A Web-based Geographic Information System (GIS) for Visualizing Cancer Disparity with Socioeconomic and Demographic Variables

Su Yeon Han[1,2], Ming-Hsiang Tsou[1,2]*, Atsushi Nara[1,2], Joseph Gibbons[1,3], Sindana Ilango[4], Caroline A. Thompson[4]

[1]The Center for Human Dynamics in the Mobile Age, San Diego State University, San Diego, USA, [2]Department of Geography, San Diego State University, San Diego, USA, [3]Department of Sociology, San Diego State University, San Diego, USA, [4]Division of Epidemiology in the Graduate School of Public Health, San Diego State University

OBJECTIVES

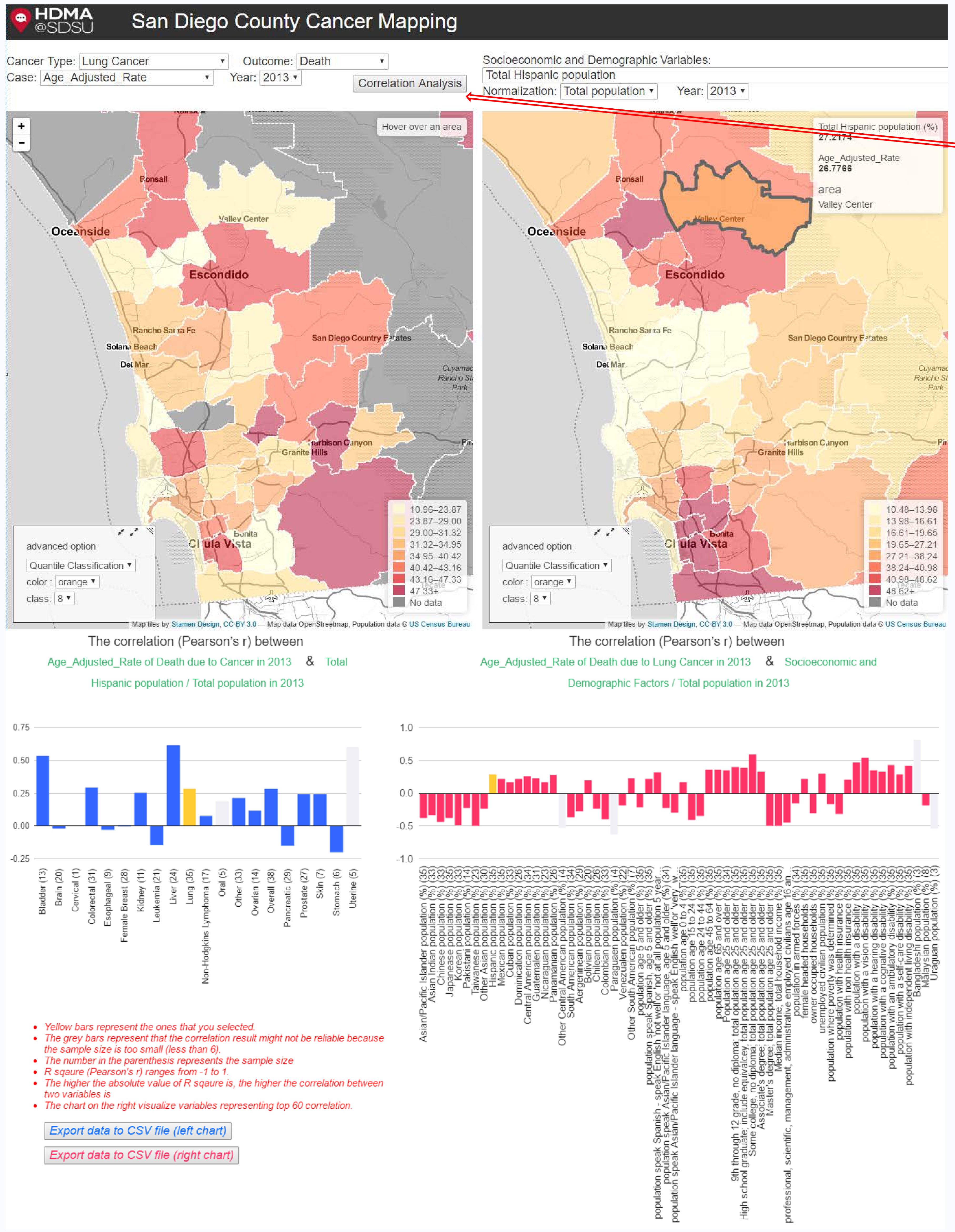
This project aims to develop a highly-interactive and user-friendly Web-based mapping application to visualize mortality and hospitalization data for 19 cancer sites in each Subregional Areas (SRAs) of San Diego County during 2010 to 2013. The application allows cancer researchers to synchronously explore 23 different types of cancer mortality rates on the left side and 96 socioeconomic and demographic factors on the right side, and help them to explore interrelationships between cancer outcomes and socioeconomic factors by computing and visualizing spatial and statistical correlations.

METHODS

Given the high dimensionality and complexity of the various cancer outcomes and 96 socioeconomic and demographic variables, our approach is based on the principles of exploratory data analysis (EDA). The principle of EDA is to let data speak for themselves free of theory by imposing no *a priori* hypothesis [1]. EDA aims to discover "potentially explicable patterns" of data [2] by utilizing statistical tools and information visualization such as cartographic maps, tables, histograms, scatter plots, and charts [3]. The results of EDA can guide users to suggest explanations, create formal hypothesis and theoretical constructs, and present data in a form that is easily understandable [4].

In population health research for cancer, a serious and often fatal disease, data are often aggregated in the same geographical unit or suppressed due to privacy issues. To make a good model for prediction of the geographical distribution of cancer outcomes, the locations of individual cancer patients without aggregation often are needed. Even if a model can be developed based on the aggregated data, the model may fail to predict the distribution of cancer outcomes at the desired spatial resolution. In this particular situation, exploratory data analysis can be used instead of constructing a prediction model [5].

FINDINGS



The correlation analysis is performed after this button click

Synchronous Visualization of Cancer Mortality and Socioeconomic & Demographic Variables

Computing and Visualizing the Correlation (Pearson's r)

The graph on the left bottom of the image shows the correlation (Pearson's r) between mortality rate of each different cancer and Hispanic population. The graph on the right bottom of the image shows the correlation between liver cancer mortality and each 96 socioeconomic and demographic factor. Some examples of findings from these graphs are:

Liver caner mortality has a positive relationship with Hispanic population (R^2 = 0.6)

Population below poverty has a positive relationship with liver cancer (R^2 = 0.7)

*The findings in the graphs are some of examples among findings that can be explored in this application. More findings can be explored by selecting each 23 different cancer types and different 96 socioeconomic and demographic variables on the top of the interface.

CONCLUSIONS

Different from existing web mapping tools for public health data, our tool can provide a side-by-side visual comparison between health data and socioeconomic data to facilitate hypothesis testing for various health disparities and epidemiology research questions.

Disclaimer

The results of correlation analysis should not be considered as causal. For example, the correlation (Pearson's r = 0.6) is computed between the age adjusted death rate from liver cancer in each SRA region and the density of Hispanic population in each SRA region. It means that mortality rate of liver cancer is likely to be high in the region where Hispanic population density is high, but it does not necessarily mean that being Hispanic or residing in an area that is predominantly Hispanic causes increases risk of death from liver cancer. Other factors that correlate with the racial composition of a particular area could also be confounding the observed relationship between Hispanic race and liver cancer mortality, such as limited access to health care or poor health behaviors like drinking alcohol, both of which are known to increase the risk of death from liver cancer. Our exploratory analysis is a data mining tool that can be used for hypothesis testing and to construct more complex, realistic models of cancer mortality risk, but this visualization alone should not be interpreted as a valid tool for prediction of cancer outcomes.

REFERENCES

1. Gould, P. 1981. Letting the data speak for themselves. *Annals of the Association of American Geographers* 71 (2): 166-76.
2. Good, I. J. 1983. The philosophy of exploratory data analysis. *Philosophy of Science*: 283-95.
3. Harris, R. L. 1999. *Information graphics: A comprehensive illustrated reference.* Oxford University Press, USA.
4. Messner, S. F., L. Anselin, R. D. Baller, D. F. Hawkins, G. Deane, and S. E. Tolnay. 1999. The spatial patterning of county homicide rates: An application of exploratory spatial data analysis. *Journal of Quantitative Criminology* 15 (4): 423-50.
5. Kaski, S., and T. Kohonen. 1996. Exploratory data analysis by the self-organizing map: Structures of welfare and poverty in the world. Paper presented at Neural Networks in Financial Engineering. *Proceedings of the Third International Conference on Neural Networks in the Capital Markets* 498-507.

ACKNOWLEDGEMENTS

Research reported in this poster was supported by the National Cancer Institute of the National Institutes of Health under award numbers:
U54CA132384 & U54CA132379

FOR MORE INFORMATION:

Visit the Website of Web-based Health Data Mapping Tools
http://vision.sdsu.edu/health/