1. Global Moran's I:
    a. Two slopes:

    Blue: the x means value itself, the y means neighbors value.

    Grey: The slope represent the value of Global Moran's I

2. SD: The output is changing according to the values of SD, radius =

    The Weighted Standard Distance extends to the following:

$$SD_w = \sqrt{\frac{\sum_{i=1}^{n} w_i(x_i - \bar{X}_w)^2}{\sum_{i=1}^{n} w_i} + \frac{\sum_{i=1}^{n} w_i(y_i - \bar{Y}_w)^2}{\sum_{i=1}^{n} w_i} + \frac{\sum_{i=1}^{n} w_i(z_i - \bar{Z}_w)^2}{\sum_{i=1}^{n} w_i}}$$

    Where w_i is the weight at feature i and $\{x_w, y_w, z_w\}$ represents the weighted Mean Center.

    ***standard(1.0/2.0/3.0)**

3. **Gival**: (z-score), for statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot spot). For statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (cold spot).
4. Zval, p-value
    a. The **p-value is a probability**. For the pattern analysis tools, it is the probability that the observed spatial pattern was created by some random process
    b. **Z-scores are standard deviations**. If, for example, a tool returns a z-score of +2.5, you would say that the result is 2.5 standard deviations.

--

TFS source code:

.\TerraFly\App\GeoCloud\GeoCloud\GeoCloudHealth

Service is running at (Old & Current Server):

IP: 131.94.133.233

Username: jing; Pass: J***3

Database: jarvis.cs.fiu.edu:10533

Command in the Sql Server: jarvis.cs.fiu.edu,10533

All the table information should be stored in the "dataset_info"

db=geo_cloud

usr=geocloudadmin

pwd=geocloudadmin

source folder: /home/guang/workplace/projects/statistic_analysis_tool/djcode/mysite

New Server

Access via jarvis.cs.fiu.edu:10722 (Linux)

username: root

pass: akula****
Database: jarvis.cs.fiu.edu:10533
source folder: /home/guang/workplace/projects/statistic_analysis_tool/djcode/mysite

How to Run(When Power outage or other things happen)
1. screen
2. python manage.py runserver 0.0.0.0:8000
Control + a + d detached screen but screen is still alive
When there exist screen already, you can kill them
1. screen -ls
2. screen -X -S [session # you want to kill] kill

Download(Mac)
scp -r mysite jing@131.94.133.233:/home/guang/workplace/projects/statistic_analysis_tool/djcode/mysite /Users/jinjing/Documents

Download(Window)
SSH Secure shell Client

How to debug:
import pdb
pdb.set_trace()

Architecture

1. using python (Django framework)

cluster, geodistribution, lib, manage.py, map, mysite, regression

2. The main important parts: cluster, geodistribution, regression, map
    a. **Cluster**: Cluster/Hotspot/Cluster and Outlier
    b. **Geodistribution**: MeanCenter/MedianCenter/Standard/Distributional trend
    c. **Regression:** Linear regression/Auto regression
    d. **Map**: disease map tool/SMR map tool
3. These modules share the codes in the **lib**.

- **Geodistribution**

1. MeanCenter

tance, and distributional trend functions. In our system, a weighted mean central is provided as follows:

$$X = \frac{\Sigma_i w_i x_i}{\Sigma_i w_i}, Y = \frac{\Sigma_i w_i y_i}{\Sigma_i w_i}, \quad (4)$$

where x_i and y_i denote the coordinate of each point (but when the dataset is polygonal, x_i and y_i indicate the center of each polygon) and w_i is the weight that corresponds in our system to mortality or incidence. Figure 8(d) shows these two types of points: one

URL: http://131.94.133.233:8000/analysis/geodistribution/meanctr/18/1/avg_income/

Output (Geojson: just one point): {"type": "FeatureCollection", "features": [{"geometry": {"type": "Point", "coordinates": [-80.308983377963045, 25.751016310751204]}, "type": "Feature", "properties": {}, "id": null}]}

2. MedianCenter

(http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-median-center-spatial-statistics-works.htm)

Identifies the location that minimizes overall Euclidean distance to the features in a dataset. The method used to calculate the Median Center is an iterative procedure: At each step (t) in the algorithm, a candidate Median Center is found (Xt, Yt) and then refined until it represents the location that minimizes the Euclidean Distance d to all features (or all weighted features) (i) in the dataset.

The Median Center is given as:

$$d_i^t = \sqrt{(X_i - X^t)^2 + (Y_i - Y^t)^2 + (Z_i - Z^t)^2}$$

where x_i, y_i, and z_i are the coordinates for feature i, and n is equal to the total number of features.

> Note: While the Median Center tool only returns a single point, there may be more than one location (solution) that would minimize the distance to all features.

URL: http://131.94.133.233:8000/analysis/geodistribution/medianctr/18/1/avg_income/

Output (Geojson: just one point): {"type": "FeatureCollection", "features": [{"geometry": {"type": "Point", "coordinates": [-80.302505041928853, 25.751773397262596]}, "type": "Feature", "properties": {}, "id": null}]}

3. Standard Distance tool

http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-standard-distance-spatial-statistic-works.htm

For two-dimensional data, the Standard Distance tool creates a new feature class containing a circle polygon **centered on the mean center.** According to the degree, append the points in a list.

Measuring the compactness of a distribution provides a single value representing the dispersion of features around the center. Each circle polygon is drawn with a radius equal to the standard distance value.

The Weighted Standard Distance extends to the following:

$$SD_w = \sqrt{\frac{\sum_{i=1}^{n} w_i (x_i - \bar{X}_w)^2}{\sum_{i=1}^{n} w_i} + \frac{\sum_{i=1}^{n} w_i (y_i - \bar{Y}_w)^2}{\sum_{i=1}^{n} w_i} + \frac{\sum_{i=1}^{n} w_i (z_i - \bar{Z}_w)^2}{\sum_{i=1}^{n} w_i}}$$

Where w_i is the weight at feature i and $\{x_w, y_w, z_w\}$ represents the weighted Mean Center.

URL: http://131.94.133.233:8000/analysis/geodistribution/stddst/18/1/avg_income/1.0/

Output(Geojson: just one polygon): {"type": "FeatureCollection", "features": [{"geometry": {"type": "Polygon", "coordinates": [[[-80.137237690761836, 25.751016310751204], [-80.13726384846187, 25.75401368628809], [-80.137342313594075, 25.757010148795352]...

4. Distribution Trend

The Distributional trends tool provides an ellipse to give a measure of the feature distribution and trends.

URL: http://131.94.133.233:8000/analysis/geodistribution/trends/18/1/avg_income/1.0/

Output(Geojson: just one polygon): {"type": "FeatureCollection", "features": [{"geometry": {"type": "Polygon", "coordinates": [[[-80.41115887822167, 25.765581075099956], [-80.411401073662077, 25.763796093560789], [-80.411619233704982, 25.762007860004459]...

- **Cluster**
  - The **p-value is a probability**. For the pattern analysis tools, it is the probability that the observed spatial pattern was created by some random process. When the p-value is very small, it means it is very unlikely (small probability) that the observed spatial pattern is the result of random processes, so you can reject the null hypothesis.
  - **Z-scores are standard deviations**. For statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (hot spot). For statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (cold spot). If, for example, a tool returns a z-score of +2.5, you would say that the result is 2.5 standard deviations.
  - **Local Moran I:** A positive value for I indicates that a feature has neighboring features with similarly high or low attribute values; this feature is part of a cluster. A negative value for I indicates that a feature has neighboring features with dissimilar values; this feature is an outlier.

1. Hotspot

The Hotspot analysis tool identifies hot spot(cluster with high value) and cold spot(cluster with low value) using the Getis-Ord Gi* statistic method.

http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/hot-spot-analysis.htm

URL: http://131.94.133.233:8000/analysis/cluster/hotspot/18/1/avg_income/

Output: {"type": "FeatureCollection", "features": [{"geometry": {"type": "Polygon", "coordinates": [[[-80.259137018055497, 25.8307150240997], [-80.259227023884506, 25.830709997867501], [-80.259228992923994, 25.831047999031799]...

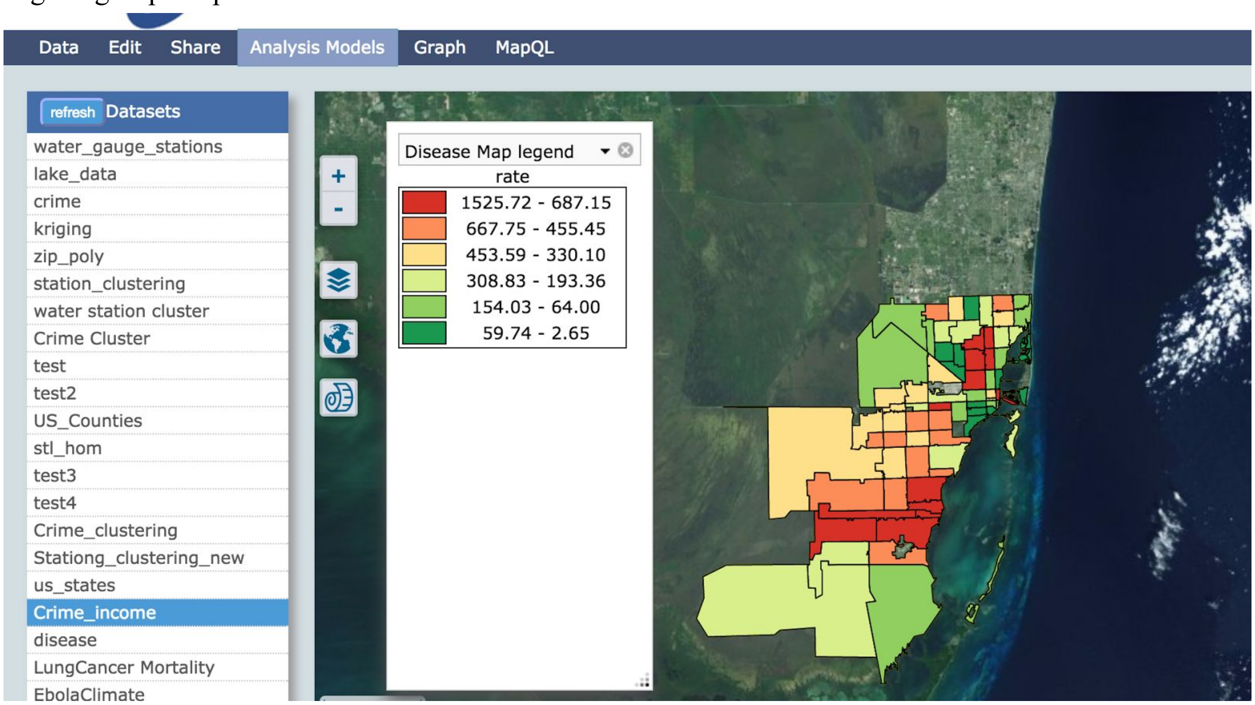
2. Cluster and Outlier

The Cluster and Outlier Analysis identifies hot spot(cluster with high value) , cold spot(cluster with low value) and outlier.This tool using Anselin Local Moran's I statistic method.
http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/cluster-and-outlier-analysis-anselin-local-moran-s.htm
URL: http://131.94.133.233:8000/analysis/cluster/ClusterandOuter/18/1/avg_income/
Output: {"type": "FeatureCollection", "features": [{"geometry": {"type": "Polygon", "coordinates": [[[-80.259137018055497, 25.8307150240997], [-80.259227023884506, 25.830709997867501]...

- **Map Tool**

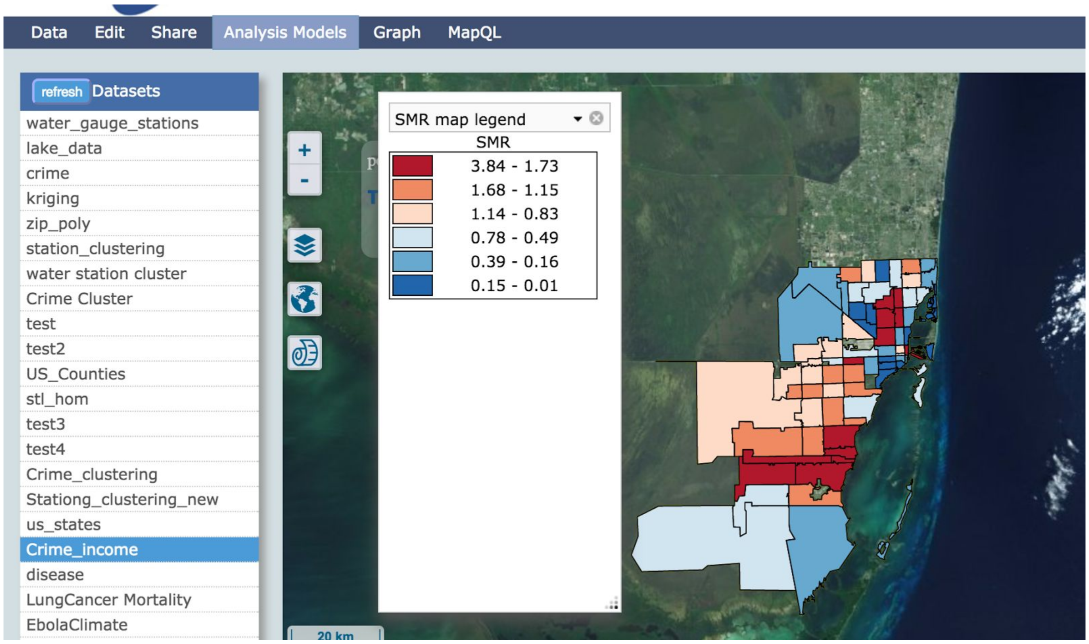
1. Disease Map tool

The input of disease map is the cases/dead number and population number of each area, and the output is the incidence/mortality.
Population: population
Case: disease number
Site Name: only represents the name of location.
Legend groups: represents the number of color will be showed.



2. SMR Map tool

Similar as the (1), but represents the local disease rate/average rate



- **Regression**

1. Linear Regression

URL: http://131.94.133.233:8000/analysis/regression/linear/18/1/crime_numb/avg_income/

Linear regression is an approach to modeling the relationship between a dependent variable y and one or more explanatory variables denoted as x.

(a) **Residual standard deviation:** The residual standard deviation is a statistical term used to describe the standard deviation of points formed around a linear function, and is an estimate of the accuracy of the dependent variable being measured.

(b) **AIC**: Akaike information criterion. It offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. In doing so, it deals with the trade-off between the goodness of fit of the model and the complexity of the model(The lower the better).

(c) **Multiple R-squared:** Multiple R-squared is used for evaluating how well your model fits the data. They tell you how much of the variance in the dependent variable (the predicted variable) can be explained by the independent variables (the predictor variables). For example, an R-squared value of 0.75 implies that the model can explain three-quarters of the variation in the outcome. It is important to know that **every time you add an independent variable to the model, the R-squared value will increase**. (if you wanted add as many parameters as observation into your model. This would give you a r-squared of 1) Why is that ? Because the model tries to capture both the information and any noise in the new variable. We do not know if the increase in R-squared value is due to actual predictive power of the new variable or due to chance alone.

(d) **Adjusted R-squared:** Adjusted R-squared also provides the same information as R-squared but adjusts for the number of terms in the model. It does not monotonically increase like R-squared but increases only when the new variable actually has an effect on the predicted value. It decreases when the new variable does not have any real impact on the predicted value.

2. Auto Regression

URL: http://131.94.133.233:8000/analysis/regression/spatialauto/18/1/crime_numb/avg_income/
Maximum likelihood estimation of spatial simultaneous autoregressive lag and mixed models of the form: Y = rho W y + X beta + e where Y is a dependent variable, W is a matrix of spatial weights, x is an independent variable, beta denotes the unknown parameters and e is an error term.

(a) Wald test: used to determine whether various parameters can be zero or not.
(b) LR test: used for testing spatial dependence;
(c) LM test: evaluating the absence of spatial autocorrelation in lag model residuals

3. Error Regression

URL: http://131.94.133.233:8000/analysis/regression/errorauto/18/1/avg_income/crime_numb/
http://www.statsref.com/HTML/index.html?sar_models.html
This model is applied when there appears to be significant spatial autocorrelation, but tests for spatial lag effects do not suggests that inclusion of the latter would provide a significant improvement.