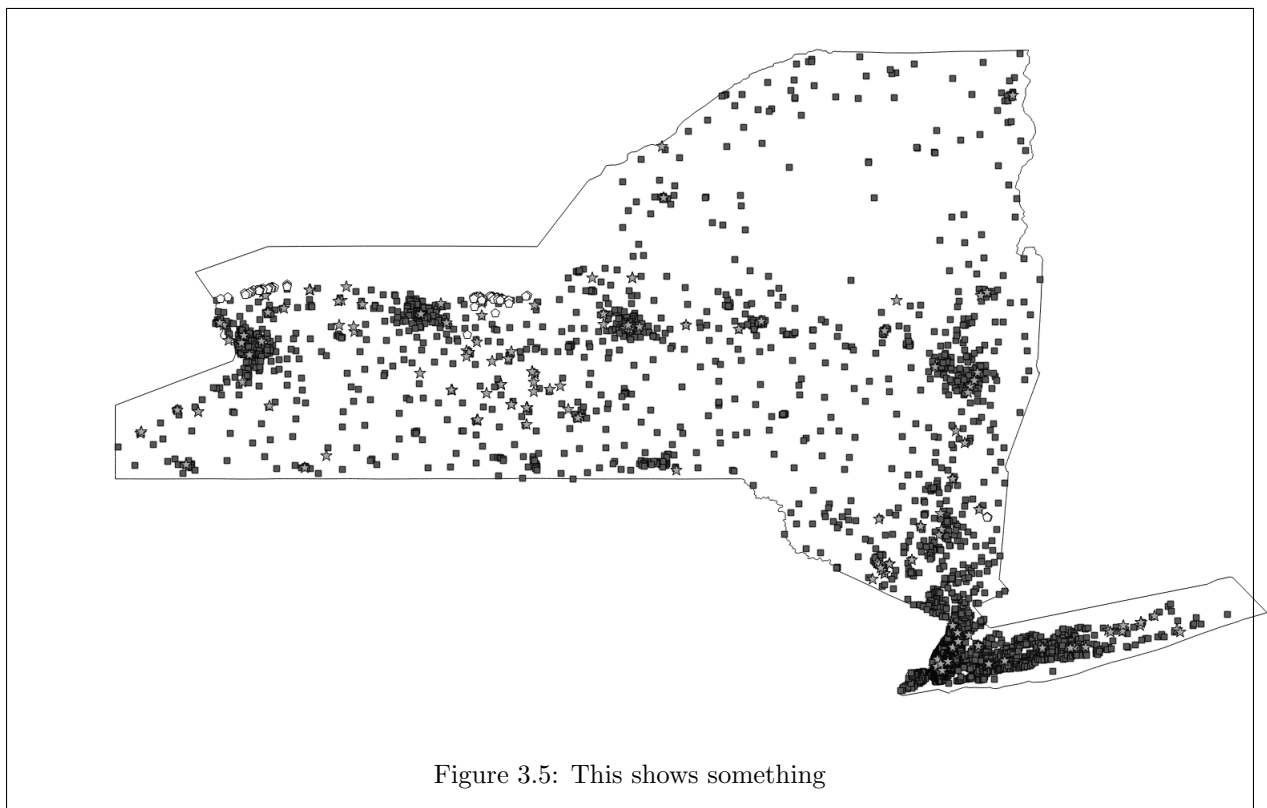
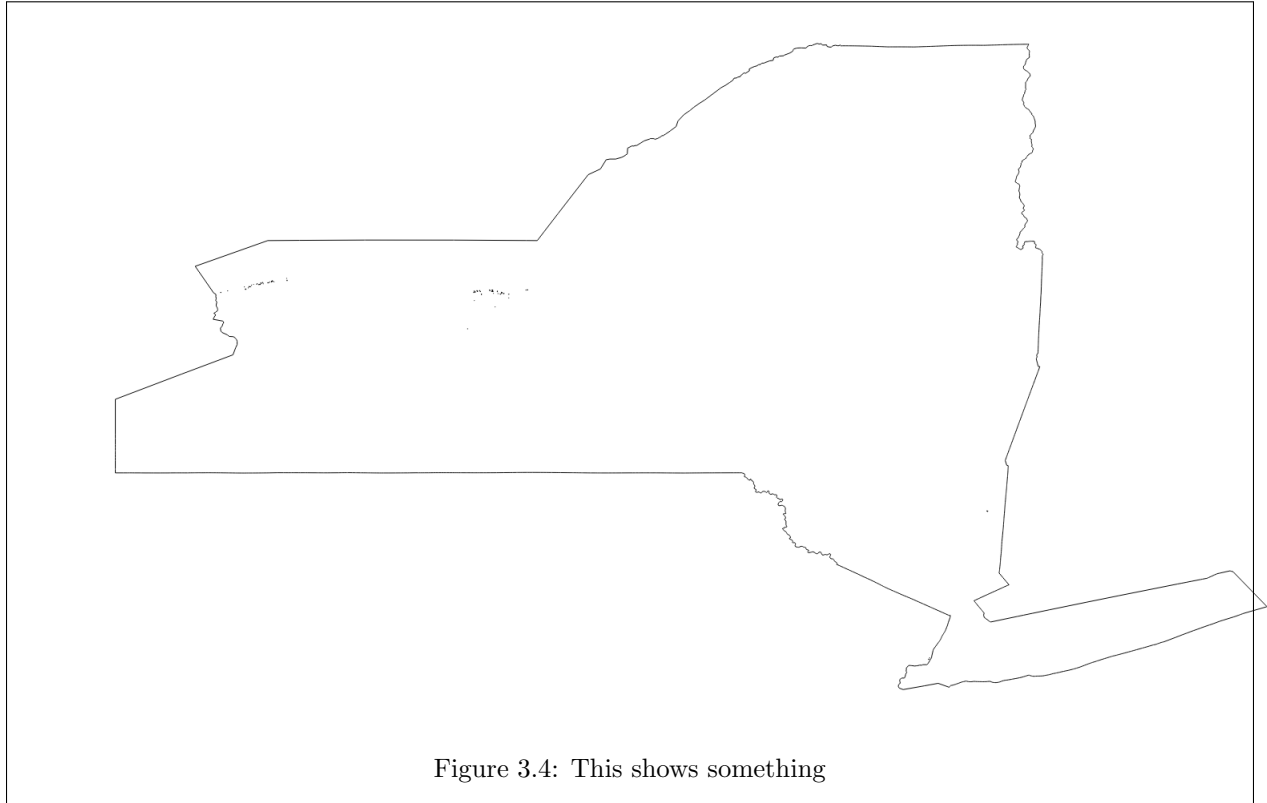
Type	Average	Variance	Deviation
Farms	0	0	0
Processors	3197.00382514	14557042.9968	3815.36931329
Store	6874.11295805	13268916.3606	3642.65237988

Table 3.3: Another table caption

Type	Max Price	Max County	Min Price	Min County
Farms	0	Niagara County	0	Niagara County
Processors	18308.3	Clinton County	0	Ontario County
Store	23301.8	Clinton County	311.5	Wayne County

Table 3.4: Another table caption





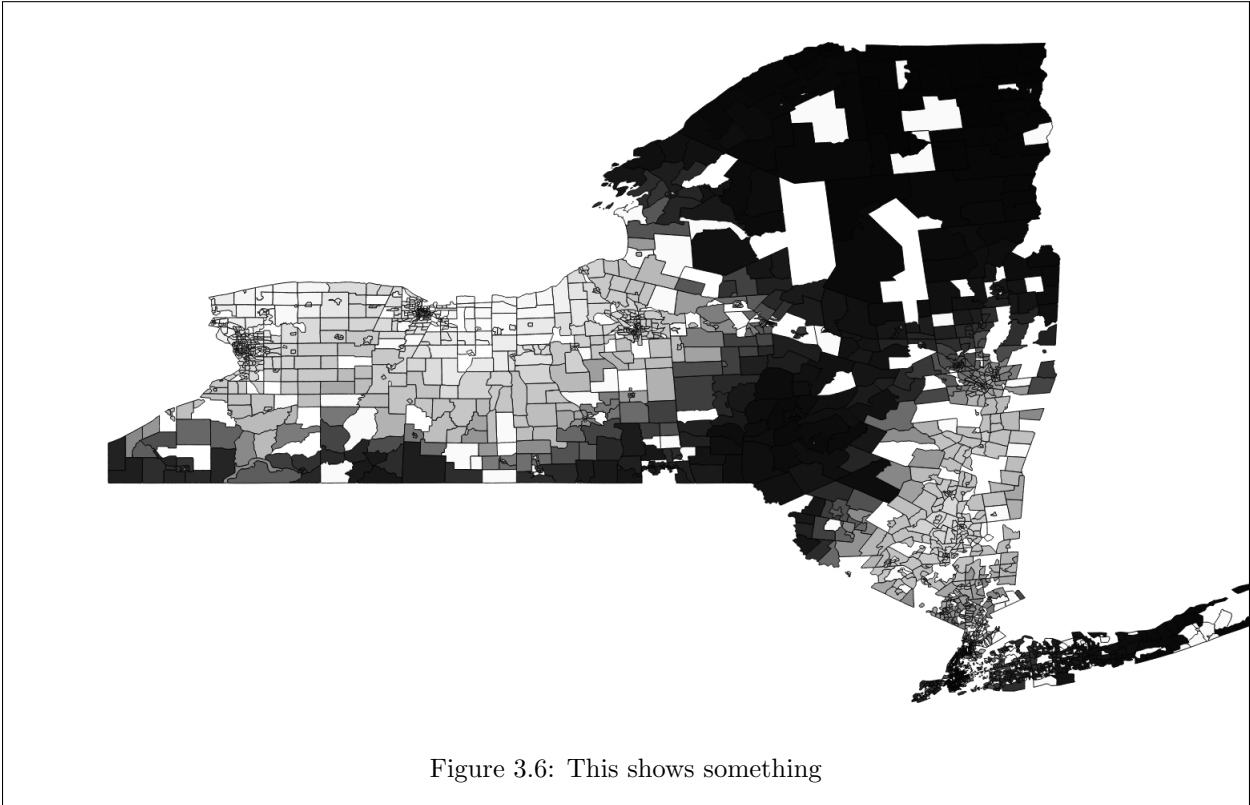


Figure 3.6: This shows something

Type	Average	Variance	Deviation
Farms	26.484884646	29195.7287858	170.867576754
Processors	4271.8295082	27027616.4373	5198.80913646
Store	7159.60167286	21412615.3937	4627.37672918

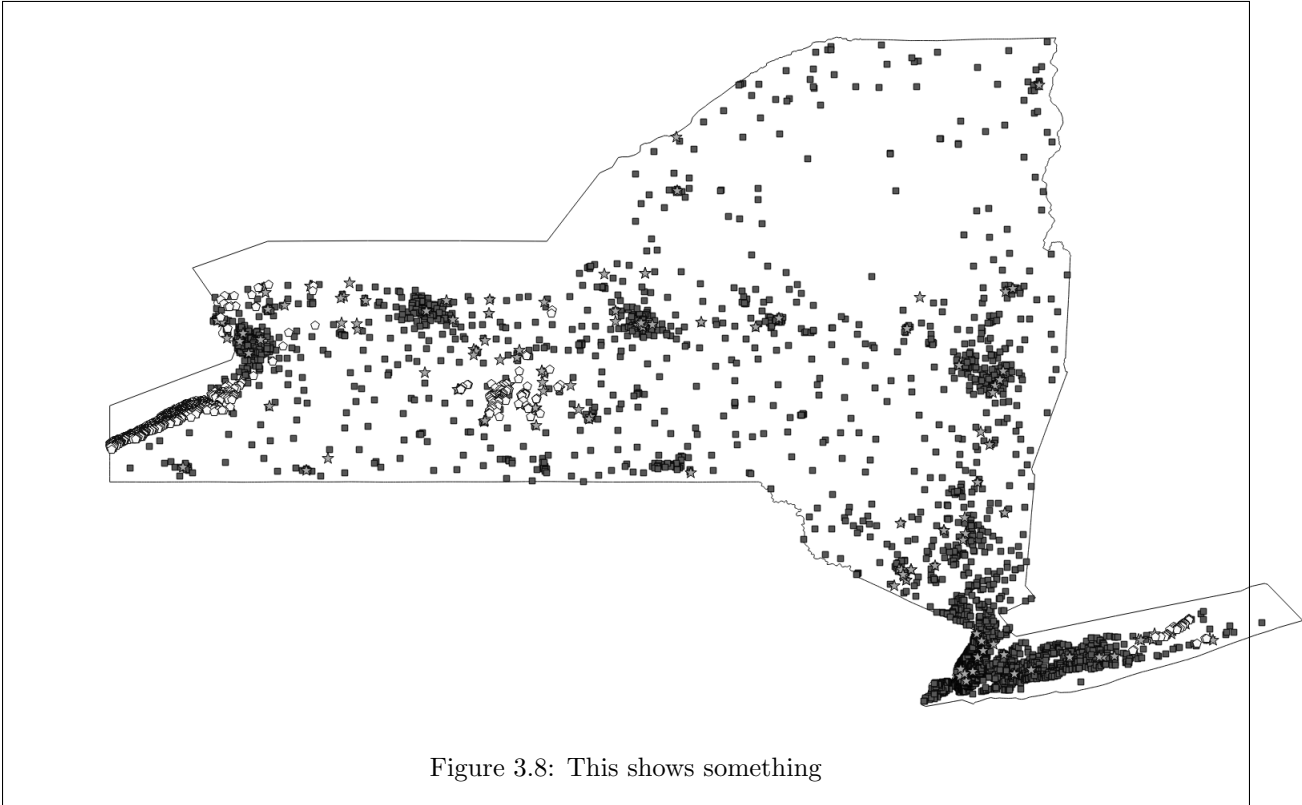
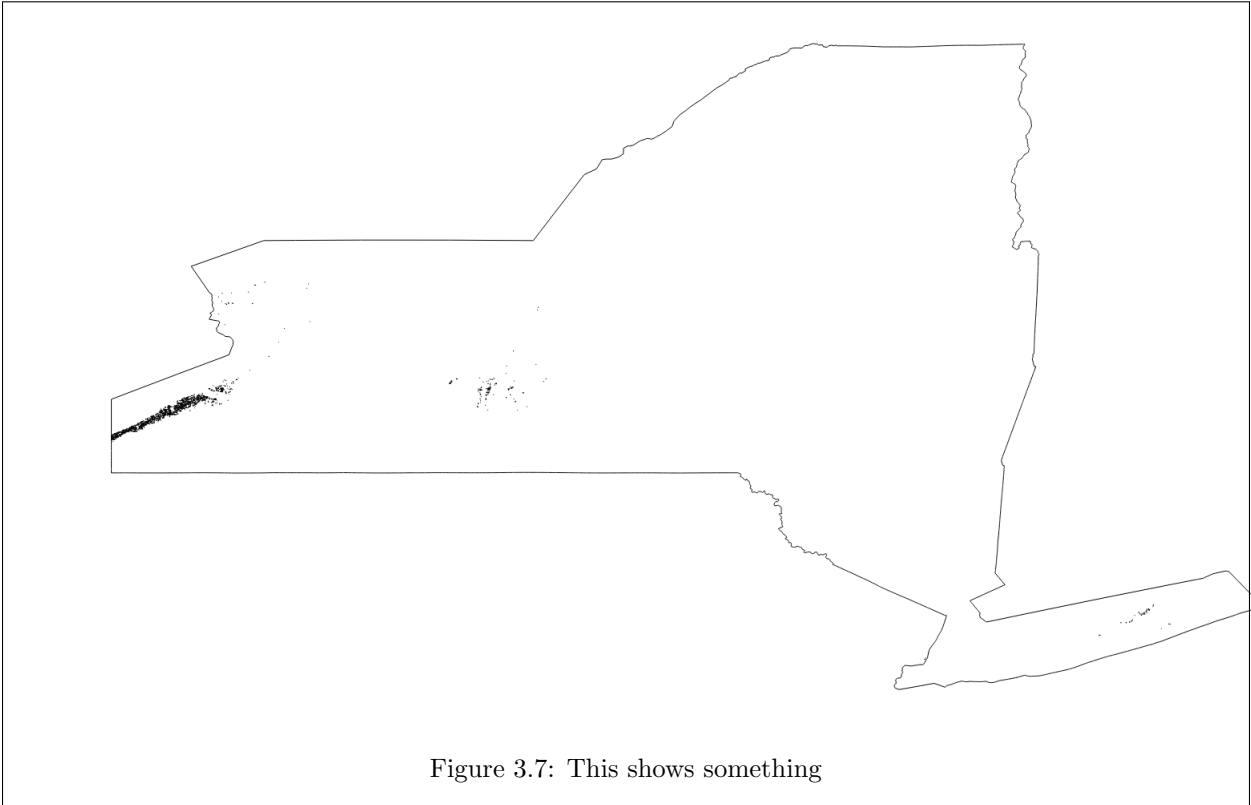
Table 3.5: Another table caption

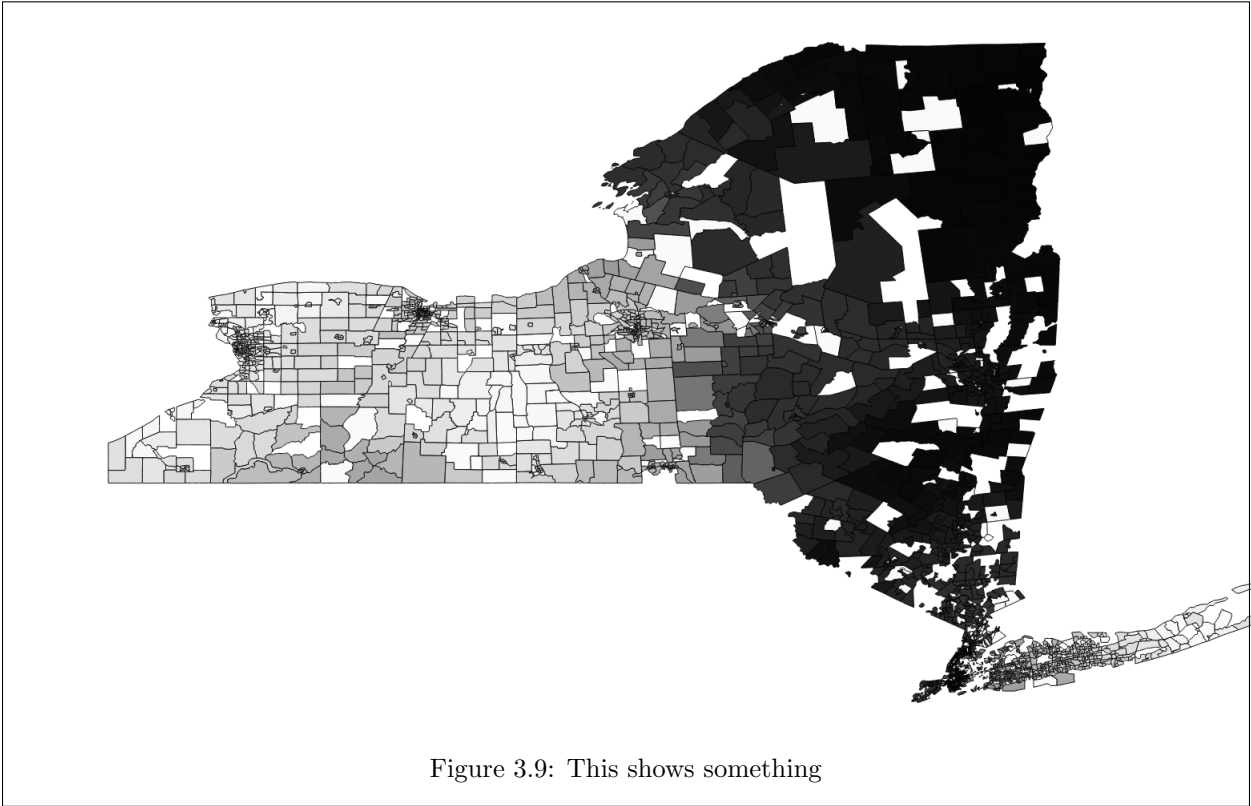
Type	Max Price	Max County	Min Price	Min County
Farms	2611.5	Suffolk County	0	Niagara County
Processors	23951.8	Clinton County	0	Suffolk County
Store	24858.6	Clinton County	123.6	Chautauqua County

Table 3.6: Another table caption

Type	Average	Variance	Deviation
Farms	25.6446601942	46718.515814	216.144664089
Processors	7088.95081967	95240598.712	9759.12899351
Store	15549.7135157	107420104.717	10364.3670678

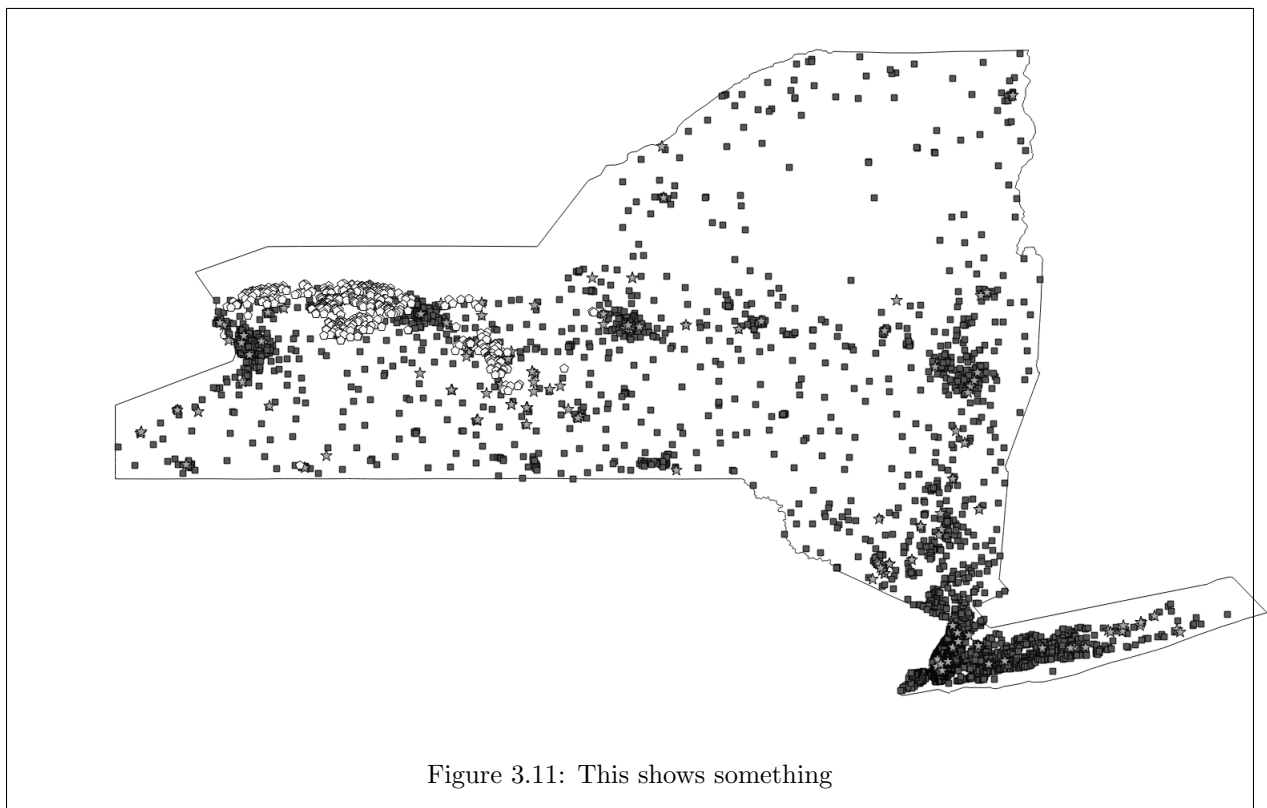
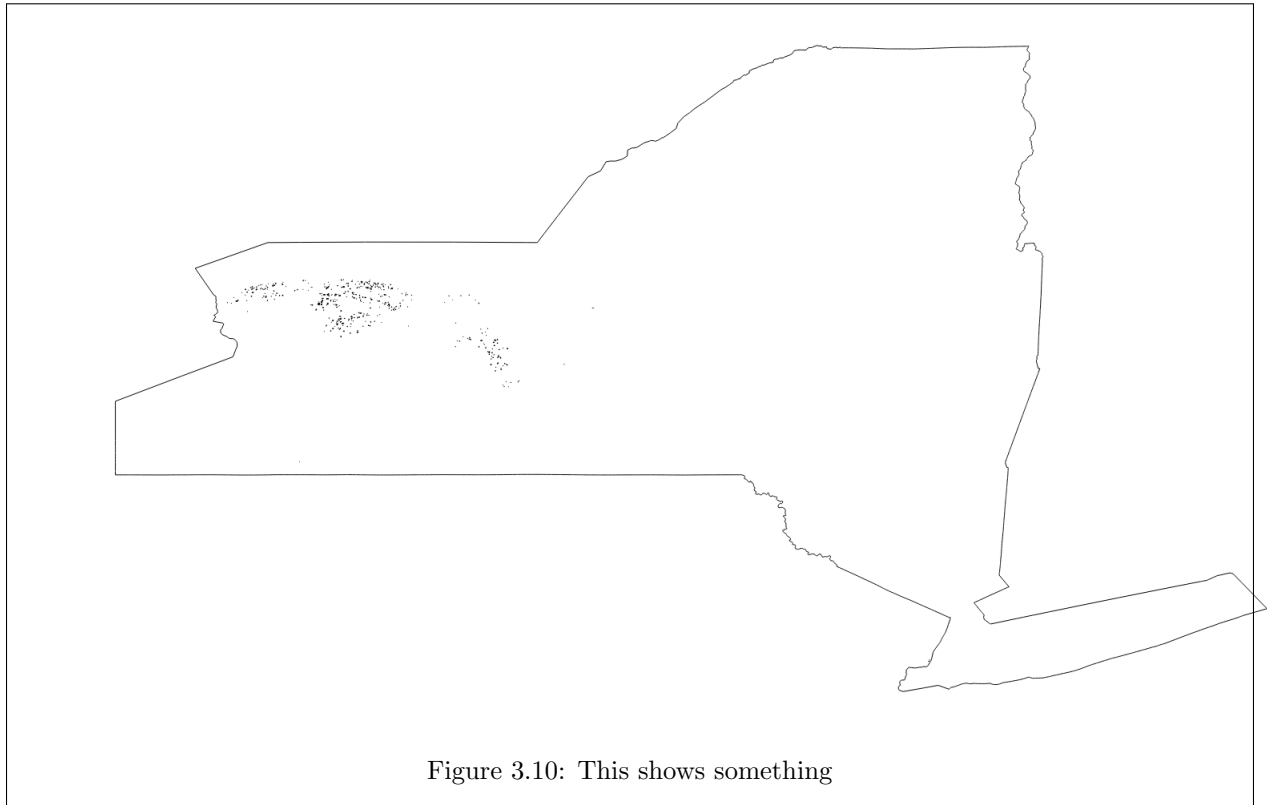
Table 3.7: Another table caption

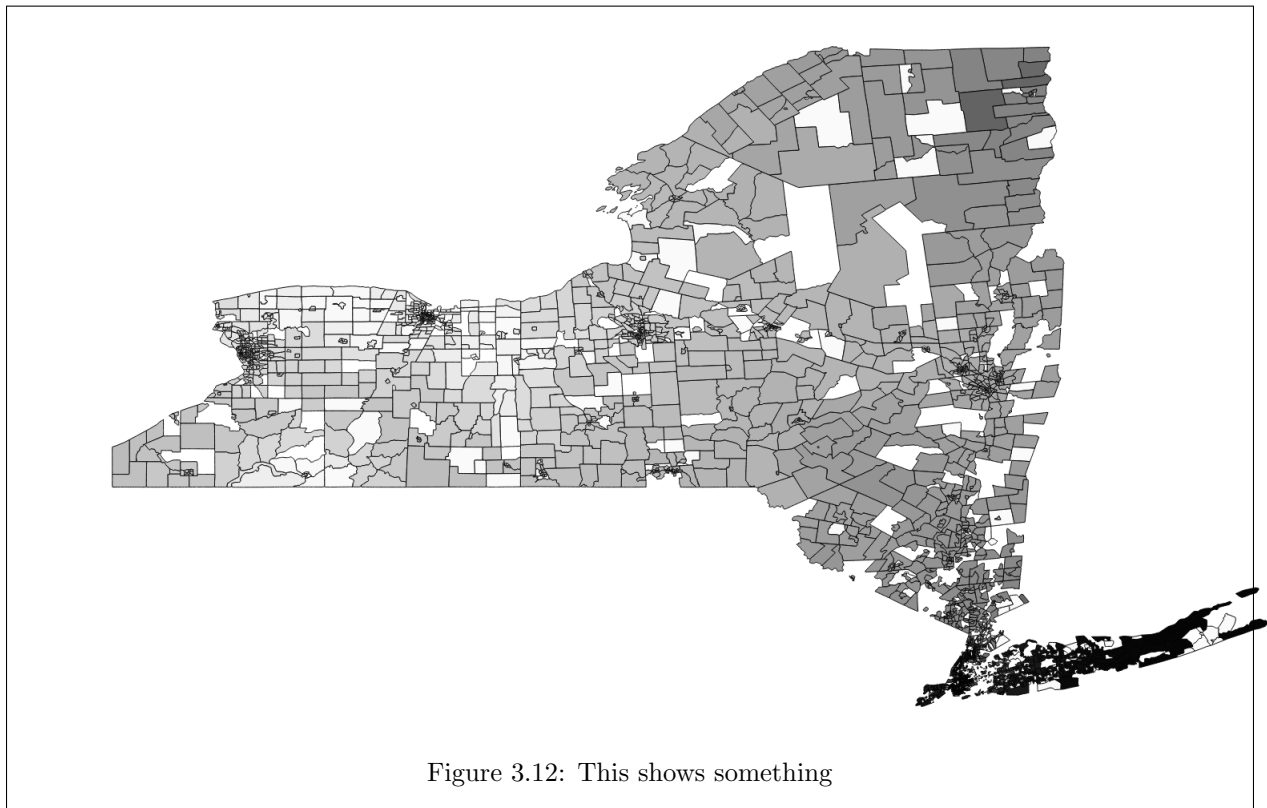




Type	Max Price	Max County	Min Price	Min County
Farms	2844.2	Onondaga County	0	Orleans County
Processors	29180.1	Suffolk County	0	Ontario County
Store	32869.4	Suffolk County	237.6	Orleans County

Table 3.8: Another table caption





Chapter 4

Discussion

4.1 Robustness of Solutions

I tried to perturb the solution to the linear program. Greater than the maximum Less than the maximum less than the minimum greater than the minimum

Although it changed the results it didn't really change things drastically. To be honest, I think the changes have more to do with rounding errors related to the way that python stores floating point numbers rather than actual differences to the solution. When the value of the objective function is parsed from the solution file into a python float the last decimal point may not be accurate. As you can expect the numbers didn't differ drastically.

I've included a table of the differences between table XX and table YY as a quick reference so you can get a sense of what I found. I can provide the other difference tables as .csv files upon request if you want to look through them in detail. They can also be generated using the git repository.

4.2 Comparing Different Results

So, all the results ended up looking pretty similar. I guess that happens because of the fact that processors and stores stay pretty the much the same for each of the goods.

So, cabbages actually being interesting. The variance in the prices calculated by the algorithm is a lot lower. Considering the that store and for the most part intermediate processors stay constant between the various bands in the problem, it becomes obvious the aspect of the network that changed is the farms.

As you can see there are more cabbage farms. They are more spread apart and there is less variance in terms of their size.

Band	Total	Average	Minimum	Maximum	Variance
49	38663100	143728.996282528	9000	8073000	331676621761.446
66	1624500	18889.5348837209	9000	78300	190736053.272039
69	89541900	71291.321656051	9000	1542600	16951135856.215
243	24317100	47309.53307393	9000	468000	3699027710.67693

Table 4.1: Another table caption

Band	Average	Variance	Deviation
49	2.38222222222	399.223332163	19.9805738697
66	0	0	0
69	26.484884646	29195.7287858	170.867576754
243	25.6446601942	46718.515814	216.144664089

Table 4.2: Another table caption

Band	Max Price	Max County	Min Price	Min County
49	209.6	Orange County	0	Jefferson County
66	0	Niagara County	0	Niagara County
69	2611.5	Suffolk County	0	Niagara County
243	2844.2	Onondaga County	0	Orleans County

Table 4.3: Another table caption

Band	Count
49	120
66	99
69	147
243	110

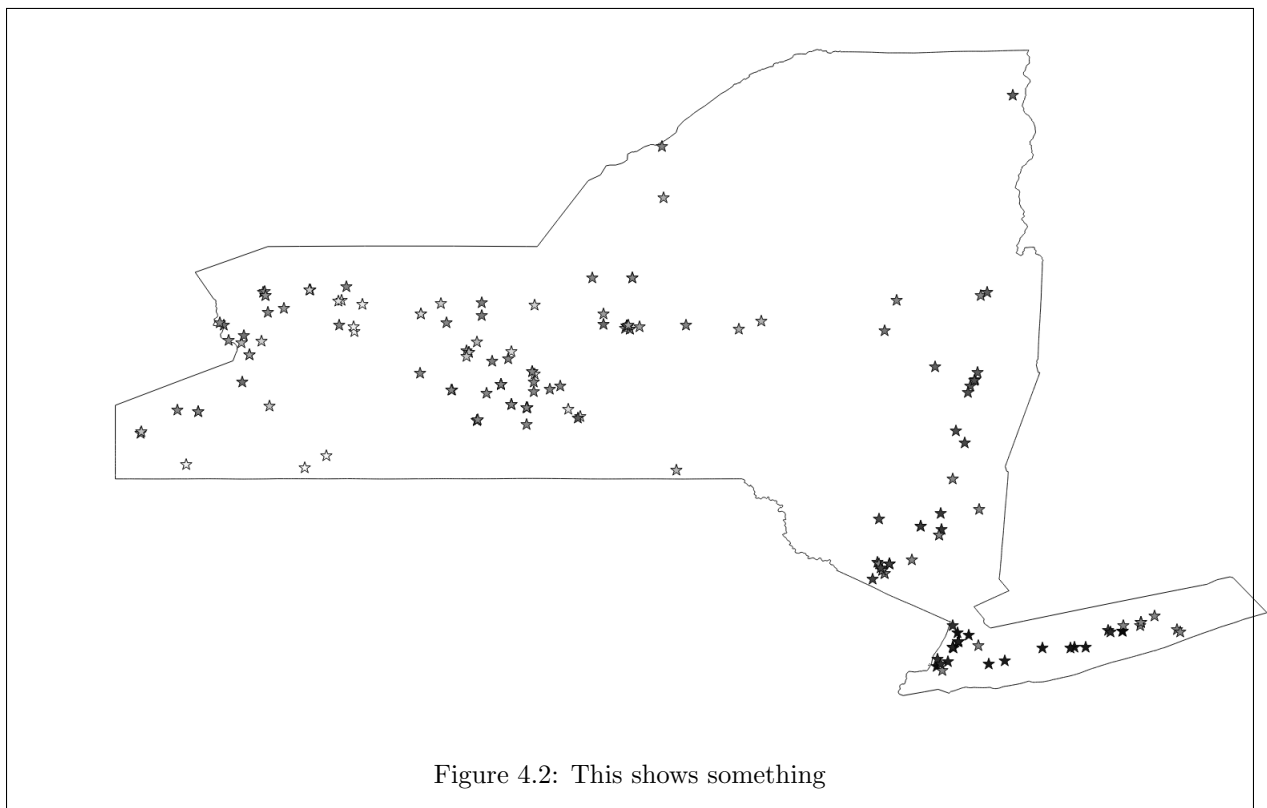
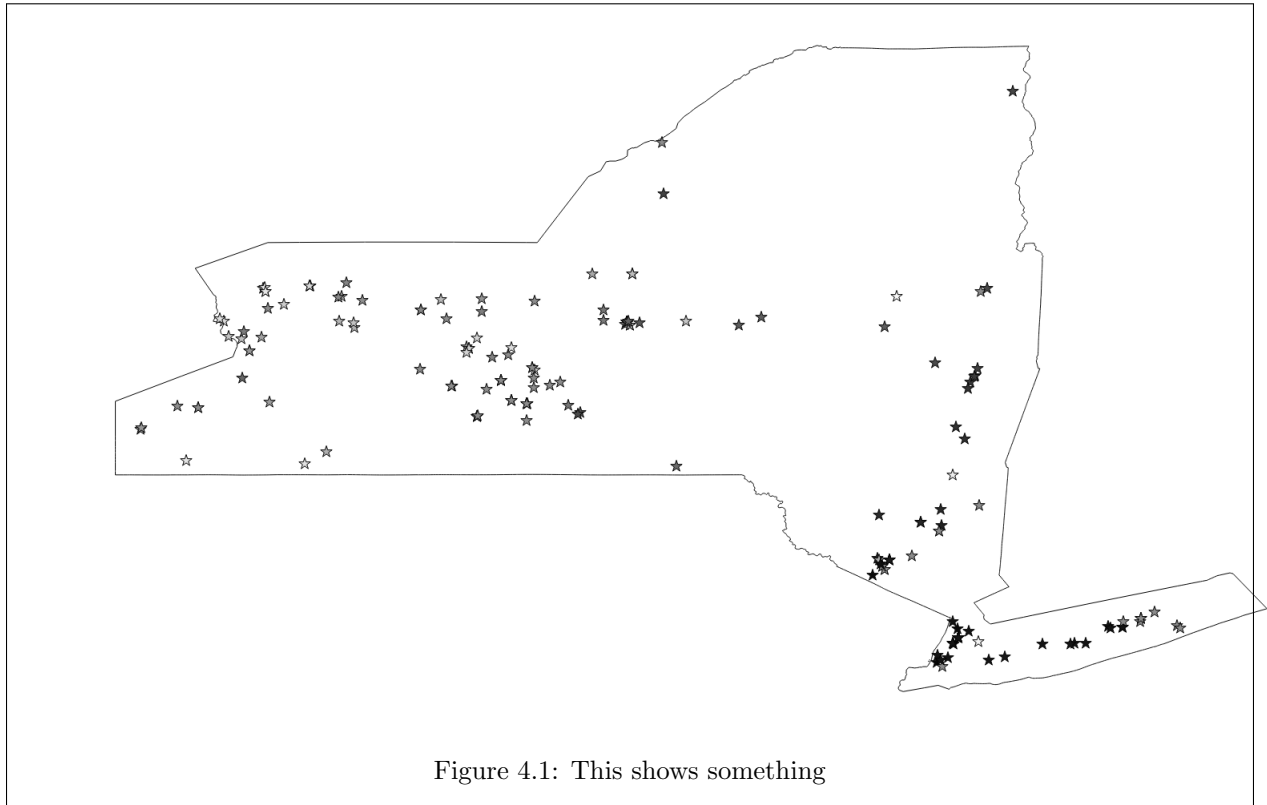
Table 4.4: Another table caption

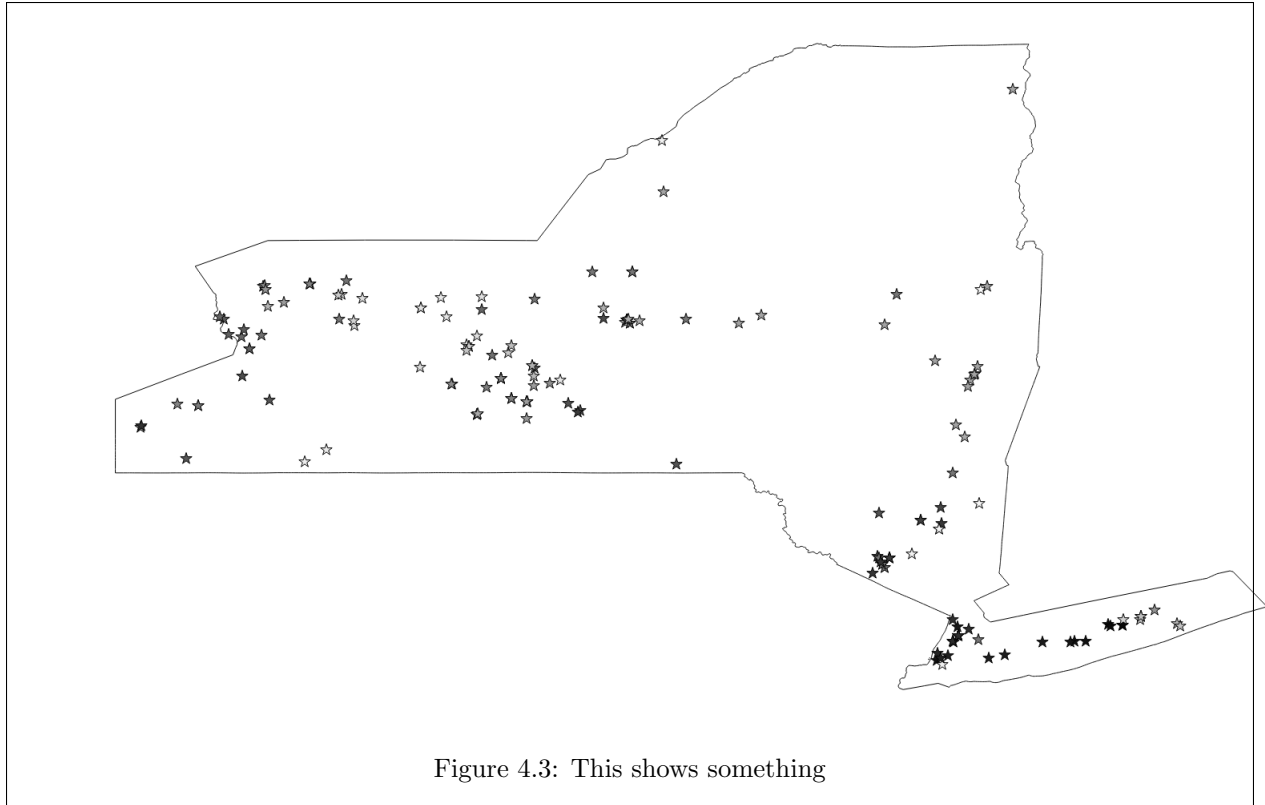
Band	Average	Variance	Deviation
49	2516.51147541	10468640.2301	3235.52781321
66	3197.00382514	14557042.9968	3815.36931329
69	4271.8295082	27027616.4373	5198.80913646
243	7088.95081967	95240598.712	9759.12899351

Table 4.5: Another table caption

Band	Max Price	Max County	Min Price	Min County
49	17243.3	Clinton County	0	Ontario County
66	18308.3	Clinton County	0	Ontario County
69	23951.8	Clinton County	0	Suffolk County
243	29180.1	Suffolk County	0	Ontario County

Table 4.6: Another table caption





Band	Average	Variance	Deviation
49	5073.48738715	9636831.91217	3104.32471114
66	6874.11295805	13268916.3606	3642.65237988
69	7159.60167286	21412615.3937	4627.37672918
243	15549.7135157	107420104.717	10364.3670678

Table 4.7: Another table caption

4.2.1 Farms

4.2.2 Intermediate Processors

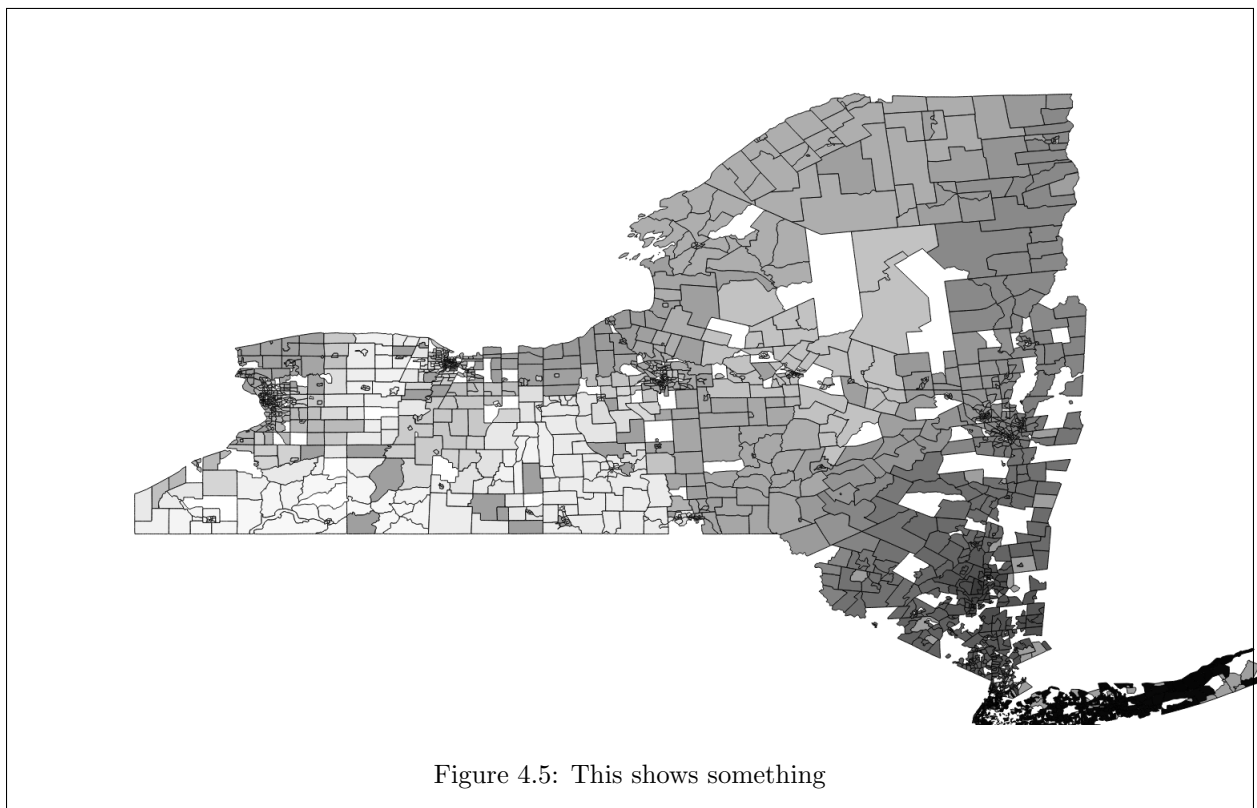
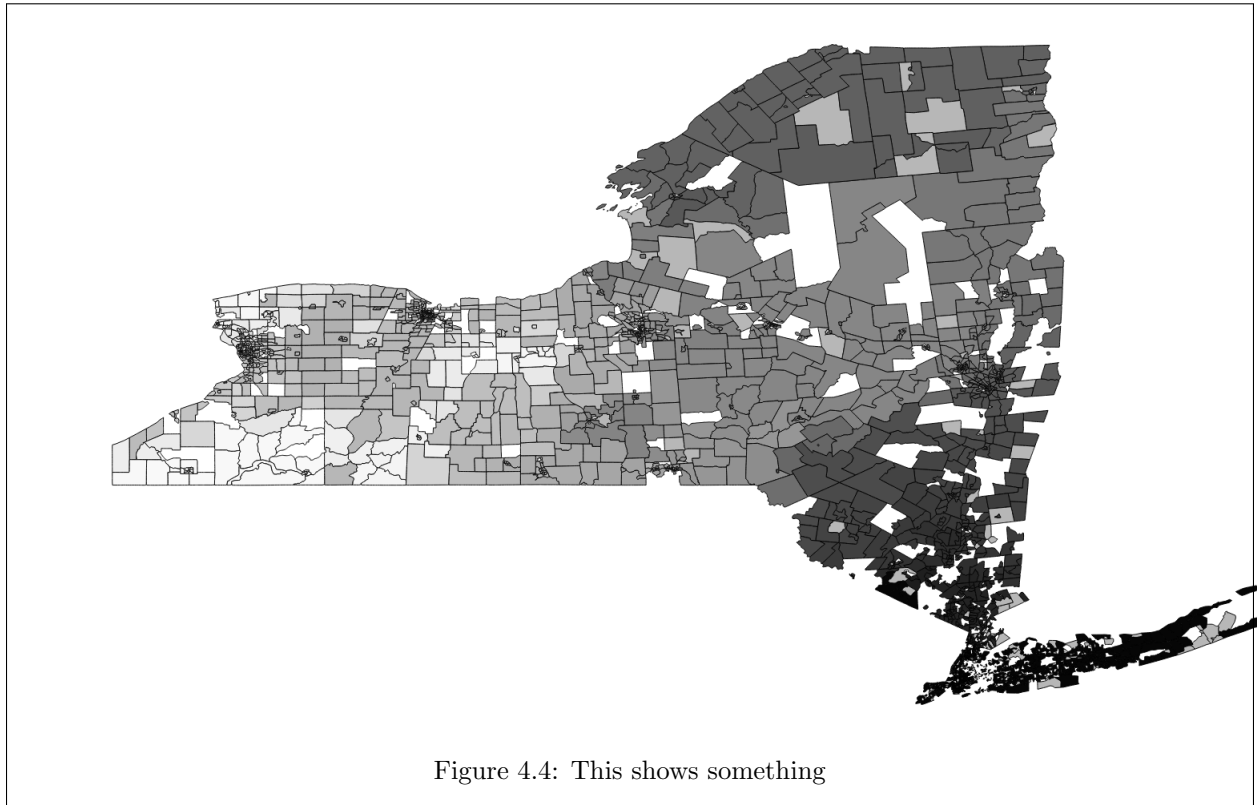
4.2.3 Stores

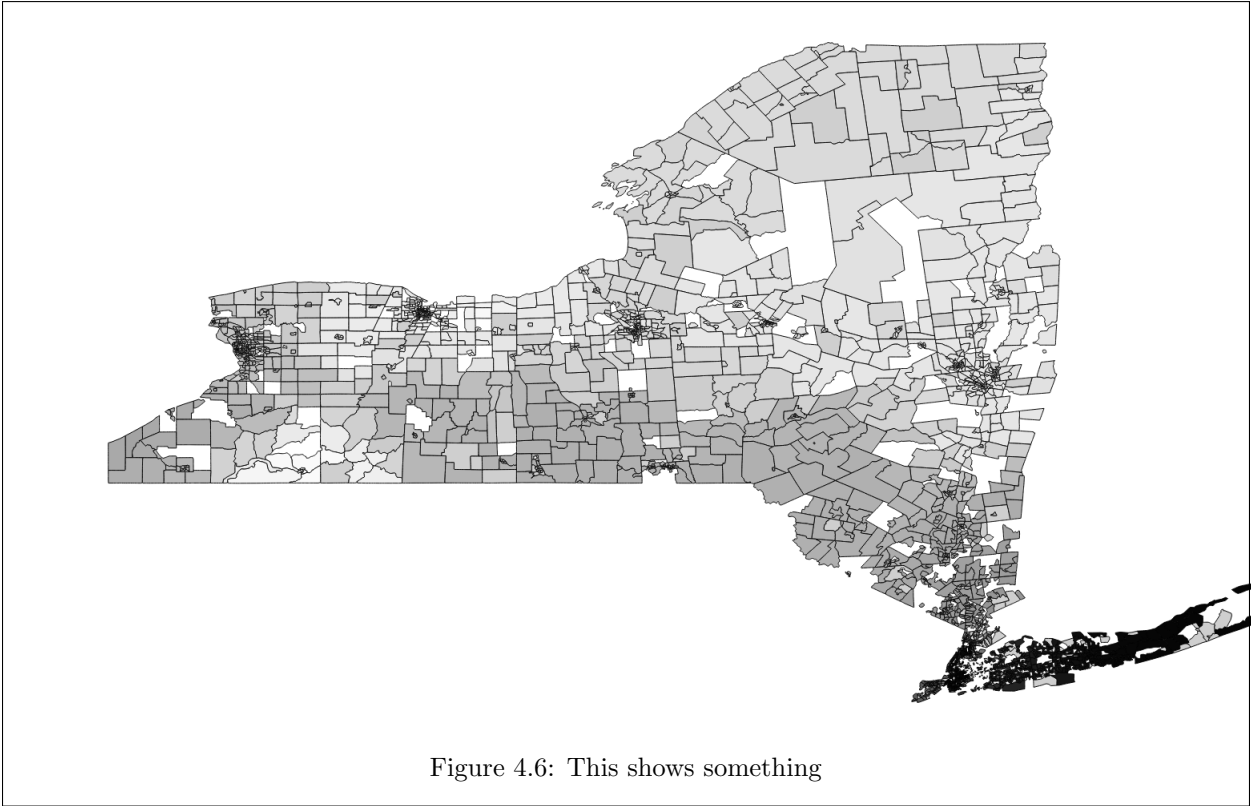
4.3 Checking the Uniqueness of Results

I think this more has to do with a rounding error, rather than an issue with the data. One key question is how I've skipped doing exclusion restrictions and the such to do my model.

The other question is how I've come up with point estimates when most empirical models come up with ranges.

The final thing is it might be foolish to look at the data in this way. I've probably lost a lot by ignoring the traditional techniques economist use make estimates about supply and demand. That being said, I think there is something here. There is a ton of potential in refining this technique more and more so that it's as polished as the techniques involved with ordinary least squares





Band	Max Price	Max County	Min Price	Min County
49	17798.3	Clinton County	171.4	Orleans County
66	23301.8	Clinton County	311.5	Wayne County
69	24858.6	Clinton County	123.6	Chautauqua County
243	32869.4	Suffolk County	237.6	Orleans County

Table 4.8: Another table caption

4.4 Transportation Problem Assumptions

Despite the possibilities, prices in a transportation problem aren't a perfect analogue to market prices. Transportation problems assume firms coordinate to reduce total transportation costs. Additionally, prices in the problem only reflect linear transportation costs. To make the problem more realistic, we can increase all prices by a constant amount to reflect inputs (assuming constant returns to scale). Under this change, the solution is still optimal from a social planning perspective.

Additionally, I am looking into better specifying the transportation problem using the Google Maps and MapQuest API. My specification makes assumptions about connections between farms, factories, and retailers that may not exist for contractual reasons. In reality, fresh produce can be altered or stored at intermediaries, there may be capacities on the intermediate nodes, and intermediate nodes that buy less than 10,000 dollars of produce.

Finally, I under-estimate costs because New York exports fresh produce. Adding demand from out-of-state increases costs in-state. That being said, the estimated prices are still economically significant without exports. They represent a minimum for prices; adding additional demand from out-of-state increases costs in-state. Additionally, most of the categories of produce have a short shelf life and are exported in insignificant quantities with the exception of apples.

Chapter 5

Conclusion

Instead of just jumping into my project. I figured I wanted to take a moment to reflect and explain why I've done what I've done. I've brought together a lot of cool tools from mapping and geographic imaging, optimization, and of course economics. I've tried my best to fit these tools together in a way that makes sense, but that's one of the hard things about looking about using a well established tool in an unconventional way. In some ways, it may not make sense. Either way, I definitely want to say it's been a fun and exciting process to see the resources that are out there beyond FRED, the BLS, and NCYS and would highly recommend to others to check it out and not be afraid when the file extension isn't .csv.

The second thing is Food is super complicated, I've created a model where I tried to capture the elements of the food transportation system that I thought were important. Obviously, there is room for improvement

Bibliography

- [1] Bernell, S., Weber, B., & Edwards, M. (2006). Restricted Opportunities, Personal Choices, Ineffective Policies: What Explains Food Insecurity in Oregon? *Journal of Agricultural and Resource Economics*, 31(2), 193-211.
- [2] Chung, Chanjin, and Samuel L. Myers. "Do the Poor Pay More for Food? An Analysis of Grocery Store Availability and Food Price Disparities." *The Journal of Consumer Affairs* 33.2 (1999): 276-96.
- [3] Cook, W. J., Cunningham, W. H., Pulleyblank, W. R., & Schrijver, A. (1998). *Combinatorial optimization* (pp. 91-126). New York: John Wiley & Sons.
- [4] Hwang, M., & Smith, M. (2012). Integrating publicly available web mapping tools for cartographic visualization of community food insecurity: A prototype. *GeoJournal*, 77(1), 47-62.
- [5] Just, R., & Weninger, Q. (1997). Economic Evaluation of the Farmers' Market Nutrition Program. *American Journal of Agricultural Economics*, 79(3), 902-917.
- [6] Larson, J., & Moseley, W. (2012). Reaching the limits: A geographic approach for understanding food insecurity and household hunger mitigation strategies in Minneapolis-Saint Paul, USA. *GeoJournal*, 77(1), 1-12.
- [7] Stewart, T., & Ittmann, H. (1979). Two-Stage Optimization in a Transportation Problem. *The Journal of the Operational Research Society*, 30(10), 897-904.