

DataPlace: Exploring Statistics about Cities

Karen Theisen
Vinq, LLC
560 S. Winchester Blvd.
San Jose CA 95125
Karen@vinq.com

Nancy Frishberg, Ph.D.
Vinq, LLC
560 S. Winchester Blvd
San Jose CA 95125
nancyf@acm.org

ABSTRACT

DataPlace™ is a website created by Vinq LLC under contract to Fannie Mae Foundation as part of its KnowledgePlex® initiative. The intended user population includes specialists in affordable housing and community development. DataPlace.org puts a full-featured front end (interactive maps, charts, ranking at several levels of geography, and area overviews of cities, states and other jurisdictions) on U.S. Federal statistical databases, including two decennial censuses, multiple years of Home Mortgage Disclosure Act data, among other resources. For the CHI2007 workshop “Imagining the City”, we review what we’ve learned about actual users, situations of use, actual and anticipated changes to the interface and website functionality.

Author Keywords

Statistics, data visualization, US Government, mapping, charting, user identification, interface revision

ACM Classification Keywords

H5.2. User Interfaces

WHAT IS DATAPLACE?

Datplace.org is a website which allows both specialists and non-specialists to look interactively at standard statistics about housing and demographics in the United States, get a profile of one or more places, and compare specific places against one another. DataPlace offers the opportunity to inspect maps of regions as small as census tracts or zipcodes, and as large as a state or the whole of the US, as well as cities, metropolitan areas, counties, and combinations of these geographies. Choropleth maps use shading and color to represent numerical data about particular indicator (as Figure 1). Users can choose to view

barcharts and scatterplots of several indicators at once or values for the same indicator over several years or across several geographic entities. DataPlace’s several types of representation have been influenced by the work of Edward Tufte [1], and draw on best practices in automated natural language text generation. Further information on some of the principles and the underlying technology for DataPlace are available [2].

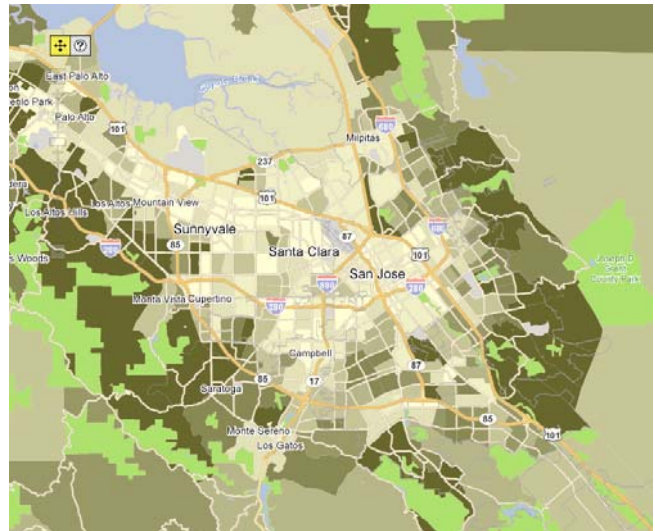


Figure 1: Map of Homeownership Rate for 2000 in San Jose, CA

Today users can insert a place name (San Jose, San Jose Metropolitan Area, Santa Clara County, 95125, California) in the autocomplete search box available on every page at any time and the system will return a profile of the selected geographic entity. Figure 2 shows that Area Overview page for San Jose, CA with a small map (in the upper left), 2 paragraphs of automatically generated text giving a few statistical highlights that distinguish this place, a set of 5 suggestions of what to do next (toward the right), and below a set of approximately 40 indicators with their numeric values and sparkline histograms [3]. These 40 indicators are editorially selected out a larger set of several thousand indicators from the data sources mentioned above.

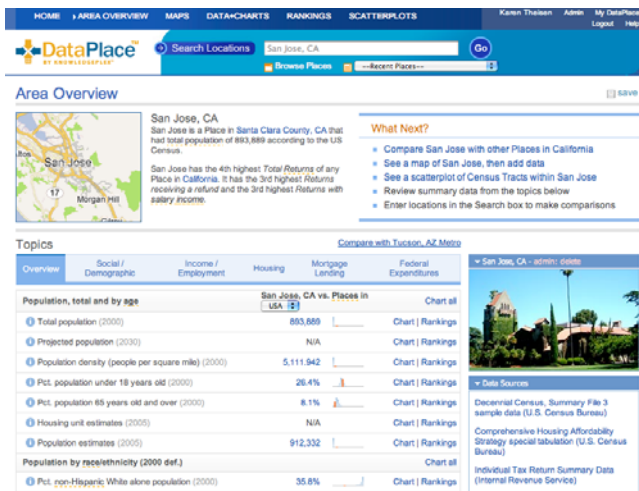


Figure 2: Area Overview for San Jose, CA

The small map is clickable, taking the viewer to a full-sized map of the region. Each of the suggestions under the heading “What Next?” is a link that brings the viewer to a view of data related to the target place.

The numerical data on the lower part of the page when clicked opens to a choropleth map showing values for the selected indicator at the target area and surrounding region. The sparkline histogram expands on mouseover to show a larger-sized histogram labeled and indicating the position of the selected geography when considered within 20 segments across the US (as in Figure 3). Additional tabs (besides Overview) offer quick views of 5 additional categories of roughly 40 indicators each (Social/Demographic; Income/Employment; Housing; Mortgage Lending; Federal Expenditures), giving the viewer at least 200 indicators within two or three clicks of their initial encounter with DataPlace.

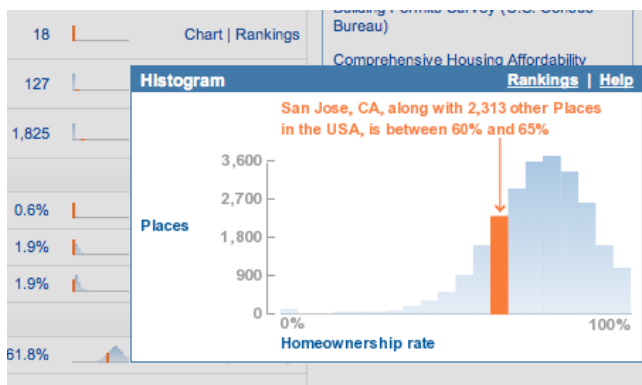


Figure 3: Histogram of Homeownership Rate in 2000 for San Jose, CA

DataPlace developers have responded to requests from expert users (demographers, urban planners, community development practitioners) and general users (business people, staff of various non-profit agencies, government employees) to include functionality such as

- Saving frequently used indicators, frequently used geographic entities, and charts, maps or other complex user-created constructions for later citation or sharing.
- Publishing specific maps or charts to a website beyond DataPlace. The publication at another site loses the interactive capabilities of the website, but allows for inspection of the static data.
- Creating “custom geographies”, i.e., user-defined geographic entities (e.g., states grouped into Federal regions; the 14 counties served by a regional foodbank; zipcodes with a criterion level of Spanish speakers within a metropolitan area). These custom geographies encourage user-defined, custom statistical summaries to be created.
- Comparing user-designated data with the validated US Census or other Federal data provided by DataPlace: “Dataset Uploading” allows a user to associate local statistics to local geographic boundaries, or make use of statistics from sources that DataPlace has not incorporated. So long as those boundaries match the census tract or zipcode boundaries (or other standard boundaries) that DataPlace already accounts for, the data can be examined with the existing tools of DataPlace: mapped on the maps provided, compared with the standardized statistics found at DataPlace.

WHAT DO WE KNOW ABOUT DATAPLACE USERS

DataPlace users are invited to register, in order to receive the weekly newsletter from its companion site, KnowledgePlex.org, and in order to save their work. By inspecting the database of registered users, we confirmed some expectations and gained some further insights into what differences there are between users who first found KnowledgePlex and those who first registered at DataPlace. Our reasoning about users derives from 4 kinds of information: the user’s job title, the organization employing them, their email address for the registration, and the role(s) a user chose from our fixed list.

Research conducted in late 2002 just before DataPlace’s launch (but after KnowledgePlex had been available for over a year) suggested that there are 4 primary audience segments:

- Policy Makers (legislators and their staffs, government agency executives; municipal administrators; regulator officials’ community advocates)
- Practitioners (Community-based organization executive directors and program officers; Federal level agency officers; bank or credit union loan officers, real estate agents or specialists, planners)

- Scholars and Researchers (Professors, graduate students, research director or research analyst at a not-for-profit organization or government agency)
- News Media Professionals (journalists, producers, communication staff for agencies)

Our hypothesis that DataPlace is attracting a more research-oriented audience, where KnowledgePlex attracts general non-profit staff or government officials, is confirmed by our analysis of the registration databases.

Registration Analysis

Registration uses the same forms and inserts the information into the same database from either DataPlace or KnowledgePlex. (A notation is made telling us which site the individual used to register from.) Users are asked both their name and email address, as well as job title and organization. In addition, users are asked to select one or more categories that fit their role.

DataPlace registrants were more likely to have job titles related to research: Planner, Researcher, Educator, Student, Economist, than were KnowledgePlex registrants. In addition, those with Marketing in their job titles were more likely to have registered at DataPlace.

A large number of people who registered with DataPlace or KnowledgePlex chose a neutral email address (such as hotmail.com, junio.com, yahoo.com, earthlink.net or gmail.com), rather than one associated with their workplace, for their subscription. Of the remaining addresses, users with .edu addresses were slightly more likely to have registered at DataPlace. Those from the non-profit realm (.org addresses) were more likely to have registered at Knowledge.Plex.

Thus from the job title analysis and the email address identification (limited though it is), we believe that DataPlace is attracting an audience oriented toward research.

SCENARIOS OF USE

We have conducted interviews and observations of prospective and actual DataPlace users over the past 2 years. In addition we have conducted at least 3 formal usability studies of specific functionality. In each case we draw most of the participants from the database of registered users, and from those who have communicated via the feedback forms on the website. We find a wide range of desired and actual uses, including many beyond the expectations of the original concept.

A scenario typical of the anticipated uses is the city official who is trying to demonstrate how his town's affordable housing stock compares in quantity and type to other cities of similar size in his state and in other states.

A few other scenarios include

- A grant writer for a public school system who wants to track the ethnicity and language used at home of the school-age population, in order to support grants she's writing
- A Habitat for Humanity research staff member who is getting support for the local chapter's work, and making sure that the building projects being undertaken are sited in locations of greatest need
- A credit union investigating where to underwrite mortgages and support redevelopment efforts within its state
- A business which is looking for opportunities to develop "brownfields" sites (environmentally compromised land), which needs information about the zoning and types of businesses around the target sites
- A city employee who wants to compare each council member's district across various statistical indicators (employment, household size, Federally supported housing units, etc.)

ANTICIPATED FUTURE DEVELOPMENTS

While we cannot fully predict where DataPlace will grow, we are currently designing options for searching by topic as well as by geographic entity, or both. The addition of topics ("welfare recipients", "substandard housing") implies a new navigation schema that allows for both topics and for visualizations.

DataPlace invites each user to save their own work, and this is the key motivator for registering at DataPlace. In a recent development, we have opened functionality of creating and communicating in groups, including the ability to publish one's results or work-in-progress to a group. In the initial release, all groups are designated as "private", so that the existence of the group and its specific communications are not revealed to other DataPlace users. Groups might be a workgroup within a city government, where several people want to review data or presentation materials in advance of sharing them more widely. Groups may also attract people who share a specific role or task, but work at different agencies or levels of government. We anticipate that with a bit of experience, groups may choose to make their communications public, or will share specific work developed using DataPlace in a public gallery. Such galleries are likely to inspire both greater collaboration and greater experimentation.

In addition, DataPlace is considering broadening its reach to include topics beyond housing and community development to encompass all categories affecting municipal life (health, nutrition, safety and justice, education, environment, etc.). The addition of an increased range of topics implies new ways of organizing and navigating topics for easy access to each semantic category. As Dan Russell (of Google) has been saying about search in

general “we’re in the business of mind-reading” [4]. DataPlace is about to enter this business also.

The challenge for us is creating a highly interactive site that is self-instructing, which reveals its capabilities sufficiently, and which is easy enough for newcomers, but has the functionality that more experienced users want and can appreciate.

ACKNOWLEDGMENTS

We acknowledge the vision of Troy Anderson of Fannie Mae Foundation in conceiving of the resource that DataPlace has become, and creating the support and team to make it happen, our FMF and Urban Institute colleagues for their contributions in realizing that vision every day. We also acknowledge Vinq’s founder and CEO, Mark Torrance, whose implementation of the vision we have lightly touched. Mark involved many others at Vinq over the course of the past 3 years, and we have benefited greatly

from the mapping and GIS expertise of the PlaceBase team. Any errors that remain in this work are our own.

REFERENCES

1. Tufte, E.R. *Visual Display of Quantitative Information* (2nd edition). Graphics Press, Cheshire, CT USA. 2001.
2. Torrance, Mark. Rapid Data Publishing. *Urban Markets Initiative Forum on Information and Urban Markets*. February 2006.
http://www.vinq.com/umi/rapid_publishing.pdf
3. Tufte, E.R. *Beautiful Evidence*. Graphics Press, Cheshire, CT USA 2006.
4. Russell, Daniel M. How do People Use Search Engines? Using Multiple Kinds of Data to Understand What People Do when Searching. ACM SIGCHI San Francisco Bay chapter meeting of December 12, 2006.
<http://www.baychi.org/calendar/20061212/#2>.