Lecture 10: Thematic Maps, Intro to Census Data, and the Basics of Map Making

I. Overview of Map Types
There are two different types of maps developed by geographers and within a GIS system. Do not confuse these map types with the data models previously discussed.

(1) **Reference Maps** are for general use. For example, the USGS has developed topographic map for the nation at fractional scales from 1:24,000 to 1:250,000. These reference maps show general features such as roads; hydrography; & political boundaries; etc. An example is the 7.5' quadrangle maps, which are the standard base maps useful for a broad variety of applications. These maps are periodically updated by the USGS. The Laredo East Quadrangle was last revised in 2010. Always be a little wary of base maps as geography changes in particularly fast-growing places such as Laredo. There is a certain truth in saying that a map becomes obsolete soon after it is finished.

(2) **Thematic Maps** conversely display specific geographic data like demographic information and population density. These maps are custom made for specific purposes. There are several types of thematic maps:

> Choropleth Maps - depict numerical information within areas or polygons

> Dot Density Maps - depict numerical information with dots within a polygon that represent a count of a feature. Generally, the dots are randomly distributed within the polygon

> Isopleth Maps - use lines to represent a specific value of an attribute such as temperature, precipitation, or elevation

> Graduated Circle Maps - use a point symbol but the symbols have different sizes in proportion to some quality that occurs at that point.

Basically, thematic maps are developed for a specific application like showing demographic information in the surrounding colonia communities or the distribution of black scorpions on the TAMIU campus.

All thematic maps are based on the concept of **data aggregation**. For example, let's imagine that you want to map population density in the United States. A US map with a dot for each of the

over 300,000,000 persons would be cluttered beyond recognition. **Data aggregation** allows you to organize your data based on class. Therefore, to map US population you could make a map by state with seven categories of population as indicated below.

Size Class	Population
1	< 1 million
2	1 to 3 million
3	3 to 5 million
4	5 to 8 million
5	8 to 12 million
6	12 to 20 million
7	> 20 million

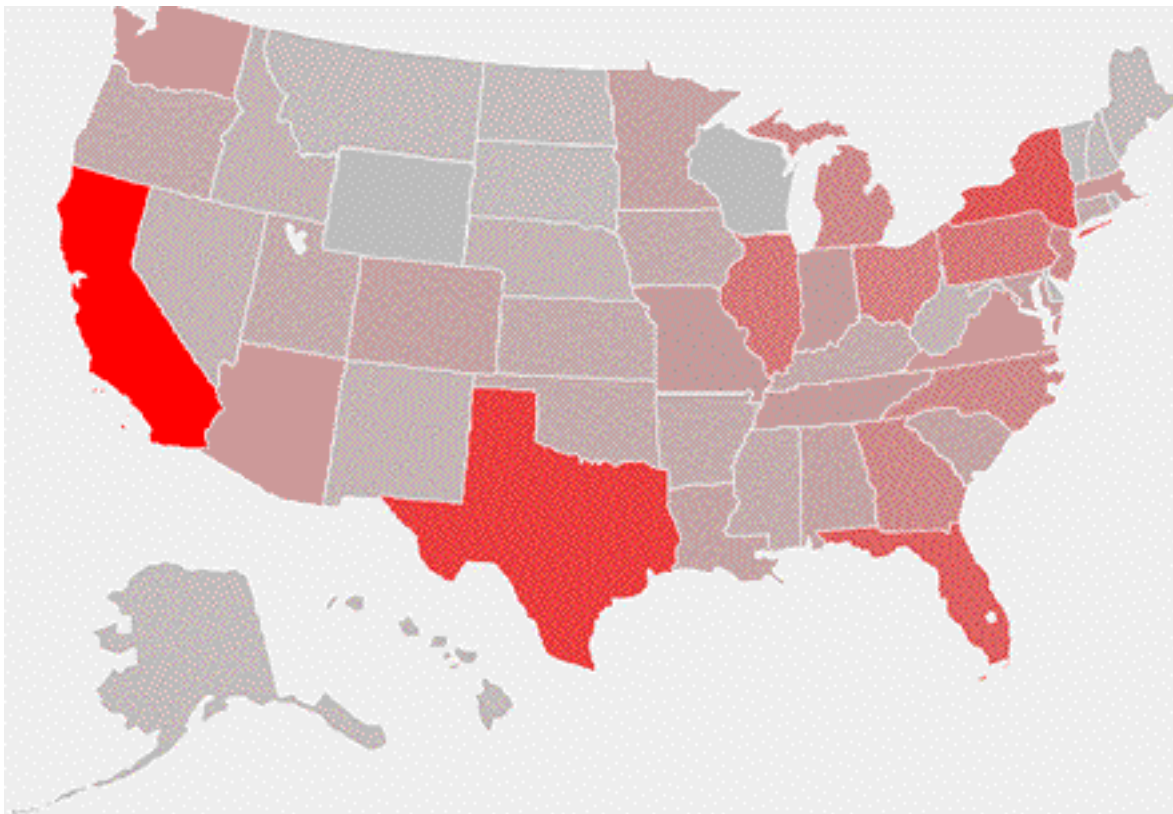For each size class you can designate a different color; as is shown in Figure 1.

Figure 1 (on previous page), Thematic map of US population aggregated based on state population. Darker red indicates greater populations.

Q: What type of thematic map is depicted by Figure 1? Answer: Choropleth

After a quick a glance of Figure 1 a user can see which states have low and high populations and can discern broad spatial trends in where people live in the country. How to make a map is important and will be discussed in the next section. Poor classification, color selection, and symbology can result in a map that is useless or in a worst-case scenario, one that is misrepresentative of the information you are attempting to depict. Figures 2 to 4 show the three other types of thematic maps.
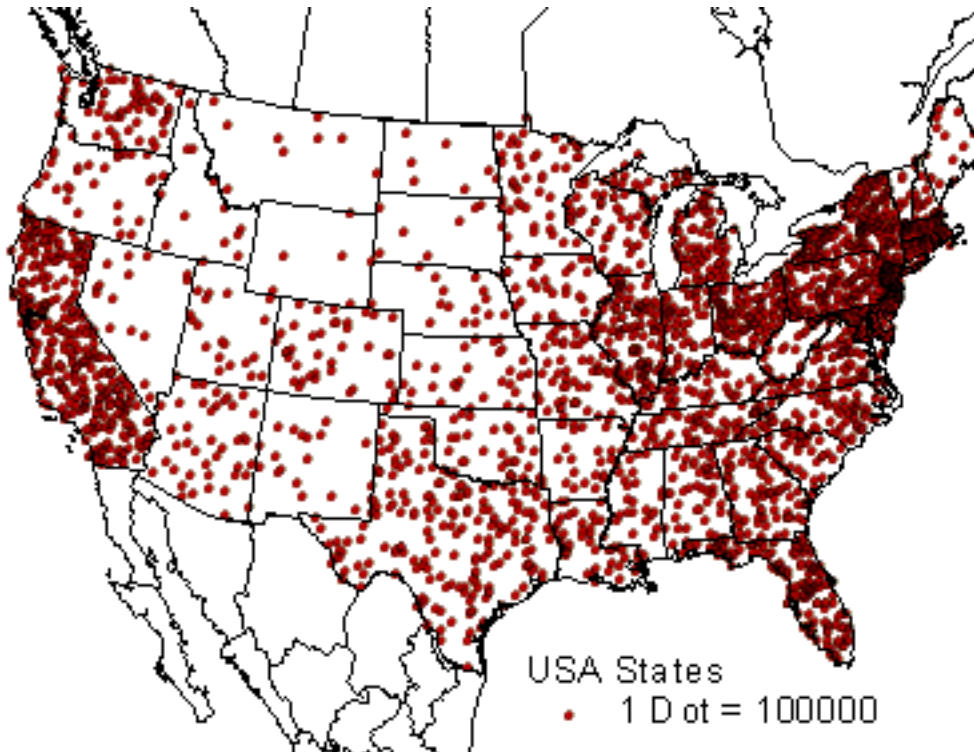

Figure 2 is a dot density map of US population. Each dot represents 100,000 people.
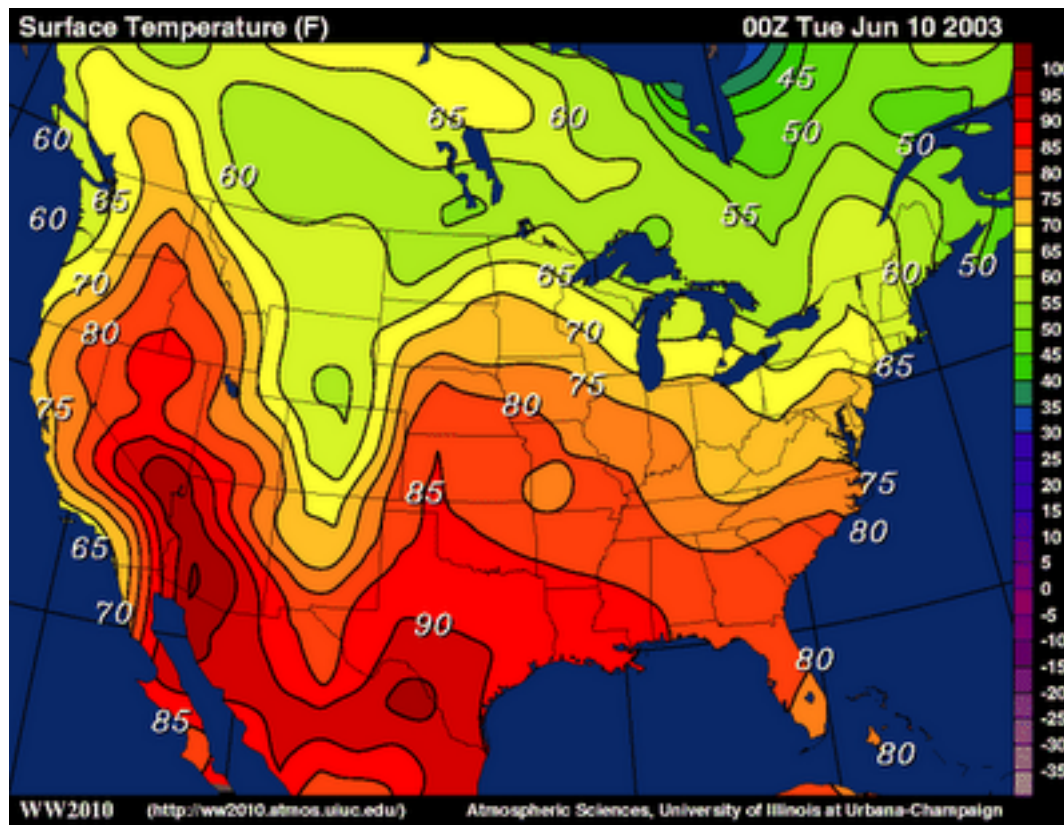
Figure 3 is an isopleth map that shows surface temperature in °F across North America.

II. Making Thematic Maps

Maps that you are likely to make in association with your research are likely to be thematic in nature. There are many options. For example, thematic maps can be developed using ratio level data including counts, rates, and densities. Additionally, ratio level data can be combined into ordinal level classes through aggregation especially when there are numerous numeric attribute values.

When creating a thematic map a geographer (and GIS user) can potentially misrepresent ratio data during the reclassification to ordinal classes. Let 's imagine that you want to make the impression that the counties in South Texas are not growing rapidly. You could select a classification scheme that will include only a few counties in this region; therefore, when an individual glances at your map you will have conveyed the mistaken impression that South Texas is not a rapidly growing region as a whole. This is how GIS professionals can **lie** with maps spinning the facts to suit a predetermined conclusion. Check out the book How to Lie with Maps! GIS is like most professions in that there is in theory a code of ethics that governs behavior, which regrettably is not always adhered to. So, do not take any thematic map at face

value; you should always ground-truth these maps in your mind by asking the question does the map make sense with what you know about the world.
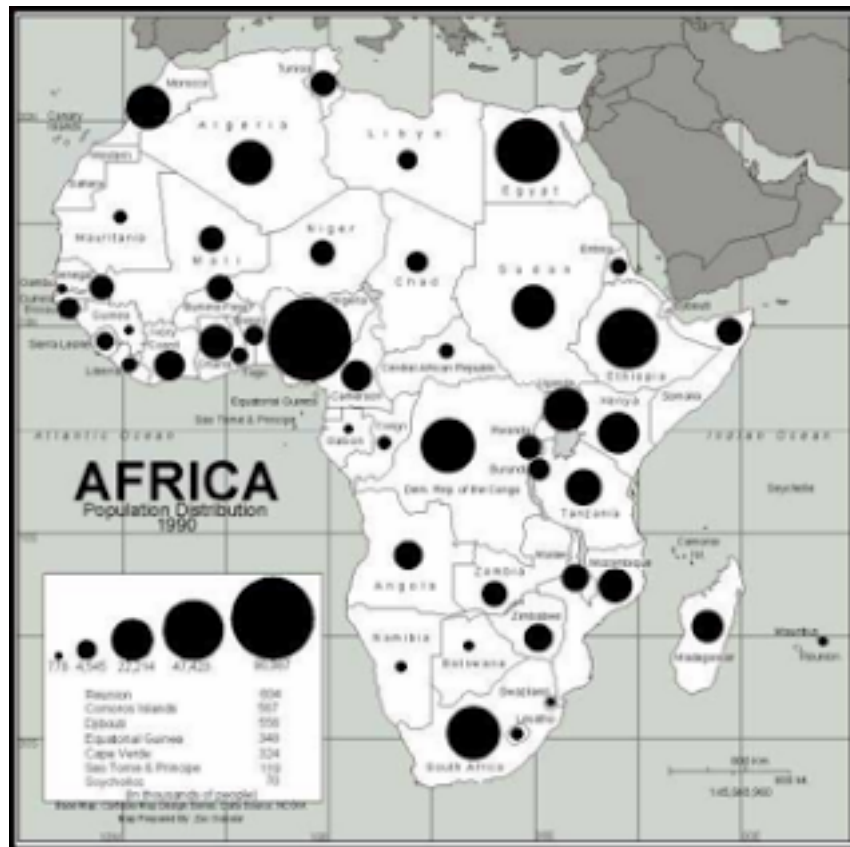

Figure 4 is a graduated circle map that shows population by country on the African continent.

When developing a map there are numerous decisions that need to be made. Probably the most important of these is the scale, size, and shape of the map you are going to make. Generally, map size is constrained to the size of the paper that you are going to use to print out the report in which your map is embedded, which is generally, 8.5" X 11" or smaller if you have margins. Your map should tightly focus only on your area of interest. Additionally, you want to have a map that has a fractional scale that is a nice even number, for example, 1:2,000. This makes it easy for an end user to use a ruler and determine real world distances from your map.

The next significant decision you are going to make is what data you are going to plot on your map. Think of your map as a story. You want to plot data that will tell the end user an interesting story. Therefore, data that is peripheral to the main message your map is attempting to convey should not be included.

Classification Methods

You have several options in terms of how you aggregate your data. Within ArcGIS Pro several classification methods that can be used to transform using ratio level data into a small number of data classes. These classifications schemes include equal interval, percentile (quintile), natural breaks, and user defined. The simplest method is natural breaks and is based on low points in the frequency distribution of your data set (Figure 5).

Read the following web page that quickly describes ArcGIS classification methods: http://pro.arcgis.com/en/pro-app/help/mapping/symbols-and-styles/data-classification-methods.htm

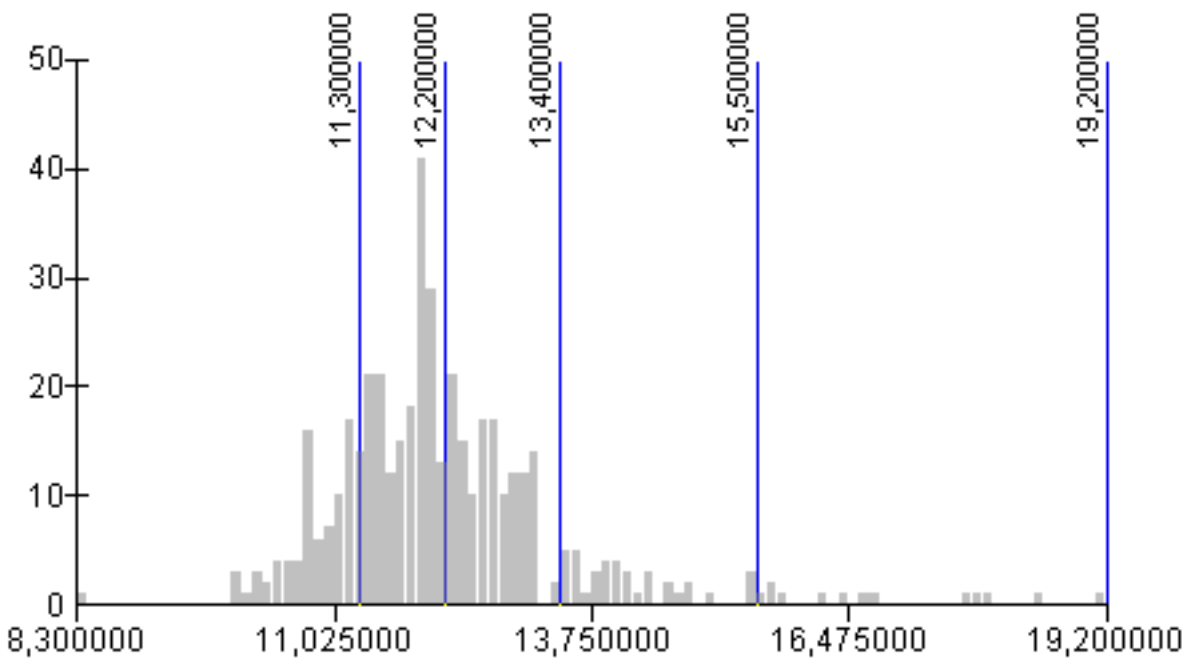**Natural Breaks** Classification

Figure 5 - Classification of data into six classes based on low points, or breaks, in the histogram of attribute values.

The second classification method is referred to as an **equal interval** classification. Continuing with the example from the previous page an equal interval classification is based on the overall range of attribute values divided by the number of classes. So (19.2000 - 8.3000) / 6 indicating that each class is 1.8167 - an equal interval; thus, the resulting equal area classification is as follows:

Class	Low Values	High Value
1	8.3000	10.1167
2	10.1167	11.9333
3	11.9333	13.7500
4	13.7500	15.5667
5	15.5667	17.3833
6	17.3833	19.2000

Note that a potential problem with this approach is the clumping on data into classes 2 and 3, which may not result in a realistic thematic depiction of the attribute.

A third classification approach is based on a **percentile (or quintile)** classification. This approach splits your attributes into classes based on ranking. Consider an assumption that the population for each of twelve counties, A to L, in State X, is to be distributed into 4 classes using the quantile method. The table below demonstrates how the distribution is achieved. Each county is ranked according to population count.

County Population Rankings	Class	Values Range
A: 6750 C: 7232 D: 7247	Class # 1	6750 - 7247
B: 8697 G: 9654 L: 9786	Class # 2	8697 - 9786
H: 10520 F: 10673 E: 11900	Class # 3	10520 - 11900
J: 12101 I: 12991 K: 13443	Class # 4	12101 - 13443

The 3 counties with the lowest population counts are assigned to Class 1; the 3 counties with the next lowest population counts are assigned to Class 2, and so on. The gaps in a quantile map are intentional and mean that there are no values between the top level of the previous class and the bottom level of the next class.

And finally, you can always make up your own classification scheme; but, if someone ask why you classified the data this way you may have difficulties in your explanation. Which method is better - the method that produces the most realistic depiction of your attributes across the

mapped area. GIS users have great power and if they are not careful can potentially misrepresent ratio data during the reclassification to ordinal classes.

Some questions to ponder What problem is there with your thematic map is you select too low of a number of classes? Too high a number of classes? Generally, five to seven is a good number. You do not want to select more than eleven classes as you will make your map too busy.

Symbology, Color & Labels - Once you have decided what data to include the next decision involves determining a symbology that will maximize your message. Check out the following website by googling (Tutorial 3-Map Symbology in ArcGIS). This pdf document shows different possible symbologies that can be applied to point, line, and polygon data. Additionally, color is important and almost always the default color selected by a GIS program needs to be changed into a color scheme that more clearly shows differences in the data layers included in your map. Your symbology selected needs to be clearly indicated in your map legend.

Another consideration is the type of labels you applied in your map. Like with color the default label type and placement is almost never optimum. The main point about labels is that you want to avoid excessive cluttering of your map that will detract from its readability. In many cases, especially when dealing with thematic maps, labels can be omitted.

Other Map elements - Other elements besides the actual map that add value include a detailed legend, north arrow, scale bar, border and neatlines, credits, and grid, which are illustrated in the following pdf (google: mapdesign.pdf)

The position of all of these elements can either enhance or detract from the overall readability of a map. Figure 5 is an example of a good map design whereas Figure 6 shows a poor map design. Note that there is little wasted space in the map on Figure 5. All geographic entities are clearly labeled. Labels are placed in a way that does not clutter. The colors selected serve to highlight the issues present in this nation. Map symbols are clear. Graphic scale is included.

Figure 6 is an example of poor map design. The colors scheme selected is awful and does not provide contrast to highlight the different attractions of this park. There's no scale to indicate distance. Additionally, no map legend is provided. The elements of this map are highly cluttered and the labeling is too dense with text description. This makes it difficult for the user to find anything within the map.
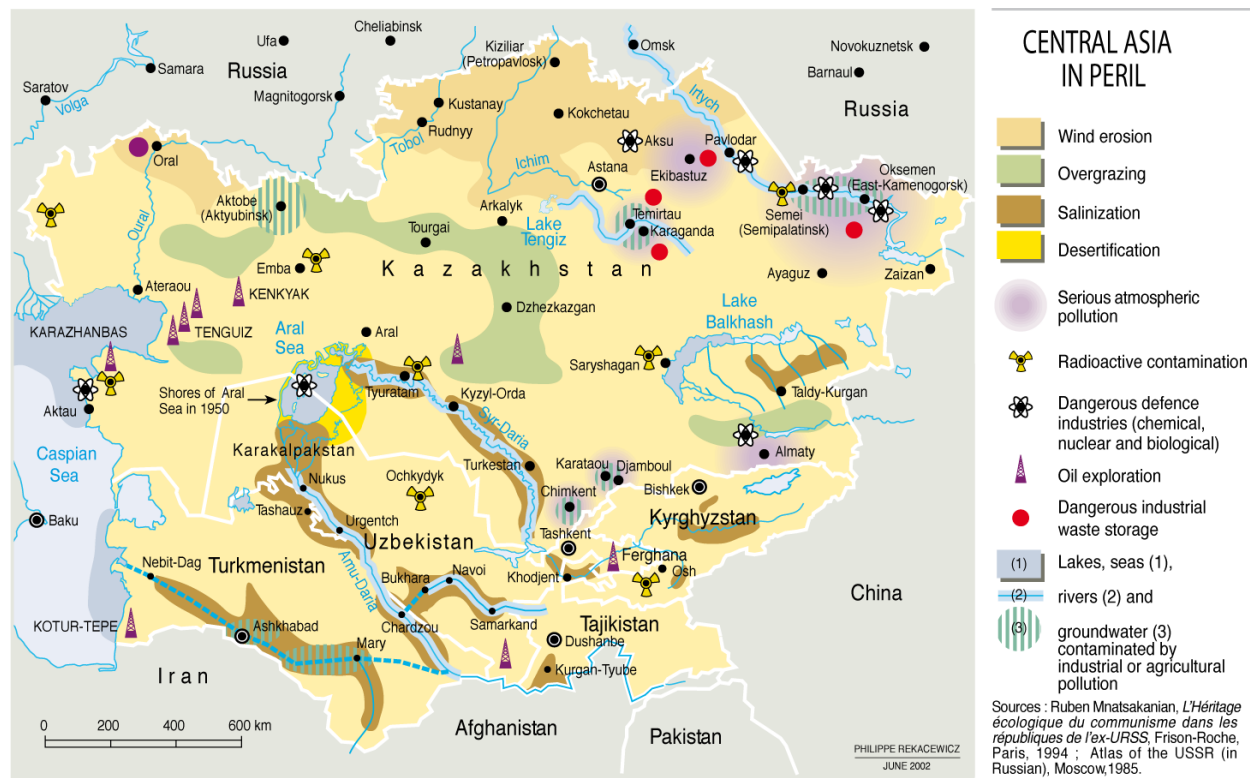
Figure 5. An example of good map design.

III. US Census Enumeration Units

Since assignment 4 will be based on making a choropleth maps of Laredo population a brief introduction to the US Census is in order. As mandated by the US Constitution every ten years the federal government is directed to develop a tabulation of all individuals residing in the country. A task as seemingly mundane as counting the populace can become a political hand grenade, as is the case right now in the State of Texas, as congressional districts will be redrawn. States can lose or gain seats in congress depending upon whether they are losing or gaining population (obviously Texas is gaining seats -four) and public services can be affected because allocations are based on the Census tallies. There is a strong motivation to craft districts to enhance representation of the party in power. This unethical tinkering of maps is referred to as gerrymandering. An ethical GIS professional would never engage in such a practice.

The census is important. A few years back report was issued by the Texas A&M Colonias Program blasting the recent Census effort (http://www.civilrights.org/census/gulf-coast-report.html). The most recent U.S. Census (2010) failed to send tabulation forms to 90% of the residences in these remote communities.

Additionally, many census workers that were deployed to the colonias were outsiders not fully trusted by colonia residents. Because of these difficulties in obtaining a total count of colonia residents non-governmental organizations estimate that there was a massive undercount of individuals in these communities probably on the order of 10,000's individuals. And with recent political developments this situation is only going to get worst.

Figure 6. An example of a very ugly map layout.

The spatial hierarchy of US census data, which defines the areas over which data can be aggregated is shown in Figure 7 and is summarized below.

State   Biggest
     County
          Census Tract
               Block Group
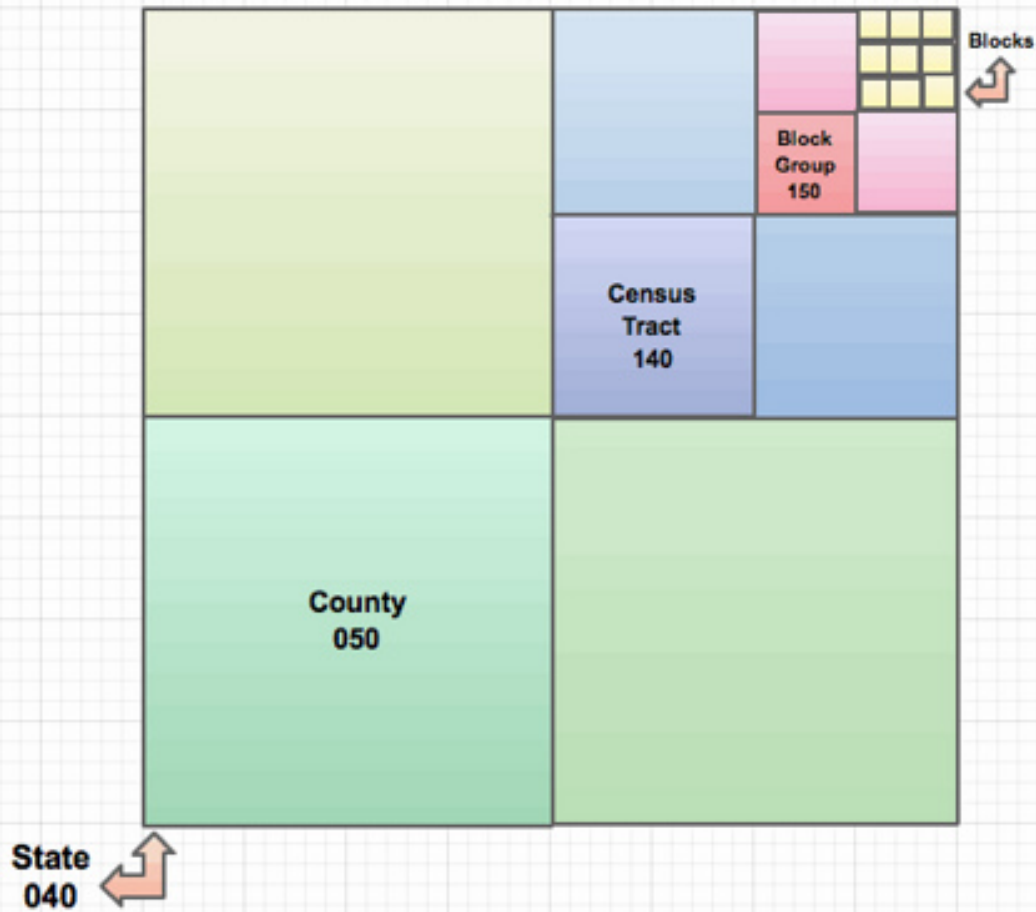   Block   Smallest

Figure 7. Hierarchy of US Census tabulation units.

Readings
GIS Commons webpage; Chapters 2.
DiBiase, D., 2014, Nature of Geographic Information Systems. Sections 3.2 to 3.3; 3.12 to 3.19.
Campbell, J and Shin, M., 2011, Essentials of Geographic Information Systems. Chp. 9.

Terms

Reference Map	Thematic Map	Data Aggregation	Choropleth Map
Dot Density Map	Isopleth Map	Graduated Circle Map	Equal Interval
Classification Method	Natural Breaks	Quintile	County
Census Track	Block Group	Block	

Concepts

Examine section 3.18 from DiBiase. From reading this section can you understand how it is possible to lie with a map

Know about the different type of thematic maps

Know about the different classification methods that can be used to aggregate data

What is the hierarchy of geographic units within the US Census

Make a list of best practices that should be followed when making a map. Use these the next time you make a map. If you don't you will lose points.

HOMWEWORK

1. Examine the dataset on the last page. Using draw lines that would divide the data using the equal interval, quintile, and natural breaks classification schemes. Following the instructions for each classification scheme given below. You should have five classes for all three records. Make three copies of the last page. Label each page equal interval, quintile, and natural breaks depending on the classification schemes used on that page.

Equal Interval Classification:

1. Determine the overall range (R) in your data

2. Determine the range within each class (r) $r = R \ / \ n$

      where n is equal to the total number of records

3. Draw lines to reflect this classification on the pre-sorted data set on the first page that should be labeled Equal Interval.

Percentile (Quintile) Classification:

1. Determine the total number of records that the data set consists of (n)

2. Determine the number of classes (NC) to be used to classify the data. In this case five.

3. Determine the number of records per class (RC). $RC = n \ / \ NC$

4. Draw lines to reflect this classification on to the pre-sorted data set on the second page that should be labeled Quintile.

.

Natural Breaks Classification:

1. Scan the pre-sorted data set and draw five lines in places that seem to have gaps in the data.

2. The number of record in your classes does not need to be equal unlike the two other classification schemes.

3. Draw lines to reflect this classification on to the pre-sorted data set on the third page that should be labeled Natural Breaks.

<u>Data Set</u>: Annual Population Growth of Selected Counties in % in South Texas (1990-2000)

Label: _____

-1.1
-1.0
-0.8
-0.7
-0.7
-0.5
-0.4
-0.3
-0.2
-0.1
0
0
0
0.1
0.2
0.3
0.4
0.4
0.5
0.5
0.5
0.6
0.9
0.9
0.9
0.9
1.0
1.1
1.2
1.3
1.3
1.3
1.3
1.4
1.5
1.6
2.0
2.5
3.2
3.3