Teaching Redistricting During a Census Rollout

William M. Harris
Department of Environmental Science and Studies
University of St. Thomas – Houston
wharris@stthom.edu

Jean-Philippe Faletta
Department of Political Science
University of St. Thomas – Houston
falettj@stthom.edu

Abstract

The departments of political science and environmental science and studies co-offered a course in redistricting in the spring of 2011, as the 2010 Census data was being distributed. The course was a combination political science and GIS course where students learned to use GIS and the redistricting extension, were taught the rules and political realities of redistricting (gerrymandering), and were able to use the 2010 Census data to make their own redistricting maps for the state of Texas. The course also proved to be a challenge for the professors when the Texas data sets came out and exceeded the million-line limit of most available software to parse out the pertinent lines of data. While challenging, the course turned out to be an excellent example of how GIS can be used 'on the fly' to enhance learning in a variety of disciplines.

Paper Prepared for Delivery at the 2012 Annual ESRI International User Conference, San Diego, CA: July 21 – 27, 2012.

History of Cooperation

In the fall of 1992, the department of environmental studies began offering BA and BS degrees in environmental studies. Students within the major spent two "common" years taking the same coursework including biology, chemistry and environmental science introductory sequences. They also, beginning in 1994 with the completion of the Robertson science building, began taking required courses in GIS.

The major has always been interdisciplinary with courses in history, biology and chemistry for the majors. For the Bachelor of Science students, the course sequence was heavy in ecology, field studies and culminated in either an internship or thesis. For the Bachelor of Arts, students took more courses within their majors, and they also completed a 12- to 13-hour sequence within a concentration area.

The initial concentration areas were communications, economics, international studies, political science and psychology. The reasoning behind the wide range of concentration areas was that to change people's behaviors, they needed to hear what was going wrong, understand that we can do better environmentally AND economically, cooperate with other countries, work within the political system at least as well as the corporations do and change people's hearts and minds. Each of the subject areas complimented the course offerings and eventually, such as with political science, we also developed joint majors. Although the department has gone through some "reinventing" through the years, we continue working closely with communications, international studies, political science and psychology.

As with many environmental programs, we suffered a "loss of faith" in the mid-2000's and a reevaluation of the program and its goals, a kind of sabbatical. At that time environmental studies was working with the political science department on the establishment of a minor in urban planning. With the expansion of environmental laws, many municipalities were finding it harder and harder to operate within the law using only volunteer and elected individuals. The urban studies initiative was designed to give students the option of learning about the functioning of urban systems (both large and small) and the impact of dysfunctional infrastructure on the environment. Unfortunately the University decided to suspend the environmental studies major at that time.

In an attempt to salvage some of the courses that had been offered, the faculty of the environmental studies program continued working with the faculty of the political science department on the urban studies initiative. We contacted members of several other departments in hopes of broadening the scope of the urban studies minor with the hopes of eventually including it in the Masters of Liberal Arts degree program. In 2008 the department was revived as the environmental science and studies program with both BA and BS degrees, which has expanded to three program tracks in the BS area and two in the BA area. The political science department has taken part of the urban studies initiative and included it in their MLA concentration area in public administration.

The cooperation between the departments (environmental studies and political science) had been cemented, though, through the development work for the urban studies minor and that cooperation continues today with the cross-listing of courses and an interest on the part of political science in GIS.

In the fall of 2010, Dr. Harris approached Drs. Faletta and Taylor (chair of political science) about offering a course on using GIS for redistricting in the spring semester of 2011, when the census data would be rolled out. This course would combine the instruction on the contentious nature of Texas redistricting (which almost always makes it to the US Supreme Court), the nature of how apportionment affects representation and how to manipulate the census data in the process of redistricting. This opportunity only comes around every decade and so this course would be as "cutting edge" in terms of using the raw data as you could get. The students in the class would be constructing their new district boundaries at the same time as the Texas legislature would be working on their maps.

As a continuation of the cooperation, the departments will be offering a course in the fall of 2012 on the use of GIS in elections. In addition to the use of the census data files, we will be using the community survey data to look at voting districts to determine how they might vote in the fall elections. Students will make projections based upon their analysis of demographic data in the same way that candidates and their campaigns will. Part of the course will include looking at the campaign results and trying to see if their projections were indeed correct, or if additional factors came into play.

2010 Census

The book used for the redistricting course was Charles S. Bullock III's *Redistricting: The Most Political Activity in America*[1]. The book looked at the redistricting process from the original apportionment through the activities of the 2010 Census (but without the 2010 census data, because the book was published before the data was released).

The US Census is mandated by Article I, Section 2 of the Constitution and attempts to count every individual (regardless of status) in the United States.

> "Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers . . . . The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct."[2]

The data produced by the census goes into the apportionment of the representatives in the US House of Representatives, and additionally factors into how federal money is distributed (billions of dollars) to local communities.

The first census of the United States was conducted in 1970, shortly after George Washington became president of the country. The data was collected and the results were made public by posting them "at two of the most public places". Access elsewhere was difficult, at best. This data included information about free white males under and over the age of 16, free white females and other free persons, and slaves. The total number of people in the country at that time was 3.9 million. This number caused an immediate increase in the number of representatives in the House from 65 to 105.[3]

In addition to the number of individuals counted in the decennial census, the Census Bureau collects a number of surveys (for the government) and an annual sampling of data from the population referred to as the "American Community Survey" (1996 to present). The census data cannot be reported to the government and/or public in any form where

the results can be traced back to any one individual, which means that the data as it is released is released in aggregate form.

Initially the Census Bureau asked for a limited amount of information. Since its inception, though, the amount of information requested has changed to include education, marital status, occupation, place of birth, income, language and Spanish/Hispanic origin. The first 100 years of data collection was very low-tech, with clerks who hand tabulated census results with pen or pencil. This meant that the census data was released years after the data was gathered, and guaranteed that the census data itself was obsolete. From 1880 to 1950, the Census Bureau used a "tabulating machine" to speed up the data using punch cards. Even with this advance in technology, some elements of the census were released at much as two years after the census was conducted.[4]

In 1950 the Census Bureau purchased a UNIVAC I, to improve the turn around time in reporting data. In 1980, the data from the individual computers at the processing centers were able to be transmitted electronically to the headquarters for final compilation.

As with any other kind of information, nothing is useful unless you can access the data. The initial results could only be utilized if you went to the location of a public-access or had a hard copy of the data and paged through reams of paper. In the mid-1920's data could be produced in punch card form and in the 1960 data could be accessed on computer tape. In the 1980's, census data was made available on diskettes, CD-ROM and laser disks for use in microcomputers.[5] Today, census data is available through the Census Bureau's website using the American Fact Finder and as direct downloads from their ftp site.[6]

In the 1990's, the Census Bureau began working with the U.S. Geological Survey in developing an electronic data base (called the TIGER system) to allow for the attachment of

census data to map files. Today, you can download a wide variety of shape files from the Census Bureau website and then link those geographic objects to census data as it is released to the public.

Today, the Census Bureau collects population and housing data every ten years as the Decennial Census, conducts an Economic Census every five years (as well as a Census of Governments), and the American Community Survey annually.

To political scientists, politicians and the state and federal governments, the greatest impact of the Decennial Census is related to reapportionment of the members of the House of Representatives. States can gain or lose representation in the House depending on the changes in their population relative to other states. For some time, the Bureau of Census collected their data and then "increased their reports on the population" by assuming an underreporting of individuals. They saw their data as being a sample from a larger population and proportionally increased their population estimates using a statistical technique. Because of a challenge mounted by "interested individuals" in the courts, the Census Bureau eventually had to end that practice and go back to reporting only the data as it was collected.[7]

The results of the 2010 Decennial Census were released in the spring of 2011. The information was released as sets of data, released in groups of states. The first reporting of the data was in raw form with information provided in data sets that could be opened with a variety of software (some programs were proprietary to governments while other people used Excel, Access, and some other database formats). While there were files available to help in the process of joining the data to the shapefiles, this information was not "user friendly" to the casual user. To link the shapefiles to the data, one had to conduct several

joins and know how to concatenate the appropriate fields within the data. Also, the column headers were listed in "governmentese" rather than as logical descriptions, making the use of the data difficult. One might view this release of data as a kind of "data vomit" out to the public. The data for the state of Texas was released on February 17, 2011.

Why Redistricting is an Important Area to Study (Especially in Texas)

As discussed in the previous section the decennial census translates into the reapportionment of congressional seats necessitating the redistricting or redrawing of district lines. The redistricting process is quite contentious nationwide and has been ever more so in the State of Texas over the last three cycles (in fact, following the 2000 census, there was a second mid-decade redistricting in 2003 after the majority Republicans believed that they did not receive results that were equal to their numbers in the state following the 2002 midterm elections). In fact, how votes translate into representation depends upon the districting format. In the state of Texas the redistricting process is controlled by the state legislature, as is the case in most states.

In the spring following the census and following legislative and public hearings, the allowance of individual legislators, groups, or private individuals submitting their own maps[8], the legislature will adopt new maps for congressional districts, state house and state senate districts, and State Board of Education districts; the maps are then sent to the governor for approval. If the governor and the legislature cannot agree, a body, known as the Legislative Redistricting Board (LRB) has to meet. The governor may also call a special session to deal with the matter and in recent redistricting cycles, the issue has been addressed by federal and state courts. Approved maps must also be subjected to federal

preclearance as mandated by Section 5 of the Voting Rights Act of 1965. The new maps are used for the following primary and general elections.[9]

Redistricting is a controversial process with the charge being levied that elected officials use the exercise to pick their voters as opposed to voters selecting their elected officials. In Texas redistricting is a highly partisan process with the party that controls the legislature usually controlling the process and in most cases the outcome. Partisanship in drawing the new maps is a bipartisan affair and has profound effects on the final outcome depending on party control not only over the legislature but the governor's office as well. In fact, one may end up with the party that receives fewer votes can still win a majority of the seats (e.g. Texas following the 1990 census).

The outcome of each redistricting cycle is also important for other reasons, especially in a state such as Texas, now the second largest state in the Union in terms of population. For example, who wins the most seats determine what interests will be represented in public policy decisions and whose priorities and values will be reflected. Much depends on whether a state is gaining or losing seats in the reapportionment where the party controlling redistricting has much to gain if gaining seats, if not, they just want to hold their own (following the 2010 Decennial Census, Texas gained four seats). Redistricting decisions impact public policy; it impacts the number of minorities in the legislature and their subsequent ascension to leadership positions, and the creation of majority-minority districts, districts where the majority of the population/voters are minorities.[10]

The impact of using partisan redistricting schemes is greatest in places such as the United States that uses single-member districts and is especially profound in the state of Texas as compared with most other states. In fact, the two professors (Harris and Faletta)

for the spring 2011 redistricting course jokingly remarked before the class began that the process should just begin in the courts as opposed to the legislature since we already knew that the final product would end up the courts (which it did). In fact, the majority party payoff is exacerbated by the single-member district, winner take all system that the United States utilizes. For example, political scientists, Andrew Gelman and Gary King (creators of the JudgeIt II software) argue that the focus regarding redistricting plans should be on bias and responsiveness.[11] They also contend that to be free of bias the proportion of the seats that a party receives (responsiveness) should be comparable to the percent of the vote they receive.

In Texas most legislative districts, especially at the congressional level, are not competitive and are safe for one party or the other (the real competition takes place in the primary if there is one). In fact, an argument has been made that our Congress is approximating a type of House of Lords and some political commentators have postulated that if all redistricting was taken out of partisan hands and put under the control of non-partisan bodies the competitiveness of districts would increase and the quality of political representation would also increase.[12]

Regardless of which process is used, there are two cardinal rules in redistricting: there must be population equality in districts and minorities must not be discriminated against in the drawing of new maps. In Texas, in addition to the partisan angle to the redistricting process, the issue of minority representation has been the other major issue, especially in the 2010-2011 cycle where groups such as the Mexican American Legal Defense and Education Fund charged that the four seat congressional gain in the size of the Texas delegation was in part due to the growth of the Hispanic population over the

previous decade and as a result there should be a commensurate increase in Hispanic representation.[13] It was against this backdrop and background that the redistricting course began at the end of January 2011.

Spring Opportunity

At a faculty social occasion housed at the political science house on campus, Dr. Harris approached the faculty of political science asking if anyone would like to co-teach a course on how to use GIS in the process of redistricting. The process and practice of redistricting had been a subject of conversation and interest between the coauthors for some time. In Texas, redistricting has always been contentious at best, and impossible at the worst. There are many jokes about Texas redistricting, all of which reference to the fact that proposals are almost always dead-on-arrival in the Senate and House and once the maps are finally released, they might as well have been printed at a courthouse. Final judgment of district boundaries must obey, or at least seem to obey legal requirements. Quite often, the courts have judged that Texas has a "special problem" with offering redistricting maps that allow for equal representation by the states ethnic groups.

In the spring of 2011, the University of St. Thomas held a town hall meeting where the city redistricting process (The city of Houston, Texas added seats in the city government) was discussed. Dr. Harris was asked to discuss the data rollout and the Voting Rights Act of 1964 at this meeting by councilman and former police chief Bradford. The seriousness of the census data rollout was underscored when one of the meeting participants accused a political party of getting "early" information from the Census Bureau and that some political elements had "already made their maps!" The reality was, of course, that no one had an early release of the census data but that some had used the

Community Survey data to project where the growth may have occurred and were testing possible scenarios in terms of where new city council district lines would be established. Dr. Harris also presented information about the release of data and the voting history of the minority-majority districts for the city, stressing that no increase of minority population would overcome a tradition of NOT participating in the political process. Texas is a state covered by the Voting Rights Act of 1964, and as such is monitored during the redistricting process to ensure minority voting rights are protected and that districts are not manipulated to decrease minority representation.

Dr. Harris has taught GIS courses at the University of St. Thomas since 1994 and had been using ESRI software in interpreting environmental data. As part of the environmental science and studies department's outreach program, the department had been recruiting students from other departments for the "Introduction to GIS" course since the late 1990's (even changing the title from "Environmental GIS" to simply "GIS"). Since the 2000's, Drs. Faletta and Taylor (political science) had indicated an interest in learning GIS, or at least having their students learn GIS. The release of the 2010 data provided a "perfect storm" of interest and the need for a class. With the release of census data being imminent there wasn't enough time to apply for the addition of a new GIS class, so the course in redistricting and GIS was taught as a "special topics" class.

The student population for the class was varied. Some of the students in the course had taken the Introduction to GIS course offered by the environmental science and studies department. The department is the only one to teach and use GIS on the campus, so the introductory course is open to all students and majors. The class size is relatively small, so there wasn't a large cohort of students who would have prior GIS knowledge for the

redistricting course. One student in the redistricting class had taken an additional, advanced course in GIS. Other students had expressed interest in GIS, but were not able to take the class because of other degree requirements or time conflicts. Many of those students looked upon the redistricting course as giving them a change to begin learning bout GIS within the context of a course that would fulfill their degree requirements in political science.

A final group of students were those who were interested in the partisan process of redistricting. In Texas the process of redistricting is a spectator sport for a subset of the population. Those students had varied interest in GIS, but understood that everyone in the class would be able to make basic redistricting maps for their class final project.

In 1994, the University of St. Thomas completed the construction of the Robertson science building. During the planning phase, the environmental studies department requested a GIS lab with five stations for exclusive student use. There were six computer stations created in the teaching laboratory (geology and environmental science) for use by students when laboratories were not in session. At the time the building was constructed, this was not only the largest student-use computer installment, but we also were the only department with an intranet and internal e-mail. The University has always been supportive of the department and the use of GIS with these computers. For most of the last 18 years, the school has purchased new computers with enhanced graphics capabilities for the department and several years ago we were allowed to purchase a 42-inch printer to make our maps. We also had installed the redistricting extension from ESRI on all of the computers when it became available.[14]

Students and GIS

As stated earlier, one of the difficulties with teaching the course was the variety of levels of familiarity of the students in the use of GIS. Two of the students had taken an introductory GIS course and were comfortable with the use of the software package. One of those students had also taken a second course in GIS entitled Advanced GIS and was able to serve as a student-mentor in GIS for the class. That student was given data and information before the other students and asked to attempt the assignments to make sure that they were possible for the other students. Tasks that were difficult for the advanced student were modified to make them simpler for the other students.

Most of the students in the class had not used any GIS software. The course was listed as an upper division and graduate level political science course because of the content and we were not able to dictate a particular level of familiarity of any student before the course. One of the students had been engaged in a political campaign and wanted to be able to work with the program in the re-election of their candidate and needed to know how to use GIS software and hadn't been able to take the GIS course. This particular student was very "gung-ho" during the course and wanted instruction beyond the others. He had to be "reigned in" at times because his questions were beyond the position of the rest of the class and lead to a bit of confusion in the case of some students who were not at the same level and who were having problems with the use of the software.

With all of the different levels of student knowledge in GIS, and because this wasn't a GIS course, we had a hard time deciding how to give the students an introduction to the software and also give them enough instruction in GIS to be able to use the redistricting

extension. Dr. Harris had been teaching GIS for many years before the redistricting course and had developed a concern about the perception of GIS instruction by the profession. During the semester that the course was offered, Dr. Harris had proposed a presentation for the ESRI 2011 Education User Conference in which he proposed that GIS isn't a single "thing," but a variety of careers. In the presentation the question was asked "What is GIS? Computer Science, Geography or Useful Tool?"[15] The answer given in the paper was "yes." Because it is actually several things, depending upon need. It is computer science in the way that it can require programming expertise with Python and other scripting available and the development of "enterprise" GIS editions. GIS is also geography and map-making. You need to understand how to make an accurate map and the proper use of map elements. At the same time, you don't have to be able to make maps of the level of the type of National Geographic to have made a useful map. Finally, within the context of GIS as a tool, the software allows people who have spatial data available to manipulate that data and display the results. Computer scientists and geographers may not understand the use of the particular data, making those using GIS as a "tool" probably the largest group of individuals using the software, albeit they may be making some of their maps in violation of geographic standards.

In the end, the professors determined that the ability to become a professional in the use of GIS could not be the main focus of the course. The history of redistricting, the 1964 Voting Rights Act, and the process of redistricting in the modern era needed to be the prime focus of the course. Dr. Harris provided a short-version of the Introduction to GIS class at the beginning of the semester with PowerPoints and assignments using a book titled "the GIS 20 Essential Skills" by Gina Clemmer (2010).[16] This book includes an

assignment which requires students to download Census data from the Census Bureau website and use that data to make several maps. This is similar to the process that the students needed to complete to obtain their data for their redistricting assignments. In addition to the book, the faculty provided the students with a wide variety of handouts and data sets to use to practice using the GIS software.

After the intensive instruction in GIS during the first several weeks of the course, the students were given a copy of a population dataset for the United States and a copy of a map showing the distribution of people by state for the country. Without further instruction, the students were asked to use the software to replicate that map in exact detail with only the data and shapefiles provided. The students were graded on this project and those who had difficulties producing the required map were given more instruction until they had developed that level of proficiency. Upon the successful completion of the assignment (a 100% correct map was the only grade available) the students were considered to be minimally proficient in the use of GIS for the course.

Problems

What had at first seemed easy ended up far from being so. Unfortunately we had already updated the computers to Install 10.0 of ArcMap on all of the computers. The version of the software that the free districting extension worked on was 9.3. Because of this, we had to uninstall 10.0 in favor of 9.3 on half the computers in the lab. This created a problem at times because three different groups of students needed the computers, most of them needing version 10.0. Also, the districting extension taxed the limits of the computers in the lab. The Texas dataset was very large and we were unable to open the entire Texas dataset and use the districting extension on the computers that were available.

As for the districting extension, the positive side was that the extension was free. For the negative side, there was no documentation for the program. This meant that you had to understand the nature of districting using the GIS extension provided by ESRI. Having had experience with pre-prepared data for districting, it was initially almost impossible to decode what was being asked by the extension and without a key for the government column headings. To use the extension, we had to find blogs, websites and posted exercises provided by the user community.

Another problem that we had using the program was a problem with the data as it was released in raw form by the Census Bureau. While smaller states' (in both size and population) raw data files were easily opened using programs such as Microsoft's Access and Excel, the raw data file size for Texas was over one million lines long. Although the Census Bureau had broken up the geographic data file and numerical file for most states so that there was fewer than one million lines, states such as Texas and California were significantly larger than the desired file size. In order to open the files we first had to split each file into smaller files that could be opened by Excel. Then we had to "repair" the files by removing the partial data from the last line of one file and add that data to the first line of the next file. The segmenting program subdivided the file based upon file size and not on a specific logical line. A repeated line would have been a problem when the data files were attached to the shapefile using a join.

The next problem had to do with the fact that the data is released as the raw population and housing data and it was released for almost any level of subdivision of the data. The data was released for the state, counties, metropolitan statistical areas, county subdivisions, census blocks, and other levels. All of the data was identified in a

"decomposed" FIPS (Federal Information Processing Standards) format as database-line information. Each portion of the data, such as the "Fileid" and "STUSAB" was identified by record length (Table 1). The meant that the user had to import the files into EXCEL by indicating all of the data records lengths and the length summaries were not always correct as to where one was at any place in the data.

Name	Type	Size
FILEID	Text	6
STUSAB	Text	2
SUMLEV	Text	3
GEOCOMP	Text	2
CHARITER	Text	3
CIFSN	Text	2
LOGRECNO	Text	7
REGION	Text	1
DIVISION	Text	1
STATE	Text	2
COUNTY	Text	3
COUNTYCC	Text	2
COUNTYSC	Text	2
COUSUB	Text	5
COUSUBCC	Text	2
COUSUBSC	Text	2
PLACE	Text	5
PLACECC	Text	2
PLACESC	Text	2
TRACT	Text	6
BLKGRP	Text	1
BLOCK	Text	4

Table 1. Headers for the raw data file P1 from 0FILE_ STRUCTURE.DBF file available from the Census Bureau. [17]

If you had to import eight files (produced by segmentation), you had to indicate the record and item lengths again and again, hoping not to make any errors.

Once the data files were input, then you had to insert the header codes for each of the fields (Table 2) and the correct code to join to a shapefile required concatenation of the data at the appropriate level (Table 3). Eventually, we found a file we referred to as the

"Rosetta Stone" (that was the name for the file as it was stored on the shared drive for the faculty and students, its original file name from the Census Bureau was "0GEOID_Construction_for_ Matching").

In this file the Census Bureau indicated what level of concatenation was necessary for each of the shapefiles that were available (Table 4). Once the GEO and data files were combined, they were joined with the shapefiles to make sure that the students could use the data. If you remember the earlier discussion, we had decided not to train our students as computer scientists, but as users of GIS. The data had to be prepared for them in a form for easy use. We were not expecting them to be computer or census experts.

DESC	NAME	LEN
File Identification	FILEID	6
State/US-Abbreviation (USPS)	STUSAB	2
Summary Level	SUMLEV	3
Geographic Component	GEOCOMP	2
Characteristic Iteration	CHARITER	3
Characteristic Iteration File Sequence Number	CIFSN	2
Logical Record Number	LOGRECNO	7
Region	REGION	1
Division	DIVISION	1
State (FIPS)	STATE	2
County	COUNTY	3
FIPS County Class Code	COUNTYCC	2
County Size Code	COUNTYSC	2
County Subdivision (FIPS)	COUSUB	5
FIPS County Subdivision Class Code	COUSUBCC	2
County Subdivision Size Code	COUSUBSC	2
Place (FIPS)	PLACE	5
FIPS Place Class Code	PLACECC	2
Place Size Code	PLACESC	2
Census Tract	TRACT	6
Block Group	BLKGRP	1
Block	BLOCK	4

Table 2. Headers available in the HEADERS_2010.DBF file. [18]

STUB	ITEM	SEGMENT
P1. RACE [71]		
Universe: Total population		
Total:	P0010001	P1
Population of one race:	P0010002	P1
White alone	P0010003	P1
Black or African American alone	P0010004	P1
American Indian and Alaska Native alone	P0010005	P1
Asian alone	P0010006	P1
Native Hawaiian and Other Pacific Islander alone	P0010007	P1
Some Other Race alone	P0010008	P1

Table 3. Item STUB for population in the P1 files from TABLES_2010.DBF. [19]

Area Type	PL 94-171 TIGER/Line Shapefile	PL 94-171 Summary Level	Fields in the 2010 Census PL 94-171 Summary Files to Concatenate to match the TIGER/Line Shapefile GEOID10
Census Tract	tl_2010_<state FIPS>_tract10.shp	140	STATE + COUNTY + TRACT
Block Group	tl_2010_<state FIPS>_bg10.shp, tl_2010_<state- county FIPS>_bg10.shp	150	STATE + COUNTY + TRACT + BLKGRP
Block	tl_2010_<state FIPS>_tabblock10.shp, tl_2010_<state-county FIPS>_tabblock10.shp	750	STATE + COUNTY + TRACT + BLOCK

Table 4. Concatenation information excerpted from 0GEOID_Construction_for_Matching.pdf. [20]

With FIPS and header codes, which obey file naming considerations being used to identify geographic units and population information, we had to instruct students to use headers ("OFILE_STRUCTURE.pdf") that were joined to the data from an external file that we created.

Budget considerations caused the replacement of computers in the GIS lab to be delayed for two years. This meant that the computers which were available had fewer processors and less operating memory. For most operations, the computers were

adequate, but GIS requires more "horsepower" than an ordinary lab computer could provide. This meant that the computers could not handle the raw data files, nor could they operate with the census data at the block level for the state after being processed. As the larger, state file was opened, the computers would freeze and had to be restarted. Even if they didn't crash, they could take up to seven hours to process a single button's request. If you tried to go back a step, the backup file would not open and data was lost. This caused us to have to process the data offsite on a personal computer and the state was subdivided into regions for redistricting on that computer. The regions that were constructed had approximately the same populations and students were asked to come up with new districts based upon population, and then construct districts based upon partisan gerrymandering using race as a proxy for political party.

The regions established included South and West Texas (Hispanic majority), East Texas (with a higher African American population), North Texas (included Dallas and Fort Worth with a larger Anglo population), Central Texas (from San Antonio to Austin with race distinctively different between the two cities) and the Houston Region (with a close minority/majority balance). Unlike the state systems where they can subdivide a region based upon population down to the street level and have districts within two or three voters of each other, our students were required to have their districts within a 5% variation based upon population. Some students were able to come within a couple of hundred individuals between their districts, while others were not able to do as well.

Solutions

Necessity, or a course deadline, is the mother of invention, or something like that. The release dates for the data were never published and data was made available five

states at a time. With each data release Texas wasn't in the list. Finally in April, with less than a month left in the semester, the Texas data was released. This provided a problem for us, but through a lot of hard work and many sleepless nights, we prevailed in providing functional data files for the students' final class assignment.

Texas has a lot of political subdivisions and that resulted in over a million lines of data. At first, any attempt to open the file ended in an "out of memory" error message. The files were parsed using an optimum file size rather than at the end of a line of data. That meant that logical lines of data were split and had to be put back together within each file. To make sure that we didn't have duplicate lines, we had to establish a "rule" that data was taken from the end of a file and placed at the beginning of the next file.

To import the data into Excel, we first had to understand the nature of the data and the logical record length for each data item. This information was provided by the "0File_Structure" file which gave us information on how to import the PL_GEOHD_2010, PL_Part1_2010 and PL_PART2 files. The length of the individual entries for each of the geographical features was 498 digits with 101 fields per record for the PL_GEOHD_2010 file alone. The "Structures" file also provided descriptions of the headers for each of the fields. Headers such as "P0010001" represented the "Total" population of the geographic item for all races (not something intuitively understood) and "P0010070" was "Population of six races". In several instances, we had to correct the pre-graded submissions of the student's files for using the wrong columns of data.

In the end, the preparation of the files for the students took up the largest portion of time dedicated to the course. We had to provide the proper TIGER/line files for the students based upon the census blocks for the regions that they were redistricting. We

then had to pare down the files by removing columns of extraneous data so that the files could be used by the students in the University's computers. This involved removing fields from both the join files before connecting them with the appropriate TIGER/line files, including removing the lines that were not at the census block data levels (removal of lines such as the state, county, subcounty and other levels). We created over ten data layers for joining with the appropriate TIGER/line files. This took over forty hours of data preparation. By the time that we had completed our files for use by students, several state agencies had released data in forms that could also be used by the students.

The original intention of the course was to require the students to use the entire state dataset and redistrict the entire state including the addition of the four new districts that were to be added. At best, the software and computers could handle four districts. We began with the entire state, and then went to ten districts and then finally four districts before we found a file size that could be used by the computers. The use of individual computers, rather than mainframe computers, made for a real constraint in what could be done with the class.

With all of the problems and concerns that we had with the release of the data and the nature of the raw data that was released we were pleasantly surprised by the level of work turned in by the students. The quality of "product" that was turned in varied widely, but some of the students did truly professional results. Figures 1 through 3 are examples of the maps that were produced by the students. The typical student spent about 20 hours completing their final assignment. One student was frustrated because the computer constantly crashed while they were performing their redistricting of the region and eventually we had to reassign him to a different region. We don't entire understand what

the problem was except that it was probably hardware oriented. The other student assigned the district was able to complete the project and the Dr. Harris was able to successfully construct new districts based upon the data.

Conclusions

The saying "fools rush in" definitely had an application in this case. While the professors had used census data before, it was "processed" data rather than raw data. It is easy to use data where the headers were already translated and the data was provided at the level needed for analysis, but it is another thing to work with the raw data. We would recommend using previous census raw-release data by anyone waiting for a data release to get use to what needs to be done to translate that into usable data.

If we had realized that the districting extension was only available for an earlier version of the software, we would have left all of the computers in the lab at the 9.3 release level. At the same time, we wanted to teach the GIS students at the latest release level of ArcMap (10.0). The mix of machines meant that the 10.0 users could only use two machines while the 9.3 users only had three machines available. 10.0 files were not compatible with 9.3, and when you opened the 9.3 files and then saved them as 10.0 files, you could not go back to the 9.3 machines. We eventually had to create signs that said "9.3" and "10.0" to place on the machines so that people would not be confused. At the end of the semester, all of the computers were upgraded to the 10.0 release.

For the city of Houston, Texas, people were being told by the city that they should be able to create their own districts for the city redistricting plan and submit their own plans

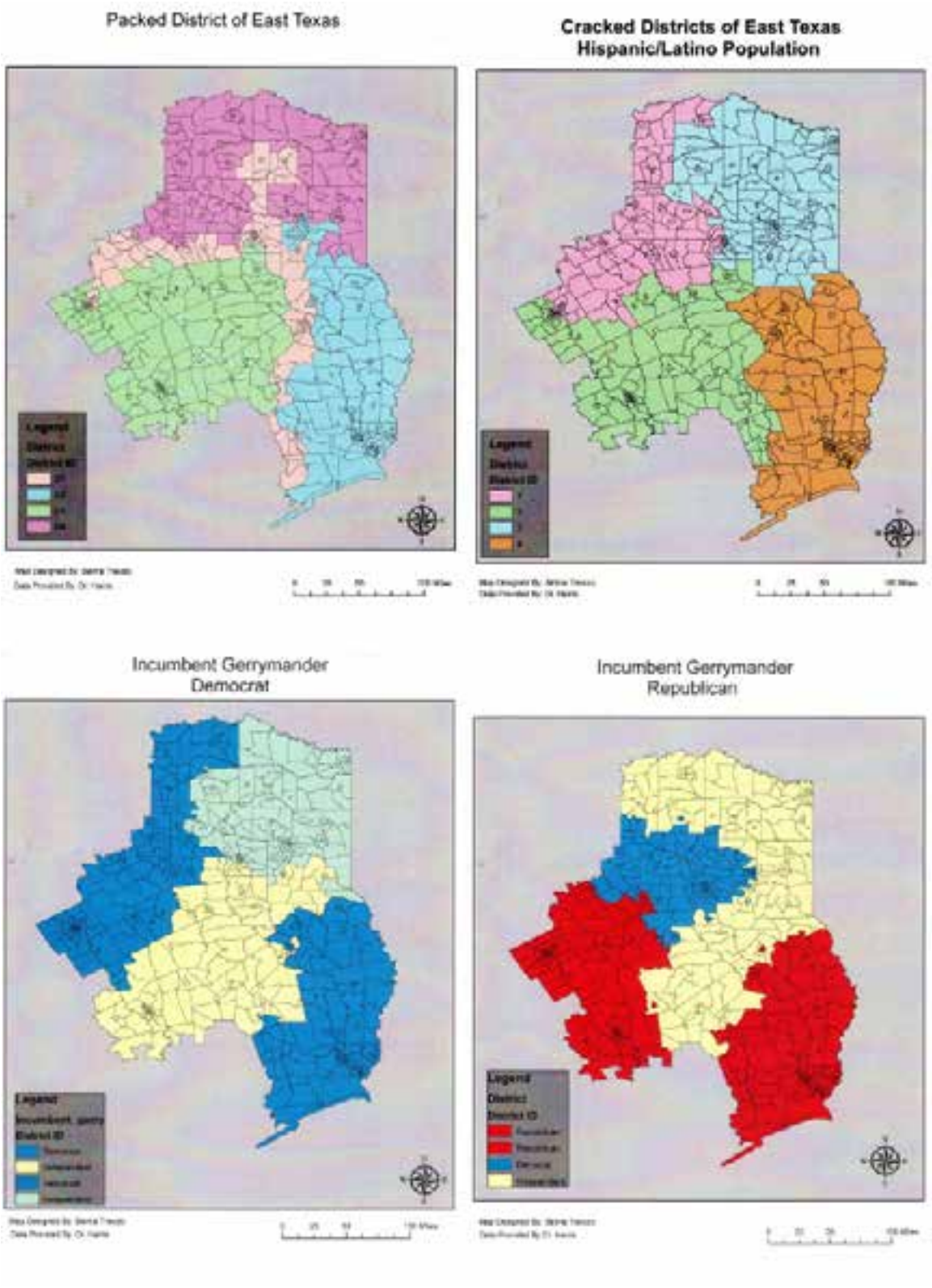

Figure 1. Districting files. Packed, Cracked and Gerrymandered. Belma Trevizo.

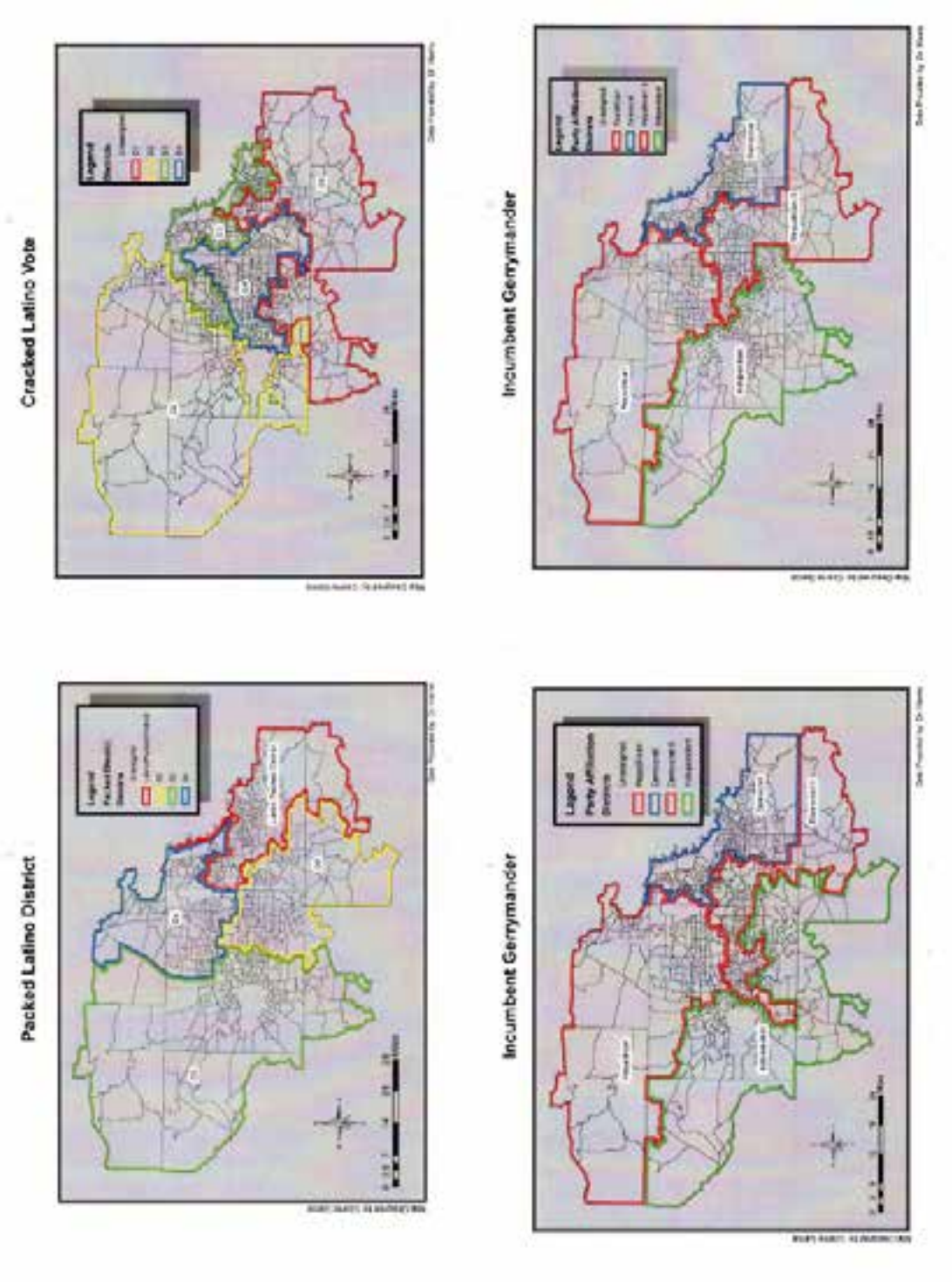

Figure 2. Districting files. Packed, Cracked and Gerrymandered. Cosme Garcia.

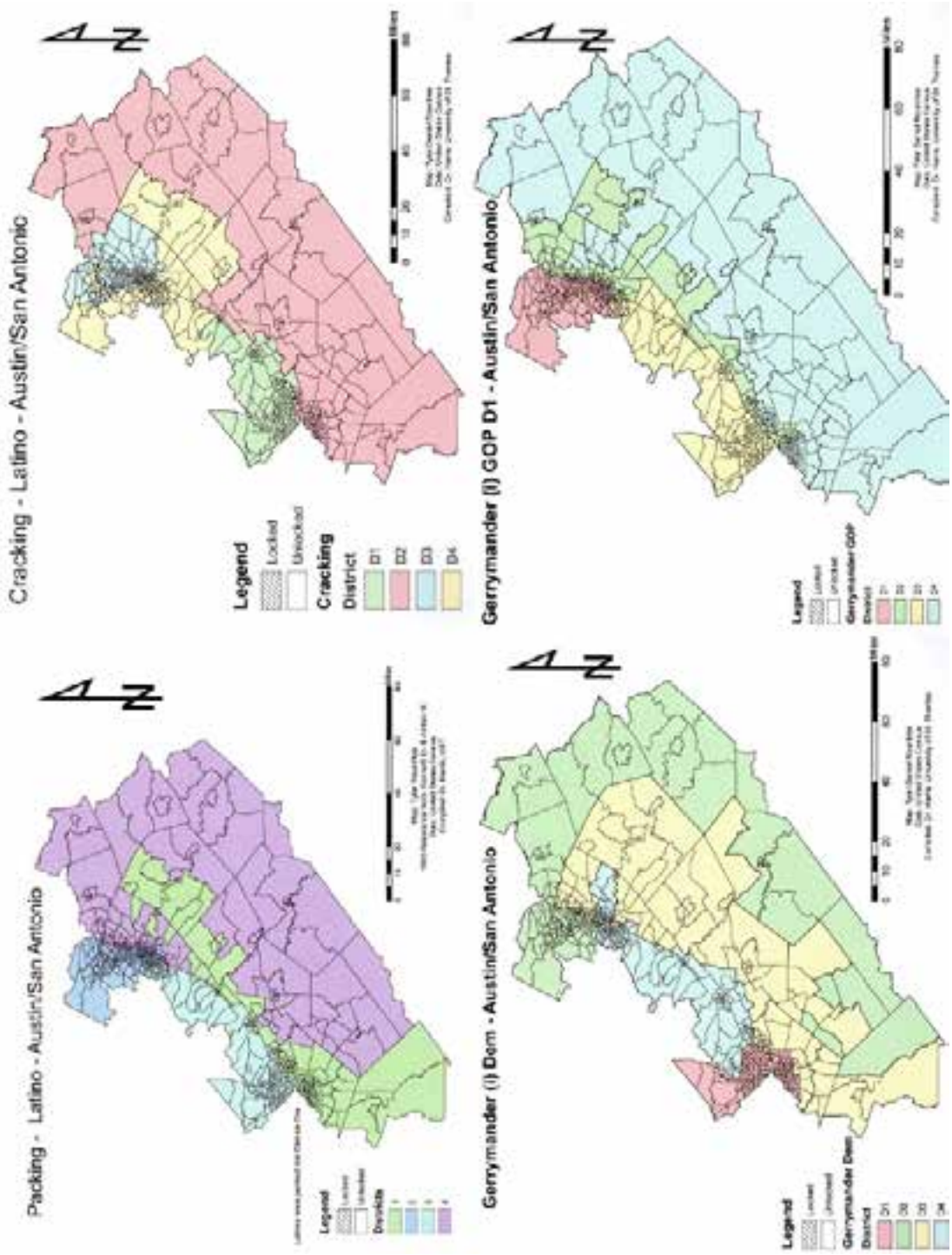

Figure 3. Districting files. Packed, Cracked and Gerrymandered. Tyler Rountree.

for consideration as if the process was "easy" and available to everyone on any computer. If you didn't have a GIS program on your own personal computer, there were websites that you could go to and then you could create districts for your own plan. Almost all of the free solutions, such as the districting GIS extension from ESRI had limitations and some had no, or almost no documentation and assumed an extensive knowledge of drawing districts and the terminology used. There are commercial solutions in terms of extensions and "add on" software packages, but they are relatively expensive and in our case we would have had to know a significant time before the class to make a supplemental funding request to buy the software. Access to the process *should* be free and easy to the common citizen, but it isn't and will almost always be more accessible to those with the money to drive their interests in the process, meaning that gerrymandering will be the rule. Districts will split communities and neighborhoods and will have strange shapes as long as the process is driven by politics and disenfranchisement, rather than producing districts that reflect their populations and interests.

"Fools are rushing in again" might be the subtitle for our new course for the fall semester of 2012. The authors are preparing to teach a course in using the GIS data from the Community Survey to try to predict election results for the 2012 presidential elections. The 2010 census data is already out of date and only reflects ethnicity and housing information and the result of the next election is also a matter of economics and quality of life considerations. The Community Survey is released each year for the previous year and is statistical in nature (a subset of the population is queried and the data is then expanded to the estimated population at that time). This is a much richer dataset and the students will be asked to determine what kind of information may reflect the voting patterns of the

population. Then they will be asked to predict the results of the 2012 election for a region before Election Day.

The proof of their model will be the results of the election in November and we may have an all-night watching party as part of the course. The day after the election we will begin looking at the factors and conduct a multiple regression study looking at the actual results and the characteristics for each voting district to refine the model.

In conclusion, we recommend that GIS professors and departments extend their scope and operation to work with various departments on campus to show how GIS can be used with many areas of study, it pays off for everyone involved. The students get to see how to use this very valuable software and the GIS professionals get to use their skills in areas where they may have not though about. One of the greatest caveats might be to either wait for a semester after a data release to have a change to work the kinks out of the data or to work with previous-release data to make sure that problems can be identified before they become an issue for the course. In this way, the students will not become rushed to complete assignments at the last minute because the data isn't available to them. This class was extremely satisfying for the faculty and students and reflected "real world" experiences for the students in "real time". Our class completed their redistricting proposals BEFORE the state legislature was able to produce their districts. Unlike the state whose new districts didn't pass muster before a court or the Voting Rights' commission, our students' maps didn't have to undergo a rigorous review. Unless you consider the grading process rigorous.

End Notes

1. Bullock III, Charles S., 2010, Redistricting: The Most Political Activity in America, Rowman & Littlefield Publishers, 234 p.
2. United States Census Bureau, 2000, Factfinder for the Nation: History of Organization, Cff-4. Page 1.
3. Ibid. Page 1.
4. Ibid. Page 11.
5. Ibid. Page 11.
6. http://www2.census.gov/
7. Vobejda, Barbara, 1998, August 25, Judges Reject Census Sampling, *Washington Post.com*, Retrieved May 20, 2012, from http://www.washingtonpost.com/wp-srv/politics/daily/aug98/census25.htm
8. GIS has facilitated this process.
9. http://www.tlc.state.tx.us/redist/process_legislature.html.
10. See for example, points made in Bullock III, Charles S., 2010, Redistricting: The Most Political Activity in America, Rowman & Littlefield Publishers.
11. http://gking.harvard.edu/judgeit/.
12. For example, see http://www.theatlantic.com/magazine/archive/2011/07/how-to-turn-republicans-and-democrats-into-americans/8521/1/.
13. http://www.texastribune.org/texas-redistricting/redistricting/hispanics-at-heart-of-texas-redistricting-debate/.
14. Downloaded from http://www.esri.com/software/arcgis/extensions/districting/download
15. Harris, Jr., William M., 2011, GIS in the University: Computer Science, Cartography/Geography or Useful Tool, Proceedings, ESRI EDUC, 2012, San Diego, California. 15 p.
16. Clemmer, Gina, 2010, The GIS 20: Essential Skills, ESRI Press, 156 p.
17. United States Census Bureau, 2011, 0FILE_STRUCTURE.DBF, Downloaded June 15, 2012 from http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/.
18. United States Census Bureau, 2011, 0GEOID_Construction_for-Matching.pdf, Downloaded June 15, 2012 from http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/.
19. Ibid. Page 17.
20. United States Census Bureau, 2011, 0GEOID_Construction_for-Matching.pdf, Downloaded from http://www2.census.gov/census_2010/01-Redistricting_File--PL_94-171/.