\*\*\* Files needed for exercise: *NC\_HeartDisease0608.dbf, and North Carolina census tract shapefile: NC\_tracts\_2010prj.shp*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Goals:** Consider and use a tool to produce age-adjusted, smoothed sub county estimates.

Interpreted the output and results of the tool.

**Objectives:**

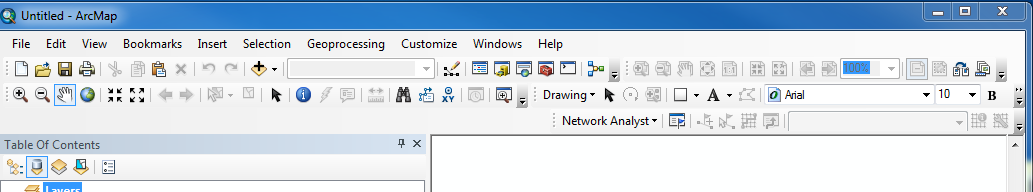
* Gain experience installing the tool and managing the user interface;
* Develop an understanding of the required data inputs; and
* Consider and interpret the output

**Problem:** You have been provided with individual level heart disease death data for the years 2006 to 2008 that has been geocoded to the US 2010 Census tract level. Using the Rate Stabilizer Tool (RST) you will generate age adjusted, 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.

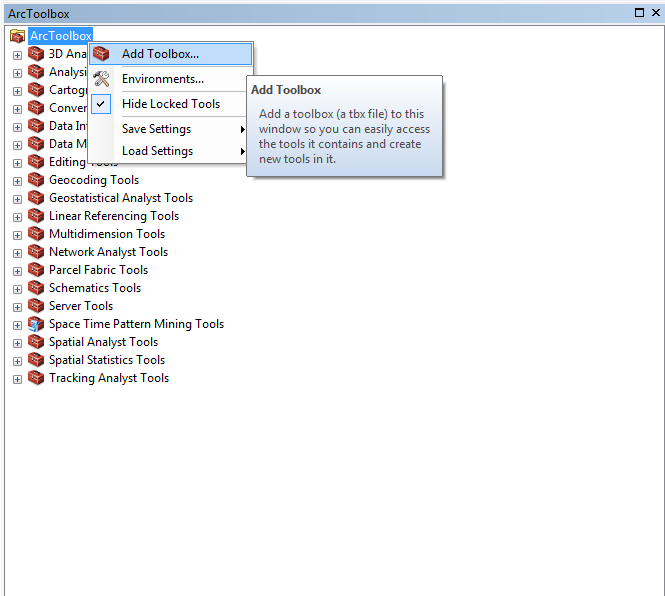
**Loading the toolbox**

1. Open a new blank map in ArcMap.

Activate the ArcToolbox window by left clicking on the toolbox icon. This tool has been built using Python and can be added as a toolbox into the ArcGIS Desktop interface.

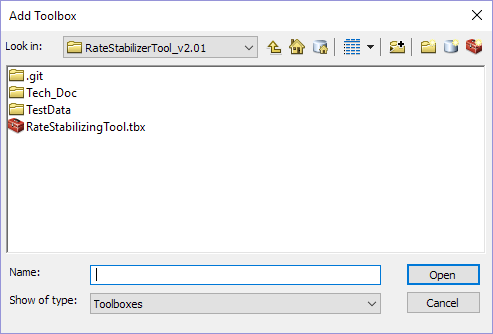


1. Right click on the ArcToolbox icon again in the ArcToolbox window and select Add toolbox…



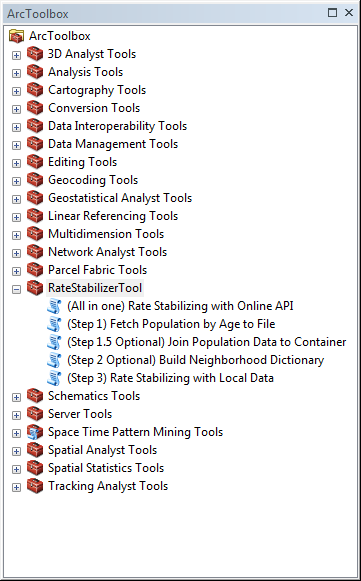
1. Navigate to the location where you have save the extracted folder locally on your machine

*RateStabilizerTool* > *RateStabilizingTool.tbx* and click *open*



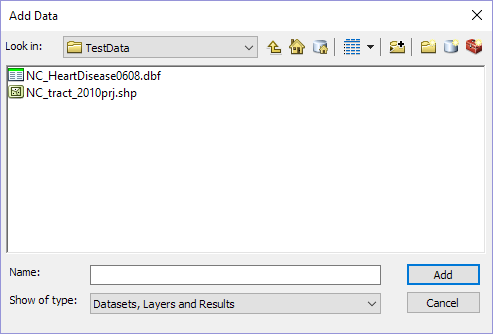
1. The ***RateStabilizingTool*** toolbox will now appear, expand the toolbox to reveal the

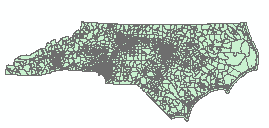
Script tools in this toolbox



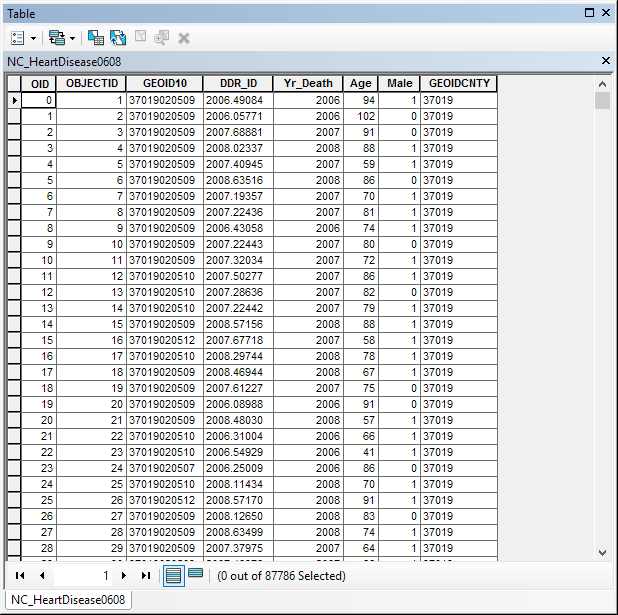
**Adding your data**

1. The *RateStabilizerTool TestData* folder contains a shapefile*:* ***NC\_tracts\_2010prj.shp****.* Add thisshapefile for some geographical context.





1. Next add the record level heart disease death data: ***NC\_HeartDisease0608.dbf***



Open this table and take a look at the attribute fields:

**OID** – an automatically generated object identifier;

**GEOID10** – the geographic identifier for the US Census tract unit that you will be calculating a stabilized age adjusted 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 adjusted rate

**Using the tool**

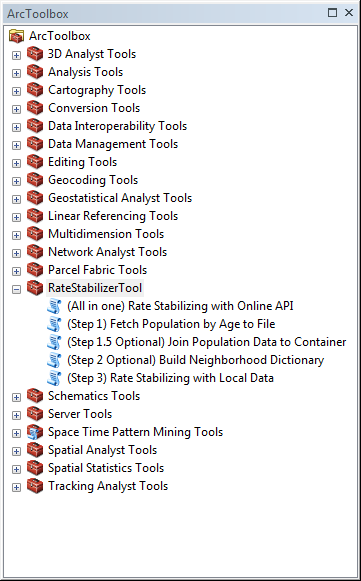
Since this tool makes use of the US Decennial Census API to obtain census population data for age adjustment calculation, we offer two options of running this tool depending on your situation.

Option One: You may run the all-in-one RST tool in an open environment (with internet access) – using census population data that has been fetched dynamically via US Decennial Census API.

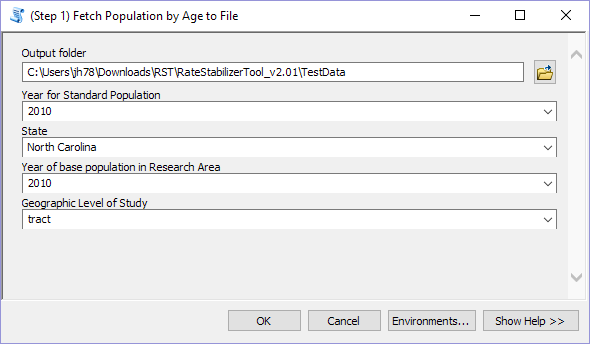
or

Option Two: You may run the 3-step RST tool in a secure environment – using census population data that has been fetched, downloaded and saved in a secure location (behind a firewall for example).

1. Let’s use Option Two assuming that you want to run the RST in a secure environment. To do this you will need to first run the *(Step 1) Fetch Population by Age to File* Script tool. Double left click on the icon to open the tool.



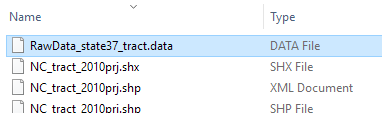
1. Let’s consider each of the parameters/inputs.
   1. **Output** **folder:** The output folder location for the results.
      1. Please single click your desired folder and then click "Add" on the lower right side of the dialog box
   2. **Year for Standard Population:** In order to adjust for age, a standard population must be selected. The adjusted weight will be generated from the standard population. Select the desired year of the standard age structure for the entire USA.
      1. We will use 2010.
   3. **State**: Select the state for which population data will be used to generate crude rates.
      1. Choose North Carolina.
   4. **Year of base population in Research Area**: Select the desired year of the chosen state’s base population that will be used for calculating crude rates.
      1. We will use 2010.
   5. **Geographic Level of Study**: Select the desired geographic level of analysis for the population data.
      1. We will perform the analysis at the tract level.



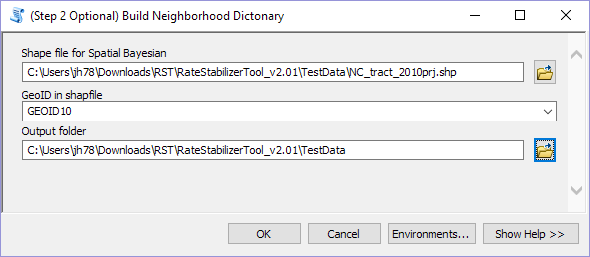
Click OK to run the tool… and you should see this:



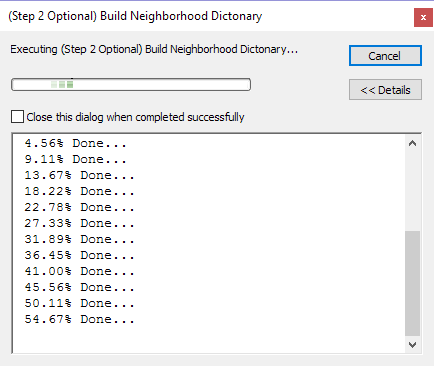
Once the tool has successfully run, it will generate a data file that can be used to run the tool behind your firewall. In our case the file will be named: RawData\_state37\_tract.data.



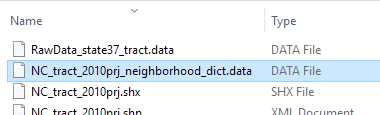
1. The RST tool can optionally perform spatial Bayesian smoothing when given a neighborhood dictionary. Run the *(Step 2 Optional) Build Neighborhood Dictionary* Script tool to generate a matrix of adjacency relationships between geographic units. Double left click on the icon to open the tool.
2. Let’s consider each of the parameters/inputs for the tool.
   1. **Shapefile for Spatial Bayesian:** Select a shapefile containing all the geographic units at the level of study for the state of interest.
      1. Add the *NC\_tract\_2010prj.shp* shapefile.
   2. **GeoID in shapefile:** Index the field name containing each geographic unit’s GeoID.
      1. Select GEOID10.
   3. **Output** **folder:** The output folder location for the results.
      1. Please single click your desired folder and then click "Add" on the lower right side of the dialog box



Click **OK** 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.



1. Now that we have the population data from the US Census to manage a proper age adjustment 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.
2. Let’s consider each of these parameters/inputs for the tool.
   1. **Input deidentified point level death data:** Select your records table for calculation. The age and GEOID fields are required for the calculation. Please de-identify the point level data before using the tool.
      1. Browse to the NC\_HeartDisease0608.dbf file.
   2. **Output** **folder:** Select the output folder location for the results.
      1. Please single click your desired folder and then click "Add" on the lower right side of the dialog box.
   3. **Geographic Identification Field**: Index the field name containing the GeoID’s.
   4. **Age field**: Index the field name of the patients' age.
   5. **Downloaded Raw Data**: Point the tool to the folder that you have placed the downloaded Census data (RawData\_state37\_tract.data).
      1. Remember in generating this data you have:
         1. Selected Year for standard US population 2010. In order to adjust for age, a standard population must be selected. The adjusted weight will be generated from standard population.
         2. Selected theselect State of interest: NC or 37.
         3. Selected the Year of base population 2010 for NC.
         4. Specified the geographic level of study: Tract .
   6. **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.
      1. For example, for the following group structure:
         * 1. Younger than 1 years,
           2. 1-5 years,
           3. 6-20 years,
           4. 21-45 years,
           5. 46-60 years,
           6. Over 61 years

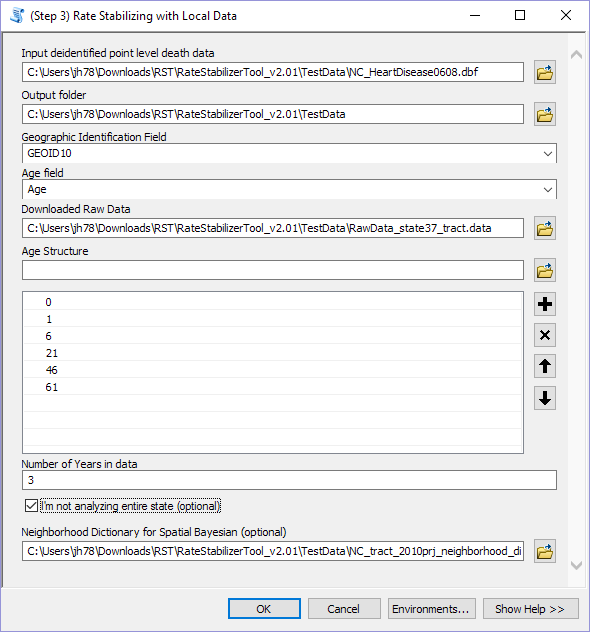
The following numbers should be entered: 0, 1, 6, 21, 46, 61 (one line at a time).

* + 1. For an age group structure:
       - 1. Younger than 5 years,
         2. 5-10 years,
         3. 10-17 years

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

* 1. **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.
     1. For our example, enter 3.
  2. **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.
     1. For our example, check this box since we are analyzing data from a select set of counties rather than from every county in North Carolina.
  3. **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.
     1. Navigate to and select the NC\_tract\_2010prj\_neighborhood\_dict.data file.

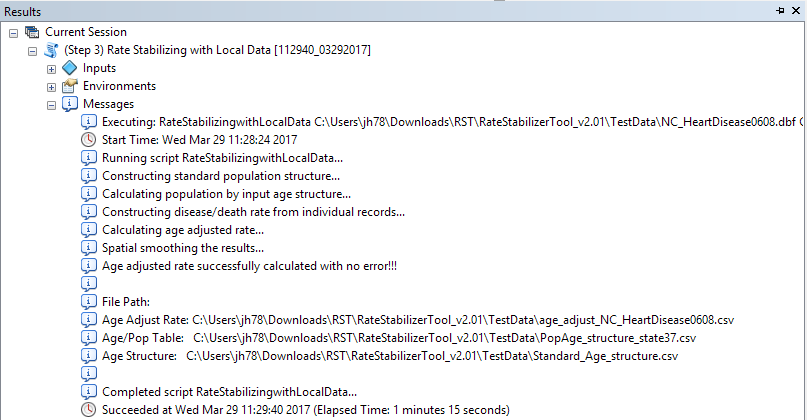
Here is what your completed test run of the tool should look once you have input the parameters:



Click **OK** to run the tool.

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 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 adjusted rates (both smoothed and non-smoothed) are calculated for each age category and standard age structure.
* If there are GeoIDs (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 Results window.



**Understanding the output:**

1. Once the tool has completed all processes, 3 tables will be places in the **Output** folder you specified. These include:
   1. Two intermediate tables required for calculations which can be used for verification:
      1. ***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.

* + 1. ***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 adjusted mortality rate.

* 1. And the table with your results: ***age\_adjust\_NC\_HeartDisease0608.csv***

1. Add the results table to ArcMap and open it up to see what has been produced.

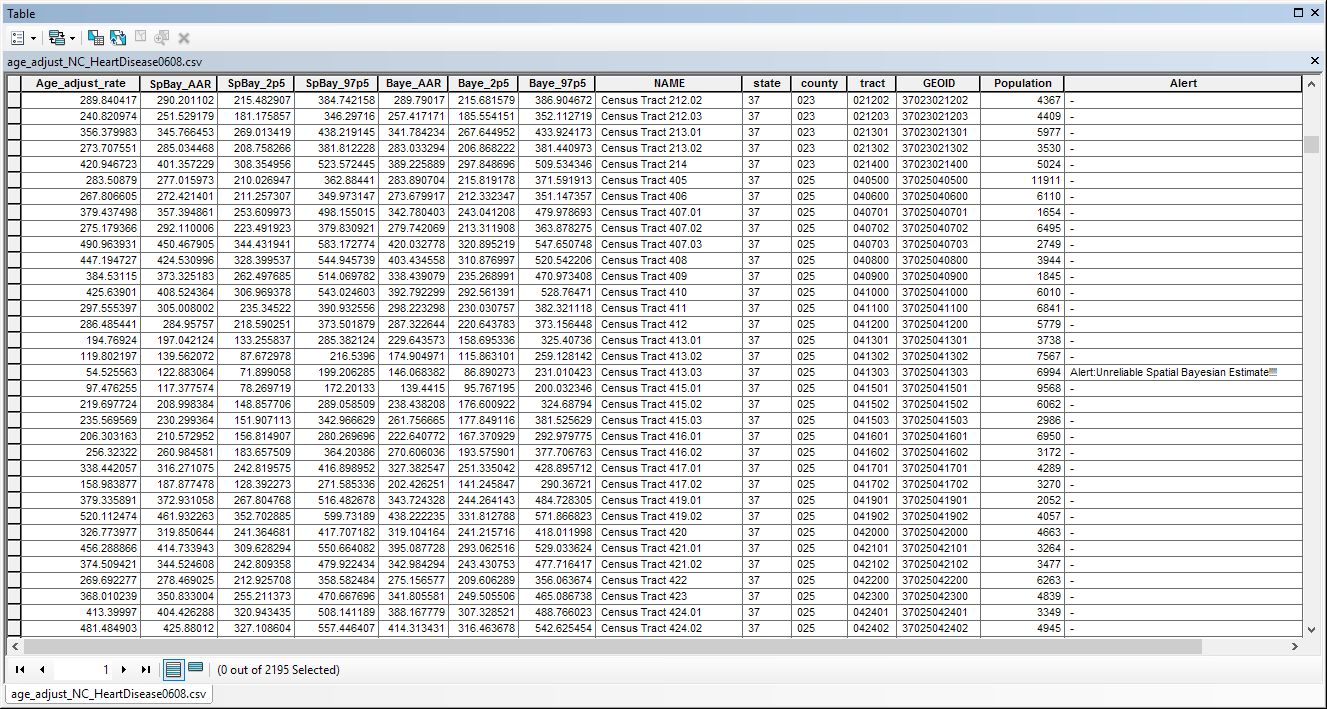
**Age\_adjust\_rate**: the non-smoothed rate per 100,000;

**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 adjusted 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 adjusted rate.



**NSpUnreli:** this field will be 0 when non-spatial Bayesian estimate is reliable.

**SpUnreli:** this field will be 0 when spatial Bayesian estimate is reliable.

**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.

1. Alert:Unreliable non-Spatial Bayesian Estimate!!!! – The empirical Bayesian estimate is not reliable in the region. In this case, **NSpUnreli** will be 1 and **SpUnreli** will be 0.
2. Alert:Unreliable Spatial Bayesian Estimate!!!! – The spatial Bayesian Estimate is not reliable in the region. In this case, **NSpUnreli** will be 0 and **SpUnreli** will be 1.
3. Alert:Unreliable Estimate!!!! – Both estimates are not reliable in the region. In this case, both **NSpUnreli** and **SpUnreli** will be 1.