

# HW #3

## Visualizing FEMA NRI and ACS Data

Ava Robillard

2026-02-11

**Key Question: How does climate hazard risk exposure vary across racial / ethnic groups in California?**

### Load packages

```
library(here)
library(tidyverse)
library(tidycensus)
library(janitor)
#library(ggridges)
```

### Import 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"))

# Import NRI data
nri_ca <- read_csv(here("data", "National_Risk_Index_Counties_807384124455672111.csv")) %>%
  clean_names() %>%
  # Remove unwanted territories and districts
  filter(state_name == "California")

```

## Data wrangling

```

# Join NRI and ACS data
full_data <- left_join(race_ethnicity, nri_ca, by = c("GEOID" = "state_county_fips_code"))

# Clean data
clean_data <- full_data %>%
  clean_names() %>%
  # Keep wanted columns
  select(geoid, name, variable, estimate, moe, label, concept, county_name, population_2020,
  # Rename label column values for plotting
  mutate(label = replace(label, label == 'Estimate!!Total', 'Estimate!!Total:!!Total')) %>%
  mutate(label = str_remove(label, ":$")) %>%
  mutate(label = str_sub(label, start = 19, end = -1)) %>%
  mutate(label = case_when(
    label == "White alone" ~ "White",
    label == "Black or African American alone" ~ "Black or African American",
    label == "American Indian and Alaska Native alone" ~ "American Indian and Alaska Native",
    label == "Asian alone" ~ "Asian",
    label == "Native Hawaiian and Other Pacific Islander alone" ~ "Native Hawaiian and Other",
    label == "Some Other Race alone" ~ "Some Other Race",
    TRUE ~ label
  ))

```

```

# Total population of each group in California
total_pop <- clean_data %>%
  filter(label != "Total") %>%
  group_by(label) %>%
  summarize(population = sum(estimate))

# Population in each group and risk rating
total_risk <- clean_data %>%
  filter(label != "Total") %>%
  group_by(label, national_risk_index_rating_composite) %>%
  summarize(sub_pop = sum(estimate))

# Join data
full_popdata <- left_join(total_risk, total_pop, by = "label") %>%
  # Create percent column of each group and risk rating
  mutate(perc_pop = (sub_pop/population)*100) %>%
  # Reorder risk levels
  mutate(national_risk_index_rating_composite = factor(national_risk_index_rating_composite,
                                                         levels = c("Very Low", "Relatively Low",
                                                       "Relatively Moderate", "Relatively High", "Very High"))

# Reorder columns by proportion of High Risk for plotting
full_popdata <- full_popdata %>%
  group_by(label) %>%
  mutate(very_high_prop = sum(perc_pop[national_risk_index_rating_composite %in% c("Very High",
  ungroup() %>%
  mutate(label = fct_reorder(label, very_high_prop))

```

## Build visualization

```

# Customize colors from NRI map
rating_colors <- c("Very High" = "#D06274",
                    "Relatively High" = "#E2867E",
                    "Relatively Moderate" = "#F0D973",
                    "Relatively Low" = "#6DA9C9",
                    "Very Low" = "#6B84C2")

# Create stacked bar chart
ggplot(full_popdata, aes(x = label, y = perc_pop,

```

```

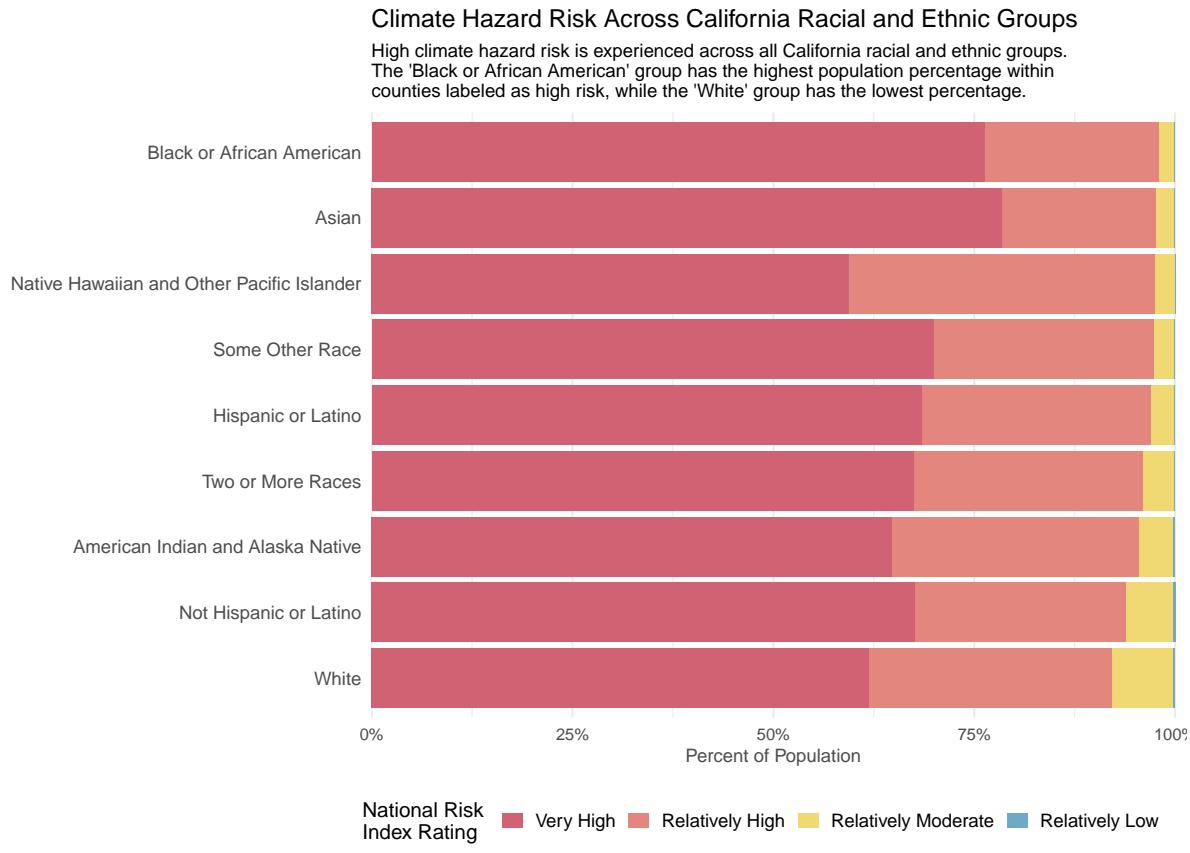
fill = national_risk_index_rating_composite))+

geom_col() +
coord_flip() +
scale_fill_manual(values = rating_colors) +
# Flip legend
guides(fill = guide_legend(reverse = TRUE)) +
# Expand plot and add percents to axis
scale_y_continuous(expand = c(0,0), labels = scales::label_percent(scale = 1)) +
labs(title = "Climate Hazard Risk Across California Racial and Ethnic Groups",
     subtitle = "High climate hazard risk is experienced across all California racial and ethnic groups",
     caption = "Data sources: FEMA National Risk Index (2025 Release) and US Census Bureau's American Community Survey (2019-2023 5-Year Estimates)",
     y = "Percent of Population",
     fill = "National Risk \nIndex Rating") +
# Customize theme
theme_minimal() +
theme(
  # legend edits ----
  legend.position = "bottom",
  legend.direction = "horizontal",
  legend.key.width = unit(0.4, "cm"),
  legend.key.height = unit(0.25, "cm"),
  legend.title = element_text(size = 9.5),
  legend.justification = "left",
  legend.box.margin = margin(l = -10),

  # axis edits ----
  axis.title.y = element_blank(),
  axis.text.x = element_text(size = 8),
  axis.title = element_text(color = "grey30", size = 9),

  # text size edits ----
  plot.title = element_text(size = 12),
  plot.subtitle = element_text(size = 9),
  plot.caption = element_text(size = 7)
)

```



## 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)?
  - Risk score rating: Categorical, ordered based on binned numeric values
  - Percent of population: Numeric
  - Race/ethnicity: Categorical
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 first determined my variable types and then used the From Data to Viz flow chart, which suggested a stacked bar plot or alternatively a grouped bar plot. I decided to go with a stacked bar plot because there were many groups to plot and 4 different categories within these groups,

so the grouped bar plot would have been more visually overwhelming. It also helped to draw the stacked bar plot and what could be placed on each axis to then wrangle the data as needed.

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

Climate hazard risk is experienced greatly across all California ethnic and racial groups, with between 75% and 100% of all group populations living within counties with National Risk Index Ratings of Very High or Relatively High. The Black or African American alone and Asian alone groups have the highest proportion of populations within high risk counties, while the Not Hispanic or Latino and White alone groups have the lowest proportion.

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

After creating the initial stacked bar chart, I flipped the plot to make the group names more legible, changed x axis ticks to percents, removed the y axis label as it was redundant with the group names, moved legend to bottom to extend plot bars, and added a visual hierarchy of text size from the title to the axis labels. For my data, I rearranged the bar plotting order based upon the highest percent of “Very High” and “Relatively High” ratings to more clearly show how groups are ranked in terms of climate risk. Lastly, I rearranged the legend to match the order of the plot, reducing how much you need to look around to determine what the colors correlate to.

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

The 100% axis tick label was cut off on the right side of the plot, despite trying to reduce the size of the tick labels. I am also not sure if making some of the group axis labels into two lines to allow the bars to take up more space would be advised or doable.