Autonomous Electric Pedestrian Ferries (Autoferry) - Efficient Connectors

FINAL REPORT

A Technical Report Submitted to the Rural Safe Efficient Advanced Transportation (R-SEAT) Center and United States Department of Transportation

February 2025

Principal Investigator:

Dr. Heather Dillon
University of Washington Tacoma
Pacific Northwest National Laboratory
E-Mail: hedillon@uw.edu

Co-Principal Investigators:

Dr. Angela Kitali

Dr. Jeff Walters

Dr. Panick Kalambay

DISCLAIMER

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Autonomous Electric Ferries (Autoferry) – Efficient Connectors		5. Report Date February 2025 6. Performing Organization Code 59-0977035
7. Author(s) Heather Dillon https://orcid.org/0000-0002-4467-2306 Jeffrey P. Walters https://orcid.org/0000-0001-6655-1233 Angela Kitali https://orcid.org/0000-0002-1962-162X Panick Kalambay https://orcid.org/0000-0002-0815-3074 Grant Lipscomb Gauge Frederick Weber https://orcid.org/0009-0007-4706-2099 Jose Sanchez-Sossa		8. Performing Organization Report No.
9. Performing Organization Name and Address University of Washington Tacoma		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 69A3552348321
12. Sponsoring Agency Name and Address Rural Safe Efficient Advanced Transportation Center 2525 Pottsdamer Street Tallahassee, FL 32310		13. Type of Report and Period Covered Final Report Period Covered: 03/01/2024 – 02/28/2025 14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This study examines the benefits and implementation of autonomous pedestrian ferries in Washington State, aiming to understand which routes and terminals should be prioritized for their deployment. Correlation, principal component, and cluster analyses were performed to identify key factors influencing transportation and classify ferry terminals into distinct clusters based on the socioeconomic and demographic characteristics of the census tracts where these terminals are located. Correlation analysis revealed that high-poverty census tracts are associated with higher proportions of disabled individuals, households without vehicles, and mobile homes. These tracts also show higher unemployment rates and lower education levels. These areas need better connectivity to job opportunities in the Seattle-Tacoma region. The principal component analysis identified five principal components that define clusters of terminals based on seventeen variables, with significant contributors including poverty, unemployment, education, and transit access. Cluster analysis highlighted five clusters with distinct characteristics. Proposed pedestrian autoferry routes using existing terminals could include Port Orchard (Cluster 1), Vashon/Tahlequah (Cluster 1), and Kingston (Cluster 3). These possible routes aim to address socioeconomic differences, improve access to essential services, and offer more frequent service due to shorter connections.

17. Key Word Autoferry, Autonomous ferry, Ferry terminal		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages	22. Price

ACKNOWLEDGEMENTS

This project was sponsored by the Rural Safe, Efficient, Advanced Transportation (R-SEAT) Center and the United States Department of Transportation. The Principal Investigators would like to thank the representatives of the R-SEAT Center for their valuable feedback throughout the project activities. The research team is also grateful to the WAV-C cluster for supporting our research, to the Numurus company for a valuable project partnership, and to the many undergraduate research students who have assisted in data collection and analysis.

EXECUTIVE SUMMARY

This study examines the benefits and implementation of autonomous pedestrian ferries in Washington State, aiming to understand which routes and terminals should be prioritized for their deployment. Correlation, principal component, and cluster analyses were performed to identify key factors influencing transportation and classify ferry terminals into distinct clusters based on the socioeconomic and demographic characteristics of the census tracts where these terminals are located. Correlation analysis revealed that high-poverty census tracts are associated with higher proportions of disabled individuals, households without vehicles, and mobile homes. These tracts also show higher unemployment rates and lower education levels. These areas need better connectivity to job opportunities in the Seattle-Tacoma region. The principal component analysis identified five principal components that define clusters of terminals based on seventeen variables, with significant contributors including poverty, unemployment, education, and transit access. Cluster analysis highlighted five clusters with distinct characteristics. Proposed pedestrian autoferry routes using existing terminals could include Port Orchard (Cluster 1), Vashon/Tahlequah (Cluster 1), and Kingston (Cluster 3). These possible routes aim to address socioeconomic differences, improve access to essential services, and offer more frequent service due to shorter connections.

Table of Contents

1.0 Introduction

Washington is currently home to the largest and most widely used ferry system in the United States [1]. In 2023, 18,661,400 Washington residents relied on these services to commute to school, work, and travel across the Puget Sound [2]. Despite providing staple transit services for years, Washington State Ferries have experienced a considerable decline in funding towards maintenance, staffing, and operation [3]. Such issues have proven to be an inconvenience for many Washington residents dependent on ferry services. The wear and age of Washington State ferries currently in service also pose several uncertainties for the future of accessible public transit across the Puget Sound [4].

Such concerns necessitate measures to meet commuter demands and alleviate the stresses on the present-day system. Our proposed solution involves the creation and operation of autonomous electric pedestrian ferries to safely and reliably provide transportation to passengers around the Puget Sound. While the proposed vessels will be designed to accommodate passengers and bicycles, introducing them alongside existing ferry services will enable greater accessibility and flexibility. Individuals seeking to travel without a personal vehicle may find this to be an economically favorable mode of transportation. The implementation of autonomous pedestrian ferry operations in the Puget Sound may result in greater commuter volume and retention, all while offloading the demands on the high-volume transportation system in service.

Autonomous electric ferries offer a unique low-carbon option to better connect rural communities in the region. For example, the residents of Anderson Island and Ketron Island in the south Puget Sound region are currently served by one ferry that connects them to the mainland. For Ketron Island, the ferry runs only four times per day and was out of service entirely for several days recently while the ferry was repaired. On the nearby peninsula, the community of Longbranch is accessible by road, but travel requires more than an hour by car for residents to reach services. Several tribal communities are included in these isolated areas.

To address the local environmental agenda, the use of sustainably sourced electricity will reduce dependence on carbon based fuels, thereby reducing emissions and providing a model for greener public transportation services. Washington state government policy has already established a commitment to a zero-carbon emission energy infrastructure by 2045 through Senate Bill 5116, known as the Clean Energy Transformation Act (CETA) [5]. This is further supported by plans for the electrification of Washington State Ferries, a prospect which will result in the construction and operation of hybrid electric ferries by 2040 [6]. The creation of infrastructure for autonomous passenger ferries provide a supplementary and serviceable addition towards this common goal.

A focus on fair, sustainable transportation and the future of autonomous ferry operations in the Puget Sound lies at the heart of our research. The objective of our research is to analyze the potential impact of possible autoferry routes and terminals to maximize accessible transportation across the Puget Sound, including rural areas. This research is focused on understanding the future of autonomous ferry operation in Washington state with the following research questions:

1. What benefits might autonomous ferries offer to provide fair transportation in Washington state?
2. What routes and terminals should be considered for fair autonomous ferries in Washington state?

2.0 Background

Autonomous marine traffic projects have been pilot tested in other regions of the world. Leslie *(5)* provides a succinct summary of the types of crafts and the advantages of autonomous marine vehicles. To understand the context of the autonomous ferry work, a literature review was performed and summarized in **Table 1**.

A few of the prior studies focused on autonomous ferries in specific locations including Norway and Sweden with a focus on cost and stakeholder perceptions (*7*, *8*). Many studies have focused on the design (*9*) and automation of ferries in other locations (*7*, *10*). Safety concerns and collision modeling have also been reviewed (*11*, *12*). One challenging aspect of the autoferry operation is docking. Several studies have focused on this as a unique concern for autonomous marine vessels (*10*, *13*, *14*). In the Salish Sea this challenge is more difficult due to strong currents, wind gusts, and large tidal swings.

Energy and environmental considerations are very important for autonomous ferries. Viola et al. (*15*) considered the modeling of wind assisted vessels and Minak (*16*) reviewed solar options. Wang et al. (*17*) calculated the life cycle cost and environmental impact of hybrid ferry designs. While these studies were not focused on autonomous ferries, the insights are applicable for all vessels.

TABLE 1 Summary of Prior Work on Autonomous Ferry Operation and Traffic Mapping

Author(s)	**Year**	**Location**	**Analysis Type**	**Summary**
Merrick et al. (*18*)	2003	California	Vessel interaction	Boat congestion analysis
Zhang (*19*)	2016	California	Traffic analysis	Considers ferries as an alternative traffic flow method
Reddy et al. (*8*)	2019	Norway	Cost	Overview of autonomous ferry in the context of transportation networks
Pham et al. (*20*)	2020	Korea	Route Efficiency	Optimization of ferry routes, longer routes are not always better
Guo et al. (*11*)	2021	Norway	Collision Model	Developed a model for risk of collision of autoferries
Puzdrakiewicz and Tarkowski (*21*)	2021	Poland	Traffic Analysis	Assesses the benefits of autonomous ferries on urban mobility and connectivity
Fan et al. (*12*)	2021	China	Collision Model	Developed a model for risk of collision of autonomous vessels
Cheemakurthy and Garme (*22*)	2022	Sweden	Stakeholders	Performance specifications for ferries
Munim et al. (*7*)	2022	Norway	Stakeholders	Public perception of autonomous ferry safety
Ringsberg (*23*)	2023	Sweden	Stakeholders	Survey analysis of rural ferry opportunities
Present work	**2024**	**Washington**	**Route Analysis**	**Statistical analysis of possible autoferry routes in the Salish Sea**

While ferry operations have been studied for many years in the literature, few of the prior works focused on rural route issues. Recently Ringsberg (*23*) surveyed ferry passengers to better understand how ferry operations could be improved in rural areas in Sweden and found that excess ferry capacity may be utilized by cargo or freight to reduce costs for transport of goods to rural areas. We build on this work by analyzing the opportunities for future autonomous ferry operations in Washington State using an economic lens for rural regions.

3.0 Methods

This section outlines the data collection and processing, and the methodological approach employed for analysis. The data processing, mapping, and analysis were performed using the R software [7].

Data Collection and Processing

The data used in this study are derived from two main sources. The locations (geographic coordinates) of ferry terminals and transit stops in Washington were obtained from the Washington Department of Transportation (WSDOT) along with the ferry routes [8]. There are 37 ferry terminals operated by the Washington State Ferries (WSF), county, provincial, tribal, and private entities. The ferry network is concentrated in the Salish Sea. Only ferry terminals listed by WSDOT are considered in this study for phase 1 of the research. **Figure 1** shows the locations of WSDOT ferry terminals and the routes of vessels providing scheduled public car ferry service in the waters of Washington State as linear features. Additionally, the map includes known private, provincial, tribal, and passenger-only ferry services, highlighting the comprehensive nature of the ferry network in the region. This extensive network in the Salish Sea plays a crucial role in regional connectivity, supporting both local commuting and tourism.

The proposed autonomous ferry operations will utilize much smaller boats than existing ferry operations designed for passengers and cars. To consider all possible future ferry routes, the research team developed a list of smaller public and private marinas and docks in the region that are not included in this study but will be the focus of future work. While additional docks beyond existing private/public could also be considered they would require a long time to build and may create unintended environmental impacts.

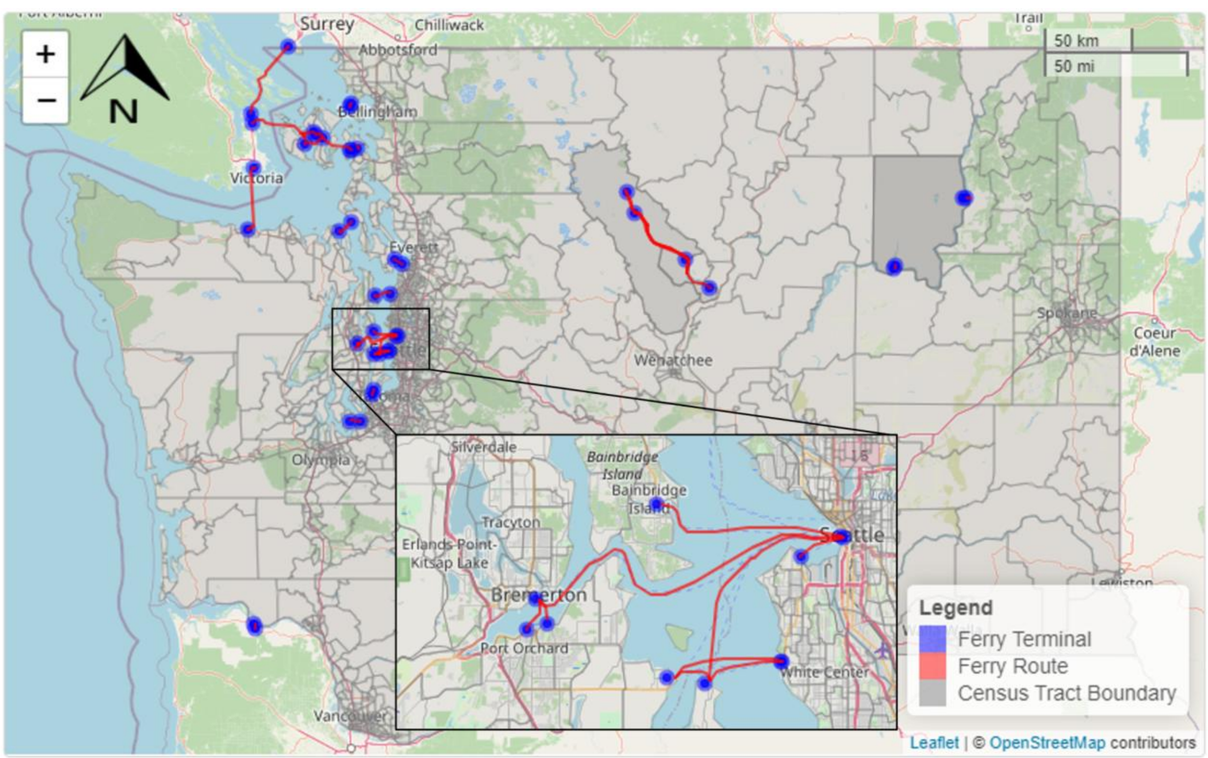

Figure 1 Ferry terminals and routes across census tracts

Note: Darker census tracts are those with ferry terminals.

The study also examined the characteristics of the areas where the ferry terminals are located. These characteristics, related to economic variables, were sourced from the Social Vulnerability Index (SVI) database [9]. Using data from the 2022 American Community Survey (ACS), the SVI database provides 2018-2022 estimates covering various aspects such as socioeconomic status, household characteristics, racial status, housing types, and transportation characteristics. Therefore, demographic variables of the thirty-four census tracts where the thirty-seven ferry terminals in Washington are located were collected.

Census tracts with ferry terminals exhibit a wide range of characteristics, both in terms of area and population. These tracts span from as small as 0.13 square miles to as large as 1,177.62 square miles, indicating dense urban areas to vast rural or coastal regions. The population in these tracts varies considerably, ranging from a minimum of 965 residents to a maximum of 6,497, with a median household count of 1,427. This variety suggests that ferry terminals serve a variety of community types, from potentially busy commuter hubs in more populated areas to vital links for remote or island communities in larger, less densely populated tracts. The substantial variations in both area and population density imply that ferry systems likely play different roles across these communities, adapting to local needs and geographic constraints. **Table 2** presents a descriptive summary of various demographic variables (17) across census tracts with ferry terminals.

TABLE 2 Descriptive Summary of Variables of Census Tracts with Ferry Terminals

Variable	Description	Min	Mean	SD	Median	Max
POV150	Percentage of persons below 150% poverty estimate	3.8	20.19	11.51	18.1	50.1
UNEMP	Unemployment Rate estimate	0	4.71	3.34	3.75	11.2
NOHSDP	Percentage of persons with no high school diploma (age 25+) estimate	0.3	6.4	4.91	5.15	21.2
AGE65	Percentage of persons aged 65 and older estimate	0	28.11	10	27.9	48.4
AGE17	Percentage of persons aged 17 and younger estimate	1.5	15.1	6.2	15.9	25.2
DISABL	Percentage of civilian noninstitutionalized population with a disability estimate	2.4	17.03	6.75	16.2	30.9
MOBILE	Percentage of mobile homes estimate	0	8.39	8.31	5.5	24.6
CROWD	Percentage of occupied housing units with more people than rooms estimate	0	3.68	4.29	2.8	22
NOVEH	Percentage of households with no vehicle available estimate	0	10.79	18.35	3.6	72.4
STOP	Number of transit stops per square mile	0	29.27	100.22	1.23	304.44
HISP	Percentage of Hispanic or Latino persons estimate	0.4	8.04	7.41	5.35	28.7
AFAM	Percentage of Black/African American, not Hispanic or Latino persons estimate	0	1.37	2.73	0.45	10.8
ASIAN	Percentage of Asian, not Hispanic or Latino persons estimate	0	3.81	6.68	2.3	31
AIAN	Percentage of American Indian or Alaska Native, not Hispanic or Latino persons estimate	0	6.79	18.43	0.3	59
NHPI	Percentage of Native Hawaiian or Other Pacific Islander, not Hispanic or Latino persons estimate	0	0.31	0.5	0.05	2.2
TWOMORE	Percentage of two or more races, not Hispanic or Latino persons estimate	0	5.41	2.94	5.15	10.9
OTHERRACE	Percentage of some other race, not Hispanic or Latino persons estimate	0	0.42	0.54	0.1	1.7

Table 2 includes key statistical measures for each variable: minimum (Min), mean (Mean), standard deviation (SD), median (Median), and maximum (Max). The variables cover a wide range of indicators

pertaining to poverty, unemployment, education, age, disability, housing, vehicle access, transit, and race/ethnicity in these census tracts. For instance, the percentage of persons below 150% of the poverty estimate (POV150) has a mean of 20.19% and ranges from 3.80% to 50.10%. Unemployment rates (UNEMP) average 4.71%, with a maximum of 11.20%. The percentage of individuals without a high school diploma (NOHSDP) averages 6.40%. Age-related metrics show that the percentage of persons aged 65 and older (AGE65) averages 28.11%, while those aged 17 and younger (AGE17) average 15.10%. Disability rates (DISABL) have a mean of 17.03%. Mobile homes (MOBILE) constitute an average of 8.39% of housing units. Overcrowded housing (CROWD) averages 3.68%. The percentage of households without a vehicle (NOVEH) averages 10.79%, with a high of 72.40%. The number of transit stops per square mile (STOP) is highly variable, with a mean of 29.27 and a maximum of 304.44. Racial and ethnic variety is represented with percentages for Hispanic or Latino persons (HISP) averaging 8.04%, Black/African American persons (AFAM) averaging 1.37%, Asian persons (ASIAN) averaging 3.81%, and American Indian or Alaska Native persons (AIAN) averaging 6.79%. Other racial categories include Native Hawaiian or Other Pacific Islander (NHPI) with a mean of 0.31%, two or more races (TWOMORE) at 5.41%, and some other race (OTHERRACE) at 0.42%. These statistics provide insight into the demographic and socioeconomic landscape of census tracts hosting ferry terminals, highlighting areas of potential inequality and variety.

By incorporating these factors, the analysis aims to understand the broader social and economic context of ferry terminal locations, emphasizing potential demographic issues that can inform strategies for more transportation planning. This comprehensive approach ensures that the study not only maps the physical infrastructure but also considers the social dimensions of accessibility in the ferry network, especially when considering which communities autonomous ferries also need to serve.

Methodological Approach

This study analyzes the demographic needs of census tracts and tribal communities with ferry terminals, adopting a multifaceted methodological approach. It combines correlation analysis and clustering/spatial analysis to assess the relationships among demographic variables, identify clusters of similar characteristics, and identify the most important factors/variables contributing to high needs that could potentially lead to transportation interventions, such as the establishment of new or enhanced autoferry routes.

Due to the relatively low sample size of the dataset, Kendall's tau was used for the correlation analysis. It is more robust and reliable for smaller datasets compared to other correlation measures like Pearson or Spearman. The Kendall method evaluates the strength and direction of association between pairs of demographic variables. The following section describes the clustering analysis as it constitutes the core of the methodology employed in this study.

Cluster analysis was performed to categorize ferry terminals based on the demographic characteristics of the census tracts where they are located. By grouping similar ferry terminals, we aim to identify distinct clusters that share common needs. This step is crucial for understanding the heterogeneity among the ferry terminals and census tracts. Ultimately, optimal ferry routes should connect terminals in census tracts with contrasting socioeconomic conditions and transportation needs. For instance, linking a terminal in a high-unemployment area to one in a low-unemployment area could provide access to job opportunities to people from these communities. The clustering algorithm employed in this study involved two steps. Initially, Principal Component Analysis (PCA) was conducted on the demographic variables to reduce their dimensionality and capture the most significant variance [10]. Thereafter, K-means was applied to the transformed data to identify distinct groups/clusters [11].

PCA summarizes information from the demographic variables by reducing dimensionality. It identifies the principal directions of data variance and expresses this as a few principal components (PCs), which are linear combinations of the original variables [10]. PCA was performed using the R package *factoextra*. Before applying PCA, the model variables were normalized, transforming them into variables with a mean of 0 and a standard deviation of 1. This standardization allows identifying PCs that account for more variance than any single original variable. An eigenvalue greater than 1 is commonly used as a cutoff point for retaining PCs [12].

After PCA-based dimension reduction, K-means clustering was performed using the package *cluster* in R. K-means, a popular unsupervised machine learning algorithm [11], classifies data points into clusters by minimizing within-cluster variation. The algorithm begins with k randomly selected centroids and iteratively optimizes their positions, defining each cluster by its centroid (the mean of its data points). This study employed the Hartigan and Wong algorithm. The total within-cluster variation is defined as the sum of squared Euclidean distances between each data point (x_i) and its corresponding cluster centroid (μ_k), as shown in **Equation 1** [13].

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \qquad (1)$$

The K-means algorithm requires specifying the number of clusters (k), which we determined using the elbow method. This approach balances the number of clusters against the total within-clusters sum of squares (SST). We ran K-means for various k values, and calculated SST for each k using **Equation 2**. The optimal k identified at the elbow is the value where the decrease in SST begins to level off [14].

$$SST = \sum_{k=1}^{K} W(C_k) = \sum_{k=1}^{K} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \qquad (2)$$

4.0 Results

Correlation Analysis

The census data for the region with ferry terminals was analyzed using a correlation matrix, as shown in **Figure 2**. This matrix reveals significant patterns that help identify high-priority and low-priority areas and inform the planning of autoferry routes from a demographic standpoint.

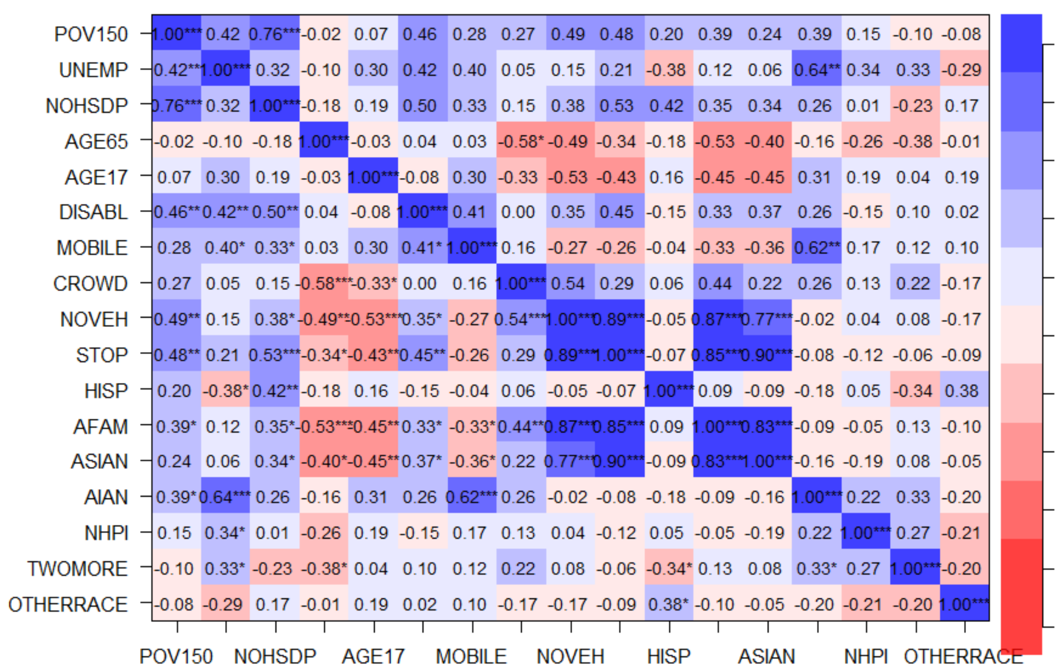

Figure 2 Correlation analysis results

*Note: *, **, *** indicate p-value <0.05, <0.01, and <0.001, respectively.*

The results indicate that census tracts with higher poverty rates (POV150) are positively correlated with higher proportions of American Indian or Alaska Native (AIAN), Black/African American (AFAM), disabled individuals (DISABL), households without vehicles (NOVEH), and mobile homes (MOBILE). They also show higher unemployment rates (UNEMP) and lower education levels (NOHSDP).

Furthermore, areas with lower educational attainment are correlated with higher proportions of Asian (ASIAN) and Hispanic (HISP) populations and mobile homes, and denser transit stops. This reflects a critical need for targeted educational and transportation support in communities with lower educational levels. Census tracts with higher proportions of elderly (AGE65) and younger individuals (AGE17) are negatively correlated with higher percentages of Asians, Blacks/African Americans, and density of transit stops (STOP). Elderly and younger populations in these tracts may face challenges due to fewer transit stops and potentially inadequate housing conditions. In fact, areas with more crowded housing conditions (CROWD) are positively correlated with higher proportions of Blacks/African Americans and households without vehicles. This highlights the need for improved housing and transportation infrastructure in densely populated and disadvantaged areas. These trends are consistent with other regions and broader trends in transportation demographics [15].

Cluster Analysis

PCA Results

PCA was employed to reduce the dimensionality of demographic variables and identify those contributing most significantly to the principal components (PCs). The K-means algorithm was then applied to these PCs. All variables were scaled during PCA, addressing two issues: (1) the different measurement scale of some variables (e.g., density for transit stops vs. proportions for others), and (2) the large differences in means and/or standard deviations among some variables (STOP vs. others).

Table 3 presents the PCA results for PCs, including their eigenvalues and proportions of variation explained (both variance and cumulative variance). Eigenvalues measure the variation retained by each PC, with larger values for initial PCs and smaller values for subsequent ones. This indicates that the first PCs correspond to directions with maximum data variation. For instance, the first eigenvalue explains 29.70% of data variation, while the first two eigenvalues combined account for 48.88%.

TABLE 3 PCA Results

PC	Eigenvalue	Variance (%)	Cumulative variance (%)
PC1	5.05	29.7	29.7
PC2	3.26	19.18	48.88
PC3	2.28	13.43	62.31
PC4	1.75	10.32	72.63
PC5	1.09	6.44	79.07
PC6	1	6.08	85.14
PC7	0.63	3.7	88.84
PC8	0.48	2.82	91.66
PC9	0.4	2.36	94.02
PC10	0.29	1.68	95.71
PC11	0.23	1.36	97.06
PC12	0.19	1.1	98.17
PC13	0.12	0.7	98.87
PC14	0.1	0.6	99.47
PC15	0.05	0.31	99.78
PC16	0.03	0.17	99.95
PC17	0.01	0.05	100

Determining the optimal number of PCs lacks a universally accepted objective method. However, eigenvalues greater than 1 are commonly used as a cutoff [12], indicating PCs that account for more variance than one original standardized variable. Based on the results in **Table 3**, PC1 to PC5 should be retained for cluster analysis, reducing the dimensions from 17 demographic variables to 5 PCs. These first five PCs explain 79.07% of the data variance.

Given that the results indicate retaining the first five PCs for cluster analysis, it is crucial to understand how each original variable contributes to these PCs. **Table 4** presents the percentage contributions of all seventeen demographic variables to each PC. This information is vital for several reasons. It reveals which variables are most influential in forming each PC, highlighting NOVEH, STOP, AFAM, ASIAN, and POV150 as the most significant contributors to PC1, for instance. The high contributions of UNEMP and MOBILE to PC2 and NOHSDP and HISP to PC3 further underscore their importance in differentiating between clusters. AGE65, DISABL, and CROWD significantly shape PC4, while POV150, NHPI, and TWOMORE dominate PC5.

TABLE 4 Contributions of Variables to PCs (%)

Variable	PC1	PC2	PC3	PC4	PC5
POV150	6.88	6.81	7.08	0.1	9.97
UNEMP	1.82	17.4	2.08	3.62	1.11
NOHSDP	6.24	4.82	18.5	0.31	0.52
AGE65	5.21	0.05	3.29	25.9	6.69
AGE17	3.31	9.49	2.54	2.71	0.09
DISABL	5.09	4.36	2.46	14	8.25
MOBILE	0.18	19.6	1.02	0.01	6.68
CROWD	5.48	0.23	3.45	12.6	0.4
NOVEH	17.4	0.98	0.37	0.07	0.8
STOP	17.1	1.12	0.47	1.86	0.17
HISP	0	0.38	18.4	22.7	1.09
AFAM	16.2	1.89	0.11	0.43	0.1
ASIAN	13.9	3.44	0.02	1.24	1.83
AIAN	0.31	21.4	1.24	0.02	0.53
NHPI	0	4.92	4.09	10.7	19.7
TWOMORE	0.33	2.76	20.3	0.9	12.8
OTHERACE	0.49	0.43	14.6	3	29.3

Note: Variable definitions and units are provided in **Table 1**.

Overall, identifying the variables that play the most significant role in defining PCs and differentiating between clusters enhances our understanding of the underlying factors contributing to issues in areas with ferry terminals. In practice, we tend to look at the first few PCs to find the most contributing factors. Correlations in **Figure 2** can be used to find additional contributing factors. Therefore, **Figure 3** displays the biplots of all the demographic variables with their importance in the first three PCs. Each arrow represents one of the seventeen variables. The direction and length of the arrow indicate how much and in which direction the variable contributes to PCs. PCs are scaled to unit length. This scaling helps compare the contributions of variables to different PCs. Variables with scores close to +1 or -1 on a PC are highly correlated with that component. Variables with scores near 0 on a PC do not contribute much to that component.

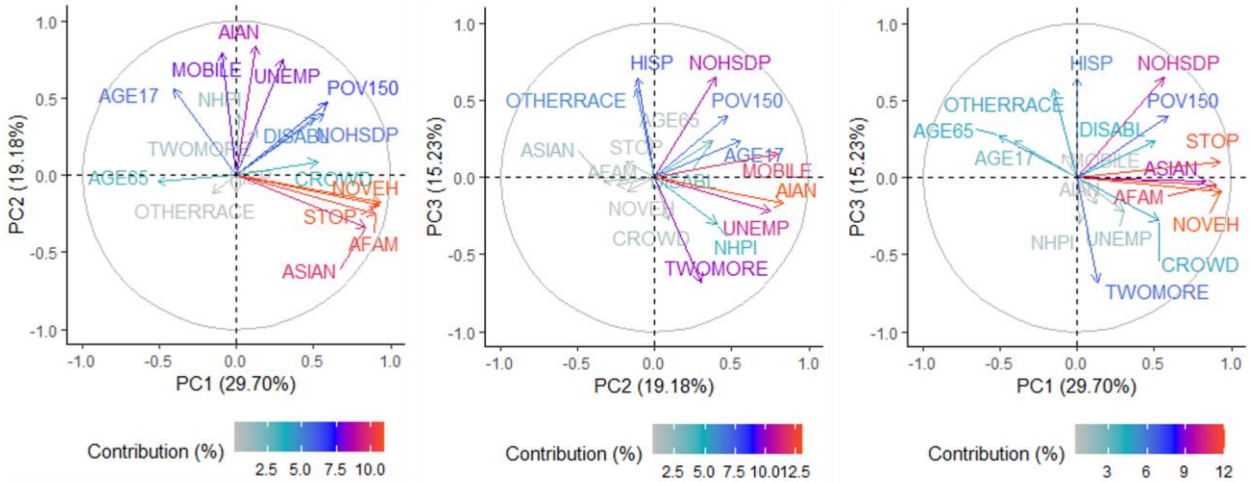

Figure 3 Contributions of variables to PC1, PC2, and PC3

From **Figure 3**, it can be concluded that the most important variables in determining PCs and, consequently, in forming the clusters are AFAM, ASIAN, AIAN, TWOMORE, HISP, MOBILE, POV150, UNEMP, NOHSDP, NOVEH, and STOP. Variables such as POV150, UNEMP, NOHSDP, and NOVEH highlight socioeconomic and educational differences. Meanwhile, variables like AFAM, ASIAN, AIAN, HISP, and TWOMORE capture the racial and ethnic variety within the census tracts. STOP and MOBILE emphasize transportation accessibility and housing characteristics. Also, the results show that AFAM, ASIAN, NOVEH, and STOP are close to each other, emphasizing their positive high pairwise correlations revealed in **Figure 2**. These variables collectively illustrate the multifaceted nature of issues within the census tracts with ferry terminals, providing insights into the various factors that influence cluster formation and the differentiation of these tracts.

Clustering and Mapping Results

As mentioned above, cluster analysis was conducted using the K-means algorithm with k=1 to k=10. The optimal cluster was determined using the elbow method. The clustering is based on the five PCs identified in PCA describing 79.07% of the demographic level variability in census tracts with ferry terminals. **Figure 4** shows how SST decreases with increasing k before leveling off at k=5 (elbow). The WSS sharply decreases from k = 1 to k = 5, and then the rate of decrease slows down significantly. This means 5 is the optimal number of clusters to consider.

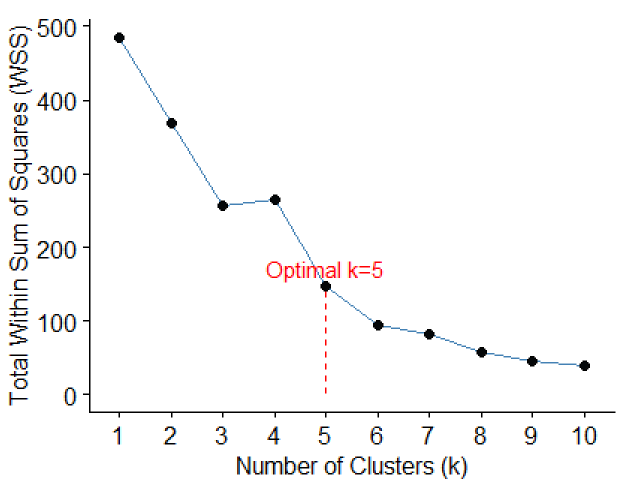

Figure 4 Optimal number of clusters (Elbow method)

The cluster analysis results are presented for all seventeen demographic variables, but the discussion focuses primarily on the variables that define the five PCs explaining the majority of data variability. **Table 5** presents a summary of the cluster analysis results, showing the mean and median values of all seventeen variables for each identified cluster. This tabulation allows for a direct comparison of how these indicators vary across the different clusters. **Figure 4** provides a visual representation of the spatial distribution of clusters of ferry terminals. This map illustrates the geographical location of each ferry terminal, the cluster to which it belongs, the ferry routes connecting these terminals, and the type of areas where they are situated. The classification of census tracts in this study follows the National Center for Education Statistics (NCES) methodology, which categorizes areas as city, town, suburban, or rural based on population and distance to urbanized areas [16]. To determine the area type of each census tract, we spatially overlaid this NCES classification dataset with the census boundary data to understand the urban-rural context of each ferry terminal location, allowing for more nuanced comparisons and analyses of how socioeconomic and transportation factors might vary across different settlement types within the ferry network. About 62% of the thirty-seven ferry terminals are in census tracts identified as rural.

TABLE 5 Mean (Median) of Demographic Variables by Cluster

Variable	Cluster				
	Cluster 1 (N = 5)	Cluster 2 (N = 2)	Cluster 3 (N = 1)	Cluster 4 (N = 5)	Cluster 5 (N = 24)
POV150	26 (32)	28 (28)	10 (10)	22 (26)	15 (14)
UNEMP	8.0 (10.4)	4.9 (4.9)	4.0 (4.0)	2.5 (2.6)	4.0 (3.8)
NOHSDP	7.6 (9.4)	11.8 (11.8)	3.3 (3.3)	11.5 (10.0)	3.6 (2.9)
AGE65	22 (25)	18 (18)	29 (29)	29 (33)	29 (28)
AGE17	21.0 (21.4)	9.7 (9.7)	13.0 (13.0)	18.1 (17.2)	14.9 (16.2)
DISABL	18.9 (20.1)	19.9 (19.9)	13.2 (13.2)	18.8 (20.5)	14.1 (14.0)
MOBILE	15 (19)	0 (0)	2 (2)	13 (13)	5 (4)
CROWD	5.06 (4.00)	4.45 (4.45)	2.70 (2.70)	2.04 (2.00)	3.13 (1.85)
NOVEH	7 (9)	41 (41)	2 (2)	1 (1)	7 (4)
STOP	22 (1)	215 (215)	4 (4)	0 (0)	71 (21)
HISP	5.4 (6.0)	4.8 (4.8)	5.1 (5.1)	19.6 (16.5)	6.7 (5.4)
AFAM	0.82 (0.40)	5.40 (5.40)	0.30 (0.30)	0.40 (0.50)	1.21 (0.75)
ASIAN	1.6 (1.5)	19.3 (19.3)	3.9 (3.9)	1.6 (2.5)	3.9 (2.3)
AIAN	35 (57)	0 (0)	1 (1)	1 (0)	0 (0)
NHPI	0.68 (0.80)	0.00 (0.00)	0.00 (0.00)	0.20 (0.20)	0.28 (0.00)
TWOMORE	8.24 (10.10)	4.20 (4.20)	10.90 (10.90)	2.82 (4.40)	5.64 (5.95)
OTHERRACE	0.16 (0.10)	0.60 (0.60)	0.00 (0.00)	1.08 (1.20)	0.33 (0.00)

Note: Variable definitions and units are provided in **Table 1**.

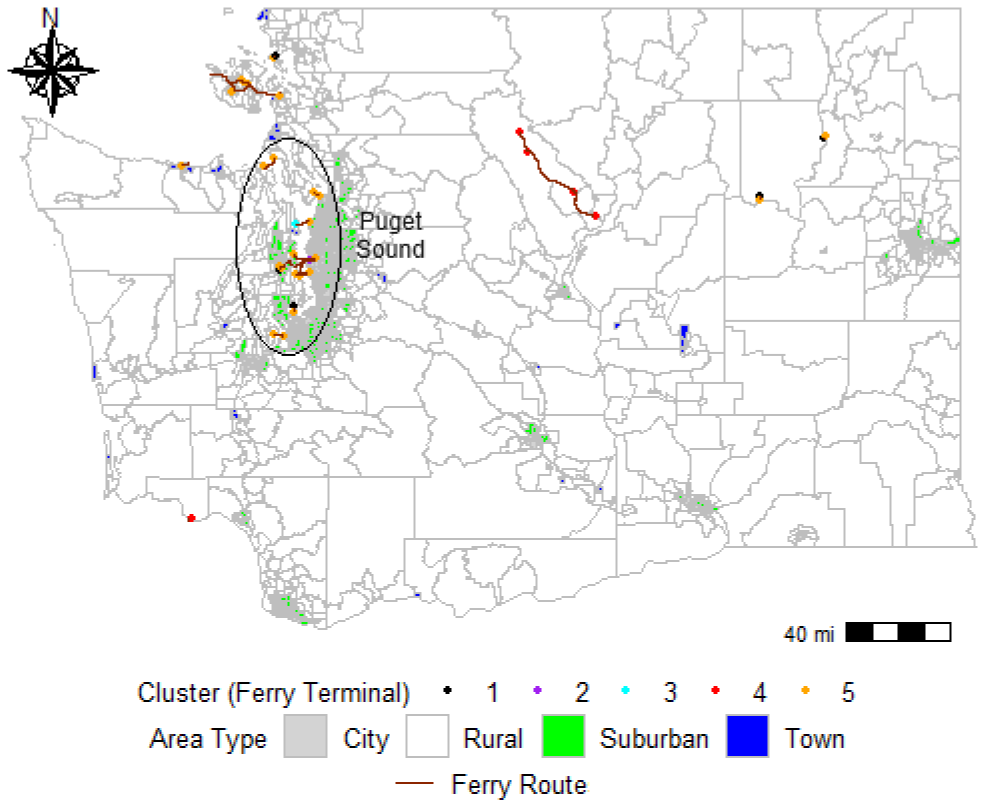

Figure 4 Map of the clusters of the ferry terminals

Cluster 1 exhibits high poverty rates and unemployment levels compared to the other clusters. The mean poverty rate in this cluster is 26%, with a median of 32%, indicating a significant portion of the population living below 150% of the poverty threshold. Similarly, the unemployment rate is relatively high, with a mean of 8.0% and a median of 10.4%. The high percentage of individuals with health challenges (mean = 18.9%) and a notable presence of mobile homes (mean = 15%) further accentuate the socioeconomic challenges faced by people living in the census tracts with ferry terminals belonging to this cluster. This cluster also has a high percentage of American Indian or Alaska Native individuals (mean = 35%). In contrast, it shows a low percentage of Black/African American and Asian individuals. The sparse number of transit stops (mean = 22) combined with a low percentage of households without vehicles available (mean = 7%) suggests limited access to public transportation and vehicular resources. Port Orchard, Tahlequah, Inchelium, Keller N, and Whatcom belong to this cluster.

Cluster 2 stands out due to its extremely high density of transit stops (mean = 215), which is significantly higher than in any other cluster. This high transit availability is coupled with a very high percentage of households without vehicles available (mean = 41%), indicating a strong reliance on public transportation. The high poverty rate (mean = 28%) and a notable percentage of persons without a high school diploma (mean = 11.8%) emphasize the socioeconomic challenges faced by this cluster. The absence of mobile homes and a relatively high percentage of Asian individuals (mean = 19.3%) are distinctive features of this cluster. The low percentages of American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander individuals further differentiate this cluster from others. Terminals at Seattle Pier 50 and West Seattle belong to this cluster.

Table 5 revealed that *Cluster 3* is characterized by relatively low poverty rates (mean = 10%) and low unemployment rates (mean = 4.0%), indicating better economic stability compared to other clusters. The high percentage of elderly individuals (mean = 29%) suggests a significant aging population. This cluster also exhibits very low percentages of mobile homes (mean = 2%) and transit stops (mean = 4). The low percentages of American Indian or Alaska Native individuals (mean = 1%), and the minimal presence of other racial groups, set this cluster apart from others. Only the Kingston ferry terminal made up this cluster.

Cluster 4 features a high percentage of elderly residents (mean = 29%), similar to Cluster 3. However, it is marked by a high percentage of Hispanic or Latino individuals (mean = 19.6%) and a significant presence of mobile homes (mean = 13%). Despite a relatively high percentage of individuals with no high school diploma (mean = 11.5%), the unemployment rate is notably low (mean = 2.5%). The absence of transit stops (mean = 0) is a key characteristic, which, combined with a very low percentage of households without vehicles (mean = 1%), suggests that public transportation is either very limited or not a primary concern in this cluster. They are all located in rural areas, which in fact are not well served in terms of public transportation due to the lack of economic viability for transportation companies. The Puget Island ferry, Stehekin, Lucerne, Lakeshore, and Chelan belong to this cluster, although the inland terminals are not the focus of the autonomous ferry efforts.

Table 5 also indicates that *Cluster 5* shows generally low poverty rates (mean = 15%) and unemployment rates (mean = 4.0%). The percentage of individuals with no high school diploma is also low (mean = 3.6%), indicating higher educational attainment. The cluster has a moderate number of transit stops (mean = 71), suggesting reasonable access to public transportation. It features a low percentage of Black/African American (mean = 1.21%) and Asian individuals (mean = 3.9%). The percentage of persons of two or more races is moderate (mean = 5.64%), and there is a relatively low presence of mobile homes (mean = 5%). This cluster's profile suggests relatively stable socioeconomic conditions with moderate variety and access to transportation. Most of the ferry terminals (65%) belong to this cluster.

In summary, the clusters reveal demographic profiles, highlighting significant differences across different regions. *Cluster 1* and *Cluster 2* face higher socioeconomic challenges, while *Cluster 3* stands out for its potential economic stability. *Cluster 4* features a high elderly population and substantial Hispanic presence, whereas *Cluster 5* reflects relatively stable conditions with moderate variety. These insights can inform targeted interventions and policies to address the specific needs of census tracts with ferry terminals and improve access to transportation services.

5.0 Conclusions

This section discusses the possible routes for ferry terminals, particularly in the Puget Sound using only existing ferry terminals. The Puget Sound region, centered around the Seattle-Tacoma metropolitan area, is a hub for ferry transportation with numerous terminals serving approximately 23.8 million passengers annually (3). This extensive marine network plays a crucial role in connecting urban centers, suburban areas, and island communities while supporting the region's thriving job market and economy. As environmental concerns grow, there is increasing interest in transitioning these high-volume routes to low-carbon electric ferries, which could significantly reduce the carbon footprint of marine transportation in the area. Figure 5 provides a detailed visualization of the existing spatial distribution of ferry terminals, categorized by their respective clusters, Additionally, it depicts the main (most frequent) ferry routes currently operating in the region.

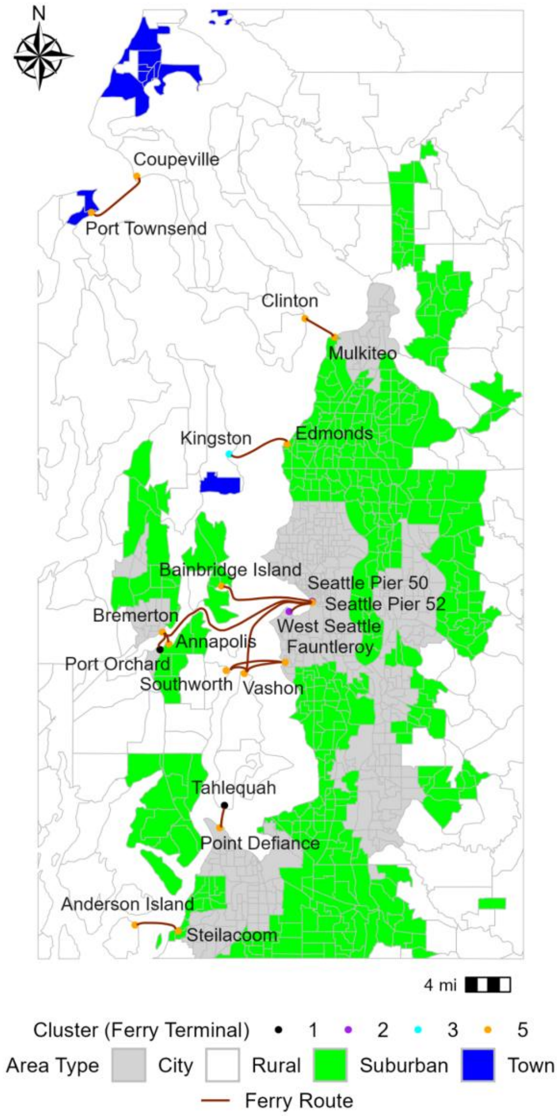

Figure 5 Map of ferry terminals in the Puget Sound by cluster

The Puget Sound ferry network not only supports the region's thriving economy but also serves as a vital link to job opportunities for residents of disconnected islands. From a demographic perspective, expanding and improving ferry services is essential to grant all communities, particularly those on isolated islands, equal access to the economic and social opportunities concentrated in the Seattle-Tacoma area. Future developments should focus on balancing service expansion, environmental sustainability, and, most importantly, fair access for all communities. Recommendations and results are discussed through this lens, aiming to create a ferry network that not only efficiently moves people but also bridges opportunity gaps across the Puget Sound region.

1. An pedestrian autoferry system could create a vital transportation link by connecting the greater Port Orchard and Bremerton area with Bainbridge Island and Seattle. Port Orchard, characterized by high poverty and unemployment rates, a significant percentage of people with health challenges and presence of mobile homes, faces substantial socioeconomic challenges (Cluster 1). This area, which is home to a substantial population of American Indians and Alaska Natives, could benefit from improved connections to more economically stable regions.
2. Another potential autoferry pedestrian route could connect rural parts of Vashon Island and Tahlequah with the greater Tacoma area, including Point Defiance and Des Moines.

Vashon/Tahlequah, situated in a rural area, faces distinct socioeconomic challenges compared to the urban area of the greater Tacoma area (Figure 5). Cluster 1, which includes Tahlequah, is characterized by challenging socioeconomic conditions, whereas Cluster 5, which includes Point Defiance, may offer more work opportunities due to proximity to Tacoma, a larger urban area. An autoferry route could hug the perimeter of Vashon Island, stopping at smaller docks on Maury Island like Dockton.

3. The pedestrian autoferry can also serve as a connector from rural areas to existing ferry terminals like Kingston that then connect passengers to existing ferry routes. Kingston belongs to Cluster 3, characterized by relatively low poverty and unemployment rates but a high percentage of elderly residents. A pedestrian ferry could take passengers from the Suquamish and Indianola areas to Kingston, which then provides connections to the greater Seattle metropolitan area.
4. Finally, the autoferry system can also include terminals within Cluster 2, connecting mainland areas with islands. This aspect of the network could address the unique challenges faced by island communities, such as limited daily ferry service. In the south sound, the islands of Ketron and Anderson Island could be connected with Longbranch, Fox Island, Steilacoom, University Place, and Gig Harbor.

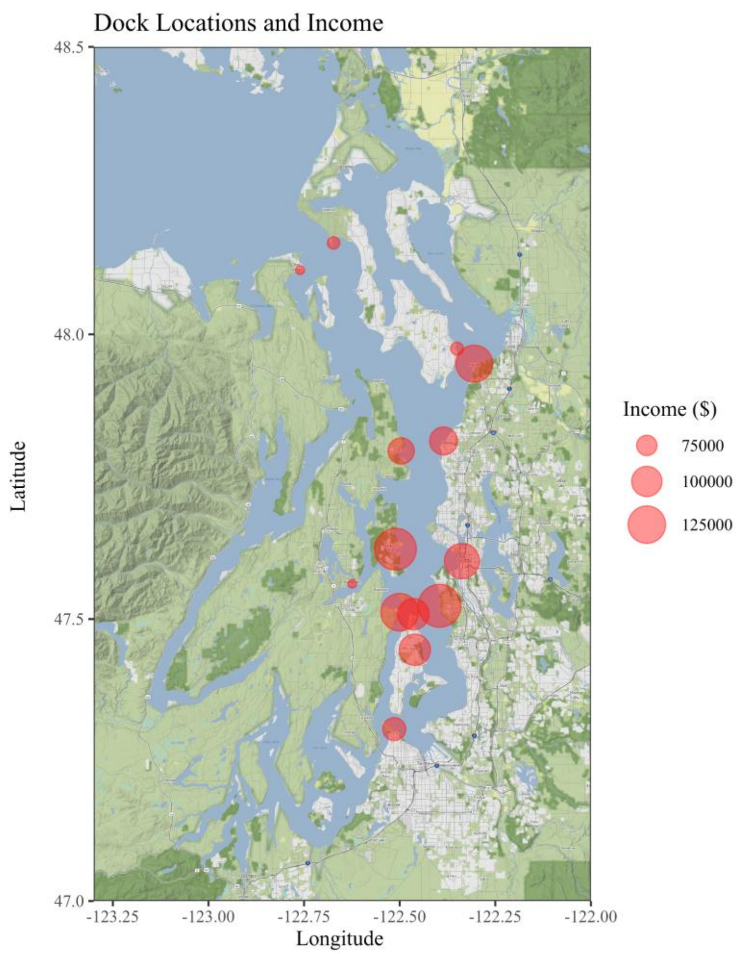

Figure 6 Map of existing ferry terminals in the Puget Sound with circles representing the median income by census track.

The cluster analysis indicates that there are suburban areas on the Kitsap peninsula that represent an opportunity for additional small ferry service that supports communities. Figure 6 shows the median income variation based on the census data for each existing ferry terminal. Whidbey Island and Port Townsend offer opportunities based on income levels. Several of the locations that represent more rural groups could provide an important opportunity for future routes.

6.0 Summary

This study examines the potential benefits and fair implementation of autonomous pedestrian ferries in Washington State, aiming to understand how these ferries can provide fair transportation across the state. To achieve this, correlation, principal component, and cluster analyses were performed to identify key factors and patterns that influence transportation demographics and to classify existing ferry terminals into distinct clusters based on socioeconomic and demographic characteristics of census tracts where these terminals are located.

The correlation analysis results indicate that census tracts with higher poverty rates are correlated with higher proportions of American Indian or Alaska Native, Black/African American, disabled individuals, households without vehicles, and mobile homes. These tracts also show higher unemployment rates and lower education levels, highlighting the concentration of marginalized groups and reliance on public assistance in high-poverty areas. Census tracts with denser transit stops are positively correlated with larger proportions of Black/African American and Asian residents, as well as high rates of households without vehicles and mobile homes. This indicates efforts to improve transportation for minority individuals who predominantly rely on public transportation, though there is still a need for better connectivity to job opportunities in the Seattle-Tacoma region. Additionally, areas with lower educational attainment are correlated with higher proportions of Asian and Hispanic populations, mobile homes, and denser transit stops, reflecting the need for targeted educational and transportation support. Tracts with higher proportions of elderly and younger individuals are negatively correlated with higher percentages of Asians, Blacks/African Americans, and transit stop density, underscoring challenges related to inadequate housing and transit infrastructure in these communities.

The PCA results reveal five PCs that define clusters of terminals based on the initial seventeen demographic variables considered. Among these variables, the most contributing factors to these PCs are the percentage of Blacks/African Americans, the percentage of Asians, the percentage of American Indians or Alaska Natives, the percentage of individuals of two or more races, the percentage of Hispanic or Latino individuals, the percentage of mobile homes, the percentage of persons below 150% poverty estimate, the unemployment rate estimate, the percentage of persons with no high school diploma, the percentage of households with no vehicle available estimate, and the number of transit stops per square mile. The cluster analysis highlighted five clusters of different characteristics. Proposed autoferry routes include several strategic connections aimed at addressing socioeconomic differences and improving access to essential services across different clusters. The Port Orchard (Cluster 1) route would link a high-poverty area with more stable economic regions, enhancing job access and services for Port Orchard residents. The Vashon/Tahlequah (Cluster 1) route would better connect Vashon Island to Des Moines and Tacoma, providing economic opportunities and access to residents. Connecting Kingston (Cluster 3) to nearby regions would benefit Kingston's residents by creating a more central regional hub.

The recommendations for pedestrian ferry routes focused on shorter connections (frequent trips) to recommend autoferry routes. Consequently, the proposed routes were recommended for their potential to offer more frequent service beyond addressing demographic issues across many areas in the Puget Sound. However, future studies can incorporate travel distance and time considerations to optimize route planning and include smaller public docks as possible ferry stops. Investigating the economic viability of these routes and their impact on local communities would also provide valuable insights for policymakers and transportation planners.

References

[1] T. Adler, Y. Dehghani, and C. Gihring, “Estimating Price Elasticities of Ferry Demand,” *Transp. Res. Rec.*, vol. 2176, no. 1, pp. 59–66, Jan. 2010, doi: 10.3141/2176-07.

[2] “Transit Statistics - American Public Transportation Association.” Accessed: Jul. 30, 2024. [Online]. Available: https://www.apta.com/research-technical-resources/transit-statistics/

[3] J. McNichols, “Washington’s ferry system is seeing the impact of decades of underfunding,” *NPR*, May 11, 2024. Accessed: Jul. 15, 2024. [Online]. Available: https://www.npr.org/2024/05/11/1250756750/washingtons-ferry-system-is-seeing-the-impact-of-decades-of-underfunding

[4] L. Denkmann and S. Leibovitz, “Could opting for diesel-powered over electric ferries get Washington new boats faster?” Accessed: Jul. 30, 2024. [Online]. Available: https://www.kuow.org/stories/could-opting-for-diesel-powered-ferries-get-washington-its-new-boats-faster

[5] “Clean Energy Transformation Act,” Washington State Department of Commerce. Accessed: Jul. 29, 2024. [Online]. Available: https://www.commerce.wa.gov/growing-the-economy/energy/ceta/

[6] “Ferry system electrification | WSDOT.” Accessed: Jul. 29, 2024. [Online]. Available: https://wsdot.wa.gov/construction-planning/major-projects/ferry-system-electrification

[7] CRAN, “R Software.” The Comprehensive R Archive Network, 2019. [Online]. Available: https://cran.r-project.org/

[8] “WSDOT Geospatial Open Data Portal.” Accessed: Jul. 29, 2024. [Online]. Available: https://gisdata-wsdot.opendata.arcgis.com/

[9] “CDC/ATSDR Social Vulnerability Index (CDC/ATSDR SVI).” Accessed: Jul. 29, 2024. [Online]. Available: https://www.atsdr.cdc.gov/placeandhealth/svi/index.html

[10] K. Pearson, “LIII. On lines and planes of closest fit to systems of points in space,” *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

[11] J. MacQueen, “Some methods for classification and analysis of multivariate observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, vol. 5.1, University of California Press, 1967, pp. 281–298. Accessed: Jul. 29, 2024. [Online]. Available: https://projecteuclid.org/ebooks/berkeley-symposium-on-mathematical-statistics-and-probability/Proceedings-of-the-Fifth-Berkeley-Symposium-on-Mathematical-Statistics-and/chapter/Some-methods-for-classification-and-analysis-of-multivariate-observations/bsmsp/1200512992

[12] H. F. Kaiser, “A Note on Guttman’s Lower Bound for the Number of Common Factors,” *Br. J. Stat. Psychol.*, vol. 14, no. 1, pp. 1–2, 1961, doi: 10.1111/j.2044-8317.1961.tb00061.x.

[13] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. STHDA, 2017.

[14] T. M. Kodinariya and P. R. Makwana, “Review on determining number of Cluster in K-Means Clustering,” 2013.

[15] C. Gayán-Navarro, M. Sanso-Navarro, and F. Sanz-Gracia, “An assessment of poverty determinants in U.S. census tracts, 1970–2010,” *Growth Change*, vol. 51, no. 3, pp. 977–999, 2020, doi: 10.1111/grow.12389.

[16] IES, “National Center for Education Statistics.” Accessed: Jul. 29, 2024. [Online]. Available: https://nces.ed.gov/programs/edge/Geographic/LocaleBoundaries