

Visualizing FEMA NRS and ACS Data

Richard Montes Lemus

2026-02-12

Load in Necessary Libraries

```
library(tidycensus)
library(tidyverse)
library(ggtext)
library(here)
library(janitor)
library(ggplot2)
```

Read in US Census Bureau Data

```
# #....Step 1a: see all available ACS variables + descriptions.....
# acs_vars <- tidycensus::load_variables(year = 2023,
#                                     dataset = "acs1")
#
# #.....Step 1b: import race & ethnicity data.....
# race_ethnicity <- tidycensus::get_acs(
#   geography = "county",
#   survey = "acs1",
#   # NOTE: you may not end up using all these variables
#   variables = c("B01003_001", "B02001_002", "B02001_003",
#                 "B02001_004", "B02001_005", "B02001_006",
#                 "B02001_007", "B02001_008", "B03002_012",
#                 "B03002_002"),
#   state = "CA",
#   year = 2023) |>
#   # join variable descriptions (so we know what's what!)
```

```
# dplyr::left_join(acs_vars, by = dplyr::join_by(variable == name))
#
# #.....Step 2: write ACS data to file.....
# readr::write_csv(race_ethnicity, here::here("data",
# "ACS-1yr-2023-county-race-ethnicity.csv"))
#.....Step 3: read in your CSV file.....
race_ethnicity <- readr::read_csv(here::here("data",
"ACS-1yr-2023-county-race-ethnicity.csv"))
```

Read in NRI Data

```
# Filter out non-US state rows
nri <- read_csv(here("data", "National_Risk_Index_Counties.csv"))
```

Clean and Filter Data

```
cal_race_ethnicity <- race_ethnicity %>%
  clean_names() %>%
  # Use "," to separate "name" column into "county" and "state"
  tidyr::separate(col = name, into = c("county_name", "state"), sep = ",") %>%
  # Remove word "County" from column
  mutate(county_name = str_remove(county_name,
                                   regex("County", ignore_case = TRUE))) %>%
  str_squish() %>% # Remove white space
  select(county_name, estimate, label, variable)
```

```
cal_nri <- nri %>%
  clean_names() %>%
  # Filter NRI data to California
  filter(state_name %in% c("California")) %>%
  select(county_name, national_risk_index_rating_composite)
```

Merge Dataframes

Upon reviewing these data sets I noticed cal_race_ethnicity from ACS is missing some counties. This makes sense, the ACS only collects data from larger counties year to year. It only collects

data from all larger and smaller counties every 5 years. This data set comes from the year 2023. Since this analysis is focused on risk information pertaining to each race/ethnicity, it makes sense to left join NRI (containing all counties) to ACS data. This ensures we avoid adding empty rows to our merged data. Also, it is important to note that the “Very Low” category does not exist in any rows remaining after the merge.

```
# Merge ACS and NRI data
nri_race_ethnicity <- cal_race_ethnicity %>%
  left_join(cal_nri)
```

Filter Out Unecesary Rows and Characters

```
# Create data frame with total population of each county as new column
total_pop <- nri_race_ethnicity %>%
  filter(variable == "B01003_001") %>%
  select(county_name, total_pop = estimate)
```

```
nri_race_ethnicity_pop <- nri_race_ethnicity %>%
  left_join(total_pop, by = "county_name") %>% # Join total population column
  filter(variable != "B01003_001") %>%
  # Remove total population, we already have a column for it
  mutate(label = label %>% # Remove special characters in label column
    str_remove("Estimate!!Total:!!") %>%
    str_remove_all(":") %>%
    str_remove("alone") %>%
    str_trim()) %>%
  filter(label != "Not Hispanic or Latino") # Exclude label fromr analysis
```

Calculate Summary Statistics

```
# Get ethnicity/race summary per risk category
pop_per_risk <- nri_race_ethnicity_pop %>%
  group_by(label, national_risk_index_rating_composite) %>%
  summarise(pop_estimate = sum(estimate)) %>%
  ungroup()
```

Plot Data

```
# Created stacked bar plot with percentage of group in each risk category
plot_data <- nri_race_ethnicity_pop %>%
  group_by(label) %>%
  mutate(total_in_group = sum(estimate),
# Calculate percentage of race/ethnicity per label
  percentage = (estimate / total_in_group) * 100) %>%
  mutate(national_risk_index_rating_composite = factor( # Order ratings
    national_risk_index_rating_composite,
    levels = c("Relatively Low",
               "Relatively Moderate",
               "Relatively High",
               "Very High"))) %>%
  mutate(label = factor(label, # Order Race/Ethnicity
    levels = c("Native Hawaiian and Other Pacific Islander",
               "White",
               "American Indian and Alaska Native",
               "Two or More Races",
               "Hispanic or Latino",
               "Some Other Race",
               "Black or African American",
               "Asian")))) %>%
# Remove white lines in stacked bar by grouping by label
  group_by(label, national_risk_index_rating_composite) %>%
  summarize(percentage = sum(percentage), .groups = "drop")

# Create stacked bar chart
ggplot(plot_data, aes(x = label,
  y = percentage,
  fill = national_risk_index_rating_composite)) +
  geom_col(position = "stack") +
  scale_fill_manual(values = c(
    "Very High" = "#d73027",
    "Relatively High" = "#fc8d59",
    "Relatively Moderate" = "#fee08b",
    "Relatively Low" = "#91cf60"
  )) +
  coord_flip() +
  labs(
    title = "California National Risk Index Rating by Race or Ethnicity",
    subtitle = "Asian communities have the highest percentage of their
```

```

population
in the Very High risk category",
x = NULL,
y = "Percentage of Population",
fill = "Risk Rating",
caption = "**Figure 1:** The following graph shows the percentage of each
race or ethnicity that falls into each FEMA risk rating category in
California. Race and ethnicity are ordered from largest to smallest
based on their FEMA risk rating in the Very High category. Data: FEMA
National Risk Index (2023 Release) and The United States Census'
American Community Survey (2023 Release)" +
# Order legend rating from highest to lowest
guides(fill = guide_legend(reverse = TRUE)) +
theme_minimal() +
theme(plot.title = element_text(size = 22,
                                face = "bold"),
      plot.caption = element_markdown(size = 18,
                                       hjust = 0,
                                       margin = margin(t = 15)),
      plot.subtitle = element_markdown(size = 15),
      axis.title = element_text(size = 17),
      axis.text = element_text(size = 15),
      legend.title = element_text(size = 16),
      legend.text = element_text(size = 15))

```

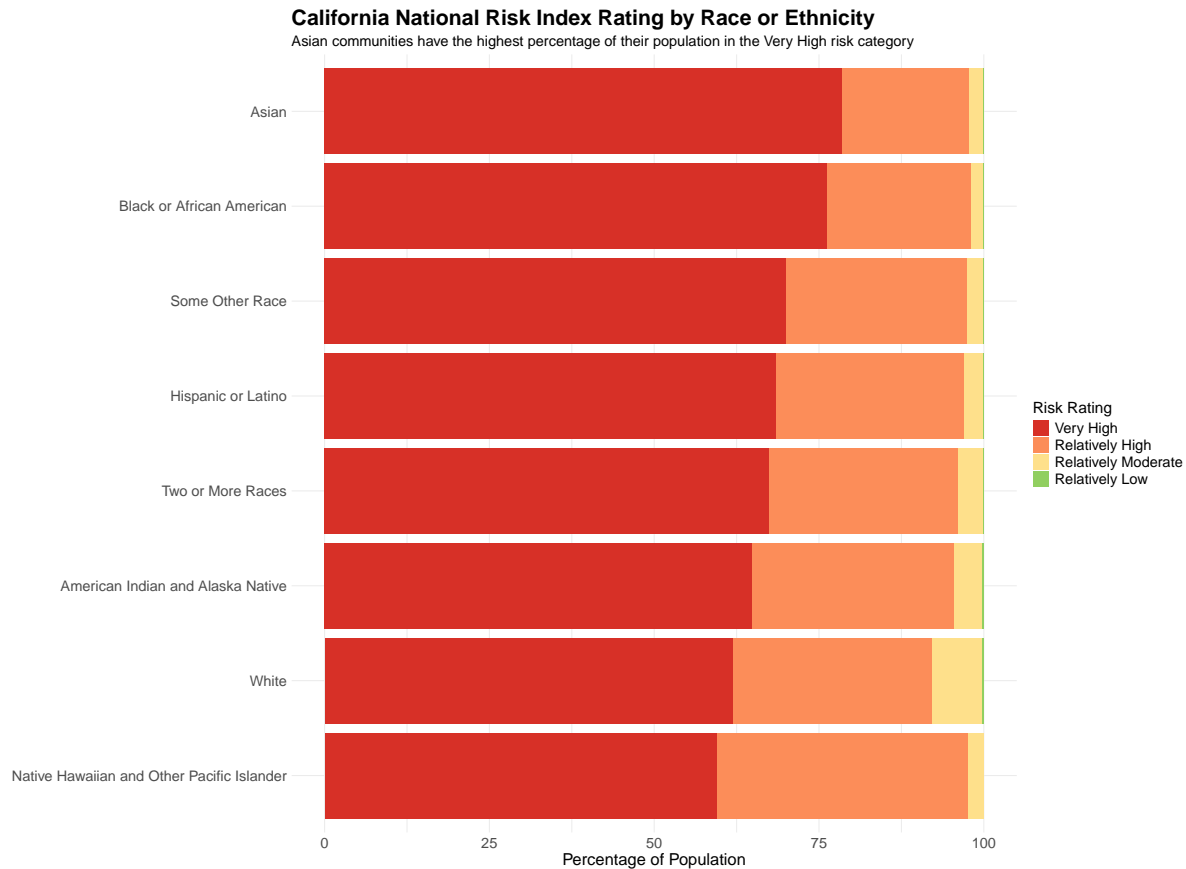


Figure 1: The following graph shows the percentage of each race or ethnicity that falls into each FEMA risk rating category in California. Race and ethnicity are ordered from largest to smallest based on their FEMA risk rating in the Very High category. Data: FEMA National Risk Index (2023 Release) and The United States Census' American Community Survey (2023 Release)

Discussion Questions

1. What are your variables of interest and what kinds of data (e.g. numeric, categorical, ordered, etc.) are they (a bullet point list is fine)?

- The first variable is “national_risk_index_rating_composite”, this is categorical and ordered.
- The second variable is “label” this is categorical.
- The third variable is “percentage”, this is numeric.

2. How did you decide which type of graphic form was best suited for answering the question? What alternative graphic forms could you have used instead? Why did you settle on this particular graphic form?

I decided which graphic form was best suited for answering the question by doing the following. First I wrote down my variables/columns of interest. Two of these columns were categorical and one was numeric. Next, I used this information to explore the data to viz website for plots containing data with several categorical variables and one numeric variable. This page gave me the idea to create a scatter plot containing groups colored by race/ethnicity or create a lollipop plot showing risk index percentile median per race or ethnicity. However, none of them seemed to communicate risk per group like a stacked bar plot. I settled on a stacked bar plot for answering my question because it allowed me to obtain normalized percentages contained to each racial/ethnic group. This meant I could easily compare risk across groups.

3. Summarize your main finding in no more than two sentences. In California, Asian communities have the highest percentage of their population in the “Very High” risk FEMA rating category.

4. What modifications did you make to this visualization to make it more easily readable? I factored the stacked bar plot order so it descended from highest to lowest percentage in the “Very High” risk rating category per race/ethnicity. Then, I also factored the legend so it was ordered from “Very High” to “Relatively Low”.

5. Is there anything you wanted to implement, but didn’t know how? If so, please describe. I don’t believe there is anything I wanted to implement but didn’t get the chance to.