JT's IBM Data Science Capstone Assignment:

Red and Blue Gerrymandering

What determines (alleged) Georgia gerrymandering?

Georgia™

(image courtesy of economicmodeling.com)

Introduction (Business Problem Summary):

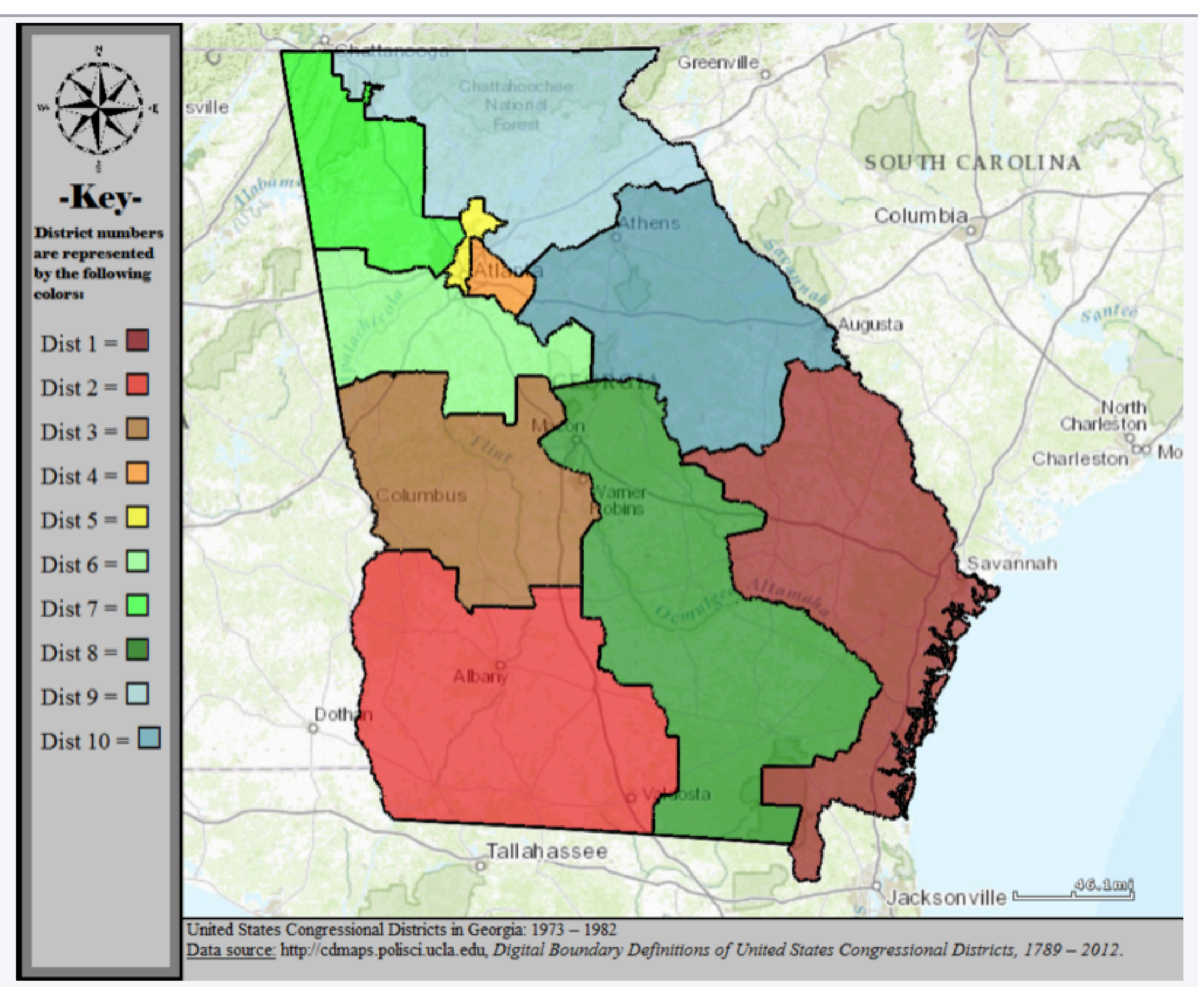
Amongst the issues affecting election campaigning, Gerrymandering is one of the most vexing and arcane. Given the contentious nature of this topic in Georgia (USA) and the close results of recent elections, it is important that political campaign management firms (as well as voters) understand the effects of this issue in order to better target candidate marketing strategies during the next Georgia elections. Demographic analysis of the Georgia districts, most of which had undergone gerrymandered re-districting, will be important to their success.

Note that ***the target audience for this analysis are political campaign management firms (often hired by congressional candidates) who do business in Georgia.*** *Such firms typically control investment strategies for the candidate, including types of adverts to place, locations to focus on, and delivery channels for important candidate communication. These strategies are aimed at achieving the highest voter turnout for their candidate,* ***therefore, demographic analysis is vital to the success and reputation of the campaign management firm*** *(as well as the candidate).*

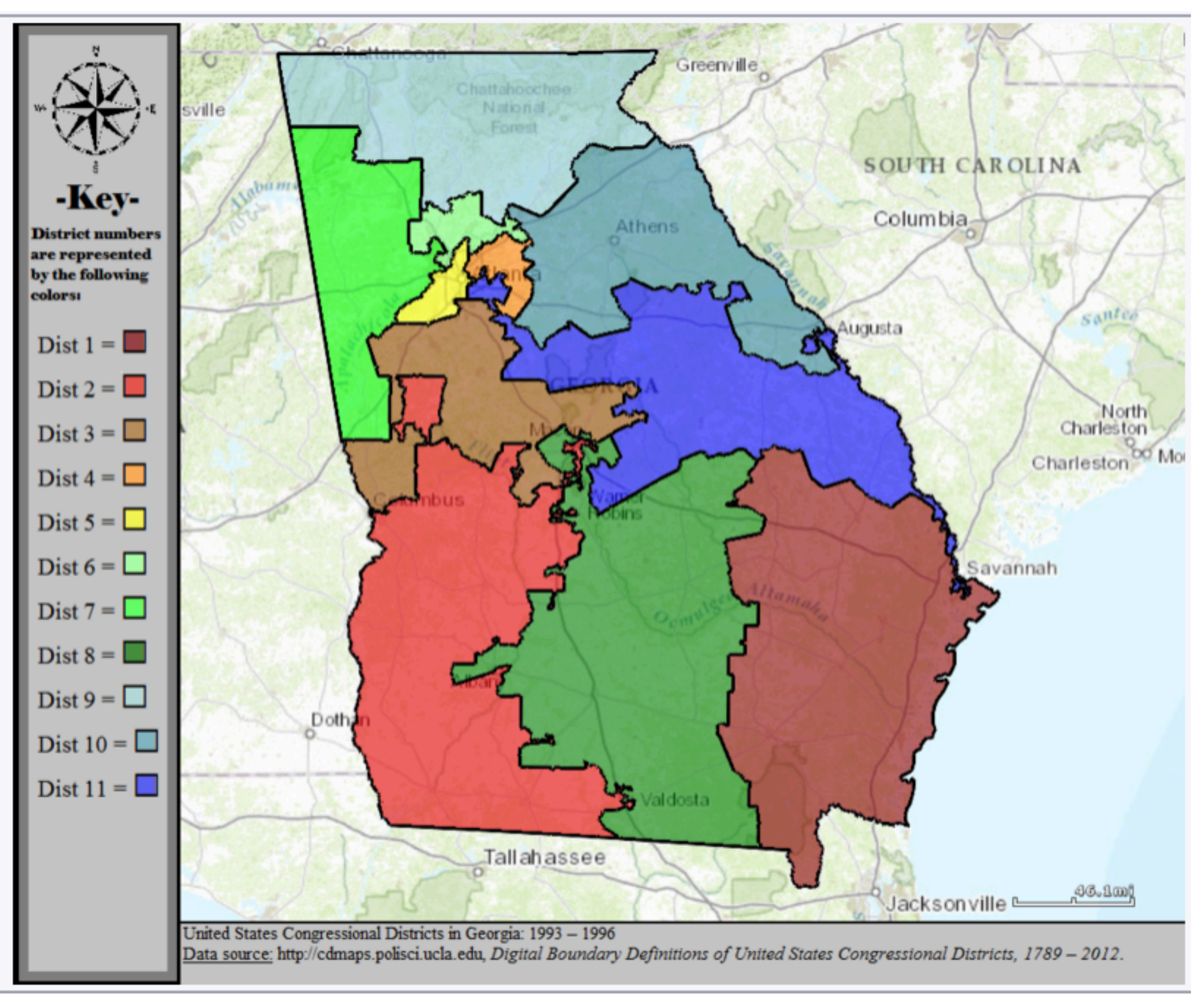
Gerrymandering is a practice intended to establish an unfair political advantage for a particular party or group by manipulating election district boundaries. "Districts" define geographical boundaries, with each district within a state being geographically contiguous and having about the same number of state voters.

In recent years, districting policies in Georgia, USA have been hotly debated recently, particularly during the 2018 gubernatorial election run-off between Stacy Abrahms and current Governor Brian Kemp. Accusations of a 'rigged process' were rife, as redistricting often resulted in varied and interesting "geographically-contiguous" shapes:

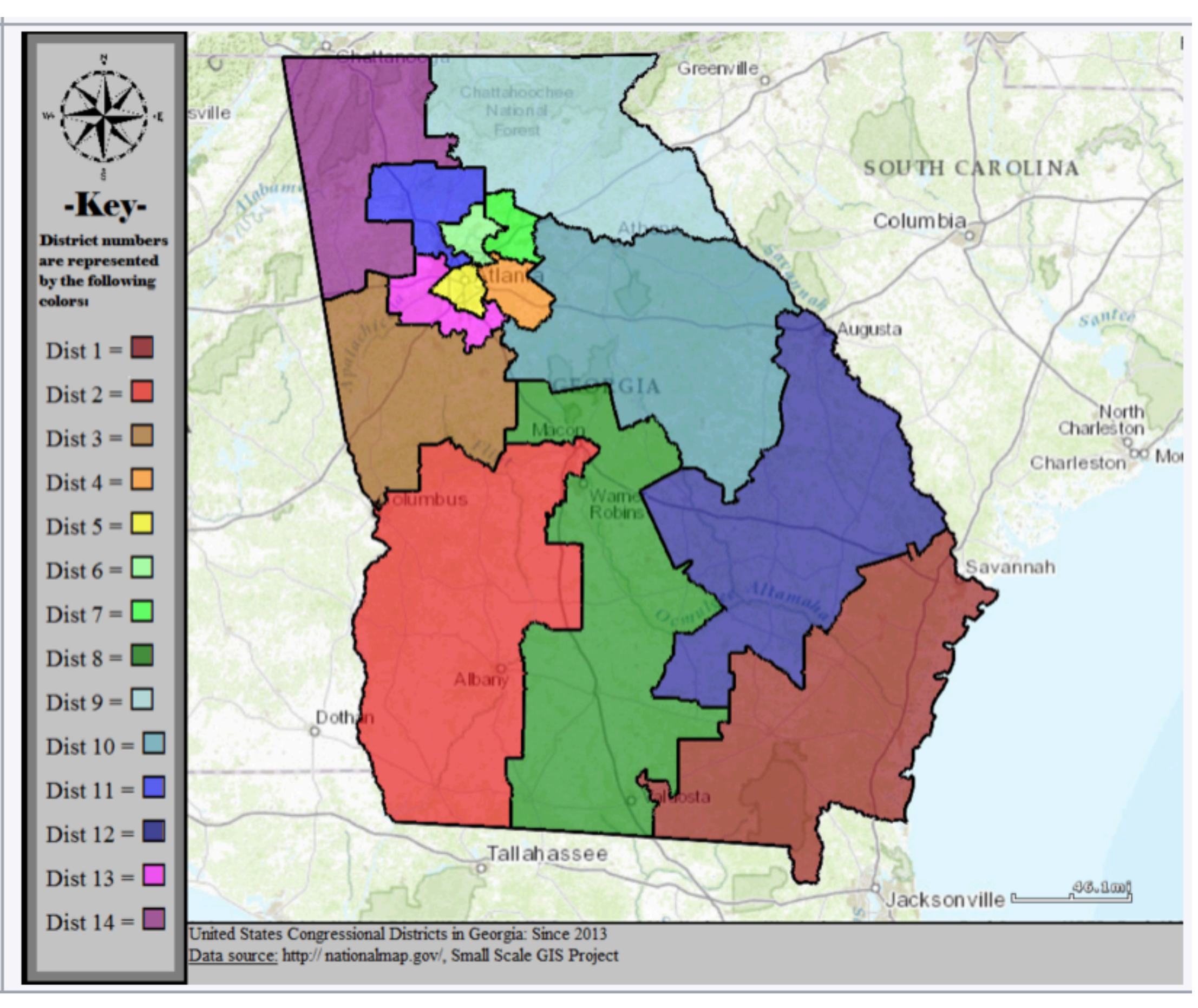
1973–1982



1993–1996



Since 2013



_(source: wikipedia - https://en.wikipedia.org/wiki/Georgia%27s_congressional_districts)_

Given that personal policital ideologies have shifted over time in given locations, understanding this phenomena is essential. We will be evaluating the demographics in Georgia including contrasting Georgian 'Red' districts (Republicans) with 'Blue' districts (Democrats) to see what comprises each type. These observations will inform investments campaign firms should consider to combat the negative effect of gerrymandering on candidate success.

Caveats: Please note that ***this is exploratory analysis*** *(in the loosest sense of the word); my results and observations could mislead at a time where accurate information ("truth") is under stress. Moreover, to conduct such analysis properly, I would need access to more data (eg, cuts of information by year pre and post redistricting, more granular income distribution and education reporting); such data is currently not freely available.*

-----END OF INTRODUCTION SECTION-----

START OF DATA SECTION:

Method and Data Requirements:

I will review certain characteristics of "red" (Republican) and "blue" (Democrat) districts:

- population voting history
- education
- age
- local amenities (venue categories)
- poverty level

These features will be used with a k-means clustering process which will present groupings that can be evaluated for strategic review and investment. These features will be sourced from following (samples included):

https://ballotpedia.org/Redistricting_in_Georgia - congressional districts by number, current representative by full name, and current party affiliation as well as term, election victoty margins, district ethnic demographics. Information is conveyed in several tables included in this one webpage. Samples:

Demographics of Georgia's congressional districts (as percentages) [hide]

District	Hispanic or Latino of any race	White	Black or African American	American Indian and Alaska Native	Asian	Native Hawaiian and other Pacific Islander	Other	Multiple races
District 1, Georgia	6.20%	59.45%	29.02%	0.44%	2.02%	0.10%	0.22%	2.55%
District 2, Georgia	5.18%	39.87%	52.13%	0.19%	1.00%	0.10%	0.08%	1.46%
District 3, Georgia	5.38%	66.58%	23.97%	0.13%	1.61%	0.01%	0.25%	2.07%

Demographics of Georgia's congressional districts (as percentages) [hide]

District	Hispanic or Latino of any race	White	Black or African American	American Indian and Alaska Native	Asian	Native Hawaiian and other Pacific Islander	Other	Multiple races
District 1, Georgia	6.20%	59.45%	29.02%	0.44%	2.02%	0.10%	0.22%	2.55%
District 2, Georgia	5.18%	39.87%	52.13%	0.19%	1.00%	0.10%	0.08%	1.46%
District 3, Georgia	5.38%	66.58%	23.97%	0.13%	1.61%	0.01%	0.25%	2.07%

https://www2.census.gov/programs-surveys/demo/tables/voting/table01.xlsx - "Number of Votes Cast, Citizen Voting-Age Population and Voting Rates for Congressional Districts: 2018" Sample:

Table 1. Number of Votes Cast, Citizen Voting-Age Population and Voting Rates for Congressional Districts: 2018

Line #	State abbreviation	State name	Congressional district	Votes cast for congressional representative for the November 6, 2018 election[1]	Citizen voting-age population[2] Estimate	Citizen voting-age population[2] Margin of error (MOE)	Voting rate[3] Estimate	Voting rate[3] Margin of error (MOE)
1	AL	Alabama	1	242,617	544,464	3,424	44.6	0.3
2	AL	Alabama	2	226,230	516,295	5,674	43.8	0.5
3	AL	Alabama	3	231,915	543,854	4,099	42.6	0.3

https://www2.census.gov/programs-surveys/demo/tables/voting/table02a.xlsx - "Characteristics (Age) of the Citizen Voting-Age Population for Congressional Districts: 2018" Sample:

Table 2A. Characteristics (Age) of the Citizen Voting-Age Population for Congressional Districts: 2018

Line #	State abbreviation	State name	Congressional district	Citizen voting-age population Estimate	Margin of error (MOE)	18-29 years old Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)	30-44 years old Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)	45-64 years old Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)
1	AL	Alabama	1	544,464	3,424	104,060	2,412	19.1	0.4	125,900	2,538	23.1	0.5	188,031	3,186	34.5	0.5
2	AL	Alabama	2	516,295	5,674	109,222	2,918	21.2	0.5	118,999	3,146	23.0	0.5	173,351	3,069	33.6	0.5
3	AL	Alabama	3	543,854	4,099	121,328	2,379	22.3	0.4	121,212	3,483	22.3	0.6	182,379	3,899	33.5	0.7
4	AL	Alabama	4	515,701	4,678	95,990	2,777	18.6	0.5	114,921	2,560	22.3	0.5	177,247	3,022	34.4	0.5

https://www2.census.gov/programs-surveys/demo/tables/voting/table02c.xlsx - "Characteristics (Educational Attainment) of the Citizen Voting-Age Population for Congressional Districts: 2018" Sample:

Table 2C. Characteristics (Educational Attainment) of the Citizen Voting-Age Population for Congressional Districts: 2018

Line #	State abbreviation	State name	Congressional district	Citizen voting-age population Estimate	Margin of error (MOE)	Less than 9th grade Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)	9th to 12 Grade, no diploma Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)	High school graduate Estimate	Margin of error (MOE)	Percent of total
1	AL	Alabama	1	544,464	3,424	13,409	1,845	2.5	0.3	57,234	4,433	10.5	0.8	185,503	6,968	34.
2	AL	Alabama	2	516,295	5,674	18,295	2,059	3.5	0.4	53,192	3,981	10.3	0.8	172,427	6,215	33.
3	AL	Alabama	3	543,854	4,099	19,578	2,504	3.6	0.5	56,833	4,646	10.5	0.9	174,209	6,786	32.
4	AL	Alabama	4	515,701	4,678	24,346	2,997	4.7	0.6	57,850	3,966	11.2	0.8	175,806	6,423	34.
5	AL	Alabama	5	551,968	2,121	14,908	2,282	2.7	0.4	43,805	4,016	7.9	0.7	151,609	5,046	27.
6	AL	Alabama	6	535,753	7,635	11,300	2,173	2.1	0.4	36,957	3,501	6.9	0.6	140,443	6,392	26.
7	AL	Alabama	7	504,177	9,052	16,616	1,923	3.3	0.4	55,234	5,319	11.0	1.0	168,045	8,111	33.
8	AK	Alaska	At-large	532,244	2,539	10,640	1,693	2.0	0.3	26,444	2,293	5.0	0.4	160,451	4,824	30.

https://www2.census.gov/programs-surveys/demo/tables/voting/table02b.xlsx - "Characteristics (Sex and Poverty) of the Citizen Voting-Age Population for Congressional Districts: 2018" Sample:

Table 2B. Characteristics (Sex and Poverty) of the Citizen Voting-Age Population for Congressional Districts: 2018

Line #	State abbreviation	State name	Congressional district	Citizen voting-age population Estimate	Margin of error (MOE)	Men Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)	Women Estimate	Margin of error (MOE)	Percent of total	Percent of total margin of error (MOE)	Poverty universe[1]	Poverty universe margin of error (MOE)
1	AL	Alabama	1	544,464	3,424	256,603	2,312	47.1	0.3	287,861	2,175	52.9	0.3	532,906	4,244
2	AL	Alabama	2	516,295	5,674	243,971	3,467	47.3	0.4	272,324	3,369	52.7	0.4	498,215	6,078
3	AL	Alabama	3	543,854	4,099	261,066	2,813	48.0	0.3	282,788	2,387	52.0	0.3	526,264	4,581
4	AL	Alabama	4	515,701	4,678	248,166	3,029	48.1	0.3	267,535	2,859	51.9	0.3	509,407	4,836
5	AL	Alabama	5	551,968	2,121	267,633	1,788	48.5	0.2	284,335	1,610	51.5	0.2	537,668	2,826
6	AL	Alabama	6	535,753	7,635	255,440	5,227	47.7	0.5	280,313	4,313	52.3	0.5	524,830	7,733
7	AL	Alabama	7	504,177	9,052	230,516	5,232	45.7	0.6	273,661	5,749	54.3	0.6	481,363	9,041

https://developer.foursquare.com/docs/build-with-foursquare/categories - And of course, FourSquare data for venue categories, with locations pulled from geopy *(if it cooperates for me).* Sample:

Documentation / Build with Foursquare / Venue Categories   Search Documentation

Build with Foursquare
Sample Apps
Pilgrim Toolkit
Venue Categories
Venue Chains
Categories Changelog
Resources and Logos

Over time, our categories list may change slightly. Visit our changelog to view updates.

Arts & Entertainment
4d4b7104d754a06370d81259

Amphitheater
56aa371be4b08b9a8d5734db

Aquarium
4fceea171983d5d06c3e9823

Arcade

-----END OF DATA SECTION-----

In [ ]: