Celebration of SCHOLARS

Breast Cancer and Access to Health Care: A Spatial and Statistical Analysis

By: Shelby Smith
Faculty Advisor: Dr. Wenjie Sun
Department of Geography of Earth Science, Carthage College
Celebration of Scholars 2013: Exposition of Student and Faculty Research, Scholarship and Creativity

Carthage

Abstract

This research examines both the spatial and non-spatial relationships between breast cancer mortality rate and percent of population living below poverty, health care facilities per 1,000 people and percent urban, and analyzed for all 58 counties in California.

ArcGIS and SPSS are used for explorative mapping, clustering analysis, and regression models. These tests are able to determine if there is a correlation between breast cancer mortality rate and the other three variables.. Based on these results, further research should be done to connect other socioeconomic variables to increased rates in breast cancer mortally.

Problem Statement and Hypotheses

Based on pre-research along with a literature review this study will be identifying the spatial correlations between breast cancer mortality rates and percent of population living below poverty, health care facilities per 1,000 people and percent urban in all California counties.

Hypothesis$_1$: Breast cancer mortality rates are lower among counties with higher percent urban population.

Hypothesis$_2$: Breast cancer mortality rates are lower among counties with a higher numbers of health care facilities per 1,000 people.

Hypothesis$_3$: Breast cancer mortality rates are higher among counties with higher poverty rates.

Methodology

Statistical Analysis

- ArcMap is a geographic information systems (GIS) software. Ordinary least squares (OLS) which is a global non-spatial regression will first be run in ArcMap. R-Squared is an output value for OLS and the higher the R-Squared the more variance of the dependent variable can be explained by the independent variables.
- Geographically weighted regression (GWR) will be done. GWR is a local spatial regression model.
- Step wise regression will be performed using SPSS 20, a statistical software program. Step wise regression will test one independent variable at a time and it creates different models showing which independent variables best explain the variance of the dependent variable.

Mapping and Spatial Statistics

- ArcMap will be used for all the mapping and spatial statistical analyses. Explorative mapping will consist of mapping each data set individually and also in pair with cancer mortality rate to become familiar with the data sets and to visualize any potential spatial patterns and relationship in the data.
- The first test is a Spatial Autocorrelation tool called Global Moran's I (Burt and Barber 1996). Global Moran's I tests for the Z-score and P-value, which evaluates whether the overall spatial pattern expressed by a variable is clustered, dispersed, or random in a statistical sense. Moran's I will be run for breast cancer mortality rate, the OLS residual values and GWR residual values.
- If clustering is determined then hot spot analysis will be done. Hot spot analysis determines where specific local clusters of high or low values are located.

2007-2009 California Counties Breast Cancer Mortality Rate Compared to Percent Urban

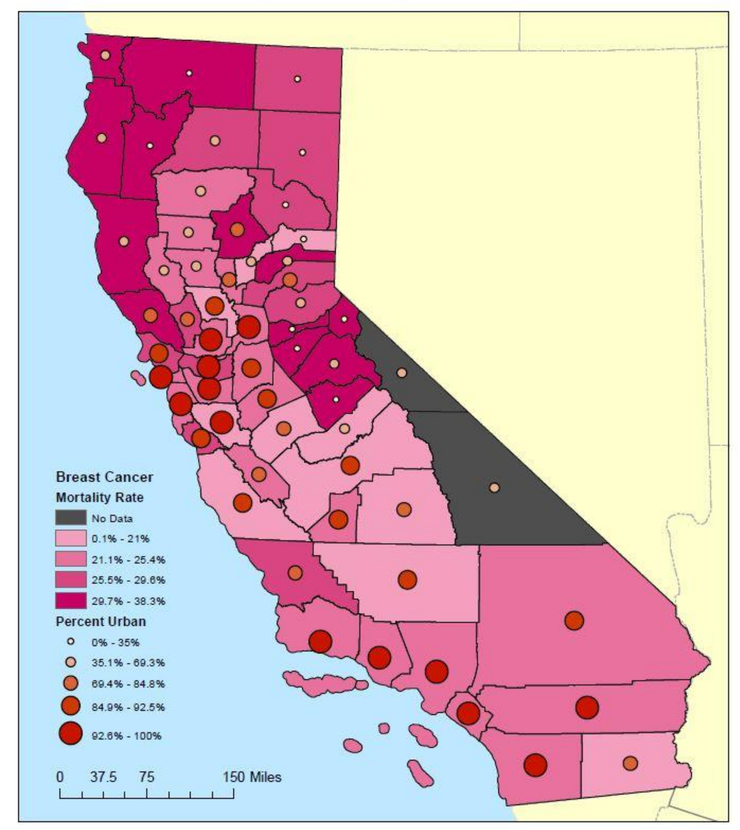

Figure 1

2007-2009 California Counties Breast Cancer Mortality Rate and Health Care Facilities per 1,000 People

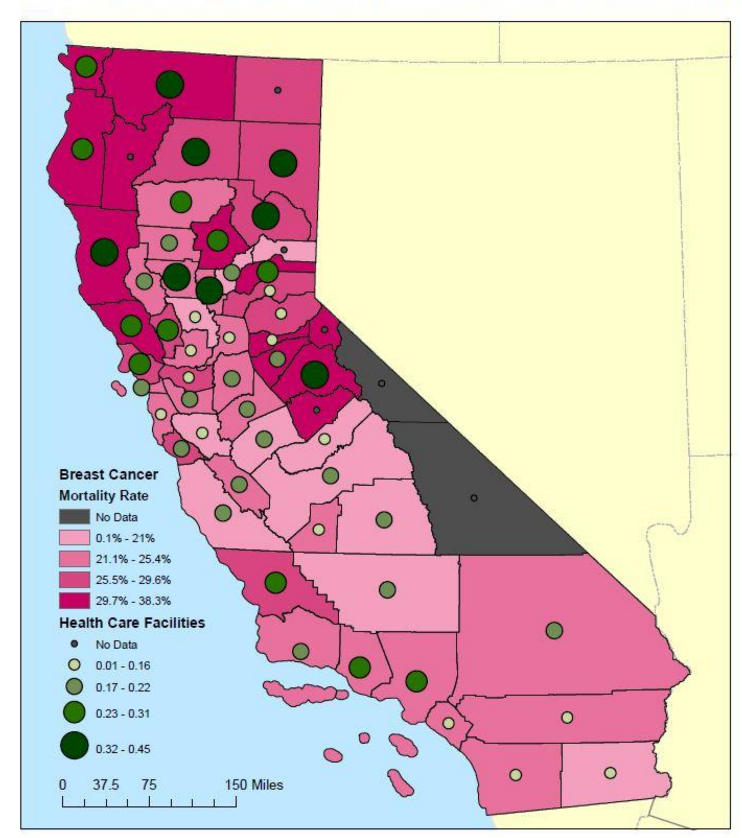

Figure 2

2007-2009 California Counties Breast Cancer Mortality Rate and Percent Population Below Poverty

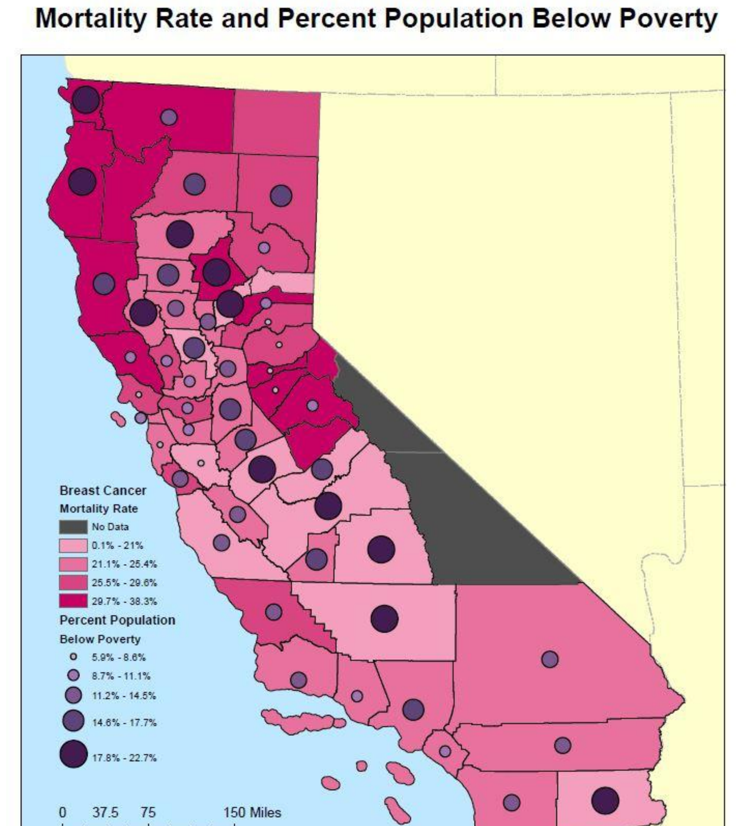

Figure 3

Results

2007-2009 California Counties Breast Cancer Mortality Rate Hot Spot Analysis

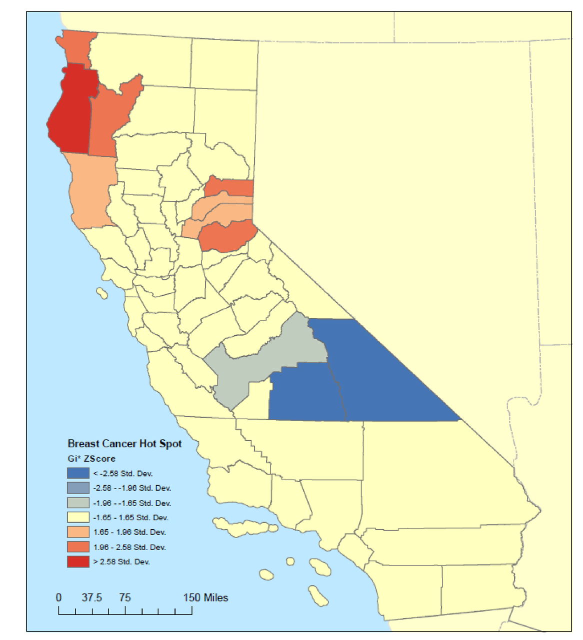

Figure 4

Ordinary Least Square Results

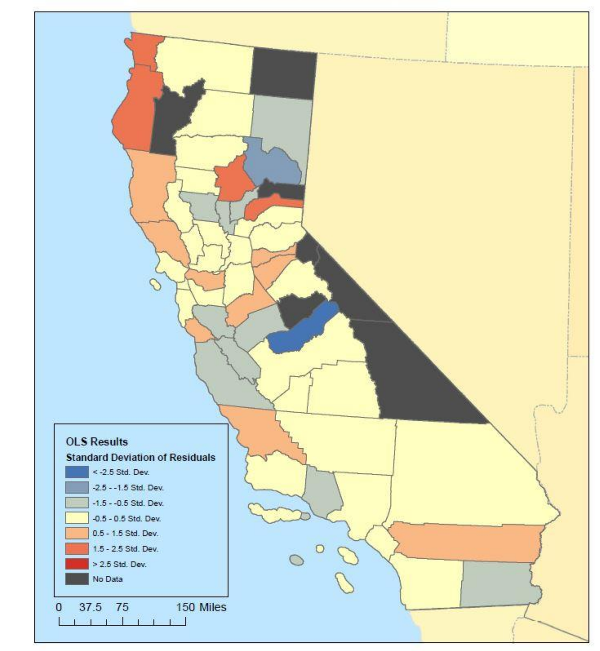

Figure 5

Geographically Weighted Regression Results Local R-Squared

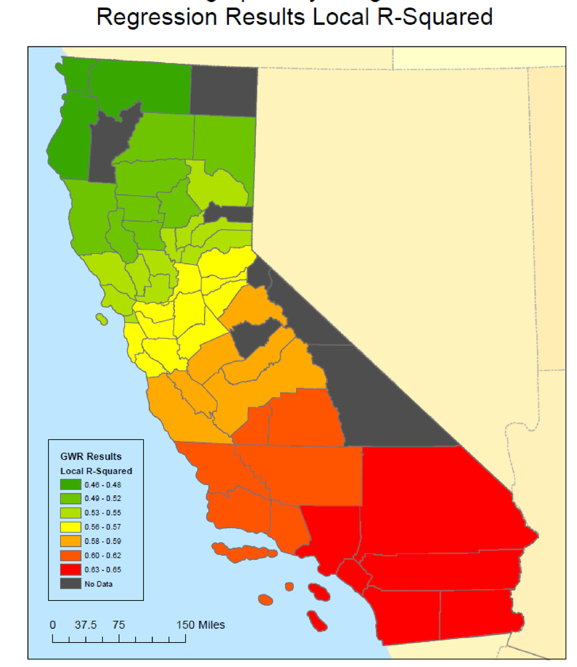

Figure 6

Geographically Weighted Regression Results Standard Deviation of Residuals

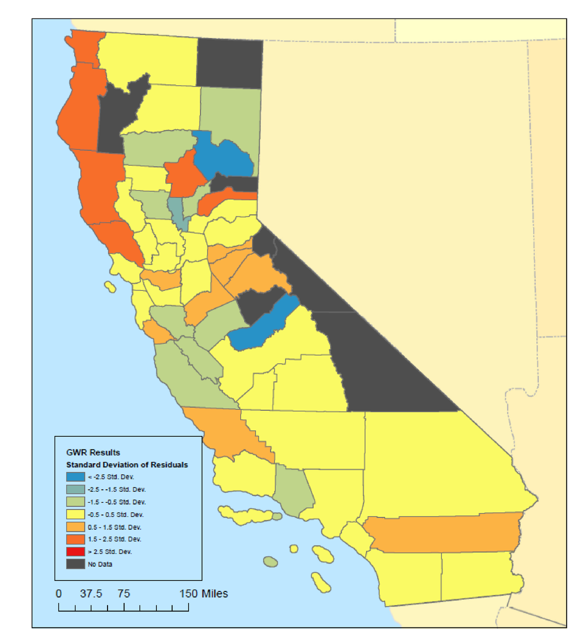

Figure 7

Ordinary Least Square Results
Adjusted R-Squared: 0.53

Geographically Weighted Regression Results
Adjusted R-Squared 0.55

Table 1: Ordinary Least Square Regression Coefficients *Statistically Significant

Variable	Coefficient	Std Error	t-Statistics	Probability
Intercept	36.24	4.04	8.97	0.00*
Facilities per 1,000 People	21.73	7.83	2.77	0.01*
Percent Living Below Poverty	-47.47	12.47	-3.81	0.00*
Percent Urban	-11.24	3.09	-3.64	0.00*

Table 2: Step Wise Regression *Statistically Significant

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Adjusted R Squared
		B	Std. Error	Beta			
1	(Constant)	32.65	2.44		13.37	0.00	
	Percent Urban	-10.36	3.25	-0.39	-3.19	0.00*	.14
2	(Constant)	27.53	2.53		10.86	0.00	
	Percent Urban	-11.53	2.91	-0.44	-3.96	0.00*	
	Facilities per 1,000 People	30.36	7.71	0.43	3.94	0.00*	.32

Figures 1, is showing that in counties with higher breast cancer mortality rate, percent urban is higher.

Figure 2, is showing a pattern that as health care facilities per 1,000 people increases, so does breast cancer mortality rate.

Figure 3, is not showing any discernible pattern between breast cancer mortality rate and percent of population below poverty.

Figure 4, is showing the hot spot analysis of breast cancer mortality rate. The clusters of reds are hot spots with higher mortality rates, while the cluster in blues are cold spots with lower mortality rates.

Figure 5, is showing the OLS results . The areas of reds are over predicted counties, while the areas in blues are under predicted by the regression model. The pattern of the residuals is random.

Figure 6, is showing the GWR local R-squared results. The northern counties have lower R-squared values and the R-squared values continue to increase the further south the counties, meaning that this model is explaining the variance of mortality rate better the further south the county is located.

Figure 7, is showing the GWR residuals. The areas of reds are over predicted counties, while the areas in blues are under predicted. The pattern of the residuals is random.

Discussion

Hypothesis$_1$ has been accepted ➔ Table 1, is showing the OLS results. Percent urban has .00 significance and a negative coefficient, which means as percent urban increases, breast cancer mortality rate is decreasing. Table 2 is showing the step-wise regression coefficients. Model 1 percent urban and is showing a .00 statistical significance and again has a negative coefficient.

Hypothesis$_2$ has been rejected ➔ The statistical significance for healthcare facilities per 1,000 people was .01 for OLS shown in Table 1 and .00 for step wise regression, shown in Table 2. Both had a positive coefficient. This means as health care facilities per 1,000 people increase so does breast cancer mortality rate.

Hypothesis$_3$ has been rejected ➔ Table 1, the OLS regression coefficients show percent of population living below poverty is statistically significant .00 and has a negative coefficient. This means as percent poverty decreases, breast cancer mortality rate increases

- OLS resulted with an adjusted R-square value of .53 meaning 53 percent of the variance of breast cancer mortality rate can be explained by this model. GWR which is a local regression model, so the map of local R-square (Figure 6) is more meaningful.
  - ➢ These results imply that 45-47 percent of the variance of the dependent variable is still unaccounted for in this study.

The reason why percent poverty would be significant in OLS but not significant enough to be included in the step wise regression could be that ArcMap and SPSS have different algorithms in calculating the regression models.

Acknowledgements & References

A huge thank you goes to my faculty advisors Dr. Wenjie Sun and Kurt Peipenberg, . I would like to give a special thanks to Professor Patricia Murphy, because of her classes and the knowledge she has of human geography, it has allowed me to find my passion in this world. Along with the rest of the Geography and Earth Science faculty and students.

The data is derived from the California Cancer Registry 2007-2009, 2010 Census, American Community Survey 2007-2009 and Cal-Atlas 2010

Anderson, Elijah. *Against the Wall* . Philadelphia: University of Pennsylvania Press, 2008.
Baldwin, Laura-Mae, et al. "Access to Cancer Services for Rural Coloractal Cancer Patients." *National Rural Health Association*, 2008: 390-399.
Burt, James E., and Gerald M. Barber. *Elementary Statistics for Geographers.* Second Edition. New York, New York: The Guilford Press, 1996.
Dai, Dajun. "Black residential segregation, disparities in spatial access to health care facilities, and late-stage breast cancer diagnosis in metropolitan Detroit." *Elsevier*, 2009: 1038-1052.
ESRI 2011. "ArcGIS Desktop: Release 10." Redlands, CA: Environmental Systems Research Institute.
Fukuda, Yoshiharu, Masahiro Umezaki, Keiko Nakamura, and Takehito Takano. "Cariations in societal characteristics of spatial disease clusters: examples of colon, lung and breast cancer in Japan." *International Journal of Health Geographies*, 2005: 4-16.
Kennedy, Betty M, Sahasporn Paeratakul, Donna H Ryan, and George A. Bray. "Socioeconomic Status and Health Disparity in the United States." *Journal of human behavior in the social environment*, 2007: 13-23.
Melnick, Alan L. *Introduction to Geographic Information Systems in Public Helath.* Gaithersburg, Maryland: Aspen Publishers, 2002.
Polednak, Anthony. "Survival of Breast Cancer Patients in Connecticut in Reltion to Socioeconomic and Health Care Access Indicators." *Journal of Urban Health*, 2002: 211-218.
Sprague, L Brian, Shaneda Warren Andersen, and Amy Trentham-Dietz. "Thyroid cancer incidence and socioeconomic indicators of health care access." *Cancer Causes Control*, 2008: 585-593.