# Capstone Project Proposal

Name: Joseph Rama

## Objective

The objective of my project is to identify geographic patterns in the distribution of LGBTQ people throughout the United States and determine any evidence of clustering based on a variety of data sources. After this, I will attempt to correlate this distribution with the availability of support systems of LGBTQ people as well as with localized legal and social support for LGBTQ lifestyles. This project is important to me as a gay man, as it will hopefully send some light on trends in and motivations for LGBTQ spatial distributions, especially considering the current Presidential administration’s overturning of many progressive policies in support of LGBTQ people.

## Impact

The hope of this project is to determine what places in the United States LGBTQ people are drawn to and why. This will not only be beneficial in identifying the safest/most accepting/most supportive places for LGBTQ people to live, but may also help find places that have large LGBTQ populations but that are lacking the support they need. Although this project will most likely not be groundbreaking or cause much change, it is certainly important to examine the social changes our country is undergoing and to try to find explanations. Additionally, I have found it difficult to locate accessible information relating locations of LGBTQ populations to locations of LGBTQ support centers/offices. This has the potential to be a valuable resource for the planning of future office locations or determining any existing locations that need additional resources.

## Dataset(s)

US Census Household Type: <https://data.census.gov/app/mdat/ACSPUMS1Y2022/vars> (2022) and <https://data.census.gov/app/mdat/ACSPUMS1Y2012/vars> (2012)

This dataset will be a reliable source for information about localized LGBTQ populations based on the “Household Type” Variable. This will give perspective into the distribution of LGBTQ citizens in the United States, and the geographic nature of the data will work well with geospatial analysis. As the data is sourced from the US Census Bureau, I find it to be generally reliable.

University of Michigan: <https://www.icpsr.umich.edu/web/DSDR/studies/37166/summary>

This dataset contains variables pertaining to LGBTQ lifestyles and habits, including numeric data about individual ratings of health and safety that LGBTQ people have self-reported. The data also has a state-by state breakdown, which will make it easier to combine with the Census data for geospatial analysis. The data comes from the UCLS Williams Institute of Law, which is well known for its studies of social justice issues, and thus I find it to be reliable.

UCLA: <https://dataverse.ucla.edu/dataset.xhtml;jsessionid=396da93010f0b4c91c9c2284cb9e?persistentId=doi%3A10.25346%2FS6%2FN3XTZB&version=&q=&fileTypeGroupFacet=&fileAccess>=

This dataset contains information about the locations and services provided by LQBTQ support organizations. It provides data about geographic location of offices, as well as categorical variables denoting the types of services provided to LGBTQ people. This will help me conduct analysis on the availability of assistance for the LGBTQ people studied in the previous datasets. Again, this will allow me to conduct geospatial analyses on the distribution of LGBTQ people and support systems. This data is again sourced from the UCLA Williams Institute of Law, and so it should be reliable.

## Approach

First, I will gather, compile, and clean my data using R. This will likely include multiple steps of connecting the data to FIPS codes for each of the states and ensuring numeric and categorical data are properly stored. This step will also include removing and making note of areas where data is unavailable or potentially incorrect. Then, I will perform exploratory data analysis on the data using R to determine possible trends I would like to study geospatially. This is the step where I will create my charts to be included in the final report. Based on my findings, I will move to step three, which will be spatial analysis using GeoDa and ArcGIS Pro. In this step, I will first create thematic maps of my data and then move on to doing cluster analysis with the various tools available in GeoDa (e.g. Global/Local Moran’s I, Getis-Ord Gi\* hotspot analysis, etc.). Based on the results from these analyses, I plan to move to geospatial regression to identify potential locations where further support for LGBTQ centers is required. Each of these steps utilizing GeoDa and ArcGIS Pro will be accompanied by maps demonstrating the significance of my cluster analysis. If possible based on my analysis, I will also attempt to predict future trends in the movement of LGBTQ populations.

## Timeline

This is a rough timeline for this project:

- (2 Weeks) Data Collection and Cleaning

- (3 Weeks) Exploratory Data Analysis and Data Visualization

- (2 Weeks) Mapping Data

- (2 Weeks) Cluster and Hotspot Analysis

- (2 Weeks) Geospatial Regression and Modeling Trends

- (1 Weeks) Compiling results

- (1 Week) Writing up the report

- (1 Week) Poster and Final Presentation

## Possible Issues

The main issue I expect to face is the limited availability of within-state LGBTQ population data, which I would like to have to compliment the US Census data. In my research, I have found very few datasets that compile this information for all states, so I will need to find and compile them myself, which has the potential to take longer than expected. Additionally, I feel it will be important that all of my data covers the same time frame. While the majority of the data I have already collected covers the last 15 years, any additional data I may need to locate may again be difficult to track down because of the limited availability of LGBTQ population data. However, there are many datasets available from the Census Bureau, Williams Institute, and Gallup that I can use to fill in the gaps.

Another potential issue is the current lack of preparation for geospatial analysis in my data. The majority of my data is not yet spatially correlated, which means I will need to prepare the data to be mapped myself, as well as ensure that all of the data is connected to the proper locations. Luckily, there are a variety of ways I can do this, so this step should not cause any issues that are impossible to fix.

Finally, given that this project is based heavily in US events that are actively changing (and using data that are based over a period of longer than a decade), my results may ultimately lack relevancy to the current social and political climate. While laws pertaining to LGBTQ individuals have remained relatively stable over the past few years, significant changes are possible even within the next few months that could make my models inaccurate. Unfortunately, this is the reality of such a rapidly changing political climate, so in the later weeks of my project I may need address these changes should new data become available.