HW #3: Visualizing FEMA NRI x ACS Data

Vedika Shirtekar

2/11/26

Build visualization

Create a data viz that helps to answer the question, *How does climate hazard risk exposure vary across racial / ethnic groups in California?* This will require some data wrangling first (including joining the NRI and ACS data).

```
library(tidyverse)
library(here)
#....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.

#.................Step 3: read in your CSV file.................
```

```r
race_ethnicity <- readr::read_csv(here::here("data", "ACS-1yr-2023-county-race-ethnicity.c
nri <- read_csv(here("data", "National_Risk_Index_Counties_807384124455672111.csv"))
```

```r
## Filter for county score in California
nri_ca <- nri %>% filter(`State Name Abbreviation`  == "CA")

## Create table to understand which code is associated with each race
race_ethnicity %>%
  select(variable, label, concept) %>% ## Select variable (group code), label (racial grou
  distinct() %>% ## Ignore duplicates of the same group code and label
  arrange(label, concept) ## Arrange in alphebetical order
```

```
# A tibble: 10 x 3
   variable   label                                                    concept
   <chr>      <chr>                                                    <chr>
 1 B01003_001 Estimate!!Total                                          Total ~
 2 B02001_004 Estimate!!Total:!!American Indian and Alaska Native alone Race
 3 B02001_005 Estimate!!Total:!!Asian alone                            Race
 4 B02001_003 Estimate!!Total:!!Black or African American alone        Race
 5 B03002_012 Estimate!!Total:!!Hispanic or Latino:                    Hispan~
 6 B02001_006 Estimate!!Total:!!Native Hawaiian and Other Pacific Islan~ Race
 7 B03002_002 Estimate!!Total:!!Not Hispanic or Latino:                Hispan~
 8 B02001_007 Estimate!!Total:!!Some Other Race alone                  Race
 9 B02001_008 Estimate!!Total:!!Two or More Races:                     Race
10 B02001_002 Estimate!!Total:!!White alone                            Race
```

```r
## Based on previous table, filter for racial groups of interest based
## on the associated code
race_ethnicity <- race_ethnicity %>%
  filter(variable %in% c(
    "B03002_012", ## Hispanic or Latino
    "B02001_002", ## White
    "B02001_003", ## African American
    "B02001_004", ## American Indian and Alaska Native
    "B02001_005", ## Asian
    "B02001_006", ## Native Hawaiian and Other Pacific Islander
    "B02001_007", ## Some other race
    "B02001_008" ## Two or more races
  )) %>%
```

```r
  ## Create a new variable to store racial groups as levels
  mutate(
    race = case_when(
    variable == "B03002_012" ~ "Hispanic or Latino",
    variable == "B02001_002" ~ "White",
    variable == "B02001_003" ~ "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"
    )
  ) %>%

  ## Select for GEOID as well to perform left join with nri_ca
  select(GEOID, NAME, race, estimate)
```

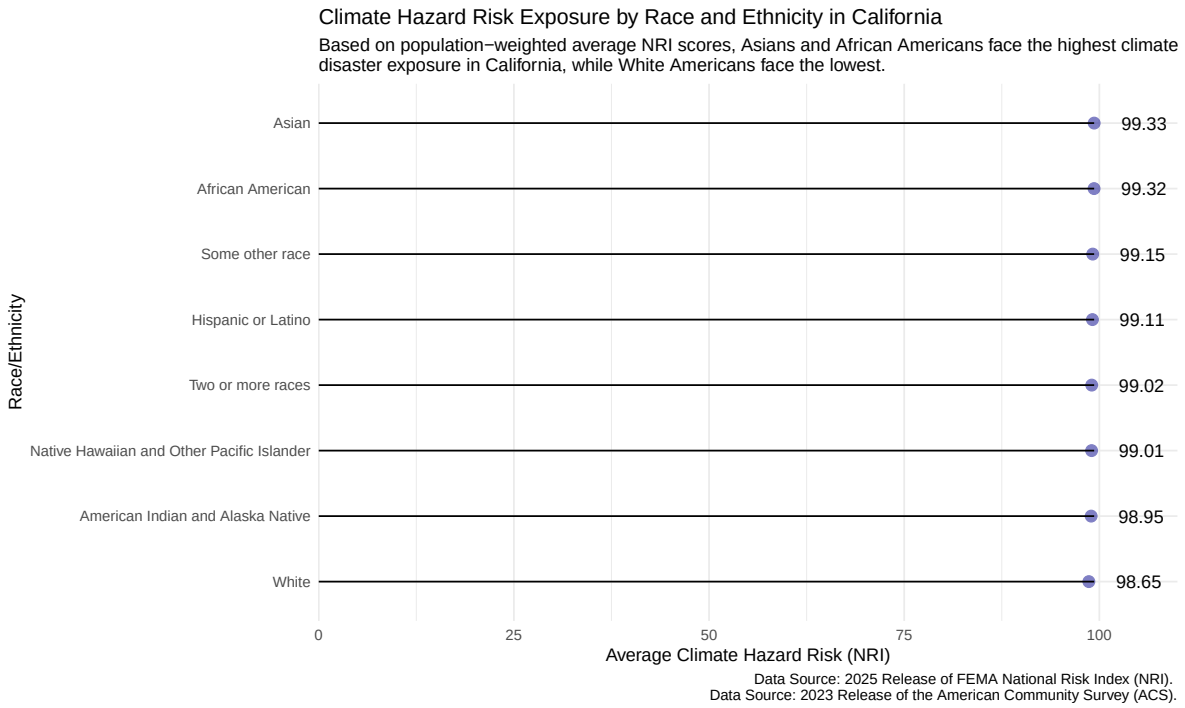
```r
## Select and rename state-county code and composite NRI score
nri_ca <- nri_ca %>%
  select(
    fips = `State-County FIPS Code`,
    nri_score = `National Risk Index - Score - Composite`
  )

## Perform merge on GEOID
race_nri <- race_ethnicity %>%
  left_join(
    nri_ca,
    by = c("GEOID" = "fips")
  )
```

```r
## Use a weighted score based on population estimate per county to
## obtain a representative NRI score that accounts for "misbalance" between
## more populous and less populous counties
race_nri %>%
  group_by(race) %>%
  summarize(
    weighted_nri = weighted.mean(nri_score, estimate, na.rm = TRUE)
  ) %>%

## Use a ranked lollipop plot to showcase the weighted score per racial group
```

```r
ggplot(aes(
  x = weighted_nri,
  y = reorder(race, weighted_nri) ## Reorder by descending weighed score
)) +
  geom_point(
    size = 3,
    color = "darkblue",
    alpha = .5
  ) +
  geom_segment(aes(x = 0, xend = max(weighted_nri), y = race, yend = race)) +
  scale_x_continuous(
    limits = c(0, 100), ## Based on min and max weighted score
    breaks = seq(from = 0, to = 100, by = 25), ## Allow at least 10 units of space
    expand = expansion(mult = c(0, 0.1))
    )+
  geom_text(
    aes(label = round(weighted_nri, 2)),
    hjust = -0.6, ## Too positive is too far left for horizontal text placement
    size = 4
  ) +
  labs(
    title = "Climate Hazard Risk Exposure by Race and Ethnicity in California",
    subtitle = "Based on population-weighted average NRI scores, Asians and African Americ
    x = "Average Climate Hazard Risk (NRI)",
    y = "Race/Ethnicity",
    caption = "Data Source: 2025 Release of FEMA National Risk Index (NRI). \nData Source:
    alt = "Ranked lollipop plot showing composite National Risk Index (NRI) scores for rac
  ) +
  theme_minimal()
```



Answer 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)?
  - `State-County FIPS Code` (categorical)
  - `National Risk Index - Score - Composite` (numeric)
  - `variable` (categorical)
  - `label` (categorical)
  - `GEOID` (treated as categorical)
  - `NAME` (categorical)
  - `estimate` (numeric)
- **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 decided to create a ranked lollipop plot to address the research question because the focus centered on comparing the average climate hazard risk exposure (composite NRI

score) across racial and ethnic groups rather than examining distributions or spatial patterns among these groups. The lollipop plot is appropriate for this purpose because it clearly emphasizes the relative differences and ordering across categories while being visually simple and easy to interpret. By displaying the population-weighted average NRI scores, the lollipop plot chart allows to quickly identify which groups experience higher or lower average exposure. While a bar graph would have conveyed the information similarly, it would place an emphasis on area. A distribution based plot, such as a boxplot, would display the spread of county-level risk scores within each group but would not explicitly communicate differences in the average exposure across groups. As a result, I opted to address the question with a horizontal lollipop plot as this kind of chart provides a direct and interpretable representation for understanding how the NRI score varies across each group when represented as a single quantity.

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

In California, Asian (99.33) and African American (99.32) populations experience the highest average climate hazard risk exposure, based on population-weighted NRI scores that account for differences in county population size and reflect where people in each group actually live. In contrast, White populations experience the lowest average climate hazard risk exposure (98.65).

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

To improve legibility, I applied a minimal theme (`theme_minimal()`) to reduce visual complexity, added direct value labels using `geom_text()` such that each population-weighted NRI score can be read without relying solely on the axis, and ordered racial and ethnic groups in descending order to ensure comparisons and ranking across groups are easier to interpret.

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

One feature I considered implementing was visually representing uncertainty in climate risk exposure for each racial and ethnic group, such as overlaying distributions of county-level NRI scores. However, because the analysis relies on county-level National Risk Index (NRI) scores and incorporates population to account for differences in county population size, I was not sure how to incorporate uncertainty in an appropriate way without overstating precision.