HW2: FEMA National Risk Index Visualization
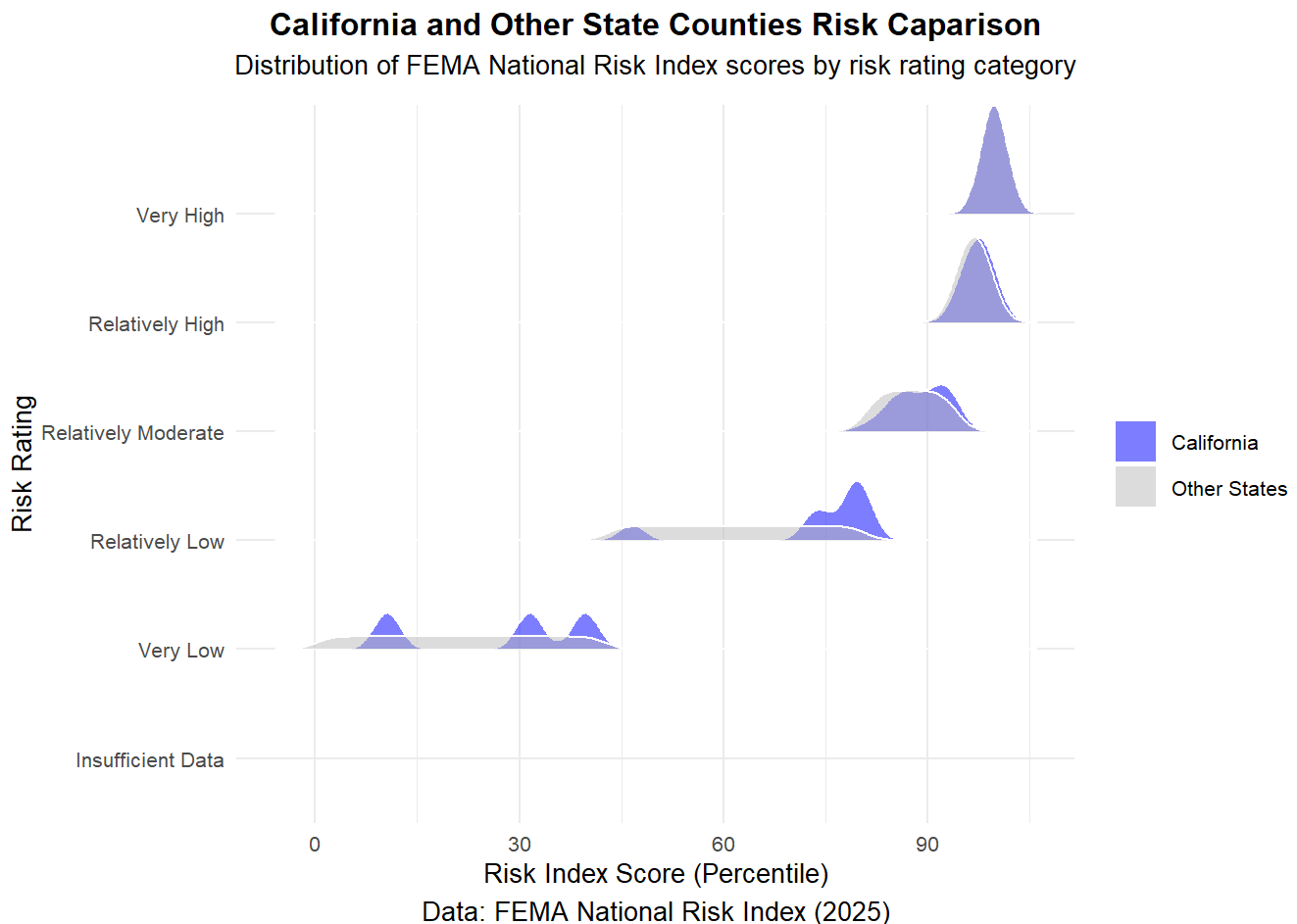
AUTHOR
Joshua Ferrer-Lozano

```r
# Import libraries
library(tidyverse)
library(janitor)
library(ggridges)
```

```r
# Read in Data
nri <- read.csv('data/National_Risk_Index_Counties.csv')
# Inspect data head
head(nri)
```

```r
# Clean data names and select columns of interest
nri_clean <- nri |> clean_names() |>
  select(
  state = state_name_abbreviation,
  county = county_name,
  risk_score = national_risk_index_score_composite,
  risk_rating = national_risk_index_rating_composite
  ) |>
# Group by state of comparison
  mutate(
  group = if_else(state == "CA", "California", "Other States"),
  risk_rating = factor(
  risk_rating,
  levels = c(
    "Insufficient Data","Very Low","Relatively Low","Relatively Moderate","Relatively High", "Very
    )
  )
)
```

```r
# Plot ridge plot
ggplot(nri_clean, aes(x = risk_score, y = risk_rating, fill = group)) +
  geom_density_ridges( alpha = 0.5,
                      scale = 1, color = "white" ) +
  scale_fill_manual( values = c( "California" = "blue",
                                "Other States" = "grey"
                                )) +
# Title/label plot
  labs(title = "California and Other State Counties Risk Caparison",
       subtitle = "Distribution of FEMA National Risk Index scores by risk rating category",
       x = "Risk Index Score (Percentile)",
       y = "Risk Rating",
```

```
        caption = "Data: FEMA National Risk Index (2025)" ) +
# Adjust text elements
  theme_minimal(base_size = 10) +
  theme(legend.title = element_blank(),
        plot.title = element_text(face = "bold", hjust = .5),
        plot.subtitle = element_text(margin = margin(b = 10), hjust = .5),
        plot.caption = element_text(size = 10, hjust = .5))
```


California and Other State Counties Risk Caparison
Distribution of FEMA National Risk Index scores by risk rating category

Data: FEMA National Risk Index (2025)

Questions Section

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 variables of interest include: - State: Categorical (Qualitative/string) - County: Categorical (Qualitative/string) - Risk Score: numerical (float/percentile) - Risk Rating: categorical (ordered classes) - Grouping column "group": categorical data

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 selected a ridge plot because it shows how the distribution of National Risk Index scores vary across the risk categories based on two groups: "California" and "Other States". Moreover, in this figure type we are able to compare how the ridges are similar or different relative to other based on risk level.

The alternative graphics include: - boxplots where comparisons of distributions are possible - violin plots where density can be shown

I chose the ridge plot because it shows multiple "dimensions" of the data, allowing me to see how California counties differ across risk categories and other state counties.For me, the distribution and shape of data was crucial to compare California to other state counties.

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

Based on the visualization, we can see that the distributions for California and other state counties are similar in shape of distribution for risk categories "Relatively Moderate", Relatively High" and "Very High". In the category "Very Low" we see about 3 small "distributions" for California, centering around levels 10, 30 and 40 index scores, whereas other counties tend to be spread throughout the "Very Low" category, there is similar shape of distributions in the "Relatively Low" category but with a higher scoring bracket.

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

To improve readability, I used custom colors for California and the other states as well as adjusted the opacity to better see overlapping distributions. Font size and limited text elements were used to focus on the bare minimum needed information to improve readability. Scale, base_size and caption hjust were adjusted to ensure centered text.

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

Something interesting to implement would be to create hazard-specific plots for various risks to view how certain risks contribute California's overall risk and measure the magnitude of loss for each hazard.