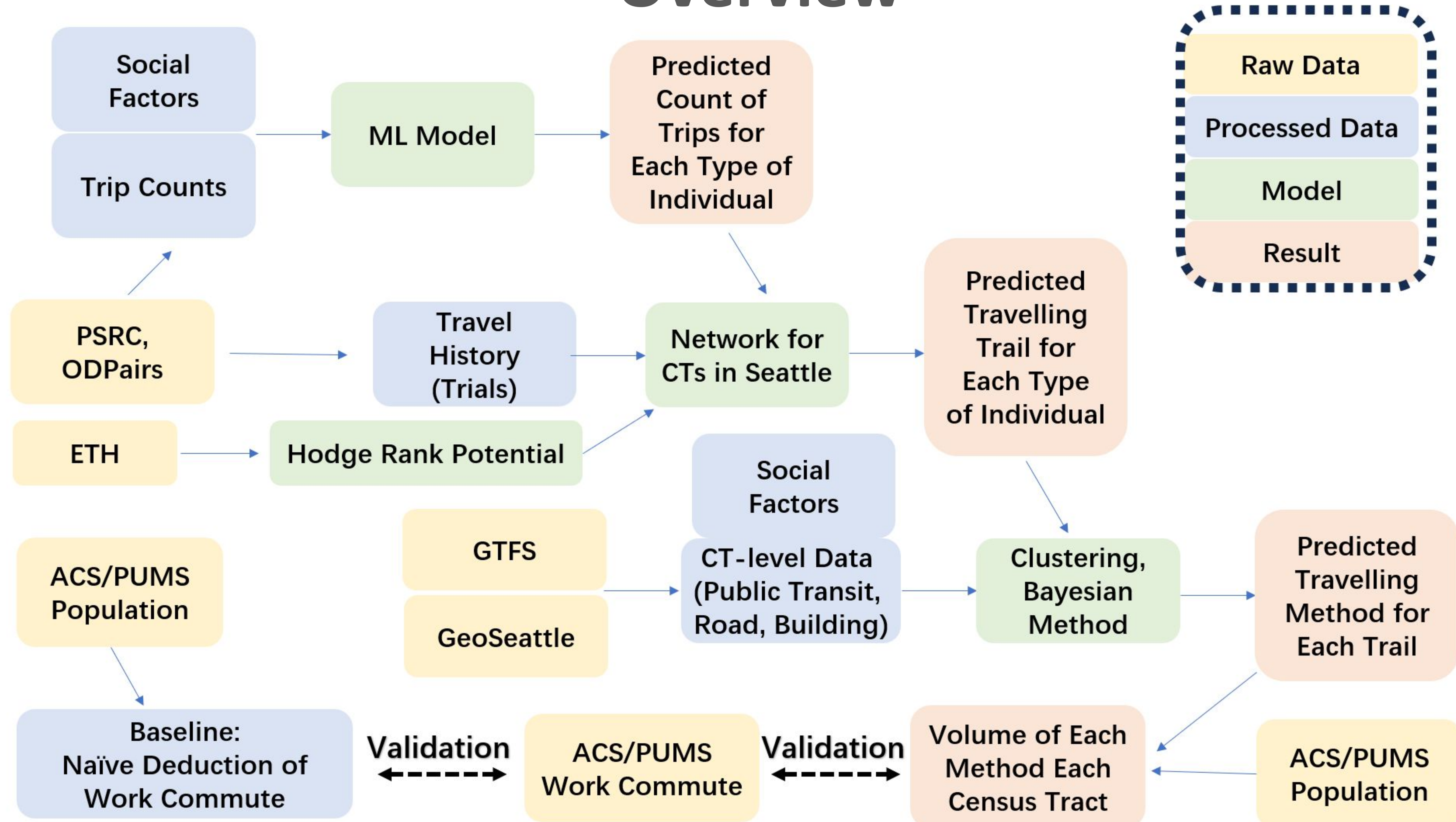
Unraveling Seattle's Travel Patterns Through Data Science Exploration

Yangyang Wang, Yifei Sun - UM-School of Information SI699, UM-Transportation Research Institute

UMTRI
SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

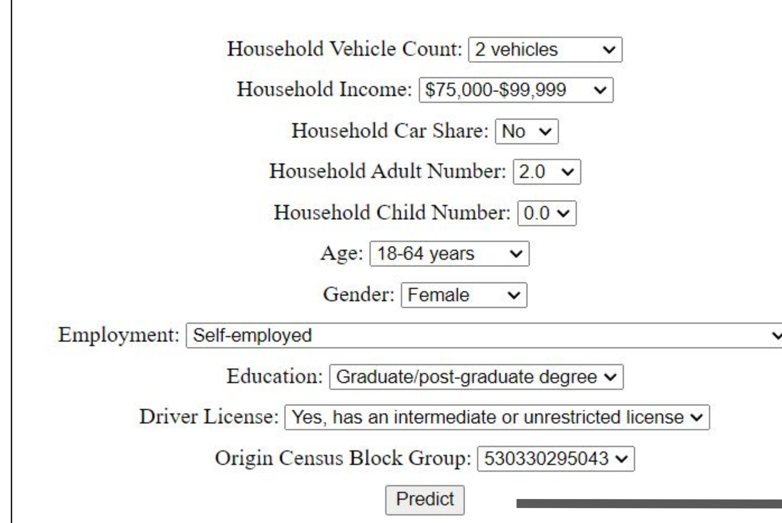
Overview

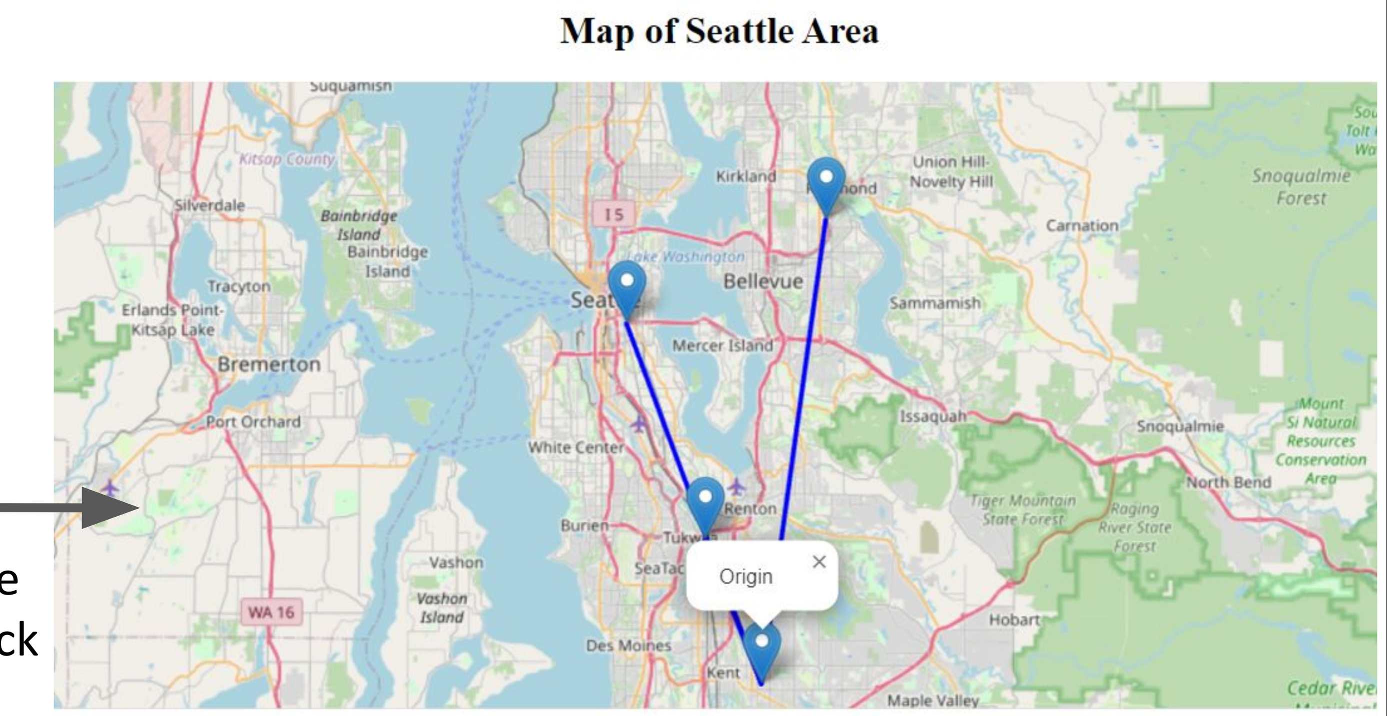


Email: wyy@umich.edu, yifeisun@umich.edu
Github: https://github.com/wyy-frank/SI699-Repo
Website: https://seattle-travel.replit.app

This study explores Seattle's travel patterns through a data-driven approach. We analyze social demographic factors, origin-destination pairs, and transportation modes to predict travel behaviors. Employing machine learning models and network analysis techniques, we uncover insights into travel behavior and network structures. Additionally, we develop a real-time user interface for interactive exploration. Validation with real-world data and population synthesis further enhances the robustness of our findings. Our study contributes to understanding travel dynamics and informs policy-making for transportation infrastructure in urban areas.

User Interface



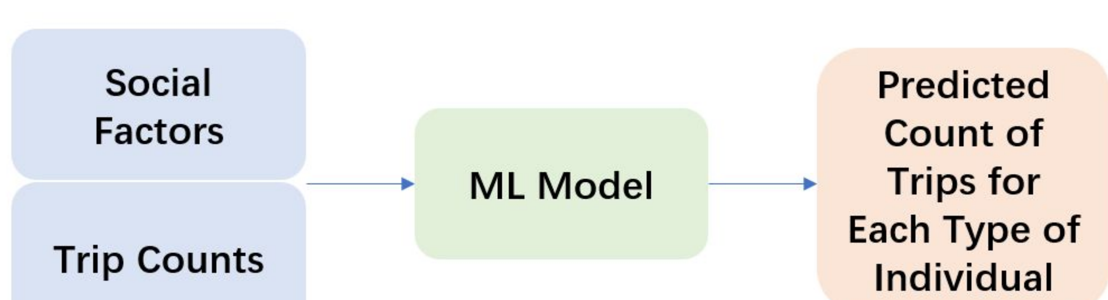


After selecting the social factors of the individual and the starting census block group, click "Predict".
The predicted trip routes would then be displayed on an interactive Leaflet map, with all the probability of corresponding travel methods indicated.

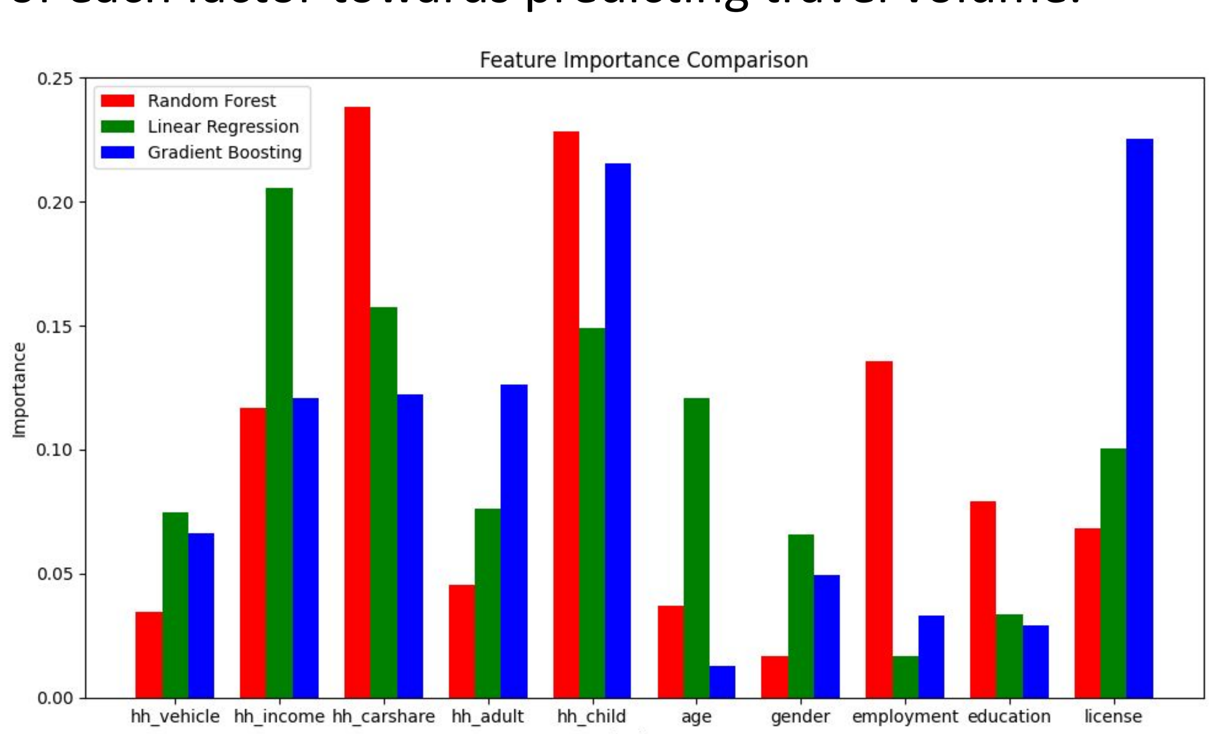
Method Probability

Destination	Bike	Carpool	Other	Drive	Public Transit	Walk
530330094005	0.01	0.33	0.0	0.58	0.0	0.08
530330228031	0.01	0.34	0.0	0.58	0.0	0.07
530330262001	0.0	0.27	0.0	0.59	0.0	0.14

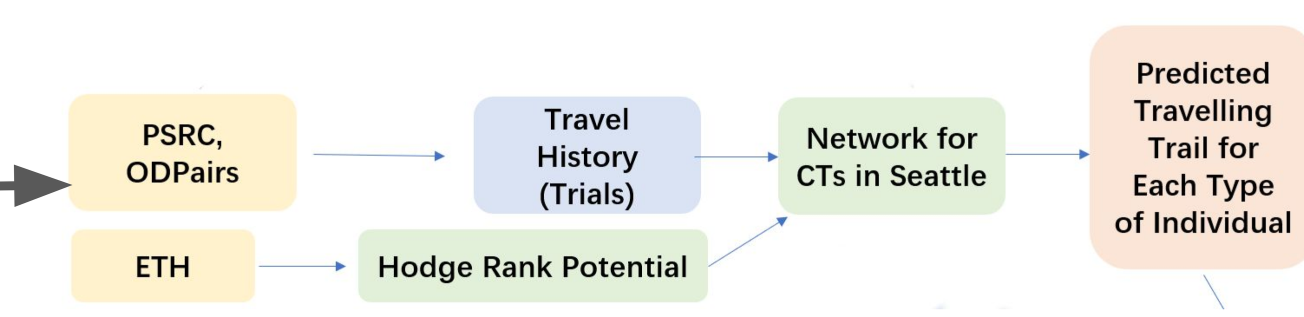
Step 1. Volume Prediction



- Utilized one-hot encoding to represent ten different household and individual-level features for each data record in the PSRC dataset.
- Applied spectral clustering to delineate distinct areas within King County, Seattle, leveraging both social and geographical attributes.
- Utilized the identified factors, comprising demographic characteristics and spatial properties, as input variables (X), with trip volume serving as the target variable (Y).
- Employed models such as Random Forest or Gradient Boosting to analyze feature importance.
- Examined feature importance using SHAP (SHapley Additive exPlanations) to understand the contribution of each factor towards predicting travel volume.

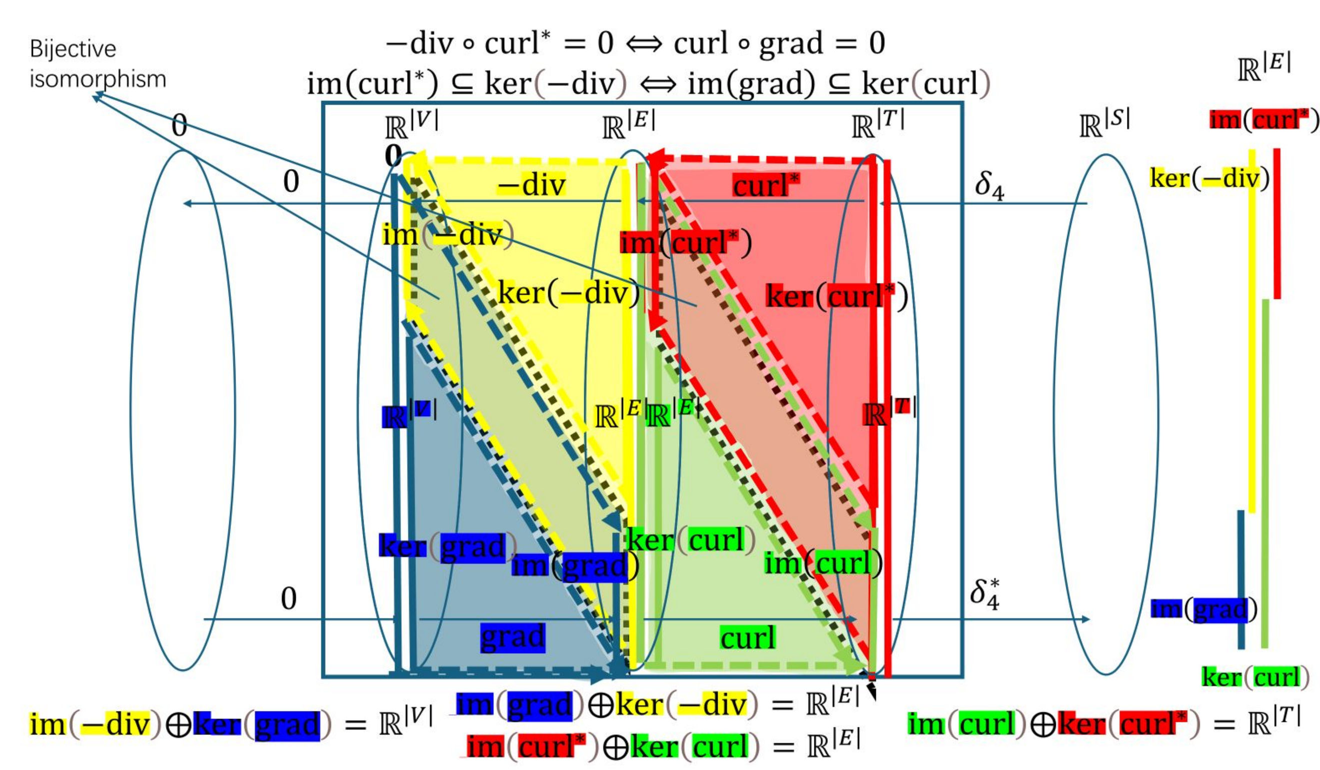


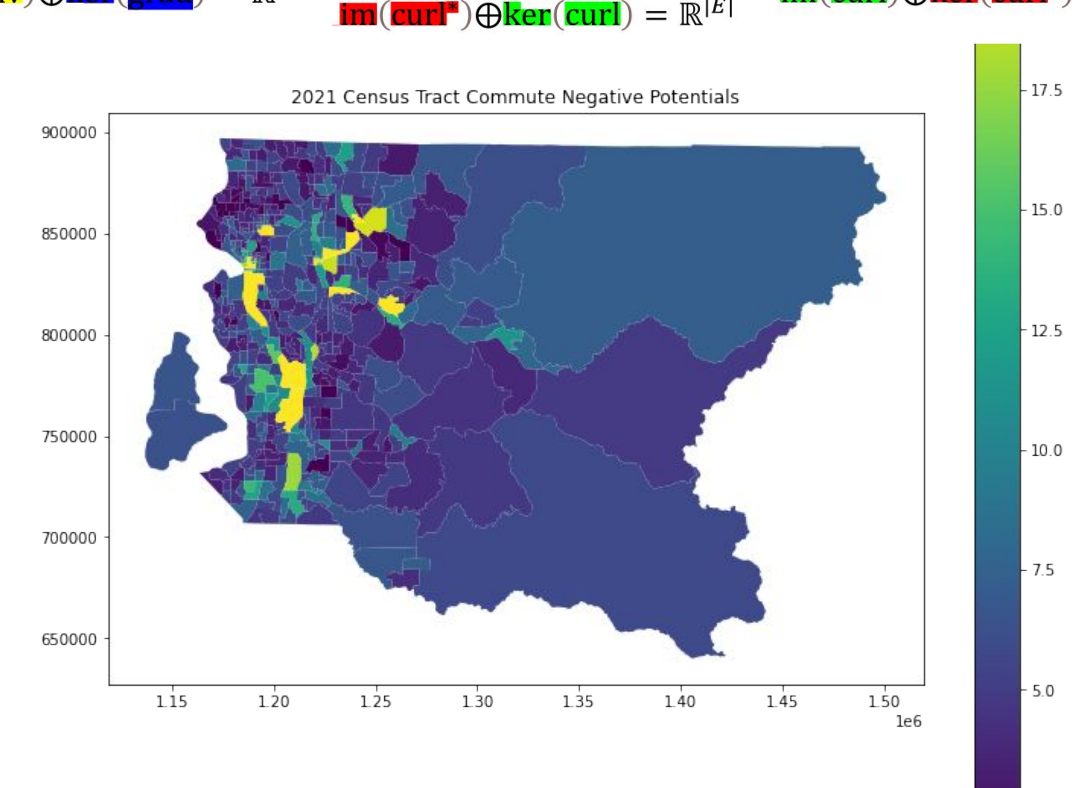
Step 2. Origin-Destination Planning



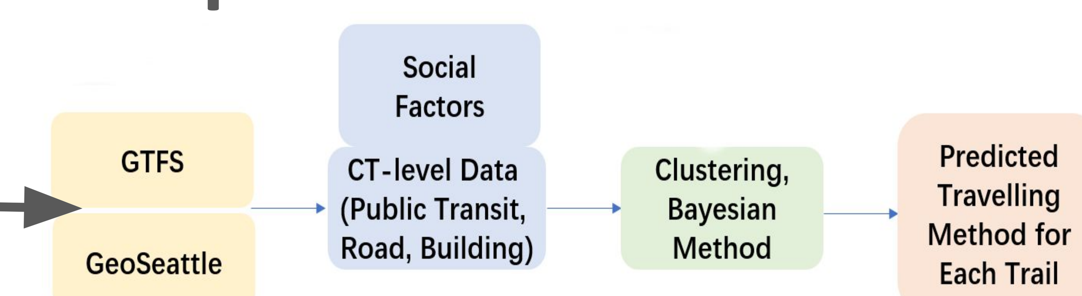
$$\mathbb{R}^{|E|} = \mathrm{im}(\mathrm{curl}^*) \oplus \ker(-\mathrm{grad}\circ\mathrm{div} + \mathrm{curl}^*\circ\mathrm{curl}) \oplus \mathrm{im}(\mathrm{grad})$$

$$\boldsymbol{f} \in \mathbb{R}^{|E|}, \quad \min_{\boldsymbol{r}\in\mathbb{R}^{|V|}} \|\boldsymbol{f} - \mathrm{grad}\,\boldsymbol{r}\|_W^2$$





Step 3. Method Inference



Spectral clustering

$$\underset{F}{\mathrm{argmin}}\,\mathrm{tr}(\boldsymbol{F}^T \boldsymbol{L} \boldsymbol{F}) = \mathrm{tr}(\boldsymbol{F}_{n\times k}{}^T \boldsymbol{L}_{n\times n} \boldsymbol{F}_{n\times k}),$$

Normalized cut:

$$(f_{A_m})_i = \begin{cases} 0, & v_i \notin A_m \\ \dfrac{1}{\sqrt{vol(A_m)}}, & v_i \in A_m \end{cases}$$

$$\boldsymbol{F}^T \boldsymbol{D} \boldsymbol{F} = \boldsymbol{I}$$

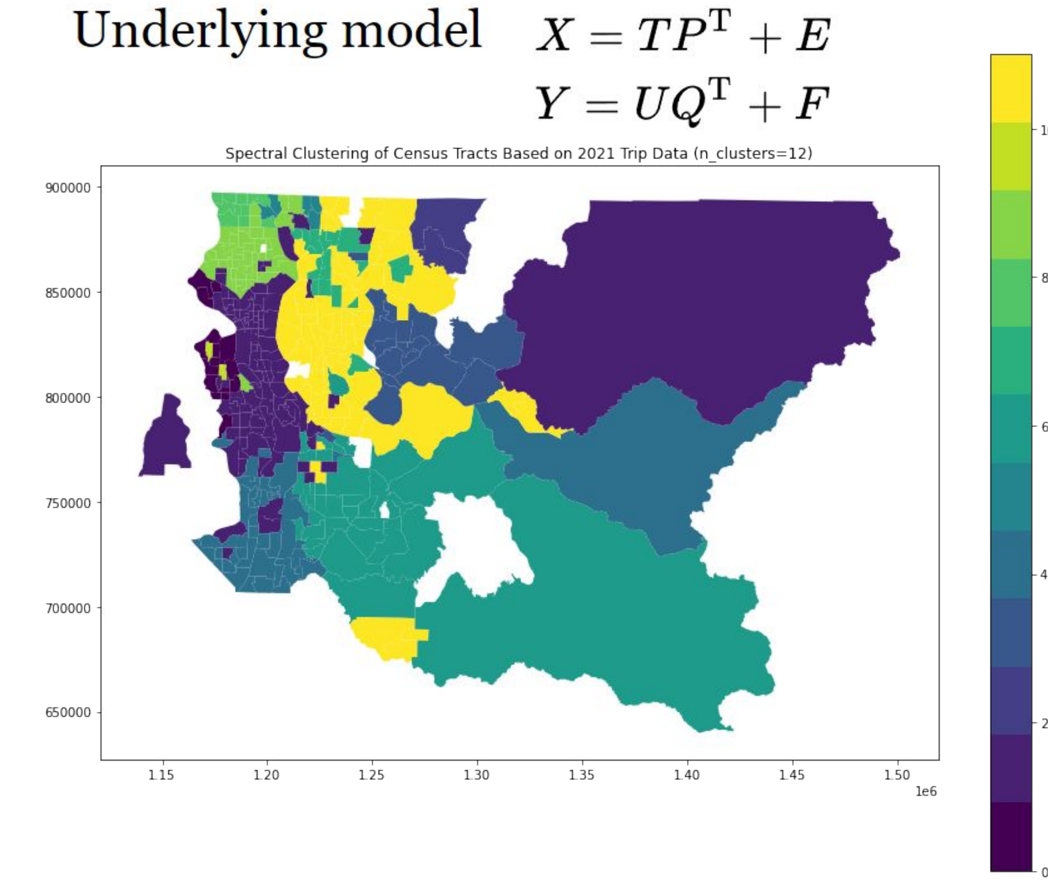
Partial least squares regression

Given n paired random samples $(\vec{x}_i, \vec{y}_i), i \in 1, \ldots, n$. In the first step $j = 1$, the partial least squares regression searches for the normalized direction $\vec{p}_j, \vec{q}_j$ such that the correlation is maximized

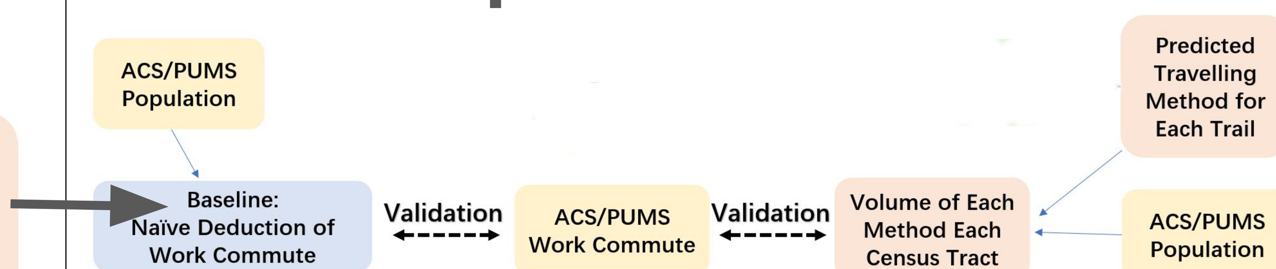
$$\max_{\vec{p}_j, \vec{q}_j} E_{\vec{X},\vec{Y}}[\underbrace{(\vec{p}_j \cdot \vec{X})}_{t_j}\underbrace{(\vec{q}_j \cdot \vec{Y})}_{u_j}]$$

Underlying model $\quad X = TP^{\mathrm{T}} + E$

$$Y = UQ^{\mathrm{T}} + F$$



Step 4. Validation



How to Validate?

- Combined with the population of individuals in each census tract to derive transportation volume.
- Compared against a naive approach distributing volume based solely on population proportion.
- Validated against Work Commute data.

Why Our Work Is Significant?

- **Accuracy Improvement:** Significantly outperformed the baseline in detecting low accuracy methods (transit, walk, and bike).
- **Granularity Enhancement:** Achieved a finer level of granularity, providing data specific to census block groups compared to the coarse census tract data in ACS.
- **Policy Insights and Future Directions:** Offered city routes and attraction analysis for better decision-making and transportation planning.