

Project Proposal

The Idea:

For my project I wish to redraw the borders of the United States based on census data. This seems to be able to be done with a simple k-means classifier. It is likely feature analysis and selection will be needed.

To increase the difficulty of the project, I will also design a random forest model based on these classifications, and allow a user to input their demographic info, and have it produce what "state" they would belong to.

However, any further feedback on ideas to extend this project to something a bit more difficult would be appreciated. I had to pivot from the music idea when I found that domain to be quite saturated, especially at the level I would be approaching it from.

Needed Resources:

Census Data by census tract

Preliminary Research:

Surprisingly, I was not able to find someone who had done this exact problem before.

[1] This link shows someone who had done a similar thing with congressional districts

[2] This paper provides an alternate method for dividing a state into various polygons of similar population, also intended for use in congressional districts

[3][4] Many seem to have done a clustering based on single or a collection of variables, but it seems none have done one on a larger variable set