



Using Python, PostgreSQL and R to Analyze NIBRS

With data from the Crime Data Explorer

Ian Thomas, Dan Liao,
Alex Harding, Alex Giarrocco

What do we know about crime in the US?

- **Uniform Crime Reporting (UCR)**
 - Administrative Law enforcement agency data
 - Collected since 1929
- **Summary Reporting System (SRS)**
 - Summary Statistics
 - *Voluntarily* submitted **counts** of reported crimes by type of crime to the Federal Bureau of Investigation (FBI)
 - 99% law enforcement agencies reporting as of 2015
 - This system prevented certain types of crime, based on the characteristics of the crime, from being explicitly reported.
 - firearm violence
 - crimes committed by gangs
 - domestic violence
 - crimes against children



NIBRS - National Incident-Based Reporting System



- **National Incident-Based Reporting System**

- Developed in 1980s
- **Incident level** reporting
- provide additional detail of each incident and allow for a broader array of crimes to be reported
- provides a level of detail and context about the crime, victim, and offender
- ~ 1/3rd law enforcement agencies reporting as of 2018

How do you access NIBRS Data?

- Published Files
 - Hierarchical
 - Incident granularity
 - Large
 - Difficult to use
 - Extract Files
 - Segment granularity
 - Smaller but also large
 - Easier to use
 - Generated from FBI released master file
 - Latest Master file: 2016
- What if I want 2017 data?

Documentation Only

SAS

SPSS

Stata

R

ASCII

Delimited

Crime Data Explorer

FEDERAL BUREAU OF INVESTIGATION

Crime Data Explorer

Home

Explorer

Documents & Downloads

About

Downloads & Documentation

The Crime Data Explorer (CDE) provides select bulk data sets for download. Incident-based data by state, summary data with estimates, and data on specific topics like assaults on law enforcement officers, hate crime, or human trafficking are available for download in CSV files below. Data is also available via the Crime Data API, a read-only web service that returns JSON or CSV data and provides experienced users access to large amounts of UCR data to use and share.

Because this data is dynamic, be aware of the time stamp that reflects the refresh date of content.

The [Uniform Crime Reporting \(UCR\) Program](#) provided updated data for 2017 on September 24, 2018.

Incident-based data by state

Download NIBRS data by state and year

Kentucky

2017

Download

<http://s3-us-gov-west-1.amazonaws.com/cg-d4b776d0-d898-4153-90c8-8336f86bdfec/2017/KY-2017.zip>

Resources

- | [Readme](#)
- | [NIBRS Data Dictionary](#)
- | [NIBRS Data Diagram](#)

- <https://crime-data-explorer.fr.cloud.gov>
- 2017 available
- Released before master file
- Relational

What You Get

File Name	Size
agencies.csv	11.2 MB
agency_participation.csv	1.3 MB
nibrs_diagram.pdf	121 KB
postgres_load.sql	3 KB
postgres_setup.sql	34 KB
README.md	14 KB
sqlite_load.sql	2 KB
sqlite_setup.sql	15 KB
REF_RACE.csv	759 bytes
REF_STATE.csv	3 KB
NIBRS_WEAPON.csv	1.7 MB
NIBRS_VICTIM.csv	16.9 MB
NIBRS_VICTIM_OFFENSE.csv	9.4 MB
NIBRS_VICTIM_OFFENDER_REL.csv	3 MB
NIBRS_VICTIM_INJURY.csv	1.1 MB
NIBRS_VICTIM_CIRCUMSTANCES.csv	249 KB
NIBRS_SUSPECTED_DRUG.csv	782 KB
NIBRS_SUSPECT_USING.csv	5.9 MB
NIBRS_PROPERTY.csv	8.6 MB
NIBRS_PROPERTY_DESC.csv	12.6 MB
NIBRS_OFFENSE.csv	11 MB
NIBRS_OFFENDER.csv	13.9 MB
NIBRS_month.csv	27.9 MB
NIBRS_incident.csv	36.7 MB
NIBRS_CRIMINAL_ACT.csv	1.2 MB
NIBRS_BIAS_MOTIVATION.csv	6.2 MB
NIBRS_ARRESTEE.csv	8.3 MB
NIBRS_ARRESTEE_WEAPON.csv	2.5 MB
NIBRS_VICTIM_TYPE.csv	288 bytes
NIBRS_WEAPON_TYPE.csv	945 bytes
NIBRS_PROP_LOSS_TYPE.csv	547 bytes
NIBRS_RELATIONSHIP.csv	953 bytes
NIBRS_SUSPECTED_DRUG_TYPE.csv	474 bytes
NIBRS_USING_LIST.csv	158 bytes
NIBRS_OFFENSE_TYPE.csv	5 KB
NIBRS_PROP_DESC_TYPE.csv	2 KB
NIBRS_INJUSTIFIABLE_FORCE.csv	505 bytes

File Name	Size
agencies.csv	11.2 MB
agency_participation.csv	1.3 MB
nibrs_diagram.pdf	121 KB
postgres_load.sql	3 KB
postgres_setup.sql	34 KB
README.md	14 KB
sqlite_load.sql	2 KB
sqlite_setup.sql	15 KB
REF_RACE.csv	759 bytes
REF_STATE.csv	3 KB
NIBRS_WEAPON.csv	1.7 MB
NIBRS_VICTIM.csv	16.9 MB

Retrieving and Loading Data From CDE

```
shutil.copyfile(load, 'postgres_load.sql')
result = subprocess.run(cfg.get_command(year, 'load'), shell=True, stdout=subprocess.PIPE, stderr=subprocess.PIPE)
stdout_str = result.stdout.decode('utf-8')
ret = True
# extracting agency/incident load counts from stdout.
if stdout_str and stdout_str != '' and int(year) >= 2016:
    stdout_lines = stdout_str.split('\n')
    copy_num_regex = re.compile(r'^COPY (\d+)$')
    for i, line in enumerate(stdout_lines):
        report_key = None
        if line.startswith("\COPY agencies FROM"):
            report_key = 'agencies_loaded'
        elif line.startswith("\COPY nibrs_incident FROM"):
            report_key = 'incidents_loaded'

        if report_key is not None:
            currentReportRow[report_key] = int(
                re.match(copy_num_regex, stdout_lines[i+1]).group(1))
if result.stderr:
    currentReportRow['errors'] = result.stderr.decode('utf-8')
    log(result.stderr, state, year)
ret = False
```

- Download files
- Extract files
- Setup database
- Load files
- Log everything
- Check for errors

The CDE NIBRS Database

pgAdmin 4

File ▾ Object ▾ Tools ▾ Help ▾

Browser



Properties

SQL

Statistics

Dependencies

Dependents

- > nibrs_criminal_act_type
- > nibrs_drug_measure_type
- > nibrs_ethnicity
- ▼ nibrs_incident
 - ▼ Columns (15)
 - data_year
 - agency_id
 - incident_id
 - nibrs_month_id
 - cargo_theft_flag
 - submission_date
 - incident_date
 - report_date_flag
 - incident_hour
 - cleared_except_id
 - cleared_except_dat
 - incident_status
 - data_home
 - orig_format
 - did

Type

Name

Foreign Key

public.nibrs_incident.nibrs_incident_agency_fk

Foreign Key

public.nibrs_incident.nibrs_incident_clear_ex_fk

Foreign Key

public.nibrs_incident.nibrs_incident_month_fk

Primary Key

public.nibrs_incident_pk

Foreign Key

public.nibrs_arrestee.nibrs_arrestee_inc_fk

Foreign Key

public.nibrs_offender.nibrs_offender_nibrs_inci_fk1

Foreign Key

public.nibrs_offense.nibrs_offense_inc_fk1

Foreign Key

public.nibrs_property.nibrs_property_inc_fk

Foreign Key

public.nibrs_victim.nibrs_victim_inc_fk

Using PgAdmin

pgAdmin 4 File Object Tools Help

Browser NIBRS_2017/postgres@RDS CDE *

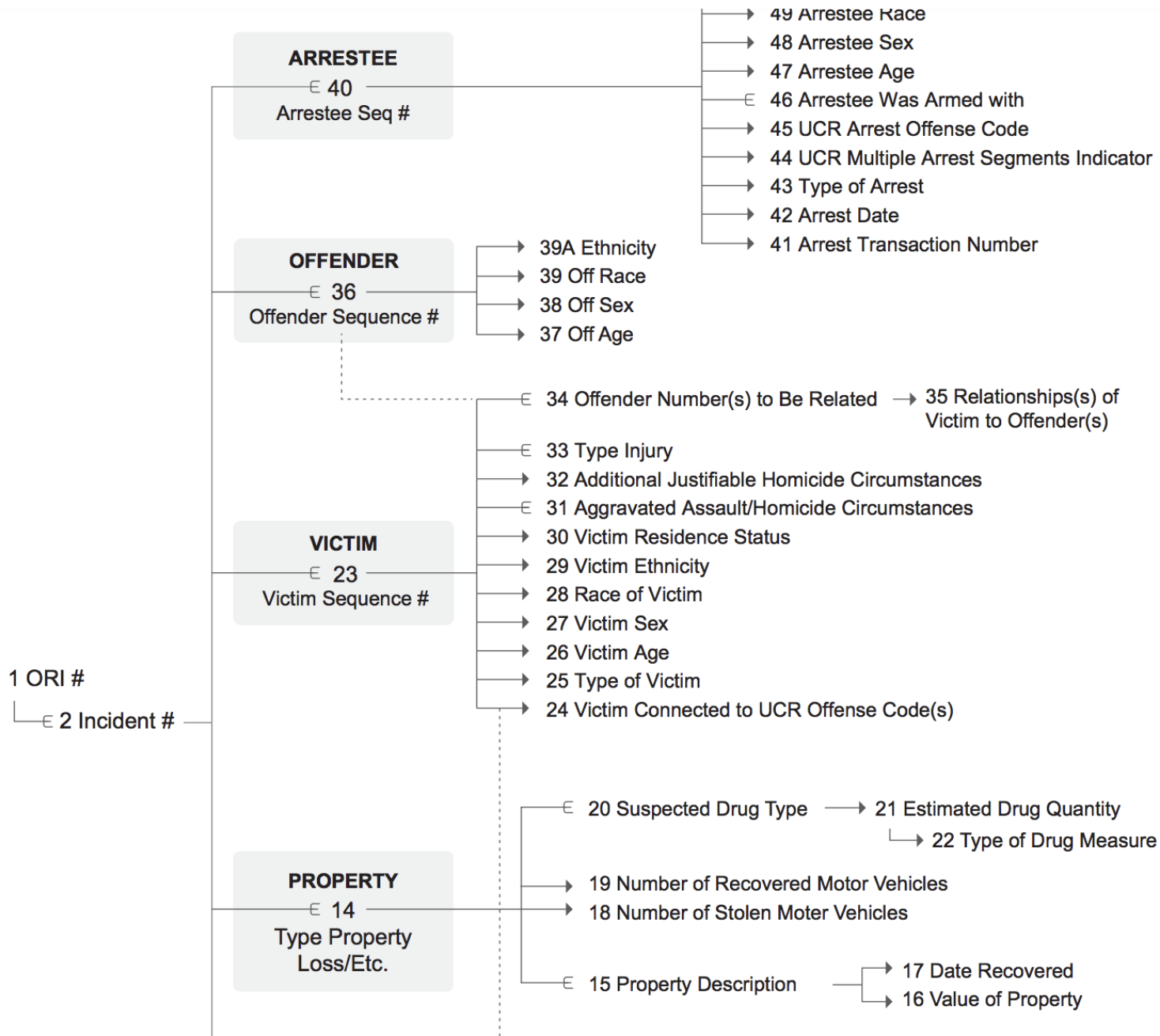
Query Editor Query History

```
1 SELECT data_year, agency_id, incident_id, nibrs_month_id, cargo_theft_flag, submission_date, incident_date, report_date_f1
2 FROM public.nibrs_incident
3 limit 1000;
```

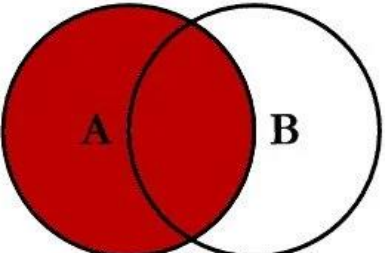
Data Output Explain Messages Notifications

	data_year integer	agency_id bigint	incident_id bigint	nibrs_month_id bigint	cargo_theft_flag character (1)	submission_date timestamp without time zone	incident_date timestamp without time zone	report_date_flag character (1)	incident_hour smallint
1	2016	61	88178288	7352848		2018-08-17 00:00:00	2016-11-15 00:00:00	R	[null]
2	2016	61	88178295	7352848		2018-08-17 00:00:00	2016-11-08 00:00:00		8
3	2016	61	88178289	7330064		2018-08-17 00:00:00	2016-10-18 00:00:00		22
4	2016	61	88178298	7352848		2018-08-17 00:00:00	2016-11-25 00:00:00		2
5	2016	61	88178297	7352848		2018-08-17 00:00:00	2016-11-29 00:00:00		17
6	2016	61	87398880	7284496		[null]	2016-08-01 00:00:00	R	[null]
7	2016	61	87399929	7284496		[null]	2016-08-27 00:00:00		1
8	2016	61	87399934	7284496		[null]	2016-08-27 00:00:00		16
9	2016	61	87399935	7284496		[null]	2016-08-27 00:00:00		18
10	2016	61	87399938	7284496		[null]	2016-08-28 00:00:00		9
11	2016	61	87399940	7284496		[null]	2016-08-26 00:00:00		18
12	2016	61	87399946	7284496		[null]	2016-08-29 00:00:00		7
13	2016	61	87399950	7284496		[null]	2016-08-29 00:00:00		8
14	2016	61	87399962	7284496		[null]	2016-08-31 00:00:00	R	[null]
15	2016	61	87399970	7307280		[null]	2016-09-01 00:00:00	R	[null]
16	2016	61	87399977	7284496		[null]	2016-08-25 00:00:00		14

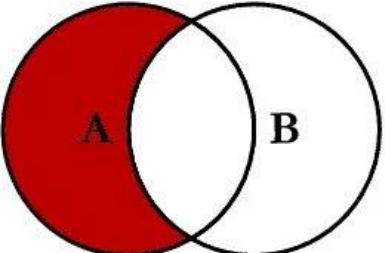
The Structure of NIBRS Data



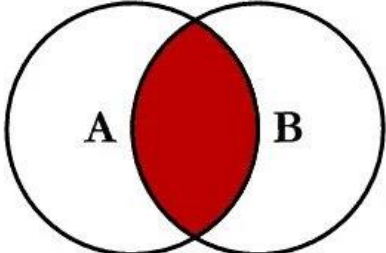
SQL JOINS



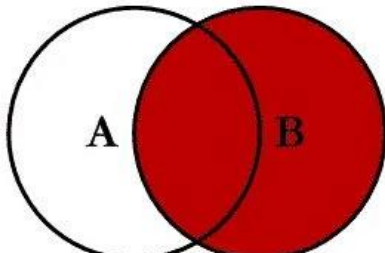
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key
```



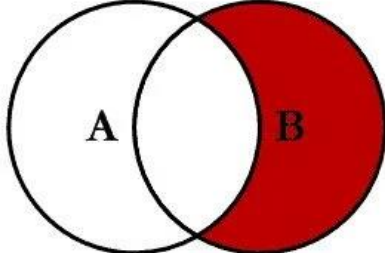
```
SELECT <select_list>  
FROM TableA A  
LEFT JOIN TableB B  
ON A.Key = B.Key  
WHERE B.Key IS NULL
```



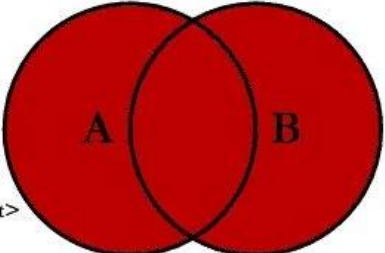
```
SELECT <select_list>  
FROM TableA A  
INNER JOIN TableB B  
ON A.Key = B.Key
```



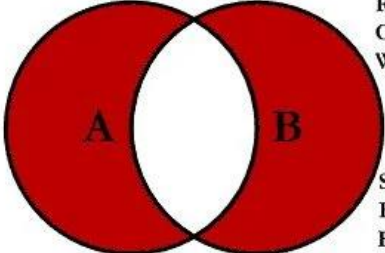
```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
RIGHT JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key
```



```
SELECT <select_list>  
FROM TableA A  
FULL OUTER JOIN TableB B  
ON A.Key = B.Key  
WHERE A.Key IS NULL  
OR B.Key IS NULL
```

Using R

```
```{r, echo=FALSE, message=FALSE}  
#install.packages("RPostgres")
#install.packages("dplyr")

library(DBI)
library(dplyr)
library(dbplyr)
library(tidyverse)
```
```

```
<!-- -----  
- -->  
<!-- Create the database connection. -->
```

```
```{r, echo=FALSE}  
con <- dbConnect(
 RPostgres::Postgres(),
 dbname = "NIBRS_2017",
 host = "thor.rtp.rti.org",
 port = 5433,
 user = 'username',
 password = 'password'
)
```
```

```
```{r}  
query1 <- "
SELECT v.victim_id, v.incident_id, v.victim_seq_num, v.victim_type_id,
 v.assignment_type_id, v.activity_type_id, v.outside_agency_id, v.age_id,
 v.age_num, v.sex_code, v.race_id, v.ethnicity_id, v.resident_status_code,
 v.age_range_low_num, v.age_range_high_num, subquery.offense_type_id as
 offense_type_id, subquery.data_year as data_year
FROM nibrs_victim v
LEFT JOIN
 (
 SELECT offense.offense_type_id, offense.incident_id, month.data_year
 FROM nibrs_incident inc
 LEFT JOIN nibrs_month month
 ON month.nibrs_month_id = inc.nibrs_month_id and month.agency_id = inc.agency_id
 LEFT JOIN nibrs_offense offense
 ON inc.incident_id = offense.incident_id
 WHERE month.data_year = 2017
) subquery
ON v.incident_id = subquery.incident_id
WHERE subquery.data_year = 2017
"

df1 <- dbGetQuery(con, query1)

dim(df1)
head(df1)
glimpse(df1)
```
```

Indicators

- NIBRS Data Dictionary
 - Quick reference for using NIBRS Data
 - Published on the CDE
- NIBRS User Manual
 - Addresses National Incident-Based Reporting System (NIBRS) policies, the types of offenses reported via NIBRS, and guidelines for an agency to become certified to submit NIBRS data to the FBI
- NIBRS Technical Specification
 - Technical details on how to submit NIBRS Incidents to the FBI

```
SELECT
    arrest_type_id,
    arrest_type_code,
    arrest_type_name,

    CASE
        WHEN arrest_type_code = 'O' THEN 1
        ELSE 0
    END
    AS is_on_view_arrest,

    CASE
        WHEN arrest_type_code = 'S' THEN 1
        ELSE 0
    END
    AS is_summoned_cited,

    CASE
        WHEN arrest_type_code = 'T' THEN 1
        ELSE 0
    END
    AS is_taken_into_custody
FROM
    "public"."nibrs_arrest_type"
```

Analysis Views

segment-incident

- Columns (83)
- incident_id
- incident_date
- cargo_theft_flag
- cleared_except_date
- is_cargo_theft
- agency_id
- data_year
- ori
- state_name
- state_abbrev
- county_name
- agency_type_name
- is_city_agency
- is_county_agency
- is_college_agency
- is_state_police_agency
- is_other_state_agency
- is_tribal_agency
- is_federal_agency
- is_other_agency
- population_group_desc
- is_city_county_100k_250k
- is_city_county_25k_99k
- is_city_county_10k_24k
- is_city_county_under_10k
- is_state_pop_group
- has_group_a_offense

NIBRS_2017/postgres@RDS CDE

Query Editor Query History

```
1 SELECT ori, incident_id,
2     agency_id, data_year, state_name,
3     has_injury, has_damaged_prop_loss, has_recovered_prop_loss,
4     incident_hour
5 FROM dbt_nibrs."segment-incident" limit 1000;
```

Data Output Explain Messages Notifications

| | ori | incident_id | agency_id | data_year | state_name | has_injury | has_damaged_prop_loss |
|----|------------------------|-------------|-----------|-----------|----------------|------------|-----------------------|
| | character varying (25) | bigint | integer | integer | character vary | integer | integer |
| 1 | TX2201200 | 91951232 | 19711 | 2017 | Texas | 0 | 0 |
| 2 | TX2201200 | 91951944 | 19711 | 2017 | Texas | 0 | 0 |
| 3 | TX2201200 | 91952393 | 19711 | 2017 | Texas | 0 | 0 |
| 4 | TX2201200 | 91952704 | 19711 | 2017 | Texas | 0 | 0 |
| 5 | TX2201200 | 91953185 | 19711 | 2017 | Texas | 0 | 0 |
| 6 | TX2201200 | 91953486 | 19711 | 2017 | Texas | 0 | 0 |
| 7 | TX2201200 | 91953834 | 19711 | 2017 | Texas | 0 | 1 |
| 8 | TX2201200 | 91954345 | 19711 | 2017 | Texas | 0 | 0 |
| 9 | TX2201200 | 91954850 | 19711 | 2017 | Texas | 0 | 0 |
| 10 | TX2201200 | 91954924 | 19711 | 2017 | Texas | 0 | 0 |
| 11 | TX2201200 | 91955210 | 19711 | 2017 | Texas | 0 | 0 |
| 12 | TX2201200 | 91955607 | 19711 | 2017 | Texas | 0 | 0 |
| 13 | TX2201200 | 91955620 | 19711 | 2017 | Texas | 0 | 1 |
| 14 | TX2201200 | 91955714 | 19711 | 2017 | Texas | 0 | 1 |

Creating Views Using DBT



Search for models...

models

- base
- indicator
 - indicator-arrestee
 - indicator-incident
 - indicator-offense
 - indicator-offense-gang
 - indicator-offense-hate
 - indicator-victim-business
 - indicator-victim-person
- recoded
- segment-agency
- segment-arrestee
 - segment-arrestee
- segment-incident
 - incident-join-year
 - segment-incident
 - segment-incident-arrestee
 - segment-incident-clearance
 - segment-incident-offense
 - segment-incident-victim

segment-incident-victim view

Details Description Columns SQL

| | |
|------------------|-------------------|
| victim_type_name | character varying |
|------------------|-------------------|

SQL

Source Compiled

```
1 SELECT
2   i.incident_id,
3   v.victim_id,
4   v.victim_type_code,
5   v.victim_type_name
6 FROM
7   "dbt_nibrs"."segment-incident" i
8   JOIN "dbt_nibrs"."segment-victim" v
9     USING (incident_id)
```

Lineage Graph



Analysis View Explorer

NIBRS Data Explorer [Main](#) [Search](#) [About](#)

segment-incident

segment-incident-offense

segment-incident-victim

ARRESTEE ⓘ

segment-arrestee

OFFENDER ⓘ

segment-offender

PROPERTY ⓘ

segment-property

segment-property-prop-desc

segment-property-suspected-drug

BASE ⓘ

segment-incident-victim

Number of Rows: 11,884,348

Joins victim segment to incident information.


Sample Data (10 rows)

SQL

Column Reference

| incident_id | victim_id | victim_type_code | victim_type_name |
|-------------|-----------|------------------|------------------|
| 86385150 | 90358596 | I | Individual |
| 85736652 | 90354158 | I | Individual |
| 86956143 | 90361126 | I | Individual |
| 87416030 | 90376848 | I | Individual |
| 87446241 | 90382583 | I | Individual |
| 84528424 | 90406187 | I | Individual |
| 83353349 | 90451854 | I | Individual |
| 83625187 | 90449185 | I | Individual |

Future Work

- 
- Refine analysis workflow
 - Work with analysts to create more useful views
 - Create views for all indicators
 - Add Quality Assurance to import
 - Explore using columnar databases
 - Add versioning and history to database

Key Take Aways

- For the latest data go to:
<https://crime-data-explorer.fr.cloud.gov>
- Download the states and years you need
- Load the data using PostgreSQL or SQLite
- Use SQL to create analysis views
 - Use 'Limit 100' to test
 - Join wisely
- Remember this data will get updated





delivering **the promise of science**
for global good



Ian Thomas
ianthomas@rti.org

Dan Liao
Alex Harding
Alex Giarrocco