Legislative County Clustering in North Carolina

Looking towards the 2020 Census

Christopher Cooper[1], Blake Esselstyn[2], Gregory Herschlag[3],
Jonathan Mattingly[3], Rebecca Tippett[4]

July 16, 2021

The decennial census is the once-a-decade count of all residents in the United States. Every decade, following the release of the census population counts, we reapportion seats in the U.S. House of Representatives and redraw the U.S. House and state House and Senate district lines ("redistricting") to reflect patterns of population growth over the decade. In a typical census cycle, we would be well into the redistricting process by now: the redistricting data file is usually released by March 31 of the year following the census. The 2020 Census, however, was significantly impacted by the Covid-19 pandemic and other considerations. Currently, the first data for use in redistricting will be released by August 16, 2021, nearly five months behind the typical schedule. Soon after that, we will perform a similar analysis to the one outlined here with the 2020 Census.

The new census counts will bring many changes to district maps in North Carolina. This document will not focus on the 14 congressional districts in North Carolina, but instead focuses on possible changes to districts for the North Carolina General Assembly, which is composed of two bodies: 50 state Senate districts and 120 state House districts.

State assembly district maps in North Carolina are subject to a pair of conflicting legal demands. On one hand, districts must be of approximately equal population to ensure that each representative speaks for essentially the same number of constituents, an extension of the *one person one vote principle*. The population size associated with equal populations across districts is known as the **ideal district size**. On the other hand, there is a state constitutional requirement that counties be kept whole when redistricting. It is typically impossible to satisfy these two requirements simultaneously.

To understand some of the issues in play, consider a county that has a population equal to 1.5 times the ideal district population. To make districts with roughly the ideal district population, one is forced to cross the county boundary with some district, thus cutting the county in two. In North Carolina, this situation was resolved by the 2002 court case *Stephenson v. Bartlett*[5] which set out an algorithm by which the counties in North Carolina should be grouped into county

---

[1] Political Science and Public Affairs, Western Carolina University
[2] FrontWater, LLC and Mapfigure Consulting
[3] Duke Mathematics Department and the Quantifying Gerrymandering Project, Duke University. We thank Alexis Sparko for help with map visualization.
[4] Carolina Demography, UNC at Chapel Hill
[5] Supreme Court of North Carolina.July 16, 2003 357 N.C. 301582 S.E.2d 247

clusterings. Each county clustering has the right number of people so that it might be split into a whole number of districts with populations within the prescribed tolerance of +/- 5% of the ideal district population.[6]

In this note, we examine possible county clusterings as specified in *Stephenson v. Bartlett*. We use 2020 populations from four different data sources ("estimates") to identify potential shifts in county clusterings since 2010. The four population sets used were:

1) **Census Bureau Population Estimate** (CBEst) – the official population estimates for 2020 from the U.S. Census Bureau based on observed births, deaths, and migration between 2010 and 2020.

2) **North Carolina Office of State Budget and Management Population Projections** (NC_OSBM) – county-level projections from the state demographer based on historical patterns of births, deaths, and migration. These projections are modified for counties with age structure impacts of specific institutions, such as colleges, universities, and military barracks.[7]

3) **Redistricting Data Hub** (RedHub) – projections incorporating 2010 Census data, American Community Survey estimates, geocoded voter files, and an unspecified commercial dataset.[8]

4) **Esri** (Esri) - forecasts which take into account data from the US Postal Service, as well as multiple commercial data sources.[9]

Each estimate is adjusted so the statewide population matches the recently released North Carolina population from the 2020 Census.[10]

The material below highlights how small fluctuations in population estimates may lead to dramatically different county clusters. The different source populations each produced very different county clusterings. These clusterings are each different from those used in the last decade based on the 2010 Census.[11] Hence, it is essentially impossible to predict the specific county clusters before the actual 2020 Census population figures are released (by August 16,

---

[6] The algorithm is outlined in Carter et al (2020) and is a greedy algorithm in that it first finds the single counties that could be clusters, then the pairs of counties, then the triplets of counties, and so on. While it gives a definite algorithm, it is not guaranteed to produce the smallest number of split counties. See Carter et al (2020) for more details.

[7] https://www.osbm.nc.gov/media/1547/download

[8] Explanation of methodolgy available at https://redistrictingdatahub.org/wp-content/uploads/2021/05/readme_csv_nc_b_proj_P1_2020tiger.txt

[9] See http://downloads.esri.com/esri_content_doc/dbl/us/J10268_Methodology_Statement_2020-2025_Esri_US_Demographic_Updates.pdf, p. 7.

[10] We adjusted the population figures by calculating the ratio of the statewide population given by the 2020 Census to the estimate's statewide population. We then multiplied each county's population by this ratio to adjust the source's county population so that they sum to a statewide population that agrees with the Census figure of 10,439,388 people.

[11] Even in using the 2010 Census figures, there were two possible clusterings for the House and four for the Senate. See Carter et al (2020) for more details.

2021). However, the findings from these four population estimates are suggestive of general trends. The main takeaways from this study are:

1. We expect that county clusters based on the 2020 Census population will be very different from those used in the last decade.
2. There is significant variation in the county clusters across the different population estimates for 2020. Hence, caution should be exercised when drawing conclusions from the specifics of the maps we have included. We note the few clusters which seem to be stable across the population estimates.
3. Using the 2010 population figures we found 2 possible clusterings for the N.C. Senate. The number of clusterings for the 2020 population estimates ranged from 12 (CBEst, NC_OSBM, and RedHub) to 33 (Esri).
4. Using the 2010 population figures we found 4 possible clusterings for the N.C. House. The number of clusterings for the 2020 population estimates ranged from 2 (CBEst) to 16 (RedHub).

Instability of Clustering

The following table gives the number of possible different clusterings which satisfy Stephenson criteria. It is striking how a slightly different population number leads to drastically different numbers of county clusters in some cases.

Adjusted 2020 Population Estimates	**Number of Different Clusterings**	
	House	**Senate**
Census Estimate	2	12
OSBM	4	12
Redistricting Data Hub	16	12
Esri	4	33

The tables below list all the county clusters seen in at least two of the population estimates. Clusters that appeared in only one set of estimates are not displayed. There are 44 clusters in the House and 31 clusters in the Senate that appear in at least two of population estimates.[12] If a county cluster appears with two different numbers of districts in the grouping it is counted twice.

[12] Overall, there were 97 different county clusters in the House and 94 in the Senate across the four different sets of estimates. This research note focuses only on the 44 and 31, respectively, that appeared in the county clusterings associated with at least two different population estimates.

The columns labeled *conserved* refer to the number of the four population estimates in which the cluster exists in all clusterings for that population estimate. For example, if a specific county cluster exists in all 16 of the different clusterings for the NC House identified with the Redistricting Data Hub population estimates, that cluster is conserved in these estimates.

The columns labeled *exists* refer to the number of the four population estimates in which the cluster exists in at least one clustering for that population estimate. For example, a value of 3 in this column indicates that the cluster exists in the clusters associated with three of the population estimates.

The first two tables give the list of county clusters that are conserved across all four of the population estimates. The first table gives the conserved county clusters for the Senate and the second for the House. Notice that only a small number of county clusters are conserved across all of the clusterings and across all of the population projections.

Senate County Grouping	**# of Districts in Grouping**	**2020 Adjusted Population Estimates**	**Does Cluster exist in Current Clustering ?**
		# Grouping Conserved	
Davidson-Davie	1	4	-
Cabarrus	1	4	-
Rowan-Stanly	1	4	Yes
Wayne-Wilson	1	4	-
Greene-Pitt	1	4	Yes
Guilford-Rockingham	3	4	-
Beaufort-Craven-Lenoir	1	4	-
Johnston	1	4	-

House County Grouping	**# of Districts in Grouping**	**2020 Adjusted Population Estimates**	**Does Cluster exist in Current Clustering ?**
		# Grouping Conserved	
Alamance	2	4	Yes
Pitt	2	4	-
Durham-Person	4	4	-
Haywood-Madison	1	4	-
Hoke-Scotland	1	4	Yes
Cherokee-Clay-Graham-Macon	1	4	Yes
Buncombe	3	4	Yes
Guilford	6	4	Yes
Lincoln	1	4	Yes
Davidson	2	4	Yes

The next two tables give a list of county clusters that are not conserved across all population estimates but appear in at least two different population estimates. For example, if a cluster is conserved in 0 estimates and exists in 4 estimates, that means the cluster shows up in clusterings associated with all of the population estimates but it is not present in all of the unique clusterings associated with any specific estimate. Both the number of population estimates for which it is conserved and for which it exists are given in the two rightmost columns.

Notice that in the House, Wake County has 12 seats for some population figures and has 13 for others. Additionally, many counties are part of a number of different clusters.

Senate County Grouping	**# of Districts in Grouping**	**2020 Adjusted Population Estimates**		**Does Cluster exist in Current Clustering?**
		# Grouping Conserved	**# Grouping Exists**	
Buncombe-Henderson-Polk	2	1	4	-
Burke-Gaston-Lincoln	2	0	4	-
Cleveland-McDowell-Rutherford	1	0	4	-
Edgecombe-Franklin-Nash	1	0	4	-
Burke-McDowell-Rutherford	1	0	4	-
Cleveland-Gaston-Lincoln	2	0	4	-
Onslow	1	3	3	-
Cumberland-Moore	2	3	3	-
Franklin-Halifax-Nash	1	0	3	-
Henderson-Polk-Rutherford	1	0	3	-
Buncombe-Burke-McDowell	2	0	3	-
Bladen-Harnett-Lee-Robeson-Sampson	2	2	2	-
Alexander-Iredell	1	2	2	-
Wake	5	2	2	-
Ashe-Avery-Caldwell-Catawba-Cherokee-Clay-Graham-Haywood-Jackson-Macon-Madison-Mitchell-Swain-Transylvania-Watauga-Yancey	3	2	2	-
Durham-Wake	7	2	2	-
Chatham-Randolph	1	2	2	-
Alleghany-Stokes-Surry-Wilkes	1	2	2	-
Alexander-Surry-Wilkes-Yadkin	1	2	2	-
Brunswick-Columbus	1	2	2	-
Forsyth-Stokes	2	2	2	-
Brunswick-Columbus-New Hanover	2	2	2	-
Forsyth-Yadkin	2	2	2	-

House County Grouping	# of Districts in Grouping	2020 Adjusted Population Estimates		Does Cluster exist in Current Clustering?
		# Grouping Conserved	# Grouping Exists	
Caswell-Orange	2	3	4	Yes
Forsyth-Stokes	5	2	4	-
Rockingham	1	3	3	-
Ashe-Watauga	1	3	3	Yes
Franklin-Granville-Vance	2	3	3	-
Anson-Union	3	3	3	Yes
Montgomery-Stanly	1	3	3	-
Carteret-Pamlico	1	3	3	-
Nash-Wilson	2	3	3	-
Burke	1	3	3	-
Cabarrus-Rowan	4	3	3	-
Mecklenburg	13	3	3	Yes[13]
Bertie-Chowan-Halifax	1	3	3	-
Chatham-Lee-Moore-Randolph-Richmond	5	2	3	Yes
Alleghany-Surry	1	2	3	Yes
Harnett-Johnston	4	2	3	-
Iredell	2	2	2	Yes
Wake	12	2	2	Yes[13]
Wake	13	2	2	Yes[13]
Bladen-Robeson	2	2	2	-
Edgecombe-Martin-Washington	1	2	2	-
Brunswick-Columbus-New Hanover	5	2	2	-
Duplin-Wayne	2	2	2	-
Avery-Cleveland-Gaston-Henderson-Jackson-McDowell-Mitchell-Polk-Rutherford-Swain-Transylvania-Yancey	8	2	2	-
Cumberland	4	2	2	Yes
Caldwell	1	2	2	Yes
Greene-Jones-Lenoir	1	2	2	-
Onslow-Pender	3	1	2	-
Alexander-Catawba-Davie-Wilkes-Yadkin	4	1	2	-
Camden-Gates-Hertford-Northampton-Warren	1	1	2	-
Johnston-Sampson	3	1	2	-
Duplin-Onslow	3	1	2	Yes
Chatham-Harnett-Lee	3	0	2	-
Davie-Forsyth	5	0	2	-

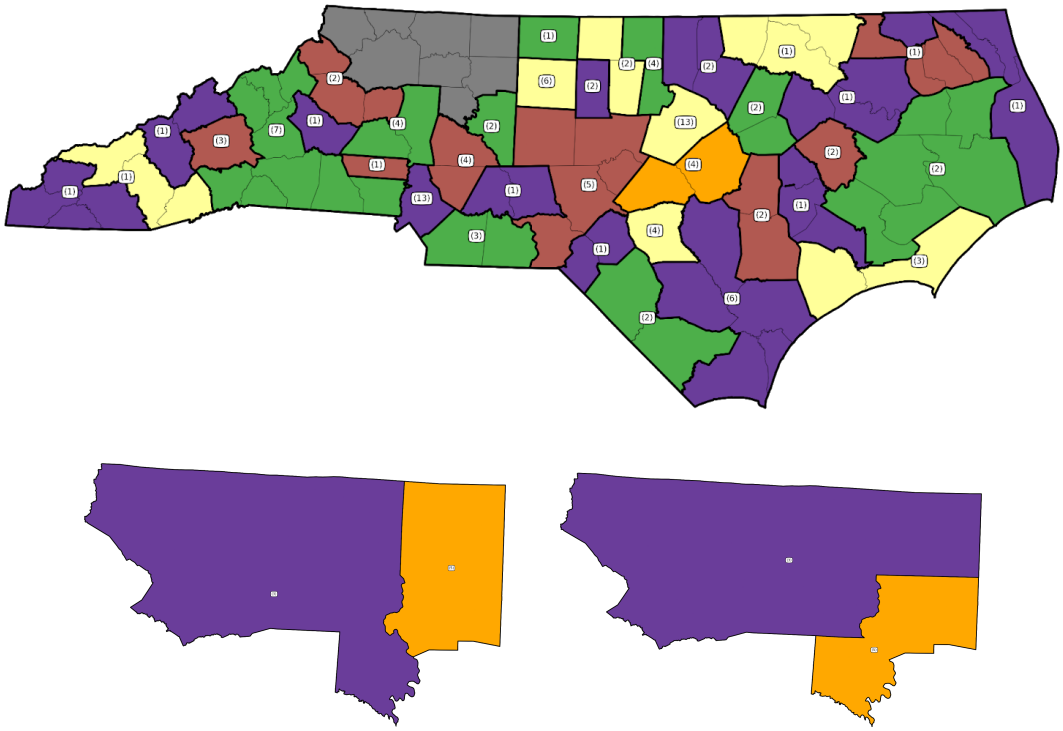
[13] The Cluster is preserved from current clustering, but the number of seats has changed. Hence the districts in this cluster will need to be redrawn.

Maps for Estimated Population Figures

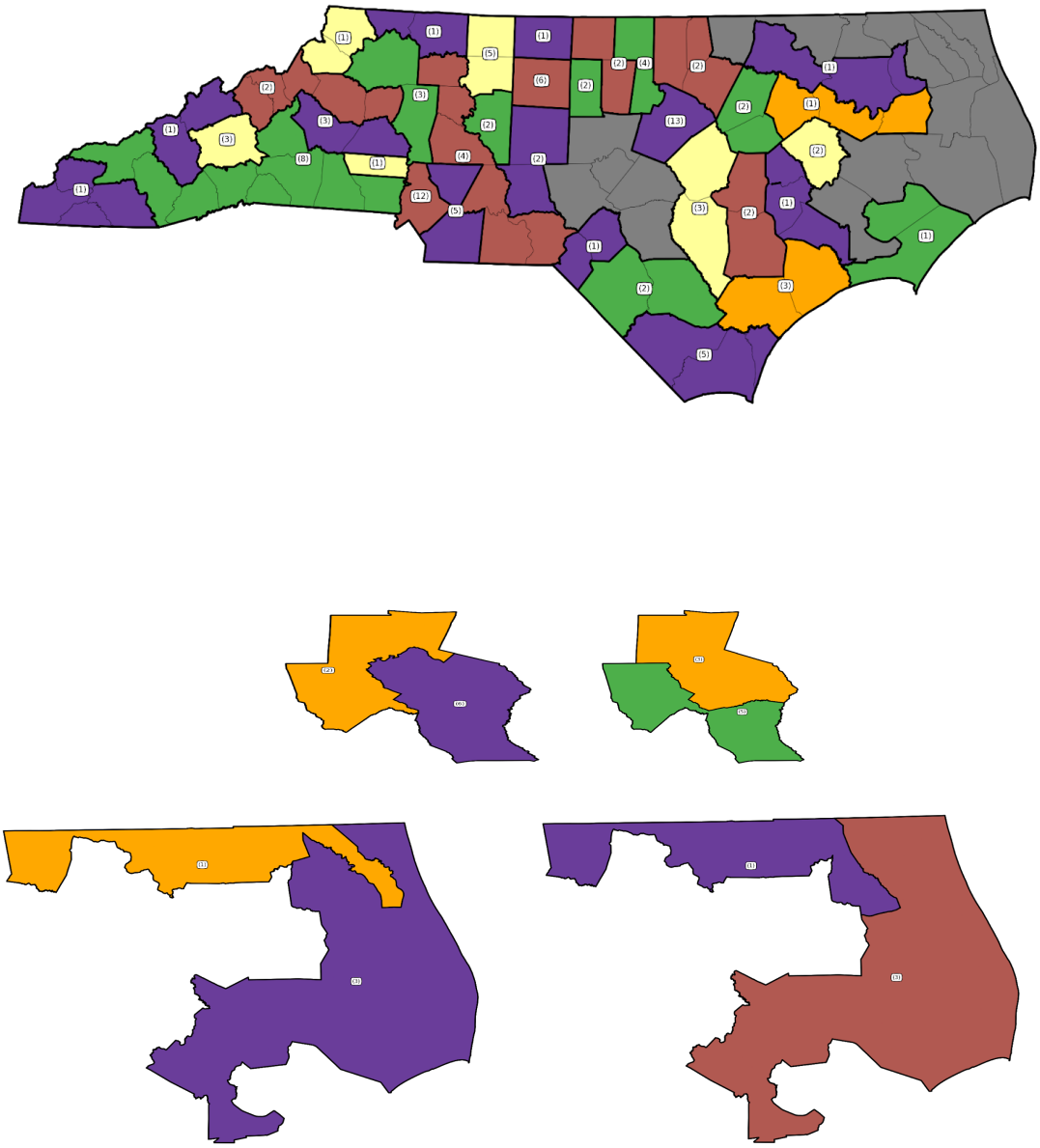
On the following pages, the main maps for each population estimate give the county clusters that are conserved across the estimate. The smaller insets give the options for the grayed-out regions, representing the choices producing different clusterings. The figures underline the volatility of the clustering across different projections. There is no *unique* clustering for a given population estimate. We emphasize that given the instability of these maps across different estimates, they should not be taken as predictions of the clusterings.

House Maps

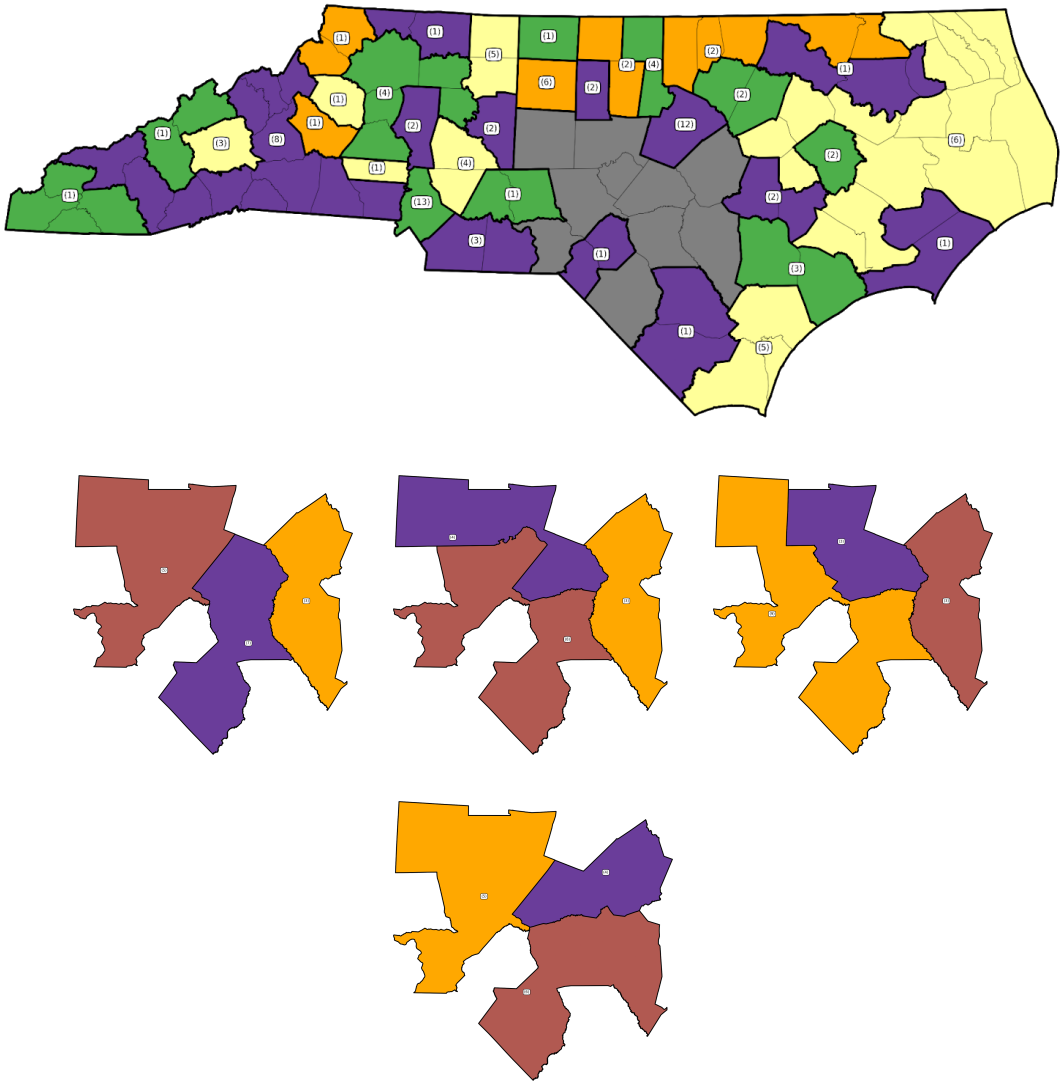
Census Estimated



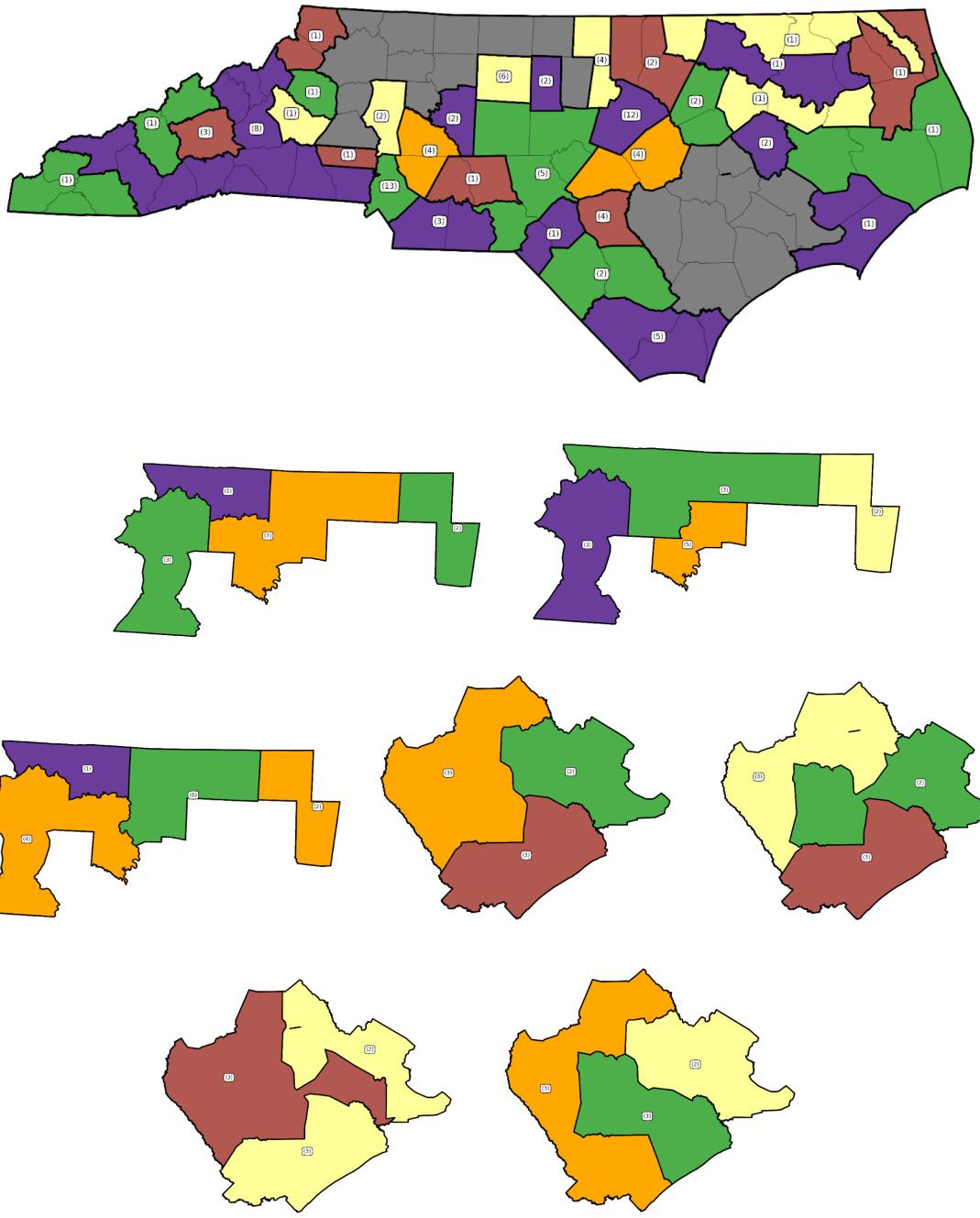
Esri



OSBM

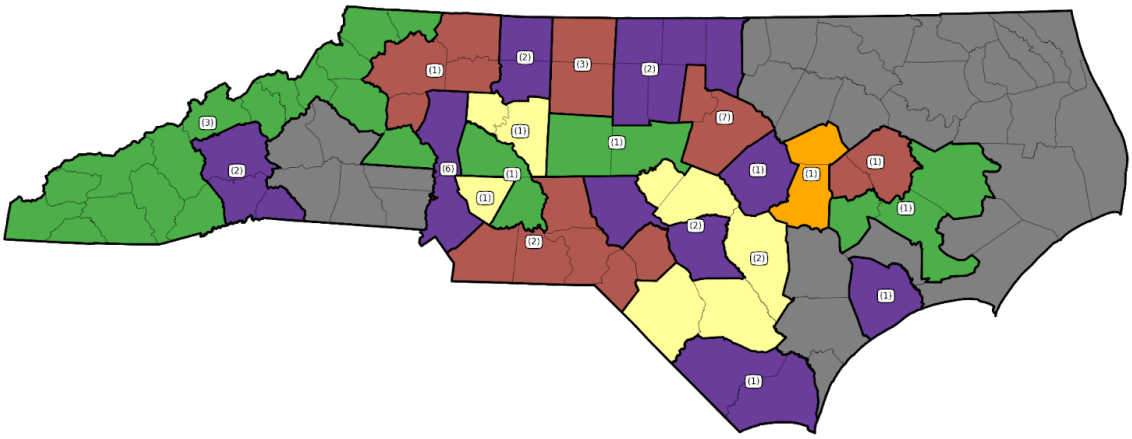


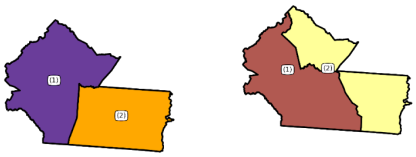
Redistricting Data Hub

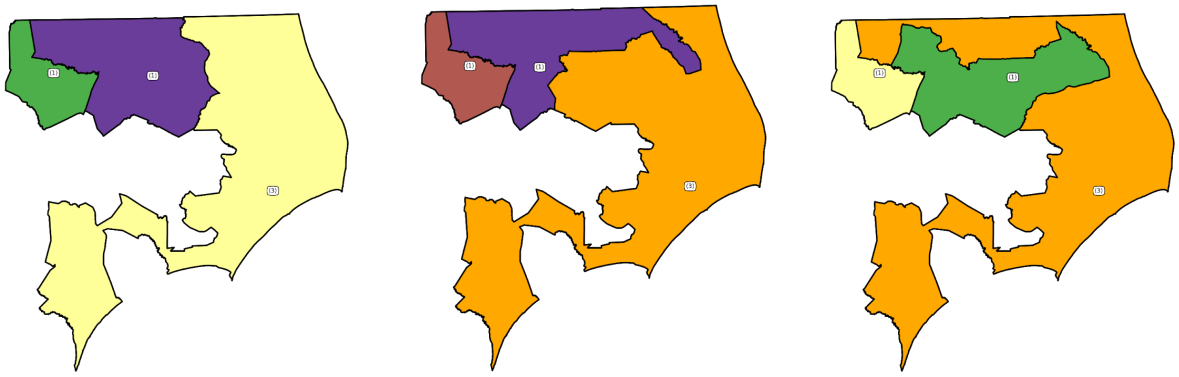


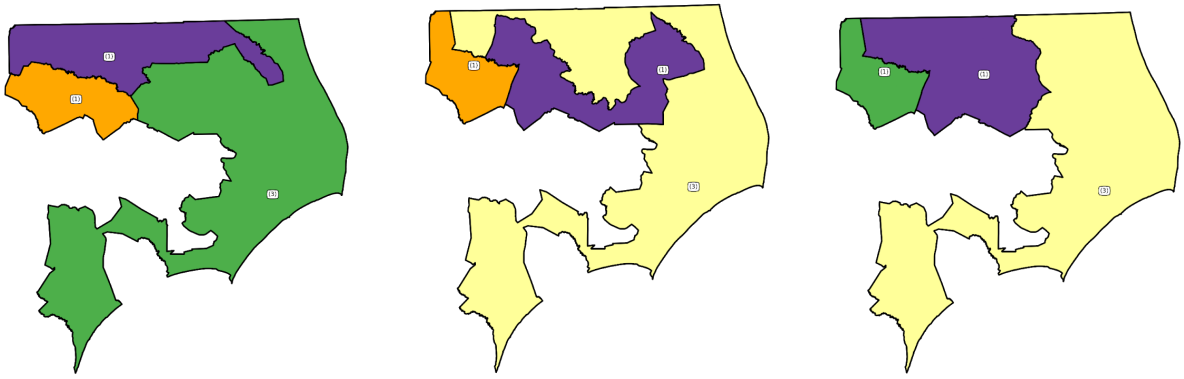
Senate Maps

Census Estimated

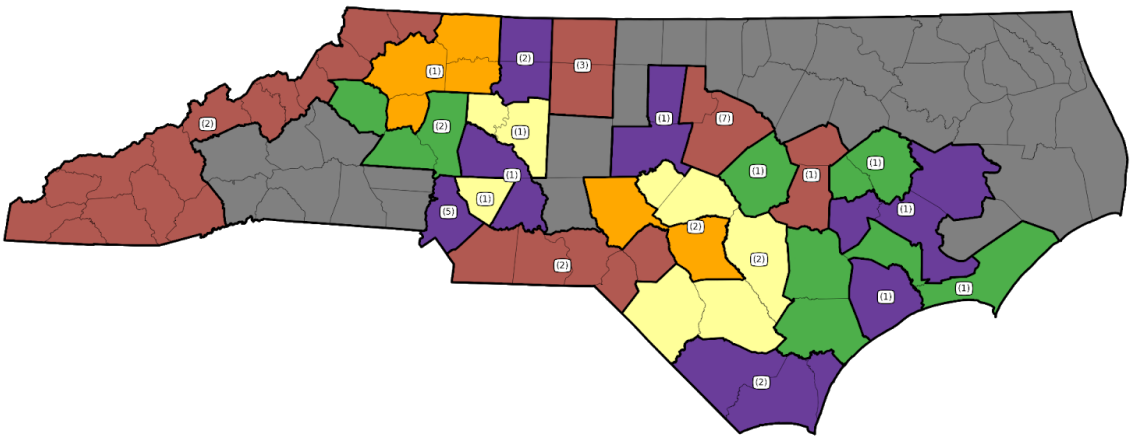


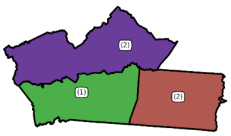


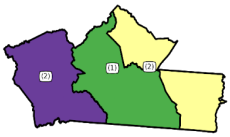


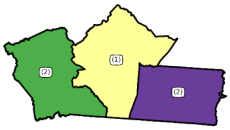


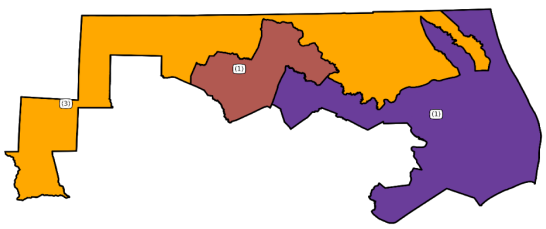
Esri

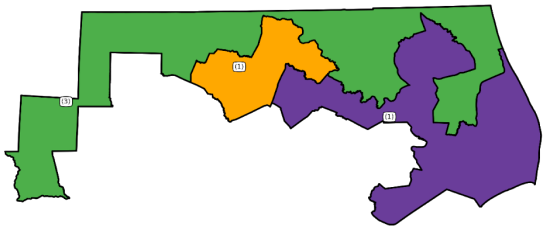


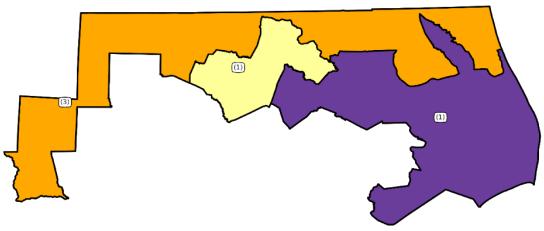


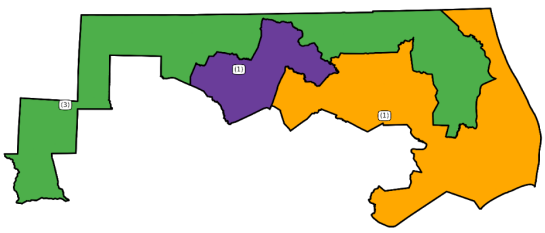


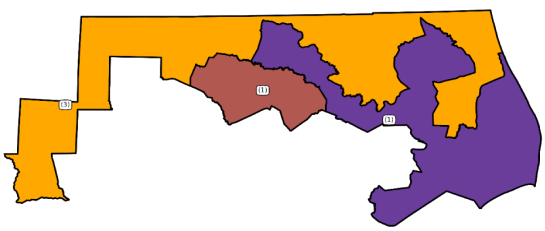


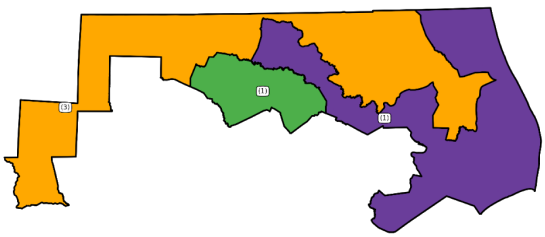


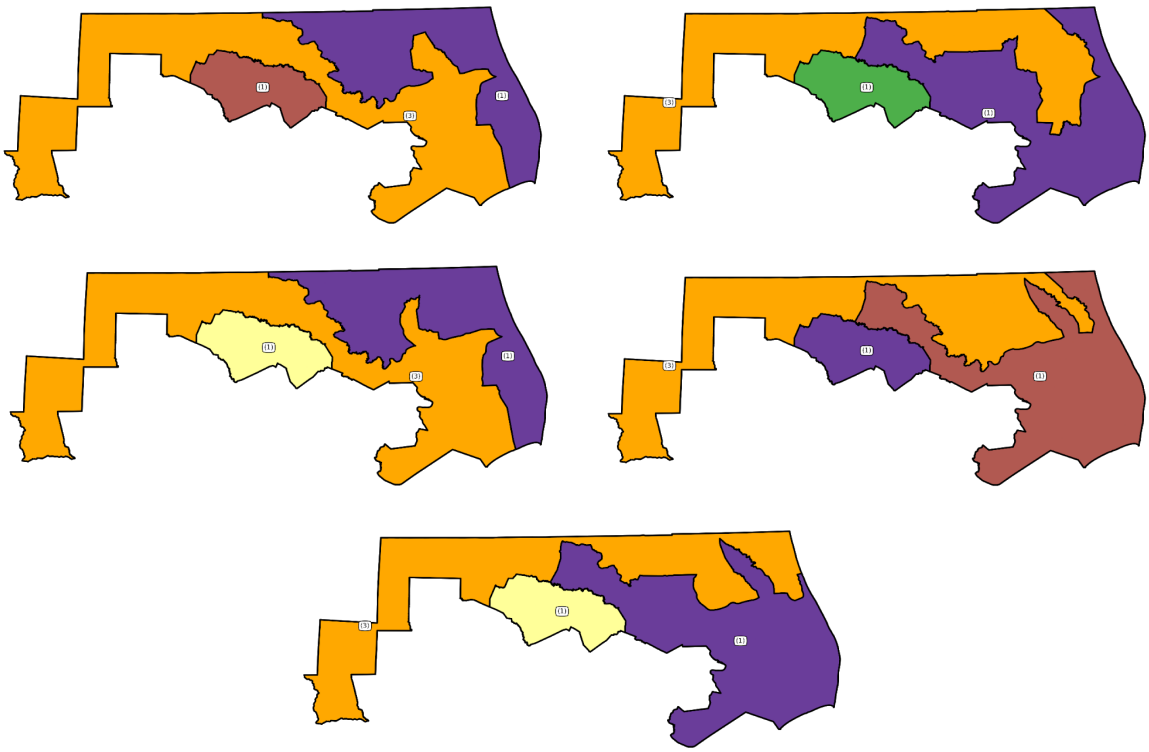




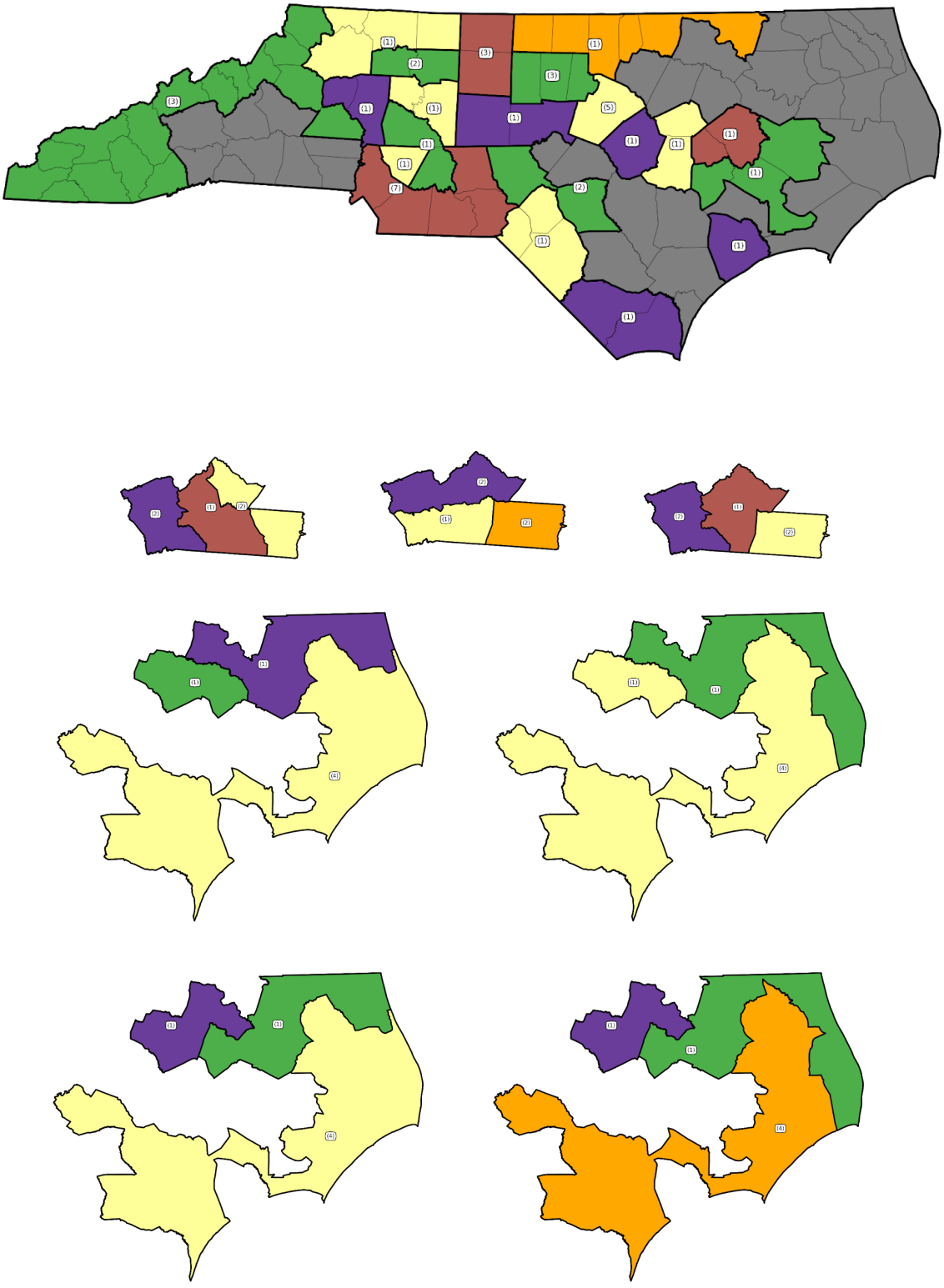
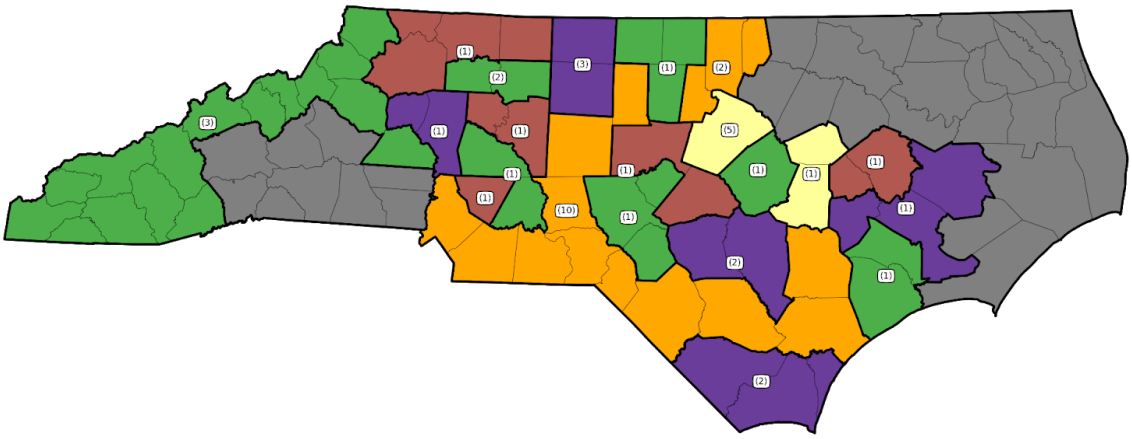
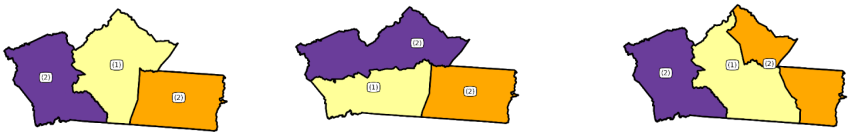
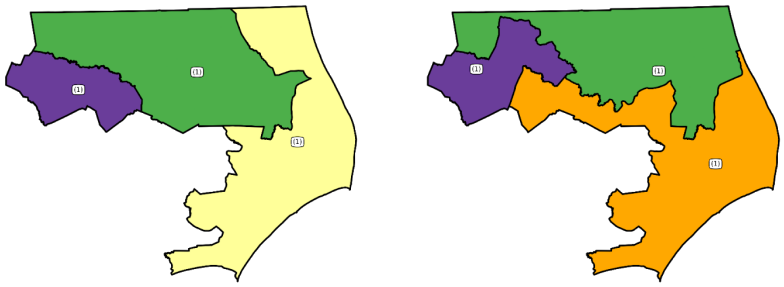
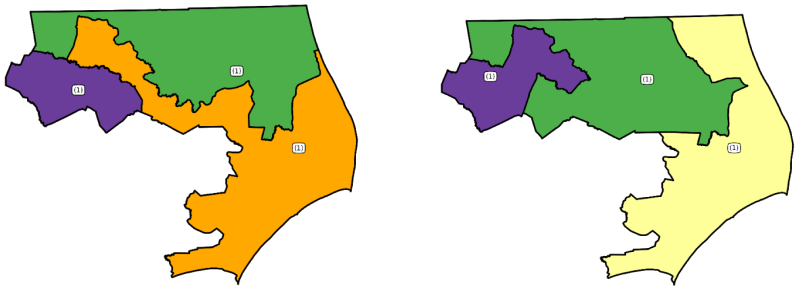








OSBM



Redistricting Data Hub