National Incident-Based Reporting System (NIBRS) Data: A Practitioner's Guide

Jacob Kaplan

2021-06-22

Contents

List of Tables

List of Figures

Chapter 1

Preface

The average American had, in 2019, about a 1 in 20,500 chance of being murdered, 1 in 1,223 chance of being robbed, and a 1 in 64 chance of having something they own stolen.[1] Getting these numbers is extremely simple. We take the number of crimes reported to the police and divide it by the number of people living in the United States that year. For example, there were about 16,000 murders in 2019 and 328 million people in the country - 16,000 / 328 million = ~1/20,500. You'll more commonly see this - in news articles, in political speeches, in research articles, on TV, etc. - reported as the rate per 100,000 people but that's just a matter of conversion, the numbers are the same. This is, however, totally wrong. It assumes - and let's now just pretend that there's no underreporting of crimes to the police so the reported number of crimes is the true number of crimes - that every single person has the exact same risk of victimization. We know this is wrong intuitively. There are the "bad parts of town" or people who "run with the wrong crowd." Research in criminology backs this up by finding that crime is generally concentrated among a small group of people and within a small geographic area (usually a very small number of streets or neighborhoods in a city). From surveys that ask if people have been victims of a crime we also know that victimization rates differ by age, race, gender, income, and city type. Indeed, think of a personal characteristic (e.g. risk tolerance, athleticism, frequency outdoors) and there will probably be large differences in the likelihood of being a victim within these groups.

So why do people so frequently talk about crime as rates per total population? Why assume that everyone has equal risk of being a victim? The major reason, I think, is that the main FBI dataset on crime, the Uniform Crime Reporting (UCR) System Data doesn't provide any information about crime victims other than for homicide and hate crime victims.[2] Victimization surveys are nearly all at state or national levels so trying to use their results

[1] This is according to data from the FBI's 2019 Crime in the United States report.

[2] There is a small amount of information available for property theft victims, but not enough to do these kinds of victimization rate calculations

for specific city-level rates will have very high levels of imprecision, likely enough to make the results unhelpful. So by necessity people tend to assume equal risk since there's not good data to do so otherwise. Doing so, however, is a large mistake in my opinion. It both overstates and understates risk, often very drastically. For example, many murders are concentrated among young minority men in impoverished areas of urban cities. Their risk of being murdered is many times higher than you'd expect from simply dividing murders by city population overall. And the risk for other people in the city is far smaller than you'd expect from the naive risk value.

While UCR data is the predominant dataset in criminology and the one that truly guides America's perceptions of crime, there is actually one that is far more detailed and allows us do a much better (though far from perfect) measure of crime victimization by a number of victim traits. This dataset, called the National Incident-Based Reporting System - and often called by its abbreviation NIBRS - provides demographics information about each victim of crime. Like the UCR this is a dataset from the FBI which standardizes information from agencies in the United States so it is comparable across agencies and over time (with some caveats). This data includes the age, race, gender, and the relationship of the victim to the offender for every known victim. We also have info about the offense such as which crimes occurred (and unlike the UCR, incidents can have multiple offenses), where and when it occurred, and what weapons (if any) were used. And we have demographic information about offenders, and whether they were arrested (including the type of arrest such as if the police had a warrant and arrested them or saw them committing the crime and arrested them at the scene). From this we can figure out victimization rates based on certain (though not all) victim characteristics.

Of course, the research opportunities from this highly detailed data extend far beyond mere victimization rates. Personally, I've used the data to examine topics ranging from marijuana decriminalization and domestic violence injuries to daylight saving time's effect on crime clearance. As of this writing in early summer 2021, Google Scholar has over 1,500 results for NIBRS research since 2017. As the FBI has said that starting in 2021 NIBRS data will completely replace UCR data, this is *the* crime dataset of the future (though I believe that UCR will still be heavily used for many years to come). For graduate students and early career researchers, this is likely the dataset that you'll use for your entire career.

1.1 Goal of the book

By the end of the book you should have a firm grasp on NIBRS data and how to use it (or as is occasionally the case, choose not to use it) properly. However, this book can't 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.[3] As the FBI has moved to only use NIBRS data starting in 2021, I expect the uses of this data - and thus the number of emails I get - to grow very quickly. This is an increasingly popular dataset used by criminologists (and by other fields studying crime) and yet there are still occasions where papers are using the data incorrectly.[4] So hopefully this book will decrease the number of misconceptions about this data, increasing overall research quality.

Since manuals are boring, I'll try to include graphs and images to try to alleviate the boredom. That said, I don't think it's possible to make it too fun so sorry in advanced. This book is a mix of facts about the data, such as how many years are available, and my opinions about it, such as whether it is reliable. In cases of facts I'll just say a statement - e.g. "NIBRS data began in 1991". In cases of opinion I'll temper the statement by saying something like "in my opinion…" or "I think".

1.2 Structure of the book

This book will be divided into eight chapters: this chapter, an intro chapter briefly summarizing each segment files (NIBRS data is broken into multiple "segments" which are basically just a set of variables covering a specific topic like victims or offenders. We'll discuss this more next chapter) and going over overall issues with NIBRS data, and seven chapters each covering one of the seven NIBRS segments. Each chapter will follow the same format: we'll start with a brief summary of the data such as its possible uses and pitfalls. And then, we'll cover the important variables included in the data and how to use them properly.

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 (2021). *National Incident-Based Reporting System (NIBRS) Data: A Practitioner's Guide*. https://nibrsbook.com/.

BibTeX format:

[3] Writing also keeps away the boredom.

[4] Though given that the data is fairly complicated and requires good programming knowledge, the bar is higher to use it. So there are far fewer bad uses of this data than there is for UCR data.

```
@Manual{ucrbook,
  title = {National Incident-Based Reporting System (NIBRS) Data: A Practitioner's Guide
  author = {{Jacob Kaplan}},
  year = {2021},
  url = {https://nibrsbook.com/},
}
```

1.4 Pronunciation

This data is usually just called NIBRS, and generally there's no distinction between segment files since they work in unison as they are pieces of the overall criminal incident (i.e. people say they use NIBRS data not that they use the Victim Segment of NIBRS). "NIBRS" is generally pronounced as "NIE-BERS". It rhymes with "HIGH-BERS". I've also heard it pronounced - usually by non-academics - using a soft i like in "timber" so it sounds like "nih-bers". I prefer the "NIE-BERS" saying but it really doesn't make a difference.

1.5 Sources of NIBRS data

There are a few different sources of NIBRS data available today. First, and probably most commonly used, is the data put together by the National Archive of Criminal Justice Data (NACJD)). 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 - NIBRS or other crime data - I highly recommend you reach out to them for answers. They are extremely helpful and can likely quickly answer any question that you have. They have a collection of data and excellent documentation available for UCR data available on their site here. They've also put together what they call "Extract Files" which are where they merged some of the NIBRS segments together, saving you the effort of doing so yourself. These extract files essentially take every potential unit of analysis - incident, victim, offender, and arrestee (some crimes have no victims, only arrestees) - and merge it with the segment which has info about the incident such as the time of day or the outcome, and information about the reporting agency. NACJD only has data through 2016 which means that the most recent years (NIBRS 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. On this site you can chose an agency and see annual crime data (NIBRS data is at the hourly-level so this is very aggregated data) for certain crimes (and not even all the crimes actually available in the data). This site has only a small subset of the available data and is already aggregated so you're dealing with a subset of data in a unit of analysis that you may not want. For example, this site lets you see the annual age of offenders for certain crimes in age brackets such as aged 20-29. NIBRS data provides the exact age (in years) of each offender so this website is much less useful than the full data. The crimes on this site are also limited to only the eight "Index Crimes" (murder, rape, robbery, aggravated assault, arson, burglary, theft, and motor vehicle theft) so are only a tiny share of the crimes actually reported in NIBRS data. For more on what Index Crimes are, please see here. This data source is potentially 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.

Finally, I have my own collection of UCR data available freely and publicly on openICPSR, a site which allows people to submit their data for public access. For each of these datasets I've taken the raw data from the FBI and read it into R. Since the data is only available from the FBI as fixed-width ASCII files, I created a setup file (we'll explain exactly how reading in this kind of data works in the next chapter) and read the data into R and saved the files in R and Stata files for easy use. The main advantage is that all my data has standard variable names and column names and can be read into modern programming languages R and Stata (this is also true of recent NACJD years, but early years come as fixed-width ASCII files). The downside is that I don't provide documentation other than what's on the openICPSR page (and in this book) and only provide data in R and Stata format so people using languages such as SAS or SPSS cannot use this data.[5]

1.5.1 Where to find the data used in this book

The data I am using in this book is the cleaned and concatenated data that I put together from the raw data that the FBI releases. The raw data that the FBI releases is available here. The data that I have released is available on the data hosting site openICPSR here. I am hosting this book through GitHub which has a maximum file size allowed that is far smaller than these data (which are millions of rows long and can get quite large), so you'll need to go to openICPSR to download the data; it's not available through this book's GitHub repo.

1.6 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 and users understand

[5] I am not sure if SAS or SPSS can read in R or Stata data files.

it. That manual, called the "2019 National Incident-Based Reporting System 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've also posted it on my GitHub page in case the FBI ever takes down the PDF.[6] The link on my GitHub page is here. 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 don't follow the manual, will provide a solid foundation for your understanding of NIBRS data.

1.7 How to contribute to this book

If you have any questions, suggestions, or find any issues, please email me at jkkaplan6 [@] gmail.com. For more minor issues like typos or grammar mistakes, you can edit the book directly through its GitHub page. That'll 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'll 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've done and then submit the changes for me to review.

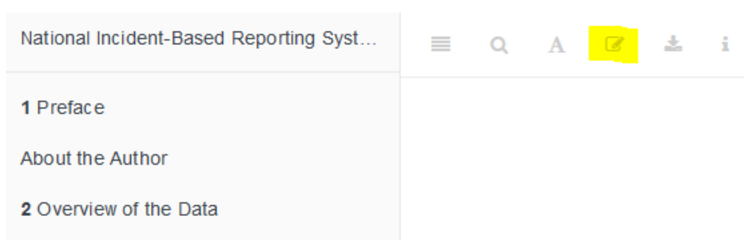

Figure 1.1: The edit button for how to make edits of this book.

[6] This is far more likely to happen as a result of standard government changing a site and forgetting to update the link rather than intentionally making the manual unavailable.

About the Author

Jacob Kaplan is the Chief Data Scientist at the Research on Policing Reform and Accountability RoPRA at Princeton University. He holds a PhD and a master's degree in criminology from the University of Pennsylvania and a bachelor's degree in criminal justice from California State University, Sacramento. His current research portfolio includes evaluating police policy and reforms, place-based crime prevention, measuring spatial crime concentration, and simulating how firing 'bad apples' affects police complaints and uses of force. In the past he's written on the effect of marijuana decriminalization on domestic violence, how increasing the number of police officers affects prison trends, how outdoor lighting affects crime and perception of safety, and public perceptions of forensic science techniques. He is the author of several R packages that make it easier to work with data, including fastDummies and asciiSetupReader. His website allows for easy visualization of crime-related data and he has released over a dozen crime data sets (primarily FBI UCR data) on openICPSR that he has compiled, cleaned, and made available to the public. He is also the author of a book on the FBI's Uniform Crime Reporting (UCR) Program data.

For a list of papers he has written (including working papers), please see here.

For a list of data sets he has cleaned, concatenated, and made public, please see here.

For a list of R packages he has created, please see here.

Chapter 2

Overview of the Data

Nearly a century ago the FBI started collecting data on crime that occurred in the United States as a way to better understand and respond to crime. This data, the Uniform Crime Reporting (UCR) Program Data, is a monthly count of the number of crime incidents (in cases where more than one crime happens per incident, only the most serious crime is included) in each police agency that reports data.[1] Other than for homicides (which provides info about each victim and offender), only the number of crimes that occurred is included. So we know, for example, the number of robberies in a city but nothing about who the victims or offenders were, when in that month (day or time of day) the robberies occurred, or the type of location where they happened. To address these limitations the FBI started a new dataset in 1991, the National Incident-Based Reporting System data - which is known by its abbreviation NIBRS - and is the topic of this book. Relative to the FBI's UCR data there are far fewer "weird things" in NIBRS data. Still, we'll cover instances of the "weirdness" in the data, such as the why crime always goes up on the 1st of the month, or why there are more crimes at noon than at nearly all other hours of the day. We'll also be discussing how much of the detailed information that should be available in the data is missing, and when that affects which questions we can answer.

NIBRS data provides detailed information on every crime reported to the police, including victim and offender demographics, whether the offender was arrested (and the type of arrest it was), what date and time of day (by hour only) it happened on, the victim-offender relationship, and the crime location (as a location type, not the exact address). It also covers a far wider range of crimes than UCR data did. With the exception of UCR data on assaults against police officers, all NIBRS data can be converted back to UCR data, making it fully backwards compatible and, therefore, comparable to UCR data. In many ways NIBRS data is a massive improvement over UCR data. This data allows for a deeper

[1] This data has been expanded since it began in 1929 to include information on arrests, hate crimes, and stolen property.

understanding of crime and it has led to an explosion of research that allows a far more detailed analysis of crime and crime-policies than the blunt UCR data.

However, there is a major limitation to this data: most agencies don't use it. According to the FBI only about 8,500 police agencies, covering about 45% of the US population, reported NIBRS data in 2019 (the latest year currently available). This is fewer than half of the about 18,000 police agencies in the United States. This is an even larger problem than it seems as the agencies that do report - especially in earlier years of the data - are disproportionately small and rural. So we're missing out of data from most major cities. A number of states don't have any agencies reporting, making this data relatively biased at least in terms of geography and city size. **Even so, the FBI has said that they are moving entirely to NIBRS data starting in 2021, and will no longer even collect UCR data.** While NIBRS can be converted to UCR data, meaning we can have consistent statistics over time, for agencies that don't report to NIBRS, we have no information on their crimes. In effect, unless the majority of agencies suddenly switch to NIBRS - which, given that the high level of detail relative to UCR data makes moving to NIBRS a costly and timely switch - we will be flying blind for most crime in the country.

2.1 Problems with NIBRS

There are three major problems with NIBRS data, with the first two related to the lack of reporting. First, we are potentially looking at a massive loss of data when UCR data ends in 2020 - it takes over a year for data to be released so even though I'm writing this in Spring 2021, 2019 UCR and NIBRS data are the latest years available. 2020 data won't be released by the FBI until September or October of this year. Considering the huge crime changes during 2020 - and the latest evidence suggests that the violent crime increase is continuing (and in places even accelerating) in 2021 - losing standardized crime data for most cities (and especially the largest cities) is a very bad thing. Moving the majority of agencies over to NIBRS so quickly may also risk the integrity of the data.[2] As they rush to comply with the FBI's order that they only will accept NIBRS data, there will likely be more mistakes made and erroneous data included in NIBRS data. This will likely include both knowledge problems with agencies not understanding how to properly report data and the simply issue of typos leading to wrong info being entered. Though the FBI does do quality assurance checks, no check is foolproof - and their checks in UCR data have still allowed clearly impossible data to be entered (e.g. millions of arsons reported in a month in

[2] "Quickly" is a bit is a misnomer as agencies were free to report to NIBRS since it began in 1991 and the FBI had announced many years ago that they'd only collect NIBRS in 2021. Still, given that the majority of agencies don't report to NIBRS and 2020 had a plague, the switch is likely to introduce issues and should be delayed.

a small town). So while I always urge caution when using any data - caution that should be accompanied by a thorough examination of your data before using it - NIBRS data from 2020 and beyond merits extra attention.

The second problem is that even if suddenly all agencies do start reporting in 2021, we'd only have a single year of data available. Even for agencies that already report, we generally don't have too many years of data for them. This really limits the kind of research since we can do since it's hard to know if a finding is based on a trend or is just a weird outlier without having many years of data available. For the agencies where 2020 is the first year, we'll likely to have to wait a few years to even figure out what "normal" crime is supposed to look like. This means that for the next several years at least we'll be mostly using NIBRS data as UCR-like datasets, aggregated to the month- or year-level so we can compare it with UCR data from the past. Luckily, this problem will be alleviated the longer we wait as more years of data will become available.

The final issue is that this data is massive. A single year of 2019 data - with <50% of agencies reporting, and few large agencies reporting - has about 6.5 million crime incidents recorded. Since each crime incident can have multiple victims, offenders, and crimes, there are more rows for these datasets.[3] Once all agencies report - though it's doubtful that'll ever occur, though we may come close - we're looking at tens of millions of rows per year. And even now if we wanted to look at a decade of data we're going to be dealing with over 50 million rows of data. So this data requires both good hardware - a strong laptop or server is necessary - and good programming skills, which most academics sorely lack. If you can, buy more RAM for your computer as that's much easier than having to write complicated code to deal with large data.[4] I want to stress this point. If you intend to work with NIBRS data for any significant amount of time you should buy the most RAM your computer can use (RAM is very cheap now) and install it. I'd recommend at least 16GB but more is better. While computers can handle NIBRS with less RAM, it'll just lead to you spending more time writing code to deal with big data and it'll inevitably still run slower than buying extra RAM.

[3] While people generally refer to NIBRS just as "NIBRS data" it is actually a collection multiple different datasets all - with a few exceptions - corresponding to a single crime incident. For example, if you care about victim info you'll look in the victim file called the "Victim Segment" (each of the datasets are called "Segments" since they are part of the whole picture of the crime incident) and likely will merge it with other data, such as when are where the crime occurred which is in the "Offense Segment". In most cases you'll merge together multiple datasets from the NIBRS collection to be able to answer the question that you have.

[4] "When in doubt, apply force" - Dean Knox.

2.1.1 NIBRS allows for different units of analysis

A major benefit of UCR data is that you have very limited choices. If you wanted to measure crime your only choice was to use their monthly aggregated police agency-level data. This makes working with the data relatively easy, even though what work you could do was limited. NIBRS data takes an opposite approach. It provides detailed data and largely leaves it up to the users for what to do with it. This flexibility is also a curse. For every use of this data you will need to decide which unit of analysis to use - and NIBRS provides a few options.

If you're interested in measuring rape you could do so in several different ways, each of which addresses a different part of crime measurement and will lead to different answers to your questions: the number of crime incidents, the number of victims, the number of offenders, and the number of crimes. Let's use an incident where four men rape a single woman as an example. Even if we somehow solve the issue of victims not reporting their rapes, we still have a few different ways of even measuring rape. First, we can follow the old UCR measure of incident-level and say that this is one rape since only one crime incident occurred (even though there were multiple offenders). Second, we could look at the victim-level, which again is one rape as there was only one victim. Or at the offender-level, which now has four rapes since each offender would be responsible the rape. Finally we could look at the offense-level. Even though the four men were involved in the rape incident, potentially not all of them would have actually committed the rape (and would have the offense in NIBRS data as something else such as assault or attempted rape if they didn't complete the act). Some could have acted as, for example, lookouts so would be involved with the incident but not the rape. So through this measure we'd have between one and four rapes, depending on the exact circumstances. Each way of measuring could lead to substantially different understandings of rape, and this is the kind of complexity that we'll have to wrangle with when using NIBRS data.

Since this data includes multiple crimes in each criminal incident, unlike the UCR which includes only the most serious crime per incident, we can also measure crime in its relationship to other crimes. In the above example we're interested in rapes. The UCR method would measure it as the number of rapes in incidents where rape is the most serious charge ("most serious" is based on the FBI's hierarchy of offenses, following what they call the Hierarchy Rule) but this undercounts crimes where rape happened alongside another, more serious, offense.[5] So we can also look at incidents where any offense that occurred was a rape. Using this method we can examine how often rape - or any crime we're interested in - co-occurs with other offenses, which provides more information on how crime happens that looking at one crime alone. For example, we could see how often burglary-rapes occur, a crime which is

[5] Based on the Hierarchy Rule, only murder is more serious.

far different than spousal-rape, and in UCR data we'd have no way of differentiating the two. In most cases, however, only one offense occurs per criminal incident (at least as reported in the data), so the opportunity to explore co-occurrence is relatively limited.

2.2 Crimes included in NIBRS

NIBRS data contains far more crime categories than in the UCR data, particularly far more than UCR crime data which contained only eight crimes (and their subcategories of crimes). It also includes several more crime categories than in the UCR arrest data which is far more expansive than the UCR crime dataset. Compared to UCR data, however, there are occasionally more steps you must take to get the same crime category. For example, UCR crime data has the number of gun assaults each month. NIBRS data has the number of aggravated assaults only, but has a variable indicating what weapon the offender used. So you can find out how many aggravated assaults used a gun, giving you the same data as in the UCR, but you need to take extra steps to get there.

Likewise the UCR arrest data has the number of people arrested for selling drugs (broken down into a few different categories of drugs). NIBRS data has if the crime type was a "drug/narcotic violation" which means any crime having to deal with drugs possession, sale, or manufacturing, and excluding drug equipment crimes. We then have to look first at the subcategory of offenses to see if the arrest was for possession, for sale, for manufacturing, or some other kind of drug crime. The final step to be comparable to UCR data is to look at the type of drug involved in the crime. You'll often have to do steps like this during NIBRS research. NIBRS data is available in multiple files that all (for the most part) correspond with each other so you'll tend to have to combine them together to get the complete data you want.

The crimes included in NIBRS are broken into two categories: Group A and Group B crimes.

2.2.1 Group A crimes

The first set of crimes included are Group A crimes and these are really the main crimes included in NIBRS. For each of these crimes we have full data on the victim, the offender, the offense, any property stolen or damaged (or for drug crimes, seized by the police), and info about the arrestee (if any). Of course, not all of this data may be available (e.g. information on the offender is unknown) so there can be significant amounts of missing data, but each crime incident does have corresponding files with this information.

The complete list of Group A crimes is below. I've bolded the Index Crimes which are a flawed, but ubiquitous measure of crime used in the UCR crime data as the main measure of crime in the United States. The Index Crimes are murder, rape (sexual assault with an object and sodomy are only considered rape using the FBI new definition that began in 2013), aggravated assault, robbery (these four are the “Violent Index Crimes”), burglary, motor vehicle theft, and theft (these are the “Property Index Crimes”. Theft here is broken down into several types of theft like purse-snatching and shoplifting. In the UCR crime dataset it is only “theft”.). Arson is also technically an Index Crime (arson is considered a property crime) but is generally excluded. Using Index Crimes as your measure of crime is a bad idea (see here for more on this) but it’s good that all of the Index Crimes are available in NIBRS so we have continuity of data from when agencies move from UCR to NIBRS.

- **Aggravated Assault**
- **All Other Larceny**
- Animal Cruelty
- Arson
- Assisting Or Promoting Prostitution
- Betting/Wagering
- Bribery
- **Burglary/Breaking And Entering**
- Counterfeiting/Forgery
- Credit Card/ATM Fraud
- Destruction/Damage/Vandalism of Property
- Drug Equipment Violations
- Drug/Narcotic Violations
- Embezzlement
- Extortion/Blackmail

- False Pretenses/Swindle/Confidence Game
- Fondling (Incident Liberties/Child Molest)
- Gambling Equipment Violations
- Hacking/Computer Invasion
- Human Trafficking - Commercial Sex Acts
- Human Trafficking - Involuntary Servitude
- Identity Theft
- Impersonation
- Incest
- Intimidation
- Justifiable Homicide
- Kidnapping/Abduction
- **Motor Vehicle Theft**
- **Murder/Non-negligent Manslaughter**
- **Negligent Manslaughter**
- Operating/Promoting/Assisting Gambling
- Pocket-Picking
- Pornography/Obscene Material
- Prostitution

- Purchasing Prostitution
- **Purse-Snatching**
- **Rape**
- **Robbery**
- **Sexual Assault With An Object**
- **Shoplifting**
- **Simple Assault**
- **Sodomy**
- Sports Tampering
- Statutory Rape
- Stolen Property Offenses (Receiving, Selling, Etc.)
- **Theft From Building**
- **Theft From Coin-Operated Machine Or Device**
- **Theft From Motor Vehicle**
- **Theft of Motor Vehicle Parts/Accessories**
- Weapon Law Violations
- Welfare Fraud
- Wire Fraud

2.2.2 Group B crimes

The other set of crimes included in NIBRS are called Group B crimes. For these crimes, only the arrestee segment is available, meaning that we have far more limited data on these incidents than for Group A crimes. Unlike Group A, we only have data here when a person was arrested for the crime, so we don't know how often they occur without an arrest made. These crimes are considered Group B rather than Group A, according to the FBI, because they are less serious or less common than Group A crimes. This isn't really true though. They are certainly less serious than the most serious Group A crimes but include offenses more serious than some Group A crimes. For example, DUIs can potentially lead to serious injury if they crash into someone (if they did then that would likely be considered a charge like manslaughter or assault, but DUIs still have the *potential* to cause great harm) and peeping toms are an invasion of privacy and can cause serious distress to their victims. Relative to crimes like shoplifting, Group B offenses can indeed be more serious. Group B crimes are also quite common, particularly the catch-all category All Other Offenses.

One way I like to think of Group B crimes is that they're mostly - excluding peeping tom - victim-less crimes, or more specifically crimes without a specific victim. For example, in DUIs there's no individual victim; public drunkenness may disturb certain people around the event but they aren't the victims of the drunkenness. There are Group A crimes where the same is true, such as drug offenses, but I think this is a helpful way of thinking about Group B crimes.

- All Other Offenses - excludes traffic violations
- Bad Checks
- Curfew/Loitering/Vagrancy Violations
- Disorderly Conduct
- Driving Under The Influence (DUI)
- Drunkenness
- Family Offenses, Nonviolent
- Liquor Law Violations
- Peeping Tom
- Runaway - only for minors (data ends in 2011)
- Trespass of Real Property

2.3 Differences from UCR data

While NIBRS data is a far more expansive and detailed dataset than the UCR data, in most cases you can convert NIBRS to UCR which allows for continuation of data over time. So the switch from UCR to NIBRS adds a lot of information but loses relatively little. That relatively little amount of difference, however, can impact the types of questions we can ask so they are detailed below.

2.3.1 NIBRS does not have unfounded crimes

In UCR data, which provides monthly counts of crimes (as well as more detailed info on hate crimes and homicides, and monthly counts of arrests), there is a count of "unfounded" crimes in each month. An unfounded crime is just one which was previously reported and then new evidence finds out that it never actually occurred (or that it isn't for the crime that was reported). For example, if you misplace your wallet but think it is stolen you may call the police and report it stolen. This would be recorded in UCR data as a theft. If you then find your wallet and tell the police, then it would be changed to an unfounded crime since the reported theft never actually happened. NIBRS data does not include unfounded data at all so you don't know how many reported crimes turn out to not be true. In practice, this doesn't matter too much as unfounded crimes are rare, constituting generally under 2% of each crime type. The major exception is in rape, where some agencies report that over 10% of rapes in certain years are unfounded. For more on this issue with rape, please see this section of my UCR book here. Given that UCR data already has major issue with rape data, including both changes in the definition of rape in 2013 and evidence that the number of cleared rapes is greatly exaggerated (see this report from ProPublica for more info on this), losing unfounded rapes means losing a not insignificant number of likely real rapes.

Unfounded crimes are also a way that the UCR used to identify justifiable homicides and when police killed someone. But that way was not always used properly and NIBRS data already includes justifiable homicide as a crime category so this isn't a problem.

2.3.1.1 NIBRS does not have negative numbers

Negative numbers in UCR data are because when a crime is reported and then later unfounded, in the month that it is unfounded it is classified as -1 crimes. This is so over the long term (i.e. more than a single month) the positive (but incorrect, and therefore later unfounded) reports and the negative reports to deal with unfounding would equal out so you have the actual number of crimes. In practice though this tended to end up confusing users - though only users who didn't read the manual. Since NIBRS does not have unfounded data, and since it isn't aggregated beyond the incident-level anyways, there are no negative numbers in NIBRS data.

2.3.2 Less information on assaults against officers

One of the UCR datasets, the Law Enforcement Officers Killed and Assaulted (LEOKA) has monthly information on the number of police officers killed and assaulted for each reporting agency. For officers killed it only tells us how many officers were killed "feloniously" which basically means they were murdered or killed accidentally, such as in a car crash. While still only monthly counts, the info on assaults is far more detailed. It includes, for example, the type of call the officer was on during the assault (e.g. responding to a robbery, transporting a prisoner), the hour it occurred, what weapon the offender used, and whether the officer was injured. The data also included information on how many people were employed by the agency as of Halloween of that year, with breakdowns by if that employee is a sworn officer or not, and by the gender of the employee. NIBRS data keeps some of this info. We know if the victim of a crime is a police officer, what injuries they sustained, and any weapon the offender was using. We don't have any info on the call type the officer was on or information about the number of employees in that department with the same breakdown as in LEOKA. So, assuming that agencies that reported data to LEOKA report at the same rate in NIBRS, we'd still be losing information about assaults on police officers. The FBI has said that they're retiring UCR data after 2020, and LEOKA is one of the UCR datasets so assumably this dataset would also be retired, leaving just NIBRS for information on crimes against police officers.[6]

2.4 A summary of each segment file

NIBRS data is often discussed - and is used - as if it was a single file with all of this information available. But it actually comes as multiple different files that each provide different information about a crime incident, including at different levels of analysis so users must clean each segment before merging them together. In this section we'll discuss each of the segments and how they are related to each other. First, keep in mind that NIBRS is at its core an incident-level dataset (hence the "Incident-Based" part of its name). Everything that we have stems from the incident, even though we can get more detailed and look at, for example, individual victims in an incident or even offenses within an incident. Figure 2.1 shows the seven main segments and how they relate to each other.[7] There are also three segments called "window segments" - there is one for arrestees, one of exceptional clearances (i.e. police could have made an arrest but didn't for some reason but still consider the case closed), and one for property - that do not have an associated segment with them, they only

[6] If you have any evidence that they're doing otherwise, please let me know.

[7] There are also segments called "batch headers" which provide information about the agency such as the population under its jurisdiction but we won't cover those since they are agency-level and the same across each incident

have the information available in the given "window" segment. We'll talk about window segments more in Section 2.4.7 below.

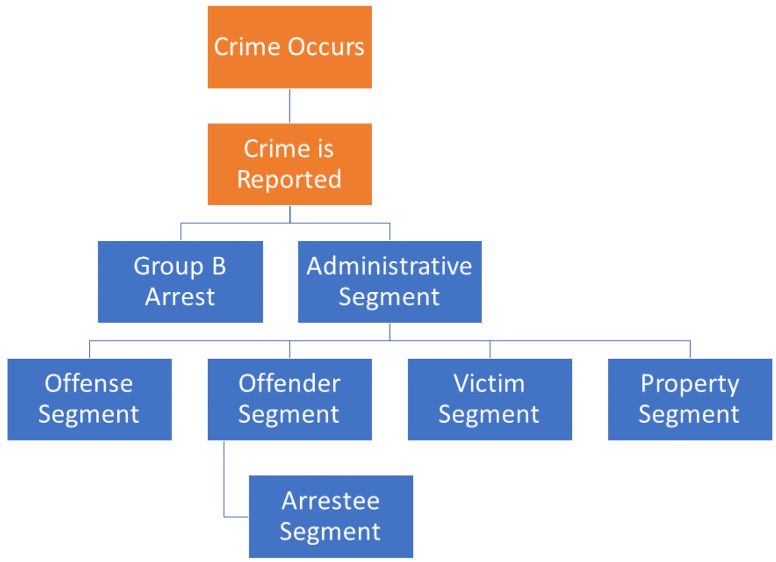

Figure 2.1: The association of each segment file in the NIBRS dataset.

The first two boxes in Figure 2.1, colored in orange, are not part of NIBRS but are part of the data generating process. First, obviously, a crime has to occur. The police then have to learn about the crime. This can happen in two ways. First, they can discover it themselves while on patrol. This is common in crimes such as drug possession or sale as well as any crime that occurs outdoors, which is largely where police are able to observe behavior. The second way is that the victim or witness to a crime reports it. So if they call the police to report a crime, the police learn about it from that call. We don't actually know from the data how the police learned of a crime but it's important to think about this data generating process when using the data. Alongside the crime being reported (or discovered) to the police, agencies must then report the crime to NIBRS. All crimes that occur in that agency's jurisdiction *should* be reported, but that's not always the case. Since reporting is voluntary (at least nationally, though some states do require agencies to report data), agencies are free to report as many or as few crimes as they wish. This often occurs when agencies report only parts of the year, excluding certain months, so you should ensure that the agency reported data for each month you're interested in.

Once a crime occurs and is reported to the police, it can be recorded to NIBRS in two ways, depending on the type of crime that occurred. If it is one of the Group B crimes, then we only get a Group B Arrestee Segment which is the same as the normal arrestee segment which we discuss in more detail below as well as in Chapter 7. In this segment we have useful variables including the type of arrest (e.g. arrested by a warrant), what crime was committed, demographics of the arrestee, and weapon use. However, we're missing a wealth of information that is available in the other segments. When the crime is one of the Group A crimes, we get all of this additional information.

For Group A crimes, we get every other segment, starting with the Administrative Segment. The Administrative Segment is largely a meta-segment - it provides information about other segments. The Administrative Segment is the only incident-level segment of the collection and provides information that is consistent across every offense in the incident such as the incident date and time (in hours of the day). It also includes the type of exceptional clearance for the incident, if the incident was exceptionally cleared. The key part of this segment, however, is that it tells you how many of the Offense, Offender, Victim, and Property segments that are associated with this incident. There are always at least one of these segments per incident, but can potentially be multiple of each segment. These other segments do exactly what their name suggests, providing information about the offenses, offender, victims, and stolen or damaged property for each crime incident. Each of these segments, including the Administrative Segment, have the agency identifier code (the ORI code which is discussed on Section 2.6) and an incident number (which is just a randomly generated unique identifier for that incident) so you can merge the files together. Please note that the incident number of only unique *within* an agency. So there can - and are - incident numbers that are identical across different agencies but are for different incidents. To avoid this issue, make sure you match based on *both* the ORI code and the incident number (or make a new variable with just combines the ORI code and incident number together).

At the bottom is the Arrestee Segment which is only available when a person was arrested for that incident. This provides a bit more detailed data than the Offender Segment for everyone who was arrested for the incident. Now, in reality arrestees aren't necessarily a subset of offenders as some people arrested may not be the ones included in the offender data. Consider, for example, a crime where police initially think two people committed it but end up arresting three people for the crime. The third person would be in the arrestee file but not the offender file. However, in this data there is never a case where there are more arrestees than offenders so it appears that if an offender is arrested who wasn't previously known to the police, they add a corresponding offender segment row for that arrestee.

2.4.1 Administrative Segment

The Administrative Segment provides information about the incident itself, such as how many victims or offenders there were. In practice this means that it tells us how many other segments - offense, offender, victim, and arrestee segments - there are for this particular incident. It also has several important variables at the incident-level such as what hour of the day the incident occurred and whether the incident date variable is actually just the date the incident was reported. Finally, it tells us whether the case was cleared exceptionally and, if so, what type of exceptional clearance it was. This can tell us, for example, how many crimes were cleared because the offender died or the victim refused to cooperate. As the UCR data doesn't differentiate between normal clearances (i.e. arrest the offender) and exceptional clearances, this provides a far deeper understanding of case outcomes.

2.4.2 Offense Segment

This segment provides information about the offense that occurred, and each incident can have multiple offenses. This data tells you which offense occurred and for a subset of offenses it also provides a more detailed subcategory of offense, allowing a deeper dive into what exactly happened. For example, for animal abuse there are four subcategories of offenses: simple/gross neglect of an animal, intentional abuse or torture, animal sexual abuse (bestiality), and organized fighting of animals such as dog or cock fights. This segment also says what date the crime occurred on, where the crime occurred - in categories such as residence or sidewalk rather than exact coordinates in a city - whether the offender is suspected of using drugs, alcohol, or a computer, and which weapon was used. In cases where the weapon was a firearm it says whether that weapon was fully automatic or not. It also provides information on if the crime was a hate crime by including a variable on the bias motivation (if any) of the offender. This is based on evidence that the crime was motivated, at least in part, by the victim's group (e.g. race, sexuality, religion, etc.). There are 34 possible bias motivations and while hate crimes could potentially be motivated by bias against multiple groups, this data only allows for a single bias motivation.

2.4.3 Offender Segment

As might be expected, the Offender Segment provides information about who the offender is for each incident, though this is limited to only demographic variables. So we know the age, sex, and race of each offender but nothing else. This means that important variables such as criminal history, ethnicity, socioeconomic status, and motive are missing. In the Victim Segment we learn about the relationship between the victim and offender, and in the Offense Segment we learn which weapon (if any) the offender used. So there is some other data on the offender in other segments but it's quite limited. This data has one row per offender so incidents with multiple offenders have multiple rows. In cases where there is no information about the offender there will be a single row where all of the offender variables will be "unknown." In these cases having a single row for the offender is merely a placeholder and doesn't necessarily mean that there was only one offender for that incident. However, there's no indicator for when this is a placeholder and when there was actually one offender but whose demographic info is unknown.

2.4.4 Victim Segment

The Victim Segment provides data at the victim-level and includes information about who the victim is and their relationship to offenders. This data tells us what "type" of victim it is with the type meaning if they are a police officer, a civilian (called an "Individual" victim and basically any person who isn't a police officer), a business, the government, etc. It also includes the standard demographics variables in other segments - age, race, sex, ethnicity - as well as whether the victim is a resident (i.e. do they live there?) of the jurisdiction where they were victimized. We also learn from this data what types of injuries (if any) the victim suffered as a result of the crime. This is limited to physical injuries - excluding important outcomes such as mental duress or PTSD - but allows for a much better measure of harm from crime than simply assuming (or using past studies that tend to be old and only look at the cost of crime) what harm comes from certain offenses. There are seven possible injury types (including no injury at all) and victims can report up to five of these injuries so we have a fairly detailed measure of victim injury.

One highly interesting variable is the relationship between the victim and the offender (for up to 10 offenders). This includes, for example, if the victim was the offender's wife, their child, employee, or if the stranger was unknown to them, with 27 total possible relationship categories. You can use this to determine which incidents were crimes by strangers, identify domestic violence, or simply learn about the victim-offender relationship for certain types of crimes. This variable is only available when the victim is a police officer or an "individual." This makes some sense though there could actually be cases where non-human victims (e.g. businesses, religious organizations) do have a relationship with the offender such as an employee stealing from a store. Related to the victim-offender relationship, this segment provides a bit of information about the motive for the crime. For aggravated assaults and homicides, there is a variable with the "circumstance" of the offense which is essentially the reason why the crime occurred. For example, possible circumstances include arguments between people, hunting accidents, child playing with weapon, and mercy killings.

2.4.5 Arrestee and Group B Arrestee Segment

The Arrestee Segment has information that is largely available in other segments. For example, it has information on the arrestee's age, sex, and race which is the same as in the Offender Segment for that individual, and adds only ethnicity and residency status (of the city, not as a United States citizen) to the available demographics variables. It also says the crime the arrestee was arrested for, the weapon used and if the weapon (if it is a firearm) was an automatic weapon, all of which is available in the Offense Segment. Of course, in incidents with multiple offenders these other segments don't directly apply because, for example, having multiple people in a crime where a gun was used doesn't mean they all used it, so the Arrestee Segment does provide more detailed info for specific arrestees. There are a few new variables, however, including the date of the arrest and the type of arrest. The type of arrest is simply whether the person was arrested by police who viewed the crime, if the arrest followed an arrest warrant or a previous arrest (i.e. arrested for a different crime and then police find out you also committed this one so consider you arrested for this one too), and whether the person was cited by police and ordered to appear in court but not formally taken into custody. Finally, for juvenile arrestees it says whether arrestees were "handled within the department" which means that they were released without formal sanctions or were "referred to other authorities" such as juvenile or criminal court, a welfare agency, or probation or parole department (for those on probation or parole).

2.4.6 Property Segment

The Property Segment provides a bit more info than would be expected from the name. For each item involved in the crime it tells you what category that items falls into, with 68 total categories of types of property (including "other") ranging from explosives and pets to money and alcohol. It also tells you the estimated value of that item. This data covers more than just items stolen during a crime. For each item it tells you what happened to that item such as if it was stolen, damaged, seized by police (such as illegal items like drugs), recovered by police, or burned during an arson.

For drug offenses it includes the drugs seized by police. For these offenses, the data tells us the type of drug, with 16 different drug categories ranging from specific ones like marijuana or heroin to broader categories such as "other narcotics". There can be up to three different drugs included in this data - if the person has more than three types of drugs seized then the third drug category will simply indicate that there are more than three drugs, so we learn what the first two drugs are but not the third or greater drugs are in these cases. For each drug we also know exactly how much was seized with one variable saying the amount the police found and another saying the units we should we reading that amount as (e.g. pills, grams, plants).

2.4.7 Window segments

The final set of segments are the "Window" segments which are partial reports meaning that the incident doesn't have all of the other segment files associated with it.[8] There are three window segments Window Arrestee, Window Property, and Window Exceptional Clearance. All three are very rare relative to non-window data and are generally no more than several thousand incidents per year (the non-window data is several million per year). Window files are here when the crime occurred before the agency started reporting to NIBRS and then the arrest happened after they switched to NIBRS.

2.5 Which agencies report data

So if this data has the same info (other than unfounded and negative crimes) as UCR data, but is also far more detailed, why do people ever use UCR data? Besides NIBRS being more complicated to use, far fewer agencies report NIBRS data than do UCR data. Nearly all agencies report crime data for UCR, though fewer do so for some of the UCR datasets such as arrests or arsons - for more, please see my UCR book. In comparison, fewer than half of agencies report to NIBRS, and these agencies are disproportionately smaller and more rural. Starting with 2021 data, the FBI has stopped collecting UCR data, instead only collecting NIBRS data. So if - and this is a very large if - many more agencies move to NIBRS in 2021, we'll start having much more detail from a very representative sample of agencies.[9] Even so, most research - especially policy analyses - requires many years of data so it'll take many years before the full potential of NIBRS data can be realized.

We'll look here at how many agencies report at least one crime each year between 1991 - the first year of data - and 2019 - the latest year of data - as well as compare NIBRS reporting to UCR reporting. Figure 2.2 shows the number of agencies each year that reported at least one incident. Keep in mind that there are about 18,000 police agencies in the United States. Only a little over 600 agencies reported in 1991. This has grown pretty linearly, adding a few hundred agencies each year though that trend accelerated in recent years. In 2019, nearly 8,200 agencies reported at least some data to NIBRS. Compared to the estimated 18,000 police agencies in the United States, however, this is still fewer than half of agencies. The data shown here is potentially an overcount, however, as it includes agencies reporting any crime that year, even if they don't report every month.

Another way to look at reporting is comparing it to reporting to UCR. Figure 2.3 shows

[8] I think the "window" part is a metaphor for while we have lots of info on other incidents, for window segments we have less info so it's like looking through a narrow window. Personally I find it a bit confusing.

[9] I say representative sample as certainly not all agencies will report, even once UCR is no longer collected. It'll take time for all agencies to report, and I doubt we'll ever get to 100% of agencies.

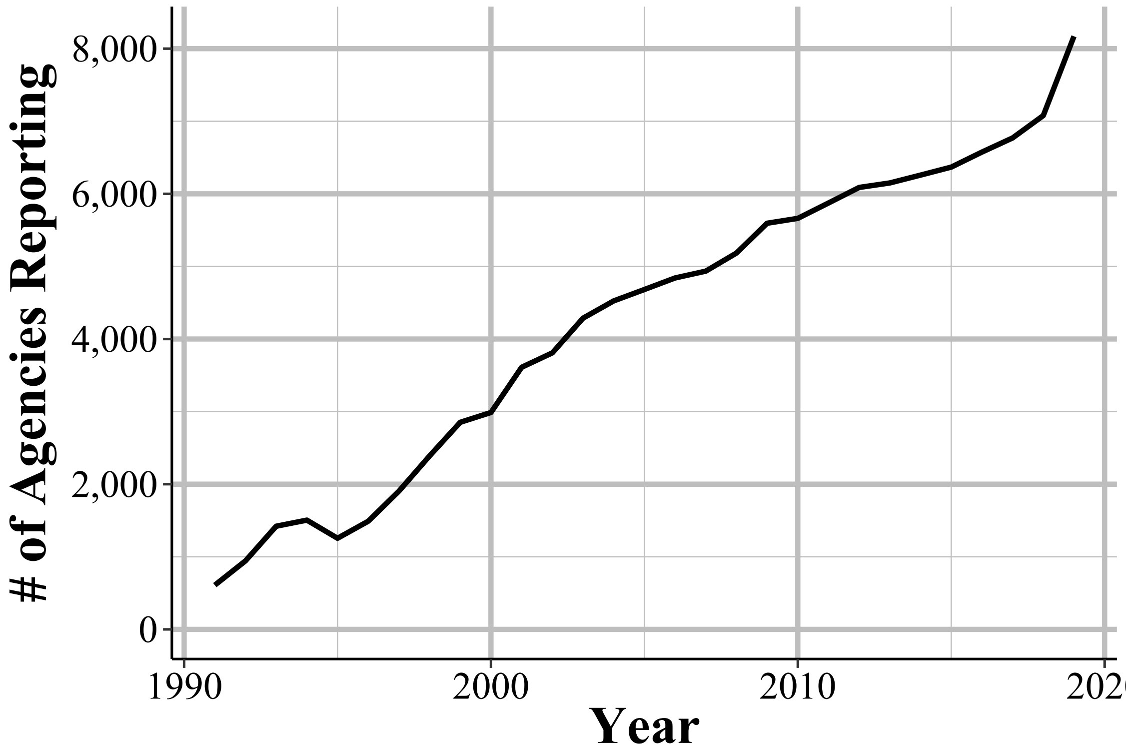

Figure 2.2: The annual number of agencies reporting at least one incident in that year.

the number of agencies in each state that report NIBRS data in 2019. Since 2019 is the year with the most participation, this does overstate reporting for previous years. This map pretty closely follows a population map of the US. Texas had the most agencies, followed by Michigan and Ohio. The southern states have more agencies reporting than the lightly populated northern states. The issue here is that a number of states are in white, indicating that very few agencies reported. Indeed, four of the most populated states - California, New York, Florida, and New Jersey - don't have any agencies at all that report NIBRS data.

Since the number of agencies in a state is partially just a factor of population, Figure 2.4 shows each state as a percent of agencies in that state that report to NIBRS that also reported to the UCR Offenses Known and Clearances by Arrest (the "crime" dataset) in 2019.[10] Not all agencies in the US reported to UCR in 2019 - and a small number reported to NIBRS but not UCR in 2019 - but this is a fairly good measure of reporting rates. Here the story looks a bit different than in the previous figure. Now we can tell that among north-western states and states along the Appalachian mountains, nearly all agencies report. In total, 18 states have 90% or more of agencies that reported to UCR in 2019 also reporting to NIBRS. Thirteen agencies have fewer than 10% of agencies reporting to NIBRS that also reported to UCR, with 5 of these having 0% of agencies reporting. The remaining states average about 56% of agencies reporting. So when using NIBRS data, keep in mind that you have very

[10] This is the UCR dataset which has the highest reporting rate.

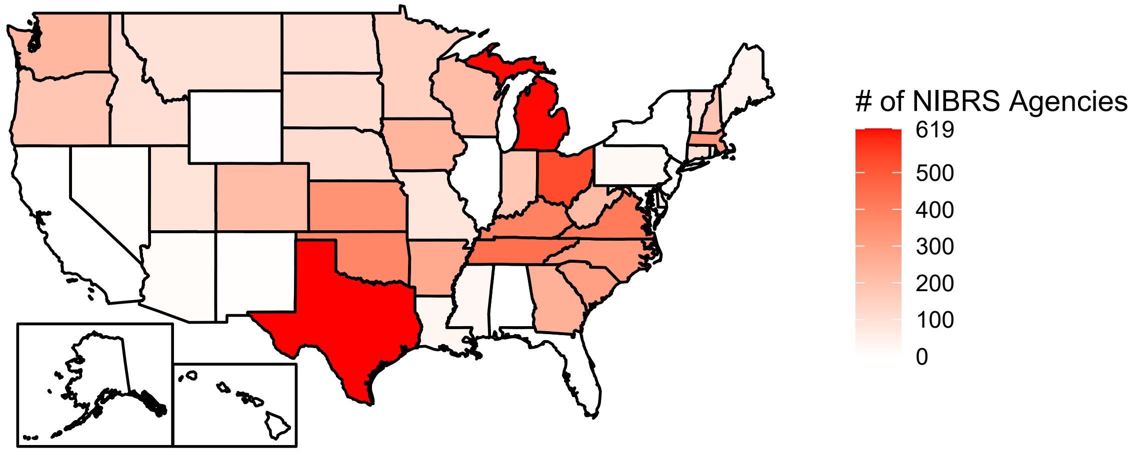

Figure 2.3: The number of agencies in each state that reported at least one crime in 2019 to NIBRS.

good coverage of certain states, and very poor coverage of other states. And the low - or zero - reporting states are systematically high population states.

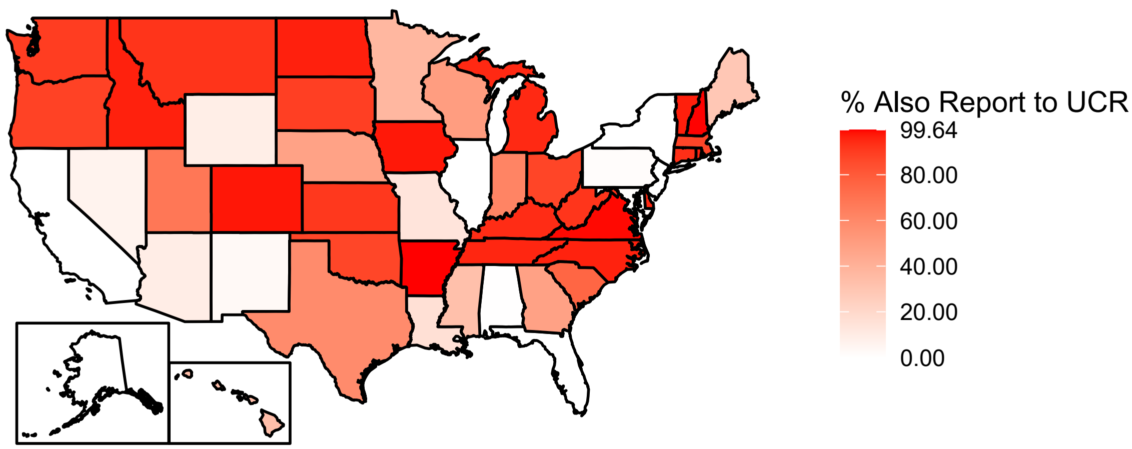

Figure 2.4: Agencies in each state reporting at least one crime to NIBRS in 2019 as a percent of agencies that reported UCR Offenses Known and Clearances by Arrests data in 2019.

For ease of reference, Table 2.1 shows the number of agencies in each state reporting to NIBRS and to UCR in 2019, and the percent shown in Figure 2.4.

Table 2.1: The number of agencies in each state reporting to NIBRS and to UCR in 2019. Also shows NIBRS reporting in each state as a percent of UCR reporting.

State	NIBRS Agencies	UCR Agencies	% of UCR Agencies
Alabama	1	352	0.284090909090909%

State	NIBRS Agencies	UCR Agencies	% of UCR Agencies
Alaska	0	33	0%
Arizona	11	114	9.64912280701754%
Arkansas	277	278	99.6402877697842%
California	0	738	0%
Colorado	217	222	97.7477477477478%
Connecticut	100	107	93.4579439252336%
Delaware	60	63	95.2380952380952%
District of Columbia	1	3	33.3333333333333%
Florida	0	678	0%
Georgia	251	522	48.0842911877395%
Hawaii	1	3	33.3333333333333%
Idaho	104	108	96.2962962962963%
Illinois	1	739	0.13531799729364%
Indiana	180	289	62.2837370242215%
Iowa	240	246	97.5609756097561%
Kansas	343	376	91.2234042553192%
Kentucky	389	413	94.1888619854721%
Louisiana	30	188	15.9574468085106%
Maine	39	135	28.8888888888889%
Maryland	2	156	1.28205128205128%
Massachusetts	317	363	87.3278236914601%
Michigan	616	650	94.7692307692308%
Minnesota	153	409	37.4083129584352%
Mississippi	26	79	32.9113924050633%
Missouri	80	576	13.8888888888889%
Montana	95	103	92.2330097087379%
Nebraska	113	236	47.8813559322034%
Nevada	4	62	6.45161290322581%
New Hampshire	187	188	99.468085106383%
New Jersey	0	578	0%
New Mexico	4	121	3.30578512396694%
New York	0	572	0%
North Carolina	320	333	96.0960960960961%
North Dakota	106	110	96.3636363636364%
Ohio	529	604	87.5827814569536%
Oklahoma	381	437	87.1853546910755%
Oregon	184	208	88.4615384615385%

State	NIBRS Agencies	UCR Agencies	% of UCR Agencies
Pennsylvania	21	1477	1.4218009478673%
Rhode Island	47	49	95.9183673469388%
South Carolina	306	405	75.5555555555556%
South Dakota	115	129	89.1472868217054%
Tennessee	443	465	95.2688172043011%
Texas	619	1053	58.7844254510921%
Utah	88	129	68.2170542635659%
Vermont	86	89	96.6292134831461%
Virginia	411	415	99.0361445783133%
Washington	231	257	89.8832684824903%
West Virginia	221	240	92.0833333333333%
Wisconsin	215	428	50.2336448598131%
Wyoming	5	55	9.09090909090909%

2.6 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.[11] If we used the agency's name we'd 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'd 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 (for some reason the FBI incorrectly puts the abbreviation for Nebraska as NB instead of NE) 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 isn't 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 on my site. Type an agency name or an ORI code into the search section and it will return everything that is a match.

[11] This is refer to this an "ORI", "ORI code", and "ORI number", all of which mean the same thing.

2.7 The data as you get it from the FBI

We'll finish this overview of the NIBRS data by briefly talking about format of the data that is released by the FBI, before the processing done by myself or NACJD 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 2.5 shows what the data looks like as you receive it from the FBI for the 1991 NIBRS data, the first year with data available. The way the FBI releases NIBRS data adds even more complications to using ASCII files. Since there are multiple segments in each NIBRS file you'd think that each segment would be its own file. But, no, the FBI gives you a single file with every segment stacked on top of each other. And, it's "stacked" essentially in a random way where row 1 may be from a certain segment while the next row is from a different segment.

You can think of this a little like chapters from a book. Normally chapter 1 is first, followed by chapter 2 and so on. And within each chapter are sentences that use proper punctuation so we can easily read it and know where one sentence ends and the next begins. The FBI's NIBRS data basically removes all punctuation, cuts up every sentence and rearranges them in a random order and then hands it to you and says "read this." This is a terrible way to present any data, but is what we've had every since data began in 1991 and seems highly unlikely to change.

The "fixed-width" part of the file type is how this works (the ASCII part basically means it's a text file). Each row is the same width - literally the same number of characters, including blank spaces. Though each segment has a different width so you'll first need to grab only the rows that correspond to a particular segment before you can read in that segment. So you must tell the software you are using to process this file - by literally writing code in something called a "setup file" which 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 two character says which segment it is (07 means the Group B Arrestee segment) and the next two characters (in the setup file written as 3-4 since it is characters 3 through 4 [inclusive]) are the state number (01 is the state code for Alabama).[12] So we can read this row as the first column indicating it is an Group B Arrestee segment, 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'll 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 don't 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

[12] We can see several rows down that we have a row starting with "01" which means it is an Administrative Segment, and then starting with "02" which is from the Offense Segment.

```
0701AL0010000-2KY0WSKHEMQ1991040101    T90Z01    41MBUU
0701AL0010000PKMANJRV8W5H1991042401    T90Z01    26MBUU
0701AL00100007HKJQKO3ZHPU1991050201    T90Z01    34MBUU
0701AL0010000-BKGBW5SH27K1991011801    T90Z01    25MBUU
0701AL0010000QYKNY0KK5AEM1991080701    T90Z01    28MBUU
0101AL0010000JZPQUFEM9CTD19910804 00010010100    N
0201AL0010000JZPQUFEM9CTD1991080413BCN  13      40
0401AL0010000JZPQUFEM9CTD1991080400113B
I17MWUR     N    01RU
0501AL0010000JZPQUFEM9CTD19910804O136MW
0701AL00100006PMJW5ST6MQ-1991080301    O90D01    31MWUR
0101AL00100006PMJWOVCSMQ-19910822   010010100    N
0201AL00100006PMJWOVCSMQ-19910822240CN  18
0301AL00100006PMJWOVCSMQ-19910822703000004000        01
0401AL00100006PMJWOVCSMQ-19910822001240
I25MWUN
0501AL00100006PMJWOVCSMQ-199108220100UU
0701AL00100006PMJW2CPDMQ-1991081401    T90Z01    40MBUR
0101AL00100006PMIWS9VZMQ-19910917 16020010100    N
0201AL00100006PMIWS9VZMQ-19910917220CN  20  F
0201AL00100006PMIWS9VZMQ-19910917290CN  20
0301AL00100006PMIWS9VZMQ-19910917416000000040
0301AL00100006PMIWS9VZMQ-19910917777000000130
0401AL00100006PMIWS9VZMQ-19910917001220290
I00UUUR
0501AL00100006PMIWS9VZMQ-199109170100UU
0101AL00100006PMRWOVSUMQ-19910425 21010010100    N
0201AL00100006PMRWOVSUMQ-19910425220CN  05  F
0301AL00100006PMRWOVSUMQ-19910425720000000136
0401AL00100006PMRWOVSUMQ-19910425001220
B
0501AL00100006PMRWOVSUMQ-1991042500
```

Figure 2.5: Fixed-width ASCII file for the 1991 National Incident-Based Reporting System (NIBRS) dataset.

called a "value label."[13] For example, in the above image we saw the characters 3-4 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 and 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. 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'm not sure why they *still* release data 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.[14] **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.[15] 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.[16] Each NIBRS segment has over a dozen columns and potentially dozens of value labels.[17] 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.[18] 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 doesn't mean our work is perfect.[19]

[13]For most fixed-width ASCII files there are also missing values where it'll have placeholder value such as -8 and the setup file will instruct the software to convert that to NA. NIBRS data, however, does not have this and does not indicate when values are missing in this manner.

[14]For those interested in how to actually read in this type of data, please see my R package asciiSetupReader.

[15]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'd guess that at least once you'd make a mistake.

[16]They do provide the instructions in a PDF but you still need to actually make the file yourself. You cannot use their instructions because setup files must be formatted in a very specific way and the PDF is not in this format/

[17]In comparison to UCR data, this is far less complicated to make a setup file for, so the risk of mistakes is far lower.

[18]Other than in my own UCR book.

[19]For evidence of this, please see any of the openICPSR pages for my data as they detail changes I've made in the data such as decisions on what level to aggregate to and mistakes that I made and later found and fixed.

Chapter 3

Administrative and Window Exceptional Clearance Segment

The Administrative Segment provides information about the incident itself, such as how many victims or offenders there were. In practice this means that it tells us how many other segments - offense, victim, offender, and arrestee segments - there are for this particular incident. It also has several important variables at the incident-level such as what hour of the day the incident occurred and whether the incident date variable is actually just the date the incident was reported. Finally, it tells us whether the case was cleared exceptionally and, if so, what type of exceptional clearance it was. This can tell us, for example, how many crimes were cleared because the offender died or the victim refused to cooperate. As the UCR data doesn't differentiate between normal clearances (i.e. arrest the offender) and exceptional clearances, this provides a far deeper understanding of case outcomes.

3.1 Important variables

In addition to the variables detailed below this segment has the tradition agency and incident identifiers: the ORI code, the agency state, the year of this data, and the incident number.

3.1.1 The incident and report date

An important variable, especially for policy analyses, is when the crime happened. NIBRS tells you both the date and the hour of the day for when the crime occurred. We'll start with the date. We can convert the date a few different ways, such as daily, weekly, monthly, quarterly. We could use this precise date to do regression discontinuity studies where we look at days just before and just after some policy change or natural experiment. In this chapter we'll look simply at the percent of crimes each month and each day of the month (overall, not within each month).

Figure 3.1 shows the percent of crimes in the 2019 NIBRS data each month. Past research has found that crimes are lowest when it is cold and hottest when it is hot (and summer also comes with many teens and young adults out of school so have more free time to offend or be victimized). We find these same results with crime lowest in February and steadily increasing until it peaks in July and August, and then decreasing as weather gets cooler in the fall and winter.

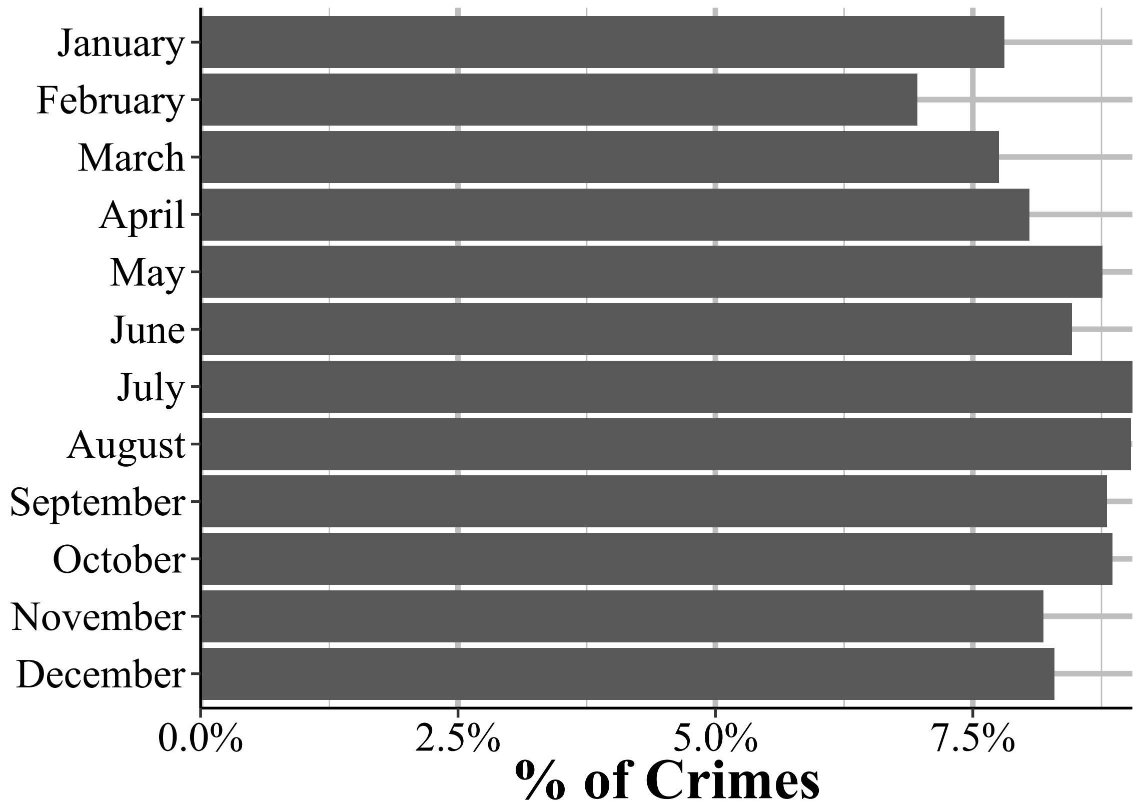

Figure 3.1: The percent of crime incidents in 2019 NIBRS by the month of incident.

We can also look at the days of the month to see if there is any variation here. Figure 3.2 shows the percent of crimes on each day of the month. There's not much variation other than a few days. The 29th and 30th day of the month have fewer crimes, and the 31st day has by far the fewest crimes. These findings are reasonable since not all months have that many days so by definition there are fewest 31st (and 29th, and 30th) days of the month for crimes to occur on. The most common day of the month is the 1st which accounts for 3.95% of all incidents. There's also a small spike on the 15th on the month with 3.39% of incidents occurring on the 15th. No other date is more than 3.35% of incidents. In this data the agencies must report an date, even if they don't know the exact date; there is no option to put "unknown date". So when these agencies don't know when a crime occurred, they put it as the 1st (or less commonly the 15th) of the month. I call this issue the "first-of-the-month error" and it is one I haven't seen discussed in papers that use daily data from NIBRS. For this reason I'd warn against including the 1st of the month in any analysis where the day of the month is relevant.

The above graph showed the days of the month where the incident was said to occur. There is also a variable that says if the date included was the incident date or the date the crime was reported to the police. Figure 3.3 replicates Figure 3.2 but now shows only report dates rather than incident date. There's still a spike for the 1st and 15th but they are very small now and comparable to other days of the month. This provides even more evidence that the incident date spike on the 1st and 15th of the month is just an artifact of agencies putting unknown dates there, rather than a true increase in crime on those days.

3.1.2 Hour of incident

An extremely important aspect of crime data is when exactly the crime occurs. If, for example, crime always spikes when the local high school ends their day that would likely indicate that high school students are involved with crime (both as victims-offenders). In my own research on daylight saving time-crime I only care about the sunset hours, which is when daylight saving time would affect outdoor lighting. When crime happens also would affect police behavior as they'd likely increase patrol during times of elevated crime. Luckily, NIBRS data does have the time of each incident, though it's only at the hour level.

Figure 3.4 shows the total number of crimes that occurred in the 2019 NIBRS data for each hour of the day. There are two key trends in this figure. First, past research has found that crime tends to increase during the night (this is especially true during weekends), drop to a daily low in the couple of hours before sunrise, and then slowly increase as the day progresses.[1] What we find here is a little different. Crime peaks at night at 5-5:59pm which is far earlier than other estimates. Since this is all crimes it could be biased by large increases of certain crimes at this time, such as people coming home from work and finding their house burgled. As crime differ in their timing (e.g. burglary happens often during the day, fights are more common at night), you'll need to merge this segment with the Offense Segment to be able to look at certain types of crimes alone.

The second key trend is the large spike at midnight-12:59am and at noon-12:59pm. The midnight spike certainly seems far too large to be real. The midnight hour is more than double that of neighboring hours, and reverses the trend in the hours before it. Crime increases more at night (though not as much here as in other studies of this topic) but not

[1] In all of the police ridealongs I've been on the police tend to stop patrolling in early morning (e.g. 3-4am) and go back to the station to do paperwork. I think this likely partially explains the findings that crime is lowest around 4-5am.

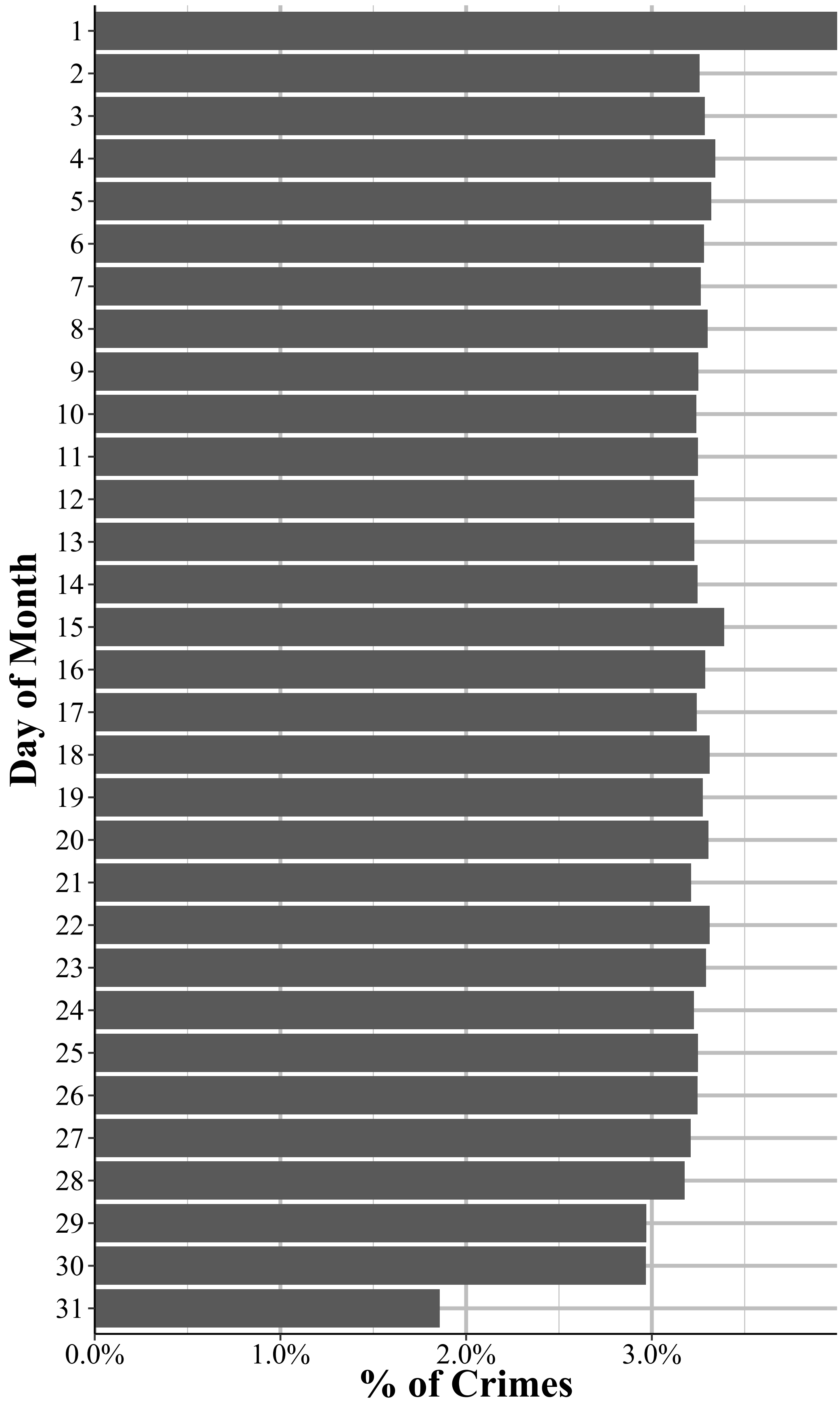

Figure 3.2: The percent of crimes that occur (the day of the incident, even if the crime wasn't reported that day) each day of the month for all agencies reporting to NIBRS in 2019.

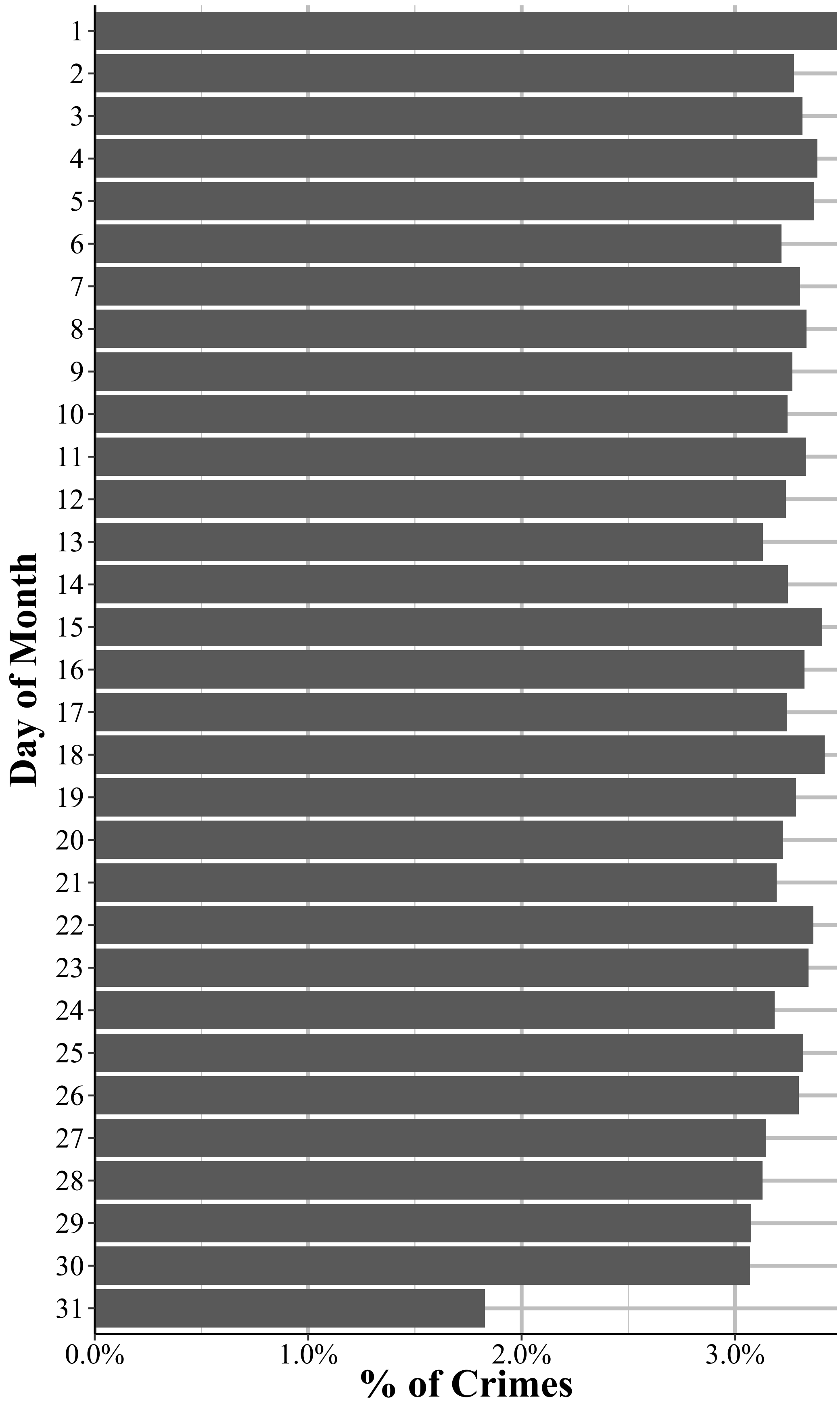

Figure 3.3: The percent of crimes that are reported (the day of the report, even if not the day of the incident) each day of the month for all agencies reporting to NIBRS in 2019.

by such a massive amount. The noon hour is about 50% larger than in the neighboring hours, so is a sizable increase though continues the trend of increasing crime during the day and is a far smaller increase than at midnight. It seems unlikely that this spike is real, though potentially some of it could be. For example, if a lot of people take their lunch at noon then there could be an increase of crime (such as pickpocketing and purse snatching) when these people gather at restaurants and malls to eat. So what are these spikes? There is no option to say that the hour is unknown so the police must put something in, and it seems that they put midnight and noon as the default "unknown hour" hour.[2] Given this, I recommend that you remove the midnight and noon hour from any analysis you do where the time of day is relevant.

3.1.3 Exceptional clearance

When we speak of clearances we generally mean that a person was arrested for the crime.[3] Cases may also be cleared "through exceptional means" which is also called an "exceptional clearance." An exceptional clearance means that the police have identified who the offender is (in cases with multiple offenders, they need to identify only one offender), have sufficient evidence and ability (e.g. know where they are) to arrest this offender but are unable to do so for some reason. Basically, if they could arrest them they would but for some reason they can't. NIBRS data tells us if the case is exceptionally cleared as well as the reason for the exceptional clearance.

Exceptional clearances are quite rare, making up only about 5% of cases. Figure 3.5 shows the breakdown of reasons why the case was cleared for these ~5% of cases that are exceptionally cleared. The most common reason, at 47.3% of exceptional clearances, is that the victim refused to cooperate with the police or prosecution. This can include cases where the victim reported their crime to the police and then later decide to stop assisting. The next most common reason, at 45.3% of exceptionally cleared cases, is that the prosecution decides to not prosecute the case. This excludes cases where the prosecution believes that there is not probably cause to make the arrest. Therefore it largely includes cases that the prosecution either doesn't believe they could win or where the agency has a policy of non-prosecution - this is likely increasingly common in jurisdiction which has "progressive prosecutors" who de facto legalize certain crimes.[4]

---

[2] To my knowledge these spikes are not discussed in papers that use NIBRS data other than in my own papers.

[3] While a more expansive definition may include a conviction in a court for that crime (including pleading guilty), NIBRS data only extends to the arrest stage so we never know the judicial outcome of the case.

[4] Of course this is an empirical question and one that could potentially be examined through NIBRS data. Unfortunately, progressive prosecutors are generally found in large urban cities (e.g. Philadelphia), which also tend to not report NIBRS data.

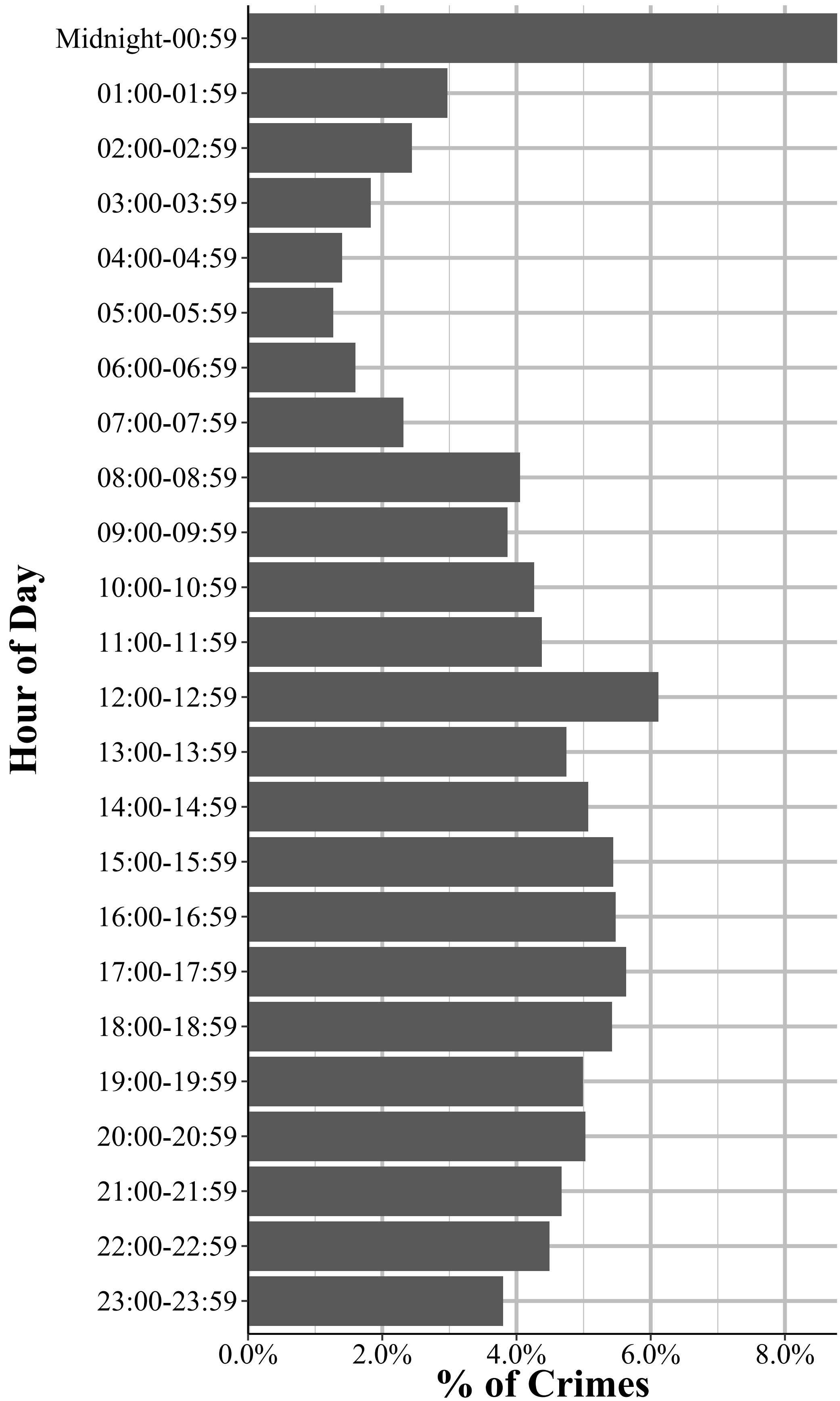

Figure 3.4: The percent of crimes that are reported each hour for all agencies reporting to NIBRS in 2019.

The next most common group is when the offender is a juvenile and the police chose to avoid arresting them due to their age. This occurs in minor offenses such as property cries and the police do give notice to the juvenile offender's parents about the situation. If the offender is in another agency's jurisdiction (including out of the country) and the agency reporting data is unable to make an arrest, including when the agency who has the offender in their jurisdiction refuses to extradite the offender, the case can be exceptionally cleared. And this occurs in 1.6% of exceptional clearances. In these cases we don't know any information about which jurisdiction the offender is in at the time of the exceptional clearance. Finally, if the offender dies (by any means) before the arrest they can't be arrested so the case is exceptionally close. This makes up about 0.9% of exceptional clearances.

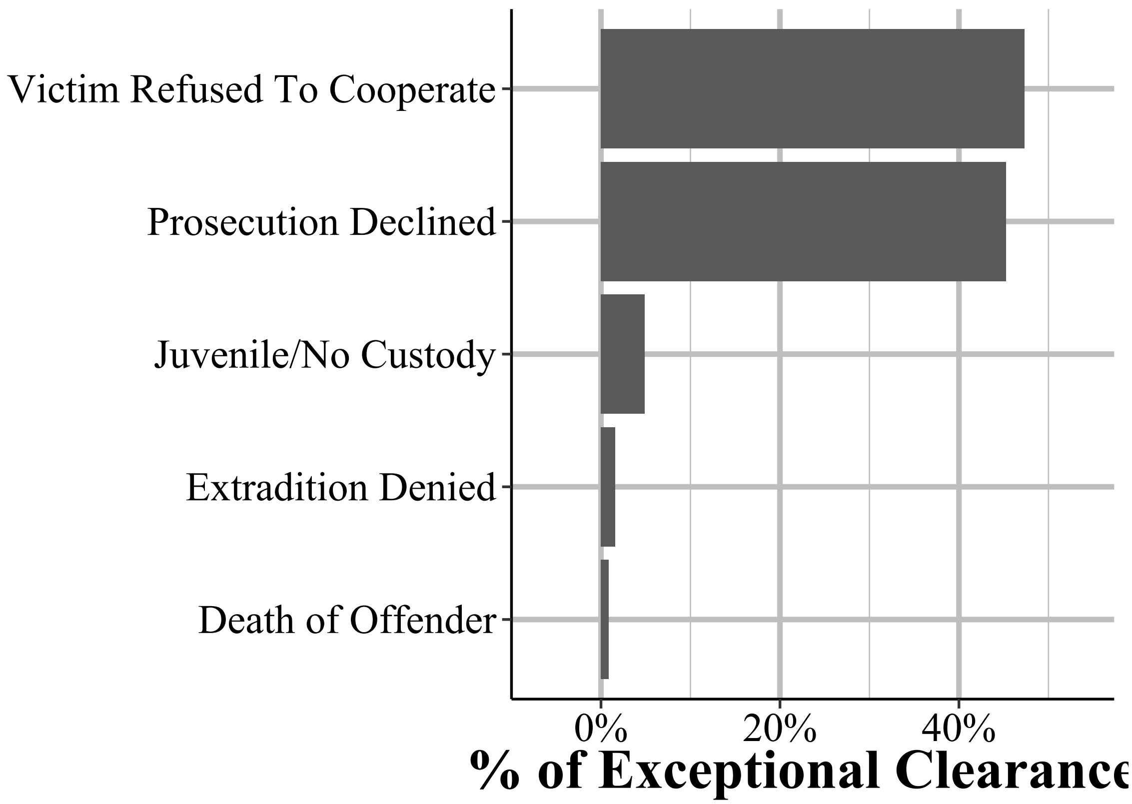

Figure 3.5: The distribution of exceptional clearances for all exceptional clearances reported to NIBRS in 2019.

3.1.4 Number of other segments

The "Administrative" part of this segment is that it tells us about other segments related to this incident. Here we know how many Offense, Offender, Victim, and Arrestee segments there are for the incident. In effect it says how many crimes were committed, offenders involved (at least the number known to police), victims involved, and people arrested for this particular incident. This can be useful both as a check to make sure you're not missing anything when looking at the other segments and to quickly subset data, such as to only single-victim or only multiple-offender incidents.

3.1.4.1 Offense Segments

This variable indicates how many offense segments there are associated with this incident, with the possible values ranging from one to nine. Each incident can have multiple offenses, so this just says how many of these crimes there were. For example, if a person is raped and robbed then there'd be two offense segments related to that incident. Figure 3.6 shows the number of offense segments - and thus the number of crimes - associated with each incident. The vast majority of incidents only have one offense reported, making up 88.4% of incidents.[5] This drops considerably to 10.3% of incidents having two offenses, 1.1% having three, and then under 0.15% of incidents having four through nine offenses.

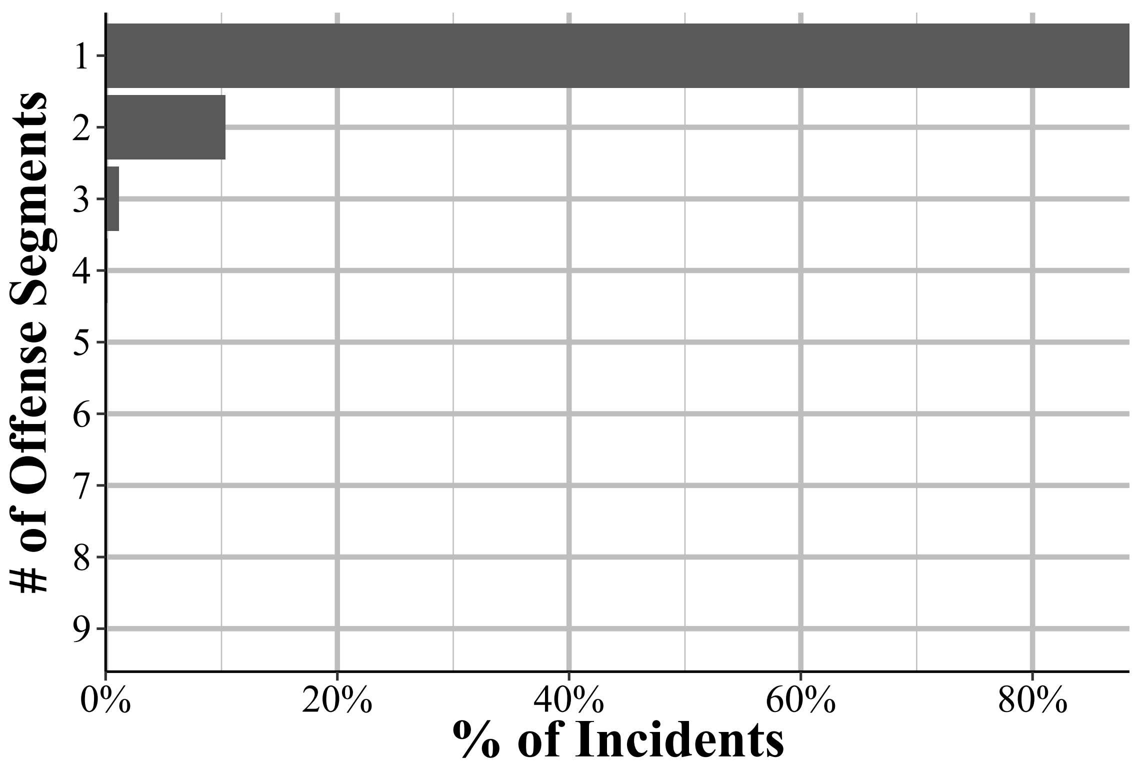

Figure 3.6: The distribution for the number of Offender Segments per incident, for all incidents in NIBRS 2019.

[5] In reality a person who commits a crime may be arrested or charged with many (often highly related) offenses related to a single criminal incident. So this data does report fewer incidents than you'd likely find in other data sources, such as if you request data from a local police agency or district attorney's office.

3.1.4.2 Offender Segments

The Administrative Segment tells you how many offenders are involved with an incident. This is, of course, an estimate because in some incidents the police don't know how many people are involved. If, for example, someone was robbed then they can tell the police how many robbers there were. But if someone comes home to find their home burglarized then they don't know how many burglars there were. If there's no video evidence (e.g. a home security camera) and neighbors didn't see anything then the police would not know how many offenders were involved in the incident. In these cases they put in a single offender and in the Offender Segment all of the information about the offender is "unknown." The remaining number of offenders are still estimates as the police may not know for sure how many offenders were involved, but this is more reliable than when there is only a single offender reported. With that major caveat in mind, Figure 3.7 shows the distribution in how many offenders there were per incident.

The vast majority of incidents have only one (or potentially an unknown number) offenders, at 89.8% percent of incidents. Incidents with two offenders make up only 7.9% of incidents while those with three make up 1.6% of incidents. No other number of offenders make up more than 0.5% of incidents. The data does have the exact number of offenders but I've top coded it to 10 for simplicity. There can potentially be a large number of offenders involved in an incidents and in the 2019 NIBRS data the incident with the higher number of offenders had 81. However, it is exceedingly rare for there to be even more than a handful of offenders.

3.1.4.3 Victim Segments

In cases where the offense is a "victimless crime" (or at least one where there's no specific victim) such as a drug offense, the victim and the offender can be the same individual. For other cases, being a victim in an incident just means that you were the victim of at least one offense committed in the incident. Figure 3.8 shows the distribution in the number of victims per incident. Like the number of offenses and offenders, this is massively skewed to the left with 90.3% of incidents having a single victim. Incidents with two victims make up 7.9% of the data while incidents with three victims are 1.2%. All remaining numbers of victims are less than one third of one percent of the data each. The data does have the exact number of victims but I've top coded it to 10 for simplicity. The incident with the most victims in 2019 had 207 victims.

3.1.4.4 Arrestee Segments

Unlike the previous segments, there may not always be an arrestee segment since not all crimes lead to an arrest. Figure 3.9 shows the distribution in the number of arrestee segments

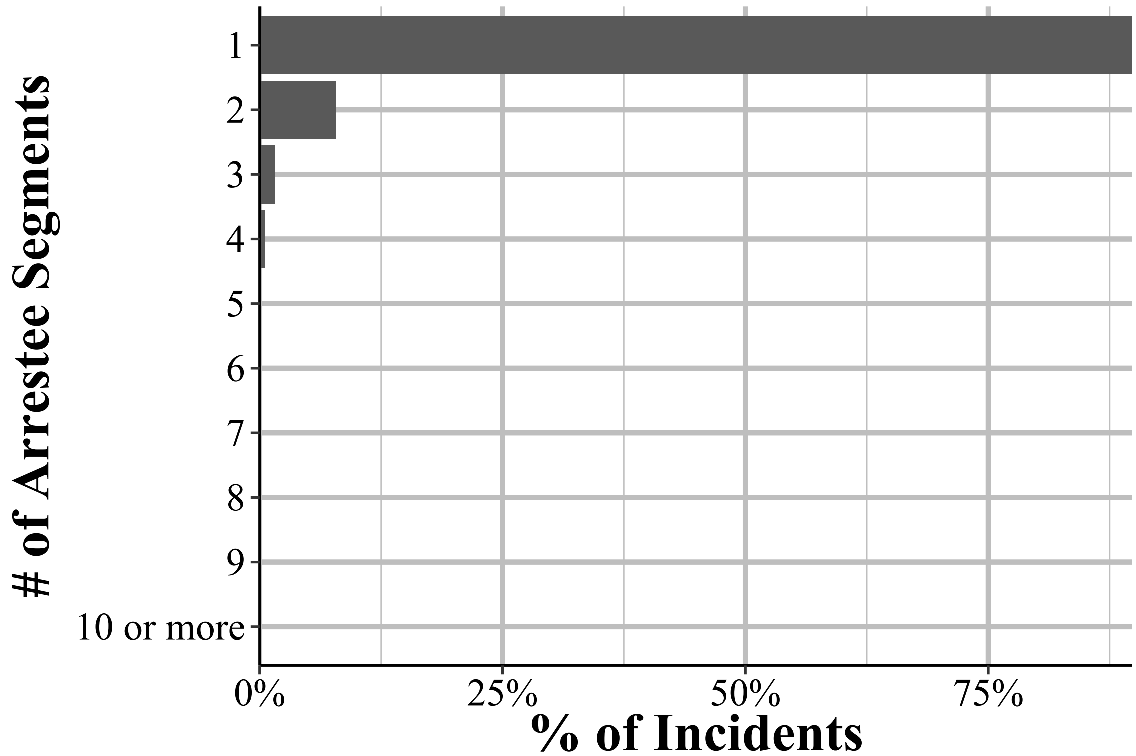

Figure 3.7: The distribution for the number of Offender Segments per incident, for all incidents in NIBRS 2019.

per incident in the 2019 NIBRS data. Indeed, the vast majority - 72.9% of incidents - didn't lead to a single arrest. In 24% of incidents a single person was arrested while in 2.5% of incidents two people were arrested. The remaining numbers of people arrested are increasingly small with a little over a third of one percent of incidents having three people arrested and no greater than 0.1% for larger numbers of people arrested. The incident with the most arrests in 2019 led to 81 people arrested.

Of course, to really understand these arrests we'd need to know how many people committed the crime. Having one arrest for an incident with one offender is good, having one arrest when there are multiple offenders means relatively low clearance rates. While we don't know the true number of offenders (as police may not know how many there actually were), we can use the Offender Segment count as a good estimate. Figure 3.10 shows the percent of people arrested for each number of offenders in the incident. Keep in mind that when there's one offender reported this can potentially mean the police don't know how many offenders there are, so it's a bit unreliable. There is wide variability in the percent of offenders arrested by the number of offenders in an incident. In cases with one offender, there was an arrest made only 25% of the time. When there are two offenders, about 36% of offenders are arrested, on average, and this steadily declines as the number of offenders increase back down to 25.7% with six offenders. This moves a percentage point or two as the number of offenders increases

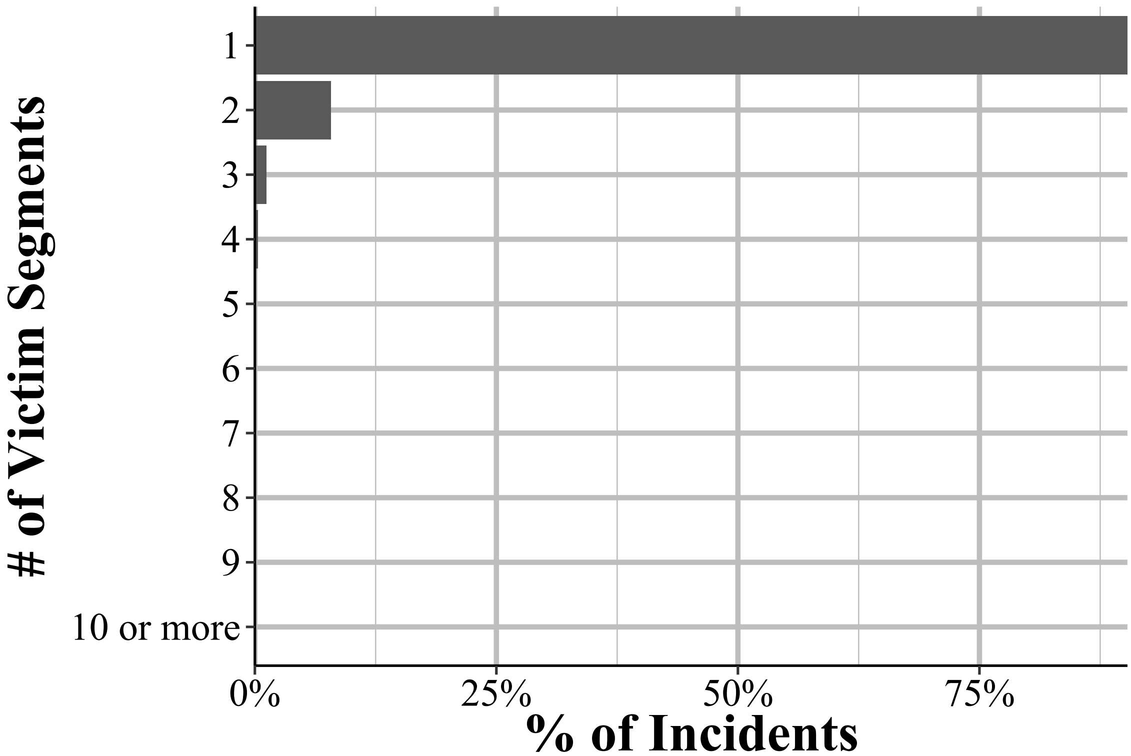

Figure 3.8: The distribution for the number of Victim Segments per incident, for all incidents in NIBRS 2019.

but not by very much.

Another way to look at this is to examine the percent of incidents, by number of offenders, where at least one offender was arrested. Here we're asking a very different question than in the above figure. Instead of asking, of all offenders who committed a crime what number are arrested, we want to know of all incidents how many led to at least one arrest. This is actually how the UCR defines clearance rates, at least one person arrested (no matter how many offenders there are) clears the entire incident.

Figure 3.11 shows this analysis by giving the percent of incidents by the number of offenders where at least one offender was arrested. As before, 25% of incidents with one offender had a person arrested. Now we see a steady decline in the percent of incidents arrested as the number of offenders increase. This figure shows that relatively few incidents ever lead to an arrest of at least one offender, and the chance that an arrest is made decreases as the number of offenders increases. These findings suggest that the larger numbers found in Figure 3.10 are because when at least one offender is arrested in an incident it's also likely that multiple offenders are arrested.

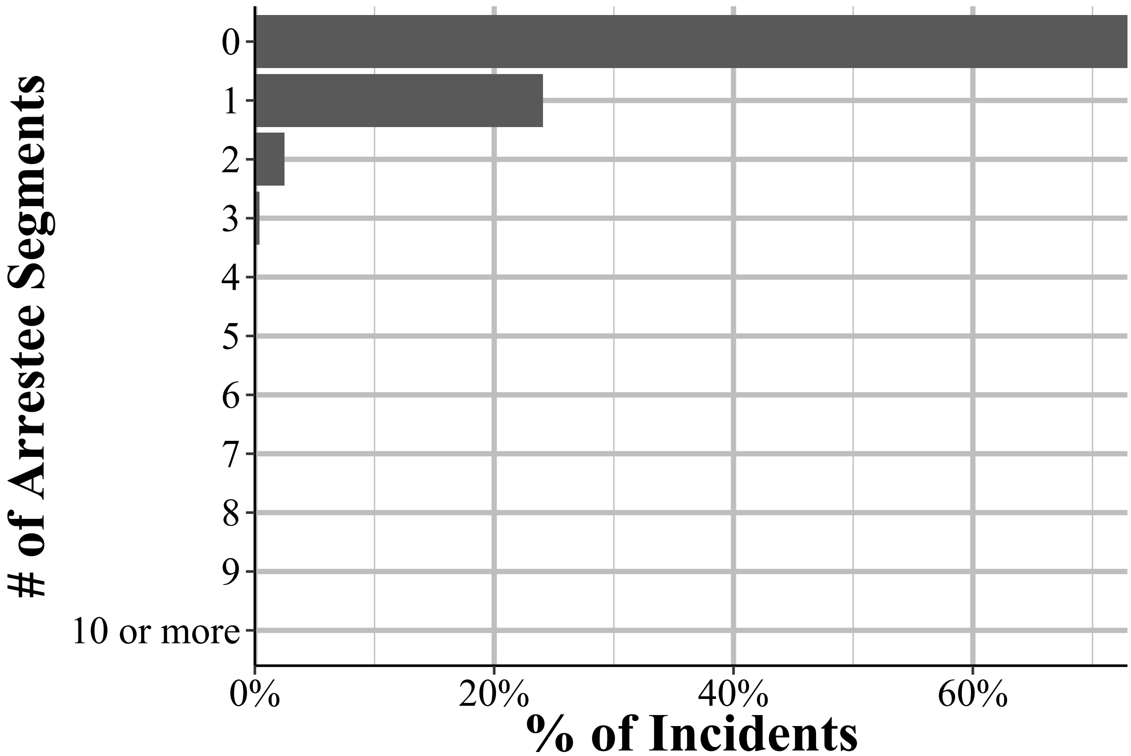

Figure 3.9: The distribution for the number of Arrestee Segments per incident, for all incidents in NIBRS 2019.

Figure 3.10: The percent of people arrested by the number of offenders in an incident.

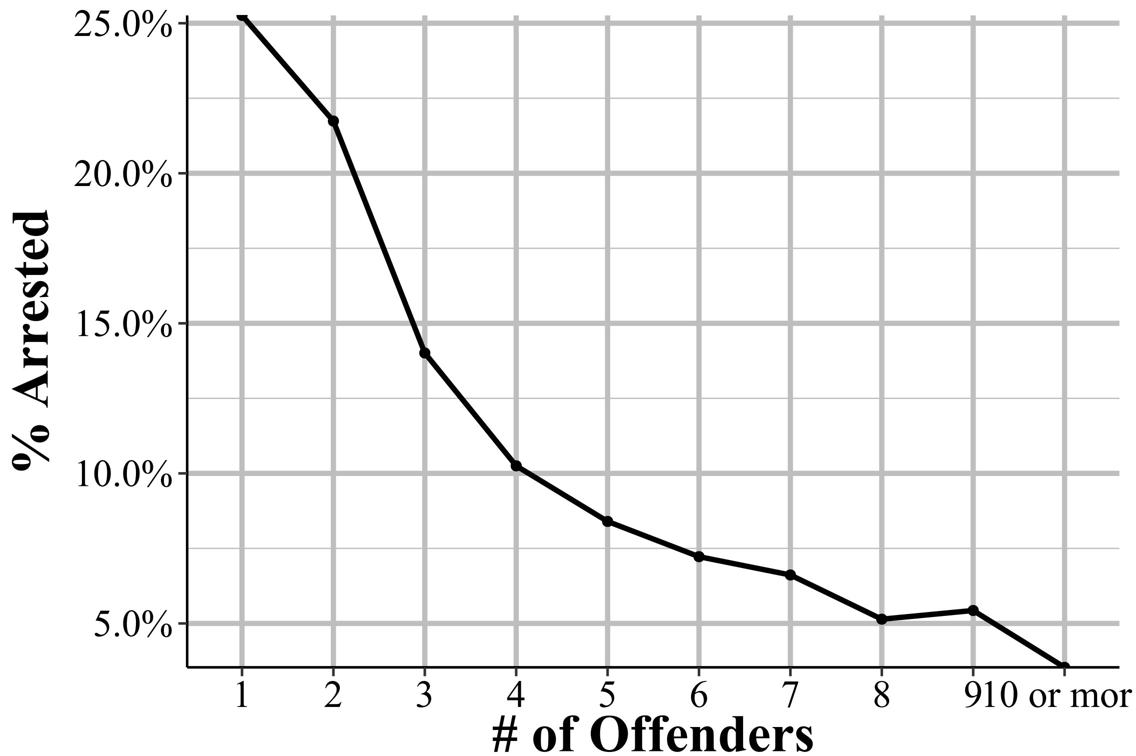

Figure 3.11: The percent of incidents by number of offenders where at least one offender is arrested.

Chapter 4

Offense Segment

This segment provides information about the offense that occurred, and each incident can have multiple offenses. This data tells you which offense occurred and for a subset of offenses it also provides a more detailed subcategory of offense, allowing a deeper dive into what exactly happened. For example, for animal abuse there are four subcategories of offenses: simple/gross neglect of an animal, intentional abuse or torture, animal sexual abuse (bestiality), and organized fighting of animals such as dog or cock fights. This segment also says what date the crime occurred on, where the crime occurred - in categories such as residence or sidewalk rather than exact coordinates in a city - whether the offender is suspected of using drugs, alcohol, or a computer, and which weapon was used. In cases where the weapon was a firearm it says whether that weapon was fully automatic or not. It also provides information on if the crime was a hate crime by including a variable on the bias motivation (if any) of the offender. This is based on evidence that the crime was motivated, at least in part, by the victim's group (e.g. race, sexuality, religion, etc.). There are 34 possible bias motivations and while hate crimes could potentially be motivated by bias against multiple groups, this data only allows for a single bias motivation.

4.1 Important variables

In addition to the variables detailed below this segment has the tradition agency and incident identifiers: the ORI code, the agency state, the year of this data, and the incident number. It also has the date of the incident, which is also included in the Administrative Segment.[1]

[1] Potentially some offenses could occur after the incident date. For example, if a incident happened at 11:50pm and it had multiple crimes involved, some of them may occur the follow day, such as at 12:01am.

4.1.1 Crime category

The first important variable in the Offense Segment is figuring out exactly what offense was committed. This tells you what crime occurred in the incident. There can be multiple crimes in a single incident so this provides info about each crime that occurred. To figure out which offenses belong together, just look at the incident number and the ORI. Within ORI, each incident number is a unique identifier for an incident. Each crime is mutually exclusive and crimes which are elements of another crime are counted only as the crime they are elements of. For example, robberies are basically theft plus assault/intimidation - it is the use of force or threat of force (assault or intimidation) to take property (theft). A case of robbery in this data would only count as robbery, not as robbery and theft and assault/intimidation. If there are these crimes together in an incident that's because that crime *also* occurred separately. For example, if someone is robbed and after the robbery is over (i.e. they hand over their belongings) they are then punched repeatedly, that could potentially be classified as a robbery and an assault.

Table 4.1 shows each possible crime in the data and how common it was in 2019. It's sorted by frequency instead of alphabetically so it's easier to see which crimes are most common. There were about 7.4 million crimes reported to NIBRS in 2019. The most common crime is simple assault - which is an assault that didn't use a weapon and didn't result in serious injury - at 12.7% of crimes, or about 944k crimes. If you think this is odd because property crimes are more common than violent crimes, you'd be right. NIBRS data is pretty specific in its crime categories so it splits up certain crimes into a number of different categories. Theft is the most common crime committed in the United States. In NIBRS it's broken into several different types of theft so you need to combine them together to actually measure theft in its entirety. Of the top 6 most common crimes, theft crimes make up ranks 3, 5, and 6 (all other larceny, theft from motor vehicle, and shoplifting).

Though each agency is supposed to report the same crimes - using the exact same definition of the crimes so the data is consistent - that isn't always true in practice. For example, animal cruelty became a NIBRS crime in 2018 (before that it wasn't an option so agencies could not report it) and likely most agencies in the US have had at least one animal abuse crime since then. In 2018, however, reporting was concentrated in a small number of states, meaning that not all agencies reported that offense. The concentration in certain states suggests that this low reporting is due to agencies in certain states deciding (or not being able to, such as if having older reporting systems which don't have animal cruelty as an option) not to report that offense at all. Reporting has increased in 2019 (though still remains highly concentrated), suggesting that over time more agencies begin reporting crimes as they are added. Therefore, I strongly suggest examining your data over time and across geographic areas to see if there are any biases before using it.

Table 4.1: The number and percent of crimes reported from all agencies in 2019, by crime category.

Crime Category	# of Offenses	% of Offenses
Simple Assault	944,601	12.70%
Destruction/Damage/Vandalism of Property	821,523	11.05%
All Other Larceny	810,138	10.89%
Drug/Narcotic Violations	710,822	9.56%
Theft From Motor Vehicle	550,202	7.40%
Shoplifting	487,848	6.56%
Burglary/Breaking And Entering	476,554	6.41%
Drug Equipment Violations	327,715	4.41%
Motor Vehicle Theft	304,964	4.10%
Intimidation	279,040	3.75%
Aggravated Assault	271,444	3.65%
Theft From Building	212,507	2.86%
False Pretenses/Swindle/Confidence Game	195,003	2.62%
Credit Card/Atm Fraud	134,527	1.81%
Weapon Law Violations	134,182	1.80%
Counterfeiting/Forgery	112,766	1.52%
Theft of Motor Vehicle Parts/Accessories	111,384	1.50%
Robbery	92,935	1.25%
Identity Theft	68,515	0.92%
Stolen Property Offenses (Receiving, Selling, Etc.)	58,449	0.79%
Impersonation	57,920	0.78%
Fondling (Incident Liberties/Child Molest)	46,141	0.62%
Rape	46,049	0.62%
Embezzlement	24,166	0.32%
Kidnapping/Abduction	22,757	0.31%
Pornography/Obscene Material	19,814	0.27%
Arson	15,500	0.21%
Wire Fraud	15,373	0.21%
Sodomy	11,046	0.15%
Animal Cruelty	9,956	0.13%
Pocket-Picking	9,527	0.13%
Purse-Snatching	7,487	0.10%
Prostitution	7,309	0.10%
Murder/Nonnegligent Manslaughter	6,095	0.08%

Crime Category	# of Offenses	% of Offenses
Extortion/Blackmail	5,583	0.08%
Statutory Rape	5,430	0.07%
Theft From Coin-Operated Machine Or Device	4,617	0.06%
Sexual Assault With An Object	4,599	0.06%
Hacking/Computer Invasion	2,765	0.04%
Assisting Or Promoting Prostitution	2,283	0.03%
Purchasing Prostitution	1,159	0.02%
Welfare Fraud	940	0.01%
Human Trafficking - Commercial Sex Acts	915	0.01%
Incest	762	0.01%
Operating/Promoting/Assisting Gambling	613	0.01%
Negligent Manslaughter	561	0.01%
Bribery	544	0.01%
Gambling Equipment Violations	351	0.00%
Justifiable Homicide	300	0.00%
Betting/Wagering	260	0.00%
Human Trafficking - Involuntary Servitude	142	0.00%
Sports Tampering	7	0.00%
Total	7,436,090	100%

4.1.2 Offense subtype

In addition to the broader crime committed, NIBRS does allow for a “subtype” of crime variable which gives us a bit more information about what crime occurred (the variable is technically called the “type of criminal activity”). This is especially useful for certain crimes where it’s not clear exactly what they’re referring to from the crime category along. For example, for drug crimes we generally differentiate possession from sale or manufacturing. NIBRS combines everything into “drug/narcotic violations (crimes for drug materials such as syringes are classified as”drug equipment violations“). So we need to use the subtype variable to figure out what type of drug crime it is. Looking at the subtype we can see if the arrest is for”distributing/selling“,”operating/promoting/assisting“,”possessing/concealing“,”transporting/transmit or”using/consuming”. There can be up to three subtypes per offense, so potentially an arrest can be related to something such as both possessing and selling drugs.

There are also some unexpected subtypes related to certain offenses. For example, there are a few dozen drug offenses that also have the subtype of “exploiting children”. This subtype is generally for cases where a child is abused, but happens here for drug offenses that don’t have any associated child abuse (or for some of them, doesn't have any other crime at all) offense. The reason, I believe, for this category is that these offenses occurred in public so could have been viewed by children, and were labeled as exploiting children for that reason. If you're studying crimes against children, pulling from this variable would likely overcount crimes so - as always - you should make sure that the data you carefully check your data to odd things like this.[2]

This data is only available for the below subset of crimes, and isn't always present even for these crimes.

- Aggravated Assault
- Animal Cruelty
- Counterfeiting/Forgery
- Drug Equipment Violations
- Drug/Narcotic Violations
- Fondling (Incident Liberties/Child Molest)
- Gambling Equipment Violations
- Intimidation
- Kidnapping/Abduction
- Murder/Non-negligent Manslaughter
- Negligent Manslaughter
- Pornography/Obscene Material
- Rape
- Robbery
- Sexual Assault With An Object
- Simple Assault
- Sodomy
- Stolen Property Offenses (Receiving, Selling, Etc.)
- Weapon Law Violations

Table 4.2: The number and percent of crime subtypes. This breakdown is only available for a subset of offenses.

Crime Subcategory	# of Offenses	% of Offenses
Possessing/Concealing	1,074,646	44.35%
None/Unknown Gang Involvement (Mutually Exclusive)	1,040,062	42.92%

[2]Whether children viewing a crime, even a drug crime, would count as a crime against children would, of course, depend on your definition.

Crime Subcategory	# of Offenses	% of Offenses
Distributing/Selling	100,708	4.16%
Using/Consuming	90,049	3.72%
Buying/Receiving	46,692	1.93%
Cultivating/Manufacturing/Publishing	26,289	1.08%
Operating/Promoting/Assisting	12,446	0.51%
Transporting/Transmitting/Importing	7,821	0.32%
Simple/Gross Neglect (Unintentionally, Intentionally, Or Knowingly Failing To Provide Food, Water, Shelter, Veterinary Care, Hoarding, Etc.)	6,996	0.29%
Other Gang	6,482	0.27%
Exploiting Children	5,448	0.22%
Intentional Abuse And Torture (Tormenting, Mutilating, Poisoning, Or Abandonment)	2,770	0.11%
Juvenile Gang Involvement	2,582	0.11%
Animal Sexual Abuse (Bestiality)	104	0.00%
Organized Abuse (Dog Fighting And Cock Fighting)	86	0.00%
Total	2,423,181	100%

We'll look in more detail about the subtype of offenses for animal cruelty. Table 4.2 shows each possible subtype for animal cruelty and how often they occur. There were about 10,000 cases of animal cruelty reporting to NIBRS in 2019 and over two-thirds are for neglect of the animal. Over a quarter are for torturing or abandoning the poor animal. And the remaining small share of offenses are for sexual abuse of the animal or for forcing them to fight other animals. These subtypes provide a lot more information about the crime that occurred, but still has some uncertainties. We don't, for example, know the type of animal involved or the severity of the abuse (other than that it was serious enough for police to become involved). Still, this is a large improvement in our understanding of this crime (and others which have subtypes), and an colossal improvement when compared with UCR data.

Table 4.3: The number and percent of crime subtypes for animal abuse.

Crime Subcategory	# of Offenses	% of Offenses
Simple/Gross Neglect (Unintentionally, Intentionally, Or Knowingly Failing To Provide Food, Water, Shelter, Veterinary Care, Hoarding, Etc.)	6,996	70.27%
Intentional Abuse And Torture (Tormenting, Mutilating, Poisoning, Or Abandonment)	2,770	27.82%
Animal Sexual Abuse (Bestiality)	104	1.04%
Organized Abuse (Dog Fighting And Cock Fighting)	86	0.86%
Total	9,956	100%

4.1.3 Offense completed

For each offense, this segment also tells you if the offense was completed or only attempted. Some offenses, such as simple and aggravated assault or homicide, are only labeled as completed. This is because an attempted murder, for example, would be classified as aggravated assault. Since crimes in NIBRS are mutually exclusive, there cannot be both attempted murder and aggravated assault, so only aggravated assault is included. This does limit the data as it is important to know when an aggravated assault is done with the intent to kill the victim and when it's just to seriously harm the victim (though measuring this would likely be extremely imprecise since it requires knowing the motives of the offender). For other crimes, we do know if each crime was completed or not. In the vast majority of offenses they are completed. Table 4.4 shows the percent of each crime category in 2019 NIBRS data that was completed or was only attempted.

Table 4.4: The percent of crimes completed or attempted, by crime category.

Crime Category	% Completed	% Attempted
Aggravated Assault	100 %	0 %
All Other Larceny	99.01 %	0.99 %
Animal Cruelty	97.34 %	2.66 %
Arson	94.83 %	5.17 %
Assisting Or Promoting Prostitution	95.84 %	4.16 %
Betting/Wagering	77.69 %	22.31 %

Crime Category	% Completed	% Attempted
Bribery	84.74 %	15.26 %
Burglary/Breaking And Entering	93.2 %	6.8 %
Counterfeiting/Forgery	95.84 %	4.16 %
Credit Card/Atm Fraud	95.83 %	4.17 %
Destruction/Damage/Vandalism of Property	99.52 %	0.48 %
Drug Equipment Violations	99.85 %	0.15 %
Drug/Narcotic Violations	99.67 %	0.33 %
Embezzlement	99.27 %	0.73 %
Extortion/Blackmail	53.88 %	46.12 %
False Pretenses/Swindle/Confidence Game	90.39 %	9.61 %
Fondling (Incident Liberties/Child Molest)	97.31 %	2.69 %
Gambling Equipment Violations	95.16 %	4.84 %
Hacking/Computer Invasion	91.21 %	8.79 %
Human Trafficking - Commercial Sex Acts	89.62 %	10.38 %
Human Trafficking - Involuntary Servitude	92.25 %	7.75 %
Identity Theft	96.21 %	3.79 %
Impersonation	93.6 %	6.4 %
Incest	97.24 %	2.76 %
Intimidation	100 %	0 %
Justifiable Homicide	100 %	0 %
Kidnapping/Abduction	94.52 %	5.48 %
Motor Vehicle Theft	97.16 %	2.84 %
Murder/Nonnegligent Manslaughter	100 %	0 %
Negligent Manslaughter	100 %	0 %
Operating/Promoting/Assisting Gambling	92.82 %	7.18 %
Pocket-Picking	99.13 %	0.87 %
Pornography/Obscene Material	97.74 %	2.26 %
Prostitution	94.16 %	5.84 %
Purchasing Prostitution	91.98 %	8.02 %
Purse-Snatching	97.4 %	2.6 %
Rape	96.66 %	3.34 %
Robbery	90.36 %	9.64 %
Sexual Assault With An Object	97.83 %	2.17 %
Shoplifting	98.81 %	1.19 %
Simple Assault	100 %	0 %
Sodomy	97.12 %	2.88 %
Sports Tampering	85.71 %	14.29 %

Crime Category	% Completed	% Attempted
Statutory Rape	97.73 %	2.27 %
Stolen Property Offenses (Receiving, Selling, Etc.)	98.84 %	1.16 %
Theft From Building	99.38 %	0.62 %
Theft From Coin-Operated Machine Or Device	90.45 %	9.55 %
Theft From Motor Vehicle	92.91 %	7.09 %
Theft of Motor Vehicle Parts/Accessories	98.8 %	1.2 %
Weapon Law Violations	98.55 %	1.45 %
Welfare Fraud	92.45 %	7.55 %
Wire Fraud	89.44 %	10.56 %

4.1.4 Drug, alcohol, or computer use

Intoxication, mainly by alcohol, is known to be a major correlate (and cause) of crime. Drunk people commit a lot of crime (even though most drunk people never commit crime). NIBRS tries to capture this by telling us if the offender is *suspected of using* drugs (just "drugs", we don't know which drug though we could look in the Property Segment to see what drug [if any] was recovered), alcohol, or "computer equipment" which also includes cell phones. For each offense there are three variables about usage of any of these so potentially the offender could have used all three. The data doesn't get any more specific than if the offender is *suspected of using* these items. So we don't know how intoxicated they are or what they used the computer equipment for. Just that they're suspected of using it. And suspected is key. We don't know for sure if they used it. If, for example, a victim says that their mugger was drunk, NIBRS will say they're suspected of using alcohol, even though there's no definitive proof such as a blood test or breathalyzer. Unless some past variables like offense subtype where it applies to only a subset of crimes, this variable is available for every crime.

Figure 4.1 shows the distribution is suspected usage for all offenses in 2019 NIBRS. This is just from the first suspected use variable for simplicity of the graph, even though there are three variables on this topic. The most common outcome is "Not Applicable" at 87.6% of offenses. Not Applicable actually just means that the offender was not suspected of using drugs, alcohol, or computer equipment. I'm not sure why it's called that but that's how NIBRS calls "none of the above". Since Not Applicable is so common, Figure 4.2 shows the distribution when excluding that option.

Drug usage is the most common thing offenders are suspected of using, at about 66% of all crimes where they are suspected of using anything. Again, we don't know what type of drug was used, only that it wasn't alcohol. Alcohol follows at 28% while computer equipment is

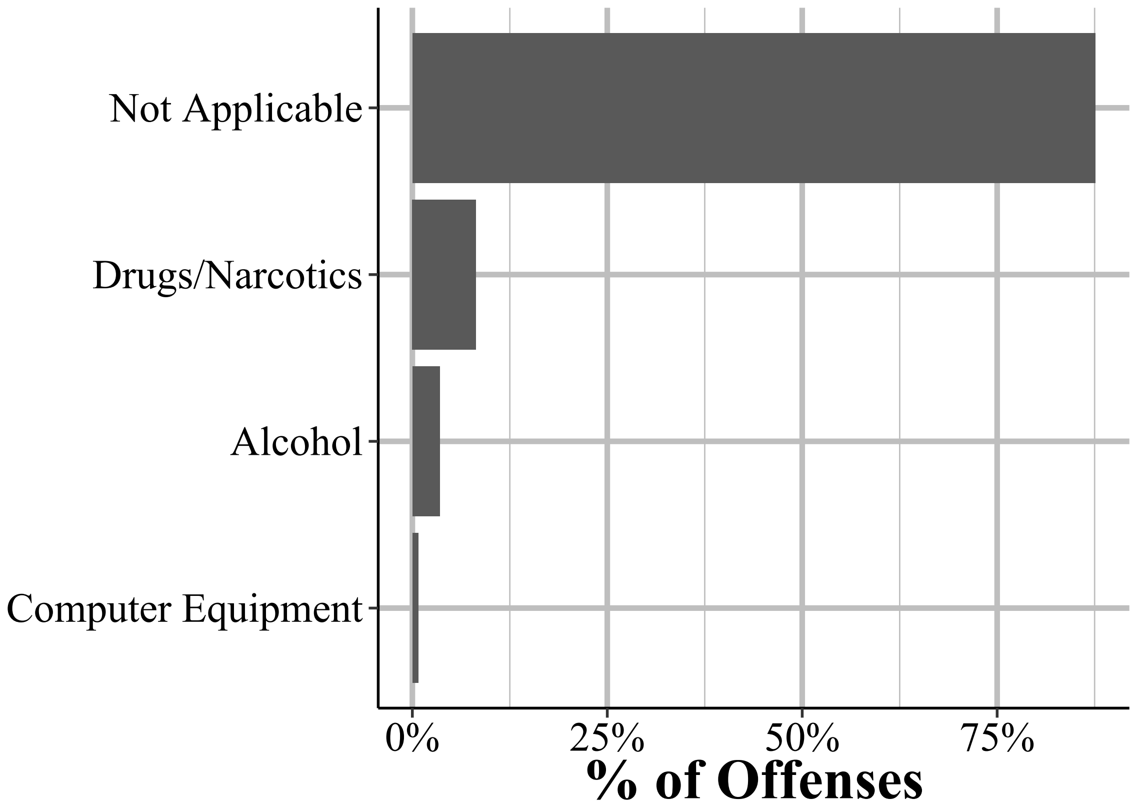

Figure 4.1: The distribution of drug, alcohol, or computer use for all offenses in 2019.

only 6.1%.

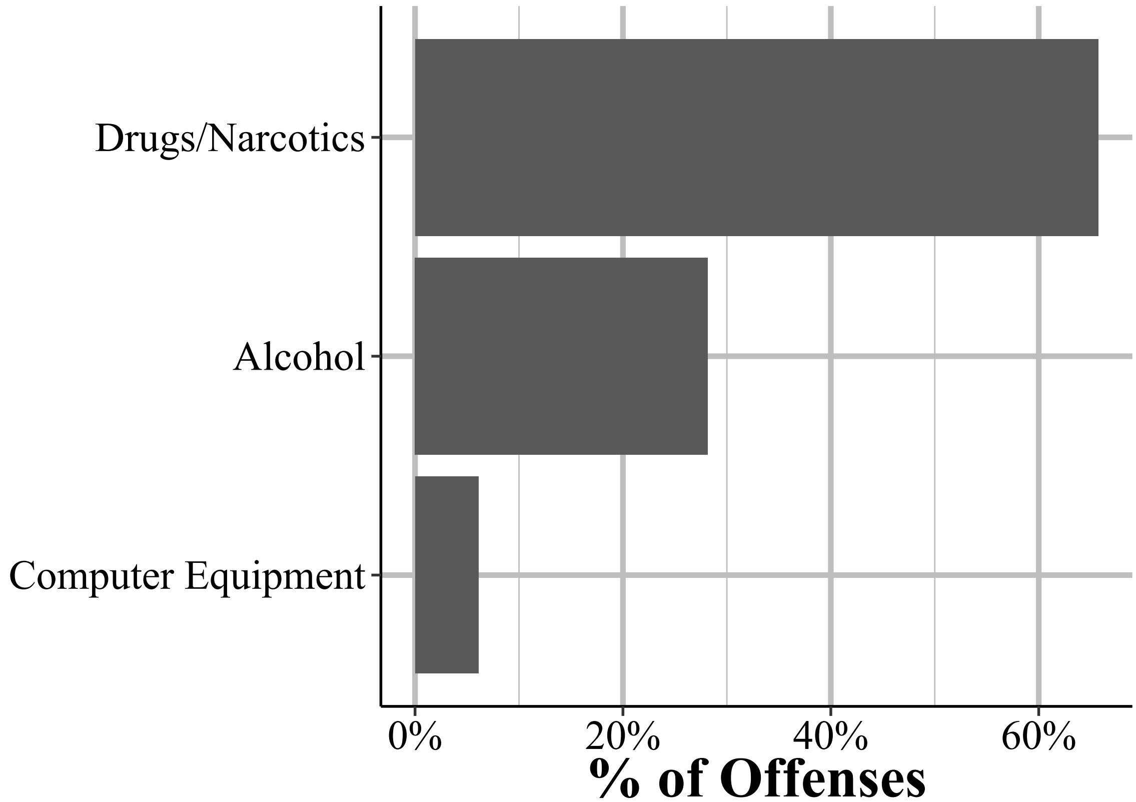

Figure 4.2: The distribution of drug, alcohol, or computer use for offenses where there was usage of one of these items. For easier viewing of how this variable is distributed, this figure excludes all offenses where there was no drug, alcohol, or computer use or the variable was NA.

4.1.5 Crime location

My own research looks a lot at how the built environment affects behavior. For example, I have a few papers looking at how outdoor lighting affects crime. Outdoor lighting largely only effects outdoor crimes (since there are generally already lights indoors) so this variable in NIBRS has been crucial for my research. NIBRS tells us where each crime happened, giving more of a type of location rather than the precise location (e.g. coordinates) where it happened. Table 4.5 shows the 46 different location types where each offense could occur, sorted by most common to least common location. The most common place for a crime to occur is in someone's own home, at 40.4% of crimes. This makes a bit of sense as people spend a lot of time at home and certain crimes, such as burglary and domestic violence, happen a lot of the victim's own home. Crimes happening on a road or alley make up the second most common location at 16% and parking lot or garage follows at 8.6%. The remaining locations only make up 5% or fewer of offense locations.

But keep in mind that some locations may be a overly specific location that fits well into a broader category that you're interested in. For example, if you care about crimes that happen in stores you'd look at “Bank/Savings and Loan”, “Restaurant”, “Bar/Nightclub” among other locations, which combined have a lot more offenses than any one individually. This is a recurring theme of NIBRS data - you have a lot of data and some of it is so specific that you need to do extra work to aggregate data into units you want.

Table 4.5: The location of crimes for all offenses reported in 2019.

Crime Location	# of Offenses	% of Offenses
Residence/Home	3,001,159	40.36%
Highway/Road/Alley	1,188,722	15.99%
Parking Lot/Garage	637,986	8.58%
Department/Discount Store	373,338	5.02%
Other/Unknown	349,477	4.70%
Convenience Store	174,952	2.35%
Grocery/Supermarket	171,702	2.31%
Commercial/Office Building	159,915	2.15%
Specialty Store (Tv, Fur, Etc.)	144,856	1.95%
Restaurant	132,347	1.78%
School - Elementary/Secondary	119,901	1.61%
Hotel/Motel/Etc.	118,313	1.59%
Service/Gas Station	103,901	1.40%
Drug Store/Doctors Office/Hospital	74,861	1.01%
Bank/Savings And Loan	67,959	0.91%
Government/Public Building	59,118	0.80%
Bar/Nightclub	55,220	0.74%
Park/Playground	54,226	0.73%
Field/Woods	40,011	0.54%
School - College/University	36,947	0.50%
Jail/Prison	35,363	0.48%
Construction Site	34,744	0.47%
Rental Storage Facility	34,630	0.47%
Shopping Mall	32,532	0.44%
School/College	31,330	0.42%
Cyberspace	29,830	0.40%
Air/Bus/Train Terminal	28,409	0.38%
Church/Synagogue/Temple	26,470	0.36%

Crime Location	# of Offenses	% of Offenses
Auto Dealership New/Used	20,643	0.28%
Liquor Store	15,237	0.20%
Industrial Site	11,044	0.15%
Gambling Facility/Casino/Race Track	9,688	0.13%
Shelter - Mission/Homeless	7,252	0.10%
Community Center	7,191	0.10%
Lake/Waterway	7,020	0.09%
Farm Facility	5,992	0.08%
Arena/Stadium/Fairgrounds/Coliseum	5,256	0.07%
Camp/Campground	5,251	0.07%
Abandoned/Condemned Structure	4,705	0.06%
Atm Separate From Bank	4,597	0.06%
Daycare Facility	4,260	0.06%
Amusement Park	3,429	0.05%
Dock/Wharf/Freight/Model Terminal	3,383	0.05%
Rest Area	2,236	0.03%
Tribal Lands	414	0.01%
Military Installation	273	0.00%
Total	7,436,090	100%

4.1.6 Weapons

Using a weapon during a crime can greatly increase the severity of the offense, as evidenced by increased sanctions for using a weapon (and particularly a gun) and the enormous amount of attention - by the media, the public, and researchers - on gun crimes.[3] Luckily, NIBRS data tells us the weapon used in certain offenses. There can be up to three different weapon types included in an offense. NIBRS data doesn't provide a weapon used for all offenses, just for the ones that they deem to the violent crimes, and thus could involve a weapon. Please note that this is the weapons used in some capacity during the crime, not necessarily to harm the victim.[4]. For example, if a gun is involved in a crime that gun may have been fired and missed the victim, fired and hit the victim, used to beat the victim, or merely brandished. From this data alone we don't know how it was used.

The list of offenses where there is data on weapon usage is below:

[3] The 2021 NYC mayor election, which as of this writing has not been completed yet, seems to hinge on an increase of gun violence in the city.

[4] The Victim Segment does have data on victim injuries though it doesn't say which weapon caused the injuries

- Aggravated Assault
- Extortion/Blackmail
- Fondling (Incident Liberties/Child Molest)
- Human Trafficking - Commercial Sex Acts
- Human Trafficking - Involuntary Servitude
- Justifiable Homicide
- Kidnapping/Abduction
- Murder/Nonnegligent Manslaughter
- Negligent Manslaughter
- Rape
- Robbery
- Sexual Assault With An Object
- Simple Assault
- Sodomy
- Weapon Law Violations

Table 4.6 shows the breakdown in the weapons used in the above offenses. There were about 1.6 million offenses reported in NIBRS in 2019 that could have used a weapon. The most common weapon used was only the offender's body at 57.8% of offenses. The "personal weapons (hands, feet, teeth, etc.) basically means that the offender used or threatened force but wasn't carrying a weapon. So this includes things like punching, kicking, biting, wrestling, and anything you may see in a boxing or MMA match. Strangulation can be done without any weapons but as strangulation is its own weapon, it is not included in the"personal weapons" category. The next most common group is the offender doesn't have any weapons, and doesn't use their body as a weapon, at a little over 10% of offenses, following by the offender using a handgun in 9.4% of offenses.

"Other" is the next most common category which just means any weapon not already included in the weapon categories. Knife/cutting instrument makes up 4.5% of offenses are is a rather broad category, composed of anything that could cut or pierce someone's body. The most likely weapon in this category is a knife, but can extend to rarer items like broken glass or a sword. The remaining weapon groups are rarer than 4% of offenses, but given that NIBRS covers so many crimes these weapons still occur in hundreds or thousands of cases.

Table 4.6: The weapon used by an offender in the crime for all offenses reported in 2019. The use means that it was part of the crime though may not have been physically discharged. For example, pointing a gun at someone even without firing the gun is still using it.

Weapon Used	# of Offenses	% of Offenses
Personal Weapons (Hands, Feet, Teeth, Etc.)	918,134	57.84%
None	161,623	10.18%
Handgun	149,146	9.40%
Other	90,385	5.69%
Knife/Cutting Instrument (Ice Pick, Screwdriver, Ax, Etc.)	71,311	4.49%
Firearm (Type Not Stated)	58,920	3.71%
Unknown	57,430	3.62%
Blunt Object (Club, Hammer, Etc.)	36,536	2.30%
Motor Vehicle	17,278	1.09%
Rifle	7,696	0.48%
Shotgun	5,896	0.37%
Asphyxiation (By Drowning, Strangulation, Suffocation, Gas, Etc.)	4,984	0.31%
Other Firearm	4,633	0.29%
Drugs/Narcotics/Sleeping Pills	1,239	0.08%
Explosives	949	0.06%
Fire/Incendiary Device	715	0.05%
Poison (Include Gas)	475	0.03%
Total	1,587,350	100%

4.1.7 Automatic weapons

When the weapon involved was a firearm there is a variable which indicates that the firearm was fully automatic. To be clear, this means that when you pull the trigger once the gun will fire multiple bullets. Semi-automatic firearms are **not** automatic firearms. Of course, saying a gun is fully automatic requires either the policing seizing the gun or the gun being fired (and for witnesses to accurately determine that it is fully automatic). Since most crimes are never solved (and even those that lead to an arrest may not lead to the gun being seized - though some guns are seized even without an arrest, such as if the gun is left at the crime scene) and most gun crimes don't actually involve the gun being fired, this variable is likely very imprecise. Still, Figure 4.3 shows the percent of firearms used in offenses in 2019 that are reported to be fully automatic. Even though there can be up to three weapons used in an offense, this figure only looks at the first weapon. The most common guns to be automatic are rifles and handguns, both with about 4.5% of all uses being of an automatic weapon. The remaining categories are all under 3% of uses.

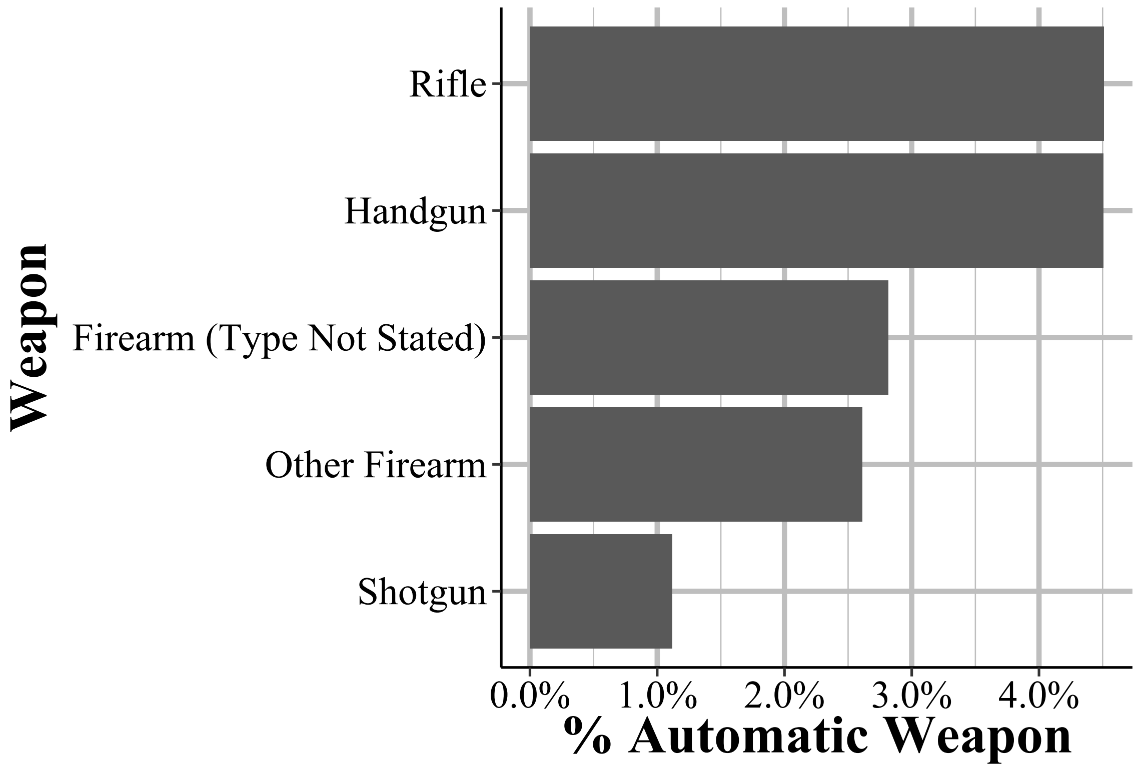

Figure 4.3: The percent of firearms used that were fully automatic, for all offenses in 2019.

4.1.8 Burglary info

For burglary offenses specifically there are two variables that provide a little more information on the offense. The first variable is the number of “premises” that the burglar entered. This is only available when the location for the offense is either hotel/motel or a rental storage facility. So the “premise” can really be thought of as a room in the building, not that they break into multiple hotels. Figure 4.4 shows the breakdown in the number of premises entered during a burglary incident. The graph is capped at 10 or more but in 2019 the highest number was 99 buildings entered, which is the maximum value the police can enter, so in reality it may have been higher. The vast majority of hotel/motel and storage facility burglaries only have one room entered, with 85% of these burglaries only being on a single room. This declines enormously to 4.7% burglarizing two rooms and then nearly halves to 2.5% burglarizing three rooms. This trend continues as the number of rooms increase.

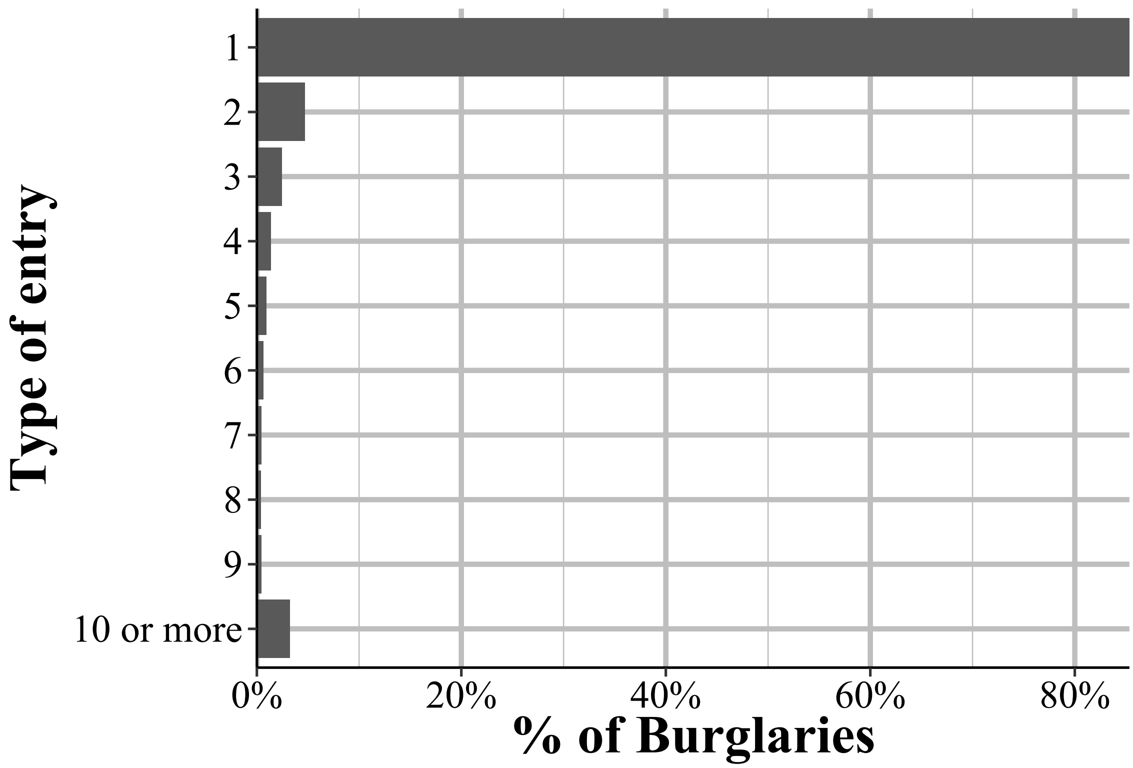

Figure 4.4: The distribution in the number of premises entered during burglaries. This info is only available for burglaries in a hotel/motel or rental storage facilities.

The second variable, and one where there is data from every burglary reported regardless of location, says whether the burglar entered the building forcibly or not. A burglary without force is one when the burglary *only* enters through unlocked doors or windows. The *only* means that if they entered through an unlocked door or window and then forced open another door or window, the entire burglary is classified as forcible entry. forcible entry is any when the burglar has to access through any means a locked door or window. This is very broad and includes actions ranging from breaking the window - which people generally think of when it comes to forcible entry - to less obvious uses of force like picking the lock or even using a passcard (e.g. a hotel room card) to unlock the door. The FBI also includes when a burglar enters a building legally and then stays past their allowed time (e.g. walk into a store and hide somewhere until past closing time).

Figure 4.5 shows the breakdown in burglaries by type of entry. The majority of burglaries, 57.6%, use force at some point in the burglary. 42.4% don't use force at all. There's no option for "unknown" if force was used so my guess is that when in doubt - that is, when there's no evidence of force - the police report that no force is used.

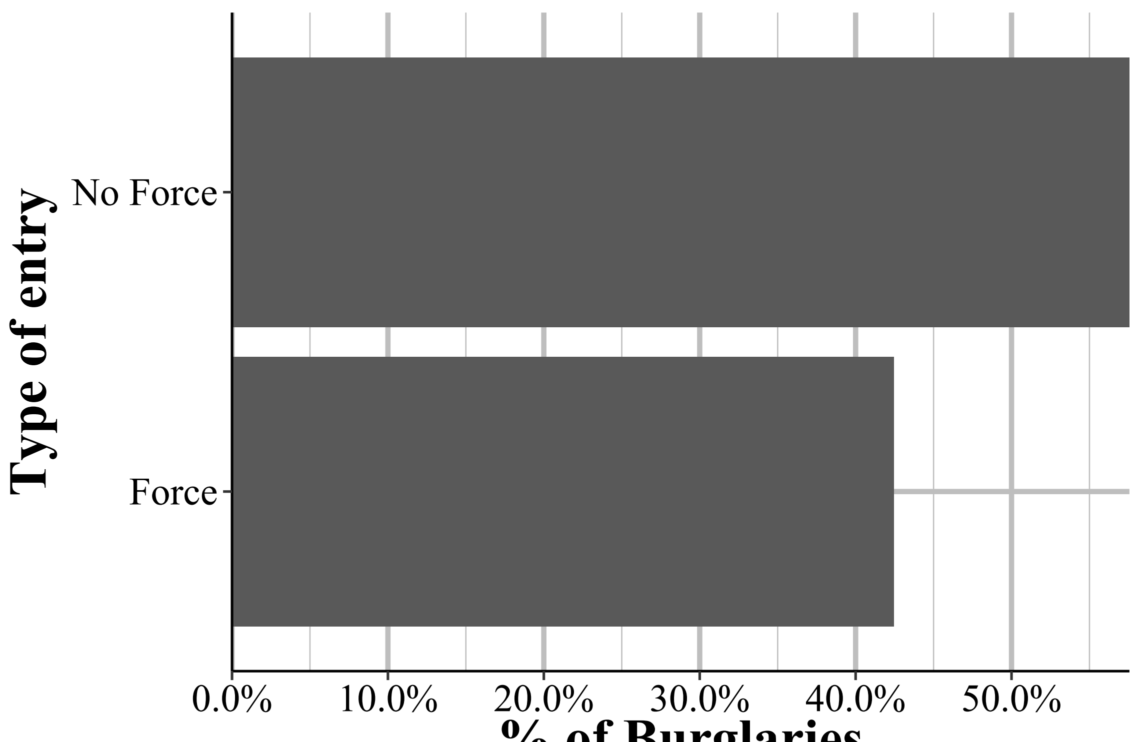

Figure 4.5: The percent of burglaries reported in 2019 where the burglary entered the structure forcibly or non-forcibly.

4.1.9 Hate crime indicator (bias motivation)

For each offense NIBRS indicates whether it had a bias motivation, which is NIBRS way of saying if it was a hate crime or not. For the police to classify an incident as a hate crime, and to assign a particular bias motivation, the police must have some evidence that the crime was motivated by hate. The victim saying that the crime is a hate crime alone is not sufficient - though if large portions of the victim's community believe that the crime is a hate crime, this is a factor in the police's assessment. The evidence required is not major, it includes things as explicit as racial slurs said during an incident and less obvious factors like the victim is celebrating their community (e.g. attending a holiday event) or the crime occurring on an important holiday for that community (e.g. Martin Luther King Day, religious holidays). The FBI also encourages police to consider the totality of the evidence even if none alone strongly suggests that the crime was a hate crime in making their determination about whether the incident was a hate crime or not.

This also means that many (likely most) hate crimes will not be recorded as hate crimes since there is no evidence that the crime is motivated by hate. For example, if a man committed a crime against Asian people for crimes because they are Asian, that would in reality be a hate crime. However, if the offender does not say anything anti-Asian to the victim, which is the mostly likely thing to indicate that this is a hate crime, the crime would not likely be recorded as a hate crime. For example, at the time of this writing (Spring, 2021), there are numerous media reports discussing an increase in anti-Asian hate crimes as a result of racism relating to the pandemic.48 This data would likely undercount both anti-Asian behavior and anti-Asian hate crimes. First, if someone walked to an Asian person and called them an anti-Asian slur, that is clearly a hateful act and would be classified as a hate crime under some organization's collections methods. However, as hateful as this incident is, this alone would not be classified as a hate crime in this dataset as a slur is not a crime. If accompanied by other criminal behavior, or if it continues to the point where it can be considered intimidation, it would then be classified as a hate crime. Second, crimes against Asian victims that are in fact hate crimes, but have no evidence that they are hate crimes would not be classified as hate crimes in this data.

Bias motivation is based on the offender's perceptions of the victim so even if they are incorrect in who their victim is, if they intended to target someone for their perceived group membership, that is still a hate crime. For example, if a person assaults a man because they think he is gay, that is a hate crime because the assault was motivated by hate towards gay people. Whether the victim is actually gay or not is not relevant - the offender perceived him to be gay so it is an anti-gay hate crime. To make this even more complicated, the offender must have committed the crime because they are motivated, at least to some degree, by their bias against the victim. Being biased against the victim but targeting them for some other reason means that the crime is not a hate crime.

Table 4.7 shows the percent of all offenses in 2019 that were classified with or without a bias motivation. Nearly all offenses - 99.14% - are without a bias motivation meaning that they are not considered hate crimes. This still leaves a large 63,876 offenses classified as hate crimes.

Table 4.7: The number and percent of offenders that had a bias motivation or not for all offenses reported in 2019.

Bias Motivation	# of Offenses	% of Offenses
No Bias Motivation	7,372,214	99.14%
Bias Motivation	63,876	0.86%
Total	7,436,090	100%

Table 4.8 shows the breakdown in the bias motivation of hate crimes, for all incidents where the crime is considered a hate crime. The most common bias motivation is anti-Black, which accounts for 29% of all hate crimes in the data. This is followed by anti-White at 12.7% and "anti-male homosexual (gay)" at almost 8% of crimes. The only other biases that make up more than 5% of hate crimes are anti-Hispanic and anti-Jewish.[5]

Some of these groups are also subsets of larger groups. For example, anti-Muslim, anti-Arab, and anti-Sikh (while Sikhs are not Muslim or Arabic, some Sikhs have been targeted by people who incorrectly believe that they are) are probably all the same bias motivation. Likewise, attacks on LGBT people are in multiple categories, which allows for a more detailed understanding of these hate crimes but requires aggregation to look at them as a group. While this aggregation is easy enough to do, accidentally missing any of the subcategories could vastly undercount offenses against the larger category.

[5]Looking at the raw percents is a rather naive measure as it assumes that all groups have equal risk of hate crimes. Certain groups, such as Jews and transgender people, make up a relatively small share of the percent of hate crimes but when considering their percent of the overall population (itself only a slightly better measure as even total population doesn't account for true opportunity to be victimized) are victimized at much higher rates than many other groups.

Table 4.8: The bias motivation (i.e. if hate crime or not and what type of hate crime) for all offenses reported in 2019 that were classified as hate crimes. For easier viewing of how hate crimes are broken down, this excludes all offenses where there was no bias motivation.

Bias Motivation	# of Offenses	% of Offenses
Anti-Black	1,309	28.99%
Anti-White	573	12.69%
Anti-Male Homosexual (Gay)	357	7.91%
Anti-Hispanic	310	6.86%
Anti-Jewish	249	5.51%
Anti-Homosexual (Both Gay And Lesbian)	213	4.72%
Anti-Other Ethnicity/National Origin	210	4.65%
Anti-American Indian Or Alaskan Native	125	2.77%
Anti-Mental Disability	107	2.37%
Anti-Asian/Pacific Islander	101	2.24%
Anti-Multi-Racial Group	97	2.15%
Anti-Islamic (Muslim)	86	1.90%
Anti-Female Homosexual (Lesbian)	70	1.55%
Anti-Transgender	68	1.51%
Anti-Sikh	66	1.46%
Anti-Arab	64	1.42%
Anti-Other Religion	63	1.40%
Anti-Gender Non-Conforming	59	1.31%
Anti-Female	51	1.13%
Anti-Eastern Orthodox (Greek, Russian, Etc.)	50	1.11%
Anti-Physical Disability	44	0.97%
Anti-Catholic	42	0.93%
Anti-Other Christian	40	0.89%
Anti-Native Hawaiian Or Other Pacific Islander	30	0.66%
Anti-Protestant	26	0.58%
Anti-Multi-Religious Group	25	0.55%
Anti-Bisexual	22	0.49%
Anti-Male	17	0.38%
Anti-Heterosexual	12	0.27%
Anti-Mormon	7	0.16%
Anti-Hindu	7	0.16%

Bias Motivation	# of Offenses	% of Offenses
Anti-Buddhist	6	0.13%
Anti-Atheism/Agnosticism	5	0.11%
Anti-Jehovahs Witness	5	0.11%
Total	4,516	100%

Chapter 5

Offender Segment

As might be expected, the Offender Segment provides information about who the offender is for each incident, though this is limited to only demographic variables. So we know the age, sex, and race of each offender but nothing else. This means that important variables such as criminal history, ethnicity, socioeconomic status, and motive are missing. In the Victim Segment we learn about the relationship between the victim and offender, and in the Offense Segment we learn which weapon (if any) the offender used. So there is some other data on the offender in other segments but it's quite limited. This data has one row per offender so incidents with multiple offenders have multiple rows. In cases where there is no information about the offender there will be a single row where all of the offender variables will be "unknown." In these cases having a single row for the offender is merely a placeholder and doesn't necessarily mean that there was only one offender for that incident. However, there's no indicator for when this is a placeholder and when there was actually one offender but whose demographic info is unknown.

5.1 Important variables

The Offender Segment is the sparsest of the available segments, provides only three new variables that are about the offender's demographics. It also includes the standard set of variables: the agency ORI, the incident number, the state the agency is in, and the incident date (though we'd need to check the Administrative Segment to see if this is actually the incident date or the report date).

5.1.1 Demographics

There are three demographics variables included in this data: the offenders' age, sex, and race. Please note that what we have here are not unique offenders as someone may be involved in multiple crimes. So be cautious when trying to compare this with some base rate such as percent of people of each age/sex/race in a population.

5.1.1.1 Age

The age variable is the suspected age of the offender. This is presented to us as whole years though agencies can input an age range if they don't know the exact age (e.g. age 20-29) and the FBI will convert that to an exact number by averaging it for us. So if the police say the offender is ages 20-29 (since they don't know for sure), we only see that the offender is 24 years old since the FBI (for some reason) rounds down. This value is top-coded to 99 years old with everyone 99 years or older being set as "over 98 years old." Figure 5.1 shows the distribution of offender ages for all known offenders in the 2019 NIBRS data. About 14% of offenders have an unknown age and are excluded from the figure.

This figure shows the percent of offenders at each age that make up known offenders in the data. If you're familiar with research on the age-crime curve, which says that crimes peak in the late teens and then rapidly decrease, this essentially replicates those findings. There are some differences between this figure and past age-crime research as crime peaks later here, in the mid-20s (the most common age is 25), but the general trend of crime being largely a "young person" phenomenon holds consistent. This also depends on exactly which crime occurs as different crimes have different age-crime trends, so you'll need to merge this segment with the Offense Segment to be able to subset by crime committed.

The spike you see at the very end of the data is due to the data maxing out possible individual ages at 98, so anyone older is grouped together. There's also a spike at age 1 - and other offenders at very young ages - , the youngest possible age. Surely very young children aren't going around attacking people, so is this a data error? Potentially yes. But it could actually be real as there are very rare cases where children harm or kill someone while playing with a gun and are included in the data. These aren't crimes as we conventionally think of them - and won't be criminally charged - but are still included in the data. However, the bulk of this, especially for age 0, is likely just a data error or the police entering age 0 instead of saying that the age is unknown.

Another indicator of guesses about age is that three of the five most common ages are 25, 30, and 20 years old. People tend to like multiples of five when making estimates, so these indicate that someone (the victim or the officer) probably didn't know the exact age and so guessed the age.

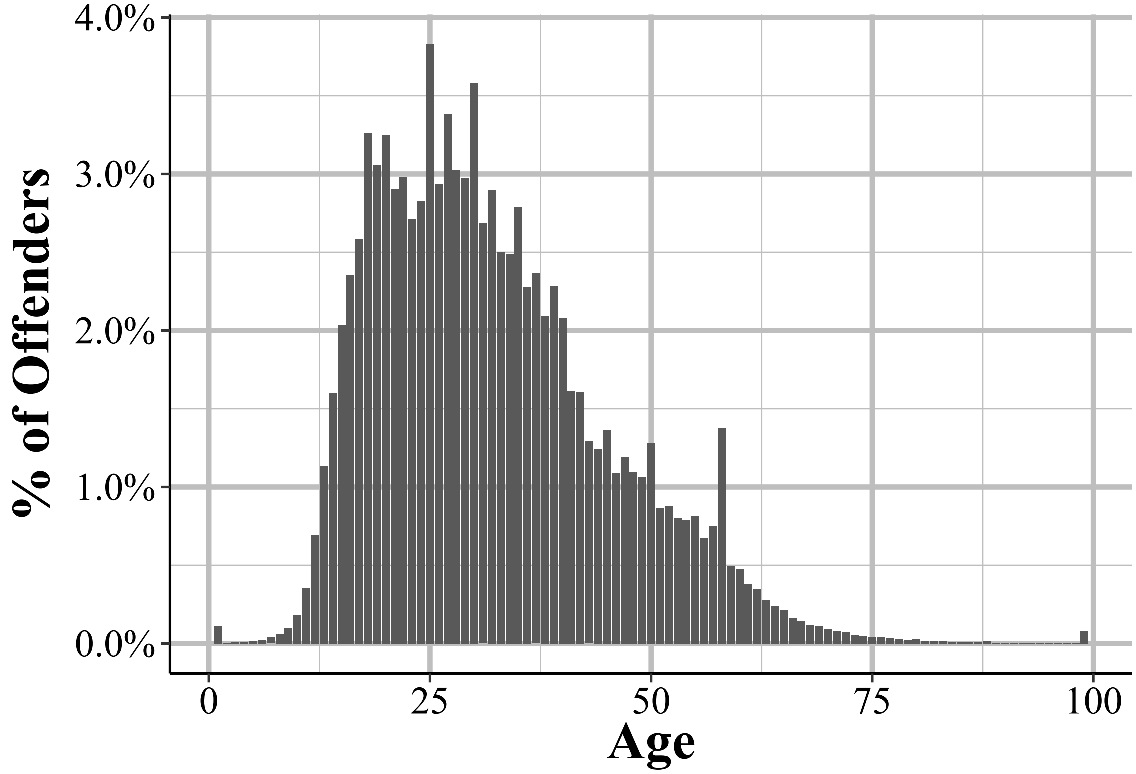

Figure 5.1: The age of all offenders reported in the 2019 NIBRS data. Approximately 39 percent of offenders have an unknown age are an not shown in the figure.

5.1.1.2 Sex

The second offender demographic variable available is the offender's sex with male and female being the only available sexes. There is no option for transgender or any other identify. Other than arrestees, where police could (though we don't know if they do) use their identification (e.g. drivers license) to determine their sex, this is the perceived sex of the offender. Figure 5.2 shows the distribution of offenders by sex. The most common sex is male, which is consistent with the literature on who commits crime. About 45% of all offenders were male. Female offenders make up nearly 19% of offenders. Over a third - 35.9% - of offenders have an unknown sex. Considering that when nothing is known about offenders (including even how many offenders there are) this data includes a single row with "unknown" for all demographic variables, this is actually an undercount of offenders who have unknown sex.

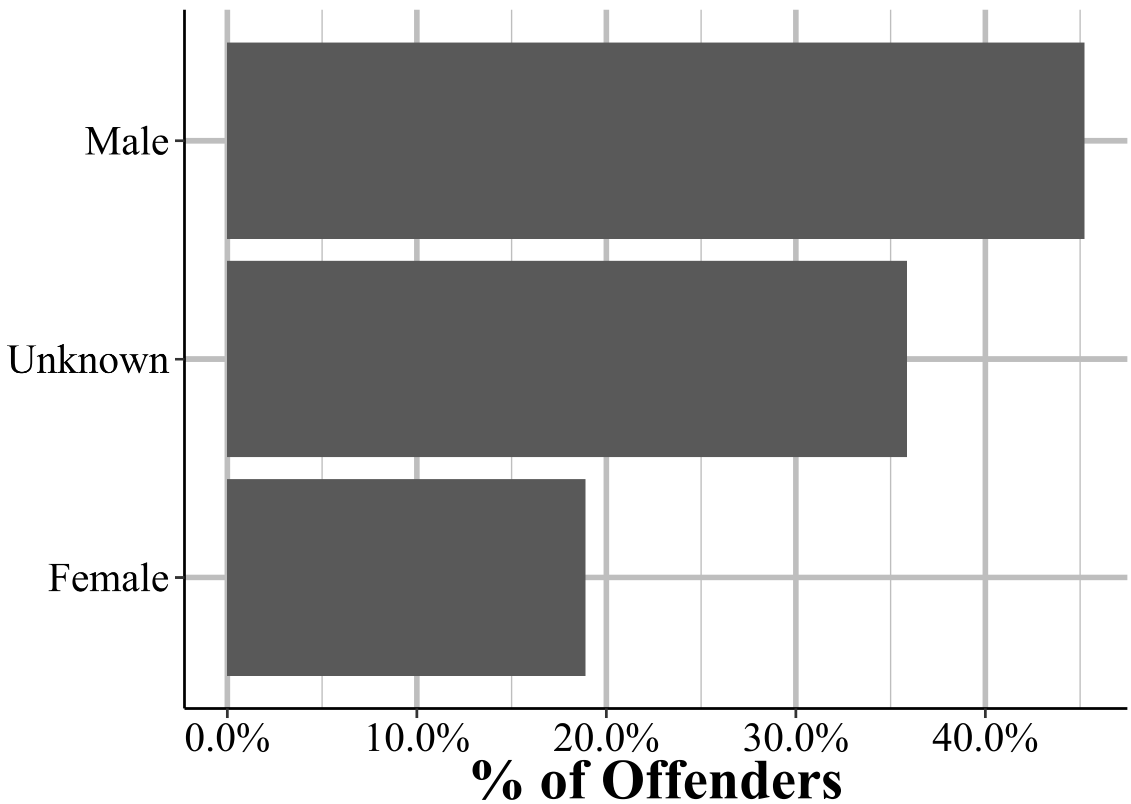

Figure 5.2: The sex of all offenders reported in the 2019 NIBRS data.

5.1.1.3 Race

The final variable available is the race of the offender. The only possible races are White, Black, American Indian/Alaskan Native, Asian/Pacific Islander, and Native Hawaiian/Other Pacific Islander. These categories are mutually exclusive so people cannot be labeled as mixed race, they must be put into one of the categories. Other than offenders who were arrested, and thus police can clearly see them and potentially ask them what race they are, this variable is likely a rough estimate of a person's race.

Figure 5.3 shows the breakdown in offender races for every offender in the 2019 data. The most common offender race is Unknown, with about 38.5% of offenders not having a known race. This 38.5% of actually an undercount as in cases where the agency doesn't known anything about the offenders they'll put down a single offender with "unknown" for every demographics variable. So there could potentially be multiple offenders represented when there is a row with a unknown offender race. This is also dependent on the type of crimes committed and when they are committed. For example, a daytime robbery would likely have a known offender race as the victim can clearly see (complexities about identifying people's race aside) the race of the robber. A daytime burglary where no one is home would likely have an unknown offender race and there'd be no witnesses.[1] Likewise, a robbery at night could have an unknown offender race because the darkness makes it harder to identify people.

The next most common offender race is White at 38.7% followed by Black at 22.1%. The remaining races make up only a little over 1.5% of offenders, with American Indian/Alaskan Native at 0.77%, Asian at 0.63%, and Native Hawaiian/Other Pacific Islander at 0.24%.

[1] When NIBRS 2020 data comes out it'll be interesting to see how people staying at home effects the amount of unknown offender info.

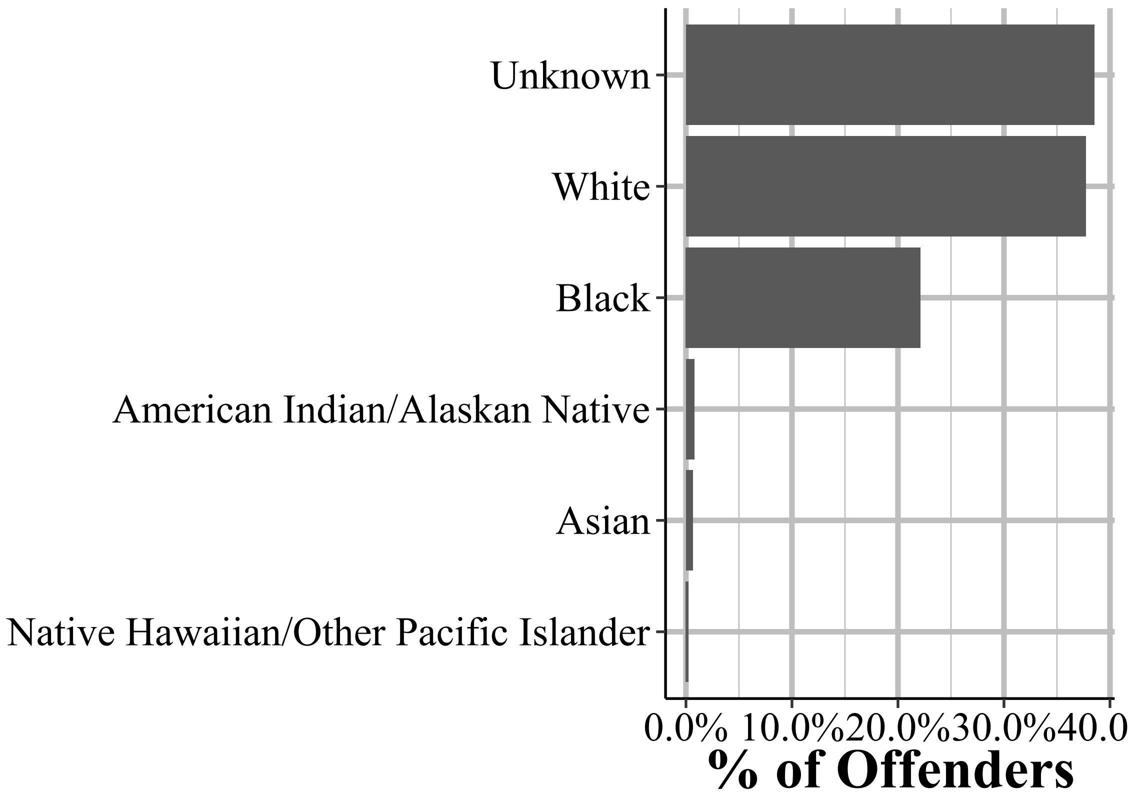

Figure 5.3: The race of all offenders reported in the 2019 NIBRS data.

Chapter 6

Victim Segment

The Victim Segment provides data at the victim-level and includes information about who the victim is and their relationship to offenders. This data tells us what "type" of victim it is with the type meaning if they are a police officer, a civilian ("Individual" and basically any person who isn't a police officer), a business, the government, etc. It also includes the standard demographics variables in other segments - age, race, sex, ethnicity - as well as whether the victim is a resident (i.e. do they live there?) of the jurisdiction where they were victimized. We also learn from this data what types of injuries (if any) the victim suffered as a result of the crime. This is limited to physical injuries - excluding important outcomes such as mental duress or PTSD - but allows for a much better measure of harm from crime than simply assuming (or using past studies that tend to be old and only look at the cost of crime) what harm comes from certain offenses. There seven possible injury types (including no injury at all) and victims can report up to five of these injuries so we have a fairly detailed measure of victim injury.

One highly interesting variable is the relationship between the victim and the offender (for up to 10 offenders). This includes, for example, if the victim was the offender's wife, their child, employee, or if the stranger was unknown to them, with 27 total possible relationship categories. You can use this to determine which incidents were crimes by strangers, identify domestic violence, or simply learn who tends to commit crimes against certain types of victims. This variable is only available when the victim is a police officer or an "individual." This makes some sense though there could actually be cases where non-human victims (e.g. businesses, religious organizations) do have a relationship with the offender such as an employee stealing from a store. Related to the victim-offender relationship, this segment provides a bit of information about the motive for the crime. For aggravated assaults and homicides, there is a variable with the "circumstance" of the offense which is essentially the reason why the crime occurred. For example, possible circumstances are arguments between people, hunting accidents, child playing with weapon, and mercy killings.

6.1 Important variables

6.1.1 Crime category

Table 6.1: The number and percent of crimes committed against each victim. For victims with multiple crimes committed against them, this shows the most serious crime.

Crime Category	# of Victims	% of Victims
Simple Assault	1,062,299	14.31%
All Other Larceny	820,257	11.05%
Drug/Narcotic Violations	710,870	9.58%
Destruction/Damage/Vandalism of Property	681,090	9.17%
Theft From Motor Vehicle	614,794	8.28%
Burglary/Breaking And Entering	551,161	7.42%
Shoplifting	498,696	6.72%
Aggravated Assault	336,840	4.54%
Intimidation	313,809	4.23%
Motor Vehicle Theft	312,692	4.21%
Theft From Building	212,308	2.86%
False Pretenses/Swindle/Confidence Game	190,933	2.57%
Robbery	125,835	1.69%
Credit Card/Atm Fraud	125,001	1.68%
Counterfeiting/Forgery	111,637	1.50%
Theft of Motor Vehicle Parts/Accessories	111,376	1.50%
Drug Equipment Violations	108,595	1.46%
Weapon Law Violations	95,283	1.28%
Identity Theft	63,639	0.86%
Stolen Property Offenses (Receiving, Selling, Etc.)	55,445	0.75%
Impersonation	55,367	0.75%
Fondling (Incident Liberties/Child Molest)	50,697	0.68%
Rape	47,233	0.64%
Embezzlement	24,310	0.33%
Pornography/Obscene Material	19,703	0.27%
Arson	17,406	0.23%
Wire Fraud	15,162	0.20%
Sodomy	11,027	0.15%

Crime Category	# of Victims	% of Victims
Kidnapping/Abduction	10,066	0.14%
Animal Cruelty	9,807	0.13%
Pocket-Picking	9,712	0.13%
Purse-Snatching	7,595	0.10%
Murder/Nonnegligent Manslaughter	6,590	0.09%
Prostitution	6,464	0.09%
Statutory Rape	5,591	0.08%
Extortion/Blackmail	5,334	0.07%
Theft From Coin-Operated Machine Or Device	4,791	0.06%
Sexual Assault With An Object	4,480	0.06%
Hacking/Computer Invasion	2,162	0.03%
Assisting Or Promoting Prostitution	1,787	0.02%
Purchasing Prostitution	1,061	0.01%
Human Trafficking - Commercial Sex Acts	854	0.01%
Incest	849	0.01%
Welfare Fraud	785	0.01%
Negligent Manslaughter	597	0.01%
Operating/Promoting/Assisting Gambling	567	0.01%
Bribery	449	0.01%
Justifiable Homicide	308	0.00%
Betting/Wagering	249	0.00%
Gambling Equipment Violations	219	0.00%
Human Trafficking - Involuntary Servitude	176	0.00%
Sports Tampering	5	0.00%
Total	7,423,963	100%

6.1.2 Victim type

6.1.3 Injury

6.1.4 Relationship to offender

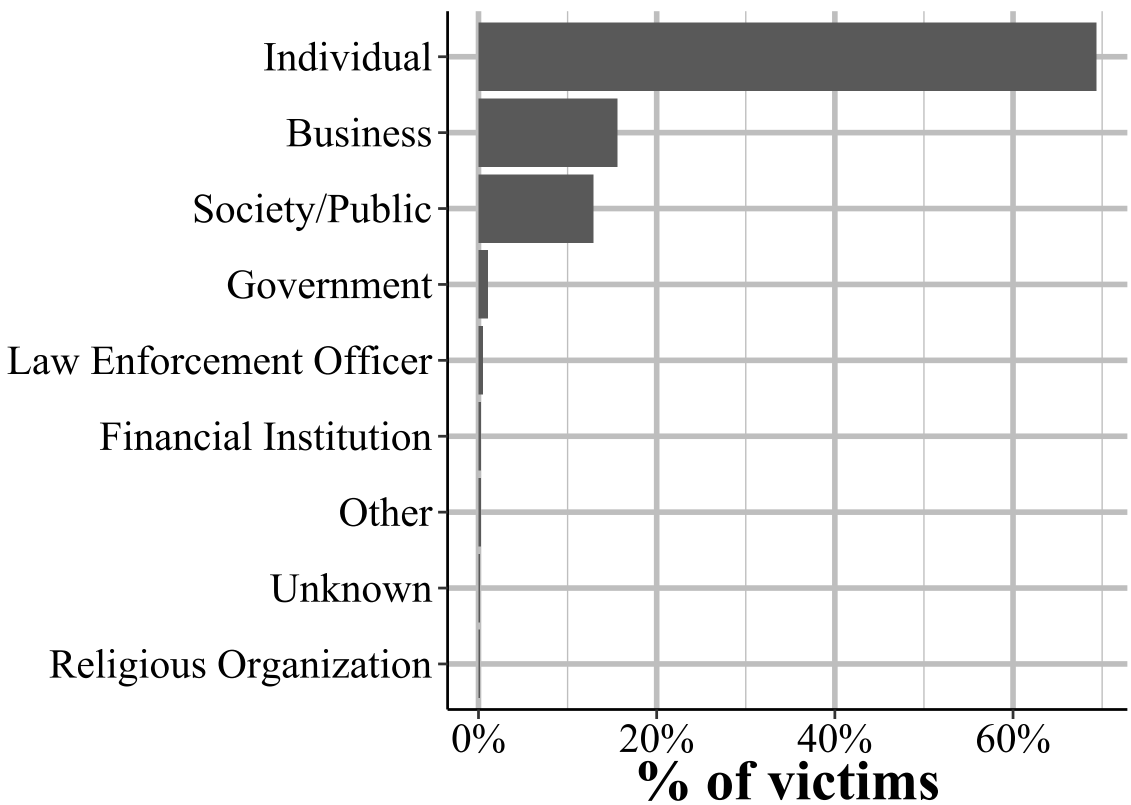

Figure 6.1: The distribution of the type of victim. Victim types are mutually exclusive.

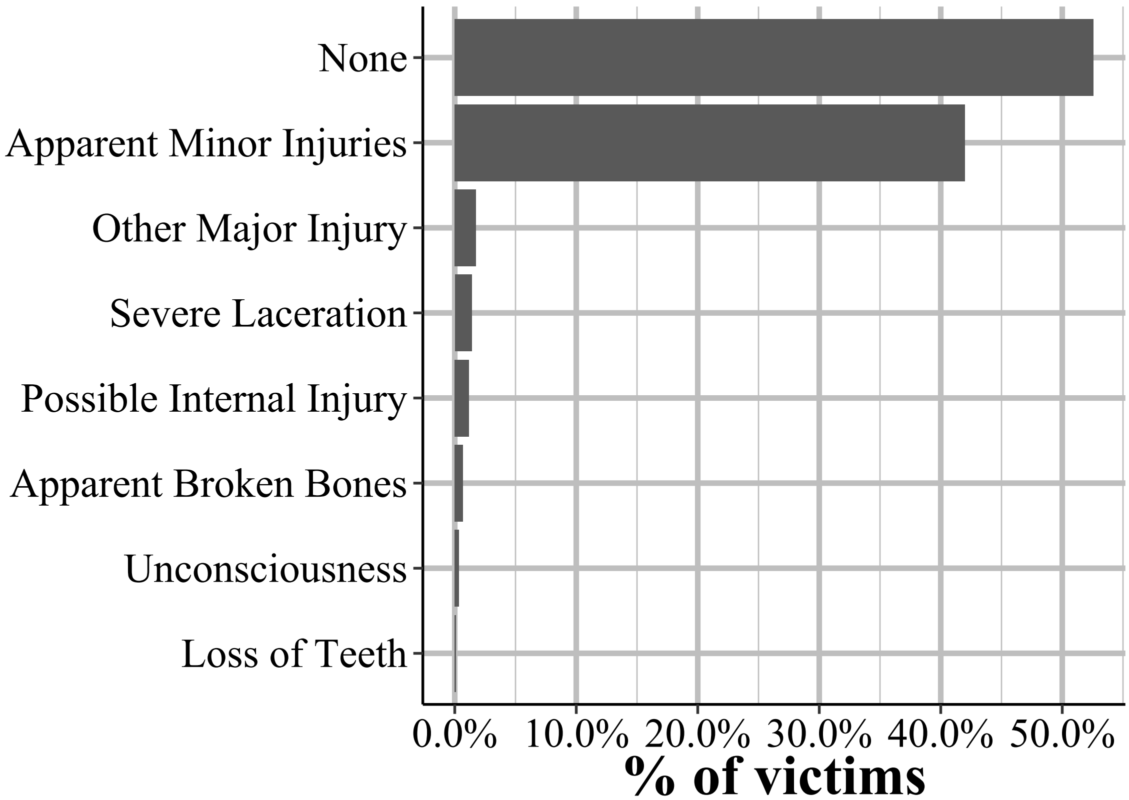

Figure 6.2: The distribution of the injury sustained by the victim. Only individual and law enforcement officer victims have this variable available.

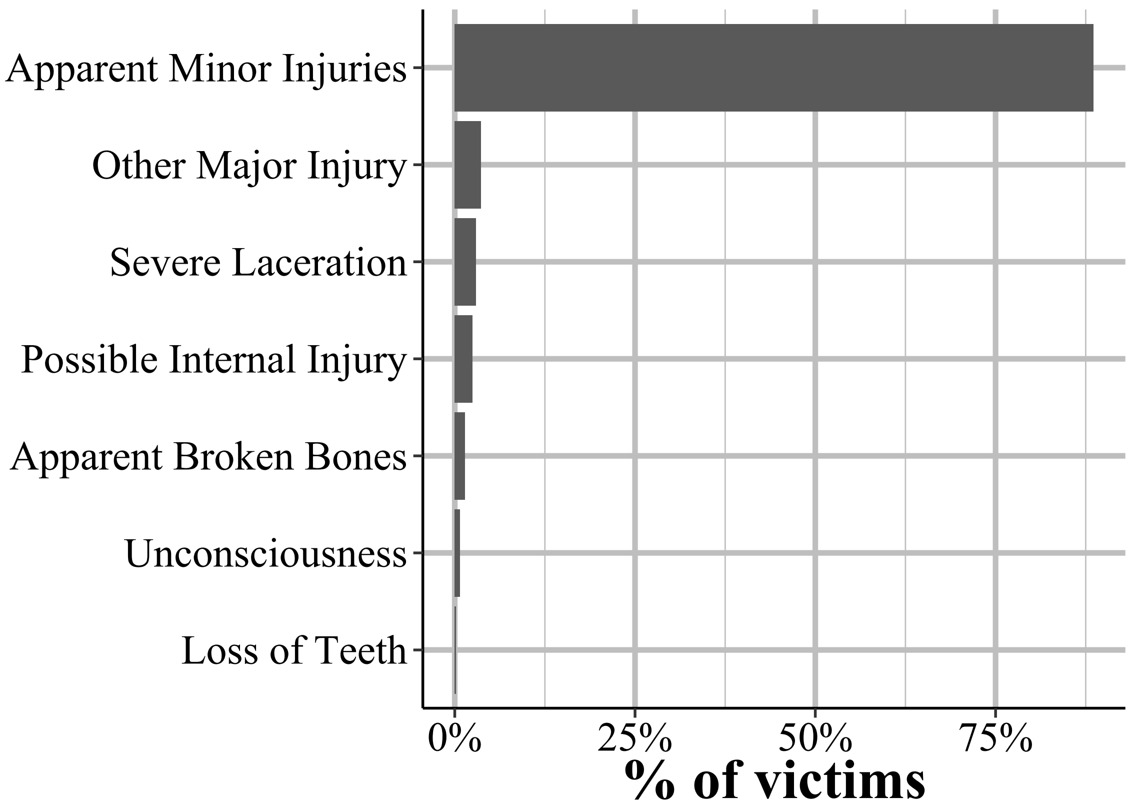

Figure 6.3: The distribution of the injury sustained by the victim for those who had an injury other than 'none'.

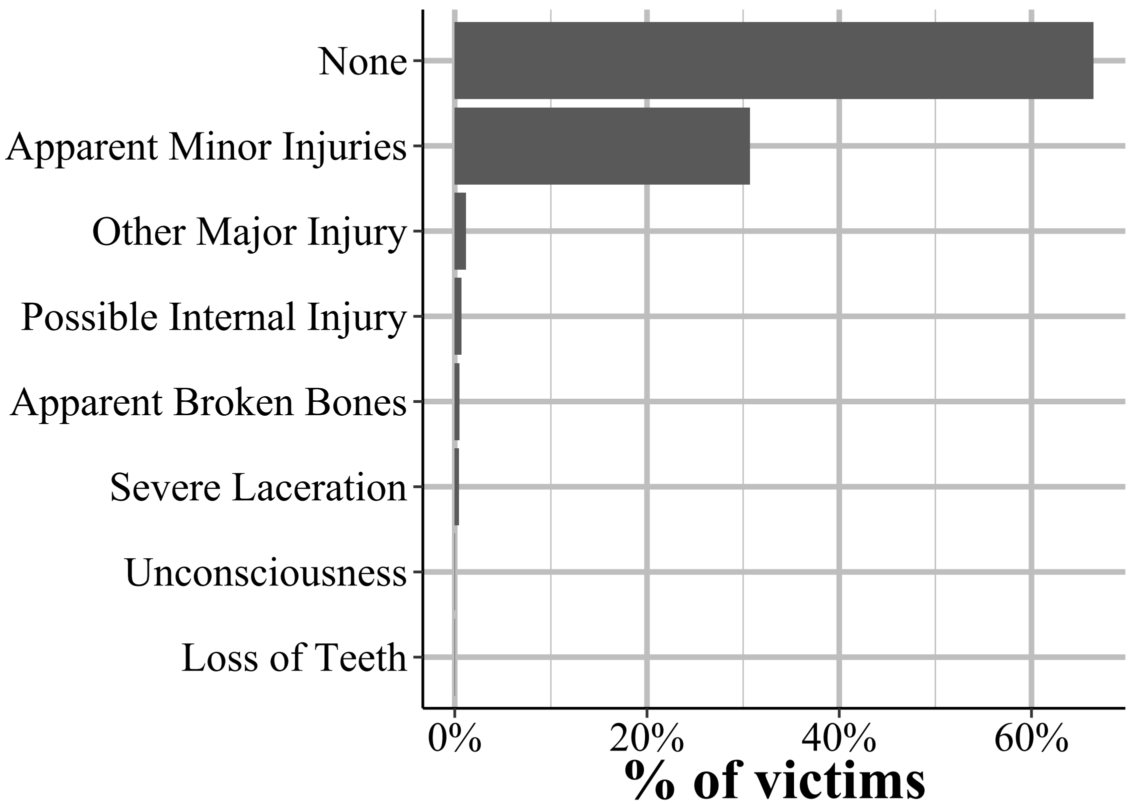

Figure 6.4: The distribution of the injury sustained by the victim for law enforcement officer victims

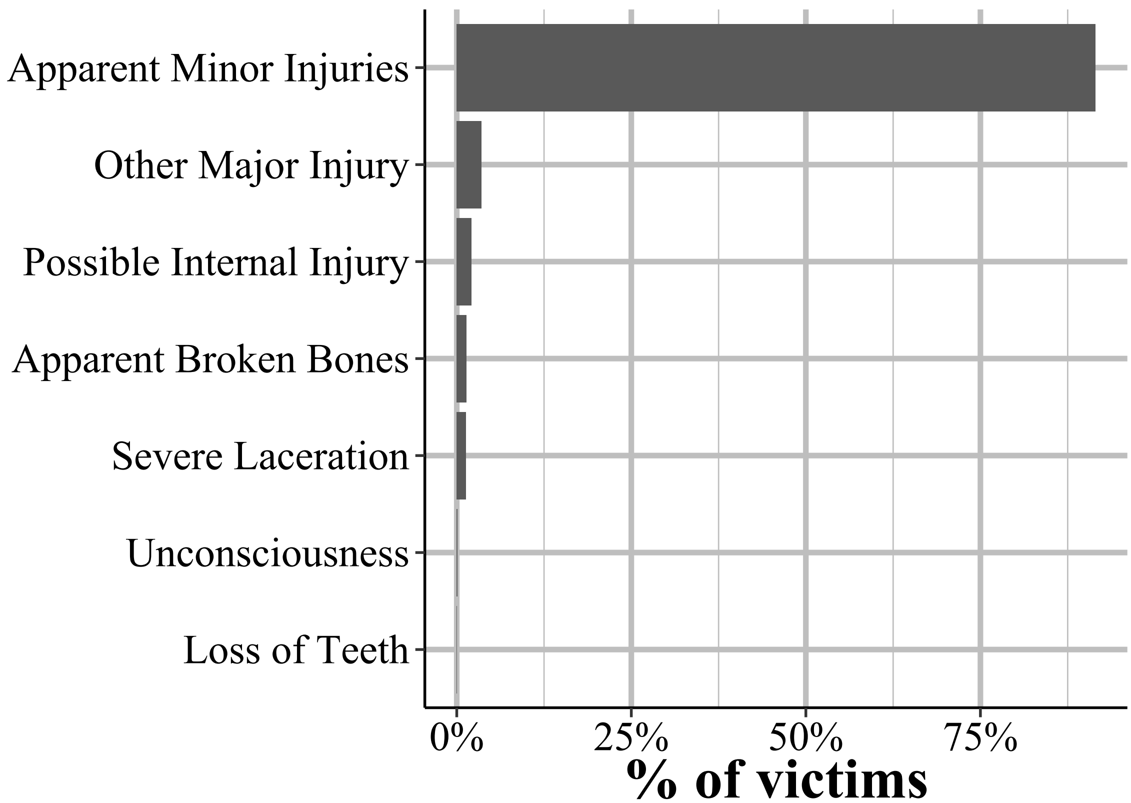

Figure 6.5: The distribution of the injury sustained by the victim for law enforcement officer victims excluding those who had no injury at all.

Table 6.2: The distribution of the relationship between the victim and the offender. Only individual and law enforcement officer victims have this variable available.

Crime Category	# of Victims	% of Victims
Relationship Unknown	436,858	19.75%
Victim Was Boyfriend/Girlfriend	317,314	14.35%
Victim Was Acquaintance	286,790	12.97%
Victim Was Stranger	282,912	12.79%
Victim Was Otherwise Known	210,300	9.51%
Victim Was Spouse	119,859	5.42%
Victim Was Parent	80,730	3.65%
Victim Was Other Family Member	69,761	3.15%
Victim Was Child	69,669	3.15%
Victim Was Sibling	56,121	2.54%
Victim Was Friend	54,207	2.45%
Victim Was Offender	52,615	2.38%
Victim Was Neighbor	36,082	1.63%
Victim Was Ex-Relationship (Ex-Boyfriend/Ex-Girlfriend)	24,865	1.12%
Victim Was Ex-Spouse	24,575	1.11%
Victim Was Common-Law Spouse	13,824	0.63%
Victim Was Step-Child	11,066	0.50%
Victim Was In-Law	10,238	0.46%
Victim Was Grandparent	8,881	0.40%
Victim Was Step-Parent	8,621	0.39%
Victim Was Child of Boyfriend/Girlfriend	8,273	0.37%
Victim Was Employee	6,816	0.31%
Victim Was Employer	6,231	0.28%
Victim Was In A Homosexual Relationship With The Offender	5,370	0.24%
Victim Was Grandchild	5,287	0.24%
Victim Was Step-Sibling	2,780	0.13%
Victim Was Babysittee (The Baby)	1,512	0.07%
Total	2,211,557	100%

Table 6.3: The distribution of the relationship between the victim and the offender for law enforcement officer victims.

Crime Category	# of Victims	% of Victims
Victim Was Stranger	19,779	58.89%
Victim Was Otherwise Known	6,294	18.74%
Relationship Unknown	6,225	18.53%
Victim Was Acquaintance	1,085	3.23%
Victim Was Employee	52	0.15%
Victim Was Boyfriend/Girlfriend	40	0.12%
Victim Was Spouse	16	0.05%
Victim Was Friend	15	0.04%
Victim Was In-Law	13	0.04%
Victim Was Other Family Member	11	0.03%
Victim Was Parent	9	0.03%
Victim Was Neighbor	9	0.03%
Victim Was Ex-Relationship (Ex-Boyfriend/Ex-Girlfriend)	8	0.02%
Victim Was Step-Sibling	6	0.02%
Victim Was Offender	6	0.02%
Victim Was Child	4	0.01%
Victim Was Sibling	4	0.01%
Victim Was Ex-Spouse	4	0.01%
Victim Was Common-Law Spouse	3	0.01%
Victim Was Employer	2	0.01%
Victim Was Step-Parent	2	0.01%
Victim Was Babysittee (The Baby)	1	0.00%
Total	33,588	100%

6.1.5 Aggravated assault and homicide circumstances

Table 6.4: The distribution of circumstances for aggravated assault and homicides.

Circumstance	Crime Category	# of Victims	% of Victims
Argument	Aggravated Assault/Murder	147,027	42.66%
Unknown Circumstances	Aggravated Assault/Murder	82,432	23.92%

Circumstance	Crime Category	# of Victims	% of Victims
Other Circumstances	Aggravated Assault/Murder	64,605	18.74%
Lovers Quarrel	Aggravated Assault/Murder	32,249	9.36%
Assault On Law Enforcement Officer(S)	Aggravated Assault/Murder	9,695	2.81%
Other Felony Involved	Aggravated Assault/Murder	4,007	1.16%
Drug Dealing	Aggravated Assault/Murder	1,718	0.50%
Gangland	Aggravated Assault/Murder	1,420	0.41%
Juvenile Gang	Aggravated Assault/Murder	588	0.17%
Other Negligent Killings	Negligent Manslaughter	490	0.14%
Criminal Killed By Private Citizen	Justifiable Homicide	181	0.05%
Criminal Killed By Police Officer	Justifiable Homicide	127	0.04%
Other Negligent Weapon Handling	Negligent Manslaughter	86	0.02%
Child Playing With Weapon	Negligent Manslaughter	18	0.01%
Mercy Killing	Aggravated Assault/Murder	6	0.00%
Gun-Cleaning Accident	Negligent Manslaughter	2	0.00%
Hunting Accident	Negligent Manslaughter	1	0.00%
Total	Aggravated Assault/Murder	344,652	100%

6.1.6 Justifiable homicide circumstance

6.1.7 Demographics

6.1.7.1 Residence status

Only for when victim is individual or law enforcement officer

6.1.7.2 Age

6.1.7.3 Sex

6.1.7.4 Race

6.1.7.5 Ethnicity

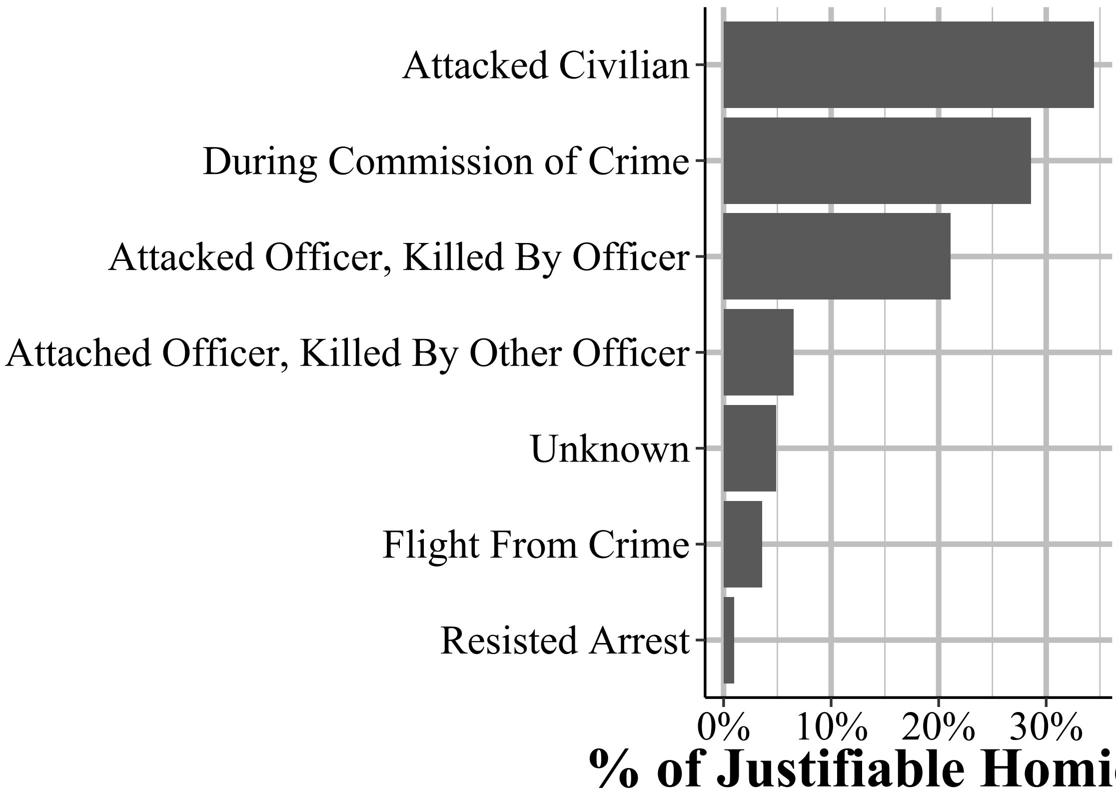

Figure 6.6: The distribution of circumstances for justifiable homicides (N = 308 in 2019 for all agencies reporting).

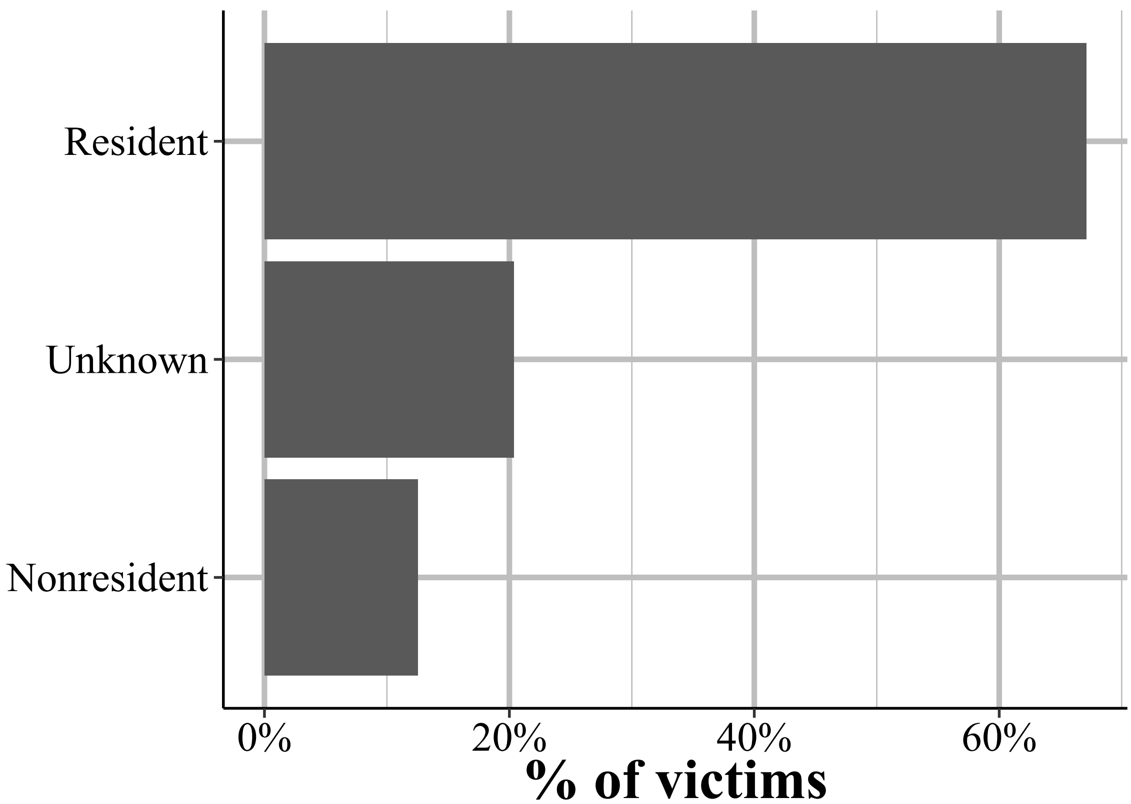

Figure 6.7: The distribution of residence status for all victims reported to NIBRS in 2019. Residence status is residence in the police agency's jurisdiction (e.g. do you live in the city you were victimized in?). It is unrelated to citizenship or immigration status.

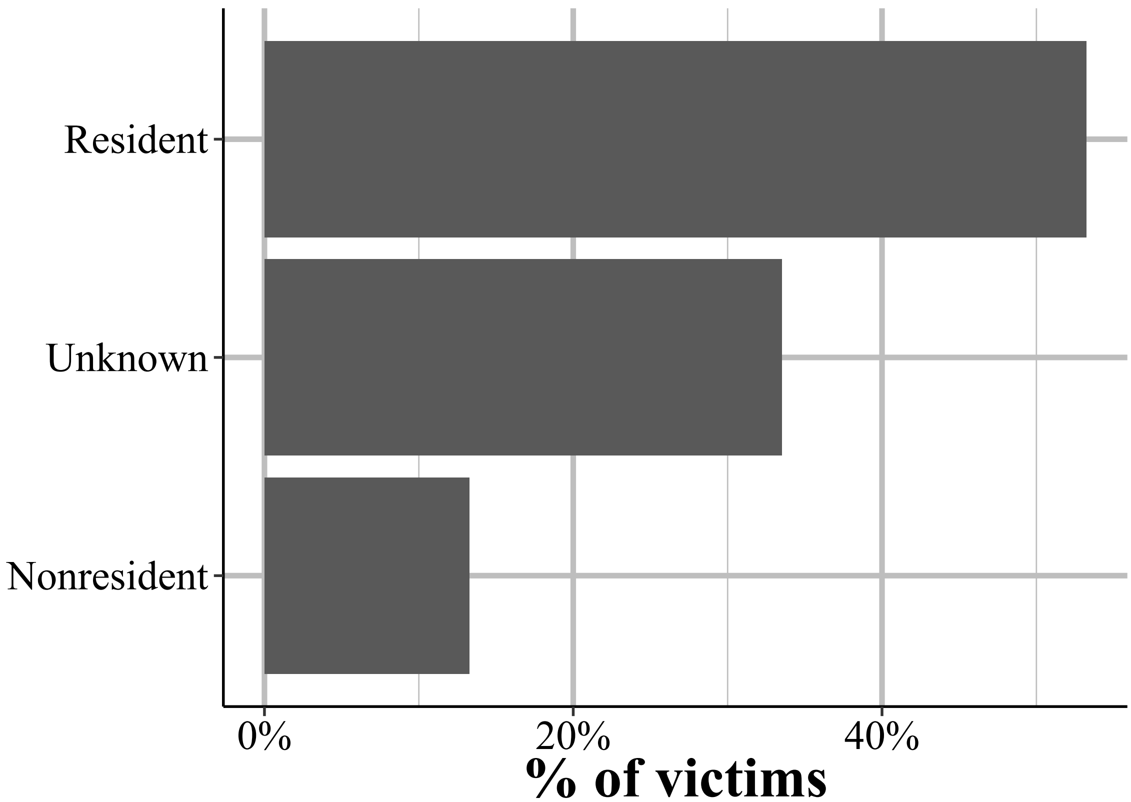

Figure 6.8: The distribution of residence status for all Law Enforcement Officer victims.

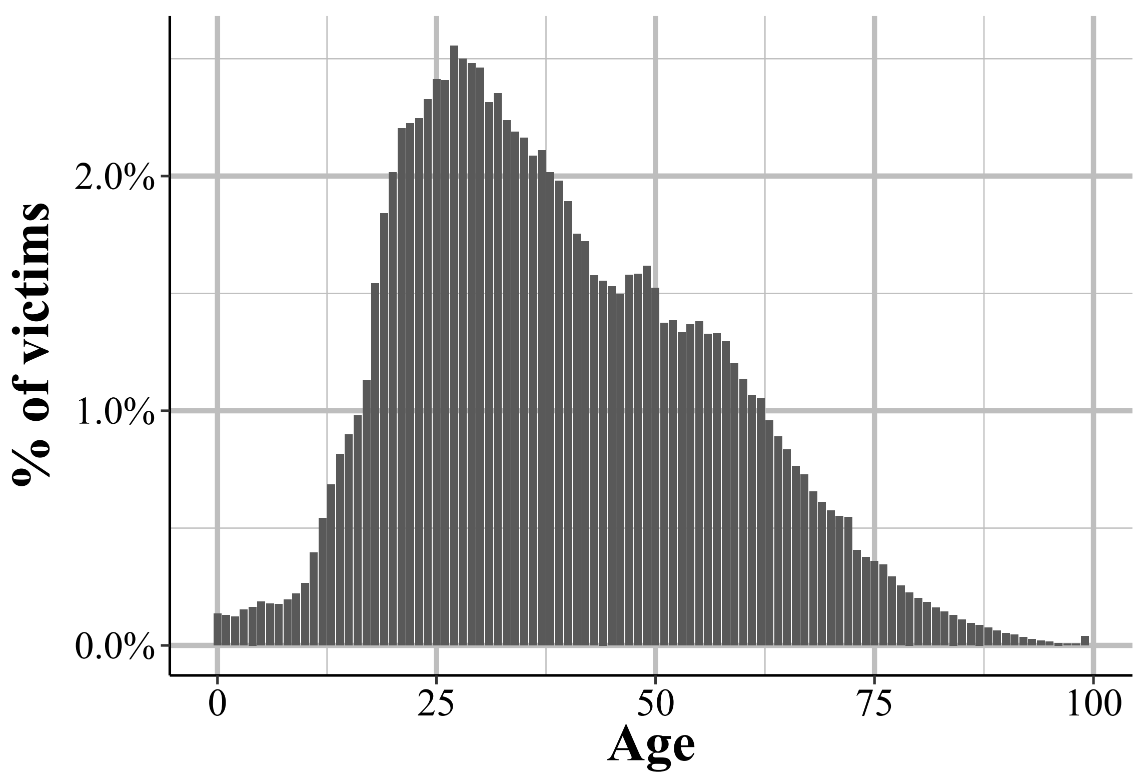

Figure 6.9: The age of all victims reported in the 2019 NIBRS data.

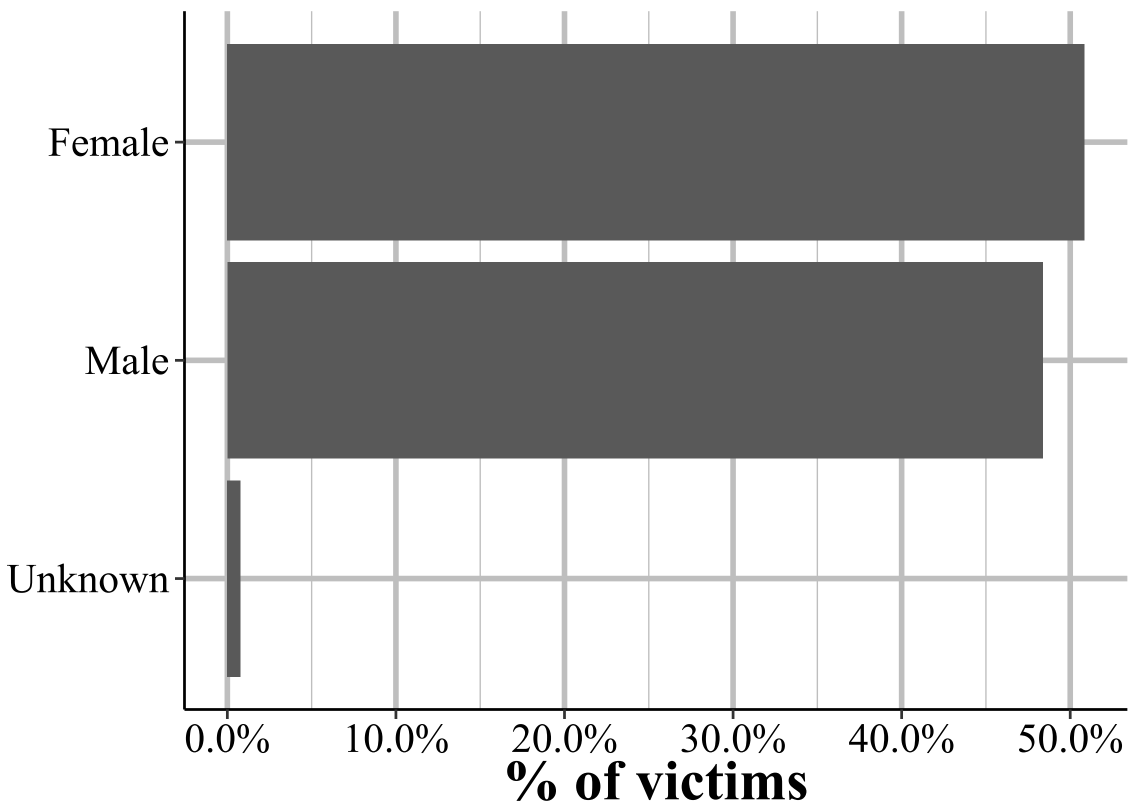

Figure 6.10: The sex of all victims reported in the 2019 NIBRS data.

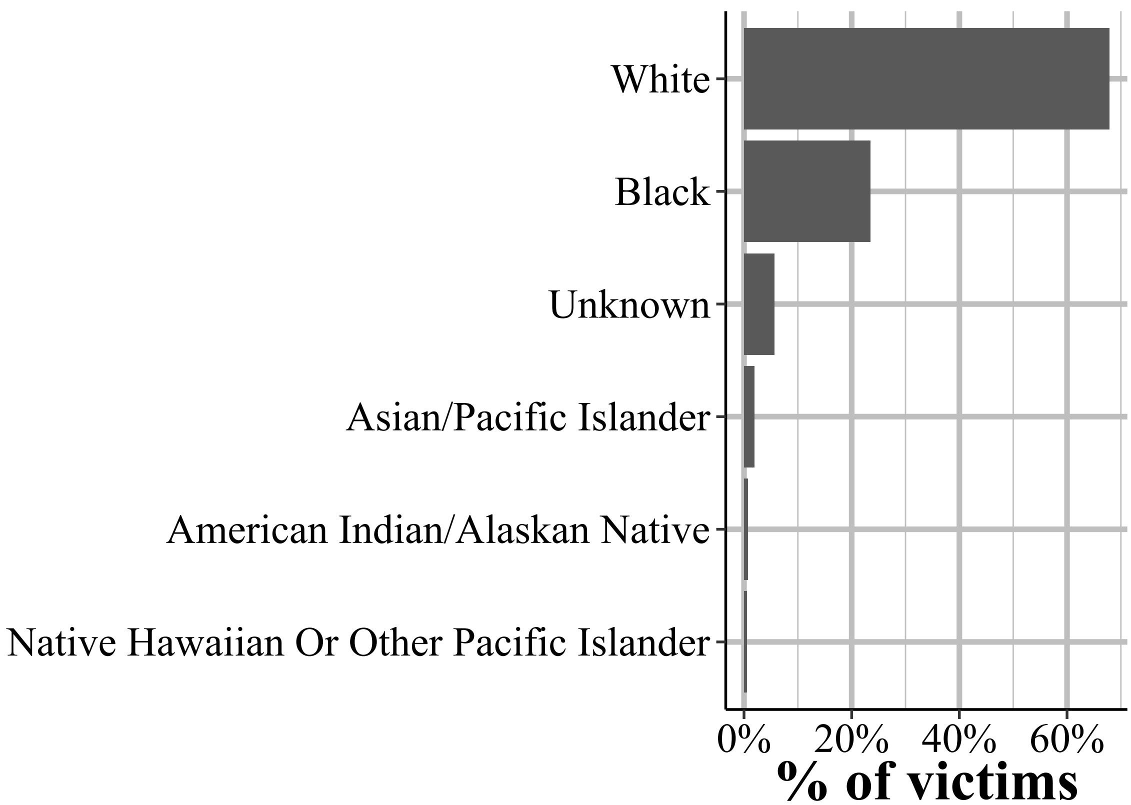

Figure 6.11: The race of all victims reported in the 2019 NIBRS data.

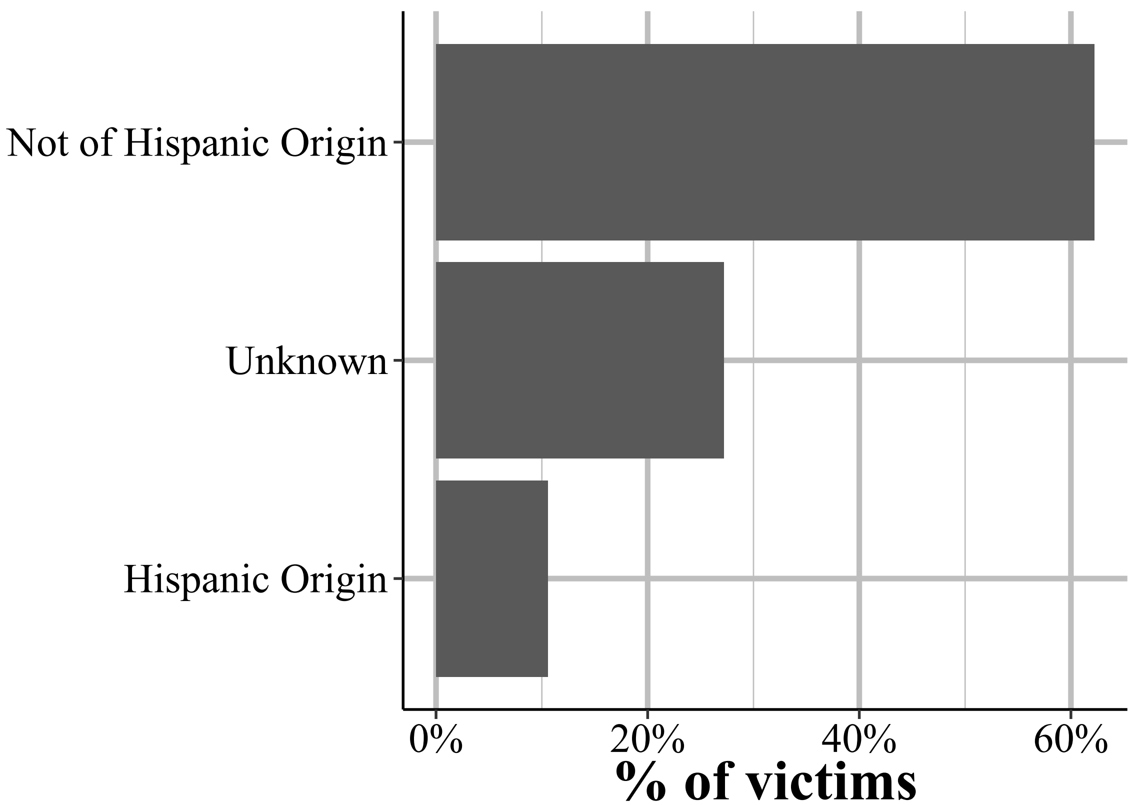

Figure 6.12: The ethnicity of all victims reported in the 2019 NIBRS data.

Chapter 7

Arrestee, Group B Arrestee, and Window Arrestee Segment

The Arrestee Segment has information that is largely available in other segments. For example, it has information on the arrestee's age, sex, and race which is the same as in the Offender Segment for that individual, and adds only ethnicity and residency status (of the city, not as a United States citizen) to the available demographics variables. It also says the crime the arrestee was arrested for, the weapon used and if the weapon (if it is a firearm) was an automatic weapon, all of which is available in the Offense Segment. There are a few new variables, however, including the date of the arrest and the type of arrest. The type of arrest is simply whether the person was arrested by police who viewed the crime, if the arrest followed an arrest warrant or a previous arrest (i.e. arrested for a different crime and then police find out you also committed this one so consider you arrested for this one too), and whether the person was cited by police and ordered to appear in court but not formally taken into custody. Finally, for juvenile arrestees it says whether arrestees were "handled within the department" which means they were released without formal sanctions or were "referred to other authorities" such as juvenile or criminal court, a welfare agency, or probation or parole department (for those on probation or parole).

This chapter also covers the Group B Arrestee Segment and the Window Arrestee Segment. The Arrestee Segment covers arrests for Group A offenses and there are corresponding Offense, Offender, and Victim segments for these incidents. Group B offenses, however, only have information about the arrest so incidents in this segment do not have any corresponding segments with it. Since Group B only has arrests without any associated incident, instead of the incident number variable like other segments have, this segment has an "arrest transaction incident number" which works the same as a normal incident number. Likewise, the Window Arrestee Segment isn't associated with any other segments as the "window" part means that they are only partial reports. The Window Arrestee Segment has the same variables as the normal Arrestee Segment but also has 10 variables on each of the offenses committed (up to 10 offenses) during the incident. This is really to try to provide a bit of information that you'd otherwise get from the other segments but don't since this is a window segment.

For the rest of this chapter I'll be using examples from the Arrestee Segment and not the Group B Arrestee (except for a table showing each Group B offense) or the Window Arrestee Segment.

7.1 Important variables

In addition to the variables detailed below this segment has the tradition agency and incident identifiers: the ORI code, the agency state, the year of this data, and the incident number. It also has an "arrestee sequence number" which is an identifier for an arrestee in an incident since incidents can have multiple arrests. This is just the number of each arrestee and to my knowledge is not associated with how involved the arrestee is. Being the 1st arrestee, for example, doesn't mean that individual played a greater role in the crime than the 2nd arrestee.

7.1.1 Crimes arrested for

This segment tells us which offense the arrestee was arrested for. There are a couple of caveats with this data. First, there can be up to 10 crimes per incident but this segment only tells you the most serious offense (based on the agency's decision of which is most serious). This can be solved partially by merging this segment with the Offense Segment and getting all of the offenses related to the incident. It's only partially solved because the crime the person is arrested for may not necessarily be the crime involved in the incident. This is because a person can be arrested for a certain crime (e.g. shoplifting) and then the police find out that there are also responsible for a more serious crime (e.g. aggravated assault). Their arrest crime is shoplifting and they'll be associated with the incident for the aggravated assault.

One interesting part of this segment is that while it's associated with Group A offenses, as someone may be arrested for a crime other than the crime in the incident, arrests can include Group B offenses. So the Group B Arrestee Segment can really be thought of as an arrest for a Group B offense where the arrestee isn't associated with a previous Group A incident (other than ones that already led to an arrest). We'll look first at the crimes people were arrested for in the main Arrestee Segment, which includes both Group A and Group B offenses as possible arrest crimes, and then at the Group B Arrestee Segment which only includes Group B offenses.

7.1.1.1 Arrestee Segment arrest crimes

Table 7.1 shows the number and percent of arrests for all arrests associated with a Group A crime incident. Perhaps unsurprising, drug crimes are the most common arrest making up a quarter of all arrests (30% when including drug equipment crimes). Simple assault (assault without a weapon or without seriously injuring the victim) is the next most common at 19% of arrests, and aggravated assault is the 4th most common arrest crime at 6.3% of arrests. Theft, which NIBRS breaks into a number of subcategories of theft such as shoplifting and “all other larceny” is among the most common arrest crimes, making up ranks 3 and 6 of the top 6, as well as several other subcategories later on. The remaining crimes are all relatively rare, consisting of under 5% of arrests each. This is due to a mix of the large concentration of drug and assault crimes as both arrests and overall crimes (more crimes of a certain category artificially inflates its share of arrests) and how the top crimes are broad categories (e.g. drug offenses ranges from simple possession to large scale sales) while other crimes are specific (e.g. purse-snatching is a very specific form of theft).

Table 7.1: The number and percent of arrests for Group A crimes for all arrests reported to NIBRS in 2019.

Crime Category	# of Offenses	% of Offenses
Drug/Narcotic Violations	523,732	25.82%
Simple Assault	385,695	19.02%
Shoplifting	228,355	11.26%
Aggravated Assault	127,192	6.27%
All Other Larceny	104,244	5.14%
Drug Equipment Violations	95,730	4.72%
Destruction/Damage/Vandalism of Property	69,153	3.41%
Burglary/Breaking And Entering	56,613	2.79%
Intimidation	52,972	2.61%
Weapon Law Violations	51,907	2.56%
All Other Offenses	49,288	2.43%
Stolen Property Offenses (Receiving, Selling, Etc.)	32,953	1.62%
Motor Vehicle Theft	28,489	1.40%
Robbery	25,590	1.26%
False Pretenses/Swindle/Confidence Game	23,183	1.14%
Theft From Motor Vehicle	18,780	0.93%
Counterfeiting/Forgery	18,022	0.89%
Theft From Building	15,899	0.78%
Disorderly Conduct	11,413	0.56%

Crime Category	# of Offenses	% of Offenses
Driving Under The Influence	10,084	0.50%
Impersonation	9,429	0.46%
Kidnapping/Abduction	8,941	0.44%
Credit Card/Atm Fraud	6,846	0.34%
Fondling (Incident Liberties/Child Molest)	6,659	0.33%
Trespass of Real Property	6,439	0.32%
Rape	6,406	0.32%
Embezzlement	6,327	0.31%
Prostitution	5,492	0.27%
Murder/Nonnegligent Manslaughter	4,788	0.24%
Liquor Law Violations	4,267	0.21%
Identity Theft	3,802	0.19%
Drunkenness	3,753	0.19%
Pocket-Picking	3,137	0.15%
Pornography/Obscene Material	3,055	0.15%
Arson	3,016	0.15%
Family Offenses, Nonviolent	2,498	0.12%
Theft of Motor Vehicle Parts/Accessories	2,026	0.10%
Animal Cruelty	1,852	0.09%
Assisting Or Promoting Prostitution	1,442	0.07%
Sodomy	1,374	0.07%
Statutory Rape	1,137	0.06%
Purse-Snatching	866	0.04%
Curfew/Loitering/Vagrancy Violations	840	0.04%
Sexual Assault With An Object	647	0.03%
Purchasing Prostitution	609	0.03%
Theft From Coin-Operated Machine Or Device	426	0.02%
Negligent Manslaughter	327	0.02%
Operating/Promoting/Assisting Gambling	262	0.01%
Betting/Wagering	262	0.01%
Extortion/Blackmail	254	0.01%
Welfare Fraud	241	0.01%
Human Trafficking - Commercial Sex Acts	232	0.01%
Bribery	230	0.01%
Bad Checks	210	0.01%
Wire Fraud	180	0.01%
Incest	150	0.01%

Crime Category	# of Offenses	% of Offenses
Runaway	90	0.00%
Gambling Equipment Violations	89	0.00%
Hacking/Computer Invasion	66	0.00%
Peeping Tom	39	0.00%
Human Trafficking - Involuntary Servitude	27	0.00%
Sports Tampering	1	0.00%
Total	2,028,028	100%

7.1.1.2 Group B Segment arrest crimes

Table 7.2 shows the number and percent of arrests for all arrests associated with a Group A crime incident. The offense categories overlap with Table 7.1 but these are for separate arrests, a single arrest cannot be in both segments. Unhelpfully, the most common Group B offense is “All other offenses” which means that it’s a crime that isn’t covered in either the Group A or the Group B crime categories. However, this can also include Group A or Group B crimes if they are not completed. So an attempted or a conspiracy to commit a Group A or B crime can go in this category. At 57% of Group B arrests, this very vague category covers a huge amount of arrests. The next most common Group B arrest is driving under the influence of drugs or alcohol, and this occurred in 18.4% - or 352k times - of arrests.

Trespassing makes up 5.7% of arrests and this is unlawfully entering someone’s property, including a building. The difference between this and burglary is that this is entering without any intent to commit theft or a felony. Disorderly conduct is a broad category ranging from indecent exposure (which should be it’s own sex offense but isn’t for some reason) to “profanity” and noise violations, and it makes up 6.2% of arrests. So be cautious with this offense as it ranges from very minor to quite serious and there’s no distinguishing what actually happened. Drunkenness and liquor law violations make up 6% and 3.6% of arrests, respectively. The difference is that drunkenness is when someone is seriously drunk in public (to the point where they can’t control their own body) and liquor law violations are about illegal making or selling of liquor. So basically bootlegging, selling alcohol without a license (or to people not allowed to drink, like minors), or avoiding paying tax on alcohol sales. “Family Offenses, Nonviolent” is also a rather vague category and includes “nonviolent abuse” (which I guess means emotional abusive) as well as neglecting the child in a few different ways like not paying alimony and deserting the child. Since these are arrests, the actions have to reach the level of criminal behavior, simply being an awful parent (or even leaving the child, as long as they have another adult to watch them) is not this offense. Runaways is an offense that only applies to people under age 18 and is no longer collected in NIBRS.

The remaining categories are all rare and none are more than 1% of arrests.

Table 7.2: The number and percent of arrests for Group B crimes for all arrests reported to NIBRS in 2019.

Crime Category	# of Offenses	% of Offenses
All Other Offenses	1,095,755	57.26%
Driving Under The Influence	351,926	18.39%
Disorderly Conduct	117,707	6.15%
Drunkenness	116,343	6.08%
Trespass of Real Property	108,546	5.67%
Liquor Law Violations	68,862	3.60%
Family Offenses, Nonviolent	31,251	1.63%
Runaway	9,535	0.50%
Curfew/Loitering/Vagrancy Violations	9,360	0.49%
Bad Checks	3,911	0.20%
Peeping Tom	414	0.02%
Total	1,913,610	100%

7.1.2 Arrest date

For each arrest we know the exact date of the arrest. As with the incident date, there is evidence that when agencies don't know the exact arrest date, they put down the first of the month. However, this is far less of a problem than with the incident date, likely because since the agency is doing the arresting they know exactly when they do it. Instead of looking at arrests by day of the month, we'll use both the arrest date and the incident date to look at how long it takes for crimes to get solved.

Figure 7.1 shows how long it takes for arrests to be made. The shortest time is zero days which means the arrest and the incident happened on the same day and the longest is 461 days after the incident. About 76.5% of arrests happen on the same day as the incident while 6.6% happen on the next day. 1.4% happen the following day and 1% on the day after this. This trend of a lower probability of the case being solved as the time from the incident continues throughout the figure. Including dates up to 461 days is a bit ridiculous since it's impossible to see trends among the early dates other than zero days, but it's a good demonstration of how massively concentrated arrests are that occur on the same day of the incident. The lesson here is that if an arrest isn't made on the day of the incident (such as at the scene of the crime), it's very unlikely that'll it'll be made at all (and most crimes never lead to an arrest).

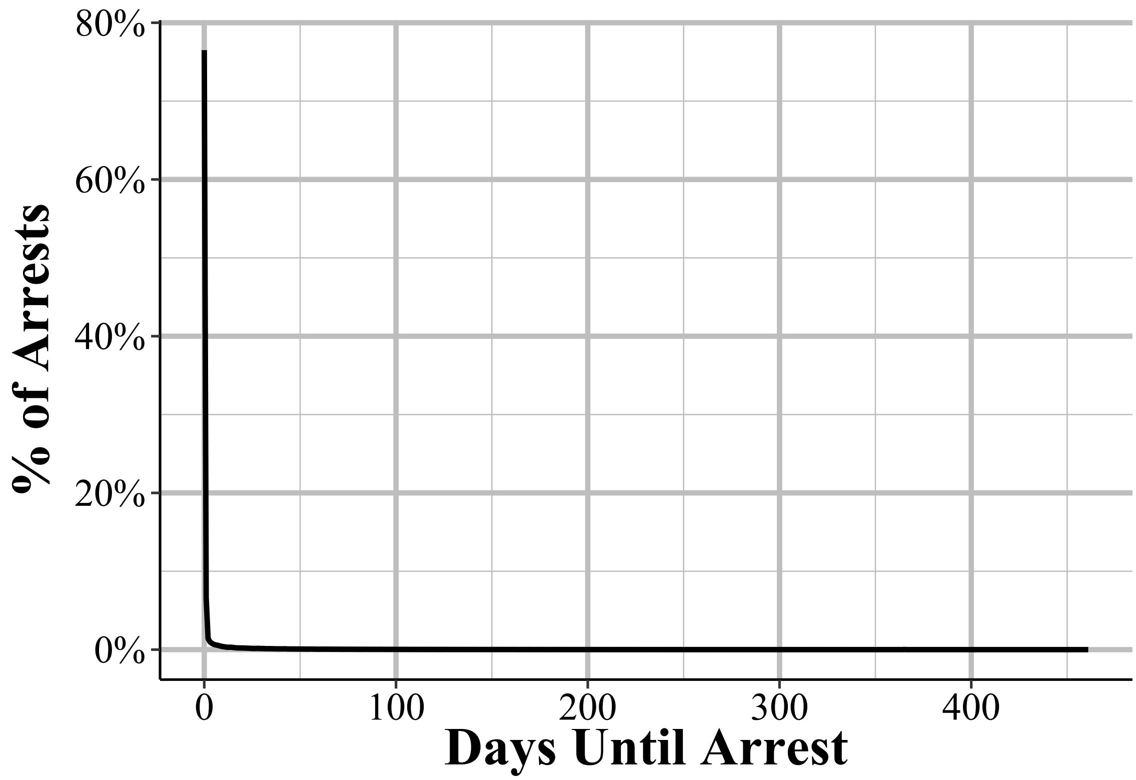

Figure 7.1: The distribution of the number of days from the incident to the arrest date. In 2019 the maximum days from incident to arrest was 461 days. Zero days means that the arrest occurred on the same day as the incident.

Figure 7.2 groups the larger number of days together to make it easier to see trends early after the incident. Here we can see much better how the percent of arrests move quickly downwards after zero days.

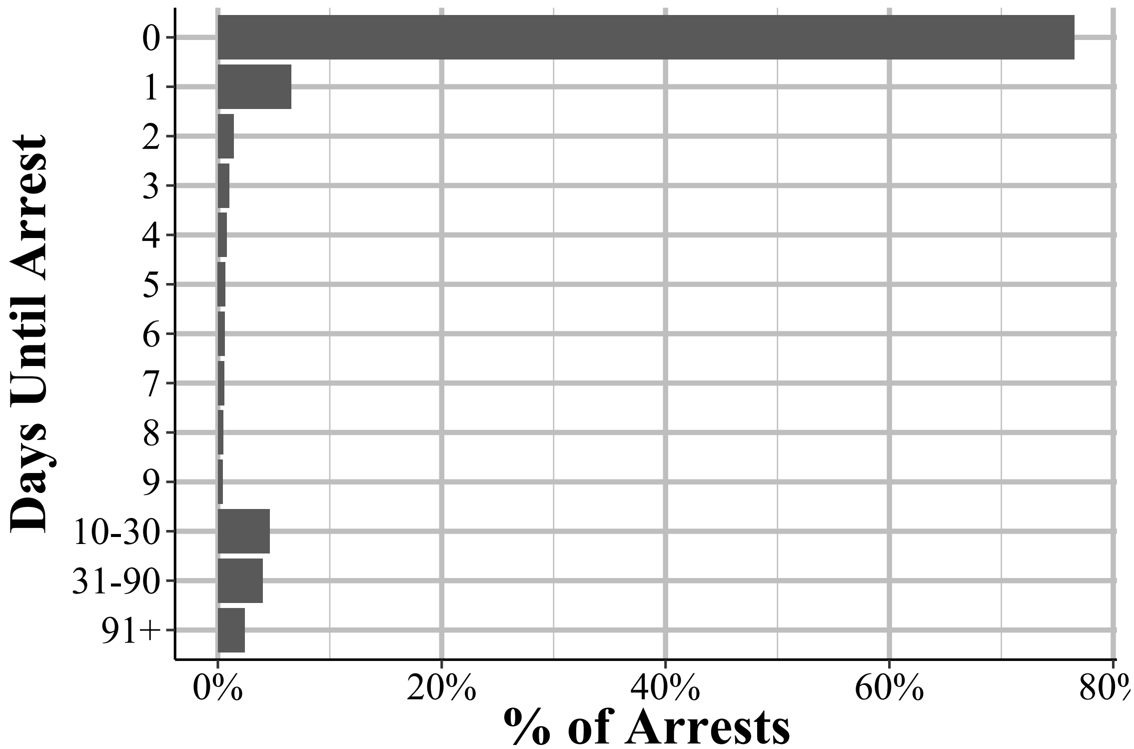

Figure 7.2: The number of days from the incident to the arrest date. Values over 10 days are grouped to better see the distribution for arrests that took fewer than 10 days. Zero days means that the arrest occurred on the same day as the incident.

7.1.3 Weapons

In the Offense Segment we get info on what weapon (if any) was used during the crime. Here, we see what weapon the person arrested was carrying *when they were arrested.* There is probably a very large overlap here, especially given that the vast majority of arrests happen on the same day as the offense, so probably happen at the scene of the crime (and we'll see exactly which ones do happen there later on in this chapter). Compared to the weapons covered in the Offense Segment - see Section 4.1.6 for more - the weapons here are only a narrow subset, and cover mostly firearms. This is partly because the most common "weapon" in the Offense Segment is that the offender used their body as a weapon (e.g. punching, kicking) and everyone arrested has a body so it doesn't make sense to count that as a weapon. Each arrestee can carry up to two weapons, but we'll only look at the first weapon for the below graphs. Please note that this is weapons found on the arrestee, and doesn't mean that they used the weapon during the arrest.

Figure 7.3 shows the breakdown in the weapon carried by the arrestee during the arrest. In 94.% of arrests, the arrestee was not carrying any weapon. Since this graph shows arrests for all crimes, it makes a good deal of sense. Most crimes are non-violent so we'd expect those people to not carry a weapon. Since so few arrestees have weapons, we'll look at the breakdown among those who were carrying a weapon in Figure 7.4.

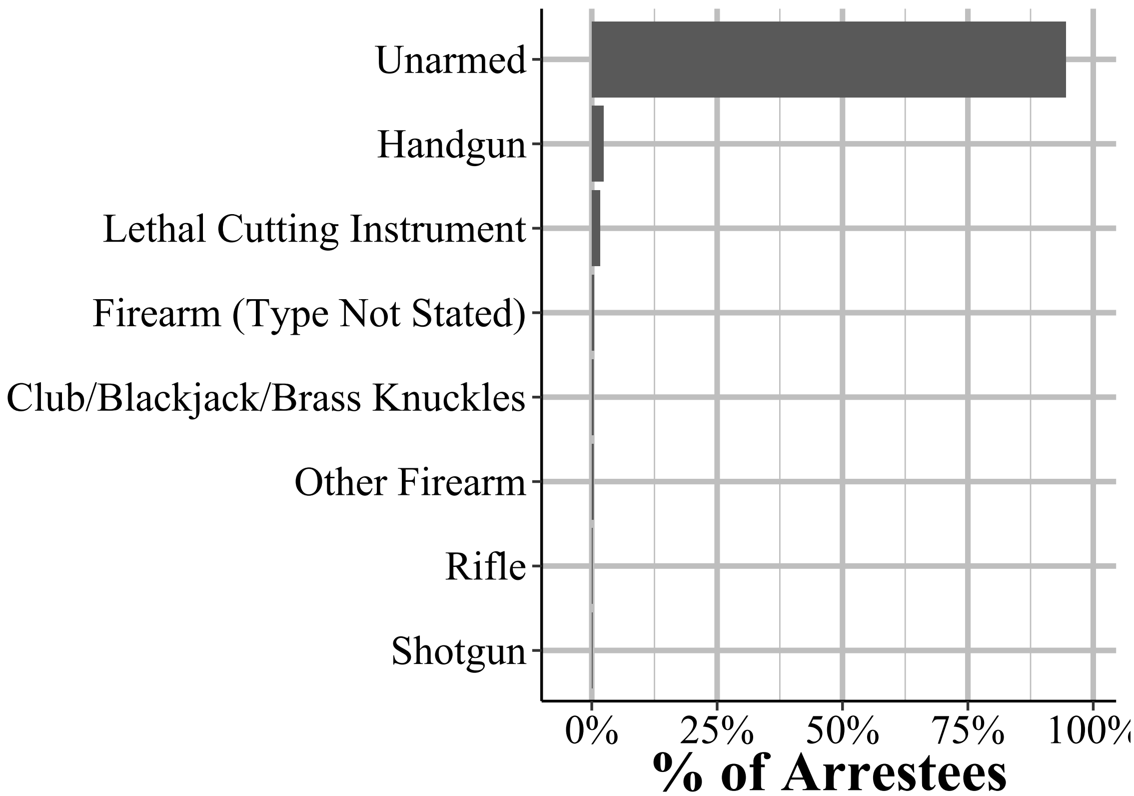

Figure 7.3: The weapon found on the arrestee for all arrestees reported in 2019.

To see the weapons carried when the arrestee had a weapon, Figure 7.4 shows the breakdown in which weapon they carried. About 43.8% of people arrested who had a weapon were carrying a handgun followed by 30% with some kind of "lethal cutting instrument" like a knife. While rifles, and especially "assault rifles", are what people (and especially politicians and the media) focus on when talking about violent crime, handguns are actually the most common gun to be used in a crime so it makes sense that handguns are the most frequently found weapon. "Firearm (type not stated)" basically means that the type of firearm used is unknown so can belong in one of the firearm categories and makes up 9% of weapons. Blunt instruments (including bats, clubs, and brass knuckles) follow at 6.9% of weapons. And the remaining weapons included are "other firearm" (so any other than ones specified) at 5.8%, rifle at 2.3%, and shotgun at 2%.

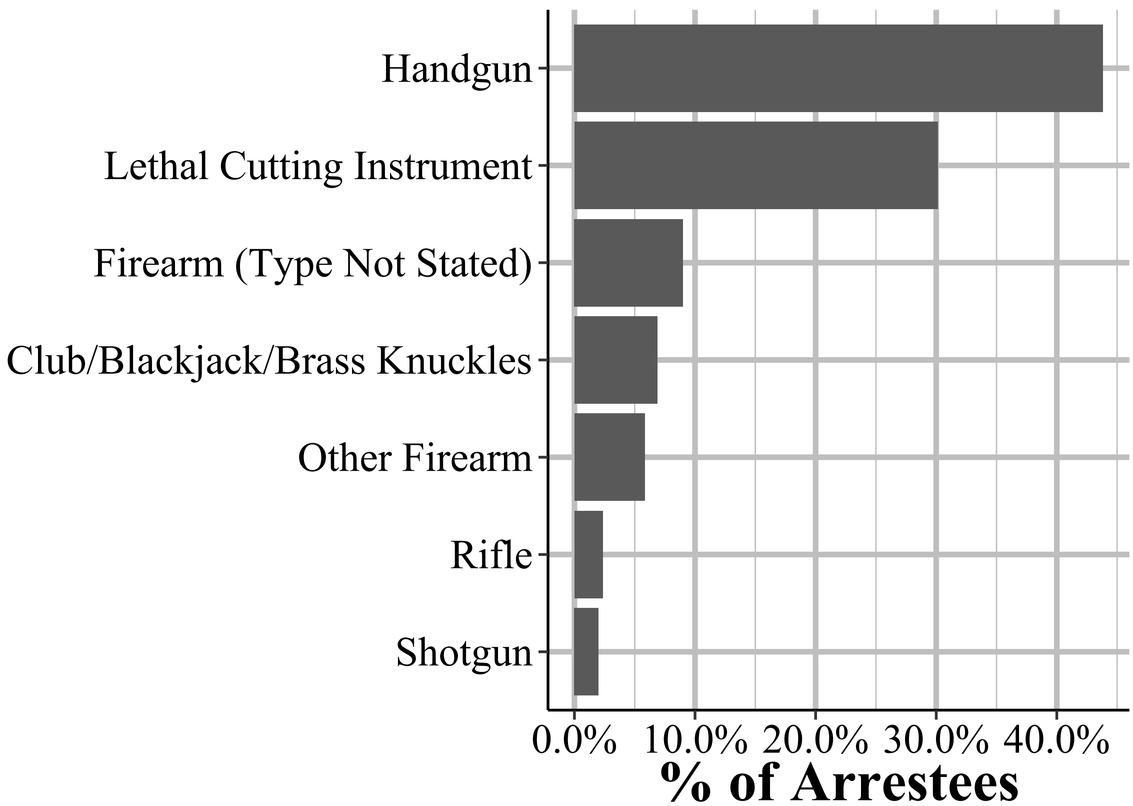

Figure 7.4: The distribution of weapon usage for all arrestees in 2019 who were arrested with a weapon (i.e. excludes unarmed arrestees).

7.1.4 Automatic weapons

This variable tells you if the weapon the arrestee was carrying was a gun whether that gun was fully automatic. To be clear, this means that when you pull the trigger once the gun will fire multiple bullets. Semi-automatic firearms are not automatic firearms. The Offense Segment also has a variable indicating if the offender used an automatic weapon but there they didn't necessarily recover the gun so it's much less reliable than in this segment where the police have the gun and are able to test it.[1] The percent of guns that are fully automatic are fairly similar between the weapons seized at arrest, as shown in Figure 7.5 and those used in the offense as shown in Figure 4.3 in Chapter 4. Figure 7.5 shows that about 5.6% of rifles seized by police during an arrest were fully automatic. About 4.9% of handguns are automatic while "firearm (type not stated) are automatic in 4.3% of cases. Shotguns and"other firearm" category are the least likely to be automatic at about 2.5% and 1.1% of weapons, respectively.

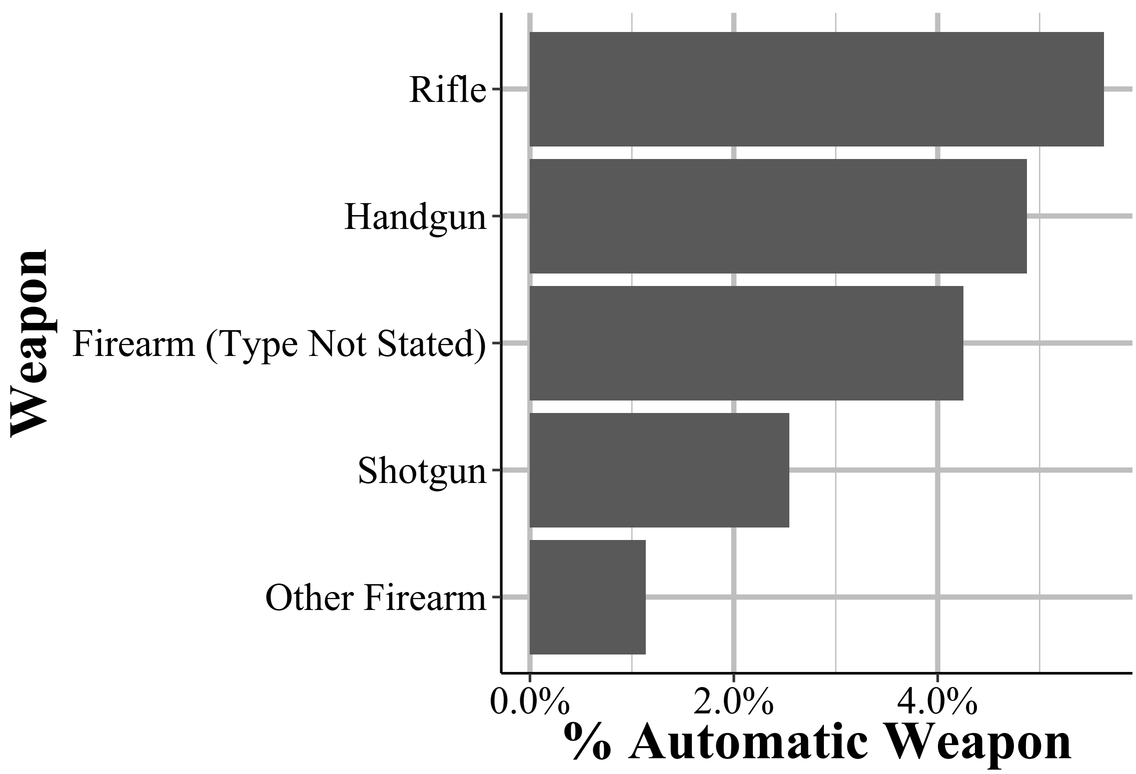

Figure 7.5: The percent of firearms the arrestee was carrying that were fully automatic, for arrestees in 2019.

[1] It's not clear whether they actually test it or simply go by the design of the gun, such as if the model allows for fully automatic firing.

7.1.5 Type of arrest

While arrests are sometimes talked about as if they are a homogeneous group (likely in big part because UCR data doesn't differentiate types of arrests), NIBRS data breaks them down into three different types of arrests. Figure 7.6 shows the distribution in the type of arrest for all arrestees in the 2019 NIBRS data. The most common type is "On-View" which means that the person is arrested at the scene by an officer. For example, if police respond to a bank robbery and nab the robbers as they run out of the bank, this is an on-view arrest. On-view arrests make up the majority - 50.9% - of arrests.

The next type of arrest is a "warrant/previous incident report" and this makes up 26.8% of arrests. In these cases the police had an arrest warrant and found the person and arrested them based on that warrant. This also includes when a person is arrested and found to have been involved in previous incidents. Then these previous incidents would be considered cleared from this single arrest. In these cases NIBRS has an indicator variable, called the "multiple arrestee indicator" to tell us that this individual is responsible for multiple incidents cleared so we avoid counting them twice (as their demographics will be the same each time). In this variable it says "count arrestee" if this is their only arrest or if this is the first arrest that is counted in cases where multiple incidents are cleared by the arrest, and "multiple" otherwise.

Finally, people can get a "summoned/cited" arrest which isn't really an arrest at all. This is when they are given a subpoena that says that must go to court to be tried for a crime, but they are not formally arrested and taken into police custody (i.e. they are never handcuffed, taken to a police precinct or to jail). About 22.2% of arrests are this form of arrest.

7.1.6 Disposition for juvenile arrestees

For juvenile arrestees - those under age 18 *at the time of the arrest* (and, by definition they'd also be under age 18 during the incident) - there is some information about the outcome of the arrest.[2] There are two possible outcomes (which NIBRS calls "dispositions"): being referred to other (that is, other than the arresting agency) authorities or handled within the arresting agency. Figure 7.7 shows this breakdown and being referred to other authorities is the most common outcome at 72.6% of juvenile arrests. This is a very broad category and the "other authorities" can range from juvenile or adult court (so the police recommend that they be prosecuted) to welfare agencies and being sent to other police agencies (such as if they committed a crime elsewhere and are being extradited). The other category, being handled within the department, means that the police release the juvenile without any

[2] There are a few people older than 18 with this variable but it's so rare that I think that they're just incorrectly inputted ages.

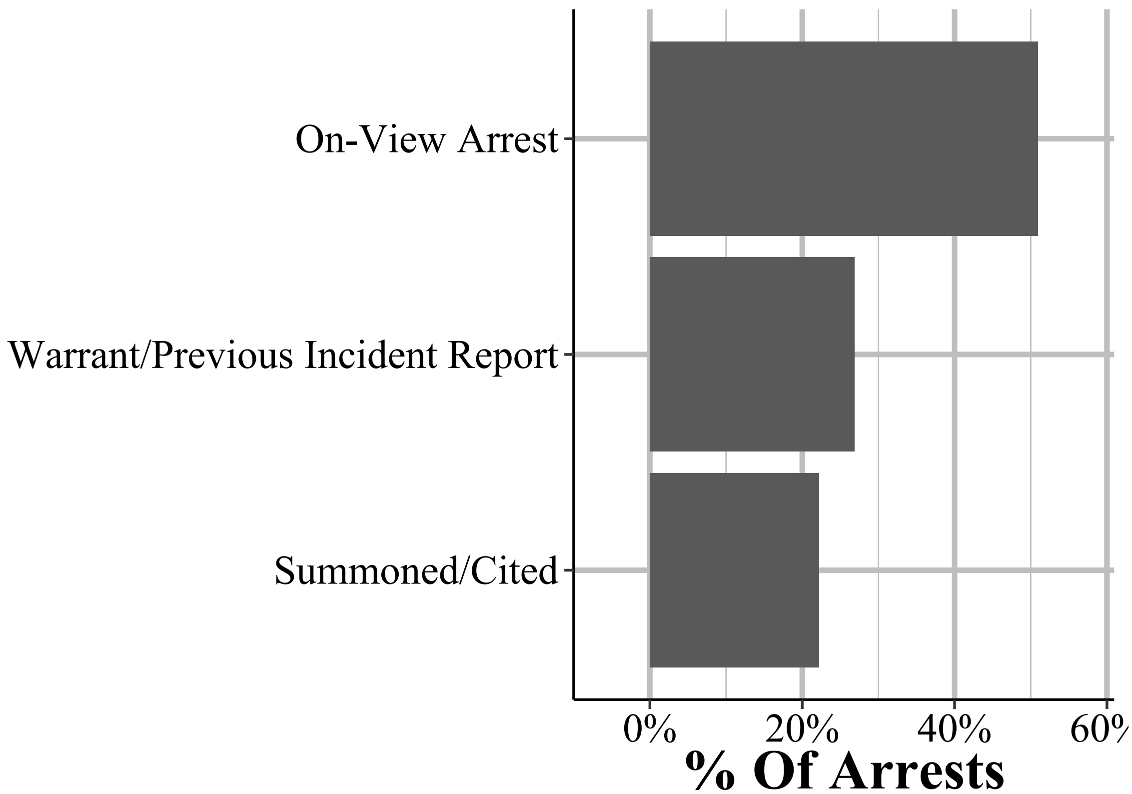

Figure 7.6: The distribution of arrests by type of arrest. Previous Incident Report includes cases where an individual was arrested for a separate crime and are then reported as also arrested for this incident.

formal action taken (but they may give the juvenile a warning). In these cases the juvenile is released to an adult (including but not limited to family members or guardians) and the case is essentially dropped. In about 0.001% of juvenile arrests the disposition is unknown.[3]

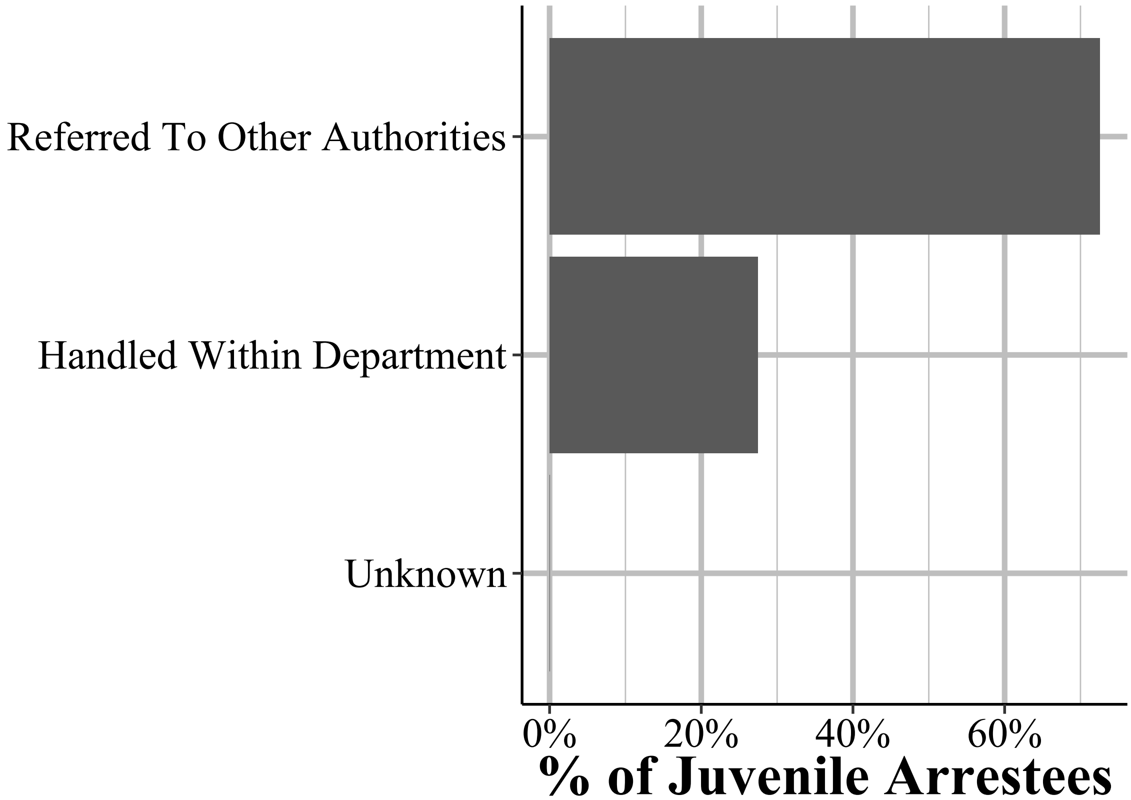

Figure 7.7: For juvenile arrestees (under age 18), the distribution of case outcomes.

7.1.7 Demographics

This segment provides several variables related to who the arrestee is. Age, race, and sex overlap with the Offender Segment but this segment also adds ethnicity and the whether they live in the jurisdiction of the agency (i.e. the city the police patrols) they were arrested by.

7.1.8 Residence status

The first variable we'll look at is the residence status for the arrestee. This is residence in the jurisdiction that arrested them and it has nothing to do with residence status or citizenship

[3] A juvenile can potentially get multiple dispositions, such as if they're initially released but later the police recommend them for prosecution. It's not clear which outcome is recorded in these cases. In UCR data, however, only the initial disposition is recorded so that is likely how it also is in NIBRS.

status in the United States. People tend to commit crimes (and the the victims of crimes) very close to where they live, so this provides some evidence for that. We don't know where the arrestee lives, but know if they live in the jurisdiction or not. This is useful because some areas (e.g. Las Vegas, Washington DC, urban city centers where people commute to work) likely have a lot more people moving into those areas during the day but who live else compared to places like rural towns or suburbs. So it's helpful to be able to distinguish locals arrested with people may be tourists or come into town just to commit the crime.[4]

One thing to be cautious about is that residence status may be an imprecise measure of where someone actually lives. How it's defined may differ by agency which could affect comparability across agencies. For example, if it's defined as official residence (such as address on a driver's license) that may be incorrect for a sizable share of the population (e.g. many college students live far from where their driver's license address is).[5] If residence status is based on asking the arrestee, they may of course lie to the police. There's also the question of how to label people who are truly transient such as homeless people who may move from city to city.

Figure 7.8 shows the percent of arrestees in 2019 who were residence or not (or whose status was unknown) of the jurisdiction that arrested them. Most people were arrested near they live, with 56.9% of arrestees being residents, compared with 23.3% not being residents. However, about one-fifth of arrestees had an unknown residence status so the percents of resident and non-resident may change dramatically if we didn't have any unknowns.

7.1.8.1 Age

This variable is the age at the arrest, which may be different than age during the crime. As in the Offender Segment we are giving the exact age (in years) but agencies can input a range of possible ages with the FBI giving up the average of this range (rounding down, not to the nearest integer) in the data. In Figure 5.1 in the Offense Segment, this can be seen in the sudden spikes in the percent of offenders of a certain age and that some of the most common ages are divisible by five (e.g. 20, 25, 30). There are also far fewer unknown ages in this data with only 0.1% of arrestees having a missing age. This is reasonable as a person arrested is present for the police to learn their age from something like a driver's license or past criminal records, or estimate the age by looking at the arrestee. Like the Offender Segment, age as a specific year is cutoff at 98 with all older ages grouped simply as "over 98 years old".

Figure 7.9 shows the percent of arrestees at every age available. Relative to Figure 5.1, this

[4] In a ridealong I was on a woman who lived over an hour from the city I was in was caught shoplifting clothes.

[5] One another ridealong a man from Florida was arrested for stealing from a local store (in California).

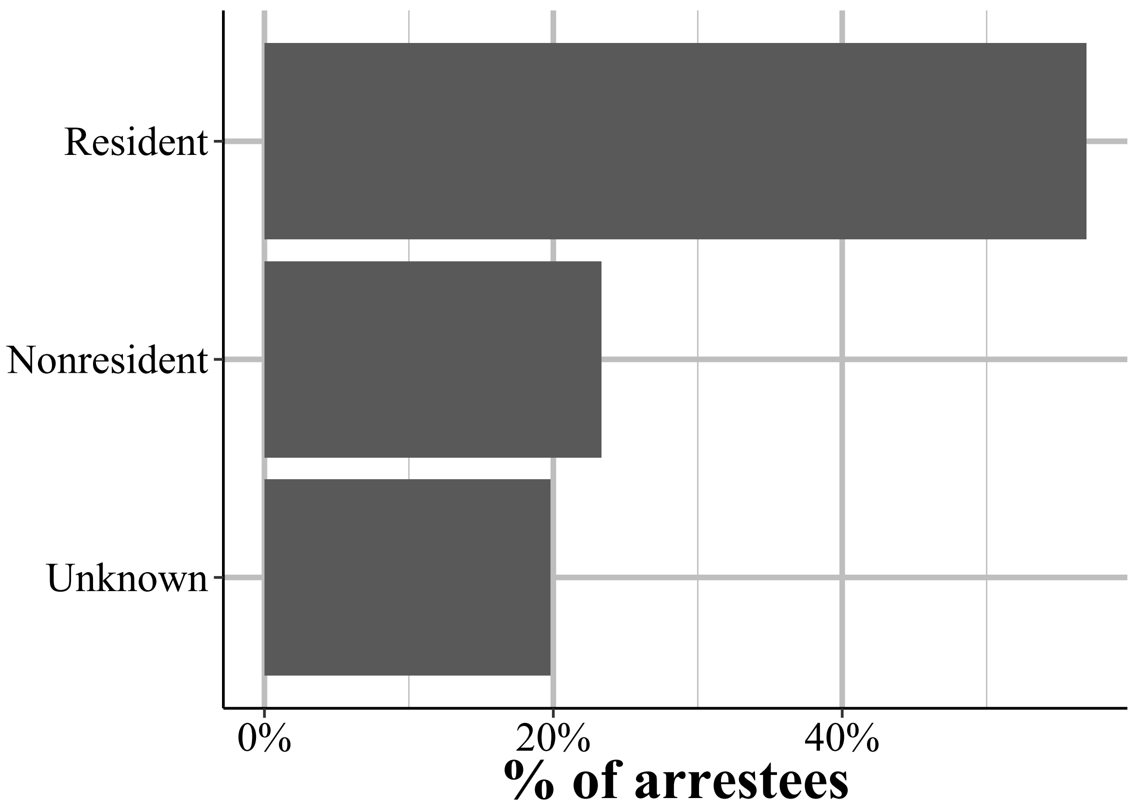

Figure 7.8: The distribution of residence status for all arrestees reported to NIBRS in 2019. Residence status is residence in the arresting agency's jurisdiction (e.g. do you live in the city you were arrested in?). It is unrelated to citizenship or immigration status.

graph is far smoother, indicating that there was less estimating ages and more knowing the actual age. While the trend is the same for both of these graphs, the arrestee data doesn't have any odd spikes with certain ages. Age we see that the percent of people arrested increases as they age, peaking in the early twenties before declining and then peaking age even higher in the late 20s. After this, there is a long steady decline as the arrestee ages.

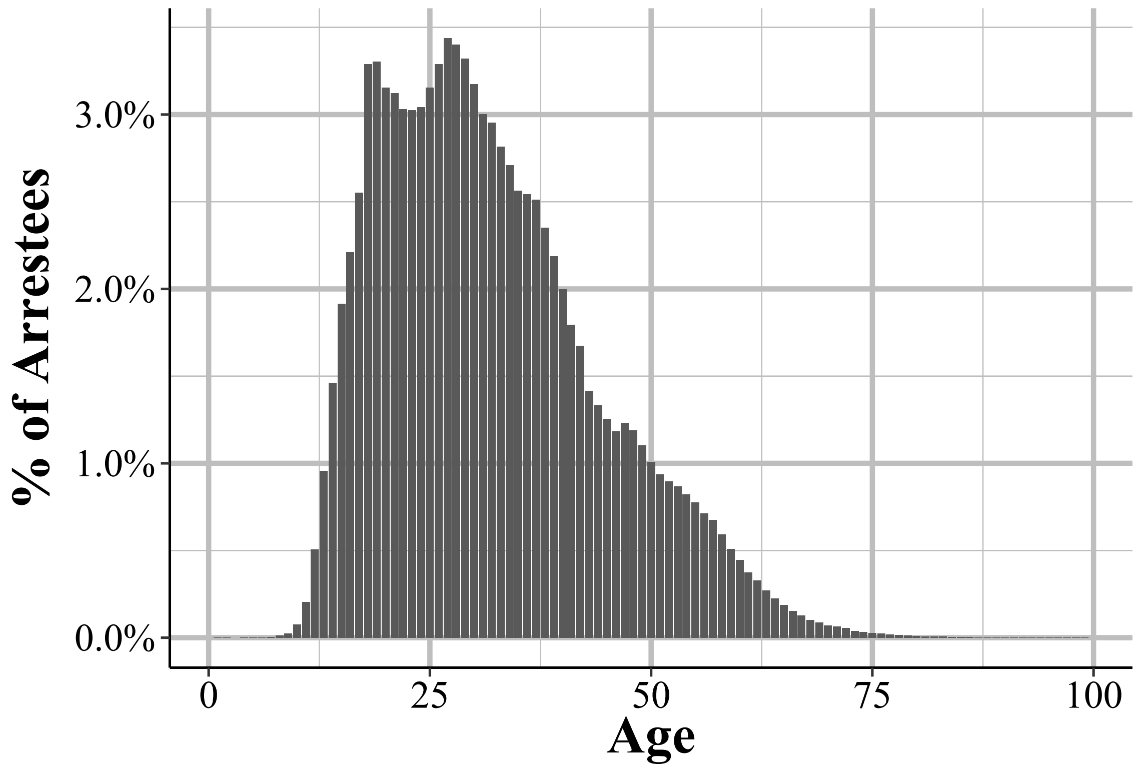

Figure 7.9: The age of all arrestees reported in the 2019 NIBRS data.

7.1.8.2 Sex

We also know the sex of the arrestee. The only options for this variable are male and female and there is never missing data so the police always choose one of these two choices. There is no option for transgender or any other identity. Figure 7.10 shows the distribution of arrestees by sex. The vast majority, 70.5% of arrestees are male and the remaining 29.5% are female. This is a higher rate of female arrestees than you might expect - past research has found that crime is largely a male-phenomenon, even greater than found here - and that's because there are differences in sex involvement by type of crime. For rape, as an example, 97.8% of arrestees in 2019 were male. Shoplifting was an even 50% split between female and male arrestees.

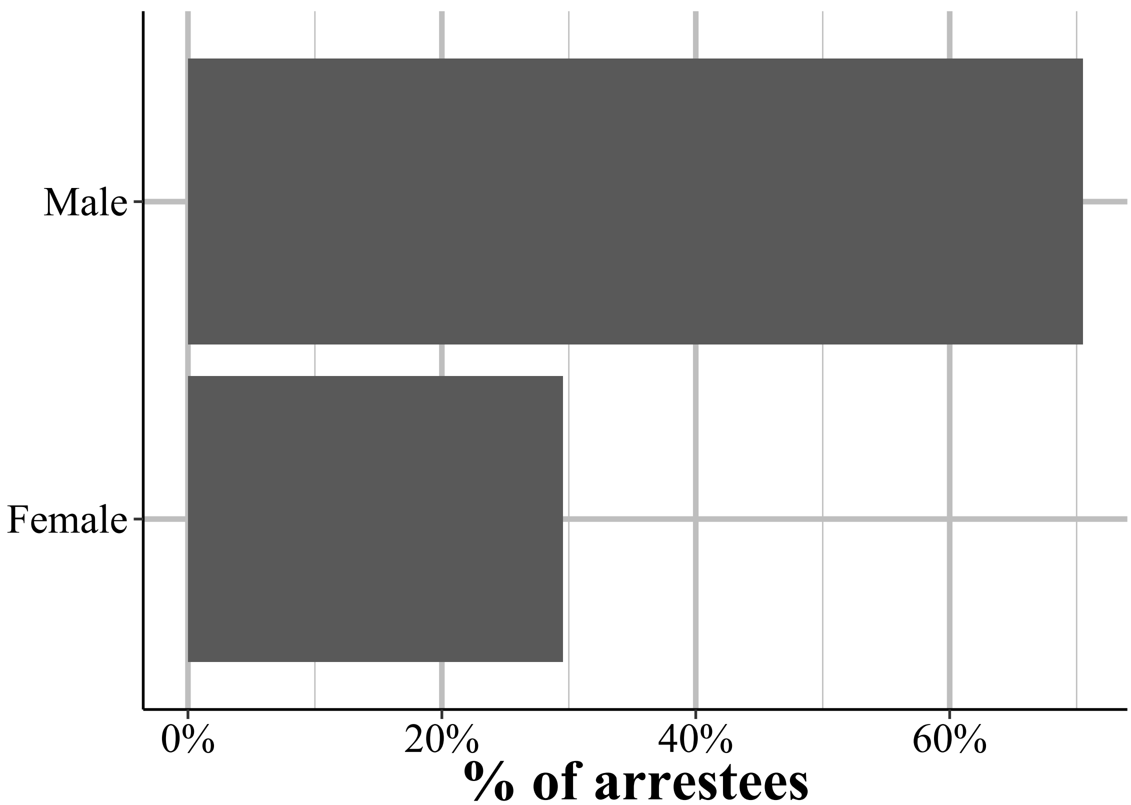

Figure 7.10: The sex of all arrestees reported in the 2019 NIBRS data.

7.1.8.3 Race

For each arrestee we know their race, with possible races being White, Black, American Indian/Alaskan Native, Asian, and Native Hawaiian/Pacific Islander. Unlike sex, the police can say that the race is unknown.[6] As each arrestee is visible to the police, and can self-report race or provide official records (e.g. criminal history or jail admission data) which may say their race, there is far less uncertainty for arrestees race than offender's race where 38.5% of the data is missing. As with any measure of race there is still some degree of uncertainty since people's race are not always obvious and may not fit tidily into each of the mutually exclusive groups available in NIBRS data (e.g. there is no option for mixed race).

Figure 7.11 shows the breakdown for the races of each arrestee. White arrestees are most common at 64.2% of arrestees, following by Black arrestees at 29.8%. Only 3.1% of arrestees have an unknown race. The remaining small share of arrestees is made up of American Indian/Alaskan Native at 1.6%, Asian at 1%, and Native Hawaiian/Pacific Islander at 0.3% of arrestees.

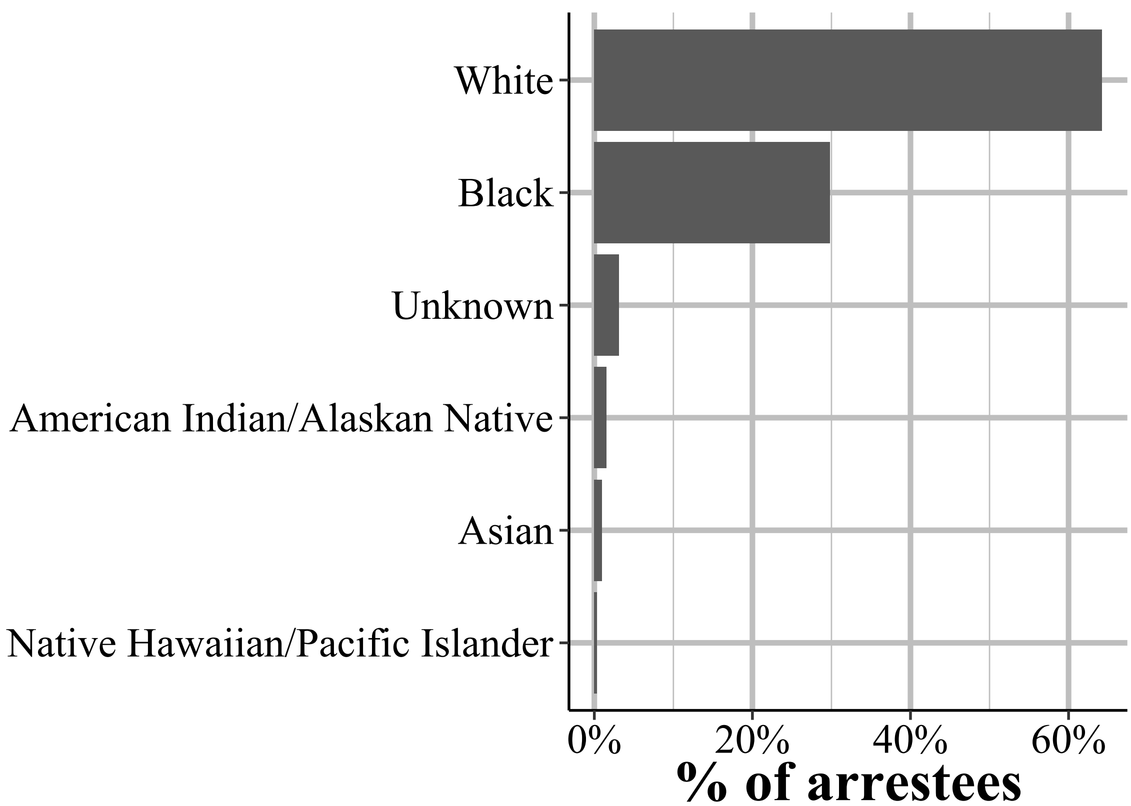

Figure 7.11: The race of all arrestees reported in the 2019 NIBRS data.

[6] I've been told that measuring race at all is itself racist so should never been done, even in research. This from a group of people who also said they have no need to actually evaluate police racial bias properly (i.e. using a regression with control variables) since they already knew the answer. If you agree with this, please don't ever do research on anything, you'll do it poorly.

7.1.8.4 Ethnicity

Finally, there is data on the race of the arrestee so we know if they are Hispanic or not. Ethnicity is so poorly used in the UCR data (e.g. UCR stopped collecting it for arrests for most years available and most agencies still don't report it) that I recommended in the UCR book against ever using it. For NIBRS there's a less less data missing so it's not as much of a problem to use ethnicity as it is with UCR data. The issue remains as to what agencies are actually reporting this data or in which scenarios this variable is reported or not even in agencies that generally do report it.

Ethnicity is an optional variable so agencies are never required to report it, so there's a greater chance that it'll be used only in non-random situations (which would make the data biased in some unknown way). There's also the question of reliability of the ethnicity data. Someone being Hispanic or not is likely just what the arrestees calls themselves or what the arresting officer perceives them to be. Both are important ways of measuring ethnicity but get at different things. Perception is more important for studies of bias, self-identification for differences among groups of people such as arrest rates by ethnicity. And the subjectivity of who is classified as Hispanic means that this measurement may differ by agency and by officer, making it imprecise.

Figure 7.12 shows the breakdown in arrests by arrestee ethnicity. Most arrestees - 72.5% - are not Hispanic. Only 10% are reported to be Hispanic but an even higher percent of arrestees - 16.8% - have an unknown ethnicity so the true percent or Hispanic or non-Hispanic arrestee may be very different than shown here.

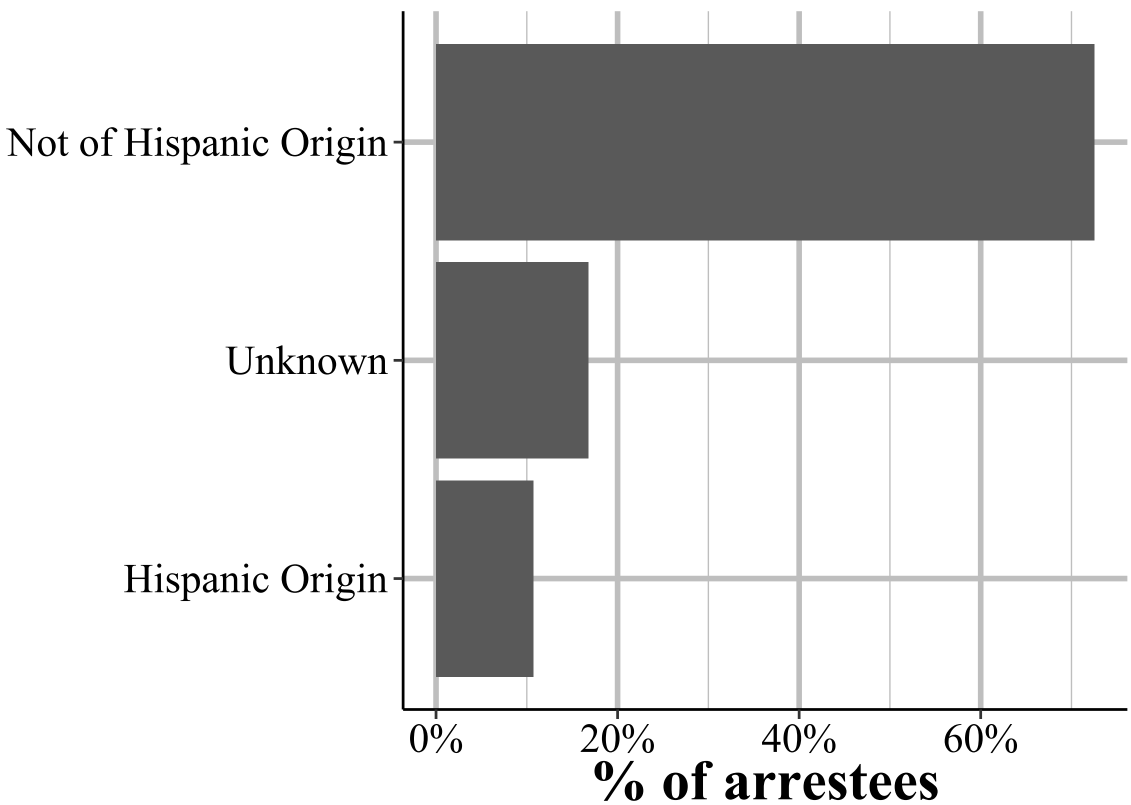

Figure 7.12: The ethnicity of all arrestees reported in the 2019 NIBRS data.

Chapter 8

Property and Window Property Segment

The Property Segment provides a bit more info than would be expected from the name. For each item involved in the crime it tells you what category that items falls into, with 68 total categories (including "other") ranging from explosives and pets to money and alcohol. It also tells you the estimated value of that item. This data covers more than just items stolen during a crime. For each item it tells you what happened to that item such as if it was stolen, damaged, seized by police (such as illegal items like drugs), recovered by police, or burned during an arson.

For drug offenses it includes the drugs seized by police. For these offenses, the data tells us the type of drug, with 16 different drug categories ranging from specific ones like marijuana or heroin to broader categories such as "other narcotics". There can be up to three different drugs included in this data - if the person has more than three types of drugs seized then the third drug category will simply indicate that there are more than three drugs, so we learn what the first two drugs are but not the third or greater drugs are in these cases. For each drug we also know exactly how much was seized with one variable saying the amount the police found and another saying the units we should we reading that amount as (e.g. pills, grams, plants).

8.1 Important variables

8.1.1 Type of property loss

8.1.2 Description of property

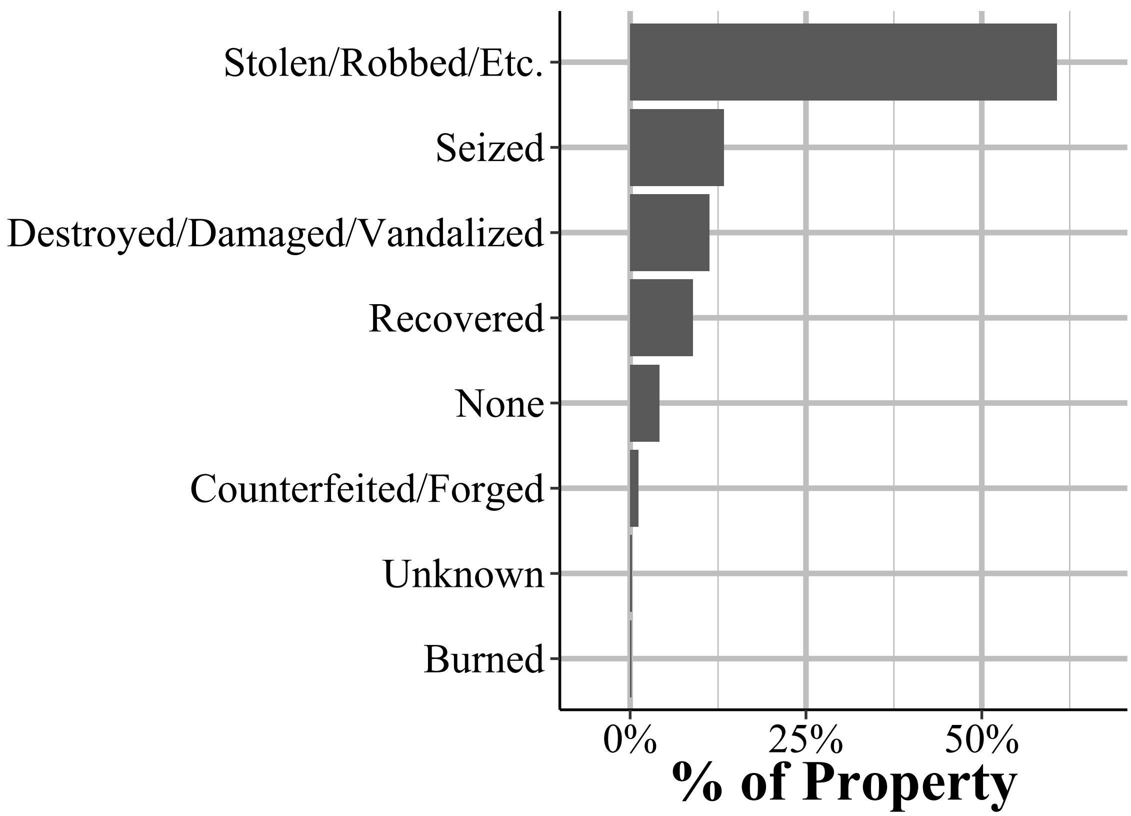

Figure 8.1: The type of loss or if the item is recovered.

Table 8.1: The number and percent of property stolen (including forcibly taken such as during a robbery) in a crime, for all offenses in 2019. Each incident can have multiple items stolen

Property	# of Property Stolen	% of Property Stolen
Other	687,291	14.31%
Money	564,574	11.75%
Purses/Handbags/Wallets	267,044	5.56%
Clothes/Furs	247,128	5.14%
Automobiles	246,591	5.13%
Credit/Debit Cards	227,460	4.74%
Tools - Power/Hand	214,257	4.46%
Vehicle Parts/Accessories	197,125	4.10%
Merchandise	190,765	3.97%
Consumable Goods	183,699	3.82%
Identity Documents	172,335	3.59%
Computer Hardware/Software	170,819	3.56%
Portible Electronic Communications	163,689	3.41%
Radios/Tvs/Vcrs	144,608	3.01%
Household Goods	127,883	2.66%
Jewelry/Precious Metals	110,807	2.31%
Firearms	92,796	1.93%
Bicycles	79,115	1.65%
Identity - Intangible	69,226	1.44%
Alcohol	53,190	1.11%
Documents - Personal Or Business	51,234	1.07%
Negotiable Instruments	44,881	0.93%
Drugs/Narcotics	44,833	0.93%
Office-Type Equipment	42,147	0.88%
Trucks	40,316	0.84%
Lawn/Yard/Garden Equipment	39,321	0.82%
Nonnegotiable Instruments	29,177	0.61%
Other Motor Vehicles	28,049	0.58%
Recreational/Sports Equipment	25,509	0.53%
Trailers	23,624	0.49%
Photographic/Optical Equipment	20,252	0.42%
Camping/Hunting/Fishing Equipment/Supplies	18,925	0.39%

Property	# of Property Stolen	% of Property Stolen
Recordings - Audio/Visual	18,680	0.39%
Building Materials	17,690	0.37%
Fuel	16,051	0.33%
Pending Inventory (Of Property)	13,327	0.28%
Heavy Construction/Industrial Equipment	12,723	0.26%
Weapons - Other	12,576	0.26%
Collections/Collectibles	10,925	0.23%
Firearm Accessories	9,993	0.21%
Musical Instruments	8,729	0.18%
Metals, Non-Precious	8,316	0.17%
Medical/Medical Lab Equipment	7,758	0.16%
Pets	6,821	0.14%
Farm Equipment	6,718	0.14%
Recreational Vehicles	5,906	0.12%
Artistic Supplies/Accessories	3,425	0.07%
Drug/Narcotic Equipment	3,221	0.07%
Chemicals	2,759	0.06%
Watercraft	2,749	0.06%
Explosives	2,444	0.05%
Gambling Equipment	2,332	0.05%
Watercraft Equipment/Parts/Accessories	2,211	0.05%
Law Enforcement Equipment	1,771	0.04%
Livestock	1,582	0.03%
Logging Equipment	1,251	0.03%
Crops	954	0.02%
Special Category	931	0.02%
Structures - Other	746	0.02%
Aircraft Parts/Accessories	483	0.01%
Structures - Single Occupancy Dwellings	411	0.01%
Buses	406	0.01%
Structures - Storage	281	0.01%
Aircraft	237	0.00%
Structures - Commercial/Business	208	0.00%
Structures - Other Dwellings	169	0.00%
Structures - Industrial Manufacturing	118	0.00%
Structures - Public/Community	116	0.00%
Total	4,803,688	100%

Table 8.2: The number and percent of property seized by police (excludes recovering property that was stolen, for all offenses in 2019. Each incident can have multiple items seized.

Property	# of Property Seized	% of Property Seized
Drugs/Narcotics	667,424	63.21%
Drug/Narcotic Equipment	288,418	27.31%
Other	27,845	2.64%
Money	21,852	2.07%
Firearms	9,664	0.92%
Portible Electronic Communications	6,460	0.61%
Automobiles	3,196	0.30%
Firearm Accessories	2,682	0.25%
Documents - Personal Or Business	2,437	0.23%
Weapons - Other	2,279	0.22%
Office-Type Equipment	2,279	0.22%
Purses/Handbags/Wallets	2,104	0.20%
Identity Documents	1,758	0.17%
Computer Hardware/Software	1,533	0.15%
Consumable Goods	1,419	0.13%
Clothes/Furs	1,286	0.12%
Alcohol	1,258	0.12%
Negotiable Instruments	1,227	0.12%
Household Goods	1,205	0.11%
Credit/Debit Cards	1,032	0.10%
Recordings - Audio/Visual	1,018	0.10%
Nonnegotiable Instruments	971	0.09%
Vehicle Parts/Accessories	945	0.09%
Explosives	910	0.09%
Tools - Power/Hand	785	0.07%
Radios/Tvs/Vcrs	549	0.05%
Heavy Construction/Industrial Equipment	395	0.04%
Merchandise	390	0.04%
Jewelry/Precious Metals	354	0.03%
Gambling Equipment	293	0.03%
Medical/Medical Lab Equipment	260	0.02%
Other Motor Vehicles	201	0.02%

Property	# of Property Seized	% of Property Seized
Photographic/Optical Equipment	175	0.02%
Trucks	166	0.02%
Pending Inventory (Of Property)	145	0.01%
Identity - Intangible	115	0.01%
Chemicals	109	0.01%
Bicycles	101	0.01%
Camping/Hunting/Fishing Equipment/Supplies	98	0.01%
Recreational/Sports Equipment	80	0.01%
Special Category	60	0.01%
Law Enforcement Equipment	57	0.01%
Metals, Non-Precious	48	0.00%
Collections/Collectibles	42	0.00%
Farm Equipment	38	0.00%
Structures - Storage	34	0.00%
Trailers	33	0.00%
Artistic Supplies/Accessories	31	0.00%
Lawn/Yard/Garden Equipment	24	0.00%
Recreational Vehicles	23	0.00%
Structures - Other	19	0.00%
Crops	19	0.00%
Building Materials	18	0.00%
Fuel	15	0.00%
Musical Instruments	14	0.00%
Aircraft	13	0.00%
Aircraft Parts/Accessories	11	0.00%
Structures - Single Occupancy Dwellings	7	0.00%
Pets	7	0.00%
Watercraft Equipment/Parts/Accessories	6	0.00%
Structures - Public/Community	6	0.00%
Livestock	6	0.00%
Watercraft	5	0.00%
Buses	4	0.00%
Logging Equipment	4	0.00%
Structures - Other Dwellings	2	0.00%
Structures - Industrial Manufacturing	2	0.00%
Structures - Commercial/Business	1	0.00%
Total	1,055,967	100%

8.1.3 Value of stolen property

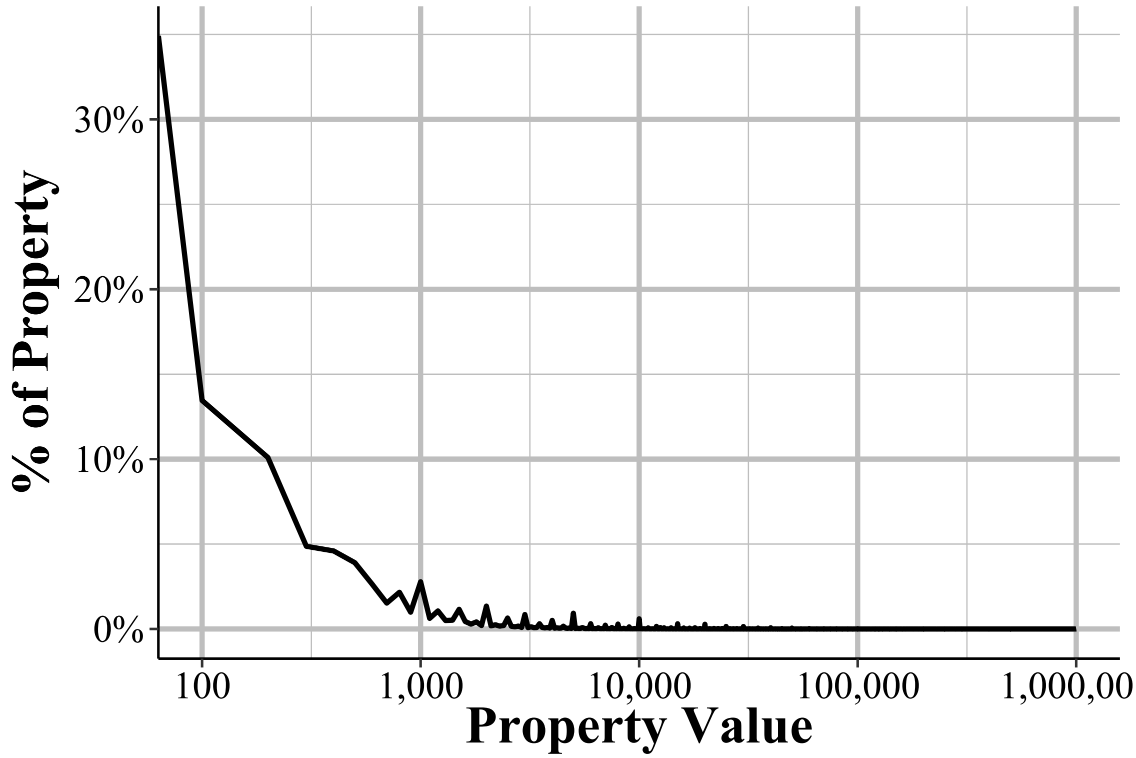

Figure 8.2: The distribution of the value of property stolen. Values are capped at $1,000,000 and each value is rounded to the nearest 100$. The x-axis is set on the log scale as this distribution is hugely right skewed.

8.1.4 Date property was recovered

8.1.5 Drugs

This segment also provides information about drugs seized by the police. This also includes cases where the police would have seized the drugs if the offender didn't destroy it somehow, such as flushing it down the toilet. For each drug seized there is information on what type of drug it was and how much of the drug was seized. There can be up to three drugs seized in an incident and data is available for each type of drug. The exception, however, is when there are more than four drugs seized in an incident. In that case, info on the third drug just says that there were more than three drugs involved. So you'd have info on the first two drugs but none on the third through however many (and it doesn't say how many) drugs. For the below table and figure I only look at the first drug seized, so the numbers shown are undercounts.

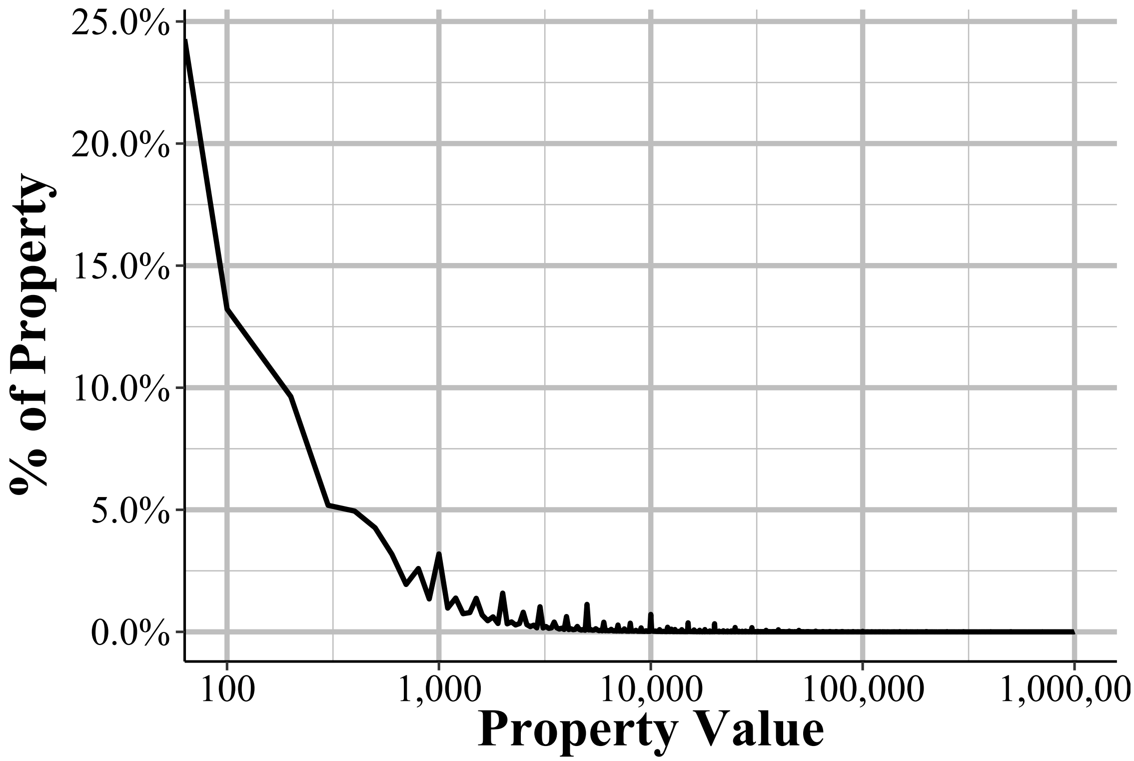

Figure 8.3: The incident-level distribution of the value of property stolen. As values are aggregated to the incident-level, these are higher than the above graph which shows each item individually. Values are capped at $1,000,000 and each value is rounded to the nearest 100$. The x-axis is set on the log scale as this distribution is hugely right skewed.

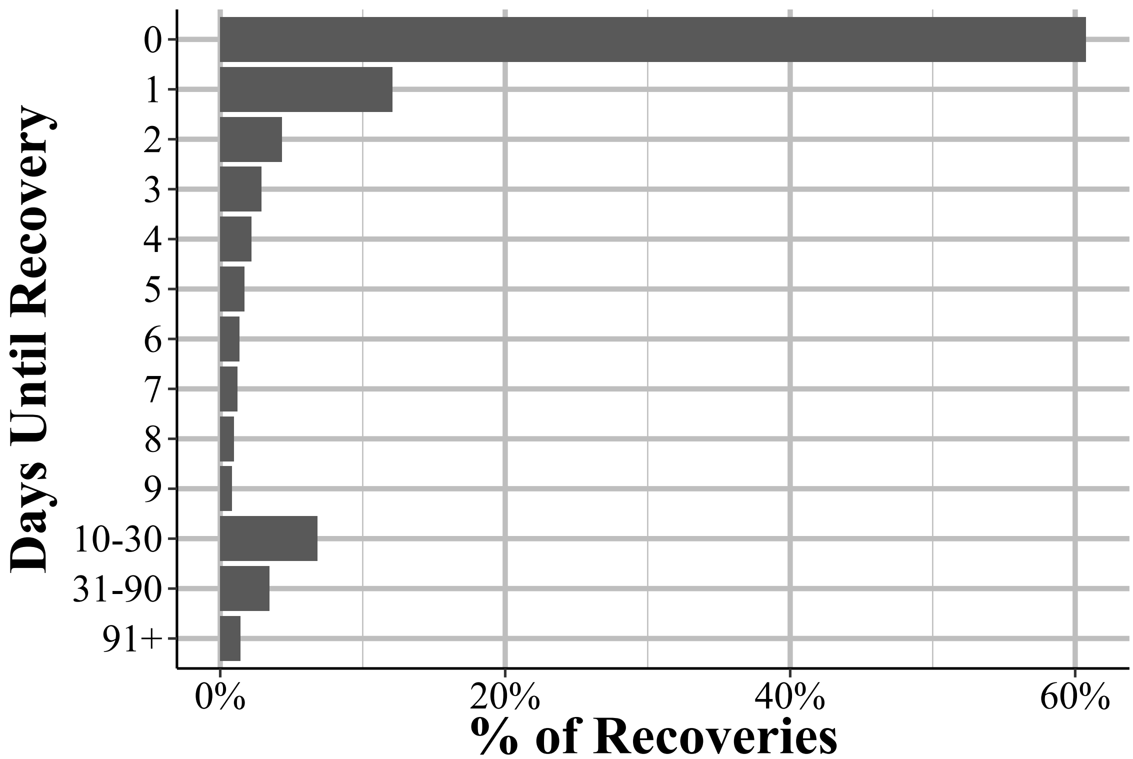

Figure 8.4: The distribution of the number of days from the incident to the property recovered date. In 2019 the maximum days from incident to arrest was 450 days. Zero days means that the arrest occurred on the same day as the incident.

The ordering of drugs when there are multiple drugs is based on how much drugs were recovered and the seriousness of the drugs.For example, heroin is probably considered more serious than marijuana, but overall ranking of drugs is a subjective assessment depending on the department. Is heroin more serious than heroin? That decision likely varies by the agency and their situation in regards to what drugs they often seize.

8.1.5.1 Suspected drug type

The drugs in NIBRS are the "suspected drug types" which means that they are what the police believe the drugs to be, even if there is not definitive proof (such as a crime lab testing for what type of drug it is) that the drug is what they say it is. There are 15 possible drug types in NIBRS (16 when including "unknown drug type") and each is shown in Table 8.3 along with how often they occur in the data. Some of these drug types are specific enough to only include a single drug while others are groups of drug types, such as hallucinogens or stimulants (and they include all of these types other than specifically stated drugs).

Not too surprising, marijuana is the most common drug seized at 47% - or 455k incidents with it seized - of the data. This is followed by amphetamines/methamphetamines (including what we'd normally just call meth) at 20.7% and then heroin at 8.5%. Interestingly, cocaine and crack cocaine (which is always separate categories) both occur in 5.09% of drugs seized. Given the large disparity in sentences for these types of drugs, and that "crack wars" were a major purported cause of violent crime in the 1980s and 1990s, I expected crack cocaine to be much more common than normal cocaine. The remaining drug types are all less than 5% of drugs seized each and has some groupings of drug types (e.g. stimulants) rather than specific drug types (though some of these categories, such as LSD, are specific drugs).

Table 8.3: The number and percent of drugs seized by police by type of drug.

Drug Type	# of Drugs	% of Drugs
Marijuana	455,429	46.69%
Amphetamines/Methamphetamines	201,716	20.68%
Heroin	82,437	8.45%
Cocaine (All Forms Except Crack)	49,699	5.09%
Crack Cocaine	49,646	5.09%
Unknown Type Drug	39,836	4.08%
Other Narcotics: Codeine, Demerol, Dihydromorphinone Or Dilaudid, Hydrocodone Or Percodan, Methadone, Etc.	37,401	3.83%

Drug Type	# of Drugs	% of Drugs
PCP	30,348	3.11%
Hashish	7,046	0.72%
Other Hallucinogrens: Bmda (White Acid), Dmt, Mda, Mdma, Mescaline Or Peyote, Psilocybin, Stp, Etc.	6,256	0.64%
Other Depressants: Glutethimide Or Doriden, Methaqualone Or Quaalude, Pentazocine Or Talwin, Etc.	5,165	0.53%
Other Stimulants: Adipex, Fastine And Ionamin (Derivatives of Phentermine), Benzedrine, Didrex, Methylphenidate Or Ritalin, Phenmetrazine Or Preludin, Tenuate, Etc.	3,368	0.35%
Opium	2,984	0.31%
LSD	1,624	0.17%
Morphine	1,297	0.13%
Barbiturates	1,276	0.13%
Total	975,528	100%

8.1.5.2 Amount of drugs

For each drug type we know exactly how much was seized (at least we do other than for the 6.7% where the amount is "not reported"). Since different drug types are measured in different ways, this data also tells us what units the amount seized is in. So you'll need to look at both the amount and the units to understand how much drugs were actually seized. The possible units are listen below:

- Dosage Unit/Items (Pills, Etc.)
- Fluid Ounce
- Gallon
- Gram
- Kilogram
- Liter
- Milliliter
- Number of Plants
- Ounce

- Pound

Once you know the units you can look at the amount of drugs seized. The amount is specific up to the thousandths place though the integer and the numbers after the decimal point are actually in different columns in the data. For example, if police seized 1.257 grams of heroin, the 1 gram and the 0.257 grams would be in separate columns. As an example, we'll look at Figure 8.5 which shows the number of grams seized for marijuana. This is only looking at the column with the amount in integers, so parts of a gram are excluded (but are available in the data). This also excludes any case where the marijuana seized was measured in a unit other than gram, such as number of plants, ounces, or pounds. Even though the data shows the number of grams as actual integers, not grouped at all, I do group the larger values together to make the graph simpler.

So with those caveats, we can see what amounts of marijuana, measured in grams, are most frequently seized. Generally, the amount of marijuana seized is in small amounts with 22.5% being 1-2 grams (since we don't include the parts of a gram we can only say that it is 1 to 1.999 grams) and 18.6% being less than one gram. As the amount of drugs increase, the percent of seizures that involve this number of drugs decreases. It's a bit curious that they include grams for values over 28 since 28.3495 grams is one ounce so it'd make sense to just start reporting in units of ounces instead of just increasingly large number of grams.

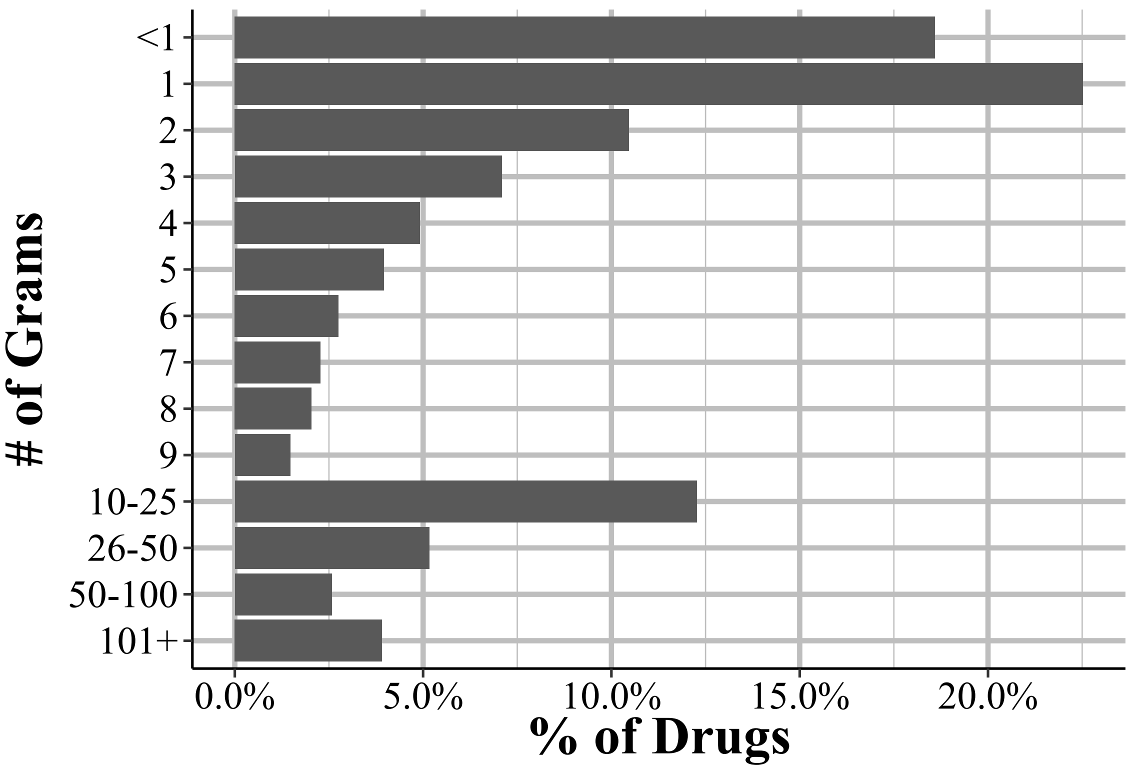

Figure 8.5: For drugs seized that are measured in grams, this figure shows the distribution in the number of grams seized. Values over 10 grams are grouped together for easier interpretation of lower values of drugs seized.