CYT-250-Threat Investigation

Final Project

Seeing VERIS in Action-Working with VCDB Data

Elaborate by:

Leandro Delgado

Student Number: 114416241

Professor: Tatiana Outkina

Introduction

In this lab, I continued working with the VCDB dataset from Verizon, following the exercises outlined in Chapter 8 of the book *Data-Driven Security*. This chapter builds upon the work from Lab 7, where the Python script from Listing 7-2 was already executed. From this point forward, all scripts are provided in R, marking a shift in tools and approach.

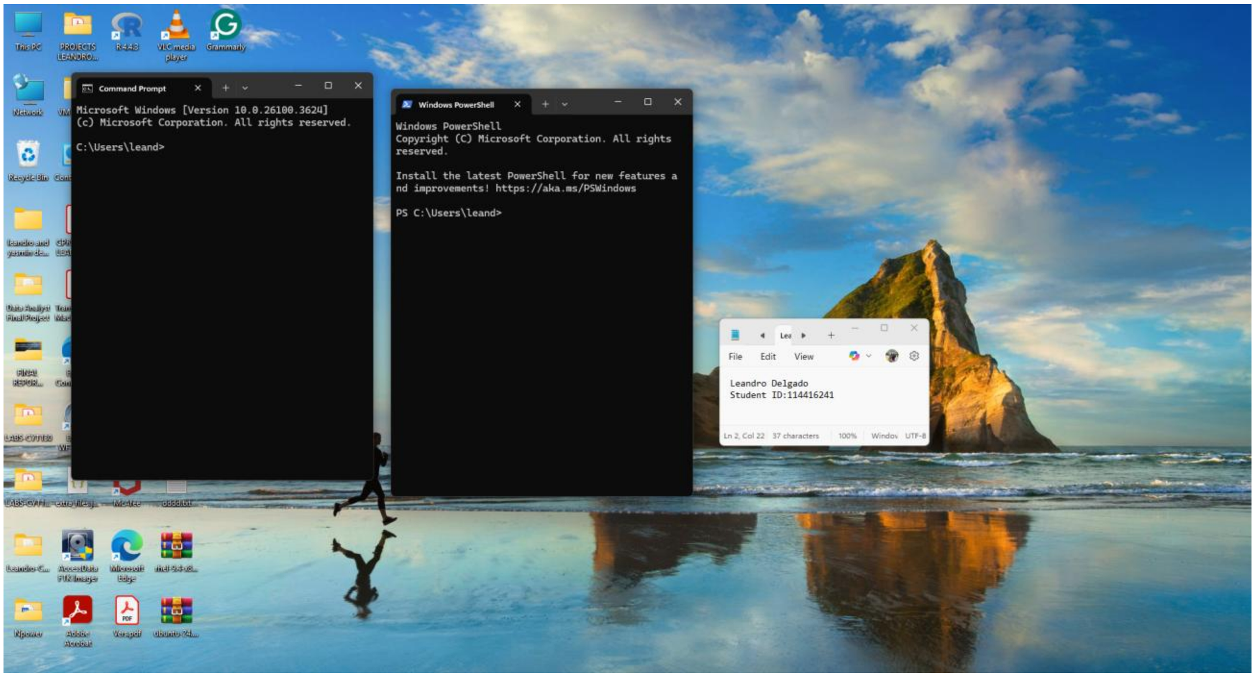
As someone new to using the R environment, this lab also served as an opportunity to explore and become comfortable with this powerful yet approachable tool. My previous experience has mostly been with Google Colab, but due to recurring issues with installing libraries and running code there, I decided to switch to RStudio. Although R is not overly complex, I've come to realize that it requires practice and familiarity both of which I'm starting to develop through this lab.

The objective here is to study the chapter thoroughly, execute all the R scripts, and document the process with screenshots, meaningful comments, and personal insights. At each step, I reflect on what I've learned, the challenges I encountered, and how I overcame them, making this not only a technical exercise but also a personal journey of growth in data analysis and cybersecurity.

Objective: Seeing VERIS in Action. Working with VCDB Data

Sources of information:

Data-Driven security, Chapter 8. This is in continuation Lab 7 work on VCDB from Verizon. You already run the Python script in Listing 7-2. All other scripts are given in R. Your task is to study the Chapter and run all scripts. Make screenshots and meaningful comments. Explain what you have learned at each step.



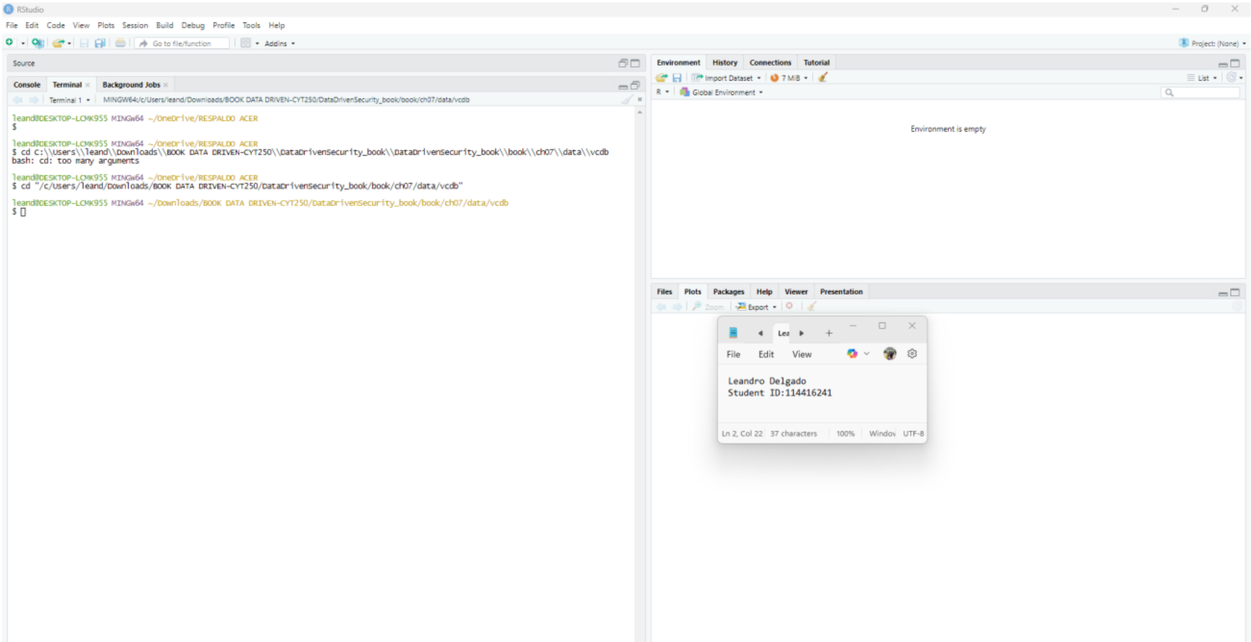
```
Microsoft Windows [Version 10.0.26100.3624]
(c) Microsoft Corporation. All rights reserved.

C:\Users\leand>
```

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights
reserved.

Install the latest PowerShell for new features a
nd improvements! https://aka.ms/PSWindows

PS C:\Users\leand>
```

Leandro Delgado
Student ID:114416241

Figure 1. 0 Screen

Task 1: Set R environment



```
leand@DESKTOP-LCMK955 MINGW64 ~/OneDrive/RESPALDO ACER
$

leand@DESKTOP-LCMK955 MINGW64 ~/OneDrive/RESPALDO ACER
$ cd C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book\\DataDrivenSecurity_book\\book\\ch07\\data\\vcdb
bash: cd: too many arguments

leand@DESKTOP-LCMK955 MINGW64 ~/OneDrive/RESPALDO ACER
$ cd "/c/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/book/ch07/data/vcdb"

leand@DESKTOP-LCMK955 MINGW64 ~/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/book/ch07/data/vcdb
$ 
```

```
Leandro Delgado
Student ID:114416241
```

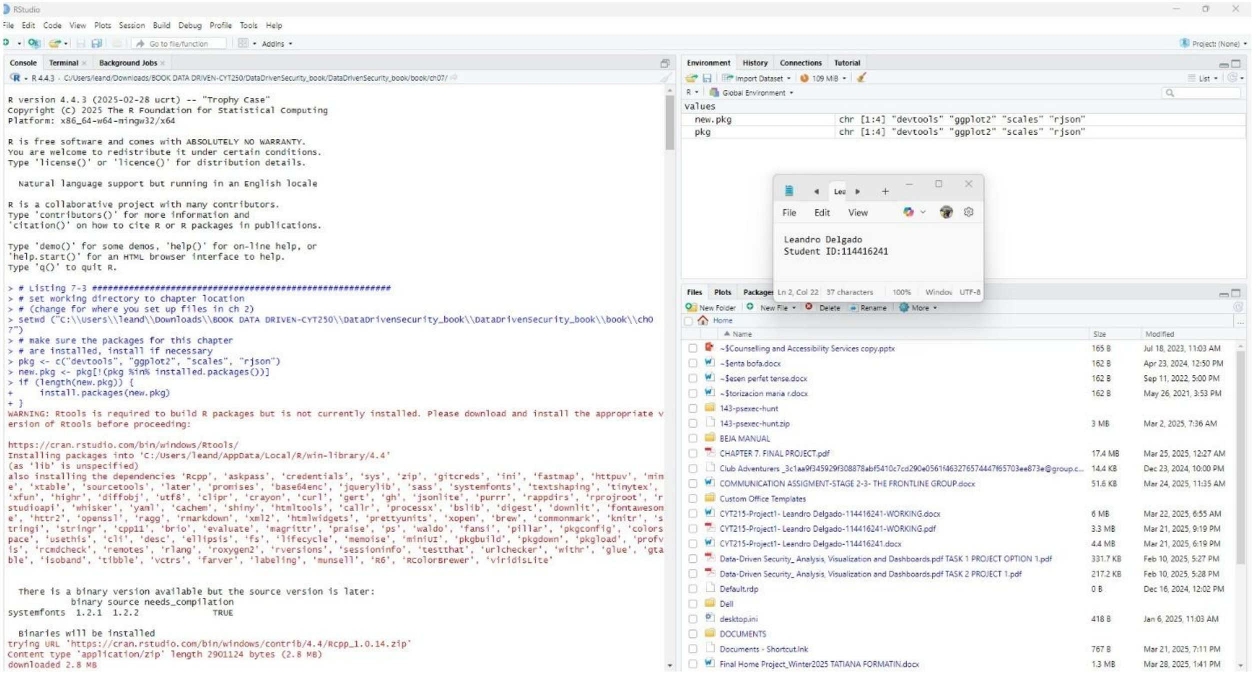
Figure 2. Set R environment

This was my first time using the R environment. My previous experience had been mainly with Google Colab. However, due to several technical issues with installing the right libraries and running the code properly there, I decided to try RStudio instead. It’s a new tool for me — not too complex, but I know I need more practice to feel confident using it. I’ve started to explore how it works and I'm gradually getting more comfortable with it.

Task 2: Study Chapter 7 material and run Listings 7-3 to 7-11

Listing 7-3

To begin, I configured the working directory so R could easily access the chapter’s data files. I then ran a script that checks for essential packages needed for the analysis. If any of them—like ggplot2 or rjson—weren’t already installed, the script handled it automatically. This step helped me get the R environment ready for the work ahead without manually managing each dependency.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console Terminal Background Jobs
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
Environment History Connections Tutorial
Import Dataset 109 MiB
List
R Global Environment
Files Plots Packages
New Folder New File Delete Rename More
Home
File Edit View
Lea

```
R version 4.4.3 (2025-02-28 ucrt) -- "Trophy Case"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> # Listing 7-3 ##########################################################
> # set working directory to chapter location
> # (change for where you set up files in ch 2)
> setwd ("C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book\\DataDrivenSecurity_book\\book\\ch0
7")
> # make sure the packages for this chapter
> # are installed, install if necessary
> pkg <- c("devtools", "ggplot2", "scales", "rjson")
> new.pkg <- pkg[!(pkg %in% installed.packages())]
> if (length(new.pkg)) {
+     install.packages(new.pkg)
+ }
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate v
ersion of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing packages into 'C:/Users/leand/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
also installing the dependencies 'Rcpp', 'askpass', 'credentials', 'sys', 'zip', 'gitcreds', 'ini', 'fastmap', 'httpuv', 'mim
e', 'xtable', 'sourcetools', 'later', 'promises', 'base64enc', 'jquerylib', 'sass', 'systemfonts', 'textshaping', 'tinytex',
'xfun', 'highr', 'diffobj', 'utf8', 'clipr', 'crayon', 'curl', 'gert', 'gh', 'jsonlite', 'purrr', 'rappdirs', 'rprojroot', 'r
studioapi', 'whisker', 'yaml', 'cachem', 'shiny', 'htmltools', 'callr', 'processx', 'bslib', 'digest', 'downlit', 'fontawesom
e', 'httr2', 'openssl', 'ragg', 'rmarkdown', 'xml2', 'htmlwidgets', 'prettyunits', 'xopen', 'brew', 'commonmark', 'knitr', 's
tringi', 'stringr', 'cpp11', 'brio', 'evaluate', 'magrittr', 'praise', 'ps', 'waldo', 'fansi', 'pillar', 'pkgconfig', 'colors
pace', 'usethis', 'cli', 'desc', 'ellipsis', 'fs', 'lifecycle', 'memoise', 'miniUI', 'pkgbuild', 'pkgdown', 'pkgload', 'profv
is', 'rcmdcheck', 'remotes', 'rlang', 'roxygen2', 'rversions', 'sessioninfo', 'testthat', 'urlchecker', 'withr', 'glue', 'gta
ble', 'isoband', 'tibble', 'vctrs', 'farver', 'labeling', 'munsell', 'R6', 'RColorBrewer', 'viridisLite'


  There is a binary version available but the source version is later:
           binary source needs_compilation
systemfonts  1.2.1  1.2.2              TRUE

  Binaries will be installed
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/Rcpp_1.0.14.zip'
Content type 'application/zip' length 2901124 bytes (2.8 MB)
downloaded 2.8 MB
```

values

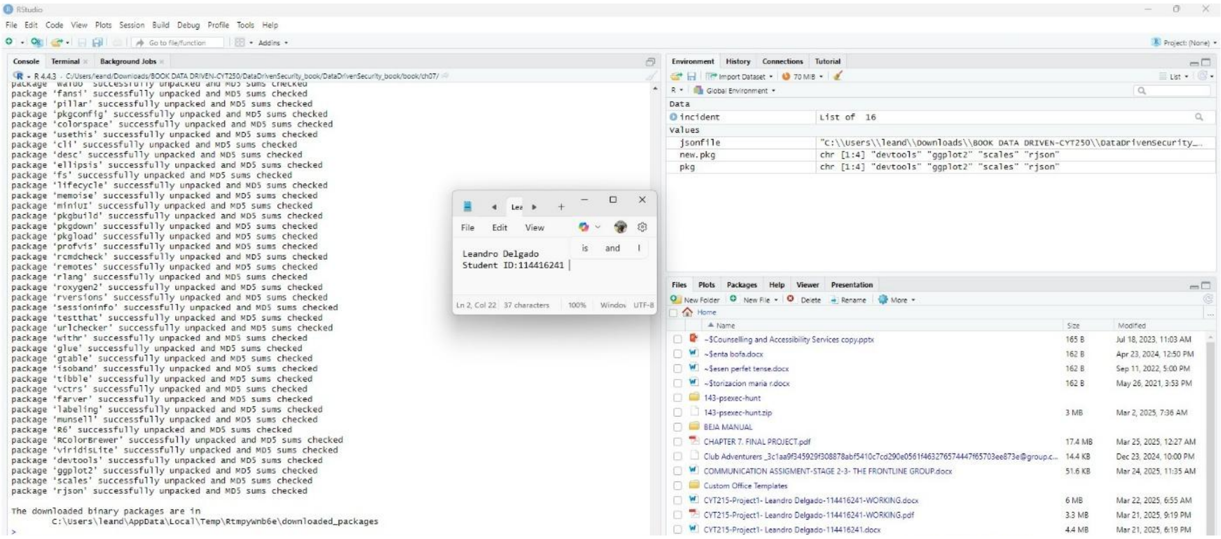
new.pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"

```
Leandro Delgado
Student ID:114416241
```

Ln 2, Col 22 | 37 characters | 100% | Windov | UTF-8

Name	Size	Modified
~$Counselling and Accessibility Services copy.pptx	165 B	Jul 18, 2023, 11:03 AM
~$enta bofa.docx	162 B	Apr 23, 2024, 12:50 PM
~$esen perfet tense.docx	162 B	Sep 11, 2022, 5:00 PM
~$torizacion maria r.docx	162 B	May 26, 2021, 3:53 PM
143-psexec-hunt		
143-psexec-hunt.zip	3 MB	Mar 2, 2025, 7:36 AM
BEJA MANUAL		
CHAPTER 7. FINAL PROJECT.pdf	17.4 MB	Mar 25, 2025, 12:27 AM
Club Adventurers _3c1aa9f345929f308878abf5410c7cd290e0561f463276574447f65703ee873e@group.c...	14.4 KB	Dec 23, 2024, 10:00 PM
COMMUNICATION ASSIGMENT-STAGE 2-3- THE FRONTLINE GROUP.docx	51.6 KB	Mar 24, 2025, 11:35 AM
Custom Office Templates		
CYT215-Project1- Leandro Delgado-114416241-WORKING.docx	6 MB	Mar 22, 2025, 6:55 AM
CYT215-Project1- Leandro Delgado-114416241-WORKING.pdf	3.3 MB	Mar 21, 2025, 9:19 PM
CYT215-Project1- Leandro Delgado-114416241.docx	4.4 MB	Mar 21, 2025, 6:19 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 1 PROJECT OPTION 1.pdf	331.7 KB	Feb 10, 2025, 5:27 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 2 PROJECT 1.pdf	217.2 KB	Feb 10, 2025, 5:28 PM
Default.rdp	0 B	Dec 16, 2024, 12:02 PM
Dell		
desktop.ini	418 B	Jan 6, 2025, 11:03 AM
DOCUMENTS		
Documents - Shortcut.lnk	767 B	Mar 21, 2025, 7:11 PM
Final Home Project_Winter2025 TATIANA FORMATIN.docx	1.3 MB	Mar 28, 2025, 1:41 PM

Figure 3. Run listing 7-3


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console Terminal Background Jobs
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
Environment History Connections Tutorial
Import Dataset 70 MiB
List
R Global Environment
Files Plots Packages Help Viewer Presentation
New Folder New File Delete Rename More
Home
File Edit View
Lea
is and I

```
package 'waldo' successfully unpacked and MD5 sums checked
package 'fansi' successfully unpacked and MD5 sums checked
package 'pillar' successfully unpacked and MD5 sums checked
package 'pkgconfig' successfully unpacked and MD5 sums checked
package 'colorspace' successfully unpacked and MD5 sums checked
package 'usethis' successfully unpacked and MD5 sums checked
package 'cli' successfully unpacked and MD5 sums checked
package 'desc' successfully unpacked and MD5 sums checked
package 'ellipsis' successfully unpacked and MD5 sums checked
package 'fs' successfully unpacked and MD5 sums checked
package 'lifecycle' successfully unpacked and MD5 sums checked
package 'memoise' successfully unpacked and MD5 sums checked
package 'miniUI' successfully unpacked and MD5 sums checked
package 'pkgbuild' successfully unpacked and MD5 sums checked
package 'pkgdown' successfully unpacked and MD5 sums checked
package 'pkgload' successfully unpacked and MD5 sums checked
package 'profvis' successfully unpacked and MD5 sums checked
package 'rcmdcheck' successfully unpacked and MD5 sums checked
package 'remotes' successfully unpacked and MD5 sums checked
package 'rlang' successfully unpacked and MD5 sums checked
package 'roxygen2' successfully unpacked and MD5 sums checked
package 'rversions' successfully unpacked and MD5 sums checked
package 'sessioninfo' successfully unpacked and MD5 sums checked
package 'testthat' successfully unpacked and MD5 sums checked
package 'urlchecker' successfully unpacked and MD5 sums checked
package 'withr' successfully unpacked and MD5 sums checked
package 'glue' successfully unpacked and MD5 sums checked
package 'gtable' successfully unpacked and MD5 sums checked
package 'isoband' successfully unpacked and MD5 sums checked
package 'tibble' successfully unpacked and MD5 sums checked
package 'vctrs' successfully unpacked and MD5 sums checked
package 'farver' successfully unpacked and MD5 sums checked
package 'labeling' successfully unpacked and MD5 sums checked
package 'munsell' successfully unpacked and MD5 sums checked
package 'R6' successfully unpacked and MD5 sums checked
package 'RColorBrewer' successfully unpacked and MD5 sums checked
package 'viridisLite' successfully unpacked and MD5 sums checked
package 'devtools' successfully unpacked and MD5 sums checked
package 'ggplot2' successfully unpacked and MD5 sums checked
package 'scales' successfully unpacked and MD5 sums checked
package 'rjson' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\leand\AppData\Local\Temp\RtmpyWnb6e\downloaded_packages
>
```

Data

incident	List of 16

Values

jsonfile	"C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
new.pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"

```
Leandro Delgado
Student ID:114416241
```

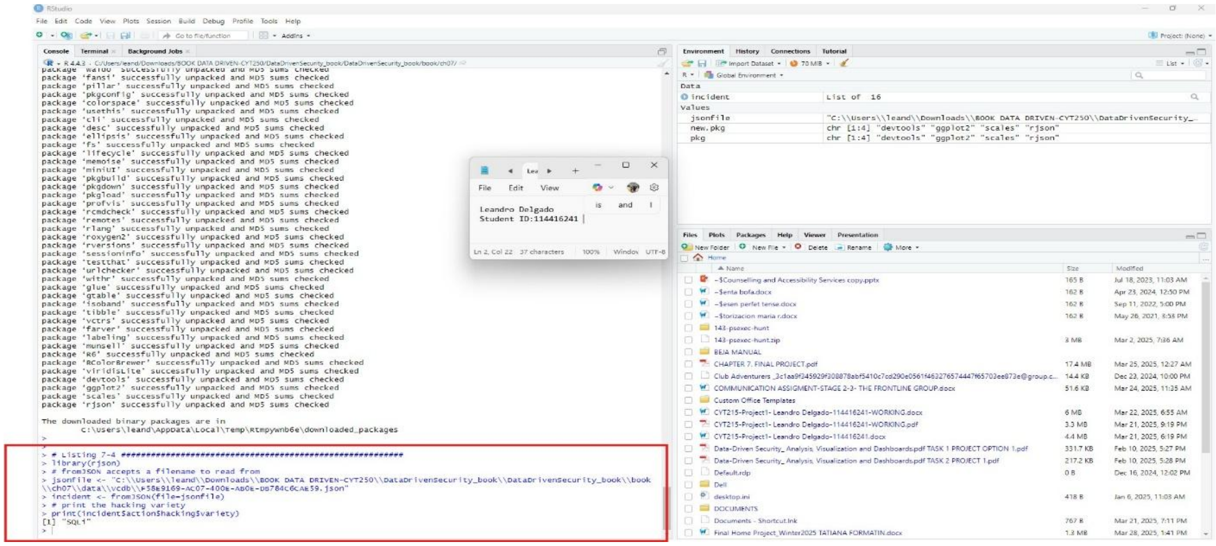
Ln 2, Col 22 | 37 characters | 100% | Windov | UTF-8

Name	Size	Modified
~$Counselling and Accessibility Services copy.pptx	165 B	Jul 18, 2023, 11:03 AM
~$enta bofa.docx	162 B	Apr 23, 2024, 12:50 PM
~$esen perfet tense.docx	162 B	Sep 11, 2022, 5:00 PM
~$torizacion maria r.docx	162 B	May 26, 2021, 3:53 PM
143-psexec-hunt		
143-psexec-hunt.zip	3 MB	Mar 2, 2025, 7:36 AM
BEJA MANUAL		
CHAPTER 7. FINAL PROJECT.pdf	17.4 MB	Mar 25, 2025, 12:27 AM
Club Adventurers _3c1aa9f345929f308878abf5410c7cd290e0561f463276574447f65703ee873e@group.c...	14.4 KB	Dec 23, 2024, 10:00 PM
COMMUNICATION ASSIGMENT-STAGE 2-3- THE FRONTLINE GROUP.docx	51.6 KB	Mar 24, 2025, 11:35 AM
Custom Office Templates		
CYT215-Project1- Leandro Delgado-114416241-WORKING.docx	6 MB	Mar 22, 2025, 6:55 AM
CYT215-Project1- Leandro Delgado-114416241-WORKING.pdf	3.3 MB	Mar 21, 2025, 9:19 PM
CYT215-Project1- Leandro Delgado-114416241.docx	4.4 MB	Mar 21, 2025, 6:19 PM

Figure 4. Result of run Listing 7-3

Listing 7-4

In this step, I loaded a specific JSON file from the VCDB dataset and parsed it using the from JSON() function. Then, I printed the “hacking variety” field, which returned "SQLi", showing that the incident involved a SQL injection. It was interesting to see how structured data from a file can be easily accessed and explored in R.



```
package 'waldo' successfully unpacked and MD5 sums checked
package 'fansi' successfully unpacked and MD5 sums checked
package 'pillar' successfully unpacked and MD5 sums checked
package 'pkgconfig' successfully unpacked and MD5 sums checked
package 'colorspace' successfully unpacked and MD5 sums checked
package 'usethis' successfully unpacked and MD5 sums checked
package 'cli' successfully unpacked and MD5 sums checked
package 'desc' successfully unpacked and MD5 sums checked
package 'ellipsis' successfully unpacked and MD5 sums checked
package 'fs' successfully unpacked and MD5 sums checked
package 'lifecycle' successfully unpacked and MD5 sums checked
package 'memoise' successfully unpacked and MD5 sums checked
package 'miniUI' successfully unpacked and MD5 sums checked
package 'pkgbuild' successfully unpacked and MD5 sums checked
package 'pkgdown' successfully unpacked and MD5 sums checked
package 'pkgload' successfully unpacked and MD5 sums checked
package 'profvis' successfully unpacked and MD5 sums checked
package 'rcmdcheck' successfully unpacked and MD5 sums checked
package 'remotes' successfully unpacked and MD5 sums checked
package 'rlang' successfully unpacked and MD5 sums checked
package 'roxygen2' successfully unpacked and MD5 sums checked
package 'rversions' successfully unpacked and MD5 sums checked
package 'sessioninfo' successfully unpacked and MD5 sums checked
package 'testthat' successfully unpacked and MD5 sums checked
package 'urlchecker' successfully unpacked and MD5 sums checked
package 'withr' successfully unpacked and MD5 sums checked
package 'glue' successfully unpacked and MD5 sums checked
package 'gtable' successfully unpacked and MD5 sums checked
package 'isoband' successfully unpacked and MD5 sums checked
package 'tibble' successfully unpacked and MD5 sums checked
package 'vctrs' successfully unpacked and MD5 sums checked
package 'farver' successfully unpacked and MD5 sums checked
package 'labeling' successfully unpacked and MD5 sums checked
package 'munsell' successfully unpacked and MD5 sums checked
package 'R6' successfully unpacked and MD5 sums checked
package 'RColorBrewer' successfully unpacked and MD5 sums checked
package 'viridisLite' successfully unpacked and MD5 sums checked
package 'devtools' successfully unpacked and MD5 sums checked
package 'ggplot2' successfully unpacked and MD5 sums checked
package 'scales' successfully unpacked and MD5 sums checked
package 'rjson' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\leand\AppData\Local\Temp\RtmpywnbGe\downloaded_packages
>
>
> # Listing 7-4 ###########################################################
> library(rjson)
> # fromJSON accepts a filename to read from
> jsonfile <- "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book\\DataDrivenSecurity_book\\book
\\ch07\\data\\vcdb\\F58E9169-AC07-400E-AB0E-DB784C6CAE59.json"
> incident <- fromJSON(file=jsonfile)
> # print the hacking variety
> print(incident$action$hacking$variety)
[1] "SQLi"
> |
```

```
Leandro Delgado
Student ID:114416241
```

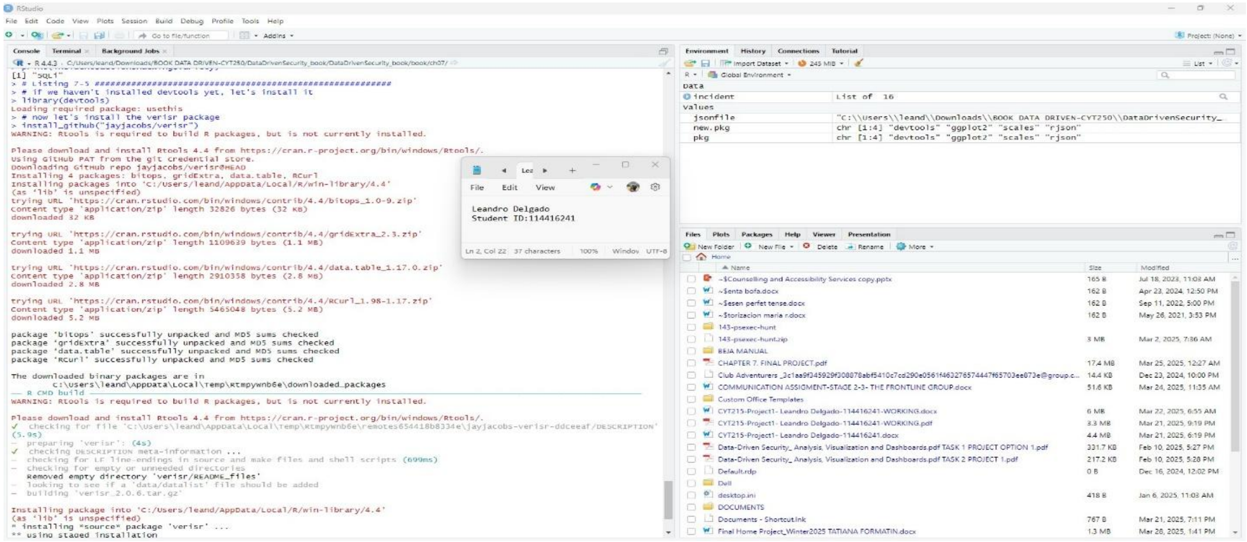
```
Data
incident                  List of  16
Values
  jsonfile                "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_
  new.pkg                 chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg                     chr [1:4] "devtools" "ggplot2" "scales" "rjson"
```

Name	Size	Modified
~$Counselling and Accessibility Services copy.pptx	165 B	Jul 18, 2023, 11:03 AM
~$enta bofa.docx	162 B	Apr 23, 2024, 12:50 PM
~$esen perfet tense.docx	162 B	Sep 11, 2022, 5:00 PM
~$torizacion maria r.docx	162 B	May 26, 2021, 3:53 PM
143-psexec-hunt		
143-psexec-hunt.zip	3 MB	Mar 2, 2025, 7:36 AM
BEJA MANUAL		
CHAPTER 7. FINAL PROJECT.pdf	17.4 MB	Mar 25, 2025, 12:27 AM
Club Adventurers _3c1aa9f345929f308878abf5410c7cd290e0561f463276574447f65703ee873e@group.c...	14.4 KB	Dec 23, 2024, 10:00 PM
COMMUNICATION ASSIGMENT-STAGE 2-3- THE FRONTLINE GROUP.docx	51.6 KB	Mar 24, 2025, 11:35 AM
Custom Office Templates		
CYT215-Project1- Leandro Delgado-114416241-WORKING.docx	6 MB	Mar 22, 2025, 6:55 AM
CYT215-Project1- Leandro Delgado-114416241-WORKING.pdf	3.3 MB	Mar 21, 2025, 9:19 PM
CYT215-Project1- Leandro Delgado-114416241.docx	4.4 MB	Mar 21, 2025, 6:19 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 1 PROJECT OPTION 1.pdf	331.7 KB	Feb 10, 2025, 5:27 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 2 PROJECT 1.pdf	217.2 KB	Feb 10, 2025, 5:28 PM
Default.rdp	0 B	Dec 16, 2024, 12:02 PM
Dell		
desktop.ini	418 B	Jan 6, 2025, 11:03 AM
DOCUMENTS		
Documents - Shortcut.lnk	767 B	Mar 21, 2025, 7:11 PM
Final Home Project_Winter2025 TATIANA FORMATIN.docx	1.3 MB	Mar 28, 2025, 1:41 PM

Figure 5. Result listing 7-4

Listing 7-5

In this step, I used the devtools package to install veris directly from GitHub. Since this package isn’t available on CRAN, installing it this way is necessary. The process required downloading several dependencies, and I was also reminded to install Rtools, which is needed to build packages from source. This step helped me understand how to bring in external R packages that aren’t part of the standard repositories.



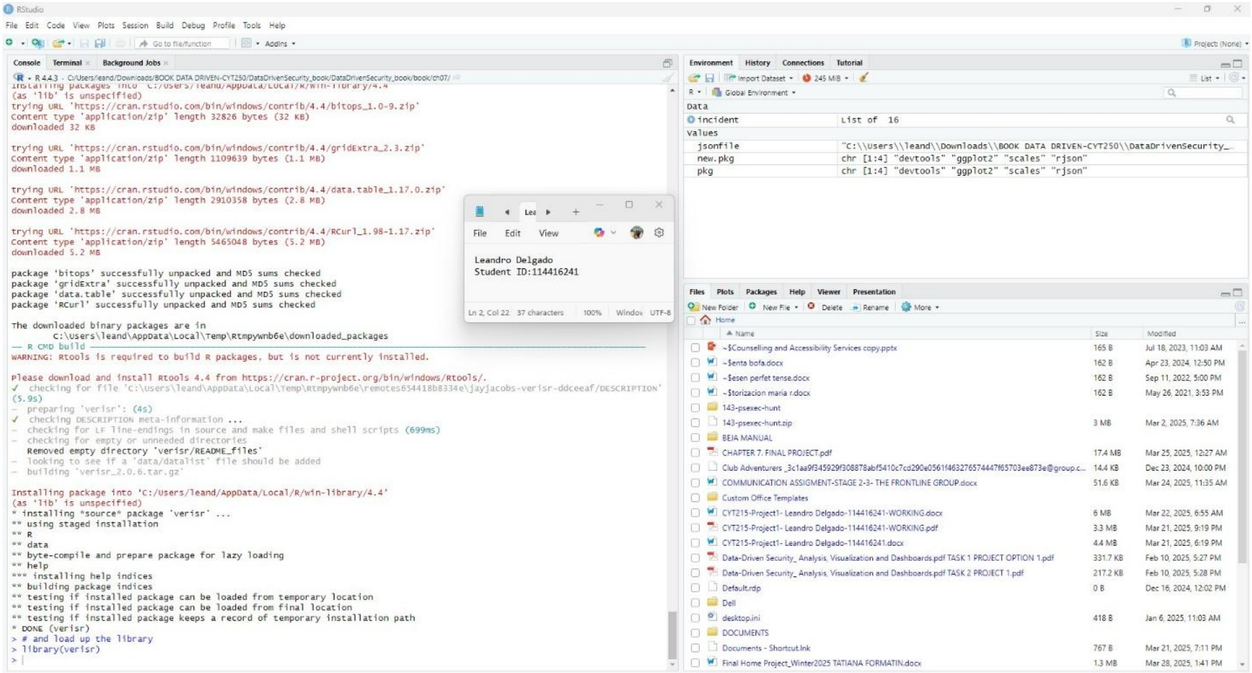
```
[1] "SQLi"
> # Listing 7-5 ###########################################################
> # if we haven't installed devtools yet, let's install it
> library(devtools)
Loading required package: usethis
> # now let's install the verisr package
> install_github("jayjacobs/verisr")
WARNING: Rtools is required to build R packages, but is not currently installed.

Please download and install Rtools 4.4 from https://cran.r-project.org/bin/windows/Rtools/.
Using GitHub PAT from the git credential store.
Downloading GitHub repo jayjacobs/verisr@HEAD
Installing 4 packages: bitops, gridExtra, data.table, RCurl
Installing packages into 'C:/Users/leand/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/bitops_1.0-9.zip'
Content type 'application/zip' length 32826 bytes (32 KB)
downloaded 32 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/gridExtra_2.3.zip'
Content type 'application/zip' length 1109639 bytes (1.1 MB)
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/data.table_1.17.0.zip'
Content type 'application/zip' length 2910358 bytes (2.8 MB)
downloaded 2.8 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/RCurl_1.98-1.17.zip'
Content type 'application/zip' length 5465048 bytes (5.2 MB)
downloaded 5.2 MB

package 'bitops' successfully unpacked and MD5 sums checked
package 'gridExtra' successfully unpacked and MD5 sums checked
package 'data.table' successfully unpacked and MD5 sums checked
package 'RCurl' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\leand\AppData\Local\Temp\RtmpywnbGe\downloaded_packages
── R CMD build ─────────────────────────────────────────────────────────────
WARNING: Rtools is required to build R packages, but is not currently installed.

Please download and install Rtools 4.4 from https://cran.r-project.org/bin/windows/Rtools/.
✓  checking for file 'C:\Users\leand\AppData\Local\Temp\RtmpywnbGe\remotes654418b8334e\jayjacobs-verisr-ddceeaf/DESCRIPTION'
(5.9s)
─  preparing 'verisr': (4s)
✓  checking DESCRIPTION meta-information ...
─  checking for LF line-endings in source and make files and shell scripts (699ms)
─  checking for empty or unneeded directories
   Removed empty directory 'verisr/README_files'
─  looking to see if a 'data/datalist' file should be added
─  building 'verisr_2.0.6.tar.gz'

Installing package into 'C:/Users/leand/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
* installing *source* package 'verisr' ...
** using staged installation
```

```
Leandro Delgado
Student ID:114416241
```

```
Data
incident                  List of  16
Values
  jsonfile                "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_
  new.pkg                 chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg                     chr [1:4] "devtools" "ggplot2" "scales" "rjson"
```

Name	Size	Modified
~$Counselling and Accessibility Services copy.pptx	165 B	Jul 18, 2023, 11:03 AM
~$enta bofa.docx	162 B	Apr 23, 2024, 12:50 PM
~$esen perfet tense.docx	162 B	Sep 11, 2022, 5:00 PM
~$torizacion maria r.docx	162 B	May 26, 2021, 3:53 PM
143-psexec-hunt		
143-psexec-hunt.zip	3 MB	Mar 2, 2025, 7:36 AM
BEJA MANUAL		
CHAPTER 7. FINAL PROJECT.pdf	17.4 MB	Mar 25, 2025, 12:27 AM
Club Adventurers _3c1aa9f345929f308878abf5410c7cd290e0561f463276574447f65703ee873e@group.c...	14.4 KB	Dec 23, 2024, 10:00 PM
COMMUNICATION ASSIGMENT-STAGE 2-3- THE FRONTLINE GROUP.docx	51.6 KB	Mar 24, 2025, 11:35 AM
Custom Office Templates		
CYT215-Project1- Leandro Delgado-114416241-WORKING.docx	6 MB	Mar 22, 2025, 6:55 AM
CYT215-Project1- Leandro Delgado-114416241-WORKING.pdf	3.3 MB	Mar 21, 2025, 9:19 PM
CYT215-Project1- Leandro Delgado-114416241.docx	4.4 MB	Mar 21, 2025, 6:19 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 1 PROJECT OPTION 1.pdf	331.7 KB	Feb 10, 2025, 5:27 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 2 PROJECT 1.pdf	217.2 KB	Feb 10, 2025, 5:28 PM
Default.rdp	0 B	Dec 16, 2024, 12:02 PM
Dell		
desktop.ini	418 B	Jan 6, 2025, 11:03 AM
DOCUMENTS		
Documents - Shortcut.lnk	767 B	Mar 21, 2025, 7:11 PM
Final Home Project_Winter2025 TATIANA FORMATIN.docx	1.3 MB	Mar 28, 2025, 1:41 PM

Figure 6. Result listing 7-5


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Console
Terminal
Background Jobs
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/

```
Installing packages into 'C:/Users/leand/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/bitops_1.0-9.zip'
Content type 'application/zip' length 32826 bytes (32 KB)
downloaded 32 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/gridExtra_2.3.zip'
Content type 'application/zip' length 1109639 bytes (1.1 MB)
downloaded 1.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/data.table_1.17.0.zip'
Content type 'application/zip' length 2910358 bytes (2.8 MB)
downloaded 2.8 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.4/RCurl_1.98-1.17.zip'
Content type 'application/zip' length 5465048 bytes (5.2 MB)
downloaded 5.2 MB

package 'bitops' successfully unpacked and MD5 sums checked
package 'gridExtra' successfully unpacked and MD5 sums checked
package 'data.table' successfully unpacked and MD5 sums checked
package 'RCurl' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\leand\AppData\Local\Temp\RtmpyWnb6e\downloaded_packages
── R CMD build ─────────────────────────────────────────────────────────────
WARNING: Rtools is required to build R packages, but is not currently installed.

Please download and install Rtools 4.4 from https://cran.r-project.org/bin/windows/Rtools/.
✓  checking for file 'C:\Users\leand\AppData\Local\Temp\RtmpyWnb6e\remotes654418b8334e\jayjacobs-verisr-ddceeaf/DESCRIPTION'
(5.9s)
─  preparing 'verisr': (4s)
✓  checking DESCRIPTION meta-information ...
─  checking for LF line-endings in source and make files and shell scripts (699ms)
─  checking for empty or unneeded directories
   Removed empty directory 'verisr/README_files'
─  looking to see if a 'data/datalist' file should be added
─  building 'verisr_2.0.6.tar.gz'

Installing package into 'C:/Users/leand/AppData/Local/R/win-library/4.4'
(as 'lib' is unspecified)
* installing *source* package 'verisr' ...
** using staged installation
** R
** data
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** testing if installed package can be loaded from temporary location
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (verisr)
> # and load up the library
> library(verisr)
> |
```

```
Leandro Delgado
Student ID:114416241
```

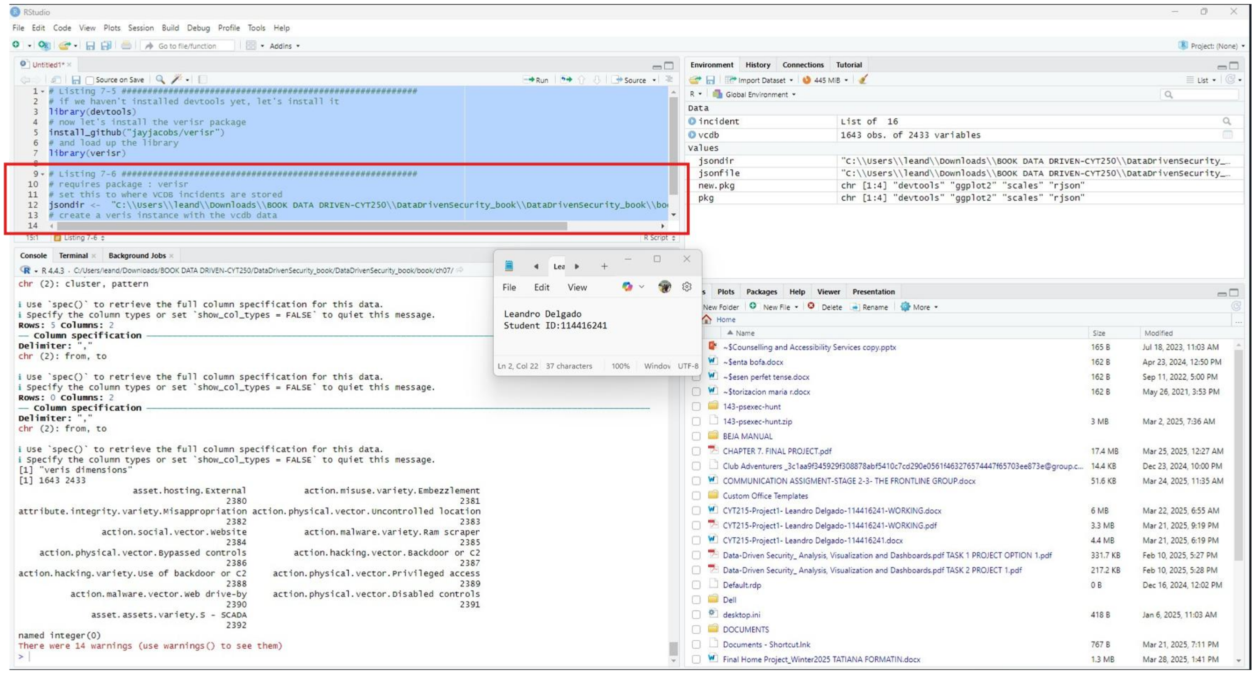
Data	
incident	List of 16
Values	
jsonfile	"C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_…
new.pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"

Name	Size	Modified
~$Counselling and Accessibility Services copy.pptx	165 B	Jul 18, 2023, 11:03 AM
~$enta bofa.docx	162 B	Apr 23, 2024, 12:50 PM
~$esen perfet tense.docx	162 B	Sep 11, 2022, 5:00 PM
~$torizacion maria r.docx	162 B	May 26, 2021, 3:53 PM
143-psexec-hunt		
143-psexec-hunt.zip	3 MB	Mar 2, 2025, 7:36 AM
BEJA MANUAL		
CHAPTER 7. FINAL PROJECT.pdf	17.4 MB	Mar 25, 2025, 12:27 AM
Club Adventurers _3c1aa9f345929f308878abf5410c7cd290e0561f463276574447f65703ee873e@group.c...	14.4 KB	Dec 23, 2024, 10:00 PM
COMMUNICATION ASSIGMENT-STAGE 2-3- THE FRONTLINE GROUP.docx	51.6 KB	Mar 24, 2025, 11:35 AM
Custom Office Templates		
CYT215-Project1- Leandro Delgado-114416241-WORKING.docx	6 MB	Mar 22, 2025, 6:55 AM
CYT215-Project1- Leandro Delgado-114416241-WORKING.pdf	3.3 MB	Mar 21, 2025, 9:19 PM
CYT215-Project1- Leandro Delgado-114416241.docx	4.4 MB	Mar 21, 2025, 6:19 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 1 PROJECT OPTION 1.pdf	331.7 KB	Feb 10, 2025, 5:27 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 2 PROJECT 1.pdf	217.2 KB	Feb 10, 2025, 5:28 PM
Default.rdp	0 B	Dec 16, 2024, 12:02 PM
Dell		
desktop.ini	418 B	Jan 6, 2025, 11:03 AM
DOCUMENTS		
Documents - Shortcut.lnk	767 B	Mar 21, 2025, 7:11 PM
Final Home Project_Winter2025 TATIANA FORMATIN.docx	1.3 MB	Mar 28, 2025, 1:41 PM

Figure 7. Result Listing 7-5

Listing 7-6

In this step, I used the veris package to read in a collection of VCDB incident files from a directory. The data was loaded into a vcdb object, which contains 1,643 records and over 2,400 variables. This gives me a structured dataset ready for analysis, with detailed information on cyber incidents.



```
# Listing 7-5 ################################################################
# if we haven't installed devtools yet, let's install it
library(devtools)
# now let's install the verisr package
install_github("jayjacobs/verisr")
# and load up the library
library(verisr)

# Listing 7-6 ################################################################
# requires package : verisr
# set this to where VCDB incidents are stored
jsondir <-  "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book\\DataDrivenSecurity_book\\bo
# create a veris instance with the vcdb data
```

Console
Terminal
Background Jobs
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/

```
chr (2): cluster, pattern

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
Rows: 5 Columns: 2
── Column specification ────────────────────────────────────────────────────
Delimiter: ","
chr (2): from, to

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
Rows: 0 Columns: 2
── Column specification ────────────────────────────────────────────────────
Delimiter: ","
chr (2): from, to

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
[1] "veris dimensions"
[1] 1643 2433
                     asset.hosting.External          action.misuse.variety.Embezzlement
                                       2380                                        2381
attribute.integrity.variety.Misappropriation action.physical.vector.Uncontrolled location
                                       2382                                        2383
                 action.social.vector.Website          action.malware.variety.Ram scraper
                                       2384                                        2385
        action.physical.vector.Bypassed controls    action.hacking.vector.Backdoor or C2
                                       2386                                        2387
action.hacking.variety.Use of backdoor or C2      action.physical.vector.Privileged access
                                       2388                                        2389
           action.malware.vector.Web drive-by      action.physical.vector.Disabled controls
                                       2390                                        2391
                asset.assets.variety.S - SCADA
                                       2392
named integer(0)
There were 14 warnings (use warnings() to see them)
> |
```

```
Leandro Delgado
Student ID:114416241
```

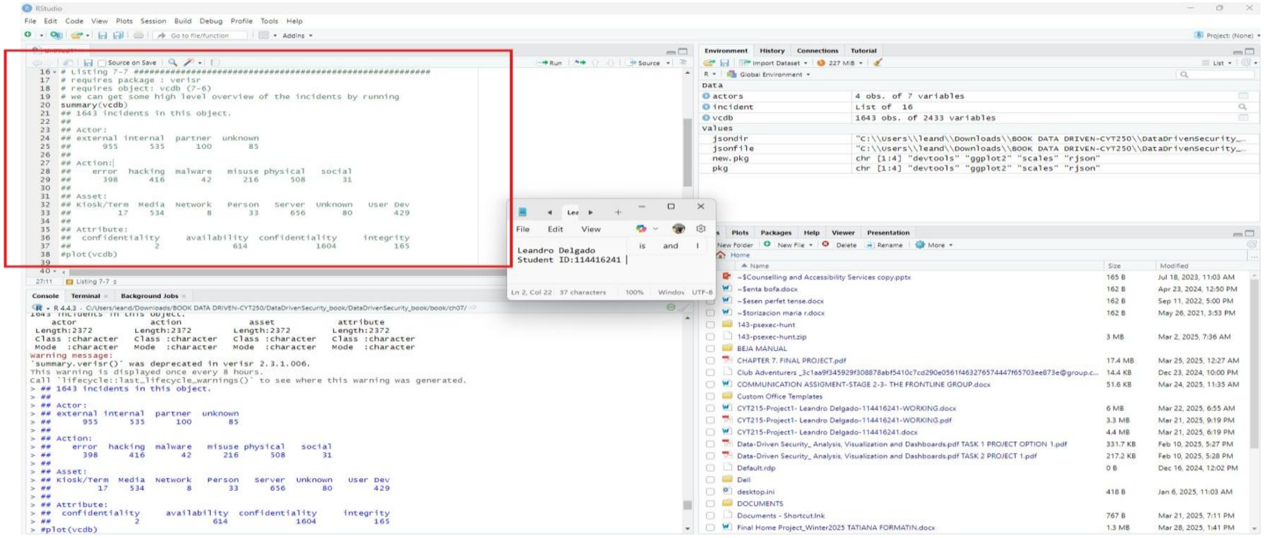
Data	
incident	List of 16
vcdb	1643 obs. of 2433 variables
Values	
jsondir	"C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_…
jsonfile	"C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_…
new.pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"

Name	Size	Modified
~$Counselling and Accessibility Services copy.pptx	165 B	Jul 18, 2023, 11:03 AM
~$enta bofa.docx	162 B	Apr 23, 2024, 12:50 PM
~$esen perfet tense.docx	162 B	Sep 11, 2022, 5:00 PM
~$torizacion maria r.docx	162 B	May 26, 2021, 3:53 PM
143-psexec-hunt		
143-psexec-hunt.zip	3 MB	Mar 2, 2025, 7:36 AM
BEJA MANUAL		
CHAPTER 7. FINAL PROJECT.pdf	17.4 MB	Mar 25, 2025, 12:27 AM
Club Adventurers _3c1aa9f345929f308878abf5410c7cd290e0561f463276574447f65703ee873e@group.c...	14.4 KB	Dec 23, 2024, 10:00 PM
COMMUNICATION ASSIGMENT-STAGE 2-3- THE FRONTLINE GROUP.docx	51.6 KB	Mar 24, 2025, 11:35 AM
Custom Office Templates		
CYT215-Project1- Leandro Delgado-114416241-WORKING.docx	6 MB	Mar 22, 2025, 6:55 AM
CYT215-Project1- Leandro Delgado-114416241-WORKING.pdf	3.3 MB	Mar 21, 2025, 9:19 PM
CYT215-Project1- Leandro Delgado-114416241.docx	4.4 MB	Mar 21, 2025, 6:19 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 1 PROJECT OPTION 1.pdf	331.7 KB	Feb 10, 2025, 5:27 PM
Data-Driven Security_ Analysis, Visualization and Dashboards.pdf TASK 2 PROJECT 1.pdf	217.2 KB	Feb 10, 2025, 5:28 PM
Default.rdp	0 B	Dec 16, 2024, 12:02 PM
Dell		
desktop.ini	418 B	Jan 6, 2025, 11:03 AM
DOCUMENTS		
Documents - Shortcut.lnk	767 B	Mar 21, 2025, 7:11 PM
Final Home Project_Winter2025 TATIANA FORMATIN.docx	1.3 MB	Mar 28, 2025, 1:41 PM

Figure 8. Result listing 7-6

Listing 7-7

In this step, I used the summary() function on the vcdb object to get an overview of the dataset. It returned useful insights, such as the number of incidents involving external actors (955), and the most common impact being on confidentiality (1,604). Although the message confirms that the dataset contains 1,643 incidents, the actual records are not shown—only a high-level summary is provided. To explore individual incidents, I would need to use functions like head(vcdb) or View(vcdb).



```
# Listing 7-7 ##################################################################
# requires package : verisr
# requires object: vcdb (7-6)
# we can get some high level overview of the incidents by running
summary(vcdb)
## 1643 incidents in this object.
##
## Actor:
## external internal  partner  unknown
##      955      535      100       85
##
## Action:
##    error  hacking  malware   misuse physical   social
##      398      416       42      216      508       31
##
## Asset:
## Kiosk/Term  Media  Network   Person   Server  Unknown   User Dev
##         17    534        8       33      656       80        429
##
## Attribute:
##  confidentiality     availability  confidentiality     integrity
##                2              614             1604           165
#plot(vcdb)
```

```
1643 incidents in this object.
    actor              action              asset               attribute
 Length:2372        Length:2372        Length:2372        Length:2372
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character
Warning message:
`summary.verisr()` was deprecated in verisr 2.3.1.006.
This warning is displayed once every 8 hours.
Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
> ## 1643 incidents in this object.
> ##
> ## Actor:
> ## external internal  partner  unknown
> ##      955      535      100       85
> ##
> ## Action:
> ##    error  hacking  malware   misuse physical   social
> ##      398      416       42      216      508       31
> ##
> ## Asset:
> ## Kiosk/Term  Media  Network   Person   Server  Unknown   User Dev
> ##         17    534        8       33      656       80        429
> ##
> ## Attribute:
> ##  confidentiality     availability  confidentiality     integrity
> ##                2              614             1604           165
> #plot(vcdb)
```

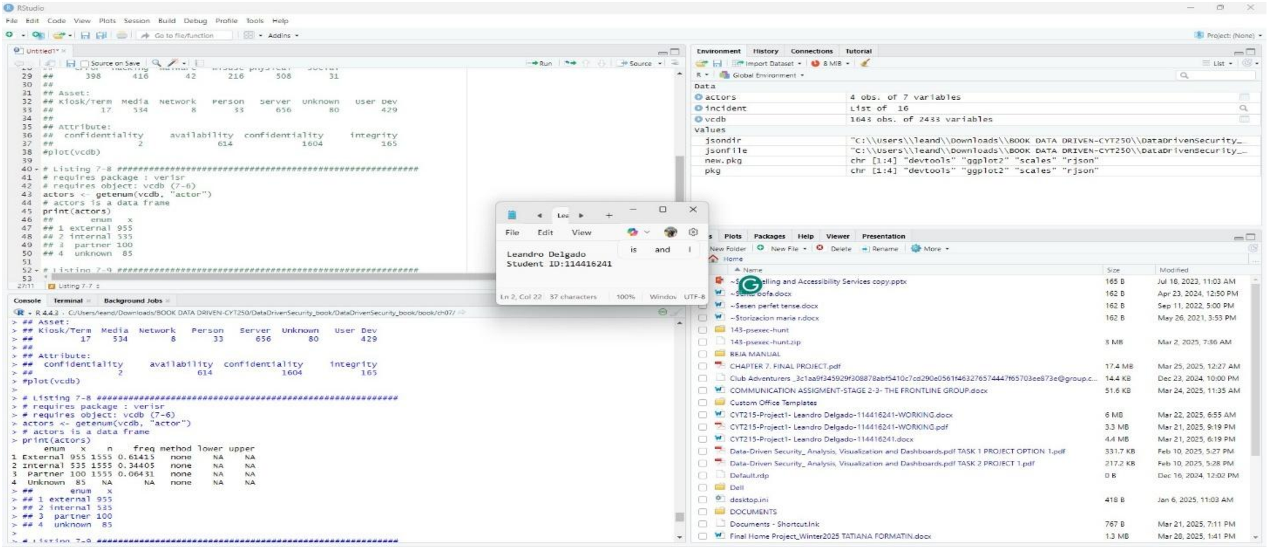
Data	
actors	4 obs. of 7 variables
incident	List of 16
vcdb	1643 obs. of 2433 variables
Values	
jsondir	"C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
jsonfile	"C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
new.pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg	chr [1:4] "devtools" "ggplot2" "scales" "rjson"

Leandro Delgado
Student ID:114416241

Figure 9.Result Listing 7-7

Listing 7-8

This step uses the getenum() function to extract and count the different types of actors involved in the VCDB incidents. The results show that most incidents were caused by external actors (955), followed by internal (535), partners (100), and unknown sources (85). The output is stored as a data frame named actors, making it easy to view and use in further analysis.



```
# Listing 7-8 ##################################################################
# requires package : verisr
# requires object: vcdb (7-6)
actors <- getenum(vcdb, "actor")
# actors is a data frame
print(actors)
##        enum   x
## 1 external 955
## 2 internal 535
## 3  partner 100
## 4  unknown  85
```

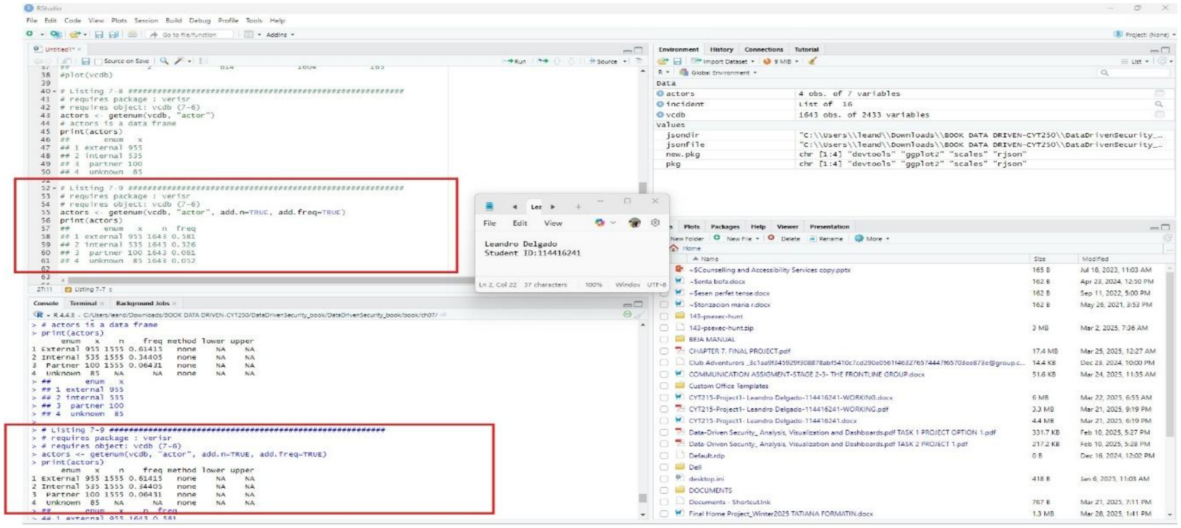
```
> # Listing 7-8 ##################################################################
> # requires package : verisr
> # requires object: vcdb (7-6)
> actors <- getenum(vcdb, "actor")
> # actors is a data frame
> print(actors)
      enum   x    n    freq method lower upper
1 External 955 1555 0.61415   none    NA    NA
2 Internal 535 1555 0.34405   none    NA    NA
3  Partner 100 1555 0.06431   none    NA    NA
4  Unknown  85   NA      NA   none    NA    NA
> ##        enum   x
> ## 1 external 955
> ## 2 internal 535
> ## 3  partner 100
> ## 4  unknown  85
```

Leandro Delgado
Student ID:114416241

Figure 10. Result 7-8

Listing 7-9

In this step, I ran an updated version of the previous command, this time adding arguments to include both the total count of incidents (n) and the frequency (freq) for each actor category. The results show that external actors were responsible for over 58% of the incidents, followed by internal actors at 32%. This gave me a clearer picture of the relative impact of each actor type within the dataset.



```
#plot(vcdb)

# Listing 7-8 ################################################################
# requires package : verisr
# requires object: vcdb (7-6)
actors <- getenum(vcdb, "actor")
# actors is a data frame
print(actors)
##        enum   x
## 1 external 955
## 2 internal 535
## 3  partner 100
## 4  unknown  85

# Listing 7-9 ################################################################
# requires package : verisr
# requires object: vcdb (7-6)
actors <- getenum(vcdb, "actor", add.n=TRUE, add.freq=TRUE)
print(actors)
##        enum   x    n  freq
## 1 external 955 1643 0.581
## 2 internal 535 1643 0.326
## 3  partner 100 1643 0.061
## 4  unknown  85 1643 0.052
```

```
> # actors is a data frame
> print(actors)
      enum   x    n    freq method lower upper
1 External 955 1555 0.61415   none    NA    NA
2 Internal 535 1555 0.34405   none    NA    NA
3  Partner 100 1555 0.06431   none    NA    NA
4  Unknown  85   NA      NA   none    NA    NA
> ##        enum   x
> ## 1 external 955
> ## 2 internal 535
> ## 3  partner 100
> ## 4  unknown  85
>
> # Listing 7-9 ##############################################################
> # requires package : verisr
> # requires object: vcdb (7-6)
> actors <- getenum(vcdb, "actor", add.n=TRUE, add.freq=TRUE)
> print(actors)
      enum   x    n    freq method lower upper
1 External 955 1555 0.61415   none    NA    NA
2 Internal 535 1555 0.34405   none    NA    NA
3  Partner 100 1555 0.06431   none    NA    NA
4  Unknown  85   NA      NA   none    NA    NA
> ##        enum   x    n  freq
> ## 1 external 955 1643 0.581
```

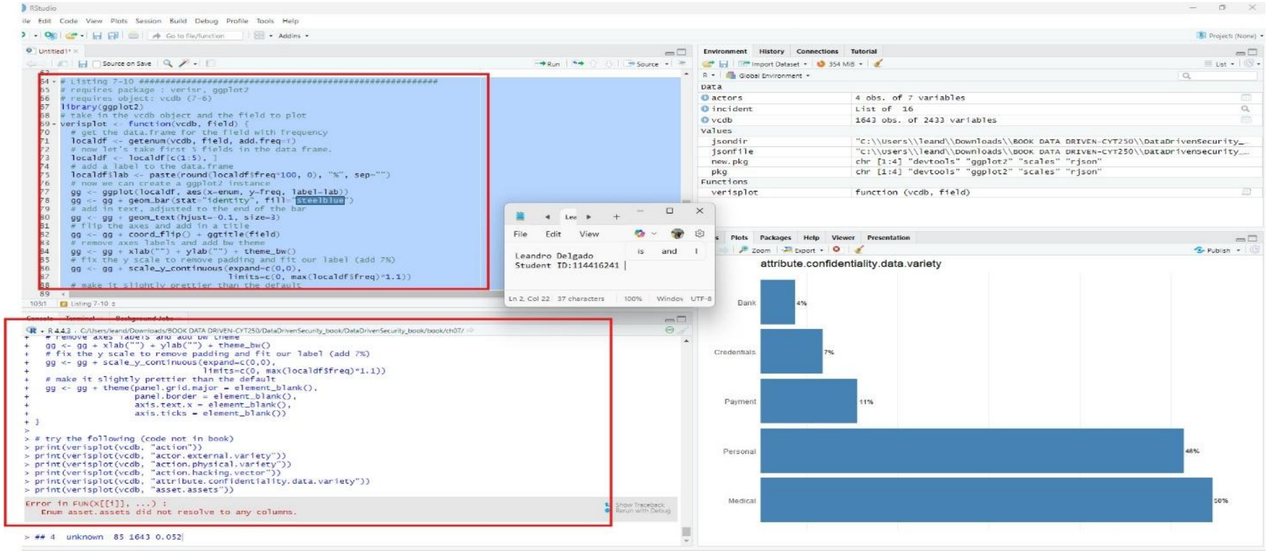
```
Data
actors      4 obs. of 7 variables
incident    List of 16
vcdb        1643 obs. of 2433 variables
Values
jsondir     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
jsonfile    "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
new.pkg     chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg         chr [1:4] "devtools" "ggplot2" "scales" "rjson"
```

Leandro Delgado
Student ID:114416241

Figure 11. Result listing 7-9

Listing 7-10

In this step, I created a custom function called verisplot() to generate bar charts from specific VCDB fields using ggplot2. I used it to visualize the different types of data affected under the attribute confidentiality.data.variety field. The chart showed that medical and personal data were the most compromised. This helped me understand how to combine data aggregation and visualization in R for clearer insights.



```
# Listing 7-10 ###############################################################
# requires package : verisr, ggplot2
# requires object: vcdb (7-6)
library(ggplot2)
# take in the vcdb object and the field to plot
verisplot <- function(vcdb, field) {
  # get the data.frame for the field with frequency
  localdf <- getenum(vcdb, field, add.freq=T)
  # now let's take first 5 fields in the data frame.
  localdf <- localdf[c(1:5), ]
  # add a label to the data.frame
  localdf$lab <- paste(round(localdf$freq*100, 0), "%", sep="")
  # now we can create a ggplot2 instance
  gg <- ggplot(localdf, aes(x=enum, y=freq, label=lab))
  gg <- gg + geom_bar(stat="identity", fill="steelblue")
  # add in text, adjusted to the end of the bar
  gg <- gg + geom_text(hjust=-0.1, size=3)
  # flip the axes and add in a title
  gg <- gg + coord_flip() + ggtitle(field)
  # remove axes labels and add bw theme
  gg <- gg + xlab("") + ylab("") + theme_bw()
  # fix the y scale to remove padding and fit our label (add 7%)
  gg <- gg + scale_y_continuous(expand=c(0,0),
                                limits=c(0, max(localdf$freq)*1.1))
  # make it slightly prettier than the default
```

```
+   # remove axes labels and add bw theme
+   gg <- gg + xlab("") + ylab("") + theme_bw()
+   # fix the y scale to remove padding and fit our label (add 7%)
+   gg <- gg + scale_y_continuous(expand=c(0,0),
+                                 limits=c(0, max(localdf$freq)*1.1))
+   # make it slightly prettier than the default
+   gg <- gg + theme(panel.grid.major = element_blank(),
+                    panel.border = element_blank(),
+                    axis.text.x = element_blank(),
+                    axis.ticks = element_blank())
+ }
>
> # try the following (code not in book)
> print(verisplot(vcdb, "action"))
> print(verisplot(vcdb, "actor.external.variety"))
> print(verisplot(vcdb, "action.physical.variety"))
> print(verisplot(vcdb, "action.hacking.vector"))
> print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
> print(verisplot(vcdb, "asset.assets"))
Error in FUN(X[[i]], ...) :
  Enum asset.assets did not resolve to any columns.
> ## 4  unknown  85 1643 0.052
```

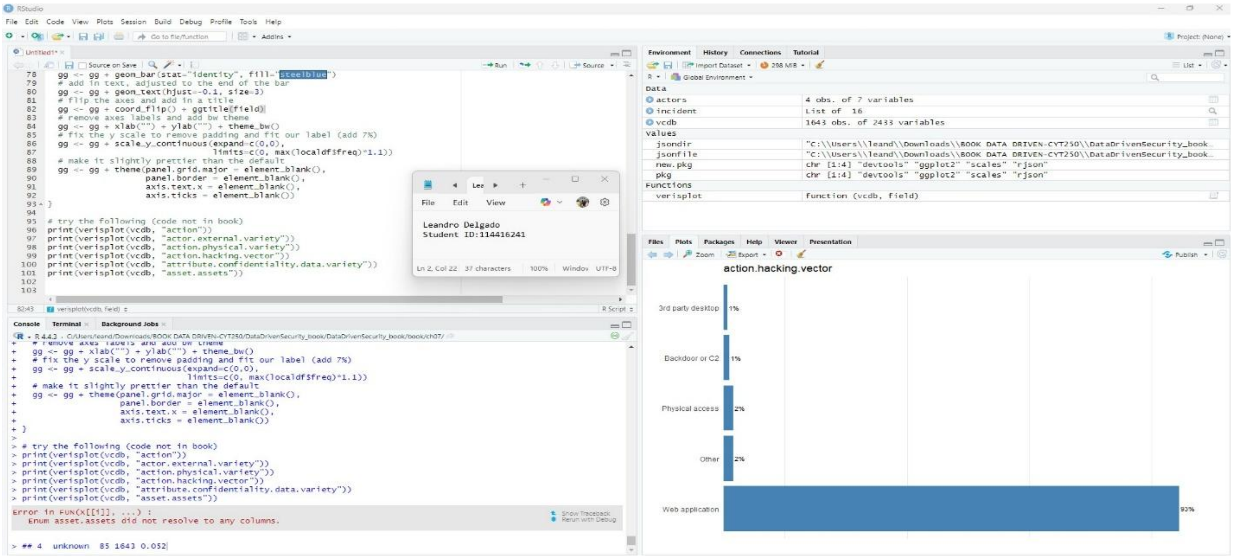
```
Data
actors      4 obs. of 7 variables
incident    List of 16
vcdb        1643 obs. of 2433 variables
Values
jsondir     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
jsonfile    "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_...
new.pkg     chr [1:4] "devtools" "ggplot2" "scales" "rjson"
pkg         chr [1:4] "devtools" "ggplot2" "scales" "rjson"
Functions
verisplot   function (vcdb, field)
```

Leandro Delgado
Student ID:114416241

Figure 12. Result listing 7-10

Listing 7-10

After defining the verisplot() function, I used it to generate several additional charts by specifying different fields from the VCDB dataset. One of the most interesting results came from the action.hacking.vector field, where 93% of the incidents involved web applications, highlighting a major attack surface. I also attempted to visualize asset. assets, but received an error indicating that this field could not be resolved. This likely means the column doesn't exist in the current version of the dataset or is named differently. It was a helpful reminder to verify field names before plotting.



```
  gg <- gg + geom_bar(stat="identity", fill="steelblue")
  # add in text, adjusted to the end of the bar
  gg <- gg + geom_text(hjust=-0.1, size=3)
  # flip the axes and add in a title
  gg <- gg + coord_flip() + ggtitle(field)
  # remove axes labels and add bw theme
  gg <- gg + xlab("") + ylab("") + theme_bw()
  # fix the y scale to remove padding and fit our label (add 7%)
  gg <- gg + scale_y_continuous(expand=c(0,0),
                                limits=c(0, max(localdf$freq)*1.1))
  # make it slightly prettier than the default
  gg <- gg + theme(panel.grid.major = element_blank(),
                   panel.border = element_blank(),
                   axis.text.x = element_blank(),
                   axis.ticks = element_blank())
}

# try the following (code not in book)
print(verisplot(vcdb, "action"))
print(verisplot(vcdb, "actor.external.variety"))
print(verisplot(vcdb, "action.physical.variety"))
print(verisplot(vcdb, "action.hacking.vector"))
print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
print(verisplot(vcdb, "asset.assets"))
```

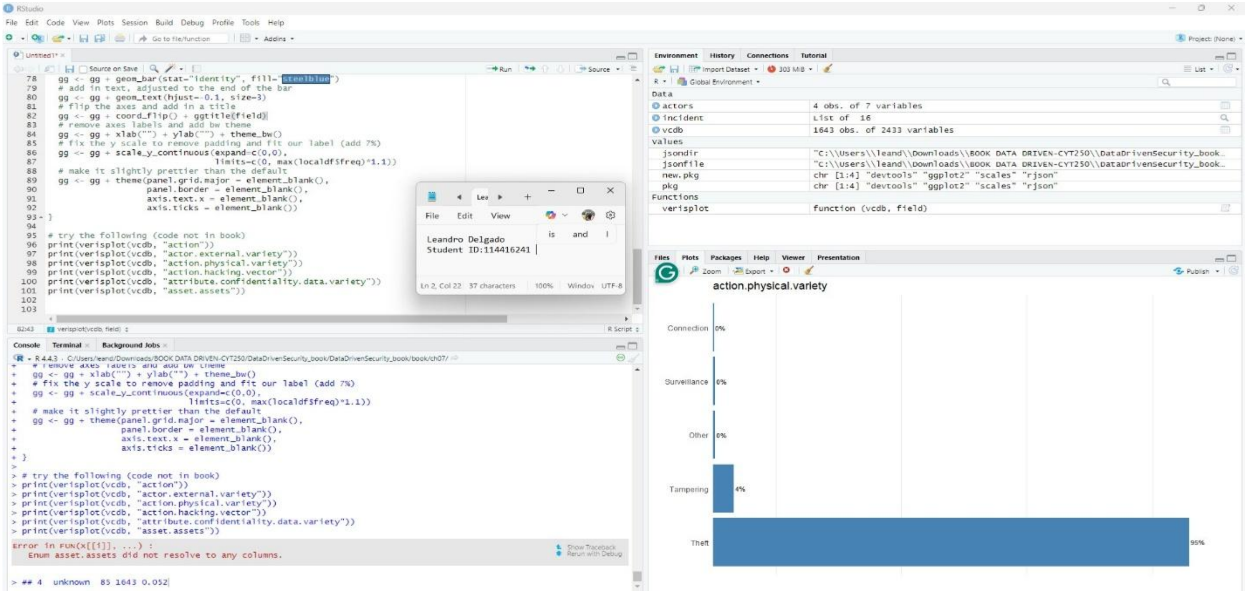
```
Leandro Delgado
Student ID:114416241
```

```
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
+   # remove axes labels and add bw theme
+   gg <- gg + xlab("") + ylab("") + theme_bw()
+   # fix the y scale to remove padding and fit our label (add 7%)
+   gg <- gg + scale_y_continuous(expand=c(0,0),
+                                 limits=c(0, max(localdf$freq)*1.1))
+   # make it slightly prettier than the default
+   gg <- gg + theme(panel.grid.major = element_blank(),
+                    panel.border = element_blank(),
+                    axis.text.x = element_blank(),
+                    axis.ticks = element_blank())
+ }
>
> # try the following (code not in book)
> print(verisplot(vcdb, "action"))
> print(verisplot(vcdb, "actor.external.variety"))
> print(verisplot(vcdb, "action.physical.variety"))
> print(verisplot(vcdb, "action.hacking.vector"))
> print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
> print(verisplot(vcdb, "asset.assets"))
Error in FUN(X[[i]], ...) :
  Enum asset.assets did not resolve to any columns.

> ## 4   unknown   85 1643 0.052
```

```
Data
  actors       4 obs. of 7 variables
  incident     List of  16
  vcdb         1643 obs. of 2433 variables
Values
  jsondir      "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  jsonfile     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  new.pkg      chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg          chr [1:4] "devtools" "ggplot2" "scales" "rjson"
Functions
  verisplot    function (vcdb, field)
```

Figure 13. Result run listing 7-10

Using the verisplot() function, I explored the action.physical.variety field. The resulting bar chart shows that theft was the most frequent physical action, accounting for 95% of all such incidents. This highlights how physical security threats like stolen devices remain a major concern in cybersecurity incidents.



```
  gg <- gg + geom_bar(stat="identity", fill="steelblue")
  # add in text, adjusted to the end of the bar
  gg <- gg + geom_text(hjust=-0.1, size=3)
  # flip the axes and add in a title
  gg <- gg + coord_flip() + ggtitle(field)
  # remove axes labels and add bw theme
  gg <- gg + xlab("") + ylab("") + theme_bw()
  # fix the y scale to remove padding and fit our label (add 7%)
  gg <- gg + scale_y_continuous(expand=c(0,0),
                                limits=c(0, max(localdf$freq)*1.1))
  # make it slightly prettier than the default
  gg <- gg + theme(panel.grid.major = element_blank(),
                   panel.border = element_blank(),
                   axis.text.x = element_blank(),
                   axis.ticks = element_blank())
}

# try the following (code not in book)
print(verisplot(vcdb, "action"))
print(verisplot(vcdb, "actor.external.variety"))
print(verisplot(vcdb, "action.physical.variety"))
print(verisplot(vcdb, "action.hacking.vector"))
print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
print(verisplot(vcdb, "asset.assets"))
```

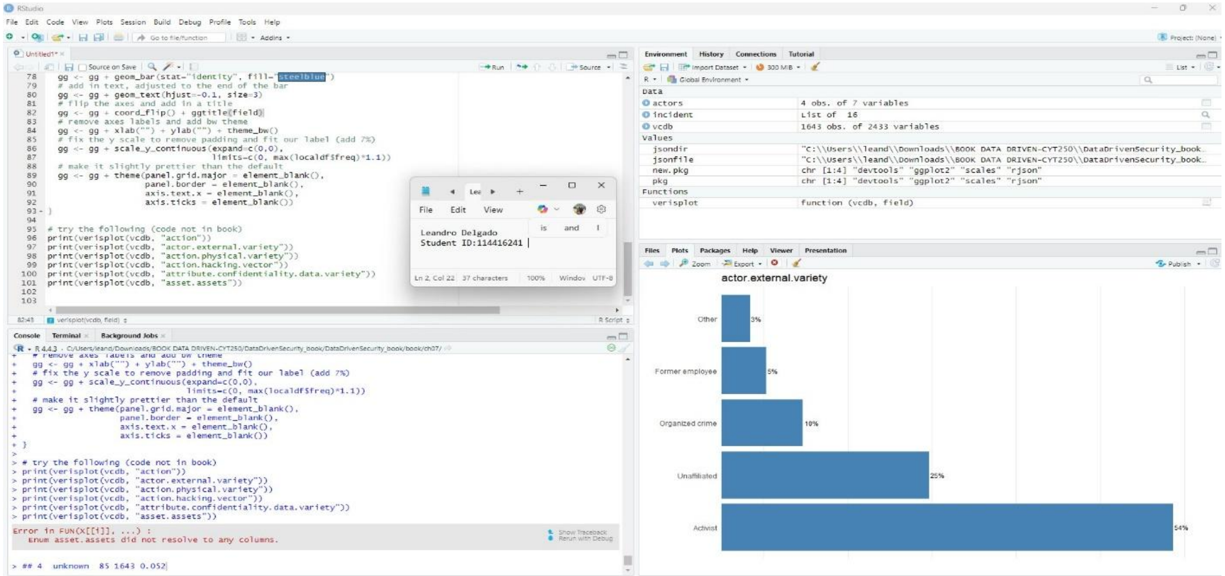
```
Leandro Delgado
Student ID:114416241
```

```
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
+   # remove axes labels and add bw theme
+   gg <- gg + xlab("") + ylab("") + theme_bw()
+   # fix the y scale to remove padding and fit our label (add 7%)
+   gg <- gg + scale_y_continuous(expand=c(0,0),
+                                 limits=c(0, max(localdf$freq)*1.1))
+   # make it slightly prettier than the default
+   gg <- gg + theme(panel.grid.major = element_blank(),
+                    panel.border = element_blank(),
+                    axis.text.x = element_blank(),
+                    axis.ticks = element_blank())
+ }
>
> # try the following (code not in book)
> print(verisplot(vcdb, "action"))
> print(verisplot(vcdb, "actor.external.variety"))
> print(verisplot(vcdb, "action.physical.variety"))
> print(verisplot(vcdb, "action.hacking.vector"))
> print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
> print(verisplot(vcdb, "asset.assets"))
Error in FUN(X[[i]], ...) :
  Enum asset.assets did not resolve to any columns.

> ## 4   unknown   85 1643 0.052
```

```
Data
  actors       4 obs. of 7 variables
  incident     List of  16
  vcdb         1643 obs. of 2433 variables
Values
  jsondir      "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  jsonfile     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  new.pkg      chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg          chr [1:4] "devtools" "ggplot2" "scales" "rjson"
Functions
  verisplot    function (vcdb, field)
```

Figure 14. Run listing 7-10

Using the verisplot() function on the actor.external.variety field, I was able to see which types of external actors were most involved in security incidents. The chart revealed that activists were responsible for the majority (54%) of incidents, followed by unaffiliated individuals (25%) and organized crime (10%). This helped me better understand the diversity of external threats represented in the dataset.



```
  gg <- gg + geom_bar(stat="identity", fill="steelblue")
  # add in text, adjusted to the end of the bar
  gg <- gg + geom_text(hjust=-0.1, size=3)
  # flip the axes and add in a title
  gg <- gg + coord_flip() + ggtitle(field)
  # remove axes labels and add bw theme
  gg <- gg + xlab("") + ylab("") + theme_bw()
  # fix the y scale to remove padding and fit our label (add 7%)
  gg <- gg + scale_y_continuous(expand=c(0,0),
                                limits=c(0, max(localdf$freq)*1.1))
  # make it slightly prettier than the default
  gg <- gg + theme(panel.grid.major = element_blank(),
                   panel.border = element_blank(),
                   axis.text.x = element_blank(),
                   axis.ticks = element_blank())
}

# try the following (code not in book)
print(verisplot(vcdb, "action"))
print(verisplot(vcdb, "actor.external.variety"))
print(verisplot(vcdb, "action.physical.variety"))
print(verisplot(vcdb, "action.hacking.vector"))
print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
print(verisplot(vcdb, "asset.assets"))
```

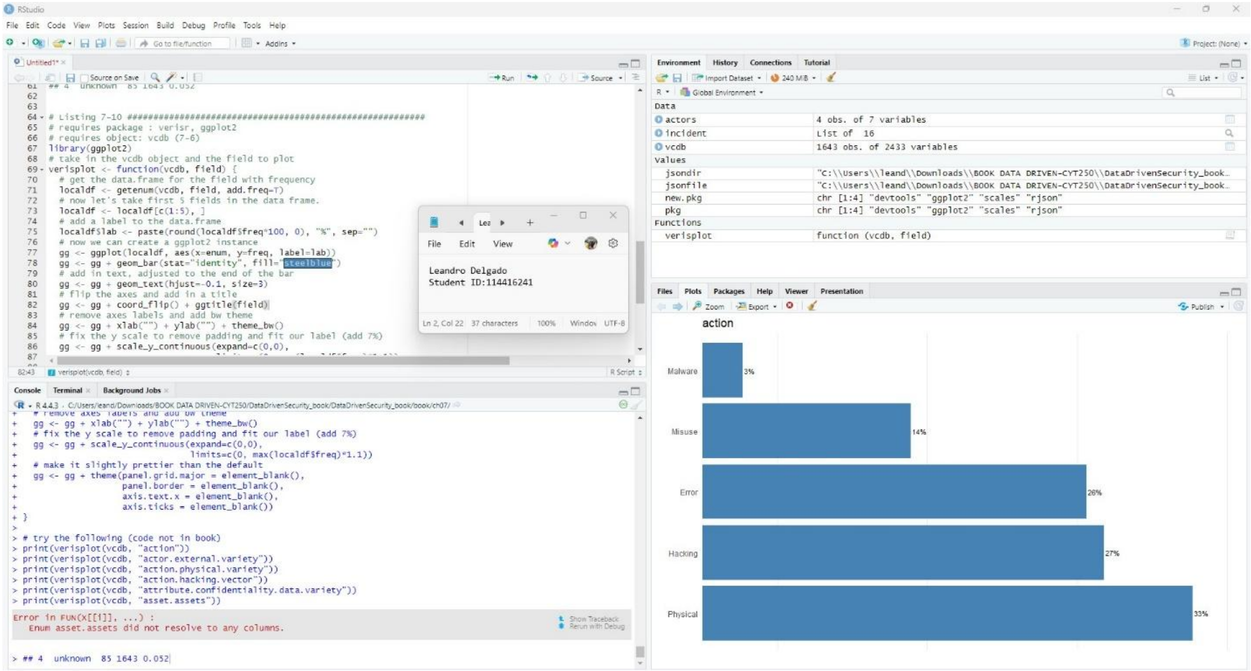
```
Leandro Delgado
Student ID:114416241
```

```
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
+   # remove axes labels and add bw theme
+   gg <- gg + xlab("") + ylab("") + theme_bw()
+   # fix the y scale to remove padding and fit our label (add 7%)
+   gg <- gg + scale_y_continuous(expand=c(0,0),
+                                 limits=c(0, max(localdf$freq)*1.1))
+   # make it slightly prettier than the default
+   gg <- gg + theme(panel.grid.major = element_blank(),
+                    panel.border = element_blank(),
+                    axis.text.x = element_blank(),
+                    axis.ticks = element_blank())
+ }
>
> # try the following (code not in book)
> print(verisplot(vcdb, "action"))
> print(verisplot(vcdb, "actor.external.variety"))
> print(verisplot(vcdb, "action.physical.variety"))
> print(verisplot(vcdb, "action.hacking.vector"))
> print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
> print(verisplot(vcdb, "asset.assets"))
Error in FUN(X[[i]], ...) :
  Enum asset.assets did not resolve to any columns.

> ## 4  unknown  85 1643 0.052
```

```
Data
actors        4 obs. of 7 variables
incident      List of  16
vcdb          1643 obs. of 2433 variables
Values
  jsondir     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  jsonfile    "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  new.pkg     chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg         chr [1:4] "devtools" "ggplot2" "scales" "rjson"
Functions
  verisplot   function (vcdb, field)
```

Figure 15.Result Listing 7-10

Using the verisplot() function on the action field, I generated a bar chart showing the types of actions involved in the incidents. The data revealed that physical actions (33%), hacking (27%), and errors (26%) were the most common, followed by misuse and malware. This helped me understand the broader categories of how breaches occurred across the dataset.



```
## 4  unknown  85 1643 0.052


# Listing 7-10 ##############################################################
# requires package : verisr, ggplot2
# requires object: vcdb (7-6)
library(ggplot2)
# take in the vcdb object and the field to plot
verisplot <- function(vcdb, field) {
  # get the data.frame for the field with frequency
  localdf <- getenum(vcdb, field, add.freq=T)
  # now let's take first 5 fields in the data frame.
  localdf <- localdf[c(1:5), ]
  # add a label to the data.frame
  localdf$lab <- paste(round(localdf$freq*100, 0), "%", sep="")
  # now we can create a ggplot2 instance
  gg <- ggplot(localdf, aes(x=enum, y=freq, label=lab))
  gg <- gg + geom_bar(stat="identity", fill="steelblue")
  # add in text, adjusted to the end of the bar
  gg <- gg + geom_text(hjust=-0.1, size=3)
  # flip the axes and add in a title
  gg <- gg + coord_flip() + ggtitle(field)
  # remove axes labels and add bw theme
  gg <- gg + xlab("") + ylab("") + theme_bw()
  # fix the y scale to remove padding and fit our label (add 7%)
  gg <- gg + scale_y_continuous(expand=c(0,0),
```

```
Leandro Delgado
Student ID:114416241
```

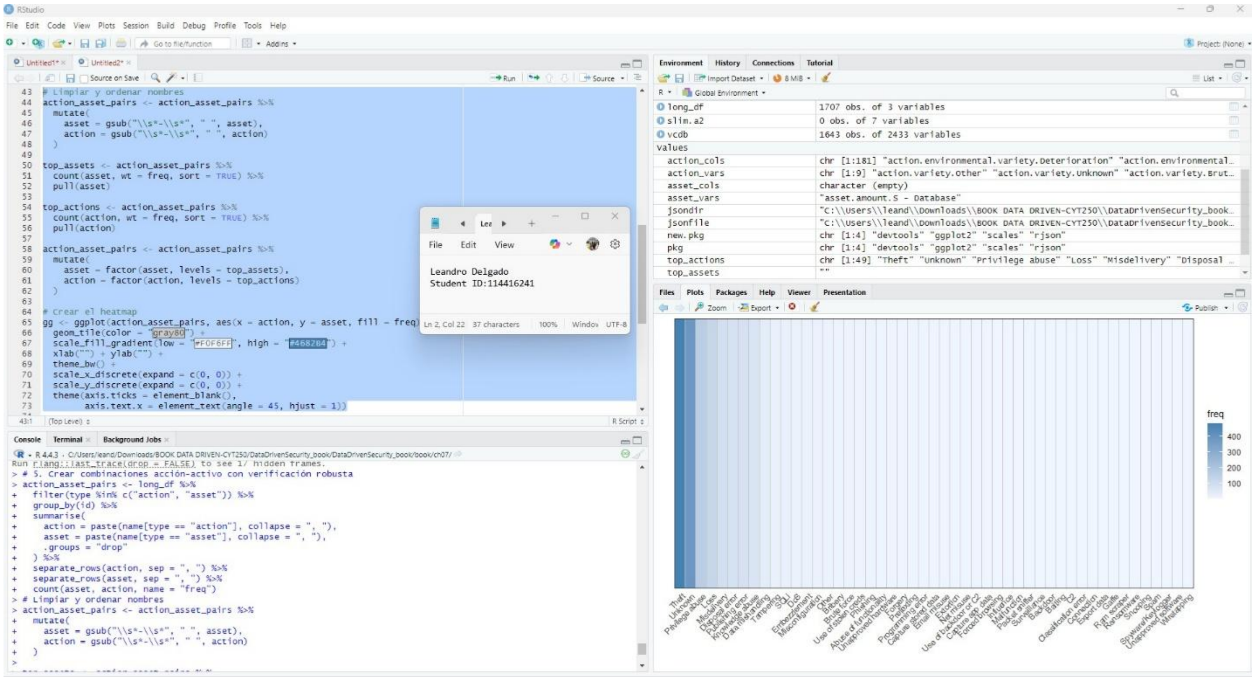
```
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
+   # remove axes labels and add bw theme
+   gg <- gg + xlab("") + ylab("") + theme_bw()
+   # fix the y scale to remove padding and fit our label (add 7%)
+   gg <- gg + scale_y_continuous(expand=c(0,0),
+                                 limits=c(0, max(localdf$freq)*1.1))
+   # make it slightly prettier than the default
+   gg <- gg + theme(panel.grid.major = element_blank(),
+                    panel.border = element_blank(),
+                    axis.text.x = element_blank(),
+                    axis.ticks = element_blank())
+ }
>
> # try the following (code not in book)
> print(verisplot(vcdb, "action"))
> print(verisplot(vcdb, "actor.external.variety"))
> print(verisplot(vcdb, "action.physical.variety"))
> print(verisplot(vcdb, "action.hacking.vector"))
> print(verisplot(vcdb, "attribute.confidentiality.data.variety"))
> print(verisplot(vcdb, "asset.assets"))
Error in FUN(X[[i]], ...) :
  Enum asset.assets did not resolve to any columns.

> ## 4  unknown  85 1643 0.052
```

```
Data
actors        4 obs. of 7 variables
incident      List of  16
vcdb          1643 obs. of 2433 variables
Values
  jsondir     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  jsonfile    "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  new.pkg     chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg         chr [1:4] "devtools" "ggplot2" "scales" "rjson"
Functions
  verisplot   function (vcdb, field)
```

Figure 16.Result Listing 7-10

Listing 7-11

In this step, I created a heatmap to explore the relationship between different types of actions and the assets they impacted. I grouped the data, counted the most frequent combinations, and used ggplot2 to visualize it. The darker the color, the more frequently that action-asset pair appeared. This helped me quickly identify which actions were most often associated with specific types of compromised assets across the dataset.



```
# Limpiar y ordenar nombres
action_asset_pairs <- action_asset_pairs %>%
  mutate(
    asset = gsub("\\s*-\\s*", " ", asset),
    action = gsub("\\s*-\\s*", " ", action)
  )

top_assets <- action_asset_pairs %>%
  count(asset, wt = freq, sort = TRUE) %>%
  pull(asset)

top_actions <- action_asset_pairs %>%
  count(action, wt = freq, sort = TRUE) %>%
  pull(action)

action_asset_pairs <- action_asset_pairs %>%
  mutate(
    asset = factor(asset, levels = top_assets),
    action = factor(action, levels = top_actions)
  )

# Crear el heatmap
gg <- ggplot(action_asset_pairs, aes(x = action, y = asset, fill = freq)
  geom_tile(color = "gray80") +
  scale_fill_gradient(low = "#F0F6FF", high = "#468284") +
  xlab("") + ylab("") +
  theme_bw() +
  scale_x_discrete(expand = c(0, 0)) +
  scale_y_discrete(expand = c(0, 0)) +
  theme(axis.ticks = element_blank(),
        axis.text.x = element_text(angle = 45, hjust = 1))
```

Leandro Delgado
Student ID:114416241

```
R 4.4.3 · C:/Users/leand/Downloads/BOOK DATA DRIVEN-CYT250/DataDrivenSecurity_book/DataDrivenSecurity_book/book/ch07/
Run rlang::last_trace(drop = FALSE) to see 17 hidden frames.
> # 5. Crear combinaciones acción-activo con verificación robusta
> action_asset_pairs <- long_df %>%
+   filter(type %in% c("action", "asset")) %>%
+   group_by(id) %>%
+   summarise(
+     action = paste(name[type == "action"], collapse = ", "),
+     asset = paste(name[type == "asset"], collapse = ", "),
+     .groups = "drop"
+   ) %>%
+   separate_rows(action, sep = ", ") %>%
+   separate_rows(asset, sep = ", ") %>%
+   count(asset, action, name = "freq")
> # Limpiar y ordenar nombres
> action_asset_pairs <- action_asset_pairs %>%
+   mutate(
+     asset = gsub("\\s*-\\s*", " ", asset),
+     action = gsub("\\s*-\\s*", " ", action)
+   )
>
```

```
long_df        1707 obs. of 3 variables
slim.a2        0 obs. of 7 variables
vcdb           1643 obs. of 2433 variables
values
  action_cols  chr [1:181] "action.environmental.variety.Deterioration" "action.environmental…
  action_vars  chr [1:9] "action.variety.other" "action.variety.Unknown" "action.variety.Brut…
  asset_cols   character (empty)
  asset_vars   "asset.amount.S - Database"
  jsondir      "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  jsonfile     "C:\\Users\\leand\\Downloads\\BOOK DATA DRIVEN-CYT250\\DataDrivenSecurity_book…
  new.pkg      chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  pkg          chr [1:4] "devtools" "ggplot2" "scales" "rjson"
  top_actions  chr [1:49] "Theft" "Unknown" "Privilege abuse" "Loss" "Misdelivery" "Disposal …
  top_assets   ""
```

Figure 17. Result listing 7-11

I created a refined heatmap to visualize the relationships between different types of assets and the types of actions that affected them. After filtering out unknown and irrelevant categories, I used geom_tile() in ggplot2 to plot a matrix where darker tiles represent higher frequencies of incidents. This allowed me to quickly identify patterns, such as how laptops are most often involved in physical attacks or how web applications are commonly linked to hacking incidents. This final heatmap provided a more complete and insightful view of how different types of cyber actions target specific assets.

Summary of the Final Project

Reading through Chapter 7 of Data-Driven Security has given me real-life on hands practice analyzing incident data in real life through R. It introduced me to many technical challenges such as package dependencies, path errors, and undefined objects, all of which have turned out to be my greatest opportunities to sharpen my troubleshooting and data-wrangling skills.

At the end of this project, I was able to import the JSON formatted VCDB data, explore its structure, do some descriptive summaries, and make visible clear views through ggplot2; some of these were frequency bar charts and heatmaps showing patterns between actors, actions, and affected assets.

Such practical learning has made it easy for me to understand VERIS data structures and how to work with and visualize complicated datasets in R. It has prepared me for future cybersecurity data analysis tasks where structured data, incident classification, and insight generation are key.