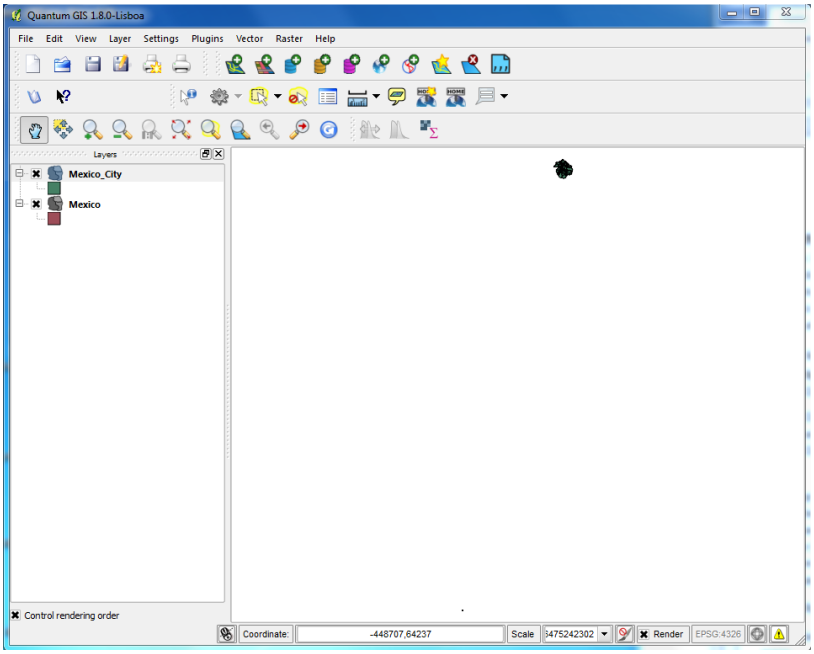
LESSON 3 Local Scale Analysis - Exploring Demographic and Air Quality Data for Mexico City

Jennifer Boehnert, National Center for Atmospheric Research

In this lesson, we will work with socio-economic, demographic, and air quality data for Mexico City municipalities. We will derive new data, based on existing attributes, learn how to join non-spatial attributes to spatial data, work with data selections, and learn how to conduct more sophisticated spatial analysis, using different types of data.

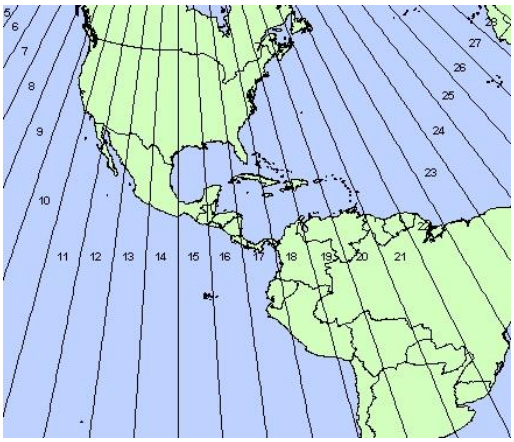
STEP 1 Working with data which are in different projections

- **Open** the QGIS Desktop application.
- Click **OK** when the QGIS tip appears.
- Click the **Add Vector Layer**  button.
- In the dialog box which opens, navigate to **<your working directory>\data** and select **Mexico.shp** hold down your Ctrl (Windows) or Command (Mac) and select **Mexico_City.shp**.
- Click **Open** to add the data.



Your map view may seem a little strange (see screenshot above). There is data on the top of your screen and data at the bottom of your screen. This is because the *Mexico_City.shp* shapefile is in a projected coordinate system called Universal Transverse Mercator (UTM) and *Mexico.shp* shapefile is in a WGS 84 coordinate system.

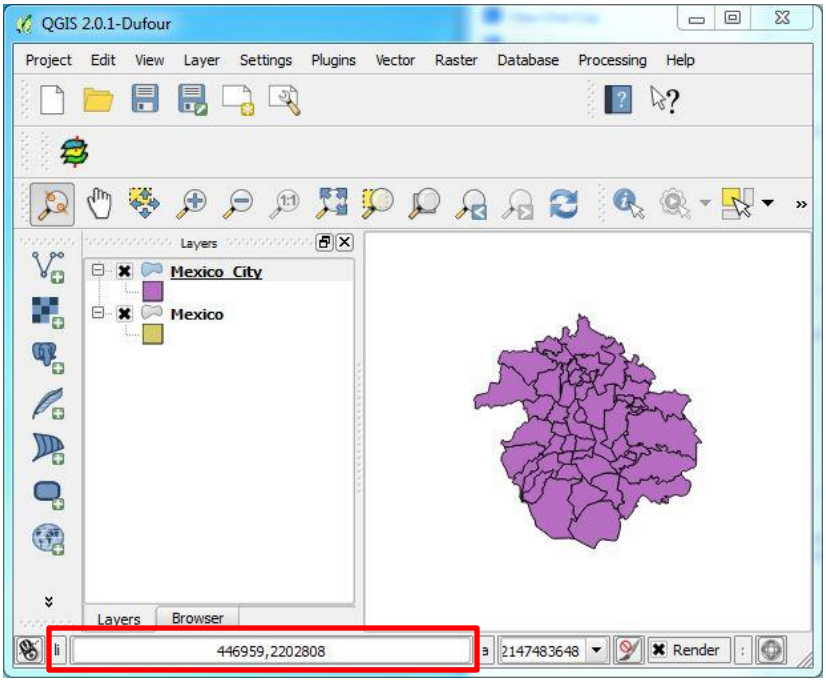
The image below is an example of UTM zones and how they span across the Americas.



The map is currently projected to WGS_1984_UTM_ Zone _14N. UTM stands for Universal Transverse Mercator coordinate system. It is a grid-based coordinate system where the world is divided up into 60 zones; each zone is 6° of longitude, with units in meters. Mexico City falls within Zone 14.

- Right-click on *Mexico_City* in the TOC and select **Zoom to Layer Extent**.

Notice that we can now see the *Mexico_City* municipalities.



- Move your mouse around in the map view window and take a look at the Coordinate values which appear at the bottom of the map.

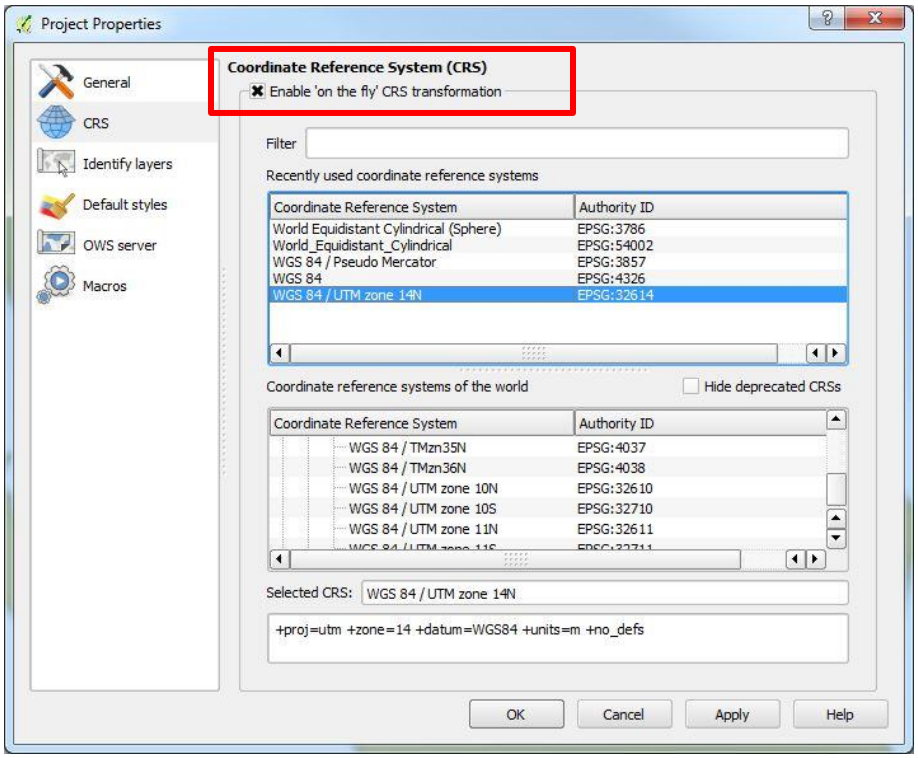
These values are not in the Geographic coordinate space of (-90, 90 and -180, 180), but are in UTM space (meters).

- Right-click on *Mexico* in the TOC and select **Zoom to Layer Extent**.
- Move your mouse around in the map view window and take a look at the Coordinate values which appear at the bottom of the map.

These values are in the Geographic coordinate space of (-90, 90 and -180, 180).

In QGIS you can set up the map so that it will overlay data from different projections.

- Click **Project > Project Properties (Windows)**.
- Click **File > Project Properties (Mac).**
- In the dialog which appears click the **CRS** option.
- Check the box to "Enable 'on the fly' CRS transformation".
- Click **WGS 84 / UTM Zone 14N.**



- Click **OK**.

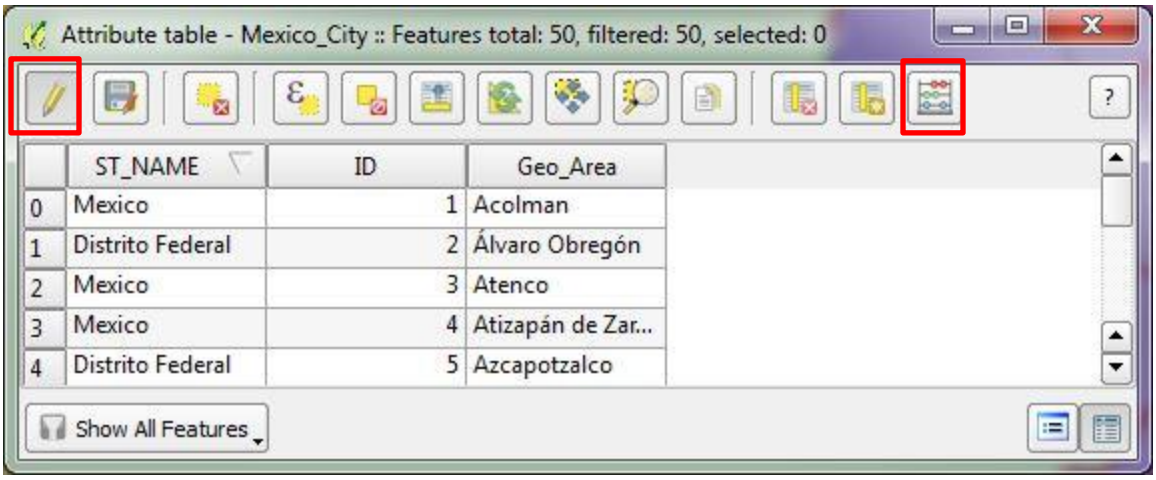
In your map, you should now see *Mexico_City* municipalities overlaid on top of the *Mexico* layer.

- Right-click on *Mexico_City* in the TOC and select **Zoom to Layer Extent**.

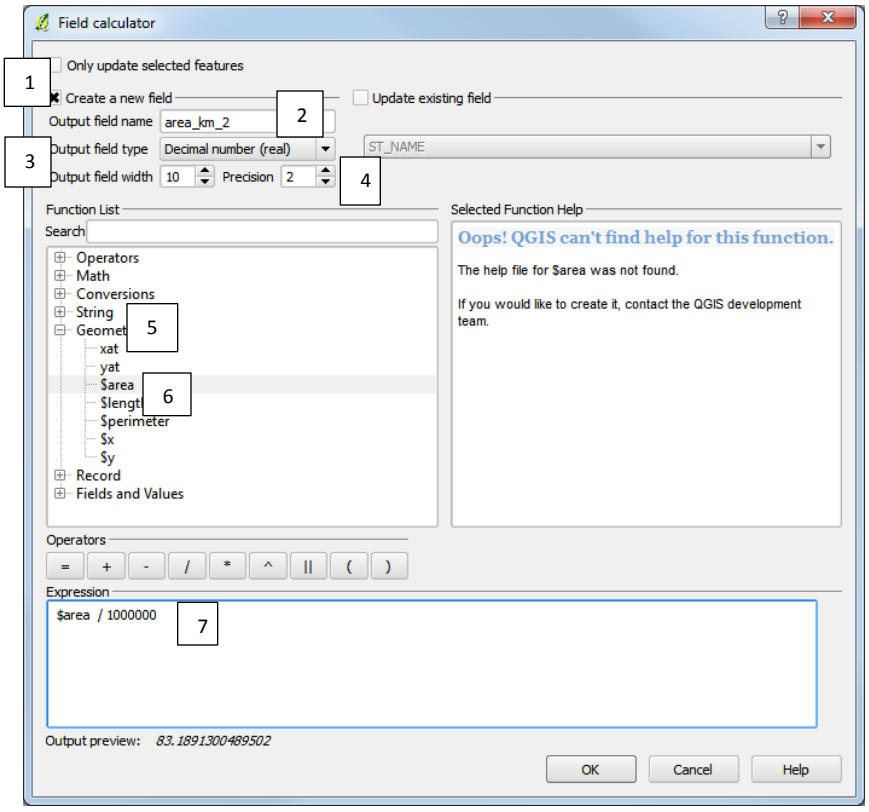
STEP 2 Create new fields and calculate values – calculate area for each municipalities in Mexico City

Now we will create new data in our dataset. These data will be stored in two new fields in the attribute table. The first will hold population density values for each municipality and the second will hold the area (in square kilometers) of each municipality.

- In the TOC, right-click on the layer *Mexico_City*, and select **Open Attribute Table** from the context menu



- In the Attribute table for *Mexico City* click the **Toggle Editing Mode** .
- Click the **Open field Calculator** button.



1. In the Field Calculator dialog box make sure *Create a new field* is checked.

2. For *Output field Name* type in **area_km_2.**

3. For *Output field type* select **Decimal number (real).**

4. For *Precision* change to **2**. This is the total number of decimal places the new field will contain.

5. In the *Function List* expand **Geometry**.

6. Double-click on **$area** so that is appears in the Expression box.

7. Type **/ 1000000**. Click **OK**.

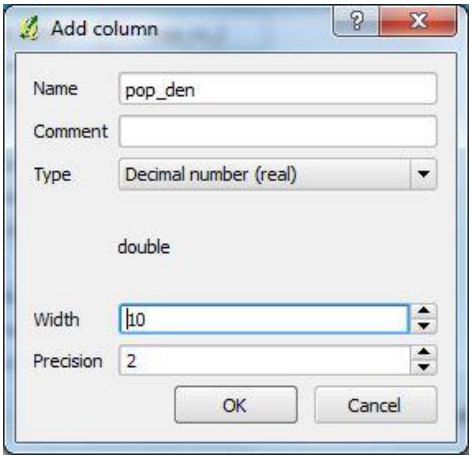
When using the geometry functions all geometry will be calculated using the units of the projections. *Mexico_City* is projected using the UTM coordinate system, with units in meters. Therefore, the area would be calculated into square meters. By applying a division by 1,000,000 we have converted the area into square kilometers ($1km^2 = 1{,}000{,}000m^2$).

Notice that the new field is added and populated with values which represent the area of each district in square kilometers.

- Click the **Save Edits** button  .

Next, we will add a population density field and calculate population density for each municipality.

- Click the **New Column**  button.



- For *Name* type in **pop_den.**
- For *Type* select **Decimal number (real).**
- For *Width* change to **10.** This is the total number of digits the new field will contain.
- For *Precision* change to **2.**
- Click **OK**.
- Click the **Save Edits** button.
- Click the **Toggle Edit Mode** button to stop editing.

Two new fields are created in the attribute table. We need a population field for each municipality before we can calculate population density. In the next step, we will join this spatial data with a non-spatial table of demographic data and then calculate population density.

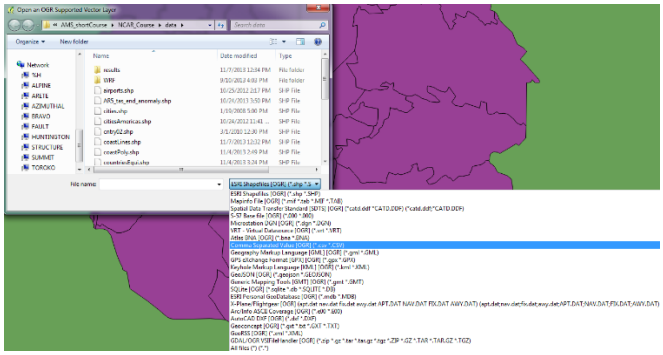
- Close the attribute table.

STEP 3 Perform a table join – joining demographic data to our Mexico City district layer

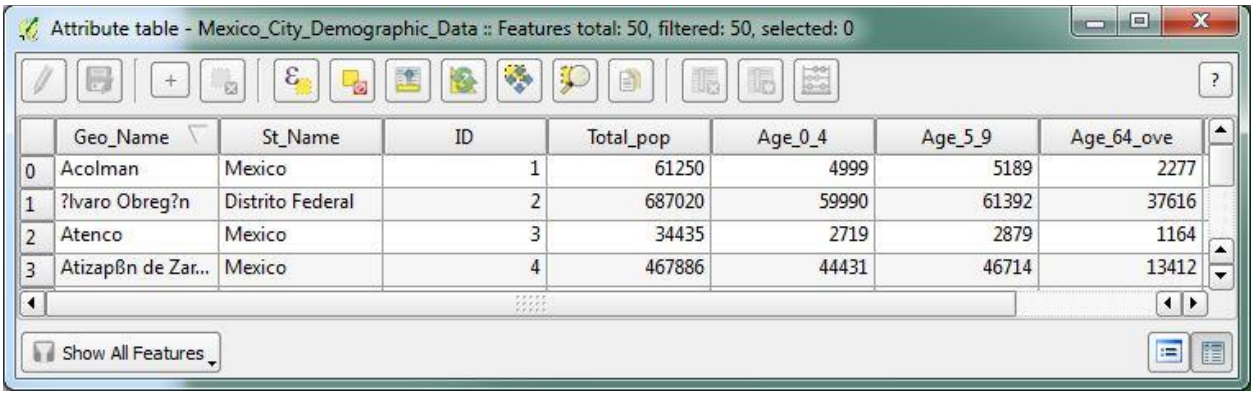
The *Mexico_City* layer has useful spatial and geometric information, but it does not contain any demographic information that we need for our analysis. We will now bring in and process demographic data in order to better understand distribution of population of different municipalities in Mexico City. We will be working with non-spatial (.csv) data which contains demographic information. The reason this data is non-spatial is because it does not contain a geometry or any location information.

To add demographic data to the existing shapefile, we will join data from two tables into one. **Table Join** is a process by which attributes from one table are appended to the attributes of another table, through a field, common to both tables.

- Click **Add vector layer** and navigate to **<your working directory>/data**.
- Change the file type to **Comma Separated Value [OGR].**



- Then double-click on **Mexico_City_Demographic_Data.csv.**
- Click **Open**.
- In TOC, open the *Mexico_City_Demographic _Data* by right-clicking on the table, and select **Open Attribute Table**.



	Geo_Name	St_Name	ID	Total_pop	Age_0_4	Age_5_9	Age_64_ove
0	Acolman	Mexico	1	61250	4999	5189	2277
1	?lvaro Obreg?n	Distrito Federal	2	687020	59990	61392	37616
2	Atenco	Mexico	3	34435	2719	2879	1164
3	Atizapßn de Zar...	Mexico	4	467886	44431	46714	13412

You will see that there are several fields of demographic information, as well as a field titled "*ID*". The "*ID*" field is a unique identifier (numbered 1-50) that is used to reference each municipality in Mexico City. Because the same identifiers are present in the *Mexico_City* layer, we will use this field as a basis for the **Table Join**.

- Close the attribute table for *Mexico_City_Demographic_Data*.
- Right-click on the layer *Mexico_City* in the TOC and select **Properties**.
- In the Layer Properties click on the **Joins** tab.

Currently, there are no joins for this layer.

- Click the button to add a join.
- In the Join dialog make sure that **Join layer** is set to *Mexico_City_Demographic _Data*.
- Change the *Join Field* to **ID**.
- Change the *Target field* to ***ID.***
- Keep all other defaults and click **OK**.
- Click the **Fields** tab.

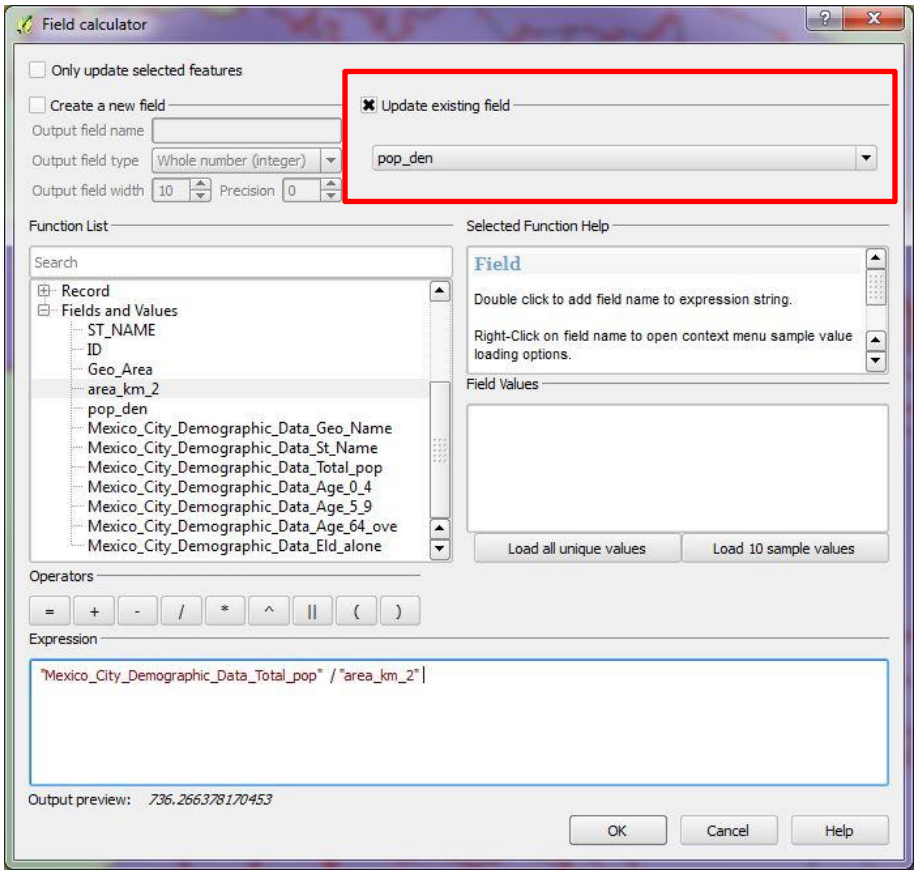
Notice now, that this layer contains all the fields from the original shapefile as well as all the attributes from the .csv. The reason the field types are specified as numbers and strings is because we created a .csvt file which defines the field types of each column.

- Click **OK** to close the Layer Properties.
- Look in **<your working Directory>\data** for Mexico_City_Demographic.csvt, and view it in your preferred text editor application.

STEP 4 Calculate fields – calculating population density from existing fields

When working with population data, it is useful to know not only the total number of population but also the population density. In this step, we will calculate the population density for each municipality in Mexico City.

- Open the attribute table for the *Mexico_City* layer.
- Click the **Toggle Edit Mode** button to start editing.



- Click the **Open field Calculator** button.

- Check the box to *Update existing field*.
- For the field to update select **pop_den**.
- Expand the **Fields and Values** category.
- Double-click on **Mexico_City_Demographic_Data.Total_pop** so that it appears in the Expression box.
- Type in **/** or click the  button.
- Double-click on **area_km_2**.

When your expressions looks like the one above, click **OK**.

- In the Attribute table dialog box, click the **Save** button to save your edits.
- Click the **Toggle Edit Mode** button to stop editing.

What is the most densely populated municipality in Mexico City and what is its population density? (HINT: use the sorting skills and selection skills you just learned)

- Close the *Mexico_City* attribute table.
- Click **File > Save Project As….**
- In the dialog which appears, navigate to **<your working directory>/maps** and save the map as **exercise3.qgs.**

STEP 5 Mapping spatial distribution of demographic data

We will now map the demographic data for Mexico City in order to better understand Mexico City's demographic makeup and its spatial distribution.

- Right-click on *Mexico_City* in your TOC and select **Zoom to Layer Extent**.
- Double-click on *Mexico_city* in the TOC in order to open the **Layer Properties**.
- Click the **Style** tab.
- Change the way the features are symbolized from **Single Symbol** to **Graduated**.
- Change the *Column* to *Mexico_City_Demographic_Data_Age_64_ove*. This column represents the number of people over the age 64.
- Change the *Color ramp* to **OrRd**.
- Leave the *Mode* to **Equal Interval**.
- Click **OK**.

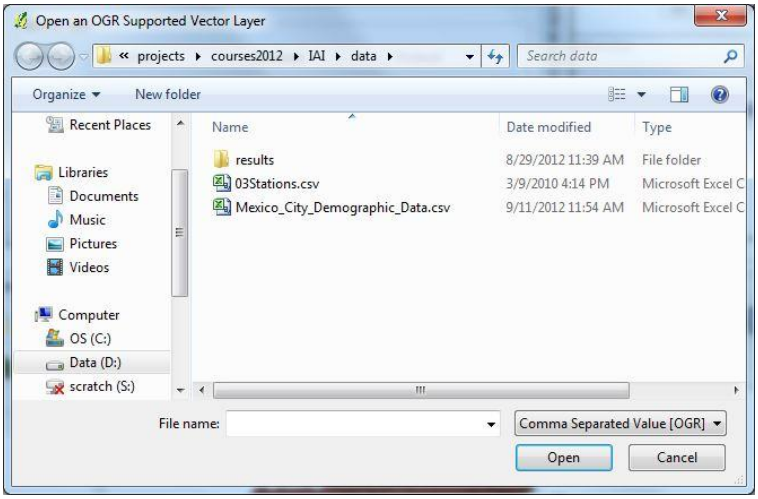
What municipality has the highest number of people 64 years old and older? (HINT: Use the Identify tool)

- Symbolize the data using the *pop_den* attribute.

STEP 6 Add X, Y data to QGIS - air quality monitoring station data

Our next step is to explore Mexico City's air quality. The World Health Organization (WHO) has recently published links between exposures to high levels of ozone and increased respiratory illness in children. We have a dataset of ozone concentrations averaged over 8 hours for the month of April (2006) for 20 monitoring stations throughout Mexico City (MILAGRO, 2006). The data are stored in a text file. We will overlay locations of air quality monitoring stations with the demographic data, and answer questions on *where* the highest ozone concentrations occur and *what* demographics may be affected by high levels of ozone.

- Click the **Add Vector Layer** button on the main toolbar



- In the Add Vector Layer dialog box click **Browse**.
- Make sure to set the *File type* to Comma Separated Value [0GR] (**.csv, *.CSV)*
- Navigate to **<your working directory>/data.**
- Click **03Stations.csv** and click **Open**.
- Click **Open** again to add the file to QGIS.

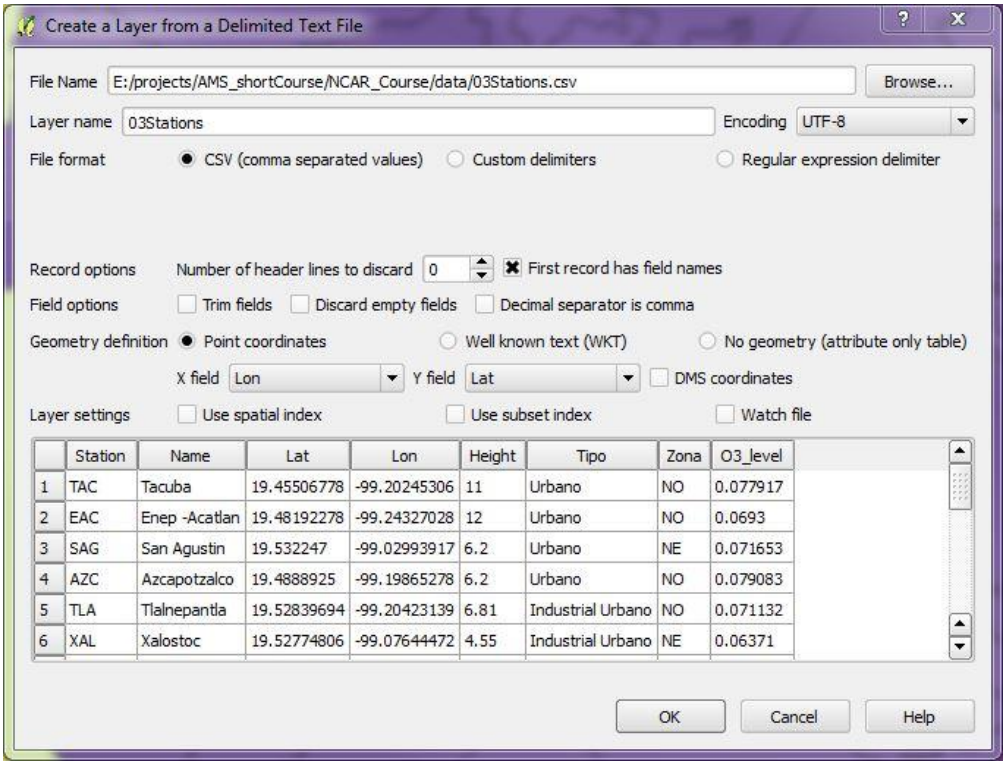
- Right-click on *03Stations* in the TOC and select **Open Attribute Table**.

This table contains 19 stations. The last column is an 8 hour average ozone concentration for the month of April.. This table is currently not spatial (it does not contain a feature or shape) but it does contain a latitude and a longitude column which we can use to plot the data on the map.

- After exploring the table, close the Attribute table.

- Click the **Add Delimited Text File** tool from the File Toolbar.
- In order to set the *File Name*, click the **Browse…** button.
- Navigate to **<your working directory>/data and select 03Stations.csv,** and click **Open.**
- For *Layer Name* type **03Stations**. This will be the name of the layer when it is added to QGIS.
- For *File format* select **CSV (comma separated values)**

- Keep all the defaults
- Once your dialog looks like the one below click **OK**.



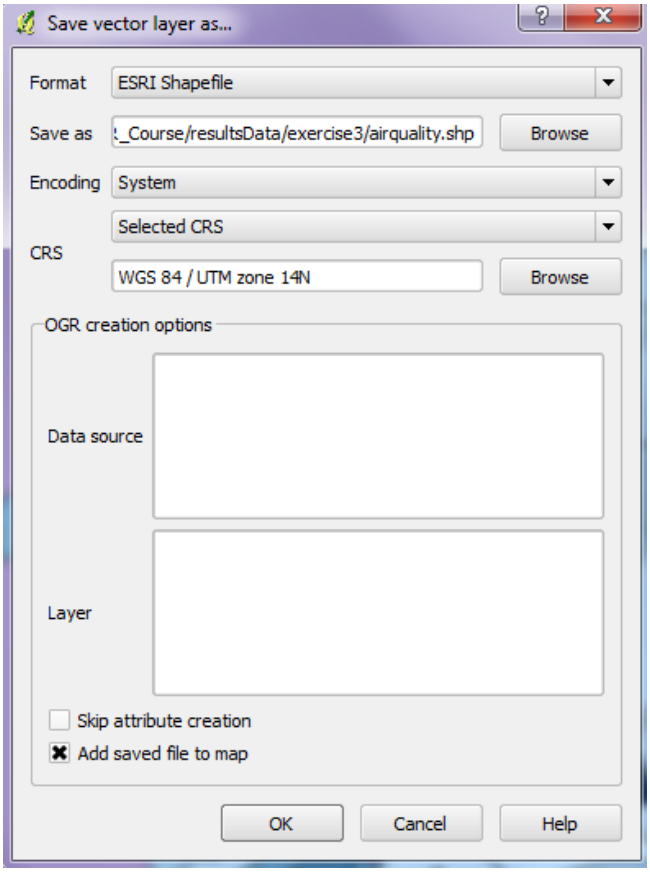
	Station	Name	Lat	Lon	Height	Tipo	Zona	O3_level
1	TAC	Tacuba	19.45506778	-99.20245306	11	Urbano	NO	0.077917
2	EAC	Enep -Acatlan	19.48192278	-99.24327028	12	Urbano	NO	0.0693
3	SAG	San Agustin	19.532247	-99.02993917	6.2	Urbano	NE	0.071653
4	AZC	Azcapotzalco	19.4888925	-99.19865278	6.2	Urbano	NO	0.079083
5	TLA	Tlalnepantla	19.52839694	-99.20423139	6.81	Industrial Urbano	NO	0.071132
6	XAL	Xalostoc	19.52774806	-99.07644472	4.55	Industrial Urbano	NE	0.06371

- In the *Coordinate Reference System Selector*, choose **WGS 84.** This data is in geographic (lat, lon).
- Click **OK** to continue adding the data.

The data, which is added to your map, are points which represent air quality monitoring stations. This layer is not stored as a spatial dataset but it is dynamically reading the latitude and longitude information from the .csv file. The air quality monitoring station data are in a Geographic Coordinate System and the Mexico City municipality data is in a projected coordinate system of UTM (Universal Transverse Mercator) zone 14. The two datasets overlay because we indicated in

QGIS to reproject data on the fly. However, since we will be doing spatial analysis with this data it is a good idea that all the layers have matching projection information. We will now export the air quality monitoring station data as a shapefile in the same projection as the municipality data.

- Right-click on *03Stations* in your TOC and select **Save As…**



- Click the **Browse** button to set the *Save as* output shapefile.
- Navigate to **<your working directory>/data/results** and name the file **airquality.shp,** and click **Save.**
- Click the **Browse** button to set the *CRS.*
- Click the option **WGS 84 / UTM zone 14N EPSG:32614.**
- Click **OK**.
- Check the option to **Add saved file to map**.
- Click **OK** to run the tool.
- Click **OK** once the tool is complete.

You will see dots scattered on your map, these are the air quality stations and the ozone measurement data from the CSV file is now associated with these points and projected in UTM coordinates.

- Remove the layer *03Stations* from your TOC.

STEP 8 Determining where ozone concentrations exceed the 8-hour WHO limits

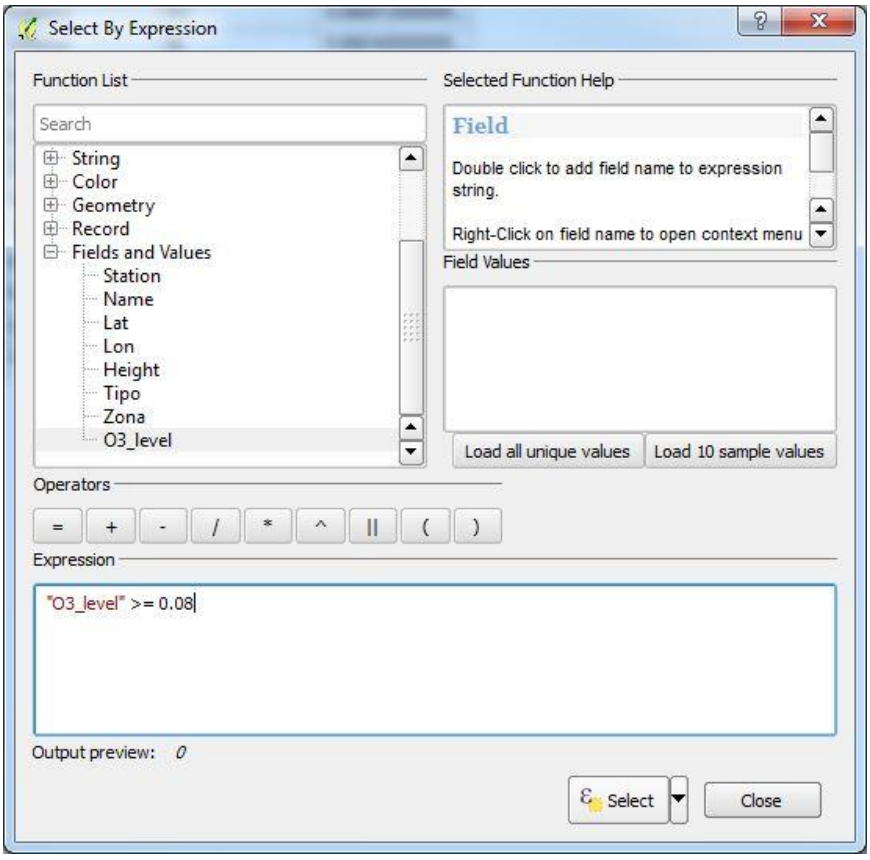
The field "03_*level*" is the 8 hour ozone concentration averaged for the month of April 2006. These values are in parts per million (PPM). The WHO recommends that countries not exceed 0.0473 PPM of ozone averaged over an 8 hour period. This level has been recommended after studies linking daily mortality and ozone levels exceeding 0.0473 PPM.

- In your TOC, Right-click on *airquality* layer and select **Open Attribute Table**.
- Click on the field *03_level*. You will now see it sorted Ascending.
- Click on the field *03_level* again to see the data sorted Descending.

Notice that all stations in this file exceed this 0.0473 level for the month of April.

Let's map where the highest levels of ozone concentration occur.

- In the Attribute table click the **Select features using an expression** button.



We are choosing to select stations which exceed 0.08 PPM over this 8 hour period because 0.08 PPM is almost double the WHO's acceptable level for ozone concentrations.

- Double-click on **03_*level*** from the Fields and Value.
- Type ***>= 0.08***
- Click **Select.**
- Click **Close.**

5 stations should be selected out of the 19 in the attribute table.

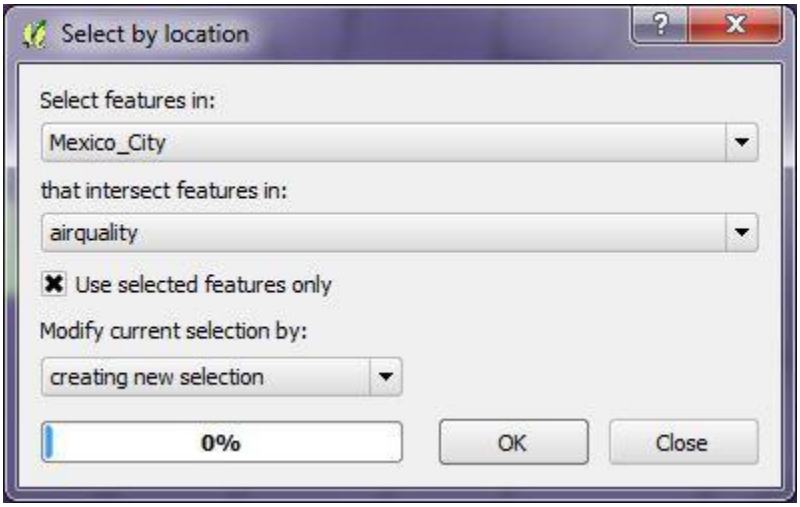
- Close the Attribute table. You should see these five stations highlighted as yellow in your map.

The highlighted yellow dots on your map show stations with an average ozone concentration of 0.08 PPM or greater (as measured in April, 2006). We will now analyze which municipalities have these highest levels of ozone by performing a selection based on location.

- Click on **Vector > Research Tools > Select By Location**.

In the **Select By Location** dialog box make sure the following parameters are set:

1. Select **Mexico_City** as the Select features in:
2. And make sure **airquality** is selected as the layer "that intersect features in:"
3. Make sure **Use selected features only** is checked.



The parameters you just set will select features (municipalities) from the *Mexico_City* layer that intersect the selected features (air quality stations that exceed 0.08ppm) from the *airquality* **layer**.

- Click **OK** to run the selection and then click **Close**.

You should now see 4 municipalities are selected on your map. Let's open the attribute table and summarize how many young children and how many people, in general, live in these municipalities.

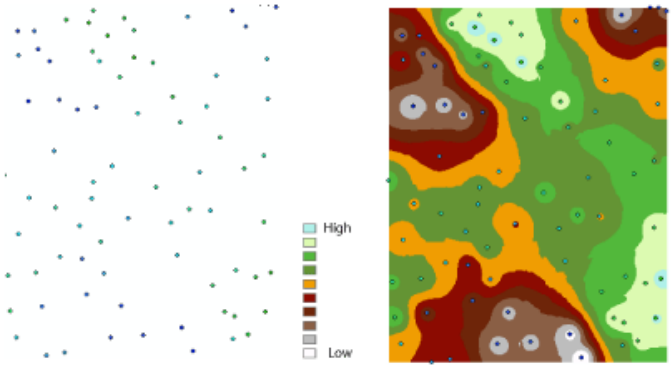
STEP 9 Interpolating air quality measurements

QGIS is able to perform data interpolation, however it requires an additional plugin.

- Click **Plugins > Manage Plugins** on the main toolbar.
- In the Plugin Manager click the **Get More** tab and type in **interpolation** in the Search box.
- Click the **Interpolation plugin**.
- Click **OK.**

To map the spatial distribution of Ozone concentrations, we will perform spatial interpolation of the Ozone measurements.

All interpolation techniques create a continuous surface by predicting values for cells in a raster layer using a limited number of sample points. Unknown values are predicted using a mathematical formula based on nearby sample points. **IDW** stands for the "**I**nverse **D**istance **W**eighted" interpolation technique. This technique uses the distance of sample points to each cell to weigh the importance of sample points on the cell being calculated. Closer points are weighted more heavily than distant points.



We will use the field "O3_level" as our interpolation variable (i.e., z-value), and interpolate the Ozone concentration across the city based on the air quality monitoring station data from nearby points. We are using vector (point) data to create a raster layer based on the value of the points.

- Click **Raster > Interpolation > Interpolation**. (The interpolation plugin added this function).
- In the Interpolation dialog which appears select **airquality** as the *Input Vector layer*.
- Select **03_level** as the *Interpolation attribute*.
- Click the **Add** button.

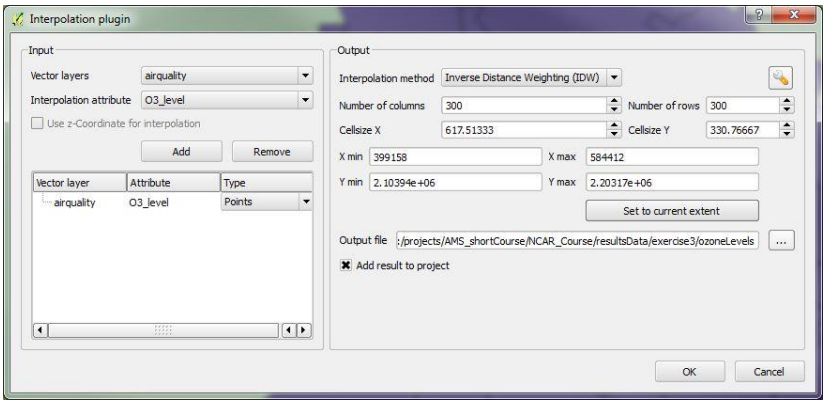
Note you can add multiple interpolation layers if desired and the interpolation will consider the data from all layers.

- Under the Output box, change the *Interpolation method* to **Inverse Distance Weighting (IDW).**
- Click the **tool** icon  to define the settings for IDW.

This will allow you to set the "Distance coefficient P". IDW is based on the assumption that the further the surface gets from the point value the less similar it becomes. Surface values are more influenced by near points and less influenced by points further away. The Distance coefficient P is used to specify the rate of influence as distance increases. A larger coefficient means it takes a larger distance for the values of the surface to become dissimilar from nearby points. Smaller numbers can produce "bull's eye" effect as the surface values change abruptly.

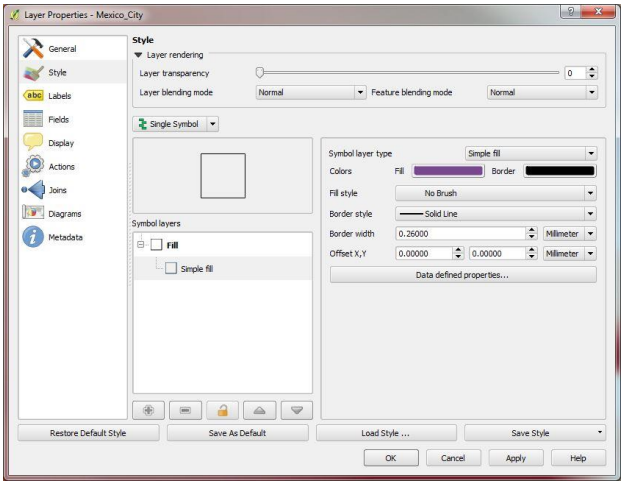
- Set this value to 3. (units are the same as the input data, in our case meters)
- Click **OK**.

- The Number of columns and Number of rows specifies how many rows and columns will be in our output raster. This parameter is linked to the X and Y cellsize. We will let the data default to 300 columns and rows.
- Click the **Set to current extent**.
- For Output file, save your raster to **<your working directory>\data\results\ozoneLevels**
- Click **OK**.
- If a dialog appears about the projection click **WGS84 / UTM zone 14N** and click **OK**. (See screenshot on the next page)



The output will appear as a single-band grey raster layer. You will need to symbolize it to visualize the results.

- Double-click on *ozoneLevels* in the Map Legend in order to open the Layer Properties.
- Click the **Style** tab.
- Change the *Render type* to **Singleband Pseudocolor**.
- Click **Classify**.
- Click **OK**.
- Double-click on *Mexico_city* in order to open the Layer Properties.
- In the **Style** tab set the symbol to **Single Symbol.**
- Click the Simple Fill.
- Set the *Fill style* to **No Brush**.
- Click **OK** twice.



- Reorder the layer so you can see the Mexico City municipalities, the airquality stations and the interpolated ozone surface.
- Click the **Deselect Features** button.
- **Save** your project.

STEP 10 Calculate Zonal Statistics for the air quality in each municipality

QGIS has a plugin for a tool called zonal statistics. This tool calculates several values (sum, mean, value, total count) for pixels by polygonal vector layer. We will use this tool to calculate a sum of population for each country which may be affected by temperature increase of 5 degrees or greater by the end of the century.

- Click **Plugins > Manage Plugins**.
- Click **Get more.**
- In the *Search* box type in **zonal**.

You should see Zonal Statistics plugin show up in the plugin list.

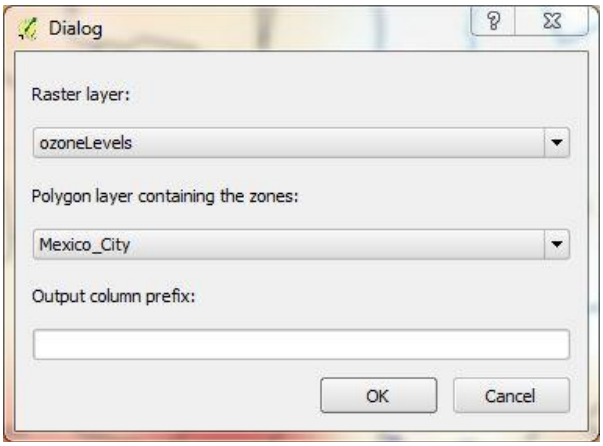
- Click **Zonal statistics plugin**.
- Click **Install plugin.**
- Once the plugin is installed click the Installed tab and make sure the Zonal Statistics plugin is checked on.
- Click **Close**.

Zonal statistics uses any zone (a polygon from a vector layer) to calculate the statistics of pixel values that are within the zone. In this case, we can use the Zonal Statistics function to calculate areal mean O_3 levels for each municipality in Mexico City based on interpolated ozone measurements.

- Click **Raster > Zonal Statistics > Zonal Statistics**.

In the Zonal Statistics dialog box,

- Set the *Raster layer***:** to **ozoneLevels.**
- Change the *Polygon layer containing the zones*: to **Mexico_City.**
- Set the *Output column prefix***:** to **ozone.**
- Click **OK**.



This tool will add statistics from the raster layer to the input Polygon layer.

- Open the Attribute table for *Mexico_City*.

The field mean is the mean ozone concentration for that municipality. Using the field sorting tools answer the question below.

What is the name of the municipality with the highest mean level of ozone?

What is the mean ozone concentration in this municipality?

How many children age 0 to 4 live in this municipality?

- Close the attribute table.
- **Save** your project in **<your working directory>/maps as MexicoCity.qgs** and close QGIS Desktop.

Summary: In this lesson we worked with local level socio-economic, demographic and air quality data. We created new data through joining attribute (non-spatial) data to spatial data, and by

calculating fields based on existing fields. We also worked with X, Y text file information in order to derive point locations and ultimately interpolate those points to a surface. These GIS analytical processes are very powerful in answering questions about our spatial relationships.