For office use only

T1 ____________

T2 ____________

T3 ____________

T4 ____________

For office use only

F1 ____________

F2 ____________

F3 ____________

F4 ____________

2016

19th Annual High School Mathematical Contest in Modeling (HiMCM) Summary Sheet

(Please make this the first page of your electronic Solution Paper.)

Team Control Number: 6209

Problem Chosen: B

Please paste or type a summary of your results on this page. Please remember not to include the
name of your school, advisor, or team members on this page.

Summary

The question requests us to find the optimal placement of warehouses both with and without tax considerations so that the minimum locations selected can cover all 48 states in the United States. In order to achieve the task, our team put forward two models: the location selection model based on *Floyd-Warshall Algorithm* and the location scoring model.

By appealing to the *Floyd-Warshall Algorithm*, we are able to calculate the shortest route between two locations, basing on the highway transportation system. Then, we use Matlab programming to help us select the qualified locations and the minimum number. Through observation, we find out that UPS is able to reach 320-370 miles in one day service, and our model tells us a slight difference in the selected distance will not have a significant influence.

The second model of us is a little complicated as we want to include all possible factors. Our group believe it is impossible to find out a solution that minimizes all customers' tax liability, therefore we must weight all locations to give out the best solution. We first develop an exponential equation containing state population and state GDP. However, as the purpose of the company to add delivery service is to earn more future benefits, we improve our model to be more concrete and have more future credibility. We optimize *Leslie Matrix* to reflect future population growth based on the proportion of female in the total population, and we replace state GDP with income gap as a better indicator of state consumption. We refer to the *Absolute Income Hypothesis* to prove the significance of income gap on our own. At last, we combine the two improved factors and tax factor into the initial equation to create a scoring system. The final model is the sum of score of each location in a solution. Our idea is that since cross state delivery does not include tax, we should select the locations in regions where there are less consumption.

The result of our first model tells the minimum number to be 17. We have obtained 3 sets of results with equivalent importance, but considering the space limit, we are not going to list here. By scoring the location with our second model, it turns out that including sales tax does not make a big difference, but the minimum number becomes 18 for the lowest score. Our suggested locations are: Barstow, Charlotte, Des Moines, Boise, Chicago, Pittsburg, Tallahassee, Phoenix, New York City, Spokane, Cheyenne, Bismarck, Dallas, Sioux Falls, Oklahoma City, Hartford, St. Louis, and Eugene.

Our group has created a realistic model for finding warehouse locations. We consider the real driving distance, the future population growth, and the average state consumption in our model. We also verify our result by putting circles on the US land map to simulate the coverage area our location solution is able to achieve, and the conclusion turns out to be positive.

Letter to Company's President

Dear President,

After thorough consideration of our company's current plan to expand warehouses, our team has developed robust models to predict the minimum number of warehouses and the locations to be built. We have the following recommendations to make:

1. Seventeen warehouses are enough to cover the 48 states within one day, assuming that the cost of building every warehouse is the same. One possible set of locations is *Eugene, Boise, Las Vegas, Barstow, Hartford, Missoula, Durango, El Paso, Bismarck, Omaha, Oklahoma City, Madison, Jackson, Indianapolis, Detroit, Tallahassee, Charlotte*.
2. If our company also wants to take advantage of the tax policy so as to minimize the tax burden of customers, we suggest building eighteen warehouses. The best set of locations are *Barstow, Charlotte, Des Moines, Boise, Chicago, Pittsburg, Tallahassee, Phoenix, New York City, Spokane, Cheyenne, Bismarck, Dallas, Sioux Falls, Oklahoma City, Hartford, St. Louis, and Eugene*.

We would like to justify our recommendations by explaining the models we relied on. Our first model is an optimization of *Floyd-Warshall Algorithm*, an algorithm primarily aiming to find the shortest route between two points. We specifically chose the data obtained from the highway transpiration system to run the algorithm because we wanted our solution to be realistic instead of simply assuming the straight-line distance between two points. In our second model, we developed a grading system for each possible location, and finally the lowest score represented the best solution. We made our scoring system effective at future predictions on purpose. Considering the primary purpose for our company to include delivery service is to gain more future benefits, we graded each location basing on the state future population and the average level of state consumption. We believed the product of the above two factors and tax would be an excellent indicator for location selection. Ultimately, we wanted our warehouse locations to be placed in region where population is smaller and the level of consumption is lower because interstate delivery is free of sales tax so that we minimize the tax liability for most of our customers.

We believe our team has provided a solution that is good for both company's future earnings and company's potential customers. Moreover, the whole operation of our model is computer-based. If there is any change in plan, we can still edit the MATLAB codes to obtain the best solution. Notice that our model only tells the locations no more specific that the city names, we also encourage other departments in our company to research the cities and figure out the practical solution for warehouse placement with respect to our theoretical model result.

At last, we are confident that our recommendations can be very helpful to the company. Therefore, we hope our efforts to be considered carefully. Thank you!

Sincerely,
Team#6209

Contents

1. Introduction

1.1 Background

With the fast development of internet, a new fashion of shopping appears in the 21st century – online shopping. This brand new form of shopping relies on heavily on the online shops and the delivery services. More and more people nowadays prefer buying online instead of going to a real shop due to the convenience of internet. However, because of the increasing demand for online goods, companies want to guarantee their products can be delivered to the customers as soon as possible. Therefore, selecting the best locations of warehouses becomes an urgent issue for online shop companies. In this paper, we would like to demonstrate our approach to the location selection process in order to ensure the delivery efficiency.

1.2 Restatement of Problem

Firstly, we need to analyze the optimal placement of warehouses as the requirement of Part I, so that the one-day ground shipping service would be able to cover the 48 states in US. We should not only build a model to predict the minimum number of warehouses, but also consider the locations of the warehouses. Each of the warehouses should be responsible for a particular area, completing the orders from customers. The key idea is to calculate the shortest route for delivery service. Secondly, we should select the locations that best reduce customers' tax liability. Warehouses should be placed either in a state with less sales tax or in a state with less potential consumers. Our model also needs to predict if any change in minimum number will occur. Thirdly, we need to consider clothing tax factor in our model. We should think of a way to consider both taxes. By any means, they should not be simply added together.

1.3 Assumptions

<u>General Assumptions</u>

1. The distance that a delivery truck is able to reach within one day is fixed.

   It is impossible for us to take all possible real world scenarios into consideration, such as the unpredictable weather and the diverse geography. Therefore, we assume the maximum distance of one-day ground shipping all over US is defined and fixed.

2. The term “one day” refers to the time from the parcel sent to the parcel delivered is within 24 hours.

   As a common sense, if someone sends the parcel at 23:00, he or she should not expect the parcel to be delivered before the end of day. Therefore, to prevent any future confusion, our group strictly define the chronological meaning of one-day ground shipping.

3. The driving time within a city is very short so that we only need to provide the name of cities when defining the locations.

   If we were to give a location even more specific than a city name, then we need to investigate each location meticulously. Since the minimum number of warehouses is around 20, such approach is inefficient and unnecessary.

4. The construction of warehouse costs the same amount of money regardless of its location.

   For the sake of business benefits, though the question does not specifically address the consideration of cost, we make such assumption in advance so that we do not need to be bothered about weighting the cost of warehouse at each location.

5. Drivers always follow the shortest route during delivery.

   We assume that drivers will take the fastest route in order to drive a maximum number of distance. Besides, it is impossible for us to predict each driver’s driving habit and take individual’s personal preference into consideration when building a model.

<u>Sub-Model Assumptions:</u>

For Leslie Matrix:

- The proportion of women population in the total population will not change significantly in the near future.
- There will not be serious diseases or natural disasters to cause a sudden decline in population

For Consumption Calculation:

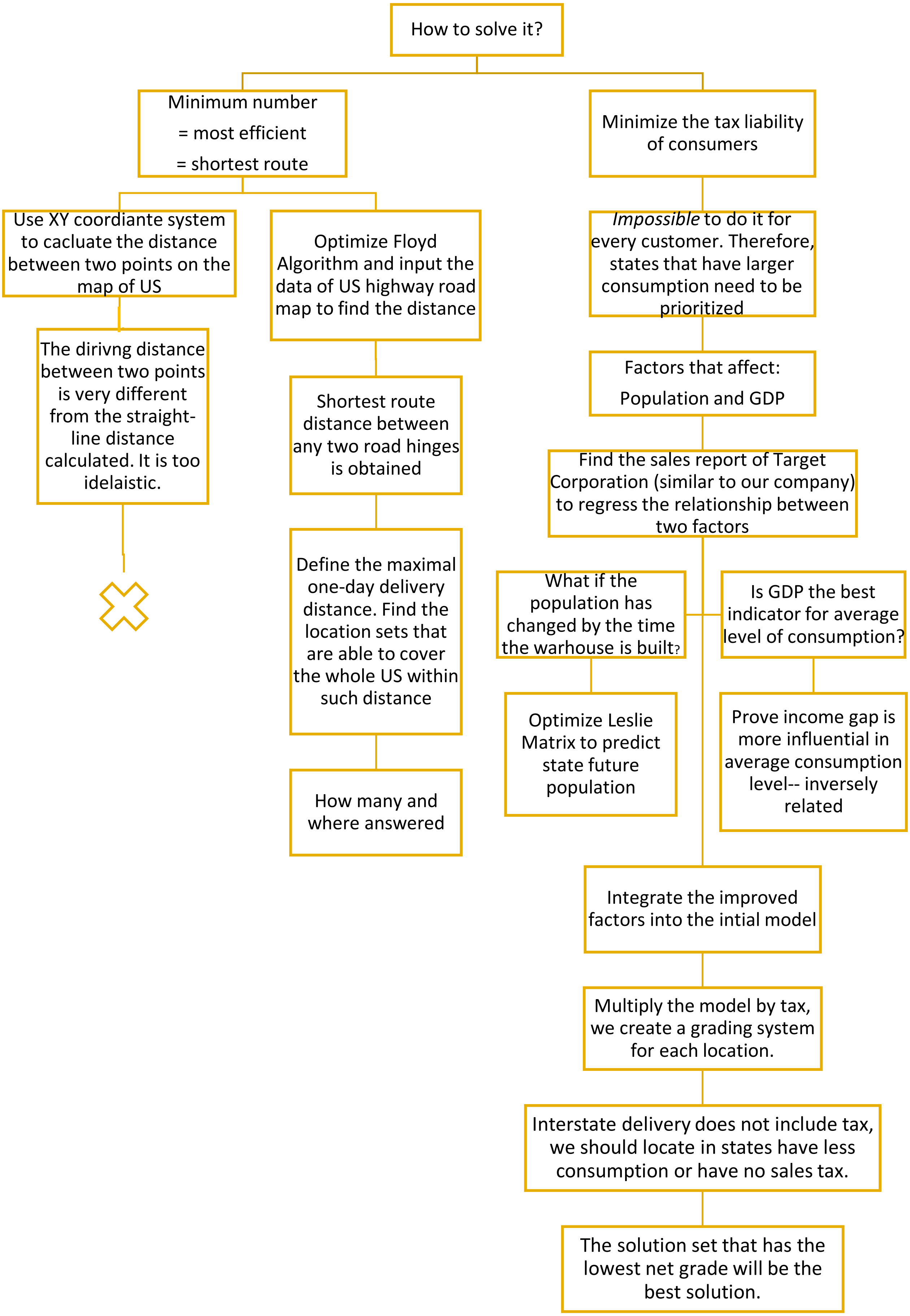
- We only need a good prediction of the near future since short-term benefits are valued more by businesses.
- The income of individuals will not change significantly in the near future.

1.4 Variable Sheet

A	The shortest route matrix calculated
$b_i(t)$	The female birth rate of the i_{th} age interval at time interval t .
C	The level of consumption
C_h	The total consumption of high income group
C_l	The total consumption of low income group
C_t	The consumption at time t
D	The distance matrix
$d_i(t)$	The female birth rate of the i_{th} age interval at time interval t
G	State Gross Domestic Product
$M\mathrm{C}_h$	The marginal propensity of high income group to consume
$M\mathrm{C}_l$	The marginal propensity of low income group to consume
N_t	The future population of US
P_{it}	The population of each state
r	The ratio of the net income of the two income groups
$x_0(t)$	The population of the i_{th} age interval at time interval t .
Y	The total income of people
Y_h	The net income of high income group
Y_l	The net income of low income group
p	The average ratio of female population versus the total population.
ω_i	The average ratio of the population of the i^{th} state versus the total population

(** All variables are explained again during future use in models)

1.5 General Thinking Process



2. Mathematical Model Part 1

2.1 Model Approach

The goal of our first model is to calculate the least number of warehouses that enables one-day ground shipping to reach all 48 states in US and to find these qualified locations. Therefore, if we were to minimize the number of warehouses, the spots we choose should be able to cover as much nearby area as possible in a one-day interval.

Initially, we attempt to set up a coordinate system and use the distance equation to analyze the question. By defining the locations through coordinates, we should be able to obtain the distances between all of the locations. However, we then realize that such approach is too idealistic because the landmass of US is of irregular shape. Besides, in a real world scenario, the driving distance between two locations is longer than the straight line distance. Finally, we decide to build our model upon the US highway transportation system data, which will help us obtain more realistic and reliable results. We optimize Floyd Algorithm to calculate the least distance between any two road junctions, and use Matlab codes to help us find out the minimum locations.

2.2 Floyd-Warshall Algorithm Introduction

The Floyd-Warshall algorithm(1) dates back to the early 60's. Warshall was interested in the weaker question of reachability: determine for each pair of vertices u and *v*, whether *u* can reach *v*. Floyd realized that the same technique could be used to compute shortest paths with only minor variations.

The genius of the Floyd-Warshall algorithm is in finding a different formulation for the shortest path subproblem than the path length formulation introduced earlier. At first the formulation may seem most unnatural, but it leads to a faster algorithm. As before, we will compute a set of matrices whose entries are $d_{ij}^{(k)}$. We will change the *meaning* of each of these entries.

For a path $\mathrm{p}=\langle v_1, v_2, \cdots, v_\ell \rangle$,we say that the vertices $v_2, v_3, \ldots, v_{\ell-1}$ are the *intermediate vertices* of this path. Note that a path consisting of a single edge has no intermediate vertices. We define $d_{ij}^{(k)}$ to be the shortest path from *i* to *j* such that any intermediate vertices on the path are chosen from the set {1, 2, . . . , k}. In other words, we consider a path from i to j which either consists of the single edge (*i, j*), or it visits some intermediate vertices along the way, but these intermediate can only be chosen from {1, 2, . . . , *k*}. The path is free to visit any subset of these vertices, and to do so in any order.

2.3 Optimizing Floyd Algorithm

Considering the importance of distance between the clients and the warehouses, we summarized the problem as: the warehouses should be responsible for a particular area and deliver the goods, finally satisfying the demands within the particular area. The key idea is to calculate the shortest distance. Based on the reality, we develop our own optimization model.

Variables

I is the number of demanded locations;

i is the serial number of the demanded location, $i = 0,1,2,3,\dots,I$, $i = 0$ is the serial number of the warehouses;

k is the serial number of the transport route;

d_{ij} is the minimum distance restriction by the real transportation route between i and j, $i/j = 0,1,2,3,\dots,I$;

M is a large integer.

Objective Function

In order to calculate the minimum length of transportation:

$$min = \sum_{k}\sum_{i=0}^{I}\sum_{j=0}^{I} d_{ij} \times x_{kij}$$

Premise/constriction

In order to maximize the efficiency, one and only one transportation route should serve a demanded location:

$$\sum_{k} y_{ki} = 1 \quad \forall i = 1,2,3,\dots,I$$

In order to ensure the in-and-out flow equilibrium of each demanded location, we have:

$$\sum_{j=0}^{I} x_{kij} = y_{ki} \quad \forall k; i = 0,1,2,3,\dots,I$$

$$\sum_{i=0}^{I} x_{kij} = y_{kj} \quad \forall k; i = 0,1,2,3,\dots,I$$

Furthermore, since only the route starts from the warehouse is efficient:

$$\sum_{i=0}^{I}\sum_{j=0}^{I} x_{kij} \leq M\times \sum_{j=0}^{I} x_{k0j} \quad \forall k; i = 0,1,2,3,\dots,I$$

At last, we need to eliminate the possible sub-loop:

$$u_{ki} - u_{kj} + M\times x_{kij} \leq M - 1 \ \ \forall k; i = 0,1,2,3,\dots,I; i \neq j$$

Decision Variables:

1. $y_{kj} = 1$ when y_{kj} can satisfy the amount of demand of demand location i; otherwise, $y_{kj} = 0$.
2. $x_{kij} = 1$ when the transportation route satisfies the amount of demand between i and j, and transportation sequence of i and j is successive; otherwise, $x_{kij} = 0$.
3. u_{ki} is the variable between 0-1, used to eliminate the possibility of the occurrence of sub-loop.

2.4 Shortest Route Calculation

In order to make the data and model reliable, we try to find out the driving distance between different locations. We choose the cities which are on the traffic hinge because it is more convenient to start a journey from a road junction.

In the figure below, we select the important traffic hinges all around the United States (99 in total), and number them from 1 to 99. We extract the real world data of road distance from the map provided by *universalmap.com* and confirm the data with *maps.google.com*. *Table 1* is the Matrix D we use for future calculation which contains all distance information we obtained from the map.

Table 1. Distance Matrix D

	1	2	3	4	5	6	7	8	9	...	99
1	0	inf	inf	inf	inf	inf	inf	inf	inf	...	inf
2	inf	0	174	inf	inf	inf	inf	inf	279	...	inf
3	inf	174	0	113	331	inf	inf	inf	359	...	inf
4	inf	inf	113	0	220	inf	478	inf	inf	...	inf
5	inf	inf	331	220	0	356	413	inf	inf	...	inf
6	inf	inf	inf	inf	356	0	220	388	inf	...	inf
7	inf	inf	inf	478	413	220	0	523	inf	...	inf
8	inf	inf	inf	inf	inf	388	523	0	inf	...	inf
9	inf	279	359	inf	inf	inf	inf	inf	0	...	inf
...	...	...	...	...	...	...	...	...	...	...	...
99	inf	inf	inf	inf	inf	inf	inf	inf	inf	...	0

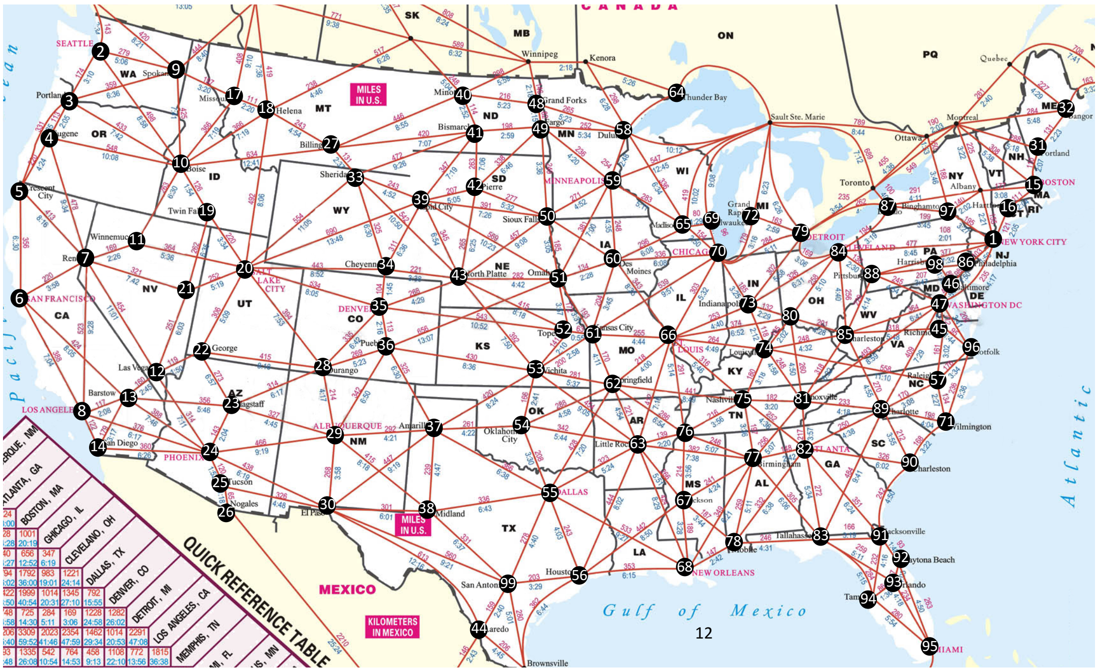


Figure 1. The transportation map containing 99 spots (2)

Let $G(V, E)$ represents the undirected liaison graph *Figure 1*;

v_i represents the point i ;

(v_i, v_j) represents the non-straight route connecting i and j;

$d(v_i, v_j)$ represents the length of the non-straight route connecting i and j.

Since *Figure 1* denotes 99 locations in total, we obtain a matrix of distances between all locations:

$$A^{(0)} = \{a_{ij}^{(0)}\}_{99\times 99}$$

The Floyd-Warshall Algorithm defines

$$a_{ij}^{(0)} \begin{cases} o & i = j \\ d(v_i, v_j) & i \neq j, (v_i, v_j) \in E \\ \infty & i \neq j, (v_i, v_j) \notin E \end{cases}$$

- If there is a route directly connecting two points, then the length of the route is the value of the correspondent element in the matrix.
- If the two points overlap, the value of the correspondent element is zero.
- If the two points are neither overlapped nor connected directly, the value of the correspondent elements will be infinitive.

  In addition, when $k = 1,2,3 \dots, n$:

$$a_{ij}^{(0)} = \min\left\{a_{ij}^{(k-1)}, a_{ik}^{(k-1)} + a_{kj}^{(k-1)}\right\}, \qquad i/j = 1,2,3, \dots, n$$

Therefore, when $k = n$, $A^{(n)} = \{a_{ij}^{(n)}\}_{99\times 99}$ calculates the shortest distance between all points.

Through the operation of our optimized Floyd Algorithm in Matlab[1], we acquire matrix A. By defining the distance d that a one-day ground shipping service can reach within one day, we will know the capacity[2] of each of the 99 spots.

Table 2. Shortest Route Distance Matrix A

	1	2	3	4	5	6	7	8	9	...	99
1	0	3051	3125	3238	3262	3069	2849	2861	2772	⋯	1866
2	3051	0	174	287	505	861	765	1249	279	⋯	2280
3	3125	174	0	113	331	687	591	1075	359	⋯	2215
4	3238	287	113	0	220	576	478	964	472	⋯	2244
5	3332	505	331	220	0	356	413	744	690	⋯	2166
6	3139	861	687	576	356	0	220	388	1046	⋯	1810
7	2919	765	591	478	413	220	0	523	950	⋯	1766
8	2915	1249	1075	964	744	388	523	0	1341	⋯	1480
9	2772	279	359	472	690	1046	857	1341	0	⋯	2119
...	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
99	inf	inf	inf	inf	inf	inf	inf	inf	inf	...	0

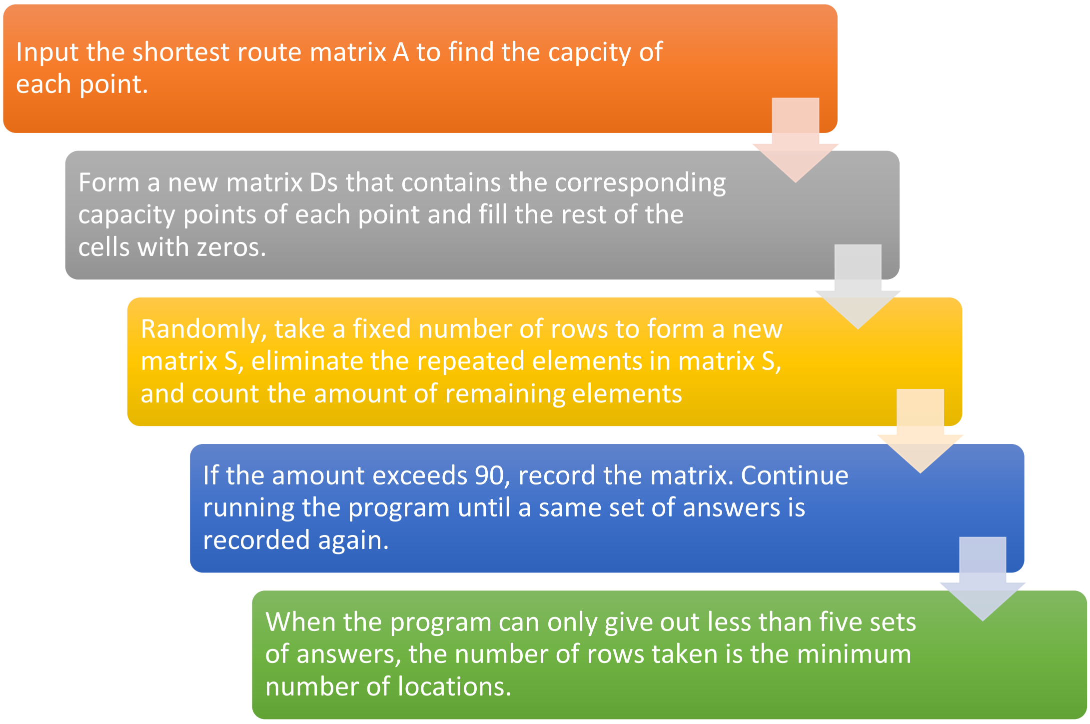
The next step is to decide the exact number of warehouses and their locations. We are able to calculate the capacity of each spot easily through "find(A<=d)", and we can even know the maximum capacity of one spot with defined distance.

We observe that the spots that have big capacities locate closely, meaning by simply adding up the large capacities until the sum exceeded the total number of locations 99 will be a failing attempt because some spots are counted repeatedly while others are not really covered. If the least amount of spots chosen are able to cover the rest of all spots, then it will be the minimum number we are looking for. Considering the complexity, we use Matlab[3] to help us achieve the task. It flows as

[1] Program codes see Appendix A

[2] Capacity refers to the number of spots the delivery service is able to reach within a defined distance in one day.

[3] Code see Appendix



In addition, through observation of the UPS ground-shipping map, we find that the approximate distance UPS delivery is able to reach is between 320-370 miles. We have run our program several times and conclude that a slight difference in distance limit will not make a big difference in the result, so that we use 360 miles as the distance limit of one-day ground shipping distance limit in our further calculations. The method we use to estimate the distance limit is illustrated below:

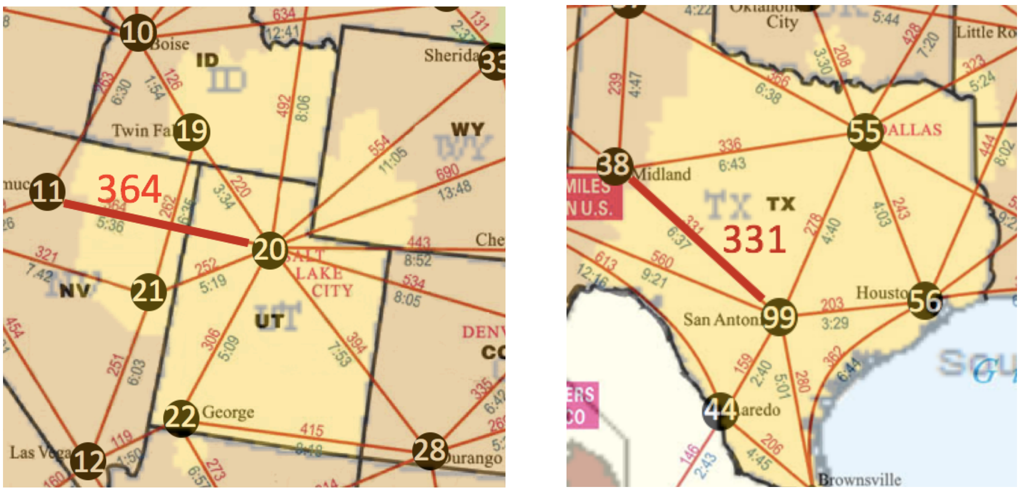


Figure 2. The method used to determine distance

By inputting a random ZIP code into the UPS map system, we are able to obtain a ground-shipping map. We then combine it with our route information map. We choose the spot that lies on the center of the “within one-day” area, and then read the distances from that spot to the spots that lie on the fringe of the “within one-day” area. The largest one will be the approximated maximum distance. The diagrams above are just two examples of how we do the approximations. We have chosen ten more different places to confirm that the approximate distance ranges between 320 to 370 miles; and in future calculation, we define d to be 360 miles.

2.5 Strengths and Weaknesses

Strengths

- Our model is based on the real highway system, instead of on hypothetical distances. It does not only apply to UPS service, but also other delivery services, as long as the maximum distance for one day is defined.
- The model may give out more than one location solution for the minimum number so that we can compare one with another to produce the best one if necessary.

Weaknesses

- The model does not contain a reasonable calculation for defining the maximum distance of “one-day ground shipping”. Instead, it is finished by manpower.
- As already stated in the assumptions, the model is still not concrete enough to predict the precise location for warehouse, but only the city where it should be placed.
- In order to calculate the minimum number of locations, we take all solutions that achieve above 90% coverage rate. In the real world scenario, we may have left few spots that are unable to be reached within one day, but still the number will be very small.

3. Mathematical Model Part 2

3.1 Model Approach

The purpose of our second model is to take sales tax into consideration when selecting warehouse locations. Since it is impossible to minimize every consumer's tax liability, we will need to sacrifice a small group of consumers so that the majority will enjoy less burden on sales tax.

Respecting the fact that different states will have a different amount of consumption on commodities, we think states that have greater consumption should be prioritized as they contain a greater amount of consumers. Ultimately, we want to build a model that evaluates the consumption of each state so that we are able to weight all possible locations and hand out the best one.

3.2 Initial Model

Fundamentally, we think that the average consumption of one state is related to the population of the state and the disbursement ability of individuals. Therefore, we come up with the equation:

$$Q = A * P * G$$

Where A is a constant coefficient, P is the population of each state, and G is the Gross Domestic Product (GDP) of each state. Then, we add two coefficient indexes α and β to the function Q to make it more reliable

$$Q = A * P^{\alpha} * G^{\beta} \quad [1]$$

In order to obtain the value of coefficients, we attempt to apply real world data to the data. We find the annual report of Target Corporation online. Target Corporation is the second-largest discount retailer in the United States. We think it is similar to the company that we analyze. Therefore, we use their 2015 sales per capita data (3) to find the regression equation of Q.

We take the natural logarithm of each side of the function so that equation [1] becomes

$$\ln(Q) = \ln(A) + \alpha \ln(P) + \beta \ln(G)$$

By doing so, we turn a two-variable exponential function into a two-variable linear equation, making it easier for us to regress the relationship between P and G. We input the state population, GDP, and sales per capita into the equation:

Table 3. The data for state population, GDP and sales

State	Population (4)	LP	GDP (5)	LG	sales	LS
OR	4,028,977	6.605194788	228120	5.358163363	175	2.243038049
MT	1,032,949	6.01407888	45799	4.660855995	250	2.397940009
CO	5,456,574	6.736920049	318600	5.503245771	350	2.544068044
WY	586,170	5.768023588	40170	4.603901832	50	1.698970004
SD	858,469	5.933724617	45415	4.657199319	250	2.397940009
LA	4,670,724	6.669384205	253517	5.404007087	125	2.096910013
AL	4,858,979	6.686545022	209382	5.320939344	125	2.096910013
GA	10,214,860	7.009232419	501241	5.700046588	175	2.243038049
NY	19,795,791	7.29657286	1455568	6.163032499	175	2.243038049
MO	6,083,672	6.784165791	290713	5.463464453	250	2.397940009
OK	391,338	5.592552021	179835	5.254874219	175	2.243038049
NC	10,042,802	7.0018549	509718	5.70732997	175	2.243038049
ND	756,927	5.879053997	53686	4.729861047	350	2.544068044
WI	5,771,337	6.761276434	300699	5.478131984	250	2.397940009
NM	2,085,109	6.319128763	90810	4.958133676	125	2.096910013
VA	8,382,993	6.923399104	480876	5.682033102	250	2.397940009
NE	1,896,190	6.277881852	112208	5.050023822	250	2.397940009
ME	1,329,328	6.123632153	55137	4.741443132	125	2.096910013
AZ	6,828,065	6.834297647	298204	5.474513465	250	2.397940009
OH	11,613,423	7.064960245	599093	5.777494245	175	2.243038049
UT	2,995,919	6.476530067	148225	5.170921459	175	2.243038049
ID	1,654,930	6.218779629	65202	4.814260917	50	1.698970004
IA	3,123,899	6.494696984	171532	5.234345151	250	2.397940009
KY	4,425,092	6.645922304	194578	5.289093735	125	2.096910013
MI	9,922,576	6.996624434	468029	5.670272764	175	2.243038049
FL	20,271,272	7.306881001	893189	5.950943366	250	2.397940009
SC	4,896,146	6.68985436	199256	5.299411408	125	2.096910013
WV	1,844,128	6.265791062	71123	4.852010067	50	1.698970004
PA	12,802,503	7.107294886	684313	5.83525479	175	2.243038049
MD	6,006,401	6.778614323	365209	5.562541471	250	2.397940009
TX	27,469,114	7.438844652	1639375	6.214678308	250	2.397940009

IL	12,859,995	7.1092408	771896	5.88755879	250	2.397940009
MA	679,442	5.832152389	478941	5.680282017	250	2.397940009
CT	3,590,886	6.555201618	262212	5.418652563	250	2.397940009
WA	7,170,351	6.855540416	449404	5.652636934	250	2.397940009
KS	2,911,641	6.464137826	149090	5.173448515	250	2.397940009
AR	2,978,204	6.473954443	123424	5.091399617	50	1.698970004
NV	2,890,845	6.461024806	141204	5.149847	250	2.397940009
MN	5,489,594	6.739540226	334780	5.524759505	350	2.544068044
TN	6,600,299	6.81956361	310276	5.491748184	175	2.243038049
MS	2,992,333	6.476009922	106880	5.028896445	50	1.698970004
IN	6,619,680	6.820836996	331126	5.519993283	175	2.243038049
NJ	8,958,013	6.952211688	579379	5.76296275	250	2.397940009
CA	39,144,818	7.592674278	2448467	6.388894255	350	2.544068044

And we use computer programming to help us regress the model. We find the values of coefficients to be A=3.437, $\alpha = -0.1828$, $\beta = 0.4134$. By replacing the coefficients with numerical values, we obtain:

$$Q = 3.437 * P^{-0.1828} * G^{0.4134}$$

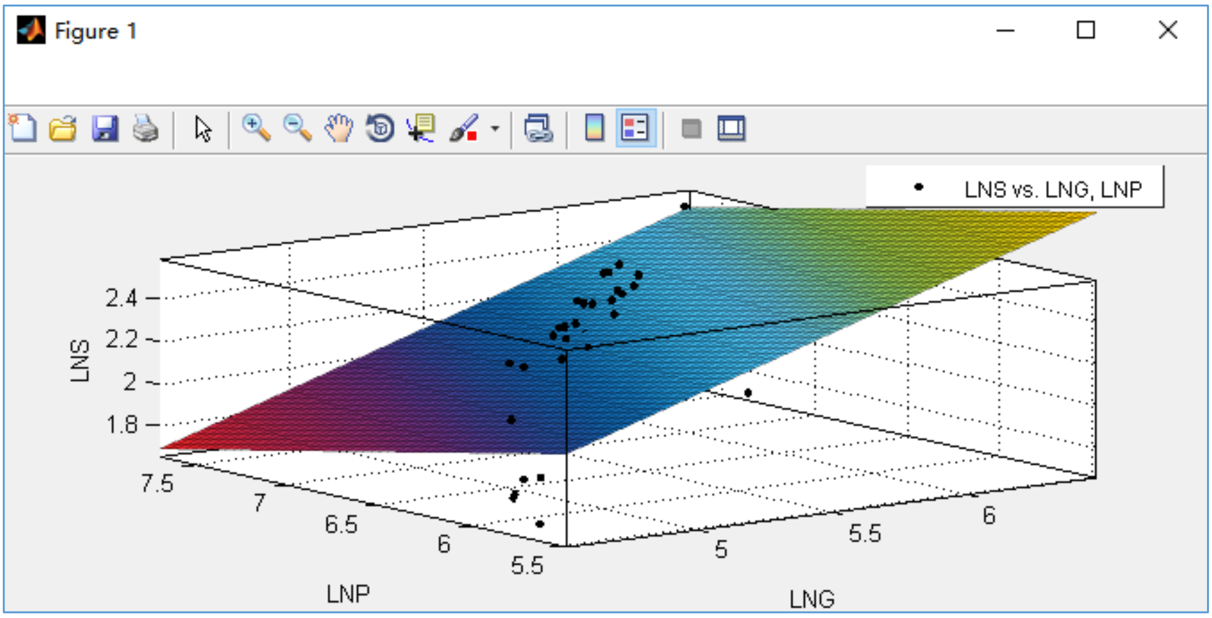


Figure 3. The graph of regression model

We consider the model as highly reliable because it is based on real sales data. However, we soon realize that by only integrating current population and GDP into the model, our model will not yield any future credibility. Since the basic purpose for the sale company of adding delivery service is to bring more future benefits, we want to delve further into the two factors to build a concrete and reliable model for future predictions.

3.3 Factor Improvements

3.3.1 Population

Our group believes that population is one of the important indicators of state consumption since it reflects the amount of potential consumers, which is connected with our goal to minimize the tax liability for most customers. In order to take future population growth into account, we studied Leslie Matrix (6) to improve our model.

Leslie matrix is used to predict the further population based on the data we get from the previous censuses. In this model, we assume that the population grows due to the reproduction of females. Therefore, we use the change of female population as the study object.

We divide the population of the U.S. into 91 groups, each of which has an interval of 1-year-old (population above 90 are counted as 90). The same as how we deal with the ages, we also make the time dispersed as time intervals. We use the data of female population, and come up with the female population projection.

$x_i(k)$is the female population of the i_{th} age interval we get from the k_{th} observation; thus we get:

$$x(t) = [x_0(t), x_1(t), \dots, x_{90}(t)]^T$$

According to Leslie Matrix: at the time interval $t+1$, the population of the 1st age is the sum of the reproduction of each age interval at time interval t. This concept can be expressed as

$$x_0(t+1) = \sum_{i=0}^{90} b_i(t)x_i(t) \quad [2]$$

In the equation above,

$x_0(t)$represents the population of the i_{th} age interval at time interval t .

$b_i(t)$represents the female birth rate of the i_{th} age interval at time interval t .

$d_i(t)$represents the female birth rate of the i_{th} age interval at time interval t .

The data above we can get from the website www.census.com.

The population of the $(t+1)^{\text{th}}$ age at time interval $(t+1)$ is the survivals of the ith age at time interval t , which is

$$x_{i+1}(t+1) = p_i(t)x_i(t), \qquad i = 0,1,2,3,\dots,90 \quad [3]$$

The matrix of female population for each age at time interval t is

$$x(t) = [x_0(t), x_1(t), \dots, x_{90}(t)]^T \quad [3.8]$$

Thus the birth rate b_i of female and the survival rate p_i make up the matrix:

$$\boldsymbol{L} = \begin{bmatrix} b_0(t) & b_1(t) & \cdots & b_{89}(t) & b_{90}(t) \\ p_0(t) & 0 & \cdots & 0 & 0 \\ 0 & p_1(t) & \cdots & 0 & 0 \\ 0 & 0 & p_2(t) & \cdots & 0 \\ & \vdots & & \vdots & \\ 0 & 0 & \cdots & p_{90}(t) & 0 \end{bmatrix}_{91\times91}$$

Combining equations [2] and [3], we obtain that:

$$x(t) = L^t x(0)$$

In the equation above, $x(t)$ is the female population for each age interval at time interval t , according to matrix L and the age at time interval t we predict the total female population at time interval t .

Therefore, if the ratio of female population versus total population is fixed, we can easily predict the future population. Applying the *Leslie Matrix* to predict the population, we calculate the total population for the whole country at time interval t is:

$$x(t) = \frac{L^t x(0)}{p}$$

In the equation above, p is the average ratio of female population versus the total population.

According to the data we researched from the US Census Bureau, in 2015, the ratio of female population versus the total population p is 0.508 (7).

Therefore, the total population $N(t)$ at time interval t should be:

$$N_t = \sum_{i=o}^{m} x_i(t) \quad [4]$$

$$P_{it} = N_t \omega_i \quad [5]$$

Notice that ω_i is the average ratio of the population of the i^{th} state versus the total population. When i = 1,2,3,…,48, we are able to calculate the population of all of the 48 states in 2016-2020.

3.3.2 Level of Consumption

Level of consumption is also another important factor to be considered. If the general level of consumption of one state is reasonably high, then we need to put more effort to find the location that best fits the interest of customers. We think the level of consumption is largely decided by the income of individuals. Moreover, with respect to the Absolute Income Hypothesis (8) by Keynes, a macroeconomic theory of consumption, our group conclude that the income gap within a state will have a significantly negative impact on the consumption of that state.

The Absolute Income Hypothesis identifies the relationship between income and consumption. The theory asserts that as income rises, consumption will also rise but not necessarily at the same rate. Besides, when applied to a cross section of a population, rich people are expected to consume a lower proportion of their income than poor people. It states that:

$$\mathrm{C}_t = MC * Y_t \quad [6]$$

where C_t is consumption at time t, MC is the marginal propensity to consume, and Y_t is income at time t. Though the theory only states the general relationship between income and consumption, through our own analysis, we find out the relationship between the income gap.

We divide the main body of consumption into two groups: the high income group and the low income group. Although in reality there are more diverse levels of income, for the sake of mathematical proof, we eliminate these possibilities. We define the net income of the high income group as Y_h, the net income of the low income group as Y_l, and the total income of people as Y. Then we will have

$$\mathrm{Y} = Y_h + Y_l$$

Meanwhile, we also define the marginal propensity to consume of each group to be MC_h and MC_l. Since according to Keynes' theory, the high income group has a lower marginal propensity to consume than the low income group does, we are able to know that $0 < \mathrm{MC}_h < \mathrm{MC}_l < 1$. We input the newly defined variable into equation [6], so that we have

$$\mathrm{C}_h = M\mathrm{C}_h * Y_h$$

$$\mathrm{C}_l = M\mathrm{C}_l * Y_l$$

$$\mathrm{C} = \mathrm{C}_h + C_l$$

Let r be the ratio of the net income of two groups, meaning $\mathrm{r} = \frac{Y_h}{Y_l}$, so a larger r will reflect a larger income gap, and r>1. Replacing Y_h and Y_l with r, we obtain that

$$Y_h = \frac{r}{1+r}\,Y$$

$$Y_l = \frac{1}{1+r}\,Y$$

Therefore, the total consumption function becomes

$$C = MC_h * \frac{r}{1+r} Y + MC_l * \frac{1}{1+r} Y$$

In order to obtain the relationship between r and C, we take the first derivate of the function C:

$$\frac{dC}{dr} = (MC_h - MC_l) \frac{1}{(1+r)^2} Y$$

Since $MC_h < MC_l$, so that $\frac{dC}{dr}$<0, meaning the total consumption and income gap have an inverse relationship. Therefore, we conclude that enlarging the income gap within a state will decrease the consumption of the state, and narrowing the income gap within a state will increase the consumption of the state.

3.4 Final Model

Now that we have developed out factors into robust models, the next step is to input the factor to our initial model. Finally, we want to build up a model that helps us weight each of the solutions so that we can find the best solution with tax consideration.

For future state populations, combing equations [4] and [5], we obtain that

$$P_{it} = \sum_{i=o}^{m} x_i(t) * \omega_i$$

For state consumptions, as we have mathematically proven the relation between income gap and consumption, we'd like to replace the initial GDP factor *G* with the consumption level of each state *C*, and we have:

$$C = \frac{k}{r}$$

k is a constant number in the above inverse function, however, we will eliminate it in the later calculation because the constant k for each state is the same, it has no influence on reflecting the differences between each location.

By congregating the above two equations into the initial model, we get

$$Q = \frac{(\sum_{i=o}^{m} x_i(t) * \omega_i)^{\alpha}}{r^{\beta}}$$

Notice that we also eliminate the constant coefficient A for the same reason as the constant number k.

We now need to make Q a determinant for us to filter out the best solution. According to the problem statement, any online order delivered to a location outside a state where a warehouse is located will NOT be taxed. Therefore, we believe the most preferable situation is when states that have higher consumptions can buy the products from nearby states, unless the states have a 0% sales tax.

In order to achieve so, we develop a scoring function:

$$W = \sum_{i=1}^{n} T_n \, Q_n$$

Where n is the minimum number of locations, T_n is the tax of the state that the warehouse located in, and Q_n is the level of consumption of the state that the warehouse located in. Substituting the Q_n with its extension, we have our final version to be:

$$W = \sum_{i=1}^{n} \frac{T_n * (\sum_{i=o}^{m} x_i(t) * \omega_i)^{-0.1828}}{(\frac{Y_h}{Y_l})^{0.4134}} * 10^2$$

The function is multiplied by a factor of 10^2 so that each individual score will lie between a 0-100 scale. The working mechanism of this evaluation function W is simple. While we are ensuring the majority of consumers has a least tax burden, we also want the rest of the people to pay as less as possible, meaning we still prefer to set up the warehouse in a state with lower sales tax rate. By observation of the population date and the income data, we find that these two sets of numbers are huge. Therefore, if the location turns out to be either in a densely populated state or in a state with high consumption, the scoring function will give out a high score, meaning the location is not an ideal one. Consequently, the solution which scores lowest will be the best solution of the problem.

3.5 Strengths and Weaknesses

Strengths

- The model is continuous at predicting the future as long as it is backed up with adequate research data.
- The model provides a good criterion for us to select the best solution since it approaches to the problem comprehensively by looking at population growth and level of consumption.

Weaknesses

- Both population growth model and income gap model can be used as excellent indicators for finding the best solution of this problem. However, these two models may not be rigorously applied to other population or consumption related issues because they are over-simplified.

4.. Results

4.1 Minimum Number and Locations

By applying Floyd Algorithm and running our computer program, we obtain the minimum number to be 17. Our program also tells the specific corresponding spots, which are found to be the following three sets:

set	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰
#1	54	79	4	12	41	16	67	17	83	51	13	89	73	30	65	28	10
#2	83	34	24	41	90	74	65	12	88	63	54	17	30	53	4	16	10
#3	24	10	73	65	30	12	78	53	17	83	88	55	34	9	89	16	41

We consider all three sets of equivalent importance; therefore, we would like to present all of the three solutions as our answer to the problem.

4.2 Considering Sales Tax

Following the Algorithm provided in section three, we obtain the scores for the above three sets of solutions.

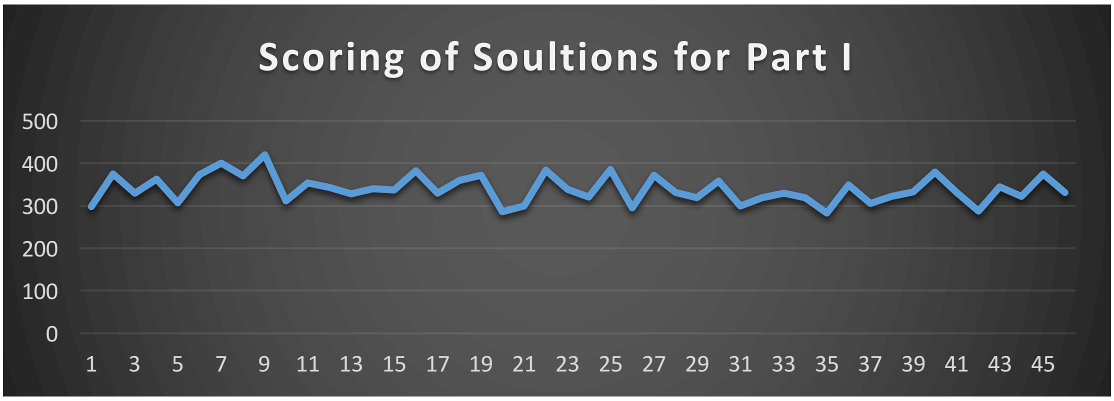
Set #1	298
Set #2	374
Set #3	328.9

Our research data[8] comes from the *May 2015 State Occupational Employment and Wage Estimates*, from the Bureau of Labor Statistics. With careful consideration, we define high wages as annual salary above 100,000 USD and define low wages as annual salary below 25,000 USD. We calculate Y_h and Y_l by summing up the product of the employment of each job and its wage.

To make sure the we find out the best solution, we decide not to choose one immediately from above, but to also include the solutions for 18 locations because it is possible that the scoring of an eighteen-location set is even lower. The computer program has shown us 43 sets of solutions for selecting 18 locations, and the scoring results are:

#4	361.9	#13	328.3	#22	383.4	#31	299	#40	379.6
#5	306.8	#14	340.5	#23	338.8	#32	319.5	#41	331.7

#6	373.6	#15	337.5	#24	319.9	#33	329.5	#42	287.6
#7	400.1	#16	381.9	#25	385.6	#34	318.3	#43	343.9
#8	370.3	#17	329	#26	295.4	#35	283.3	#44	321.8
#9	419.1	#18	359	#27	372.2	#36	349.3	#45	375.4
#10	311.3	#19	371.3	#28	330.6	#37	304.9	#46	330.4
#11	353.7	#20	286.6	#29	318.4	#38	322.4		
#12	343.8	#21	299.1	#30	357.7	#39	332.3		



According to the chart above, we choose the set of solution that scores the lowest, which is the set#35, containing the locations below:

13	89	68	10	70	88	83	24	1	9	34	41	55	50	54	16	66	4

It also shows that for the purpose of lowering tax liability of customers, we need to increase the minimum number of locations to 18.

4.3 Considering Apparel Tax

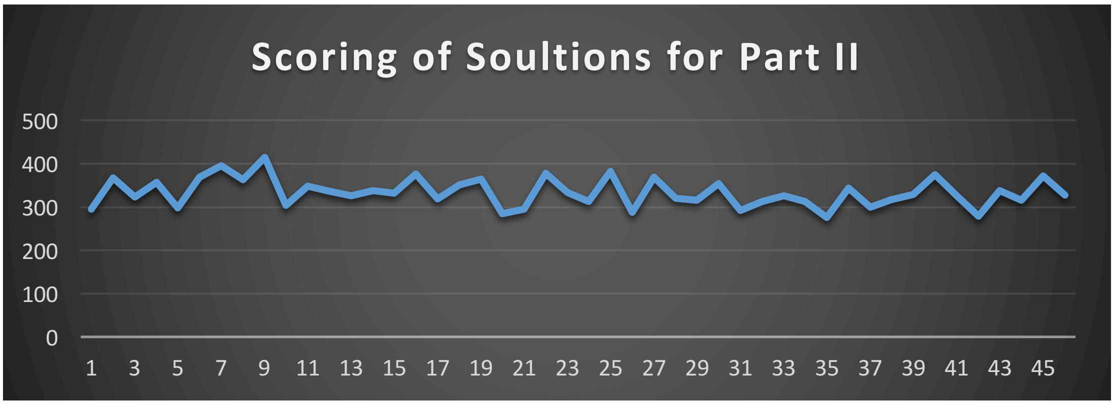
This part is almost the same as the second part, however, we need to take the clothing tax into consideration as well. In fact, people do not spend same amount of money on commodities and apparel goods. Therefore, the value of T_n will need to be considered in a more complicated manner than it was in Part II.

We think the most credible way to balance the two taxes is to look at the average percentage of money the Americans spend on commodities and apparel goods. According to our research data(9) from the Bureau of Labor Statistics, August 2016, it shows that, every month, an average of 16.247% of the urban consumers' expenditure is spent on commodities, while 3.097% of their expenditure is spent on apparels. Therefore the way we calculate the mixed tax is:

$$T_n = \frac{16.247}{16.247 + 3.097} T_s + \frac{3.097}{16.247 + 3.097} T_c$$

And the new scorings we obtain are:

#1	295.7	#11	348.4	#21	295.1	#31	292.2	#41	325.8
#2	368.1	#12	337.3	#22	378.4	#32	313.1	#42	279.3
#3	322.9	#13	326.3	#23	333.9	#33	326.7	#43	338.4
#4	356.8	#14	338.7	#24	312.8	#34	313.6	#44	315.4
#5	297.8	#15	332.2	#25	383.1	#35	275.4	#45	372
#6	369.1	#16	376.4	#26	287.6	#36	344	#46	327.9
#7	396.2	#17	319	#27	369.2	#37	300		
#8	363.7	#18	350.9	#28	320.4	#38	317.2		
#9	415.6	#19	364.4	#29	315.6	#39	329.3		
#10	304	#20	284.4	#30	354.5	#40	375.7		



According to the new scores graded by our improved model, set #35 still scores the least. Therefore, the addition of apparel tax will *not* affect the solution for locations.

5. Analysis and Conclusion

5.1 Result Analysis

In attempt to verify our results, we use circles with radius equivalent to 360 miles on map to represent the area a warehouse location can cover in one day. The center of each circle is placed on our selected locations.

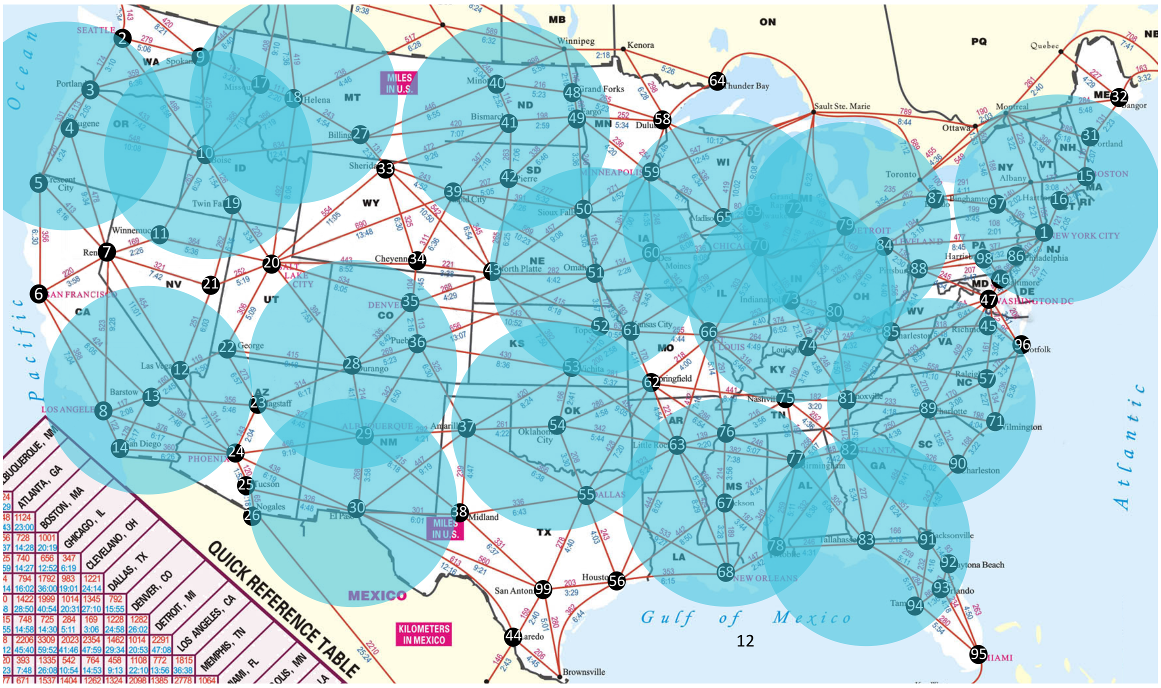


Figure 4. The coverage area for our solution to Part I

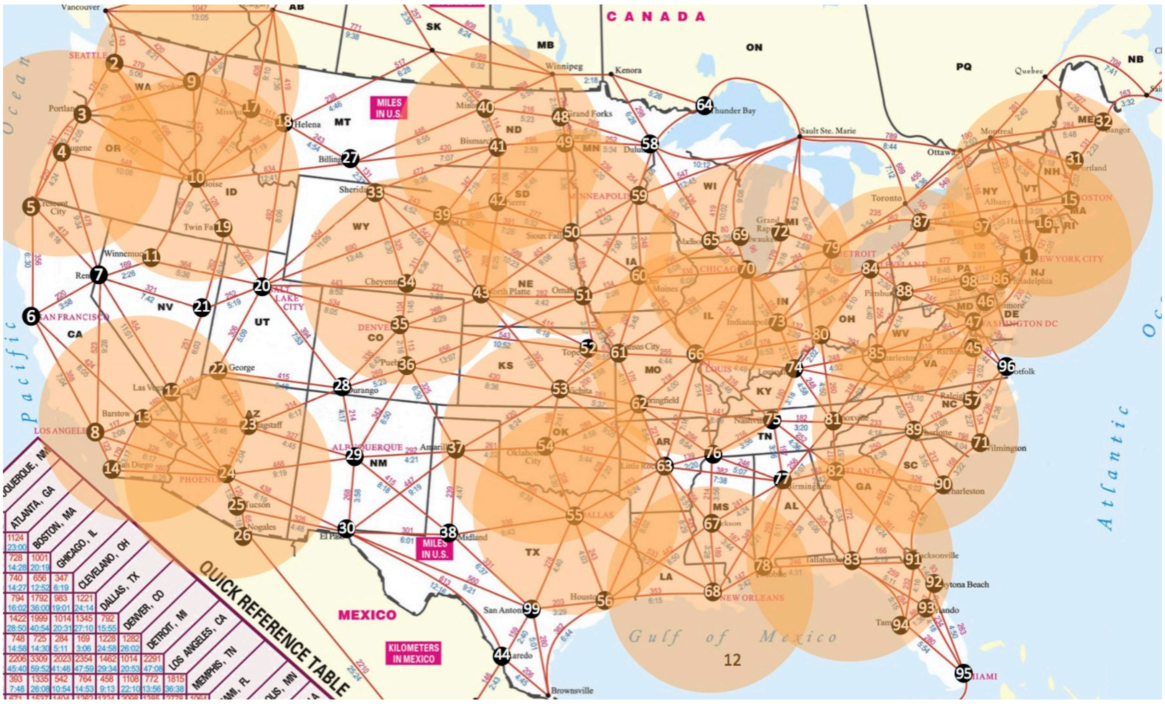


Figure 5. The coverage area for our solution to Part II and III

We observe that there is a little space left uncovered on the graph, but it does not mean our solutions have failed. Since our solutions are based on the road hinges we numbered for simplicity sake, in reality the warehouse does not have to be strictly placed in that specific point. Therefore, we slightly move certain circles to test if these city locations can really cover the 48 continents of US. And we obtain:

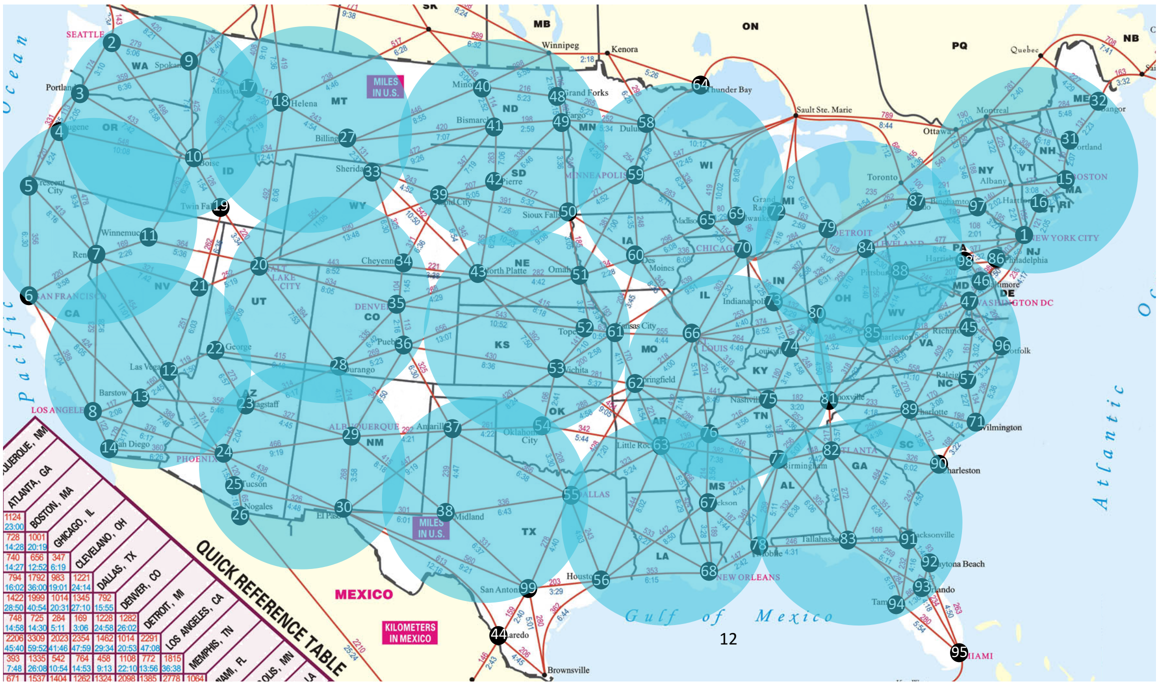


Figure 6. The edited coverage area for our solution to Part I

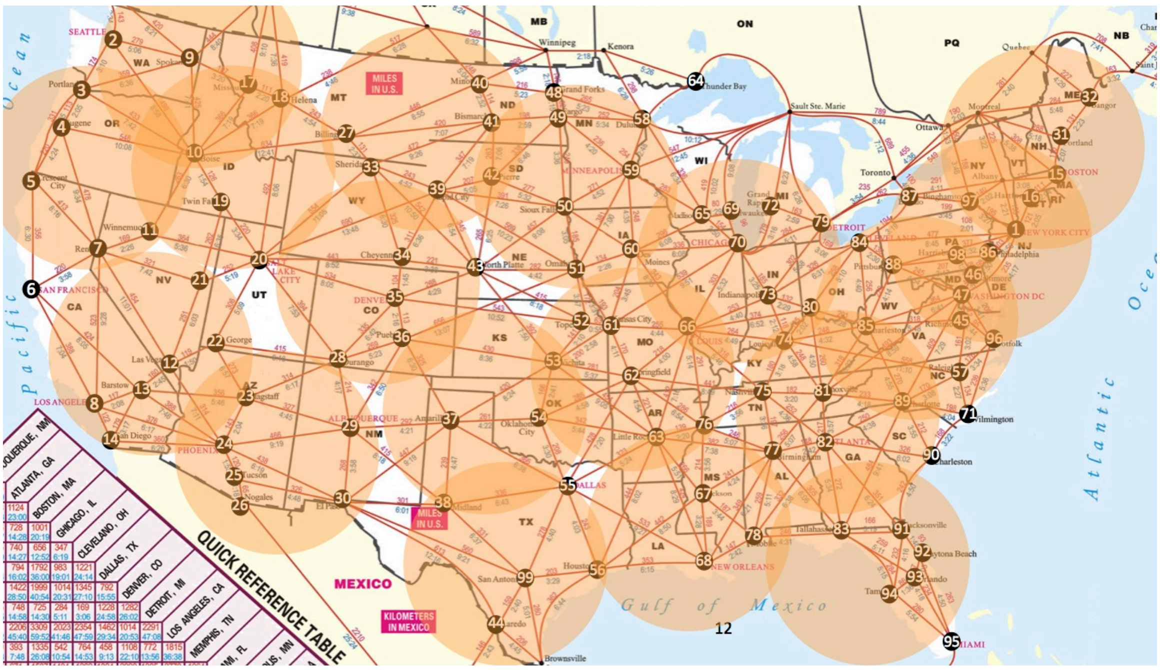


Figure 7. The edited coverage area for our solution to Part II and III

From the new graphs we conclude that our solution is reliable and achievable. Since at first our model is aimed to realize a more than 90% percent coverage, it explains why there is still little space uncovered. In general, we have achieved the task to give out the minimum number of locations and the best solution.

5.2 Sensitivity Analysis

Besides verifying the locations, we also make sensitivity analysis of our second model to prove our grading system is reliable. If the model varies accordingly with the change in population and income gap, then our model is sensitive to change and is reasonable. Graphs below are the result of our analysis:

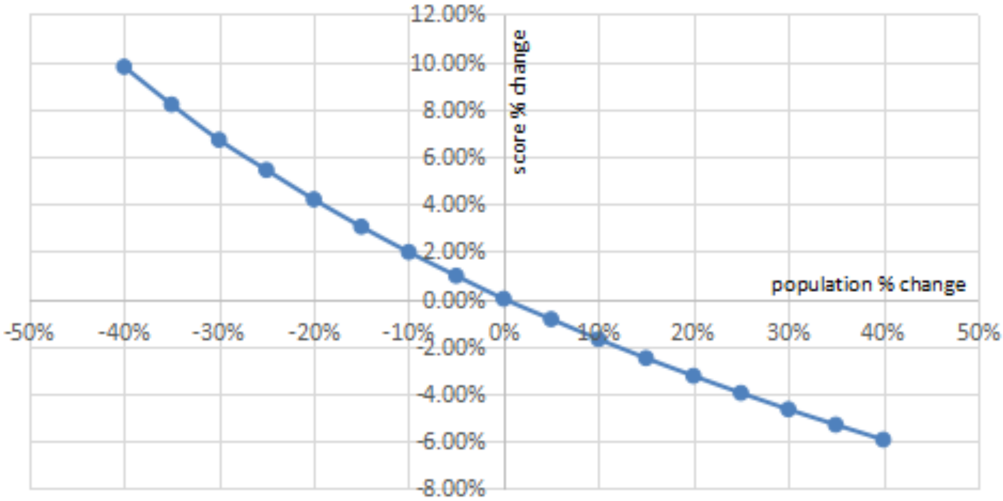


Figure 8. Change of scores if population changes

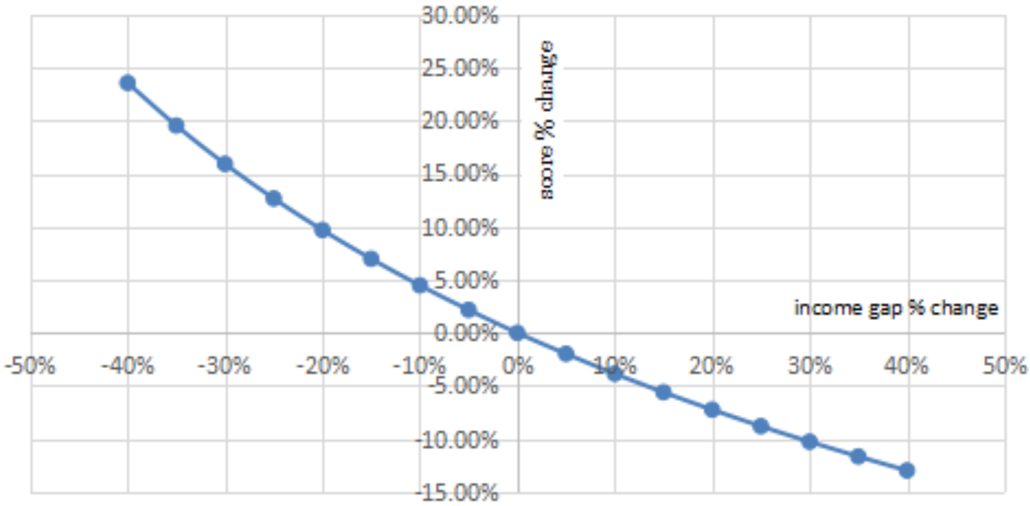


Figure 9. Change of scores if income gap changes

Therefore, we conclude that our grading system is reliable.

5.3 Conclusion

The results of our model show that the minimum location number without tax consideration is 17; the minimum location number with respect to sales tax is the same as that to the addition of apparel tax, which is 18. We have three set of locations for question Part I. Since listing all of them will be time-consuming, we would provide one example of locations to be: Eugene, Boise, Las Vegas, Barstow, Hartford, Missoula, Durango, El Paso, Bismarck, Omaha, Oklahoma City, Madison, Jackson, Indianapolis, Detroit, Tallahassee, Charlotte. The minimum locations for question Part II and Part III are Barstow, Charlotte, Des Moines, Boise, Chicago, Pittsburg, Tallahassee, Phoenix, New York City, Spokane, Cheyenne, Bismarck, Dallas, Sioux Falls, Oklahoma City, Hartford, St. Louis, and Eugene.

Generally speaking, we put forward two models to solve the problem; one is to determine the minimum number of locations, and the other is to determine the best locations. We think the good part of our model is that it is based on the highway transportation system so that our result will be realistic. Notice that our model does not only apply to UPS service. As we consider the vast choices of delivery service, as long as the maximum one-day distance of service is acquirable, the company can use our model to predict the minimum number of warehouse locations. We have also made deep consideration to measure the future consumption of each state so that our solution will make sure it is profitable to place warehouses in certain locations. The defect of our model is that it can only show the possible city location for warehouses. In fact, knowing which city to place the warehouse is not enough for the company because we do not know which specific area in the city is suitable for building a warehouse, or whether it is fiscally profitable to build a warehouse in that city. However, if we do so, it will simply explode our burden of research, and there are still many situations that we are unable to predict.

Reference

[1] Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L. (1990). *Introduction to Algorithms* (1st ed.). MIT Press and McGraw-Hill. ISBN 0-262-03141-8.

[2] http://www.hydegroup.net/Item/Show.asp?m=1&d=264

[3] 2015 Annual Report. Target Brands, Inc. https://corporate.target.com/annual-reports/2015/financials/2015-sales-per-capita

[4] State population by rank. U.S. Census Bureau, Population Division.

[5] "Gross domestic product (GDP) by state (millions of current dollars)". U.S. Department of Commerce. Bureau of Economic Analysis.

[6] XIE Sheng-lan，KE Zhong-yi，DING Xiao-tong. Discussion on predicting the future population growth. Department of Mathematical Sciences，Huizhou University，Huizhou，China

[7] State population by rank. U.S. Census Bureau, Population Division.

[8] May 2015 State Occupational Employment and Wage Estimates. U.S. Bureau of Labor Statistics. http://stats.bls.gov/oes/current/oessrcst.htm

[9] Consumer Price Index—October 2016. U.S. Bureau of Labor Statistics. http://www.bls.gov/news.release/pdf/cpi.pdf

Appendix

Floyd Algorithm

```
A=Data;
n=length(A);
path=zeros(n);
for k=1:n
    for i=1:n
        for j=1:n
            if A(i,j)>A(i,k)+A(k,j)
                A(i,j)=A(i,k)+A(k,j);
                path(i,j)=k;
            end
        end
    end
end
A
path
```

Location Selection

```
X=size(99,750);
z=750;
for i=1:z
    X(N(i),i)=M(i);
    i=i+1;
end
X;
clc
j=1;
n=17;
E=size(100,n);
S = size(Ds,1);
for i=1:nchoosek(44,n);
SampleRows = randperm(S);
SampleRows = SampleRows(1:n);
```

```
SampleDs = Ds(SampleRows,:);
Size=size(unique(SampleDs));
a=Size(1)
if a>=90
    for k=1:n;
        E(j,k)=SampleRows(k);
        k=k+1;
    end
    disp('it works!')
    j=j+1;
end
    i=i+1;
end
```

Regress

```
clc
LNS=[2.24303804900000,2.39794000900000,2.54406804400000,1.6989700040
0000,2.39794000900000,2.09691001300000,2.09691001300000,2.2430380490
0000,2.24303804900000,2.39794000900000,2.24303804900000,2.2430380490
0000,2.54406804400000,2.39794000900000,2.09691001300000,2.3979400090
0000,2.39794000900000,2.09691001300000,2.39794000900000,2.2430380490
0000,2.24303804900000,1.69897000400000,2.39794000900000,2.0969100130
0000,2.24303804900000,2.39794000900000,2.09691001300000,1.6989700040
0000,2.24303804900000,2.39794000900000,2.39794000900000,2.3979400090
0000,2.39794000900000,2.39794000900000,2.39794000900000,2.3979400090
0000,1.69897000400000,2.39794000900000,2.54406804400000,2.2430380490
0000,1.69897000400000,2.24303804900000,2.39794000900000,2.5440680440
0000];
LNP=[6.60519478800000,6.01407888000000,6.73692004900000,5.7680235880
0000,5.93372461700000,6.66938420500000,6.68654502200000,7.0092324190
0000,7.29657286000000,6.78416579100000,5.59255202100000,7.0018549000
0000,5.87905399700000,6.76127643400000,6.31912876300000,6.9233991040
0000,6.27788185200000,6.12363215300000,6.83429764700000,7.0649602450
0000,6.47653006700000,6.21877962900000,6.49469698400000,6.6459223040
0000,6.99662443400000,7.30688100100000,6.68985436000000,6.2657910620
0000,7.10729488600000,6.77861432300000,7.43884465200000,7.1092408000
```

```
0000,5.83215238900000,6.55520161800000,6.85554041600000,6.4641378260
0000,6.47395444300000,6.46102480600000,6.73954022600000,6.8195636100
0000,6.47600992200000,6.82083699600000,6.95221168800000,7.5926742780
0000];
LNG=[5.35816336300000,4.66085599500000,5.50324577100000,4.6039018320
0000,4.65719931900000,5.40400708700000,5.32093934400000,5.7000465880
0000,6.16303249900000,5.46346445300000,5.25487421900000,5.7073299700
0000,4.72986104700000,5.47813198400000,4.95813367600000,5.6820331020
0000,5.05002382200000,4.74144313200000,5.47451346500000,5.7774942450
0000,5.17092145900000,4.81426091700000,5.23434515100000,5.2890937350
0000,5.67027276400000,5.95094336600000,5.29941140800000,4.8520100670
0000,5.83525479000000,5.56254147100000,6.21467830800000,5.8875587900
0000,5.68028201700000,5.41865256300000,5.65263693400000,5.1734485150
0000,5.09139961700000,5.14984700000000,5.52475950500000,5.4917481840
0000,5.02889644500000,5.51999328300000,5.76296275000000,6.3888942550
0000];
X=[ones(44,1) LNP' LNG'];
[B,bint,r,rint,stats] = regress(LNS',X);
B
```