We utilize a constrained version of the popular k-means clustering algorithm to assign census block groups to congressional districts in each US state. In ordinary clustering, the number of clusters, k, is subject to choice. In congressional districting, k is stipulated by apportionment, recently completed for the 2020 US Census. Results are demonstrated for the ten most populous states in the US. Click on PDFs for visualization.

For each state, cluster populations are approximately equal and total population-weighted squared geodistance (per longitude and latitude) from block groups to population-weighted cluster centroids is minimized. Census block groups are between larger census tracts and atomic census blocks, a reasonable level of granularity. Ghost block groups with zero population are ignored.

Beside visual PDFs, we provide block group cluster assignment details, our data extraction and algorithmic R-scripts, and algorithmic run details. We downloaded state data files from <https://www2.census.gov/programs-surveys/decennial/2020/data/01-Redistricting_File--PL_94-171/> and trimmed them using modified versions of R-scripts from the Census Bureau <https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html#P1>. Large, downloaded state PL data files are not provided here, nor are trimmed input files, as they are easily reproduced.

Our results show that sensible congressional districts can be formed by algorithm, with no touch by human hands. Until we remove humans, and particularly politicians, from the districting process, we will never end up with fair representation in the US House of Representatives.

Bob Agnew (raagnew1@gmail.com, www.raagnew.com)

Reference Links

<https://github.com/raagnew/Constrained-K-Means-Clustering-in-R>

<https://towardsdatascience.com/k-means-a-complete-introduction-1702af9cd8>c