Tables and Attribute Data

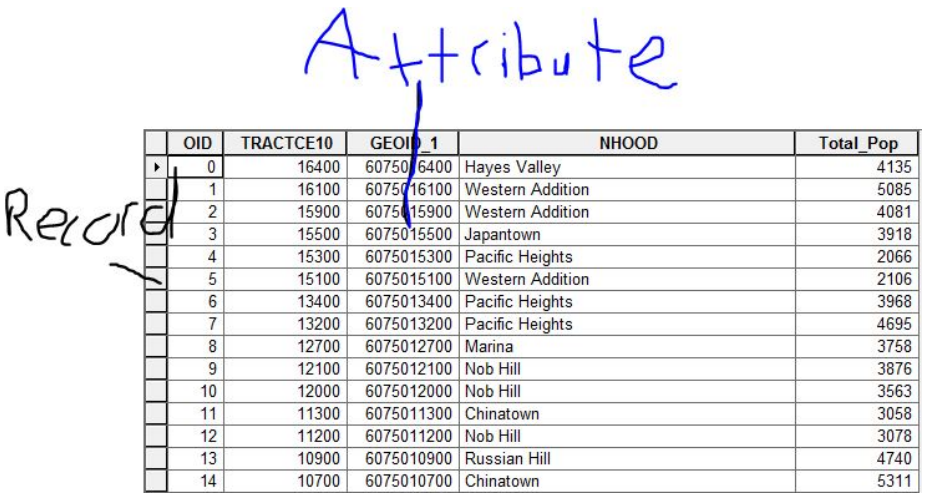
Bradley Gardener

Purpose: The purpose of this lab is to become familiar with the relational database model. It allows us to join tables based on attributes, a valuable way to integrate new data into a GIS. This is a skill that you will use frequently in your GIS practice. Students will also learn how to calculate the population density of enumeration units. To calculate area students will learn how to use the tool - calculate geometry. The field calculator will also be covered. In this example we will explore population density in San Francisco neighborhoods.

Acquiring Data and Understanding Relational Databases

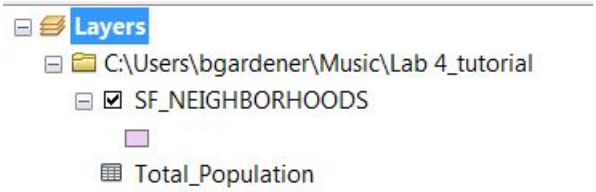
The data for neighborhoods in San Francisco originally comes from https://datasf.org/opendata/. The data has been manipulated for the purpose of this lab.

As we transition from GIS novices to GIS experts we will become intimate with data tables. This is what you see when you open an attribute table. We will manipulate them in a way that helps us analyze the specific spatial problems we want to solve. Tables are based on the relational database model. Please ask your professor about relational data tables or feel free to conduct your own research. Arc GIS makes sure we stay in what is called normal form. This way we cannot break the connection between a record and its attributes.



	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

Add 'SF_NEIGHBORHOODS' And 'TOTAL_POPULATION' from the S Drive to ARC Map and examine them. One is an attribute table (Total_Population) and the other is a shapefile (SF_NEIGHBORHOODS). Notice how the table has a grid next to it and SF_Neighborhoods appears like previous shapefiles.

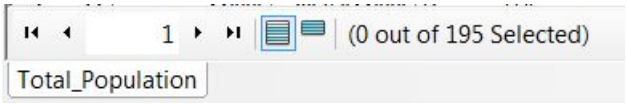


If you right click your mouse on both files you will see that you have different options. Understanding the difference between these types of files is very important.

Understanding how data are organized in the relational database model

1. Right click on Total_Population
2. Select Open
   a. Draw your attention to the column Total_Pop. Each one of the entries in the column is an attribute of a different Census Tract in San Francisco. Notice that each Census Tract is attached to a neighborhood.
3. Scroll down to the bottom of the table, until you've reach the last or 195th entry. You can see how many rows or enumerations units are on the table.

I am getting this number because I am looking towards the bottom center of the table window. You should see that 0 out of 195 cases are selected.



Notice that the last entry in the table has the `NHOOD` attribute Mission and the TRACTCE10 22803

Right click on Total Population (the column/field). You should see options such as Field Calculator and Summarize. We will use those later. You will also see Calculate Geometry grayed out. We will use that tool later to calculate the area of each enumeration unit. We cannot calculate area because we don't have the geographical information to do so in a table. To calculate area we will need to two things.

First we must have a `.shp` file. Second, we will need a coordinate system.

For now right click Total_Pop and select 'Sort Descending'. Now use 'Select by Attribute' on the `TRACTCE10` column and select `22803`.

OID	TRACTCE10	GEOID_1	NHOOD	Total_Pop
112	30102	6075030102	Inner Sunset	5229
171	21500	6075021500	Noe Valley	5129
186	25403	6075025403	Bernal Heights	5125
125	25701	6075025701	Portola	5116
154	26003	6075026003	Excelsior	5097
1	16100	6075016100	Western Addition	5085
43	11100	6075011100	Nob Hill	5078
142	45100	6075045100	Inner Richmond	5053
95	40200	6075040200	Inner Richmond	5045
164	22901	6075022901	Mission	5024
167	22801	6075022801	Mission	4988
182	16700	6075016700	Haight Ashbury	4960
194	22803	6075022803	Mission	4930
80	12800	6075012800	Marina	4908
191	23102	6075023102	Bayview Hunters Point	4870
193	23001	6075023001	Bayview Hunters Point	4835
124	25702	6075025702	Portola	4831
45	10800	6075010800	Nob Hill	4779
60	32601	6075032601	Sunset/Parkside	4766
71	16300	6075016300	Hayes Valley	4748
13	10900	6075010900	Russian Hill	4740
113	30101	6075030101	Inner Sunset	4728
85	20900	6075020900	Mission	4723
187	25300	6075025300	Bernal Heights	4718
7	13200	6075013200	Pacific Heights	4695
100	33100	6075033100	Sunset/Parkside	4677
155	26301	6075026301	Excelsior	4672
120	26004	6075026004	Excelsior	4658
62	12401	6075012401	Tenderloin	4613
175	20600	6075020600	Castro/Upper Market	4607
63	12201	6075012201	Tenderloin	4576
173	21000	6075021000	Mission	4556
146	35202	6075035202	Sunset/Parkside	4551
79	13000	6075013000	Marina	4548
59	32602	6075032602	Sunset/Parkside	4540
53	61100	6075061100	Chinatown	4488
78	13300	6075013300	Presidio Heights	4482
66	61000	6075061000	Bayview Hunters Point	4474
138	23003	6075023003	Bayview Hunters Point	4369
30	31301	6075031301	Oceanview/Merced/Ingleside	4363
50	10300	6075010300	Russian Hill	4346
96	40100	6075040100	Inner Richmond	4346
166	21600	6075021600	Noe Valley	4335
160	23200	6075023200	Bayview Hunters Point	4319

Notice how the location of the row has changed, but most importantly – what is attached to it – its attributes - are intact. They haven't changed position. This is important and what fundamentally allows us to complete our next step – joining our attribute table.

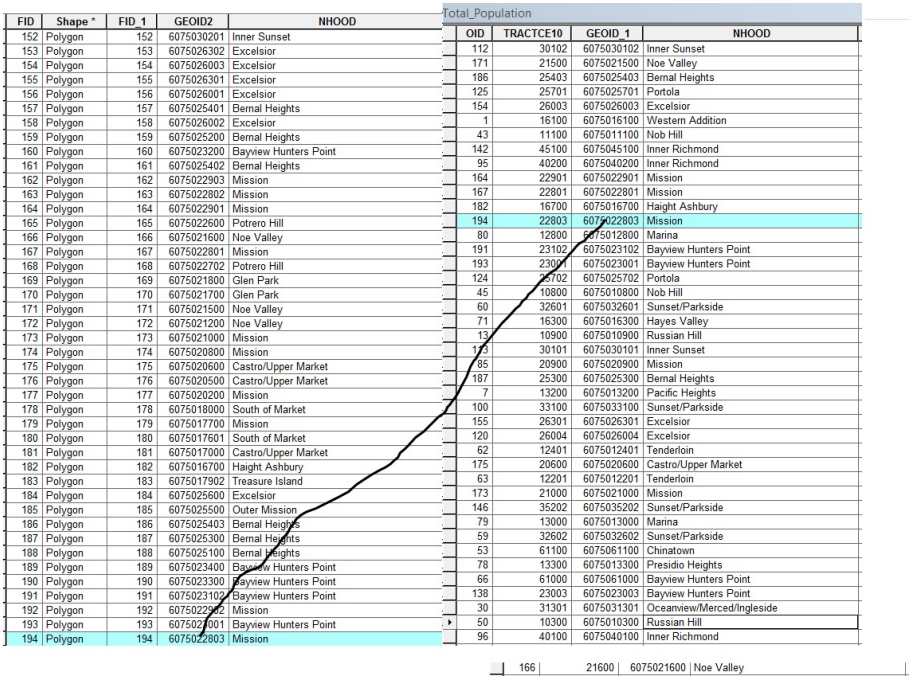
Problem: In order to calculate population density I need both total population and the ability to calculate area. Presently they are in different files. The table and the .shp file. We need both in the .shp file to calculate population density.

Solution: Join the table to the shapefile.

Essentially you are taking the information from the table Total_Population and adding it to SF_NEIGHBORHOODS. The rules of relational databases require that we match record to record from table to .shp file in order to join. Again, the goal is transfer the information from the table to the .shp file.

We will accomplish this task by using the 'joins and relates' dialogue on the SF_NEIGHBORHOODS .shp file.

First, we should go over some rules of joining.



FID	Shape *	FID_1	GEOID2	NHOOD
152	Polygon	152	6075030201	Inner Sunset
153	Polygon	153	6075026302	Excelsior
154	Polygon	154	6075026003	Excelsior
155	Polygon	155	6075026301	Excelsior
156	Polygon	156	6075026001	Excelsior
157	Polygon	157	6075025401	Bernal Heights
158	Polygon	158	6075026002	Excelsior
159	Polygon	159	6075025200	Bernal Heights
160	Polygon	160	6075023200	Bayview Hunters Point
161	Polygon	161	6075025402	Bernal Heights
162	Polygon	162	6075022903	Mission
163	Polygon	163	6075022802	Mission
164	Polygon	164	6075022901	Mission
165	Polygon	165	6075022600	Potrero Hill
166	Polygon	166	6075021600	Noe Valley
167	Polygon	167	6075022801	Mission
168	Polygon	168	6075022702	Potrero Hill
169	Polygon	169	6075021800	Glen Park
170	Polygon	170	6075021700	Glen Park
171	Polygon	171	6075021500	Noe Valley
172	Polygon	172	6075021200	Noe Valley
173	Polygon	173	6075021000	Mission
174	Polygon	174	6075020800	Mission
175	Polygon	175	6075020600	Castro/Upper Market
176	Polygon	176	6075020500	Castro/Upper Market
177	Polygon	177	6075020200	Mission
178	Polygon	178	6075018000	South of Market
179	Polygon	179	6075017700	Mission
180	Polygon	180	6075017601	South of Market
181	Polygon	181	6075017000	Castro/Upper Market
182	Polygon	182	6075016700	Haight Ashbury
183	Polygon	183	6075017902	Treasure Island
184	Polygon	184	6075025600	Excelsior
185	Polygon	185	6075025500	Outer Mission
186	Polygon	186	6075025403	Bernal Heights
187	Polygon	187	6075025300	Bernal Heights
188	Polygon	188	6075025100	Bernal Heights
189	Polygon	189	6075023400	Bayview Hunters Point
190	Polygon	190	6075023300	Bayview Hunters Point
191	Polygon	191	6075023102	Bayview Hunters Point
192	Polygon	192	6075022502	Mission
193	Polygon	193	6075023001	Bayview Hunters Point
194	Polygon	194	6075022803	Mission

Total_Population

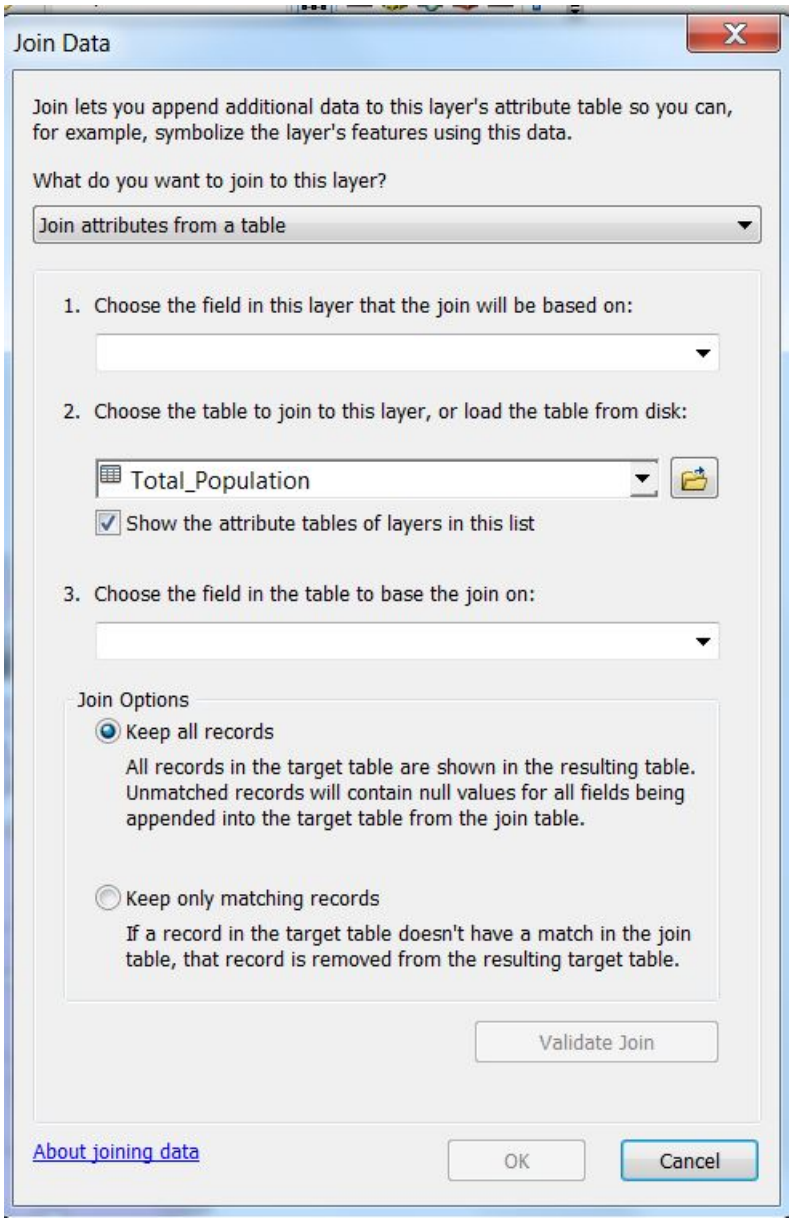
OID	TRACTCE10	GEOID_1	NHOOD
112	30102	6075030102	Inner Sunset
171	21500	6075021500	Noe Valley
186	25403	6075025403	Bernal Heights
125	25701	6075025701	Portola
154	26003	6075026003	Excelsior
1	16100	6075016100	Western Addition
43	11100	6075011100	Nob Hill
142	45100	6075045100	Inner Richmond
95	40200	6075040200	Inner Richmond
164	22901	6075022901	Mission
167	22801	6075022801	Mission
182	16700	6075016700	Haight Ashbury
194	22803	6075022803	Mission
80	12800	6075012800	Marina
191	23102	6075023102	Bayview Hunters Point
193	23001	6075023001	Bayview Hunters Point
124	25702	6075025702	Portola
45	10800	6075010800	Nob Hill
60	32601	6075032601	Sunset/Parkside
71	16300	6075016300	Hayes Valley
13	10900	6075010900	Russian Hill
113	30101	6075030101	Inner Sunset
185	20900	6075020900	Mission
187	25300	6075025300	Bernal Heights
7	13200	6075013200	Pacific Heights
100	33100	6075033100	Sunset/Parkside
155	26301	6075026301	Excelsior
120	26004	6075026004	Excelsior
62	12401	6075012401	Tenderloin
175	20600	6075020600	Castro/Upper Market
63	12201	6075012201	Tenderloin
173	21000	6075021000	Mission
146	35202	6075035202	Sunset/Parkside
79	13000	6075013000	Marina
59	32602	6075032602	Sunset/Parkside
53	61100	6075061100	Chinatown
78	13300	6075013300	Presidio Heights
66	61000	6075061000	Bayview Hunters Point
138	23003	6075023003	Bayview Hunters Point
30	31301	6075031301	Oceanview/Merced/Ingleside
50	10300	6075010300	Russian Hill
96	40100	6075040100	Inner Richmond
166	21600	6075021600	Noe Valley

At this stage we will not bog you down in the cumbersome and often complicated nature of 1 to many and many to 1 joins. In this class we will work with 1 to 1 or basic joins. I will also attempt to keep it simple here as joining and how it works in ARC is best demonstrated through a drawn illustration. I would encourage you to ask your instructor about how the joining process works. An illustration can illuminate.

If you right click on SF_Neighborhoods you will see an option called Joins and Relates.

Click on **Joins**

The following menus should come up:



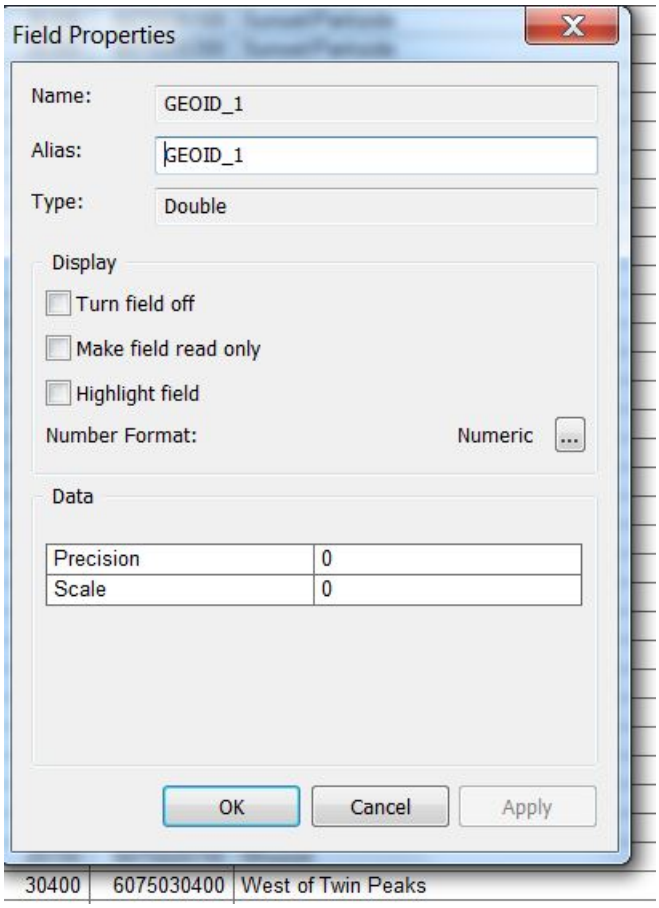
Before we fill in our parameters we should know the following information about joins:

1. When you are joining a table to a .shp file always enter the join dialogue by right clicking on the shape file in this case SF_NEIGHBORHOODS
2. Numbers 1 and 3 in the joins dialogue must be similar
    a. They should refer to the same geographic unit (Census Tract, Neigh-

borhood etc)

b. They should uniquely identify each enumeration unit
    i. This unique identification is often a combinations of numbers and letters
    ii. A unique identification introduces a system of data organization in which a value can only refer to one case. If there is any possibility that there are 2 of the same entries in the column it is not the one you are looking for.
    iii. There can be several unique identifier columns in a table
    iv. OID or FID is NEVER a unique identification column. This is the order which ARC enters each entry. This is often random and does not refer to the unique geographic case or enumeration unit you are interested in.
c. These unique identification columns must exist both in the .shp file and the table. You are essentially playing an elaborate matching game.

3. The columns that you choose to join must have the same type
    a. Right click on GEOID_1 (in Total_Population) and select properties

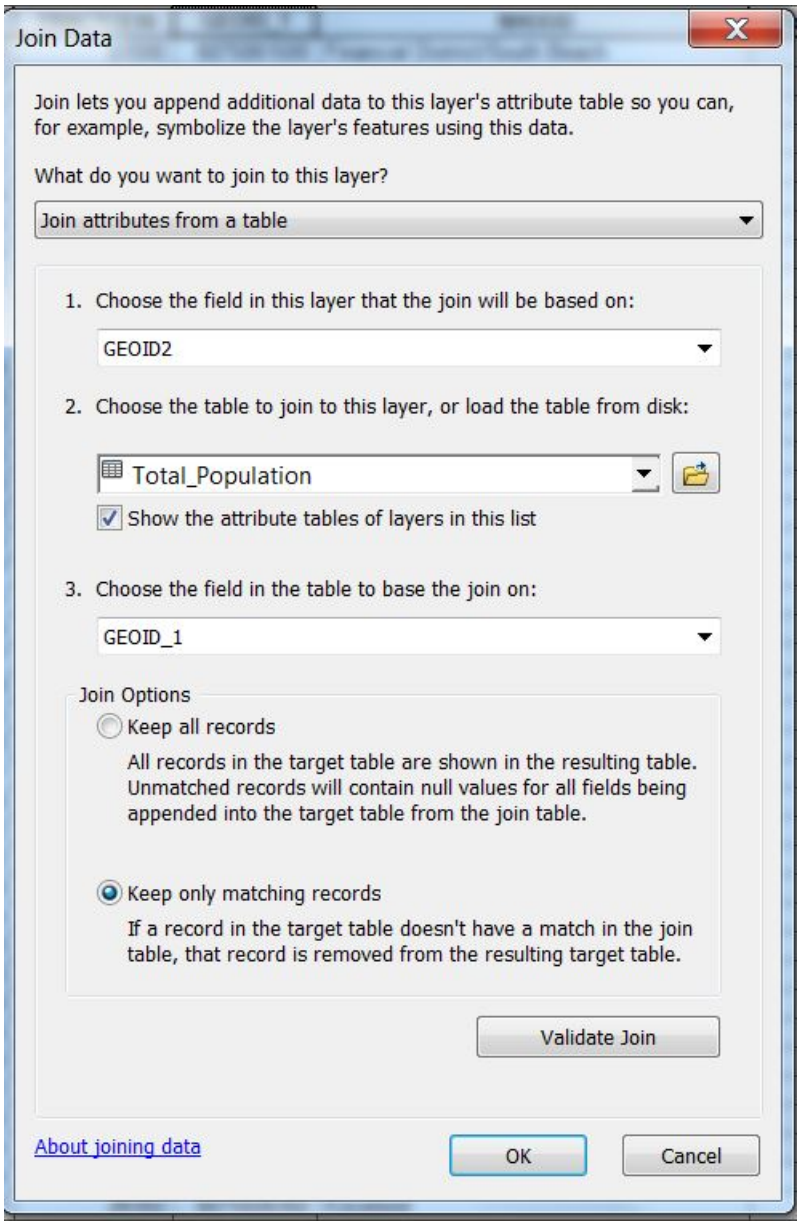


b. Under Alias you will see the category type - it says double
c. You will have to ensure that the column in Total_Population you are going to join to SF_N
d. Types signal to ARC what type of values you will be feeding or giving it. You are telling
    i. Double is the type for GEOID_1. This is good if you are creating a value that has mar
    ii. Short Integer Type. Use this for short numbers like age, which has 2 or 3 places.
    iii. Click on Total_Pop. This column is a long integer. It has 4 places.
    iv. Click on NHOOD. This is a string. This is a different type because it holds letters.
    v. Float is longer than long integer but shorter than double.
    vi. There is also a date option that we will not be using in this lab.
    vii. A lot of this is computer programming. If you would like to do more research on thi

Time to Join

1. Right click on 'SF_Neighborhoods'
2. Then 'Joins' and 'Joins and Relates'

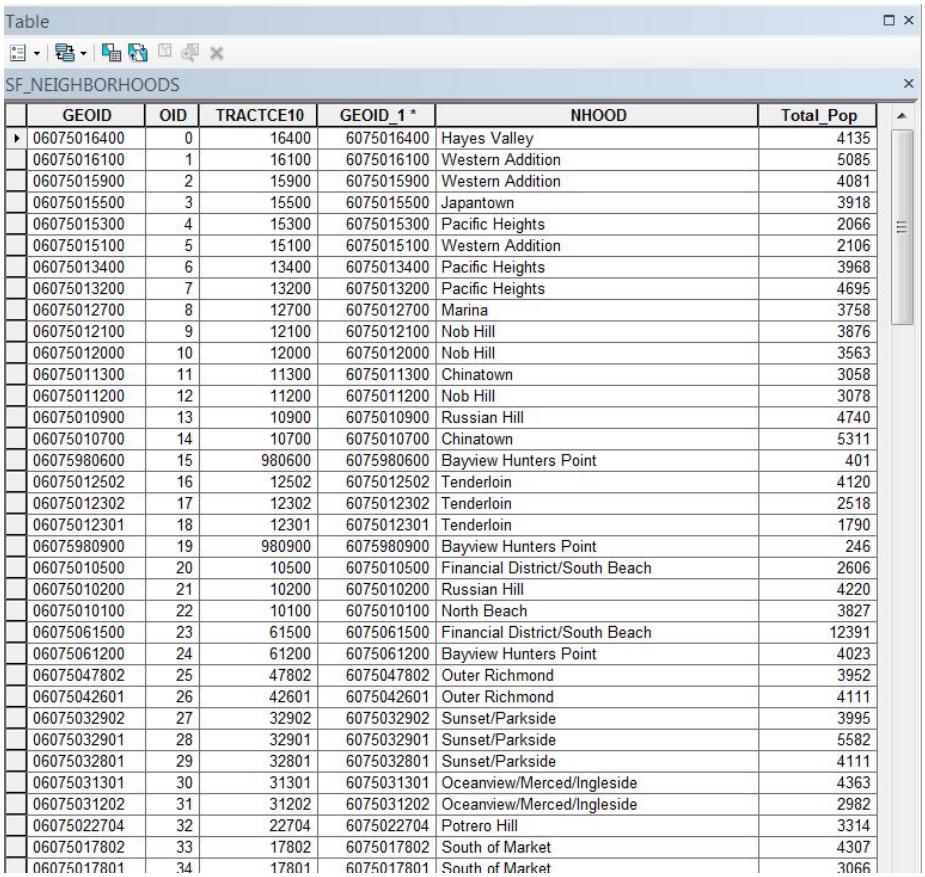
   a. Once we get to the familiar Join menu, for number 1 choose the field GEOID2
   b. Make sure you choose the Total Populations table for number 2
   c. For number 3 Choose GEOID_1
3. Click 'Keep Only Matching Records' – it won't matter here but I find it makes the data cleaner to select this option. This option prevents you from getting null values if some cases don't join.
4. Press OK. You might be asked if you want to index the join. Choose yes. On larger data sets the index helps you located individual cases.



Open up the attribute table in SF_Neighborhoods. You should see that the information from the table is now joined to your shapefile. If it isn't you probably entered the information in the join dialogue wrong. If something is wrong right click on total population and then joins and relates – click 'remove join' and choose 'Total Population' and the table.

I must emphasize that you once you have a bad join you cannot make that

correct it without starting from scratch. If you don't get what you want please remove all joins.



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

1. Make the join permanent (some wacky things can happen in ARC if you don't save it as a new shapefile.
2. Right click on SF_neighborhoods and click on export data. Save as a new shapefile and call it 'Neighborhood_Pop'.

Why this Join Worked

If you were joining without my help you would:

1. Scan the table and shape file to look for a matching unique identification column. That means they have to be on both tables or else the join can't happen. (I'm looking at Total Population) `NHOOD` can't work because there are multiple entries. Neighborhoods contain several Census Tracts. Joining here will confuse ARC. OID will be false because like I said earlier this is

the random order in which ARC enters your cases. The important thing is that records stay with their attributes and ARC does this without fail. Total Pop Is not in the other table. This is why we are joining in the first place. `GEOID_1` is the correct column.

2. Check the type of GEOID_1 by right clicking on it and highlighting properties. It should say that its type is a double. This is important because you are going to have to find a column that looks very similar, with the same amount of characters and the same type as on the column for the .shp file.
3. This is confusing, but unfortunately this happens a lot, especially with Census data. There appears to be many columns that fit the bill. This is important. THE NAME OF THE COLUMNS DON'T HAVE TO MATCH – what is important is that the uniquely identify each case, they have the same number of characters and they have the same type.
4. GEOID looks like it could work but the type is string we don't want that one
5. You'll see that GEOID2 has a similar number of characters, uniquely identifies each record, and has the same type. This is the column we will join to.

Summarizing and Joining

Now that you have joined for the first time I am going to let you practice again. Bring in 'SF_Dissolved' into ARC MAP. This file is a dissolved version of SF_Neighborhoods. Now instead of Census Tracts we only have neighborhoods. You will learn how to dissolve later in the semester.

Please note that all of the .shp files are in the proper coordinate system (In this case California State Plane)

Our Problem: We need to Join Total_Population to SF_Dissolved but the unique identification column we used earlier doesn't exist on the table. Also, there are many fewer neighborhoods than Census Tracts.

Solution: Summarize Total_Population.

In order to follow our rules of joining and add population data for each neighborhood to SF_Dissolved we have to summarize total_population. Please ask your instructor to illustrate this concept for you on the dry erase board. Essentially ARC is summarizing common values in a specific field for us. In this case instead of having many entries for the neighborhood Pacific Heights (for example) we will have only one.

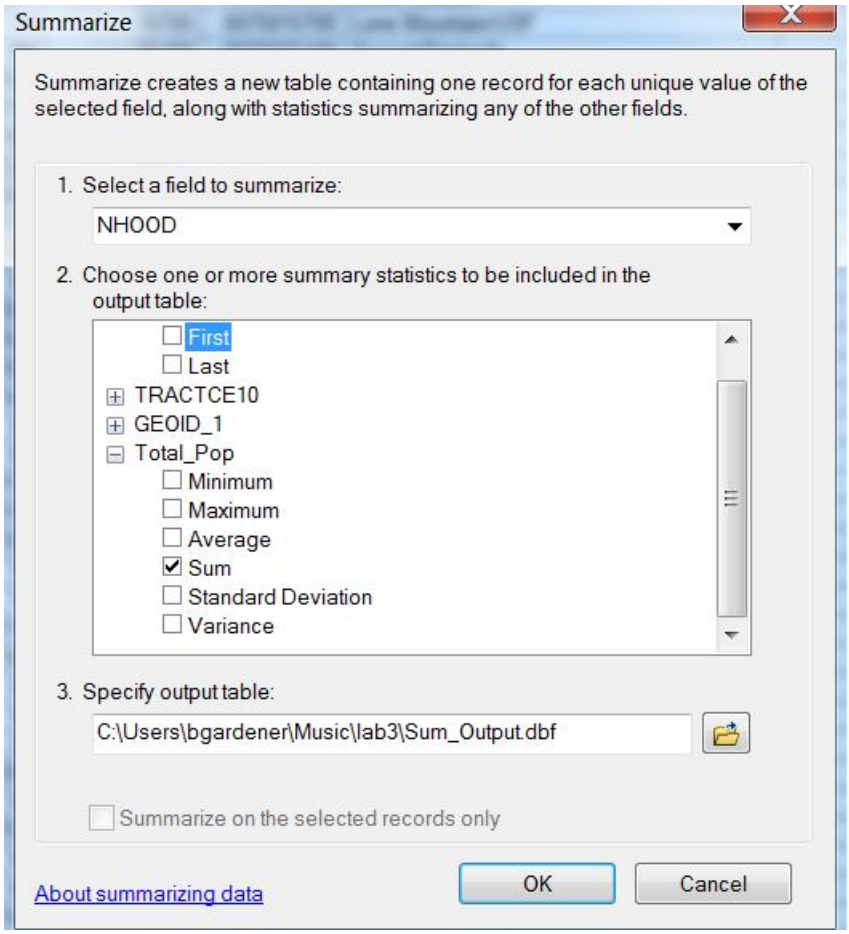
Summarize will count how many times each neighborhood appeared in the column. We can also add together the population counts for each Census Tract

and include them in our neighborhood summary table.

This is how it works:

1. Open 'Total_Population' with a right click and 'Open Attribute Table'
2. Right Click on `NHOOD`. This is the column we want to summarize because we want to join it to SF_Dissolved which has the column neighborhood. This will be the unique identification column in this case.
3. Click 'Summarize'
    a. Our field to summarize is already NHOOD. Great
4. For number 2 uncollapse `Total_Pop` (Click on the plus sign)
    a. You will have a number of options. When ARC condenses or summarizes all the rows or entries that have the same neighborhood name – for example – Pacific Heights, it is asking us what we want to do with the attributes connected to these cases. It cannot express all of the Total_Pop counts for all of the Census Tracts in Pacific Heights – there is only room for one number now.

If there are 7 Census Tracts to a neighborhood, those tracts will disappear and merge into one neighborhood row. We are only interested in saving or keeping total population information. We want the total population per neighborhood. So we are going to add together the individual populations of each Census Tract. These numbers will be tallied based on what neighborhood they are located in. 1. Check the sum box under Total_Pop. Save it to a place that makes sense and call it 'Neighborhood_Pop'. 2. Add it to the map and click on the attribute table


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

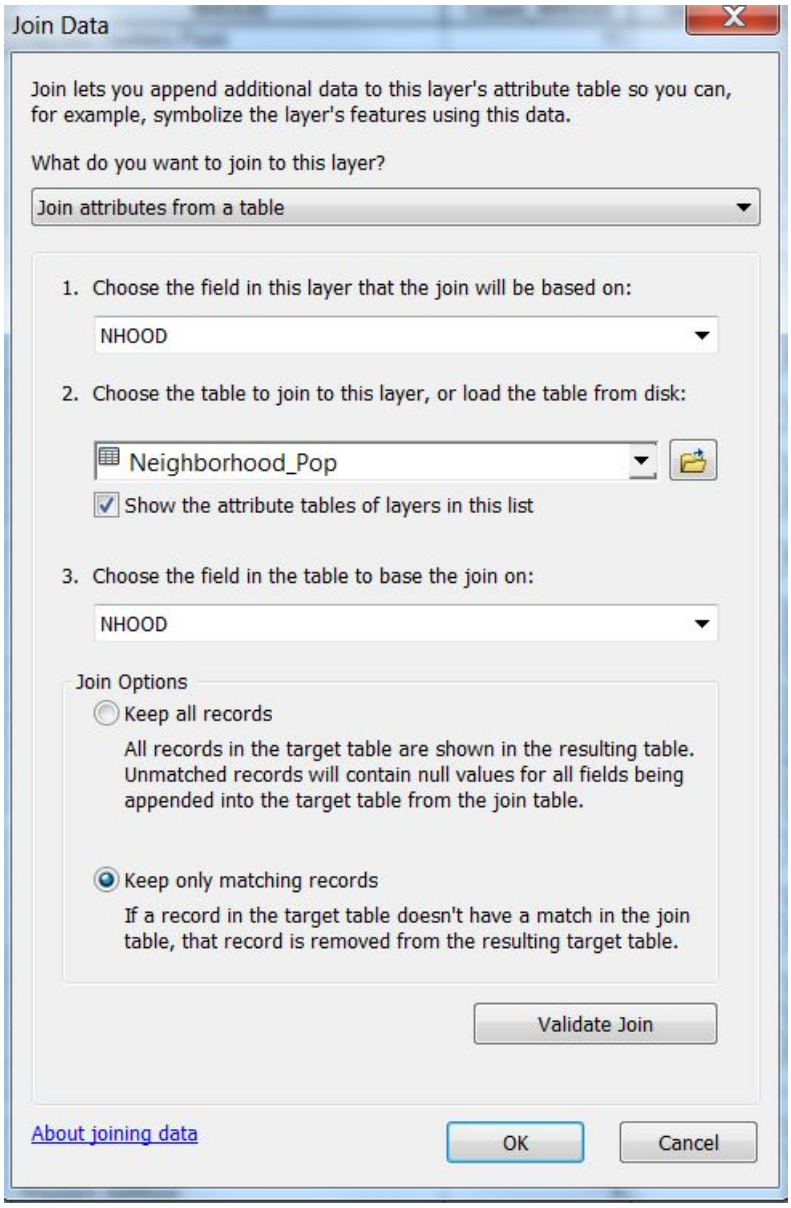
Notice that the summary table has one entry for each neighborhood. Notice there are only 41 entries.

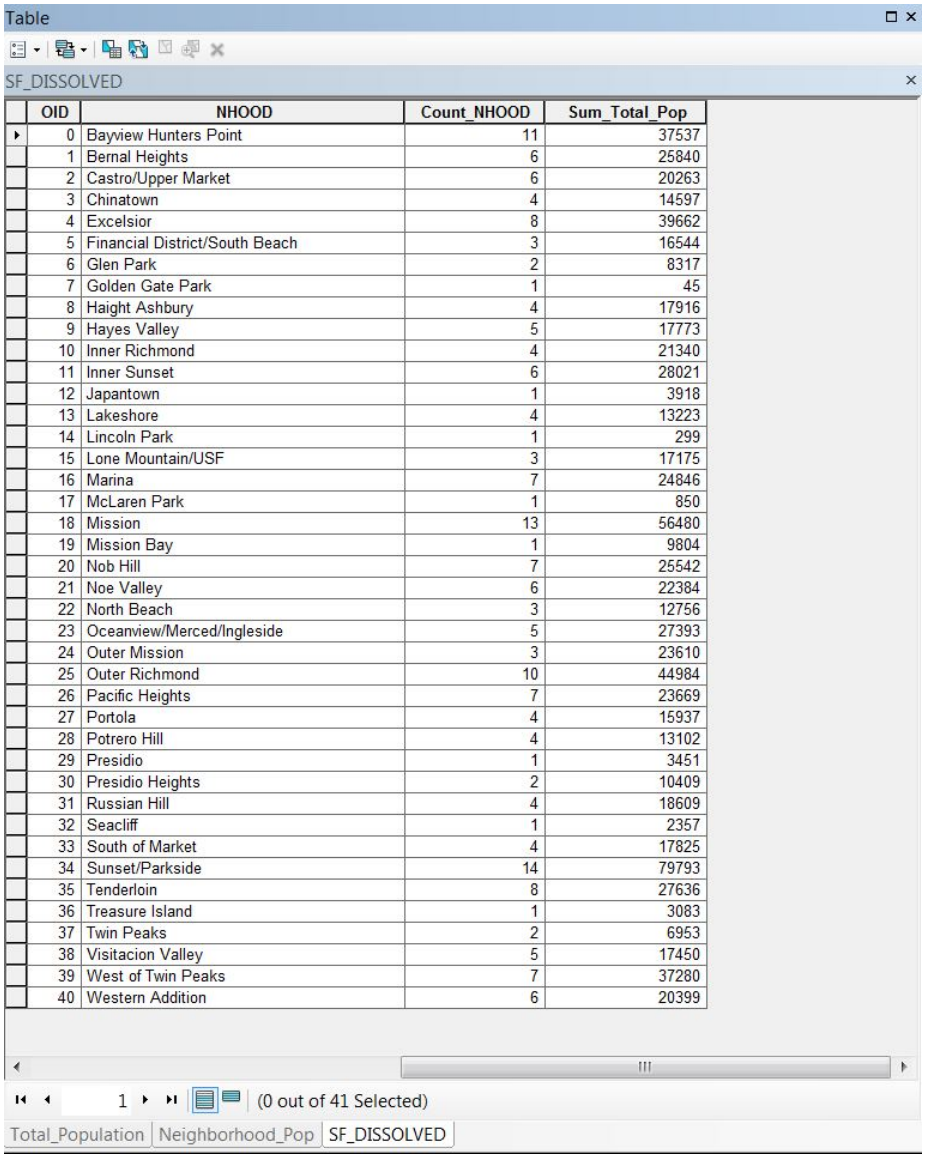
Notice there is a Count_NHOOD column that tells how many times each neighborhood was on the original table. This column is automatically included in summary and will be useful to use when you map homicides per police district in the lab/assignment.

For example – there were 11 Census Tracts in Bayview Hunters Point on the previous table (Total_Population). The last column (Sum_Total_Pop) contains the sum of populations of individual Census Tracts. The Census Tracts on Total_Population that had the NHOOD value Bayview Hunters Point.

Now we can join it to SF_Dissolved using `NHOOD` as the join field because the characters are similar, the column is a unique identifier and the type is the same.

1. Right click on SF_Dissolved – joins and relates – joins
    a. For number 1 choose `NHOOD`
    b. For number 2 choose `Neighborhood Pop`
    c. For number 3 choose `NHOOD`
2. Choose ok
3. Select 'Keep only matching'


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



Table

SF_DISSOLVED

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

1 (0 out of 41 Selected)

Total_Population | Neighborhood_Pop | SF_DISSOLVED

4. Right click on your joined file in `SF_DISSOLVED`
5. Export your data and save the shape file as `SF_DISSOLVED_POP`

Population Density

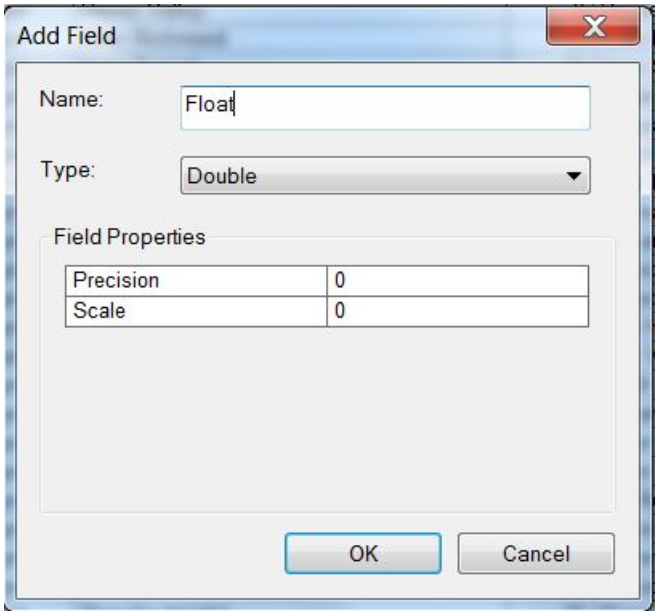
To calculate population density (the amount of people per square unit – a relative measure of how relatively concentrated a population is given an enumeration

unit) we need to find the area of each neighborhood. In this case we will choose meters.

Add Field

To calculate area we need to create a new field:

1. Open the attribute table for SF_DISSOLVED_POP
2. Click on options in the upper left hand corner of the table and select 'Add Field'.
3. Name the field `Area` and make it a double type



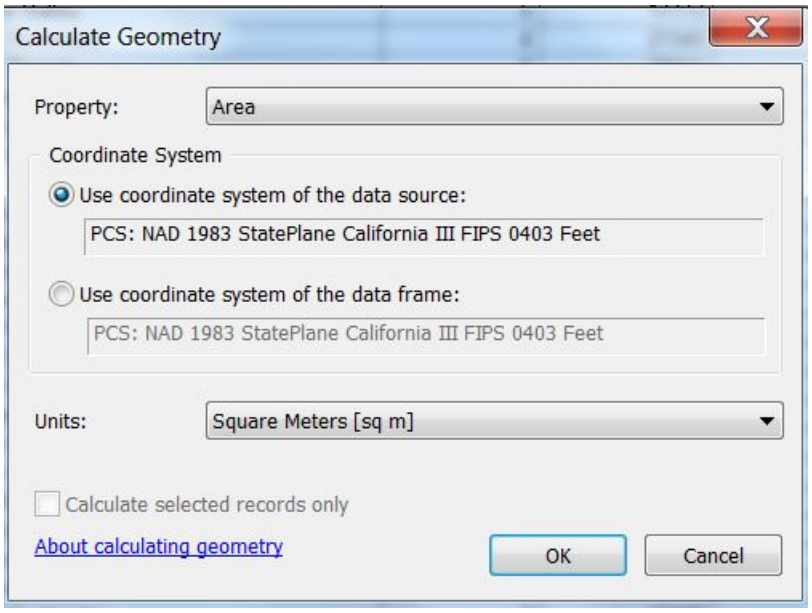
Your attribute table should show your new column - the last entry on the right side.

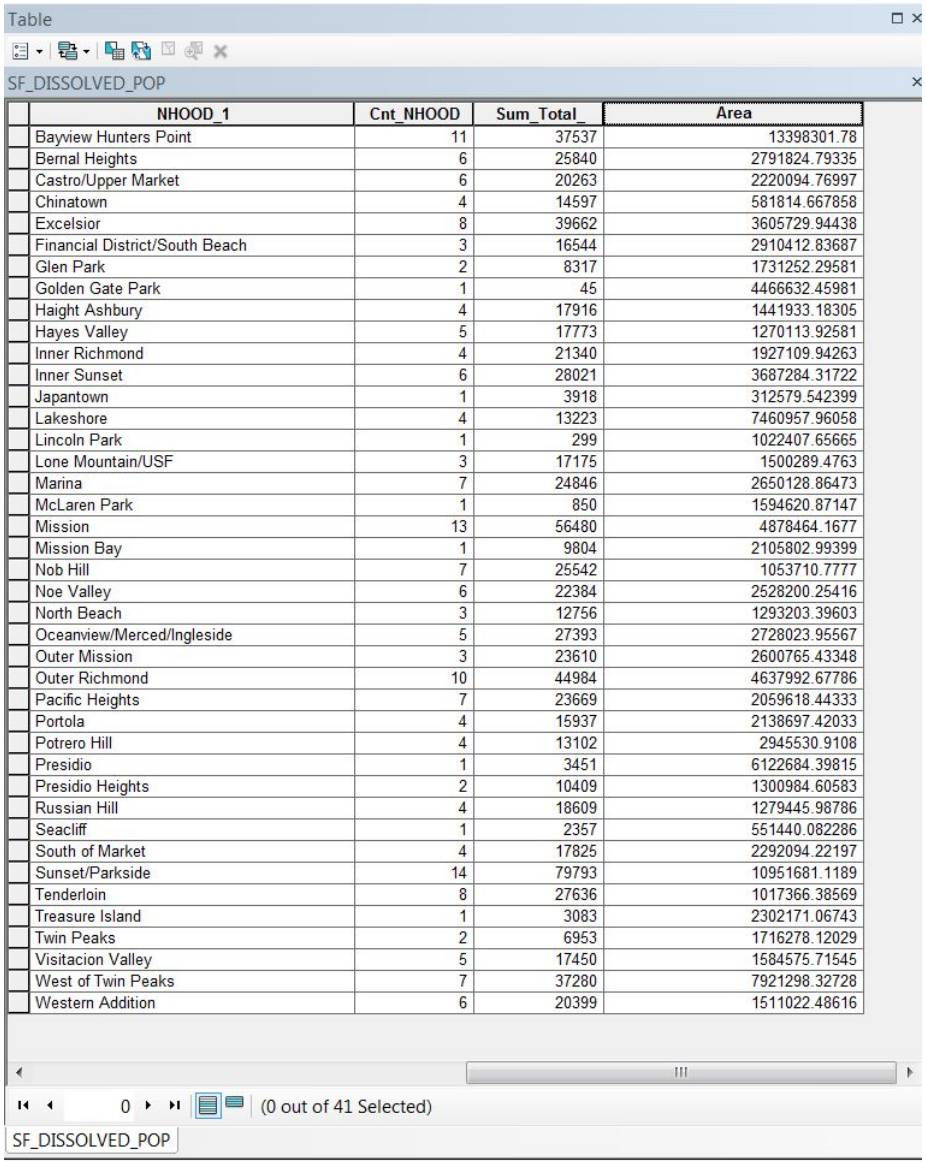
Calculate Geometry

1. Right click on your new column area and select 'calculate geometry'
    a. ARC asks you if you would like to make changes outside of an edit session - please do so
        i. Note that you will use the coordinate system you already have in place. Note that a coordinate system can determine area calculations and if you are not using a reasonable one (based on many factors – talk to your Professor about it) it can greatly

distort your results. If you have no coordinate systems you cannot calculate area. Choose square kilometers as your unit of measurement.

2. Press ok and your area column should be populated.





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

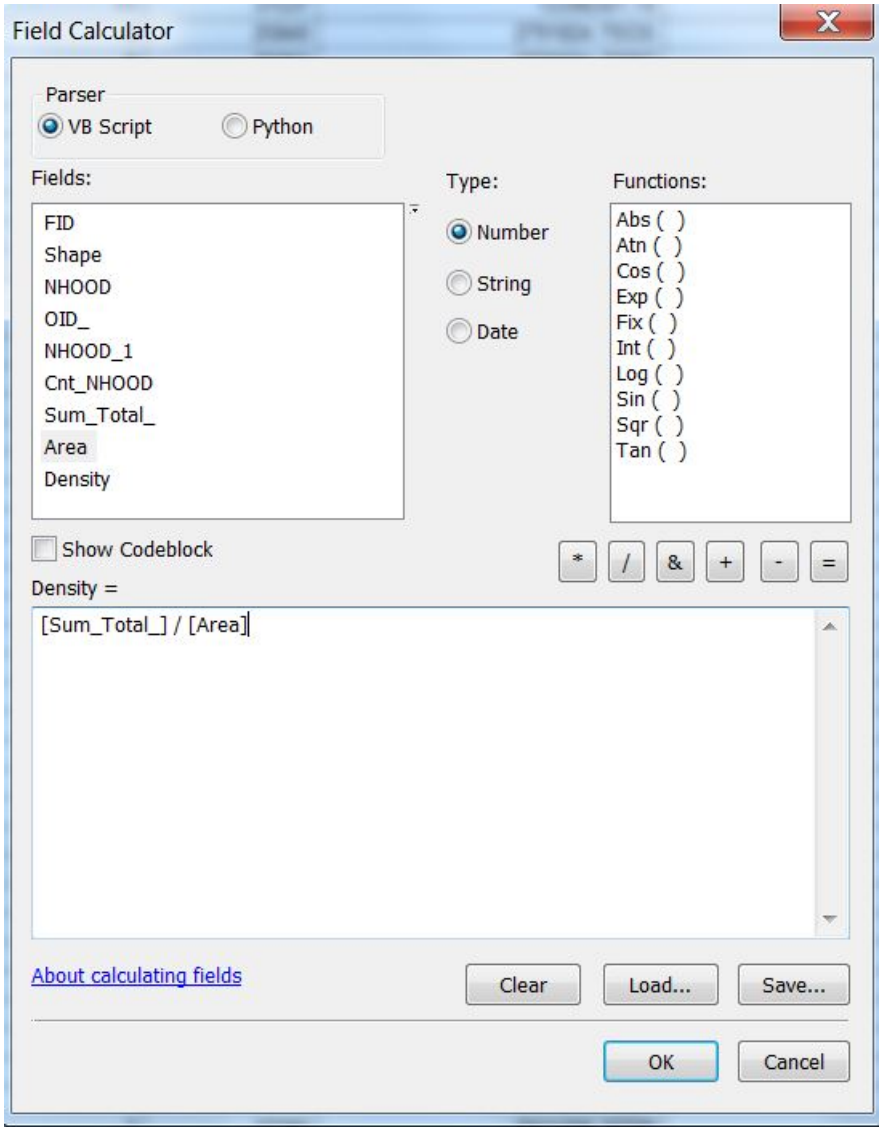
0 (0 out of 41 Selected)

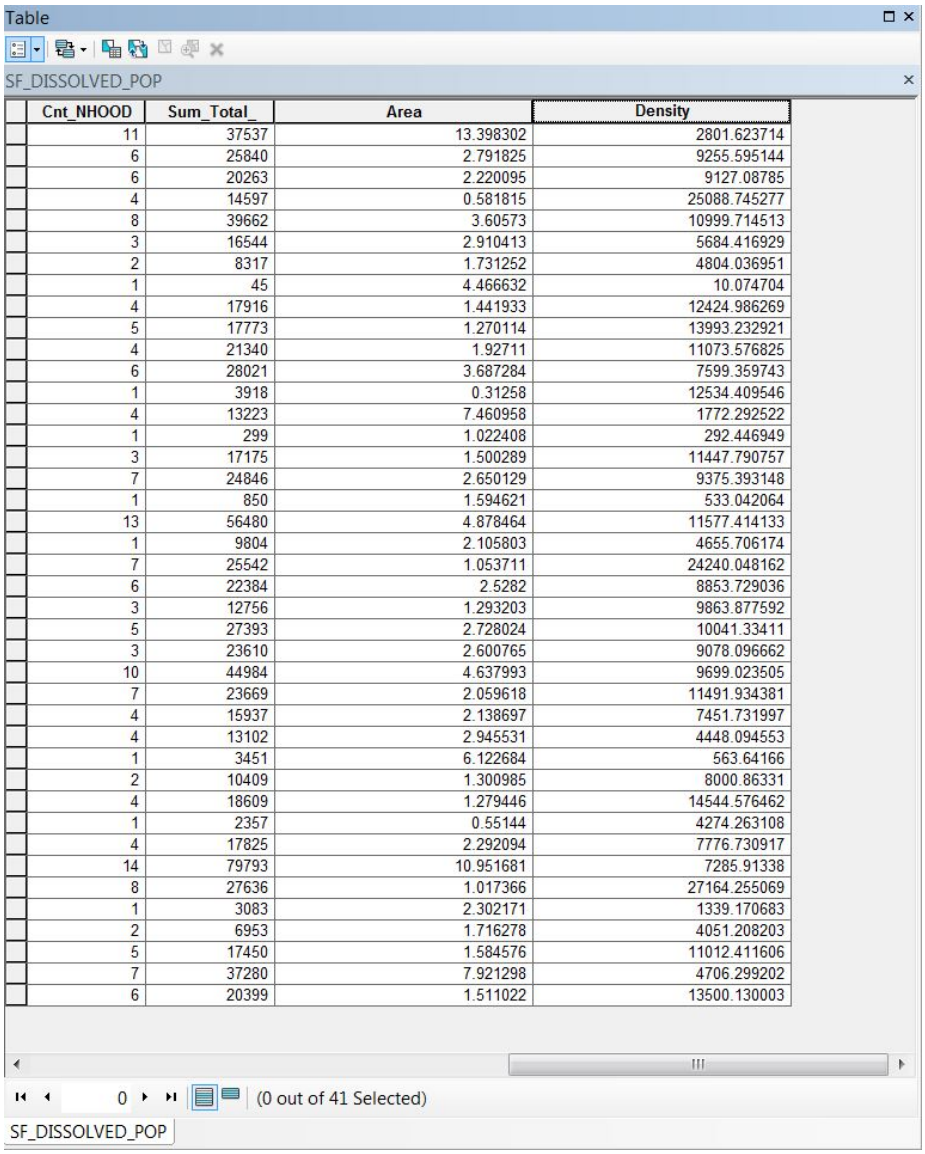
SF_DISSOLVED_POP

Field Calculator

1. Add a new field and make it double and call it density
2. Right click on the new column density and choose 'field calculator'
3. Click on the name in the field that corresponds to your population column
4. To the right there will be some operator buttons - choose the division symbol (a forward slash)

5. Choose the field under fields with the area information.
6. Divide the population by the area
7. Press ok





Cnt_NHOOD	Sum_Total_	Area	Density
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
1	850	1.594621	533.042064
13	56480	4.878464	11577.414133
1	9804	2.105803	4655.706174
7	25542	1.053711	24240.048162
6	22384	2.5282	8853.729036
3	12756	1.293203	9863.877592
5	27393	2.728024	10041.33411
3	23610	2.600765	9078.096662
10	44984	4.637993	9699.023505
7	23669	2.059618	11491.934381
4	15937	2.138697	7451.731997
4	13102	2.945531	4448.094553
1	3451	6.122684	563.64166
2	10409	1.300985	8000.86331
4	18609	1.279446	14544.576462
1	2357	0.55144	4274.263108
4	17825	2.292094	7776.730917
14	79793	10.951681	7285.91338
8	27636	1.017366	27164.255069
1	3083	2.302171	1339.170683
2	6953	1.716278	4051.208203
5	17450	1.584576	11012.411606
7	37280	7.921298	4706.299202
6	20399	1.511022	13500.130003

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

Try creating a map of population density using the skills you learned last week in thematic mapping.

You don't have to submit anything here, but it's good to refresh your mapping skills. You will need them in the lab/assignment.

Assignment, Objectives, and Deliverables

Part I

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 (But for real there was just a drive by in front of my friend's apartment in Chicago). You want to find out where the geographic concentration of murders is so your boss can distribute police resources appropriately. She wants you to create a map that shows the density of murders in each police district per square mile.

Your final answer should be a choropleth map that shows density of murders per square mile in each district.

Part II

Your boss is also interested in the relationship between crime and Airbnb listings. The total numbe of reviews are a proxy for activity in the Airbnb market. Please create the following tables and maps so we can appropriately consider this question.

She would like:

a) A table that summarizes the average Airbnb price for each neighborhood in Chicago. (Column Price)

And

b) A map that shows the density of total Airbnb reviews per police district. (The name of the column is 'Number of') The density should be per square mile.

Getting Started

Each point file contains districts. You will have to join each table to the shape file Police_Districts_Chicago.

Start by summarizing your data.

Remember that the crimes data table contains crimes other than homicides. You will want to isolate homicides before you continue.

Remember that summarize counts for you. Use this knowledge to your advantage.

In summarize you have a choice of what type of mathematical operation you perform on your data. We used sum in the tutorial – what other operator might you use?

In your report, be sure to include the following information (see the syllabus for lab report guidelines and formatting):

Introduction:

State the research question.

Data and Methods:

The data comes from 'Inside Airbnb' a watch dog for Airbnb. Please visit the website. It's pretty interesting if you have ever used Airbnb. They have collected data for cities all around the world.

The crime data comes from the city of Chicago. Please email bradley.gardener@temple.edu if you want to look at the entire crime data set. Many cities Open Data websites have similar data.

The structure of this lab is very similar to the tutorial. You will be summarizing data and joining it. You will be calculating density in both cases - per square mile - using the tools add field, calculate geometry, and field calculator. You will map population density using the techniques of choropleth mapping from previous labs. Please use a classification scheme that makes sense given what you have already learned.

Results:

You should include the following:

1. A map of murders per police district area (square miles) in Chicago.
2. A map of the total number of Airbnb reviews per police district area (square miles) in Chicago.
3. A table that shows the average Airbnb price of each neighborhood in Chicago.

Discussion:

In your final analysis, look at both of the maps you created and answer the following questions:

1. Is there a geographic relationship between the density of Airbnb reviews and the density of homicides?
2. What neighborhoods have the highest Airbnb prices? Why do you think this is? Do some Internet investigations. Look at Yelp. What types of restaurants are in these neighborhoods? Why is sample size important here?
3. What other crimes would you like to map? Where do you think their spatial distributions would look like?
4. What other crime data do you think it would be important to map?
5. What crimes would be difficult to map?

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.