

Homework 3

Emily Wapman

```
library(janitor)
library(tidyverse)
library(dplyr)
#....Step 1a: see all available ACS variables + descriptions.....
# acs_vars <- tidycensus::load_variables(year = 2023,
#                                         dataset = "acs1")
#
# #.....Step 1b: import race & ethnicity data.....
# race_ethnicity <- tidycensus::get_acs(
#   geography = "county",
#   survey = "acs1",
#   # NOTE: you may not end up using all these variables
#   variables = c("B01003_001", "B02001_002", "B02001_003",
#                 "B02001_004", "B02001_005", "B02001_006",
#                 "B02001_007", "B02001_008", "B03002_012",
#                 "B03002_002"),
#   state = "CA",
#   year = 2023) |>
#   # join variable descriptions (so we know what's what!)
#   dplyr::left_join(acs_vars, by = dplyr::join_by(variable == name))
#
# #.....Step 2: write ACS data to file.....
# readr::write_csv(race_ethnicity,
#                   here::here("data", "ACS-1yr-2023-county-race-ethnicity.csv"))

#.....Step 3: read in your CSV file.....
race_ethnicity <- readr::read_csv(here::here("data",
                                              "ACS-1yr-2023-county-race-ethnicity.csv"))

risk <- readr::read_csv(here::here("data", "National_Risk_Index_Counties.csv"))
```

Create a data vizualization to answer:

How does climate hazard risk exposure vary across racial/ethnic groups in California?

```
# join NRI and ACS data

# Filter out California from risk and remove C before key
risk_ca <- risk |>
  clean_names() |>
  filter(state_name == "California") |>
  mutate(national_risk_index_id = str_remove(national_risk_index_id, "^C")) |>
  select("national_risk_index_id", "national_risk_index_score_composite") |>
  rename(GEOID = national_risk_index_id)

head(risk_ca)

# A tibble: 6 x 2
  GEOID national_risk_index_score_composite
  <chr>                <dbl>
1 06001                 99.8
2 06003                 10.6
3 06005                 79.2
4 06007                 93.9
5 06009                 86.6
6 06011                 81.1

# Join datasets

risk_ethnicity_joined <- inner_join(risk_ca, race_ethnicity, by = "GEOID")

# Clean the ethnicity column
library(stringr)

risk_ethnicity_joined$label <- gsub("^Estimate!!Total:!!", "", 
                                      risk_ethnicity_joined$label)

risk_ethnicity_joined$label <- gsub(":$", "", risk_ethnicity_joined$label)

risk_ethnicity_cleaned <- risk_ethnicity_joined |>
  filter(label %in% c("White alone", "Black or African American alone",
                     "American Indian and Alaska Native alone", "Asian alone",
                     "Native Hawaiian and Other Pacific Islander alone",
```

```

    "Some Other Race alone", "Two or More Races",
    "Hispanic or Latino"))

risk_ethnicity_cleaned$label <- trimws(gsub("\\balone\\b", "",
                                              risk_ethnicity_cleaned$label))

# Find the proportion of the number in each ethnic group
# Out of the total population surveyed for each county
risk_ethnicity_cleaned <- risk_ethnicity_cleaned |>
  group_by(GEOID) |>
  mutate(prop = estimate / sum(estimate)) |>
  ungroup()

# Weighted risk for each group in each county
risk_ethnicity_cleaned <- risk_ethnicity_cleaned %>%
  mutate(weighted_risk = prop * national_risk_index_score_composite)

```

Build Boxplot

```

library(viridis)
library(ggplot2)
highlight_group <- "White"
risk_ethnicity_cleaned <- risk_ethnicity_cleaned %>%
  mutate(box_color = ifelse(label == highlight_group, "red", "black"))

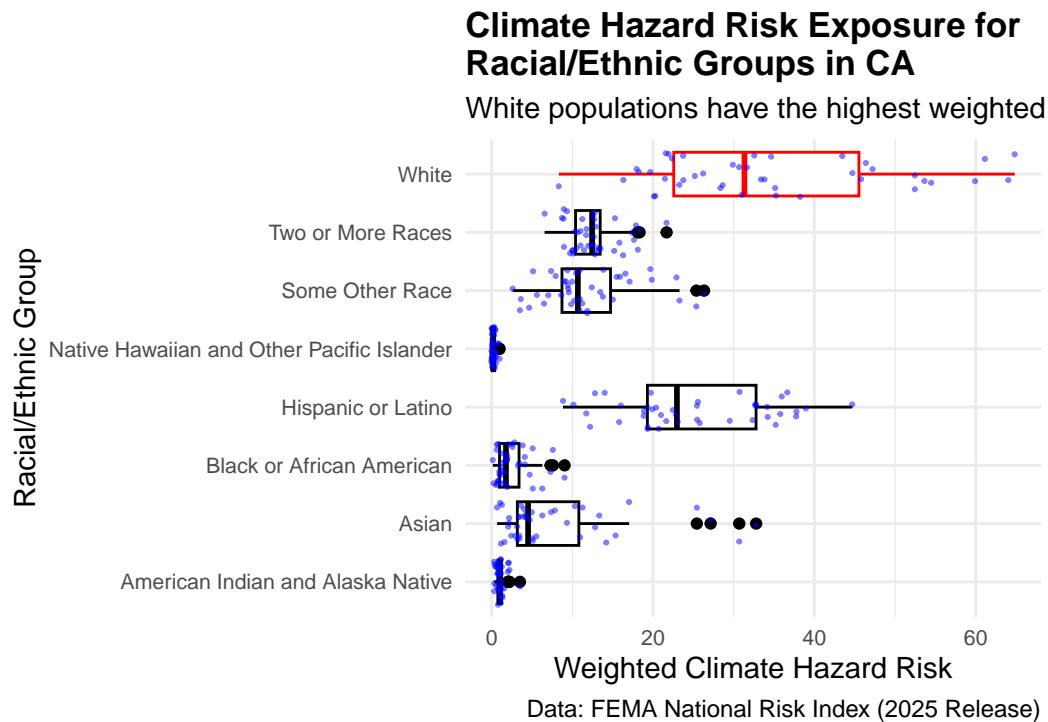
ggplot(risk_ethnicity_cleaned, aes(x = label, y = weighted_risk,
                                    color = box_color)) +
  geom_boxplot() +
  scale_color_identity() +
  labs(
    x = "Racial/Ethnic Group",
    y = "Weighted Climate Hazard Risk",
    title = str_wrap("Climate Hazard Risk Exposure for Racial/Ethnic Groups in
                     CA", width = 40,
                     ),
    subtitle = "White populations have the highest weighted risk index score.",
    caption = "Data: FEMA National Risk Index (2025 Release)") +
  theme_minimal() +
  coord_flip() +
  geom_jitter(color="blue", size=0.4, alpha=0.5) +
  theme(

```

```

plot.title = element_text(face = "bold"),
axis.text.x = element_text(size = 8), # X-axis numbers
axis.text.y = element_text(size = 8)
)

```



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)?
 - Categorical: racial/ethnic groups
 - Weighted hazard risk scores: 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 used the from Data to Viz website. I chose this plot because I had one categorical variable and one numeric variable with multiple observations per category. I could have also done a violin or ridgeline plot.

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

The graph indicates that white populations have a higher mean weighted climate hazard risk.

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

I added the jittered points, adjusted font size, and made the highest group red. I made the points blue and reduced their alpha value so they would be more readable. I also played with the aspect ratio of my layout so that it wouldn't be squished in my final rendered document.

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

I would like to have tried to use different fonts.