US County FEMA Risk Rating and Index Analysis

Richard Montes Lemus

2026-01-29

Analysis Description: The following analysis leverages the Federal Emergency Management Agency's (FEMA) risk index and rating data to visualize US state's county's potential for negative impacts due to natural disasters. It specifically compares California county's risk index and rating to the rest of the United States' counties.

```
# Load in Necessary Libraries
library(tidyverse)
library(janitor)
library(here)
library(ggtext)
```

Load in and Clean Data

```
# Filter out non-US state rows
nri <- read_csv(here("data", "National_Risk_Index_Counties.csv")) %>%
  clean_names() %>%
  filter(!state_name %in% c("Puerto Rico",
                            "District of Columbia",
                            "Guam",
                            "Northern Mariana Islands",
                            "American Samoa",
                            "Virgin Islands"))
```

Conduct Summary Statistics Across States Groups

```r
# Calculate the percentage of counties in each risk rating category per state
heatmap_nri <- nri %>%
  group_by(state_name_abbreviation, national_risk_index_rating_composite) %>%
  count() %>%
  ungroup() %>%
  group_by(state_name_abbreviation) %>%
# Calculate percentage of counties in each category
  mutate(prop_risk = n/sum(n)*100) %>%
  ungroup() %>%
# Fill in empty cells for risk rating with zero
  complete(state_name_abbreviation,
           national_risk_index_rating_composite,
           fill = list(n = 0, prop_risk = 0))
```

```r
# Generate a column of ordered states for re factoring heat map data
nri_ordered <- nri %>%
  group_by(state_name_abbreviation) %>%
# Calculate median risk percentile per state
  summarise(p_risk_median = median(national_risk_index_score_composite)) %>%
# Order based on median risk percentile
  arrange(p_risk_median) %>%
  mutate(order = row_number()) %>%
  select(state_name_abbreviation, order)
```

Reorder Factor Levels

```r
# Reorder columns in order to detect heat map patterns more easily
heatmap_order <- heatmap_nri %>%
  left_join(nri_ordered) %>%
# Reorder states based on each state's median risk percentile
  mutate(state_name_abbreviation = fct_reorder(.f = state_name_abbreviation,
                                               .x = order)) %>%
# Reorder ratings based on increasing levels
  mutate(national_risk_index_rating_composite = factor(national_risk_index_rating_composite,
                        levels = c("Very Low", "Relatively Low",
                                   "Relatively Moderate",
```

```
                                              "Relatively High",
                                              "Very High")))
```

Create Heatmap

```
# Create a heat map demonstrating risk patterns across US counties per state
heatmap_order %>%
  ggplot(aes(x = national_risk_index_rating_composite,
             y = state_name_abbreviation, fill = prop_risk)) +
  geom_tile() +
  scale_fill_viridis_c() +
  annotate("text", x = 3, y = "CA",
           label = "← California",
           hjust = -4,
           fontface = "bold",
           size = 4,
           color = "#9A1311") +
  coord_cartesian(clip = "off") +
  labs(title = "US County Risk Rating and Index by State",
       subtitle = "California has the highest risk index percentile median of all US states a
       caption = "**Figure 1:** The following heatmap shows the percentage of counties per st
       x = "Risk Rating",
       y = "State",
       fill = "% of Counties") +
  theme_minimal() +
  theme(plot.margin = margin(5, 80, 30, 5,),
        plot.caption = element_markdown(hjust = 0, size = 12),
        plot.title = element_text(face = "bold"),
        axis.title.x = element_text(face = "bold"),
        axis.title.y = element_text(face = "bold"))
```

US County Risk Rating and Index by State

California has the highest risk index percentile median of all US states and visually appears to have more counties in high risk rating categories
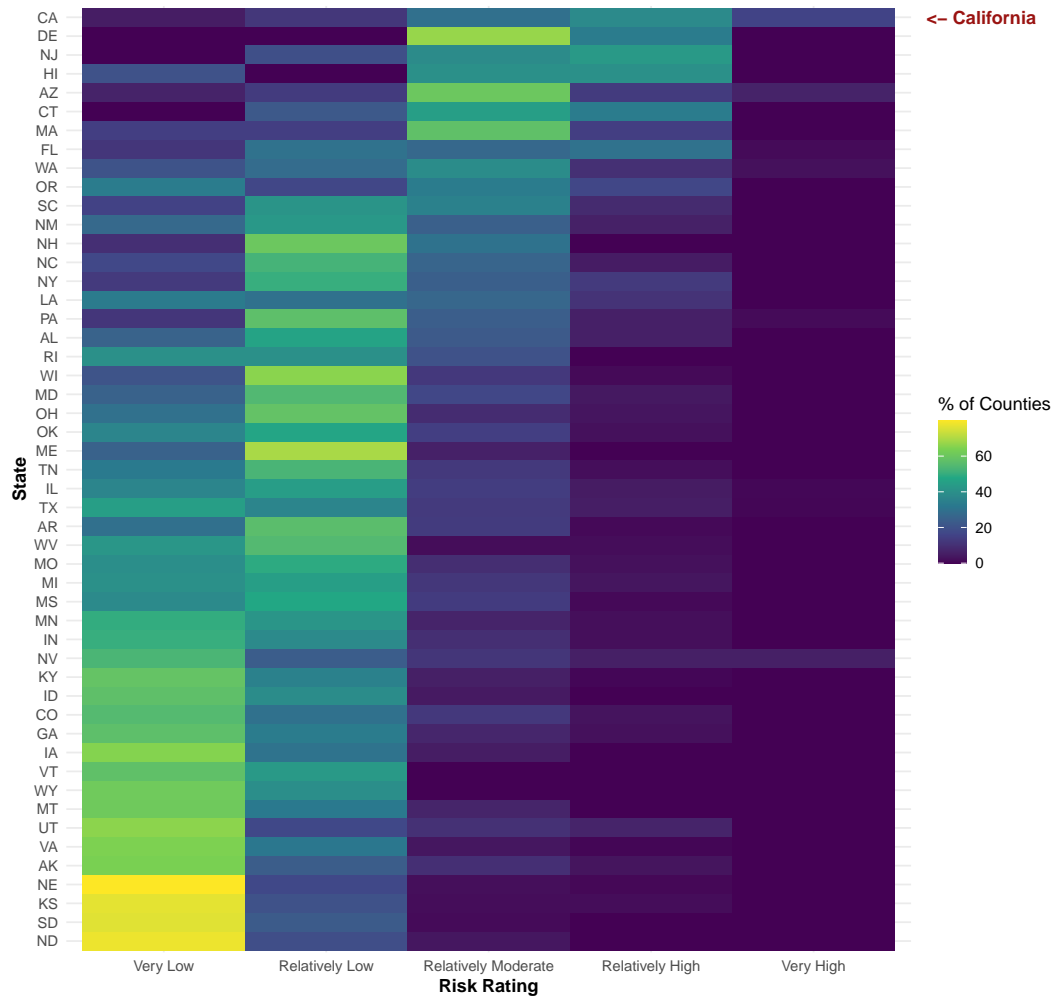
Figure 1: The following heatmap shows the percentage of counties per state in each FEMA risk rating category by color. States are ordered from largest to smallest based on their FEMA risk index percentile median.

Questions about the analysis:

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)?

- The first variable is "national_risk_index_rating_composite", this is categorical and ordered.

- The second variable is "state_name_abbreviation" this is categorical.

- The third variable is "national_risk_index_score_composite", this is 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? First, I decided what variables I wanted to use for this analysis. I knew I had to represent county risk by state, so the state column was necessary. Then I came across the risk rating column and thought that would be a great way to compare risk across states since its normalized. Then I visited the data to viz site and looked through the plots types listed as representing multiple categorical variables. As I was looking through them I remembered how well the UFO info-graphic represented sightings across US states using shaded tiles. I realized a heat map would be able to represent a similar amount of information without being overwhelming and decided that was the one I wanted to use. After plotting it I wanted to order the states in a way that showed a pattern on the heat map so I choose the risk index percentile column for this. I could have also used two donut plot showing risk rating across all 5 levels for California and another one for the rest of the US states. Grouping all the remaining US states for this visualization, however, would have caused my plot to lose state to state comparisons. Ultimately, I chose the heat map because allowed to represent all US states' risk at once without it looking too cluttered.

3. Summarize your main finding in no more than two sentences. This analysis showed that California is the state with the highest risk index percentile median for its counties. Visually through shading, it also showed that California tends to have a greater percentage of its counties in higher risk rating categories when compared to the rest of the US.

4. What modifications did you make to this visualization to make it more easily readable? I added an annotation that pointed to the row containing California (our reference for this analysis). Furthermore, I ordered the Y and X axis so they both started at low values near the origin and increased outward, this reduces eye movement for the viewer. I also adjusted the aspect ratio and width so the X and Y axis labels weren't cluttered.

5. Is there anything you wanted to implement, but didn't know how? If so, please describe. I wanted to flip the legend color bar title horizontally so it didn't stick out so awkwardly but I struggled to figure out how to do it.