DEVELOPMENT OF AN ALTERNATIVE APPROACH TO TRANSIT DEMAND MODELING

BY

ILYA K. CHISTYAKOV

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Urban Planning in Urban Planning
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Master's Committee:

Associate Professor Arnab Chakraborty
Associate Professor Bumsoo Lee

ABSTRACT

Development of transit systems around the nation faces limitations in funding and strict scrutiny of the proposed projects and their potential impact on urban environment. Potential ridership of the proposed transit route becomes one of the key indicators for analysis of investment projects. Transit demand depends on many multifaceted parameters affecting the mode choice of individual commuters. The urban planning as a field faces the demand in creation of a universal model which would allow to estimate transit demand of the areas of different scales and geographies, be simple to interpret and to replicate in any conditions.

The research is discussing the process of development of a model able to predict potential transit demand under provision of a certain level of service based on the socio-economic parameters of the area within walking distance of a transit station. The modeling approach is based on the analysis of real transit ridership of rail stations in Chicago, Los Angeles, and Denver and the parameters possibly contributing to the number of passengers using them. The selection of the variables of the model was based on the most recent research in the field and relied on the multidimensional approach including regional and local scales of socio-economic and transit data. The resulting model included ten independent variables with R^2 of 0.59 with multiple statistical tests confirming the assumptions of the model and statistical significance of the results with some limitations in accuracy of predictions.

The project included creation of a GIS and online mapping tools for deeper analysis of interconnections between built environment and transit demand. The created Transit Demand Index can be used for the analysis of spatial distribution within metropolitan areas to identify the locations where transit investment would have the most significant outcome. The possible applications of this model include preliminary justification of transit projects, small area plans and corridor studies, long-range transportation plans and implementations in Travel Demand Modeling software.

TABLE OF CONTENTS

1. INTRODUCTION

Development of sustainable transportation systems in urban agglomerations is one of the key directions in contemporary urban and regional planning. Rapid transit systems inevitably become the backbones of urban agglomerations and transit-oriented development (TOD) is considered by a number of planners as an alternative way to accommodate future urban growth (Bernick & Cervero, 1996). Analysis and understanding of interconnections that exist between transportation systems and land-use is one of the quintessential objectives of urban planning.

Investments in transit infrastructure in the United States are very limited compared to other parts of the world. Unfortunately, transit-oriented development projects are quite often located in the wrong place and are not based on the existing conditions in certain metropolitan areas. There is a lack of comprehensive and easy tools of transit-oriented development or transit demand level evaluation. Estimation of these levels would significantly support more effective TOD planning, since assessment of the project's feasibility and forecast of the future ridership remain one of the wicked issues faced by urban planners around the world. Since transit ridership is impacted by the significant number of factors, the quantitative analysis and modeling would be the most effective tools to substantiate and justify development of transit systems. Nevertheless, the transit demand modeling remains one of the least reliable directions in travel demand modeling (TDM) with very limited explanatory power.

The research objective of the thesis project is to identify a set of publicly available socio-economic parameters for units of local and regional scales which would provide high level of explanatory power for transit ridership in urban agglomerations. Essential objective is to create a model which would allow to estimate transit demand index or TOD-ness of the areas of different scales and geographies, which could be easily replicable in any metropolitan area.

The first step of the research attempts to conduct the analysis of the existing methods in estimation of transit demand. The fundamental basis of the interconnectedness between built environment and transit was built by a number of researchers studying the process of transit-oriented development as a process of urban redevelopment of the area in proximity of a rapid transit station, accompanied by the increase in population density, increased variety of land uses and economic activities, development of the street design convenient for pedestrians and cyclists, and the fact that these changes in turn increase transit ridership (Bernick & Cervero, 1996; Cervero & Kockelman, 1997). One of the main characteristics of transit-oriented development is that it is a self-accelerating process. This idea justifies a double approach to the transit-oriented development planning: to make the urban environment more transit-oriented and to develop public transportation in the areas with high potential for TOD (Singh, Fard, Zuidgeest, Brussel, & Maarseveen, 2014).

This idea has initiated a number of researches attempting to quantify the connections between transit demand and the built environment using a wide range of approaches, which are discussed in the second chapter of the thesis research. The literature review provides analysis of such approaches to estimations as density thresholds, relative comparison, regression analysis and four-step travel demand modeling. Based on the limitations of the existing approaches, the research would attempt to build a regression model which would be based on publicly and frequently updated data, would be able to scale to different geographies and would be built on multiple metropolitan areas.

The research focuses only on rapid transit systems - based on the classification of Vukan Vuchick the transit systems of categories ROW-A and ROW-B, i.e. systems with partial or complete grade separation, like heavy rail, commuter rail, light rail and bus rapid transit. According to existing research, rapid transit systems due to their higher carrying capacity and operating speeds generate significantly stronger impulses of transit-oriented development than regular buses, thus, these systems have stronger interconnections with the built environment around the public transit stations.

Identification of the set of parameters for the model was based on the theory of "5Ds" of Robert Cervero: density, diversity, design, distance to transit and destination access (Cervero, 2011). The selected parameters can be split in two parts since transit demand is affected by both local parameters of the area within walking distance of the station and regional parameters affecting the entire metropolitan area. Thus, the methodology of the research is based on two consequent parts – regional and local transit demand estimation. The regional analysis part includes creation of the nationwide rapid transit systems database, to identify a set of regional transit and socio-economic parameters affecting the transit demand such as general transit accessibility and coverage or size of the metropolitan area. The collected data was used for selection of the case study regions for local transit demand with the goal to include into analysis metropolitan areas with different regional characteristics attempting to create a universal model, which could be suitable for any regional conditions. Chicago, Los Angeles and Denver were selected to be the metro areas the transit demand model would be built on.

On the next step, the research attempts to identify a set of local built environment and transit parameters affecting transit demand of the areas within walking distance of the transit stations. There was created a geographic database designed to merge transit parameters for the existing transit stations within selected urban agglomerations and the wide range of socio-economic parameters, which can explain transit ridership of these stations. The resulting database is used for the building of model, i.e. for the final selection of parameters with the strongest explanatory power. The resulting model includes a Regional TOD parameter developed during the previous step of the research.

The resulted regression analysis is used for creation of the transit demand index forecast tool, which is discussed in the Results and Limitations chapter. The chapter demonstrates resulted GIS and interactive products for analysis of Transit Demand Index (TDI) as well as provides the review of potential applications of the model in urban and transportation planning.

2. LITERATURE REVIEW[1]

2.1. Transit-oriented development

The contemporary urban and transportation planning has reached the consensus that the further increasing of highways capacity in the metropolitan areas has a very limited impact on the congestion due to the induced travel demand (Jaffe, 2015). Over last decades the Federal Highway Administration and many Departments of Transportation across the nation admitted the impacts of induced demand and have started to change the transportation policy in the largest metropolitan areas shifting investments towards transit and active transportation modes. The main issue faced by the cities is that public transit systems require completely different urban environment for effective operations. The suburban systems solidified by transportation network, Euclidean zoning, density limitations, and parking requirements restrict development of effective transit systems. The average number of passengers per bus in the U.S. in 2009 was equal to 8.8 making it less energy-efficient in terms of MJ per passenger-mile than a private car (Davis, Diegel, & Boundy, 2009). On average, transit operations are characterized by high operating costs and low level of service imposing time costs on passengers, which makes transit a less competitive choice compared to driving.

To provide a higher level of service many metropolitan areas have started to consider rapid transit systems such as bus rapid transit (BRT), light rail transit (LRT) and heavy rail to provide higher speeds of commute and make a trip more comfortable for passengers, thus competitive with a private car. At the turn of the 20th and 21st centuries, the transportation systems of the U.S. cities began to undergo some local changes. New rapid transit systems came into operation in the largest metropolitan areas: the number of such transit systems has doubled over the past 30 years (Chistyakov, 2015b).

At the same time, as a parallel process in the end of 1980s, there was introduced a concept of the New Urbanism and Smart-Growth as an alternative to suburbs urban form with the design following the principles of walkability, connectivity, mixed-use and diversity, and higher density (Katz, 1993; "Urbanism Principles," n.d.). The principles behind New Urbanism were supported by many planners, however they remained mainly a design concept with very limited economic justification to become an alternative to suburbanization. Nevertheless, developed first as a design concept, New Urbanism combined with the

[1] The chapter contains some of the materials of Economics for Planners UP 509 term paper on Evaluation of transit-oriented development in the US agglomerations (Chistyakov, 2015a)

growth of transit systems around the nation became a basis for the concept of transit-oriented development (TOD).

For the first time the concept of TOD was introduced by Peter Calthorpe, the founding member of the Congress for New Urbanism, defining it as “moderate and high-density housing, along with complementary public uses, jobs, retail and services, concentrated in mixed-use developments at strategic points along the regional transit systems” (1993). Soon enough, the further research of Michael Bernick and Richard Cervero studied areas around the stations in the San Francisco - Oakland metropolitan area and the impact of the rapid transit system BART (1996). They have identified and described the process of urban redevelopment of the area in proximity to a rapid transit station, accompanied by the increase in population density, increased variety of land uses and economic activities, the development of the urban landscape and design convenient for pedestrians and cyclists (Bernick & Cervero, 1996).

In the later research, Robert Cervero identified and described the interconnections between land use, built environment and transit demand based on the case studies in the U.S. and Latin America. In 1997, the characteristics of the transit-oriented development were summarized as “3Ds” of density, diversity and design and later were extended to “5Ds” with additional distance to transit and destination access (Cervero & Kockelman, 1997; Cervero, Sarmiento, Jacoby, Gomez, & Neiman, 2009). Based on the analysis of research on TOD around the nation, Cervero has proven increased elasticity of transit ridership induced by the “5Ds” (see Table 1).

Dimension	*Metric*	*# Studies*	*Elasticity*
Density	Population Density	10	0.07
	Job Density	6	0.01
Diversity	Land Use Mix (0-1)	6	0.12
Design	Intersections/Street Density	4	0.23
	Connectivity (4-way inter.)	5	0.21
Distance to Transit	Distance	3	0.29

Table 1: Elasticities from regressions and logits (Ewing & Cervero, 2010)

One of the key characteristics of transit-oriented development is that it is a self-accelerating process. Transit accessibility and reliability starts the processes of redevelopment the space. Development of sustainable urban systems requires not only transformation of the existing urban environments in accordance with TOD principles but also bringing transit supply to areas where there is not enough transit connectivity but where physical characteristics correspond to TOD standards and have a potential for induced growth of transit demand. This idea justifies a double approach to the transit-oriented development planning: to make the urban environment more transit-oriented and to develop public transit in the areas with high potential for TOD. (Singh et al., 2014)

The process of investments in transit in the United States is complicated by a significant counterweight of highway lobby. Any transit development project has to compete for the federal or state funds and go through a strict scrutiny process. Basically, transit planners do not have a right for a mistake, however the success of transit is complicated by many factors. Analysis of this factors based on the existing conditions and the potential development is crucially important for allocation of transit investments and their justification. Based on the continuing expansion of rapid transit systems across the nation there is demand for comprehensive tools of transit demand level evaluation to support more effective development of transit systems and TOD planning. The concept of transit-oriented development and 5Ds became a fundamental basis for this research and a number of techniques described in the following chapter.

2.2. ANALYSIS OF EXISTING METHODS

The main goal of this chapter is to analyze existing techniques of transit demand estimation. Over last 20 years many researchers have dedicated their work to development of different approaches to TOD evaluation. Based on the analysis of the existing research there were identified four types of the methods:

A. *Density Thresholds*.

Density threshold are used to forecast required level of transit supply based on the density of population, jobs or housing units. Since the density of population and employment is one of the main drivers of transit demand, this simple technique is useful for row estimation of transit demand with very limited data available. For example, this method is being used in a number of guidelines of DOTs ("TRB," 2014) and in work of private consulting companies like Nelson/Nygaard Consulting Associates – the example can be found in Table 2 (Newmark, 2012). The main limitation of this approach is simplification of the complexity of interrelations between built environment and land use – higher density is not the only driving force of transit demand.

Class	Corridor Service Type	Key Service Features	Density Threshold (persons + jobs per acre)
A+	High Capacity Transit	Local bus and LRT/BRT operation at high service level. 5 minutes peak, 10 minutes off-peak combined frequency minimum.	60 or more
A	Rapid Transit	Local bus and limited-stop operations at high service level. 10 minutes peak, 15 minutes off-peak combined frequency minimum.	40 to 60
B	Arterial Transit	Local bus operations at enhanced service level. 15 minutes peak, 30 minutes off-peak minimums.	20 to 40
C	Local Transit	Local bus and shuttle operations at baseline levels. 30 minutes peak, 60 minutes off-peak minimums.	10 to 20
D	Shuttle Transit	Employee shuttles and commuter bus services. Circulators and flex-route operations.	5 to 10
E	Demand Response	Ridesharing (carpool, vanpool, bus pool or van share). Demand response services.	5 or less

Table 2: Nelson/Nygaard Consulting Associates threshold system for transit demand estimation (Newmark, 2012)

B. Relative Comparison

Relative comparison is similar in its approach to density thresholds but it takes into consideration more variables relating to the built environment, land use and socio-economic parameters of the area. For example, Evans and Pratt selected and classified a set of statistical indicators “that might likely make up such a TOD index” (Evans IV, Pratt, Stryker, & Kuzmyak, 2007) A more comprehensive quantitative approach was introduced by research group at University of Twente, the Netherlands. They developed a TOD index based on the 5Ds, which measures a set of geographical indicators analyzed under the Spatial Multiple Criteria Analysis (SMCA) platform. SMCA is a complex model of geographical data analysis allowing the user to create weights for different criteria. (Singh et al., 2014)

This approach provides flexibility in selection of the parameters and prioritizing certain variables in their impact on the TODness of the area (see Table 3). One of the limitations of this approach is that it is difficult to justify the resulting weights of the model and to estimate and substantiate the impact of selected variables on transit demand of the area (see Table 4).

Criteria for measuring Potential TOD Index.

Potential TOD Index	
Criteria	**Indicators**
What are the various densities?	Residential density Employment density Commercial intensity/density
How diverse is the land use?	Land use diversity
Does the design of urban space encourage walking and cycling?	Level of mixed-ness of land uses w.r.t residential land use Quality and suitability of streetscape for walking Quality and suitability of streetscape for cycling Density of controlled intersections/ street crossings
What is the current level of economic development?	Private investment in the area Number of business establishments Tax earnings of municipality Unemployment levels

Table 3: Criteria for measuring Potential TOD Index based on Spatial Multiple Criteria Analysis platform developed by University of Twente research group (Singh et al., 2014)

Criterion weights and standardisation table.

Criteria	Rank order	Resulting weights	Indicator	Contribution to criterion's weight
Level of density	1	0.35	Residential density Commercial density	50% 50%
Level of land use diversity	1	0.35	Land use diversity	100%
Level of mixed use	2	0.20	Mixed use	100%
Level of economic development	3	0.10	Number of business establishments	100%

Table 4: Criterion weight and standardization table developed by University of Twente research group (Singh et al., 2014)

C. Multiple Regression

The logical development of the previous approach is implementation of statistical methods to justify and estimate the impact of built environment parameters on transit ridership. By far, this method has proven to be the most popular with dozens of case studies around the nation. Reid Ewing and Robert Cervero in their 2010 article "Travel and Built Environment" have selected 50 case studies around the nation that relate to statistical analysis of interconnections between transit demand and built environment (Ewing & Cervero, 2010). A number of metropolitan planning organizations created in-house travel demand forecast model based on regression analyses (DVRPC, 2007; Newmark, 2012). The regression analysis allows to create coefficients for independent variables and forecast transit demand with relatively small statistical errors. Nevertheless, there are several limitations to the existing methods:

- *Availability of data*. Every research implements a different set of independent variables to forecast transit ridership. Some of the independent variables have higher explanatory power in certain metropolitan areas compared to other. At the same time, several researches tend to use variables which are not available in other metropolitan areas, most often connected with design features and land-use characteristics. It makes the approaches difficult to replicate in a different metropolitan area and to compare the outcomes. Some of the selected variables are not updated as frequently as others, thus imposing a limit on the updates of the model. Another difficulty is that the researches use different dependable variables for regression models: American Community Survey data for a number or share of commuters taking public transportation to work (commuting data), the transit forecasts of travel demand models, and the boarding and alighting counts per station, among others.
- *Scale of research and units*. The studies differ not only by metropolitan area but also by a scale – from a neighborhood or a few selected stations to the entire metropolitan area or state. As a unit of research, the studies might rely on traffic analysis zones (TAZs), Census geographic units (Census Tracts, Census Block Groups, etc.) or even grids of squares or hexagons, which makes data more difficult to compare.
- *Regression model approach*. Finally, since the set of independent variables and their explanatory power varies per study, the researchers use different methods for regression analysis and controls, ultimately receiving different results for the model.

More limitations of the regression analysis will be discussed in Chapter 4.

D. *Travel Demand Modeling*

Four-step Travel Demand Modeling approach is an essential tool for analysis of regional transportation systems. Even though travel demand models can predict traffic flows with relative accuracy, it is much more difficult to forecast transit ridership. The Mode Choice of the four-step travel demand modeling process includes comparison of trip costs for the different modes of transportation. The costs estimation for transit usually includes the time costs, fare box costs, park & ride availability and costs, first and last miles costs estimations.

Robert Cervero in his discussion on limitations of TDM refers to the fact that they were never meant to be used to analyze the interzonal trips, such as trips within area around transit stations (Cervero, 2006). Basically, the TDM considers the trips between centroids of TAZs, however, it is more difficult to predict what is the share of trips within the neighborhood, which is directly connected with elasticities of travel demand discussed by Robert Cervero, i.e. there is a higher probability that a commuter would choose transit instead of a car if the design of the area follows the TOD principles.

Another limitation of travel demand models is that they require implementation of proprietary software with significant amount of input data, which can be collected and maintained by specially dedicated staff. In addition to that, it might be difficult to explain outcomes of the travel demand model to the public, city officials or developers.

2.3. Limitations of the existing methods

The literature review has demonstrated that there is a significant demand from planners and developers in the United States to develop a comprehensive TOD index tool. Even though there was accumulated a significant research basis in this field with a number of approaches to transit demand forecasting discussed in the previous section, there are various limitations to the application of existing models. First, the lack of standardized and frequently updated set of land use, socio-economic parameters, and transit parameters. Second, limited applicability and replicability of existing techniques to different scales and geographies. Finally, lack of the models including the regional or transit operating parameters affecting transit ridership.

Based on the above-mentioned limitations, the research will attempt to create a new approach to estimation of the transit demand based on the criteria that it would rely on publicly available and frequently updated socio-economic, land-use and transit parameters, it would scale to different geographies and applied to metropolitan areas with different potential of transit development, finally, it would be simple to apply and interpret. The regression analysis is considered as the most appropriate approach for the stated objective.

The outcomes and identified interconnections between built environment and transit demand might be implemented in four-step Travel Demand Modeling. Regardless the limitations of the four-step TDM, it remains the most comprehensive and wide-spread tool for analysis of urban transportation systems, trip generation and comparison between different modes of transportation. The accuracy of transit forecasts in TDM software could be enhanced based on the inclusion of transit demand elasticities, which are estimated in this research.

3. METHODOLOGY

Transit demand, i.e. the probability of commuters in a certain area to use transit, depends on the significant number of factors, which can be split in two major categories:

1. Regional factors – characteristics of the metropolitan area and overall transit system affecting the mode choice of every commuter in the urban agglomeration;
2. Local factors – characteristics of transit and the built environment within the walking distance of the station affecting mode choice of local residents, employees and visitors.

Based on this assumption, the proposed transit demand model would consist of three parts, which will be described in the following sections of the chapter.

3.1. Regional factors in transit demand estimation

Regional factors would include a group of parameters which would have a relatively equal impact on the mode choice in all parts of the metropolitan area. The first step of the research attempts to conduct the analysis of metropolitan areas across the nation, to create a nationwide database of rapid transit systems in the United States and the regional parameters impacting the transit demand.

3.1.1. Regional transit system characteristics

One of the most critical parameters in estimation of potential transit demand for the proposed infrastructure is state of the existing transit system in the metropolitan area. Several researches have confirmed that rapid transit systems affect the mode choice of a commuter to a significantly larger degree than regular bus system (Bernick & Cervero, 1996; Cervero & Kockelman, 1997). Rapid transit system can provide higher level of service, competitive speeds of travel and frequencies of operation. Significant TOD impulses generated by a public transport station are explained by operational specifics of mass transit systems: significant distance between stations in the residential area and technical capabilities to carry significant ridership (see Figure 1).

Technology		Commercial Speed (mi/hr)
Minibuses, taxi, mixed traffic		12 – 20
Buses in mixed traffic		10 – 12
Buses in exclusive lane		15 – 18
Streetcars in mixed traffic		10 – 12
Buses in BRT w/passing		15 – 30
Light Rail in exclusive lane		15 – 30
Metro	**Elevated**	30 – 40
	Underground	30 – 40

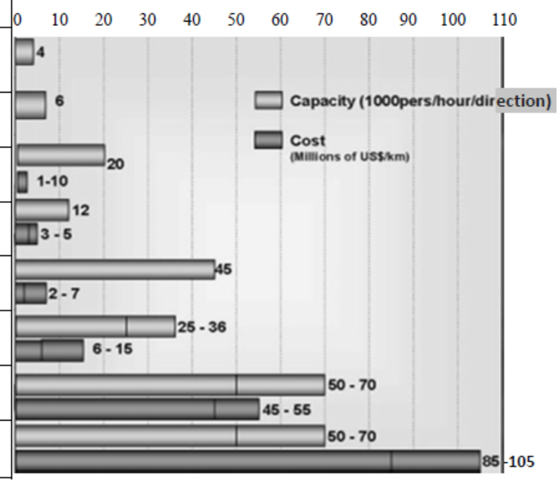

Figure 1: Approximate Commercial Speeds, Costs and Capacity for different modes of transit (Ouyang, 2017)

To define the scope of modes for further analysis, this research relies on the urban transit systems classification created by Vukan Vuchik, which is based on right-of-way category (ROW). According to the classification, there are following categories of transit (Vuchic, 2005):

ROW-C - no separation, i.e. surface street with mixed traffic (regular buses);

ROW-B - partial separation, i.e. longitudinally physically separated, with at-grade crossings (e.g. LRT, BRT);

ROW-A - full separation and separate alignment (for example, subway, elevated rail, commuter rail).

Based on this classification and advantages of rapid transit systems, it is possible to limit the range of urban agglomerations where this type of development is possible. The further research will mainly focus on categories A and B, i.e. heavy rail, commuter rail, light rail and bus rapid transit systems, which will be referred as rapid or mass transit systems in the further research.

On the next step of the research, there was created a database of all rapid transit systems existing in the U.S. (see Appendix A). Currently, there is no publicly available database of transit systems in the U.S. including ridership, operational and financial parameters. The Federal Transit Administration of the U.S. DOT maintains the National Transit Database, which includes frequently updated ridership and financial data, but does not include such operational characteristics like year of opening, length and number of routes, and number of stations, which are crucially important for comparative analysis of transit networks ("NTD," 2015). The missing data was collected using various sources like transit agencies websites and Wikipedia, which has proven to be a rather reliable source of information for this type of data ("List of United States commuter rail systems by ridership," 2017, "List of United States light rail systems by ridership," 2017, "List of United States rapid transit systems by ridership," 2017).

Since the bus rapid transit (BRT) systems are more complicated to differentiate from regular buses or express buses and there are many transitional BRTs with only some of the design elements implemented, it was decided to leave them out of scope of this research. In addition to that, the ridership of BRT lines is far lower than on rail transit, thus, it will not significantly affect the outcomes of the research.

Collection of this data allowed to calculate average spacing (distance between stations) and passengers per mile by each individual transit system (see Appendix A). The transit systems were aggregated by metropolitan areas they serve. Since some of the commuter rails serve multiple metropolitan areas, they were aggregated by the primarily served area, for example, even though Metro North commuter railway serves New Haven, it is associated only with New York metro, where most of its operations are concentrated.

Thus, there were identified 83 rail rapid transit systems in the U.S. operating in 37 metropolitan areas, allowing to compare transit systems and metropolitan areas and to estimate transit accessibility and connectivity within metropolitan areas.

3.1.2. Regional socio-economic parameters

Transit system characteristics have to be analyzed in the complex with regional socio-economic parameters:

- *Total population of the metropolitan area*. The larger metropolitan areas have more potential to have complex transportation systems. At the same time, average population density of metropolitan areas does not necessarily have direct correlation with transit demand, which should be considered on the local scale (Walker, 2010).

- *Population growth over last five years*. Significant population growth over last years could direct new development into densification of areas around the transit stations if we assume implementation of sustainable planning policies and significant pressure from real estate developers.
- *Share of jobs within Central Business District (3 miles radius).* This parameter indicates the employment density and concentration within downtown areas of urban agglomerations relative to suburban employment. Several researchers have confirmed correlation between transit demand potential and success of transit systems with higher shares of jobs within CBDs, which is explained by the fact that absolute majority of transit network in the U.S. primarily serve commuters between a CBD and residential areas. Another aspect of that is supply and cost of parking within CBD – the denser the employment, the more expensive it is to provide parking (Brown & Neog, 2012; Kneebone, 2013).
- There are many other parameters not included in the scope of this research such as traffic congestion rate (percent of peak VMTs), transit fares, parking costs in CBD, quality of service and climate factors, among others (Taylor & Fink, 2003).

On the next step, the data by transit system was aggregated by metropolitan area (see Appendix B and C). There was calculated a ratio between the total weekday ridership of rapid transit and the total population of the metropolitan area. This parameter is considered as having the highest explanatory power on regional scale, basically representing the service coverage and transit connectivity within metropolitan area. A particular transit station will have significantly larger transit demand if it provides access to existing transit network. At the same time, the impact might be significantly limited if a rapid transit system has to be developed from scratch.

The collected parameters were used for selection of three metropolitan areas for creation of the transit demand model on local scale. The choice of the metro areas was based on the goal to compare transit systems in different regional conditions and different timespan of development. Since the analysis on local scale is conducted by station and with creation of regression model, the sample of stations has to be statistically significant. Based on this criteria, *Chicago, Los Angeles* and *Denver* metropolitan areas were selected for analysis of travel demand on local scale. The selected metro areas represent different sizes, different growth pattern and history of transit development. For example, Chicago has an L system for more than a hundred years with developing going on around the stations. On the opposite end, Denver has opened its first rapid transit system in 1993 (see Appendixes A, B and C).

3.2. Local factors in transit demand estimation

The main objective of this chapter is to identify a set of publicly available socio-economic, land-use and transit parameters on local scale for the further creation of the transit demand model. This step of the research is based on creation of geographic database in ArcMap software with the main goal to intersect local transit data and socio-economic parameters under a single geographic unit – the half-mile circle with centroid in public transit station. A group of researchers at UC Berkeley have proven that it is a reasonable catchment area of transit-oriented development impulses of transit station since it is an average distance people are willing to walk (Guerra, Cervero, & Tischler, 2012).

3.2.1. Local transit characteristics

The main data source for local transit characteristics used in the research is the General Transit Feed Specification (GTFS), a most common format for geographic information on routes, stops and schedules for transit systems developed by Google ("GTFS Static Overview | Static Transit," n.d.). The GTFS files are available for most of the transit systems around the world and can be uploaded through the TransitFeeds portal or directly through transit agencies ("TransitFeeds," 2017). The GTFS files contain the data allowing to calculate several operating parameters by station based on the schedule of routes, such as frequency of trains or buses per weekday, average headway per route or maximum wait time. The BetterBusBuffers tool was used to transform the GTFS files for individual transit systems into shapefiles of transit stops with calculated operating parameters (Morang & Wasserman, n.d.). Weekday transit ridership per station is going to be used as a dependent variable in the regression model. The parameter is calculated as an average between number of boardings and alightings at the station. Unfortunately, this data is not publicly available for most of the metropolitan areas. CTA in Chicago and RTD in Denver have this data available online ("RTAMS," 2016, "RTD," 2017), however the data for Metro Rail in Los Angeles had to be requested directly through the transit agency.

Based on the collected data on the later stage of the research, it was identified that such parameters as transit mode and frequency have the strongest explanatory power for estimation of transit demand. Since there is a significant variance of frequencies among transit stations and difficulties to forecast accurate frequency of the future transit line in the model, the parameter was transformed into level of service classes using the following thresholds:

Frequency	48	96	144	192	288	360	480
LOS	1	2	3	4	5	6	7

Table 5: Level of service (LOS) classes based on frequency of trains in both directions

The main criteria for creation of this thresholds is that on average people stop taking into consideration timetable of transit if the headway is equal to five or less minutes. It is achieved with average of 480 trains per day in both directions, which was considered as maximum level of service by frequency ("TRB," 2014). Another aspect is the mode of transit, since BRT, LRT and heavy rail have different commercial speeds and general comfort of travel for passengers. It was assumed that stations with heavy rail service would receive additional coefficient of 2 for the LOS, LRT would add 1, and BRT would add 0. Thus, the resulting LOS parameter explains both frequency and quality of transit service provided at a particular station within the range of 0 to 9 – sum of frequency and mode components.

Another critical parameter of transit network impacting transit ridership is a terminal or transfer hub status. On average the terminal stations have significantly higher ridership since they serve passengers from a larger area, often being a hub for feeder buses or having a large park and ride (P&R) facility. The transfer stations have a significantly larger ridership because of the passengers transfer from one route to another. Based on transit routes network, the terminal/transfer parameter was included in the model as a dummy variable.

There are several transit parameters affecting transit demand forecast which were not included in the scope of the research, such as availability and size of P&R facility, driving accessibility or a number and ridership of feeder buses. Based on that, it was decided to exclude the commuter rail systems from analysis on local scale since more than 90% of passengers drive or carpool to the station, thus the built environment around the station has a limited impact on these passengers.

Thus, three transit parameters are included in the further model building: weekday ridership (dependent variable), level of service and terminal/transfer status.

3.2.2. Local socio-economic parameters

The selection of the parameters was based on the literature review, i.e. the parameters which have demonstrated strong connection with transit demand in previous research. Since the criteria for this research is replicability of the research, availability of data and its frequent updates, there were used only datasets provided by the U.S. Census Bureau (see Table 6).

#	PARAMETER	SOURCE*	DATA CODE IN THE SOURCE
1	Total Population	ACS	B01001e1
2	Population under 18	ACS	Sum of B01001e3, B01001e4, B01001e5, B01001e6, B01001e27, B01001e28, B01001e29, B01001e30
3	Total number of commuters	ACS	B08134e1
4	Number of commuters taking transit	ACS	B08134e101
5	Number of commuters walking to work	ACS	B08134e61
6	Total number of housing units	ACS	B25001e1
7	Number of multifamily housing units (10 or more units in structure)	ACS	Sum of B25024e7, B25024e8, B25024e9
8	Median household income	ACS	B19013e1
9	Total Employment	LEHD	C000
10	Number of jobs in Retail Trade	LEHD	CNS07
11	Number of jobs in Arts, Entertainment, and Recreation	LEHD	CNS17
12	Number of jobs in Accommodation and Food Services	LEHD	CNS18

Table 6: The selection of initial socio-economic parameters for the model. *ACS: American Community Survey 5-Year Estimates — Geodatabase Format 2011-2015 (U.S. Census Bureau, 2015); LEHD: Longitudinal Employer-Household Dynamics Origin-Destination Employment Statistics (LODES) Workplace Area Characteristics (WAC) JT00 2014 (U.S. Census Bureau Center for Economic Studies, 2014)

The Table 6 represents a narrowed down list of preliminary pre-transformation parameters which were selected on the model building step of the research. Some of these parameters are needed solely for transformation of others, like total number of commuters is required to calculate a share of people taking transit or walking to work. On the preliminary phase of the research there were uploaded 40 data categories, which were used to create 69 transformed parameters (sums, shares, logarithms and interactions of the parameters), which were narrowed down on the further steps of the research. The parameters not included in the final model are population above 65, share of no-car households, and car ownership rates, among others as having lower explanatory power.

The abovementioned data was collected for three metropolitan areas based on the Census Block Group (BGs), the smallest geographical unit for which the U.S. Census Bureau publishes sample data. The BGs combine blocks with relatively similar socio-economic pattern and with average total population of 600 to 3000 people (U.S. Census Bureau, 2010). The data was uploaded by respective states using TIGER Selected Demographic and Economic Data Geodatabase files (U.S. Census Bureau, 2015).

On the following stage of the research the socio-economic data by block groups is being aggregated by 0.785 mi^2 circles with transit stations as centroids representing the transit data parameters. On the first step, the 0.5-mile-radius buffers are created for every transit station. The tool “Intersect” in ArcGIS is used to create a layer collecting both socio-economic and transit data. Since some of the block groups are being split by the transit circles, the socio-economic parameters are calculated as a proportion of the block group area within the circle. In this case, we assume that distribution of the parameters within block groups is uniform. On the following step, the block group parts inside the buffers are being aggregated into a single geographic unit containing a sum of all socio-economic parameters and respective transit parameters. Since the methodology involves a number of analogous data manipulations, there was created a Python script allowing to automate the process of data collection.

As a result, the selected data was aggregated for 142 CTA rail stations in Chicago, 94 Metro Rail stations in Los Angeles and 52 RTD rail stations for Denver (280 stations in total) providing a significant sample for creation of the model based on the regression analysis.

3.3. MODEL BUILDING

Creation of the regression model to predict transit ridership was complicated by a number of issues with the created database. First, most of the selected independent variables as well as transit ridership are distributed nonlinearly and have many outliers. Second, many of the selected parameters have strong correlation between each other, which increases the impacts of multicollinearity on the model. Finally, it is difficult to track some of the interactions between parameters, which might lead to changes of the β-coefficient parameter or even to the opposite sign of β-coefficient compared to the correlation. The further analysis and model building was performed in R Studio software (R Core Team, 2016; Wickham, Hester, & Francois, 2016).

To avoid these issues, the selected variables went through a number of transformations like ratios, logarithms and interactions of variables (like product of a ratio and an absolute number). The transformations were performed in accordance with analysis of distribution plots to achieve more linear distribution of the variables with minimal number of outliers (see Appendix D). The process of model building involved comparison of dozens of sets of variables as well as different methods of regression

	CODE	SELECTED VARIABLE DESCRIPTION
1	PE	Total population and employment
2	U18	Natural logarithm of population under 18
3	TR	Share of commuters taking transit to work
4	WL	Product of share of commuters walking to work and their absolute number
5	MF	Product of share of multifamily housing and total number of multifamily units (10 or more units in structure)
6	SE	Natural logarithm of combined employment in retail, entertainment, accommodation and food industries (service)
7	MHI	Ratio of median household income to the median of the metro area
8	LOS	Level of service
9	TT	Terminal/transfer status (0/1)
10	REG	Regional TOD: ratio of total rapid transit ridership to total population of the metropolitan area
D	RID	Natural logarithm of weekday transit ridership at the station (dependent variable)

Table 7: Selected variables for the regression model

Residuals:

Min	1Q	Median	3Q	Max
-1.52	-0.42	-0.03	0.37	1.59

Coefficients:

<table>
<tr><th></th><th>β Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(>|t|)</th><th></th></tr>
<tr><td>(Intercept)</td><td>5.053</td><td>0.39</td><td>12.64</td><td><2e-16</td><td>***</td></tr>
<tr><td>PE</td><td>1.03E-06</td><td>0.00</td><td>0.74</td><td>0.46</td><td></td></tr>
<tr><td>U18</td><td>-0.015</td><td>0.04</td><td>-0.46</td><td>0.64</td><td></td></tr>
<tr><td>TR</td><td>1.009</td><td>0.47</td><td>2.15</td><td>0.03</td><td>*</td></tr>
<tr><td>WL</td><td>3.00E-05</td><td>0.00</td><td>0.15</td><td>0.88</td><td></td></tr>
<tr><td>MF</td><td>2.00E-05</td><td>0.00</td><td>0.66</td><td>0.51</td><td></td></tr>
<tr><td>SE</td><td>0.158</td><td>0.04</td><td>5.78</td><td>0.00</td><td>***</td></tr>
<tr><td>MHI</td><td>-0.263</td><td>0.11</td><td>-2.94</td><td>0.00</td><td>**</td></tr>
<tr><td>LOS</td><td>0.164</td><td>0.03</td><td>6.37</td><td>0.00</td><td>***</td></tr>
<tr><td>TT</td><td>0.793</td><td>0.09</td><td>8.79</td><td><2e-16</td><td>***</td></tr>
<tr><td>REG</td><td>3.252</td><td>1.57</td><td>2.38</td><td>0.02</td><td>*</td></tr>
</table>

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5831 on 279 degrees of freedom

Multiple R-squared: 0.6037, Adjusted R-squared: **0.5895**

F-statistic: 42.5 on 10 and 279 DF, p-value: < 2.2e-16

Table 8: Regression model parameters (Ridge regression). R Studio with package “mass” (R Core Team, 2016; Venables & Ripley, 2002).

analysis: stepwise regression (backwards and forward), principal components analysis (PCA), and generalized additive models (GAM), among others.

As a result of the model building, there were selected ten independent variables (see Table 7). The Ridge regression was considered as the most appropriate approach for the model to decrease multicollinearity of the data as well as to normalize the β-coefficients among independent variables. The outcomes of the regression analysis are demonstrated in Table 8. The building of the model was based on analysis of the statistical analysis of the regression and the database:

1. Multicollinearity checks:

   One of the issues affecting the model is that a number of selected parameters are interconnected among each other even though they affect transit ridership and contribute to the model in a different way. To deal with multicollinearity there were used various transformations. The preliminary check was based on the correlation analysis of the selected transformed variables (see Table 9).

	PE	U18	TR	WL	MF	SE	MHI	LOS	REG	TT	RID
PE	1.00	-0.17	0.04	0.77	0.53	0.65	0.38	0.32	0.20	0.13	0.46
U18	-0.17	1.00	0.50	-0.20	0.03	-0.19	-0.19	0.21	0.23	-0.10	0.06
TR	0.04	0.50	1.00	-0.00	0.19	0.01	0.06	0.34	0.71	-0.14	0.32
WL	0.77	-0.20	-0.00	1.00	0.67	0.61	0.38	0.26	0.25	0.06	0.40
MF	0.53	0.03	0.19	0.67	1.00	0.65	0.19	0.32	0.11	0.08	0.47
SE	0.65	-0.19	0.01	0.61	0.65	1.00	0.49	0.37	-0.01	0.09	0.53
MHI	0.38	-0.19	0.06	0.38	0.19	0.49	1.00	0.34	0.27	0.10	0.26
LOS	0.32	0.21	0.34	0.26	0.32	0.37	0.34	1.00	0.31	0.10	0.55
REG	0.20	0.23	0.71	0.25	0.11	-0.01	0.27	0.31	1.00	-0.07	0.31
TT	0.13	-0.10	-0.14	0.06	0.08	0.09	0.10	0.10	-0.07	1.00	0.37
RID	0.46	0.06	0.32	0.40	0.47	0.53	0.26	0.55	0.31	0.37	1.00

Table 9: Pairwise correlation analysis of the selected variables (R Core Team, 2016)

   The analysis of pairwise correlation coefficients proves that the selected variables do not have any strong correlations among each other with the highest of 0.77 between population-employment variable and walking share estimate. This confirms statistical independence of the data.

   Multicollinearity has been analyzed based on the calculation of the Condition number (kappa) and variance inflation factor (VIF) using "car" package in R (Fox & Weisberg, 2011; R Core Team, 2016). Kappa value for the selected dataset was equal to 35.67, while the "rule of thumb" is that any value below 100 is appropriate for the model. The mean VIF value was equal to 2.63, while anything below 4 is considered to be acceptable for the regression model. Thus, the created model addresses the multicollinearity issues of the initial dataset.

2. Normal distribution checks:

   Even though a number of transformations with the initial data were performed, the nature of the data is that is has significant variance and multiple outliers making it impossible to bring it to the normal distribution. This is demonstrated by analysis of skewness and kurtosis statistical parameters in Table 10 using package "moments" in R. The "rule of thumb" for both parameters is that the values should be below 3. However, we can see that some of the variables have non-linear distribution, like population-employment (which was not transformed), walking estimate and multifamily estimate.

	PE	U18	TR	WL	MF	SE	MHI	LOS	TT	REG	RID
SKEWNESS	3.45	-1.29	0.58	3.53	2.76	0.27	0.72	-0.91	0.03	1.77	0.04
KURTOSIS	14.92	6.51	2.46	16.13	12.70	2.54	2.58	3.66	1.02	4.12	2.69

Table 10: Skewness and kurtosis estimates (Komsta & Novomestky, 2015; R Core Team, 2016)

   Nevertheless, the normal distribution of variables is not required for creation of a reliable model. It is more important to analyze the distribution of the standardized residuals for the model (Figure 2).

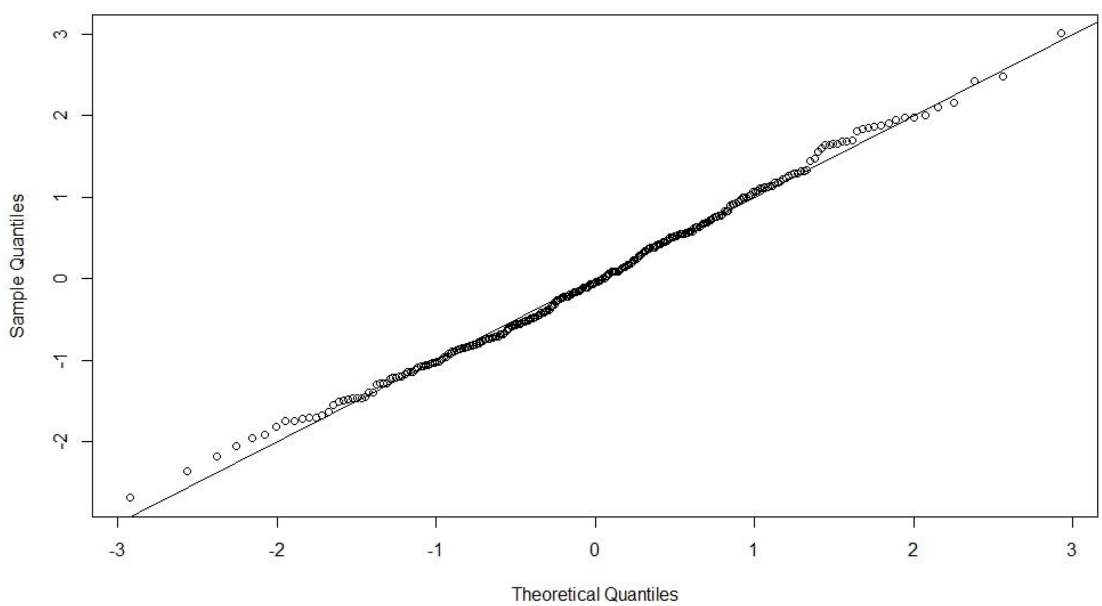

Figure 2: Normal probability plot: Quantile - Quantile (Q-Q) plot (R Core Team, 2016)

Normal probability plot demonstrates close to linear distribution of the standardized residuals on the Figure 2 with only few outliers. An additional analysis of standardized residuals distribution in Figure 3 proves the assumption.

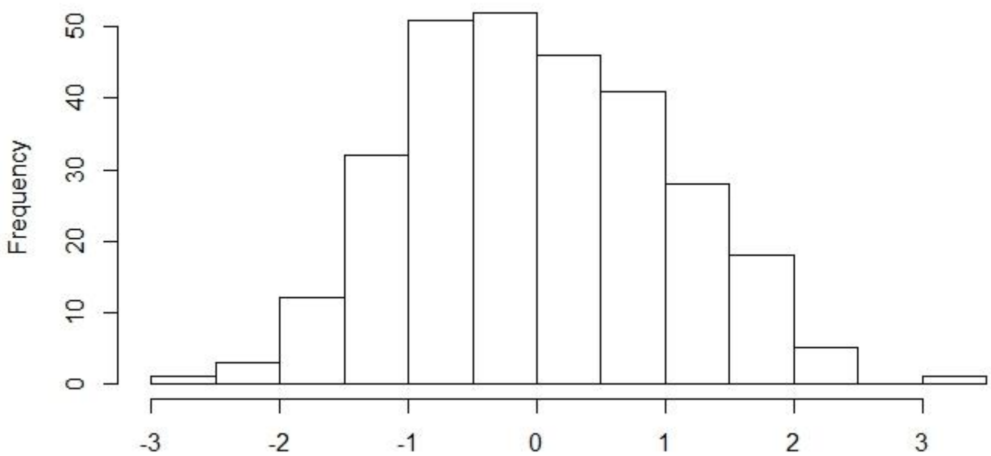

Figure 3: Distribution of standardized residuals for the model (R Core Team, 2016)

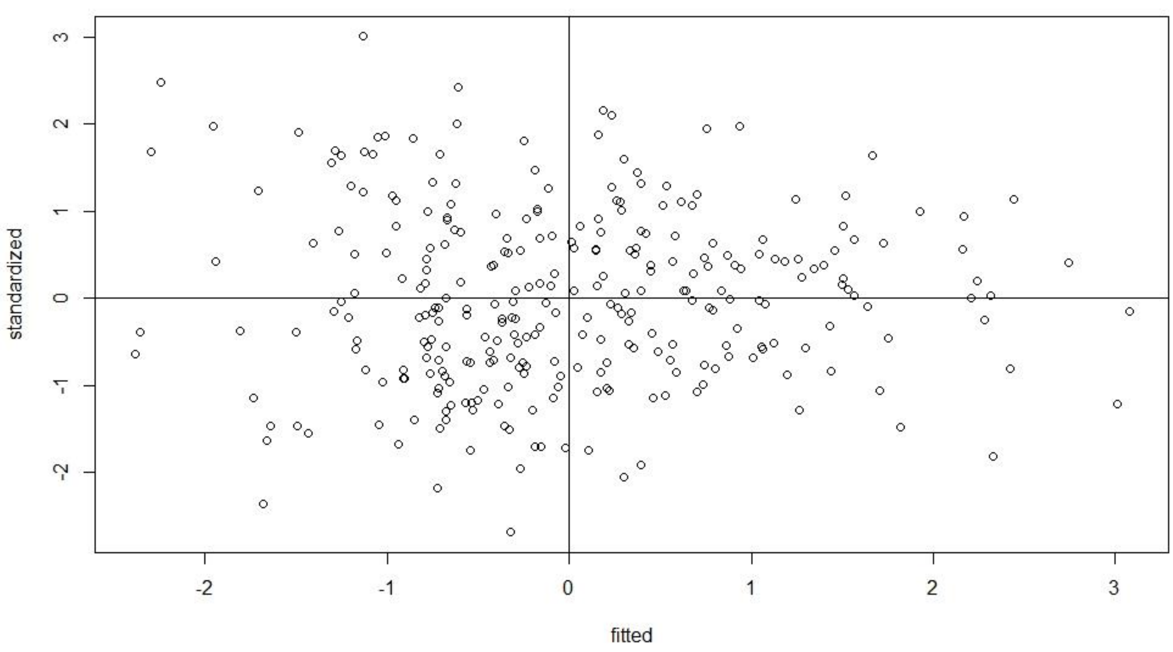

Figure 4: Plot of residuals vs. fitted values (R Core Team, 2016)

3. Linear and additive relationship between dependent and independent variables.

   Plot of residuals vs. predicted (fitted) values demonstrates symmetrical distribution of point among quarters (Figure 4). Even though there are a few outliers, the plot confirms the assumption of linearity and additive relationship of the model, as well as its homoscaditisity and homogeneity.

3.4. VALIDATION OF THE MODEL AND FORECAST OF THE TRANSIT DEMAND

The conducted analysis of regression model assumptions has proven that the created model is not affected by any of the issues of the data like multicollinearity or non-linear distribution due to the nature of transit demand distribution within metropolitan areas. The created model has demonstrated an adjusted R^2 of 0.59, which is considered as statistically significant.

The regression coefficients were used to create a transit demand model:

$$\begin{aligned} TDI = EXP(&5.053 + (1.03\mathrm{E}-06)\times\mathrm{PE} - 0.015\times\mathrm{U18} + 1.009\times\mathrm{TR} \\ &+ (3.00\mathrm{E}-05)\times\mathrm{WL} + (2.00\mathrm{E}-05)\times\mathrm{MF} + 0.151\times\mathrm{SE} \\ &- 0.263\times\mathrm{MHI} + 0.164\times\mathrm{LOS} + 0.793\times\mathrm{TT} + 3.252\times\mathrm{REG}) \end{aligned} \qquad \text{(Formula 1)}$$

Where: TDI – Transit Demand Index

PE – Total population and employment

U18 – Natural logarithm of population under 18

TR – Share of commuters taking transit to work

WL – Product of share of commuters walking to work and their absolute number

MF – Product of share of multifamily housing and total number of multifamily units (10 or more units in structure)

SE – Natural logarithm of combined employment in retail, entertainment, accommodation and food industries (service)

MHI – Ratio of median household income to the median of the metro area

LOS – Level of service

TT – Terminal/transfer status (0/1)

REG – Regional TOD: ratio of total rapid transit ridership to total population of the metropolitan area

The Transit Demand Index (TDI) represents a forecasted transit demand per 0.785 mi^2 area – a half-mile circle with a transit center as a centroid – which would be generated under a certain level of service. The regression was built based on 280 station circles in Chicago, Denver and Los Angeles. On the following step, the model was validated by calculation of forecasted ridership for these stations and comparison to real weekday ridership (previously used as the dependent variable in the regression model). The resulted forecast has provided overall 0.79 correlation between real and forecasted ridership. CTA rail stations had correlation of 0.79, Los Angeles Metro Rail - 0.82 and Denver RTD rail – 0.84. The analysis of residuals has demonstrated that median residual is equal to 79 passengers, minimum is -13,950 and maximum is 9,838. For 80% confidence interval of stations the residuals were estimated within interval of -2250 and 1650 passengers (see Figure 5).

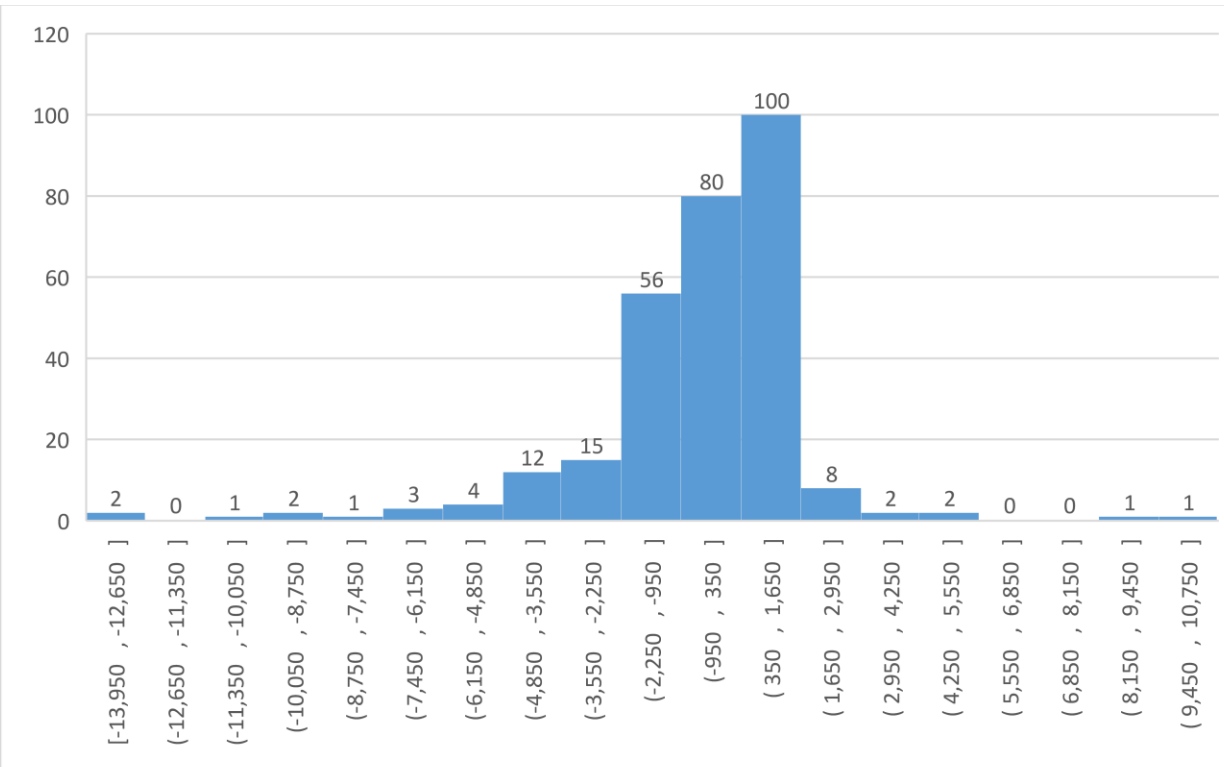

Figure 5: Analysis of residuals for the transit demand model validation

The final step of the research uses the model to predict the transit demand index for every block group in Chicago, Los Angeles and Denver. Since the model includes transit operating parameters, they were assumed to be uniform across the metropolitan area as well as the Regional TOD. The LOS parameter was assumed to be equal to 5 across Chicago and equal to 6 in Los Angeles and Denver to compensate underestimation of the Regional TOD for these metropolitan areas. The terminal/transfer status is assumed to be zero. Thus, the estimation of the travel demand within metropolitan area depends only on the socio-economic parameters of block groups. The resulted distribution of the TDI was added to geodatabase in ArcMap software for further analysis of results.

4. DISCUSSION OF RESULTS AND LIMITATIONS

The main result of the research is that there was created a model able to predict potential transit demand under provision of a certain level of transit service based on the socio-economic parameters of the area within walking distance from the station. The model is based on the publicly available and frequently updated data. The inclusion of the Regional TOD and the LOS parameters allows to use this model for transit demand forecast in other metropolitan areas, besides Chicago, Los Angeles, and Denver. Even though, there is a high variance of parameters for the selected metro areas, a sample of 280 stations makes this model statistically significant, which was proven by multiple tests in the previous chapter, allowing to predict row transit demand estimates for any part of the city.

To demonstrate applicability of the created model it was tested on the sample metropolitan areas of Chicago, Denver and Los Angeles. The resulting maps are included in maps demonstrating existing spatial distribution of the transit demand index and existing rapid transit network in the cities (see Appendixes E, F, and G). The maps allow not only to analyze spatial distribution of TOD-ness of the area, but also to compare the areas of metros among each other. Since the maps include some of the transit routes which were opened after 2015[2], like A and B lines in Denver or Expo Line in Los Angeles, the model can be used to assess the impact of transit on built environment.

One of the most important applications of this research is the ability to analyze the impact of different demographic, economic and land-use components on transit ridership. Based on the regression analysis, some variables have much higher explanatory power than others (see Table 11).

VARIABLE	PE	U18	TR	WL	MF	MHI	SE	LOS	TT	REG
t-value	0.74	-0.46	2.15	0.15	0.66	-2.94	5.78	6.37	8.79	2.38
Significance			*			**	***	***	***	*

Table 11: T-values of the independent variables for the model

It is obvious that transit operating parameters of level of service (LOS) and terminal/transfer status (TT) would have the highest explanatory power since they to some degree represent transit supply which is supposed to balance the transit demand. The transit agency is flexible to increase frequencies (LOS) in case there is increase of transit demand along the route. The Regional TOD parameter has a t-value of 2.38, which validates a destination access component of “5Ds” as a critically important factor in development of transit.

[2] While the most recent 5-year ACS data used in the model is available for 2011-2015

Since the transit operating parameters take up the significant part of the explanatory power of the model, the socio-economic parameters seem to have less impact. Nonetheless, since we consider LOS and TT uniform at the stage of data forecast, the socio-economic parameters remain the ones that make the difference in distribution of transit demand. The analysis of these parameters has provided a rather interesting insight - the employment in accommodation, food, retail and entertainment industries within station area is the main indicator of the transit demand with the t-value of 5.78. It does not necessarily mean that these employees are using transit more often and create additional transit demand, it highlights that the areas with significant employment in these industries tend to have the structure of built environment supporting high transit demand. None of the research analyzed over the literature review phase have used this parameter to predict transit demand. The RTA research on regional TDI included retail employment, however the accommodation and food industry component has significantly higher correlation with transit ridership than retail in this research (Newmark, 2012). Even though retail employees have one of the lowest salaries in the nation, they do not necessarily use transit to go to work since most of the retail in the U.S. has moved to suburban locations having very limited transit accessibility. On the other hand, hotel and food industries tend to concentrate in the most walkable parts of the cities. This is inherently connected to the "5Ds" theory of Robert Cervero, representing design and diversity aspects of transit-oriented development.

Median household income (normalized by median household income in the metro area) has demonstrated negative coefficient in the model with t-value of -2.94. Even though this parameter has a positive correlation of 0.26 with the transit ridership (see Table 9), the interaction with other parameters of the regression model has resulted in a strong negative β-coefficient, which can be backed up by many studies confirming negative correlation between income and transit demand – low-income population tends to use transit more often (Maciag, 2014).

Share of commuters taking transit to work is intended to demonstrate the existing transit usage by residents within the area and has a t-value of 2.15. Even though the rest of the parameters have the t-value less than 1, they are equally important for the model, since their t-value is significantly decreased by explanatory power of other parameters in the model. Population and employment density, multifamily housing and walking estimates reflect density, diversity and design of the area and have positive β-coefficients in the model, while population under 18 has negative coefficient since this parameter is proven to indicate areas with higher share of driving commuters.

The analysis of impact of different demographic, economic and land-use components on transit ridership is especially important for planners and decision-makers, since they can affect these parameters creating policies, regulations and design that support transit development in the most suitable areas. For example, a number of metropolitan areas have implemented lower parking standards regulations in the areas around transit stations (Vence, 2015). Expansion of similar land-use policies to the identified areas with high transit demand would support transit-oriented development and expansion of more sustainable urban environment with redirection of urban growth from suburbs into densification and transformation of central parts of the cities.

To analyze the distribution of these parameters there were created online maps for Chicago, Los Angeles and Denver demonstrating the transit demand as well as the parameters impacting the forecast. The interactive maps were created in R Studio using packages "leaflet" and "htmlwidgets" (Cheng, Karambelkar, & Xie, 2017; Ramnath Vaidyanathan, Yihui Xie, JJ Allaire, Joe Cheng, & Kenton Russell, 2016). The maps have been uploaded to the RPubs online public storage and can be accessed through the following links:

Chicago Metro - http://rpubs.com/ilyachistyakov/chicago
Los Angeles Metro - http://rpubs.com/ilyachistyakov/los-angeles
Denver Metro - http://rpubs.com/ilyachistyakov/denver

The application of the created transit demand model has a number of limitations. First, transit ridership at any station consists mainly of two parts: passenger who walk to the station and passengers who take other modes to access rapid transit, like car, feeder bus, bicycle, or taxi. This model relies mainly on the parameters describing the area within walking distance of the station and can not cover the entire service area of the station. The forecast for the second part of commuters is complicated by high variance of costs and travel times and involves analysis of origins and destinations of commuters, which is impossible without implementation of travel demand modeling software. This is the main reason why commuter rails had to be excluded from the analysis, since around 90% of their passengers drive or carpool to the station. Nevertheless, since the created model is based on the real ridership by station, the final TDI includes a certain average number of people not walking to transit.

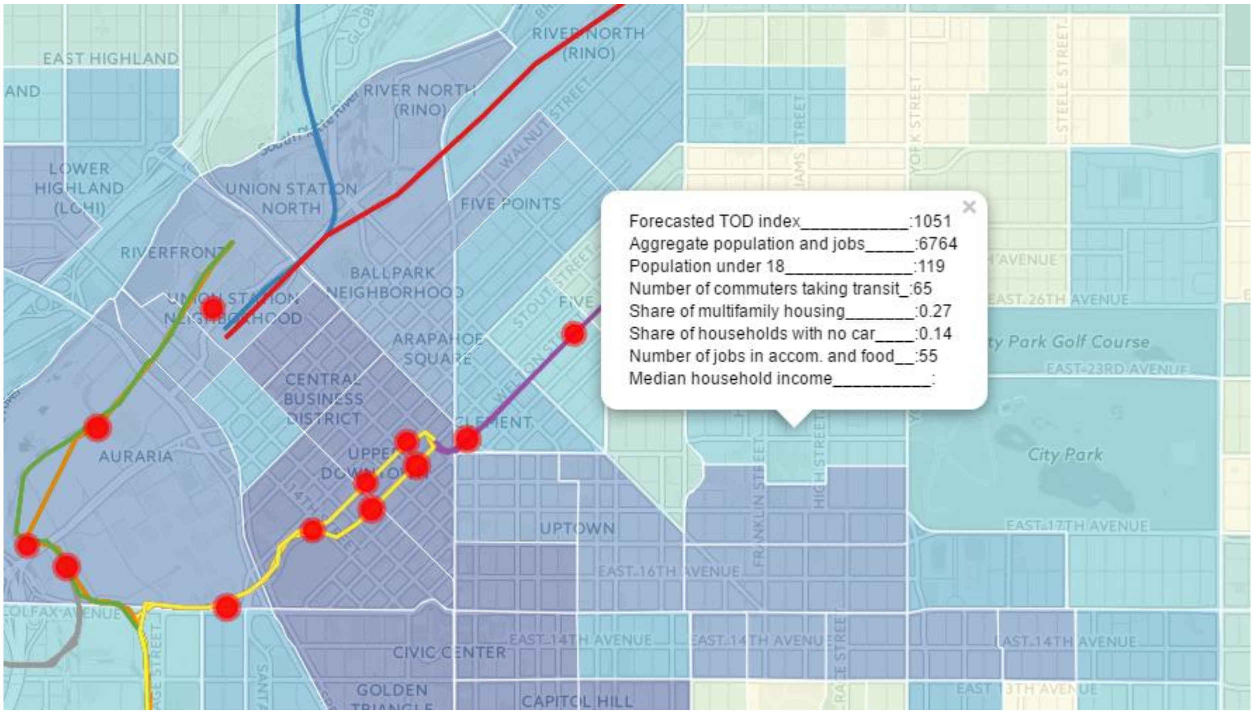

Figure 6: Demonstration of the online map tool for Denver, CO with a pop-up for one of the block groups (http://rpubs.com/ilyachistyakov/denver)

Second, the model does not allow to estimate the redistribution of transit ridership from existing to a new transit line. The total sum of forecasted transit demand for all block groups would be largely overestimating the real transit demand. Basically, the model assumes that the new station would generate the number of passengers equal to TDI per 0.785 mi^2 without taking into consideration that this station might decrease ridership of the nearby transit stop.

Third, the model underestimates ridership connected with some specific destinations, like airports, bus transit centers, large P&R facilities, universities, stadiums, etc. These objects generate significant transit demand however are not covered by any of the selected variables of the model.

Fourth, the model underestimates explanatory power and β-coefficients of socio-economic parameters because of the impact of transit operating variables (LOS and transfer/transit). Even though the model was built on the statistically significant sample (280 stations), the large number of independent variables (10) leads to unavoidable interactions between them, which makes building of the model more complicated and leads to selection of those parameters that might have smaller explanatory power but would not affect other parameters due to multicollinearity or Simpson's paradox. The other limitations of the regression model approach were discussed in the Literature Review chapter.

Finally, the model is static and represents transit demand built on ACS 2011-2015 data, while comparing the stations which has been affecting development of the areas around them for different periods of time. Some of the “L” stations have been in constant operation since 1892 generating the transit-oriented development impulses affecting surrounding built environment. On the other end, some of the stations analyzed in the research did not exist in 2015, thus the transit-oriented development did not affect the socio-economic parameters around them.

5. CONCLUSION

As a result, all the stated objectives were achieved - there was created a model able to predict potential transit demand under provision of a certain level of service based on the socio-economic parameters of the area. The selection of the variables of the model was based on the most recent research in the field and replied on the multidimensional approach including regional and local scales of socio-economic and transit data. The model was built on the most recent weekday ridership estimates for 280 rail stations in three metropolitan areas. The local group of parameters within a half-mile radius of the stations has proven to be the core of the model with seven selected parameters representing density, diversity and design of the areas within walking distance of transit. The parameters were selected over the model building process with implementation of statistical tests like correlation, multicollinearity, and normal distribution of residuals, among others. These tests confirmed the assumptions of the created Ridge regression model as statistically valid. The resulting model included ten independent variables with R^2 of 0.59, which is considered as significant value for prediction of such a multifaceted indicator as a transit demand. The resulting Transit Demand Index can be used for the analysis of spatial distribution within metropolitan areas to identify the locations where transit investment would have the most significant outcome.

The results of the model have a number of applications in the field of urban and transportation planning. First, the created interactive GIS for Chicago, Los Angeles and Denver could be used as a tool for analysis of the existing transit demand with assessment of impact of each individual variable contributing to the model. Impacts of different demographic, economic and land use components could be used to create more sustainable policies, zoning regulations and planning documents. Understanding the impact allows to create better design for the areas around transit and to stimulate transit-oriented development.

Second, the spatial distribution of transit demand could be used for long-term planning of transit development in the metropolitan area and to substantiate prioritization of selected areas. The model can be used for comparison of different metropolitan areas. Potentially, the cluster analysis of the stations could be used to identify the areas with similar transit and socio-economic characteristics. This analysis could become a basis for creation of guidelines for transit-oriented development and design based on the classification.

Finally, the identified coefficients of the parameters with the strongest explanatory power could be implemented into existing Travel Demand Models (TDM) in the metropolitan areas to increase the accuracy of the transit forecast. The identified connections between socio-economic parameters and transit justify implementation of special coefficients on the step of the mode choice and trip assignment.

Yet, since the created transit demand model is a simplification of a multifaceted commuting problem of a mode choice, especially complicated in larger metropolitan areas, it can not be the only guideline for decision-making and has to be used with consideration of a number of limitations. First of all, the model takes into consideration only ten variables and a number of other possible impacts are left out of the scope. For example, some popular destinations, like airports, stadiums, universities, and other objects can not be fully reflected by seven local variables, which causes underestimation for some of the outlier-stations. Second, the model has limited ability to predict differences in distribution of transit demand generated outside of the walking distance of the station, i.e. for commuters driving, taking a bus or other mode of transportation to get to transit. Third, the model is not based on the real trip generation and does not take into consideration redistribution of passengers between transit routes. Thus, the model has limitations to the degree of accuracy and could be used on the preliminary steps of transit demand analysis.

Based on these limitations, the future development of the model might include expansion of the sample size and possible set of the variables. Every new metropolitan area would add its own regional specifics to the model. Currently, Regional TOD parameter is built only on three metropolitan areas, being a significant simplification of the regional model. The expansion of the sample size would contribute to the increased accuracy and explanatory power, which is currently below of some of the case studies done by other researchers. The research could add a time dimension into the set of variables, since there is clear connection between intensity of transit-oriented development around the station and the time this station remains in operation.

The interactive mapping application could be enhanced with a tool, which would be able to draw a buffer around a proposed transit station and calculate the proposed transit ridership based on the block groups within the buffer, although it requires shift of the application to ArcGIS online server, which was not available for this research.

Regardless of the limitations, this model is considered to be a successful attempt contributing to the spectrum of available tools to analyze transit investment projects and their outcomes. It has a number of advantages over the existing models available at the moment in its uniqueness of multiscale approach and a wide range of variables covering transit operation and socio-economic parameters of urban environment.

APPENDIX A: TRANSIT SYSTEMS DATABASE

Metropolitan Area	Transit system	Type*	Year open ***	Weekday Ridersh.**	Length (mi) ***	Stations ***	Routes ***	Spacing (mi)	Pax/mi
Albuquerque, NM	New Mexico Rail Runner Express	CR	2006	2,900	97	13	1	7.46	30
Atlanta-Sandy Springs-Roswell, GA	Atlanta Streetcar	SR	2016	n/a	2.7	12	1	0.23	
	MARTA rail system	HR	1979	213,800	47.6	38	4	1.25	4,492
Atlanta-Sandy Springs-Roswell, GA Total				213,800	50.3	50	5		
Austin-Round Rock, TX	Capital MetroRail	CR	2010	2,800	32	9	1	3.56	88
Baltimore-Columbia-Towson, MD	Baltimore Light Rail	LR	1992	22,800	33	33	3	1.00	691
	Baltimore Metro Subway	HR	1983	36,800	15.5	14	1	1.11	2,374
Baltimore-Columbia-Towson, MD Total				59,600	48.5	47	4	1.03	
Boston-Cambridge-Newton, MA-NH	Ashmont–Mattapan High Speed Line	LR	1929	4,637					
	MBTA Commuter Rail	CR	1973	122,100	368	127	13	2.90	332
	MBTA Green Line	LR	1897	227,645	26	74	5	0.35	8,756
	MBTA Subway (Blue, Orange, and Red Lines)	HR	1901	552,500	38	53	4	0.72	14,539
Boston-Cambridge-Newton, MA-NH Total				908,182	580	266	23		
Buffalo-Cheektowaga-Niagara Falls, NY	Buffalo Metro Rail	LR	1984	17,100	6.4	14	1	0.46	2,672

Table 12: Transit Systems Database

Charlotte-Concord-Gastonia, NC-SC	CityLynx Gold Line	SR	2015	1,500	1.5	6	1	0.25	1,000
	Lynx Blue Line	LR	2007	17,600	9.6	15	1	0.64	1,833
Charlotte-Concord-Gastonia, NC-SC Total				19,100	11.1	21	2		1,721
Chicago-Naperville-Elgin, IL-IN-WI	Chicago "L"	HR	1892	749,700	102.8	146	8	0.70	7,293
	Metra	CR	1984	292,000	487.7	241	11	2.02	599
	NICTD South Shore Line	CR	1903	14,200	90	20	1	4.50	158
Chicago-Naperville-Elgin, IL-IN-WI Total				1,055,900	680.5	407	20		
Cincinnati-Middletown–Wilmington, OH-KY-IN	Cincinnati Bell Connector	SR	2016	3,163	3.6	18	1	0.20	879
Cleveland-Elyria, OH	Blue and Green Lines	LR	1913	12,400		34	2	-	
	RTA Rapid Transit (Red Line)	HR	1955	17,637	19	18	1	1.06	928
Cleveland-Elyria, OH Total				30,037	19	52	3		1,581
Dallas-Fort Worth-Arlington, TX	A-Train	CR	2011	1,900	21	6	1	3.50	90
	Dallas Area Rapid Transit (DART)	LR	1996	104,800	93	64	4	1.45	1,127
	Dallas Streetcar	SR	2015	n/a		6	1	-	
	McKinney Avenue Transit Authority	SR	1989	n/a		6	1	-	
	Trinity Railway Express	CR	1996	8,100	34	10	1	3.40	238
Dallas-Fort Worth-Arlington, TX Total				114,800	148	92	8		

Table 12: Transit Systems Database (cont.)

Denver-Aurora-Lakewood, CO	RTD (A Line and B Line)	HR	2016	25,724	30.8	8	2	3.85	835
	RTD Light Rail	LR	1994	76,600	48	46	6	1.04	1,596
Denver-Aurora-Lakewood, CO Total				102,324	78.8	54	8		1,299
Houston-The Woodlands-Sugar Land, TX	METRORail	LR	2004	60,600	22.7	44	3	0.52	2,670
Kansas City, MO-KS	KC Streetcar	SR	2016	6,800	2.2	10	1	0.22	3,091
Los Angeles-Long Beach-Anaheim, CA	Metro Rail	LR	1990	189,700	98.5	80	4	1.23	1,926
	Metro Rail (Purple and Red Lines)	HR	1993	143,000	17.4	16	2	1.09	8,218
	Metrolink	CR	1992	40,500	388	55	7	7.05	104
Los Angeles-Long Beach-Anaheim, CA Total				373,200	503.9	151	13		
Miami-Fort Lauderdale-West Palm Beach, FL	METRORail	LR	2004	70,300	24.4	23	2	1.06	2,881
	Tri-Rail	CR	1987	14,200	70.9	18	1	3.94	200
Miami-Fort Lauderdale-West Palm Beach, FL Total					95.3	41	3	2.32	887
Minneapolis-St. Paul-Bloomington, MN-WI	METRO: Blue & Green lines	LR	2004	72,900	21.8	37	2	0.59	3,344
	Northstar Line	CR	2009	2,500	40	6	1	6.67	63
Minneapolis-St. Paul-Bloomington, MN-WI Total					61.8	43	3		
Nashville-Davidson--Murfreesboro--Franklin, TN	Music City Star	CR	2006	750	32	6	1	5.33	23
New Haven-Milford, CT	Shore Line East	CR	1990	2,000	59	13	1	4.54	34

Table 12: Transit Systems Database (cont.)

New Orleans-Metairie, LA	New Orleans Streetcars	SR	1835	22,900	22.3		4		1,027
New York-Newark-Jersey City, NY-NJ-PA	Hudson–Bergen Light Rail	LR	2006	66,299	23.25	41	5	0.57	2,852
	MTA Long Island Rail Road	CR	1834	333,600	321	124	11	2.59	1,039
	MTA Metro-North Railroad	CR	1983	304,800	385	122	6	3.16	792
	New York City Subway	HR	1904	8,918,400	233	472	24	0.49	38,276
	Newark Light Rail (NJ Transit)	LR	1935	66,299		41	5		
	NJ Transit Rail	CR	1983	323,400	530	164	11	3.23	610
	Port Authority Trans-Hudson (PATH)	HR	1908	270,900	13.8	13	4	1.06	19,630
	Staten Island Railway	HR	1860	32,200	14	22	1	0.64	2,300
New York-Newark-Jersey City, NY-NJ-PA Total				10,315,898	1520.05	999	67		
Orlando-Kissimmee-Sanford, FL	SunRail	CR	2014	2,627	32	12	1	2.67	82
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	Keystone Service	CR	1976	4,600	104.6	12	1	8.72	44
	PATCO Speedline	HR	1936	36,500	14.2	13	1	1.09	2,570
	River LINE (NJ Transit)	LR	2004	9,014	34	20	1	1.70	265
	SEPTA (Broad Street, Market–Frankford, and Norristown High Speed Lines)	HR	1907	311,800	36.7	75	3	0.49	8,496
	SEPTA light rail	LR	1906	111,600	68.4	100	8	0.68	1,632
	SEPTA Regional Rail	CR	1983	134,300	280	153	13	1.83	480

Table 12: Transit Systems Database (cont.)

Philadelphia-Camden-Wilmington, PA-NJ-DE-MD Total					537.9	373	27		
Phoenix-Mesa-Scottsdale, AZ	Valley Metro Rail	LR	2008	58,700	26.3	35	1	0.75	2,232
Pittsburgh, PA	Pittsburgh Light Rail	LR	1984	22,281	26.2	53	2	0.49	850
Portland-Vancouver-Hillsboro, OR-WA	MAX Light Rail	LR	1986	122,900	60	97	5	0.62	2,048
	Portland Streetcar	SR	2001	15,248	7.35	76	2	0.10	2,075
	Westside Express Service	CR	2010	1,700	15	5	1	3.00	113
Portland-Vancouver-Hillsboro, OR-WA Total				139,848	82.35	178	8		
Sacramento--Roseville--Arden-Arcade, CA	Sacramento RT Light Rail	LR	1987	45,300	42.9	53	3	0.81	1,056
Salt Lake City, UT	S Line	SR	2013	1,087	2	7	1	0.29	544
	TRAX	LR	1999	67,300	44.8	50	3	0.90	1,502
	UTA FrontRunner	CR	2008	17,600	88	16	1	5.50	200
Salt Lake City, UT Total				85,987	134.8	73	5		
San Diego-Carlsbad, CA	NCTD Coaster	CR	1995	4,800	41	8	1	5.13	117
	San Diego Trolley	LR	1981	123,300	53.5	53	4	1.01	2,305
	Sprinter	LR	2008	8,900	22	15	1	1.47	405
San Diego-Carlsbad, CA Total				137,000	116.5	76	6		
San Francisco-Oakland-Hayward, CA	Bay Area Rapid Transit (BART)	HR	1972	446,200	104	45	5	2.31	4,290

Table 12: Transit Systems Database (cont.)

	Caltrain	CR	1987	56,900	77	32	1	2.41	739
	Muni Metro	LR	1980	180,500	35.7	152	9	0.23	5,056
San Francisco-Oakland-Hayward, CA Total				683,600	216.7	229	15		
San Jose-Sunnyvale-Santa Clara, CA	Altamont Corridor Express (ACE)	CR	1998	5,000	86	10	1	8.60	58
	Santa Clara VTA Light Rail	LR	1987	33,400	42.2	62	3	0.68	791
San Jose-Sunnyvale-Santa Clara, CA Total				38,400	128.2	72	4		
San Juan-Caguas-Guaynabo	Tren Urbano	HR	2004	30,400	10.7	16	1	0.67	2,841
Seattle-Tacoma-Bellevue, WA	Central Link (Sound Transit)	LR	2009	69,125	20.35				3,397
	Seattle Streetcar	SR	2007	1,800	3.8	21	2	0.18	474
	Sounder Commuter Rail	CR	2000	15,000	83	9	2	9.22	181
	Tacoma Link (Sound Transit)	LR	2003	3,447			1		
Seattle-Tacoma-Bellevue, WA Total				89,372	107.15	30	5		834
St. Louis, MO-IL	Metrolink	CR	1992	49,500	46	37	2	1.24	1,076
Tucson, AZ	Sun Link	SR	2014	4,000	3.9	22	1	0.18	1,026
Virginia Beach-Norfolk-Newport News, VA-NC	The Tide	LR	2011	5,800	7.4	11	1	0.67	784
Washington-Arlington-Alexandria, DC-VA-MD-WV	DC Streetcar	SR	2016	n/a	2.4	8	1	0.30	
	MARC Train	CR	1984	33,800	187	43	3	4.35	181

Table 12: Transit Systems Database (cont.)

	Virginia Railway Express	CR	1992	11,983	90	18	2	5.00	133
	Washington Metro	HR	1976	748,800	117	91	6	1.29	6,400
Washington-Arlington-Alexandria, DC-VA-MD-WV Total				794,583	396.4	160	12		

Table 12: Transit Systems Database (cont.)

*Abbreviations: CR – commuter rail; HR – heavy rail; LR – light rail; SR – streetcar

** Weekday ridership data base on the FTA Monthly Module Raw Data Release for January 2016 or American Public Transportation Association (APTA) for 4Q 2016 ("APTA," 2016, "Monthly Module Raw Data Release," 2016).

***Year open, number of stations, routes and length based on various sources in Wikipedia ("List of United States commuter rail systems by ridership," 2017, "List of United States light rail systems by ridership," 2017, "List of United States rapid transit systems by ridership," 2017)

APPENDIX B: TRANSIT DEMAND POTENTIAL BY METRO AREA

#	Metropolitan Statistical Area	Population Estimate July 1,2016 *	Population change (5 years) *	Weekday Ridership **	Regional TOD (rid. by pop.)
1	New York-Newark-Jersey City, NY-NJ-PA	20,153,634	2.0%	10,315,898	0.512
2	Boston-Cambridge-Newton, MA-NH	4,794,447	4.0%	908,182	0.189
3	San Francisco-Oakland-Hayward, CA	4,679,166	6.4%	683,600	0.146
4	Washington-Arlington-Alexandria, DC-VA-MD-WV	6,131,977	6.1%	760,783	0.124
5	Chicago-Naperville-Elgin, IL-IN-WI	9,512,999	0.2%	1,055,900	0.111
6	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	6,070,500	1.2%	607,814	0.100
7	Salt Lake City, UT	1,186,187	7.1%	85,987	0.072
8	Portland-Vancouver-Hillsboro, OR-WA	2,424,955	7.3%	139,848	0.058
9	San Diego-Carlsbad, CA	3,317,749	5.6%	137,000	0.041
10	Atlanta-Sandy Springs-Roswell, GA	5,789,700	7.7%	213,800	0.037
11	Denver-Aurora-Lakewood, CO	2,853,077	9.7%	102,324	0.036
12	Baltimore-Columbia-Towson, MD	2,798,886	2.3%	93,400	0.033
13	Los Angeles-Long Beach-Anaheim, CA	13,310,447	2.8%	373,200	0.028
14	Seattle-Tacoma-Bellevue, WA	3,798,902	8.6%	89,372	0.024
15	San Jose-Sunnyvale-Santa Clara, CA	1,978,816	5.9%	43,300	0.022
16	Minneapolis-St. Paul-Bloomington, MN-WI	3,551,036	4.8%	75,400	0.021
17	Sacramento--Roseville--Arden-Arcade, CA	2,296,418	5.6%	45,300	0.020
18	New Orleans-Metairie, LA	1,268,883	4.5%	22,900	0.018
19	St. Louis, MO-IL	2,807,002	0.5%	49,500	0.018
20	Dallas-Fort Worth-Arlington, TX	7,233,323	10.1%	114,800	0.016
21	Buffalo-Cheektowaga-Niagara Falls, NY	1,132,804	-0.3%	17,100	0.015
22	Cleveland-Elyria, OH	2,055,612	-0.6%	30,037	0.015
23	San Juan-Carolina-Caguas, PR	2,157,729	-7.1%	30,400	0.014
24	Miami-Fort Lauderdale-West Palm Beach, FL	6,066,387	6.5%	84,500	0.014
25	Phoenix-Mesa-Scottsdale, AZ	4,661,537	9.7%	58,700	0.013
26	Pittsburgh, PA	2,342,299	-0.7%	22,281	0.010
27	Houston-The Woodlands-Sugar Land, TX	6,772,470	11.8%	60,600	0.009
28	Charlotte-Concord-Gastonia, NC-SC	2,474,314	9.7%	19,100	0.008
29	Tucson, AZ	1,016,206	2.9%	4,000	0.004
30	Virginia Beach-Norfolk-Newport News, VA-NC	1,726,907	2.4%	5,800	0.003
31	Kansas City, MO-KS	2,104,509	3.9%	6,800	0.003
32	Albuquerque, NM	909,906	1.5%	2,900	0.003
33	New Haven-Milford, CT	856,875	-0.8%	2,000	0.002

Table 13: Estimation of a Regional TOD of the metropolitan areas

34	Cincinnati, OH-KY-IN	2,165,139	2.0%	3,163	0.001
35	Austin-Round Rock, TX	2,056,405	15.5%	2,800	0.001
36	Orlando-Kissimmee-Sanford, FL	2,441,257	12.2%	2,627	0.001
37	Nashville-Davidson--Murfreesboro--Franklin, TN	1,865,298	9.8%	750	0.000

Table 13: Estimation of a Regional TOD of the metropolitan areas (cont.)

* American Community Survey: PEPANNRES Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016 (U. S. Census Bureau, 2016)

** Weekday ridership data base on the FTA Monthly Module Raw Data Release for January 2016 or American Public Transportation Association (APTA) for 4Q 2016 ("APTA," 2016, "Monthly Module Raw Data Release," 2016).

APPENDIX C: REGIONAL TRANSIT RIDERSHIP FOR THE U.S. METROPOLITAN AREAS

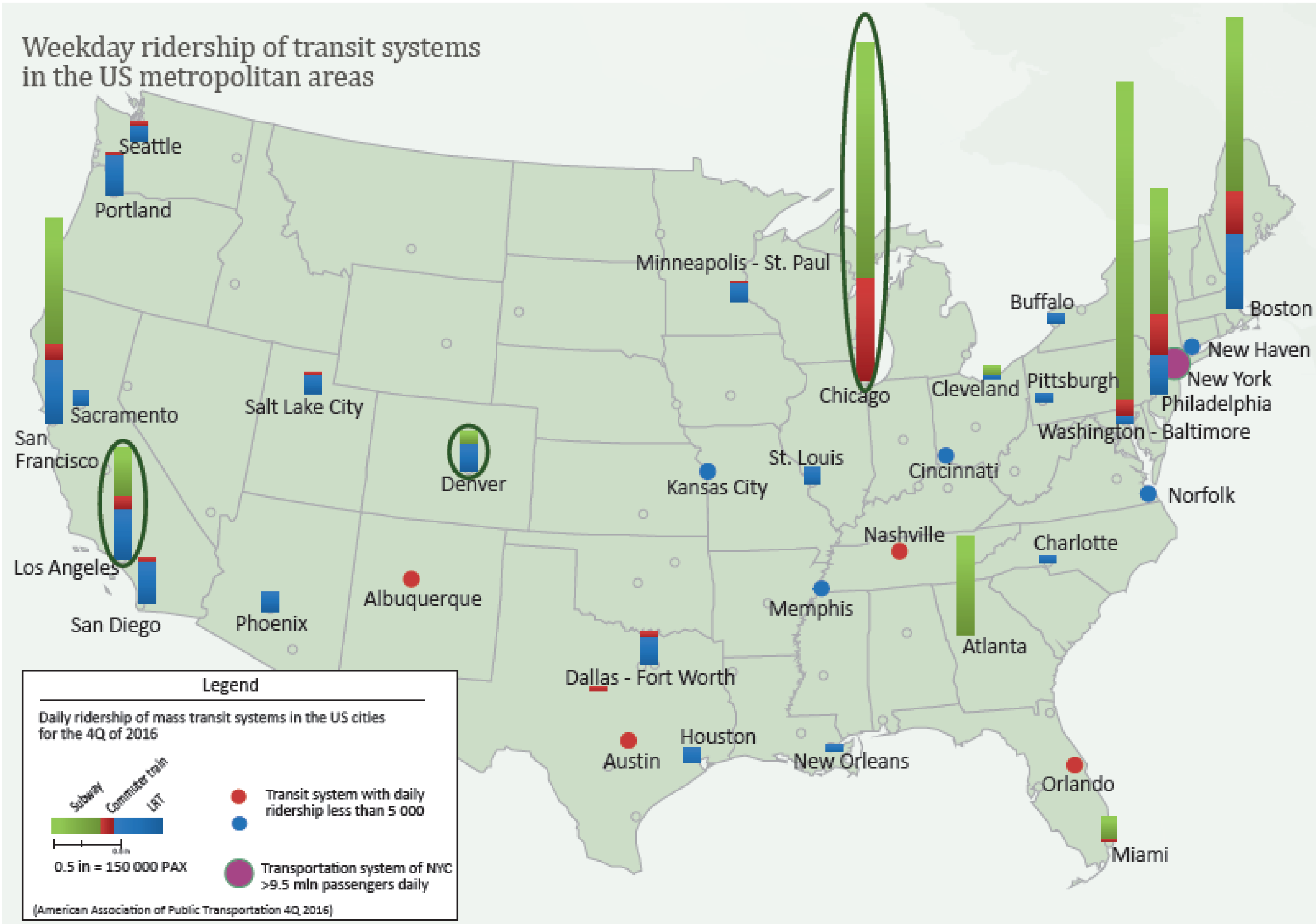

Figure 7: Regional Transit Ridership for the U.S. Metropolitan Areas ("APTA," 2016, "Monthly Module Raw Data Release," 2016)

APPENDIX D: PLOTS OF VARIABLES DISTRIBUTION

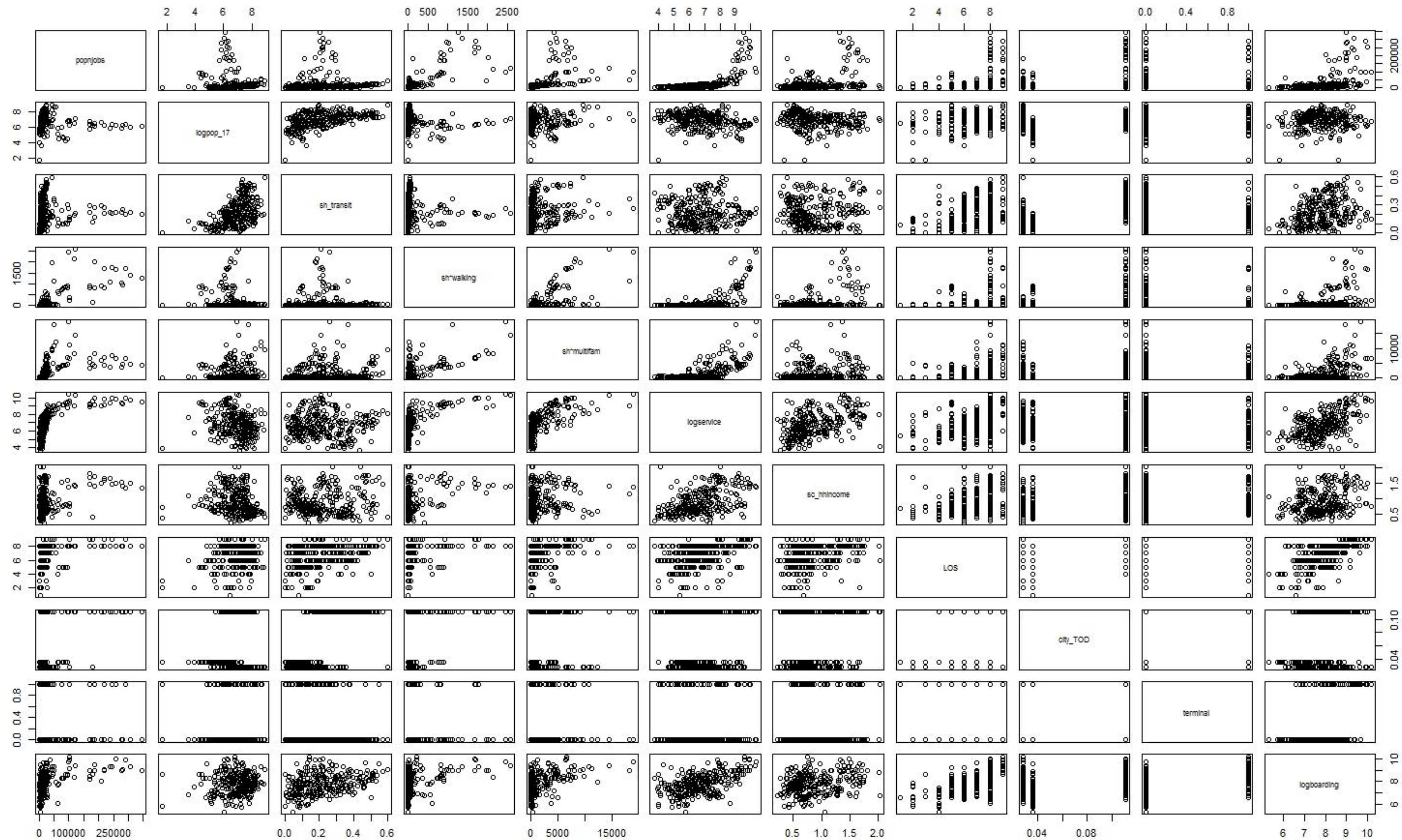

Figure 8: Plots of variables distribution. Created in R Studio (R Core Team, 2016)

APPENDIX E: TRANSIT DEMAND INDEX FOR CHICAGO

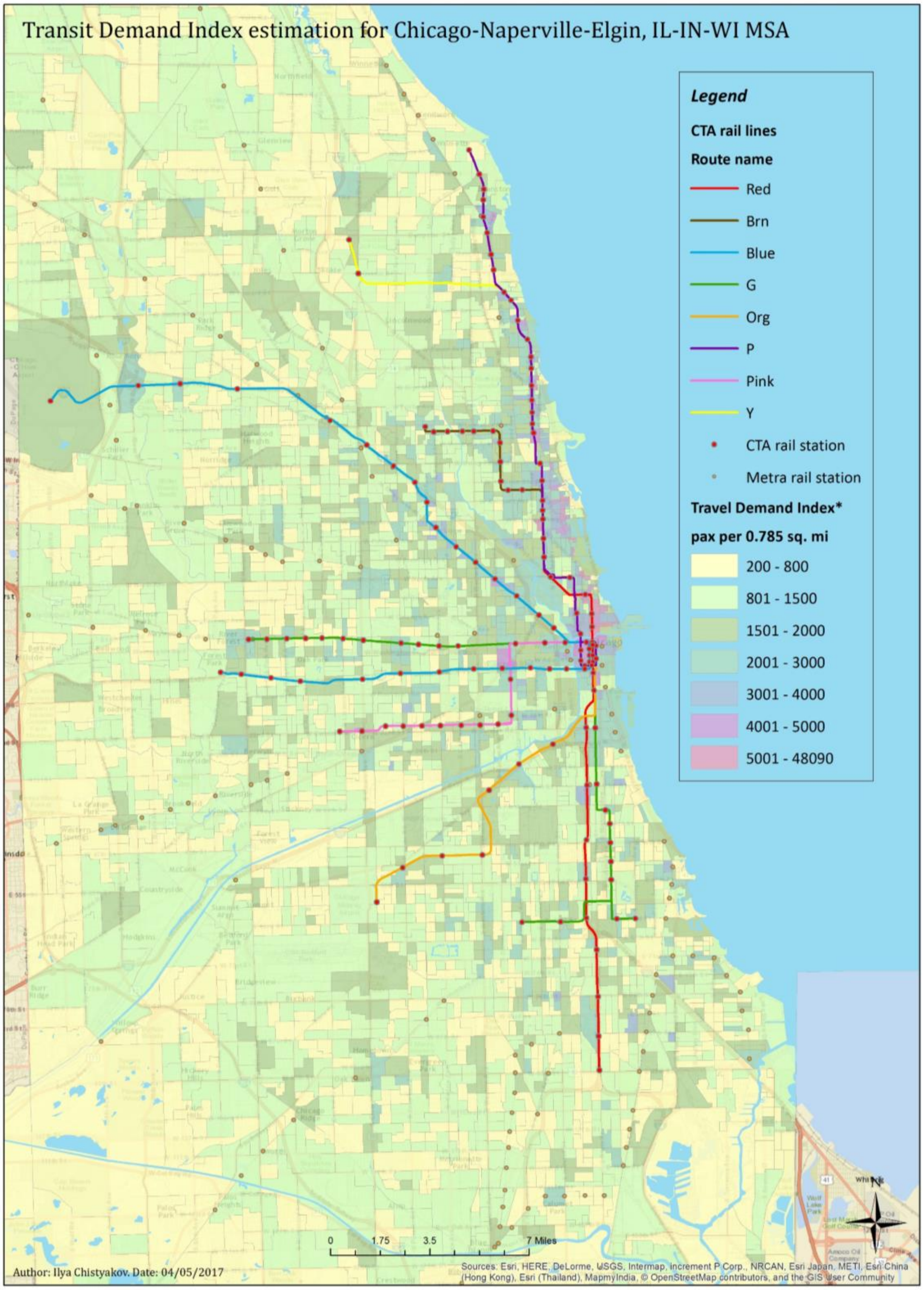

Figure 9: Transit Demand Index for central part of Chicago Metropolitan Area

APPENDIX F: TRANSIT DEMAND INDEX FOR LOS ANGELES

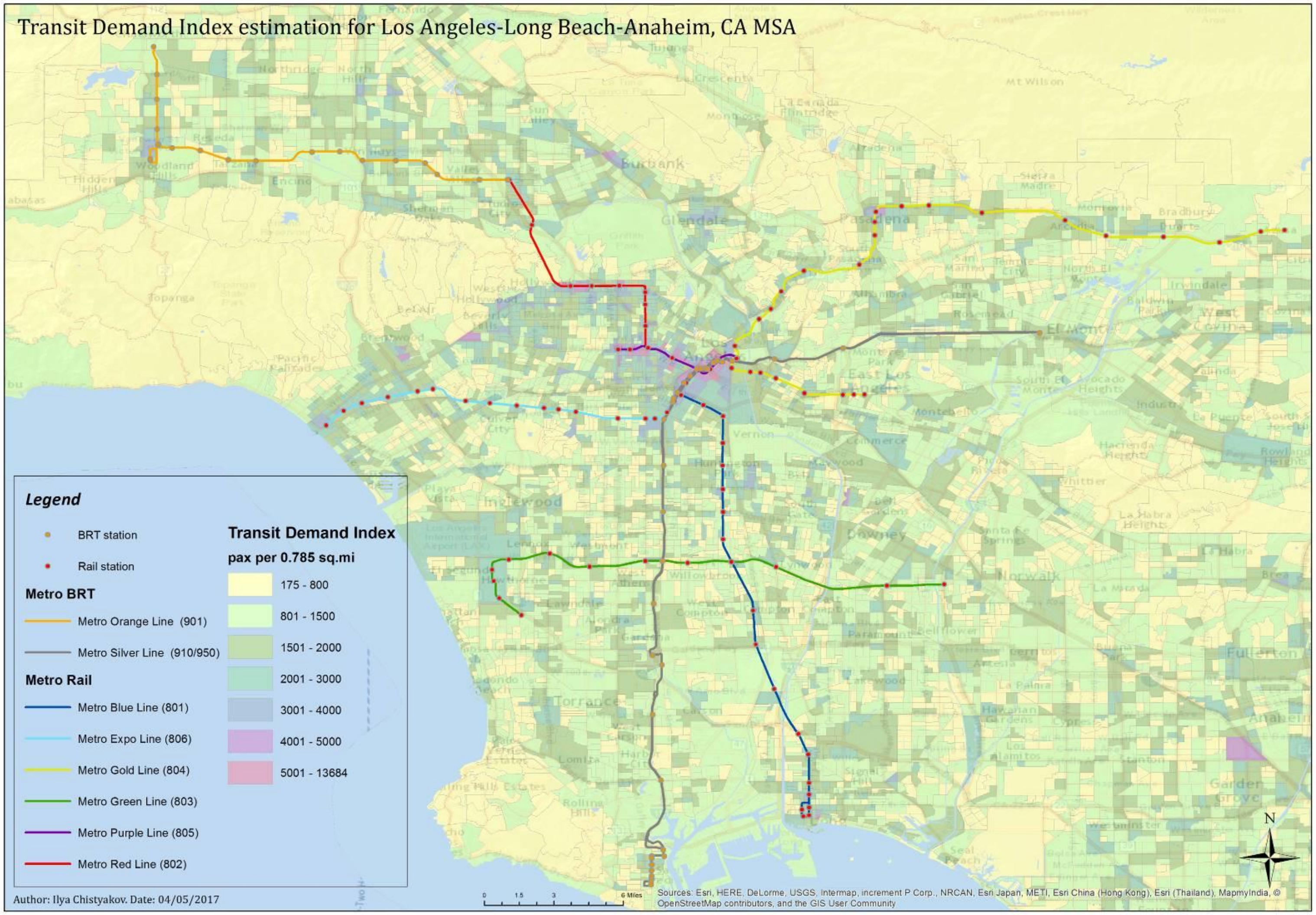

Figure 10: Transit Demand Index for central part of Los Angeles Metropolitan Area

APPENDIX G: TRANSIT DEMAND INDEX FOR DENVER

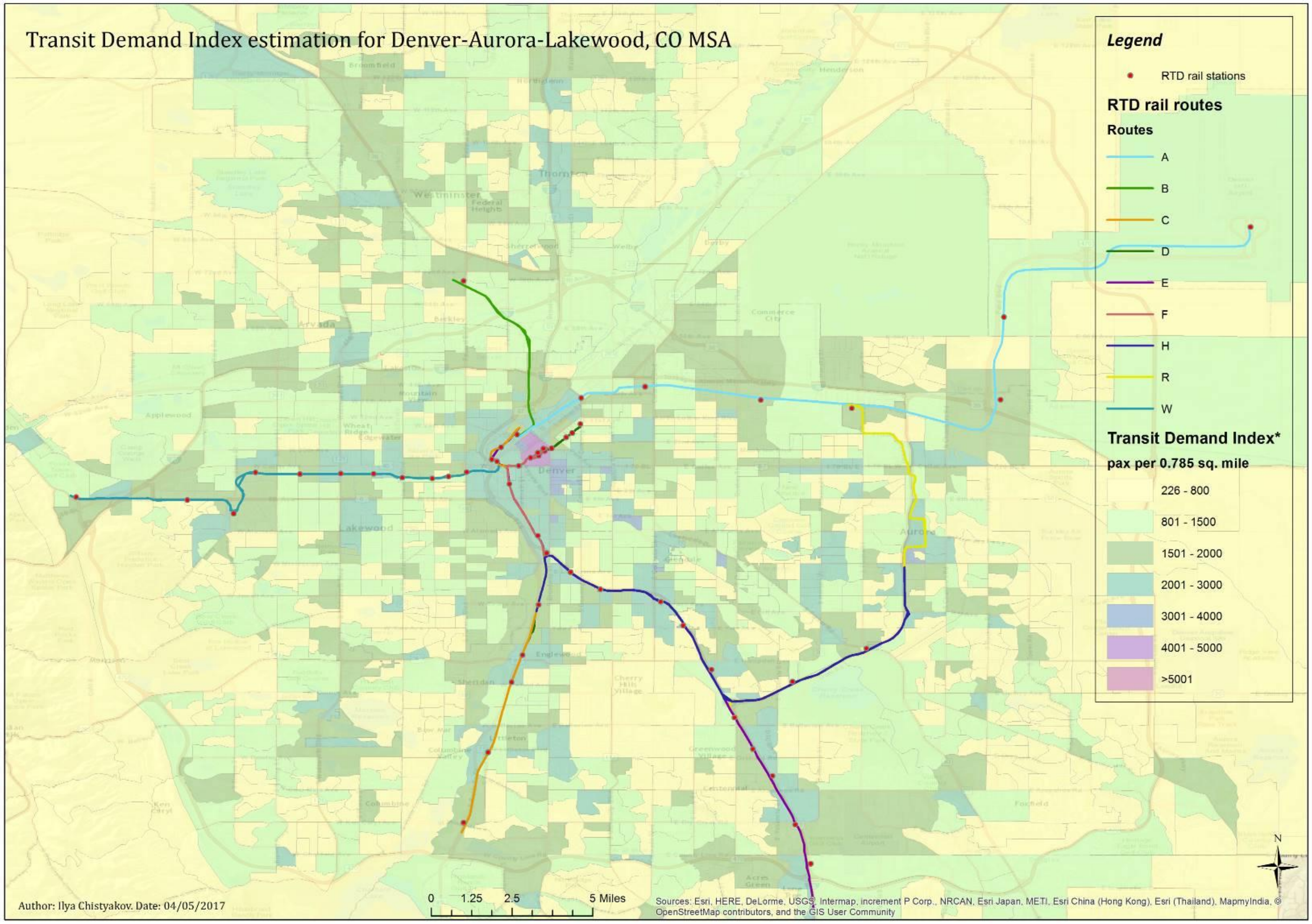

Figure 11: Transit Demand Index for central part of Denver Metropolitan Area

REFERENCES

Bernick, M. S., & Cervero, R. B. (1996). *Transit Villages in the 21st Century*. New York: McGraw-Hill Companies.

Brown, J., & Neog, D. (2012). Central Business Districts and Transit Ridership: A Reexamination of the Relationship in the United States. *Journal of Public Transportation*, *15*(4). https://doi.org/http://dx.doi.org/10.5038/2375-0901.15.4.1

Calthorpe, P. (1993). *The Next American Metropolis: Ecology, Community, and the American Dream*. Princeton Architectural Press.

Cervero, R. (2006). Alternative Approaches to Modeling the Travel-Demand Impacts of Smart Growth. *Journal of the American Planning Association*, *72*(3), 285–295. https://doi.org/10.1080/01944360608976751

Cervero, R. (2011, November). *5 Ds of Urban Development & Rapid Transit Performance*. Retrieved from https://umanitoba.ca/faculties/.../cervero_-_5Ds_and_transit-Winnepeg-Nov2011.pdf

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. https://doi.org/10.1016/S1361-9209(97)00009-6

Cervero, R., Sarmiento, O., Jacoby, E., Gomez, L., & Neiman, A. (2009). Influences of Built Environments on Walking and Cycling: Lessons from Bogota. *World Transit Research*. Retrieved from http://www.worldtransitresearch.info/research/2474

Cheng, J., Karambelkar, B., & Xie, Y. (2017). *leaflet: Create Interactive Web Maps with the JavaScript “Leaflet” Library*. Retrieved from https://CRAN.R-project.org/package=leaflet

Chistyakov, I. (2015a). *Evaluation of transit-oriented development in the US agglomerations. Prepared for the Course of Economics for Planners UP509*.

Chistyakov, I. (2015b). Sustainable development of urban transport systems. Analysis of transit-oriented development in the USA. *Proceedings of the 2nd Russian-Japanese Collaboration Seminar for Sustainable Environment / Hatoyama K., Yu.L. Mazurov and A.A. Pakina (Eds.). – Tokyo, University of Tokyo Press 2015*.

Davis, S. C., Diegel, S. W., & Boundy, R. G. (2009). *Transportation Energy Data Book* (Edition 28/ ORNL-6984 (Edition 28 of ORNL-5198)). US Department of Energy.

DVRPC. (2007). *Creating a Regional Transit Score Protocol: Full Report*. Retrieved from http://www.dvrpc.org/Products/07005/

Evans IV, J., Pratt, R., Stryker, A., & Kuzmyak, J. (2007). Transit-Oriented Development -- Traveler Response to Transportation System Changes. *World Transit Research*. Retrieved from http://www.worldtransitresearch.info/research/3058

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, *76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Fourth Quarter 2016 Ridership Report. (2016). Retrieved April 7, 2017, from http://www.apta.com/resources/statistics/Pages/ridershipreport.aspx

Fox, J., & Weisberg, S. (2011). An {R} Companion to Applied Regression (Version Second). Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

GTFS Static Overview | Static Transit. (n.d.). Retrieved April 8, 2017, from https://developers.google.com/transit/gtfs/

Guerra, E., Cervero, R., & Tischler, D. (2012). Half-Mile Circle. *Transportation Research Record: Journal of the Transportation Research Board*, *2276*, 101–109. https://doi.org/10.3141/2276-12

Jaffe, E. (2015). California’s DOT Makes a Rare Admission: More Roads Mean More Traffic. Retrieved April 2, 2017, from http://www.citylab.com/commute/2015/11/californias-dot-admits-that-more-roads-mean-more-traffic/415245/

Katz, P. (1993). *The New Urbanism: Toward an Architecture of Community* (1 edition). New York: McGraw-Hill Education.

Kneebone, E. (2013, April). Job Sprawl Stalls: The Great Recession and Metropolitan Employment Location. Retrieved March 4, 2016, from http://www.brookings.edu/research/reports/2013/04/18-job-sprawl-kneebone

Komsta, L., & Novomestky, F. (2015). *moments: Moments, cumulants, skewness, kurtosis and related tests*. R. Retrieved from https://CRAN.R-project.org/package=moments

List of United States commuter rail systems by ridership. (2017, April 2). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=List_of_United_States_commuter_rail_systems_by_ridership&oldid=773375646

List of United States light rail systems by ridership. (2017, April 2). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=List_of_United_States_light_rail_systems_by_ridership&oldid=773494240

List of United States rapid transit systems by ridership. (2017, April 1). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=List_of_United_States_rapid_transit_systems_by_ridership&oldid=773315028

Maciag, M. (2014, February 25). Public Transportation's Demographic Divide. *Governing.com*. Retrieved from http://www.governing.com/topics/transportation-infrastructure/gov-public-transportation-riders-demographic-divide-for-cities.html

Monthly Module Raw Data Release. (2016, June 1). [Text]. Retrieved April 7, 2017, from https://www.transit.dot.gov/ntd/data-product/monthly-module-raw-data-release

Morang, M., & Wasserman, D. (n.d.). BetterBusBuffers - Using GTFS Data in ArcGIS Network Analyst. Retrieved April 8, 2017, from http://www.transit.melindamorang.com/overview_BetterBusBuffers.html

Newmark, G. (2012, July). *Developing the Transit Demand Index (TDI)*. Presented at the Transport Chicago, Chicago, Ill.

Ouyang, Y. (2017, January). *Introduction and Overview presentation. CEE 418 Public Transportation Systems*.

Ramnath Vaidyanathan, Yihui Xie, JJ Allaire, Joe Cheng, & Kenton Russell. (2016). *htmlwidgets: HTML Widgets for R*. Retrieved from https://CRAN.R-project.org/package=htmlwidgets

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. R, Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

RTAMS - Rail Ridership Station Detail. (2016). Retrieved April 8, 2017, from http://webcache.googleusercontent.com/search?q=cache:O3Fz9cFjljEJ:www.rtams.org/rtams/ridershipDetail.jsp%3Fdataset%3DctaRail+&cd=2&hl=en&ct=clnk&gl=us

RTD - Ridership Statistics by rail station. (2017). Retrieved April 8, 2017, from http://www.rtd-denver.com/ServiceDevelopment.shtml

Singh, Y. J., Fard, P., Zuidgeest, M., Brussel, M., & Maarseveen, M. van. (2014). Measuring transit oriented development: a spatial multi criteria assessment approach for the City Region Arnhem and Nijmegen. *Journal of Transport Geography*, *35*, 130–143. https://doi.org/10.1016/j.jtrangeo.2014.01.014

Taylor, B. D., & Fink, C. N. Y. (2003). The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature. *University of California Transportation Center*. Retrieved from http://escholarship.org/uc/item/3xk9j8m2

TCRP Report 165 - Transit Capacity and Quality of Service Manual, Third Edition. (2014, March 30). Retrieved April 2, 2017, from http://www.trb.org/Main/Blurbs/169437.aspx

The National Transit Database (NTD). (2015, August 21). [Text]. Retrieved April 7, 2017, from https://www.transit.dot.gov/ntd

TransitFeeds - Public transit feeds from around the world. (2017). Retrieved April 8, 2017, from https://transitfeeds.com/

Urbanism Principles. (n.d.). Retrieved April 2, 2017, from http://www.newurbanism.org/newurbanism/principles.html

U.S. Census Bureau. (2010). 2010 Geographic Terms and Concepts - Block Groups. Retrieved April 8, 2017, from https://www.census.gov/geo/reference/gtc/gtc_bg.html

U.S. Census Bureau. (2015). TIGER/Line with Data: American Community Survey 5-Year Estimates. Retrieved December 2, 2016, from https://www.census.gov/geo/maps-data/data/tiger-data.html

U. S. Census Bureau. (2016). American Community Survey: PEPANNRES Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016. Retrieved April 7, 2017, from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2016_PEPANNRES&src=pt

U.S. Census Bureau Center for Economic Studies. (2014). LEHD Origin-Destination Employment Statistics (LODES). Retrieved December 2, 2016, from http://lehd.ces.census.gov/data/

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. R, New York. Retrieved from http://www.stats.ox.ac.uk/pub/MASS4

Vence, S. (2015, September 28). New TOD Ordinance Will Bring Parking-Lite Development to More of Chicago. Retrieved from http://chi.streetsblog.org/2015/09/28/new-tod-ordinance-will-bring-parking-lite-development-to-more-of-chicago/

Vuchic, V. R. (2005). *Urban Transit : Operations, Planning and Economics* (1 edition). Hoboken, N.J: Wiley.

Walker, J. (2010, October 2). Can We Make Density Make Sense? Retrieved from http://humantransit.org/2010/10/can-we-make-density-make-sense.html

Wickham, H., Hester, J., & Francois, R. (2016). *readr: Read Tabular Data*. Retrieved from https://CRAN.R-project.org/package=readr