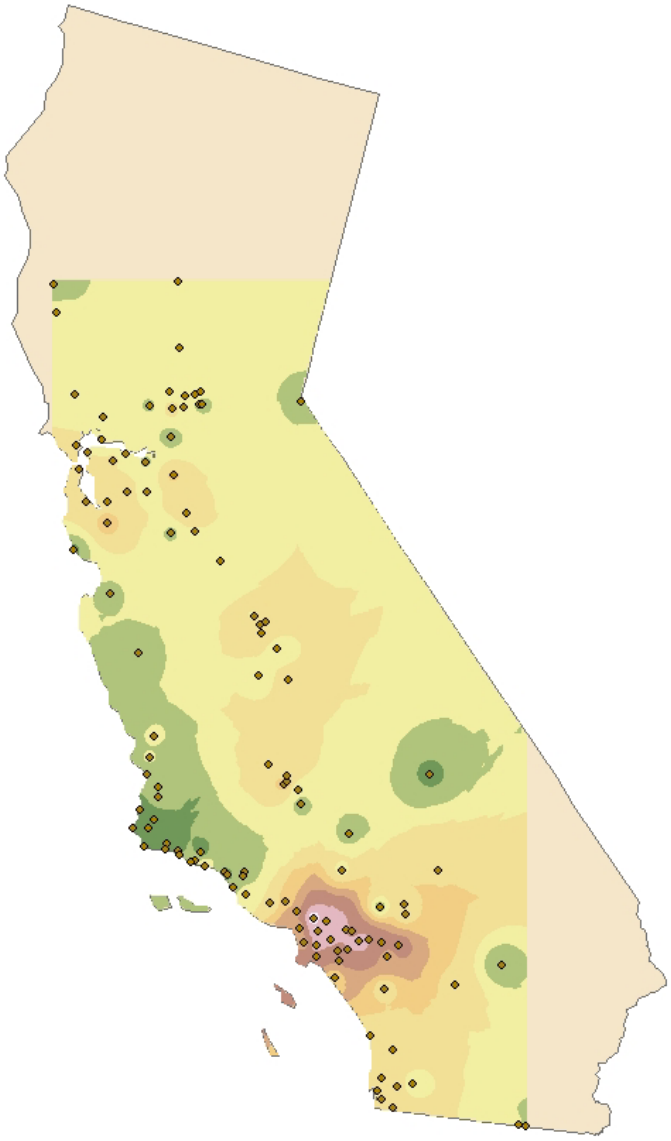
Lab 2: Spatial Analysis for Public Health
Dahl Winters
11/3/06

Part 1: Nearest Neighbor Analysis
Do the exercise called "Nearest Neighbor Analysis" on the course Blackboard site under Assignments.

Part 2: Raster Analysis in ArcGIS

Interpolating a Surface
Download the following dataset to practice with these tools- http://www.unc.edu/~emch/gisph/CAdata.zip and unzip it. First use the IDW interpolation tool to interpolate a surface of the NastyWater variable. Use the mask tool and mask the result to the California boundary area. Print the result. Where is the water nastiest in California based on this interpolation?.



Settings:
Output cell size 2024.1148
Power = 2
Search Radius = Variable
Number of Points = 12

The redder colors signify nastier water, while the greenest colors represent the cleanest water. The reddest colors are right around Los Angeles. Thus, based on this interpolation, this area has the nastiest water in all of California.

Calculating Densities

Using the CAdata calculate the density using the population field and a 50,000 meter search radius. Print the map. Now do the same using the clinic field. Then using the raster calculator, calculate a map that is the ratio of clinic density to population density. Are there some areas that don't have a lower ratio than others?

Population density using kernel method

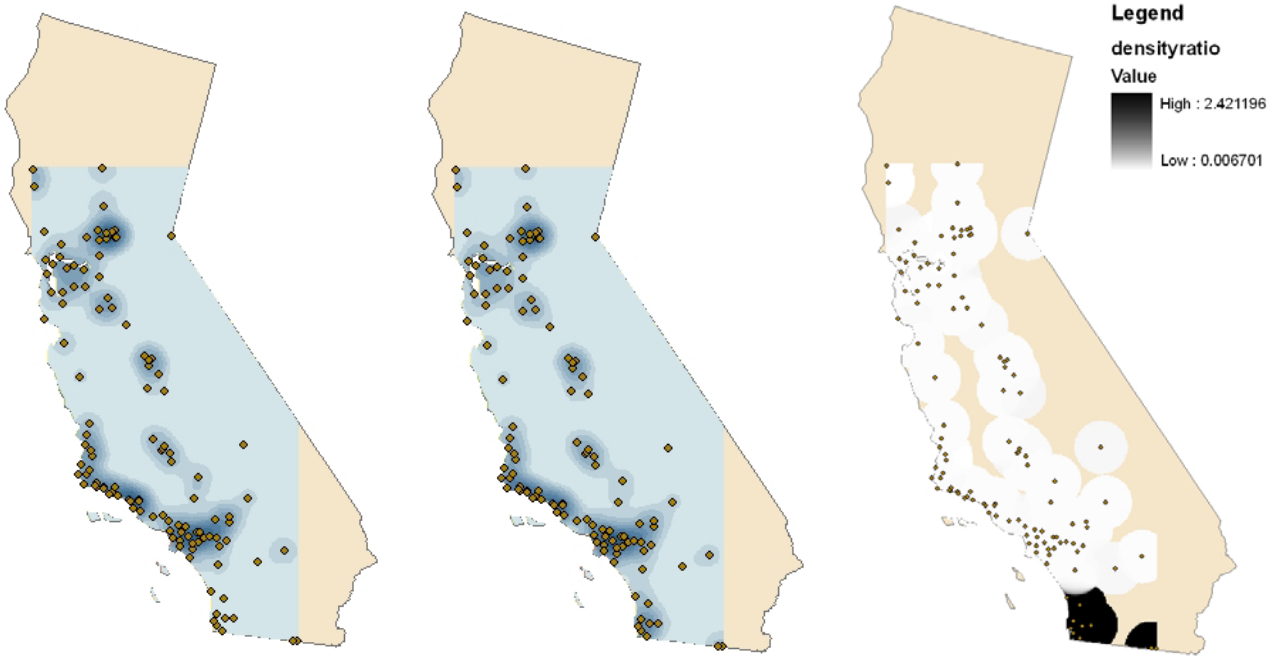
Kernel density tool used
Population field = Population
Output cell size = 2024.1148
Search radius = 50000
Units = Square meters

Clinic density using kernel method

Population field = Clinics
Same settings for everything else

Ratio of clinic density to population density

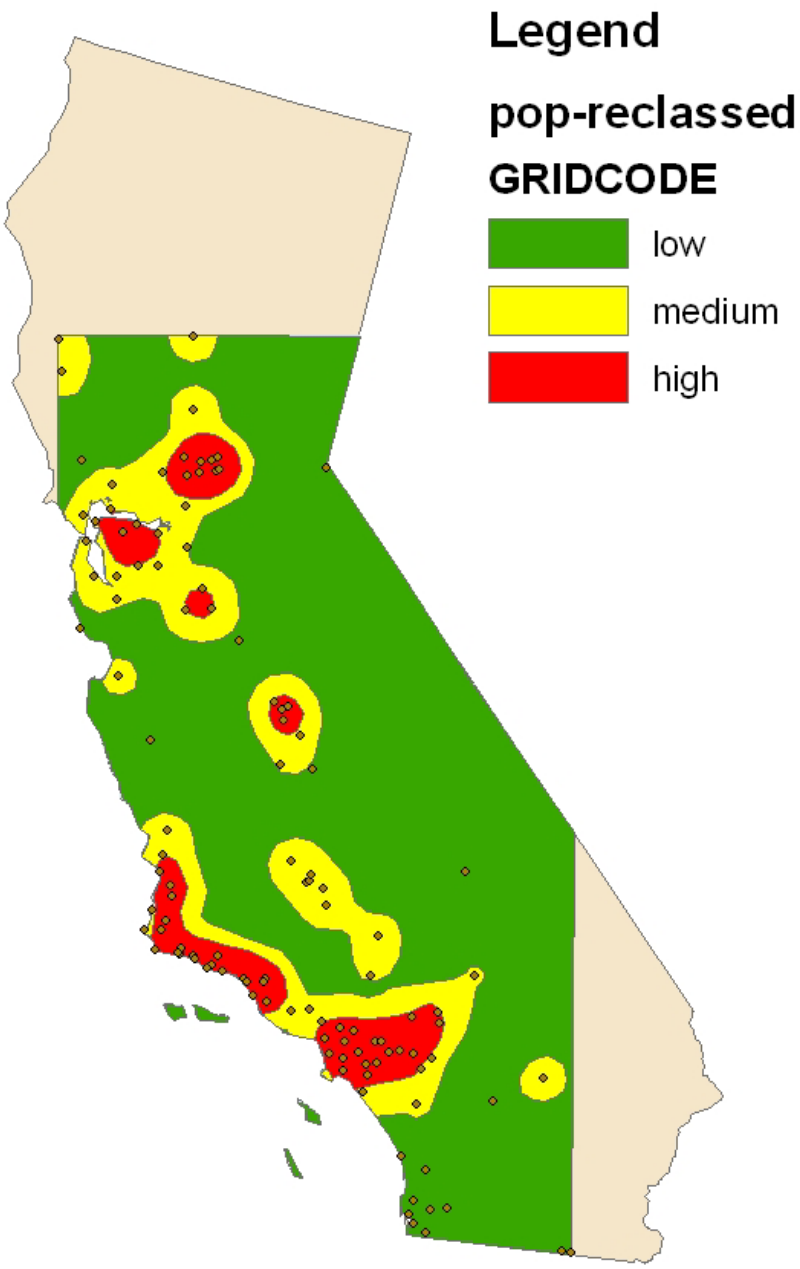
Raster calculator: Clinic density layer / Population density layer



There are some areas that don't have a lower clinic/population ratio than others. These are the black areas at the southern tip of California. These areas have relatively few clinics for the amount of population they have.

Reclassification and Vector Conversion

Reclassify the population density raster into 3 classes: high, medium, and low based on natural breaks in the data. Then convert the dataset into vector, make a map that labels the different density categories, and print the result.



Part 3: Spatial Statistics in ArcGIS

Measuring Geographic Distributions

In ArcGIS under the "Measuring Geographic Distributions" tools of the "Spatial Statistics" toolbox run the following programs using the towns point layer: 1. Central feature, 2. Mean Center (without weighting), 3. Mean Center (weighted by the population field), 4. Directional Distribution (Standard Deviational Ellipse) using 1 SD, 5. Standard Distance using 1 SD. Make a map or maps that labels all of the features that you created. Then describe what each one means.

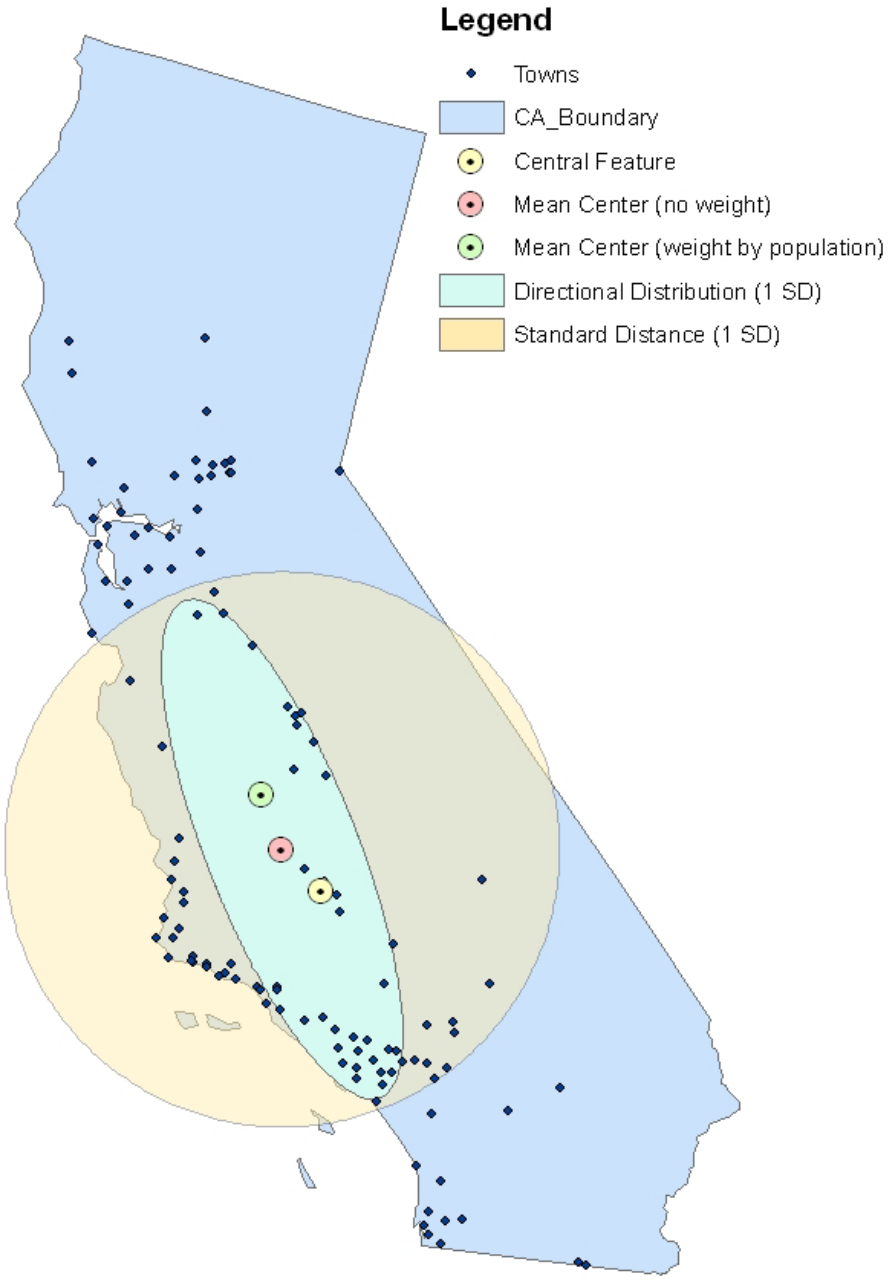
1) Central Feature – this is the town with the shortest distance to all other towns.

2) Mean Center (without weighting) – This is the average x-y coordinate of all the towns in the layer.

3) Mean Center (weighted by population) – This is the same as the above, except that to find the average x-y coordinate, each town is weighted according to its population. Thus, the weighted mean center will be closer to the most populated towns.

4) Directional Distribution using 1 SD – The resulting ellipse describes a NW to SE distributional trend of towns, roughly paralleling the California coastline.

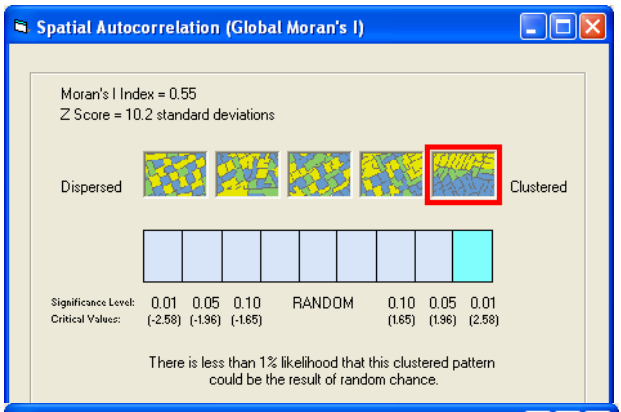
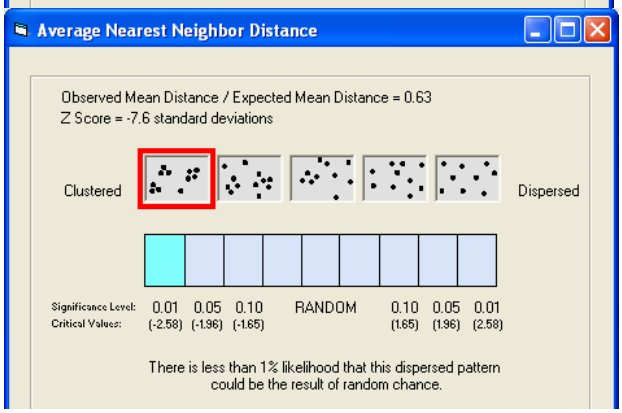
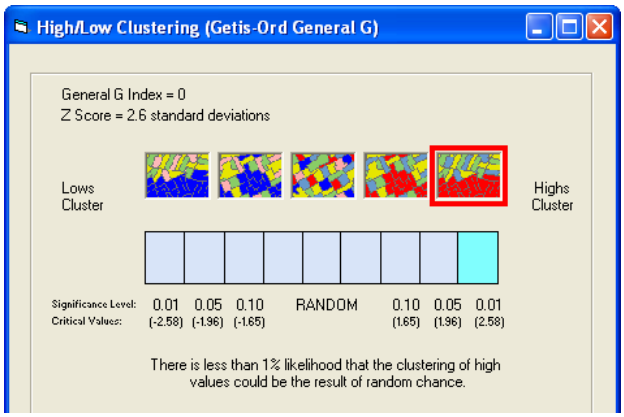
5) Standard Distance using 1 SD – this circle is a measure of the compactness of the distribution of towns, and is centered on the mean center.



Analyzing Spatial Patterns

In ArcGIS under the "Analyzing Spatial Patterns" tools of the "Spatial Statistics" toolbox run the global Moran's I program using the NastyWater variable and print the result. Is NastyWater spatially autocorrelated in California? If so, how much and what does that mean? Calculate the average nearest neighbor distance on towns. Are the towns clustered or dispersed? What is the average nearest neighborhood value and Z-score? Calculate the Getis-Ord General G Index on NastyWater. What is the value and Z-score? Is there high or low clustering?.

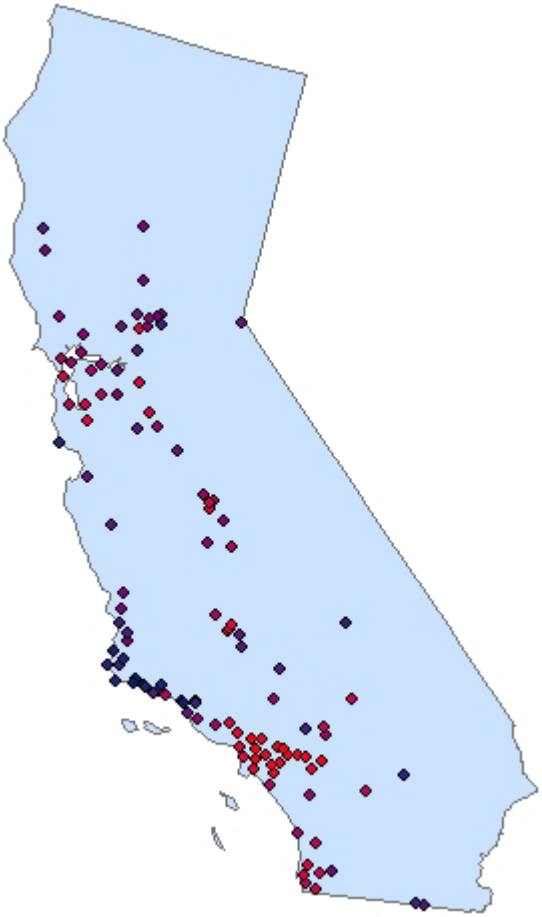
All three methods of analyzing spatial patterns conclude that the towns are highly clustered with regard to the NastyWater variable.

Output	**Graphical Output**	**Conclusions**
Global Moran's I: Moran's Index 0.55027372764353799 Expected Index -0.0085470085470085479 Variance 0.0029656716394875506 Z Score 10.26150310975752		From this, it appears NastyWater is spatially autocorrelated in California, given the high level of clustering.
Average nearest neighbor distance: Nearest Neighbor Observed Mean Distance 19877.593304735183 Expected Mean Distance 31156.800961313649 Nearest Neighbor Ratio 0.63798569466154509 Z Score -7.523113585793479 Standard Deviations		The towns are strongly clustered. The average nearest neighborhood value is 19.877 km, and the Z score is -7.52 standard deviations.
Getis-Ord General G: Observed General G 8.8955622163691837e-006 Expected General G 7.5106673185965468e-006 General G Variance 2.7128132633895175e-013 Z Score 2.6589297618698469 Standard Deviations		The value and Z-score are 8.8955 x10^{-6} and 2.6589 standard deviations, respectively. There is high clustering.

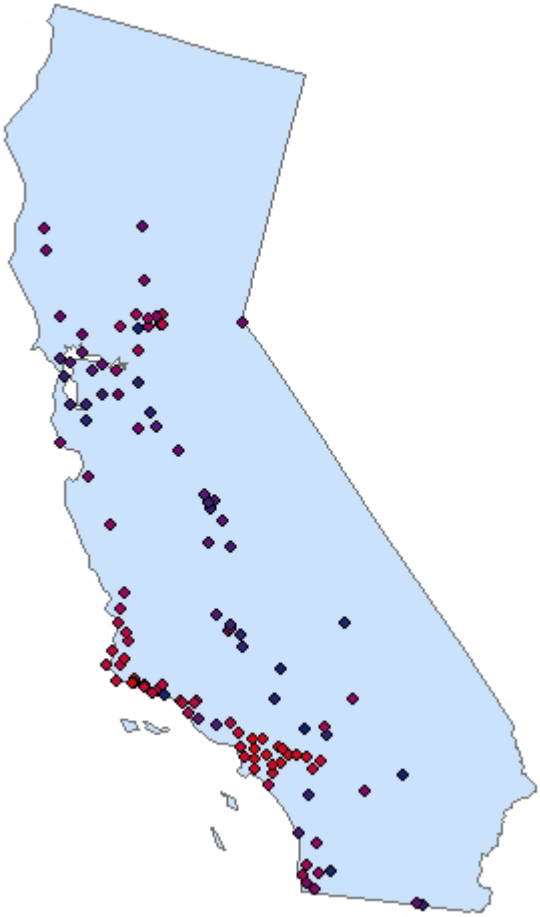
Mapping Clusters

In ArcGIS under the "Mapping Clusters" tools of the "Spatial Statistics" toolbox run the program called "Hot Spot Analysis: Getis-Ord Gi" for the NastyWater variable Use a distance band of 10,000. The result is a table that has a new field in it called Gi10000. Map that out with the small numbers (cool spots) in blue points and the large numbers (hot spots) in red. Does there seem to be spatial clustering of those hot spots? Then calculate the Anselin's Local Moran's I values for the NastyWater variable using the Town layer. Map the local Moran's I value using red and blue? Where is there high local spatial autocorrelation of NastyWater and what does that mean?*

Getis-Ord Gi*
Used distance band of 10,000



Anselin's Local Moran's I



Both methods show spatial clustering of the NastyWater hot spots. The Getis-Ord Gi* shows clustering especially around the Los Angeles area and to a lesser extent around the Bay area farther north. This makes sense since water quality is likely to be poor around these major metropolitan areas due to polluting land uses. The Anselin Moran's I method shows hot spot clustering around the Los Angeles area and an adjacent area to the northwest.

Lab Deliverable Summary: Print out all of the outputs, put your name on them, and give them to the instructor. They include:

Part 1: printout of results of nearest neighbor analysis and paragraph describing distribution.

Part 2:

- a map of the interpolated nasty water surface and a short description of the spatial pattern,
- a raster map of clinic density/ population density and short description of the spatial distribution,
- a reclassified vector map that labels the different density categories.

In Part 2, organize the deliverables by the three sections that are listed.

Part 3:

- map or maps showing the 3 geographic distribution summary statistics and a description of what each means,
- Moran's I result printout and description of what the result means,
- average nearest neighbor calculation result and description of what it means,
- Getis-Ord General G Index result and description of what it means.

In Part 3, organize the deliverables by the three sections that are listed.