

Using PostGIS to Manage the CTPP Dataset

With an Application to Mammography Access Research

JULIA WOLANSKI, PROFESSIONAL SCIENCE MASTER'S in GIS

LEE HACHADOORIAN, DEPARTMENT of GEOGRAPHY & URBAN STUDIES
TEMPLE UNIVERSITY

Background

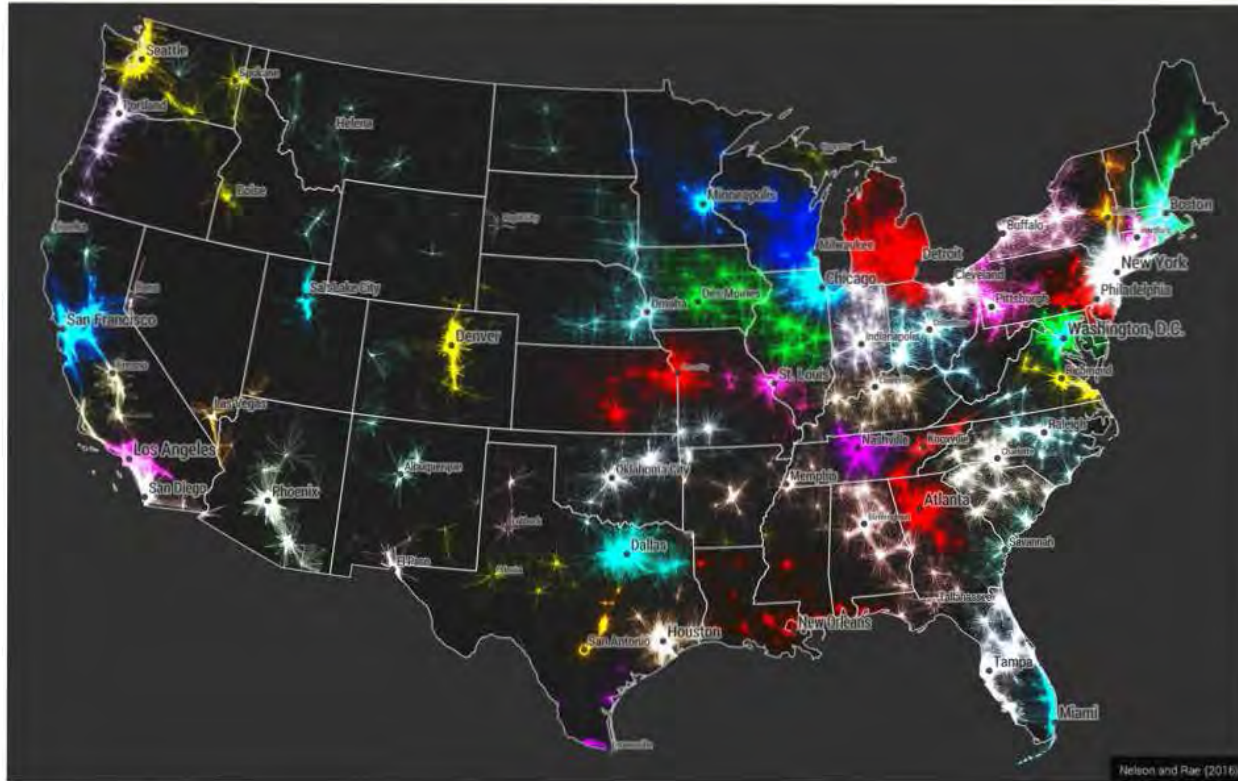
census-postgres

The screenshot shows the GitHub repository page for `leehach/census-postgres`. The browser tabs include "Inbox (3) - tug30201@tr...", "CTPP IREG Presentation", and "GitHub - leehach/census-postgres". The address bar shows the URL `https://github.com/leehach/census-postgres`. The repository page features a navigation bar with links for Features, Business, Explore, Marketplace, and Pricing. Below the repository name, there are statistics: 14 Watchers, 96 Stars, and 39 Forks. The repository description is "PostgreSQL schema and import scripts for recent US Census data". The commit history shows 41 commits, 1 branch, 0 releases, and 1 contributor. The latest commit by leehach is titled "Updated acs2011_5yr scripts to create sequences with geoid field and ..." and is dated May 22, 2014. A table of files and their commit dates is provided below.

File	Description	Time
acs2010_5yr	Added single view which returns all tracts in MSAs with associated MS...	3 years ago
acs2010_spt	Added state-specific import scripts to acs2010_5yr; Cleanup of delete...	5 years ago
acs2011_5yr	Updated acs2011_5yr scripts to create sequences with geoid field and ...	3 years ago
acs2012_1yr	Added data dictionary of ACS 2012 1-year, will allow meta-script import.	4 years ago



Commuter Megaregions of the United States



Source: Nelson and Rae, 2016



College of Liberal Arts

Department of Geography and Urban Studies

PROFESSIONAL SCIENCE MASTER'S IN GEOGRAPHIC INFORMATION SYSTEMS

- Professional degree, designed to train highly competent technicians ready to join the workforce
- Can be completed in one calendar year of full-time coursework, or two academic years part-time
- Curriculum includes business ethics, and professional development
- Prepares students for GISP Certification

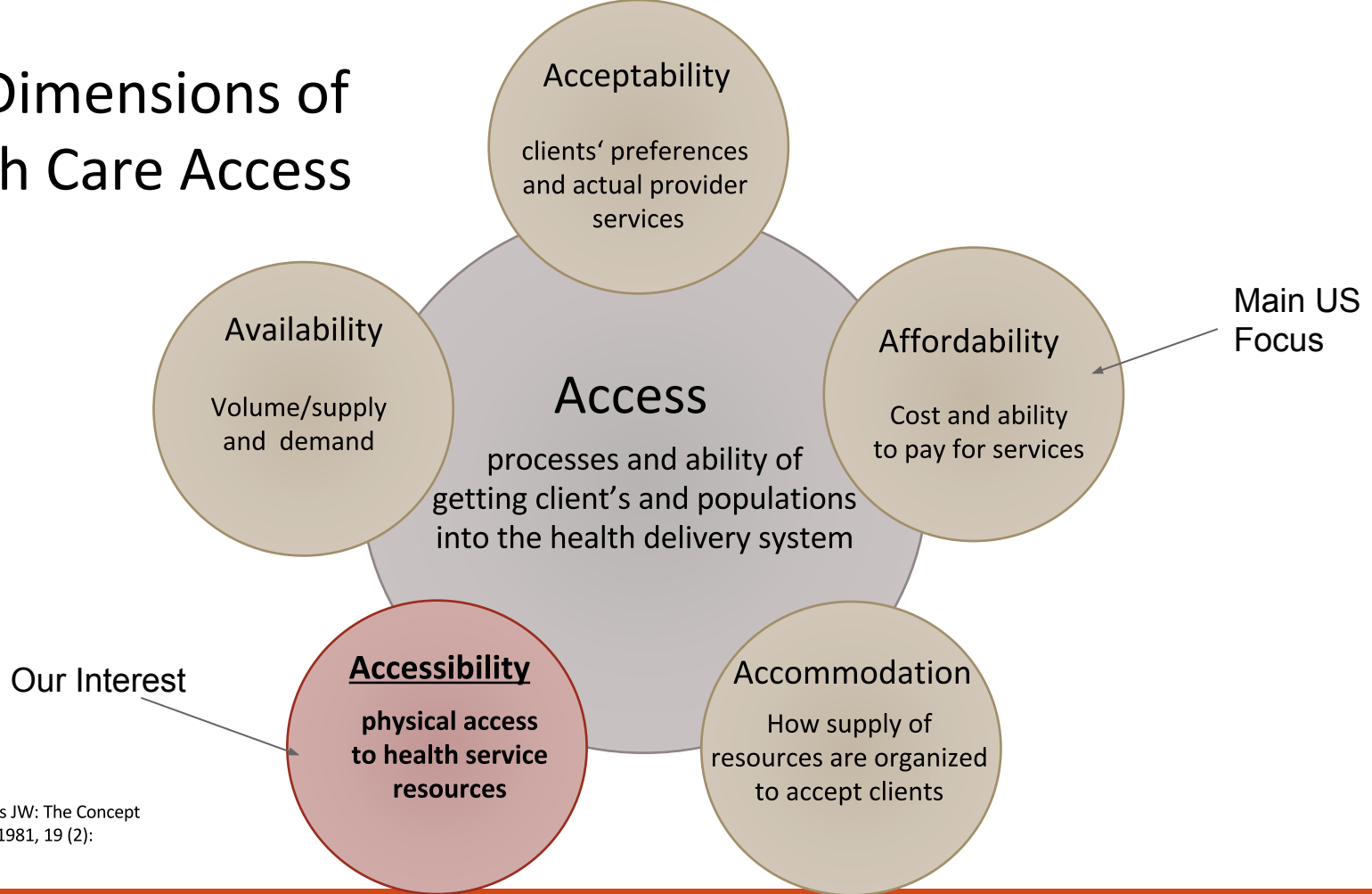
GUS 8067: Spatial Database Design

- PSM-required course
- Culminates in data management project using real-world dataset



Research Project Overview

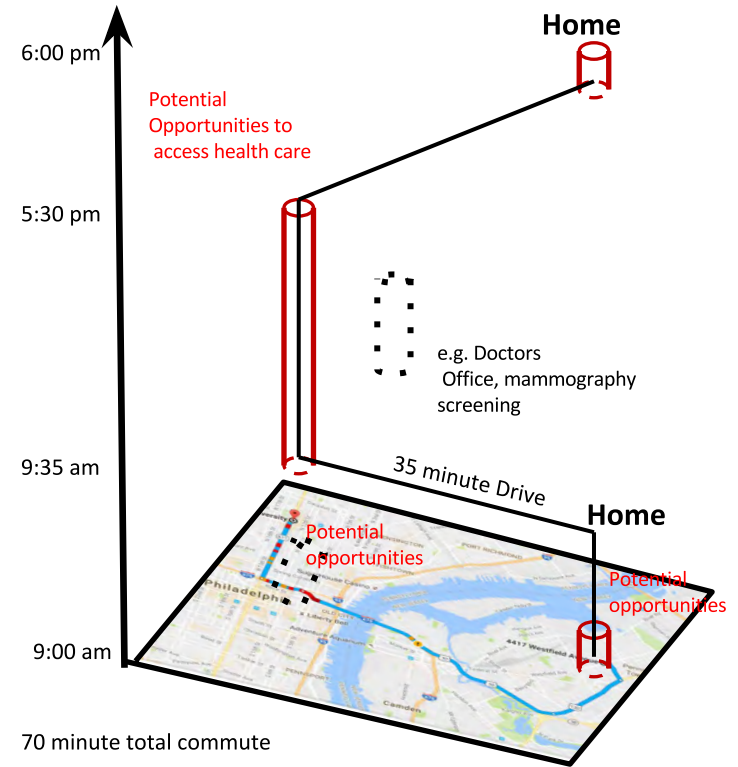
Five Dimensions of Health Care Access



Penchansky R, Thomas JW: The Concept of Access. Med Care. 1981, 19 (2): 127-140.

Geographic Access to Healthcare

- The majority of studies examining population-based geographic access to health care utilize census residential populations
 - Only night populations considered
- For some health services it is important to also consider geographic access based on where populations are during the day.
 - Preventative Care and Screening



There are potential opportunities to access health care services from both home and work

Study Questions

- Question 1:
 - Are there differences in travel times and geographic access measures (i.e. supply vs. demand) to mammography screening facilities for residential (night) and daytime populations among women in Pennsylvania?
- Question 2:
 - Are there differences in travel time to the closest mammography screening facility from the origin (residence) versus destination (work, residence, or same tract)?

CTPP Project Work

Census Transportation Planning Product

Part 1: Residence

Part 2: Workplace

Part 3: Flows

CTPP Subjects

Industry

Occupation

Earnings

Demographics: Race, Hispanic Origin, Sex, Age

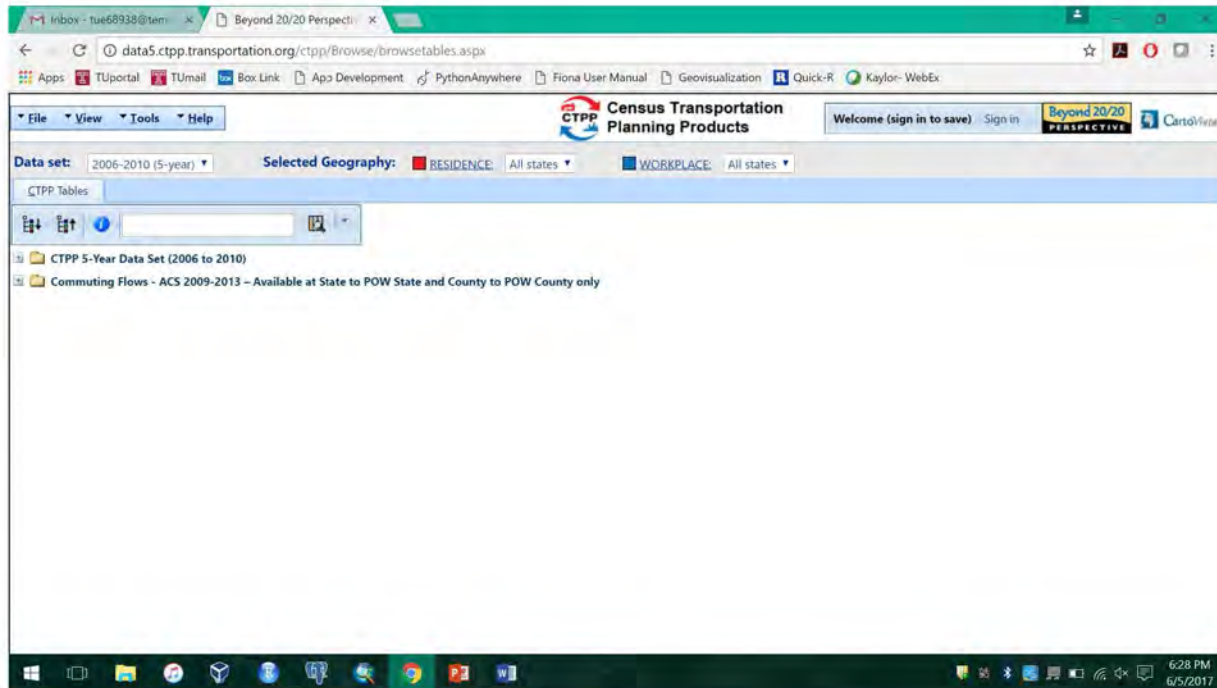
Means of Transportation

Time leaving home

Time arriving

Travel time

Original Website



Data Output

data5.cttp.transportation.org/cttp/view/disview.aspx?ReportId=250

Apps TUportal TEmail Box Link App Development PythonAnywhere Fiona User Manual Geovisualization Quick-R Kaylor- WebEx

File View Help

Census Transportation Planning Products

Welcome (sign in to save) Sign in Beyond 20/20 PERSPECTIVE Cartovista

Data set: 2006-2010 (5-year) Selected Geography: RESIDENCE: All states WORKPLACE: All states

View as: Table Chart Map

Show CTPP Tables

A116200 - Employment status (7) by Sex (3) (Persons 16 years and over)

Current date: 6/5/2017 6:29:53 PM (Eastern Daylight Time)
U.S. Census Bureau, American Community Survey 2006-2010 Five-year estimates. Special Tabulation: Census Transportation Planning

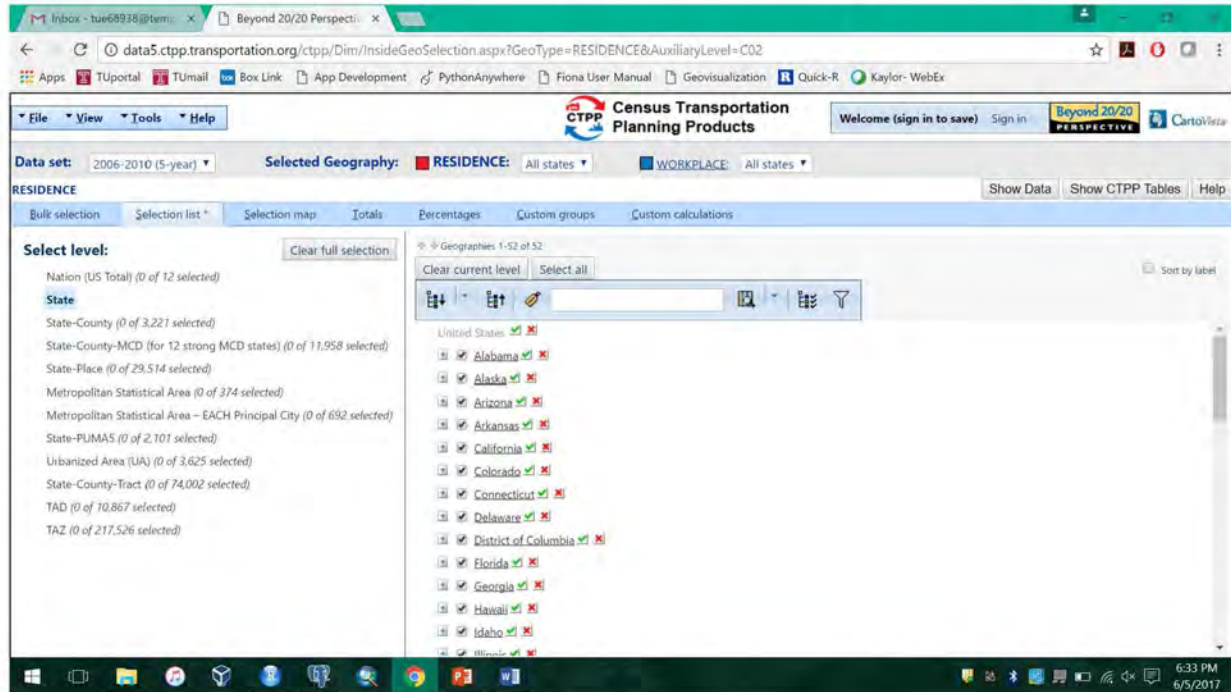
Rows 1-52 of 52 Columns 1-42 of 42

Measures: Population 16 and Over Drag dimensions here so they do not show as a row or column in table

Output	Total employment status						Civilian Employed - At Work					
	Both sexes		Male		Female		Both sexes		Male		Female	
	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error	Estimate	Margin of Error
RESIDENCE												
Alabama	3,714,505	1,589	1,775,410	1,482	1,939,095	1,294	1,984,555	6,106	1,054,515	3,998	930,035	
Alaska	528,190	575	274,420	652	253,770	630	317,050	2,231	168,805	1,639	148,245	
Arizona	4,813,495	1,503	2,375,570	1,335	2,437,925	1,156	2,679,750	7,572	1,451,010	5,118	1,228,740	
Arkansas	2,247,625	1,177	1,088,165	1,139	1,159,460	1,047	1,223,075	5,006	644,630	3,401	578,445	
California	28,445,585	4,159	14,033,980	3,021	14,411,605	3,533	16,129,860	18,556	8,818,780	11,094	7,311,080	
Colorado	3,820,290	1,699	1,903,040	1,361	1,917,250	1,188	2,399,365	6,468	1,296,995	4,188	1,102,365	
Connecticut	2,820,835	1,327	1,353,275	976	1,467,560	1,012	1,717,300	5,308	892,025	3,567	825,275	
Delaware	700,375	764	334,650	472	365,725	530	408,290	2,982	208,745	1,949	199,545	
District of Columbia	493,400	396	230,105	363	263,300	290	290,275	2,577	141,225	1,573	149,050	

6:30 PM 6/5/2017

Cleaning and Downloading Data



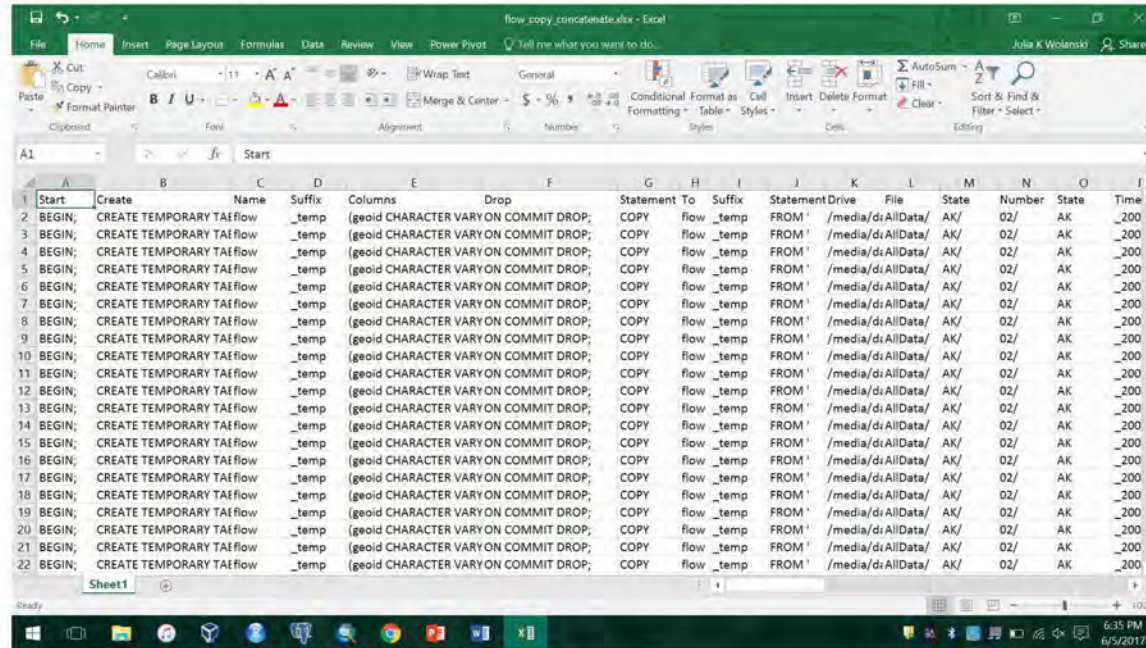
ETL Procedures

1. Download files—1 per subject table per state = $342 \times 53 = 18+$ thousand files
2. Bulk load into TEMP tables
3. Parse GEOID
 - a. Extract summary level
 - i. Parts 1 & 2: geographic level, such as state, county, TAZ, tract
 - ii. Part 3: flow between geographic levels, can include county→county AND county→tract
 - b. Extract geographic identifier, e.g., state FIPS, county fips
 - i. Part 3: split into separate residence and workplace geographic identifier columns
4. Clean up large number of nonnumeric values in population and flow counts, such as dashes, asterisks (for footnotes), or topcoded values

Destination Table Structure

- Summary Level
- Geoid (Residence and Workplace for Flow datasets)
- Subject Table
- Line Number (“variable”)
- Estimate
- Standard Error

Metacoding: SQL Generation in Excel*



Start	Create	Name	Suffix	Columns	Drop	Statement	To	Suffix	Statement Drive	File	State	Number	State	Time
2	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
3	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
4	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
5	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
6	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
7	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
8	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
9	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
10	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
11	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
12	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
13	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
14	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
15	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
16	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
17	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
18	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
19	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
20	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
21	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200
22	BEGIN;	CREATE TEMPORARY TABLE	_temp	(geoid CHARACTER VARY ON COMMIT DROP;		COPY	flow _temp		FROM ^	/media/ds/AllData/	AK/	02/	AK	_200

* Python is the future

Summary Table

Ed4 Data - Kropotkin Julia (kropotkin@temple.edu) - sumlevel - cpg.sumlevel

File Edit View Tools Help

100 rows

	sumlevel [PK] character varying	descr character varying	geoid_variable character varying	format character varying	res_dig integer	work_dig integer
14	C22	CTPP Place of Work - Place	st	C2300USasccc		
15	C23	CTPP Place of Work - State-County	st/cty	C2300USasccc		
16	C24	CTPP Place of Work - County Subdivision	st/cty/mcd	C2400USasccccmmmm		
17	C25	CTPP Place of Work - State-Place	st/pl	C2500USasppppp		
18	C27	CTPP Place of Work - Metropolitan Statistical Area	cbasa	C2700USbbbbbb		
19	C28	CTPP Place of Work - Metropolitan Statistical Area	cbasa/st/pl	C2800USbbbbbsppppp		
20	C29	CTPP Place of Work - State-POW Public Use Microdata	st/powpuma	C2900USasuuuuu		
21	C30	CTPP Place of Work - Outside United States	power	C3000USr		
22	C31	CTPP- State-County-Tract	st/cty/tr	C3100USasccctttttt		
23	C32	CTPP- MPO-TAD	mpo/tad	C3200USoooooooooooo		
24	C33	CTPP- State-County-TAZ	st/cty/taz	C3300USascccczzzzzz		
25	C42	CTPP Flow - State-to-State	st/st	C4200USasss	2	2
26	C43	CTPP Flow - County-to-County	st/cty/st/cty	C4300USascccccccc	5	5
27	C44	CTPP Flow - MCD-to-MCD	st/cty/mcd/st/cty/mcd	C4400USasccccmmmmasccccmmmm	10	10
28	C45	CTPP Flow - Place-to-Place	st/pl/st/pl	C4500USaspppppsppppp	7	7
29	C48	CTPP Flow - Metro Area-State-Principal City- to M	cbasa/st/pl/cbasa/st/pl	C4800USbbbbbspppppbbbbbsppppp	12	12
30	C49	CTPP Flow - PUMA- to -POW PUMA	st/puma5/st/powpuma	C4900USasuuuuuassuuuuu	7	7
31	C51	CTPP Flow - County-to-Place	st/cty/st/pl	C5100USasccccasppppp	5	7
32	C52	CTPP Flow - MCD-to-Place	st/cty/mcd/st/pl	C5200USasccccmmmmasppppp	10	7
33	C53	CTPP Flow - PUMA-to-Place	st/puma5/st/pl	C5300USasuuuuuassppppp	7	7
34	C54	CTPP Flow- Tract-to-Tract	st/cty/tr/st/cty/tr	C5400USasccctttttttasccctttttt	11	11
35	C55	CTPP Flow- TAD-to-TAD	mpo/tad/mpo/tad	C5500USoooooooooooo	16	16
36	C56	CTPP Flow- TAZ-to-TAZ	st/cty/taz/st/cty/taz	C5600USascccczzzzzzascccczzzzzz	13	13
37	C57	CTPP Flow- TAD-to-TAZ	mpo/tad/st/cty/taz	C5700USoooooooooooo	16	13
38	C58	CTPP Flow- TAZ-to-TAD	st/cty/taz/mpo/tad	C5800USascccczzzzzzasccccccccoooo	13	16
39	C59	CTPP Flow-	st/cty/taz/st/pl	C6000USascccczzzzzzasppppp	13	7
*						

39 rows.

6:39 PM
6/5/2017

Integration of CTPP Data into Research

Study Questions

- Question 1:
 - Are there differences in travel times and geographic access measures (i.e. supply vs. demand) to mammography screening facilities for residential (night) and daytime populations among women in Pennsylvania?
- Question 2:
 - Are there differences in travel time to the closest mammography screening facility from the origin (residence) versus destination (work, residence, or same tract)?

Travel Time to Mammography Facilities

- Estimated travel time from each census tract centroids to the nearest FDA certified mammography facility (for daytime and residence populations)
 - Average travel times to the two and three closest facilities
- Mean travel time estimates are population weighted based on women 40-70 years of age
- Network Analyst, ESRI, NAVTEQ Street Networks (2010)

Enhanced Two - Step Floating Catchment Area (E2SFCA)

- Luo and Qi, 2009
- CTPP Census Populations
- FDA Certified Mammography Facilities, # of machines per facility
- 2010 Census Tract Files
- Network Analyst, ESRI Street Networks (2010)

$$A_i^F = \sum_{j \in (d_{kj} \in D_r)} R_j W_r = \sum_{j \in (d_{ij} \in D_r)} \left(\frac{S_j}{\sum_{k \in (d_{kj} \in D_r)} P_k W_r} \right)$$

Question 1

Daytime Populations

- Census Transportation Planning Project (CTPP) Tables
 - A202111F: Sex of Workers, Ages 16+
 - A116200F: Employment Status by Sex, Ages 16+
- Day Population Calculation = $A202111 F + (A116200F - A116200L)$
- Populations proportionally adjusted for women of ages 40 - 70

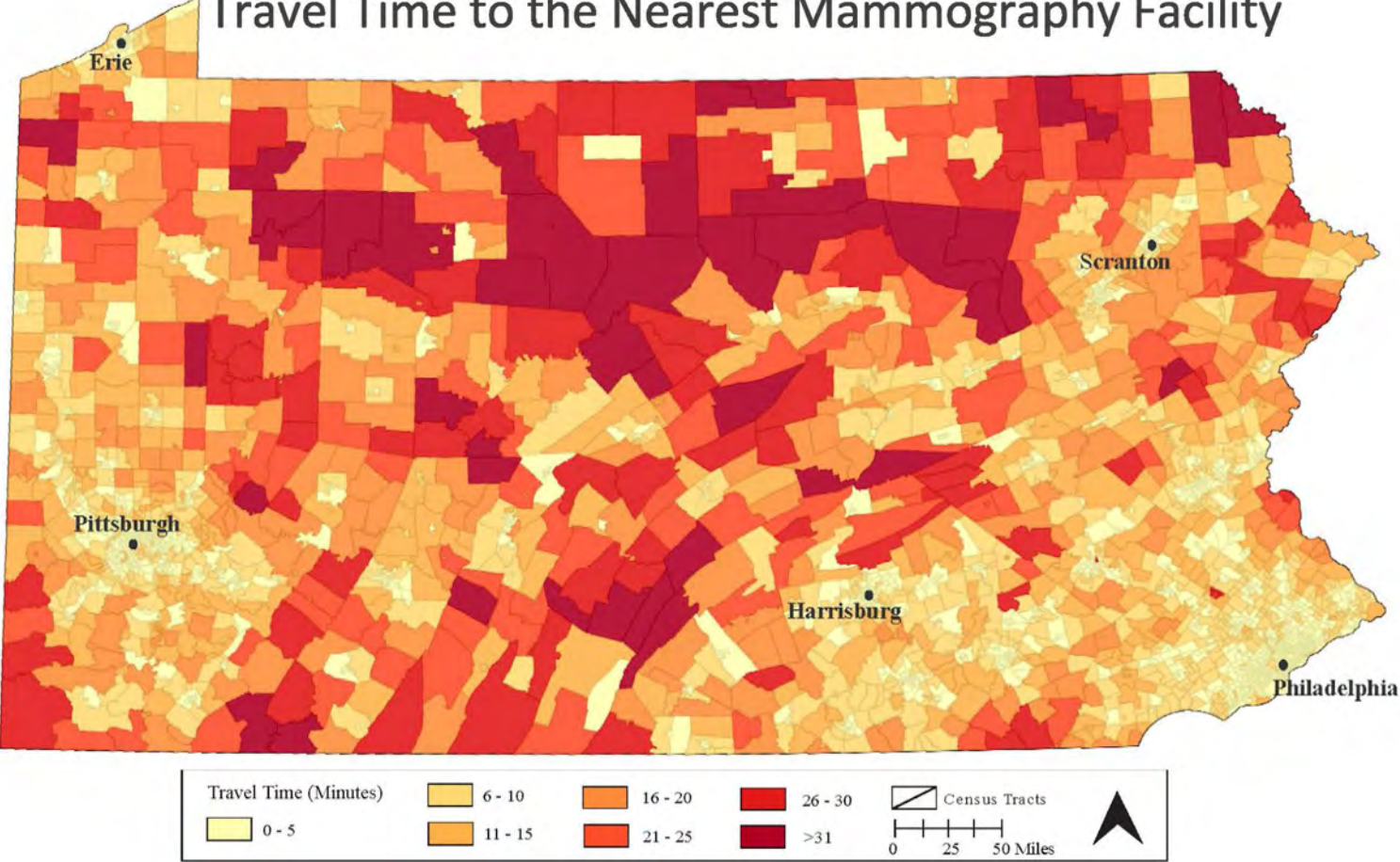
A202111 F	A116200 F	A116200 L
Females working in tracts	Total tract Population	Females at work (in another tract)

Results

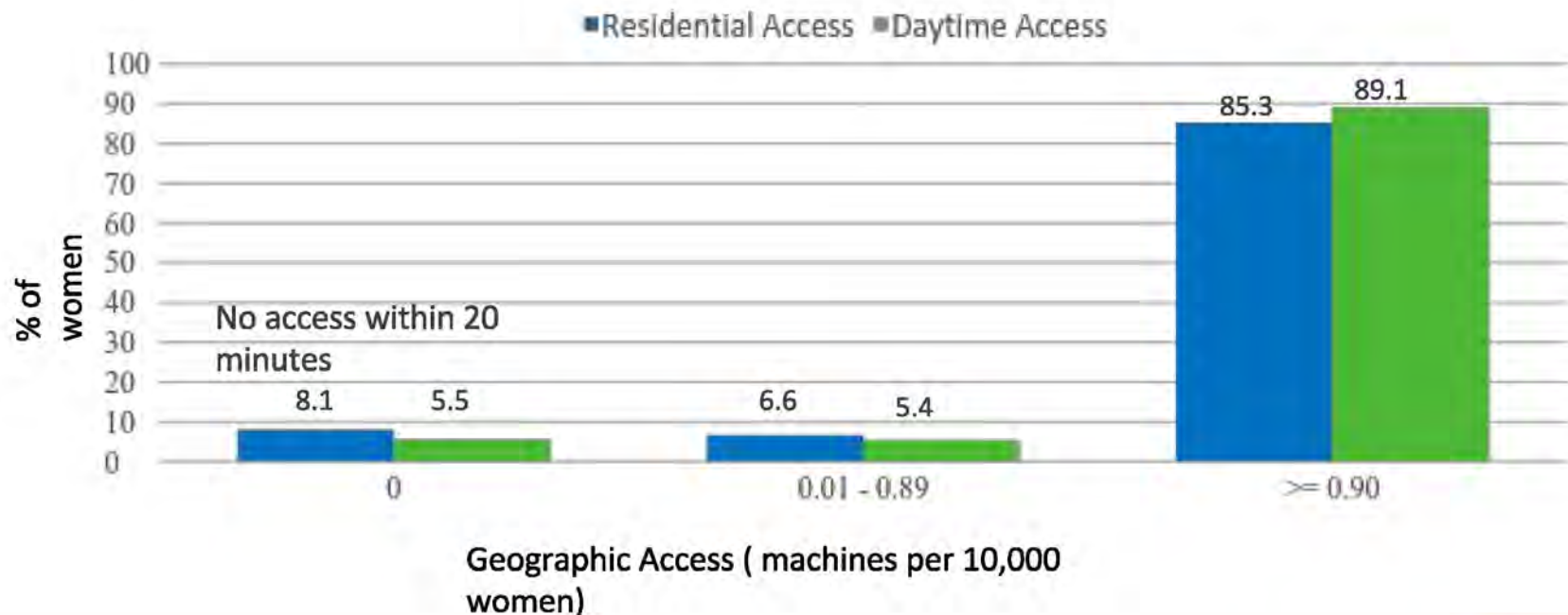
(publication submission forthcoming)

Question 1

Travel Time to the Nearest Mammography Facility

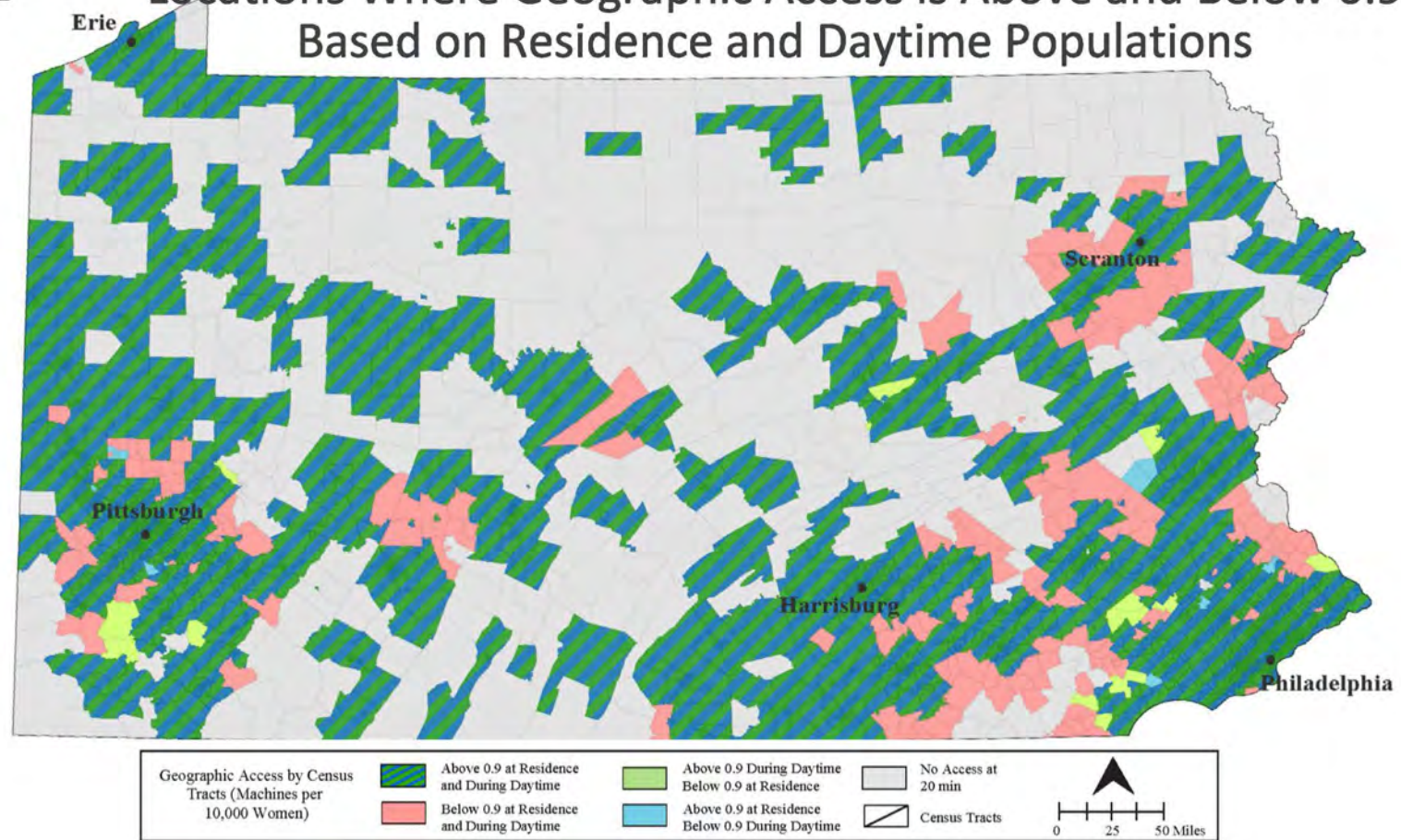


Distribution of Residential and Daytime Access



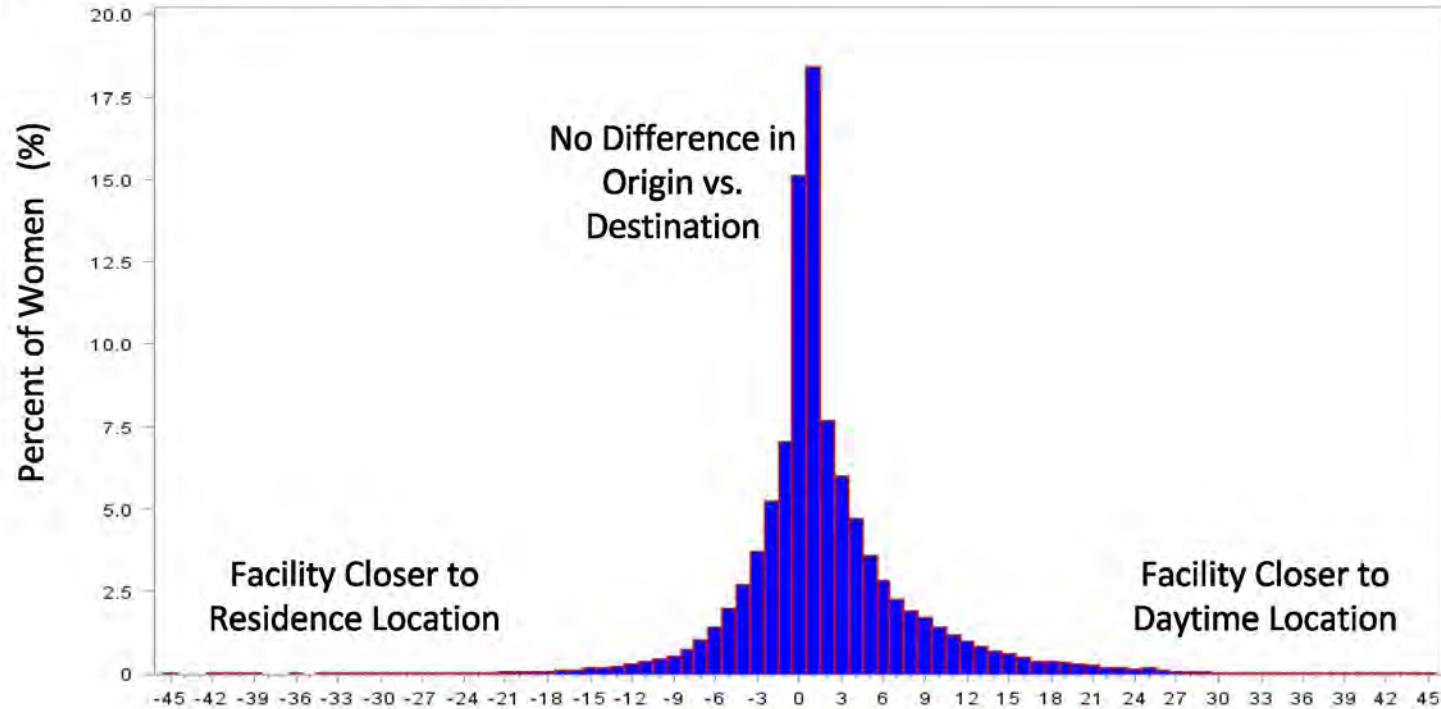
Question 1

Locations Where Geographic Access is Above and Below 0.9 Based on Residence and Daytime Populations



Question 2

Difference between Residence and Daytime Travel Times (Working Populations Only)



Discussion / Conclusion

- Question 1: Are there differences in travel times and geographic access measures to mammography screening facilities for residential and daytime populations among women in Pennsylvania?
 - >**90%** of both populations are **20** or less minutes to nearest facility
 - Travel times between both populations and nearest facilities **do not** vary significantly
 - **Minimal** geographic access differences between both populations
- Question 2: Are there differences in travel time to the closest mammography screening facility from the origin versus destinations locations?
 - **Minimal** differences in both populations considering non-working and working population
 - Rural residents **closer** to the nearest facility from their **workplace**
 - Saturation of facilities in PA **diminishing differences** between origin and destinations

Future of the Dataset

Notes

Scripts and future work will be posted to Lee Hachadoorian's GitHub account at https://github.com/leehach/ctpp_postgres.

The application to mammography, coauthored by Kevin Henry and Julia Wolanski, is a forthcoming publication currently undergoing edits.

Please feel free to contact the presenters at:

- Lee.Hachadoorian@temple.edu
- juliawolanski@gmail.com

References

- Elkin, E.B., Ishill, N.M., Snow, J.G., Panageas, K.S., Bach, P.B., Liberman, L., Wang, F., Schrag, D. (2010). Geographic access and the use of screening mammography. *Medical Care* 48(4), 349-356. Retrieved March 30, 2017, from <http://dx.doi.org/10.1097/MLR.0b013e181ca3ecb>.
- Luo, W., Qi, Y. (2009). An enhanced two-step floating catchment area (E2SFCA) methods for measuring spatial accessibility to primary care physicians. *Health & Place* 15, 1100-1107. Retrieved February 02, 2017, from <http://dx.doi.org/10.1016/j.healthplace.2009.06.002>.
- Luo, W., Wang, F. (2003). Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning B: Planning and Design* 30, 865-884. Retrieved November 03, 2016, from <http://dx.doi.org/10.1068/b29120>.
- Luo, W., Whippo, T. (2012). Variable catchment sizes for the two-step floating catchment area (2SFCA) method. *Health & Place* 18, 789-795. Retrieved January 20, 2017, from <http://dx.doi.org/10.1016/j.healthplace.2012.04.002>.
- Nelson, G., Rae, A. (2016). An economic geography of the United States: From commutes to megaregions. *PLoS ONE* 11(11): e0166083. <https://doi.org/10.1371/journal.pone.0166083>.
- Ni, J., Wany, J., Rui, Y., Qian, T., Wang, J. (2015). An enhanced variable two-step floating catchment area method for measuring spatial accessibility to residential care facilities in Nanjing. *International Journal of Environmental Research and Public Health* 12, 14490-14504. Retrieved January 20, 2017, from <http://dx.doi.org/10.3390/ijerph121115590>.
- Peipins, L.A., Miller, J., Richards, T.B., Bobo, J.K., Liu, T., White, M.C., Joseph, D., Tangka, F., Ekwueme, D.U. (2012). Characteristics of US counties with no mammography capacity. *J Community Health* 1. Retrieved November 03, 2016, from <http://dx.doi.org/10.1007/s10900-012-9562-z>.
- Schuurman, N., Berube, M., Crooks, V.A. (2010). Measuring potential spatial access to primary health care physicians using a modified gravity model. *The Canadian Geographer* 54(1), 29-45. Retrieved February 01, 2017, from <http://dx.doi.org/10.1111/j.1541-0064.2009.0030.x>.
- Shi, X., Alford-Teaster, J., Onega, T., & Wang, D. (2012). Spatial access and local demand for major cancer care facilities in the United States. *Annals of the Association of American Geographers* 102(5), 1125-134. Retrieved November 01, 2016, from <http://dx.doi.org/10.1080/00045608.2012.657498>.
- Wan, N., Zhan, F.B., Zou, B., Chow, E. (2012). A relative spatial access assessment approach for analyzing potential spatial access to colorectal cancer services in Texas. *Applied Geography* 32, 291-299. Retrieved February 03, 2017, from <http://dx.doi.org/10.1016/j.apgeog.2011.05.001>.
- Wan, N., Zou, B., Sternberg, T. (2012). A three-step floating catchment area method for analyzing spatial access to health services. *International Journal of Geographical Information Science* 26(6), 1073-1089. Retrieved February 02, 2017, from <http://dx.doi.org/10.1080/13658816.2011.624987>.
- Wang, F. (2015). Quantitative methods and socio-economic applications in GIS. Chapter 5. Boca Raton, FL.: CRC Press.
- Wang, F., Luo, W. (2005). Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place* 11, 131-146. Retrieved January 19, 2017, from <http://dx.doi.org/10.1016/j.healthplace.2004.02.003>.
- Widener, M.J., Farber, S., Neutens, T., & Horner, M.W. (2013). Using urban commuting data to calculate a spatiotemporal accessibility measure for food environment studies. *Health & Place* 21, 1-9. Retrieved September 21, 2016, from <http://dx.doi.org/10.1016/j.healthplace.2013.01.004>.