

HW2: NIR Data

William Mullins

2026-01-29

Load Required Packages

```
library(tidyverse)
library(here)
library(gggridges)
library(sysfonts)
library(showtext)
library(forcats)
library(scales)
```

Read in and Clean NRI Data

```
# Read in Raw Data
nri_raw <- read.csv(here::here("data", "National_Risk_Index_Counties_807384124455672111.csv"),
  janitor::clean_names())

# Remove Territories and Add Column to Distinguish California
nri_clean <- nri_raw %>% filter((state_fips_code <= 56) & (state_fips_code != 11)) %>%
  mutate(highlight = ifelse(state_name == "California", "California", "Not California"))

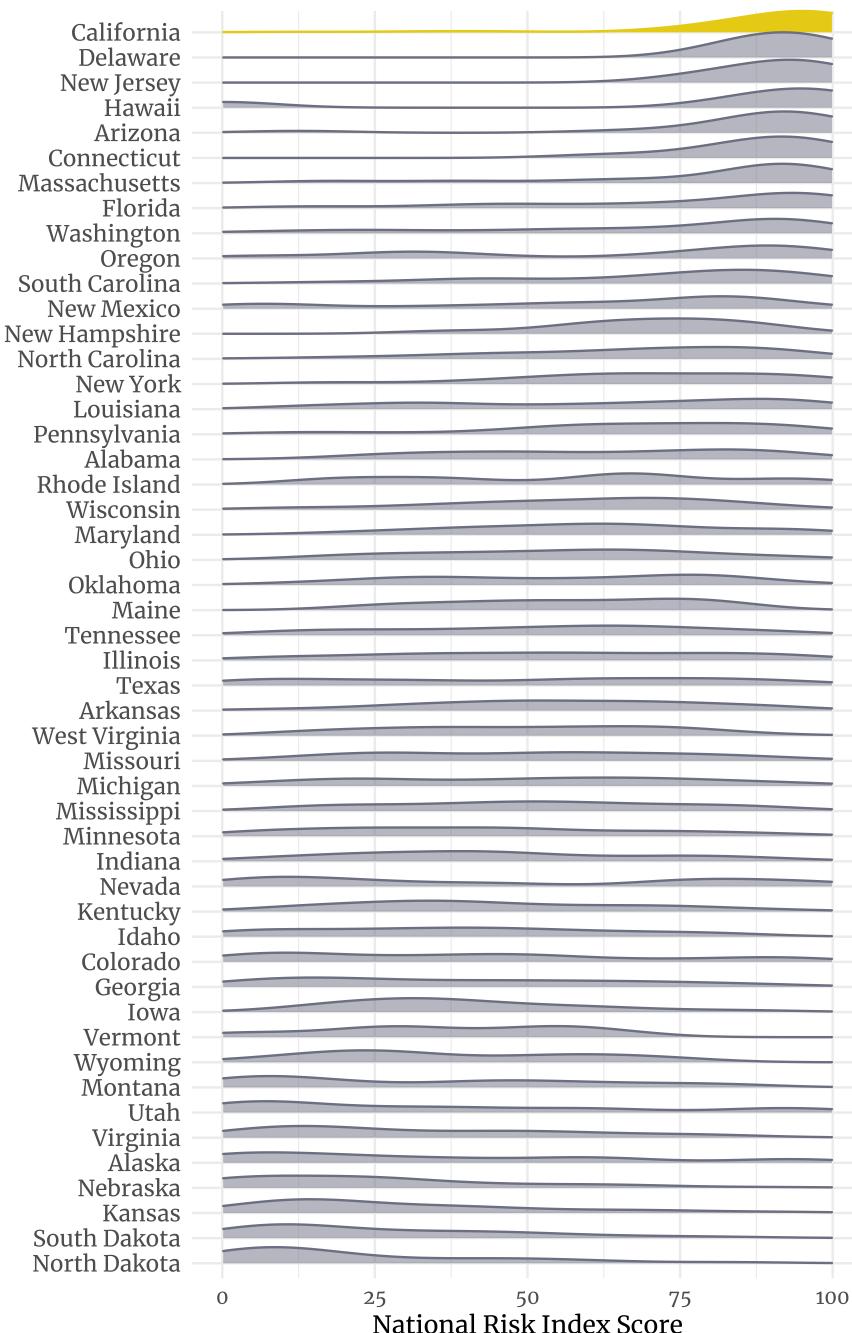
# Add Font For Styling
font_add_google(name = "Merriweather", family = "Merriweather")
showtext_auto()
```

Ridgeline Plot

```
ggplot(nri_clean,
       aes(y = fct_reorder(state_name,
                            national_risk_index_score_composite,
                            .fun = median,
                            na.rm = TRUE),
           x = national_risk_index_score_composite,
           fill = highlight,
           color = highlight,
           alpha = highlight)) +
  scale_x_continuous(limits = c(0, 100)) +
  geom_density_ridges(scale = 1) +
  labs(title = "California Counties Face the Highest Risk \nto Natural Hazards",
       subtitle = "Distribution of County NRI Scores by State With California \nHighlighted",
       caption = "Data: FEMA National Risk Index (2025 Release)",
       x = "National Risk Index Score",
       y = NULL) +
  scale_alpha_manual(values = c("California" = 1, "Not California" = 0.5)) +
  scale_fill_manual(values = c("California" = "#E4CA14", "Not California" = "#6C6F81")) +
  scale_color_manual(values = c("California" = "#E4CA14", "Not California" = "#6C6F81")) +
  coord_cartesian(clip = "off") +
  theme_minimal() +
  theme(plot.title = element_text(face = "bold", size = 16, lineheight = 1),
        plot.title.position = "plot",
        plot.subtitle = element_text(color = "gray10", size = 12, lineheight = 1),
        plot.caption = element_text(hjust = 1, color = "grey40", size = 8),
        legend.title = element_blank(),
        axis.text.y = element_text(size = 10),
        legend.position = "none",
        text = element_text(family = "Merriweather"))
```

California Counties Face the Highest Risk to Natural Hazards

Distribution of County NRI Scores by State With California Highlighted in Yellow



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)

- National Risk Index Composite Score: A numeric variable, ranging from 0 to 100, representing the cumulative risk of a community to the 18 natural hazards included in the index.
- State: Nominal categorical variable used to group together counties

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?

When choosing a graphic for this assignment I wanted to pick something that would allow for me to display the data from all of the states without introducing too much visual clutter. To do this I considered doing either a ridgeline plot, as I ended up doing, or a column graph showing the median NRI scores by county. I decided on the ridgeline plot as I think being able to see the distribution of the values is more interesting than just seeing the median.

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

California has the highest median National Risk Index score among all 50 US states, with most of the counties having a score above 75. This indicates that counties in California face substantially risk from natural hazards when compared to counties in other states.

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

- I increased the aspect ratio to make the graph taller. This allows for the ridgelines to be more easily visible.
- I formatted all of the text by adjusting the size and color.
- I used color blind friendly yellow and gray values when highlighting the data.
- I removed the legend and y axis labels as they do not add anything to the plot. Doing this helps remove clutter from the graph.
- I ordered the states by median NRI score

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

I did not have anything that I wanted to implement that I wasn't able to. I'm fairly happy with how this graph turned out.