Tutorial 5: Summarizing Tabular Data

This tutorial will introduce you to the following:

- Identifying Attribute Data Sources for Environmental Health and Demographic Data
- Creating New Attributes/Fields (Review)
- Summarizing Attribute Data

One of the most important and also one of the most challenging uses of GIS is to collect, distribute, and assess demographic data. This tutorial introduces you to demographic data, in the context of an environmental justice study. Specifically, is there inequality in exposure to air toxins in New Jersey, and particularly, are black or Hispanic minority populations at higher risks for cancer, asthma, or neurological disorders caused by air pollution? This is a type of question that we can begin to assess, using GIS, information from the EPA, and census data.

Part 1. Identifying Attribute Data Sources

National-Scale Air Toxics Assessment (NATA):

NATA is a series of studies completed by the EPA (for 1996, 1999, 2002, and 2005) that assesses the contribution of air pollution to risks of cancer, neurological disorders, and respiratory disorders. The website for NATA is here: http://www.epa.gov/nata/. These data have already been downloaded for you, and posted on Google Drive, in GIS format.

U.S. Census:

The US Census houses enormous amounts of data on demographic measures, such as total population, race/ethnicity, education, and income. The main census page is at: http://www.census.gov. The data acquired for you come from the American Fact Finder Download Center for the US Census: http://factfinder2.census.gov/faces/nav/jsf/pages/download_center.xhtml#. The census also hosts shapefiles of census geographic units, such as states, counties, and census tracts, and census blocks, see for example: http://www.census.gov/geo/maps-data/data/cbf/cbf_tracts.html.

Download Files:

The file NATA2005.zip is a shapefile of census tracts in New Jersey, and the NATA scores for each tract for cancer risks. Download and unzip this file, and add it to a blank ArcMap. Another file, called aff_download.zip contains several spreadsheets from the census for New Jersey census tracts for 2000. Download and unzip this folder. Add the file 2000SF1_NJTracts.csv to your ArcMap. Please feel to explore the files in the aff_download folder. One of the files contains the metadata, or the codes and names of the census variables.

Part 2. Joining Tables and Creating New Fields (Review)

*Step 1. Create a Field for Joining Data*:

To join the NATA and census tables together, we need a common identifier. Fortunately, there is one: FIPSNO (for NATA2005) and GISID2 (for 2000SF1_NJTracts). Unfortunately, they are in two different formats: double and string. To join, they need to be the same format (as some of you discovered in pervious exercises). In the NATA2005 file, first open the attribute table, and click Add New Field (do you remember how to do this? If not, you might consult the instructions from the previous tutorials). Name the new field something TractID, because it's an ID number, and make sure the type for the new field is DOUBLE. Next, use the field calculator, to set the new field equal to FIPSNO. Click OK to the error message – there are some polygons with no FIPSNO, but you may proceed. Now you have two attributes that are equal to FIPSNO, but the new field is in the right format. Next, join NATA2005 to the 2000SF1_NJTracts.csv, using your new value, e.g. TractID, and the GISID2 field from the census data.
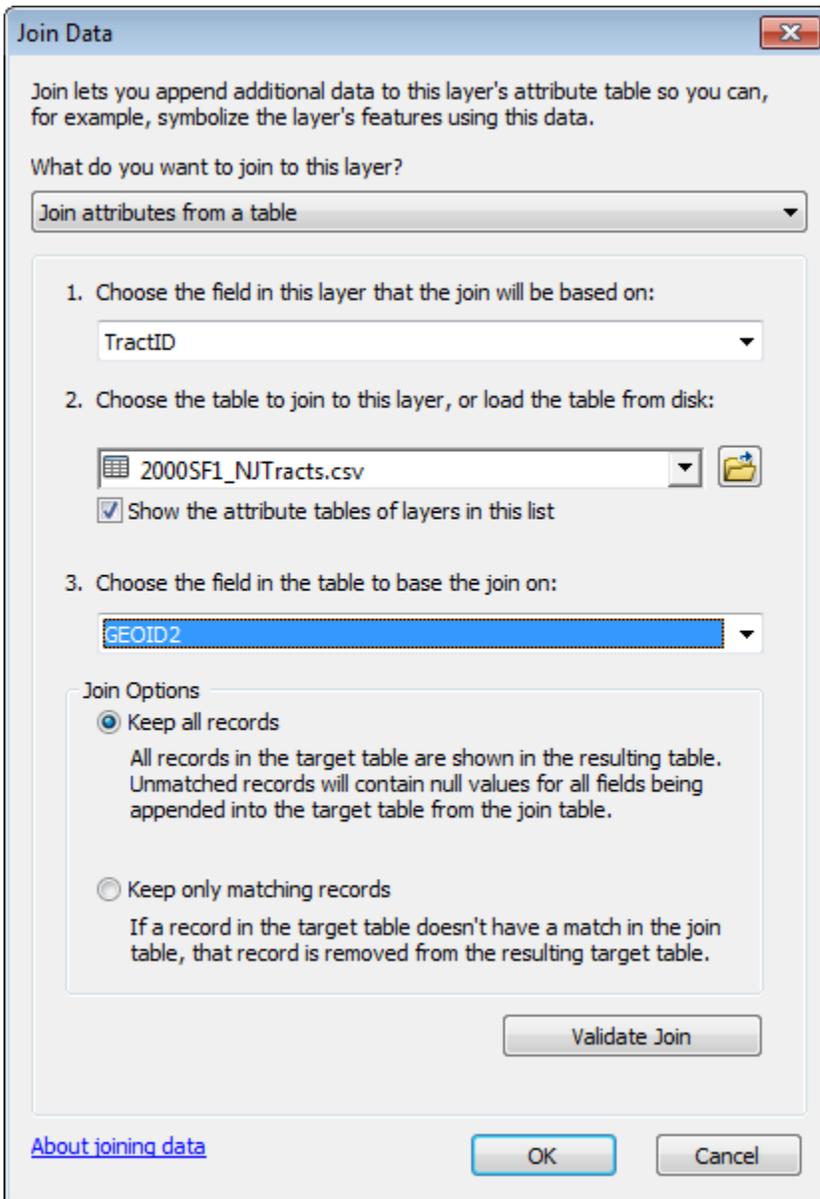
Step 2. Create a Field to Assess Data:

Suppose you want to compare demographic characteristics of areas most impacted by air pollution. There are many ways to do this. One way is to look at the areas in the top 10 percent of the state, for risks associated with air pollution.

Look at the layer properties > symbology, to show quantities for the field Total_Canc (this is the total carcinogenic risks from air pollution for each census tract). Then, click classify, and change the method to quantile (10), and you should see a chart similar to the one below.
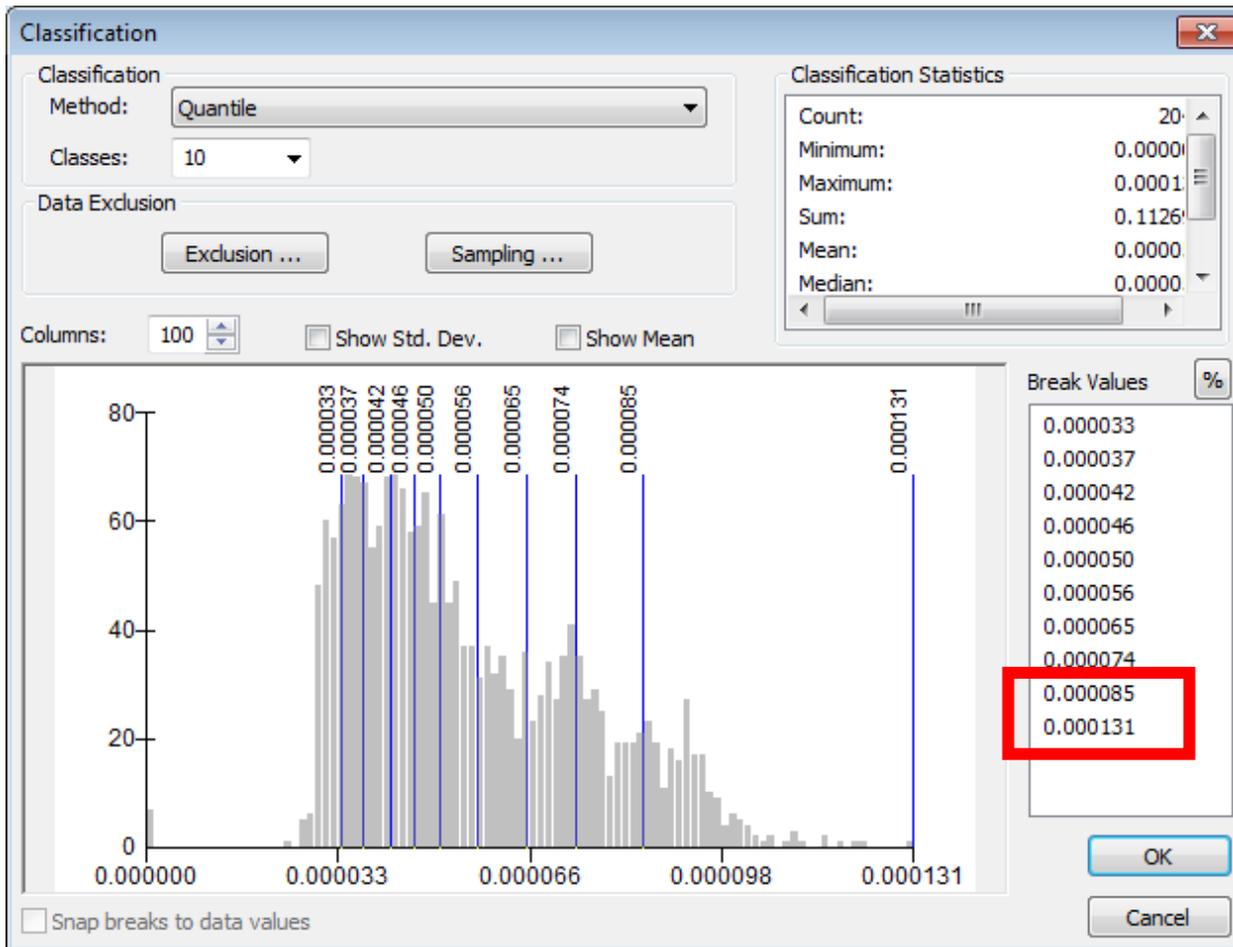
Ten quantiles also give you deciles, so you know that the values in the range 0.000085 to 0.000131 are the top ten percent for the state.

Finish creating a thematic map, then open the attribute table for NATA2005. Using the select by attribute button, choose the tracts with Total_Cancer risks that are greater than or equal to .000085.

Select by Attributes

Enter a WHERE clause to select records in the table window.

Method : Create a new selection

"NATA2005.Background"
"NATA2005.Secondary"
"NATA2005.Total_Canc"
"NATA2005.1_1_2_2_TE"
"NATA2005.1_1_2_TRIC"

=	< >	Like
>	> =	And
<	< =	Or
_ %	( )	Not

Is

8.4881299999999997e-005
8.4917299999999994e-005
8.4933999999999997e-005
8.4976100000000001e-005
8.4996900000000006e-005
8.5191499999999999e-005

Get Unique Values | Go To:

SELECT * FROM NATA2005_csv WHERE:

"NATA2005.Total_Canc" >= 8.5191499999999999e-005

Clear | Verify | Help | Load... | Save...

Apply | Close

Then, create a new variable called TopDecile.
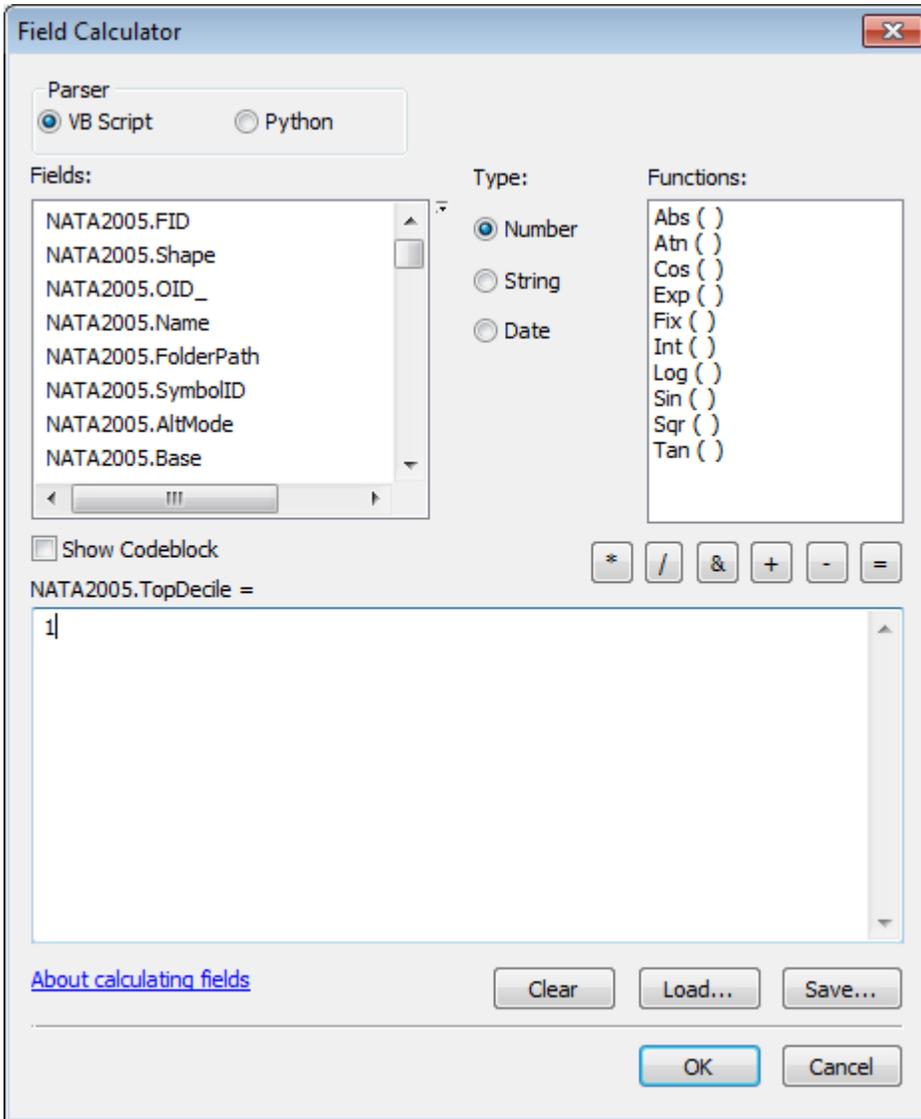
Add Field

Name: TopDecile

Type: Short Integer

Field Properties

| Precision | 0 |

OK | Cancel

Using the field calculator, set TopDecile equal to one. You will see that the field calculator, automatically, will set TopDecile equal to one only for values that are selected.

After you calculate TopDecile, unselect the census tracts with clear selection button.

Part 3. Displaying & Summarizing Attribute Data - Online

Navigate back to the TopDecile variable in the attribute table, and right-click it. This time select Summarize. Here we will summarize variables, based on if they are in the top decile for carcinogenic risk due to air pollution. Scroll down to choose demographic measure to study. To assess the relative impact on the total population (HC01_VC01), black population (HC01_VC30), and Hispanic population (HC01_VC56). Check to find the Sum.
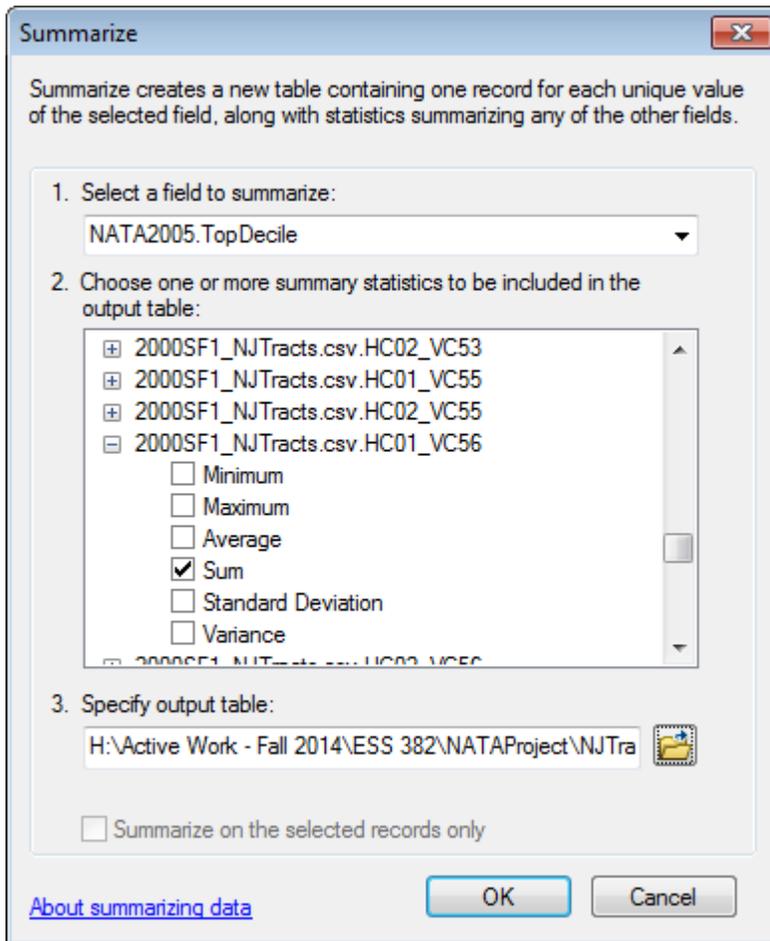
Save the output in an appropriate location, as a dBase file. Once the summary is complete, do not add the table to your map. Instead, navigate to the file on your computer and open it in excel.

Step 1. Create a Comparison Table

The summary output summarized the data by census tracts in the top decile (1) and not in the top decile (0) - example shown below.

	A	B	C	D	E	F
1	TopDecile	Cnt_TopDe	Sum_HC01	Sum_HC01	Sum_HC01_2	
2	0	1841	8042832	1037800	833421	
3	1	203	819758	131685	299646	
4						

Cnt_TopDec shows the count or number of census tracts in each category. There are 1,841 tracts that are not in the top decile, and there are 203 that are. Sum_HC01_C gives the total population in each category. So, 8 million people in New Jersey do not live in a census tract in the top decile, and about 131 thousand do live in a tract in the top decile. It is possible to add cells C2 + C3 to find the total New Jersey population in 2000: 8.9 million. Sum_HC01_1 gives the total black population in each census tract type, and Sum_HC01_2 gives the total Hispanic population in each census tract type. The findings then show that both Hispanic and black populations in New Jersey are disproportionately exposed to carcinogenic risks from air pollution.

In Microsoft Excel, dress up your table a bit (example below). You can copy and paste it into ArcGIS in Layout View.

Table 1. Demographic Characteristics in New Jersey Census Tracts with the Highest Carcinogenic Risks from Air Pollution

TopDecile	Number of Census Tracts	Total Population	%	Black Population	%	Hispanic Population	%
no	1,841	8,042,832	91%	1,037,800	89%	833,421	74%
yes	203	819,758	9%	131,685	11%	299,646	26%
	2,044	8,862,590		1,169,485		1,133,067	

Step 5. Add Additional Layers and Finishing Touches

Create two thematic maps: one to show carcinogenic risk from air pollution, and another to show a demographic variable. Also, add your summary table, legends for each map, and a title. An example is shown below. You've created a demographic analysis, useful as an environmental justice study. Hispanic populations in New Jersey experience much higher exposures to carcinogenic risks due to air pollution than other demographic groups.

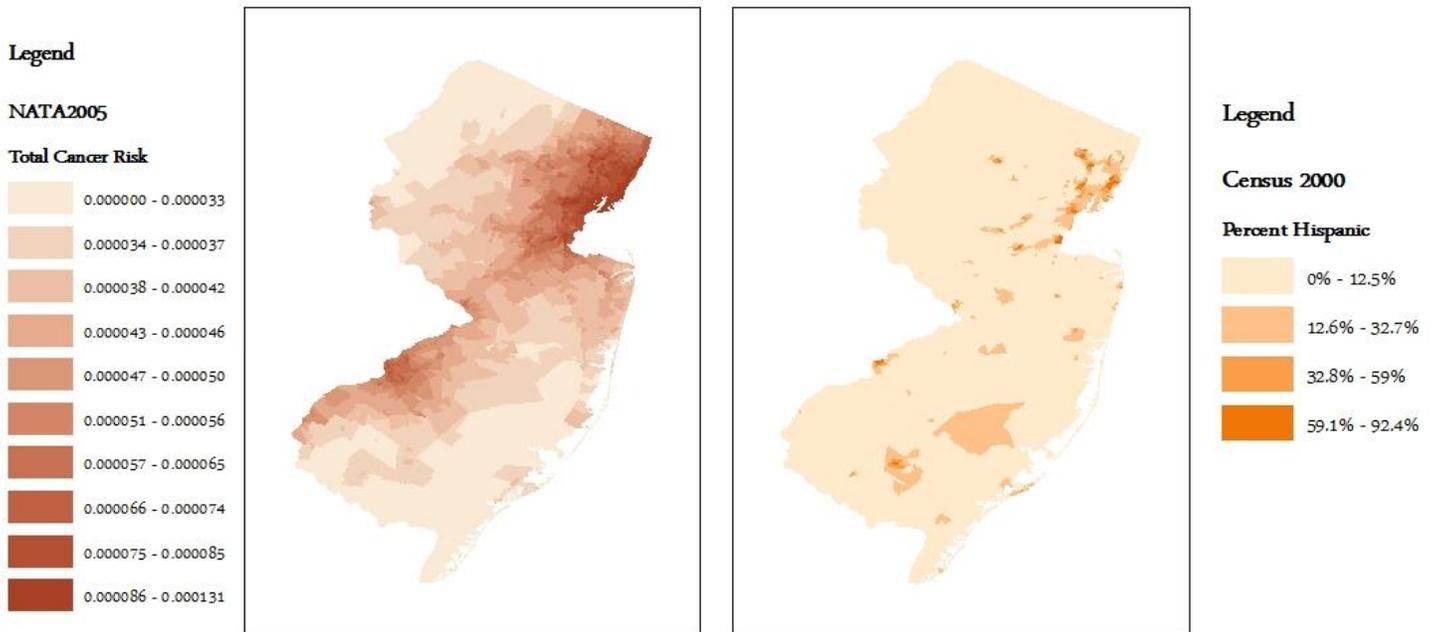Environmental Justice Study of Carcinogenic Risks from Air Pollution