
CSTDM09 - California Statewide Travel Demand Model

Model Development

Travel Behavior Datasets

Final System Documentation: Technical Note

ULTRANS
Institute of Transportation
Studies, UC Davis
Davis, California

HBA Specto Incorporated
Calgary, Alberta

May 2011

Table of Contents

1. Introduction	8
2. Survey Data Processing	8
2.1 Trips.....	9
2.1.1 Trip File	9
2.1.2 Time Periods	10
2.1.3 Trip Purposes.....	10
2.1.4 Linked Trips.....	11
2.2 Activities	13
3. Development of the Master Database.....	20
3.1 Aggregation of Data.....	20
3.1.1 Define Household and Person Table Attributes Uniformly	20
3.1.2 Eliminate “Out of Scope” / Incomplete Survey Data	21
3.1.3 Create New Attributes Based on Available Attribute Values	21
3.1.4 Create One Master Database with Three Tables	23
3.1.5 Develop Expansion Factors for Each Household & Person	25
3.2 Data Expression of Master Database	25
3.2.1 A Note on the Ordinal Income Category for SANDAG Data.....	27
4. Tour Table Preparation	37
4.1 NumberOfTripsPerPersonPerDay	38
4.2 TourStart.....	38
4.3 TourID.....	38
4.4 NumberOfToursPerPersonPerDay	38
4.5 TourPurpose	39
4.6 WhetherCompleteIncompleteTour	40
4.7 TourModeCoding	42
4.8 TourModeName	44
4.9 TourOrigin.....	45
4.10 TourDestination	45
4.11 PurposeOfOtherTour	45
4.12 WhetherSubtourExists	45

4.13 NumberOfSubTours.....	45
4.14 ForwardBackwardPass.....	45
4.15 MoreThan100Mile.....	48
4.16 TourModeAcceptance.....	49
4.17 AcceptedTour	49
5. Tour and Day Pattern Development.....	50
5.1 Specifications.....	50
5.2 Methodology	52
5.3 Tour Data Format and Structure	55
5.4 California Journey Characteristics	56
5.4.1 Key Indicators	56
5.4.2 Tour Purpose	56
5.4.3 Tour "Overview"	57
5.4.4 Tour Occurrence Frequency	58
5.4.5 Travel Activities	60
5.4.6 Time of day Patterns	64
6. Tour Statistics	62
7. Notes on Travel Activities	70
7.1 Step 1:	71
7.2 Step 2:	72
Appendix A: Travel Survey Expansion.....	74

List of Tables

Table 1: Sample Activity File Records.....	11
Table 2: Sample Trip Records.....	12
Table 3: Single Trip from a Sequence of Records.....	12
Table 4: Activity Matching between Surveys	14
Table 5: Activity Categories	15
Table 6: Trip Mode Definition and Its Coding	25
Table 7: Trip Activity Definition and Its Coding.....	26
Table 8: Dwelling Type of a Household.....	26
Table 9: Tenure of a House	26
Table 10: Ordinal Income Category of a Household	27
Table 11: SANDAG & SCAG Income Comparison.....	27
Table 12: Synthetic Household Frequencies by Persons, Workers and Income... 	31
Table 13: Probability of a Household Earning \$10-15k with One Worker.....	32
Table 14: Summary of SANDAG Record Reclassification	33
Table 15: Final Income Categories	34
Table 16: Work Status of a Person	34
Table 17: Type of Work Status	34
Table 18: Work Industry of a Worker	35
Table 19: Work Occupation of a Worker	36
Table 20: Student Status of a Person.....	36
Table 21: Type of Student Status.....	36
Table 22: School Type of Student.....	37
Table 23: Person Type of All Persons in the Database.....	37

Table 24: A Typical Trip-Tour Table.....	41
Table 25: Trip Modes and Corresponding Rank.....	42
Table 26: Tour Mode Name and Code	44
Table 27: A Typical Trip-Tour Table.....	47
Table 28: A Trip-Tour Table with Trips Over 100 Miles	49
Table 29: Time of Day Definition	50
Table 30: Travel Activity Definition.....	51
Table 31: Tour Purpose Definition.....	51
Table 32: Tour Mode	52
Table 33: Tour Data File Structure.....	55
Table 34: Tour Proportions by Purpose	57
Table 35: Tour Overview Proportions.....	57
Table 36: Tour Occurrence Frequency.....	58
Table 38: Proportion of Travel Activities.....	60
Table 39: Percentage Share of Trip Records by Trip Data Source	63
Table 40: Percentage Share of Tour Records by Trip Data Source	64
Table 41: Distribution of Tours by Tour Purpose and Tour Completeness	64
Table 42: Distribution of Persons by Number of Tours/Person/Day.....	65
Table 43: Distribution of Tours by Number of Tours/Person/Day.....	65
Table 44: Distribution of Tours by Number of Trips/Tour.....	65
Table 45: Distribution of Tour Modes	66
Table 46: Distribution of Tours by Tour Purpose and Tour Completeness	67
Table 47: Distribution of the Number of Trips per Tour	67
Table 48: Distribution of the Number of Persons by Number of Tours	68
Table 49: Percentage Share of Tour Mode.....	68

Table 50: Percentage Share of Tour Mode by Tour Purpose	69
Table 51: Percentage Share of the Existence of Subtour by Tour Purpose.....	70
Table 52: Distribution of the Number of Subtours	70
Table 53: Trip Table for Travel Activity	72
Table 54: Trip Shares by Travel Activity	73
Table 55: Socioeconomic Groups.....	75
Table 56: Number of Survey Households by Socioeconomic Group	76
Table 57: Target and Survey Households by Geographic Region	77
Table 58: Expansion Factors by Socioeconomic Group	82

List of Figures

Figure 1: SANDAG TAZ Districts	30
Figure 3: Top 20 Occurrence Tours by Person Type	59
Figure 4: Travel activities by Person Type.....	61
Figure 5: Outbound Time Distribution.....	62
Figure 6: Return Time Distribution	62
Figure 7: Geographic Regions in California	78
Figure 8: Statewide geographic group map.....	80
Figure 9: Geographic groups in San Francisco and Sacramento areas	81
Figure 10: Geographic groups in LA / San Diego area	81
Figure 11: Expansion Factors by Socioeconomic Group (1 person households). 83	
Figure 12: Expansion Factors by Socioeconomic Group (2 person households). 84	
Figure 13: Expansion Factors by Socioeconomic Group (3 person households). 84	
Figure 14: Expansion Factors by Socioeconomic Group (4 person households). 85	
Figure 15: Expansion Factors by Socioeconomic Group (5+ person households)	85
Figure 16: Expansion Factors by Socioeconomic Group (senior only households)	86
Figure 17: Expansion Factors by Geographic Region	88
Figure 19: Expansion Factors in the Los Angeles Region	89
Figure 20: Ranges of Expansion Factors.....	90
Figure 21: Distribution of Expansion Factors.....	92

1. Introduction

The travel behavior dataset for short-distance personal trips was assembled from multiple sources, which had the advantage of improved data quality but presented challenges related to dataset combination. Ultimately, what is required from this dataset is an understanding of how people travel in California, including where they are traveling to and from, for what purpose, during what time of day, and using what mode of transportation. This information was gathered from four travel surveys: the California Statewide Travel Survey, San Diego Association of Governments (SANDAG) Travel Survey, the Southern California Association of Government (SCAG) Travel Survey, and the Metropolitan Transportation Commission (MTC) Bay Area Travel Survey.

This document explains how the information obtained in these four travel surveys was extracted and reassembled into the format necessary for the Short Distance Personal Trip Model (SDPTM) in the California Statewide Travel Demand Model (CSTDM). Due to the different sources of data, some aspects of the surveys were more detailed or more complete than others, and some surveys used slightly different classifications or data definitions. For example, the travel survey data may be obtained either in activity table format or in trip table format. If the data is in activity table format, then it is necessary to convert different activities into trip activities (i.e. creating trip table) to understand the trip characteristic of a person made on a particular day.

2. Survey Data Processing

The California Statewide Travel Survey provides some coverage of the entire state, but with a strong emphasis on rural and smaller urban areas. Over half of the households in the statewide survey are from counties with fewer than 450,000 residents; these counties have 16% of the total statewide population. In contrast, the counties with over 1.5 million residents have 60% of statewide population, but less than 22% of the surveys (about 3700 households for 20 million people). In other words, there is one surveyed person for every 218 people in counties with fewer than 250,000 people, one for every 528 people in counties between 250K and one million, and one of every 2413 people was surveyed in counties over a million. Because of this underrepresentation in

large urban areas, additional surveys were used to complement these records and provide a more complete representation of travel in California. The surveys available are the SANDAG, SCAG, and MTC surveys.

2.1 Trips

To produce a transportation model, it is necessary to have an appropriate dataset that contains relevant travel information. In particular, it is important to understand the nature of the trips and tours performed by the individuals in a household in the dataset. While survey datasets vary in nature from one survey to another, an essential feature is a file where each record corresponds to a single trip, including the nature of the origin and destination of the trip.

The, SCAG and SANDAG datasets contain a “trips file”, but this file is really centered on the activities at the end of the trip. A more accurate trip file needed to be created, cataloguing the actual trips themselves. For clarity, this document will call the existing “trips file” the “trip (activity) file”.

2.1.1 Trip File

The new trip file contains the key information already in the source database about each trip activity. These are

- The household ID and person record ID
- The origin and destination locations, including place names and TAZ
- The activities at the origin and destination location
- The trip start time and end time
- The mode for the trip, and the other travel information (transit fare, parking cost, etc.)
- The travel party information, including composition

The new trip file also contains newly calculated fields:

- The trip midpoint time – the time halfway between the start and end time (note that the wrap at midnight needs to be taken into account; with midnight coded as

24:00 as in the statewide database, a trip running from 12:50 AM to 1:30 AM should be coded as 1:15 AM. However, the times are entered as 24:50 and 1:30, which would create a midpoint of 13:40, or 1:40 PM)

- The trip time period, as described below
- The trip base purpose, as described below

2.1.2 Time Periods

The SDPTM considers 5 time periods for modeling and analysis purposes (and preservation of the trip midpoint value permits future adjustments of these):

- 3:00:00 AM – 5:59:59 AM: Early Nonpeak
- 6:00:00 AM – 9:59:59 AM: AM Peak
- 10:00:00 AM – 2:59:59 PM: Midday
- 3:00:00 PM – 6:59:59 PM: PM Peak
- 7:00:00 PM – 2:59:59 AM: Late Nonpeak

2.1.3 Trip Purposes

There are 9 preliminary trip purposes as described below. The place name and activities at the origin and destination end were included in the dataset to allow for the possibility of further refinement in the future.

- Home to work, H-W
- Work to home, W-H
- Home to School, H-SC
- School to Home, SC-H
- Home to Other, H-O
- Other to Home, O-H
- Work to Other, W-O
- Other to Work, O-W
- Non-Home Based (where neither trip end is home or work), NHB.

Further processing added more detail to the trip record, including linking it to tours. To produce these trip records, consecutive “trip (activity) file” records needed to be examined and connected.

2.1.4 Linked Trips

For example, consider the following hypothetical “trip (activity) file” records in Table 1 below (note that the file does not exactly look like this; fields have not been shown and coding has been replaced by explicit descriptions for simplicity):

Table 1: Sample Activity File Records

SAMPNO	PERNO	LOCNO	LOCATION	TAZ	ACTIVITY	MODE	VEHNO	ARRIVAL	DEPART
Household ID number	Person ID number	Location ID num.	Location description	TAZ assigned	Activity purpose	Travel mode	Vehicle number	Arrival time	Departure time
10001	1	51319	HOME	354	Home			3:00 AM	6:45 AM
10001	1	48914	Office	521	Work	Drive	2	7:15 AM	12:00 PM
10001	1	62105	Quizno's	524	Eat	Walk		12:10 PM	12:45 PM
10001	1	48914	Office	521	Work	Walk		1:00 PM	4:20 PM
10001	1	66711	Safeway	636	Shop	Drive	2	4:35 PM	5:35 PM
10001	1	51319	HOME	354	Eat	Drive	2	5:55 PM	7:00 PM
10001	1	51319	HOME	354	Home			7:00 PM	2:59 AM
10001	2	51319	HOME	354	Home			3:00 AM	9:25 AM
10001	2	51319	HOME	354	TV			9:25 AM	6:00 PM
10001	2	51319	HOME	354	Eat			6:00 PM	7:00 PM
10001	2	51319	HOME	354	Home			7:00 PM	2:59 AM
10002	1	50010	HOME	335	Home			3:00 AM	9:50 AM
10002	1	32223	UC Davis	332	School	Bike		10:10 AM	6:30 PM
10002	1	62022	Sophia's	320	Eat	Bike		6:40 PM	8:05 PM
10002	1	78584	Varsity Theater	320	Leisure	Walk		8:15 PM	10:00 PM
10002	1	50010	HOME	335	Home	Bike		10:10 PM	2:59 AM

These can be processed into the following five real trip records, as shown in Table 2 below, (note that other data present in the database, such as the location name, are kept but are not shown here for simplicity).

Table 2: Sample Trip Records

SAMPNO	PERNO	OLOC	DLOC	OTAZ	DTAZ	OACT	DACT	MODE	VEHNO	START	END	MID	PERIOD	T_PURP
Household ID number	Person ID number	Origin location ID	Destination location ID	Origin TAZ	Destination TAZ	Origin activity	Destination activity	Travel mode	Vehicle number	Trip start time	Trip end time	Trip mid-point	Time period	Trip purpose
10001	1	51319	48914	354	521	Home	Work	Drive	2	6:45 AM	7:15 AM	7:00 AM	AM Peak	H-W
10001	1	48914	62105	521	524	Work	Eat	Walk		12:00 PM	12:10 PM	12:05 PM	Midday	W-O
10001	1	62105	48914	524	521	Eat	Work	Walk		12:45 PM	1:00 PM	12:53 PM	Midday	O-W
10001	1	48914	66711	521	636	Work	Shop	Drive	2	4:20 PM	4:35 PM	4:28 PM	PM Peak	W-O
10001	1	66711	51319	636	354	Shop	Eat	Drive	2	5:35 PM	5:55 PM	5:45 PM	PM Peak	O-H
10002	1	50010	32223	335	332	Home	School	Bike		9:50 AM	10:10 AM	10:00 AM	Midday	H-Sc
10002	1	32223	62022	332	320	School	Eat	Bike		6:30 PM	6:40 PM	6:35 PM	PM Peak	NHB
10002	1	62022	78584	320	320	Eat	Leisure	Walk		8:05 PM	8:15 PM	8:10 PM	Late	NHB
10002	1	78584	50010	320	335	Leisure	Home	Bike		10:00 PM	10:10 PM	10:05 PM	Late	O-H

Some trip records may reflect part of an actual journey; for instance, a person may have a series of records as shown in Table 3 below. But he/she is basically having one trip.

Table 3: Single Trip from a Sequence of Records

LOCATION	TAZ	ACTIVITY	MODE	VEHNO	ARRIVAL	DEPART
Home	1102	Home			3:00 AM	7:30 AM
Church & 24 th St Muni Stop	1092	Transfer	Walk		7:37 AM	7:43 AM
Powell St BART Station	1015	Transfer	Transit		7:52 AM	7:58 AM
Office	1414	Work	Transit		8:13 AM	4:27 PM

In this case, only one trip took place, and there should be only one record, indicating a trip departing home at 7:30 AM, and arriving at the office at 8:13 AM, with walk access transit as the mode. Some activities, such as “Change mode of transportation” or “wait for / get on a vehicle” indicate this travel pattern. Investigation into the frequency of such activities helped identify how common it is in the database. It is important to adjust

these records; it is much easier to model travel to the real destination (where there are jobs and so on), rather than to intermediate places like transit stops where the person is only visiting en route to the destination. The MTC survey contains detailed documentation on trip linkages.

2.2 Activities

Activities in each of these surveys are categorized by type. Each survey has a slightly different approach to activity classification and Table 4 summarizes these differences.

For modeling purposes, there are 10 activity types; two are “special” types of home and travel, two are the traditional mandatory types of work and school, and the remaining six are often considered under the “other” purpose. These are eat, escort, personal business, recreation/entertainment, and shop.

Table 5 below shows the coding for each of the 111 activity types in the four source surveys.

Table 4: Activity Matching between Surveys

Activity	Statewide	SCAG	SANDAG	MTC
Home	7 explicit "at home" activities	2 "at home" activities: working, other	4 "at home" activities: working; shopping; entertainment; other	1 "at home" activity: Shopping at home
Work	Work including regular; volunteer; travel while at work	Work including regular; volunteer; travel at work	Work including regular; volunteer; eat at work; travel at work; personal business at work	Work or work-related
School	School K-12; school PSE; daycare	Classes; extracurricular; childcare	Classes; extracurricular	School or school-related
Travel	Four mode change/access activities	Change mode	Change mode; loop trip	Travel activity; change mode
Shopping	Major shop; incidental shop; buy gas	Shopping; quick stop (ATM, gas, coffee, etc.)	Shopping; quick stop (ATM, gas, coffee, etc.)	Shopping
Escort	Pick up / Drop off separate, explicit, accompany activity	Pick up / Drop off separate, explicit, accompany activity	Pick up / Drop off combined	Pick up / Drop off combined
Meals	Eat meal	Eat meal	Eat meal	Meal activity
Volunteer, Civic and Religious	Three separate activities (regular volunteer with work)	Three separate activities (regular volunteer with work)	Civic and religious combined (regular volunteer with work)	VCR combined into one activity
Personal Business	Medical; ATM /post office / banking / utilities; other HH/personal business	Banking / post office / bills; ATM included with quick shopping stops	Personal business; ATM included with quick shopping stops	Personal services / bank / government; Sick or ill / medical appointment
Recreation, Fitness and Entertainment	Three separate activities	Entertainment; recreation	Entertainment; recreation/ fitness	Combined into one activity
Social	Visit with friends or relatives	Visit with friends or relatives	Visit with friends or relatives	Social activities

Table 5: Activity Categories

Survey	Code	Activity	Activity Description
MTC	8	Home	Shopping (at home)
MTC	1	Home	Home
MTC	2	Home	Household chores and personal care (childcare, care of others), meal preparation, home maintenance
MTC	5	Home	Sleep
SANDAG	1	Home	Working at home/telecommuting (job-related for pay)
SANDAG	2	Home	Eating/preparing meal at home
SANDAG	3	Home	Watching TV, DVD, or Video at home
SANDAG	4	Home	Shopping by phone, TV, or Internet at home
SANDAG	5	Home	Banking by phone or Internet at home
SANDAG	6	Home	Exercising at home
SANDAG	7	Home	Other at home, specify
SCAG	23	Home	Working at home (related to main or second job)
SCAG	24	Home	Other at home activities
Statewide	1	Home	Working at home (related to main or second job)
Statewide	2	Home	Eating/preparing meals at home
Statewide	3	Home	Watching TV/Videos at home
Statewide	4	Home	Shopping by phone/TV/Internet at home
Statewide	5	Home	Exercising at home
Statewide	6	Home	Other at home;
MTC	1	Travel	Driving, riding, walking, biking, flying
MTC	17	Travel	Change mode
SANDAG	23	Travel	Change mode of transportation (get on/off bus or train, park car, etc.)
SANDAG	24	Travel	Loop Trip that begins and ends at the same place

Survey	Code	Activity	Activity Description
SCAG	1	Travel	Change mode of transportation
Statewide	7	Travel	Wait for/get on a vehicle
Statewide	8	Travel	Leave/park a vehicle
Statewide	9	Travel	Boarding activities for airplane, rail, intercity bus
Statewide	10	Travel	Getting off airplane, rail, intercity bus
Statewide	34	Travel	SCAG - Change mode of transportation
MTC	6	Work	Work or work-related
SANDAG	8	Work	Work (including regular volunteer work)
SANDAG	9	Work	Eat meal at work
SANDAG	10	Work	Personal business by phone/Internet at work
SANDAG	21	Work	Work-related (meeting, sales call, delivery, etc.)
SCAG	7	Work	Work (include regularly scheduled volunteer work)
SCAG	8	Work	Work-related (sales call, meeting, errand, etc.)
Statewide	13	Work	Work (includes regularly scheduled volunteer work)
Statewide	14	Work	Work-related (sales calls, meetings, errands, etc.)
MTC	7	School	School or school-related
SANDAG	11	School	Attending class
SANDAG	12	School	School-related sports or extra-curricular activity
SCAG	9	School	School (attending classes)
SCAG	10	School	Other school activities (sports, extra-curricular)
SCAG	11	School	Childcare, daycare, after-school care
Statewide	15	School	School (Preschool, K-12th)
Statewide	16	School	School (Postsecondary College, vocational)
Statewide	17	School	Childcare, daycare, after-school care;
Statewide	35	School	SCAG - School (attending classes)
Statewide	36	School	SCAG - Other school activities

Survey	Code	Activity	Activity Description
MTC	3	Eat	Meals
SANDAG	17	Eat	Eat meal outside of home or work
SCAG	12	Eat	Eat meal (restaurant, drive through, take out)
Statewide	18	Eat	Eat out (restaurant, drive through, etc.)
MTC	16	Escort	Pick up / drop off
SANDAG	22	Escort	Drop off or pick up passenger
SCAG	2	Escort	Pick up someone or get picked up
SCAG	3	Escort	Drop off someone or get dropped off
SCAG	21	Escort	With another person at their activity out of home
Statewide	11	Escort	Pick up someone or get picked up
Statewide	12	Escort	Drop off someone or get dropped off
Statewide	32	Escort	Be with another person at their activity
MTC	0	Personal Business	NA/Unknown
MTC	10	Personal Business	Personal services / bank / government
MTC	14	Personal Business	Sick or ill / medical appointment
MTC	15	Personal Business	Internet use (non-work, non-shopping)
MTC	990	Personal Business	Out of town
MTC	996	Personal Business	Other
SANDAG	16	Personal Business	Personal business (medical/dental, dry cleaning, errands, etc.)
SANDAG	97	Personal Business	Other (specify)
SCAG	6	Personal	Banking, post office, pay bills

Survey	Code	Activity	Activity Description
		Business	
SCAG	13	Personal Business	Medical
SCAG	22	Personal Business	Other personal (specify)
SCAG	97	Personal Business	Other activity
SCAG	99	Personal Business	DK/RF (Don't know/Refused to answer)
Statewide	0	Personal Business	No other activities
Statewide	19	Personal Business	Medical
Statewide	30	Personal Business	ATM, banking, post office, utilities
Statewide	31	Personal Business	Other personal or household business
Statewide	97	Personal Business	Other, specify
Statewide	99	Personal Business	DK/RF (Don't know/Refused to answer)
MTC	4	Recreation / Entertainment	Recreation / entertainment
SANDAG	18	Recreation / Entertainment	Entertainment (movie, concert, etc.)
SANDAG	19	Recreation / Entertainment	Recreation, fitness
SCAG	14	Recreation / Entertainment	Fitness activity (playing sports, gym, bike ride)

Survey	Code	Activity	Activity Description
SCAG	15	Recreation / Entertainment	Recreational (vacation, camping, etc.)
SCAG	16	Recreation / Entertainment	Entertainment (watching sports, movies, dance, bar, etc.)
Statewide	20	Recreation / Entertainment	Fitness activities (Gym/Health club/participating in sports)
Statewide	21	Recreation / Entertainment	Recreational (vacation, camping, etc.)
Statewide	22	Recreation / Entertainment	Entertainment (movies, dance club, bar, spectator sports, etc.)
MTC	9	Shop	Shopping (not at home)
SANDAG	13	Shop	Quick stop for: gas, coffee, ATM, etc.
SANDAG	14	Shop	Shopping (grocery, clothes, etc.)
SCAG	4	Shop	ATM, buy gas, quick stop for coffee, newspaper, etc.
SCAG	5	Shop	Shopping
Statewide	27	Shop	Buy gas
Statewide	28	Shop	Incidental shopping (groceries, house wares, medicine, etc.)
Statewide	29	Shop	Major shopping (furniture, clothes, autos, etc.)
Statewide	33	Shop	SCAG – Shopping
Statewide	37	Shop	SCAG - ATM, buy gas, quick stop for coffee, newspaper, etc.
MTC	11	Social	Social activities
MTC	12	Social	Relaxing / resting
MTC	13	Social	Volunteer / civic / religious
SANDAG	15	Social	Visit friends or relatives
SANDAG	20	Social	Civic or religious activities
SCAG	17	Social	Visit friends/relatives

Survey	Code	Activity	Activity Description
SCAG	18	Social	Community meetings, political/civic event, public hearing
SCAG	19	Social	Occasional volunteer work
SCAG	20	Social	Church, temple, religious meeting
Statewide	23	Social	Visit friends/relatives
Statewide	24	Social	Community meetings, political or civic event, public hearing, voting, etc.
Statewide	25	Social	Occasional volunteer work
Statewide	26	Social	Church, temple, religious meeting

3. Development of the Master Database

Based on the above methods, all survey data from each source were examined and a linked trip table was developed. TAZs were allocated to trip end locations, based on coded coordinates or block groups for trip origins and destinations. Trip files were created, and preliminary tabulations prepared. Fields with “missing” data were identified (e.g. household income, TAZ of trip origin or destination).

3.1 Aggregation of Data

In this section, the steps required to prepare datasets for calibration of the short distance personal travel model are described.

3.1.1 Define Household and Person Table Attributes Uniformly

The data definitions were often different across different data sources. Therefore a uniform data definition was required as well as its value type across different data sources as described below.

- Trip-related Attributes: trip mode name, trip mode code, origin/destination activity name, and its code.
- Household-related attributes: household income, dwelling type, and tenure.

- Person-related attributes: gender, employed/unemployed, full-time/part-time worker, work industry, work occupation, student status, and school type

See section 3.2 Data Expression of Master Database for definition of various attributes

3.1.2 Eliminate “Out of Scope” / Incomplete Survey Data

The procedure to eliminate out of scope and/or incomplete trips is described in this section and is as described below.

- Eliminate survey data for surveys conducted in July and August, on weekend days, or weekday holidays. The CSTDM09 is designed to predict “typical weekday” travel in spring/fall, when schools are in session.
- Eliminate second weekday trip records, if two consecutive weekday trip records are available.
- Eliminate survey data for persons where there is incomplete travel data, e.g., no TAZ data for origin or destination, missing trip mode etc.
- Eliminate survey data for persons where the personal data is incomplete, e.g., age of the person, school type is “don’t know/refused” even if the person is identified as student, or work industry and work occupation is “don’t know/refused” even if the person is identified as worker.
- Eliminate survey data for households, and all persons in the household and all trips of made by persons in that household, where:
 - a. There is incomplete household data, e.g., size of household, dwelling type, owning type, and vehicles; and
 - b. Absence of household TAZ identification (ID) number

3.1.3 Create New Attributes Based on Available Attribute Values

Some new household/person related attributes were needed for the model, which we were able to create from the available values of different attributes.

Household table:

- Actual number of household workers

- Actual number of household students
- Identifying households where all persons age is 65+
- Total heads in household where age is between 0-5
- Total heads in household where age is between 6-15
- Ordinal household income category

Person table:

- Person Type: A 'Person Type' field was needed to classify each person into one of the 8 person types used in the model (pre-schooler; student < 16 years; high school student 16+; post secondary student; full time worker; part time worker; adult other; senior 65+ not worker or part-time student)

3.1.4 Create One Master Database with Three Tables

Consolidated tables are required with details of households, persons, and person “linked” trips that combine the data from each of the 4 travel surveys. Attributes (i.e. various data fields) for each table are given below.

Household (HH) Data:

- Source Survey ID
- Original Survey HH ID
- Consolidated HH ID
- Survey Day ID
- HH TAZ
- Household Size (Persons)
- Household Income (In three different format)
- # Private Vehicles Owned
- Tenure
- Dwelling Type
- Actual number of HH workers
- Actual number of HH Students
- Identifying HH, where all persons age is 65+
- Total heads in household where Age is between 0-5
- Total heads in household where Age is between 6-15
- Total heads in household where Age is between 16+

Person Data:

- Source Survey ID
- Original Survey HH ID
- Original Survey Person ID
- Consolidated HH ID
- Consolidated Person ID
- Survey Day ID
- Age
- Work / Not Work Status
- Full / Part Time Worker

- Worker Industry
- Worker Occupation
- Student / Not Student Status
- Full / Part Time Student Status
- School / Student Type
- Person Type

Linked Trip Data:

- Source Survey ID
- Original Survey HH ID
- Original Survey Person ID
- Consolidated HH ID
- Consolidated Person ID
- Survey Day ID
- Trip Number
- Trip Start Time
- Trip End Time
- Trip Mid Point Time
- Trip Time Period (based on midpoint time)
- Trip Origin TAZ
- Trip Destination TAZ
- Trip Mode
- Trip Origin Purpose
- Trip Destination Purpose

Note that trips need to be “linked” to combine multi-modal trips into 1 trip with appropriate multi-modal coding; fields need to be added to identify the survey source; and data are required on person tour patterns and person daily activity patterns that can be tabulated from the above data.

3.1.5 Develop Expansion Factors for Each Household and Person

Expansion factors needed to be developed to aggregate individual household and person observations to the total households and persons in California in the year 2000.

The key features include:

- The 2000 census data to provide control totals of population;
- Attributes by geographic area; and
- Socioeconomic attributes of people and households.

See the appendix for further explanation of the travel survey expansion factor.

3.2 Data Expression of Master Database

To prepare the uniform trip, person and household tables for the master database, which originated from several data sources, some attributes needed to be reclassified and recoded. The following tables (Tables 6, 7, 8, 9, and 10) show the definition of various attributes in the master database and their respective coding.

Tables 6 and 7 are for trip tables.

Table 6: Trip Mode Definition and Its Coding

Trip Mode Name	Mode Code
SOV- Single Occupant Vehicle	1
HOV-2	2
HOV-3	3
Transit- Walk Access	4
Transit-Car Access	5
Transit-No Access Specified	6
Walk	7
Bicycle	8
School Bus	9
HOV-Not Specified the number of occupancy	10
Air	11
Other	12

The database reflects modes like HOV-Not Specified the number of occupancy, Transit-No Access Specified, Air and Other. These modes are not considered in short-distance-personal-trip model.

Table 7: Trip Activity Definition and Its Coding

Origin/Destination Activity of a Trip	Trip Code
Home	1
Personal Business	2
Recreation/ Entertainment	3
Shop	4
Social	5
Travel	6
Eat	7
Escort	8
School	9
Work	10

Tables 8, 9, and 10 are household tables.

Table 8: Dwelling Type of a Household

Dwelling Type	Code
Unattached Single Family Home	1
Duplex	2
Apartment/ Condominium or Town house	3
Mobile home or Trailer	4
Other	5

Table 9: Tenure of a House

Tenure Type	Code
Own	1
Rent	2
Other	3

Table 10: Ordinal Income Category of a Household

Household Ordinal Income category (Income per annum)	Code
Earning less than \$10,000	1
\$10,000-\$24,999	2
\$25,000-\$34,999	3
\$35,000-\$49,999	4
\$50,000-\$74,999	5
\$75,000-\$99,999	6
\$100,000-\$149,999	7
\$150,000 and Up	8

3.2.1 A Note on the Ordinal Income Category for SANDAG Data

The treatment of income in the four source surveys differed significantly. In all cases, income was reported in ranges, and in all cases, each bin was defined in a “round down” fashion, where (for example) \$25-35K means income from \$25,000 to \$34,999. However, the definitions of the ranges varied from survey to survey. The Statewide survey was consistent with the SCAG breakdowns. The MTC survey provided a more detailed income breakdown, but one that could be aggregated with the Statewide and SCAG groups. The SANDAG survey, however, was inconsistent in its definition of incomes, in particular for those records under \$60,000. (Households over \$60,000 were defined in more detailed categories that could be aggregated to the Statewide groupings.) Table 11 below shows the income ranges as used for the two surveys.

Table 11: SANDAG & SCAG Income Comparison

From	To	SANDAG	SCAG/Statewide/MTC
\$ 0	\$ 4,999	1	1
\$ 5,000	\$ 9,999	1	1
\$ 10,000	\$ 14,999	1	2
\$ 15,000	\$ 19,999	2	2
\$ 20,000	\$ 24,999	2	2
\$ 25,000	\$ 29,999	2	3

\$ 30,000	\$ 34,999	3	3
\$ 35,000	\$ 39,999	3	4
\$ 40,000	\$ 44,999	3	4
\$ 45,000	\$ 49,999	4	4
\$ 50,000	\$ 54,999	4	5
\$ 55,000	\$ 59,999	4	5
\$ 60,000	\$ 64,999	5	5
\$ 65,000	\$ 69,999	5	5
\$ 70,000	\$ 74,999	5	5
\$ 75,000	\$ 79,999	6	6
\$ 80,000	\$ 84,999	6	6
\$ 85,000	\$ 89,999	6	6
\$ 90,000	\$ 94,999	6	6
\$ 95,000	\$ 99,999	6	6
\$ 100,000	\$ 104,999	7	7
\$ 105,000	\$ 109,999	7	7
\$ 110,000	\$ 114,999	7	7
\$ 115,000	\$ 119,999	7	7
\$ 120,000	\$ 124,999	7	7
\$ 125,000	\$ 129,999	8	7
\$ 130,000	\$ 134,999	8	7
\$ 135,000	\$ 139,999	8	7
\$ 140,000	\$ 144,999	8	7
\$ 145,000	\$ 149,999	8	7
\$ 150,000	\$ 154,999	9	8
\$ 155,000	\$ 159,999	9	8
\$ 160,000	\$ 164,999	9	8
\$ 165,000	\$ 169,999	9	8
\$ 170,000	\$ 174,999	9	8

\$ 175,000	\$ 179,999	9	8
\$ 180,000	\$ 184,999	9	8
\$ 185,000	\$ 189,999	9	8
\$ 190,000	\$ 194,999	9	8
\$ 195,000	\$ 199,999	9	8
\$ 200,000	max	10	8

As can be seen from the above table, the SANDAG survey divided the area under \$60,000 into four equal ranges of \$15,000 and the Statewide survey provided a different set of ranges.

To enable a consistent treatment of income across the entire dataset, it was decided to recode the SANDAG data into the common set of ranges, a task which required imputing new incomes for the SANDAG records under \$60,000. The data source used for this was the full year 2000 synthetic population for the SANDAG jurisdiction.

The imputation of the incomes under \$60,000 was based on four factors, depending on data availability. The four factors were:

- SANDAG reported income category
- Household size
- Number of workers
- Geographic district

The first factor was maintained strictly; a household reported to earn under \$15,000 could only be imputed to have an income in the appropriate Statewide categories of under \$10,000 or \$10,000 to \$24,999. Because of the layout of the categories, there were always two alternatives for each of the first four SANDAG income categories. This was convenient; each SANDAG range had probabilities developed for matching with the appropriate lower and the higher Statewide income categories.

Household size (the number of persons in the household) and the number of workers were used to match the survey household with appropriate households in the synthetic

population. The number of workers was top-coded at 3. For households of 1, 2 or 3 persons, which were over 80% of the households needing imputation, the geographic district was further used to determine the probabilities. There are six districts, numbered 61-66 in the SANDAG area, shown on the map below.

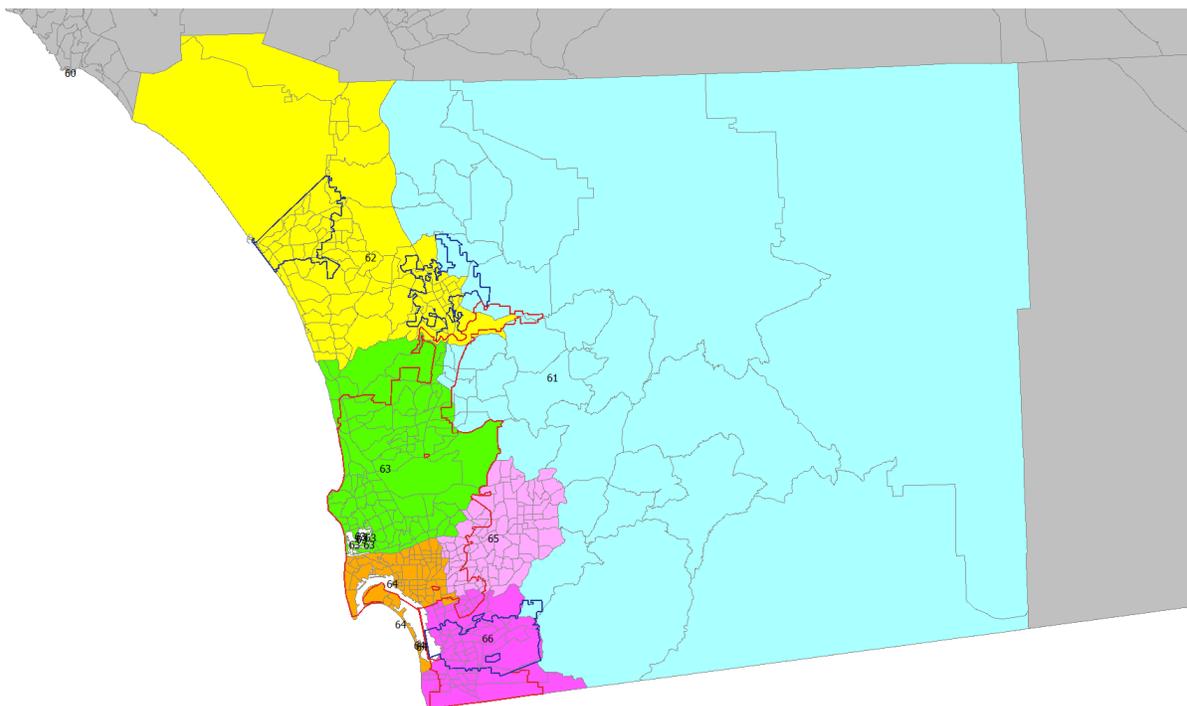


Figure 1: SANDAG TAZ Districts

District 61 is the cyan district comprising the eastern 80% or so of San Diego County; 62, in yellow, is the North County area including Escondido and Oceanside (shown with blue outlines). 63, shown in green, is the northern part of San Diego (red outline); 64, in orange, is the central part of the city as well as Coronado. 65, in pink, is the eastern fringe of San Diego and includes the East County suburbs such as El Cajon, La Mesa and Santee. Finally, 66, shown in magenta, includes the South Bay area as well as Chula Vista (shown with a blue outline).

The synthetic population was queried to find the count of households with each combination of SANDAG income group, Statewide income group, household size, number of workers, and district. The aggregations described above were performed (across districts for households with 4+ residents, and combining households with 3 or

more workers into a 3+ worker category). The resulting distributions were used to establish probabilities for each SANDAG grouping. As an example, the table below describes some frequencies of the households earning under \$15,000 in the synthetic population.

Table 12: Synthetic Household Frequencies by Persons, Workers and Income

Persons	Workers	Count of HH in synth. pop.		Prop \$10-15K
		Under \$10K	\$10-15K	
1	0	17661	77312	18.6%
	1	7062	58775	10.7%
2	0	6921	9658	41.7%
	1	4043	3465	53.8%
	2	1048	613	63.1%
3	0	2500	4838	34.1%
	1	2562	2309	52.6%
	2+	825	616	57.3%
4 or 5	0	3088	5039	38.0%
	1+	4714	3391	58.2%
6+	0	933	1256	42.6%
	1+	1450	1123	56.4%
Overall		52807	168395	23.9%

Overall, about ¼ of the households earning under \$15K in the synthetic population earn between \$10,000 and \$14,999. However, this is mostly due to single person households, who are much less likely to earn \$10-15K in this range; larger households are more likely to be placed the higher bracket. Similarly, the presence of workers increase the odds that the household is in the higher bracket – except the single households, where households without workers are actually more likely to be in the higher \$10-15K bracket. This is likely an effect of seniors on fixed income government support, where minimums prevent earning less than \$10K. Note that the probabilities actually used for imputation are less aggregated than these.

The role of districts in disaggregating groups is fairly small in general, although for a few situations there is a notable effect; Table 13 below describes the probability of a household earning under \$15K, with one person who works being assigned to the \$10-15K range by district.

Table 13: Probability of a Household Earning \$10-15k with One Worker

District	Proportion \$10-15K
61	36.2%
62	6.1%
63	16.9%
64	8.1%
65	25.4%
66	12.7%

With the probabilities calculated, the households had their income imputed and were placed in an appropriate group. Table 14 summarizes the reclassification of the SANDAG records.

Table 14: Summary of SANDAG Record Reclassification

Original SANDAG Group	Recoded income group								Grand Total
	1	2	3	4	5	6	7	8	
	\$100								
	<\$10K	\$10-25K	\$25-35K	\$35-50K	\$50-75K	\$75-100K	-150K	\$150K+	
1 - <\$15K	186	65							251
2 - \$15-30K		253	153						406
3 - \$30-45K			151	232					383
4 - \$45-60K				119	240				359
5 - \$60-75K					257				257
6 - \$75-100K						580			580
7 - \$100-125K							374		374
8 - \$125-150K							199		199
9 - \$150-200K								169	169
10 - \$200K+								127	127
Grand Total	186	318	304	351	497	580	573	296	3105

The final set of income categories are shown in Table 15.

Table 15: Final Income Categories

Group	Range
1	Earning less than \$10,000
2	\$10,000 - \$24,999
3	\$25,000 - \$34,999
4	\$35,000 - \$49,999
5	\$50,000 - \$74,999
6	\$75,000 - \$99,999
7	\$100,000 - \$149,999
8	\$150,000 and up

Tables 16 to 23 below are the person tables.

Table 16: Work Status of a Person

Whether Worker or Not	Code
Yes	1
No	2

Table 17: Type of Work Status

Whether Worker	Full/Part Time	Code
Full Time Worker		1
Part Time Worker		2

Table 18: Work Industry of a Worker

Work Industry	Code
Agriculture/Forestry/Fishing	11
Mining	21
Utilities	22
Construction	23
Manufacturing-Non Durable Goods	31
Manufacturing- Durable Goods	32
Wholesale-Trade	41
Retail Trade	44
Transportation and Warehousing	48
Information	51
Finance or Insurance	52
Real Estate	53
Professional, Scientific and Technical Services	54
Management of Companies and Enterprises	55
Administrative and Supportive Services	56
Educational Services	61
Health Care and Social Assistance	62
Arts, Entertainment and recreation	71
Accommodations and Food Services	72
Other Services (Excluding Public administration)	81
Public administration	92
Other	97
Absence of Information	-9999

Table 19: Work Occupation of a Worker

Work Occupation	Code
Executive, Admin and Managerial	1
Professional Specialty	2
Technicians or related support	3
Sales	4
Administrative Support, Clerical	5
Private household	6
Protective Services	7
Service except protective and household	8
Farming, forestry or fishing	9
Precision, production, craft or repair	10
Machine operator, assembler or inspector	11
Transportation or material moving	12
Handler, equipment, cleaner, helper or barourer	13
Other, Specify	14
Absence of Information	-9999

Table 20: Student Status of a Person

Whether Student or Not	Code
Yes	1
No	2

Table 21: Type of Student Status

Whether Full/Part Student	Code
Full-Time Student	1
Part-Time Student	2

Table 22: School Type of Student

School Type	Code
Daycare/ Preschool	1
K to 12 th Grade	2
College/ University	3
Post graduate	4
Trade/ Vocational	5
Other	6
Absence of Information	-9999

Table 23: Person Type of All Persons in the Database

Person Type	Code
Preschooler	1
K12 Student < 16 Years of Age	2
K12 Student 16+	3
Post Secondary Student	4
Full-Time Worker	5
Part-Time Worker	6
Adult Other	7
Senior (65+)	8

4. Tour Table Preparation

This section explains the different data fields that exist in a tour table (or trip-tour table) of the master database and the methodologies for their creation, which are adopted to derive tour information from the trip table. The following are the fields in the tour table.

1. Number of trips per Person
2. TourStart
3. TourID
4. NumberOfToursPerpersonPerDay
5. TourPurpose
6. WhetherCompleteIncompleteTour

7. TourModeCode
8. TourModeName
9. TourOrigin
10. TourDestination
11. PurposeOfOtherTour
12. WhetherSubTourExists
13. NumberOfSubTour
14. ForwardBackward Pass
15. Whether distance between tour origin and destination is moreThan100Mile
16. TourModeAcceptance
17. AcceptedTour
18. AcceptedCompleteTour

4.1 NumberOfTripsPerPersonPerDay

NumberOfToursPerpersonPerDay gives the information of total number of trips a person had on a typical working day.

4.2 TourStart

This data field identifies the start of a tour. The start of a tour is identified as 1. It also acts as a *flag* for identifying the tours (both complete and incomplete tour) in the tour table.

4.3 TourID

This field identifies both complete and incomplete tours with a unique tour ID number.

4.4 NumberOfToursPerPersonPerDay

This data field gives the information of the number of tours a person had on a typical working day. It counts both complete and incomplete tour.

4.5 TourPurpose

This data field identifies the purpose of each tour (both complete and incomplete tour) in the database. The following codes are used to identify the tour purpose:

Work Tour= 1

Other Tour = 2

School Tour = 3

In a tour, if the destination activity of at least one trip is recognized as “work” and the destination activity of no other trips is recognized as “school”, then the tour is considered as “work” tour. In the present study, the destination TAZ ID of the first trip with destination activity “work” is recognized as tour’s primary destination if there is not explicit information of work place, where persons come for regular work activity and not for work related activities. In the definition of origin/destination purpose of a trip, it is assumed that a person first comes to a work place and then he or she goes to some “other” or “work related” place. But, it should be mentioned that in the California Statewide Travel Survey and SCAG Survey, there has been an explicit definition of regular work places and work related places, which define distinct regular work activities and work-related activities. In these cases the explicitly-defined places are used to identify the tour destination.

On the other hand, if the destination activity of at least one trip is recognized as “school” and the destination activity of no other trips is recognized as “work”, then the tour purpose is considered as “school” tour. In the present study, the destination TAZ ID of the first trip with destination activity “school” is recognized as primary tour destination, if there is no explicit information of a school place, where classes are attended. It is assumed that a person first comes to a school place and then he or she goes to some “other” or “academic activity” related places. But, again, it should be mentioned that in the case of SCAG and SANDAG Travel Survey, there have been an explicitly defined school places (by school activity), where classes are attended and other places, where school-related activities take place. So, in these cases, this additional information is

used to identify the exact tour destination i.e. location of school, where classes are attended.

There are some instances where the destination activities of different trips constituting a tour could be both “work” and “school”. In these cases, it is difficult to identify the exact definition of tour purpose. In such cases some person-specific attributes are used to decide purpose of the tour. These person-specific attributes include: whether the tour maker (i.e. the person who had the tour) is full time worker or full time student, person type, school type and age.

In a tour, if destination activity of any trip is not recognized as either of “work” and “school”, then the tour purpose of this tour is considered as “other” tour. Currently, the destination of first trip of the tour is considered as tour destination of “other” tour purpose and corresponding activity is recognized as the actual purpose (there are 8 different “other” tour purposes defined in the present study) of the “other” tour.

4.6 WhetherCompleteIncompleteTour

While deriving the tour information from the trip table, it has been observed that there are many incomplete tours that exist in the database. A tour is said to be complete if it starts at home (more specifically a TAZ with home activity) and also ends at the same home (i.e. at the same TAZ with home activity). Based on the definition of complete tour, three different kinds of incomplete tour are identified.

1. The start of the tour is not at home
2. The end of the tour is not at home
3. The start TAZ ID does not equal the end TAZ ID

In the third case of different start and end TAZ IDs, these tours are recognized as incomplete even though they look like a complete tour (based on the sequence of trip activities). There are 212 incomplete tours of this type in the database.

The procedure for identifying a complete/ incomplete tour is explained in Table 24 below.

Table 24: A Typical Trip-Tour Table

ID	Trip ID	Origin TAZ	Destn. TAZ	Origin Act.	Destn. Act.	Tour ID	Tour Purpose	Whether Complete/Incomplete Tour
1	1	1101	1109	Per. Business	Shop	1	1	0
2	2	1109	1550	Shop	Work	1	1	0
3	3	1550	1174	Work	Home	1	1	0
333	1	3508	3517	Home	Shop	333	1	0
334	2	3517	3550	Shop	Work	333	1	0
335	3	3550	3117	Work	Per. Business	333	1	0
450	1	6771	6722	Home	Per. Business	390	1	0
451	2	6722	6788	Per. Business	Work	390	1	0
452	3	6788	6800	Work	Home	390	1	0

The above trip-tour table shows 3 different tours with Tour IDs 1, 333 and 390. All of these tours are incomplete in nature. The complete/incomplete tours are identified by the data field called Whether Complete/Incomplete tour. A complete tour is identified by code 1 whereas an incomplete tour is identified by 0. It can be seen from the above table that tour 1 does not start at home but ends at home, whereas tour 333 though starts at home but ends at somewhere else. These two tours are considered incomplete. The tour 390 looks complete as it starts at home and also ends at home. But if we compare the start home TAZ ID with the end home TAZ ID, then these two TAZ IDs do not match each other even though their corresponding activities are same i.e. home. These types of tour are also recognized as incomplete.

In the database, there are 93,906 tours. Out of these tours only 4,058 tours are recognized as incomplete in nature. It should be mentioned that irrespective of the completeness of a tour, each tour has a tour origin and a tour destination (obviously based on tour purpose). The tour origin of each incomplete tour is recognized as the household TAZ of the person who had the tour.

4.7 TourModeCoding

In the database, the following different types of trip mode are reflected: AIR, SOV, HOV-2, HOV-3, HOV-NO OCCU SPECIFIED, SCHOOL BUS, TRANSIT-CAR ACCESS, TRANSIT- WALK ACCESS, TRANSIT- NO ACCESS SPECIFIED, OTHER, BICYCLE, and WALK.

For each tour (both complete and incomplete) in the database, a tour mode is defined. In the course of defining the tour mode, Table 25 below provides the ranking of different modes followed to derive the definition of tour mode for each tour in the database.

Table 25: Trip Modes and Corresponding Rank

Mode	Rank
Air	1
SOV	2
HOV - 2	3
HOV - 3	4
HOV - No Occupancy Specified	5
School Bus	6
Transit - Car Access	7
Transit - Walk Access	8
Transit - No Access Specified	9
Other	10
Bicycle	11
Walk	12

The process of defining tour mode is explained with two typical examples.

Example 1. Suppose a tour consists of three trips. The trip modes of these three trips are SOV, HOV-2, HOV-3. In this case, the trip mode with the highest rank is SOV. Therefore, the tour mode is SOV.

Example 2. Suppose a tour consists of three trips and the trip modes for these trips are Air, SOV and Transit-Car Access. In this case, the definition of the tour mode will be Air, as its rank is highest among the three trip modes.

It should be mentioned that the above 12 modes are reflected as tour modes in the master database, but only 10 tour modes are considered for the final model development. These 10 different tour modes are SOV, HOV-2, HOV-3, HOV-No occupant specified, school bus, transit – car access, transit – walk access, transit – no access specified, bicycle, and walk.

The source surveys were not entirely consistent in their treatment of HOV travel and of transit travel; for some of the data the travel mode is only recorded as HOV or transit, and it is not possible to determine if the specific HOV mode is HOV2 or HOV3+, or whether the transit mode is walk access transit or drive access transit.

An alternative nesting method by setting different availability in the mode choice estimation data sets was applied to overcome this lack of information and keep these survey data in the estimation file. This applies to all mode choice models, including trip and tour modes.

- For HOV unspecified: it is nested in the same branch with HOV2 and HOV3+. When the HOV unspecified is chosen, then HOV2 and HOV3+ are not available in choice set; otherwise, if HOV2 or HOV3+ is chosen, then HOV unspecified is not available in choice set.
- For transit unspecified access mode: it is nested in the same branch with WT and DT. When the transit unspecified access mode is chosen, then WT and DT

are not available in choice set; otherwise, if WT or DT is chosen, then transit unspecified access mode unspecified is not available in choice set.

4.8 TourModeName

This data field contains the tour mode name in text form. The following table contains the tour mode name and its corresponding tour mode code.

Table 26: Tour Mode Name and Code

Tour Mode Name	Tour Mode Code
SOV	1
HOV-2	2
HOV-3	3
Transit-Walk Access	4
Transit- Car access	5
Transit- No Access Specified	6
Walk	7
Bicycle	8
School Bus	9
HOV-Not Specified the Occupancy	10
Air	11
Other	12

4.9 TourOrigin

This data field contains the origin TAZ ID of a tour. It is essentially the TAZ ID of a zone, where the trip maker was having home activity.

4.10 TourDestination

This is the primary destination of a tour. The process identifying the tour's primary destination is explained in *TourPurpose*.

4.11 PurposeOfOtherTour

This data field contains the destination purpose/activity information of the trip corresponding to the primary destination of the "Other" tour.

4.12 WhetherSubtourExists

Identification of the existence of a subtour is based on the following two criteria:

1. If the destination activity of a trip matches with the destination activity of the primary destination of the tour
2. If the destination TAZ ID of the same trip matches with the TAZ ID of the primary destination of the tour.

The above two criteria are followed to find out the existence of a subtour and also the number of subtours.

4.13 NumberOfSubTours

Once the subtour is identified based on the above approach, the number of subtours made by a person on his/her tour is calculated.

4.14 ForwardBackwardPass

The datasets for estimating trip-level intermediate stop locations are prepared using the individual trip records from the database. All trips on a half-tour are used except the final one - the trip to the primary destination on the forward half-tour and the trip to the home location on the backward half-tour. Even though each trip has its own reported purpose, they are grouped together using the main purpose of the tour. For example, the

intermediate stop location choice model for work tours includes all trips made on work tours, which do not necessarily have work purpose.

The following example, in Table 27, shows the presence of subtours, calculation of the number of subtours and identification of Intermediate Stop Location/ Secondary Destination.

Table 27: A Typical Trip-Tour Table

ID	Trip ID	Origin TAZ	Destn. TAZ	Origin Act.	Destn. Act.	Tour ID	Tour Purpose	Sub tour Exists	Num of Sub tours	Forward BackWard Pass
1	1	1111	1167	Home	Work	122	Work	0	0	
2	2	1167	1225	Work	Shop	122	Work	0	0	Backward
3	3	1225	1111	Shop	Home	122	Work	0	0	
102	1	1203	1305	Home	Escort	325	Work	0	0	Forward
103	2	1305	1326	Escort	Per. Business	325	Work	0	0	Forward
104	3	1326	1309	Per. Business	Work	325	Work	0	0	
105	4	1309	1204	Work	Eat	325	Work	0	0	Backward
106	5	1204	1203	Eat	Home	325	Work	0	0	
328	1	5505	5267	Home	Work	448	Work	1	1	
329	2	5267	5550	Work	Per. Business	448	Work	1	1	LoopForward
330	3	5550	5267	Per. Business	Eat	448	Work	1	1	LoopForward
331	4	5267	5267	Eat	Work	448	Work	1	1	
332	5	5267	5550	Work	Per. Business	448	Work	1	1	Backward
333	6	5550	5415	Per. Business	Shop	448	Work	1	1	Backward
334	7	5415	5505	Shop	Home	448	Work	1	1	
888	1	6606	6788	Home	Pers. Business	555	Work	1	2	Forward
889	2	6788	6217	Pers. Business	Work	555	Work	1	2	
890	3	6217	6217	Work	Pers. Business	555	Work	1	2	LoopForward
891	4	6217	6217	Pers. Business	Work	555	Work	1	2	

892	5	6217	6888	Work	Work	555	Work	1	2	LoopForward
893	6	6888	6217	Work	Pers. Business	555	Work	1	2	LoopForward
894	7	6217	6217	Pers. Business	Work	555	Work	1	2	
895	8	6217	6788	Work	Shop	555	Work	1	2	Backward
896	9	6788	6606	Shop	Home	555	Work	1	2	

In the above table, we find that we have four tours (Tour unique IDs are 122, 325, 448 and 555). All these tours have the same tour purpose, i.e. work purpose. The tours 122 and 325 do not have any subtours, but the tours 448 and 555 do have a subtour. Based on the criteria of subtour analysis, it is found that the tour 448 has a single work-based subtour, whereas tour 555 has two work-based subtours. The data field ForwardBackwardPass shows the identification of trips, which are required to prepare a dataset for the intermediate/secondary destination choice model of a tour. The terms “Forward” and “Backward” identify the forward and backward trips of a tour with respect to home and primary destination. The term “LoopForward” identifies the intermediate trips of a work- (or school-) based subtour. The intermediate stop locations of a subtour are identified for work and school tour purposes only.

4.15 MoreThan100Mile

This data field identifies the tours in the database, where trip length is more than 100 miles (considering TAZ centroid to centroid distance). This type of tour is identified with code 0 and all remaining tours are identified with code 1. All tours coded with 0 are not considered in the logit model development process. An example is shown in Table 28 below.

Table 28: A Trip-Tour Table with Trips Over 100 Miles

ID	TripID	originTAZ	DestinationTAZ	TourID	MILE	MoreThan100Mile
111	1	3303	3304	44	10	0
112	2	3304	3217	44	98	0
113	3	3217	3303	44	106	0
555	1	4500	4675	98	8	1
556	2	4675	4577	98	15	1
557	3	4577	4500	98	20	1

In the above table, we find there are two tours, i.e. TourID 44 and 98. The MoreThan100Mile data field is coded with 0 for tour ID 44, because trip length of one of the trips constituting of the same tour, is more than 100 Mile. This tour is not considered in our model development.

4.16 TourModeAcceptance

This data field identifies the tours with tour mode Air, Other and code them with 0. The remaining tours are coded with 1. Tours coded with 0 are not considered for logit model development.

4.17 AcceptedTour

In the present study, many tours are ignored in the final selection of the model development process. The information pertaining to these tours are no longer used in any of the choice models. The accepted tours are coded with 1 while non-accepted tours are coded with 0. The non-acceptance of tours is decided based on the data fields called “TourModeAcceptance” and “MoreThan100Miles”.

5. Tour and Day Pattern Development

This section provides an overview of the tour and day pattern development in CSTDM09. The tour as a unit of analysis in the development of a tour-based model represents a closed or half-closed chain of trips starting and ending at home or at work. Each tour includes at least one destination and at least two successive trips. From the travel surveys, the number of trips that an individual makes, along with attributes of the trips, was collected. Tour and day pattern development connects the person-trips to a trip chain by time of day, travel activities, and stop sequence. The tour records were used for tour-based model development and estimation.

Tour and day pattern development was based on 4 travel surveys (California Statewide Travel Survey, SCAG, SANDAG and MTC) and was assembled using Microsoft Excel Macro VBA.

5.1 Specifications

A tour is composed of travel activity and a time period within a day. Both of them have a coding scheme for simple description (see Table 29 and 30).

Table 29: Time of Day Definition

Code	Time Period
1	Early Offpeak (3AM – 6 AM)
2	AM Peak (6 AM – 10 AM)
3	Midday (10 AM – 3 PM)
4	PM Peak (3 PM – 7 PM)
5	Late Offpeak (7 PM – 3 AM)

Table 30: Travel Activity Definition

Code	Activity	
O	Home	Home
W	Work	Mandatory
S	School	
H	Shop	Non-mandatory maintenance
P	Personal Business	
C	Escort	
T	Eat	Non-mandatory discretionary
L	Social	
R	Recreation	

The mode and purpose of each tour are also identified. As shown in Table 31, the tour purpose is grouped into three categories: HBW (Home based work), HBS (Home based school), and HBO (Home based other).

Table 31: Tour Purpose Definition

Code	Purpose	Description
1	HBW	Home based work, primary activity is work
2	HBS	Home based school, primary activity is school
3	HBO	Home based other activities

The tour mode is the mode selected for the overall tour, and is always the "fastest" mode used among all the trips of a given tour (see Table 32). So a person that takes auto access transit to work, walks to lunch and back, then takes walk access transit home is considered to be on an auto access transit tour. Even if ten of the trips are walking, and one trip is by HOV 3+, the tour is an HOV 3+ tour. For tour records where the access mode or occupancy number could not be determined, the codes Transit-No access Specified (6) and HOV-Not Specified the number of occupancy (10) were used.

These two special modes were appropriately managed in the mode choice model, as described above in section 4.7.

Table 32: Tour Mode

Mode	Name
1	SOV
2	HOV2
3	HOV3+
4	Transit- Walk Access
5	Transit- Car Access
6	Transit-No Access Specified
7	Walk
8	Bicycle
9	School Bus
10	HOV-Not Specified the Number of Occupancy

5.2 Methodology

From the household travel survey, we collected the personal weekday travel diaries, which include the number of trips that individual makes along with the attributes of the trips. Tour and day pattern development connects the person-trip-destination activities to a trip chain by time of day and stop sequence. It was used for tour based model development and estimation.

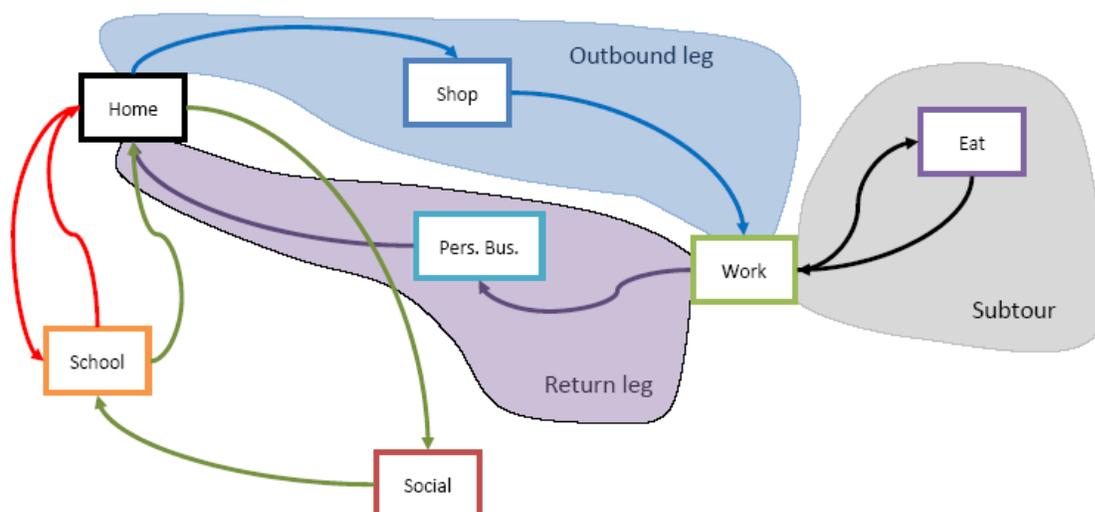


Figure 2: A Typical Day Tour

Three home based tours are shown in Figure 2. The red tour is a simple tour to school and back. Simple tours are those with only one destination; that is, they are tours from home to a location, and then back home without any additional trips. The green tour is a complex tour, i.e. one with more than one stop. In this case, the tour goes from home to a social activity, then to school, then back home. This tour would be considered a school tour, because of the presence of a school activity. If the school activity was replaced by another kind of activity (except work), such as leisure, then the tour would be considered a social tour because the first destination is the social visit. In this tour, school is the primary destination, and the social event is a secondary destination.

The third tour is shown in blue, black and purple. The tour goes from home to a shopping activity to work, then to an eating activity and back to work, then to a personal business activity and finally back to home. This is, of course, a complex tour, and the work location is the primary destination. This tour can be divided into three legs; an outbound leg, a subtour and a return leg. The outbound leg is the one shown in blue, from home to work via a shopping location. The subtour is shown in black, and involves going to eat and returning to work. Only work and school activities can have subtours, and only the first subtour is considered.

All tours must start and end at home; once a person travels home, they have ended their tour and started a new tour. For records with a partial tour, for instance if a person starts or ends their day outside the home, the partial tour that is in the record is considered to be a tour.

To develop the tour and day pattern, the first step is to identify the destination activity and the time of day for each trip. Next, the primary destination for each tour is determined, along with secondary destinations or subtours if present.

In the tour development process, the following rules are applied:

1. For work tours, the primary destination is work location even if the person stops for another activity on the way to work. The trip from home to work is the outbound leg and the trip from work to home is the return leg. The presence of a work-based subtour should be checked. The tour purpose is home-based work.
2. For school tours without work activity, the primary destination is school location. The trip from home to school is the outbound leg and from school to home is the return leg; presence of a school-based subtour should be checked. The tour purpose is home-based school.
3. A tour with both work and school activities should refer to the person type of the traveler. If the traveler is a student, then the tour is a home-based school tour, otherwise it is a home-based work tour.
4. For a work or school subtour, check the TAZ ID of subtour origin and destination. If the out and back trip are in the same TAZ, then it is a subtour, otherwise the trips are a normal work tour for outbound/return leg without a subtour.
5. The primary destination for non-work, non-school tours is the first stop on the tour. These tours should have the first trip as the outbound leg, and the latter ones on the return leg.
6. A tour with more than three successive work activities should be incorporated into one work activity keeping the first and last time of day code, since these people are probably delivery workers. The short distance commercial vehicle model with the short distance personal travel model should cover this case.

7. In the case of incomplete tours with only one home end there are two possibilities. If this tour includes more than three trips, it is treated as a complete home-based tour with outbound/subtour/return legs. Otherwise, it is simply an outbound or return based on the position of "O".
8. For a tour "overview", if there are work or school activities, select "W" or "S" with the outbound and return time of day; otherwise, use the first stop activity with the outbound and return time of day. Each tour has an overview record, which describes the overall nature of the tour.

5.3 Tour Data Format and Structure

The tour data file is an .xlsx data file. It is readable and writable directly from Excel, and also could be imported to Access database for cross-table query to get the household and person information for each tour. See Table 33 below.

Table 33: Tour Data File Structure

No.	Name	Type	Description
1	TourID	Number	Tour ID
2	HouseholdID	Number	Household ID in the survey database
3	PersonID	Number	Person ID marked in the survey database
4	HB_Tour	Text	Home-based tour
5	Purpose	Number	Tour purpose
6	Overview	Text	Tour overview
7	Outtime	Number	Tour outbound leg time of day
8	Returtime	Number	Tour return leg time of day
9	Priactivity	Text	Primary activity of a tour
10	Priacticode	Number	Code for the primary activity of a tour
11	Outbound	Text	Outbound leg of the tour
12	Subtour	Text	Subtour
13	Return	Text	Return leg of the tour
14	Mode	Number	Tour mode

5.4 California Journey Characteristics

5.4.1 Key Indicators

In the four travel surveys, there are a combined 252,595 trip records representing 65,629 travelling persons. Around 16% of the sampling population, 12,663 persons, stay at home on a typical weekday. Key indicators of tour and day patterns are summarized as below:

- **Number of tours:** from the travel surveys, 93,906 tours, including half tours (outbound or return) and entire tours (outbound and return), are observed for all purposes. Of these tour records, 93,711 tours have at least one leg (one home end).
- **Tour rates:** the average tour rate for travelling persons is 1.431 tours/person; the average tour rate for all surveyed person is 1.199 tours/person.
- **Tour rates by purpose:** the average tour rate for home based work (HBW) purpose is 1.074 tours/person; the average tour rate for home based school (HBC) purpose is 1.04 tours/person; the average tour rate for home based other (HBO) purpose is 1.446 tours/person. Tour rates for HBW and HBS are significantly lower than HBO.
- **Trip rates:** the average trip rate for travelling persons is 3.849 trips/person; the average trip rate for all surveyed person is 3.226 trips/person.
- **Number of trips per tour:** the average number of trips in a tour is 2.69 trips/tour.

The detailed tables and graphics of the weekday California statewide journey characteristics are presented below focusing on the regional household travel behavior which will be used during the model development.

5.4.2 Tour Purpose

For weekday travel, non-mandatory maintenance and non-mandatory discretionary travel activities account for around half of the total travel demand. The next largest proportion of travel is home based work, which accounts for 34.3% of all travel demand. Table 34 below presents absolute number of tours by purpose.

Table 34: Tour Proportions by Purpose

Purpose	Count	Percentage
HBO	46,507	49.5%
HBS	14,963	15.9%
HBW	32,241	34.3%
None home ends	195	0.2%
Total	93,906	100.0%

5.4.3 Tour "Overview"

As stated above, each tour has an overview record, which includes the primary tour activity with outbound and return time of day patterns. Tour overview describes the overall nature of the tour. In California, the trip overview "2W4" (a work tour which begins from home during the AM Peak 6AM-10AM and returns home during the PM Peak 3PM-7PM; see Table 12 and Table 13 for code definitions) accounts for 17.2% of all trips. Other common tour overviews and their proportions are outlined in Table 35 below.

Table 35: Tour Overview Proportions

NO.	Tour Overview	Count	Percentage
1	2W4	16,188	17.2%
2	2S4	7,294	7.8%
3	2S3	4,835	5.1%
4	2W5	3,421	3.6%
5	3H3	3,056	3.3%
6	2C2	3,008	3.2%
7	2W3	2,928	3.1%
8	3P3	2,661	2.8%
9	4H4	2,150	2.3%
10	3W4	2,135	2.3%

5.4.4 Tour Occurrence Frequency

11,345 categories of home based tour represents 93,906 tours. Table 36 illustrates the top 10 tour occurrences by frequency. The first most common tour pattern is "O2W4O". It starts at home, leaving home on AM Peak (6 AM – 10 AM) to work and back to home on PM Peak (3PM– 7PM).

Table 36: Tour Occurrence Frequency

No.	Tour	Count of Tours	Percentage
1	O2W4O	9,147	9.74%
2	O2S4O	5,089	5.42%
3	O2S3O	4,079	4.34%
4	O2C2O	2,513	2.68%
5	O3H3O	1,983	2.11%
6	O4C4O	1,701	1.81%
7	O4H4O	1,671	1.78%
8	O3C3O	1,667	1.78%
9	O2W3O	1,652	1.76%
10	O3P3O	1,357	1.45%

Figure 3 illustrates the top 20 occurrence tours by person type. More than 90% of "O2W4O" tours are made by full-time workers. Students are the key participants in "O2S4O", "O2S3O" and "O4R4O", which means parents usually take them to attend some recreation activities after school time when parents are off work.

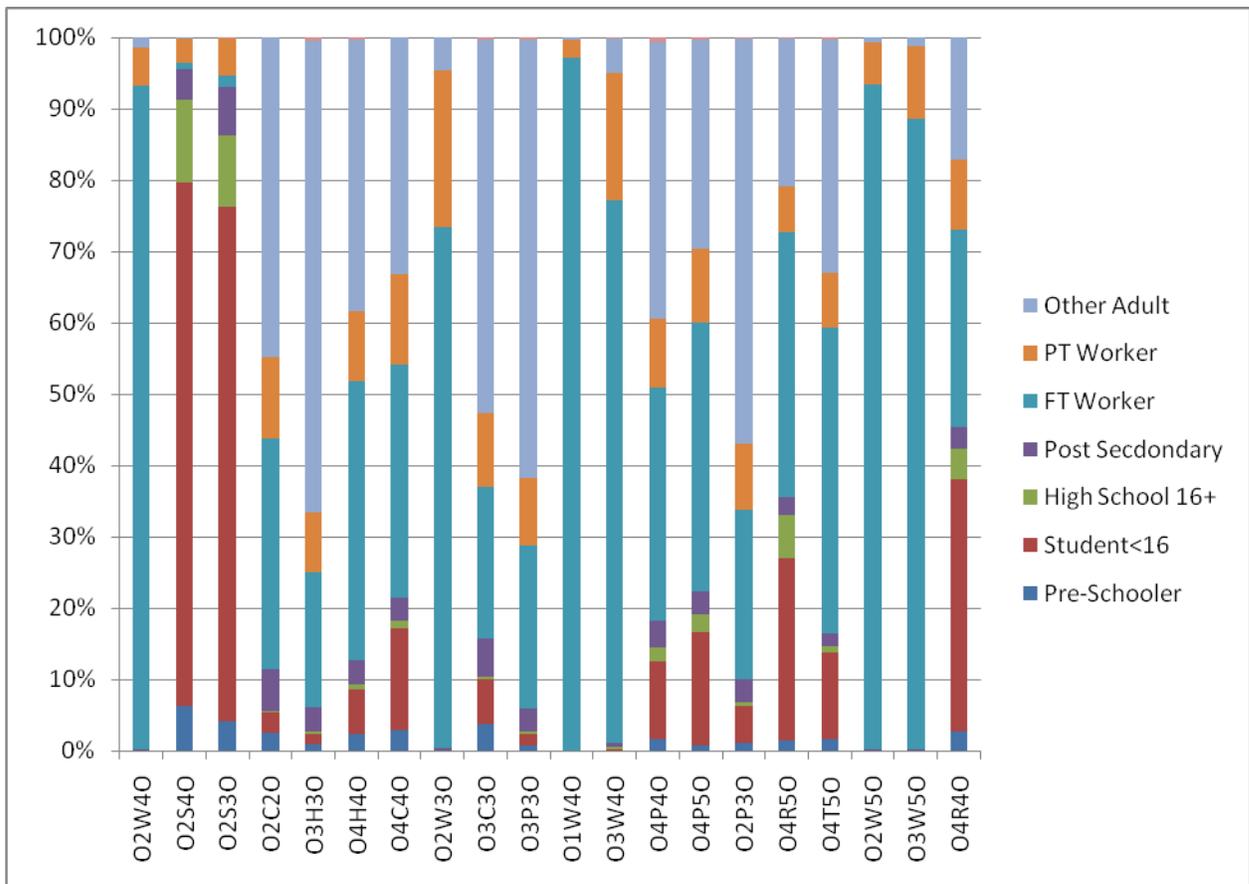


Figure 3: Top 20 Occurrence Tours by Person Type

Table 37: Tours by Person Type

Person type	Count	Percentage
Adult other	11,930	12.7%
Full-time worker	40,800	43.4%
K12 Student < 16 years old	14,289	15.2%
K12 Student 16+ years old	2,737	2.9%
Part-time workers	6,305	6.7%
Postsecondary student	6,237	6.6%
PreSchooler	3,258	3.5%
Senior (65+)	8,350	8.9%
Total	93,906	100.0%

5.4.5 Travel Activities

Table 38 shows the activity distribution for all tours. The most common activity is work (34.3%), followed by school (15.9%) and escort (11.2%).

Table 38: Proportion of Travel Activities

Activity	Count	Percentage
W	32,241	34.3%
S	14,963	15.9%
C	10,518	11.2%
H	9,815	10.5%
P	9,892	10.5%
L	5,941	6.3%
R	5,833	6.2%
T	4,508	4.8%
Non-home ends	195	0.2%
Total	93,906	100.0%

Figure 4 illustrates the travel activities for each person type. It shows the proportion of different travel activities made by each person type.

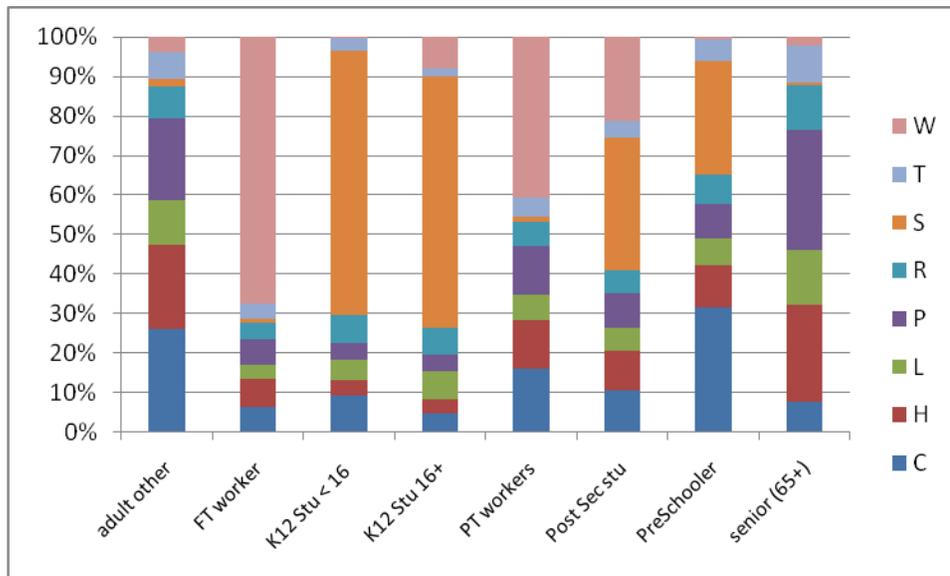


Figure 4: Travel activities by Person Type

5.4.6 Time of Day Patterns

Figure 5 depicts outbound time patterns by activities, and Figure 6 provides the tour return time patterns distribution. Household generation of individual tours for maintenance activities have a more even distribution with regard to time of day. 41% of escorting travel (drop off or pick up passengers) leave home on morning peak hour and 29% would be back to home on morning peak hour without further trip.

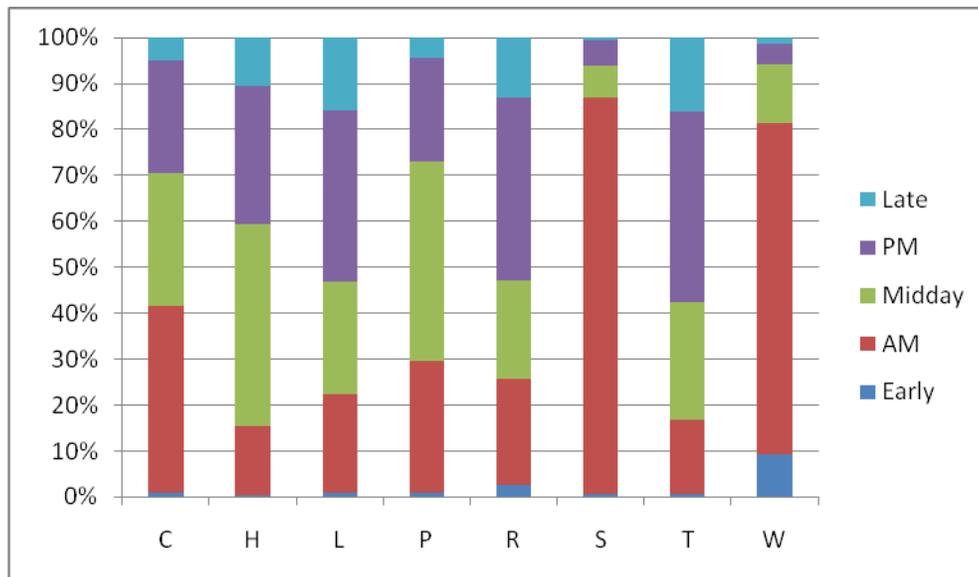


Figure 5: Outbound Time Distribution

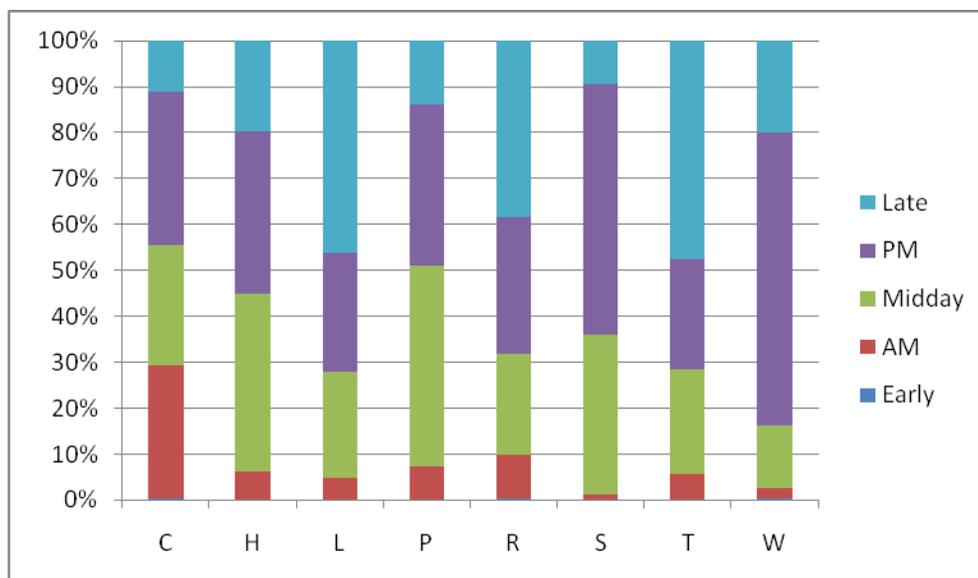


Figure 6: Return Time Distribution

6. Tour Statistics

This document focuses on the descriptive statistics of the tours, which are used in the proposed development of the Tour-based Short Distance Person Trip Model for the entire state of California, as part of the CSTDM09. The tours are derived from the trip table. Note that the tables in this section are the raw numbers of records, without

expansion. The trip table of the master database is constituted of trip records from four sources of travel survey data. These sources are:

1. California State Wide Travel Survey
2. MTC (Bay Area) Travel Survey
3. SANDAG Travel Survey
4. SCAG Travel Survey.

The total number of trip records that are used to derive the tour information is 252,595. The percentage share of the trip records from the different sources of travel survey data is given in Table 39 below.

Table 39: Percentage Share of Trip Records by Trip Data Source

Source of Travel Data	Count of Trip records	Percentage
California Statewide Travel Survey Data	96,120	38%
MTC (Bay Area) Travel Survey	65,546	26%
SANDAG Travel Survey	24,695	10%
SCAG Travel Survey	66,234	26%
Total	252,595	100%

These trip records are represented by 65,629 trip makers (persons), who had trips on a typical working day when all schools are supposed to be in session. The 65,629 persons are represented by 33,395 households in the State of California. The trip table of the master database is converted into a tour table.

There are 93,906 tours derived from the trip table. Out of 93,906 tours, 4,058 tours are found to be incomplete household-based tours and the remaining 89,848 tours are found to be complete household-based tours. Table 40 shows the percentage share of the tour records from the different sources of travel survey data.

Table 40: Percentage Share of Tour Records by Trip Data Source

Source of Travel Data	Count of Tour Records	Percentage
California Statewide Travel Survey Data	33,729	35.92%
MTC (Bay Area) Travel Survey	25,307	26.95%
SANDAG Travel Survey	9,105	9.70%
SCAG Travel Survey	25,765	27.44%
Total	93,906	100%

Table 41 table shows the distribution of tours by tour purpose and whether the tour is complete or incomplete.

Table 41: Distribution of Tours by Tour Purpose and Tour Completeness

Tour Purpose	Tour Purpose Code	Whether Complete Incomplete Tour	Count of Tours	Total Count Of Tours
Work	1	Incomplete	1,823	32,331
		Complete	30,508	
Other	2	Incomplete	1791	46,593
		Complete	44,802	
School	3	Incomplete	444	14,982
		Complete	14,538	
	Total			93,906

Table 42 shows the distribution of the number of persons by number of tours per person per day. This count includes both complete and incomplete tours.

Table 42: Distribution of Persons by Number of Tours/Person/Day

Number of Tours Per Person	Number of Persons	Percentage
1	44,281	67.47%
2	16,109	24.55%
3	3,937	6.00%
4	993	1.51%
5	239	0.36%
6	61	0.09%
7	9	0.01%
Total	65,629	100%

Table 43 shows the distribution of the number of tours per person per day in the master database.

Table 43: Distribution of Tours by Number of Tours/Person/Day

Number of Tours Per Person	Number of Tours	Percentage
1	44,281	47.2%
2	32,218	34.3%
3	11,811	12.6%
4	3,972	4.2%
5	1,195	1.3%
6	366	0.4%
7	63	0.1%

Table 44 shows the distribution of the number of trips per tour.

Table 44: Distribution of Tours by Number of Trips/Tour

Number Of Trips Per Tour	Count of tour	Percentage
1 (Incomplete Tour)	1,853	1.97%
2	59,376	63.23%
3	16,084	17.13%
4	8,509	9.06%
5+	8,084	8.61%
Total	93,906	

Table 45 shows the distribution of tour mode in the tour table.

Table 45: Distribution of Tour Modes

Tour Mode Name	Count of Tours	Percentage
Air	3	0.00%
Bicycle	843	0.90%
SOV	50,542	53.82%
HOV-2	16,172	17.22%
HOV-3	9,176	9.77%
HOV-Unspecified	8,269	8.81%
School Bus	1,177	1.25%
Transit-Car Access	430	0.46%
Transit-Walk Access	960	1.02%
Transit-No Access Specified	665	0.71%
Other	191	0.20%
Walk	5,478	5.83%
Total	93,906	

It is observed from the above table that there are some tours in the master database where tour modes are “Air” and “Other”. In the first phase of the tour mode choice model, these two tour modes are not considered. The SDPTM also does not consider trips with trip length more than 100 miles (considering distance between centroids of origin and destination TAZ). Therefore, a number of tours are omitted from the consideration of the proposed SDPTM travel demand model development.

1. Tours with tour mode “air”
2. Tours with tour mode “other”
3. Tours with tour mode “walk”, where trip length is more than 5 mile
4. Tours with tour mode “bike”, where trip length is more than 30 mile
5. Tours with tour mode “school bus”, where tour purpose is “work”
6. Tours with tour mode “school bus”, where tour purpose is “other”
7. Any trip length of a tour that is more than 100 miles
8. Any tour with a tour origin and primary destination distance that is more than 100 miles (considering centroid to centroid distance of origin and destination TAZ)

The total number of tours that are finally considered as inputs in the travel model is 92,727. These tours are represented by 64,898 persons and by 33,151 households. Table 46 shows the distribution of the tours by tour purpose and whether the tour is complete or incomplete.

Table 46: Distribution of Tours by Tour Purpose and Tour Completeness

Tour Purpose	Whether Complete Incomplete Tour	Count of Tours	Percentage
Work	Incomplete Tour	1,733	5%
Work	Complete Tour	30,292	95%
Other	Incomplete Tour	1,580	3%
Other	Complete Tour	44,295	97%
School	Incomplete Tour	433	3%
School	Complete Tour	14,394	97%
Total Tours		92,727	

Table 47 shows the distribution of the number of trips per tour in the refined tour table.

Table 47: Distribution of the Number of Trips per Tour

Number Of Trips Per Tour	Count of Tours	Percentage
1	1,735	2%
2	58,719	63%
3	15,916	17%
4	8,416	9%
5+	7,941	9%
Total Tours	92,727	

Table 48 shows the distribution of tours per person per day in the refined tour table.

Table 48: Distribution of the Number of Persons by Number of Tours

Number Of Tours Per Person Per Day	Count of Persons	Percentage
1	43,594	67.17%
2	16,074	24.77%
3	3,930	6.06%
4	991	1.53%
5	239	0.37%
6	61	0.09%
7	9	0.01%
Total	64,898	

Table 49 shows the modal share of 92,727 refined tours.

Table 49: Percentage Share of Tour Mode

Tour Mode	Count of Tours	Percentage
SOV	50,142	54.1%
HOV-2	16,042	17.3%
HOV-3	9,126	9.8%
HOV-Unspecified	8,172	8.8%
School Bus	1,132	1.2%
Transit-Car Access	426	0.5%
Transit-Walk Access	956	1.0%
Transit-No Access Specified	664	0.7%
Bicycle	833	0.9%
Walk	5,234	5.6%
Total Tours	92,727	

Table 50 shows modal share by tour purpose. This table will help construct the tour mode choice model for each tour purpose.

Table 50: Percentage Share of Tour Mode by Tour Purpose

Tour Purpose	Tour Mode Name	Count of Tours	Percentage
Work	SOV	26,981	84.2%
Work	HOV-2	2,013	6.3%
Work	HOV-3	441	1.4%
Work	HOV-Unspecified	596	1.9%
Work	Transit-Car Access	344	1.1%
Work	Transit-Walk Access	539	1.7%
Work	Transit-No Access	269	0.8%
Work	Bicycle	265	0.8%
Work	Walk	577	1.8%
Total		32,025	
Other	SOV	20,832	45.4%
Other	HOV-2	10,769	23.5%
Other	HOV-3	5,662	12.3%
Other	HOV-Unspecified	4,867	10.6%
Other	Transit-Car Access	47	0.1%
Other	Transit-Walk Access	227	0.5%
Other	Transit-No Access	288	0.6%
Other	Bicycle	333	0.7%
Other	Walk	2,850	6.2%
Total		45,875	
School	SOV	2,329	15.7%
School	HOV-2	3,260	22.0%
School	HOV-3	3,023	20.4%
School	HOV-Unspecified	2,709	18.3%
School	School Bus	1,132	7.6%
School	Transit-Car Access	35	0.2%
School	Transit-Walk Access	190	1.3%
School	Transit-No Access	107	0.7%
School	Bicycle	235	1.6%
School	Walk	1,807	12.2%
Total		14,827	

Tables 51 and 52 show the subtour-related information about the tours in the refined tour table. The subtour information is extracted only for work tour and school tour purposes. Table 51 shows the distribution of the existence of subtour.

Table 51: Percentage Share of the Existence of Subtour by Tour Purpose

Tour Purpose	Whether Sub Tour Exists	Count of Tours	Percentage
Work	No	27,766	86.70%
Work	Yes	4,259	13.30%
School	No	14,795	99.78%
School	Yes	32	0.22%

Table 52 shows the distribution of the total number of subtours per tour.

Table 52: Distribution of the Number of Subtours

Tour Purpose	Number Of Subtours	Count of Tours	Percentage
Work	0	27,766	86.70%
Work	1	3,727	11.64%
Work	2	399	1.25%
Work	3+	133	0.42%
School	0	14,795	99.78%
School	1	29	0.20%
School	2	2	0.01%
School	3	1	0.01%

Currently, 92,727 tours (i.e., tour observations) are considered for Short-Distance-Person-Travel model. But there is still a possibility that some tour observations might be excluded from the estimation of demand model based on the trade-off of values among different attributes for each tour.

7. Notes on Travel Activities

The trip table of the master database reflects the activity called “Travel” (activity code in the master database is 6), both at the origin end and at the destination end of a trip. Its presence in the database is considerably less. The following paragraphs explain why this activity still exists in the master database. It is explained in two steps: Step 1 describes the existence of travel activity in the destination end of a trip and Step 2 describes the existence of travel activity in the origin end of a trip.

7.1 Step 1:

There are 223 trip records where destination activity of a trip is "Travel". Out of 223 trips, 217 trip records come from the SANDAG database. The SANDAG database has two activities that come under travel activity. One is called "change mode of transportation" (source activity code 23), and the other is called "loop trip that begins and ends at the same place" (Source activity code is 24). The original linked trip table of the SANDAG database does not reflect activity 23 (Change mode of transportation), but it does reflect activity code 24 (Loop trip that begins and ends at the same place) as an origin/destination activity of a trip. Therefore, the final database also reflects trip end activity of these trips as Travel activity. The activity code 24 does not imply any kind of mode change rather it implies a kind of activity that can be termed "personal business" i.e., the person had a loop trip because of his personal business/other. But currently this activity is reflected as "Travel" in the database. As said before, there are 217 such trip records exist in the SANDAG portion of the master database, where destination activity of a trip record is travel. Out of 217 trip records, 109 trip records belong to a complete household-based tour whereas remaining 108 trip records belong to an incomplete household-based tour. There are still 4 more trip records from California Statewide Travel Survey and 2 more trip records from MTC Travel Survey, where destination activity is travel. In the case of the California Statewide Survey, all 4 trip records are part of a "single-trip-incomplete-HH based tour". So, deleting these trip records from the trip table is not an option as it will delete the incomplete tour information that the persons actually had. It may also delete the person from person table.

In the case of MTC travel survey, the 2 trip records are part of a complete household-based tour. In both cases the destination activity is recognized as "Change mode of transportation". But if we look carefully at the trip table, it is not exactly that the person changes mode, but rather he or she had a trip for a different purpose (e.g. personal business) because it is unlikely that a person traveled to different zone just to change mode. Instead, personal business activity will better represent the destination activity instead of using change of mode. For example trip ID 232995 and ID 232996 are two consecutive trips of a person (PERNO 1 of HHID 119925). These two trips constitute a

complete household based tour. He/she went from home (Zone ID 1182) to other (travel) (Zone ID 1185) by Car (mode code 27) and came back by Transit-Rail (mode code 36). It is true that the person changed mode on his/her way back but actually he or she went to zone 1185 for some other purpose, may be some personal business. The trip IDs 243649, 243650, 243651 constitute a complete household-based tour. The trip table of this tour is given below in Table 53.

Table 53: Trip Table for Travel Activity

Trip ID	HHID	Perno	Source_O_ACT	Source_D_ACT	O_TAZ	D_TAZ	Source_Mode
243649	139971	1	1	17 (change of mode)	1831	1040	41
243650	139971	1	17	996 (not defined kind of personal business)	1040	1831	27
243651	139971	1	996	1	1831	1831	27

Two code replacements will be updated in the final database: activity 17 (change of mode) should be replaced with activity 13 (social activities) and mode code should be 27 instead of 41.

7.2 Step 2:

Now, if all the tours are deleted from the trip table, where destination activity of a trip (a tour is comprised of two or more trips) is “travel”, then we find as many as 69 trip records where origin activity of the trip is travel. The share of these 69 trip records in the database by different sources of travel survey data is given below in Table 54.

Table 54: Trip Shares by Travel Activity

Data Source	Share
California Database	66
MTC Database	1
SANDAG Database	2

All these trips belong to incomplete household-based tours and they become the first trip of that incomplete tour. In many cases, a single trip forms the incomplete tour. If such trip records are deleted from the master database, then we may lose the information of that person from the person table. Besides, the trip records of an incomplete tour are only used to find out the primary destination of the tour, provided the tour purpose is work or school. In both the cases, the origin of the tour is considered to be home, with the assumption that the person was in home sometime on that trip day. Based on this assumption, those trip records will be retained in the database, where origin activity was found as travel.

It should be mentioned that incomplete tour information is analyzed only to find out tour origin and tour destination (i.e. primary destination), provided the tour purpose is work or school. Incomplete tour information is not analyzed to find out intermediate destinations (i.e. secondary destination) and primary destination of the “other-purpose” tour.

Appendix A: Travel Survey Expansion

1. Introduction

To prepare a dataset suitable for model estimation and calibration in California, several datasets were combined to produce a single composite survey file, comprised of four household surveys.

As the data collected represents only a sample of all households and all travel in California on any given day, the data collected must be expanded to represent the universe of household travel. The methodology developed to expand the data collected from each establishment involved scaling each establishment across three variables; household size in terms of the number of members, household income level, and geographic location. The total number of households within each variable was used to determine the individual expansion factors.

2. Overview of process

The household expansion uses the Fratar/Furness method for expansion. This process works by adjusting each dimension of targets and applying them to each household. In general, each category of each dimension is assigned a weight, which is continually updated through each iteration, being set to the target for that weight divided by the scaled number of households in that category. A Python script was used to produce household level expansion factors through this iterative process.

2.1 Targets overview

The targets used were developed from the 2000 Synthetic Population, documented elsewhere. There were two dimensions of targets; one was a size/economic dimension and the other a geographic one. The size/economic dimensions were kept together to reflect the cross-correlation between these two aspects; this establishes separate targets for large and small low income households, for instance. The other set of dimensions was geographic; there were a total of 426 geographic groups. Each survey

household is located within one of the geographic and one of the socioeconomic groups, so each household gets an expansion factor that is the product of the appropriate geographic and the appropriate socioeconomic factor.

2.2 Socioeconomic dimension

The socioeconomic dimension was defined as a combination of household size (number of persons) and annual household income. Additionally, households where all members were over 65 were considered as separate categories. Each category is coded with a name, where the integer part indicates the household size (and if the household has all members 65+) and the decimal is the income category. The socioeconomic groups are summarized below:

Table 55: Socioeconomic Groups

Household Size Category	Income range (\$000, year 2000 dollars)							
	<10	10-25	25-35	35-50	50-75	75-100	100-150	150+
1 person	1.12		1.3	1.4	1.5	1.6	1.78	
2 people	2.1	2.2	2.3	2.5	2.5	2.6	2.7	2.8
3 people	3.12		3.3	3.4	3.5	3.6	3.7	3.8
4 people	4.12		4.3	4.4	4.5	4.6	4.7	4.8
5 + persons	5.12		5.34		5.56		5.78	
1 person, 65+	6501.12		6501.3	6501.4	6501.5678			
2+, all 65+	6502.12		6502.34		6502.5	6502.678		

The following table shows the number of survey households in each of these groups:

Table 56: Number of Survey Households by Socioeconomic Group

Household Size Category	Income range (\$000, year 2000 dollars)							
	<10	10-25	25-35	35-50	50-75	75-100	100-150	150+
1 person	1691		993	1412	1546	649	605	
2 people	317	1057	1035	1546	2670	2033	1881	958
3 people	789		540	679	1294	1019	810	428
4 people	600		469	600	1200	882	834	452
5 + persons	549		871		1160		544	
1 person, 65+	1225		484	380	487			
2+, all 65+	376		972		560	548		

2.3 Socioeconomic target regions

For the actual expansion process, the whole state was divided into 10 regions, and the socioeconomic targets and expansion factors developed independently for each region; because the geographic areas never crossed these boundaries, this was, in effect, ten individual expansions, each matching a part of the state. The reason for this was twofold; on one hand, there are regional variations in the frequencies of these groups; for instance, group 1.78 (one person under 65 households earning over \$100,000) represents 0.6% of the rural North/Sierra/Coast population, but 2.3% of the urban Central Bay population. The second reason is that there are different surveys present in different areas, with their own methodologies and sampling biases; as a minimum, regions were needed to represent the MTC Bay area, the SCAG area, the SANDAG survey area and the remaining parts of the state, which had only the statewide survey coverage. The regions are summarized in the table below:

Table 57: Target and Survey Households by Geographic Region

Additional Surveys	Region	Num. TAZ	Num. geo. Groups	Target HH	Survey HH
Statewide only	■ North / Sierra / Coast	388	66	960,545	4835
	■ SACOG / Central Valley	947	76	1,803,111	3805
MTC	■ North Bay	289	60	714,625	3313
	■ Central Bay	335	24	824,048	2660
	■ South Bay	437	35	1,016,698	4259
SANDAG	■ San Diego	453	45	1,079,287	4050
SCAG	■ Ventura / NW LA	390	28	958,360	2590
	■ Central LA	1007	35	2,527,248	4810
	■ Inland Empire	545	46	1,126,007	4859
	■ Orange Co.	340	11	963,306	1964

These regions are shown in the map below. Note that the North/Sierra/Coast region is not contiguous.

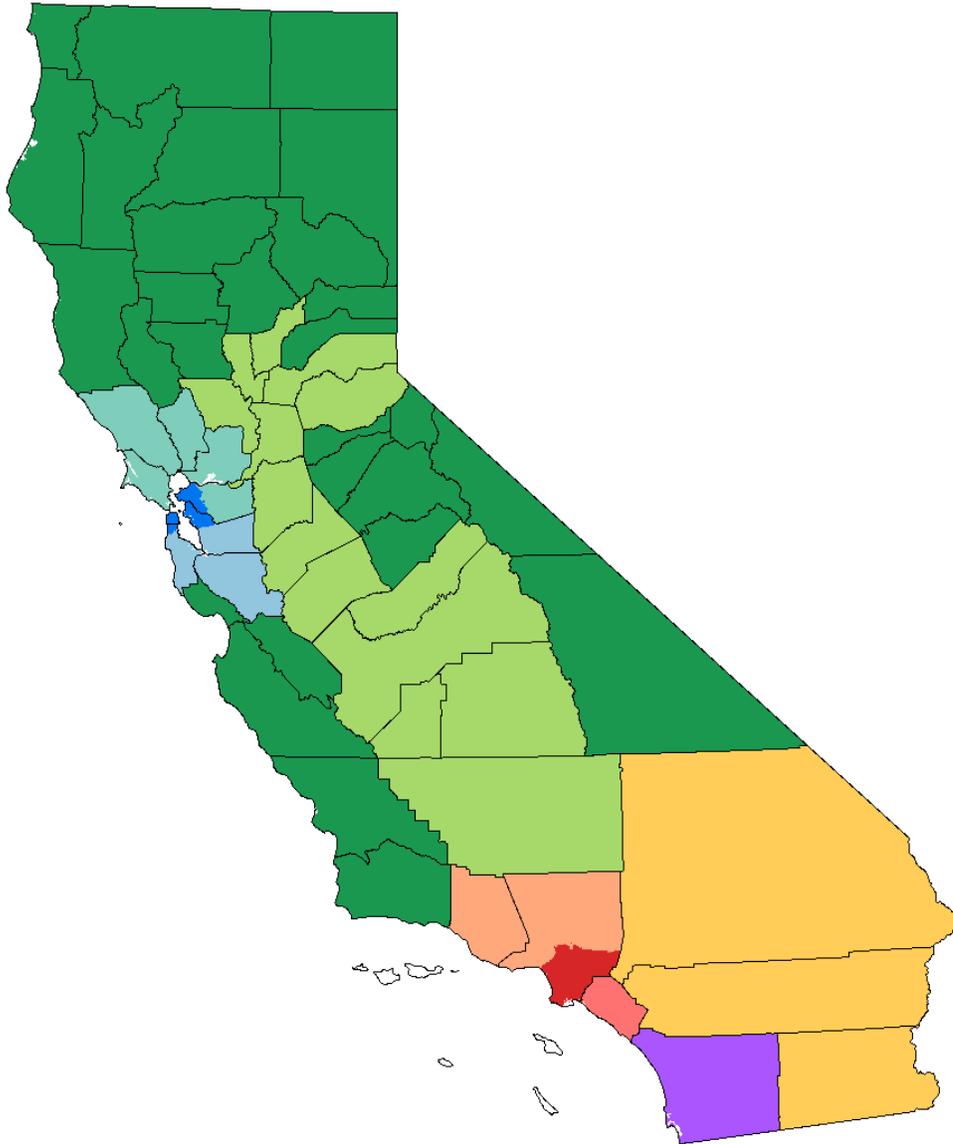


Figure 7: Geographic Regions in California

2.4 Geographic Groups

The geographic dimension was represented with a total of 426 geographic groups. These began based upon the LUZ definitions, although these were combined in a few areas where survey data was sparse and/or uneven; these areas include Sacramento, the San Gabriel Valley portion of Los Angeles and Fresno. In a few areas where survey data was especially frequent, existing LUZ were split, mostly in Ventura County. Note that for the purposes of the expansion, the old (Sept. 2009) TAZ system was used. As

the geographic groups were aggregations of TAZ, the subdivision of these TAZ would have minimal effect. The geographic groups had between 2 and 479 survey households, with 90% of the geographic groups in the 10-250 HH range.

The geographic groups are shown in the following three maps.

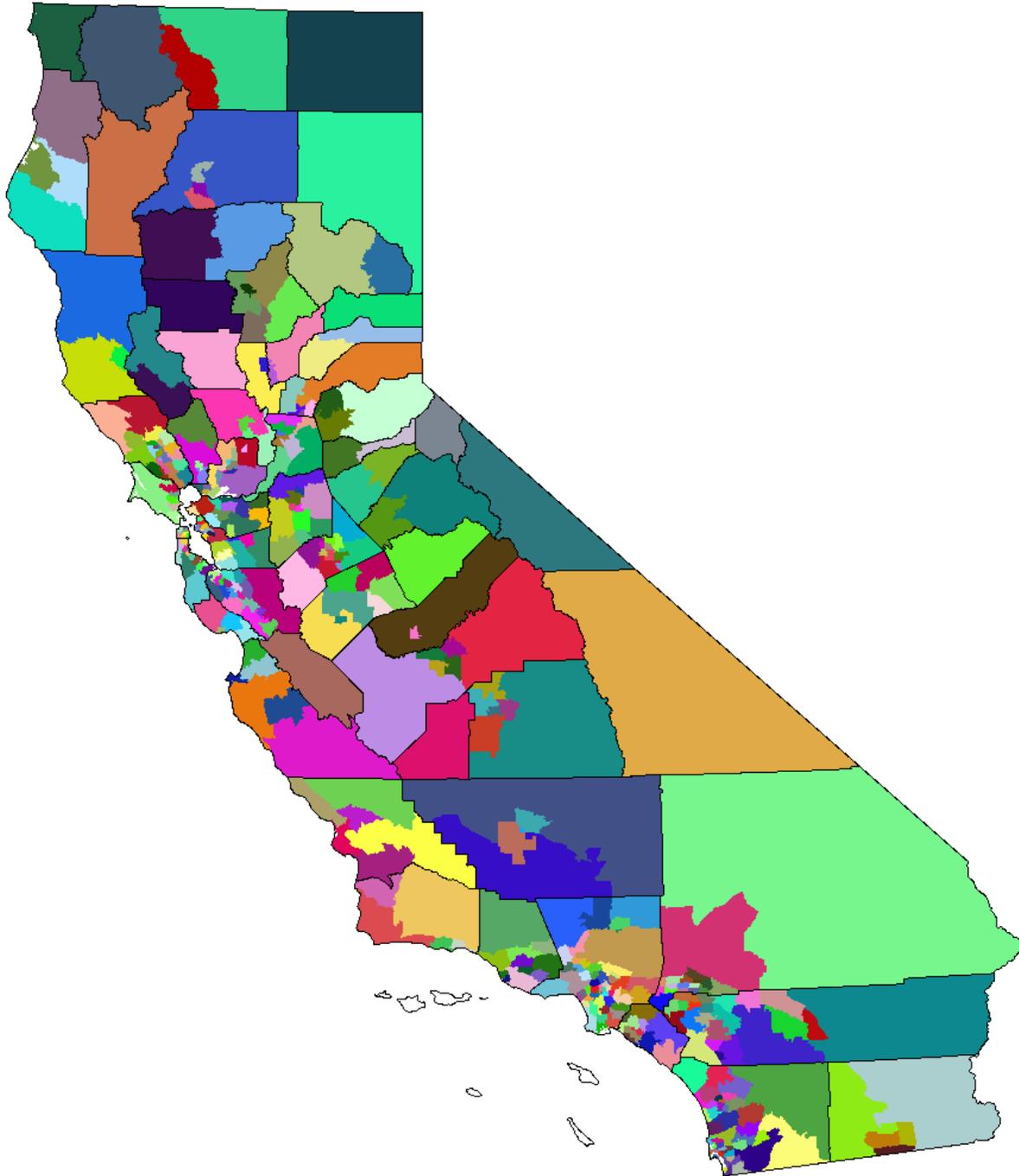


Figure 8: Statewide geographic group map.

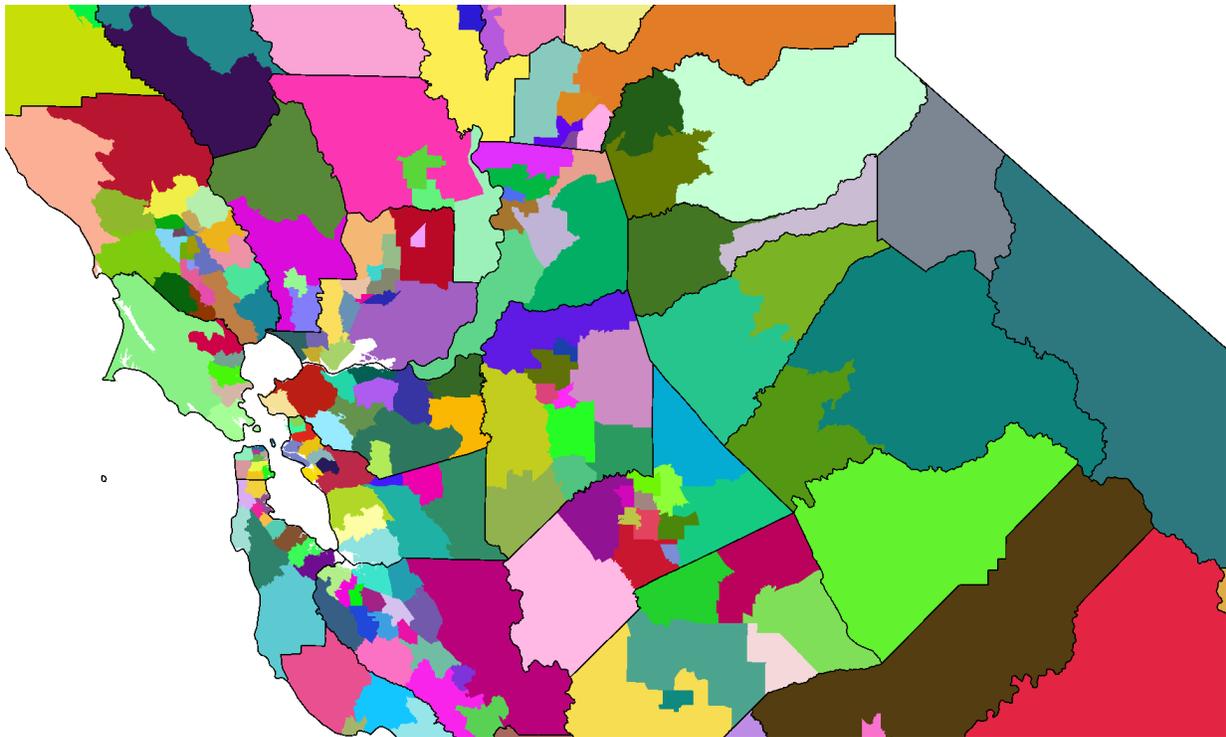


Figure 9: Geographic groups in San Francisco and Sacramento areas

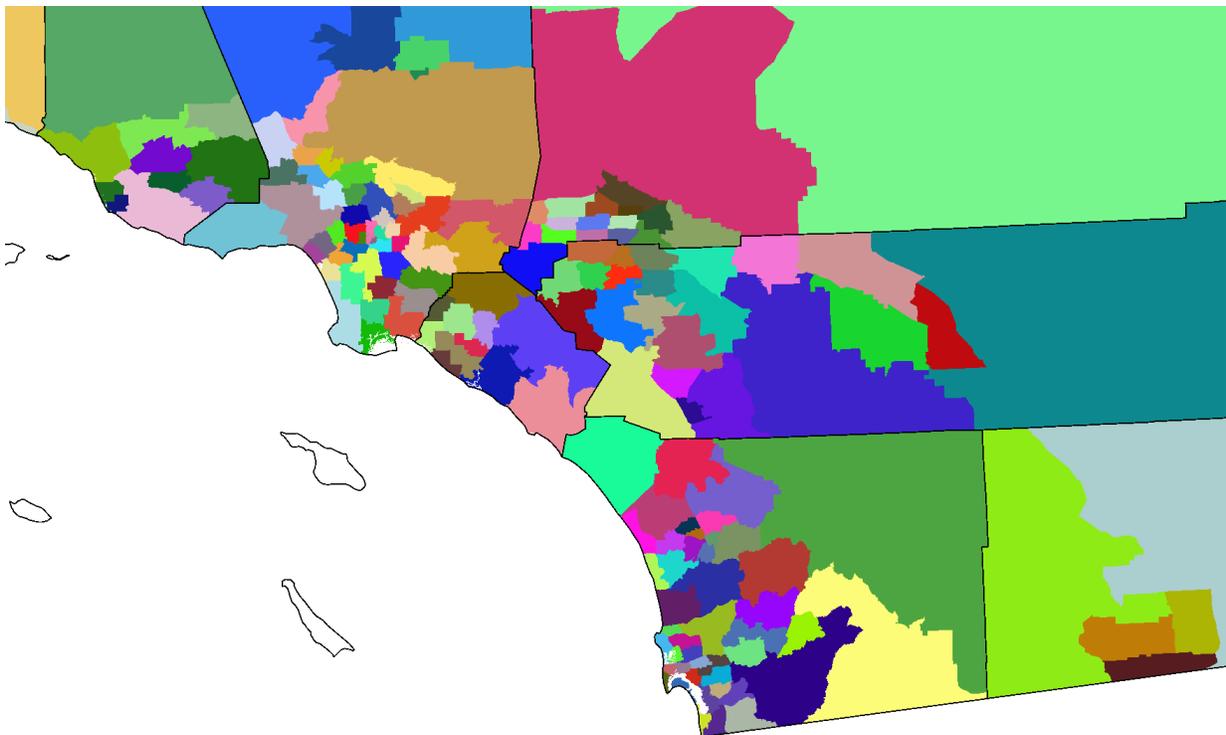


Figure 10: Geographic groups in LA / San Diego area

3. Expansion Results

For the expansion, 75 iterations of proportional fitting were used. The process worked well, and all targets were matched within the limits of machine precision (the largest error was geographic group 5600 in Orange County, with a target of 160962 households and an expanded survey total of 160961.999999, or a difference of 1E-6 households). Because of this exact match, no graphics are necessary to describe the fit versus the target. Instead, the expansion factor components will be discussed.

3.1 Socioeconomic Expansion Factors

The expansion factors for the socioeconomic classes are shown, by the 10 socioeconomic regions, in the table below.

Table 58: Expansion Factors by Socioeconomic Group

Socio Type	North / Sierra / Coast	SACOG / Central Valley	North Bay	Central Bay	South Bay	Ventura / NW LA	Central LA	Inland Empire	Orange Co.	San Diego
1.12	31.1	47.8	37.6	34.2	54.6	30.6	33.4	36.0	48.4	65.5
1.3	10.5	18.3	11.5	12.6	15.1	18.9	14.9	11.3	15.9	15.7
1.4	10.4	15.7	9.4	9.1	10.1	12.2	14.7	12.1	19.9	16.5
1.5	7.7	10.1	9.3	8.0	10.0	11.5	10.1	7.3	12.4	12.2
1.6	13.2	11.8	10.7	10.3	9.8	11.5	10.9	5.8	13.2	6.7
1.78	13.4	12.5	17.1	14.4	10.5	7.7	11.7	7.1	12.3	9.5
2.1	28.1	48.2	29.2	73.2	95.0	76.9	31.2	21.5	19.8	48.0
2.2	15.5	29.3	24.9	21.4	25.6	27.0	21.6	13.3	23.8	22.3
2.3	10.2	13.1	15.8	22.2	20.1	19.8	14.9	11.5	15.1	17.7
2.4	12.0	18.5	10.8	11.7	12.9	19.2	18.3	11.8	22.2	19.9
2.5	9.1	11.8	9.1	12.7	10.2	12.8	13.6	10.1	14.7	15.8
2.6	7.9	11.3	8.6	9.4	9.5	11.9	14.5	9.7	12.1	9.2
2.7	9.6	12.6	10.0	8.4	8.4	13.1	12.3	8.5	14.9	6.4
2.8	12.9	17.0	14.5	12.5	13.2	14.5	13.7	12.9	17.5	7.3
3.12	25.1	38.7	40.2	54.1	63.8	27.8	34.7	15.8	30.6	27.5
3.3	11.6	19.3	29.7	49.4	40.6	19.9	24.2	13.8	20.1	23.3
3.4	15.1	23.3	18.0	27.9	23.4	25.3	22.2	16.7	30.0	24.3
3.5	9.7	17.8	14.8	14.0	15.1	15.6	24.5	12.8	20.9	20.1
3.6	10.0	14.8	13.0	17.1	12.4	18.1	16.4	11.1	22.4	10.3
3.7	11.6	23.5	12.3	18.8	13.3	31.2	18.7	15.4	23.0	8.3
3.8	19.8	28.4	17.3	22.9	17.8	13.5	28.4	12.9	18.2	8.9
4.12	22.4	39.6	38.4	49.2	64.9	49.0	30.1	18.2	29.1	28.4
4.3	10.4	20.6	25.6	27.2	49.6	25.9	21.8	12.9	28.5	16.8
4.4	14.5	26.8	12.6	21.1	15.6	36.7	33.3	19.8	31.9	27.3
4.5	13.0	18.2	11.3	19.4	14.6	23.0	23.1	12.5	24.7	19.3

4.6	12.2	21.4	13.3	12.5	11.2	15.4	28.1	19.2	26.7	13.4
4.7	18.8	19.3	13.0	13.1	11.6	27.4	38.8	16.2	25.0	6.7
4.8	15.4	17.0	17.8	16.3	15.9	18.2	29.9	18.7	28.6	9.3
5.12	27.4	65.8	49.4	80.9	51.0	58.3	47.0	24.6	28.5	55.3
5.34	29.4	39.2	22.7	38.1	28.1	61.7	50.4	21.4	58.3	34.6
5.56	21.1	27.5	18.2	49.3	28.3	47.8	47.3	22.0	51.7	22.2
5.78	24.1	30.9	22.1	50.4	26.5	38.7	66.0	24.8	51.3	14.1
6501.12	17.2	38.1	28.0	47.9	32.4	29.6	37.3	24.0	33.2	14.8
6501.3	9.9	14.6	16.2	18.6	18.7	12.9	14.1	13.8	15.5	7.4
6501.4	14.4	15.7	10.5	15.4	11.4	12.0	14.5	10.9	20.6	6.8
6501.568	13.1	15.9	18.9	17.0	18.8	8.1	13.9	11.5	14.3	6.4
6502.12	16.9	29.3	29.7	55.1	52.9	22.2	39.3	23.9	37.5	15.4
6502.3	11.9	17.7	12.4	18.0	16.0	12.4	14.5	11.3	13.8	11.9
6502.5	9.2	12.3	13.9	14.3	9.5	5.9	12.7	11.7	17.1	10.3
6502.678	11.3	18.2	15.5	14.4	16.6	9.5	22.3	20.1	11.8	6.4

These expansion factors are shown in the following six graphs.

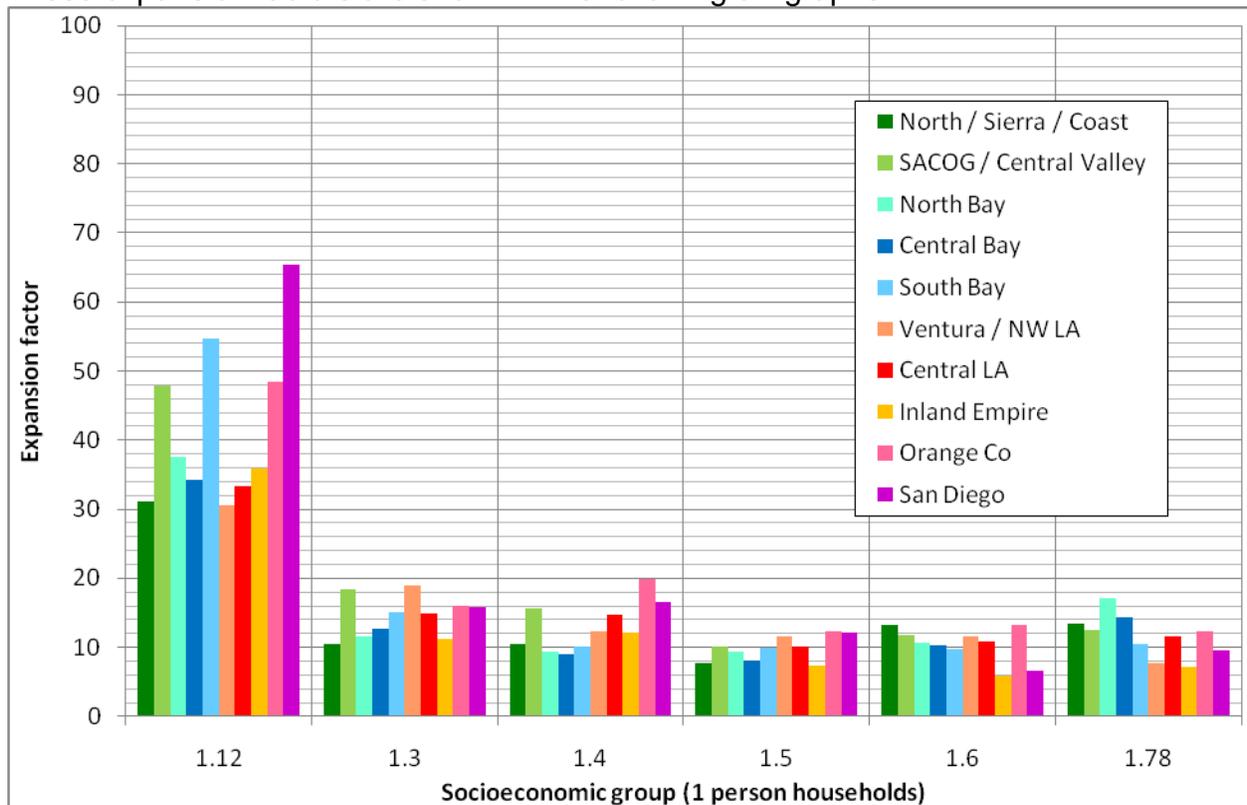


Figure 11: Expansion Factors by Socioeconomic Group (1 person households)

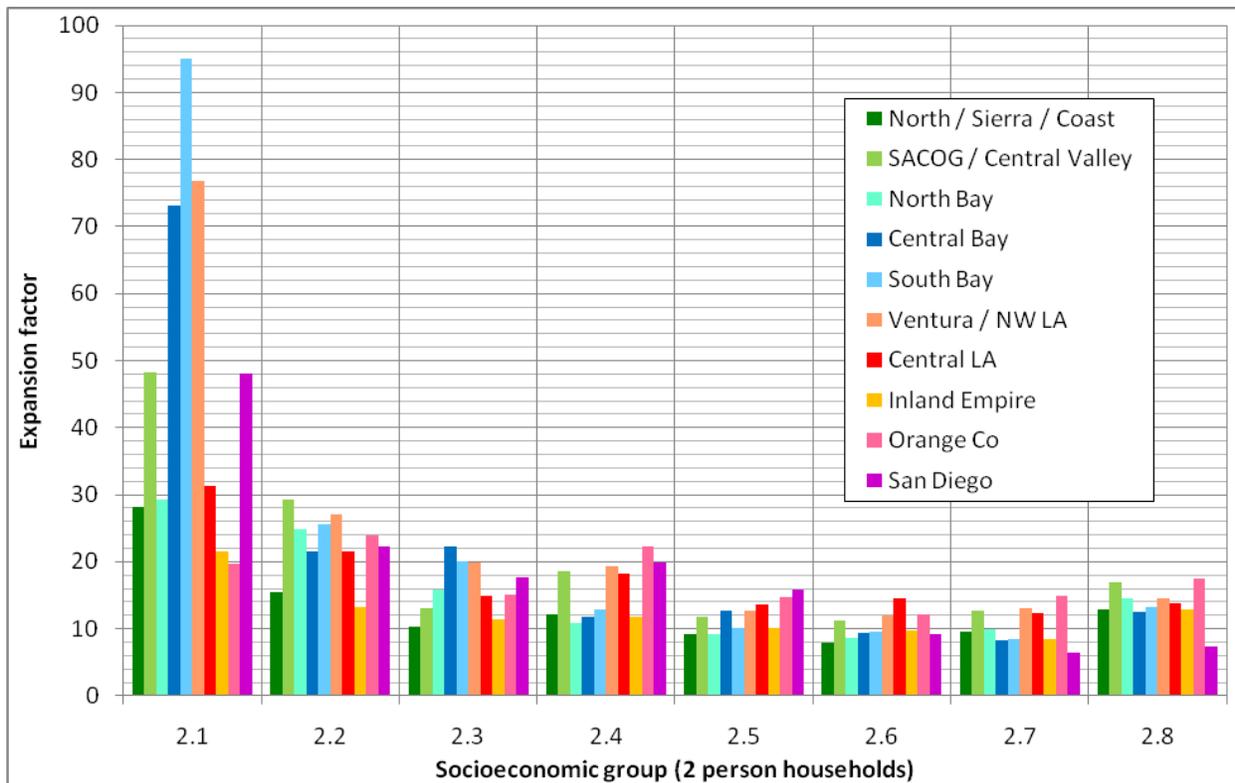


Figure 12: Expansion Factors by Socioeconomic Group (2 person households)

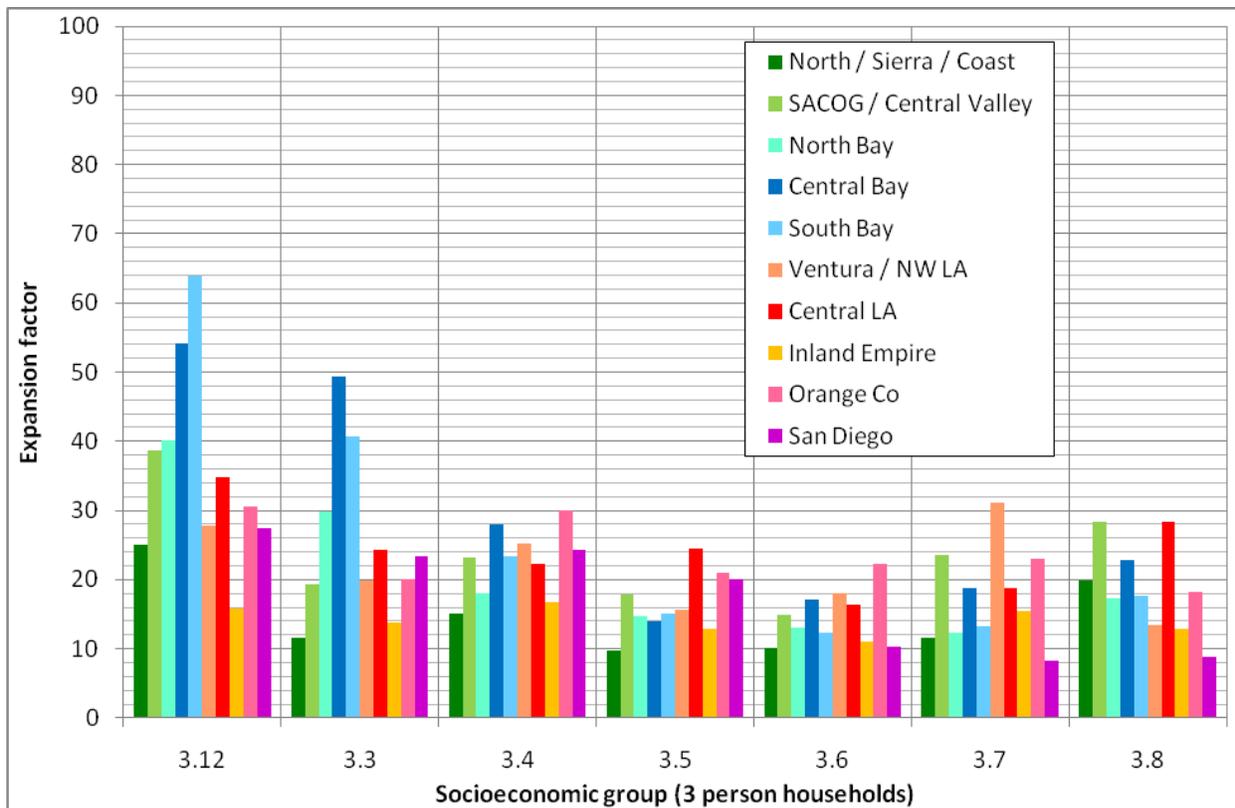


Figure 13: Expansion Factors by Socioeconomic Group (3 person households)

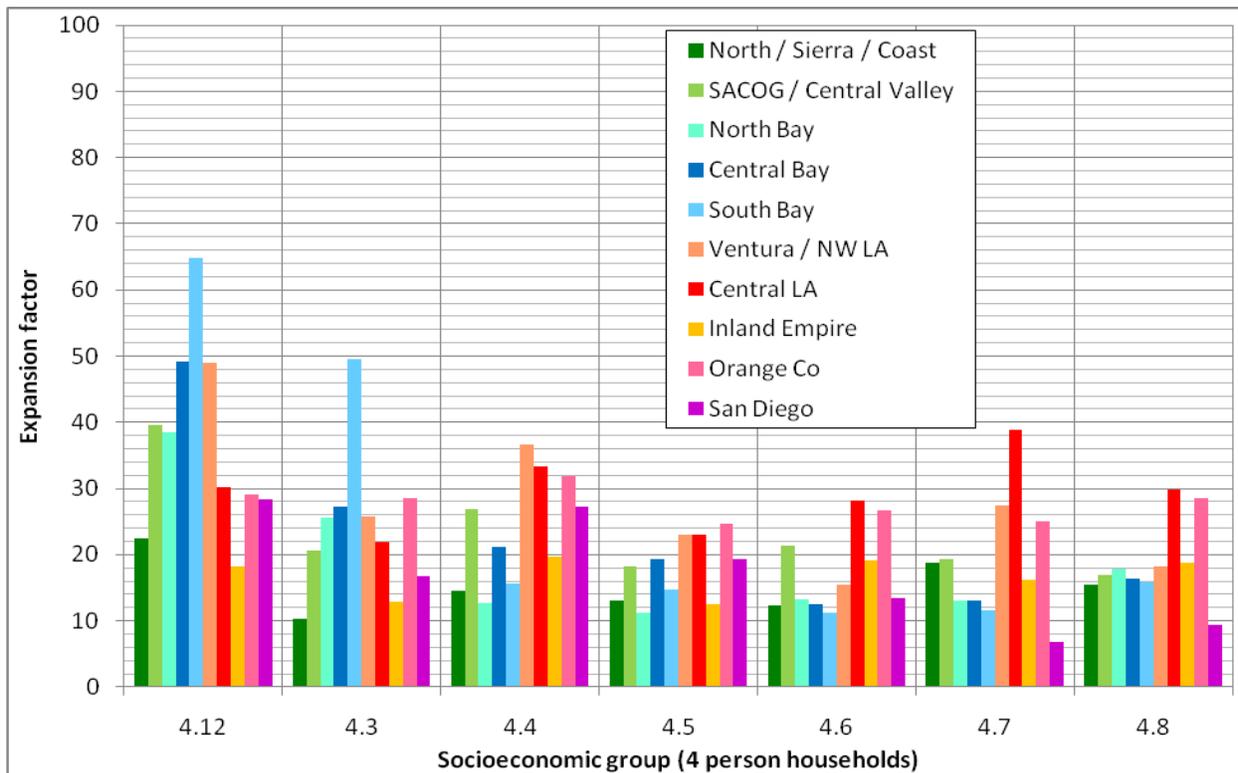


Figure 14: Expansion Factors by Socioeconomic Group (4 person households)

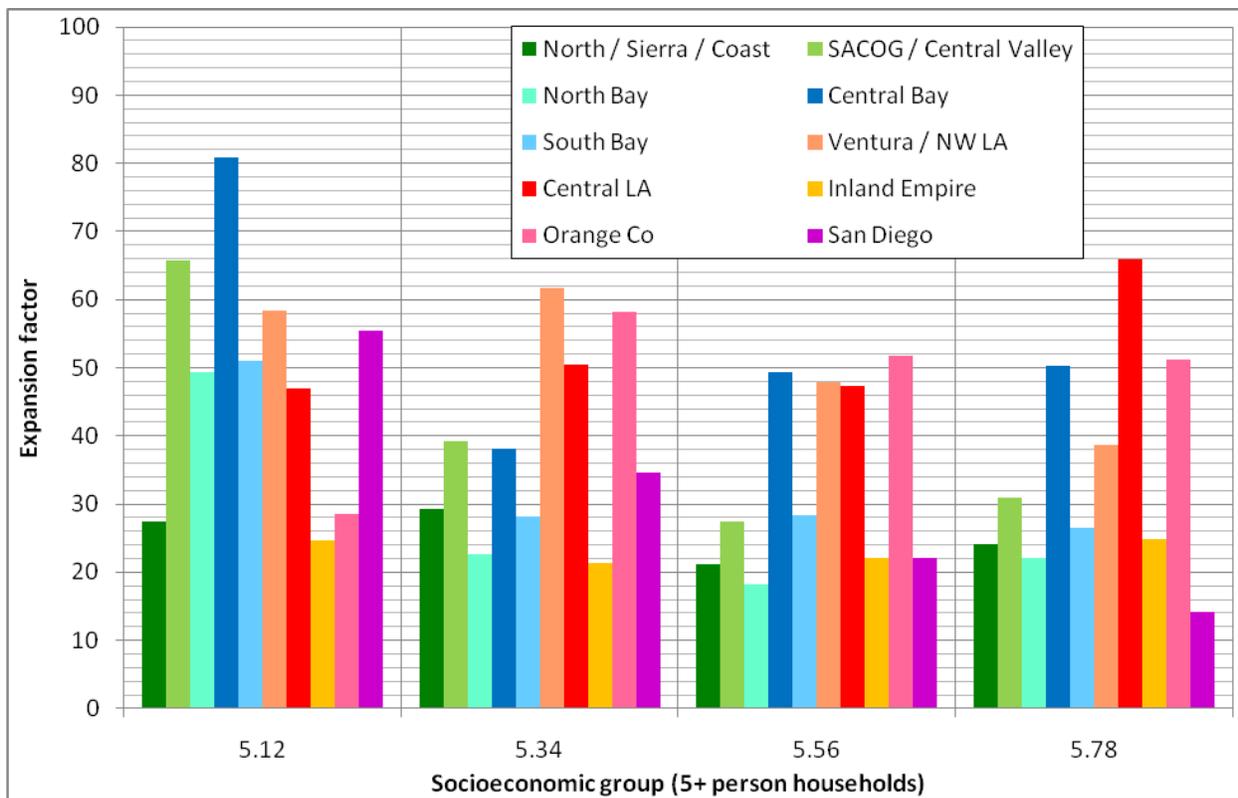


Figure 15: Expansion Factors by Socioeconomic Group (5+ person households)

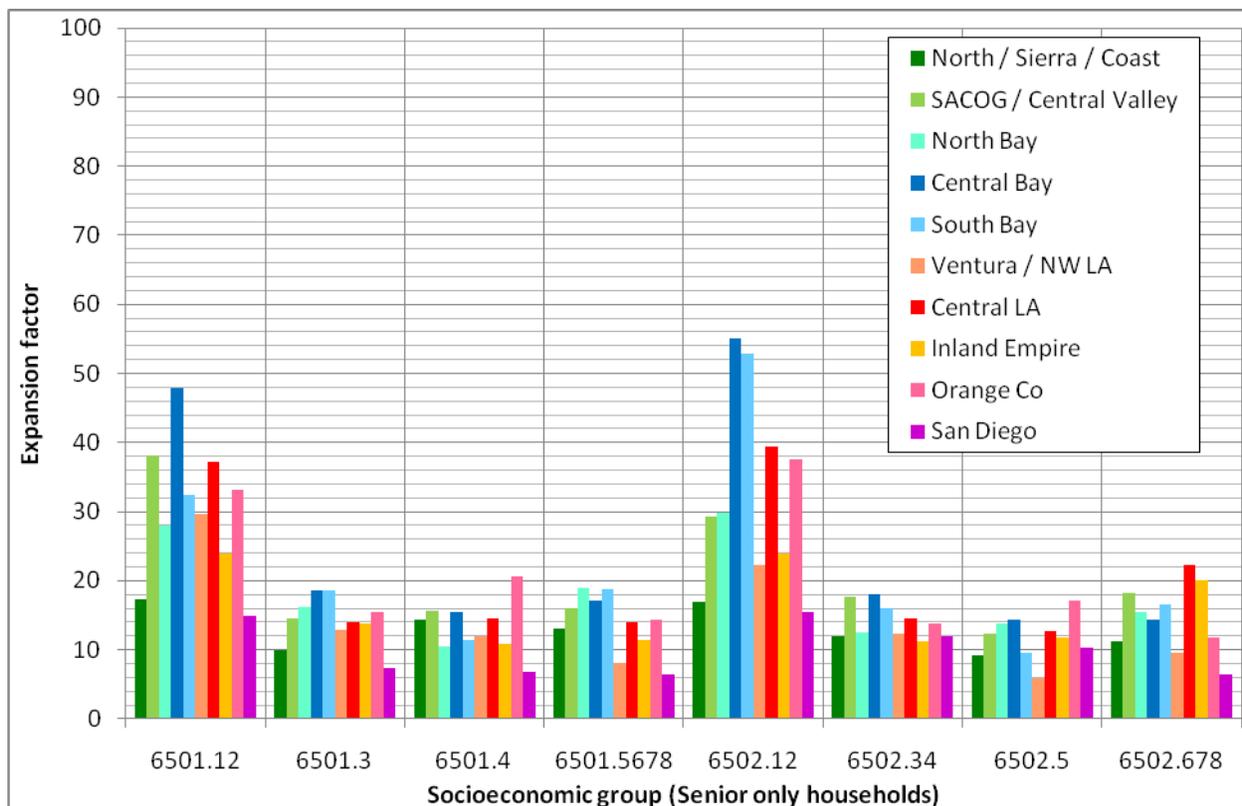


Figure 16: Expansion Factors by Socioeconomic Group (senior only households)

In general, it can be seen that the highest expansion factors tend to be for low-income households. Larger households (4 and 5+ persons) tend to have larger expansion factors across all income levels, as compared with the smaller households. The household type expansion factors range from a minimum of 5.84 to a maximum of 95.03, with a median of 16.94 and an average of 21.44. Overall, 90.8% of the expansion factors were between 8 and 50, and 72.0% were between 10 and 30.

3.2 Geographic Expansion Factors

The geographic expansion factors have a much larger range than the socioeconomic factors; this is primarily because the surveys have different sampling rates in different areas; for instance, the statewide survey sought around 500 households for each of Shasta, Sacramento and Los Angeles counties – although Shasta has 63K households, Sacramento 450K and Los Angeles 3.1 million. Los Angeles had additional records from the SCAG survey, but Sacramento did not.

The expansion factors range from 1.99 (in Imperial county) to 98.59 (in Sacramento), with a median of 16.68 and an average of 20.44. Overall, 92.7% of the expansion factors were between 5 and 50, and 63.6% between 10 and 25. These expansion factors are shown in the three maps below.

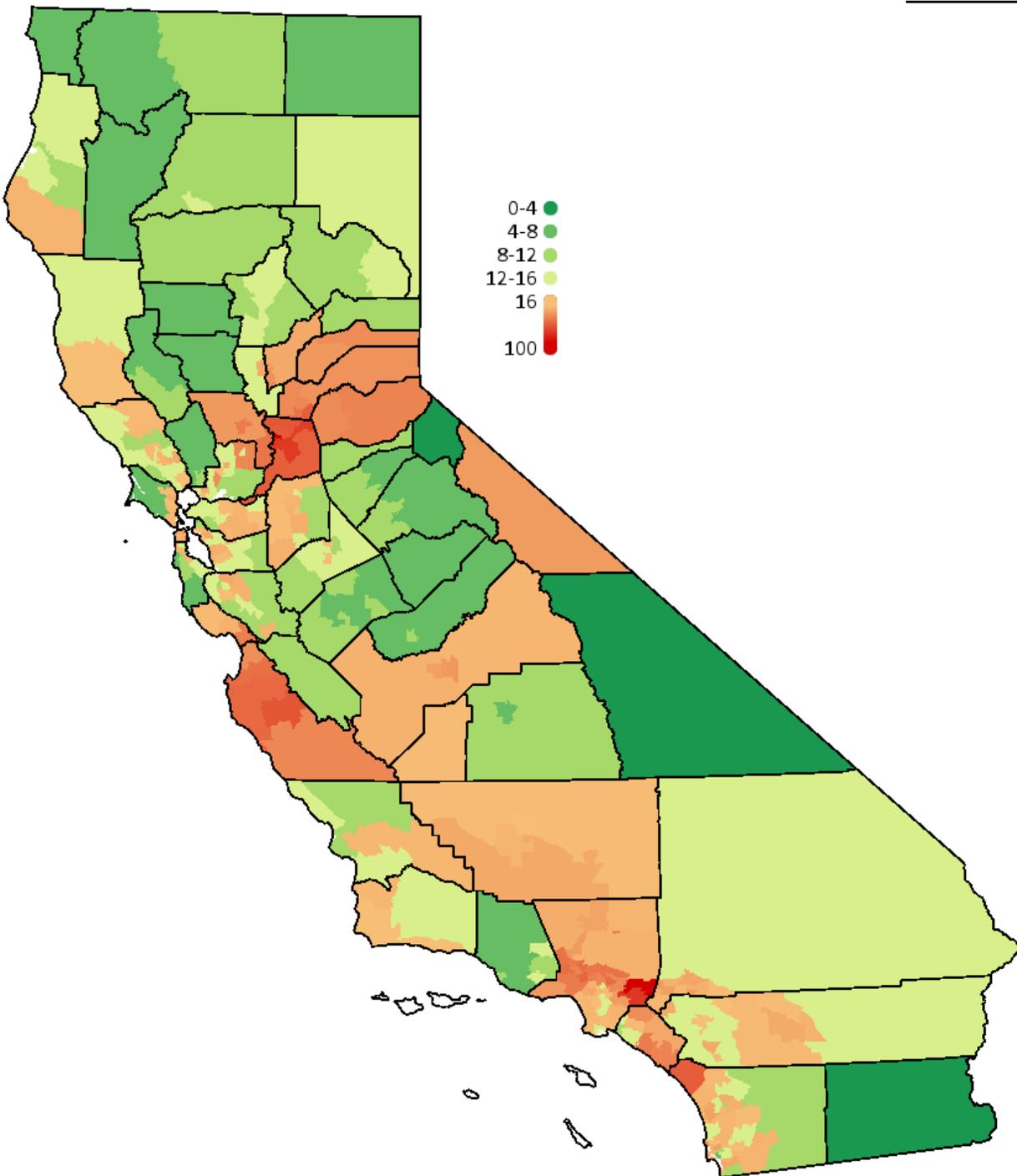


Figure 17: Expansion Factors by Geographic Region

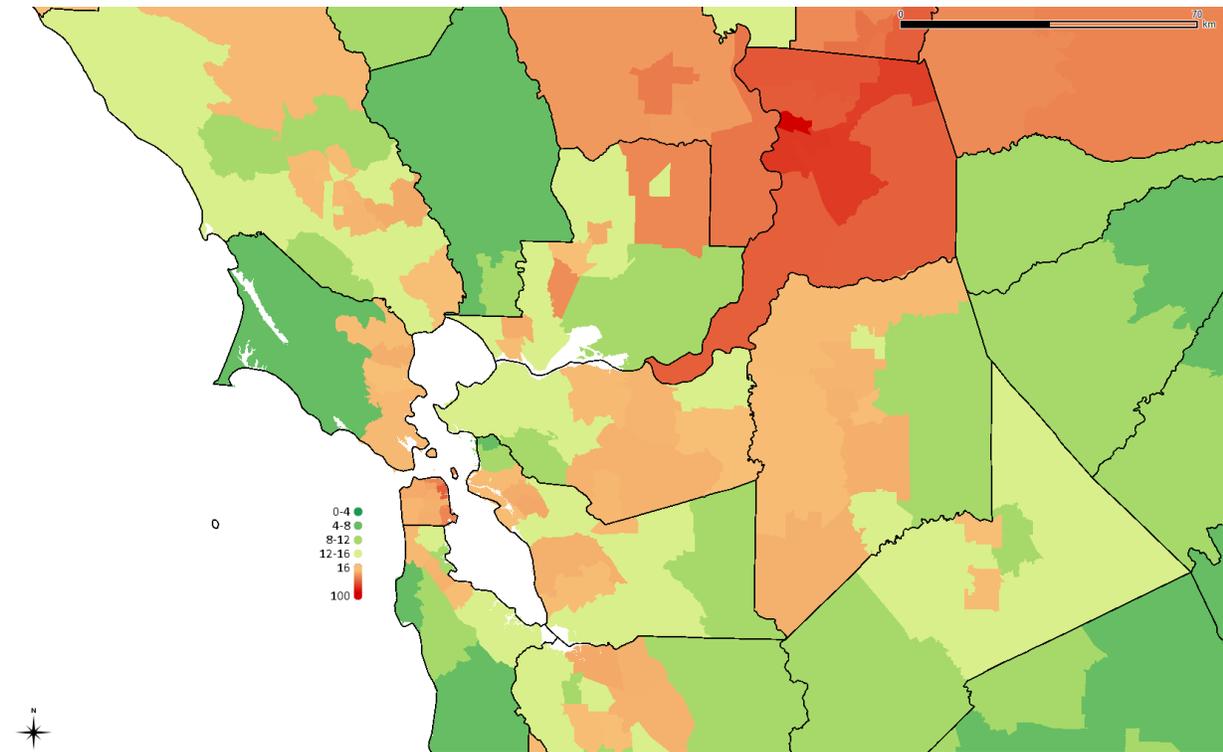


Figure 18: Expansion Factors in the Bay Area

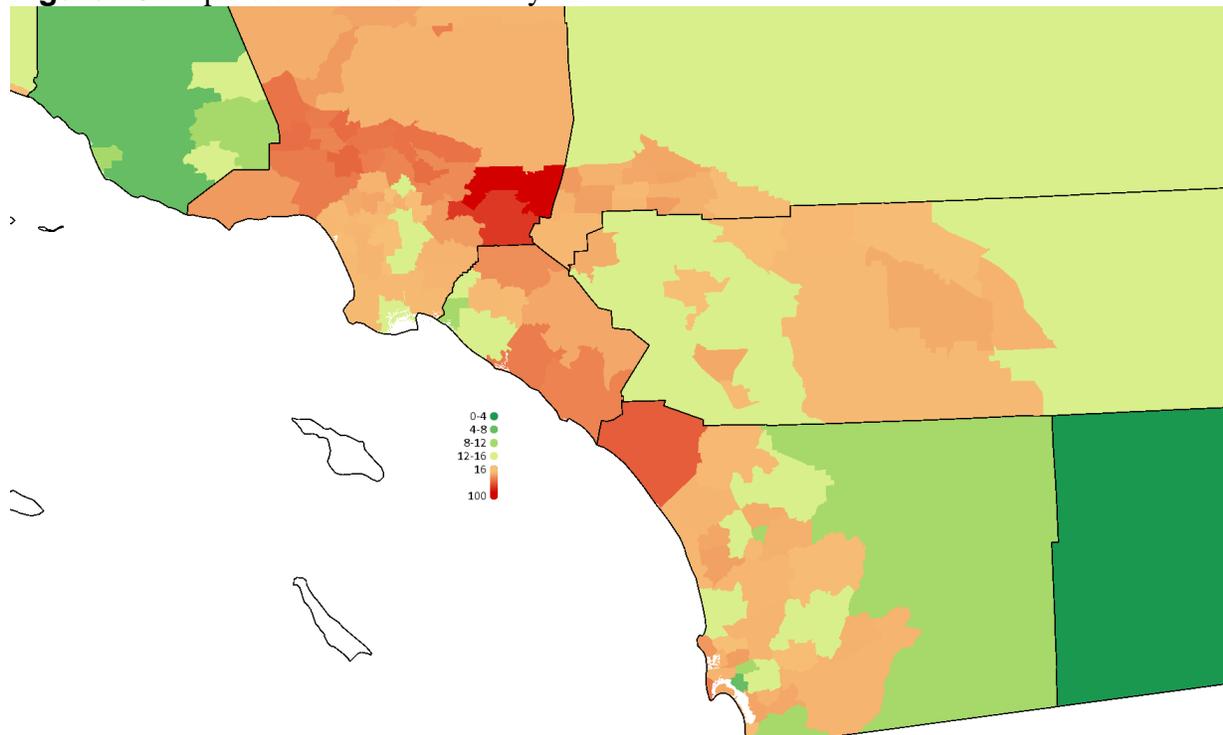


Figure 19: Expansion Factors in the Los Angeles Region

3.3 Expansion Factors

There are two sets of expansion factors as a result of the household expansion; a socioeconomic and a geographic factor. The ranges of these factors are shown in the graph below.

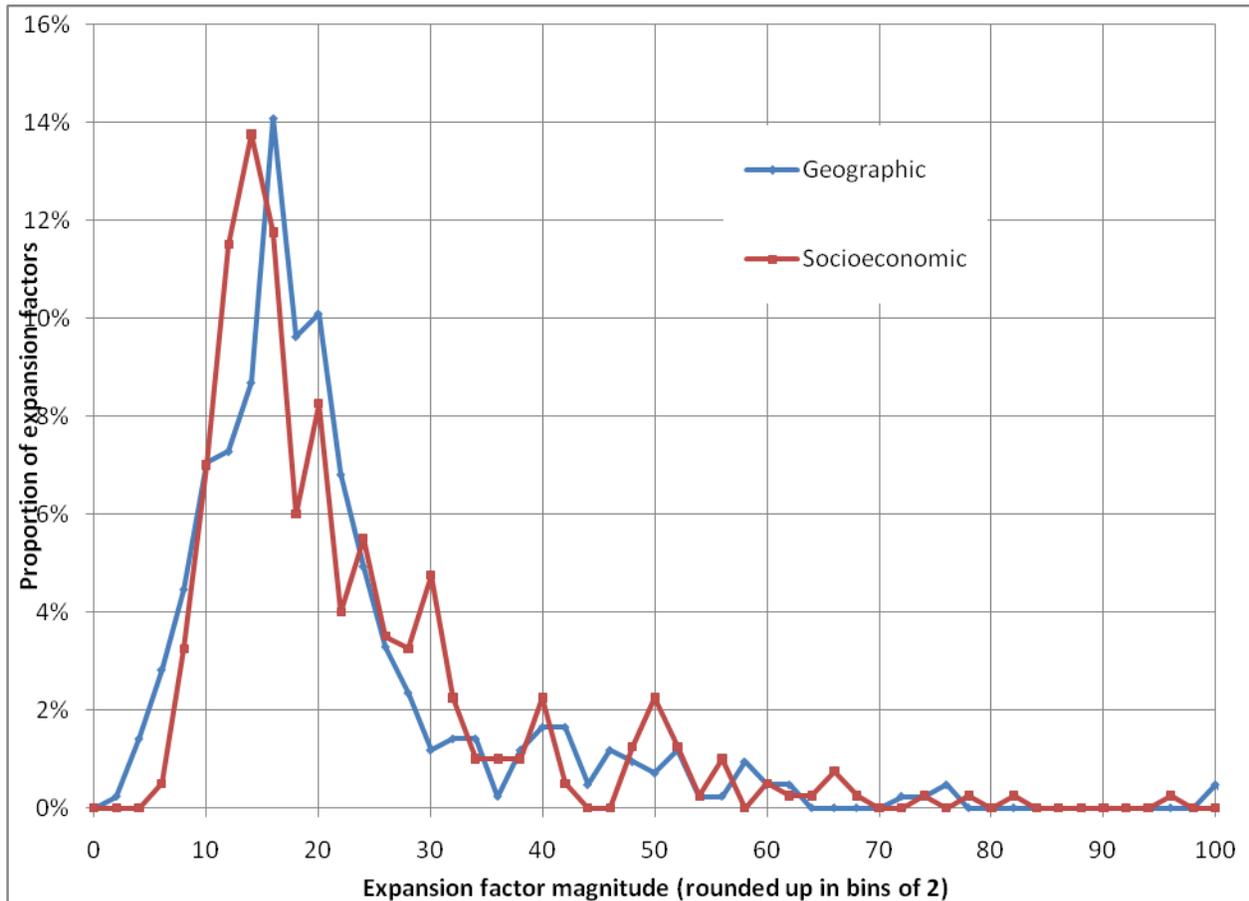


Figure 20: Ranges of Expansion Factors

The two sets of expansion factors described above are applied to each household in the survey:

$$\text{combined expansion factor} = \text{socioeconomic} \times \text{geographic}$$

These combined expansion factors exist for each household. The theoretical range for these is 11.6 (the lowest socioeconomic and lowest geographic) to 9368.7 (the product of the highest socioeconomic and geographic factors). In the actual set of survey households, the range was from 14.6 to 6494.9. The median combined expansion factor

is 229.8, and the average is 322.3. The average is particularly skewed by a number of very large expansion factors representing very large and very poor households in the Sacramento and San Gabriel Valley areas. 90.3% of surveyed households have expansion factors between 75 and 1000, and 66.0% between 100 and 400. The distribution of the expansion factors is shown in the graph below.

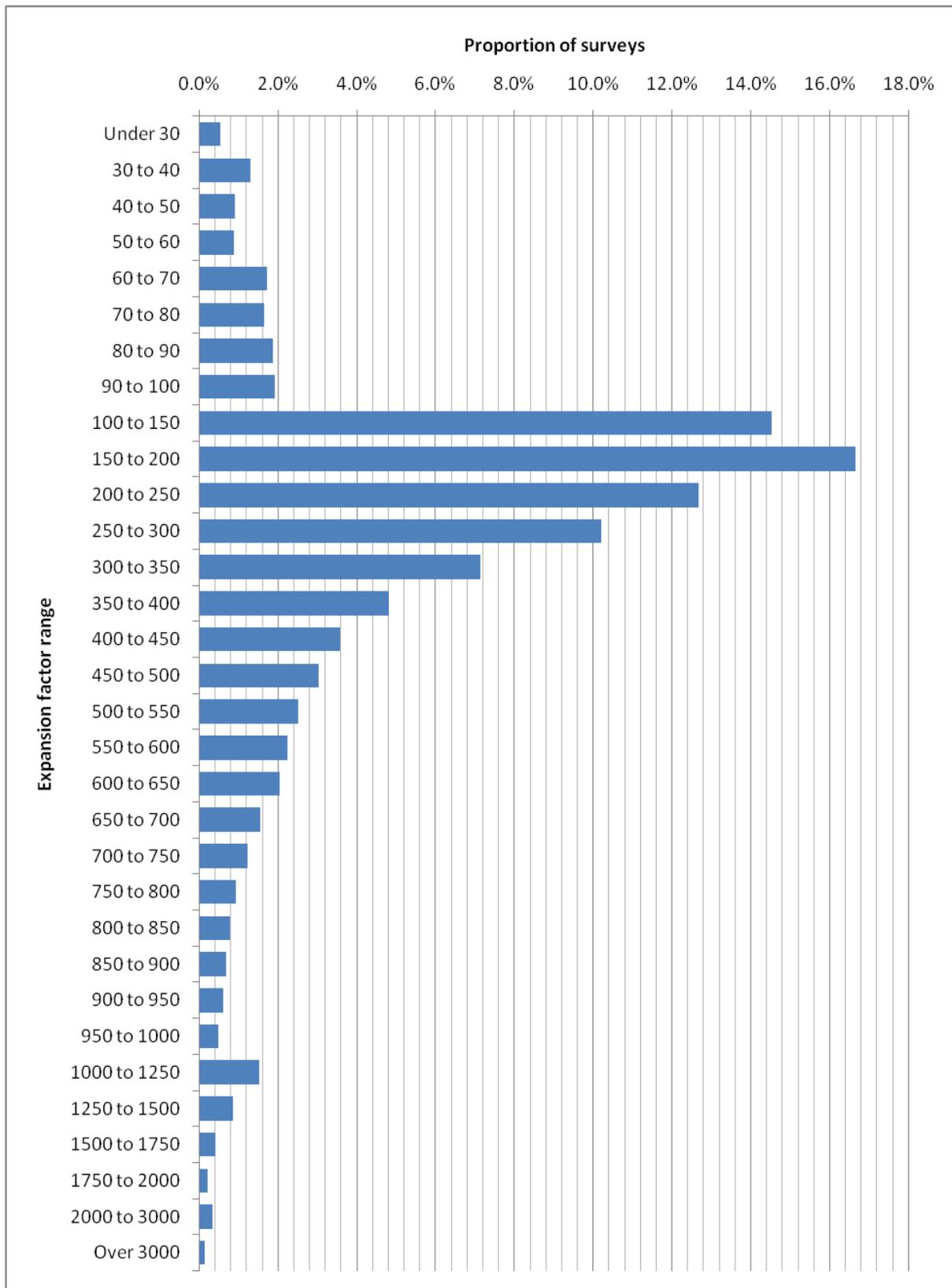


Figure 21: Distribution of Expansion Factors