

# hw3

Sarayu Ramnath

2026-01-25

## Set Up

```
library(tidyverse)
library(ggplot2)
library(janitor)
library(dplyr)
library(readr)

# #....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"))
```

## Objective

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

```
# nri_ca_race_raw <- nri_raw |>  
#   janitor::clean_names()|>  
#   select(1:17) |>  
#   filter(state_fips_code == "06") |>  
#   left_join(  
#     race_ethnicity,  
#     by = c("state_county_fips_code" = "GEOID")  
#   )  
#  
#  
# #clean up nri_ca_race set and then save to data  
# readr::write_csv(nri_ca_race_raw, here::here("data", "nri_ca_race_raw.csv"))  
  
nri_ca_race_raw <- read_csv(here::here("data", "nri_ca_race_raw.csv"))
```

Tidy data further

```
#1. Clean column names  
nri_ca_race_tidy <- nri_ca_race_raw |>  
  janitor::clean_names() |>  
  select(  
    county_name,  
    state_county_fips_code,  
    population_2020,  
    national_risk_index_value_composite,  
    national_risk_index_score_composite,  
    national_risk_index_rating_composite,  
    national_risk_index_state_percentile_composite,  
    variable,  
    estimate,  
    label,  
    concept
```

```

) |> # 2. Keep only relevant columns
filter(concept %in% c("Race", "Hispanic or Latino Origin by Race")) |> # 3. Only race/ethnicity
mutate( # 4. Clean up race/ethnicity labels
  race_ethnicity = case_when(
    grepl("White alone", label) ~ "White",
    grepl("Black or African American", label) ~ "Black or African American",
    grepl("Asian alone", label) ~ "Asian",
    grepl("American Indian", label) ~ "American Indian / Alaska Native",
    grepl("Native Hawaiian", label) ~ "Native Hawaiian / Pacific Islander",
    grepl("Some Other Race", label) ~ "Other Race",
    grepl("Two or More Races", label) ~ "Two or More Races",
    grepl("Hispanic or Latino", label) ~ "Hispanic or Latino",
    grepl("Not Hispanic or Latino", label) ~ "Not Hispanic or Latino",
    TRUE ~ "Other"
  )
) |>
group_by(county_name) |> # 5. Compute race proportion
mutate(race_pct = estimate / sum(estimate, na.rm = TRUE)) |>
ungroup() |>
mutate(weighted_nri = race_pct * national_risk_index_value_composite)|># 6. Weighted NRI
select(-label) #7. drop label column

# Save tidy dataset
write_csv(nri_ca_race_tidy, here::here("data", "nri_ca_race_tidy.csv"))

```

Circular plot

```

library(tidyverse)
library(RColorBrewer)

# -----
# Aggregate by race/ethnicity
# -----
nri_race_summary <- nri_ca_race_tidy |>
  group_by(race_ethnicity) |>
  summarize(total_weighted_nri = sum(weighted_nri)) |>
  arrange(desc(total_weighted_nri)) |>
  mutate(id = row_number(),
         race_ethnicity = factor(race_ethnicity, levels = race_ethnicity)
  )

# -----

```

```

# Set color palette
# -----
race_palette <- brewer.pal(8, "Paired")
names(race_palette) <- nri_race_summary$race_ethnicity
# -----
# Plot
# -----
ggplot(nri_race_summary, aes(x = as.factor(id), y = total_weighted_nri, fill = race_ethnicity))
  geom_bar(stat = "identity", width = 1) +
  scale_fill_manual(values = race_palette) +
  coord_polar(start = 0) +
  ylim(-max(nri_race_summary$total_weighted_nri) * 0.2,
       max(nri_race_summary$total_weighted_nri) * 1.1) +
  theme_void() +
  theme(
    legend.position = "right",
    legend.title = element_text(face = "bold", size = 11),
    legend.text = element_text(size = 9),
    legend.key.size = unit(0.4, "cm"), # Small legend keys
    legend.spacing.y = unit(0.1, "cm"), # Tight spacing
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
    plot.subtitle = element_text(size = 11, hjust = 0.5, color = "gray30"),
    plot.caption = element_text(size = 8, hjust = 1, color = "gray50"),
    plot.margin = unit(c(1, 1, 0.5, 0.5), "cm"),
    aspect.ratio = 1
) +
  labs(
    title = "Climate Hazard Exposure by Race/Ethnicity in California",
    subtitle = "Total Weighted National Risk Index Across All Counties",
    caption = "Data: FEMA National Risk Index (2023)",
    fill = "Race/Ethnicity"
)

```

## rd Exposure by Race/Ethnicity in California

| Weighted National Risk Index Across All Counties



### Race/Ethnicity

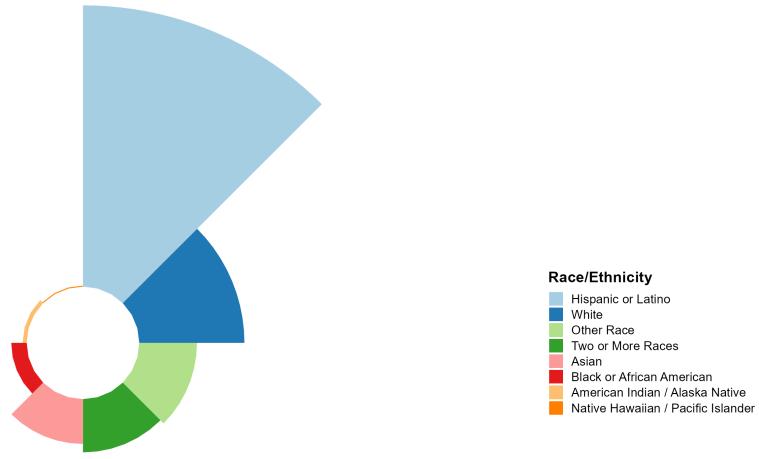
- [Light Blue] Hispanic or Latino
- [Dark Blue] White
- [Green] Other Race
- [Red] Two or More Races
- [Pink] Asian
- [Orange] Black or African American
- [Yellow] American Indian / Alaska Native
- [Dark Orange] Native Hawaiian / Pacific Islander

Data: FEMA National Risk Index (2023)

```
ggsave("nri_circular_plot.png",
       width = 12, height = 12, dpi = 300, bg = "white")
```

### Climate Hazard Exposure by Race/Ethnicity in California

Total Weighted National Risk Index Across All Counties

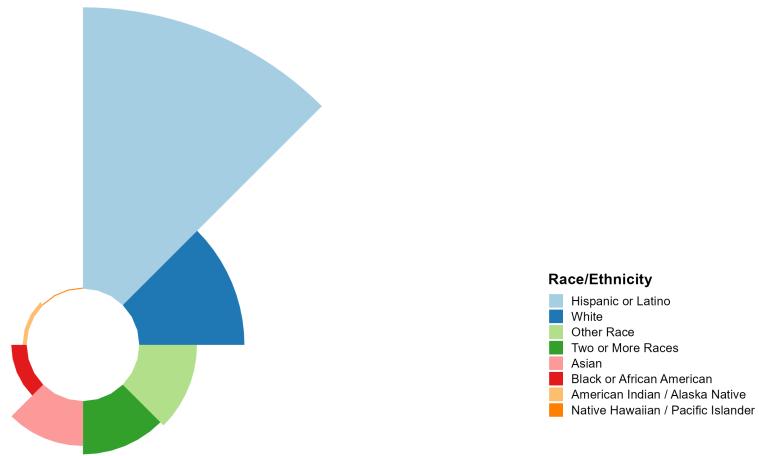


Data: FEMA National Risk Index (2023)

Figure 1: Alt-Text: Circular bar chart of climate hazard exposure across racial/ethnic groups in California. Hispanic or Latino populations show the highest exposure (15.7 billion weighted NRI), followed by White (5.9 billion), with Native Hawaiian/Pacific Islander showing the lowest (57 million). Eight distinct colored segments represent each group in descending order.

### Climate Hazard Exposure by Race/Ethnicity in California

Total Weighted National Risk Index Across All Counties



Data: FEMA National Risk Index (2023)

Figure 2: Circular bar chart of climate hazard exposure across racial/ethnic groups in California. Hispanic or Latino populations show the highest exposure (15.7 billion weighted NRI), followed by White (5.9 billion), with Native Hawaiian/Pacific Islander showing the lowest (57 million). Eight distinct colored segments represent each group in descending order.

## 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)?
- 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 looked at racetrack plot but I need to log10 scale and it skewed other NRI to be too big
  - Looked at lollipop but it felt like a barchart to me
  - looked at stacked bar plot and while it would show it the best
  - felt like you could understand magnitude of weighted nri even with the circular plot
- 3. Summarize your main finding in no more than two sentences.
- 4. What modifications did you make to this visualization to make it more easily readable?
- 5. Is there anything you wanted to implement, but didn't know how? If so, please describe.