Where in the World? Demographic Patterns in Access Data[1]

Mimi M. Recker[1], Beijie Xu[1], Sherry Hsi[2], and Christine Garrard[3]
mimi.recker@usu.edu, beijie.xu@aggiemail.usu.edu
sherryh@berkeley.edu, chrisg@leupold.gis.usu.edu
[1]Department of Instructional Technology & Learning Sciences, Utah State University
[2]Center for Technology Innovation, Lawrence Hall of Science, UC Berkeley
[3]RS/GIS Laboratory, Utah State University

Abstract. Standard webmetrics tools record the IP address of users' computers, thereby providing fodder for analyses of their geographical location, and for understanding the impact of e-learning and teaching. In this paper, we describe how two web-based educational systems were engineered to collect geo-referenced data. This is followed by a description of joining these data with demographic and educational datasets for the United States, and mapping different datasets using geographic information system (GIS) techniques to visually display their relationships. We conclude with results from statistical analyses of these relationships that highlight areas of significance.

1 Introduction

Web-based learning environments can be engineered to collect detailed data about their users and their activities (generally called *webmetrics*). In addition, standard webmetrics tools record the IP address of users' computers, thereby providing fodder for analyses of their geographical location, and for understanding the impact and the effectiveness of e-learning and teaching. In our study, we used two web-based learning environments – the Instructional Architect (IA) and the Exploratorium Learning Resources Collection (ELRC) – to explore how to collect geo-referenced data, and analyze such datasets by joining them with the demographic and national educational datasets, and using geographic information system (GIS) to visually display the relationships.

2 Collecting Geo-Referenced Data

The IA is an educational digital library service developed to support the instructional use of the online learning resources. With the IA, teachers are able to create web-based instructional activities, called IA projects, and give their students exclusive access to them. The ELRC is a digital library of over 700 teacher-tested science activities and instructional resources inspired and created from the Exploratorium's exhibits, public program events, and teacher professional development programs. Both websites have been engineered to collect detailed online usage data using Google Analytics (GA). As part of their standard reporting tool, GA estimates the visitors' location using the client computers' IP addresses, in a process called geo-location [1].

[1] For full paper, figures, and references, please visit http://edm.usu.edu/publications/demo.pdf.

3 Results

The GA reporting tool generates geographic maps of website visitors. The IA's map has darker shades in Utah, New York and Michigan, three areas that have received IA's teacher development programs; similarly, the ELRC's map has darker shade in California, where the ELRC library locates. The maps show that both groups are successful in local dissemination activities.

GA location data can be joined with other demographic and geographical datasets. In this study, we extracted population, number of schools, number of school districts, per capita income, and median family income from the ESRI 9.3 GIS software, and the National Center for Education Statistics. These were then overlayed with our GA location data. For example, Figure 1 shows IA traffic overlayed on one of the datasets –median family income. Such maps can help visually reveal relationships between site usage and demographic or school characteristics.

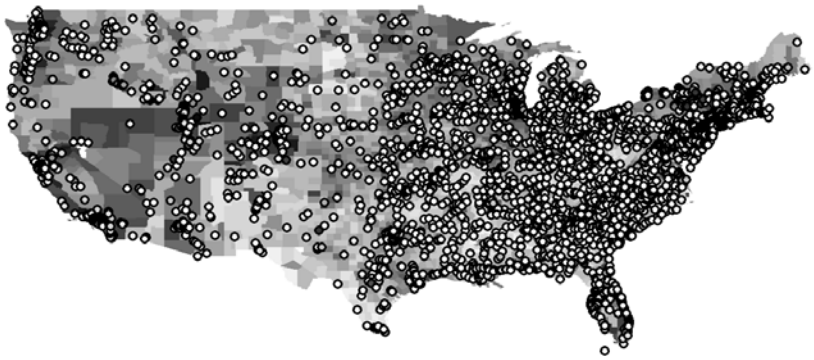

Figure 1. U.S. map showing IA visits (darker dots indicate more visits) overlayed with median family income over 1 year.

We also examined statistical relationships between the demographic predictors and the number of site visits per location as reported by GA. Due to the high correlation between some of the predictors, only three out of the five were selected, which were population, number of school districts, and per capita income. A negative binomial regression (to handle the skewed data) showed that *population*, and *per capita income* were statistically significant predictors of site visits to both the IA and the ELRC, and *number of school districts* was also a significant predictor for the ELRC.

We interpret these results to mean that online visitors to these sites came from, not surprisingly, more densely populated areas. In addition, the relationship with per capita income may be a function of the amount of resources (i.e., computing) available in the local schools and communities.

References

[1] Muir, J. A., & Oorschot, P. C. V. Internet Geolocation: Evasion and Counterevasion. *ACM Computing Surveys (CSUR)*, 2009, 42(1). Article 4.