Visual Analytics for Open Government Data

Francisco Artigas
Meadowlands Environmental Research Institute
New Jersey Meadowlands Commission
1 Dekorte Park Plaza, Lyndhurst New Jersy. 07071
1 202-460-2801
Francisco.Artigas@njmeadowlands.gov

Soon Ae Chun
CUNY College of Staten Island
2800 Victory Blvd
Staten Island, NY 10314
1-718-982-2931
Soon.chun@csi.cuny.edu

ABSTRACT
The Open Government Data initiative accelerated publishing and sharing of diverse government data with enterprises, software developers and citizens. The key challenge with this unprecedented amount of government data is to make sense of them in a timely manner i.e. find the meaning of the data to assist in decision making. The analytical tools and apps to filter, sort, aggregate for summary along different spatio-temporal dimensions will help find trends and outliers. The capabilities to combine diverse data from different sources may also help users to understand the information and knowledge embedded in data sets. One of most powerful data exploration and analytics is data visualization. In this tutorial, we present two different tools for visual exploration and analytics for citizens and researchers, making use of diverse data sets from different sources, such as government, social media and enterprise or personal collections. Specifically, we focus on Geographic Information Systems (GIS) for mapping data using ArcGIS Explorer, a cloud based data integration and visualization tool and Google Fusion Tables for tabular data. The tutorial is geared toward social science researchers, students and citizens in general.

Categories and Subject Descriptors
H3.5 [**Online Information Services**]: data sharing, H.5.2 [**User Interfaces**]

General Terms
Visualization, Visualization Systems and Tools, E-Government

Keywords
GIS Geographic Information Systems, Visualization, Visual analytics,. Tutorial, Google Fusion Tables, Google Chart Tools .

1. INTRODUCTION
Geographical Information Systems offer a variety of methods and tools to communicate geographically using spatially explicit data. Researchers may be familiar with GIS's visually appealing results and its analytical powers and may have ideas on how to apply this tool to their own research but have not had the chance to learn how to use GIS. This tutorial offers a brief introduction to ArcGIS Explorer Desktop which is a free and new powerful GIS application. After a brief introduction on map making the tutorial moves quickly to hands-on learning of the use of this tool to create maps and conduct spatial analysis based on pre-existing GIS data layers available on the WWW and on government sites. The hands on exercises have increasing levels of complexity as each explores a greater range of functionalities offered by the application. As a result, students walk away having experienced a range of functionalities from importing data, overlaying data, performing spatial analysis and preparing layouts and presentations to show findings. Students gain enough insight and practice to apply these lessons to their own problems in their corresponding fields of expertise. The second set of exercises geared toward visualization of the structured and semi-structured data using light-weight integration and charting tools such as Google Fusion Tables and Google Chart Tools.

dg.o 2013, Jun 17-20 2013, Quebec City, QC, Canada
ACM 978-1-4503-2057-3/13/06.

2. VISUAL ANALYSIS OF MAPPING DATA SETS
A Geographic Information System (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. GIS is a computer system that allow users to analyze spatial information, edit data in maps and represent the result of interactive queries in the form of geographic maps or visualizations. GIS can relate unrelated information by using location as the key index variable. Locations in Earth space–time may be recorded as dates/times of occurrence, and x, y, and z coordinates representing, longitude, latitude, and elevation, respectively. GIS data represents real objects on the surface of the earth such as roads, parcel properties and trees.

Traditionally there are two broad methods for storing and analyzing GIS data: raster images and vectors. Satellite images and aerial photographs are usually represented in a raster system where each pixel holds the X and Y coordinate information plus a Z value (color of the pixel) that represents the attribute of the pixel such as texture type, land use type, elevation, etc. A vector system on the other hand uses points, lines and polygons to represent objects on the surface of the earth. Objects where the area is not important such as a telephone pole location, voting stations, crime locations etc., are represented by a point. Roads, elevation curves, travelling trajectories etc. are represented by lines and lakes, land use types and objects where the area is important are represented by polygons.

We introduce lightweight interactive GIS tools such as ArcGIS Explorer Desktop [1] and Geo Viewer [2] for geodata search, exploration and visualization with overlays with other data sets. The particular datasets will include: geospatial data sets from Data.gov and other sources, such as USGS hydrology data, emergency management data, Crimes data , Environmental data, US Census data, etc.

2.1 Methods

This GIS-based visual analytics tutorial is divided into four sections. In the first section students are presented with the basics of GIS data types, reference datum and map projections. These topics are touched upon only briefly and just enough to understand how to look for and represent compatible data. The second section, explores the different functionalities of ArcGIS Explorer Desktop. Student learn how to navigate the application and how to find and use some of the most important tools such as adding data, overlaying GIS layers, assigning legends and symbols, performing spatial analysis via queries and buffer operations and finally creating layouts and presentations to show the results. The last section consist of three specific exercises with increasing levels of difficulty that forces students to explore and use wide range of GIS tools.

3. VISUAL ANALYSIS OF TABULAR DATA SETS

Government and personal data are structured data sets such as Spreadsheet data, CSV (Comma Separated Values), or the XML semi-structured data sets such as KML (Keyhole Markup Language). We introduce visualization tools that are made for non-technical people to search, integrate and analyze data with visualization. Specifically, participants will explore Google Fusion Tables and Google Public Data Explorer Tool.

Google Fusion Tables [6] is a cloud-based service to gather data from different sources, upload them, integrate them with possibility of human collaborations, and visualize them. The data manipulation and visualization tools include charts, maps, timelines, heat maps, etc. Google Chart Tools [5] allow users to create various charts, e.g. line, pie, bar charts as well as more advanced visualization of data. In addition, the tool generates embedded codes that allow sharing charts on their web pages. Data sets will be gathered from Google Public Datasets [3] and data sets from the Data.gov site [4].

4. RESULTS

As a result of this tutorial, participants should be able to find and import spatially explicit data into ArcGIS Explorer Desktop. They should be able to manipulate the imported data into concise visualizations by organizing data into layers and assigning symbols and legends. Participants should be able to conduct spatial queries by using the query tool to select subsets of objects or objects that have the same attribute and/or intersect with a different layer. They should be able to use the buffer tool to select and map features that fall within a given distance from an original location. Finally, they should be able to organize their findings into well labeled and balance layouts and presentations to show the final results. Similarly, the participants search and upload open government data to Google Fusion Tables or Google Chart Tools for combining, manipulating and visualizing as well as sharing the visualization results on the Web. Our objective is to allow people with minimal technical background be able to explore and analyze a vast array of data through integration and visualization to get meaningful insights as well as detect significant facts or events quickly using state-of-the-art visual analytic tools.

5. AUTHOR BIOS

Dr. Francisco Artigas is the Director of the Meadowlands Environmental Research Institute (MERI) at New Jersey Meadowlands Commission, a New Jersey state agency in charge of environmental and urban planning of the Meadowlands District of New Jersey. He is also an adjunct faculty member of the Earth and Environmental Sciences and the Biology Department of Rutgers University-Newark. He earned his Ph.D. degree in Environmental Science at the Ohio State University specializing in Geographical Information Systems, ecosystem ecology, hydrology, soil sciences and computer modeling. The GIS group of his Institute supports GIS training, mapping and planning, land use and urban development for 14 Municipalities. He served as a Research Associate Professor at the Rutgers University Center for Information Management Integration and Connectivity (CIMIC) where he coordinated the activities of the Rutgers University NASA Regional Application Center. Sponsored by the U.S. Agency for International Development, the International Development Bank and the European Economic Community, he has worked on development of environmental, agricultural and food security oriented Geographical Information Systems in Tanzania, Brazil, El Salvador and Bolivia.

Dr. Soon Chun is an Associate professor of Information Systems and a doctoral faculty member in Computer Science at City University of New York. She is currently on sabbatical leave at Columbia University affiliated with the security research group. Her research interests include Semantic Web, Workflow and Web Services, Security, Privacy and Data integration. Her research application areas include digital government, health informatics, environmental science and e-learning. Her research in digital government includes the ontology based service composition and automation using workflow technology, the security and privacy framework for government information sharing, and the semantic information integration. Her research has been sponsored by NSF, NOAA and New Jersey State agency and CUNY PSC. She serves as a board member of the Digital Government Society and has served as conference chairs and organization member for the digital government. She also serves as organization member and PC member for many conferences in information science field.

6. ACKNOWLEDGMENTS

We acknowledge the NJMC for making their GIS data available for the tutorial.

7. REFERENCES

[1] ArcGIS Explorer Desktop: http://www.esri.com/software/arcgis/explorer

[2] Data.gov Geo Viewer: http://www.data.gov/new-features

[3] Google Public Data: http://www.google.com/publicdata/directory#!st=DATASET

[4] Data.gov https://explore.data.gov/

[5] Google Chart Tools: https://google-developers.appspot.com/chart/interactive/docs/quick_start

[6] Hector Gonzalez, Alon Halevy, Christian S. Jensen, Anno Langen, Jayant Madhavan, Rebecca Shapley, and Warren Shen. 2010. Google fusion tables: data management, integration and collaboration in the cloud. In Proceedings of the 1st ACM symposium on Cloud computing (SoCC '10). ACM, New York, NY, USA, 175-180.