HW2

Sarayu Ramnath

2026-01-25

Objective

Create a data viz that helps to answer the question, How do FEMA National Risk Index scores for counties in California compare to those in other states?

Sketching out the Idea

- Pick one clear measure to compare California counties with other states

    - Community resilience score: which could be interesting because a low resilience score can make a moderate hazard risk even worse.

    - Visual Objective: Show how CA counties rank relative in community resilience score to all the other counties in the United States. maybe histogram to show density

Data Wrangling

```
library(tidyverse)
library(ggplot2)
library(janitor)

nri_raw <- read_csv("data/nri_counties.csv") %>%
  clean_names()

nri_tidy <- nri_raw %>% #keep state, county, population, community scores and percentile
  select(
    state = state_name,
    county = county_name,
    population = population_2020,
```

```r
    resilience_score = community_resilience_score,
    resilience_rating = community_resilience_rating,
    resilience_percentile = community_resilience_state_percentile
  )

#filter out territories and keep only US states
us_states <-c(state.name) #R vector of US state names

nri_us <-nri_tidy %>%
  filter(state %in% us_states) %>%
  mutate(
    ca_county = if_else(state == "California", "California Counties", "Rest of U.S. Counties"
  )

#doing boxplot comparing CA Counties resilience risk to rest of US means that I should have a

# Population-weighted mean per group
weighted_means <- nri_us %>%
  group_by(ca_county) %>%
  summarise(weighted_score = weighted.mean(resilience_score, population, na.rm = TRUE))
```
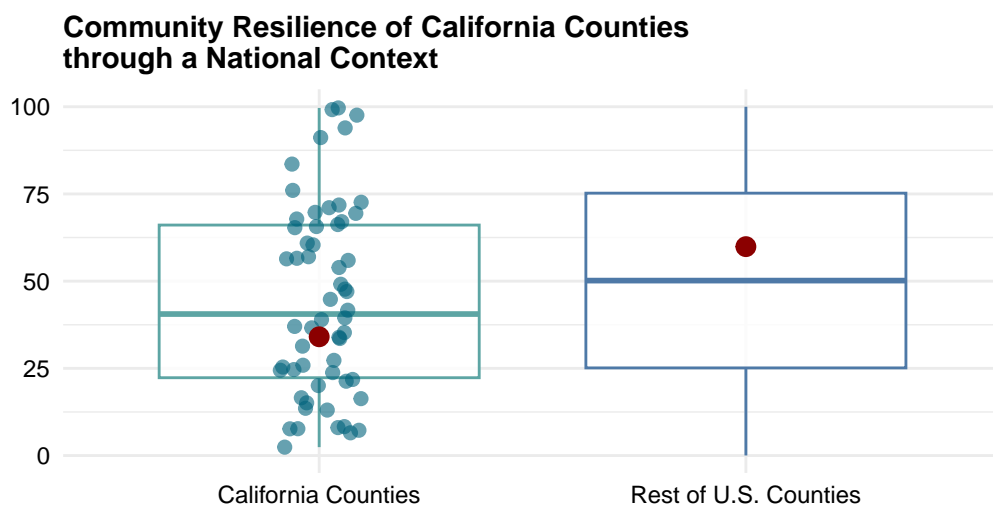
Box Plot

```r
ggplot(nri_us, aes(x = ca_county, y = resilience_score, color = ca_county)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.8) +
  geom_jitter(
    data = filter(nri_us, ca_county == "California Counties"),
    color = "#00637C",width = 0.1, height = 0, alpha = 0.6, size = 2
  ) +
  geom_point(
    data = weighted_means,
    aes(y = weighted_score),
    color = "darkred",
    size = 3
  ) +
  scale_color_manual(
  values = c(
    "California Counties" = "#5DA5A4",
    "Rest of U.S. Counties"  = "#4E79A7"))+
    labs(
```

```
    title = "Community Resilience of California Counties \nthrough a National Context",
    x = "",
    y = "",
    caption = "Red points show population-weighted mean, blue dots show individual CA countie
  ) +
theme_minimal()+
theme(
  legend.position = "none",
  axis.text = element_text(color = "black"),
  plot.title = element_text(size = 11, face = "bold"),
  plot.caption = element_text(size = 9)
)
```

**Community Resilience of California Counties
through a National Context**

Red points show population–weighted mean, blue dots show individual CA counties.
California's average community resilience score is lower than the national average.

Data: FEMA National Risk Index (2025 Release)

Alt-Text: Boxplot of community resilience scores (0-100) comparing California counties to the rest of the counties in the United States. California counties are shown as teal jittered points with a light green boxplot, while other U.S. counties are summarized in a light blue. Dark red points show population-weighted means (CA:34; U.S. 60). California's median and overall range are lower than the national distribution.

Question 1 : What are your variables of interest and what kinds of data are they?

- Community resilience score: Numerical (quantitative, continuous). Values range from 0 to 100 and represent magnitude.

- Population-weighted mean resilience score: Numerical (quantitative, continuous). This was a summary statistic taken from the resilience score and population variables in the data set.

- County group (California vs. Rest of U.S. Counties): Categorical.

Question 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?

- Since I was working with a single numeric variable and comparing two groups (CA counties vs the rest of the U.S.) along a 0-100 scale, a box plot made most sense for the ease of reading. I played around with other ways to visualize the community resilience scores, like histograms and ridge plots, but it ended looking too messy and busy.

- Alternative graphic forms I could have included were the aforementioned histograms, violin plots, or dot plots to show the distribution of scores, but I felt this was harder to compare across groups or less practical for highlighting the key points. I liked the boxplot, and I feel like adding the jittered points of individual CA county scores along with a population-weighted mean resilience score for both groups added some good context while showing the median, range, and outliers.

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

- California counties have lower community resilience scores compared to the rest of the U.S., with a median of 38 and narrower overall range (~22-70) versus 25-75 for other counties in the U.S. Population-weighted mean resilience is also lower in CA (34) compared to the national average (60), highlighting a stark gap in resilience.

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

- Overlaid jittered points for individual CA counties and dark red points for population-weighted means to highlight additional context without cluttering the plot.

- Removed the legend, simplified axis labels, used outline colors of a complementary palette to make the differences between California and the other counties visually clear.

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

- I tried adding marginal density plots to show the full distribution of resilience scores for each CA county, with horizontal density curves above the boxplot to show the ranges. But I couldn't get them to display clearly by group or align with the boxplots so I ditched it.