

STATISTICAL GRAPHICS FOR VISUALIZING DATA

Tables and Figures, I

William G. Jacoby
Michigan State University and ICPSR

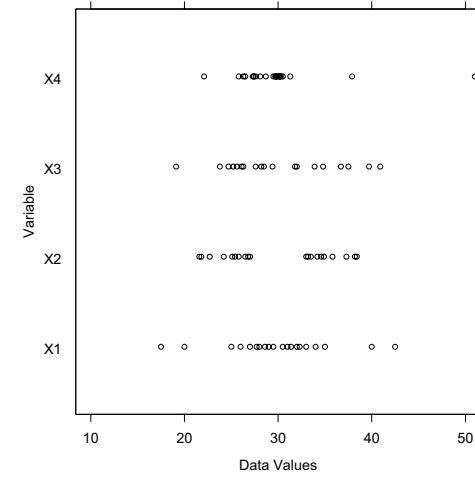
University of Illinois at Chicago
October 14-15, 2010

<http://polisci.msu.edu/jacoby/uic/graphics>

Table 1: Hypothetical Data Matrix Containing Four Variables and Twenty Observations.

Observation:	X ₁	X ₂	X ₃	X ₄
1	32.3	33.2	24.7	29.7
2	28.0	34.2	29.4	30.2
3	31.4	27.0	28.5	28.7
4	29.5	33.0	25.6	27.3
5	40.0	35.8	27.6	31.3
6	20.0	34.6	32.0	29.5
7	26.0	24.2	28.2	26.3
8	28.6	34.9	40.9	29.9
9	27.7	25.1	37.5	29.8
10	27.0	37.3	26.3	30.1
11	17.5	22.7	33.9	37.9
12	31.0	25.4	36.7	27.6
13	32.0	25.8	25.2	30.3
14	30.5	38.2	23.8	22.1
15	34.0	26.5	26.1	28.1
16	42.5	38.4	28.2	26.5
17	35.0	26.8	31.8	30.5
18	29.0	21.6	39.7	27.4
19	25.0	33.5	19.1	51.0
20	33.0	21.8	34.8	25.8
Mean:	30.0	30.0	30.0	30.0
Standard Deviation:	5.8	5.8	5.8	5.8

Figure 1: Univariate Scatterplots for Distributions of Hypothetical Variables.



Data Source: Table 1

Figure 2: Example of a Graphical Display for Presentation, Rather Than Analysis.

1992 Presidential Vote Totals:

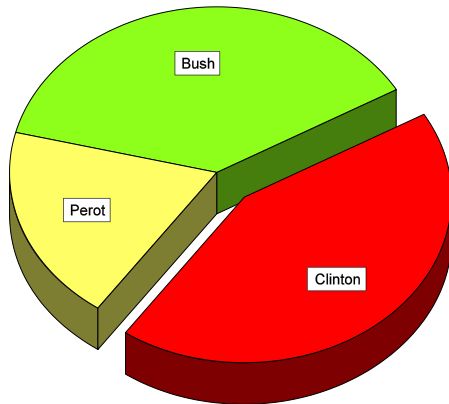


Figure 3: Candidate Vote Percentages in 2000 U.S. Presidential Election

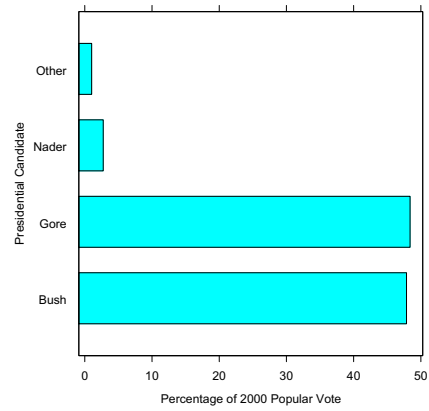
A. Tabular Display

Presidential Candidate	Percentage of Popular Vote	Popular Vote (in Millions)
Bush	47.87	50.46
Gore	48.39	51.00
Nader	2.74	2.88
Other	1.01	1.06
Total	100.00	105.40

Data Source: U.S. Federal Election Commission

Figure 3: Candidate Vote Percentages in 2000 U.S. Presidential Election

B. Graphical Display



Data Source: U.S. Federal Election Commission

Figure 4: Partisanship of State Electorates, 1992.

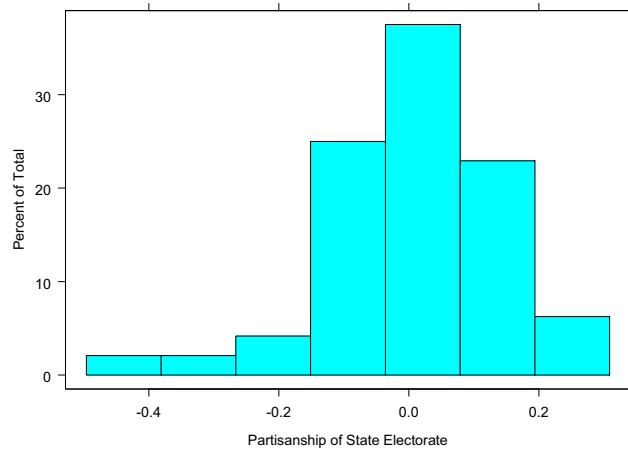
A. Tabular Display

States and Partisanship Scores*			
WY	-0.488	CA	0.023
UT	-0.322	NM	0.029
ID	-0.260	AL	0.029
SD	-0.161	MN	0.031
NE	-0.140	OH	0.031
AZ	-0.133	PA	0.042
VT	-0.109	WI	0.046
DE	-0.091	MO	0.052
NV	-0.090	NY	0.053
VA	-0.089	TX	0.055
ND	-0.068	ME	0.092
NH	-0.064	IA	0.100
OR	-0.054	NC	0.107
MS	-0.045	MT	0.108
NJ	-0.044	RI	0.128
MI	-0.044	WV	0.130
CT	-0.022	TN	0.142
IN	-0.008	OK	0.156
SC	-0.007	MA	0.158
KS	0.006	GA	0.180
FL	0.007	MD	0.185
IL	0.013	LA	0.261
CO	0.019	AR	0.267
WA	0.021	KY	0.301

Data Source: Gerald C. Wright.

Figure 4: Partisanship of State Electorates, 1992.

B. Graphical Display



Data Source: Gerald C. Wright.

Figure 5: Party Affiliation by Age Groups within the American Electorate, 1992.

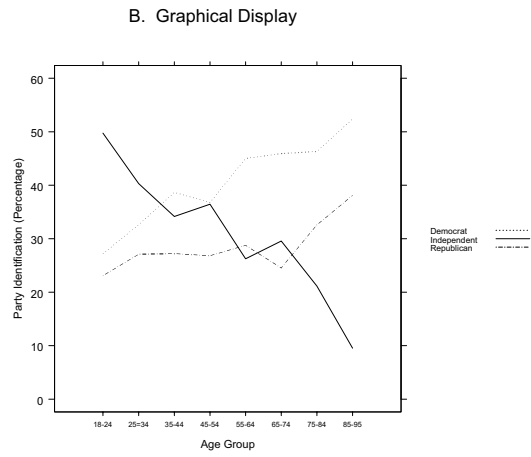
A. Tabular Display

Age Group:	Party Affiliation:*		
	Democrats	Independents	Republicans
18-24	27.18	49.74	23.08
25-34	32.62	40.29	27.09
35-44	38.65	34.15	27.20
45-54	36.74	36.46	26.80
55-64	45.00	26.25	28.75
65-74	45.91	29.57	24.51
75-84	46.29	21.14	32.57
85-94	52.38	9.52	38.10

* Table entries are row percentages.

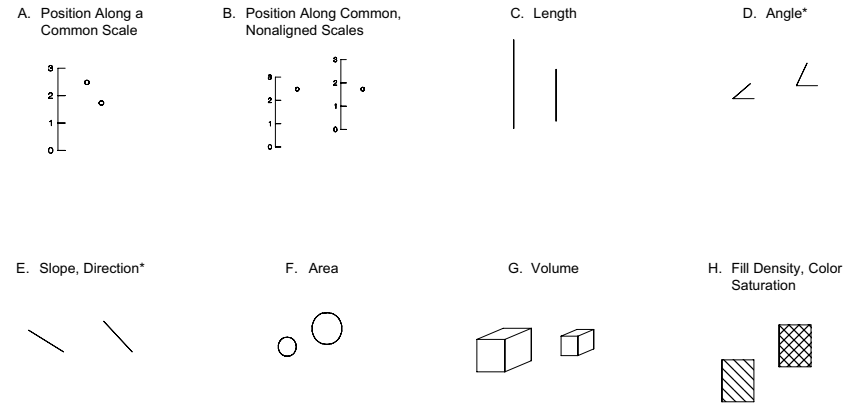
Data Source: CPS 1992 National Election Study.

Figure 5: Party Affiliation by Age Groups within the American Electorate, 1992.



Data Source: CPS 1992 National Election Study.

Figure 6: Graphical Perception Tasks, Ordered from Most Accurate to Least Accurate.



* Perceptual judgments about angles and slopes/directions are carried out with equal accuracy, so their relative ordering in this figure is arbitrary.

Source: Created from information provided in Cleveland (1994).

STATISTICAL GRAPHICS FOR VISUALIZING DATA

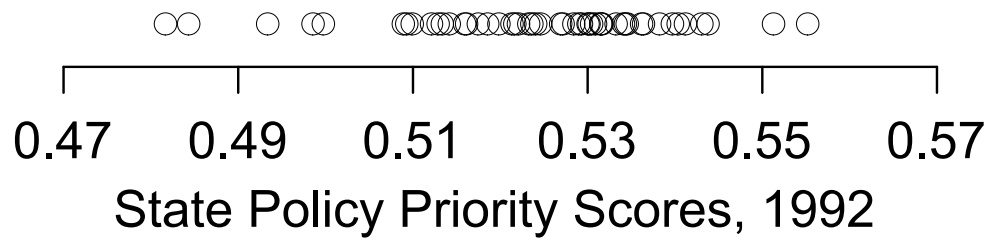
Tables and Figures, II

William G. Jacoby
Michigan State University and ICPSR

University of Illinois at Chicago
October 14-15, 2010

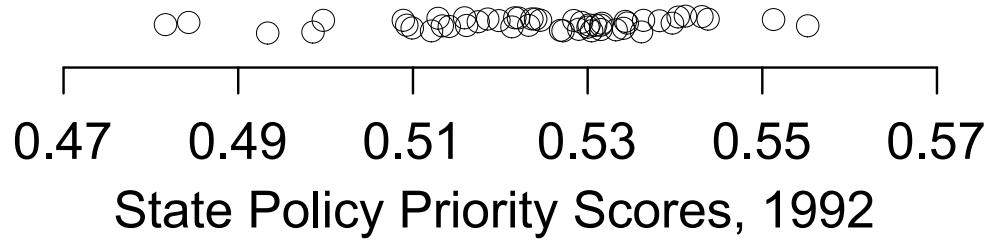
<http://polisci.msu.edu/jacoby/uic/graphics>

Figure 8D: Unidimensional Scatterplot of 1992 State Policy Priority Scores.



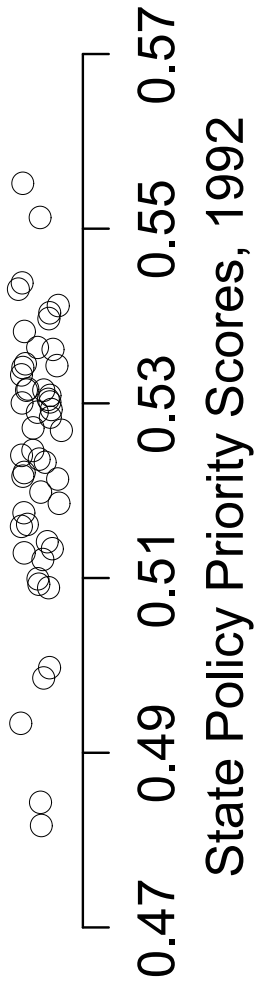
Data Source: Schneider and Jacoby (2004).

Figure 8E: Unidimensional Scatterplot of 1992 State Policy Priority Scores.



Data Source: Schneider and Jacoby (2004).

Figure 8F: Unidimensional Scatterplot of 1992 State Policy Priority Scores.



Data Source: Schneider and Jacoby (2004).

Figure 8G: Unidimensional Scatterplot of 1992 State Policy Priority Scores.

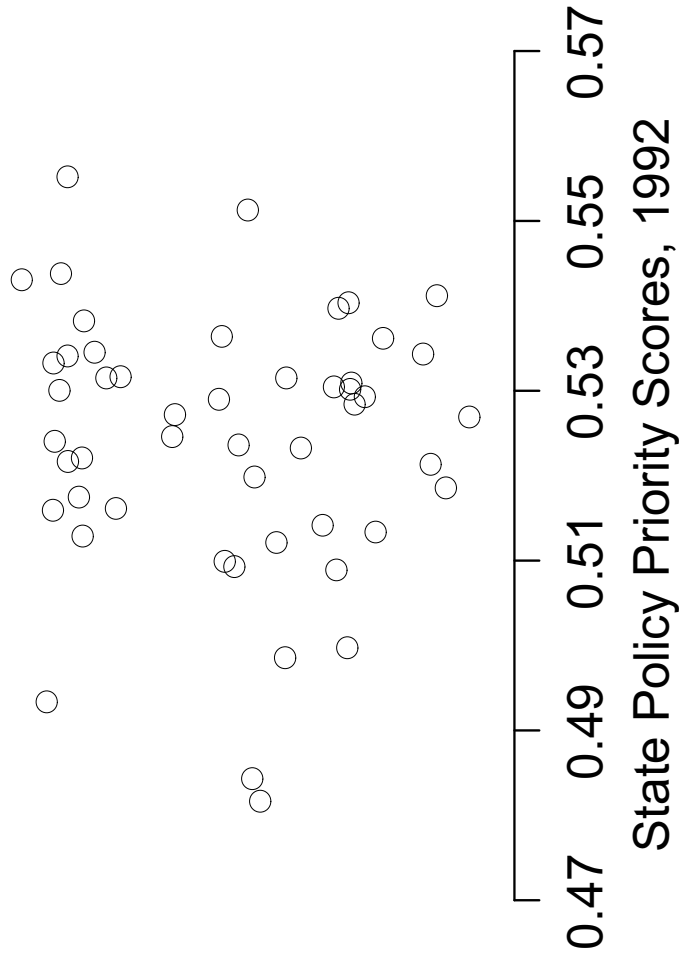
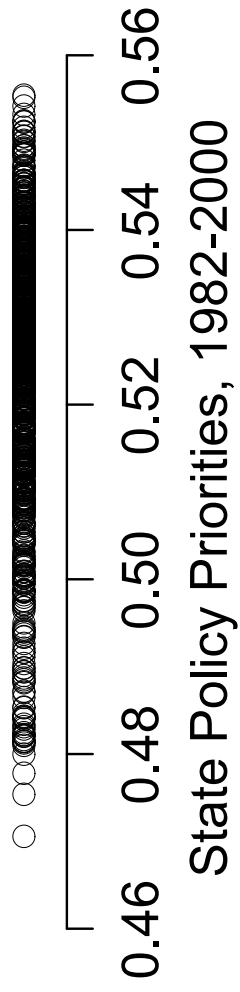
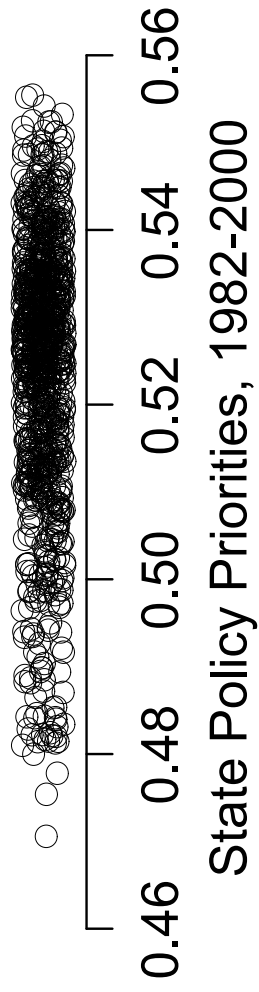


Figure 9A: Unidimensional Scatterplot of State Policy Priority Scores, 1982-2000



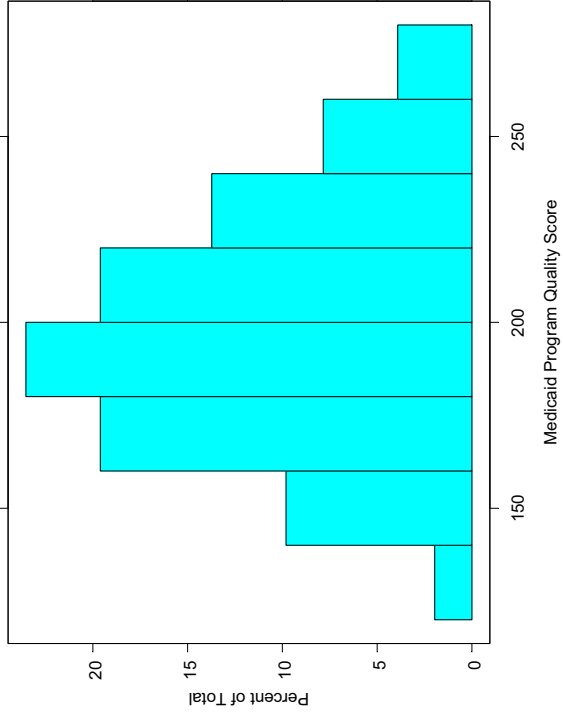
Data Source: Schneider and Jacoby (2004).

Figure 9B: Unidimensional Scatterplot of State Policy Priority Scores, 1982-2000



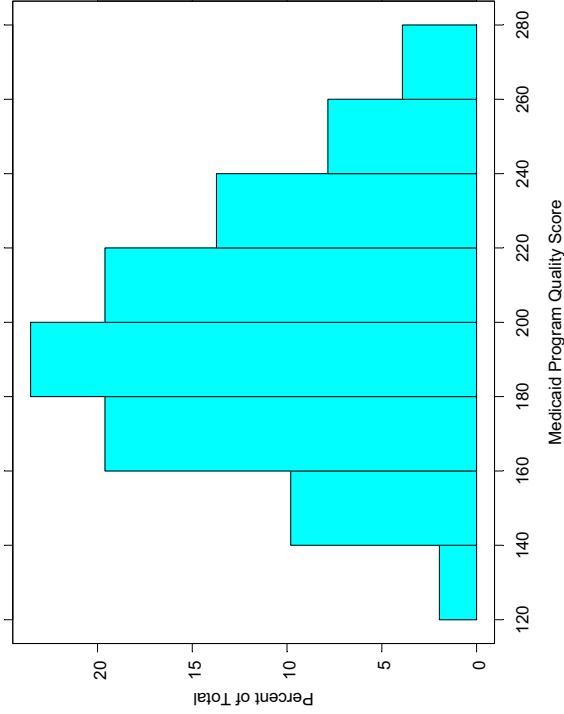
Data Source: Schneider and Jacoby (2004).

Figure 10B: Histogram of 1986 State Medicaid Program Quality Scores.



Data Source: Public Citizen Health Research Group

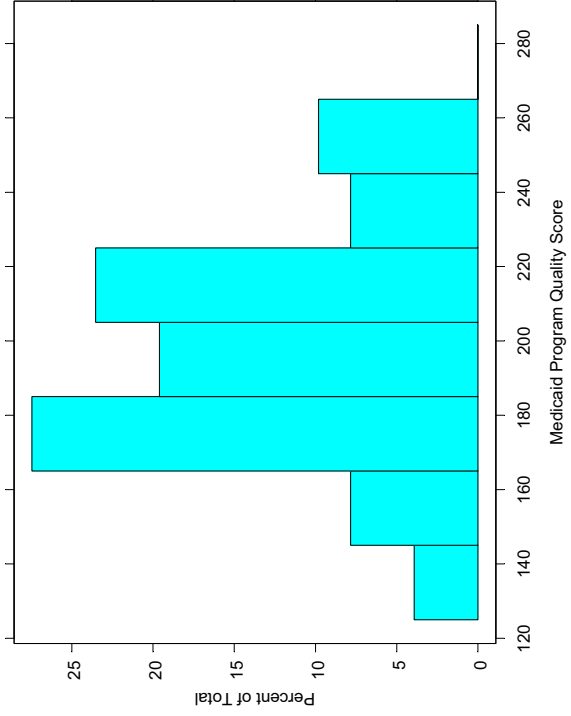
Figure 10C: Histogram of 1986 State Medicaid Program Quality Scores.



Note: Bin origin at 120, bin width is 20.

Data Source: Public Citizen Health Research Group

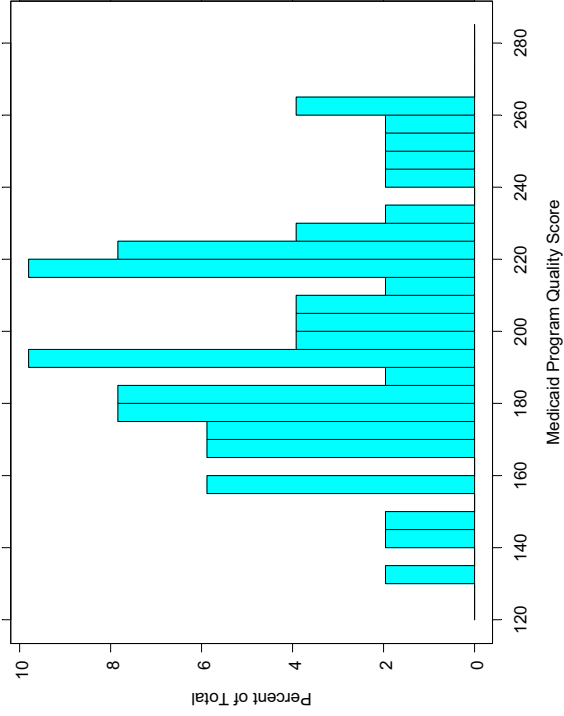
Figure 10D: Histogram of 1986 State Medicaid Program Quality Scores.



Note: Bin origin at 125, bin width is 20.

Data Source: Public Citizen Health Research Group

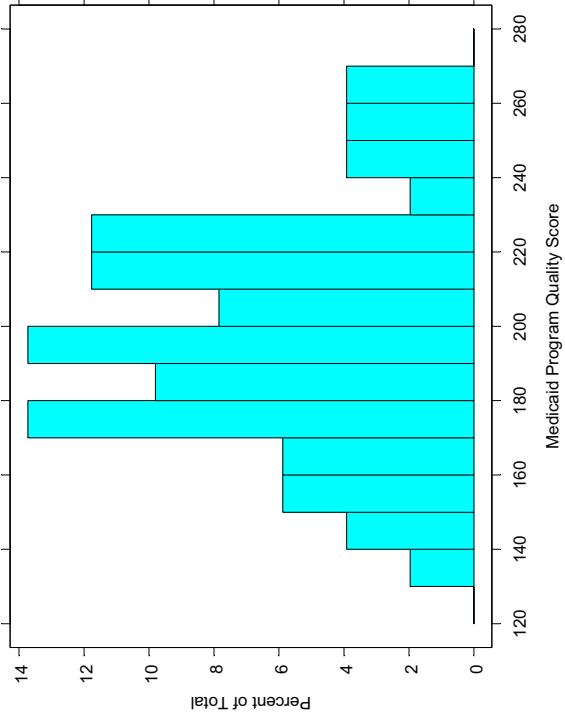
Figure 10E: Histogram of 1986 State Medicaid Program Quality Scores.



Note: Bin origin at 120, bin width is 10.

Data Source: Public Citizen Health Research Group

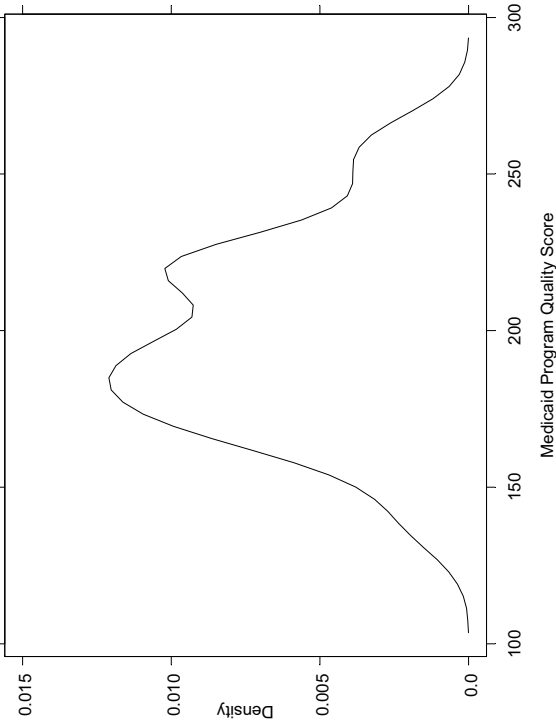
Figure 10F: Histogram of 1986 State Medicaid Program Quality Scores.



Note: Bin origin at 120, bin width is 10.

Data Source: Public Citizen Health Research Group

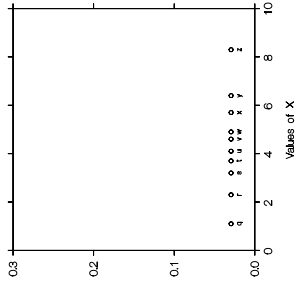
Figure 11: Smoothed Histogram of 1986 State Medicaid Program Quality Scores.



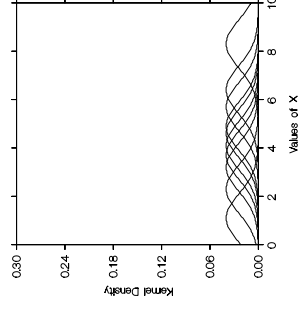
Data Source: Public Citizen Health Research Group

Constructing a Smoothed Histogram from Hypothetical Data:

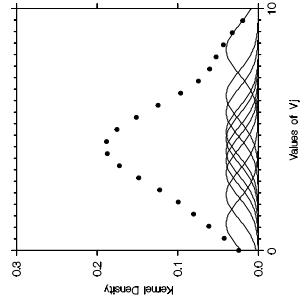
A. Univariate Scatterplot of 10 Data Points



B. Data Shown as Kernel Densities



C. Summing Heights of Kernel Densities



D. Smoothed Histogram

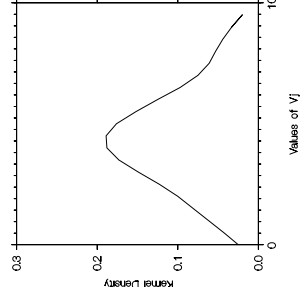
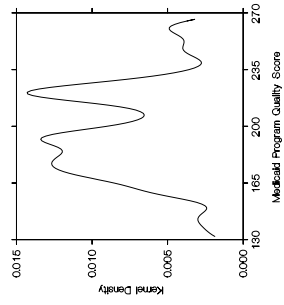
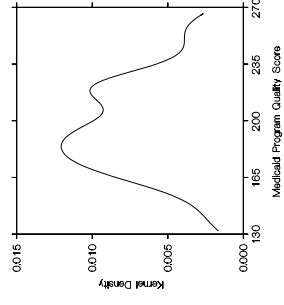


Figure 13: Changing the Bandwidth on Smoothed Histograms of Medicaid Program Quality Data.

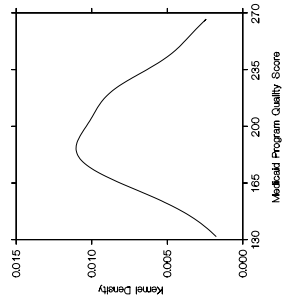
A. Bandwidth $h = 5$



B. Bandwidth, $h = 10$



C. Bandwidth, $h = 15$



D. Bandwidth, $h = 20$

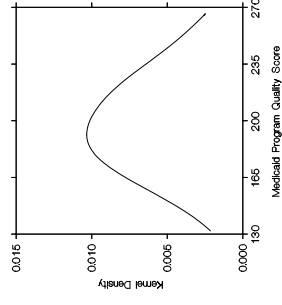


Figure 14: The Box Plot– State Policy Priorities, 1982-2000 .

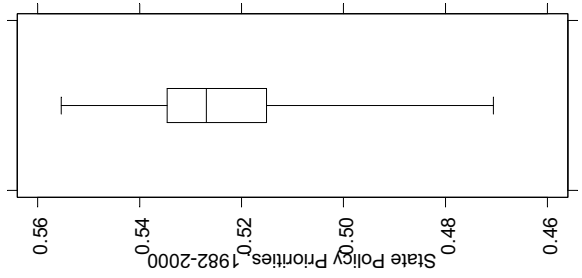


Figure 15A: Box Plots and Outliers

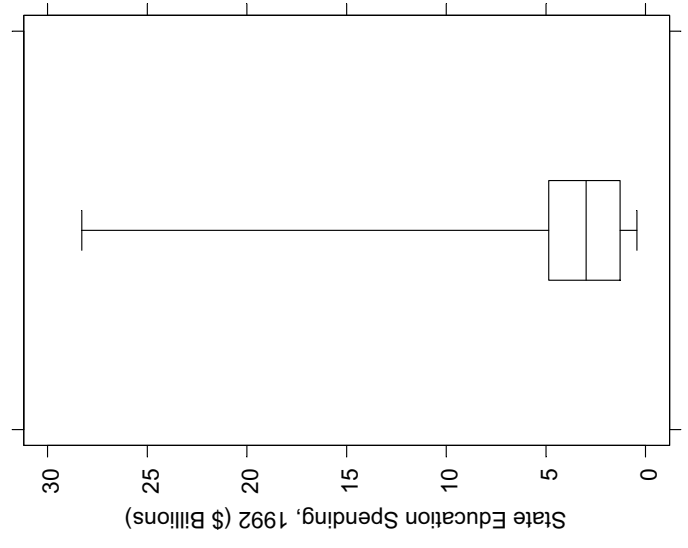


Figure 15B: Box Plots and Outliers

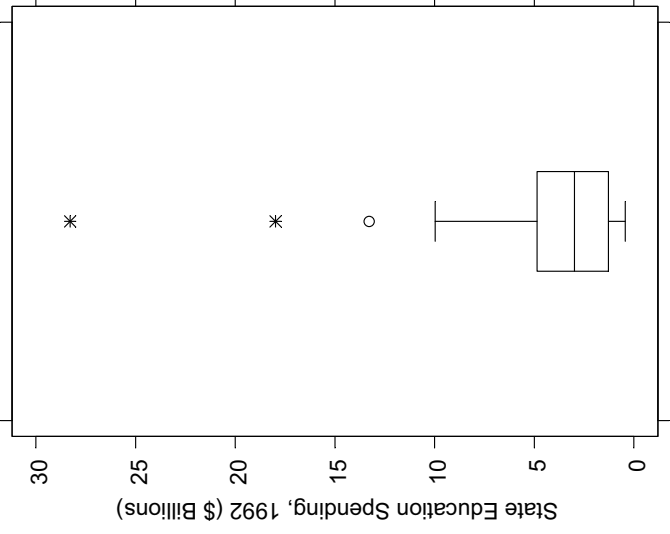
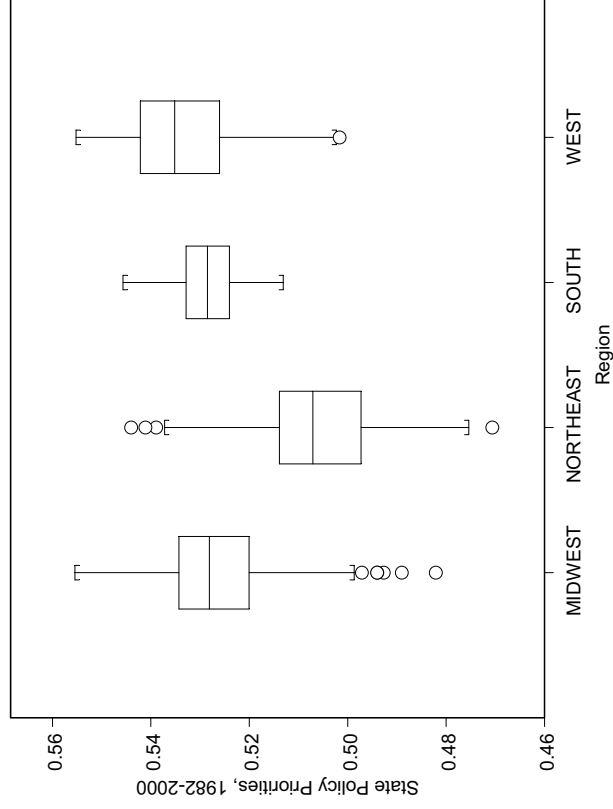


Figure 16: Parallel Box Plots of 1982-2000 State Policy Priorities by Region.



Data Source: Schneider and Jacoby (2004)

Figure 17: Dot Plot of State Medicaid Program Quality Scores.

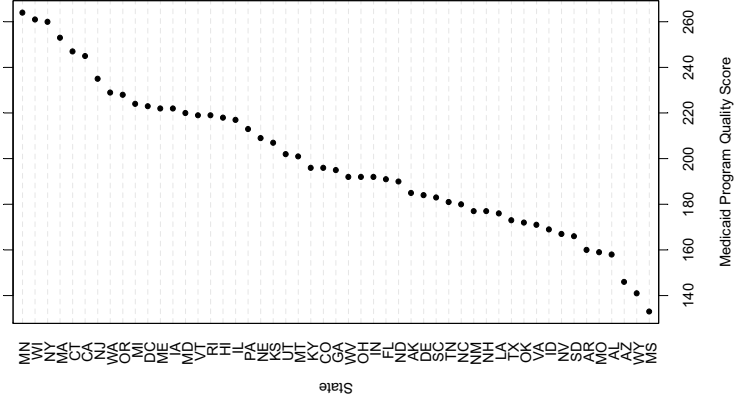


Figure 18: Dot Plot of Median State Medicaid Program Quality Scores within Regions

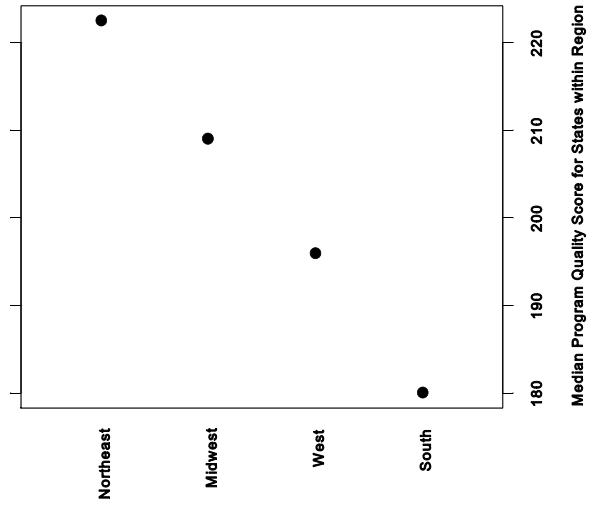
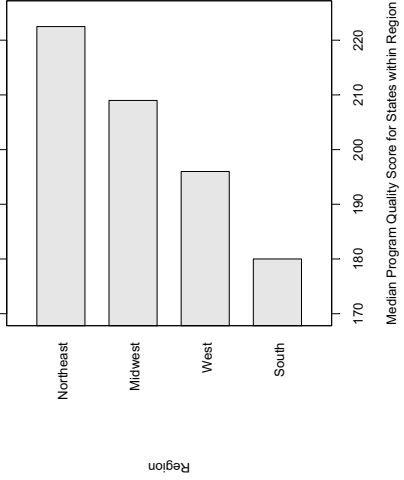


Figure 19: Bar Charts, Scale Limits, and Visual Perception

A. Scale Minimum Value Set Arbitrarily



B. Minimum Scale Value Set to Zero

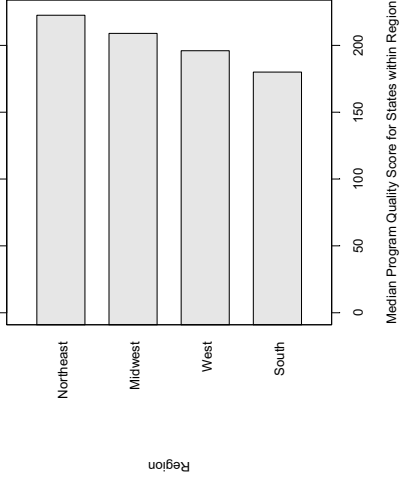


Figure 20: Dot Plot Showing 1992 Social Welfare Expenditures in the American States.

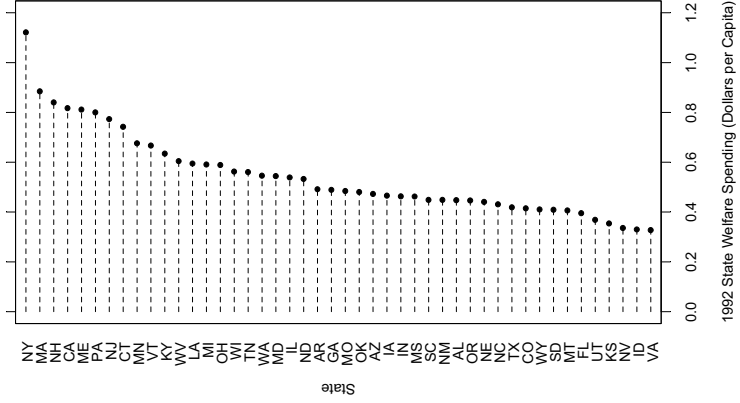
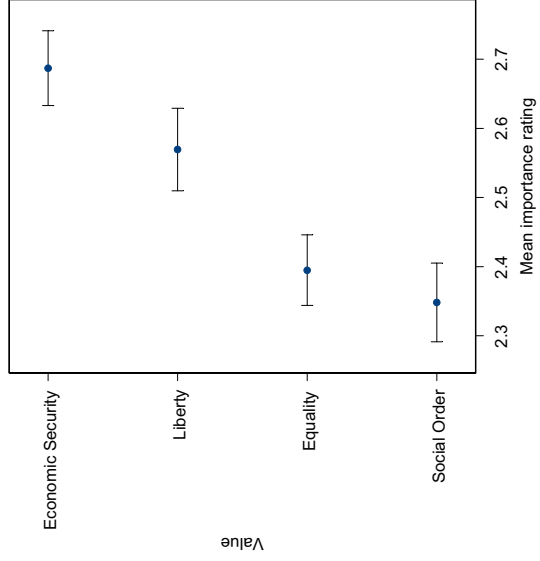


Figure 21: Value Importance Ratings in the American Public.



Note: The plotted points correspond to the mean importance rating assigned to each of the values, on a scale ranging from a minimum of zero to a maximum of three. The horizontal error bars show 95% confidence intervals for each of the means. Nonoverlapping confidence intervals imply that the means are reliably different from each other, that is, their difference is probably *not* merely due to sampling error.

Data Source: 1994 Multi-Investigator Study.

STATISTICAL GRAPHICS FOR VISUALIZING DATA

Tables and Figures, III

William Jacoby
Michigan State University and ICPSR

University of Illinois at Chicago
October 14-15, 2010

<http://polisci.msu.edu/jacoby/uic/graphics>

Table 1: The following table shows four bivariate datasets. Each one contains eleven hypothetical observations on two variables. The variables are designated X_1 and Y_1 in the first dataset, X_2 and Y_2 in the second dataset, and so on. These datasets have a remarkable feature: They produce identical regression results. When the Y variable is regressed on the X variable in each of these datasets, the following equation is produced (the figures in parentheses are standard errors):

$$Y_i = 3.00 + 0.50 X_i + e_i \quad R^2 = 0.67$$

(1.12) (0.12)

Because the OLS estimates are identical for each of the four datasets, one could easily conclude that they all have the same bivariate structure.

X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

Data Source: Anscombe, F. J. (1973) "Graphs in Statistical Analysis." *American Statistician* 27: 17-21.

Figure 1: The following bivariate scatterplots show the relationships within each of the four datasets listed in Table 1. Clearly, these datasets differ markedly from each other, contrary to the conclusions that one might reach upon the OLS regression estimates, alone.

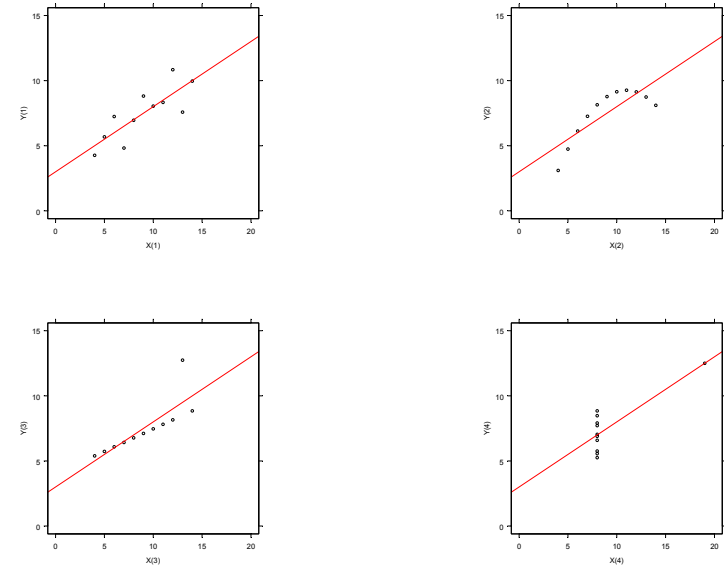


Figure 2: A poorly-constructed scatterplot.

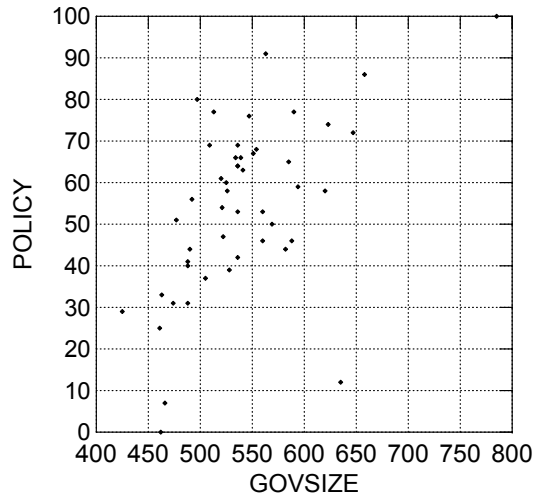


Figure 3: A Better Version of the scatterplot

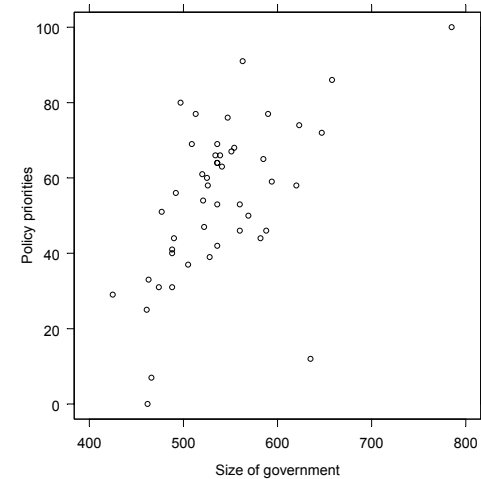
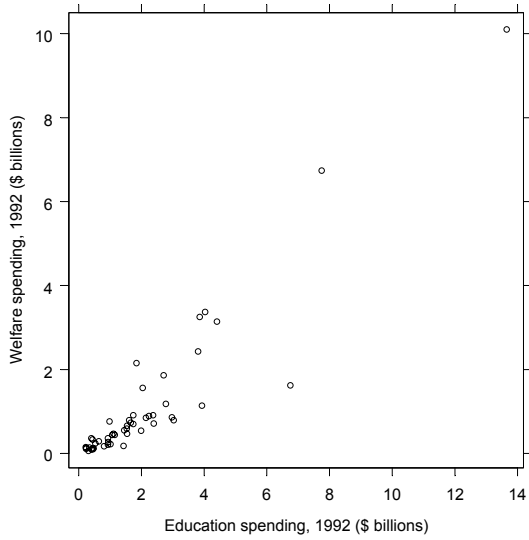
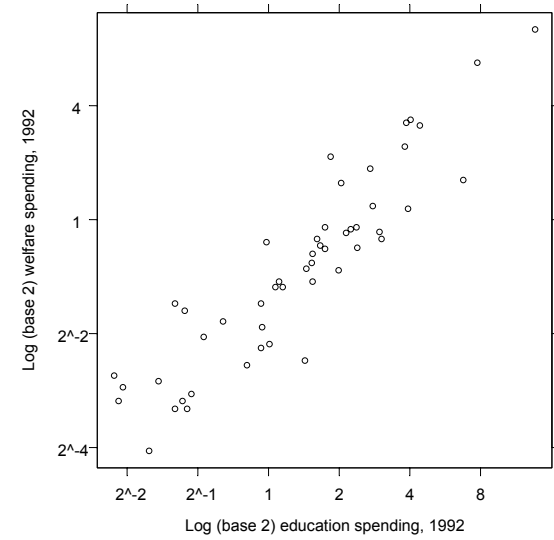


Figure 4A: Transforming variable values in order to improve visual resolution in a scatterplot.



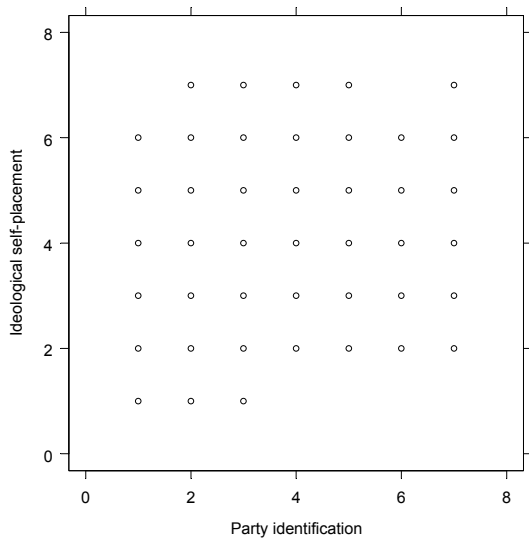
Data Source: 1994 *Statistical Abstract of the United States*.

Figure 4B: Transforming variable values in order to improve visual resolution in a scatterplot.



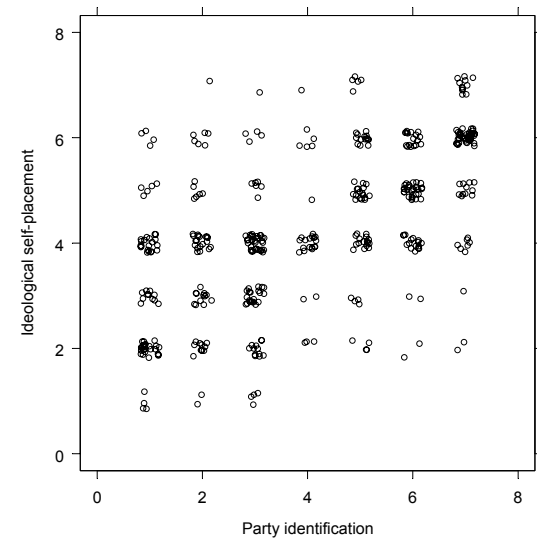
Data Source: 1994 *Statistical Abstract of the United States*.

Figure 5A: Jittering to reduce overplotting of data points



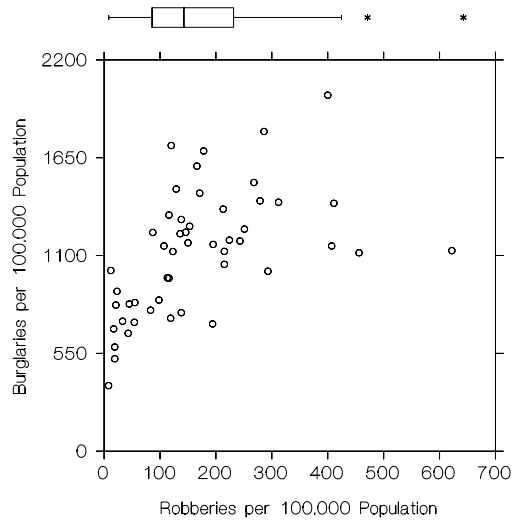
Data Source: CPS 2000 American National Election Study.

Figure 5B: Jittering to reduce overplotting of data points



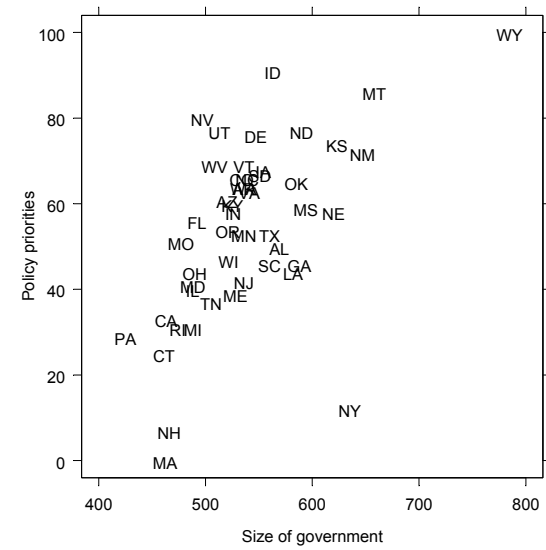
Data Source: CPS 2000 American National Election Study.

Figure 6: Robbery Rates Versus Burglary Rates in the American States, 1991.



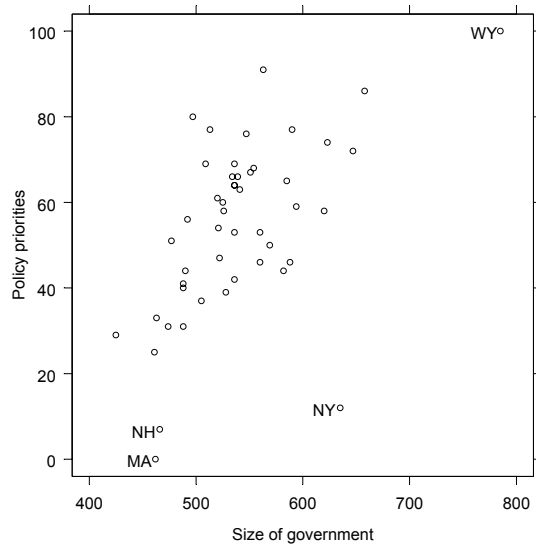
Source: 1992 Statistical Abstract of the United States.

Figure 7A: Point labels in scatterplots.



Data Source: Jacoby and Schneider (2001)

Figure 7B: Point labels in scatterplots.



Data Source: Jacoby and Schneider (2001)

Figure 8: Slicing a Scatterplot of the Relationship Between 1980 GNP per Capita and Infant Mortality Rates.

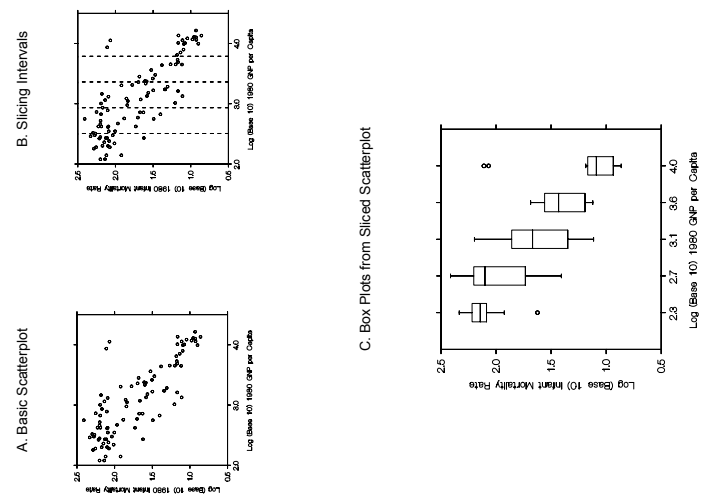
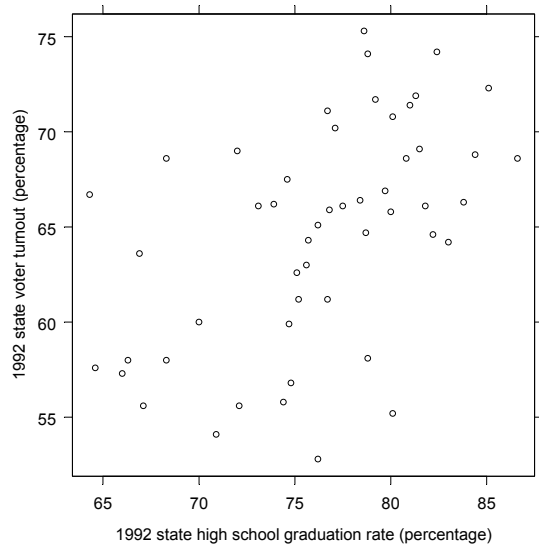
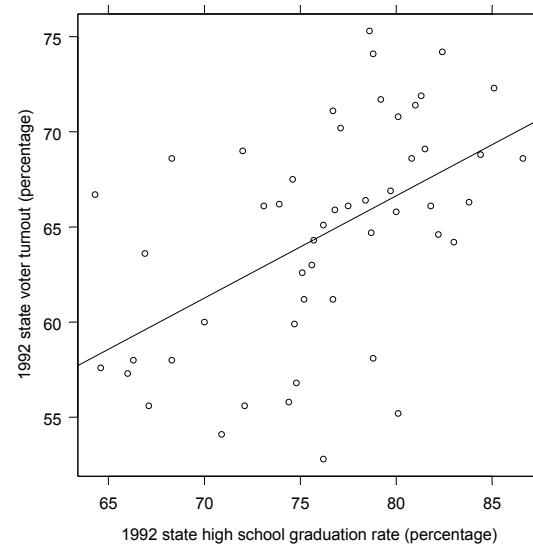


Figure 9A: The relationship between education and voter turnout in the American states, 1992.



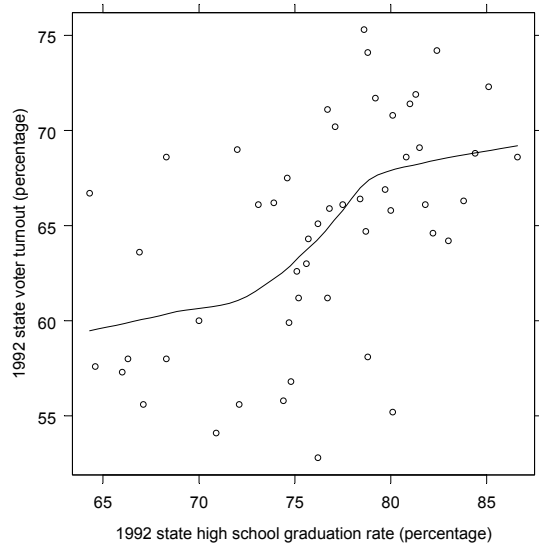
Data Source: 1994 Statistical Abstract of the United States

Figure 9B: The relationship between education and voter turnout in the American states, 1992.



Data Source: 1994 Statistical Abstract of the United States

Figure 9C: The relationship between education and voter turnout in the American states, 1992.



Data Source: 1994 Statistical Abstract of the United States

Figure 10: Example of a loess "local fitting window" and tricube weights (hypothetical data)

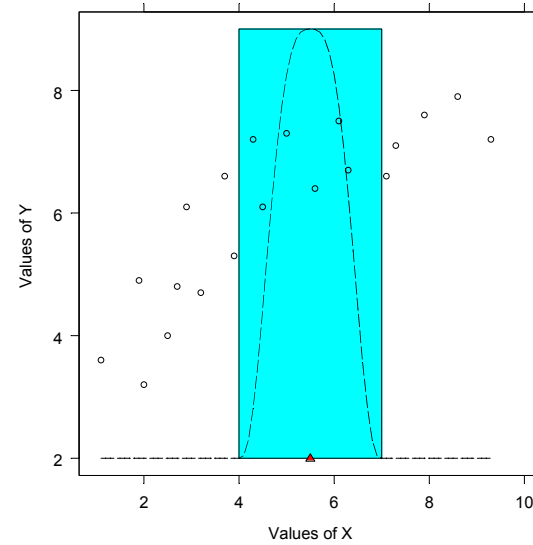
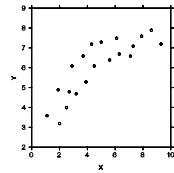
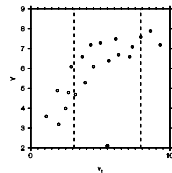


Figure 11: Illustration of Loess Fitting Procedure Using Hypothetical Data.

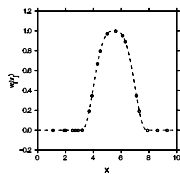
A. Hypothetical Data for Loess Fit.



B. Window for $\tau_1=5.5$ and $\alpha=0.6$.



C. Tricube Neighborhood Weights.



D. Initial Regression Line and Fitted Value.

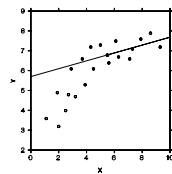
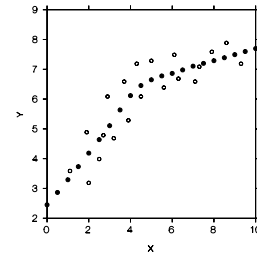


Figure 11: Illustration of Loess Fitting Procedure (Continued).

E. Complete Set of Fitted Values.



F. Loess Curve

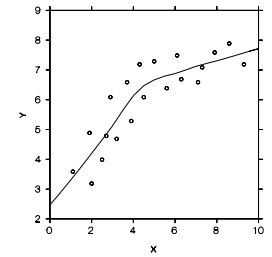
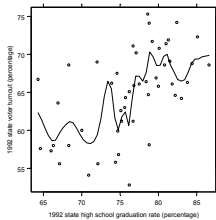
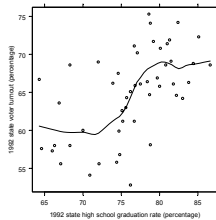


Figure 12: Effect of the α Parameter on the Loess Smooth Curve.

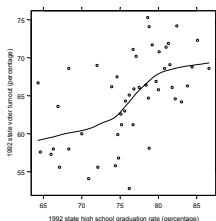
A. Loess Curve with $\alpha = 0.15$



B. Loess Curve with $\alpha = 0.35$



C. Loess Curve with $\alpha = 0.75$



D. Loess Curve with $\alpha = 1.00$

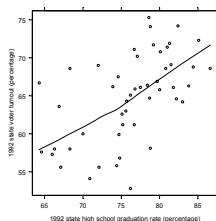


Figure 13A: Residual Plot from Loess Curve Fitted to State Education and Voter Turnout Data

A. Residuals from curve fitted with $\alpha = .65$

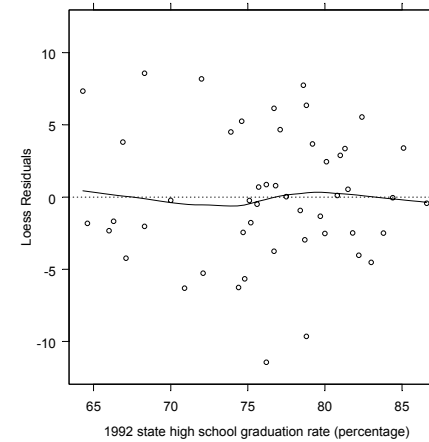
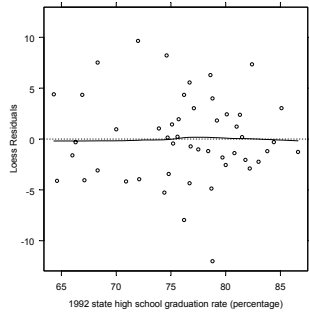


Figure 13: Residual Plots for Loess Curves Fitted with Different α Values.

B. Residual Plot for Loess Curve Fitted with $\alpha = 0.15$



C. Residual Plot for Loess Curve Fitted with $\alpha = 1.00$

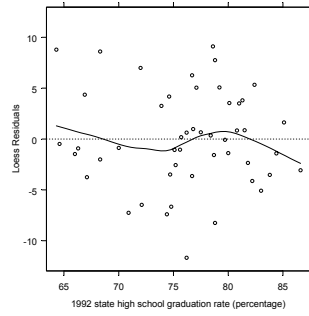
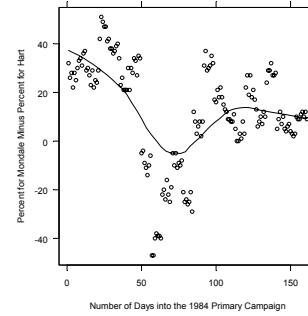
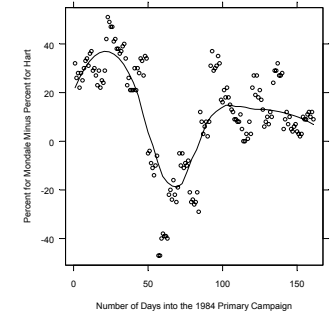


Figure 14: Effect of the λ Parameter on a Loess Smooth Curve.

A. Loess Curve Fitted with $\lambda = 1$ and $\alpha = 0.50$

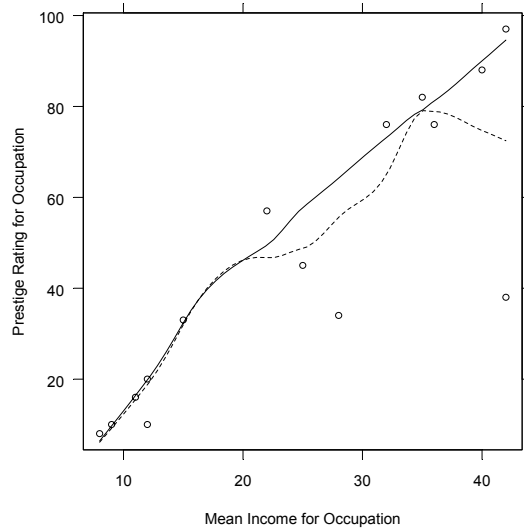


B. Loess Curve Fitted with $\lambda = 2$ and $\alpha = 0.50$



Note: The data show public preferences between Walter Mondale and Gary Hart (among Democrats only) during the 1984 Presidential primary campaign. Data source is the 1984 CPS Continuous Monitoring Survey.

Figure 15: The Effect of Including Robustness Weights in the Loess Fitting Process.



Note: The data are fifteen observations sampled from the Duncan Occupational Prestige Dataset. The solid line in the figure represents the robust loess fit. The dotted line shows the loess fit obtained without robustness iterations.

Figure 6 Residual-Fit Spread Plot for Robust Loess Fit, Occupational Prestige and Income

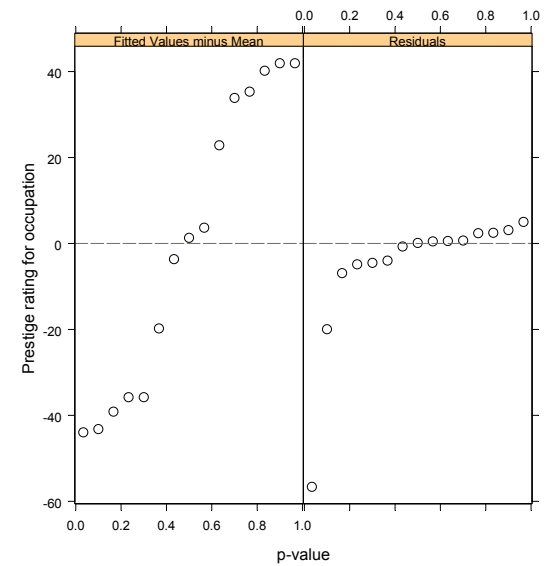


Figure 7 Residual-Fit Spread Plot for OLS Fit, Policy Priorities and Government Size

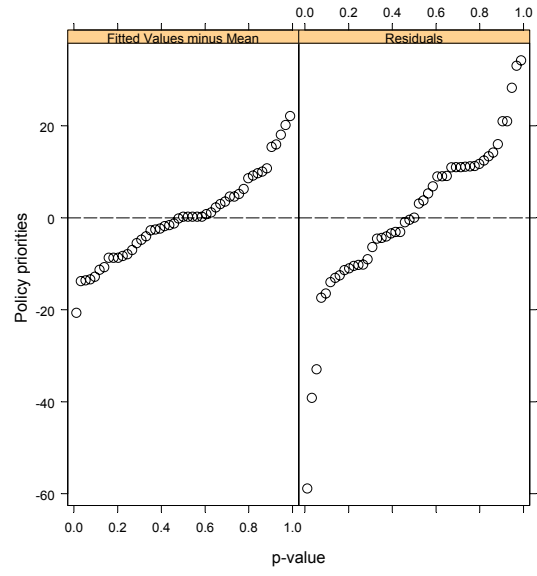
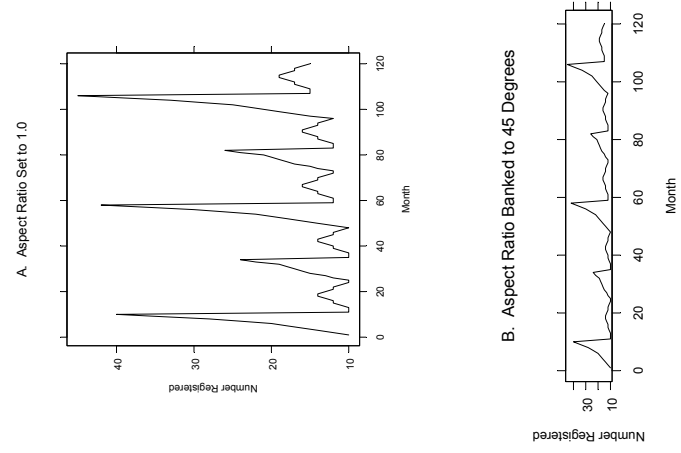


Figure 8

Plot of Hypothetical Data on Voter Registration Figures by Month, to show the Effects of Aspect Ratio on Visual Perception.



STATISTICAL GRAPHICS FOR VISUALIZING DATA

Tables and Figures, IV

William Jacoby
Michigan State University and ICPSR

University of Illinois at Chicago
October 14-15, 2010

<http://polisci.msu.edu/jacoby/uic/graphics>

Figure 1: Bivariate scatterplots showing two possible influences on incumbent margin of victory in the 2006 congressional elections.

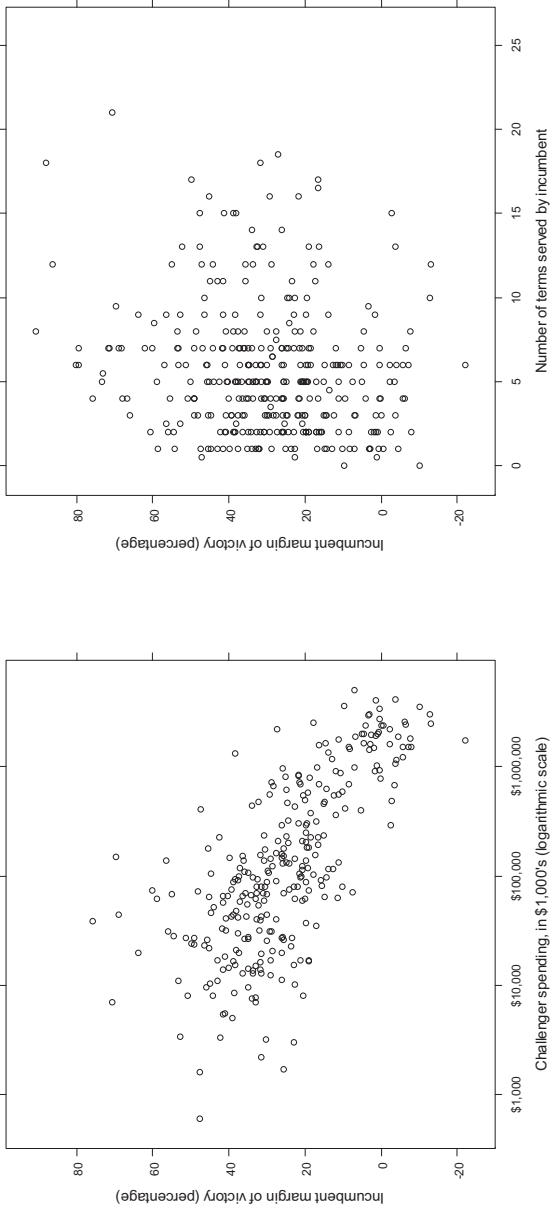


Figure 2: Three-dimensional scatterplot showing the trivariate relationship between incumbent margin of victory, challenger spending, and incumbent longevity in office.

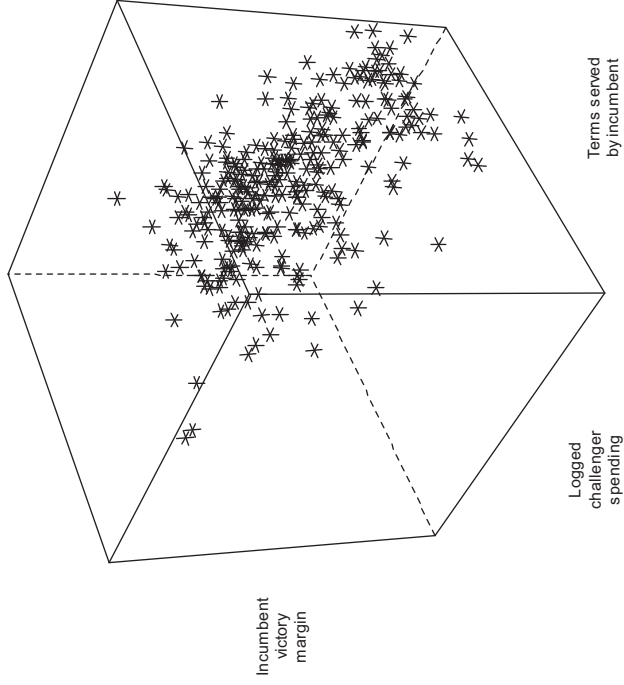


Figure 3: Adding motion to a three-dimensional scatterplot to enhance visual perception of three-dimensional structure.

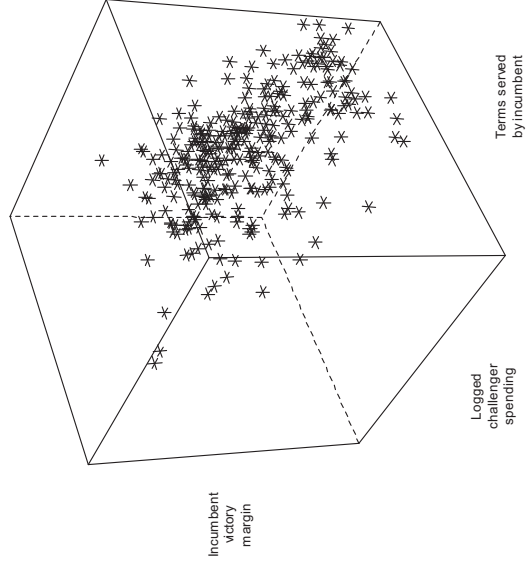


Figure 3: Adding motion to a three-dimensional scatterplot to enhance visual perception of three-dimensional structure.

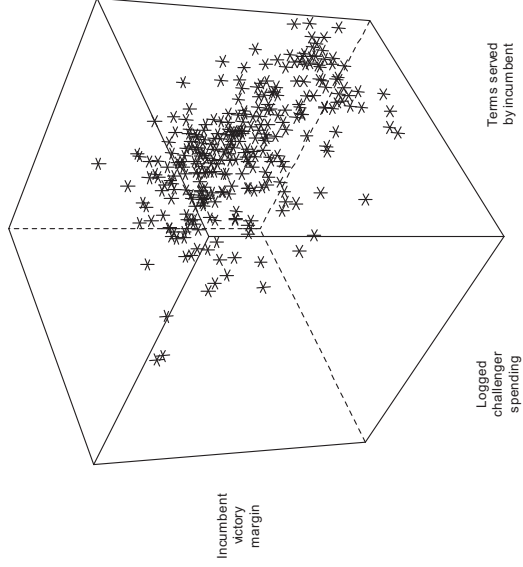


Figure 4: Three-dimensional plot showing the OLS surface for the regression of incumbent margin of victory on challenger spending, and incumbent longevity in office.

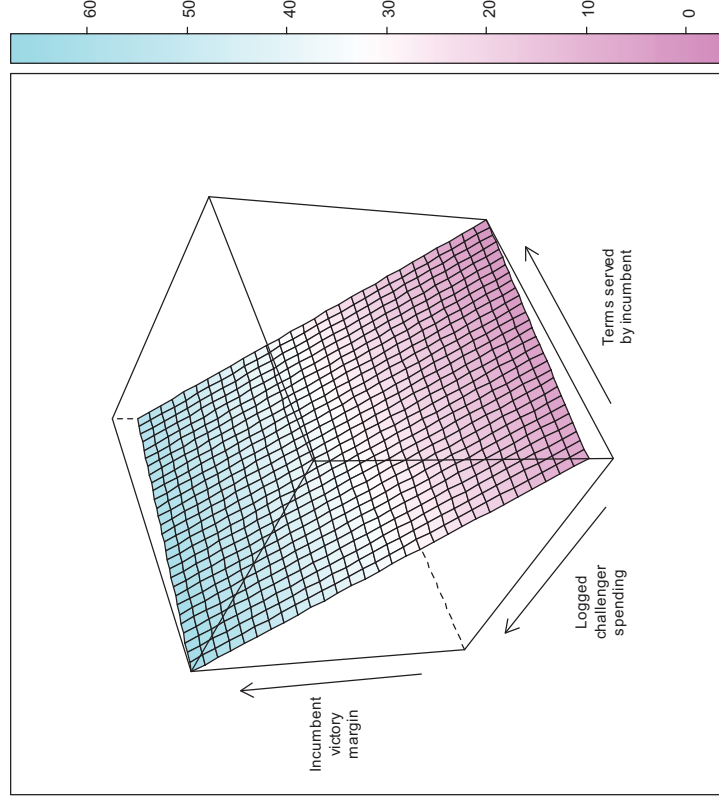


Figure 5: Three-dimensional plot showing fitted surface for the nonparametric regression (loess) of incumbent margin of victory on challenger spending, and incumbent longevity in office.

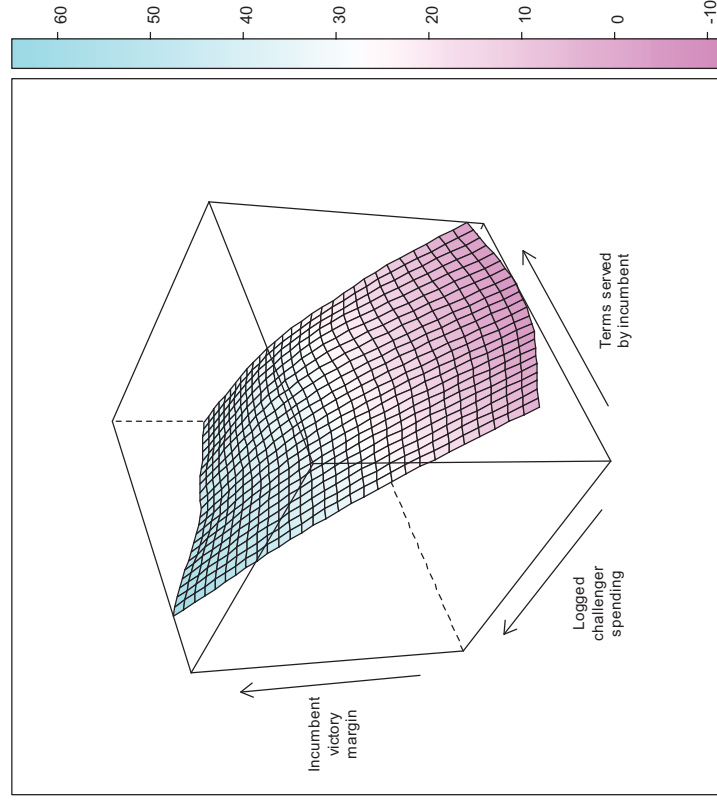


Figure 6: Three-dimensional scatterplot showing the trivariate relationship between incumbent margin of victory, challenger spending, and incumbent longevity in office (Figure 2, repeated).

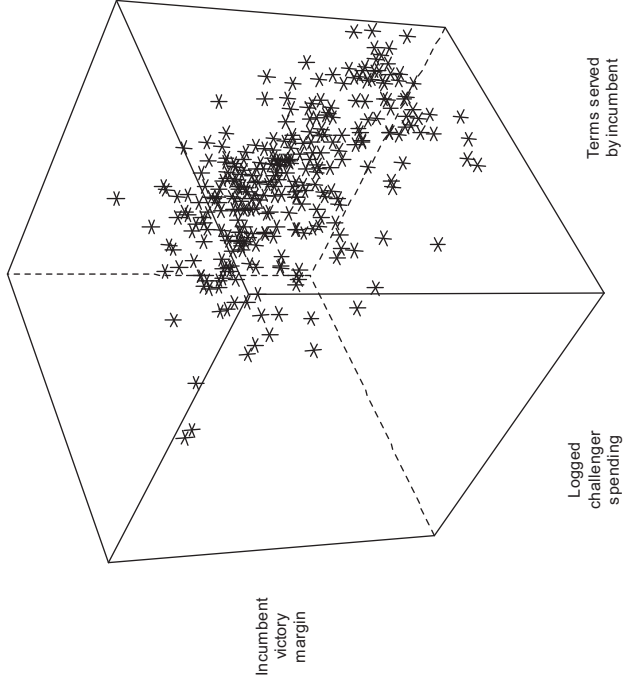


Figure 7: Scatterplot matrix for data on 2006 congressional elections.

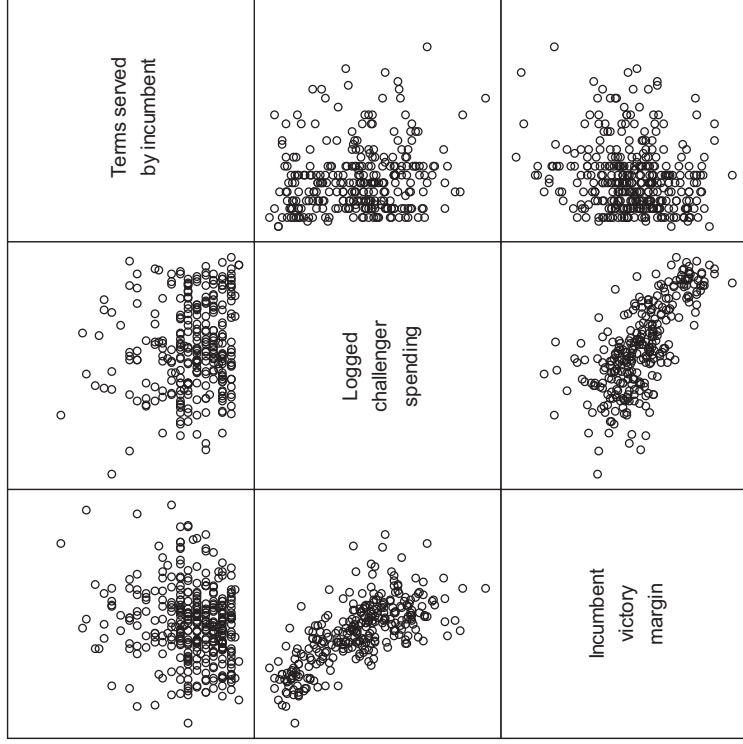


Figure 8: Scatterplot matrix with bivariate OLS lines for data on 2006 congressional elections.

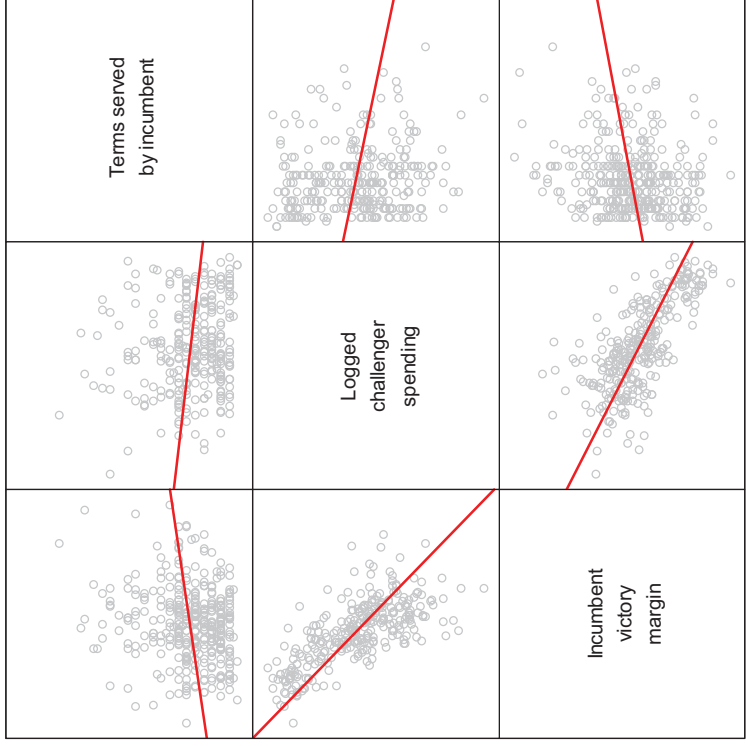


Figure 9: Scatterplot matrix with bivariate loess curves for data on 2006 congressional elections.

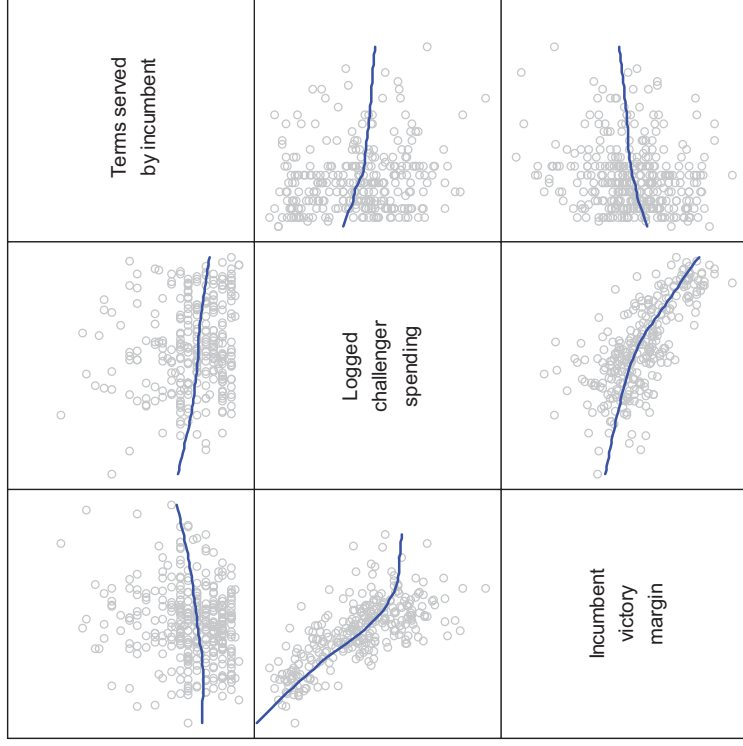


Figure 10: Creating conditioning slices from variable measuring number of incumbent terms.

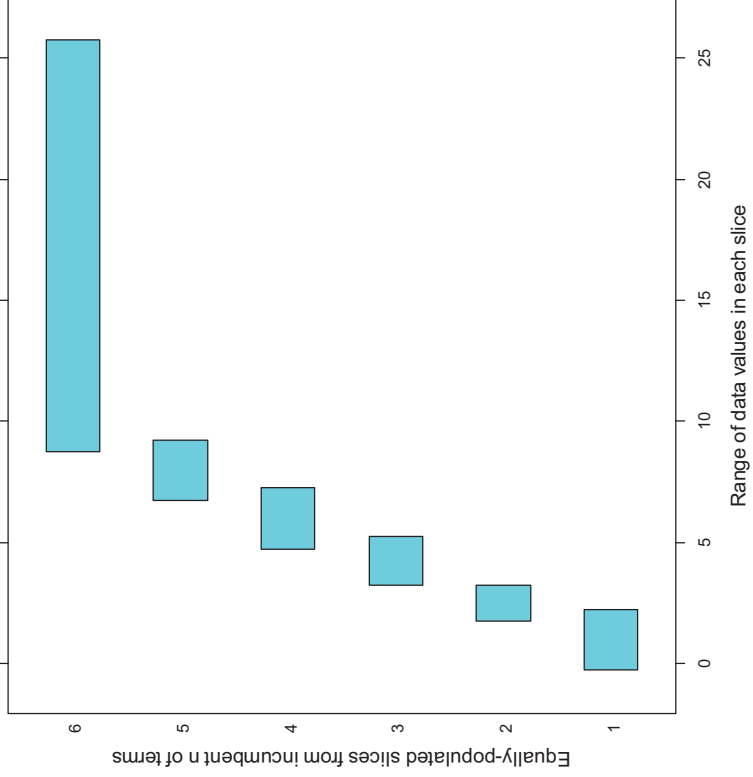


Figure 11: Trellis display showing incumbent victory margin versus challenger spending, conditioned on incumbent number of terms in office.

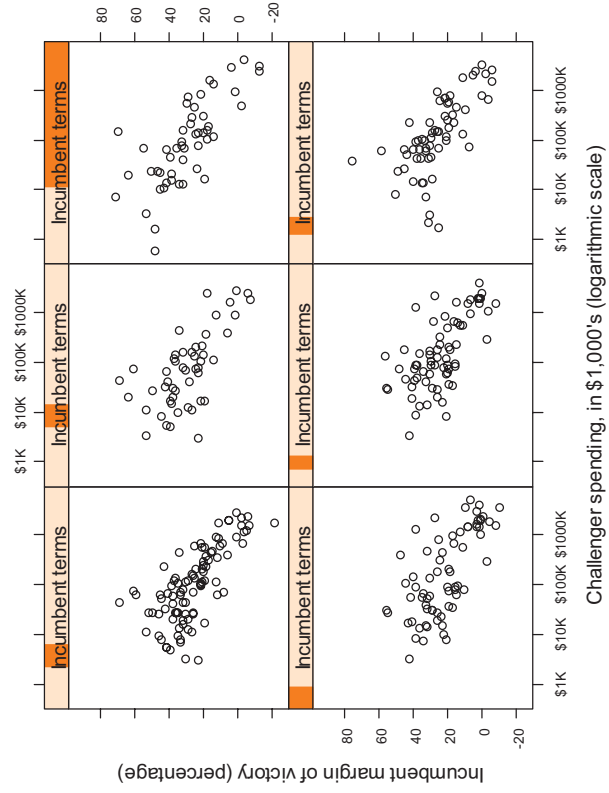


Figure 12: Trellis display showing incumbent victory margin versus challenger spending, conditioned on incumbent number of terms in office (bivariate loess curves added to each panel).

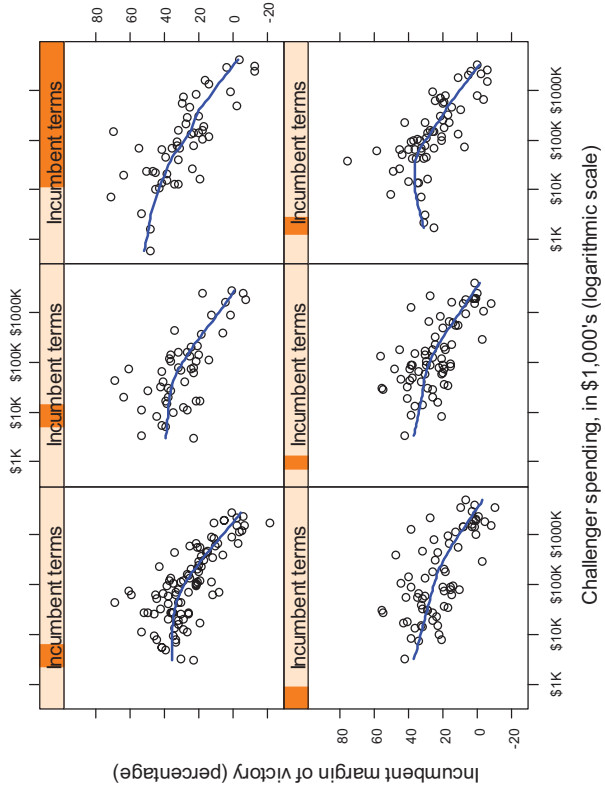


Figure 13: Trellis display showing incumbent victory margin versus incumbent number of terms, conditioned on logged challenger spending (bivariate loess curves added to each panel).

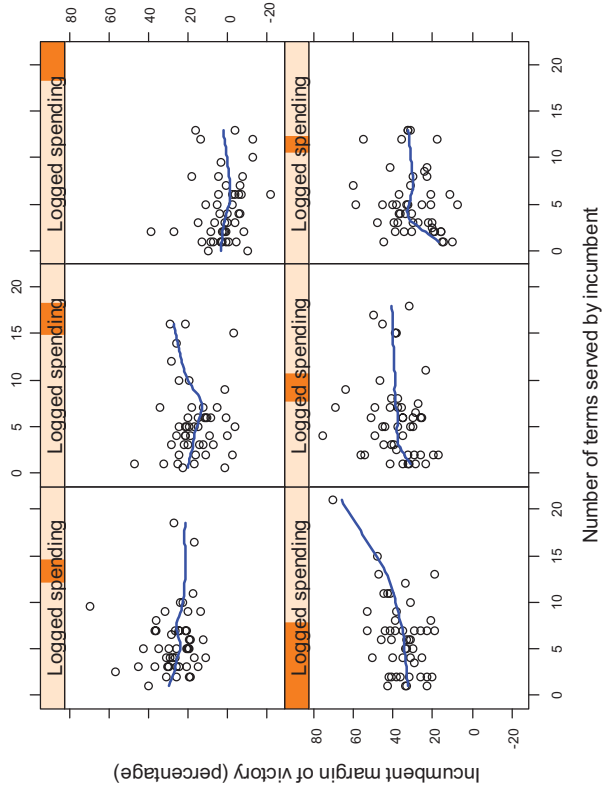


Figure 14: Bivariate scatterplots showing two possible influences on 1992 state policy priorities.

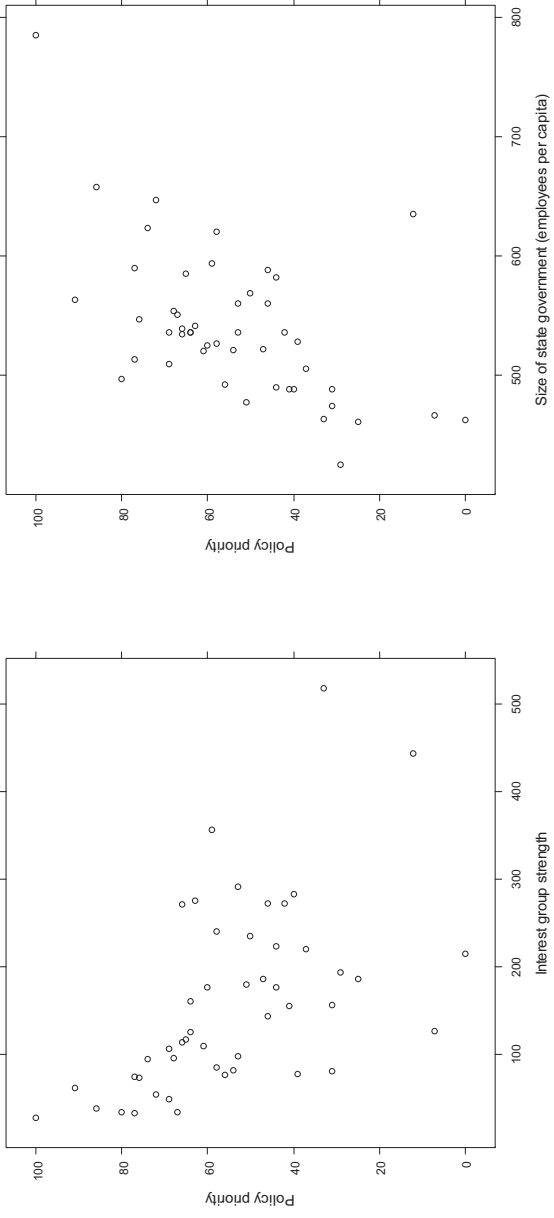


Figure 15: Conditioning slices from variable measuring size of state government, with two-thirds overlap between adjacent slices.

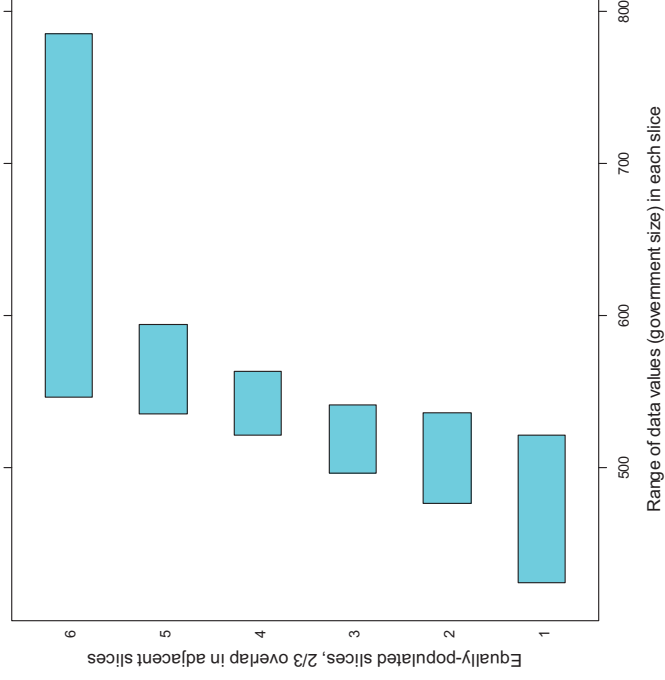


Figure 16: The relationship between 1992 state policy priorities and interest group strength within state, conditioned on size of state government (Bivariate OLS line shown within each panel).

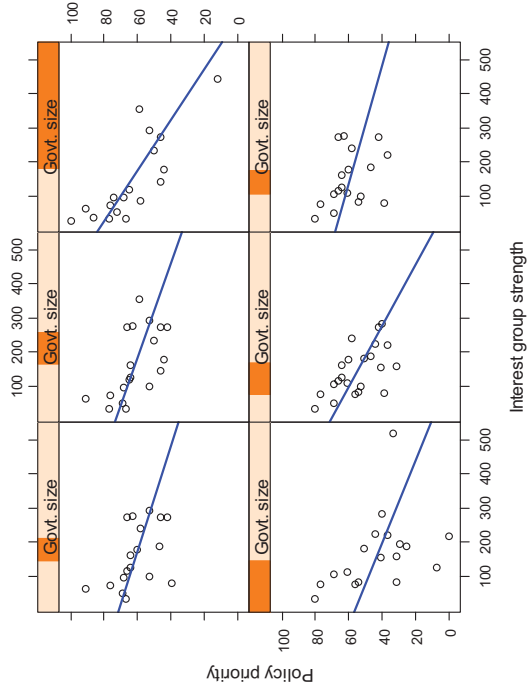


Figure 17: The relationship between 1992 state policy priorities and state government size, conditioned on interest group strength (Bivariate OLS line shown within each panel).

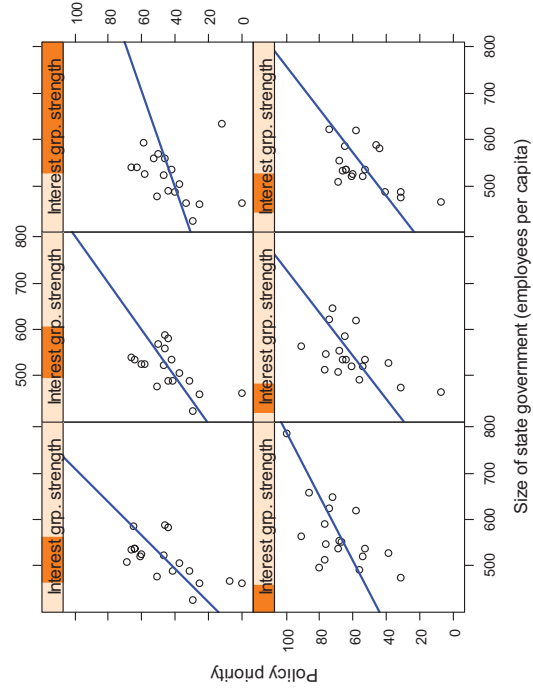


Figure 18: Scatterplot matrix for 1992 state policy priority data.

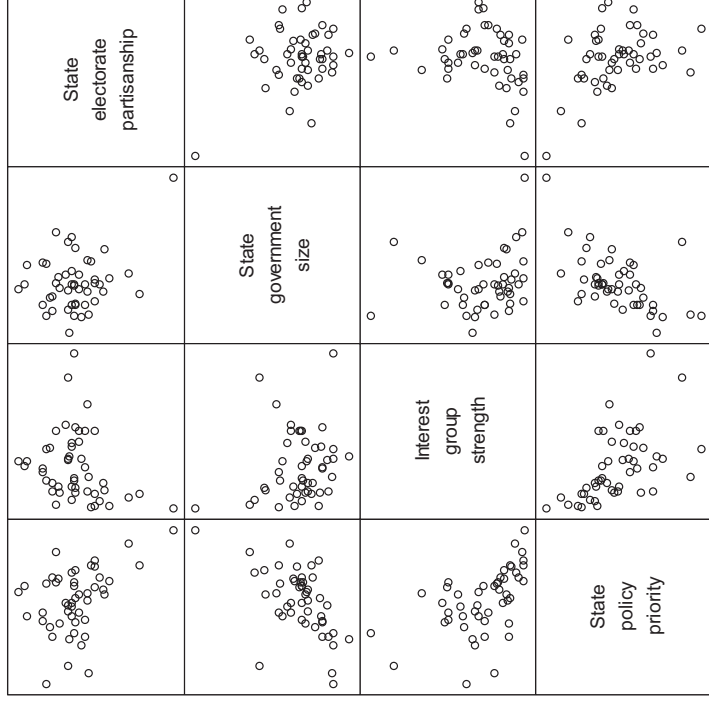


Figure 19: Relationship between 1992 state policy priorities and interest group strength, conditioned on size of state government and electorate partisanship (bivariate OLS line shown in panels).

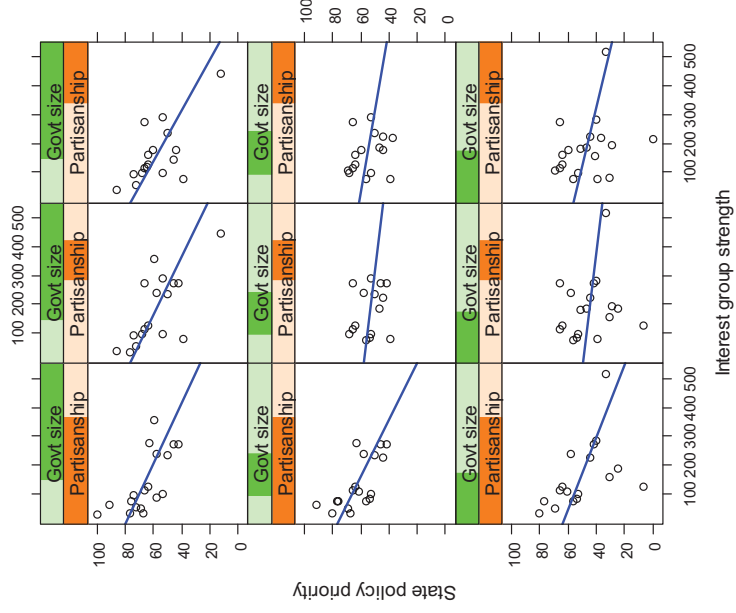


Figure 20: Relationship between 1992 state policy priorities and state government size, conditioned on interest group strength and electorate partisanship (bivariate OLS line shown in panels).

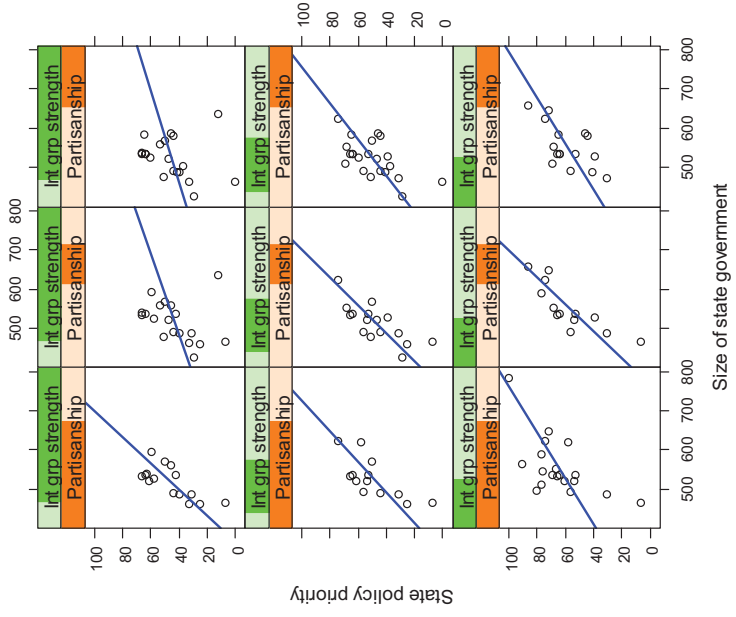


Figure 21: Relationship between 1992 state policy priorities and electorate partisanship, conditioned on interest group strength and state government size (bivariate OLS line shown in panels).

