TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.
UMTA/TX-86/1081-1F

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
The Use of Profiles for Metropolitan
Transit Authority Analysis

5. Report Date
April 1986

6. Performing Organization Code

7. Author(s)
Katie N. Womack, Mark Von Tress, and
Julie Fesenmaier

8. Performing Organization Report No.
Technical Report 1081-1F

9. Performing Organization Name and Address
Texas Transportation Institute
The Texas A&M University System
College Station, Texas 77843

10. Work Unit No.

11. Contract or Grant No.
Study No. 3-10-85-1081

12. Sponsoring Agency Name and Address
Texas State Department of Highways and Public
Transportation; Transportation Planning Division
P. O. Box 5051
Austin, Texas 78763

13. Type of Report and Period Covered
Final - September 1984
April 1986

14. Sponsoring Agency Code

15. Supplementary Notes
Research performed in cooperation with DOT, UMTA.
Technical Study Title: Metropolitan Transit Authorities in Texas

16. Abstract

This report describes a methodology whereby transit systems were categorized into descriptive profiles on the basis of local socioeconomic and performance factors. The study showed that 267 transit systems in the U. S. can be profiled into seven categories based on nine socioeconomic/performance factors. Texas systems, with and without authority organization, were classified as average in general.

17. Key Words
transit authorities, performance,
socioeconomic data, profiles.

18. Distribution Statement
No restrictions. This document is
available to the public through the
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

19. Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
57

22. Price

Form DOT F 1700.7 (8-69)

THE USE OF PROFILES FOR METROPOLITAN TRANSIT AUTHORITY ANALYSIS

by

Katie N. Womack

Mark Von Tress

and

Julie Fesenmaier

Technical Report 1081-1F
Technical Study Number 3-10-85-1081
Metropolitan Transit Authorities in Texas

Sponsored by

U. S. Department of Transportation

and

Urban Mass Transportation Administration

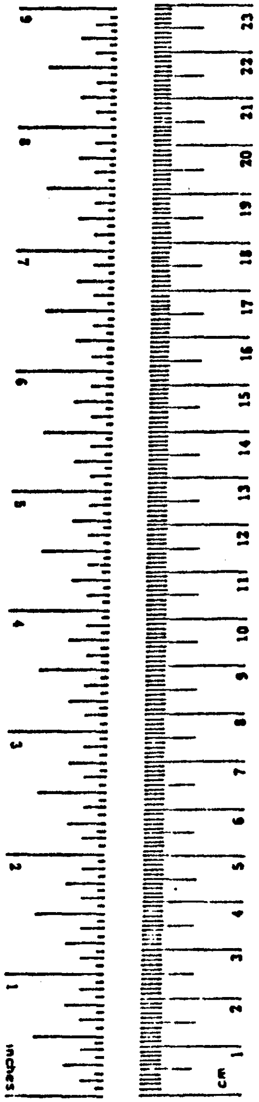
April 1986

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas

METRIC CONVERSION FACTORS

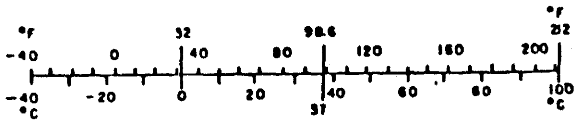
Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
	LENGTH			
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
	AREA			
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
	MASS (weight)			
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
	VOLUME			
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
	TEMPERATURE (exact)			
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
	LENGTH			
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
	AREA			
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
	MASS (weight)			
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
	VOLUME			
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
	TEMPERATURE (exact)			
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F



ACKNOWLEDGEMENTS

The authors would like to express appreciation to Mr. Jack Housworth, Engineer of Transportation Planning, State Department of Highways and Public Transportation, for serving as project contact for this research study. The authors also acknowledge the managerial and editorial assistance of Mr. Dock Burke, Manager of Policy and Management Program, TTI. Finally, the authors would like to thank Ms. Barbara Evins for her assistance in the preparation of this manuscript.

EXECUTIVE SUMMARY

It is important to examine the effectiveness and efficiency of Metropolitan Transit Authority (MTA) operations. However, simple comparative evaluations have generally been discounted because nationwide average data usually masks important differences between localities. The present study groups transit systems on the basis of socioeconomic and performance indicators, so that scores within these groups or profiles may be comparatively assessed.

Factor analysis followed by cluster analysis (using a selection of 24 socioeconomic variables from census data and 23 transit performance variables from UMTA Section 15 data) produced a reasonable grouping of 267 transit systems into seven profiles based on nine factors. In addition to the profile of average operations, six categories described systems with above or below average features. These included: 1) three systems with high labor performance, 2) eight MTA's that operate in large cities, 3) 45 systems that scored similarly on performance and operate in cities of dense growth, 3) three systems with high vehicle performance, 5) 18 systems that clustered on the basis of high social effectiveness, and 6) three small systems that had high service per subsidy dollar ratios.

The seventeen Texas transit systems included in the analysis separated into two profiles. Ten were in the "average" group. The other seven systems, located primarily along the south and west borders, were in the profile that described growing, dense cities. For the most part, factor scores were similar within and across the two sets of Texas systems. Some differences in vehicle and labor performance were found.

IMPLEMENTATION STATEMENT

This study resulted in the development of a methodological approach for evaluating transit performance that incorporates the environment in which systems operate. Use of this approach should assist transit operators and other decision-makers in evaluating individual system performance based on the performance of comparable operations. The methodology is useful for detecting systems with high performance that may be used as models for cities considering expansion or MTA establishment.

DISCLAIMER

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration or the State Department of Highways and Public Transportation. This report does not constitute a standard, specification, or regulation.

There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine, manufacture, design, or composition of matter, or any new and useful improvement thereof, or any variety of plant which is or may be patentable under the patent laws of the United States of America or any foreign country.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Midway into the decade of the 1980's, several cities in Texas are exploring alternative mechanisms for providing transit service to their increasing urban population. Prospects of reduced Federal assistance, limited State appropriations, and inadequate fare box revenues propel the search for new ways of organizing, financing, and delivering transit service to metropolitan areas of Texas. One way is to establish a transit authority, which effectively allows the State of Texas to collect a sales tax to partially fund the authority. There are several issues of importance for cities contemplating the establishment of a region-wide transit system, including the type and level of service to provide, and structure and size of a governing board, and the tax level required. As more urban areas of Texas seek legislative approval to establish Metropolitan Transit Authorities (MTA's), objective information and analyses are needed to clarify the advantages and disadvantages of this organizational unit to deliver transit services.

Study Purpose

A principal purpose of this research was to develop current, objective analyses that would describe information valuable for decisions involving the expansion, establishment, and performance of MTA's in Texas. This purpose was approached by means of examining data from transit systems nationwide. The research was designed to answer the following questions:

1) Can a profile be developed for transit system performance and socioeconomic conditions?
2) What transit systems are similar on the basis of this profile and why are they similar?
3) Can the similarities be represented geographically?
4) Can distinctions be made on the basis of organizational structure or taxing capabilities?
5) Where do Texas transit systems in general and MTA's in particular fit into the profiles, and how do they compare with other systems in the U.S.?

Transit Authority History

The first MTA enabling legislation for Texas was passed in 1973. This legislation authorized Metropolitan Rapid Transit Authorities for areas of at least 1.2 million people. The law was subsequently amended so that currently (1985) the population requirement is for areas with a density of at least 250 persons per square mile and at least 51 percent of the incorporated territory must be comprised of a city with at least 230,000 people.

Regional Transportation Authorities (RTA's) were created in 1979. RTA's differ legally from MTA's in several ways, the most notable difference being the population requirement. RTA's may be created in metropolitan statistical areas larger than 500,000 population, but not more than 60 percent of the population can reside in cities larger than 300,000. This legislation also enabled the creation of Subregional Transportation Authorities, which can be created in cities of at least 150,000 population that are adjacent to a city with a regional authority.

Both laws begin with the rationale for establishing authorities. The five justifying elements are (Vernon's Civil Statutes, 1984):

1. A large part of the State's population is located in growing metropolitan areas which often include several local jurisdictions and counties.
2. Highly concentrated populations are associated with high concentrations of motor vehicles that emit pollutants, resulting in dangers to the public health and welfare.
3. The concentration of motor vehicles places an undue burden on existing roadways, resulting in serious traffic congestion, retarding mobility and adversely affecting the health and welfare of the citizens and the economic life of the area.
4. Traffic congestion is primarily caused by the unavailability of efficient, inexpensive mass transit. It is in the public interest to provide mass transit for the benefit and convenience of the people, to improve air quality and reduce traffic congestion.

5. The right to use the air for natural purposes does not give any person the right to pollute the air by artificial means, and such artificial use is subject to regulation and control by the State.

The two Acts consist of 45 sections that relate to authority creation and organization, management, governing boards, powers and duties, elections, financing, annexation, vehicle emission taxes, bonds and notes, sales and use taxes, tax exemptions, contracting authority, and security personnel, among others. During the past 12 years there have been 53 bills filed that directly relate to MTA's. Many of these pertained to either territory or population requirements; many others addressed issues concerning taxation and finance and powers and duties of the authority unit. Twelve different amendments have made effective 53 changes in the two laws.

In other parts of the U.S., metropolitan transit authorities have been in existence since the late 1940's. The first transit authorities were established in major metropolitan areas, including: Chicago, Cleveland, Boston, New York, Los Angeles, Alameda-Contra Costa, San Francisco, and Fresno. Researchers at Berkeley (1961) described the characteristics of these first transit authorities. With the exception of the MTA in Cleveland, all of the agencies were established by state legislation. Most systems provided service to the central city and surrounding areas and were not restricted to type of transit service. The New York City Transit Authority, however, operated only within the city limits. The Bay Area Rapid Transit District was restricted to providing only intercity rapid transit. These systems supplied all or most of the local transit service in their territories. The governing boards of these authorities varied in size from 3 to 16 members, most of whom were appointed for terms of 4 to 7 years. The methods of funding ranged from operating revenues alone to a variety of methods including limited taxing power and the issuance of revenue bonds. Most of the agencies operated independently of state regulation but were controlled by local governments. The establishment of these authorities was thought to offer a workable approach to solving the problem of providing and operating area-wide transit systems.

As the number of area-wide transit systems has grown, so has the tax support at the local and state levels. Today there are at least 188 transit systems in the U.S. that receive operating assistance from dedicated taxes.

Table 1. Operating Subsidy from Dedicated Taxes
By System Size, 1980

Source of Operating Subsidy	Number of Vehicles Under 25 (N=71)	25-49 (N=46)	50-99 (N=30)	100-249 (N=24)	250-499 (N=10)	500-400 (N=5)	1000+ (N=2)	All Systems (N=188)
Percent of Total Operating Subsidy: Dedicated Local Taxes	17.2	10.9	18.3	8.0	21.1	14.8	57.7	28.0
Percent of Dedicated Local Taxes:								
Sales	75.1	26.3	74.9	36.1	23.5	30.9	25.5	30.6
Property	15.1	4.5	25.1	6.3	41.4	69.1	0	14.0
Other	9.8	69.2	0	57.7	35.1	0	74.5	55.5
Percent of Total Operating Subsidy: Dedicated State Taxes	2.6	2.0	1.7	6.2	7.8	0	2.8	3.7
Percent of Dedicated State Taxes:								
Sales	68.0	68.1	38.0	87.5	92.4	0	100.0	88.7
Gasoline	14.1	.5	0	1.9	0	0	0	.9
Other	18.0	31.4	62.0	10.6	7.6	0	0	10.3
Percent of Total Revenue: Taxes Levied Directly by Bus System	.6	2.0	1.3	.9	1.2	17.2	0	3.9

Source: UMTA Section 15 Annual Report, FYE 6/81 (1982)

Table 1 indicates the increased use of dedicated taxes to support transit such that systems of all sizes now receive this benefit. Table 1 also shows the various types of taxes that contribute to transit subsidy. The predominant type of tax is on sales, making up 30.6 and 88.7 percent of local and state dedicated funds for transit.

The predominant tax for area-wide transit systems in Texas is sales tax. Property tax may not be levied to pay for transit. Table 2 lists the sales tax amounts that have been collected from the two Texas MTA's as of 1983. These taxes are collected locally, and no State taxes are used for transit operating assistance.

The financial benefit to local transit that results from a dedicated tax source is obvious from the data shown in Table 2. A spokesman for the transit industry in Texas pointed out: "The primary reason that authorities are needed is because there are no State funds for operating assistance. Also, the Federal operating assistance continues to be in question. So, there must be a way for the local community to meet these needs, and the best way is to have the authority to levy a tax on themselves to pay for local transit" (Heil, 1984).

Table 2. Metropolitan Transit Authority Sales Tax Collections (in millions)

Year	Houston MTA Sales Tax Collections	Total Expenditures (Operating)	San Antonio MTA Sales Tax Collections	Total Expenditures (Operating)
1979	$ 97.9	$ 44.8	$13.7	$22.3
1980	129.9	64.8	17.0	24.5
1981	153.7	86.1	19.9	30.2
1982	166.8	104.2	20.8	31.4
1983	157.8	102.7	23.6	32.5

Source: Comptroller of Public Accounts (Mendez, 1984); Texas Transit Statistics (1980, 1982, 1983).

This authority, as noted previously, has been extended to an increased number of Texas urban areas. As the MTA law continues to be scrutinized, the option for smaller cities to include a tax in the creation of a transit authority can be expected to reappear in future sessions of the Texas Legislature. However, when put to the voters, the MTA option has not always been successful in the past. Specifically, in June 1983, voters in Houston turned down a $5.2 billion, eight-year bus and rail development plan that offered an 18.5 mile heavy rail passenger system. METRO officials searched for answers as to why the voters rejected the bond issue and concluded that: 1) the plan offered too little transit to too few commuters at too large an expense, and 2) the City of Houston and Harris County constituencies were too disenchanted with the existing system to support further expenditures of the magnitude planned.

Albeit disenchanting to voters, the existing system has allocated funds to its operating budget in greater proportion to other expenditures each year. The percent of public subsidy that went to METRO operations in 1981 was 78.3 percent. The amount of sales tax used for operations (at the expense of the capital program) is expected to increase every year, reaching 87 percent in 1995. One criticism often levied is that subsidy fosters inefficiency. Another is that expenditures rise to meet the income available. METRO and VIA have at times been cited as examples of this axiom. The fact is that Houston METRO has collected excess funds from the city sales tax since 1978 that total approximately $200 million (see Table 2). The original 1978 plan called for part of this fund to go to transit capital improvements as well as to street improvement projects, but a high proportion has consistently been used for operating subsidization.

While the MTA concept may provide a financial mechanism for making public transportation viable, it is yet to be proved that this form of public ownership will be successful at lowering costs and simultaneously improving service. It is important to examine the effectiveness and efficiency of MTA operations.

Chapter 2 describes the methodological approach used to examine effectiveness and efficiency by developing transit profiles. Chapter 3 identifies the profiles and their constituent systems.

2. METHOD: Analysis of Transit Data

Evaluating transit performance by peer group comparison has always been plagued with problems. The most frequently cited problem is that of nonuniformity among systems and systems' goals and objectives. Thus, uniform applications of performance indicators are often criticized as unfair. Hence, it is important to determine similar characteristics among transit systems and the environments in which they operate, furthering the effort to analyze comparable systems.

In the present study, a methodology was designed to identify a typology for transit systems to examine the operational and performance characteristics of transit systems in Texas. Two statistical procedures were used to produce a classification scheme for U.S. and Texas transit systems based on performance, socioeconomic, and climate data. With this procedure, the effects of relevant factors within the local environment that impinge on transit performance were ascertained, and a technique to enhance system comparability was produced.

Data Description

Two data sets were collected and merged for 294 transit systems throughout the United States. Twenty-six socioeconomic indicators and 32 transit performance measures were selected to develop the transit profiles. The socioeconomic data was collected from the U.S. Census County and City Data Book, 1983, and the State and Metropolitan Area Data Book, 1982, for the year 1980. Table 3 lists the socioeconomic variables and their abbreviated names used in the analysis.

The 26 variables in Table 3 were used to describe these nine different indicators of local environmental conditions for areas served by transit.

1. Population Size was measured by the total population of the MSA (metropolitan statistical area) and by the population of the central city of the MSA. For cities of less than 50,000 population, the total population and the population in the central city were the same. For county transit systems, the population of the county was used.

Table 3. Socioeconomic Variables

Variable Description	Variable Name
Total population	TPOP
Population in central city	POPCC
Population per square mile	POPMI
Population per square mile in central city	POPMICC
Persons per household	PHOUSE
Percent of permits for more than 5 unit residences	H5
Percent of housing owner occupied	OWNOCC
Average annual population change between 1970 and 1980	PCHANGE
Number of new housing permits	NEWPERM
Percent of permits for single unit residences	H1
Percent of population over 65 years	P65
Total population over 18 years	P18
Median age	MEDAGE
Crime rate	CRIMERT
Unemployment rate	UNEMP
Per capita money income	INCOME
Total gross revenue	TGR
Total taxes	TTAX
Journey to work - percent who drive alone	DRALONE
Journey to work - percent who use public transportation	PUBTR
Form of government	GOVT
Mean annual temperature	CLMN
Mean annual maximum temperature	CLMAX
Mean annual minimum temperature	CLMIN
Average annual precipitation (inches)	PRECIP
Average annual snowfall (inches)	SNOW

2. Population Density was measured directly with two variables - population per square mile in the total service area, and population per square mile in the central city served. Additionally, persons per household was used as a measure of internal density, and percent of owner occupied housing was considered an indirect measure of density. Also, the percent of new houses built that contained five or more units was an indicator of dense growth.

3. Population Age was measured by three variables - the percent of population over 65, the total population over 18, and the median age of the population.

4. Population Growth was expressed by the average annual change in population between 1970 and 1980, by the number of new residences built in 1980, and by the percent of single unit housing permits.

5. Crime Rate was measured by the number of serious crimes reported per 100,000 resident population in 1980. Serious crimes included murder, rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft.

6. The Economic Health of the local community was quantified using four variables - unemployment rate, per capita money income, the total gross revenue of the city government (1981), and the total taxes collected for the city government (1981). The latter two variables were combined into a single ratio of taxes as a portion of gross revenue.

7. Commuter data consisted of the two journey to work variables, percent who drive alone and percent who use public transportation.

8. The Type of Government was a classification of five types of organization as of 1982: 1) mayor with a city council; 2) city council with a city manager; 3) commission; 4) county; and 5) metropolitan statistical area.

9. Climate was described in terms of the average high and low temperatures and the overall average temperature for the locale over a 20 year period. Average annual rainfall and snowfall were also included.

The transit performance data were taken from the National Urban Mass Transit Statistics, Section 15 Report for the fiscal year which ended between July 1980 and June 1981 (1982). This report consists of data collected from 319 U.S. transit systems, and was furnished by UMTA in disk form.

Thirty-two variables were selected initially to describe three types of transit performance. The three concepts of performance used were delineated by Fielding (1983) -- cost efficiency, service effectiveness, and cost effectiveness. As Fielding explains: "Cost efficiency measures the resources expended to produce transit service (e.g., labor cost per hour); service effectiveness measures the extent to which service provided is used (e.g., passengers per hour); and cost effectiveness measures the service used against the resources expended (e.g., passengers per dollar operating cost)." Table 4 lists the performance variables by concept selected for the analysis.

In addition to the 30 variables listed in the Table 4, two others - the proportion of capital assistance that came from local general revenue and the proportion that came from local dedicated taxes were included. These two variables were named GREVTAX and DEDTAX.

The variables listed in Table 4 were further refined in Fielding's work as to the performance measured. Specifically, 12 performance indicators identified were:

1. Labor efficiency (I. 1,2,3)*
2. Vehicle efficiency (I. 4,5)
3. Fuel efficiency (I. 6,7)
4. Maintenance efficiency (I. 8,9)
5. Output per dollar cost (I. 10,11)

---

*Numbers in parentheses identify variable concept and number in Table 4.

Table 4. Transit Performance Variables

Performance Concept and Variable Description	Variable Name
I. COST EFFICIENCY	
1. Vehicle miles per employee	TVMEMP
2. Peak vehicles per operations personnel	PVEHOP
3. Peak vehicles per maintenance, support, and servicing personnel	PVEHMNT
4. Total vehicle hours per peak vehicle requirement	TVHPVEH
5. Total vehicle miles per peak vehicle requirement	TVMPVEH
6. Gallons of fuel consumed per vehicle mile	FUELTVM
7. Revenue vehicle miles per gallon of fuel	RVMFUEL
8. Total vehicles per $1000 maintenance expense	TVEHMEXP
9. Total annual vehicle miles per $1000 maintenance expense	TVMMNT
10. Operating expense per revenue vehicle hour	OEXPRVH
11. Revenue vehicle hours per operations labor expense	RVHOWAG
II. SERVICE EFFECTIVENESS	
1. Passenger trips per revenue vehicle hour	TPASRVH
2. Passenger trips per revenue vehicle mile	TPASRVM
3. Passenger trips per peak vehicle	TPASPVH
4. Revenue vehicle hours per service area population	RVHPOP
5. Passengers per service area population	TPASPOP
6. Passengers per elderly population	TPASELD
7. Passenger revenue per peak vehicle	REVPVEH
8. Passenger revenue per revenue vehicle hour	REVRVH
9. Operating revenue per revenue vehicle hour	TREVRVH
10. Passenger revenue per total operating assistance	REVOSUB
11. Passengers per total operating assistance	PASOSUB
12. Revenue vehicle hours per total operating assistance	RVHOSUB
13. Revenue vehicle hours per accident	RVHACC
14. Accidents per million vehicle miles	ACCTVM
III. COST EFFECTIVENESS	
1. Total operating expense per passenger	OEXPPAS
2. Passengers per total labor expense	PASTWAG
3. Passengers per gallon of fuel	PASFUEL
4. Operating revenue per total operating expense	REVOEXP
5. Total revenue per total operating expense	REVTEX

6. Utilization of service (II. 1,2,3)
7. Social effectiveness (II. 4,5,6)
8. Revenue generation (II. 7,8,9)
9. Public assistance (II. 10,11,12)
10. Operating safety (II. 13,14)
11. Service consumption per expense (III. I,2,3)
12. Revenue generation per expense (III. 4,5)

These 12 indicators were used by Fielding to classify individual performance measures. This classification scheme provided a useful means of labeling the variables used in the present study.

Statistical Method

A general statistical methodology was developed to organize and classify the transit system variables. Specifically, a two-step methodology was used on various combinations of the variables. First, factor analysis was used to reduce the data into a smaller set of new variables referred to as factors or dimensions.

The factor analysis used principal components for the initial extraction and a varimax rotation of the principal components. It was expected that variables such as those used in this analysis would be multicollinear in nature and factor analysis was used to eliminate this multicollinearity. Large multivariate data sets typically contain groups of variables that are correllated internally but the groups are independent of each other. Principal components identifies the independent groups of correlated variables and the varimax rotation emphasizes the separation of the groups. The groups are considered representative of latent constructs in the original data because of collinearity of the variables in the group. In this study, each factor or dimension represents a linear combination of variables that summarized or accounted for the original set of variables, known as transit profiles.

For the second phase of analysis, a cluster analysis was performed to identify similar profiles and multivariate extreme points. As described in the following section, the extreme points were removed and the factor and cluster analyses were repeated until a meaningful factor and cluster pattern were found. The outcome of this procedure was a classification scheme for transit systems, based on performance, socioeconomic, and climate data.

Next, cluster means and standard deviations were used to identify unusual factor scores for individual systems. Analysis of Variance (ANOVA) was used to compare Texas systems with other systems.

Data Screening

A data screening step was performed by examining bar graphs and descriptive statistics for the variables in the data. After errors were detected and corrected, the final data set contained many outliers and extremes. However, these were meaningful and correct observations. For instance, the New York Transit Authority serves an area of nine million inhabitants. Although such extremes weaken the assumption of multivariate normality for factor analysis, they were not deleted from the data set because the objective of the analysis was to find a profile that would identify differences, including outliers, among transit systems.

However, several variables were deleted from the analysis due to missing or unreliable data. GRVTAX (local general revenue tax) and DEDTAX (local dedicated tax) were deleted after the first factor analysis because only 188 systems reported these data, none of which were Texas systems. Furthermore, the frequency of the reported value of 0.0 was problematic in that there was no way to determine if no taxes were allocated or no taxes were reported. The factor analysis performed on the 188 systems did not reveal summary factors that included either tax variable as a significant contributor.

Missing values were also frequent for the variables TPASPVH (passenger trips per peak vehicle), REVPVEH (passenger revenue per peak vehicle), OEXPRVH (operating expense per revenue vehicle hour), OEXPPAS (operating expense per passenger), REVTEX (total revenue per total operating expense), TVMMNT (total annual vehicle miles per maintenance expense), and PVEHMNT (peak vehicles per maintenance, support, and servicing personnel). Initial factor analyses revealed that each of these variables contributed to a separate factor, illustrating multicollinearity with other variables. Therefore, their removal from the data set did not result in a significant loss of information. The deletion of these seven variables was preferable to the elimination of transit systems where these data were missing from the data set.

Two Texas systems had several missing data points even after the above variables were deleted. Because of the importance of the inclusion of the Texas systems these variables were estimated using mean values for the variables over all systems. The variables estimated were: REVOEXP (revenue per operating expense) and TVEHMEXP (vehicles per maintenance expense) for El Paso; and PASFUEL (passengers per gallon of fuel), RVMFUEL (revenue miles per gallon of fuel), REVOEXP (revenue per operating expense), RVHOWAG (revenue vehicle hours per operations labor), PASTWAG (passengers per labor expense), TVEHMEXP (vehicles per maintenance expense), and RVHACC (vehicle hours per accident) for the Dallas Transit System.

Finally, the Form of Government variable was deleted for the factor analysis because its values were not continuous. Thus, the final data set consisted of 24 socioeconomic variables and 23 performance variables for 267 transit systems (see Table 5).

3. RESULTS: Transit Profiles

The previous chapter described a methodological approach for examining transit operations by analyzing a large number of socioeconomic, climatic, and operational variables. The first step of the analysis reduced these variables to a meaningful pattern.

Factor Analysis

The set of 47 socioeconomic and performance variables was reduced to nine factors that represent the underlying latent constructs present in the original set. The SAS Procedure Factor (SAS Institute, 1982) was used to derive the factor matrix, from which the new variables were taken. The factor matrix was rotated so that the reference axes of the factor were turned about the origin and the best linear combinations were found.

The decision of the number of dimensions or linear combinations used to explain the initial set of variables is usually determined by the latent root criterion. More specifically, the latent root criterion identifies significant factors based on the eigenvalues of the sample correlation matrix. The sum of the eigenvalues of the sample correlation matrix represent a general measurement of the variability in the original data set. The latent root criterion selects the number of significant factors if a suitable percent of the total variability is explained by the factors, and if all of the latent roots (eigenvalues) are greater than one (Johnson, 1982). When factors have eignvalues less than one, then it may be assumed that those factors cannot explain the variance of a least one variable, and thus, those factors can be disregarded.

In this case, the nine factors selected for the final iteration explained 68 percent of the variance (see Table 5). Although 13 factors had eigenvalues greater than one and explained 79 percent of the variance, the last four factor groupings had little meaning with high loadings on single variables. The variable groupings for these factors were difficult to interpret due to their incompatibility and high single loadings. They were sensitive to corrections in the data because of their narrow definition. Therefore, they were eliminated to achieve a more stable and meaningful factor profile. Table 5 contains the variable groupings, the names chosen

Table 5. Results of Varimax Rotation

Factor Name and Percent Variance Explained		Variable Name	Variable Description	Factor Loading
1. Size	21.1%	T18	Total population over age 18	.9124
		TPOP	Total population	.9094
		POPCC	Population in center city	.8265
		PUBTR	Percent who use public transportation	.7515
		POPMICC	Population per square mile in center city	.6341
		REVRVH	Passenger revenue per revenue vehicle hour	.5975
		NEWPERM	Number of new housing permits	.5827
		TREVRVH	Operating revenue per revenue vehicle hour	.5736
		DRALONE	Percent who drive alone	-.6879
2. Climate	9.9%	CLMN	Mean annual temperature	.9488
		CLMAX	Mean annual maximum temperature	.9464
		CLMIN	Mean annual minimum temperature	.9310
		PCHANGE	Percent change in population, 1970-1980	.5345
		crimert	Crime rate	.4428
		SNOW	Average annual snowfall	-.8138
3. Public Assistance	8.2%	RVHOSUB	Revenue vehicle hours per operating assistance	.8896
		REVOSUB	Passenger revenue per operating assistance	.8346
		PASOSUB	Passengers per operating assistance	.8060
		REVOEXP	Operating revenue per operating expense	.6629
4. Peak Performance	5.8%	TVHPVEH	Total vehicle hours per peak vehicle	.7733
		TVMPVEH	Total vehicle miles per peak vehicle	.7364
		PVEHOP	Peak vehicles per operating personnel	.6825
		TPASRVH	Passenger trips per revenue vehicle hour	.6062

Table 5. (continued)

		TPASRVM	Passenger trips per revenue vehicle mile	.5695
		rvhacc	Revenue vehicle hours per accident	-.3927
5. Social Effectiveness	5.4%	RVHPOP	Revenue vehicle hours per population	.8325
		TPASPOP	Passengers per total population	.8220
		TPASELD	Passengers per elderly population	.8091
6. Density	5.2%	POPMI	Population per square mile	.7768
		H5	New housing permits for more than 5 units	.5308
		OWNOCC	Percent housing owner occupied	-.5784
		H1	New housing permits for 1 unit	-.7294
7. Age	4.4%	MEDAGE	Median age	.9153
		P65	Percent of population over 65 years	.8594
		PHOUSE	Persons per household	-.7269
8. Labor Performance	4.3%	RVHOWAG	Revenue vehicle hours per operations labor expense	.9565
		PASTWAG	Passengers per total labor expense	.8948
		TVEHMEXP	Total vehicles per maintenance expense	.6296
		tvmemp	Vehicle miles per employee	.3521
9. Vehicle Performance	3.8%	RVMFUEL	Revenue vehicle miles per gallon of fuel	.6555
		PASFUEL	Passengers per gallon of fuel	.5835
		ACCTVM	Accidents per million vehicle miles	.5359
		fueltvm	Gallons of fuel consumed per vehicle mile	.4198
		income	Per capita money income	.3598
		unemp	Unemployment rate	-.2707
		precip	Average annual precipitation	-.3973

for each factor, cumulative variance explained, and the factor loadings for each variable.

Factor loadings represent the correlation between the original variables and its respective factor. Factor loadings greater than ±.50 were considered significant and were used to describe a particular factor. The choice of ±.50 is arbitrary. However, this choice is conservative and represents a clearly interpretable factor pattern. Table 5 shows that only seven of the 46 variables in this study were not useful in identifying the latent constructs in the data. Four of the seven were contained in the last factor, which explained only 3.8 percent of the variance in the data.

The nine factors can each be described by their variable composites. Factor 1 (F1) is defined by variables that represent population size. As would be expected, a positive association is shown between area population, center city population and density, and numbers of new houses. High revenue generation is also associated with transit operations in large, dense cities where a greater percent of the population use public transportation to get to work and fewer commute by car. The Size factor is heavily loaded and accounts for over 21 percent of the variance in the data.

Factor 2 (F2) summarizes the Climate variables. Population change is also included in this factor and reflects growth in the warmer temperatures of the sunbelt. There was a minor loading by the Crime Rate variable for this factor. However, its loading was not significant enough to be descriptive.

Factor 3 (F3) represents the Public Assistance variables. Included with the three operating subsidy variables is the passenger per operating expense ratio.

The fourth factor (F4) was named Peak Performance as three of its variable constituents are measures of performance for peak operations. The two other variables of significance in this factor, passenger trips per revenue vehicle hour and passenger trips per revenue vehicle mile, reflect heavy usage during peak periods.

Factor 5 (F5) is composed of the three variables described as Social Effectiveness measures. This factor summarizes transit service provided in terms of the population of the area served.

Factors 6 (F6) and 7 (F7) are demographic dimensions measuring density and age. Factor 6 is a measure of housing density with positive loadings for persons per square mile and increased number of apartment buildings, and negative loadings for new and currently occupied single-family owned homes. The two direct measures of Age are median age and the percent of population over 65 years. Persons per household is an indirect indicator of the age of the population. A large number of persons per household is usually indicative of a greater number of children per family, which has the effect of reducing median age and the percent of the population over 65 years.

Factor 8 (F8) is a summary of variables that are related to Labor Performance. The number of hours and passengers per dollar of labor are heavily loaded in this factor. To a less extent, the number of vehicles per maintenance dollar contributes to this dimension. Because maintenance expenses are somewhat dominated by maintenance labor, this ratio encompasses the labor performance concept.

Factor 9 (F9) is defined by two fuel ratios and a measurement of operating safety. Although the two fuel ratios are classified as measures of fuel efficiency, the three variables as a summary factor are called measures of Vehicle Performance, reflecting the simultaneous relationship of passengers and miles per gallon of fuel and accidents per mile.

The nine factors represent clear and logical groupings of the variables of this study. An interesting characteristic of the factor pattern is that the socioeconomic and transit performance data loaded on the same factor in only one of the nine factors retained. The Size factor (F1) contained information from both data sets that measured population size in terms of totals. Factor 9 (F9) also contained a mixture of socioeconomic and performance variables; however, the variables income, unemployment, and precipitation all had loadings of less that ±.5 and were not influential to the factor summary. This separation of socioeconomic and performance variables suggests that transit system performance is, to a large extent, independent of the local socioeconomic conditions of the system.

The above nine dimensions were used to categorize the 267 transit properties. The second phase of analysis was to categorize the systems by using cluster analysis on the factor scores.

Cluster Analysis

Cluster analysis is a classification technique than can be used to display and summarize data by describing the natural groupings or clusters that are known to exist within the data set. Anderberg offers several philosophical guidelines to the utilization of cluster analysis (Anderberg, 1973, 22-23). Principally, he postulates that clustering methods simultaneously impose a structure on the data, and reveal the structure or "natural groups" that exist in the data.

The 267 transit systems were segmented into meaningful groupings that manifest similar characteristics for the nine factor dimensions described above. The factor scores produced by the factor analysis for each transit system were the input for this cluster analysis. The clustering algorithm is designed to minimize the within group variance and concurrently maximize the between group variance. When the within group variance is minimized, the systems in the group are closer to the nine dimensional cluster mean. When the between group variance is maximized, the distance between the groups is made larger. The clustering algorithm classifies systems so that these two criteria are satisfied. This technique provides a highly useful approach to arrive at a meaningful description of the data, in this case transit systems, and to discover unsuspected clusterings.

One of the most important uses of cluster analysis is the flagging of systems with incorrect data. Such systems tend to be classified into clusters having only one member. Much of the preliminary analysis of this study was the detection and correction of this data. Also, a part of the final analysis of this study was to check the data for clusters with small memberships.

The maximum number of clusters that are to be used in the cluster analysis must be specified *a priori*. This initial decision appears to be subjective in nature. There is no reliable statistical method to determine the optimal number of clusters that will best classify the data (Everitt, 1977; Anderberg, 1973). The clustering algorithm computes the cubic clustering criterion and some analysts (Everitt, 1977; Anderberg, 1973) argue that the breakpoint of the cubic clustering criterion is an indicator or the "best" cluster pattern. This statistic follows a linear trend until some optimal cluster pattern is achieved. A break in the clustering criteria suggests that at the point the different group or clusters are clearly distinct. There can be several breakpoints and therefore this statistic should be used in conjunction with a priori theory. Cluster analysis is a descriptive tool; hence, rational theory based on a priori information about the subject sample is the usual criterion for determining the best descriptive cluster pattern.

Findings from the cluster analysis suggested a classification pattern for the 267 systems. Table 6 identifies the cluster sizes and the mean values for each cluster. Those clusters with high positive or negative factor means can be interpreted as having a strong correlation with the factor. The factor scores for most of the clusters have an average near zero and a standard deviation near 0.75, so an unusually large average factor score is outside the interval of ±1.50. However clusters C3 and C5 had all factor score averages within the interval ±1.5. Cluster C3 was named on the basis of its largest mean factor score of 1.09 for F6. None of the factor score means were significantly different from zero for Cluster C5 on the basis of a 90 percent confidence interval about the mean.

The first cluster isolated three systems that had very high labor performance ratios, as evidenced by the high score for Factor 8. These three systems -- Phoenix (AZ) Transit; City of Rochester (MN) Transit; and Fargo (ND) Area Transit, are located on the map in Figure 1.

Cluster C2 represents the large transit systems. Each factor and cluster iteration performed resulted in the isolation of these eight systems on the basis of Size (F1). They are mapped in Figure 1.

The third cluster is composed of 45 systems that correlate most highly on Density (F6). Included in this cluster are smaller systems in New York (smaller than the regional authority). This cluster also showed a high negative value for the Age factor (F7). Seven Texas transit systems were in this cluster. They are: Brownsville, Corpus Christi, El Paso, Fort Worth, Galveston, Laredo, and San Antonio. The 45 systems that comprise Cluster C3 are shown in Figure 2.

Cluster C4 had a membership of three systems -- City of Glendale, Arizona; Golden Gate Bridge HTD; and Cape Ann Transportation in Massachusetts. These three systems distinguished themselves on the basis of vehicle performance (F9). The systems in Glendale and Cape Ann each have a small fleet of gasoline powered vehicles and therefore scored high on fuel

Table 6. Cluster Means

Cluster	Number of Systems	Factor Means F1	F2	F3	F4	F5	F6	F7	F8	F9
C1	3	+0.29	-0.13	-0.04	-0.23	-0.30	+0.00	-0.17	+8.73	+0.32
C2	8	+3.97	+0.12	-0.31	+0.43	+1.16	-0.53	+0.11	-0.09	-0.01
C3	45	-0.04	+0.23	+0.10	+0.03	-0.37	+1.09	-0.74	-0.12	-0.20
C4	3	+1.00	+0.37	+0.41	-.077	-1.40	-0.20	+0.21	-0.61	+6.18
C5	187	-0.13	-0.07	-0.12	-0.04	-0.15	-0.31	+0.15	-0.09	-0.08
C6	18	-0.32	+0.27	-0.17	+0.42	+2.05	+0.76	+0.12	-0.12	+0.30
C7	3	-0.74	-0.53	+7.86	-0.55	+1.35	-0.49	-0.16	+0.33	+0.88

Figure 1. Location of Clusters C1 and C2 Systems

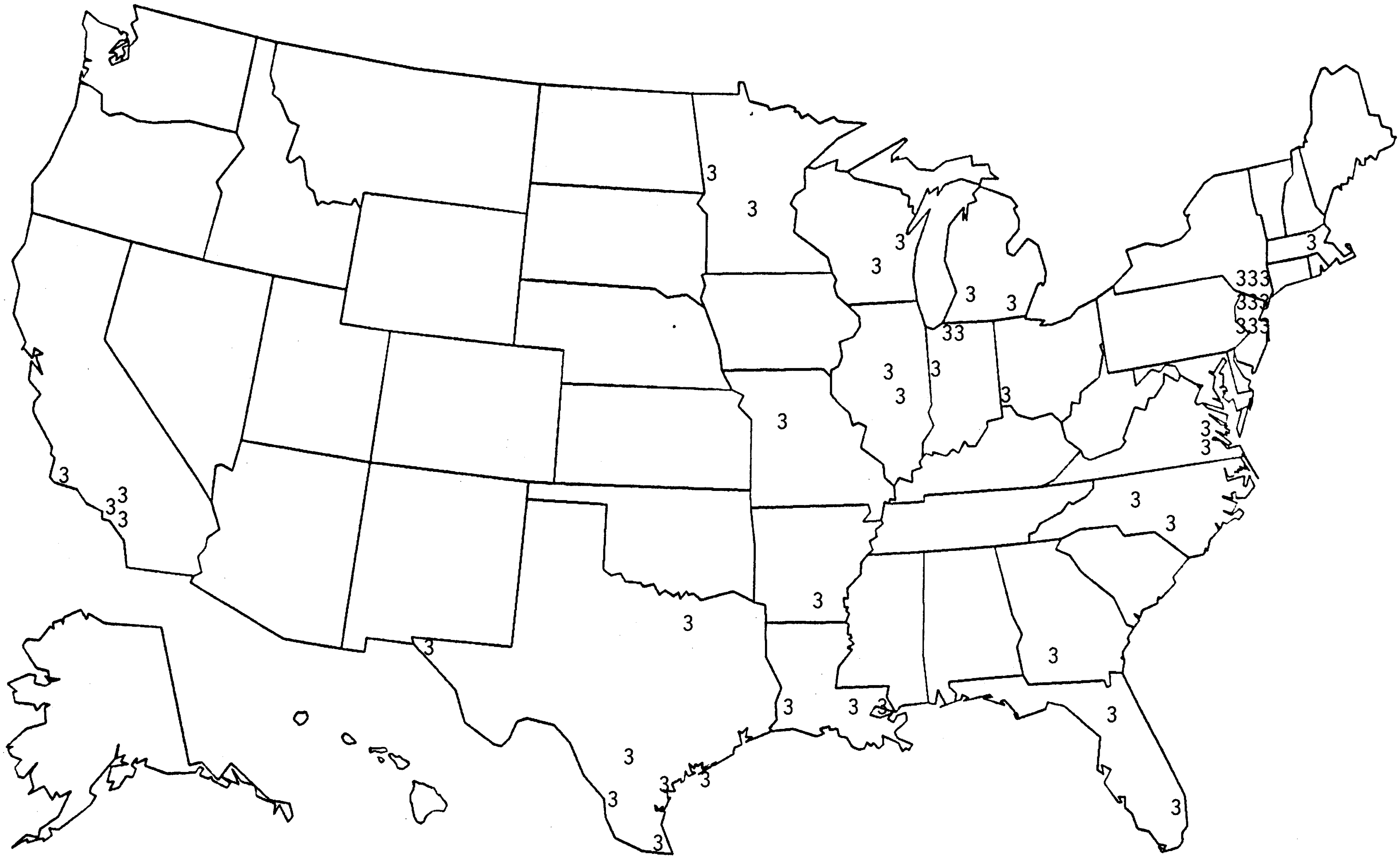


Figure 2. Location of Cluster C3 Systems

Figure 3. Location of Clusters C4, C5, and C6 Systems

efficiency. The Golden Gate Bridge system operates four ferry boats in addition to its bus fleet. These increased its overall passenger per gallon of fuel ratio. Cluster C4 systems are located in Figure 3.

The fifth cluster (C5) basically represents the average metropolitan transit system. Seventy percent of the system cities were classified in this group. The means for this cluster were near zero, reflecting average values for the nine factors. Ten Texas systems were located in this cluster. They were: Abilene, Amarillo, Austin, Dallas, Houston, Lubbock, Port Arthur, San Angelo, Waco, and Wichita Falls. No discernable geographic pattern was evidenced for this cluster with membership throughout the United States.

The chief similarity among the 18 systems in the sixth cluster (C6) was a greater than average degree of social effectiveness (F5). A large proportion of these systems were located in California (see Figure 3).

The subsidy factor scores were high for the three systems in Cluster C7--Pelham Parkway Bus Service and Riverdale Transit, both in Mt. Vernon, New York; and Campus Bus Service in Kent, Ohio. These cities also clustered with very low scores on the Size factor. (See Figure 3 for their location.)

Table 7 provides a descriptive summary of the seven clusters.

Table 7. Cluster Descriptions

Cluster	Members	Description
C1	3	High labor performance ratios
C2	8	Large systems in large cities
C3	45	Smaller systems in dense growth cities; younger population
C4	3	High vehicle performance
C5	187	Average systems
C6	18	High social effectiveness
C7	3	High performance to public assistance ratios

Analysis of Clustering of Texas Cities

As mentioned above, ten Texas systems were represented in Cluster C5 and seven were represented in Cluster C3. Figure 4 identifies the location of the 17 Texas systems by cluster membership. The map illustrates that Cluster C5 systems were found primarily in the eastern section of the State, while Cluster C3 systems were located primarily in the western and southwestern section of the State.

Several comparisons of the Texas systems were performed. Three approaches were used to further investigate the performance of Texas systems. These three approaches examined the following:

1. How did Texas transit systems load on factor scores? Examination of the factor scores revealed for each variable how each Texas system compared with the other systems within its same cluster.
2. What are the differences between the two clusters of Texas systems? A one-way Analysis of Variance (ANOVA) was used on the nine sets of factor scores that identified the reasons for the clustering in terms of which factors were significantly different.
3. How do Texas systems compare with all other systems sampled? This comparison was a two-way ANOVA using clusters and states (Texas and all others) as classification variables.

In the first comparison to determine how Texas systems compared to other systems within their profiles, the systems within a cluster were compared to an interval on two standard deviations about the cluster average. These intervals should contain 95 percent of the factor scores of the cluster, and systems with factor scores outside the interval may be considered unusual for the cluster. Table 8 gives Texas systems' scores for each of the nine factors. The intervals for Clusters C3 and C5 are given in Tables 9 and 10.

The analysis showed one unusual factor for the Texas systems in Cluster C3. The Laredo system had an unusually small age factor score (F7) of -3.582 (Table 8), indicative of lower median age than the average for the cluster and higher than the cluster average for number of persons per household. Otherwise, the systems in Cluster C3 had individual factor scores (Table 8) within the intervals found in Table 9.

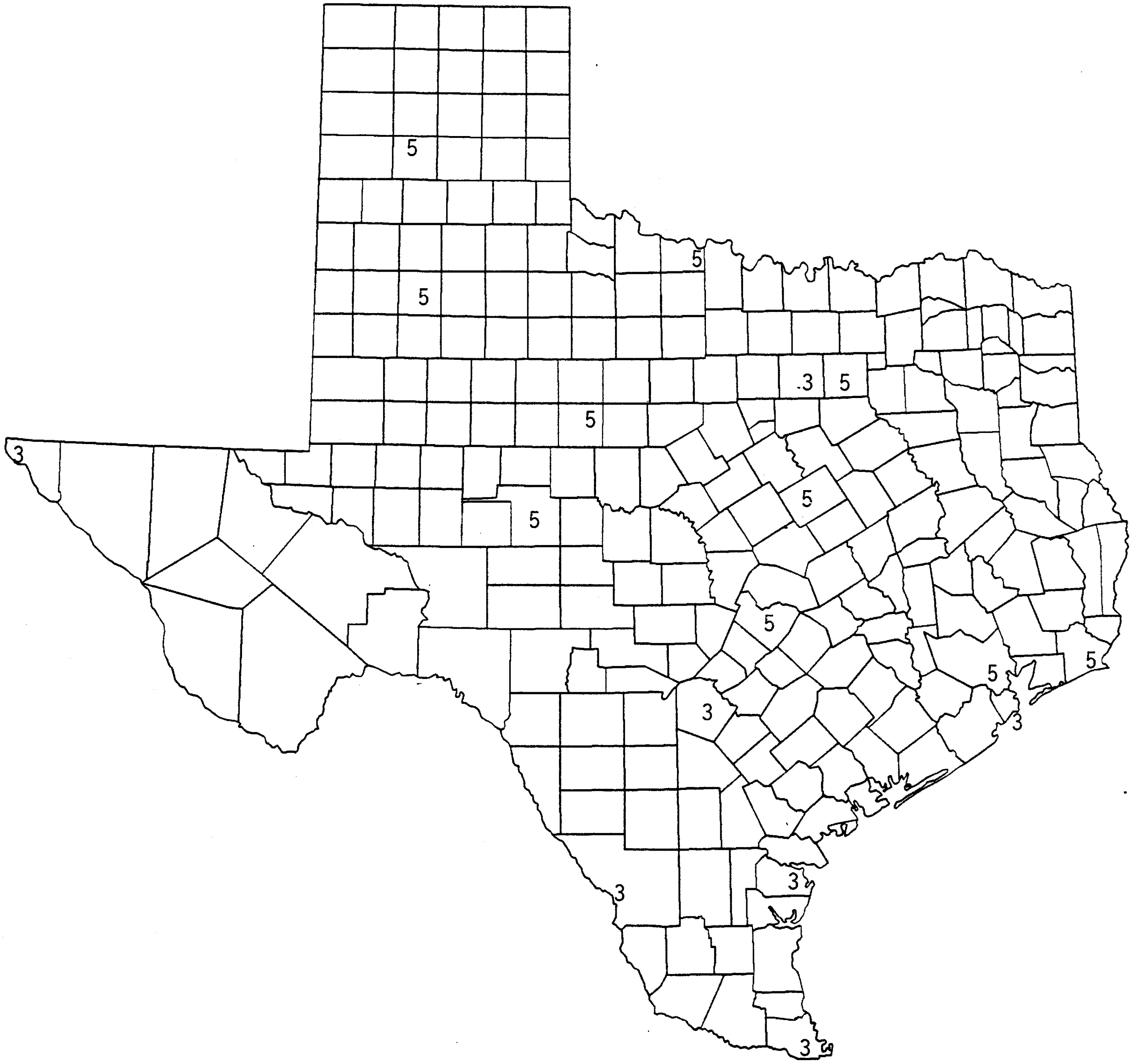


Figure 4. Location of Clusters C3 and C5 Texas Systems

Table 8. Cluster and Factor Scores for Texas Systems

City	Cluster	F1	F2	F3	F4	F5	F6	F7	F8	F9
Brownsville	C3	+0.240	+2.050	+0.383	+0.393	-0.906	-0.137	-2.485	-0.477	-1.409
Corpus Christi	C3	-0.568	+1.641	-0.273	-0.279	-0.580	+0.619	-1.456	-0.157	+0.143
El Paso	C3	+0.099	+0.906	+0.232	+0.233	-0.020	-0.349	-2.597	-0.093	+0.110
Ft. Worth	C3	-0.535	+1.209	+0.001	+0.285	+0.197	+0.837	+0.038	-0.105	+0.053
Galveston	C3	-0.898	+1.593	+0.095	+0.803	-0.022	+1.855	+0.404	-0.104	-0.279
Laredo	C3	+0.013	+1.586	+1.107	+1.107	-0.398	-0.369	-3.582	-0.394	-1.101
San Antonio	C3	+0.062	+1.513	-0.407	-0.407	+0.155	+0.275	-1.232	-0.047	-0.034
Abilene	C5	-0.972	+0.755	-0.547	-0.547	-0.396	+0.031	-0.055	+0.333	+0.252
Amarillo	C5	-1.133	+0.120	-0.561	-0.561	-0.070	+0.437	-0.157	+0.485	+0.589
Austin	C5	-0.311	+1.558	-0.128	-0.128	+0.075	+0.175	-0.511	+0.061	+0.548
Dallas	C5	+1.312	+1.640	+0.264	+0.264	+0.243	-1.281	-0.394	+0.729	+1.411
Houston	C5	+2.06	+2.119	-1.593	-0.593	-0.279	-1.422	-0.434	+0.141	+1.364
Lubbock	C5	-0.720	+0.489	-0.129	-0.129	+0.023	-0.207	-1.130	+0.117	+1.617
Port Arthur	C5	+0.105	+1.191	+0.044	+0.044	-0.593	-1.249	+0.123	-0.555	-1.596
San Angelo	C5	-0.456	+0.934	-0.198	-0.198	+0.144	-0.191	+1.095	+0.200	-0.624
Waco	C5	-0.826	+1.275	-0.176	-0.176	-0.594	+0.518	+0.091	-0.090	-0.225
Wichita Falls	C5	-0.221	+0.851	+0.640	+0.640	+0.004	-0.919	+0.108	-0.072	-0.405

Table 9. Cluster C3 Factor Score Ranges

Factor	Average Score	Standard Deviation	Interval Lower Limit	Interval Upper Limit
F1. Size	-0.038	0.926	-1.89	1.81
F2. Climate	0.230	1.028	-1.83	2.29
F3. Public Assistance	0.102	0.691	-1.28	1.484
F4. Peak Performance	0.030	0.828	-1.63	1.69
F5. Social Effectiveness	-0.370	0.620	-1.61	0.87
F6. Density	1.090	1.500	-1.91	4.09
F7. Age	-0.736	1.004	-2.74	1.27
F8. Labor Performance	-0.121	0.257	-0.64	0.39
F9. Vehicle Performance	-0.202	0.636	-1.47	1.07

Table 10. Cluster 5 Factor Score Ranges

Factor	Average Score	Standard Deviation	Interval Lower Limit	Interval Upper Limit
F1. Size	-0.131	0.511	-1.15	0.89
F2. Climate	-0.069	0.976	-2.02	1.88
F3. Public Assistance	-0.124	0.450	-1.02	0.78
F4. Peak Performance	-0.036	0.962	-1.96	1.89
F5. Social Effectiveness	-0.151	0.524	-1.20	0.90
F6. Density	-0.311	0.516	-1.34	0.72
F7. Age	0.146	0.975	-1.80	2.10
F8. Labor Performance	-0.088	0.245	-0.58	+0.40
F9. Vehicle Performance	-0.079	0.680	-1.44	1.28

Cluster C5 had several unusual factors scores for Texas systems. This cluster was a profile that described average U.S. systems. However, the Houston and Dallas systems had high factor scores for size (F1), and Houston had a high factor score for climate (F2). Houston also had a low density (F6) score. Regarding transit performance, two systems had high labor performance (F8) scores - Amarillo and Dallas. Port Arthur had a low vehicle performance (F9) score. Dallas, Houston, and Lubbock had high vehicle performance (F9) scores. All of the Cluster C5 Texas systems were in the average range for the public assistance (F3), peak performance (F4), age (F7), and social effectiveness (F5) factors.

Analysis of variance was used to look at the Texas systems as a group compared to their respective cluster member systems. Demographic differences were found that were not surprising. Tables 11 and 12 illustrate that in both Cluster C3 and C5, Texas systems overall had higher scores on climate (F2). The Texas systems in Cluster C3 served a younger population (F7) than other states. Texas systems as a group had higher average labor (F8) and vehicle performance scores (F9) than other systems in Cluster C5.

The results of the analysis of factor scores for Texas systems in Cluster C3 compared with those in Cluster C5 are summarized in Table 13. The systems are considered statistically different if a value of less than 0.1 is found for the probability that the means of the two clusters are equal. In this case, the clusters have significantly different density (F6), age (F7), and labor performance (F8) factor scores. The Cluster C3 systems tend to serve more densely populated and younger areas. Cluster C3's youth and density scores for Texas may be explained by the fact that six of the seven Texas systems in C3 are southern border cities that have higher than average numbers of persons per household, which is also usually indicative of greater than average number of children per family. The Cluster C5 systems had a higher labor performance factor score average than Cluster C3, meaning they have higher values for the number of hours, passengers, and vehicles per dollar of labor expense.

Table 11. Summary of One-Way ANOVA for Texas Cluster C3 Systems and Other Systems

Factor	State	Mean	Standard Error of Mean	Probability that means are equal
F1	Texas	-0.227	.353	.5636
	Other	-0.003	.151	
F2	Texas	+1.497	.332	.0002
	Other	-0.003	.142	
F3	Texas	+0.164	.264	.7987
	Other	+0.090	.113	
F4	Texas	+0.372	.311	.2379
	Other	-0.033	.134	
F5	Texas	-0.231	.236	.5243
	Other	-0.395	.101	
F6	Texas	+0.387	.562	.1801
	Other	+1.220	.241	
F7	Texas	-1.556	.359	.0169
	Other	-0.585	.154	
F8	Texas	-0.197	.097	.4025
	Other	-0.107	.042	
F9	Texas	-0.362	.242	.4763
	Other	-0.173	.104	

Table 12. Summary of One-Way ANOVA for Texas Cluster C5 Systems and Other Systems

Factor	State	Mean	Standard Error of Mean	Probability that means are equal
F1	Texas	-0.116	.162	.9241
	Other	-0.132	.038	
F2	Texas	+1.093	.097	.0001
	Other	-0.134	.070	
F3	Texas	-0.138	.142	.9167
	Other	-0.123	.034	
F4	Texas	-0.163	.305	.5680
	Other	-0.029	.072	
F5	Texas	-0.144	.166	.9686
	Other	-0.151	.039	
F6	Texas	-0.411	.163	.5302
	Other	-0.305	.039	
F7	Texas	-0.226	.308	.2155
	Other	+0.167	.073	
F8	Texas	+0.135	.076	.0028
	Other	-0.101	.018	
F9	Texas	+0.273	.214	.0927
	Other	-0.099	.051	

Table 13. Summary of One-Way ANOVA for Cluster 3 and 5 Texas Systems

Factor	Cluster	Mean	Standard Error of Mean	Probability that means are equal
1	3	-0.227	0.320	0.7944
	5	-0.116	0.268	
2	3	1.497	0.192	0.1289
	5	1.093	0.161	
3	3	0.164	0.165	0.1799
	5	-0.139	0.138	
4	3	0.372	0.306	0.2001
	5	-0.163	0.256	
5	3	-0.231	0.132	0.6222
	5	-0.144	0.110	
6	3	0.387	0.289	0.0516
	5	-0.411	0.242	
7	3	-1.555	0.364	0.0134
	5	-0.226	0.304	
8	3	-0.196	0.110	0.0349
	5	0.135	0.092	
9	3	-0.362	0.342	0.1747
	5	0.273	0.286	

The third comparison of factor scores was a two-way ANOVA to test for interaction between the cluster and State groups. This technique allowed a simultaneous analysis of each of the factor scores of the two Texas groups in comparison with the other 232 systems in the two clusters. The two-way ANOVA incorporates the hypothesis of the first two comparisons into a single test procedure that uses all of the systems in Cluster C3 and Cluster C5. It also allows a test for uniformity of the conclusions across the clusters. That is, if differences are found in the average factor scores for the Texas systems in Clusters C3 and C5, then the average factor scores for the other states also may be compared simultaneously in Cluster C3 and C5.

Figures 5 and 6 illustrate the only two factor scores that had statistically significant interactions - Density (F6) and Labor Performance (F8). Figure 5 depicts the interaction of the Labor Performance (F8) factor mean scores. Labor performance averages for systems in other States were not significantly different among the Cluster C3 and Cluster C5 groupings, and the Texas systems within Cluster C3 did not differ from them significantly. However, Cluster C5 factor scores for Texas systems were significantly higher on labor performance than all other systems.

Figure 6 shows that the factor mean scores for Density (F6) among Cluster C5 systems were not different between Texas and other states' systems. Mean scores for Density (F6) for Cluster C3 Texas systems were higher than all other Cluster C3 and C5 Density scores.

This third analysis further substantiates that within the Average system cluster, Texas (C5) systems scored above average on labor performance when compared to other Texas and other average systems. Furthermore, Texas transit system cities characterized by high density and high density growth were distinctive in this respect compared to other Texas cities and other similar cities in the U.S. in 1980.

Summary

To summarize, the comparison of Texas transit systems to each other and to other systems in the U.S. resulted in several rather obvious general conclusions. However, some specific findings pertaining to performance were discovered as well.

Figure 5. Interaction Effect of Labor Performance Between Clusters C3 and C5 Texas Systems

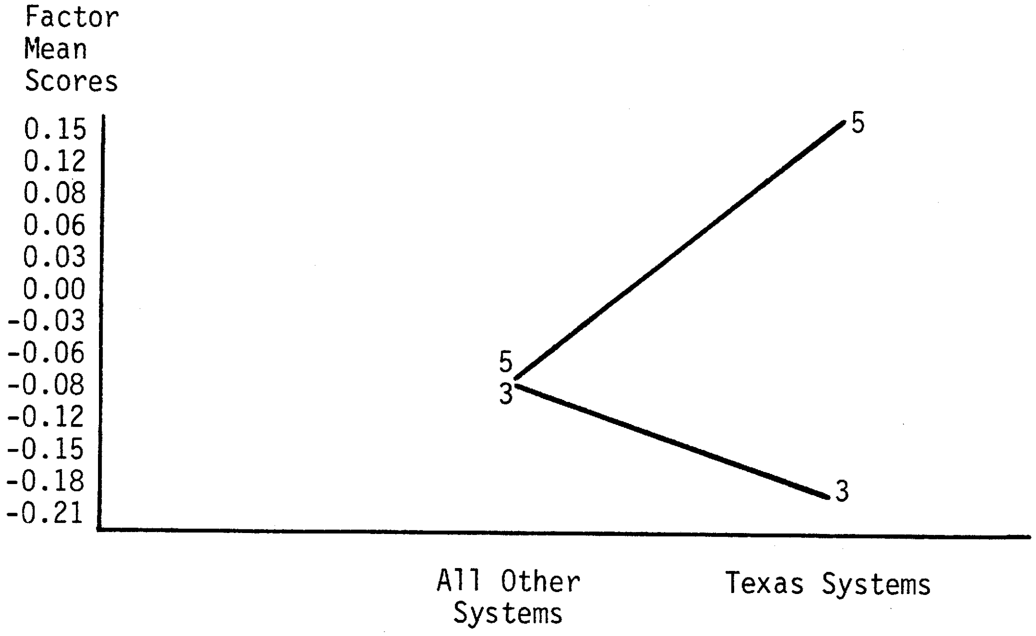
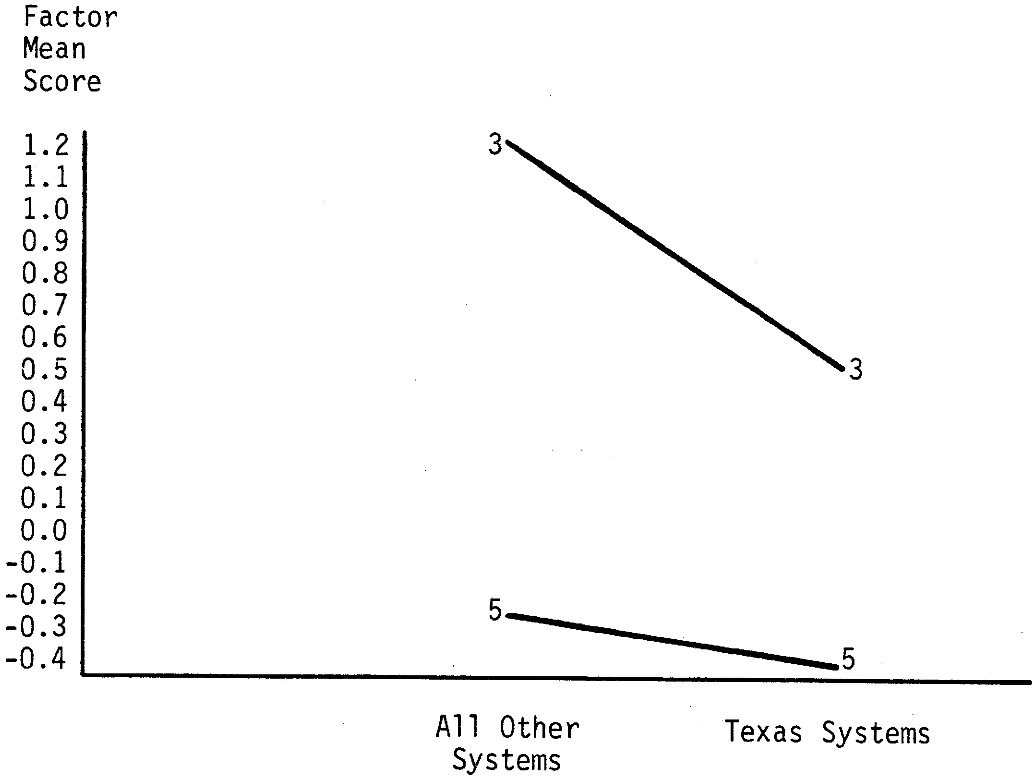


Figure 6. Interaction Effect of Density Between Clusters C3 and C5 Texas Systems



All the Texas systems were classified in the two largest profile groups or clusters. Texas systems are more similar to other systems nationwide than they are different. In fact, 10 of Texas' 17 systems were categorized in the profile of the average system. The remaining seven systems were very similar to each other in terms of performance. They distinguished themselves from other Texas systems primarily on the basis of two factors, higher density and lower age values.

Although considered average overall, the 10 systems in Texas that were in Cluster C5 scored higher as a group on labor and vehicle performance. In particular, Houston, Dallas, and Lubbock had higher than average factor scores on vehicle performance. Dallas and Amarillo had higher than average factor scores on labor performance. Port Arthur reported lower than average vehicle performance.

4. SUMMARY AND CONCLUSIONS

This study was undertaken to develop current and objective analyses that would describe information useful to transit operators and public administrators. Specifically, data from transit systems nationwide were analyzed to simultaneously examine the performance of transit systems within the context of the environment in which they perform. One objective was to determine if transit systems can be classified into groups of similar performers, or profiles, based on geographic and/or demographic differences. If so, the next step was to determine the distinctions between systems, particularly distinctions based on organizational structure or taxing capabilities. The major objective of the study was to find out where Texas transit systems in general and MTA's in particular fit into the profiles, and how they compare with other systems in the U.S.

Summary of Findings

The analysis revealed that U.S. cities that provide transit service could be distinguished on the basis of socioeconomic data, and to a lesser extent by several operational aspects of their transit systems. Factor analysis summarized the final selection of 24 socioeconomic variables and 23 performance variables into nine dimensions or factors. These factors were:

F1. Size
F2. Climate
F3. Public Assistance
F4. Peak Performance
F5. Social Effectiveness
F6. Density
F7. Age
F8. Labor Performance
F9. Vehicle Performance

Using cluster analysis, the 267 transit systems that reported complete data to UMTA in 1980 were categorized into seven clusters. These clusters represented meaningful groupings of systems that manifest similar characteristics for the nine factors. The seven clusters were described as the following:

C1. Systems with high labor performance ratios
C2. Large systems in large cities
C3. Smaller systems in dense growth cities with a younger population
C4. Systems with high vehicle performance ratios
C5. Average systems
C6. Systems ranking high on social effectiveness
C7. Systems with high performance to public assistance ratios

The techniques used in this analysis were very sensitive to the number of variables and systems included in the study. Several steps were required to find acceptable factor and cluster patterns. These methods were useful in the detection of outlier and extreme points. Although variances and means are very sensitive to extreme points in the formulation of factors, it should be pointed out that the first three factors (F1, F2, and F3) were isolated in all of the data screening steps. The other factors tended to change order, but also appeared in the last two data screening steps. Hence, the factor pattern was reasonably stable. The cluster pattern was more sensitive to the addition or deletion of systems from the data set. This is, however, the nature of cluster analysis since it is often used to detect multivariate extreme points. The final cluster pattern grouped the transit systems in a reasonable way, so it was accepted.

There were no clusters with only a single member. The clusters having fewer members centered around a single large factor score average so they can be considered good examples of systems having the properties of the variables used to construct the factor. The large clusters represented systems with nearly average performance and socio economic factor scores. The clusters were not chosen to bias the results in favor of Texas systems which could be achieved by taking advantage of the sensitivity of the clustering algorithm, i.e., deleting systems.

No Texas transit system was classified into profiles that characterized those systems with exceptional transit performance. Generally, Texas transit systems had average performance and socioeconomic scores. Ten Texas systems were classified as average on the basis of their membership in Cluster C5, the largest cluster. Seven Texas systems were located in the second largest cluster (C3) that grouped cities with dense, younger populations that experienced growth in multi-family housing in 1980. Six of these seven Texas systems were located in the south and west portions of the State, primarily along the border.

With respect to operational performance, the 10 systems in Texas belonging to Cluster C5 scored higher as a group on labor (F8) and vehicle performance (F9) than did other U.S. systems in the same cluster (C5). In particular, Houston, Dallas, and Lubbock had higher than average factor scores on vehicle performance (F9) while Port Arthur had a lower than average score on this factor. Dallas and Amarillo had higher than average factor scores on labor performance (F8).

No apparent differences in performance were found on the basis of organizational structure or the presence of an MTA in Texas cities. In other parts of the country, the very large transit systems that operate in very large cities (or authorities) separated themselves from smaller systems operating in the same cities on indicators of per mile, per passenger, and per hour service. For example, in New York City the Regional Authority coexists with several smaller, independent systems that serve various segments of the area. The authority and smaller systems differ on performance. Texas cities with MTA's do not have this authority and independent system mixture.

Socioeconomic and performance data loaded together in only one factor, (F1), which summarizes Size. This separation of socioeconomic and performance variables suggests that transit system performance is, to a large extent, independent of local socioeconomics and climate.

Implementation and Recommendations

The procedures used in this study to reduce a sufficiently large amount of Section 15 data and census data into meaningful profiles yielded techniques that can be used to make comparisons of transit systems. The clustering analysis technique created a basis for within-group evaluation of system performance. In addition, comparisons of factor scores for individual systems were produced.

In addition to comparisons of performance, factor analysis combined with cluster analysis can be used as a method for detecting outlier systems; i.e., systems with exceptionally good scores on performance factors. These systems can be used as models for existing system improvements as well as examples for new systems.

Future studies of this type will benefit from more accurate and complete Section 15 data. Data for this study was taken from the report of the second year that systems receiving Federal monies were required to report to UMTA. As data in subsequent years becomes available and is more reliable, comparisons over time both within and between similar transit systems should be helpful in identifying emerging problem areas and evaluating performance overall.

This study focused on a quantitative analysis of MTA and other types of transit systems nationwide. Research is needed to investigate further the more qualitative aspects of transit authority structure.

REFERENCES

1. Anderberg, Michael K. Cluster Analysis for Applications. New York: Academic Press, 1973.

2. County and City Data Book, 1983. U.S. Department of Commerce, Bureau of the Census.

3. Everitt, B.C. "Cluster Analysis." In The Analysis of Survey Data: Volume 1, Exploring Data Structures, pp. 63-88, C.A. O'Muircheartaigh and Clive Payne, eds. London: John Wiley and Sons, 1977.

4. Fielding, Gordon J. and Shirley C. Anderson. "Public Transit Performance Evaluation: Application to Section 15 Data." Transportation Research Record 947, 1983.

5. Heil, Larry. "Urban Areas in Texas: Public Transportation," Testimony presented to Transportation Committee of the Texas House of Representatives, March 1984.

6. Johnson, R.A. and D. W. Wichern. Applied Multivariate Statistical Analysis, Prentice-Hall: New Jersey, 1982.

7. Kennedy, N. and Homburger, W.S. The Organization of Metropolitan Transit Agencies, Institute of Transportation and Traffic Engineering, University of California, Berkeley, 1961.

8. Mendez, Maria C. "Can Mass Transit Deliver in Texas?" Fiscal Notes, Issue 84:3, Office of the State Comptroller, Austin, TX, March 1984.

9. National Urban Mass Transportation Statistics: Second Annual Report, Section 15 Reporting System. UMTA, U.S. Department of Transportation, Rept. UMTA-MA-06-0107-82-1, 1982.

10. SAS Institute, Inc., SAS User's Guide: Statistics, 1982 Edition, Cary, NC: 1982.

11. State and Metropolitan Area Data Book, 1982. U.S. Department of Commerce, Bureau of the Census.

12. Texas Transit Statistics. State Department of Highways and Public Transportation, 1980, 1982, 1983.

13. Vernon's Civil Statutes, Articles 1118x and 1118y, 1984, pp. 394-448.

APPENDIX A

SOCIOECONOMIC DATA FOR TEXAS CITIES

	POPULATION					AGE					HOUSING			
CITY	Total	Per Square Mile	In Center City	Per Sq. Mile In Center City	Annual Percent Change 1970-80	Percent 65+	Total 18+	Median	Persons per House-hold	Crime Rate	Percent Owner Occupied	New Permits	Percent 1 Structure	Percent 5+ Structures
Abilene	139192	51	98315	1295	+1.30	12.4	100	28.7	2.68	4503	68.2	956	59.6	36.9
Amarillo	173669	190	149230	1861	+1.80	9.7	124	28.6	2.64	5946	67.5	1273	48.5	47.8
Austin	536688	191	345496	2978	+4.00	7.8	394	26.6	2.61	7130	54.9	8269	50.9	33.7
Brownsville	209727	232	146528	2184	+4.00	9.6	129	25.0	3.56	6044	64.7	2504	63.8	34.0
Corpus Christi	326228	212	231999	2031	+1.40	8.4	217	26.9	3.08	7130	62.2	3635	45.2	51.3
Dallas	1556390	357	902619	2836	+2.20	8.3	1107	28.6	2.72	8267	62.3	17422	47.5	49.1
El Paso	479899	473	425259	1778	+2.90	6.6	311	25.0	3.32	5887	59.4	3148	73.0	12.0
Ft. Worth	385164	1604	385164	1604	-0.21	11.8	281	29.3	2.60	11833	61.1	2255	50.3	35.7
Galveston	61902	1284	61902	1284	+0.02	13.1	46	29.8	2.49	9672	45.2	419	22.7	74.2
Houston	2905353	430	1595138	2866	+3.70	6.2	2019	27.5	2.79	6222	58.8	38339	52.5	46.4
Laredo	99258	30	91449	4642	+3.10	8.4	60	23.6	3.79	5918	62.0	488	77.0	10.0
Lubbock	211651	235	173979	920	+1.70	7.9	151	25.5	2.76	7239	60.8	1236	76.5	10.0
Port Arthur	61251	1076	61251	1076	+0.68	13.5	44	30.4	2.74	5836	68.2	99	100.0	0.0
San Angelo	84784	56	73240	1932	+0.18	12.1	62	28.9	2.67	5596	64.4	758	68.9	25.1
San Antonio	1071954	426	786023	2992	+1.90	9.0	731	27.4	2.97	6411	64.0	8300	43.5	51.8
Waco	170755	166	101261	1367	+1.50	13.2	125	29.3	2.65	7147	63.3	589	43.5	43.1
Wichita Falls	130664	77	94201	1915	+0.20	11.8	96	29.2	2.63	7547	66.8	435	94.9	0.0
Texas Sample Average	506149	417	336650	2092	+1.79	10.0	353	27.7	2.86	6960	62.0	5301	59.9	33.0
U.S. Sample Average	653758	1792	309773	4219	+1.11	11.2	466	30.2	2.72	6435	62.5	3047	56.6	29.1

APPENDIX A - (continued)

SOCIOECONOMIC DATA FOR TEXAS CITIES

CITY	Per Capita Money Income (1979)	Unemployment rate	1976/77 GROSS REVENUE		JOURNEY TO WORK		CLIMATE				
			Total (Mil.)	Total Taxes (Mil.)	Percent Who Drive Alone	Percent Using Public Transit	Average Temp.	Mean Maximum Temp.	Mean Minimum Temp.	Average Annual Rainfall (inches)	Average Annual Snow (inches)
Abilene	6683	3.4	58	23	74.9	0.7	64.5	75.9	53.0	24.1	4.4
Amarillo	7779	4.2	110	45	74.7	1.0	57.2	70.0	44.3	20.4	14.6
Austin	7363	3.7	285	118	67.1	4.4	68.1	78.6	57.5	33.3	1.0
Brownsville	4336	10.1	119	31	66.9	2.1	73.5	82.2	64.8	26.8	0.0
Corpus Christi	6506	6.0	234	87	72.5	1.4	71.2	78.6	63.8	27.3	0.1
Dallas	8325	4.5	1710	796	67.5	8.3	65.7	76.1	55.2	32.0	3.5
El Paso	5367	9.2	253	80	68.0	4.2	63.8	76.7	50.9	8.5	4.6
Ft. Worth	7336	4.1	155	70	69.2	3.6	65.7	76.1	55.2	32.0	3.5
Galveston	7292	4.3	52	11	58.3	4.5	69.8	74.7	64.9	44.5	0.3
Houston	8950	4.2	1611	896	68.4	4.8	67.7	78.7	56.7	46.9	0.4
Laredo	3980	11.3	52	15	63.1	4.6	70.6	83.4	57.9	20.7	0.1
Lubbock	6929	4.4	107	47	73.9	1.2	59.8	73.5	48.0	17.7	9.7
Port Arthur	6538	10.3	29	16	69.6	1.5	68.7	77.3	60.1	52.5	0.5
San Angelo	6848	3.8	36	14	70.3	0.6	65.3	78.5	52.1	20.4	3.2
San Antonio	6227	6.6	577	194	67.6	6.0	69.0	79.5	58.5	27.8	0.5
Waco	6236	5.4	83	24	74.5	1.5	67.3	77.9	56.7	33.5	1.5
Wichita Falls	7099	4.0	72	26	67.9	1.1	64.2	76.4	51.9	26.6	5.8
Texas Sample Average	6694	5.8	326	147	69.1	3.03	62.4	72.7	52.2	27.5	3.1
U.S. Sample Average	7571	6.9	716	331	63.5	8.33	55.7	65.4	45.5	35.7	25.1

APPENDIX B

TRANSIT PERFORMANCE DATA FOR TEXAS SYSTEMS*

System City	RVH/ OWAG	OEXP/ RVH	TREV/ RVH	REV/ RVH	TPAS/ RVM	OEXP/ PAS	PAS/ TWAG	TPAS/ RVH	REV/ PVEH	TVM/ PVEH	TVH/ PVEH	PVEH/ OP	FUEL/ TVM	RVM/ FUEL
Abilene	.117	16.00	2.06	1.95	1.15	0.82	2.27	19.55	7.45	65229	3825	.46	.28	.28
Amarillo	.144	17.20	4.70	2.47	0.85	1.55	1.60	11.13	8.75	58095	3829	.59	.16	.20
Austin	.067	51.60	6.59	5.32	2.54	9.49	1.96	29.12	20.55	44224	3595	.29	.47	.32
Brownsville	.055	26.70	11.66	9.48	3.45	0.62	2.40	43.40	NA	NA	NA	NA	.20	.20
Corpus Christi	.080	29.80	7.08	4.37	1.09	4.92	1.22	15.24	15.05	42162	2970	.34	.37	.23
Dallas	.087	NA	11.07	9.11	3.09	NA	2.30	44.59	19.80	71360	4947	.78	.60	.29
El Paso	.075	NA	11.20	10.67	2.28	NA	2.38	31.73	40.51	54880	3824	.25	.34	.35
Ft. Worth	.059	28.70	10.49	8.65	2.90	0.78	2.16	36.77	19.15	32953	2780	.38	.30	.35
Galveston	.095	20.80	10.03	9.68	2.84	0.70	2.81	29.71	40.20	43384	4152	.32	.29	.29
Houston	.027	57.70	14.57	13.80	3.29	1.29	1.19	44.78	40.62	45858	3383	.20	.36	.41
Laredo	.101	18.00	11.27	11.04	3.03	0.57	3.18	31.35	50.43	47250	4623	.26	.24	.24
Lubbock	.112	81.80	3.69	3.14	1.17	44.56	1.84	16.33	9.31	41100	2992	.39	1.28	.37
Port Arthur	.076	23.80	4.72	4.70	1.40	1.22	1.48	19.52	NA	NA	NA	NA	.15	.15
San Angelo	.117	17.80	3.49	2.78	1.76	0.73	2.84	24.18	NA	NA	NA	NA	.15	.15
San Antonio	.068	NA	9.11	6.61	2.21	NA	2.14	31.47	18.68	78297	4895	.38	.35	.30
Waco	.091	35.10	8.51	4.92	1.84	6.63	2.12	23.35	12.98	32667	2510	.37	.42	.21
Wichita Falls	.088	17.10	6.70	6.16	0.68	1.75	0.86	9.77	NA	NA	NA	NA	.15	.15
Texas Sample Average	.086	31.58	8.05	6.76	2.09	5.40	2.04	27.18	23.34	50574	3717	.38	.36	.26
U.S. Sample Average	.115	48.91	11.86	10.83	2.30	3.66	2.60	30.00	34.33	NA	NA	.34	.35	.28

APPENDIX B - (continued)

TRANSIT PERFORMANCE DATA FOR TEXAS SYSTEMS

System City	PAS/ FUEL	REV/ OSUB	PAS/ OSUB	RVH/ OSUB	RVH/ POP	TPAS/ POP	TPAS/ ELD	PVEH/ MNT	TVEH/ MEXP	TVM/ EMP	REV/ OEXP	REV/ TEX	RVH/ ACC	ACC/ TVM
Abilene	.24	.09	0.93	.047	.16	3.22	0.12	NA	.13	30.11	.13	1.45	3279	17.9
Amarillo	.23	.19	0.86	.077	.35	3.85	0.40	NA	.19	34.05	.27	1.04	1625	37.4
Austin	.12	.32	1.80	.062	.42	12.37	1.59	1.90	.09	14.17	.29	1.04	1640	117.8
Brownsville	.06	.62	2.83	.065	.20	8.74	0.91	NA	.07	7.49	.44	1.01	1456	54.5
Corpus Christi	.21	.33	1.16	.076	.33	4.99	0.59	1.72	.11	16.39	.37	1.42	731	144.4
Dallas	.33	1.71	8.36	.187	.57	25.39	3.06	4.59	.17	27.69	.34	1.00	1876	235.2
El Paso	.15	.63	1.86	.059	.59	18.83	2.85	1.32	.17	13.58	.34	NA	2019	34.3
Ft. Worth	.12	.49	2.08	.057	.52	19.22	1.63	1.94	.10	12.55	.36	1.00	1428	47.1
Galveston	.10	.84	2.56	.086	.67	19.93	1.52	1.33	.06	13.99	.48	1.04	1122	85.3
Houston	.12	.30	0.98	.022	.36	16.14	2.60	0.82	.04	9.41	.25	0.40	328	195.6
Laredo	.08	1.65	4.69	.150	.92	28.85	3.43	1.25	.03	12.27	.63	1.00	643	30.2
Lubbock	.32	.22	1.16	.071	.36	5.94	0.75	2.01	.20	16.29	.22	1.06	1571	150.2
Port Arthur	.11	.24	1.00	.051	.23	4.54	0.34	NA	.10	11.67	.20	1.02	2377	196.9
San Angelo	.08	.20	1.78	.074	.22	5.24	0.43	NA	.10	21.27	.20	0.96	2623	27.4
San Antonio	.13	.29	1.38	.044	.96	30.36	3.37	1.90	.10	16.50	.38	1.39	1120	81.0
Waco	.11	.32	1.54	.066	.20	4.69	0.35	1.86	.16	12.50	.39	1.10	612	131.8
Wichita Falls	.22	1.17	1.85	.189	.15	1.45	0.26	NA	.14	17.68	.39	0.70	3230	21.2
Texas Sample Average	.16	.56	2.17	.081	.42	12.57	1.42	1.88	.11	16.92	.33	1.04	1628	94.6
U.S. Sample Average	.32	.91	2.27	.075	.56	19.40	1.84	1.78	.17	18.59	.34	1.06	1927	86.0

***Key to Heading Abbreviations:**

RVH/OWAG	revenue vehicle hours per operations labor expense
OEXP/RVH	operating expense per revenue vehicle hour
TREV/RVH	operating revenue per revenue vehicle hour
REV/RVH	passenger revenue per revenue vehicle hour
TPAS/RVM	passenger trips per revenue vehicle mile
OEXP/PAS	operating expense per passenger
PAS/TWAG	passengers per total labor expense
TPAS/RVH	passenger trips per revenue vehicle hour
REV/PVEH	passenger revenue per peak vehicle
TVM/PVEH	total vehicle miles per peak vehicle
TVH/PVEH	total vehicle hours per peak vehicle
PVEH/OP	peak vehicles per operating personnel
FUEL/TVM	gallons of fuel consumed per vehicle mile
RVM/FUEL	revenue vehicle miles per gallon of fuel
PAS/FUEL	passengers per gallon of fuel
REV/OSUB	passenger revenue per total operating assistance
PAS/OSUB	passengers per total operating assistance
RVH/OSUB	revenue vehicle hours per total operating assistance
RVH/POP	revenue vehicle hours per service area population
TPAS/POP	passengers per service area population
TPAS/ELD	passengers per elderly population (65+)
PVEH/MNT	peak vehicles per maintenance, support, and servicing personnel
TVEH/MEXP	total vehicles per $1000 maintenance expense
TVM/EMP	vehicle miles per employee
REV/OEXP	operating revenue per total operating expense
REV/TEX	total revenue (including public assistance) per total operating expense
RVH/ACC	revenue vehicle hours per accident
ACC/TVM	accidents per million vehicle miles