Final Project

TOTAL: 10pts

PURPOSE

To understand similarity, better, using multi-dimensional scaling.

GOAL

Analyze the most similar cities from all prior projects in one comprehensive model to identify which cities are truly similar to D.C.

DATA

Build a dataset from the findings from all prior projects:

- Project 2: Firearm violence
- Project 3: Demographics
- Project 4: Hate crime
- Project 5: Public health

Use your findings from each project to construct a city-level dataset. First, create a one-column table of just city names:

```
data.rankings = data.frame("City" = c("Washington DC",
                                      "Baltimore MD",
                                      "Miami FL",
                                      "Atlanta GA",
                                      "Etc",
                                      "Etc etc"))
```

Second, add a column for the rankings of these cities for Project 2:

```
data.rankings$project2 <- c(0,1,2,3,4,5)
```

Third, repeat for projects 3-5.

```
data.rankings$project3 <- c(0,5,4,2,3,1)
```

NOTES:

- Keep the values in order of the list of cities, so everything matches up
- MSAs from Project 3 will need to be "translated" into cities
- Not every city is in each project
    - For example, Baltimore may be a similar city from Project 2, but not Project 4
    - You will need to find an approximate rank for these 'missing' cities across projects
- When complete, your dataset may have as many as 21 total cities/rows
    - Washington, D.C., plus five cities per project
    - Your dataset can have *less* than 21; this will happen if the same city was Top 5 in multiple projects. That's completely fine
- Rankings are fixed, and generic, but usable. If you prefer, you can use specific metrics from the datasets in each project instead.
    - It is not recommended to use both rankings and metrics - this will create likely create spurious relationships across redundant fields
- If you want to re-evaluate any rankings from prior projects and update for this one, go for it.

METHODS

First, here's a completely unrelated example of how multi-dimensional scaling works, but hopefully helps make sense of this concept.

Second, this script is a more relevant class example, based on the data from Project 5.

ANALYSIS

When conducting your analysis, consider this perspective:

1. State a finding
    a. This is the new thing that you've identified from the data
2. Provide your evidence
    a. This is typically a statistical reference
3. Add the context
    a. This is the integration of the stat(s) and the finding
    b. Compare your finding to the norm/average, or a similar measure in the data.

HERE IS THE EXAMPLE SCRIPT FROM CLASS ON 16 APRIL 2025

FORMAT

Submit your script as a usable HTML file, via RMarkdown. You create this output as a .Rmd file in RStudio, but only submit the .html file.

Use this file as a template.

HERE IS THE EXAMPLE MARKDOWN FILE FROM CLASS ON 23 APRIL 2025

SUBMISSION

Once your analysis is complete, please submit your project as either an HTML file (as the output/export/knit of your RMarkdown script), via Canvas.

GRADES

Data: **1pt**

- Provide a table of your created dataset in your Markdown output.

Analysis: **6pts**

- Describe three analytic insights from your data (2pts each).
    - An analytic insight should be 3-4 sentences and describe something new, interesting, unique, or otherwise noteworthy about your data
    - Each insight should focus on at least *one city* that isn't D.C., and at least *two datasets*

Visuals: **2pts**

- One good graph
    - Clean titles, legend, labels, and theme
    - Clearly supports your analysis

Formatting: **1pt**

- Error-free writing. No typos, run-on sentences, or sentence fragments. Proper punctuation. Real words.
    - Need help paraphrasing dense content? Try this. And here's a great reference, too.
- Project created in RMarkdown and submitted as an HTML file via Canvas.

Please email me with any questions.