

Tables and Attribute Data

Learning Objective

To become familiar with the relational data model, and its implementation in a GIS software package, including table join, data export, sorting, summarizing, creating new fields, and calculating values for fields based on table and geometric calculations. As a demonstration, you will analyze spatial relationships among data on census tracts, neighborhoods, crime, and businesses in San Francisco, California and Chicago, Illinois.

TUTORIAL

Acquiring the Data

You are provided the data for this lab in the Lab_4_tutorial.zip file, which includes the following files:

- SF_Tracts is a polygon shapefile representing the US Census Bureau tracts in San Francisco. A tract is statistical reporting unit used by the US Census Bureau to publish population data (i.e. it does not necessarily correspond to political or administrative units like cities or wards). The neighborhood assignments in San Francisco originally comes from DataSF, San Francisco's open data portal. The data has been manipulated for the purpose of this lab.
- Total_Population is a table in dBase format that contains the total population of each tract.
- SF_Dissolved is a polygon shapefile representing the neighborhoods in San Francisco. Neighborhoods are defined as one or more adjacent tracts that together compose a named neighborhood.

Move all files to your workspace folder, and unzip any zip files.

GIS and the Relational Data Model

An attribute table in ArcGIS is composed of records (rows) and fields (columns):

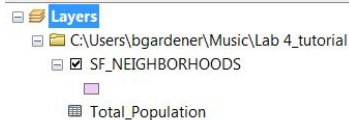
Attribute

Record →

	OID	TRACTCE10	GEOID_1	NHOOD	Total_Pop
▶	0	16400	6075016400	Hayes Valley	4135
	1	16100	6075016100	Western Addition	5085
	2	15900	6075015900	Western Addition	4081
	3	15500	6075015500	Japantown	3918
	4	15300	6075015300	Pacific Heights	2066
	5	15100	6075015100	Western Addition	2106
	6	13400	6075013400	Pacific Heights	3968
	7	13200	6075013200	Pacific Heights	4695
	8	12700	6075012700	Marina	3758
	9	12100	6075012100	Nob Hill	3876
	10	12000	6075012000	Nob Hill	3563
	11	11300	6075011300	Chinatown	3058
	12	11200	6075011200	Nob Hill	3078
	13	10900	6075010900	Russian Hill	4740
	14	10700	6075010700	Chinatown	5311

Explore the SF_tracts and the TOTAL_POPULATION data sets in ArcCatalog.

Open ArcMap and add both the SF_Tracts layer and the Total_Population table. In the Table of Contents (TOC), notice that Total_Population has a grid icon indicating it is tabular data (with no spatial information) and the SF_Tracts layer has an icon indicating it is a polygon spatial data layer.



Right-click on each file in the TOC. You will see that you have different menu options for tables versus spatial data layers.

Understanding How Tables Are Organized

1. In the TOC, right-click on Total_Population
2. Select Open
 - a. Draw your attention to the **Total_Pop** field. This field stores the total population of each tract according to the 2010 Census.
 - b. Notice also the **NHOOD** field. This field stores the name of the neighborhood within which each tract falls, i.e. each neighborhood wholly contains one or more census tracts.
- c. At the bottom of the table it should indicate that 0 out of 195 records (each record is a census tract) are currently selected.



Right-click on the Total_Pop field name. You should see something like this:

Total_Population				
	OID	TRACTCE10	GEOID_1	NHOOD
	0	16400	6075016400	Hayes Valley
	1	16100	6075016100	Western Addition
	2	15900	6075015900	Western Addition
	3	15500	6075015500	Japantown
	4	15300	6075015300	Pacific Heights
	5	15100	6075015100	Western Addition
	6	13400	6075013400	Pacific Heights
	7	13200	6075013200	Pacific Heights
	8	12700	6075012700	Marina
	9	12100	6075012100	Nob Hill
	10	12000	6075012000	Nob Hill
	11	11300	6075011300	Chinatown
	12	11200	6075011200	Nob Hill
	13	10900	6075010900	Russian Hill
	14	10700	6075010700	Chinatown
	15	980600	6075980600	Bayview Hunters Point
	16	12502	6075012502	Tenderloin
				4120

The various options you see are tools you can use to manipulate the data contained in the field you’ve selected in various ways. For now, explore the two sorting tools on your own; they’re self-explanatory. Below, you’ll learn how the Field Calculator, Summarize and Calculate Geometry tools work.

Close the table.

Joining a Table to a Spatial Data Layer

Conceptually, a join operation can be understood as adding attributes (columns) from one table to another based on a common field, such that the output of a join displays a single table with fields from both tables. For example, if you have a table with the unemployment rate by state, and another table with the poverty rate by state, you would join them to get one table that includes both unemployment rate and poverty rate by state.

In ArcGIS, the tables that we want to join are referred to as the **target table** and the **join table**. The target table is the table that you want new attributes to be added to, and the join table is the table that is “lending” new attribute fields. In the GIS world, the target table is almost always a spatial data layer, and the join table is often a nonspatial table.

To perform a join, a field in each of the tables must be identified that contains matching values. These **key** fields allow the software to properly link the records from one table with the analogous records from the other table. In the target table, the key field must be a **primary key** or **candidate key** (field that *could* be used as a primary key). This is a unique field which *identifies* the rows (features) in the table. For example, a table of US states might have the state name, two-letter state postal codes, and population. No US state has exactly the same population, so all of the fields are unique. But you wouldn’t use the population to *identify* the state. (That is, you wouldn’t say “I’m going to

37,253,956 this week”, you would say “I’m going to California.”) You could use either the state name or the postal code, both of which are unique, as the primary key field.

In the join table, the field must be a **foreign key**, which means a field with values matching a candidate key in the target table.

Here, we will join the Total_Population table to the SF_Tracts spatial data layer attribute table. Since the Total_Population is a nonspatial table, this will allow us to (a) map the total population of each tract, and (b) generate a population density variable by dividing the total population by the area of each tract, which we can calculate from the SF_Tracts layer.

First, let’s explore our data. Open the SF_Tracts attribute table. Find the **GEOID2** field. This field is a unique identifier used by the U.S. Census Bureau for each tract. No two tracts have the same **GEOID2** value.

Each field has a specific data type, which indicates the types of values it can hold. Common data types include text, which holds letters and/or numbers as text, and numeric data types. Numeric data types include integers (whole numbers) as well as so-called “floating point” types such as ‘double,’ and ‘float’.

Right-click on the name of the **GEOID2** field in the table and click on Properties. Note the data type, which is double (a numeric type).

Now open the Total_Population table. Find the **GEOID_1** field. This field is also a unique identifier for each tract. Even though the name of the field differs from the **GEOID2** field in the SF_Tracts table, it contains the same information—a specific tract will have the same value in each table. These two fields will facilitate the join operation.

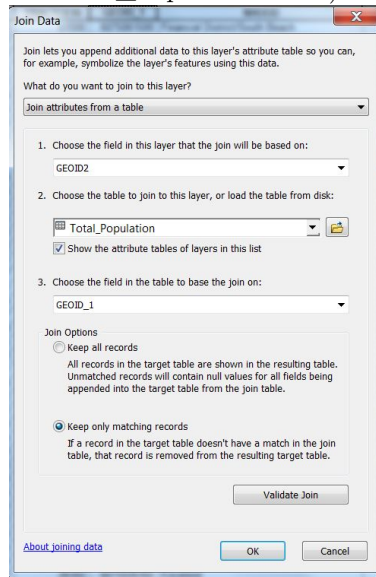
Close both tables.

Keep in mind the following information about joins:

1. Always begin a join operation with the target table (by right-clicking on the target table in the TOC).
2. The key fields in both the target and join table must contain similar information.
 - a. They should both refer to the same entity.
 - b. They should both identify each entity using a common value.
 - c. They should be the same data type. Usually this means text or integer.
 - d. They do not have to have the same field name (though they can). Also, just because they have the same field name doesn’t mean they have matching values. In one table a **state** field might have state *names*, while in another table a **state** field might have state *postal codes*. You *must* become familiar with the data you are working with.
 - e. The **OID** or **FID** fields should not generally be used as a join field (even though internally they are used as the primary key by ArcGIS).

To execute the join operation, in the TOC, right-click on SF_Tracts (the target or destination), and select Joins and Relates→Join. Use the following settings:

1. In dropdown 1, choose the field GEOID2 (the key identifying each tract in the SF_Tracts table)
2. In dropdown 2, choose the Total_Population table (the join table)
3. In dropdown 3, choose GEOID_1 (the foreign key identifying each tract in the Total_Population table).



Press OK. You might be asked if you want to index the join. Choose yes. On larger data sets the index helps speed processing.

Open up the attribute table in SF_Tracts and scroll to the right. You should see that the fields from the Total_Population table are now joined to the SF_Tracts attribute table.

SF_NEIGHBORHOODS						
GEOID	OID	TRACTCE10	GEOID_1 *	NHOOD	Total_Pop	
06075016400	0	16400	6075016400	Hayes Valley	4135	
06075016100	1	16100	6075016100	Western Addition	5085	
06075015900	2	15900	6075015900	Western Addition	4081	
06075015500	3	15500	6075015500	Japantown	3918	
06075015300	4	15300	6075015300	Pacific Heights	2066	
06075015100	5	15100	6075015100	Western Addition	2106	
06075013400	6	13400	6075013400	Pacific Heights	3968	
06075013200	7	13200	6075013200	Pacific Heights	4695	
06075012700	8	12700	6075012700	Marina	3758	
06075012100	9	12100	6075012100	Nob Hill	3876	
06075012000	10	12000	6075012000	Nob Hill	3563	
06075011300	11	11300	6075011300	Chinatown	3058	
06075011200	12	11200	6075011200	Nob Hill	3078	
06075010900	13	10900	6075010900	Russian Hill	4740	

Note that while the join is displayed as though the two tables have become one, the join is not permanent, i.e. the storage of the SF_Tracts attribute table has not changed.

To permanently store the SF_Tracts layer with the joined attributes, export the layer to a new shapefile. Right-click on SF_Tracts in the TOC and choose Data→Export Data. Save as a new shapefile and call it ‘Tract_Pop’. Add the new shapefile to ArcMap, open the attribute table, and scroll to the right. You will see that this shapefile includes the joined attributes – not as a join, but encoded as a part of the shapefile attribute table.

You can remove a join you no longer need by right-clicking SF_Tracts in the TOC, going to Joins and Relates and choosing Remove Join(s). Sometimes you will perform a join incorrectly—for example, by joining the tables in the wrong direction; or by selecting fields that don’t quite match. Since joins don’t modify the data on disk, there is no harm in creating an incorrect join. Just remove the undesired join and try again.

Summarizing and Joining

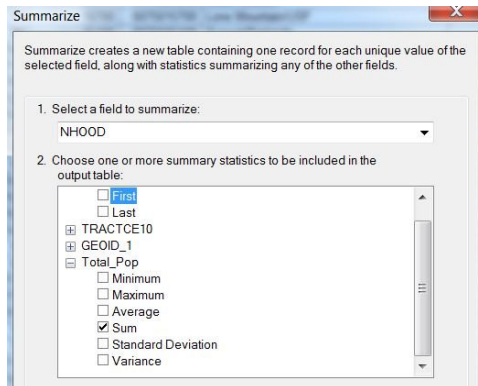
It is possible to summarize one field by another. This means calculating summary statistics (such as the sum or average) of a quantitative field by the unique values in a nominal (categorical) field.¹ The Total_Population table has population by tract. The field with unique values that we want to Summarize by is the neighborhoods. We will sum the total population for the set of tracts in each neighborhood, to yield the total population of each neighborhood.

The output table from the summarize operation will also show the number of records summarized for each category, which in this case is the number of tracts in each neighborhood.

To calculate the total population of each neighborhood:

1. Open the ‘Total_Population’ table.
2. Right-click on the NHOOD field and choose Summarize to open the Summarize dialog box. Recall that this field is a unique identifier of each neighborhood. This is the field we are summarizing on.
3. The ‘Select a field to summarize:’ box should have NHOOD entered
4. For ‘Choose one or more summary statistics to be included in the output table:’ expand the Total_Pop field by clicking on the plus sign. Click Sum. This indicates that we want the sum of the total population for all the tracts that are in the same neighborhood.

¹See Bolstad 5e, pp. 39-40 for a refresher on levels of measurement.



5. For 'Specify output table' choose a path to your workspace and the file name 'Neighborhood_Pop'. Make sure to set the Save as type to "dBase Table". ArcGIS will automatically add the .dbf extension.

Add the new table to ArcMap and open it.

Table				
Neighborhood_Pop				
	OID	NHOOD	Count_NHOOD	Sum_Total_Pop
▶	0	Bayview Hunters Point	11	37537
	1	Bernal Heights	6	25840
	2	Castro/Upper Market	6	20263
	3	Chinatown	4	14597
	4	Excelsior	8	39662
	5	Financial District/South Beach	3	16544
	6	Glen Park	2	8317
	7	Golden Gate Park	1	45
	8	Haight Ashbury	4	17916
	9	Hayes Valley	5	17773
	10	Inner Richmond	4	21340
	11	Inner Sunset	6	28021
	12	Japantown	1	3918
	13	Lakeshore	4	13223
	14	Lincoln Park	1	299

Notice there are 41 records—one record for each neighborhood.

The **Sum_Total_Pop** field contains the total population of each neighborhood.

Notice there is a **Count_NHOOD** field that indicates the number of records in the **Total_Population** table that were summed for each record in the **Neighborhood_Pop** table, i.e. the number of tracts in each neighborhood. For example, there are 11 Census tracts in the Bayview Hunters Point neighborhood in the **Total_Population** table. The **Sum_Total_Pop** field contains the sum of the populations of those 11 Census tracts. This count field is created automatically by the summarize operation, and will be useful in the lab assignment.

We will use this new table to map the population density of each neighborhood.

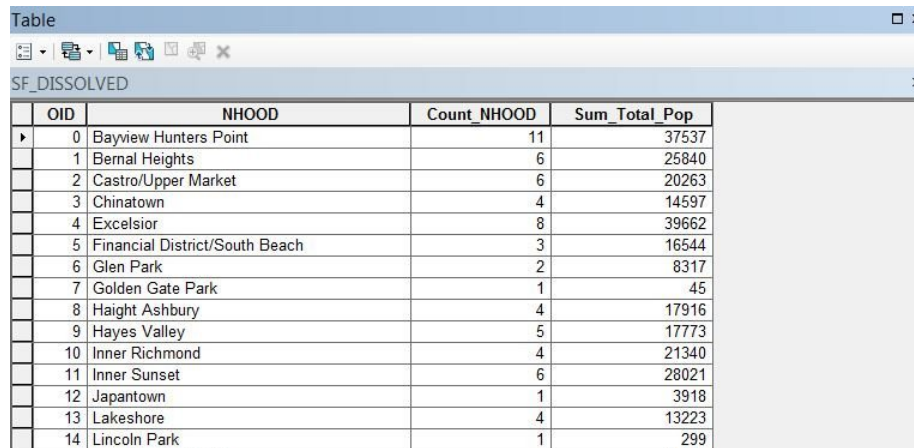
For this, we need a neighborhood spatial data layer.

Add the 'SF_Dissolved' data layer to ArcMap. Each polygon in this layer is a **neighborhood**—an aggregation of one or more adjacent tracts that compose a neighborhood.

We can join the Neighborhood_Pop table to the SF_Dissolved layer's attribute table using NHOOD as the join field in both tables, because they both contain common values that identify each neighborhood.

1. Right-click on SF_Dissolved. Select Joins and Relates→Join
 - a. For number 1 choose NHOOD.
 - b. For number 2 choose Neighborhood_Pop.
 - c. For number 3 choose NHOOD.
2. Click OK.

Check to see if your join was successful by opening the SF_Dissolved attribute table and seeing if the Sum_Total_Pop field is there and the population data are displayed.




	OID	NHOOD	Count_NHOOD	Sum_Total_Pop
▶	0	Bayview Hunters Point	11	37537
	1	Bernal Heights	6	25840
	2	Castro/Upper Market	6	20263
	3	Chinatown	4	14597
	4	Excelsior	8	39662
	5	Financial District/South Beach	3	16544
	6	Glen Park	2	8317
	7	Golden Gate Park	1	45
	8	Haight Ashbury	4	17916
	9	Hayes Valley	5	17773
	10	Inner Richmond	4	21340
	11	Inner Sunset	6	28021
	12	Japantown	1	3918
	13	Lakeshore	4	13223
	14	Lincoln Park	1	299

To preserve the join permanently, export the SF_Dissolved layer to its own layer and call the new layer SF_Dissolved_POP. Add it to ArcMap and open the attribute table to ensure the join and export worked properly. Note that when you do the export the field names in the new data layer will be truncated to 11 characters, so Sum_Total_Pop will be unhelpfully renamed to Sum_Total_.

Calculating Population Density

To calculate population density, we need to find the area of each neighborhood. We will calculate the area in square kilometers, so that we can ultimately calculate the population density as people per square kilometer.

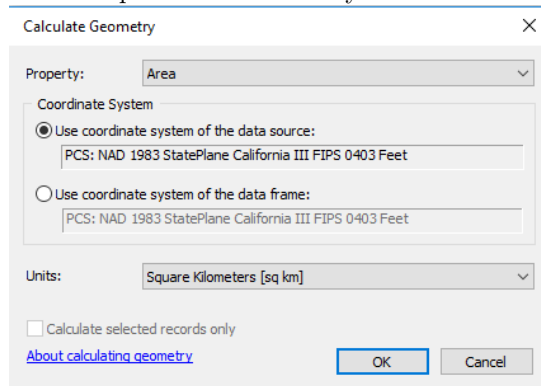
First, we need to create a new field to hold the area value:

1. Open the attribute table for SF_Dissolved_POP.
2. In the table, click on the Table Options button  in the upper left hand corner of the table and select 'Add Field'.
3. Name the field **Area_km** and set the data type to double.

Your attribute table should show your new field as the last entry on the right side.

To calculate the area of each neighborhood:

1. Right-click on your new **Area_km** field at the top of the column and select 'Calculate Geometry'. (ArcMap asks you if you would like to make changes outside of an edit session—please do so).
2. Choose square kilometers as your unit of measurement.



The 'Calculate Geometry' dialog box is shown. The 'Property' dropdown is set to 'Area'. Under 'Coordinate System', the radio button 'Use coordinate system of the data source:' is selected, with the text 'PCS: NAD 1983 StatePlane California III FIPS 0403 Feet' displayed below it. The 'Units' dropdown is set to 'Square Kilometers [sq km]'. There is an unchecked checkbox for 'Calculate selected records only' and a link for 'About calculating geometry'. 'OK' and 'Cancel' buttons are at the bottom right.

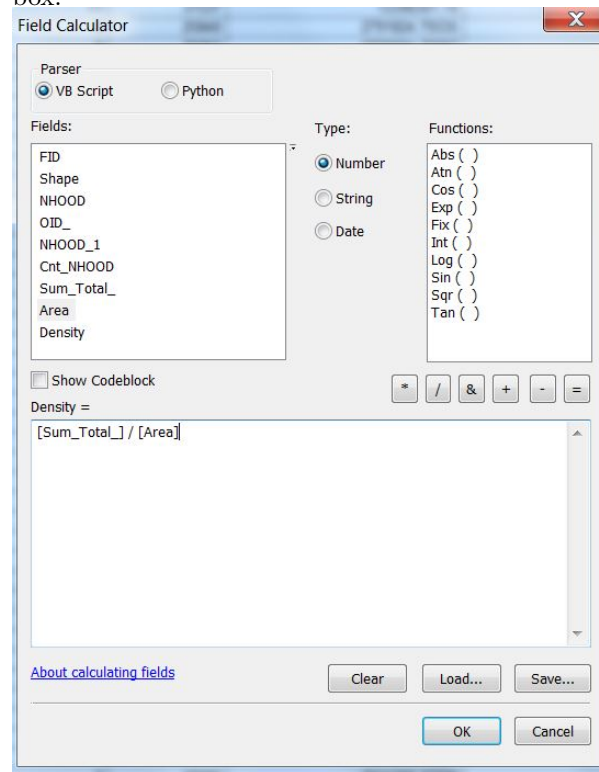
Click OK. You should see that the **Area_km** column has been updated with the calculated values:

Table				
SF_Dissolved_POP				
	NHOOD_1	Cnt_NHOOD	Sum_Total	Area_km
	Bayview Hunters Point	11	37537	13.398302
	Bernal Heights	6	25840	2.791825
	Castro/Upper Market	6	20263	2.220095
	Chinatown	4	14597	0.581815
	Excelsior	8	39662	3.60573
	Financial District/South Beach	3	16544	2.910413
	Glen Park	2	8317	1.731252
	Golden Gate Park	1	45	4.466632
	Haight Ashbury	4	17916	1.441933
	Hayes Valley	5	17773	1.270114
	Inner Richmond	4	21340	1.92711
	Inner Sunset	6	28021	3.687284
	Japantown	1	3918	0.31258
	Lakeshore	4	13223	7.480958
	Lincoln Park	1	299	1.022408

Now that we have the area of each neighborhood encoded, we can calculate the

population density of each neighborhood:

1. Add a new field to the SF_Dissolved_Pop table, using the double data type, and call it **popden**.
2. Right-click on the new field **popden** at the top of the column and choose 'Field Calculator'.
3. Create an equation in the text box where the neighborhood population is divided by the neighborhood area:
 - a. Double-click field name that holds the population data: **Sum_Total_**
 - b. To the right there will be some operator buttons. Choose the division symbol (a forward slash).
 - c. Double-click the field name that holds the area data: **Area_km**.
 - d. The equation **[Sum_Total_] / [Area_km]** should appear in the text box.



Click OK. You should see that the **popden** field has been updated with the calculated values:

	Cnt_NHOOD	Sum_Total_	Area_km	popden
	11	37537	13.398302	2801.623714
	6	25840	2.791825	9255.595144
	6	20263	2.220095	9127.08785
	4	14597	0.581815	25088.745277
	8	39662	3.60573	10999.714513
	3	16544	2.910413	5684.416929
	2	8317	1.731252	4804.036951
	1	45	4.466632	10.074704
	4	17916	1.441933	12424.986269
	5	17773	1.270114	13993.232921
	4	21340	1.92711	11073.576825
	6	28021	3.687284	7599.359743
	1	3918	0.31258	12534.409546
	4	13223	7.460958	1772.292522
	1	299	1.022408	292.446949
	3	17175	1.500289	11447.790757
	7	24846	2.650129	9375.393148

You have now encoded the population density for each neighborhood in San Francisco.

Create a choropleth map of population density using the skills you learned previously.

ASSIGNMENT

Objective

You have been hired as a GIS crime analyst for the city of Chicago. Your boss is upset about how presidential candidates judge your whole city based on the murder rate. She is also interested in how homicides might impact tourism in the city.

You have two objectives.

1. To describe the spatial distribution of the density of homicides by police district in Chicago.
2. To compare the spatial patterns of homicides, Airbnb listings, and Airbnb prices in Chicago.

Deliverables

Turn in a report in the format described in the syllabus.

Be sure to include the following information:

1. A choropleth map that shows the spatial distribution of the density of homicides (per square mile) in Chicago by police district, displayed in an appropriate UTM coordinate system. **Note:** This requires you to change the CRS of the file you are given.
2. A choropleth map that shows the spatial distribution of the density of Airbnb reviews (per square mile) in Chicago by police district, displayed in an appropriate UTM coordinate system.
3. A table that summarizes the average Airbnb price for each police district in Chicago.

The **Introduction** section should state the research objective.

The **Data and Methods** section should state the data sets used in the analysis (only the data used for the assignment), from where those data were acquired, and the GIS operations employed to investigate the research objective.

The **Results** section should state the results (i.e. the spatial distribution of the density of homicides by police district and the spatial patterns of homicides, Airbnb listings, and Airbnb prices). The maps and table should be cited in the text here (e.g. Figure 1, Figure 2, Table 1).

The **Discussion** section should state an interpretation of the results. Consider:

1. Where are the homicides in Chicago concentrated?
2. Where are the Airbnb reviews in Chicago concentrated?
3. Which police districts have the highest average Airbnb prices? Which have the lowest?
4. Is there a relationship, or pattern, between homicides and Airbnb reviews and/or prices?
5. If so, what is the form of this relationship and why might it occur?
6. The limitations of the analysis, and how the analysis could be improved or expanded.

The **Tables and Figures** section should contain the table and maps noted above, each on a separate page with a caption. The table and maps should be cited in the text.

Data

Several data sets are provided to you in the Lab_4_assignment.zip file.

- Police_Districts_Chicago is a polygon shapefile of the police districts in Chicago. The DISTRICT field is a unique number used by the police

department that identifies each police district.

- `Selected_Crimes_in_Chicago` is a point shapefile of the crime incidents in Chicago. The `PRIMARY_DE` field indicates the type of crime (i.e. homicides versus other types of crime). The `DISTRICT` field is a unique number used by the police department that identifies each police district.
- `AirBnB_Data` is a point shapefile of the Airbnb listings in Chicago. The `price` field contains the nightly price for the listing. The `number_of_` field contains the number of reviews for that listing.

It is not uncommon for data of unknown location to be geocoded oddly. For example, many of the points in the crimes layer appear at Null Island. Most likely this means that when these crimes were reported, the exact location was unknown, and the database applied a default value of 0° Latitude, 0° Longitude. Three of the Airbnb properties are at the latitude of Chicago, but for some reason their longitude is 0° (the Prime Meridian).

For this assignment you can ignore these misplaced features. Use **Zoom to Layer** on the police districts layer to zoom the map canvas to our area of interest. Optionally, you could extract only the features you are interested in by selecting them in the map canvas, then exporting the layers to new shapefiles.

Getting Started

You will need to use several GIS operations you have learned from this lab and previous labs:

1. Use operations in projections and coordinate systems to transform your data to UTM (research which UTM zone Chicago is in).
2. Use operations in selection and data export to create a spatial data layer of only homicides (not including other crime types).
3. Use the Summarize operation to calculate the number of homicides and Airbnb reviews, and average price, for each police district.
4. Use the Field Calculator to calculate the density of homicides and density of Airbnb reviews for each police district.
5. Use your thematic mapping skills to create the choropleth maps.