We propose a project to find the optimal public transportation system for a city, given its demographics, city centers, etc.

First, we need to learn what makes a transportation system "good", using the population's transit ridership percentage as the main figure of merit. We intend to look at factors such as the locations of each station with respect to the surrounding features (residential areas, business districts, parks, sports stadiums, etc.), the demographics of the people around at each station, which kind of transit system generates the most ridership and/or revenue, the timing of the stops, etc. We will train our algorithm both on cities where a high percentage of the population uses public transit (for example New York, San Francisco) and cities that have poor percentage ridership (for example Los Angeles). With any luck, our algorithm will learn what traits of a city correlate highly with ridership.

Our next step is to take the model we've learned and apply it - given a new city with data on its neighborhoods, demographics, etc., our algorithm will return the public transportation system that maximizes ridership while minimizing cost (where cost is defined as cost of maintaining infrastructure per mile minus the revenue generated from ticket sales).

Of course, to simplify the model we will be making many assumptions - we won't be taking into account the cost of building the system (only maintaining it) or relocating the people (as more often than not it becomes a political issue), that the fare of each transit system is fixed, and most importantly we will assume that transit culture is the same across North America (we assume that Angelinos are just as willing to take public transit as New Yorkers if the transit system is optimal).

We will be using the Ridership Report Archives collected by the American Public Transportation Association (http://www.apta.com/resources/statistics/Pages/RidershipArchives.aspx) as our source of data on ridership, revenue, and cost. To obtain transit data of bus routes and stops, we will be making use of Google Map's public transit data feed (https://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds). We will map both of these against census data (http://www.census.gov/) for demographic information.