3-3-2017

Big Data and the Future of Travel Modeling

Greg Macfarlane
Parsons Brinckerhoff

Big Data and the Future of Travel Modeling

Gregory Macfarlane, PhD, PE

3 March, 2017

Me

- Graduated from BYU, Georgia Tech
- Past: Utah Transit Authority, National Center for Sustainable Transportation
- Now: Systems Analysis Group at WSP | Parsons Brinckerhoff, Transport Foundry

Outline

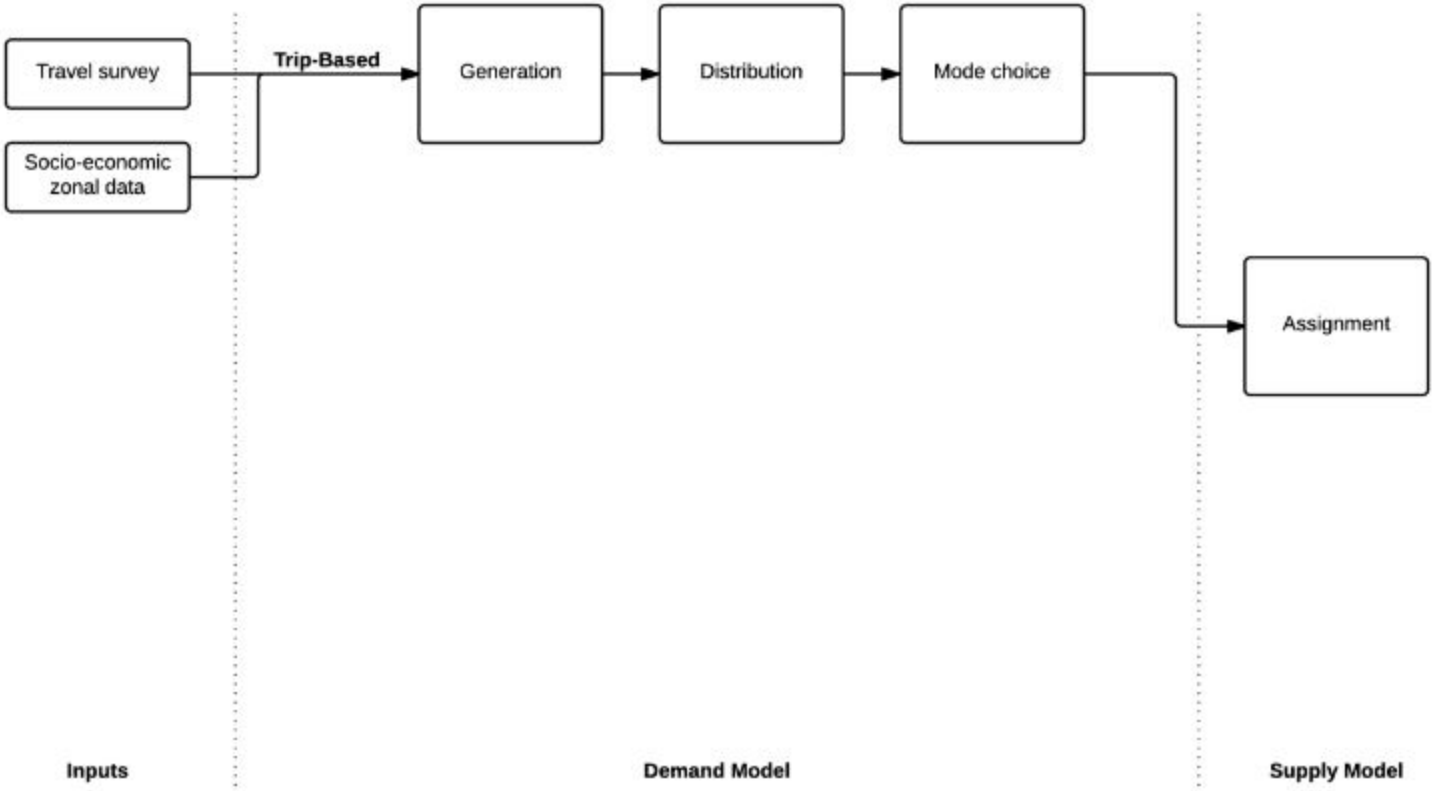
- Where we've been in travel modeling
- Confronting current challenges with data
- What's coming next?

A brief history of travel modeling

What is the point of a travel model?

"The purpose of travel forecasting is not to predict the future, but to make informed decisions now." — Eric Miller

Aggregate Trip-Based Model



Prehistory (1950-1970)

- Interstate Highway System
- Detroit and Chicago
- "Three-step" models

Mode Choice (1970-1990)

Multinomial logit model (McFadden). Person n will pick highest-utility option. Observed and unobserved:

$$U_{ni} = X_{ni}\beta + \epsilon, \forall i \in J$$

Probability of i highest, if assume equally random ϵ:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in J} e^{V_{nj}}}$$

Mode Choice (1990-2005)

- Transit investment
- Nested logit: group similar modes into nests

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \left(\sum_{j \in l} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l \in K} \left(\sum_{j \in l} e^{V_{nj}/\lambda_l} \right)^{\lambda_l}}$$

Choice Model Mania (2005-present)

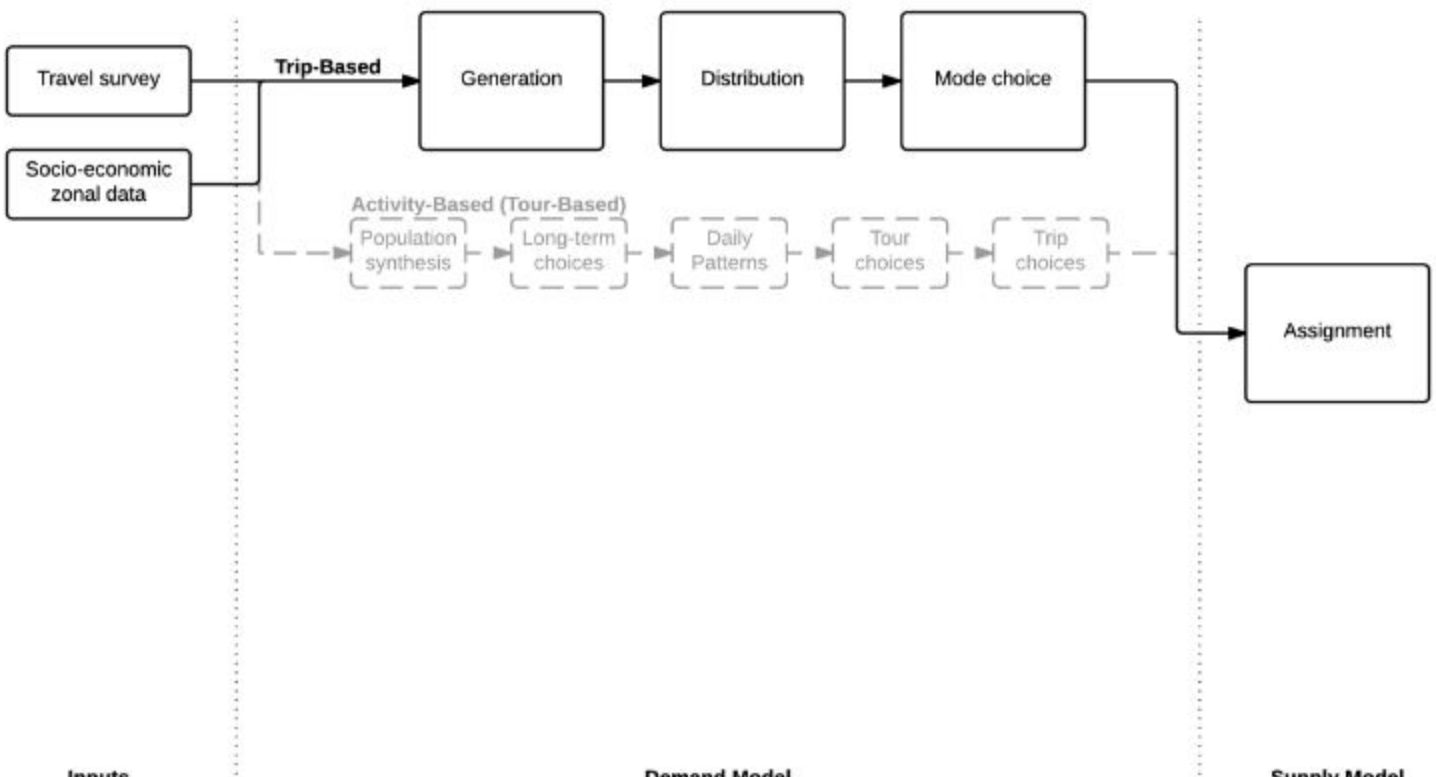
New econometric forms:

- Cross-nested logit
- Mixed logit
- New probit estimators
- Spatial/social dependence
- Discrete-continuous models

New applications:

- Destination choice
- Auto ownership
- Daily activity pattern choice
- Coordinated choice

Activity-Based Model



How is now different from the 1960's?

- Constant infrastructure while behavior is changing
- There's lots of data
- Machine learning/artificial intelligence

Big Data

What on earth is Big Data?

"High-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making." -Gartner

Big data in transportation

- Cellular/GPS traces
- origin-destination
- probe data
- targeted marketing/credit reports
- administrative records
- Android location data
- others?

New insights into old problems - 1

Who is failing emissions tests? And how can we use emissions testing to improve air quality?

Joined emissions test database with target marketing records, and modeled test failures as a function of socioeconomic data.

Policy	Missed/Avoided failures	Cost per avoided failure	Benefit
Extended exemptions	(2645)	*tricky*	$1.8 M, mostly to wealthy
Maintenance subsidy	430	$19k	$8 M, entirely to poor
Cash-for-Clunkers	95	$59k	$5.9 M, mostly to wealthy

New insights into old problems - 2

Does the built environment influence travel behavior?

Or do people choose their built environment to suit their prefered behavior?

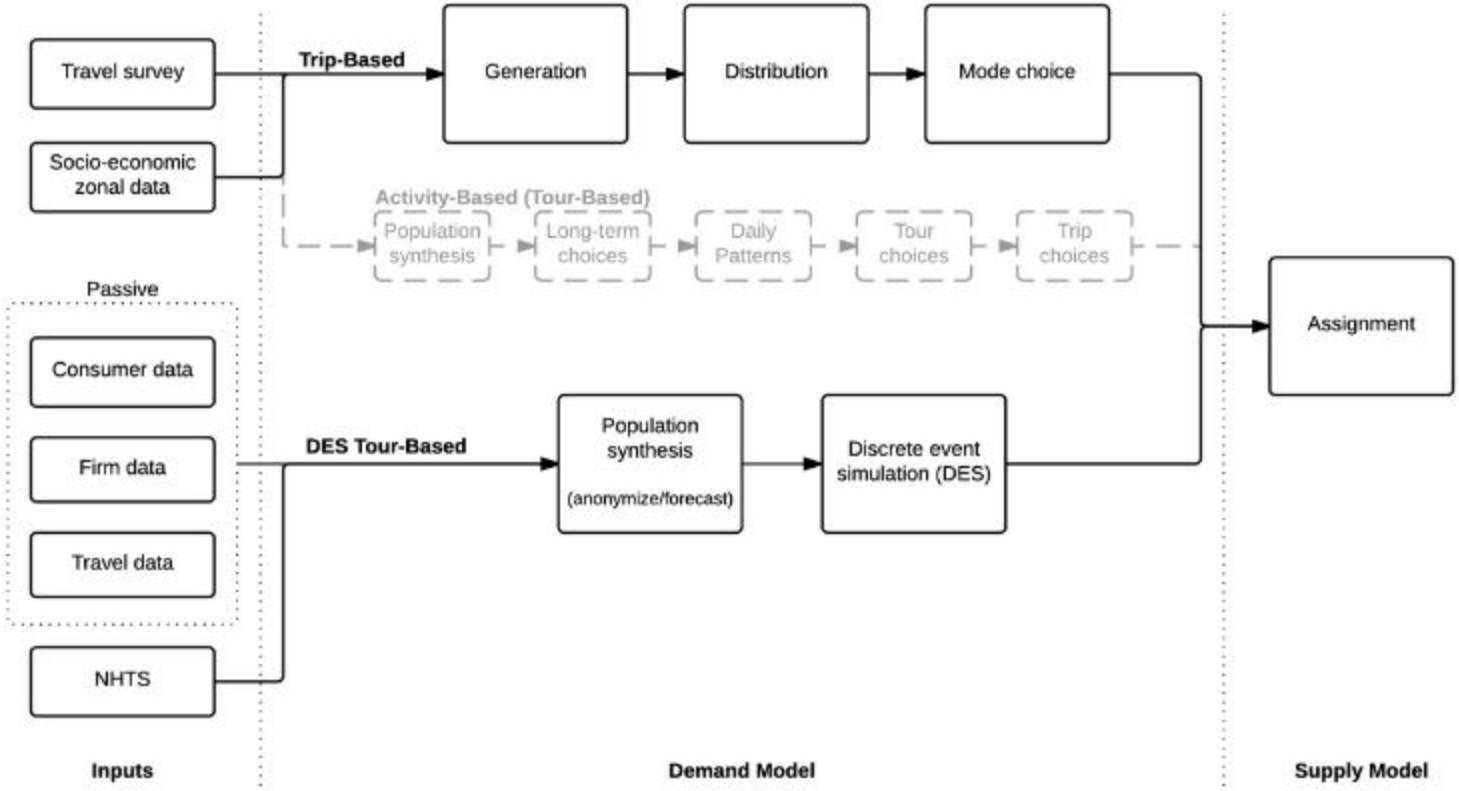
Or… does the built environment affect people's preferences?

Got address histories from the credit records of 250,000 households in Atlanta, and modeled vehicle ownership.

Scenario	Total Vehicles	Δ from Reference
Reference	452,170	
Past equals present	457,537	1.2
Mean past	446,748	-1.2
Random past	452,016	-0.034
1^{st} percentile	478,380	5.8
5^{th} percentile	469,393	3.8
95^{th} percentile	437,524	-3.2
99^{th} percentile	425,169	-6.0

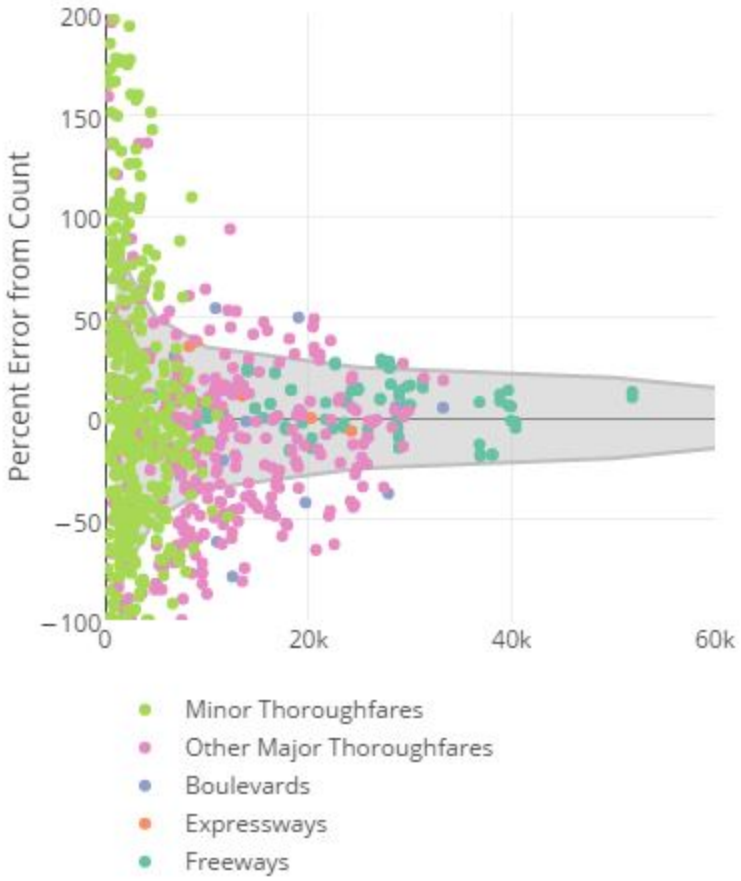
A new model?

Passive Data Tour-Based Model

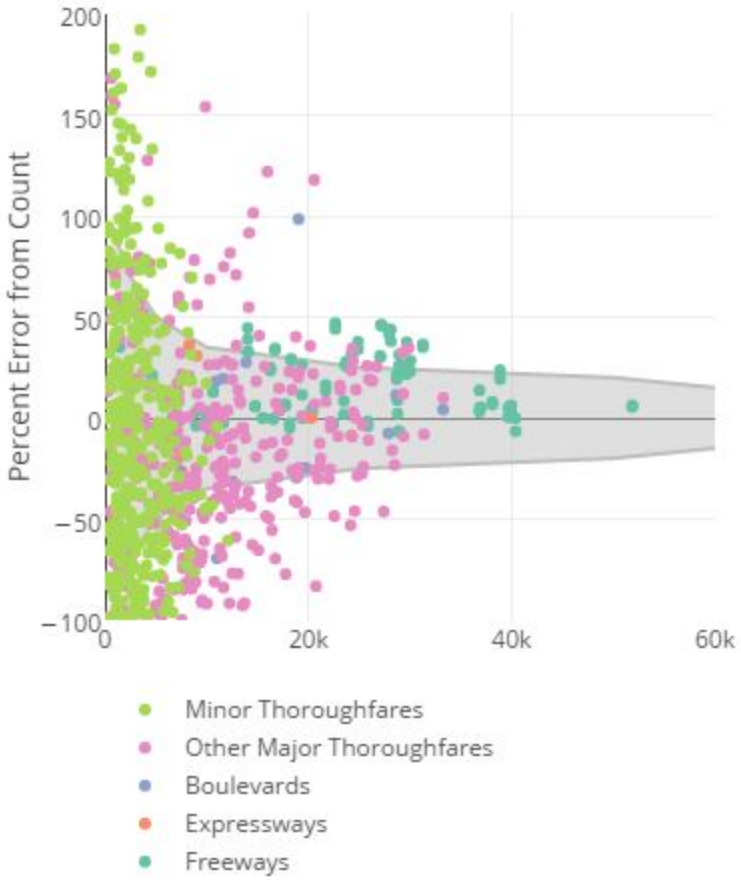


Assignment

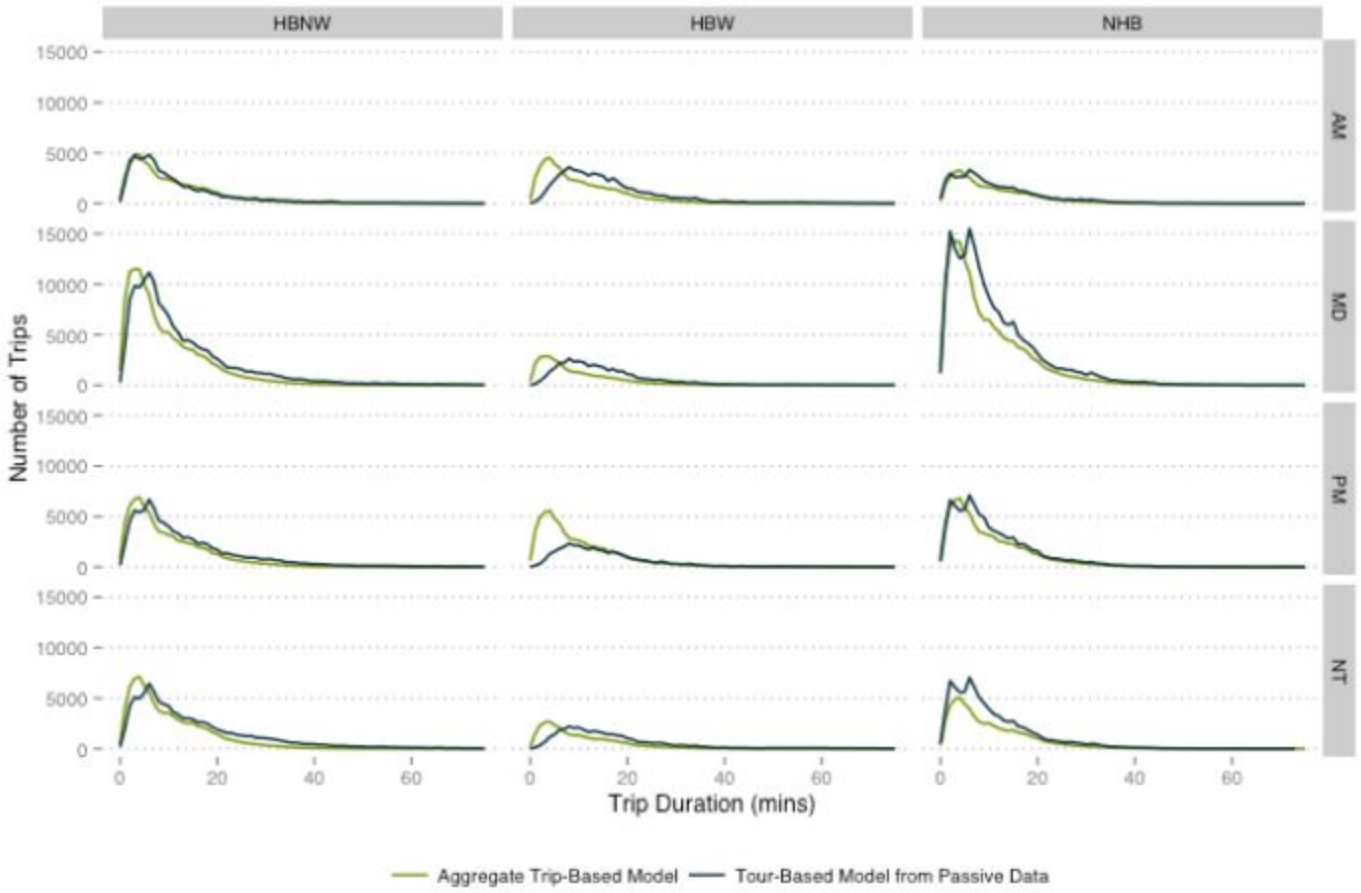
Trip-Based Model



DES Model

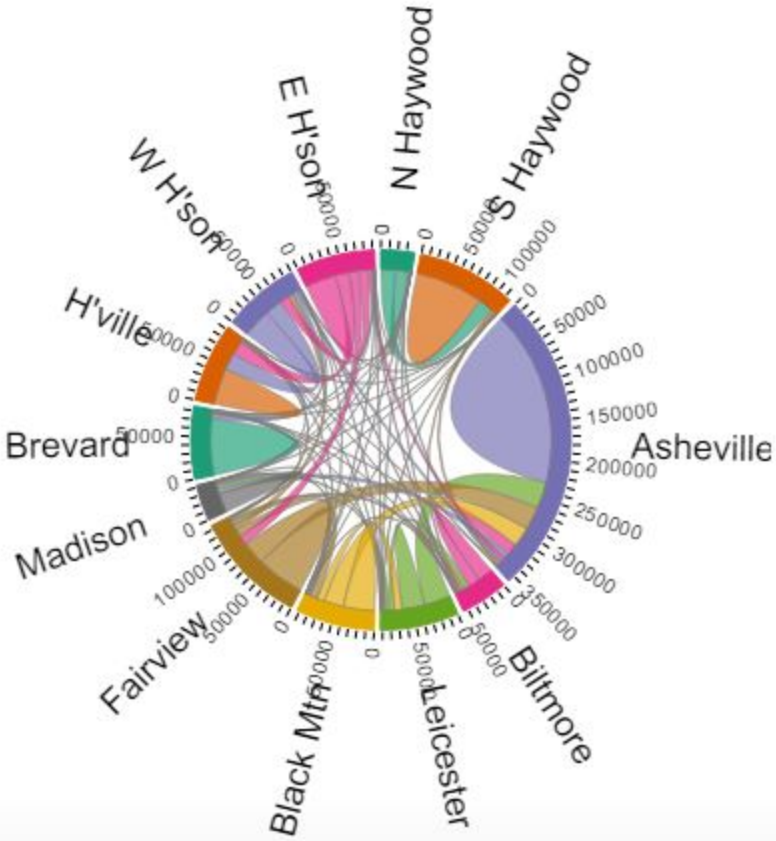


Distribution: Trip Length

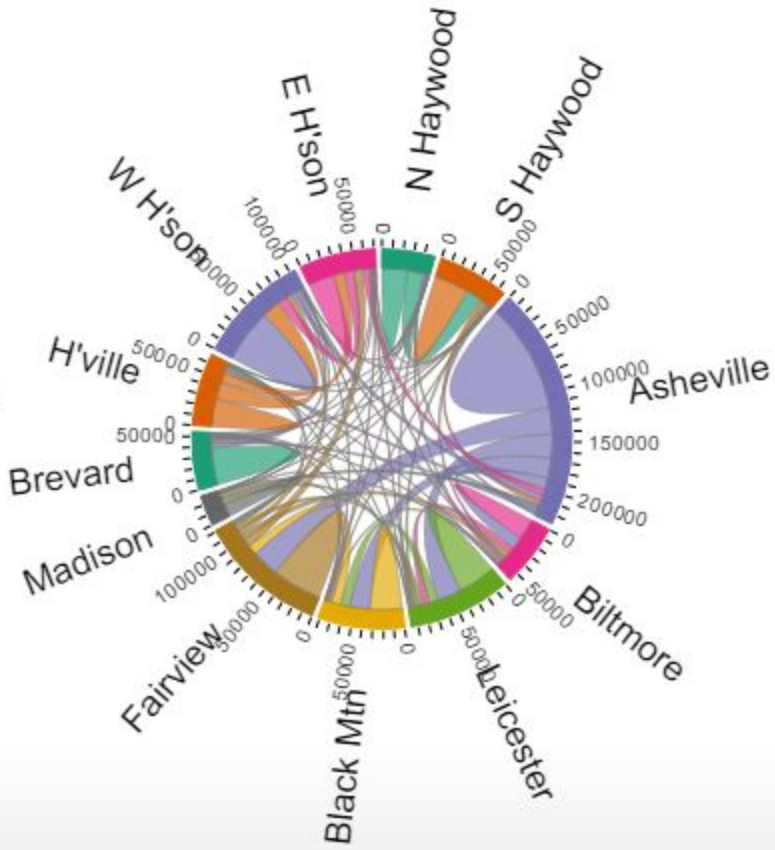


Distribution: Demand Flows

Trip-Based Model



Tour-Based Model



MATSim integration

6:45 AM

7:00 AM

Where might we go from here?

- Adjust underlying probability distributions to reflect uncertainty
- Use AI to generate responsive OD matrices

Questions