Exploratory Data Analysis

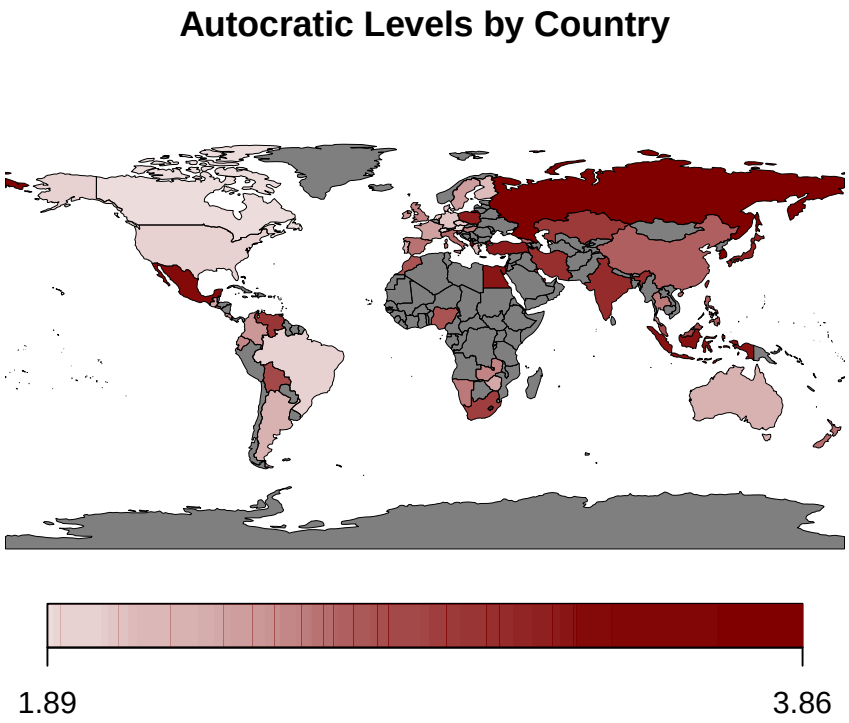
Ethan Allavarpu · Raymond Bai · Jaclyn Chiu · Ariel Chow · Carlie Lin · Dara Tan
Explore the GLOBE

Fall 2021

The Data

The data set we have chosen to analyze is the **Dana Landis Leadership** dataset, which comes from the GLOBE Research Survey. The data provided in the folder had survey results for (1) leadership and (2) societal and culture data and a PDF describing the nature of the survey, but nothing else. To glean more information, we found the two questionnaires (alpha and beta) described in the informational PDF to get the original questions. While we do not have a "codebook" in the traditional sense, the original questions asked may help guide us in understanding what each variable means and how the survey represents respondents' answers numerically. The survey is on a 1 to 7 scale, with 1 being a negative response, 4 a "neutral" score, and 7 positive.

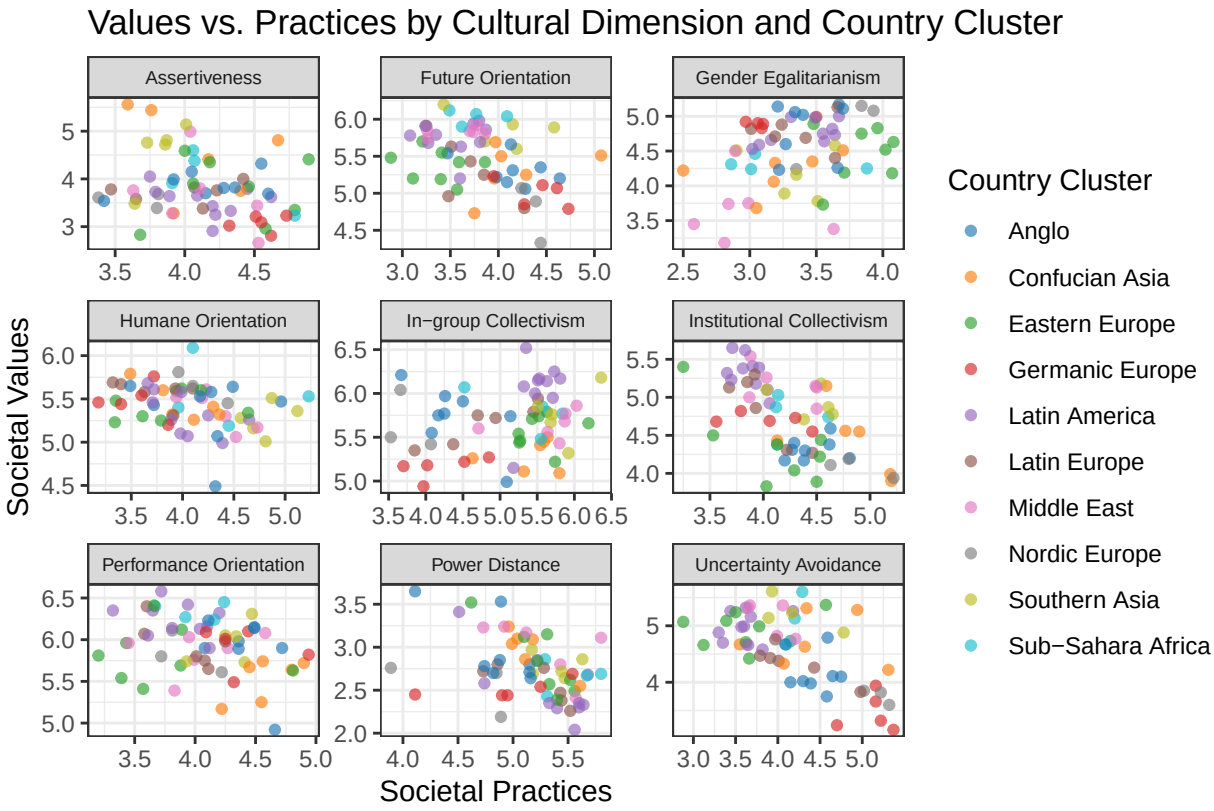
Leadership



The above global heatmap presents the beliefs about whether autocracy helps (higher values) or hinders (lower values) leaders in a given country. Multiple similar graphics were generated for a variety of leadership variables, ultimately leading us to recognize that different regions of the world may favor certain leadership qualities over others and that certain leadership qualities may be liked/disliked "together" (i.e., they have associations).

From boxplots we generated (not pictured), the difference by country cluster (as defined by the dataset) seems visible, prompting us to consider clustering countries to determine how they may be separated. We do want to note that there are only 62 observations–this totals to fewer than 62 countries because Germany and South Africa have two observations each (West vs. East and White vs. Black, respecively). The limited number of observations may limit us in the scope of our analysis because we have less than a third of the total countries (demonstrated by the vast swaths of grey on the world map for which there is no data) and clustering would further reduce block/group sizes.

Societal Values and Practices



Note: Countries with missing country clusters were excluded.

The plot grid demonstrates how, contrary to what *should* happen, the societal values of a country do not correlate well with the societal practices (what *should be* does not align with what *is*). This may be a source of further exploration, as we may want to see whether there is a significant difference between the values and practices. When we look at the per-region divisions though (by color), the trend (or lack thereof) seems to persist for these concepts.

Problem Statements

From the above exploratory analysis, we see that the data have been segregated into two sets: those regarding leadership beliefs (i.e., what makes a good leader) and those about society (values, practices, and beliefs). Thus, to mirror this, we have two main problem statements we want to consider:

1. Which characteristics or traits do countries tend to group together when determining “good" leadership values?
2. Do societal practices and societal values align?
    - If they do not, which practices and values deviate most significantly?

Potential Issues

While promising, there are a few issues with the data that we have to consider. First, the number of observations: 62 observations is a rather small dataset, which suggests that any analysis will likely be limited. Another potential issue is running multiple significance tests: with nine societal practice vs. value concepts to consider, maintaining the $\alpha = 0.05$ significance level would increase the experimentwise error rate and thus the likelihood of a false positive ($P(\text{False Positive}) = 1 - P(\text{No False Positive}) = 1 - 0.95^9 \approx 0.3698$). To remedy this issue, we may either consider employing a *Bonferroni correction* or *Tukey's method (HSD).*