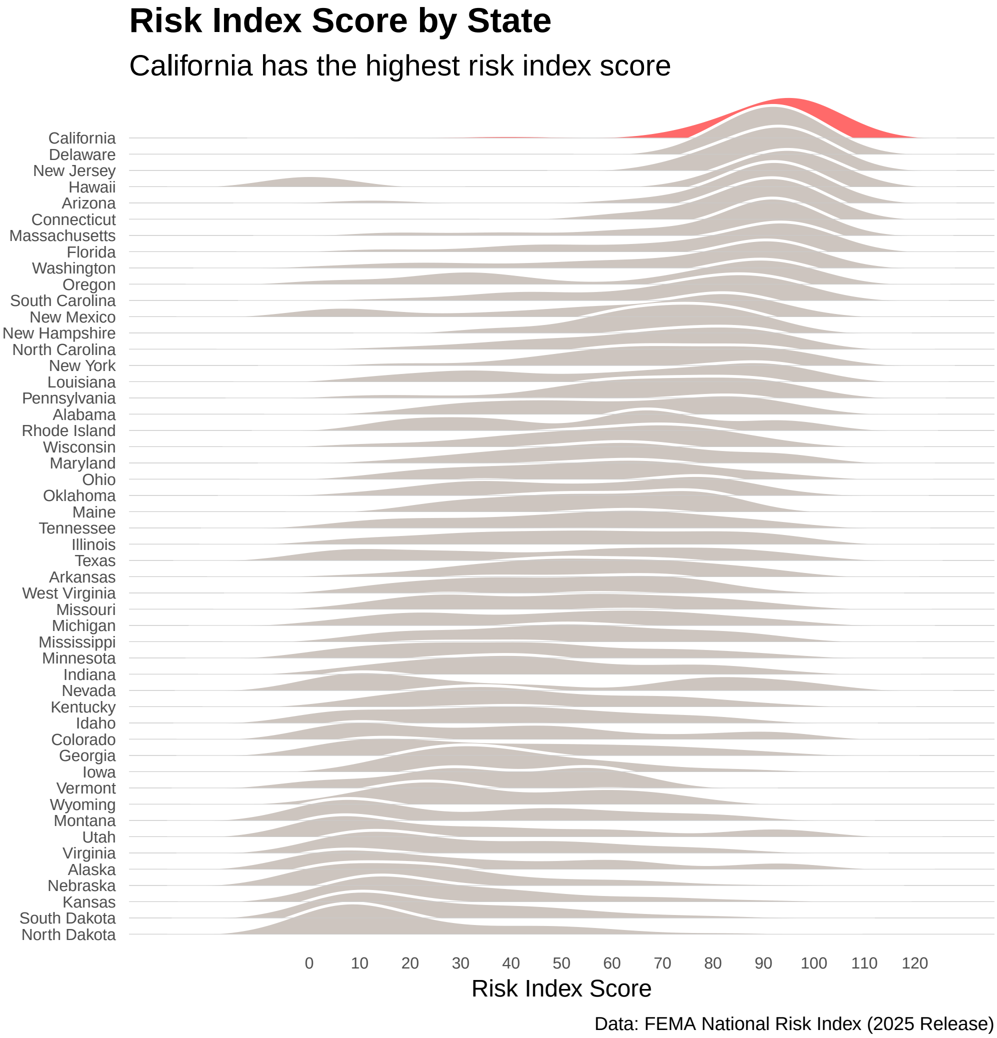
Homework 2

Emily Wapman

```r
# Import libraries
library(tidyverse)
library(ggplot2)
library(here)
library(janitor)
library(ggridges)
library(dplyr)
library(gghighlight)


# Read in data
risk_data <- read_csv(here('data', 'National_Risk_Index_Counties.csv'))

# Clean data
risk_data_clean <- risk_data |>
  clean_names() |>
  select(state_name, national_risk_index_score_composite) |>
  filter(!state_name %in% c("American Samoa", "Guam", "District of Columbia",
                            "Northern Mariana Islands", "Puerto Rico", "Virgin Islands"))

# Data frame with means
risk_data_means <- risk_data_clean |>
  group_by(state_name) |>
  summarise(mean_risk = mean(national_risk_index_score_composite, na.rm = TRUE)) |>
  arrange(desc(mean_risk)) |>
  pull(state_name)



# reorder state_name
risk_data_clean <- risk_data_clean |>
  mutate(state_name = factor(state_name, levels = risk_data_means))
```

```
#| fig-alt: "Ridgeline plot of risk index scores for the 50 US States.
#| Ridges are arranged from lowest risk index score (bottom left) to
#| highest risk index score (top right). California has the highest risk index score
#| and is highlighted in orange."

#Create ridgeline plot

# Add a new column of colors to distinguish CA as orange
risk_data_clean |>
  mutate(
    my_color = case_when(
      state_name == "California" ~ "#FF6A6A",
      TRUE ~ "#CDC5Bf"
    )
  ) |>
ggplot(aes(x = national_risk_index_score_composite,
           y = fct_reorder(state_name, national_risk_index_score_composite, median),
           fill = my_color)) + # reorder the factors based on median to be arranged
                              #from low to high
  geom_density_ridges(rel_min_height = 0.001, scale = 3, color = "white") +
  # use the min height to remove any values lower than 0.001
  scale_fill_identity() +
  # Fill the ridges with the colors designated earlier
  coord_cartesian(clip = "off") + # avoid cutting off the top ridge
  theme_minimal() + # Remove grey background
  scale_x_continuous(breaks = seq(0, 120, by = 10)) + # Add more x axis values
  labs(title = "Risk Index Score by State",
       y = "",
       x = "Risk Index Score",
       subtitle = "California has the highest risk index score",
       caption = "Data: FEMA National Risk Index (2025 Release)") +
  theme(
      panel.grid.major.x = element_blank(),
  panel.grid.minor.x = element_blank(),
  panel.grid.major.y = element_line(color = "grey80", linewidth = 0.1),
  panel.grid.minor.y = element_blank(),
  plot.title = element_text(face = "bold"),
  plot.caption = element_text(size = 7),
  axis.text.x = element_text(size = 6),  # X-axis numbers
  axis.text.y = element_text(size = 6),
  axis.title.x = element_text(size = 9)
  ) +
  #Add these lines to get rid of the blank spaces on the
  #horizontal lines caused by the ridges
    geom_hline(
    yintercept = seq_along(unique(risk_data_clean$state_name)),
    color = "grey80",
    linewidth = 0.1
  )
```



Risk Index Score by State

California has the highest risk index score

Data: FEMA National Risk Index (2025 Release)

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)?

The variables of interest are the state_name and national_risk_index_score_composite. The state_name is a factor and the risk index score 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?

I decided to use this graph because I had one numeric and one categorical variable with several observations per group. I could have also used a boxplot or violin plot. I chose the ridgeline plot because I thought it showed the overall trend well despite the large amount of data. The boxplot appeared too cluttered.

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

The distribution of California county risk index scores is higher than other US states, meaning California counties on average have greater social vulnerability, lower community resilience, and greater expected annual loss due to natural disasters than counties in other states.

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

I removed unnecessary gridlines and the grey background and adjusted font sizes on the axis value labels to make them not run into each other. I also reordered the ridges from low to high medians to emphasize the overall trend.

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

I wanted to try using different fonts.