

## Regression analysis of demographic factors on the public transit ridership

### 1. Motivation & Goals:

This proposal mainly wants to deal with some problems about the public transit planning and design. One important issue about bus stop design is to estimate how many potential passengers exist in each area. The actual demand of a bus stop is recorded as the ridership in that stop and there might be many factors related to that ridership. This project aims to detect the relationship between several demographic factors (i.e. population, income, age, etc.) and the actual ridership around each bus stop.

### 2. Methods

This study first aggregated the ridership data of each bus stop to each census block group level. Therefore, for each block group, there will be a dependent variable ridership, and a few independent variables: population, age, income, etc. The multi-linear regression, the spatial lag model and the geographically weighted regression will be conducted to examine the relationship. The contribution of each variable will be analyzed and the most significant variables will be identified.

### 3. Results

Among the three models used, the spatial lag model obtained a smallest AIC value of 34433. The GWR model is the second best with an AIC value of 34440 and the multiple-linear regression model gained an AIC value of 34442. By comparing the R-squared of the three models, the GWR has the highest R-squared. In general, the spatial relationship between the dependent variable and the independent variables are not very strong and there might be caused by some data accuracy issues. However, our result can still prove that including the spatial consideration in the modeling process can help improve the result.

