Climate Hazard Risk Exposure by Ethnic Group in California

Melannie Moreno Rolón

2026-02-12

```
library(tidyverse)
library(janitor)
library(scales)
```

III. Import & save ACS 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
```

```
Rows: 420 Columns: 7
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (5): GEOID, NAME, variable, label, concept
dbl (2): estimate, moe

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

IV. How does climate hazard risk exposure vary across racial / ethnic groups in California?

Data Cleaning

```
# Load raw NRI data
nri_raw <- read_csv(here::here("data", "National_Risk_Index_Counties.csv"))
```

```
Warning: One or more parsing issues, call `problems()` on your data frame for details,
e.g.:
  dat <- vroom(...)
  problems(dat)
```

```
Rows: 3232 Columns: 467
-- Column specification --------------------------------------------------------
Delimiter: ","
chr  (67): National Risk Index ID, State Name, State Name Abbreviation, Stat...
dbl (396): OBJECTID, Population (2020), Building Value ($), Agriculture Valu...
lgl   (4): Coastal Flooding - Number of Events, Earthquake - Number of Event...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

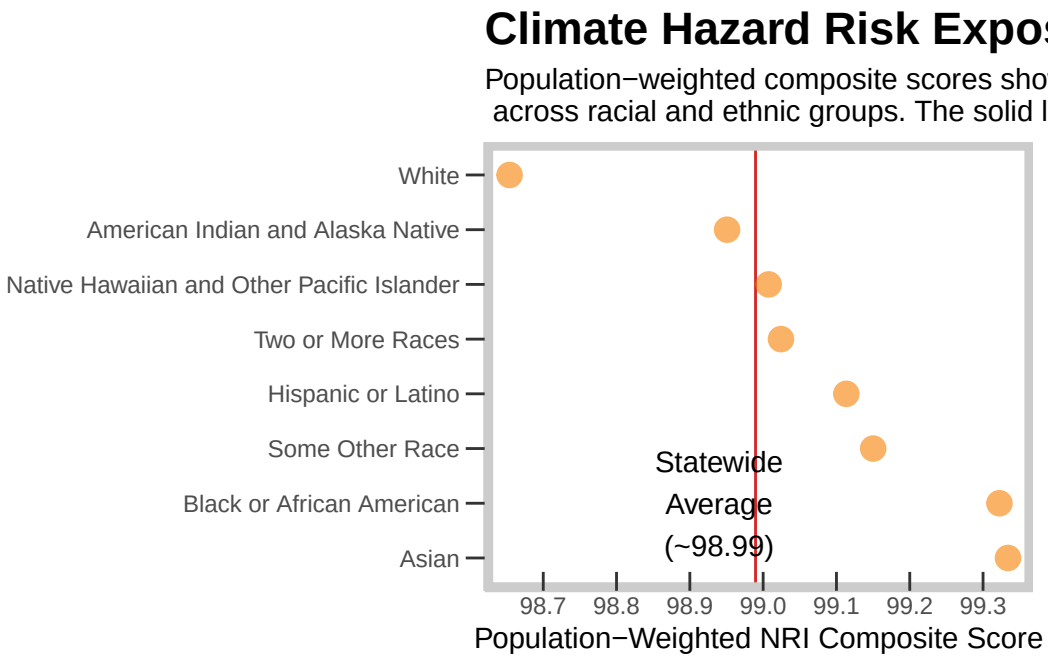
```r
# Keep only CA NRI values
nri_ca <- nri_raw %>%
  filter(`State Name Abbreviation` %in% c("CA")) %>%
  clean_names()

# Standardize race_ethnicity data
race_eth_clean <- race_ethnicity %>%

  # Split NAME column into Country and State Name
  separate(col = "NAME", sep = ", ", into = c("county_name", "state_name")) %>%

  # Remove the word County from county_name
  mutate(
    county_name = str_remove(county_name, pattern = " County$"),
    county_name = str_trim(county_name)) %>%

  # Clean the label column values of interest
  mutate(race_group = case_when(
    label == "Estimate!!Total:!!White alone" ~ "White",
    label == "Estimate!!Total:!!Black or African American alone" ~ "Black or African Americar
    label == "Estimate!!Total:!!Asian alone" ~ "Asian",
    label == "Estimate!!Total:!!Native Hawaiian and Other Pacific Islander alone" ~ "Native H
    label == "Estimate!!Total:!!Hispanic or Latino:" ~ "Hispanic or Latino",
    label == "Estimate!!Total:!!American Indian and Alaska Native alone" ~ "American Indian a
    label == "Estimate!!Total:!!Some Other Race alone" ~ "Some Other Race",
    label == "Estimate!!Total:!!Two or More Races:" ~ "Two or More Races",
    TRUE ~ NA) ) %>%
  filter(!is.na(race_group)) %>%
  clean_names()


# Merge ethnicity and NRI data
nri_ethnicity <- left_join(x = race_eth_clean,
                           y = nri_ca,
                           by = c("geoid" = "state_county_fips_code"))
```

```r
# Prepare plotting data frame
# Find the total population from estimates and calculate weighted NRI composite score
plot_data <- nri_ethnicity %>%
  group_by(race_group) %>%
  summarise(
    total_population = sum(estimate, na.rm = TRUE),
```

```
    weighted_nri_score =
      sum(national_risk_index_score_composite * estimate, na.rm = TRUE) /
      sum(estimate, na.rm = TRUE),
    .groups = "drop"
  )

# Arrange the plot dataframe to order the largest score at the top and the smallest at the bo
plot_data <- plot_data %>%
  arrange(weighted_nri_score) %>%
  mutate(race_group = factor(race_group, levels = rev(race_group)))

# Calculate the state average NRI score based on 2020 census population estimates for our ver
state_avg <- nri_ethnicity %>%
  distinct(geoid, national_risk_index_score_composite, population_2020) %>%
  summarise(
    state_avg = sum(national_risk_index_score_composite * population_2020, na.rm = TRUE) /
                sum(population_2020, na.rm = TRUE)
  ) %>%
  pull(state_avg)
```

```
# Build Dot Plot
ggplot(plot_data,
       aes(x = weighted_nri_score,
           y = race_group)) +
  geom_vline(xintercept = state_avg, linetype = "solid", color = "#D62828") +
  geom_point(size = 4, color = "#F77F00", alpha = 0.6) +

  labs(
    x = "Population-Weighted NRI Composite Score",
    y = "Race/Ethnic Group",
    title = "Climate Hazard Risk Exposure by Ethnic Group in California",
    subtitle = "Population-weighted composite scores show differences in average climate haza
  ) +

  annotate("text", x = 98.94, y = "Black or African American",
           label = "Statewide\nAverage\n(~98.99)") +

  theme_bw() +

  theme(plot.background = element_rect(fill = "white"),
        panel.background = element_rect(fill = "white"),
        panel.border = element_rect(linewidth = 3,colour = "grey80"),
```

```
        panel.grid = element_blank(),
        axis.line = element_line(linewidth = 0.5, colour = "grey80"),
        plot.title = element_text(size = rel(1.5), face = "bold"),
        axis.ticks.length.y = unit(0.25, "cm"),
        axis.ticks.length.x = unit(-0.25, "cm"),
        axis.title.y = element_blank()) +

  scale_x_continuous(breaks = pretty_breaks(n = 8))
```



Part V

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)?

Our variables of interest and data types were: - Race/ethnic Group (`race_group`, obtained from the `label` column). Type: categorical (nominal) - NRI Composite Score (`national_risk_index_composite_score`). Type: Numeric (continuous) which ranges from 0 through 100. - 2020 Census Population (`population_2020`, used to calculate the state average reference line). Type: Numeric (discrete) - Population estimates at the county level (`estimate`). Type: Numeric (discrete)

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 selected a dot plot because the graphic showed relative differences in magnitude and ordering the groups by score made the comparison intuitive. Alternative graphic forms that could have been included are a bar chart, a scatterplot, and a clorpleth map.I settled on the dot plot because it was the most direct way to compare group-level exposure values without implying variability that did not exist in the summarized data.

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

Population-weighted climate hazard exposure varies across racial and ethnic groups in California. Asian and Black populations experience the highest average exposure, while White populations experience the lowest, though overall differences are relatively small at the county level.

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

To make the visualization more readable, I reordered ethnic groups by exposure. I also used a horizontal layout to accommodate the long category labels. Other modifications included adding a vertical dashed line to indicate the statewide average for context, and zooming on the x-axis to highlight the subtle differences in exposure levels.

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

I initially attempted a few different plots, such as a heatmap, a parallel coordinate scatter plot, and a boxplot. To create a grouped boxplot, I found that what I wanted to show, two categorical variables (race and NRI category), was not viable since the county-level data structure repeated the same risk score across race groups. I would have liked to implement confidence intervals or uncertainty bands around the weighted exposure estimates, changing the font type of the plot.