Stats 140XP: Final Project

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

1 December 2021

Contents

Abstract

We wanted to answer two main questions: which leadership qualities do countries tend to view similarly and if countries align their perceptions societal practices and values. For determining leadership qualities and similar countries, we used principal component analysis (PCA) then k-means clustering to create four "clusters" of countries with similar leadership beliefs. We used the _____ method to _____. We found that _____. We also looked at _____. In the future, we recommend looking into _____. Some limitations to our project are _____.

Problem Statements

1. Which characteristics or traits do countries tend to group together when determining "good" leadership values?

   - Which countries have similar perceptions of these leadership values?

2. Do societal practices and societal values align?

   - If they do not, which practices and values deviate most significantly?

Description of Dataset

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 more. To glean more information, we found the two questionnaires (alpha and beta) described in the informational PDF to get the original questions asked in the surveys. While we do not have a "codebook" in a 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.

Here is a look at 6 full and complete observations from the leadership survey:

Here is a look at 6 full and complete observations from the social and cultural survey:

Description of Variables

Visualization and Exploratory Data Analysis

Analysis

Leadership Values

For the leadership values problem statement, the first objective is to collapse the data into the first four principal components through principal component analysis (PCA). PCA finds the directions which capture the most variability in the data, so the first four account for the maximum variation. PCA allows us to visualize trends in the leadership values: countries tend to have similar sentiments about variables that "point" in the same direction (had principal component values that aligned).

Before performing PCA, though, we remove the second-order factor analysis variables due to the heavy correlation with the original predictor variables and a more difficult interpretation of these variables. Since our goal is to understand the relationship between certain leadership characteristics, keeping these complicated variables might reduce our understanding of some characteristics.

After performing PCA, we visualize the directions of the first two principal components:
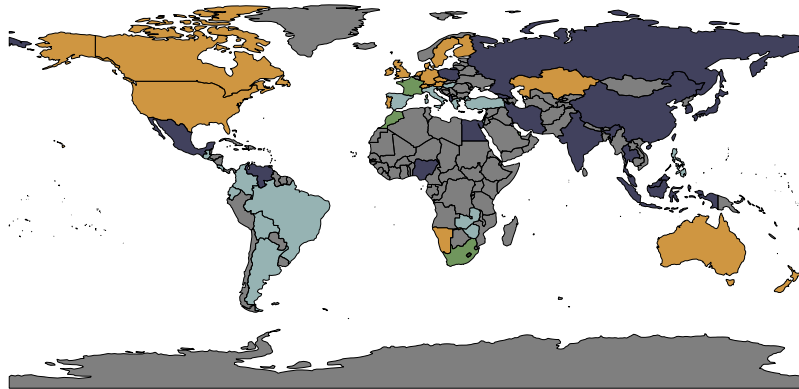
First Two Principal Components with Each Variable's Loading
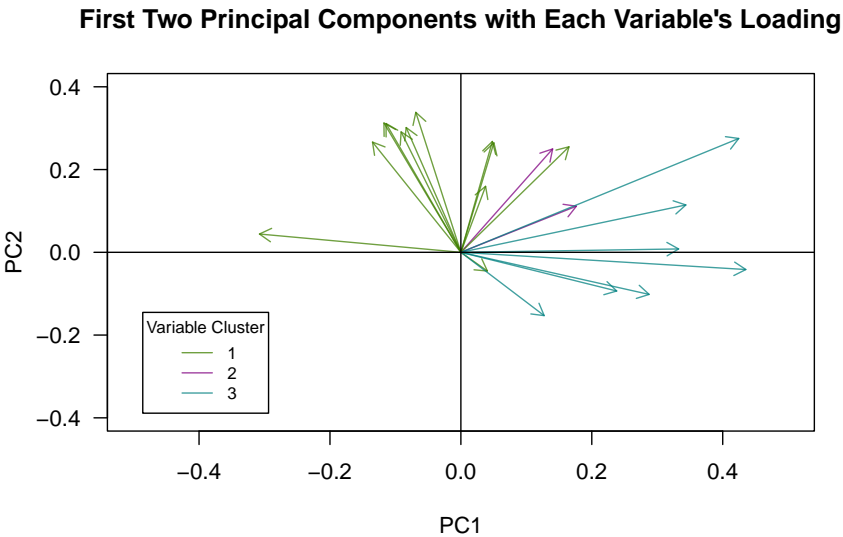
On the first four principal components, we perform k-means clustering to determine the "groups" of leadership characteristics that have similar perceptions:

PCA Coefficients for Each Variable


Note: Variables colored by cluster
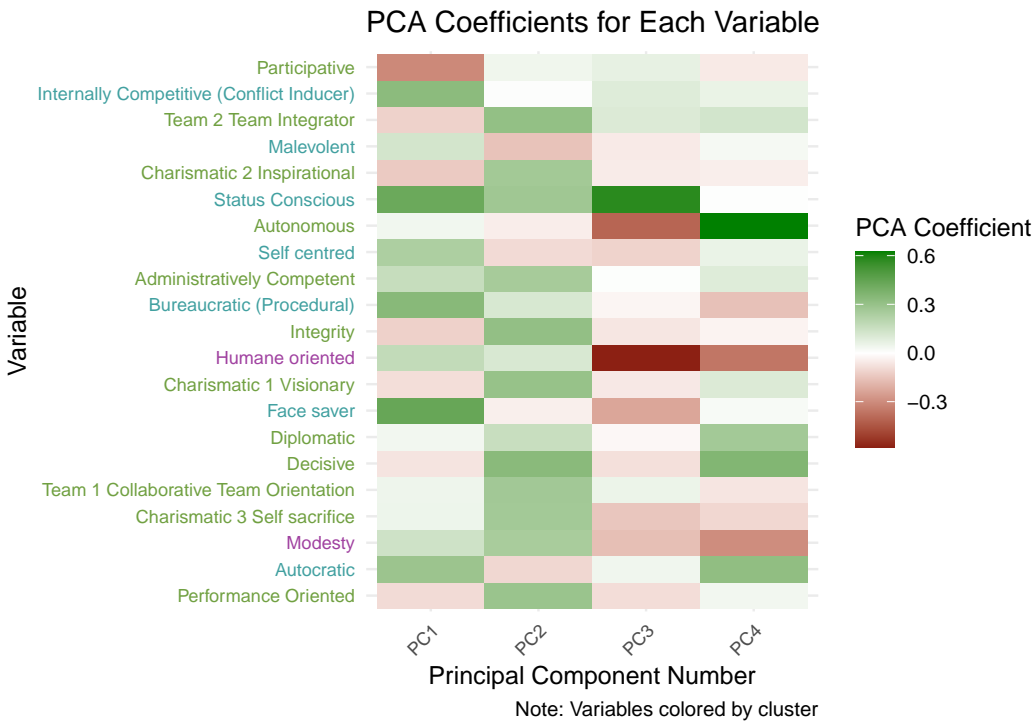
After considering the leadership characteristics, we cluster countries based on similar perceptions. We do this by using the variables of the countries transformed into the first four principal components, then running PCA on those components. The clustering results segregate the countries into the following segments:

Country Clusters with Similar Leadership Values

Country Clusters by First Two Principal Components

Median Survey Scores for Leadership Categories by Cluster
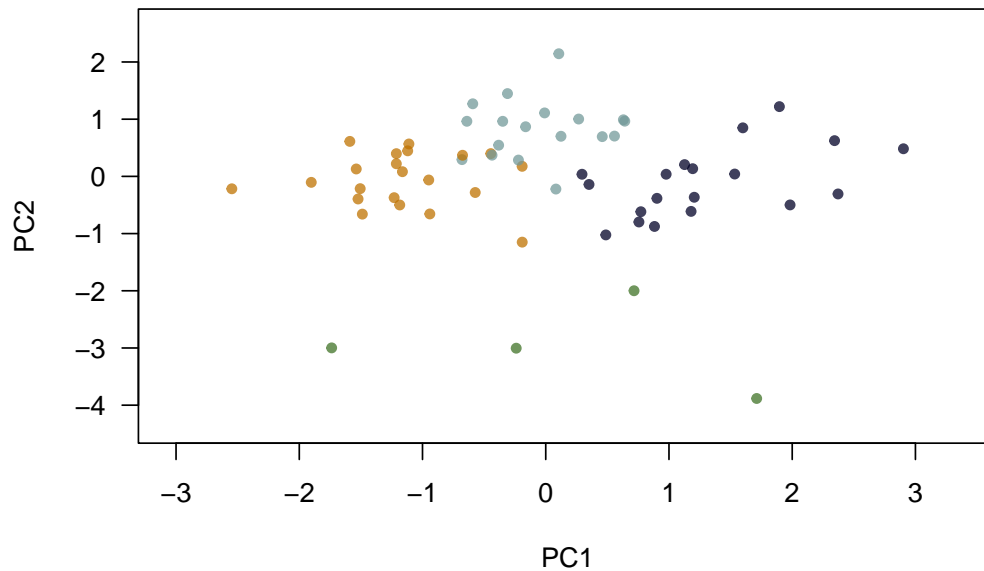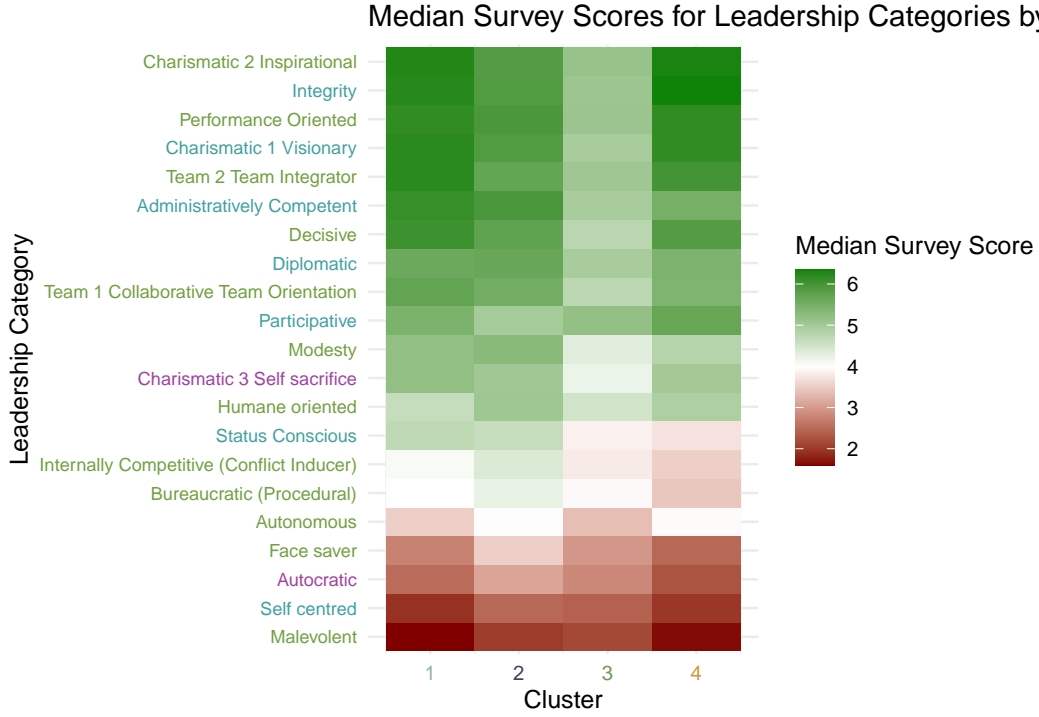
Societal Practices and Values

For this section, since we consider nine societal practice vs. value concepts, so maintaining a significance level of $\alpha = 0.05$ increases the experimentwise error rate: $P(\text{Any False Positive}) = 1 - P(\text{No False Positives}) = 1 - 0.95^9 \approx 0.3698$. To remedy this issue, we employ a *Bonferroni correction* to change the significance threshold from $\alpha = 0.05$ to $\alpha = \frac{0.05}{9} \approx 0.0056$.

Conclusions

Suggestions for Further Research

Limitations

While promising, there are a few issues with the data we have to consider. First, the number of observations: 62 observations is a rather small dataset, which suggest that any analysis will likely be limited. Also, the data silences are important and should not be overlooked. Why were certain countries not surveyed?