

# HW #3: ACS Data

Ixel M.

## Visualization message goal

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

## Packages and Data uploads

```
# Upload necessary libraries
library(tidyverse)
library(janitor)
library(here)
library(tidycensus)

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

#.....Step 1b: import race & ethnicity data.....
#race_ethnicity <- 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"))

nr_index <- read_csv(here("data/National_Risk_Index_Counties_807384124455672111.csv"))

```

## First data exploration and wrangling

```

# Data Exploration
head(race_ethnicity)
glimpse(race_ethnicity)
colnames(race_ethnicity)

# Check racial/ethnic groups
race_ethnicity %>%
  distinct(variable, label) %>%
  print(n = Inf)

# Clean & filter for racial/ethnic variables
race_ethnicity_clean <- race_ethnicity %>%
  filter(variable != "B01003_001", # Remove total population
         variable != "B03002_002") %>% # Remove "Not Hispanic or Latino"
# Clean labels to be shorter
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"
)) %>%
# Select relevant columns
select(GEOID, county = NAME, race_ethnicity, population = estimate)

```

```

# Create a matching GEOFID for NRI data to join by
nr_index_ca <- nr_index %>%
  clean_names() %>%
  filter(state_name_abbreviation == "CA",
        county_type == "County") %>%
  select(GEOFID = state_county_fips_code,
         county = county_name,
         nri_score = national_risk_index_score_composite) %>%
  mutate(GEOFID = as.character(GEOFID)) # making sure its chr type

# Join census demographics with risk scores
climate_risk_race <- race_ethnicity_clean %>%
  left_join(nr_index_ca, by = "GEOFID")

# Summarize to get the weight.mean for populations

exposure_summary <- climate_risk_race %>%
  group_by(race_ethnicity) %>%
  summarize(
    # weight.mean assigns different levels of importance to each data point
    avg_risk = weighted.mean(nri_score, w = population, na.rm = TRUE),
    total_population = sum(population, na.rm = TRUE)
  )

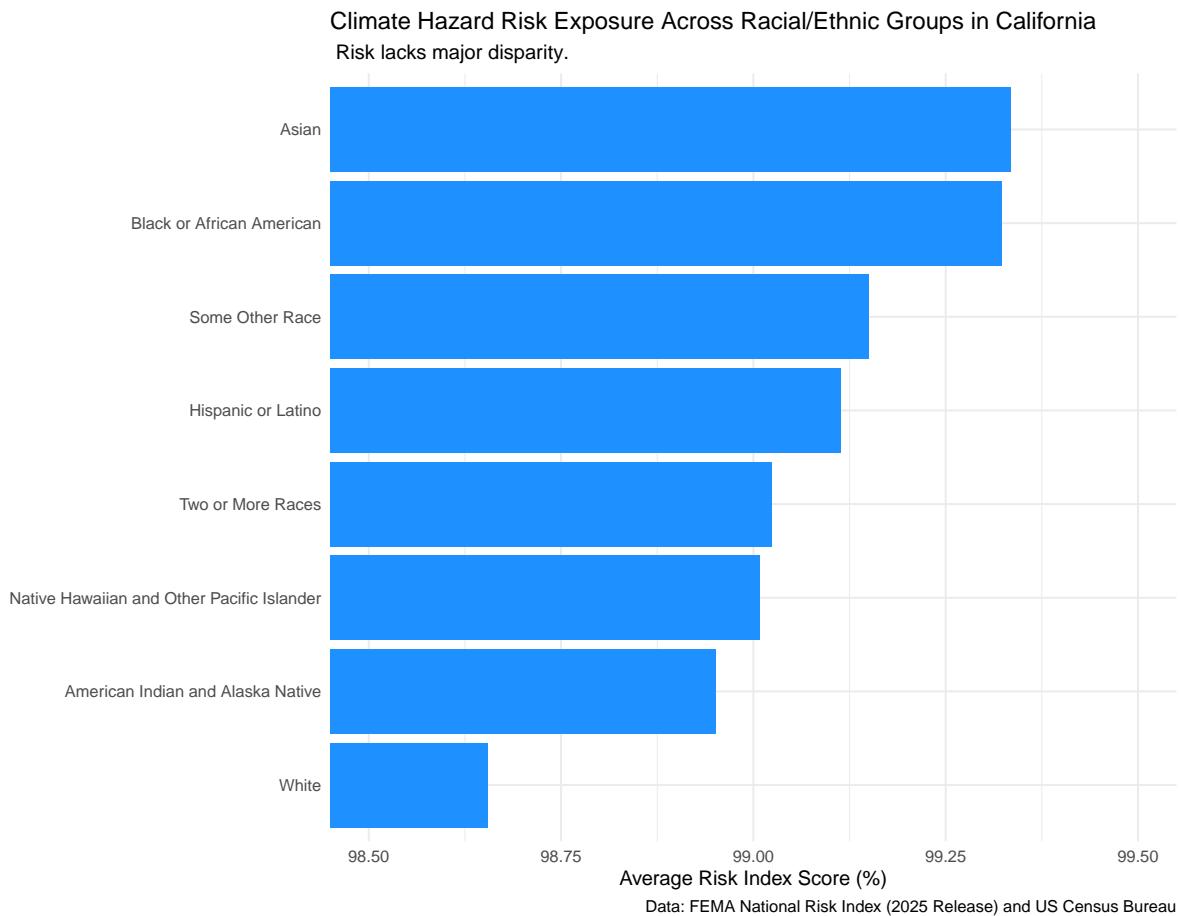
```

## Build the plot

```

ggplot(exposure_summary, aes(x = reorder(race_ethnicity, avg_risk), y = avg_risk)) +
  geom_col(fill = "dodgerblue") +
  coord_flip(ylim = c(98.5, 99.5)) +
  labs(
    title = "Climate Hazard Risk Exposure Across Racial/Ethnic Groups in California",
    subtitle = " Risk lacks major disparity. ",
    x = NULL,
    y = "Average Risk Index Score (%)",
    caption = "Data: FEMA National Risk Index (2025 Release) and US Census Bureau"
  ) +
  theme_minimal()

```



## V. Answer some 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 following variables of interest are;

- Race/ethnicity (categorical: White, Black or African American, Asian, Hispanic or Latino, etc.)
- Population by race/ethnicity (numeric, discrete)
- National Risk Index Score (numeric, continuous - percentile from 0-100)
- County identifier/GEOID (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 started by thinking about the question and understanding what it's asking. Then I determined which variables I need from the data to address the questions. I then do some data exploration and identify the data types - whether numeric or categorical - and visualized how these variables would map to a plot. In this case, I'm comparing climate hazard risk exposure (represented as a numeric score) across racial/ethnic groups (categorical). A bar chart works well for comparing a continuous variable across discrete categories. I also consulted Visual Vocabulary and data-to-viz.com to validate my approach and explore alternatives like grouped bar charts or dot plots, but settled on a simple bar chart for clarity.

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

Climate hazard risk exposure is uniformly high across all racial and ethnic groups in California, with a narrow range of variation. Average scores clustered between the 98.9th and 99.3rd percentile. This narrow range aligns with findings from the first visualization, which showed California counties have consistently elevated risk scores with low variability. Though Asian and Black or African American populations show slightly higher exposure the differences are minimal. These minimal differences suggest climate risk affect all communities nearly equally across the state.

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

I used `coord_flip()` to display racial/ethnic group names horizontally for easier reading. I also adjusted the y-axis scale to zoom in on the narrow range of values to highlight the small differences in exposure levels.

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

I wanted to create a stacked bar chart showing the distribution of each group across different risk categories (Very High, High, Moderate, Low) but it was really hard to see those difference categories and wasn't sure how to adjust for that. I also wanted to add error bars or confidence intervals to show uncertainty in the weighted averages but wasn't sure how and if it would just cause the plot to look cluttered.