

# **Backward Mapping: Exploring Questions of Representation via Spatial Analysis of Historical Congressional Districts**

Michael H. Crespin  
[crespin@uga.edu](mailto:crespin@uga.edu)

M.V. (Trey) Hood III  
[th@uga.edu](mailto:th@uga.edu)

Department of Political Science  
The University of Georgia  
104 Baldwin Hall  
Athens, GA 30602

Paper prepared for presentation at the Annual Meeting of the American Political Science Association, Chicago, IL., August 30-Sept. 2. 2007.

## **Introduction**

An important feature of our democratic system of government is the idea that individuals living in legally defined geographic areas will elect representatives to serve their interests in the federal government. As political scientists, we strive to understand how members of Congress behave once elected and how well their behavior matches up with the preferences of constituents. In the larger project, we will begin to study the influence of geography, broadly defined, on congressional representation during a time when our political system and national infrastructure were still developing. More specific to this particular paper, we will first introduce a new dataset of historical congressional districts that was created using geographic information systems (GIS). We then briefly demonstrate a possible use of these data by looking at how geography and choices of electoral systems influenced member-district ideological congruence following the Presidential election of 1828.

Although many papers have studied geographic aspects of districting such as apportionment (see Erikson 1972, Tufte 1973, Cox and Katz 2002) or compactness (Niemi et al. 1990), most of these focused on relatively modern times.<sup>1</sup> Arguably, one reason for the lack of attention to historical eras is the scarcity of appropriate data to answer meaningful questions. Although census data are available from 1790 through the current day, it is recorded at the county, not the district level. This is because the United States does not require congressional districts to match up with pre-existing state political boundaries.

To solve this problem, we are in the process of creating a dataset that will match up counties to congressional districts beginning with the initial set of district boundaries

---

<sup>1</sup> For exceptions, see Engstrom 2005, 2006 and Altman 1998.

that were created for the first Congress in 1788. This will allow us and others to take advantage of historical census data as well as presidential election returns when trying to study representation. In addition, and where we hope to make our main contribution, we have started to digitize the boundary files so they are readable by GIS programs such as ArcMap. Once the boundary files are in electronic form, we can more easily study how certain aspects of district geography can influence members of Congress. For example, do members of Congress who represent small densely populated districts represent their constituents “better” compared to a member whose constituents are spread over a wide geographic area? Also, does distance between a member’s district and the capital matter? Before commercial aviation became feasible, travel and communication with constituents far from the seat of government was slow. Again, we might expect “better” representation from some representatives compared to others. Since districts are not constant over time, the historical nature of this project will provide for an interesting set of quasi-experiments.

### **Geography and Representation**

Now that congressional scholars (Bianco, Spence, and Wilkerson 1996, Carson and Engstrom 2005) have found evidence of an electoral connection as early as 1816, it seems reasonable to study the factors that can influence how well a member is able to vote their district’s preferences in the legislature. We feel geography should have an especially strong influence on representation when travel times were long and information costs were high. In an era well before the advent of the Tuesday-Thursday Club, we can safely speculate that travels back and forth to the district were probably

infrequent occurrences. Further, news or instructions from “back home” probably took some time before it reached a representative in New York, Philadelphia or Washington D.C., depending on the location of the U.S. Capital. Of course, we would expect this to be a larger problem for a member from Kentucky compared a representative whose district is closer to the seat of government.

Previous to the 1960s and the “one person, one vote ruling” in *Wesberry v. Sanders*, there was quite a lot of variation in the number of constituents in congressional districts (Engstrom 2005). This was true both within and across states. We expect that the number of constituents in a district should also have a strong influence on how well a member’s voting matches up with the opinion of the voters. All else equal, as the population of a district increases it should be harder to measure preferences since there are more voters to canvas.

This effect may also be present if the voters are dispersed over a wide geographic area. Although some districts in urban areas were quite small, some were massive in size. For example, the entire state of Maine (then part of Massachusetts) had a population of over 84,000 spread over 31,438 square miles. When population density is low, a member should have a more difficult time representing district opinion. Of course, it is possible that for each of these geographic variables, the causal arrow could run in the opposite direction. It might be the case that constituents living in districts that are far from the capital, have a large number of voters, or reside in a sparsely populated one will have trouble learning what their representative is doing. This would give a member the ability to “shirk” from the voters’ preferences and not be punished. Since we do not look

at electoral effects in this paper, we will not be able to untangle this particular question at this time.

### **Mapping Historical Congressional Districts**

In order to create the dataset used in this and future projects, we relied on two key sources of information to help us match up counties with congressional districts. First, Parsons, Beach and Hermann (1978) *United States Congressional Districts 1788-1841*, provided us with a paper map for each state's congressional districts and their underlying county boundaries. Second, Carville et al's (1999) *Historical United States County Boundary Files 1790 - 1999* served as the base layer for each of the GIS maps.

Figure 1 uses North Carolina as an example to illustrate the basic techniques employed to create this new dataset. The initial step in the process was to import the appropriate county file base map into ArcMap. For the First Congress (1789-1791), we used the county boundary file for the 1790 census. Since the base layers are only available for each census decade, we used the new county file as appropriate after each ten year period.<sup>2</sup> The top map in Figure 1 shows the county boundaries for North Carolina in 1790. Then, in the data attribute table, we coded each county into the appropriate congressional district (middle map). Finally, we used the dissolve command in ArcMap to “erase” the county boundaries for each of the five districts leaving only the district boundaries depicted in the bottom map of Figure 1. So, instead of a map with 54 county polygons, we have a new map with five district polygons. Once the new

---

<sup>2</sup> Unfortunately, new counties were created and old ones were changing on a yearly basis for some states. Generally, the new counties were created out of one parent-county. However, some new counties had multiple parent counties. We relied on the appendix in Parsons, Beach and Hermann (1978) and various web-pages as sources to match up new counties with congressional districts.

boundary files were created for each state, we merged them all together to create a map for each Congress.

We can also take advantage of other features in the GIS program to make the maps more useful. Since our base layer uses county FIPS as identifiers, we can merge in census and other data that was collected at the county level and aggregate it up to the district level. We have also calculated the area and perimeter of each district so various measures of compactness can be created. The program also allows us to measure distance between a point in the district and anywhere else on the map. Finally, we can also designate certain points of interest such as the various placement of the U.S. Capital.

Generally, the maps are quite similar for each apportionment decade, but occasionally states would redistrict between census periods. For example, New York redistricted between the 5<sup>th</sup> (1797-1799) and the 6<sup>th</sup> (1797-1799) Congresses during the same apportionment period and then drew new boundaries again for the 8<sup>th</sup> Congress (1803-1805) after the census of 1800 when it gained 7 additional representatives.<sup>3</sup>

States were also experimenting with electoral systems at this time so the maps could also change as election methods changed. To determine the type of electoral system the state used, we relied on Parson, Beach and Herman (1978) as well as Dubin (1998). If there was a discrepancy between the two sources, we went with the information from Dubin as it is more up to date.

As displayed in Figure 2, we can see that during the 1<sup>st</sup> Congress, three different systems were used. Massachusetts (which included current state of Maine), Rhode Island, New York, Delaware, Virginia (which included West Virginia and Kentucky), North Carolina, and South Carolina all elected their representatives using the now

---

<sup>3</sup> For more on historical redistricting, see Engstrom (2006).

traditional single-member districts. Members representing New Hampshire, Connecticut, New Jersey and Pennsylvania were elected at-large from a statewide vote. The remaining two states, Maryland and Georgia, used a mixed system where members were elected from a statewide vote but they had to be residents of the specific districts designated by state law.<sup>4</sup>

While some states were quite consistent in their methods for electing representatives, others routinely switched back and forth (see Figures 2 through 5). Pennsylvania, for example, used an at-large method for the 1<sup>st</sup> Congress, single member districts for the 2<sup>nd</sup> (1791-1793), at-large for the 3<sup>rd</sup> (1793-1795), and then a mixed system for the 4<sup>th</sup> (1795-1797). In the mixed system, 11 of the 12 districts were single member, except for the 4<sup>th</sup> district where two members were elected. Other states such as Massachusetts and New York occasionally used this type of electoral system for some districts. Because of the variations in electoral systems, the sporadic intra-decade redistricting and new states entering the union, we will provide a map for each Congress.

## **Testing a Spatial Model of Legislator Responsiveness**

### *Data and Measures*

To test the influence of geography and malapportionment on representation, we need a variable that can capture this arguably abstract concept. As our theory suggests, we expect that a member's voting behavior in Congress will deviate from their district median depending on factors related to their district. As such we need to compare measures at both the member and district levels.

---

<sup>4</sup> See Dubin 1998:3 *notes* 1 and 2 for a more detailed description of Georgia and Maryland's electoral systems.

As a proxy for constituency-level preferences, we elected to employ Andrew Jackson's share of the presidential vote in each congressional district for the 1828 Presidential election between Jackson and John Quincy Adams.<sup>5</sup> We chose this election because it was the first election contested between two parties with a sufficient number of states using a popular vote to elect presidential electors. See Figure 6 below for a mapping of the returns at the district level. Overall, Dubin (2002) lists returns for just over 82% of the counties in place by the election. We had results for at least one county in every district, except for three districts, all in Tennessee. In two states, Delaware and South Carolina the electors were chosen by the state legislature so they were not included in the analysis. For the remaining 22 states, we had at least some county level returns that could be aggregated up to the district level.<sup>6</sup> In our analysis below, we also excluded Maine and Massachusetts because many of their districts crossed county borders and we could not disaggregate the county level returns to the sub-county level. We were also forced to eliminate Indiana from our study because the map of the congressional districts in Parsons, Beach and Hermann (1978:223) appears to be a duplicate of the map for the 1830 apportionment decade so we are unable to put many of the northern counties into the correct districts.<sup>7</sup> Finally, there were two counties in Vermont and one in Maryland that crossed district boundaries so the presidential returns were not included in a congressional district. In the future, we hope to accurately split these counties and assign

---

<sup>5</sup> The advantage of employing district presidential vote is that it provides a more direct measure of the partisan or general ideological predisposition of each congressional district separate from the popularity of the incumbent representing the district (Ansolabehere, Snyder, and Stewart 2000, 2001; Jacobson 2000).

<sup>6</sup> Although some states used special districts to select electors, returns were still given at the county level.

<sup>7</sup> The map given for the 21<sup>st</sup> Congress depicts seven congressional districts even though Indiana only had three representatives at the time. We are in the process of contacting state archives to get better information so we can include accurate mappings of these states. However, our results are similar if we include these states in our analysis based on the information that we currently possess.

data in a correct fashion. Although there were some problems, we feel we have enough data to move forward with our analysis.

Next, we measure member's voting behavior using the standard first dimension DW-NOMINATE score (Poole and Rosenthal 1997, 2007) for the Congress immediately following the 1828 election, the 21<sup>st</sup>. DW-NOMINATE theoretically ranges from -1 to 1 with negative scores corresponding to more liberal voting behavior. In order to determine if members were responsive to our measure of district preferences during this early time period, we ran an initial regression with a member's nominate score as the dependent variable and percent Jackson as the independent variable. We found that as the vote for Jackson in a district increases, the member's voting behavior becomes significantly more liberal.<sup>8</sup> As such, we feel we can make a valid claim that presidential vote in the district is providing at least some pressure in terms of a member's voting behavior in the House.

Because we need to measure how far a member deviates from his district median, and presidential vote and DW-NOMINATE are on two different scales, a simple difference is probably not the correct way to create our dependent variable for our models. Instead, we first created z-scores for both measures and then calculated the absolute difference between the two z-scores. As this new variable gets bigger, a member is further away from his district median.<sup>9</sup> This measure serves as the dependent variable in both of our regressions.

---

<sup>8</sup> The coefficient was -.918 and the standard error was .099. We also examined a scatter plot between these two variables and found most members of the Jackson Party were clustered in one section of the graph while the other members of Congress were clustered towards the other end.

<sup>9</sup> We must, of course, be careful in presenting marginal effects since our dependent variable lacks meaningful units of measurement.

In our first model, we have three independent variables that we think influence how far a member will deviate from their district median. Each of the three variables taps into a member's ability to gauge district preferences (or the constituents' ability to keep track of their member's behavior). Our first variable is the *total population* in the member's district measured using the 1830 census. Before the Supreme Court's landmark 1962 decision in *Baker v. Carr* there was significant variation in district population across districts. For example, New York's 15<sup>th</sup> district has 35,870 constituents while the neighboring 14<sup>th</sup> had 71,326 residents. Both of these are "up-state" districts so these differences are not necessarily a function of an urban-rural divide. If we take into account states with at-large districts, the variation becomes even larger. We expect that as the total population in a district increases, the deviation from the district median should increase.

To create the next variable, *population density*, we calculated the ratio of the total population to the size of the district measured in square miles. We expect that members who represent a densely populated district should be closer to the district median so as density increases, our dependent variable should decrease. The third and final variable in our first regression is a measure of the total *distance* between a member's district and the capital. Here, we take the natural log of the number of miles between the geographic centroid of each district and Washington D.C.<sup>10</sup> We argue that as a member gets further away from their district, it becomes harder to gauge district preferences. Therefore, as distance between the district and the capital increases, so should the differences between a member's voting behavior and our measure of district preferences.

---

<sup>10</sup> This is "as the crow flies distance." Ideally we would have a measure that takes into account both land and water routes but those data are unavailable at this time.

Our next regression looks into election methods instead of strictly geographic measures. Here we have three measures: *at-large district*, *single member district* and *multi-member district*. For the 21<sup>st</sup> Congress, five states (Connecticut, Georgia, New Hampshire, New Jersey and Rhode Island) with a total of 27 members elected multiple representatives with a state-wide vote. Two states, (New York and Pennsylvania) had a few districts where multiple members, 20 all together, were elected to serve the same district. The remaining 133 members in our dataset represented single-member districts.

We expect that members from at-large districts should deviate more from their district median compared to single member districts for two reasons. First, it is possible for different members in at-large district to form different winning coalitions the same way senators can be elected from two parties to serve the same state (Fiorina 1992, Brunell and Grofman 1998). Second, due to the relatively large size of at-large districts, members may have more difficulty in estimating district preferences. Our expectations regarding multi-member districts are not as clear. In some ways, these districts are like small at-large districts so members can also form different election constituencies. However, because these districts are relatively smaller, it may be easier to determine the district ideal point. To facilitate interpretation, we run this regression with all three variables included and suppress the constant.

## **Findings**

The results of the first regression model outlined are in Table 1. Both the total population of the district and the distance from the district to the capital are positively and significantly related to increased levels of ideological deviance between the member

and their constituents. In an age where both communication and transportation were relatively slow, it may well be that constituents further removed from Washington, geographically speaking, may have had a more difficult time tracking representative behavior at the capitol. This information asymmetry created by distance may have allowed representatives a greater amount of voting leeway.

<Table 1 about here>

Likewise, legislators representing districts that contained comparatively higher population counts were also more likely to deviate from their constituents' preferences. In addition to violating the *one person, one vote* standard, malapportioned districts might also produce differential effects related to representation. As the federal courts have refused to allow any deviance in regard to population counts for congressional districts within states since the early 1960's, examining the linkage between constituent preferences and legislator behavior can only be studied in the historical realm. This analysis offers some insight into the manner in which malapportionment may have affected congressional representation in substantive terms.

Finally, population density is negatively related to the ability of legislators to diverge from the preferences of their district as measured by the presidential vote percentage cast for Andrew Jackson. In this case, population density may be viewed as a proxy for the degree to which an electoral district is urbanized.<sup>11</sup> Higher levels of urbanization should, in turn, produce the ability for greater intra-district communication via newspaper production and dissemination and by word-of-mouth. We might also note the model explains only 14% of the variance in the ideological deviance measure, an

---

<sup>11</sup> Indeed, if percent urban is substituted for population density the coefficient on the variable is also negative and statistically significant. However, we feel that population density better captures this concept.

indication that other factors, outside geography, certainly play a role in legislator responsiveness.

Next, we specify a second model designed to examine how district typology affects the ideological divergence of House members from their constituents. Due to issues relating to multicollinearity, we were unable to include both the spatial variables presented in Table 1, along with the variables designed to denote the three types of congressional districts utilized in 1828. By suppressing the intercept in the regression model we are able to include all three election-type indicators, the results which are reported in Table 2.

<Table 2 about here>

All three of the election-type variables were shown to be significant predictors of ideological congruence between representatives and their constituents. As measured by the size of the coefficient and consistent with our expectations, representatives elected from at-large districts showed the largest amount of ideological deviance from their constituents. Somewhat counter to expectations, in comparison to members elected from single-member districts, those elected from multi-member districts actually demonstrated closer congruence with district preferences.

The results presented in Tables 1 and 2 indicate that both spatial indicators and the type of electoral system utilized have a direct effect on the relationship between legislators' voting patterns in Congress and preferences of their constituents. In order to more fully flesh out the tentative relationship found to exist based on the 1828 election, a more stringent set of control variables should be utilized along with the addition of more temporal observations (presidential elections). We would expect to find that as

technology improves that can enhance communication between representatives and their constituents, the effects of geography should dissipate.

## **Discussion**

Recently, political scientists have begun to use Geographic Information Systems (GIS) to study political phenomenon. Through these studies we have learned about interstate conflict (Berry and Baybeck 2005), electoral competition (Crespin 2005) and turnout (Darmofal 2006). However, these studies have either been tied to a particular time-period (i.e. the modern era) or a specific geographic context (i.e. the county). In regard to congressional studies in particular, much of the empirical literature in this area focuses on the post World War II-era as data are not readily available from earlier historical time-periods for analysis.

The overall goal of this project is to create a dataset suitable for spatial analyses of congressional districts. In addition to literally mapping districts in an electronic format, we also plan to match political and demographic data to these districts, making historical empirical analyses a possibility. Only by taking these steps will congressional researchers be able to compare theories developed in the post-World War II time period with previous historical periods. In addition, some questions, such as the one briefly analyzed in this manuscript relating to malapportionment and representation can only be analyzed in an historical setup.

## References

- Altman, Micah. 1998. "Traditional Districting Principles: Judicial Myths vs. Reality." *Social Science History* 22: 159-200.
- Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart, III. 2000. "Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage." *American Journal of Political Science* 44 (January): 17-34.
- Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart, III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45 (January): 136-159.
- Berry, William, and Brady Baybeck. 2005. "Using Geographic Information Systems to Study Interstate Competition." *American Political Science Review*, 99:505-19.
- Bianco, William T., David B. Spence, and John D. Wilkerson. 1996. "The Electoral Connection in the Early Congress: The Case of the Compensation Act of 1816." *American Journal of Political Science* 40: (February):145-71.
- Brunell, Thomas L. and Bernard Grofman. 1998. "Explaining divided U.S. Senate delegations. 1988-1996: A realignment approach." *American Political Science Review* 92:391-411.
- Carson, Jamie L. and Erik Engstrom. 2005. "Assessing the Electoral Connection: Evidence from the Early United States," *American Journal of Political Science* 49(4): 746-757
- Cho, Wendy K. 2003. "Contagion Effects and Ethnic Contribution Networks." *American Journal of Political Science* 47(2): 368-87.
- Cox, Gary W. and Jonathan N. Katz. 2002. *Elbridge Gerry's Salamander: The Electoral Consequences of the Reapportionment Revolution*. New York: Cambridge University Press.
- Crespin, Michael H. 2005. "Using Geographic Information Systems to Measure District Change, 2000-02," *Political Analysis* 13(3): 253-260.
- Darmofal, David. 2006. "The Political Geography of Macro-Level Turnout in American Political Development." *Political Geography* 25(2): 123-150.
- Dubin, Michael J. 1998. *United States Congressional Elections, 1788-1997: The Official Results*. Jefferson, North Carolina: McFarland & Company.

- Dubin, Michael J. 2002. *United States Presidential Elections, 1788-1860: The Official Results by County and State*. Jefferson, North Carolina: McFarland & Company.
- Earle, Carville, Changyong Cao, John Heppen, and Samuel Otterstrom. 1999. *The Historical United States County Boundary Files 1790 - 1999* on CD-ROM. Geoscience Publications. Louisiana State University.
- Engstrom, Erik. 2006. "Stacking the States, Stacking the House: The Partisan Consequences of Redistricting in the 19th Century." *American Political Science Review* 100:419-427.
- Engstrom, Erik. 2005. "The Partisan Impact of Malapportionment on the 19<sup>th</sup> Century and Early 20<sup>th</sup> Century House of Representatives." Paper presented at the Midwestern Political Science Association Annual Meeting, Chicago. IL.
- Erikson, Robert S. 1972. "Malapportionment, Gerrymandering, and Party Fortunes." *The American Political Science Review* 66: 1234-1245.
- Fiorina, Morris P. 1992. *Divided Government*. New York: Macmillan.
- Gimpel, James G., Frances E. Lee and Joshua Kaminski. 2006. "The Political Geography of Campaign Contributions in American Politics," *The Journal of Politics* 68(3): 626-639.
- Jacobson, Gary C. 2000. "Party Polarization in National Politics: The Electoral Connection." In *Polarized Politics: Congress and the President in a Partisan Era*. Jon R. Bond and Richard Fleisher, editors. Washington, DC: CQ Press, 9-30.
- Niemi, Richard G., Bernard Grofman, Carl Carlucci, and Thomas Hofeller. 1990. "Measuring Compactness and the Role of a Compactness Standard in a Test for Partisan and Racial Gerrymandering," *The Journal of Politics*, 52(4): 1155-1181.
- Parsons, Stanly B., William W. Beach and Dan Hermann. 1978. *United States Congressional Districts 1788-1841*, Westport, CT: Greenwood Press:.
- Poole, Keith T. and Howard Rosenthal. 1997. *Congress: A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Poole, Keith T. and Howard Rosenthal. 2007. *Ideology and Congress*. New Brunswick, N.J.: Transaction Publishers.
- Tufte, Edward R. 1973. "The Relationship between Seats and Votes in Two-Party Systems." *The American Political Science Review* 67:540-554.

Table 1. Explaining Legislator Ideological Deviance using Geography

	Coefficient
Total Population	.0000024 *** (.0000005)
Population Density	-.000143 *** (.000037)
Distance to Capitol (Logged)	.1616* (.0748)
Constant	.1856 (.3979)
R <sup>2</sup>	.14
N	180

Notes: Entries are regression coefficients with robust standard errors in parentheses.

\*p<.05; \*\*p<.01; \*\*\*p<.001

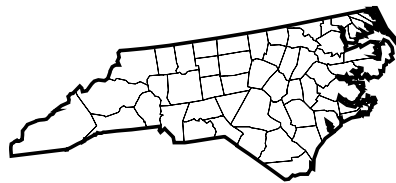
Table 2. Explaining Legislator Ideological Deviance using District Typology

	Coefficient
At-Large	2.0088*** (.1395)
Multi-Member	.8589*** (.1175)
Single-Member	1.2522*** (.0661)
R <sup>2</sup>	.78
N	180

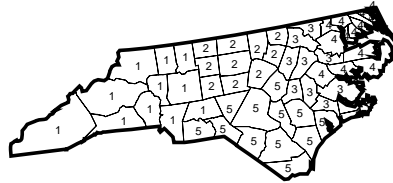
Notes: Entries are regression coefficients with robust standard errors in parentheses. Model intercept is suppressed.  
\*p<.05; \*\*p<.01; \*\*\*p<.001

Figure 1 - District Map Example

North Carolina - 1790 Counties



North Carolina 1790 - Counties by Congressional District



North Carolina Districts - 1st Congress

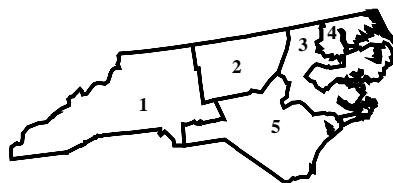


Figure 2 - 1st Congress (1789 - 1791)

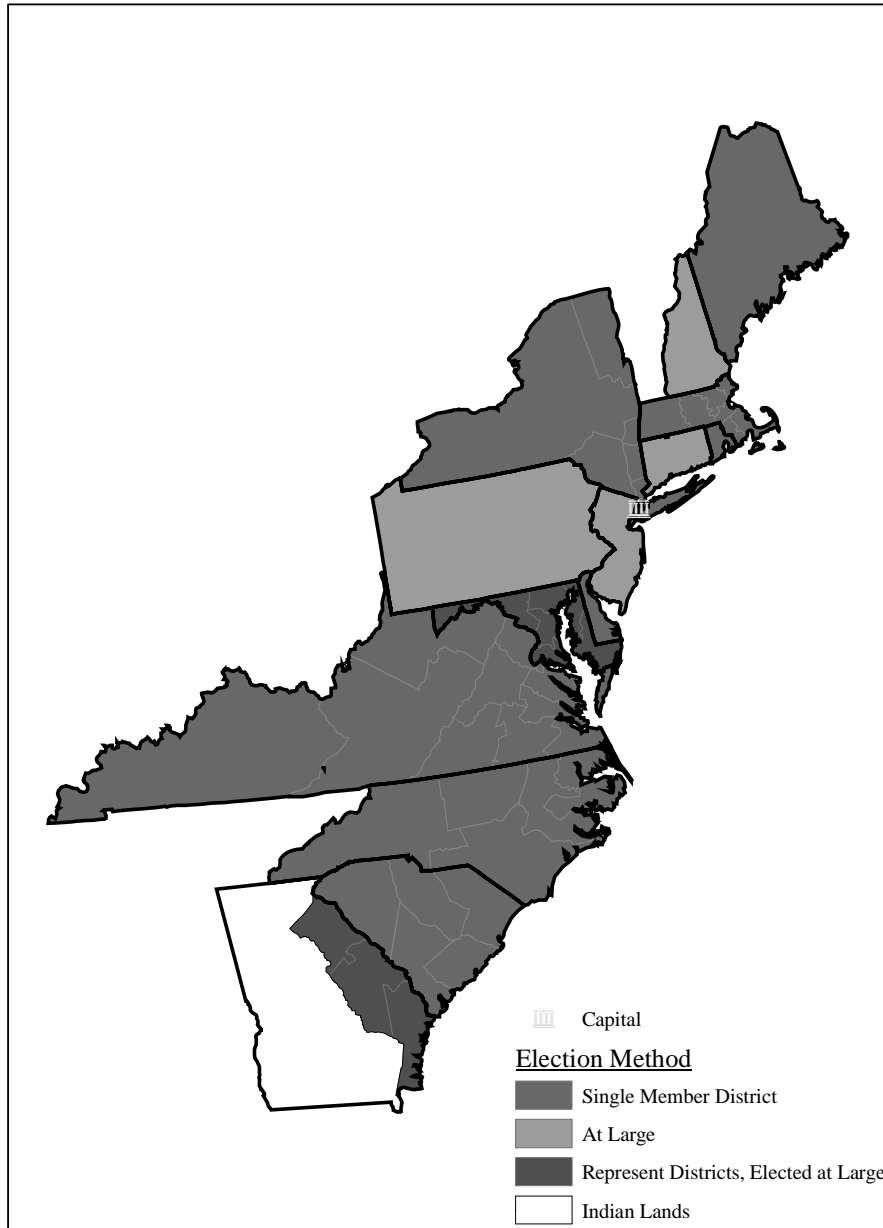
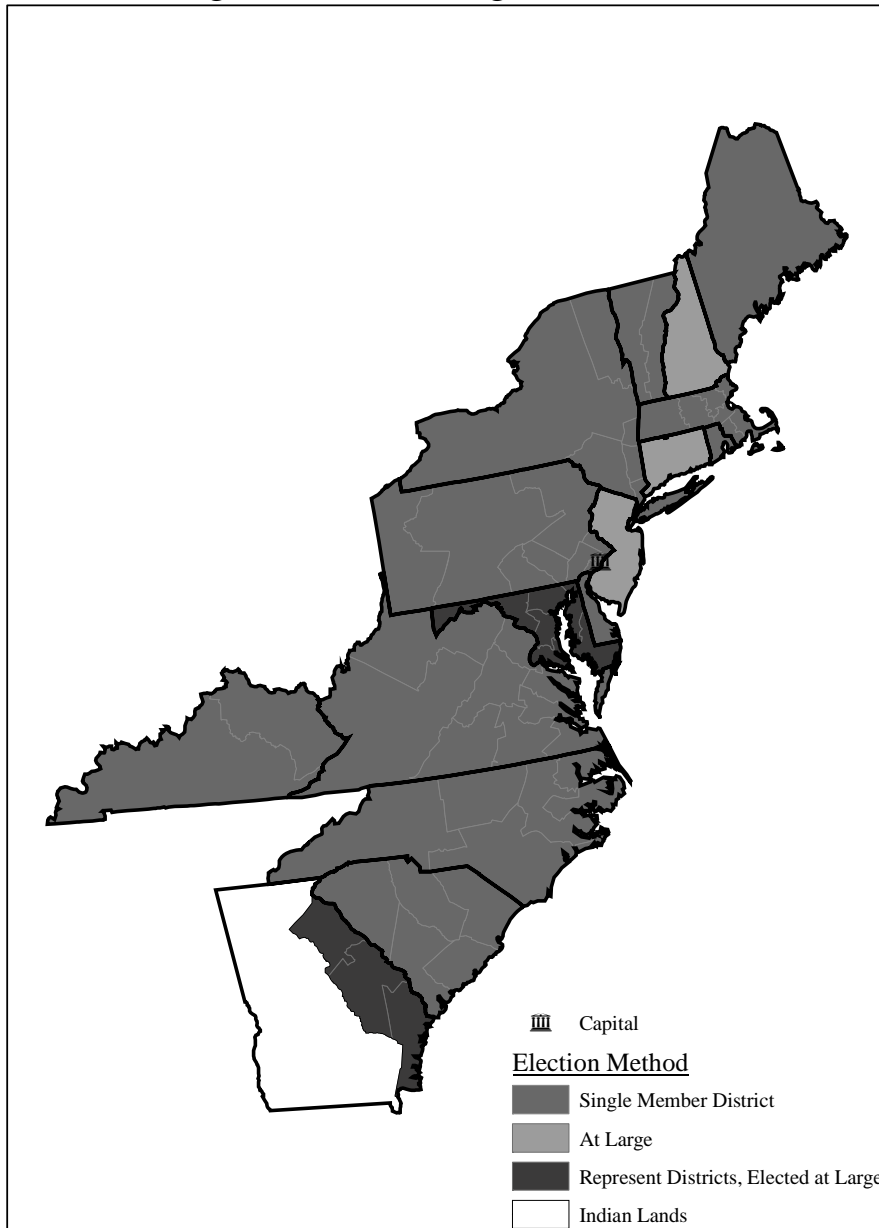


Figure 3 - 2nd Congress (1791-1793)



Note - Kentucky became a state on June 1, 1792 and held elections in Sept. of that year. Previously it was the 2nd district of Virginia. Vermont became a state on March 4, 1791.

Figure 4 - 3rd Congress (1793 - 1795)

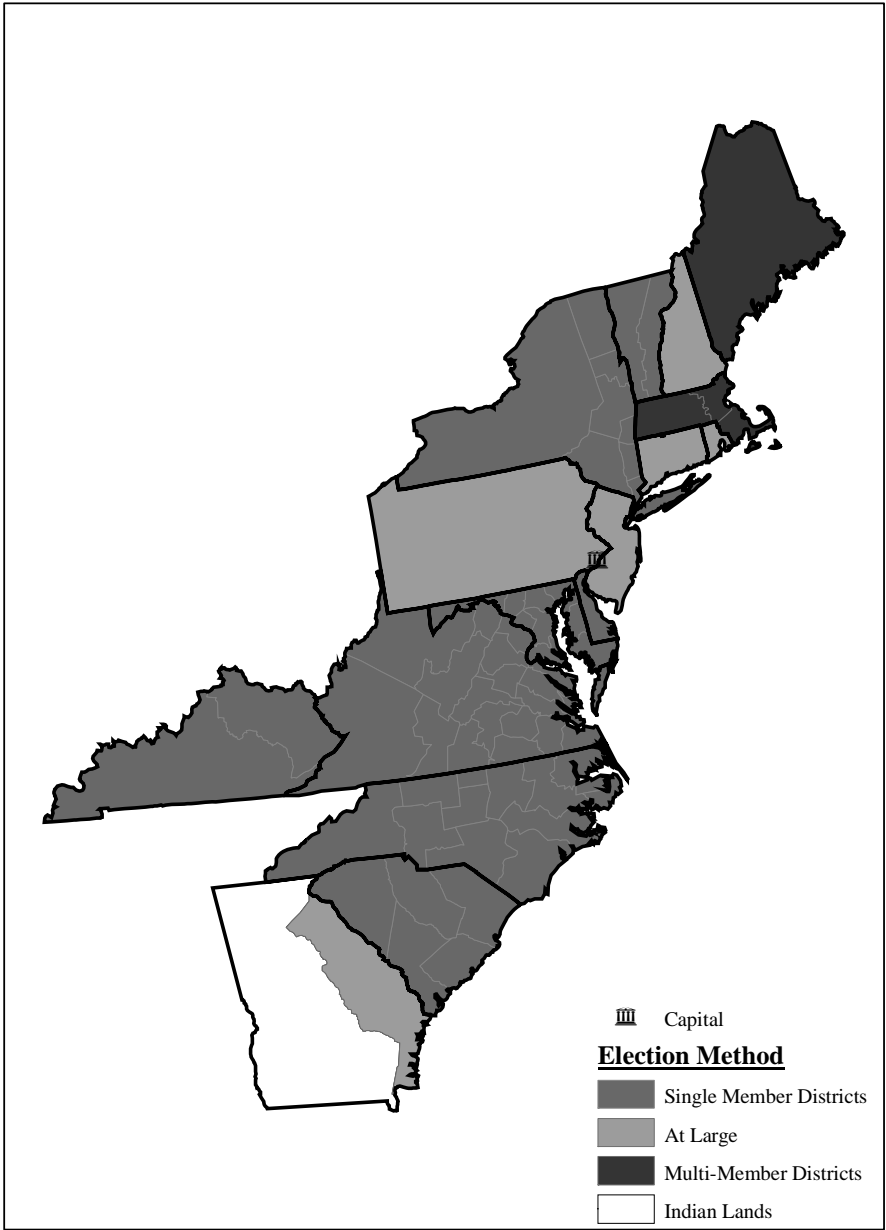


Figure 5 - 4th Congress (1795 - 1797)

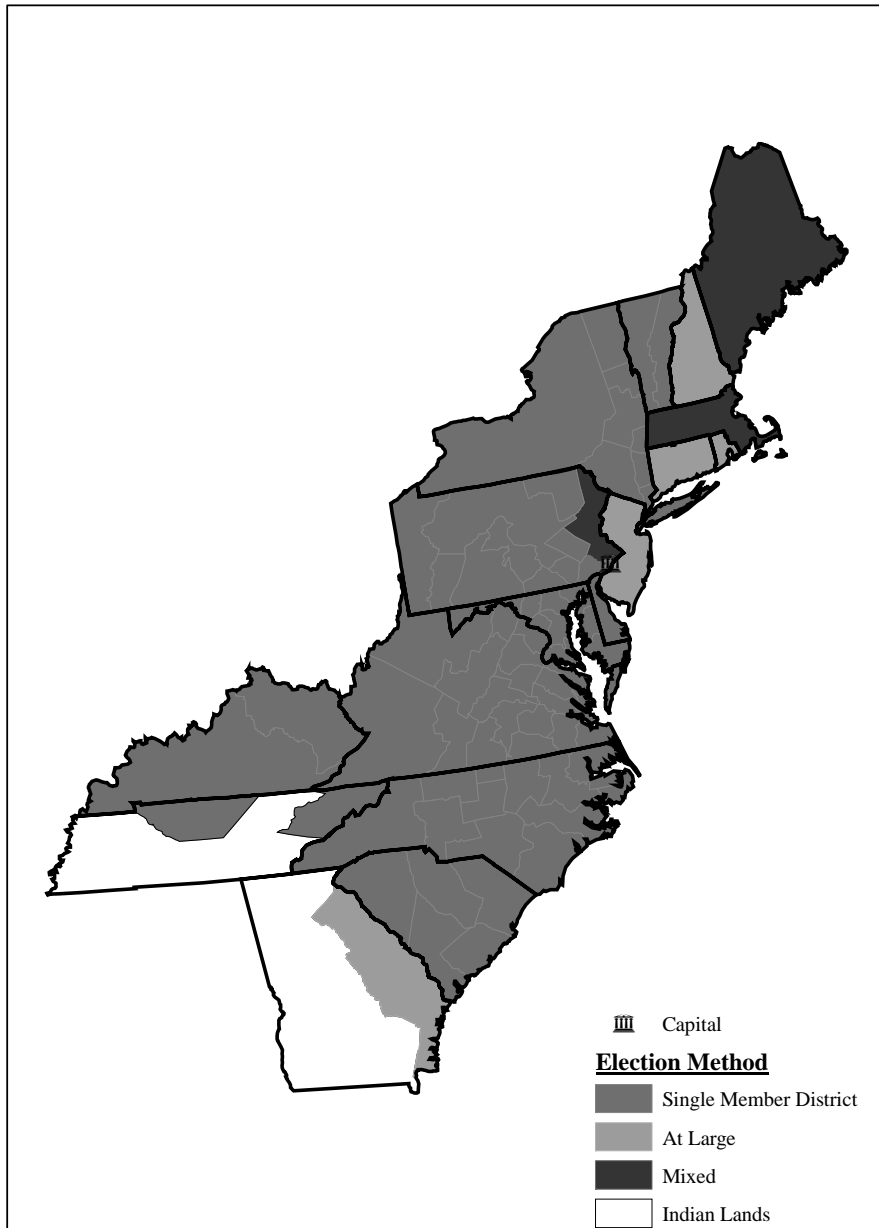
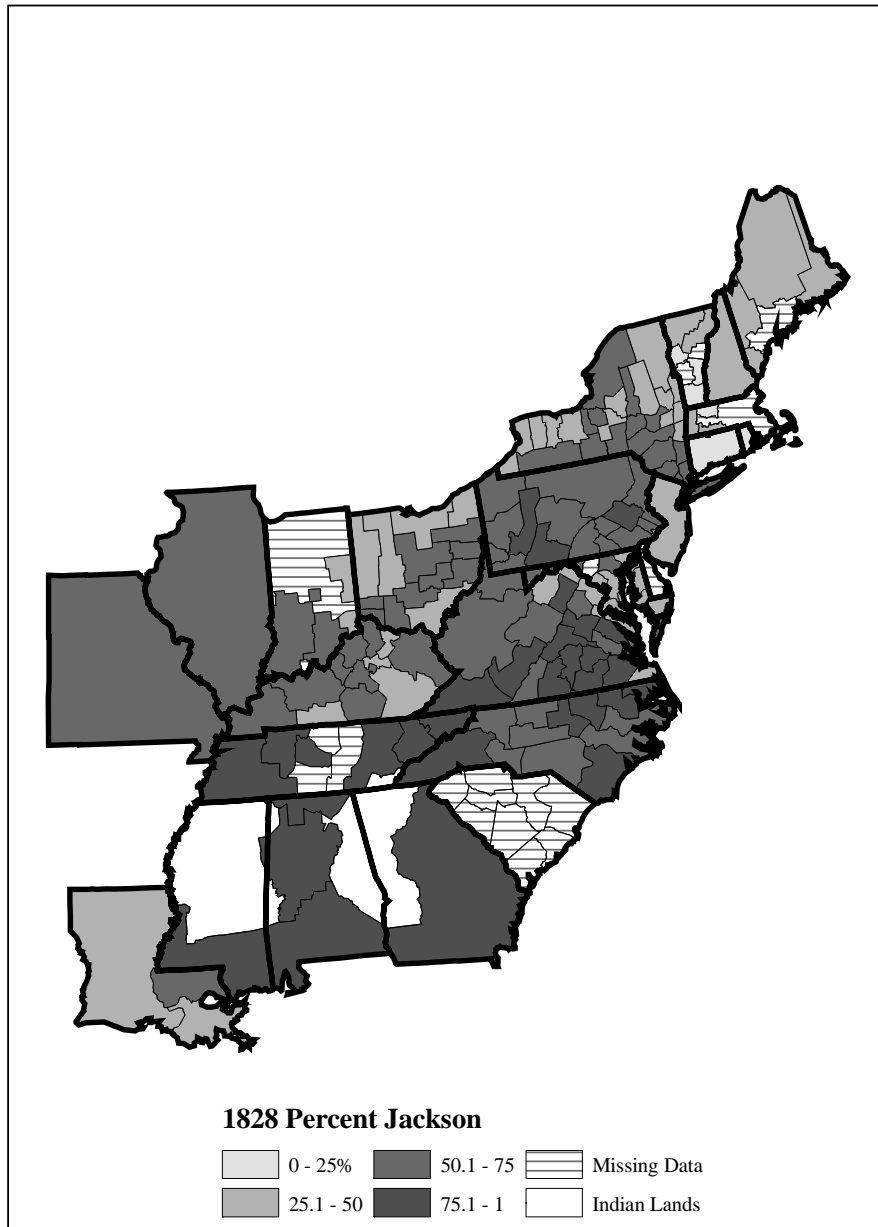


Figure 6 - 1828 Presidential Election Results by 21st Congressional Districts



Note - Electors for Delaware and South Carolina chosen by the legislature.  
 Electors for Tennessee chosen by Presidential Elector Districts with county results not given for all districts.  
 Maine, Massachusetts, Maryland and Vermont contained Congressional districts that split county lines.  
 Indiana district boundaries are only partially available for this Congress.