Data sources, dependencies, scripts, and outputs

# Running / reproducing this work

1. Download and install dependencies.
2. Open the R Project.
3. Open and run the script(s) of your choosing.
4. Use the outputs.

# Data sources

Data sources for this project include the 1990, 2000, and 2010 Decennial Census downloaded from the [National Historical GIS](https://www.nhgis.org/) (NHGIS), and 2006-2010 and 2013-2017 ACS 5-Year Estimates downloaded using the Census API. Census tract boundaries are different in the 1990 and 2000 Censuses from 2010 tract boundaries. NHGIS has aggregated some, but not all, 1990 and 2000 tables from census blocks to 2010 census tract boundaries. When NHGIS geographically-standardized data were not available, the [Longitudinal Tract Database](https://s4.ad.brown.edu/Projects/Diversity/Researcher/Bridging.htm) (LTDB) was used to interpolate 1990 and 2000 census tract data to 2010 census tract boundaries. See **Table 1** for a list of the 11 variables used in this study and their data sources.

Household income and home value figures are reported in 2017 inflation-adjusted dollars. Because the CPI-U inflation adjustment factor does not account for variation in geography or spending category, inflation-adjusted values serve as an approximation of changes in income and housing values over time. Inflation adjustment factors were as follows: 1990-2017, 1.84; 2000-2017, 1.42; 2010-2017, 1.12.

When looking at data over time, there are sometimes notable changes in variables between 2000 and 2010. This discontinuity can partially be attributed to changes in data sources. While the Decennial Census was used for 1990 and 2000, Decennial data was not available for all variables in 2010, and data from 2006-2010 ACS 5-Year Estimates was substituted for the household income, incomes relative to the Federal Poverty Level, and unemployment variables. Because ACS Estimates are a totally different type of data product from the Decennial Census, please **interpret changes in income, poverty, and unemployment from 2000 to 2010 with caution.**

# Dependencies

1. **R and RStudio, incl. extrafont, haven, here, sf, tidycensus, tidyverse, and units packages.** R is used to download and process data, compute changes between data years, and create summary measures and charts.
2. **STATA.** The only program that can run Longitudinal Tract Database scripts to harmonize 1990 and 2000 Census datasets to 2010 tract boundaries. Warning: DVRPC does not have STATA software on any computer.
3. **ArcMap.** Used to create maps of change from 1990 to 2017.

# Scripts and outputs

process\_raw.R

* Purpose: Merges and standardizes all datasets from the NHGIS and the ACS and exports standardized files. This script was created first, and the other four scripts rely on its outputs. While the project folder now contains all tract-harmonized data as .dta files, these .dta files were imports and exports from Longitudinal Tract Database scripts in STATA, which is not a program that DVRPC owns. **Do not delete files with the** .dta **extension. Because no DVRPC computer has STATA, you will not be able to reproduce results if you delete any** .dta **files.**
* Input
  + nhgis0011\_ts\_geog2010\_tract.csv
  + nhgis0011\_ds120\_1990\_tract.csv
  + nhgis0011\_ds123\_1990\_tract.csv
  + nhgis0012\_ds120\_1990\_tract.csv
  + nhgis0011\_ds151\_2000\_tract.csv
  + nhgis0012\_ds146\_2000\_tract.csv
* Output: Standardized files are saved in the “final” folder as merg\_<YEAR>.csv, where the year can be 90 (1990), 00 (2000), 10 (2010), or 17 (2017).

change.R

* Purpose: Computes changes from 1990 to 2017 for all variables.
* Input: merg\_90.csv, merg\_17.csv
* Output: change.shp

distance\_plots.R

* Purpose: Computes as-the-crow-flies distance from census tract centroids to City Hall by income bracket and year. Don’t like the fonts and colors? They’re embedded in the script to generate each plot. Font is Segoe UI, color is hexadecimal color #666666. Other colors are #FCA50A, #DD513A, #932667, and #420A68. It’s easy to trawl through the script and change these parameters if you’d like.
* Input
  + merg\_90.csv
  + merg\_00.csv
  + merg\_10.csv
  + merg\_17.csv
* Output: Four .png images are exported as lixdr\_<YEAR>.png, where <YEAR> is the year.

indices.R

* Purpose: Computes six measures of segregation and inequality from 1990 to 2017. The [dissimilarity index](https://www.dartmouth.edu/~segregation/IndicesofSegregation.pdf) is a measure of the spatial evenness of a population and can range from 0 to 1, where smaller values mean greater spatial evenness. Dissimilarity indices are calculated for the racial minority, ethnic minority, low-income (households with incomes 0-199% FPL), poverty (households with incomes 0-99% FPL), and unemployment variables. The [Theil index of economic inequality](https://utip.lbj.utexas.edu/tutorials/theil_guide.doc) is computed for median household income (2017 dollars) from 1990 to 2017.
* Input
  + merg\_90.csv
  + merg\_00.csv
  + merg\_10.csv
  + merg\_17.csv
* Output
  + racial\_dissim\_idx.csv
  + ethnic\_dissim\_idx.csv
  + low\_income\_dissim\_idx.csv
  + poverty\_dissim\_idx.csv
  + unemployment\_dissim\_idx.csv
  + income\_theil\_idx.csv

time\_series.R

* Purpose: Creates time series graphs of all variables, by county and by geographic subregion (Philadelphia, PA Suburban Counties, NJ Suburban Counties). Don’t like the fonts and colors? They’re embedded in the script to generate each plot. Font is Segoe UI, color is hexadecimal color #666666. The lines in each plot are generated using the scale\_color\_viridis\_d command, though this could be replaced with scale\_color\_manual followed by the color codes for the number of colors you need. The scale\_color\_manual function is used in distance\_plots.R.
* Input
  + merg\_90.csv
  + merg\_00.csv
  + merg\_10.csv
  + merg\_17.csv
* Output: 22 unique plots in the format d\_<VARIABLE>\_<GEOGRAPHIC UNIT>.png, where the following codes apply:
  + VARIABLE
    - Ethnic minority = em
    - Number of housing units = hu
    - Median household income = mhi
    - Median home value = mhv
    - Percentage of owned housing units (Tenure) = own
    - Total population = pop
    - Percentage of residents with incomes below FPL = pov99
    - Percentage of residents with incomes below 199% FPL = pov199
    - Racial minority = rm
    - Percentage unemployment = unemp
    - Vacant housing units = vhu
  + GEOGRAPHIC UNIT
    - By County = cty
    - Aggregated to subregion = agg

**Table 1**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Year** | **Source** |
| Population | 1990 | NHGIS geographically standardized |
| Population | 2000 | NHGIS geographically standardized |
| Population | 2010 | NHGIS |
| Population | 2017 | ACS B01003 |
| No. Housing Units | 1990 | NHGIS geographically standardized |
| No. Housing Units | 2000 | NHGIS geographically standardized |
| No. Housing Units | 2010 | NHGIS |
| No. Housing Units | 2017 | ACS B25002 |
| Tenure (Pct. Own) | 1990 | NHGIS geographically standardized |
| Tenure (Pct. Own) | 2000 | NHGIS geographically standardized |
| Tenure (Pct. Own) | 2010 | NHGIS |
| Tenure (Pct. Own) | 2017 | ACS B25003 |
| Median Home Value | 1990 | NHGIS adjusted for inflation and harmonized with LTDB |
| Median Home Value | 2000 | NHGIS adjusted for inflation and harmonized with LTDB |
| Median Home Value | 2010 | ACS adjusted for inflation |
| Median Home Value | 2017 | ACS B25077 |
| Pct. Vacant Housing Units | 1990 | NHGIS geographically standardized |
| Pct. Vacant Housing Units | 2000 | NHGIS geographically standardized |
| Pct. Vacant Housing Units | 2010 | NHGIS |
| Pct. Vacant Housing Units | 2017 | ACS B25002 |
| Pct. Racial Minority | 1990 | NHGIS geographically standardized |
| Pct. Racial Minority | 2000 | NHGIS geographically standardized |
| Pct. Racial Minority | 2010 | NHGIS |
| Pct. Racial Minority | 2017 | ACS B02001 |
| Pct. Ethnic Minority | 1990 | NHGIS geographically standardized |
| Pct. Ethnic Minority | 2000 | NHGIS geographically standardized |
| Pct. Ethnic Minority | 2010 | NHGIS |
| Pct. Ethnic Minority | 2017 | ACS B03002 |
| Pct. Residents with Incomes below 199% FPL | 1990 | NHGIS harmonized with LTDB |
| Pct. Residents with Incomes below 199% FPL | 2000 | NHGIS harmonized with LTDB |
| Pct. Residents with Incomes below 199% FPL | 2010 | ACS B05010 |
| Pct. Residents with Incomes below 199% FPL | 2017 | ACS B05010 |
| Pct. Residents with Incomes below 100% FPL | 1990 | NHGIS harmonized with LTDB |
| Pct. Residents with Incomes below 100% FPL | 2000 | NHGIS harmonized with LTDB |
| Pct. Residents with Incomes below 100% FPL | 2010 | ACS B05010 |
| Pct. Residents with Incomes below 100% FPL | 2017 | ACS B05010 |
| Median Household Income | 1990 | NHGIS adjusted for inflation and harmonized with LTDB |
| Median Household Income | 2000 | NHGIS adjusted for inflation and harmonized with LTDB |
| Median Household Income | 2010 | ACS adjusted for inflation |
| Median Household Income | 2017 | ACS B19013 |
| Unemployment Rate | 1990 | NHGIS harmonized with LTDB |
| Unemployment Rate | 2000 | NHGIS harmonized with LTDB |
| Unemployment Rate | 2010 | ACS B17005 |
| Unemployment Rate | 2017 | ACS B17005 |