METHODOLOGY FOR ESTIMATING POPULATION PER ELECTION DISTRICT (ED)

Notes:
- This method assumes that population is evenly distributed across census blocks. Because of this assumption, some "people" (population counts) will inevitably be misallocated to election districts. As a result, there may be some election districts in which ultimately the number of votes cast is greater than the estimated population (and thus will yield a number greater than 1 when normalized).

(1) Download election district shapefiles (hereafter ED) from year of interest (2017: https://data.cityofnewyork.us/City-Government/Election-Districts/h2n3-98hq/data 2009: https://www1.nyc.gov/site/planning/data-maps/open-data/bytes-archive.page?sorts[year]=0&page=20&offset=380)

(2) Download census block groups shapefiles (hereafter BG) from year closest to year of interest (here, 2016 and 2010) https://www.census.gov/geo/maps-data/data/cbf/cbf_blkgrp.html

(3) Download population estimates per BG (2016 5-year ACS estimates, 2010 5-year estimates).

(4) Merge population estimates data onto BG polygons.

(5) Calculate area of each BG: Create a new variable called "areaBG". Use calculate geometry tool to populate field and select square meters as units.

(6) Make a new dataset where polygons are the "UNION" of BG and ED. Allow one meter tolerance and allow gaps. (These datasets were produced by different agencies and so the lines don't overlap exactly perfectly). http://pro.arcgis.com/en/pro-app/tool-reference/analysis/union.htm

(7) Calculate area of each new polygon: Create a new variable called "areaNew". Use calculate geometry tool to populate field and select square meters as units.

(8) Create new variable called "portionPop": Populate field with "areaNew" divided by "areaBG". This tells us what portion of the population from "areaBG" to allocate to "areaNew" (remember, we're assuming population is evenly distributed across BGs). NB: A few observations were marginally greater than one. This is just due to rounding error and the tolerance. I enforced that nothing be greater than one (1 if !portionPop! >= 1 else !portionPop!).

(9) Create a new variable called "adjPop". Populate field with "estimatesTotal" (population estimates variable) multiplied by "portionPop".

(10) Export data as a csv (preserve only election districts ("elecDist") and adjusted population ("adjPop") variables. Use "Export Feature Attribute to ASCII" tool.

(11) Import into Stata. (See "collapse_election_districts.do") Collapse data by ED, summing up the adjusted population. This gives us a rough total of the population per ED. (Maybe there's a way to do this in ArcGIS too). Add a random OID field. Export as a .csv ("adjusted_pop_YEAR.csv")

(12) Import csv back into ArcGis.

(13) Merge "adjusted_pop" onto original ED data (pre-union data) by "elecDist".

(14) Visualize "totalBallots", normalized by "adjPop".