GEOG 2100: Introduction to GIS

Lab 5 Attribute data operations

Goals:

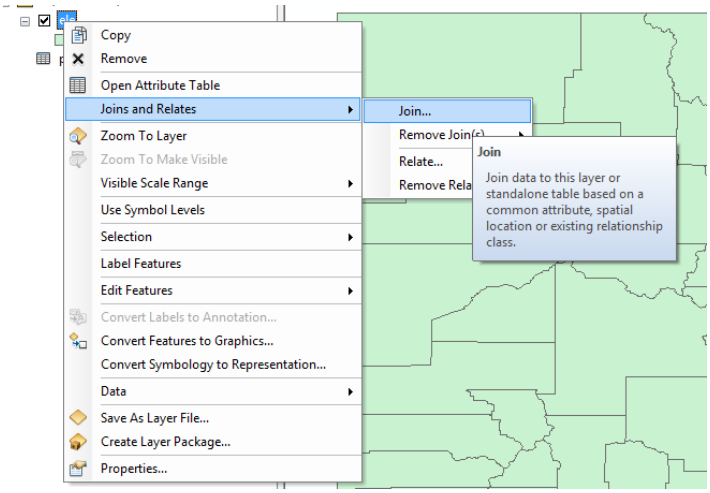
- To demonstrate that you know how to join attribute data to cartographic data
- To demonstrate that you understand the process of performing attribute selection

Datasets:

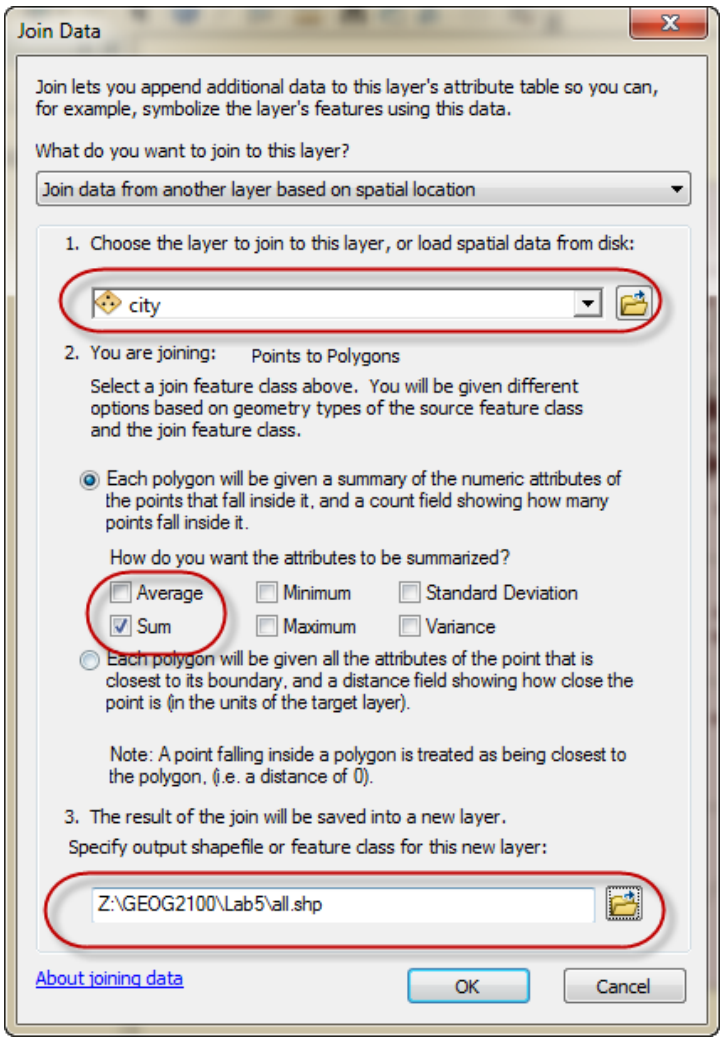
- Presidential election results in Colorado: http://nationalatlas.gov/mld/popul08.html by county.
    - Attributes
        - VOTE_DEM
        - VOTE_REP
        - VOTE_OTH
        - TOTAL_VOTE
- Census population data in Colorado: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.
    - Population group: age >18
    - Attributes
        - GEOID: FIPS id of a county
        - Name: county name
        - Total: total population, age>18
        - Male: male population, age>18
        - Female: female population, age>18
        - HighEdu: people have bachelor degree

Note: There are **a few errors** in the raw data (e.g., the total population is less than the population of a category). For this lab assignment, you don't need to deal with the errors in the raw data.

1. Data preparation. Download and unzip data.
2. Data preprocessing (**2 points**).
    a. Open ArcMap and load population data (**pop.dbf**), **city.shp** and election data (**ele.shp**). View the attribute tables and identify the **common** fields for the join operation. You may need to perform a field calculation to convert the common field so both fields are of the same data type.
    b. Perform an **attribute join** to join population data with the election data.
        i. Right click **ele.shp** and select **Joins and Relates→ Joins**

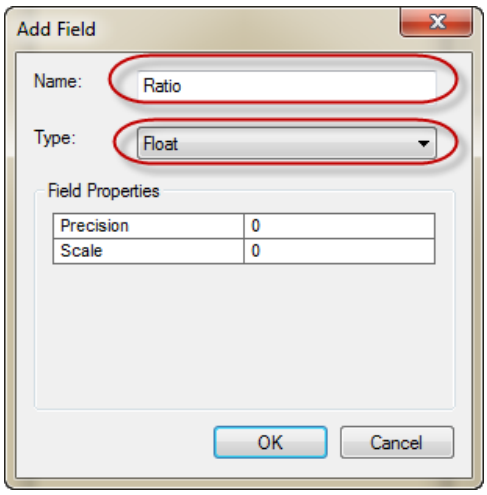


        ii. Set up attribute join and specify the fields (**a common field**) for joining.
        iii. View the attribute table of the **ele.shp**.
        iv. Save the joined results by exporting the dataset. Right click the ele.shp→Data→Export Data. You should specify the datatype as **shapefile**. Name the shapefile as '**final.shp**'
    c. Perform a spatial join to identify cities located in each county.
        i. Right click the final layer "**final.shp**" and select **Join and Relates**
        ii. Configure the join**.** Switch the join option to **"based on spatial location".** Change the summary **to Sum.** Change the name of the output shapefile to "**all.shp**".

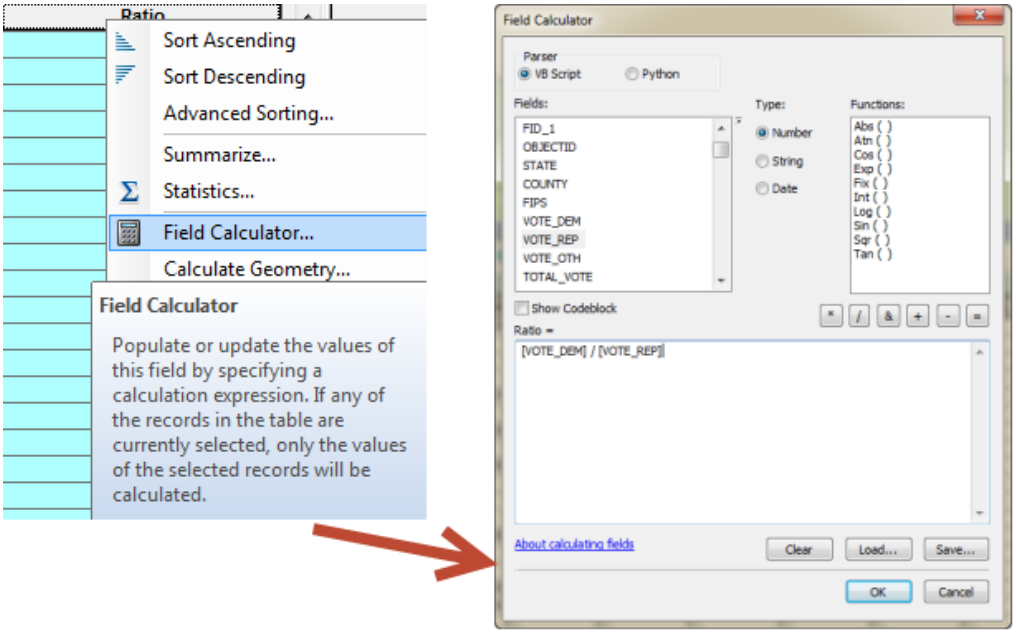


3. Use field calculator to identify the winner party of a county.
    a. Add "**all.shp**" to the map. Add a field to store the election results.
        i. Open the attribute table of **all.shp**. Right click the **table options**

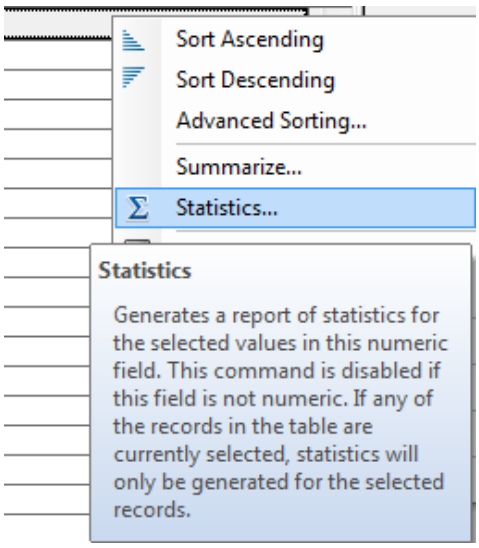
in the top left corner of the table and select **Add Field**. Specify the name ([**Ratio**]) and type of the attribute field (**Float**).



ii. Right click the field and select Field Calculator. Write the expression **[VOTE_DEM]/[VOTE_REP]**. You can select the fields by clicking the field from the Fields list. If the value is greater than 1, the winner party is DEM. If the value is less than 1, the winner party is REP.



b. View the statistics of the field. Right click the field and select **Statistics**.



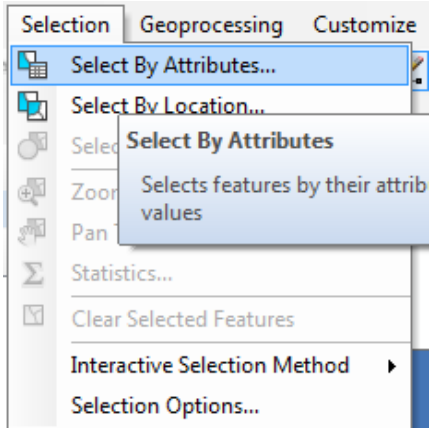
c. Create a ***choropleth map*** and add text description of the spatial distributions of election results in the map. You will need to use different symbols (e.g., Blue or Red) to show the winner of each county. **Label** counties with their names. You can refer to Lab 1 for labeling and symbology. Describe the patterns in the map (**2 points**).

d. Develop another method to identify the winner party (e.g., difference) (**3 points**)

4. Perform queries on the tables and answer the following questions (points may vary). **After you perform selection, right click the layer, Selection→Create a layer from selected records to save the selection results temporarily.**

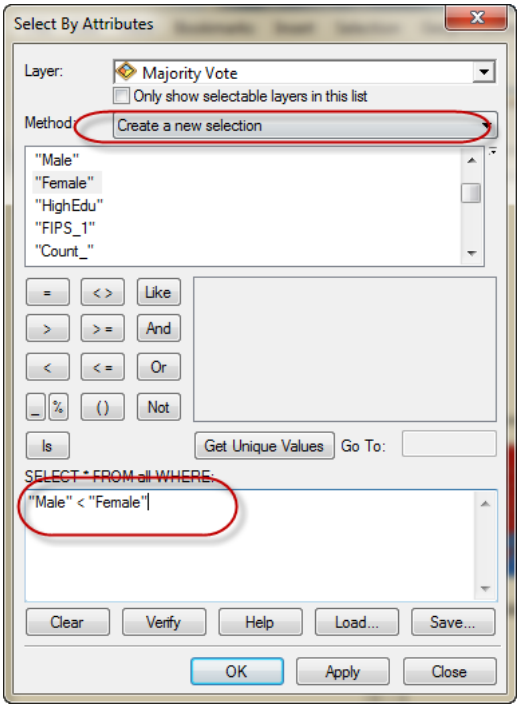
   a. **Sex**. **How many counties that have more female population than male population? What about the election results of these counties?** (**1 point)**

      i. Go to menu **→Selection→Select by Attributes**

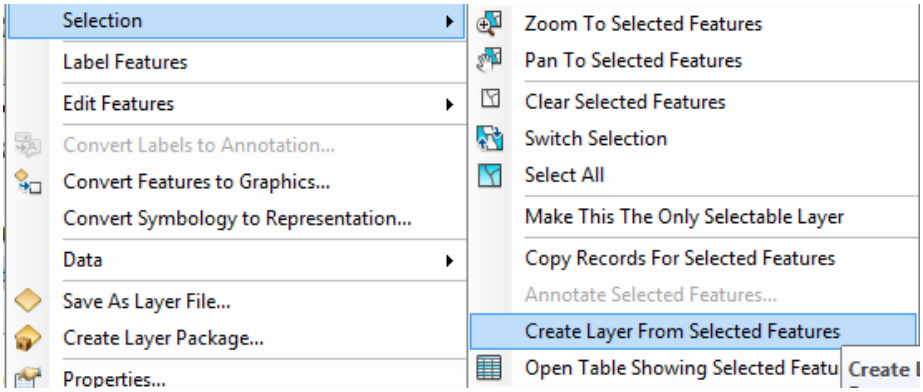


      ii. Set up the query and click OK. The expression should be **"Male" < "Female"**. You can select the field name from a list of fields in

the query window. You should see a few counties highlighted in the data view.

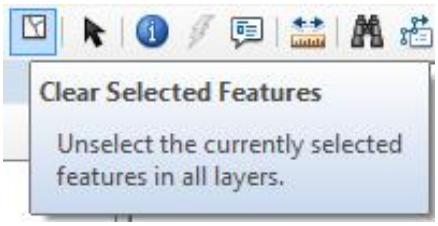


iii. Open attribute table to view selected records. Save the result of selection. This is a temporary saving solution. Right select the layer, Selection→**Create Layer From Selected Features**.



b. **Education**. Calculate the percentage of higher education population over total population. **Identify the top 5 counties** in term of **percentage** of higher education population. **Among those counties, do you observe any relationships between education and election results? (1 point)**

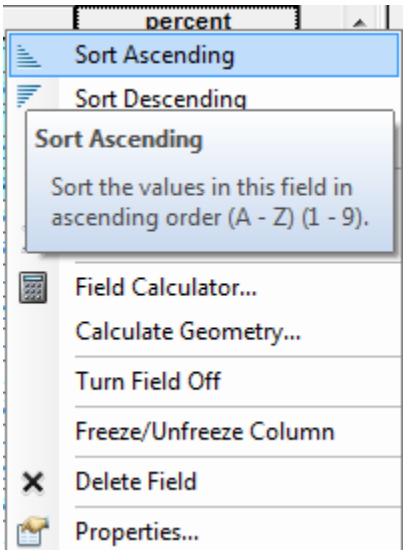
i. Clear selection (The first icon in the figure below)



ii. Add a **field** to store the percentage of higher education population of each county.

iii. Use **Field Calculator** to calculate the percentage. Expression: **[HighEdu] / [total]**

iv. Sort the field by descending.



v. Select **the top 5 most educated counties**. When you want to select multiple counties, you should press **Ctrl** and select those counties.

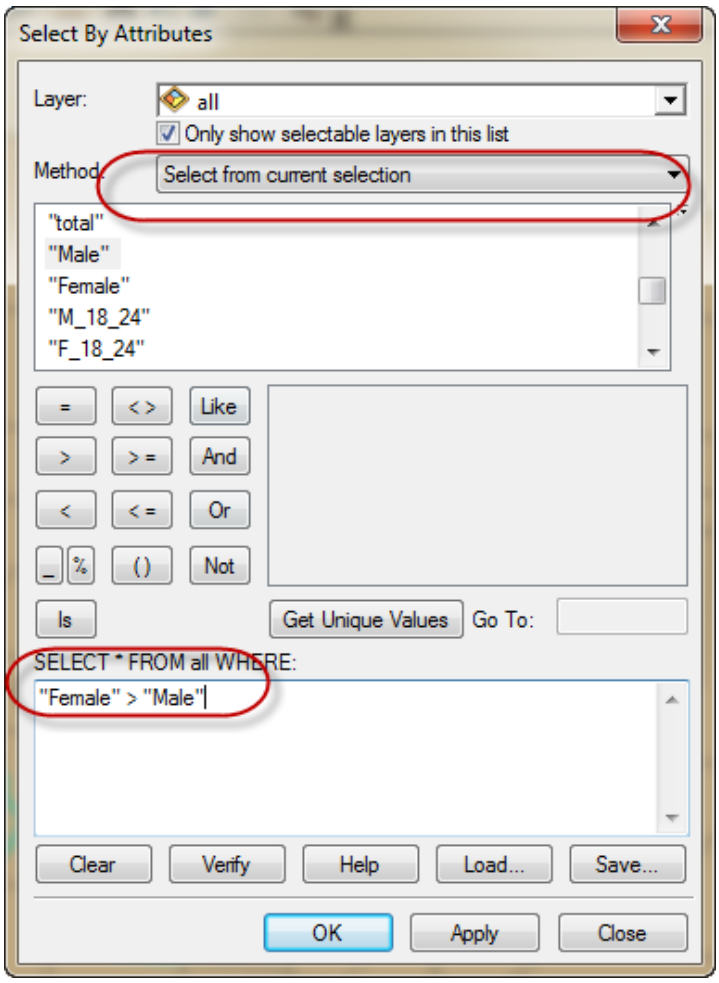
vi. View the field [“**Ratio”**] of those counties or view the map. **Describe the results of selection**. **Do not clear selection**.

c. Combination of selection. **For those highly educated counties, how many counties have more female population than male population**? **What are they? How about the winner party of those counties? (1 point)**

i. Use **Select By Attribute**.

ii. Method should be **“Select from current selection”.**

iii. Expression: "Female" > "Male"



vi. View the results. Try switch selection. How does switch selection change the results? Clear Selection.

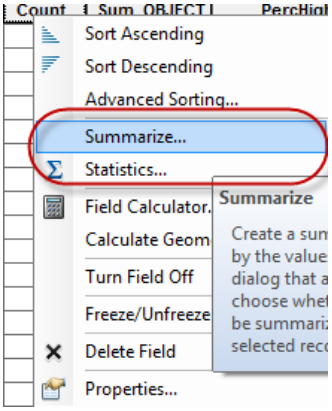
d. Combination of selections. Surveys show that people living in metropolitan areas tend to vote for Demographic party. **For those highly populated area OR counties with more than 4 cities, what about the election result? (1 point)**
   i. Use Select By Attributes.
   ii. Use "**Create a new selection**". Expression: **"total" >100000 OR "Count_" >4**

e. (**4 points**) Design and implement three attribute queries (one should use logical connectors (e.g., OR, AND), 2 points for this one). Write the queries, the processes of implementing queries and the results. You can use all possible attribute operations to implement the queries.

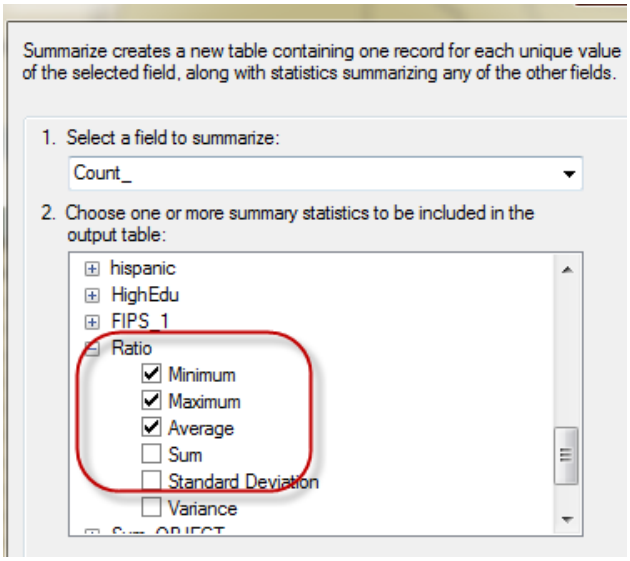
5. Summary statistics. Categorize counties based on the number of cities in the county and examine the election results. **What is the average ratio of counties with 1 city? What are the minimum and maximum ratios for counties with 0 city**? (**1 point**)

i. Open Attribute Table of "**all.shp**"

ii. Find "**Count_"** field and right click the field. Choose "**Summarize**"



iii. Select "**Minimum**", "**Maximum**" and "**Average**" as the statistical values.



iv. Open the attribute table of the summary table.

6. Questions:
   a. (**2 points**) What is the difference between summary statistics (**Summarize**) and **Statistics**? Provide examples (can refer to the previous steps) when you want to use each.
   b. (**2 points**) Use examples to explain the difference between the following two selection options:
      i. Select from current selection
      ii. Add to current selection

Items to turn in:

- A word document for your answers to questions.
- A complete map showing the winner party of each county.