EPIMAP EXERCISE

MAPPING & PIVOTING EXCEL DATA

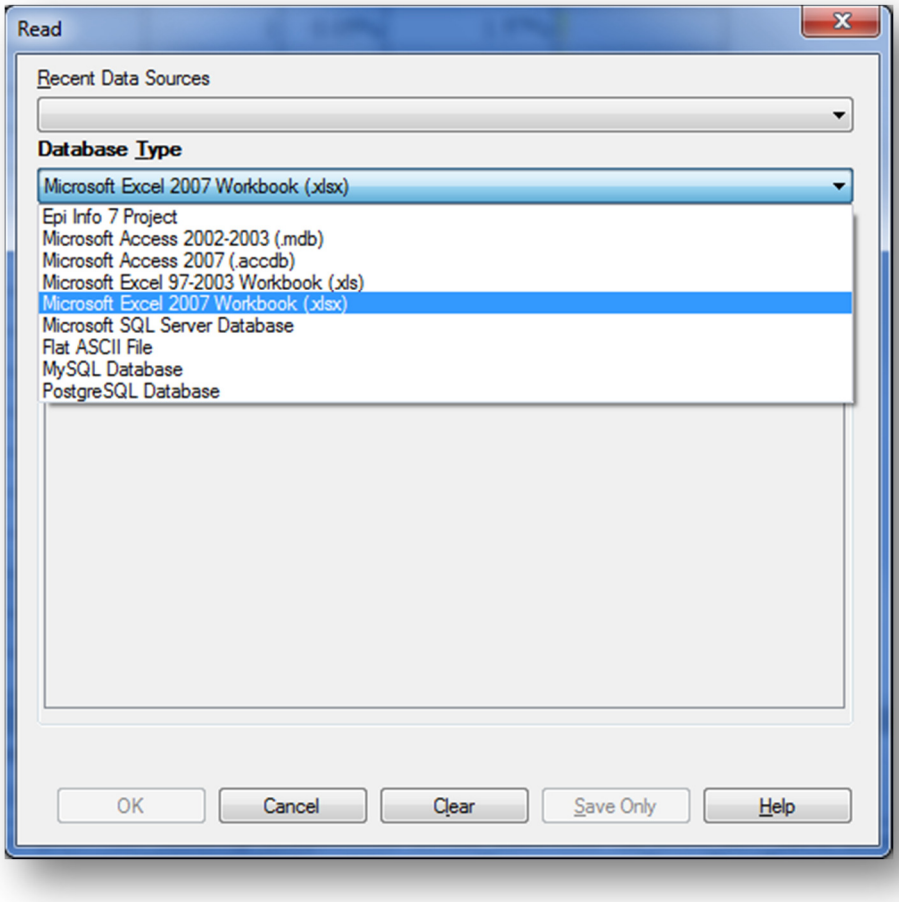
Document Version 1.0 **Tuesday, November 01, 2016**

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

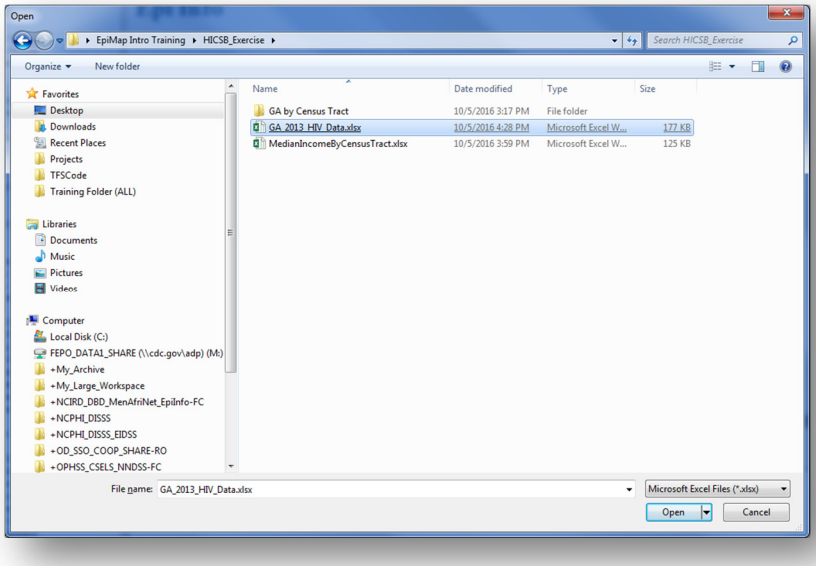
Exercise – Pivoting Excel Data

stateno	status_fla	hiv_aids_dx_dt	report_sta	rsd_state_	rsd_count	_rsd_county_fips	_rsd_county_na	rsd_census_tra	rsd_geogr	rsd_addre	site	_race	birth_sex	_mi_trans_	hiv_aids_age_yrs
1	W	20130110	GA	GA	USA	121	FULTON CO.	13121003100	1	1	GA	6	M	01	41
2	A	20130129	GA	GA	USA	089	DEKALB CO.	13089023418	1	1	GA	6	F	02	16
3	A	20130217	GA	GA	USA	067	COBB CO.	13067031108	1	1	GA	1	M	05	45
4	A	20130313	GA	GA	USA	089	DEKALB CO.	13121009200	1	1	GA	6	M	01	56
5	A	20130714	GA	GA	USA	089	DEKALB CO.	13089023428	1	1	GA	4	M	02	40
6	W	20130418	GA	GA	USA	067	COBB CO.	13067030227	1	1	GA	1	M	03	30
7	A	20130614	GA	GA	USA	033	BURKE CO.	13245010708	1	1	GA	1	F	01	54
8	A	20130410	GA	GA	USA	033	BURKE CO.	13245010708	1	1	GA	4	M	02	19
9	A	20130226	GA	GA	USA	073	COLUMBIA CO.	13073030201	1	1	GA	6	M	05	18
10	A	20130214	GA	GA	USA	089	DEKALB CO.	13121009402	1	1	GA	1	M	01	28
11	W	20130509	GA	GA	USA	121	FULTON CO.	13121011414	1	1	GA	6	M	01	18
12	A	20130501	GA	GA	USA	009	BALDWIN CO.	13009970701	1	1	GA	2	M	01	33
13	A	20130320	GA	GA	USA	009	BALDWIN CO.	13009970701	1	1	GA	6	M	02	23
14	A	20130418	GA	GA	USA	009	BALDWIN CO.	13009970701	1	1	GA	6	M	05	28
15	W	20130430	GA	GA	USA	059	CLARKE CO.	13059130300	1	1	GA	1	M	01	32
16	A	20130629	GA	GA	USA	067	COBB CO.	13067031409	1	1	GA	6	F	01	31
17	A	20130425	GA	GA	USA	185	LOWNDES CO.	13185010101	1	1	GA	6	F	05	34

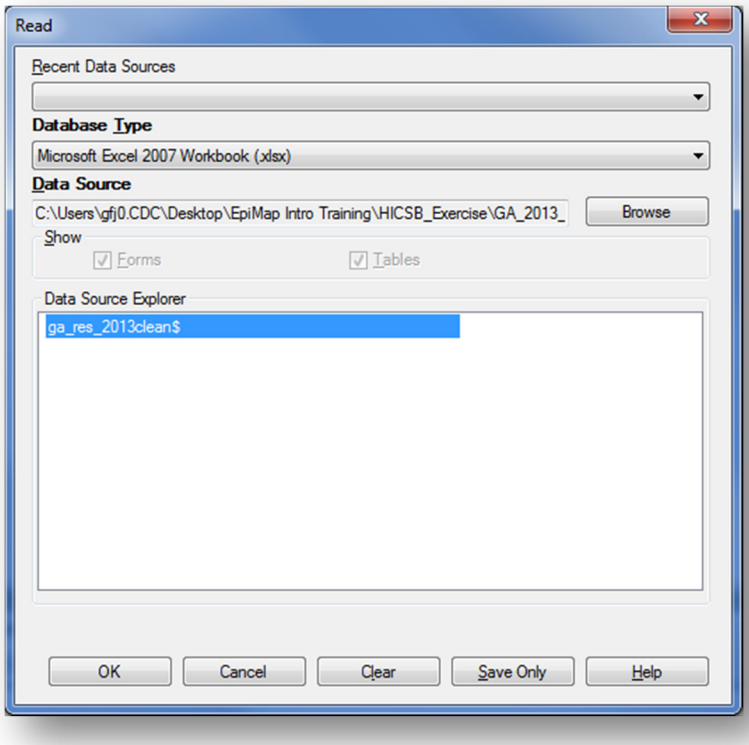
The instructor will demonstrate how to pivot Excel data. Below is an extract of spreadsheet of **2172** records of individual HIV case data. EpiMap requires aggregate or summary data and a numeric count field. Classic Analysis can pivot the Excel data into aggregate data.



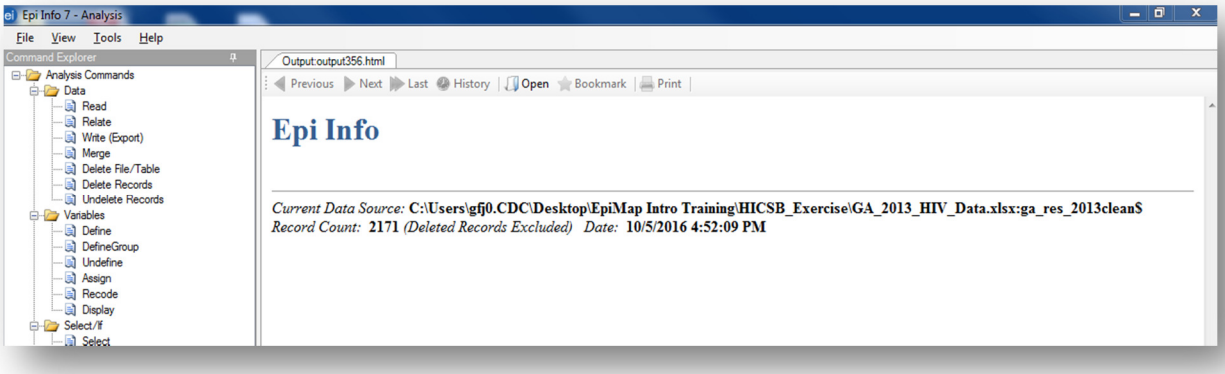
1. Read the Excel spreadsheet **GA_2013_HIV_Data.xlsx.** Select data type XLSX.



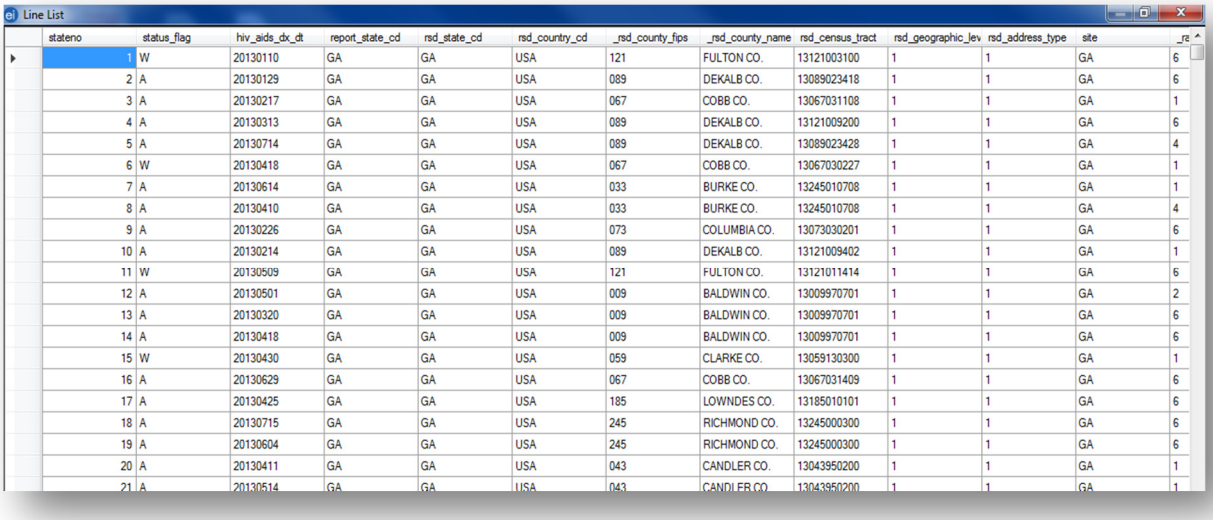
2. Read **GA_2013_HIV_Data.xlsx** file.



3. Select **ga_clean_data** sheet from the Excel file. A record count of **2171** is indicated.

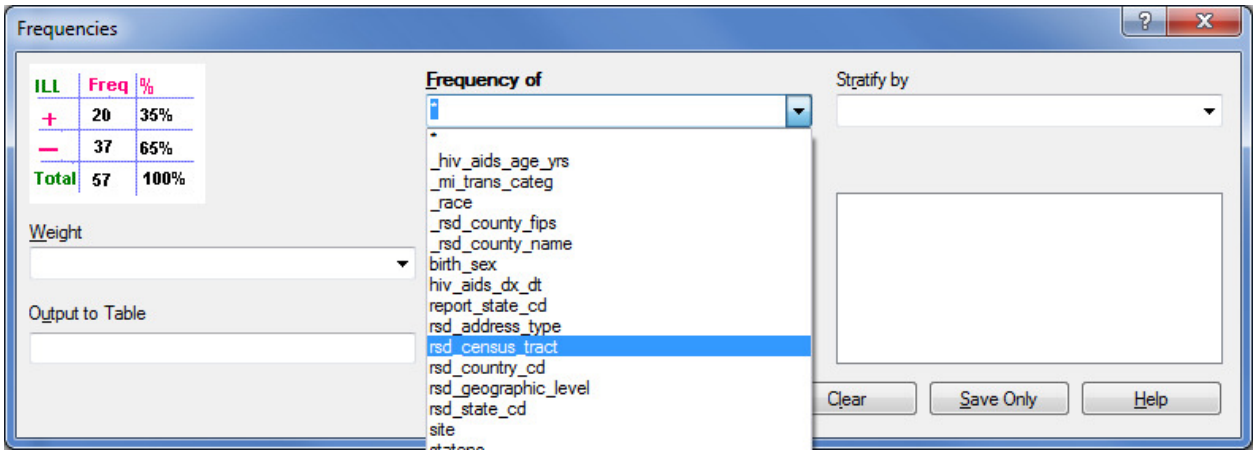


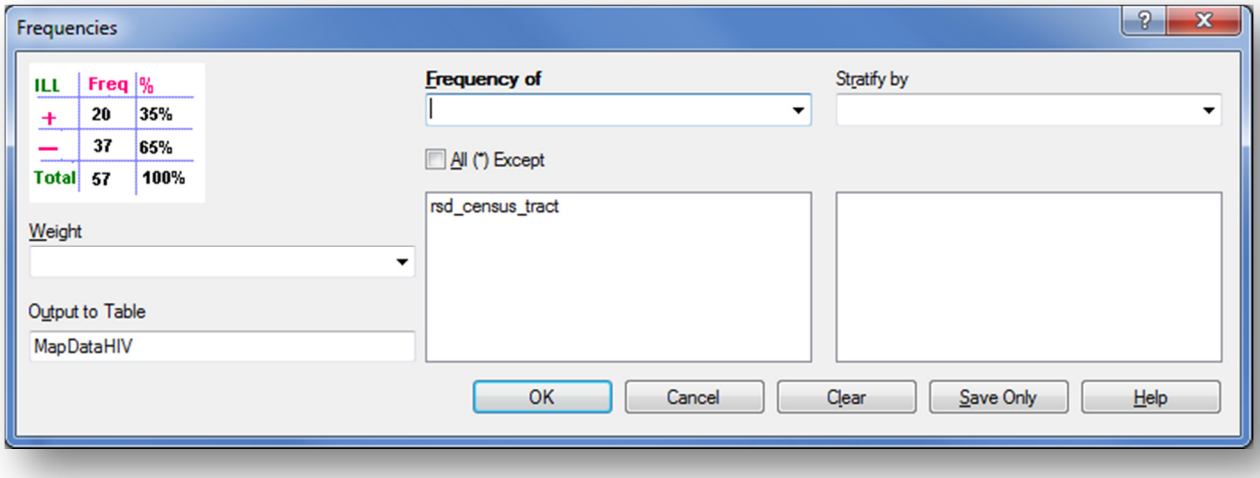
4. Select **List** form the command bar to get a spreadsheet view of the dataset.



	stateno	status_flag	hiv_aids_dx_dt	report_state_cd	rsd_state_cd	rsd_country_cd	_rsd_county_fips	_rsd_county_name	rsd_census_tract	rsd_geographic_lev	rsd_address_type	site	_re
▸	1	W	20130110	GA	GA	USA	121	FULTON CO.	13121003100	1	1	GA	6
	2	A	20130129	GA	GA	USA	089	DEKALB CO.	13089023418	1	1	GA	6
	3	A	20130217	GA	GA	USA	067	COBB CO.	13067031108	1	1	GA	1
	4	A	20130313	GA	GA	USA	089	DEKALB CO.	13121009200	1	1	GA	6
	5	A	20130714	GA	GA	USA	089	DEKALB CO.	13089023428	1	1	GA	4
	6	W	20130418	GA	GA	USA	067	COBB CO.	13067030227	1	1	GA	1
	7	A	20130614	GA	GA	USA	033	BURKE CO.	13245010708	1	1	GA	1
	8	A	20130410	GA	GA	USA	033	BURKE CO.	13245010708	1	1	GA	4
	9	A	20130226	GA	GA	USA	073	COLUMBIA CO.	13073030201	1	1	GA	6
	10	A	20130214	GA	GA	USA	089	DEKALB CO.	13121009402	1	1	GA	1
	11	W	20130509	GA	GA	USA	121	FULTON CO.	13121011414	1	1	GA	6
	12	A	20130501	GA	GA	USA	009	BALDWIN CO.	13009970701	1	1	GA	2
	13	A	20130320	GA	GA	USA	009	BALDWIN CO.	13009970701	1	1	GA	6
	14	A	20130418	GA	GA	USA	009	BALDWIN CO.	13009970701	1	1	GA	6
	15	W	20130430	GA	GA	USA	059	CLARKE CO.	13059130300	1	1	GA	1
	16	A	20130629	GA	GA	USA	067	COBB CO.	13067031409	1	1	GA	6
	17	A	20130425	GA	GA	USA	185	LOWNDES CO.	13185010101	1	1	GA	6
	18	A	20130715	GA	GA	USA	245	RICHMOND CO.	13245000300	1	1	GA	6
	19	A	20130604	GA	GA	USA	245	RICHMOND CO.	13245000300	1	1	GA	6
	20	A	20130411	GA	GA	USA	043	CANDLER CO.	13043950200	1	1	GA	1
	21	A	20130514	GA	GA	USA	043	CANDLER CO.	13043950200	1	1	GA	1

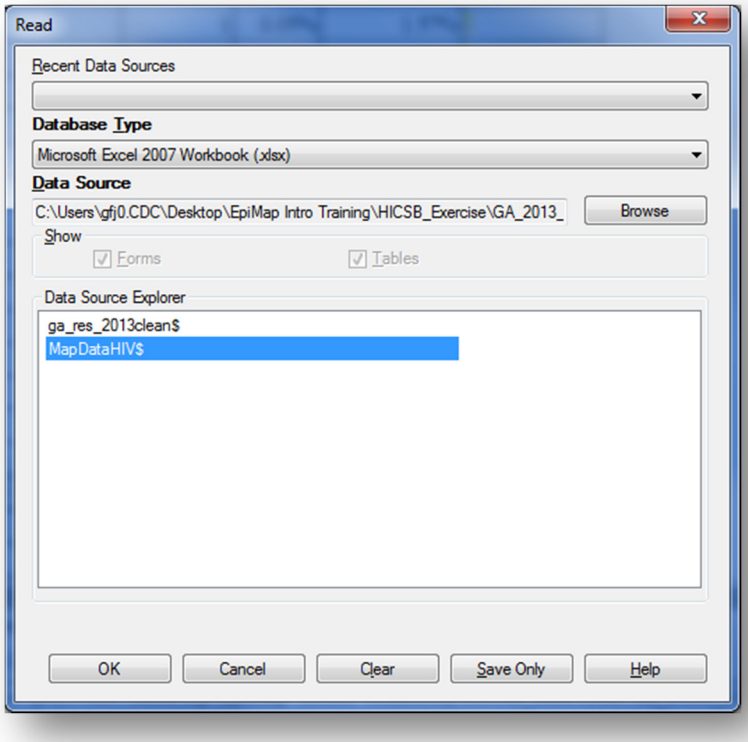
5. Select Frequency from the command bar. Select **rsd_census_tract** as the Frequency Of variable.



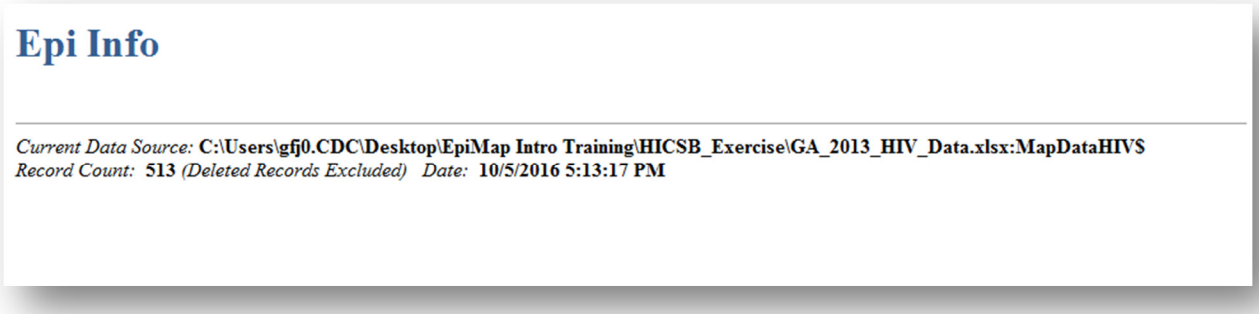


6. Type “MapDataHIV” as the Output Table name. Click OK to create the frequency table.

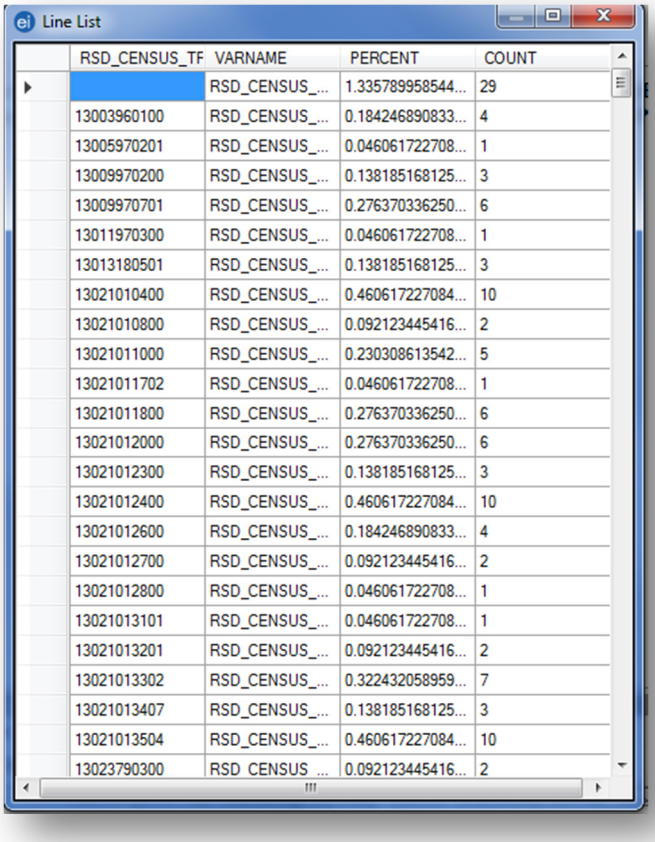
RSD_CENSUS_TRACT	Frequency	Percent	Cum. Percent	
	29	1.34%	1.34%	
13003960100	4	0.18%	1.52%	
13005970201	1	0.05%	1.57%	
13009970200	3	0.14%	1.70%	
13009970701	6	0.28%	1.98%	
13011970300	1	0.05%	2.03%	
13013180501	3	0.14%	2.16%	
13021010400	10	0.46%	2.63%	
13021010800	2	0.09%	2.72%	
13021011000	5	0.23%	2.95%	
13021011702	1	0.05%	2.99%	
13021011800	6	0.28%	3.27%	
13021012000	6	0.28%	3.55%	
13021012300	3	0.14%	3.68%	
13021012400	10	0.46%	4.15%	
13021012600	4	0.18%	4.33%	
13021012700	2	0.09%	4.42%	
13021012800	1	0.05%	4.47%	
13021013101	1	0.05%	4.51%	



7. Repeat the steps done earlier to read the newly created Excel spreadsheet tab MapDataHIV$.



8. The record count shows **513** records.



9. Select List from the command bar to display the aggregate counts of by census tract.

The dataset is now pivoted and ready to be mapped.

Creating a Choropleth Map of Household Income and a Dot Density Map of HIV cases by Census Tract

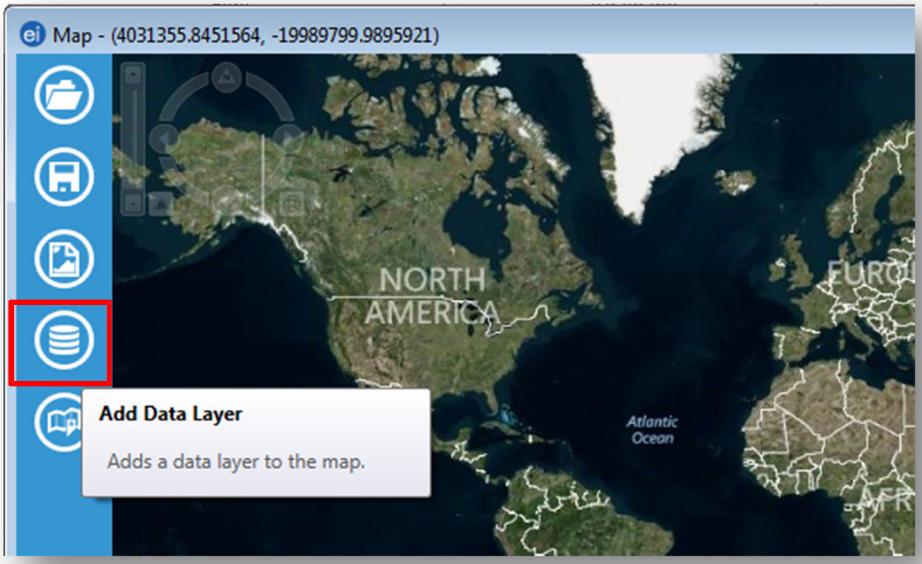
In this lesson, you will create and customize a choropleth map and save it as a .map7 interactive map file.

The dataset pivot from case to aggregate data of HIV cases is ready for mapping. To correlate cases of HIV and income, you will create a choropleth map using income data and then map cases of asthma from newly created dataset. This will help you determine if a link exists between income and HIV. First, create the choropleth map and then overlay the dot density data on to the map.

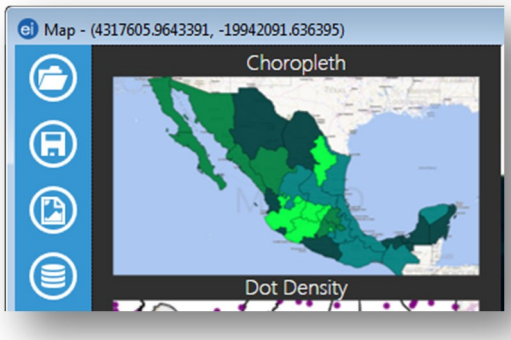
New EpiMap interface is part of 7.2. Our team updates often but not the interface.

1. Open Epi Map.

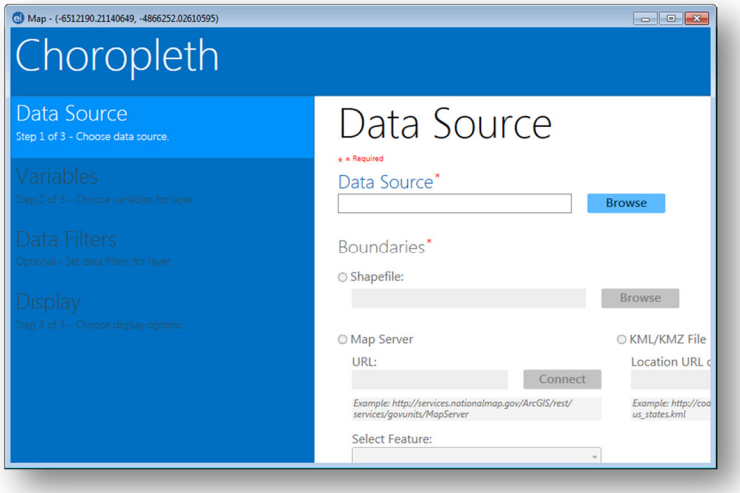
2. Click the Add Data Layer menu option



3. Click on Choropleth.



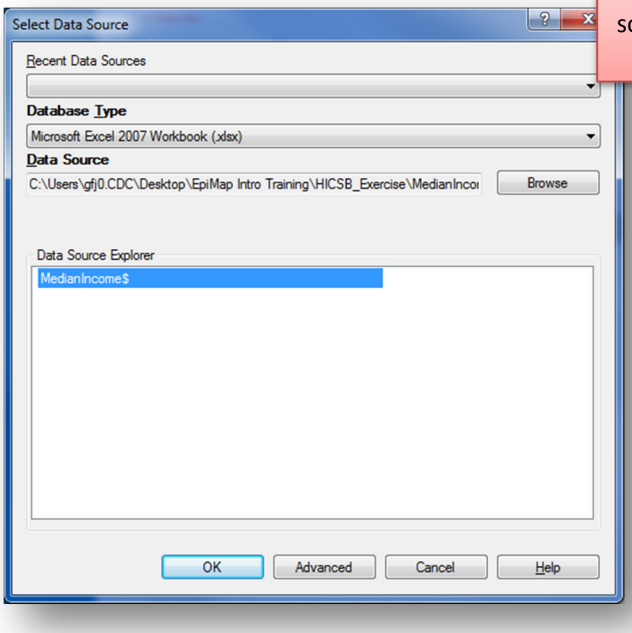
4. The Choropleth Map Configuration window opens.



5. The excel file need some revision. The ACS Excel file is very large. Actually to learge to map.

5. Click on the **Browse** button. The Select Data Source window opens.

As we saw earlier, you can map data sources other than just Epi Info™ 7!



6. Select **MS Excel** as the Database Type. Click the **Browse** button for the **Data Source.** Locate the MS Excel file **MedianIncomeByCensusTract.xlsx** in the folder.

Desktop\EpiMap Intro Training\HICSB_Exercise

7. Click OK.
8. Make sure the **"First row contains header information"** is checked.

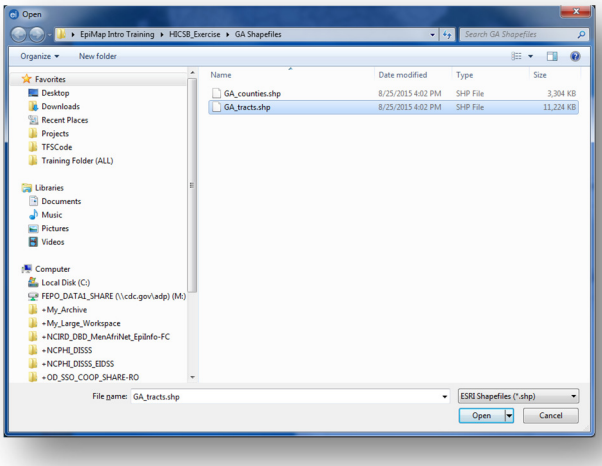
   - The demographic project has two tables. One contains income data and the other race and ethnicity data. You want to create a choropleth map using the income data. The income table contains a series of zip codes and the median income and per capita income for those zip codes. The median income is the value of family income with 50% of families above the value and 50% of the families below the value. Per capita income is the total income for that zip code divided by the population.

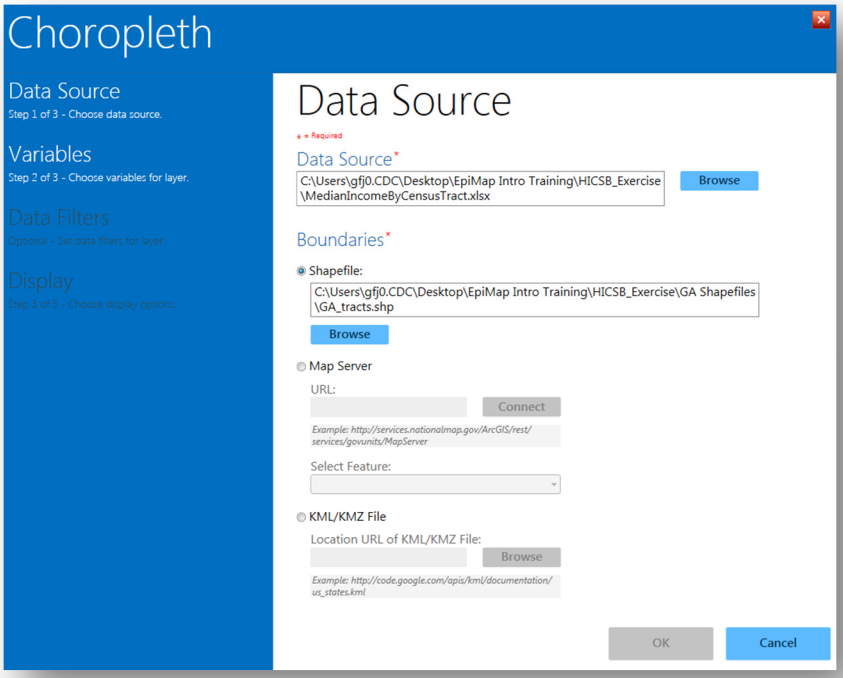
9. Select MedianIncome from the Data Source Explorer.
10. Click OK.
11. Click on the **Browse** button next to the Shape File radio button in the Boundaries section.

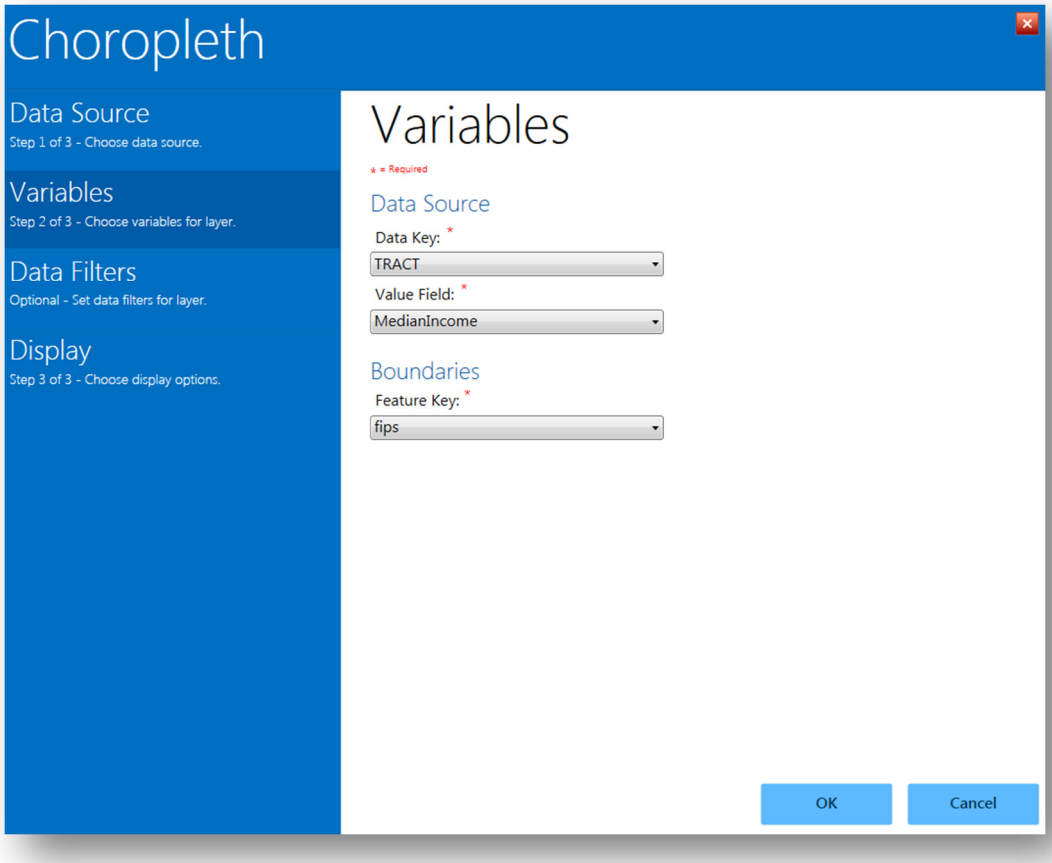
12. Locate the file **cb_2013_13_tract_500k** in the

    Epi Info 7/Resources/Training Projects/Community Health Assessment folder.

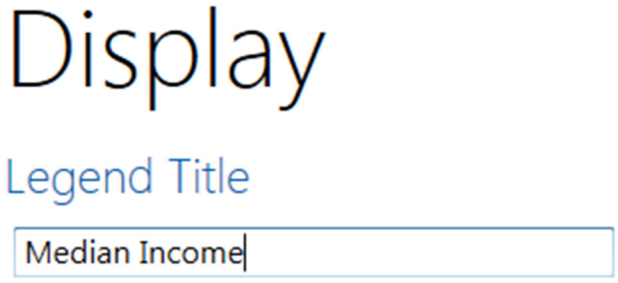
13. Click Open.




Choropleth
Data Source
Step 1 of 3 - Choose data source.
Variables
Step 2 of 3 - Choose variables for layer.
Data Filters
Optional - Set data filters for layer.
Display
Step 3 of 3 - Choose display options.
Data Source
* = Required
Data Source*
C:\Users\gfj0.CDC\Desktop\EpiMap Intro Training\HICSB_Exercise\MedianIncomeByCensusTract.xlsx
Browse
Boundaries*
Shapefile:
C:\Users\gfj0.CDC\Desktop\EpiMap Intro Training\HICSB_Exercise\GA Shapefiles\GA_tracts.shp
Browse
Map Server
URL:
Connect
Example: http://services.nationalmap.gov/ArcGIS/rest/services/govunits/MapServer
Select Feature:
KML/KMZ File
Location URL of KML/KMZ File:
Browse
Example: http://code.google.com/apis/kml/documentation/us_states.kml
OK
Cancel

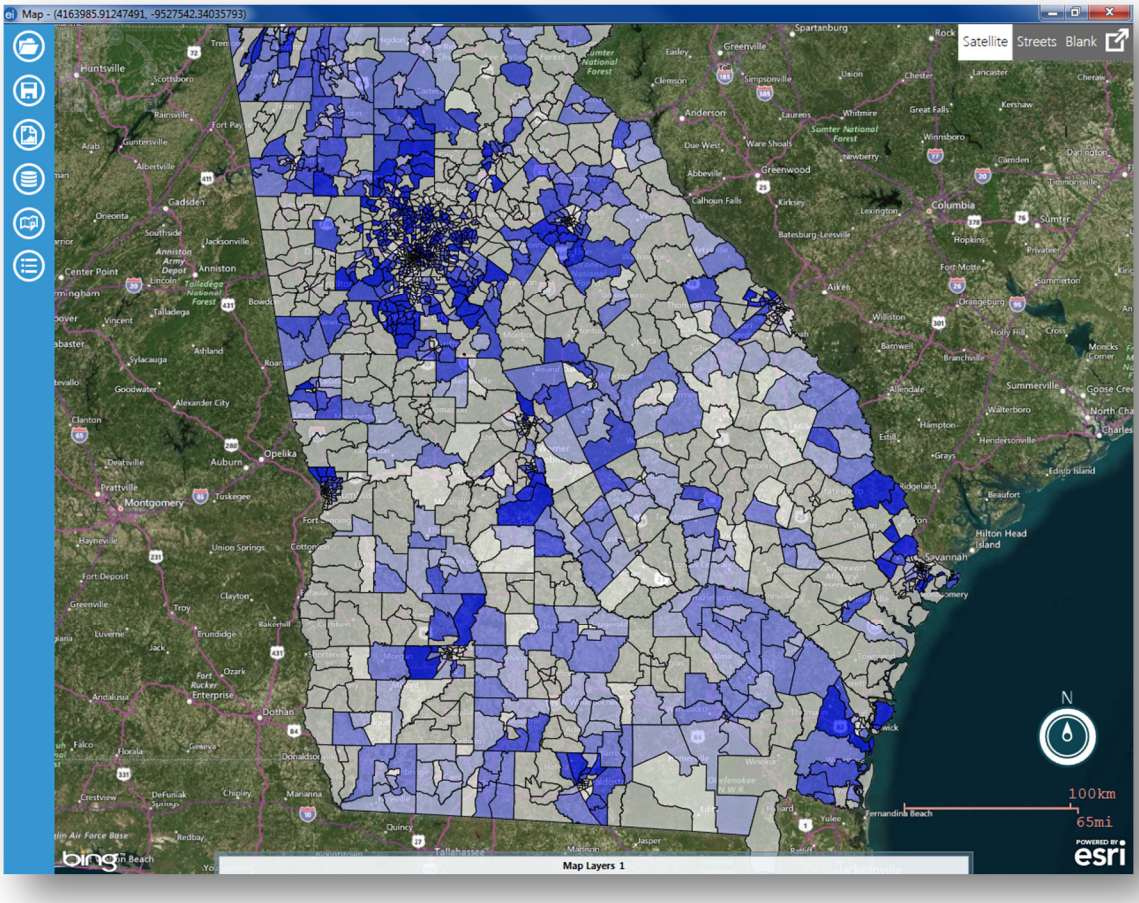


14. In the Data Source section, select **TRACT** from the Data Key Field drop down list.
    - ❖ This is the field inside your data that matches the key field in your shape file.

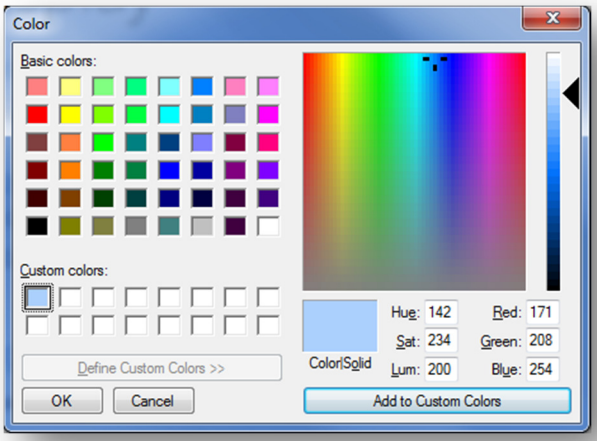


15. In the Data Source section, select **MEDIANFAMILY** From the Value Field drop down list,
    - ❖ This is the data field to be mapped.
16. From the Boundaries section, select **GEOID** from the Feature Key Field drop down list.
    - ❖ This is the field inside the shape file that matches key field in your dataset.
17. Keep the number of classes at 5. The number of classes shows how many divisions or groups into which your data will be divided on the map.
18. Type **"Median Income"** as the Legend Title.

19. Click OK. The choropleth map appears. Your map should look like the one below.



20. Position the cursor (**hover**) above the Map Layers manager and click the map configuration icon (Edit data layer).
21. Click the Start Color from Display menu. Change the Start Color to a custom color of light blue by clicking the **Define Custom Color** icon.

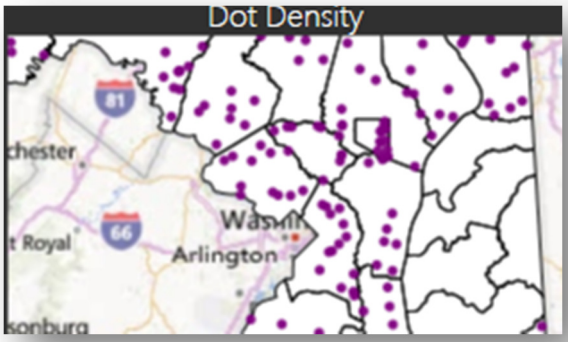


22. Click the **Add Custom Colors** icon, select **Custom color** and click **OK**.
23. Change the color of **Missing / Excluded** from grey to white.

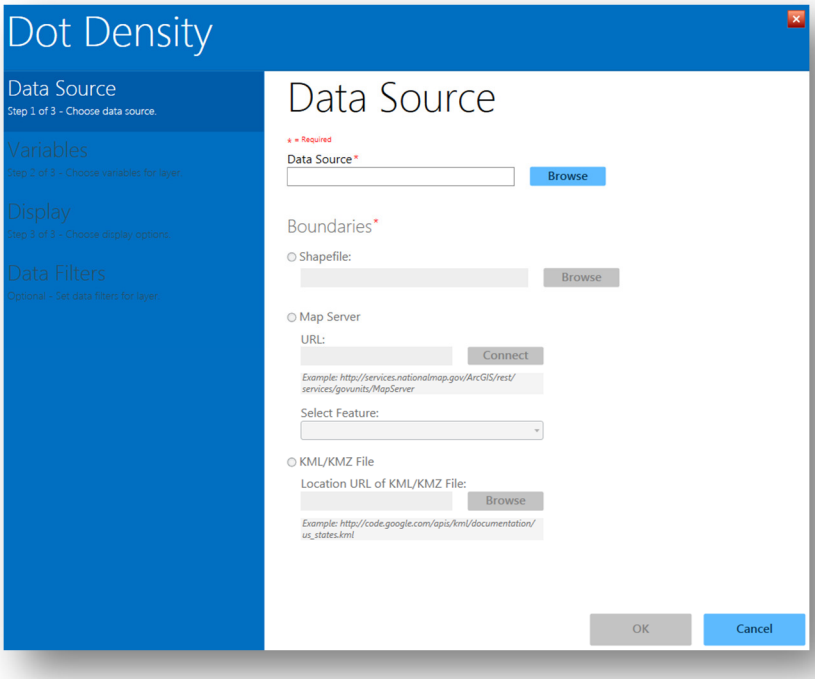
Adding the Dot Density Map

1. Click the Add Data Layer menu option.

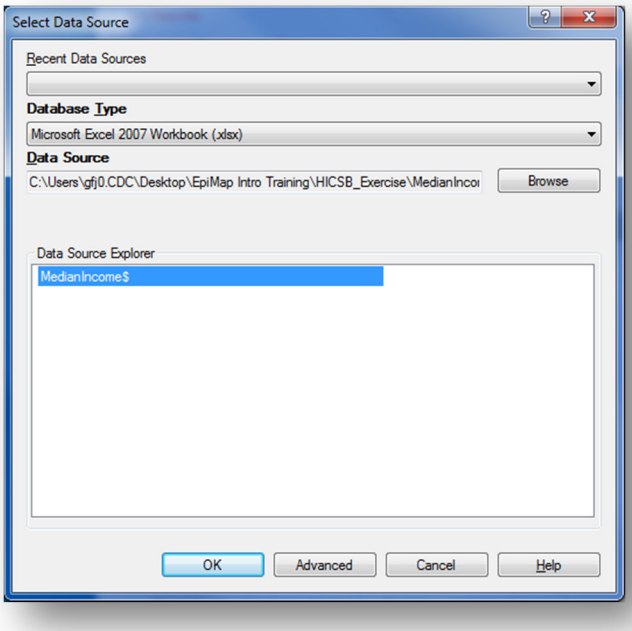
2. Select Dot Density as the map type.



3. The Dot Density Map Configuration window opens.



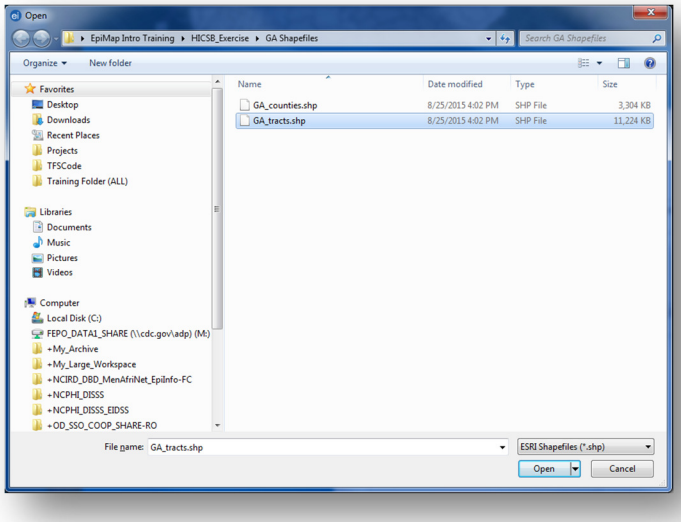
4. Click on the **Browse** button. The Select Data Source window opens.



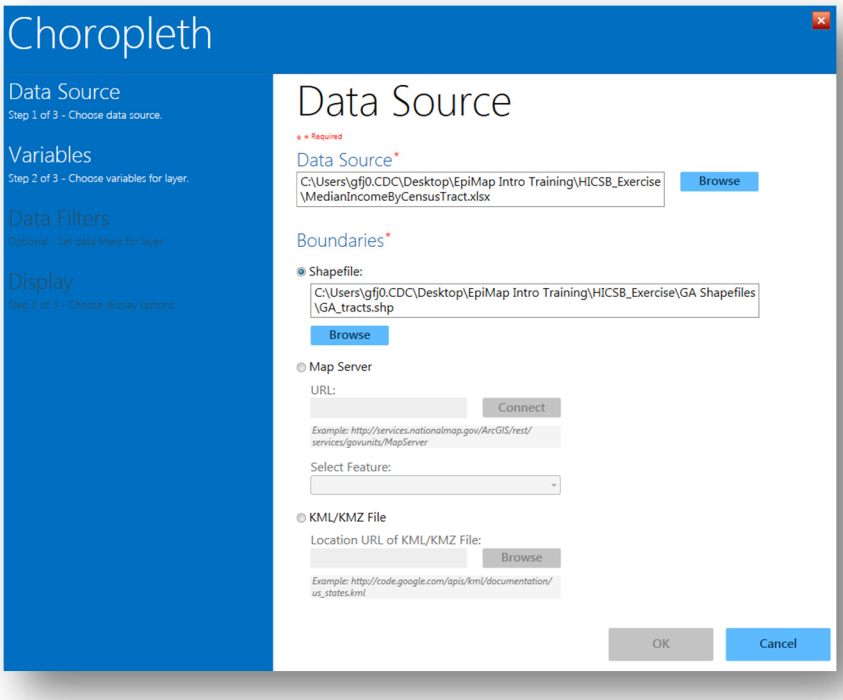
5. Select **MS Excel** as the Database Type. Click the **Browse** button for the **Data Source.** Locate the MS Excel file **GA_2013_HIV_Data.xlsx** in the folder.

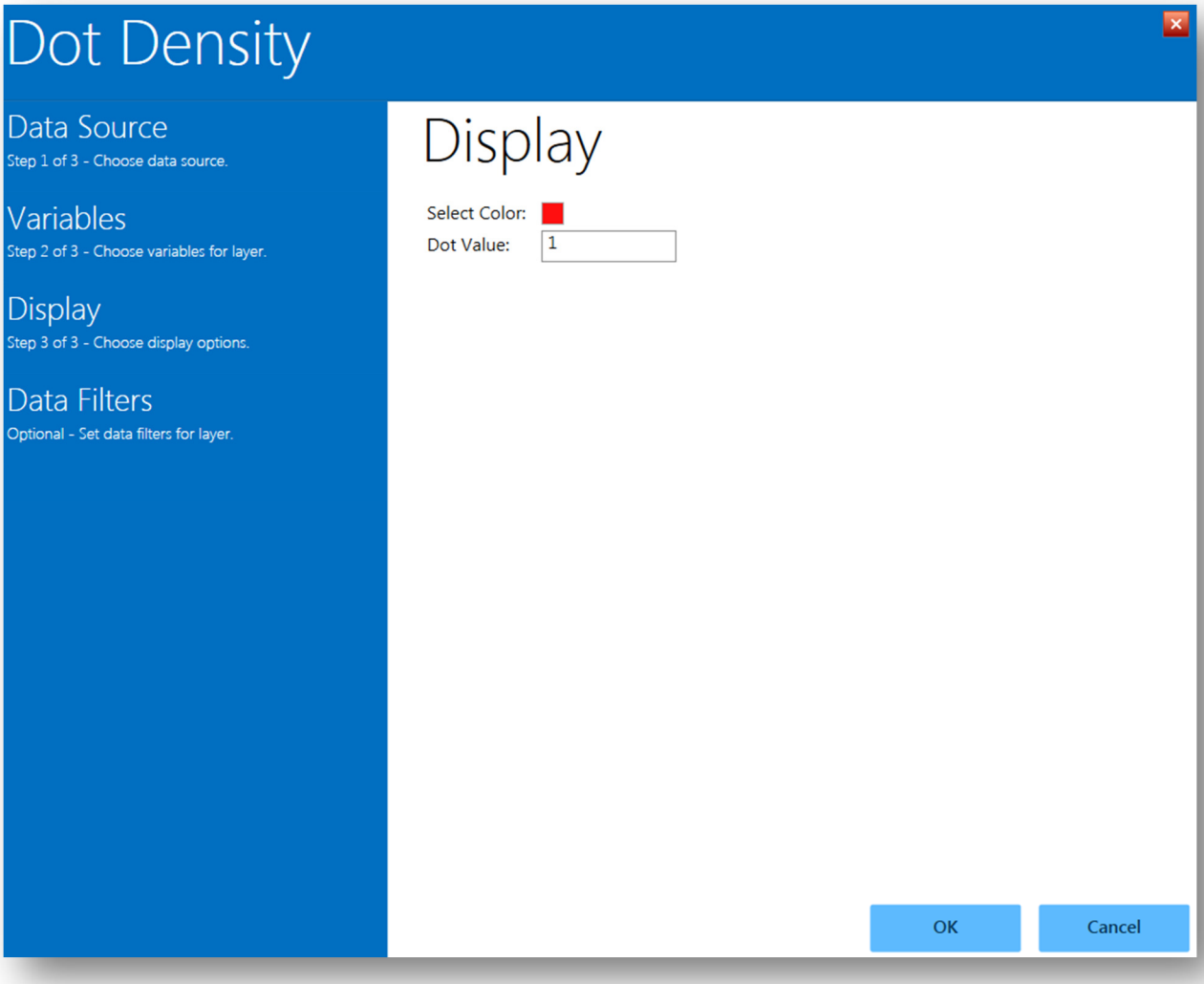
   Desktop\EpiMap Intro Training\HICSB_Exercise

6. Make sure the "First row contains header information" is checked.
7. Select the **MapDataHIV** tab and click OK.
8. Click on the **Browse** button next to the Shape File radio button in the Boundaries section.
9. The steps are the same as the choropleth map previously.



10. Locate the file **cb_2013_13_tract_500k** in the Epi Info 7/Resources/Training Projects/Community Health Assessment folder.
11. Click Open.





12. In the Data Source section, select **RSD_CENSUS_TRACT** from the Data Key Field drop down list.
    a. This is the field inside your data that matches the key field in your shape file.
13. In the Data Source section, select **COUNT** From the Value Field drop down list,
    a. This is the data field to be mapped.
14. From the Boundaries section, select **GEOID** from the Feature Key Field drop down list.
    a. This is the field inside the shape file that matches key field in your dataset.
15. Change the color to red on the display tab.
    - ❖ *Processing for the Dot Density map will take a few seconds to complete.*
16. Select **Blank** as the map view.

Appendix:

HIV Surveillance Data – Data Dictionary

Variable	Format	Example
STATUS_FLAG	1-Character variable	A
REPORT_STATE_CD	2-Character variable	GA (for Georgia)
STATENO	Every line entry should have a stateno associated with it.	
CITYNO	Similar to STATENO	**For NY City only**
HIV_AIDS_DX_DT	yyyymmdd 8-character variable	20140210
RSD_COUNTY_FIPS	3-Character variable	121
RSD_COUNTY_NAME	Character variable	FULTON
RSD_COUNTRY_CD	3-Character variable	USA or PRI (for Puerto Rico)
RSD_STATE_CD	2-Character variable	GA
RSD_GEOGRAPHIC_LEVEL	1-Character variable **1**=Census tract **2**=Zip code **3**=County **4**=Insufficient	1
RSD_ADDRESS_TYPE (Note: This is not related to eHARS ADDRESS_TYPE_CODE variable, see Section 3 Obtaining Address Type) ¥	1-Character variable **1**=Residential **2**=Post office box **3**=Insufficient **4**=Missing **5**=Corrections **6**=Military **7**=Homeless **8**=Other	2
RSD_CENSUS_TRACT*	11-Character variable	13121123456
_RACE	1-Character variable **1**=Hispanic, all races **2**=Not Hispanic, Amer Indian/Alaska Nat **3**=Not Hispanic, Asian **4**=Not Hispanic, Black **5**=Not Hispanic, Nat Hawaiian/Pac Isl **6**=Not Hispanic, White **7**=Not Hispanic, Legacy Asian/Pac Isl **8**=Not Hispanic, Multi-race **9**=Unknown	6
_MI_TRANS_CATEG	2-Character variable **01**=Adult MSM **02**=Adult IDU **03**=Adult MSM & IDU **04**=Adult recd. Clotting factor	02

	05=Adult Hetero. Contact **06**=Adult rcd. Transf/transplant **07**=Perinatal Exposure age 13+ years **08**=Adult other confirmed risk **09**=Adult NIR **10**=Adult NRR **11**=Child recd. Clotting factor **12**=Perinatal Exposure **13**=Child rcd. Transf/transplant **18**=Child other confirmed risk **19**=Child NIR **20**=Child NRR **99**=Risk factors selected w/no age at dx	

The US Census TIGER/Line web site is a good source of US Shape files with population variables:

- **https://www.census.gov/cgi-bin/geo/shapefiles2010/main**
- **http://www.census.gov/geo/maps-data/data/tiger-data.html**
- **http://www.census.gov/cgi-bin/geo/shapefiles2010/main**
- **http://factfinder2.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t**

For shape files for other countries, you can try:

- **http://www.diva-gis.org/gData**