Visualizing World Bank Indicators through Google Earth

William Murakami-Brundage, M.S.; Jennifer Bopp, M.A.;
Megan Finney, M.S.; Joselito Abueg, M.A.

ABSTRACT

Abstract— The purpose of this project was to develop a large visual data resource for Google Earth using major education, gender, and health datasets. With global data increasingly being made public by organizations such as the World Bank, global data modeling has been a significant development in information visualization and geographical information systems. While there is a considerable amount of publicly owned and open-source data sets available, there has been minimal development beyond proof-of-concept ideas. The research project modeled four major domains of the World Bank's global datasets: Education, women and gender issues, health care access, and life expectancy and wellness, with each comprised of 15 or more sub-elements. A data translation key has been developed for data migration into Google Earth's KML format. The resulting KML files will be made available for public use. It is hoped that a greater global awareness will develop by using the World Bank/Google Earth data. Additionally, data development will be easier once the data key is published.

Index Terms— Data and knowledge visualization, Spatial databases and GIS, Visualization systems and software

BACKGROUND

The mission of the World Bank is to reduce poverty by supplying resources and knowledge that lead to improvements in infrastructure and sustained development. To achieve this goal, the World Bank engages in partnerships with public and private sector entities and promotes awareness of issues related to global development. The World Bank publishes reports, statistics, and data sets to support program management, accountability, and academic research. These include over 2,000 economic and human development statistics for 209 countries. In April 2010, the World Bank made most of this data available on the Internet under its Open Data initiative.

On its Labs site, Google has made available some of the World Bank's indicators as well as several visualization tools. Google Earth's compelling interface and the fact that its software and data are freely available for different platforms, make it an important, educational tool to promote global awareness.

OBJECTIVES

Our primary goal is to make the World Bank indicators readily available to users in a compelling form. We explore ways in which the user could visualize the data without the need for programming and with tools that are freely available.

Starting with the *GenderStats* data collection, we first modeled indicators for *Expected Years of Schooling, Male; Expected Years of Schooling, Female;* and, *Gross National Income per Capita, Atlas Method* [1].

One final goal for this project is to establish a streamlined and cost-free way to convert data into a KML format so that the technology is accessible to a wider user population.

MATERIALS & METHODS

```
<Placemark>
 <name>Zambia - 970</name>
 <Style>
  <IconStyle><Icon>
    </Icon></IconStyle>
  <LabelStyle><color>ff00ff80</color><scale>0.7</scale></LabelStyle>
 </Style>
  <Point>
    <altitudeMode>relativeToGround</altitudeMode>
    <coordinates>27.84,-13.13, 79104.3</coordinates>
  </Point>
 </Placemark>
</Folder>
</Document>
</kml>
```

Fig. 1. The text fragment above showed the end of the Gross National Income per capita file. Each World Bank Indicator (e.g., GNI per capita) data set (values by country) is geocoded such that a specific location for each county is determined. The geocoded data are then encoded into a Keyhole Mark Language format document, which is readable by Google Earth.

During the course of data transformation, the developer discovered an incompatibility of ISO country codes. World Bank data identifies countries based upon the ISO 3166-3 standard of three letters (e.g., USA). Google Earth labels countries according to the ISO 3166-2 standard of two letters (e.g., US). As a solution, a master index was created in Microsoft Access.

One of the project's major goals also proved to be one of the major challenges: discovering publicly available free software to perform data processing and transformation. The intention is to inspire open-source systems.

A Brazilian application provided a solution. GE-Graph is a publicly available "freeware" (free of charge) program developed by Ricardo Sgrillo of the Escola Superior de Agricultura Luiz de Queiroz (ESALQ) of the University of Sao Paolo (http://www.sgrillo.net/googleearth/gegraph.htm). GE-Graph is intended to aid in the import and export of KML data, the file format used by the Google Earth program.

Once processed, GE Graph output the *GenderStats* data in KML format. As previously discussed, KML (Keyhole Markup Format) is a geographically enhanced extension of the XML (Extensible Markup Language) standard.

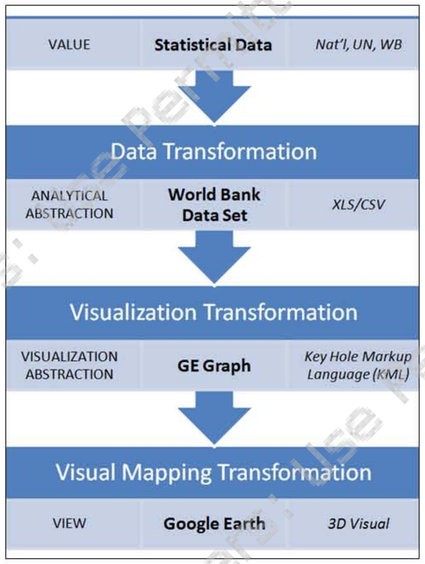

Fig. 2. The Data State Model applied to the World Bank visualization project. Work by project members begins at the Analytical Abstraction stage. [Legend: Uppercase-DATA STAGE, Bold-**Object/Tool** used in transformation. Italics-*Format of data presented*]

RESULTS

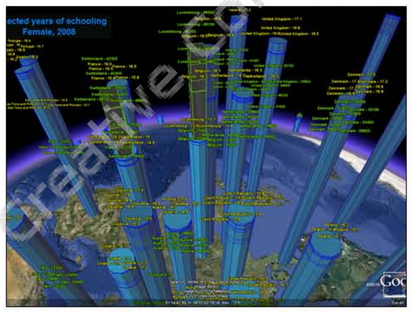

Figure 3: A display of national GNI and expected years of schooling for females. The date range is 1999-2009, with the label showing the year 2008. Income is labeled with green text, while years of schooling have yellow text. Even with the graphic difference between variables, it is impossible to differentiate between years without referring to the data set itself. General trends, such as lower income in Eastern Europe, can still be identified.

The World Bank, anticipating the importance of data visualization, provided ways to map and graph the data. However, the limited methods by which these data are represented are not without issues.

Overlaying data on a map makes it easy to visualize the importance of location/country on the variable of interest, and is an obvious choice of data coded by country.

That said, maps are understood to carry with them certain biases that can influence how their users perceive the world [2]. With respect to the mission of the World Bank, centering a flat map on the African subcontinent or other regions of the developing world might yield greater impact.

The project developer explored different ways of representing indicator data. Three dimensional rods were chosen from a variety of options for this project. The height of each rod correlated to the value of the variable processed by the GE-Graph program (e.g., income, education level). Due to development tools, graphics were restricted to simple polygonal shapes, but height, width, and color could be altered when data processing occurred.

Fig. 4. Close-up of Gross National Income (GNI) per capita in 2009. Note the high incomes in Europe relative to North Africa and the Middle East. Also, note the difficulty in seeing tall rods from directly above.

CONCLUSIONS

Using a virtual model of the world such as Google Earth allows the World Bank data to be displayed in geographic context and should provide a greater understanding of that data through locating it. While it is beyond the scope of this project to evaluate the benefits of virtual spaces (3D representations), we suspect that the ways in which a virtual world matches our mental model would make a stronger sense of space and context than would a two dimensional, static model.

Shneiderman's [3] visual information seeking mantra, "Overview first, zoom and filter, then details-on-demand" is instructive for considering what would benefit these visualizations. The overview may be the most successful aspect of the visualization; as the user turns the globe, the variable heights of the rods give a general impression of different regions of the globe. The user is able to filter by checking or un-checking years or countries in the data folder, though this can be cumbersome.

Displaying time series data proved to be another significant difficulty. The user can go into the data folder and check or uncheck as many years of data as desired. If multiple years are displayed, the labels positioned along the rod appear somewhat chaotic. In the case of closely positioned countries, the density of labels makes them nearly illegible. Making sense of the information conveyed by the labels is further complicated by the fact that the corresponding year is not included with each.

Programming the dataset with Google Earth's timespan feature is the next step for the World Bank variables. The first step to modeling data over time is to develop all the necessary years in Google Earth. After this, the data will be joined together to form a cohesive, timespan-capable file.

REFERENCES

References

1. The World Bank (2010, December). *GenderStats* [Online]. Available: http://data.worldbank.org/data-catalog/gender-statistics
2. J.B. Hartley, "Deconstructing the Map," *Cartographica: The International Journal for Geographic Information and Visualization*, vol. 26, no. 2, pp. 1-20, Summer 1989.
3. B. Shneiderman. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," In *Proc. of IEEE Symposium on Visual Languages*, Los Alamos, pp. 336-343, 1996.

ACKNOWLEDGEMENTS

Acknowledgements

We would like to acknowledge the iSchool at Drexel University, especially Prof. Chaomei Chen. Without his guidance, this work would not have been possible.

Disclaimer

The official terms of usage for the World Bank datasets can be found at http://go.worldbank.org/OJC02YMLA0. This research takes advantage of the World Bank's generosity in sharing their data sets for public use, and is not affiliated with the World Bank in any way.