

## Content

CrimeRisk is a block group and higher level geographic database consisting of a series of standardized indexes for a range of serious crimes against both persons and property. It is derived from an extensive analysis of several years of crime reports from the vast majority of law enforcement jurisdictions nationwide. The crimes included in the database are the “Part 1” crimes and include murder, rape, robbery, assault, burglary, theft, and motor vehicle theft. These categories are the primary reporting categories used by the FBI in its Uniform Crime Report (UCR), with the exception of Arson, for which data is very inconsistently reported at the jurisdictional level. Part II crimes are not reported in the detail databases and are generally available only for selected areas or at high levels of geography.

In accordance with the reporting procedures using in the UCR reports, aggregate indexes have been prepared for personal and property crimes separately, as well as a total index. While this provides a useful measure of the relative “overall” crime rate in an area, it must be recognized that these are unweighted indexes, in that a murder is weighted no more heavily than a purse snatching in the computation. For this reason, caution is advised when using any of the aggregate index values.

## Methodology

The primary source of CrimeRisk is a careful compilation and analysis of the FBI Uniform Crime Report databases. On an annual basis, the FBI collects data from each of about 16,000 separate law enforcement jurisdictions at the city, county, and state levels and compiles these into its annual Uniform Crime Report (UCR). The latest national crime reports can be obtained from the FBI web site in Excel format. While useful, the UCR provides detailed data only for the largest cities, counties, and metropolitan areas.

Recently, some large cities have begun posting very detailed crime reports to their websites that are available for download. These data have very specific locations which is useful for matching against specific block groups. If the data is detailed enough and is available for the last five years, we use it alongside the FBI UCR data. Currently only data for Chicago and Baltimore fits this criteria, but hopefully more cities will qualify as time goes by.

We are now using UCR data from 2008-2012. A considerable effort was made to correct a number of problems that are prevalent within the FBI databases, including:

- The standardization of jurisdictional names: the FBI does not employ Census bureau codes in its databases and the jurisdictional names contain numerous typographical errors and format discrepancies which needed to be manually corrected
- Reporting by individual jurisdictions can be inconsistent from year to year, in that data for some jurisdictions is missing for one or more years and required handling
- Reporting for some crime types is inconsistent between jurisdictions. The FBI handles this by simply suppressing the statistics entirely for those areas. This primarily affects the rape category for Illinois, where statistics are suppressed for all but the largest jurisdictions. These missing values were handled via the modeling process, in which rape estimates were prepared

- for these jurisdictions by using a model which related rape incidence to other crime types
- The standardization of the database to account for jurisdictional overlaps. For example, the California Highway Patrol has jurisdiction over only state and interstate highways in urban areas.
  - Crime rates in general have been declining over the past several years, so it was necessary to adjust the historical data to reflect current crime rates.

Once this correction and standardization effort was completed, the database consisted of a time series of five years of data covering:

- All cities and towns which have their own police agency
- All cities and towns where policing for the local jurisdiction is contracted to a higher level agency but which tracks statistics separately (e.g. the city of Thousand Oaks, California contracts with the Ventura County Sheriff's Department for police services, but the incident reports are separately compiled)
- A record for each state which covers the population not covered by either of the two cases above. This is normally a combination of County Sheriff (or equivalent) and State level jurisdictions.

For a very limited number of areas, such as New York City, the local jurisdiction spans several counties.

The initial models were undertaken using a subset of this database. In the smallest cities, a single murder will have a profound effect on the crime rate per 100,000 population that would severely distort the resulting models. A wide range of Census and current year demographic attributes were extracted from AGS' databases for the remaining areas (approximately 10,000 separate "jurisdictions"). This database was then used as the primary modeling database and was used later for scaling purposes.

Each of the seven crime types was modeled separately, using an initial range of about 100 socio-economic characteristics taken from the Census and AGS' current year estimates. The models constructed typically accounted for over 85% of the variance in crime rates at this "jurisdiction" level, although it should be noted that the results for property crimes were generally more reliable than for personal crimes.

The results of these models were then applied to the block group level using the same demographic attributes compiled at the block group level. The resulting estimates were then scaled to match the master database of jurisdictions. The block groups within each jurisdiction were scaled to match the crime totals for that jurisdiction. For block groups outside of these areas, results were normalized to match the state totals minus the jurisdiction totals within each state.

The final crime rate estimates were then weighted by population and aggregated to the national totals. The results were then scaled to match the 2012 preliminary estimates (at a state level) and converted to indexes relative to the national total.