GEO STAC Student Lesson

Lesson Title: Sampling and Bias

Lessons Summary: You will use a map to gather some population data about Oregon for three samples. You will place the data in a spreadsheet to calculate basic statistics about each population set. Finally, you will identify potential bias in each sample by comparing your results to those of Oregon.

Before you begin using this module, you will need to know about using the ArcGIS Server Web-based GIS viewer. You can do this by watching the server tutorial video or working through the tutorial. The tutorial video, student activity, and Web-based GIS Tutorial Viewer can be found at http://gis.lanecc.edu. Go to the "Modules" tab and click the Tutorial link. You will need a high speed Internet connection and a current version of Adobe Flash Player.

Prior Skills: You will need to know how to turn layers on and off, use the ID tool and, zoom in and out of the map, toggle from layers to the legend, and perform a search (Boolean) query. You should also be familiar with spreadsheets, be able to organize data, and perform simple calculations.

Remember, computer steps are indicated by a ▶ symbol and questions you need to answer are numbered.

▶ Open up the web browser go to Lane's GIS page http://gis.lanecc.edu
▶ Click Modules Tab
▶ Click Sampling Bias
▶ Click Link to the web-based GIS map.

▶ The page will take a minute or two to load. The base map will load first then the other data layers. If it stalls at 55%, just click the **refresh** button on your browser.

▶ Use the Zoom Tool to focus the map on the state of Oregon.

▶ In the table of contents on the left, uncheck Lakes, Rivers, National Parks, Federal Lands, and County Names.
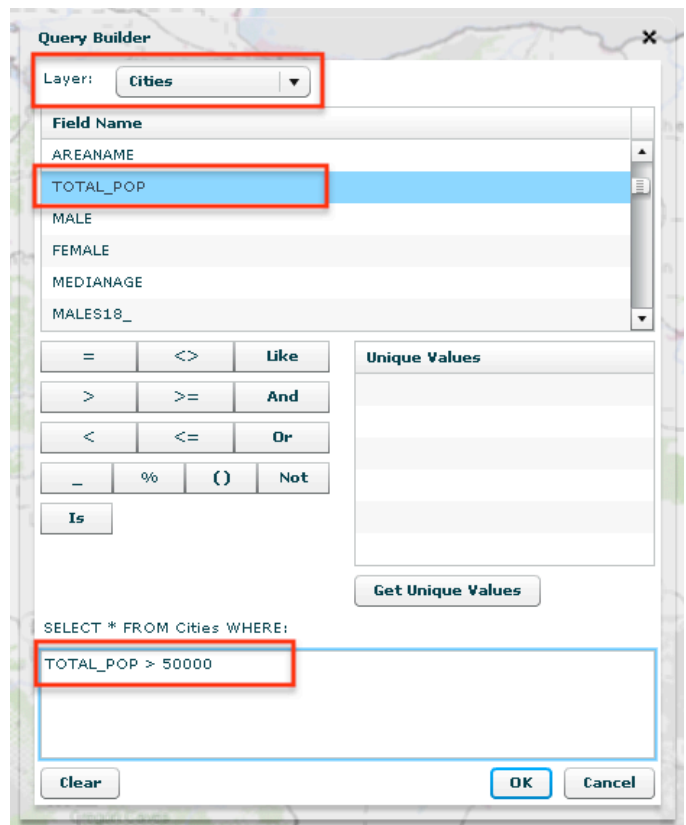
Note that all of the purple areas (polygons) represent incorporated cities. Based on your observations:

1) Describe the spatial pattern of most cities in Oregon. Where are most of the cities and what regions (areas) appear to have no cities?

To gather information about the cities, you can perform a "Query" that will select places based on the specific information.

- ► Click the **Selection** link at the top of the page.

- ► Choose **Select by Attribute.**

- ► In the Query Builder, set the Layer to **Cities.**

- ► Drag the slider to find T**OTAL_POP.** Double click to add it to the box below.

- ► Click the greater than sign > to add it to the box below.

- ► Type 50000 in the box below after the >. Do not use a comma separator.

- ► Your query should look exactly like the one shown

- ► Choose OK

  You will see a table open up in your web browser. In the map, they will be highlighted in neon blue.

2) How many selections are in your table?   At the top of the table it will say Query Results.

3) Look at the map and describe the location of most of the larger cities in Oregon. Are there any anomalies or larger cities where you would not expect to see a larger city?

- ► Copy the information provided about the cities from your query.

- ► Open a spreadsheet such as Excel and paste the information in the spreadsheet. Save your spreadsheet.

- ► Clear your map selections by clicking Selection → Clear Selected Features.

▶ Using the same set of procedures to perform a new query to find the cities with a population less than or equal to 145. If you do not remember how to do this, go back and work through the previous steps, but substitute <= and 145.

4) Describe the patterns on the map that associated with the location of these cities. Are there any anomalies?

▶ Copy and paste the information provided about these cities into your spreadsheet. *Leave at least four rows of space below your previous data.*
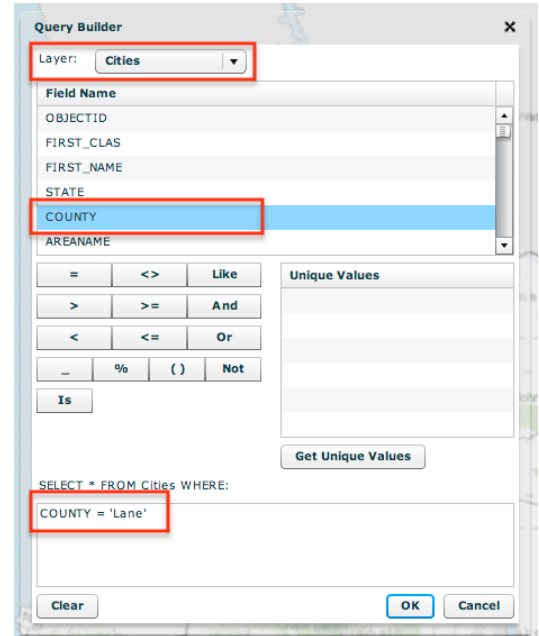
Perform one more query. This time only find at cities in Lane County

▶ Be careful to type in Lane (properly capitalized) and make sure to use single quotation marks. This indicates you are looking for text not a number

5) Describe the patterns of cities in Lane County. How does this pattern compare to the pattern of the other queries?

▶ Copy and paste the information provided about these cities into your spreadsheet. Leave at least four rows of space below your previous data.

As is often the case, there is more data than we need in the table. There are several columns of data you can delete to trim down the file.

▶ In Excel, right-click on the **column letter** heading and choose delete. DELETE columns with the heading listed below.
▶ Go carefully! If you mistakenly delete the wrong column use Control Z or go to EDIT > UNDO.

These are the headings that you should delete:

Rec	FILIPINO	PUERTO_RIC
FID	JAPANESE	CUBAN
OBJECTID	KOREAN	TOTAL_POPU
FIRST_CLAS	VIETNAMESE	LIVEALONE
STATE	HAWAIIAN	HSEHOLD65_
AREANAME	GUAM__CHAM	HSEHOLDW18
MALE	SAMOAN	HSEHOLDW65
FEMALE	ALLWHITE	TOTAL_HOUS
MALES18_	ALLBLACK	VACANT_HOU
FEMALE18_	ALLNATAM	OWNER_OCCU
MALE65_	ALLASIAN	RENTER_OCC
FEMALE65_	RACE_OTHER	Shape_Leng

INDIAN_A	TOTALPOP	Shape_Area
CHINESE	HISPANIC_A	

This is the list of columns that you should have still have. Make the columns wide enough to easily read them.

FIRST_NAME
TOTAL_POP
MEDIANAGE
WHITE
BLACK
AMERICAN_I
ASIAN
MEXICAN
FAMILYSIZE
COUNTY

► In the rows directly below **each set** of data…
A) Calculate TOTALs for Population Data.
B) Calculate AVERAGEs for median age and Family Size[1]
C) Calculate the PERCENTs of population for each population ethnicity.[2]

Remember that in a spreadsheet, all formulas begin with "=", and that you can click on cells rather than re-type numbers.

[SCREEN SHOT shows the results of the first set]

	FIRST_NAME	TOTAL_POP	MEDIANAGE	WHITE	BLACK	AMERICAN_I	ASIAN	MEXICAN	FAMILYSIZE	COUNTY
2	Beaverton	76129	32.6	59615	1324	507	7349	6184	3.07	Washington
3	Bend	52029	34.8	48897	145	412	521	1725	2.92	Deschutes
4	Eugene	137893	33	121546	1729	1281	4916	4712	2.87	Lane
5	Gresham	90205	32.5	74619	1707	848	3007	8666	3.17	Multnomah
6	Hillsboro	70186	29.7	54391	858	577	4585	10924	3.28	Washington
7	Medford	63154	37	56834	313	677	720	4715	2.99	Jackson
8	Portland	529121	35.2	412241	35115	5587	33470	25136	3	Washington
9	Salem	136924	33.6	113746	1750	2064	3304	16636	3.1	Marion
10	Springfield	52864	32.1	47386	374	730	588	2853	3.03	Lane
11	**TOTAL**	1208505		989275	43315	12683	58460	81551		
12	**AVERAGE**		33.38888889						3.047777778	
13	**PERCENT**			81.9%	3.6%	1.0%	4.8%	6.7%		

► Compare the results that you calculated for the three different data sets.

---

[1] Technically, we might want to use a weighted average (so that cities with larger populations were 'worth more' in the results, but we will just use a straight average for the purpose of this activity.
[2] Note that the percentages will not add up to 100% because some people have more than one race and others have races not included in our subset.

6) How are they (median age, diversity, number persons living together in a single house) different? How are they similar? Based on the location of the different samples, what does this tell you about Oregon?

▶ Think about the potential bias of each sample if this group of cities was used to represent Oregon in a study knowing that the statistics for the ENTIRE state of Oregon are as follows:

        Population: 3,432,399
        Median Age: 36.3
        % White: 86.6
        % Black: 1.6
        % American Indian: 1.3
        % Asian: 3.0
        % Mexican: 6.3
        Family Size: 3.02

7) List at least two potential biases that would have to be addressed for each sample if it was to be used to represent Oregon in a research study. One of your stated biases for each sample should be based on the statistics that you calculated. The other should be based on the geography of the sample.

Turn in your responses to all seven questions and a copy of your spreadsheet.

Lesson Extension:
Find out what makes something a "city" in Oregon. What portion of Oregon's population was not included in the map? Why might this be a problem for statisticians? Why do you think that they were not included?

What about where you live? If the general area where you live was used as a sample to represent your entire state, what biases would you expect to find? Age? Income? Race? Employment?

Career Extension:
A field biologist is one of the many careers in which scientists regularly complete research projects that involve samples. If a team of biologists need to determine how many Douglas Fir trees are in a particular 10,000 acre forest, how should the researchers proceed? Give at least three ideas for samples together with their potential drawbacks.

Career Extension Option:
▶ Go to the website http://esri.com/industries.html

▶ Find a career from the list that is of interest to you. For example: In the Natural Resources list, there is a link to forestry. In the forestry link, there are several job descriptions.

▶ List four ways GIS is used in the career you choose.

▶ Conduct an internet search to find information about salary ranges and possible job locations.