

# Showing the Big Picture: Examples of Blending Data From Different Sources

Laurie Brown  
United States, Social Security Administration

## Agenda

- ✍ Simple explanations of complex methodologies
- ✍ Definitions and measurements
- ✍ Making a match
- ✍ It's all a matter of time

## Simple Explanations of Complex Methodologies

- ✍ How can data be presented to help nontechnical users accurately answer questions without overburdening them with metadata and methodology?
  - ✍ Traditionally, metadata had been presented in an independent section or appendix
  - ✍ For nontechnical users, need to integrate with data presented

## Simple Explanations of Complex Methodologies

### How do we know who commits crime?

Three major sources provide information about the kinds of persons who commit crimes:

?**Official records** compiled by police, courts, jails, and prisons have the advantage that they offer information on the more serious crimes and criminals. However, these records are limited to only the crimes and criminals that come to the attention of law enforcement officials.

?**Self-report surveys**, in which people are asked whether they had committed crimes, can provide more complete information than official records about crimes and criminal whether or not they are detected or apprehended. But there is the danger that people will exaggerate, conceal, or forget offenses. Many self-report surveys are limited to people who are in correctional custody.

?**Victim surveys**, such as the National Crime Survey obtain information from crime victims including their observations of the age, race, and sex of assailants. Victim surveys give information not only about crime reported to the police but also about unreported crimes. A disadvantage is that crime of stealth (such as burglary, and auto theft) victims seldom ever see who committed the crime. Also, many victims of crime fail to tell interviewers about being victimized by relatives and other nonstrangers.

Source: *Report to the Nation on Crime and Justice, Second Edition*, Bureau of Justice Statistics, U.S. Department of Justice, 1988

## Simple Explanations of Complex Methodologies



## Definitions and Measurements

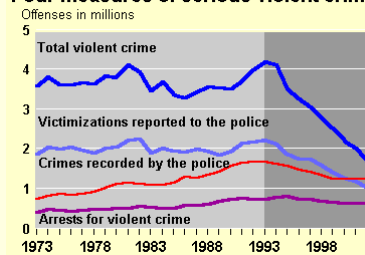
- ⌘ What to do when the sources use the same term for different things?
  - ⌘ Energy Information Administration standardized data definitions
    - ⌘ Begin definitions with a generic statement
    - ⌘ Limit supplementary descriptive information
  - ⌘ National Center for Health Statistics
    - ⌘ Present definitions by survey

## Definitions and Measurements

How can different measures of the same item be compared?

- Uniform Crime Reports
- National Crime Victimization Survey

Four measures of serious violent crime



The measures are:

**Total serious violent crime**

The number of homicides recorded by police plus the number of rapes, robberies, and aggravated assaults from the victimization survey whether or not they were reported to the police.

**Victimizations reported to the police**

The number of homicides recorded by police plus the number of rapes, robberies, and aggravated assaults from the victimization survey that victims said were reported to the police.

**Crimes recorded by the police**

The number of homicides, forcible rapes, robberies, and aggravated assaults included in the Uniform Crime Reports of the FBI excluding commercial robberies and crimes that involved victims under age 12.

**Arrests for violent crimes**

The number of persons arrested for homicide, forcible rape, robbery or aggravated assault as reported by law enforcement agencies to the FBI.

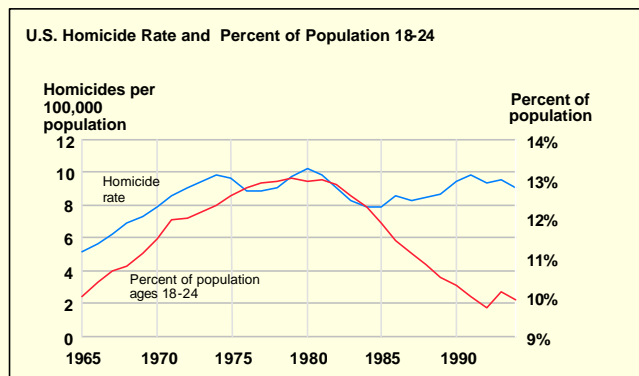
## Definitions and Measurements

What should be considered when selecting the most appropriate measure from combined or overlapping sources?

- What is the question being answered?
- Which source is more appropriate or accurate?

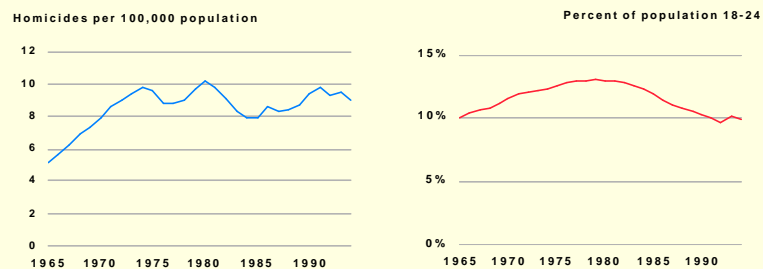
## Making a Match

- How can findings from two or more sources be presented when the units of analysis are not the same?



## Making a Match

- Small multiples



## It's All a Matter of Time

---

### ✍ What to do when the time periods covered do not correspond?

Information was gathered from an array of sources including:

- National Crime Victimization Survey (1992-2001)
- School Crime Supplement to the National Crime Victimization Survey (1995, 1999, and 2001)
- Youth Risk Behavior Survey (1993, 1995, 1997, 1999, and 2001)
- School Survey on Crime and Safety (2000)
- School and Staffing Survey (1993-94 and 1999-2000).

Highlights include the following:

- Students age 12-18 were victims of about 1.2 million crimes of theft and 764,000 nonfatal crimes of violence or theft at school in 2001.
- Data on homicides and suicides at school show there were 32 school-associated violent deaths in the United States between July 1, 1999 and June 30, 2000, including 24 homicides, 16 of which involved school-age children.
- Between 1995 and 2001, the percentage of students who reported being victims of crime at school decreased from 10 percent to 6 percent.

## Conclusion

---

- ✍ What do users really need to know in order to understand the data?
- ✍ How can that information best be presented, both in terms of publication design and the use of terms and concepts users will understand?