Exploring Transit-Oriented Development in Chicago: A Spatial Analysis from an Urban Design Perspective

Jizhe Yang
Department of Urban and Regional Planning
University of Illinois at Urbana-Champaign

Room 111 Temple Buell Hall
611 Taft Drive, Champaign, IL 61820
Phone: 217.979.2996
E-mail: yang98@illinois.edu

The paper should be part of the student paper competition.
Forum preference: Oral presentation before an audience

Abstract

In the recent years, transit-oriented development has become a popular solution to urban transportation problems caused by auto-dependence. Recent studies on designing and evaluating bus transit systems often employ Geographic Information Systems (GIS) and statistical tools to predict ridership. However, these technical approaches often ignore the fact that users' experience of public transportation takes a human-scale and involves intimate interactions with the surrounding built environment. This paper fills this void by combining spatial statistical analysis with a human-scaled urban design perspective while answering the key question: how does urban design influence transit ridership?

Through GIS and observation, the paper establishes spatial correlations between the ridership at Chicago CTA bus stops on a weekday and elements that represent good urban design considerations, such as green spaces, land use and building density around CTA stations.Adjacent sites with contrasting counts of ridership are further investigated through street views.

By exploring the transit systems through regression analysis with an urban design theme and visualizing street views around the bus stops, the study identifies urban design features that are significant to the success of transit-oriented development. However, the conclusion drawn from the spatial analysis is limited due to available data accuracy. Data for more weekdays and weekend ridership, businesses, vacancy, as well as age group and ADA status that may improve the regression model should also be probed for better understanding.

Introduction and Background

In the recent decades, automobile dependence has been fueling urban sprawl and a myriad of transportation problems such as pollution and congestion. As gasoline prices have been rising dramatically in recent years, dependence on private cars can increasingly hinder people's mobility, especially that of lower-income groups. Transit-oriented development (TOD) seeks to create a physical environment compatible with transit systems through enhancing built density, promoting mixed use, creating pedestrian-friendly environment and integrating various modes and routes of transportation (Sung, 2011). By doing so, TOD could promote an affordable, reliable and efficient alternative to automobile and minimize the negative externalities caused by private cars. Cities across the United States and worldwide have been adopting TOD to increase the efficiency of transportation, energy and land use (Lin, 2008).

Numerous researches have evaluated the impact of TOD and the effectiveness of its design factors using regression models. The direct measurement of the impact of TOD on promoting transit usage is ridership. Studies have found that TOD can significantly increase the level of ridership; however, the amount of increase and the specific design factors that have caused the increase vary across cities and regions, often due to discrepancies in local economy as well as cultural and historical traditions of land use patterns. For example, mixed use is often regarded as a vital factor in promoting transit ridership in U.S. cities, but has minimal impact on ridership in Asian cities because Asian cities have a long tradition of mixed use development that is insensitive and not exclusive to TOD (Lin, 2008). Similarly, while density should be emphasized for successful TOD in many U.S. cities, it is not a significant factor for many Asian cities that have already established high density historically (Sung, 2011). Therefore, factors that contribute to successful TOD are not universal and should be identified on a place-by-place basis, because discrepancies in local history, demographics and economy often lead to different urban forms.

As one of the cities in the U.S. actively pursuing TOD, Chicago has a long tradition of transit systems dating back to the late 1900s and runs the second largest transit system in the nation (Chicago Transit Authority, 2011). As of 2012, the Chicago Transit Authority (CTA) serves 11,493 bus stops and 143 train stations (CTA, 2011). Recently the CTA has designated 150 bus shelters to have Digital

Bus Tracker installed to display real-time bus arrival information (CTA, 2011). Despite Chicago's active practical efforts in enhancing TOD, there are few recent studies on what design factors are essential to promote ridership in the specific case of Chicago. Therefore, the purpose of this study is to identify design factors that are most influential in the context of Chicago to inform effective investment and planning decisions.

Research Design

Identifying Explanatory Factors

Studies have designed various schemes of classifying and identifying variables that affect ridership. One of the most appropriate scheme for this study would be Cervero's framework (2010), in which attributes and TOD design factors are categorized into three groups of independent variables: (1) attributes of location and neighborhood (eg. demographics, density, mixed use indices, distance to the nearest feature, etc.); (2) attributes of stops (eg. bus shelter, real-time information display, park-and-ride, etc.); and (3) attributes of service (eg. frequency, number of routes, operating speeds, etc.). The study will use Cervero's scheme as a basis to identify initial independent variables, but will modify the grouping of variables to suit local conditions and data availability.

This study explores the relationship between the ridership at an individual CTA bus stop and its surrounding environment in terms of TOD design factors. Table 1 lists the explanatory factors in terms of variables that the study is able to quantify and examine. The explanatory factors are categorized into three groups based on Cervero's scheme.

Data Sources

For the purpose of this study, data for ridership by bus stop, poverty rate, travel time to work, location of bus stops, location and area of open/green spaces, building density, location of stops with real-time information display, location of Metra and CTA rail stations, as well as numbers of bus routes serving each bus stop are needed. The data for poverty rate and travel time to work are available from U.S. Census Bureau, and all the other data can be obtained

Table 1. Candidate Variables for the Multiple Regression Model

Variable Group	Purpose	Variables
Attributes of location and neighborhood	Measure the impact of demographics of the area (by census tract) where the bus stop is situated	• Poverty rate • Travel time to work
Attributes of stops	Measure the impact of land use and urban design elements near the bus stop	• Distance from the bus stop to its nearest open/green space • Area of the nearest open/green space • Building density in the situated neighborhood (by census tract)
Attributes of service	Measure the impact of bus stop infrastructure and design	• Availability of real-time information display at the bus stop • Distance to the nearest transfer station • Number of bus routes serving the bus stop

from City of Chicago Data Portal. Ridership is used as the dependent variable, and all the other data are used as independent variables.

Preparing the Independent Variables

In order to convert these data into usable variables for regression analysis, several techniques in ArcGIS were used: display x-y data, calculate geometry, near, field calculator and spatial join. A discussion of how these techniques prepared each of the data for regression analysis is as follows:

1. Display x-y data: The data for CTA bus ridership by stop and the location of real-time information display are excel files with geographic coordinates. These excel files are projected based on geographic coordinates, resulting in a point dataset of 5268 bus stops with boarding and alighting numbers and a point dataset of all the stops with real-time information display.
2. Calculate geometry: All data of which the size of area is needed in the study underwent calculating geometry using square feet as the unit. This attribute would be later passed on to other datasets using spatial join for further calculation.
3. Near: The input feature is the dataset of bus stops, and the near features are independent variables for which the nearest is to be found for each bus stop. The operation returns the distance between each entry in the input feature and its closest near features. While only one input feature can be specified, multiple near features can be entered. In this study, three sets of near features are used one at a time, each set belonging to the same category of variables:
   i. Stop infrastructure: Stops with real-time information display;
   ii. Transfer connections: Metra stations and CTA rail stations.
4. Field calculator: Density and routes serving each bus stop are obtained using field calculator.
   i. Ridership: This field is calculated as the average of the numbers of boarding and alighting in the bus stop dataset.
   ii. Density: First the sizes of individual buildings are obtained by calculating geometry using the building footprint dataset. Next, buildings in the central business district (CBD) is spatially selected and exported as a separate dataset, of which the floor areas are multiplied by the number of stories to obtain ground floor area of each building. The reason of this special treatment is the availability of data: only 818 out of 2533 of the buildings in the CBD lack information about the number of stories, and many of these buildings are attached to each other and have similar heights, allowing for estimation using average building height. In contrast, city-wide 392801 out of 820154 of the buildings have no data for the numbers of stories, of which the sheer number could not guarantee a reasonable estimation. In addition, since the buildings inside the CBD tend to be significantly higher than buildings elsewhere, the results of regression analysis could be distorted if the number of floors is not accounted for by the density variable. Then the building footprint dataset of buildings in CBD and the dataset of all buildings in the study area are respectively spatially joined to census tract dataset with "sum" specified for the added fields. The resulted datasets have fields indicating the total ground floor area (the result of CBD dataset) and ground coverage (the result of all buildings) within each census tract respectively. Finally, these fields are divided by the area of their corresponding census tracts to obtain density indices used as independent variables in this study.
   iii. Number of routes serving each bus stop: the

dataset for bus stops has a field named "routes" that contains the names of routes serving each bus stop separated with a comma. The number of routes is calculated using Python script introduced by UP 519 class: Num_routes=(!routes!.count(","))+1.

5. Merge: The datasets of boulevards, habitats, mall plaza, parks, campus parks are merged as one dataset for open/green space. The reason why neighborhood gardens are not included in this category is that neighborhood gardens are represented as point data with area as a numeric attribute, while all the other datasets of open/green space are polygons.

6. Spatial join: This study uses bus stop dataset as the target features.

   i. In order to join demographic information and other census tract-level information such as poverty rate, travel time to work and density indices, the census tract dataset is specified as the join features and "within" is used as the match option.

   ii. To join information about open/green spaces and neighborhood gardens, the "closest" match option is used. The merged open/green space dataset is spatially joined to the resulted dataset of 6(a), and the product of this operation is further spatially joined by the neighborhood garden dataset. The reason why the technique of "spatial join" is used instead of "near" is that "spatial join" can not only calculate the distance to the nearest feature but also transfer the attributes of the nearest feature to the target feature, while all attributes will be lost using "near" technique. Since this study also needs the areas of open/green spaces and neighborhood gardens as independent variables, the use of spatial join is mandated for these two datasets.

The goal of using these techniques is to transfer all of the needed independent variables into the bus stop dataset each as a field, so that the regression analysis can access the needed information.

Regression Analysis

Regression analyses are performed in R to evaluate the significance and influence of demographic and design factors on CTA bus ridership. Ridership is used as the dependent variable for all regression analysis.

The initial ordinary least squares (OLS) model includes all independent variables:

1. NearestNei: distance to the nearest neighborhood garden
2. NeiG_Area: area of the nearest neighborhood garden
3. Near_OPSP: distance to the nearest neighborhood garden
4. OPSP_area: area of the nearest neighborhood garden
5. NearDispla: distance to the nearest stop with real-time information display
6. NearTransf: distance to the nearest Metra or CTA rail station
7. TTW_MINS: the travel time to work of people living in the census tract where the stop is located
8. POV_RATE: the poverty rate of the census tract where the stop is located
9. density: the building density of the census tract where the stop is located
10. num_routes: number of routes serving the stop

Street View Analysis

The stops are classified into three categories according to ridership level using natural breaks. The streetscapes of nine adjacent bus stops along the Jackson Boulevard are examined through Google Street View (see Figure 2).

Table 2. Results of the OLS Regression Model for CTA Bus Stops

```
Call:
lm(formula = sqrt(ridership) ~ NearestNei + NeiG_Area + NearDispla +
    NearTransf + Near_OPSP + OPSP_area + TTW_MINS + POV_RATE +
    density + num_routes, data = rider.df)

Residuals:
    Min      1Q  Median      3Q     Max
-19.370  -3.089  -0.904   1.947  33.001

Coefficients:
                   Estimate      Std. Error t value  Pr(>|t|)
(Intercept)  4.240668474073  0.452839220961   9.365   < 2e-16 ***
NearestNei   0.000016891235  0.000020215200   0.836  0.403434
NeiG_Area   -0.000023481879  0.000008693020  -2.701  0.006930 **
NearDispla  -0.000368434234  0.000031496779 -11.698   < 2e-16 ***
NearTransf  -0.000140812722  0.000033215258  -4.239 0.0000228 ***
Near_OPSP   -0.000080408831  0.000100453277  -0.800  0.423480
OPSP_area    0.000000023528  0.000000008618   2.730  0.006355 **
TTW_MINS     0.004409186647  0.006904418698   0.639  0.523109
POV_RATE     0.019258290309  0.005239049731   3.676  0.000239 ***
density      8.769277951974  0.871810327873  10.059   < 2e-16 ***
num_routes   2.182981022018  0.061359605565  35.577   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.826 on 5257 degrees of freedom
Multiple R-squared: 0.289,      Adjusted R-squared: 0.2876
F-statistic: 213.7 on 10 and 5257 DF,  p-value: < 2.2e-16
> extractAIC(model.1)
[1]    11.00 16594.22
```

Table 3. Results of the Modified OLS Regression Model for CTA Bus Stops

```
Call:
lm(formula = sqrt(ridership) ~ log(NearDispla) + log(NearTransf) +
    OPSP_area + POV_RATE + density + num_routes, data = rider.df)

Residuals:
    Min      1Q  Median      3Q     Max
-19.605  -2.956  -0.729   1.966  31.867

Coefficients:
                       Estimate      Std. Error t value  Pr(>|t|)
(Intercept)     21.221571894075  0.797603116845  26.607   < 2e-16 ***
log(NearDispla) -1.410541034245  0.063909191759 -22.071   < 2e-16 ***
log(NearTransf) -0.901330481803  0.072257445662 -12.474   < 2e-16 ***
OPSP_area        0.000000025031  0.000000008052   3.109   0.00189 **
POV_RATE         0.019744573805  0.004632725284   4.262 0.0000206 ***
density          6.978776523040  0.802156009382   8.700   < 2e-16 ***
num_routes       1.877629777439  0.059744145333  31.428   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.603 on 5261 degrees of freedom
Multiple R-squared: 0.3525,     Adjusted R-squared: 0.3518
F-statistic: 477.4 on 6 and 5261 DF,  p-value: < 2.2e-16
> extractAIC(model.2)
[1]     7.00 16092.95
```

Results and Discussions

Regression Analysis

The results of the initial model for bus stops in the entire study area indicate that three factors are not significant: NearestNei (distance to the nearest neighborhood garden), Near_OPSP (distance to the nearest open/green space) and TTW_MINS (travel time to work). After eliminating the insignificant factors and using log transformation for NearDispla and NearTransf to boost normality (which renders NeiG_Area insignificant and subject to removal), the modified regression model yields the following values: Adjusted R-squared: 0.3518, p-value: < 2.2e-16, median residual: -0.729 (see Table 3).

To ensure that the model does not violate the underlying assumptions of OLS, several additional tests are conducted. Variance inflation factor tests reveal that the model has no problems with multicollinearity. The results of Shapiro-Wilk test and bp test indicate that the residuals of the regression model are not normally distributed and do not exhibit constant error variance (see Table 3). Therefore, robust.R is used to adjust the calculations for the presence of heteroscedasticity. The result shows that all independent variables in the modified regression model are statistically significant at 0.001 level of confidence.

The Moran's I statistic is then used to test the residuals of the modified regression model for the presence of spatial autocorrelation. The test returns a Moran's I statistic of 0.33, which indicates a moderate degree of positive spatial correlation, and a highly significant p-value, meaning that a spatial regression model is warranted. The following test using LaGrange Multiplier Statistics reveals that a spatial error model should be used, because the error LM statistic is the most significant (see Table 6). However, the spatial error model has a much higher AIC value than the modified OLS model (32334>16100.62), indicating that either spatial relationships do not matter or spatial relationships are not correctly conceptualized. Therefore, we can proceed with the modified OLS model as an approximate option.

The regression analysis for city-wide pattern is

Table 4. Tests to Warrant the Modified OLS Model

```
> vif(model.2)
log(NearDispla) log(NearTransf)        POV_RATE         density
       1.143883        1.125570        1.161657        1.179433
     num_routes
       1.136604
> bptest(model.2, varformula=NULL, studentize=TRUE, data= rider.df)

        studentized Breusch-Pagan test

data:  model.2
BP = 300.7851, df = 5, p-value < 2.2e-16
```

Table 5. Estimate Robust Standard Errors

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	21.37403136	1.072290102	19.933068	2.102693e-88
log(NearDispla)	-1.43018394	0.085511745	-16.725000	8.617613e-63
log(NearTransf)	-0.88127787	0.101643156	-8.670312	4.309382e-18
POV_RATE	0.01823464	0.004422548	4.123107	3.737957e-05
density	6.49354899	0.841985588	7.712185	1.236812e-14
num_routes	1.91983322	0.090223997	21.278521	1.795161e-100

Table 6. Moran's I Test and Model Choice

```
                    Global Moran's I for regression residuals

data:
model: lm(formula = sqrt(ridership) ~ log(NearDispla) +
log(NearTransf) + POV_RATE + density + num_routes, data = rider.df)
weights: nbRider

Moran I statistic standard deviate = 41.1672, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Observed Moran's I        Expectation           Variance
     0.33030880326     -0.00083257072      0.00006470304

> result <- lm.LMtests(model.2, nbRider, test=c("LMerr",
+ "LMlag","RLMerr", "RLMlag"))
> tresult <- t(sapply(result, function(x) c(x$statistic, x$parameter,
+ x$p.value)))
> colnames(tresult) <- c("Statistic", "df", "p-value")
> printCoefmat(tresult, signif.stars=TRUE, has.Pvalue=TRUE)
```

(Continue)

```
      Statistic df    p-value
LMerr   1679.491  1  < 2.2e-16 ***
LMlag   1185.371  1  < 2.2e-16 ***
RLMerr   511.537  1  < 2.2e-16 ***
RLMlag    17.417  1 0.00003002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:
errorsarlm(formula = sqrt(boardings) ~ NEAR_DIST + SQ_FT + num_routes,
    data = rider.df, listw = nbRider, method = "LU", quiet = TRUE)
Residuals:
     Min       1Q   Median       3Q      Max
-22.1672  -3.1457  -1.0938   2.0354  34.1223

Type: error
Coefficients: (asymptotic standard errors)
                Estimate   Std. Error z value       Pr(>|z|)
(Intercept)  5.794446842  0.266215394 21.7660      < 2.2e-16
NEAR_DIST   -0.000328113  0.000054434 -6.0277 0.000000001663
SQ_FT       -0.000034680  0.000014255 -2.4329        0.01498
num_routes   2.273147950  0.080881602 28.1046      < 2.2e-16

Lambda: 0.49476, LR test value: 707.22, p-value: < 2.22e-16
Approximate (numerical Hessian) standard error: 0.016812
    z-value: 29.429, p-value: < 2.22e-16
Wald statistic: 866.04, p-value: < 2.22e-16

Log likelihood: -16161.16 for error model
ML residual variance (sigma squared): 25.784, (sigma: 5.0778)
Number of observations: 5268
Number of parameters estimated: 6
AIC: 32334, (AIC for lm: 33040)
```

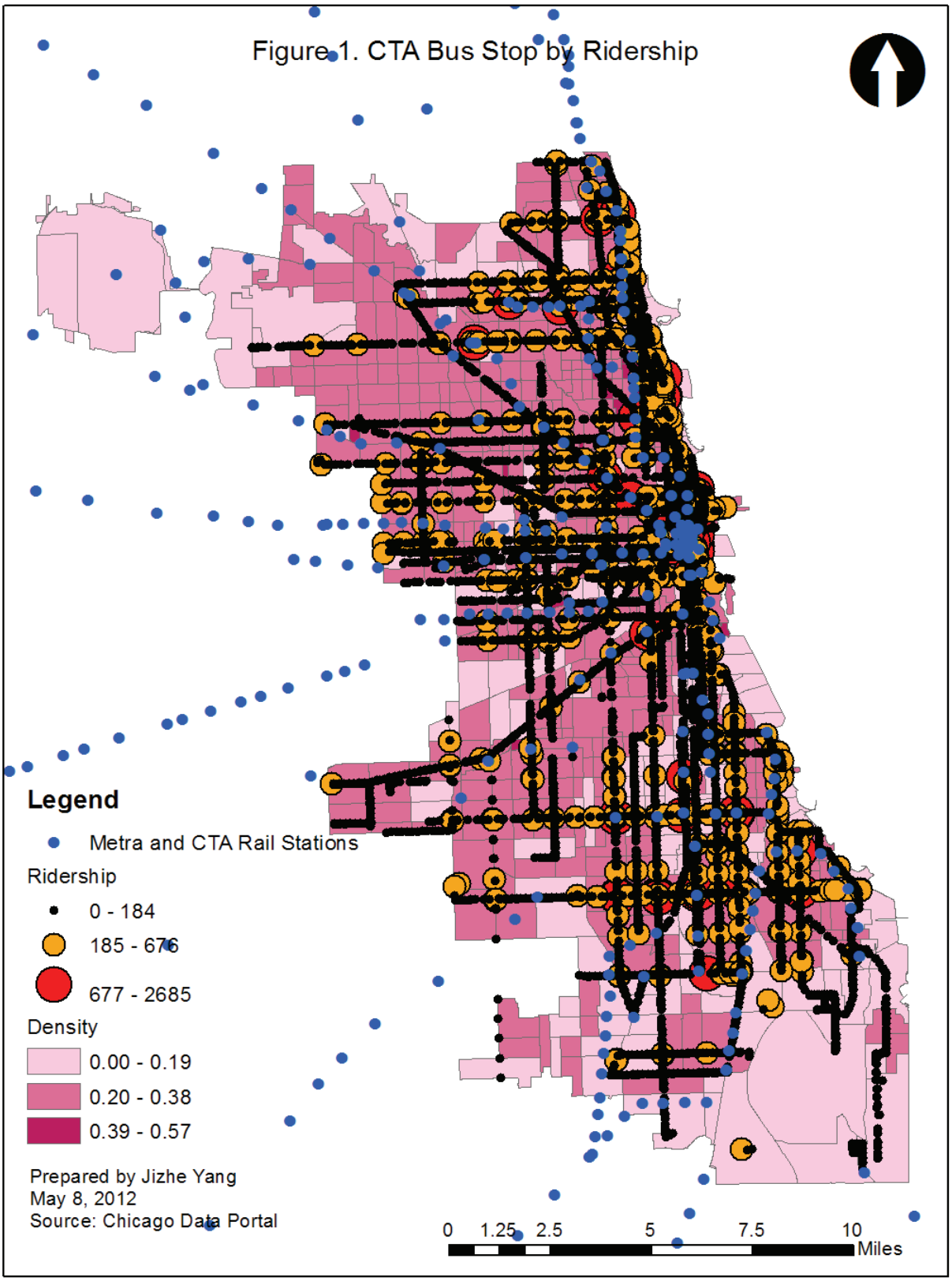


Figure 1. CTA Bus Stop by Ridership
Legend
Metra and CTA Rail Stations
Ridership
0 - 184
185 - 676
677 - 2685
Density
0.00 - 0.19
0.20 - 0.38
0.39 - 0.57
Prepared by Jizhe Yang
May 8, 2012
Source: Chicago Data Portal
0 1.25 2.5 5 7.5 10
Miles

Figure 2. Comparison of Street Views

Stops with High Levels of Ridership

Jackson & Peoria

Jackson & Halsted

Jackson & Clinton

Jackson & Ashland

Stops with Low Levels of Ridership

Jackson & Aberdeen

Jackson & Sangamon

Jackson & Paulina

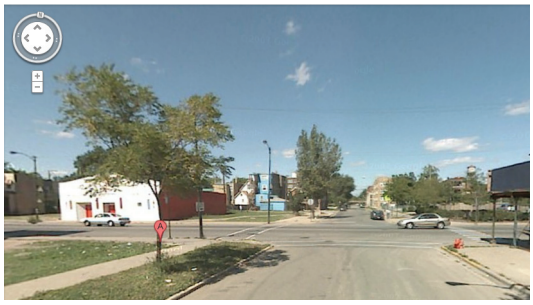
Jackson & Aberdeen

Figure 2. Comparison of Street Views (Continue)

Roosevelt & Pulaski

Roosevelt & Keeler

Roosevelt & Independence



Roosevelt & Karlov

Roosevelt & Homan

Roosevelt & Springfield

Roosevelt & Kedzie

Roosevelt & Lawndale

Figure 2. Comparison of Street Views (Continue)

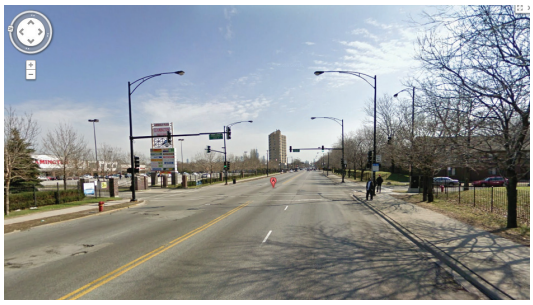
Roosevelt & Racine

Roosevelt & Central Park

Roosevelt & Western

Roosevelt & St. Louis

Roosevelt & Halsted



Roosevelt & Spaulding

Table 7. Street View Analysis

Stops with High Levels of Ridership	Reason for High Levels of Ridership	Visibility	Adjacent Stops with Low Levels of Ridership	Reasons for Low Levels of Ridership	Visibility
Jackson & Peoria	street facing on all four corners, architectural details	sign	Jackson & Aberdeen	vacant lot, no street-facing buildings (has windows but no doors)	sign
Jackson & Halsted	street facing on all four corners, architectural details	sign	Jackson & Sangamon	large parking lots; street facing but vacant business buildings	sign
Jackson & Clinton	street facing on three corners, architectural details, GRAND buildings, extremely wide sidewalks, outdoor seating	sign	Jackson & Paulina	large underused parking lot; street facing but vacant building; large parking structure that generates no interest to pedestrians	sign
Jackson & Ashland	street facing on three corners, well-maintained green space, architectural details	shelter	Jackson & Ogden	one corner with outdoor restaurant seating, but large fenced and underused parking lots on other three corners	sign
Roosevelt & Pulaski	street facing business on two corners, one as vacant lot, one as vacant building	shelter	Roosevelt & Keeler	only one corner has business activity; the rest are vacant lot, parking lot and a deterioriating building under construction	sign
Roosevelt & Independence	well-managed green space on two corners, gas station and restaurant with shelter on the other two corners	shelter	Roosevelt & Karlov	two corners as empty lots; one as parking lot; one with a deterioriating building under construction	sign
Roosevelt & Homan	street-facing buildings on three corners, business on four corners	shelter	Roosevelt & Springfield	vacant lots on two corners; deterioriating and vacant business on two other corners	sign
Roosevelt & Kedzie	street-facing buildings near three corners, residential green space on one corner	shelter	Roosevelt & Lawndale	vacant lots on two corners; street end facing vacant buildings	sign
Roosevelt & Racine	street-facing buildings on two corners; large full parking lot; large vacant lot	shelter	Roosevelt & Central Park	vacant lot on one side; vacant building on the other side	shelter
Roosevelt & Western	street-facing businesses near all four corners	shelter	Roosevelt & St. Louis	parking lot; vacant lot; vacant building; building that is not street-facing	
Roosevelt & Halsted	large buildings near two corners, well-maintained green spaces	shelter	Roosevelt & Spaulding	large underused parking lots on two corners; residential green space on others	sign

statistically significant with a median residual close to 0 and can explain roughly 35% of the variations in ridership using the six independent variables in the model. The model also reveals that density and number of bus routes serving the bus stop have a statistically significant and numerically influential positive impact on increasing ridership. For every one unit of increase in density, the ridership will increase by almost 7 units. For every one more route serving the bus stop, the ridership will increase by roughly 2 people. The presence of real-time information display and Metra or CTA rail station close by can also significantly boost bus ridership. Although the area of nearest open space and poverty rate around the site of bus stop are statistically significant, they only exert minimal influence on ridership (see Table 3).

Street View Analysis

The pictures obtained from Google Street View illustrate several recurring themes of the physical environment around bus stops that are associated with high and low levels of ridership. Although the 22 stations are almost adjacent to one another along Jackson Boulevard and Roosevelt Road, some have contrasting level of ridership. Visibility of the bus stops plays a significant role in increasing ridership, as most of the stops examined with high levels of ridership are equipped with shelter, while all but one stop with low ridership level have only signs (see Table 7). This finding is consistent with the results of regression analysis.

Street view analysis also reveals findings that regression analysis is unable to discover due to data unavailability. Most CTA bus stops are located at streets intersections. By observing and documenting the appearance of the environment around each bus stop, the study finds that stops with high levels of ridership mostly associate with key words such as "street facing on most corners" (which means 3 to 4 the corners of the street intersection are occupied with operating buildings of which the entrance faces the street), "architectural details" on buildings, as well as "well-maintained green space" (which means the lawn is trimmed and the trees are orderly planted). The keyword "street-facing" on most corners renders a sense of enclosure and human activities, and "architectural details" and "well-maintained green space" convey a sense of consistent care and monitor by people and aesthetic values (see Table 7). Conversely, the surroundings of bus stops with low levels of ridership are characterized by "vacant lot", "vacant building", "large parking lots", "underused parking lots" on most corners of street intersection, imposing the perception of lack of human activities, care and security. This indicates that vacancy at the street intersection near a bus stop deters ridership (see Table 7, Figure 2).

Conclusions and Policy Implications

The results of regression analysis helps decision makers identify what urban design elements are essential in promoting CTA bus ridership and what are negligible, so that resources can be allocated effectively. Since this study shows that having neighborhood gardens and open/green spaces close to the bus stop does not matter at all in terms of promoting ridership, investment should not be wasted in related projects.

The study also reveals that building density is the single most influential contributor to promoting ridership (see Table 3 and Figure 1). It is obvious that having more open/green spaces contradicts with enhancing density; therefore, assumptions can be made that people prefer to feel a sense of enclosure rather than openness when waiting at bus stops. Higher building density creates a space adequately bounded by buildings through which people can derive a sense of security. Higher building density also generates more pedestrian activities, as it can trigger higher residential, commercial and employment density, all of which may induce more riders for bus transit system.

The importance of building density discovered by regression analysis is substantiated and calibrated by street view analysis. The majority of CTA bus stops locate at street intersections. The study discovered through street view images show that building density and activity level on the four corners of street intersections should be emphasized to

give bus riders a sense of enclosure, shelter and economic vitality. In other words, if raising the density level is not viable on a large scale, the TOD planning can strategically intensify the land use just on the four corners at the street intersection with street-facing operating buildings, because these four vertices are the focal points that alter human perceptions of whether riding the buses is a safe and welcoming travel option.

The study also identifies a key design element of bus stops through street views analysis: bus shelter. Since bus stops with shelter generally attract greater ridership, bus shelter should be installed for more bus stops so as to increase the visibility of the bus stops and to create a sense of enclosure. Another important design element of bus stops is the real-time information display, which allows riders to make flexible plans and have a sense of assurance by knowing when the buses will arrive.

Finally, the study confirms the importance of facilitating intermodal and cross-route transfers, as the proximity to Metra and CTA rail stations as well as the number of bus routes at one bus stop are positively and strongly associated with the level of ridership. More efforts should be targeted at integrating rail transit system and bus transit system in terms of spatial proximity and compatible scheduling to enhance the performance of TOD.

Limitations and Caveats

The lack of accurate data may have caused the relatively low-moderate goodness of fit. To further optimize the regression model and to identify a more comprehensive list of key urban design factors of TOD, future studies should consider the following suggestions:

- Gather more days of ridership data, both for weekdays and weekends: the ridership by bus stop data is only available online for September 1, 2010. Due to time and resource constraints, building the regression model by relying on the ridership on one single day is precarious and does not fully account for everyday situation.
- Collect data regarding automobiles: Since automobile is a competing alternative to transits, data describing automobile ownership and usage may improve the explanatory power of the regression model. The data may include automobile ownership by block, traffic volume by street, street width, etc.
- Collect a complete and detailed set of spatial data for businesses: the current business data provided by Chicago Data Portal only includes a limited number of businesses, which does not reflect the reality of business distribution near the bus stops. Therefore, this study had to eliminate "number of businesses within walking distance" from the list of independent variables, which may have reduced the goodness of fit of the model. In order to improve the model by accounting for the impact of nearby businesses, a complete list of businesses is required. The businesses can be further classified by type to better inform decisions on what businesses should be located near the transit stations and what should be avoided in order to promote ridership.
- Gather spatial data of vacant buildings and vacant lots: as a common sense, bus stops near such elements would be perceived as unsafe or not welcoming by potential riders. However, formal statistical evidence is needed to be more persuasive when informing policy.
- Take age groups and ADA status into consideration: people of certain age groups and people with special assistances require special considerations for accessibility. By examining these groups' use level of transit, future studies could evaluate how accessible is the current TOD design for these particular groups and identify ways of improvement.

Although the findings of this study best apply to the context of Chicago, the procedures and methods are transferrable to other study areas. The approach of

examining Google Street View or on-site observations is particularly worthwhile for transit researches in other study areas, as pictures can convey much richer contextual information than statistics and regression results. The images, despite prone to subjectivity, trigger intimate and instinctive insights to what design elements are desirable for successful TOD, thus demonstrating the necessity of integrating technical and statistical analysis with site observation.

Bibliography

1. Cervero, R. (2010). Direct Ridership Model of Bus Rapid Transit in Los Angeles County, California. Transportation Research Record. 2145(-1), 1-7.
2. Lin, J. (2008). Does Transit-oriented development affect metro ridership?: Evidence from Taipei, Taiwan. Transportation research record, 2063(-1), 149-158.
3. Sung, H. "Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea." Cities 28.1 (2011):70-82.
4. Data used in this study is acquired from Chicago Data Portal, U.S. Census Bureau and Google Map; Images used in this study are acquired from Google Street View.