Updated BJA Dataset Analysis

Amy Duan

2025-03-10

Library Setup and Data Importation

```
library(tidyverse)
library(here)
library(stringi)
library(knitr)

bja_file_path <- here("data", "3102025_cleaner_bja_tbl.csv")
bja_dataset <- read_csv(bja_file_path)

tcp_file_path <- here('data', 'aggregate_data.csv')
tcp_data <- read_csv(tcp_file_path)
```

Third City Project Dataset Cleaning

```
tcp_data <- tcp_data %>%
  mutate(
    press_release_system = case_when(
      c_collection_type != "Webscrape Table" ~
        str_split_fixed(file, "_", 3)[, 1],
      TRUE ~ NA_character_
    )
  ) %>%
  mutate(
    c_system_abbr = case_when(
      c_collection_type != "Webscrape Table" & is.na(c_system_abbr) ~
        press_release_system,
      TRUE ~ c_system_abbr
    )
  )
```

BJA Dataset Cleaning

```
#Rename BJA dataset variables and selects relevant variables
bja_dataset_renamed <- bja_dataset %>%
  rename(dod_year = `Calendar Year Death`,
         cod = `Manner of Death`,
         death_loc = `Location of Death`,
         dod = `Date of Death`,
         tod = `Time of Death`,
         age = `Age`,
         ci_type = `Location Type`,
```

```
         full_id = `Concat ID`,
         gender = `Gender`,
         system = `State`,
         middle_name = `Middle Name`,
         dob_year = `Birth Year`,
         race = `Race`,
         last_name = `Decedent`,
         first_name = `First Name`,
         ethnicity = `Ethnicity`,
         cod_description = `Brief Circumstances`)
bja_dataset_sub <- bja_dataset_renamed %>%
  select(system, full_id, first_name, middle_name, last_name, dod, cod,
         cod_description, death_loc, tod, age, gender, race,
         ethnicity, dob_year, ci_type)
```

```
bja_dataset_sub <- bja_dataset_sub %>%
  mutate(cod = case_when(
    cod == "Death attributed to use of force by a law enforcement or corrections officer" ~ "Homicide by
    cod == "Homicide (e.g., an incident between two or more incarcerated individuals resulting in a deat
    cod == "Unavailable, investigation pending" ~ "Pending",
    cod == "Natural causes" ~ "Natural",
    TRUE ~ cod
  )) %>%  #recode to standardized causes of death
  filter(stri_enc_isascii(first_name)) %>%
  filter(stri_enc_isascii(middle_name)) %>%
  filter(stri_enc_isascii(last_name)) %>%
  mutate(first_name = str_to_title(first_name),
         middle_name = str_to_title(middle_name),
         last_name = str_to_title(last_name)) %>%
  mutate(first_name = case_when(
          first_name == "N/A" ~ NA,
          TRUE ~ first_name
         ),
         middle_name = case_when(
          middle_name == "N/A" ~ NA,
          TRUE ~ middle_name
         ),
         last_name = case_when(
          last_name == "N/A" ~ NA,
          TRUE ~ last_name
         ),
         cod_description = case_when(
          cod_description == "N/A" ~ NA,
          TRUE ~ cod_description
         ),
         death_loc = case_when(
          death_loc == "N/A" ~ NA,
          TRUE ~ death_loc
         )
  ) %>% #clean first, middle, and last names
  mutate(dod = as.Date(dod, "%m/%d/%Y")) %>%
  mutate(tod = format(strptime(tod, format = "%H:%M"), "%I:%M %p")) %>%
  filter(ci_type %in% c("Prison", "State/Local Law Enforcement"))
```

BJA Analysis

```
prison_systems <- c("AK", "AL", "AR", "AZ", "BOP", "CA", "CO", "CT", "DE", "DC", "FL", "GA", "HI", "ICE
missing_prison_systems <- c(setdiff(prison_systems, tcp_data$system),
                            setdiff(bja_dataset_sub$system, prison_systems))
overlapping_prison_systems <- setdiff(prison_systems, missing_prison_systems)
```

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 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_abbr) %>%
  count() %>%
  rename(count_tcp = n,
         system = c_system_abbr)

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
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

system	count_bja	count_tcp
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

```
dcra_variables <- c("c_system_abbr", "c_ind_full_name", "c_ind_first", "c_ind_last", "c_ind_dob_year",
```

```
tcp_data <- tcp_data %>%
  mutate(c_ind_first =
           case_when(
             c_ind_first == "N/A" ~ NA,
             TRUE ~ c_ind_first),
         c_ind_last =
           case_when(
             c_ind_last == "N/A" ~ NA,
             TRUE ~ c_ind_last
           )
        )

dcra_dataset <- tcp_data %>%
  mutate(c_ind_full_name = ifelse(!is.na(c_ind_first) & !is.na(c_ind_last),
                                  paste(c_ind_first, " ", c_ind_last),
                                  NA),
         c_ind_cod_avail = ifelse(c_ind_cod_avail != "Listed", NA,
                                  c_ind_cod_avail)) %>%
  select(all_of(dcra_variables)) %>%
  rename(system = c_system_abbr) %>%
  mutate(c_ind_full_name = str_trim(c_ind_full_name))

bja_dcra_dataset <- bja_dataset_sub %>%
  mutate(first_name = toupper(first_name),
         last_name = toupper(last_name)) %>%
  mutate(c_ind_full_name = ifelse(!is.na(first_name) & !is.na(last_name),
                                  paste(first_name, " ", last_name),
                                  NA)) %>%
  mutate(c_ind_full_name = str_trim(c_ind_full_name))

dcra_dataset <- dcra_dataset %>%
  filter(!is.na(c_ind_full_name))
```

```r
bja_dcra_dataset <- bja_dcra_dataset %>%
  filter(!is.na(c_ind_full_name))
```

```r
dcra_dataset_overlapping <- dcra_dataset %>%
  filter(system %in% overlapping_prison_systems)
bja_dcra_dataset_overlapping <- bja_dcra_dataset %>%
  filter(system %in% overlapping_prison_systems)
```

```r
joined <- bja_dcra_dataset %>%
  inner_join(dcra_dataset, by = c("c_ind_full_name", "system"))
```

```r
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>, ...
```

Subanalysis of BJA/Third City Project Datasets (2019-2024)

```r
dcra_dataset_sub <- dcra_dataset %>%
  mutate(ind_dod_year = year(c_ind_dod_ymd)) %>%
  filter(ind_dod_year %in% c(2019, 2020, 2021, 2022, 2023, 2024))
```

```r
dcra_dataset_overlapping_sub <- dcra_dataset_sub %>%
  filter(system %in% overlapping_prison_systems)
bja_dcra_dataset_overlapping <- bja_dcra_dataset %>%
  filter(system %in% overlapping_prison_systems)

joined_sub <- bja_dcra_dataset_overlapping %>%
  inner_join(dcra_dataset_overlapping_sub, by = c("c_ind_full_name", "system"))
```

```r
joined_sub %>%
  filter(dod == c_ind_dod_ymd)
```

```
## # A tibble: 1,405 x 30
##    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 23 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>, ...
```