

Population Mapping and Modeling for Manhattan

—by *Monika Calef*

Introduction

Problem

Worldwide, the coastal zone, which is vulnerable to sea level rise, tsunamis, and storms, is currently home to 1.2 billion people, a number that might grow to as much as 5.2 billion people by the 2080s (IPCC 2007). One globally significant city in the coastal zone is New York City, located at the mouth of the Hudson River and currently inhabited by nearly 8.5 million people (US Census 2008). As the attacks of September 11, 2001, showed, any major upheaval in this city causes global shockwaves.

Every city works to enable and secure the livelihoods of its people. Part of doing that requires that people apply what they know from the past and present to project a reasonable image—a model—of the future. This lab addresses one element of that model, population, to answer the question, What is the estimated population of Manhattan in 2100 at the census tract level?

Location

Manhattan, New York, USA

Time to complete the lab

30–45 minutes

Prerequisites

Intermediate geographic information system (GIS) experience

Keywords: environmental studies; population prediction; population modeling; US Census data; TIGER data; vector data; TIGER conversion; select by attribute; project; join table; add and calculate field

Data used in this lab

Census tract data for New York in projection Latitude/Longitude, 1927 datum at http://www.census.gov/geo/www/cob/bdy_files.html Census2000.xlsx (see <http://www.census.gov/> or <http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml>)

Student activity

This exercise consists of three tasks:

- Mapping current population in Manhattan by census tract
- Estimating future population (year 2100) by census tract
- Mapping population for 2100 by census tract

To complete the above tasks, you first need to determine the present population of Manhattan and its growth rate.

The US Census Bureau collects detailed information on the entire US population every 10 years. Data is organized by several hierarchical levels ranging from nations to blocks. You will use census tracts as a trade-off between detail and data manageability (figure 1). Since you have no way of knowing exactly how large the population in New York City will be in 2100, you will have to develop a model based on past census counts. As long as you don't interpolate too far into the future, the estimate should be fairly accurate.

To determine the population in Manhattan in 2100, you have to make some assumptions since you cannot know for sure what the future holds: (1) Assume that the city population will continue to grow at the same rate as it has displayed since 1940, and (2) assume that the population in each census tract will grow at the same citywide rate. The first assumption makes statisticians unhappy, as you are using less than 70 years (1940–2008, figure 2) to project 90 years into the future (2008–2100). This means your predictive accuracy is not great. It could be improved by using more historic data, but population growth rates prior to World War II might not be truly representative of the New York City of the twenty-second

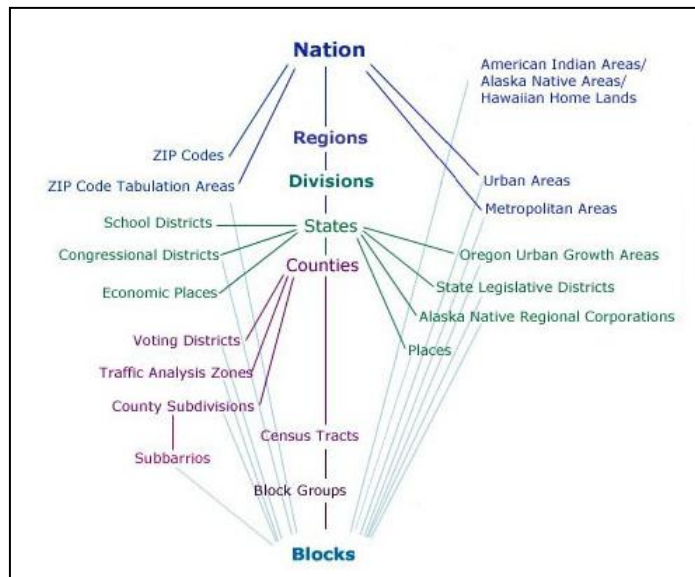


Figure 1: Geographic hierarchy of census data (source: <http://www.census.gov>).

century. The second assumption is very likely wrong, as it is probable that different tracts increase in population at variable rates and that some tracts simply cannot grow beyond a certain number of people. But short of creating population projection models for each individual census tract and determining maximum occupancy rates, this assumption will have to do. Averaged over the entire city, your prediction will probably arrive at a very similar value as a tract-by-tract estimate.

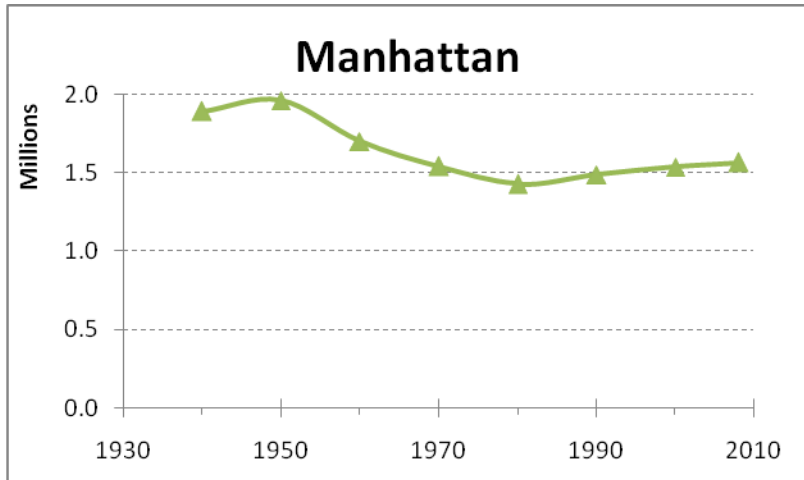


Figure 2: Population of Manhattan from 1940 to 2008 according to the US Census.

To predict future population, you need to create a mathematical model. Natural populations are assumed to grow at an exponential growth rate (J-curve); however, humans don't always abide by those laws (at least not at a city scale). Using a program like Excel, you can create four differently fit trendlines with their respective equations (figure 3).

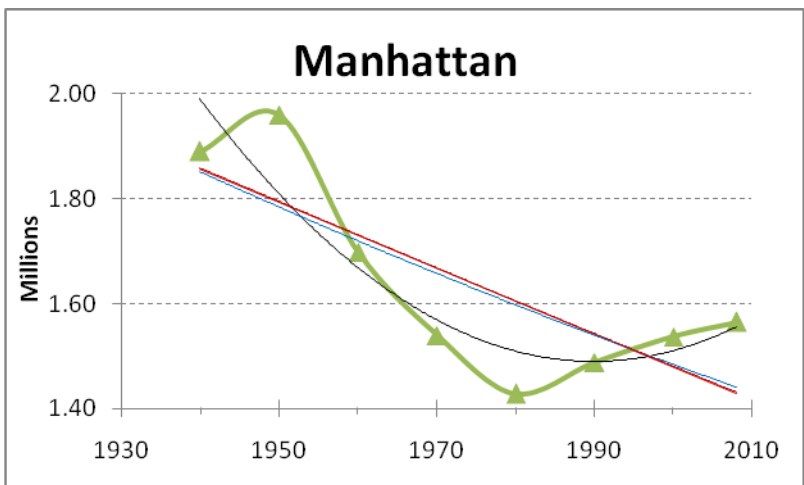


Figure 3: Population models for Manhattan using four mathematical approaches (note that the minimum value on the y-axis does not reach zero):

- $y = 6,282x + 14,043,805$, $R^2 = 0.61$ (linear; exactly underneath logistic line),
- $y = 2,349,868,542 e^{0.00x}$, $R^2 = 0.60$ (exponential),
- $y = -12,438,393 \ln(x) + 96,022,313$, $R^2 = 0.61$ (logistic), and
- $y = 201x^2 - 800,650x + 798,090,262$, $R^2 = 0.84$ (second order polynomial).

R^2 values are a simple tool for measuring how much variability is explained by the model or how closely the model fits the data points. Thus, the closer this value is to 1.0, the better the fit. Based on this simple comparison, the second order polynomial model ($y = 201x^2 - 800,650x + 798,090,262$) is much better than the other three, with an R^2 value of 0.84.

Using this equation, you can now estimate the future population for Manhattan by substituting years for x . For 2100, the equation reads:

$$Y = 201 * (2100)^2 - 800,650 * (2100) + 798,090,262 = 3,929,062$$

Again, using a simple program like Excel, you can create a figure showing future population growth by substituting 2010 to 2100 in decadal increments for x (figure 4).

Since the population in 2000 was 1.5 million, and predicted population for 2100 is 3.9 million, you can calculate a direct growth rate of $3.9 \text{ million} / 1.5 \text{ million} = 256$ percent. This value provides you with a quick way to estimate population for 2100 for each census tract.

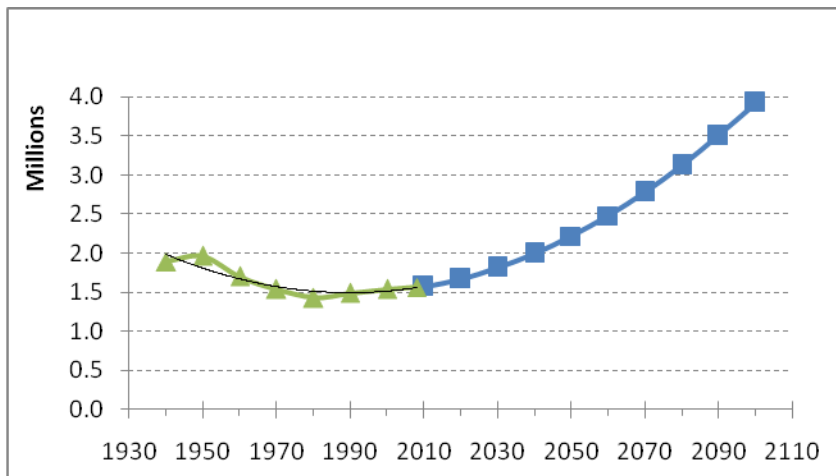


Figure 4: Historic population growth in Manhattan (green line and triangles) and predicted population growth until 2100 (blue line and squares) based on a second order polynomial model (black line).

Project overview

This project consists of three parts: (1) downloading appropriate datasets, (2) projecting and data processing to extract the information you are seeking, and (3) computing population estimates to answer the research question on population distribution in Manhattan in 2100.

COLLECT DATA

CARTOGRAPHIC BOUNDARY FILES

The US Census Bureau creates its own cartographic boundary files, also called Topologically Integrated Geographic Encoding and Referencing (TIGER) data. Census 2000 data is available for download for the entire United States in shapefile format.

- 1 Before you get started downloading data, use either ArcCatalog or Windows Explorer to create a folder called 01_PopModeling under C:\Student. Use this folder for the rest of this exercise to store your data.
- 2 Go to http://www.census.gov/geo/www/cob/bdy_files.html and choose *Census Tracts: 2000*.
- 3 On the next screen, scroll down to *Census 2000 in ArcView Shapefile (.shp) format* and select *New York tr36_d00_shp.zip*.

Open the ZIP file by clicking it and clicking *Open* or by using a program like WinZip or 7-Zip to first extract the three files and copy them to your student folder. The data has an undefined projection because it is unprojected, that is, it is in latitude-longitude form. New York City officially includes four counties in New York State (FIPS code 36): Bronx, Kings, New York, and Queens. You will focus this analysis on New York County (i.e., Manhattan, FIPS code 61), but you are welcome to repeat the analysis for any of the other counties.

POPULATION COUNTS

Population data can be downloaded from the US Census site at <http://www.census.gov/> or <http://factfinder.census.gov>, but it is tedious to find the right link and clean up the data so that it is formatted correctly.

This lab includes a user-friendly version of the population 2000 data by census tract in Excel format. The file is called *census2000.xlsx* and contains one worksheet called *Manh* with two columns: *Tract* and *Pop2000*. File cleanup consisted of removing tract from the name column, setting the population numbers so they would be recognized as numbers, removing text wrapping and commas, and so on. Also, if you are importing worksheets in .xls format, any column with numbers will automatically be converted to numeric format. This causes a problem when joining files by tract because in the geographic boundary files, tracts are treated as text. Check out this link in the webhelp: “Understanding how to use Microsoft Excel files in ArcGIS” (<http://resources.arcgis.com/en/help/main/10.1/index.html#//005s0000001w000000>).

PROCESS DATA

CARTOGRAPHIC BOUNDARY FILES

- 1 Start a new map in ArcMap™ and create a folder connection to C:\Student\01_PopModeling. Then add the *tr36_d00* shapefile.
- 2 Click *OK* and proceed past the “Unknown Spatial Reference” warning. Your screen should show all census tracts in New York State.
- 3 To decrease the amount of data you have to process, extract Manhattan from the statewide dataset. Open your toolbox and select *Analysis Tools » Extract » Select*. Use *tr36_d00* as your input and click the *SQL* icon. On the next menu, either select or type “*COUNTY*” = ‘061’ as the expression (make sure you use spaces and appropriate quotation marks). Save the output as *manh_LL*.

Undefined projection is terrible to work with, since it does not provide useful ground units (feet, meter, miles, kilometers) for the data. Therefore, the next step is to project *manh_LL* to universal transverse Mercator (UTM).

- 4 Go to *Toolbox » Data Management Tools » Projections & Transformations » Feature » Project*.
- 5 Select *manh_LL* as your input, click the button for *Input Coordinate System*, and then click *Select » Geographic Coordinate Systems » North America » NAD 1983.prj*
- 6 Call the output file *manh_pop*, select a predefined coordinate system for your output, and then click *Projected Coordinate Systems » UTM » NAD 1983 » NAD 1983 UTM Zone 18N.prj* and click *OK* twice.
- 7 Right-click *manh_pop* in the table of contents and select *Zoom to Layer* to see only Manhattan. Remove *manh_LL* from your table of contents.

If you want to change the data frame's projection to UTM, go to the main navigation toolbar and click *View » Data Frame Properties » Coordinate System » Select a Coordinate System » Layers » manh_pop » NAD_1983_UTM_Zone18N*. To change the dataframe display units from decimal degrees to meters, right-click *Layers* in your table of contents, go to *General » Units » Display*, scroll to *Meters*, and click *OK*.

- 8 Change the symbology to outline only.

Double-click the symbol for *manh_pop* in the table of contents and choose *Hollow* from the symbology menu that pops up.

Your map should look like figure 5.

POPULATION COUNTS

Add the *Manh\$* worksheet from the *02_data* folder to your view and look at the data in the table. Compare it with the attribute table for *manh_pop*. Then join the two files (*manh_pop*) using *tract*.

- 9 Right-click *manh_pop* and click *Join and Relates » Join*.
- 10 Select *join attributes from a table*, base the join on the field *NAME*, and then use *Manh\$* as the table to join and *Tract* as the field to base the join on. Click *OK*.

If you now open the attribute table for *manh_pop*, you will see *Tract* and *Pop2000* at the very right of your table.

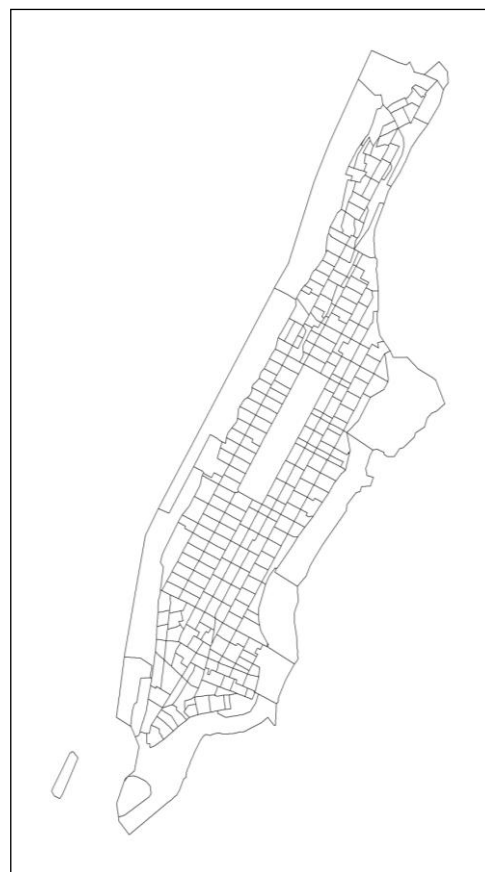


Figure 5: Manhattan census tracts.

If you are having problems with the join, for example, either *Pop2000* or *Tract* does not show up as an option in the wizard, you probably have incorrect field properties. Right-click *Manh\$* and go to *Properties » Fields*. *Tract* must be of type *text*, while *Pop2000* must be of type *double* (this refers to precision) with *numeric* number format.

ANALYZE AND VISUALIZE

You can now map population distribution in Manhattan by census tract.

- 1 Right-click *manh_pop*, choose *Properties*, and click the *Symbology* tab.
- 2 Click *Quantities » Graduated Colors*. Choose *Pop2000* for the value field and keep the default color ramp (Yellow to Red).
- 3 For seven classes, Natural Breaks classification method and the labels formatted (zero number of decimal spaces and show thousands separators), your map should look similar to figure 6.

Usually, you would not use a choropleth map to show raw population numbers. However, in this case, you want to see which tracts contain large numbers of people versus which tracts are almost empty of people. A more appropriate use of choropleth mapping here would be a conversion of the raw population numbers into population density. Since there is an area column in the shapefile's attribute table, you could simply normalize your value field (see previous step) *Pop2000* over *AREA*. Or you could calculate the actual area in square kilometers in your UTM units by right-clicking the *AREA* field heading in the attribute table and clicking *Calculate Geometry » Area*.

Now you are ready to update your census tract information in ArcMap.

- 4 First, make a clean backup copy of the population shapefile, just in case.
- 5 Also insert another data frame and copy/paste the *manh_pop* layer to it. Activate the new data frame and do the remaining steps. This way you will have both maps in the same map document for comparison.
- 6 Open your *manh_pop* attribute table in ArcMap, click *Table Options* in the top left corner, and then select *Add Field*. Call this new field *Pop2100*; the field type must be *Long Integer* to allow sufficient storage space for numbers in the millions.
- 7 Right-click the *manh_pop.Pop2100* field in your table and choose *Field Calculator*. Calculate the field by using the following expression and make sure there is a space on each side of the asterisk: $[Manh$.Pop2000] * 2.56$.

Now you can map your estimated population for 2100, which should look similar to figure 7 (using seven classes and the *Natural Breaks* classification method).

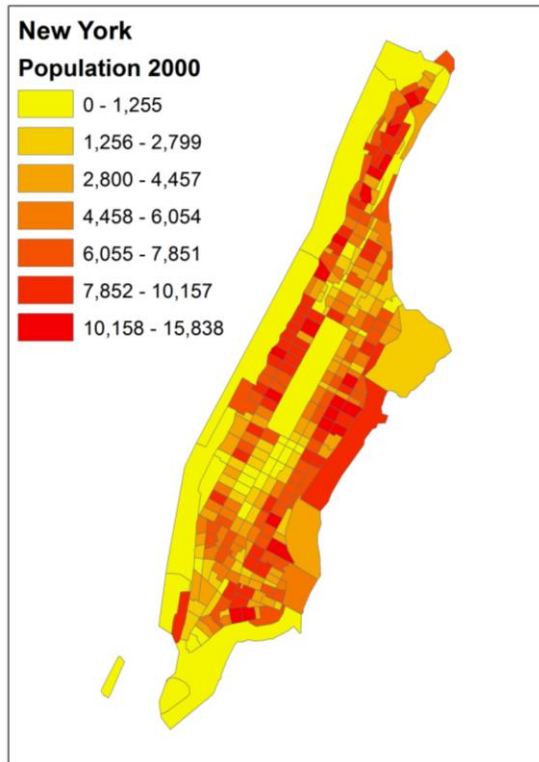


Figure 6: Population distribution in Manhattan by census tract in 2000.

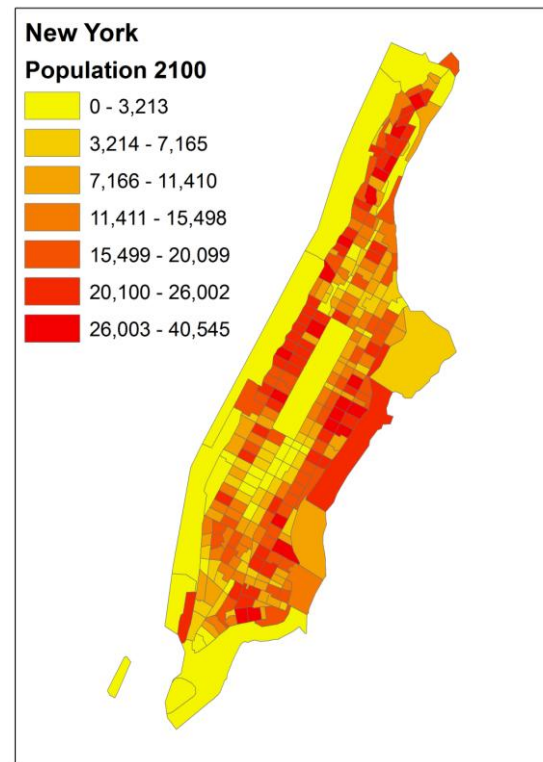


Figure 7: Projected population for Manhattan in 2100 by census tract.

Figure 7 now provides the answer to your research question on population distribution in Manhattan at the census tract level for the year 2100.

Since both maps are using raw populations and an identical color scheme, they should look identical. This is not very helpful for conveying information on population change but provides a quick error check that you have performed the conversions correctly. It would be better if figure 7 was mapped using an color scheme which differs sufficiently from figure 6 to show that populations are much higher in 2100 than 2000.

A comparison of the two maps shows that there are still tracts with zero population in the future while, for example, the tract with the highest population has increased from 15,838 to 40,545 persons. This is not surprising, since you have used a 256 percent growth rate (again, this is a quick and easy way to check for any errors).

FURTHER ANALYSIS

Another interesting comparison between the two years would be a calculation of population density and a comparison of the resultant maps. Which tracts are the most densely populated? Are these the same tracts that also contain the highest population?

Credits

Images

Figure 1, courtesy of U.S. Census.

Sources of supplied data

Census2000.xlsx, created by the author using data derived from the U.S. Census.

Instructor resources

Updates for 2013

Since the 2010 census boundaries were not available yet, the lab was checked to make sure 2000 census data was still available at the same internet addresses. A few minor edits were done to the wording to improve clarity the legend text for figures 6 and 7 were cleaned up.

Additional information

This exercise is one of three complementary exercises. *Population Mapping and Modeling for Manhattan* estimates the current population of Manhattan and the population for the year 2100. *Impacts of Sea Level Rise and Storms on Manhattan* estimates the land area threatened to flood from climate change-induced sea level rise as well as a hurricane-induced storm surge. *People of Manhattan at Risk* estimates the number of people at risk from hurricane-induced floods and sea level rise now and in 2100.

To assess resource needs for the future or threats by natural disaster, it is critical to compute population projections. The model in this exercise is very simple and has statistical flaws, but in its simplicity, it is useful for predicting population distribution in this one location for 2100. This becomes important in the third exercise of this series, *People of Manhattan at Risk*, in which the number of people at risk of flooding during a storm surge and during future sea level rise is calculated.

References

Intergovernmental Panel on Climate Change Working Group II. *Impacts, Adaptation and Vulnerability*. Edited by Parry, M. L., O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson. New York: Cambridge University Press, 2007. Available at http://www.ipcc.ch/publications_and_data/ar4/wg2/en/contents.html.