

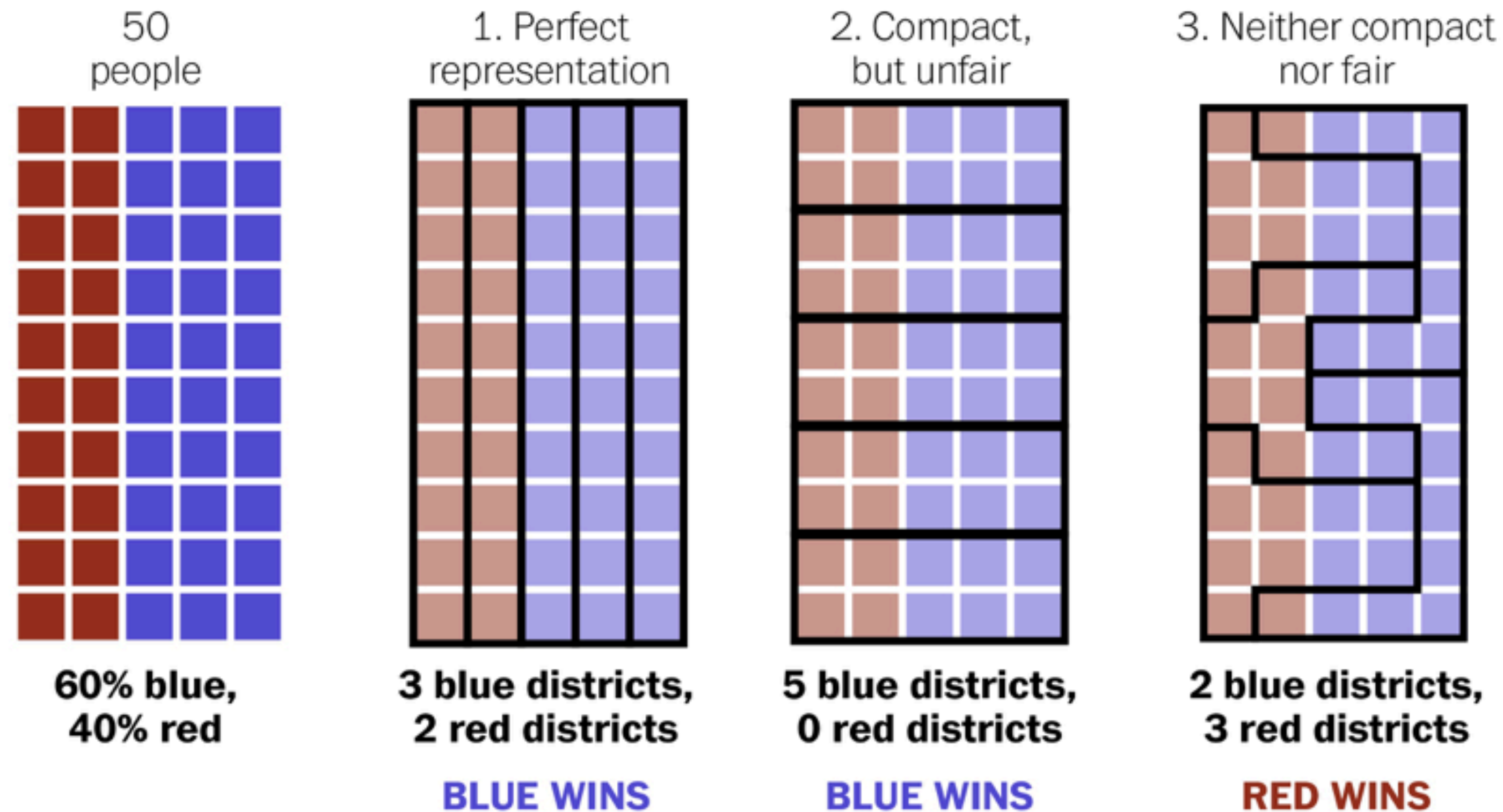
Hybrid Tree Visualizations for Analysis of Gerrymandering

Chenguang Xu (Shine)
Oklahoma City University
shine.xu@okcu.edu

Gerrymandering

Manipulate the boundaries of an electoral constituency so as to favor one party or class.

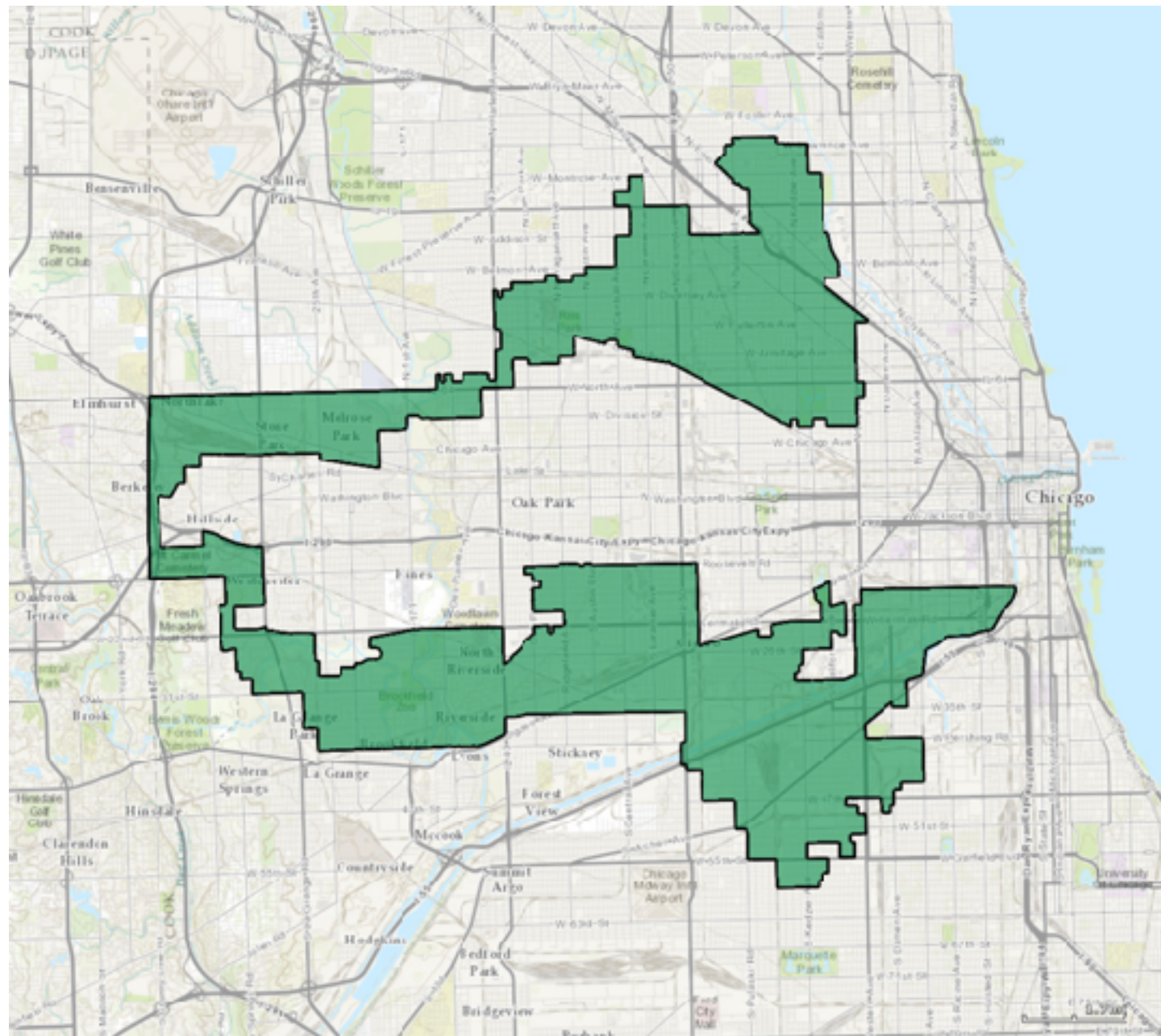
Three different ways to divide 50 people into five districts



<https://www.washingtonpost.com/news/wonk/wp/2015/03/01/this-is-the-best-explanation-of-gerrymandering-you-will-ever-see/>

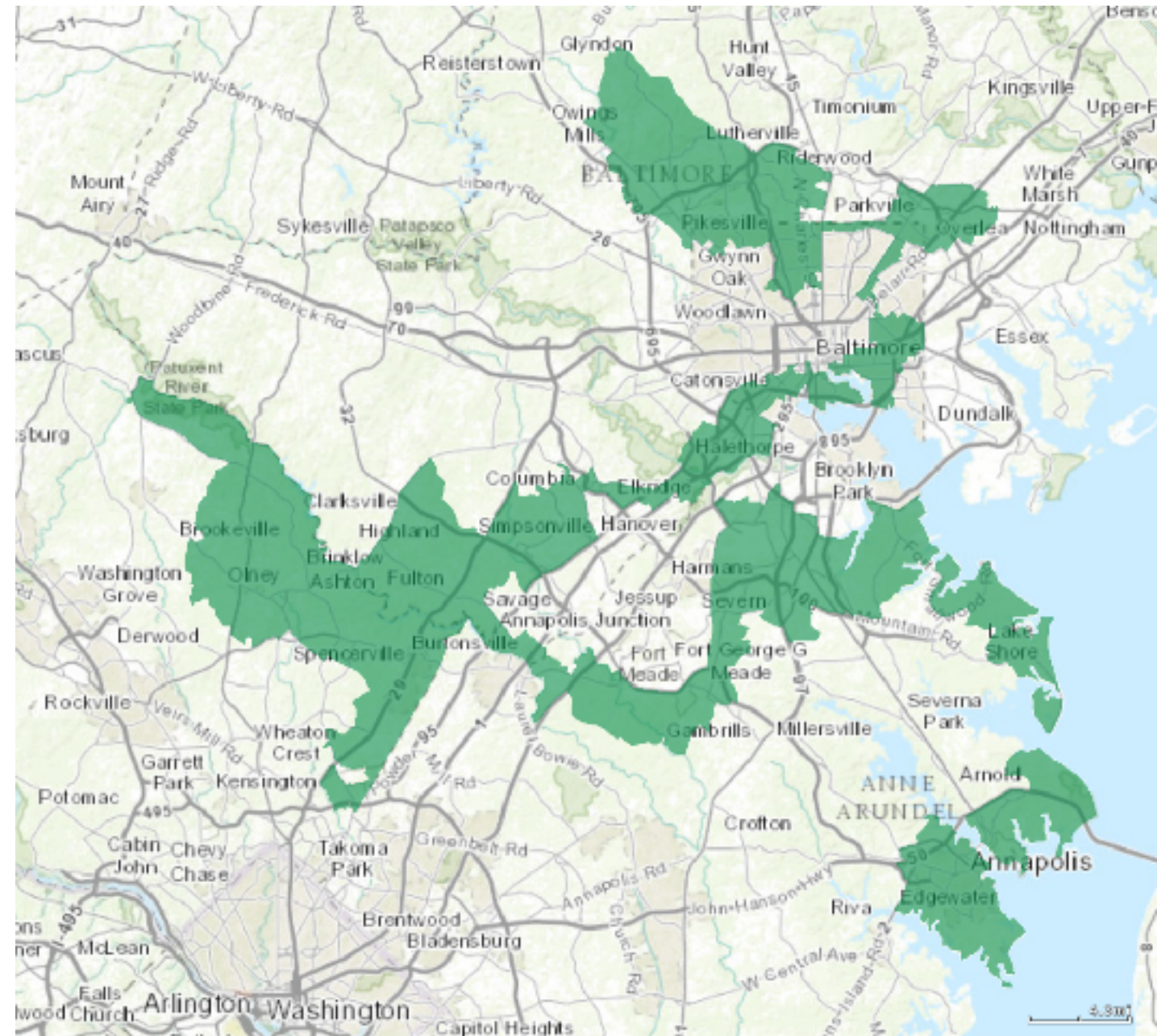
District shapes look weird

Illinois's 4th congressional district



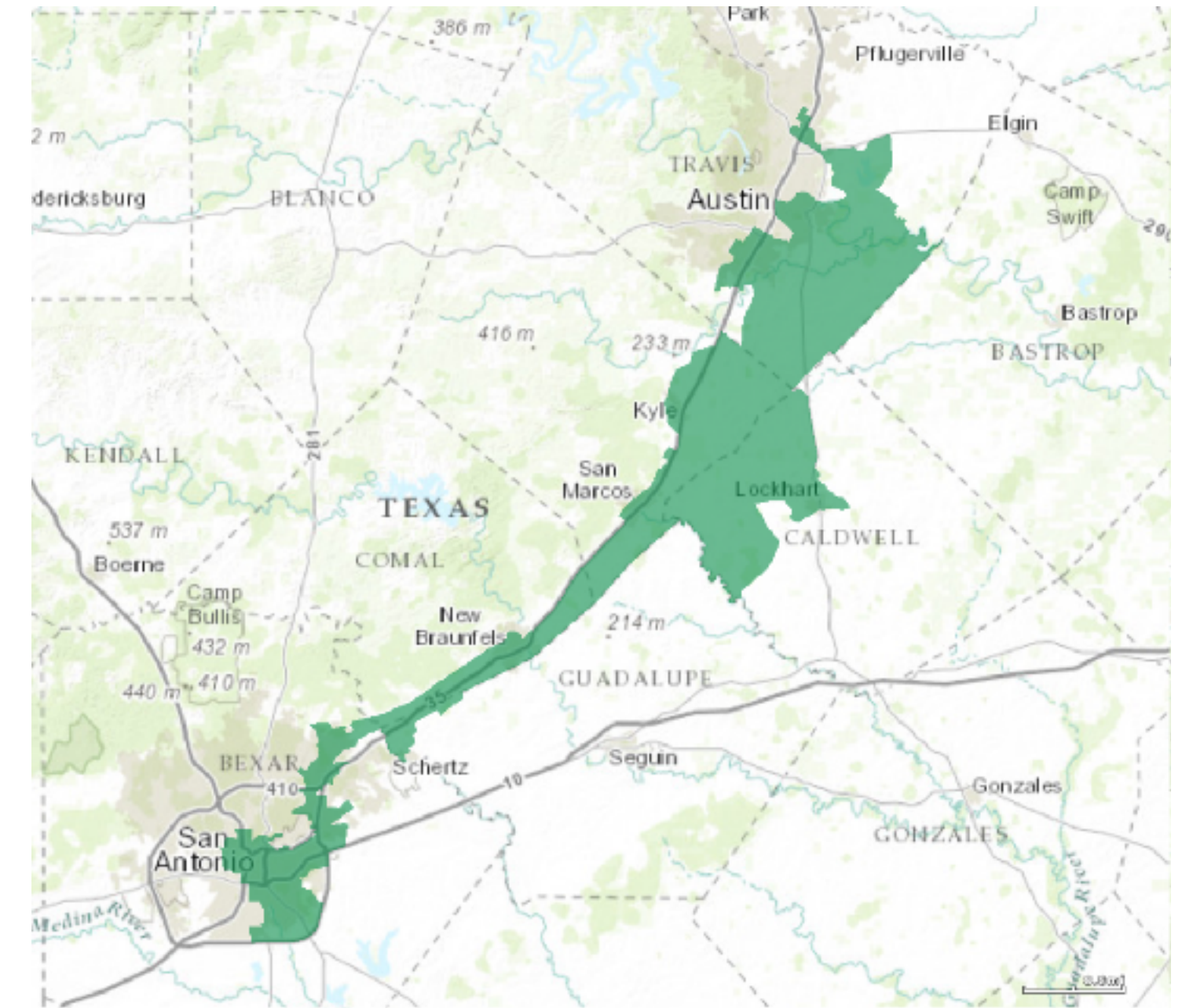
2013-2023

Maryland's 3rd congressional district



2013-2023

Texas's 35th congressional district

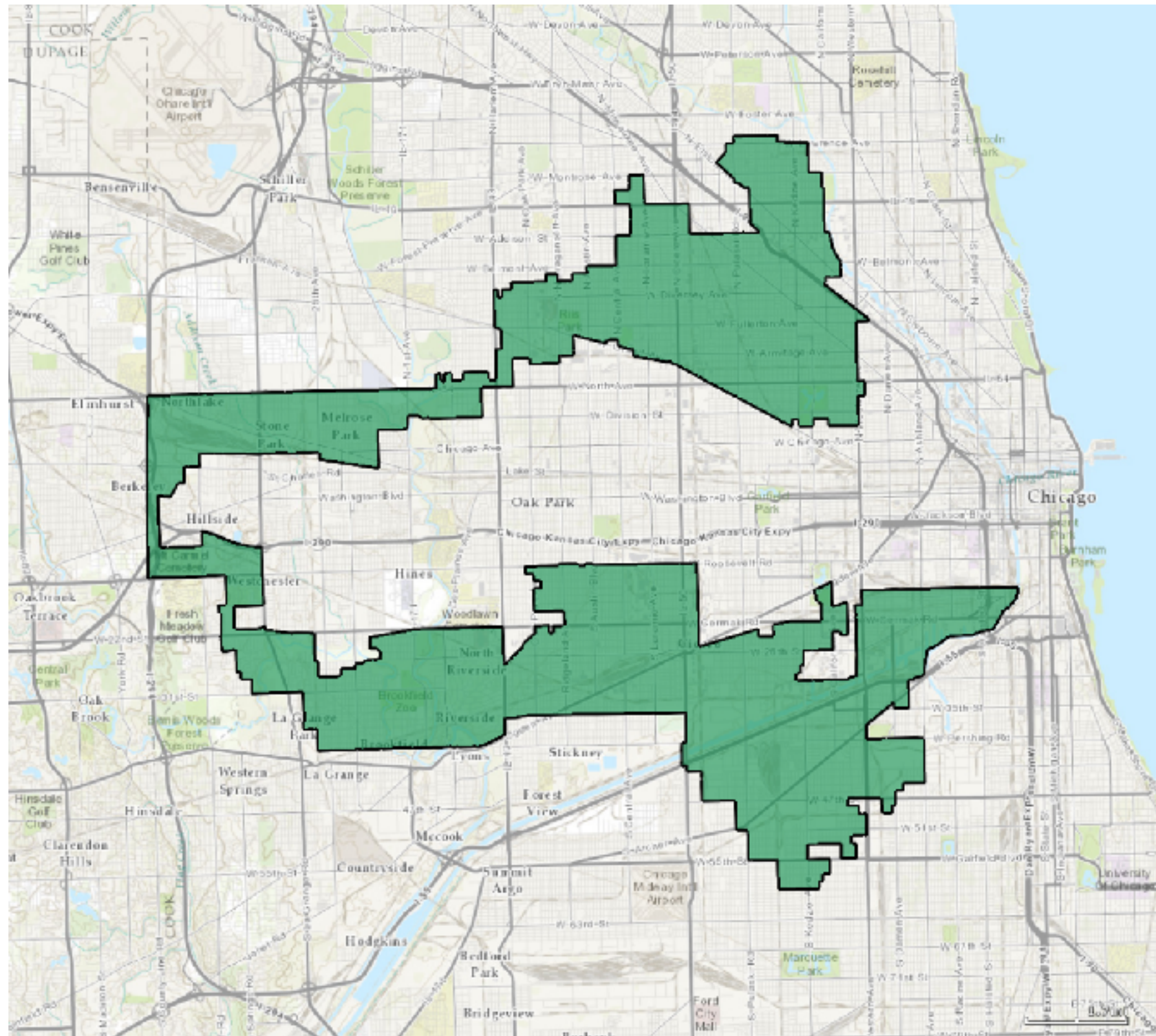


2013-2023

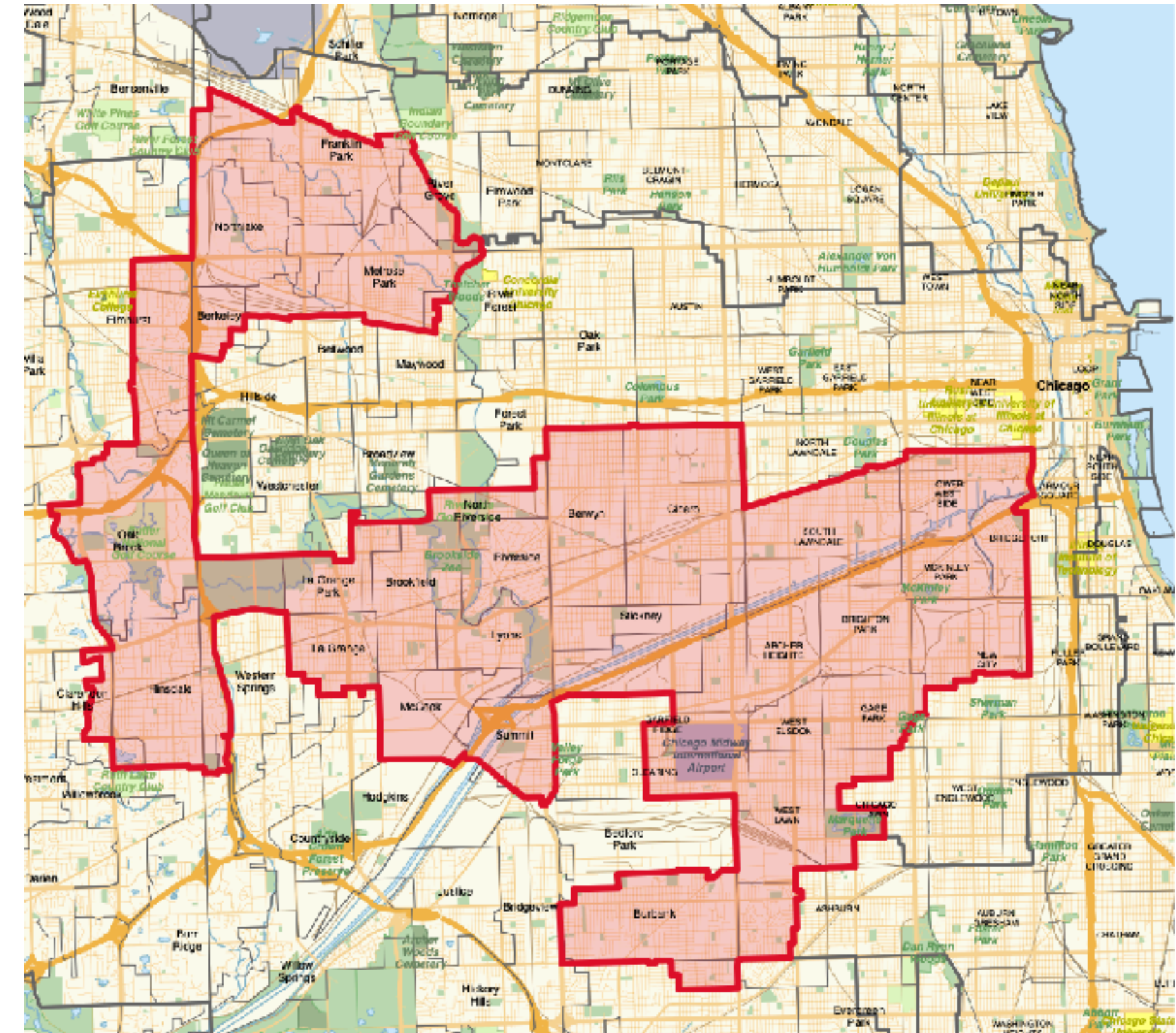
Data Source: Wikipedia

Redistricting changes district's shape

Illinois's 4th congressional district



2013-2023



2023-present

What are the rules?

Equal Population

Contiguity: connected, hole-free region

Compact: reasonable shape

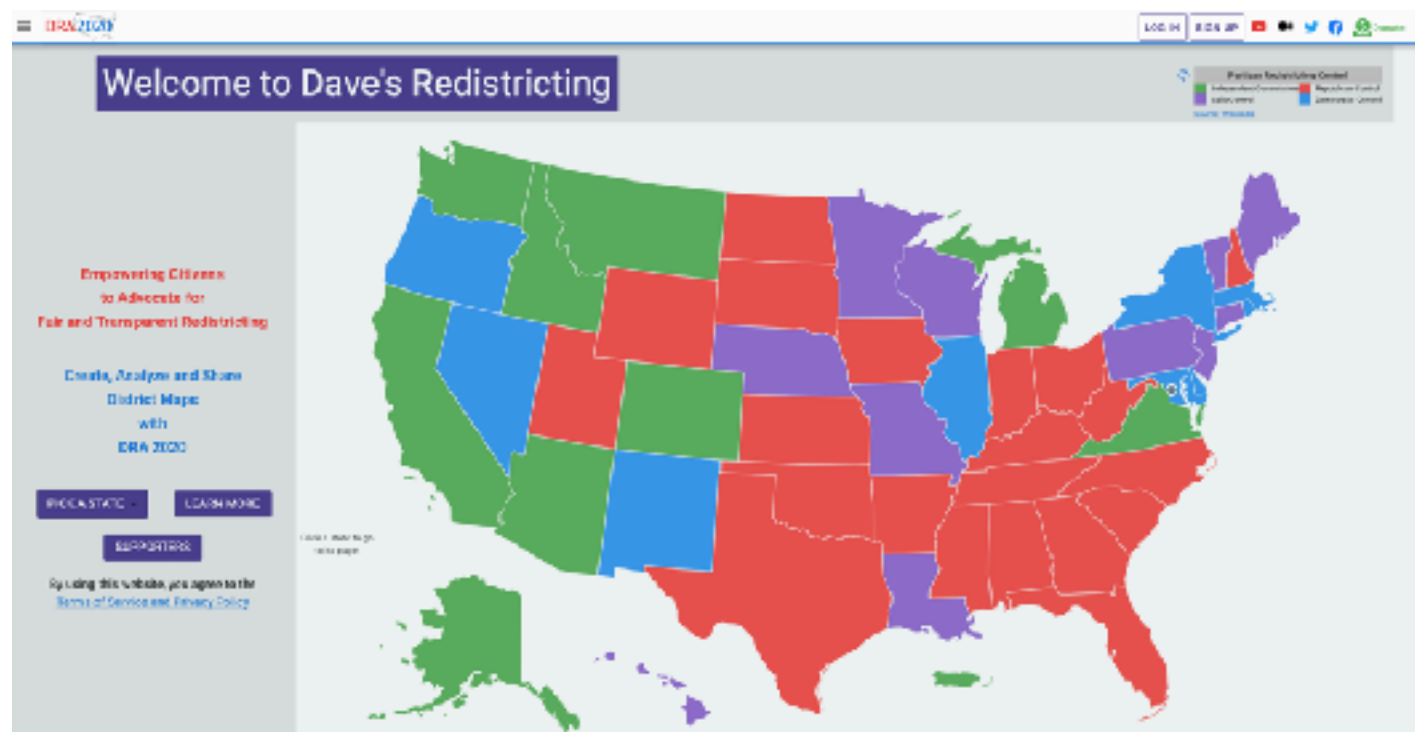
Avoid splitting counties/cities

Avoid splitting “communities of interest”

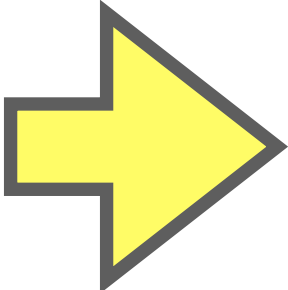
Respect Voting Rights Act – “opportunity” for minority groups

.....

Precinct-Level Data



extract data

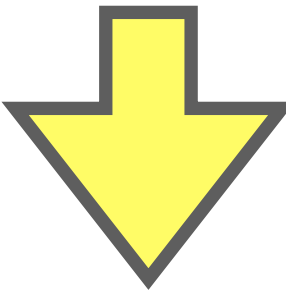


2010 Districting Data

GEOID	District	Dem	Rep
40083000305	3	206	592

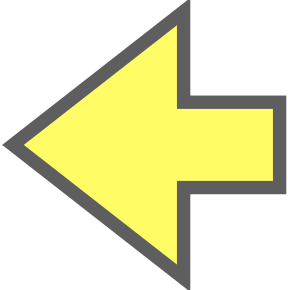
2020 Districting Data

GEOID	District	Dem	Rep
40083000303_part1_51	3	38	106
40083000303_part2_125	5	168	486



merge data

reshape



District 2010 Cycle	District 2020 Cycle	GEOID20	Party	Votes 2016	Votes 2020
2	2	40079000102	DEM	75	82
2	2	40079000102	REP	385	412
3	3	40113000113	DEM	17	14
3	3	40113000113	REP	37	40
3	3	40113000102	DEM	31	24
3	3	40113000102	REP	126	167

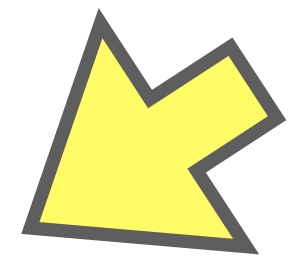
GEOID	District_2010	District_2020	Dem	Rep
40083000305	3	n.a.	206	592
40083000305_part1_51	n.a.	3	38	106
40083000305_part2_125	n.a.	5	168	486

Dave's Redistricting App.
<https://davesredistricting.org/maps#home>

Statistics in Gerrymandering Analysis

Voting Share Percentage(VSP) :

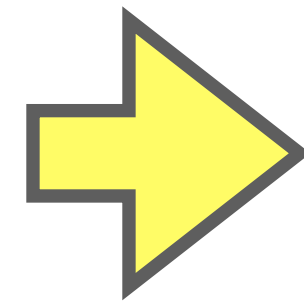
$$\frac{\sum (\text{votes} \wedge \text{party})}{\sum (\text{votes})}$$



District 2010 Cycle	District 2020 Cycle	GEOID20	Party	Votes 2016	Votes 2020
2	2	40079000102	DEM	75	82
2	2	40079000102	REP	385	412
3	3	40113000113	DEM	17	14
3	3	40113000113	REP	37	40
3	3	40113000102	DEM	31	24
3	3	40113000102	REP	126	167

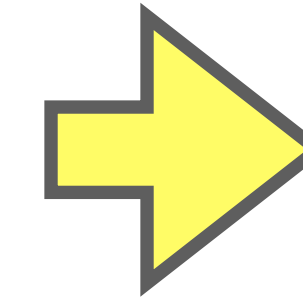
Relationship(R) :

$$VSP_{\text{Dem}} > VSP_{\text{Rep}}$$



Distance (d) :

$$d(R_{\text{state}}, R_{\text{district}})$$

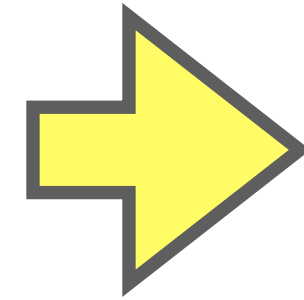


Summary (S) :

$$S = \text{average}(d)$$

VoteMargin(M) :

$$|VSP_{\text{Dem}} - VSP_{\text{Rep}}|$$



Summary (S) :

$$S = \text{count}(M)$$

How do we represent data visually?

Nominal/ Ordinal

Party

District Cycle

District

Quantitative

Voting Share Percentage (VSP)

Relationship (R)

Distance (d)

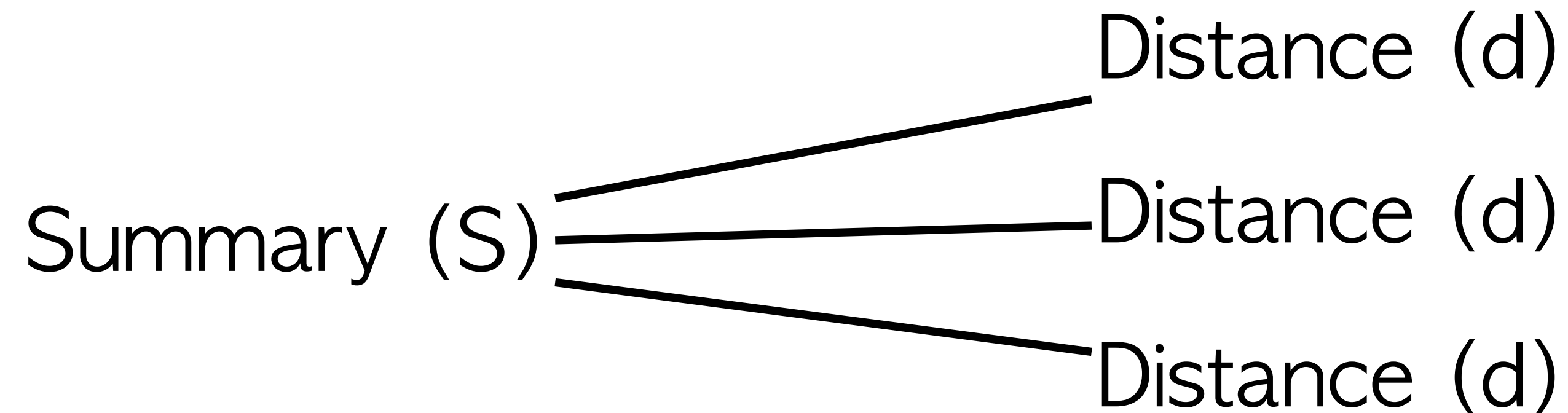
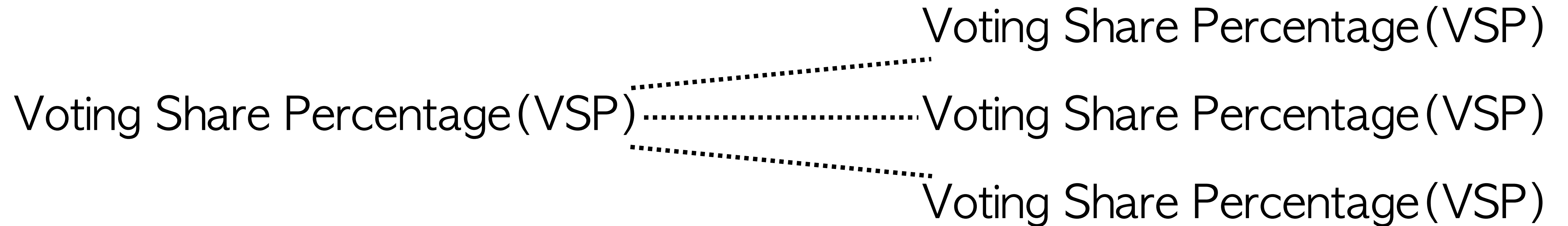
Summary (S)

Vote Margin (M)

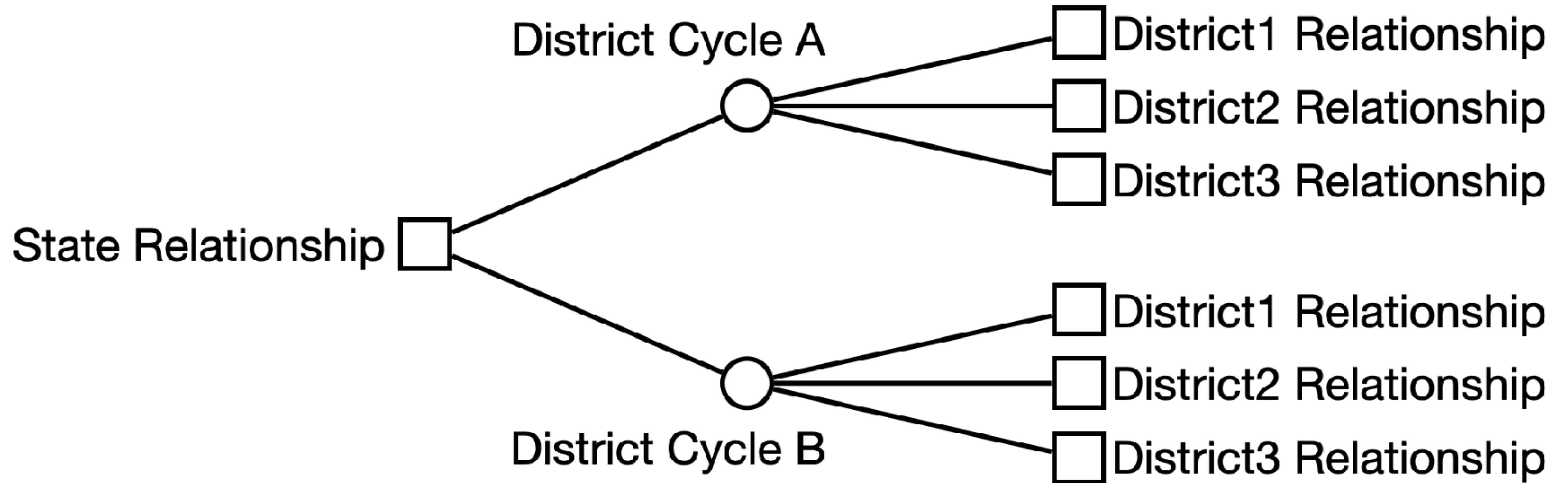
How do we represent data visually?

State-level

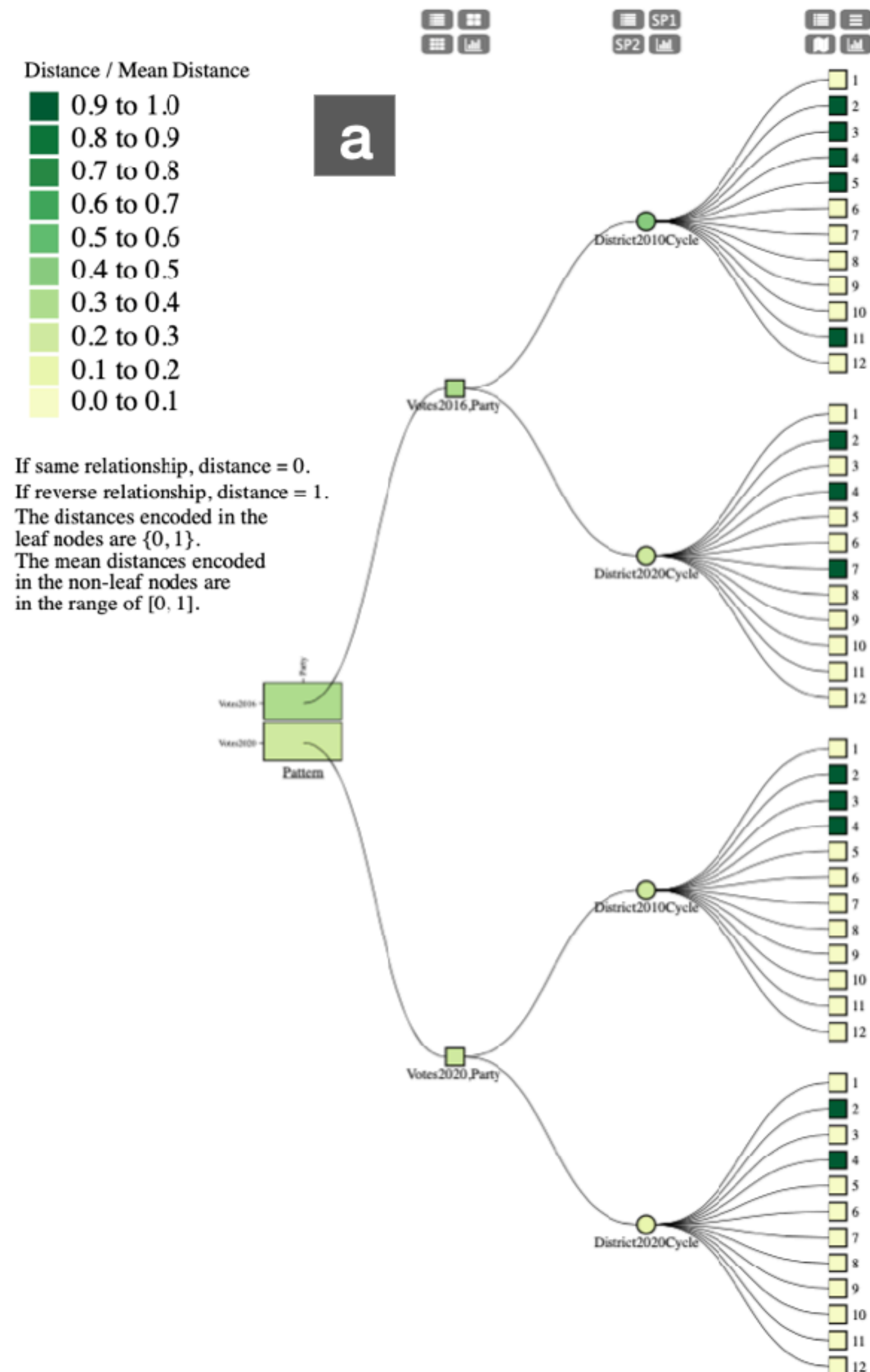
District-level



The hierarchical view design



Representing Data Visually



Nominal/ Ordinal

District Cycle

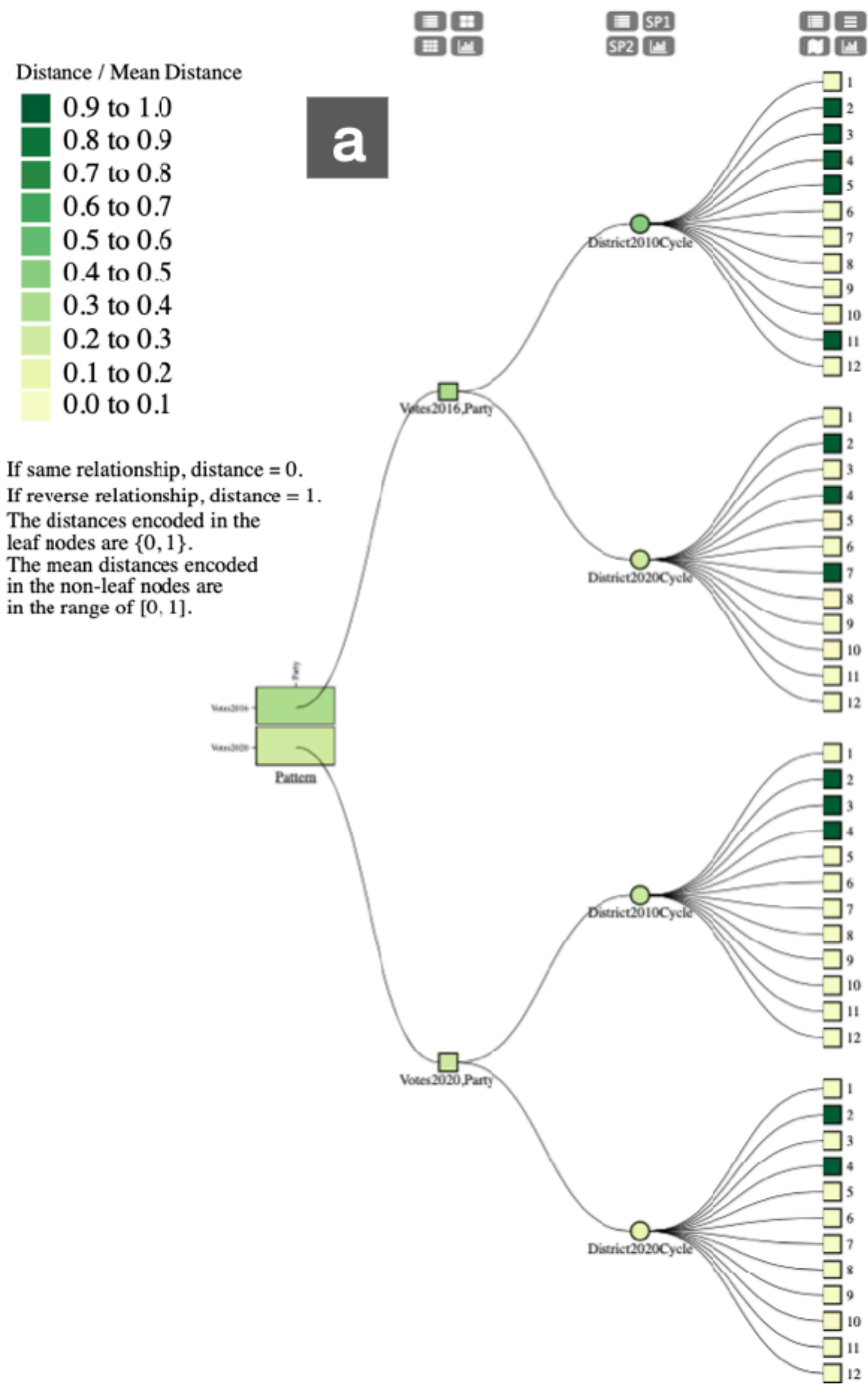
District

Quantitative

Distance (d)

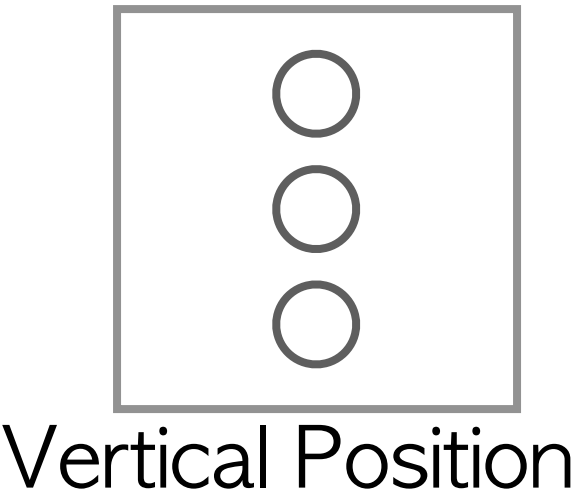
Summary (S)

Representing Data Visually



Nominal/ Ordinal

District Cycle



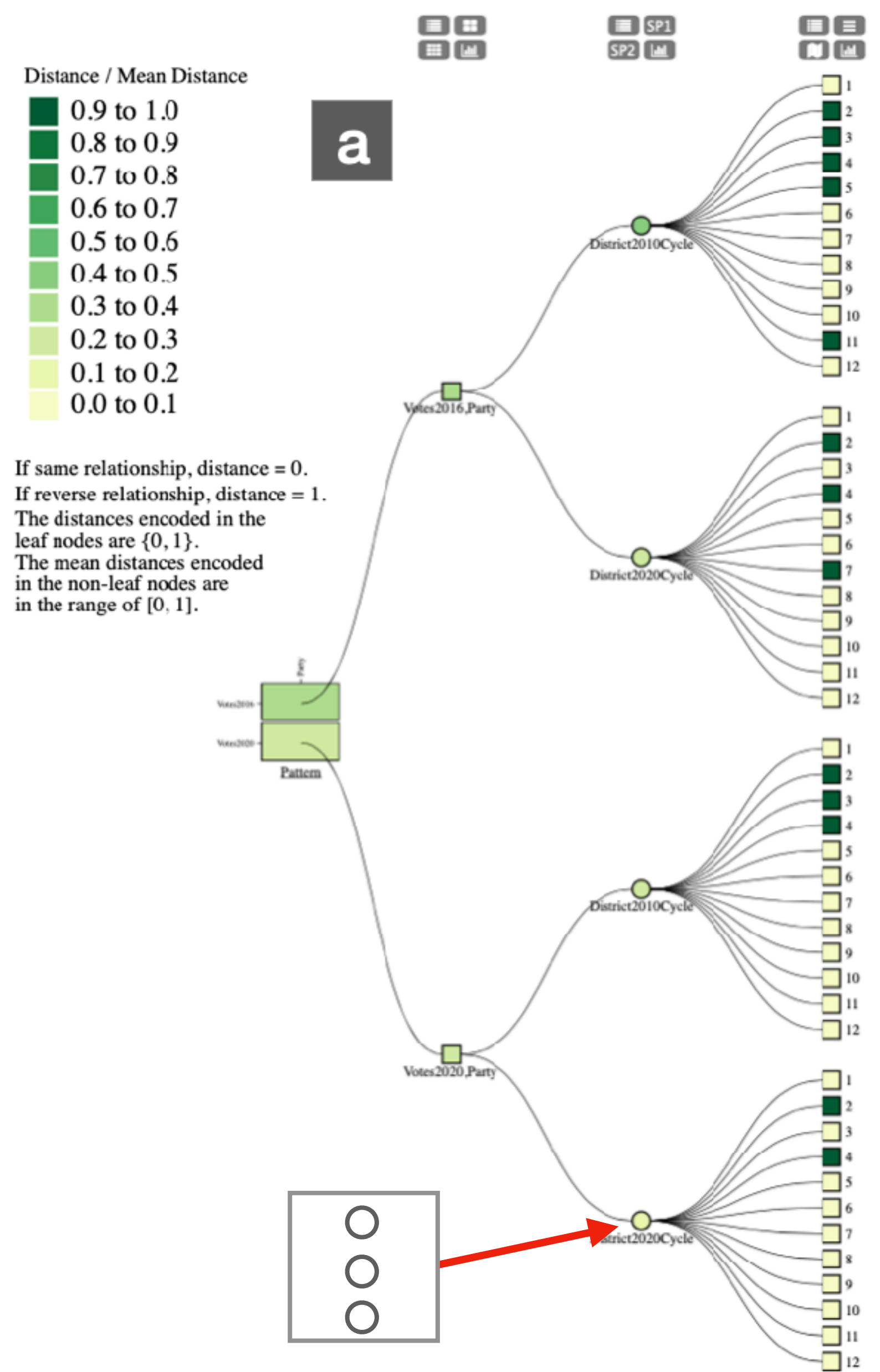
District

Quantitative

Distance (d)

Summary (S)

Representing Data Visually



Nominal/ Ordinal

District Cycle

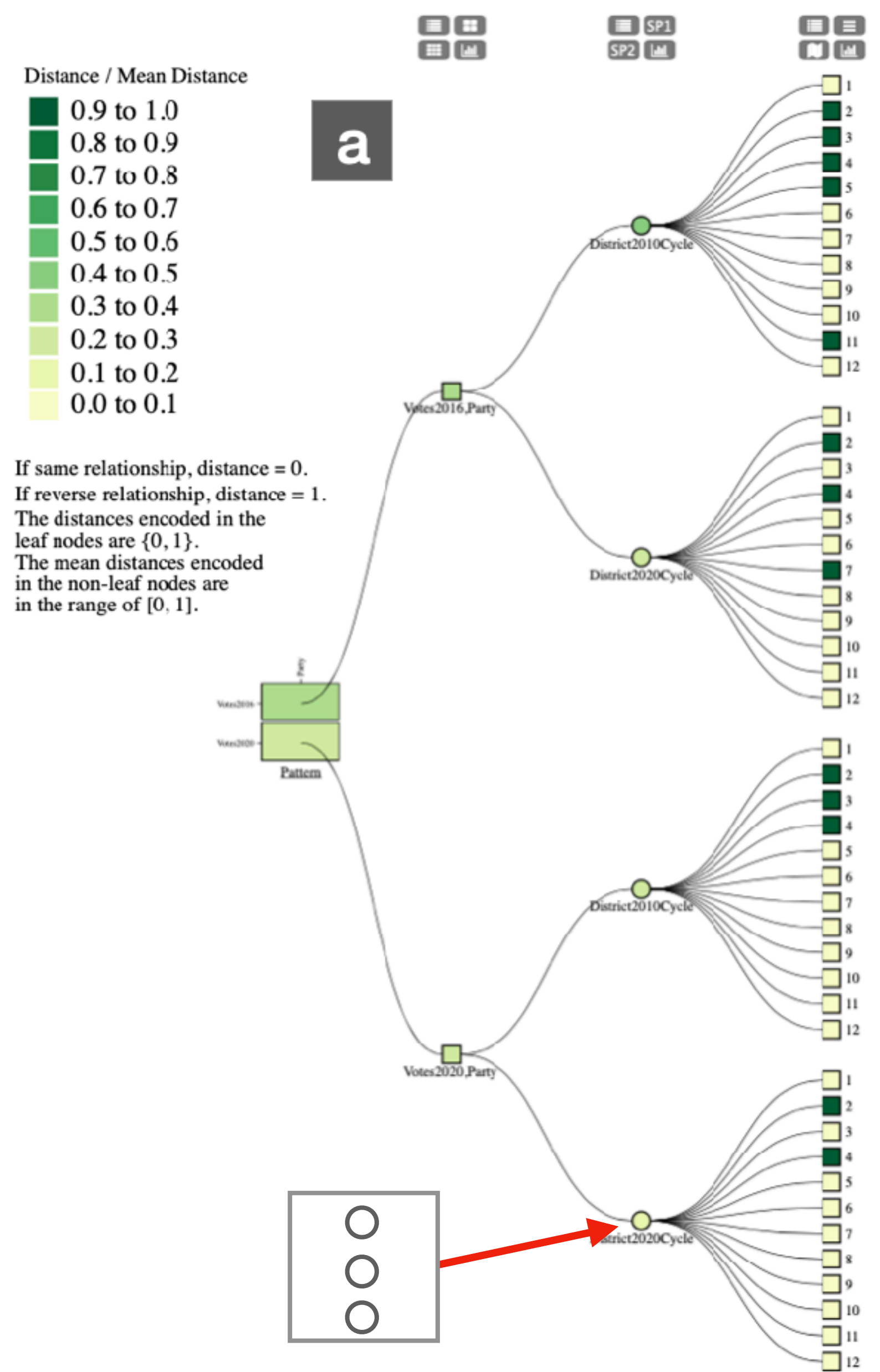
District

Quantitative

Distance (d)

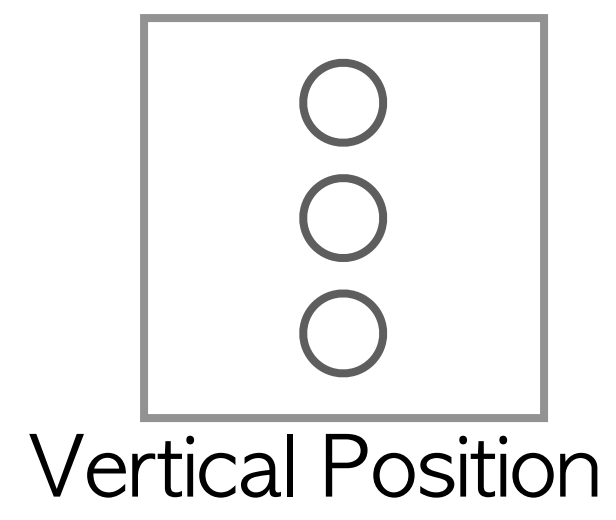
Summary (S)

Representing Data Visually

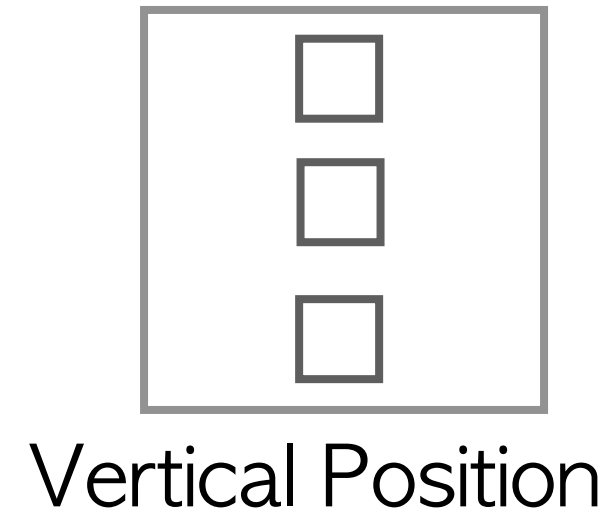


Nominal/ Ordinal

District Cycle



District

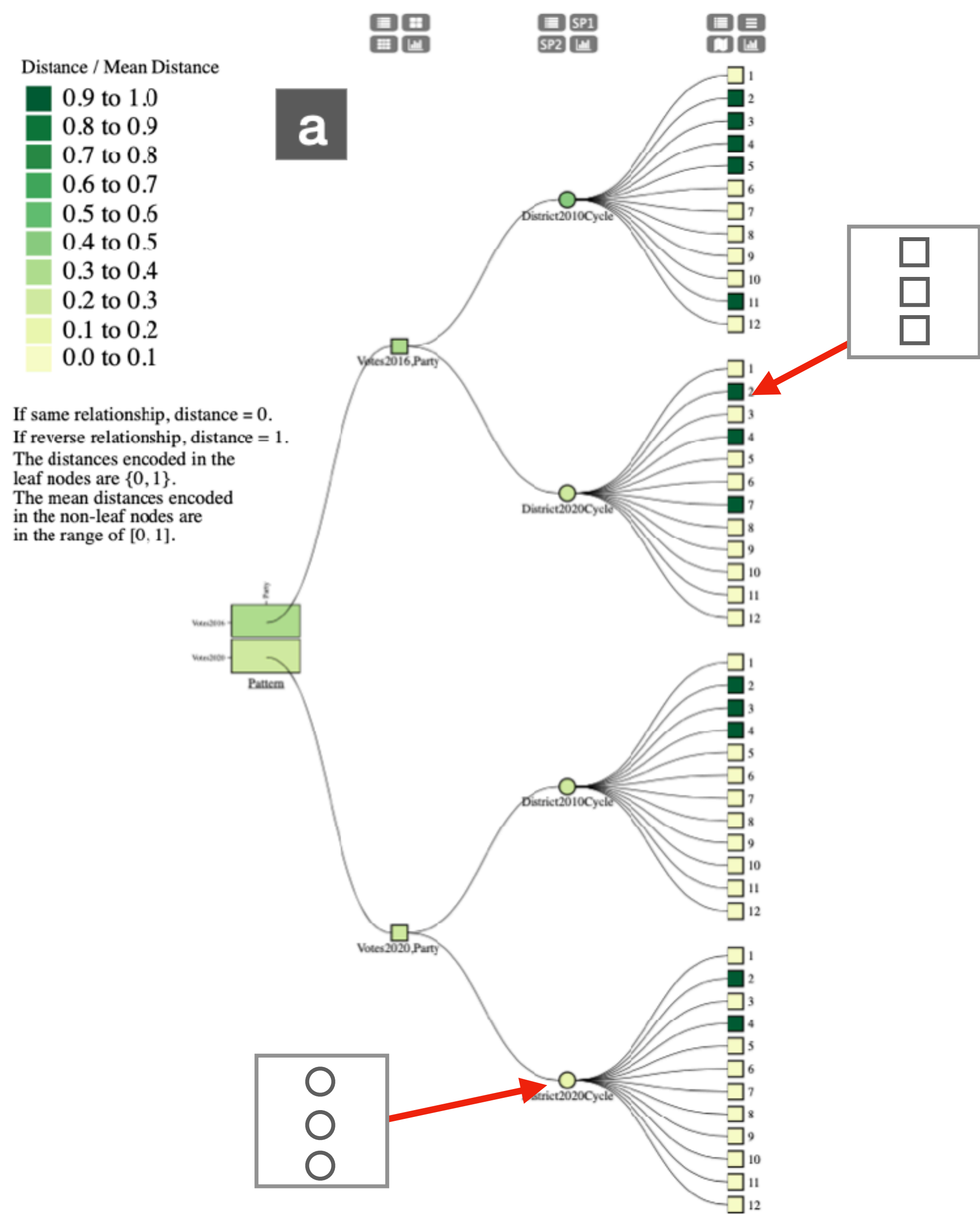


Quantitative

Distance (d)

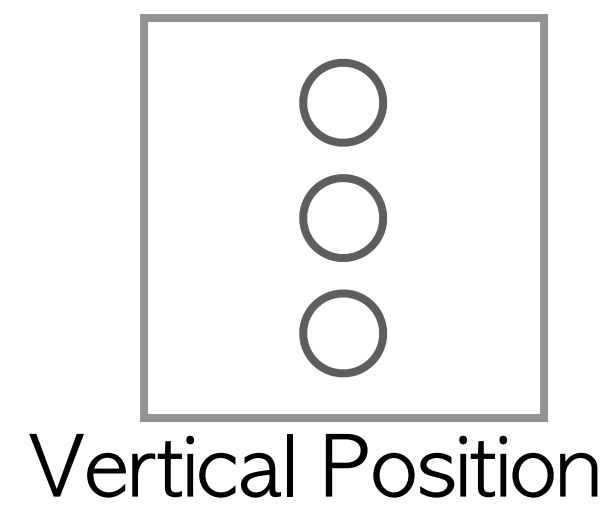
Summary (S)

Representing Data Visually

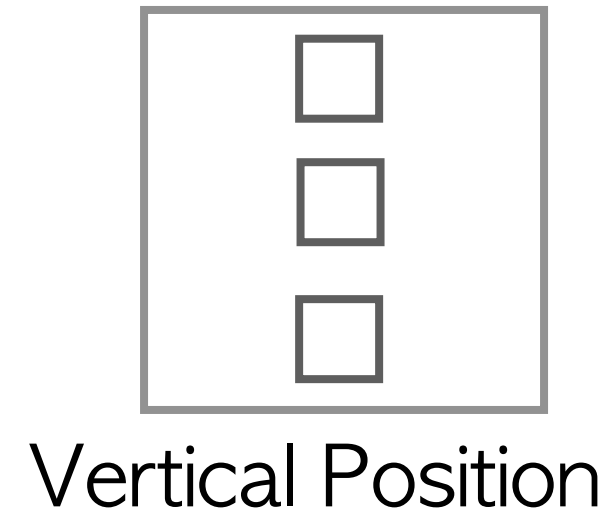


Nominal/ Ordinal

District Cycle



District

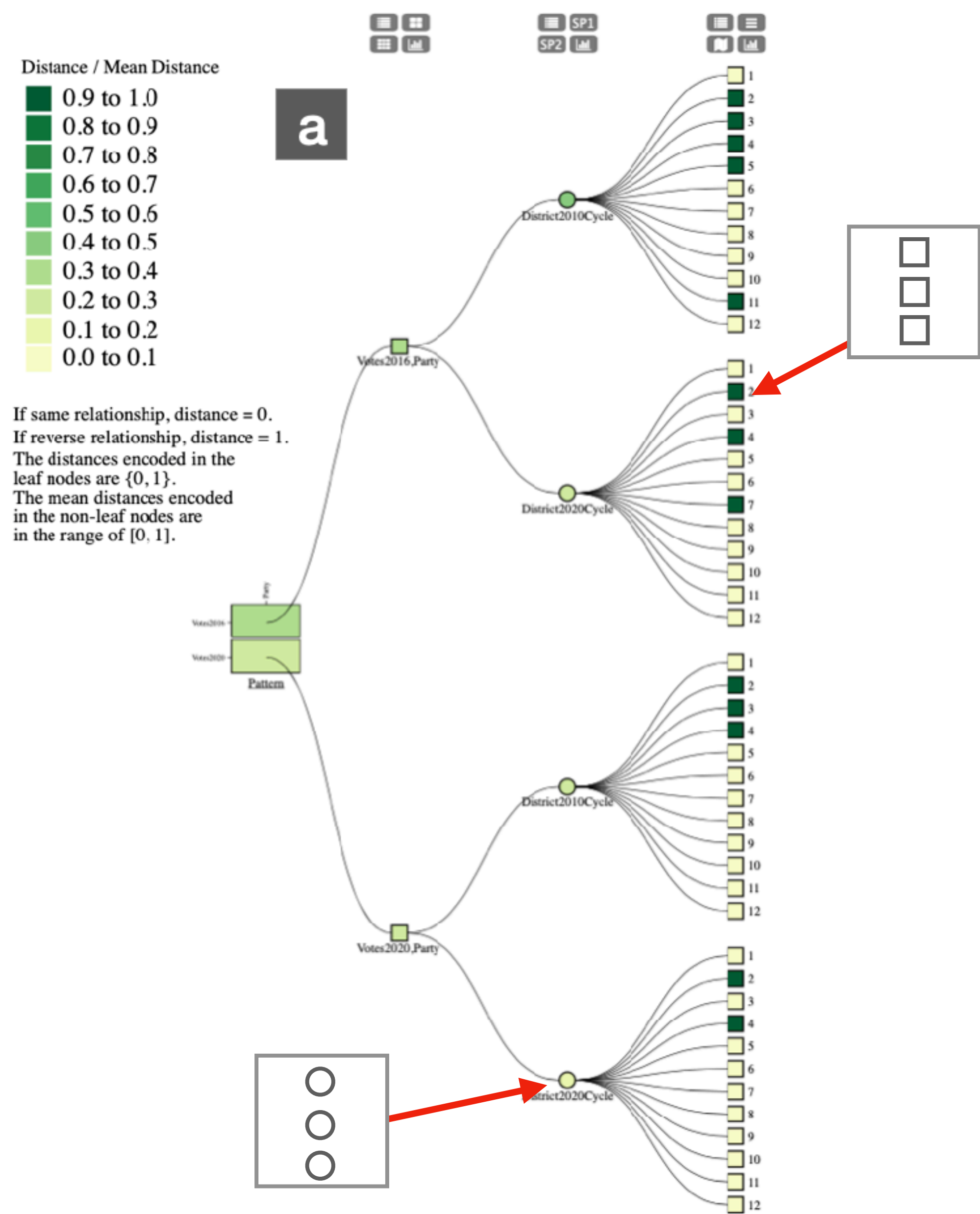


Quantitative

Distance (d)

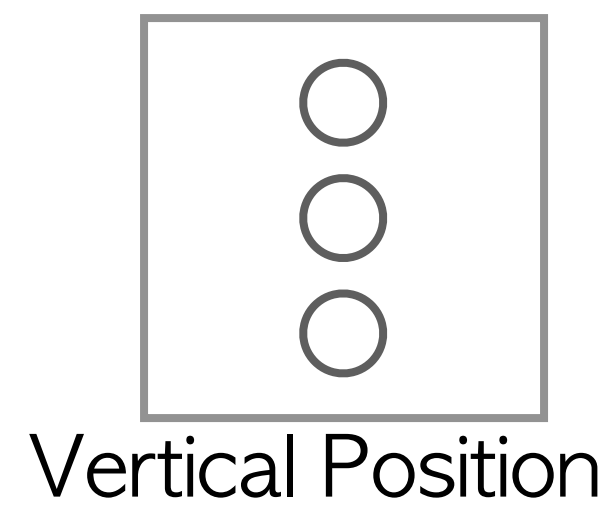
Summary (S)

Representing Data Visually

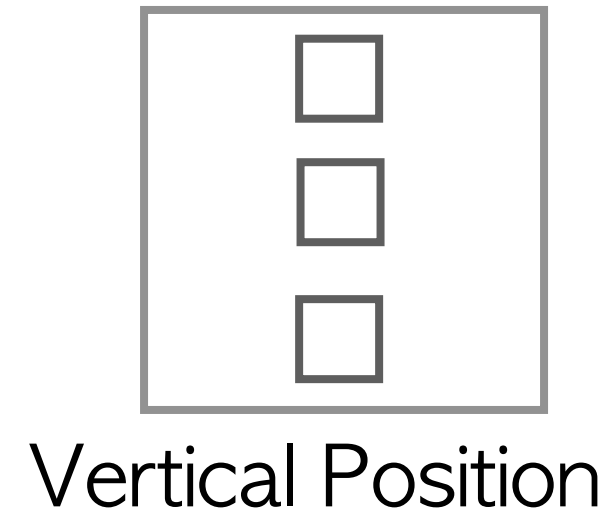


Nominal/ Ordinal

District Cycle

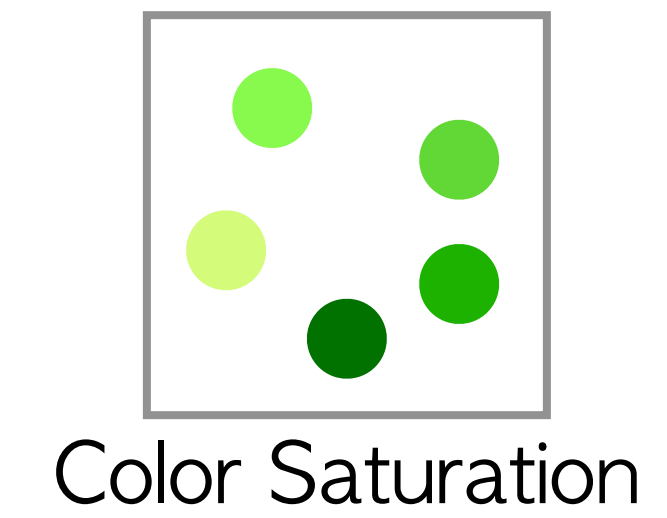


District



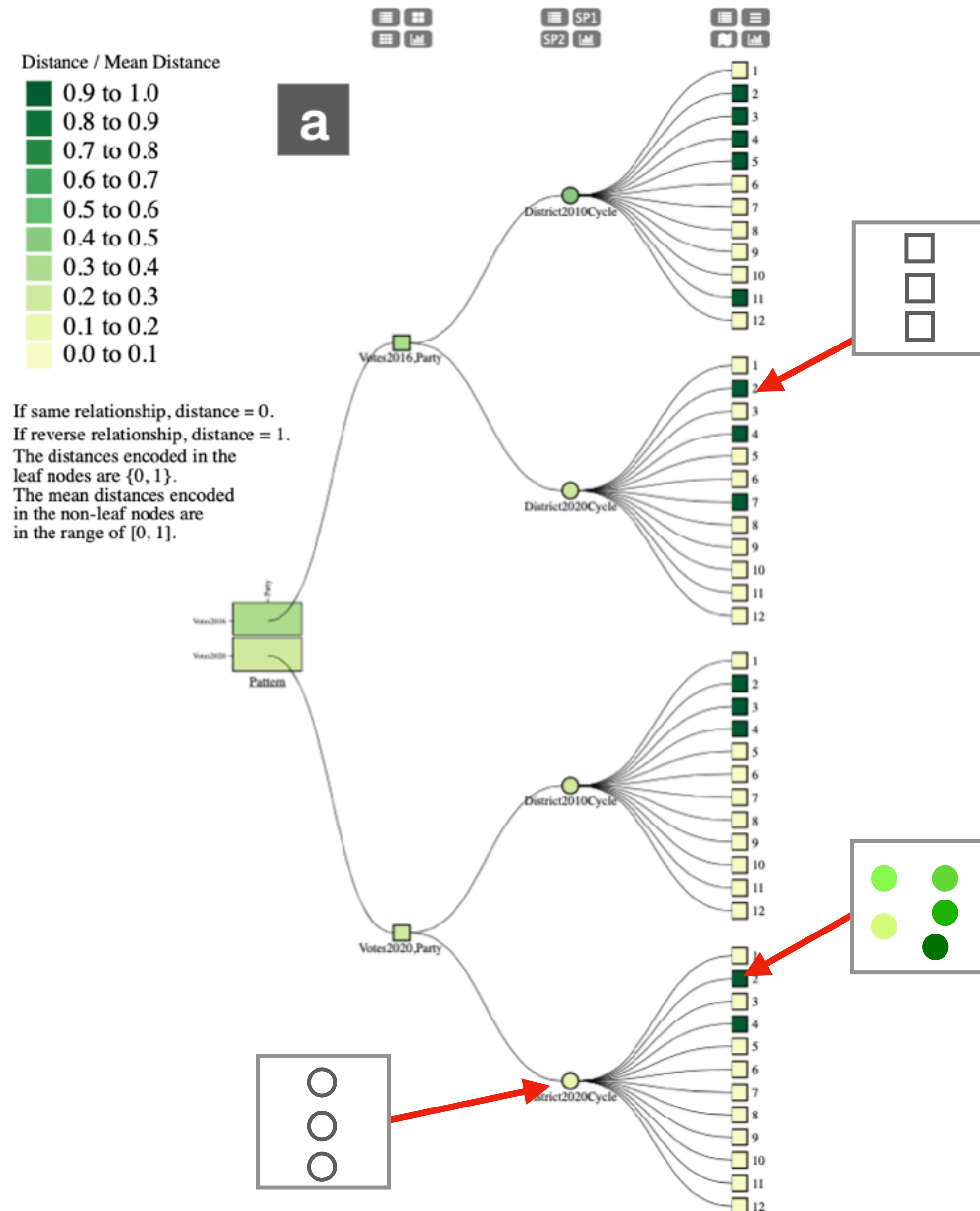
Quantitative

Distance (d)



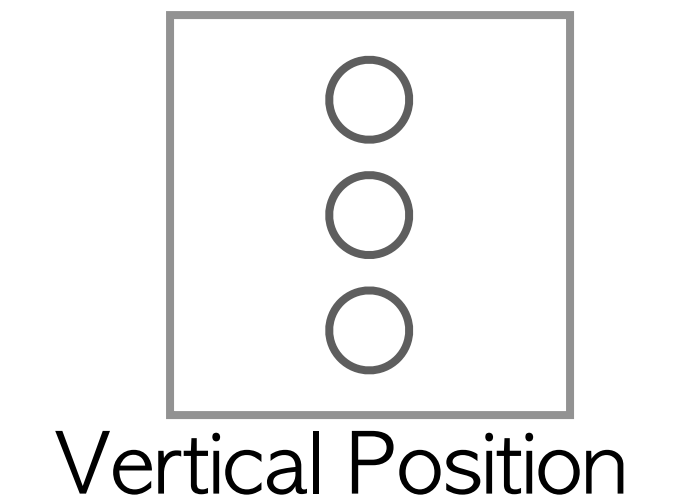
Summary (S)

Representing Data Visually

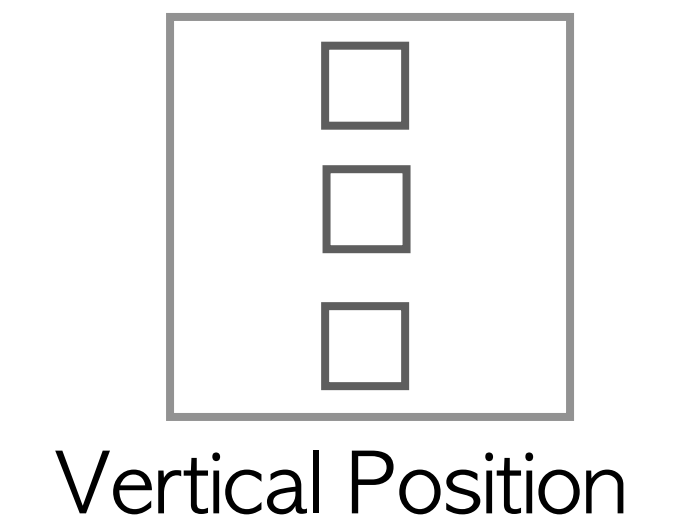


Nominal/ Ordinal

District Cycle

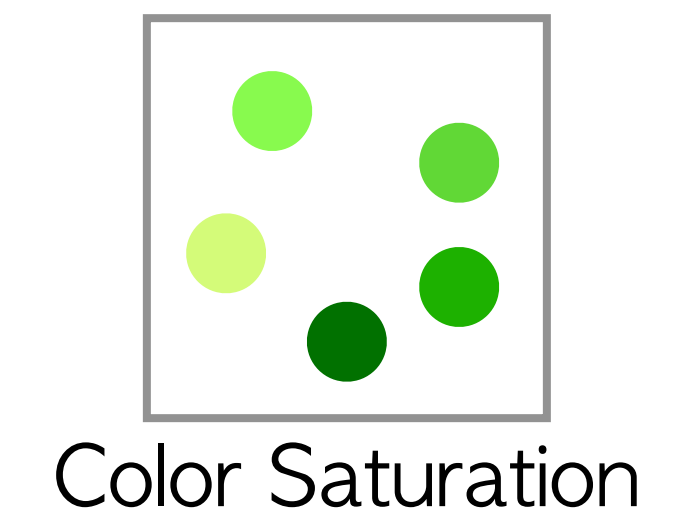


District



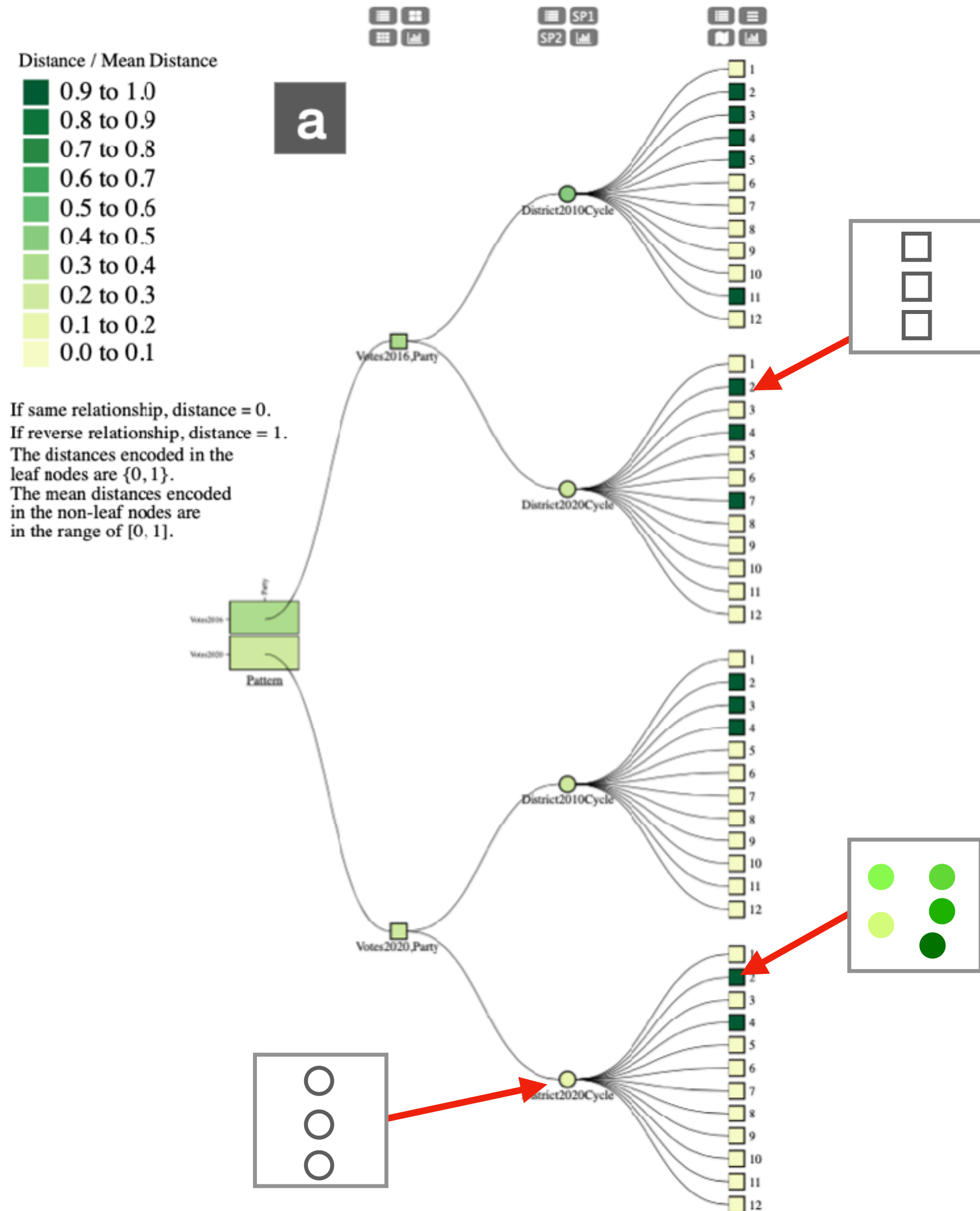
Quantitative

Distance (d)



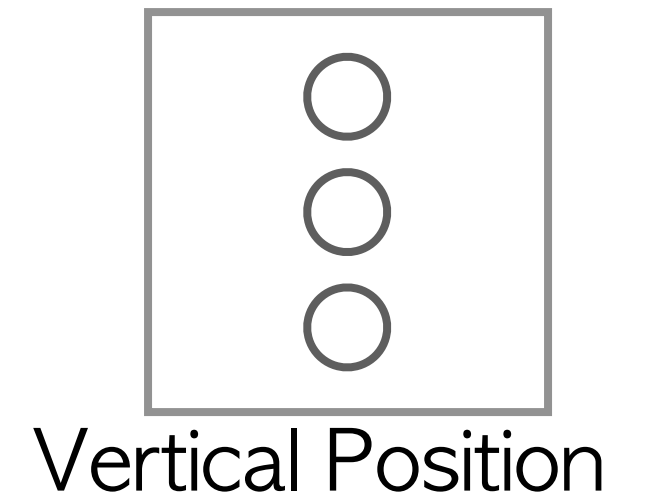
Summary (S)

Representing Data Visually

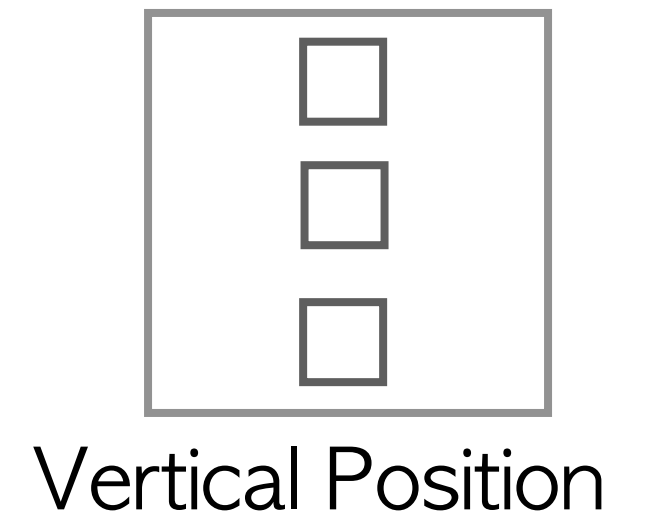


Nominal/ Ordinal

District Cycle

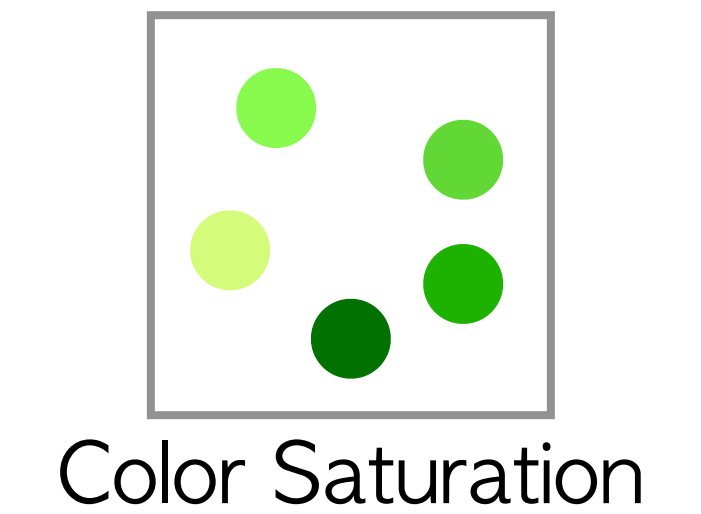


District

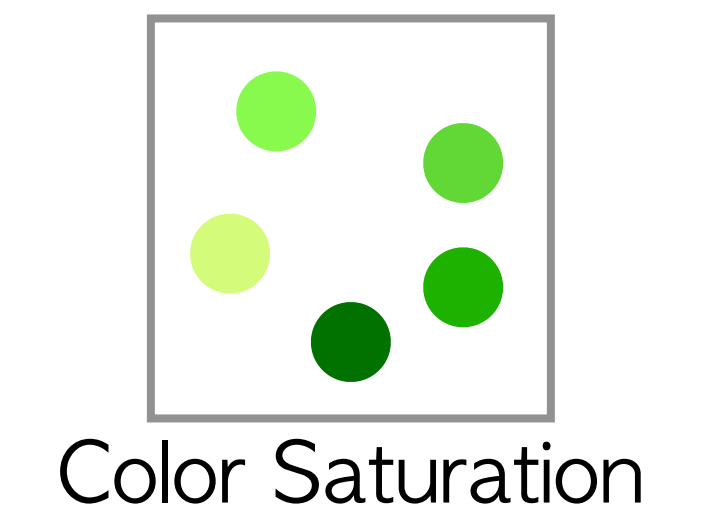


Quantitative

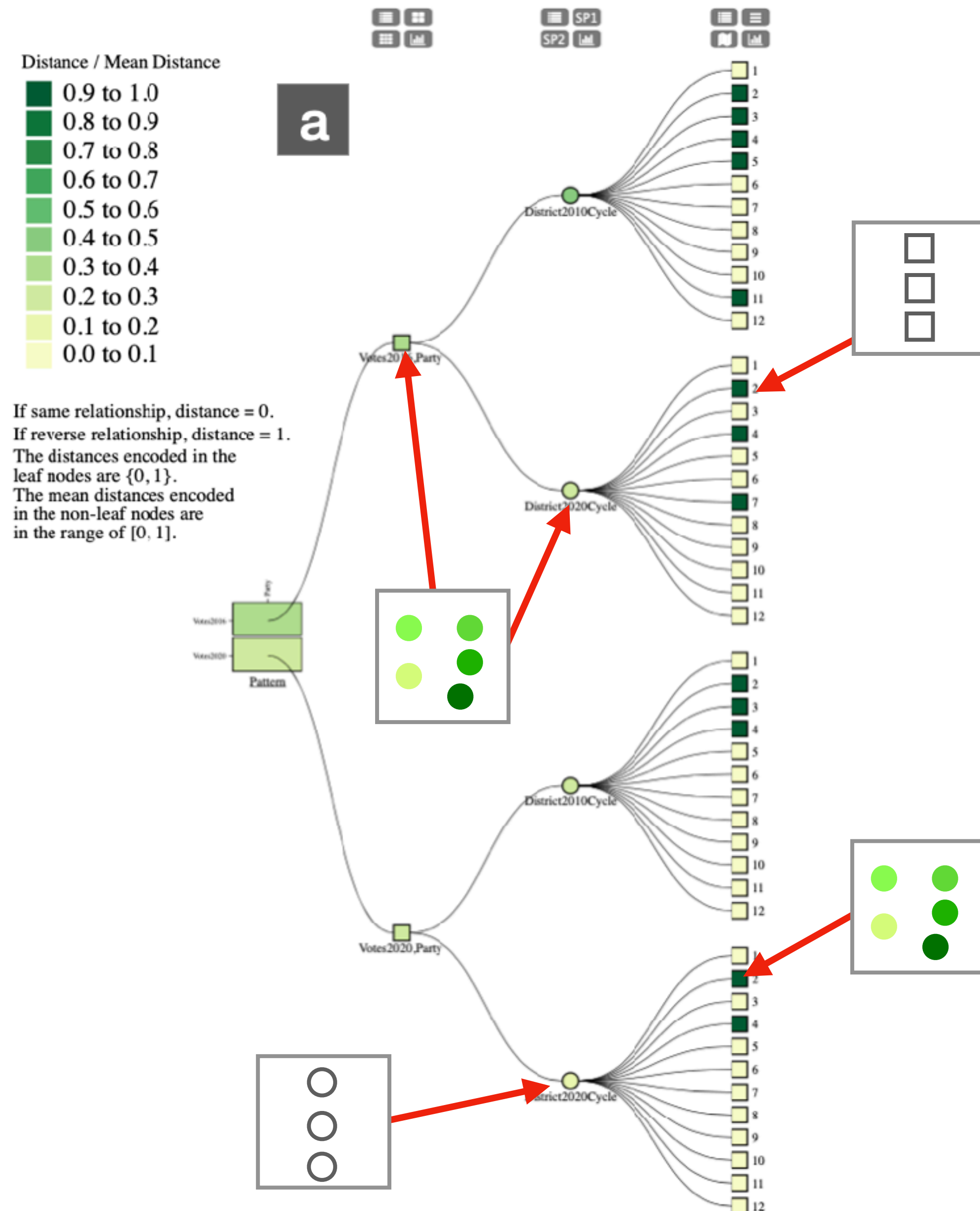
Distance (d)



Summary (S)

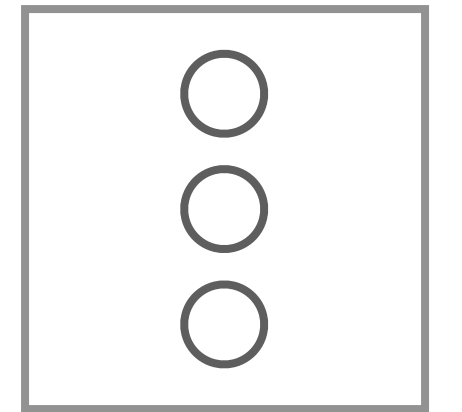


Representing Data Visually



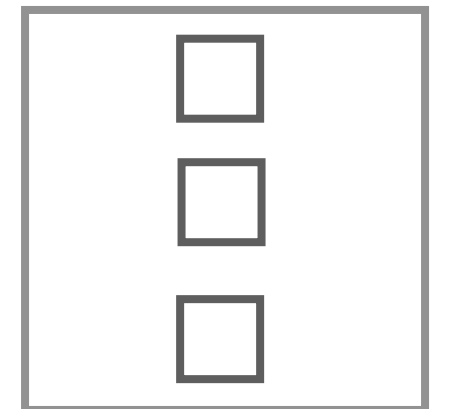
Nominal/ Ordinal

District Cycle



Vertical Position

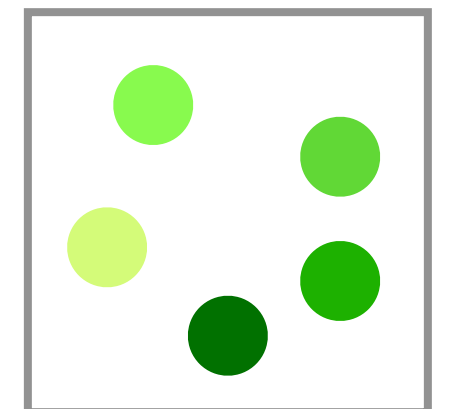
District



Vertical Position

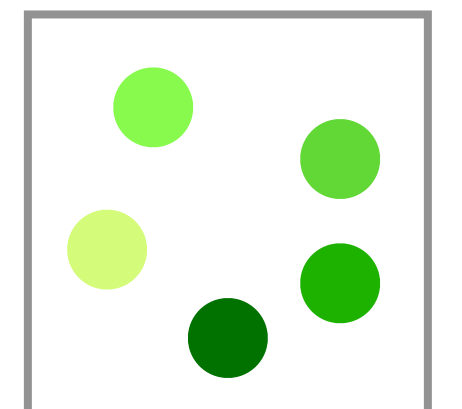
Quantitative

Distance (d)



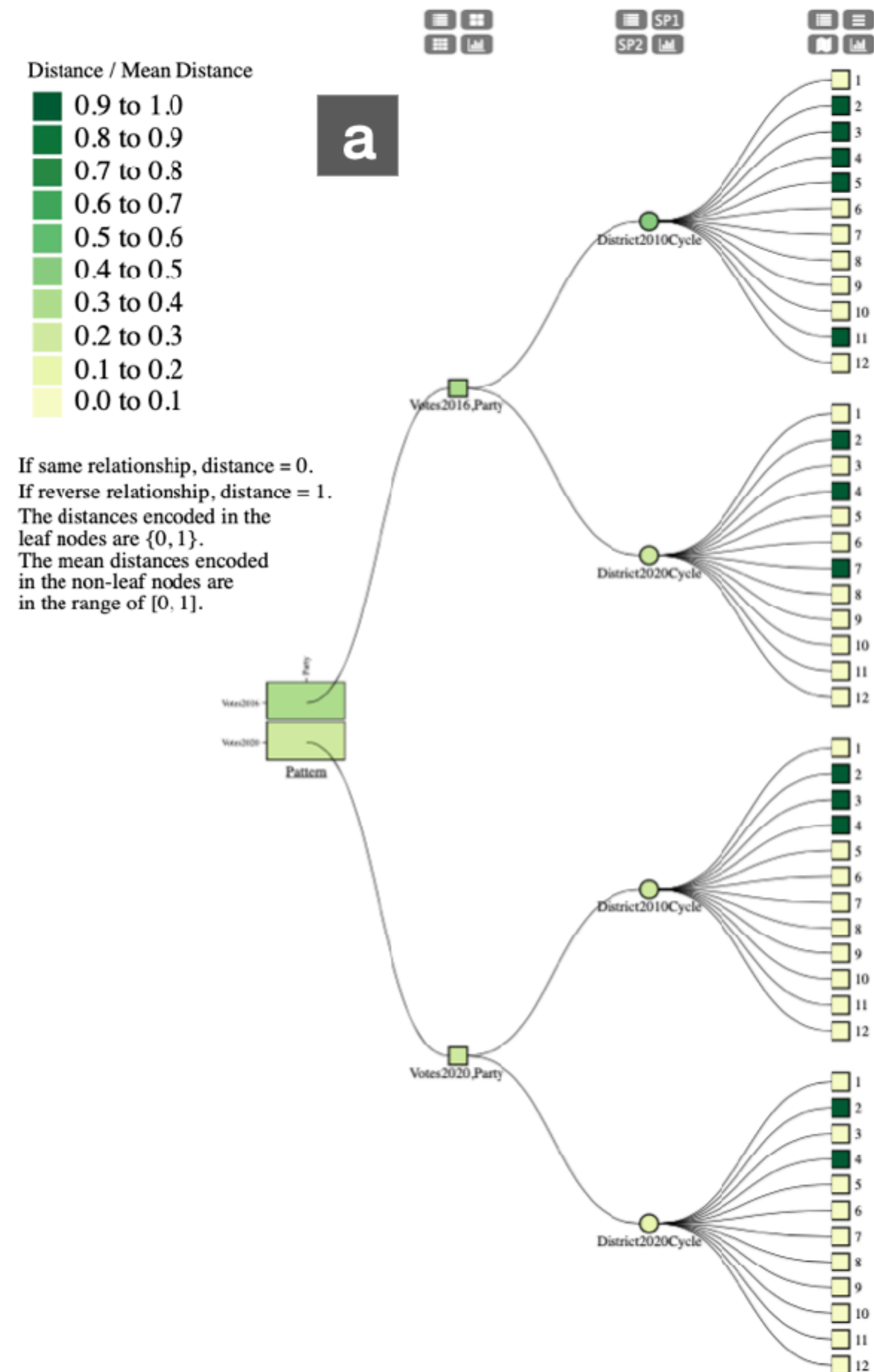
Color Saturation

Summary (S)

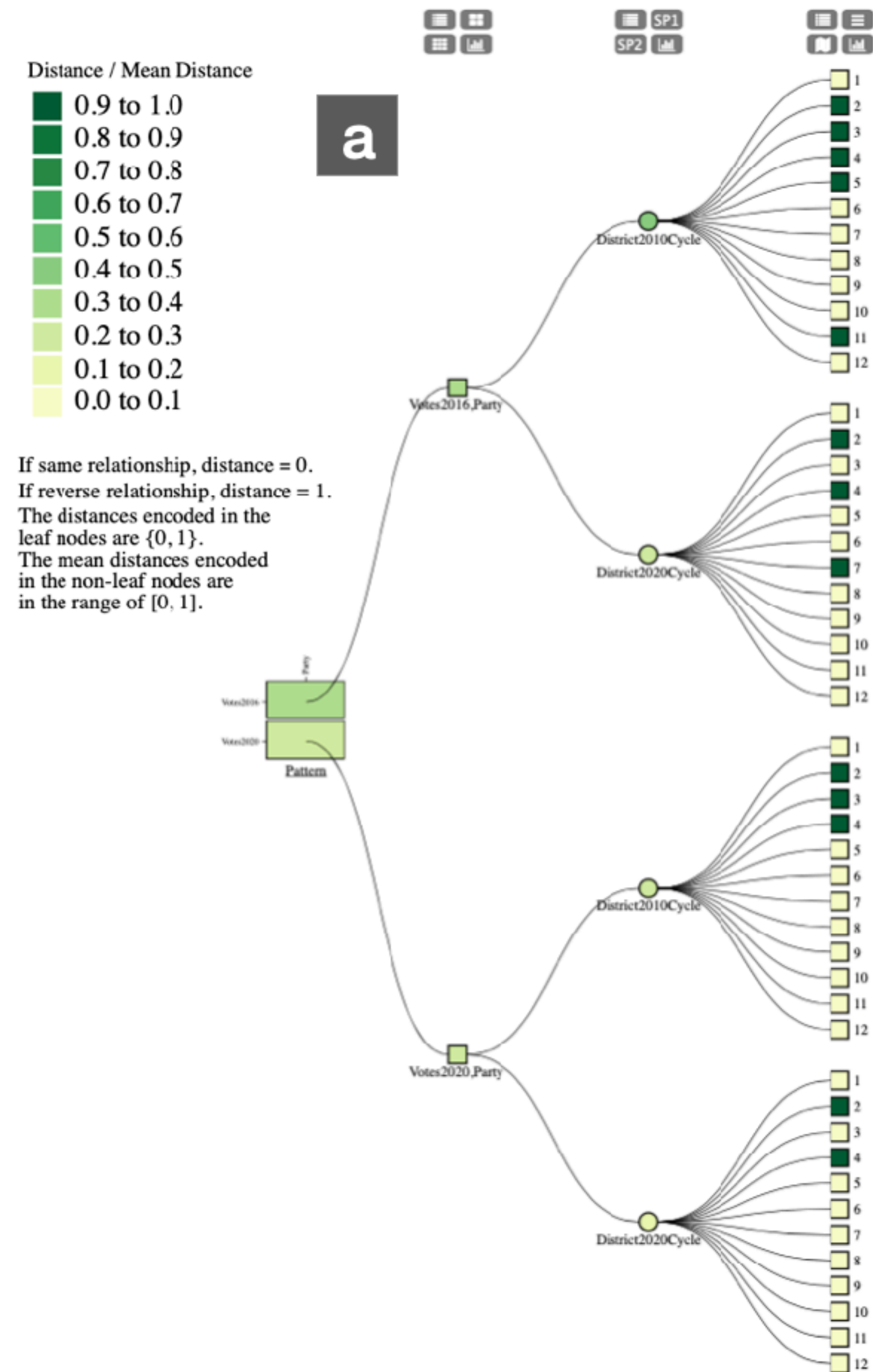


Color Saturation

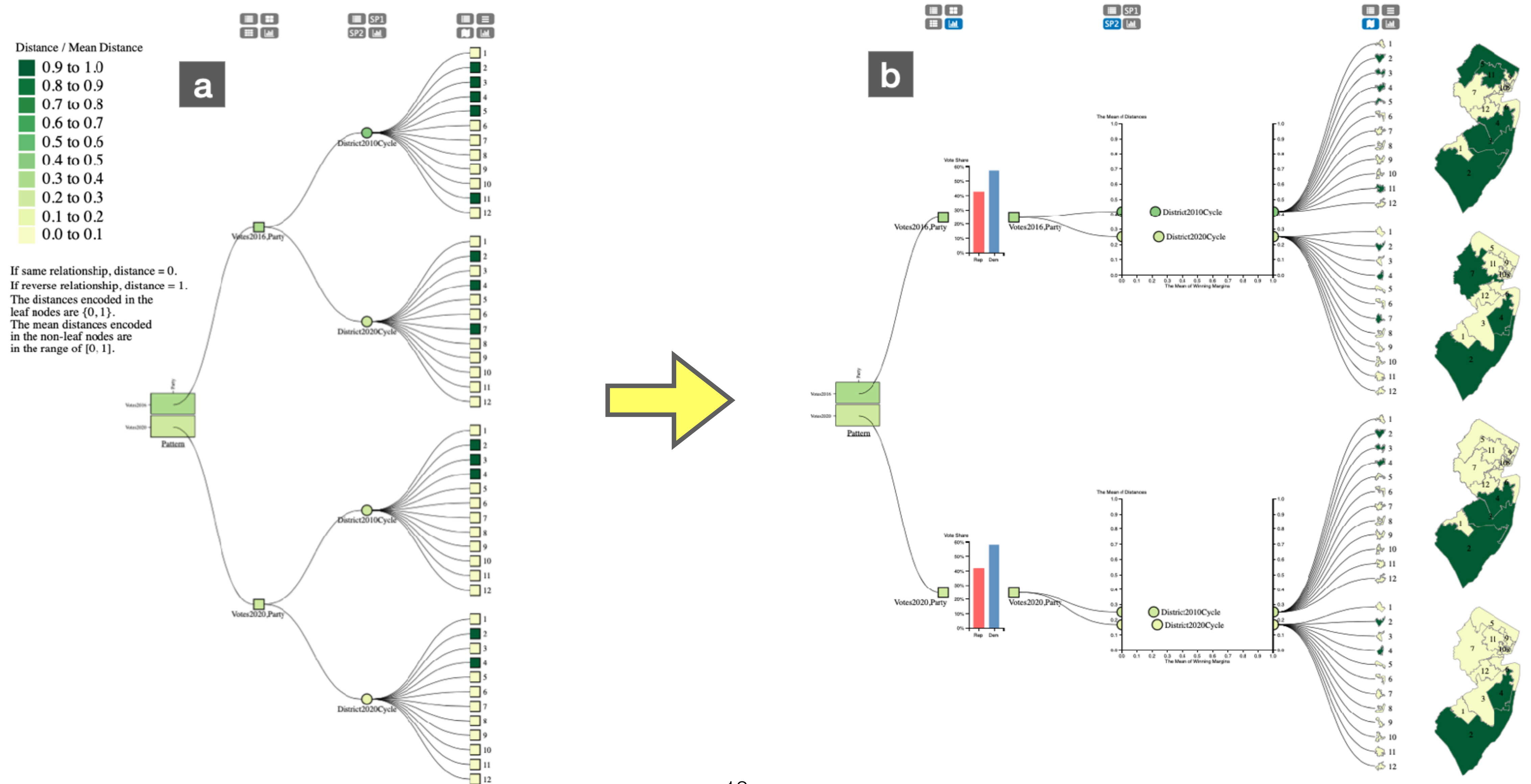
Embed views in tree Visualization



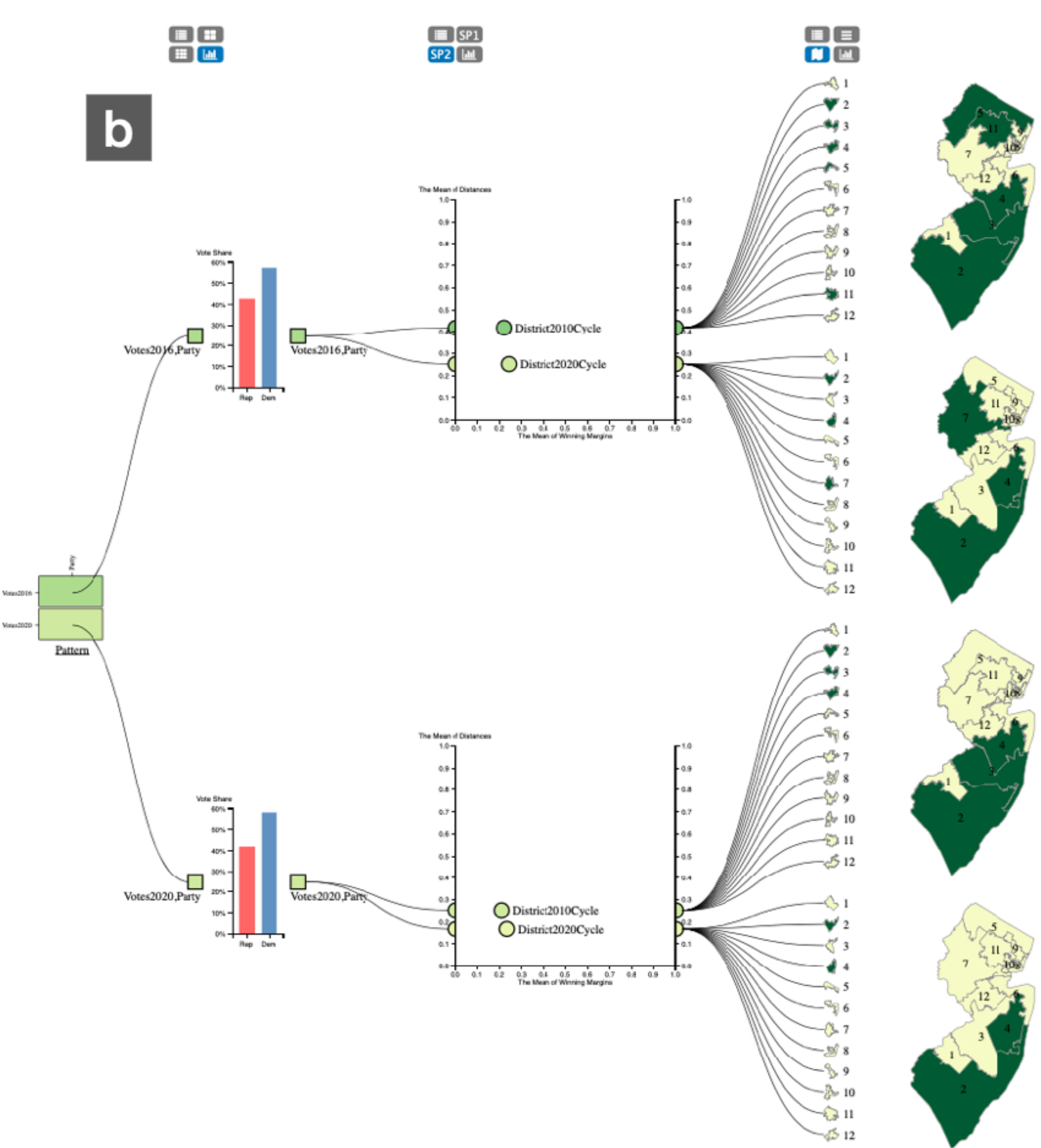
Embed views in tree Visualization



Embed views in tree Visualization



Representing Data Visually



Nominal/ Ordinal

District Cycle

District

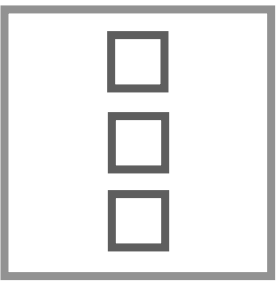
Party

Quantitative

Distance (d)

Summary (S)

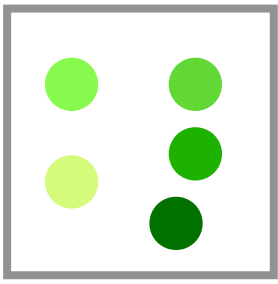
Voting Share % (VSP)



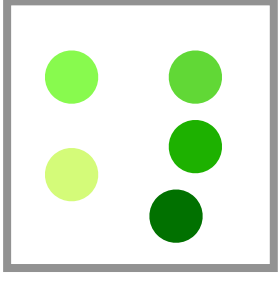
Vertical Position



Vertical Position

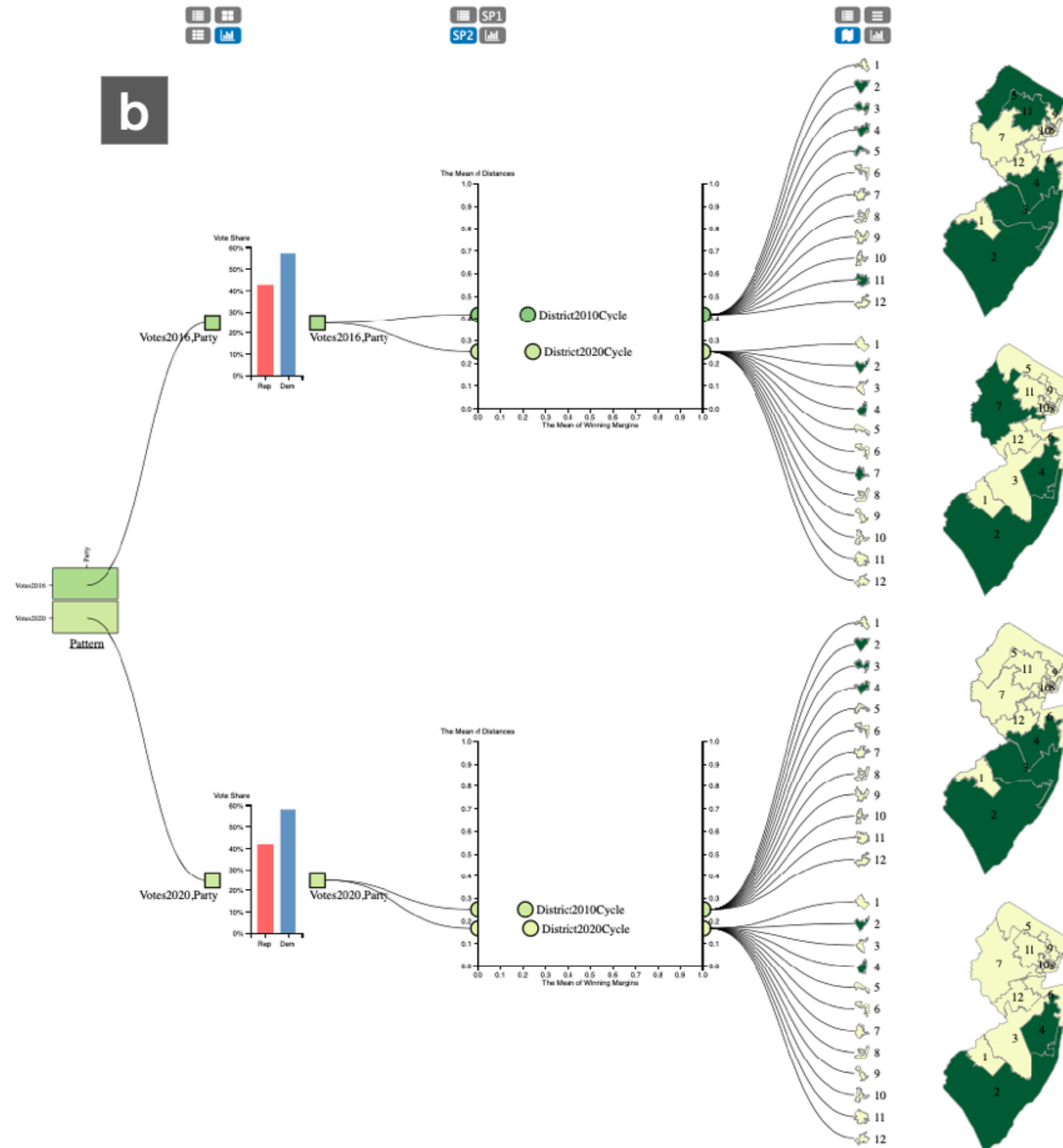


Color Saturation



Color Saturation

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

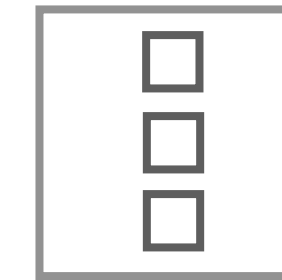
Party

Quantitative

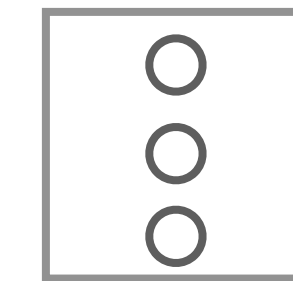
Distance (d)

Summary (S)

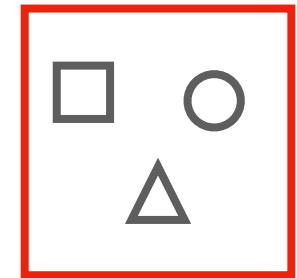
Voting Share % (VSP)



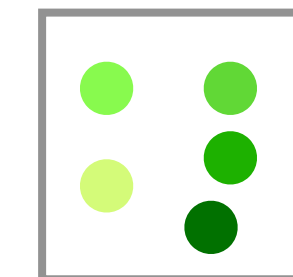
Vertical Position



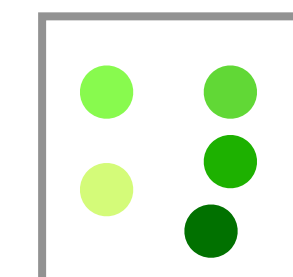
Vertical Position



Shape

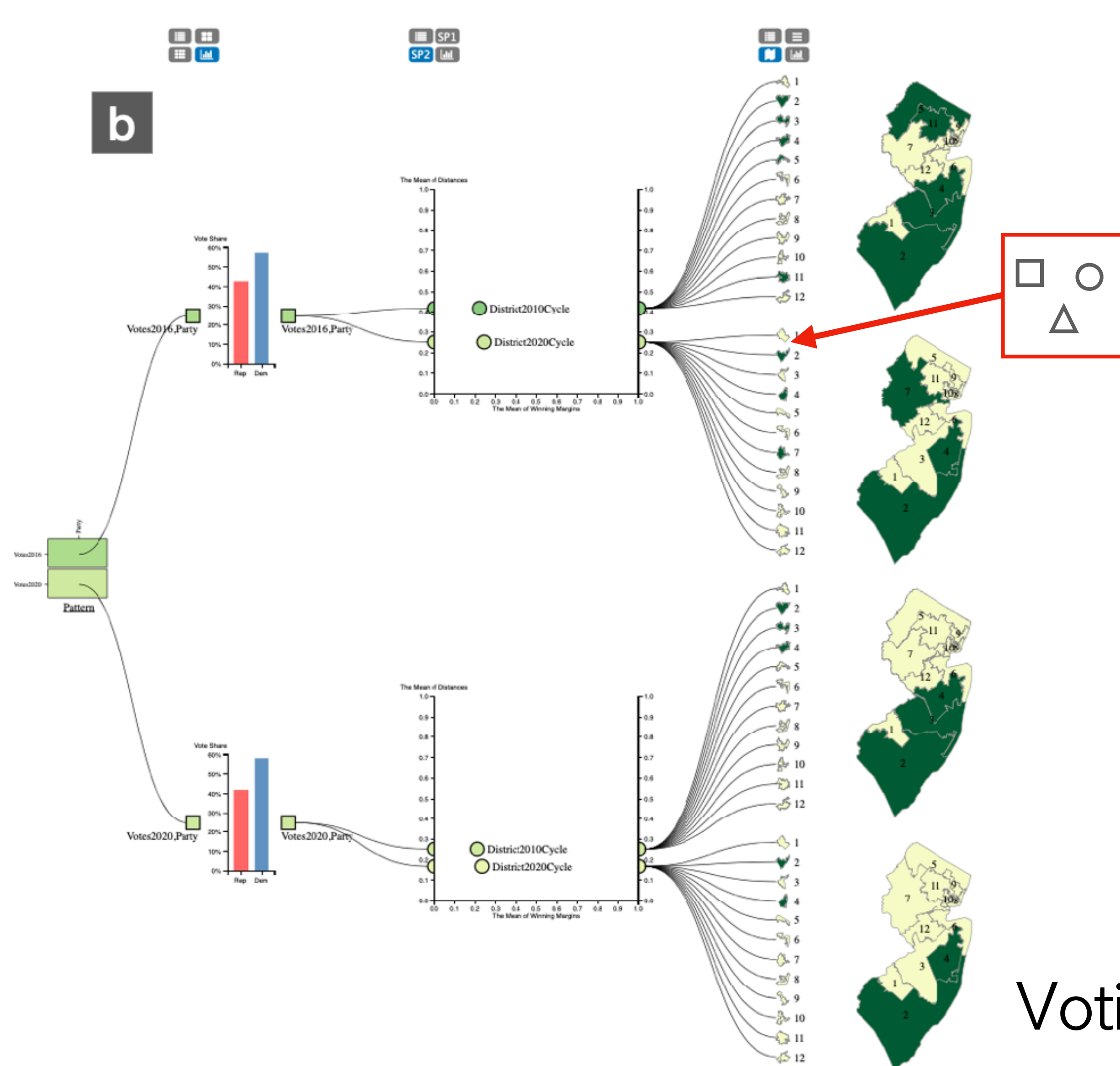


Color Saturation



Color Saturation

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

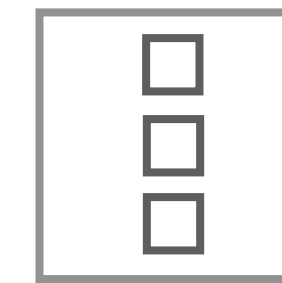
Party

Quantitative

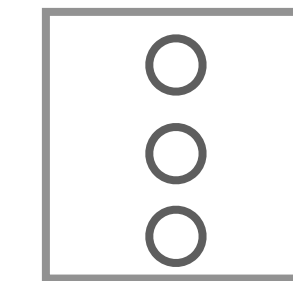
Distance (d)

Summary (S)

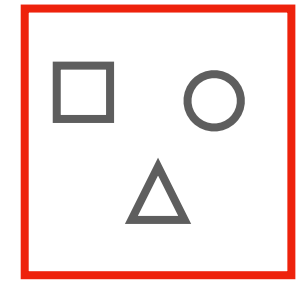
Voting Share % (VSP)



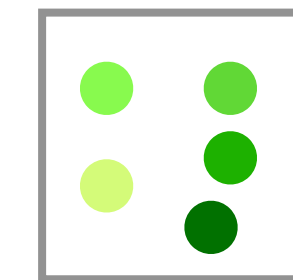
Vertical Position



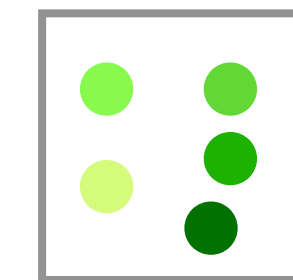
Vertical Position



Shape

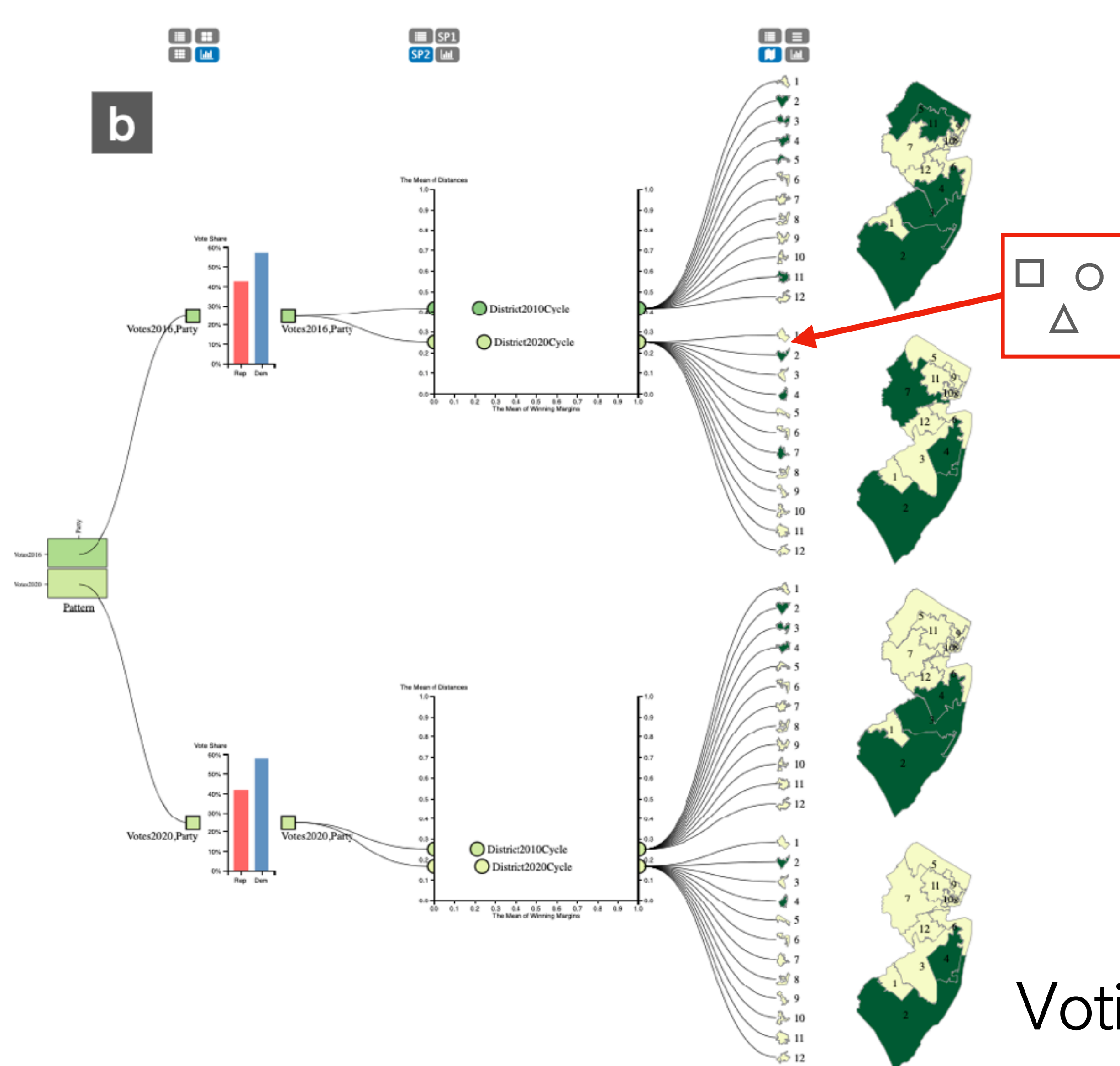


Color Saturation



Color Saturation

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

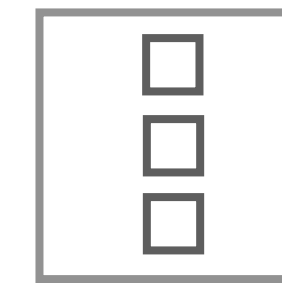
Party

Quantitative

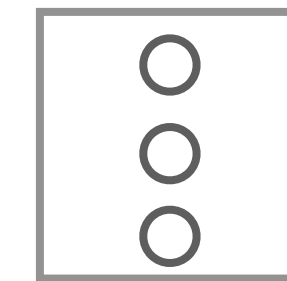
Distance (d)

Summary (S)

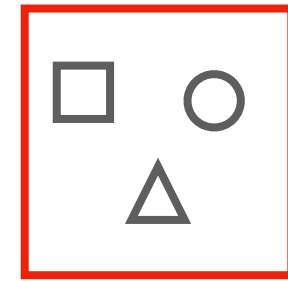
Voting Share % (VSP)



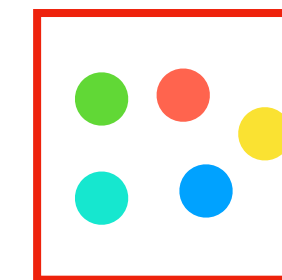
Vertical Position



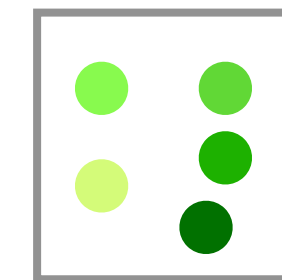
Vertical Position



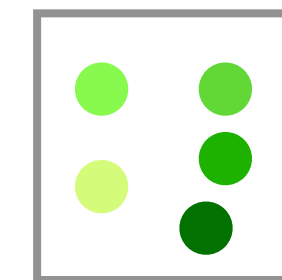
Shape



Color Hue

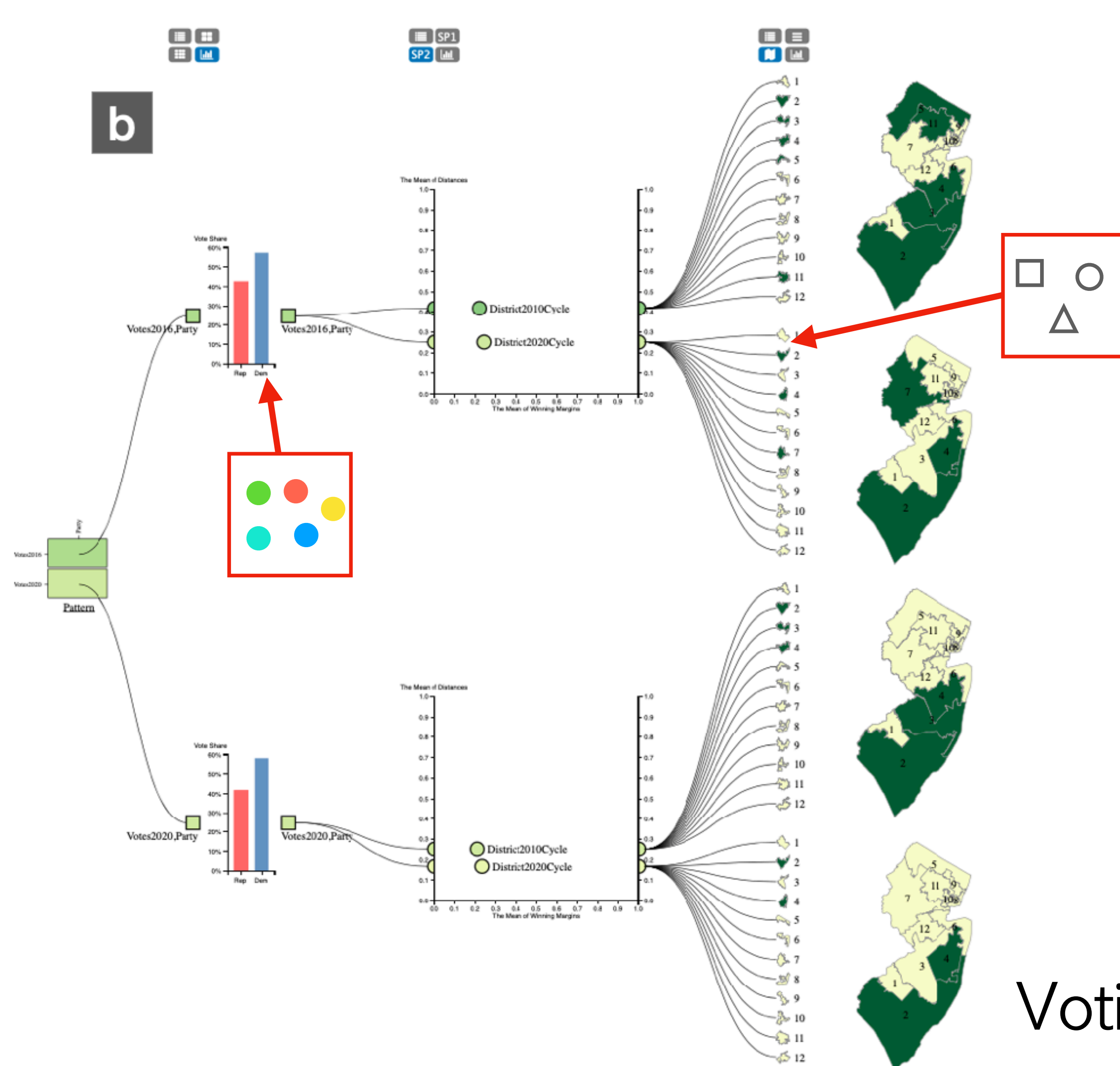


Color Saturation



Color Saturation

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

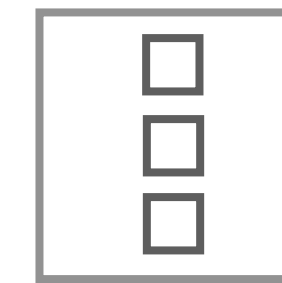
Party

Quantitative

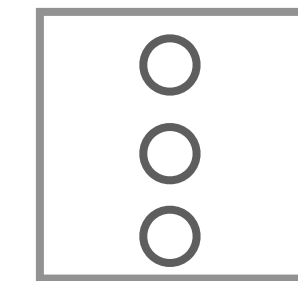
Distance (d)

Summary (S)

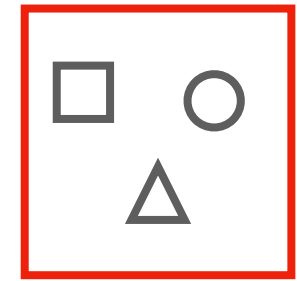
Voting Share % (VSP)



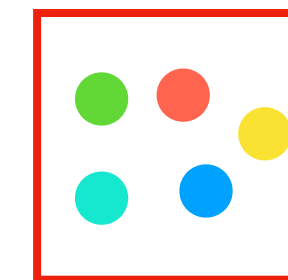
Vertical Position



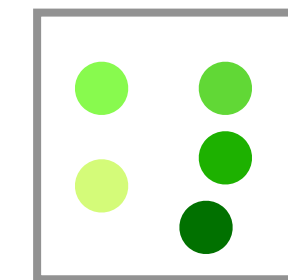
Vertical Position



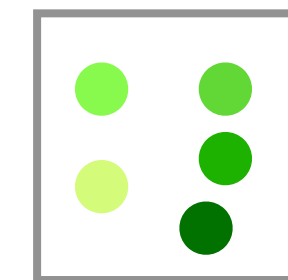
Shape



Color Hue

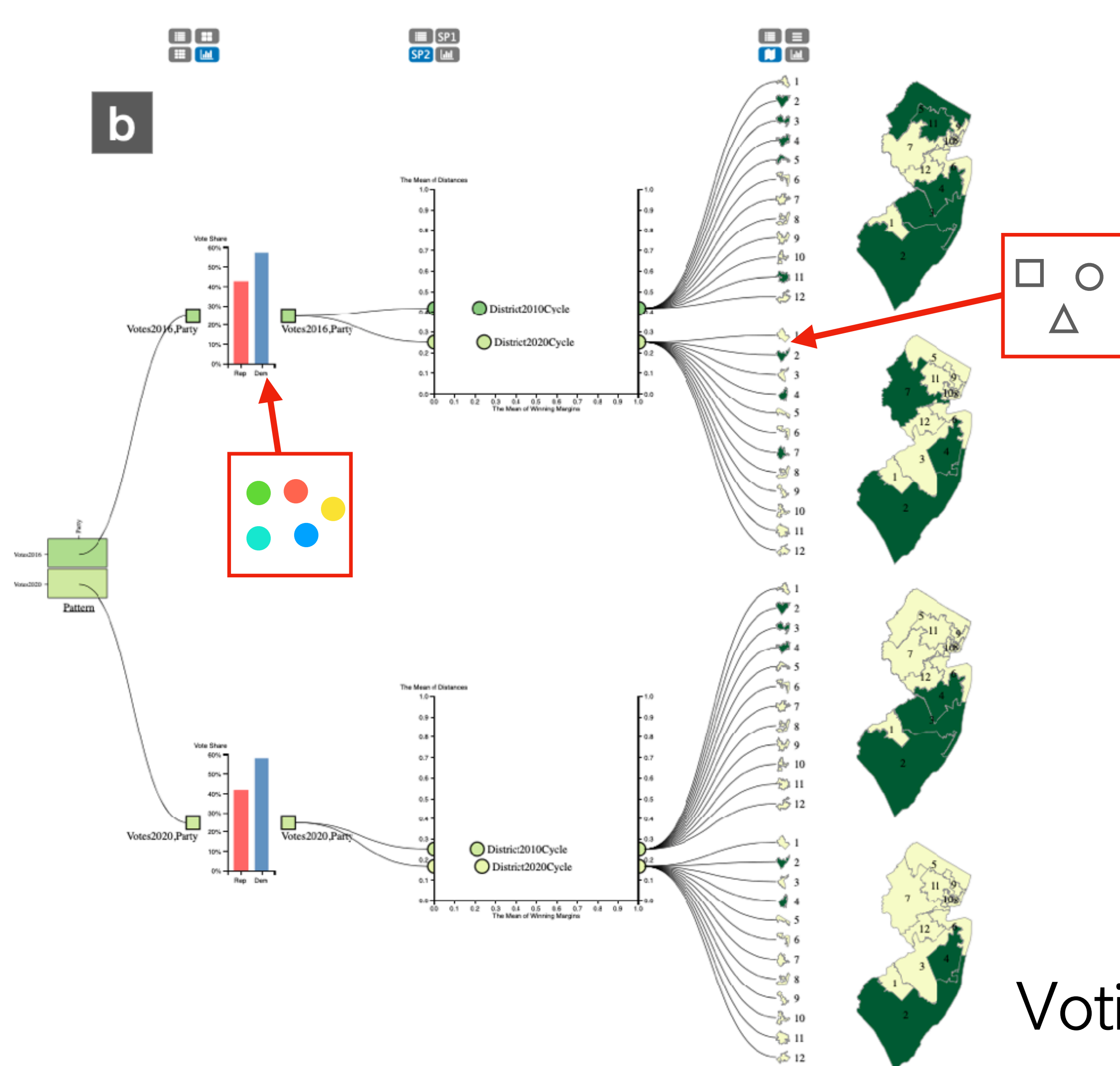


Color Saturation



Color Saturation

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

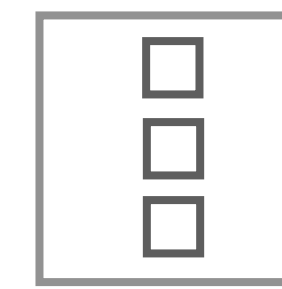
Party

Quantitative

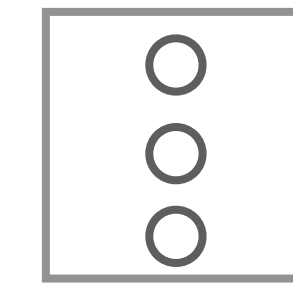
Distance (d)

Summary (S)

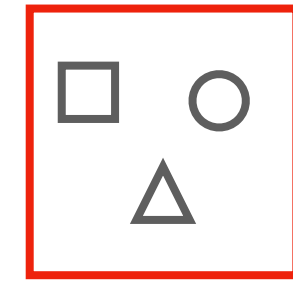
Voting Share % (VSP)



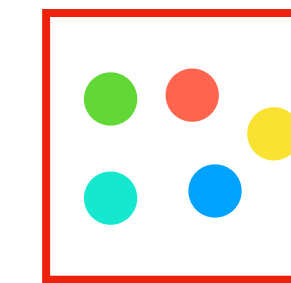
Vertical Position



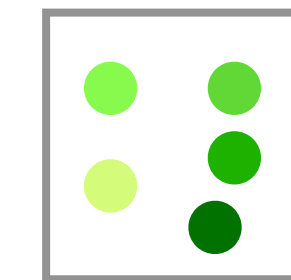
Vertical Position



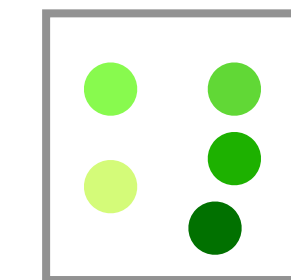
Shape



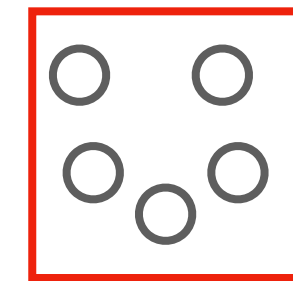
Color Hue



Color Saturation

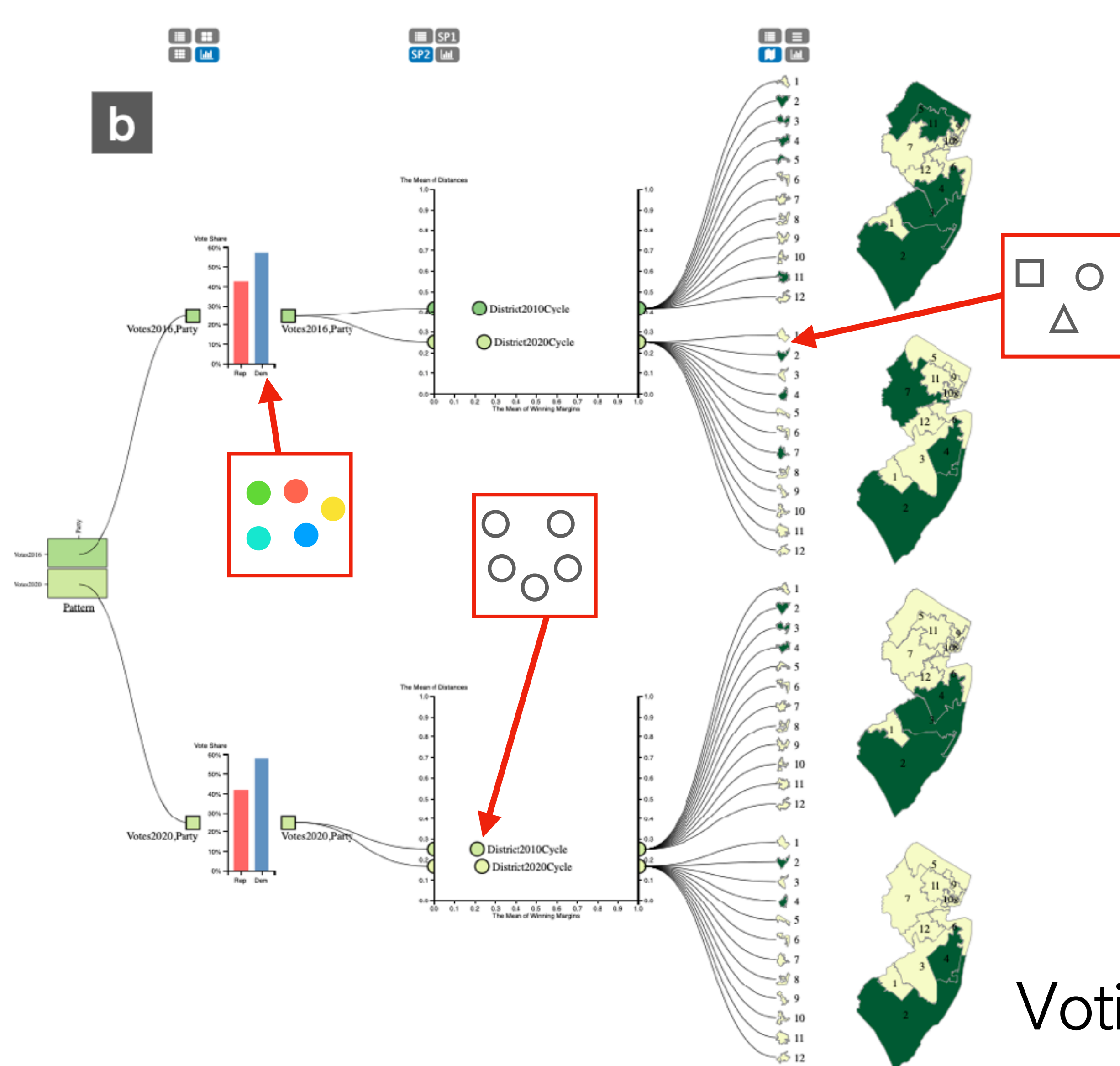


Color Saturation



Position

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

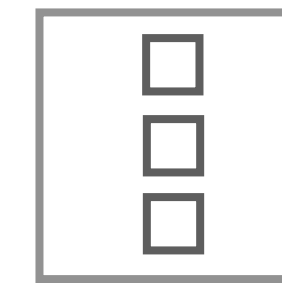
Party

Quantitative

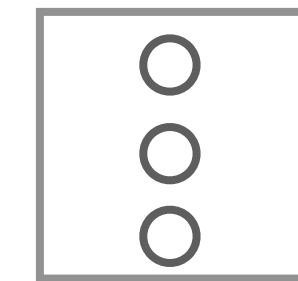
Distance (d)

Summary (S)

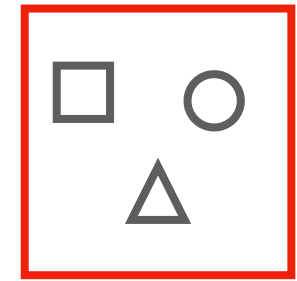
Voting Share % (VSP)



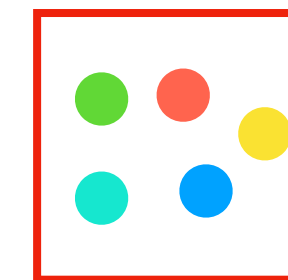
Vertical Position



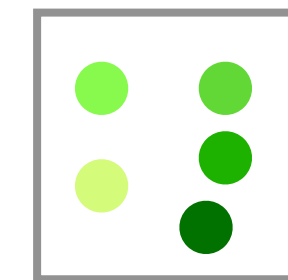
Vertical Position



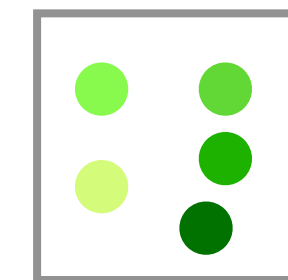
Shape



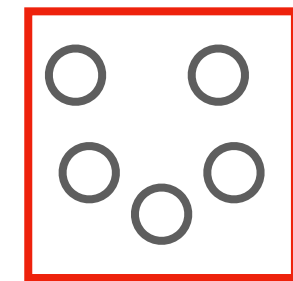
Color Hue



Color Saturation

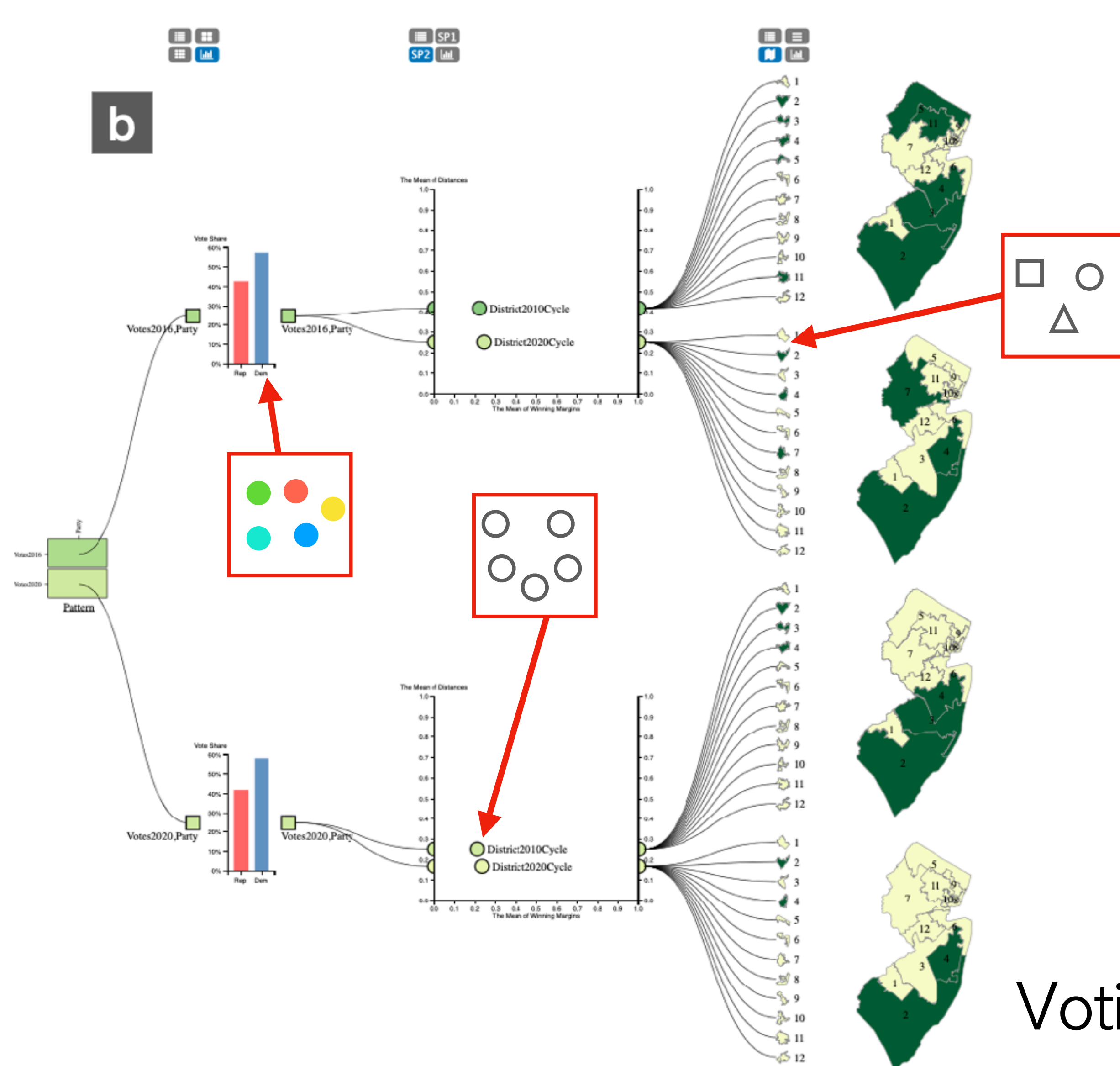


Color Saturation



Position

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

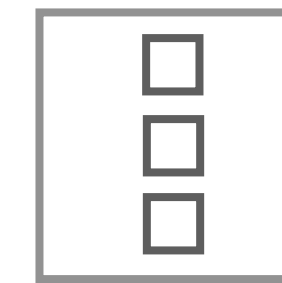
Party

Quantitative

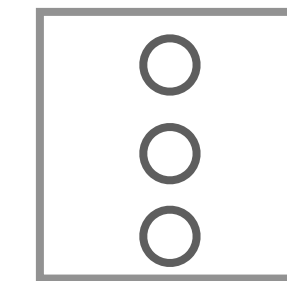
Distance (d)

Summary (S)

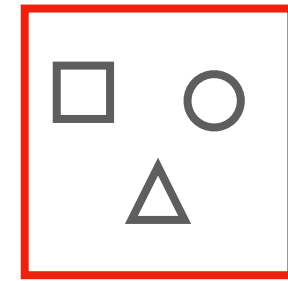
Voting Share % (VSP)



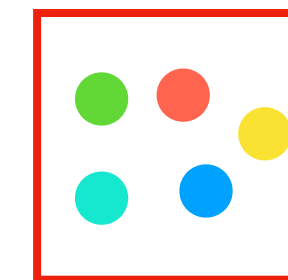
Vertical Position



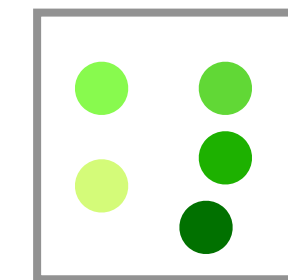
Vertical Position



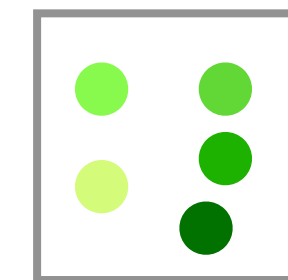
Shape



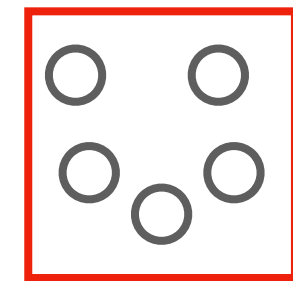
Color Hue



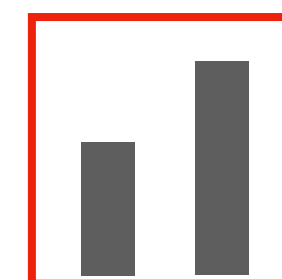
Color Saturation



Color Saturation

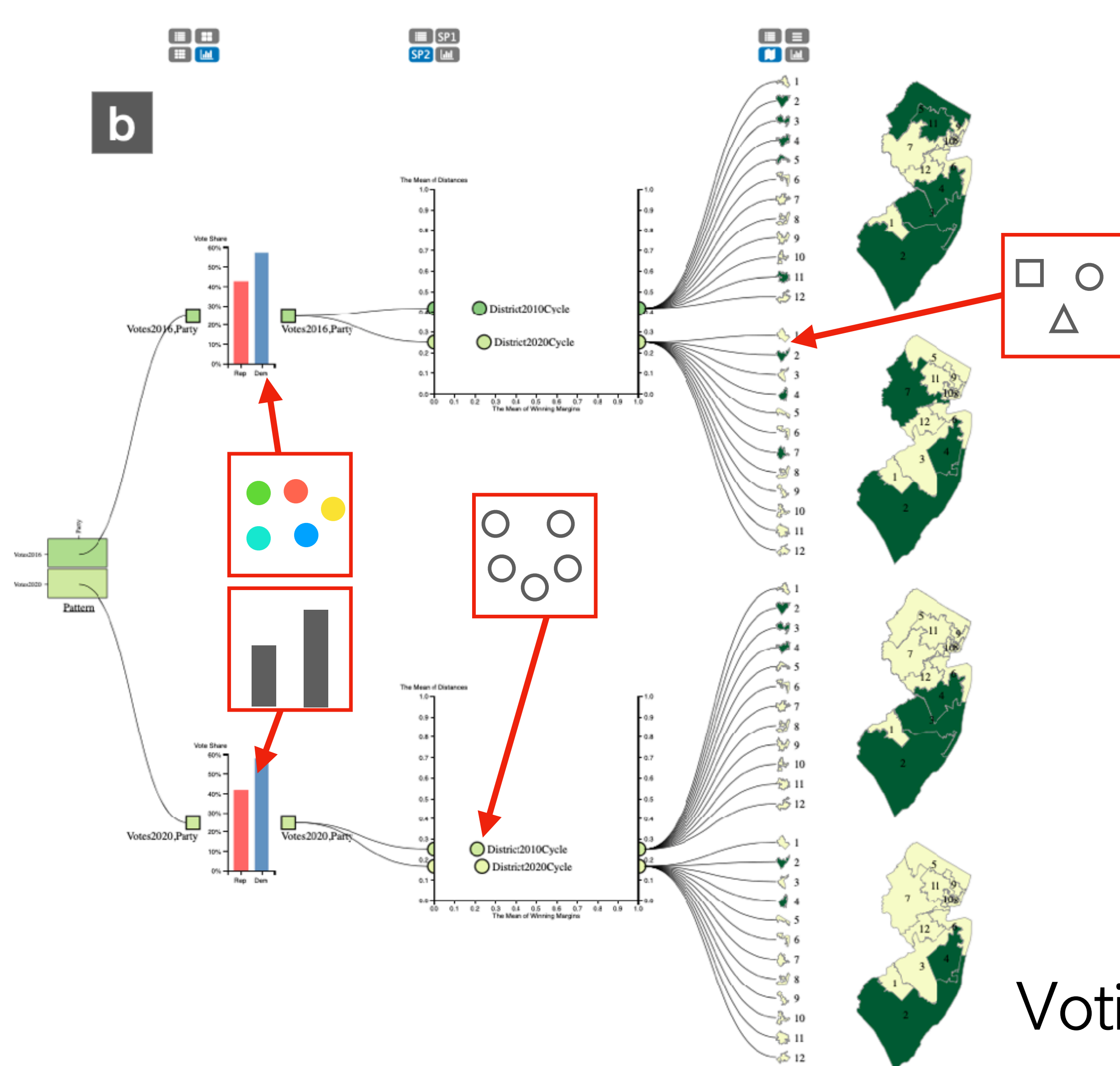


Position



Length

Representing Data Visually



Nominal/ Ordinal

District Cycle

District

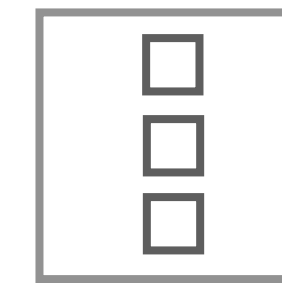
Party

Quantitative

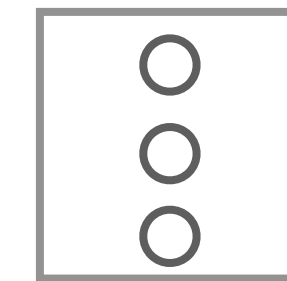
Distance (d)

Summary (S)

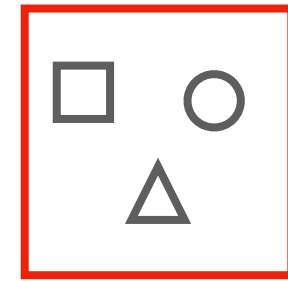
Voting Share % (VSP)



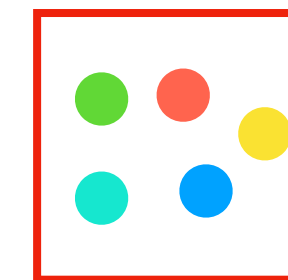
Vertical Position



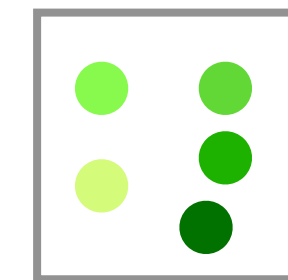
Vertical Position



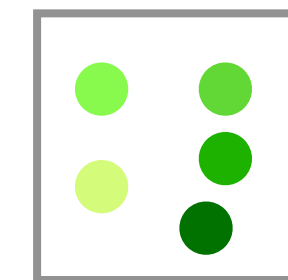
Shape



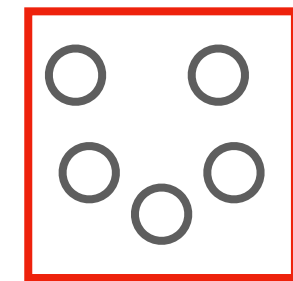
Color Hue



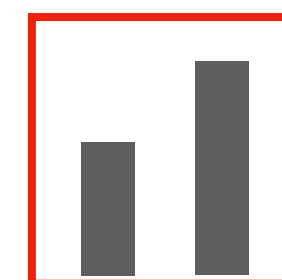
Color Saturation



Color Saturation



Position

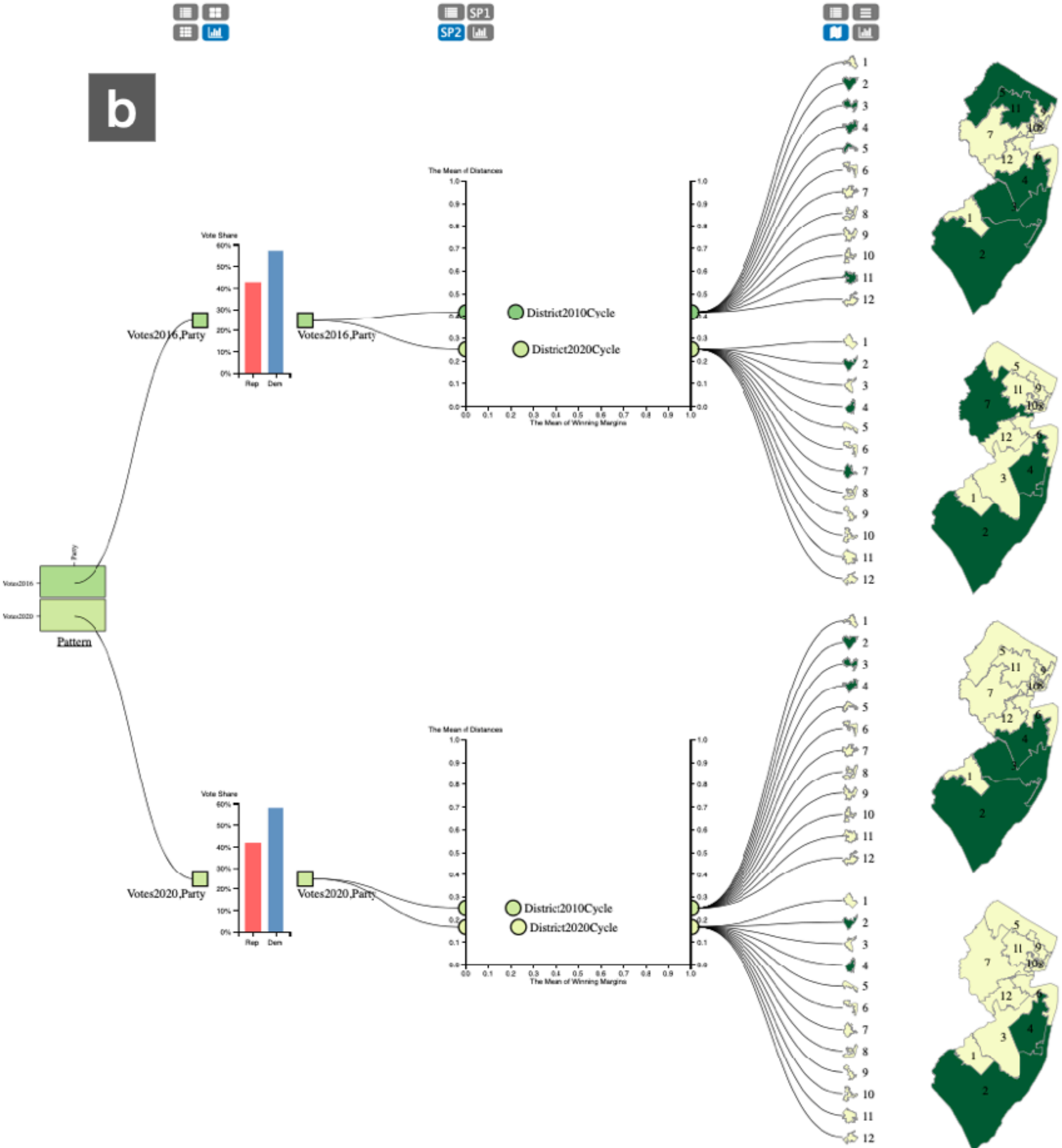


Length

Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

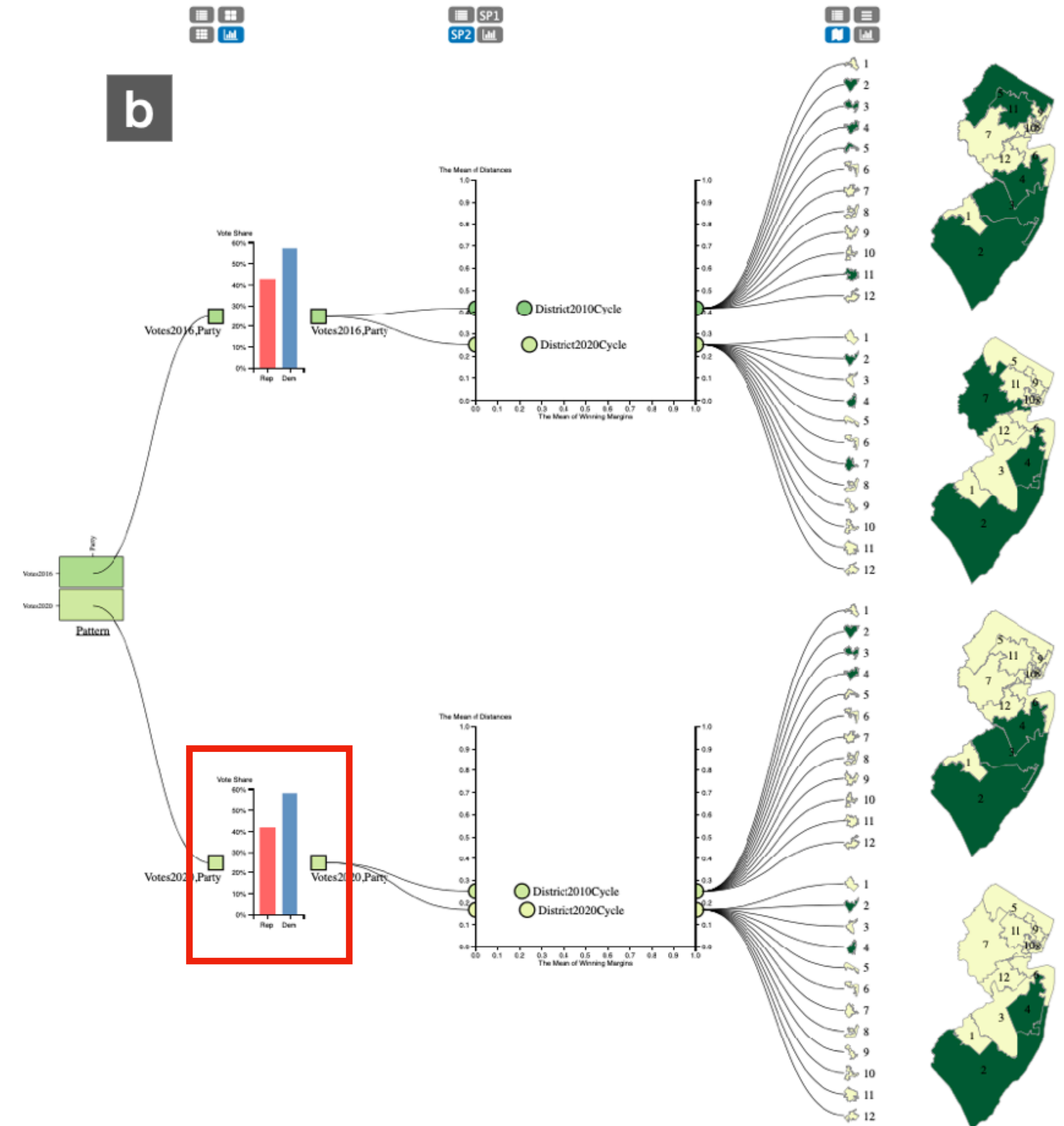
Republican:



Evaluating the efficiency gap

2020 presidential votes and 2020 district cycle

Republican:

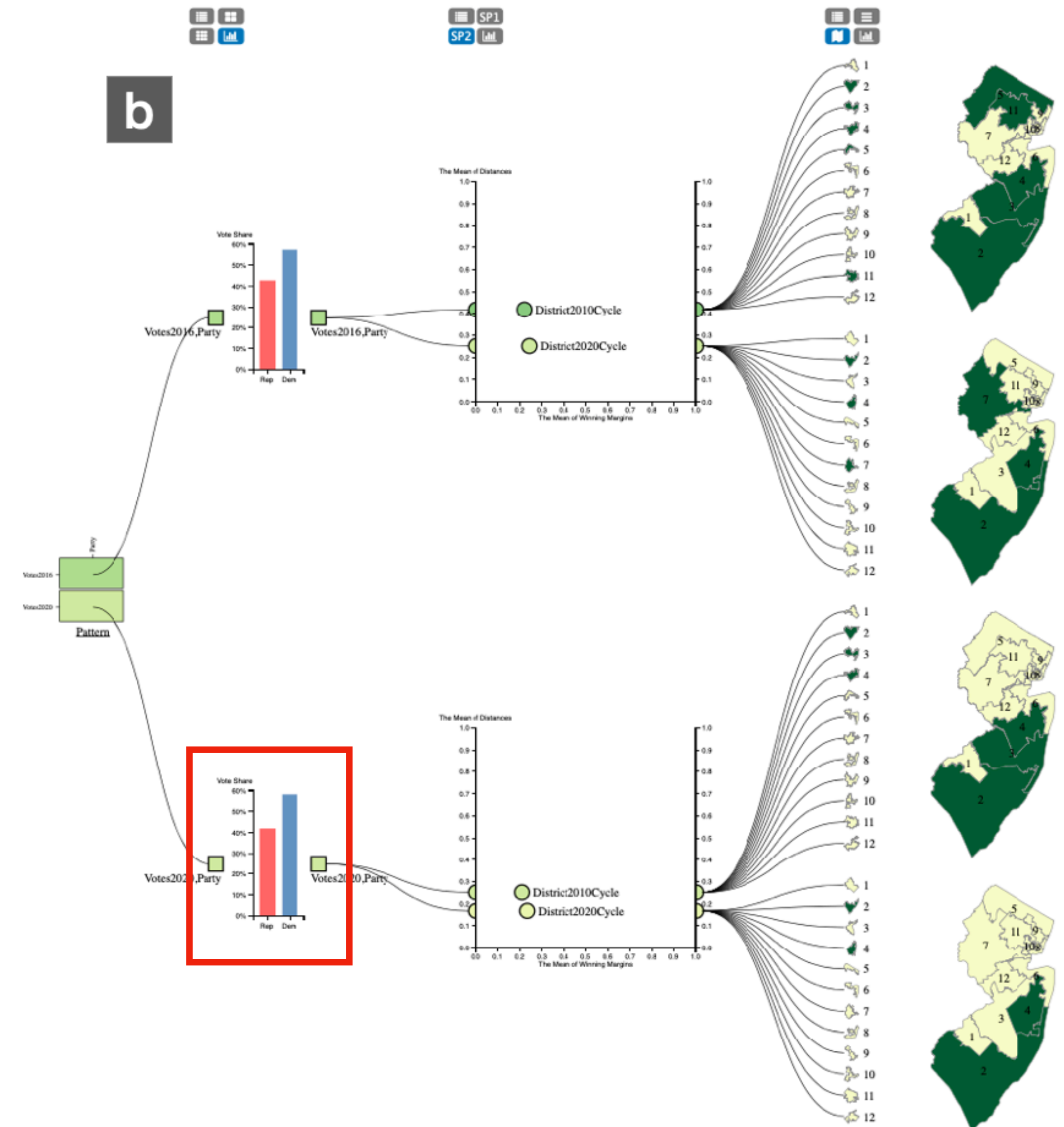


Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

Republican:

Voting share percentage (VSP):

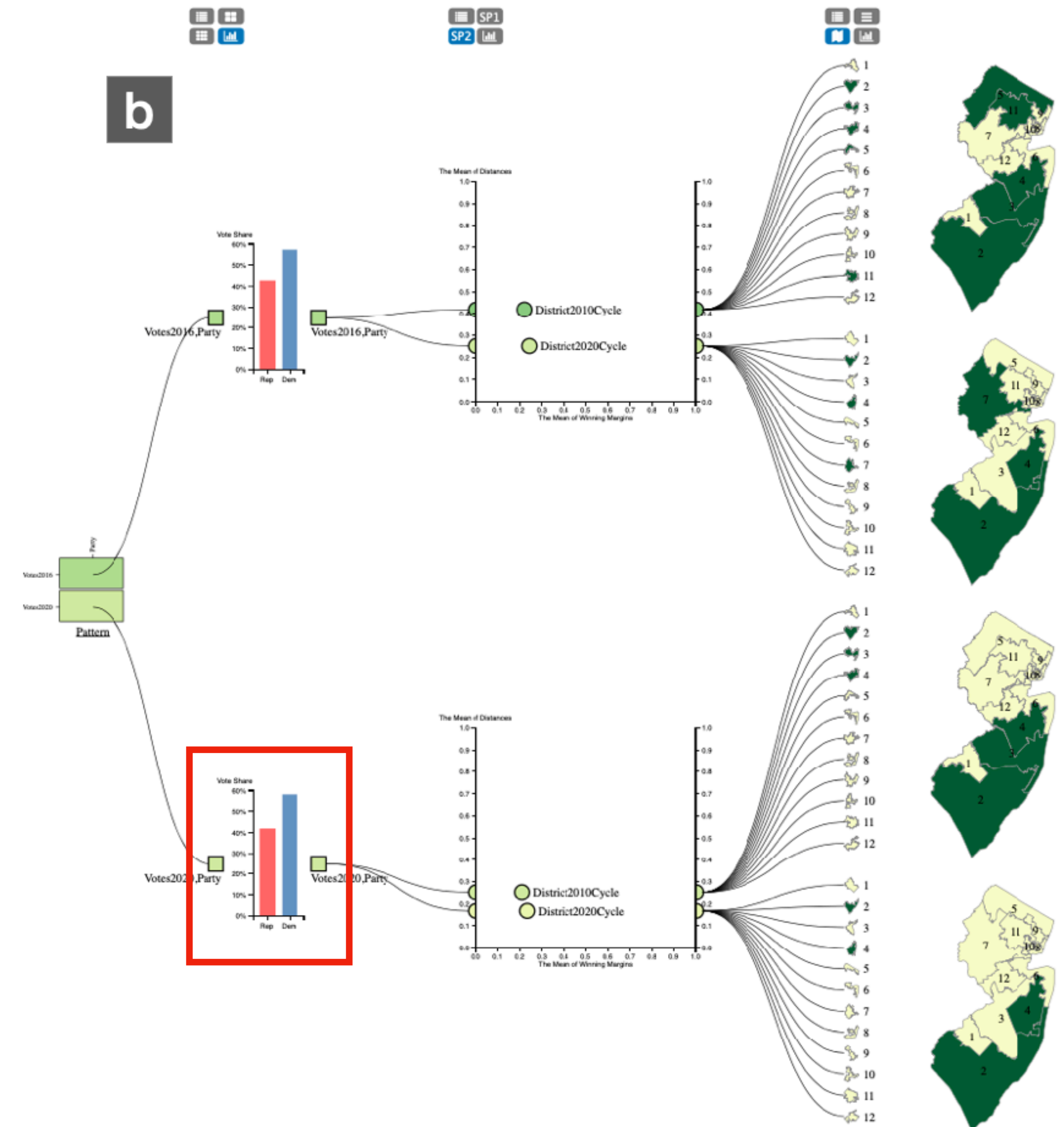


Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

Republican:

Voting share percentage (VSP): 42%

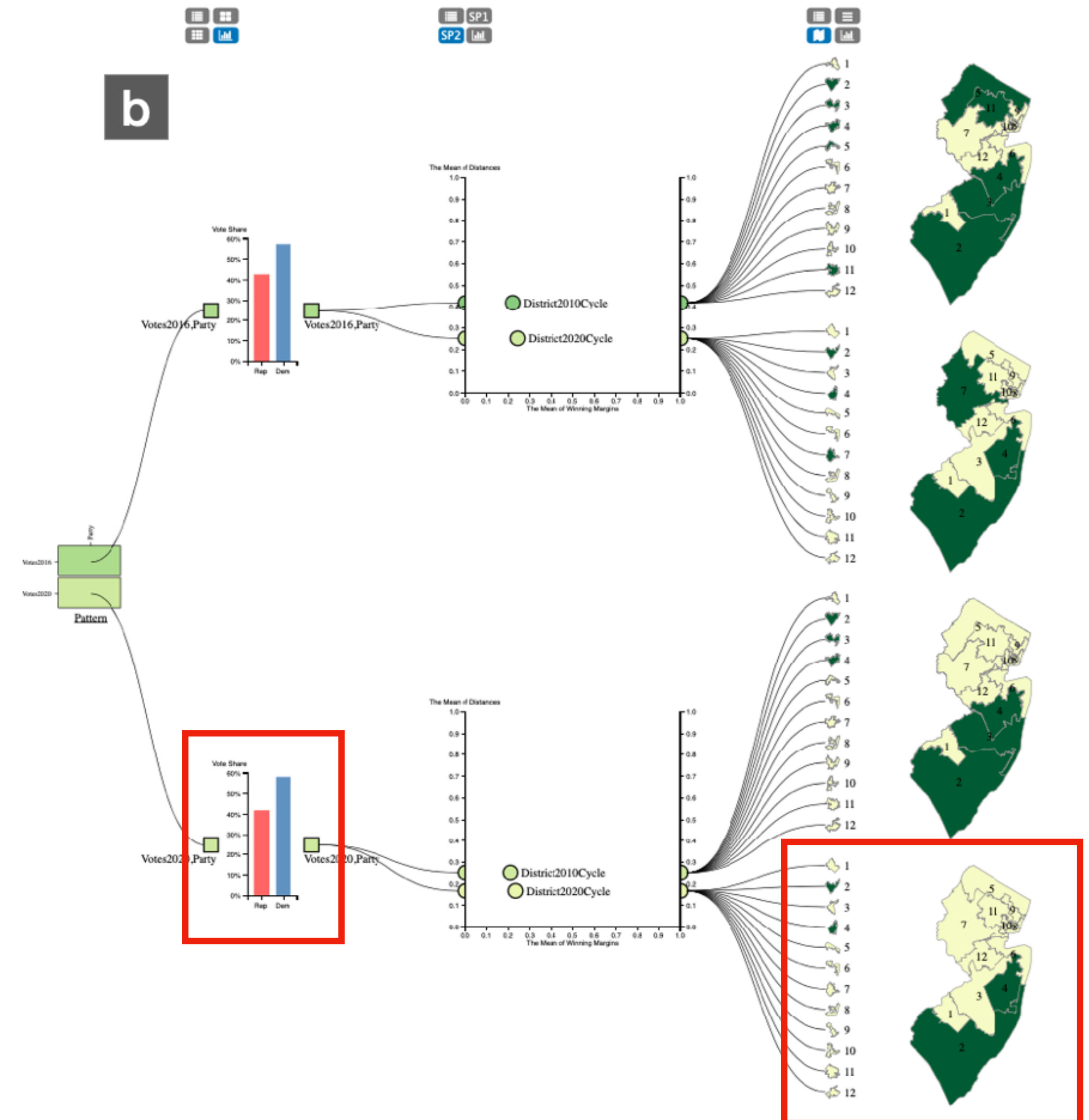


Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

Republican:

Voting share percentage (VSP): 42%



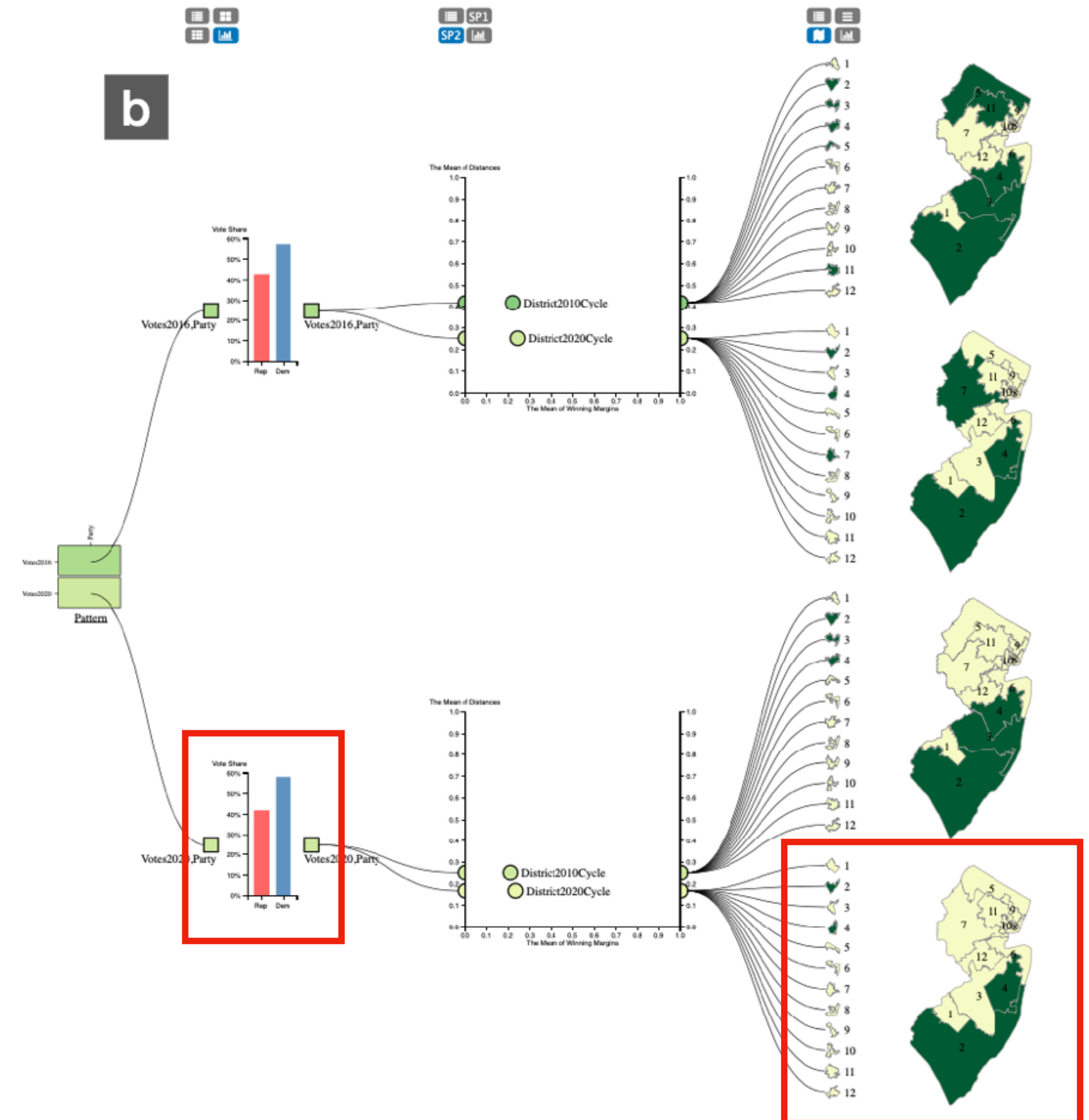
Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

Republican:

Voting share percentage (VSP): 42%

Winning Seat Ratio:



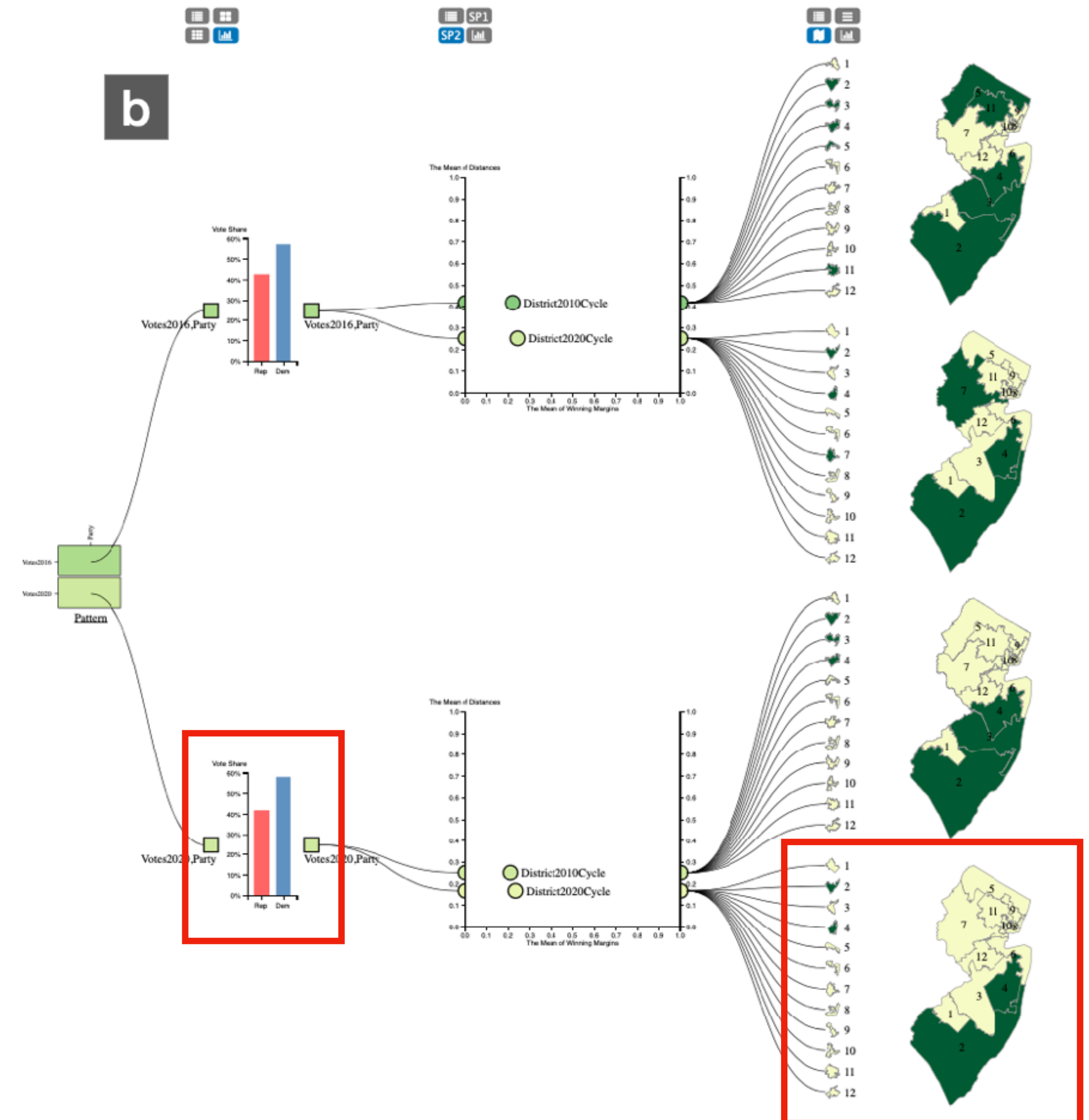
Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

Republican:

Voting share percentage (VSP): 42%

Winning Seat Ratio: $2/12 = 16.7\%$



Evaluating the efficiency gap

2020 presidential votes and
2020 district cycle

Republican:

Voting share percentage (VSP): 42%

Winning Seat Ratio: $2/12 = 16.7\%$

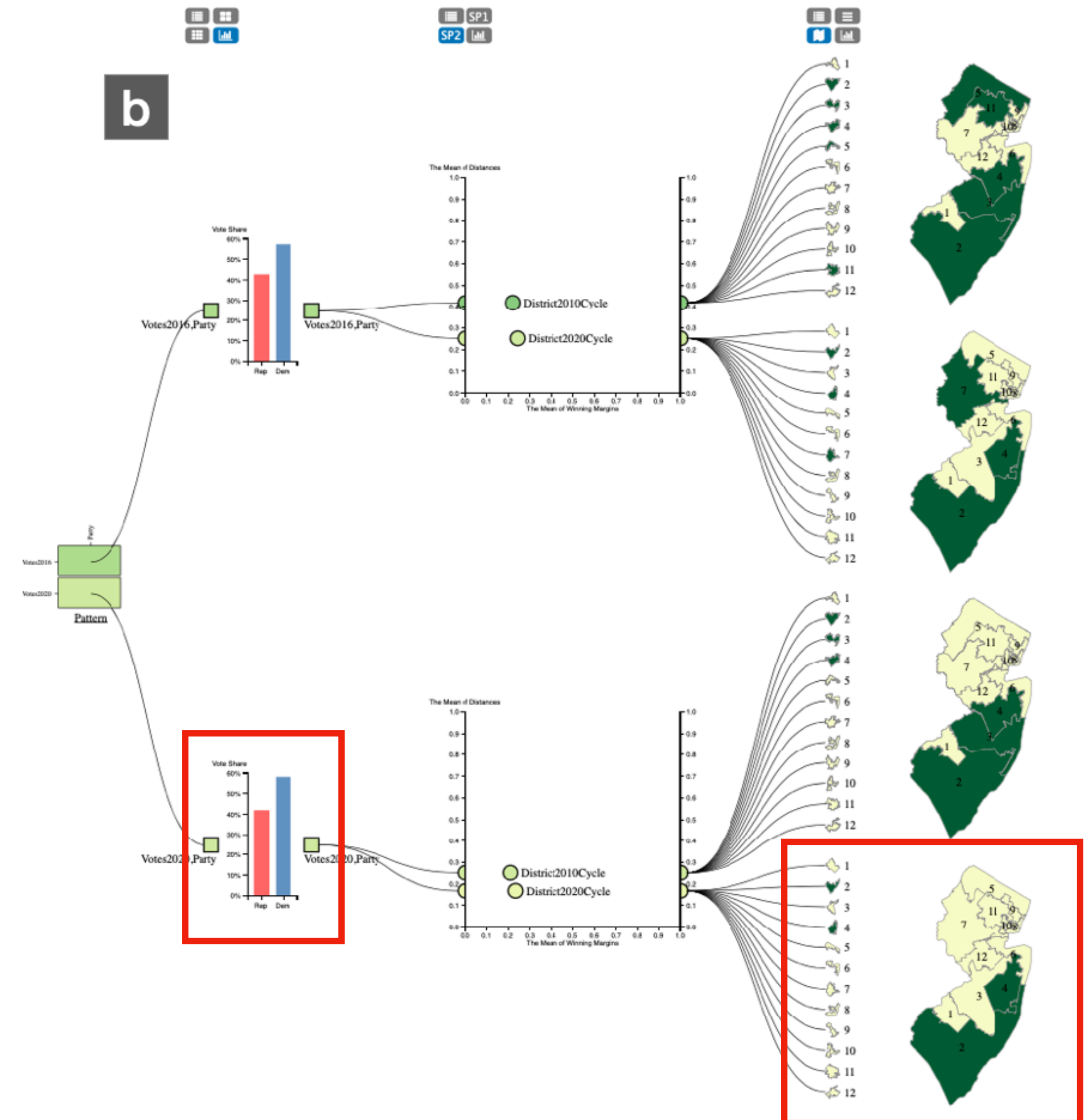
Efficiency Gap = Seat Margin - (2 x Vote Margin)

Republican

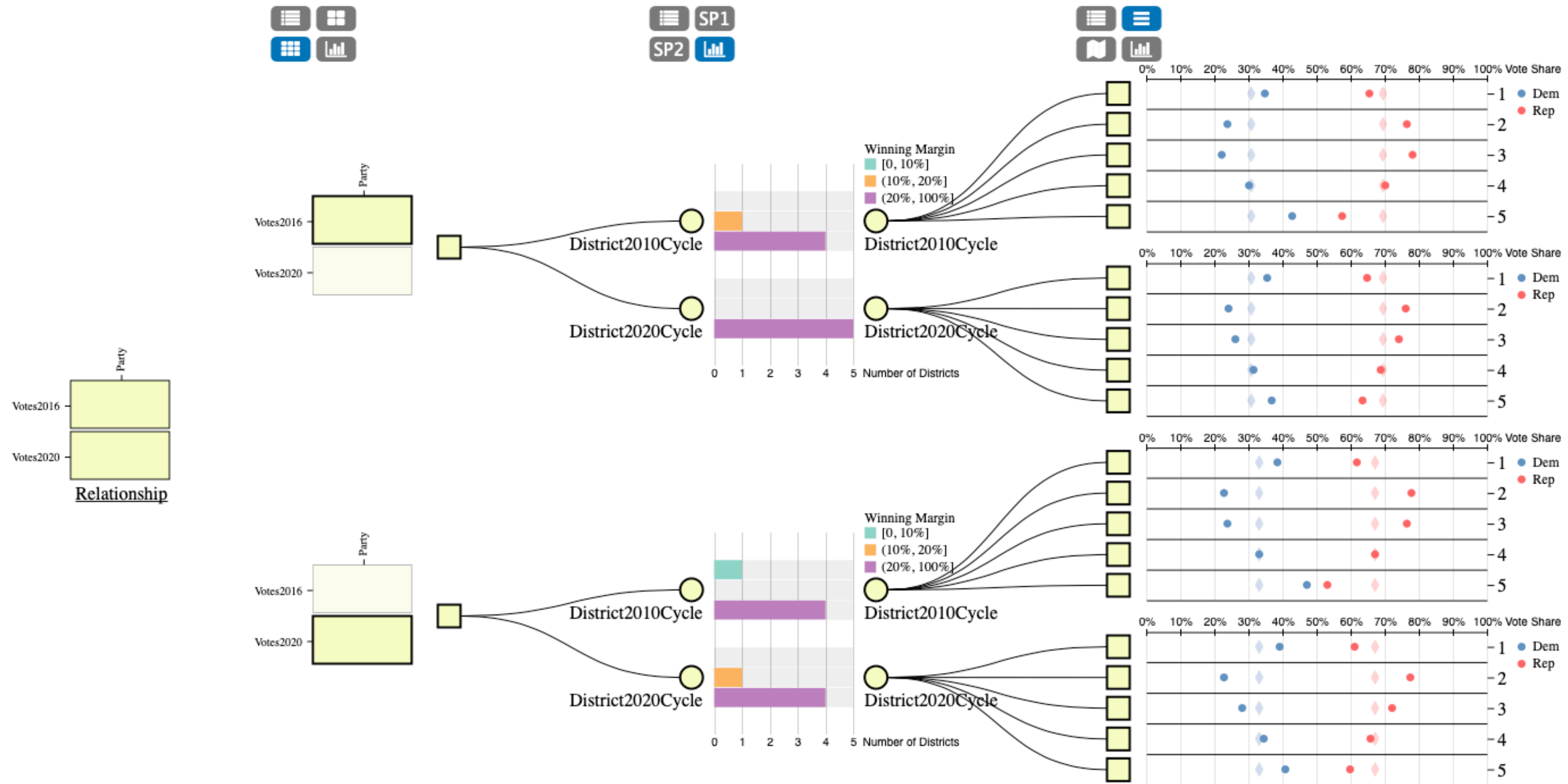
$$= (2/12 - 50\%) - (2 \times (42\% - 50\%))$$

$$= (-33\%) - (-16\%)$$

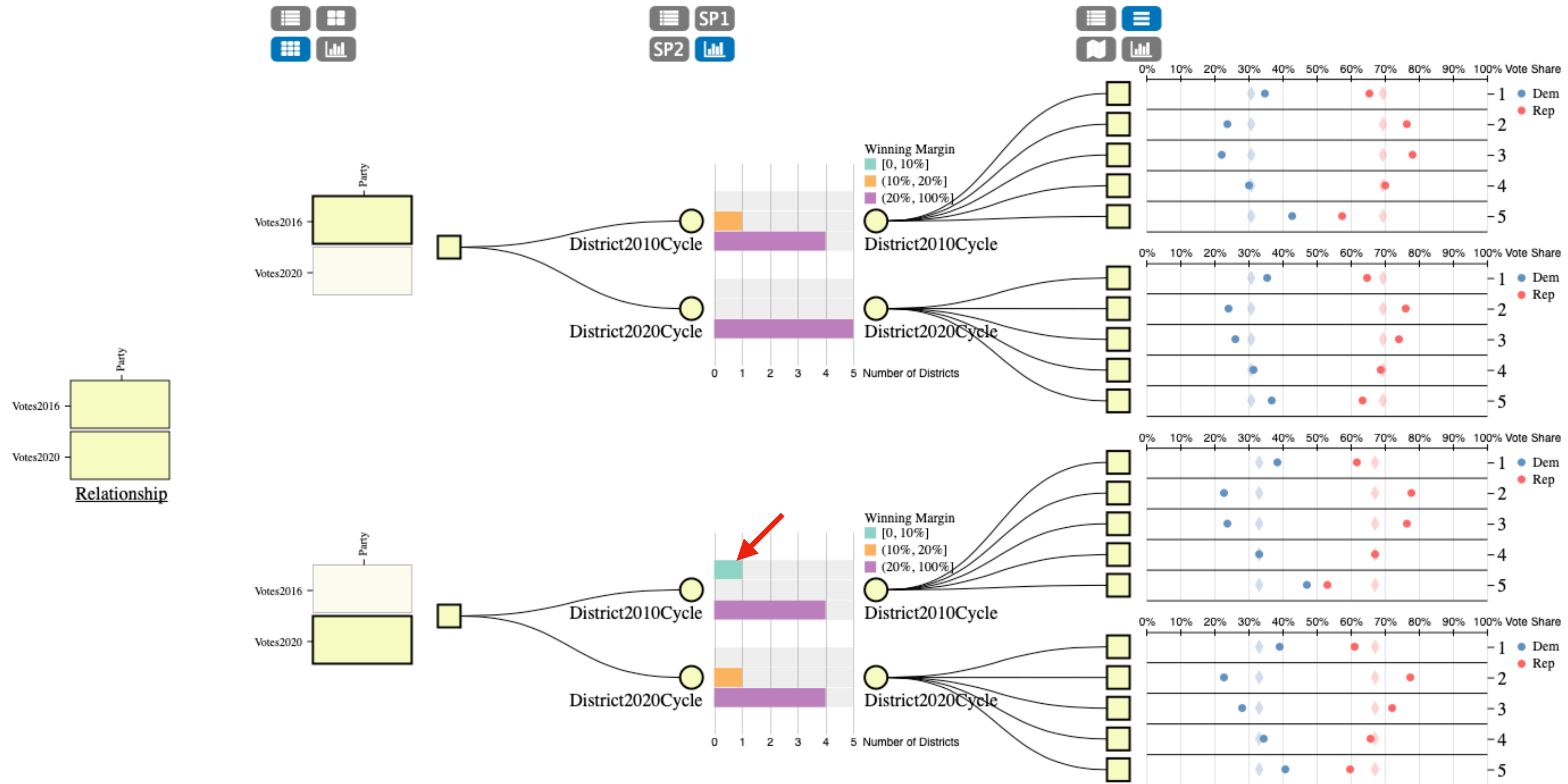
$$= -17\%$$



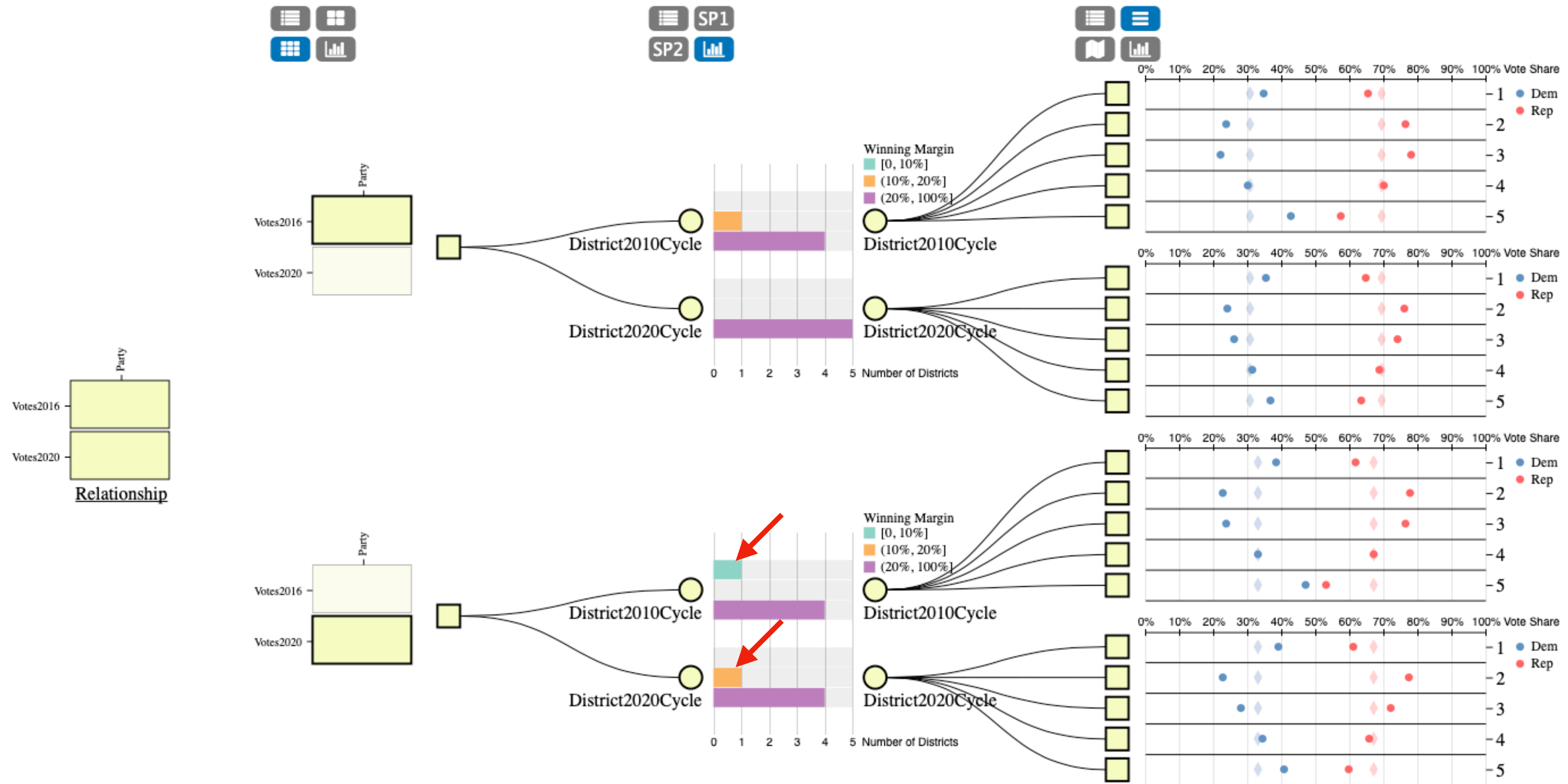
Assessing electoral competition



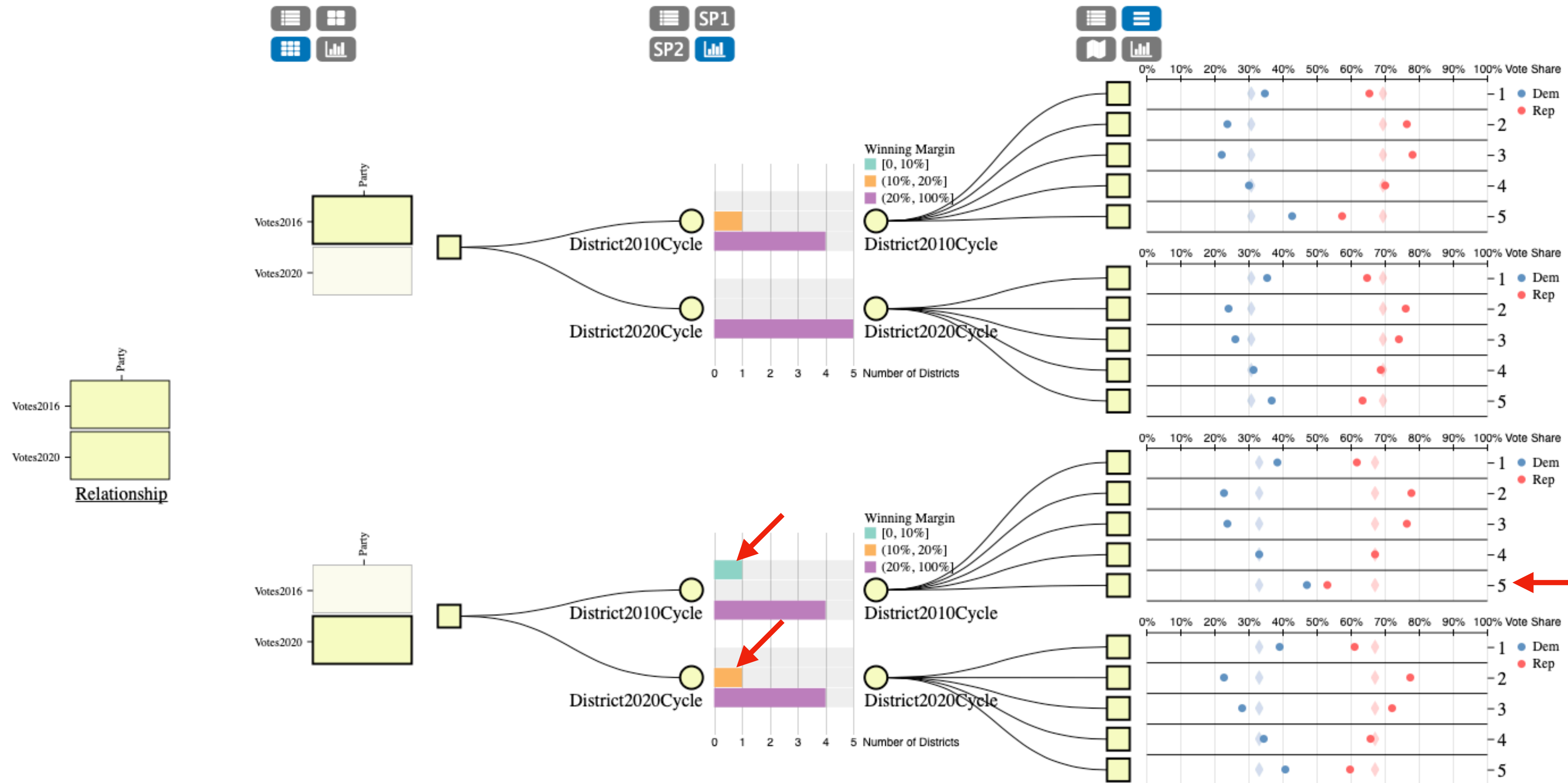
Assessing electoral competition



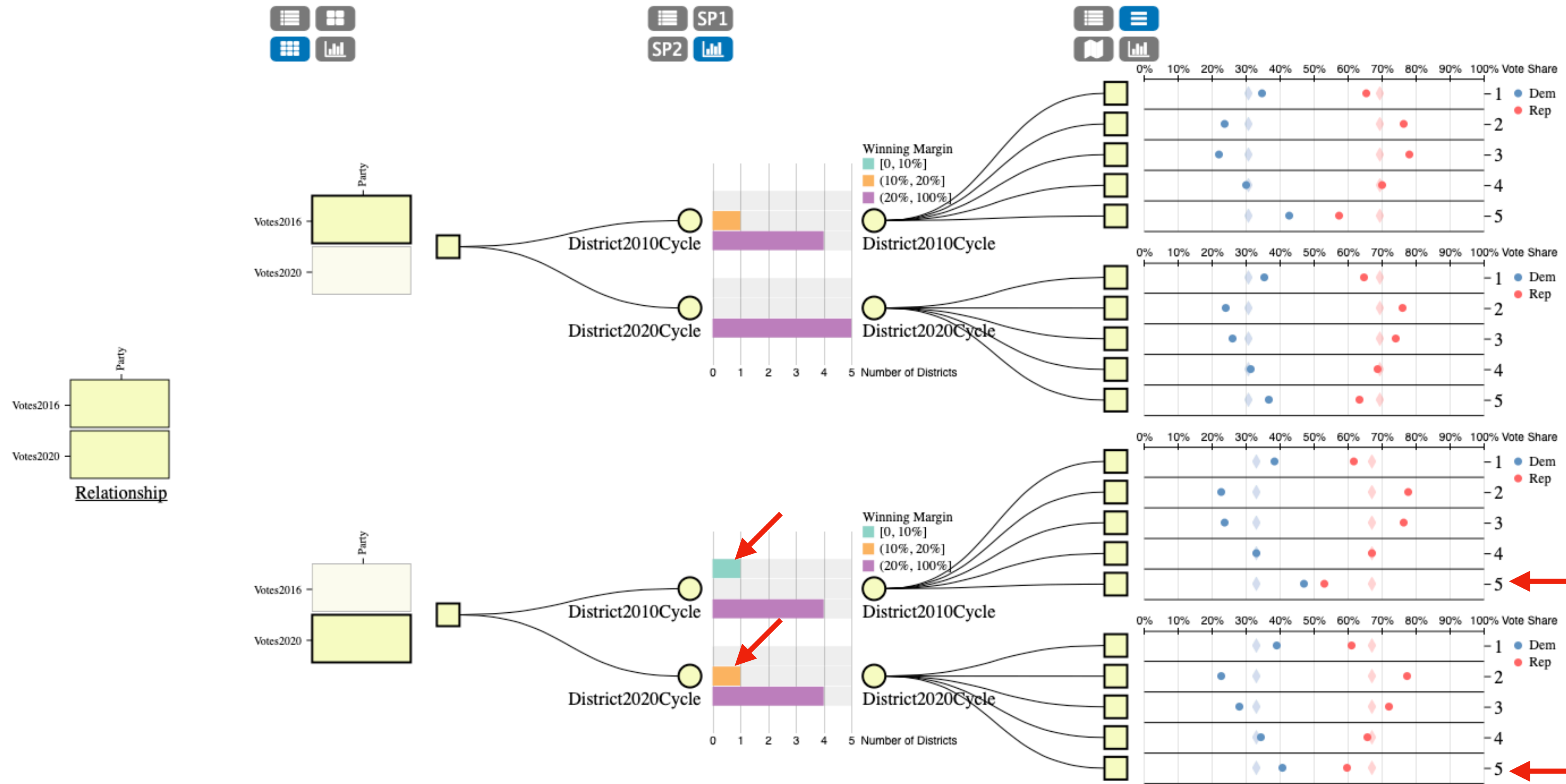
Assessing electoral competition



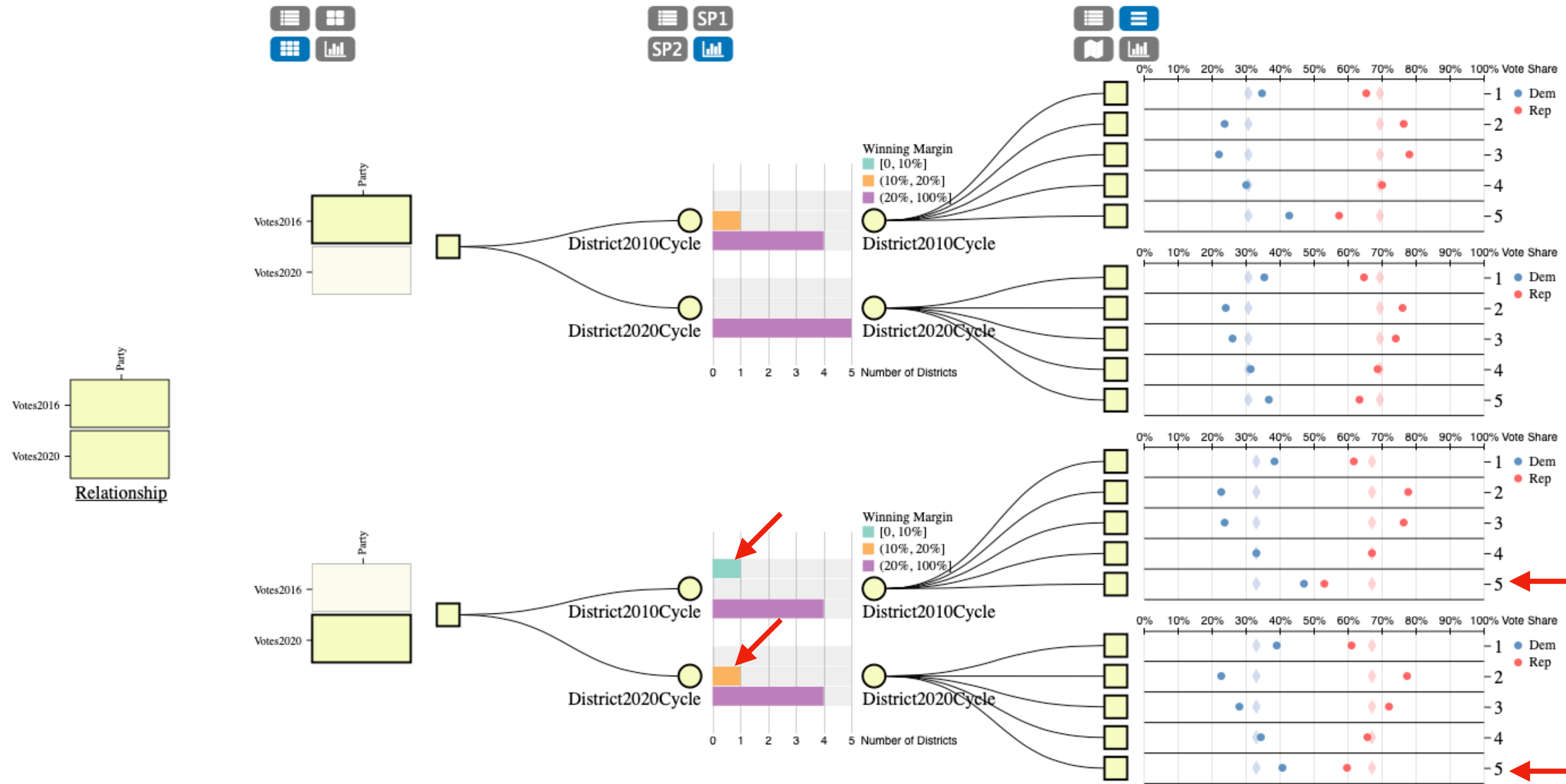
Assessing electoral competition



Assessing electoral competition

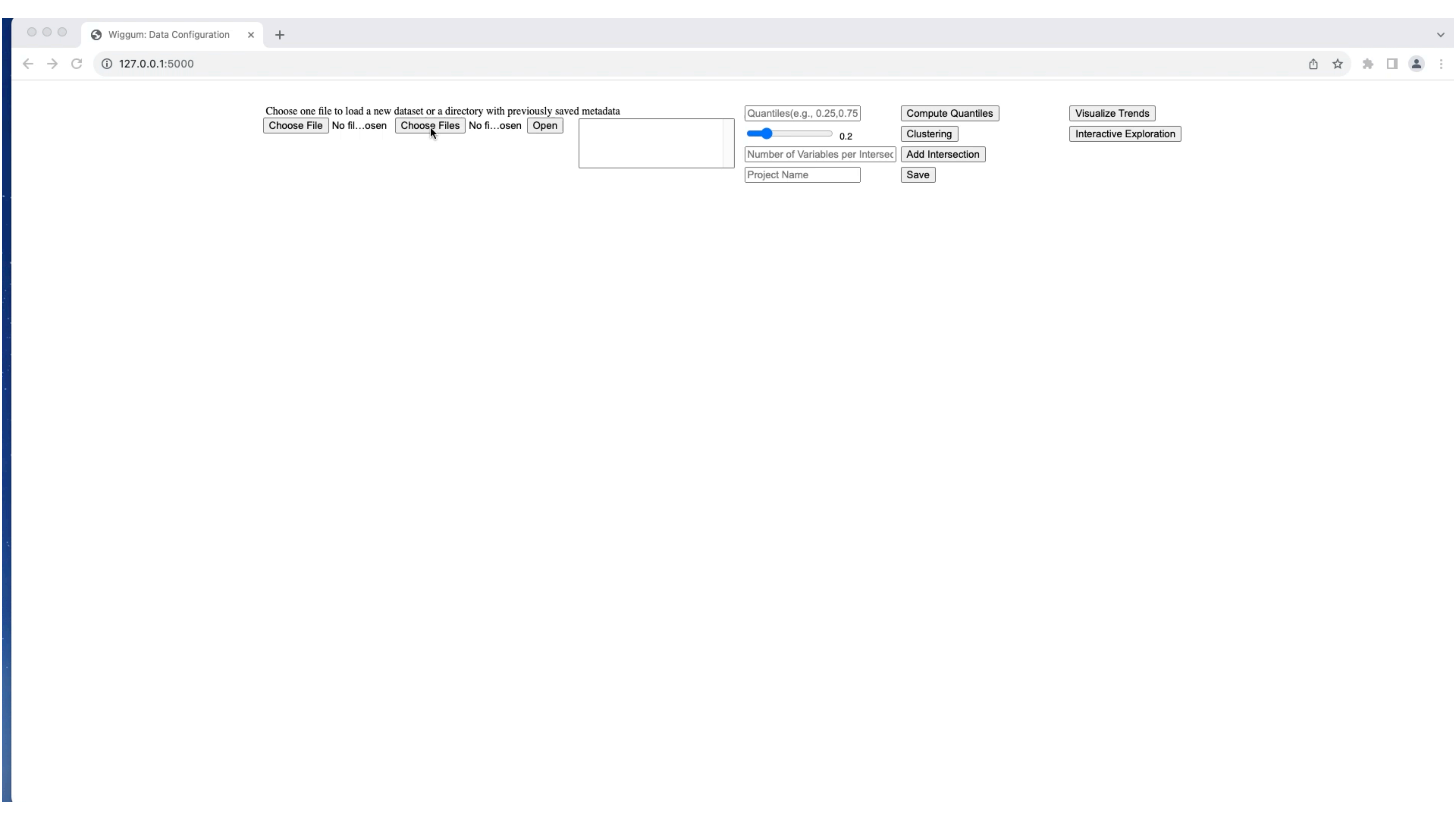


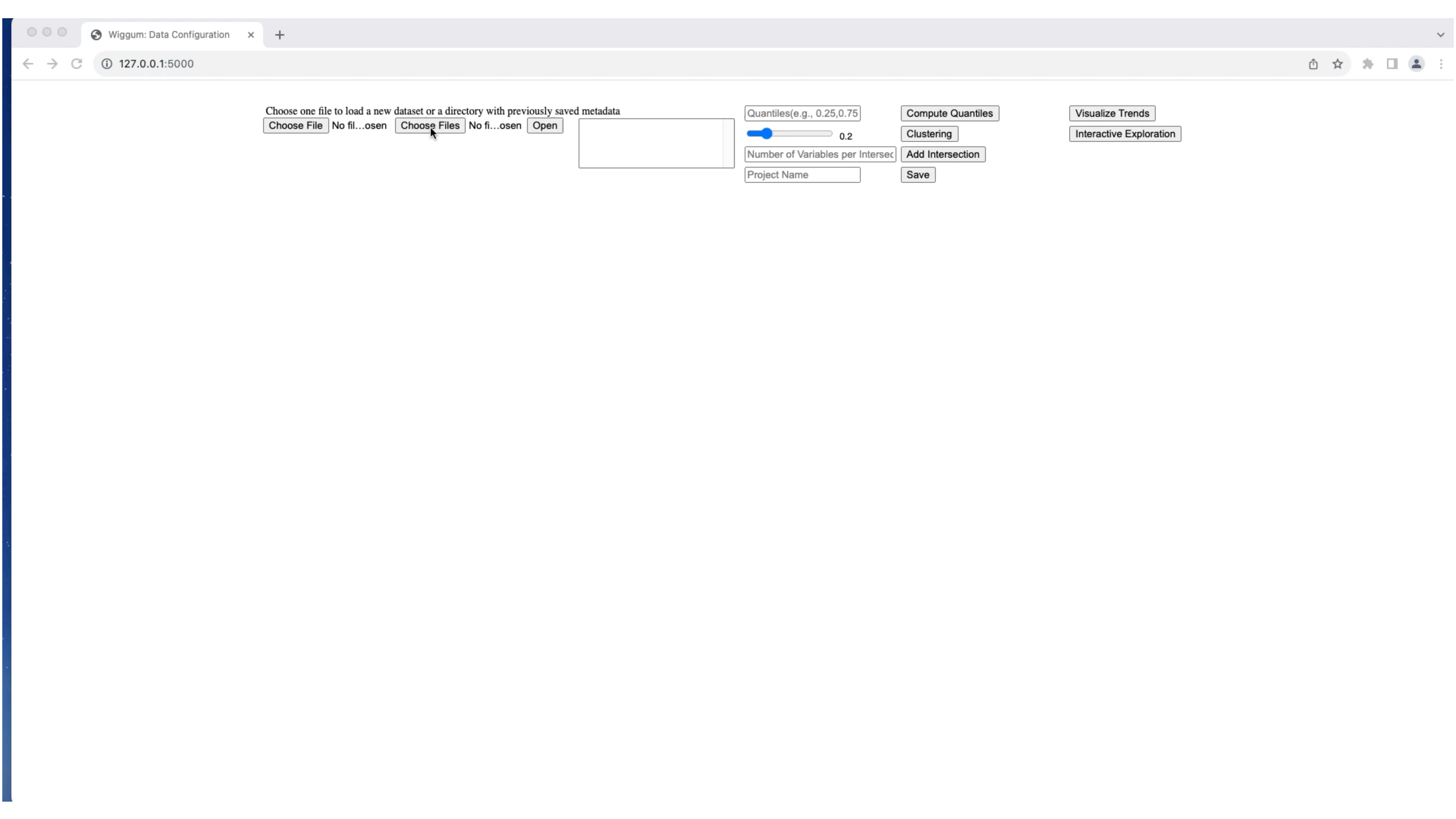
Assessing electoral competition

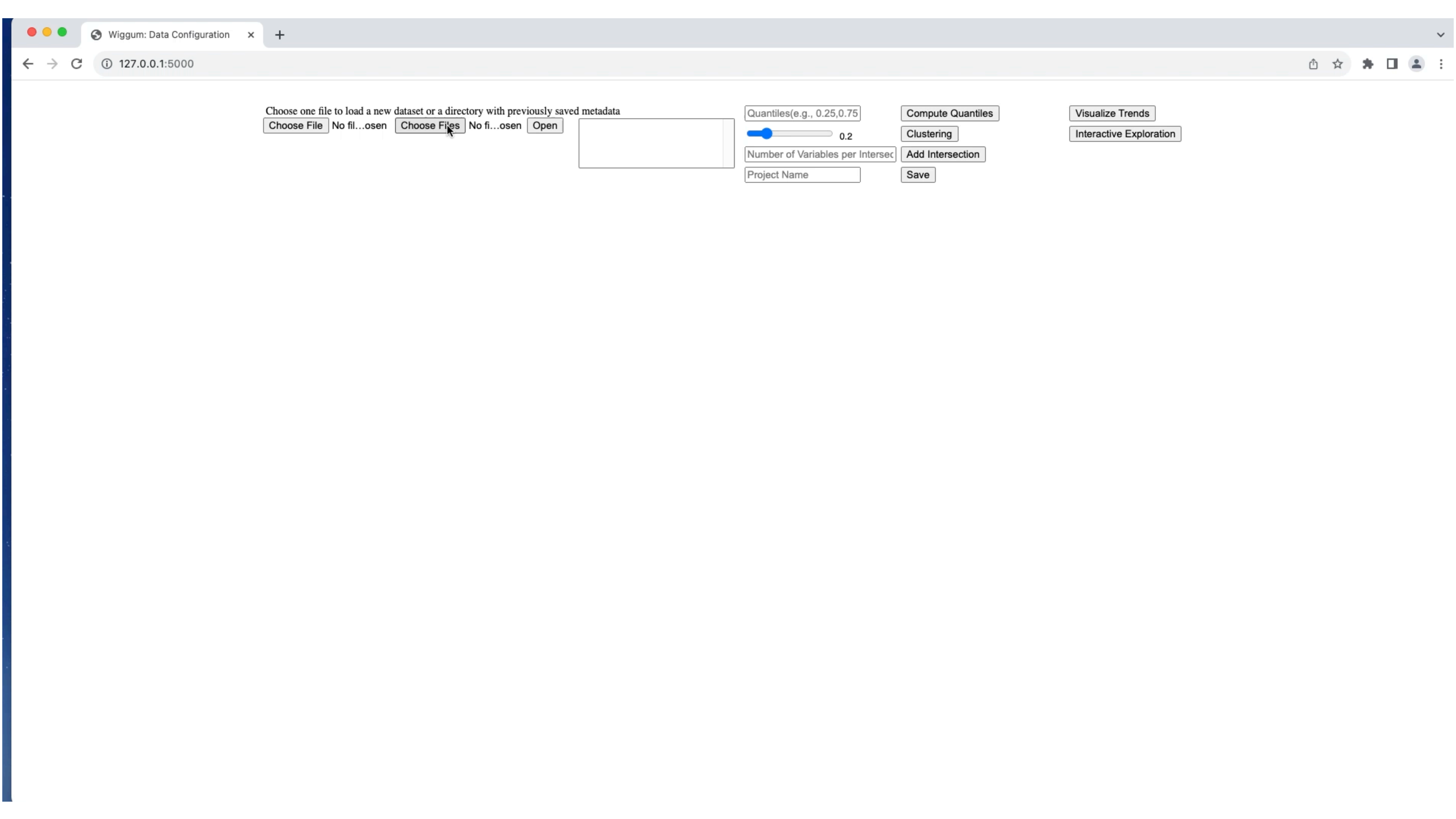


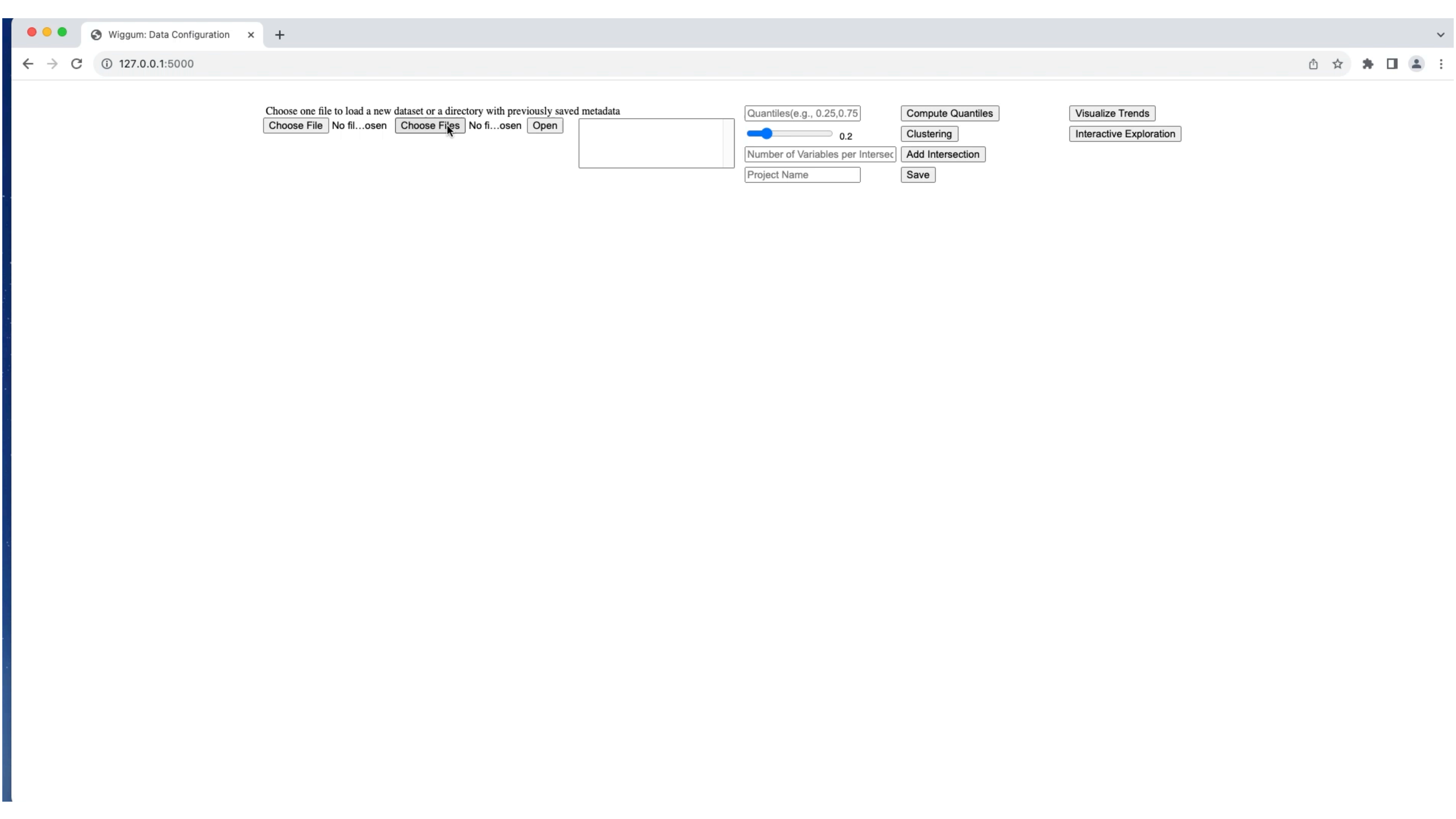
Our findings in Oklahoma is a strong sign that gerrymandering can occur even without overturning the election.

Demo





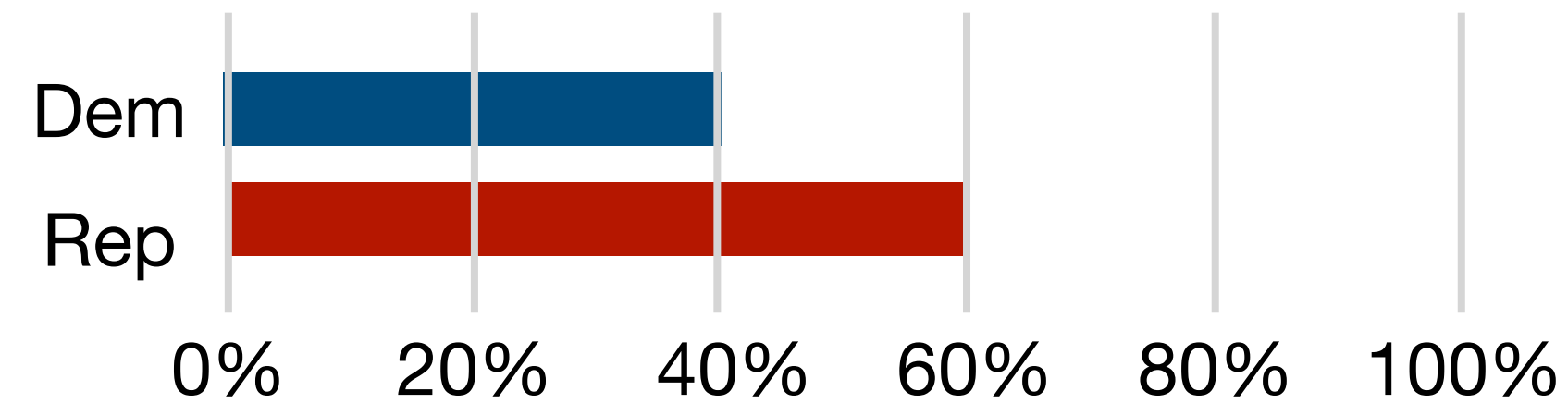




Thanks!

Which Pattern is in Gerrymandering?

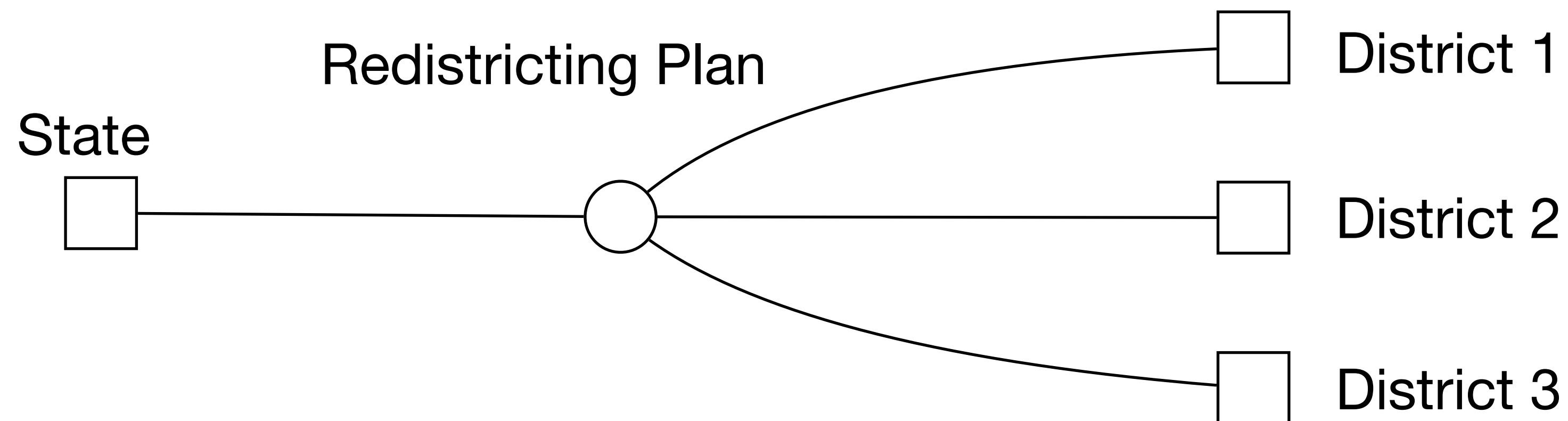
Voting Share Percentage (Dem) < Voting Share Percentage(Rep)



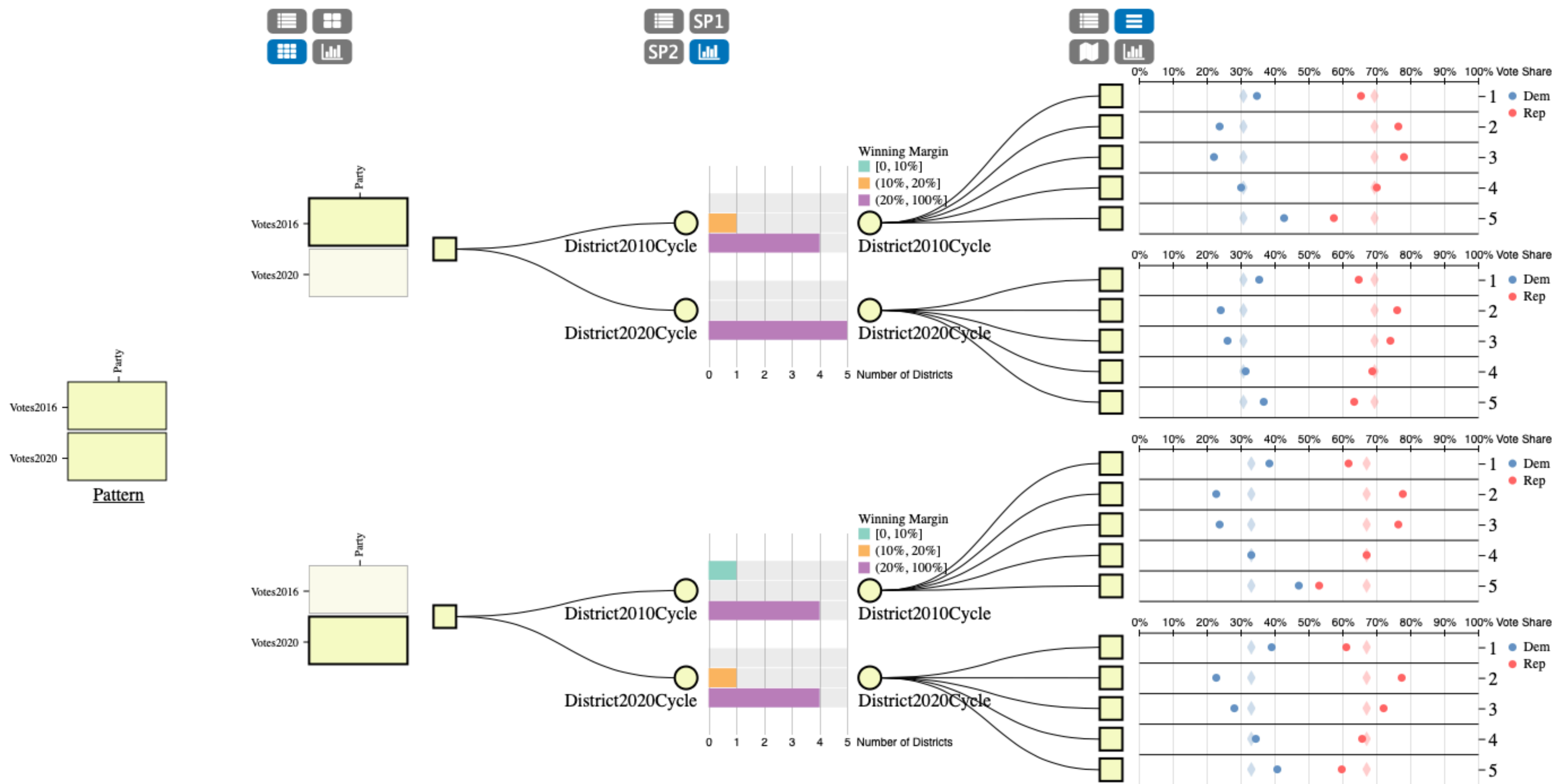
Aggregate Data Pattern

Splitby

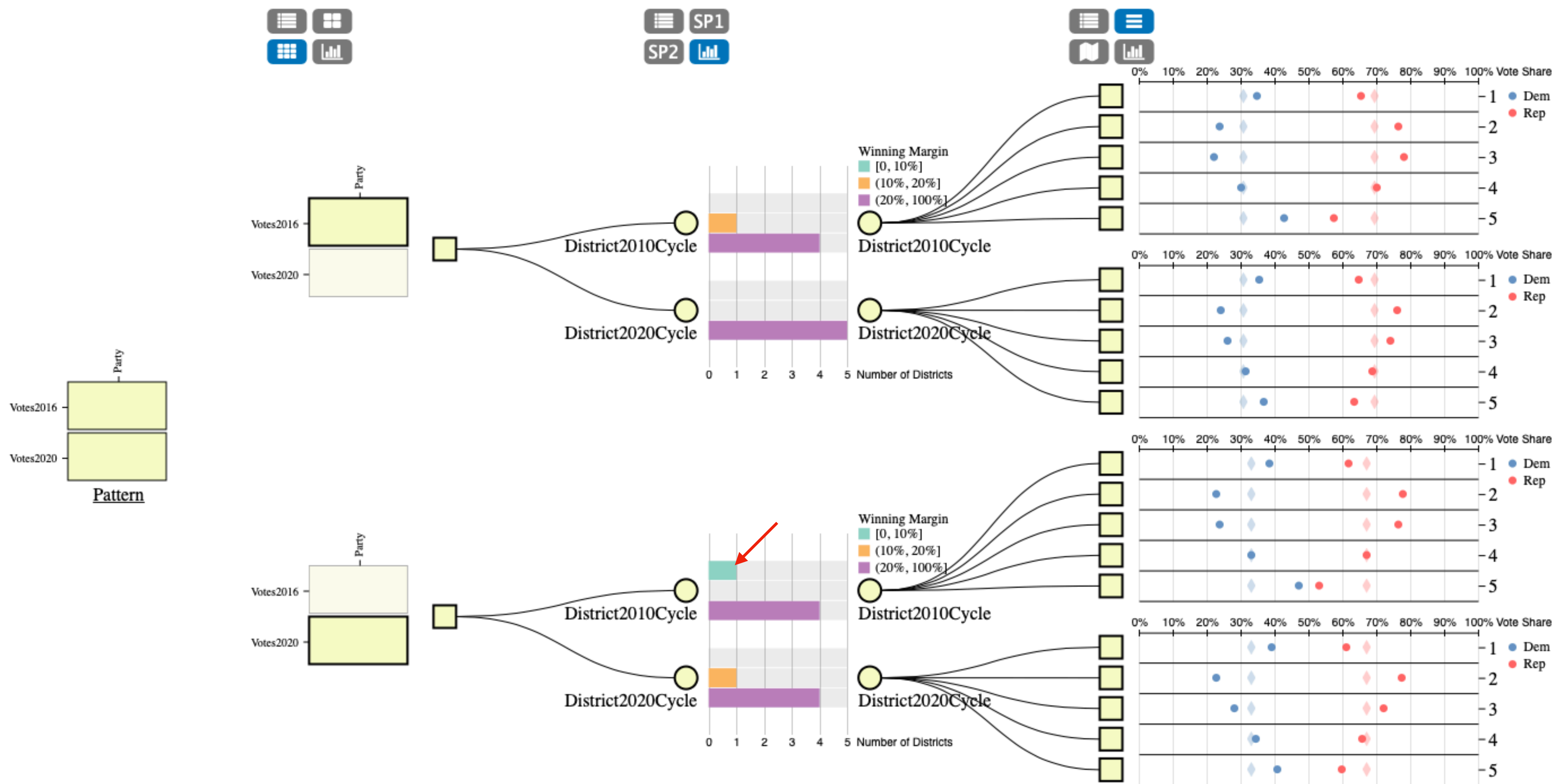
Subgroup Pattern



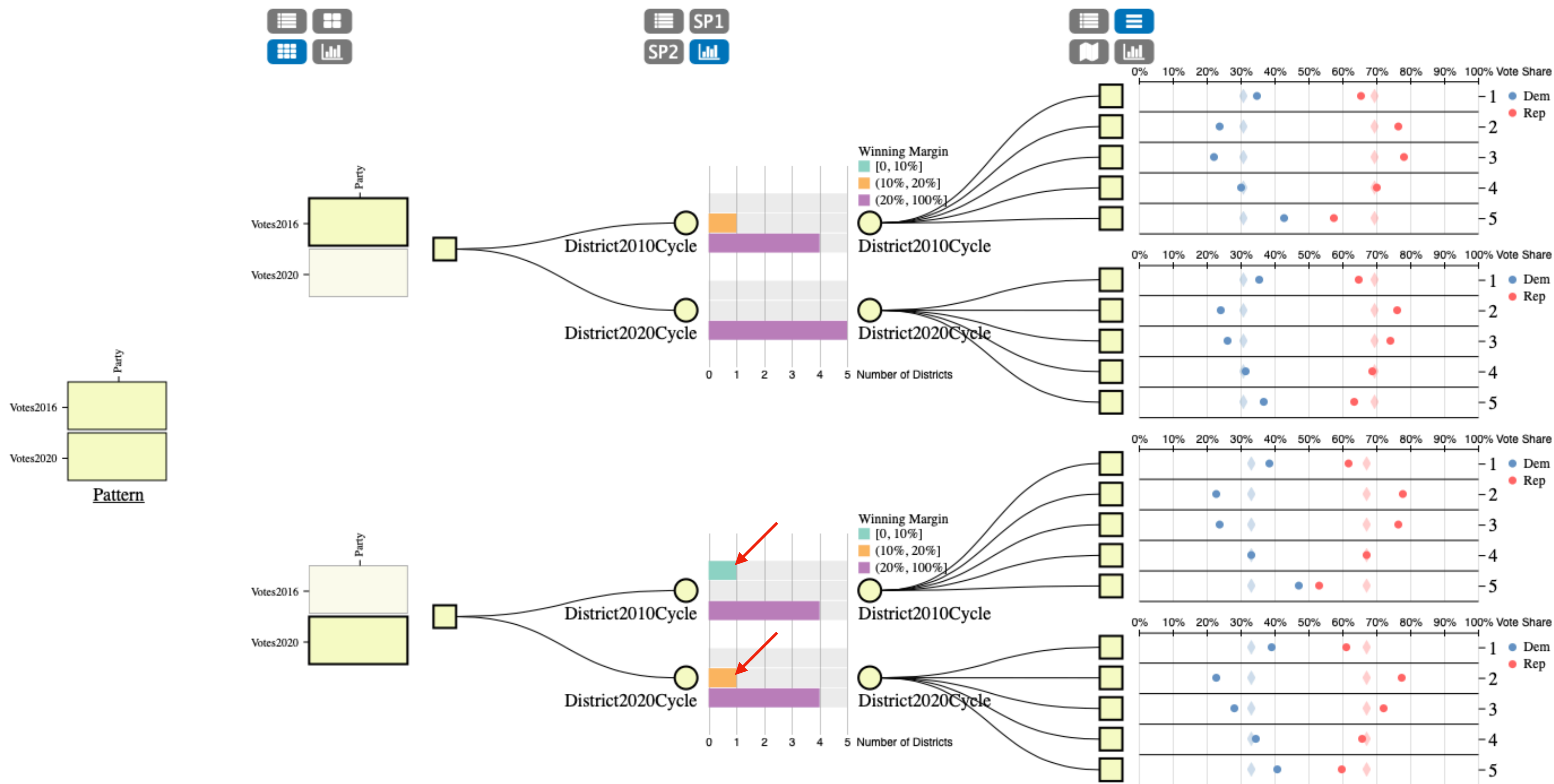
Assess the Electoral Competition



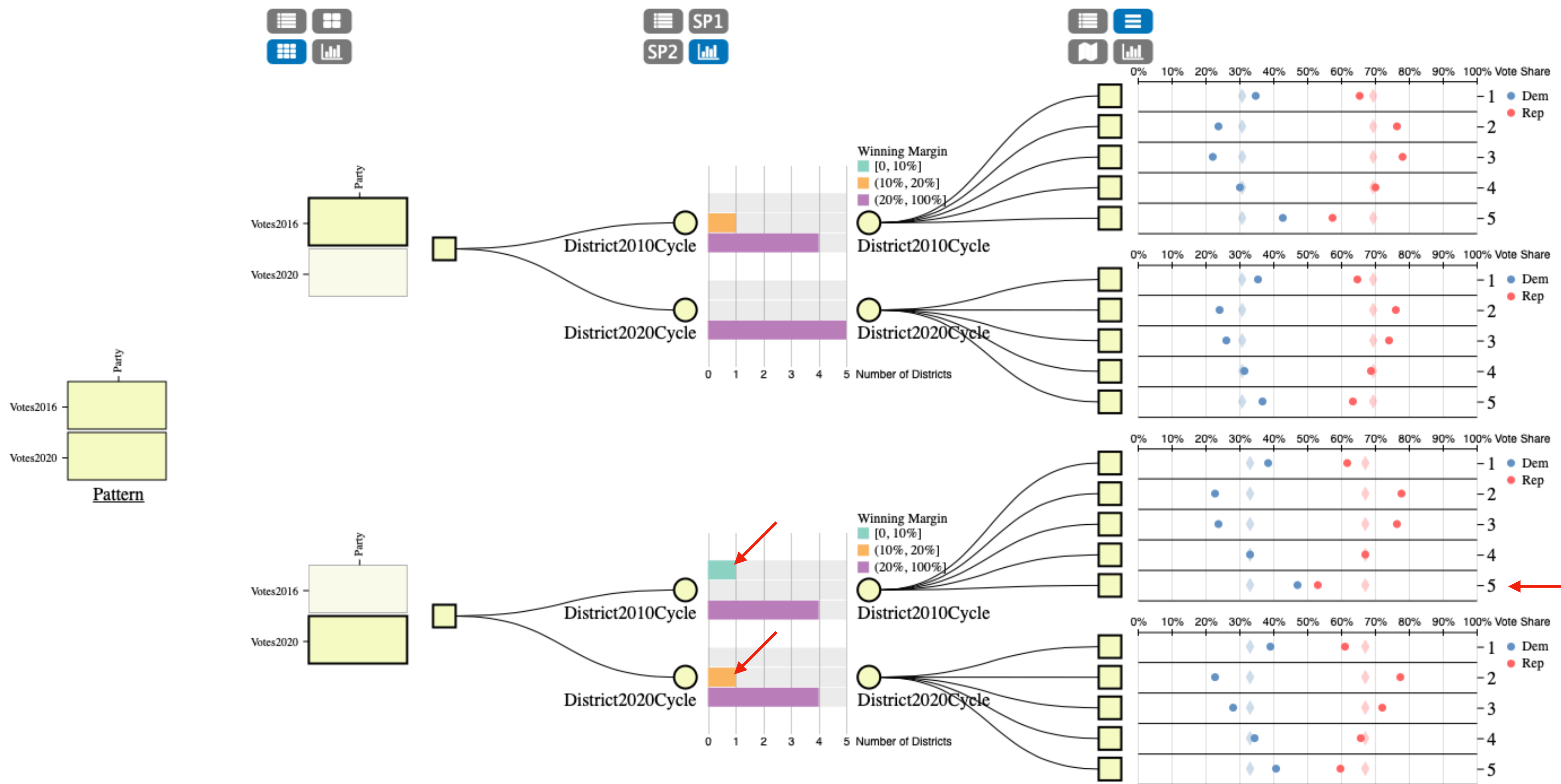
Assess the Electoral Competition



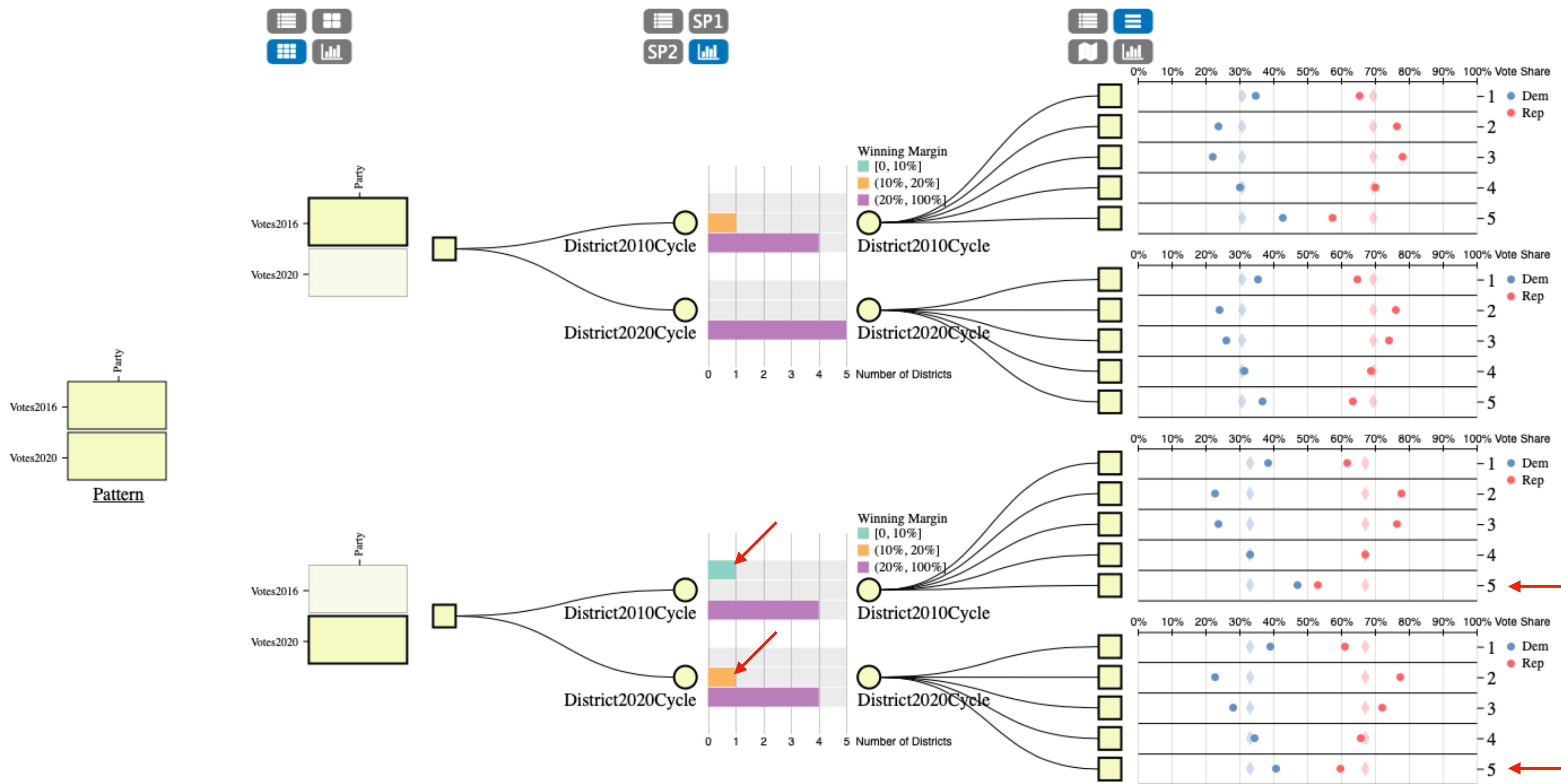
Assess the Electoral Competition



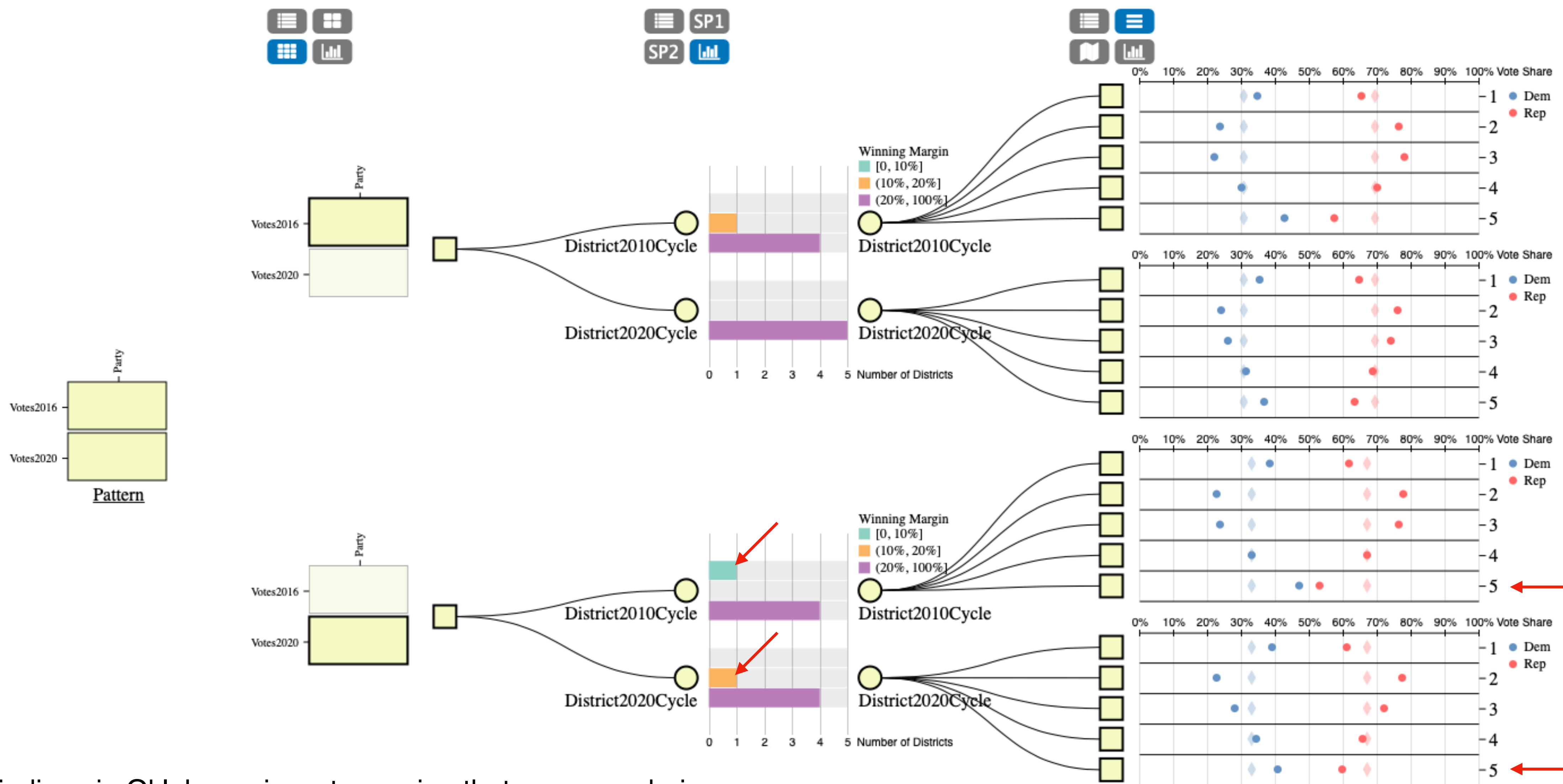
Assess the Electoral Competition



Assess the Electoral Competition

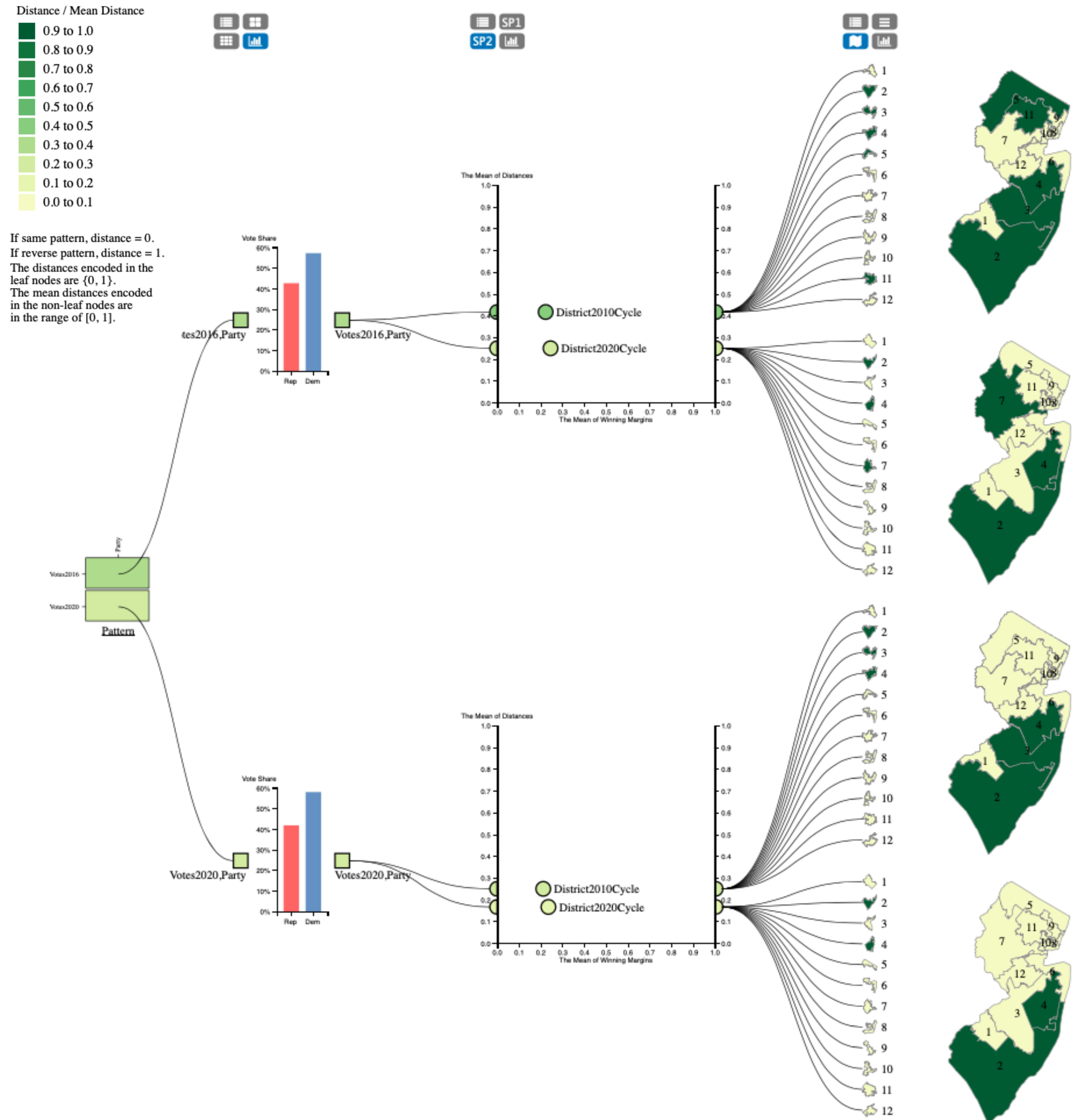


Assess the Electoral Competition



Our findings in Oklahoma is a strong sign that gerrymandering can occur even without overturning the election.

Evaluate Efficiency Gap



Evaluate Efficiency Gap

2020 presidential votes and 2020 district cycle

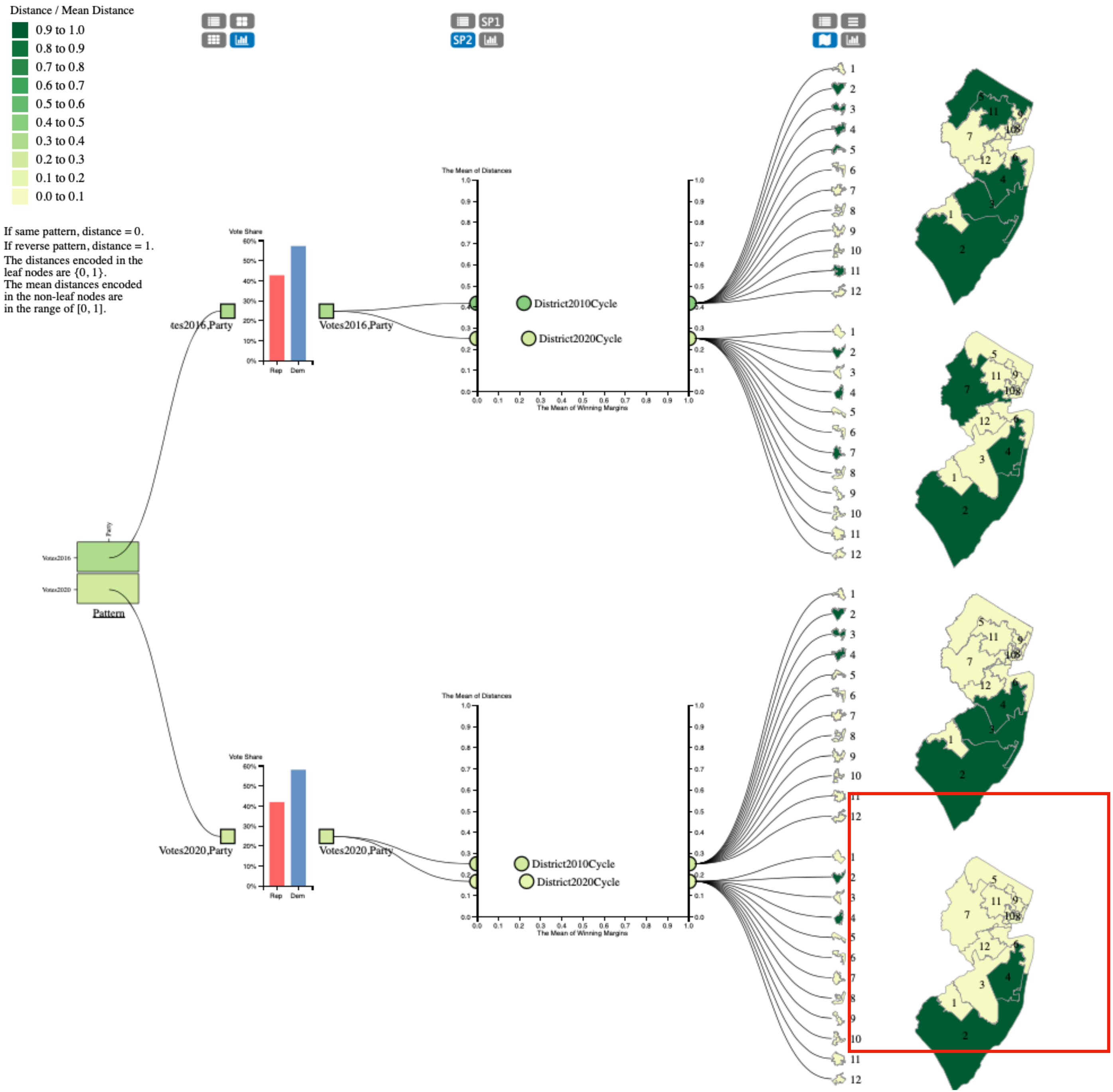
Efficiency Gap = Seat Margin - (2 x Vote Margin)

Democratic

= (10/12-50%) - (2 x (58%-50%))

= 33% - 16%

= 17%



Evaluate Efficiency Gap

2020 presidential votes and 2020 district cycle

Efficiency Gap = Seat Margin - (2 x Vote Margin)

Democratic

= (10/12-50%) - (2 x (58%-50%))

= 33% - 16%

= 17%

Efficiency Gap = Seat Margin - (2 x Vote Margin)

Republican

= (2/12-50%) - (2 x (42%-50%))

= (-33%) - (-16%)

= -17%

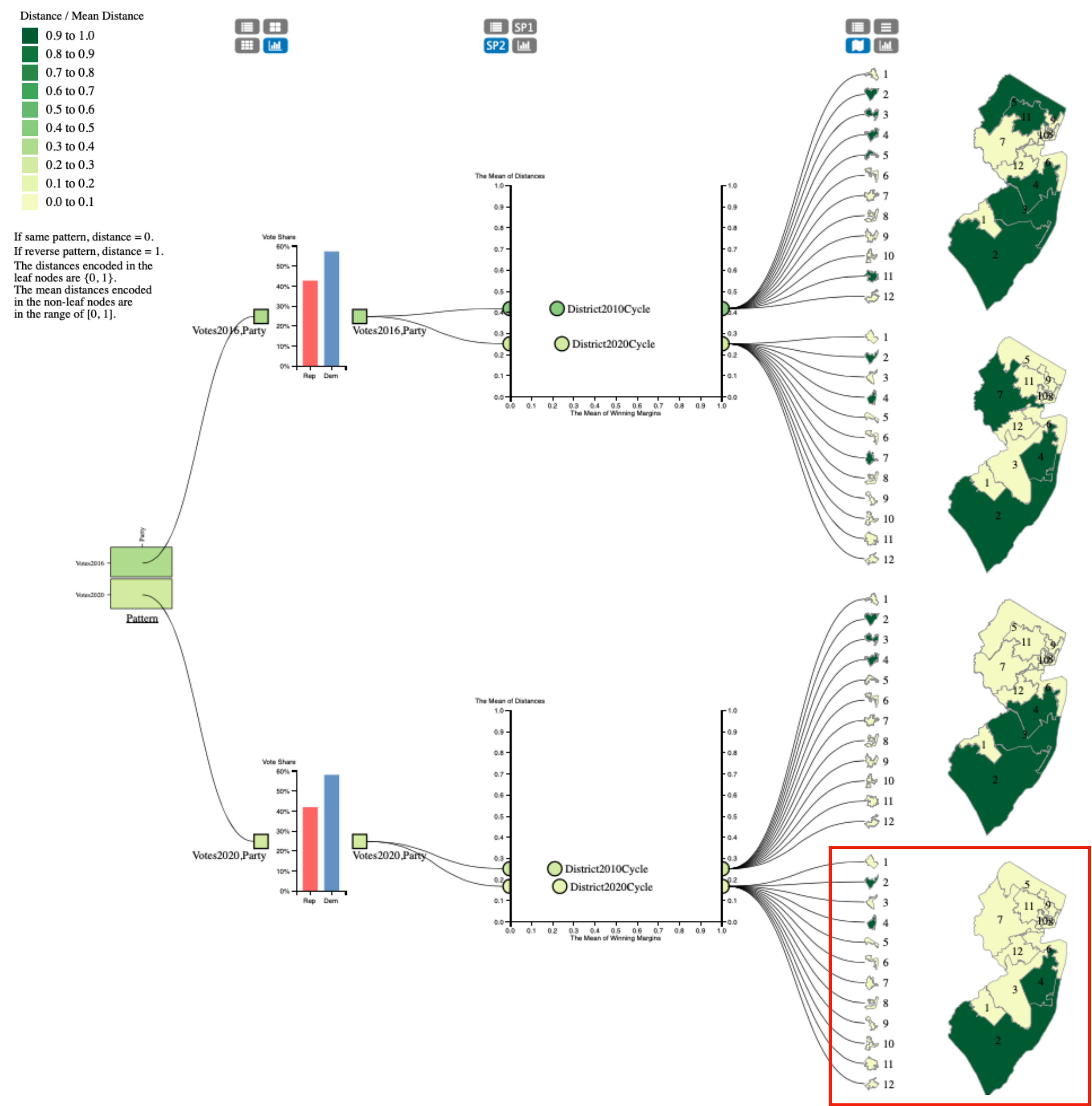
If democratic party wins 8 seats:

Efficiency Gap = Seat Margin - (2 x Vote Margin)

= (8/12-50%) - (2 x (58%-50%))

= 16% - 16%

= 0%



Precinct-level Data in DRA

Unchanged Precinct: (1 to 1)

2010 Districting Data

GEOID	District	Dem	Rep
40079000102	2	82	412

2020 Districting Data

GEOID	District	Dem	Rep
40079000102	2	82	412

Splitting Precinct Case 1: (1 to n)

2010 Districting Data

GEOID	District	Dem	Rep
40083000305	3	206	592

2020 Districting Data

GEOID	District	Dem	Rep
40083000303_part1_51	3	38	106
40083000303_part2_125	5	168	486

Splitting Precinct Case 2: (m to n)

2010 Districting Data

GEOID	District	Dem	Rep
40131000030_part1_1	1	3	11
40131000030_part2_73	2	511	2056

2020 Districting Data

GEOID	District	Dem	Rep
40131000030_part1_30	1	78	316
40131000030_part2_44	2	436	1751

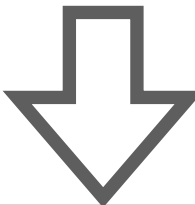
2010 Districting Data

GEOID	District	Dem	Rep
40079000102	2	82	412

2020 Districting Data

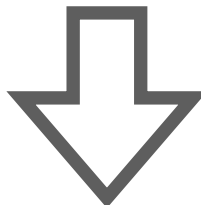
GEOID	District	Dem	Rep
40079000102	2	82	412

Merge on **GEOID**



GEOID	District_2010	District_2020	Dem	Rep
40079000102	2	2	82	412

Melt



GEOID	District_2010	District_2020	Party	Votes
40079000102	2	2	Dem	82
40079000102	2	2	Rep	412

Splitting Precinct Case 1

2020 President Data

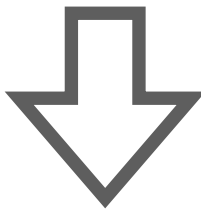
2010 Districting Data

GEOID	District	Dem	Rep
40083000305	3	206	592

2020 Districting Data

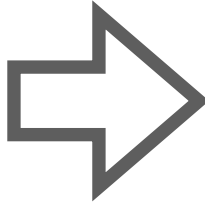
GEOID	District	Dem	Rep
40083000303_part1_51	3	38	106
40083000303_part2_125	5	168	486

Merge on GEOID



GEOID	District_2010	District_2020	Dem	Rep
40083000305	3	n.a.	206	592
40083000305_part1_51	n.a.	3	38	106
40083000305_part2_125	n.a.	5	168	486

Melt



GEOID	District_2010	District_2020	Party	Votes
40083000305	3	n.a.	Dem	206
40083000305	3	n.a.	Rep	592
40083000303_part1	n.a.	3	Dem	38
40083000303_part1	n.a.	3	Rep	106
40083000303_part2	n.a.	5	Dem	168
40083000303_part2	n.a.	5	Rep	486

Splitting Precinct Case 2

2020 President Data

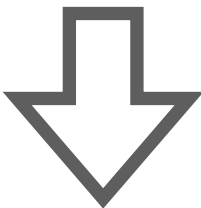
2010 Districting Data

GEOID	District	Dem	Rep
40131000030_part1_1	1	3	11
40131000030_part2_73	2	511	2056

2020 Districting Data

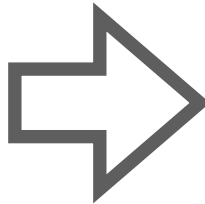
GEOID	District	Dem	Rep
40131000030_part1_30	1	78	316
40131000030_part2_44	2	436	1751

Merge on GEOID



GEOID	District_2010	District_2020	Dem	Rep
40131000030_part1_1	1	n.a.	3	11
40131000030_part2_73	2	n.a.	511	2056
40131000030_part1_30	n.a.	1	78	316
40131000030_part2_44	n.a.	2	436	1751

Melt



GEOID	Dist_2010	Dist_2020	Party	Votes
40131000030_part1_1	1	n.a.	Dem	3
40131000030_part1_1	1	n.a.	Rep	11
40131000030_part2_73	2	n.a.	Dem	511
40131000030_part2_73	2	n.a.	Rep	2056
40131000030_part1_30	n.a.	1	Dem	78
40131000030_part1_30	n.a.	1	Rep	316
40131000030_part2_44	n.a.	2	Dem	436
40131000030_part2_44	n.a.	2	Rep	1751

Aggregate Pattern Calculation

Drop n.a. by column Dist_2010, then sum

GEOID	Dist_2010	Dist_2020	Party	Votes
40131000030_part1_1	1	n.a.	Dem	3
40131000030_part1_1	1	n.a.	Rep	11
40131000030_part2_73	2	n.a.	Dem	511
40131000030_part2_73	2	n.a.	Rep	2056
40131000030_part1_30	n.a.	1	Dem	78
40131000030_part1_30	n.a.	1	Rep	316
40131000030_part2_44	n.a.	2	Dem	436
40131000030_part2_44	n.a.	2	Rep	1751