

Updated BJA Dataset Analysis

Amy Duan

2025-03-10

General BJA Dataset Observations:

- BJA dataset contains 2,297 entries for systems that don't exist in our dataset.
- BJA dataset ranges from 2019-2024.

Missing BJA Systems:

- BOP
- ICE
- TN

Missing Third City Project Systems:

- IL
- IN
- ME
- MD
- MA
- MI
- NJ
- NY
- ND
- OH
- PR
- RI
- WV
- WI
- GU
- AmSa

BJA vs. Third City Project Collected Variables

```
#manual mapping of overlapping BJA to TCP variables
bja_cols <- c("Manner of Death", "Concat ID", "Gender", "Location of Death",
             "Middle Name", "Race", "State", "Time of Death", "Age",
             "Birth Year", "Brief Circumstances", "Calendar Year Death",
             "Date of Death", "Decedent", "Ethnicity", "First Name")
tcp_cols <- c("c_ind_cod_type", "ind_id", "c_ind_gender", "ind_deathloc",
             "ind_middle", "c_ind_race", "c_system_abbr", "ind_tod",
             "c_ind_age", "c_ind_dob_year", "combo_COD_c", "c_ind_dod_ymd",
             "c_ind_dod_ymd", "c_ind_last", "c_ind_ethnicity", "c_ind_first")

mapped_cols <- data.frame(bja_cols, tcp_cols)
kable(mapped_cols)
```

bja_cols	tcp_cols
Manner of Death	c_ind_cod_type
Concat ID	ind_id
Gender	c_ind_gender
Location of Death	ind_deathloc
Middle Name	ind_middle
Race	c_ind_race
State	c_system_abbrev
Time of Death	ind_tod
Age	c_ind_age
Birth Year	c_ind_dob_year
Brief Circumstances	combo_COD_c
Calendar Year Death	c_ind_dod_ymd
Date of Death	c_ind_dod_ymd
Decedent	c_ind_last
Ethnicity	c_ind_ethnicity
First Name	c_ind_first

Addition Relevant BJA Columns/Variables:

- Location Type
- Zip Code
- Age Range
- City

Additional Relevant TCP Columns/Variables:

- c_ind_fachoused
- c_ind_offense_date_ymd

BJA vs. Third City Project Number of Recorded Individual Deaths by System

```

bja_system_count <- bja_dataset_sub %>%
  group_by(system) %>%
  count() %>%
  rename(count_bja = n)
tcp_system_count <- tcp_data %>%
  group_by(c_system_abbrev) %>%
  count() %>%
  rename(count_tcp = n,
         system = c_system_abbrev)

system_count_totals <- full_join(bja_system_count, tcp_system_count,
                                by = "system")
kable(system_count_totals)

```

system	count_bja	count_tcp
AK	27	40
AL	489	11
AR	22	37
AZ	74	1301
AmSa	1	NA
CA	1027	283

system	count_bja	count_tcp
CO	173	7
CT	26	4
DC	25	17
DE	40	125
FL	1546	3211
GA	579	110
GU	3	NA
HI	43	96
IA	142	154
ID	122	50
IL	24	NA
IN	426	NA
KS	127	20
KY	65	1
LA	178	2
MA	81	NA
MD	158	NA
ME	5	NA
MI	372	NA
MN	75	10
MO	218	3
MS	90	4
MT	60	64
NC	456	138
ND	7	NA
NE	112	147
NH	8	11
NJ	202	NA
NM	15	2
NV	12	795
NY	126	NA
OH	70	NA
OK	2	9
OR	155	31
PA	453	133
PR	10	NA
RI	2	NA
SC	422	56
SD	33	196
TX	1617	180
UT	75	19
VA	20	12
VT	2	27
WA	8	16
WI	225	NA
WV	92	NA
WY	34	28
BOP	NA	311
ICE	NA	58
TN	NA	20
NA	NA	8848

BJA vs. Third City Project Number of Recorded Individual Deaths by COD

```
tcp_cod_count <- tcp_data %>%
  group_by(c_ind_cod_type) %>%
  count() %>%
  rename(cod_count_tcp = n,
         cod = c_ind_cod_type)
bja_cod_count <- bja_dataset_sub %>%
  group_by(cod) %>%
  count() %>%
  rename(cod_count_bja = n)

cod_count_totals <- full_join(bja_cod_count, tcp_cod_count, by = "cod")
kable(cod_count_totals)
```

cod	cod_count_bja	cod_count_tcp
Accident	515	NA
Execution	6	750
Homicide	397	654
Homicide by LEO	1247	13
Natural	5590	8650
Other	460	NA
Pending	1295	856
Suicide	866	769
COVID-19	NA	562
Drug / Alcohol	NA	17
Unintentional non-Drug Injury	NA	996
Unknown	NA	3320

Mapping BJA to Third City Project

```
#filter out entries that have conflicting dod values
joined %>%
  filter(dod == c_ind_dod_ymd)
```

```
## # A tibble: 1,405 x 29
##   system full_id          first_name middle_name last_name dod      cod
##   <chr>   <chr>          <chr>     <chr>     <chr>   <date>   <chr>
## 1 FL     RAMIREZADELADEL CAR~ ADELA     Del Carmen~ RAMIREZ   2020-07-25 Natu~
## 2 FL     DAVILAJOSEFASOCORRO~ JOSEFA     Socorro    DAVILA    2020-06-25 Natu~
## 3 FL     ROMANESLOTIATFL2020~ LOTIA      T          ROMANES   2020-04-09 Natu~
## 4 FL     SOLOMONCHINETTEANDR~ CHINETTE   Andrea     SOLOMON   2019-11-04 Natu~
## 5 FL     MUSIALRUTHMARIEFL20~ RUTH       Marie      MUSIAL    2019-10-08 Natu~
## 6 FL     WILLIAMSTYRALATIFAF~ TYRA       Latifa     WILLIAMS  2020-06-25 Natu~
## 7 FL     FAULKSPAMELADENISEF~ PAMELA     Denise     FAULKS    2020-08-04 Natu~
## 8 FL     GREENBARBARAJEANFL2~ BARBARA    Jean       GREEN     2020-05-12 Natu~
## 9 FL     TETERBRITTANYELIZAB~ BRITTANY   Elizabeth  TETER     2020-05-12 Natu~
## 10 FL    COLLINSKIMBERLYJEAN~ KIMBERLY   Jean       COLLINS   2020-02-11 Natu~
## # i 1,395 more rows
## # i 22 more variables: cod_description <chr>, death_loc <chr>, tod <chr>,
## #   age <dbl>, gender <chr>, race <chr>, ethnicity <chr>, dob_year <dbl>,
## #   ci_type <chr>, c_ind_full_name <chr>, c_ind_first <chr>, c_ind_last <chr>,
## #   c_ind_dob_year <dbl>, c_ind_gender <chr>, c_ind_cod_type <chr>,
```

```
## #   c_ind_race <chr>, c_ind_ethnicity <chr>, c_ind_dod_ymd <date>,  
## #   ind_tod <chr>, ind_deathloc <chr>, c_ind_fachoused <chr>, ...
```