

San Francisco crime data correlations—comparing Linear Regression with Geographically Weighted Regression methods.

EXECUTIVE SUMMARY

I compared two regression methods to identify associations between crime factors and their spatial relationships using crime data from San Francisco. The methods included Multi Linear Regression and Geographically Weighted Regression (both fixed-distance and adaptive kernels).

I hypothesized originally that *drug* crimes would be highly correlated with other factors, including *robbery*, *vandalism* and *car theft* as predictors.

Testing suggested a more complex set of relationships among factors, particularly between *drug* crimes & *robbery* and *robbery* & *vandalism*, pairs which showed relatively strong correlations. Compared to the Multi Linear Regression, the Geographically Weighted Regression appears to explain more of the variation and confirms that *robbery* generally had the highest R^2 values and Regression Coefficient. Mapping the data confirms the spatial relationships between *drug* crimes and *robbery*, especially, and shows that *vandalism* and *car theft* incidents were more widely distributed which may explain why they didn't factor into the results as significantly.

