Crime Data State-by-Year Data Set Documentation

Overview

The state-by-year crime data set includes data on several major crime categories reported each year by local police agencies to the FBI. These records are collected as part of the Uniform Crime Reports (UCR) and then aggregated to the state-by-year level and reported as both counts and rates per 100,000 state residents. These records are available for all 50 states for the years 2008 to 2022 (the District of Columbia is not included).

Data set is `.csv` formatted – to load into R, use the `read_csv()` function from `tidyverse`.

- Each row of the data set corresponds to a state and year, so rows can be uniquely identified using a combination of `statefip` or `state_name` and `year`.
- Documentation below lists all variables and includes definitions and coding notes.
    - Rate variables for each crime category are reported as crime rates per 100,000 people and are calculated as:

```
Rate = (Total Crimes in Given Category / State Population) * 100,000
```

- The data below is available from the FBI's Crime Data Explorer website as the Summary Reporting System (SRS) files.
    - You can read more about the coding and definitions of variables here.
    - When a criminal act is committed that involves multiple categories of crimes, these offenses are categorized using a hierarchy rule ranking crimes by severity – if someone murders a person and also robs them, this event is recorded in this data set as a murder and ***not*** a robbery or assault.

Variable Name	*Definition*	*Coding and Notes*
`year`	Year (covers 2008-2022)	
`statefip`	State FIPS code – numeric ID for states	See link here for list of states and codes (you can also use the `state_name` variable to ID states)
`state_name`	State name (string format)	
`total_population`	Total state population	Total population for each state in each year (used to construct crime rate per 100,000 variables – see definition above).
`violent_crime`	Total number of violent crimes	Includes homicides, robberies, and aggravated assault as well other types of crimes not reported in this data set (in particular, rape is not reported as a separate category in this data set because of changes in reporting over time).

Variable Name	*Definition*	*Coding and Notes*
`homicide`	Total number of homicides	Includes murder and manslaughter.
`robbery`	Total number of robberies	Robbery is theft committed in the presence of the victim and involves some type of assault – given the hierarchy rule (see definition on page 1 above), assaults during a robbery are ***not*** counted separately.
`aggravated_assault`	Total number of aggravated assaults	Aggravated assault entails attacking another person with the purpose of inflicting severe or aggravated bodily injury.
`property_crimes`	Total number of property crimes	Includes burglary, larceny, and motor vehicle theft as well as other types of property crimes not reported in this data set.
`burglary`	Total number of burglaries	Burglary entails the unlawful entry of a structure (home, business, etc.) for the purpose of committing a felony or theft.
`larceny`	Total number of larcenies	Larceny is synonymous with theft – the unlawful taking of someone else's property. Does ***not*** include motor vehicle theft (counted separately). Given the hierarchy rule, instances of robbery are ***not*** counted within the larceny category.
`motor_vehicle_theft`	Total number of motor vehicle thefts	Motor vehicle theft is larceny specifically involving motor vehicles (cars, trucks, etc.) and is counted separately from the larceny category above. Note that given the hierarchy rule, motor vehicle theft in the presence of the victim (sometimes referred to as "carjackings") are counted as robberies and ***not*** included here.
`rate_X`	Rate of crime category X per 100,000 people	To save space, I've abbreviated the documentation for the rate variables. Rates are available for each of the crime categories listed above and named according to the syntax `rate_X` for crime category `X`. Crime rates are calculated using the definition on the first page of this documentation file.