1

PopGen – 1.1

SETTING UP AND SYNTHESIZING POPULATION USING CENSUS DATA

2

Instructions for setting up a project in PopGen

- This presentation contains the screen shots of each of the steps involved in the setting up a project for population synthesis in PopGen using census data

3

Project Details

4

Project Details

To create a new project, go to ***File*** ➡ ***New Project*** on the menu bar at the top, or by clicking the ***New Project*** icon at the top left of the page

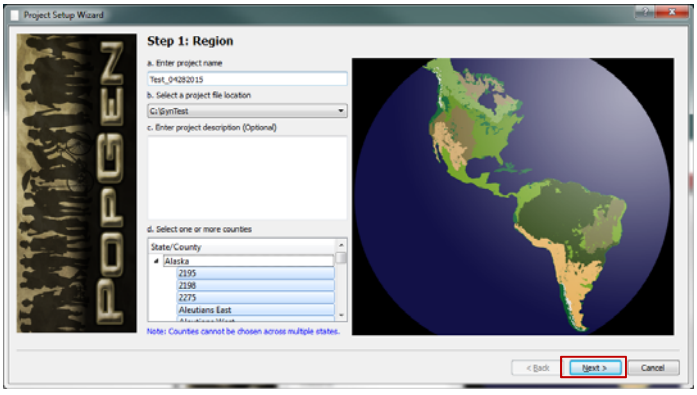
5

Project Details

- Choose a name for your project – (Example: ***Test_04282015***)
- Select a project file location on your computer
- Enter a project description (Optional)
- Select all counties in a state for which you want to generate synthetic population
  - If you want to select multiple counties, hold the **Ctrl** button while selecting. All selected counties must be within a single state
  - To select all counties in a state, double click on the name of the state

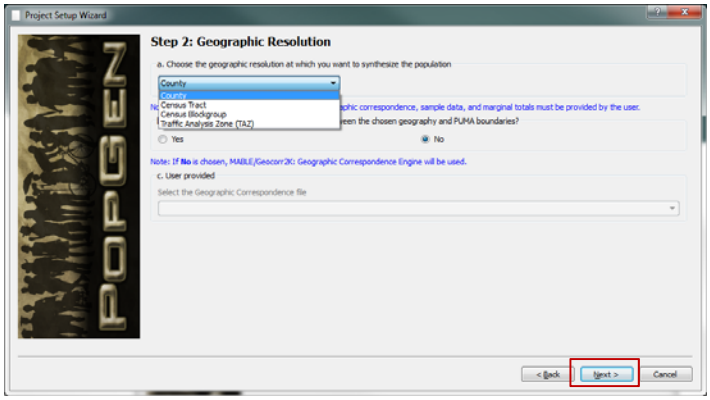
6

Project Details



7

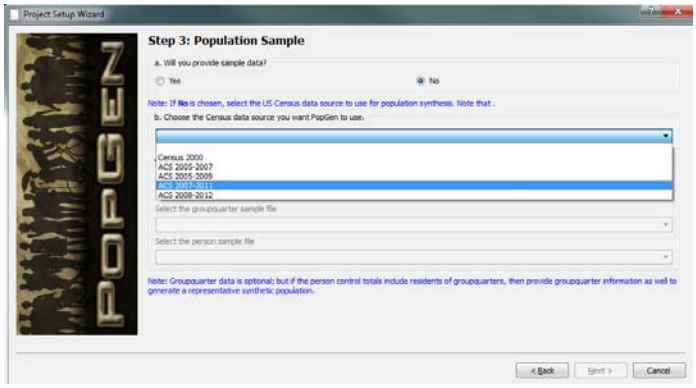
Project Details



- Select the geographical resolution at which you want to generate the synthetic population
- Using data from census, synthetic population can be generated at county, census tract and blockgroup levels
- For generating data at the TAZ level, a file that provides correspondence between the geography and PUMA (from where the population will be sampled) should be provided by the user

8

Project Details

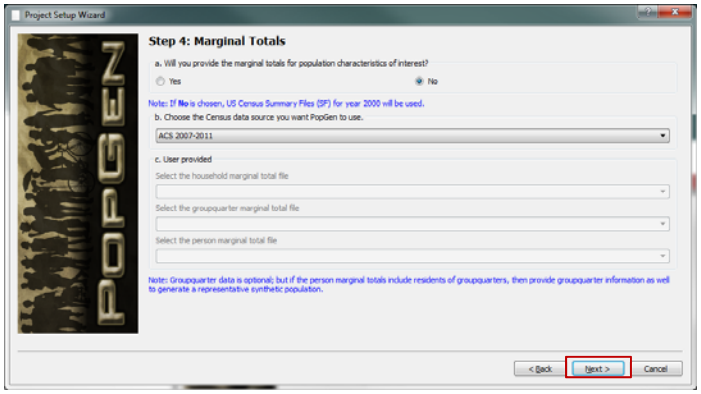


Select the source for sample data

Note: **ACS 2007-2011** sample data will be used when **ACS 2008-2012** is selected, as the 2012 sample data is coded to the 2010 PUMA system whereas PopGen 1.1 requires sample data coded to the **2000 PUMA** system. The marginal data however will be from 2008-12.

9

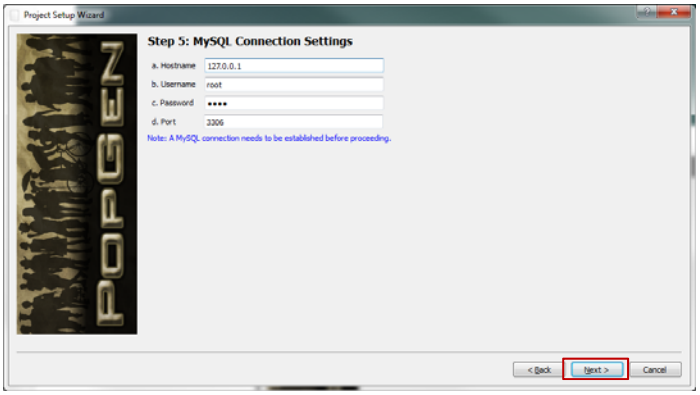
Project Details



Select the source for marginal data

10

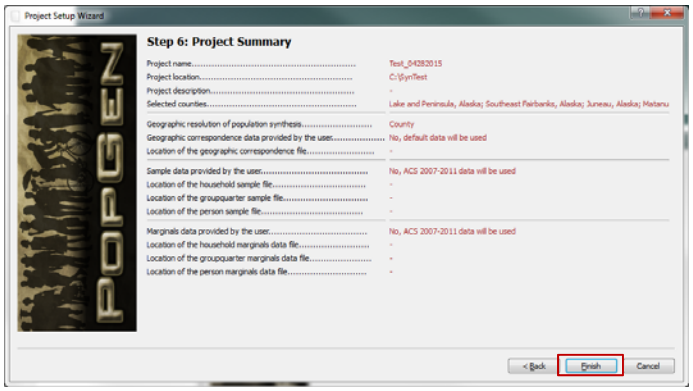
Project Details



- The hostname should be set to 127.0.01
- The default password is '1234'. *Do Not* change this
- The default port is '3306'

11

Project Details



- A window showing all the previous steps with the selected options is displayed
- Review and ensure that everything is correct
- Click ***Finish*** to complete the project setup process

12

Data Import

13

Data Import

Go to ***Data*** ➡ ***Import*** at the top of the screen to start importing data or click the ***Data Import icon*** in the toolbar

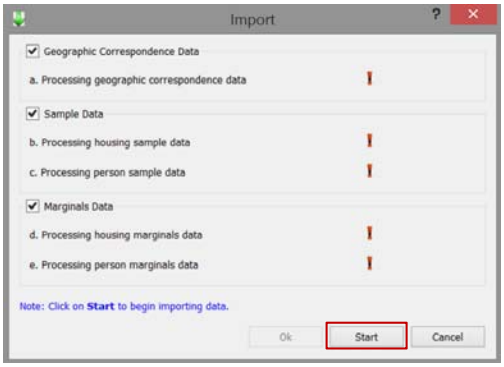
14

Data Import

- Click ***Start*** to begin the import process
- This step downloads the census files, parses them, and then uploads them into the MySQL database as tables
- While importing, you will be prompted to overwrite any files that already exist, and to select specific variables to be imported into the database

15

Data Import



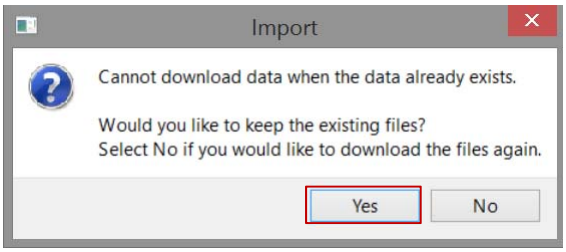
16

Data Import

- If you are prompted to keep data files that already exist, you may choose to download the data again (Choose No), or keep the existing data (Choose Yes) that has already been downloaded
  - Click **Yes**, in case you are running the software on a machine that is offline (assuming you have already copied the data to the pertinent location)
- If you are prompted to replace existing files, you may choose to keep the old files (**No to All**) or replace them (**Yes to All** → suggested)

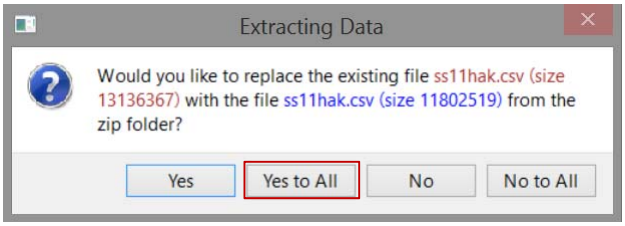
17

Data Import



18

Data Import



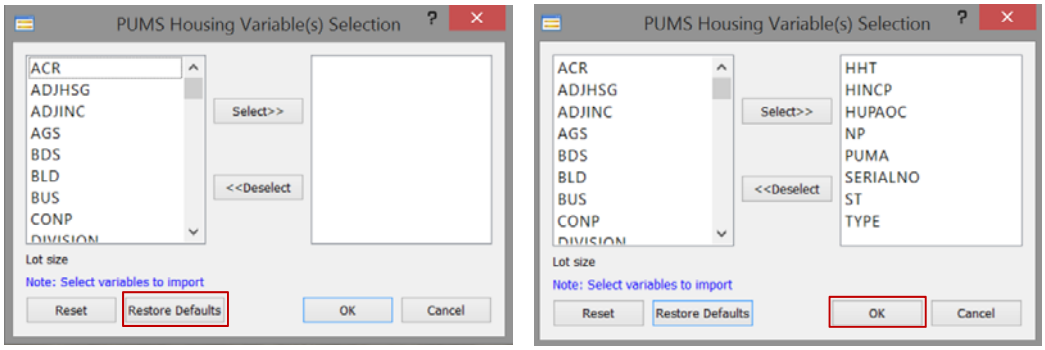
19

Data Import

- While importing the data, there will be a series of variable selection windows that will appear
- For each one of these windows, select the ***Restore Defaults*** option at the bottom left corner of the window
- This option will automatically select the default variables from the master list of variables

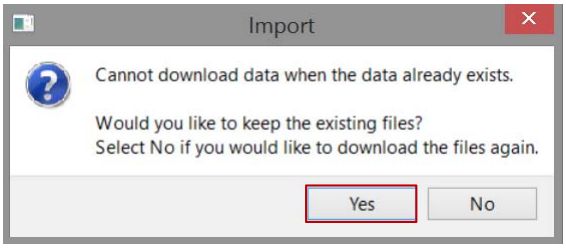
20

Data Import



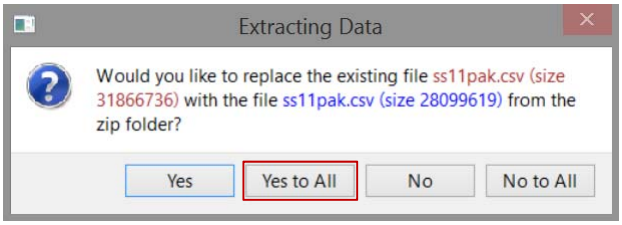
21

Data Import



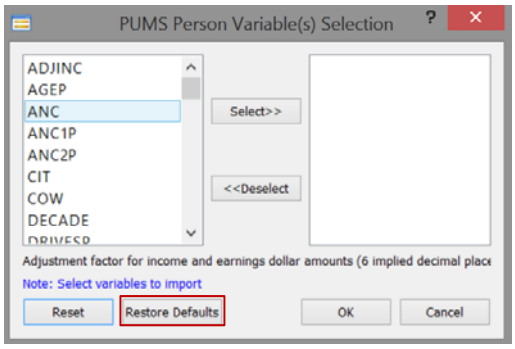
22

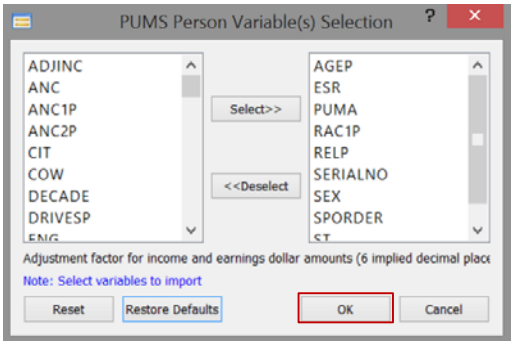
Data Import



23

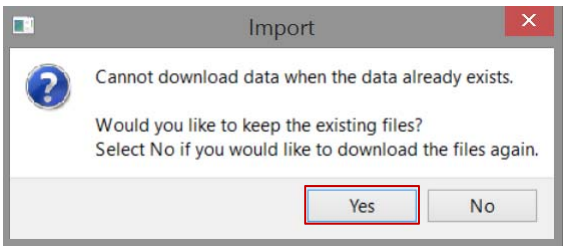
Data Import





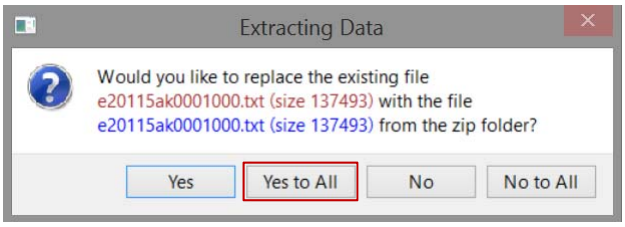
24

Data Import



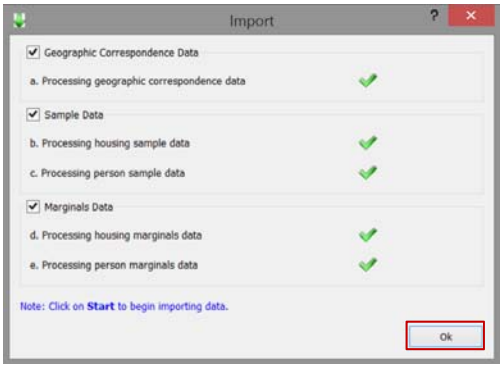
25

Data Import



26

Data Import



27

Data Transformation

28

Data Transformation

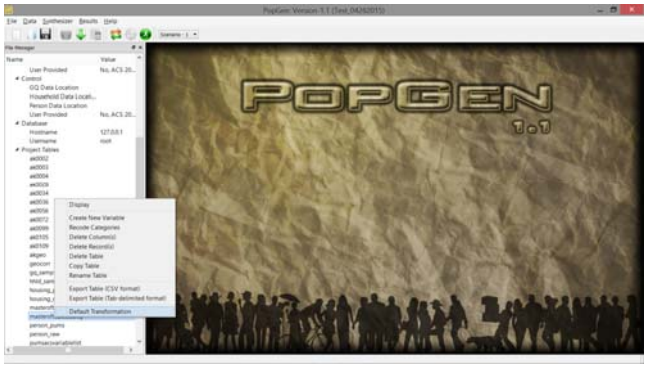
- Right-click on the table titled *'housing_pums'* under *'Project Tables'* in the *'File Manager'*
- Select ***Default Transformation*** at the bottom
- The default transformation will take a few moments to complete
- Repeat this step for the *'mastersftable<geographical_resolution> and person_pums'* (in the same order) tables located in the *'Project Tables'* list

29

Data Transformation

30

Data Transformation



31

Data Transformation

32

Data Transformation

Default Control Variables from Census Data

Household	Person	Group quarter
Household Income	Employment Status*	Institutional Group Quarter
Household Size	Age	Non-institutional Group Quarter
Presence of Children*	Gender	
# Workers in the Household*	Race	
Householder Age		
Household Type		

* Marginal data not available at the census block-group level

33

Group Quarter (GQ) Population Marginals at the Blockgroup level

1. Group Quarter (GQ) population is available at the county and census tract levels in ACS Summary files. However, GQ population is not readily available at the block-group level.
2. Therefore, GQ population is obtained at the block-group level using the following equation:
   - total population – (hhldsize1*1 + hhldsize2*2 + hhldsize3*3 + hhldsize4*4 + hhldsize5*5 + hhldsize6*6 + hhldsize7*7.98)
3. GQ population is categorized into two groups (Institutional and non-institutional) using the following steps:
   - Compute the proportion of institutional and non-institutional GQs for each PUMA (from ACS sample data)
   - Obtain the institutional and non-institutional GQ population totals in a geography (county, tract, or block-group) by multiplying the total GQ population (from step 2) with the proportions computed from above

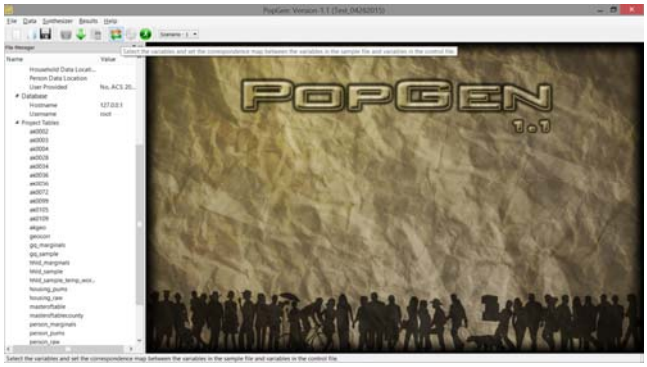
Note: All zones (county, tract, or block-group) that fall within a PUMA get the same proportion of institutional and non-institutional GQ population

34

Data Correspondence

35

Data Correspondence



Go to ***Synthesizer*** ➡ ***Set Corresponding Variables*** in the menu bar at the top or click the icon

36

Data Correspondence

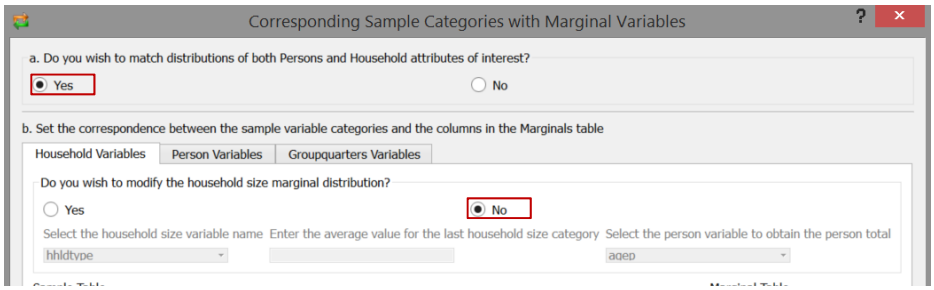
- Select ***Yes*** for "Do you wish to match distributions of both Persons and Household attributes of interest?"

  Note: If you select ***No***, the synthesizer will not control for person attributes of interest

- Under the '*Household Variables*' tab, ***No*** will be selected by default
  - If you select ***Yes***, the program will try to mitigate person total inconsistencies by the modifying household marginal distributions

37

Data Correspondence



38

Data Correspondence

- To create correspondences between the variables, select the appropriate variables from the *'Sample Variable'* column and click ***Select***
- The variables now appear in the *'Selected Variable'* column and are ready to be assigned correspondences
- This process is meant to establish correspondences between variables and their categories in the sample files and variables and their categories in the marginal files

39

Data Correspondence

- Next, select a variable in the *'Selected Variable Categories'* column along with its matching variable in the *'Marginal Variables'* column and click ***Add Correspondence***
- Repeat this step for all of the variables until they are all given correspondences
- The same procedure needs to be repeated for the *'Person Variables'* and *'Groupquarters Variables'* before continuing
- Select ***OK*** when all correspondences have been set

40

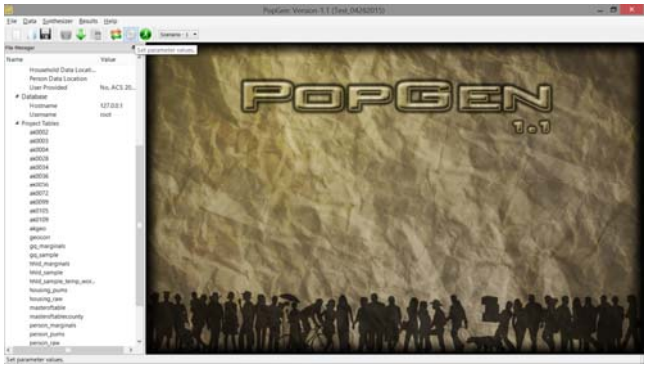
Data Correspondence

41

Parameters

42

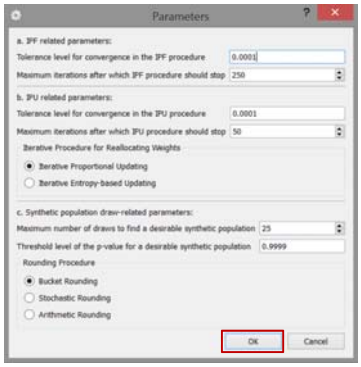
Parameters



Go to ***Synthesizer*** ➡ ***Parameters/Settings*** in the menu bar at the top or click the ⚙ icon

43

Parameters



44

Running the Synthesizer

45

Running the Synthesizer

Go to ***Synthesizer*** ➡ ***Run*** in the menu bar at the top or click the icon in the toolbar

46

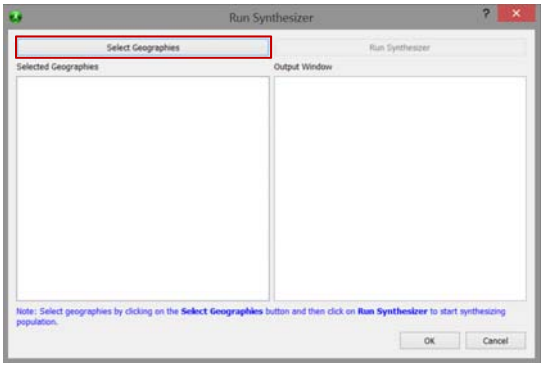
Running the Synthesizer

- Once the synthesizer window opens, click ***Select Geographies***
- A new window opens with a list of all census tracts (or individual geographies corresponding to the chosen geographic resolution)
- Select all the geographies you intend to run by holding the ***Ctrl*** button or select all geographies using the ***Ctrl + A***
- Click ***OK*** to continue

Note: You can choose as few or as many geographies as you like.

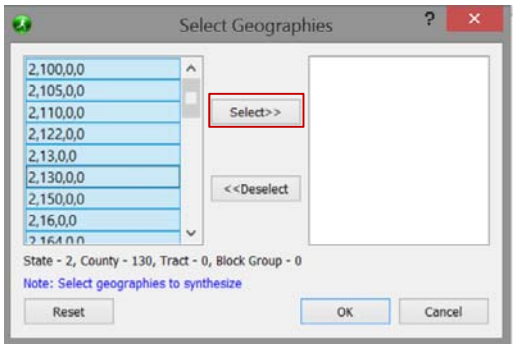
47

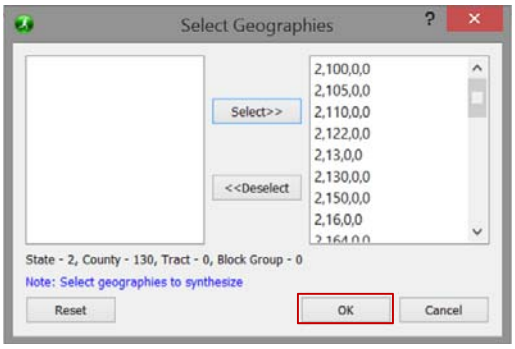
Running the Synthesizer



48

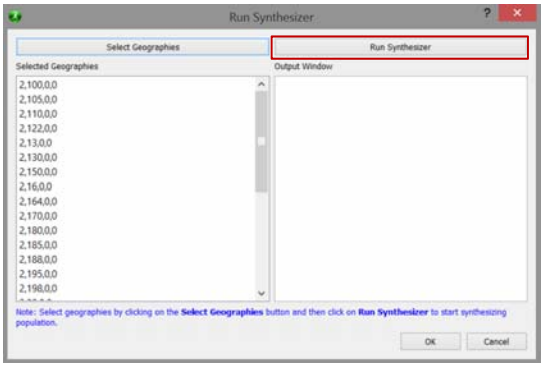
Running the Synthesizer





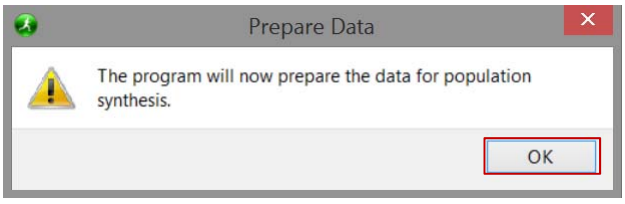
49

Running the Synthesizer



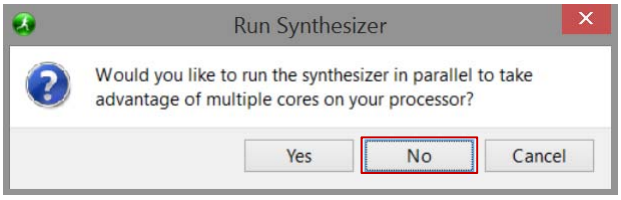
50

Running the Synthesizer



51

Running the Synthesizer



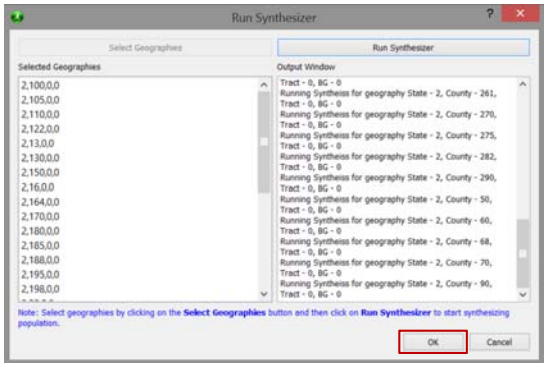
52

Running the Synthesizer

The synthesizer window appears frozen when the run is in progress. This is *OK!*

53

Running the Synthesizer



Once the run is completed, the '*Output Window*' on the right hand side of the *'Run Synthesizer Window'* will display each geography that was processed

54

Thank You!