Philadelphia Children's Health & Well-being

Methods

METHODS INTRODUCTION

The purpose of this project is to analyze and visualize the risks and assets for children's health and well-being across Philadelphia, by city council district. The primary project outcomes are specific reports for each city council district and an overall city summary report. The following describes the data used and analytics. The data used for this project are from 2014–2015, but this data does not typically change significantly from year to year.

TECHNOLOGIES

Python; R; Flask; CSS/HTML; Docker

DATA AND TECHNICAL CONSIDERATIONS

In order to make this work transparent and replicable, each aspect of the project, from data analysis. This approach allows other cities or regions to adapt this project to their area. It also allows Scattergood to more easily duplicate this work at a later date to generate temporal comparisons.

The analysis and scripts themselves are tailored to this project and any user seeking to duplicate this work with updated data or in a new geography must follow the Getting Started directions and edit the Python and R scripts to be specific to their data.

GETTING STARTED

- update/build the docker container ./scripts/dockerbuild.sh (this may take a little while if you haven't built it yet)
- enter into docker container ./scripts/container.sh
- enter a census API key into ./src/data/census.py

SCATTERGOOD

THINK DO SUPPORT

- run the overlay python -m src.run
- (optional) refresh report copy from csv python -m src.run.generate_json

GENERATING GRAPHICS

The bar charts and maps are generated via an R script which is not run from within the Docker container. Program and package versions that were used during this project are provided below. The script may work with older or newer versions but will require testing.

- R: 3.5.2
- tidyverse: 1.2.1
- tidyr: 0.8.2
- ggthemes: 4.0.1
- sf: 0.7-2
- svglite: 1.2.1
- httr: 1.4.0
- classInt: 0.3-1
- ggsn: 0.5.0

Once appropriate dependencies are installed, insert the correct working directory into the script. Then run R/plotting.R.

Generating Reports

See the [Flask App documentation] (flask-app/README.md)

METHODOLOGY AND DATA

This project generates scores for each risk and asset at the census tract level. Scores are then aggregated to each city council district. Risks are subtracted from assets to compute the final cumulative score. The scores at the census tract level are percentiles. Because the aggregation to city council district takes the mean, the scores at the city council district level will not represent a scale of 0 to 100. This cumulative score is also computed at the census tract level in order to visualize the variance across a more local level.

The methodology for computing scores, as percentiles, is taken from the Place Matters report (available at scattergoodfoundation.org). This allows the risk scores, which now incorporate updated American Community Survey (ACS) data, to be compared to the 2016 study.

azavea

AGGREGATING TRACT SCORES TO DISTRICTS

Census tracts were assigned to districts based on the location of each tract's centroid. Scores were then aggregated to districts by taking the mean of the scores for the census tracts, which were assigned to each district.

CREATING RISK AND ASSET SCORES FROM VARIABLES

The Place Matters report converts values for each variable to percentiles and takes the mean of each percentile for each variable to compute the risk scores.

For example, a poverty percentage is found for each census tract. This list of poverty percentages is then converted to percentiles. The highest poverty percentage is the 100th percentile, the lowest poverty percentage is the 1st percentile. All other values are added as percentiles along this rank.

Once all variable scores are converted to percentiles, the mean of those percentiles is taken. Due to this aggregation, this report does not report the percentiles directly at the census tract level, so the scores at the city council district level will not represent a scale of 0 to 100. Risk scores per district and tract are the mean of the percentiles for crime, low education status, poverty, unemployment, and Adverse Childhood Experiences (ACEs). Asset scores per district and tract are the mean of the percentiles for SNAP utilization, behavioral health center utilization, walkable access to healthy food, access to parks, and school quality. Information about how these variables are computed is below.

CREATING CUMULATIVE SCORES

To create the cumulative scores, which took into account how assets might mitigate risks, the risk scores were subtracted from the asset scores at the district level. This generated a cumulative score for each council district in the city, allowing each district to be ranked against one another. The same cumulative score calculation was computed at the census tract level for visualization purposes.

RISK VARIABLES

Crime

• Shooting victims, from 2014 onward, per 10,000 people.

Education

• Percent of the population that is 25 years and over, with less than a 9th grade education.

Census API variable: S1501_C02_007E

Poverty

• Percent of families in poverty with related children of the householder under 18 years old.

Census API variable: S1702_C01_002E

Unemployment

• The unemployment rate for the population 16 years and over.

Census API variable: S2301_C04_001E

Adverse Childhood Experiences (ACEs)

- 2016 Philadelphia Expanded ACE Survey conducted by the Philadelphia Health Management Corporation (PHMC).
- PHMC delivered a weight-adjusted file that was based on age, sex, race, and poverty distribution to get the proper sample size per census tract. This census tract weight adjusted file was used for census tract calculations.
- PHMC provided an additional weight to be applied when ACE data was aggregated to the district level. When aggregating risks from census tracts to districts, this weight was applied.

ASSET VARIABLES

SNAP Utilization

 Households receiving food stamps which, in the past 12 months, have been below the poverty level

Census API Variable: S2201_C03_021E

 Households not receiving food stamps which, in the past 12 months, have been below the poverty level

Census API Variable: S2201_C05_021E

 SNAP utilization metric was generated as a percentage of total households below poverty that are receiving food stamps.

Behavioral Health Usage

- This dataset is a behavioral health usage percentage by zip code.
- Usage by zip code was disaggregated to the census tract level by computing an area weighted join. This resulted in a usage metric per census tract.

Walkable Access to Healthy Food

 This dataset provides a block level assessment of access to healthy grocery stores. It was aggregated to the census tract level via centroid.

Park Assets

- Calculated a kernel density for existing and planned Rebuild locations
- Calculated a kernel density for park
 assets by sites
- Each kernel density was aggregated to each census tract. This provides two scores per census tract, one which represents the density of parks and recreation assets, the other representing the density of Rebuild locations. The average of the two scores was taken, producing one score per census tract.

School Performance

- School performance scores are at the catchment level and were disaggregated to tracts using an area weighted join.
- Averaged scores for all three levels (elementary, middle, high) to get an overall metric, providing one score per census tract.

Social Mobility Data

The data represented as social mobility is obtained from Opportunity Insights, from the All Outcomes by Census Tract, Race. Gender and Parental Income Percentile table. Data is provided at the census tract level. The variable used was kir_pooled_pooled_mean, this is a pool of all races and genders. Kir is a code for the outcome defined as "Mean percentile rank (relative to other children born in the same year) in the national distribution of individual income (i.e. just own earnings) measured as mean earnings in 2014-2015 for the baseline sample". The data was aggregated to districts following the methodology outlined above.

Life Expectancy and Social Mobility Charts

Life expectancy and social mobility data are provided for context for each district and for the city as a whole. Neither of these datasets are incorporated into the risk, asset, or cumulative scores.

Life Expectancy Data

The data is provided as life expectancy, in years, per census tract. The data was aggregated to districts following the methodology outlined above.

DATA SOURCES

American Community Survey, 2016 5-Year Estimates

- SNAP households
- Educational Status
- Poverty
- Unemployment

The Census Bureau Tiger Database

Census tract boundaries, 2010

City of Philadelphia

- Crime data, 2006-2018
- Park assets, date not available
- Health centers, 2018
- Streets, 2018
- Walkable access to healthy food, 2012-2014
- District boundaries, 2016
- School performance data, 2016-2018
- School locations, 2018
- School catchments, 2017-2018
- Community school locations, 2018
- Rebuild locations, 2018

PASDA

- Park boundaries, 2016
- Libraries, 2012

Public Health Management Corporation's Community Health Database

• Adverse Childhood Experiences (ACEs) data, 2016

Community Behavioral Health, 2015

Behavioral health usage by zip code, based on unique eligible members. 2015

Opportunity Insights

 Social mobility data (https://opportunityinsights.org/data/)
 Data is provided for children born between 1978 and 1983. Data gathered by Azavea in 2018.

National Center for Health Statistics. U.S. Small-Area Life Expectancy Estimates Project (USALEEP)

Life Expectancy Estimates Files (https://www.cdc.gov/nchs/nvss/ usaleep/usaleep.html), 2010-2015. National Center for Health Statistics. 2018.