A New Software Prototype for Geographic Profiling

Mike O'Leary

Towson University

2011 Crime Mapping Conference
Miami, FL
October 19-21, 2011

Acknowledgments

- This work has been supported by the National Institute of Justice, through grant 2009-SQ-B9-K014
- We thank the Baltimore County Police Department for providing the data used in this study
- We thank Phil Canter from the Baltimore County Police Department for his assistance.

The Geographic Profiling Problem

- The geographic profiling problem is to estimate the location of the home base of a serial criminal from the known locations of the elements of the offender's crimes.
  - The home base is also called the anchor point of the offender. It may be the offenders home, the home of a relative, a place of work, or even a favorite bar.

- We have developed a new tool for the geographic profiling problem.
  - It is free for download and use, and is entirely open source.
    - http://pages.towson.edu/moleary/Profiler.html
  - The tool is still under very active development, and should be considered a prototype.
  - Effectiveness and accuracy testing has only just begun.

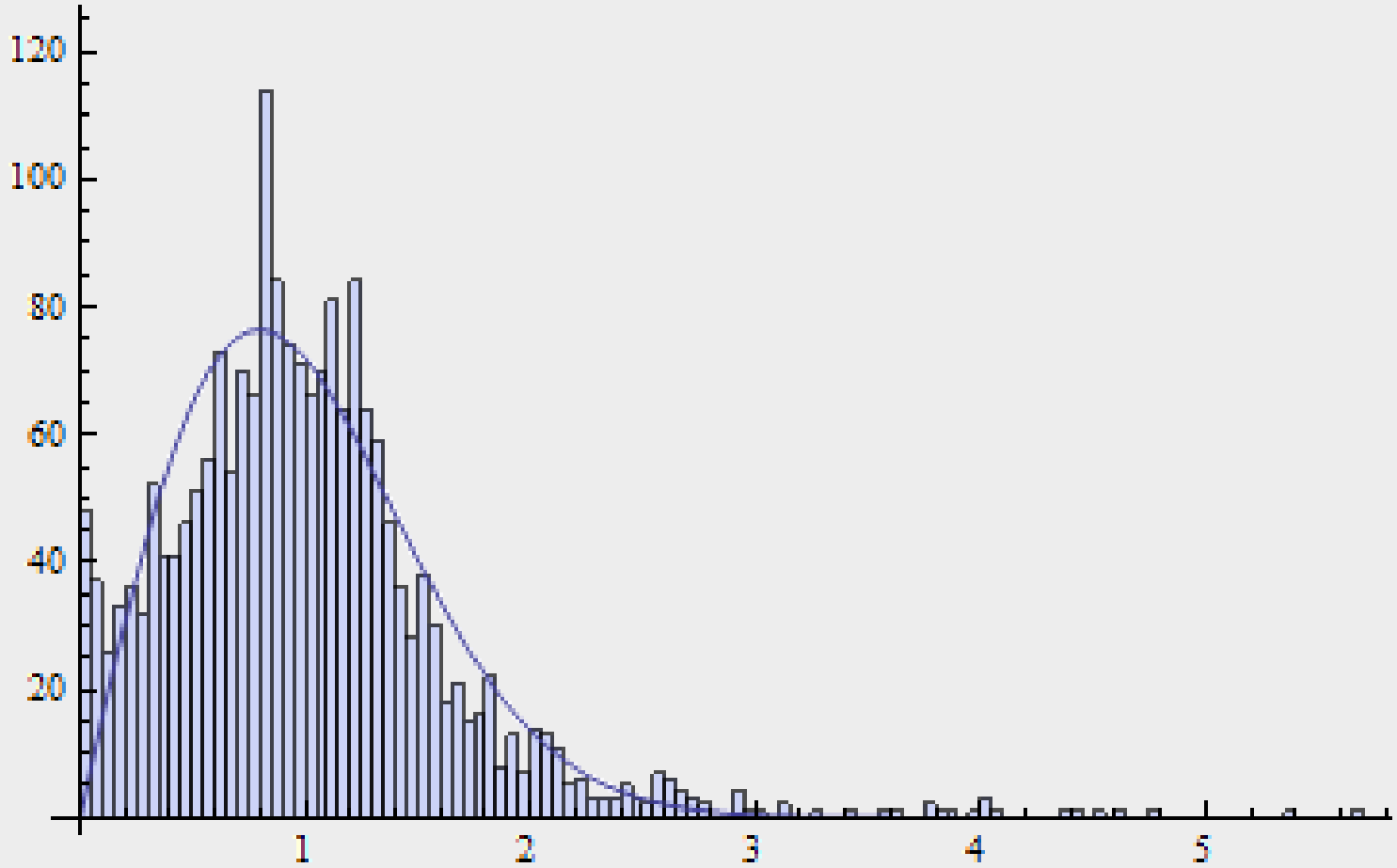
How Does it Work: The Mathematics

- We start by constructing a mathematical model for offender behavior; this gives $\mathrm{P}(\mathbf{x}|\mathbf{z})$, the probability that an offender with anchor point $\mathbf{z}$ offends at location $\mathbf{x}$.
- We assume that the probability that an offender commits a crime is:
  - Proportional to a distance decay function of the distance between the anchor point and the crime location, and
  - Proportional to the number of past crimes that have occurred at or near that location.
- We assume that crime sites are selected independently of one another
- We suppose that different offenders have different average distances they travel to offend.

Distance Decay

- We assume that the distance decay can be modeled by a two-dimensional normal distribution, whose corresponding one-dimensional model is Rayleigh.
- We assume that each offender has an average offense distance α, but that it varies between offenders.
- There is evidence to support this model.
  - Under these assumptions, the ratio of the distance to the crime site divided by the offender's average offense distance should follow a Rayleigh distribution with shape parameter equal to 1.
  - We have calculated this ratio for residential burglaries in Baltimore County. Once commuters are excluded, the Rayleigh distribution is a close match for the observed data.

Distance Decay

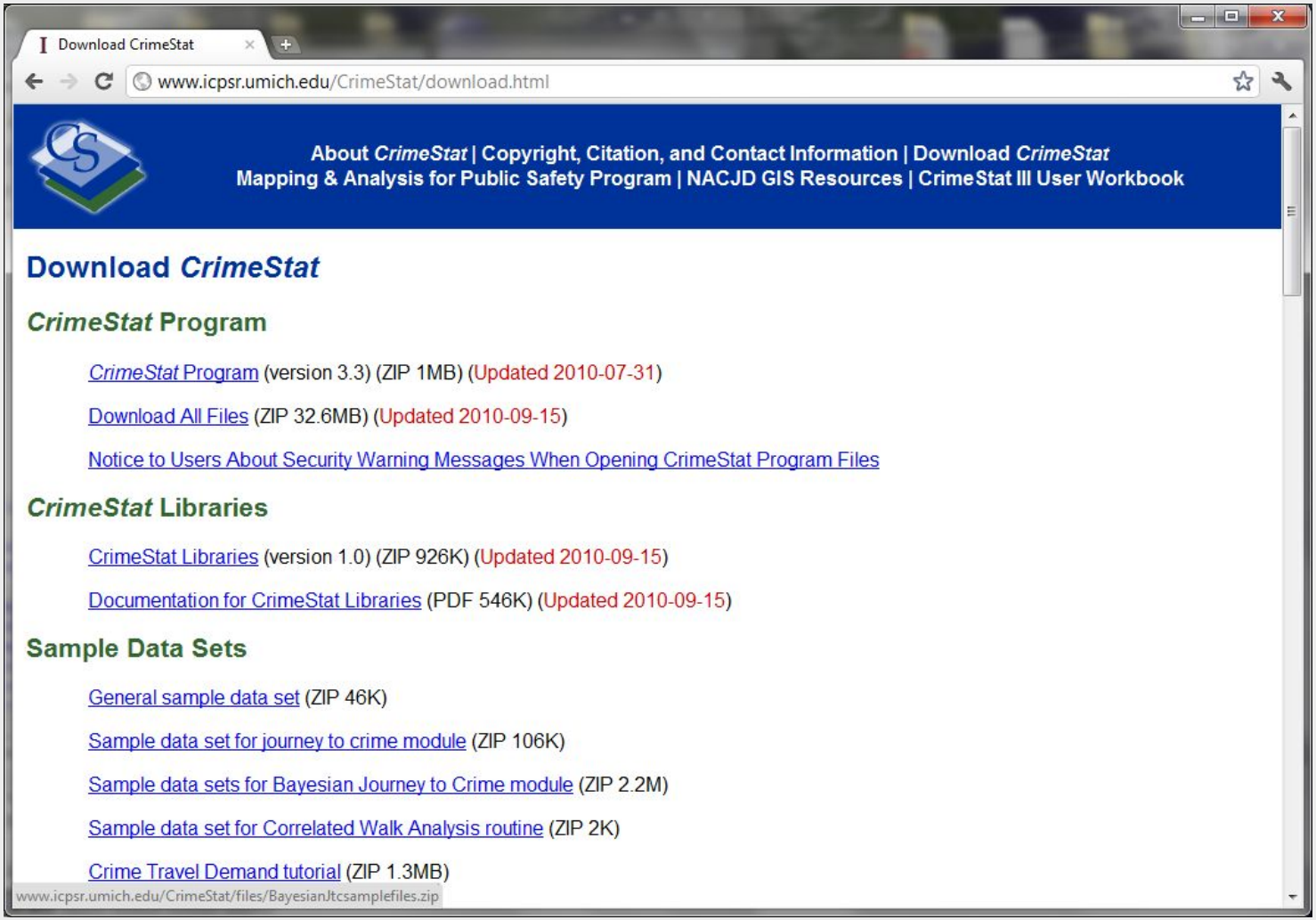


How Does it Work: The Mathematics

- Our assumptions mean that $\mathrm{P}(\mathbf{x}|\mathbf{z}) \propto \mathrm{D}(\mathrm{d}(\mathbf{x}, \mathbf{z}), \alpha)\mathrm{G}(\mathbf{x})$.
- Bayes' Theorem then let's us calculate the probability $\mathrm{P}(\mathbf{z}|\mathbf{x})$, the probability that an offender who has committed a crime at $\mathbf{x}$ has anchor point at $\mathbf{z}$.
  - To do so, we need a prior estimate for the location of the offender's anchor point, before information from the crime location is used.
    - For this purpose, reasonable choices include the local population density or the distribution of offenders from other crimes.
  - We also need to estimate the offender's average offense distance.
    - We again use Bayes' Theorem.
    - For the prior, we use the distribution of average offense distances for offenders, as calculated from historical data.
    - For the posterior, we use the distances between the crime locations.
- Provided we assume that crime sites are selected independently, we can extend this to series, and estimate the probability $\mathrm{P}(\mathbf{z}|\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$.

How Does it Work: The GUI

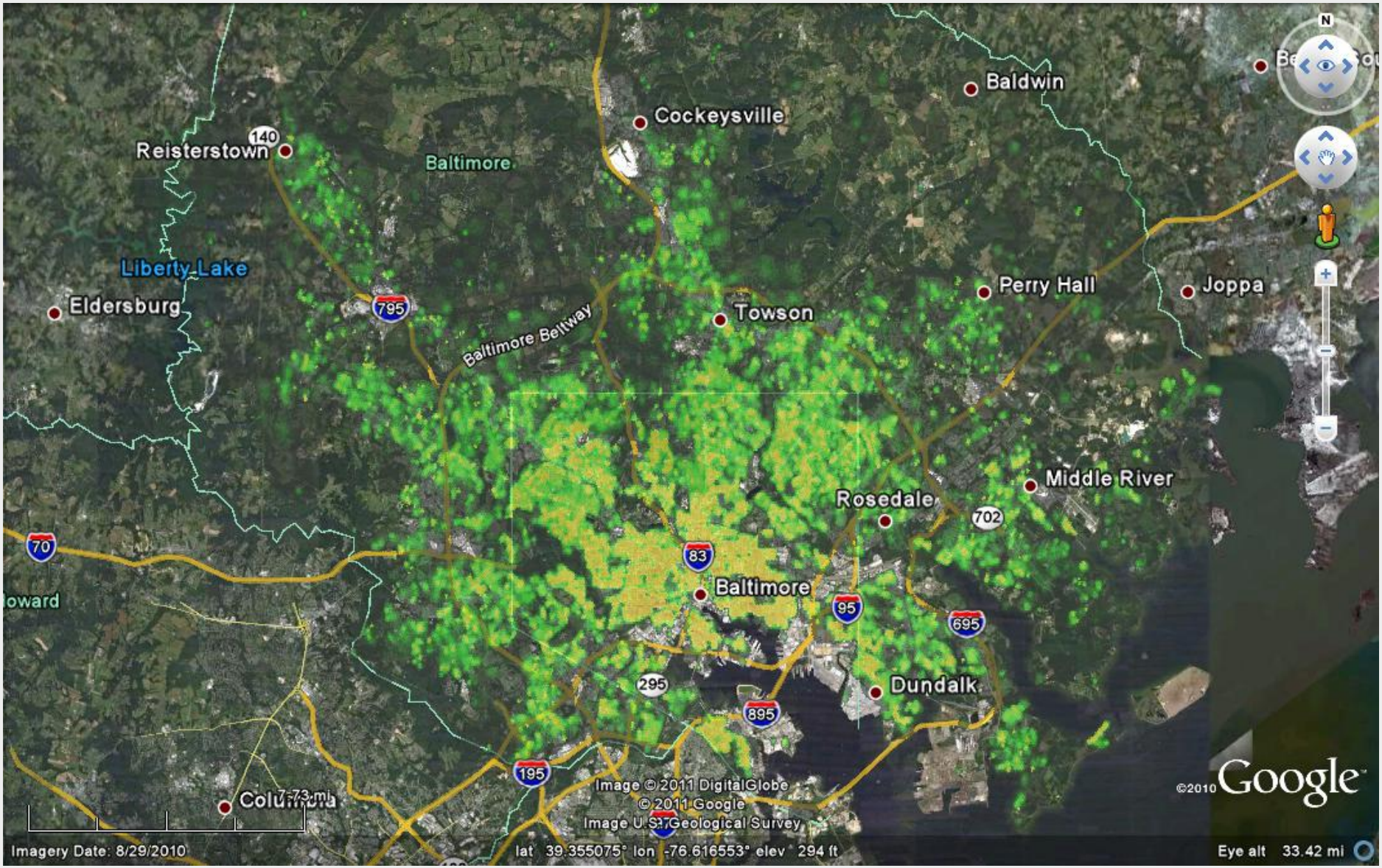
- We will use data obtained from Ned Levine's Crime Stat program:



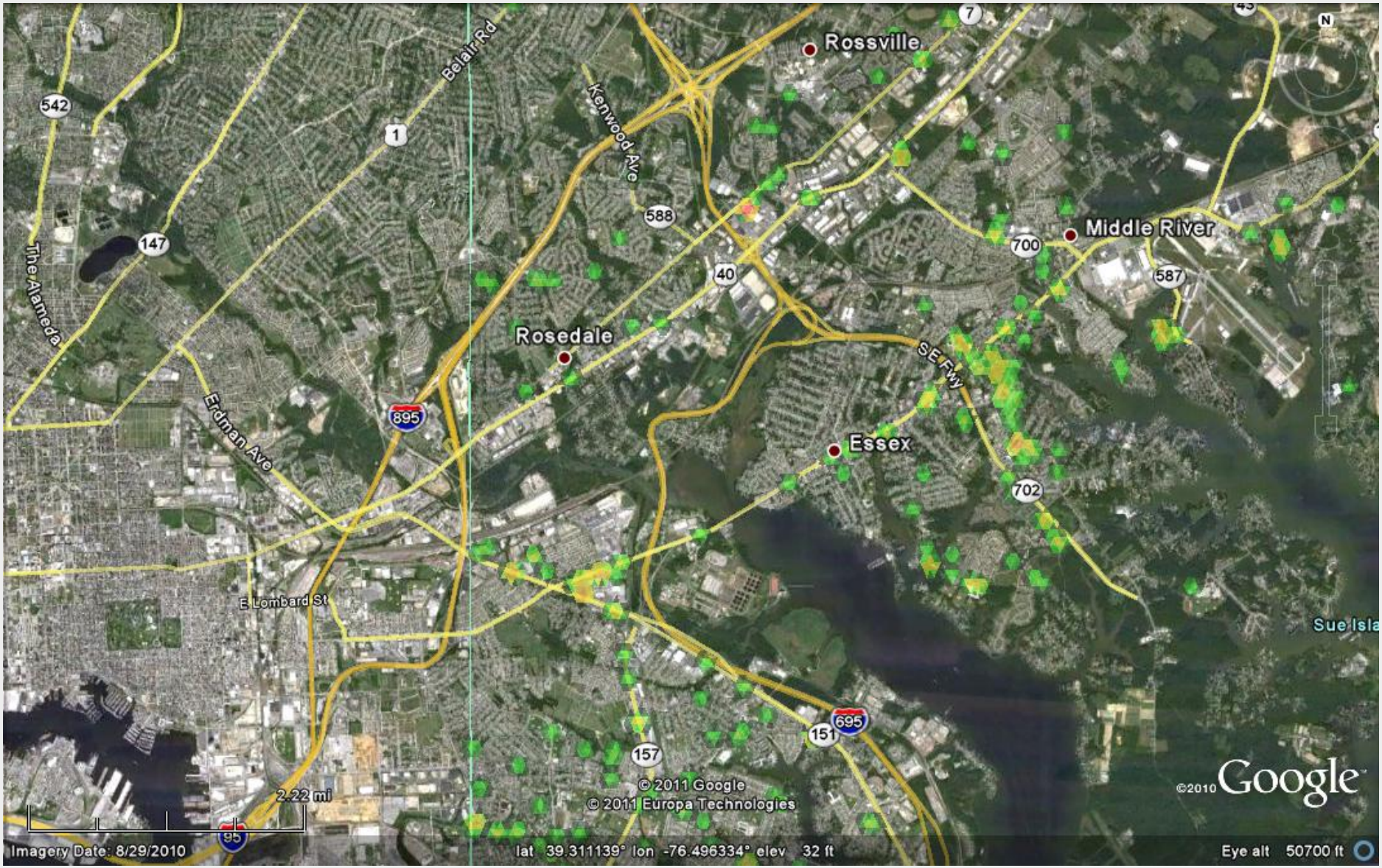
How Does it Work: The GUI

- Let's try the tool...

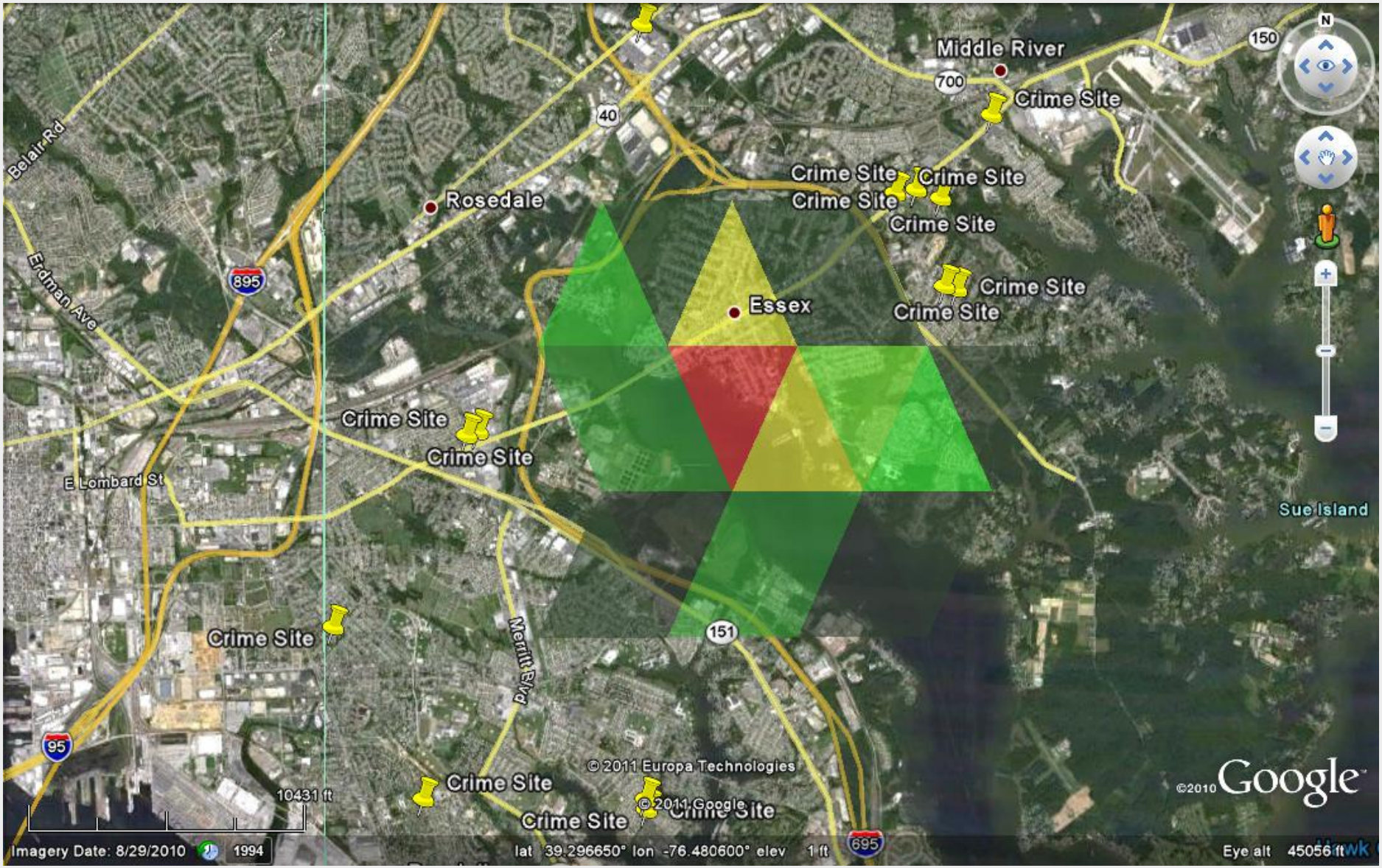
Results: Population Density



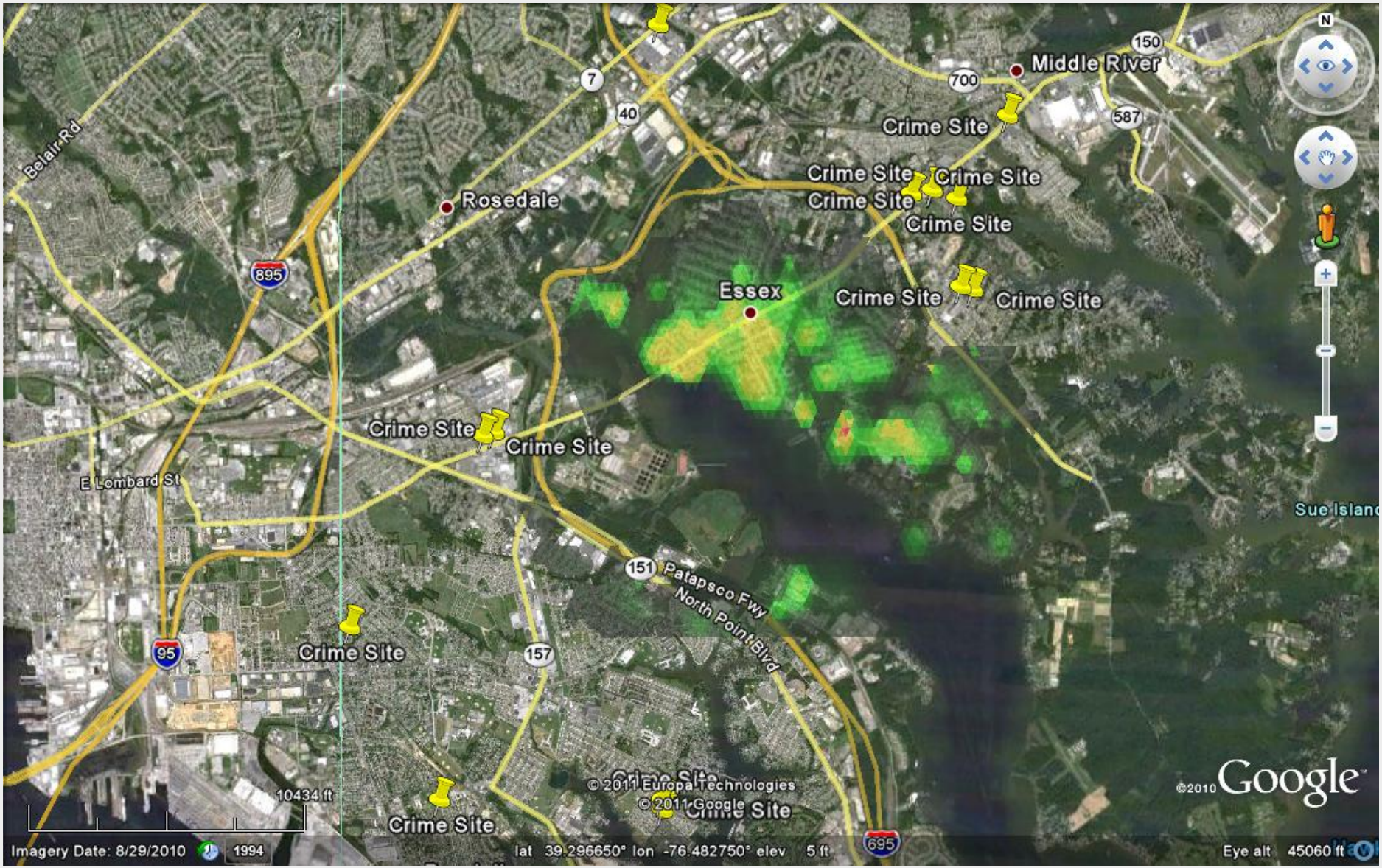
Results: Target Density



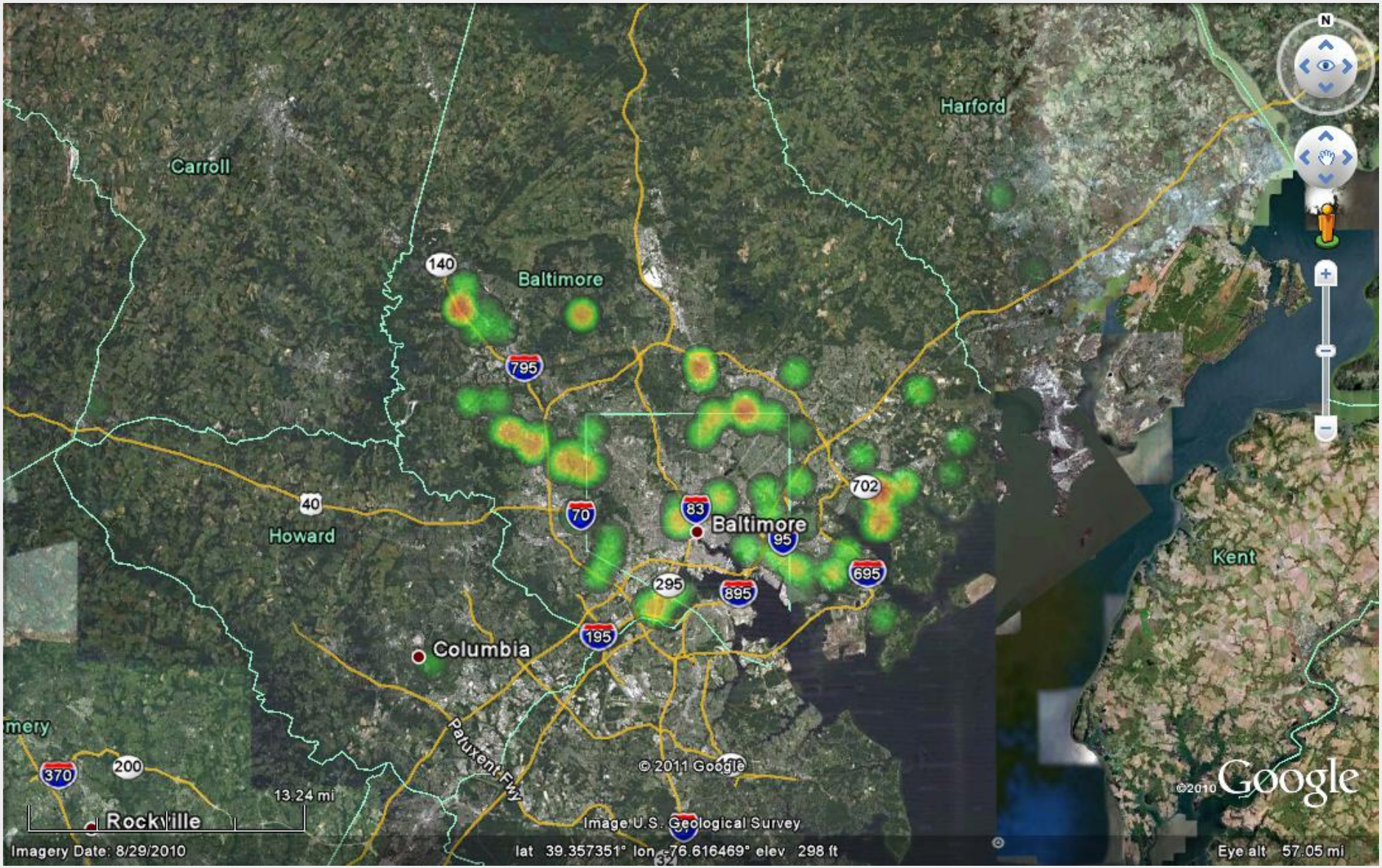
Results: Coarse Results



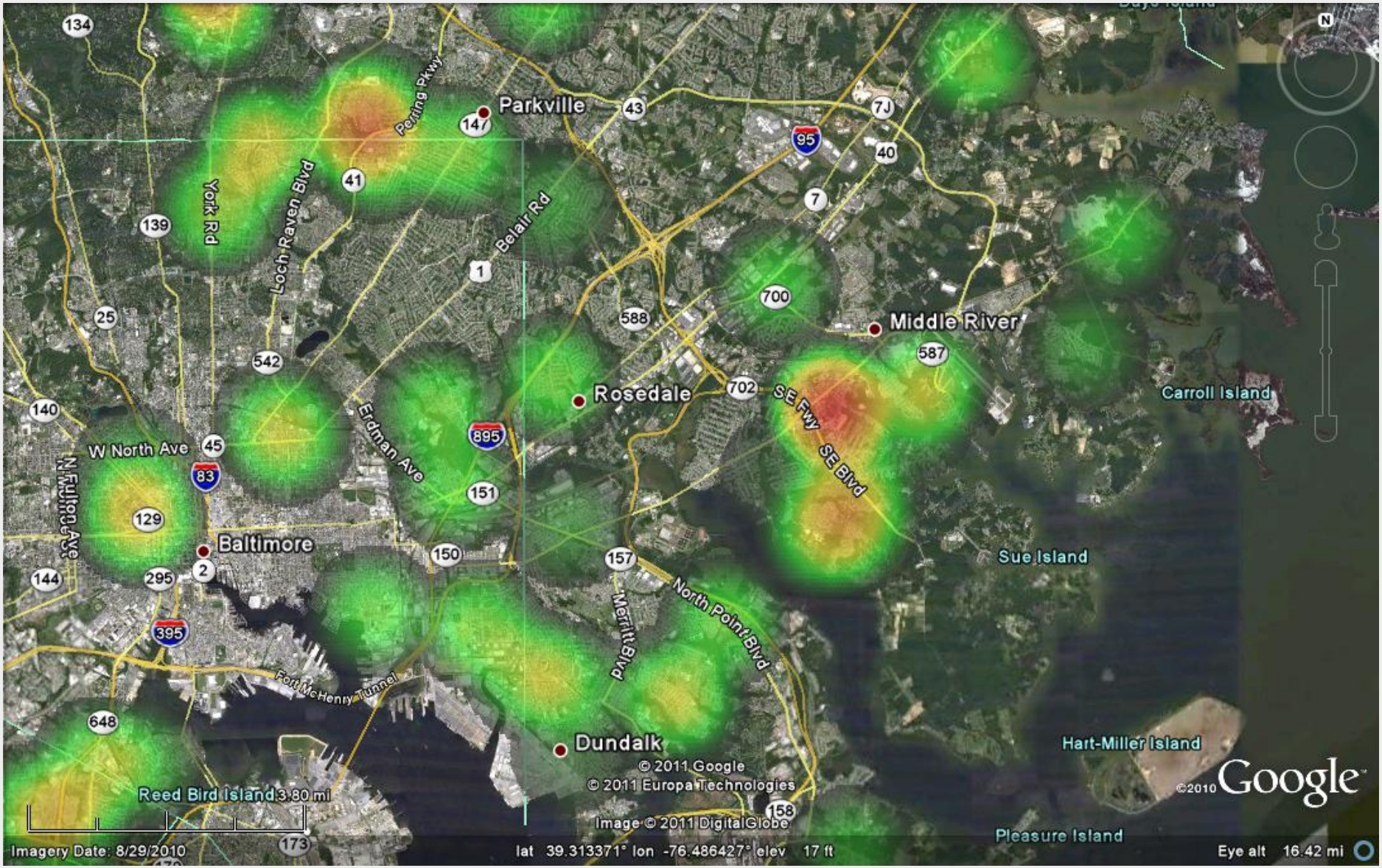
Results: Fine Results



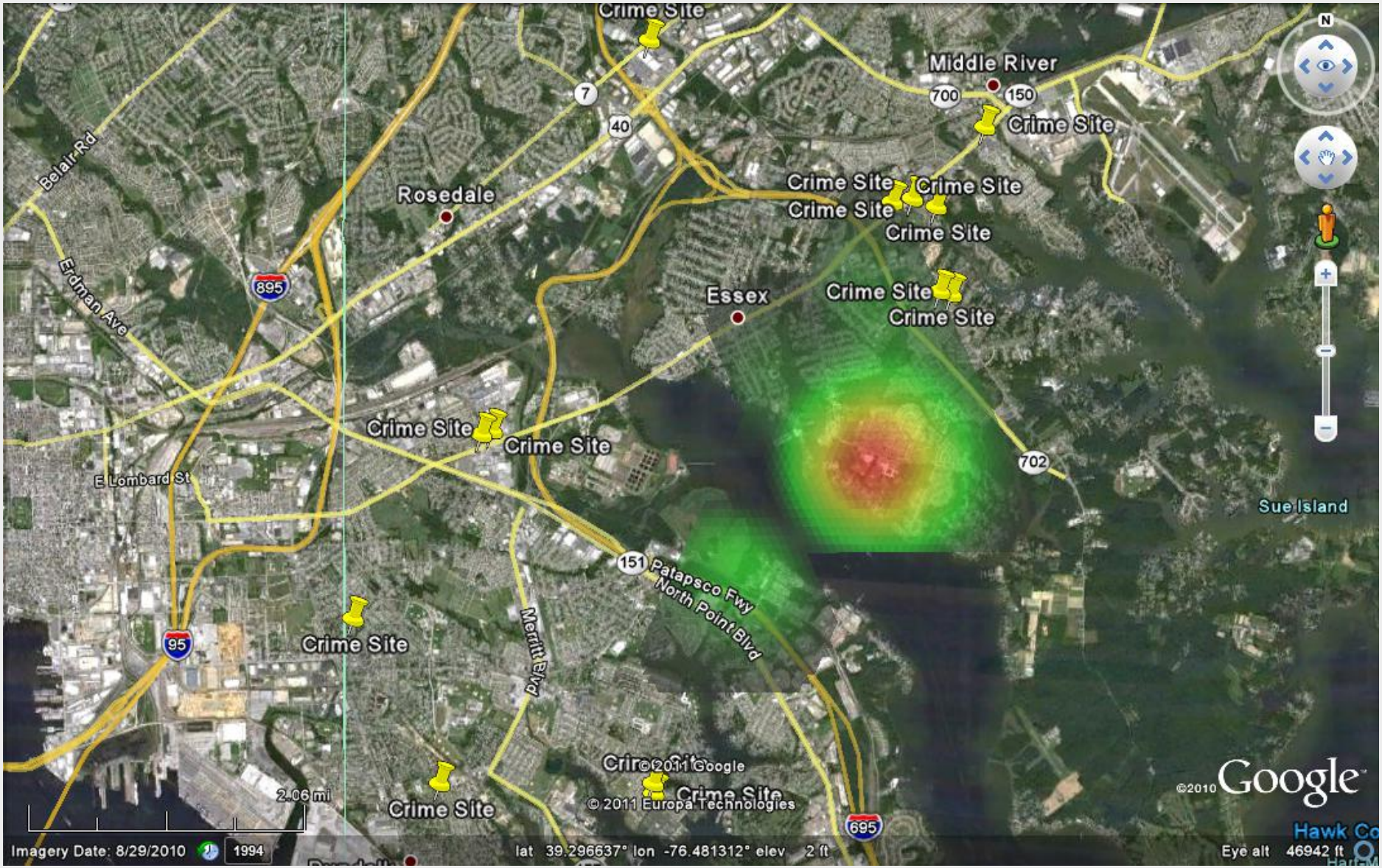
Results: Offender Density



Results: Offender Density



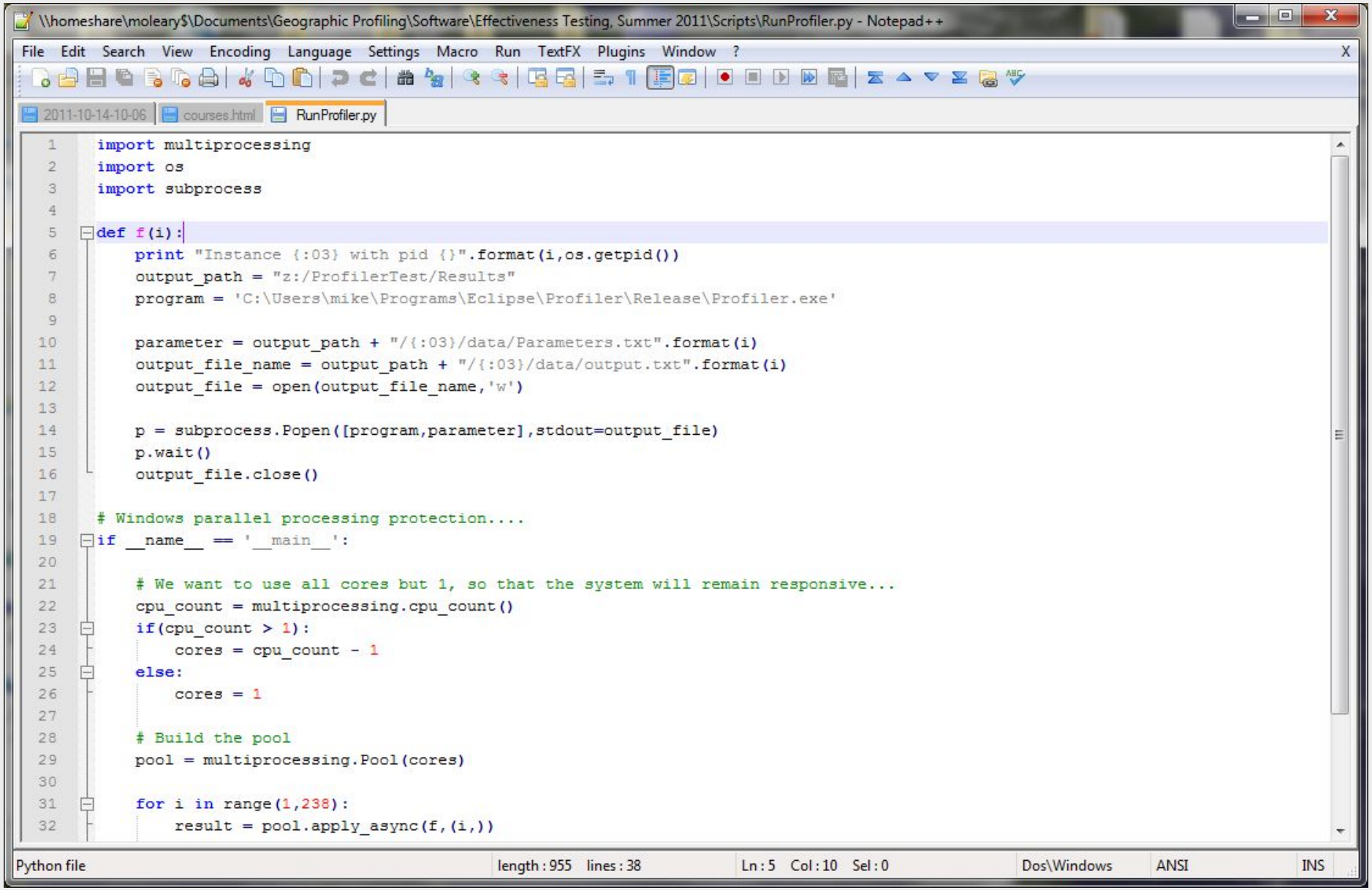
Results: Fine Results



How Does it Work: The Code

- There are two programs:
  - The graphical user interface
    - Written in C++ using the Qt Toolkit (4.7) and MinGW GCC (4.5.2)
    - Requires the Shapefile C Library (1.2)
      http://shapelib.maptools.org/
    - The source code is available
    - It is commented, but the documentation is not yet complete.
  - The analysis engine
    - Written in C++ with the MinGW GCC (4.5.2)
    - Requires the Shapefile C Library (1.2)
      http://shapelib.maptools.org/
    - Requires the Lapack++ Library (2.5.1)
      http://lapackpp.sourceforge.net/
    - The source code is available
    - It is fully commented, and the documentation is nearly complete
    - Is scriptable, and can be called from programs other than the graphical user interface

How Does it Work: Scripting



Strengths

- Strengths of this approach include:
  - The ability to incorporate geographic distribution of crimes
  - The ability to account for variations in offender distance decay patterns
  - The use of use variations in population density or locations of prior offenders
  - A completely open mathematical model, with all assumptions about offender behavior made explicit
  - A completely open source code base, available for others to examine and modify

Weaknesses

- Weaknesses of this approach include:
  - Effectiveness testing is not yet complete.
    - It is unclear what are the best choices for various parameters and model components
    - The preliminary results do not show significant improvements over existing methods
  - Computation times can be excessive
  - Population data currently only uses 2000 US Census data- though the modifcations to allow the use of 2010 data have begun