

# Automated Creation of High-Quality Maps Using SAS and Python

Jingsong Lu M.S., Christopher Goranson, Kevin Konty M.S., Farzad Mostashari, M.D. M.S.

Bureau of Epidemiology Services, New York City Department of Health and Mental Hygiene.

## OBJECTIVE

The objective was to minimize the amount of time spent on routine, daily analysis of syndromic data, integrate additional spatial analysis, create better maps, and cut response times to outbreaks.

## BACKGROUND

New York City ED syndromic surveillance data uses SaTScan to detect spatial signals. SaTScan analysis has been integrated into SAS since 2002, and signal maps have been generated from SAS since 2003<sup>[1]</sup>. Signal maps are created occasionally to investigate a severe outbreak based on the SaTScan results. Previous use and integration of additional GIS analysis in ArcGIS has been done manually, requiring more time, and running the risk of being less consistent than an automated method. This script now integrates the SAS, SaTScan and spatial analysis from ArcGIS to create high-quality maps in an automated procedure.

## METHODS

We developed a way to complement the strengths of SaTScan's spatial cluster analysis with that of ESRI's Spatial Analyst extension. Using Python, work flows from ArcGIS were first modeled using ModelBuilder and then exported as a geoprocessing script written in Python<sup>[2]</sup>. A second script was written in SAS that used the nested Python scripts to activate the necessary ArcGIS spatial analysis tools. The SaTScan analysis was also invoked through the SAS script using the X command. The final script was scheduled daily using Windows Scheduled Task Wizard, which launched a batch file, SSAW.bat. Three processes were run by the program SSAW.sas:

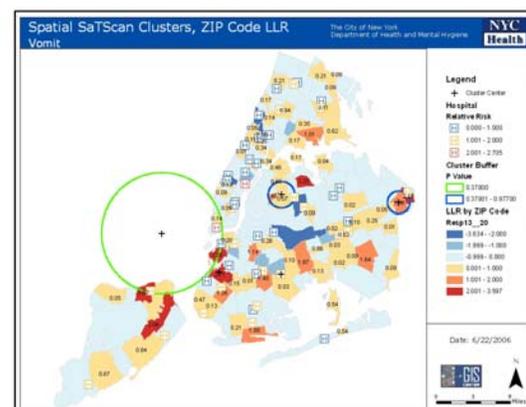
1. Data preparation and analysis in SAS.
2. SaTScan spatial analysis results were exported to a .dbf file. The option to run analysis by age group was included in the script. The .dbf contained relative risk, observed and expected cases, and p-values for each ZIP code and hospital; cluster information (including cluster center, radius, cluster expected and observed values, p-value; and intermediate SaTScan calculations, including zip and hospital log-likelihood ratios).
3. Nested Python script was run using a SAS file procedure. SAS macro variables within the script were used to control the path and file name of the output feature classes which allowed us to use separate Python scripts for different syndromes dates

while basic geo-processing steps remained the same. We used the X command to run Python from within SAS, created the basic GIS feature classes, performed the join between the GIS feature class and the .dbf, ran the ArcGIS tools, and finally buffered the results.

The resulting feature classes were exported to a personal geodatabase on a daily basis, and supporting .dbfs were saved in a date-stamped folder. Maps were generated using customized templates built for syndromic analysis output; results were sent via e-mail to various users and saved to a network drive.

## RESULTS

The process to map the same syndrome for consecutive days and related syndromes is now automated. The timeliness of response during an outbreak is improved - when retrospective analysis is required for an outbreak, the script cuts the time required to run analysis from 3-6 hours to under 10 minutes. Daily analysis that used to take 1 hour a day now is also automated with no user intervention required. Additionally, the script allows users unfamiliar with GIS tools to run and customize the analysis. Final maps are standardized regardless of the user.



## CONCLUSIONS

Timeliness of both daily analysis and responding to outbreaks is improved by the script, allowing users to focus on additional analysis or other tasks.

## REFERENCES

- [1] Heffernan R., Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D. 2004. Syndromic Surveillance in Public Health Practice, New York City. *Emerging Infectious Diseases*, Vol. 10, No. 5.
- [2] Tucker, C. 2004. Writing Geoprocessing Scripts with ArcGIS. Redlands, CA: Environmental Systems Research Institute, Inc.