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

Jacob Kaplan

2021-06-09

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 murder, 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 county - 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 - 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) we used. And we have demographic information about offenders, and whether they were arrested (including the type of arrest such as warrant arrest). From this we can figure out victimization rates based on certain victim characteristics, though not all of them that we may think are important.

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 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 nine chapters: this chapter, an intro chapter briefly summarizing each segment files and going over overall issues with NIBRS data, and seven chapters each covering one of the seven UCR datasets. 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 (including when to not use them at all) - this will be the bulk of each chapter.

1.3 Citing this book

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

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

BibTeX format:

```
@Manual{ucrbook,
  title = {National Incident-Based Reporting System (NIBRS) Data: A Practitioner's Guide
```

[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.

```
  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. "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 UCR 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 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. This source 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 day-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. As the data provides the exact age (in years) of each offender, this 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 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 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, 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. 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

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

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.

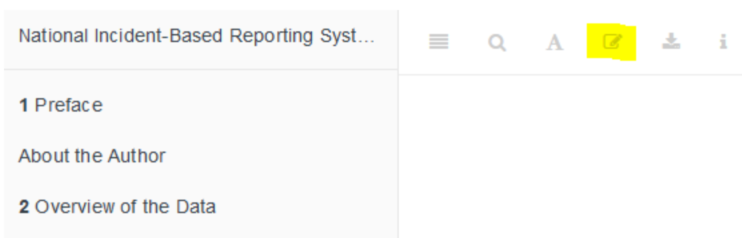


Figure 1.1: The edit button for how to make edits of this book.

[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 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. He is the Chief Data Scientist at the Research on Policing Reform and Accountability RoPRA at Princeton University. 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.

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, 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 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.

NIBRS data provides detailed information on every crime reported to the police, including victim and offender demographics, whether the offender was arrested (and type of arrest or type of "exceptional clearance"), the crime date and hour, 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 to UCR data. We now have a far deeper understanding of crime and this 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, this 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), fewer than half of the about 18,000

[1]This data has been expanded since it began in 1929 to include information on arrests, hate crimes, and stolen property. For more on this, please see my book Uniform Crime Reporting (UCR) Program Data: A Practitioner's Guide at ucrbook.com.

police agenices in the United States. This is an even larger problem that it seems are 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 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 the high level of detail relative to UCR data is a costly and timely switch - we will be flying blind for most crime in the country.

So there are really three major problems with NIBRS data, both 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 is Spring 2021, 2019 UCR and NIBRS data are the latest years available. Especially given the huge crime changes during 2020 - and whose violent crime increases and continuing into 2021 - losing standardized crime data for most cities is a very bad thing. 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. 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. Once all agencies report - though it's doubtful that'll ever occur - 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.[2]

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

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

"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. This means that you'll need to understand how to deal with multiple datasets, and subset and merge them as needed.

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.

A word of caution. To date fewer than half of agencies report NIBRS data. 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 later than is covered in this book, which ends with 2019 data as 2019 is the most recent year available. 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.

2.0.1 Choosing a unit of analysis

A major benefit of UCR data is that you have very limited choices. If you wanted to measure crime you only choice was to use their monthly aggregated city-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.

Consider, for example, that you want to measure rape in a city. Even if we somehow solve the issue of victims not reporting their rapes, we still have a few different ways of even measuring rape. Let's use an incident where four men rape a single woman as an example. If you're interested in measuring robbery 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. 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 robbery. Finally we could look at the offense-level. This data actually makes it unclear what act each offender actually committed so we'd likely measure it as one offense per offender in each incident.[3]

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 understand crimes where rape happened alongside another, more serious, offense. 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. In most cases, however, only one offense occurs per criminal incident, so the opportunity to explore co-occurrence is relatively limited.

2.1 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 was 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 combine them together to get the complete data you want.

[3] For four people involved in a rape, do we care if all four actually committed the crime? Does it matter if only three committed the rape and one merely stood by?

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

2.1.1 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 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

2.1.2 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

2.2 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.

2.2.1 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 very, 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 here 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.

2.2.1.1 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 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.

2.2.2 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 employeeby 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.

2.3 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 comes as multiple different files that each provide different information about a criminal 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 segments and how they relate to each other.[4]. There are also three segments called "window segments" - there is one for arrestees, one of exceptional clearances, and one for property - that do not have an associated segment with them, they only have the information available in the given window segment. We'll talk about window segments more at the end of this section.

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. Alongside being report (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

[4]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

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) and an incident number (which is just a randomly generated unique identifier for that incident) so you can merge the files 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 arrestee 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.

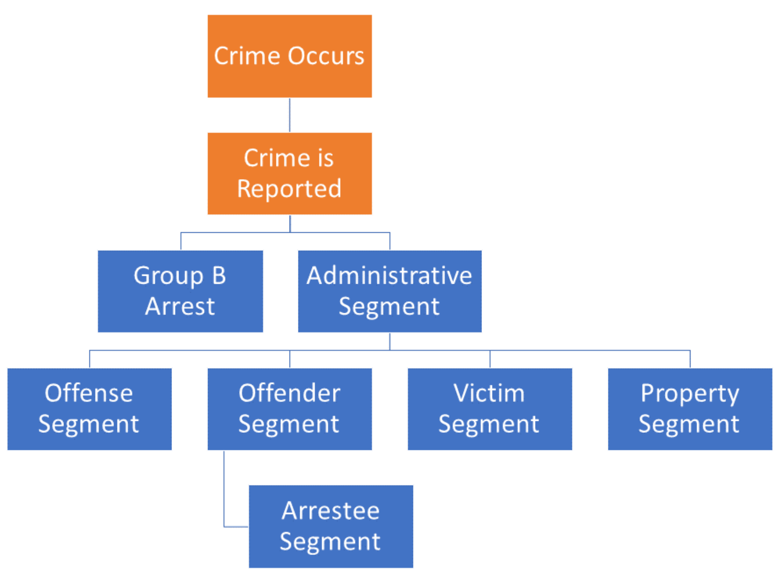


Figure 2.1: The association of each segment file in the NIBRS dataset.

2.3.1 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, 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 was 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.

2.3.2 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.

2.3.3 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.

2.3.4 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 ("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.

2.3.5 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. 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).

2.3.6 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 (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).

2.3.7 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.[5] 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.

2.4 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 - agencies move to NIBRS in 2021, we'll start having much more detail from a very representative sample of agencies.[6]. 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, assuming most agencies report data starting in 2021.

We'll look here at how many agencies report at least one crime 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 reporting 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 This is an overcount, however, this constitutes relatively low participation among agencies. And, as agencies may report for part of the year meaning that we actually have fewer fully reporting agencies that this

[5] 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.

[6] 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.

graph applies.

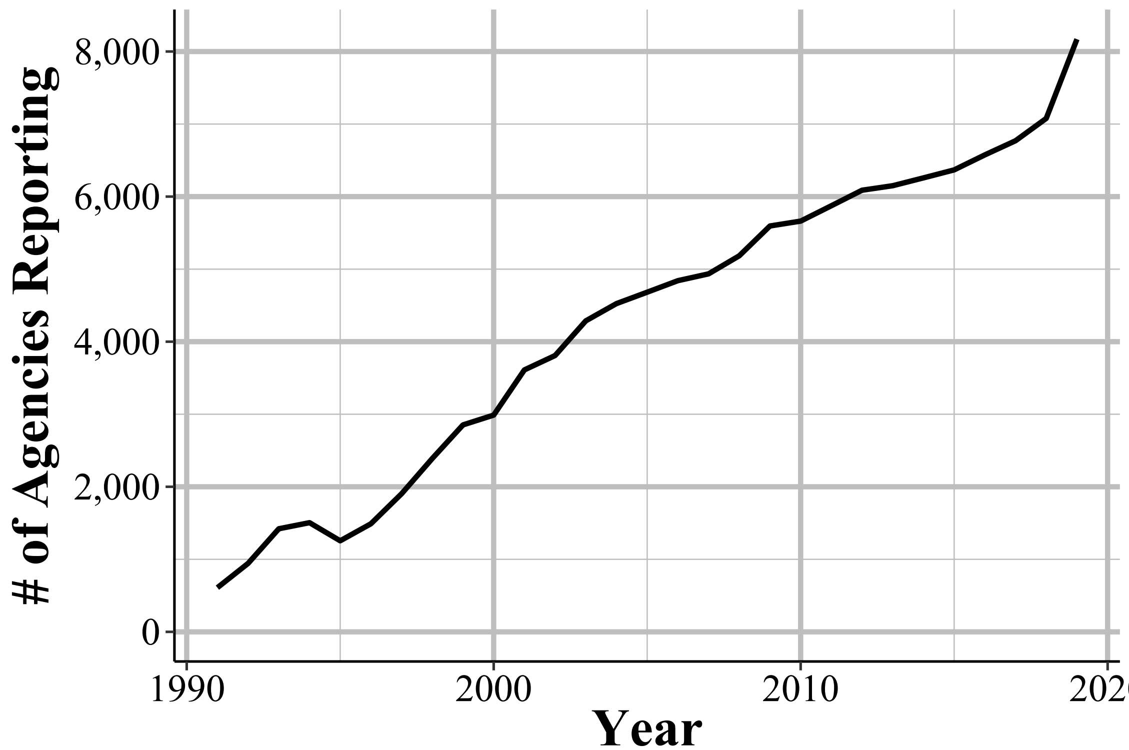


Figure 2.2: The annual number of agencies reporting at least one incident in that year.

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

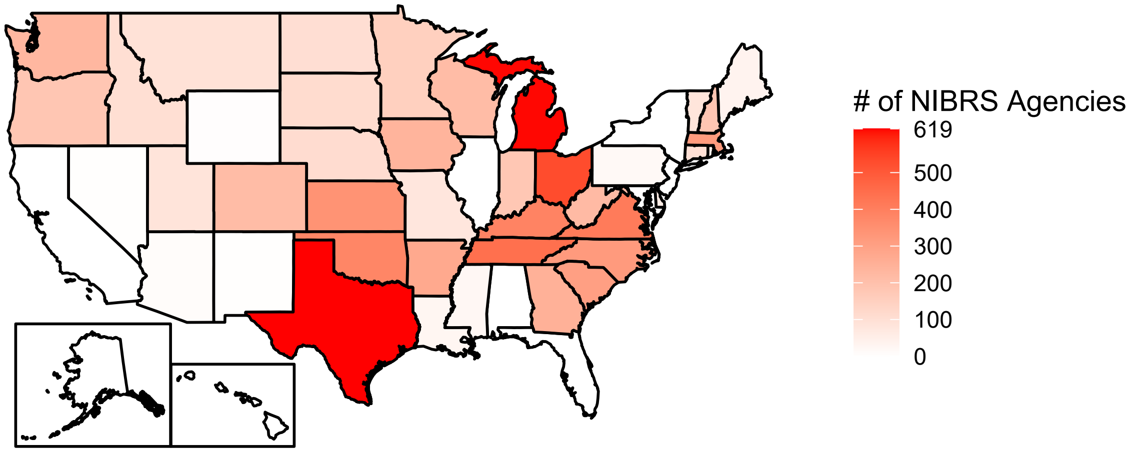


Figure 2.3: The number of agencies in each state that reported at least one crime in 2019 to NIBRS.

56% of agencies reporting. So when using NIBRS data, keep in mind that you have very good coverage of certain states, and very poor coverage of other states. And the low - or no - reporting states are systematically high population states.

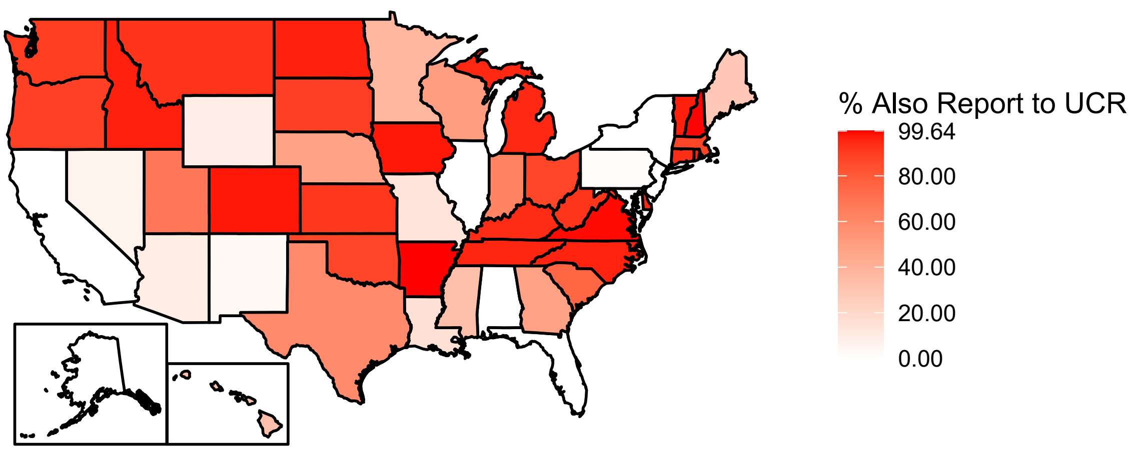


Figure 2.4: Agencies in each state reporting at least one crime to NIBRS in 2019 as a percent of agencies that reported UCR Offenses Known and Clearances by Arrests data in 2019.

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	percent
Alabama	1	352	0.28	%
Alaska	0	33	0.00	%
Arizona	11	114	9.65	%
Arkansas	277	278	99.64	%
California	0	738	0.00	%
Colorado	217	222	97.75	%
Connecticut	100	107	93.46	%
Delaware	60	63	95.24	%
District of Columbia	1	3	33.33	%
Florida	0	678	0.00	%
Georgia	251	522	48.08	%
Hawaii	1	3	33.33	%
Idaho	104	108	96.30	%
Illinois	1	739	0.14	%
Indiana	180	289	62.28	%
Iowa	240	246	97.56	%
Kansas	343	376	91.22	%
Kentucky	389	413	94.19	%
Louisiana	30	188	15.96	%
Maine	39	135	28.89	%
Maryland	2	156	1.28	%
Massachusetts	317	363	87.33	%
Michigan	616	650	94.77	%
Minnesota	153	409	37.41	%
Mississippi	26	79	32.91	%
Missouri	80	576	13.89	%
Montana	95	103	92.23	%
Nebraska	113	236	47.88	%
Nevada	4	62	6.45	%
New Hampshire	187	188	99.47	%
New Jersey	0	578	0.00	%
New Mexico	4	121	3.31	%
New York	0	572	0.00	%

State	NIBRS Agencies	UCR Agencies	% of UCR Agencies	percent
North Carolina	320	333	96.10	%
North Dakota	106	110	96.36	%
Ohio	529	604	87.58	%
Oklahoma	381	437	87.19	%
Oregon	184	208	88.46	%
Pennsylvania	21	1477	1.42	%
Rhode Island	47	49	95.92	%
South Carolina	306	405	75.56	%
South Dakota	115	129	89.15	%
Tennessee	443	465	95.27	%
Texas	619	1053	58.78	%
Utah	88	129	68.22	%
Vermont	86	89	96.63	%
Virginia	411	415	99.04	%
Washington	231	257	89.88	%
West Virginia	221	240	92.08	%
Wisconsin	215	428	50.23	%
Wyoming	5	55	9.09	%

2.5 How to identify a particular agency (ORI codes)

In the UCR 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.[7] 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.

[7] I will refer to this an "ORI", "ORI code", and "ORI number", all of which mean the same thing.

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.

2.6 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" but is basically just instructions for whatever software you use (R, SPSS, Stata, SAS can all do this) - which characters are certain columns. For example, in this data the first 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).[8] So we can read this

[8] 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.

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).

```
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
0501AL0010000JZPQUFEM9CTD19910804 0136MW
0701AL00100006PMJW5ST6MQ-1991080301    O90D01    31MWUR
0101AL00100006PMJWOVCSMQ-19910822   010010100    N
0201AL00100006PMJWOVCSMQ-19910822240CN  18
0301AL00100006PMJWOVCSMQ-199108227030000040 00        01
0401AL00100006PMJWOVCSMQ-19910822001240
I25MWUN
0501AL00100006PMJWOVCSMQ-19910822 0100UU
0701AL00100006PMJW2CPDMQ-1991081401    T90Z01    40MBUR
0101AL00100006PMIWS9VZMQ-19910917 16020010100    N
0201AL00100006PMIWS9VZMQ-19910917220CN  20  F
0201AL00100006PMIWS9VZMQ-19910917290CN  20
0301AL00100006PMIWS9VZMQ-19910917416000000040
0301AL00100006PMIWS9VZMQ-19910917777000000130
0401AL00100006PMIWS9VZMQ-19910917001220290
I00UUUR
0501AL00100006PMIWS9VZMQ-19910917 0100UU
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.

The second important thing to know about reading in a fixed-width ASCII file is something called a "value label."[9] 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 as efficient as possible, it often replaces longer

[9]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. UCR data, however, does not have this and does not indicate when values are missing in this manner.

values with shorter one and provides a translation for the software to use to convert it to the proper value when reading it. "Alabama" is more characters than "01" so it saves space to say "01" and just replace that with "Alabama" later on. So "01" would be the "value" and "Alabama" would be the "label" that it changes to once read.

Fixed-width ASCII files may seem awful to you reading it today, and it is awful to use. But it appears to be an efficient way to store data back many decades ago when data releases began but now is extremely inefficient - in terms of speed, file size, ease of use - compared to modern software so I'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.[10] **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.[11] 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.[12] Each NIBRS segment has over a dozen columns and potentially dozens on value labels.[13] 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.[14] 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.[15]

[10] For those interested in reading in this type of data, please see my R package asciiSetupReader.

[11] 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.

[12] They do provide the instructions in a PDF but you still need to actually make the file yourself. You cannot use their instructions.

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

[14] Other than in my own UCR book.

[15] For evidence of this, please see any of the openICPSR pages for my detail 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 was 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

3.1.1 The incident report date

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

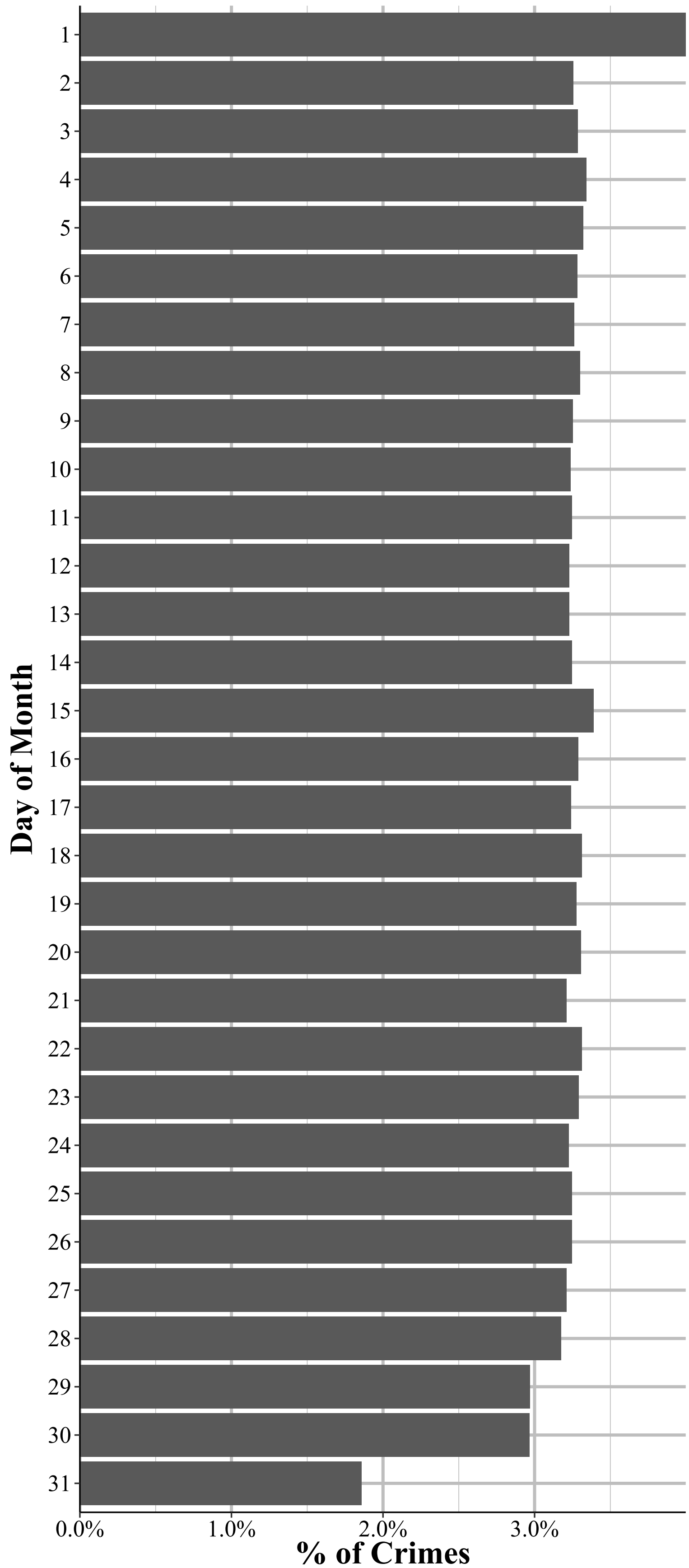


Figure 3.1: The percent of crimes that occur (the day of the incident, even if the crime wasn't reported that day) each day of the month for all agencies reporting to NIBRS in 2019.

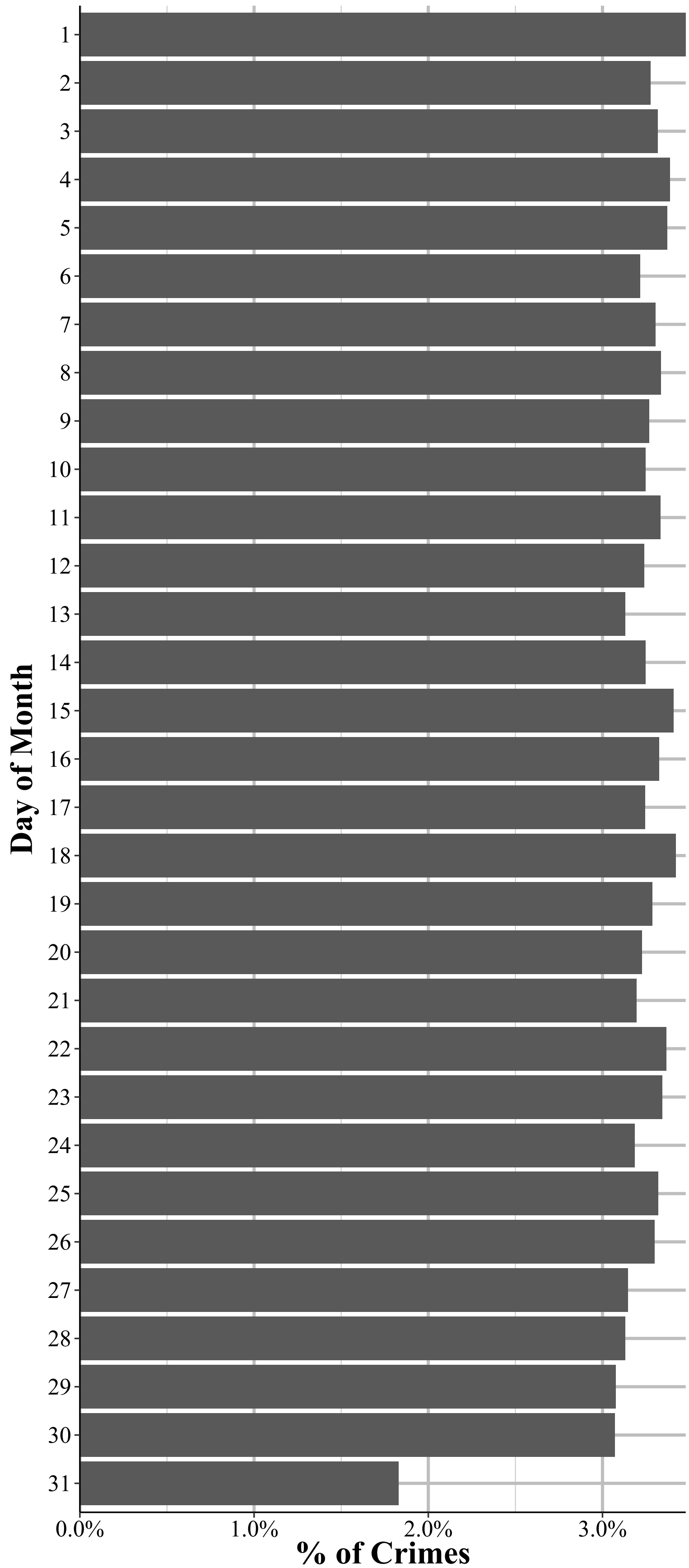


Figure 3.2: The percent of crimes that are reported (the day of the report, even if not the day of the incident) each day of the month for all agencies reporting to NIBRS in 2019.

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.

3.1.3 Exceptional clearance

3.1.4 Number of other segments

3.1.4.1 Offense segments

3.1.4.2 Victim segments

3.1.4.3 Offender segments

3.1.4.4 Arrestee segments

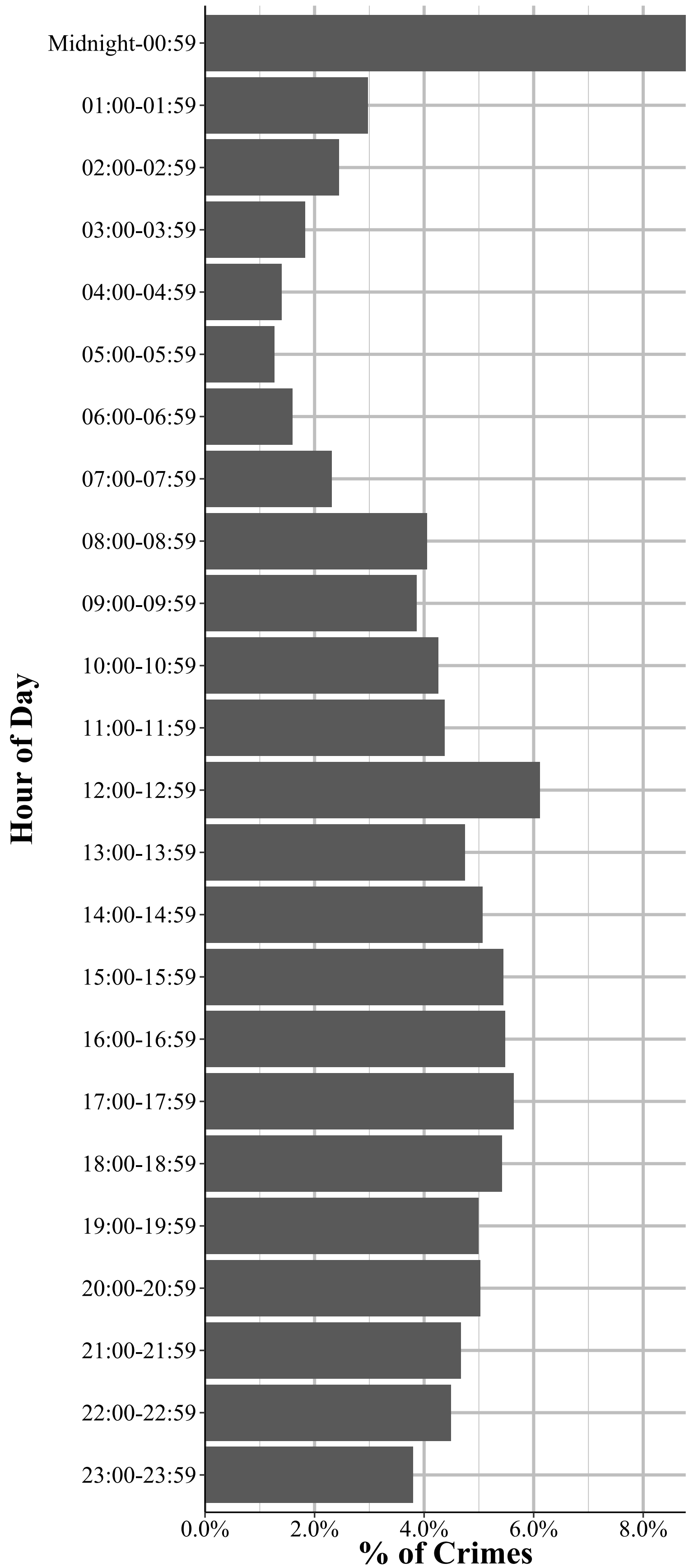


Figure 3.3: The percent of crimes that are reported each hour for all agencies reporting to NIBRS in 2019.

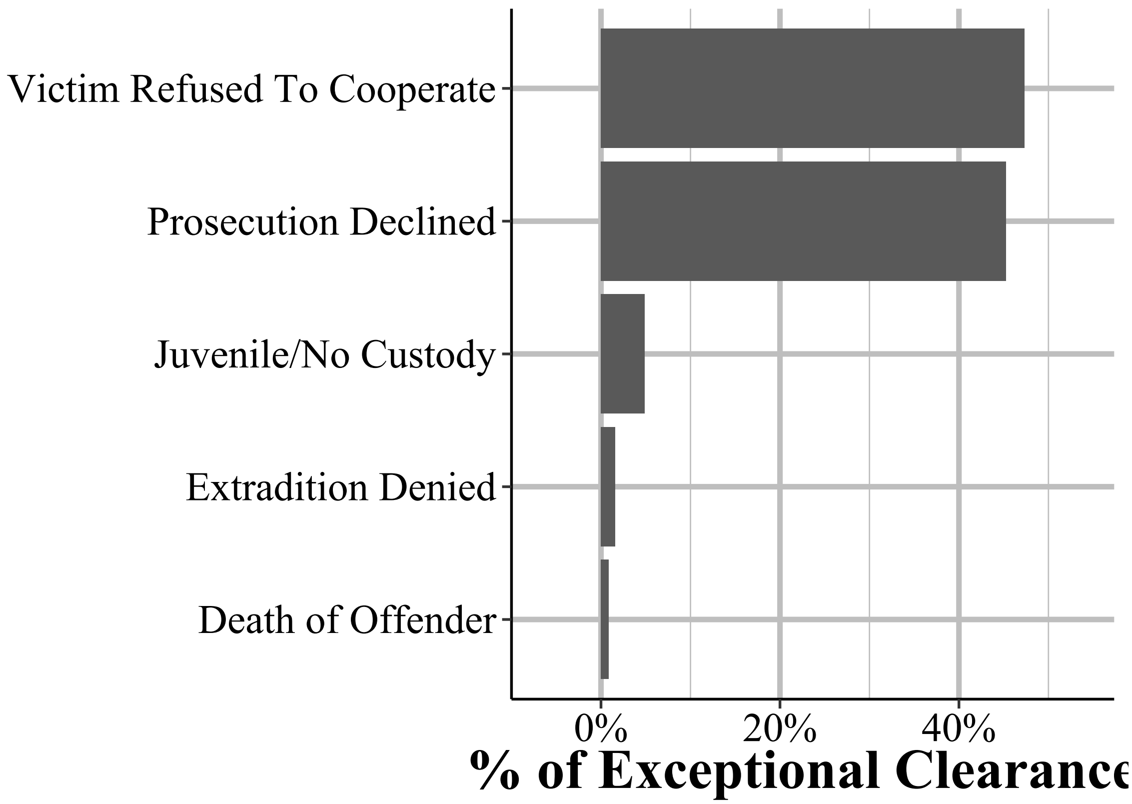


Figure 3.4: The percent of crimes that are reported each hour for all agencies reporting to NIBRS in 2019.

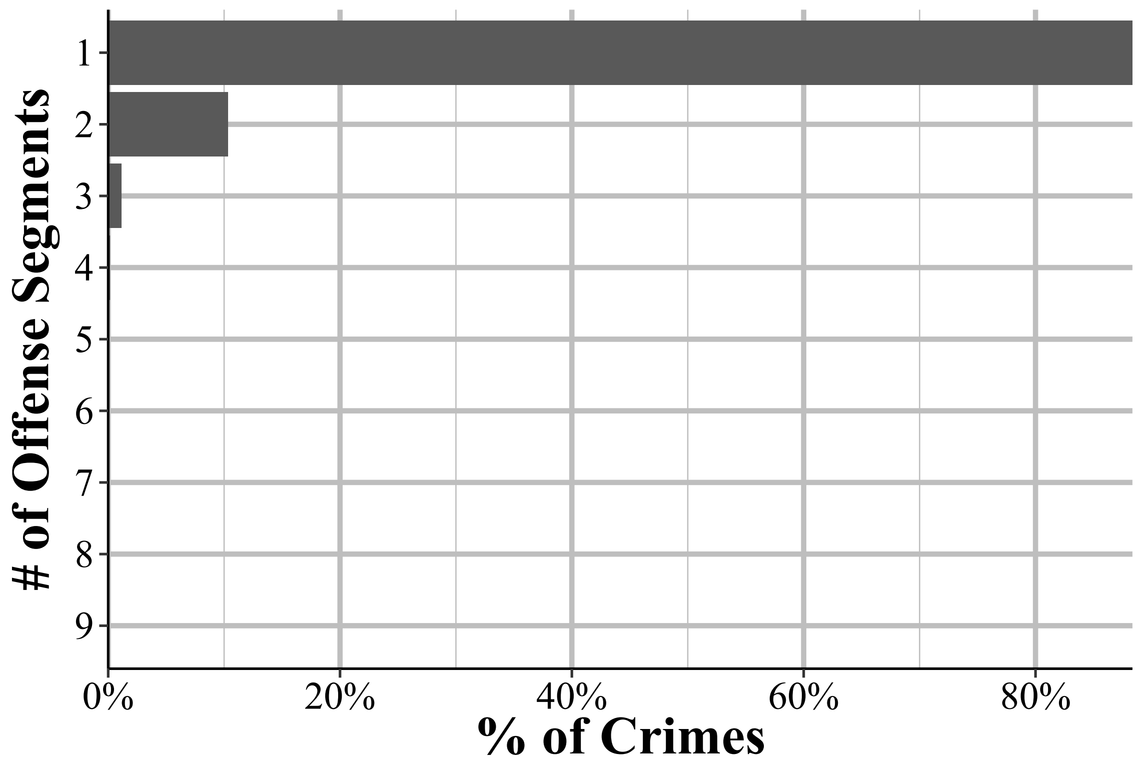


Figure 3.5: The percent of crimes that are reported each day of the month for all agencies reporting to NIBRS in 2019.

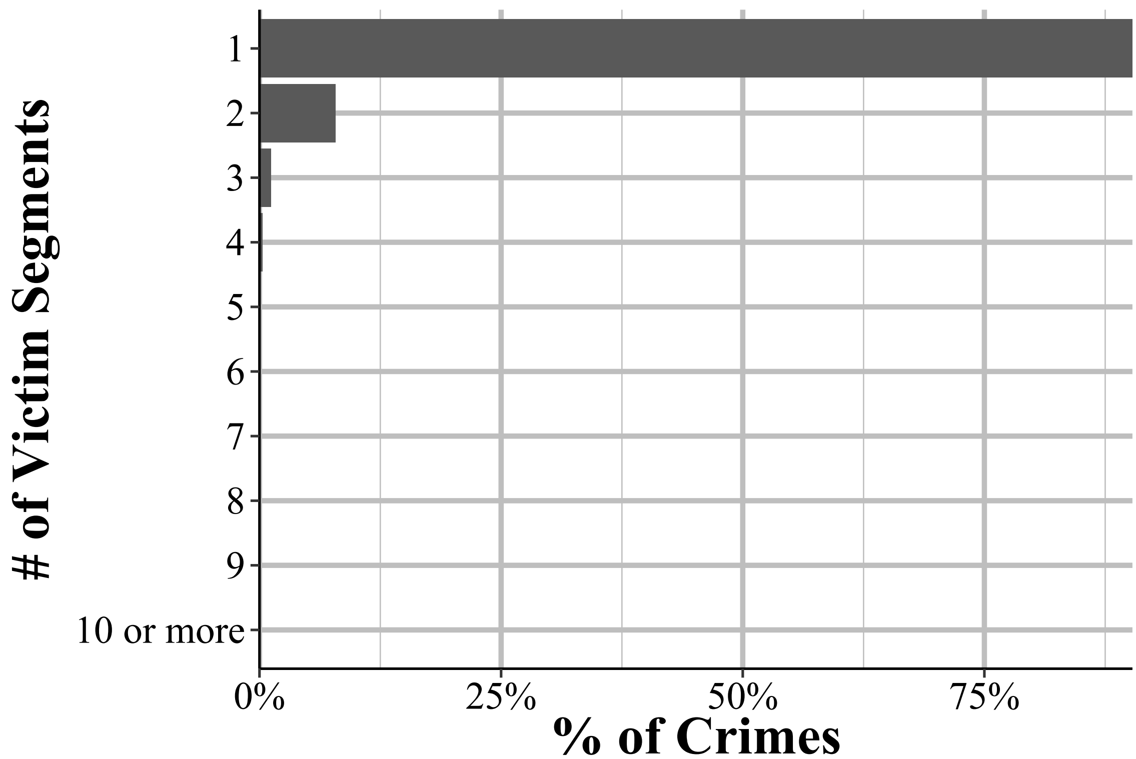


Figure 3.6: The percent of crimes that are reported each day of the month for all agencies reporting to NIBRS in 2019.

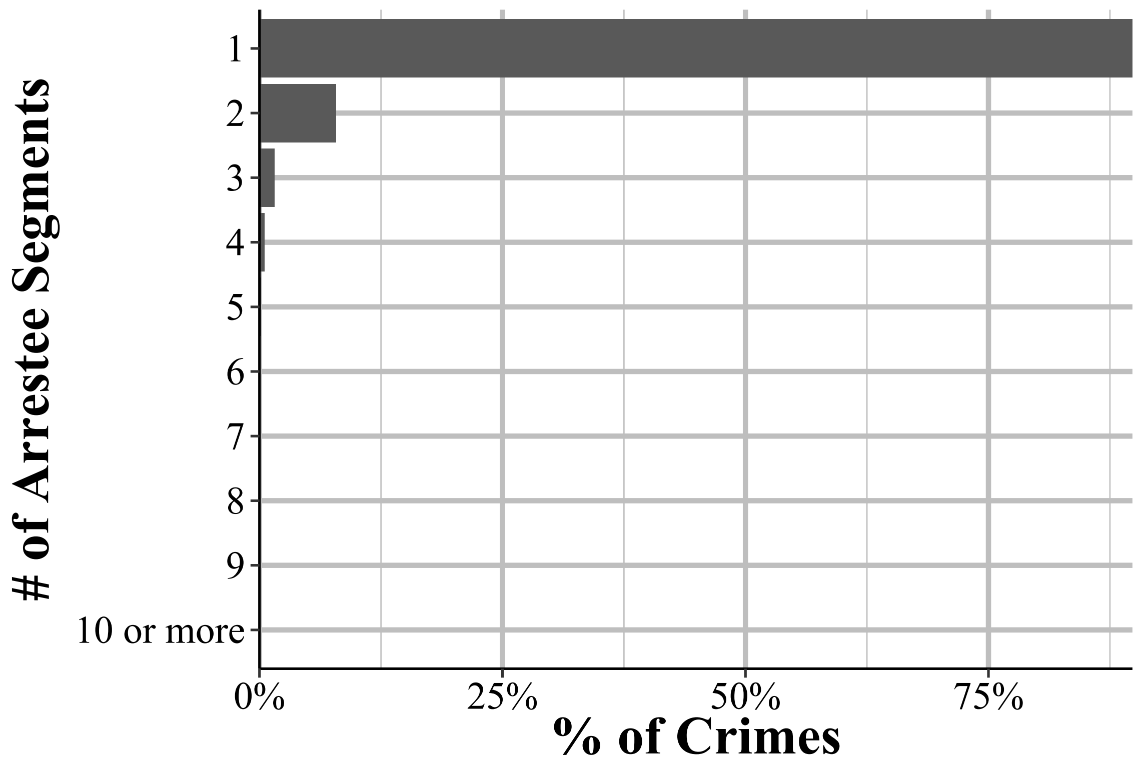


Figure 3.7: The percent of crimes that are reported each day of the month for all agencies reporting to NIBRS in 2019.

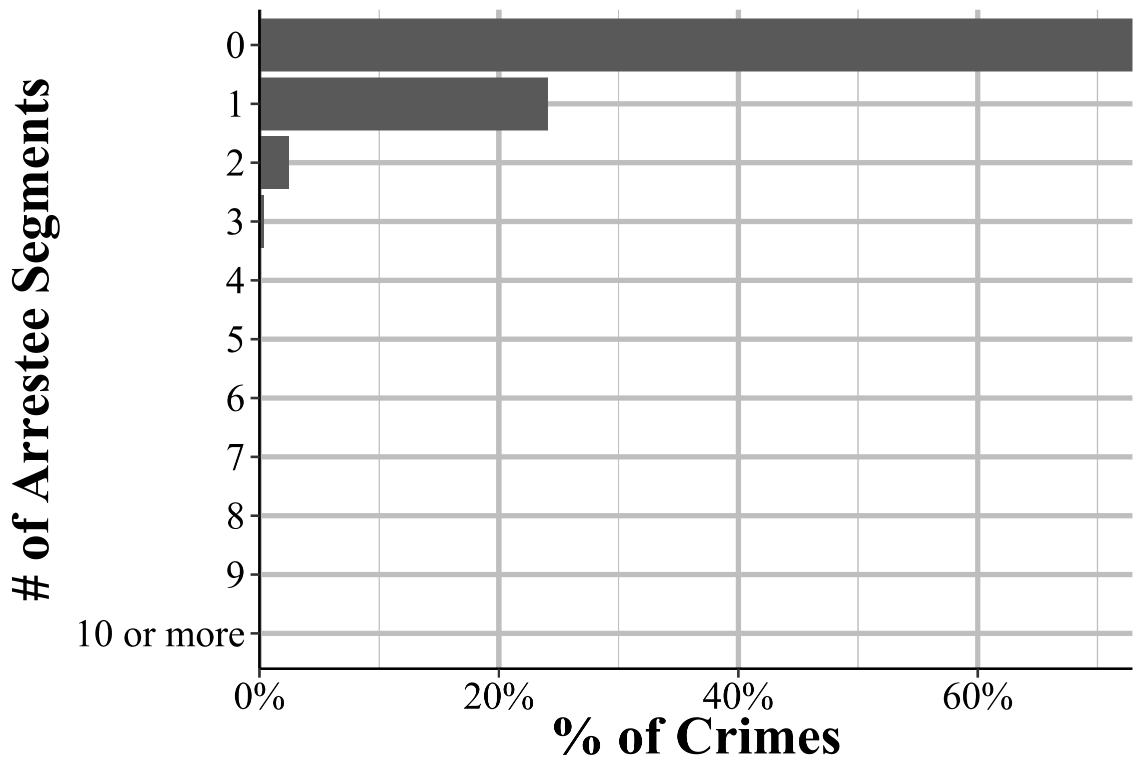


Figure 3.8: The percent of crimes that are reported each day of the month for all agencies reporting to NIBRS in 2019.

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

4.1.1 Crime category

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

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%
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%

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	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%

- 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 4.3: 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 Drug, alcohol, or computer use

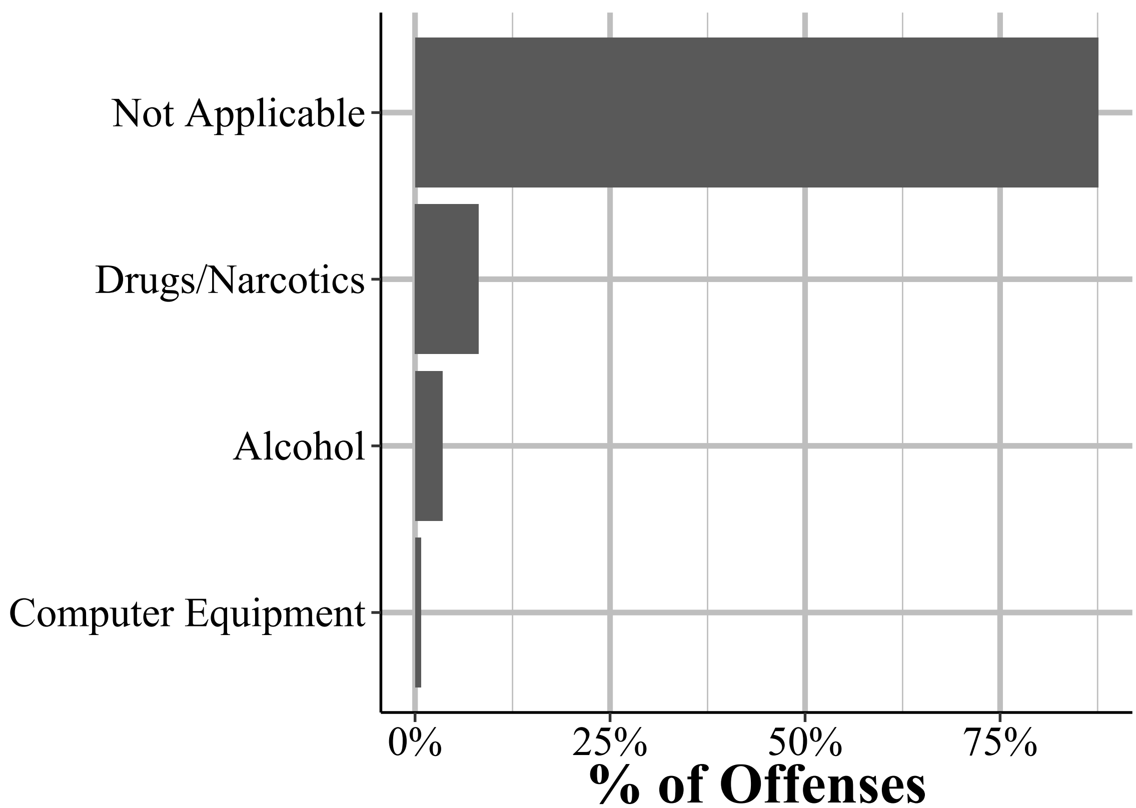


Figure 4.1: The distribution of drug, alcohol, or computer use for all offenses in 2019.

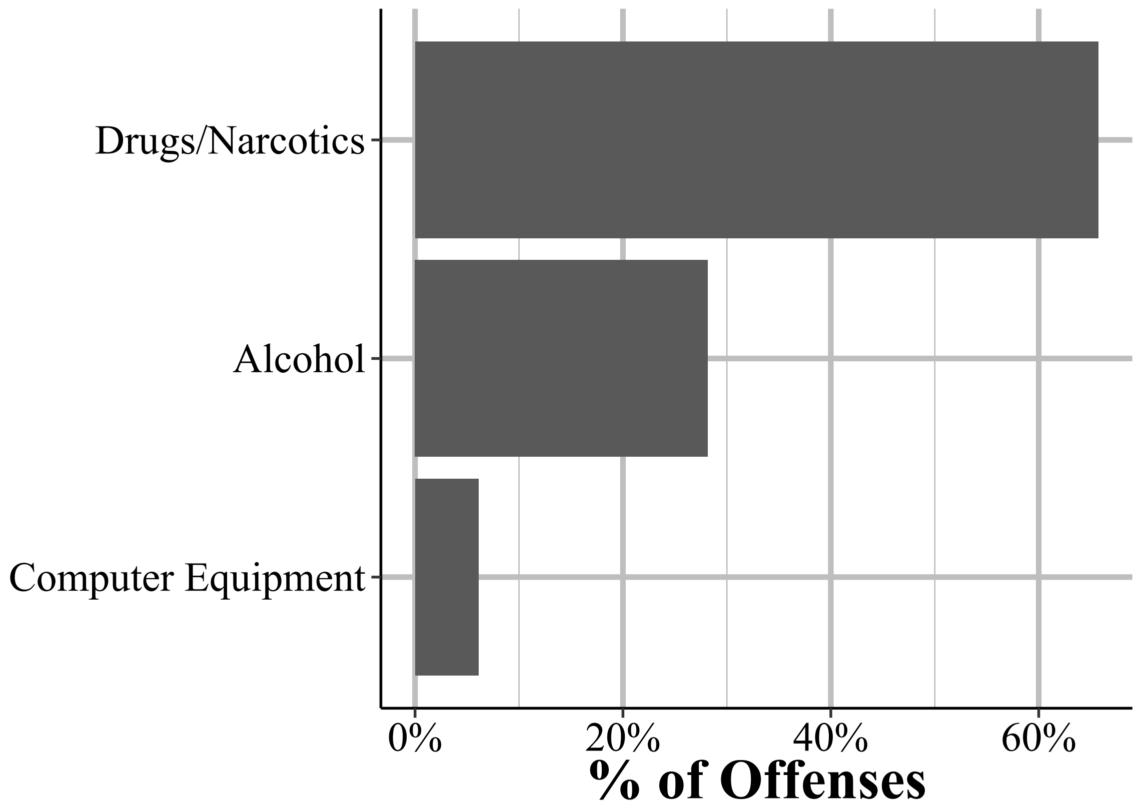


Figure 4.2: The distribution of drug, alcohol, or computer use for offenses where there was usage of one of these items. For easier viewing of how this variable is distributed, this figure excludes all offenses where there was no drug, alcohol, or computer use or the variable was NA.

4.1.4 Crime location

Table 4.4: 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%
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%

Crime Location	# of Offenses	% of Offenses
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%

4.1.5 Weapons

Table 4.5: 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%

Weapon Used	# of Offenses	% of Offenses
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%

- 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

4.1.6 Automatic weapons

This variable only tells you if the weapon is automatic

4.1.7 Hate crime indicator (bias motivation)

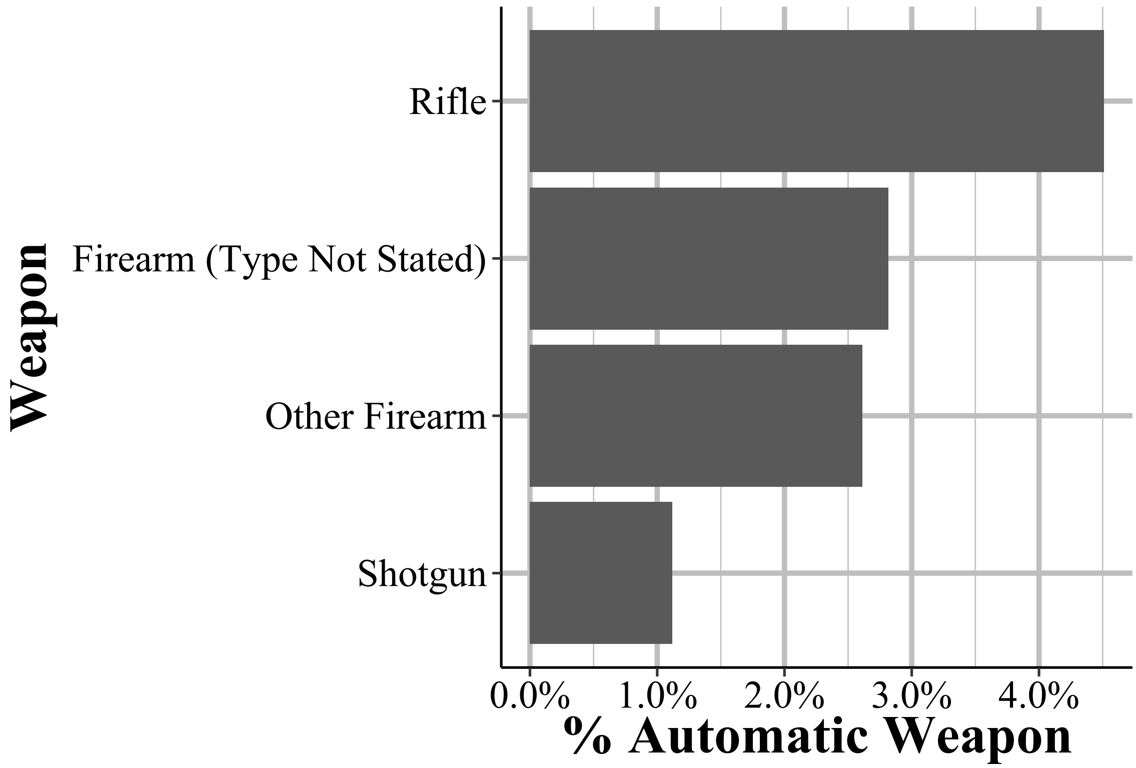


Figure 4.3: The percent of firearms used that were fully automatic, for all offenses in 2019.

Table 4.6: The bias motivation (i.e. if hate crime or not and what type of hate crime) for all offenses reported in 2019.

Bias Motivation	# of Offenses	% of Offenses
No Bias Motivation	7,372,214	99.14%
Unknown Bias Motivation	59,360	0.80%
Anti-Black	1,309	0.02%
Anti-White	573	0.01%
Anti-Male Homosexual (Gay)	357	0.00%
Anti-Hispanic	310	0.00%
Anti-Jewish	249	0.00%
Anti-Homosexual (Both Gay And Lesbian)	213	0.00%
Anti-Other Ethnicity/National Origin	210	0.00%
Anti-American Indian Or Alaskan Native	125	0.00%
Anti-Mental Disability	107	0.00%
Anti-Asian/Pacific Islander	101	0.00%
Anti-Multi-Racial Group	97	0.00%

Bias Motivation	# of Offenses	% of Offenses
Anti-Islamic (Muslim)	86	0.00%
Anti-Female Homosexual (Lesbian)	70	0.00%
Anti-Transgender	68	0.00%
Anti-Sikh	66	0.00%
Anti-Arab	64	0.00%
Anti-Other Religion	63	0.00%
Anti-Gender Non-Conforming	59	0.00%
Anti-Female	51	0.00%
Anti-Eastern Orthodox (Greek, Russian, Etc.)	50	0.00%
Anti-Physical Disability	44	0.00%
Anti-Catholic	42	0.00%
Anti-Other Christian	40	0.00%
Anti-Native Hawaiian Or Other Pacific Islander	30	0.00%
Anti-Protestant	26	0.00%
Anti-Multi-Religious Group	25	0.00%
Anti-Bisexual	22	0.00%
Anti-Male	17	0.00%
Anti-Heterosexual	12	0.00%
Anti-Mormon	7	0.00%
Anti-Hindu	7	0.00%
Anti-Buddhist	6	0.00%
Anti-Atheism/Agnosticism	5	0.00%
Anti-Jehovahs Witness	5	0.00%
Total	7,436,090	100%

Table 4.7: The bias motivation (i.e. if hate crime or not and what type of hate crime) for all offenses reported in 2019 that were classifed 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%

Bias Motivation	# of Offenses	% of Offenses
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%
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

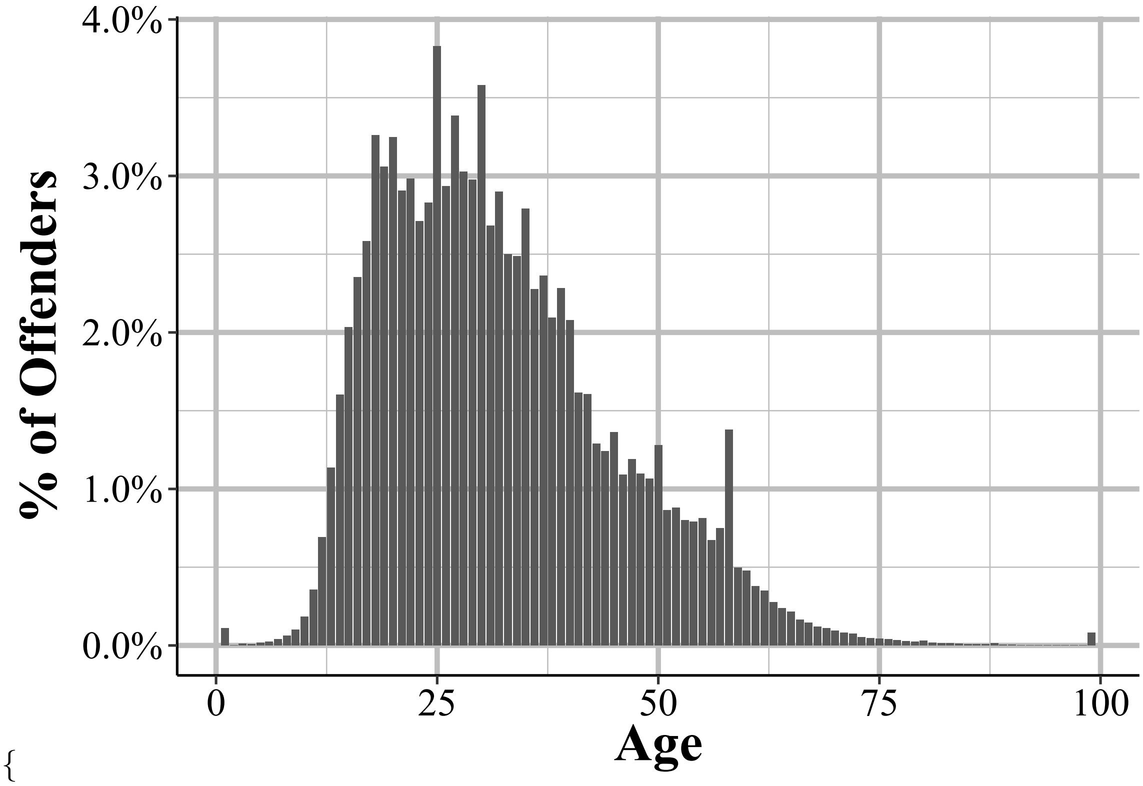
As might be expected, the Offender Segment provides information about who the offender is for each incident, though this is limited to only demographic variables. So we know the age, sex, and race of each offender but nothing else. This means that important variables such as criminal history, ethnicity, 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

5.1.1 Demographics

5.1.1.1 Age

\begin{figure}



{

}

\caption{The age of all offenders reported in the 2019 NIBRS data. Approximately 39% of offenders have an unknown age are an not shown in the figure.} \end{figure}

5.1.1.2 Sex

5.1.1.3 Race

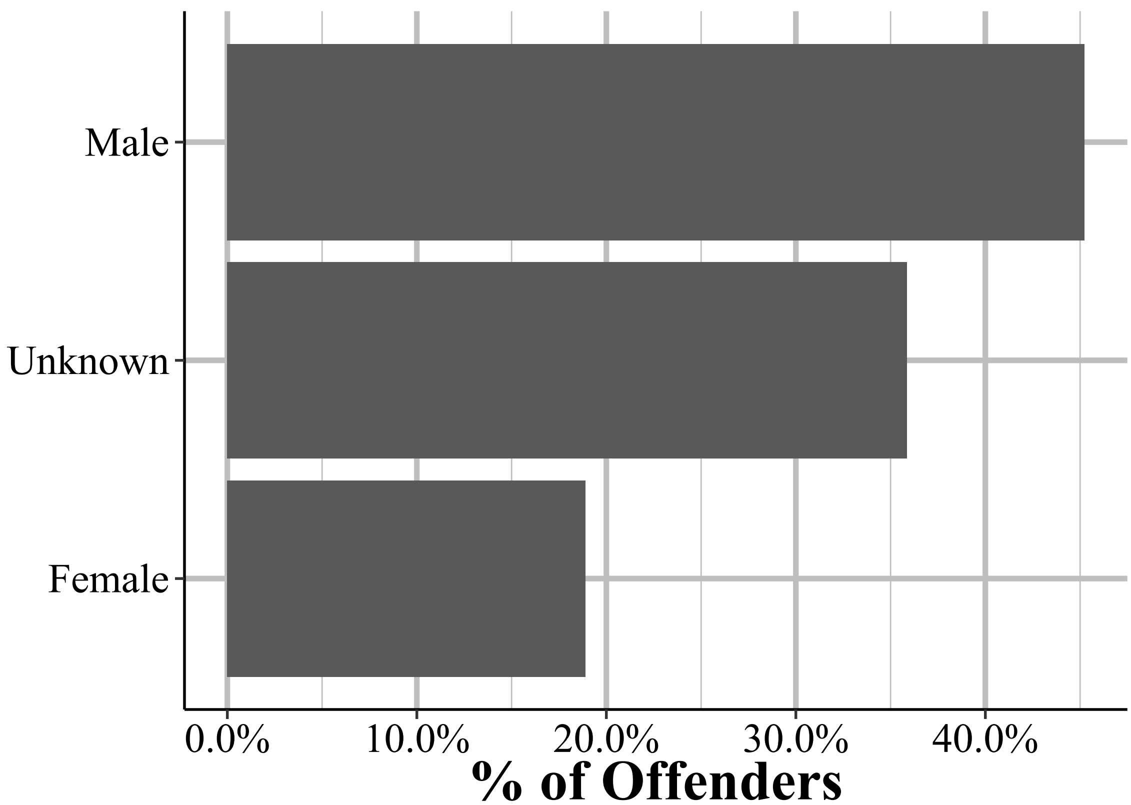


Figure 5.1: The sex of all offenders reported in the 2019 NIBRS data.

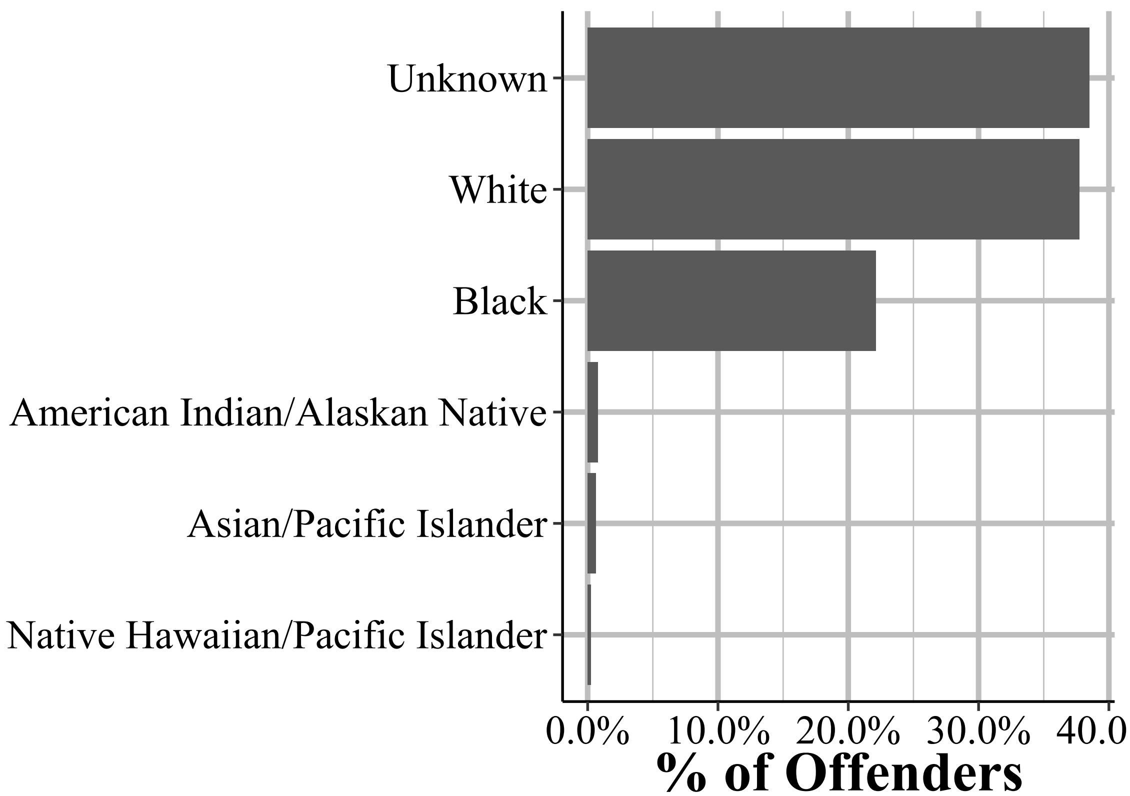


Figure 5.2: The race of all offenders reported in the 2019 NIBRS data.

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

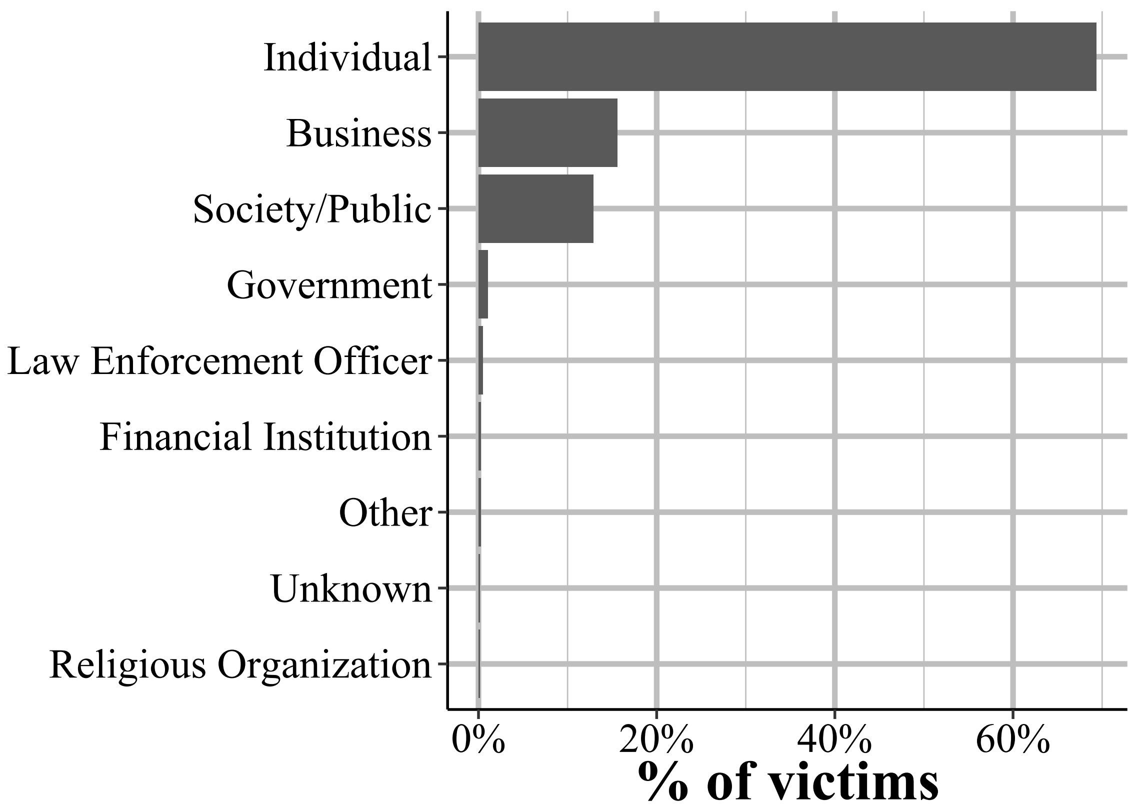


Figure 6.1: The distribution of the type of victim. Victim types are mutuaully exclusive.

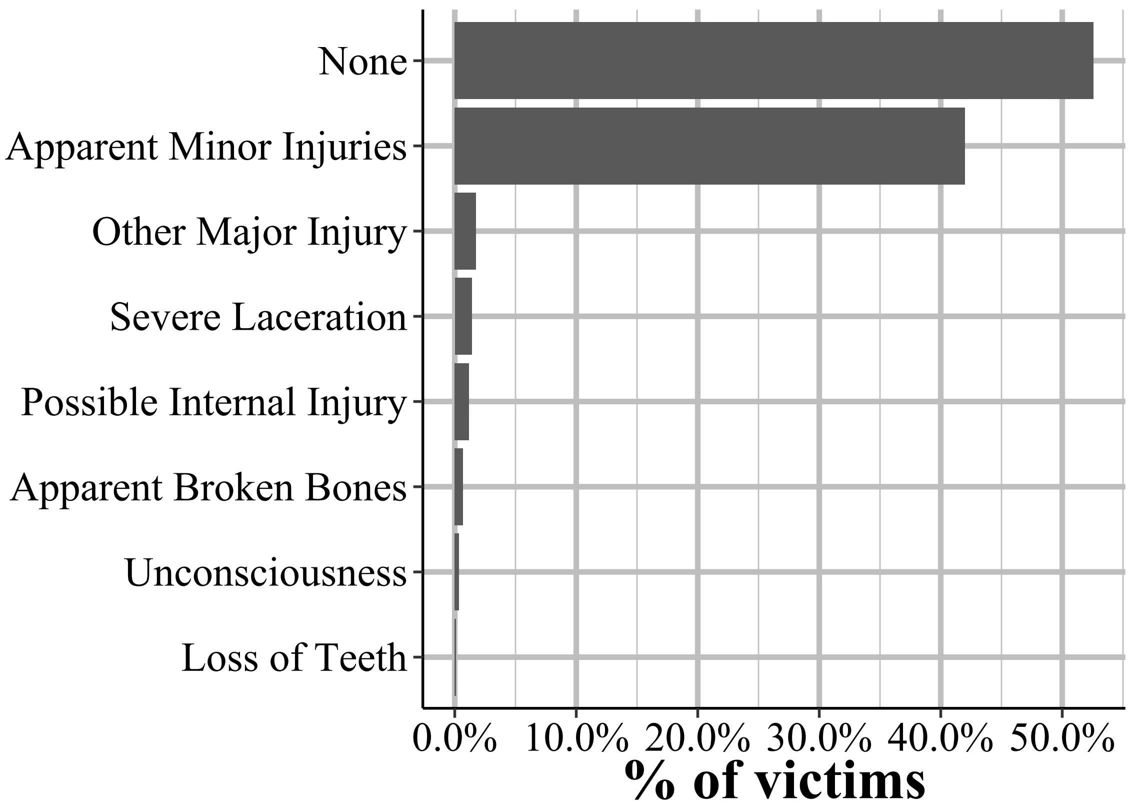


Figure 6.2: The distribution of the injury sustained by the victim. Only individual and law enforcement officer victims have this variable available.

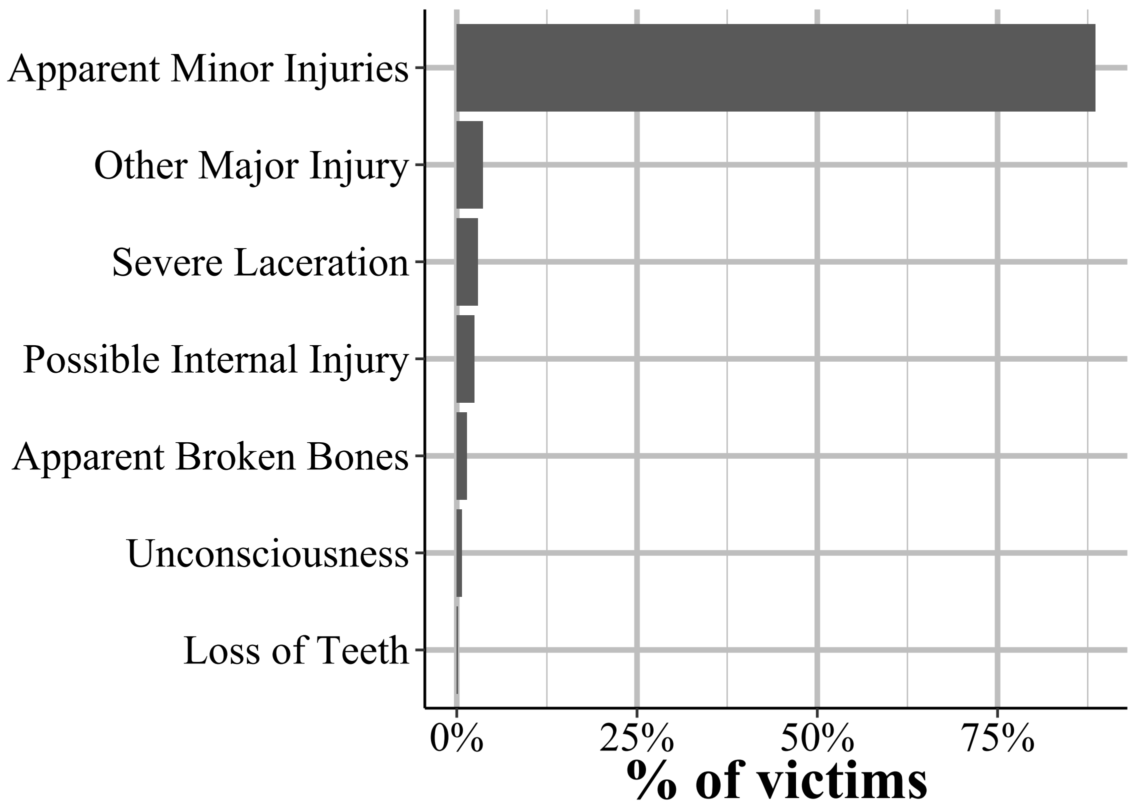


Figure 6.3: The distribution of the injury sustained by the victim for those who had an injury other than 'none'.

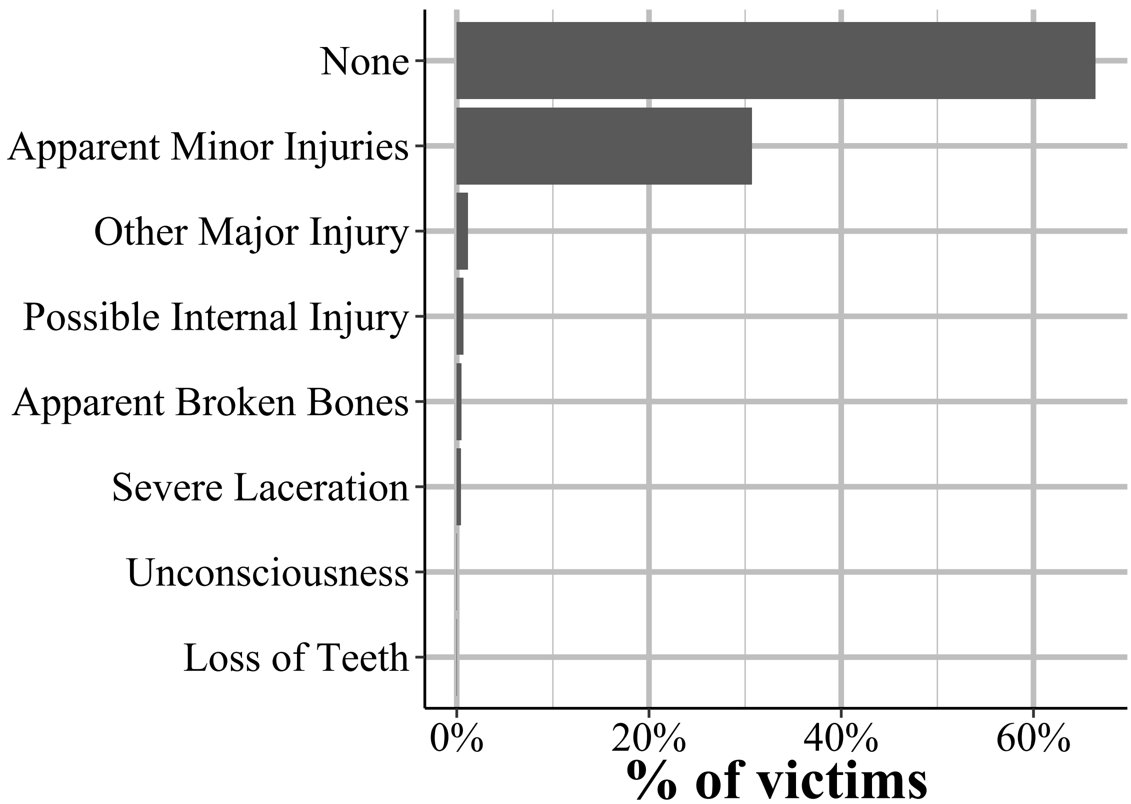


Figure 6.4: The distribution of the injury sustained by the victim for law enforcement officer victims

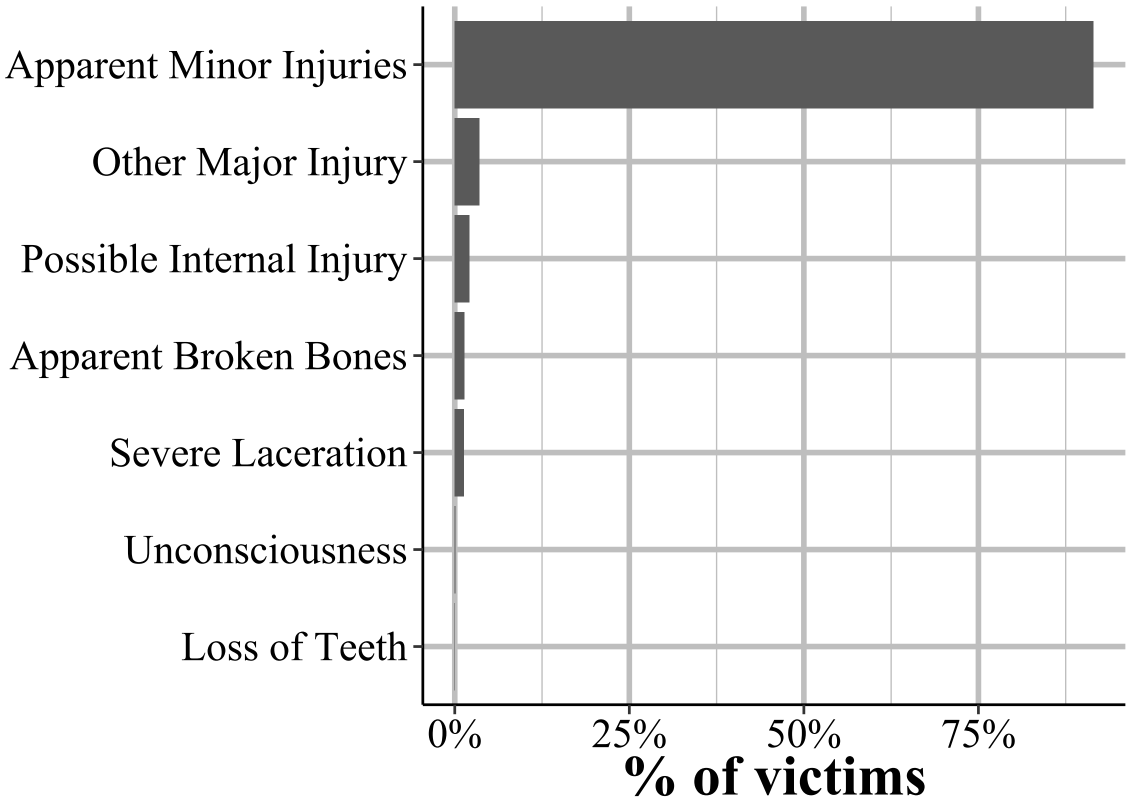


Figure 6.5: The distribution of the injury sustained by the victim for law enforcement officer victims excluding those who had no injury at all.

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 Residence status

Only for when victim is individual or law enforcement officer

victim$resident_status_of_victim

6.1.6 Aggravated assault and homicide circumstances

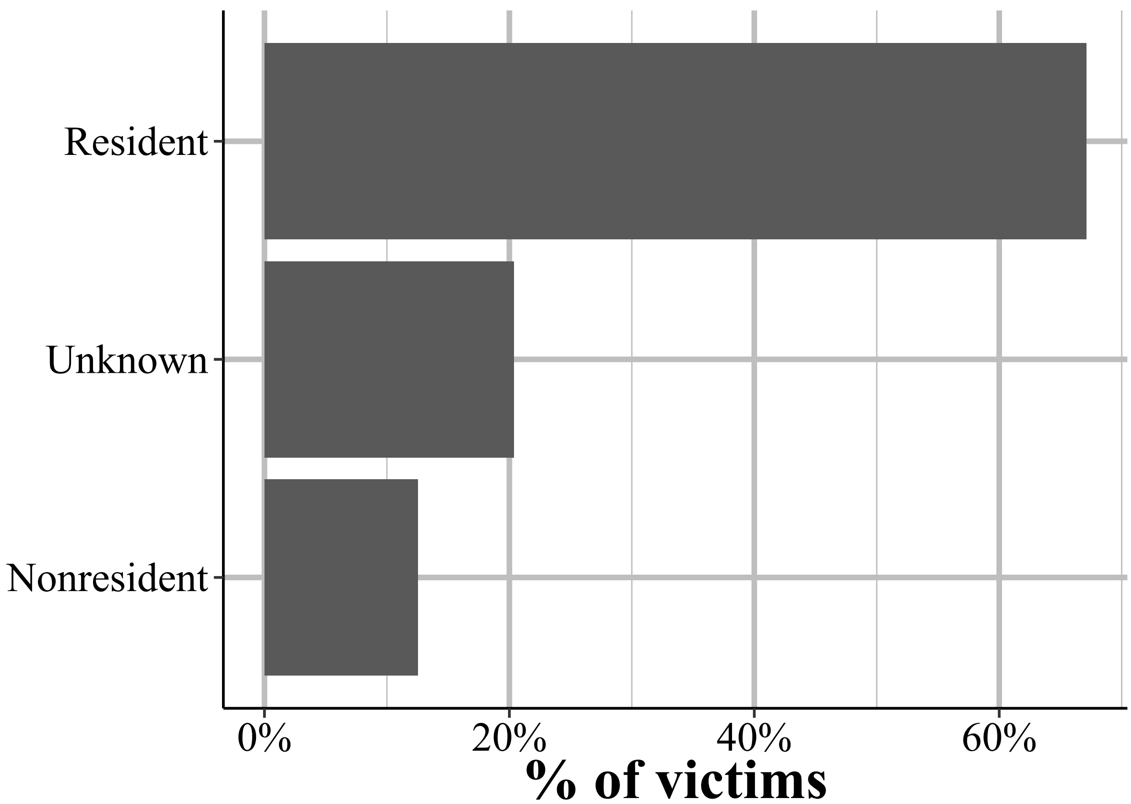


Figure 6.6: The distrubution of residence status for all victims reported to NIBRS in 2019. Residence status is residence in the police agency's jurisdiction (e.g. do you live in the city you were victimized in?). It is unrelated to citizenship or immigration status.

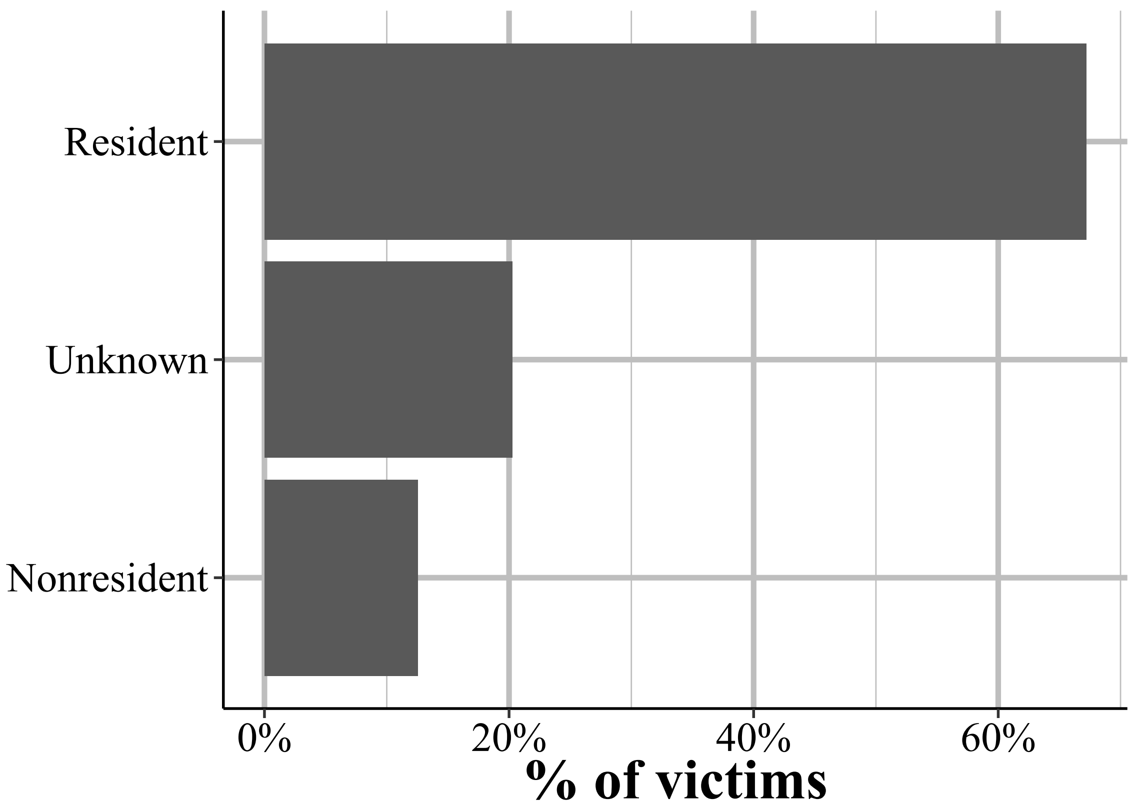


Figure 6.7: The distrubution of residence status for all Individual (i.e. person who is not law enforcement officer) victims.

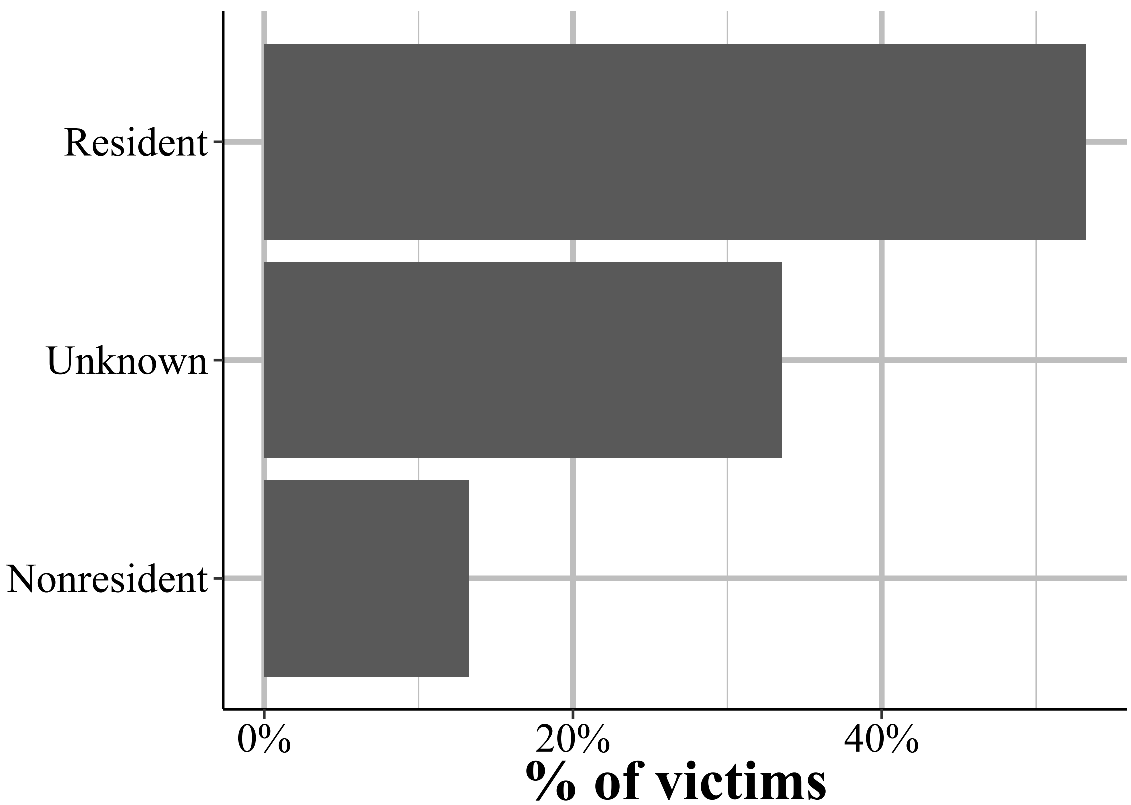


Figure 6.8: The distrubution of residence status for all Law Enforcement Officer victims.

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%
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%

6.1.7 Justifiable homicide circumstance

6.1.8 Age

6.1.9 Sex

6.1.10 Race

6.1.11 Ethnicity

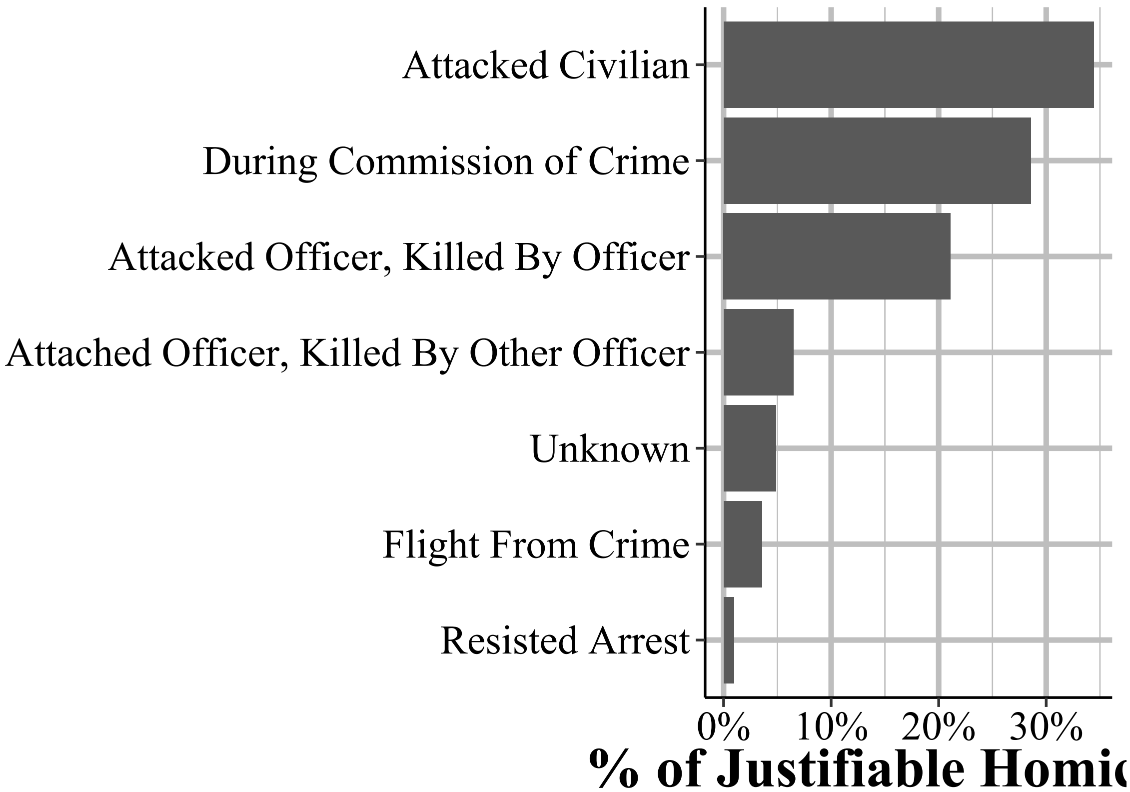


Figure 6.9: The distribution of circumstances for justifiable homicides (N = 308 in 2019 for all agencies reporting).

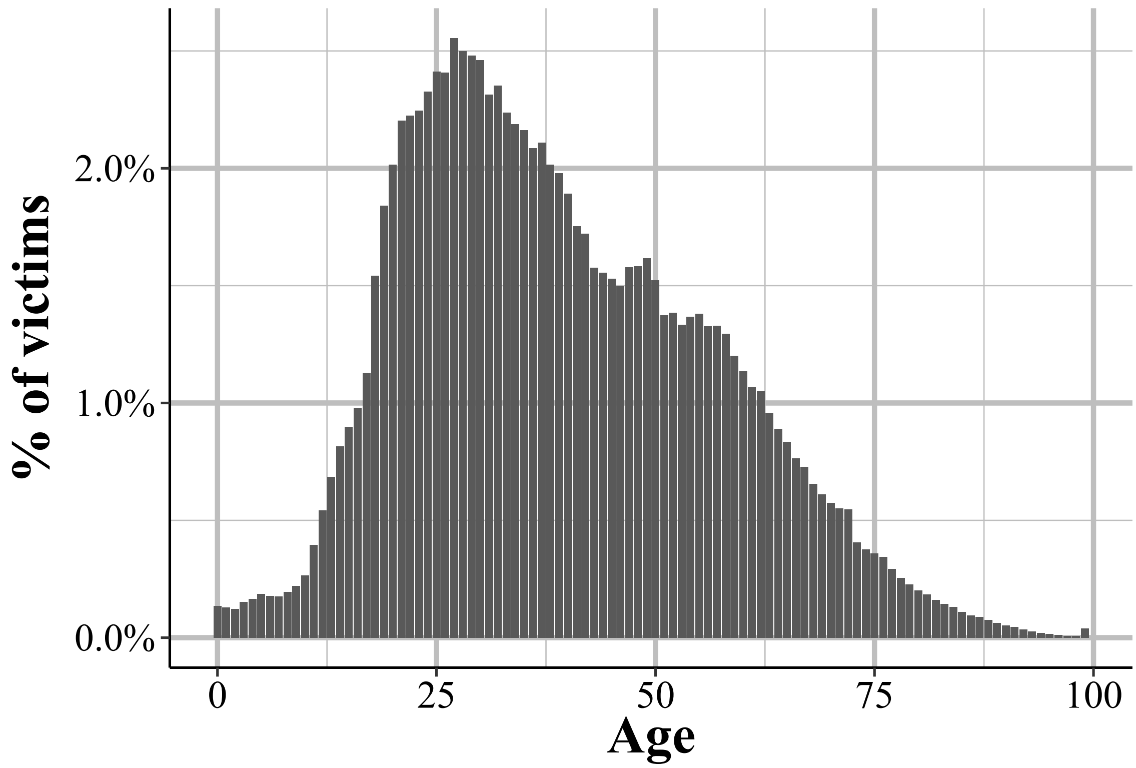


Figure 6.10: The age of all victims reported in the 2019 NIBRS data.

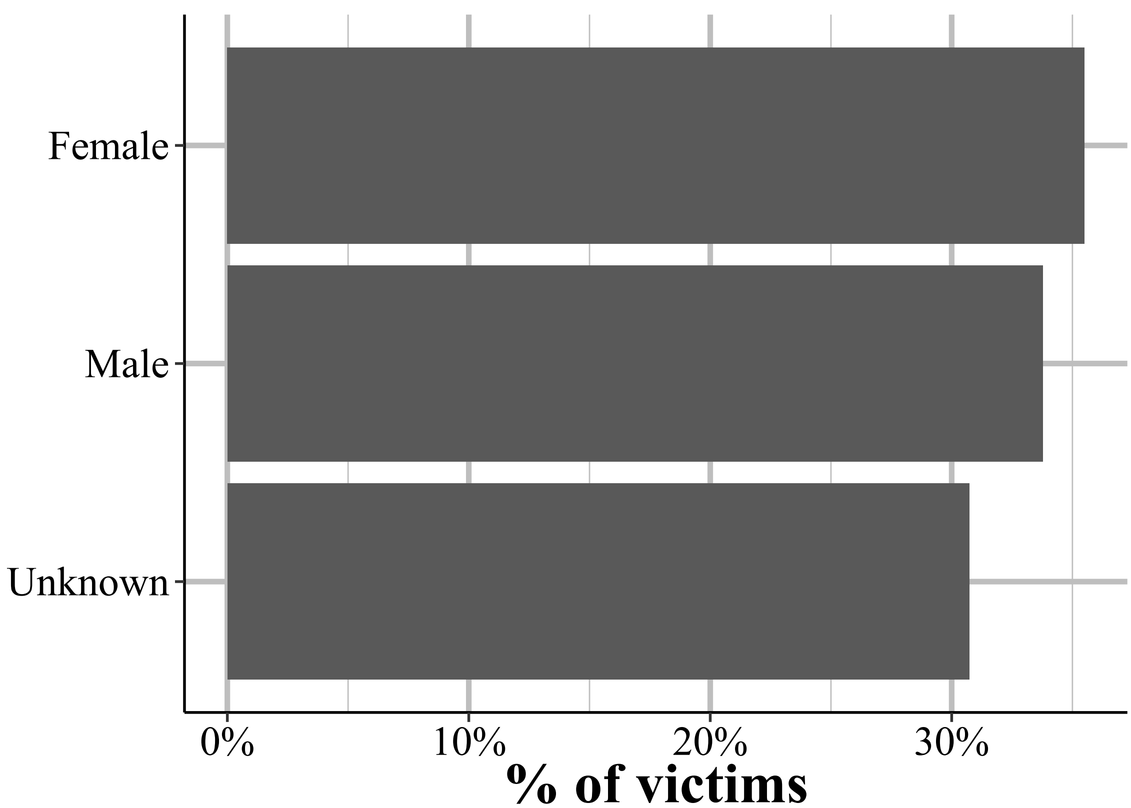


Figure 6.11: The sex of all victims reported in the 2019 NIBRS data.

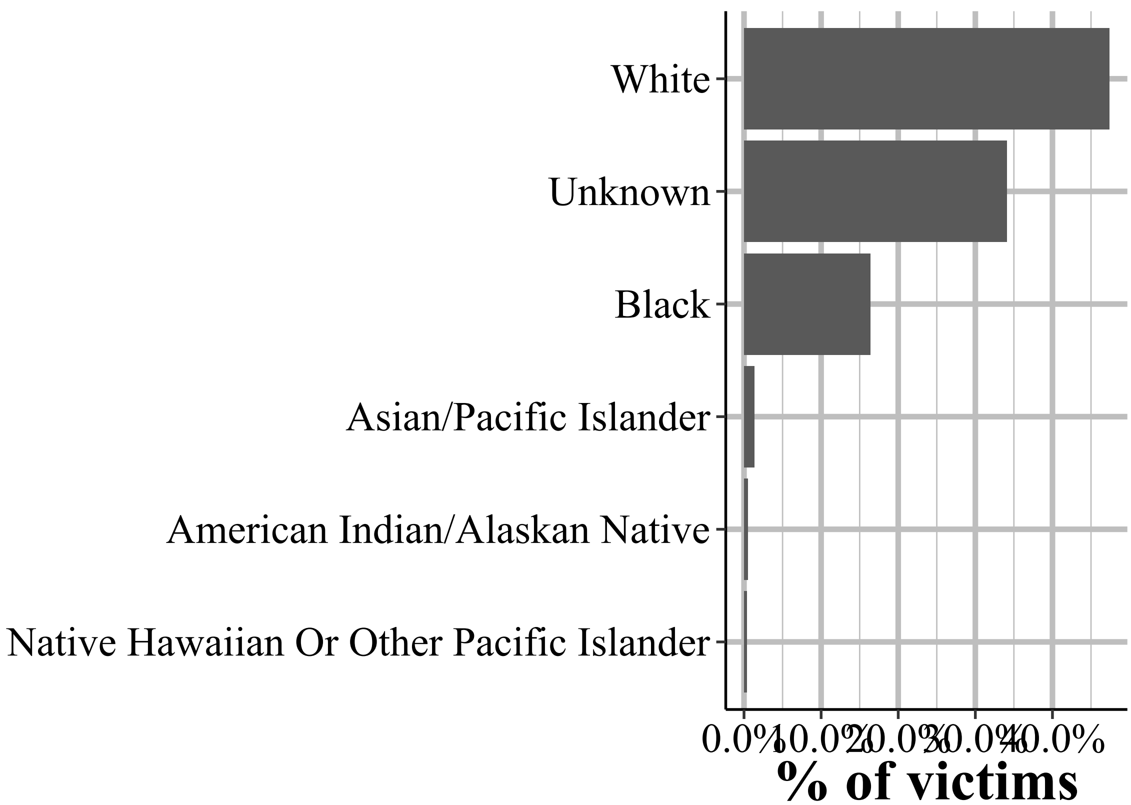


Figure 6.12: The race of all victims reported in the 2019 NIBRS data.

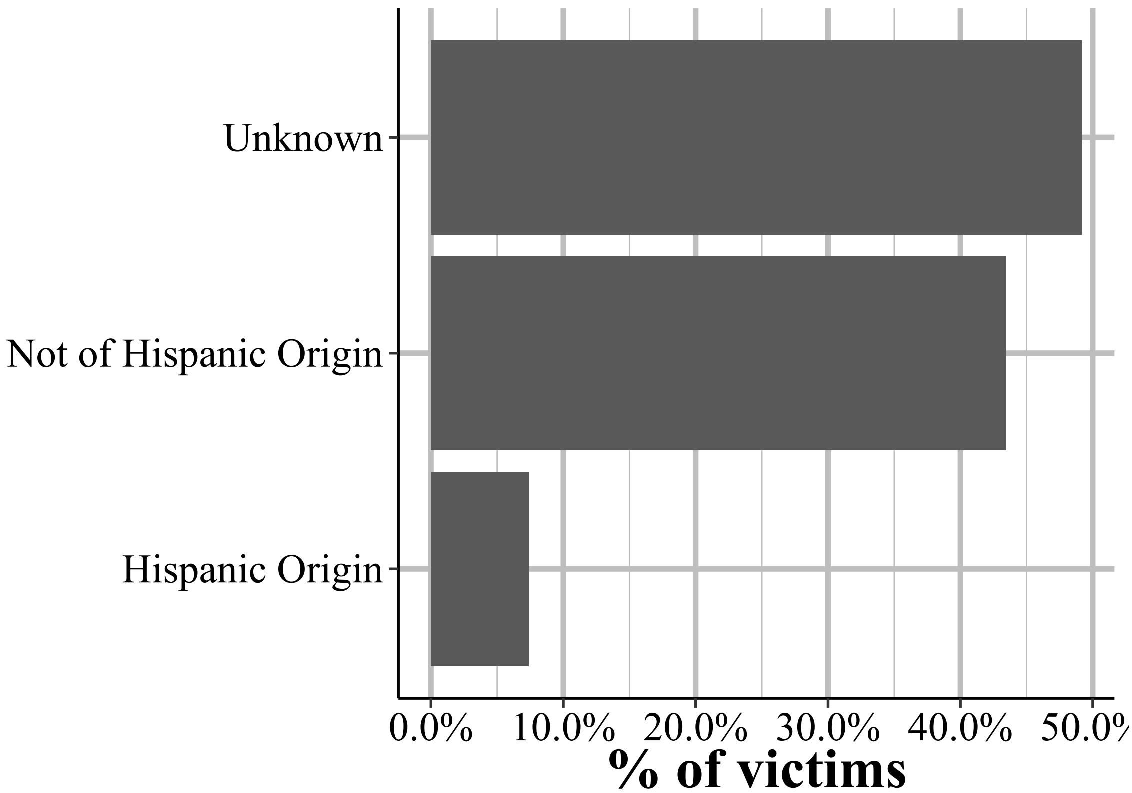


Figure 6.13: The ethnicity of all victims reported in the 2019 NIBRS data.

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).

7.1 Important variables

7.1.1 Crime arrested for (excluding Group B arrests)

As Table 7.1 shows...

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%
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%

Crime Category	# of Offenses	% of Offenses
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%
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%

7.1.2 Group B crimes arrested for

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%

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).

7.1.3 Arrest date

7.1.4 Weapons

All crimes can potentially have a weapon

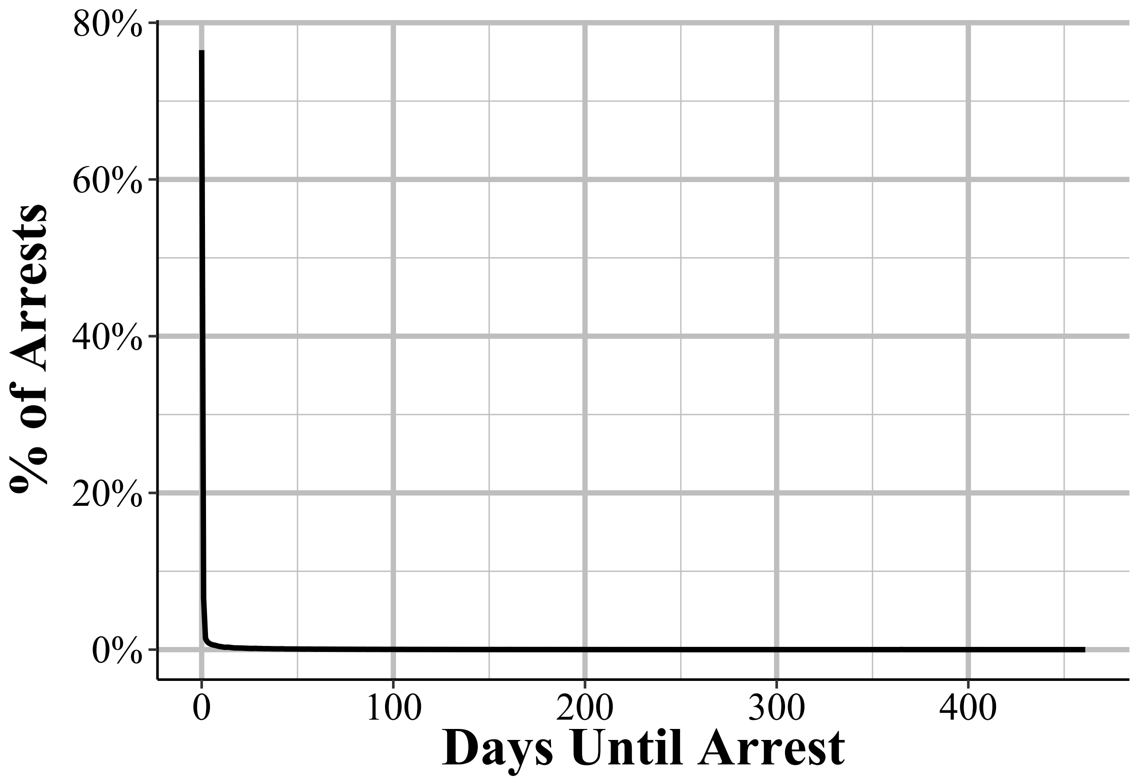


Figure 7.1: The distribution of the number of days from the incident to the arrest date. In 2019 the maxiumum days from incident to arrest was 461 days. Zero days means that the arrest occurred on the same day as the incident.

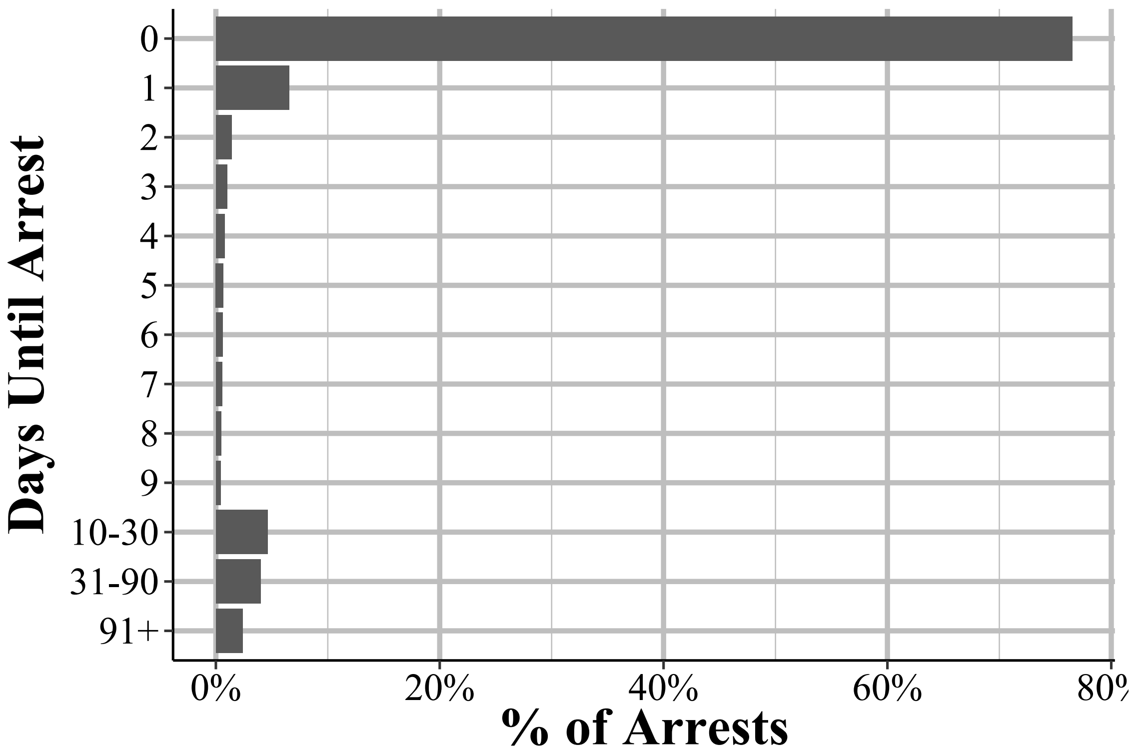


Figure 7.2: The number of days from the incident to the arrest date. Values over 10 days are grouped to better see the distribution for arrests that took fewer than 10 days. Zero days means that the arrest occurred on the same day as the incident.

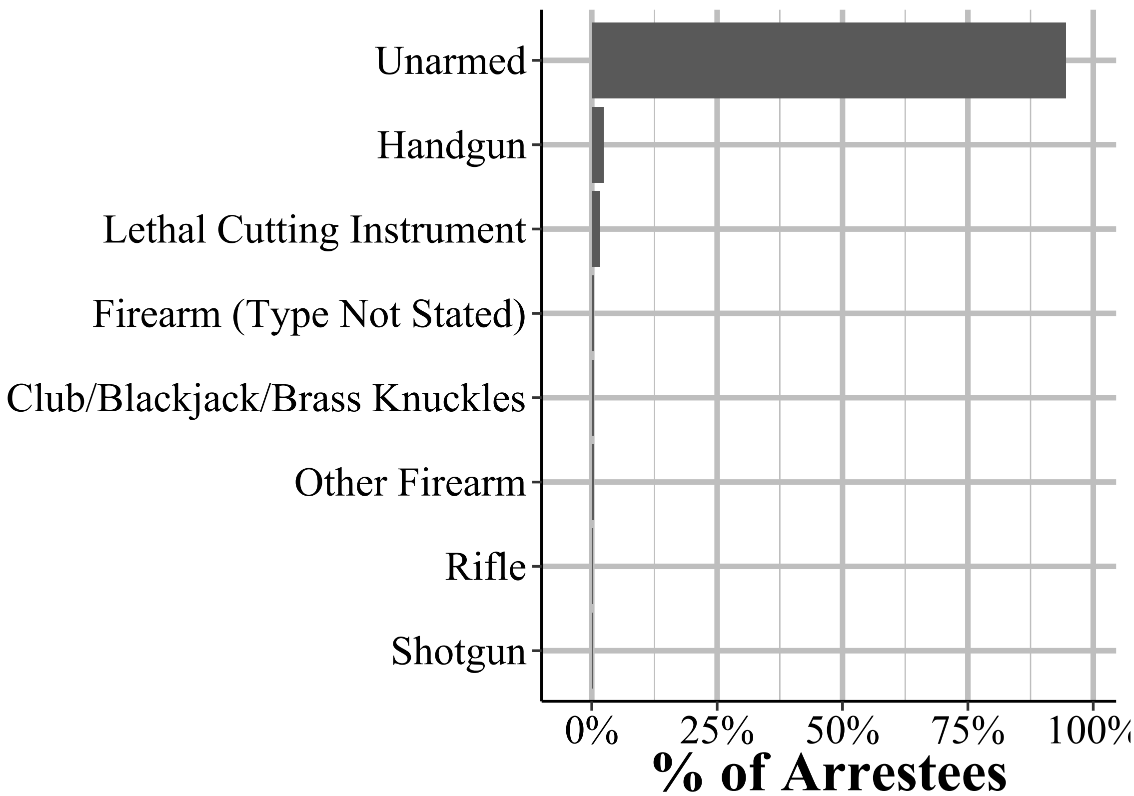


Figure 7.3: The weapon used in the crime for all arrestees reported in 2019. Unarmed can still mean that the arrestee used force, such as by punching the victim, but that they did not use any weapon.

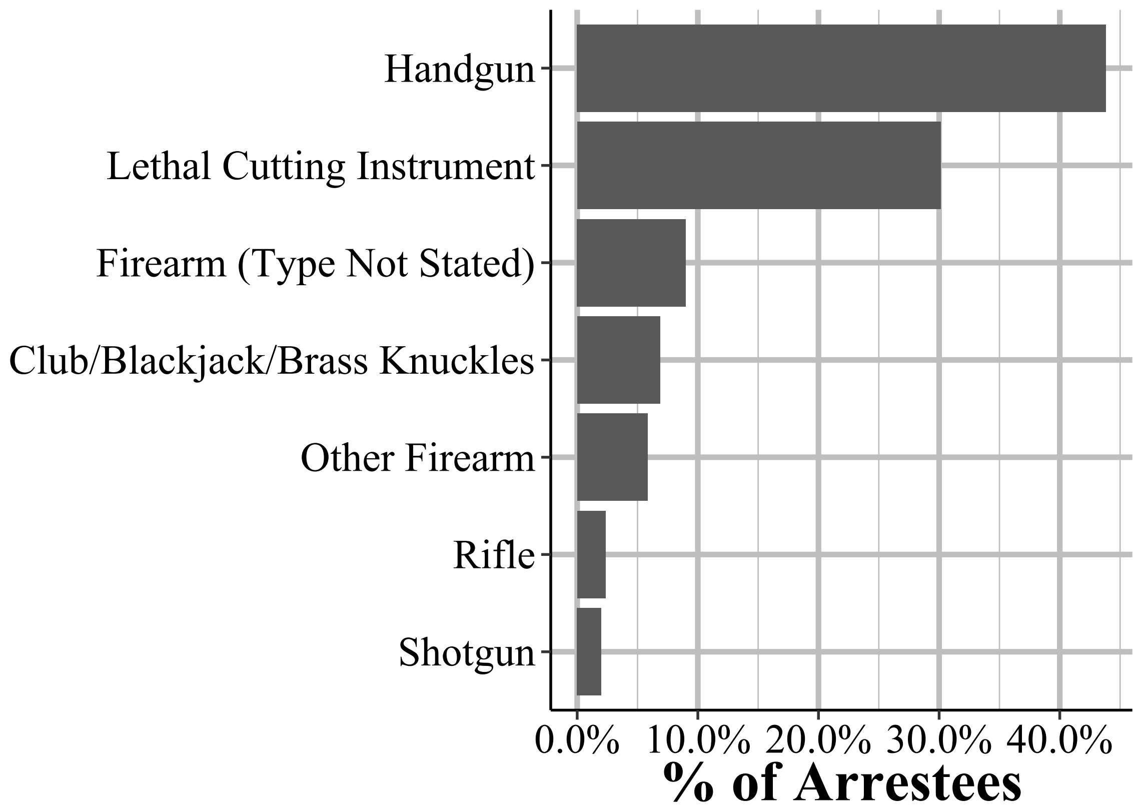


Figure 7.4: The distribution of weapon usage for all arrestees in 2019 who used a weapon (i.e. excludes unarmed offenders).

7.1.5 Automatic weapons

This variable only tells you if the weapon is automatic

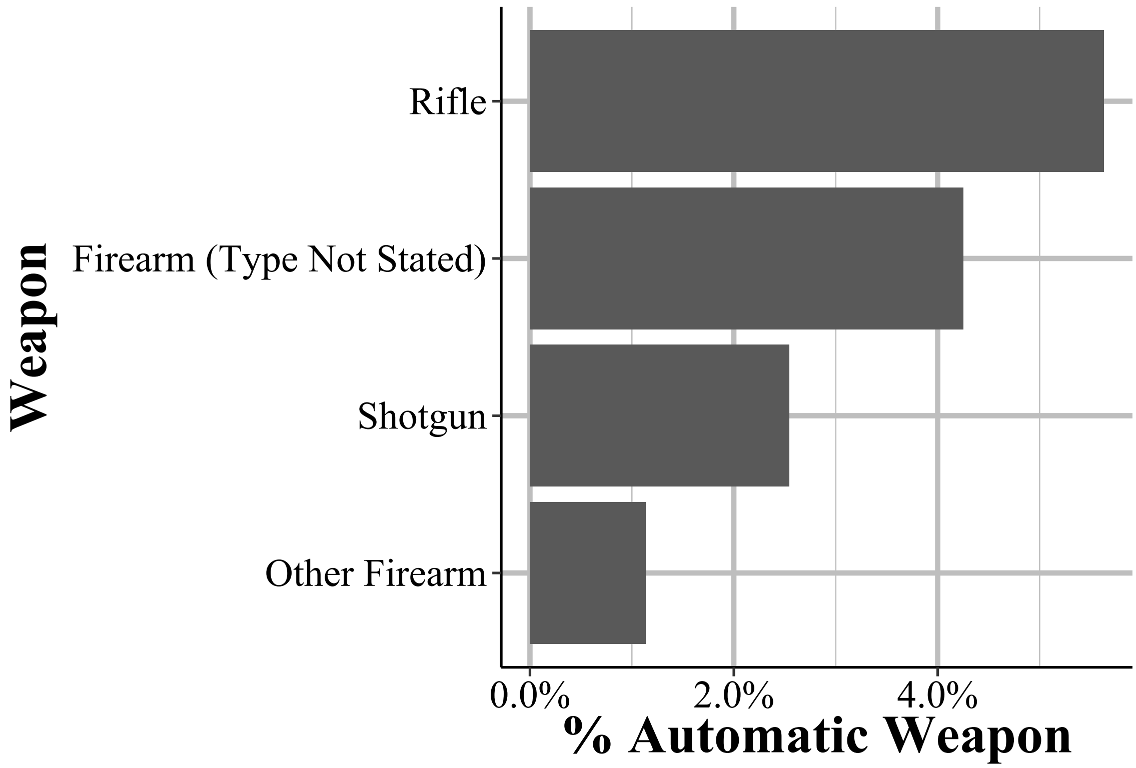


Figure 7.5: The percent of firearms used that were fully automatic, for arrestees in 2019.

7.1.6 Type of arrest

7.1.7 Disposition for juvenile arrestees

7.1.8 Residence status

7.1.9 Demographics

7.1.9.1 Age

7.1.9.2 Sex

7.1.9.3 Race

7.1.9.4 Ethnicity

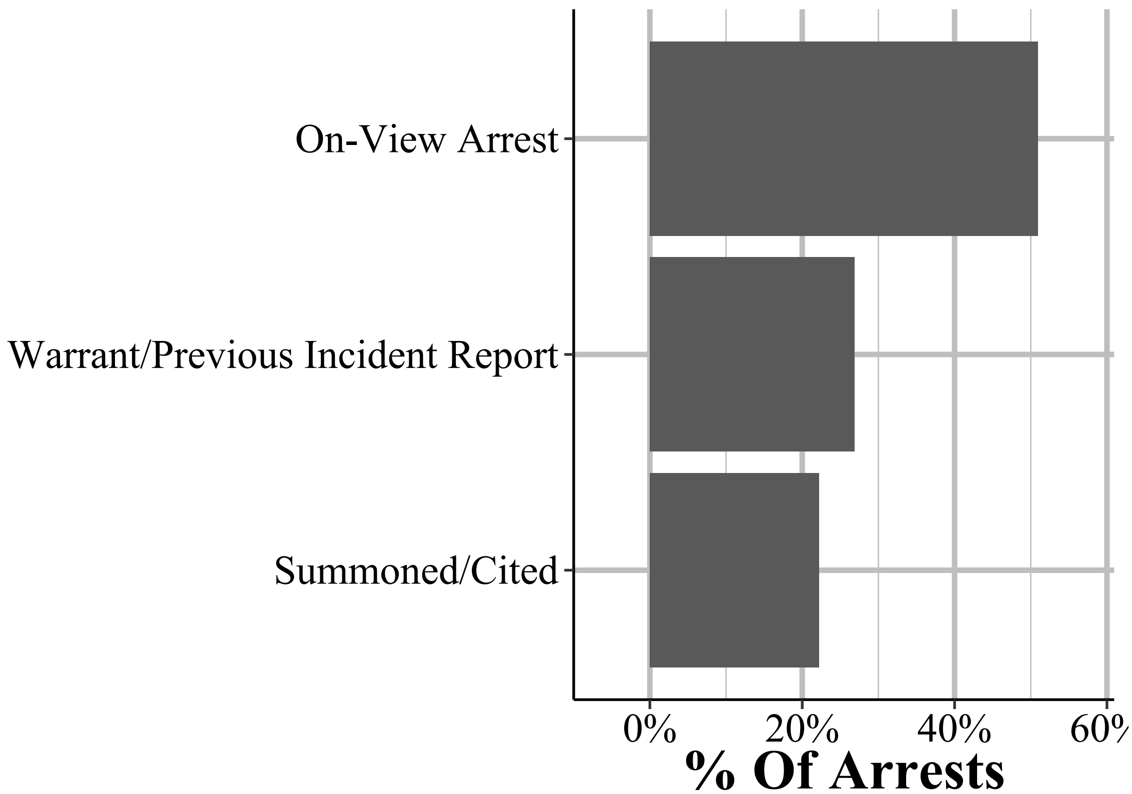


Figure 7.6: The distrubution of arrests by type of arrest. Previous Incident Report includes cases where an individual was arrested for a separate crime and are then reported as also arrested for this incident.

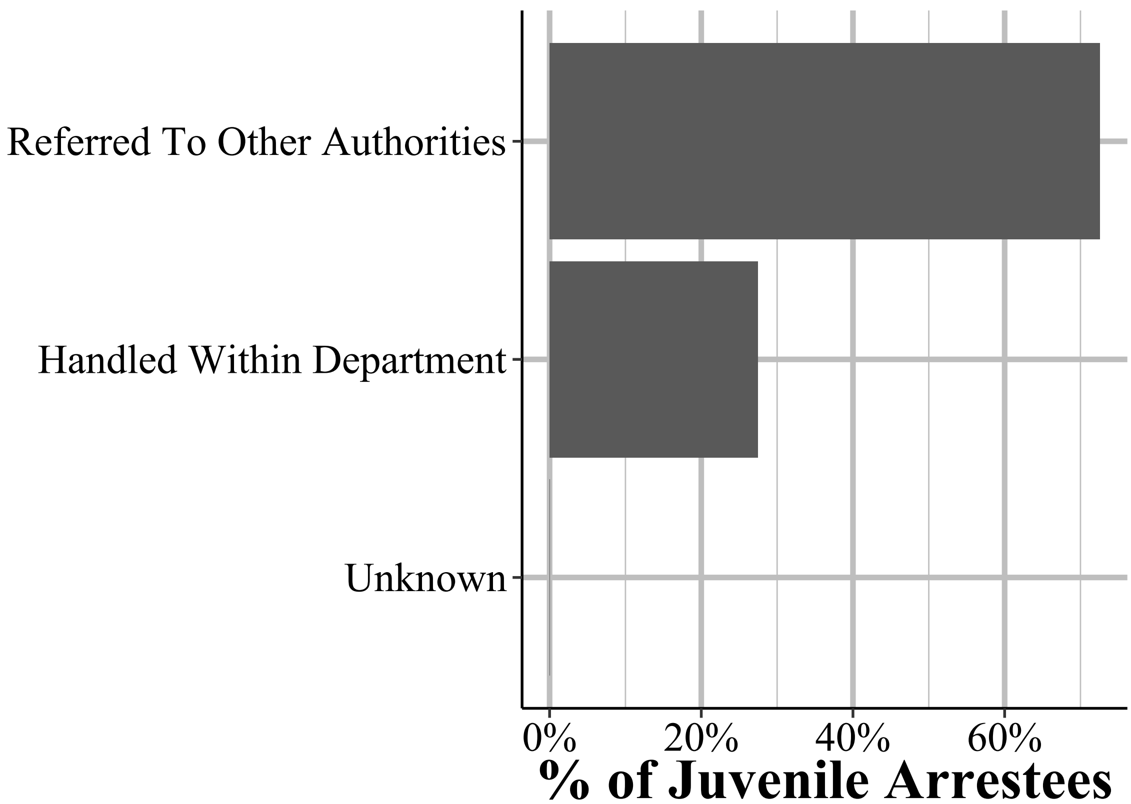


Figure 7.7: For juvenile arrestees (under age 18), the distribution of case outcomes.

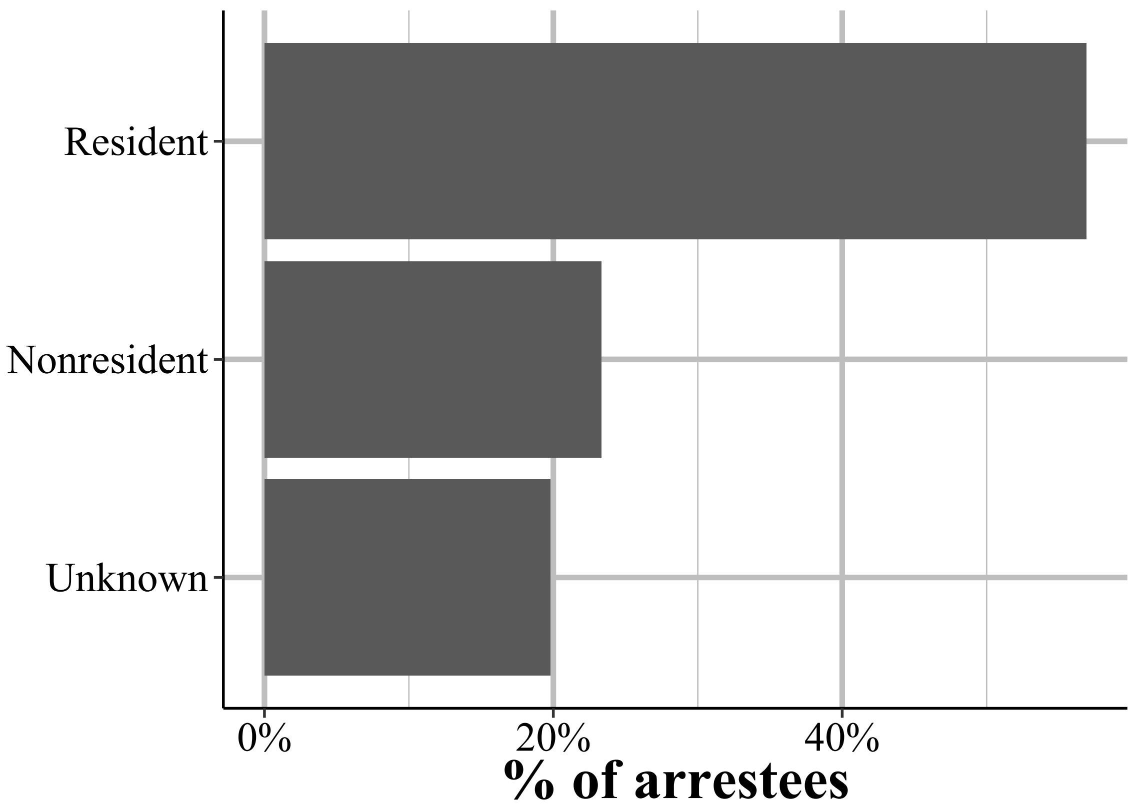


Figure 7.8: The distrubution of residence status for all arrestees reported to NIBRS in 2019. Residence status is residence in the arresting agency's jurisdiction (e.g. do you live in the city you were arrested in?). It is unrelated to citizenship or immigration status.

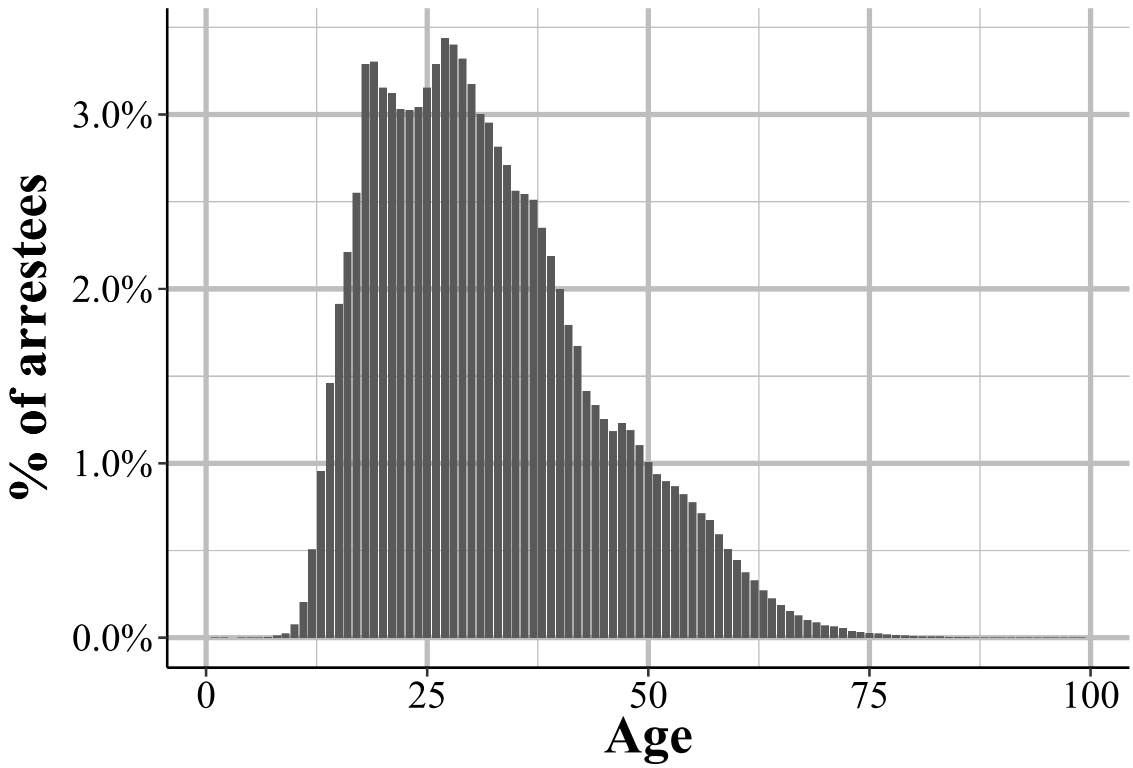


Figure 7.9: The age of all arrestees reported in the 2019 NIBRS data.

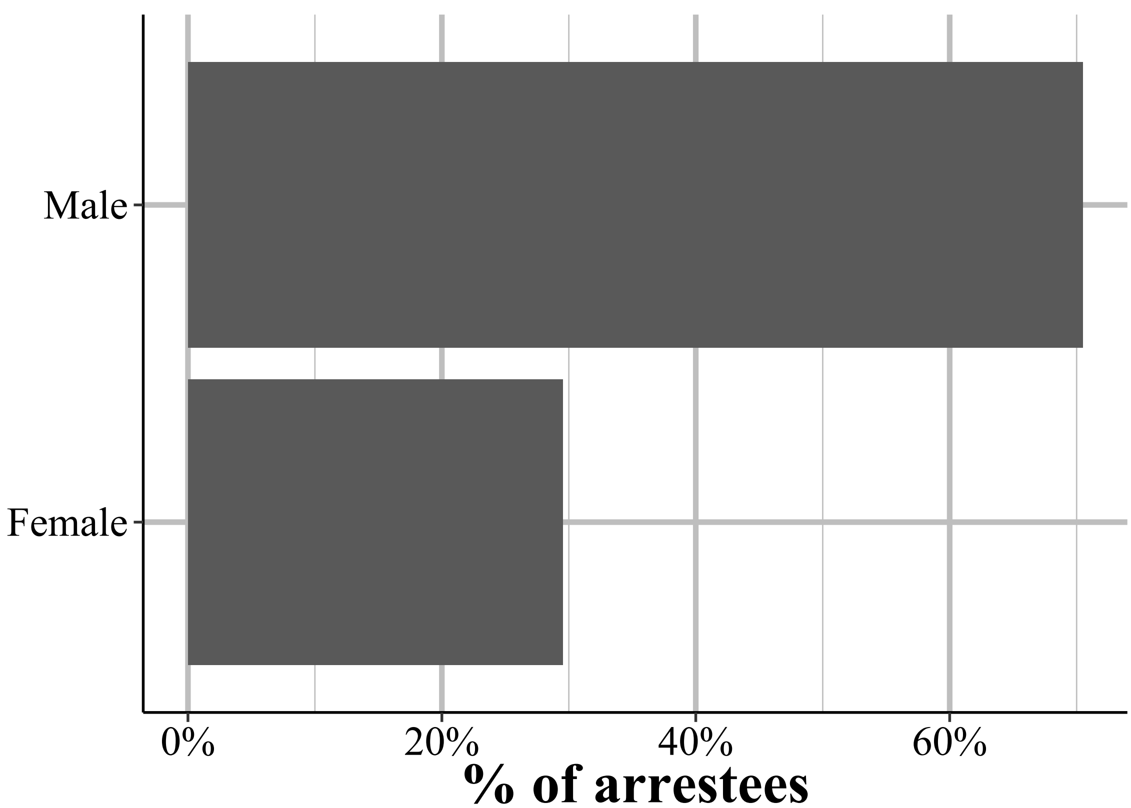


Figure 7.10: The sex of all arrestees reported in the 2019 NIBRS data.

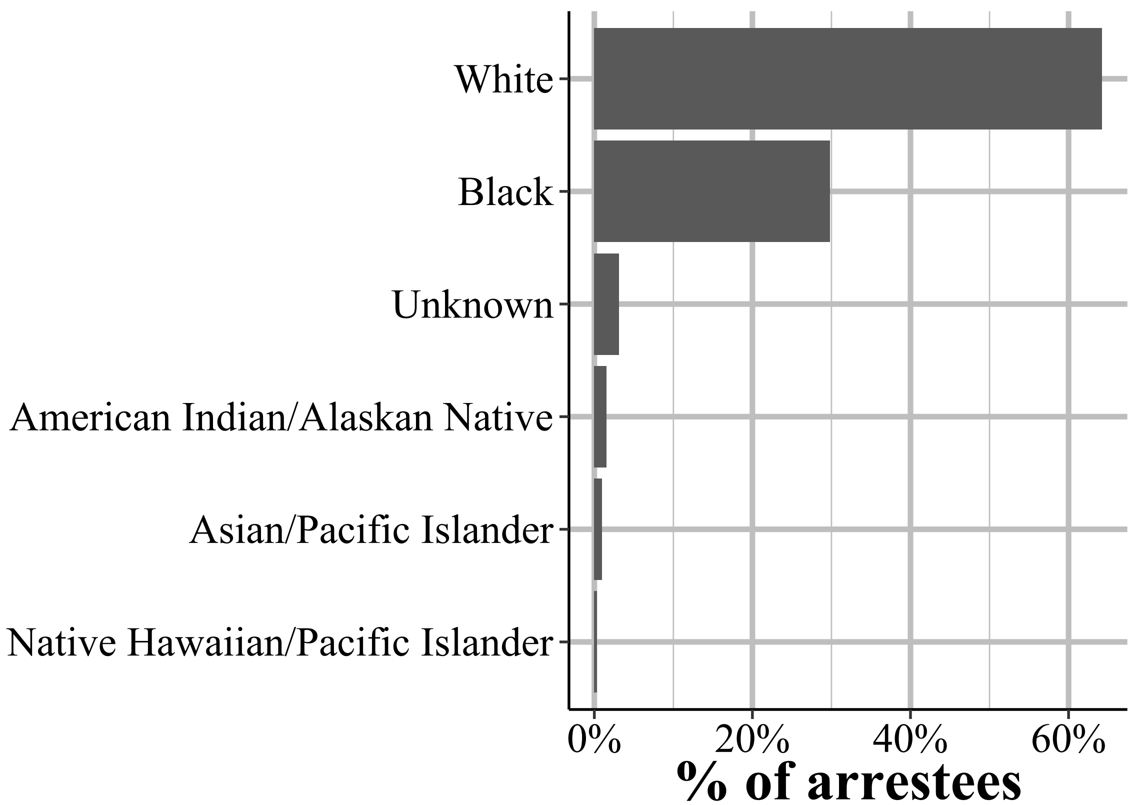


Figure 7.11: The race of all arrestees reported in the 2019 NIBRS data.

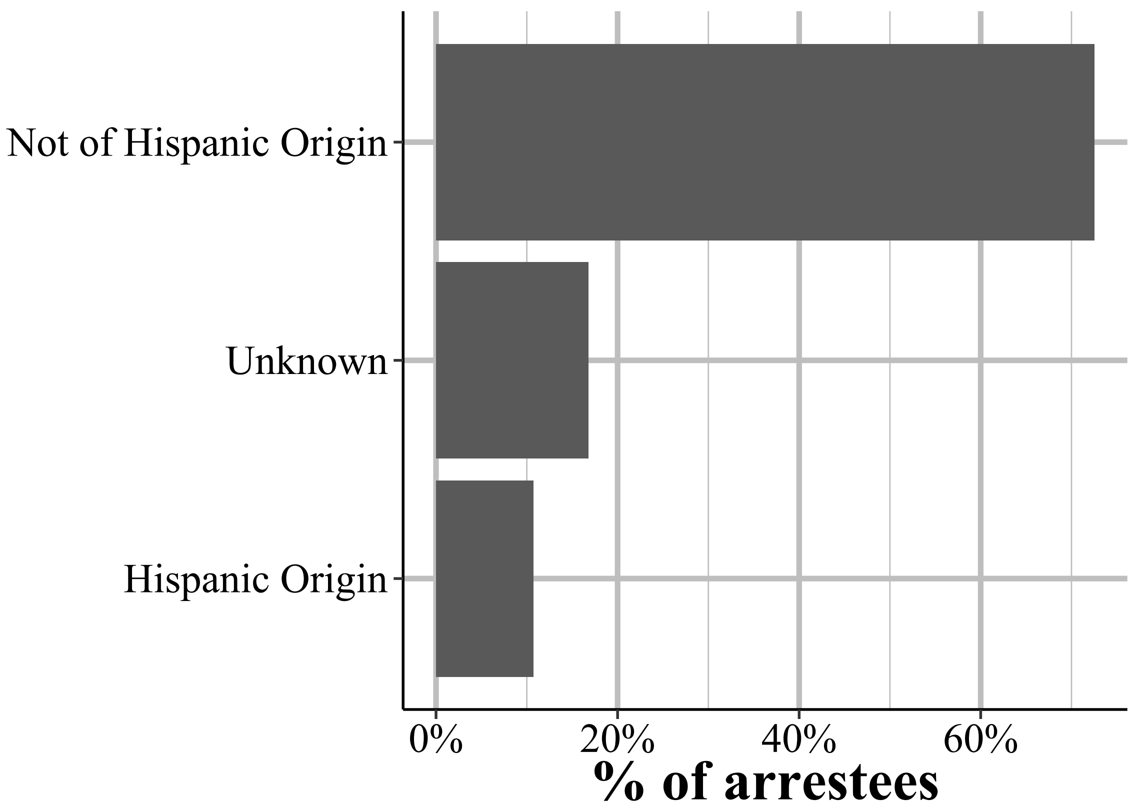


Figure 7.12: The ethnicity of all arrestees reported in the 2019 NIBRS data.

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 Description of property

Table 8.1: The number and percent of property stolen (including forcibly taken such as during a robbery) in a crime or seized by police (generally in drug arrests) for all offenses in 2019. Each offense can have multiple items stolen/seized.

Property	# of Property Stolen	% of Property Stolen
Other	1,005,616	13.29%
Drugs/Narcotics	714,797	9.45%
Automobiles	645,115	8.52%
Money	637,046	8.42%
Vehicle Parts/Accessories	328,970	4.35%
Clothes/Furs	312,506	4.13%
Drug/Narcotic Equipment	292,097	3.86%
Purses/Handbags/Wallets	291,694	3.85%
Merchandise	266,945	3.53%
Credit/Debit Cards	242,981	3.21%
Tools - Power/Hand	240,156	3.17%
Consumable Goods	236,104	3.12%
Portible Electronic Communications	201,787	2.67%
Computer Hardware/Software	192,970	2.55%
Household Goods	192,923	2.55%
Identity Documents	185,098	2.45%
Radios/Tvs/Vcrs	169,518	2.24%
Jewelry/Precious Metals	123,742	1.64%
Firearms	117,039	1.55%
Bicycles	89,936	1.19%
Trucks	83,796	1.11%
Negotiable Instruments	73,327	0.97%
Identity - Intangible	71,331	0.94%
Structures - Single Occupancy Dwellings	64,507	0.85%
Alcohol	64,339	0.85%
Documents - Personal Or Business	62,883	0.83%
Building Materials	54,091	0.71%
Office-Type Equipment	53,707	0.71%
Lawn/Yard/Garden Equipment	48,398	0.64%
Other Motor Vehicles	46,222	0.61%
Nonnegotiable Instruments	41,596	0.55%

Property	# of Property Stolen	% of Property Stolen
Structures - Other	33,265	0.44%
Trailers	30,124	0.40%
Recreational/Sports Equipment	29,946	0.40%
Structures - Commercial/Business	27,401	0.36%
Structures - Other Dwellings	23,936	0.32%
Photographic/Optical Equipment	23,579	0.31%
Recordings - Audio/Visual	23,358	0.31%
Camping/Hunting/Fishing Equipment/Supplies	22,040	0.29%
Heavy Construction/Industrial Equipment	17,546	0.23%
Fuel	17,013	0.22%
Weapons - Other	16,690	0.22%
Pending Inventory (Of Property)	14,262	0.19%
Firearm Accessories	14,053	0.19%
Collections/Collectibles	12,763	0.17%
Structures - Public/Community	11,351	0.15%
Musical Instruments	10,047	0.13%
Recreational Vehicles	10,020	0.13%
Metals, Non-Precious	9,951	0.13%
Medical/Medical Lab Equipment	9,212	0.12%
Farm Equipment	8,831	0.12%
Pets	8,071	0.11%
Structures - Storage	7,019	0.09%
Artistic Supplies/Accessories	4,705	0.06%
Watercraft	3,775	0.05%
Explosives	3,671	0.05%
Chemicals	3,549	0.05%
Gambling Equipment	3,226	0.04%
Watercraft Equipment/Parts/Accessories	2,628	0.03%
Law Enforcement Equipment	2,610	0.03%
Livestock	2,114	0.03%
Crops	2,094	0.03%
Buses	1,962	0.03%
Logging Equipment	1,437	0.02%
Structures - Industrial Manufacturing	1,398	0.02%
Special Category	1,277	0.02%
Aircraft Parts/Accessories	761	0.01%
Aircraft	455	0.01%

Property	# of Property Stolen	% of Property Stolen
Total	7,567,377	100%

8.1.2 Type of property loss

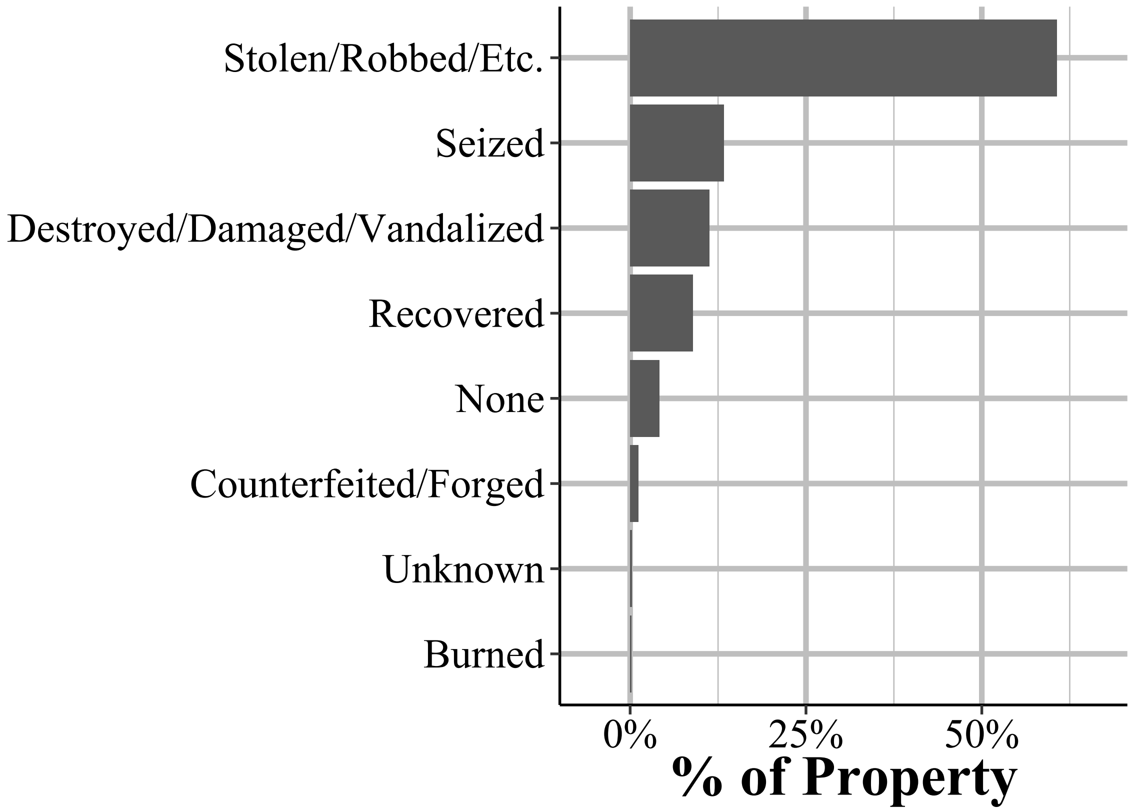


Figure 8.1: The type of loss or if the item is recovered.

8.1.3 Value of stolen property

8.1.4 Date property was recovered

8.1.5 Drugs

8.1.5.1 Suspected drug type

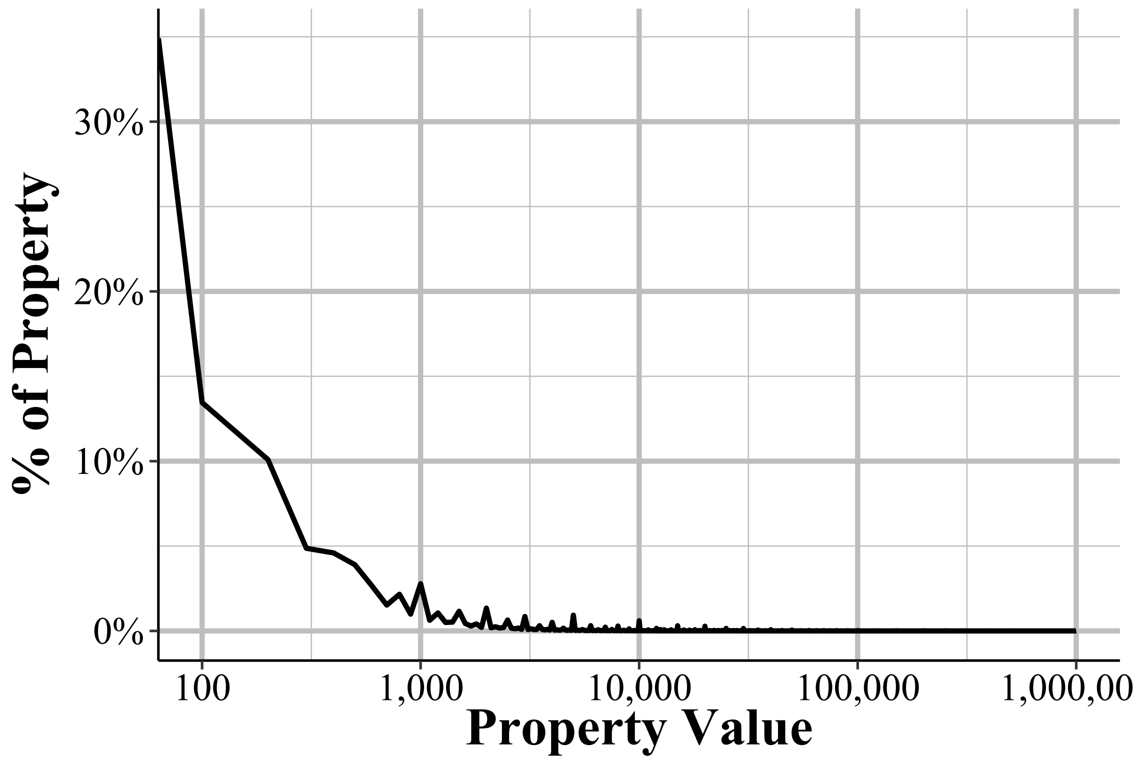


Figure 8.2: The distribution of the value of property stolen. Values are capped at $1,000,000 and each value is rounded to the nearest 100$. The x-axis is set on the log scale as this distribution is hugely right skewed.

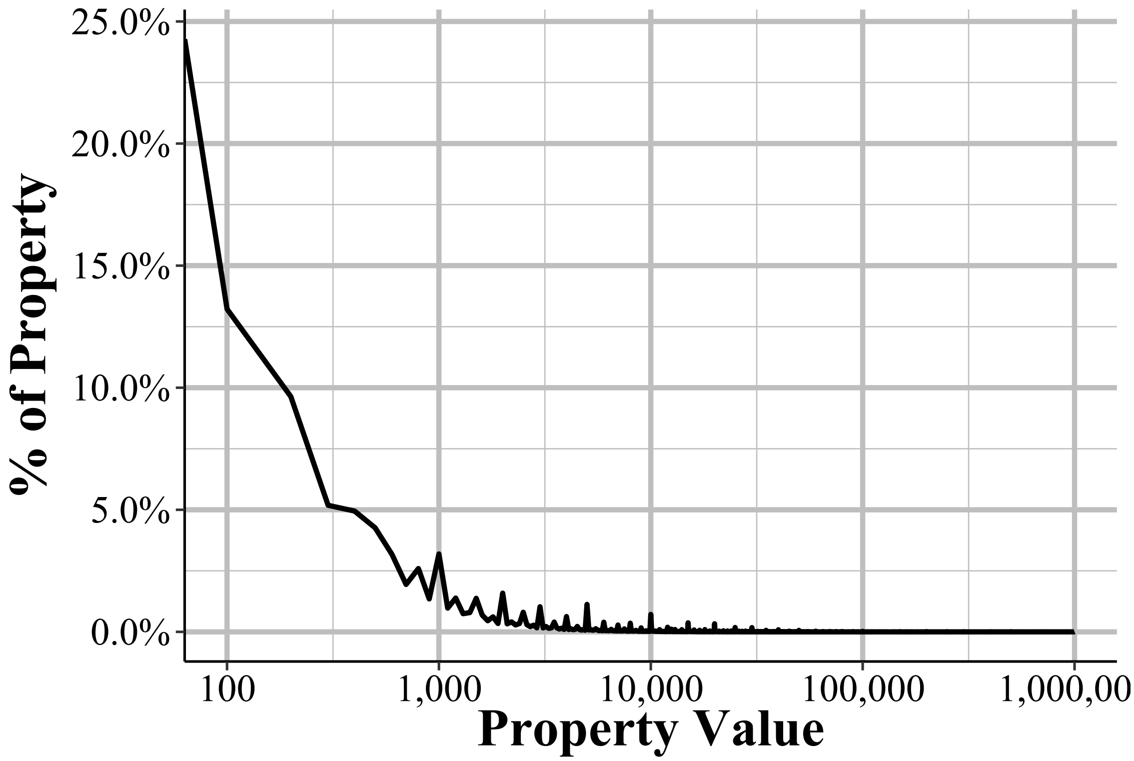


Figure 8.3: The incident-level distribution of the value of property stolen. As values are aggregated to the incident-level, these are higher than the above graph which shows each item individually. Values are capped at $1,000,000andeachvalueisroundedtothenearest$100. The x-axis is set on the log scale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

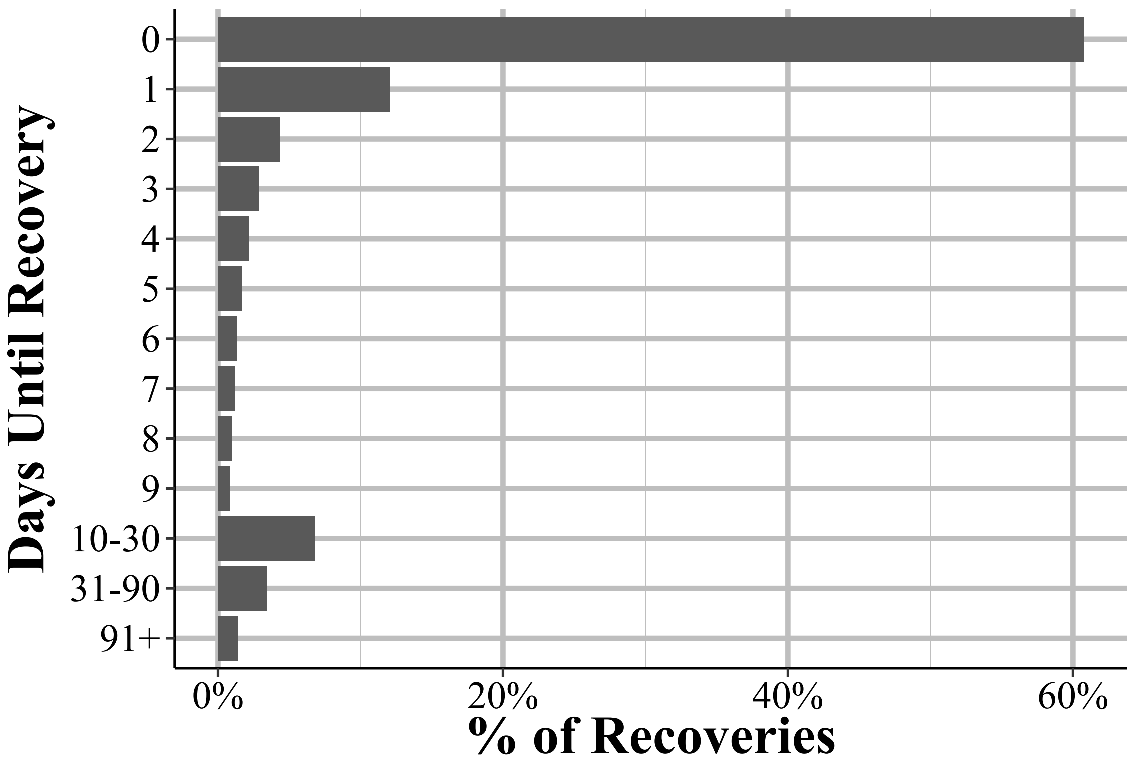


Figure 8.4: The distribution of the number of days from the incident to the property recovered date. In 2019 the maxiumum days from incident to arrest was 450 days. Zero days means that the arrest occurred on the same day as the incident.

Table 8.2: 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%
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

- Dosage Unit/Items (Pills, Etc.)
- Fluid Ounce
- Gallon
- Gram
- Kilogram
- Liter
- Milliliter
- Not Reported

- Number of Plants
- Ounce
- Pound

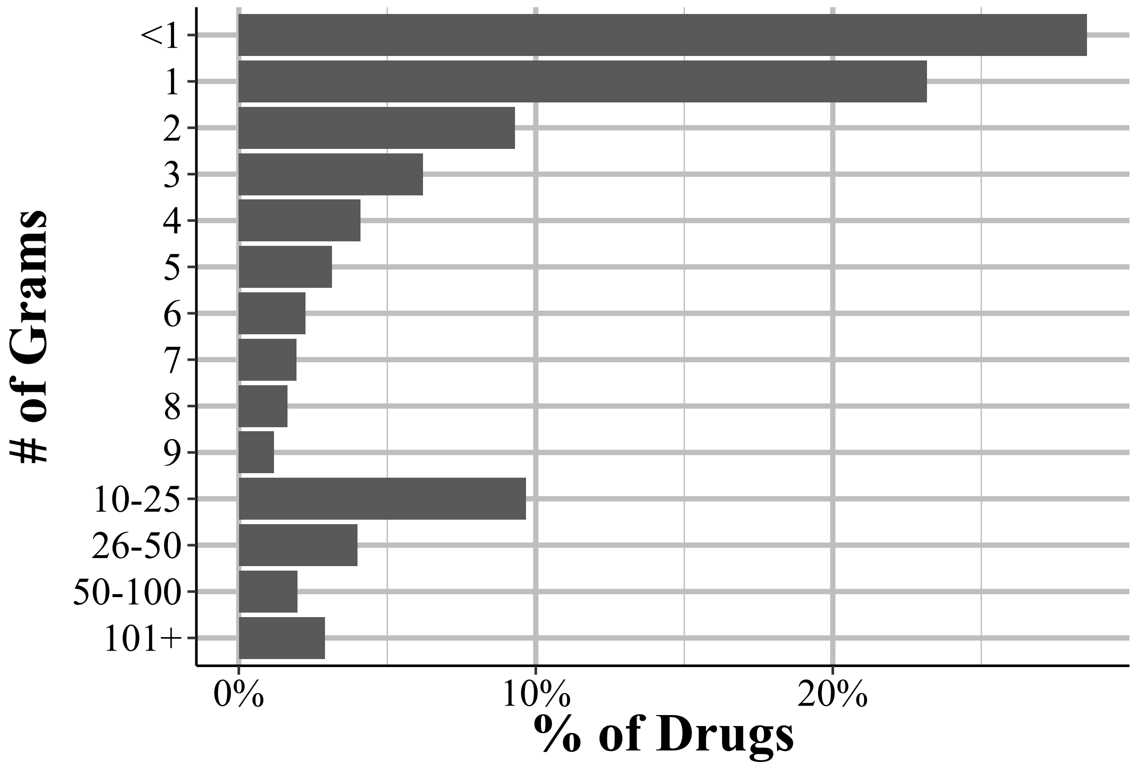


Figure 8.5: For drugs seized that are measured in grams, this figure shows the distribution in the number of grams seized. Values over 10 grams are grouped together for easier interpretation of lower values of drugs seized.