

HW #3: Visualizing FEMA NRI Data

Emily Miller

2026-02-05

Introduction

This analysis explores FEMA National Risk Index (NRI) scores across US counties, comparing California counties to other states. The NRI measures community risk from natural hazards, considering exposure, annual loss, and social vulnerability.

Data loading

```
# Load necessary libraries
library(tidycensus)
library(tidyverse)
library(here)
library(janitor)
library(showtext)

# Step 1a: see all available ACS variables + descriptions
# acs_vars <- load_variables(year = 2023, dataset = "acs1")

# Step 1b: import race & ethnicity data from Census API
# race_ethnicity <- get_acs(
#   geography = "county",
#   survey = "acs1",
#   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) |>
#   left_join(acs_vars, by = join_by(variable == name))
```

```

# Step 2: write ACS data to file for offline use
# write_csv(race_ethnicity, here("data", "ACS-1yr-2023-county-race-ethnicity.csv"))

# Step 3: read in saved CSV file
race_ethnicity <- read_csv(here("data", "ACS-1yr-2023-county-race-ethnicity.csv"))

# Load NRI data
nri_data_raw <- read_csv(here("data", "National_Risk_Index_Counties_807384124455672111.csv"))

```

Data wrangling and merging

```

# Filter to California counties only and select relevant NRI columns
nri_data <- nri_data_raw %>%
  clean_names() %>%
  filter(state_name == "California") %>%
  select(state_county_fips_code,
         county_name,
         state_name,
         nri_score = national_risk_index_score_composite,
         risk_rating = national_risk_index_rating_composite) %>%
  rename(geoid = state_county_fips_code)

# Process ACS data: separate county/state and map variable codes to readable names
race_ethnicity_clean <- race_ethnicity %>%
  clean_names() %>%
  # Split combined "County, State" into separate columns
  separate(name, into = c("County", "State"), sep = ", ") %>%
  # Convert ACS variable codes to human-readable race/ethnicity labels
  mutate(race_eth_group = case_when(
    variable == "B03002_002" ~ "White",
    variable == "B02001_003" ~ "Black or African American",
    variable == "B02001_004" ~ "American Indian and Alaska Native",
    variable == "B02001_005" ~ "Asian",
    variable == "B02001_006" ~ "Native Hawaiian and Other Pacific Islander",
    variable == "B02001_007" ~ "Some Other Race",
    variable == "B02001_008" ~ "Two or More Races",
    variable == "B03002_012" ~ "Hispanic or Latino",
    TRUE ~ NA_character_
  )) %>%
  # Remove variables we're not analyzing

```

```

filter(!is.na(race_eth_group)) %>%
  select(geoid, County, State, race_eth_group, estimate)

# Join NRI risk ratings with ACS demographic data by FIPS code
merged_data <- race_ethnicity_clean %>%
  left_join(nri_data, by = "geoid") %>%
  # Convert risk rating to ordered factor for proper visualization ordering
  mutate(risk_rating = factor(risk_rating,
                               levels = c("Very Low", "Relatively Low",
                                         "Relatively Moderate", "Relatively High",
                                         "Very High"),
                               ordered = TRUE))

```

Visualization

```

# Load custom font for cleaner appearance
font_add_google("Raleway", "raleway")
showtext_auto()

# Aggregate population by race/ethnicity and risk level, then calculate percentages
viz_data <- merged_data %>%
  # Sum population for each group in each risk category
  group_by(race_eth_group, risk_rating) %>%
  summarize(total_pop = sum(estimate, na.rm = TRUE), .groups = "drop") %>%
  # Calculate what percent of each group lives in each risk category
  group_by(race_eth_group) %>%
  mutate(percent = total_pop / sum(total_pop)) %>%
  ungroup()

# Order racial/ethnic groups by their Very High risk exposure (lowest to highest)
group_order <- viz_data %>%
  filter(risk_rating == "Very High") %>%
  arrange(percent) %>%
  pull(race_eth_group)

# Create stacked bar chart
viz_data %>%
  # Apply ordering to y-axis
  mutate(race_eth_group = factor(race_eth_group, levels = group_order)) %>%
  ggplot(aes(x = percent, y = race_eth_group, fill = risk_rating)) +

```

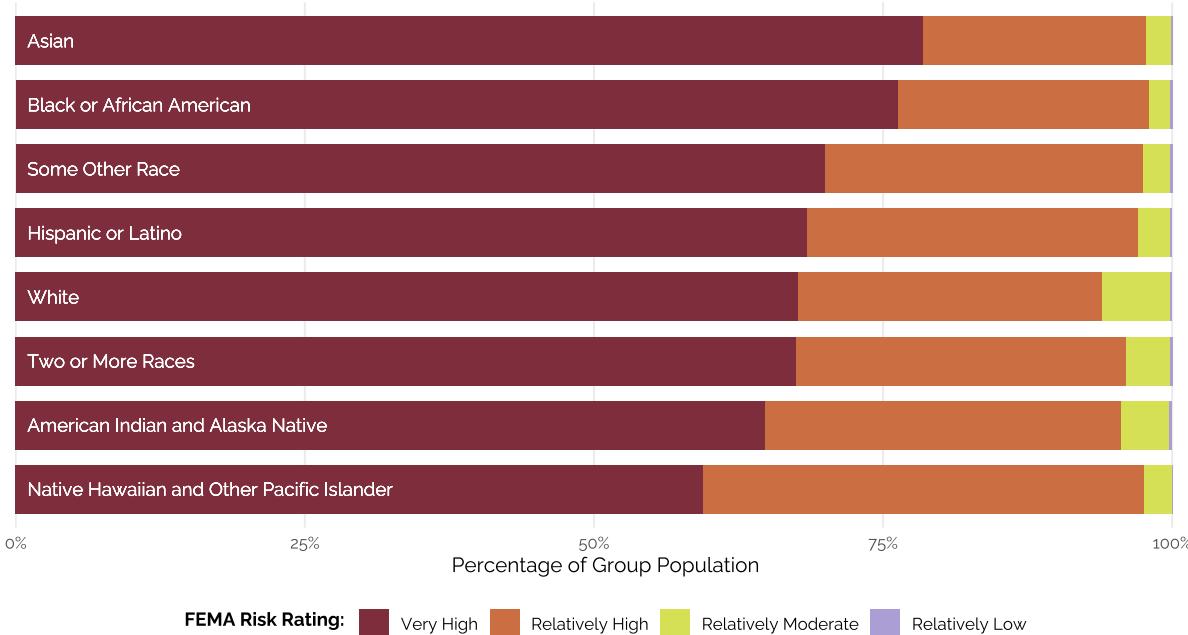
```

geom_col(width = 0.75) +
# Embed group labels directly on bars
geom_text(aes(label = race_eth_group, x = 0.01),
          hjust = 0, color = "white",
          size = 3.5, family = "raleway") +
# Apply color palette
paletteer::scale_fill_palettes("rockthemes::janelle",
                                name = "FEMA Risk Rating: ") +
# Reverse legend order to match bar stacking
guides(fill = guide_legend(reverse = TRUE)) +
scale_x_continuous(labels = scales::label_percent(),
                    expand = expansion(mult = c(0, 0.02))) +
labs(
  title = "Climate Hazard Risk Exposure Across Racial Groups in California",
  subtitle = "Asian, Black, and Some Other Race populations show highest concentration in",
  caption = "Data: FEMA National Risk Index (2023 Release) & ACS (2023)",
  x = "Percentage of Group Population",
  y = NULL
) +
theme_minimal(base_family = "raleway") +
theme(
  axis.text.y = element_blank(), # Hide y-axis text (labels are on bars)
  legend.position = "bottom",
  legend.title = element_text(face = "bold", size = 10,
                               margin = margin(b = 5)),
  legend.text = element_text(size = 9),
  panel.grid.minor = element_blank(),
  panel.grid.major.y = element_blank(),
  plot.title = element_text(face = "bold", size = 14),
  plot.subtitle = element_text(size = 11, margin = margin(b = 10)),
  plot.caption = element_text(size = 8, hjust = 0, margin = margin(t = 10)),
  plot.margin = margin(10, 10, 10, 10)
)

```

Climate Hazard Risk Exposure Across Racial Groups in California

Asian, Black, and Some Other Race populations show highest concentration in very high risk counties.



Data: FEMA National Risk Index (2023 Release) & ACS (2023)

Figure 1: Proportional climate risk exposure by racial and ethnic groups in California.

Questions

1. What are your variables of interest and what kinds of data are they?

- `race_eth_group`: Categorical - racial/ethnic group identity
- `risk_rating`: Categorical - FEMA risk level (Very/Relatively Low to Very High)
- `estimate/total_pop`: Numeric - population counts
- `percent`: Numeric - proportion of group in each risk category

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?

The question asks about variation across groups, which requires comparing proportional distributions. I initially considered a dodged bar chart to show each risk level separately, but this made it difficult to see the full risk profile for each group. I ended up going for a stacked bar chart instead which makes the comparison of how risk is distributed across groups easier while also showing each group sums to 100%. I also considered faceted plots showing each risk category separately, but this would have made more work for the viewer. I think the

stacked bar chart best answers “which groups face the most high-risk exposure” at a glance while preserving the context of each group’s risk distribution.

3. Summarize your main finding in no more than two sentences.

Asian, Black or African American, and Some Other Race populations show the highest proportions living in Very High risk counties (approximately 75-80%). Native Hawaiian and Other Pacific Islander populations show the lowest exposure to Very High risk counties (approximately 50%), though they still face substantial climate hazard exposure with most residing in Relatively High or Very High risk areas.

4. What modifications did you make to this visualization to make it more easily readable?

I ordered groups by their Very High risk exposure so the disparity is immediately apparent. I positioned the race/ethnicity labels directly on the bars in white text for a less cluttered and cleaner look. I also changed the font to be nicer to read and look at. I selected an intuitive color palette that progresses from lighter blue/yellow (lower risk) to darker red (higher risk). I bolded the title of the legend to distinguish it from the legend labels, and the plot title itself so it would stand out against the subtitle.

5. Is there anything you wanted to implement, but didn’t know how?

I wanted to make the legend easier to understand, I feel like when everything is all in a row it can be hard to immediately determine which are labels and which is the title of the legend. I tried putting it vertical to the side of the plot but I didn’t like that either. I am not sure exactly how I would make it better but I feel like it could be better.