

Rate Stabilizing Tool (RST)

*** Files needed for exercise: *NC_HeartDisease0608_simulate.dbf*, and *North Carolina census tract shapefile: NC_tracts_2010prj.shp*

Goal: Use the Rate Stabilizing Tool (RST) to produce easily mapped age-standardized, smoothed sub-county estimates.

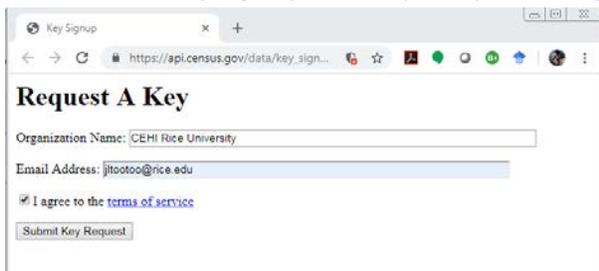
Objectives:

- Gain experience installing the tool and managing the user interface;
- Develop an understanding of the required data inputs; and
- Interpret and map the output

Problem: You have been provided with individual level heart disease death data for three years (2006 to 2008) that have been geocoded to the US 2010 Census tract level. Using the Rate Stabilizing Tool (RST), you will generate age-standardized, smoothed rates. Using the information generated by the RST, you will be able to map tract level death rates for heart disease for the state of North Carolina.

Obtain a US Census API key

1. Go to Census Key Sign up site: https://api.census.gov/data/key_signup.html



The screenshot shows a web browser window with the URL https://api.census.gov/data/key_signup.html. The page title is "Request A Key". The form contains the following fields and elements:

- Organization Name: CEHI Rice University
- Email Address: jitootoo@rice.edu
- A checked checkbox labeled "I agree to the [terms of service](#)".
- A "Submit Key Request" button.

Fill all blanks; in a few minutes you should receive an e-mail like this:

Hello!

Thank you for your interest in the Census Data API. Your API key is *de5117a6a7856a2278eb58afc38961991c4d5568*. Please [click here to activate your key](#).

Save this email for future reference.

Have Fun,

The Census Bureau API Team

Follow @uscensusbureau on twitter for API updates.

Activate your key and you are ready to use the Census API.

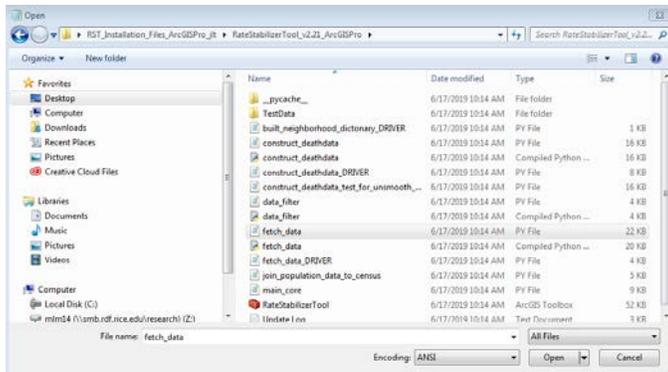
You will need to use this key after you have downloaded the RST toolbox.

Rate Stabilizing Tool (RST)

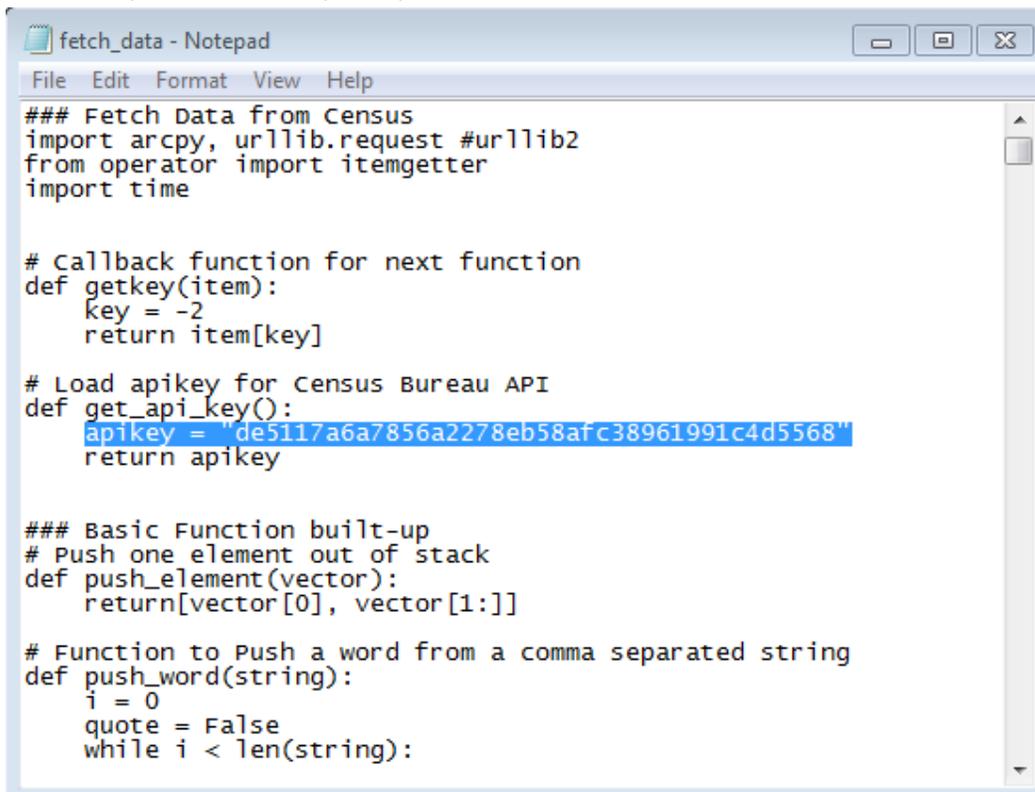
Loading the toolbox

1. Download the tool.
 - a. Save the zipped file to your workspace folder
 - b. Extract zipped file to your workspace folder

This tool has been built using Python and can be added as a toolbox into the ArcGIS Desktop interface.
2. Add your API key.
 - a. In your workspace folder RST folder, use notepad to open *fetch_data.py*



- b. In line 14, paste the API key that you received from the US Census.



```

### Fetch Data from Census
import arcpy, urllib.request #urllib2
from operator import itemgetter
import time

# callback function for next function
def getkey(item):
    key = -2
    return item[key]

# Load apikey for Census Bureau API
def get_api_key():
    apikey = "de5117a6a7856a2278eb58afc38961991c4d5568"
    return apikey

### Basic Function built-up
# Push one element out of stack
def push_element(vector):
    return[vector[0], vector[1:]]

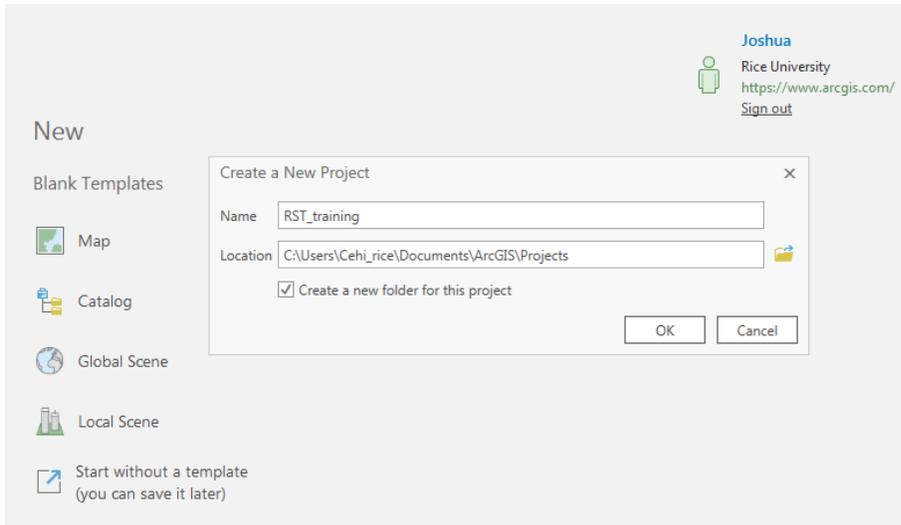
# Function to Push a word from a comma separated string
def push_word(string):
    i = 0
    quote = False
    while i < len(string):

```

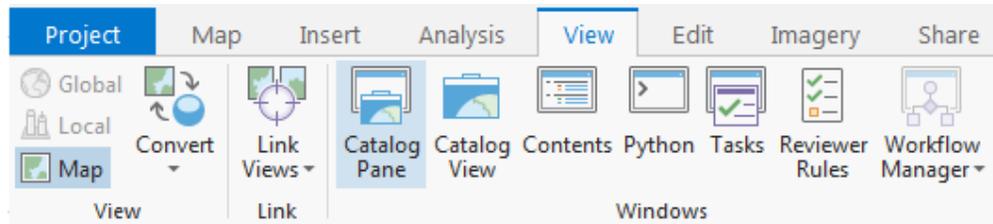
Close the *fetch_data.py* and save your changes.

Rate Stabilizing Tool (RST)

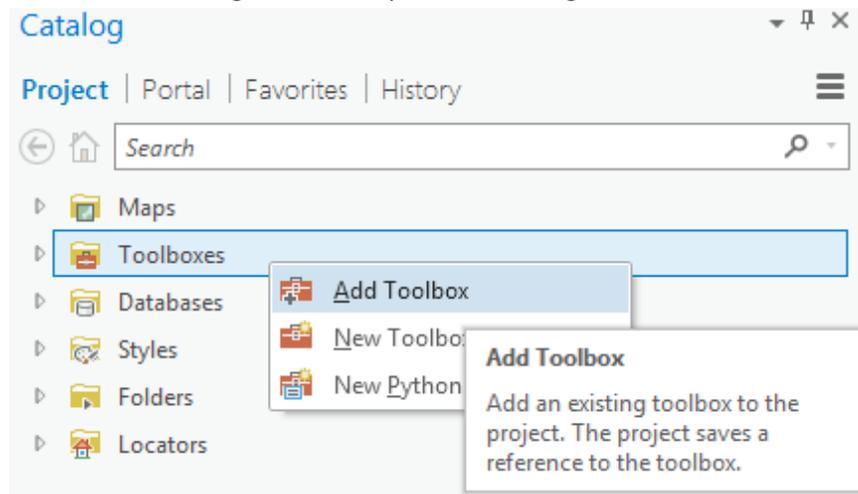
3. Create a new map project in ArcGIS Pro.



4. Activate the Catalog pane by left clicking on the **View** tab on the top ribbon and click on **Catalog Pane**.

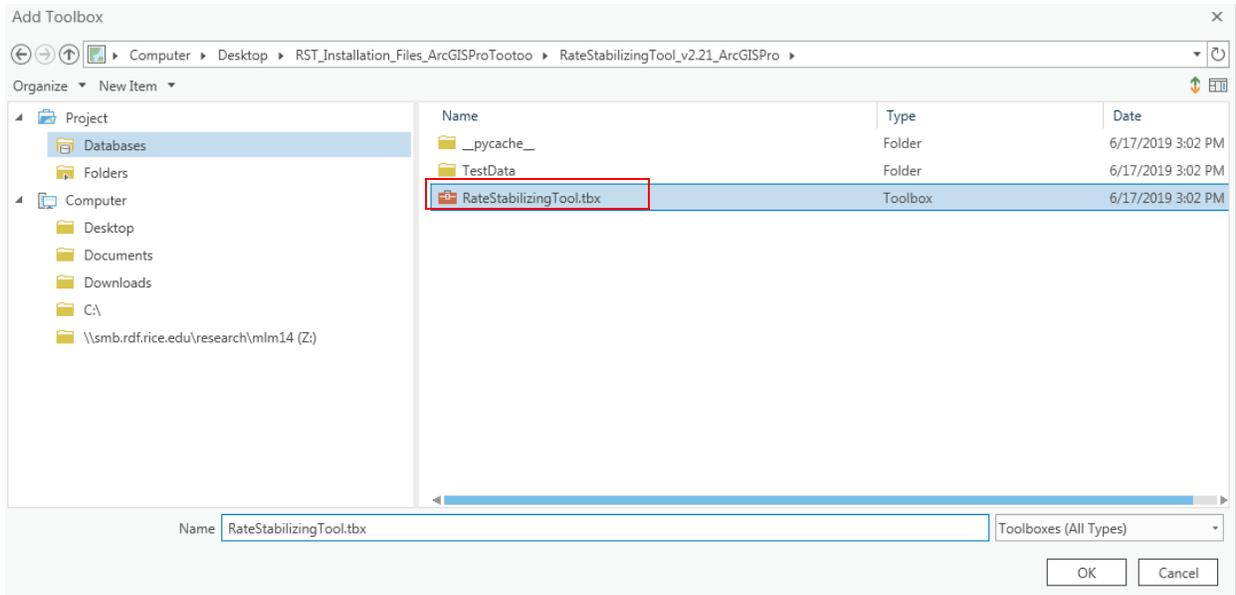


5. Right click on the **Toolboxes** icon again at the top of the **Catalog** and select Add toolbox...

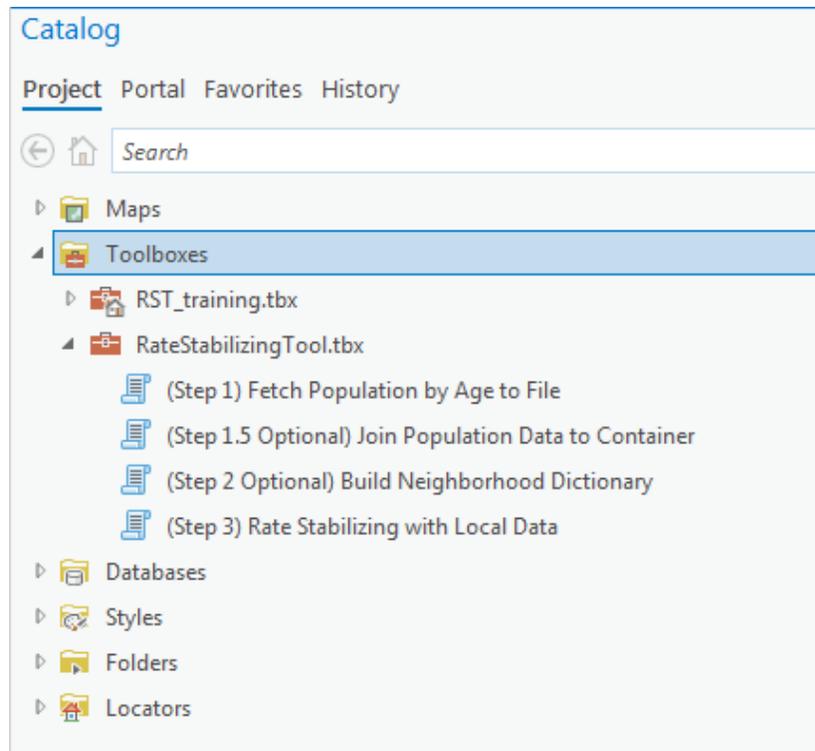


Rate Stabilizing Tool (RST)

- Navigate to your workspace location (where you extracted the toolbox)
RateStabilizingTool_v2.20_ArcGISPro > *RateStabilizingTool.tbx* and click **OK**.



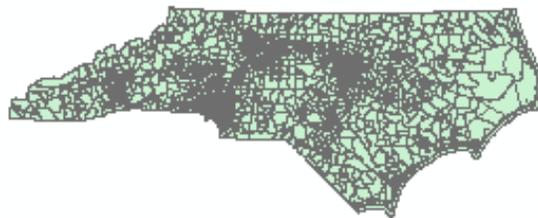
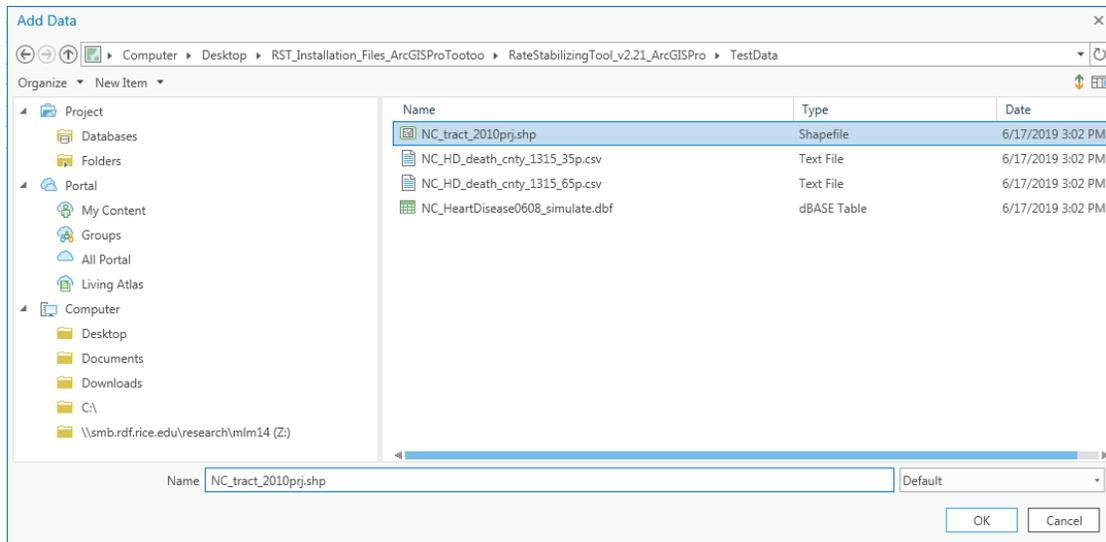
- The **RateStabilizingTool** toolbox will now appear. Expand the toolbox to reveal all of the Script-based tools in this toolbox.



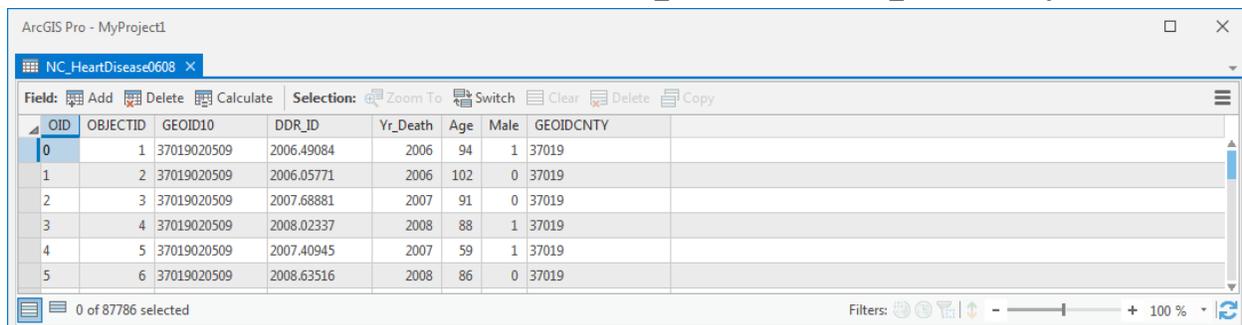
Rate Stabilizing Tool (RST)

Adding your data

1. Select Add Data in the Map Ribbon. The *RateStabilizingTool_v2.20_ArcGISPro TestData* folder contains a shapefile: **NC_tracts_2010prj.shp**. Add this shapefile for some geographic context.
 - a. You can download geographic boundaries for your state from the Census website (<https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>). Be sure the boundaries have **not** been generalized or simplified for cartographic display.



2. Next add the record level heart disease death data: **NC_HeartDisease0608_simulate.dbf**



OID	OBJECTID	GEOID10	DDR_ID	Yr_Death	Age	Male	GEOIDCNTY
0	1	37019020509	2006.49084	2006	94	1	37019
1	2	37019020509	2006.05771	2006	102	0	37019
2	3	37019020509	2007.68881	2007	91	0	37019
3	4	37019020509	2008.02337	2008	88	1	37019
4	5	37019020509	2007.40945	2007	59	1	37019
5	6	37019020509	2008.63516	2008	86	0	37019

Right click on the table name to view this table and take a look at the attribute fields:

OID – an automatically generated object identifier;

OBJECTID – automatically generated unique identifier;

Rate Stabilizing Tool (RST)

GEOID10 – the geographic identifier for the US Census tract unit that you will be calculating a stabilized age-standardized rate, this is a required field for the tool;

DDR_ID – the unique-identifier for each record;

Yr_Death – the year of the death record;

Age – the age at which the individual represented by the record died, this is also a required field for the tool;

Male – the sex of the decedent, with 1 indicating male and 0 indicating female (to be used in a future version of the tool); and

GEOIDCNTY – alternatively, the geographic identifier for the US Census county unit can also be used to calculate a stabilized age-standardized rate.

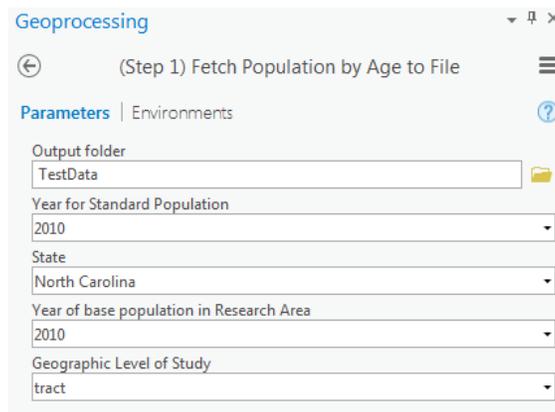
Using the tool

This tool makes use of the US Decennial Census API to obtain census population data for age-standardization calculation.

1. First, run the *(Step 1) Fetch Population by Age to File Script* tool. Double left click on the icon to open the tool.

Name	Type
(Step 1) Fetch Population by Age to File (C	Script
(Step 1.5 Optional) Join Population Data to	Script
(Step 2 Optional) Build Neighborhood Dict	Script
(Step 3) Rate Stabilizing with Local Data	Script

2. Let's consider each of the parameters/inputs.
 - a. **Output folder:** The output folder location for the results. Select your workspace folder as the output folder and then click "OK" on the lower right side of the dialog box
 - b. **Year for Standard Population:** To standardize for age, you need to select a standard population. A standardized rate weight will be generated from the standard population. Select 2010 as the desired year of the standard age structure for the entire USA.
 - c. **State:** Select North Carolina as the state for which population data will be used to generate crude rates.
 - d. **Year of base population in Research Area:** Select 2010 since our data was collected from 2006 - 2008. 2010 is the year of the chosen state's base population that will be used for calculating crude rates.
 - e. **Geographic Level of Study:** Select US Census Tract as the desired geographic level of analysis for the population data.

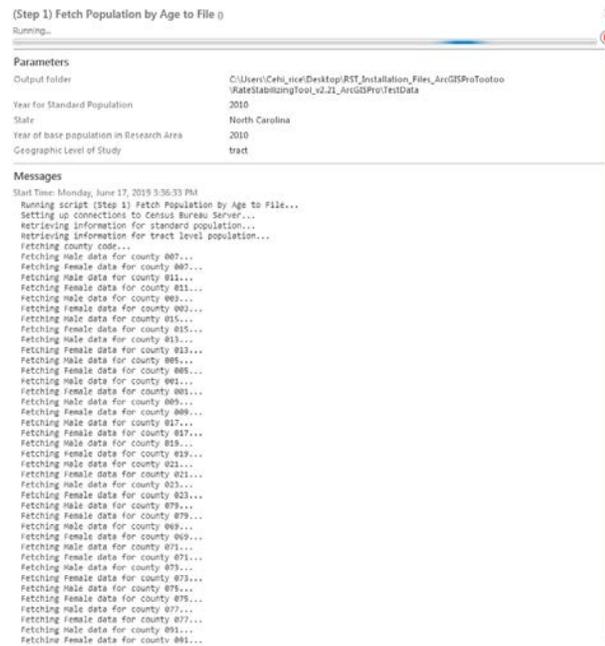


The screenshot shows the 'Parameters' tab of the tool. The settings are as follows:

- Output folder: TestData
- Year for Standard Population: 2010
- State: North Carolina
- Year of base population in Research Area: 2010
- Geographic Level of Study: tract

Rate Stabilizing Tool (RST)

Click **Run** to run the tool... and expand to show messages. You should see this:



Once the tool has successfully run (the dialogue box will say “completed script” and “succeeded”), it will generate a data file that can be used to run the tool behind your firewall. In our case, the file will be named: *RawData_state37_tract.data*. This file can be re-used for multiple analyses.

Name	Type
 RawData_state37_tract.data	DATA File
 NC_tract_2010prj.shx	SHX File
 NC_tract_2010prj.shp	XML Document
 NC_tract_2010pri.sho	SHP File

Build Neighborhood Dictionary Script: The RST tool can optionally perform spatial Bayesian smoothing when given a neighborhood dictionary. Run the *(Step 2 Optional) Build Neighborhood Dictionary* Script tool to generate a matrix of adjacency relationships between geographic units. Double left click on the icon to open the tool.

3. Let’s consider each of the parameters/inputs for the tool.
 - a. **Shapefile for Spatial Bayesian:** Select a shapefile containing all the geographic units at the level of study for the state of interest. The neighborhood dictionary will be built based on this input shapefile – **select: NC_tract_2010prj.shp shapefile.**

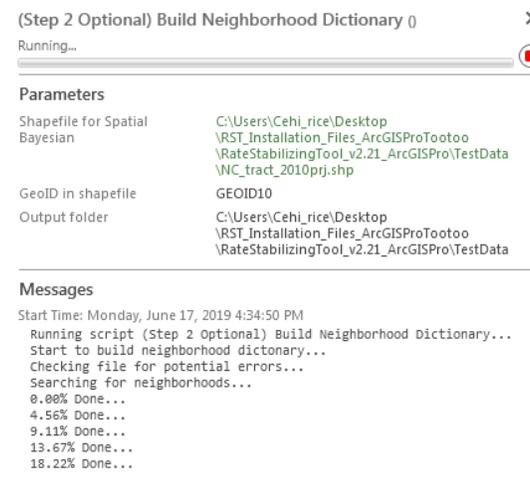
*You can download geographic boundaries for your state from the Census website (<https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>). Be sure the boundaries have **not** been generalized or simplified for cartographic display.*

Rate Stabilizing Tool (RST)

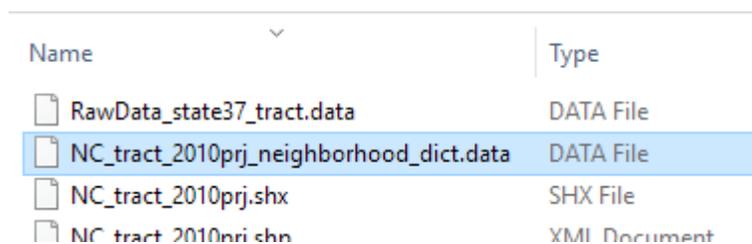
- b. **GeoID in shapefile:** Index the field name containing each geographic unit's GeoID. In order to generate a valid neighborhood dictionary, you need a field containing a unique identifier for each census geography that aligns with the census ID – **select: GEOID10.**
- c. **Output folder:** The output folder location for the results. Single click your workspace folder and then click **Add** on the lower right side of the dialog box.



Click **Run** to run the tool. You should see something like this:



After the tool has successfully run, it will generate a new neighborhood_dict.data file. In our case, the file will be named: *NC_tract_2010prj_neighborhood_dict.data*.



4. Now that we have the population data from the US Census to manage a proper age standardization tool locally and a neighborhood dictionary for optional spatial smoothing, we can run the *(Step 3) Rate Stabilizing with Local Data* Script tool. Double left click on the icon to open the tool.

Rate Stabilizing Tool (RST)

5. Let's consider each of these parameters/inputs for the tool.
 - a. **Input de-identified record level death data:** Select your records table for calculation. The age and GEOID fields are required for the calculation. Please de-identify the data before using the tool.
 - i. Browse to the *NC_HeartDisease0608_simulate.dbf* file.
 - b. **Output folder:** Select the output folder location for the results.
 - i. Please single click your desired folder and then click **Add**.
 - c. **Geographic Identification Field:** Index the field name containing the GeoIDs.
 - d. **Age field:** Index the field name of the patients' age.
 - i. In the case of the test data: Age_Mask
 - e. **Downloaded Raw Data:** Point the tool to the folder that you have placed the downloaded Census data (*RawData_state37_tract.data*).
 - i. Remember in generating this data you have:
 1. Selected Year for standard US population as 2010. In order to adjust for age, a standard population must be selected. The adjusted weight will be generated from this standard population.
 2. Selected the State of interest: NC or 37.
 3. Selected the Year of base population as 2010 for NC.
 4. Specified the geographic level of study: Tract.
 - f. **Age Structure:** Specify your own age group structure. Enter the lower-bound of each age group. One negative value can be used to control the max age of consideration.
 - i. For example, for the following group structure:
 - a. Younger than 1 years,
 - b. 1-5 years,
 - c. 6-20 years,
 - d. 21-45 years,
 - e. 46-60 years,
 - f. Over 61 years

The following numbers should be entered: 0, 1, 6, 21, 46, and 61 (pressing enter to add the next value).

- ii. For an age group structure:
 - a. Younger than 5 years,
 - b. 5-10 years,
 - c. 10-17 years

The following numbers should be entered: 0, 5, 10, and -18.

Note: The test dataset that comes with the tool is simulated data with masked age (i.e., all ages are changed to one specific value in their corresponding age group). If you are using the test data, please use the following age structure:

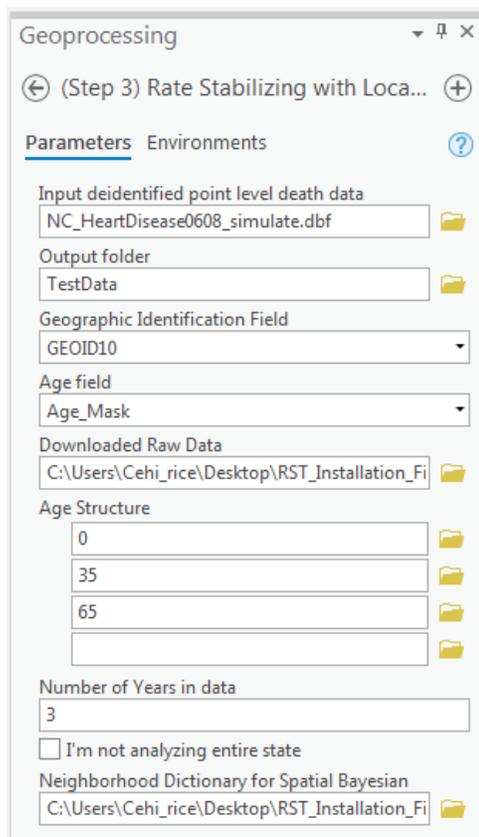
- a. Younger than 35 years,
- b. 35-65 years,
- c. Over 65 years

Rate Stabilizing Tool (RST)

The following numbers should be entered: 0, 35, 65.

- g. **Number of Years in data:** Enter in the total number of years represented by the input death data. Decimals can be used to represent partial years.
 - i. For our example, enter 3.
- h. **I'm not analyzing the entire state:** Check this if your input data only reflects a portion of the state's geographies rather than the entire state. Why? If we tell the tool that the base population is for the entire state, it will smooth to that population. If the data only comes from select counties, we only want to consider the base population for these counties combined.
 - i. For our example, **uncheck** this box since we are analyzing data from every county in North Carolina.
- i. **Neighborhood Dictionary for Spatial Bayesian (optional):** If you would like to include spatial Bayesian smoothing, select the neighborhood dictionary data file built in Step 2. Otherwise, leave blank.
 - i. Navigate to and select *the NC_tract_2010prj_neighborhood_dict.data* file and select "Open".

Here is what your test run of the tool should look once you have input the example parameters:



The screenshot shows the 'Geoprocessing' window for the tool '(Step 3) Rate Stabilizing with Local Data'. The 'Parameters' tab is active, showing the following settings:

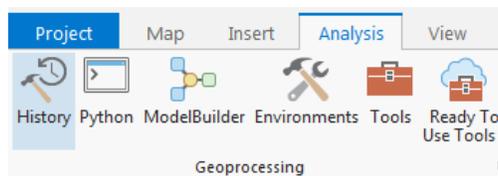
- Input deidentified point level death data:** NC_HeartDisease0608_simulate.dbf
- Output folder:** TestData
- Geographic Identification Field:** GEOID10
- Age field:** Age_Mask
- Downloaded Raw Data:** C:\Users\Cehi_rice\Desktop\RST_Installation_Fi
- Age Structure:** 0, 35, 65
- Number of Years in data:** 3
- I'm not analyzing entire state:** (unchecked)
- Neighborhood Dictionary for Spatial Bayesian:** C:\Users\Cehi_rice\Desktop\RST_Installation_Fi

Click **Run** to run the tool.

Rate Stabilizing Tool (RST)

Let's talk about what is going on while the script is running:

- The tool is using the Census data file you created to calculate the population in each age group and the standard age structure.
- The tool is calculating crude rates for each age category you specified. Once these are calculated, age-standardized rates (both smoothed and non-smoothed) are calculated for each age category and standard age structure.
- If there are GeolIDs (census identifiers in this case) that are missing, or cannot be recognized in census data, you will receive a warning.
- All output file paths can be viewed under the Messages section in the **Analysis > History Pane**.



(Step 3) Rate Stabilizing with Local Data () X

Completed Today at 5:17:35 PM

Parameters

Input deidentified point level death data	C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData\NC_HeartDisease0608_simulate.dbf
Output folder	C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData
Geographic Identification Field	GEOID10
Age field	Age_Mask
Downloaded Raw Data	C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData\RawData_state37_tract.data
Age Structure	0;35;65
Number of Years in data	3
I'm not analyzing entire state	
Neighborhood Dictionary for Spatial Bayesian	C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData\NC_tract_2010prj_neighborhood_dict.data

Messages

Start Time: Monday, June 17, 2019 5:15:24 PM

Running script (Step 3) Rate Stabilizing with Local Data...
 constructing standard population structure...
 Calculating population by input age structure...
 Constructing disease/death rate from individual records...
 Calculating age adjusted rate...
 Spatial smoothing the results...
 Age adjusted rate successfully calculated with no error!!!
 The average rate for the area is 178.18882447062816 cases per 100,000

File Path:
 Age Adjust Rate: C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData\age_adjust_NC_HeartDisease0608_simulate.csv
 Age/Pop Table: C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData\PopAge_structure_state37.csv
 Age Structure: C:\Users\Cehi_rice\Desktop\RST_Installation_Files_ArcGISProTootoo\RateStabilizingTool_v2.21_ArcGISPro\TestData\Standard_Age_structure.csv
 completed script (Step 3) Rate Stabilizing with Local Data...

Succeeded at Monday, June 17, 2019 5:17:35 PM (Elapsed Time: 2 minutes 10 seconds)

Rate Stabilizing Tool (RST)

Understanding the output:

1. Once the tool has completed all processes, 3 tables will be placed in the **Output** folder you specified. These include:
 - a. Two intermediate tables required for calculations which can be used for verification:
 - i. **PopAge_structure_state37.csv**
This file includes the population for each geographic unit in the state of interest by age category. You defined these age categories when you set up the tool.
 - ii. **Standard_Age_structure.csv**
This file includes the proportion of the population in each age category in the standard year of choice. This age structure was calculated using the population data from the whole United States. The structure works as the weight when calculating the weighted average for age-standardized mortality rate.
 - b. And the table with your results: **age_adjust_NC_HeartDisease0608.csv**
2. Add the results table to your map and open it up to see what has been produced.

Age_adjust_rate: the non-smoothed rate per 100,000;

SpBay_AAR: the spatially smoothed rate per 100,000;

SpBay_2p5 & SpBay_97p5: the lower and upper boundaries for 95% confidence interval for spatially smoothed age-standardized rate;

Baye_AAR: the smoothed rate per 100,000; and

Baye_2p5 & Baye_97p5: the lower and upper boundaries for 95% confidence interval for smoothed age-standardized rate.

Potential Alert Messages: **When the width of the confidence interval (Upper limit – Lower limit) is larger than the estimate, the estimate is unreliable. These rates should NOT be mapped.**

- a. Alert:Unreliable Empirical Bayesian Estimate!!!! – The empirical Bayesian estimate is not reliable in the region.
- b. Alert:Unreliable Spatial Bayesian Estimate!!!! – The spatial Bayesian Estimate is not reliable in the region.
- c. Alert:Unreliable Estimate!!!! – Both estimates are not reliable in the region.

NspUnreli: New after version 2.15. This field indicates the reliability of non-spatial Bayesian estimates. Non-spatial Bayesian will be reliable if NspUnrel is 0.

SpUnreli: New after version 2.15. This field indicates the reliability of spatial Bayesian estimates. Spatial Bayesian will be reliable if SpUnrel is 0.