—Overview—

-This GitHub Site hosts data from a survey, conducted of Arizona and Austrian citizens, shortly after the 2022 midterm cycles of the national elections. The survey gathered socio-economic, behavioral and insights about respondents, as well as data from more complex composite scales.

This GH repo was created as part of the CAS502 course at ASU, by Joshua Abbott and Markkus Pfirman Schlosser.

—Features—

-There are a number of Jupyter Notebooks that can be used to demonstrate various steps in the data cleaning and preparation process, including one that re-scales a number of the composite scores, which were on various scales: 3-21, 0-5, etc. into a 0-1 score, with decimal points re-scaled to make these composite variables more comparable.

-Jupyter Notebook (mostly Markdown) exploring the survey questions, scales, and data format, some simple descriptive stats

-Original CSV file with data in the un-altered state (not all measures included as in original survey)

-Jupyter Notebook demonstrating the process of using MinMax to re-scale composite data points

-CSV with re-scaled data, and composite scores MinMaxed

-Jupyter Notebook demonstrating Testing of Data for row and column completeness, testing of MinMax re-scaling, etc.

-Jupyter Notebook allowing users to generate visualizations of networks based on:

-Gender: Male, Female, Other

-Election (2022): Voted Democrat, Voted Republican, Did not vote.

After users choose a starting variable, they can choose a “similarity threshold” which will reduce the edges of the network (above .8 is most effective). The Notebook also allows users to input a specific respondent’s ID number and generate a list of other nodes that have the highest “similarity score”, as calculated by a function in the notebook. Similarity scores are calculated based on the cosine similarity of two respondents, with vectors representing their results ranging from -1 to +1 (-1 indicating a low similarity, +1 indicating the highest similarity).

\*\*\*\*NOTE\*\*\*\* This feature is not completely functional because it is currently calculating based on ALL of the composite scores, and ideally, a user could choose specific variables to compare, or exclude others.

-Jupyter Notebook with Analysis of Crisis Categories COMING SOON: This notebook will demonstrate a similar exploration as the above notebook, but looking specifically at the categorical crisis responses that individuals gave. The comparative variables might change here as well, instead of looking at Gender and Voting history as predictors, we might look at income or education levels.

—Reporting Issues—

-Please feel free to contribute observed issues in data or code through the pull request feature. This is also a channel to communicate ideas for future features and functionality that might enhance this work.