BJS NIBRS Extracts: An introduction and how-to guide

Webinar presentation for the Justice Research Statistics Association
March 29, 2022
SJS Program: Core-Capacity Building Projects

- Up to $75,000/year (12-36 month project period).

- Final deliverables must be submitted to BJS as part of the final performance report, including performance measures, associated reports, evaluations, or website enhancements completed.
  - May include travel funds for up to 2 people to attend up to 2 SJS–focused conferences/meetings approved by BJS.
  - Must describe current capabilities, how project will enhance them, and how enhancements will be sustained.
Conduct research and analysis of **NIBRS-compliant incident-based crime data** that will reflect the utility of these data for studying criminal justice issues.

Projects proposed under this area should include:

- An assessment and review of data quality
- Publication quality final report to be disseminated nationally that includes a thorough discussion of data quality
NIBRS/IBR: Areas of Interest to BJS

- Law enforcement agency recording of the relationship(s) between victim(s) and offender(s) for violent offenses.
- Analyses and data quality assessments of data on property crimes.
- Analyses that examine differences in how LE agencies have recorded assault over time and across reporting agencies.
  - Example: have assault rates increased due to laws or policies mandating arrest for certain types of assaults?
- Analyses that compare IBR practices across reporting agencies over time to understand changes data entry and potential impacts on reporting.
- Projects that examine the utility of linking incident-based crime reports to criminal history records.
  - Must be conducted in partnership with the state’s criminal history record repository
  - Must include an assessment and review of the quality of the criminal history record information used.
Let’s talk about NIBRS

• Purpose of today’s webinar
  – What is a NIBRS extract?
  – History of BJS with NIBRS extract files and ICPSR
  – What are their purpose and value? Why use an extract?
  – How does BJS use them?
  – Value for researchers, SACs, local agencies
  – Current work and BJS plans for future extract files
What are NIBRS Extract Files?

Each NIBRS segment provides information on a single aspect of an incident such as property loss, victims, offenders, arrestees.

Segments are combined to produce the NIBRS extract files: incident, victim, arrestee.

Extract files are in wide format to produce a single observation per record that includes related data elements from other NIBRS segments.

Each extract reflects a specific perspective on the incident by allowing users to examine incidents using several units of analysis (i.e., incident, victim, arrestee) available from other NIBRS segments.

Easily import NIBRS data into standard statistical packages for analysis and data visualization.
Background: NIBRS Extracts at the NACJD

• Historically, the NACJD at the University of Michigan’s Inter-university Consortium for Political and Social Research (ICPSR) developed the process to create the NIBRS extract files from 1991 to 2016.

• NIBRS extract files and codebooks for 1991-2016 are available for download at ICPSR: National Incident-Based Reporting System (NIBRS) Series (umich.edu)

• Extract files allow for more efficient analyses, but are complex and require detailed knowledge of NIBRS to analyze the data properly.
  – Example: connecting weapons, clearance, location and timing, and information about property losses and recoveries to the victim segment, and defining car-jackings and mass violence incidents.
Key Terms and Definitions

- **Segments** – basic format for how NIBRS data elements are grouped together.
  - Single record for a given type of data; 7* NIBRS segments

- **Extracts** – extract files consist of segment data combined to analyze data from a particular segment perspective and includes summarized information from other segments to make analysis easier
  - Victim, Arrestee, Incident extracts

- **NIBRS master file** - contains each NIBRS segment within a single file in long format. Each record can be any segment but is prefixed with a leading indicator code

- **Transformation** - process by which segments are prepared to create extract files

- **Long format** – segments are in a long format which means they contain multiple records per incident
  - Victim, property, arrestee, etc.

- **Wide format** – taking long format data and structuring it so that you create a single record for a given extract/perspective

- **Join** – the process of combining segments based on a logical key to produce extract files
  - Linkage variables in ICPSR language
  - Sometimes referred to a ‘concatenated’ file.

- **Shell script** – a list of commands in a computer program that are run by the operating system
Step 1: Review NIBRS standards, business rules

Step 2: Integrate standards into NIBRS-compliant RMS

Step 3: Officers input data into the RMS System

Step 4: Agencies submit NIBRS data to state UCR programs

Step 5: State UCR program submits NIBRS data to FBI

Step 6: FBI releases NIBRS via the CDE

Step 7: BJS NIBRS extracts sourced from CDE

Key challenge: How can we benchmark data without merging data from multiple states?
NIBRS Segments Used to Create BJS Extracts

- These 7 segments are the primary input data for the NIBRS extract files:
  - Batch Header
  - Administrative
  - Offense
  - Property
  - Victim
  - Offender
  - Arrestee

- [https://www.icpsr.umich.edu/web/pages/NACJD/NIBRS/varlist.html](https://www.icpsr.umich.edu/web/pages/NACJD/NIBRS/varlist.html)
<table>
<thead>
<tr>
<th>BH001</th>
<th>BH002</th>
<th>BH003</th>
<th>BH004</th>
<th>BH005</th>
<th>BH006</th>
<th>BH007</th>
<th>BH008</th>
<th>BH009</th>
</tr>
</thead>
<tbody>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010100</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Anchorage</td>
<td>AK</td>
<td>1C</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010200</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Fairbanks</td>
<td>AK</td>
<td>4</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010300</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Juneau</td>
<td>AK</td>
<td>4</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010400</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Ketchikan</td>
<td>AK</td>
<td>6</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010500</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Kodiak</td>
<td>AK</td>
<td>6</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010600</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Nome</td>
<td>AK</td>
<td>6</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010700</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Petersburg</td>
<td>AK</td>
<td>6</td>
</tr>
<tr>
<td>BH</td>
<td>50</td>
<td>AK0010800</td>
<td>0</td>
<td>20200630</td>
<td>0</td>
<td>Seward</td>
<td>AK</td>
<td>6</td>
</tr>
</tbody>
</table>
## Victim segment
Contains NIBRS offense codes, victim type, victim demographics, victim injury, victim-offender relationships, homicide circumstances, residence status of the victim, multiple records per incident.

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>V4001</strong></td>
<td><strong>V4002</strong></td>
<td><strong>V4003</strong></td>
<td><strong>V4004</strong></td>
<td><strong>V4005</strong></td>
<td><strong>V4006</strong></td>
<td><strong>V4007</strong></td>
<td><strong>V4008</strong></td>
<td><strong>V4009</strong></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7282X</td>
<td>20171230</td>
<td>1</td>
<td>23F</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7284J4</td>
<td>20171222</td>
<td>1</td>
<td>250</td>
<td>26A</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7284XO</td>
<td>20171227</td>
<td>1</td>
<td>250</td>
<td>NA</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU72878I</td>
<td>20171231</td>
<td>1</td>
<td>23F</td>
<td>NA</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7287TI</td>
<td>20171231</td>
<td>1</td>
<td>23F</td>
<td>NA</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU728G2I</td>
<td>20171231</td>
<td>1</td>
<td>23F</td>
<td>NA</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU728GTI</td>
<td>20171231</td>
<td>1</td>
<td>23F</td>
<td>NA</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU728Y04</td>
<td>20171231</td>
<td>1</td>
<td>23F</td>
<td>NA</td>
</tr>
</tbody>
</table>
## Offense Segment

Contains information on offense codes, location type, weapon and force used, suspected offender drug use, bias motivation, premises entered, attempts vs completed crimes, multiple records per incident.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7282X</td>
<td>20171230</td>
<td>23F</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7284J4</td>
<td>20171222</td>
<td>26A</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7284J4</td>
<td>20171222</td>
<td>250</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7284XO</td>
<td>20171227</td>
<td>250</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU72878I</td>
<td>20171231</td>
<td>23F</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU7287TI</td>
<td>20171231</td>
<td>23F</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU728G2I</td>
<td>20171231</td>
<td>23F</td>
<td>C</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>AL0011200</td>
<td>2W14PU728GTI</td>
<td>20171231</td>
<td>23F</td>
<td>C</td>
<td>N</td>
</tr>
</tbody>
</table>
Incident Extract

- Consists of 1 incident per record.
- Up to N records from other segments joined to the incident.
  - For example, up to 3 records from the offense segment are joined in but there could be greater than 3 offense records for a given incident.
- Joined data is converted to wide format for analytic purposes.
- Enables more detailed analysis of the circumstances of incidents.
## Incident Extract Example

<table>
<thead>
<tr>
<th>RECSOFS</th>
<th>RECSPRP</th>
<th>RECSVIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>N OFFENSE (02) RECORDS PER ORI-INCIDENT NUMBER</td>
<td>N PROPERTY (03) RECORDS PER ORI-INCIDENT NUMBER</td>
<td>N VICTIM (04) RECORDS PER ORI-INCIDENT NUMBER</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>V20061</th>
<th>V20062</th>
<th>V20063</th>
<th>V20071</th>
<th>V20072</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCR OFFENSE CODE – 1</td>
<td>UCR OFFENSE CODE – 2</td>
<td>UCR OFFENSE CODE – 3</td>
<td>OFFENSE ATTEMPTED / COMPLETED – 1</td>
<td>OFFENSE ATTEMPTED / COMP</td>
</tr>
<tr>
<td>220</td>
<td>290</td>
<td>351</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>280</td>
<td>351</td>
<td>352</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>131</td>
<td>238</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>233</td>
<td>261</td>
<td>351</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>220</td>
<td>238</td>
<td>262</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>132</td>
<td>238</td>
<td>240</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>132</td>
<td>250</td>
<td>351</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
## Table 3. Percentage of single- and multiple-offense incidents, by type of co-occurring offense and agency population served, 2016

<table>
<thead>
<tr>
<th>Incident Type</th>
<th>All Agencies</th>
<th>&lt;100,000</th>
<th>100,000–199,999</th>
<th>≥200,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Single-Offense Incidents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Part 1 offense</td>
<td>51.1</td>
<td>48.8</td>
<td>53.5</td>
<td>54.8</td>
</tr>
<tr>
<td>Single non-Part 1 offense</td>
<td>49.0</td>
<td>51.2</td>
<td>46.5</td>
<td>45.2</td>
</tr>
<tr>
<td>Multiple-Offense Incidents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple Part 1 offenses only</td>
<td>1.1</td>
<td>1.1</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Single Part 1 offense + at least 1 non-Part 1 offense</td>
<td>5.1</td>
<td>4.8</td>
<td>5.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Multiple non-Part 1 offenses only</td>
<td>5.1</td>
<td>5.9</td>
<td>4.6</td>
<td>3.7</td>
</tr>
<tr>
<td>Multiple Part 1 offenses and at least one non-Part 1 offense</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>Number of Incidents</td>
<td>5,207,481</td>
<td>3,109,283</td>
<td>761,858</td>
<td>1,336,340</td>
</tr>
</tbody>
</table>

Multiple Offense Incidents Provide Clues....
Victim Extract

- Consists of 1 victimization per record.
- Up to N records from other segments are joined to the victim.
- Joined data is converted to wide format for analytic purposes.
- The same incident numbers repeat if multiple victims present.
- Enables analysis of data from the victim’s perspective.

- **Objective**: users can explore victimizations that have additional details about the incident appended to an individual, or other victim type.

- **Examples include**: robbery victimization records with weapons and injury codes attached, proceeds from the robbery, and whether an offender was arrested.
NIBRS Victim Extract in SPSS: A brief look
Arrestee Extract

- Consists of 1 arrestee per record
- Up to N records from other segments joined to the victim
- Joined data is converted to wide format for analytic purposes.
- Incident numbers could repeat if multiple arrestees per incident
- Enables analysis of data from an arrestee perspective as the unit of analysis, with additional data elements from other segments.
NIBRS Arrestee Extract in SPSS: A brief look
Where to start?

We had to determine...
Sourcing NIBRS Data: State standards and federal NIBRS data on the CDE
Unpacking the Black Box: How were the 1991-2016 ICPSR extract files created?

ICPSR Extract Creation Process
Tech stack used to create NIBRS extract files:
Open-source resources

- R Studio
- BASH
- PostgreSQL

SED Command in Linux/Unix

```
~$ sed 's/Unix/Linux/' file.txt
```
Logic for Creating Extract Files

1. Download Master File from FBI
2. Shell script separates each segment into unique text files
3. R code maps raw values to ICPSR numeric values for each segment
4. R code pivots segments from long format to wide format
5. R code filters segments and joins based on desired extract file (incident, victim, arrestee)
6. R code special handling based on QA process (substituting descr values for spaces and other special cases)
7. R code embeds lookup descriptions based on ICPSR numeric values
8. R code writes target output format: Stata, R, or SPSS
9. QA process to compare to known good or analyze variance from year-to-year, unexpected values, etc
10. Output files delivered to BJS
Process Flow for NIBRS Data Sourced from the CDE

1. Download long format data from CDE Database
2. R code pivots from long format to wide format for each segment & fills missing values
3. R code maps raw values to ICPSR numeric values for each segment
4. R code pivots segments from long format to wide format
5. R code filters segments and joins based on desired extract file (incident, victim, arrestee)
6. R code special handling based on QA process (substituting descr values for spaces and other special cases)
7. R code embeds lookup descriptions based on ICPSR numeric values
8. R code writes target output format: Stata, R, or SPSS
9. QA process to compare to known good or analyze variance from year-to-year, unexpected values, etc
10. Output files delivered to BJS

Everything else down-stream is the same as with the master file-based process.
Moving Forward

• CDE relational database driven extract process
  – Faster delivery of analysis files rather than using NIBRS master files which are released later in the year from the FBI
  – Opportunity to create analysis files based on most recently released data

• Open Sourcing of extract creation process
  – Opportunity for community input and use

• Annual cadence for extract data release

• Concluding remarks
Kimberly Martin, PhD
Statistician
Law Enforcement Incident-Based Statistics Unit, BJS

Kimberly.Martin@usdoj.gov
(202)616-4592
Resources: Example R code to create

• Consists of 2 scripts
  – A simple shell script to separate segments into distinct files
  – An R script that performs the transformations and joins to create a particular extract file.

• Resources available: Resources on using code in R to create extracts, data dictionaries, business rules and logic for joins.
Key Terms and Definitions

- **Segments** – basic format for how NIBRS data elements are grouped together.
  - Single record for a given type of data; 7* NIBRS segments

- **Extracts** – extract files consist of segment data combined to analyze data from a particular segment perspective and includes summarized information from other segments to make analysis easier.
  - Victim, Arrestee, Incident extracts

- **NIBRS master file** - contains each NIBRS segment within a single file in long format. Each record can be any segment but is prefixed with a leading indicator code.

- **Transformation** - process by which segments are prepared to create extract files.

- **Long format** – segments are in a long format which means they contain multiple records per incident.
  - Victim, property, arrestee, etc.

- **Wide format** – taking long format data and structuring it so that you create a single record for a given extract/perspective.

- **Join** – the process of combining segments based on a logical key to produce extract files.
  - Linkage variables in ICPSR language
  - Sometimes referred to a ‘concatenated’ file.

- **Shell script** – a list of commands in a computer program that are run by the operating system.