Interactive Detection of Spatial Clusters
Frank Hardisty1, PhD, Jamison Conley2, PhD
1GeoVISTA Center, Department of Geography, Pennsylvania State University
2Department of Geology and Geography, West Virginia University

OBJECTIVE
Geographic visualization methods allow analysts to visually discover clusters in multivariate, spatially-referenced data. Computational and statistical cluster detection techniques can automatically detect spatial clusters of high values of a variable of interest. The authors propose that the two approaches can be complementary; and present an integration of the GeoViz Toolkit and Proclude software suites as proof-of-concept.

BACKGROUND
Geographic visualization (or geovisualization) refers to a set of tools and techniques for supporting geospatial data analysis through interactive visualization. [1] The GeoViz Toolkit is one platform for doing geographic visualization [2].

Proclude includes a suite of four cluster detection techniques [3]. The most computationally efficient of these is a genetic GAM. This approach has been found to be both computationally tractable and able to cover the solution space [4].

METHODS
We combined our toolkits, allowing for interactive “round-tripping”: results of Proclude clustering methods are fed into the GeoViz Toolkit as a subset selection highlighting the data observations contained within the cluster. Conversely, selections in the GeoViz Toolkit are passed to Proclude to be evaluated by a fitness function, such as the observed to expected ratio.

RESULTS
We found that combining multivariate geographic visualization with cluster detection encourages enables a more thorough investigation into clustering. It does so by enabling rapid exploration into the spatial and attribute structure of the clusters uncovered by the cluster detection techniques, and conversely, by allowing for the rapid evaluation of spatial and attribute subsets of the problem space.

CONCLUSIONS
This is a promising combination of techniques. Further research is needed to uncover the optimal connections between geographic visualization and clustering software.

REFERENCES
Further Information:
Frank Hardisty, hardisty@psu.edu
http://www.geovista.psu.edu/grants/cdcesda/software/


ACKNOWLEDGMENT
This work was partially supported by grant CA95949-01 from the National Cancer Institute.