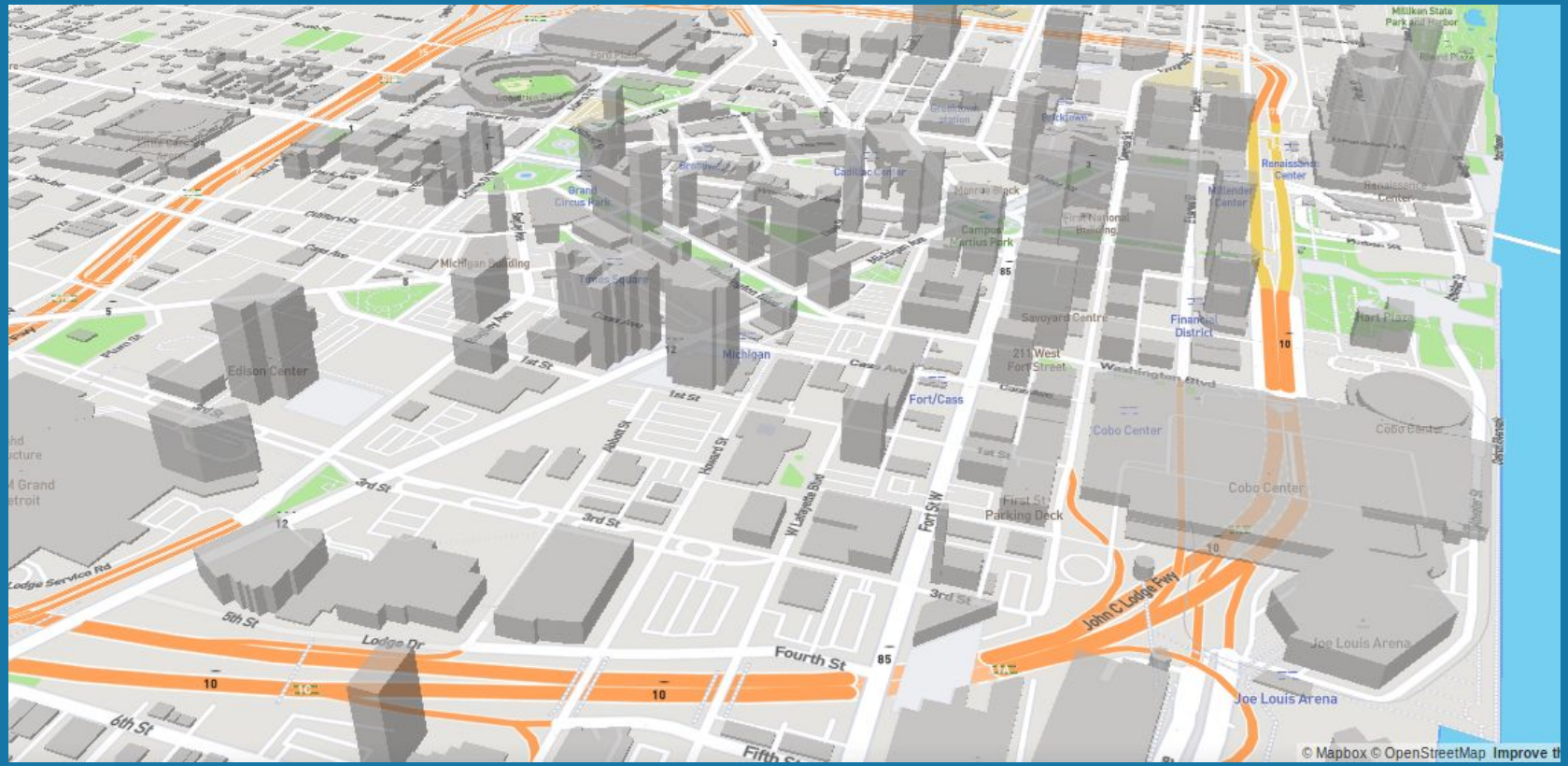
SEMCOG-UrbanSim Bi-weekly



July 11, 2019

Areas of work

1. Iterating on regionally-estimated specifications, conducting runs, and comparing loss scores / case-study indicators with runs 4036
    a. Diagnosing Households and Employment distribution Issues on run 4074
    b. Setting up Template-based model estimations and Runs
    c. Estimating HLCMs without unplaced households and demolished/redeveloped buildings
2. Setting up calibration script to calibrate each location choice submodel, with the official forecast as the target

Other work related to project: Pandana performance diagnosing (colleagues Federico and Sam M.)

- https://github.com/UDST/pandana/tree/enhancement/optimize-string-comparisons
- Working on ***v0.4.2*** release with optimizations

Comparison

	SPECIFICATIONS			
	REGIONAL	LARGE AREA CONTROL	REGIONAL	LARGE AREA CONTROL
zones	3,522,233	947,148	1,523,257	1,167,831
large_areas	36,552	1,294	122,736	199
cities	495,160	472,554	547,454	261,910
counties	40,603	1,407	137,888	217
	3 vs 4 (Regional vs large-area controlled models)		*6 vs 7 (Regional vs large-area controlled models)*	

- LARGE-AREA CONTROLLED MODELS SHOW DECREASE OF MSE IN ALL GEOGRAPHIES
- The most relevant improvements are shown in the Large-area and the county-level

Comparison

	SAMPLING-UNIT					
	BUILDING-LEVEL	UNIT-LEVEL	BUILDING-LEVEL	UNIT-LEVEL	BUILDING-LEVEL	UNIT-LEVEL
zones	1,167,831	555,068	3,522,233	1,317,909	3,522,233	496,055
large_areas	199	201	36,552	32,943	36,552	34,152
cities	261,910	296,566	495,160	495,681	495,160	500,153
counties	217	218	40,603	36,500	40,603	37,746
	7 vs 7B (building-level vs unit-level models)		*3 vs 3B (building-level vs unit-level models)*		*3 vs 3C (building-level vs unit-level models)*	

- The biggest improvements using Unit-level LCMS are shown in the zone-level
- The employment location choice model seem to have a better performance when estimating at a building level in all geographies except zones.
- MSE in cities increase with Unit-Level models , specifically with ELCM unit-level.
- At the city level, variables such as: hh size > 4 p, hh age > 65 and jobs in sector 8, 9, 14 and 16 have a higher MSE when estimating with unit-level

RUN 4074

CLIENT RUN 4074 SEMCOG AS IS WITH LARGE AREA CONTROL and UNIT LEVEL HLCMS AND ELCMS - CLUSTERING VARS	Mean MSE
zones	271,060
large_areas	70
cities	57,055
counties	77

1 - RUN 4036

	Overall MSE
zones	1,612,099
large_areas	68
cities	140,845
counties	75

%: MSE(4074) - MSE(4036) / MSE(4036)

% Improvement in Run 4074

- Unit-level Improvement on a zone and city-level. But, poor indicators in pop age and hh size =3.
- Large-Areas con counties indicators does not change or benefit from unit-level sampling.
- Very poor performance on residential units count in all geographic levels in run 4074 compared to run 4036.
- Run 4074 shows poor performance in job spaces for large-areas and counties

RUN 4074

CLIENT RUN 4074 SEMCOG AS IS WITH LARGE AREA CONTROL and UNIT LEVEL HLCMS AND ELCMS - CLUSTERING VARS	
	Mean MSE
zones	271,060
large_areas	70
cities	57,055
counties	77

Large-Area Price Models

	Run 7 Building-Level	*Run 7 B Unit-Level*
	BUILDING-LEVEL	UNIT-LEVEL
zones	1,167,831	555,068
large_areas	199	201
cities	261,910	296,566
counties	217	218

Regional Price Models

- Unit-level Improvement on a zone. However, poor indicators in pop age and hh size =3.
- For cities Unit-level sampling shows poor indicator in job sectors (again!)
- Large-Area price models show improvement over Regional price models

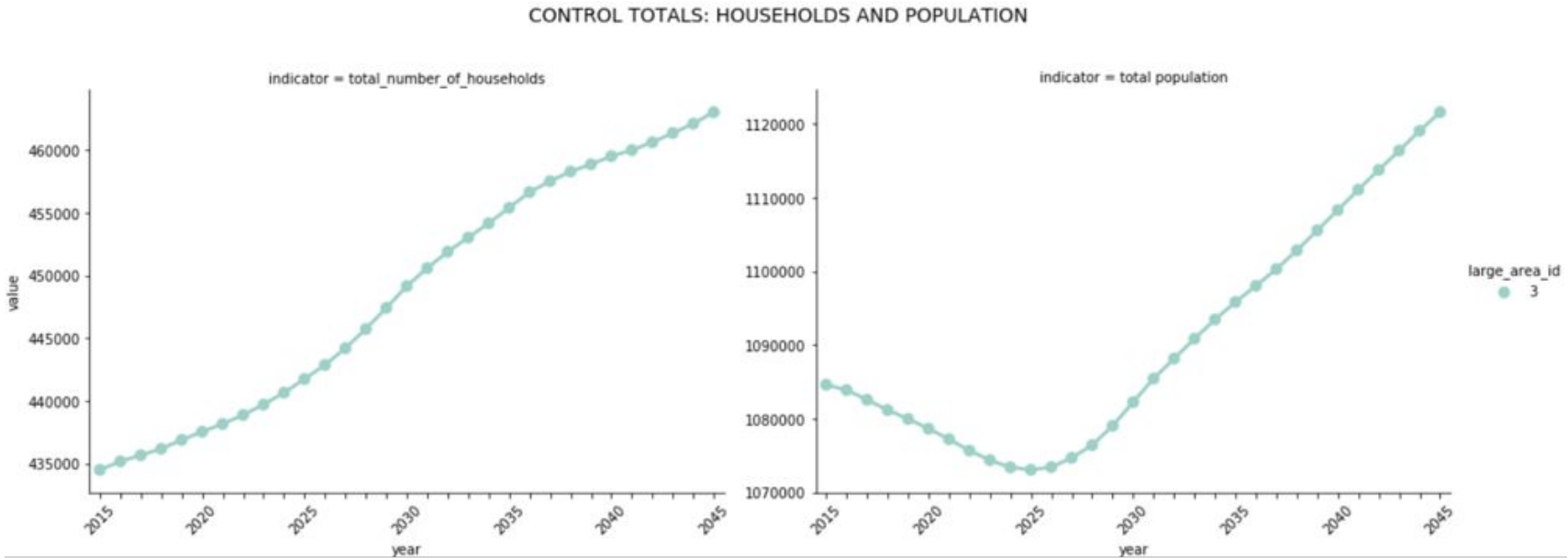
RUN 4074: HLCM Diagnosis

- Dearborn (cityid=1025)(large_Area_id = 3), a community with large HHsize and large immigrant population is losing population substantially. The situation is worse than SEMCOG asis scenario. Same is the case with many older but stable communities – Troy (2270), Dearborn Heights (1030). Communities like West Bloomfield (2285), Southfield (2250) are developing at much slower pace than they should. All these are very desirable communities with good access to amenities, high property values.
- On the flip side, growing communities, ones that are on the fringes of these stable communities and are now are considered suburban communities, saw good growth in last 5 year, are growing at a much faster rate compared to the asis scenario. eg. Macomb Twp (3060), Lyon Twp (2145), Canton (1020), Van Buren Twp (1200) Huron Twp (1105).
- From the initial glance through, it appears that the model is moving people from older communities to these growing communities.

RUN 4074: HLCM Diagnosis

- Transition Models: Annual Control Totals: Households and Population
- Relocation Model: relocation probabilities by income and age of head.
- Distribution of Data in the base year
- Hlcm variables

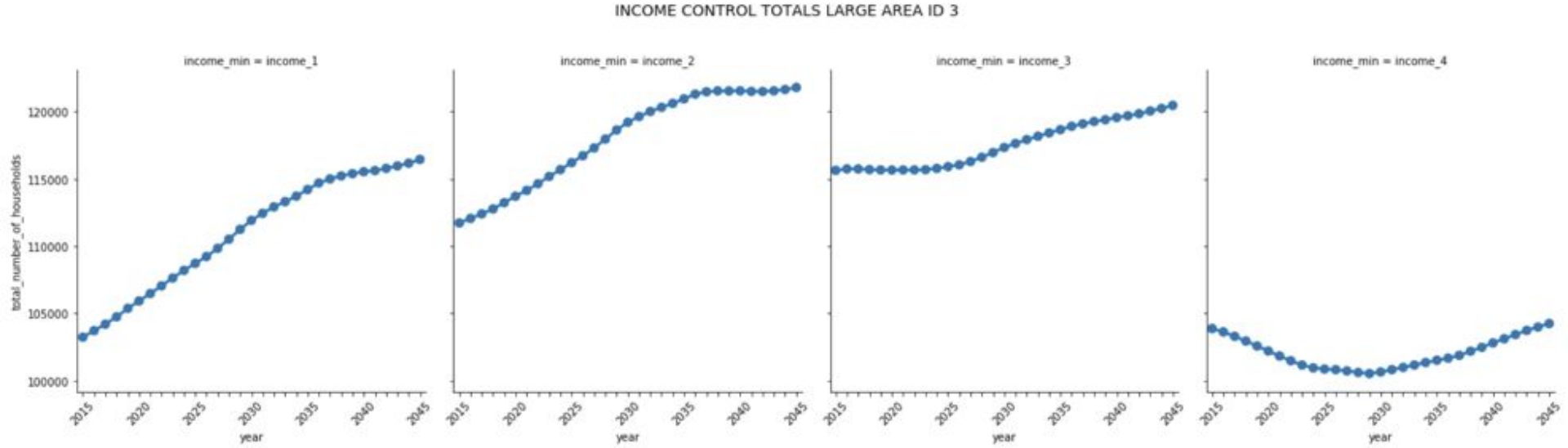
RUN 4074: Households Diagnosis



LARGE AREA 3

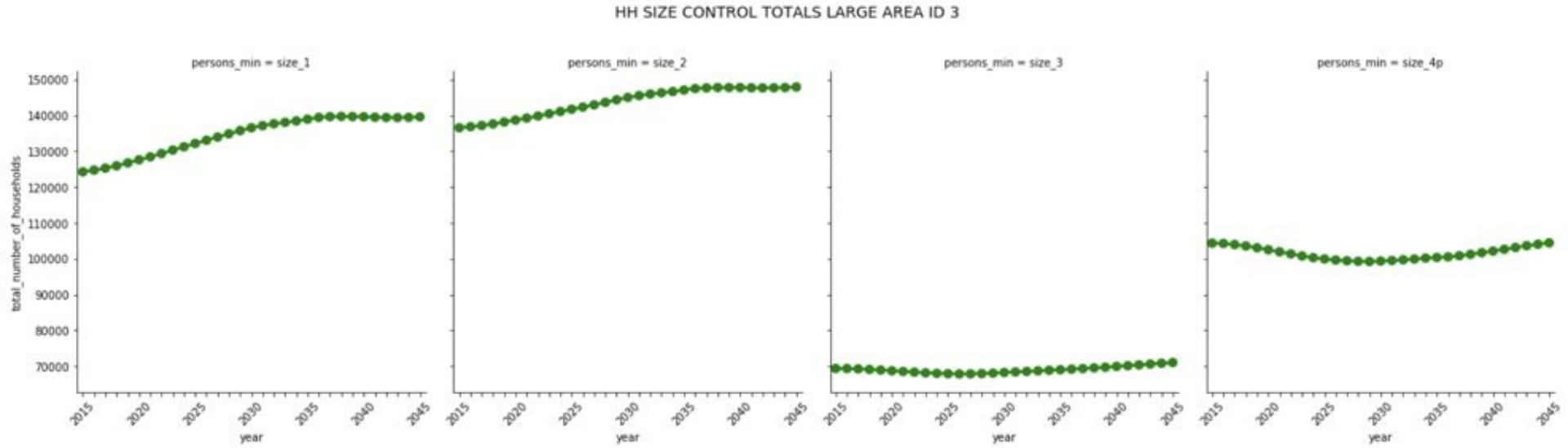
indicator	year perc_growth_2045
total households	11.366618
total population	6.802737

RUN 4074: Households Diagnosis



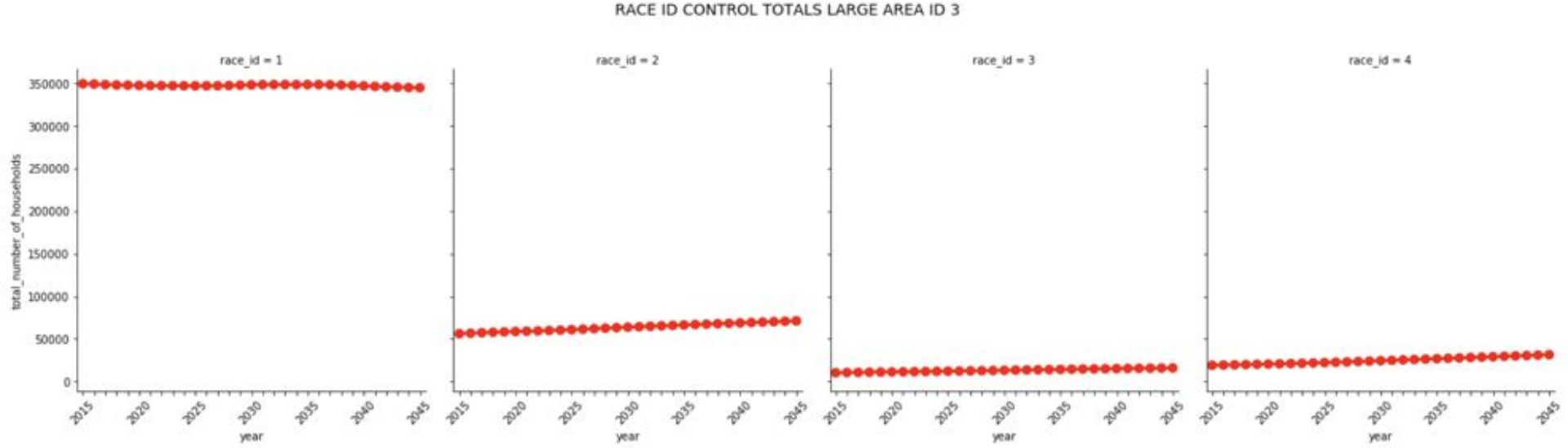
		perc_growth_2045
	income_min	
total_number_of_households	income_1	12.802332
	income_2	9.013210
	income_3	4.178268
	income_4	0.360997

RUN 4074: HLCM Diagnosis



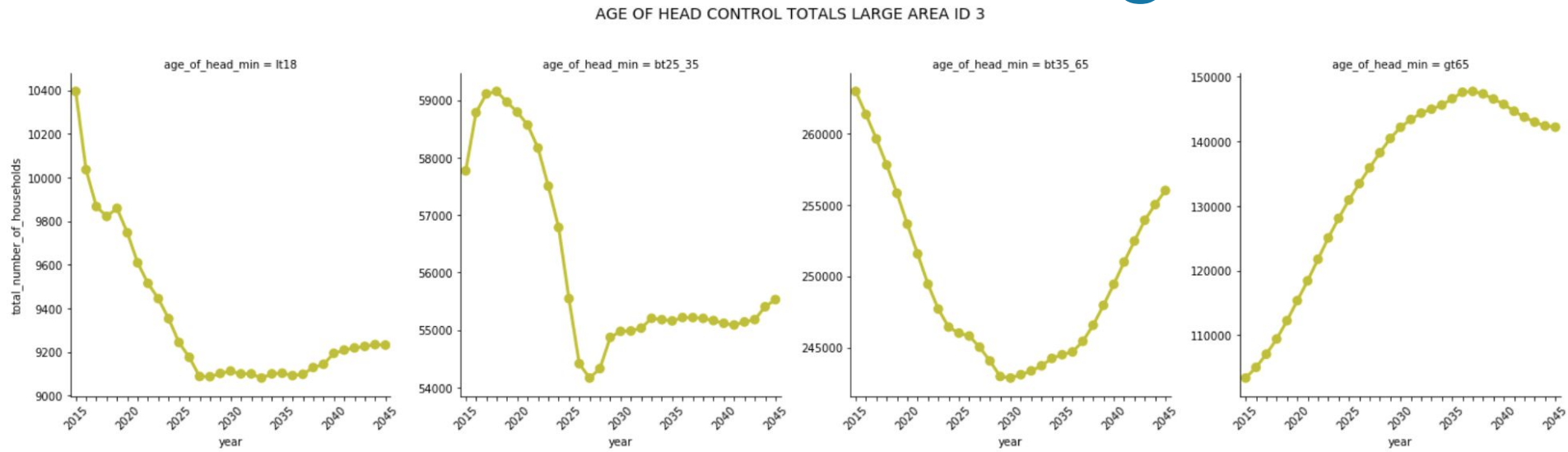
		perc_growth_2045
	persons_min	
total_number_of_households	size_1	12.354830
	size_2	8.334554
	size_3	2.399077
	size_4p	0.094885

RUN 4074: Households Diagnosis



	race_id	perc_growth_2045
total_number_of_households	1	-1.322635
	2	26.944355
	3	53.877315
	4	65.334875

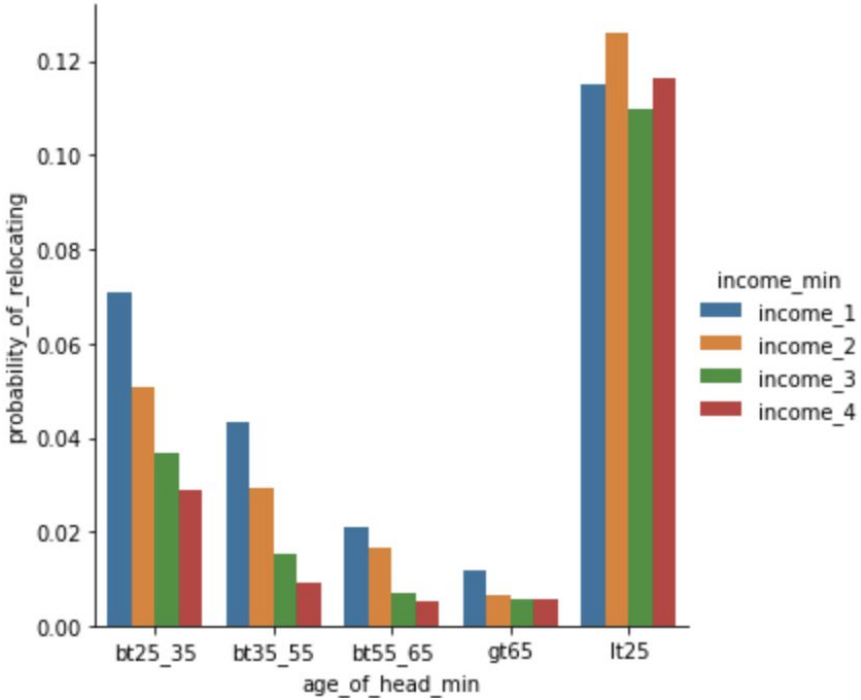
RUN 4074: Households Diagnosis



		perc_growth_2045
	age_of_head_min	
total_number_of_households	bt25_35	-3.875781
	bt35_65	-2.645849
	gt65	37.592639
	lt18	-11.188071

RUN 4074: HLCM Diagnosis

RELOCATION TOTALS



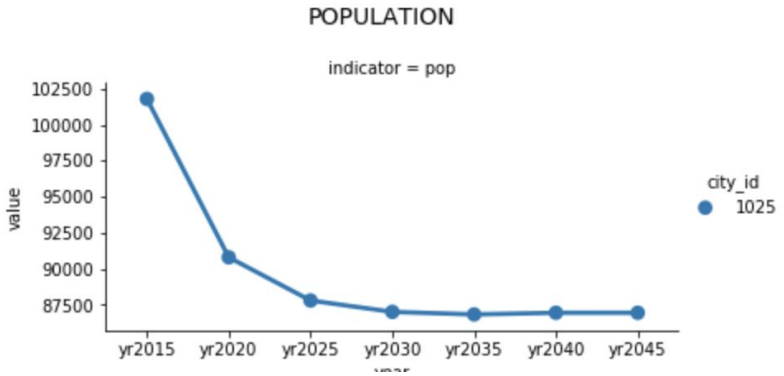
ANNUAL CONTROL TOTALS IN LARGE AREA 3:

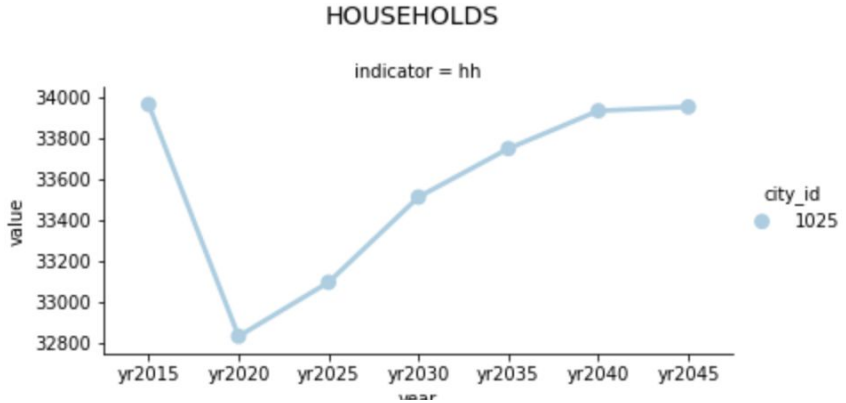
- Increase of low income households
- Increment of Elderly population and decrease of all other ages
- Increase of races 3 and 4, small decrease of race 1
- No increase of large HH sizes, Increase of households with one person

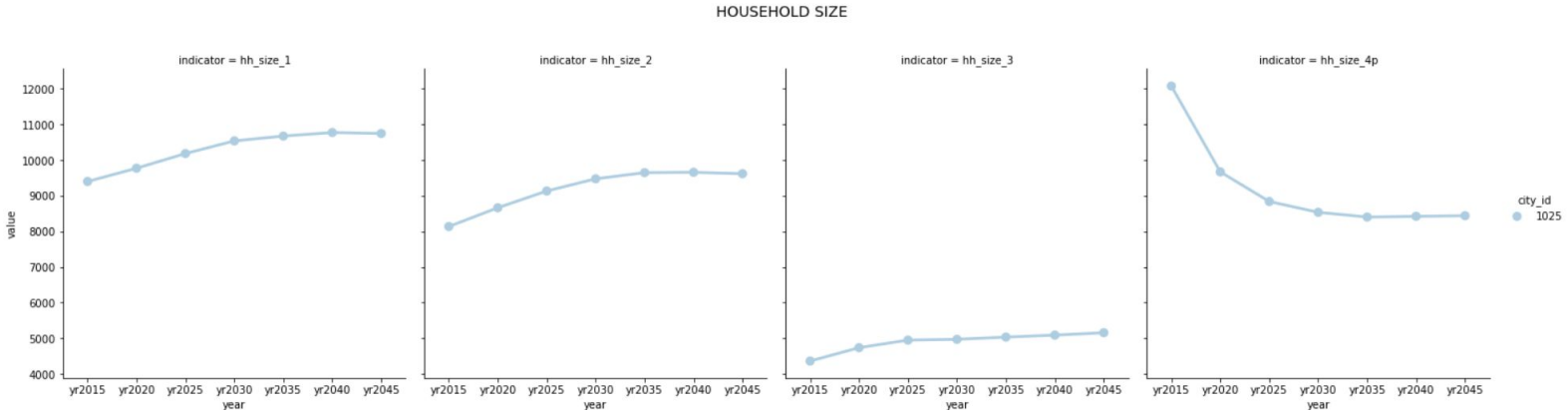
RELOCATION MODEL:

-
- HH with age under 25 have higher prob of relocating
- Low Income HH have higher prob of relocating

RUN 4074: Dearborn Diagnosis

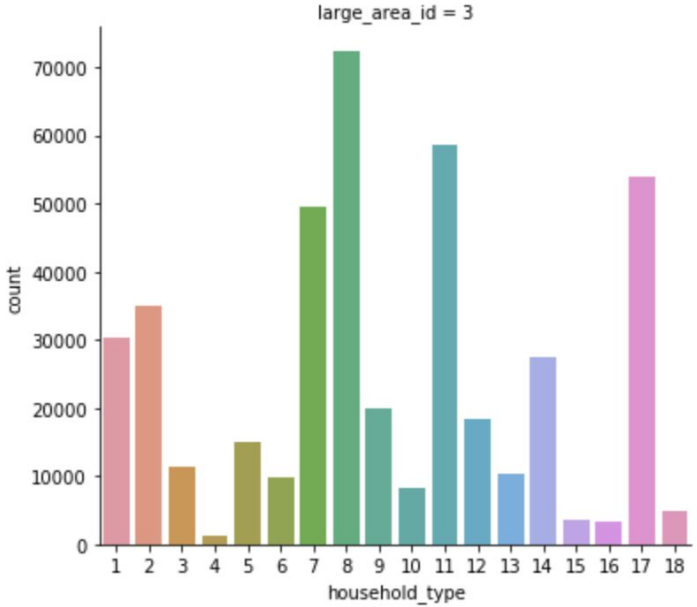




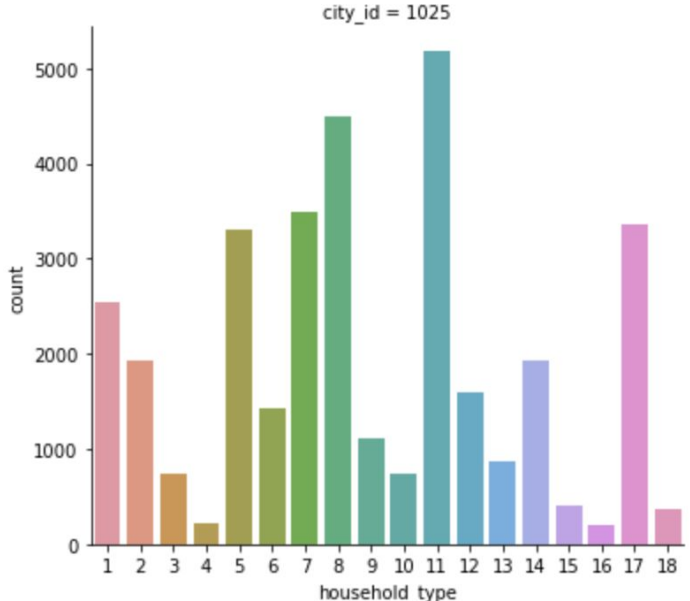


Base Data: HLCM

Base Year Data




```
1: 'income_quartile ==1 & persons <=2 & age_of_head >= 65',
2: 'income_quartile ==1 & persons <=2 & age_of_head >= 35 & age_of_head < 65',
3: 'income_quartile ==1 & persons <=2 & age_of_head < 35',
4: 'income_quartile ==1 & persons > 2 & age_of_head >= 65',
5: 'income_quartile ==1 & persons > 2 & age_of_head >=35 & age_of_head < 65',
6: 'income_quartile ==1 & persons > 2 & age_of_head < 35',
7: 'income_quartile in [2,3] & persons <=2 & age_of_head >= 65',
8: 'income_quartile in [2,3] & persons <=2 & age_of_head >= 35 & age_of_head < 65'
9: 'income_quartile in [2,3] & persons <=2 & age_of_head < 35',
10: 'income_quartile in [2,3] & persons > 2 & age_of_head >= 65',
11: 'income_quartile in [2,3] & persons > 2 & age_of_head >=35 & age_of_head < 65'
12: 'income_quartile in [2,3] & persons > 2 & age_of_head < 35',
13: 'income_quartile ==4 & persons <=2 & age_of_head >= 65',
14: 'income_quartile ==4 & persons <=2 & age_of_head >= 35 & age_of_head < 65',
15: 'income_quartile ==4 & persons <=2 & age_of_head < 35',
16: 'income_quartile ==4 & persons > 2 & age_of_head >= 65',
17: 'income_quartile ==4 & persons > 2 & age_of_head >= 35 & age_of_head < 65',
18: 'income_quartile ==4 & persons > 2 & age_of_head < 35'}
```



Dearborn have higher HH types: 17, 11, 8, 7, 5 and 1.
These means middle aged/elderly households, lower income hh, large hh sizes

RUN 4074: HLCM Variables

LOW INCOME HOUSEHOLDS:
- st_zones_prop_income_1

LARGE SIZE HOUSEHOLDS:
- st_zones_mean_hh_size

HIGH INCOME HOUSEHOLDS:
- st_zones_z_total_jobs
- st_zones_prop_income_4

ELDERLY AGE OF HEAD HOUSEHOLDS:
- st_zones_mean_age_of_head

RUN 4074: Employment Issues

%: MSE(4074) - MSE(4036) / MSE(4036)

Mean MSE	retail_buildings	financial_buildings	office_buildings	hospital_buildings	industrial_buildings
zones	23%	2%	-3%	-100%	12%
large_areas	-42%	128%	-73%	-100%	-78%
cities	35%	-33%	194%	-100%	-1%
counties	-45%	447%	-24%	-100%	-78%

Mean MSE	job_spaces	jobs
zones	16%	-89%
large_areas	804%	0%
cities	-55%	77%
counties	2276%	0%

Mean MSE	jobs_sector10	jobs_sector11	jobs_sector12	jobs_sector14	jobs_sector16	jobs_sector17	jobs_sector2	jobs_sector3	jobs_sector4	jobs_sector5	jobs_sector6	jobs_sector8	jobs_sector9
zones	-86%	-48%	-1%	-81%	-97%	-99%	-66%	-69%	-88%	-82%	-90%	-84%	-98%
large_areas	0%	-0%	-0%	-0%	0%	0%	0%	-0%	-0%	0%	-0%	-0%	-0%
cities	-32%	-92%	9%	-53%	-74%	-24%	-52%	-49%	75%	-44%	86%	140%	-76%
counties	0%	0%	-0%	0%	0%		-0%	-0%	0%	-0%	0%	-0%	0%

HLCM: RE-ESTIMATION

Estimating HLCMs without unplaced households and demolished/redeveloped buildings.

After running the transition, relocation and developer models.

Calibration- conceptual framing

1. Here's the data on changes over time I actually observed
2. What should the parameters of the simulator have been to match the observed data on changes over time?

- **Instead of asking which parameter values were most likely to generate a cross-sectional dataset**
- **Instead ask which parameter values were most likely to generate some longitudinal outcome**

Methodology: backpropagation

Calibration update

- HLCM calibration loop extended to all large areas, not just Oakland County
- Next step: Instead of running test simulation just within Oakland County, run test simulation region-wide with the newly calibrated HLCM coeffs for all large areas