

Spatial and temporal issues in health related databases: geo-statistical methods to make data more robust, relevant, and useful.

Presented by:
Paul Robinson, PhD
The Charles R. Drew Medical GIS Laboratory
(CDMGIS lab)
paulrobinson@cdrewu.edu
probinson@ucla.edu

Whats is GISc?

Geographic Information Science

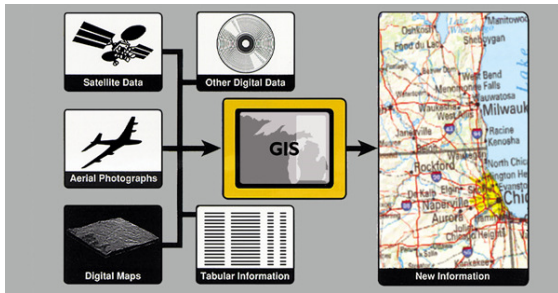
- GISc is the theoretical foundation for Geographic Information Systems (GIS) and related technologies
- GISc theory, principles and methods should be at the core of spatial inquiry.
- GIS is merely the the toolbox of the geo-scientist
- Health research and practice has yet to be become fully infused with geo-spatial truths

What is GIS?

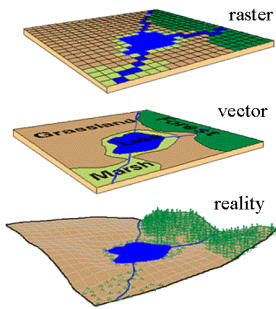
GIS defined:

"A geographic information system (GIS) is a computer based information system that uses a set of procedures and tools to store, integrate, manipulate, query, model, and display geographically referenced, and associated attribute (tabular) data, for the purpose of supporting decision making in the resolution of real-world problems."

Where does GIS data come from?



Sources & Types of Geographic Data



Analyzing Geographic Data

What type of geo data do you have?

Zip Codes

Administrative districts

X,Y points

Imagery

Street Addresses

GPS coordinates

Other geographic data considerations

Temporal stamps - does the data have time stamps?

Temporal Stationarity - do the data characteristics (mean, variance, etc) change over time?

Geographic relationships and movement – Does the data describe moving objects, or phenomena that are linked across space?

Data helps determine approach

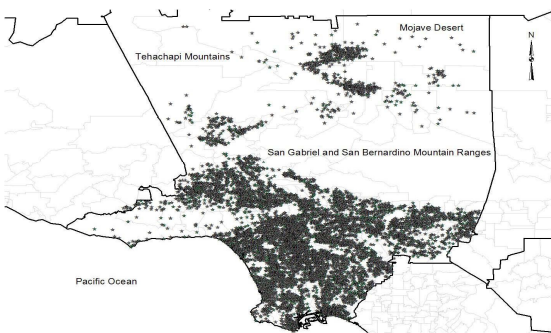
Basic GIS looks at one time period at a time. Temporal changes are recorded in different map layers.

If you are interested in how phenomena move through space in real time or in some short time frames you would consider the network modeling functions of GIS.

In health we often large amounts of data with time slices. In these instances use space time cubes.

Appendix 1

Census Blocks containing one or more Los Angeles Health Survey Respondents in 2002 and/or 2005



Space time Cubes

Puts data in space time data bins

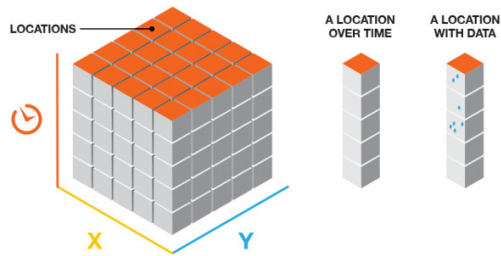
Time Series

Time Slices

Counts

Must have values (or estimates) in every bin

Space Time Cubes



Temporal step assignment

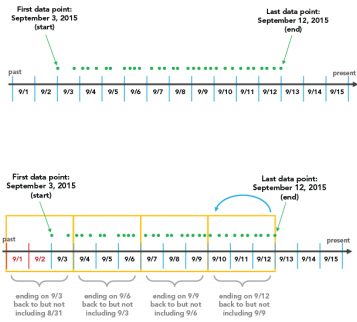
Must assign appropriate temporal ranges in order to avoid time based bias.

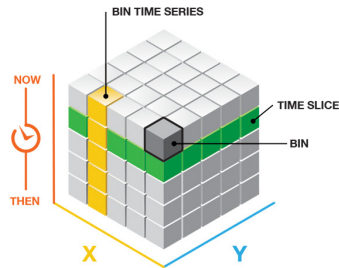
End-time vs Start time

Reference times

Template Cubes

Temporal steps and and bias





Mann-Kendall trend test

The Mann-Kendall trend test is performed on every location with data as an independent bin time-series test. The Mann-Kendall statistic is a rank correlation analysis for the bin count or value and their time sequence. The bin value for the first time period is compared to the bin value for the second. If the first is smaller than the second, the result is a +1. If the first is larger than the second, the result is -1. If the two values are tied, the result is zero. The result for each pair of time periods compared are summed. The expected sum is zero, indicating no trend in the values over time. Based on the variance for the values in the bin time series, the number of ties, and the number of time periods, the observed sum is compared to the expected sum (zero) to determine if the difference is statistically significant or not. The trend for each bin time series is recorded as a z-score and a p-value. A small p-value indicates the trend is statistically significant. The sign associated with the z-score determines if the trend is an increase in bin values (positive z-score) or a decrease in bin values (negative z-score).

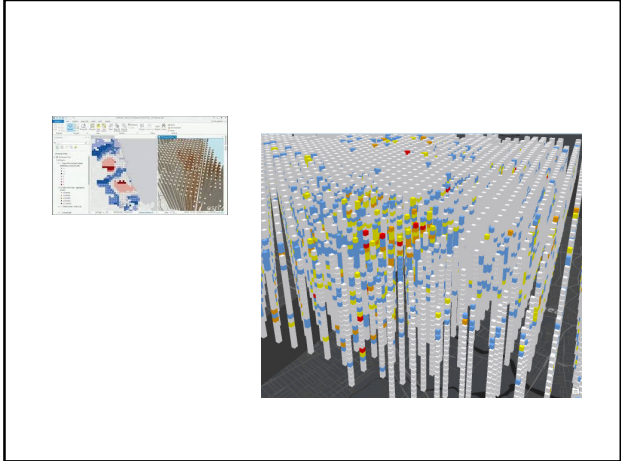
```
----- Space Time Cube Characteristics -----
Input feature time extent      2000-03-12 00:00:00
                               to 2009-03-12 00:00:00
Number of time steps          10
Time step interval            1 year
Time step alignment           Start
First time step temporal bias  0.405
                               on or after
First time step interval      2000-03-12 00:00:00
                               to before
                               2001-03-12 00:00:00
Last time step temporal bias   49.553
                               on or after
Last time step interval       2009-03-12 00:00:00
                               to before
                               2010-03-12 00:00:00

Cube extent across space      (coordinates in us feet)
Min X                          918816.0100
Max X                          92467.3259
Min Y                          912720.0100
Max Y                          912720.0100
Rows                            13
Columns                          17
Total bins                      2210

----- Summary Field - YEAR_MIN_SPACETIME -----
Total number of locations     321
Locations with at least one point
- associated bins              54
- % non-zero (sparseness)     16.30

----- Overall Data Trend - COFF -----
Trend direction                Not Significant
Trend statistic                 -0.7913
Trend p-value                  0.4658

----- Overall Data Trend - YEAR_MIN_SPACETIME -----
Trend direction                Not Significant
Trend statistic                 0.0000
Trend p-value                  1.0000
Completed script CreateSpaceTimeCube...
```



Emerging Hotspot analysis

The Emerging Hot Spot Analysis tool identifies trends in your data. It finds new, intensifying, diminishing, and sporadic hot and cold spots, for example. It takes as input a space-time cube created using the Create Space Time Cube tool. It then uses the Neighborhood Distance and Neighborhood Time Step parameter values you provide to calculate the Getis-Ord G_i^* statistic (Hot Spot Analysis) for each bin. Produces a Z score and P-value

Local Outlier analysis

Identifies statistically significant clusters and outliers in the context of both space and time. This tool is a space-time implementation of the Anselin Local Moran's I statistic.

<https://www.youtube.com/watch?v=d170TZgduy0>
