

## Further Size-Density Research

Table 15-1, adapted from my 1970 doctoral dissertation, shows a similar distribution of county seats by 2°. quadrant. In that work I suggested it might be a topic for further research, paralleling the study of animal distributions.<sup>[1]</sup>

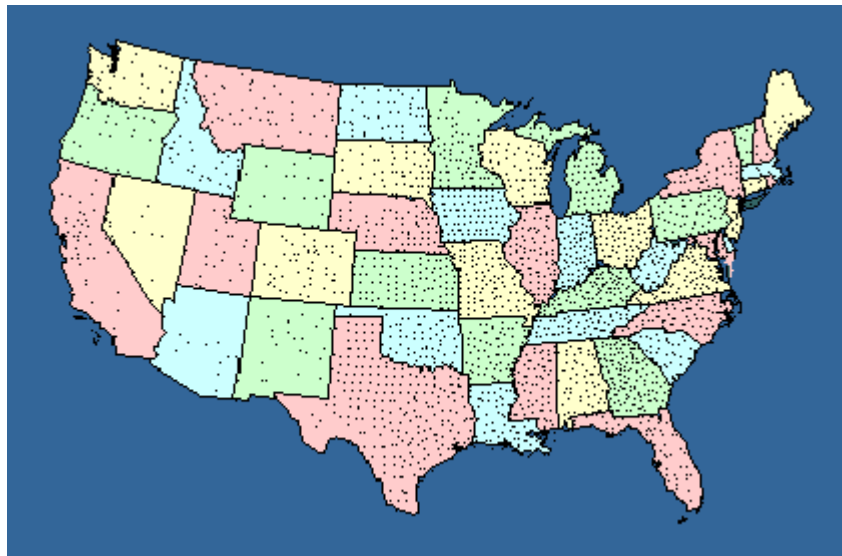
Furthermore, the study could be conducted historically since I had the founding date of each county.<sup>[3]</sup> All that would be missing would be those few counties which had been created but were no longer in existence (like Umpqua County, Oregon). Again, this wasn't much of a problem since the number

of counties ever dissolved, though unknown, was probably very small in relation to the 3,000-plus still surviving.

What I wanted to do was to get from the usual state- by-state distribution (Fig. 15-1a) to a quadrant-by-quadrant one (Fig. 2-a). It took a decade to get around to it.

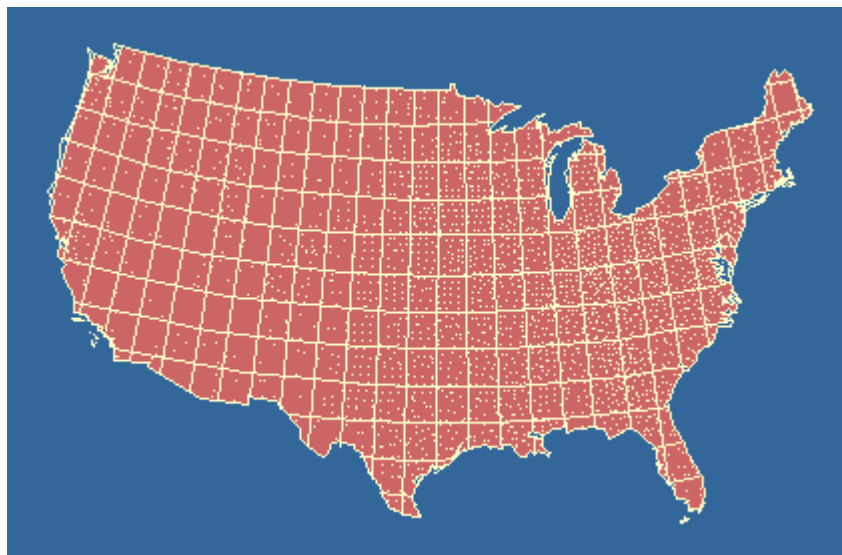
**FIG. 15-1a. COUNTY SEATS IN EACH STATE**

each dot is a county seat



**FIG. 15-1b. COUNTY SEATS IN EACH 2° QUADRANT**

each dot is a county seat



In 1981 Rob Moe, by then a graduate student,<sup>[4]</sup> was looking for a thesis topic. He took on the geographic quadrants project. Quadrant areas were computed as indicated in Chapter 11. The Demographic Research Lab in our Sociology Department had accumulated all the county populations from

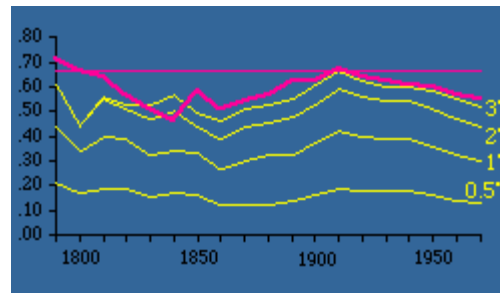
1790 on. So quadrant "county densities" could be regressed against quadrant "population densities." The results are shown in Table 15-2 and Figs. 15-2,3.

**TABLE 15-2. DENSITY-DENSITY SLOPES  
FOR GEOGRAPHIC QUADRANTS**

year	0.5°	1°	2°	3°	4°	5°	6°	states
1790	.21	.44	.61	.61	.70	.75	.93	.71
1800	.17	.34	.44	.44	.47	.47	.46	.66
1810	.19	.40	.55	.56	.63	.64	.60	.64
1820	.19	.39	.51	.53	.62	.67	.57	.57
1830	.15	.32	.47	.53	.58	.64	.62	.51
1840	.17	.34	.50	.57	.56	.66	.64	.46
1850	.16	.33	.44	.49	.51	.62	.53	.58
1860	.12	.27	.39	.46	.46	.49	.58	.51
1870	.12	.30	.44	.51	.52	.58	.55	.54
1880	.12	.32	.45	.53	.55	.61	.63	.57
1890	.14	.32	.48	.55	.60	.65	.68	.62
1900	.16	.37	.53	.61	.64	.67	.70	.62
1910	.19	.42	.59	.66	.70	.73	.74	.67
1920	.18	.40	.56	.62	.65	.70	.71	.64
1930	.18	.39	.54	.60	.63	.68	.68	.62
1940	.18	.39	.54	.60	.63	.68	.67	.61
1950	.16	.36	.51	.58	.61	.67	.65	.60
1960	.14	.32	.47	.55	.58	.65	.64	.57
1970	.13	.30	.44	.52	.56	.63	.61	.55

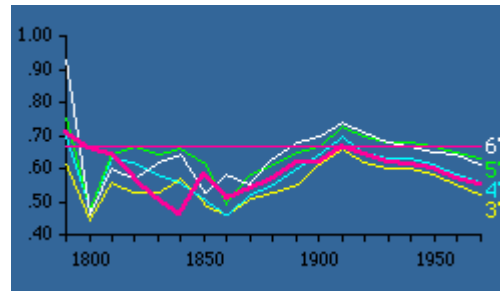
As can be seen, very small quadrants show little change in slopes over time. In fact, small quadrants produce slopes uniformly near zero. The reason for this is that, at an extremely small size, a quadrant has either 1 or 0 county seats. Zero values can't be included in logarithmic transformation, so the number of seats per usable quadrant tends toward a constant. As we increase quadrant size, greater variation becomes possible. By the time we reach quadrants of 2-3° the pattern becomes stable, mirroring the earlier pattern found using states as units of analysis (the pink line in Fig. 15-2,3).

**FIG. 15-2. DENSITY-DENSITY SLOPES  
0.5-3° QUADRANTS**



We could, of course go on increasing quadrant size, but the result would only be a return toward zero since, again, there would be little variation in the dependent variable. The ideal size, for our purposes at least, seems to be quadrants of about 3-5°.

**FIG. 15-3. DENSITY-DENSITY SLOPES  
3-6° QUADRANTS**



These results show that essentially the same methods employed in analyzing predator-prey densities (Chapter 11) can be employed in studying the distribution of a social "species", county governments, in relation to the distribution of its "food supply", the human population which gives it life through taxation (ultimately of its time resources).

Perhaps more importantly, they demonstrate that the far easier approach used in earlier studies, with data aggregated into such "artificial" units as nations or states, give virtually identical outcomes. Simultaneously, we have clearly avoided any charge of "tautology" with this approach.

### **A Non-Ratio Test**

The four equations derived in Chapter 10 were each derived from

$$(*10-10) N = P^{2/3} A^{1/3}$$

In the Summer and Fall of 1983 I supervised thesis research by Dan Dorman which tested the logarithmic form of this equation

$$(1) \quad \log N = K + \frac{2}{3} \log P + \frac{1}{3} \log A$$

Here we have two independent variables; the obvious test is through multiple regression.<sup>[5]</sup>

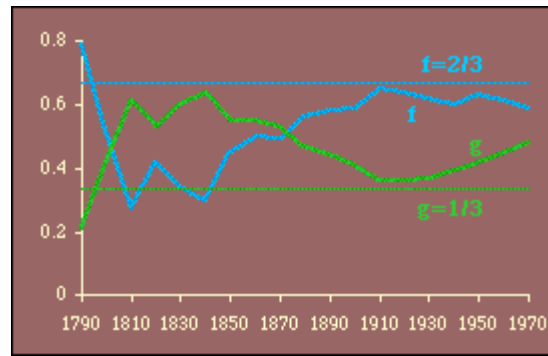
The results, computed with P representing the *rural* population, are shown in Table 15-3 and Figs. 15-4,5 and 6.

**TABLE 15-3. STATISTICS FROM THE MULTIPLE REGRESSION TEST**

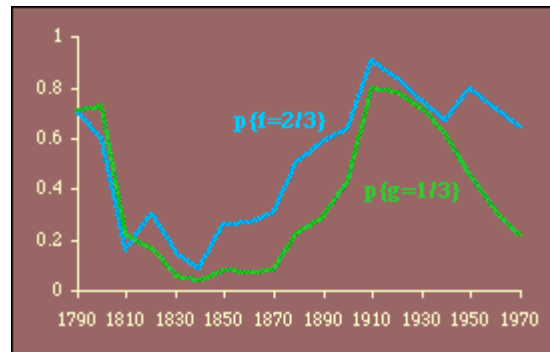
year	f	p{f=2/3}	g	p{g=1/3}	R <sup>2</sup> <sub>N,PA</sub>	r <sup>2</sup> <sub>PA</sub>
1790	.79	.70	.21	.71	.80	.52
1800	.51	.59	.41	.73	.82	.61
1810	.27	.16	.61	.22	.88	.79
1820	.42	.30	.53	.17	.81	.37
1830	.34	.15	.60	.06	.82	.53
1840	.30	.09	.64	.04	.84	.62
1850	.45	.26	.55	.08	.84	.32
1860	.50	.27	.55	.07	.85	.40
1870	.49	.31	.53	.08	.85	.40
1880	.56	.51	.47	.23	.88	.41
1890	.58	.59	.44	.29	.89	.22
1900	.59	.64	.41	.43	.91	.28
1910	.65	.91	.36	.80	.91	.39
1920	.64	.84	.36	.78	.89	.37
1930	.62	.75	.37	.72	.86	.37
1940	.60	.67	.39	.61	.85	.36
1950	.63	.80	.42	.45	.85	.30
1960	.61	.72	.45	.31	.82	.27
1970	.59	.65	.48	.22	.80	.21

It is evident that these results generally conform to patterns observed earlier. That is, the best theoretical fits are prior to territorial expansion and at the close of the frontier, around the years 1800 and 1900. The multiple correlation coefficient is quite high throughout the period, peaking in 1900 and 1910 and eroding since that time.

**FIG. 15-4. REGRESSION COEFFICIENTS**  
f is for log P; g is for log A

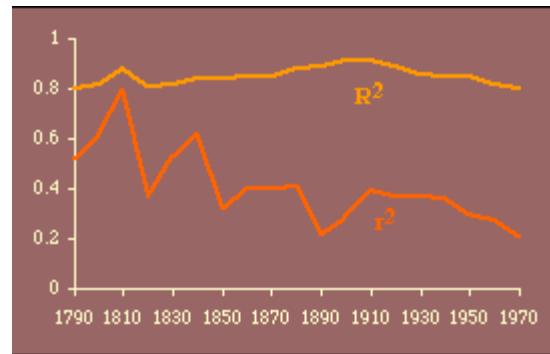
**FIG. 15-5. COEFFICIENT PROBABILITIES**

f is for log P; g is for log A



One difficulty introduced by this test is the possibility of multicollinearity.<sup>[6]</sup> High multicollinearity would result in wide confidence intervals for the coefficients and, therefore, a lessened probability of rejecting an erroneous hypothesis. Every year the independent variables are positively related,  $p\{r^2=0\} < .05$ . Whether this represents high multicollinearity is difficult to say, though it may explain why at its highest value (in 1810,  $r^2 = .79$ ) the hypothesis is supported by the probability measures  $p\{f\}$  and  $p\{g\}$  even though their actual values are opposite from expectations (i.e.,  $f$ , which should equal  $2/3$ , is .27 while  $g$ , which should be  $1/3$ , is .61).

**FIG. 15-6. COEFFICIENTS  
OF DETERMINATION (r)  
AND MULTIPLE CORRELATION (R)**



### A Test with "Raw Data"

In another project with Dan Dorman, we summed both sides of Eq. 10-10

$$(2) \quad \sum N = k \sum P^{2/3} A^{1/3}$$

From this we obtained a value for k

$$(3) \quad k = \frac{\sum N}{\sum P^{2/3} A^{1/3}}$$

producing an expression for  $N^*$ , the **expected** value of N, given values for P and A.

$$N^*_i = \frac{\sum N}{\sum P^{2/3} A^{1/3}} P_i^{2/3} A_i^{1/3}$$

$$(4) \quad N^*_i = k P_i^{2/3} A_i^{1/3}$$

This seemed a fairly severe test — exponents are simply imposed on P and A; and, though k comes from the data, there is no guarantee of error-minimization (as there would be, e.g., with least-squares estimators). We used state rural populations for P; N and A were the number of counties and square miles in the state; results are shown in Table 15-4 and Fig. 15-7.

**TABLE 15-4. A TEST WITH "RAW DATA"**

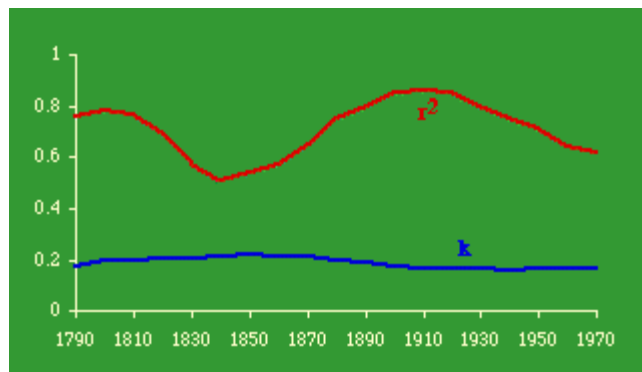
year	k	$r^2$
1790	0.000178	0.76
1800	0.000195	0.78

1810	0.000201	0.77
1820	0.000203	0.70
1830	0.000208	0.57
1840	0.000216	0.51
1850	0.000218	0.54
1860	0.000214	0.57
1870	0.000212	0.65
1880	0.000194	0.75
1890	0.000189	0.80
1900	0.000177	0.85
1910	0.000169	0.86
1920	0.000169	0.85
1930	0.000165	0.80
1940	0.000159	0.75
1950	0.000165	0.71
1960	0.000167	0.64
1970	0.000169	0.62

The value of  $k$  doesn't change appreciably, though lower values tend to be associated with the two periods of best fit in earlier tests, around 1800 and again in 1900. The  $r^2$  values show very clear conformity to the earlier pattern.

**FIG. 15-7. A TEST WITH "RAW DATA"**

$k$  is the constant ( $\times 1000$ );  $r$  is the correlation coefficient



The results of the non-logarithm, non-regression test and the multivariate test were published together.[7]

[Next Chapter](#)

## NOTES:

[1] The Wynne-Edwards map (Fig. 11-3) and Table 15-1 appear on pages 104 and 105 of the dissertation. It is more than an analogy to say that the taxable population provides the food supply for the (predatory? parasitic?) species, county



government.

[2] Bill Gossman, a Computer Science undergraduate and the son of my good friend and colleague Charles Gossman, had coded the locations of all 3,064 county seats for me as a summer project.

[3] from Joseph Nathan Kane, *The American Counties* (rev. ed.), New York: Scarecrow Press, 1962.

[4] As an undergraduate in 1978 he had helped me with the study of city areas (Chapter 9).

[5] *multiple regression* will be covered in the appendix

[6] correlation among  $\log P$  and  $\log A$ , the presumably "independent" variables (which should be independent of one another).

[7] G. Edward Stephan and Dan Dorman, "Testing Size-Density Relationships without Ratio-Variables, Multicollinearity, Logarithmic Transformations, or Regression Analysis." *Journal of Regional Science*, 25:427-35, 1985.