Where are Things Changing?

Using data from the American Community Survey for Historical Analyses

By Doug Hillmer

Independent consultant

Where are Things Changing?

Topics I will cover today

- 14 years of ACS data covering whole U.S. population – how can we use these data?
- Why I chose educational attainment as the topic to investigate
- Why I chose counties
- Why I chose to use a single ACS table (and its iterated version) as my data source
- How I chose to measure change
- What I found
- What I learned

Where are Things Changing?

14 years of ACS data covering whole U.S. population – how can we use these data?

- Should we use the 1-year data or the 5-year data?
- Should we look for changes over the entire 14-year period or a sub-period?

Where are Things Changing?

Why I chose educational attainment

- Became interested when working on a project for a highly rural congressional district
- Discovered that an increasing proportion of white men were failing to complete high school
- The opposite was true for white women in the district
- Educational Attainment is well-studied at the national and state levels
- Wanted to look at the two extremes of the educational attainment distribution:
  - No high school completion (including GED)
  - Bachelor's degree or higher

Measures used in study
two measures of educational attainment by gender for 4 groups (total population, Black alone, White alone, Hispanic origin plus 8 ratio measures)

variable name	description
bach_all_f_pct	total population: females with a Bachelors degree or higher as a proportion of all females 25 years or older
bach_all_m_pct	total population: males with a Bachelors degree or higher as a proportion of all males 25 years or older
bach_all_ratio	total population: ratio of men to women with bachelors degree or higher
bach_black_f_pct	Black alone population: females with a Bachelors degree or higher as a proportion of all females 25 years or older
bach_black_m_pct	Black alone population: males with a Bachelors degree or higher as a proportion of all males 25 years or older
bach_black_ratio	Black alone population: ratio of men to women with bachelors degree or higher
bach_hisp_f_pct	Hispanic population: females with a Bachelors degree or higher as a proportion of all females 25 years or older
bach_hisp_m_pct	Hispanic population: males with a Bachelors degree or higher as a proportion of all males 25 years or older
bach_hisp_ratio	Hispanic population: ratio of men to women with bachelors degree or higher
bach_white_f_pct	White alone population: females with a Bachelors degree or higher as a proportion of all females 25 years or older
bach_white_m_pct	White alone population: males with a Bachelors degree or higher as a proportion of all males 25 years or older
bach_white_ratio	White alone population: ratio of men to women with bachelors degree or higher
no_hs_all_f_pct	total population: females with no high school completion as a proportion of all females 25 years or older
no_hs_all_m_pct	total population: males with no high school completion as a proportion of all males 25 years or older
no_hs_all_ratio	total population: ratio of men to women no high school completion
no_hs_black_f_pct	Black alone population: females with no high school completion as a proportion of all females 25 years or older
no_hs_black_m_pct	Black alone population: males with no high school completion as a proportion of all males 25 years or older
no_hs_black_ratio	Black alone population: ratio of men to women no high school completion
no_hs_hisp_f_pct	Hispanic population: females with no high school completion as a proportion of all females 25 years or older
no_hs_hisp_m_pct	Hispanic population: males with no high school completion as a proportion of all males 25 years or older
no_hs_hisp_ratio	Hispanic population: ratio of men to women no high school completion
no_hs_white_f_pct	White alone population: females with no high school completion as a proportion of all females 25 years or older
no_hs_white_m_pct	White alone population: males with no high school completion as a proportion of all males 25 years or older
no_hs_white_ratio	White alone population: ratio of men to women no high school completion

Where are Things Changing?

Why I chose counties

- Counties have local governments
- People know the name of their county and the names of nearby counties
- County boundaries change infrequently
- PUMAs were considered
  - Using detailed tables from 1-year ACS – too much loss due to data quality filtering
  - PUMA definitions changed in 2012 because of 2010 Census
  - In many states PUMAs have no meaning to most people
- Cities were considered
  - Too many boundary changes
  - Fewer cities than counties meet the 65,000+ population threshold
  - Most of the cities that meet the threshold have total population < 150,000
  - Most of the counties that meet the threshold have population >150,000

Where are Things Changing?

Why I chose to use a single ACS table (and its iterated version)

- Simpler than using estimates from multiple tables
- I found a simple table that existed unchanged throughout the period
  - Not always true: Subject Table S1501 existed throughout, but changed from 2006 to 2019
- I used only C15002 and C15002A-I
- “Full table”, B15002, B15002A-I had more detail than I needed – more likely to be suppressed

C15002			**SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER**
C15002			*Universe: Population 25 years and over*
C15002	1	C15002_001	Total:
C15002	2	C15002_002	Male:
C15002	3	C15002_003	Less than 9th grade
C15002	4	C15002_004	9th to 12th grade, no diploma
C15002	5	C15002_005	High school graduate (includes equivalency)
C15002	6	C15002_006	Some college, no degree
C15002	7	C15002_007	Associate's degree
C15002	8	C15002_008	Bachelor's degree
C15002	9	C15002_009	Graduate or professional degree
C15002	10	C15002_010	Female:
C15002	11	C15002_011	Less than 9th grade
C15002	12	C15002_012	9th to 12th grade, no diploma
C15002	13	C15002_013	High school graduate (includes equivalency)
C15002	14	C15002_014	Some college, no degree
C15002	15	C15002_015	Associate's degree
C15002	16	C15002_016	Bachelor's degree
C15002	17	C15002_017	Graduate or professional degree
C15002A			**SEX BY EDUCATIONAL ATTAINMENT FOR THE POPULATION 25 YEARS AND OVER (WHITE ALONE)**
C15002A			*Universe: White alone population 25 years and over*
C15002A	1	C15002A_001	Total:
C15002A	2	C15002A_002	Male:
C15002A	3	C15002A_003	Less than high school diploma
C15002A	4	C15002A_004	High school graduate (includes equivalency)
C15002A	5	C15002A_005	Some college or associate's degree
C15002A	6	C15002A_006	Bachelor's degree or higher
C15002A	7	C15002A_007	Female:
C15002A	8	C15002A_008	Less than high school diploma
C15002A	9	C15002A_009	High school graduate (includes equivalency)
C15002A	10	C15002A_010	Some college or associate's degree
C15002A	11	C15002A_011	Bachelor's degree or higher

Where are Things Changing?

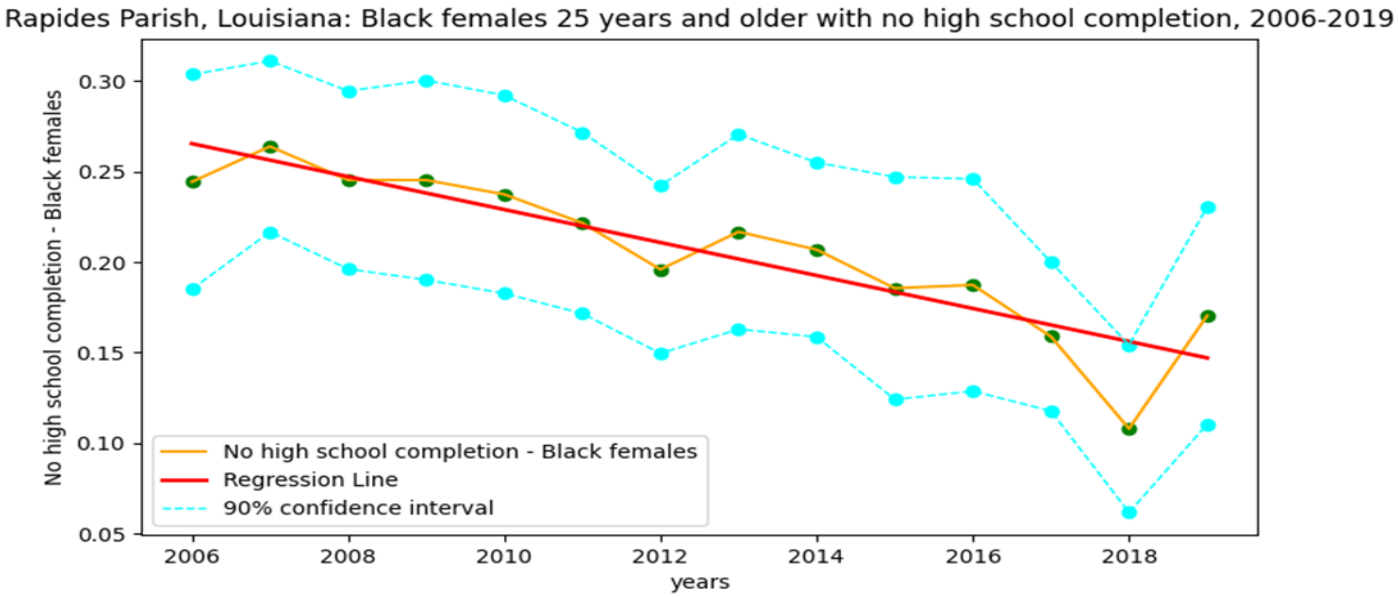
Brief outline of my data processing steps:

1. Determined that the tables C15002 and C15002A-I existed unchanged throughout 2006-2019
2. Created and ran program to download the data needed from the Census API
3. Created and ran programs to create measures needed (along with the MOEs)
4. Created reports summarizing results
5. Created graphs for some examples of results

Where are Things Changing?

How I chose to measure change

- My interpretation of “change over time” : an approximately monotonic curve going up or down. Here is an example. Note: none of the year-to-year changes were statistically significant at the 90% level.



Where are Things Changing?

How I chose to measure change

- My interpretation of “change over time” : an approximately monotonic curve going up or down
- How do we measure this?
  - Mann-Kendall nonparametric trend test
  - Most often used in weather-related and hydrology applications
  - Contains an actual hypothesis testing capability
  - There is a trend test based on regression, but it was not clear tome how to use it. Mann-Kendall seemed more straightforward.

Where are Things Changing?

How I chose to measure change

How Mann-Kendall (MK) works:

Starts with a simple measure of the year-to-year changes in the values:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(X_j - X_k) \quad \text{with} \quad \operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

For the 2006-2019 time period n=14. S must be in the interval [-91,91]. A consistently upward curve will get close to 91 and a consistently downward curve will get close to -91.

Where are Things Changing?

How I chose to measure change

How Mann-Kendall (MK) works:

MK allows for hypothesis testing at any pre-defined confidence level using the p value. A test statistic (a type of Z statistic) is created and tested to see if the value returned is less than the pre-defined p-value. If it is, a trend is determined to exist and the direction of the trend is determined by the S value. Otherwise, the null hypothesis is not rejected and “no trend” is reported.

I used a p-value of 0.001 for the MK hypothesis testing.

Where are Things Changing?

What I found

- results for the total population and for the White alone population in 817 counties.
- Could only get results for the Black alone population in 489 counties
- For the Hispanic origin population in 513 counties.
- However, for each group, these counties contained about 80% of that group's U.S. total.
- Here is a summary of the results showing trends in about 10% of the cases examined.

	Trend result	Avg S value	Avg p value	Number of occurrences
Results across all measures and all counties	decreasing	-69.78	0.0003	875
	increasing	71.69	0.0002	879
	no trend	-3.41	0.3306	16,030
			Total	**17,784**

Results for 24 measures of educational attainment for counties published from the 1-year ACS tables, 2006-2019

Notes: 1) The columns labelled "Avg S values" refer to the average of the Mann-Kendall S estimates for assessing a trend.

2) Only counties with 10 or more years of published results are included.

3) A p-value threshold of 0.001 was used in the Mann-Kendall test to test the hypothesis that a trend existed.

	decreasing		increasing		no trend		Total	
Measure	**Counties**	**Avg S value**	**Counties**	**Avg S value**	**Counties**	**Avg S value**	**Counties**	**Avg S value**
bach_all_f_pct			317	73.00	500	44.78	817	55.73
bach_all_m_pct			102	70.92	715	29.36	817	34.55
bach_all_ratio	40	-67.85			777	-27.81	817	-29.77
bach_black_f_pct			33	69.55	456	23.52	489	26.62
bach_black_m_pct			6	68.00	483	14.15	489	14.81
bach_black_ratio					489	-5.91	489	-5.91
bach_hisp_f_pct			39	70.95	474	20.80	513	24.61
bach_hisp_m_pct			19	67.63	494	17.68	513	19.53
bach_hisp_ratio	1	-63.00			512	-7.73	513	-7.84
bach_white_f_pct			286	71.72	531	42.82	817	52.94
bach_white_m_pct			75	69.88	742	27.62	817	31.50
bach_white_ratio	31	-68.94			786	-26.55	817	-28.16
no_hs_all_f_pct	212	-70.20			605	-41.53	817	-48.97
no_hs_all_m_pct	149	-69.70			668	-37.30	817	-43.21
no_hs_all_ratio					817	9.56	817	9.56
no_hs_black_f_pct	45	-70.96			444	-27.63	489	-31.62
no_hs_black_m_pct	27	-69.59			462	-23.97	489	-26.49
no_hs_black_ratio	1	-63.00			488	6.81	489	6.67
no_hs_hisp_f_pct	61	-71.72			452	-21.44	513	-27.42
no_hs_hisp_m_pct	47	-70.83			466	-22.31	513	-26.76
no_hs_hisp_ratio	3	-62.67			510	-7.42	513	-7.75
no_hs_white_f_pct	164	-69.66			653	-38.60	817	-44.83
no_hs_white_m_pct	94	-68.32	1	65.00	722	-34.38	817	-38.16
no_hs_white_ratio			1	65.00	816	11.46	817	11.53

Results of the 24 measures for the metro parts of states published from the 1-year ACS data, 2006-2019

Note: The column labelled "Avg S value" refers to the Mann-Kendall S statistic for estimating a trend

	Descreasing		Increasing		No Trend		Total	
Measure	**Avg S value**	**States**	**Avg S value**	**States**	**Avg S value**	**States**	**Avg S value**	**States**
bach_all_f_pct			84.27	52			84.27	52
bach_all_m_pct			78.32	41	49.36	11	72.19	52
bach_all_ratio	-76.21	38			-42.29	14	-67.08	52
bach_black_f_pct			77.24	25	27.81	26	52.04	51
bach_black_m_pct			70.33	12	25.38	39	35.96	51
bach_black_ratio	-67.00	1			-13.78	50	-14.82	51
bach_hisp_f_pct			76.88	16	36.06	36	48.62	52
bach_hisp_m_pct			72.83	12	30.15	40	40.00	52
bach_hisp_ratio	-67.00	2			-18.00	50	-19.88	52
bach_white_f_pct			83.12	50	57.00	2	82.12	52
bach_white_m_pct			75.49	37	46.47	15	67.12	52
bach_white_ratio	-74.19	37			-40.87	15	-64.58	52
no_hs_all_f_pct	-81.52	50			-53.00	2	-80.42	52
no_hs_all_m_pct	-80.25	48			-51.50	4	-78.04	52
no_hs_all_ratio			71.00	6	25.57	46	30.81	52
no_hs_black_f_pct	-79.07	28			-21.83	23	-53.25	51
no_hs_black_m_pct	-76.52	25			-24.65	26	-50.08	51
no_hs_black_ratio			75.00	1	16.82	50	17.96	51
no_hs_hisp_f_pct	-75.29	21			-35.71	31	-51.69	52
no_hs_hisp_m_pct	-74.73	22			-35.37	30	-52.02	52
no_hs_hisp_ratio					-19.79	52	-19.79	52
no_hs_white_f_pct	-79.91	46			-54.67	6	-77.00	52
no_hs_white_m_pct	-78.42	45			-51.86	7	-74.85	52
no_hs_white_ratio			68.50	8	30.50	44	36.35	52

Results of the 24 measures for the non-metro parts of states published from the 1-year ACS data, 2006-2014

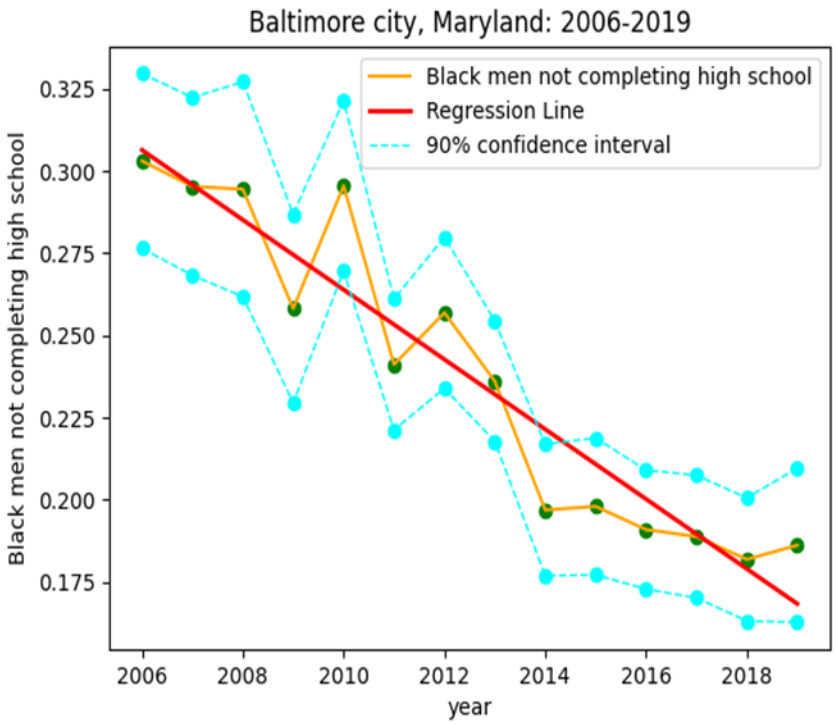
Note: The column labelled "Avg S value" refers to the Mann-Kendall S statistic for estimating a trend

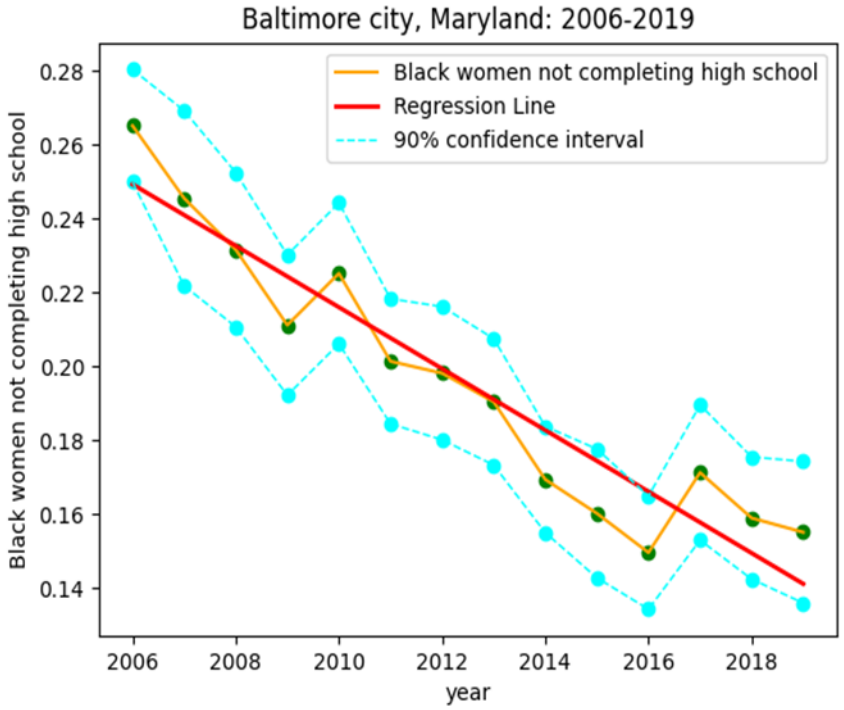
	decreasing		increasing		no trend		Total	
Measure	**Avg S value**	**States**	**Avg S value**	**States**	**Avg S value**	**States**	**Avg S value**	**States**
bach_all_f_pct			72.79	29	41.31	13	63.05	42
bach_all_m_pct			70.00	4	33.89	38	37.33	42
bach_all_ratio	-63.00	1			-31.05	41	-31.81	42
bach_black_f_pct			71.00	2	13.70	27	17.66	29
bach_black_m_pct					14.34	29	14.34	29
bach_black_ratio					1.31	29	1.31	29
bach_hisp_f_pct					11.76	42	11.76	42
bach_hisp_m_pct					19.02	42	19.02	42
bach_hisp_ratio					-14.19	42	-14.19	42
bach_white_f_pct			72.46	26	40.13	16	60.14	42
bach_white_m_pct			67.00	4	30.63	38	34.10	42
bach_white_ratio					-14.95	42	-14.95	42
no_hs_all_f_pct	-75.33	30			-46.17	12	-67.00	42
no_hs_all_m_pct	-74.78	27			-43.67	15	-63.67	42
no_hs_all_ratio			66.60	5	29.76	37	34.14	42
no_hs_black_f_pct	-69.86	7			-22.95	22	-34.28	29
no_hs_black_m_pct	-67.00	3			-28.58	26	-32.55	29
no_hs_black_ratio					12.28	29	12.28	29
no_hs_hisp_f_pct	-72.00	2			-19.25	40	-21.76	42
no_hs_hisp_m_pct	-70.33	3			-21.77	39	-25.24	42
no_hs_hisp_ratio					-4.33	42	-4.33	42
no_hs_white_f_pct	-74.52	29			-40.54	13	-64.00	42
no_hs_white_m_pct	-72.11	27			-33.80	15	-58.43	42
no_hs_white_ratio			65.50	4	27.63	38	31.24	42

Where are Things Changing?
Findings

- My results are consistent with results on both measures at the state and national level. (references are given in Appendix 3 of the paper)
- A limitation of the high school completion results: based solely on the population 25 years and older. So, a downward trend (= improvement in high school completion) may not be due to any program or policy implemented by a school system or county. For example, migration of people could affect these results.
- It would be helpful to further disaggregate these results by age groups (e.g. 25-34, 35-50, etc.). However, that may lead to so much data loss and reduced statistical quality in measures at the county level to force the analysis to be done at a higher geographic level.
- Many more “no trend” results in the non-metro portions of states versus the metro portions is a result requiring further analysis. For example, the proportion of Hispanic women not completing high school showed a decreasing trend in the non-metro parts of only 7 states, whereas the same measure had a decreasing trend in the metro parts of 32 states. Can this be because the metro portions of states are much more populous and thus the measure estimates are less variable?

Where are Things Changing?
Findings - Baltimore





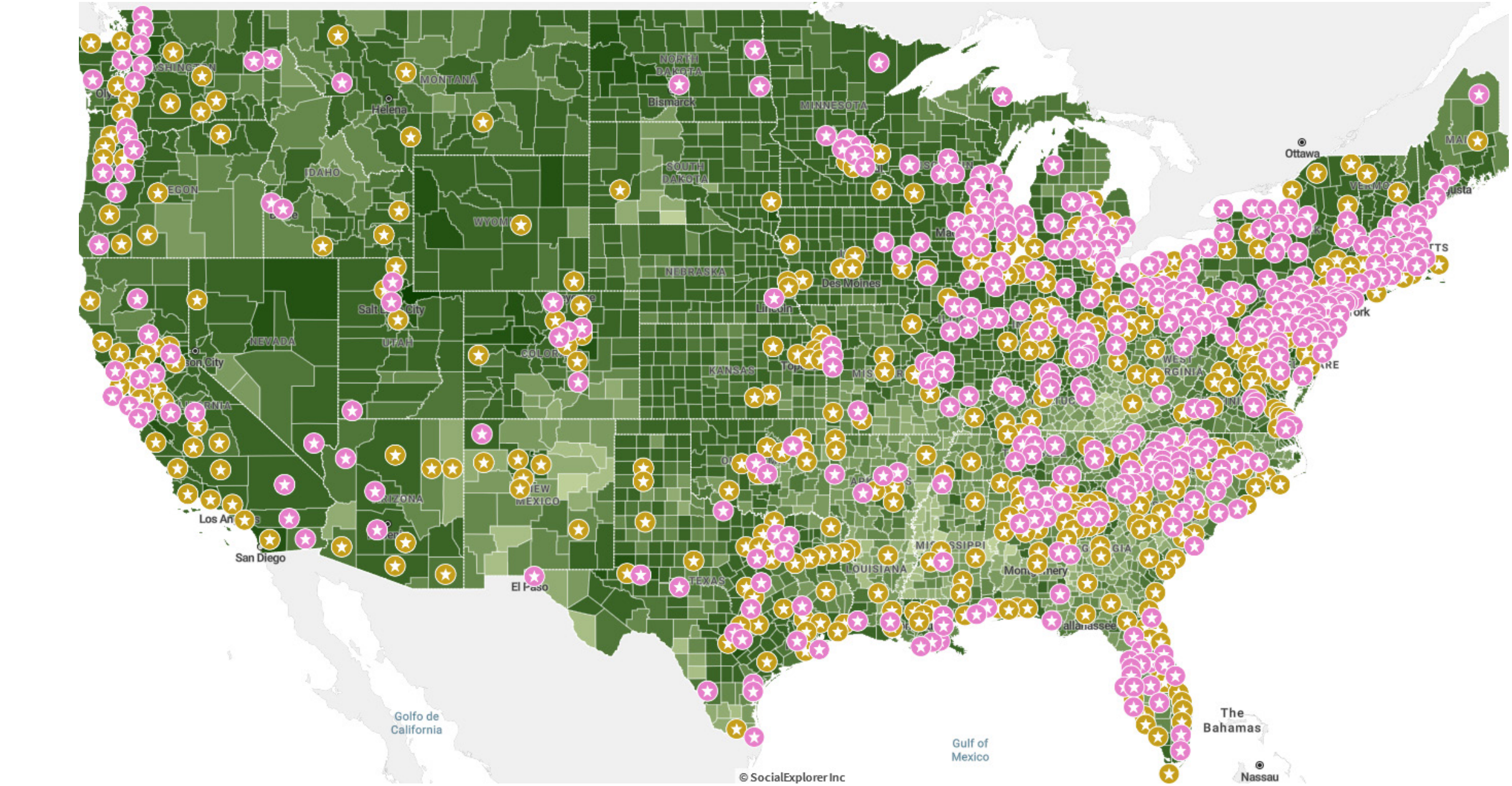
Where are Things Changing?

What I learned – questions to answer for next project

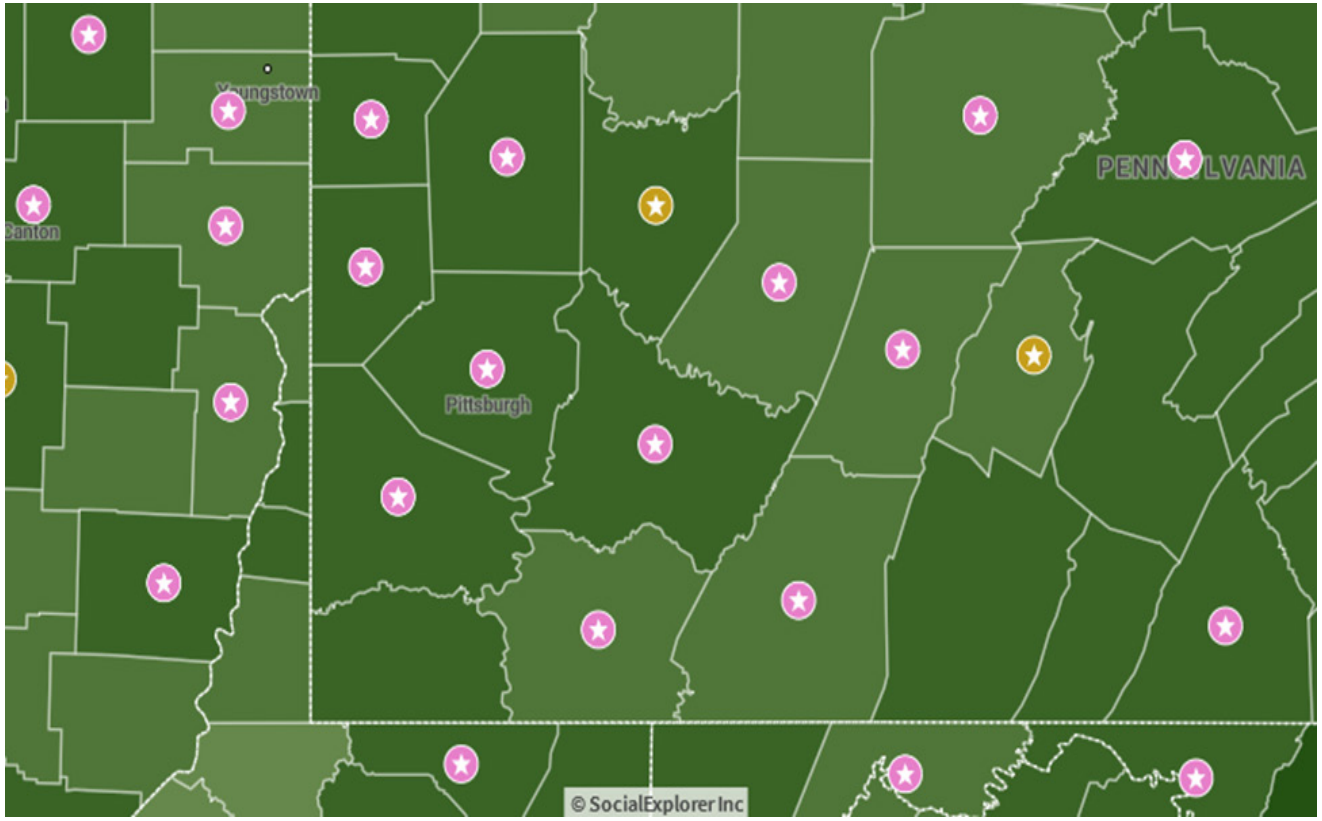
When I do my next "trend detection" project, I will try to incorporate what I have learned and answer these questions before I even start to download data:

a. Can I use the pre-aggregated data at the PUMA geographic level; i.e., am I using a time period in which the PUMA boundary definitions do not change?
b. Is there a table in one of the ACS data products that supports what I want to measure and does not change throughout the time period? If not, can I use the PUMS microdata to create my estimates without sacrificing too much in reliability?
c. Is there a way for me to check my results against some already created data product, perhaps at the margins of a table or only at a higher geographic level?
d. Should I use the 1-year ACS data products, or are there enough non-overlapping 5-year periods allowing me to use the 5-year ACS estimates? A related question: is there some way to use the 5-year estimates to "fill in the hole" of the counties too small for 1-year estimates? Or, can I "fill the hole" using the 1-year PUMS data for each year in the time period?
e. Is the population subgroup I plan to study distributed fairly evenly across geography, or in a very skewed manner (as is the case for the Black alone population)?

county level median household income map from Social Explorer with markers for counties analyzed for trend for measure no_hs_white_f_pct (pink markers indicate decreasing trend)



Detail from the full U.S. map showing a number of counties near Pittsburgh, PA and Youngstown, OH with trend results.



Where are Things Changing?

Lots of details are in the paper that could not be covered in this talk. There are also references in an appendix.

Thank you