

1.1: Introduction to Crime Data Analysis, R and RStudio

Data Analysis in the Criminal Justice System

Data analysis in the criminal justice system is often associated with policing. However, intelligence from analysis is a cornerstone across various stages of the criminal justice system, not just within policing. Crime analysts are essential players, providing an objective understanding of the current status and data-driven strategies that inform decision-making and improve effectiveness. Let's examine how crime analysts can contribute to various parts of the criminal justice system.

First, crime analysts are decisive in facilitating investigations and supporting law enforcement agencies in solving crimes. Through meticulous data analysis, they uncover trends and connections that help investigations and lead to the apprehension of offenders. They can also contribute to creating problem-solving strategies to prevent future crimes, allowing agencies to stay ahead of potential threats. For example, analysts can identify hot spot areas and analyze why crime is more concentrated (e.g., motels being used for drug use and prostitution).

Once the underlying problem is revealed, crime analysts can discuss with government agencies to take action to address it (e.g., a nuisance abatement lawsuit can be filed by prosecutors against the motel owner whose property has been used for illegal activities repeatedly). Beyond law enforcement, crime analysts can contribute to broader public safety initiatives and improve overall community well-being. Crime analysts can identify the patterns and provide information to increase the quality of internal operations and resource allocation within agencies. The information from crime analysts enables agencies to address chronic problems more efficiently. Crime analysts can also enhance traffic safety and community quality of life by constructing and implementing a model that allows cars to reach their destination quickly. Furthermore, crime analysts can play significant roles by providing educational materials informing the public of crime-related information and prevention/intervention strategies. They may share the results from their analyses using scientific data. Sharing data-driven information can help communities understand and actively participate in efforts to stop crime and promote the effectiveness of agency-led programs.

Also, crime analysts play vital roles in correctional settings and courts. They examine inmate behavior and trends in correctional settings to identify security risks, create intervention strategies, and improve overall facility safety. Also, they may assist in evaluating the effectiveness of rehabilitation programs and informing decision-making related to inmate management. Crime analysts can also provide expert testimony and statistical analysis to support court legal proceedings. They may examine crime information to assess the impact of proposed legislation or policies, evaluate recidivism rates, and inform sentencing decisions. By providing data-driven information, crime analysts add to the fair and effective administration of justice within the legal system.

In short, the work of crime analysts underscores the difficult role of information in promoting safer societies and enhancing the effectiveness of the criminal justice system as a whole, encompassing law enforcement correctional settings and courts.

Importance of Statistics and Statistical Software Programs for Crime Analysts

Statistics and statistical software are essential tools for crime analysts, making them central topics in this book. Proficiency in these areas empowers crime analysts to examine information, identify patterns, effectively make informed decisions, forecast future trends, evaluate interventions, and communicate findings with law enforcement agencies and communities. This book aims to provide comprehensive guidance on these essential skills to improve crime analysts' analytical capabilities and effectiveness in their work. Crime analysts are crucial in generating essential information for the decision-making of criminal justice agencies. They use statistical software to extract valuable insights from diverse data sources relevant to criminal activities. Crime analysts deliver refined insights to criminal justice agencies by meticulously transforming raw data into meaningful information. For example, the information provided by crime analysts can significantly impact police operations. Once absorbed and understood, such information evolves into actionable knowledge crucial for shaping police actions and strategies, forming the foundation for effective law enforcement initiatives.

The transition from raw data to actionable knowledge involves two interconnected processes. First, data undergoes thorough analysis to extract patterns, trends, and relevant insights. Applying statistical principles and software programs facilitates this transformation from data to structured information. Subsequently, this information is effectively communicated, elevating it to actionable knowledge. Without a solid understanding of statistics and proficiency in statistical software, analysts will encounter substantial obstacles in extracting actionable knowledge from raw data. Crime analysts can use crime data that are collected from various sources. Many criminal justice agencies collect firsthand data and make them available to the public. For example, the Bureau of Justice Statistics collects the Police-Public Contact Survey, which includes data regarding US residents who had contact

with the police and the nature of this interaction. This dataset is available to the public. If a crime analyst is interested in the circumstances regarding traffic stops, they can use this dataset to analyze the patterns.

Instead of using existing datasets, crime analysts can gather primary data through firsthand methods. For instance, they may conduct community surveys or organizational surveys for police officers to identify crime-related problems or concerns. Crime analysts can also work within correctional settings and study inmate behaviors. By analyzing infraction reports, crime analysts may identify potential security threats such as gang activity, contraband distribution, or inmate conflicts. They can also examine incident reports to observe trends in inmate misconduct, violence, or disciplinary infractions to help prison staff anticipate and prevent future incidents.

The greater the familiarity crime analysts have with statistics and software programs, the more effectively they can bridge the gap between raw data and informed action. Their skillful use of statistics and statistical software programs ensures that the valuable insights from analysis result in informed decision-making within criminal justice agencies. In this book, I will explore a range of statistical analyses, from chi-square tests to multiple regression analysis. I will also demonstrate how to manage and handle data effectively using the R programming language to perform these analyses.

What Is R and RStudio?

R is a computer software that can be used for crime data analysis. R is popular among people who handle and analyze data for several reasons. First, it is free. Anybody can use this software without having to pay a subscription fee. Second, R allows people to conduct various statistical analyses and compute graphical images, from very simple to advanced statistics. A problem with R is that it is not easy to use, especially for those who have not used it before. Writing codes can be overwhelming, but R's user interface does not help make it easy. RStudio is a software that makes R more user-friendly by providing various functions and support.

Setting Up R and RStudio

Now that I have discussed why data analysis is important in the criminal justice field, let us analyze the data to see the value of analysis. You will need to install R and RStudio, and you can use [R Is for Rams](#) to follow the instructions and complete the installation process successfully.

Calculating Using R

While R is incredibly flexible and supports numerous statistical packages, at its core, it functions as a big, powerful calculator. The basic use of R helps you analyze criminal justice data. I believe most of you have not used R before, but if you follow the instructions below, you will not have major trouble producing the same outcome as mine. Since R is a big calculator, we will do some calculations.

Step 1. Launch RStudio

Open the RStudio application on your computer.

Step 2. Choose Script or Console

You can work with R in RStudio using either the script editor or the console. The console lets you execute commands interactively, while the script editor lets you write and save scripts for more complex tasks. Today, we will use the console.

Step 3. Type R Code

In the R console of RStudio, you can directly type R code.

```
#calculate multiplications
5*4*3*2*1
## [1] 120

#you can also do the same thing using the following syntax
factorial(5)
## [1] 120
```

Hashtag # for Annotation

Have you noticed I added a hashtag—"#"—before certain descriptions? The hashtag signals to R that the text following it on the same line is a comment and should not be processed or saved as part of the code. Comments are not essential for code execution but are valuable for explaining and recalling the purpose of the code. Annotating code is considered a best practice in programming.

A vector is used to store collected elements of the same data type in R. These elements include numbers and characters. It may be easier to see examples to understand the concept of vector.

```
#calculate multiplications
5*4*3*2*1
## [1] 120
#you can also do the same thing using the following syntax
factorial(5)
## [1] 120
#you can create numeric vectors
a1 <- 1
a2 <- 2
#you can create character vectors
b1 <- "criminal"
b2 <- "justice"
```

Did you notice that "<-" was used here? In R, the <- symbol is used as an assignment operator to assign values to variables. It is used to create or update vectors by assigning a value or expression to a vector name. This operator is often referred to as the "left arrow" operator. Here, we created two vectors, a1 and a2, and assigned 1 and 2, respectively. As you can see above, when you create character vectors, you need to enclose the text or characters within double quotation marks. If you want to check if vector values are properly assigned, you can type a vector name.

```
#you can double check the vector values
a1
## [1] 1
a2
## [1] 2
b1
## [1] "criminal"
b2
## [1] "justice"
```

Let us continue to use R as a calculator. We can calculate using the vectors we created earlier. Using vectors makes our analysis much easier and clearer. A vector can contain multiple elements, and we can write simple code to perform more complex calculations.

```
#you can add numeric vectors
a1+a2
## [1] 3
#you can multiply numeric vectors
a1*a2
## [1] 2
```

```
#you can subtract vectors
a1-a2
## [1] -1
#you can divide vectors
a1/a2
## [1] 0.5
```

Please note that you cannot calculate using character vectors. For example, `b1+b2` would not work in R. You can combine multiple values of the same data type into a single vector using `'c()'` function.

```
#create a numeric vector
x <- c(1,2,3,4,5)
#you can check the outcome of this function
x
[1] 1 2 3 4 5
#create a character vector
y <- c("Life", "is", "good")
```

You can use a function to vectors in R. A function refers to a block of code that allows us to perform different types of tasks. Let me give you an example demonstrating how we can use R to simplify a task.

```
# There are many ways to calculate the arithmetic mean of a numeric
vector. You can type.
(1+2+3+4+5)/5
## [1] 3
# But, an easier way to accomplish the same task is to apply a function to the vector
we created earlier.
mean(x)
## [1] 3
```

The `mean()` function calculates the arithmetic mean or average of a numeric vector or a group of numeric values. You can compute the arithmetic mean by dividing the sum of all values by the total number of values.

R Packages (How To Install and Load Packages)

The basic R functions that come with the software are highly versatile but have limitations. Additional functions are available in packages developed by researchers and contributors worldwide to extend R's capabilities, which are then integrated into the R open-source platform. Throughout this textbook, we will leverage many of these packages. One widely used package we will incorporate is the `tidyverse`[JC1] package. The `tidyverse` encompasses various R packages tailored for data science and statistical analysis – simply it is a package that contains multiple packages. These packages synergistically interact to offer a unified framework for data manipulation, visualization, and analytics. Let's practice installing a package and operating it. Before using a package, you must first install it. You can accomplish the installation by using the R function

```
install.packages(), as illustrated below:
Install.packages("tidyverse")
```

After installing the `tidyverse` package, it needs to be loaded for use. Unlike the installation process, each time you wish to use a package, it must be loaded beforehand.

```
Library(tidyverse)
# Load the tidyverse package
```

```
library(tidyverse)
```

To briefly demonstrate how you can use the loaded package, I will use the built-in ‘USArrests’ dataset. You do not need to download this since the USArrests dataset is actually embedded in the basic R package. The dataset includes information on the number of arrests per 100,000 residents for assault, murder, and rape from the 50 US states in 1973. Additionally, it provides for the percentage of the population residing in urban areas. Using the tidyverse package, I filter the dataset to include only certain states and visualize the relationship between the two variables below. Specifically, I wanted to filter the dataset to include only states with a murder rate above 7 and then visualize the number of assaults per 100,000 annually and the percent of the state population living in urban areas using a scatter plot.

```
# Print the first few rows of the dataset
print(head(USArrests))

# Filter the dataset to include only states with a murder rate above
7
filtered_data <- USArrests %>%
  filter(Murder > 7)
15

# Print the filtered dataset
print(filtered_data)

# Visualize the relationship between Assault and UrbanPop
ggplot(filtered_data, aes(x = Assault, y = UrbanPop)) + geom_point() +
  labs(title = "Relationship between Assault and UrbanPop",
       x = "Assault", y = "UrbanPop") + theme_minimal()
```

Conclusion

This chapter has provided a comprehensive introduction to the role of crime analysts and has emphasized the importance of understanding statistics and R programming. It has outlined the fundamental concepts of R and RStudio, along with basic code examples. In the upcoming chapter, we will delve more deeply into data transformation and visualization techniques.

This page titled [1.1: Introduction to Crime Data Analysis, R and RStudio](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Jaeyong Choi](#) ([The Pennsylvania Alliance for Design of Open Textbooks \(PA-ADOPT\)](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.