

University of Denver

Digital Commons @ DU

Geography and the Environment: Graduate
Student Capstones

Department of Geography and the Environment

2-20-2012

MapUtopia: An Improvement over Existing Census Mapping Applications

Daniel Trone

Follow this and additional works at: https://digitalcommons.du.edu/geog_ms_capstone



Part of the [Geographic Information Sciences Commons](#), and the [Human Geography Commons](#)

Recommended Citation

Trone, Daniel, "MapUtopia: An Improvement over Existing Census Mapping Applications" (2012).
Geography and the Environment: Graduate Student Capstones. 21.
https://digitalcommons.du.edu/geog_ms_capstone/21

This Capstone is brought to you for free and open access by the Department of Geography and the Environment at Digital Commons @ DU. It has been accepted for inclusion in Geography and the Environment: Graduate Student Capstones by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.

MapUtopia: An Improvement Over Existing Census Mapping Applications

Daniel Trone

University of Denver Department of Geography

Capstone Project

for

Master of Science in Geographic Information Science

February 20, 2012

Frank Orr
Capstone Advisor

Steven R. Hick
Academic Director

ABSTRACT

The United States Census public data portal and other existing demographic mapping applications do a suboptimal job of serving two prime data user groups; the casual data user, and the GIS data user. The Census Bureau in particular has produced a user-unfriendly product with a steep learning curve. In response to the lack of an existing platform which could efficiently handle common user requests, the author has created MapUtopia.com.

MapUtopia is an interactive census data mapping application which allows for quick user access to commonly requested demographic data, and allows for user download of that data in the commonly utilized KML format. MapUtopia is built on the principle of usability, and leverages the Google Maps API and open source software.

Table of Contents

Thesis	1
An Overview of Common Census Terminology	1
US Census Data	2
American Community Survey	2
Review of Existing Products (Literature Review)	3
Census Website	3
Other Options	6
How Will this Product Improve on Earlier Products?	8
An Introduction to the Concepts and Technologies Driving MapUtopia	9
Usability	10
Web 2.0	10
AJAX	11
Open Source Software	13
LAMP	13
XHTML	14
CSS	14
PHP	14
Javascript	15
jQuery	16
MySQL	17
Developing Tools / Environment	18
Notepad++	18
XAMPP	18
Utilities	19
MapShaper	19
Shp2mysql	19
Interface	20
Google Maps API V3	20
SlideOut Tab	21

JQueryUI	22
File Form ats	22
K M L	22
W K T	23
Shapefile	23
TIGER Shapefiles	23
Generalized Shapefiles	23
Origin of the Shapefile	24
Implementation in M apUtopia	24
Product in Action	25
Startup	25
Options Slide-Out Panel.....	26
About Button	26
Search Box.....	27
Boundary Select	27
Data Table	28
M ap Variable	29
by Variable	29
Census Geometry Levels.....	30
Redraw Off.....	32
Refresh Button.....	32
Export K M L.....	33
Legend Slide Out Panel.....	34
Transparency.....	34
Color Scheme	34
Legend Values, Lock Button	35
Refresh.....	36
Using the Application	36
Navigation	36
Tooltips	37
Loading Graphic.....	38
Zoom In / Zoom Out / Pan	38

Info Window (custom reports).....	38
Extra Capabilities provided by the Google Maps Interface	39
Business Model	40
Costs Low by Using Open Source.....	40
Sustained by Web Traffic	41
Search Engine Optimization	41
Website Path	41
Tooltips	42
Accessory Products that Leverage the Database	42
Works Cited	44
Websites	44
Books	46

Thesis

The main public access portal for United States Census Bureau data is its homepage (www.census.gov). For most demographic data seekers, the question of 'where' to find the data has never been a problem. It is the frustrating process of extracting this data that has always been the primary roadblock. When using tools provided by the Census Bureau such as the American Fact Finder 2, or the Census interactive map, one quickly comes to the conclusion that these applications were not built for the casual user. Even the intended audience, presumably social data researchers, would undoubtedly find gathering data for use in a Geographic Information System (GIS) to be a daunting proposition. On the other end of the spectrum, a less determined user encountering the Census' homepage for the first time is likely to spend hours just trying to understand the process involved in downloading data for creating a basic map. Most likely, this user will give up before completing their task. It is evident that the US Census Bureau has created a suboptimal product for the vast majority of its users, a product that could easily be improved upon to allow for broader public access to demographic data.

An Overview of Common Census Terminology

US Census Data

The decennial census has traditionally been the primary and most trusted source for American demographic data. Prior to 2010, the decennial census could be counted on for information on a variety of subjects including income, poverty status, race, ethnicity, age, mobility, marriage, and many more. Typically the census questionnaire came in short and long form varieties. All households were given the 'short' form of the census questionnaire. The short form contained questions on a limited selection of categories, including race, ethnicity, age, and homeownership. A random set of households was also given the 'long form' census questionnaire. The long form was conducted as a sample, and had a large variety of questions that spanned from income to housing values to occupational questions. The long form was not administered in 2010, as it has been effectively replaced by the American Community Survey (ACS).

American Community Survey

In the last several years, the American Community Survey (ACS) has taken the place of the long form census questionnaire. The ACS is an ongoing survey that provides updated demographic statistics every year. Previous to the implementation of ACS, information would commonly get stale. For instance, in early 2010 the best information available for many

common demographic categories was still 2000 census data –10 years old at that time).

For larger areas such as populous cities, metropolitan areas, and states, new estimates can now be obtained every year. For small areas of geography such as census block groups and tracts (the ACS does not tract at the block level – as it is too small for a reliable estimate), 5 year estimate averages are reported. Thus, to find the most accurate demographic information for a census tract, one would consult the latest 5 year estimate average for that area (currently ACS 2006-2010 data).

Review of Existing Products (Literature Review)

U.S. Census Bureau Website (www.census.gov)

The typical process involved in visualizing a census dataset (provided one already has an off-the-shelf GIS product such as ESRI ArcMap) entails visiting two websites; the Census' TIGER Shapefiles page (www.census.gov/geo/www/tiger/), and the primary Census Data lookup page; The American Fact Finder 2 (factfinder2.census.gov). The TIGER page contains geometric 'shape' data, while the American Fact Finder 2 contains demographic data in tabular form. Common fields provide a link between the two datasets, to allow for data visualization in a GIS.

The TIGER shapefiles page, while not 'engaging' from a user interface standpoint, is a fairly straightforward and streamlined process. If one is familiar with the basic organization of census geographic objects (such as blocks, block groups, tracts, etc), it will be relatively straightforward to download a shapefile of the desired geographic area.

In contrast to downloading a shapefile, finding one's target demographic data is a far more troublesome process. The American Fact Finder 2 is a very powerful, but very user-unfriendly tool for gathering tabular data. While the Census does provide information on how to use the tool, this seems like an unnecessary burden for the majority of users who would likely prefer to choose amongst a smaller group of more commonly requested datasets. If the user decides to bypass the training, they encounter an unfamiliar tool that does not resemble any other user interface they have ever experienced online.

To get the information that one is looking for, they must figure out that they are being taken through a process whereby they whittle down a set of all possible data products to arrive at precisely the one they need. If a user is looking for a simple product such as 'the number of persons age 55+ in Denver County census tracts', they may be alarmed to discover that there is no 'Age' table, but instead a vast number of tables that contain 'Age' in their title, and that even among Census 2010 data (if they were observant

enough to discover that there is a small dropdown button that allows for the selection of a dataset), there are an overwhelming array of choices.

On delving into the subject matter further (and perhaps asking questions on a message board or calling a demographic expert), they might discover that the table they actually want is titled 'Sex by Age'. Furthermore, they discover that they can't get a singular statistic that contains the age of the population 55+. The Census tabulates data by gender first, then by age. Furthermore, a statistic for age 55+ is not an option, rather, they will have to add up the set of fields for male and female where age is 55-59, 60-64, 65-69, etc.

The user will then have to download the data in the correct format. If they download to Microsoft Excel (perhaps the most common software available for dealing with spreadsheet data), they will discover that the software will eliminate the leading zero of their unique identifier field, and thus they will be unable to join the data to their shapefile. To be fair, the census does provide information on how to bypass these common stumbling blocks. However, the point should be that the process shouldn't be so complex that it requires these instructions.

The number of potential pitfalls is astounding, thus only the most dedicated will arrive at their desired final result – a map of the senior population. Needless to say, it is the author's opinion that the process could

not only be easier, but it could be drastically reduced to less than a dozen clicks of the mouse in total.

Other Options

If the user would prefer not to work through the tedious process of gathering data through the US Census Bureau's website, they will discover that there are various examples of census mapping applications that are more user friendly than that of the Census Bureau. Specific to the Denver Metropolitan area, the Piton Foundation (www.piton.org) has created a local implementation of a data mapping application. Their site is equipped with a fast and intuitive interface (which makes use of the Google Maps API) that not only maps 2010 census data, but has the ability to map the absolute or percent changes in demographics since Census 2000.

The disadvantages are several: the demographics able to be mapped are limited to Total Population and Race/Ethnicity Categories. While they have the ability to map subcategories, such as the population under or over 18, the program is otherwise limited. The legend is static and cannot be changed, and there is no option for exporting data in any format.

The New York Times (projects.nytimes.com/census/2010/map) has a very user friendly application for mapping 2010 census data as well as another for American Community Survey Data

(projects.nytimes.com/census/2010/explorer). This application, like Piton's, is also powered by the Google Maps interface. While many of the visuals are stunning (such as a creative dot density based mapping of race), it is limited to a few predefined data sets that mostly relate to race and income. Despite 2006-2010 American Community Survey data having been available for several months as of the time of writing, the New York Times was still using 2005-2009 data. Similarly to the Piton Foundations application, as well as the US Census, there is no direct method to output a Shapefile or KML with ready to use data.

The best example of a robust full-featured census mapping program is Social Explorer (<http://www.socialexplorer.com>). Social Explorer uses dynamic mapping technologies and datasets that stretch all the way back to 1940 at the Census Tract level, and even further (1790) at the county and state level.

Social Explorer has the ability to map a wide range of demographic data including Race, Income, Family Structure, Marital Status, Education, and many more. Social Explorer also grants the ability to create maps and save statistical reports.

Social Explorer is not without its disadvantages however. The basemap is not based on the Google Maps API, and is cartographically unpleasant to look at. The map was programmed in Adobe Flash, rather than Javascript, creating a product that many users will likely have to download a plugin to

use (if they refuse the download, the application will not work). The legends on Social Explorer are hard-coded, preventing the ability to create custom data categorization. While the ability to create maps is present, the application does not provide the ability to download data in any GIS related format.

Also, while much data is included, Social Explorer does not provide unrestricted access to mapping the American Community Survey on the free version of its application (this ability is available for a price).

How Will this Product Improve on Earlier Products?

The previous products tend to suffer similar flaws. The first and foremost is that there is no tie-in to the GIS community. One of the most common complaints amongst GIS professionals is that they cannot easily retrieve the data that they need. In fact, only the Census Bureau itself, out of the programs I have surveyed, provided the ability to download data, and that process was too cumbersome for all but expert users. Of the graphical user interfaces that I did mention (Piton, New York Times, Social Explorer), neither of them had the ability to export data.

A second flaw I noticed was either an overwhelming abundance of data (US Census, Social Explorer), or a very limited selection of data (NY Times, Piton). A good census mapping application should contain data on a wide

range of commonly researched topics, however, it should not cover it in such detail that it becomes difficult to find what one is looking for. Conversely, having few datasets available similarly reduces an applications utility.

A third flaw I noticed was the lack of ability for the user to customize a legend. While some applications allowed different color schemes to be used (which can be a fun accessory), none allowed the custom describing of data ranges. There are no 5 rigid ranges that can adequately describe data on any topic across the country, yet the user was forced to see the data mapped in such a way. For example, if one was trying to map areas of poverty to determine if there were well defined pockets of extreme poverty, a household income category hard-coded to define the lowest income category as 'less than \$25,000' is not useful.

A fourth flaw was that none of the programs gave the user an option to map a specific area of geography only. For example, if a user wanted to discover the distribution of race in Denver County (but was not highly familiar with the area), mapping the entire area around Denver County could be slightly confusing if one did not understand precisely where the county boundaries were.

An Introduction to the Concepts and Technologies Driving MapUtopia

Usability

The concept for MapUtopia came about during the course of the authors work as a GIS Technician for the City and County of Denver. Tasked with gathering 2005-2009 American Community Survey data for consumption by various city departments, as well as compiling 2010 Census data for the City Council Redistricting Effort, it became clear that the process for compiling data contained too many steps and had a steep learning curve. At worst, the difficulty of the process provided a barrier that would deter most casual data users. As I came to be regarded as a point of contact for census information within the city, I came across many individuals who had gone to the Census webpage first, only to get discouraged and give up.

The basic concept behind MapUtopia is that there should be a resource able to handle 90% of typical data requests (the other 10% presumably being researchers of niche subjects) in a simple, user friendly manner that requires virtually no prior experience in GIS or detailed instructions to use the application. Ideally, this could be done by leveraging other commonly used technologies that a typical user is most likely familiar with, such as Google Maps and ESRI ArcMap.

Web 2.0

The phrase 'Web 2.0' can be defined in a multiplicity of ways. In Wikipedia, Web 2.0 is characterized as 'web application features that facilitate participatory information sharing, interoperability, user-centered design, and collaboration on the World Wide Web' (http://en.wikipedia.org/wiki/Web_2.0). Within the last decade, the online experience has evolved from merely utilizing the internet to 'look up' information on static websites (now considered Web 1.0), to a point where the user expects an interactive experience with dynamic sites that behave much more like desktop applications. Instead of simply reading the news, we expect to be able to share and comment on it. Instead of looking up an address of a company, we can purchase products or find out which stores locally have an item in stock. Instead of posting pictures of our vacations in an online photo album, we can quickly post our pictures to our Facebook 'Wall', and then discuss them with friends and family.

Many of these features were not even possible in the earlier ages of the internet, limited by bandwidth and less advanced technology. Since about 2005 there has been an explosion of interactive content that has revolutionized our collective internet experience. Much of this advancement has come from a technology referred to as AJAX.

AJAX

AJAX is a programmatic technique that gives a website the ability to continually load new web content from a remote server without having to completely reload a page or redirect to a new page. Although AJAX had been used in some form for many years previously, it was an obscure technique before a seminal article appeared on the subject in 2005, which also coined the term.

AJAX, which ultimately revolutionized web design, was originally popularized in part by Google Maps, (the application which serves as the backbone for MapUtopia). Previous web mapping applications utilized an interface where with each mouse movement, the user had to wait for a new page to load to see updated content. In retrospect, this can only be described as 'clunky' and makes for a frustrating user experience.

The acronym AJAX itself originally stood for 'Asynchronous Javascript And XML' to describe the technical mechanism by which it worked. Asynchronous refers to the ability of a process to run concurrently with other processes. Thus, in the example of internet mapping, the application is not placed in a 'frozen' state while waiting for data to load.

While the 'X' of AJAX refers to XML (Extensible Markup Language), in reality, a number of different data sources can be used in its place. In the case of MapUtopia, a more appropriate acronym might be AJAJ, for 'Asynchronous Javascript and Javascript', as the returned data from the server is in the form of Javascript statements.

Open Source Software

Whenever possible, the author attempted to utilize Open Source software. 'Open Source' refers to a freely available software product, usually owned and/or maintained by a dedicated community of developers. The author was successful in only using open source software with one notable exception; the compiling and processing of census data was done with Microsoft Access 2007. While this process could have conceivably been done with an open source database program, the time savings by using Access were enormous. For the 2010 Census, the US Census Bureau provided Access Database 'shells', which allowed for easy importing of data. Before purchasing Office '07, the author attempted using the freeware solution provided by Open Office (www.openoffice.org), but was unsuccessful.

LAMP

LAMP stands for Linux, Apache, MySQL, PHP. Not only is this combination of products free for use, it is also an extremely powerful and widely used package for web development in general. Many web hosting companies provide this suite as an option for their customers, usually by default.

XHTML

HTML (HyperText Markup Language) is the basic language of the web page, and XHTML can be defined in a basic sense as HTML code that is written to well-formed XML standards. XHTML can be considered the backbone of a webpage, and is largely responsible for the structure and setup of a website. Page layout, basic controls, and linking can all be performed without any additional scripting language, and is often all that is necessary to produce most static websites.

CSS

CSS (Cascading Style Sheets) is a 'style sheet' language that is responsible for 'styling' web pages. CSS works on the various abstract elements of a website, such as object names and classes, and defines their attributes such as color, size, and more. CSS is used extensively in MapUtopia to define the symbology color schemes, spacing, and font style in the document. Additional CSS snippets are used by plugins (such as JQuery) to stylize custom controls.

PHP

PHP (PHP: Hypertext Preprocessor) is a server side programming language. In essence, if a webpage has PHP code within it, it is sent to the server first to be processed. The output from the server is then inserted back onto the webpage, where it is then served to the user. PHP allows complex dynamic processes to occur. As a basic example, a PHP code block could query the server for the date and time, and deliver that code back to the webpage, and then to the user. A static HTML page would not be capable of this behavior. MapUtopia uses PHP scripts to power its data retrieval procedures from a MySQL database located on the web server. The individual PHP files written for MapUtopia contain procedures to retrieve shapes and demographic data from the database, populate the boundary select dropdown box with a list of eligible areas of geography, and write and export KML files.

JavaScript

JavaScript is a client side scripting language. This means that the javascript code is processed by each individual users browser (be it Internet Explorer, Firefox, Chrome, etc). This has the potential to cause a number of difficulties, since there are a wide variety of web browsers in common use throughout the world, and not all of them interpret JavaScript in the same way. Thankfully, most browsers implement a very similar brand of

javascript. However, they are not perfect, and often browsers implement certain features differently. Javascript is used extensively in MapUtopia to interact with the GoogleMaps API, validate user control values, populate select box dropdowns with hard-coded values, add auto-complete functionality to the location search box, and send out AJAX requests. It also provides the building blocks for the JQuery Library.

JQuery

To help deal with browser compatibility issues noted above, and to help tap the power of client side scripting, the author makes extensive use of JQuery (jquery.com). JQuery is an example of a Javascript library. In its most basic sense, JQuery is a collection of Javascript functions which allow a programmer the ability to harness Javascript in a much more concise way. Of major importance, not only does JQuery simplify code writing, it also takes care of the detailed intricacies of dealing with the bugs and quirks inherent in the many currently operating web browsers, thereby providing a uniform platform that is not browser-specific. Though JQuery is essentially Javascript, its conventions and structure are unlike Javascript, and are more akin to coding in an entirely different programming language. The author makes extensive use of JQuery in interacting with web controls, for example

by using JQuery to get and set values with a minimum of native Javascript code writing.

MySQL

MySQL (www.mysql.com) was the author's database of choice. In retrospect, PostgreSQL / PostGIS (www.postgis.org) may have been a more GIS-friendly choice due to its enhanced support for spatial data. However, MySQL did have the functionality to do exactly what it was intended to do; store and display polygon information. MySQL has the ability to store spatial data within the traditional column and row structure of a database by storing each shape as a text representation of the object. Traditional database capabilities such as indexing are applicable to the spatial data as well. A number of built-in common spatial queries are available, such as the ability to query for shapes based on a bounding box area, which is the technique used in MapUtopia.

Aside from the standpoint of compatibility with spatial data, MySQL was a convenient choice because of the free availability of the PHP MyAdmin program, which is a powerful, yet user friendly interface for handling large quantities of data.

A weakness of MySQL though, is it's lack of support for dealing with a variety of common file types. While PostGIS has routines for handling

export to the industry standard Shapefile format, MySQL does not have any predefined routine or open source library for doing the same.

Developing Tools / Environment

Notepad++

Notepad++ (notepad-plus-plus.org) is a simple scripting environment that allows for writing code in HTML, Javascript, PHP and more. A prime advantage of Notepad++ is the tabbed editing environment, allowing for easy editing of multiple open documents. The program is currently free to use. Some features are: color coding of keywords, annotation and string values for easy reading of code. It also includes the ability to save a file with any of the main web development extensions (.php, .html, etc), and the inclusion of many common word processing features such as find/replace, and printing. The author conducted 100% of his scripting within Notepad++.

XAMPP

XAMPP is a package consisting of an Apache Server, MySQL database, and PHP/Perl compiler. In laymans terms, XAMPP is a full-fledged server environment that one can use to test website performance locally, rather

than uploading each edit to the main server. Included is a full MySQL database installation with PHP MyAdmin included, along with the capability for full customization of the server environment.

The advantage of XAMPP is the ability to quickly test website features and conduct debugging without having to disrupt your published website with changes that could potentially cause problems. XAMPP is free to use.

Utilities

MapShaper

MapShaper (www.mapshaper.org) is a free website that features the ability to generalize shapefiles. The program allows the user to upload their own shapefile, and then responds to user input to remove polygon vertices that may be excessive with respect to the desired map scale. The most important feature is the user's ability to view the shapefile at every step of the process, to determine whether they have chosen to remove too many vertices (and thus have shapes that are too distorted to be acceptable for their intended use). Topological integrity is maintained with MapShaper.

Shp2mysql

Shp2MySQL is a command line program that takes a number of user parameters to convert shapefiles to WKT (Well Known Text) format embedded in an easy to upload SQL file. The author took every shapefile downloaded through TIGER and ran the data through Shp2mysql before the data was able to be uploaded to the MySQL database.

Interface

Google Maps API V3

The latest version of the Google Maps API (v3) was used in the making of this website (developers.google.com/maps/). While the Google interface dominates the application, the actual amount of code that MapUtopia uses to access the API is fairly small. Thus it is easy to tout the 'power' of the API for being able to do so much with so little code, and it is likely why there is a prevalence of Google Maps driven applications online.

The Google Maps API has the advantage of having numerous built in controls and features that are not available with other API's, chief among them the ability to use 'Street View' or toggle between satellite imagery and a road map. Also, the Google basemap is instantly recognizable and, in the authors opinion, is the most cartographically 'pleasant' of any other readily available basemap.

However, in retrospect the author should have taken a closer look at utilizing OpenLayers instead (openlayers.org), a freeware API with similar features. An unfortunate limitation of using GoogleMaps instead of an open source option is that exporting an image of a map created in MapUtopia.com (which could be extremely useful to an end user) would be a potential copyright violation. This is doubly unfortunate, as from a business standpoint, the open circulation of maps with a MapUtopia.com logo on them would have been a valuable free advertisement opportunity.

SlideOut Tab

The slideout tab used in MapUtopia is an example of a JQuery 'plugin'. This particular plugin was developed by web blogger William Paoli (<http://wpaoli.building58.com/>). It was built using JQuery, and provides an easy set of functions and methods to quickly implement for any application.

The slideout tab allows for placement docked anywhere on the screen, and lets the developer customize the buttons that uncover each tab. The author has used slide out tab to house most of the web controls available to the user. The main advantage of the slide out tab is that the user is able to collapse or expand the tab at will, revealing the entire map with an absence of 'clutter'.

JQueryUI

The JQuery UI (JQuery User Interface, jqueryui.com) is a library of Javascript functions and CSS libraries that enhance and improve upon the standard 'off the shelf' web controls such as standard HTML radio buttons and drop down menus. Combined with JQuery 'themes', they provide an enhanced user interface that embellishes standard web controls while keeping a consistent interface for users.

The author specifically used the JQuery UI for the buttons and transparency control in the application. (The select controls, while visually similar, were made with CSS due to an inherent conflict in styling a select box that was located on a slide out tab). The author utilized the 'vader' theme for uniform styling.

File Formats

KML

KML (Keyhole Markup Language) has recently been popularized by being extensively used by Google (in Google Maps and Google Earth) and is widely accepted by the general public. It is an XML styled document that can be written in plain text editors, and therefore is easily parsed. MapUtopia

makes use of the KML format as an easy to write file format that is compatible with many common data platforms.

W K T

W K T stands for Well Known Text format. It is a geographic file format that can be read on plaintext editors, and is the only supported spatial format (aside from its cousin W K B – Well Known Binary) for MySQL databases.

Shapefile

TIGER Shapefiles

The Census Bureau's TIGER shapefiles page contains all shapefiles used in this web application. The author chose the following more commonly used levels of geography: (block group, tract, county, state, place, cbsa, csa). The author had originally intended on also providing data at the census block level, but the amount of data processing and the demand it placed on the web hosting service made this level of detail unfeasible.

Generalized Shapefiles

The shapefiles provided by the TIGER page are very detailed and are able to be viewed at a high resolution. For the authors purposes, storing (for example) a state shapefile at its original resolution is an inefficient use of bandwidth considering it is only viewed at a comparably large scale. Thankfully, TIGER also provides generalized versions of many shapefiles (www.census.gov/geo/www/cob/). Those that weren't available for generalization were run through the MapShaper program.

Origin of the Shapefile

The Shapefile is a file format developed by ESRI and is widely used in most commercial GIS mapping applications. It consists of three main files, a shapefile (.shp) that contains geometry information, a database file (.dbf) containing attribute information, and an index file (.shx) which links the two previous files. While the shapefile is an essential file format output for MapUtopia, it is a difficult file format to write to. The feature to output data to a shapefile will be included in future versions of the website, but was left out of the present version.

Implementation in MapUtopia

The shapefile is used currently in MapUtopia as the main source of input geometry. The shapefiles of every geometry level were downloaded

from the US Census TIGER shapefiles page, generalized, and then converted and stored in a MySQL database (via the shp2mysql program) as type WKT.

Product in Action

The official url of the website is www.maputopia.com/capstone.html.

Startup

On load, geo-location is used to pinpoint where in the country the user is located in. Geo-location is the act of using information about the user found either in the browser itself, or through the users IP address. The geo-location routine the author used was downloaded straight from Google and modified only slightly. The initial display will reflect this location by displaying a map centered on where the user is located.

The initial map presented to the user is a map of average household size. This was in alternative to a blank map or a map of just geography outlines. This way, the user is exposed on first contact to the mapping capabilities of the program. This sets an expectation as to what type of product the application delivers, and is a stepping stone into the other features that are available.

Options Slide-Out Panel

On visually inspecting the application window, the first thing that will jump out to the user is a panel of controls in the upper-right area of the screen. On startup this panel defaults to the 'visible' position, (in contrast to the Legend panel, which defaults to 'hidden'). The contents of this panel include an about button, a search box, and several drop down selection boxes. Mid way down the control box is a series of buttons that correspond to census geometry levels. This is followed by a toggle button titled 'Redraw Off', a normal button titled 'Refresh' and lastly, an 'Export KML' button.

The control panel itself is a slide-out tab control. It can be toggled on and off by clicking on the 'Options' text. Toggling off gives the advantage of more screen real estate from which to view the map.

About Button

The about button displays simple information in a familiar pop up dialog box control. For now, the text reads only basic information about the author, project, and university, along with a contact email. Future plans could be to link to, or embed a youtube video explaining some of the features of the website.

Search Box

The search box should be an intuitive control for most users. Upon typing into the box, place locations are auto-populated into the search results. The code snippet for this feature was found on Google (source), and utilizes a Google place name database to populate the place names that appear and change with every keystroke by the user. Upon clicking an item in the search results, the user will be taken to their location of choice. Since it triggers map movement, the map will be entirely redrawn at the location of choice, along with whichever data variables had been previously chosen.

Boundary Select

The Boundary Select allows the user to determine the bounding area that they would like to map. By default it is set to 'Map Area', which is the entire extent of the map visible to the user. However, the user can also elect to choose a geography feature from the drop down menu, which includes an auto-populated list of counties and micro and metropolitan regions in the immediate area. Upon selecting a boundary area, the map will pan to the geographic center of that feature geometry, and constrain the map to the extent of the feature geometry. The option to confine mapping to a specific boundary is a unique control that is not found in any competing mapping

application product, but which could be very useful to a user looking to create a demographic map of only a specific area.

Data Table

The Data Table select control allows the user to choose which dataset they would like to use in their map. The list is hard-coded with a number of popular choices for both the 2010 Census, and the 2006-2010 American Community Survey. Rather than overwhelm the user with the original detailed census data tables, many data tables are custom derivations, such as Age (the original Census Table would be Sex by Age). This was done because in many cases the original census tables have data in a form which is too granular for most users.

As of the time of writing, the ACS data is confined only to the state of Colorado. In the future this dataset will be expanded to the entire United States. Ideally, MapUtopia will also contain information from previous US Census. Future plans also call for an additional select dropdown whereby a user chooses their specific dataset first, and then uses the data table dropdown to choose their information subject. (As it would be obvious, adding too many additional data tables and data sets into one dropdown would quickly make this control unwieldy).

Map Variable

Once a data table is selected, a user can choose a variable to map. For instance, in the 'Age' data table, the variables include Total Population, followed by a list of age range breakdowns, then followed by a male, female breakdown. The general scheme of the data table is that the main population of interest is listed without a prefix, while the components that sum to the main population of interest are preceded by a underscore character.

Upon selecting a map variable, nothing immediately changes (the map will only redraw upon moving the map, choosing a new location in the search bar, changing the Boundary Area select control, or pressing the Refresh button). This is so the user can exactly customize their settings before drawing the map of their choice. If an auto redraw was set for every change, the user could experience frustration, as some densely populated areas could take up to half a minute to load even on a fast connection (for example, New York City, NY has hundreds of census tracts and takes quite a while to load).

by Variable

Optionally, a second variable can be chosen from the same data table. This is hugely important because it allows for mapping ratio data, rather than just raw totals. This is similar to the simple mapping capabilities available in standard GIS software. Choosing a second variable allows mapping in the form of Var1/Var2. Thus if the first variable chosen is the age group from `_Age_18_to_24`, while the second variable chosen (the 'by Variable') is `Total_Pop_2010`, the end result will be a mapping of the ratio of the total population that is of an age from 18 to 24.

Census Geometry Levels

MapUtopia supports demographic information at 7 distinct geographic levels; Block Group, Tract, County, State, Place, CBSA, CSA. The Block level was excluded due to the significant overhead in data loading at such a fine level.

Users are able to choose which level of geography they would like to see mapped by selecting an available level from the Geographic selection button group. Due to the possibility of selecting too large of a dataset, some limitations have been placed on which levels of geography can be selected at specific zoom levels. For example, the block group level cannot be selected if a user is zoomed out to an area where multiple states are visible. To allow

this would require the sending of 10's of megabytes of information, which would require data load times that are impractical.

Instead, selectable geographic levels are indicated by buttons that are a darker shade of grey. Un-selectable levels have a 'greyed out' appearance similar to behavior in other applications. The currently selected level of geography is indicated by a black button with white type.

As a user zooms in and out, if a zoom level does not support showing the currently selected level of geography, the geography level will change to the most similar level of geography available. Similar geographies are classified in two separate groupings, the first being the familiar Block Group, Tract, County, State grouping. This is a logical grouping because each smaller level of geography is contained in a larger level of geography in this grouping, without any overlap. The next grouping is not a logical grouping, and contains the geographic levels of Place, CBSA, and CSA. While not a logical grouping per-se, these levels of geography group data at either a city or regional level.

In some more extreme cases, on a zoom change the geographic level will 'jump' groups. This is unavoidable in some situations since the lowest level of geometry in the second logical grouping (Place, CBSA, CSA) is the 'Place', which still does not support drawing at the most zoomed in levels.

Redraw Off

The 'Redraw Off' button is a useful option, though perhaps the least intuitive item on the interface. Its main purpose is to tell the program not to retrieve any more data from the server, but yet, to not clear the data on the screen as well. Thus the user is able to zoom in and out on the dataset without having to wait for the data to be retrieved from the server after every pan or zoom. In complex areas, such as Chicago, Los Angeles or the New York City area, loading times can be significant for the smaller levels of geography. Having the option to suspend auto-redraw allows the user to examine the data in a much faster environment.

The 'Redraw Off' button is a toggle button, meaning it can be used to click the auto-redraw setting on or off at any time. Future iterations of the program may automatically employ caching to do this, so it is possible that this button will be superfluous at some point in the future.

Refresh Button

As the program was purposefully designed so that not every setting change triggers a redraw (to avoid a constantly reloading interface), a refresh button was created so that the user can determine when they are

done selecting their data table, variable, and geography level options before redrawing the screen.

Used in concert with the 'Redraw Off' toggle button, the refresh button can allow the user to redraw on demand, rather than by map movement.

Export KML

The Export KML button is the primary difference between currently existing census mapping applications, and MapUtopia. Export KML gives the user the ability to take data that they have mapped online, and be able to visualize it in a familiar environment, such as a GoogleEarth, GoogleMaps, or any other compatible GIS program.

Under most conventional browser configurations, MapUtopia will ask the user if they want to save the .kml file to a location on their computer, or if they want to open it directly with the default kml application (typically Google Earth).

The entire area of the map window is exported in the file, unless a Boundary other than 'Map Area' is chosen. All geographic and symbolic information (color scheme, legend values, and transparency) are retained by the file, allowing the user to immediately view the data without having to customize any settings after the data is downloaded.

Legend Slide Out Panel

Unlike the Options Slide Out Panel, the Legend Panel starts out in 'auto hide' mode. Clicking on the 'Legend' text-button will draw out the entire panel and reveal a number of mapping options.

Transparency

The transparency slider controls how much the base map is visible beneath the data symbology. Values can range all the way from 0 (only the geometry outlines are visible) to 100 (the base map is not visible beneath the opaque data symbology). MapUtopia defaults at a value of 50, which allows the user the perspective of being able to clearly see the pattern of the data being mapped, while at the same time retaining an understanding of perspective via the familiar Google basemap.

Color Scheme

The Color Scheme select drop-down allows the user to change the symbolic color scheme. MapUtopia defaults to a "Purple to Yellow" scheme, but other color schemes ("Brown to Sky Blue", "Orange to Yellow", and "Blue to Pink") are available, as well as the option ("None") to draw only the

geometric outlines (similar to setting transparency to 0, or both MapVariables to "None").

Legend Values, Lock Button

The ability to customize legend values is another feature of MapUtopia that is absent in other online census mapping applications. By default, MapUtopia maps data in quantiles. Thus, 20% of all data values will correspond to each color value. The legend is updated with each move of the map or press of the refresh button, thus the legend values are constantly being recalculated.

To turn on customization, the user will need to press the 'Lock' toggle button. This 'locks' the legend values from being recalculated with each move.

The way the legend values are then manipulated (provided that 'Lock' is toggled, is similar to GIS desktop programs. In MapUtopia, the minimum value in the dataset is located in the text box in the upper left, while the maximum value is located in the lower right. The user is allowed to manipulate all the text box values on the right hand side except for the maximum value. This permits the user to set the maximum value for any of the categories. The start of the next category will automatically be set to the previous value plus .001.

Refresh

A duplicate refresh button is placed on the Legend slide out tab, so that when the user changes legend values, they have a convenient button close by to test their new classification schema. This Refresh Button runs the same routine as the Refresh button on the Options slide out tab, and they are functionally the same in every way.

Using the Application

Once the user has chosen the appropriate data table and variables to map, and have chosen their preferred unit of geography, they are now free to explore the resulting mapped data.

Navigation

Navigation can be controlled either with the native Google Maps navigation controls in the upper left corner of the map, or alternately (and highly preferred) through a two button mouse with a scroll wheel. The scroll wheel will have the effect of controlling the zoom level, while the mouse can be used to pan around the screen by clicking to 'grab' a location on the map, and moving the grabbed location to either direction to reveal new areas of

the map. You will notice that with each mouse movement the map will be redrawn, constantly gathering data from the server to populate the map with information about the topic of interest you have chosen.

Since the program is instructed to redraw the data upon an idle state, the completion of each mouse movement or scroll movement will trigger the redraw of demographic data. Unfortunately, at times this may mean that even with a small mouse movement, all of the data on the map will have to be re-retrieved from the server. In the future, caching may be implemented to take care of this undesirable lag. For now, users are free to disable auto-redraw, meaning the map will only retrieve information when the user presses the Refresh button (or until the auto-redraw button is toggled again).

Tooltips

If a user needs more information about any of the controls on the slide-out tabs, tooltips have been implemented to provide a short explanation. A tooltip is a text information snippet that pops up when the user hovers the mouse over a control. This contributes to the ease of the user interface, by providing a direct answer when the user poses the question; 'what does this do?'

Loading Graphic

The loading graphic appears when the web page is communicating with the server and retrieving information. It is a friendly reminder to the user that the program is not being non-responsive, but rather, that data is loading. Without it, the user may question if the button they pressed actually 'did' anything.

Zoom In / Zoom Out / Pan

The application has the standard navigation capabilities that a user would come to expect from a Google Maps interface. Zoom in can be triggered by the slider bar in the upper left corner of the screen, or by scrolling in an upwards direction on the mouse. Zooming out can be triggered in much the same way, though with a downward mouse scroll. Panning (moving around the map without zooming) is achieved by the grab and move technique, whereby a user 'grabs' a part of the map, and moves the grabbed location to a new location, revealing previously uncovered portions of the map.

Info Window (custom reports)

By clicking anywhere on the map that has shows feature symbology, an InfoWindow will pop up. An InfoWindow is a way of providing information on the specific geographic area that was clicked on. For instance, if the user is mapping at the tract level, and then they click on one of the census tracts in the map, an InfoWindow will appear displaying the specific statistics available in that area for whichever Data Table is selected in the Data Table select dropdown.

Extra Capabilities provided by the Google Maps Interface

The Google Maps interface also provides several other features which deserve mention. In the upper right hand corner of the map there is an option to use either a roadmap, or satellite imagery. Users are free to toggle to whichever view they prefer. Typically, the roadmap is easier to display overlay information with, however, the user may have a specific need to use satellite imagery to gain a better understand the area.

Secondly, Google Maps 'Streetview' capability is also present by default. This can be activated by dragging the Google Earth Street Viewer (an icon of a man which is located in the upper left of the screen) to any area of the map. Google will populate the street areas with a blue outline if street view is available for that area. If the user 'drops' the icon on an

eligible area (via a mouse-up click), a Street View image will be presented. Navigation is also possible within the StreetView image itself.

Business Model

MapUtopia was conceived to be fully self-supporting and ideally a profitable venture. Among the chief concerns of the author is the ability to derive revenue at least equal to, but preferably exceeding the costs of web hosting.

Seeing that the concept of the map is to make census data more available to the common user, it would not make sense to charge in any way for using MapUtopia. (This would not be possible at any rate, since Google does not permit free usage of its GoogleMaps API in the context of a pay-per service or membership).

Instead, the recuperation of costs could theoretically be achieved by gaining a critical mass of users and a modest, non-intrusive source of revenue driven by advertisements from a service such as Google AdSense (www.google.com/adsense).

Costs Low by Using Open Source

In order to keep development costs as low as possible, a decision to use open source software was made. Indeed, there were many more off-the-shelf products that could have facilitated a much quicker development process. (ArcGIS Server). However, it was readily apparent that the revenue derived from advertisements would never be enough to pay for the licenses of the software that would create and support it. Knowing that the margins were low, free software such as MySQL and Notepad++ were used (covered in previous sections).

Sustained by Web Traffic

The hope is for the application to be successfully supported by a critical mass in site traffic. Thus the more useful the site can be to the average user, the higher the likelihood of repeat visits and word-of-mouth advertising to catch on. In addition, a number of SEO (Search Engine Optimizations) were conducted to help enhance the site's prospects of high search engine ranking, which will be critical in drawing traffic to the site.

Search Engine Optimization

Website Path

By using a site path such as www.MapUtopia.com/census-maps/demographics.html, a search engine can glean information about the website, its content and its relevance from the path itself. By using keywords within the url address, a higher search ranking can be achieved. Future, improved versions of MapUtopia will follow this scheme.

Tooltips

Search engines rely on a website's text content to extract information about the subject and relevance of a website. Tooltip information is not only helpful to the user to identify the purpose of controls, tooltip information can also be a potentially significant source of information about a website that otherwise does not publish any text on the site. Dynamic websites such as MapUtopia are generally at a disadvantage to most other websites in terms of search rankings because they do not publish much text, which is what drives search rankings. By using tooltips that contains text related to the purpose and functionality of the website, the disadvantage due to a lack of text can be slightly mitigated.

Accessory Products that Leverage the Database

If the previously mentioned ad-revenue is insufficient, it is anticipated that a related product can be sold on the site's landing page,

www.maputopia.com . Using the traffic drawn from the free portion of the website could drive sales of related pay-per products such as demographic reports.

Works Cited

Websites

Block, Matthew, and Harrower, Mark. "MapShaper.com." MapShaper.com .
mapshaper.com/test/MapShaper.swf (accessed February 3, 2012).

Kobben, Barend. "(deprecated) RIMapper." Temporary server .
http://kartoweb.itc.nl/RIMapper/rimapper_old.html (accessed
February 3, 2012).

Oracle. "MySQL :: MySQL 5.5 Reference Manual :: 11.17 Spatial
Extensions." MySQL :: Developer Zone.
<http://dev.mysql.com/doc/refman/5.5/en/spatial-extensions.html>
(accessed February 3, 2012).

Oracle. "MySQL :: The world's most popular open source database." MySQL
:: The world's most popular open source database.
<http://www.mysql.com> (accessed February 3, 2012).

"Notepad++ Home." Notepad++ Home. <http://notepad-plus-plus.org/>
(accessed February 3, 2012).

"PHP: Hypertext Preprocessor." PHP: Hypertext Preprocessor.
<http://www.php.net> (accessed February 3, 2012).

"World Wide Web Consortium (W3C)." World Wide Web Consortium (W3C).

<http://www.w3.org> (accessed February 3, 2012).

Apache Friends. "apache friends - xampp." apache friends - very easy apache, mysql, php and perl installation without hassles.

<http://www.apachefriends.org/en/xampp.html> (accessed February 3, 2012).

Garrett, Jesse James. "Ajax: A New Approach to Web Applications - Adaptive Path." Adaptive Path. N.p., n.d. Web. 18 Feb. 2012.

<<http://www.adaptivepath.com/ideas/ajax-new-approach-web-applications>> .

"Geolocation - Google Maps API Family - Google Code." Google Code. N.p., n.d. Web. 18 Feb. 2012.

<<http://code.google.com/apis/maps/articles/geolocation.html>> .

"Mapping America" Census Bureau's 2005-9 American Community Survey." New York Times. N.p., n.d. Web. 15 Feb. 2012.

<projects.nytimes.com/census/2010/explorer> .

"Mapping the 2010 U.S. Census." New York Times. N.p., n.d. Web. 15 Feb.

2012. <projects.nytimes.com/census/2010/map> .

"Piton Foundation 2010 Census Project." Home. N.p., n.d. Web. 19 Feb.

2012. <<http://www.piton.org/census2010/>> .

"Places Library - Google Maps JavaScript API V3 - Google Code." Google Code. N.p., n.d. Web. 18 Feb. 2012.

<<http://code.google.com/apis/maps/documentation/javascript/places>.

html#places_autocomplete>.

"Demographic Maps and Reports." Social Explorer. N.p., n.d. Web. 19 Feb. 2012. <<http://www.socialexplorer.com/pub/maps/home.aspx>>.

Books

Boehm, Anne. *Murach's HTML, XHTML, and CSS: training & reference*. Fresno, Calif.: Mike Murach & Associates, 2010.

Harris, Ray. *Murach's JavaScript and DOM scripting*. Fresno, CA: Mike Murach & Associates, 2009.

Lengstorf, Jason. *Pro PHP and jQuery*. New York, N.Y.: Apress :, 2010.

Murach, Joel, and Ray Harris. *Murach's PHP and MySQL: training & reference*. Fresno, Calif.: Mike Murach & Associates, 2010.

Purvis, Michael, Jeffrey Sambells, and Cameron Turner. *Beginning Google maps applications with PHP and Ajax from novice to professional*. Berkeley: Apress ;, 2006.

Svennerberg, Gabriel. *Beginning Google Maps API 3*. New York: Apress, 2010.

Udell, Sterling. *Beginning Google Maps mashups with maplets, KML, and GeoRSS: from novice to professional*. Berkeley, CA: Apress ;, 2009.

Viney, David. *Get to the top on Google: tips and techniques to get your site*

to the top of the search engine rankings - and stay there. London:

Nicholas Brealey, 2009.