

An Individual-Oriented Typology of Social Areas in the United States

Joe Tuccillo

Jessica Moehl

Geospatial Sciences & Human Security Division

ACS Data Users Conference 2021

May 19, 2021

ORNL is managed by UT-Battelle, LLC for the US Department of Energy

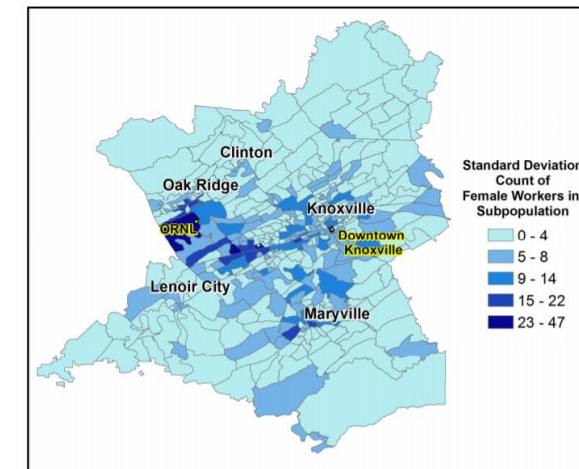
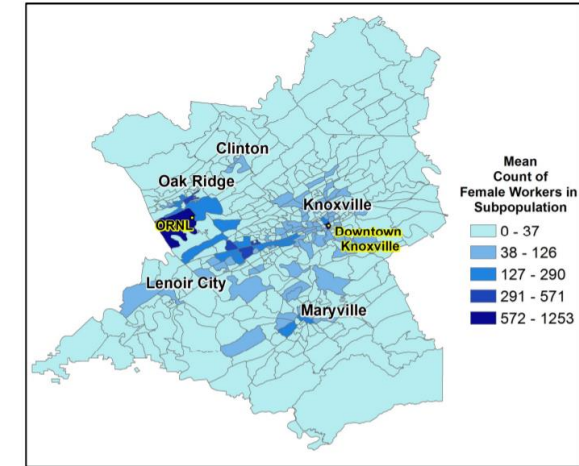
This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (<https://energy.gov/downloads/doe-public-access-plan>).



U.S. DEPARTMENT OF
ENERGY

UrbanPop: the “Backbone” of our Approach

- UrbanPop (UP) is a **spatial microsimulation framework** based on intercensal data from the American Community Survey.
 - Produces **daytime/nighttime synthetic populations**.
 - **Updated annually** (ACS 5-Year Estimates)
 - **High spatial resolution** (block groups)
 - Coverage for the **full United States** (UP2017; UP2019 in progress)
 - **Customizable demographic profiles** for specific applications (“packages”)

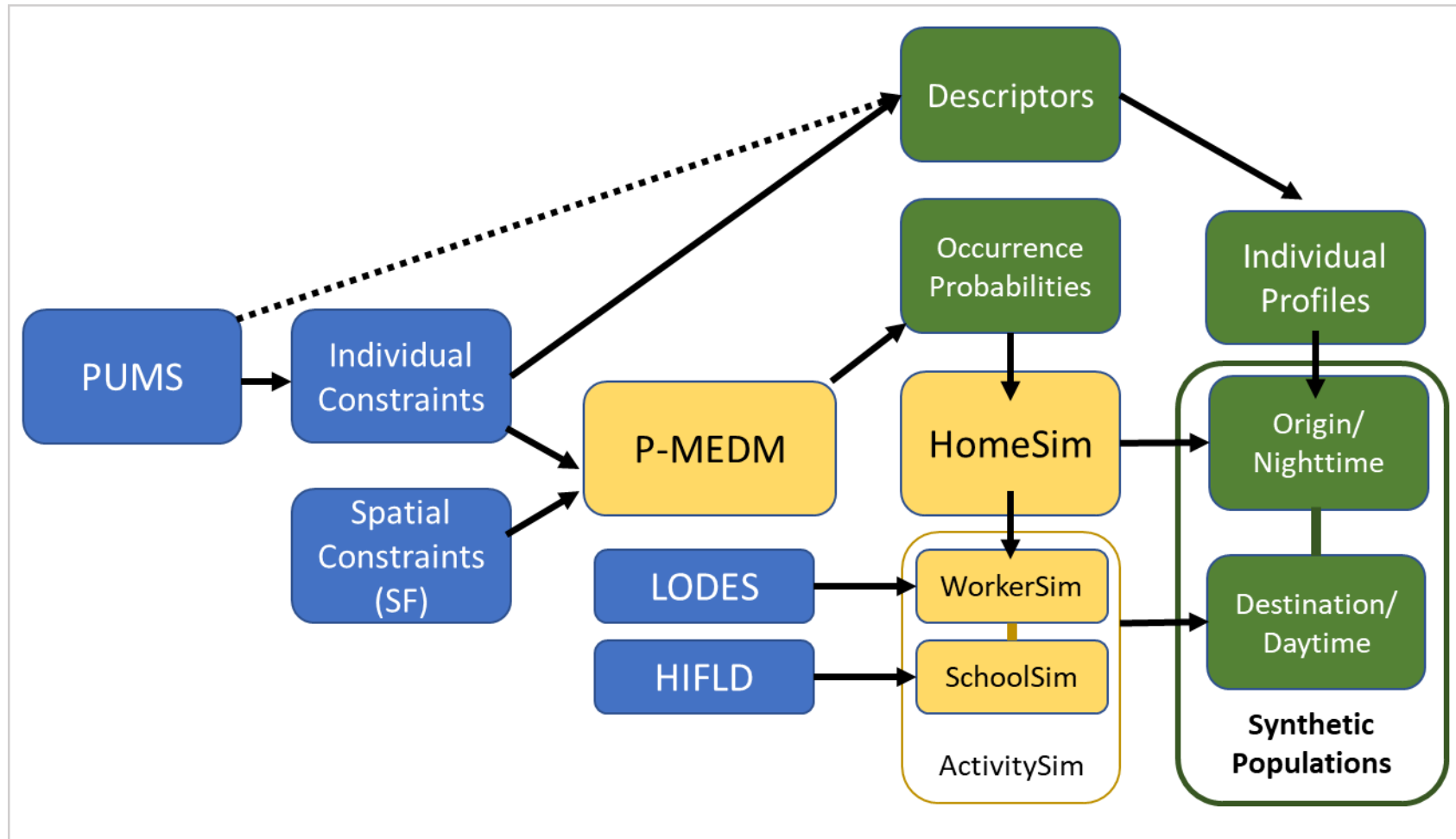


Morton et. al 2017

History of UrbanPop

- **2015:** UP begins as an effort to scale a population synthesizer for ACS data (Nagle et al 2014) and extend it to travel simulation with Pitney Bowes commute data.
- **2017-2018:** Application-specific UP packages are developed (transportation, energy).
- **2018:** UP goes open-source, swapping Pitney Bowes for LODES.
- **2020:** UP Epidemiology package is developed to aid COVID-19 situational awareness effort.
- **2021:** UPgrading to 2019 ACS/2018 LODES; development of new school travel behavior model.

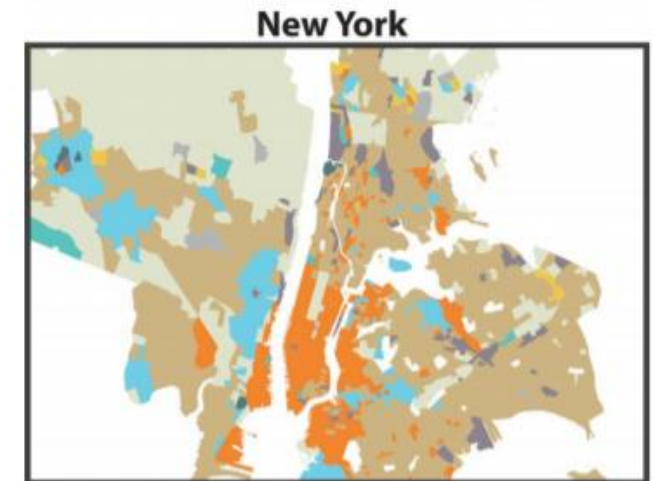
UrbanPop Core Workflow



Re-“packaging” UrbanPop

- UrbanPop synthetic populations are highly customizable but also a challenge to use out of the box.
- **High-level information on people + social mix** is needed to support planning and spatial policy interventions (i.e., hazards, transportation, energy, public health).
- One way to achieve this is to integrate UP with **geodemographics**, a framework for generalizing “where people are” + “how they live” (Harris et. al 2005)

- A: Hispanic and Kids
- B: Wealthy Nuclear Families
- C: Middle income, single family homes
- D: Native American
- E: Wealthy Urbanites
- F: Low Income and Diverse
- G: Old, Wealthy White
- H: Low Income Minority Mix
- I: African–American Adversity
- J: Residential Institutions, Young People

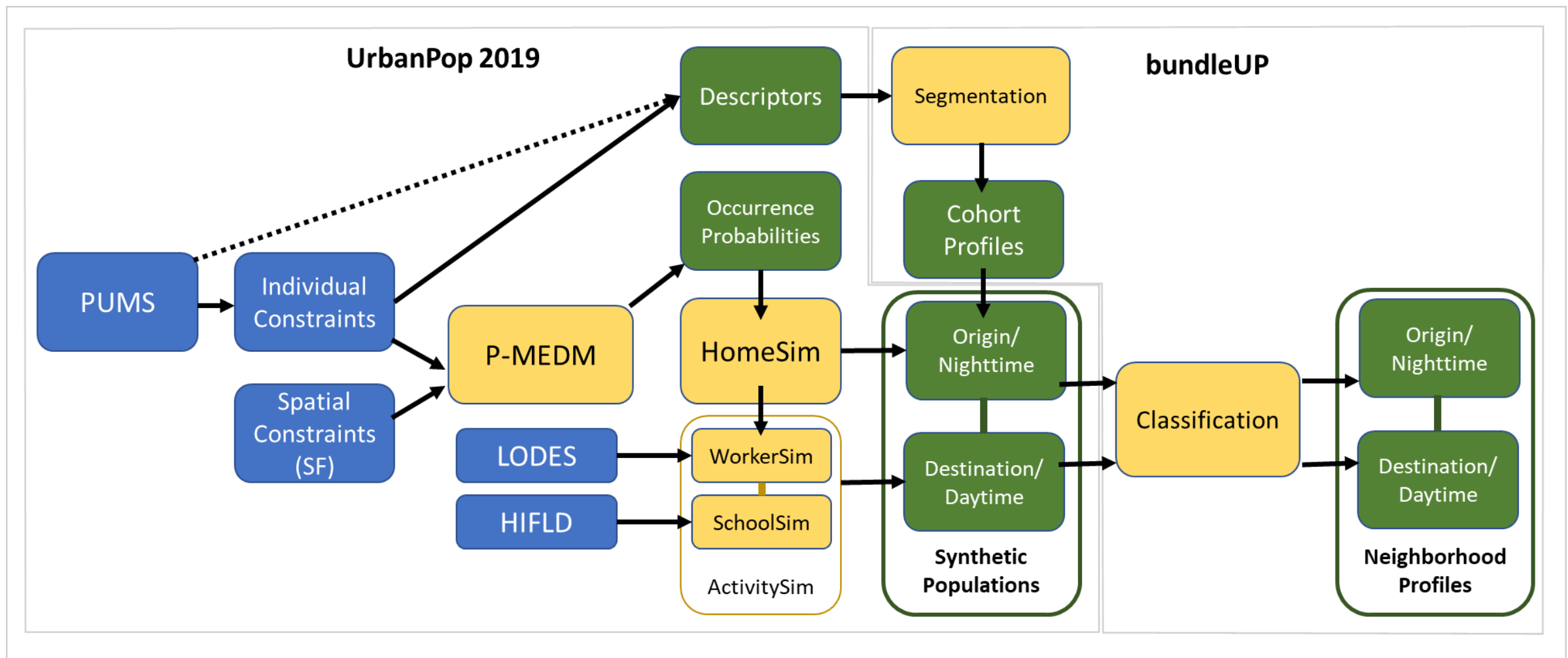


Spielman and Singleton 2015

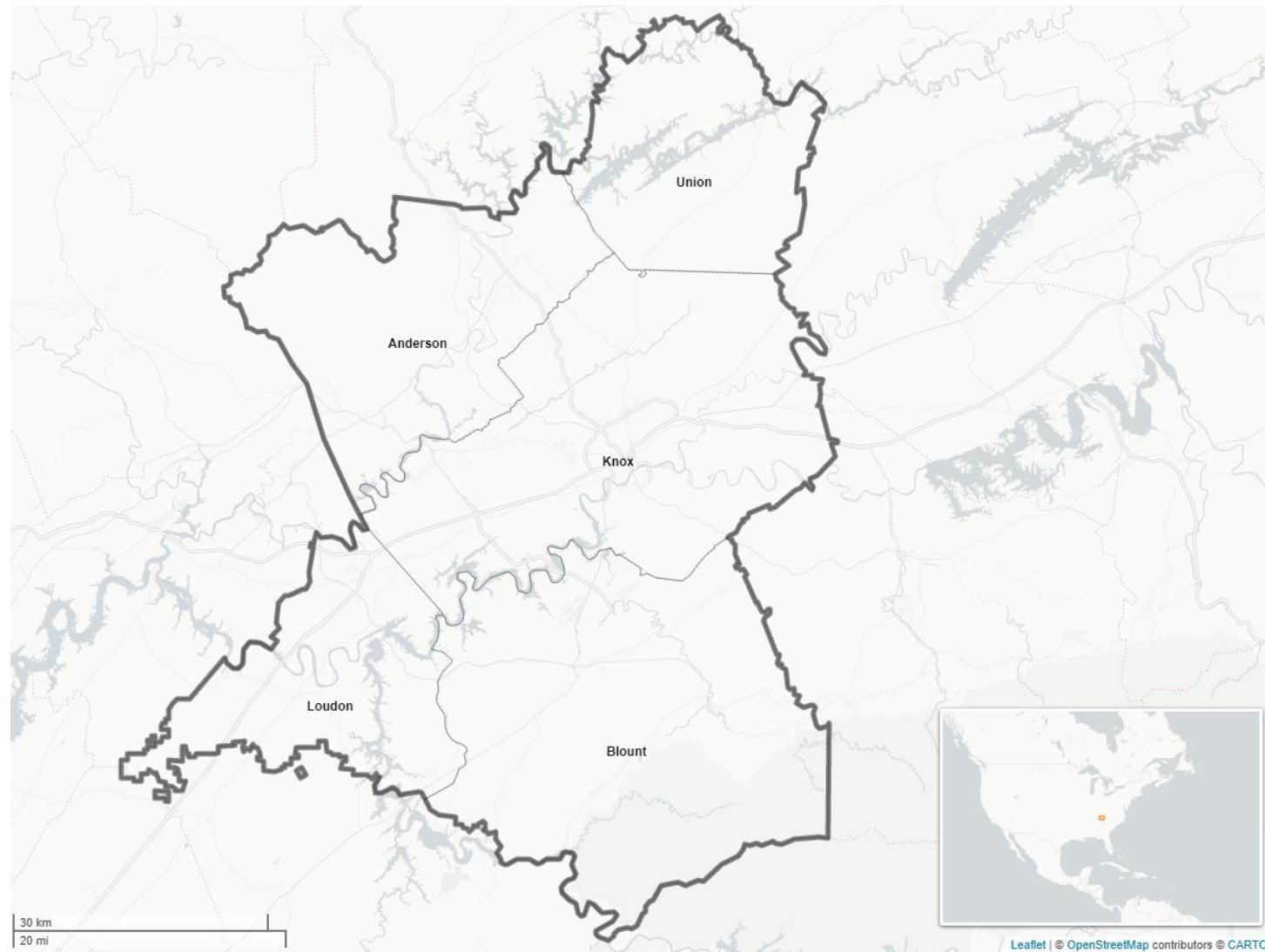
bundleUP: A Geodemographic Extension for UrbanPop

- Neighborhoods/social areas are *"bundle[s] of spatially based attributes associated with clusters of residences...produced by the same actors that consume them..."* (Galster 2001)
 - **UP + Geodemographics** → “bundling” of neighborhood actors (residents, workers, students) at varying times of day
 - Extends the notion of “how people live” + “where people live” to “how/where they interact”

Combined Workflow

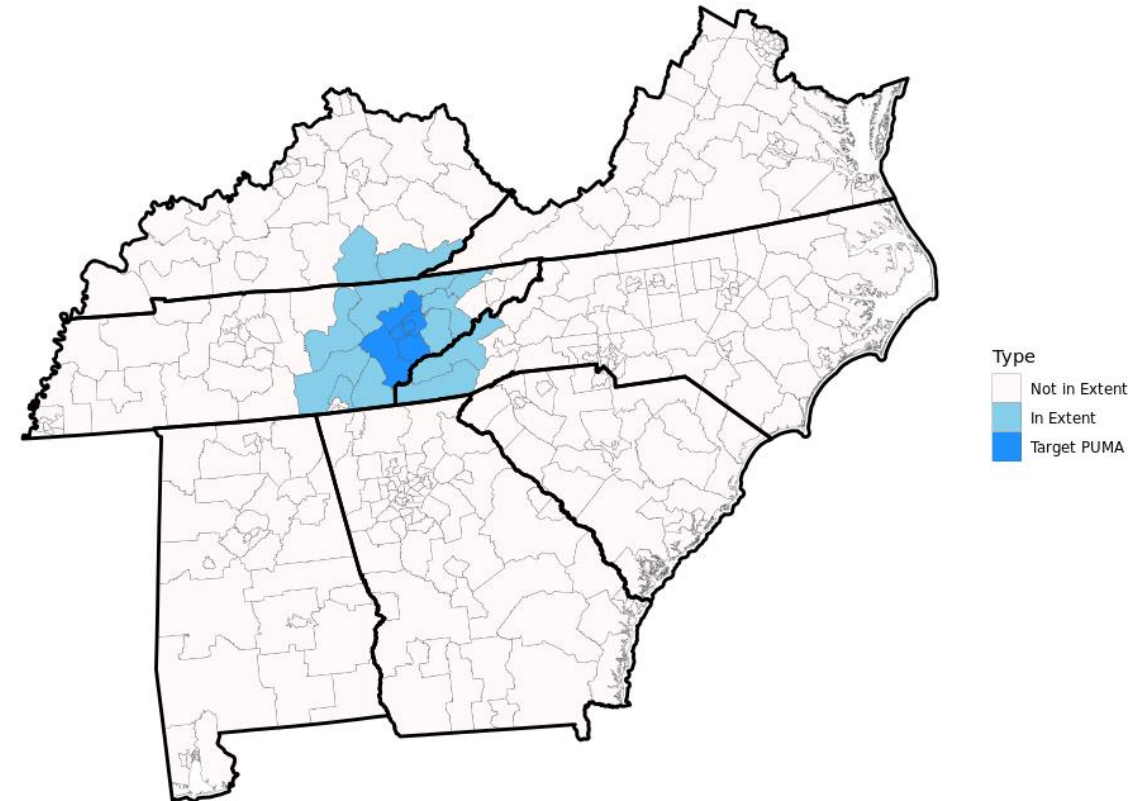


Proof of Concept: Knoxville, TN MSA



Model Design

- **“Basic Demographics, SES, and Accessibility”** mock-package
- “On the Fly” MSA-specific methodology
 - Assumption: there is a threshold distance from which regular daytime populations (workers, students) outside MSA will travel to target PUMAs
 - Include all first-order neighbors of Knox MSA PUMAs at a lag distance of 100km (n = 18)



Population Synthesis with UrbanPop

- **Penalized Maximum Entropy Dasymetric Modeling (P-MEDM)** approach (Leyk et al 2013; Nagle et al 2014).
- P-MEDM: ACS PUMS → block groups by recreating both ACS Summary File (SF) estimates *and* error variances.
- Target variables from ACS PUMS + SF are known as “constraints”.

ACS SF Constraints for P-MEDM

Table	Type	Level	Description
B01001	Universe	Person	Total Population
B09019	Universe	Person	Group Quarters Population
B25001	Universe	Household	Housing Units
B25003	Universe	Household	Occupied Housing Units
B02001	SES	Person	Race
B03003	SES	Person	Hispanic/Latino Ethnicity
B23025	SES	Person	Employment Status
B24080	SES	Person	Sex by Class of Worker
C24010	SES	Person	Sex by Occupation
C24030	SES	Person	Sex by Industry
B14007	SES	Person	Grade Level Attending
B23009	Demographics	Household	Presence of own children
B11001	Demographics	Household	Family Type
B25003	Housing	Household	Tenure
B25024	Housing	Household	Units in Structure
B25034	Housing	Household	Year Built
B08303	Mobility	Person	Commute Time to Work

Describing Individuals in the Synthetic Population

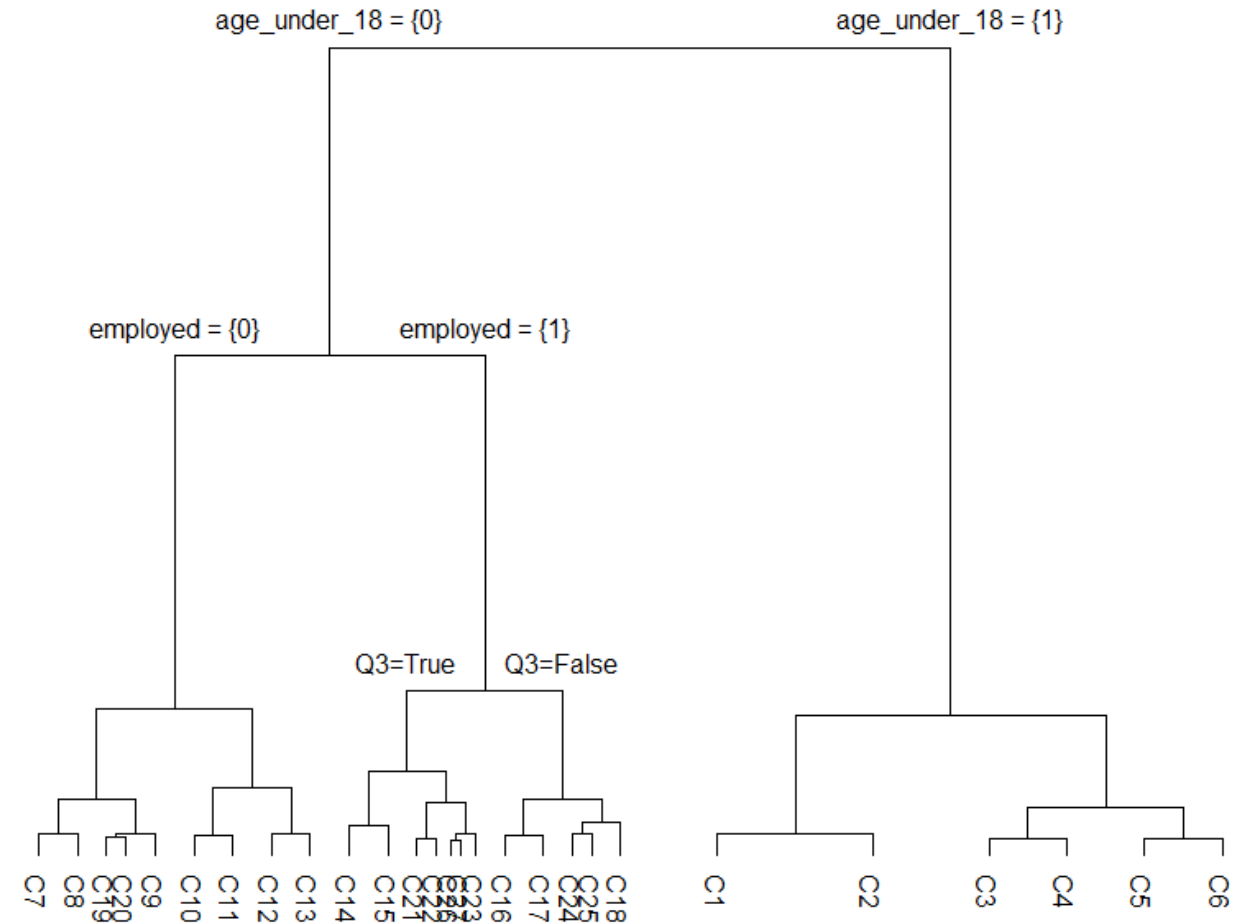
- **Descriptors** are derived from all PUMS files in the study extent (n = 18) to account for intra-MSA and extra-MSA travel on a typical day.
- **Segmentation** performed on Daytime/Nighttime synthetic populations for Knox MSA PUMAs with descriptors attached

Theme	Descriptor
Demographics	Age >=60
Demographics	Age <18
SES	White
Demographics	Married Couple HH
SES	Employed
SES	Skilled Occupation*
SES	In Poverty
SES	In School
SES	K-12 Student
SES	College/University Student
SES	Bachelor's Degree or Higher
Mobility	Vehicle Available in Household
Mobility	Commute <=30m

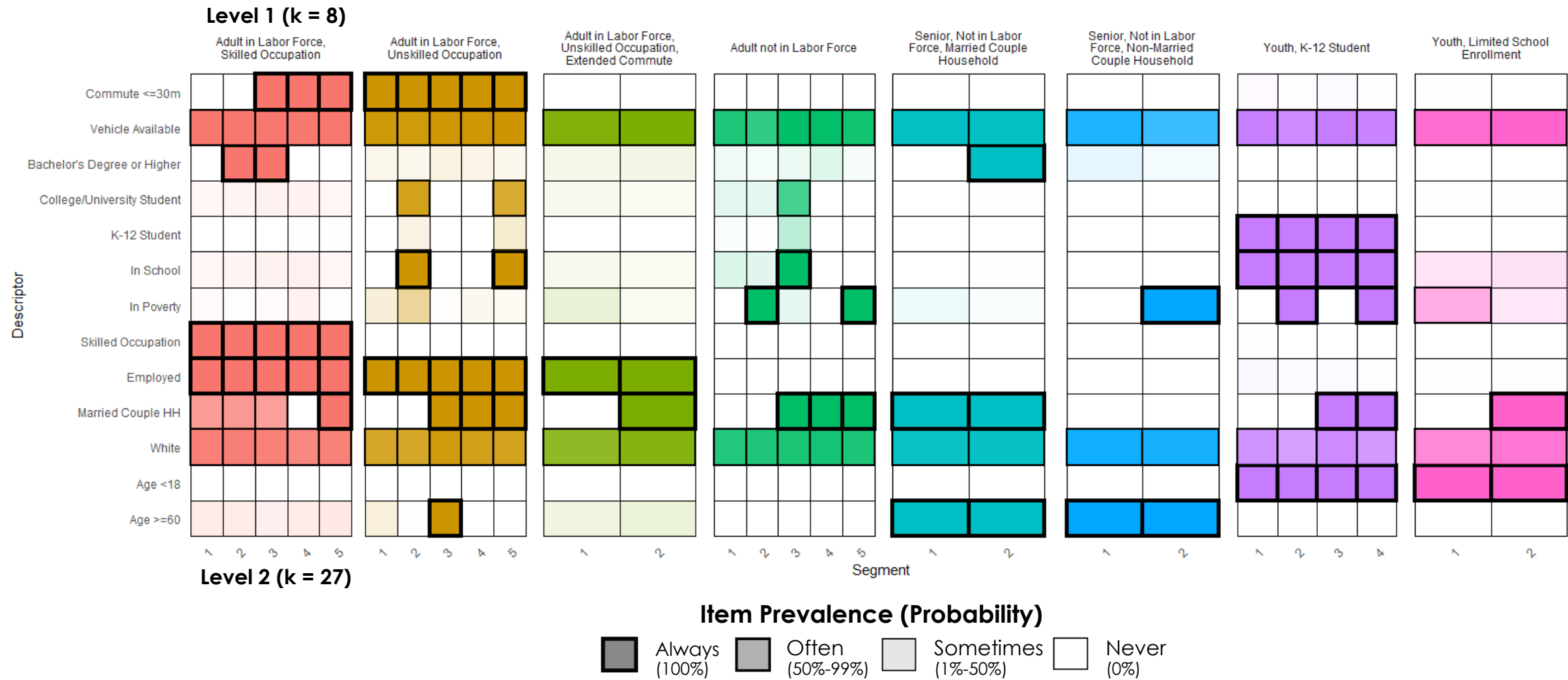
*Management, Finance, Science, Engineering, Social Services, Arts/Athletics, Medical, Law Enforcement

bundleUP: Segmentation

- Monothetic Divisive Clustering (DIVCLUST) (Chavent et al 2007) on PUMS inputs.
- DIVCLUST → Decision tree where each split is an individual attribute.
- Each Cohort Profile is characterized by the presence/absence of specific attributes.



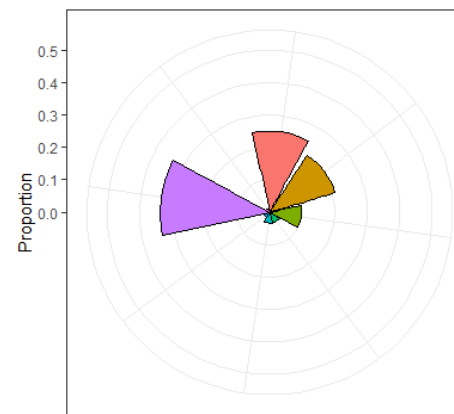
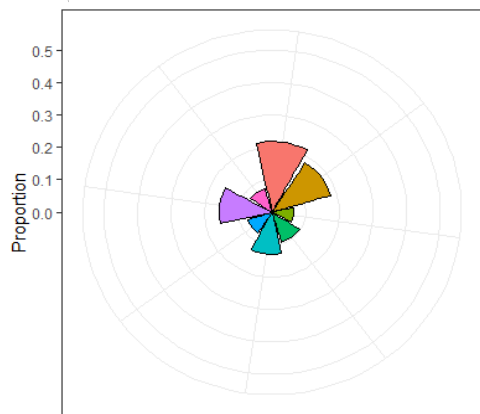
bundleUP: Cohort Profiles



Characterizing Synthetic Populations

Nighttime/
residential

Daytime



**Cohort Profiles
(Level 1)**

Adult in Labor Force,
Skilled Occupation

Adult in Labor Force,
Unskilled Occupation,
Extended Commute

Senior, Not in Labor
Force, Married Couple
Household

Youth, K-12 Student

Adult in Labor Force,
Unskilled Occupation

Adult not in Labor Force

Senior, Not in Labor
Force, Non-Married
Couple Household

Youth, Limited School
Enrollment

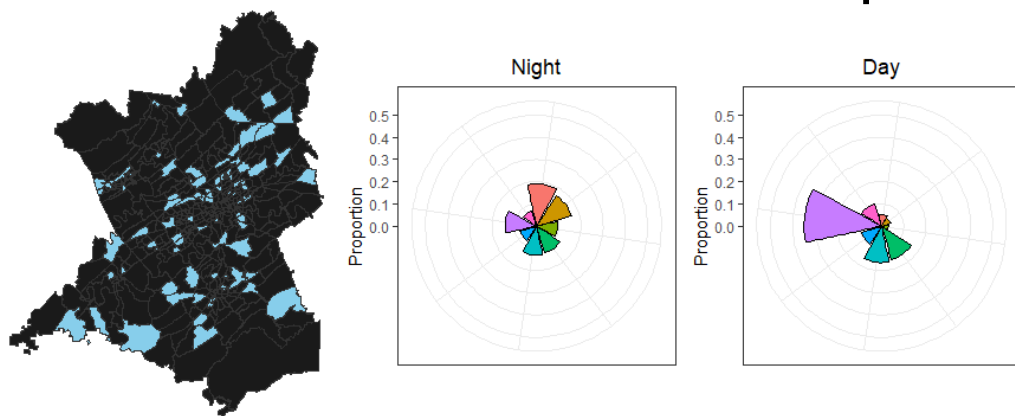
Knox County
West/Farragut Town
PUMA (4701604)

Glyph plot methodology
adapted from Lu et. al
(2014), Gollini et. al
(2015).

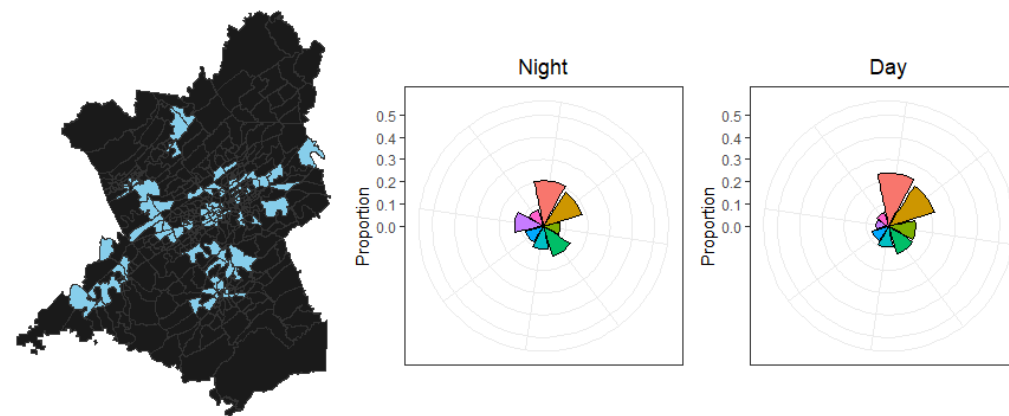
bundleUP: Block Group Typology

- High-level catalog of synthetic populations/activities.
- Aggregate (block group) level DIVCLUST with inputs:
 - Nighttime/Residential population (Cohort Profile Lv2 prevalence)
 - Nighttime/Daytime difference in Cohort Profile Lv2 prevalence
- The resulting typology accounts for:
 - Block group social mix
 - Large-scale activity patterns (i.e., worker, student commutes)

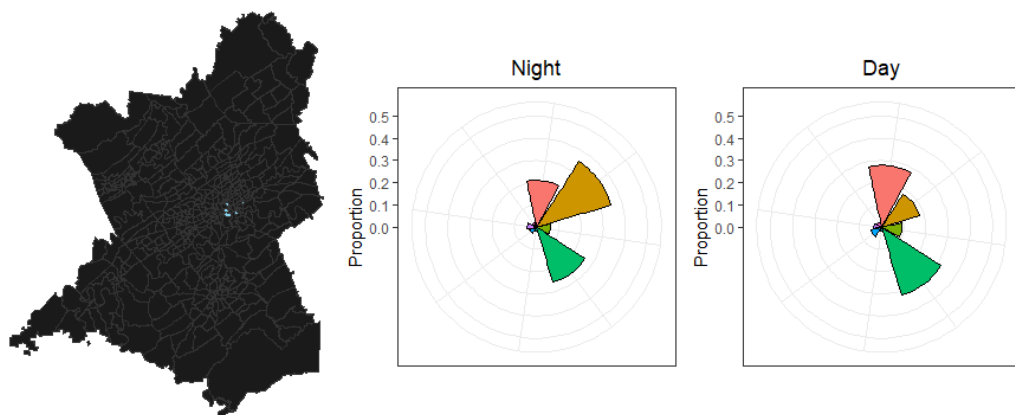
bundleUP: Block Group Profiles



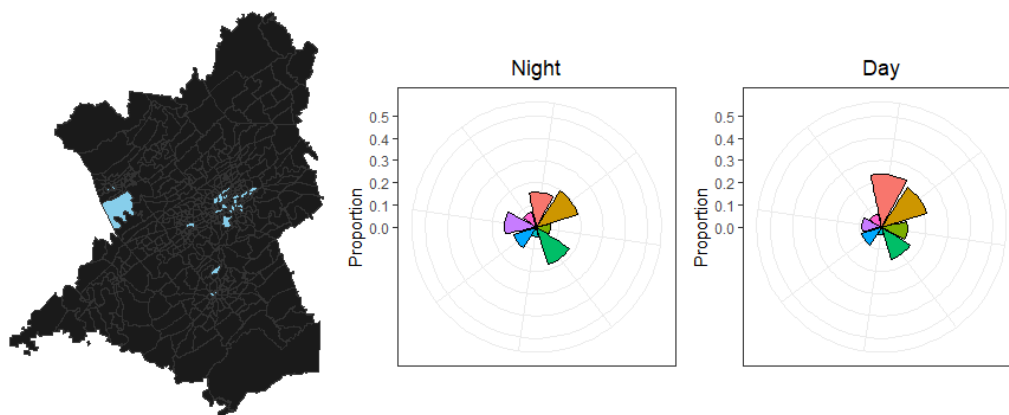
Workers leave, influx of K-12 Students, unemployed adults remain.



Moderate influx of skilled workers, high influx of unskilled workers; K-12 students leave.

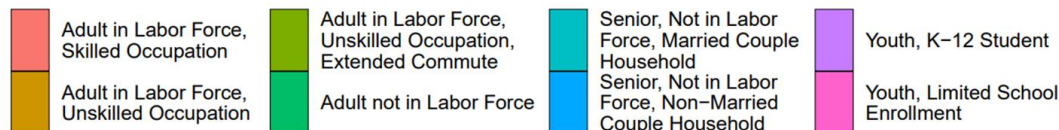


University campus areas: influx of skilled workers and unemployed adults (students).



High influx of skilled workers, unskilled workers with extended commutes.

Cohort Profiles (Level 1)



Discussion

- Emergence provides “a way of capturing important features of a system using far less information than an exhaustive description would entail...and still say[ing] quite a lot about [that system].” (Carroll 2019)
 - **UP + bundleUP** → geodemographics as the study of emergent “states” of social areas comprising an urban/regional system, with respect to a condition/process of interest.
 - Orients tasks like planning, spatial policy interventions toward critical needs of specific groups by time of day.

Outlook

- **Near term:**

- Tailor-made profiles for application-specific UP packages (environmental hazards, health and place, mobility/energy use)
- Deep learning/AI to catalog people and places at scale

- **Mid term:**

- Public-facing tools (web portal, API)

- **On the horizon:**

- UP/bundleUP episodic layers with point of interest (POI), time-use data

Thank You!

Questions?

References

1. Morton, April M., Piburn, Jesse O., Nagle, Nicholas N., Aziz, H M Abdul, Duchscherer, Samantha E., & Stewart, Robert N. A Simulation Approach for Modeling High-Resolution Daytime Commuter Travel Flows and Distributions of Worker Subpopulations. United States.
2. Nagle, N. N., Battenfield, B. P., Leyk, S., & Spielman, S. (2014). Dasymetric modeling and uncertainty. *Annals of the Association of American Geographers*, 104(1), 80-95.
3. Harris, R., Sleight, P., & Webber, R. (2005). *Geodemographics, GIS and neighbourhood targeting* (Vol. 8). John Wiley & Sons.
4. Spielman, S. E., & Singleton, A. (2015). Studying neighborhoods using uncertain data from the American community survey: a contextual approach. *Annals of the Association of American Geographers*, 105(5), 1003-1025.
5. Galster, G. (2001). On the nature of neighbourhood. *Urban studies*, 38(12), 2111-2124.
6. Leyk, S., Nagle, N. N., & Battenfield, B. P. (2013). Maximum entropy dasymetric modeling for demographic small area estimation. *Geographical Analysis*, 45(3), 285-306.
7. Chavent, M., Lechevallier, Y., & Briant, O. (2007). DIVCLUS-T: A monothetic divisive hierarchical clustering method. *Computational Statistics & Data Analysis*, 52(2), 687-701.
8. Isabella Gollini, Binbin Lu, Martin Charlton, Christopher Brunsdon, Paul Harris (2015). GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software*, 63(17), 1-50. URL <http://www.jstatsoft.org/v63/i17/>.
9. Binbin Lu, Paul Harris, Martin Charlton, Christopher Brunsdon (2014). The GWmodel R package: further topics for exploring spatial heterogeneity using geographically weighted models. *Geo-spatial Information Science*, 17(2), 85-101. URL <https://www.tandfonline.com/doi/abs/10.1080/10095020.2014.917453>
10. Carroll, S. (2019). *Something deeply hidden: quantum worlds and the emergence of spacetime*. Dutton.

Acknowledgements

April Morton, Robert Stewart, Amy Rose, Nathan Trombley

Contact

tuccillojv@ornl.gov