

National Risk Index and ACS in California

AUTHOR
Peter Vitale

PUBLISHED
February 11, 2026

Load Libraries

```
pacman::p_load('here',  
               'janitor',  
               'tidyverse',  
               'ggthemes',  
               'ggbeeswarm',  
               'patchwork',  
               'gghighlight',  
               'ggribes')
```

Read Data with api key

```
#....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(ac_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.....  
race_ethnicity <- readr::read_csv(here::here("data", "ACS-1yr-2023-county-race-ethnicity"),  
  clean_names() %>%  
  rename(nri_id = geoid)
```

Read NRI csv

```
nri_df <- read_csv(here('data',
                        'National_Risk_Index_Counties_807384124455672111.csv')) %>%
  clean_names()
```

Clean NRI Data

```
nri_clean <- nri_df %>%
  select(c(1,2,3,4,6,7,10, 15)) %>%
  clean_names() %>%
  filter(state_name == 'California') %>% # Keep only California
  separate(national_risk_index_id,
            into = c(NA, 'nri_id'),
            sep = 'C')
```

Join NRI and race_ethnicity

```
nri_acs_joined <- left_join(nri_clean,
                            race_ethnicity,
                            by = 'nri_id') %>%
  filter(!is.na(name))
```

Create a data viz that helps to answer the question, How does climate hazard risk exposure vary across racial / ethnic groups in California?

```
nri_county_race <- nri_acs_joined %>%
  mutate(label = case_when( # Mutate the groupings to more easy to use labels
    label == "Estimate!!Total" ~ 'Overall Population',
    label == "Estimate!!Total:!!White alone" ~ 'White',
    label == "Estimate!!Total:!!Black or African American alone" ~ 'Black',
    label == "Estimate!!Total:!!American Indian and Alaska Native alone" ~ 'Native',
    label == "Estimate!!Total:!!Asian alone" ~ 'Asian',
    label == "Estimate!!Total:!!Native Hawaiian and Other Pacific Islander alone" ~ 'Paci',
    label == "Estimate!!Total:!!Two or More Races:" ~ 'Multi-Ethnic',
    label == "Estimate!!Total:!!Hispanic or Latino:" ~ 'Hispanic',
    TRUE ~ 'Other Race'
  )) %>%
  # Remove where those labels are of the entire population in a county
  filter(label != 'Overall Population')
```

Make the plot

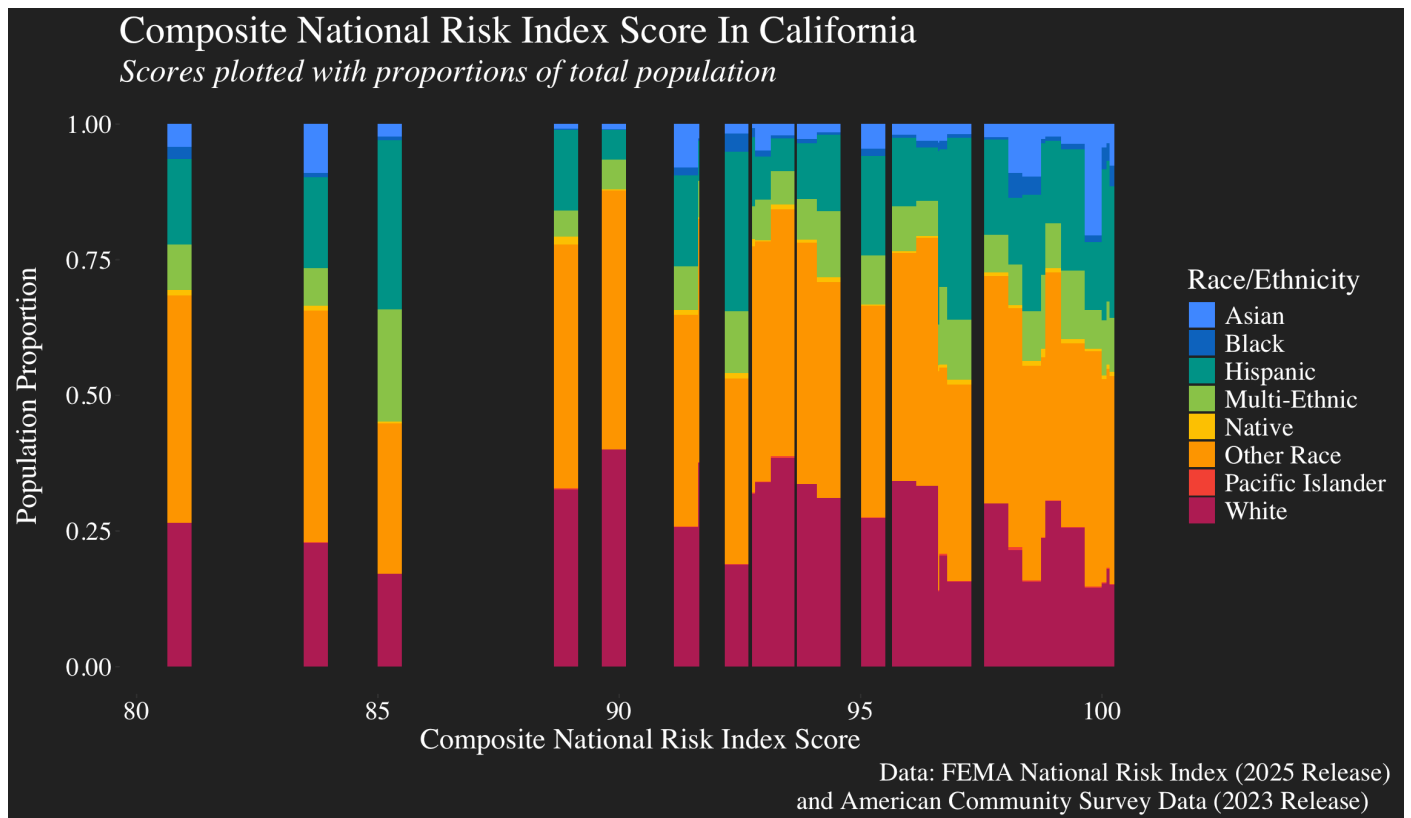
```
ggplot(nri_county_race,
       aes(y = estimate,
           x = national_risk_index_score_composite,
```

```

    width= .5,
    fill = label)))+
geom_histogram(position="fill", # I use fill and identity to make a proportion
               stat="identity")+
scale_fill_manual(values = c("#448aff", # Scale the color with colors color scale
                             "#1565c0",
                             "#009688",
                             "#8bc34a",
                             "#ffc107",
                             "#ff9800",
                             "#f44336",
                             "#ad1457")))+

# Update the labels
labs(x = 'Composite National Risk Index Score',
     y = 'Population Proportion',
     fill = 'Race/Ethnicity',
     title = 'Composite National Risk Index Score In California',
     subtitle = 'Scores plotted with proportions of total population',
     caption = 'Data: FEMA National Risk Index (2025 Release) \n and American Community
theme_tufte()+
# I want dark colors so I change the background to black and the text to white
theme(plot.background = element_rect(fill = "#222222",
                                     color = "#222222"),
      plot.title = element_text(color = 'white',
                                size = 24,
                                family = 'serif'), # Times new roman
      plot.subtitle = element_text(color = 'white',
                                   size = 20,
                                   face = "italic",
                                   family = 'serif'),
      axis.title = element_text(color = 'white',
                                size = 18,
                                family = 'serif'),
      legend.title = element_text(color = 'white',
                                   size = 18,
                                   family = 'serif'),
      plot.caption = element_text(color = 'white',
                                  size = 16,
                                  hjust = 1.45),
      axis.title.y = element_text(margin = margin(r = 15,
                                                  unit = "pt")), # Add margin on x
      axis.text = element_text(color = 'white',
                               size = 16),
      legend.text = element_text(color = 'white',
                                 size = 16))

```



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)?

I am interested in the `label`, `national_risk_index_score_composite`, and `estimate`

- `label`: The categorical character race/ethnicity label provided by the ACS
- `national_risk_index_score_composite`: Numeric county specific index score.
- `estimate`: Numeric estimate of population in each `label`

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 played around a couple of ideas. My first thought was a streamgraph, however the inputs to that need a date. I could have also gone in the donut direction but I had a vision of that streamgraph aesthetic.

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

There is no discernable trend in population makeup with increasing NRI score.

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

I changed the font family, as well as making the plot and QMD a dark background. I did this because I wanted to use a bright color palette for the histogram, but when rendered it was almost blinding. Dark backgrounds often help me in terms of readability, so I hope the effect is the same to others.

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

Like I mentioned, I wasn't able to make a streamgraph due to not having a date. I tried to coerce the counties into dates (like days of the year) but that was not a scalable solution. Ultimately I shifted gears, but I am counting down the days until I can run a streamgraph.