Hub Location and Routing Optimization

INEG 3613 – Introduction to Operations Research

April 30, 2020

Benjamin Barron, Samuel McKinney and Parker Tankersley

Section 1: Preliminary Analysis

Annual Cost No Routing

First, we calculated the cost to meet all demand from city to city without any routing through any hubs. This was done to get a baseline cost without any optimization applied as well as validate an early pre-routing AMPL model. For this calculation we can use the data that represents the distance from city i to city j and the demand from city i to city j. These two variables can be denoted as the following:

$$h_{i,j} = demand\ from\ origin\ city\ i\ to\ destination\ city\ j, i = 1..42, j = 1..42$$

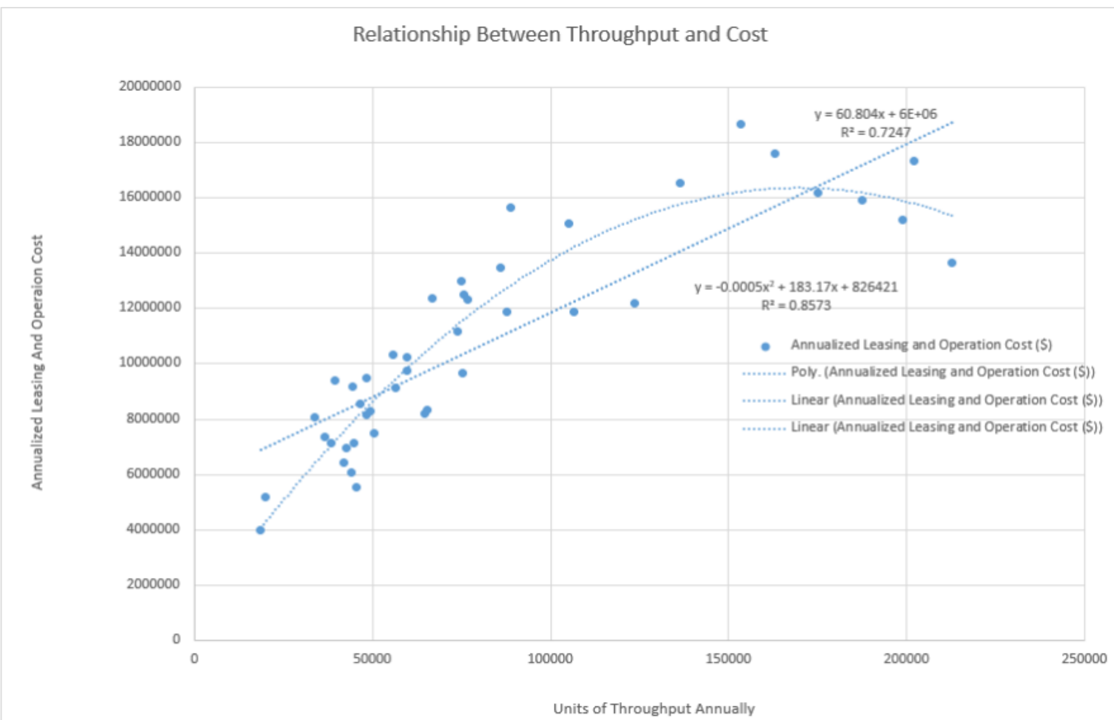
$$d_{i,j} = distance\ from\ city\ i\ to\ city\ j, i = 1..42, j = 1..42$$

Using these two variables the cost can be computed. Recall that the cost to ship one unit from a non-hub city to another non hub city is \$0.40 per mile. Thus, the calculation for the total cost will be as follows.

$$\text{cost} = \sum_{i=1}^{42} \sum_{j=1}^{42} 0.4 \cdot h_{i,j} \cdot d_{i,j} = 155{,}120{,}200.40$$

Cost to Build A Hub in Each City

Shown below is a graph showing the relationship between hub size (throughput) and its associated cost. We had the option of a linear or polynomial relationship. We chose the linear relationship because it appears that the polynomial fitted line was skewed by a few outliers that were not very realistic.



Using the result from the graph above we can create a function that takes in a throughput value and returns an annual cost. For this calculation we will also use the same demand and distance matrices that were used in the first preliminary analysis. All of these can be represented as follows.

$$c(x) = 60.804 \cdot x + 6 \cdot 10^6$$

$$h_{i,j} = demand\ from\ origin\ city\ i\ to\ destination\ city\ j, i = 1..42, j = 1..42$$

$$d_{i,j} = distance\ from\ city\ i\ to\ city\ j, i = 1..42, j = 1..42$$

We can then compute the throughput for each city j using the following equation:

$$T_j = \sum_{i=1}^{42} h_{i,j} + \sum_{i=1}^{42} h_{j,i} \quad \forall\, j = 1..42$$

After computing the cost to build a hub in every city, we must also consider the cost to ship required units to each city. The assumptions that were made in the first part of the preliminary analysis still hold for this analysis. Because each city is now a hub, the cost to ship one unit one mile will no longer be \$0.40 but instead \$0.05. This the final calculation will be as follows:

$$C = \sum_{j=1}^{42} c(T_j) + \sum_{i=1}^{42} \sum_{j=1}^{42} 0.05 \cdot h_{i,j} \cdot d_{i,j} = \$339{,}794{,}525.05$$

Section 2: Scope the problem for optimization

This optimization problem requires several decisions to be made by the CPLEX solver. A major decision regarding hub location is what cities should be home to hubs and which should not. Another major decision is how many of these city hubs need to be opened in order to meet the projected demand for the least amount of money. Finally, a decision will have to be made as to when cargo should route through a hub or go directly to a destination city.
Decisions will need to be made in accordance with a specific set of criteria. One influencer for decision making is the distance between cities. Another factor that will affect decisions is the cost of moving cargo between hubs and cities. Finally, the cost of hub opening, and operating hubs will be a major factor when making decisions.
Meeting forecasted demand is going to be the only limiting factor, or constraint, for decision making. There must be adequate hubs to accommodate the large amount of forecasted demand from city to city. What we must do is find the cheapest and most effective option for hub location and cargo transactions from city to city.

Section 3: Model Formulation

First Model
The first model we created was an over simplified model that could then be expanded upon. We created a normal form minimization linear program with NxN variables where N is a parameter equal to the number of cities. Additional parameters that were used are the distance from city i to

city j which is denoted as $distance_{i,j}$, and the units of demand from city i to city j which is denoted as $demand_{i,j}$. The final notation for these parameters is as follows:

$$N = number\ of\ cities$$

$$demand_{i,j} = demand\ from\ city\ i\ to\ city\ j, i = 1..N, j = 1..N$$

$$distance_{i,j} = distance\ from\ city\ i\ to\ city\ j, i = 1..N, j = 1..N$$

Decision variable $v_{i,j}$ represents the units of product shipped from city i to city j where i and j both range from 1 to N and is denoted as follows:

$$v_{i,j} = units\ transported\ from\ city\ i\ to\ city\ j\ directly, i = 1..N, j = 1..N$$

The objective of this model is to minimize the total annual cost which is calculated by multiply the cost per mile of $0.40 times the distance from the origin to destination times the number of units shipped.

$$minimize\ z = \sum_{i=1}^{N}\sum_{j=1}^{N} 0.4\ \cdot\ distance_{i,j}\ \cdot\ v_{i,j}$$

The main constraint for this model states that the units shipped from city i to city j must be greater than or equal to the demand from city i to city j. Units shipped must be a non-negative value.

$$v_{i,j} \geq\ demand_{i,j}\ \ \forall i = 1..N\ \forall j = 1..N$$

$$v_{i,j} \geq 0\ \ \forall i = 1..N\ \forall j = 1..N$$

<u>Second Model</u>

A second model was developed to handle the various ways of routing between hubs and cities in order to get a better solution than the first model. A breakdown of the second model has been provided below.

New parameters for the second model include S , hub_cost_s, and $hub\_capacity_s$. Parameter S represents the number of hub sizes available to utilize in any given city. The parameter hub_cost_s represents the cost to build hubs of size 1 through S. Finally, the parameter $hub\_capacity_s$ represents the maximum capacity that a hub of size S can handle. All other parameters represent the same factors as in the first model.

$$N = number\ of\ cities$$

$$S = number\ of\ different\ hub\ sizes$$

$$hub_{cost_s} = cost\ to\ build\ hub\ of\ size\ s, s = 1..S$$

$$hub\_capacity_s =\ capacity\ for\ hub\ of\ size\ s, s = 1..S$$

$$demand_{i,j} = demand\ from\ city\ i\ to\ city\ j, i = 1..N, j = 1..N$$

$$distance_{i,j} = distance\ from\ city\ i\ to\ city\ j, i = 1..N, j = 1..N$$

New decision variables $x_{i,h,k,j}$, and $y_{i,s}$ were added to the model on top of the existing decision variable $v_{i,j}$. Decision variable $x_{i,h,k,j}$ represents the forecasted demand units transported from city i to hub h, hub h to hub k, and then finally from hub k to city j. Decision variable $y_{i,s}$ is a binary variable indicating if a hub of size S is in city i. These variables are denoted as:

$$v_{i,j} = units\ transported\ from\ non-hub\ city\ i\ to\ non-hub\ city\ j\ directly, i = 1..N, j = 1..N$$

$$x_{i,h,k,j} = units\ transported\ from\ city\ i\ to\ hub\ h\ to\ hub\ k\ to\ city\ j, i = 1..N, h = 1..N, k = 1..N, j = 1..N$$

$$y_{i,s} = \begin{cases} 1, & hub\ of\ size\ s\ is\ located\ in\ city\ i \\ 0, & hub\ of\ size\ s\ isn't\ located\ in\ city\ i \end{cases} \quad i = 1..N, s = 1..S$$

However, the variable $x_{i,h,k,j}$ has a few different interpretations depending on the value of the i, h, k and j indices. For example, $x_{i,i,i,j}$ indicates that a unit is going from a hub city to a non-hub city and will cost $0.22. In the table below, the different variations and respective representations of the x variable can be seen in the table below.

X Variable Variations	Interpretation
$x_{i,i,i,j}$ $where\ i \neq j$	Initial city is hub and destination is non-hub. Cost = $0.22
$x_{i,j,j,j}$ $where\ i \neq j$	Initial city is non-hub and destination is hub. Cost = $0.22
$x_{i,i,j,j}$ $where\ i \neq j$	Initial city is hub and destination is hub. Cost = $0.05
$x_{i,h,h,j}$ $where\ i \neq h \neq j$	Initial city is non-hub, routed through hub, destination is a non-hub. Cost = $0.22
$x_{i,h,j,j}$ $where\ i \neq h \neq j$	Initial city is non-hub, routed through hub, destination is hub. Cost = $0.22 for initial to hub then $0.05 from hub to destination.
$x_{i,i,h,j}$ $where\ i \neq h \neq j$	Initial city is hub, routed through hub, destination is non-hub. Cost = $0.05 for initial to hub then $0.22 from hub to destination.
$x_{i,h,k,j}$ $where\ i \neq h \neq k \neq j$	Initial city is non-hub, routed through first hub, routed through second hub, arrive at non-hub destination. Cost = $0.22 for initial to first hub, $0.05 from first hub to second hub, $0.22 from second hub to non-hub destination.

The new model's objective function is still seeking to minimize the annual cost of meeting forecasted demand, but it also considers the cost of building the hubs of various sizes in select cities. The original $0.40/mile can be seen for shipping directly from city to city using the original summation technique, as well as the $0.22/mile it costs to ship from a city to a hub, and the $0.05/mile it costs to ship from hub to hub also utilizing this technique. The cost to build the required number of hubs is done by taking the sum over all cities and all hub sizes and multiplying the associated hub cost by $y_{i,s}$, which indicates the existence of a hub of size S, giving us the total cost for building all required hubs for an optimal solution.

$$minimize\ z = \sum_{i=1}^{N}\sum_{j=1}^{N} 0.22 \cdot\ distance_{i,j} \cdot (x_{i,i,i,j} +\ x_{i,j,j,j})\ such\ that\ i \neq j\ \ +$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} 0.05 \cdot\ distance_{i,j} \cdot (x_{i,i,j,j})\ such\ that\ i \neq j\ +$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} 0.40 \cdot\ distance_{i,j} \cdot (v_{i,j})\ such\ that\ i \neq j\ +$$

$$\sum_{i=1}^{N}\sum_{h=1}^{N}\sum_{j=1}^{N} x_{i,i,h,j} \cdot (0.05 \cdot distance_{i,h} + 0.22 \cdot distance_{h,j})\ \ such\ that\ i \neq h \neq j\ \ +$$

$$\sum_{i=1}^{N}\sum_{h=1}^{N}\sum_{j=1}^{N} x_{i,h,j,j} \cdot (0.22 \cdot distance_{i,h} + 0.05 \cdot distance_{h,j})\ \ such\ that\ i \neq h \neq j\ \ +$$

$$\sum_{i=1}^{N}\sum_{h=1}^{N}\sum_{j=1}^{N} x_{i,h,h,j} \cdot (0.22 \cdot distance_{i,h} + 0.22 \cdot distance_{h,j})\ \ such\ that\ i \neq h \neq j\ \ +$$

$$\sum_{i=1}^{N}\sum_{h=1}^{N}\sum_{k=1}^{N}\sum_{j=1}^{N} x_{i,h,k,j} \cdot (0.22 \cdot\ distance_{i,h} + 0.05 \cdot distance_{h,k} + 0.22\ \cdot distance_{k,j})\ such\ that\ i \neq h \neq k \neq j +$$

$$\sum_{i=1}^{N}\sum_{s=1}^{S} hub\_cost_{s}\ \cdot\ y_{i,s}$$

The first constraint listed below ensures that only one hub or one size is constructed in each hub assigned city. The second constraint ensures that routing through any one hub does not exceed its assigned capacity based on its size. For the second constraint we believe there still might be an edge case when the h and k indexes equal each other but were unable to get it to work properly when we tried to fix it. Regardless we don't think it is large enough of an issue to cause a large difference in the answer to the problem.

$$\sum_{s=1}^{S} y_{i,s} \leq 1\ \ \forall i = 1..N$$

$$\sum_{i=1}^{N}\sum_{k=1}^{N}\sum_{j=1}^{N}\left(x_{i,h,k,j} + x_{i,k,h,j}\right)\ \leq\ \sum_{s=1}^{S} hub\_capacity_{s} \cdot\ y_{h,s}\ \ \forall h = 1..N$$

The following constraints ensure that unit demand from city to city is met by either direct shipping techniques or through one of the two forms of routing through hubs.

$$v_{i,j} + \sum_{h=1}^{N}\sum_{k=1}^{N} x_{i,h,k,j} \geq demand_{i,j} \;\; \forall i = 1..N \; \forall j = 1..N$$

$$x_{i,i,i,i} = 0 \;\; \forall \; i = 1..N$$

The following constraints were added at the end of the model in order to ensure non-negative shipping values as well as binary values indicating if a hub was constructed or not.

$$v_{i,j} \geq 0 \;\; \forall i = 1..N \; \forall j = 1..N$$

$$y_i \; binary \; \forall i = 1..N$$

Section 4: Gather and Synthesize the Data

Parameters that are fed into the models are the number of cities present, the different sizes of hubs available, the forecasted demand from city to city and the distance from city to city. The last two parameters were obtained from the provided data sheet that accommodated this optimization problem. The number of cities will be populated with the number forty-two to represent each city involved in the problem. The parameter 'S' representing the sizes of hubs is used to provide the model with different sizes of hubs to build to accommodate the forecasted demand.

$$N \; = \; Number \; of \; cities \; = \; 42$$

$$S \; = \; number \; of \; different \; sized \; hubs \; = \; 10$$

$$demand_{i,j} = (located \; in \; s5\_model2\_data.dat)$$

$$distance_{i,j} = (located \; in \; s5\_model2\_data.dat)$$

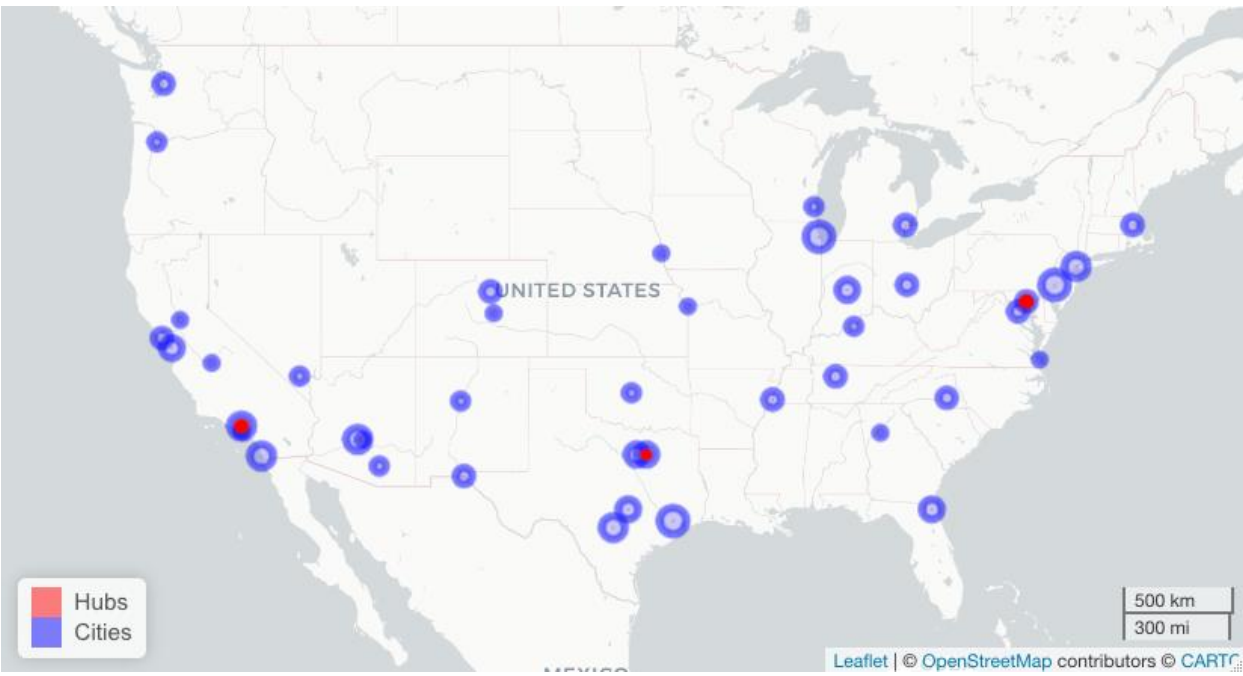
$$hub\_capacity_s = \; (located \; in \; s5\_model2\_data.dat)$$

$$hub\_cost_s = \; (located \; in \; s5\_model2\_data.dat)$$

Section 5: Solve the Model

After running the final model outlined in section 3 using the parameters outlined in section 4, for a total of 12 hours it was found that the objective value of the optimal solution that minimized annual cost while meeting forecasted demand was \$118,923,130.20. This is significantly lower than the anticipated cost of building a hub in every city and shipping directly which was \$339,794,525.05. It was also significantly lower than the annual cost of building no hubs and shipping all units directly which was \$155,120,400.40. The model resulted in a relative gap value of 0.227506 which, while not at the optimal relative gap of 0, gives a good ballpark idea of where the 'true' optimal solution lies.

As seen in the map below, out of the 42 cities, 3 of them were made into hubs. The hub on the west coast is in Los Angeles, California as a size 4 hub, which means its max capacity of throughput is set to 120,000 units. The hub in Los Angeles routed 118,684 units through it. The hub on the east coast is in Baltimore, Maryland as a size 4 hub which means its max capacity of throughput is set to 120,000 units as well. The Baltimore hub routed 114,177 units through it. The last hub is in Dallas, Texas as a size 2 hub which enables it to have a max throughput of 60,000 units. The hub in Dallas had 59,070 units that were routed through it. The larger the city's circle is on the map equates to how many units started and finished in that city. As seen in the map, the three hub locations are validated because they are all located in those higher demand regions with large city circles. The hubs are also spread out across the country and not in all on one region, which is another good indicator that the hubs were opened in correct locations.



Routing Data		
Routing Designation	*Units Transported*	*Proportion of Total*
No Routing	200,784	57.61%
Routing with one hub	31,591	9.06%
Routing with two hubs	116,117	33.33%
Total	348,552	100.00%

It can be observed from the table above that routing with either one or two hubs is done by approximately 71% of the total units transported throughout the network of cities. Conversely there is only about 29% of the total units transported that move directly from city to city. From this we can conclude that routing plays a vital role in meeting the forecasted demand from city to city in a cost-effective manner.

	Base Run	***Doubled Demand***	***Halved Shipping Costs***	***Halved Hub Costs***
Optimal Cost	$31,986,059.86	$61,123,406.92	$17,953,153.4	$22,924,402.87
No Routing	42786	68,338	80286	30313
Routing Through One Hub	13325	42872	0	30476
Routing Through Two Hubs	21996	49362	0	19497
Hubs Built	3	4	0	5

In the table above, changes as a result of parameter alteration can be observed. Effects on the model were observed by using a smaller dataset, with only 16 cities instead of 42, in order to shorten the time required for analysis. When the demand between cities was doubled it was observed that the overall cost was higher, an additional hub was constructed, and unites were routed through hubs about 12% more than the base case. When the cost per mile for all types of shipping was cut in half, the objective cost was lowered by approximately 45% and routing was not utilized at all. With these reduced costs there was no justifications for any hubs to be constructed. Finally, when the hub operating costs were halved the objective cost was lowered by 29%, two additional hubs were constructed, and routing was utilized 62% of the time.

Section 6: Extend the Model

To extend the model, the focus was to determine a way reduce the run time while not losing a large degree of optimality. To do this we decided to explore the idea of creating two separate models: the first to determine hub locations for a set number of hubs, and the second to determine the optimal way to utilize these hubs. The hope is that if we could reduce the complexity of the model by removing the decision of where to build the hubs, the model would be able to solve more optimally and in less time.

We achieved this goal by creating a p-median model to determine the hub locations that minimized the total demand weighted distance. It is important to note that our p-median model did not solve the issue of determining hub sizes, this was still handled by the second model. The parameters for this model were slightly different that the previous model. The demand for each city is equal to the units of cargo that originates at that city plus the units of cargo that are destined for that city and is denoted as $h_{i,}$. The number of demand nodes is denoted as I, the number of candidate hub locations is denoted as J, and the number of hubs to build is denoted as P. For this analysis we decided upon P being equal to 3 because that was the number of hubs that the model in section 5 decided on and it seemed like a reasonable number of hubs.

$$I = 42,\ J = 42,\ P = 3$$

$$h_i = demand\ for\ node\ i\ \forall i = 1..I$$

$$d_{i,j} = distance\ from\ demand\ node\ i\ to\ canidate\ facility\ j\ \forall i = 1..I\ \forall j = 1..J$$

The decision variables for this model are a binary variable the represents whether or not to build a hub at each candidate facility, and a two-dimensional binary variable that represents whether of not a demand node i is assigned to a facility j. These decision variables can be denoted as follows:

$$x_j = \begin{cases} 1, facility\ is\ built\ as\ canidate\ site\ j \\ 0, facility\ is\ not\ built\ at\ canidate\ site\ j \end{cases} \forall j = 1..J$$

$$y_{i,j} = \begin{cases} 1, demand\ node\ i\ is\ assigned\ to\ facility\ j \\ 0,\ \ demand\ node\ i\ isn't\ assigned\ to\ facility\ j \end{cases} \forall\ i = 1..I\ \ \forall\ J = 1..J$$

The objective function for the p-median model is to minimize the total demand weighted distance and is denoted as follows:

$$z = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j} \cdot d_{i,j} \cdot h_i$$

For the p-median model there are three constraints. The first constraint is to ensure that exactly P hubs are built. The second is to ensure that every demand node is assigned to one facility. The third is to ensure that a demand node cannot be assigned to a facility that does not exist. These three constrains can be written as follows:

$$\sum_{i=1}^{J} x_i = P$$

$$\sum_{j=1}^{J} y_{i,j} = 1\ \ \forall i = 1..I$$

$$y_{i,j} \leq x_j\ \ \forall i = 1..I\ \ \forall j = 1..J$$

$$y_{i,j}\ binary\ \ \forall i = 1..I\ \ \forall j = 1..J$$

$$x_j\ binary\ \ \forall j = 1..J$$

This model ran to full optimality in less than 5 seconds and the output was relatively similar to the results obtained from our model that was outlined in section five. The previous model decided to build hubs of sizes four, two and four were built in Los Angeles, Dallas, and Baltimore respectively. The output of this model was to build hubs in Los Angeles, Columbus and Fort Worth. We then took the output from this model and fed it into our original model as input through a parameter called hubs which is written as:

$$hubs_i = \begin{cases} 1, & do\ build\ hub\ in\ city\ i \\ 0, & don't\ build\ hub\ in\ city\ i \end{cases} \forall i = 1..N$$

In addition to this we needed a constraint that make sure a hub was built in each city that was supposed to have a hub built in it without specifying the size of the hub. To do this we altered our already existing $c\_hub\_per\_city$ constraint to look like the following:

$$\sum_{s=1}^{S} y_{i,s} = hub_i \ \ \forall i = 1..N$$

The output of this model was to build a size 5 hub in Los Angeles, a size 6 hub in Columbus and a size 4 hub in Fort Worth. The hub in Los Angeles used 135,150 out of its 150,000 units of capacity, the hub in Columbus used 161,944 out of its 180,000 units of capacity and the hub in Fort Worth used 115,933 out of its 120,000 units of capacity. The annual cost for this operation was found to be $113,678,150.50 which is over five million dollars cheaper than the solution found in section five. In addition, each of these two models solved to full optimality in less than 5 seconds each. A visual representation of the results can be seen in the map below. The red circles represent hubs and the blue dots represent cities. The hubs are scaled based on the size of hub to be built and the cities are scaled based on the total demand that originates and total demand that is destined for each city.