



## Spatial Perspectives on Analysis for Curriculum Enhancement

### Downloading map data from the WWW, processing and classifying

The purpose of this exercise is to familiarize you with different methods of classing choroplethic map information, and to determine for yourself what the effect is of varying number of classes and type of classing schemes on the visual presentation of your data. But first we will be using the Web to access census data. Your workshop assistants will show you various Web sites where you can download population data. In the next part of the exercise we will use ArcGIS again, and focus on creating choropleth maps with the classified census data.

#### Population data access and preprocessing

The World Wide Web is a convenient place to access digital data. Unfortunately, because there is so much information out there, we want to make sure that we are retrieving reliable data collected by trust worthy sources. It is very important to take note of who manages the Web site, and to investigate the origins of the data. For this exercise we will use Web sites that publish data from the U.S. Census Bureau. If you have never explored the availability of spatial data on the Web, this is a great opportunity for you. Your workshop assistants can give you instructions on how to perform your own searches on the Web. Check out the links page off the workshop exercise web site for starting points. We will be using the following site to download data for this exercise:

<http://plue.sedac.ciesin.org/plue/ddviewer/>

**Please Note:** This site has only 1990 US census data. The 2000 data is available on the Census Factfinder site:

<http://factfinder.census.gov/servlet/BasicFactsServlet>

#### Directions for the Web site

Take note of the Web address (URL), of who is responsible for the site, and the date the data was collected. You may need to navigate around the site a bit to find this information. This is crucial source information that you must show on your final maps. Then, do the following from the main page (URL printed above).

- Launch Non-Java Edition v 2.0
- Select **California** by clicking on the map
- Select Geographic Area: Choose option **The Whole State (all counties)** from the list box (first item), then click **SUBMIT** (top of page)
- Select geographic summary level: **county** AND select a **variable** (e.g. “persons on farms”). Select any variable you want to map **EXCEPT** population or population density). If you choose a variable that is **not** standardized, also select the standardization variable with it (e.g. “persons on farms” AND “total population” to compute a ratio of farmers per total population in percent). To select multiple variables at once hold down the CNTRL key and click on variables. In the “Select Variable” box, you need to choose both your variable of interest and your standardization variable.

**Please note:** For choropleth maps you should ALWAYS use standardized data. Just for a moment think about standardization. For most cases, you can standardize by county size, and you can always download the data for county size in square kilometers. You may want to standardize by some other measure, in some cases for example number of physicians per square mile isn't logical, but number of physicians per 1000 population is logical. Data on ethnic breakdowns are conventionally standardized by total population, data on household economic measures are standardized by total number of households, and so forth.

You can ignore the rest of the options on the Web page. Scroll to the bottom of the page. Under the heading Output Format Options select **Attribute Data** → **ASCII** and click SUBMIT.

- Data and Descriptions: click on **Attribute Data**
- You should see a table with your variables...
  - 1st column - polygon identifier
  - 2nd column - state level code
  - 3rd column - area name
  - 4th column - county code
  - 5th column – variable 1
  - 6th column – variable 2 etc.

Don't forget to include the standardization variable in your selection, if you chose to map a non-standardized variable. Go to File → Save As... in your Web Browser, navigate to your directory and save this data file as a **text file**. Ok, now you can quit your Web Browser.

### **Data pre-processing in EXCEL**

Open MS EXCEL. Go to File → Open and navigate to your data text file in your directory (be sure to show files of type = All files).

A window labeled Text Import Wizard – Step 1 of 3 will pop-up. Under Original Data Type select **Fixed width**. Start import at **row 4** (header row). Click on the NEXT button. You will see your data table in a preview window with dividers you can move left and right. Check your table by scrolling up and down to make sure that the dividers do not cut any data. For example, you will want to move one of the dividers from about 50 to 55. If some of the data is still cut off, you can clean it up later. Click on the NEXT button. You want to make sure that all the data are now neatly separated in columns, and then complete process with the FINISH button. County names that have spaces and span several columns need to be put in one. Scroll down your table, and check that there are no gaps in your columns, or repetitive header information. Furthermore, we don't need all of these columns. The format we need to import into the GIS includes a minimum of 3 data columns:

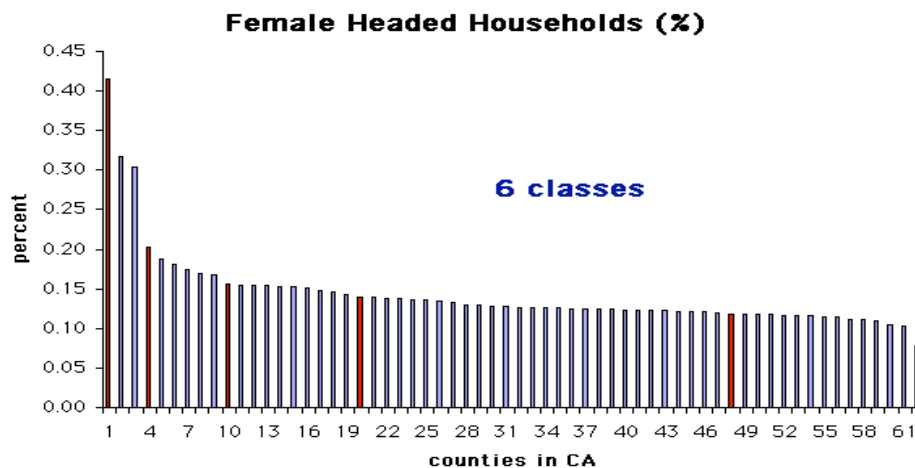
1. a polygon identifier (POLID)
2. the variable you want to classify
3. a standardizing variable i.e. polygon size in square kilometers, total population, or any other numeric standardizing information. (If your data is already standardize, e.g. percent, density, etc., you do not need to add a standardizing variable)

Make sure that the variable labels **do not** have spaces in them (keep them short, max. 8 characters)! To move columns around (e.g. switch order) use the EDIT: COPY and PASTE commands. To

delete an entire column, click on the column letter (A, B, C, D, etc.) and go to EDIT: DELETE. Keep the header row with the column labels. Delete all empty rows and other unwanted items. To delete an entire row, click on the row number, go to EDIT: DELETE (not DELETE Sheet...!). Check out an extract of a correctly formatted data file on the “Agenda” web page of the SPACE web site. Pay special attention to the bottom of the file, where the column headers repeat—delete these and make sure all the data lines up correctly.

When the file is complete go to SAVE AS... and save it as Text **Tab delimited**. Give it a new file name to represent the variable you have chosen to work with. For example, if your variable is Single Head of Households, you could call your file HHOLDS.txt. You will get a warning about Excel compatibility—click Yes to accept the warning and continue.

Ok, now let’s look at the distribution of your chosen variable. We would like to find out what classification scheme to use, and how many classes the maps should contain, to most appropriately represent our data. There are many ways to go about this, and it depends mostly on the characteristics of your data, and purpose of the map. Here are a couple of questions to guide your decision-making: How is the data distributed? Is it following a normal distribution, or is it unevenly distributed over the geographic area? Do you want to emphasize differences or clusters in the data? To answer these questions it is a good idea to visually inspect your data distribution using a graphing tool. We will look into this in more detail during next lab using ArcGIS. For now, we need to get an idea of how many classes to use. Again, there are many ways to go about this. We will apply the visual inspection method for this lab. Select the column containing your chosen variable. Go to menu DATA → Sort. In this dialog box, select the radio button on ‘Expand the selection’ and click Sort. Select the variable to be sorted from the drop down menu, and select the radio buttons ‘Descending’ and ‘Header row’ and click OK. Make a chart of this sorted list by clicking on the Chart Wizard in the toolbar (e.g. button with a little chart icon and a wand on top of it) or click Insert → Chart. Make a simple column chart. Now look at your chart, where do you see natural breaks occurring? You can resize the chart to inspect the breaks better. Here is an example...



How many breaks can you identify? Determine the number of breaks that represents your own data best, and take a note of it. Use the labeling tool in Excel and put the number you just identified on the chart (e.g. “number of classes = X”, where X is the number of your choice). You will need this information for the next part of the lab, so save it to a new Excel file. You don’t want the chart to be in the attribute file you will load into the GIS. Ok, now you can quit Excel. Save the files you created for this portion of the exercise. You will need them for next part of this exercise.

## Optional

If you want the most recent 2000 Census population data you need to navigate to the U.S. Census Web page: <http://www.census.gov/>. Click on the link “Your Gateway to Census 2000” (top of page). In the section “Data Highlights” (including a pull-down menu) select California from the pull-down menu and click “go”. Select “State by County” in the “Population, Housing Units, Area” section (second block from the top). You should see a table with the desired information. Now select the blue/white “download” button from the pull-down menu at the top of the page. Choose the “Tab delimited file” (.lst file) option. Choose to Save this file onto your ZIP disk and label it “ca\_2000.txt”.

Now you can import the 2000 data into your previously saved EXCEL file. Open the file with the 2000 data you just downloaded (e.g. “ca\_2000.txt”) with EXCEL. The import wizard will help you format the file correctly. Make sure that the radio button **Delimited** is selected, and start importing at the row number where you see the first data row in the preview window (e.g. just after the line with the COUNTY text string). Click NEXT. Check that the data looks neatly divided in the preview window. Click NEXT and FINISH importing. Now open the 1990 data file you saved earlier. You should have two tables now open in EXCEL. In the 2000 data table, select and copy just the data values in the second column (just after the labels) for all the counties. Paste this selection into column an empty columns of the 1990 data starting at row 2. Add a label for that column in row 1 (e.g. yr2000). Save the 1990 data file onto your own ZIP disk under a new file name, and quit out of EXCEL.

## Relevant Web pointers with spatial data

US Census FactFinder (download geography and attributes)

- <http://factfinder.census.gov/servlet/BasicFactsServlet>

ESRI Census Watch (census data portal and information)

- <http://www.esri.com/industries/localgov/censuswatch/datares.html>

ESRI Geography Network – Tiger Line data (download geography)

- <http://www.geographynetwork.com/data/tiger2000/>
- [http://www.geographynetwork.com/data/tiger2000/tiger\\_download.cfm?uid=40031](http://www.geographynetwork.com/data/tiger2000/tiger_download.cfm?uid=40031)

© for CSISS/SPACE, by sara f., 2004.