Table of Contents

[Introduction 2](#_Toc368302643)

[Deliverable 2](#_Toc368302644)

[Theoretical Background and Model Framework 3](#_Toc368302645)

[Software Installation and Usage 6](#_Toc368302646)

[Python Installation 6](#_Toc368302647)

[ObesityPrevalence Simulator 7](#_Toc368302648)

[Using ObesityPrevalence Simulator 7](#_Toc368302649)

[Sample Data 10](#_Toc368302650)

[Software Usage 11](#_Toc368302651)

[Analysis of Output Data 13](#_Toc368302652)

[Data Analysis Example 17](#_Toc368302653)

[Simulation Example 18](#_Toc368302654)

[Concluding Remarks 19](#_Toc368302655)

[References Cited 20](#_Toc368302656)

# Introduction

Future obesity prevalence has been predicted by statistical models and simple dynamic models that predict only the size of the obese population as a whole without further distinguishing the population to various levels of obesity. The simple models from current literature (e.g., Finkelstein et al., 2012; Wang et al., 2008) are often too simplified in two ways:

1. modeling future trends of obese population at a geographic scale that is often too coarse to be useful in revealing area disparities,
2. missing important factors, such as death rates, birth rates of the population, and more importantly, and
3. lumping all levels of normal weight/overweight/obese/extremely obese subpopulations into one.

As such, the results of statistical analysis and predictions have little practical use in assisting policy-making process by public health districts when designing and implementing more geographically- and temporally-focused intervention programs.

# Deliverable

A CD-ROM (e.g., X:\ drive) contains the following folders is submitted along with this report:

* [X]:\Report\report
  + This report.
* [X]:\Software\ObesityPrevalence
  + *ObesityPrevalence Simulator*
* [X]:\Software\Python 2.7
  + Mac 32bits
  + Mac 64bits
  + Windows 32bits
  + Windows 64bits

# Theoretical Background and Model Framework

This project aims at developing a tool that allows predictive modeling of obesity prevalence at a small geographic scale for revealing area disparities of obesity prevalence in Summit County, Ohio. While most studies of obesity use census tracts as the geographical unit of analysis, it is often still too coarse in understanding how disparities in obesity prevalence at neighborhood level. As such, this project adopted census block groups as the geographic unit of analysis.

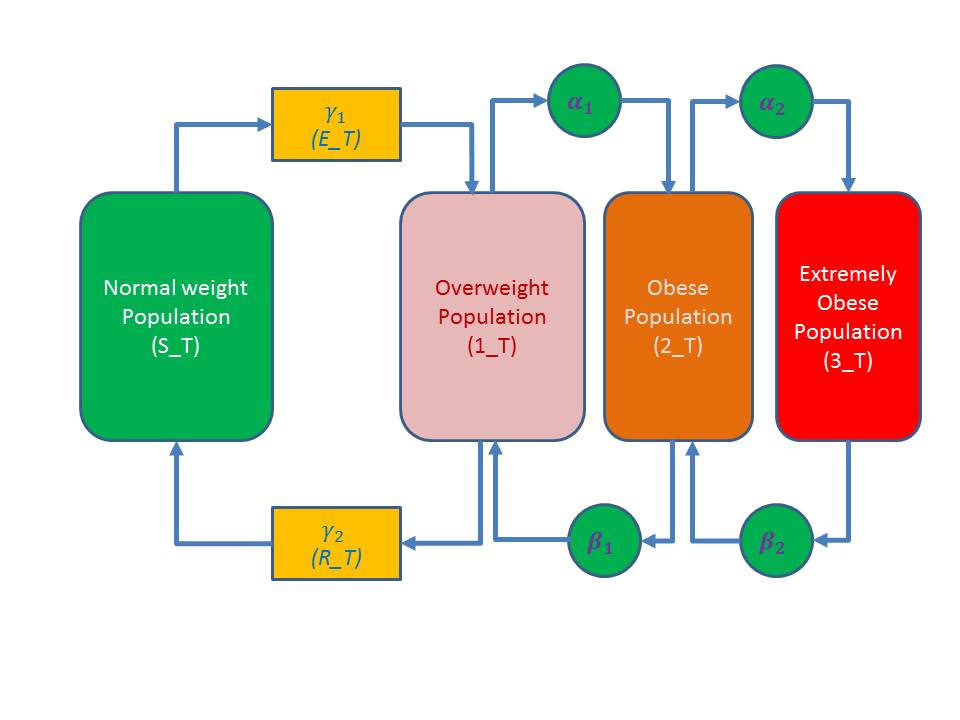
With 2011 data from American Community Survey (Census Bureau, 2011) and the 2008-2013 BMI data from the Bureau of Motor Vehicles, Summit County has 452 census block groups with a wide spectrum of obesity ratios (ranging from 16 per 1,000 population to 549 per 1,000 population) and overweight ratios (ranging from 32 per 1,000 population to 541 per 1,000 population).

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Number of Obese Population  per 1,000 | Number of Overweight Population  per 1,000 |

*Data Sources: BMI data from Ohio Bureau of Motor Vehicles, 2008-2013; Population data from American Community Survey of the US Census Bureau, 2011.*

As can be seen in the two maps above, (1) obese population, though still are in lower ratios than those of overweight population, does seem to have a geographic clustering patterns in the county, (2) overweight population prevails in most of the county with exceptions of only a few census block groups, and (3) the use of census block groups as the unit for geographic analysis indeed reveals more detail of how obesity prevails in the county.

We adopted the concept of the susceptible, infected, and recovered (SIR) framework to divide a population into subpopulations categorized as normal weight, overweight, obese, and extremely obese by BMI data. To estimate the population moving between these categories, we use a simulation approach that allow analysts to specify the ratios that subpopulations change in between categories. The relationships and potential movements between subpopulations are shown in the diagram below:



In each neighborhood, population is categorized into six (6) subpopulations:

* Normal weight (*S\_T*),
* Overweight (*1\_T*),
* Obese (*2\_T*),
* Extremely Obese (*3\_T*),
* Exposed (*E\_T,* or *S\_T 🡪 1\_T*), and
* Recovered (R\_T, or *1\_T 🡪 S\_T*).

The ratios that define how subpopulations move in between categories are

* *α1 (1\_T* 🡪 *2\_T),*
* *α2 (2\_T* 🡪 *3\_T),*
* *β1 (3\_T* 🡪 *2\_T),*
* *β2 (2\_T* 🡪 *1\_T),*
* *ϒ1 (S\_T* 🡪 *1\_T),* and
* *ϒ2 (1\_T* 🡪 *S\_T)*.

Following Thomas et al. (2013):

* Total population at *time0* (*TotalPopulation*) = *S\_T* + *1\_T* + *2\_T* + *3\_T* + *E\_T* + *R\_T*
* The exposed subpopulation (*E\_T*) are individuals who were exposed to either social or non-social influences that lead to weight gain and these individuals will eventually become overweight.
* The subpopulation (*R\_T*) are individuals who have weight loss under social or non-social influences.
* Social interactions between compartments are governed by the law of mass action and modeled by multiplying the population numbers in each class.
* Estimated subpopulations at *time1* can be derived as solutions for *α1, α2, β1, β2, ϒ1,* and *ϒ2* from a set of differential equations as proved in Thomas et al. (2013).
* For the purpose of modeling and simulations, initial values for model parameters are estimated from publications in the obesity literature:
* The probability of being born in obesogenic environment is set to be ***0.55*** of females of reproductive age who are overweight or obese, based on Balcan et al. (2010).
* Birth rate is set to be ***0.0144***, based on Jacobson et al. (2007).
* Baseline prevalence rates are set to be ***0.32*** for overweight, ***0.22*** for obese, ***0.03*** for strictly obese, based on Flegal et al. (2010).
* Social influence by overweight and obese are set to be ***0.4*** for overweight subpopulation and ***0.2*** for obese subpopulations, both are based on fitting to initial trends as discussed in Flegal et al. (2010).
* Spontaneous rate of weight gain to each class are set to be: exposed (***0.05***), overweight (***0.14***), obese (***0.08***), and extremely obese (***0.014***), also based on Flegal et al. (2010).
* Rate of weight loss to each class are set to be: extremely obese to obese (***0.05***), obese to overweight (***0.03***), and overweight to normal weight (***0.033***), also based on Flegal et al. (2010).
* Rate of weight regainers transitioning from normal weight to overweight is set to be ***0.04***, also beased on Felgal et al. (2010).
* Death rate of obese and extremely obese populations is set to vary between ***16.5*** to ***22*** per 1,000 population as suggested by Oizumi (2013).

Needless to say, any of the parameter values in this model can be changed to reflect the conditions of the simulated area. This will be further discussed in the **Software Installation and Usage** section below.

Essentially, we implemented the model described by Thomas et al. (2013) for each neighborhood (census block groups) in Summit County. We developed the simulator by using years as the temporal unit of analysis. The modeling process as described in Thomas et al. (2013) was repeated for each neighborhood. With this approach, the developed simulator allows users to

* Observe the spatial distribution of obesity prevalence at any given year.
* Observe the changes in each neighborhood’s obesity prevalence over time.
* Observe the spatio-temporal patterns by neighborhoods by changing one or more parameter values.
* Each round of simulation will generate an output file.

For example, the table below shows the simulated obesity prevalence by neighborhoods from 2013 to 2019. As shown in this table, obesity prevalence does seem to plateau into future years.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 2013 | 2014 | 2015 | 2016 |
|  |  |  |  |
| 2017 | 2018 | 2019 |  |

Note: Increase rates are calculated with reference to baseline figures in 2013.

# Software Installation and Usage

The *ObesityPrevalence Simulator* was developed in this project as the deliverable. Below are two sections that discuss the software installation with files on the accompanying CD-ROM.

1. Python Installation – discusses the installation of Python programming language, which is needed to run the *ObesityPrevalence Simulator*.
2. *ObesityPrevalence Simulator* – discusses the installation and usage of the developed software.

## Python Installation

The *ObesityPrevalence Simulator* was developed in Python programming language. Python is a versatile language that is free to acquire and install. Python is also a cross-platform programming language, which means a python script can be used by computers with Windows operating systems or Mac computers, for both 32bit and 64bit versions. In addition, many libraries that process GIS and other forms of data have been developed and freely available in public domain. This allows further improvements and updates to be carried out easily.

To download python, visit <http://www.python.org/download/>. Alternatively, Python installers can be found in the accompanying CD-ROM.

Please note that the *ObesityPrevalence Simulator* was developed using Python 2.7.

A total of four Python installers are included in the accompanying CD-ROM:

* Windows 32bit Python 2.7
* Windows 64bit Python 2.7
* Mac 32bit Python 2.7
* Mac 64bit Python 2.7

## ObesityPrevalence Simulator

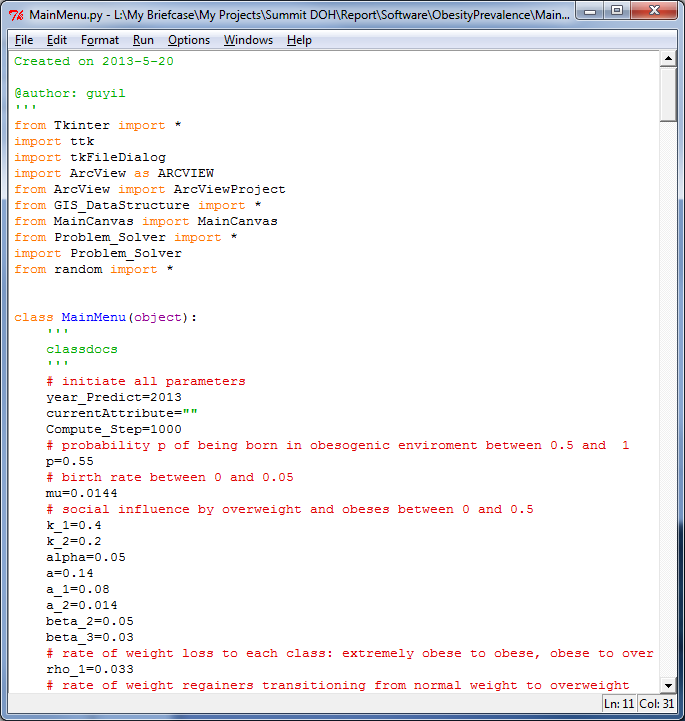
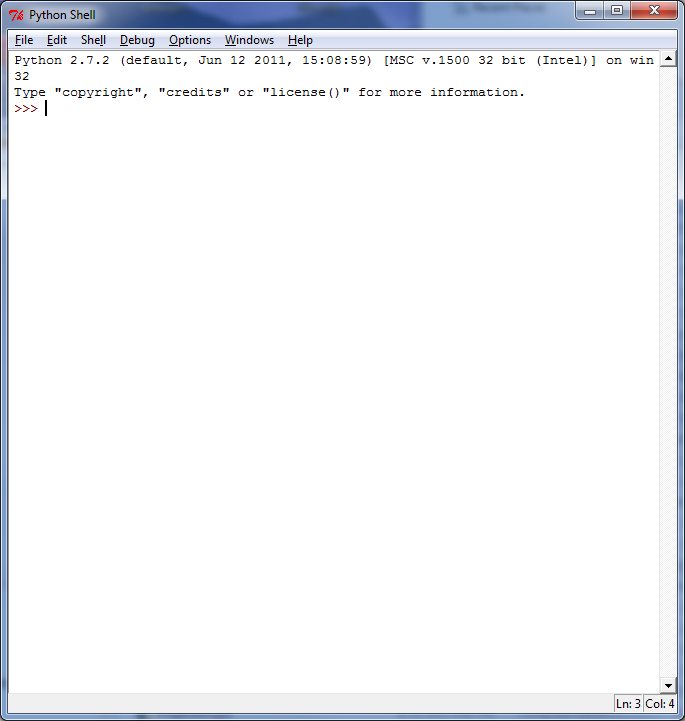
The implementation of our approach to simulating obesity prevalence has resulted in the *ObesityPrevalence Simulator*. This software was developed using Python programming language and can be used by itself without any GIS or statistical software.

There are three ways to use the *ObesityPrevalence Simulator*.

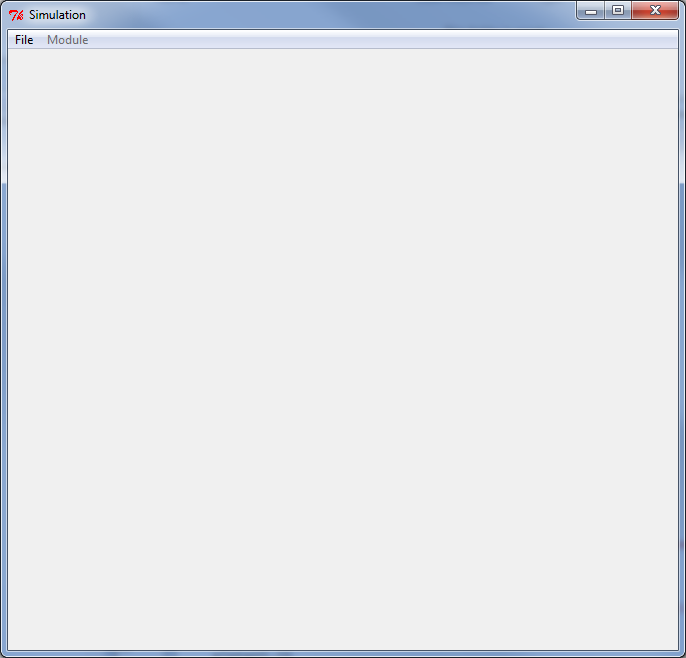
1. Users can download Python programming language and install it on the computer before running *ObesityPrevalence Simulator*. Installing this will enable the Windows to recognize the .py file extension as a Python script.
2. If using a Windows-based PC that is installed with ArcGIS 10.X, Python 2.7 would already be installed with ArcGIS. *ObesityPrevalence Simulator* can be used without additional installation of Python.
3. Users can just copy the entire content of the obesityprevalence folder to the computer’s hard drive. Once this is done, *ObesityPrevalence Simulator* can be initiated from within the copied folder by double-clicking at MainMenu.py file.

### Using ObesityPrevalence Simulator

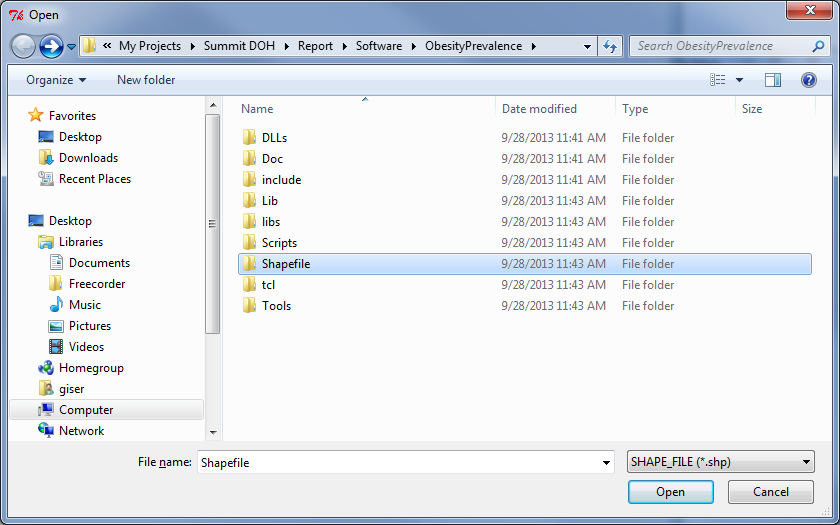
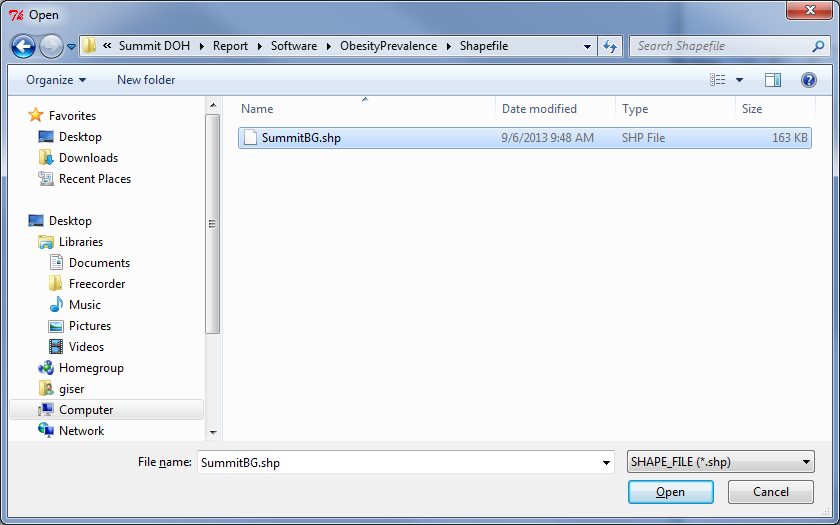
To initiate *ObesityPrevalence Simulator*, navigate to the **obesityprevalence** folder and double-click at **MainMenu.py**. Two windows will appear as below:



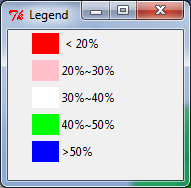
In the **MainMeny.py** window, use **Run** > **Run Module**, or press the **F5** key. The *ObesityPrevalence Simulator* will start with a Simulation window as below:



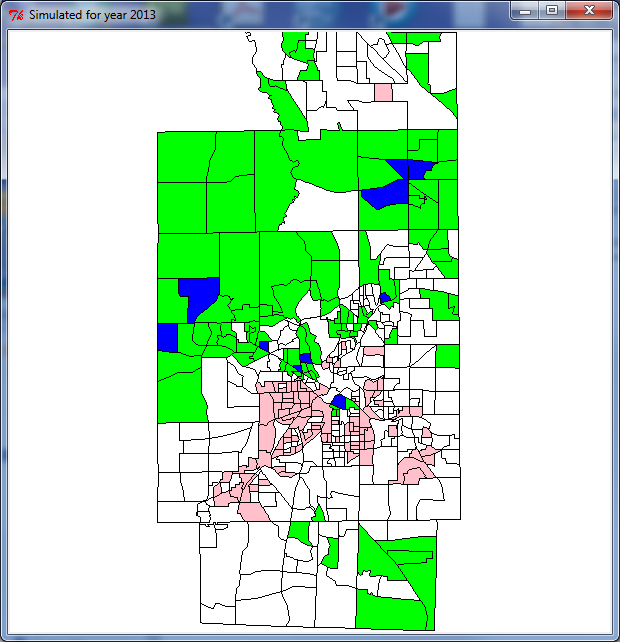
From the Simulation window, choose **File** > **Open** to open the **Shapefile** folder, and then select **SummitBG.shp**

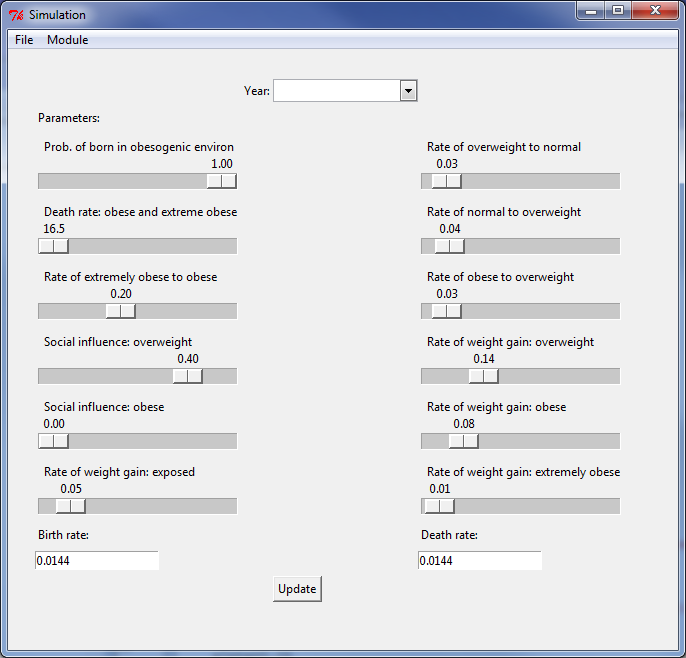
Upon opening the **SummitBG.shp** data layer, three windows appear as below:

A small **Legend** window shows the 2013 obesity ratios in 5 colors: 

A **Simulated for year 2013** windows shows the current obesity ratios by neighborhoods:



A simulation control panel, entitled **Simulation**, shows the various simulated year, parameters, and the **Update** button as below:

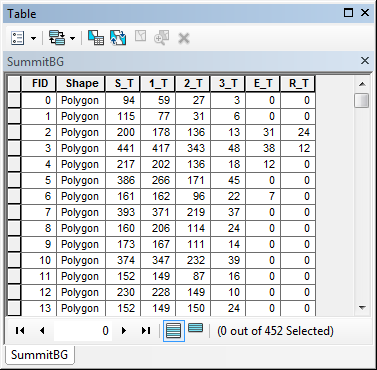


Please note that the parameters are set to their initial values (default values), which can be changed in simulation runs.

## Sample Data

The ObesityPrevalence simulator comes with a sample data file in shapefile format (ESRI, Inc., Redlands, California). Users of the *ObesityPrevalence Simulator* can use it to work with any customized shapefile data. The only requirement for the shapefiles is to have the following columns in the attribute table:

* ***S\_T***: the number of people in each neighborhood who are in normal weight range (BMI <= 25)
* ***1\_T***: the number of people in each neighborhood who are considered overweight (20 < BMI <= 30)
* ***2\_T***: the number of people in each neighborhood who are considered obese (30 < BMI <= 40)
* ***3\_T***: the number of people in each neighborhood who are considered extremely obese (BMI > 40)
* ***E\_T***: the number of people in each neighborhood who are exposed to possibility of changing from normal weight to overweight
* ***R\_T***: the number of people in each neighborhood who may have weight loss so to return from overweight to normal weight.



The simulator expects to read data from columns named as listed above. Therefore, please be sure to rename your attribute columns if needed.

In the obesityprevalence folder, a **shapefile** subfolder holds a set of shapefiles, entitled **SummitBG**. This can be used to test run the *ObesityPrevalence Simulator*. Please note that the boundary data for block group polygons were downloaded from <http://www.esri.com>. Figures for the *S\_T*, *1\_T*, *2\_T*, and *3\_T* subpopulations were calculated using height/weight data derived from drivers’ license data from the Ohio Bureau of Motor Vehicles. *E\_T* and *R\_T* data were derived from geographically weighted regression of the following relationships:

*E\_T* = *function*(*S\_T*, density non-fresh food outlets)

*R\_T* = *function*(*1\_T*, Distance to nearest fitness centers)

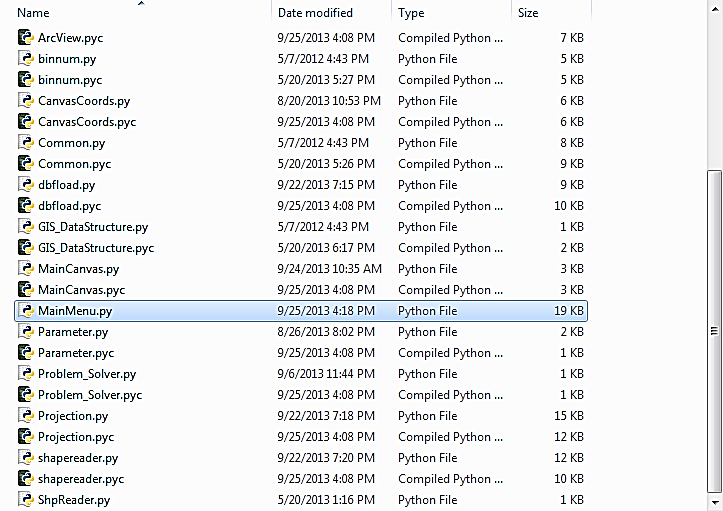
It should be noted that estimations for *E\_T* and *R\_T* with the above regression are provided here purely for the purpose of demonstrating the usage of *ObesityPrevalence Simulator*. Additional studies and analysis may be needed in order to derive better or more precise estimates.

The estimates for *E\_T* and *R\_T* should be done so each neighborhood has its own estimates. The examples included in the sample shapefile were derived using the relationships

* between *S\_T* and the density of non-fresh food outlets in each neighborhood for estimating *E\_T* and
* between *1\_T* and the distance to the nearest fitness centers from the neighborhood center for estimating *R\_T*.

## Software Usage

From within the **obesityprevalence** folder, double-click **MainMenu.py** to initiate the simulator.



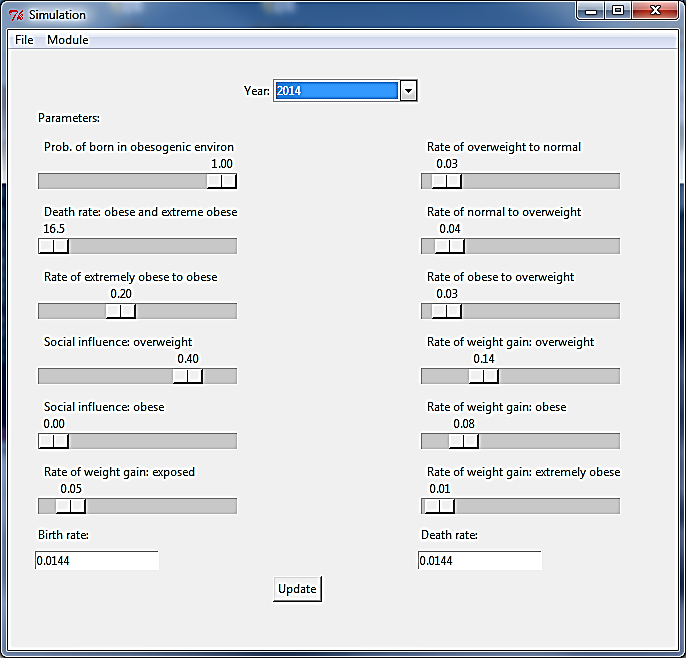
Once the simulator is started, two windows will be opened. The simulator is in the window with the title of Simulation.

To load the data layer, please click at the **File** menu and then the **Open** menu item. An **Open** dialogue box will be open. Please navigate to the folder where you have your shapefile stored. In our example, the **Shapefile** folder is inside the **obesityprevalence** folder. The sample data layer is **SummitBG.shp**.

* Click at the **SummitBG.shp** shapefile to open it.

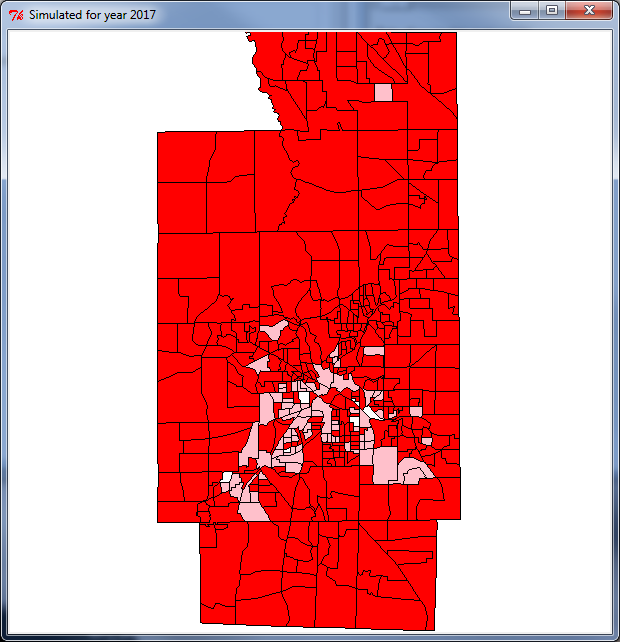
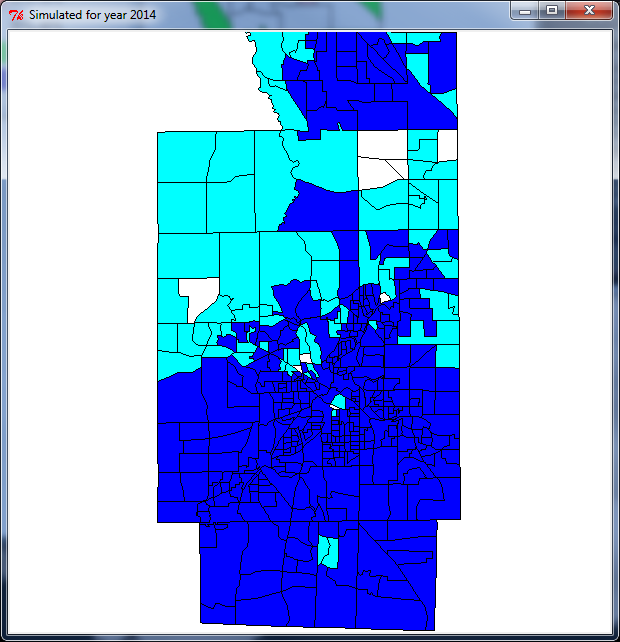
Three windows will be opened:

* A **Legend** window is opened to display the numerical ranges of the ratios of normal weight population in each neighborhood in the base year.
* A second window, entitled **Simulated for year 2013**, displays a map showing the spatial pattern of normal weight population.
* The third window is the control panel for setting the values of the simulation parameters.
  + When initiated, each model parameter is set to be in its initial value as outlined in the previous section. Users can click at the button below each parameter description and drag the button to the left or the right to change the parameter’s value.



In the control panel,

* At the top of the control panel, a target year can be selected.
* At the bottom of the control panel, different birth rate and death rate can be entered.
* Once all parameter values are set, click at the **Update** button to calculate simulated obesity prevalence.
* Each update will generate a new map window to show the simulated outcomes.
* Each update will also generate an output text file that records simulated values for *S\_T*, *1\_T*, *2\_T*, *3\_T*, *E\_T*, and *R\_T*. These textfiles are places in the **obesityprevalence** folder.



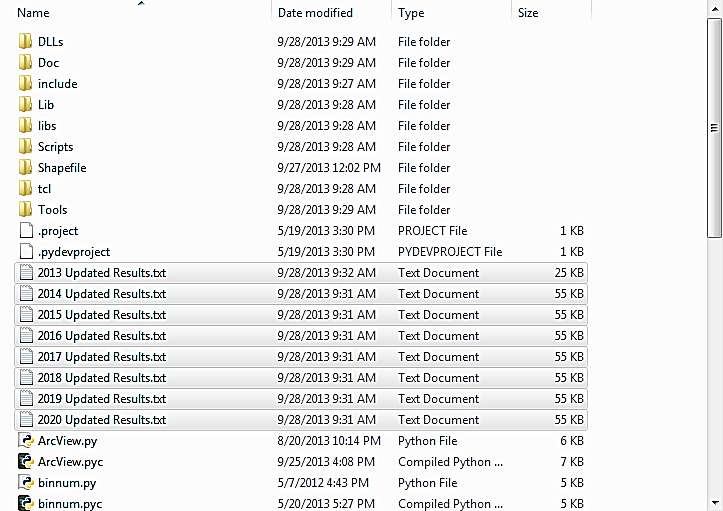
Upon completion of simulation, close all windows to exit from the simulator.

## Analysis of Output Data

Each time the Update button is clicked; new estimates for subpopulations in all neighborhoods would be calculated for the year selected. In addition, an output file will also be created and saved to the **obesityprevalence** folder. For example, the figure below shows the results from calculating estimates of subpopulations for all neighborhoods from 2013 to 2020.

Please note that to perform sensitivity analysis of a particular model parameter, follow these steps:

1. Decide the value range of *p*, which is to be tested for sensitivity analysis
2. Set all parameter values, including *p* = initial value
3. Run Simulation, and rename output file to the value of *p*.
4. Change *p* to the next value (add or subtract an increment value)
5. Check to see if *p*’s current value equals to *p*’s ending value. If so, stop.
6. Go to step 3.

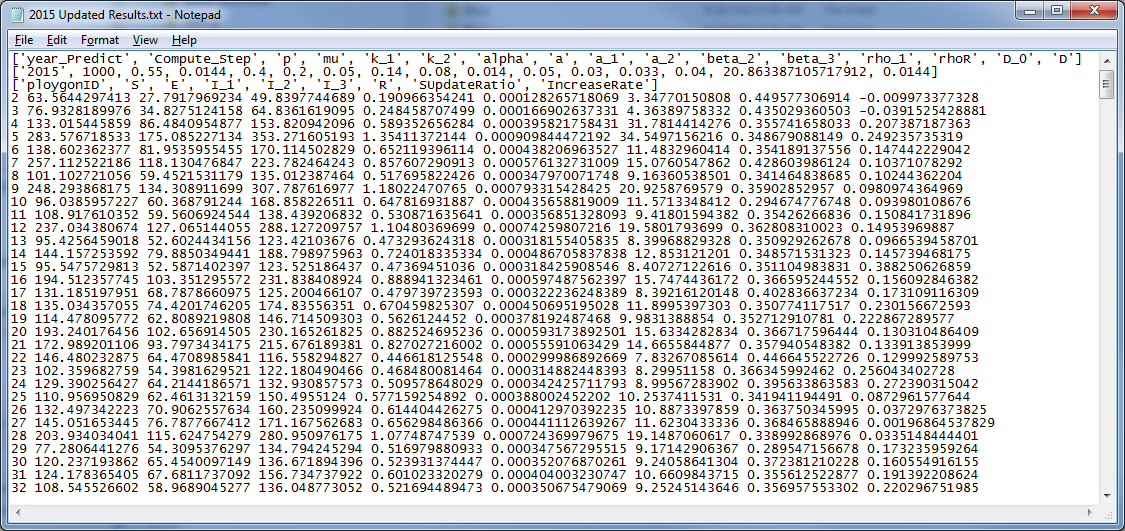


The output files are all in plain text format. In each output text file, the first two records (lines) report the parameter values used for that specific simulation run. The third record contains the names of the data items included in the subsequent records:

polygonID: Feature ID of the polygons in the shapefile. This column can be used to join estimated data to the shapefile for mapping purposes.

* **S**: *S\_T*, or the estimated subpopulation of normal weight for the simulated year
* **E**: *E\_T*, or the estimated exposed subpopulation for the simulated year
* **I\_1**: *1\_T*, or the estimated overweight subpopulation for the simulated year
* **I\_2**: *2\_T*, or the estimated obese subpopulation for the simulated year
* **I\_3**: *3\_T*, or the estimated extremely obese subpopulation for the simulated year
* **R**: *R\_T*, or the estimated recovered subpopulation for the simulated year
* **SUpdateRate**: the ratio of *S\_T* over total population
* **IncreaseRate**: the ratio of changed *S\_T* over *S\_T* of the base year (2013)

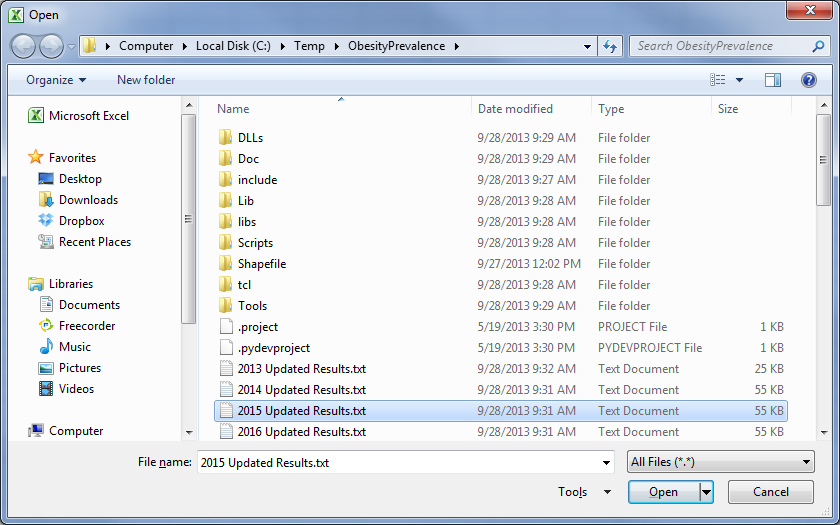
For example, the top portion of the output data file for 2015 Updated Results.txt is as below:



For convenience, we suggest that these data files may be opened by using spreadsheet programs such as Microsoft Excel™ in the following sequence:

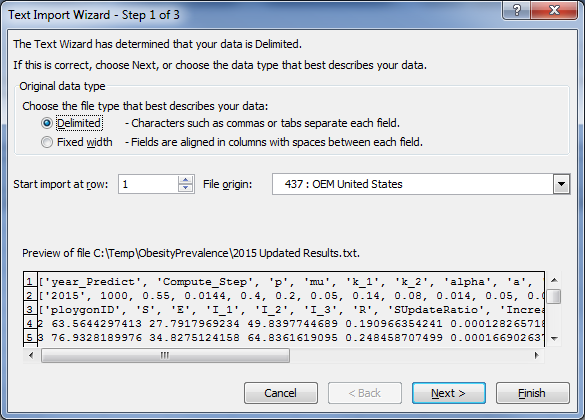
Start Microsoft Excel™ and use the **File**/**Open** menu item to open the **2014 Updated Results.tx**t from within the **obesityprevalence** folder.

Please note that in the **Open** dialog box, the file type filter may need to be set to **All Files (\*.\*)** so to display .txt files.



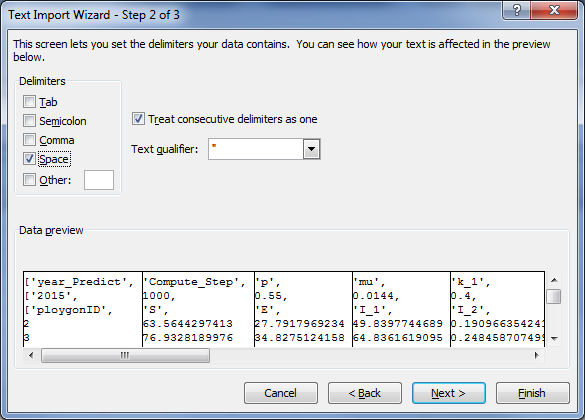
Click at the text file and then click at the **Open** button to continue.

In the **Text Import Wizard – Step 1 of 3** dialog box, make sure the **Delimited** option is checked and the **Fixed with** option is unchecked.



Click at the **Next >** button to continue.

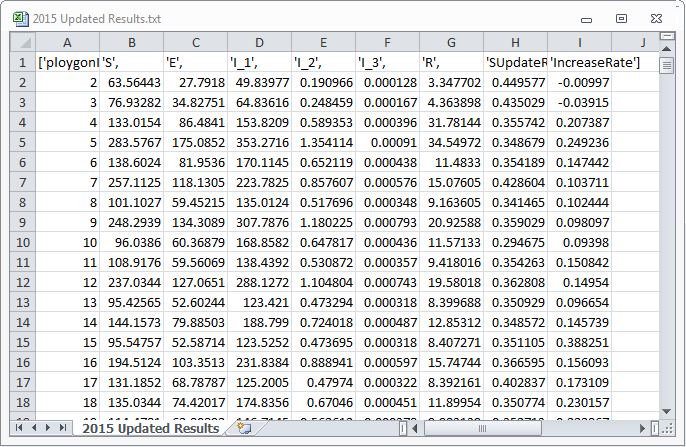
In the **Text Import Wizard – Step 2 or 3** dialog box, check off the **Tab** delimiters and check on the **Space** delimiters. At this point, you can see the data items



Click at the **Next >** button to continue.

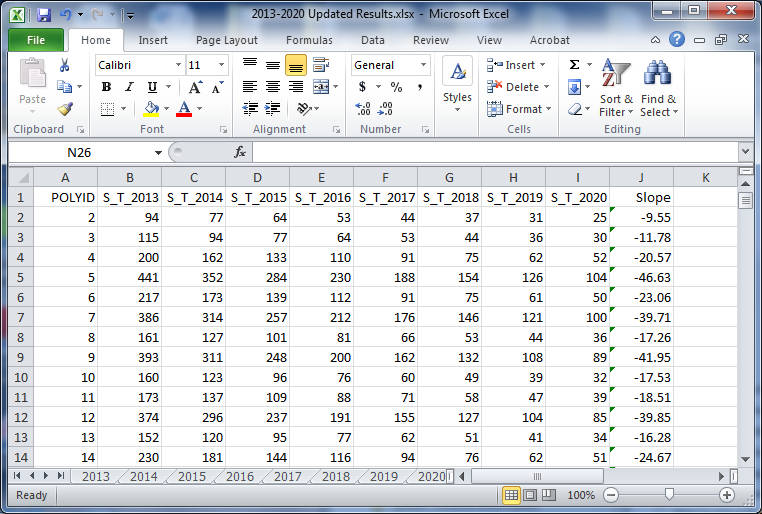
In the **Text Import Wizard – Step 3 of 3** dialog box, click at the **Finish** button to complete the import process.

In the opened spreadsheet, delete the first two rows. Rename the columns (in the first row) if desired, so the spreadsheet would look like:



## Data Analysis Example

In the obesityprevalence folder, a spreadsheet file, **2013-2020 Updated Results.xlsx**, contains output results generated with default parameter values for years from 2013 to 2020, as below:



In this spreadsheet, updated results from running simulations for years 2013 to 2020 were integrated into this file as worksheets. The *S\_T* subpopulation figures from each simulated year were then copied and pasted into the last worksheet.

Subsequently, a new column (J), entitled **Slope**, was created using the math function SLOPE in the spreadsheet to calculate the rate of decreasing normal weight population by the default parameter values. This is just one example for what can be done with the simulated results.

## Simulation Example

The advantage of using to estimate obesity prevalence is the ability to change values of model parameters by holding all others constant while varying only one or only a few parameter values. In the figure below, obesity prevalence is simulated for year **2018**, by setting social influence value to be **0.2**, **0.3**, **0.4**, and **0.5**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 0.2 | 0.3 | 0.4 | 0.5 |  |
| Simulated Obesity Prevalence, 2018  With varying levels of social influence on overweight subpopulation  While holding social influence on obese at 0.00 | | | | |

As can be seen in the progressive changes of obesity prevalence by increasing social influences on overweight and holding that influence on obese constant, above figure shows that higher levels of social influence seem to be important in shaping simulated obesity prevalence.

As a comparison, the figure below shows the insensitivity of social influence on obese subpopulation while that influence on overweight is held constant at **0.20**.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 0.2 | 0.3 | 0.4 | 0.5 |  |
| Simulated Obesity Prevalence, 2018  With varying levels of social influence on obese subpopulation  While holding social influence on overweight at 0.20 | | | | |

# Concluding Remarks

Obesity is an exceedingly complex public health problem with hypothesized causes at multiple interacting levels that are embedded in the very structure of society. This complexity appears to be the reason that one-dimensional preventive or therapeutic interventions are not very successful. Given this, new approaches are needed to fully understand the complexities associated with obesity. The *ObesityPrevalence Simulator* developed in this project is a new, more comprehensive, decision support tool for policy makers. The implementation of policies that effectively combat obesity would improve the health and well-being of a high percentage of the population, including both adults and children, as well as greatly reducing associated economic costs to society such as obesity-related health care expenses and loss of productivity.

Based on the susceptible, infected, and recovered (SIR) framework, *ObesityPrevalence Simulator* is featured by categorizing the population into subpopulations of normal weight, overweight, obese, and extremely obese. Furthermore, *ObesityPrevalence Simulator* allows population to be moved between subpopulations. Such movements can be defined by any reasoning from the various physical environment, food environment, built environment, and socio-economic environments of the neighborhoods.

Beyond the features of categorizing a population to subpopulations and allowing people to move between subpopulations, *ObesityPrevalence Simulator* also allows users to set a suite of model parameters in estimating future obesity prevalence. These parameters do affect how estimations are calculated. However, the parameters as defined by the local conditions allow the simulations to be executed with spatial variations and with localized conditions.

Finally, *ObesityPrevalence Simulator* provides a means of studying obesity prevalence at a very fine geographic scale. By using census block groups as neighborhoods, *ObesityPrevalence Simulator* goes beyond the conventional approaches of studying obesity prevalence at the scale of census tracts. The additional details reveal by using smaller geographic units certainly allow us to better understand spatial patterns and processes of obesity prevalence.

Beyond the scope of this project, studies that compare how simulated obesity prevalence levels react to different values of the model’s parameters would be valuable to engage. By fixing all but one parameter to vary in simulations, estimated obesity prevalence patterns can be used to related to how that particular parameter changes. If desired, multiple parameters can be allowed to change simultaneously so observations can be made to see how they affect obesity prevalence as a whole.

# References Cited

Balcan, D., B. Goncalves, H. Hu, J. J. Ramasco, V. Colizza, A. Vespignani. 2010. Modeling the spatial spread of infectious disease: the Global Epidemic and Mobility computational model. *J. Comput. Sci*. 1(3): 132-45.

Finkelstein, E. A., O. A. Khavjou, H. Thompson, J. G. Trogdon, L. Pan, B. Sherry et al. 2012. Obesity and severe obesity forecasts through 2030. *Am J. Prev Med*. 42(6): 563-70.

Flegal, K. M., M. D. Carroll, C. L. Ogden, and L. R. Curtin. 2010. Prevalence and trends in obesity among US adults, 1999-2008. *JAMA* 303(3): 235-41.

Jacobson, P., J. S. Torgerson, L. Sjostrom, and C. Bouchard. 2007. Spouse resemblance in body mass index: Effects on adult obesity prevalence in the offspring generation. *Am. J. Epidemiol*. 165(1): 101-8.

Thomas, D. M., M. Weedermann, B. F. Fuemmeler, C. K. Martin, N. V. Dhurandhar, C. Bredlau, S. B. Heymsfield, E. Ravussin, and C. Bouchard. 2013. Dynamic model predicting overweight, obesity, and extreme obesity prevalence trends. *Obesity*. Doi: 10.1002/oby.20520.

Wang, Y., M. A. Beydoun, L. Liang, B. Caballero, and S. K. Kumanyika. 2008. Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. *Obesity* 16(10: 2323-30.