International Insights: Analysis of United Nations General Assembly Voting Data

By Jorge Tenorio
Introduction to Data Science
Springboard Starting Cohort of August 2016

Background

The United Nations General Assembly (UNGA) serves as a global forum where nations all around the world express their views and vote on issues of global importance. The issues brought forward on the UN stage range from human rights violations to economic development initiatives. Every country and its citizens are affected by the outcome of UNGA votes. As the world becomes more and more globalized, even non-member countries can be indirectly impacted by a UNGA resolution.

When the United Nations was first developed, it was meant to promote world peace and open the world to global conversation as opposed to armed conflict. Opinions on whether the UNGA has achieved its goals vary from country to country. Aside from the individual impacts, UNGA votes have a regional and global impact in terms of international relations. By analyzing the voting data collected from the outcomes of UNGA resolutions, international relationships can be analyzed and quantified.

Although countries hold varied opinions on global issues within themselves, they ultimately have to present a unified voice through their UNGA vote. Ideally, this voice is consistent with the current views of the country, and intend to advance or protect the country's interests. Interests for one country might align with those of another for a variety of reasons: they could share the same cultural values, strong economic ties, or even the same allies or foes. Analysis of UNGA voting data helps to identify and understand how a country votes based on a particular topic.

The Problem

The intent of the UNGA Voting Analysis project is to identify topics of importance or of frequent appearance on the UNGA stage. Secondly, the analysis identifies particular voting blocs based on these topics. Voting blocs are derived by similar voting patterns on particular topics and a country's liberal or conservative tendencies. Using inputs indicating these voting patterns, countries can be grouped by similarity. Once grouped by similarity, a third party can begin to assess how exactly the groups are similar (i.e. liberal, vote similar on topics of peace, etc.). Such an analysis would help answer the following questions:

- Are regions viable voting blocs with significantly similar voting tendencies? What kind of "region" is more indicative of a voting bloc (e.g. Americas, North/South America, Latin America, etc.)?
- Are there factors other than regions more indicative of the outcome of a vote?
- Is there a combination of factors that could yield an optimal method, formula, or algorithm for assessing the outcome of a vote based on the topic?

By gaining insight into the factors most related to the outcome of a vote-based on topic, the United Nations body and its members can assess the dynamics between various countries. Ultimately, the goal of the UN is to break down barriers to allow open discussion, foster tolerance, and promote peace. Analysis of UNGA voting data would help identify voting blocs and their probable causes (i.e. geography, socioeconomic status, language, etc.). Identifying the "blocs" would help illuminate potential barriers the UNGA could seek to understand, resolve, and dissolve (if possible).

The Data

United Nations General Assembly Data is available from 1946 to 2014 and comes in the form of four datasets:

1. Raw Voting Data: contains the raw voting data for each country on each resolution
2. Vote Description: contains a short and long text description for each resolution and categorizes each using six issue codes.
3. Ideal Point Data: contains ideal point data for each country. Ideal points are used as a measure for estimating liberal/conservative inclination
4. Codebook Dyadic Data: contains data on the absolute distance between ideal points of each country as well as affinity data (although affinity data will not be used)

In addition to the data provided in the combined package by Erik Voeten, other global datasets can be used to include additional measures relevant to the study. Data can be acquired from credible databanks from the United Nation, the World Economic Forum, or the World Bank. The only additional data included is each country's regional label as designated by the UN.

Outline of Methodology and Approach

1. **Data Scrubbing:** Data included a single vote for each of the 192 member countries for multiple issues, across multiple sessions, throughout multiple years. In total there were 700k rows to potentially evaluate. Data scrubbing resulted in a simplified dataset with no missing values and additional info extracted into separate columns.
2. **Text Mining**: Text mining resulted in additional columns for classification of vote types. Although the dataset provided these classifications for some votes (e.g. Palestine, nuclear weapons, economy, etc.), many votes did not fall into one of these categories. In total, 14 columns were used to assign voting types.
3. **Data Simplification**: The data was simplified by eliminating countries without enough representative votes across all UN sessions. The data was then aggregated into a single representative point for each country to simplify the interpretation of the clustering.
4. **Clustering**: K-means clustering was used to assign each of the 151 countries to a specific cluster. A sample of 5 clustering results was used to determine the final cluster for each data point.
5. **Cluster Characterization**: After each country was assigned to a cluster, the clusters were characterized based on the inputs used for clustering (ideal points and voting weights on each category).

Results

The following table shows the list of each country in each of the 6 clusters of the k-means clustering:

Cluster 1 (23)	Cluster 2 (44)		Cluster 3 (36)		Cluster 4 (10)	Cluster 5 (26)		Cluster 6 (12)
Australia	Algeria	Malaysia	Afghanistan	Nicaragua	Albania	Angola	Suriname	Antigua & Barbuda
Austria	Benin	Mali	Argentina	Pakistan	Belarus	Bahamas	Swaziland	Belize
Belgium	Burkina Faso	Mauritania	Bolivia	Panama	Bulgaria	Bahrain	Malta	Cape Verde
Canada	Burundi	Mauritius	Brazil	Paraguay	Czechoslovakia	Bangladesh		Comoros
Denmark	Cambodia	Morocco	Chile	Peru	Hungary	Barbados		Grenada
Finland	Cameroon	Nepal	Colombia	Philippines	Mongolia	Bhutan		Seychelles
France	Central African Republic	Niger	Costa Rica	Saudi Arabia	Poland	Botswana		Solomon Islands
Greece	Chad	Nigeria	Cuba	Syria	Romania	China		St. Lucia
Iceland	Congo	Rwanda	Dominican Republic	Thailand	Russia	Djibouti		St. Vincent & Grenadines
Ireland	Cyprus	Senegal	Ecuador	Uruguay	Ukraine	Equatorial Guinea		Turkey
Israel	DR Congo	Sierra Leone	Egypt	Venezuela		Fiji		Vietnam
Italy	Gabon	Singapore	El Salvador	Yemen Arab Republic		Gambia		Zimbabwe
Japan	Ghana	Somalia	Ethiopia	Yugoslavia		Guinea-Bissau		
Luxembourg	Guinea	Sri Lanka	Guatemala			Guyana		
Netherlands	Indonesia	Sudan	Haiti			Lesotho		
New Zealand	Ivory Coast	Tanzania	Honduras			Malawi		
Norway	Jamaica	Togo	India			Maldives		
Portugal	Jordan	Trinidad and Tobago	Iran			United Arab Emirates		
South Africa	Kenya	Tunisia	Iraq			Mozambique		
Spain	Kuwait	Uganda	Lebanon			Oman		
Sweden	Laos	Zambia	Liberia			Papua New Guinea		
United Kingdom	Libya		Mexico			Qatar		
United States	Madagascar		Myanmar			Samoa		

Table 1: The table above shows the countries within each cluster as assigned by the k-means algorithm.

The data is further visualized by country in the map below:

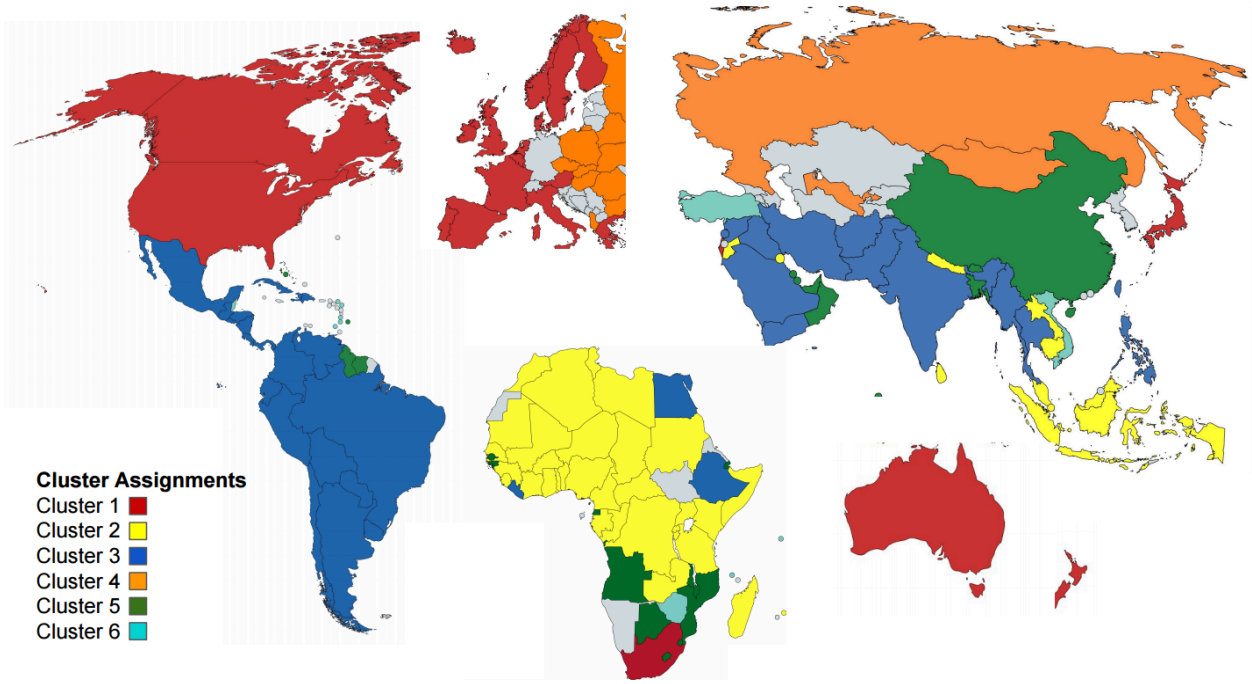


Figure 1: The map above visually displays the cluster each country belongs to; some geographic similarity is apparent, though not completely consistent.

Visually, we can see there is some geographic similarity within some of the clusters, but others are less well defined by geography. In order to better assess why the k-means algorithm placed these countries within the same cluster, the cluster averages were compared to the population average.

Variable	Population Avg	Population Stdv
ideal points (+ = Liberal, - = conservative)	-0.098245637	0.774758455
palestinian conflict (me)	0.165831962	0.058868005
nuclear weapons/material (nu)	0.109071812	0.04327139
arms control & disarmament (di)	0.134858076	0.04527727
colonialism (co)	0.123526114	0.042301779
human rights (hr)	0.13376266	0.044910194
economic development (ec)	0.068612894	0.02447119
procedural/structural (ps)	0.090469911	0.024780702
africa (af)	0.017194751	0.007899059
security council (sc)	0.007615187	0.002653147
special initiatives/votes (sp)	0.013852036	0.004567776
UN Action (un)	0.023697865	0.007768654
International (int)	0.022352494	0.007098877
UN budget (bu)	0.014963991	0.006823065
peace measures (pc)	0.019635338	0.006106505

Table 2: The table above shows the population average and standard deviation for each input variable of the k-means clustering.

From the population averages of the vote categories (me, nu, di, etc.) one key insight is apparent: in general, votes presented to the United Nations General Assembly tend to pass. The values for each of the categories could range from -1 if the country votes against the measure, and +1 if the country votes in favor of the measure. Because all the population averages are positive, on average the general assembly tends to vote in favor of all the votes presented. The main differentiation between the clusters will be how likely they are to vote in favor of a particular vote.

Cluster Characterization

The clusters can be more easily characterized by assigning them a relative value to assess how likely they are to vote in favor of a vote. Relative values between 1 to 4 were assigned to each voting category for each cluster by comparing the cluster's average value to the population average. The table below shows the intervals used to assign the relative values to each of the voting category average values for each of the clusters:

Classification Scale	Interpretation	Classification Interval
4	Very likely to approve vote	$y \geq x + \sigma$
3	Moderately likely to approve vote	$x \leq y < x + \sigma$
2	Likely to approve vote	$x - \sigma \leq y < x$
1	Barely likely to approve vote	$y < x - \sigma$

y: variable average
x: population average
σ: population standard deviation

Table 3: The table shows the scale used to assign the relative values to each of the voting categories used as inputs to the k-means clustering.

Below are the results for each cluster based on the relative scale show in Table 3.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Ideal Points (+ = Liberal, - =conservative)	4	2	3	1	2	2
Palestinian Conflict (me)	1	3	2	2	3	4
Nuclear Weapons/Material (nu)	1	3	2	1	3	3
Arms Control & Disarmament (di)	1	3	2	1	3	4
colonialism (co)	1	3	2	2	3	3
human rights (hr)	1	3	3	3	3	3
economic Development (ec)	1	3	2	2	3	3
procedural/structural (ps)	2	3	4	1	2	1
africa (af)	1	3	3	3	3	3
security council (sc)	1	3	3	2	2	2
special initiatives/votes (sp)	2	3	3	2	2	1
UN Action (un)	2	3	2	1	4	3
Addressing the International Community (int)	1	3	3	1	3	3
UN Budget (bu)	2	3	3	1	3	1
Peace Measures (pc)	1	3	3	1	3	3

Table 4: The table above shows the likelihood of each cluster voting in favor of particular voting category based on the relative scale shown in Table 3. Note the following scale for Ideal Points: 1 = very conservative, 2 = conservative, 3 = liberal, 4 = very liberal.

Using the relative values above, insights were gained from each cluster's voting pattern on particular topics, as well as their liberal/conservative leanings from the ideal points. Some clusters have particular topics they very likely vote in favor of (clusters 3, 5, and 6), while others are barely likely to vote in favor of any topic at all (clusters 1 and 2). We also have clusters tending to vote in favor of just about any vote and on any topic (cluster 2). A summary of the results is shown in the table below:

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Liberal or Conservative	Very Liberal	Conservative	Liberal	Very Conservative	Conservative	Conservative
Overall Vote Average	1.29	3.00	2.64	1.64	2.86	2.64
Overall Vote Interpretation	Least likely to vote in favor of a vote.	Most likely to vote in favor of a vote.	Likely to vote in favor of a vote.	Note very likely to vote in favor of a vote.	Likely to vote in favor of a vote.	Likely to vote in favor of a vote.
Regions	Westernized Civilizations	Africa	Latin America, Middle East, Southwest Asia	Eastern Europe	Mixed	Island Nations
Summary	The most liberal cluster, but the most cautious, reserved, and selective of which votes it approves.	A conservative cluster that is fairly likely to approve a vote.	A liberal cluster and a strong supporter of procedural/structural measures.	The most conservative cluster and the most likely to vote against measures on procedural/structural changes, the UN budget, and peace measures.	Conservative Cluster that tends to vote in favor of most votes, especially if they involve a UN call to action.	A conservative cluster and a strong supporter of measures related to Palestinian conflict and arms control/disarmament

Table 5: The table above shows the summary of each cluster based on their voting patterns and liberal/conservative ideologies.

Conclusion

Using k-means clustering, six clusters of countries were identified based on their average approval likelihood for each category type and their ideal points. Each of the six identified clusters can be considered as a voting bloc. Although some geographic similarities can be observed, the clusters are characterized by their voting patterns and ideals. By identifying the voting blocs of the United Nation's voting countries, a third party can begin to assess which countries are strongly aligned and why. Since the k-means clustering was run multiple times, the sampling results can be used to determine which countries are most likely to deviate from their voting bloc (i.e. countries on the "edge" of their cluster).

Overall, the analysis of the United Nations voting data can help us determine which factor make countries similar and quantify that similarity. In the future additional variables can be included to gain more insights. For example, do GDP, population size, or average household income affect the clustering of these countries? By understanding the similarity and differences between the voting blocs and individual countries, the United Nations can better assess why there might be conflict between particular countries and why. The UN can also assess which issues are most important to all countries to ensure topics of the highest priority are those which countries are most passionate about. By using data the United Nations can better assess how to use international relations as a tool for improving economic development, protecting human rights, and advancing world peace.