

MISSING PEOPLE



DETECTIVES SHARVARI, SUMAIYA, AND YUZE

Data Analysis

CONTENT

01. Research Question
02. Introduction to Dataset
03. Data wrangling
04. EDA
05. Data visualizations
06. References

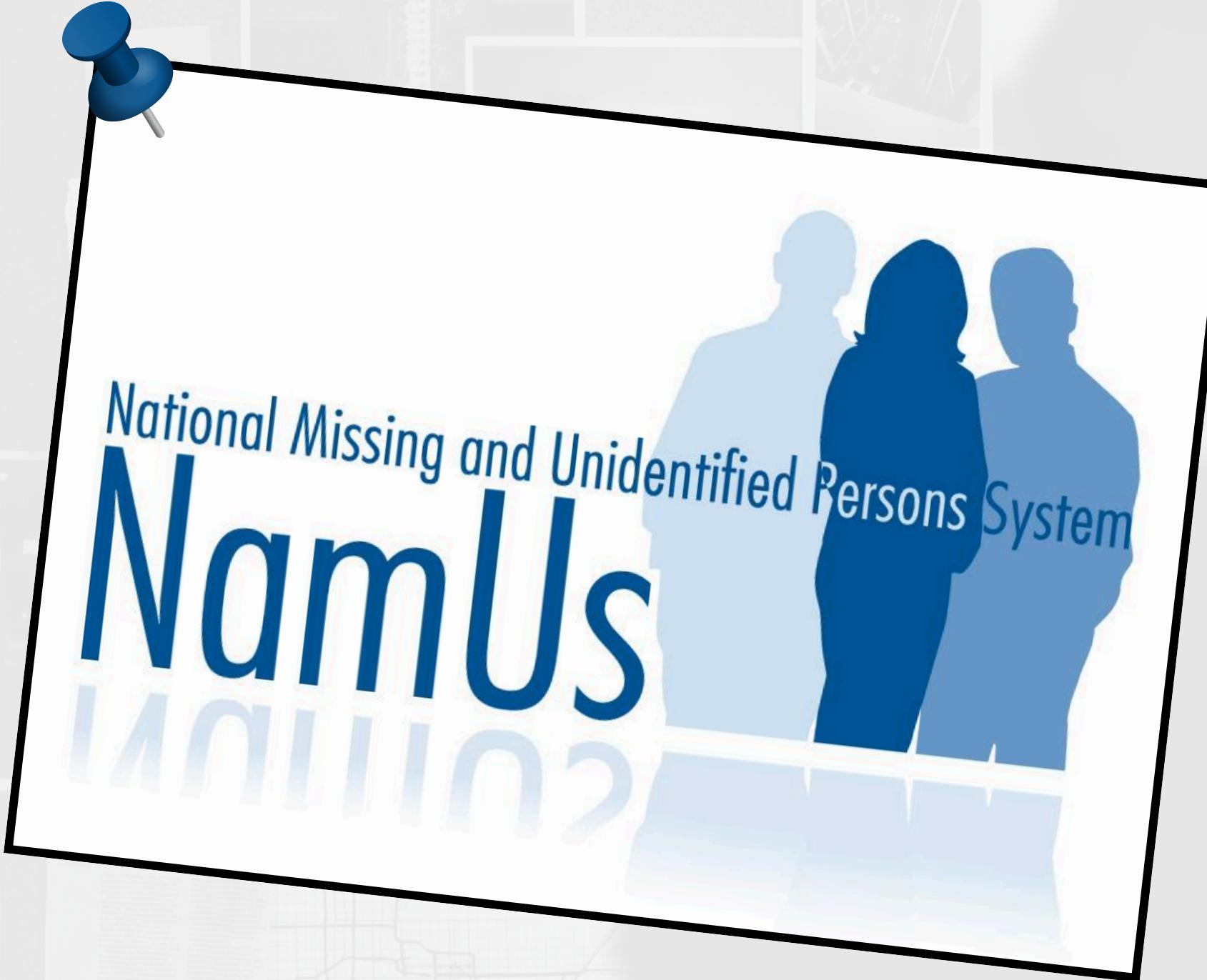


RESEARCH QUESTION

Do missing persons cases in Pennsylvania and Ohio from 1969 to 2024 exhibit overrepresentation across specific demographic groups, timespans and spatial contexts?

DATASET INVESTIGATED

- Maintained by the U.S. Department of Justice, ensuring **data credibility and reliability**
- Uses standardized reporting, allowing **consistent comparison** across states and counties
- Includes detailed **demographic, temporal, and geographic** variables
- Publicly accessible, making the dataset **transparent and reproducible** for analysis



OBTAINING DATA

- No web-scraping involved
- Register an account on NamUS, filter by tags, and download csv
- Dataset is tidy where each individual missing person is a case



Case Number	IDLC	Legal Last Name	Legal First Name	Missing Age	City	County	State	Biological Sex	Race / Ethnicity
MP2620	8/8/2004	Vega	Antonio	73 Years	Philadelphia	Adams	PA	Male	White / Caucasian
MP2905	6/27/97	Fawcett	Michele	35 Years	Donegal	Somerset	PA	Female	White / Caucasian
MP2508	12/12/1971	Lande	Elizabeth	21 Years	Philadelphia	Philadelphia	PA	Female	White / Caucasian
MP2477	9/27/74	Jones	Katherine	29 Years	Clearfield	Clearfield	PA	Female	White / Caucasian
MP2659	3/18/65	Shea	Kathleen	6 Years	Tyrone	Blair	PA	Female	White / Caucasian
MP2509	6/24/75	Thorne	Edna	15 Years	Philadelphia	Philadelphia	PA	Female	White / Caucasian
MP1561	9/8/2004	Morales	Tayna	3 Years	Reading	Berks	PA	Female	White / Caucasian
MP2496	12/19/78	Coyle	Andrea	30 Years	Altoona	Blair	PA	Female	White / Caucasian
MP2503	9/2/1983	Manley	Myra	19 Years	Philadelphia	Philadelphia	PA	Female	White / Caucasian
MP2908	6/7/1984	Mackerley	Louis	7 Years	Allentown	Lehigh	PA	Female	White / Caucasian
MP2945	3/5/2005	Westover	Ronald	63 Years	Westover	Clearfield	PA	Male	White / Caucasian
MP808	2/11/2004	Thomas	Phylicia	22 Years	Sweet Valley	Luzerne	PA	Male	White / Caucasian
MP784	11/11/1985	Carbaugh	Emerson	64 Years	Robertsdale	Huntingdon	PA	Female	White / Caucasian
MP1737	3/1/1979	Flickinger	Nellie	29 Years	Erie	Erie	PA	Male	White / Caucasian
MP806	2/19/05	Imbo	Danielle	34 Years	Philadelphia	Philadelphia	PA	Female	White / Caucasian
MP847	4/11/2006	Jackson	Lonnett	46 Years	Pittsburgh	Allegheny	PA	Female	White / Caucasian
MP2983	6/22/07	Benjamin	Michelle	35 Years	Saint Marys	Elk	PA	Female	White / Caucasian
MP4608	9/19/58	Triska	Rebecca	15 Years	Ambridge	Beaver	PA	Female	Black / African Amer
MP11745	2/4/2011	Miller	Maria	34 Years	Towanda	Bradford	PA	Female	White / Caucasian
MP12519	9/26/11	Peterson	Jamie	30 Years	Pittsburgh	Pittsburgh	PA	Female	White / Caucasian
MP6968	12/3/1999	Mohn	Kathleen	48 Years	Allegheny	Allegheny	PA	Female	Asian
MP7046	5/5/2001	Scott	Diane	39 Years	King of Prussia	Montgomery	PA	Female	White / Caucasian
MP7220	1/1/1988	Lango	John	17 Years	Harrisburg	Dauphin	PA	Female	White / Caucasian
MP7238	12/21/94	Allison	Robert	42 Years	Pottsville	Schuylkill	PA	Female	White / Caucasian
MP6764	9/25/95	Vanderhorst	Ke'Shaun	2 Years	Waynesburg	Greene	PA	Male	Black / African Amer
MP7236	2/26/94	Robinson	William	50 Years	Philadelphia	Philadelphia	PA	Male	White / Caucasian
MP14759	9/24/99	Procopio	Maria	34 Years	Philadelphia	Philadelphia	PA	Male	White / Caucasian
MP14787	1/31/06	Smatlak	Donald	25 Years	Philadelphia	Philadelphia	PA	Male	Black / African Amer
MP8249	2/20/03	Scott	Anita	58 Years	North Versailles	Allegheny	PA	Female	Black / African Amer
MP8588	4/23/85	Green	Christine	16 Years	Upper Providence T	Montgomery	PA	Male	White / Caucasian
MP10083	9/6/2006	Leonetti	Kristin	23 Years	Philadelphia	Philadelphia	PA	Female	White / Caucasian
MP9312	10/28/09	Ariza	Joel	24 Years	Glenolden	Delaware	PA	Female	White / Caucasian
MP8897	9/29/73	Hopter	Sandra	18 Years	Chester	Chester	PA	Female	Black / African Amer
MP19148	1/1/1990	Conrad	Mary	51 Years	Scranton	Lackawanna	PA	Male	White / Caucasian
MP19363	11/28/83	Stalker	Doris	54 Years	Plymouth	Luzerne	PA	Female	White / Caucasian
					Emlenton	Venango	PA	Female	White / Caucasian

Race / Ethnicity	Biological Sex	age_group	Number of Missing
Asian	Female	above 60	
Asian	Female	below 20	
Asian	Female	between 20-40	
Asian	Female	between 40-60	
Asian	Male	above 60	
Asian	Male	below 20	
Asian	Male	between 20-40	
Asian	Male	between 40-60	
Black / African American	Female	above 60	
Black / African American	Female	below 20	
Black / African American	Female	between 20-40	
Black / African American	Female	between 40-60	
Black / African American	Male	above 60	
Black / African American	Male	below 20	
Black / African American	Male	between 20-40	
Black / African American	Male	between 40-60	
Black / African American	Male	below 20	
Black / African American	Male	between 20-40	
Black / African American	Male	between 40-60	
Black / African American, Asian	Female	above 60	
Black / African American, Asian	Female	below 20	
Black / African American, Asian	Female	between 20-40	
Black / African American, Asian	Female	between 40-60	
Black / African American, Asian	Male	above 60	

```
missing_table_1 <- MissingPersons %>%
  mutate(
    age_num = readr::parse_number(`Missing Age`),
    age_num = if_else(is.na(age_num), 0, age_num),
    age_group = case_when(
      age_num < 20 ~ "below 20",
      age_num < 40 ~ "between 20-40",
      age_num < 60 ~ "between 40-60",
      TRUE ~ "above 60"
    )
  ) %>%
  count(
    `Race / Ethnicity`,
    `Biological Sex`,
    age_group,
    name = "Number of Missing People") %>%
  complete(
    `Race / Ethnicity`,
    `Biological Sex`,
    age_group,
    fill = list("Number of Missing People" = 0))
```

State	County	City	Number of Missing People
OH	Allen	Cridersville	1
OH	Allen	Lima	1
OH	Ashland	Ashland	1
OH	Ashtabula	Andover	1
OH	Ashtabula	Ashtabula	1
OH	Ashtabula	Cleveland	1
OH	Ashtabula	Jefferson	2
OH	Ashtabula	Orwell	1
OH	Athens	Athens	1
OH	Belmont	Bellaire	1
OH	Belmont	Martins Ferry	1
OH	Belmont	Shadyside	1
OH	Butler	Hamilton	5
OH	Butler	Liberty Township	1
OH	Butler	Middletown	7

```
# 4) Sort by location
location_table <- MissingPersons %>%
  count(State, County, City, name = "Number of Missing People")
arrange(State, County, City)
```

DLC	Number of Missing People
1/1/02	1
1/1/03	1
1/1/05	2
1/1/08	
1/1/11	
1/1/17	
1/1/18	2

```
# 5) Sort by date of last contact
date_table <- MissingPersons %>%
  count(DLC, name = "Number of Missing People") %>%
  arrange(DLC)
```

DATA WRANGLING

1. Each case is defined by a combination of sex, race, and age group
2. Each case is organized by state, county and city
3. Each case is associated with a specific date of last contact

EDA



Spatial
Patterns

Demographic
Patterns

Temporal
Patterns



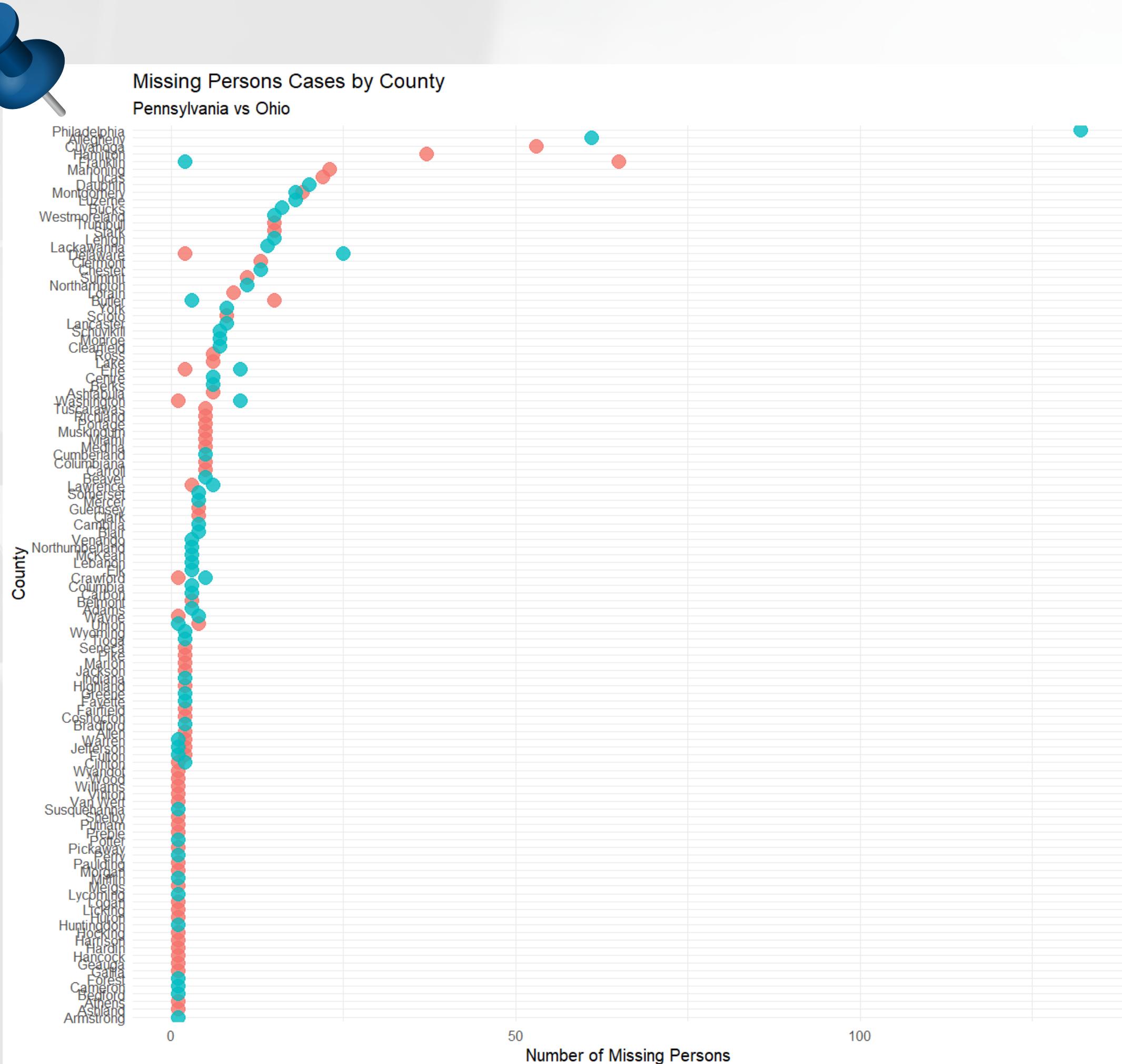
SPATIAL PATTERNS

- Bar Chart of Missing Persons cases by State
- Pennsylvania has more Missing Person Cases than Ohio



SPATIAL PATTERNS

- Dot-Plot of Missing Persons Cases by County
- Pennsylvania is more spread out while Ohio is more clustered
- Outlier: Philadelphia, suggesting great spatial overrepresentation

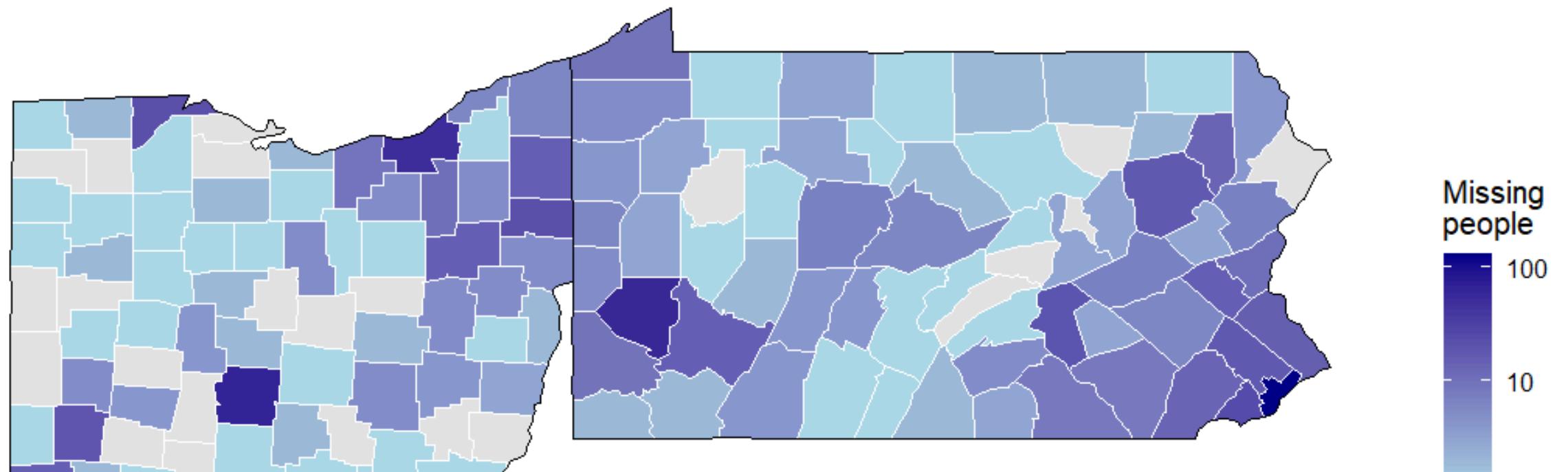


SPATIAL PATTERNS

- Generally uneven spatial distribution of Missing Person Cases across PA and OH
- Missing Person Cases concentrated in specific locations



Number of Missing Persons per County
Pennsylvania & Ohio



Location

Transfer

COUNTY MAP CODE

- County-level mapping code adapted from:
“Making Maps with R”
by E. Qande
(eriqande.github.io)

Qande, E. (n.d.). Making maps with R. Retrieved December 8, 2025, from <https://eriqande.github.io/rep-res-web/lectures/making-maps-with-R.html>

```
library(tidyverse)
library(maps)

#Filter by County
county_summary <- MissingPersons %>%
  filter(State %in% c("PA", "OH")) %>%
  group_by(State, County) %>%
  summarise(
    NumberMissing = n(),
    .groups = "drop"
  )

View(county_summary)

#Extract states and counties data
states   <- map_data("state")
counties <- map_data("county")

ohpa_states  <- states  %>% filter(region %in% c("pennsylvania", "ohio"))
ohpa_counties <- counties %>% filter(region %in% c("pennsylvania", "ohio"))

#Attach every point of polygon to missing persons numbers by county
county_summary_for_join <- county_summary %>%
  mutate(
    region   = recode(State,
                        "PA" = "pennsylvania",
                        "OH" = "ohio"),
    subregion = tolower(County)
  )

#Redefine data by region (state) and subregion
ohpa_map_df <- ohpa_counties %>%
  left_join(county_summary_for_join,
            by = c("region", "subregion"))

#Plot missing persons data by county
ditch_the_axes <- theme(
  axis.text = element_blank(),
  axis.line = element_blank(),
  axis.ticks = element_blank(),
  panel.border = element_blank(),
  panel.grid = element_blank(),
  axis.title = element_blank()
)

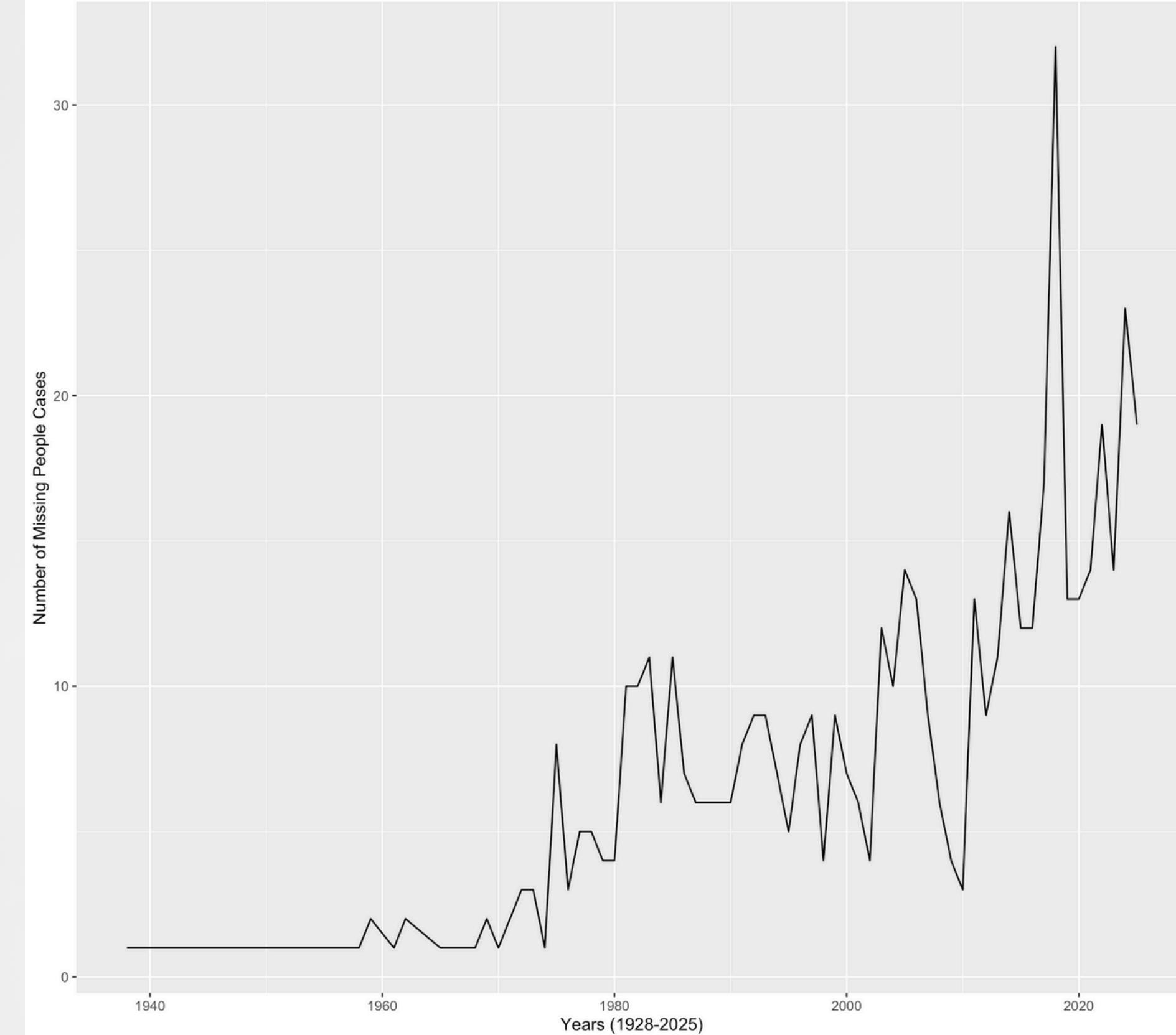
ohpa_base <- ggplot(ohpa_states,
                     aes(x = long, y = lat, group = group)) +
  coord_fixed(1.3) +
  geom_polygon(color = "black", fill = "gray80")

elbow_room <- ohpa_base +
  geom_polygon(
    data = ohpa_map_df,
    aes(fill = NumberMissing),
    color = "white"
  ) +
  geom_polygon(color = "black", fill = NA) +
  theme_bw() +
  ditch_the_axes

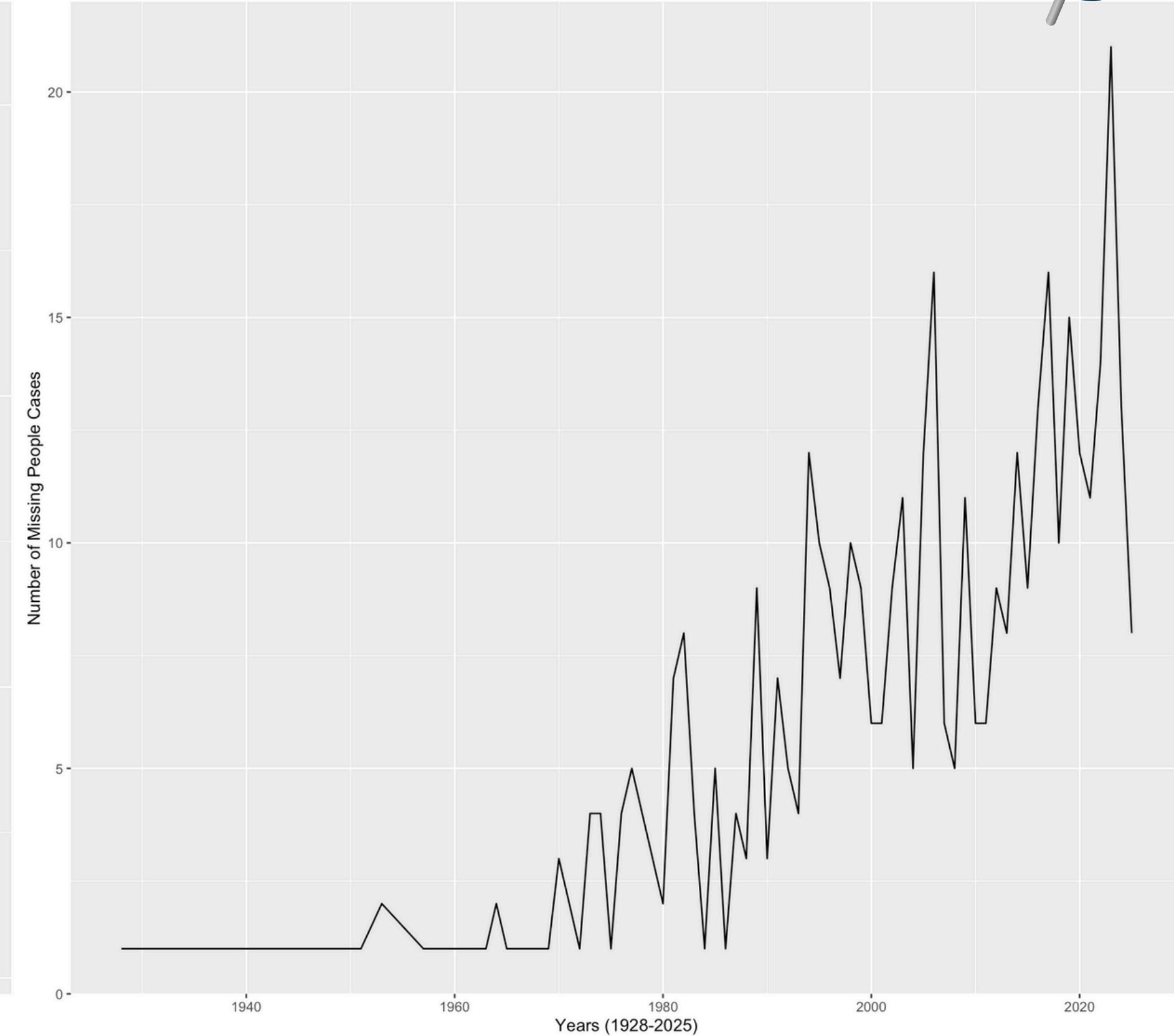
elbow_room +
  scale_fill_gradient(
    trans = "log10",
    low = "lightblue",
    high = "darkblue",
    na.value = "grey90"
  ) +
  labs(
    title = "Number of Missing Persons per County\nPennsylvania & Ohio",
    fill = "Missing\npeople"
  )
```

MISSING COUNTS OVER TIME

Missing People Per Year In PA



Missing People Per Year In OH

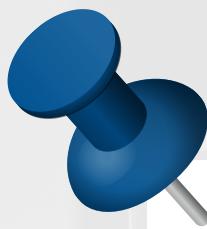


TIME SERIES CODE

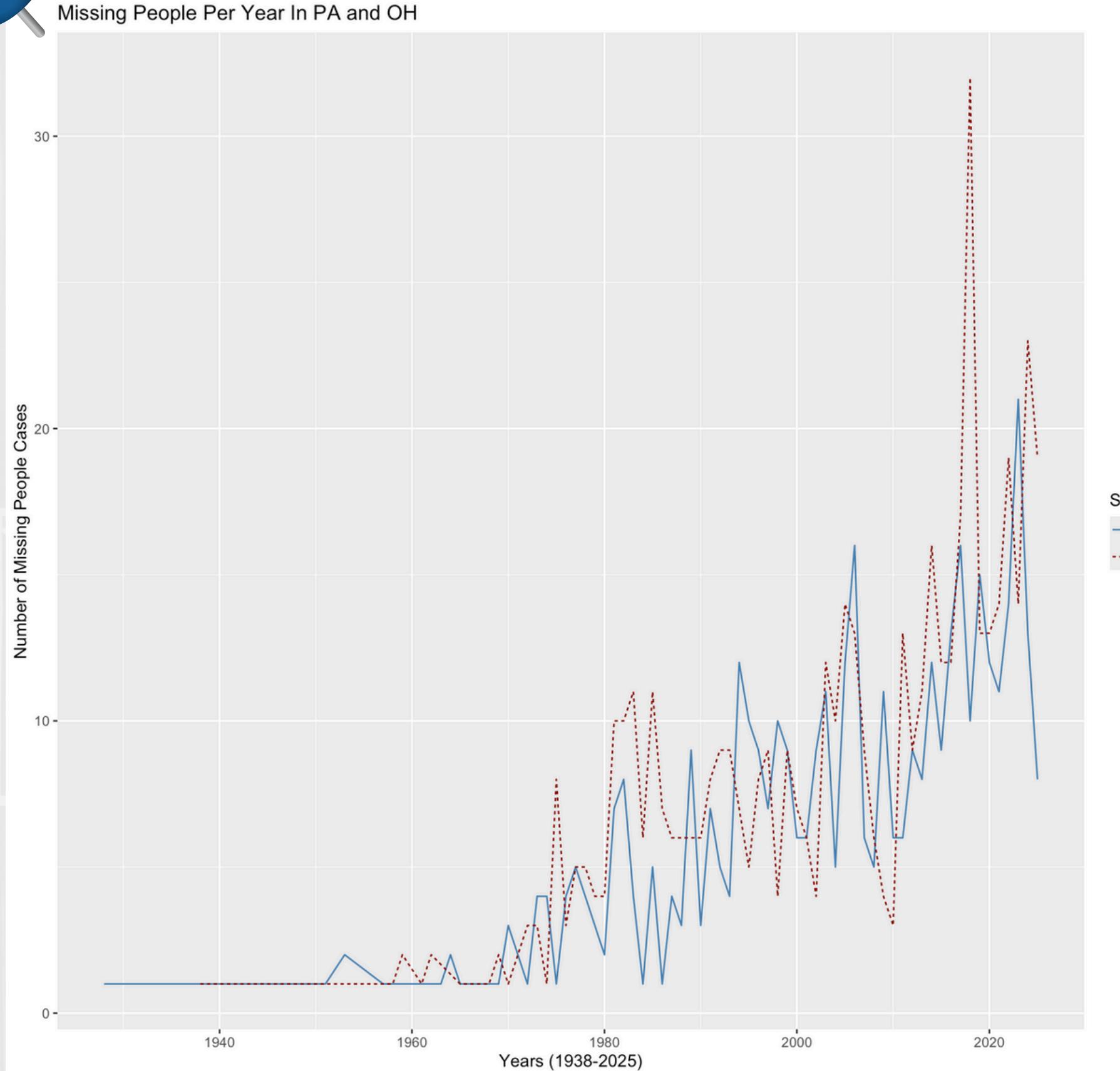
```
1 # Data wrangling ----  
2 ## Creating a time sensitive data frame via dates  
3 time_sensitive_missing <- missingPersons |>  
4   mutate(date = as.Date(DLC, format = "%m/%d/%y")) |>  
5 ## Summarize the data frame where each case is a year and an attribute is missing count  
6 ### cut() rounds every date down to year-01-01; all dates in a year will have that label  
7   mutate(year = as.Date(cut(date, breaks = "year"))) |>  
8   group_by(year, State) |>  
9   summarize(missing_count = n(),  
10             .groups = "drop")  
11 ## Make sure there's no data from the future  
12 time_sensitive_missing$year <- futureDeleteer(time_sensitive_missing$year)  
13  
14 ## Pythonian caveman function that solves problems with as.Date()  
15 ### Function takes parameter dates (column of a data frame)  
16 futureDeleteer <- function(dates) {  
17   strings = as.character(dates)  
18   for (i in seq_along(strings)) {  
19     d = strings[i]  
20     if (substr(d, 1, 2) == "20" && as.numeric(substr(d, 3, 4)) > 25) {  
21       new_str = paste0("19", substr(d, 3, 10))  
22       strings[i] = new_str  
23     }  
24   }  
25   return(as.Date(strings))  
26 }
```

R Documentation. (2019). `as.Date` [R Documentation].

The original form of the DLC column was in the form of strings of characters, therefore, changing the data type to something workable in a graph was a small challenge.



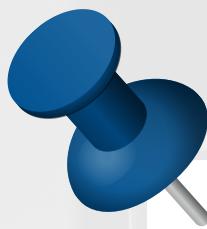
Missing People Per Year In PA and OH



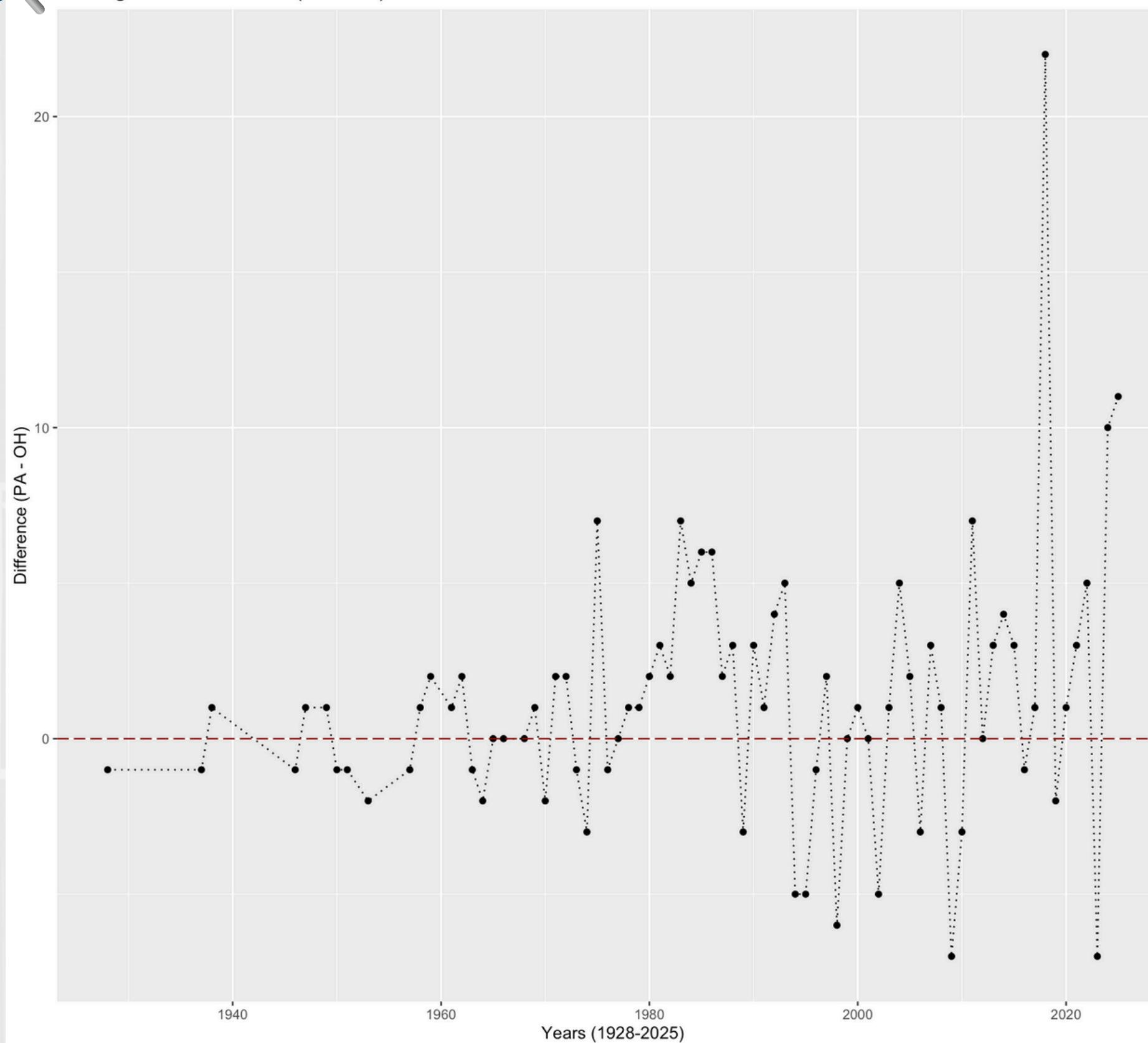
SIDE BY SIDE COMPARISON

It seems that the upward trend fits both states pretty well.

To go further examine the difference in growth of the two states, we should take the difference between them.



Missing Cases Difference (PA - OH)



```
> mean(difference_time_series$difference)
```

```
[1] 1.197368
```

DIFFERENCE IN MISSING COUNTS

There is a “fanning-out” effect as we view over time of the difference between the two.

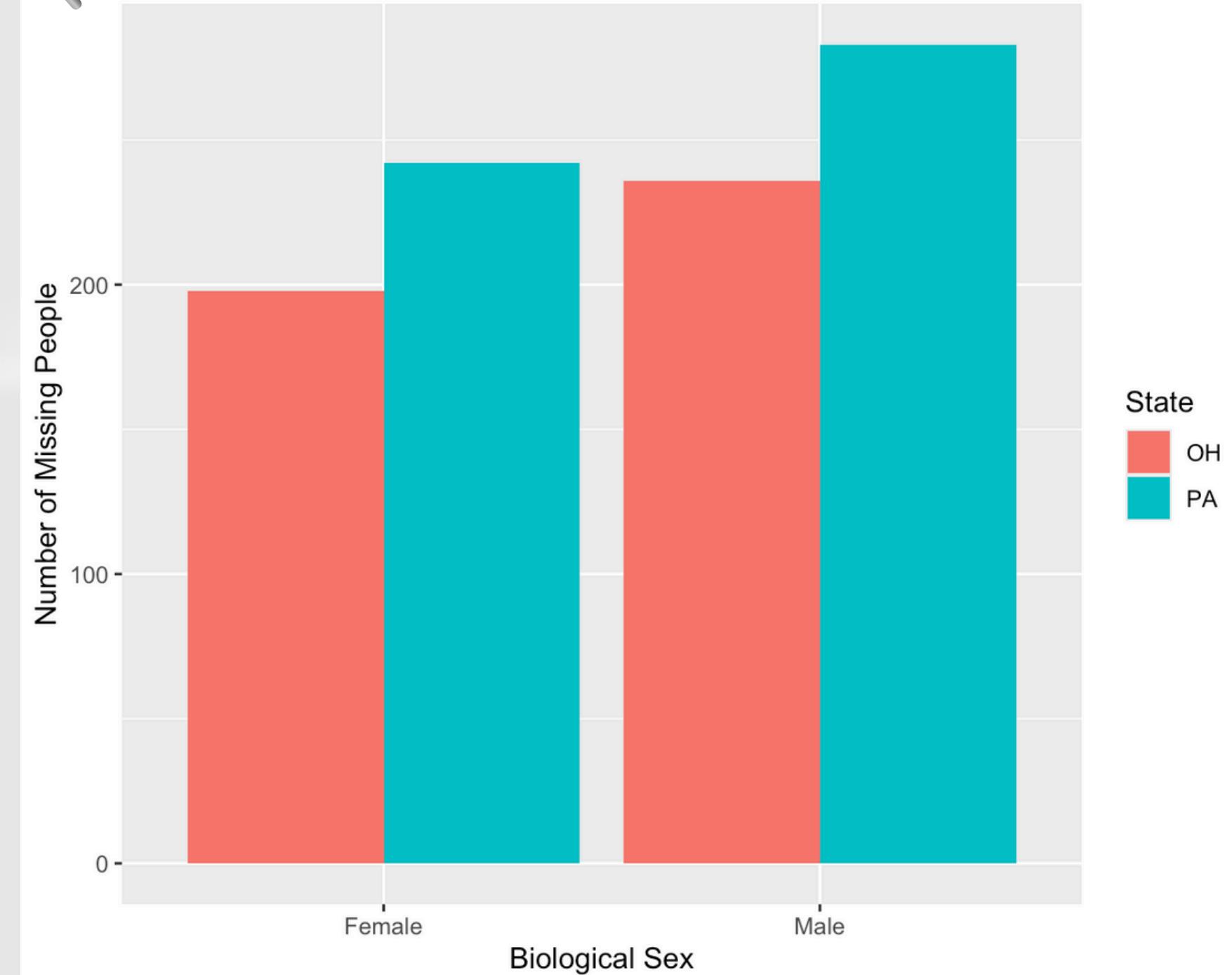
On average, the mean difference (mean value of all data points) is around 1.1974. Meaning that PA on average has a slightly higher missing count.

DEMOGRAPHIC PATTERNS

- Both PA and OH show a somewhat even split between male and female cases.
- Males are slightly higher in both states, but not by much.
- Overall counts are a bit higher in Pennsylvania.

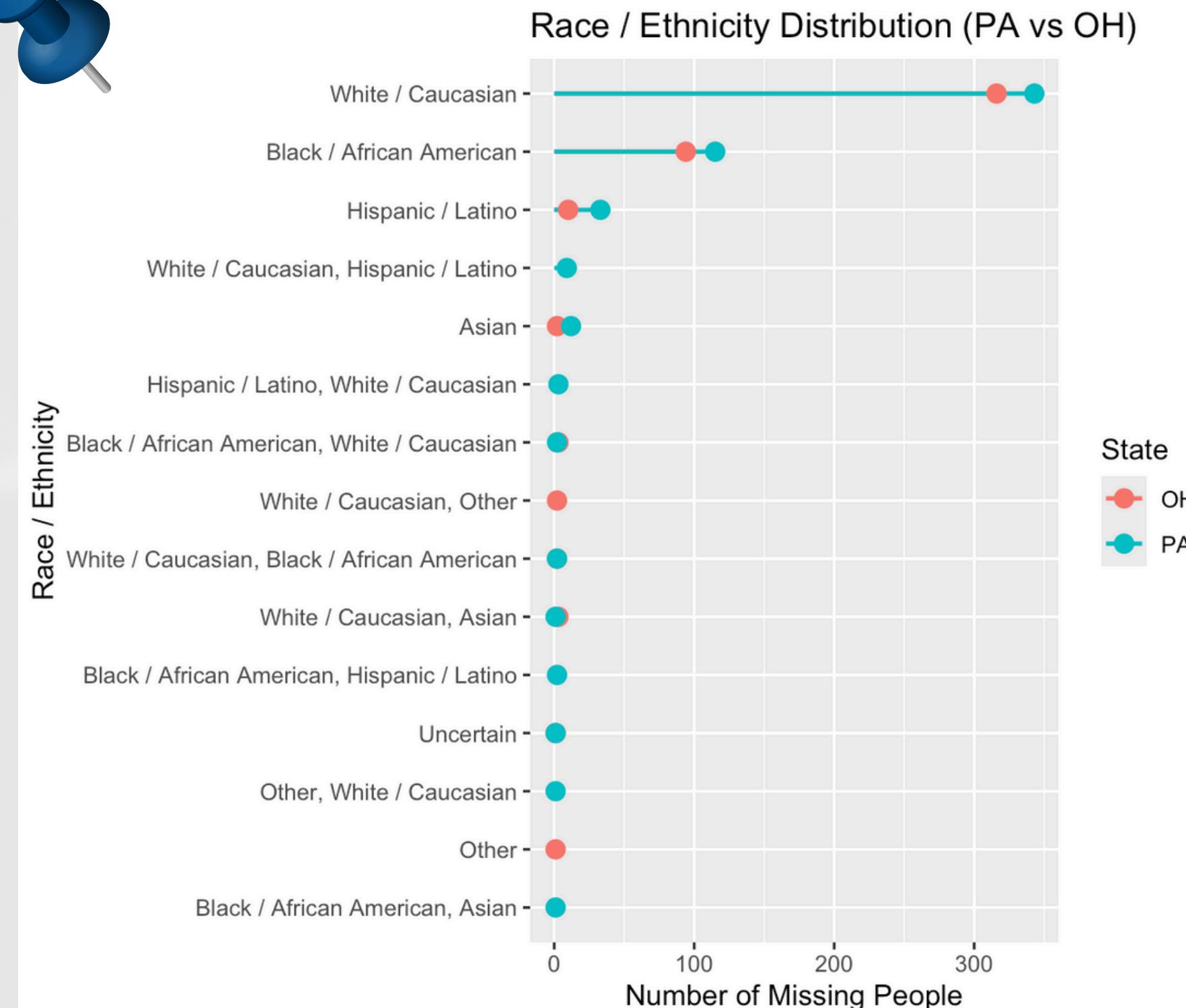


Biological Sex Distribution (PA vs OH)



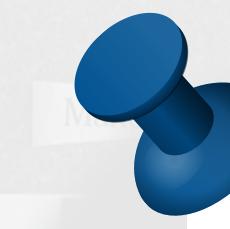
DEMOGRAPHIC PATTERNS

- White/Caucasian individuals make up the majority of missing-person cases in both PA and OH.
- Black/African American cases are the next largest group.
- All other racial categories appear in much smaller, scattered numbers.
- The pattern is basically the same across both states.

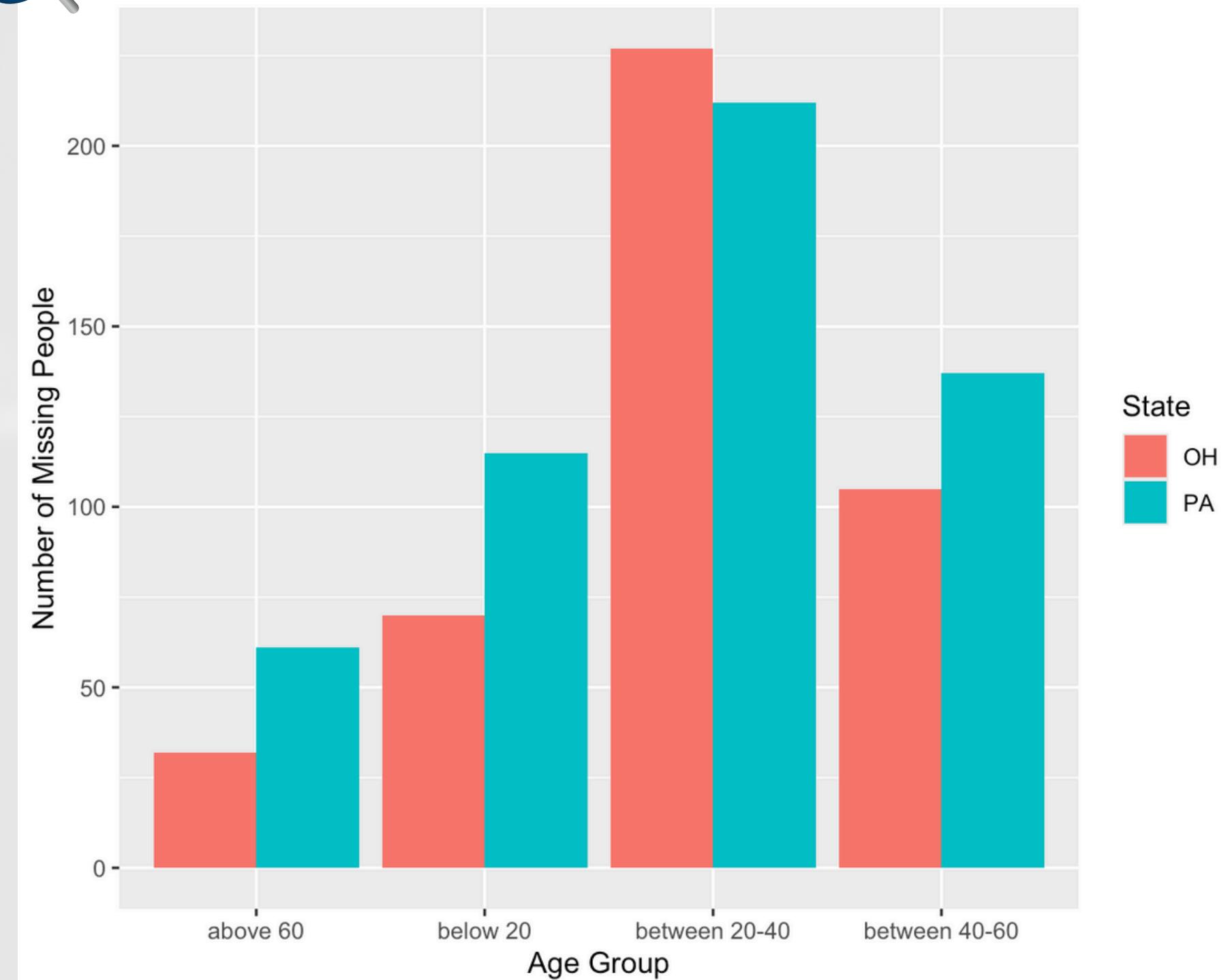


DEMOGRAPHIC PATTERNS

- Most missing-person cases fall in the 20–40 age group.
- The 40–60 group is the second largest.
- Under 20 and above 60 have noticeably fewer cases.
- PA tends to have slightly higher counts overall, but both states follow the same trend.



Age Distribution of Missing People (PA vs OH)



CONCLUSIONS

• SPATIAL AND TEMPORAL

- In the scale of states, since the counts aren't that high, there are lots of variation from year to year.
- There is a general upwards trend of missing peoples count that might seem disproportional to population growth.
- For the case of PA and OH, there aren't lots of difference over time.

• DEMOGRAPHIC GROUPS

```
### find the single demographic group with the highest missing count
missing_table_1 %>%
  arrange(desc(`Number of Missing People`)) %>%
  slice(1)
```

```
# A tibble: 1 × 4
  `Race / Ethnicity` `Biological Sex` age_group `Number of Missing People`
  <chr>                <chr>           <chr>            <int>
1 White / Caucasian   Female          between 20-40      160
> |
```

OTHER TAKEAWAYS

• IMPROVEMENTS

- Make charts more color-blind friendly and readable
- Normalize the dataset against overall state demographic populations
- Comparing the state-level data with national data

• FUTURE POTENTIAL

- Looking into city based comparisons for spatial analysis
- Looking into seasonal patterns for temporal analysis
- Comparisons with other variables (National Crime Index, Global Night Light index, etc) for checking correlations



DETECTIVES SHARVARI, SUMAIYA, AND YUZE

THANK YOU

3?

Don't try
to find help.
I'll come
for you anyway.
Just wait.
I'll come very
soon!
B.B.

Manch

Death at sea and a

Death at sea and a

My dear friend!
Here are some nice ideas.
Let it be so! the truth then the
For by the sacred audience of the time,
the Christians of Heath and the night,
by all the species of life
How we can do what we can do
I attach all my personal care
Myself and my wife and we
We have had this. To ease the burdens
Our mother has a lot more
some for a while all of us know
as well as our child and the others
than my mother's daughter

See you soon
B.B.

98C

U.S. troops in northeast Syria

4

'I cannot remain silent'