The Success of New York City Public Transport: a Comparative study with other major cities in the United States

Prepared for:
Professor Scott D. Campbell

Prepared by:
Siu Kee CHAN
21 April 2003
Introduction: City-rail and Urban Planning

Transportation is always an issue in Urban Planning. Major cities in the world are suffered from various degrees of traffic congestion problem and solving such problem is a long lasting issue for city planners.

Although there are many ways to deal with transportation issues—from extending road capacity and introduction of car pool lane, to the extensive of bus and train system—Intra-city massive transit railways system, in form of underground, surface, elevated or a mixture of the above, has been regarded as the most environmental friendly options, and of highest maximum capacity.

Although most of the cities in the US have some provision of public transportation, most of them are only limited bus services. Among them, only a handful of cities have established an extensive and comprehensive public transportation system that includes rail-type transit system such as intra-city train, light rail or subways. New York, the US city of the highest percentage of public transport, is always considered to be the only exceptional case for the success of public transportation in this country. However, this maybe true if we considering only the absolute numbers since New York also has the most exceptional subway train system in the US! Are there any other cities in the US that have a similar level of public transportation utilization—after factoring into account of the physical system? Will the physical character of the city-rail system relate (and results) to the utilization rate of public transit in the city? This is my desire to investigate such relationship in this project.

Research Question and Hypothesis:

In order to understand the scenario across the board, I set the research question as:

*Does the density of station distribution and complexity (coverage) of the subway/city-train system lead to the utilization of the Public Transport in a City?*

And the following Hypotheses:

1. The higher the density of station distribution (physical coverage) in the city, the more people are willing to use the public transport

2. The more comprehensive service coverage (frequency of service) of the city-train system,
the more people are willing to use the public transport.

Data Collection and Analysis

Population and Use of Public Transport

I selected city/metropolitan areas as the unit of analysis since this is the typical scale of coverage of a city-rail system. Data for area, population and transportation utilization is from the US Census 2000 through the Census website (http://www.census.gov). I ranked the cities and metropolitan areas (I excluded CMSA entries since it covered an area too large for a valid analysis) according to their population and then find out the top 10 cities the HAVING a city-rail system.

Rail Systems

I search the official websites of each of the 10 city-rail systems to get their system map and detailed timetable of the whole system.

From the system map, I can determine the number of stations, number of lines and number of interchange stations, the three variables I considered to be a representation of the coverage and complexity/usability\(^1\) of the system. (I converted the graphical data to interval quantitative data for the analysis)

From the timetable, I extracted the frequency of service of every lines in a system and average them in order to get the frequency of the whole system during peak hour. I selected only peak hours since the data from the census in considering people commuting to work. By doing this, I am hoping to have a more accurate result by matching the time of travel with type of travel. (I tried to combine the individual line data to represent the whole system)

\(^1\) A complex system does not mean it is difficult to navigate. On the other hand, it allows passengers to get from their origin and to their destination with minimal walking distance and allow faster transportation time through changing lines (short cut). Therefore, in this research, I assume the complexity of system as an advantage.
### Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Concepts</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Stations</strong></td>
<td>The spatial coverage of the system in the cities. Assuming the more stations the higher the coverage. Indirectly, this can relate to the accessibility of the system, which is of inverse relationship of walking distance to/from the stations and origin/destination.</td>
<td>The total number of stations of the whole system from the system map</td>
</tr>
<tr>
<td><strong>Number of Lines</strong></td>
<td>The spatial coverage and complexity of the system. Since each station have a limited area of coverage (i.e. walking distance, except with efficient feeder shuttle service), a single line extend across an area will left out people further away from the line. If two or more lines running in similar direction serve an area but each separated in certain distance, then more people can be included in the system.</td>
<td>The total number of Lines (or routes) of the whole system, as defined by the rail system operator.</td>
</tr>
<tr>
<td><strong>Number of Interchanges</strong></td>
<td>The accessibility and effectiveness of the system through working as a “Network”. In order to allow people to commute effectively in a system, it is essential to have sufficient interchange point to allow change of lines and try to minimize the additional “out of the way” travel distance in order to change lines.</td>
<td>The total number of stations that allows a change of line/route of service in the system</td>
</tr>
<tr>
<td><strong>Frequency during Rush Hours</strong></td>
<td>The “Time” coverage of the system. The more frequency the service, the fewer waiting time (and higher system capacity) and the more efficient the system in terms of traveling time.</td>
<td>Average of the frequency all lines of service during rush hours.</td>
</tr>
</tbody>
</table>

### Statement of Methods Used

At first I would like to use regression since it provides a equation expressing the relationship of all the dependent variables to explain the independent variable. However, it is simply impossible to have enough cases to perform a regression test for this topic because there are only a handful of cities in the US with city-rail system. Therefore, I selected to use scatter plots, which provides a both theatrically and visually clear representation of the relationship between variables.

In order to test the hypothesis, independent variables (No. of Station, No. of Lines, No. of Interchange, and Frequency during Rush Hours) are plotted against the dependent variable (Percentage Population Using Public Transportation) respectively in scatter plots.

The Success of New York City Public Transport: a Comparative study with other major cities in the United States
For the frequency of service, I converted the number of minutes per train, as in the timetables into “Number of Trains per Hour”, in order to match the hypothesis “the more frequent of the service (coverage of service), the more people using it”

**Outliners / Excluded Case**

According to Census 2000 data, Detroit, MI is the sixth largest cities in the US by population that have a city-rail system—The People Mover—the elevated monorail system that connects major convention/exhibition and other areas in the downtown area. However, I decided to exclude Detroit in my study because of the following factors:

- The People Mover is a **single-track** system running in a **loop**, instead of the typical double-track system running in opposite directions.
- The coverage of People Mover is **limited** to a small area of downtown area instead of city-wide. It serves more like a downtown shuttle service rather than an extensive, long range massive transit system, as of other systems.
- From the station location, it appears to be **tailored to visitors** (of major convention/entertainment events) rather than daily commuting use of local residents. Most stops are visitor areas but not residential areas.
- The passenger capacity of the People Mover, in terms of individual train or system wise, is **much smaller** than a typical city-train system when compared with other cities that I included in this analysis.
- Base on the current **unit of analysis** (city/metropolitan areas), the People Mover will produce a **misleading result**. Detroit's population is spread out mainly in the suburbs outside the downtown, while the People Mover serving **ONLY** the downtown, therefore it cannot represent Detroit.

Based on the above reasons, I decided to exclude Detroit in my statistical analysis and this bringing my total number of cases down from 10 to 9 cities.

This special scenario of Detroit, however, can show as (qualitatively) that Detroit, being named the “World Capital of Automobiles”, is **not planned as public transit dominant**. The investment by the city to such limited use monorail system instead of city-wide massive transit can tell something about that. In fact, Detroit also have a limited bus system to its suburbs, but this will be out of the scope of this analysis.

---

*Have Subway? Will Travel! No Hassles!!*
New York, although is an outlier in terms of absolute values of all variables, it is actually have the similar relationship and fits well in the trend line and therefore stayed in the analysis.

SUMMARIZED Data Table

<table>
<thead>
<tr>
<th>City/Metropolitan Area</th>
<th>Total Population</th>
<th>Percent in car pool</th>
<th>Percent of Working Population using public transportation</th>
<th>Percent worked outside county of residence</th>
<th>Number of Stations</th>
<th>Number of Lines</th>
<th>Number of Interchange</th>
<th>Frequency (trains/hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Los Angeles, CA</td>
<td>9,519,338</td>
<td>15.1</td>
<td>6.6</td>
<td>7.3</td>
<td>50</td>
<td>3</td>
<td>3</td>
<td>6.67</td>
</tr>
<tr>
<td>New York, NY</td>
<td>9,314,235</td>
<td>8.3</td>
<td>47</td>
<td>46.3</td>
<td>428</td>
<td>25</td>
<td>45</td>
<td>8.57</td>
</tr>
<tr>
<td>Chicago, IL</td>
<td>8,272,768</td>
<td>11</td>
<td>12.5</td>
<td>24.1</td>
<td>139</td>
<td>6</td>
<td>4</td>
<td>6.67</td>
</tr>
<tr>
<td>Philadelphia, PA</td>
<td>5,100,931</td>
<td>10.1</td>
<td>9.7</td>
<td>38.2</td>
<td>53</td>
<td>2</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>4,923,153</td>
<td>13.4</td>
<td>11.2</td>
<td>51.4</td>
<td>83</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Atlanta, GA</td>
<td>4,112,198</td>
<td>13.6</td>
<td>3.7</td>
<td>49.2</td>
<td>38</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Dallas, TX</td>
<td>3,519,176</td>
<td>14.3</td>
<td>2.4</td>
<td>28.5</td>
<td>34</td>
<td>2</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Boston, MA/NH</td>
<td>3,406,829</td>
<td>8.2</td>
<td>13.9</td>
<td>39</td>
<td>126</td>
<td>4</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Cleveland, OH</td>
<td>2,250,871</td>
<td>9</td>
<td>4.1</td>
<td>21.6</td>
<td>49</td>
<td>4</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

Result of Data Analysis

Number of Stations

![Graph showing the relationship between Number of Stations in System and Percentage Working Population Using Public Transit](image)

The Success of New York City Public Transport: a Comparative study with other major cities in the United States
From this plot, we can see a strong positive linear correlation between the Number of Stations (i.e. spatial coverage of the system) and the Percentage Working Population Using the Public Transit (Utilization) with $R^2=0.979$. That means more the stations in the system has leads to more utilization of public transport.

**Number of Interchanges**

![Graph showing the relationship between Number of Interchange in System and Percentage Working Population Using Public Transit (OLD Counting Method)](image)

This is the plot using the old interchange counting method, in which I will count all interchange stations even if they are adjacent to each other. From this plot, we can see a strong positive linear correlation between the Number of Interchange (i.e. accessibility and “network” of the system) and the Percentage Working Population Using the Public Transit (Utilization) with $R^2=0.898$.

**Dallas, TX is an outlier:** Under the old counting method, Dallas has 10 interchanges but only 2.4% of working population using public transit. I tried to do the scatter plot again without Dallas, TX and results $R^2=0.964$, as shown below:
After the comments from initial finding poster, I re-interpret the data and counting **neglecting all adjacent interchange stations** (i.e. if multiple interchange of the same line combination are adjacent to each other, only count as 1). The new result is reflected both in the *summarized data table* and the graph below:
From this plot, we can see a **strong positive linear correlation** between the Number of Interchange (i.e. accessibility and “network” of the system) and the Percentage Working Population Using the Public Transit (Utilization) with $R^2=0.954$. *That means more interchanges in the system has leads to more utilization of public transport.*
**Number of Lines**

![Graph showing the relationship between number of lines and percentage working population using public transit.](image)

From this plot, we can see a **strong positive linear correlation** between the Number of Lines (i.e. spatial coverage and complexity of the system) and the Percentage Working Population Using the Public Transit (Utilization) with $R^2=0.979$. That means the more the lines the system has leads to more utilization of public transport.
Frequency of Service in Rush Hours

From this plot, we can see a Neglect-able (i.e. NO) correlation between the Frequency of Service during Rush Hours (i.e. time coverage of the system) and the Percentage Working Population Using the Public Transit (Utilization) with $R^2=0.095$. On the surface, this means the frequency of service has no effects to the utilization of public transport.

However, this may be a misleading result since there is a typical range of service frequency for a city-rail system. First, there is a hardware (physical) limitation of the maximum frequency of service and cannot go beyond that under safety condition. On other hand, there are certain fixed operating costs of such system that it will not save significant money through lower the service level beyond certain point. Also, even in a system there is variation of service frequency among different lines—by averaging the lines in a system to become system frequency masking the difference among different systems (cities).

If this is a laboratory experiment, we can investigate this variable by using the same system but changing the frequency and note the resulted changed in utilization, but it is unrealistic to do such experiment in a real city and thus obtaining such data is infeasible.
Interpretation

From the analysis, there is a strong positive linear correlation of the Number of Stations, Number of Lines and Number of Interchange in a city-rail system (independent variables) to the Percent of Working Population using Public Transportation (dependent variable), all with R-square values higher than 0.9. This is a statistically significant result prove such explanatory relationship. On the other hand, the Frequency of Service during Rush Hours showed insignificant relationship to the dependent variable.

From the observation of the three variables, we can found the trend that, in the US, a system with higher spatial coverage (density) and network complexity will leads to more utilization. And with the linear relationship among the data (cities) being analyzed, the success of New York is not unpredictable—The success$^2$ of New York’s Public Transportation is not because it is a strange case, but it has a far more developed, dense, complex subway system, than all other US cities, therefore linear proportionally having more people to use it.

Conclusion

The result of analysis supports the first hypothesis that:

*The higher the density of station distribution (physical coverage) in the city, the more people are willing to use the public transport*

but failed to support the second one, which is:

*The more comprehensive service coverage (frequency of service) of the city-train system, the more people are willing to use the public transport.*

Therefore:

*Does the density of station distribution and complexity (coverage) of the subway/city-train system lead to the utilization of the Public Transport in a City?*

**Spatial Coverage? Yes**

**Time Coverage? No**

---

$^2$ Success defined as the percentage of working population using public transportation.

The Success of New York City Public Transport: a Comparative study with other major cities in the United States

Have Subway? Will Travel! No Hassles!!
Future Improvement/Research:

- More Comprehensive Calculation of frequency of service. First, obtain the number of passengers served per line during rush hours. Then formulate the weighted-average (according to passengers served) of frequency of all lines to become the system’s frequency.

- Spatial Analysis (using GIS?) of Network Accessibility through macro data (census track/block level), for the population located at buffer area of stations.

- Instead of using information of city/metropolitan areas, use census track data by selecting only tracks that is covered by the city-rail and recalculate the result, this can solve the possibly mismatch between the city-wide data coverage and the actual city-rail coverage.

- It is actually my intention to compare “public transportation” use with the city rail, but this relationship may be questionable since there are other forms of public transportation (e.g. Bus system). In order to investigate this, we can do the same analysis with data Number of Passenger served which may be obtainable through individual system operators, then compare with the population value.
# References


## Websites:

<table>
<thead>
<tr>
<th>Name of Organization/Website</th>
<th>Universal Resources Locator (URL):</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Census</td>
<td><a href="http://www.census.gov">http://www.census.gov</a></td>
</tr>
<tr>
<td>American Public Transportation Association</td>
<td><a href="http://www.apta.com">http://www.apta.com</a></td>
</tr>
<tr>
<td>Domestic Metro Area Travel by Train</td>
<td><a href="http://www.quakerbonnet.com/t-usa-c.htm">http://www.quakerbonnet.com/t-usa-c.htm</a></td>
</tr>
<tr>
<td>Los Angeles Metropolitan Transportation Authority</td>
<td><a href="http://www.mta.net">http://www.mta.net</a></td>
</tr>
<tr>
<td>MTA New York City Subway</td>
<td><a href="http://www.mta.nyc.ny.us">http://www.mta.nyc.ny.us</a></td>
</tr>
<tr>
<td>Chicago Transit Authority</td>
<td><a href="http://www.transitchicago.com">http://www.transitchicago.com</a></td>
</tr>
<tr>
<td>Southeastern Pennsylvania Transportation Authority</td>
<td><a href="http://www.septa.org">http://www.septa.org</a></td>
</tr>
<tr>
<td>Washington Metropolitan Area Transit Authority</td>
<td><a href="http://www.wmata.com">http://www.wmata.com</a></td>
</tr>
<tr>
<td>Detroit People Mover</td>
<td><a href="http://www.thepeoplemover.com">http://www.thepeoplemover.com</a></td>
</tr>
<tr>
<td>Metropolitan Atlanta Rapid Transit Authority</td>
<td><a href="http://www.itsmarta.com">http://www.itsmarta.com</a></td>
</tr>
<tr>
<td>Dallas Area Rapid Transit</td>
<td><a href="http://www.dart.org">http://www.dart.org</a></td>
</tr>
<tr>
<td>Massachusetts Bat Transportation Authority</td>
<td><a href="http://www.mbta.com">http://www.mbta.com</a></td>
</tr>
<tr>
<td>Greater Cleveland Regional Transit Authority</td>
<td><a href="http://www.riderta.com">http://www.riderta.com</a></td>
</tr>
</tbody>
</table>

Have Subway? Will Travel! No Hassles!!