A Framework for the Comparative Analysis of Diverse Mobility Data

Jessica Embury & Atsushi Nara, PhD
Department of Geography, San Diego State University
March 26, 2023 @ AAG Annual Meeting

HDMA @SDSU

SDSU | San Diego State University

Are we in a “Golden Age” of mobility data?

- Constant collection from mobile devices
- More geospatial detail
- Plentiful sources
- Supporting computational advances

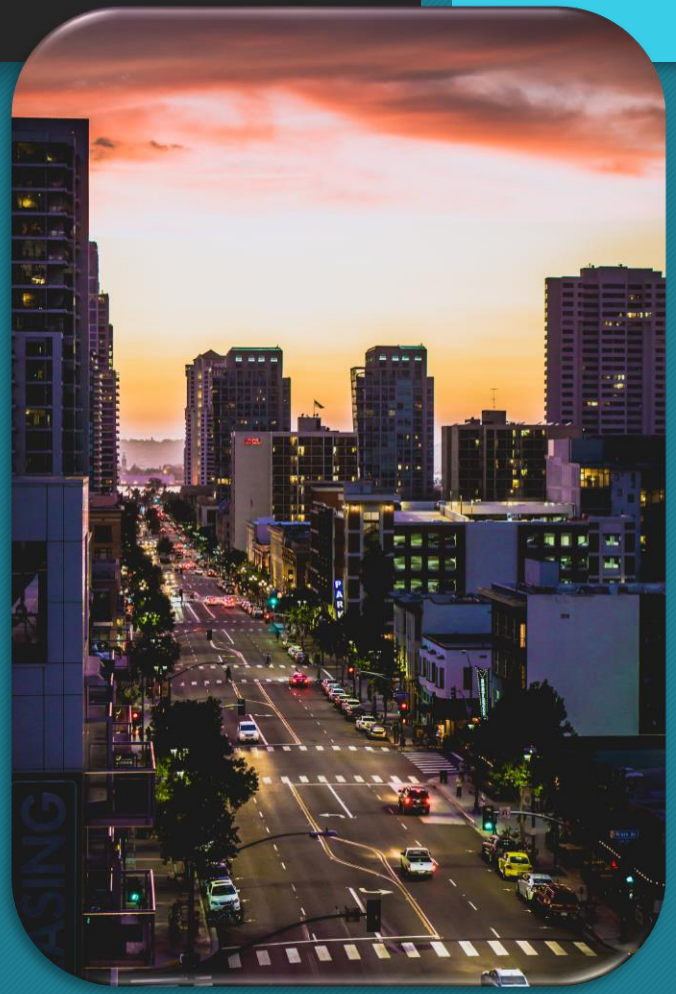
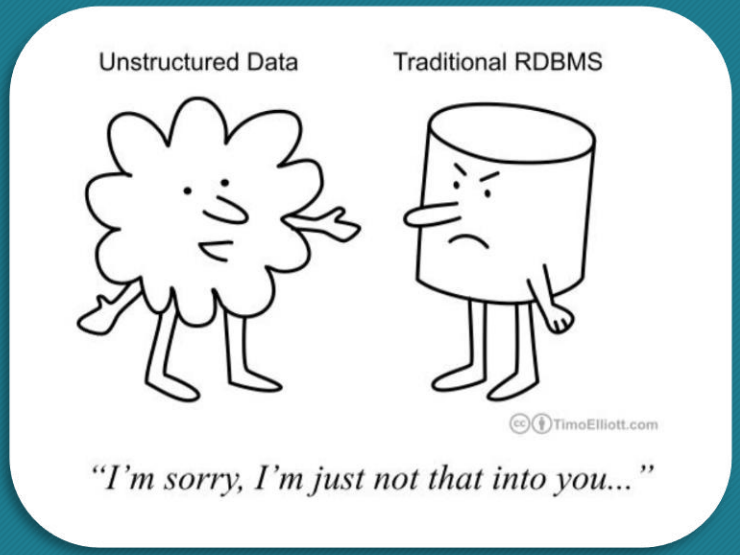
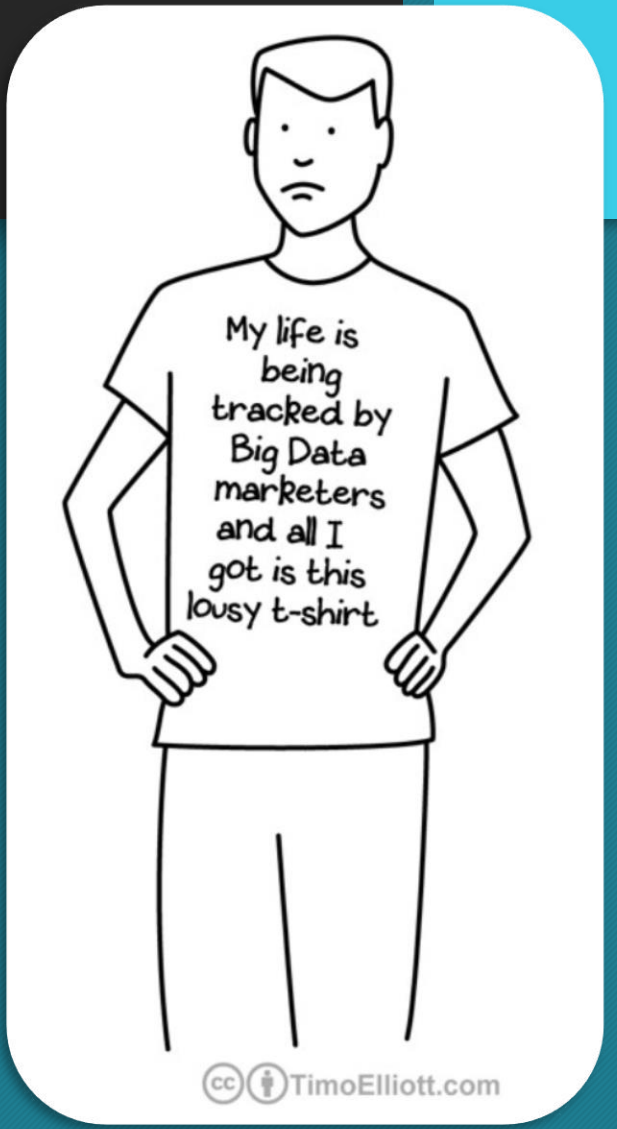


Photo by Derek Story on Unsplash

What is the downside?

- Size and structure
- Privacy and ethics

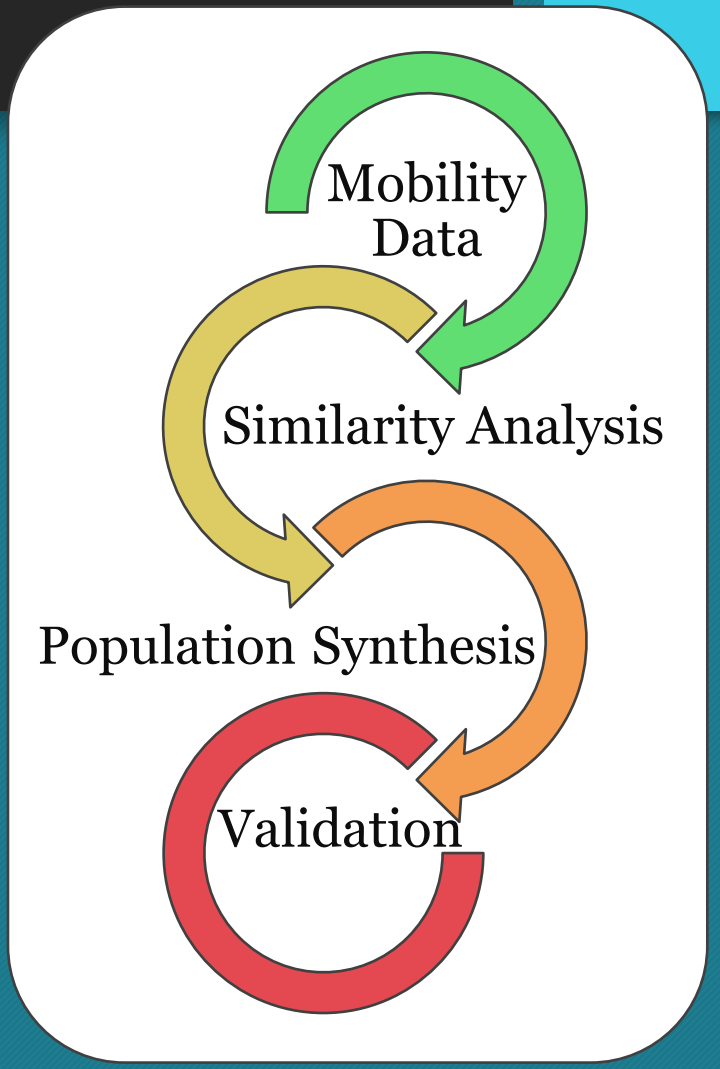




Cartoons by Timo Elliott at timoelliott.com

What are the research goals?

1. Discover how to synthesize complementary data to maximize insights
2. Investigate the activity contexts that motivate mobility in the study area ("urban" San Diego County, CA, US)
3. Validate results with independent data



Data Selection

- Public and private
- Different spatial & temporal scales
- 1758 Census Block Groups (CBGs)/34 Sub-Regional Areas (SRAs), 5749 travel diaries, 152988 points of interest (POIs)
- Different data purposes & biases

American Community Survey (ACS, 2017)

LEHD Origin-Destination Employment Statistics (LODES, 2017)

Caltrans National Household Travel Survey (NHTS, 2017)

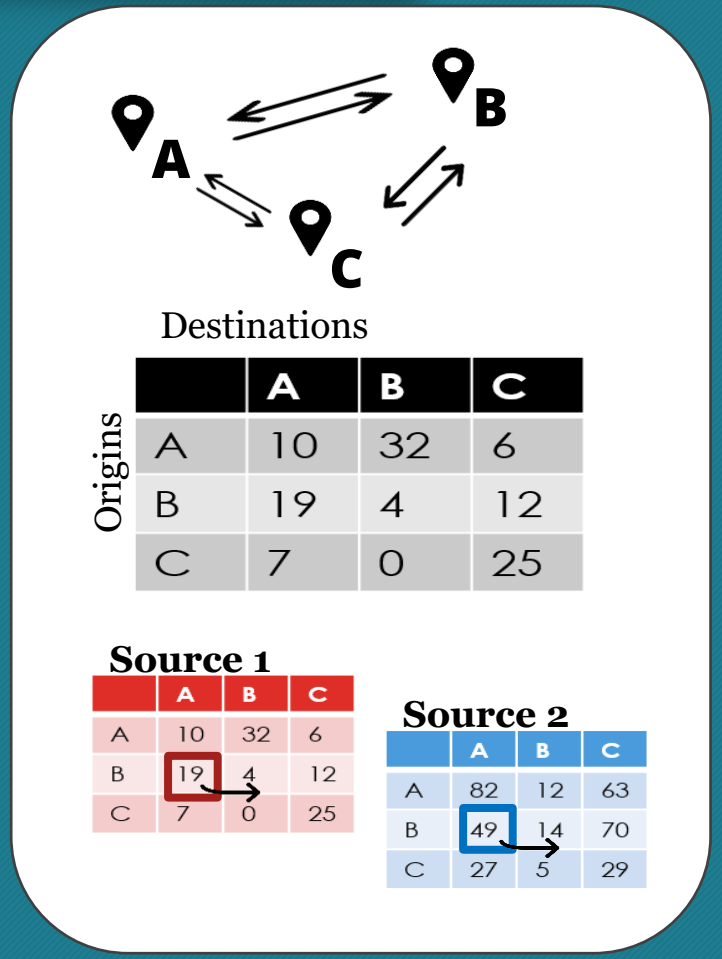
SafeGraph Monthly Patterns & "Social Distancing" (SG, 2019*)

* Earliest available

Similarity Analysis

How are the data alike and different?

- Spatially weighted structural similarity index (Embury et al., 2022; Jin et al., 2019)
- Normalizes origin-destination flows
- Compares mobility flows of similar distance



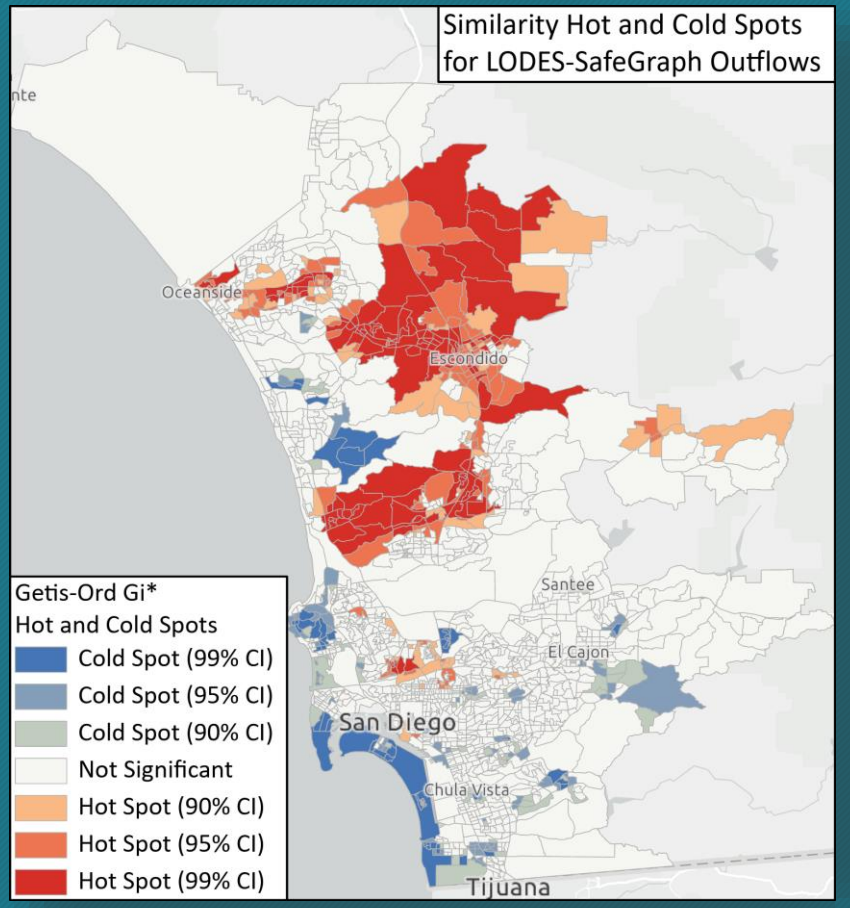
	A	B	C
A	10	32	6
B	19	4	12
C	7	0	25

	A	B	C
A	82	12	63
B	49	14	70
C	27	5	29

Similarity Analysis

LODES-SafeGraph Results

- In North County: Higher percent of outgoing trips related to commuting
- Along the coast: Lower percent of outgoing trips related to commuting
- What about the NHTS travel survey? (~21,000 trips by ~4,500 individuals)



Population Synthesis

Incorporating ACS & NHTS Data

- Create a synthetic population of residents that is representative of the community (3.2M individuals, 1.1M HHs, 1.4M workers)
- Iterative Proportional Updating (Ye et al., 2009)

Photo by Jacek Dylag on Unsplash

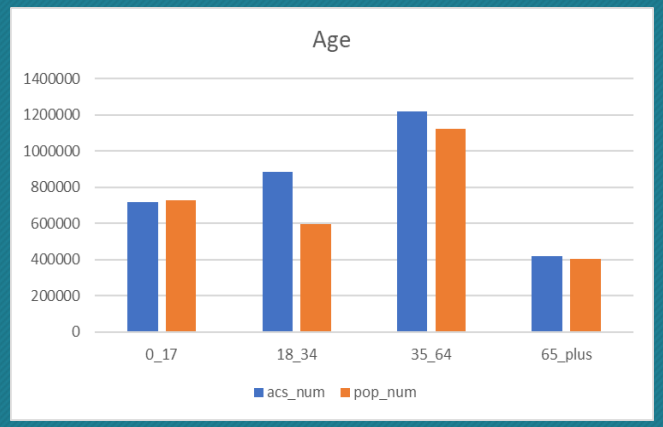
Age
Education

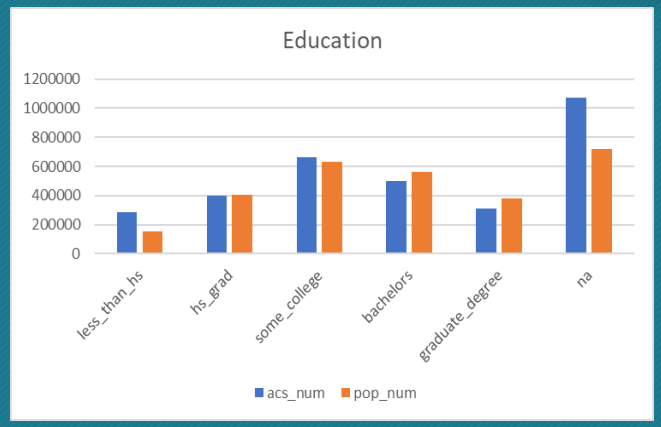
Size
Income

Validation

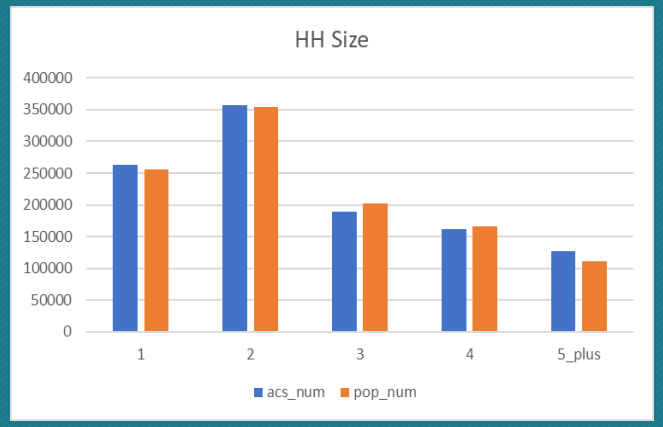
Individual & Household Attributes

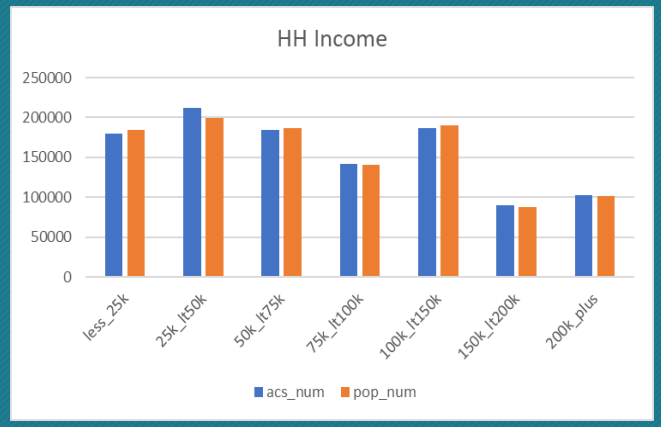
Age (L)
Education (R)





Size (L)
Income (R)



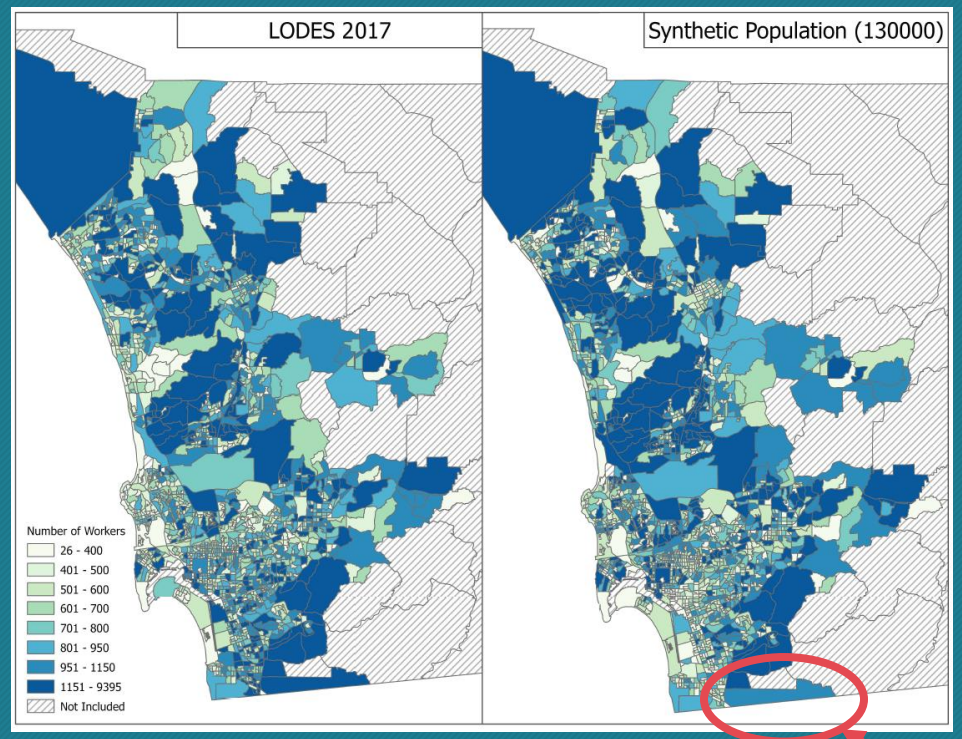


ACS / Census

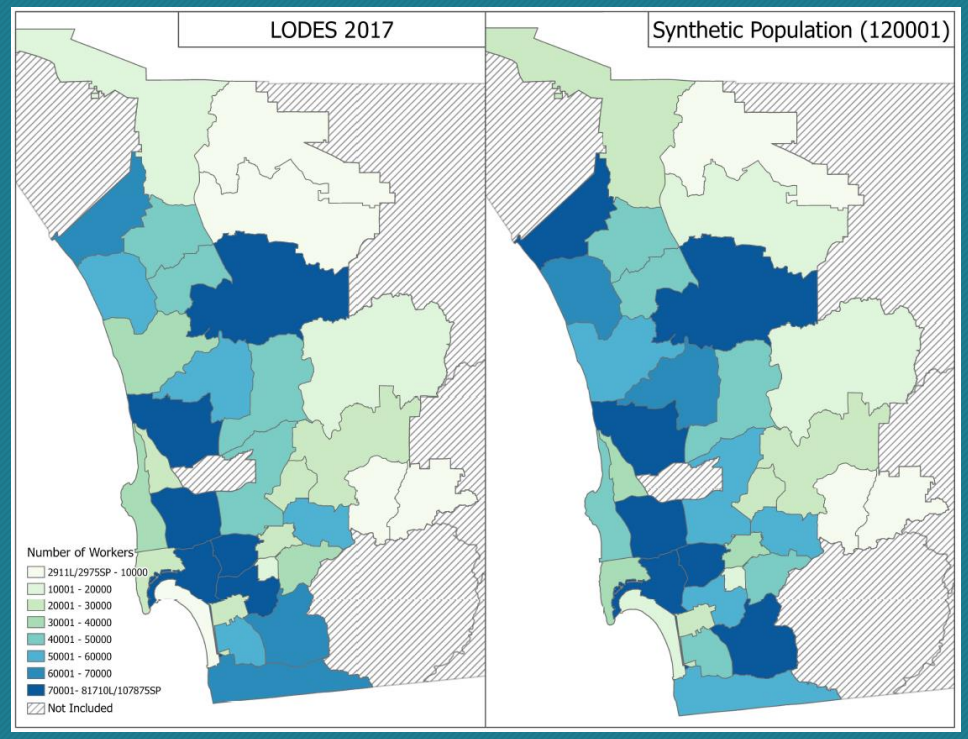
Synthetic Population

Validation

Number of Commuters Per Origin CBG/SRA: LODES (2017) (L) & Synthetic Population (R)



Pearson r = 0.883, p-value < 0.001

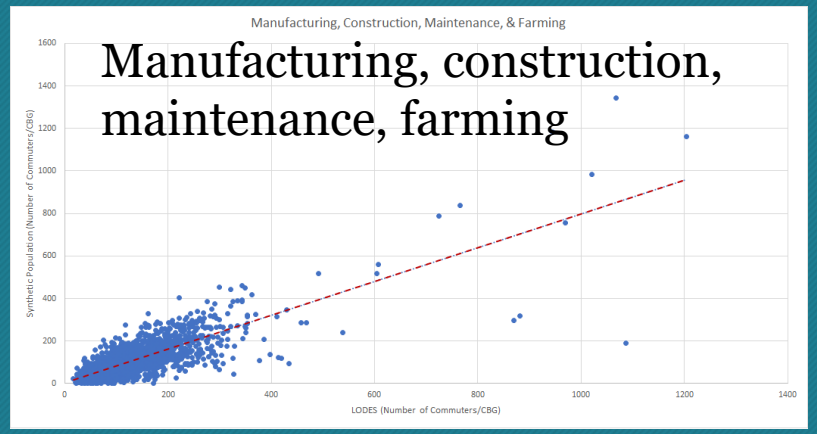


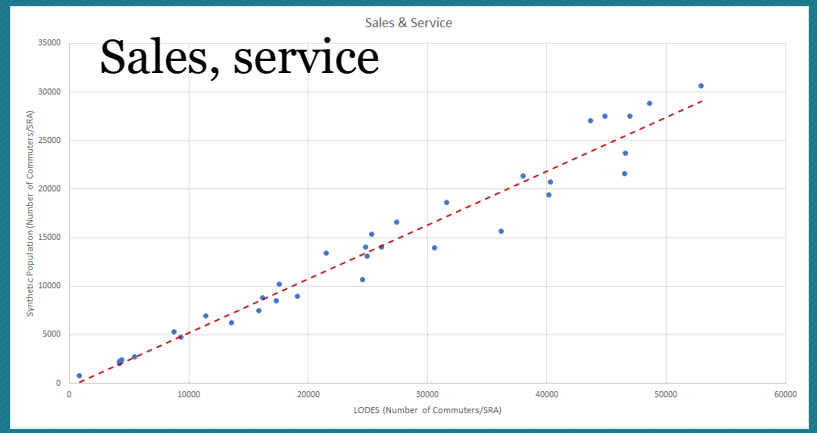
Pearson r = 0.946, p-value < 0.001

Border dynamics?

Validation

Comparing Commuters by Work Industry: LODES (2017) (X) & Synthetic Population (Y)



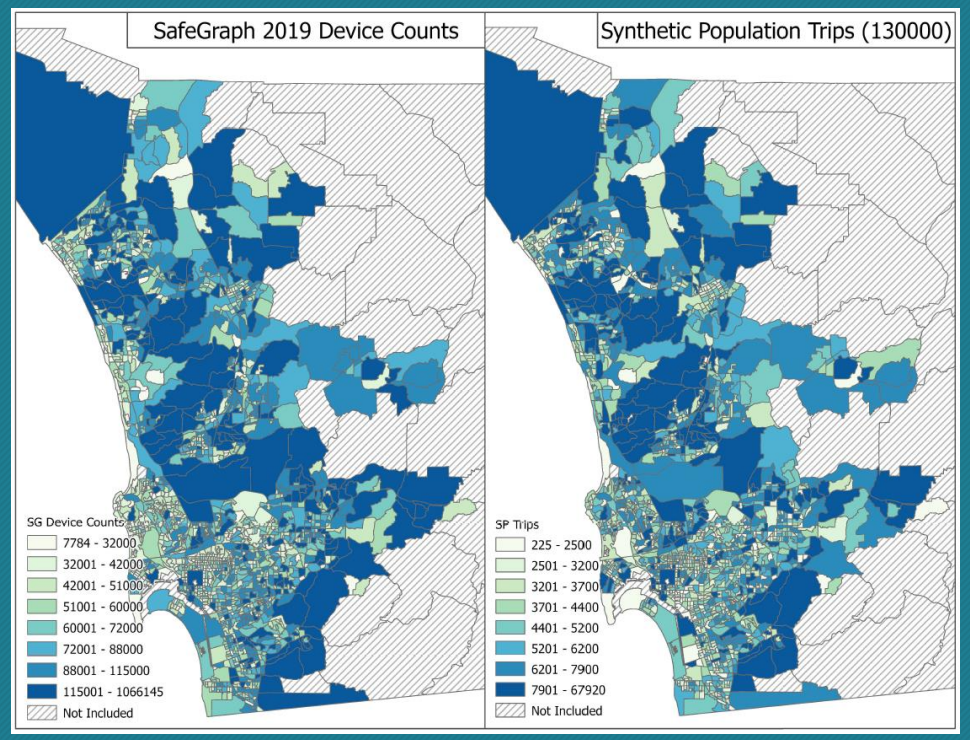


Work Industry	**Pearson r** (CBG, *n*=1758)	**Pearson r** (SRA, *n*=34)
Clerical, administrative	0.799***	0.942***
Manufacturing, construction, maintenance, farming	0.801*** (Upper)	0.951***
Professional, management, technical	0.923***	0.959***
Sales, service	0.879***	0.979*** (Lower)

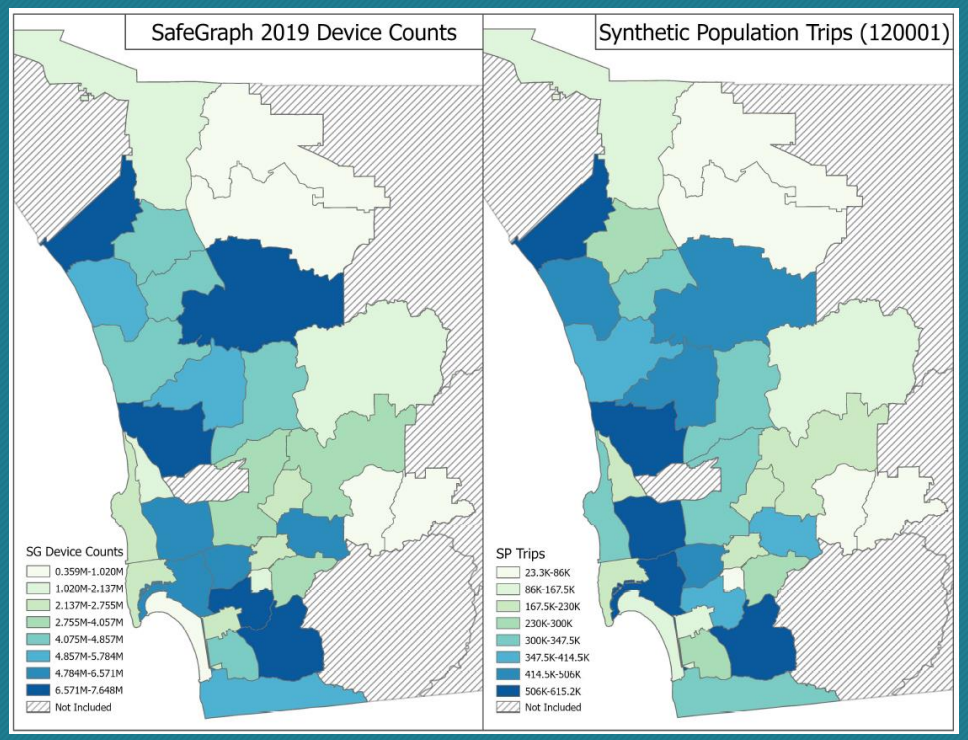
*** *p*-value < 0.001

Validation

Number of Trips Per Origin CBG/SRA*: SafeGraph (2019) (L) & Synthetic Population (R)



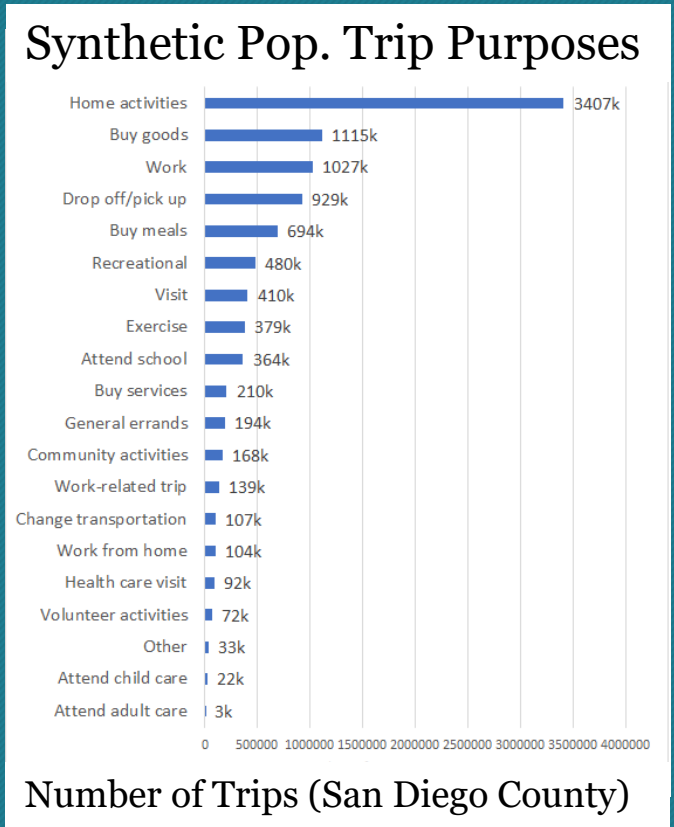
Pearson r = 0.914, p-value < 0.001



Pearson r = 0.942, p-value < 0.001

Validation

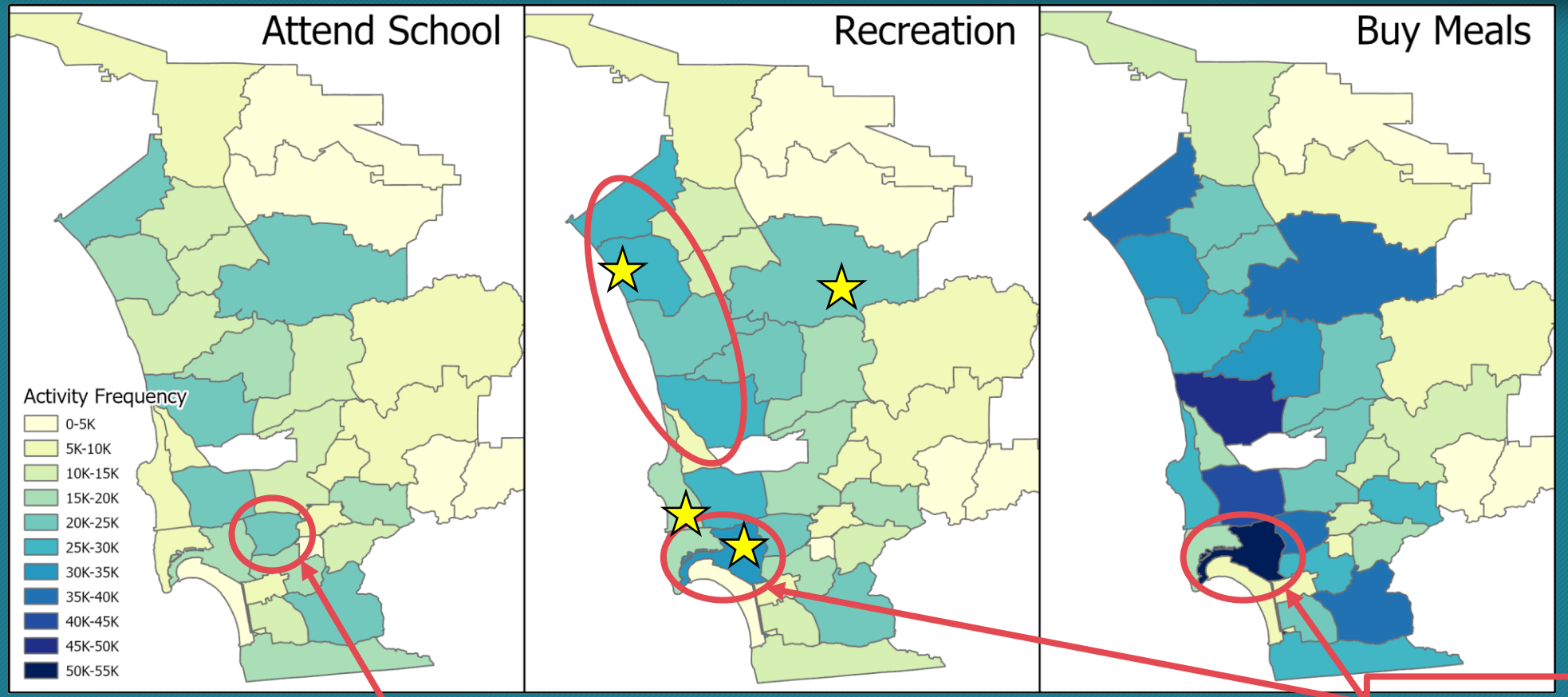
Comparing Trips by Activity Type: SafeGraph (2019) & Synthetic Population

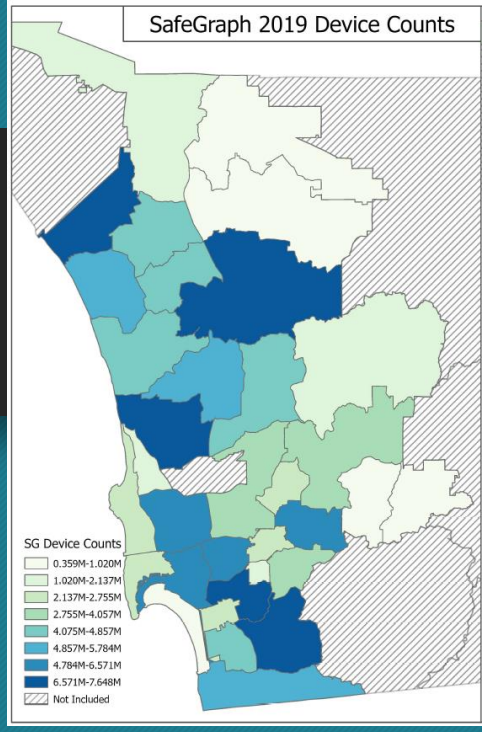


All *p*-values < 0.001

NHTS Trip Description	Pearson r (CBG)	Pearson r (SRA)
Attend school as a student	0.890	0.906
Buy meals (go out for a meal, snack, carry-out)	0.903	0.925
Recreational activities (visit parks, movies, bars, museums)	0.861	0.917
Buy goods (groceries, clothes, appliances, gas)	0.890	0.900

How do activities contextualize mobility?





Future Research

Activity Scheduling

- Similarity analysis of origins *and destinations*

Activity-based Modeling

- Incorporate interactions → complex behaviors
- EX: Traffic affecting travel times and trip locations

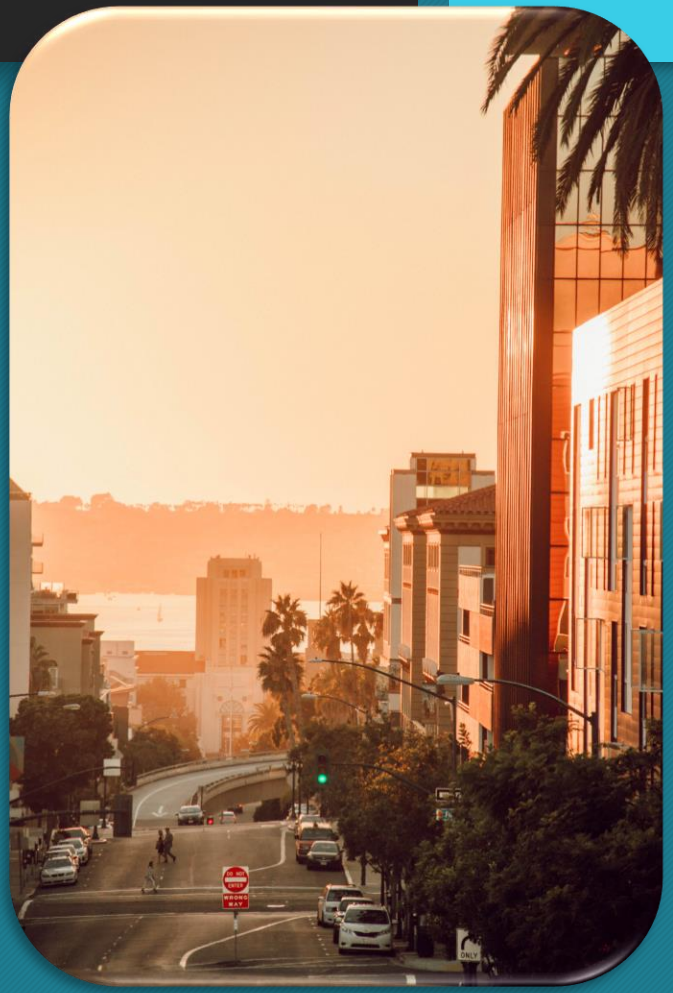


Photo by Lital Levy on Unsplash

References

Embury, J., Nara, A., Jin, C. (2022). Spatially weighted structural similarity index: A multiscale comparison tool for diverse sources of mobility data. In HANIMOB'22: The 2nd ACM SIGSPATIAL International Workshop on Animal Movement Ecology and Human Mobility Proceedings. https://dl.acm.org/doi/10.1145/3557921.3565542

Jin, C., Nara, A., Yang, J.-A., Tsou, M.-H. (2019). Similarity measurement on human mobility data with spatially weighted structural similarity index (SpSSIM). Transactions in GIS, 24, 104-122. https://doi.org/10.1111/tgis.12590

Ye, X., Konduri, K., Pendyala, R. M., Sana, B., & Waddell, P. (2009). A Methodology To Match Distributions Of Both Household And Person Attributes In The Generation Of Synthetic Populations. *88th Annual Meeting of the Transportation Research Board, January 11-15, 2009.*

Thank you!

Contact me:

Jessica Embury, jembury@sdsu.edu
Department of Geography, San Diego State University

Acknowledgements: This work was supported in part by the National Science Foundation (NSF) under Grant No. 2031407, CS4All project titled “Collaborative Research: Encoding Geography Scaling up an RPP to achieve inclusive geocomputational education” and the Center for Human Dynamics in the Mobile Age (HDMA) at SDSU. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or HDMA.