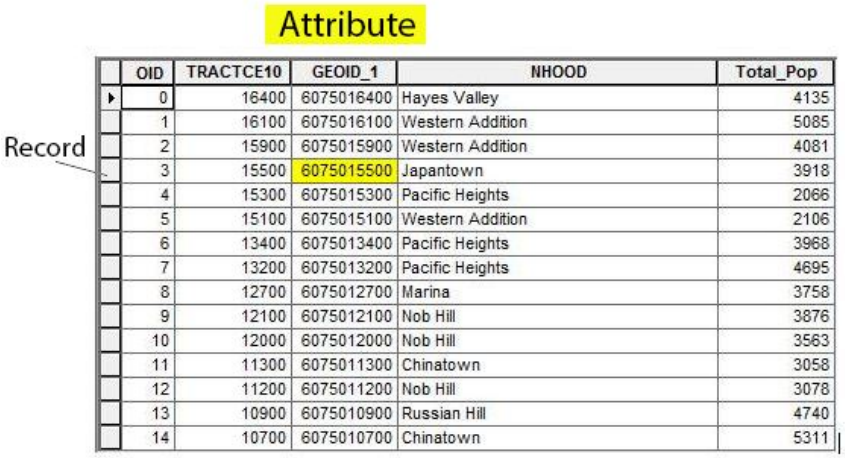
Tables and Attribute Data

Bradley Gardener

Purpose: The purpose of this lab is to become familiar with the relational data model, and its implementation in a GIS software package. Topics covered include table join, data export, sorting, summarizing, creating new fields, and calculating values for fields based on table and geometric calculations. This lab uses city level spatial data on census tracts, neighborhoods, crime, and businesses.

GIS and The Relational Data Model

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



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

Your instructor will provide the data for this lab, which includes the 'SF_Tracts' spatial data layer and the 'TOTAL_POPULATION' table.

The SF_Tracts spatial data layer is a shapefile representing the US Census Bureau tracts in San Francisco. A tract is statistical reporting unit (i.e., it does not correspond to political or administrative units like cities or wards) used by the US Census Bureau to publish population data. 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.

The Total_Population table is a dBase format table that contains the total population of each tract.

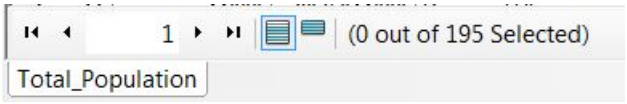
Add both the SF_Tracts layer and the Total_Population table to ArcMap and examine them. In the Table of Contents (TOC), notice how Total_Population has a grid icon indicating it is tabular data (with no spatial information) and the SF_Tracts layer has an icon indicating it is a polygon spatial data layer.

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.
   b. Notice also the `NHOOD` field. This field stores the name of the neighborhood within which that tract falls, i.e. each neighborhood contains one or more census tracts.

c. At the bottom of the table it should indicate that 0 out of 195 records, or tracts, are currently selected.



Right-click on the total population field. You should see options such as Field Calculator and Summarize. We will use those later. You will also see Calculate Geometry, but it is grayed out. We will use that tool later to calculate the area of each enumeration unit.

For now right-click `Total_Pop` and select 'Sort Descending'. The order of records changes so that they are listed from high to low population. Choose 'Sort Ascending' to list the records in order from low to high population.

Close the table.

Joining a Table to a Spatial Data Layer

Conceptually, a join operation can be understood as adding attributes 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 displaying 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).

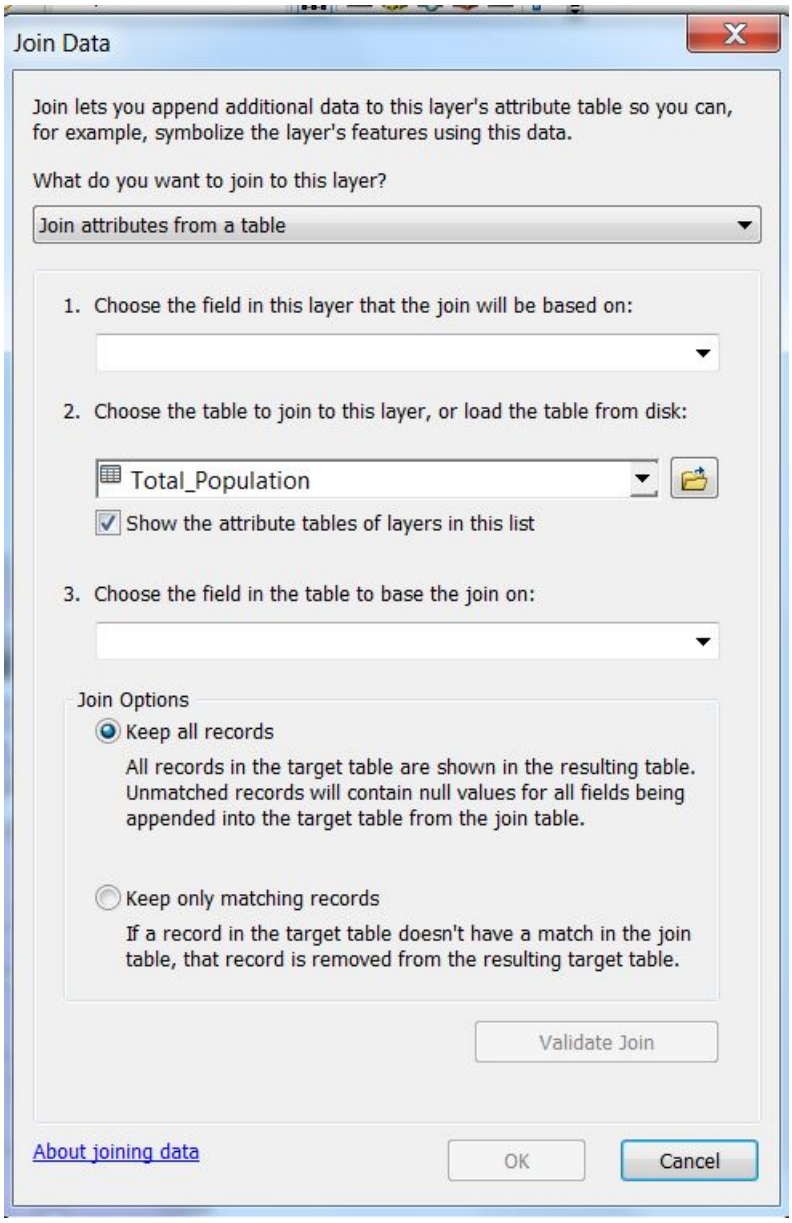
Now open the Total_Population table. Find the `GEOID_1` field. This field is also a unique identifier for each tract. Even though the name of the field differs from the `GEOID2` field in the SF_Tracts table, it contains the same information—a specific tract will have the same value in each table. These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

Close both tables.

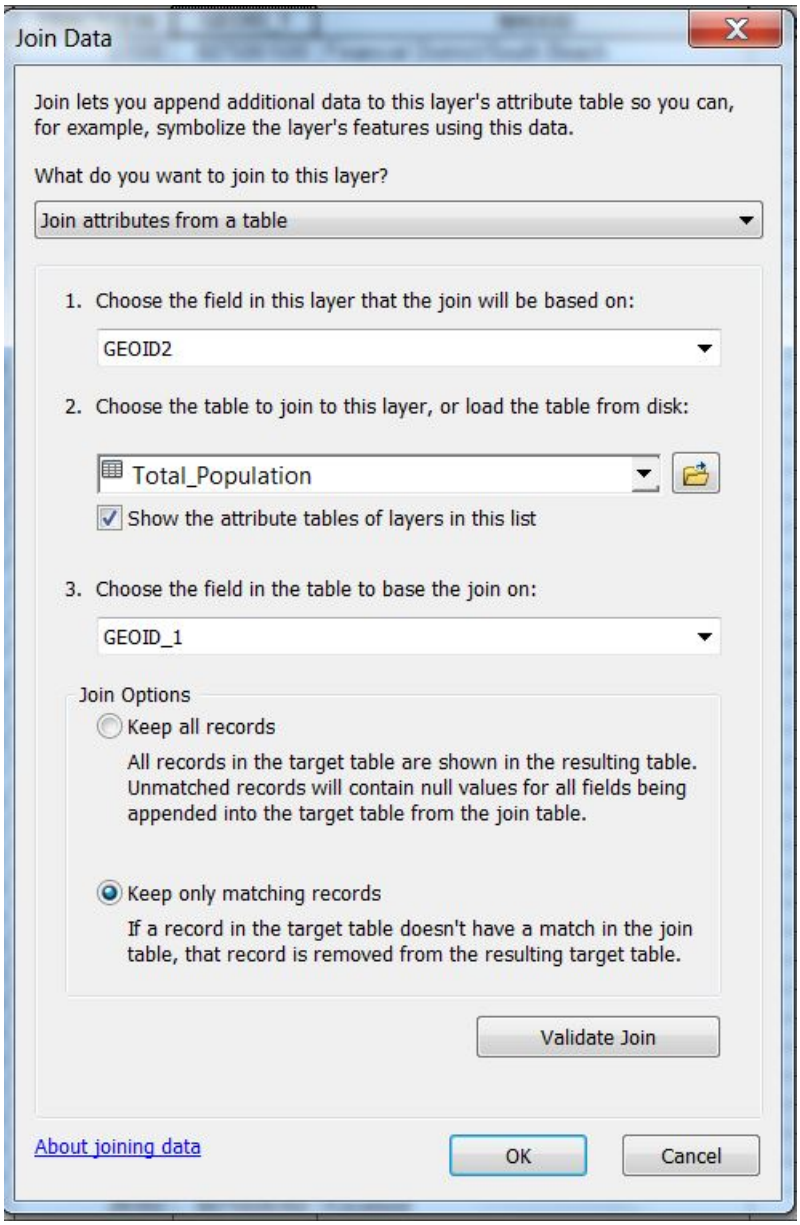
Keep in mind the following information about joins:

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. *It is rarely a good idea to use floating point data (double or float) for exact matching.*
   d. They do not have to have same field name in each table (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 destination), and select Joins and Relates→Join. You should see the following dialog box:



1. In dropdown 1, choose the field `GEOID2` (the key identifying each tract in the SF_TRACT 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.

If the join did not work you can remove the join by right-clicking SF_TRACT in the TOC, going to Joins and Relates and choosing Remove Join(s) and try the join again.

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
	06075010700	14	10700	6075010700	Chinatown	5311
	06075980600	15	980600	6075980600	Bayview Hunters Point	401
	06075012502	16	12502	6075012502	Tenderloin	4120
	06075012302	17	12302	6075012302	Tenderloin	2518
	06075012301	18	12301	6075012301	Tenderloin	1790
	06075980900	19	980900	6075980900	Bayview Hunters Point	246
	06075010500	20	10500	6075010500	Financial District/South Beach	2606
	06075010200	21	10200	6075010200	Russian Hill	4220
	06075010100	22	10100	6075010100	North Beach	3827
	06075061500	23	61500	6075061500	Financial District/South Beach	12391
	06075061200	24	61200	6075061200	Bayview Hunters Point	4023
	06075047802	25	47802	6075047802	Outer Richmond	3952
	06075042601	26	42601	6075042601	Outer Richmond	4111
	06075032902	27	32902	6075032902	Sunset/Parkside	3995
	06075032901	28	32901	6075032901	Sunset/Parkside	5582
	06075032801	29	32801	6075032801	Sunset/Parkside	4111
	06075031301	30	31301	6075031301	Oceanview/Merced/Ingleside	4363
	06075031202	31	31202	6075031202	Oceanview/Merced/Ingleside	2982
	06075022704	32	22704	6075022704	Potrero Hill	3314
	06075017802	33	17802	6075017802	South of Market	4307
	06075017801	34	17801	6075017801	South of Market	3066

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 and choose Data→Export Data. Save as a new shapefile and call it 'Tract_Pop'.

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

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

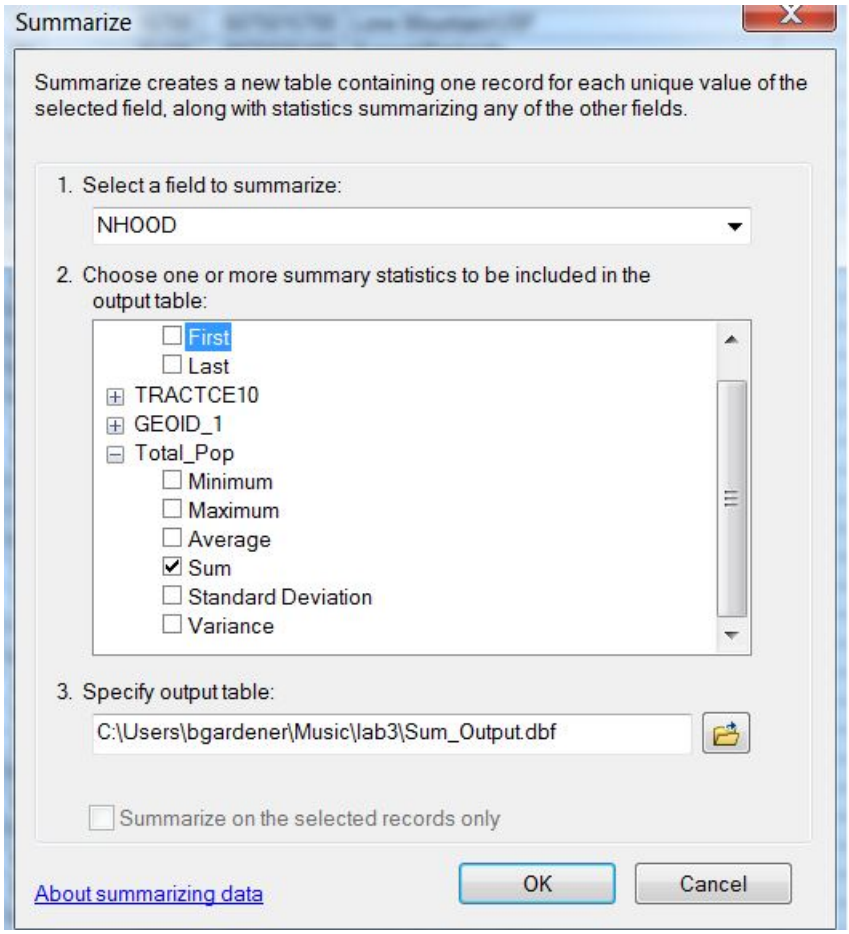
neighborhoods. We will sum the total population for the set of tracts in each neighborhood, to yield the total population of each neighborhood.

The output table 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.
5. For 'Specify output table' choose a path and the file name 'Neighborhood_Pop'.

Add the new table to ArcMap and open it.


Summarize
Summarize creates a new table containing one record for each unique value of the selected field, along with statistics summarizing any of the other fields.
1. Select a field to summarize:
NHOOD
2. Choose one or more summary statistics to be included in the output table:
First
Last
TRACTCE10
GEOID_1
Total_Pop
Minimum
Maximum
Average
Sum
Standard Deviation
Variance
3. Specify output table:
C:\Users\bgardener\Music\lab3\Sum_Output.dbf
Summarize on the selected records only
About summarizing data
OK
Cancel

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
15	Lone Mountain/USF	3	17175
16	Marina	7	24846
17	McLaren Park	1	850
18	Mission	13	56480
19	Mission Bay	1	9804
20	Nob Hill	7	25542
21	Noe Valley	6	22384
22	North Beach	3	12756
23	Oceanview/Merced/Ingleside	5	27393
24	Outer Mission	3	23610
25	Outer Richmond	10	44984
26	Pacific Heights	7	23669
27	Portola	4	15937
28	Potrero Hill	4	13102
29	Presidio	1	3451
30	Presidio Heights	2	10409
31	Russian Hill	4	18609
32	Seacliff	1	2357
33	South of Market	4	17825
34	Sunset/Parkside	14	79793
35	Tenderloin	8	27636
36	Treasure Island	1	3083
37	Twin Peaks	2	6953
38	Visitacion Valley	5	17450
39	West of Twin Peaks	7	37280
40	Western Addition	6	20399

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

The `Sum_Total_` field contains the total population of each neighborhood.

Notice there is a `Count_NHOOD` (or `Cnt_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

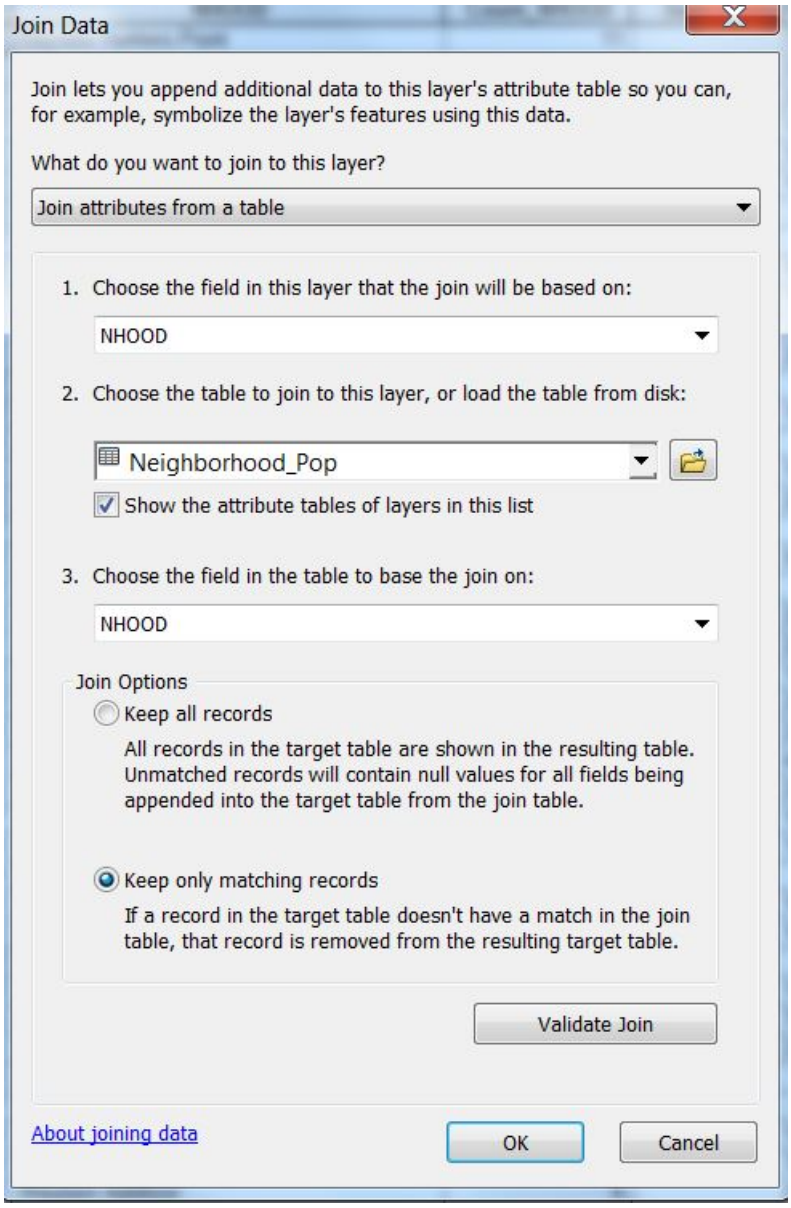
example, there are 11 Census tracts in the Bayview Hunters Point neighborhood in the Total_Population table. The `Sum_Total_` field contains the sum of the populations of those 11 Census tracts. This count field is created automatically by the summarize operation, and will be useful in the lab assignment.

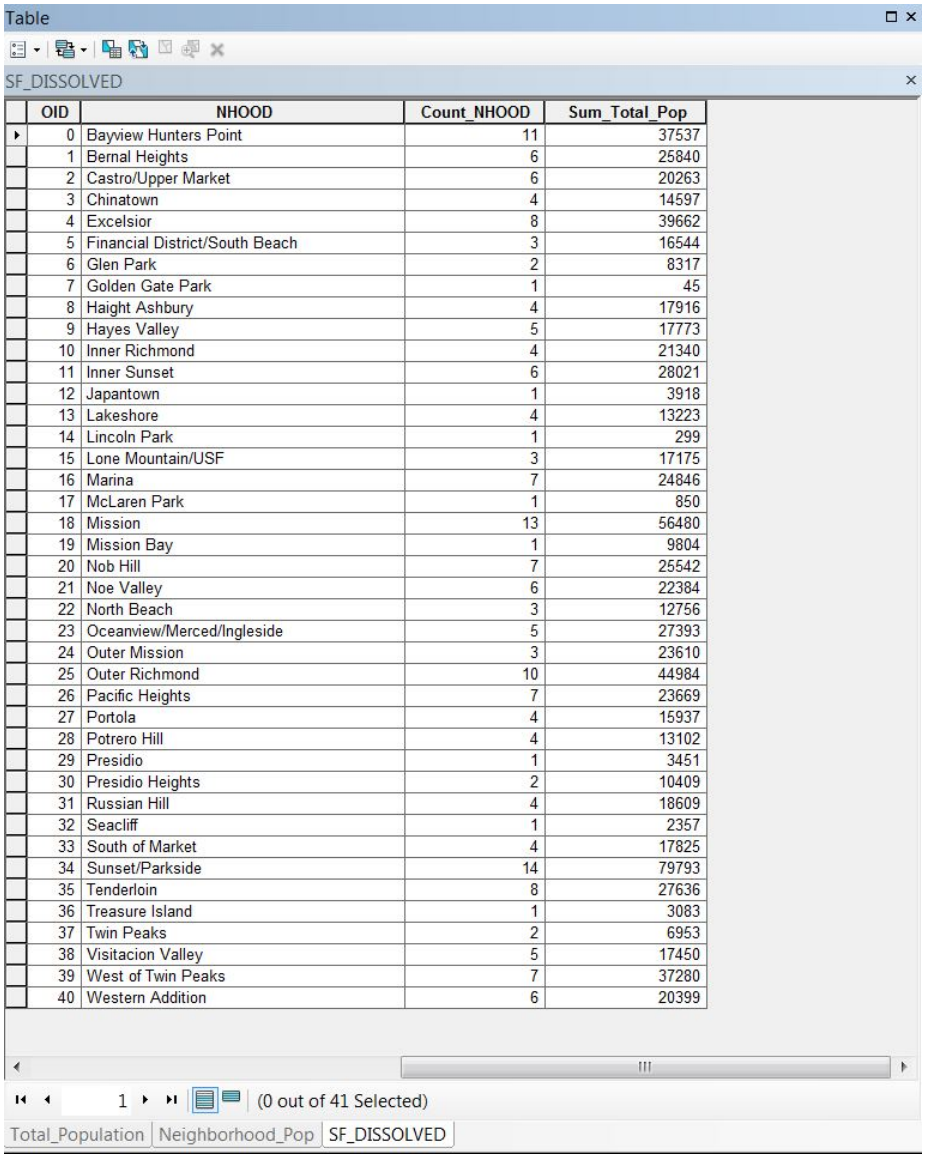
We will use this new table to map the total population of each neighborhood. For this, we need a neighborhood spatial data layer.

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(s)
    a. For number 1 choose `NHOOD`.
    b. For number 2 choose `Neighborhood_Pop`.
    c. For number 3 choose `NHOOD`.
2. Click OK.


Join Data
Join lets you append additional data to this layer's attribute table so you can, for example, symbolize the layer's features using this data.
What do you want to join to this layer?
Join attributes from a table
1. Choose the field in this layer that the join will be based on:
NHOOD
2. Choose the table to join to this layer, or load the table from disk:
Neighborhood_Pop
Show the attribute tables of layers in this list
3. Choose the field in the table to base the join on:
NHOOD
Join Options
Keep all records
All records in the target table are shown in the resulting table. Unmatched records will contain null values for all fields being appended into the target table from the join table.
Keep only matching records
If a record in the target table doesn't have a match in the join table, that record is removed from the resulting target table.
Validate Join
About joining data
OK
Cancel



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
15	Lone Mountain/USF	3	17175
16	Marina	7	24846
17	McLaren Park	1	850
18	Mission	13	56480
19	Mission Bay	1	9804
20	Nob Hill	7	25542
21	Noe Valley	6	22384
22	North Beach	3	12756
23	Oceanview/Merced/Ingleside	5	27393
24	Outer Mission	3	23610
25	Outer Richmond	10	44984
26	Pacific Heights	7	23669
27	Portola	4	15937
28	Potrero Hill	4	13102
29	Presidio	1	3451
30	Presidio Heights	2	10409
31	Russian Hill	4	18609
32	Seacliff	1	2357
33	South of Market	4	17825
34	Sunset/Parkside	14	79793
35	Tenderloin	8	27636
36	Treasure Island	1	3083
37	Twin Peaks	2	6953
38	Visitacion Valley	5	17450
39	West of Twin Peaks	7	37280
40	Western Addition	6	20399

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

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.

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 options 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 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.
3. Click OK.

Table

SF_DISSOLVED_POP

	NHOOD_1	Cnt_NHOOD	Sum_Total_	Area
	Bayview Hunters Point	11	37537	13398301.78
	Bernal Heights	6	25840	2791824.79335
	Castro/Upper Market	6	20263	2220094.76997
	Chinatown	4	14597	581814.667858
	Excelsior	8	39662	3605729.94438
	Financial District/South Beach	3	16544	2910412.83687
	Glen Park	2	8317	1731252.29581
	Golden Gate Park	1	45	4466632.45981
	Haight Ashbury	4	17916	1441933.18305
	Hayes Valley	5	17773	1270113.92581
	Inner Richmond	4	21340	1927109.94263
	Inner Sunset	6	28021	3687284.31722
	Japantown	1	3918	312579.542399
	Lakeshore	4	13223	7460957.96058
	Lincoln Park	1	299	1022407.65665
	Lone Mountain/USF	3	17175	1500289.4763
	Marina	7	24846	2650128.86473
	McLaren Park	1	850	1594620.87147
	Mission	13	56480	4878464.1677
	Mission Bay	1	9804	2105802.99399
	Nob Hill	7	25542	1053710.7777
	Noe Valley	6	22384	2528200.25416
	North Beach	3	12756	1293203.39603
	Oceanview/Merced/Ingleside	5	27393	2728023.95567
	Outer Mission	3	23610	2600765.43348
	Outer Richmond	10	44984	4637992.67786
	Pacific Heights	7	23669	2059618.44333
	Portola	4	15937	2138697.42033
	Potrero Hill	4	13102	2945530.9108
	Presidio	1	3451	6122684.39815
	Presidio Heights	2	10409	1300984.60583
	Russian Hill	4	18609	1279445.98786
	Seacliff	1	2357	551440.082286
	South of Market	4	17825	2292094.22197
	Sunset/Parkside	14	79793	10951681.1189
	Tenderloin	8	27636	1017366.38569
	Treasure Island	1	3083	2302171.06743
	Twin Peaks	2	6953	1716278.12029
	Visitacion Valley	5	17450	1584575.71545
	West of Twin Peaks	7	37280	7921298.32728
	Western Addition	6	20399	1511022.48616

0 (0 out of 41 Selected)

SF_DISSOLVED_POP

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

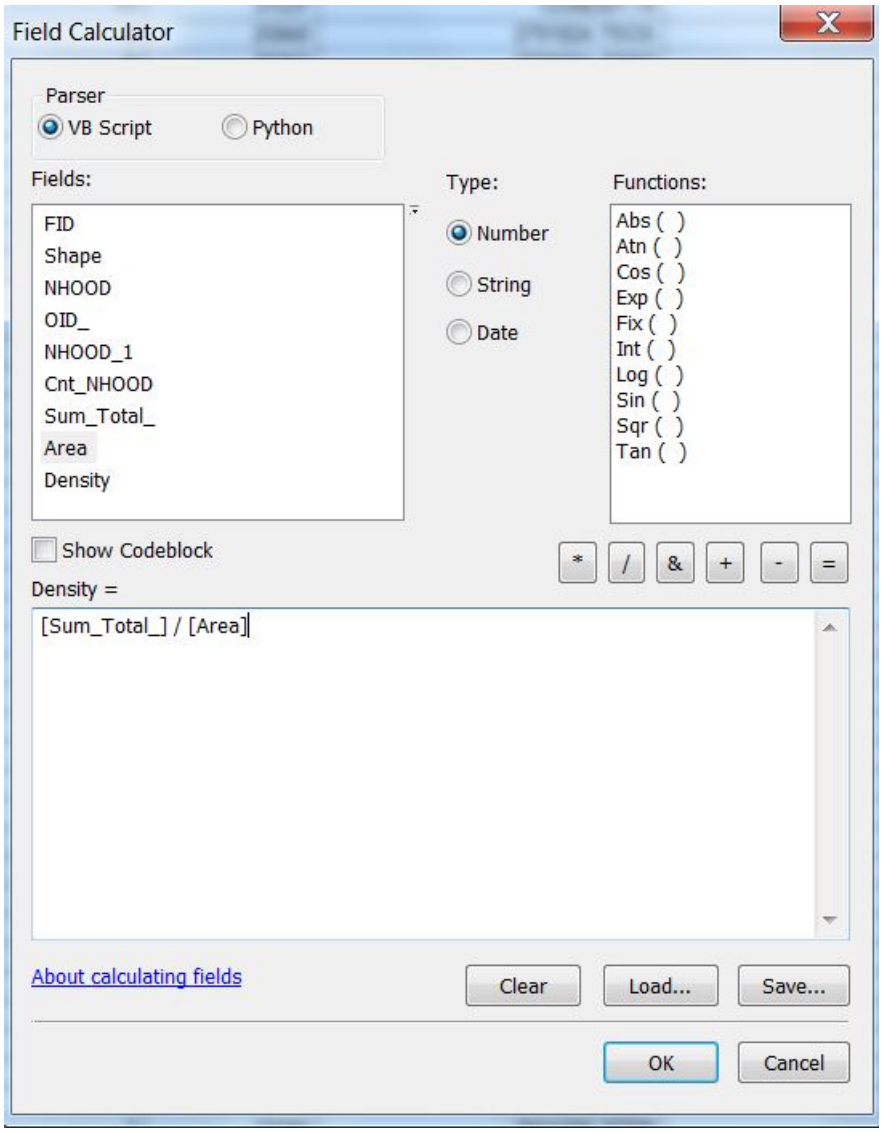
1. Add a new field and make it double and call it popden
2. Right-click on the new field popden and choose 'Field Calculator'
3. Create an equation in the text box where the neighborhood population is divided by the neighborhood area:
    a. Double click field name that holds the population data: `Sum_Total_`
    b. To the right there will be some operator buttons. Choose the division

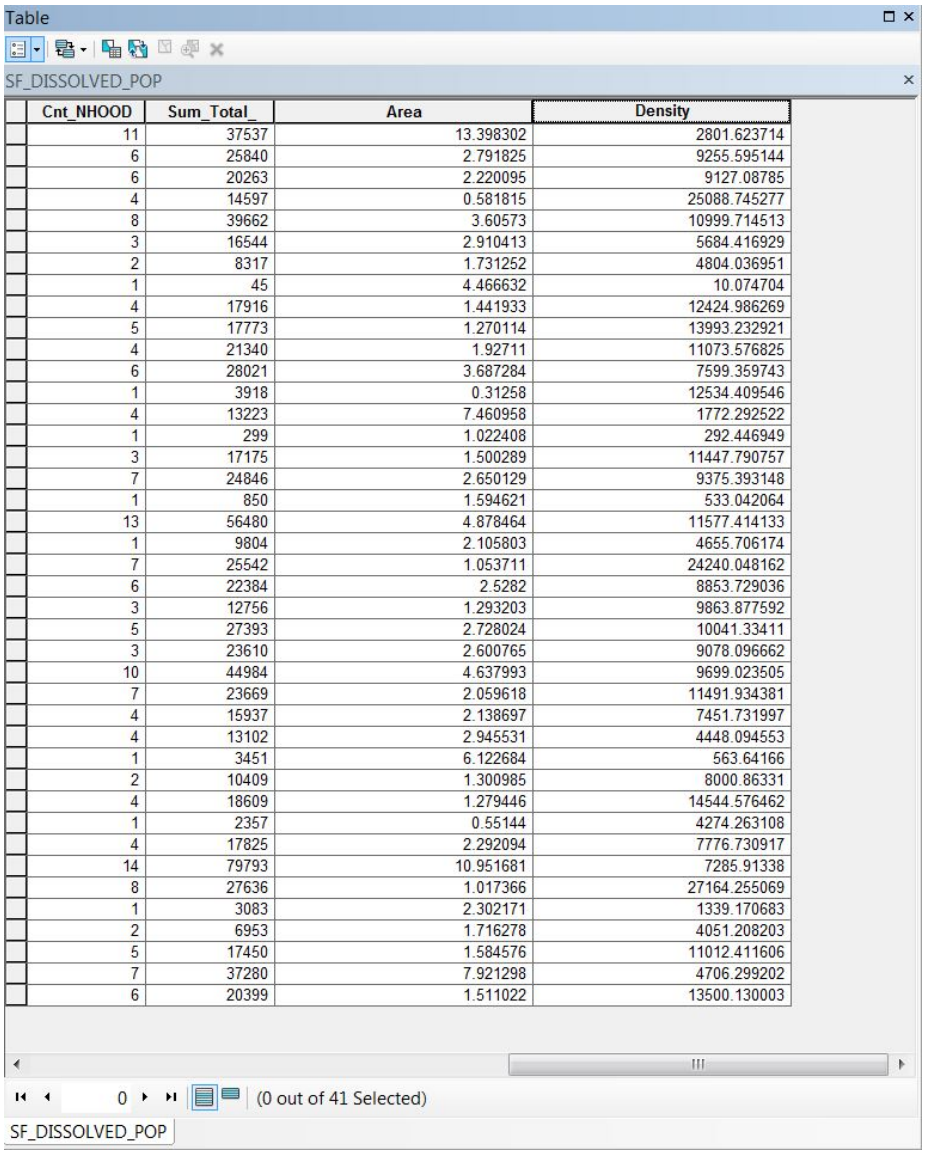
symbol (a forward slash).

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.

4. Click OK.





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

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

Assignment

Objectives

Consider that 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

Complete a lab report in the format detailed in the course syllabus. The report should include

1. A choropleth map that shows the spatial distribution of the density of homicides (per square mile) in Chicago by police district.
2. A choropleth map that shows the the spatial distribution of the density of Airbnb reviews (per square mile) in Chicago by police district.
3. A table that summarizes the average Airbnb price for each police district in Chicago.

Both maps should be displayed in an appropriate UTM coordinate system.

Data

Several data sets are provided to you.

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.

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 other including other crime types).
3. Use the Summarize operation to calculate the number of crimes 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.

In your report, be sure to include the following sections:

- **Introduction:** State the research question.
- **Data and Methods:** State the data layers and the analytical steps taken.
- **Results:** Report on the expected deliverables noted above.
- **Discussion:** Based on the maps and tables, 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?
- **Tables and Figures:** Insert all tables and figures (including maps) at the end of the report, each on a separate page, with a label (e.g. Figure 1). Be sure to cite each table and figure included in the body of the report text.