hw3

AUTHOR
Sarayu Ramnath

PUBLISHED
February 12, 2026

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?

Data Tidying pt 1: Combining Hazard Risk with Race and Ethnicity

1. Prepped NRI data
   - clean names
   - select first 17 column variables (core) NRI variables

- filter to CA counties using state FIPS code
2. Join race and ethnicity census data
    - Left join using county-level FIPS code
    - Kept all counties from NRI dataset and attached race and ethnicity estimates.

```
# 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"))
```

Data Tidying Pt 2: Race-weighted National Risk Index by County

Possible visualization to answer the question d: Which racial and ethnic groups are more represented in counties with higher overall hazard risk when accounting for the portion of the population?

1. Select relevant demographic and risk variables
    - keep county identifiers, population totals, NRI composite metrics, race/ethnicity fields.
2. Filter for race and ethnicity concepts
    - keep census records related to race and Hispanic/Latino origin.
3. Recode census labels to standardize categories
4. Calculate race/ethnicity proportions
    - Group data by county
    - convert population estimates to proportions of total county population.
    - To have fair comparison across counties with different population sizes.

```
# 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/ethnic:
#    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. Calculate 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"))
nri_ca_race_tidy <- read_csv(here::here("data", "nri_ca_race_tidy.csv"))
```

Pivot: County-level race/ethnicity weighted NRI values were highly skewed with a small number of counties taking over the plot and overshadowing broader patterns. Now shifting from county-level to statewide aggregation by race and ethnicity but emphasizing cumulative population-weighted hazard burden.

1. Aggregate across counties by race and ethnicity.

   - Group by race & ethnicity only.

   - sum race-weighted NRI values across all counties

   - sum population counts for demographic context.

2. Rank racial & ethnic groups by cumulative burden

   - convert race and ethnicity categories to ordered factors.

   - Add ID numbers to plot function order.

```
library(tidyverse)
library(RColorBrewer)
```

```r
library(showtext)

# ----------------------------
# Aggregate by race/ethnicity
# ----------------------------
nri_race_summary <- nri_ca_race_tidy |>
  group_by(race_ethnicity) |>
  summarize(
    total_weighted_nri = sum(weighted_nri, na.rm = TRUE),
    total_population = sum(estimate, na.rm = TRUE)
    ) |>
  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

pal_palette <- c("#6A0304", "#F02224", "#F85550","#5F8A80", "#B5CEA7", "#ABAE39", "#4C4A21", "#31
names(pal_palette) <- nri_race_summary$race_ethnicity
# ----------------------------
# Set fonts
# ----------------------------
font_add_google("Source Sans Pro", "sourcesans")
font_add_google("Merriweather", "merriweather")
showtext_auto()
```

ggplot Circular Bar Plot

```r
# ----------------------------
# 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 = pal_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(
    text = element_text(family = "sourcesans"),
    plot.title = element_text(face = "bold", size = 16, hjust = 0.5),
    plot.subtitle = element_text(family = "merriweather", size = 12, hjust = 0.5, color = "gray30"
    plot.caption = element_text(family = "merriweather", size = 8, hjust = 1, color = "gray50"),
    plot.margin = unit(c(1, 1, 0.5, 0.5), "cm"),
     legend.position = "right",
```

```
    legend.title = element_text(face = "bold", size = 11, family = ),
    legend.text = element_text(size = 9),
    legend.key.size = unit(0.4, "cm"),
    legend.spacing.y = unit(0.1, "cm")
  ) +
  labs(
    title = "Who Bears Climate Hazard Risk in California?",
    subtitle = "Population-weighted National Risk Index across California",
    caption = "Data: FEMA National Risk Index (2023)",
    fill = "Race/Ethnicity"
  )
```

Who Bears Climate Hazard Risk in California?
Population-weighted National Risk Index across California
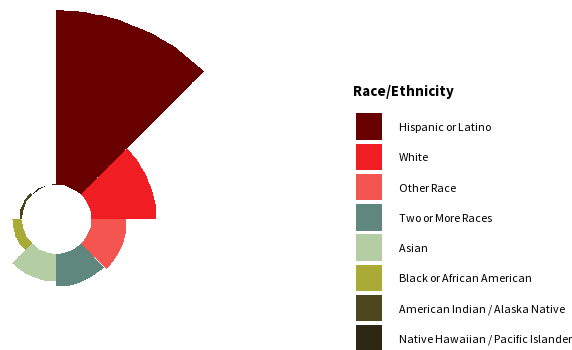
Data: FEMA National Risk Index (2023)

Alt-Text: "Circular bar chart of climate hazard exposure across racial/ethnic groups in California. Hispanic or Latino populations have 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. Note: This visualization shows cumulative population-weighted hazard burden. Smaller populations may experience high per-capita risk even if their total weighted NRI appears low."

(Side note, this alt-text is in the chunk header and it is not rendering so I am also adding it here).

Questions

---

1. What are your variables of interest and what kinds of data (e.g. numeric, categorical, ordered, etc.) are they?

- Race/ Ethnicity: Categorical

- Total Weighted NRI: Numeric (continuous)

- Id: Numeric/ ordered index assigned to each group after sorting. Helped order bars in plots.

- Race_palette: categorical mapping to colors based on group.

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?

- Racetrack (radial) plot: I initially considered, but the highest NRI groups dominated and skewed the visualization, hiding the smaller values and log10 scaling didn't help so I left it.

- Lollipop chart: Visually looked similiar to a bar chart but didn't provide any additional clarity.

- Stacked bar plot: would clearly show distribution and magnitude but visually it was less engaging.

- Ultimately I chose circular barplot because you can see the relative magnitudes of cumulative weighted NRI across groups. Even with the wide differences in values the circular layout made it easier to interpret the scale and rank order of exposure.

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

Hispanic or Latino population bear the highest cumulative hazard burden, due to a larger population and residence in higher-risk counties. Smaller populations such as American Indian/ Alaska Native and Native Hawaiian/ Pacific Islander communities may face high per-capita risk even if their total weighted NRI appears lower.

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

- Reordered categories by descending weighted NRI to emphasize magnitude and rank.
- Adjusted margins and spacing to prevent crowding and improve balance.
- Adjusted font sizes for text for clarity.
- Selected distinct color palette to differentiate groups clearly.

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

- I wanted to make the circular barplot larger because there is a lot of white space in the figure, but I couldn't figure out how to fully adjust the plot size in combination with ylim, plot margins, and saving the figure in ggplot2. I experimented with all of them but couldn't get a larger plot.