Climate Hazard Risk Exposure Across Racial/Ethnic Groups in California
HW #3: Visualizing FEMA NRI x ACS Data

Lucian Scher

2026-02-10

Setup

```r
library(tidyverse)
library(tidycensus)
library(here)
library(janitor)
```

I. Import & Save ACS Data

```r
# Import ACS data (commented out after initial download)
# 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)
#
# acs_vars <- load_variables(year = 2023, dataset = "acs1")
# race_ethnicity <- race_ethnicity |>
#   left_join(acs_vars, by = join_by(variable == name))
#
```

```r
# write_csv(race_ethnicity, here("data", "ACS-1yr-2023-county-race-ethnicity.csv"))

# Read ACS data
race_ethnicity <- read_csv(
  here("data", "ACS-1yr-2023-county-race-ethnicity.csv"),
  show_col_types = FALSE
)
```

II. Data Wrangling

```r
# Prepare ACS data
acs_data <- race_ethnicity |>
  mutate(
    race_ethnicity = case_when(
      variable == "B02001_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_ # Handles other variables
    ),
    county_fips = str_sub(GEOID, 1, 5)
  ) |>
  filter(!is.na(race_ethnicity)) |>
  select(county_fips, NAME, race_ethnicity, estimate)

# Load NRI data
nri_raw <- read_csv(
  here("data", "National_Risk_Index_Counties_807384124455672111.csv"),
  show_col_types = FALSE # Reduces console noise
)

nri_cleaned <- nri_raw |> # Clean column names
  janitor::clean_names()

# Filter to California using state abbreviation
nri_data <- nri_cleaned |>
```

```r
  filter(state_name_abbreviation == "CA") |>
  mutate(
    county_fips = str_pad(as.character(state_county_fips_code),
                          width = 5,
                          side = "left",
                          pad = "0"),
    risk_index = national_risk_index_score_composite
  ) |>
  filter(!is.na(county_fips), !is.na(risk_index)) |>
  select(county_fips, risk_index)

# Combine ACS and NRI data
combined <- acs_data |>
  left_join(nri_data, by = "county_fips") |>
  drop_na(risk_index)

# Calculate weighted average risk by race/ethnicity group
risk_by_group <- combined |>
  group_by(race_ethnicity) |>
  summarize(
    avg_risk = sum(risk_index * estimate, na.rm = TRUE) / sum(estimate, na.rm = TRUE),
    total_pop = sum(estimate, na.rm = TRUE)
  ) |>
  filter(!is.na(avg_risk), is.finite(avg_risk)) |>
  arrange(desc(avg_risk)) |>
  mutate(
    race_ethnicity = factor(race_ethnicity, levels = race_ethnicity)
  )
```

III. Visualization

```r
# Calculate overall California average risk (weighted by population)
ca_avg_risk <- weighted.mean(combined$risk_index, w = combined$estimate, na.rm = TRUE)

# Calculate deviation from average for each group
diverging_data <- risk_by_group |>
  mutate(
    deviation = avg_risk - ca_avg_risk,
    above_avg = deviation > 0
```

```r
  ) |>
  arrange(desc(deviation))

# Find max deviation for axis limits
max_dev <- max(abs(diverging_data$deviation)) * 1.15  # Add 15% padding for labels

# Create diverging bar chart
p <- diverging_data |>
  ggplot(aes(x = reorder(race_ethnicity, deviation), y = deviation, fill = above_avg)) +
  geom_col(alpha = 0.85, width = 0.7) +
  geom_hline(
    yintercept = 0,
    linetype = "solid",
    color = "gray40",
    linewidth = 1,
    alpha = 0.8
  ) +
  geom_text(
    aes(label = paste0(ifelse(deviation > 0, "+", ""), round(deviation, 2))),
    hjust = ifelse(diverging_data$deviation > 0, -0.15, 1.15),
    size = 4,
    fontface = "bold",
    color = "gray10"
  ) +
  scale_fill_manual(
    values = c("FALSE" = "#1E319E", "TRUE" = "#E35205"),
    labels = c("Below Average", "Above Average"),
    guide = guide_legend(
      title = NULL,
      override.aes = list(alpha = 1),
      nrow = 1
    )
  ) +
  scale_y_continuous(
    limits = c(-max_dev, max_dev),
    labels = function(x) paste0(ifelse(x > 0, "+", ""), round(x, 2)),
    breaks = scales::pretty_breaks(n = 6)
  ) +
  coord_flip() +
  labs(
    title = "Deviation from California Average Risk Score",
    subtitle = paste0(
```

```r
      "State average: ", round(ca_avg_risk, 2),
      " | Shows which ethnicities face above/below average exposure"
    ),
    x = NULL,
    y = "Deviation from State Average",
    caption = "Data: FEMA National Risk Index (2023 Release), US Census Bureau American Commu
) +
theme_minimal(base_size = 13) +
theme(
  plot.background = element_rect(fill = "white", color = NA),
  panel.background = element_rect(fill = "white", color = NA),
  plot.title = element_text(
    size = 18,
    face = "bold",
    margin = margin(b = 8),
    lineheight = 1.2,
    color = "#1a1a1a"
  ),
  plot.subtitle = element_text(
    size = 12,
    color = "gray40",
    margin = margin(b = 20),
    lineheight = 1.3
  ),
  plot.caption = element_text(
    size = 9,
    color = "gray50",
    hjust = 0,
    margin = margin(t = 12),
    lineheight = 1.4
  ),
  axis.title.x = element_text(
    size = 12,
    face = "bold",
    margin = margin(t = 10),
    color = "#2a2a2a"
  ),
  axis.text.y = element_text(
    size = 11,
    color = "#2a2a2a"
  ),
  axis.text.x = element_text(
```
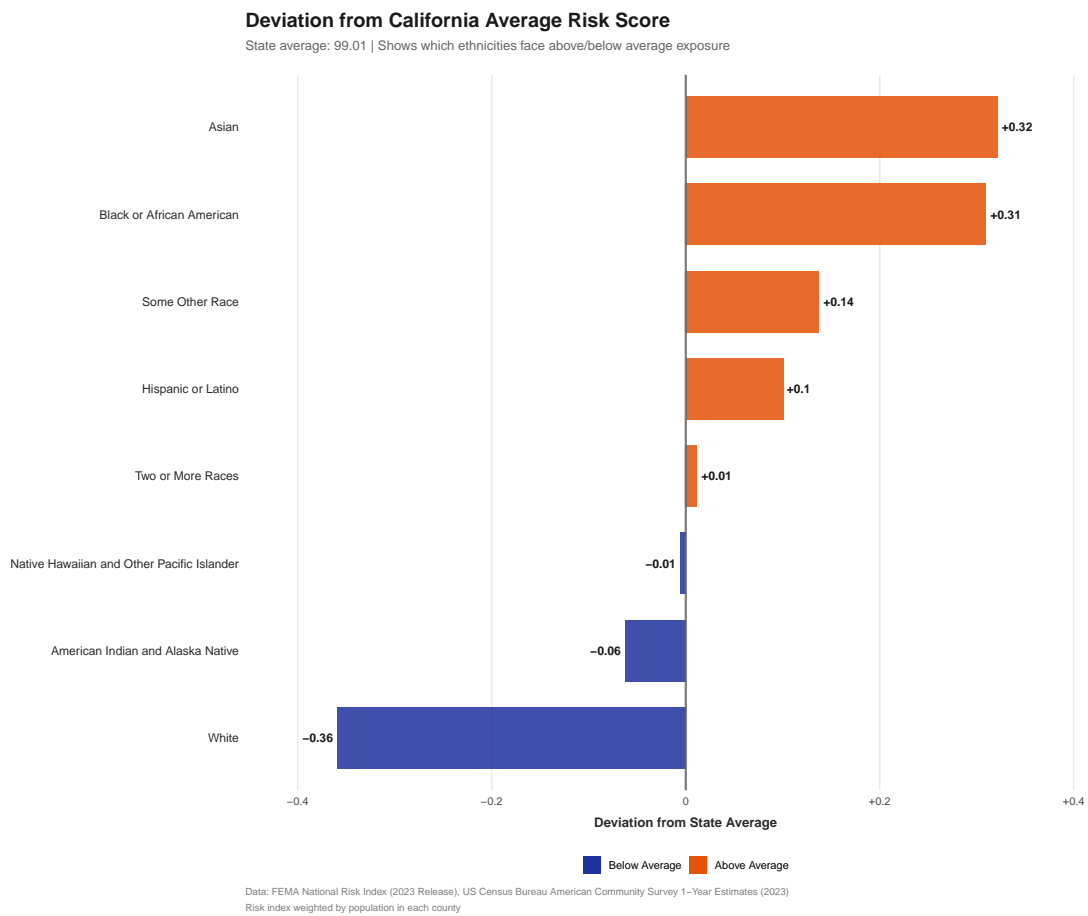
```
    size = 10,
    color = "#4a4a4a"
  ),
  panel.grid.major.y = element_blank(),
  panel.grid.minor = element_blank(),
  panel.grid.major.x = element_line(color = "gray90", linewidth = 0.5),
  legend.position = "bottom",
  legend.margin = margin(t = 10),
  plot.margin = margin(25, 45, 20, 20)
 )

print(p)
```

Deviation from California Average Risk Score

State average: 99.01 | Shows which ethnicities face above/below average exposure

Ethnicity	Deviation
Asian	+0.32
Black or African American	+0.31
Some Other Race	+0.14
Hispanic or Latino	+0.1
Two or More Races	+0.01
Native Hawaiian and Other Pacific Islander	−0.01
American Indian and Alaska Native	−0.06
White	−0.36

Deviation from State Average

Below Average   Above Average

Data: FEMA National Risk Index (2023 Release), US Census Bureau American Community Survey 1–Year Estimates (2023)
Risk index weighted by population in each county

6

IV. Answer Questions

1. Variables of Interest and Data Types

- **Race/Ethnicity Group**: Categorical (nominal) - eight distinct groups
- **Risk Index**: Numeric (continuous) - FEMA's composite climate hazard risk score
- **Population Count**: Numeric (discrete) - number of people in each group by county
- **County FIPS Code**: Categorical (nominal) - geographic identifier for joining datasets

2. Graphic Form Decision

I chose a **diverging bar chart** to show how climate hazard risk exposure varies across ethnic groups in California because by showing deviation from the state average, the chart immediately reveals which groups face above- or below-average exposure relative to California as a whole. The length of bars clearly shows the magnitude of deviation, making it easy to see that some groups face substantially higher or lower risk than the state average especially when the actually numbers are very close when compared on a 0-100 scale.

I also considered a standard bar chart and dot plot but neither showed enough deviation. I settled on the diverging bar chart because it directly answers "how does risk vary across groups" by showing each group's position relative to the California average, making differencies immediately clear

3. Main Finding

Asian and Black or African American communities in California face the highest climate hazard risk exposure, with risk indices 0.31 points above the state average (99.32 and 99.33 respectively, compared to the state average of 99.01). White communities face the lowest exposure, with a risk index 0.36 points below the state average (98.65), representing a 0.68 point gap between the highest and lowest risk groups.

4. Modifications for Readability

Bars are ordered by deviation magnitude to highlight disparities, and each bar is labeled with its deviation value (e.g., "+0.31", "-0.36") to eliminate estimation from the axis. A solid horizontal reference line at zero marks the state average, while vertical grid lines help to read values. The visualization uses a clean, professional theme, and a clear hierarchy of title, subtitle (including state average), and caption provides context at multiple levels.

5. Implementation Challenges

I want to add something highlighting the magnitude of the gap between highest and lowest groups (0.68), but the diverging format already makes this somewhat clear, and I could not think of a good way to do it. I experimented with different color schemes and found that the deep orange / deep blue worked well for clearly distinguishing above- versus below-average exposure while being accesisble to many color-blindness viewers.