Immigration Detention Centers Analysis

Michael Ale

2025-04-26

1. Importing Necessary Libraries

```
#tinytex::install_tinytex(force = TRUE)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
```

2. Loading the Dataset

```
df <- read.csv("immigration-detention.csv")
head(df)
```

```
##
## 1      These statistics are made available to the public pursuant to H.R. 1158 Sec. 218 - Department
## 2
## 3
## 4 This list is limited to facilities that have a population count of greater than or equal to 1 as th
## 5
## 6
##    X.1   X.2     X.3     X.4     X.5     X.6                     X.7
## 1
## 2
## 3
## 4
## 5
## 6 City State Level A Level B Level C Level D Last Inspection End Date
```

3. Cleaning the Dataset

3.1 Removing Headers in First Few Rows

```
df <- read.csv("immigration-detention.csv", skip = 6)
head(df)
```

```
##                                   Name       City State     Level.A    Level.B
## 1             ADAMS COUNTY DET CENTER    NATCHEZ    MS 1876.461078 266.431138
## 2       ADELANTO ICE PROCESSING CENTER   ADELANTO    CA    6.401198   4.323353
## 3   ALAMANCE COUNTY DETENTION FACILITY     GRAHAM    NC    4.329341   3.461078
## 4         ALEXA$NDRIA STAGING FACILITY ALEXANDRIA    LA  137.520958  47.413174
## 5                 ALLEGANY COUNTY JAIL    BELMONT    NY    1.221557   0.018000
## 6 ALL%EN PARISH PUBLIC SAFETY +COMPLEX    OBERLIN    LA  101.311377  30.251497
##     Level.C   Level.D Last.Inspection.End.Date
## 1  6.724551  4.257485                   45673
## 2 22.604790 32.634731                   45491
## 3  6.083832  6.185629                   45554
## 4 76.005988 52.485030                   45533
## 5  0.000000  0.000000                    <NA>
## 6 33.766467 10.664671                   45638
```

```
colnames(df)
```

```
## [1] "Name"                    "City"
## [3] "State"                   "Level.A"
## [5] "Level.B"                 "Level.C"
## [7] "Level.D"                 "Last.Inspection.End.Date"
```

```
dim(df)
```

```
## [1] 134   8
```

3.2 Handling Special characters in Name

```
df$Name<- iconv(df$Name,from = "",to = "UTF-8", sub = "")
df$Name <- gsub("[^A-Za-z ]", "", df$Name)#this ensures only A to Z and a to z character is in the data
head(df$Name)
```

```
## [1] "ADAMS COUNTY DET CENTER"            "ADELANTO ICE PROCESSING CENTER"
## [3] "ALAMANCE COUNTY DETENTION FACILITY" "ALEXANDRIA STAGING FACILITY"
## [5] "ALLEGANY COUNTY JAIL"               "ALLEN PARISH PUBLIC SAFETY COMPLEX"
```

3.3 Handling Blanks in Name,City and State

```r
# Name
df <- df %>% mutate(Name = if_else(Name == "", "Unknown", Name))

# City
df$City <- trimws(df$City)
df <- df %>%
  mutate(City = if_else(Name == "GEAUGA COUNTY JAIL" & City == "", "CHARDON", City))

# State
df$State <- trimws(df$State)
df <- df %>%
  mutate(State = if_else(City == "ATLANTA", "GA", State),
         State = if_else(City == "ENCINAL", "TX", State))  # <-- Corrected to TX
```

3.4 Handling Dates

```r
df <- df %>%
  mutate(Last.Inspection.End.Date = as.Date(as.numeric(Last.Inspection.End.Date), origin = "1899-12-30"
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `Last.Inspection.End.Date =
##   as.Date(as.numeric(Last.Inspection.End.Date), origin = "1899-12-30")`.
## Caused by warning in `as.Date()`:
## ! NAs introduced by coercion
```

```r
# Fix Nevada separately
df <- df %>%
  mutate(Last.Inspection.End.Date = if_else(
    Name == "NEVADA SOUTHERN DETENTION CENTER",
    as.Date("2024-09-19"),
    Last.Inspection.End.Date
  ))
```

4.Analyze the Data

4.1 Create Total Population Column

```r
df <- df %>%
  mutate(Total.Population = Level.A + Level.B + Level.C + Level.D)
```

4.2 Top 10 Facilities by Population

```r
top_10_facilities <- df %>%
  arrange(desc(Total.Population)) %>%
  slice_head(n = 10)
```

4.3 Top 10 States by Number of Detention Centers

```
Each_state_detention_center <- df %>%
  group_by(State) %>%
  summarise(Detention_center_count = n()) %>%
  arrange(desc(Detention_center_count)) %>%
  slice_head(n = 10)
```

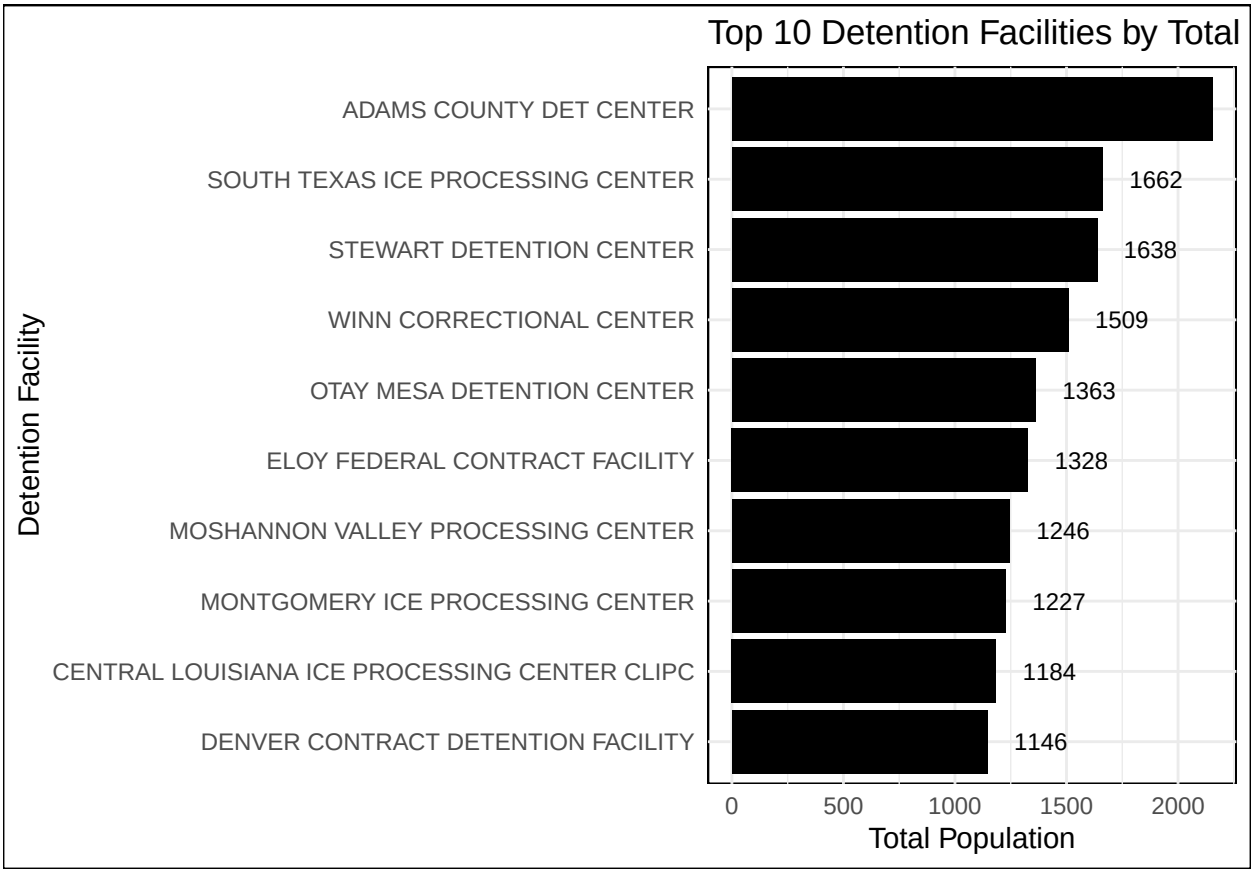
4.4 Top 10 States By Population in Detention Centers

```
State_total_population <- df %>%
  group_by(State) %>%
  summarise(Total = sum(Total.Population, na.rm = TRUE)) %>%
  arrange(desc(Total)) %>%
  slice_head(n = 10)
```

5. Visualization

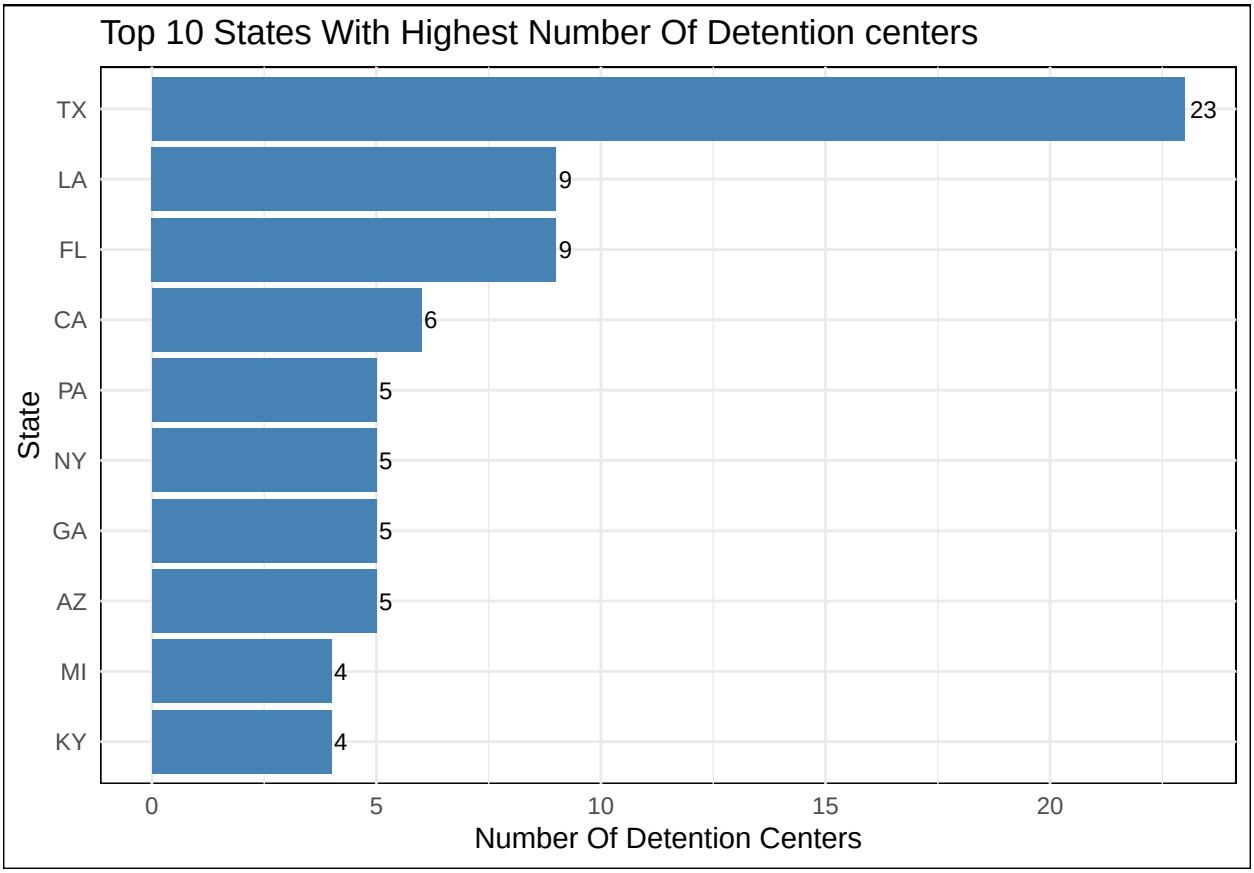
5.1 Top 10 Largest Detention Facilities by Population

```
ggplot(data = top_10_facilities, mapping =  aes(x = reorder(Name, Total.Population), y = Total.Populati
  geom_bar(stat = "identity", fill = "black") +
  geom_text(aes(label = round(Total.Population)), hjust = -0.5, size = 3) +
  coord_flip() +
  labs(title = "Top 10 Detention Facilities by Total Population",
       x = "Detention Facility",
       y = "Total Population") +
  theme_minimal() +
  theme(
    panel.background = element_rect(fill = "white"),#while saving image i wasn't getting clear plotting
    plot.background = element_rect(fill = "white")
  )
```



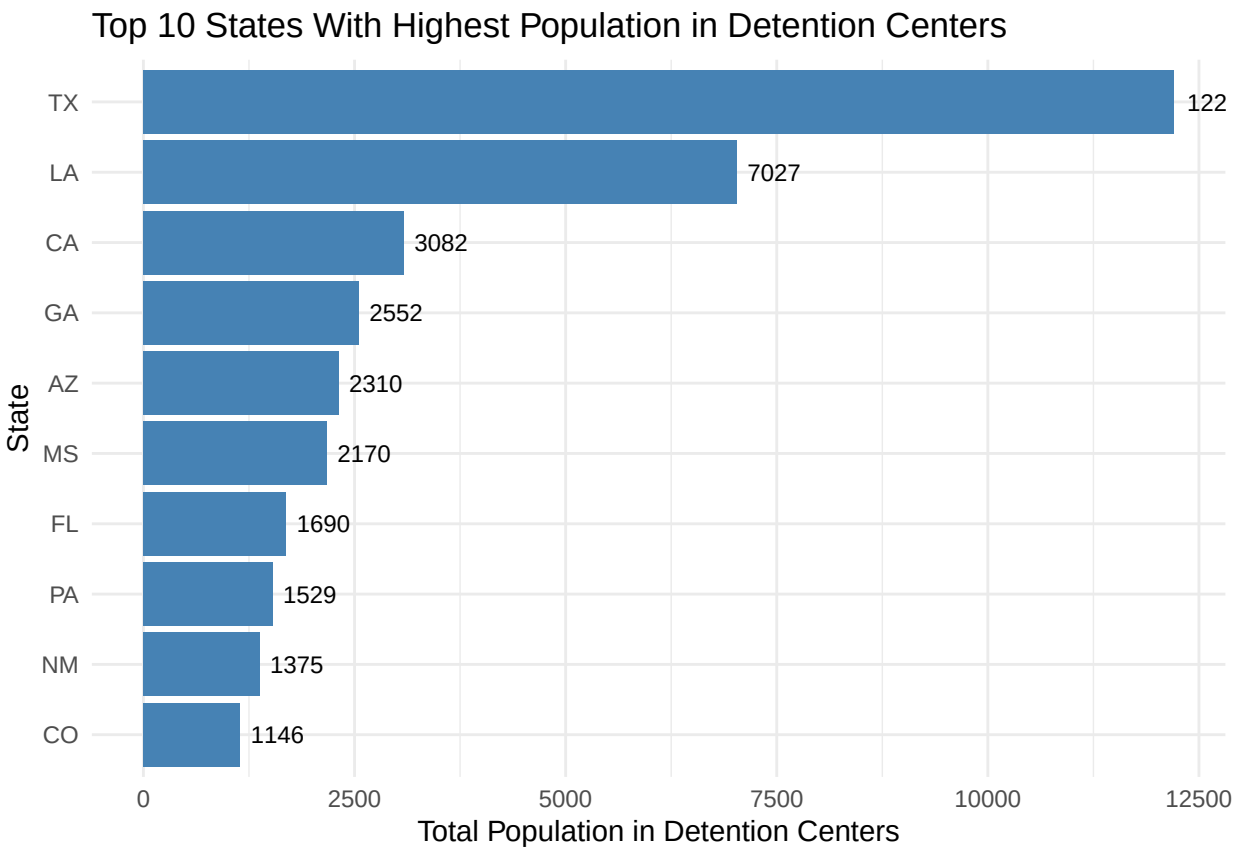
5.2 Top 10 States by Number of Detention Centers

```
ggplot(data = Each_state_detention_center,mapping =  aes(x = reorder(State, Detention_center_count), y =
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = Detention_center_count), hjust = -0.2, size = 3) + #adding text on each bar
  coord_flip() + #it flips the axes as when names were not clear.
  labs(title = "Top 10 States With Highest Number Of Detention centers",
       x = "State",
       y = "Number Of Detention Centers") +
  theme_minimal()+
  theme_minimal() +
  theme(
    panel.background = element_rect(fill = "white"),#while saving image i wasn't getting clear plotting
    plot.background = element_rect(fill = "white")
  )
```



5.3 Top 10 States by Population in Detention Centers

```
ggplot(data = State_total_population, aes(x = reorder(State, Total), y = Total)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  geom_text(aes(label = round(Total)), hjust = -0.2, size = 3) +
  coord_flip() +
  labs(title = "Top 10 States With Highest Population in Detention Centers",
       x = "State",
       y = "Total Population in Detention Centers") +
  theme_minimal()
```


Top 10 States With Highest Population in Detention Centers
State
TX
LA
CA
GA
AZ
MS
FL
PA
NM
CO
122
7027
3082
2552
2310
2170
1690
1529
1375
1146
0
2500
5000
7500
10000
12500
Total Population in Detention Centers