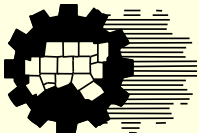# Data Fusion for Market Segmentation

Applying Census Data for Transportation: 50 Years of Transportation Planning Data Progress Workshop

Kansas City, MO
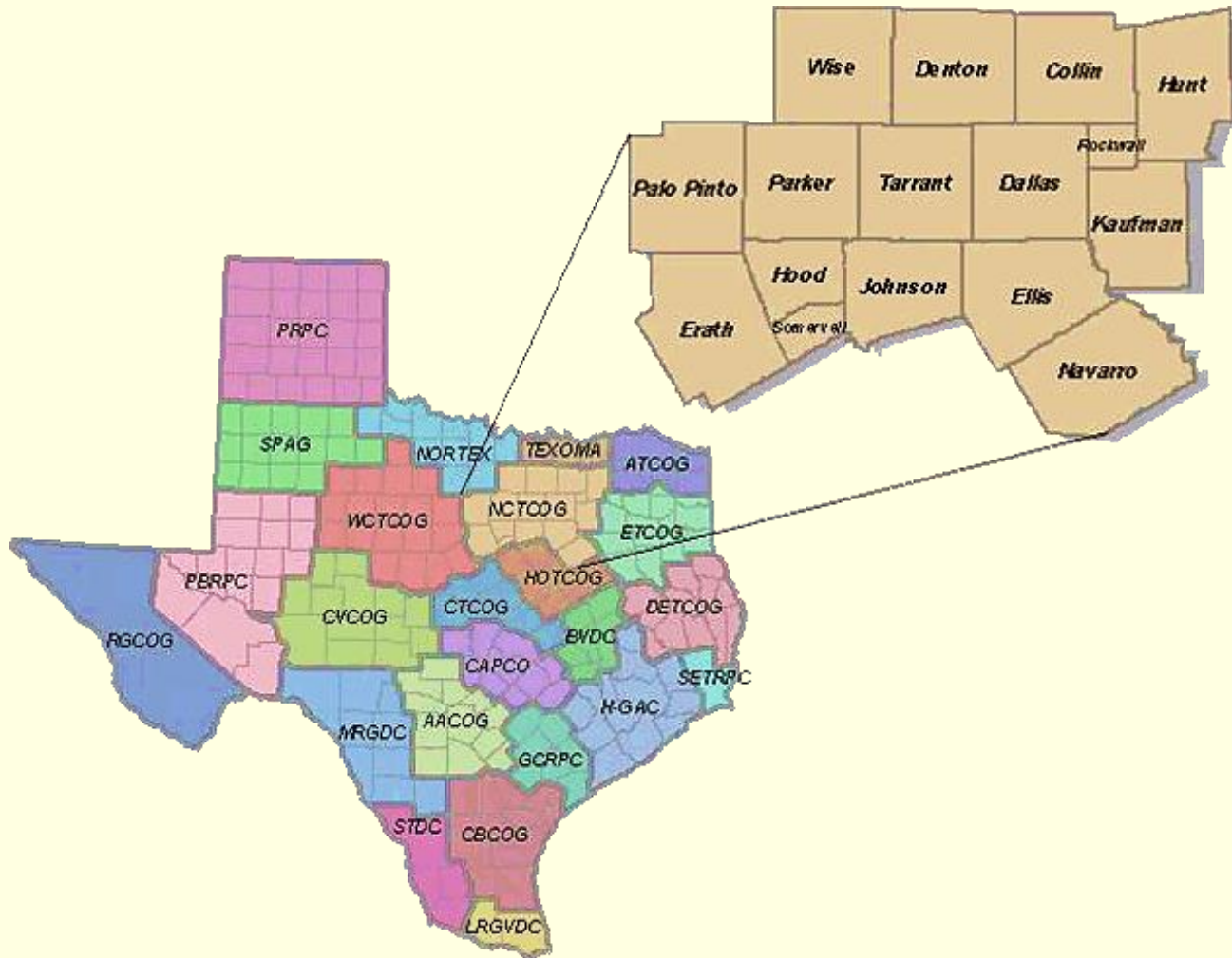
Nov 15, 2017

Arash Mirzaei, Kathleen Yu, Liang Zhou

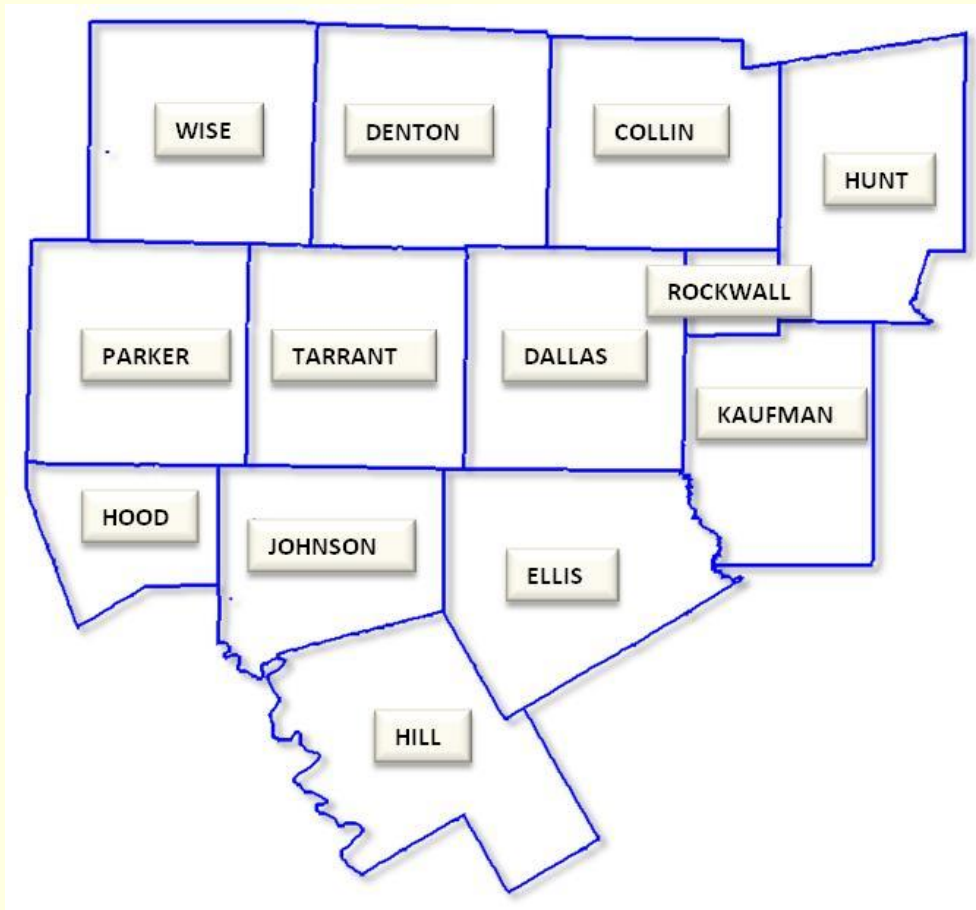North Central Texas Council of Governments

Outline

- Background
  - NCTCOG Region
  - CTPP Geography
- Use of CTPP in Market Segmentation
  - Market Segmentation of Households
  - Market Segmentation of Employment
- Final Thoughts

North Central Texas Council of Governments (NCTCOG)

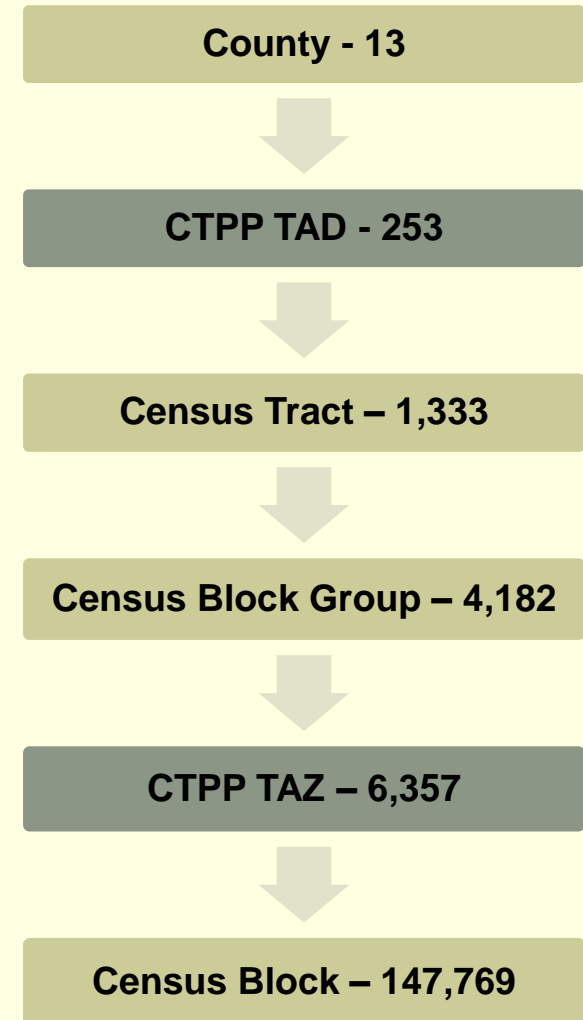# NCTCOG Modeling Area

Modeling Area:
12 County + Hill

Area:
10,000 square miles

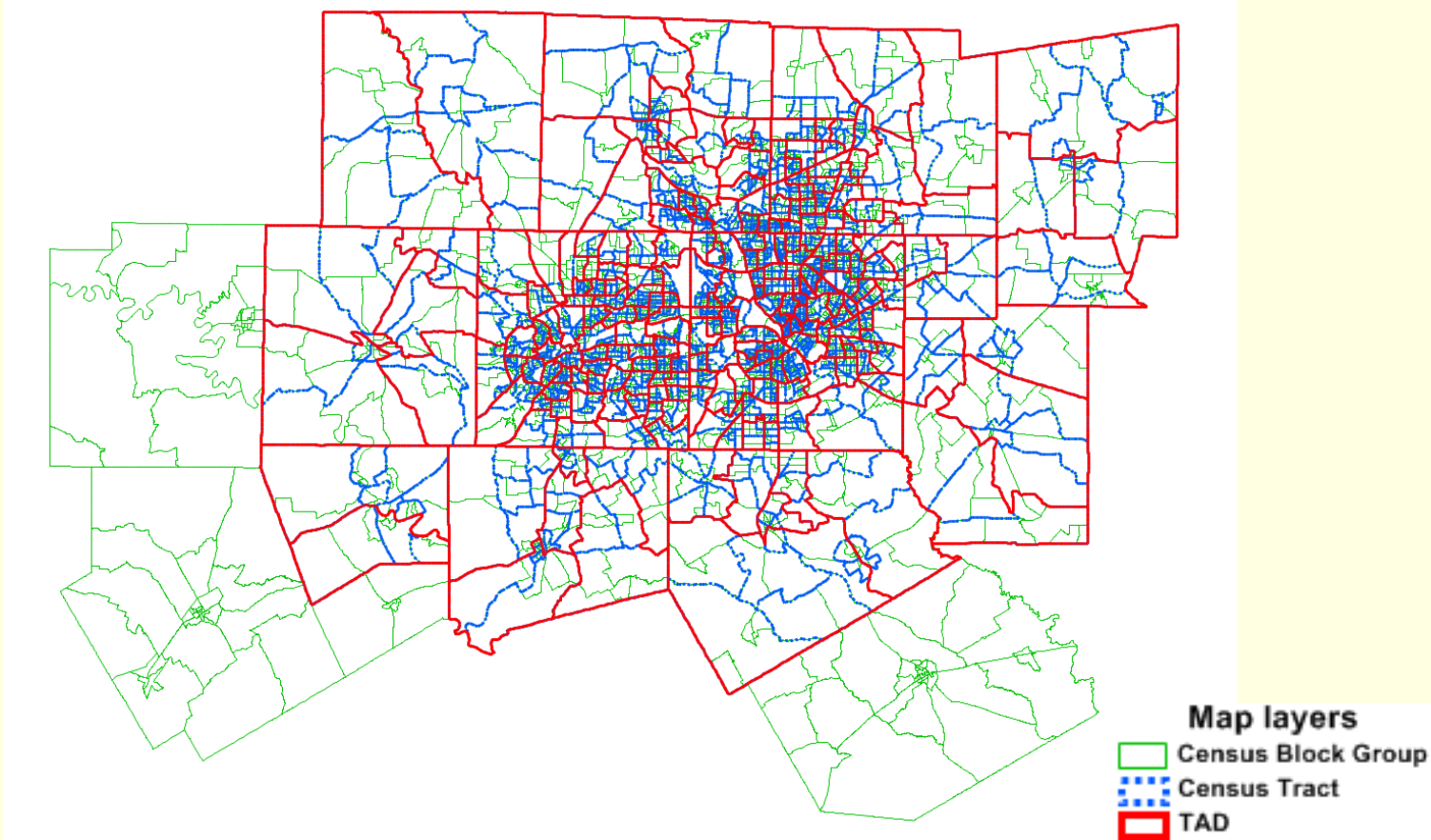**2014 Households:**
2,466,263

2014 Household Population:
6,860,993

CTPP Geography Hierarchy

- CTPP geographies are Travel Analysis Districts (TAD) and Travel Analysis Zones (TAZ).
- ACS Geographies include County, Tract, Block Group, and Block.
- NCTCOG defined the TADs based on census tracts. TADS nest in counties.
- NCTCOG defined the TAZs to nest in block groups.

County - 13

↓

CTPP TAD - 253

↓

Census Tract – 1,333

↓

Census Block Group – 4,182

↓

CTPP TAZ – 6,357

↓

Census Block – 147,769

CTPP Geography Hierarchy

Map layers
- Census Block Group
- Census Tract
- TAD

USE OF CTPP IN HOUSEHOLD MARKET SEGMENTATION

HH Market Segmentation Goal

- NCTCOG investigated various household characteristics to determine the best market segmentation for home-based trip purposes.

- For Home-Based Work, it was the combination of
    - Number of Workers in Household
    - Vehicles Available
    - Household Income

- As a result, our goal was to get the distribution of households for Number of Workers in Household by Vehicles Available by Household Income for each block group.

HH Market Segmentation Data

- The data needed for our goal was not available directly, so need to combine various data sources.

- Creating the market segments
  - Iterative Proportional Fitting
  - Fusion of data sources
    - CTPP 2006-2010 residence-based data for seed
    - ACS 2014 5-year for segmentation totals

Row Control Total

Step 1a: Row control targets from ACS 2014 5-year, B08203,Number of Workers (4) x Number of Vehicles (4)

- Census Tract Geography
- Block groups inherit the distribution from the census tract

	Inc1	Inc2	Inc3	Inc4	Row Control Targets
W0V0					
W0V1					
…					
W3V3					
Col Control Targets					

Column Control Total

Step 1b: Column control targets from ACS 2014 5-year, B19001, Household Income (4) in Past 12 months

	Inc1	Inc2	Inc3	Inc4	Row Control Targets
W0V0					
W0V1					
…					
W3V3					
Col Control Targets					

Seed

Step 2: Seed matrix from Census Transportation Planning Products (CTPP) 2006-2010, A112310, Number of Workers (4) x Vehicles (4) x Household Income (4)

 - TAD Geography

 - Block groups inherit the same distribution from TADs

	Inc1	Inc2	Inc3	Inc4	Row Control Targets
W0V0					
W0V1					
…					
W3V3					
Col Control Targets					

Iterative Proportional Fitting

Step 3: IPF – Use the seed to match the control totals

	Inc1	Inc2	Inc3	Inc4	Row Control Targets
W0V0	x	x	x	x	x
W0V1	x	x	x	x	x
…	x	x	x	x	x
W3V3	x	x	x	x	x
Col Control Targets	x	x	x	x	

USE OF CTPP IN EMPLOYMENT MARKET SEGMENTATION

Employment Segmentation Goal

In Employment Segmentation, we are trying to understand the locations of jobs for various segments of workers.

Goal – To get a distribution of the workers at the workplace by their household income by industry group for each TAZ.

Data Sources Used:

- CTPP 2006-2010 - Workplace-based tables which give workers broken down by their workplace.
- Bureau of Economic Analysis (BEA) 2013
- 2010 Longitudinal Employer Household Dynamics (LEHD)

Employment Segmentation Process

Step 1: Identify the row and column control targets from Census Transportation Planning Products (CTPP) and other data sources.

- Row Marginal: Industry (3)
  - Use 2013 BEA county data to get employment by industry group.
  - Convert county data to TAZ using 2010 LEHD, modified by local data.

- Column Marginal: Income (4)
  - A203101 - Household Income in the past 12 months (2010$)(4) - TAD from Part 2 CTPP
  - TAZ inherits its distribution from the TAD to which it belongs, but is scaled by the number of HH in the TAZ.

Employment Segmentation Process

Step 2: Find an initial seed matrix from Census Transportation Planning Products (CTPP) –Industry (3) x Income (4)

- Use CTPP Workplace-based tables which give workers broken down by their workplace by TAD.
  - A202205C - Industry (3) by Earnings in the past 12 months (2010$)(4)
  - A203202C - Household income in the past 12 months (2010$)(4) by Earnings in the past 12 months (2010$) (4)

- These two tables are combined to produce a table of Industry(3) x Income (4) for each TAD. For each earning category, that the industry distribution will be the same for all income groups.

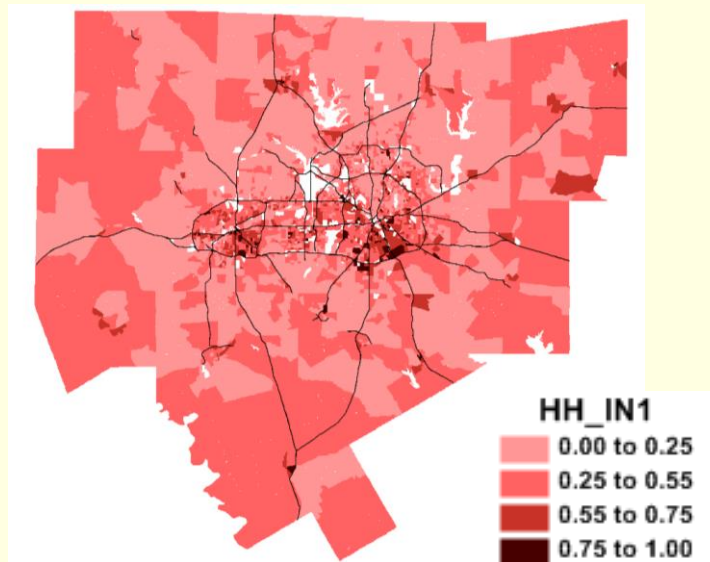- TAZ inherits its distribution from the TAD to which it belongs.

Step 3: IPF

Connecting Home to Work

Using the segmentation of households and workers, we can connect households of low income households to low income employment.
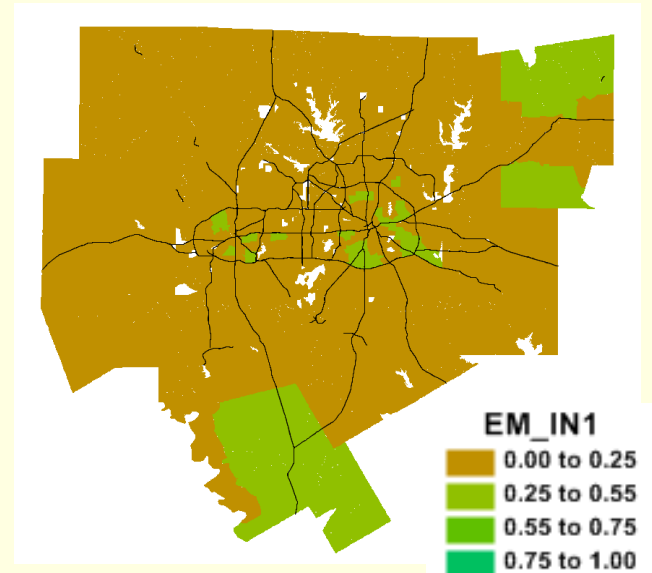

500 Households

500 Income 1 Households

500 Employment

500 Income 1 Employment
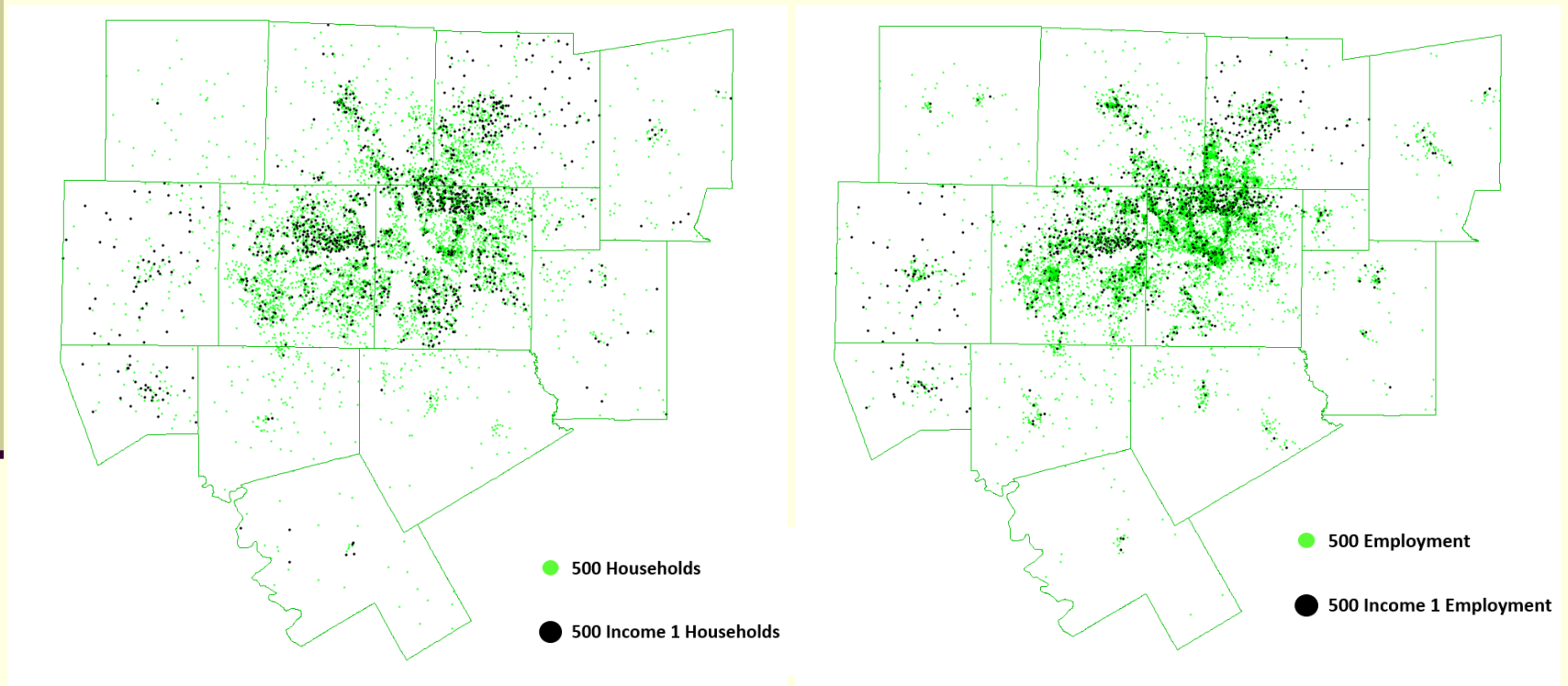
Connecting Home to Work

Location of

Low Income
Households

Location of

Low Income
Employment

HH_IN1
- 0.00 to 0.25
- 0.25 to 0.55
- 0.55 to 0.75
- 0.75 to 1.00

EM_IN1
- 0.00 to 0.25
- 0.25 to 0.55
- 0.55 to 0.75
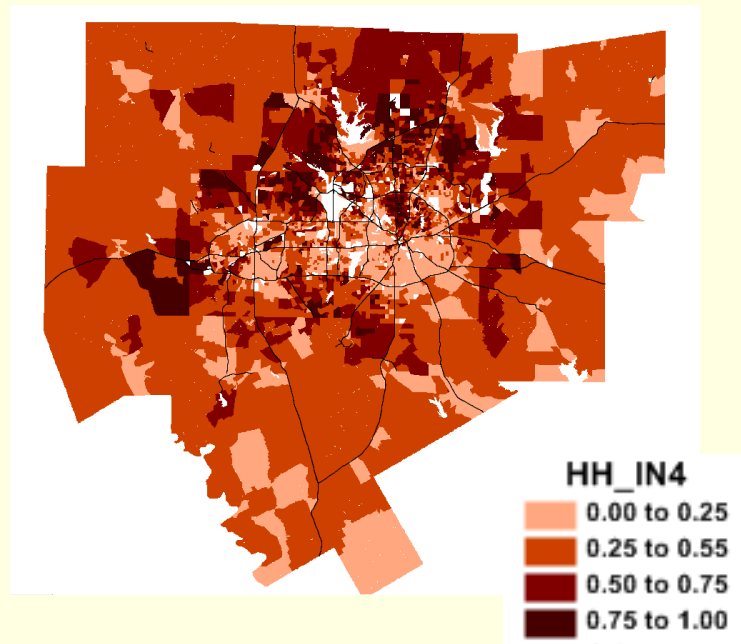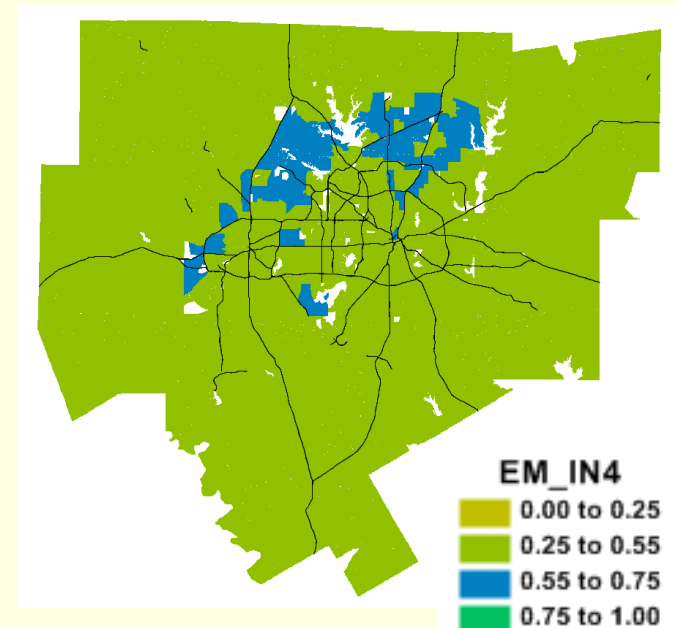- 0.75 to 1.00

Connecting Home to Work
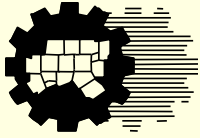
Location of

High Income Households

Location of

High Income Employment

Final Thoughts

- CTPP could use a table of industry by income for workplace-based tables.

- In CTPP Data, for purposes of stability and the reduction of sampling error, NCTCOG used the TAD level data instead of the TAZ level data. We could try to use any tables available at the Census Tract level.

Contact Information

- Arash Mirzaei   amirzaei@nctcog.org
- Kathy Yu   kyu@nctcog.org
- Liang Zhou   lzhou@nctcog.org