Central Hub Train Station

Makayli Taylor

Spring 2024

Introduction

Washington State will be considering implementing a high speed train station that allows people to get to and from selected major cities. The goal is to find the optimal location for a central hub that secondary stations can lead to.

The secondary stations will be located in selected major cities and college towns throughout the state. With at least one in each region of the state.

Olympic	Northwest	North Central	Eastern	South Central	Southwest
Tacoma	Seattle Everett Bellingham	Wenatchee	Spokane Pullman	Yakima Kennewick Ellensburg	Vancouver

Because of the irregular shape of Washington, each city will be using their longitude and latitude coordinates as (x, y) respectively (App A). Where (g, h) represents the ideal location of the new hub.

Developing Our Objective Function

Because we are developing a linear program to solve the configured problem, some modifications were required. The general distance function between two coordinates, with respect to our context, is as follows:

$$d = \sqrt{(x_i - g)^2 + (y_i - h)^2}$$

Resulting in the initial program below.

$$\begin{aligned} \min_{a,b} \quad & z = \sum_{i=1}^{55} \left[\sqrt{(x_i - g)^2 + (y_i - h)^2} \right] \\ \text{s.t.} \quad & x_i, y_i, g, h \in \mathbb{R} \end{aligned}$$

Because this is clearly not a linear function some modifications needed to be made. The first was to find an approximate linear equation that described with minimal error to the general distance equation. This approximation was done via linear fitting. For the experimental data points the absolute value of the differences in degrees longitude and latitude for each location were used (p_i and q_i respectively). Using these points resulted in 55 sets of (x,y) values (App. B). These points come from the differences in latitude and longitude between each of our considered cities

It was important to note that the difference in degrees latitude and longitude between locations was the same (i.e. degrees longitude between Seattle and Pullman and Pullman and Seattle was 5.156167° both ways). So it was redundant to consider both values. Similarly, the distance between Seattle and Seattle was also not needed as it would be zero for both latitude and longitude.

$$\begin{aligned} \min_{a,b} \quad & z = \sum_{i=1}^{55} \left[(ap_i + bq_i + c) - \left(\sqrt{p_i^2 + q_i^2} \right) \right] \\ \text{s.t.} \quad & a, b \in \mathbb{R} \end{aligned}$$

Where $(ap_i + bq_i + c)$ is our approximate distance equation, and $\sqrt{p_i^2 + q_i^2}$ is the general distance equation. So, we are trying to minimize the difference between the correct distance value and the approximate one.

The main method used to devise a linear approximation is through finding stationary points of our objective function. This can be done simply using the gradient of z, and solving the system for our coefficients

$$\nabla z(a,b,c) = \begin{bmatrix} \frac{dz}{da} \\ \frac{dz}{db} \\ \frac{dz}{dc} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{55} \left((ap_i + bq_i + c) - \sqrt{p_i^2 + q_i^2}\right) p_i \\ \sum_{i=1}^{55} \left((ap_i + bq_i + c) - \sqrt{p_i^2 + q_i^2}\right) q_i \\ \sum_{i=1}^{55} (ap_i + bq_i + c) - \sqrt{p_i^2 + q_i^2} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{55} ap_i^2 + \sum_{i=1}^{55} bp_iq_i + \sum_{i=1}^{55} +cp_i - \sum_{i=1}^{55} p_i\sqrt{p_i^2 + q_i^2} \\ \sum_{i=1}^{55} ap_iq_i + \sum_{i=1}^{55} bq_i^2 + \sum_{i=1}^{55} cq_i - \sum_{i=1}^{55} q_i\sqrt{p_i^2 + q_i^2} \\ \sum_{i=1}^{55} ap_i + \sum_{i=1}^{55} bq_i + \sum_{i=1}^{55} c - \sum_{i=1}^{55} \sqrt{p_i^2 + q_i^2} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{55} p_i^2 & \sum_{i=1}^{55} p_iq_i & \sum_{i=1}^{55} p_i \\ \sum_{i=1}^{55} p_iq_i & \sum_{i=1}^{55} q_i^2 & \sum_{i=1}^{55} q_i \\ \sum_{i=1}^{55} p_i & \sum_{i=1}^{55} q_i & \sum_{i=1}^{55} 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{55} p_i\sqrt{p_i^2 + q_i^2} \\ \sum_{i=1}^{55} q_i\sqrt{p_i^2 + q_i^2} \\ \sum_{i=1}^{55} \sqrt{p_i^2 + q_i^2} \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{55} p_i^2 & \sum_{i=1}^{55} p_iq_i & \sum_{i=1}^{55} p_i \\ \sum_{i=1}^{55} p_iq_i & \sum_{i=1}^{55} q_i^2 & \sum_{i=1}^{55} q_i \\ \sum_{i=1}^{55} p_i & \sum_{i=1}^{55} q_i & \sum_{i=1}^{55} 1 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{55} p_i\sqrt{p_i^2 + q_i^2} \\ \sum_{i=1}^{55} q_i\sqrt{p_i^2 + q_i^2} \\ \sum_{i=1}^{55} \sqrt{p_i^2 + q_i^2} \end{bmatrix}$$

$$= \begin{bmatrix} 459.748562 & 133.4154904 & 132.2825126 \\ 133.4154904 & 82.15674627 & 55.5140928 \\ 132.2825126 & 55.5140928 & 55 \end{bmatrix}^{-1} \begin{bmatrix} 489.36209 \\ 169.56659 \\ 153.3967657 \end{bmatrix}$$

$$= \begin{bmatrix} 0.8509 \\ 0.5675 \\ 0.1697 \end{bmatrix}$$

These found coefficients provide us with or linear fit to the existing data. Creating the following linear approximation:

$$0.8509p_i + 0.5675q_i + 0.1697$$

Where $p_i = |x_i - g|$ and $q_i = |y_i - h|$. From the used data set there is an error of approximately 8.27% between our approximation and the general distance function.

Considered Constraints

Equality Constraints

Stated in our original problem statement the proposed location of a hub is allowed to be at the same coordinates as one of our cities. Though this was relatively unlikely for any cities not within the North or South Central regions of Washington.

Though there were no initial equality constraints to be accounted for, because of the substitution involved in our objective function there were absolute value terms to account for.

$$z = \sum_{i=1}^{11} 0.8509|x_i - g| + 0.5675|y_i - h| + 0.1697$$

This was done by adding equality constraints. Let $j_i + k_i = |x_i - g|$ and $m_i + n_i = |y_i - h|$ such that,

$$j_i + k_i = \begin{cases} x_i - g & \text{if } x_i \geq g \\ g - x_i & \text{if } x_i < g \end{cases}, \quad m_i + n_i = \begin{cases} y_i - h & \text{if } y_i \geq h \\ h - y_i & \text{if } y_i < h \end{cases}$$

This results in more matrix friendly equality constraints

$$\begin{aligned} j_i - k_i &= x_i - g & \qquad m_i - n_i &= y_i - h \\ \Rightarrow g + j_i - k_i &= x_i & \Rightarrow h + m_i - n_i &= y_i \end{aligned}$$

The objective function after this implementation simply substitutes the sums that replace our absolute value terms. This relates the newly added values in the objective function

$$z = \sum_{i=1}^{11} 0.8509(j + k) + 0.5675(m + n) + 0.1697$$

Box Constraints

In terms of box constraints for each variable:

g is the longitude coordinate and thus must be between -180 and 180 degrees. Because we are looking in the state of Washington only we can use the minimum and maximum longitudinal values of the state itself. This is not necessarily but it ensures on quick inspection that the proposed location is within Washington.

$$-125 \leq g \leq -115$$

h is the latitude coordinate and similarly must be between 0 and 90 degrees. We can also constraint h to the minimum and maximum latitudes of the state similar to above.

$$40 \leq h \leq 50$$

$\boldsymbol{j_i, k_i, m_i, n_i}$ These variables were created to adjust for the absolute value aspect of this problem. Thus, these values must be greater than 0.

$$0 \leq h \leq \infty$$

Proposed Solution

Complete Program

This program has two general equality constraints, with each having 11 options as x_i and y_i are vectors with size 11. The final linear program is as follows

$$\begin{aligned}
\min_{g,h} \quad & z = \sum_{i=1}^{11} 0.8509 j_i + 0.8509 k_i + 0.5675 m_i + 0.5675 n_i + 0.1697 \\
\text{s.t.} \quad & g + j_i - k_i = x_i \\
& h + m_i - n_i = y_i \\
& g, h, j_i, k_i, m_i, n_i \in \mathbb{R} \\
& i = 1, 2, \ldots, 11
\end{aligned}$$

Unweighted Solution

Because our initial solution considers the city coordinates without the weight of population coming into effect there are no further modifications that are needed from the above linear program. We then have:

$$
\begin{aligned}
\min_{x} \quad & z = c^T x \\
\text{s.t.} \quad & Ax = b \\
& x \in \mathbb{R}
\end{aligned}
$$

With the following matrices:

$$
c = \begin{bmatrix} 0 & 0 & 0.8509 & \cdots^{22} & 0.8509 & 0.5675 & \cdots^{22} & 0.5675 \end{bmatrix}
$$

$$
A = \begin{bmatrix}
1_1 & 0_1 & 1_1 & 0 & 0 & 1_1 & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & 0 & \ddots & 0 & 0 & \ddots & 0 & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1_{11} & 0_{11} & 0 & 0 & 1_{11} & 0 & 0 & 1_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 \\
0_1 & 1_1 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1_1 & 0 & 0 & 1_1 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & 0 & \ddots & 0 & 0 & \ddots & 0 \\
0_{11} & 1_{11} & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1_{11} & 0 & 0 & 1_{11}
\end{bmatrix}
$$

$$
x = \begin{bmatrix} g & h & j_1 & \cdots & j_{11} & k_1 & \cdots & k_{11} & m_1 & \cdots & m_{11} & n_1 & \cdots & n_{11} \end{bmatrix}
$$

$$
b^T = \begin{bmatrix} \cdots & x_i & \cdots & \cdots & y_i & \cdots \end{bmatrix}
$$

Solving this linear program with software we get the coordinates of $-120.510842°$ longitude and $46.997063500000003°$ degrees latitude. This puts our central hub station at approximately 2111 E Third Ave, Ellensburg, WA 98926.

Weighted Solution

The weighted solution is more complicated as it requires additional modifications to our original program. For this section the importance" of location was determined by the population of each respective city. Because of this the first thing that needed to be done was calculate the relative percentages of population for each of our secondary hub cities.

For reference, since the objective function is the sum of distances between each city and our central hub location, these weights will create coefficients for for central hub coordinates.

City	Population	Percent Sum Pop
Seattle	779,200	39.74%
Spokane	340,100	17.35%
Pullman	33,060	1.69%
Bellingham	95,960	4.89%
Ellensburg	20,900	1.07%
Vancouver	199,600	10.18%
Tacoma	222,400	11.34%
Kennewick	86,470	4.41%
Yakima	98,650	5.03%
Walla Walla	34,310	1.75%
Wenatchee	50,060	2.55%
Sum Pop	1,960,710	

Letting w_i be the vector of population weights for each city i our objective function now becomes

$$z = \sum_{i=1}^{11} 0.8509 w_i |x_i - g| + 0.5675 w_i |y_i - h| + 0.1697$$

With the absolute value substitutions done in the same way as previously our objective function reads:

$$z = \sum_{i=1}^{11} 0.8509 w_i j_i + 0.8509 w_i k_i + 0.5675 w_i m_i + 0.5675 w_i n_i + 0.1697$$

Though this changes our z value drastically it does alter the differences relative to each other which is the important part of these alterations. For example the population of Seattle is 39.74% of our population sum, and Wenatchee makes up 2.55% of our population sum. This results in the distance between Seattle and our central hub is a much smaller difference than the distance between Wenatchee and our hub.

$$\begin{aligned} distance_{\text{city}_i} < distance_{\text{city}_l} \quad \text{If } Pop_{\text{city}_i} < Pop_{\text{city}_l} \\ distance_{\text{city}_i} > distance_{\text{city}_l} \quad \text{If } Pop_{\text{city}_i} > Pop_{\text{city}_l} \end{aligned}$$

Since there was no change to the absolute value elements of our objective function so no constraints will require alterations nor will any need to be added. The equality constraint matrix will be identical to that of the unweighted solution, as will the rest aside from our coefficient vector c^t.

$$c = \begin{bmatrix} 0 & 0 & 0.8509w_i & 0.8509w_i & 0.5675w_i & 0.5675w_i \end{bmatrix}$$

Using this new coefficient vector after input matrices into software we get that the ideal location for this central hub is at $-122.330062°$ longitude and $47.6038321°$ latitude. This puts the hub along the Puget Sound coast in Seattle, at approximately 601 5th Ave, Seattle, WA 98104, United States.

Optimality Analysis

The main idea we are after is to find the best location. Without considering beyond mention the geographical circumstances of the proposed location, there is are two possible ways of seeing how ideal the location might be.

The first of these just being an average of the direct distance from each city to the proposed central hub location. The second is a weighted average based on population. This average is the average distance to the central hub per person. This would in theory provide a ratio of the number of people that would be using this train system by city.

Using the formula for distance between two coordinates on earth

$$\begin{aligned} d = \cos^{-1}\Bigg(&\cos\left(\frac{(90-h)\pi}{180}\right) * \cos\left(\frac{(90-y_i)\pi}{180}\right) + \sin\left(\frac{(90-h)\pi}{180}\right) \\ &* \sin\left(\frac{(90-y_i)\pi}{180}\right) * \cos\left(\frac{(g-x_i)\pi}{180}\right)\Bigg) * E_R \end{aligned}$$

We can calculate this distance using our provided coordinates of each city and the coordinates of the central hub that were solved for.

Unweighted Analysis

From the unweighted linear program we found that based on the populations of each city the ideal location for the central hub was at $-122.330062°$ longitude and $47.6038321°$ latitude. Logistically this exact location is not going to be suitable because this is a residential area. For the sake of argument and negligible distance to the nearest possible and realistic location, we will still be considering this location for optimality analysis. The direct average

Table 1: Unweighted: Average Distance per Person to Hub (mi)

City	Dist to (g,h)	Dist.$\times$ Pop
Seattle	94.93	73967717.9
Spokane	151.50	51524831.99
Pullman	158.60	5243308.568
Bellingham	151.61	14548312.95
Ellensburg	1.61	33740.20401
Vancouver	139.81	27905413.64
Tacoma	92.173	20498300.16
Kennewick	85.54	7396473.661
Yakima	27.31	2694040.533
Walla Walla	109.22	3747420.754
Wenatchee	30.91	1547223.7
Average	94.84	106.65

in Table 1 tells us that on average each secondary hub is 94.84 miles away from the proposed unweighted hub location. We also get that the average travel distance per person was approximately 106 miles to the central hub.

In general these averages make sense, because the closes cities to the proposed location are also some of the smallest cities so with less of a travel contribution in these small cities the ones that are further away and larger will begin to dominate this average.

Weighted Analysis

From the weighted linear program we found that based on the populations of each city the ideal location for the central hub was at $-122.330062°$ longitude and 47.6038321° latitude. As mentioned previously for the sake of argument, though this location is in a very commercial area, for analysis purposes we will still be considering this location. Table 2 Tells us that the average

Table 2: Weighted: Average Distance per Person to Hub (mi)

City	D to (g,h)	Dist.×Pop
Seattle	0.003	2209.666731
Spokane	228.31	77647113.6
Pullman	249.40	8245033.57
Bellingham	79.57	7635516.194
Ellensburg	93.49	1953958.136
Vancouver	137.21	27388040.47
Tacoma	25.25	5615883.576
Kennewick	179.45	15517374.63
Yakima	110.00	10851116.44
Walla Walla	204.10	7002828.698
Wenatchee	95.01	4756220.399
Average	127.44	84.98

distance between a given city and the weighted proposed hub location is approximately 127 miles. The Weighted average of travel distance on the other hand is only about 85 miles per person.

With the weighted solution we see that the greater of the two averages and now occurred in the average city distance to the hub rather than the average travel distance per person. Though there are quite a few cities within 100 miles of our proposed location, the cities that are not are closer to 250 miles away. This allows for a major skew in the average, though the populations of those cities is much smaller than the closer ones.

Conclusion

The average distances per person to our unweighted hub location is approximately 106 miles while our weighted hub location is approximately 85 miles.

In this situation, providing a weighted average for the travel distance per person may be more ideal for optimality consideration since a larger population generally implies more users of the train station coming from those cities. It is also worth considering that the weighted location does provide a much less convenient solution for many people beyond Ellensburg. Because of this, the best location may lie somewhere between our weighted and unweighted proposed locations.

References

[1] **WSDOT Region Boundaries**
https://www.arcgis.com/apps/mapviewer/index.html?layers=1c0845bd12944017b3bb3631bda09e0e

[2] **April 1, 2023 Population of Cities, Towns, and Counties: State of Washington**
https://ofm.wa.gov/washington-data-research/population-demographics/population-estimates/april-1-official-population-estimates

[3] **Earth's Radius by Latitude Calculator**
https://rechneronline.de/earth-radius/

[4] **GPS Coordinates**
https://gps-coordinates.org/

[5] **The Optimal Location of New Facilities Using Rectangular Distances**
https://pubsonline.informs.org/doi/pdf/10.1287/opre.19.1.124

Appendices

A Long and Lat Coordinates of Each City

Table 3: City Coordinates and Populations

City	Longitude (x)	Latitude (y)	Population	Notes
Seattle	-122.33	47.60	779,200	
Spokane	-117.42	47.66	340,100	includes Spokane Valley[1]
Pullman	-117.17	46.73	33,060	
Bellingham	-122.48	48.75	95,960	
Ellensburg	-120.55	47.00	20,900	
Vancouver	-122.67	45.63	199,600	
Tacoma	-122.44	47.25	222,400	
Kennewick	-119.12	46.21	86,470	
Yakima	-120.51	46.60	98,650	
Walla Walla	-118.52	46.21	34,310	
Wenatchee	-120.31	47.42	50,060	includes East Wenatchee[2]
			1,910,650	Population Sum

[1] Includes census data for both the city of Spokane and the suburb Spokane Valley

[2] Includes census data for both the city of Wenatchee and the city of East Wenatchee.

B Long and Lat Differences between cities

Table 4: Difference in Degrees Longitude (x)

	Sea	Spo	Pul	Bel	Ell	Van	Tac	Ken	Yak	Wal	Wen
Sea	-	-	-	-	-	-	-	-	-	-	-
Spo	4.907	-	-	-	-	-	-	-	-	-	-
Pul	5.156	0.250	-	-	-	-	-	-	-	-	-
Bel	0.149	5.055	5.305	-	-	-	-	-	-	-	-
Ell	1.785	3.122	3.371	1.934	-	-	-	-	-	-	-
Van	0.344	5.251	5.501	0.196	2.129	-	-	-	-	-	-
Tac	0.108	5.015	5.264	0.040	1.893	0.236	-	-	-	-	-
Ken	3.210	1.696	1.946	3.359	1.425	3.555	3.318	-	-	-	-
Yak	1.819	3.087	3.337	1.968	0.034	2.164	1.927	1.391	-	-	-
Wal	3.813	1.094	1.344	3.961	2.028	4.157	3.921	0.602	1.993	-	-
Wen	2.020	2.887	3.136	2.168	0.235	2.364	2.128	1.190	0.200	1.793	-

Table 5: Difference in Degrees Latitude (y)

	Sea	Spo	Pul	Bel	Ell	Van	Tac	Ken	Yak	Wal	Wen
Sea	-	-	-	-	-	-	-	-	-	-	-
Spo	0.053	-	-	-	-	-	-	-	-	-	-
Pull	0.873	0.927	-	-	-	-	-	-	-	-	-
Bell	1.148	1.095	2.021	-	-	-	-	-	-	-	-
Ell	0.607	0.660	0.267	1.755	-	-	-	-	-	-	-
Van	1.973	2.026	1.100	3.121	1.366	-	-	-	-	-	-
Tac	0.358	0.412	0.515	1.506	0.248	1.615	-	-	-	-	-
Ken	1.395	1.448	0.522	2.543	0.788	0.578	1.037	-	-	-	-
Yak	1.002	1.056	0.129	2.150	0.396	0.971	0.644	0.393	-	-	-
Wal	1.398	1.452	0.525	2.546	0.792	0.575	1.040	0.003	0.396	-	-
Wen	0.180	0.234	0.693	1.328	0.426	1.793	0.178	1.215	0.822	1.218	-