Uniform Crime Reporting (UCR) Program Data: An Opinionated Guide to FBI Data

Jacob Kaplan, Ph.D.

2024-09-17

Table of Contents

# 1 Preface

If you have read an article about crime or arrests in the United States in the last half century, in most cases it was referring to the FBI’s Uniform Crime Reporting Program Data, otherwise known as UCR data. UCR data is, with the exception of the more detailed data that only covers murders, a *monthly number of crimes or arrests reported to a single police agency* which is then gathered by the FBI into one file that includes all reporting agencies. It is actually a collection of different datasets, all of which have information about crimes and arrests that occur in a particular jurisdiction. Think of your home town. This data will tell you how many crimes were reported for a small number of crime categories or how many people (broken down by age, sex, and race) were arrested for a (larger) set of crime categories in that city (if the city has multiple police agencies then each agency will report crimes/arrests under their jurisdiction though the largest agency - usually the local police department - will cover the vast majority of crimes/arrests in that city) in a given month.

This is a very broad measure of crime, and its uses in research - or uses for understanding crime at all - is fairly limited. Yet is has become over much of the last century - and will likely remain among researchers for at least the next decade - the most important crime data in the United States.

UCR data is important for three reasons:

1. The definitions are standard, and all agencies (tend to) follow them so you can compare across agencies and over time.[[1]](#footnote-20)
2. The data is available since 1960 (for most of the datasets) so there is a long period of available data.[[2]](#footnote-21)
3. The data is available for most of the 18,000 police agencies in the United States so you can compare across agencies.

More than many other datasets, there will be times when using UCR data that you will think “that is weird”. This book will cover this weirdness and when we think the weirdness is just an odd - but acceptable - quirk of the data, and when it is a sign of a big problem in the data or in that particular variable and that we should avoid using it. For most of this book we will be discussing the caveats of the above reasons - or, more directly, why these assumptions are wrong - but these are the reasons why the data is so influential.

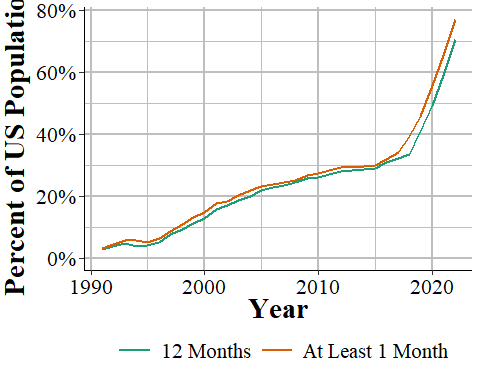


Figure 1.1: The annual percent of the United States population that is covered by an agency reporting data to NIBRS.

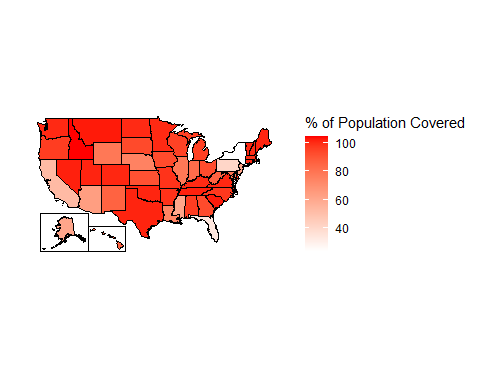


Figure 1.2: The percent of each state’s population that is covered by police agencies reporting at least one month of data to NIBRS, 2022.

## 1.1 Goal of the book

By the end of each chapter you should have a firm grasp on the dataset that is covered and how to use it properly. However, this book cannot possibly cover every potential use case for the data so make sure to carefully examine the data yourself for your own particular use.

I get a lot of emails from people asking questions about this data so my own goal is to create a single place that answers as many questions as I can about the data. Again, this is among the most commonly used crime datasets and there are still many current papers published with incorrect information about the data (including such simple aspects like what geographic unit data is in and what time unit it is in). So hopefully this book will decrease the number of misconceptions about this data, increasing overall research quality.

## 1.2 Structure of the book

This book will be divided into ten chapters: this chapter, an intro chapter briefly summarizing each dataset and going over overall issues with UCR data, and seven chapters each covering one of the seven UCR datasets. The final chapter will cover county-level UCR data, a commonly used but highly flawed aggregation of UCR data that I recommend against using. Each chapter will follow the same format: we will start with a brief summary of the data such as when it first because available and how it can be used. Next we will look at how many agencies report their data to this dataset, often looking at how to measure this reporting rate a couple of different ways. Finally, we will cover the important variables included in the data and how to use them properly (including not using them at all) - this will be the bulk of each chapter.

## 1.3 Citing this book

If this data was useful in your research, please cite it. To cite this book, please use the below citation:

Kaplan J (2024). *Uniform Crime Reporting (UCR) Program Data: An Opinionated Guide to FBI Data*. <https://ucrbook.com/>.

BibTeX format:

@Manual{ucrbook,  
 title = {Uniform Crime Reporting (UCR) Program Data: An Opinionated Guide to FBI Data},  
 author = {{Jacob Kaplan}},  
 year = {2024},  
 url = {https://ucrbook.com/},  
}

## 1.4 Sources of UCR data

### 1.4.1 My own collection

#### 1.4.1.1 openICPSR

#### 1.4.1.2 [Crimedatatool.com][<https://crimedatatool.com/>]

### 1.4.2 NACJD

### 1.4.3 FBI (raw data)

### 1.4.4 Raw data

### 1.4.5 Crime Data Explorer

### 1.4.6 Crimes in the United States report

### 1.4.7 FBI (Crime Data Explorer)

### 1.4.8 FBI (Crimes in the United States Report)

There are a few different sources of UCR data available today. First, and probably most commonly used, is the data put together by the [National Archive of Criminal Justice Data (NACJD)](https://www.icpsr.umich.edu/web/pages/NACJD/index.html)). This a team out of the University of Michigan who manages a huge number of criminal justice datasets and makes them available to the public. If you have any questions about crime data - UCR or other crime data - I highly recommend you reach out to them for answers. They have a collection of data and excellent documentation available for UCR data available on their site [here](https://www.icpsr.umich.edu/web/NACJD/series/57). One limitation to their data, however, is that each year of data is available as an individual file meaning that you will need to concatenate each year together into a single file. Some years also have different column names (generally minor changes like spelling robbery “rob” one year and “robb” the next) which requires more work to standardize before you could concatenate. They also only have data through 2016 which means that the most recent years (UCR data is available through 2019) of data are (as of this writing) unavailable.

Next, and most usable for the general public - but limited for researchers - is the FBI’s official website [Crime Data Explorer](https://crime-data-explorer.fr.cloud.gov/). On this site you can chose an agency and see annual crime data (remember, UCR data is monthly so this is not as detailed as it can be) for certain crimes (and not even all the crimes actually available in the data). This is okay for the general public but only provides a fraction of the data available in the actual data so is really not good for researchers.

It is worth mentioning a final source of UCR information. This is the annual Crimes in the United States report released by the FBI each year around the start of October. As an example, here is the [website for the 2019 report](https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019). In this report is summarized data which in most cases estimates missing data and provides information about national and subnational (though rarely city-level) crime data. As with the FBI’s site, it is only a fraction of the true data available so is not a very useful source of crime data for quality research. Still, this is a very common source of information used by researchers.

## 1.5 Recommended reading

While this book is designed to help researchers use this data, the FBI has an excellent manual on this data designed to help police agencies submit their data. That manual, called the “Summary Reporting System (SRS) User Manual” provides excellent definitions and examples of many variables included in the data. In this book when I quote the FBI, such as defining a crime, I quote from this manual. The manual is available to download as a PDF on the FBI’s site and I have also posted it on my GitHub page [here](https://github.com/jacobkap/ucrbook/blob/main/FBI%20Uniform%20Crime%20Reporting%20(UCR)%20Program%20User%20Manual.pdf) for convenience. I highly recommend that you read this manual before using the data. That manual, alongside this book which tries to explain when and how the agencies do not follow the manual, will provide a solid foundation for your understanding of UCR data.

## 1.6 How to contribute to this book

If you have any questions, suggestions (such as a topic to cover), or find any issues, please make a post on the [Issues page](https://github.com/jacobkap/ucrbook/issues) for this book on GitHub. On this page you can create a new issue (which is basically just a post on this forum) with a title and a longer description of your issue. You will need a GitHub account to make a post. Posting here lets me track issues and respond to your message or alert you when the issue is closed (i.e. I have finished or denied the request). Issues are also public so you can see if someone has already posted something similar.

For more minor issues like typos or grammar mistakes, you can edit the book directly through its GitHub page. That will make an update for me to accept, which will change the book to include your edit. To do that, click the edit button at the top of the site - the button is highlighted in the below figure. You will need to make a GitHub account to make edits. When you click on that button you will be taken to a page that looks like a Word Doc where you can make edits. Make any edits you want and then scroll to the bottom of the page. There you can write a short (please, no more than a sentence or two) description of what you have done and then submit the changes for me to review.

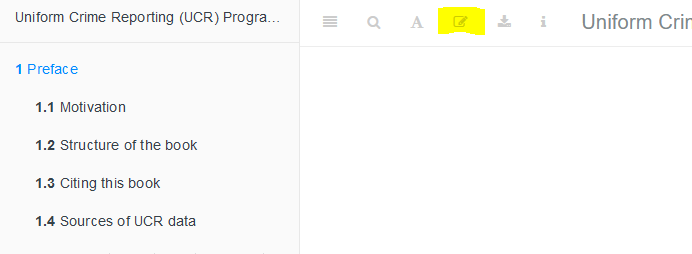


Figure 1.3: The edit button for how to make edits of this book.

Please only use the above two methods to contribute or make suggestions about the book. While it is a bit more work for you to do it this way, since you will need to make a GitHub account if you do not already have one, it helps me organize all the questions in one place and update the book if I decide to add answers to certain questions.

## 1.7 How to identify a particular agency (ORI codes)

In NIBRS and other FBI data sets, agencies are identified using **OR**iginating Agency **I**dentifiers or an ORI. An ORI is a unique ID code used to identify an agency.[[3]](#footnote-58) If we used the agency’s name we would end up with some duplicates since there can be multiple agencies in the country (and in a state, those this is very rare) with the same name. For example, if you looked for the Philadelphia Police Department using the agency name, you would find both the “Philadelphia Police Department” in Pennsylvania and the one in Mississippi. Each ORI is a 9-digit value starting with the state abbreviation[[4]](#footnote-59) followed by 7 numbers. In the UCR data (another FBI data set) the ORI uses only a 7-digit code - with only the 5 numbers following the state abbreviation instead of 7. So the NIBRS ORI codes are sometimes called ORI9. For nearly all agencies, the only difference between the UCR ORI and the NIBRS ORI is that the NIBRS ORI has “00” at the end so it is technically 9 characters long but is not any more specific than the 7-character UCR ORI code.

When dealing with specific agencies, make sure to use the ORI rather than the agency name to avoid any mistakes. For an easy way to find the ORI number of an agency, use [this page](https://crimedatatool.com/crosswalk.html) on my site. Type an agency name or an ORI code into the search section and it will return everything that is a match.

## 1.8 The data as you get it from the FBI

We will finish this overview of the SRS data by briefly talking about format of the data that is released by the FBI, before the processing done by myself or [NACJD](https://www.icpsr.umich.edu/web/pages/NACJD/index.html) that converts the data to a type that software like R or Stata or Excel can understand. The FBI releases their data as fixed-width ASCII files which are basically just an Excel file but with all of the columns squished together. As an example, Figure ?? shows what the data looks like as you receive it from the FBI for the Offenses Known and Clearances by Arrest dataset for 1960, the first year with data available. In the figure, it seems like there are multiple rows but that is just because the software that I opened the file in is not wide enough - in reality what is shown is a single row that is extremely wide because there are over 1,500 columns in this data. If you scroll down enough you will see the next row, but that is not shown in the current image. What is shown is a single row with a ton of columns all pushed up next to each other. Since all of the columns are squished together (the gaps are just blank spaces because the value there is a space, but that does not mean there is a in the data. Spaces are possible values in the data and are meaningful), you need some way to figure out which parts of the data belong in which column.

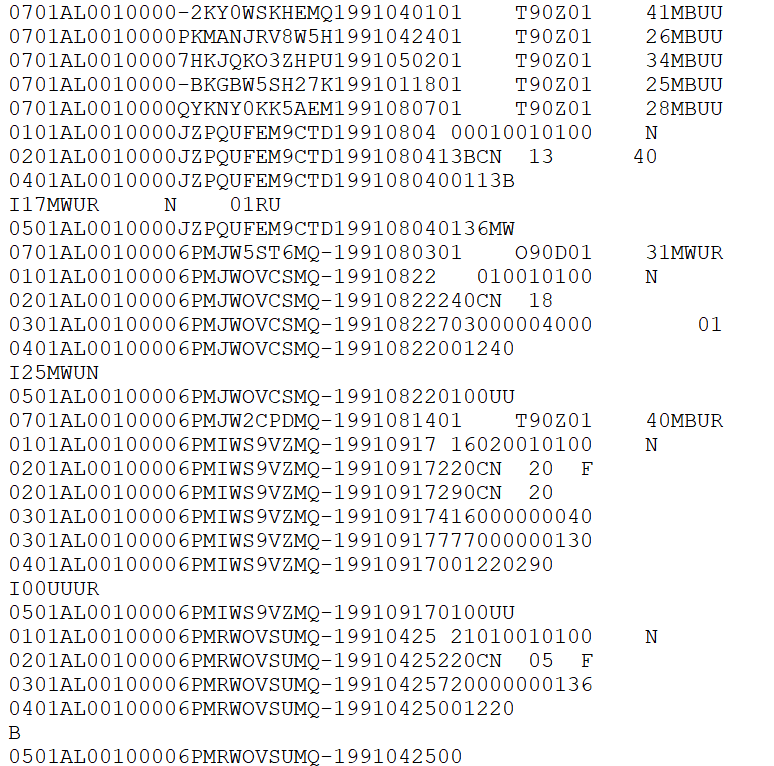


Figure 1.4: Fixed-width ASCII file for the 1991 National Incident-Based Reporting System (NIBRS) dataset.

The “fixed-width” part of the file type is how this works (the ASCII part basically means it is a text file). Each row is the same width - literally the same number of characters, including blank spaces. So you must tell the software you are using to process this file - by literally writing code in something called a “setup file” but is basically just instructions for whatever software you use (R, SPSS, Stata, SAS can all do this) - which characters are certain columns. For example, in this data the first character says which type of SRS data it is (1 means the Offenses Known and Clearances by Arrest data) and the next two characters (in the setup file written as 2-3 since it is characters 2 through 3 [inclusive]) are the state number (01 is the state code for Alabama). So we can read this row as the first column indicating it is an Offenses Known data, the second column indicating that it is for the state of Alabama, and so on for each of the remaining columns. To read in this data you will need a setup file that covers every column in the data (some software, like R, can handle just reading in the specific columns you want and do not need to include every column in the setup file).

The second important thing to know about reading in a fixed-width ASCII file is something called a “value label.”[[5]](#footnote-66) For example, in the above image we saw the characters 2-3 is the state and in the row we have the value “01” which means that the state is “Alabama.” Since this type of data is trying to be as small as efficient as possible, it often replaces longer values with shorter one and provides a translation for the software to use to convert it to the proper value when reading it. “Alabama” is more characters than “01” so it saves space to say “01” and just replace that with “Alabama” later on. So “01” would be the “value” and “Alabama” would be the “label” that it changes to once read.

Fixed-width ASCII files may seem awful to you reading it today, and it is awful to use. But it appears to be an efficient way to store data back many decades ago when data releases began but now is extremely inefficient - in terms of speed, file size, ease of use - compared to modern software so I am not sure why they *still* release data in this format. But they do, and even the more *modern* NIBRS data comes in this format. For you, however, the important part to understand is not how exactly to read this type of data, but to understand that people who made this data publicly available (such as myself and the team at NACJD) must make this conversion process.[[6]](#footnote-67) **This conversion process, from fixed-width ASCII to a useful format is the most dangerous step taken in using this data - and one that is nearly entirely unseen by researchers.**

Every line of code you write (or, for SPSS users, click you make) invites the possibility of making a mistake.[[7]](#footnote-68) The FBI does not provide a setup file with the fixed-width ASCII data so to read in this data you need to make it yourself. Since some SRS data are massive, this involves assigning the column width for thousands of columns and the value labels for hundreds of different value labels.[[8]](#footnote-69) A typo anywhere could have potentially far-reaching consequences, so this is a crucial weak point in the data cleaning process - and one in which I have not seen anything written about before. While I have been diligent in checking the setup files and my code to seek out any issues - and I know that NACJD has a robust checking process for their own work - that does not mean our work is perfect.[[9]](#footnote-70) Even with perfection in processing the raw data to useful files, decisions we make (e.g. what level to aggregate to, what is an outlier) can affect both what type of questions you can ask when using this data, and how well you can answer them.

## 1.9 Common issues

In this section we will discuss issues common to most or all of the SRS datasets. For some of these, we will come back to the issues in more detail in the chapter for the datasets most affected by the problem.

### 1.9.1 Population

Each of the SRS datasets include a population variable that has the estimated population under the jurisdiction of that agency.[[10]](#footnote-72) This variable is often used to create crime rates that control for population. In cases where jurisdiction overlaps, such as when a city has university police agencies or county sheriffs in counties where the cities in that county have their own police, SRS data assigns the population covered to the most local agency and zero population to the overlapping agency. So an agency’s population is the number of people in that jurisdiction that is not already covered by a different agency.

For example, the city of Los Angeles in California has nearly four million residents according to the US Census. There are multiple police agencies in the city, including the Los Angeles Police Department, the Los Angeles County Sheriff’s Office, the California Highway Patrol that operates in the area, airport and port police, and university police departments. If each agency reported the number of people in their jurisdiction - which all overlap with each other - we would end up with a population far higher than LA’s four million people. To prevent double-counting population when agency’s jurisdictions overlap, the non-primary agency will report a population of 0, even though they still report crime data like normal. As an example, in 2018 the police department for California State University - Los Angeles reported 92 thefts and a population of 0. Those 92 thefts are not counted in the Los Angeles Police Department data, even though the department counts the population. To get complete crime counts in Los Angeles, you would need to add up all police agencies within in the city; since the population value is 0 for non-LAPD agencies, both the population and the crime sum will be correct.

The SRS uses this method even when only parts of a jurisdiction overlaps. Los Angeles County Sheriff has a population of about one million people, far less than the actual county population (the number of residents, according to the Census) of about 10 million people. This is because the other nine million people are accounted for by other agencies, mainly the local police agencies in the cities that make up Los Angeles County.

The population value is the population who reside in that jurisdiction and does not count people who are in the area but do not live there, such as tourists or people who commute there for work. This means that using the population value to determine a rate can be misleading as some places have much higher numbers of non-residents in the area (e.g. Las Vegas, Washington D.C.) than others.

### 1.9.2 Voluntary reporting

When an agency reports their data to the FBI, they do so voluntarily - there is no national requirement to report.[[11]](#footnote-74) This means that there is inconsistency in which agencies report, how many months of the year they report for, and which variables they include in their data submissions.

In general, more agencies report their data every year and once an agency begins reporting data they tend to keep reporting. The SRS datasets are a collection of separate, though related, datasets and an agency can report to as many of these datasets as they want - an agency that reports to one dataset does not mean that they report to other datasets. Figure 1.5 shows the number of agencies that submitted at least one month of data to the Offenses Known and Clearances by Arrest data in the given year. For the first decade of available data under 8,000 agencies reported data and this grew to over 13,500 by the late 1970s before plateauing for about a decade. The number of agencies that reported their data actually declined in the 1990s, driven primarily by many Florida agencies temporarily dropping out, before growing steadily to nearly 17,000 agencies in 2010; from here it kept increasing but slower than before.

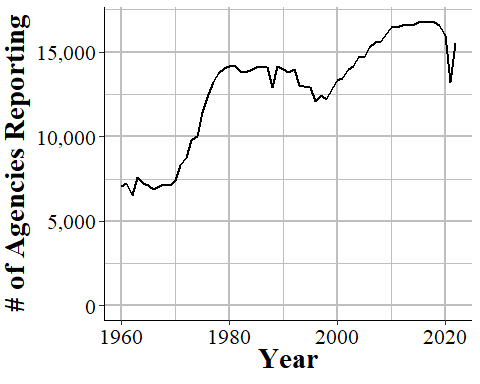


Figure 1.5: The annual number of agencies reporting to the Offenses Known and Clearances by Arrest dataset. Reporting is based on the agency reporting at least one month of data in that year.

There are approximately 18,000 police agencies in the United States so recent data has reports from nearly all agencies, while older data has far fewer agencies reporting. When trying to estimate to larger geographies, such as state or national-level, later years will be more accurate as you are missing less data. For earlier data, however, you are dealing with a smaller share of agencies meaning that you have a large amount of missing data and a less representative sample.

Figure 1.6 repeats the above figure but now including only agencies with 100,000 people or more in their jurisdiction. While these agencies have a far more linear trend than all agencies, the basic lesson is the same: recent data has most agencies reporting; old data excludes many agencies.

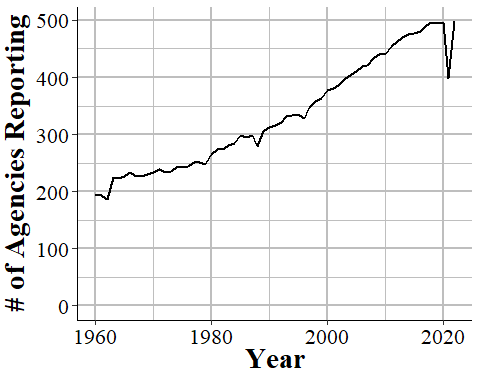


Figure 1.6: The annual number of agencies with a population of 100,000 or higher reporting to the Offenses Known and Clearances by Arrest dataset. Reporting is based on the agency reporting at least one month of data in that year.

This voluntariness extends beyond whether they report or not, but into which variables they report. While in practice most agencies report every crime when they report any, they do have the choice to report only a subset of offenses. This is especially true for subsets of larger categories - such as gun assaults, a subset of aggravated assaults, or marijuana possession arrests which is a subset of drug possession arrests. As an example, Figure 1.7 shows the annual number of aggravated assaults with a gun in New York City. In 2003 the New York Police Department stopped reporting this category of offense, resuming only in 2013. They continued to report the broader aggravated assault category, but not any of the subsections of aggravated assaults which say which weapon was used during the assault.

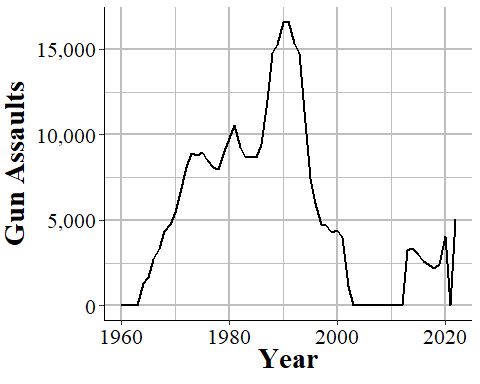


Figure 1.7: Monthly reports of gun assaults in New York City, 1960-2022.

Given that agencies can join or drop out of the SRS program at will, and report only partial data, it is highly important to carefully examine your data to make sure that there are no issues caused by this.

Even when an agency reports SRS data, and even when they report every crime category, they can report fewer than 12 months of data. In some cases they simply report all of their data in December, or report quarterly or semi-annually so some months have zero crimes reported while others count multiple months in that month’s data. One example of this is New York City, shown in Figure 1.8, in the early-2000s to the mid-2010s where they began reporting data quarterly instead of monthly.

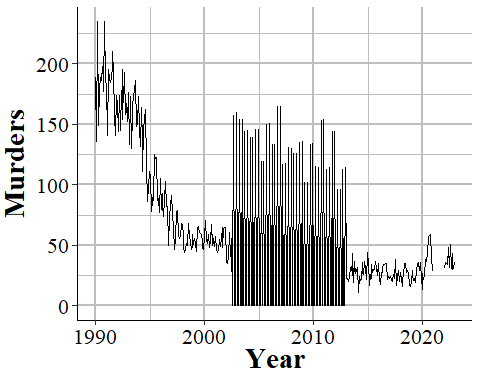


Figure 1.8: Monthly murders in New York City, 1990-2022. During the 2000s, the police department began reporting quarterly instead of monthly and then resumed monthly reporting.

When you sum up each month into an annual count, as shown in Figure 1.9, the problem disappears since the zero months are accounted for in the months that have the quarterly data. If you are using monthly data and only examine the data at the annual level, you will fall into the trap of having incorrect data that is hidden due to the level of aggregation examined. While cases like NYC are obvious when viewed monthly, for people that are including thousands of agencies in their data, it is unfeasible to look at each agency for each crime included. This can introduce errors as the best way to examine the data is manually viewing graphs and the automated method, looking for outliers through some kind of comparison to expected values, can be incorrect.

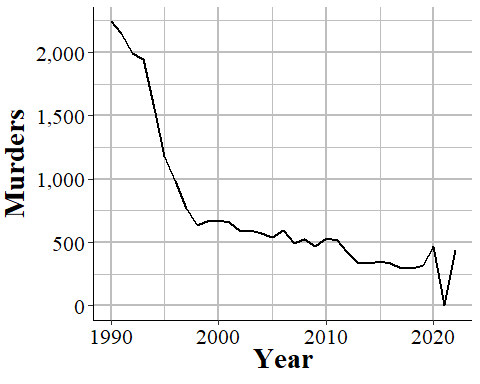


Figure 1.9: Annual murders in New York City, 1990-2022.

In other cases when agencies report fewer than 12 months of the year, they simply report partial data and as a result undercount crimes. Figure 1.10 shows annual murders in Miami-Dade, Florida and has three years of this issue occurring. The first two years with this issue are the two where zero murders are reported - this is because the agency did not report any months of data. The final year is in 2018, the last year of data in this graph, where it looks like murder suddenly dropped significantly. That is just because Miami-Dade only reported through June, so they are missing half of 2018.

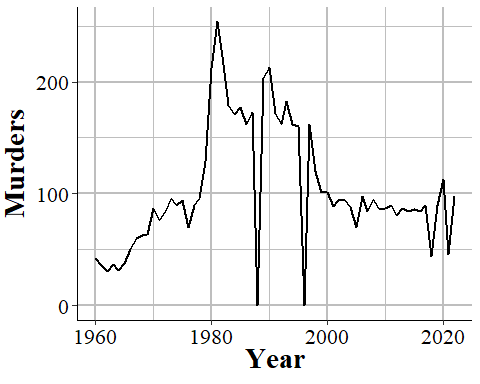


Figure 1.10: Annual murders in Miami-Dade, Florida, 1960-2022.

### 1.9.3 Zero crimes vs no reports

When an agency does not report, we see it in the data as reporting zero crimes, not reporting NA or any indicator that they did not report. In cases where the agency says they did not report that month we can be fairly sure (not entirely since that variable is not always accurate) that the zero crimes reported are simply that the agency did not report. In cases where the agency says they report that month but report zero crimes, we cannot be sure if that is a true no crimes reported to the agency or the agency not reporting to the SRS. As agencies can report some crimes but not others in a given month and still be considered reporting that month, just saying they reported does not mean that the zero is a true zero.

In some cases it is easy to see when a zero crimes reported is actually the agency not reporting. As Figure 1.7 shows with New York City gun assaults, there is a massive and sustained drop-off to zero crimes and then a sudden return years later. Obviously, going from hundreds of crimes to zero crimes is not a matter of crimes not occurring anymore, it is a matter of the agency not reporting - and New York City did report other crimes these years so in the data it says that they reported every month. So in agencies which tend to report many crimes - and many here can be a few as 10 a year since going from 10 to 0 is a big drop - a sudden report of zero crimes is probably just non-reporting.

Differentiating zero crimes and no reports becomes tricky in agencies that tend to report few crimes, which most small towns do. As an example, Figure 1.11 shows the annual reports of rape in Danville, California, a city of approximately 45,000 people. The city reports on average 2.8 rapes per year and in five years reported zero rapes. In cases like this it is not clear whether we should consider those zero years as true zeros (that no one was raped or reported their rape to the police) or whether the agency simply did not report rape data that year.

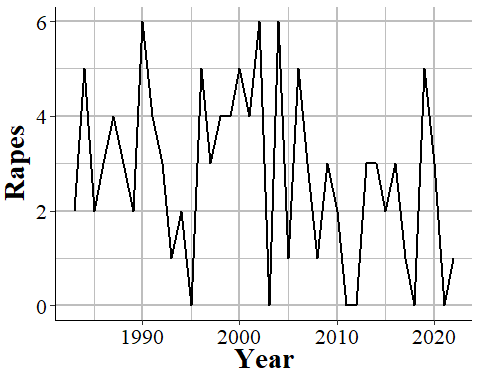


Figure 1.11: Annual rapes reported in Danville, CA, 1960-2022.

### 1.9.4 Agency data covered by another agency

# About the Author

**Jacob Kaplan**

I am a Professional Specialist at the School of Public and International Affairs (SPIA) and a member of Criminal Justice @ SPIA at Princeton University. My research focuses on law enforcement, including its impact on violent crime, the influence of removing ‘bad apple’ officers on reducing complaints against officers, the extent to which police forces represent the civilian populations they serve, and the role of race and political affiliations in shaping officer behavior. In addition to this, I conduct methodological research focused on the quality and usability of crime data, with a special emphasis on the FBI’s Uniform Crime Reporting (UCR) Program.

I am the author of [*A Criminologist’s Guide to R: Crime by the Numbers (Chapman & Hall/CRC The R Series, 2022)*](https://www.routledge.com/A-Criminologists-Guide-to-R-Crime-by-the-Numbers/Kaplan/p/book/9781032244075), an introductory textbook on the R programming language tailored for crime research, with a special focus on data cleaning and analysis. In addition, I have developed several R packages, including [fastDummies](https://jacobkap.github.io/fastDummies/), [asciiSetupReader](https://jacobkap.github.io/asciiSetupReader/), and [predictrace](https://jacobkap.github.io/predictrace/), to streamline the data analysis process for researchers. My website, [Crime Data Tool](https://crimedatatool.com/), offers users an interactive platform to explore crime data from thousands of agencies across hundreds of variables (e.g., arrests, offenses, demographics)—no data or programming skills required.

My research has been published in leading academic journals, such as the *Journal of Quantitative Criminology*, *Journal of Research in Crime and Delinquency*, *Journal of Interpersonal Violence*, and *American Political Science Review*. I hold a B.S. in Criminal Justice from California State University, Sacramento, and a M.S. and Ph.D. in Criminology from the University of Pennsylvania. I previously served on the FBI’s Criminal Justice Information Services (CJIS) Advisory Policy Board (APB) Uniform Crime Reporting (UCR) Subcommittee.

# 2 SRS Overview

One of the first, and most important, questions I think people have about crime is a simple one: is crime going up? Answering it seems simple - you just count up all the crimes that happen in an area and see if that number of bigger than it was in previous times.

However, putting this into practice invites a number of questions, all of which affect how we measure crime. First, we need to define what a crime is. Not philosophically what actions are crimes - or what should be crimes - but literally which of the many thousands of different criminal acts (crimes as defined by state or federal law) should be considered in this measure. Should murder count? Most people would say yes. How about jaywalking or speeding? Many would say probably not. Should marital rape be considered a crime? Now, certainly most people (all, I would hope) would say yes. But in much of the United States it was not a crime until the 1970s ([Bennice and Resick 2003](#ref-bennice2003marital); [McMahon-Howard, Clay-Warner, and Renzulli 2009](#ref-mcmahon2009criminalizing)). What about marijuana possession, an act that is illegal nationally and in some states, but legal elsewhere?

Next, we have to know what geographic and time unit to measure crimes at since these decisions determine how precise we can measure crime and when it changed. That is, if you are mugged on Jan 1st at exactly 12:15pm right outside your house, how do we record it? Should we be as precise as including the exact time and location (such as your home address)? Out of privacy concerns to the victim, should we only include a larger time unit (such as hour of the day or just the day without any time of day) or a larger geographic unit (such as a Census Tract or the city)? And what about when there are questions about geographic jurisdiction such as a local police and sheriff overlapping in their patrol area? Or when a crime happens over the course of several hours (e.g. a length burglary or kidnapping)? Should we count the start time, the end time, or somewhere in the middle?

The final question is that when a crime occurs, how do we know? That is, when we want to count how many crimes occurred do we ask people how often they’ve been victimized, do we ask people how often they commit a crime, do we look at crimes reported to police, crimes charged in a criminal court? Each of these measures will likely give different answers as to how many crimes occurred. And what about crimes where a victim reports it and the police investigate and decide that it did not occur (e.g. victim reports an act which was not actually illegal)? Or where the police say a crime occurred by the local District Attorney declines to prosecute?

The FBI answered all of these questions in 1929 when they began the Uniform Crime Reporting (SRS) Program Data, or SRS data for short. **Crime consists of eight crime categories - murder, rape, robbery, aggravated assault, burglary, motor vehicle theft, theft, and simple assault - that are reported to the police and is collected each month from each agency in the country.** So essentially we know how many of a small number of crime categories happen each month in each city (though some cities have multiple police agencies so even this is more complicated than it seems). These decisions, born primarily out of the resource limitations of 1929 (i.e. no computers), have had a major impact on criminology research. The first seven crime categories - known as “Index Crimes” or “Part 1 crimes” (or “Part I” sometimes) - are the ones used to measure crime in many criminology papers, even when the researchers have access to data that covers a broader selection of crimes than these.[[12]](#footnote-113) These are also the crimes that the news uses each year to report on how crime in the United States compares to the previous year. The crime data actually also includes the final crime, simple assault, though it is not included as an index crime and is, therefore, generally ignored by researchers - a large flaw in most studies that we will discuss in more detail in Section 3.1.2. In fact excluding simple assault is such a large issue that doing so has led us to exclude the most common violent crime from most government, news, and academic reports, leading to an undercount of violence in this country by approximately 50%.

In this chapter we will provide a brief overview of the datasets that make up the FBI’s Summary Reporting System (SRS) which you can think of as the “aggregate” data collection that is part of their Uniform Crime Reporting (UCR) Program Data. These datasets are primarily focused on aggregate data, mostly having monthly counts of crimes without many details about any particular crime. In comparison, the other half of the UCR collection, the National Incident-Based Reporting System (NIBRS) data is focused entirely on providing detail on every individual crime reported to the police. You can actually convert NIBRS data to SRS in most cases - and FBI does for agencies that report only NIBRS data - but there is no way to go from SRS to NIBRS.

## 2.1 Crimes included in the SRS datasets

SRS data covers only a subset - and for the crime data, a very small subset - of all crimes that can occur. It also only includes crimes reported to the police. So there are two levels of abstraction from a crime occurring to it being included in the data: a crime must occur that is one of the crimes included in one of the SRS datasets (we detail all of these crimes below) *and* the victim must report the incident to the police. While the crimes included are only a small selection of crimes - which were originally chosen since at the time the SRS was designed these were considered important offenses and among the best reported - this is an important first step to understanding the data.

SRS data should be understood as a loose collection of data that seeks to understand crimes and arrests in the United States. There are seven datasets in the SRS collection that each cover a different topic or a subset of a previously covered topic: crimes, arrests, property crimes specifically, homicides specifically, police officers killed or assaulted, arson specifically, and hate crimes.[[13]](#footnote-114) In this section we will go over the crimes included in the two main SRS datasets: Offenses Known and Clearances by Arrest and (which I like to call the “crime” dataset) and the Arrests by Age, Sex, and Race (the “arrests” dataset). These are the most commonly used SRS datasets and the stolen property and homicide datasets are simply more detailed subsets of these datasets. The hate crime dataset can cover a broader selection of offenses than in the crimes or arrests data, so we will discuss those in Chapter @ref(hate\_crimes).

## 2.2 A summary of each SRS dataset

The SRS collection of data can be roughly summarized into two groups: crime data and arrest data. While there are several datasets included in the SRS data collection, they are all extensions of one of the above groups. For arrest data, you have information about who (by race and by age-gender, but not by race-gender or race-age other than within race you know if the arrestee is an adult or a juvenile) was arrested for each crime. For crime data, you have monthly counts of a small number of crimes (many fewer than crimes covered in the arrest data) and then more specialized data on a subset of these crimes - information on homicides, hate crimes, assaults or killings of police officers, and stolen property.

Each of these datasets will have its own chapter in this book where we discuss the data thoroughly. Here is a very brief summary of each dataset which will help you know which one to use for your research. I still recommend reading that data’s chapter since it covers important caveats and uses (or misuses) of the data that would not be covered below.

### 2.2.1 Offenses Known and Clearances by Arrest (1960 - present)

The Offenses Known and Clearances by Arrest dataset - often called Return A, “Offenses Known” or, less commonly, OKCA - is the oldest and most commonly used dataset and measures crimes reported to the police. For this reason it is used as *the* main measure of crime in the United States, and I tend to call it the “crimes dataset.” This data answers the most basic questions about crimes: how many occurred? If you see crime data referenced in a news or academic article it is usually this data. This data also includes the number of crimes solved (though with a weaker definition of “solved” than you may think) and the number of crimes in which the agency concluded did not actually occur which they call an “unfounded” crime. This data has the monthly number of crimes - for a select group of crimes types - that occurred in an agency, as well as how many the police investigated and decided did not occur, and the number “cleared” by an arrest. The data uses something called a Hierarchy Rule which means that in incidents with multiple crimes, only the most serious is recorded - though in practice this affects only a small percent of cases, and primarily affects property crimes.

### 2.2.2 Arrests by Age, Sex, and Race (1974 - present)

The Arrests by Age, Sex, and Race dataset - often called ASR, or the “arrests data”, or sometimes the “Arrests by Age, Sex, Race, and Ethnicity though this is really misleading since most years do not even report ethnicity data - includes the monthly number of arrests for a variety of crimes and, unlike the crime data, breaks down this data by age and gender. This data includes a broader number of crime categories than the crime dataset (the Offenses Known and Clearances by Arrest data) though is less detailed on violent crimes since it does not breakdown aggravated assault or robberies by weapon type as the Offenses Known data does.

For each crime it says the number of arrests for each gender-age group with younger ages (15-24) showing the arrestee’s age to the year (e.g. age 16, age 17) and other ages grouping years together (e.g. age 25-29, 30-34, “under 10”). It also breaks down arrests by race-age by including the number of arrestees of each race (American Indian, Asian, Black, and White are the only included races) and if the arrestee is a juvenile (<18 years old) or an adult. The data does technically include a breakdown by ethnicity-age (e.g. juvenile-Hispanic, juvenile-non-Hispanic) but almost no agencies report this data (for most years zero agencies report ethnicity at all) so in practice the data does not include ethnicity. As the data includes counts of arrestees, people who are arrested multiple times are included in the data multiple times - it is not a measure of unique arrestees.

### 2.2.3 Law Enforcement Officers Killed and Assaulted (LEOKA) (1960 - present)

The Law Enforcement Officers Killed and Assaulted data, often called just by its acronym LEOKA (“LEE-OH-KUH”), has two main purposes.[[14]](#footnote-118) First, it provides counts of employees employed by each agency - broken down by if they are civilian employees or sworn officers, and also broken down by sex (male and female are the only options). And second, it measures how many officers were assaulted or killed (including officers who die accidentally such as in a car crash) in a given month. The assault data is also broken down into shift type (e.g. alone, with a partner, on foot, in a car, etc.), the offender’s weapon, and type of call they are responding to (e.g. robbery, disturbance, traffic stop). The killed data simply says how many officers are killed feloniously (i.e. murdered) or died accidentally (e.g. car crash) in a given month. The employee information is at the year-level so you know, for example, how many male police officers were employed in a given year at an agency, but do not know any more than that such as if the number employed changed over the year. This dataset is commonly used as a measure of police employees and is a generally reliable measure of how many police are employed by a police agency. The second part of this data, measuring assaults and deaths, is more flawed with missing data issues and data error issues (e.g. more officers killed than employed in an agency).

### 2.2.4 Supplementary Homicide Reports (SHR)

The Supplementary Homicide Reports dataset - often abbreviated to SHR - is the most detailed of the SRS datasets and provides information about the circumstances and participants (victim and offender demographics and relationship status) for homicides. For each homicide incident it tells you the age, gender, race, and ethnicity of each victim and offender as well as the relationship between the first victim and each of the offenders (but not the other victims in cases where there are multiple victims). It also tells you the weapon used by each offender and the circumstance of the killing, such as a “lovers triangle” or a gang-related murder. As with other SRS data, it also tells you the agency it occurred in and the month and year when the crime happened.

One important point of clarification: this is not the number of murders, though it does track that. This data also includes the number of homicides that are manslaughter by negligence (e.g. children playing with a gun, hunting accident) and justifiable homicides (i.e. not criminal). So be carefully when speaking about this data. It is murders but not only murders so you want to speak precisely.

### 2.2.5 Hate Crime Data (1991 - present)

This dataset covers crimes that are reported to the police and judged by the police to be motivated by hate. More specifically, they are (1) crimes which were (2) motivated - at least in part - by bias towards a certain person or group of people because of characteristics about them such as race, sexual orientation, or religion. The first part is key, they must be crimes - and really must be the selection of crimes that the FBI collects for this dataset. Biased actions that do not meet the standard of a crime, or are of a crime category not included in this data, are not considered hate crimes. For example, if someone yells at a Black person and uses racial slurs against them, it is clearly a racist action. For it to be included in this data, however, it would have to extend to a threat since “intimidation” is a crime included in this data but lesser actions such as simply insulting someone is not included. For the second part, the bias motivation, it must be against a group that the FBI includes in this data. For example, when this data collection began in 1991 crimes against transgender people were not counted so if a transgender person was assaulted or killed because they were transgender, this is not a hate crime recorded in the data (though it would have counted in the “Anti-Lesbian, Gay, Bisexual, Or Transgender, Mixed Group (LGBT)” bias motivation which was always reported).

### 2.2.6 Property Stolen and Recovered (Supplement to Return A) (1960 - present)

The Property Stolen and Recovered data - sometimes called the Supplement to Return A (Return A being another name for the Offenses Known and Clearances by Arrest dataset, the “crime” dataset) - provides monthly information about property-related offenses (theft, motor vehicle theft, robbery, and burglary), including the location of the offense (in broad categories like “gas station” or “residence”), what was stolen (e.g. clothing, livestock, firearms), and how much the stolen items were worth.[[15]](#footnote-122) It also includes robberies so is really the “stuff taken during a crime” dataset than the dataset about property crimes. The “recovered” part of this dataset covers the type and value of property recovered so you can use this, along with the type and value of property stolen, to determine what percent and type of items the police managed to recover.

Like most other SRS datasets this is at the agency-month level so you can, for example, learn how often burglaries occur at the victim’s home during the day, and if that rate changes over the year or differs across agencies. The data, however, provides no information about the offender or the victim (other than if the victim was an individual or a commercial business [[16]](#footnote-123)). The value of the property stolen is primarily based on the victim’s estimate of how much the item is worth (items that are decreased in value once used - such as cars - are supposed to be valued at the current market rate, but the data provides no indication of when it uses the current market rate or the victim’s estimate) so it should be used as a very rough estimate of value.

### 2.2.7 Arson (1979 - present)

The arson dataset provides monthly counts at the police agency-level for arsons that occur, and includes a breakdown of arsons by the type of arson (e.g. arson of a person’s home, arson of a vehicle, arson of a community/public building) and the estimated value of the damage caused by the arson. This data includes all arsons reported to the police or otherwise known to the police (e.g. discovered while on patrol) and also has a count of arsons that lead to an arrest (of at least one person who committed the arson) and reports that turned out to not be arsons (such as if an investigation found that the fire was started accidentally). This is essentially the Offenses Known and Clearances by Arrest data but only for arsons. The data even follows the same definitions for categories such as what counts as a cleared or unfounded crime. The primary additional variable is the estimated damage in dollars caused by the arson.

# 3 Offenses Known and Clearances by Arrest

The Offenses Known and Clearances by Arrest dataset - often called Return A, “Offenses Known” or, less commonly, OKCA - is the oldest and most commonly used dataset and measures crimes reported to the police. For this reason it is used as *the* main measure of crime in the United States, and I tend to call it the “crimes dataset.” This data answers the most basic questions about crimes: how many occurred? If you see crime data referenced in a news or academic article it is usually this data. This data also includes the number of crimes solved (though with a weaker definition of “solved” than you may think) and the number of crimes in which the agency concluded did not actually occur which they call an “unfounded” crime. This data has the monthly number of crimes - for a select group of crimes types - that occurred in an agency, as well as how many the police investigated and decided did not occur, and the number “cleared” by an arrest.

The Offenses Known data has been **the** crime data of record for nearly a century, and will likely still be for the next couple of decades at least. And that is due to its simplicity. This data is (with some exceptions we will get into) just the monthly number of crimes reported or otherwise known (e.g. discovered while on patrol) to a police agency for a small number of crimes. If you want to know, for example, how many murders or burglaries happened in your city last year, this is the dataset to turn to. This simplicity allows a much wider group of people to use the data; since it is monthly counts of crimes (with no breakdown by location, race, age, injury, etc.) you do not need much programming or analytic skills to use it.

I argued earlier in this book that NIBRS is a far superior dataset than anything in SRS, and I stand by that. NIBRS is superior in nearly every way. But for most people - the general public, reporters, many academics, etc. - what is important is ease of use and very basic descriptive statistics such as how many crimes happened in my city last year[[17]](#footnote-128), and is it getting more dangerous than it used to be? The Offenses Known data answers that, at least to a point. You can answer more complex questions using NIBRS but most people do not need those questions answered. They’re content with what can be answered using the Offenses Known data. And for good reason. “Is crime increasing in my city” is really the first and most important question about crime that people have.

Part of the superiority of NIBRS is that while it can answer much more complex questions than SRS datasets, it can also answer the same questions that the SRS can. That is because, by design, NIBRS collects the same information as SRS and can be converted to SRS data. Many agencies that submit NIBRS data do not submit SRS data at all, and the FBI converts the NIBRS data to its SRS counterpart for release to the public. This is an almost perfect comparison but not all of the same variables are collected, which will cause some issues that we will discuss in this chapter.

One thing you may have heard about this data is that it uses something called the Hierarchy Rule. Basically the rule says that when an incident involves more than one crime, such as a robbery and a murder, you only count the most serious crime. And most serious is based on the hierarchy the FBI has created. This is true and therefore the number of crimes - other than murder which is the most serious - recorded in this data is always an undercount. But you really should not be too concerned about this. As we will get into below, the Hierarchy Rule affects only a relatively small share of incidents so does not undercount crime by too much - and when it does it is primarily undercounting property crime.

## 3.1 Which crimes are included?

This dataset contains information on the number of “Index Crimes” (sometimes called Part I crimes) reported to each agency.[[18]](#footnote-129) These index crimes are a collection of eight crimes that, for historical reasons based largely by perceived importance and reliability of their reporting in the 1920s when the UCR program was first developed, are used as the primary measure of crime today. Other data sets in the UCR, such as the Arrests by Age, Sex, and Race data and the Hate Crime data have more crimes reported.

The crimes are, in order by the Hierarchy Rule (which we will discuss next):

1. Homicide
   * Murder and non-negligent manslaughter
   * Manslaughter by negligence
2. Rape
   * Rape
   * Attempted rape
3. Robbery
   * With a firearm
   * With a knife of cutting instrument
   * With a dangerous weapon not otherwise specified
   * Unarmed - using hands, fists, feet, etc.
4. Aggravated Assault (assault with a weapon or with the intent to cause serious bodily injury)
   * With a firearm
   * With a knife of cutting instrument
   * With a dangerous weapon not otherwise specified
   * Unarmed - using hands, fists, feet, etc.
5. Burglary
   * With forcible entry
   * Without forcible entry
   * Attempted burglary with forcible entry
6. Theft (other than of a motor vehicle)
7. Motor Vehicle Theft
   * Cars
   * Trucks and buses
   * Other vehicles
8. Arson
9. Simple Assault

Arson is considered an index crime but is not reported in this data - you need to use the separate Arson dataset of the UCR to get access to arson counts. See Chapter ?? for an overview of the Arson data. The ninth crime on that list, simple assault, is not considered an index crime but is nevertheless included in this data.

Each of the crimes in the list above, and their subcategories, are included in the UCR data. In most news and academic articles, however, you will see them reported as the total number of index crimes, summing up categories 1-7 and reporting that as “crime.”[[19]](#footnote-130) These index crimes are often divided into violent index crimes - murder, rape, robbery, and aggravated assault - and property index crimes - burglary, theft, motor vehicle theft. For more on index crimes, and the drawbacks of using them, please see Section 3.1.2.

### 3.1.1 Hierarchy Rule

This dataset uses what is called the Hierarchy Rule where only the most serious crime in an incident is reported (except for motor vehicle theft and arson, which are always included).[[20]](#footnote-131) For example if there is an incident where the victim is robbed and then murdered, only the murder is counted as it is considered more serious than the robbery. That the data uses the Hierarchy Rule is a frequently cited (by academics, reporters, random people on Twitter) criticism of the data that is, in my opinion, overblown.

In practice, the Hierarchy Rule has only modest effects on the data, undercounting few crimes. Though the Hierarchy Rule does mean this data is an undercount, data from other sources indicate that it is not much of an under count. NIBRS data contains every crime that occurs in an incident (i.e. it does not use the Hierarchy Rule). Using this we can measure how many crimes the Hierarchy Rule excludes. In approximately 90% of incidents, only one crime is committed.[[21]](#footnote-132) Additionally, when people talk about “crime” they usually mean murder which, while incomplete to discuss crime, means the UCR data here is accurate on that measure.

The FBI released a report [available here](https://ucr.fbi.gov/nibrs/2014/resource-pages/effects_of_nibrs_on_crime_statistics_final.pdf) in 2015 that directly examined this issue by taking NIBRS data from 2014 and examined how NIBRS data (which does not use the Hierarchy Rule) compares to using the Hierarchy Rule and keeping only the most serious crime. Figure 3.1 is a screenshot from their report showing the percent increases in crimes when including all crimes in an incident relative to following the Hierarchy Rule. They find that 10.6% of incidents have multiple crimes occurring. For violent crime, murder and rape have no change; for the remaining violent crimes - robbery and aggravated assault - crimes increased by 0.6%.[[22]](#footnote-134) Burglary increased by 1% and the largest increases came from theft and motor vehicle theft, increasing by 2.6% and 2.7%, respectively. Curiously, motor vehicle theft increased even though the FBI’s documentation for this data says that it is exempt from the Hierarchy Rule and should always be reported. This discrepancy suggests either non-compliance or errors in the FBI’s manual.

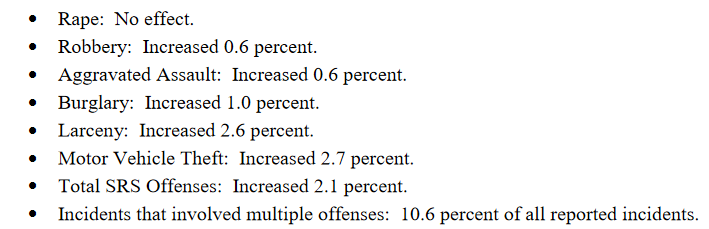


Figure 3.1: The FBI’s findings of how crime reporting changes when using the Hierarchy Rule using NIBRS 2014 data.

In Table ?? I replicate the FBI’s table using NIBRS 2022 data. Results are fairly close. Homicide and rape and unchanged; robbery and aggravated assault both increase by <1%; my count for theft and burglary are a bit smaller, and motor vehicle theft is almost triple the FBI’s 2014 number. But these numbers are mostly consistent - particularly so for violent crime - and I expect the differences are just that 2014 and 2022 data are different. So using the Hierarchy Rule does undercount crime, but this is a small undercounting and is primarily led by property crime. Violent crime is only slightly undercounted. And compared to the number of crimes not counted because the victim does not reports it to the police, this is a very small share of crimes.

### 3.1.2 Index (“Part 1”) crimes

One of the first things that people tend to learn about SRS crime data is that it covers something called an “index crime.”[[23]](#footnote-140) Index crimes, sometimes written as Part 1 or Part I crimes, are the seven crimes originally chosen by the FBI to be included in their measure of crimes as these offenses were both considered serious and generally well-reported so would be a useful measure of crime. Index crimes are often broken down into property index crimes - burglary, theft, and motor vehicle theft (and arson now, though that is often not included and is reported less often by agencies) - and violent index crimes (murder, rape, robbery, and aggravated assault). The “index” is simply that all of the crimes are summed up into a total count of crimes (violent, property, or total) for that police agency.

The biggest problem with index crimes is that it is simply the sum of 8 (or 7 since arson data usually isn’t included) crimes. Index crimes have a huge range in their seriousness - it includes, for example, both murder and theft - so summing up the crimes gives each crime equal weight. This approach is flawed as 100 murders is more serious than 100 thefts. This is especially a problem as less serious crimes (theft mostly) are far more common than more serious crimes. In 2017, for example, there were 1.25 million violent index crimes in the United States. That same year had 5.5 million thefts. So using index crimes as your measure of crimes fails to account for the seriousness of crimes. Looking at total index crimes is, in effect, mostly just looking at theft. Looking at violent index crimes alone mostly measures aggravated assault. This is especially a problem because it hides trends in violent crimes. As an example, San Francisco, shown in Figure 3.2, has had a huge increase - about 50% - in index crimes in the last several years. When looking closer, that increase is driven almost entirely by the near doubling of theft since 2011. During the same years, index violent crimes have stayed fairly steady. So the city isn’t getting more dangerous - at least not in terms of violent index crimes increasing - but it appears like it is due to just looking at total index crimes.

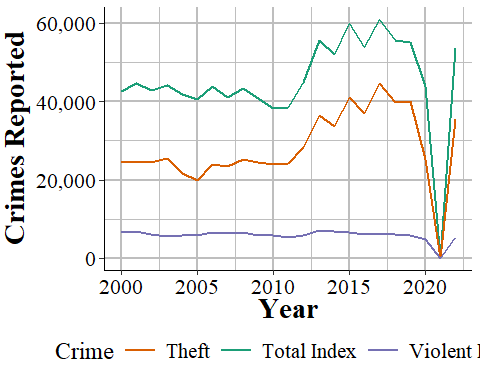


Figure 3.2: Thefts and total index crimes in San Francisco, 2000-2022.

Many researchers divide index crimes into violent and nonviolent categories, which helps but even this is not entirely sufficient. Take Chicago as an example. It is a city infamous for its large number of murders. But as a fraction of index crimes, Chicago has a rounding error worth of murders. Their 653 murders in 2017 is only 0.5% of total index crimes. For violent index crimes, murder made up 2.2% of crimes that year. As seen in Figure 3.3, in no year where data is available did murders account for more than 3.5% of violent index crimes; and, while murders are increasing as a percent of violent index crimes they still account for no more than 2.5% in most recent years. What this means is that changes in murder are very difficult to detect. If Chicago had no murders this year, but a less serious crime (such as theft) increased slightly, we could not tell from looking at the number of index crimes, even from violent index crimes. As discussed in the below section, using this sample of crimes as the primary measure of crimes - and particularly of violent crimes - is also misleading as it excludes important - and highly common relative to index crimes - offenses, such as simple assault.

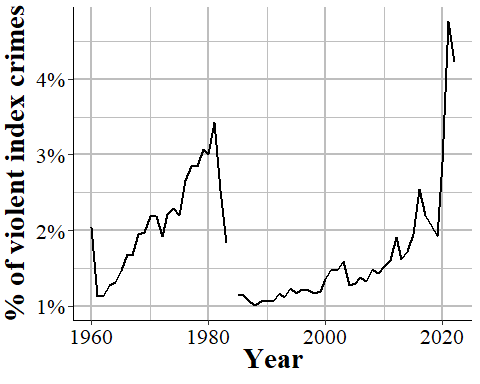


Figure 3.3: Murders in Chicago as a percent of violent index crimes, 1960-2022.

#### 3.1.2.1 What is a violent crime?

An important consideration in using this data is defining what a “violent crime” is. One definition, and the one that I think makes the most sense, is that a violent crime is one that uses force or the threat of force. For example, if I punch you in the face, that is a violent crime. If I stab you, that is a violent crime. While clearly different in terms of severity, both incidents used force so I believe would be classified as a violent crime. The FBI, and most researchers, reporters, and advocates would disagree. Organizations ranging from the [FBI itself](https://SRS.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/violent-crime) to [Pew Research Center](https://www.pewresearch.org/fact-tank/2020/11/20/facts-about-crime-in-the-u-s/) and advocacy groups like the [Vera Institute of Justice](https://arresttrends.vera.org/data-sources-methodology) and the [ACLU](https://www.aclu.org/report/tale-two-countries-racially-targeted-arrests-era-marijuana-reform) all define the first example as a non-violent crime and the second as a violent crime. They do this for three main reasons.

The first reason is that simple assault is not an index crime, so they do not include it when measuring violent crime. It is almost a tautology in criminology that you use index crimes as the measure of crime since that is just what people do. As far as I am aware, this is really the main reason why researchers justify using index crimes: because people in the past used it so that is just what is done now.[[24]](#footnote-153) This strikes me as a particularly awful way of doing anything, more so since simple assault data has been available almost as long as index crime data.[[25]](#footnote-154)

The second reason - and one that I think makes sense for reporters and advocates who are less familiar with the data, but should be unacceptable to researchers - is that people do not know that simple assault is included, or at least do not have easy access to it. Neither the FBI’s annual report [page](https://SRS.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/home) not their official [crime data tool website](https://crime-data-explorer.fr.cloud.gov/) include simple assault since they only include index crimes. For people who rely only on these sources - and given that using the data itself requires at least some programming skills, I think this accounts for most users and certainly nearly all non-researchers - it is not possible to access simple assault crime data (arrest data does include simple assault on these sites) from these official sources.

The final reason is that it benefits some people’s goals to classify violent crime as only including index crimes. This is because simple assault is extremely common compared to violent index crimes - in many cities simple assault is more common than all violent index crimes put together - so excluding simple assault makes it seem like fewer arrests are violent than they are when including simple assault. For example, a number of articles have noted that marijuana arrests are more common than violent crime arrests ([Ingraham 2016](#ref-ingraham2016); [Kertscher 2019](#ref-kertscher2019); [Devito 2020](#ref-devito2020); [Earlenbaugh 2020](#ref-earlenbaugh2020); [Edwards et al. 2020](#ref-aclu2020)) or that violent crime arrests are only 5% of all arrests ([Neusteter and O’Toole 2019](#ref-neusteter2019every); [Speri 2019](#ref-speri2019)).[[26]](#footnote-156) While true when considering only violent index crimes, including simple assault as a violent crime makes these statements false. Some organizations call the violent index crimes “serious violent crimes” which is an improvement but even this is a misnomer since simple assault can lead to more serious harm than aggravated assault. An assault becomes aggravated if using a weapon or there is the *potential* for serious harm, even if no harm actually occurs.[[27]](#footnote-157)

As an example of this last point, Figure 3.4 shows the number of violent index crimes and simple assaults each year from 1960 to 2018 in Houston, Texas (simple assault is not reported in SRS until 1964, which is why 1960-1963 show zero simple assaults). In every year where simple assault is reported, there are more simple assaults than aggravated assaults. Beginning in the late 1980s, there are also more simple assaults than total violent index crimes. Excluding simple assault from being a violent crime greatly underestimates violence in the country.

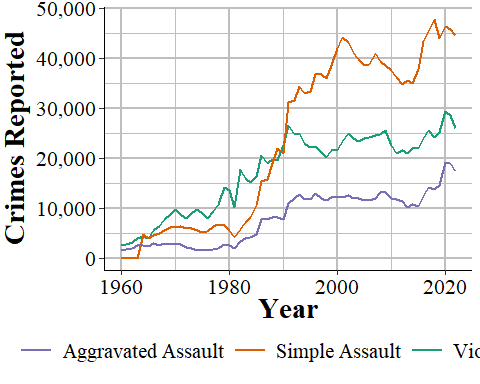


Figure 3.4: Reported crimes in Houston, Texas, from 1960 to 2018. Violent index crimes are aggravated assault, rape, robbery, and murder.

## 3.2 Agencies reporting

Figure 3.5 shows the annual number of police agencies that reported at least one month that year. With data starting in 1960, there were a little under 7,500 agencies reporting a year until 1970 when the number of agencies began increasing. This continued until the late 1970s when about 14,000 agencies reported, and this remained steady for over a decade before declining in the 1990s until about 12,500 in the latter half of the decade. Then the number of agencies reporting increased steadily until about 16,500 agencies reported in 2010. The number of agencies has slowly increased since then, adding a few hundred agencies from 2010 to 2020. The drop and then partial recovery you see in 2021 and 2022 is from the FBI stopping collection of this data and then restarting that collection in 2021. Most agencies report to NIBRS but the only that still report SRS, and therefore could not report in 2021, caused the drop.

There are actually two lines throughout this entire figure, though they are nearly identical until 2018. That is because there are two ways of measuring how many months an agency reports data. The primary one - and the one the FBI itself uses - is through a variable in the data called the “last\_month\_reported.” This is, as it sounds, the last month the agency sent data in. So if an agency reports data in December the variable will have “December” as the last month. If that agency only reported it December the variable will still say “December.” Most people use this as the number of months that the agency reported. So a December value is 12 months reported, even though in our example it was the only month with data.

In the data there are 12 columns - one for each month - that says whether the agency reported data in that month. That is what I use in the green line to measure how many months of data that agency reported. I refer to this in the figure and in the data I have released as the “number of months missing.” When looking at agencies reporting only a single month the lines are nearly identical, though the last month reported measure is nearly always larger. This changes in 2018 as a result of the data changing, meaning I needed to use different columns to check starting in that year. That means that post-2018 data may not be comparable to 2018 and earlier using this variable.

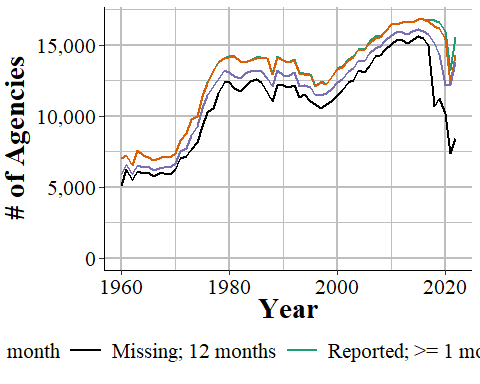


Figure 3.5: The annual number of agencies reporting at least one month of data and a full 12 months of data, 1960-2022.

Usually when you are looking at crime data you want annual data, so having agencies report a full year’s of data is more important than them submitting just partial data. This is especially important when comparing an agency over time or two different agencies to each other. Obviously, an agency with 6 month of data will have fewer crimes reported than one with 12 months reported, all else being equal. But this is something easy to overlook as it is an easy assumption that agencies will report a full years-worth of data. Unfortunately, this is always true. Some agencies do not report any data and others report only part of the year - though if an agency reports one month they usually do report all 12. Figure ?? repeats Figure 3.5 but now showing only agencies reporting 12 months of data, using both of our measures. Since 1960 every year has had fewer agencies reporting full data using the “number of months missing” method than the “last month reported” method.

## 3.3 Negative numbers

One of the most common questions people have about this data is why there are negative numbers, and if they are a mistake. Negative numbers are not a mistake. The SRS data is monthly so police agencies report the number of monthly crimes that are known to them - either reported to them or discovered by the police. In some cases the police will determine that a crime reported to them did not actually occur - which they called an “unfounded crime.” An example that the FBI gives for this is a person reports their wallet stolen (a theft) and then later finds it, so a crime was initially reported but no crime actually occurred.

How this works when the police input the data is that an unfounded crime is reported both as an unfounded crime and as a negative actual crime - the negative is used to zero out the erroneous report of the actual crime. So the report would look like 1 actual theft (the crime being reported), -1 theft (the crime being discovered as not have happened), and 1 unfounded theft. If both incidents occurred in the same month then this would simply be a single unfounded theft occurring, with no actual theft - literally a value of 1 + -1 = 0 thefts.

Negative values occur when the unfounding happens in a later month than the crime report. In the theft case, let us say the theft occurred in January and the discovery of the wallet happens in August. Assuming no other crimes occurred, January would have 1 theft, and August would have -1 thefts and 1 unfounded theft. There is no way of determining in which month (or even which year) an unfounded crime was initially reported in. When averaging over the long term, there should not be any negative numbers as the actual and unfounded reports will cancel themselves out. However, when looking at monthly crimes - particularly for rare crimes - you will still see negative numbers for this reason. Since crimes can be unfounded for reports in previous years, you can actually see entire year’s crime counts be negative, though this is much rarer than monthly values.[[28]](#footnote-170)

So using the far more common last month reported method will overestimate how many agencies report a full year of data. In practice, though, this affects very little data. At least when aggregating nationally - which I generally advise against since local crime matters a lot more than national averages. Still, let us look the increase in the total number of crimes reported by 12-month reporting agencies increases from the “number of months missing” measure to the “last month reported” measure, shown in Figure 3.6. The answer is an extremely small increase, averaging (mean = 0.93, mean = 1.55) about a one percent point increase each year. The differences in measuring how many months are reported can matter a great deal at the agency-level, but does very little when aggregating nationally.

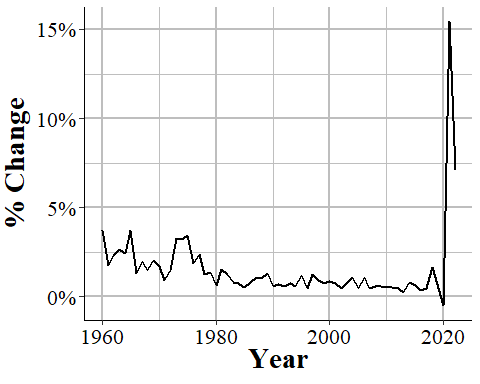


Figure 3.6: The percent change in the number of murders reported in the United States each year when moving from the number of months missing measure to the last month reported measure, 1960-2022.

## 3.4 Important variables

Now let us talk about some important variables in this data such as what it actually measures. For each crime we have four different categories indicating the number of crimes actually committed, the number cleared, and the number determined to not have occurred. Like other UCR data, there are also variables that provide information about the agency - ORI codes, population under jurisdiction - the month and year that the data covers, and how many months reported data.

### 3.4.1 Actual crimes

This is the number of offenses that *actually* occurred - where *actually* means that a police investigation found that the crime report was accurate. Occurred is actually a bit of a strong word. These are crimes that come to the police’s attention in that month, even if the crime actually occurred before. For example, if someone walks into a police department on February, 2025 and says “last year I killed five people” that would count as five murders in February, 2025 (that is assuming that the police believe that the person is telling the truth). It would not be counted in 2024 when the person says the murders took place.

Crimes that are reported that the police find did not occur (e.g. report of an arson but turns that the fire began accidentally) are called “unfounded” crimes. So this variable is the one people use to measure *crime*. For example if 10 people are murdered in a city the number of “actual murders” would be 10. A crime is a crime incident, regardless of how many offenders there were. If there are multiple victims in a case, such as a double murder, then it would count as multiple crimes.

Figure 3.7, for example, shows the number of murders per 100,000 population in Newark, NJ, for 1960-2022. One things stands out. Or does not stand out, in a bad way. Newark Police did not report a full year of data in 2015; they reported only 11 months. It is imperceptible in the figure but if you look at the number of months reported in that year - using either the last month reported or the number of months missing measure - you can see that December is missing. While visualizing the data is often a good way to look for outliers or missingness, it is not enough alone. You need to look at the raw data as well to be safe.

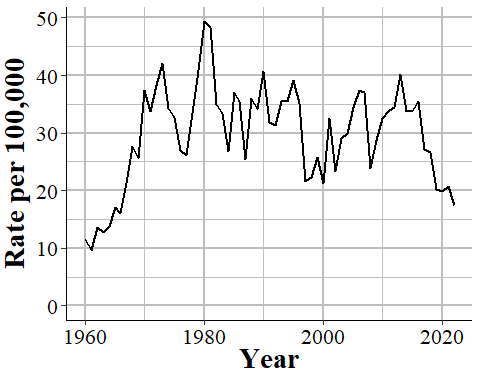


Figure 3.7: The annual murder rate per 100,000 people in Newark, NJ, 1960-2022.

Since this is the number of crimes reported and found to occurred, it undercounts the total number of reported crimes. To get that number you will need to add actual crimes to unfounded crimes, which we discuss in Section @ref({unfounded}). However, unfounded crimes are increasingly not being reported as agencies move to NIBRS reporting which does not capture unfounded crimes.

### 3.4.2 Total cleared crimes

A crime is cleared when an offender is arrested or when the case is considered cleared by exceptional means. To be more specific, this data counts crime as a crime incident, regardless of the number of offenders. For example, if three people robbed a person, that is one crime of robbery, not three separate crimes. This crime is cleared when one of the three robbers is arrested - no matter the outcome for the other two robbers. Arresting all three still counts as a single robbery cleared. If those three people had robbed two people that would be reported as two robberies. The first year with clearance data is in 1963, though that is extremely rare; the vast majority of agencies started reporting this data in 1964.

Even though this dataset is formally named “Offenses Known and Clearances by Arrest” it does include clearances where no one is arrested, which are called “exceptional clearances” or “clearances by exceptional means.” For exceptional clearances, police must have identified the offender, have sufficient evidence to arrest that offender, know where they are (so they can actually be apprehended) and only then be unable to make the arrest. Exceptional clearances include cases where the offender dies before an arrest (by any means, including suicide, accidental, illness, killed by someone including a police officer) or when the police are unable to arrest the person since they are already in custody by another jurisdiction (including out of the country or in custody of another agency) and extradition is denied. Two other potential causes of exceptional clearance are when prosecution of the case cannot go forward because the district attorney refuses to prosecute the case, for reasons other than lack of evidence, or when a victim refuses to assist the prosecution in the case.

Unfortunately, this data does not differentiate between clearances by arrest or by exceptional means so there is no way to determine how many clearances mean the offender is actually arrested - and even more problematic, how many clearances have all of the offenders arrested. NIBRS data does differentiate these types of clearances, another advantage of using it. There is also some evidence that at least some police agencies report classify large numbers of clearances as clearances through exception means (again, we have no way to tell which is which using this data) even though exceptional clearances should be rare given that times where the offender is known but cannot be arrested are likely far less common than when they are arrested. For an investigation into this issue, see the Yeung et al. ([2018](#ref-yeung2018comes)) article available on ProPublica’s website [here](https://www.propublica.org/article/when-it-comes-to-rape-just-because-a-case-is-cleared-does-not-mean-solved).

Clearances are reported in the month that they occur, regardless of when the crime they are clearing occurred. In practice, however, most crimes are cleared in the month that they occur. According to the 2019 NIBRS, it takes on average 7 days between the incident and the arrest (median = 0 days) date when averaging across all crimes - for individual crimes these values will be different. This means that most of the clearances will be for the same month as the initial crime - though this will be less common for crimes that happen at the end of a month. Of course, police agencies can solve older cases - and even target cold cases to be solved - so this is still a degree of uncertainty for which month these clearances are for.

This means that there are occasionally months - and even years - where there are more reported crimes cleared than crimes that occur.[[29]](#footnote-182) Figure 3.8, for example, shows the number of actual and cleared murders from the Los Angeles Police Department. In 2013 there were more murders cleared (271) than actual murders (251) In 2020, both values are zero as the LAPD did not report data that year.

This is actually a good check to see when people who use this data do not actually understand how it works. I have seen published academic papers that say that having more clearances than actual crimes is a data error; clearly they declined to read the official manual (or this book) before they, their editor, and their anonymous reviewers published the paper.

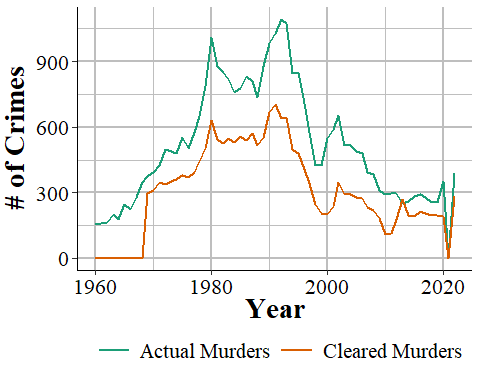


Figure 3.8: The annual number of actual and cleared murders from the Los Angeles Police Department, 1960-2022.

### 3.4.3 Crimes cleared where all offenders are under 18 years old

This variable is a subset of the Total Cleared variable and only includes clearances for offenses in which **every** offender is younger than age 18. Since this requires that the police know, or at least believe, the age of every offender, it is probably fairly inaccurate. This category includes cases where the juvenile is given a citation to show up in court for their trial and is not formally arrested and taken into custody.

### 3.4.4 Unfounded crimes

An unfounded crime is one in which a police investigation has determined that the reported crime did not actually happen. The first year of data that included unfounded crimes was 1979, though most agencies began reporting in 1983.

For example, I observed during a ride-along a report of a burglary where the homeowners said that they came home, and the front door was open and they thought it might have been their son who forgot to close it but were worried that it could be a burglar, so they called the police just in case. This would be recorded as a burglary and if it turned out to be the son, the police would then record this as an unfounded burglary.

Other unfounded crimes would include when someone reports a crime but later says that the report was not true. For example, a person could report a burglary to the police to collect insurance money on the items they claim was stolen. If the police discover this they would unfound the case - and the lying to the police and fraud would not be counted as neither of those are crimes included in this dataset. The way that the police do this in the data is to record an unfounded crime as a negative actual crime. If there are 10 burglaries already reported and then the above example occurred they would take those 10, add 1 for the report, and then subtract one for when they decide the crime is unfounded. This evens out the data to “erase” the initial report. If the unfounding occurs in a different month than the reported crime then this could lead to negative crimes being reported. This is just another quirk of SRS data and is another good check to see if a person using the data actually understands it as I have also seen people say negative crimes is a data error.

Figure 3.9 provides one example of this by showing the number of burglaries that the Frankenmuth Police, MI, say actually occurred from 1960-2022. In 1977 they reported -1 burglaries, the result of having more cleared than actual burglaries in that year.

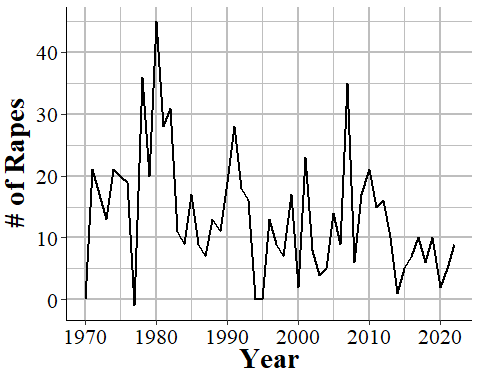


Figure 3.9: The number of actual burglaries reported by the Frankenmuth Police Department, MI, 1960-2022.

While this is a useful variable, it is not captured in NIBRS data. Instead the number of unfounded crimes is always reported as zero. For example, Figures 3.10 and 3.11 show the annual number of unfounded crimes (of any crime type) in Memphis, TN, and Denver, CO, which are two of the earliest large agencies to adopt NIBRS. Memphis started in reporting to NIBRS in 2000 and Denver did so in 2005. These agencies stopped reporting any unfounded crimes either in that or the following year.[[30]](#footnote-193)

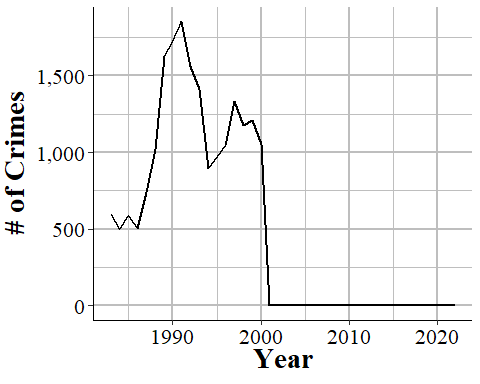


Figure 3.10: The annual number of unfounded crimes in Memphis, TN, 1983-2022.

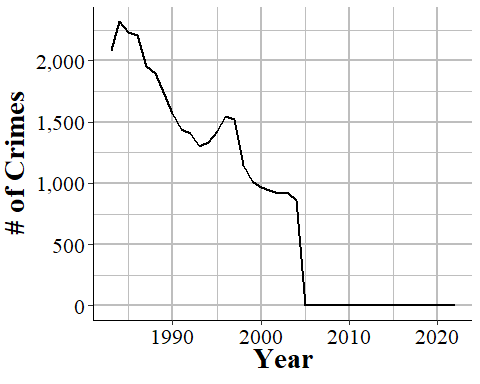


Figure 3.11: The annual number of unfounded crimes in Denver, CO, 1983-2022.

## 3.5 Important changes

There are two major changes in recording practices over the life of this dataset: an expansion of what counts as rape, and a reduction in what counts as manslaughter.

### 3.5.1 Rape definition change

The FBI changed the definition of rape for UCR data starting in 2013 to a broader definition than the older definition, which is commonly called the “legacy definition” or “legacy” or “historical” rape. The legacy definition is “the carnal knowledge of a female **forcibly** and against her will” (emphasis added). This means that only rape is only included in UCR data when it is a female (of any age, there is no differentiation for child victims) forcibly vaginally penetrated by a penis. This is a narrow definition and excludes a number of sexual acts that people may consider rape such as forced oral or anal sex, and cases with a male victim.

Starting in 2013, rape has a new, broader definition in the UCR to include oral and anal penetration (by a body part or object) and to allow men to be victims. The new definition is: “Penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim.” The previous definition included only forcible intercourse against a woman. This definition is far broader and is effectively any non-consensual sexual act. It also includes male victims though the data does not differentiate between male or female (or any other gender) victims.

Both the current and legacy definitions exclude statutory rape and incest other than forcible incest.[[31]](#footnote-204) They both also include lack of consent as cases where the victim cannot give consent, such as if they are too young or are mentally or physically incapacitated - the FBI specifically give the example of being temporarily incapacitated through drugs or alcohol.

As this revised definition is broader than the original one post-2013, rape data is not comparable to pre-2013 data. 2013, however, is simply the year that the FBI required that agencies report using the new definition. As might not be too surprising, not all agencies followed this requirement. We will look at four examples to show when there is clear evidence that the agency did change their definition in 2013, when it is clear they did so a year later, when it is unclear exactly when they made the change, and when the agency seems to not follow the change at all.

We will start with the Philadelphia Police Department shown in Figure 3.12. It is declining slowly but steadily over the 2000-2012 time period until spiking sharply in 2013. Since the rape definition change in 2013 is far broader than previous year’s definition, this makes sense. A broader definition should lead to a sudden increase in reported rapes if the agency is reporting correctly.

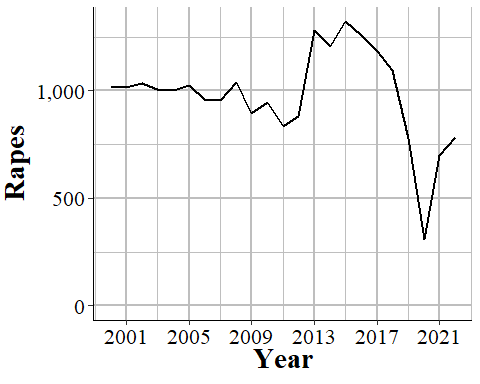


Figure 3.12: The annual number of rapes reported in Philadelphia, Pennsylvania, 2000-2022.

In comparison, New York City has the sudden spike a year later, which indicates that they did not start using the new definition until 2014. Figure 3.13 shows that rape is fairly steady, though increasing, in the years leading up to 2013 and has almost no change from 2012 to 2013, but a huge increase in 2014 and then steadily increases from there, spiking again in 2018. This seems like a fairly clear indicator that NYC simply did not follow the new definition until 2014.

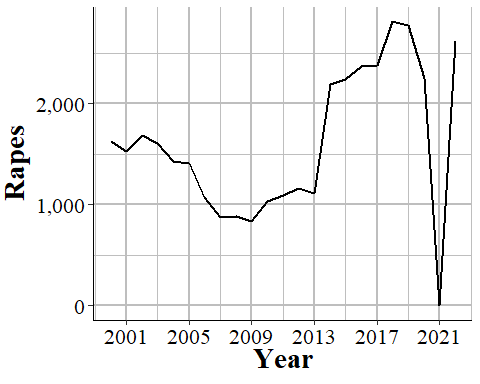


Figure 3.13: The annual number of rapes reported in New York City, 2000-2022.

Less clear is what is happening in San Francisco, shown in Figure 3.14. Here we do see an increase in 2013 which while it appears small on the graph is actually a 49% increase from 2012. Then there is a much larger spike in 2014 - a 120% increase - which may suggest that part of the agency started following the new definition in 2013 and the remainder followed in 2014. Or maybe some months used the old definition and others the new definition in 2013, while all of 2014 used the new definition However, increases or decreases are relatively common in San Francisco so it could also be that the agency only switched to the new definition in 2014 and the spike in 2013 is just a coincidence.

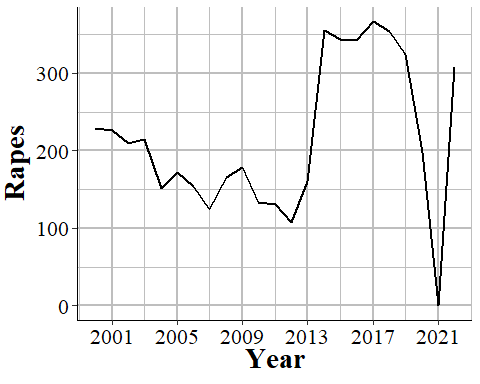


Figure 3.14: The annual number of rapes reported in San Francisco, California, 2000-2022.

Finally, we will look at Jackson Police Department in Mississippi where the definition change seems to have had no effect. As seen in Figure 3.15, reported rapes start to undulate in 2010 with 2013 data perfectly in line with the before and after trends - no sign that there is a change in reporting. This suggests that Jackson simply did not follow the definition change and continues to report using the old definition.

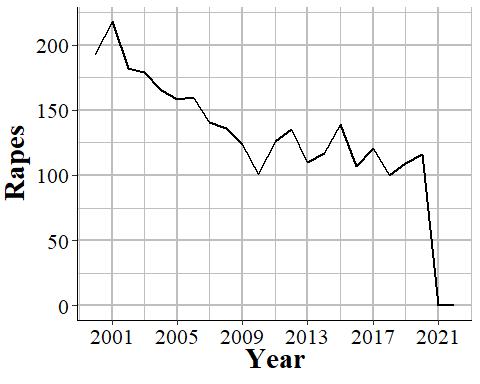


Figure 3.15: The annual number of rapes reported in Jackson, Mississippi, 2000-2022.

My takeaway from this is that rape should not be used at all for years after 2012. While the definition change makes pre-2013 and 2013+ years non-comparable, the differences in agency responses to this change - i.e. if they follow the rules or not - is such a mess that the data is too flawed to use.

### 3.5.2 The decline of manslaughter

This data contains two different crime subcategories for homicide: “murder and non-negligent manslaughter,” and “manslaughter by negligence.” The first is our measure of murder, and it includes everything we traditionally think of when it comes to murder - shootings, stabbings, strangulation, basically any intentional killing of another person.[[32]](#footnote-222) Suicides, killing a fetus, and accidental killings (e.g. car crashes) are not considered murders.[[33]](#footnote-223) The second, manslaughter by negligence - usually called just “manslaughter” - is when someone kills another person through “gross negligence” but does not kill them intentionally. This can include accidental killings when the death was caused by gross negligence. The FBI provide examples of this as kids playing with guns and shooting each other (and not knowing the gun was loaded) or a hunter accidentally shooting someone while hunting. After the late 1970s this excludes deaths from traffic accidents caused by negligence, such as hit and runs or DUIs. However, prior to this, these deaths were included, which is the cause of the very high number of manslaughters in the 1960s and 1970s.

Figure 3.16 shows the annual number of murders, manslaughters, and the sum of the two nationwide from 1960-2022. This just sums up the total reported counts from every agency each year so part of the increase is simply due to more agencies reporting as the year gets closer to the present day - so please pay attention to the diverging paths of each crime, not the trend for the individual crime over time. Murder is always more common than manslaughter, but these values are not that far apart in the early decade of data and manslaughter does not become rare until the end of the 1970s.

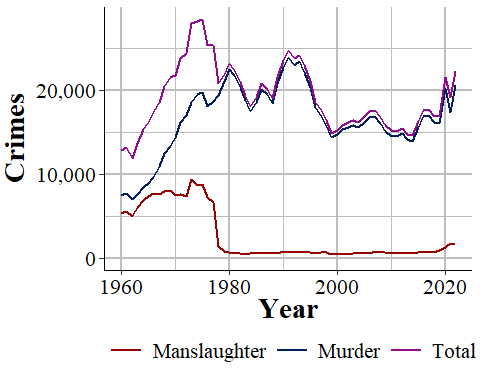


Figure 3.16: The annual number of murder and non-negligent manslaughter, manslaughter by negligence, and the sum of the two, nationwide from 1960-2022.

Figure 3.17 shows another way to look at this data: manslaughter as a percent of reported murder. In the early years of our data manslaughter was fairly common, with about 70-80% as many manslaughters reported as murders. This declined sharply in the mid-1960s until there were around 45% as many manslaughters as murders in the mid-1970s. Again, this declined until it was about 4% in 1980, and it has remained around there ever since. As police behavior could reduce traffic fatalities - and arrests for DUIs and traffic tickets are designed to improve public safety - it is unfortunate the we no longer have data on traffic deaths.

Manslaughter increased to over 1,000 for the first time since 1978 in 2020, increased against to over 1,700 in 2021 and continued at around that number in 2022. This is possibly related to the increase in murders over the last few years of available data. Unfortunately, this dataset does not allow us to do almost anything at figuring out more information than monthly or annual counts. NIBRS, in comparison, allows us to do this kind of deep dive, and for curious readers NIBRS also has manslaughter so you can investigate this question yourself.

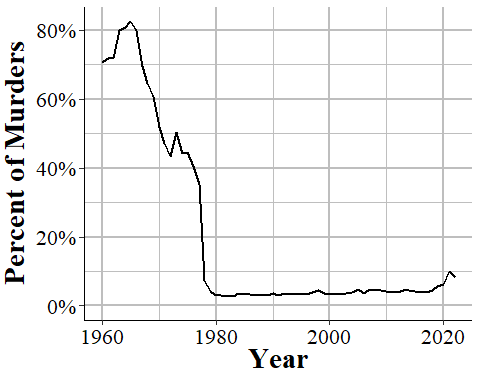


Figure 3.17: Reported manslaughter by negligence as a percent of reported murder and non-negligent manslaughter, nationwide 1960-2022.

Bennice, Jennifer A, and Patricia A Resick. 2003. “Marital Rape: History, Research, and Practice.” *Trauma, Violence, & Abuse* 4 (3): 228–46.

Devito, Lee. 2020. “Marijuana Arrests on the Decline but Still Outnumber Violent Crime Arrests, According to FBI Data.” City Beat; https://www.citybeat.com/news/blog/21145113/marijuana-arrests-on-the-decline-but-still-outnumber-violent-crime-arrests-according-to-fbi-data.

Earlenbaugh, Emily. 2020. “More People Were Arrested for Cannabis Last Year Than for All Violent Crimes Put Together, According to FBI Data.” Forbes; https://www.forbes.com/sites/emilyearlenbaugh/2020/10/06/more-people-were-arrested-for-cannabis-last-year-than-for-all-violent-crimes-put-together-according-to-fbi-data/?sh=7dd59ad6122f.

Edwards, Ezekiel, Emily Greytak, Brooke Madubuonwu, Thania Sanchez, Sophie Beiers, Charlotte Resing, Paige Fernandez, and Galai Sagiv. 2020. “A Tale of Two Countries: Racially Targeted Arrests in the Era of Marijuana Reform.” American Civil Liberties Union; https://www.aclu.org/report/tale-two-countries-racially-targeted-arrests-era-marijuana-reform.

Ingraham, Christopher. 2016. “Police Arrest More People for Marijuana Use Than for All Violent Crimes — Combined.” The Washington Post; https://www.washingtonpost.com/news/wonk/wp/2016/10/12/police-arrest-more-people-for-marijuana-use-than-for-all-violent-crimes-combined/.

Kertscher, Tom. 2019. “More Arrests in the US for Marijuana Possession Than Violent Crimes, 2020 Hopeful Cory Booker Says.” Politifact; https://www.politifact.com/factchecks/2019/jul/11/cory-booker/more-arrests-us-marijuana-possession-violent-crime/.

McMahon-Howard, Jennifer, Jody Clay-Warner, and Linda Renzulli. 2009. “Criminalizing Spousal Rape: The Diffusion of Legal Reforms.” *Sociological Perspectives* 52 (4): 505–31.

Neusteter, S Rebecca, and Megan O’Toole. 2019. “Every Three Seconds: Unlocking Police Data on Arrests.” Very Institute of Justice; https://www.vera.org/publications/arrest-trends-every-three-seconds-landing/arrest-trends-every-three-seconds/overview.

Speri, Alice. 2019. “Police Make More Than 10 Million Arrests a Year, but That Doesn’t Mean They’re Solving Crimes.” The Intercept; https://theintercept.com/2019/01/31/arrests-policing-vera-institute-of-justice/.

Yeung, B, M Greenblatt, M Fahey, and E Harris. 2018. “When It Comes to Rape, Just Because a Case Is Cleared Does Not Mean It Is Solved.” *ProPublica*.

1. We will see many examples of when agencies do not follow the definitions, which really limits this data. [↑](#footnote-ref-20)
2. While the original UCR data first reported in 1929, there is only machine-readable data since 1960. [↑](#footnote-ref-21)
3. This is referred to as an “ORI”, “ORI code”, and “ORI number”, all of which mean the same thing. [↑](#footnote-ref-58)
4. The abbreviation for Nebraska is “NB” rather than the more commonly used “NE.” [↑](#footnote-ref-59)
5. For most fixed-width ASCII files there are also missing values where it will have placeholder value such as -8 and the setup file will instruct the software to convert that to NA. SRS data, however, does not have this and does not indicate when values are missing in this manner. [↑](#footnote-ref-66)
6. For those interested in reading in this type of data, please see my R package asciiSetupReader. [↑](#footnote-ref-67)
7. Even highly experienced programmers who are doing something like can make mistakes. For example, if you type out “2+2” 100 times - something extremely simple that anyone can do - how often will you mistype a character and get a wrong result? I would guess that at least once you would make a mistake. [↑](#footnote-ref-68)
8. With the exception of the arrest data and some value label changes in hate crimes and homicide data, the setup files remain consistent so a single file will work for all years for a given dataset. You do not need to make a setup file for each year. [↑](#footnote-ref-69)
9. For evidence of this, please see any of the openICPSR pages for my detail as they detail changes I have made in the data such as decisions on what level to aggregate to and mistakes that I made and later found and fixed. [↑](#footnote-ref-70)
10. Jurisdiction here refers to the boundaries of the local government, not any legal authority for where the officer can make arrests. For example, the Los Angeles Police Department’s jurisdiction in this case refers to crimes that happen inside the city or are otherwise investigated by the LAPD - and are not primarily investigated by another agency. [↑](#footnote-ref-72)
11. Some states do mandate that their agencies report, but this is not always followed. [↑](#footnote-ref-74)
12. Arson is also an index crime but was added after these initial seven were chosen and is not included in the crimes dataset (though is available separately) so is generally not included in studies that use index crimes. [↑](#footnote-ref-113)
13. There is also a human trafficking dataset though this is a newer dataset and rarely used so I will not cover it in this book. [↑](#footnote-ref-114)
14. This data is also sometimes called the “Police Employees” dataset. [↑](#footnote-ref-118)
15. It also includes the value of items stolen during rapes and murders, if anything was stolen. [↑](#footnote-ref-122)
16. based on the location of the incident - “bank”, “gas station”, etc. [↑](#footnote-ref-123)
17. Even some academic papers are little more than a difference between two or more cities over time and can be very good research even if the data is not that complex. [↑](#footnote-ref-128)
18. While some people capitalize “Index Crime,” I prefer the term in lowercase which is how I will write it. [↑](#footnote-ref-129)
19. I have encountered a shocking number of academic papers and researchers who seem to believe that the subcategories are not included in the data at all. [↑](#footnote-ref-130)
20. FBI documentation actually differs on whether motor vehicle theft is always reported with some documentation saying it is while others placing it in the hierarchy. According to their “Summary Reporting System (SRS) User Manual” Version 1.0 released 6/20/2013, “The offenses of justifiable homicide, motor vehicle theft, and arson are exceptions to the Hierarchy Rule.” [↑](#footnote-ref-131)
21. According to 2022 NIBRS, 88.15% of incidents have only a single offense. [↑](#footnote-ref-132)
22. Murder is not shown in this figure since murder is always reported so cannot change. [↑](#footnote-ref-134)
23. Index crimes are sometimes capitalized as “Index Crimes” though I have seen it written both ways. In this book I keep it lowercase as “index crimes.” [↑](#footnote-ref-140)
24. I have also seen the justification that aggravated assault is more serious since it uses a weapon, though as the SRS subcategory of aggravated assault without a weapon clearly shows, aggravated assault does not require use of a weapon. [↑](#footnote-ref-153)
25. Simple assault is first available in 1964. Index crime data is available since 1960. [↑](#footnote-ref-154)
26. The FBI’s report for arrests does include simple assault so the second reason people may not include it does not apply here. [↑](#footnote-ref-156)
27. SRS data provides no information about the harm caused to victims. The new FBI dataset NIBRS actually does have a variable that includes harm to the victim so if you are interested in measuring harm (an understudied topic in criminology), that is the dataset to use. [↑](#footnote-ref-157)
28. From 1960-2022, there were 39 agency-years with a negative count of murders. [↑](#footnote-ref-170)
29. In about 1% of agency-years since 1964, the year most agencies started reporting this data, there were more cleared murders than actual murders. [↑](#footnote-ref-182)
30. For Memphis, as agencies can report both SRS and NIBRS, that agency may have reported both in 2000 which is why we still see unfounded data that year. [↑](#footnote-ref-193)
31. Both of these are recorded in NIBRS. [↑](#footnote-ref-204)
32. Attempted murder is usually classified as an aggravated assault. [↑](#footnote-ref-222)
33. Even the intentional killing of a fetus is classified as an aggravated assault against the mother, not a murder of the fetus. [↑](#footnote-ref-223)