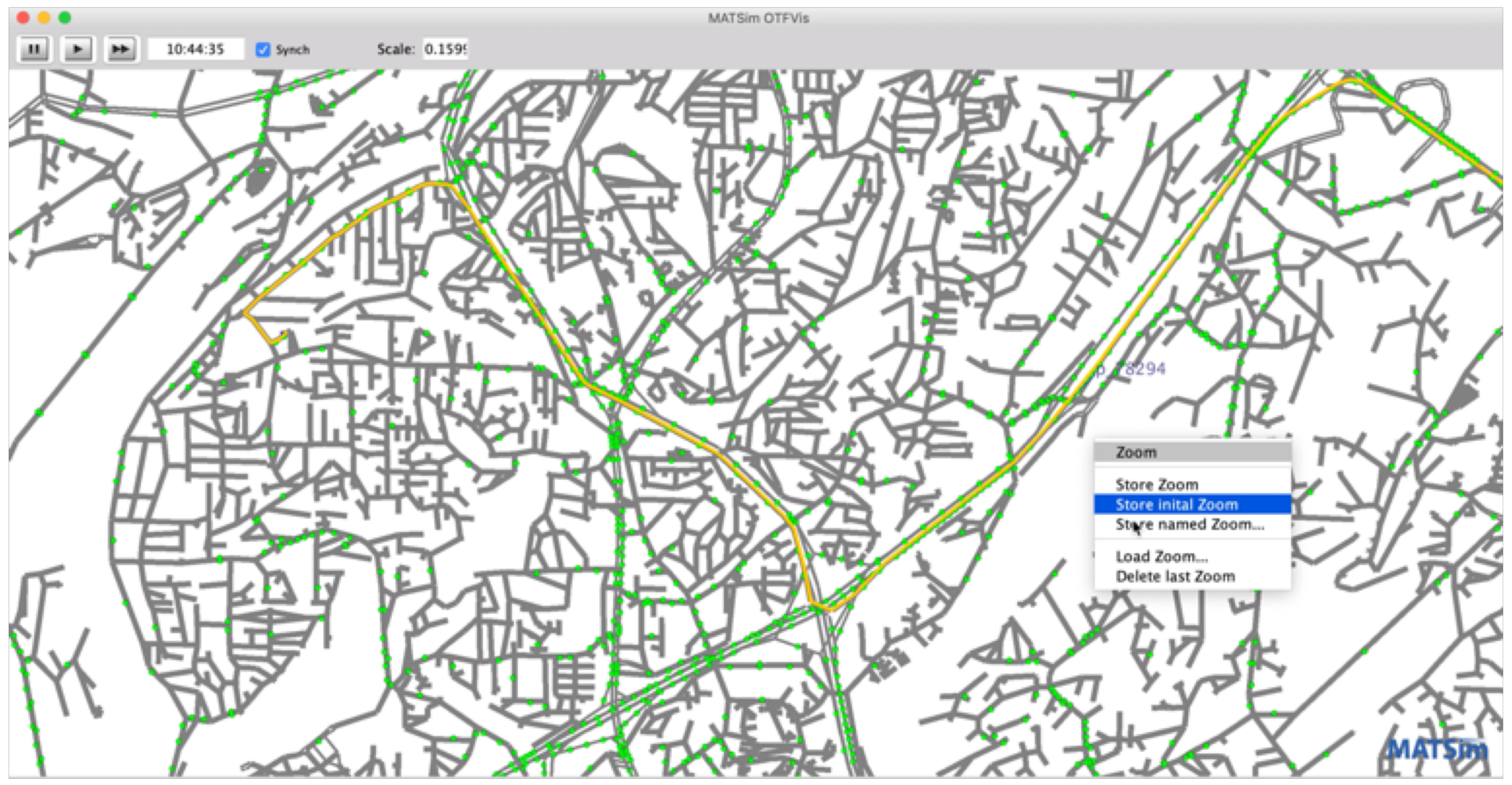
Realistic Transport Simulation with Open Data

Guimu Guo*, Jalal Khalil, Da Yan, Virginia Sisiopiku

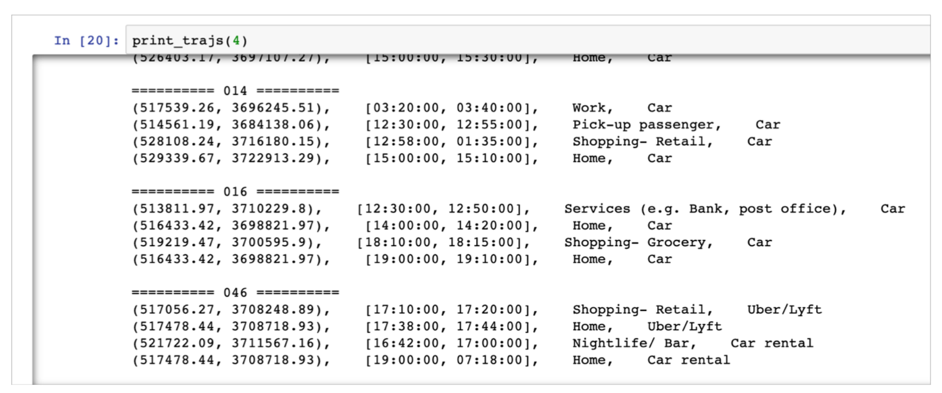
The University of Alabama at Birmingham, Contact: guimuguo@uab.edu

The Problem

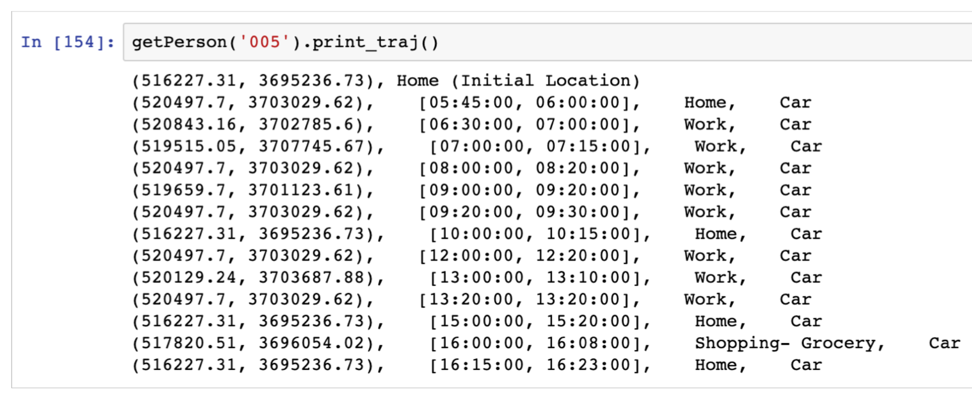
MATSim: agent-based transport simulation



User survey: 451 people with their day plans



Persons with 4 Legs



Legs of a Uber Driver

The Problem: to generate a realistic population from the survey as an input to MATSim for simulation

The simulation model can be used to study traffic problems such as road congestions, and impact of Uber traffic flow

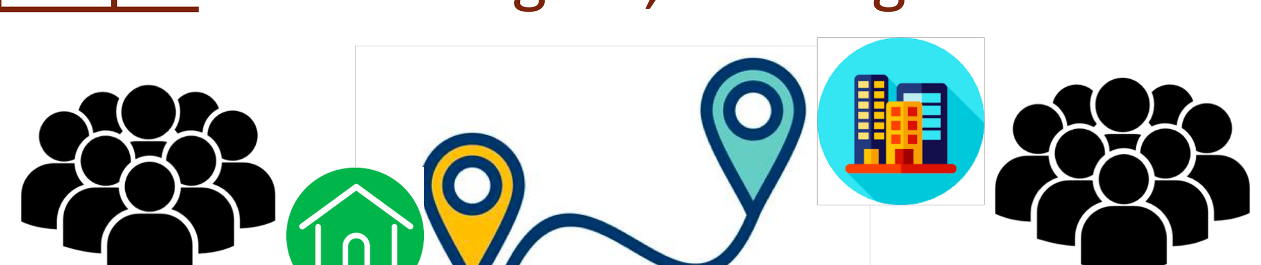
The Challenges

Conventional Approach: Iterative Proportional Fitting (IPF) according to agents' attribute distribution in the US Census

Challenges: the small data problem (tackled with open data)

IPF only generates # of people in each region; missing details like locations and activities

Sampling day plans?

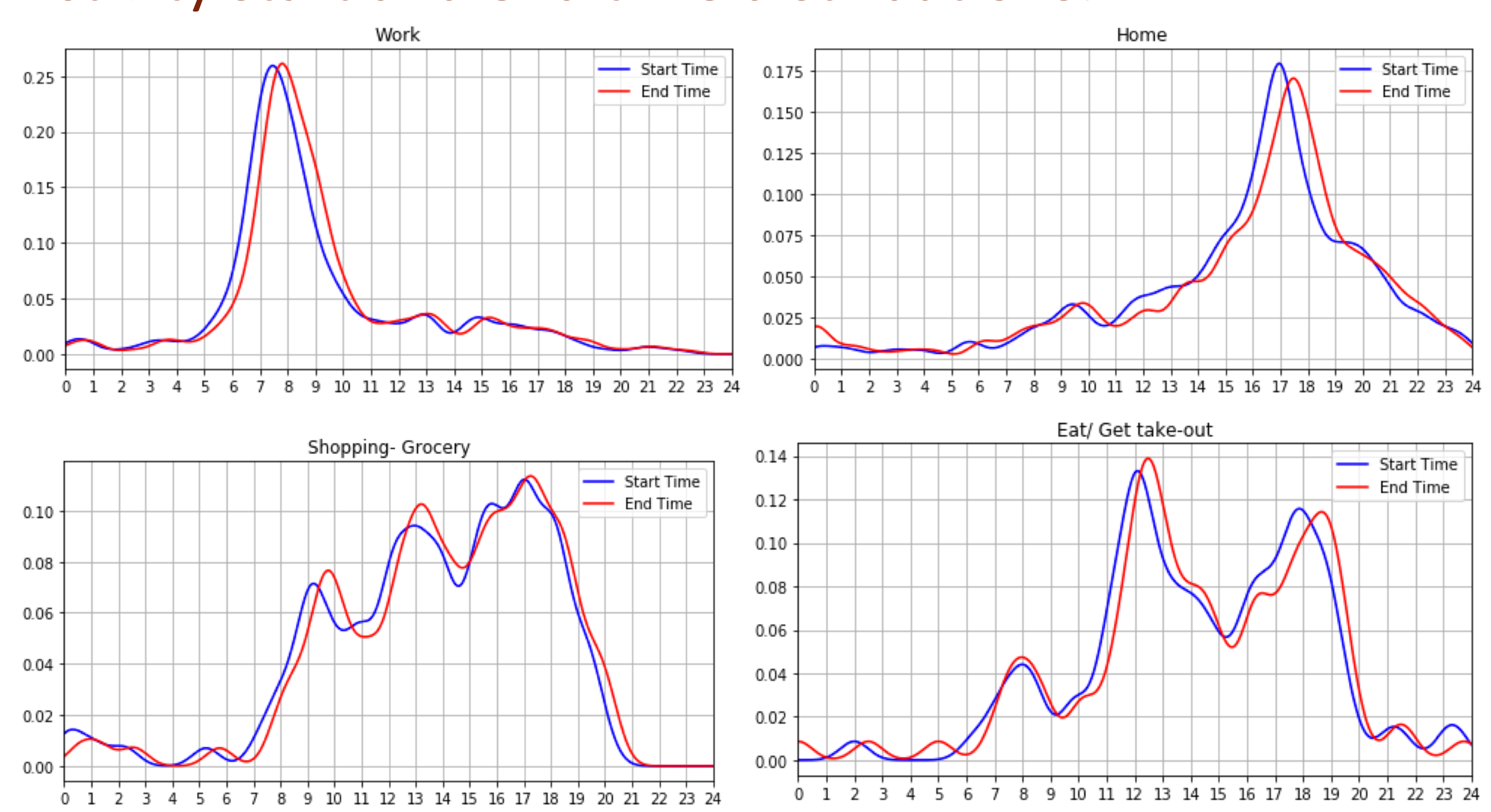


Our Solution

Data-Driven Approach: modeling (1) time, (2) location, (3) activities and (4) mode using generative machine learning models, with the help of open data.

Time Modeling using Kernel Density Estimation (KDE)

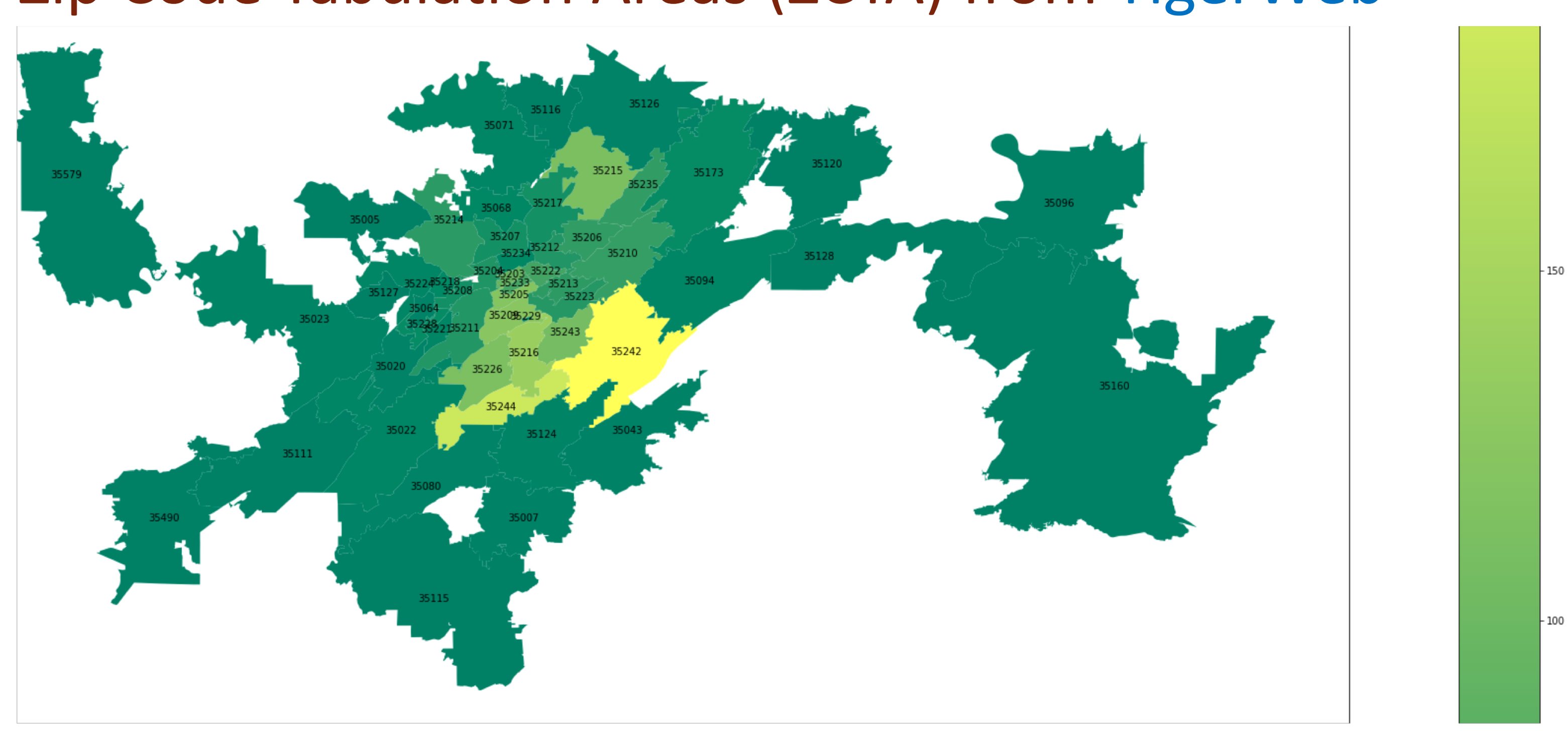
Activity start and end time distributions:



We can model the travel time using KDE similarly

Location Modeling with Open Data

Zip Code Tabulation Areas (ZCTA) from TigerWeb



Real addresses from OpenAddresses

IPF Source ZCTA marginal:

From American Community Survey (ACS), Variable "P03_0018E: COMMUTING TO WORK"

	NAME	DP03_0018E	zip code tabulation area
0	ZCTA5 35401	11636	35401
1	ZCTA5 35071	7141	35071
2	ZCTA5 35005	2777	35005
3	ZCTA5 35213	6409	35213
4	ZCTA5 35215	20017	35215
5	ZCTA5 35111	7809	35111
6	ZCTA5 35124	12974	35124
7	ZCTA5 35203	947	35203
8	ZCTA5 35206	6862	35206
9	ZCTA5 35209	16154	35209
10	ZCTA5 35218	2583	35218

Destination ZCTA marginal: Birmingham Business Alliance

BBA Top Employer List

	35215	35173	35206	35209	35226	35205	35242	35223	35207	35233	...
35215	6979	997	2493	499	0	499	0	0	0	997	...
35173	2458	4916	1229	1229	0	1229	0	0	0	0	...
35206	2357	295	589	295	295	0	295	0	589	589	...
35209	522	0	0	4441	2090	1567	2090	784	261	1045	...
35226	446	0	0	2674	4457	446	1783	0	446	891	...
35205	210	210	0	1468	0	3145	419	0	0	1887	...
35242	0	281	281	1684	1403	561	16557	281	0	1123	...
35223	0	0	0	227	0	453	227	1812	0	227	...
35207	0	0	277	277	0	0	0	0	555	0	...
35233	45	0	23	136	23	181	0	23	0	113	...
35211	1381	0	691	691	1381	0	0	0	0	691	...
35235	3337	667	667	0	0	0	667	0	0	0	...
35222	181	0	181	363	181	363	181	181	0	363	...
35243	0	0	167	1004	167	1004	2175	335	0	836	...

IPF Output Matrix

Generating agents' day plans using the IPF output matrix

Evaluation

Comparison with real traffic data

MATSim ID	Interstate	Monitored Speed	Simulation Speed
50779678_0	I-65 N	59	58.17957611
50779764_0	I-65 N	57	55.39014001
50780374_1	I-20 W/ I-59 S	67	67.23911964
119869822_0	I-20 W/ I-59 S	63	68.11513108

Future Work

OpenStreetMap POI

```
============= (33.46527270638653, -86.81967834785836) =============
>> NN 1 : dist =  0 min
supermarket (Publix): 33.465192, -86.819874
>> NN 2 : dist =  2 min
supermarket (aldi): 33.4639637, -86.8213989
>> NN 3 : dist =  3 min
supermarket (Walmart Neighborhood Market): 33.4678637, -86.8215147

============= (33.40365486174883, -86.80835731664584) =============
>> NN 1 : dist =  0 min
supermarket (Publix): 33.403505, -86.8082925
>> NN 2 : dist =  27 min
supermarket (Whole Foods Market): 33.3771436, -86.8007887
>> NN 3 : dist =  29 min
supermarket (Walmart Neighborhood Market): 33.428559, -86.791833
```

```
In [222]: root.print_maxDepth(3)
+- ROOT, 443
   +- Home, 422
   |  +- Home, 23
   |  |  +- Home, 23
   |  +- Work, 170
   |  |  +- Home, 80
   |  |  +- Shopping- Grocery, 16
   |  |  +- Eat/ Get take-out, 10
   |  |  +- Work, 10
   |  |  +- Pick-up passenger, 9
   |  |  +- Shopping- Retail, 8
   |  |  +- Services (e.g. Bank, post office), 9
   |  |  +- Drop-off passenger, 4
   |  +- Shopping- Grocery, 49
   |  |  +- Home, 23
   |  |  +- Shopping- Grocery, 8
   |  +- Services (e.g. Bank, post office), 43
   |  |  +- Home, 12
   |  |  +- Shopping- Grocery, 6
```

Video Demo

https://youtu.be/ZIm0WsmKB4E