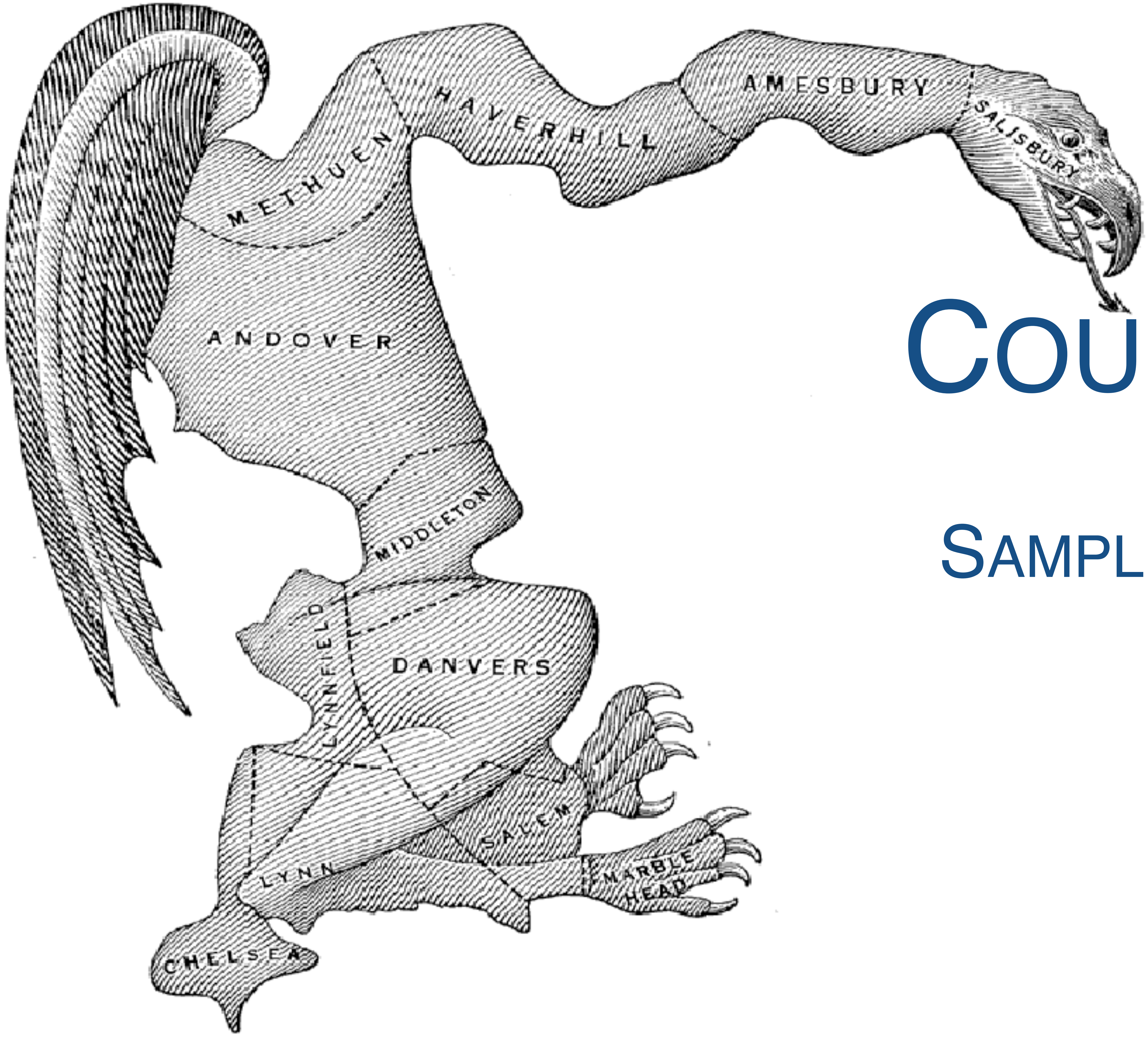


COUNTY PRESERVATION

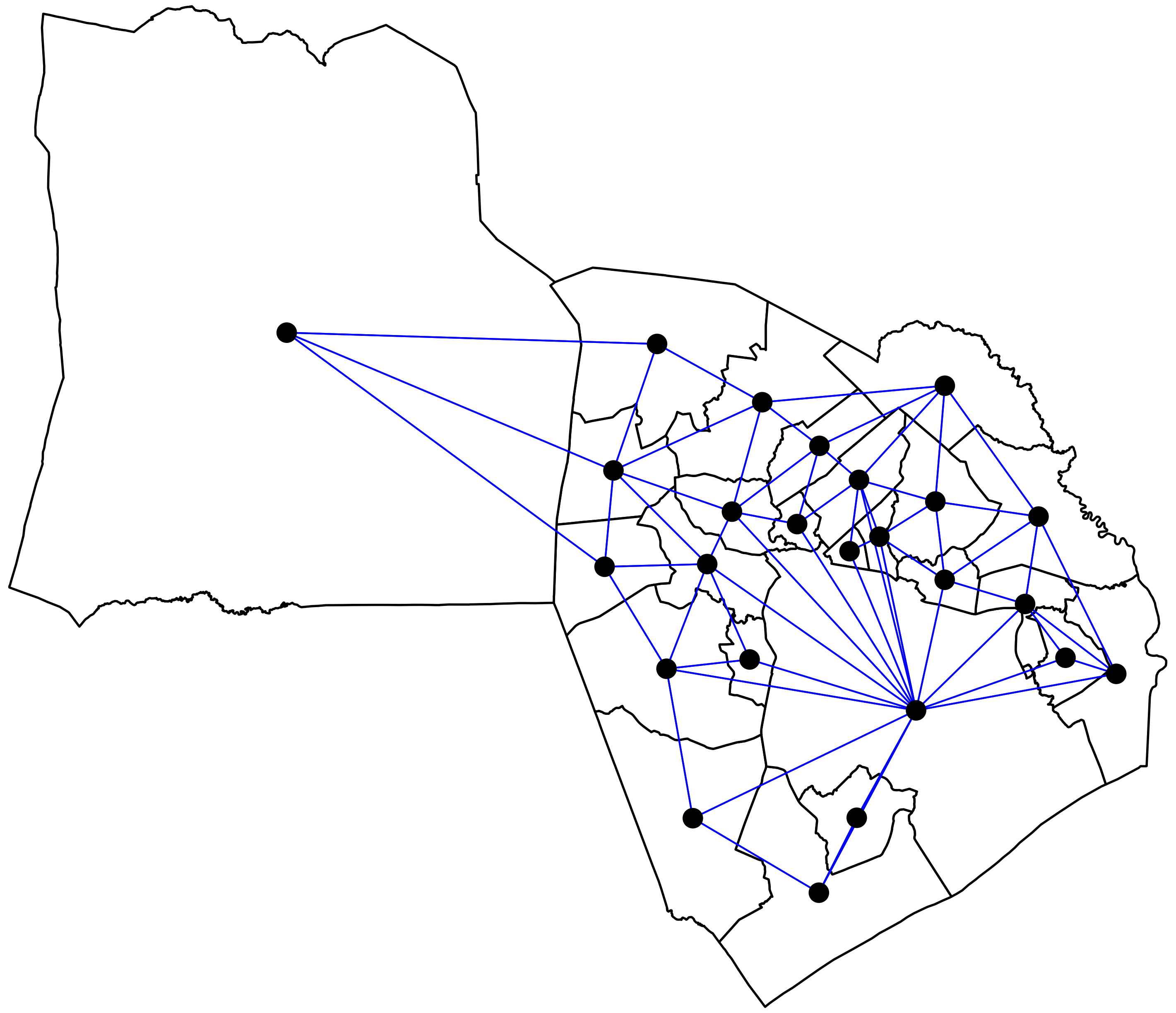
SAMPLING, CLUSTERING, CONSTRAINING

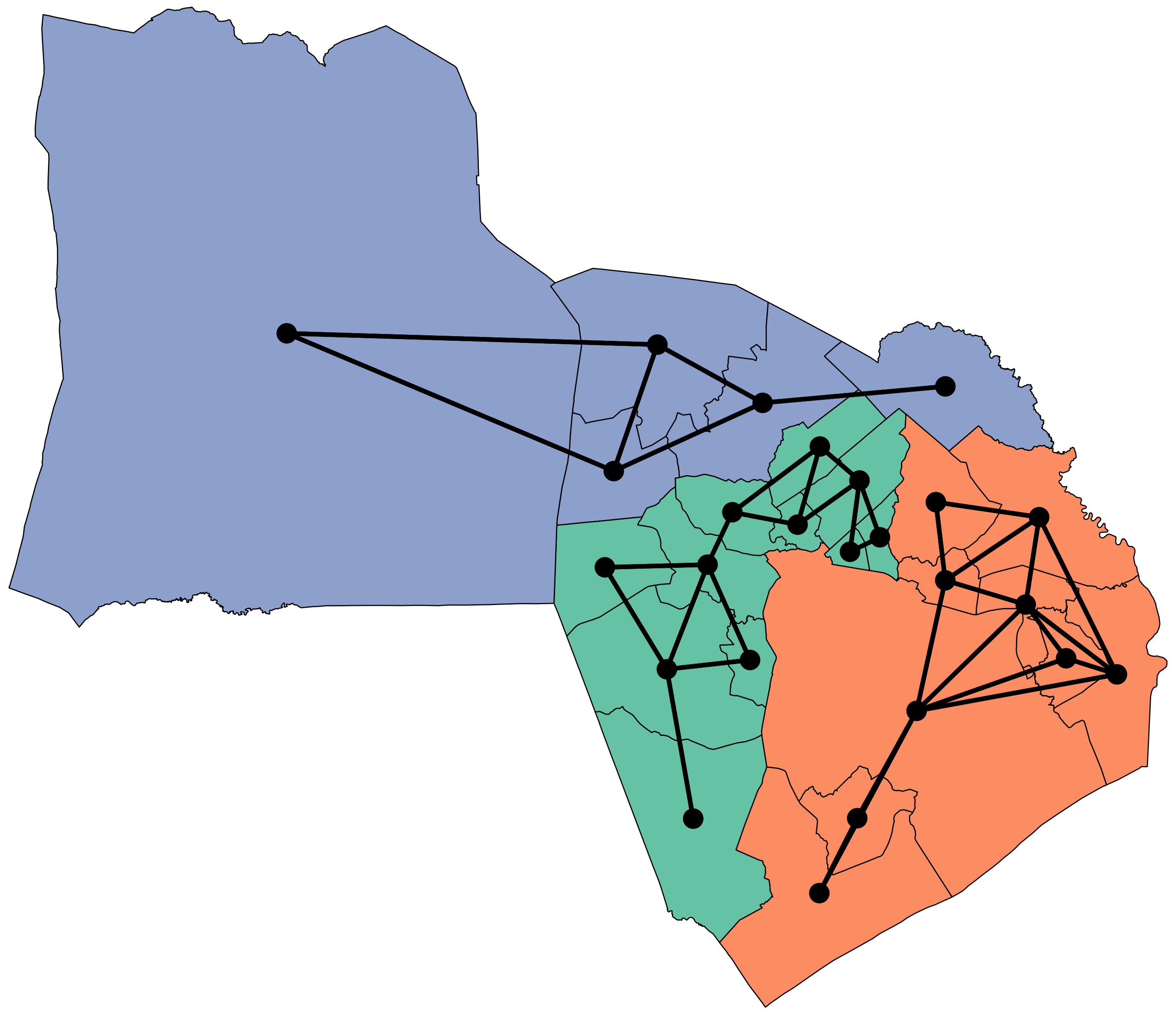
QUANTIFYING GERRYMANDERING
MARCH 3RD, 2020

GREGORY HERSCHLAG, JONATHAN MATTINGLY,
DANIEL CARTER, ZACH HUNTER, ERIC AUTRY
+THE TEAM AT DUKE (AND BEYOND)

If n districts:

ξ: (**precincts**) $\longrightarrow \{1, 2, \cdots, n\} \iff \xi = (\xi_1, \ldots, \xi_n)$ **partition into sub-graphs**

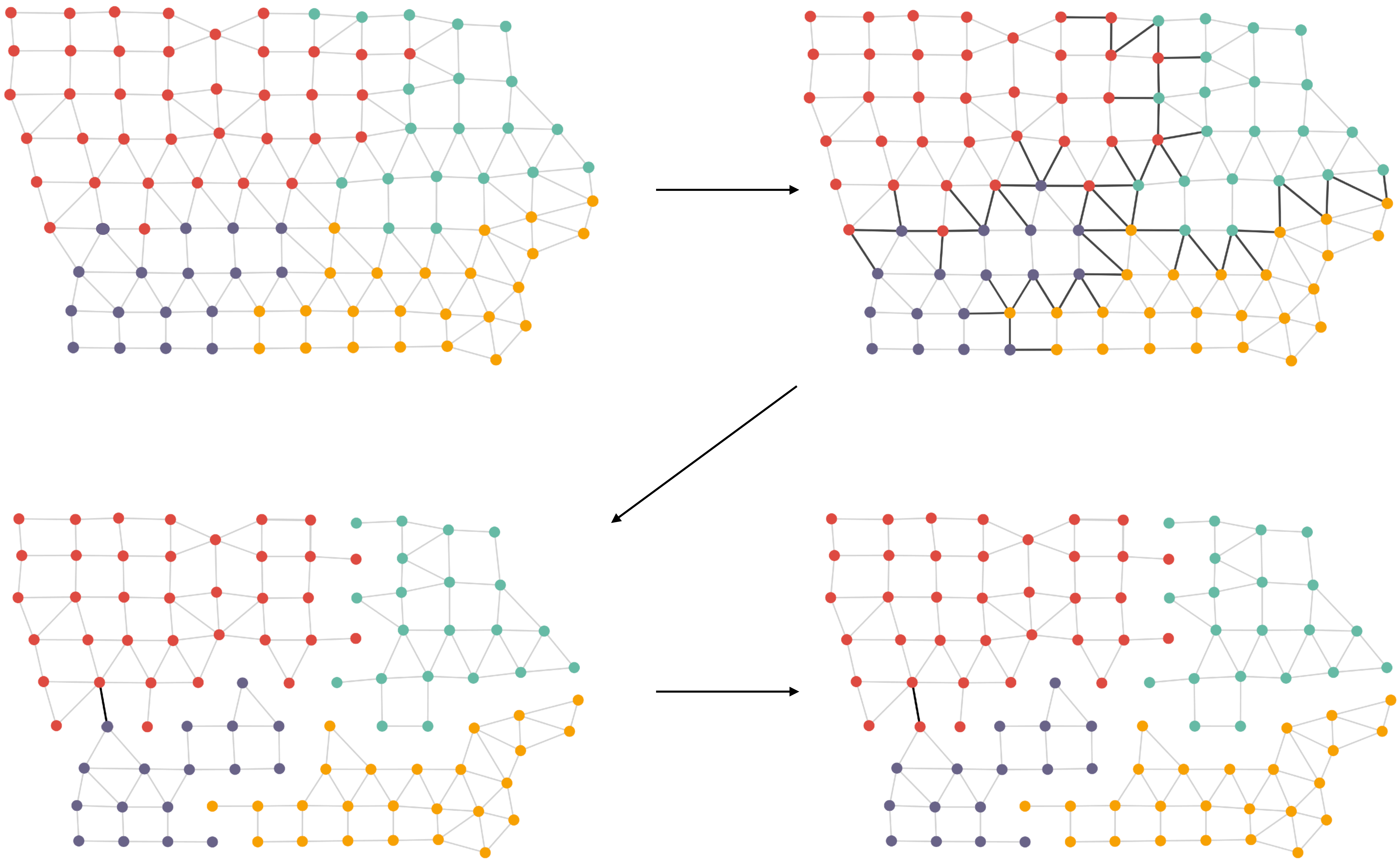




(**Probability of plan**) $\propto e^{-\beta(\textbf{Plan Score})}$

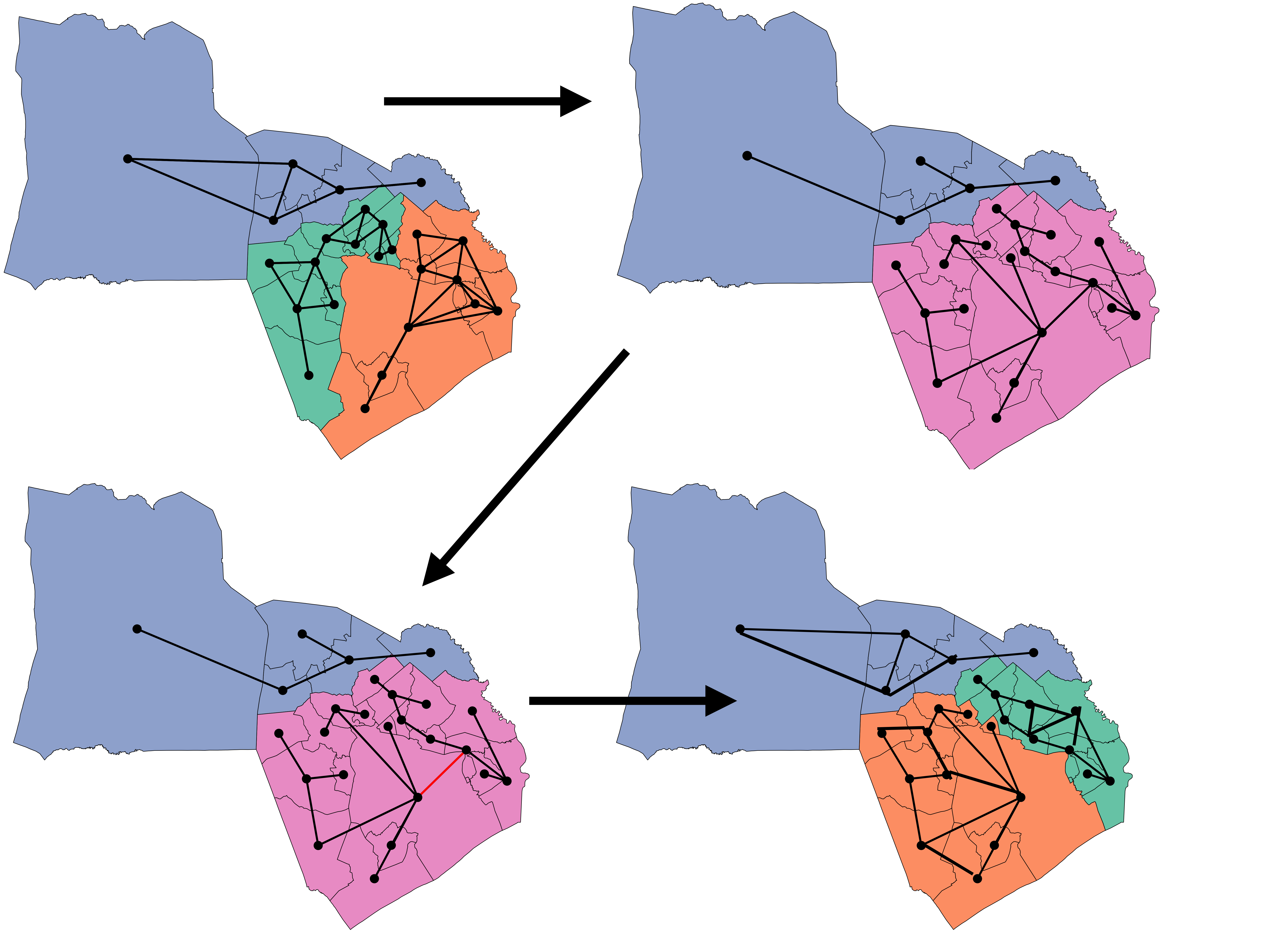
$$\mu(\xi) = \frac{1}{Z} e^{-\beta J(\xi)}$$

Single Node Flip Markov Chains



<u>Then accept/reject according to a score function</u>

One Step of ReCom Markov Chain



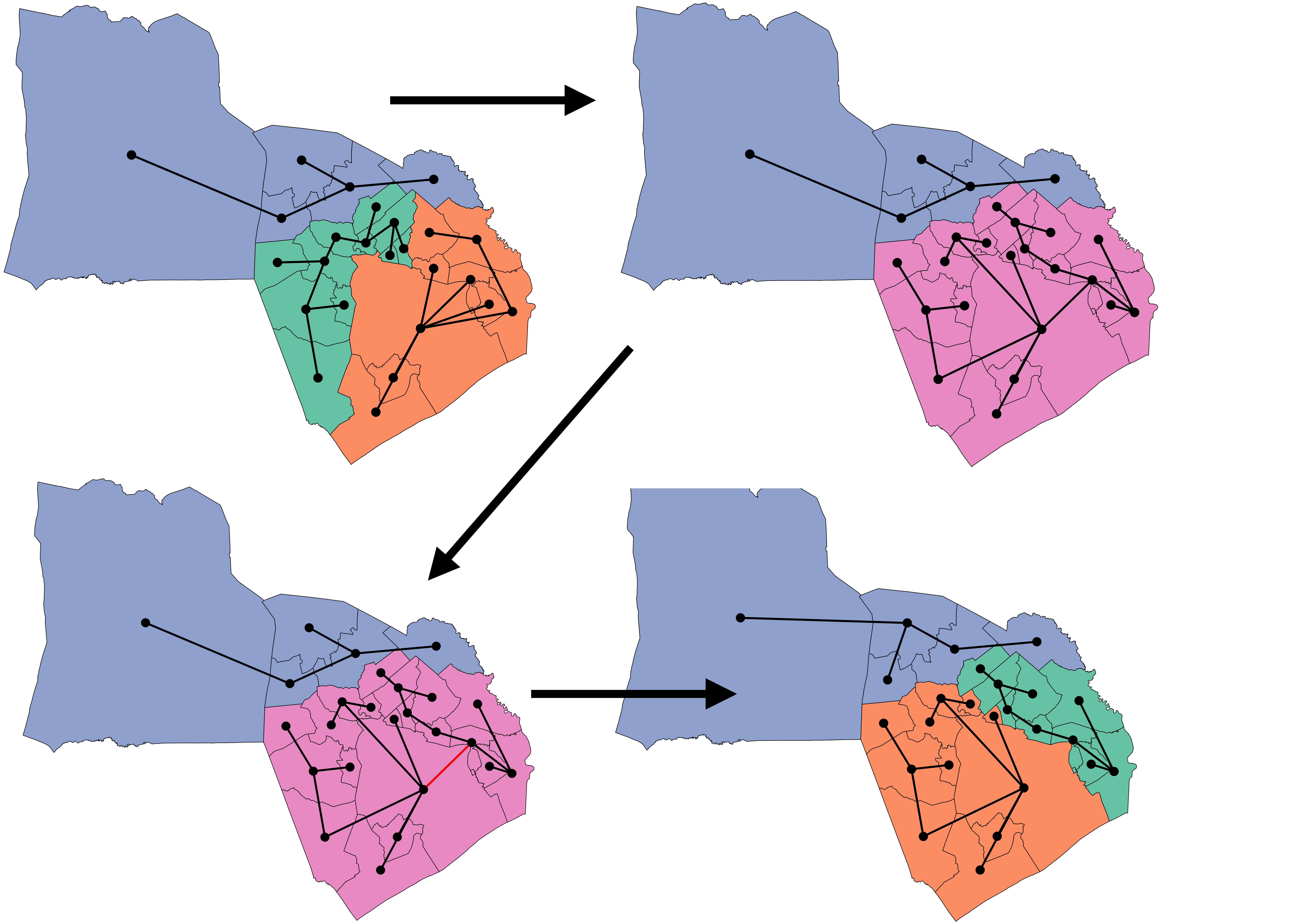
DeFord, Duchin, Solomon

ReCom Algorithm

1. Pick adjacent pair of districts to merge
2. Draw Spanning tree on merged graph (Willson's Alg)
3. Find permissible cuts (e.g. within Pop constraint)
4. cut in two, return new subgraphs

https://arxiv.org/abs/1911.05725

Or expanded state space in Merge Split



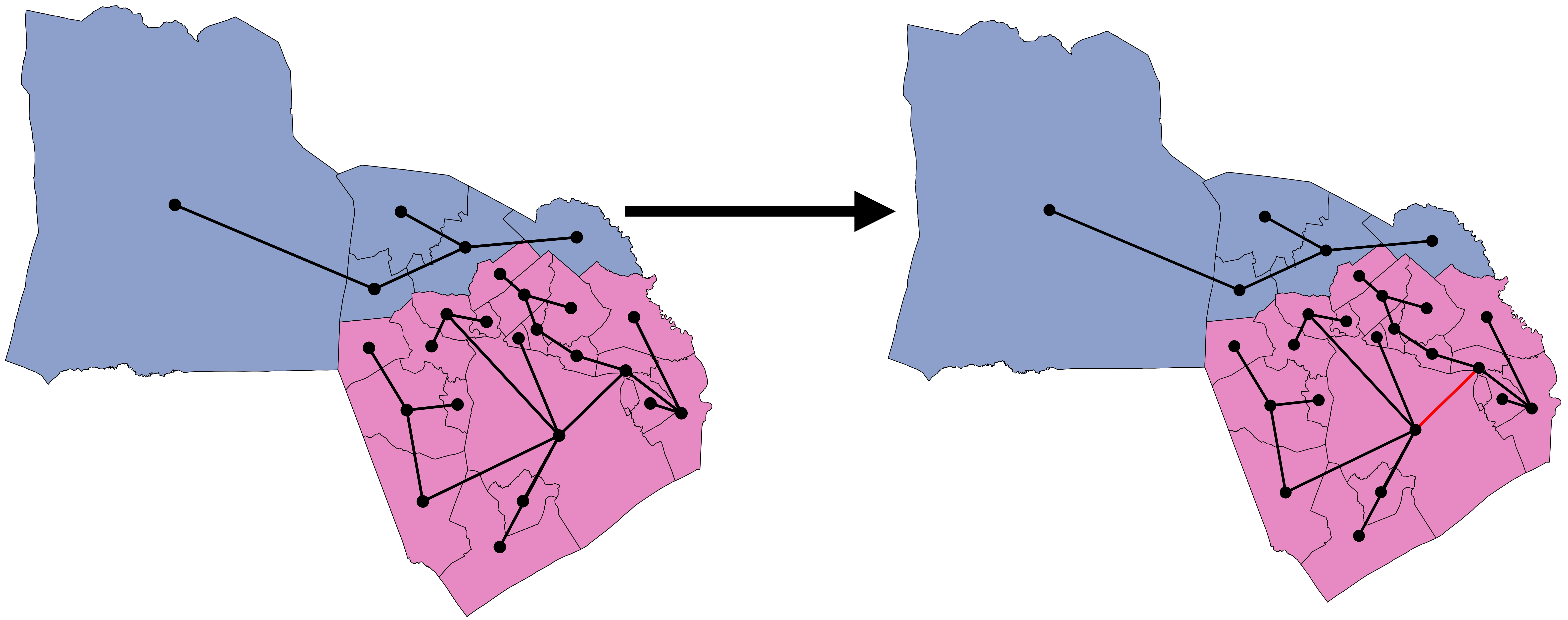
Carter, GH, **Hunter, Mattingly**

Merge-Split Algorithm

1. **Pick adjacent pair of districts to merge**
2. **Draw Spanning tree on merged graph (Willson's Alg)**
3. **Find permissible cuts (e.g. within Pop constraint)**
4. **Cut in two, return new subgraphs**

https://arxiv.org/abs/1911.01503

Aside on Recom and improvements



$$P_{tree}(T_A, T_B, e)P_{cut}(e) = \frac{1}{\tau(A \cup B)} \frac{1}{|E_{cut}|}$$

Needed for Recom (computationally intractable)

$$\sum_{T_A \in ST(A)} \sum_{T_B \in ST(B)} \sum_{e \in A \cup B} P_{tree}(T_A, T_B, e)P_{cut}(e) = \sum_{T_A \in ST(A)} \sum_{T_B \in ST(B)} \sum_{e \in A \cup B} \frac{1}{\tau(A \cup B)} \frac{1}{|E_{cut}|(T_A, T_B, e)}$$

Merge-Split

$$\sum_{e \in A \cup B} P_{tree}(T_A, T_B, e)P_{cut}(e) = \sum_{e \in A \cup B} \frac{1}{\tau(A \cup B)} \frac{1}{|E_{cut}|(T_A, T_B, e)}$$

Reversible Recom

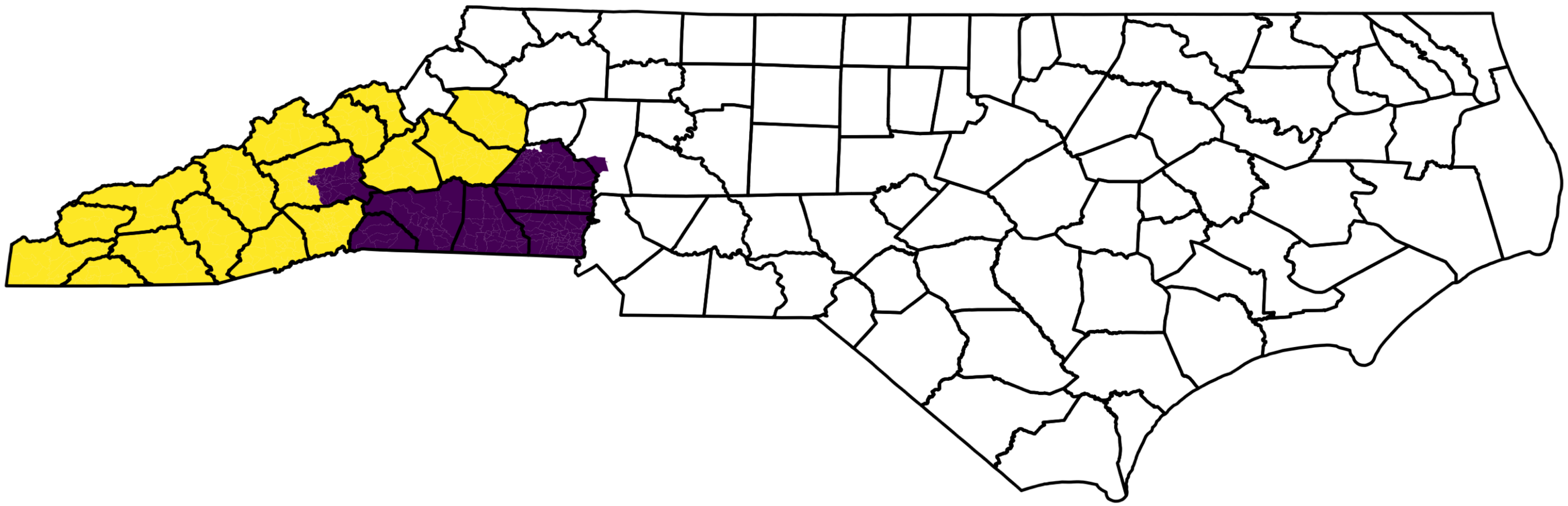
$$\sum_{T_A \in ST(A)} \sum_{T_B \in ST(B)} \sum_{e \in A \cup B} P_{tree}(T_A, T_B, e)P_{cut}(e) = \sum_{T_A \in ST(A)} \sum_{T_B \in ST(B)} \sum_{e \in A \cup B} \frac{1}{\tau(A \cup B)} \frac{1}{m}$$

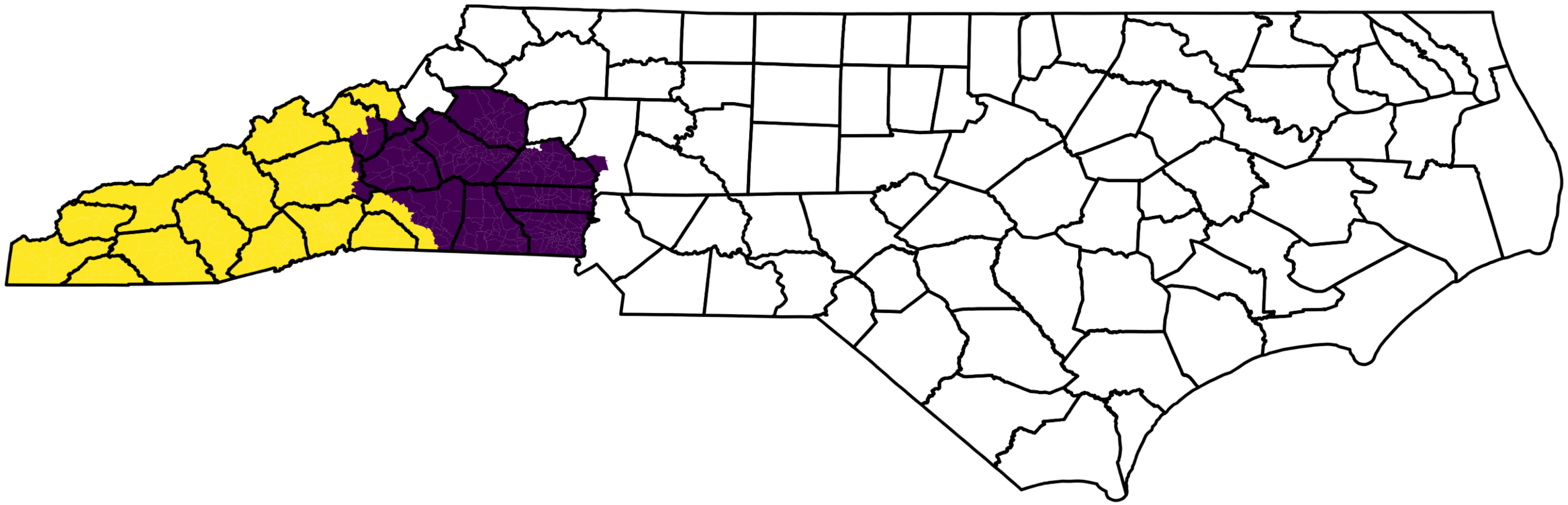
County preservation in North Carolina

- ***Congregational***
  - ***2016 Criteria: "Division of counties shall only be made for reasons of equalizing population, consideration of incumbency and political impact. Reasonable efforts shall be made not to divide a county into more than two districts."***
- ***Legislative***
  - ***County clusters***
  - ***Minimization of traversals***
  - ***When counties are split, keep them together as much as possible***

https://www.ncleg.gov/Redistricting/Process2016

Difficulty of county preservation





What are the minimal number of county splits?

Theorem: d districts must introduce at least d-1 county splits (when we can’t evenly partition counties)

County splits: the county splits of a county are the number of districts a county intersects minus 1

Conjecture: In nearly all redistricting problems, the bound from the theorem will be tight

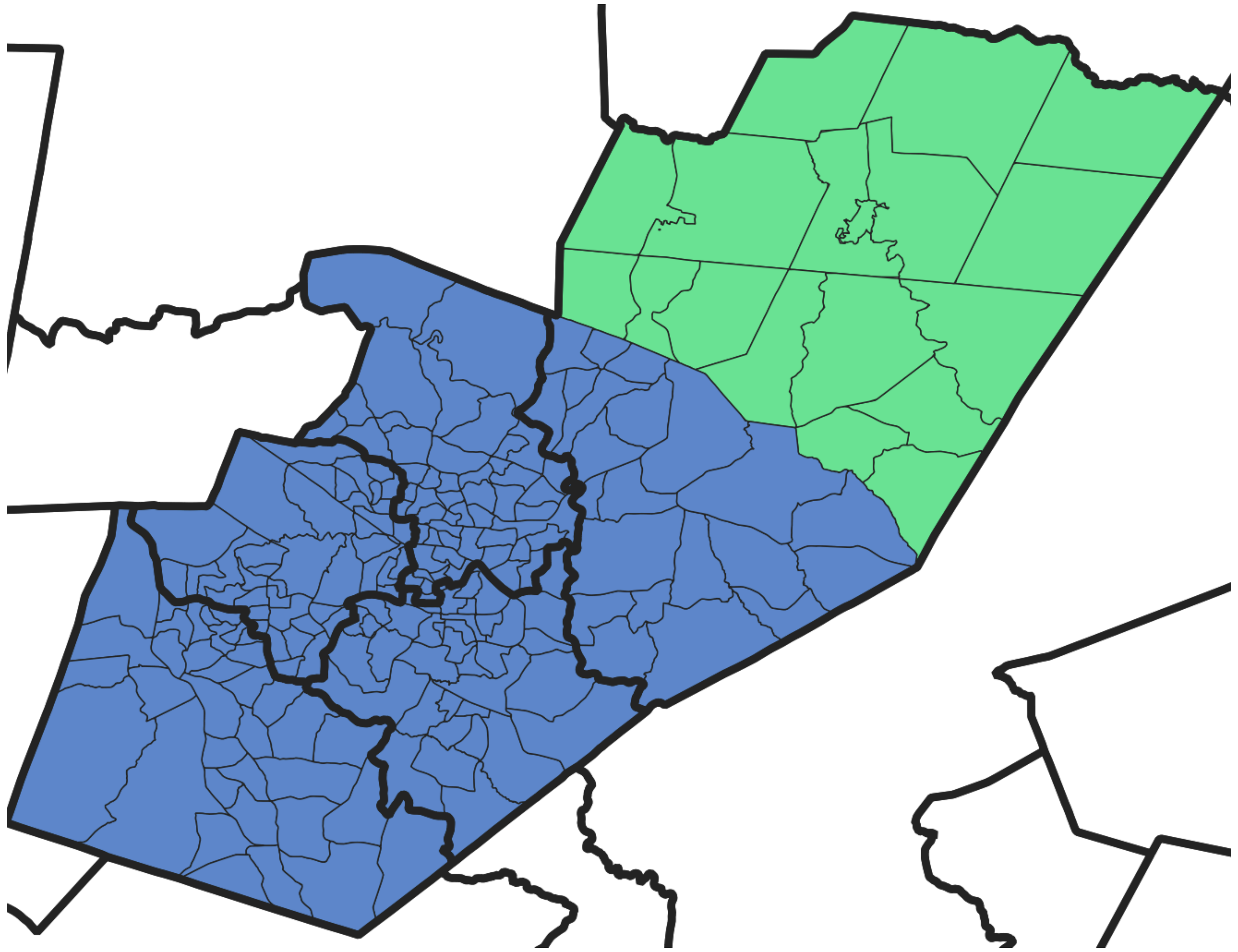
Carter, GH**, Hunter, Mattingly**
https://arxiv.org/abs/1908.11801; under review

What are the minimal number of county splits?

<u>Theorem</u>: d districts must introduce at least d-1 county splits (when we can't evenly partition counties)

<u>County splits</u>: the county splits of a county are the number of districts a county intersects minus 1

<u>Conjecture</u>: In nearly all redistricting problems, the bound from the theorem will be tight



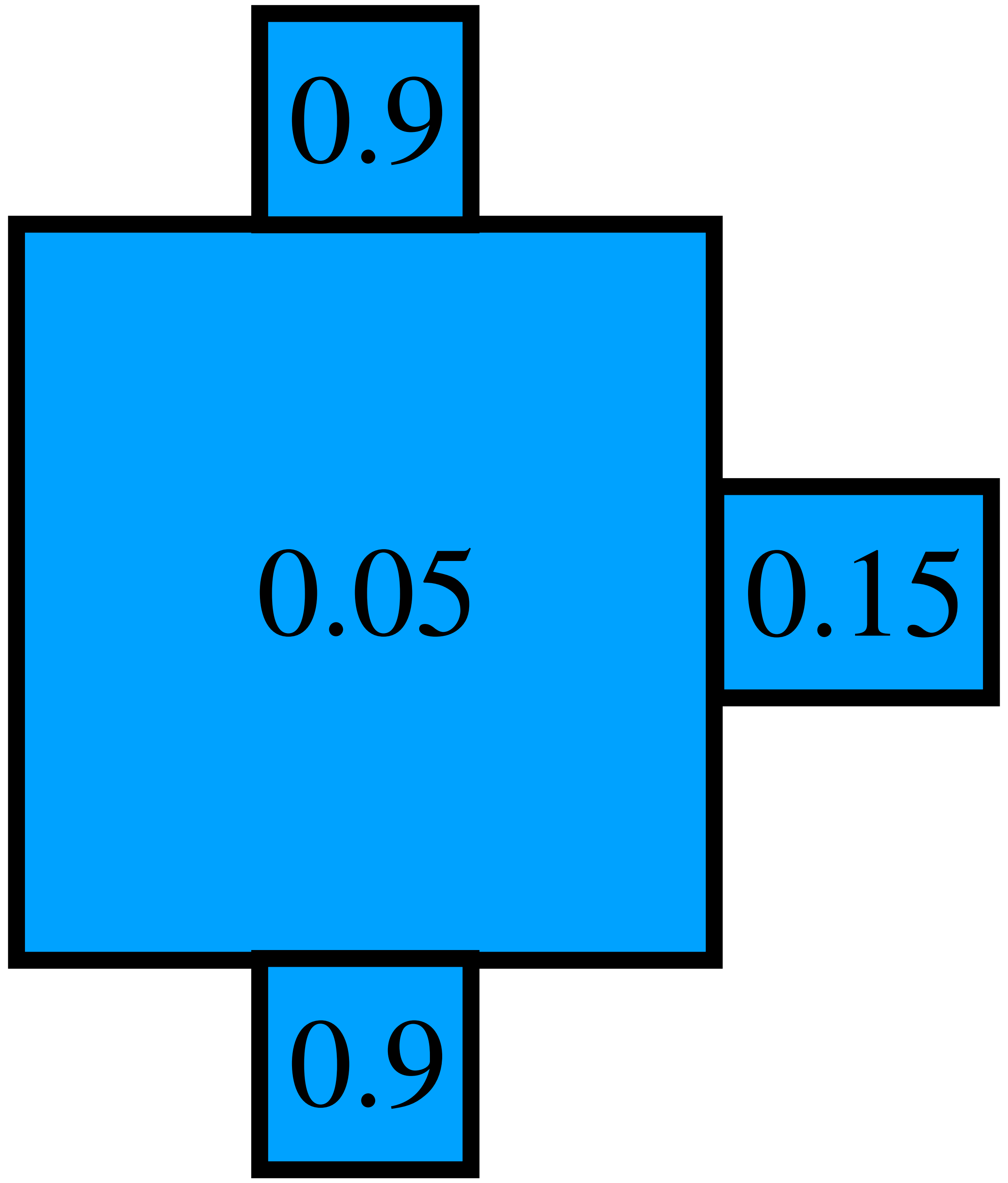
Carter, GH**, Hunter, Mattingly**
https://arxiv.org/abs/1908.11801; under review

What are the minimal number of county splits?

<u>Theorem</u>: **d districts must introduce at least d-1 county splits (when we can't evenly partition counties)**

<u>County splits</u>: **the county splits of a county are the number of districts a county intersects minus 1**

<u>Conjecture</u>: **In nearly all redistricting problems, the bound from the theorem will be tight**



Fraction of a district population

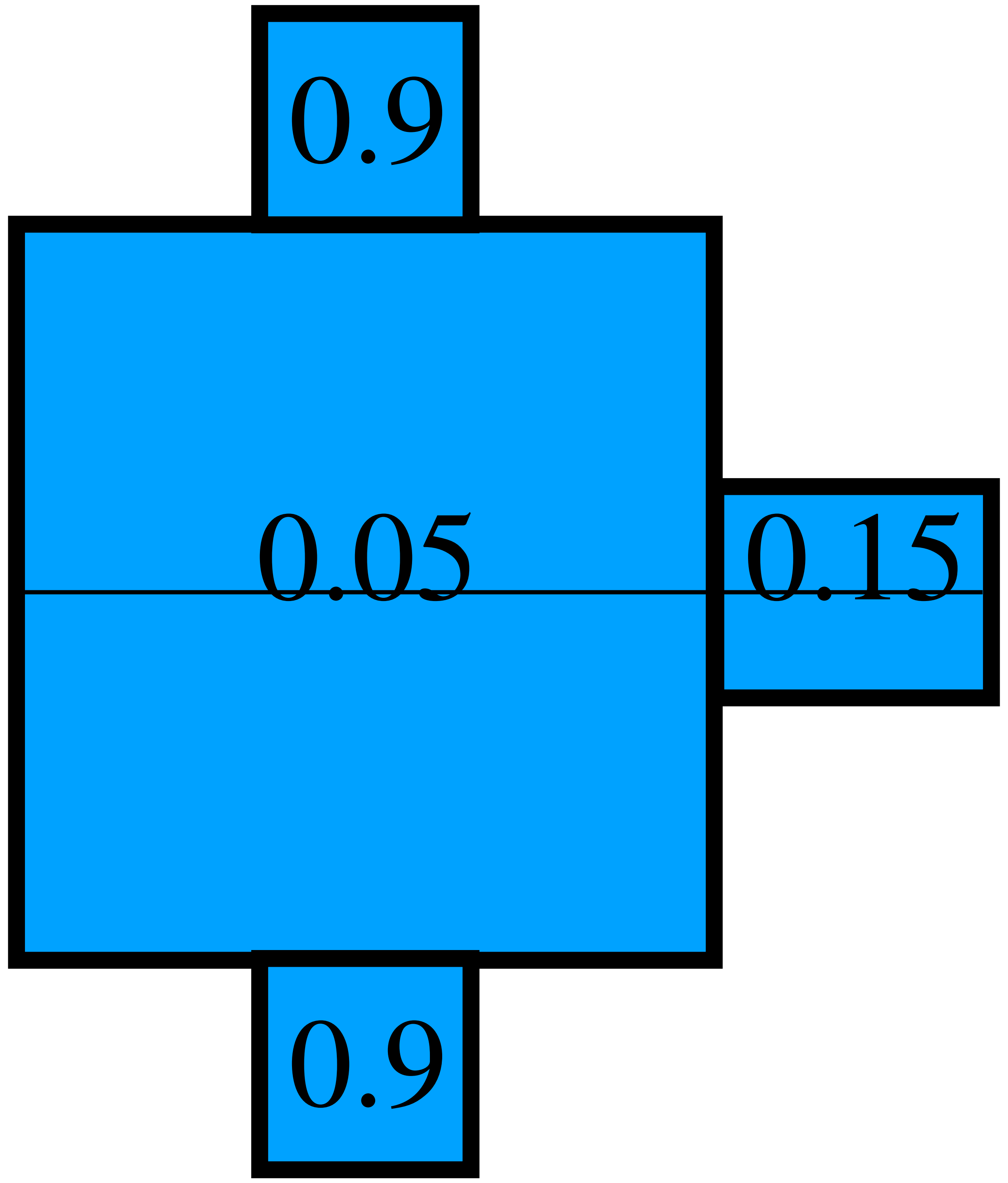
Carter, GH, **Hunter, Mattingly**
https://arxiv.org/abs/1908.11801; under review

What are the minimal number of county splits?

<u>Theorem</u>: d districts must introduce at least d-1 county splits (when we can't evenly partition counties)

<u>County splits</u>: the county splits of a county are the number of districts a county intersects minus 1

<u>Conjecture</u>: In nearly all redistricting problems, the bound from the theorem will be tight



Two districts must split the central county and one of the satellites

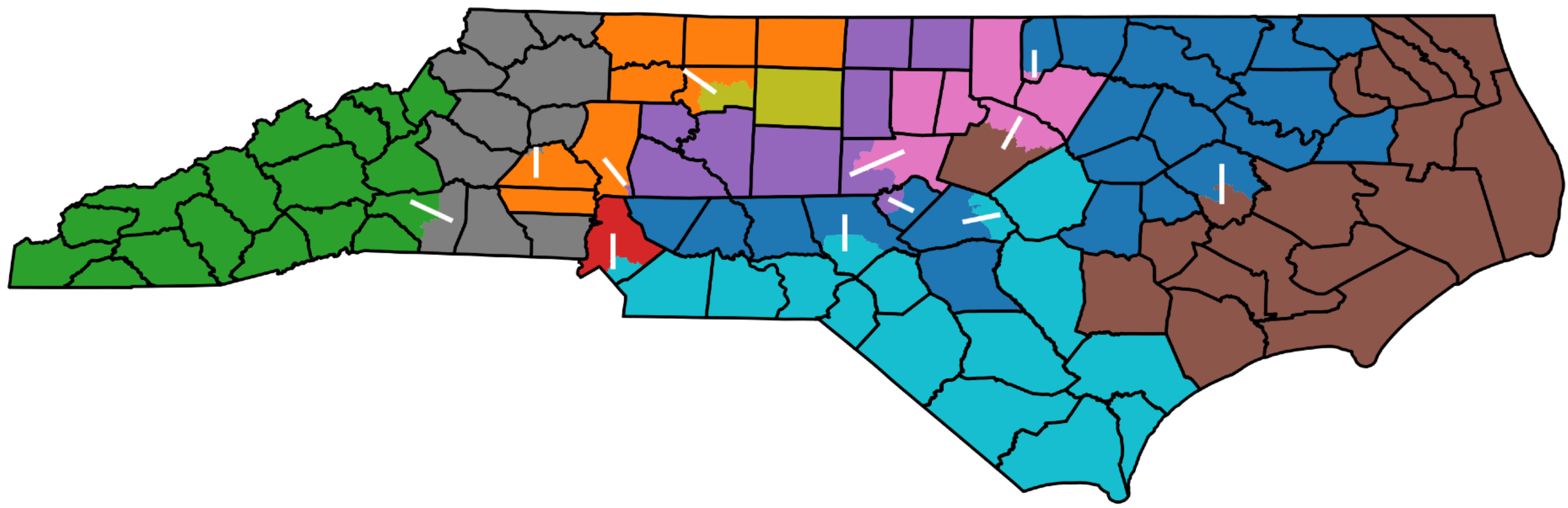
Carter, GH, Hunter, Mattingly
https://arxiv.org/abs/1908.11801; under review

What are the minimal number of county splits?

<u>Theorem</u>: d districts must introduce at least d-1 county splits (when we can't evenly partition counties)

<u>County splits</u>: the county splits of a county are the number of districts a county intersects minus 1

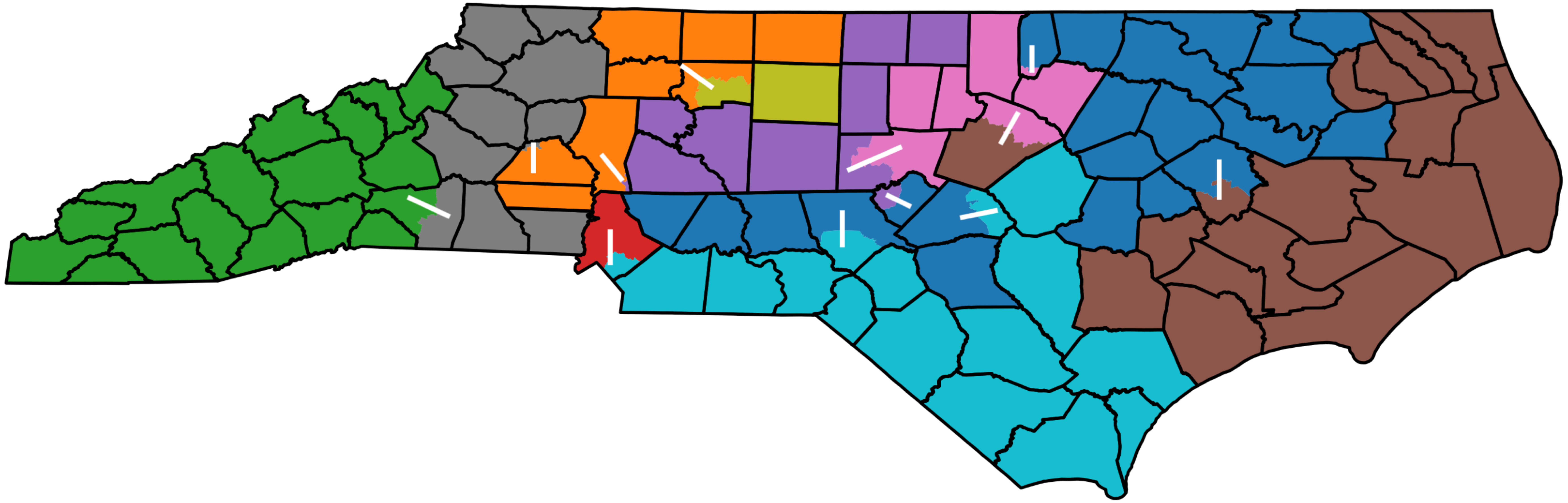
<u>Conjecture</u>: In nearly all redistricting problems, the bound from the theorem will be tight

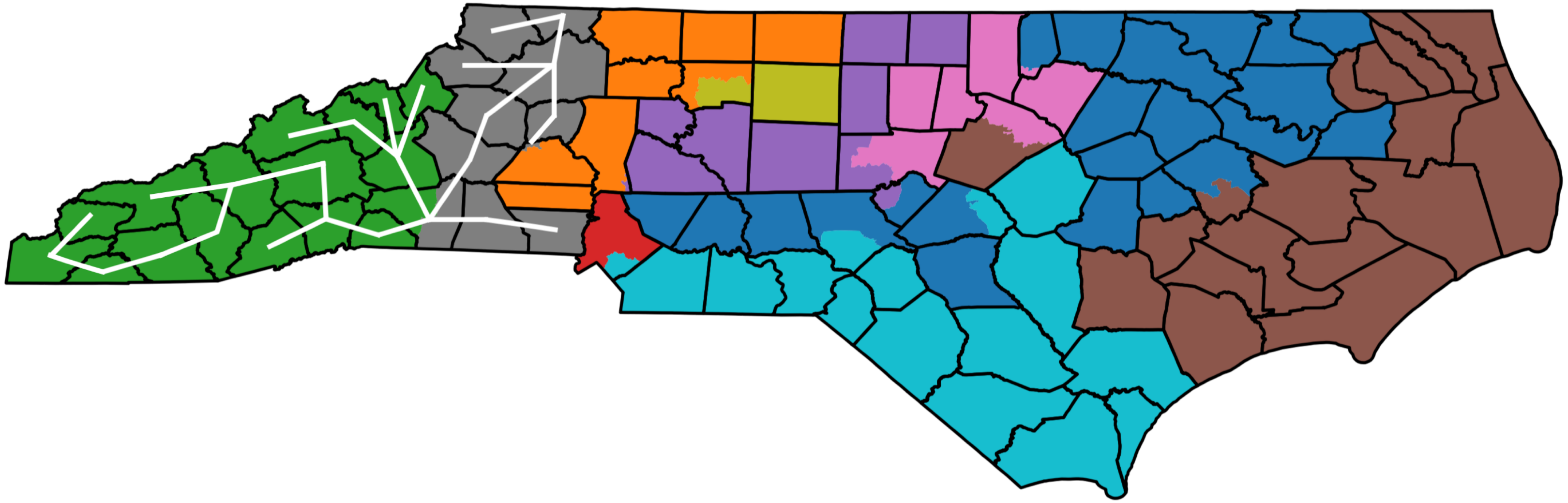


Enacted Plan (2019)

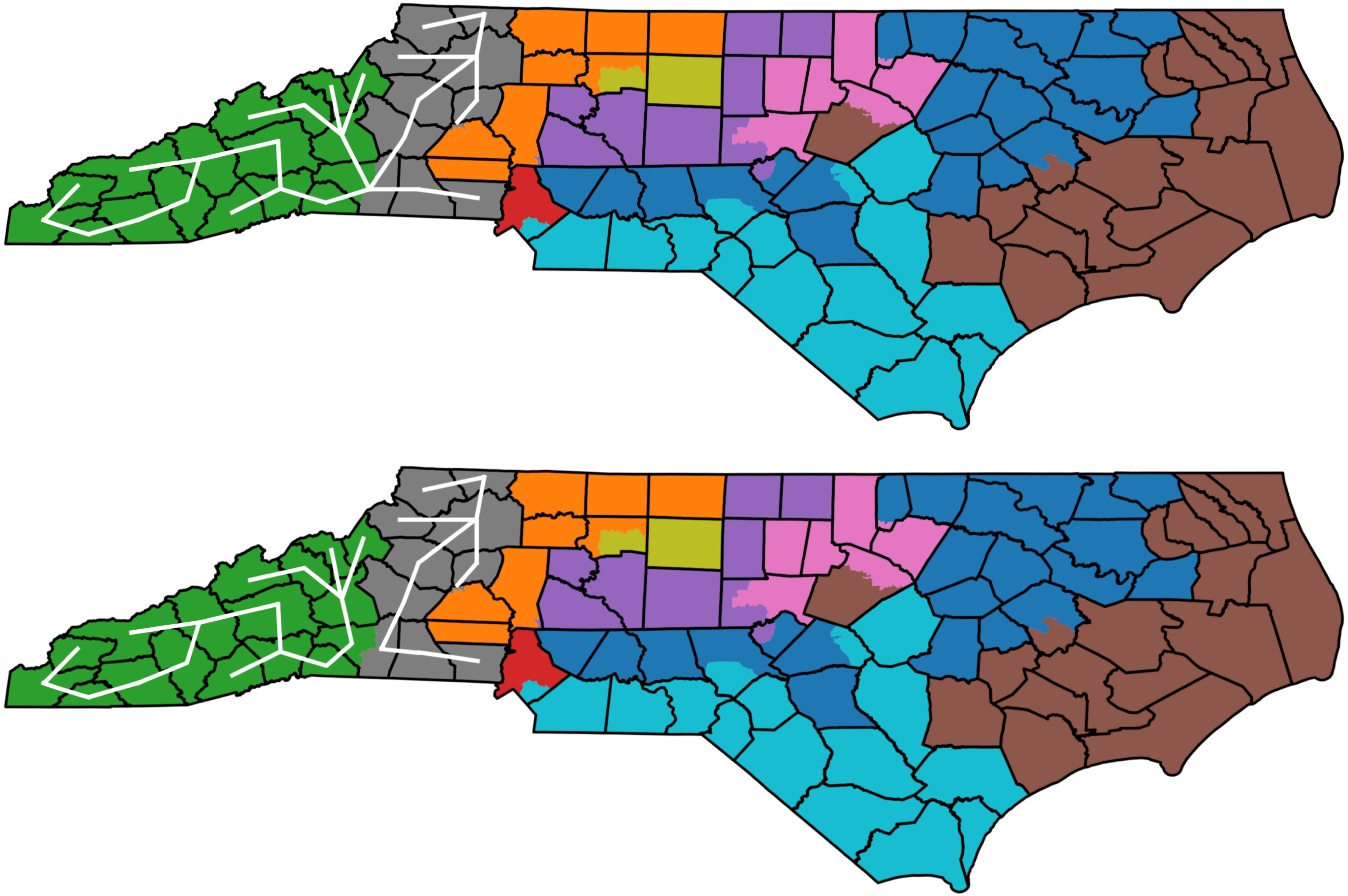
Carter, GH, Hunter, Mattingly
https://arxiv.org/abs/1908.11801; under review

Draw a merged tree on counties now

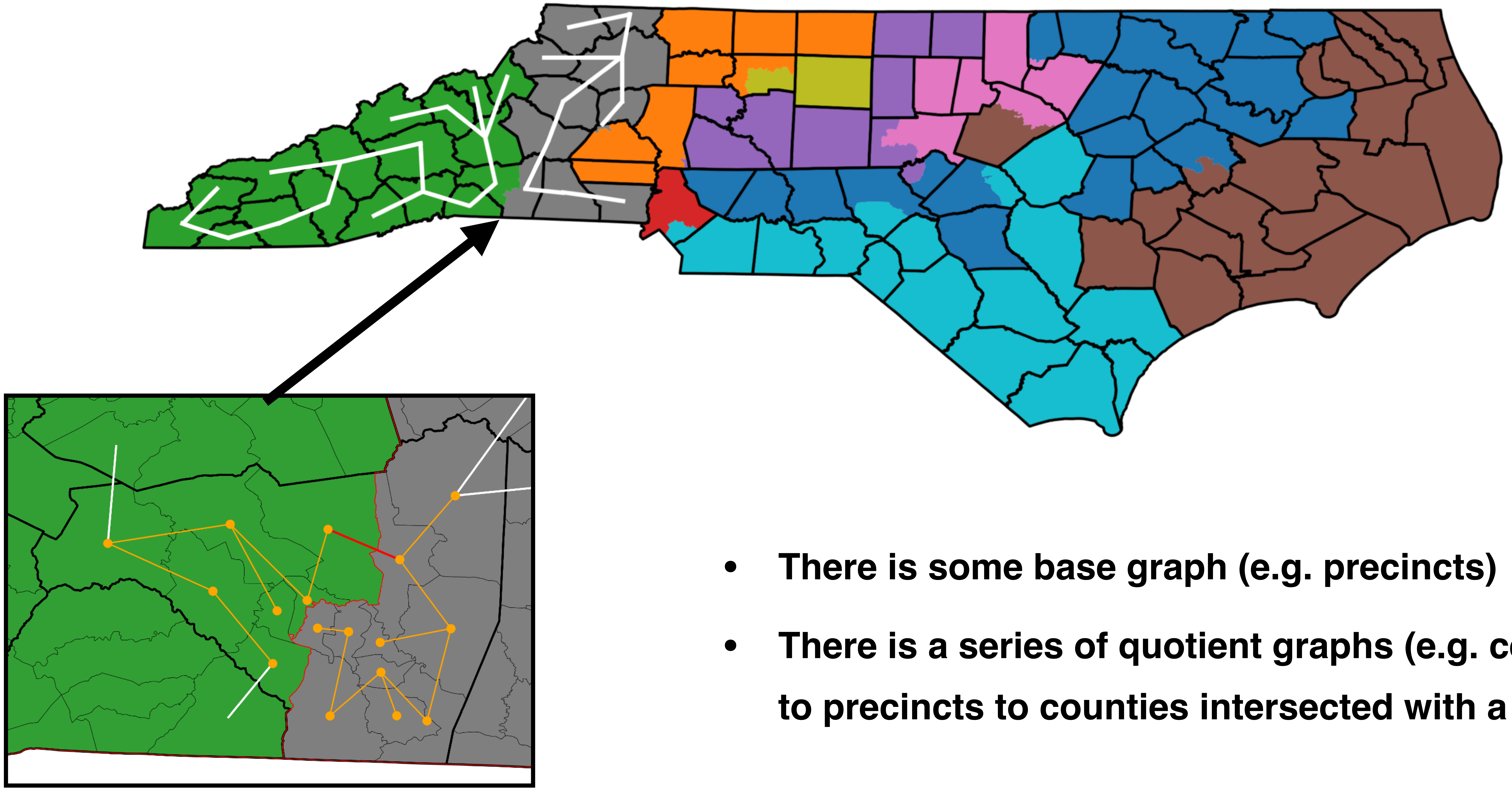




Splitting on Nodes rather than Edges

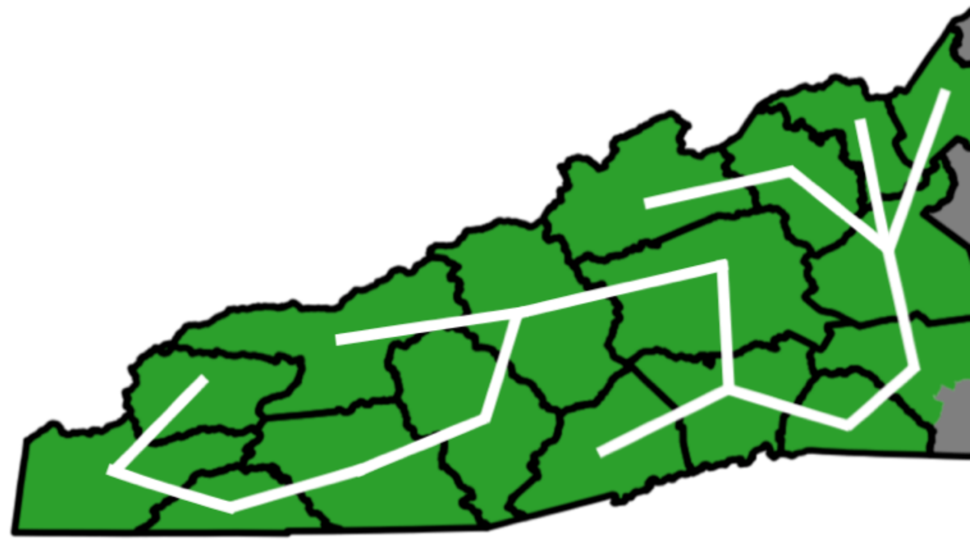


Nodes as Nested Trees

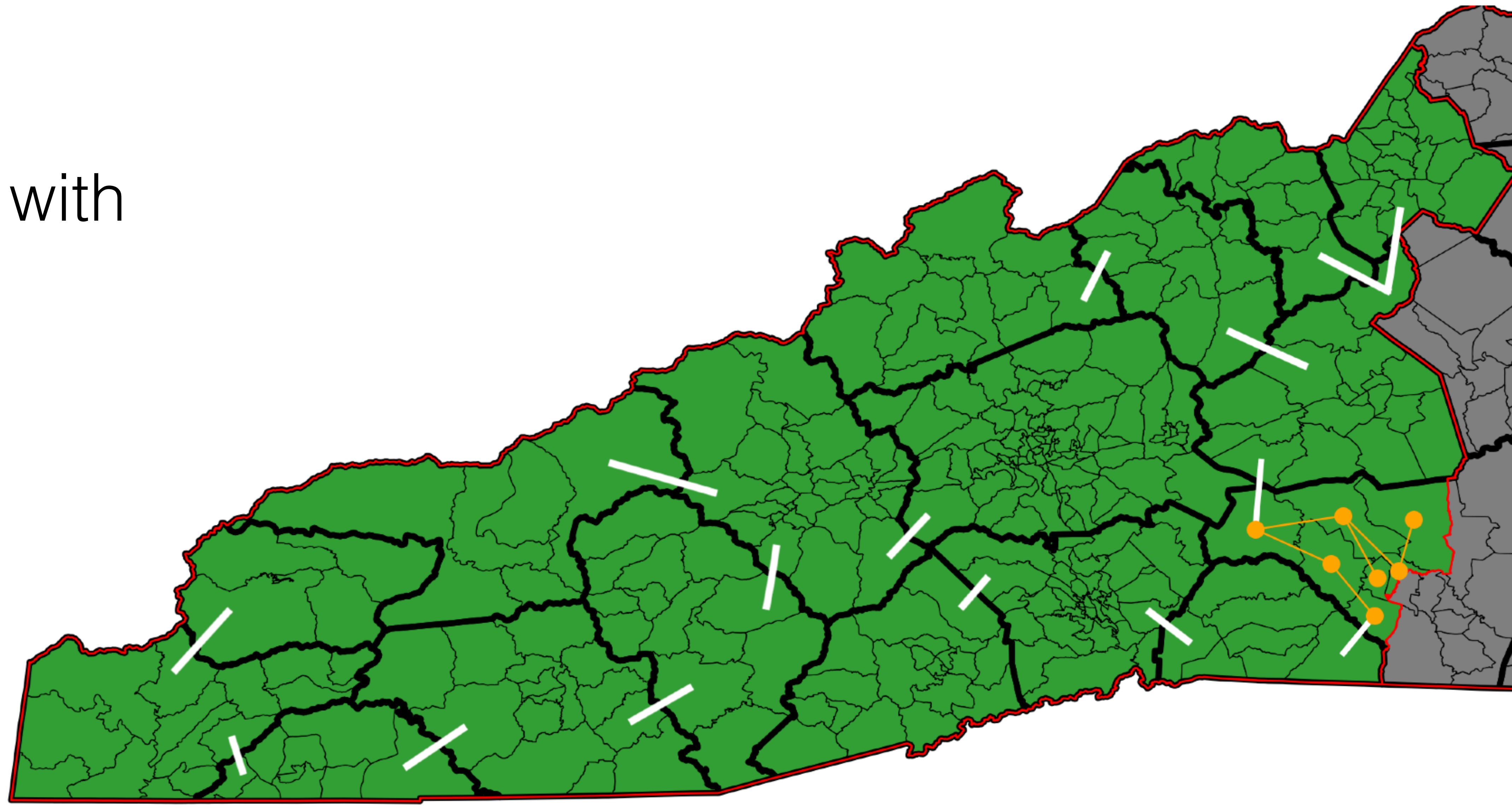


- **There is some base graph (e.g. precincts)**
- **There is a series of quotient graphs (e.g. census blocks to precincts to counties intersected with a district)**

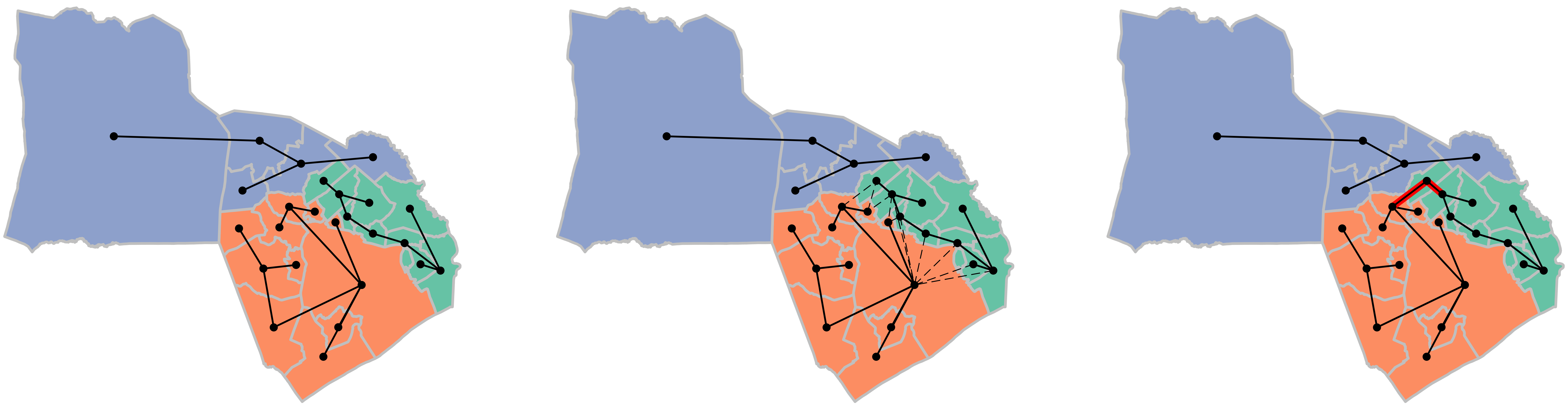
The State Space



- The state space is a forest at the finest level
- Each district is a tree on the finest scale such that
    - There is a simple tree graph at the county level when intersected with the district
    - Every nested node within the county/district graph has a tree with specified edges the top level tree

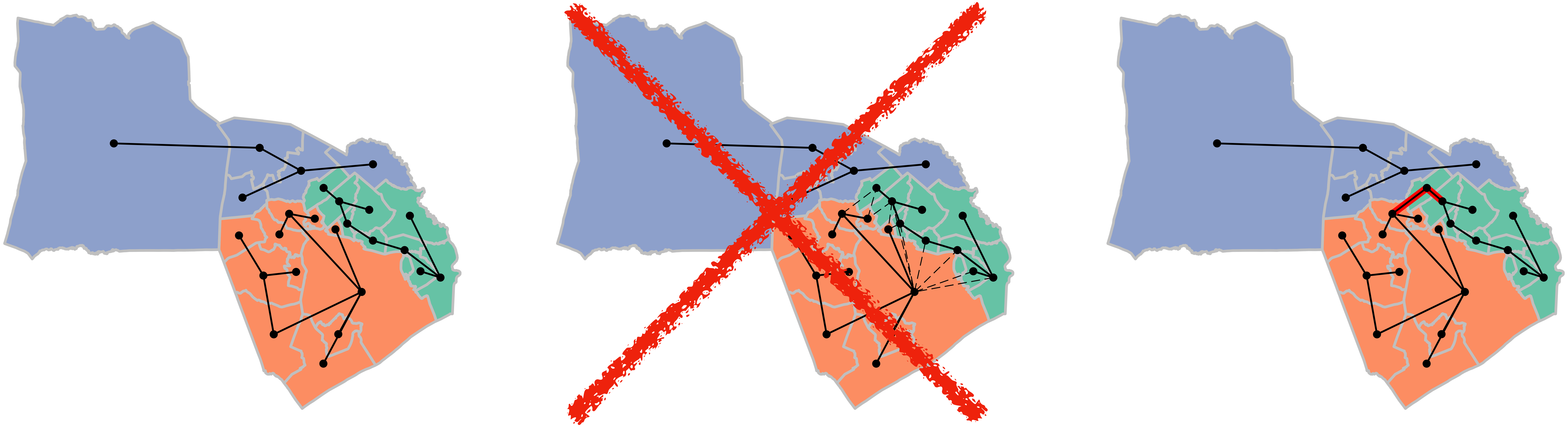


The State Space



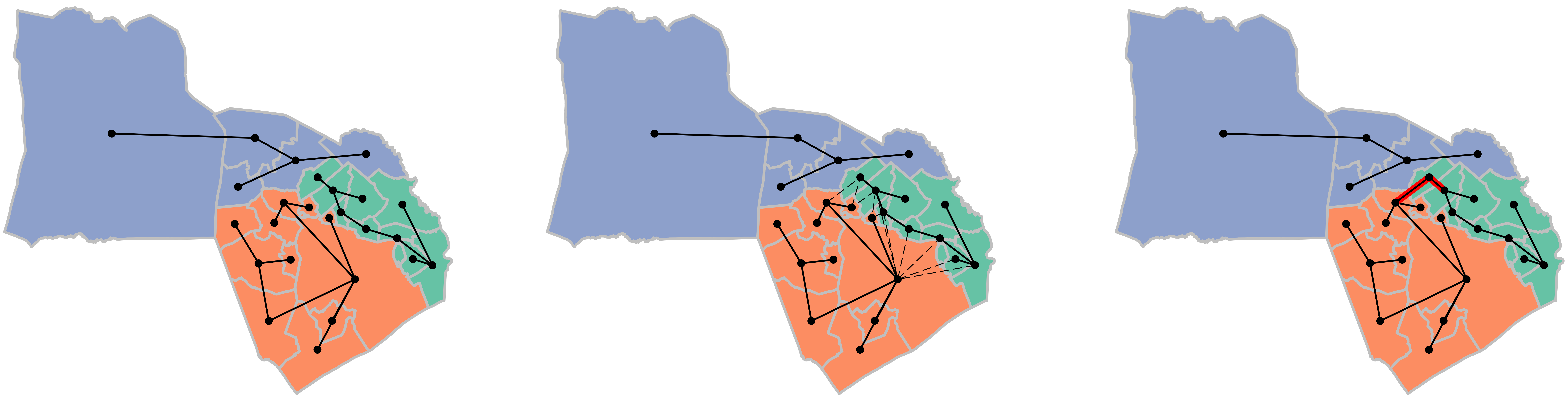
Reverse Probability in Merge-Split requires reconnecting every edge that could have drawn the merged tree

Adding Persistent Connections



Keep one of these edges in the state; only need to compute one cut set

Adding Persistent Connections



Merge-Split

$$\sum_{e \in A \cup B} P_{tree}(T_A, T_B, e) P_{cut}(e) = \sum_{e \in A \cup B} \frac{1}{\tau(A \cup B)} \frac{1}{|E_{cut}|(T_A, T_B, e)}$$

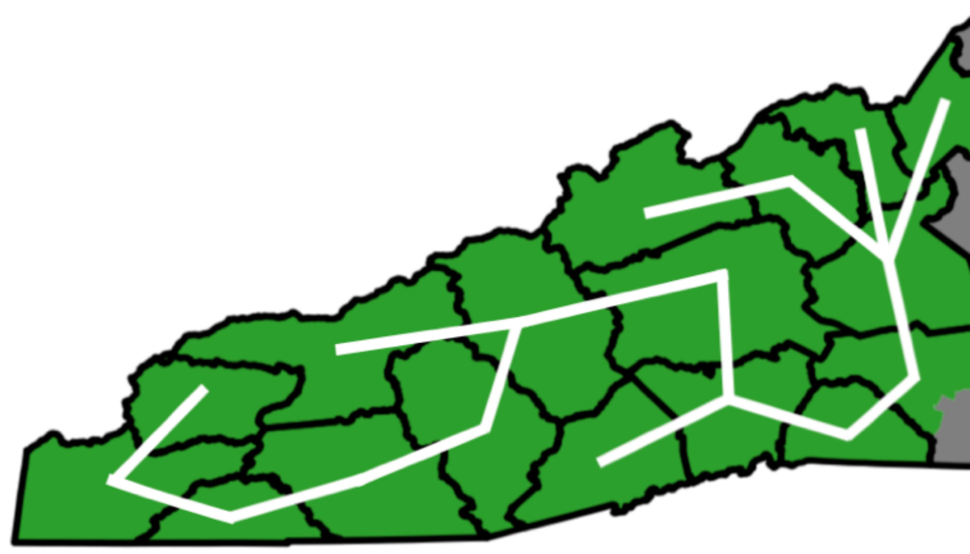
Reversible Recom

$$\sum_{T_A \in ST(A)} \sum_{T_B \in ST(B)} \sum_{e \in A \cup B} P_{tree}(T_A, T_B, e) P_{cut}(e) = \sum_{T_A \in ST(A)} \sum_{T_B \in ST(B)} \sum_{e \in A \cup B} \frac{1}{\tau(A \cup B)} \frac{1}{m}$$

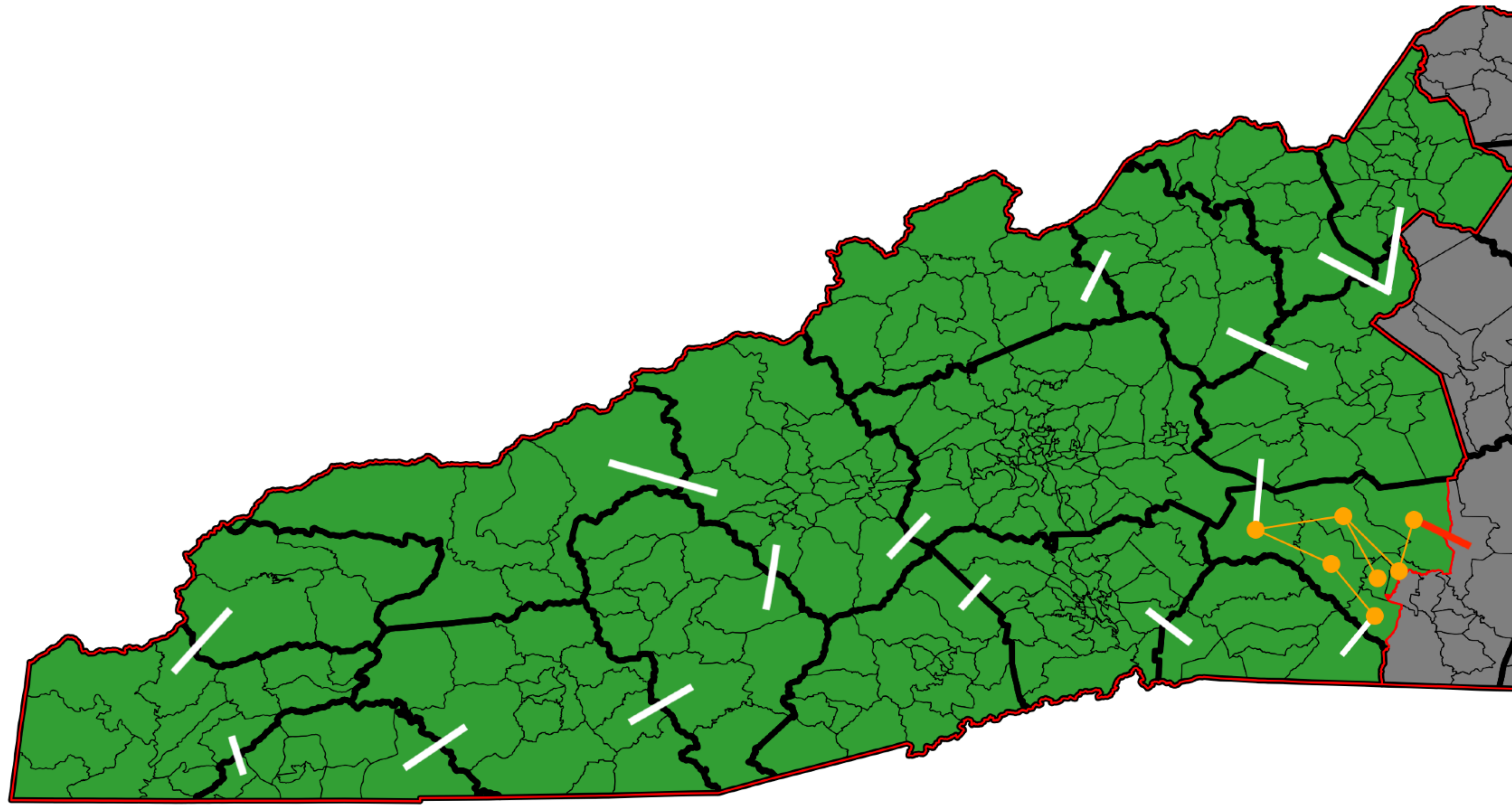
Persistent Edge Merge-Split

$$P_{tree}(T_A, T_B, e) P_{cut}(e) = \frac{1}{\tau(A \cup B)} \frac{1}{|E_{cut}|(T_A, T_B, e)}$$

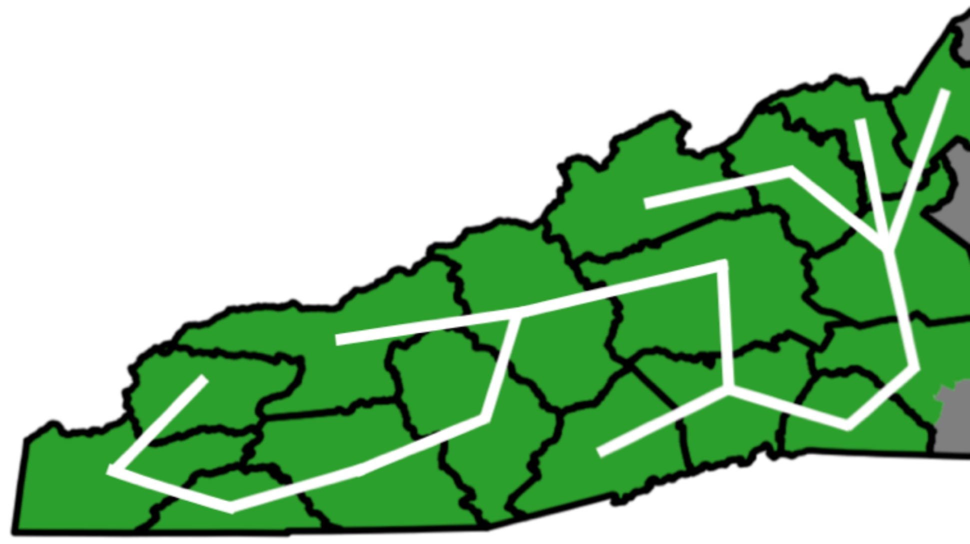
The State Space



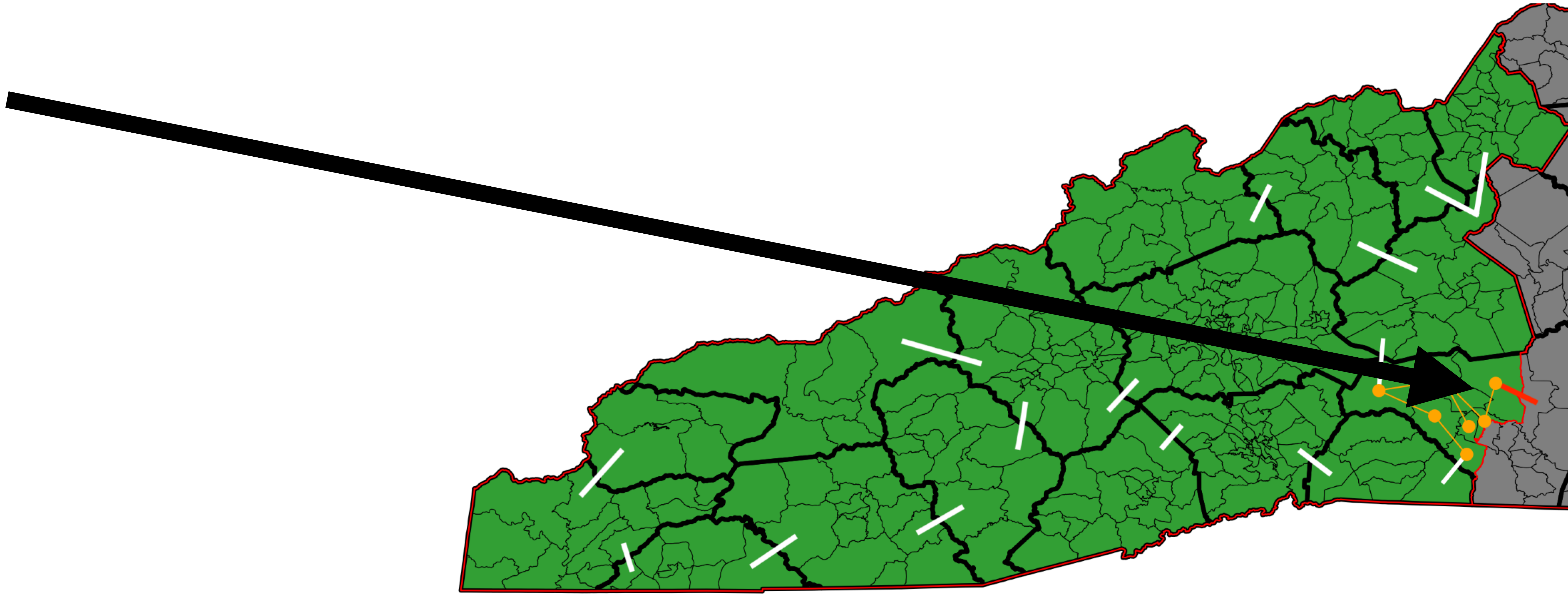
- The state space is a forest; each tree is a district
- Each district is a tree on the finest scale
- We keep edges between certain districts; call them <u>persistent edges</u>



The State Space

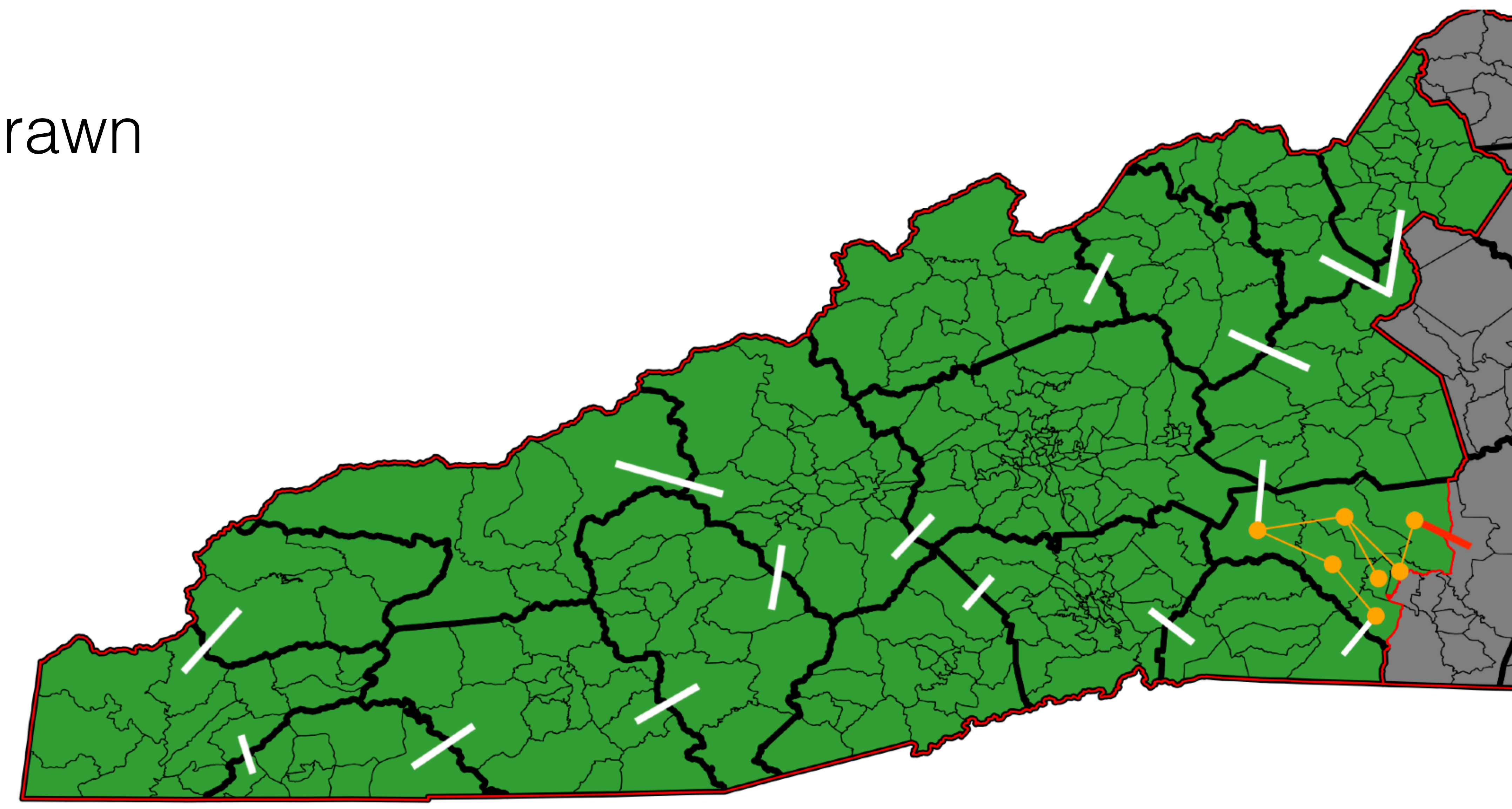


- The state space is a forest; each tree is a district
- Each district is a tree on the finest scale
- There are edges between certain districts



The number of states associated with a plan

- For each district:
    - A product of the number of trees that can be drawn on each level of the hierarchy
- For each persistent edge:
    - The number of other persistent edges that could have been drawn



The number of states associated with a plan

We could sample with a uniform measure over the extended state space:

$$\pi(\vec{T}, E_p) \propto 1$$

Or adapt by modding out by the number of similar plans with different persistent edges and trees

$$\pi(\vec{T}, E_p) \propto \left[\Big[\prod_{d \in D} \frac{1}{\tau(g_c(d))} \prod_{c \in C} \frac{1}{\tau(g_p(c,d))} \cdots \Big] \prod_{e \in E_p} \frac{1}{|\mathscr{P}(e, \vec{T})|} \right]^{\gamma}$$

$\vec{T}$**: the forest**

E_p**: persistent edges**

D**: district set**

C**: county set**

$g_C(d)$**: county graph restricted to district**

$g_p(c,d)$**: the precinct graph of a county and district**

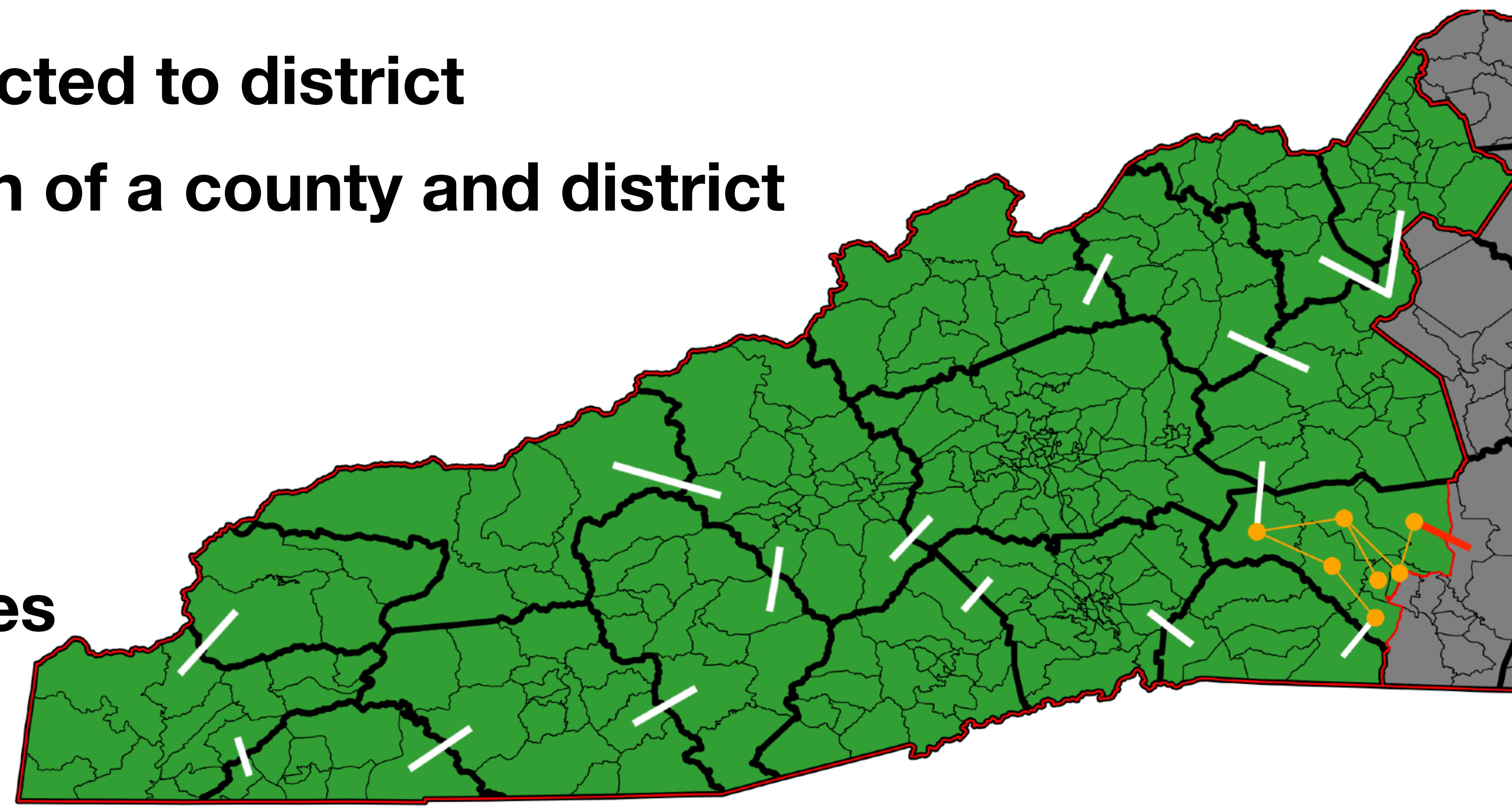
τ**: number of spanning trees**

$\mathscr{P}$**: set of possible persistent edges**

$\gamma \in [0,1]$

$\gamma = 0$**: Uniform over product space of trees and persistent edges**

$\gamma = 1$**: Uniform over partitions**

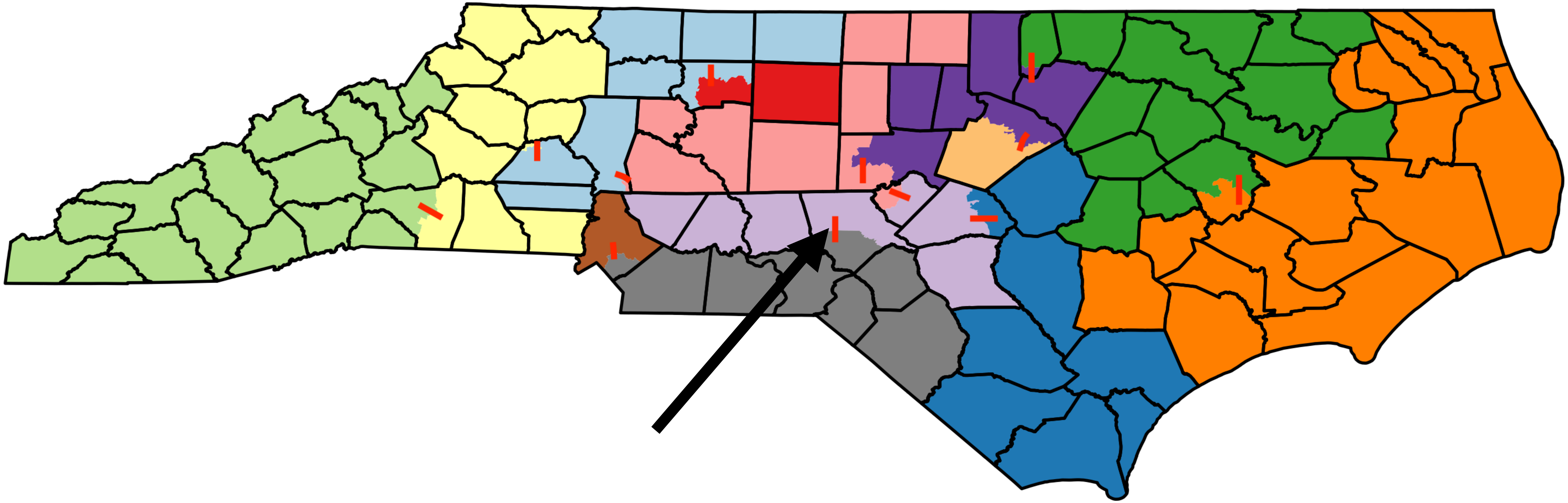


The proposal

1. Choose a persistent edge $\propto \frac{1}{|E_p|}$
2. Merge the district county graphs and draw a uniform tree on the resulting multigraph $\propto \frac{1}{\tau(g_c(d1 \cup d2))}$
3. Find edges and nodes that can be cut
4. On each cuttable node, draw a new uniform tree on the next level down (specify coarse edges where needed) and repeat steps 3 and 4 $\propto \frac{1}{\tau(g_p(c, d1 \cup d2))}$
5. Aggregate all edges that can be cut across all levels. Pick one uniformly; this is the new persistent edge. $\propto \frac{1}{|E_{cut}|}$

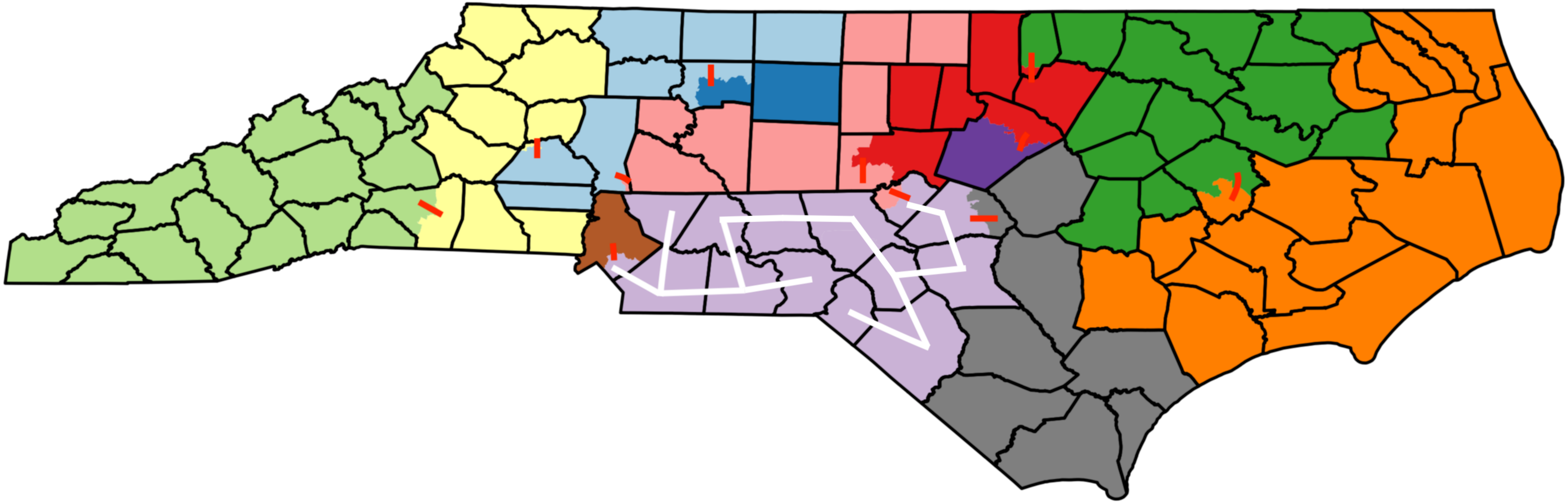
1. Choose a Persistent Edge

1. Choose a persistent edge $\propto \frac{1}{|E_p|}$



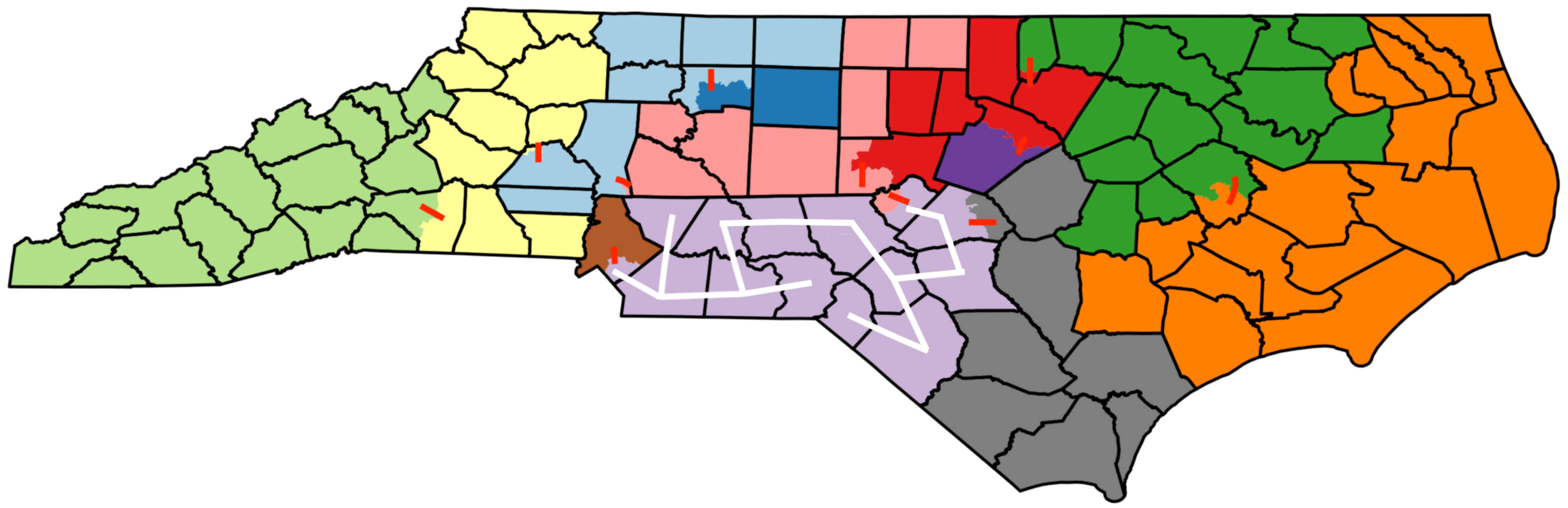
2. Merge districts; draw tree at the County level

2. Merge the district county graphs and draw a uniform tree on the resulting multigraph $\propto \frac{1}{\tau(g_c(d1 \cup d2))}$



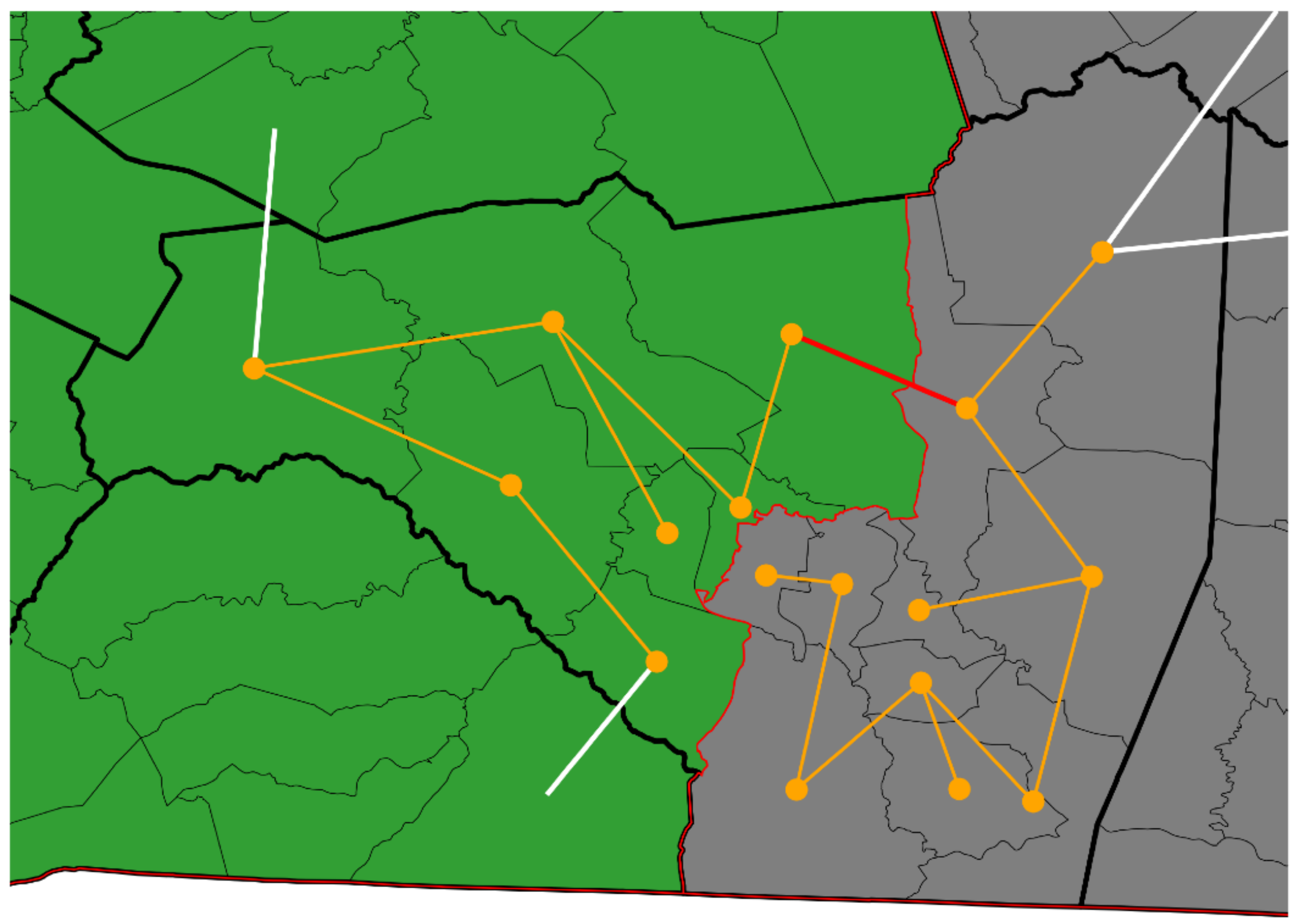
The proposal

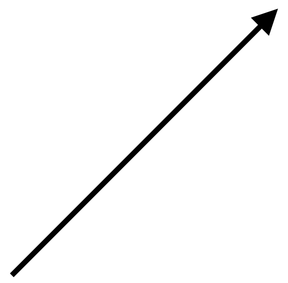
1. Choose a persistent edge $\propto \frac{1}{|E_p|}$
2. Merge the district county graphs and draw a uniform tree on the resulting multigraph $\propto \frac{1}{\tau(g_c(d1 \cup d2))}$
3. Find edges and nodes that can be cut
4. On each cuttable node, draw a new uniform tree on the next level down (specify coarse edges where needed) and repeat steps 3 and 4 $\propto \frac{1}{\tau(g_p(c, d1 \cup d2))}$
5. Aggregate all edges that can be cut across all levels. Pick one uniformly; this is the new persistent edge. $\propto \frac{1}{|E_{cut}|}$



The proposal

1. Choose a persistent edge $\propto \frac{1}{|E_p|}$
2. Merge the district county graphs and draw a uniform tree on the resulting multigraph $\propto \frac{1}{\tau(g_c(d1 \cup d2))}$
3. Find edges and nodes that can be cut
4. On each cuttable node, draw a new uniform tree on the next level down (specify coarse edges where needed) and repeat steps 3 and 4 $\propto \frac{1}{\tau(g_p(c, d1 \cup d2))}$
5. Aggregate all edges that can be cut across all levels. Pick one uniformly; this is the new persistent edge. $\propto \frac{1}{|E_{cut}|}$





The proposal (computational acceleration)

- Even though we have only explicitly drawn trees on the cuttable, nodes, we have implicitly drawn them everywhere.
- The persistent edge gives a tree on the merged graph. We no longer need to iterate over all other trees!

$$\frac{Q(T_1, T_2, e)}{Q(T'_1, T'_2, e')} = \frac{|E'_{cut}|(T'_1, T'_2, e')}{|E_{cut}|(T_1, T_2, e)}$$

With persistent edges

$$\frac{Q(T_1, T_2, e)}{Q(T'_1, T'_2, e')} = \frac{\sum_{e \in \mathscr{P}(T_1, T_2)} \frac{1}{|E_{cut}|(T_1, T_2, e)}}{\sum_{e \in \mathscr{P}(T'_1, T'_2)} \frac{1}{|E_{cut}|(T'_1, T'_2, e)}}$$

Without persistent edges

Probability Ratio

$$\pi(\overrightarrow{T}, E_p) \propto \left[\left[\prod_{d \in D} \frac{1}{\tau(g_c(d))} \prod_{c \in C} \frac{1}{\tau(g_p(c, d))} \dots \right] \prod_{e \in E_p} \frac{1}{|\mathscr{P}(e, \overrightarrow{T})|} \right]^{\gamma}$$

$$\frac{Q(T_1, T_2, e)}{Q(T_1', T_2', e')} = \frac{|E_{cut}'|(T_1', T_2', e')}{|E_{cut}|(T_1, T_2, e)}$$

$$\frac{\pi(\overrightarrow{T}, E_p)}{\pi(\overrightarrow{T}', E_p')} = \frac{\tau(g_c(T_1'))\tau(g_c(T_2'))}{\tau(g_c(T_1))\tau(g_c(T_2))} \frac{\tau(g_p(c_{n'}, d_1'))\tau(g_p(c_{n'}, d_2'))}{\tau(g_p(c_{n'}))} \frac{\tau(g_p(c_n))}{\tau(g_p(c_n, d_1))\tau(g_p(c_n, d_2))} \frac{|\mathscr{P}(e')|}{|\mathscr{P}(e)|} \prod_{e_i \in E_p' \setminus e} \frac{|\mathscr{P}(e_i, \overrightarrow{T}')|}{|\mathscr{P}(e_i, \overrightarrow{T})|}$$

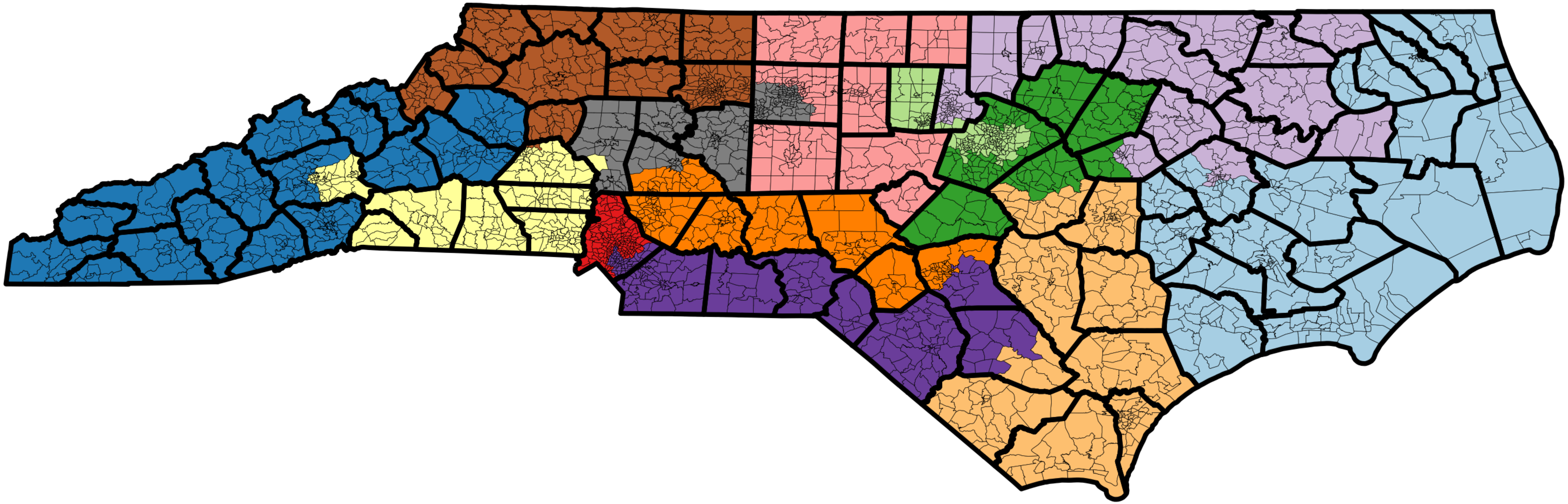
c_n', c_n **Cut counties on new and old districts, resp.**

e', e **Persistent edge for new and old districts, resp.**

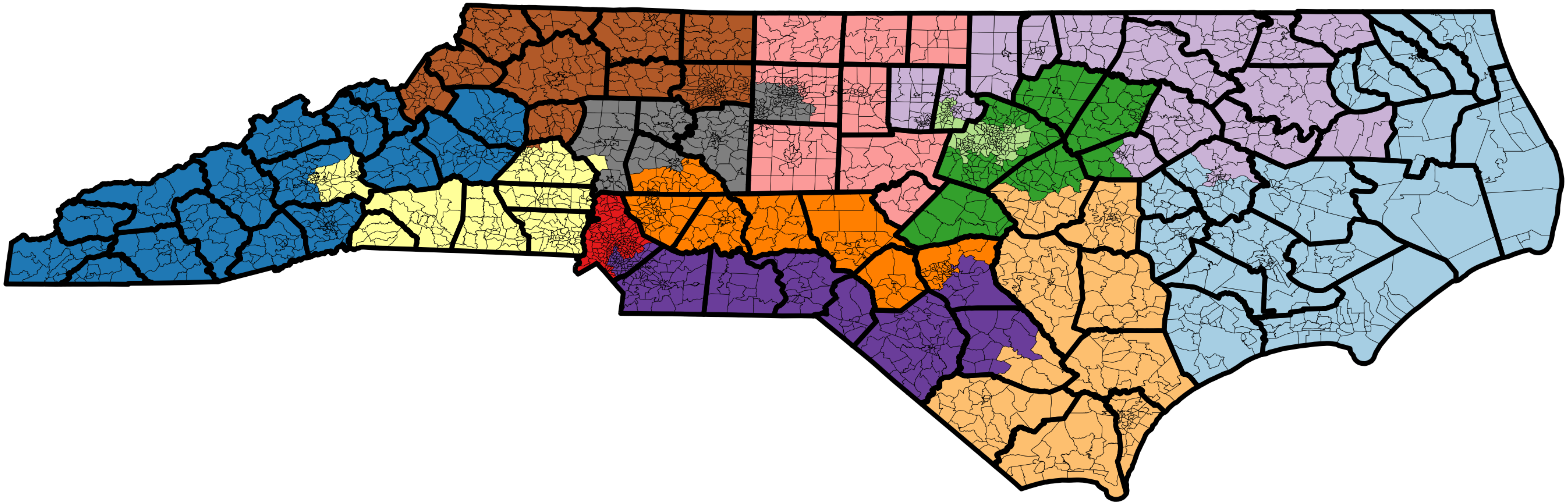
Remarks

- Must ensure there are not two persistent edges linking the same two districts.
- The number of cut counties is bounded, from above, by the number of persistent edges.
- In our implementation, we do not allow nodes to be cut into three districts.

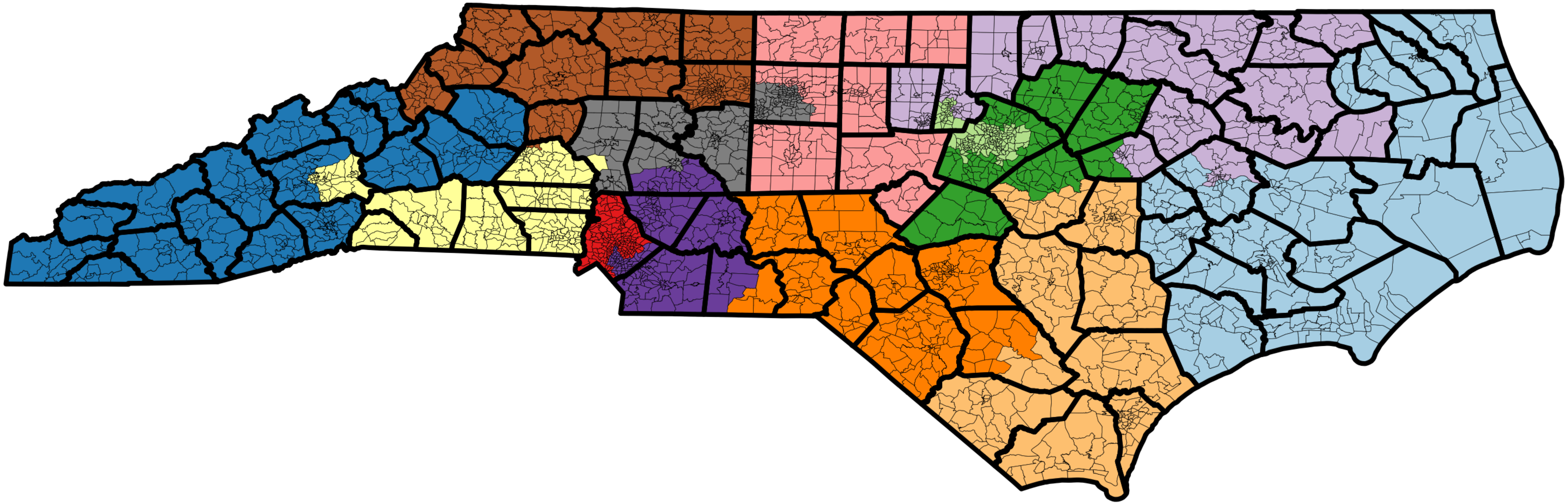
Results



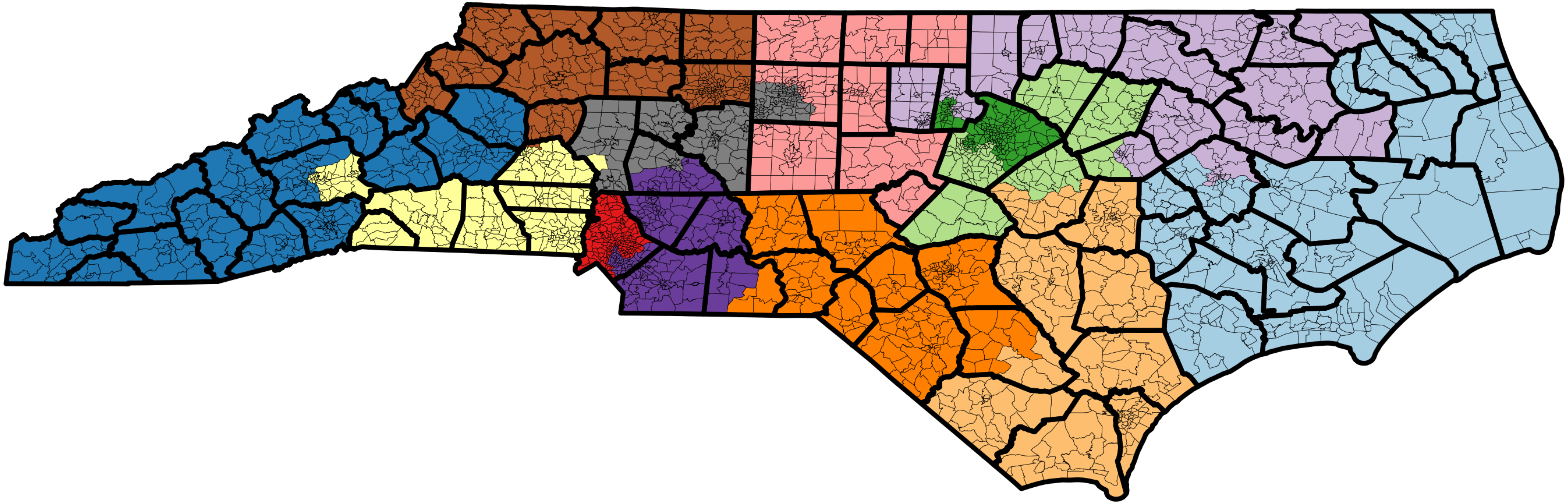
Results



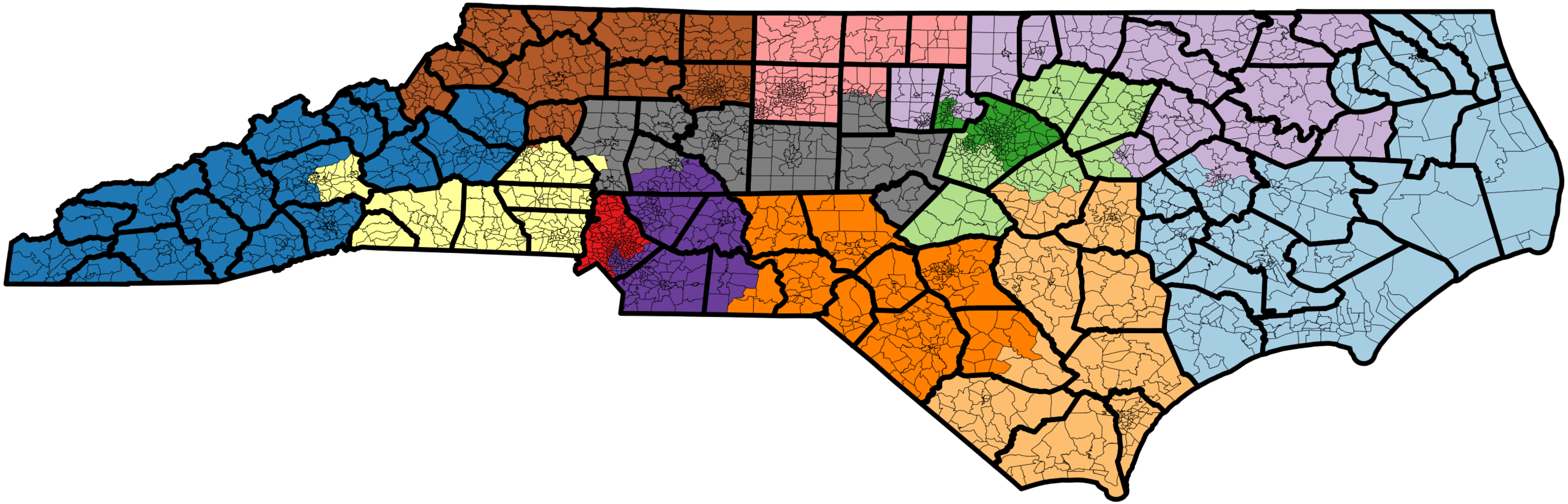
Results



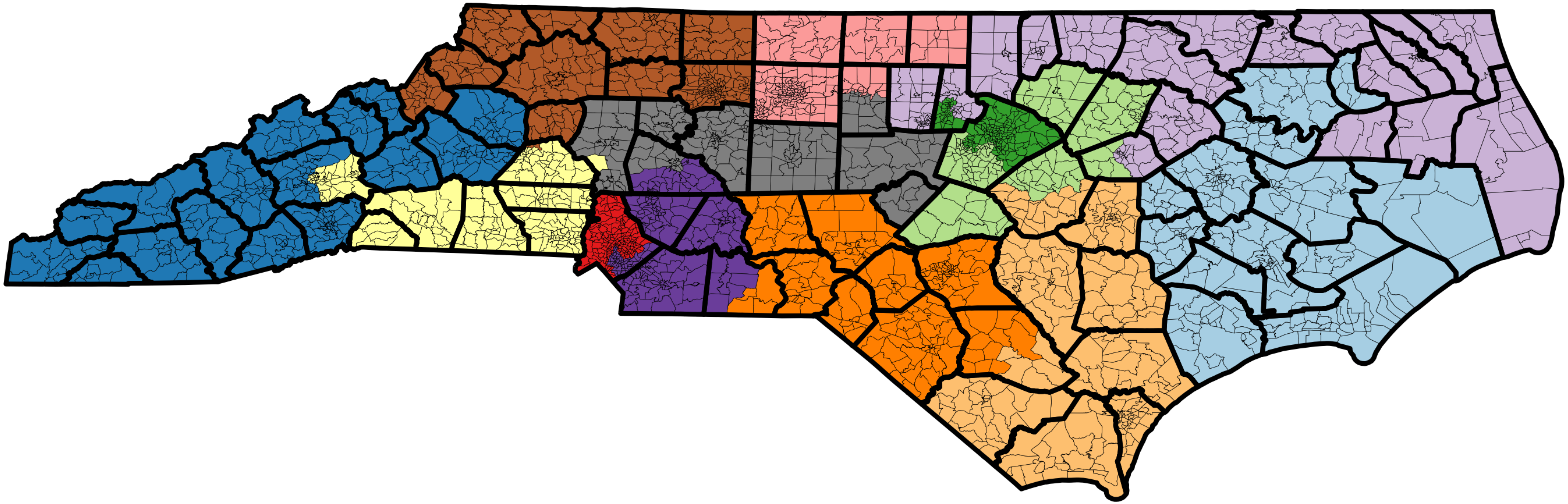
Results



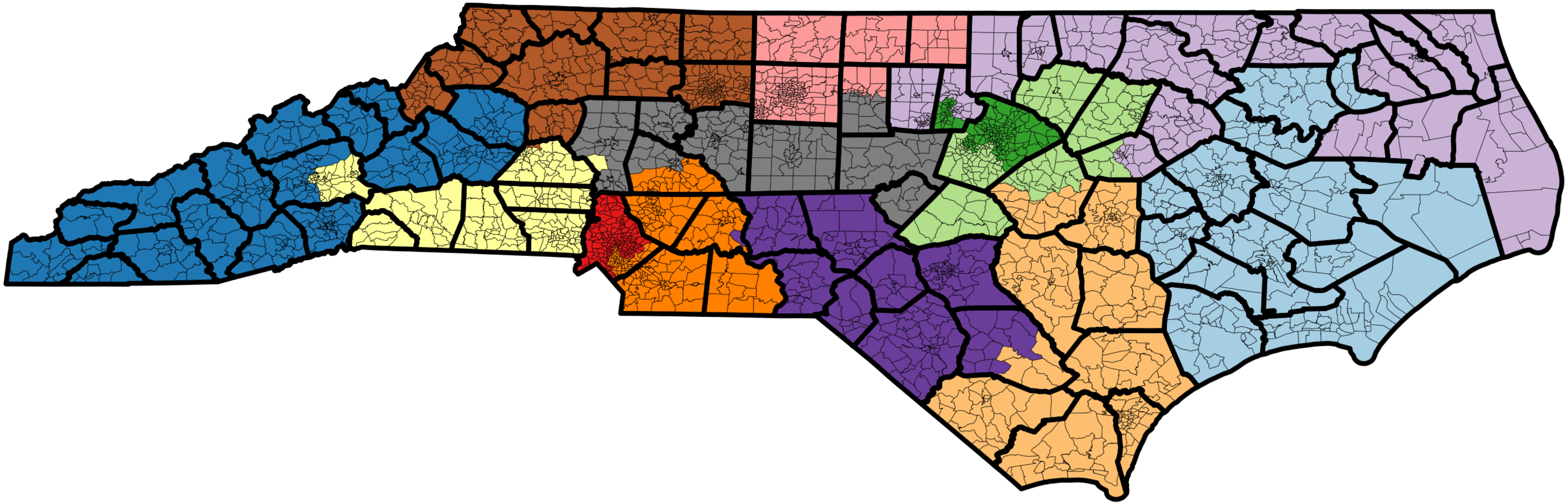
Results



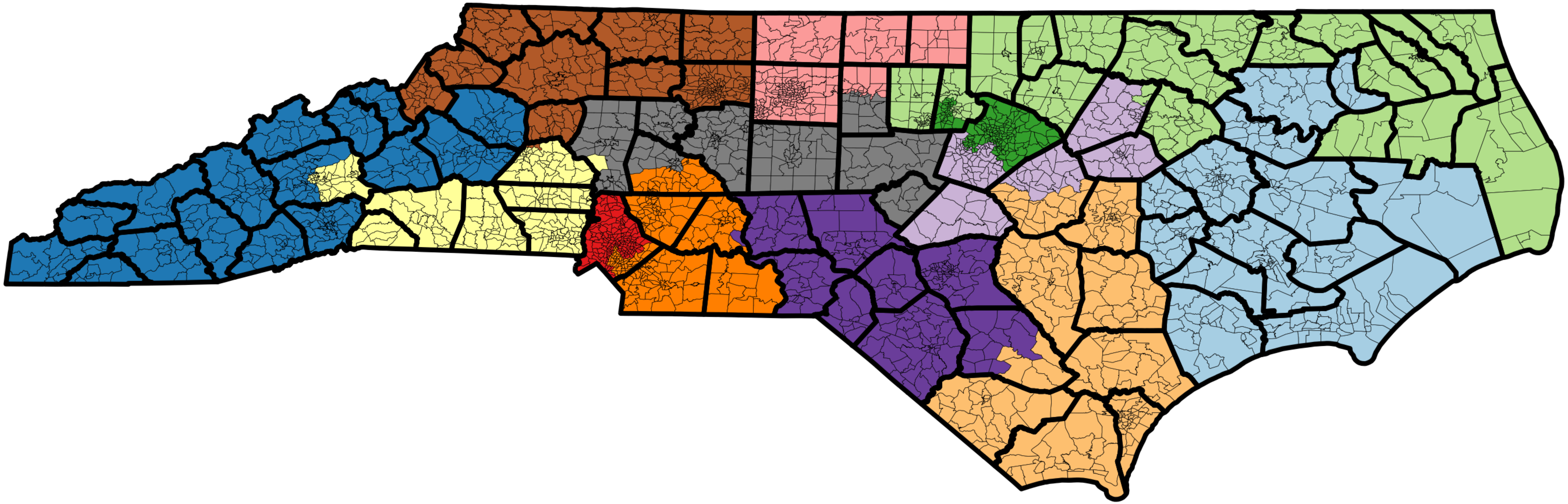
Results



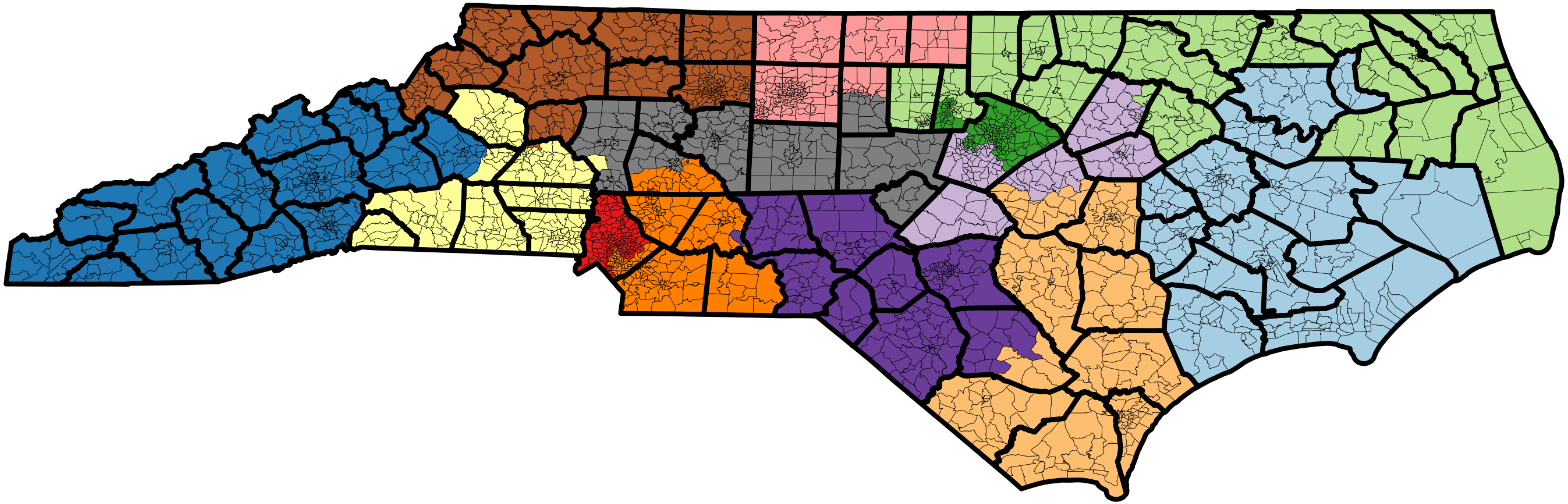
Results



Results

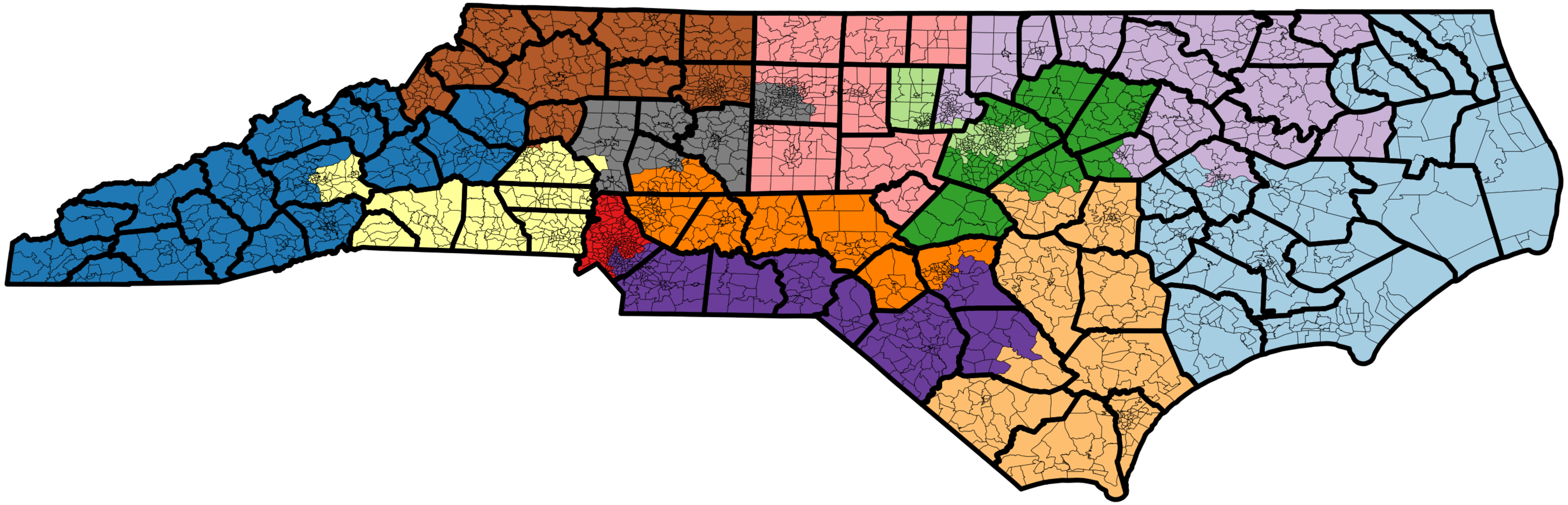


Results

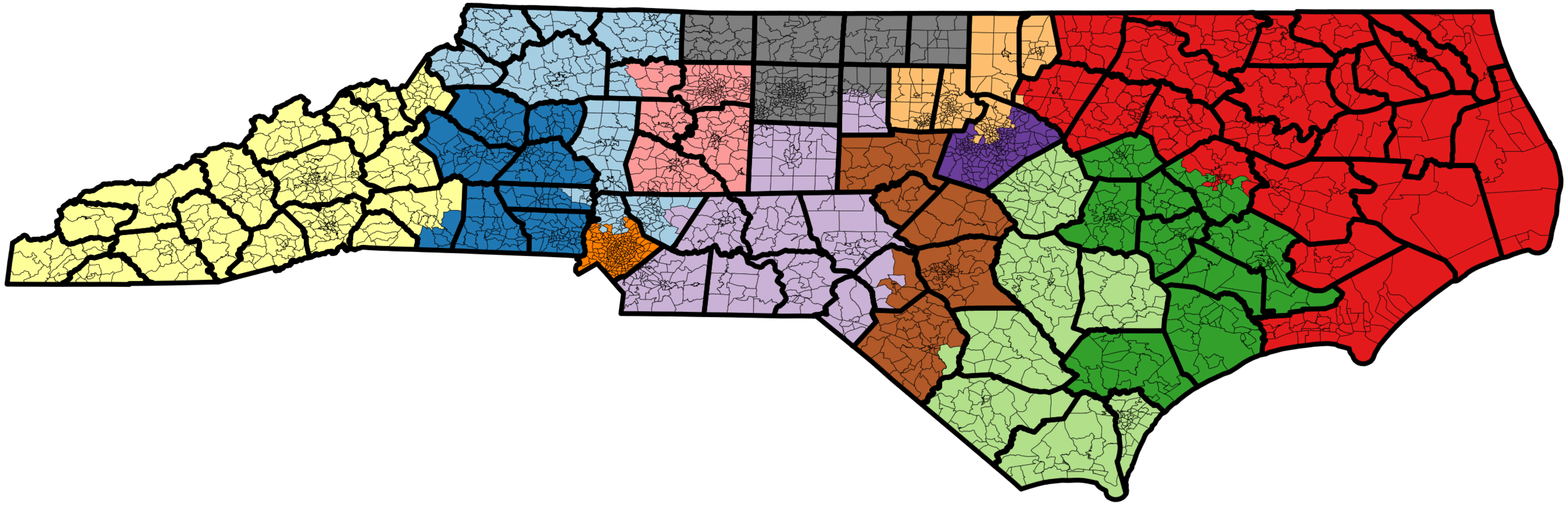


Results

Initial

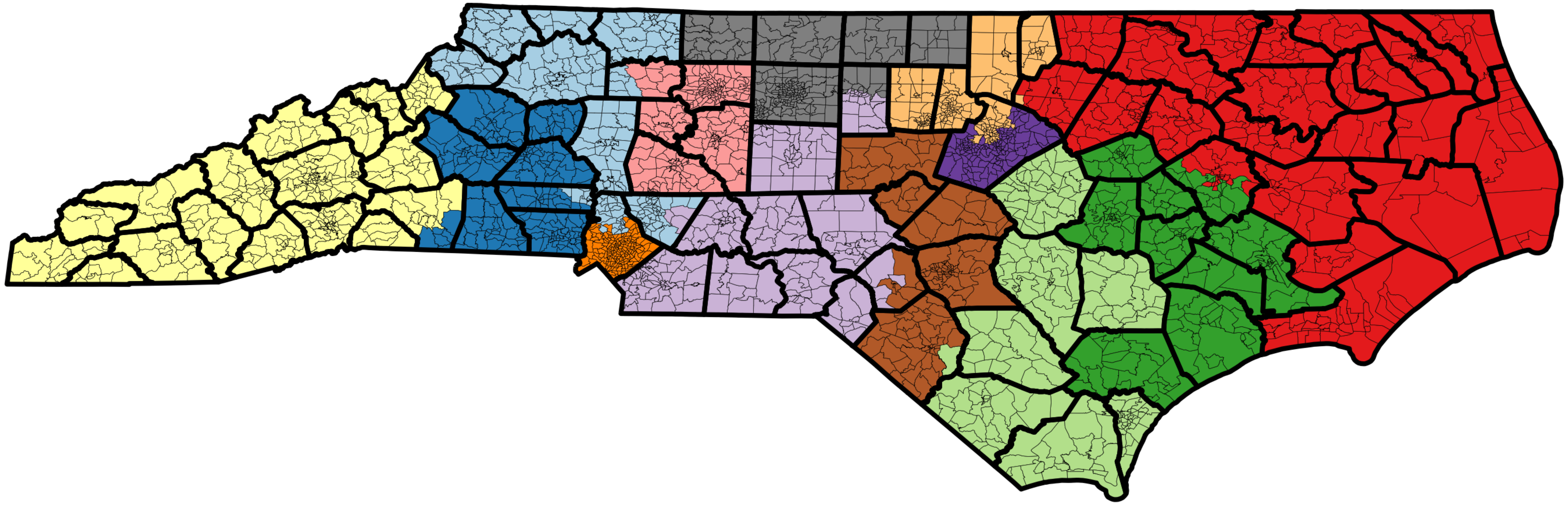


10K Proposals

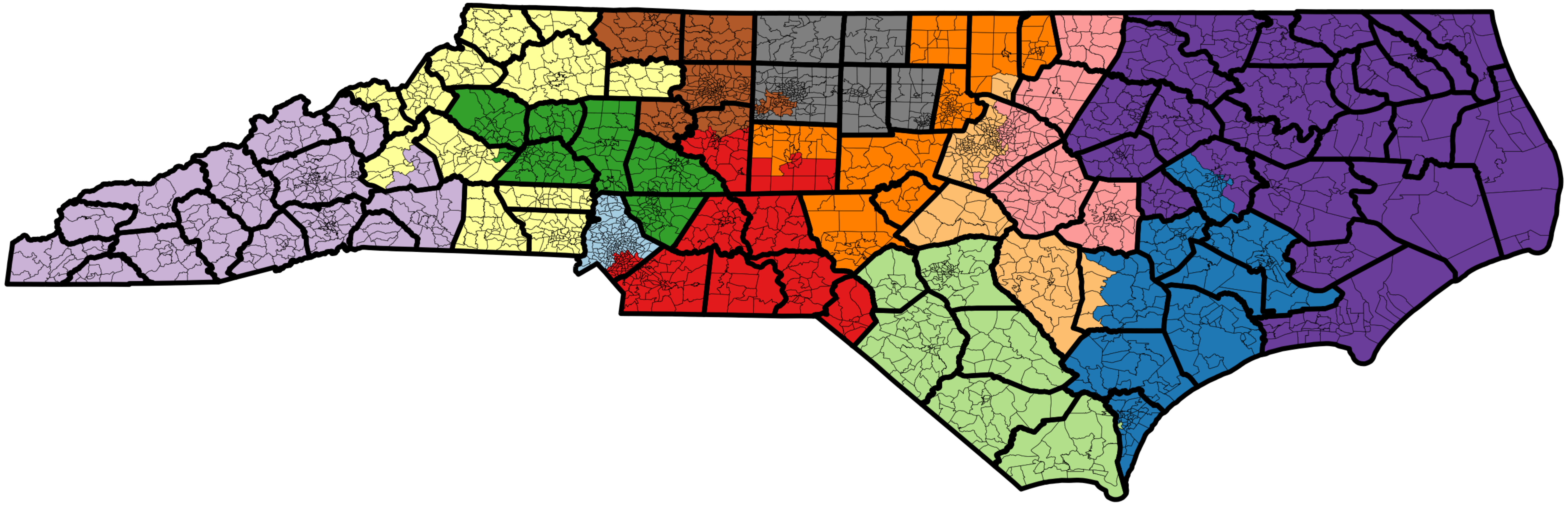


Results

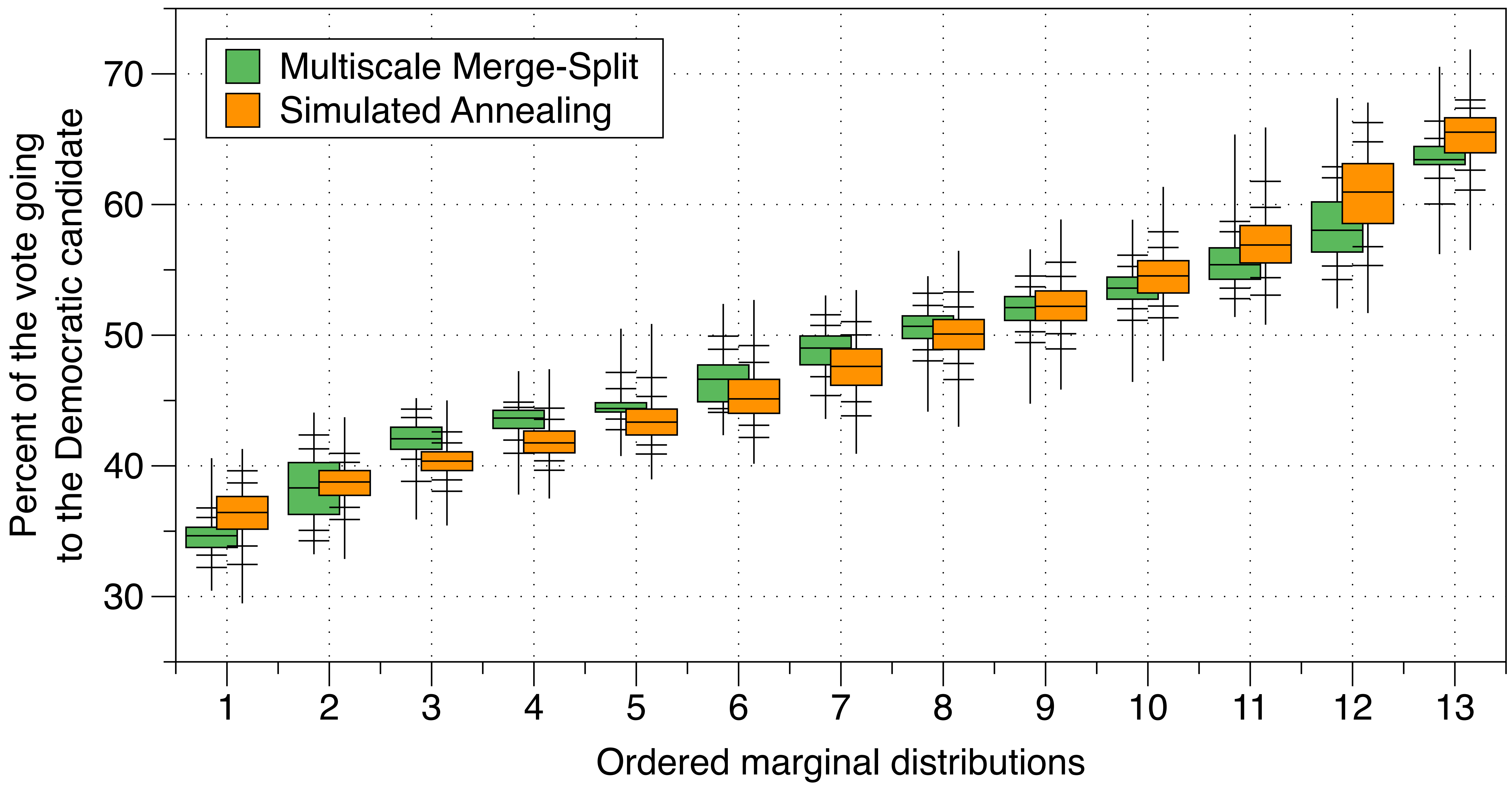
10K Proposals



15K Proposals



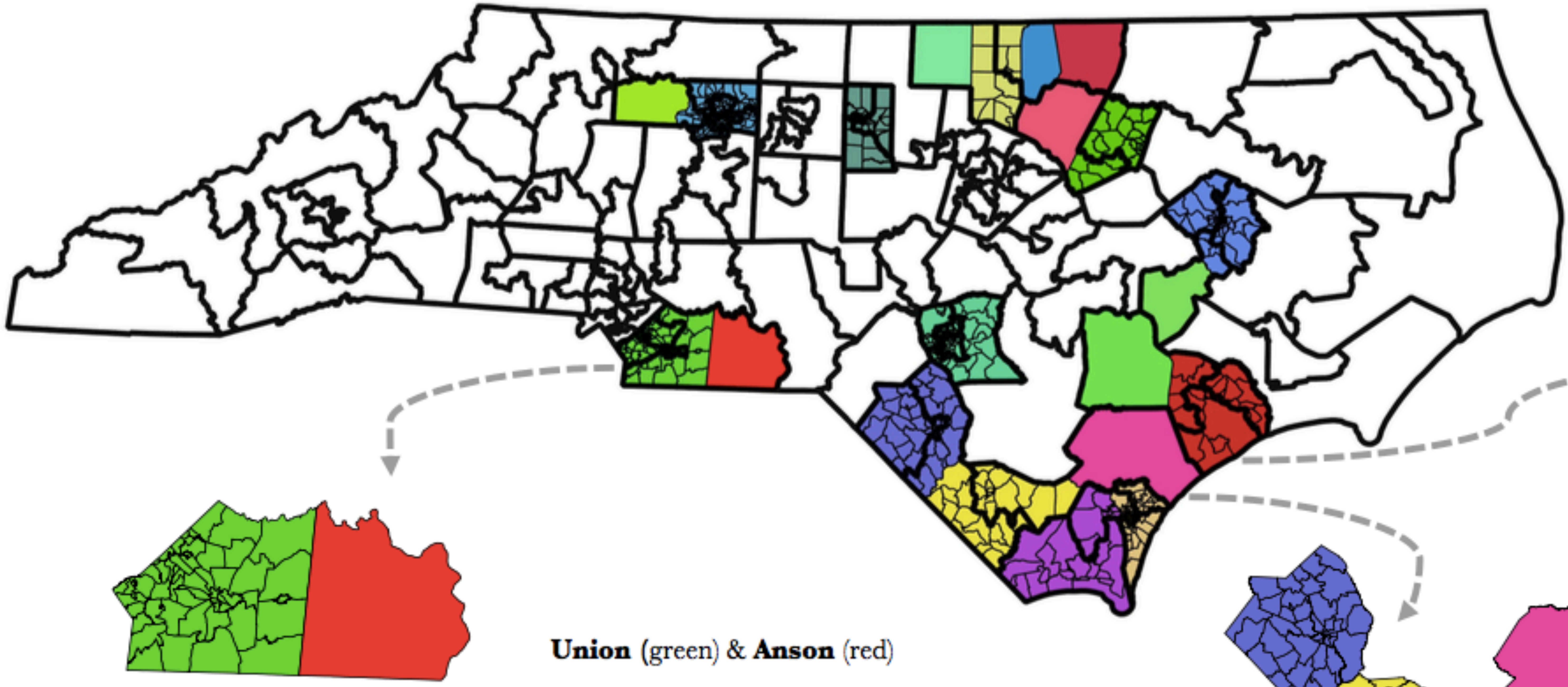
Results

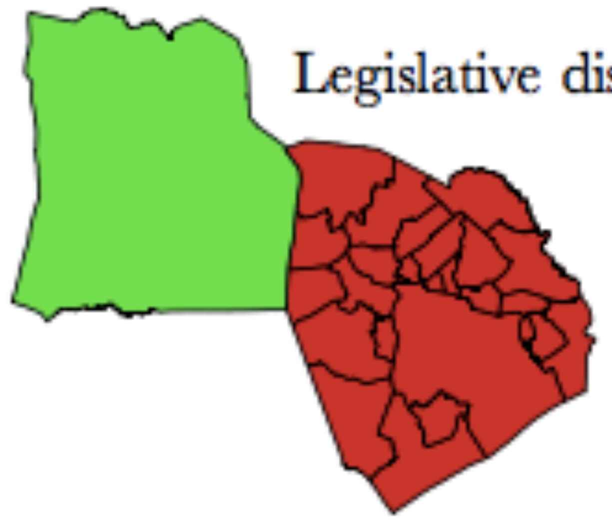


Multiscale Merge-Split
Simulated Annealing
Percent of the vote going to the Democratic candidate
30
40
50
60
70
1
2
3
4
5
6
7
8
9
10
11
12
13
Ordered marginal distributions

County Clusters

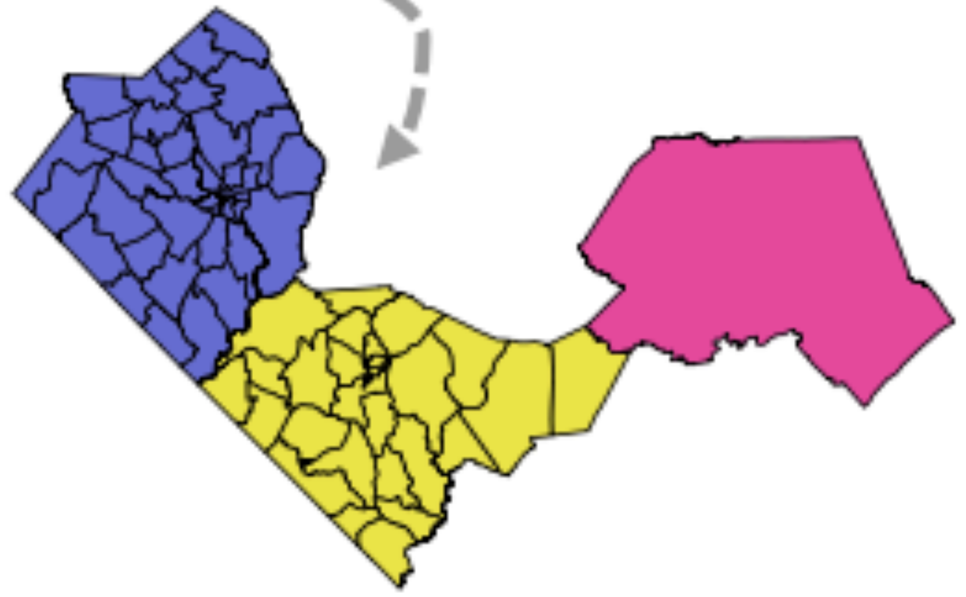
NC Legislature





Duplin (green) & **Onslow** (red) form a county cluster which must be divided into 3 individual NC House Legislative districts.

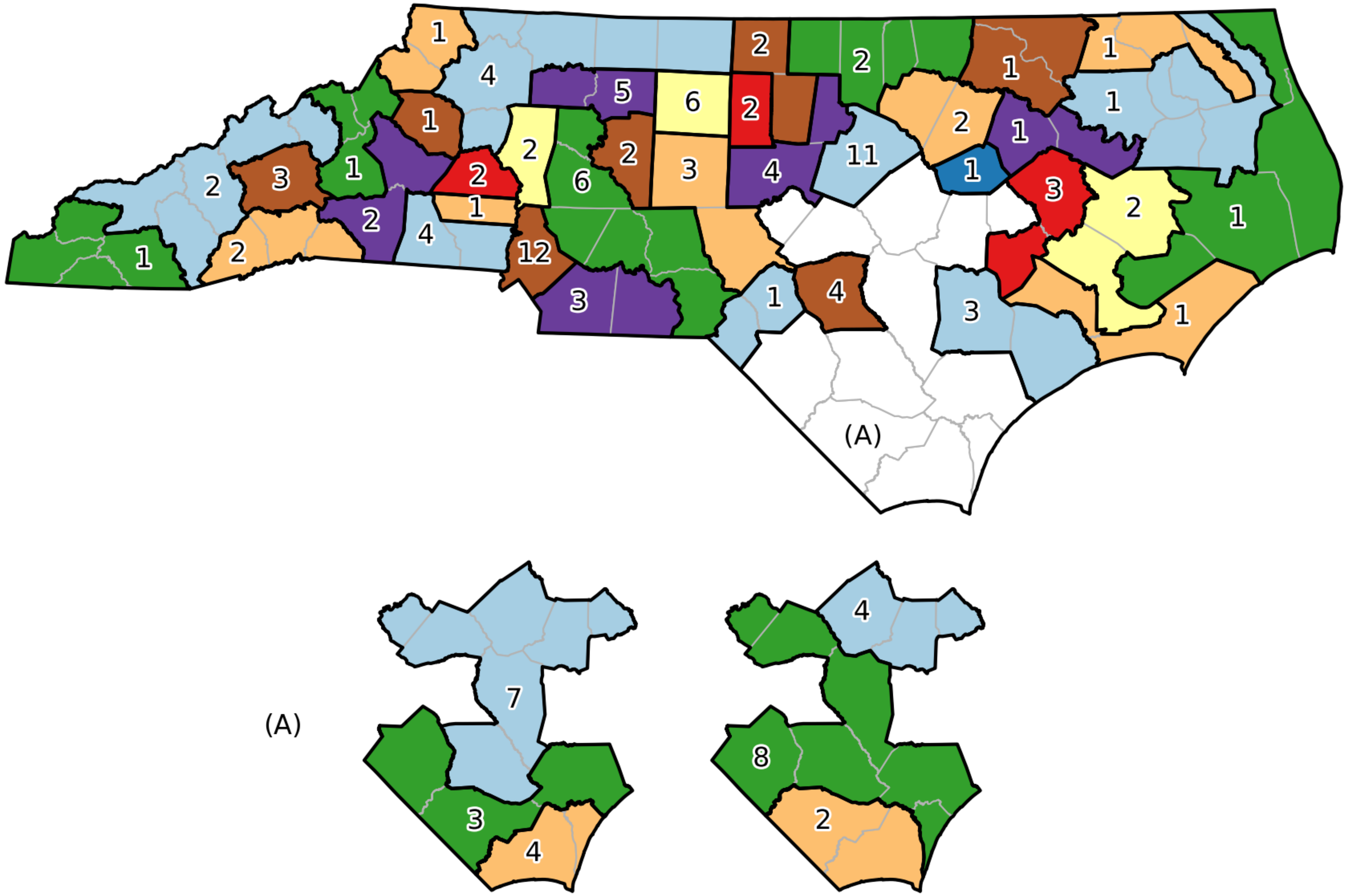
Union (green) & **Anson** (red) form a county cluster with two counties making up 3 individual NC House Legislative districts.



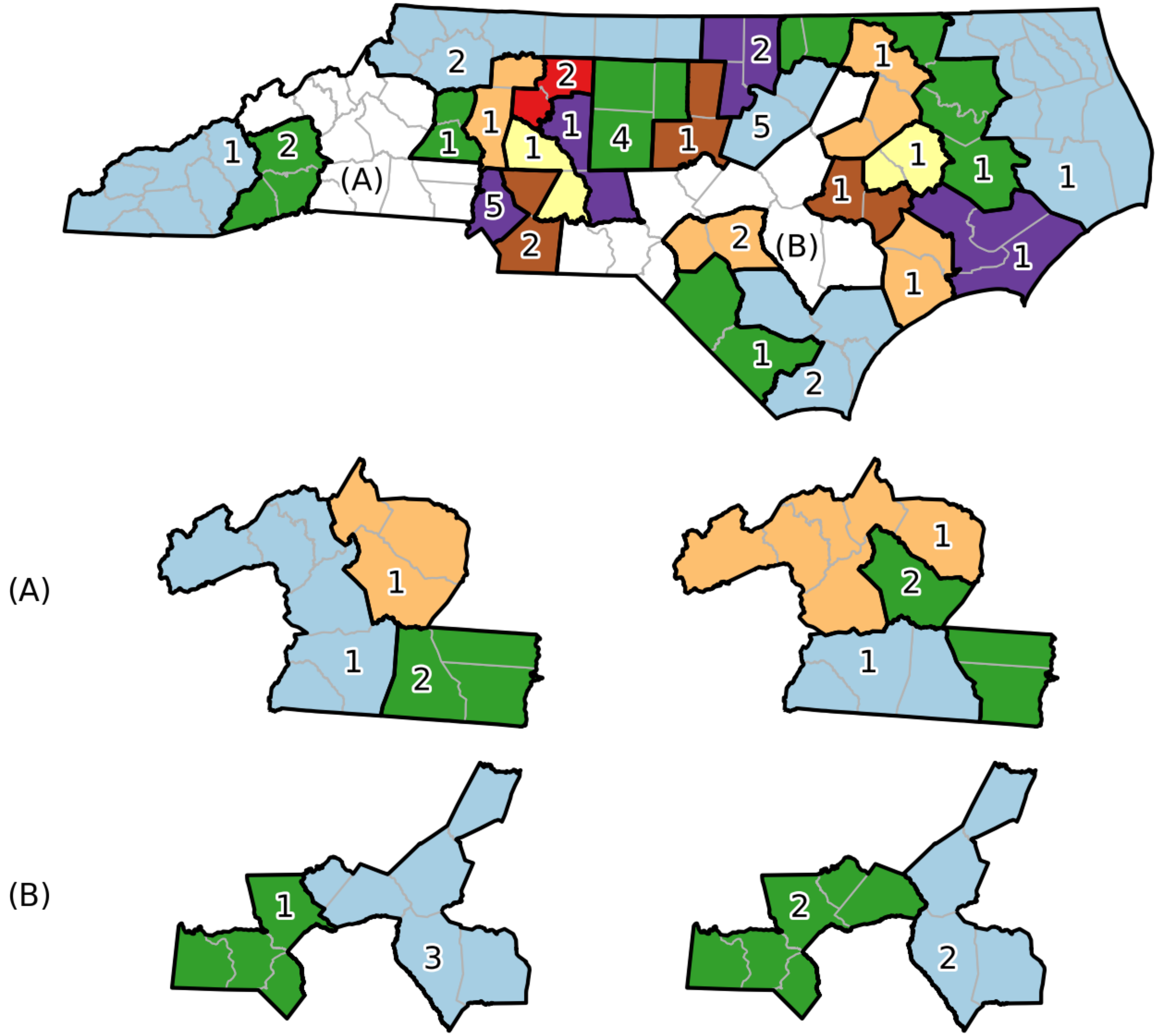
Pender (purple), **Robeson** (blue), & **Columbus** (yellow) form a county cluster which must be divided into 3 individual NC House Legislative districts.

Court order: Draw most single-county clusters, then the most two-county clusters, and so on. Maximize the number of clusters.

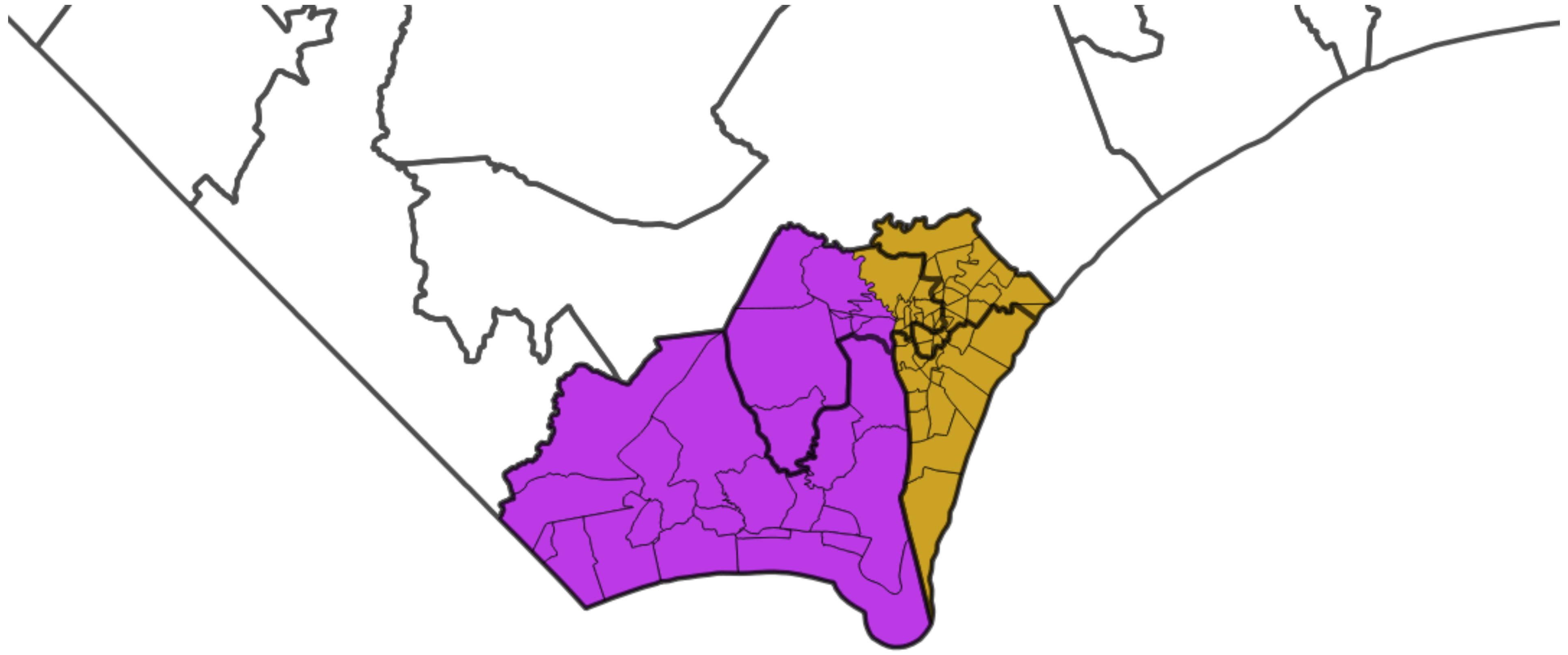
Optimal Clusters



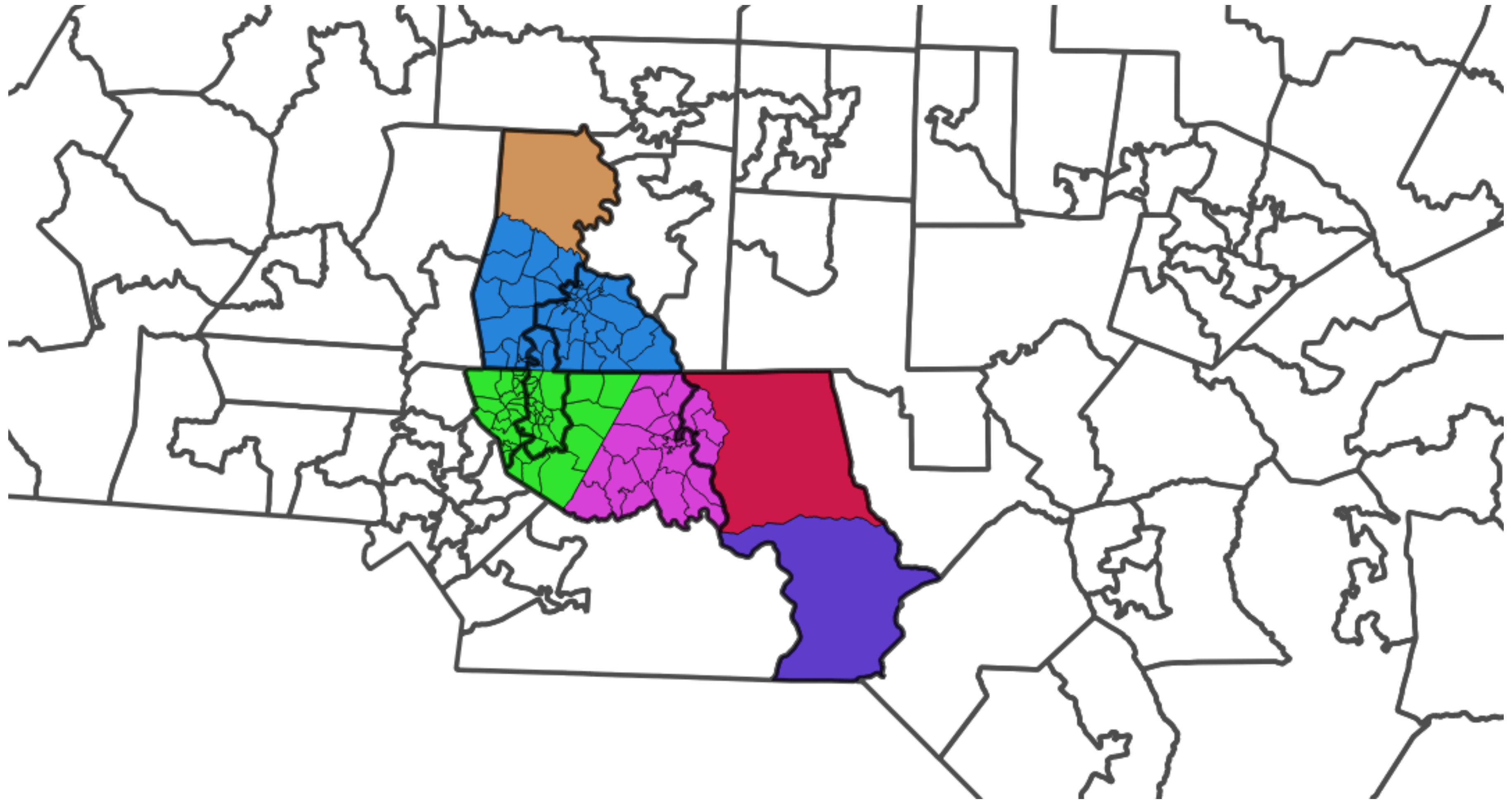
Optimal Clusters



Implicit Rules

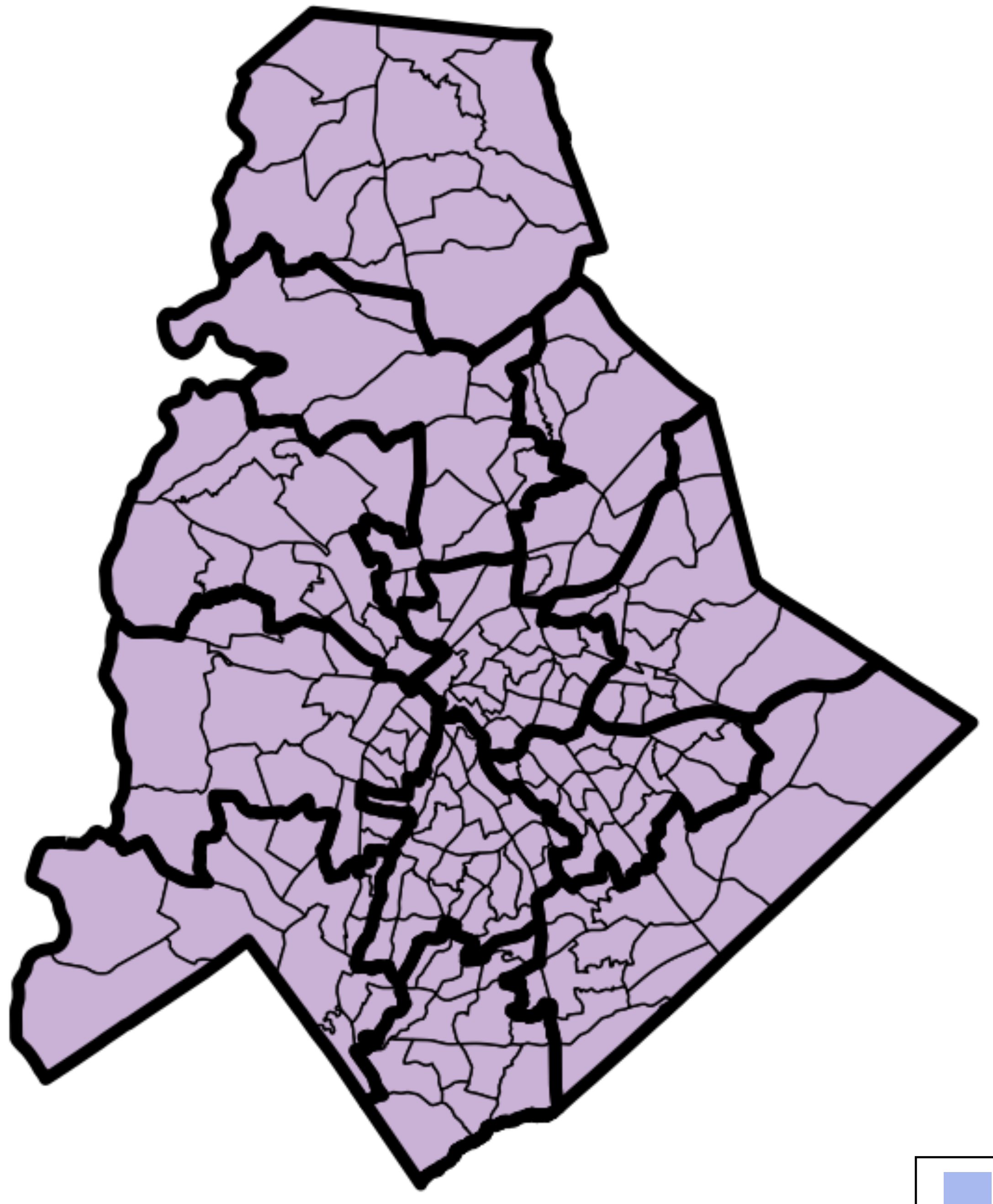


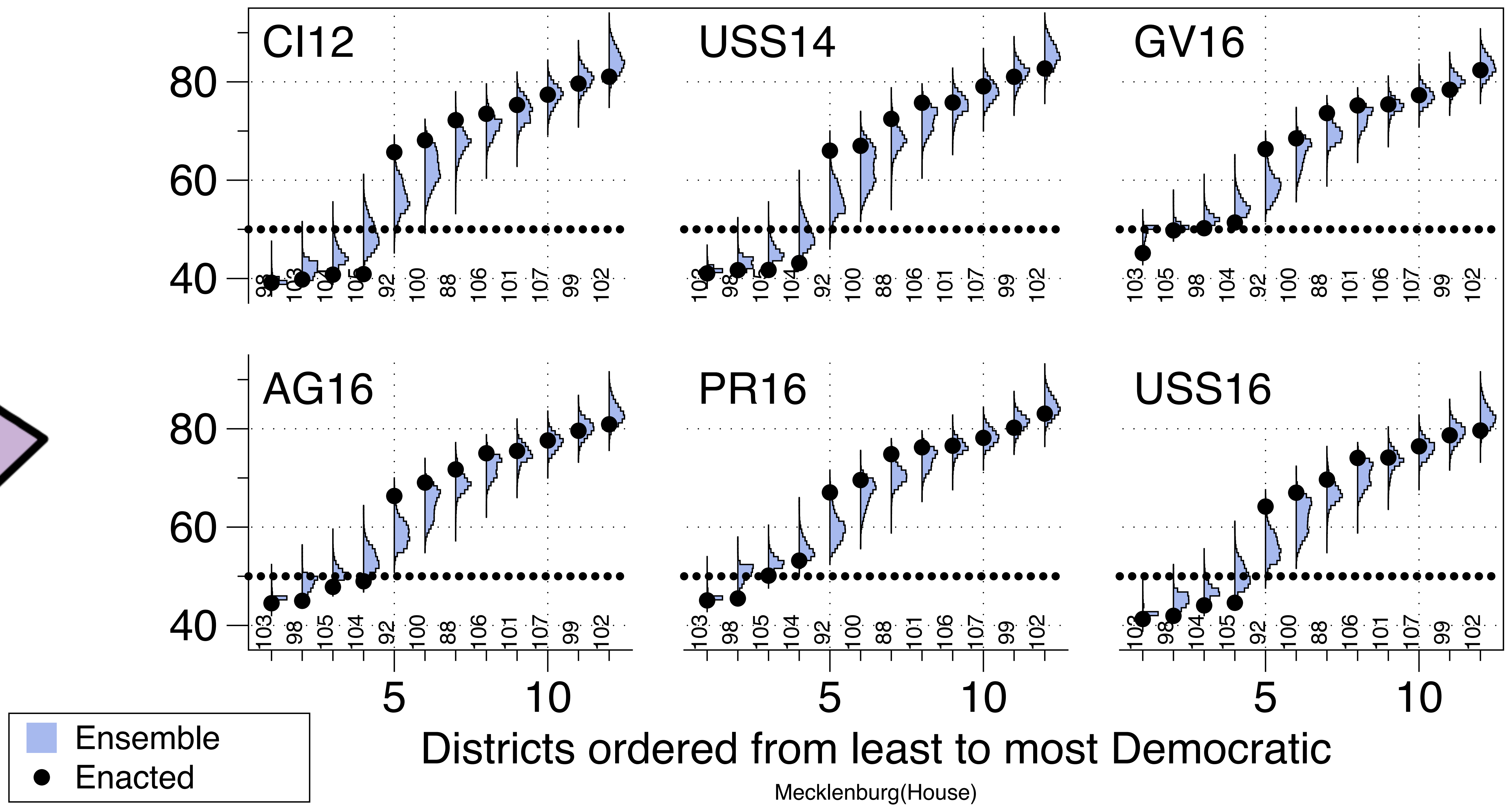
Counties boundaries are only crossed by a single district



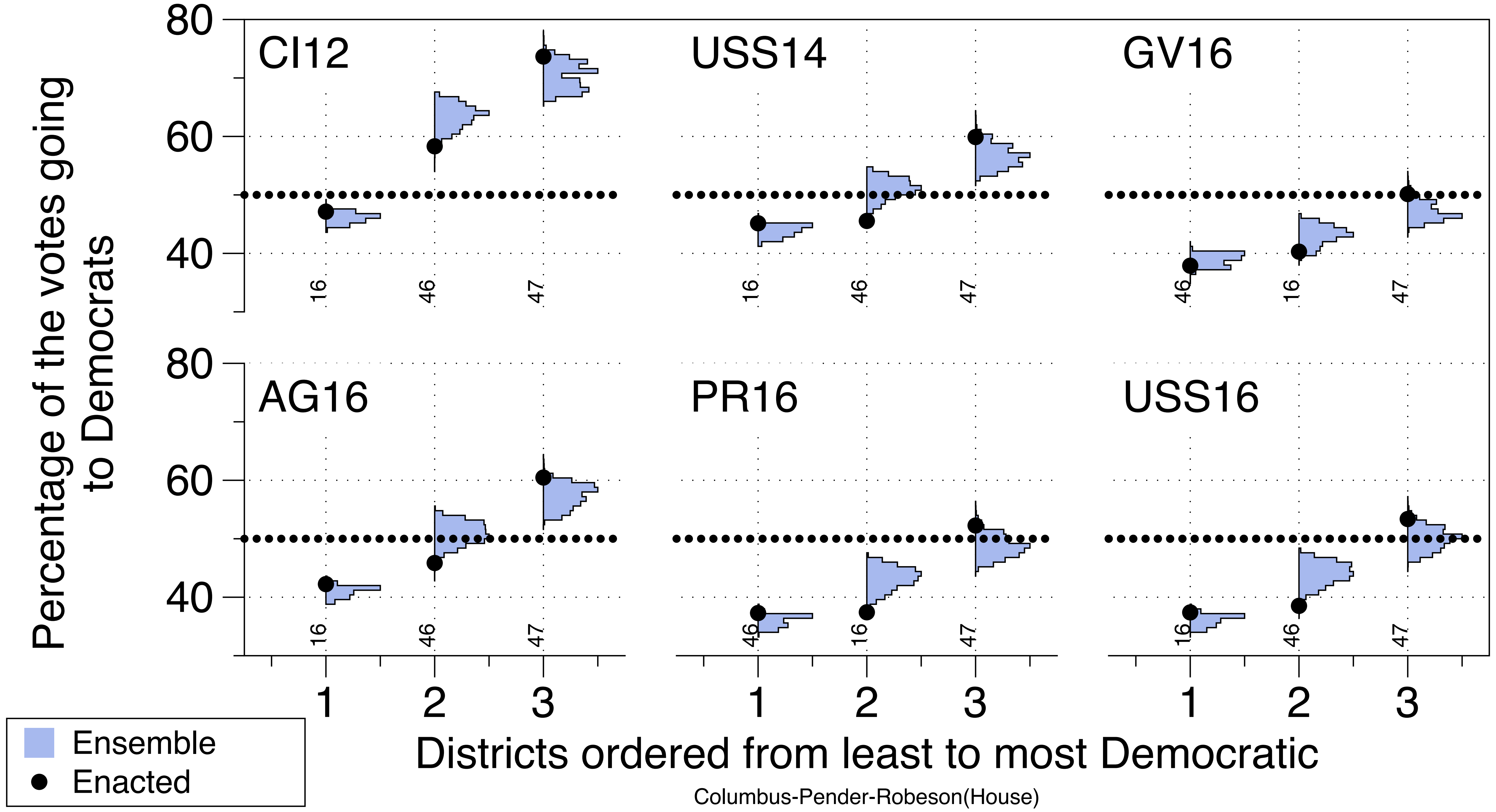
When possible districts only span two counties

Counties are kept intact to the extent possible even when split

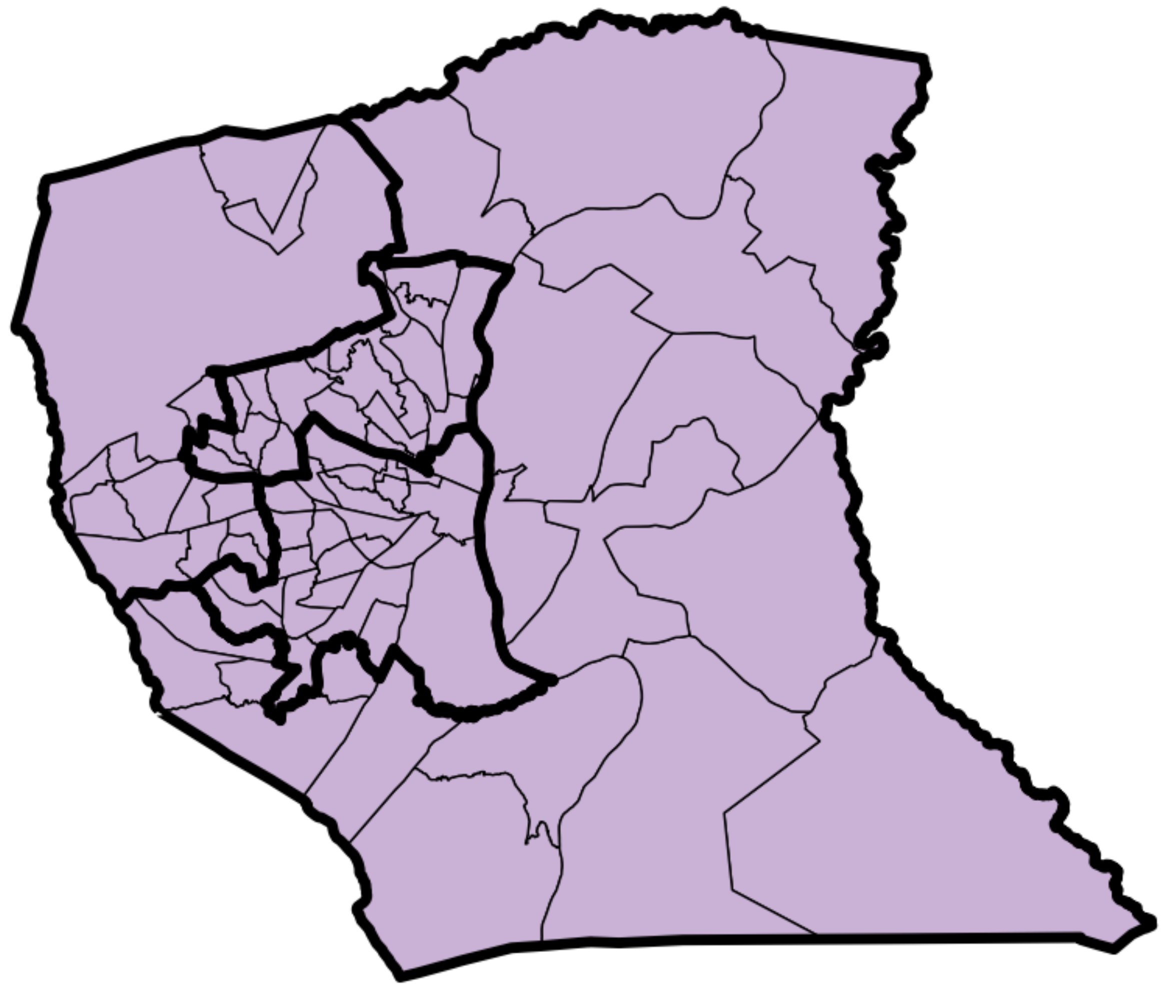




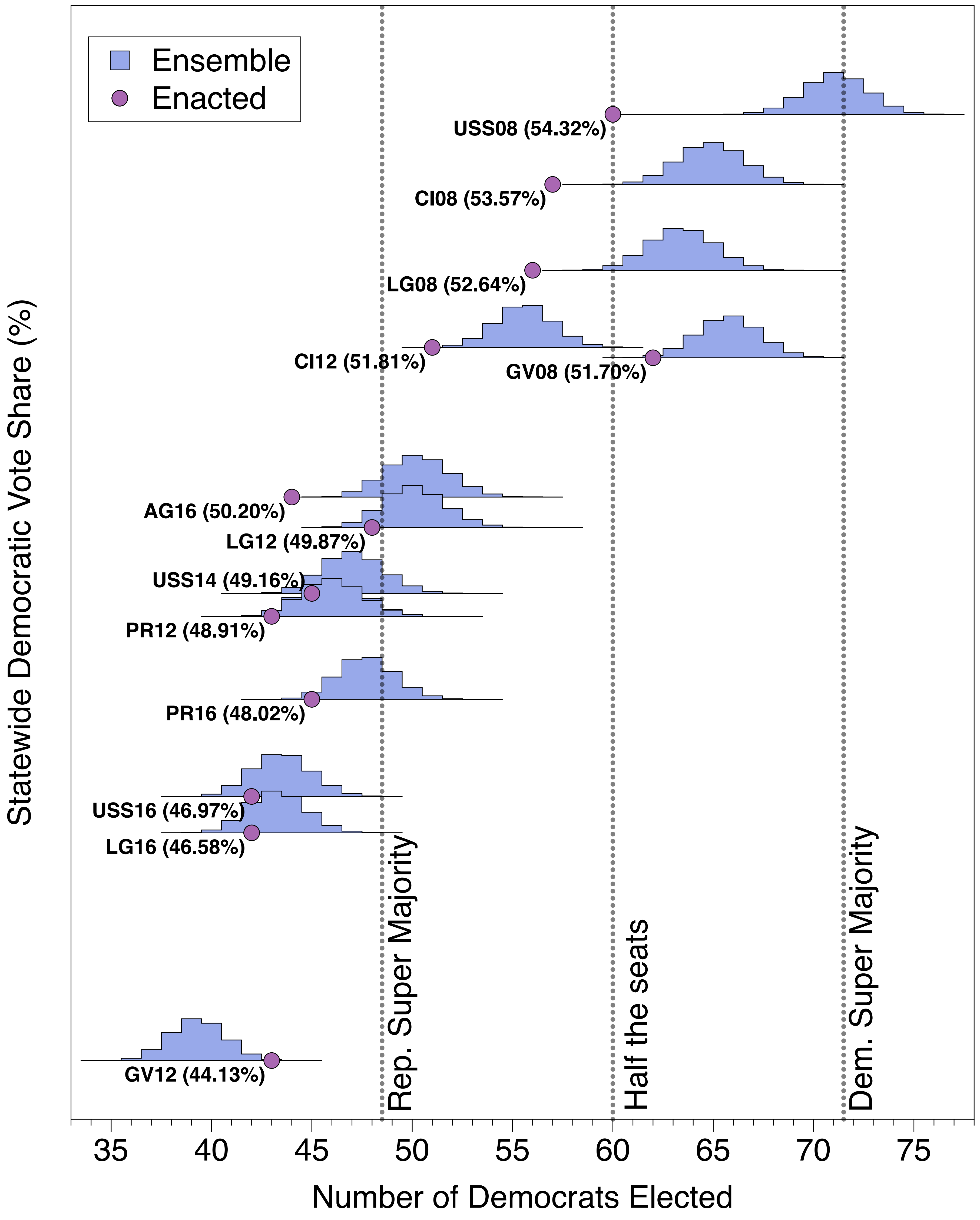
Mecklenburg(House)

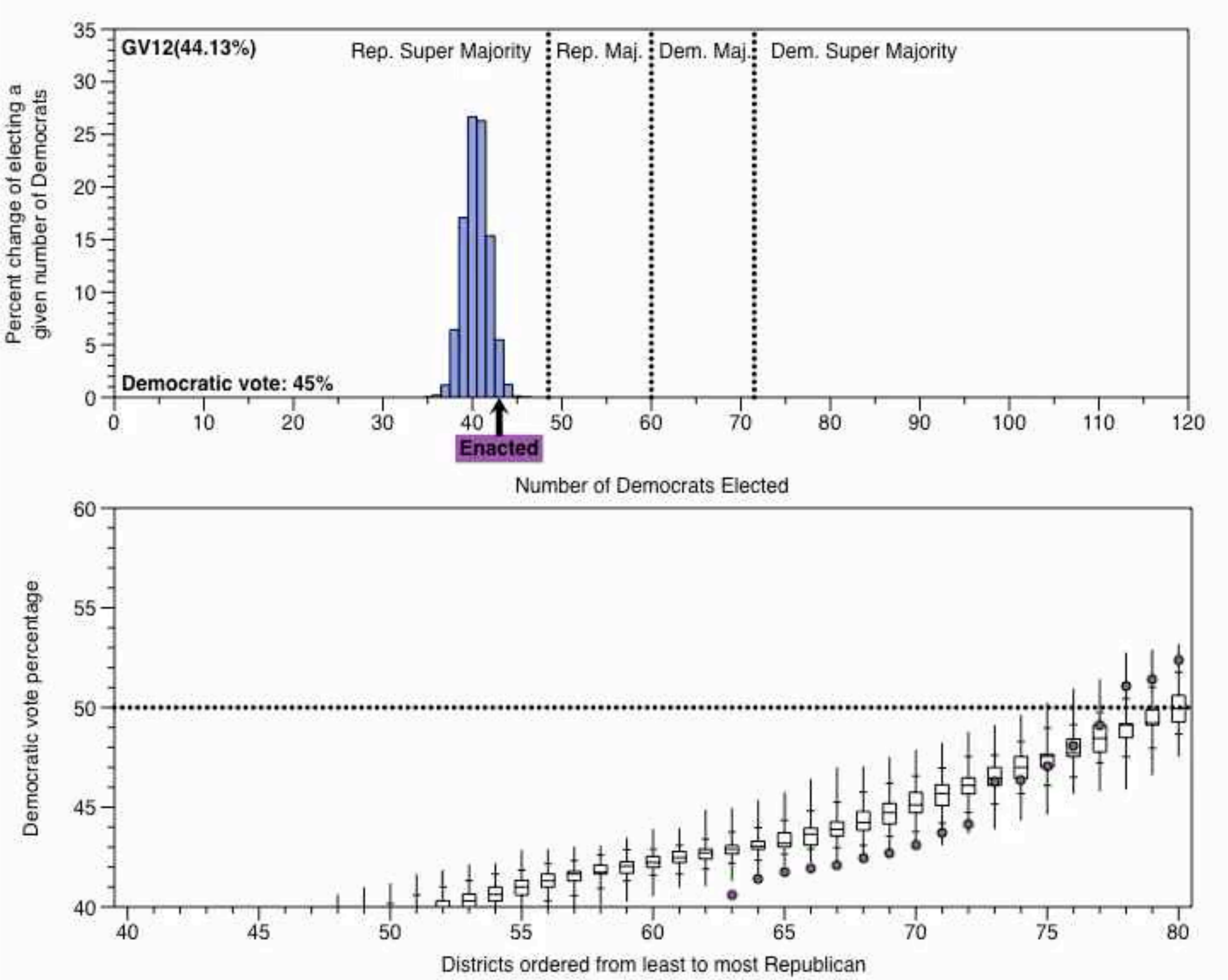


Columbus-Pender-Robeson(House)



Across many elections





The Team is Constantly Growing

Christy Vaughn Graves (UG; 2013-2016)
Sachet Bangia (UG; 2016-2017)
Sophie Guo (UG; 2016)
Bridget Duo (UG; 2016)
Hansung Kang (UG; 2016-2017)
Justin Luo (UG; 2016-2017)
Michael Kepler (MS; 2018)
Sam Eure (UG; 2018-2019)
Mike Bell (GS; 2017-2019)
Rahul Ramesh (UG; 2018-Present)
Lisa Lebovich (MS; 2018-Present)
Robert Ravier (GS; 2016-Present)
Andrew Chin (2018-Present)
Zach Hunter (2019)
Daniel Carter (2019)
Matthias Sachs (2019-Present)
Eric Autrey (2019-Present)
Jonathan Mattingly (2013-Present)
Gregory Herschlag (2016-Present)

Bass Connections Class
(year long UG class)

Claire Weibe	Samuel Eure
Ella Van Engen	Tiffany Mei
Jay Patel	Luke Farrell
Gillian Samios	Jake Shulman
Mitra Kiciman	Vinay Kshirsagar
Isaac Nicchitta	Rahul Ramesh
Nima Mohammadi	Haley Sink
Yashas Manjunatha	Jacob Rubin
Rayan Tofique	Chris Welland
Samuel Eure	Lynn Fan
Tiffany Mei	Jake Shulman
Luke Farrell	Vinay Kshirsagar

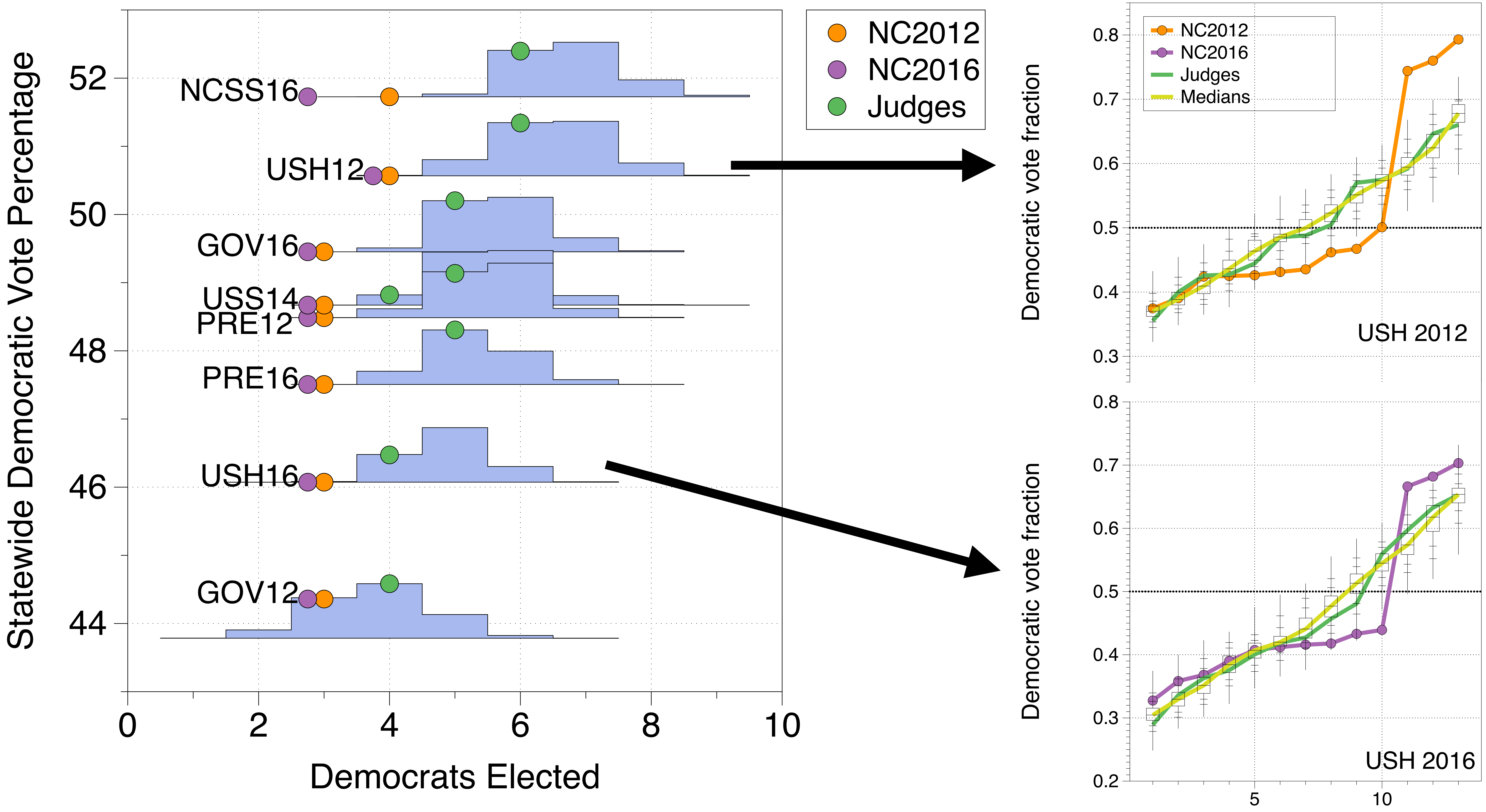
Collaborators

Guy-Uriel Charles
Janice McCarthy
Lydia Kwee
Andrew Chin
Colin Rundel
Jason Parsely
Adam Graham-Squire
Stephen Schecter
Wes Pegden

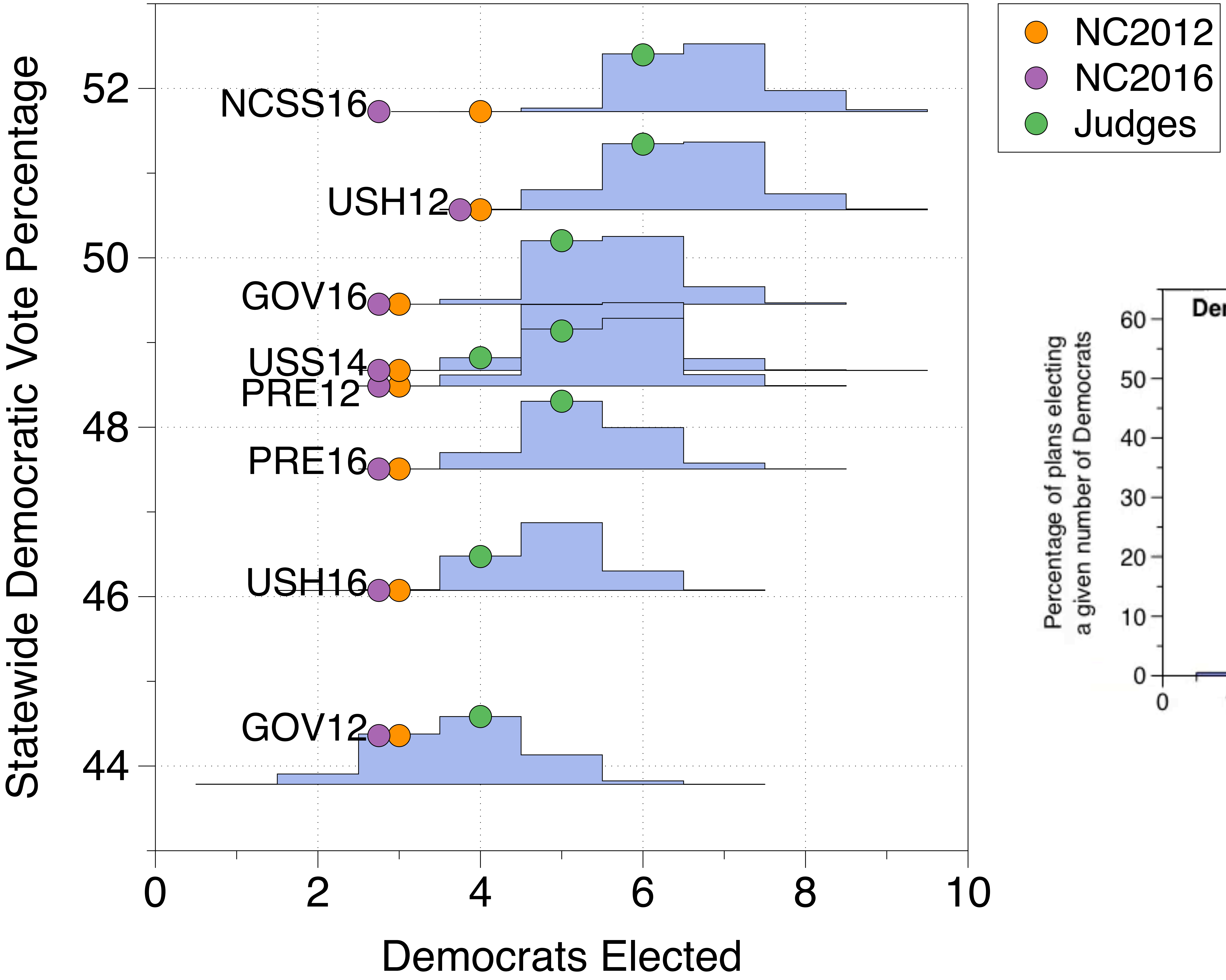
Duke UNIVERSITY MATH · Data+ · iiD · Duke UNIVERSITY | POLIS

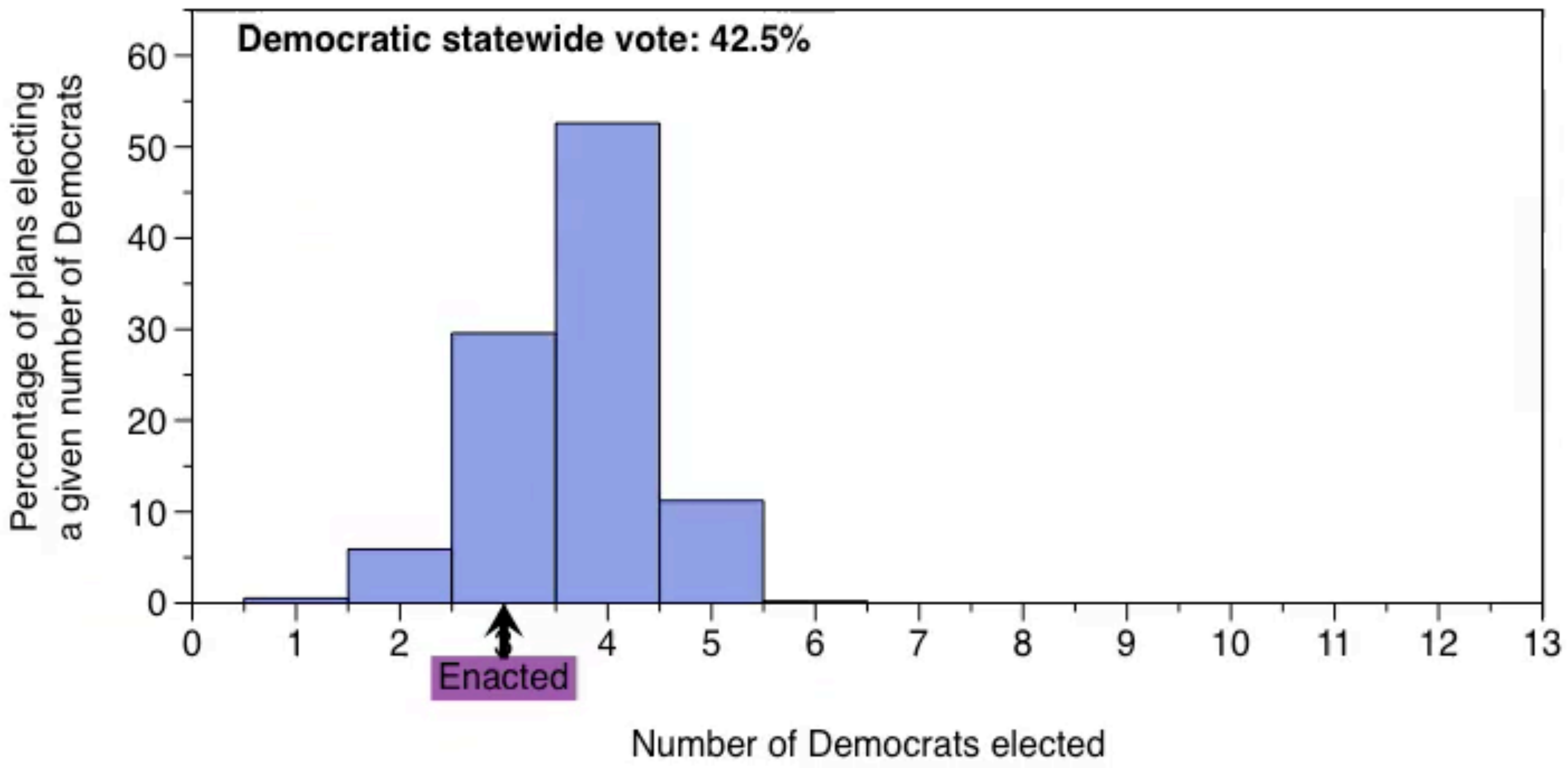
Blog: https://sites.duke.edu/quantifyinggerrymandering/

Stagnating election results



Across many elections





Partition v. Spanning Forest

Initial Partition: $\xi = (\xi_1, \ldots, \xi_i, \ldots, \xi_j, \ldots, \xi_n)$

Target Partition: $\xi' = (\xi_1, \ldots, \xi_i', \ldots, \xi_j', \ldots, \xi_n)$

ReCom:

$$A(\xi, \xi') = e^{-\beta[J(\xi')-J(\xi)]}\frac{Q(\xi',\xi)}{Q(\xi,\xi')}\frac{\tau(\xi)^{\gamma-1}}{\tau(\xi')^{\gamma-1}}$$

$Q(\xi,\xi')$ & $Q(\xi',\xi)$ **expensive**

Merge-Split:

$$A(T, T') = e^{-\beta[J(\xi')-J(\xi)]}\frac{Q(T',T)}{Q(T,T')}\frac{\tau(\xi)^{\gamma}}{\tau(\xi')^{\gamma}}$$

$Q(\xi,\xi')$ & $Q(\xi',\xi)$ **cheaper**

$$Q(\xi,\xi') : \{\xi_i, \xi_j\} \xrightarrow[\textbf{deterministic}]{\textbf{many-to-one}} \xi_{ij} \xrightarrow[\textbf{random}]{\textbf{one-to-many}} T_{ij}' \xrightarrow[\textbf{random}]{\textbf{one-to-a-few}} \{T_i', T_j'\} \xrightarrow[\textbf{random}]{\textbf{many-to-one}} \{\xi_i', \xi_j'\}$$

$$Q(T.T') : \{T_i, T_j\} \xrightarrow[\textbf{deterministic}]{\textbf{many-to-one}} \xi_{ij} \xrightarrow[\textbf{random}]{\textbf{one-to-many}} T_{ij}' \xrightarrow[\textbf{random}]{\textbf{one-to-a-few}} \{T_i', T_j'\}$$