

Statistical Graphics for Univariate and Bivariate Data

Excerpts from text by:

William A. Jacoby

Sage Publications, Thousand Oaks, CA, copyright 1997

Statistics 586 Spring 2015

TABLE 1.1
Hypothetical Data Matrix Containing 4 Variables and 20 Observations

Observation Number	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

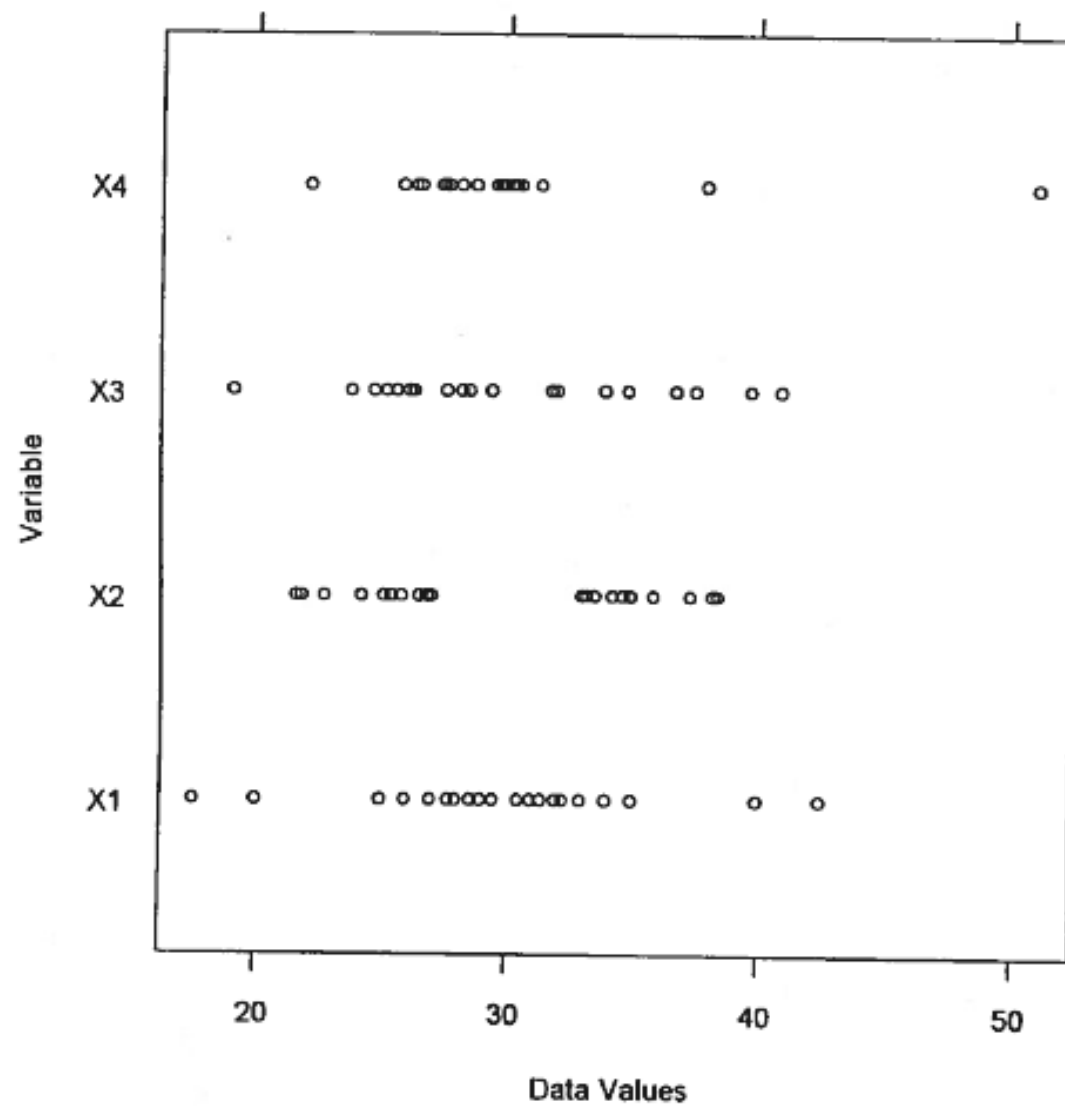
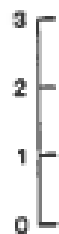
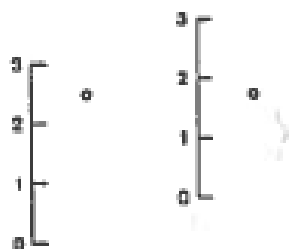


Figure 1.1. Univariate Scatterplots Comparing Data Distributions of Four Hypothetical Variables
 SOURCE: Data from Table 1.1.

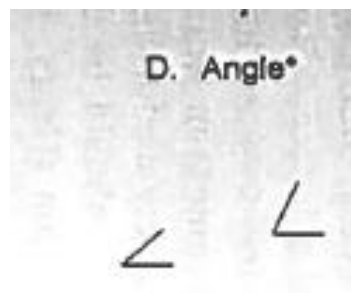
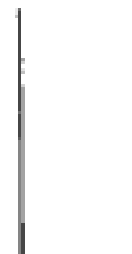
A. Position Along a Common Scale



B. Position Along Common, Nonaligned Scales



C. Length



E. Slope, Direction*



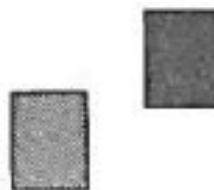
F. Area



G. Volume



H. Fill Density, Color Saturation



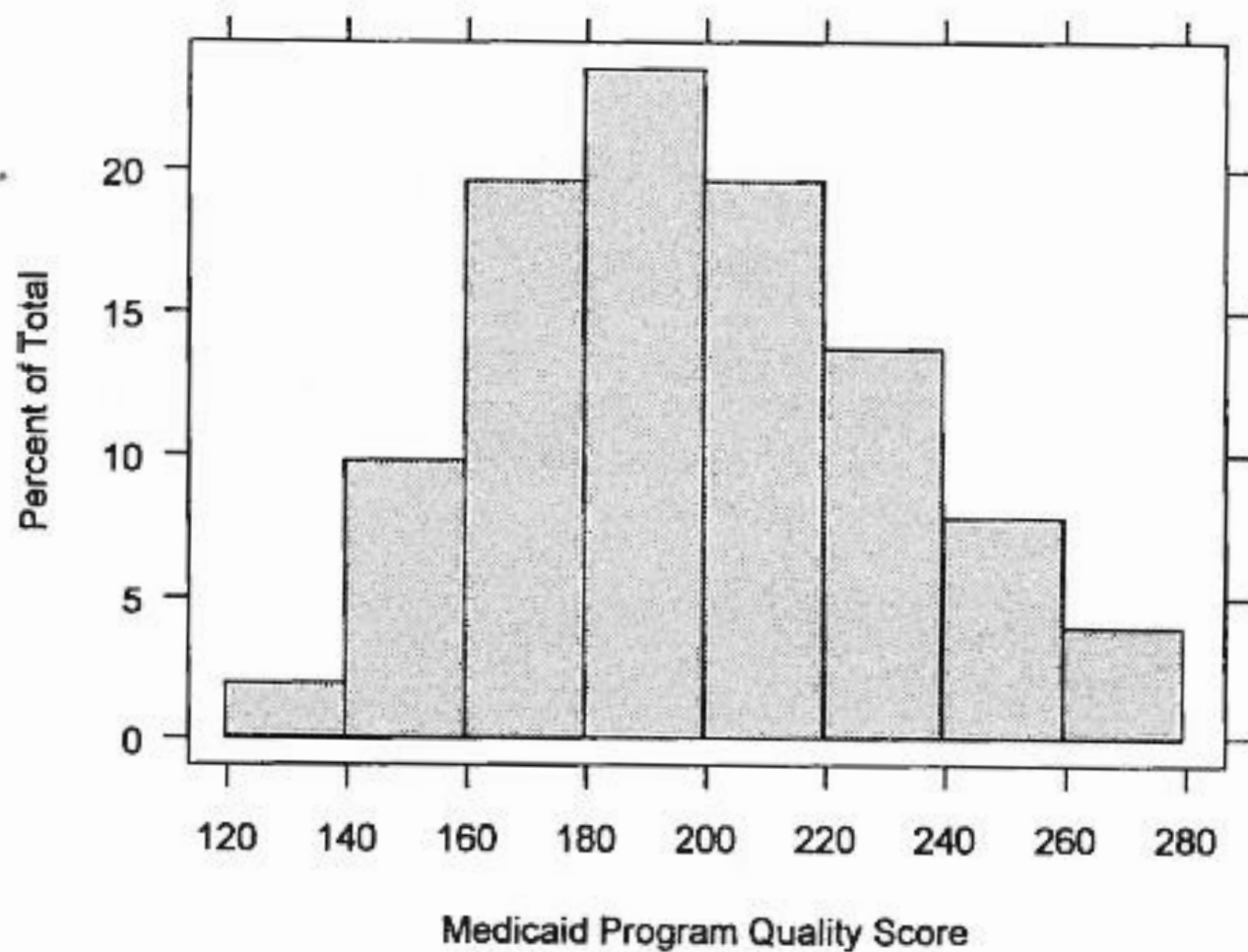
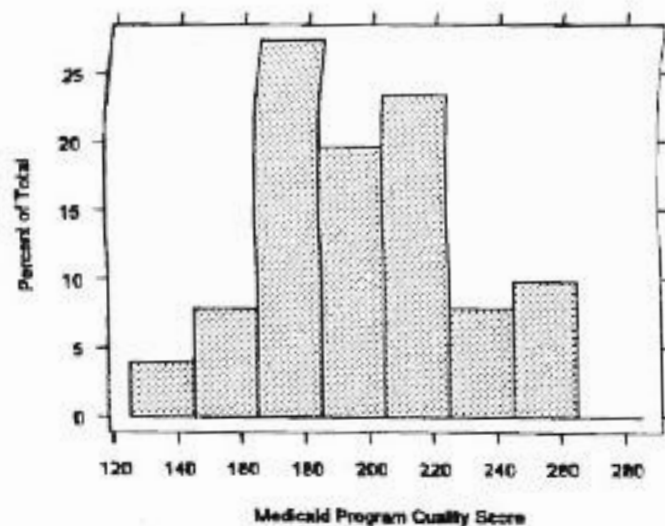


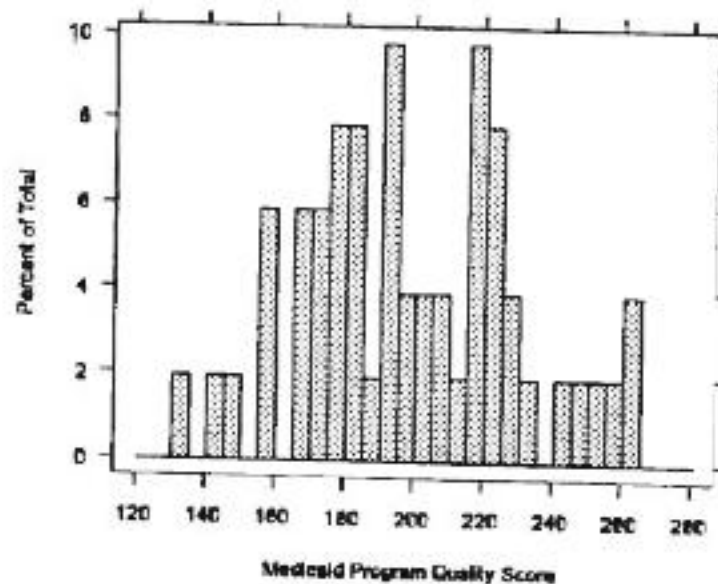
Figure 2.1. Histogram of 1986 Medicaid Program Quality Scores Within the United States

SOURCE: Public Citizen Health Research Group.

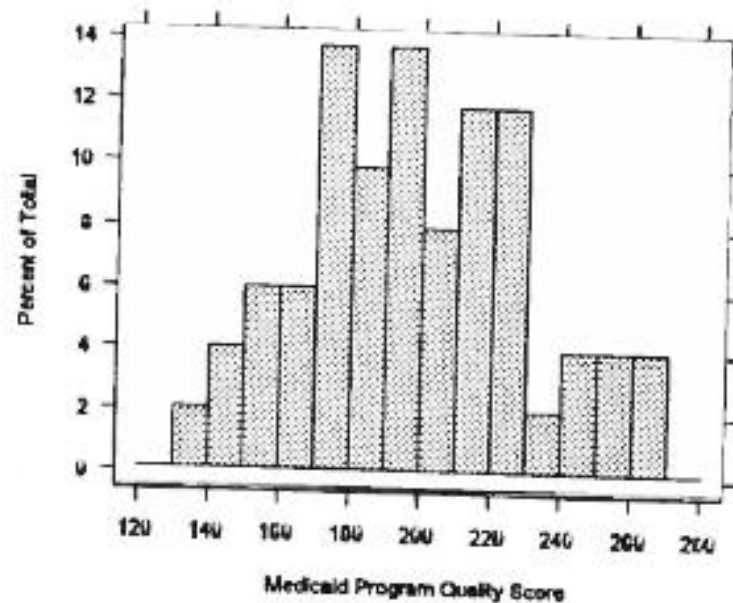
A. Bin Origin at 125, Bin Widths of 20.



B. Bin Origin at 120, Bin Width of 5.



C. Bin Origin at 120, Bin Width of 10.



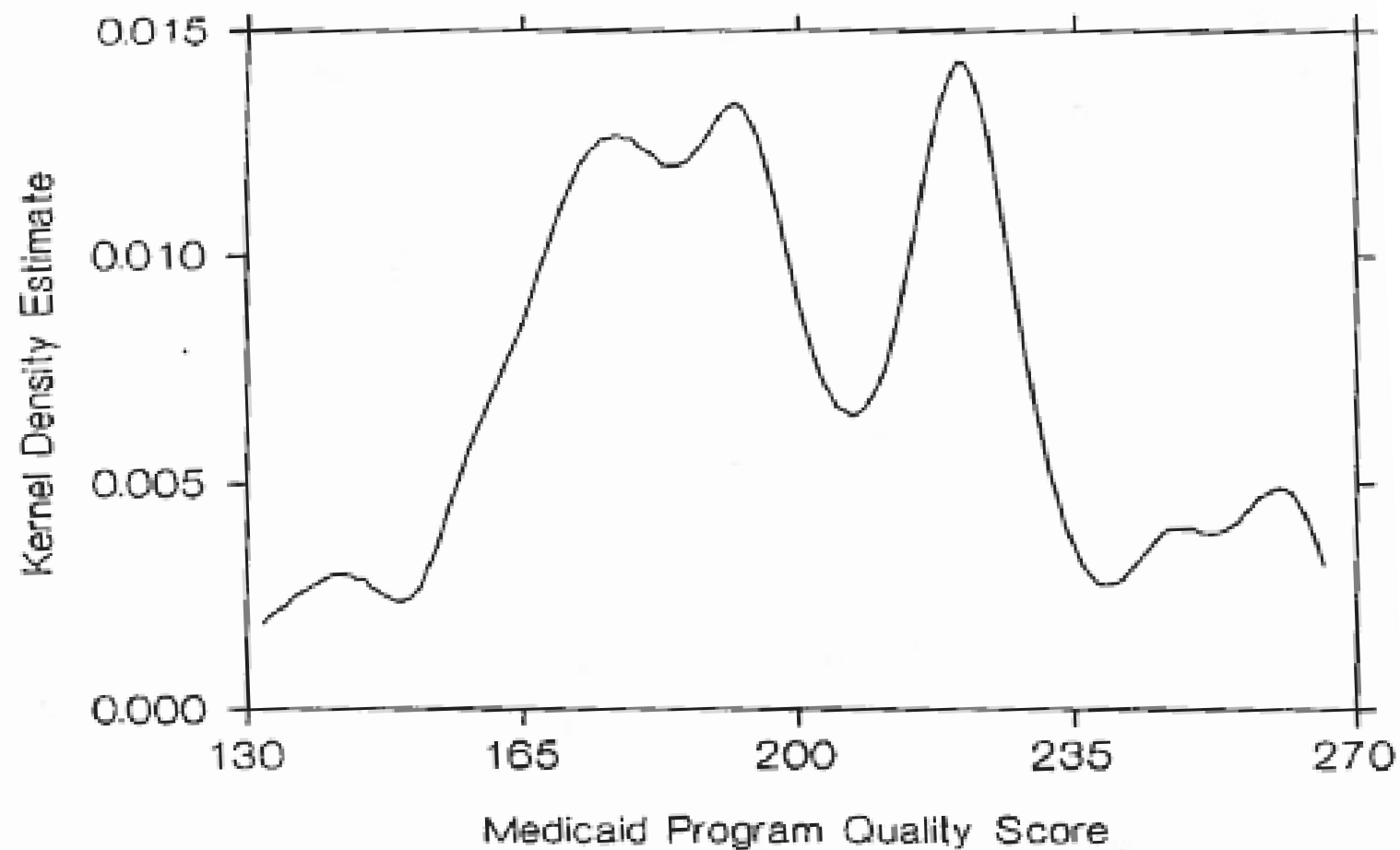
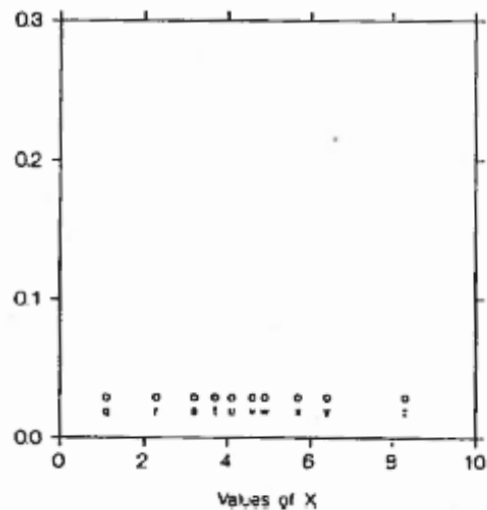
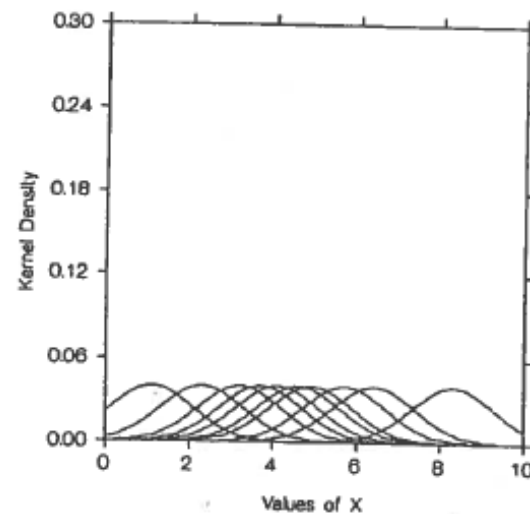


Figure 2.3. Smoothed Histogram of 1986 Medicaid Program Quality Scores
SOURCE: Public Citizen Health Research Group.

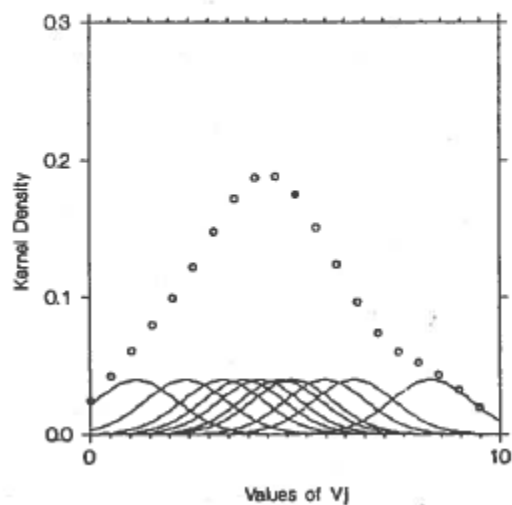
A. Unidimensional Scatterplot of 10 Data Points



B. Data Points Shown as Kernel Densities



C. Summing Kernel Densities at the 20 v_j



D. Final Smoothed Histogram

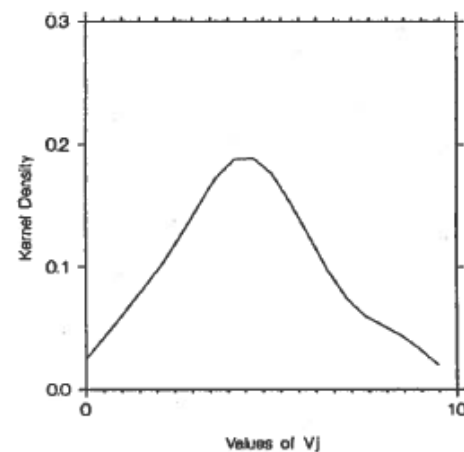


Figure 2.4. Constructing a Smoothed Histogram From Hypothetical Data

$$\hat{f}(v_j) = \frac{1}{hn} \sum_{i=1}^n K[z_{ij}] \quad (2.2)$$

$$z_{ij} = \frac{1}{h} (v_j - x_i) \quad (2.3)$$

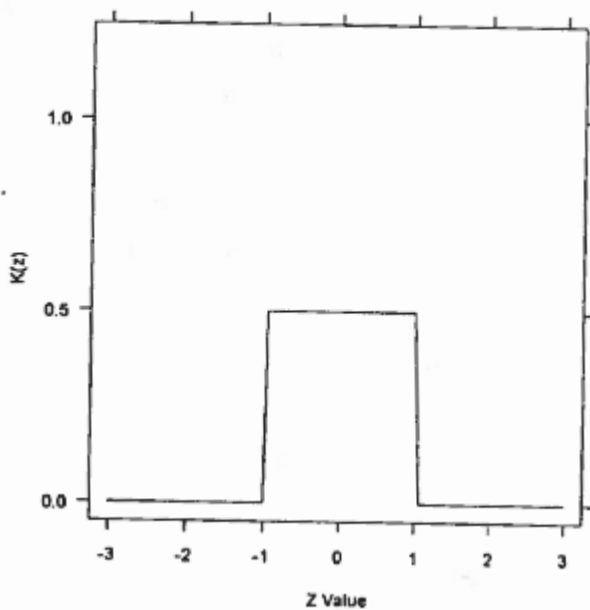
$$\text{Rectangular: } K_R(z) = \begin{cases} \frac{1}{2} & \text{for } |z| \leq 1.0 \\ 0 & \text{Otherwise} \end{cases} \quad (2.4)$$

$$\text{Triangular: } K_T(z) = \begin{cases} 1 - |z| & \text{for } |z| \leq 1.0 \\ 0 & \text{Otherwise} \end{cases} \quad (2.5)$$

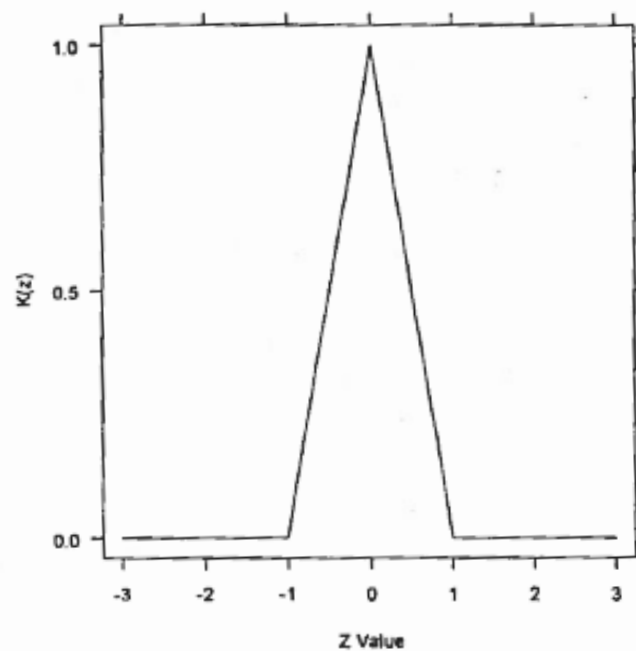
$$\text{Gaussian: } K_G(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad (2.6)$$

$$\text{Epanechnikov: } K_E(z) = \begin{cases} \frac{3}{4\sqrt{5}} \left[1 - \frac{z^2}{5} \right] & \text{for } |z| \leq \sqrt{5} \\ 0 & \text{Otherwise} \end{cases} \quad (2.7)$$

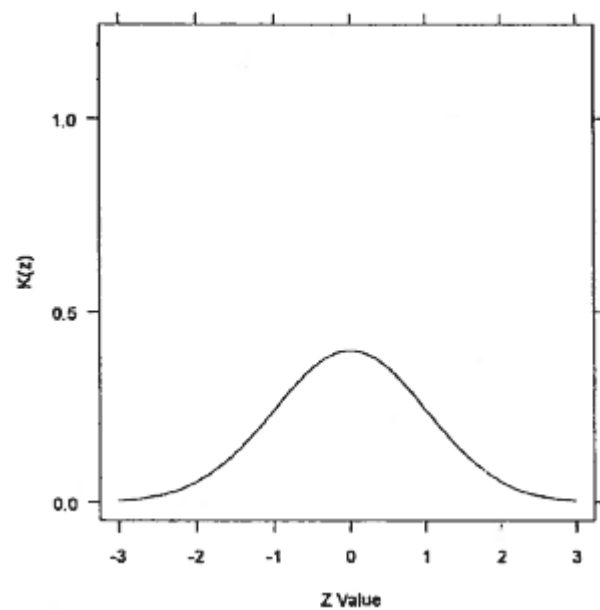
A. Rectangular Kernel



B. Triangular Kernel



C. Gaussian Kernel



D. Epanechnikov Kernel

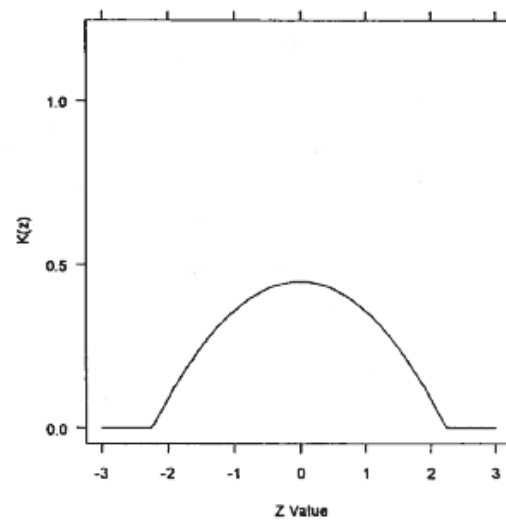
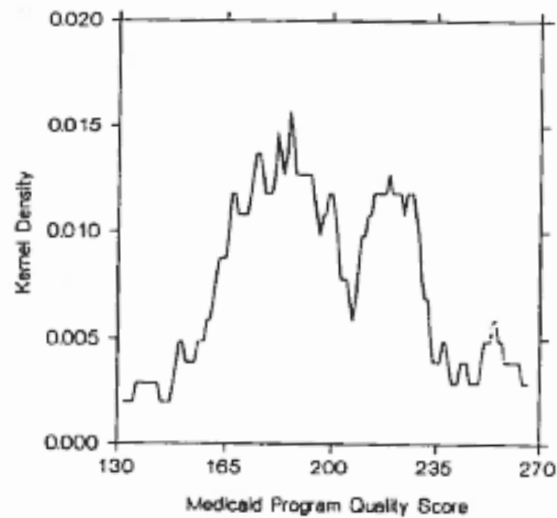
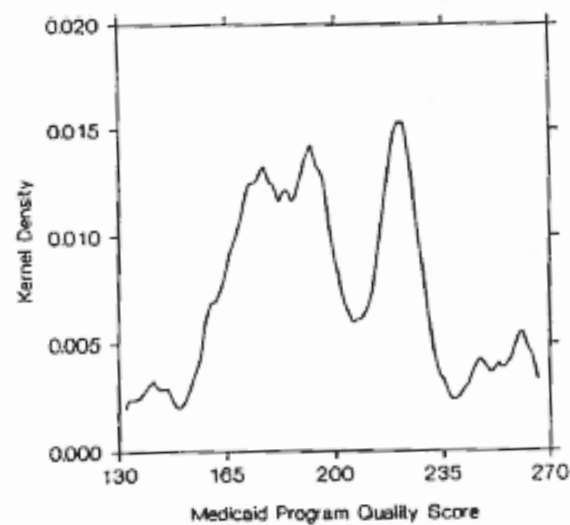


Figure 2.5. The Shapes of Various Kernel Functions

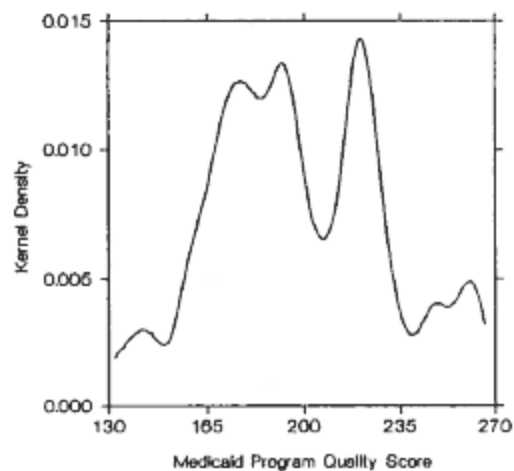
A. Rectangular Kernel



B. Triangular Kernel



C. Gaussian Kernel



D. Epanechnikov Kernel

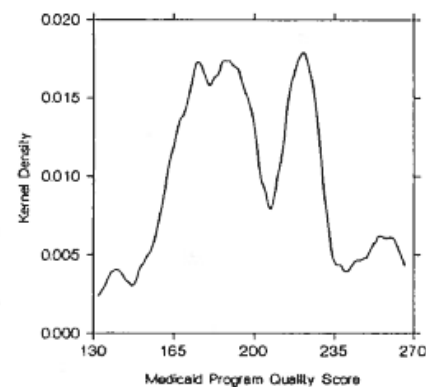
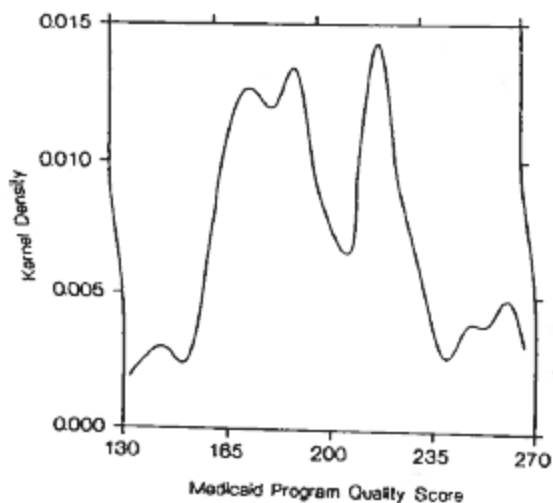
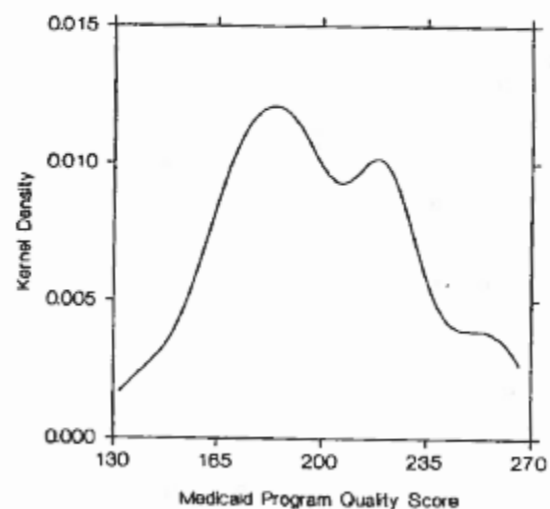


Figure 2.6. Smoothed Histograms of Medicaid Program Quality Scores Created With Different Kernel Density Functions

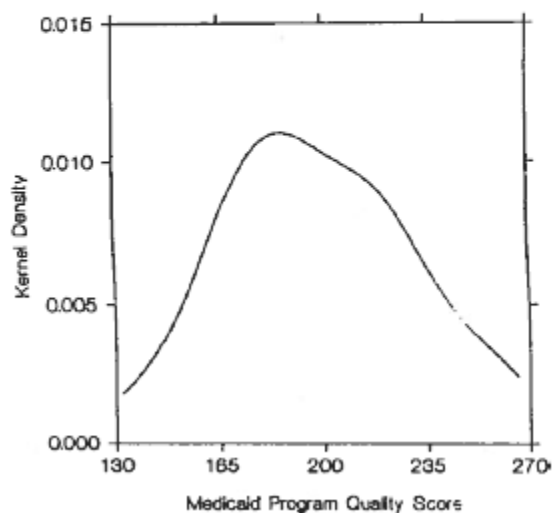
A. Bandwidth $h = 5$



B. Bandwidth, $h = 10$



C. Bandwidth, $h = 15$



D. Bandwidth, $h = 20$

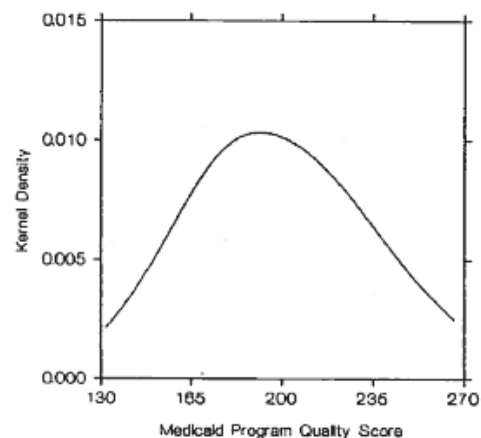


Figure 2.7. Changing the Bandwidth on Smoothed Histograms of Medicaid Program Quality Data
SOURCE: Public Citizen Health Research Group.

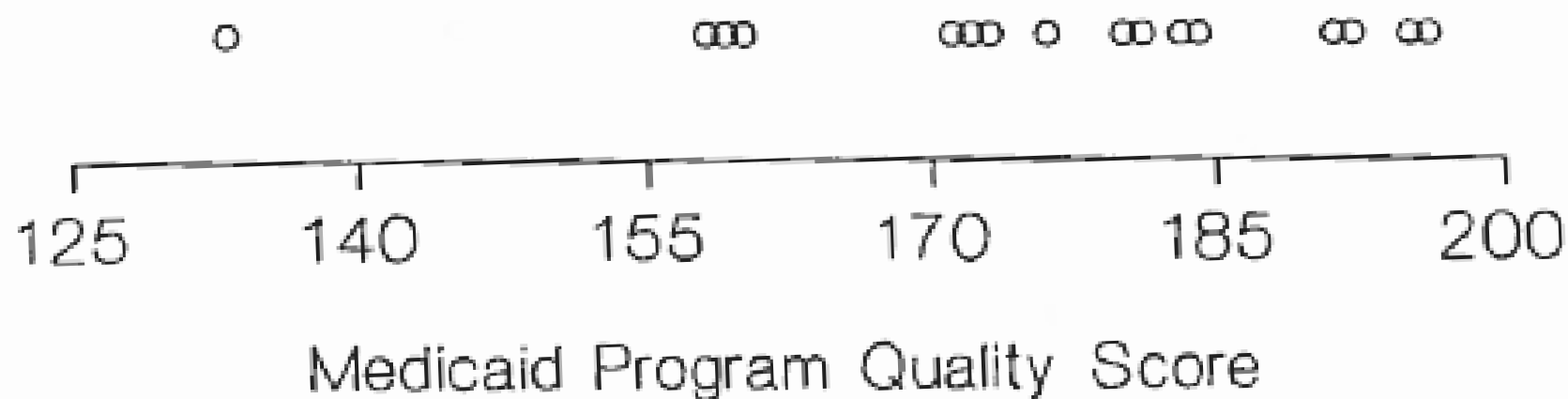
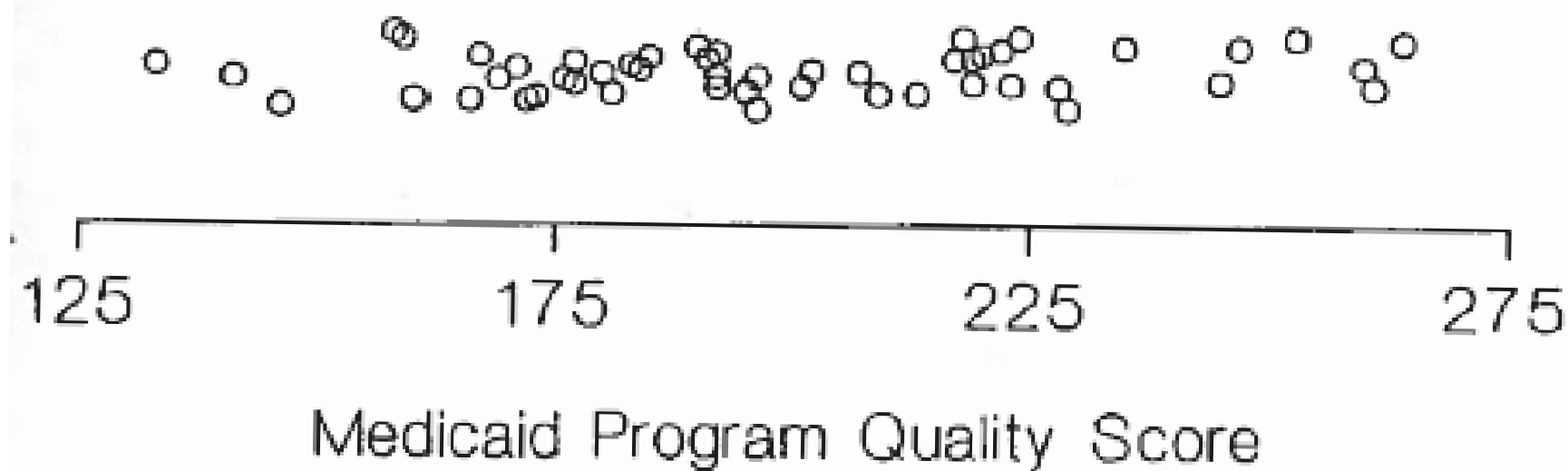


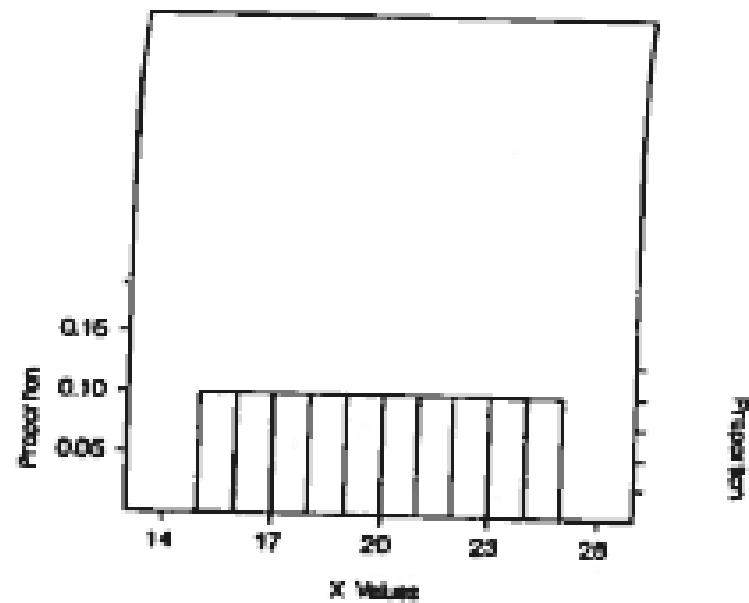
Figure 2.8. Univariate Scatterplot of Medicaid Program Quality Scores for Southern States

SOURCE: Public Citizen Health Research Group.

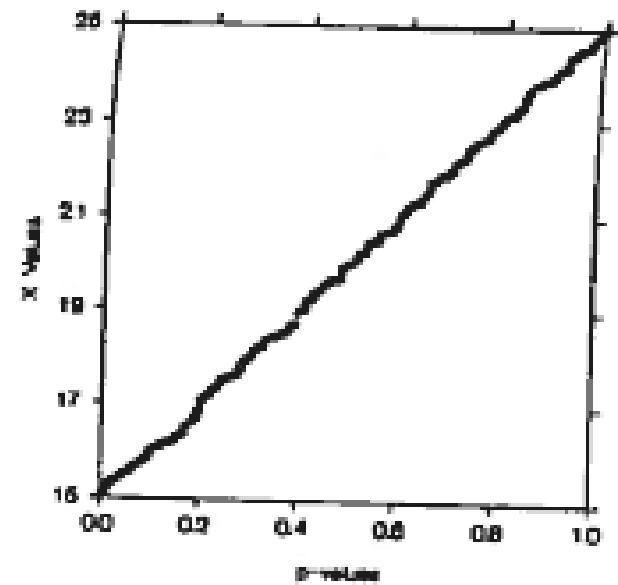


A. Uniform Distribution

Histogram

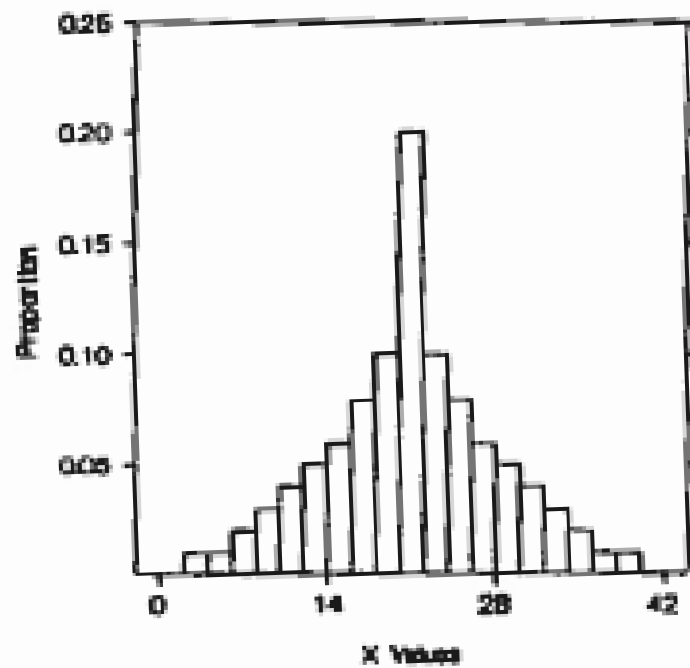


Quantile Plot

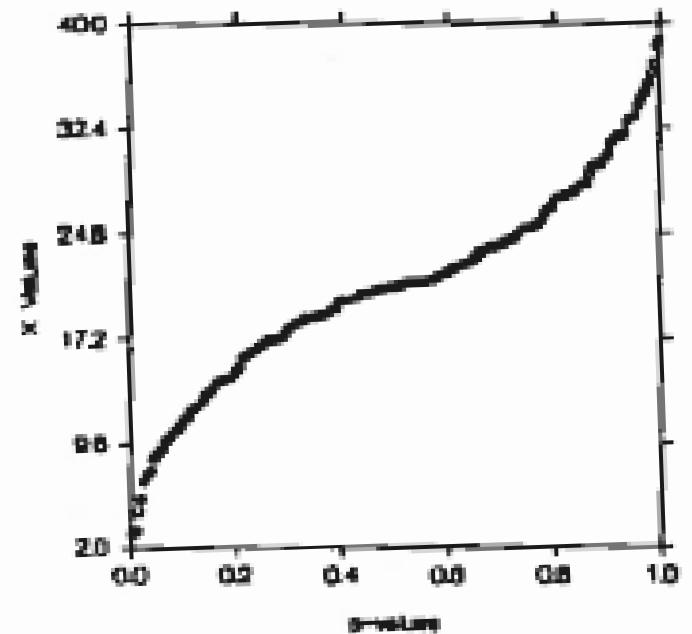


B. Symmetric, Bell-Shaped Distribution

Histogram

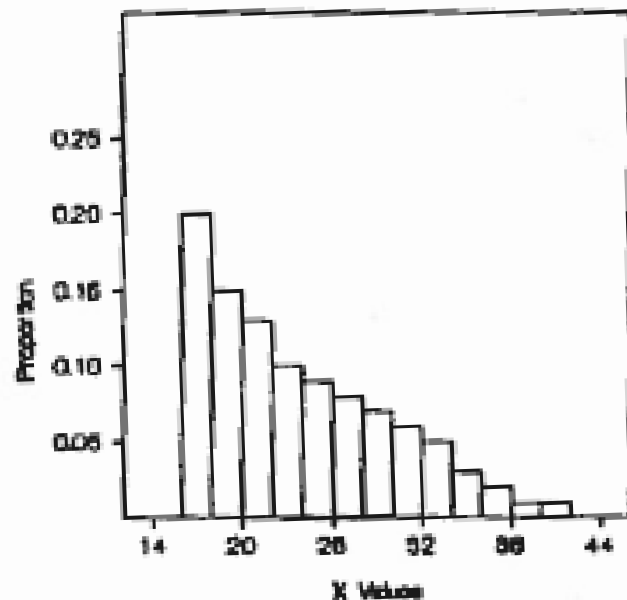


Quantile Plot



C. Positively Skewed Distribution

Histogram



Quantile Plot

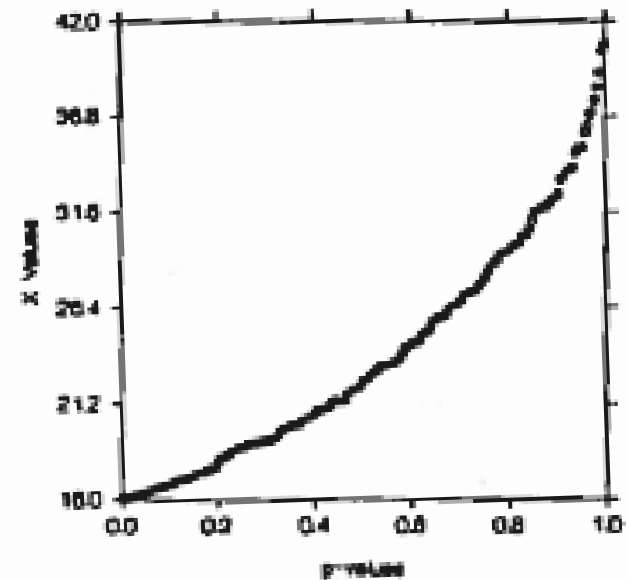


Figure 2.10. Comparison of Histograms and Quantile Plots for Differently Shaped Data Distributions

TABLE 2.1
Quantiles From the Distribution of
1986 Medicaid Program Quality Scores

State	Position Within Ordered Data (i)	Program Quality Scores $X_{(i)}$	p Values $p_i = (i - .5) / 51$
Minnesota	51	264	.990
Wisconsin	50	261	.971
New York	49	260	.951
Massachusetts	48	253	.931
Connecticut	47	247	.912
California	46	245	.892
New Jersey	45	235	.873
Washington	44	229	.853
Oregon	43	228	.833
Michigan	42	224	.814
District of Columbia	41	223	.794
Maine	40	222	.775
Iowa	39	222	.755
Maryland	38	220	.735
Vermont	37	219	.716
Rhode Island	36	219	.696
Hawaii	35	218	.676
Illinois	34	217	.657
Pennsylvania	33	213	.637
Nebraska	32	209	.618
Kansas	31	207	.598
Utah	30	202	.578
Montana	29	201	.559
Kentucky	28	196	.539
Colorado	27	196	.520
Georgia	26	195	.500
West Virginia	25	192	.480
Ohio	24	192	.461
Indiana	23	192	.441
Florida	22	191	.422
North Dakota	21	190	.402
Alaska	20	185	.382
Delaware	19	184	.363
South Carolina	18	183	.343
Tennessee	17	181	.324
North Carolina	16	180	.304
New Mexico	15	177	.284

TABLE 2.1 Continued

State	Position Within Ordered Data (i)	Program Quality Scores $X_{(i)}$	p Values $p_i = (i - .5) / 51$
New Hampshire	14	177	.265
Louisiana	13	176	.245
Texas	12	173	.225
Oklahoma	11	172	.206
Virginia	10	171	.186
Idaho	9	169	.167
Nevada	8	167	.147
South Dakota	7	166	.127
Akansas	6	160	.108
Missouri	5	159	.088
Alabama	4	158	.069
Arizona	3	146	.049
Wyoming	2	141	.029
Mississippi	1	133	.010

SOURCE: Public Citizen Health Research Group.

(continued)

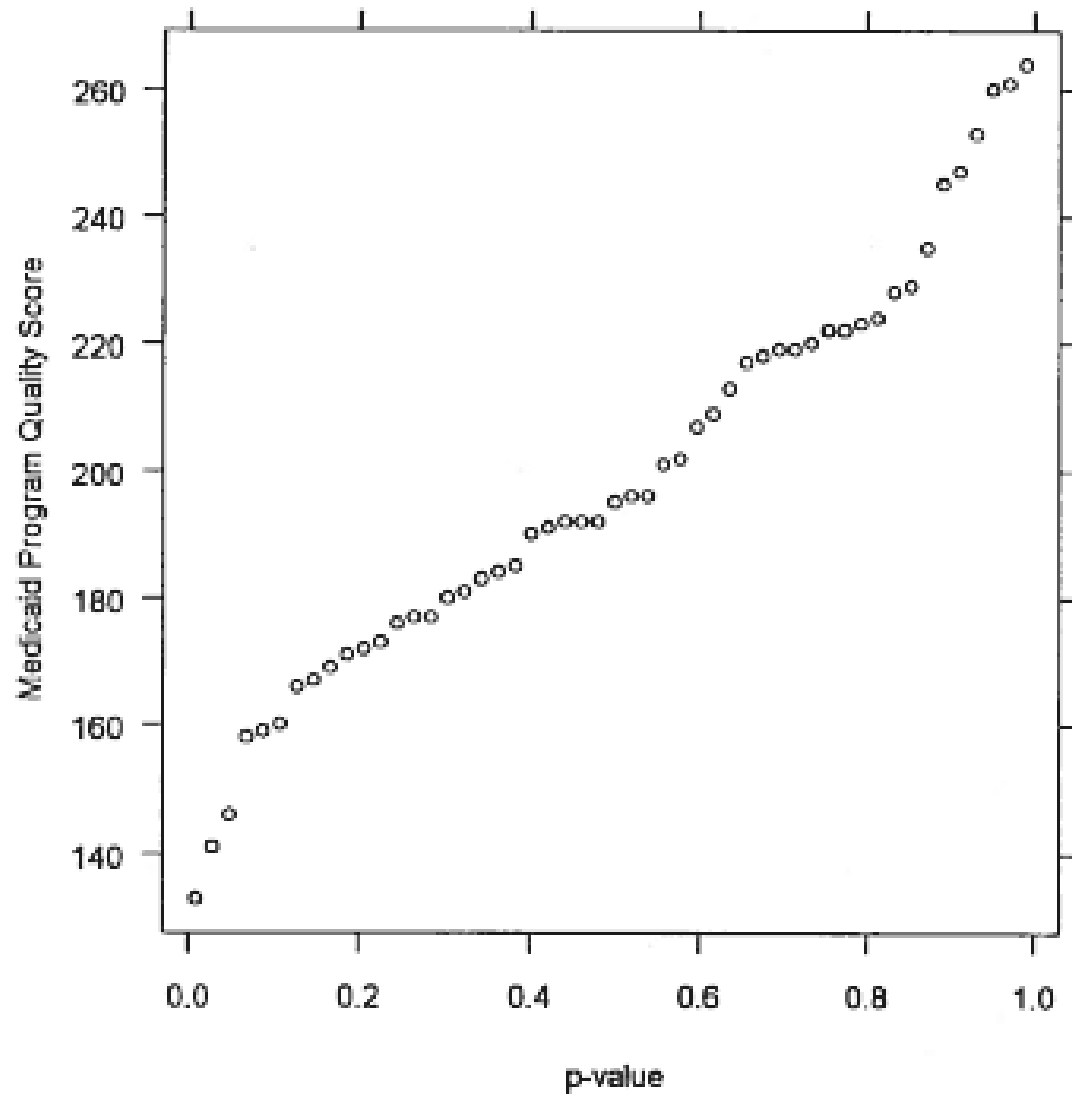


Figure 2.11. Quantile Plot of 1986 Medicaid Program Quality Scores
SOURCE: Public Citizen Health Research Group.

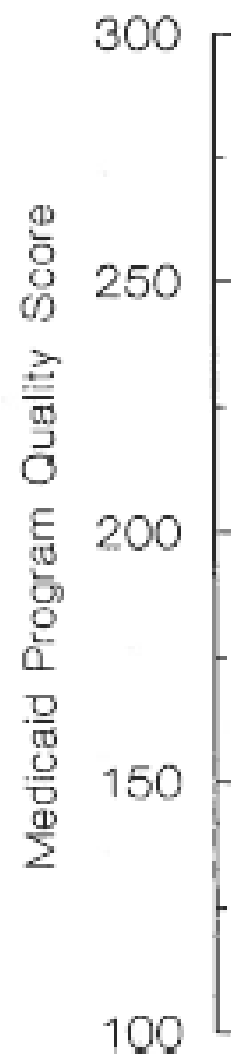


Figure 2.12. Box Plot of 1986 Medicaid Program Quality Scores
SOURCE: Public Citizen Health Research Group.

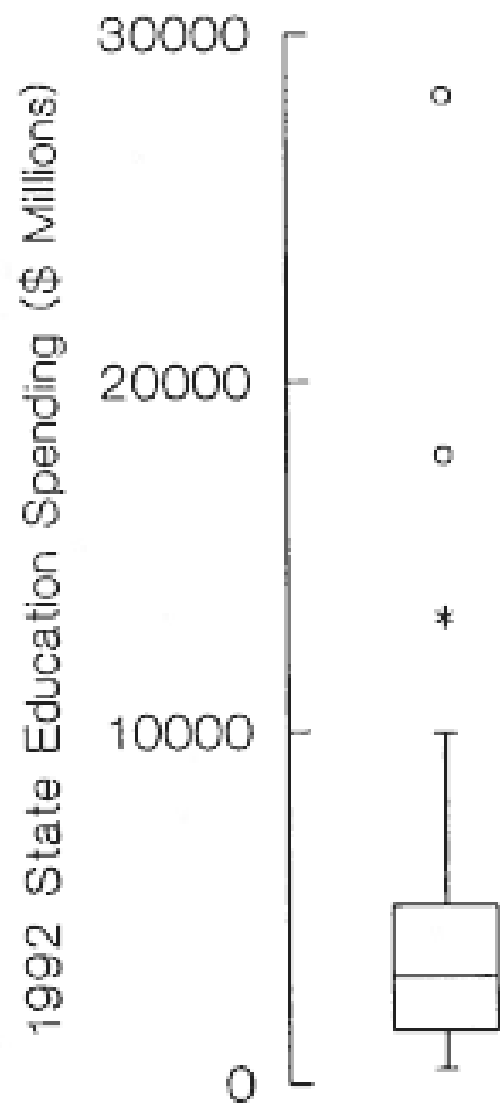


Figure 2.13. Box Plot of 1992 Public Education Expenditures in the United States
SOURCE: Public Citizen Health Research Group.

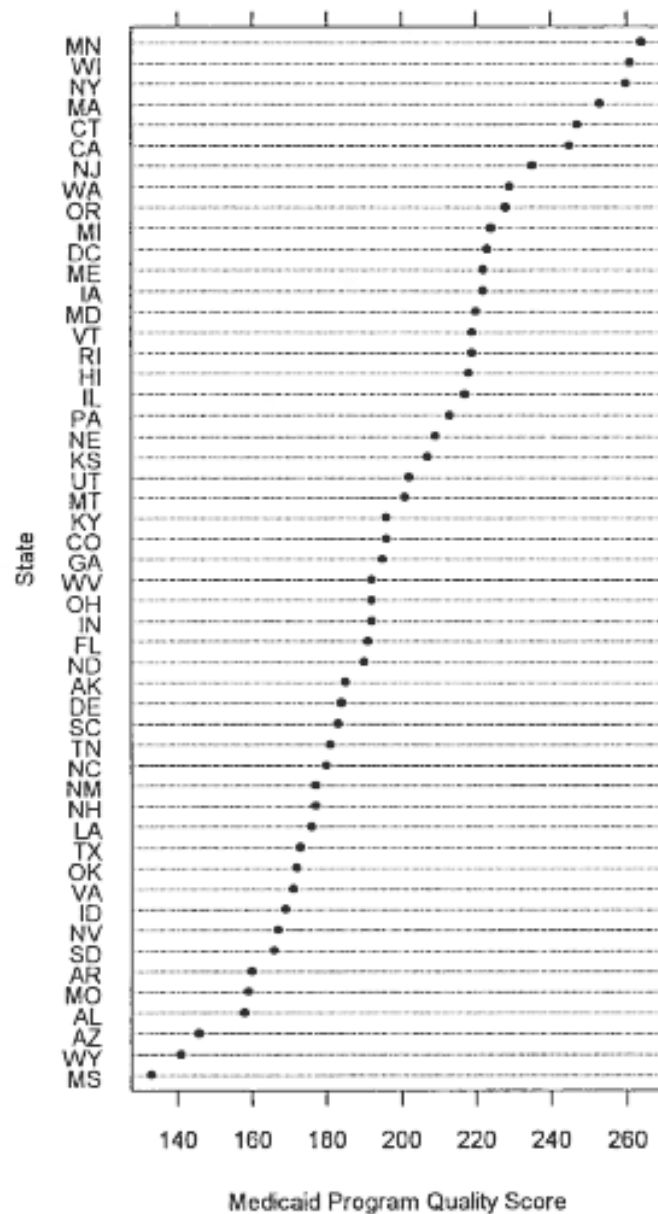


Figure 2.14. Dot Plot of State Medicaid Program Quality Scores
 SOURCE: Public Citizen Health Research Group.

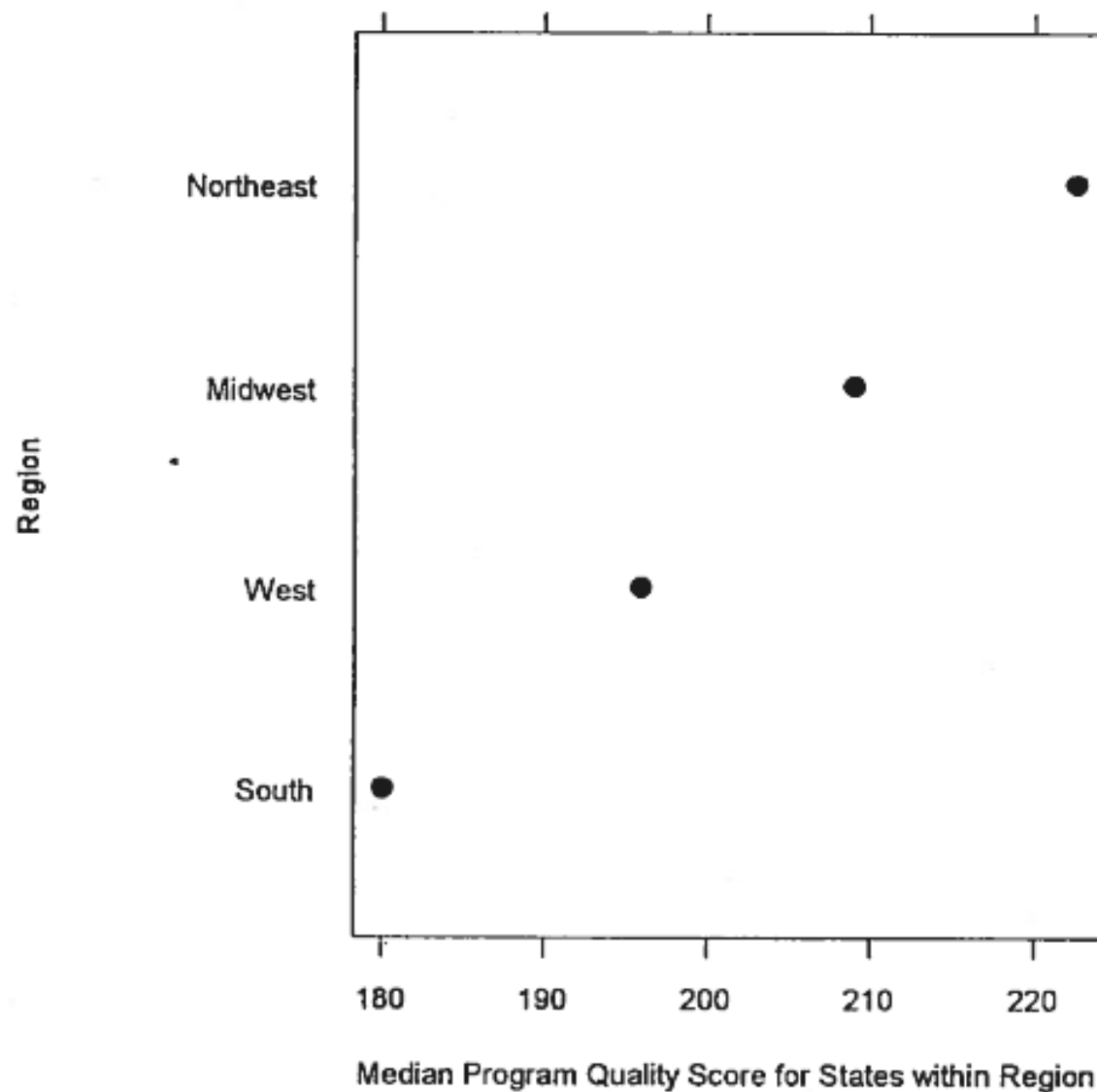
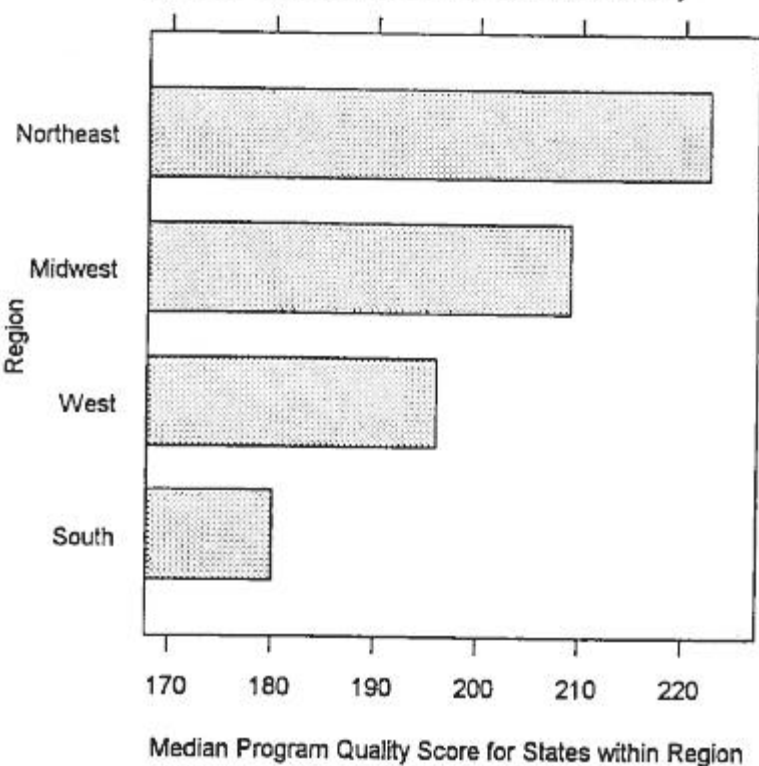


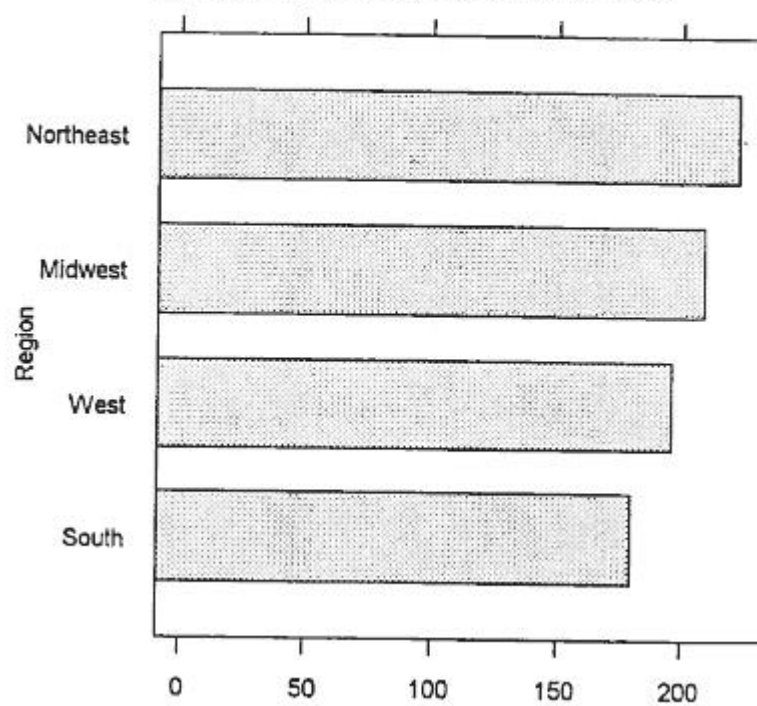
Figure 2.15. Dot Plot of Median State Medicaid Program Quality Scores Within Regions of the United States

SOURCE: Public Citizen Health Research Group.

A. Scale Minimum Value Set Arbitrarily



B. Minimum Scale Value Set to Zero



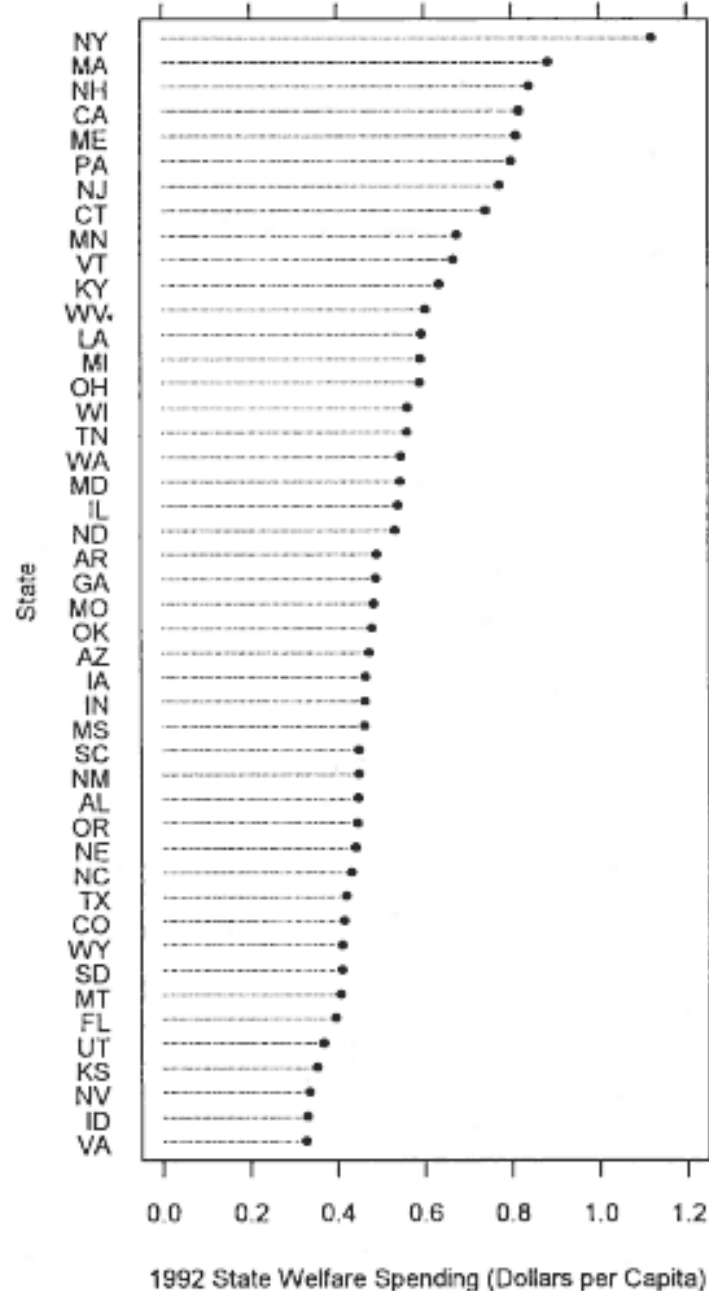
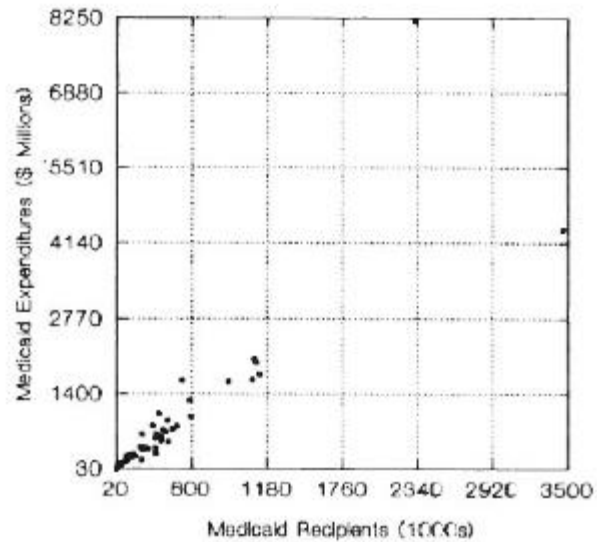


Figure 2.17. Dot Plot Showing 1992 Social Welfare Expenditures in the United States

SOURCE: 1993 Statistical Abstract of the United States.

A. Poorly-Constructed Scatterplot



B. Better Version of the Scatterplot

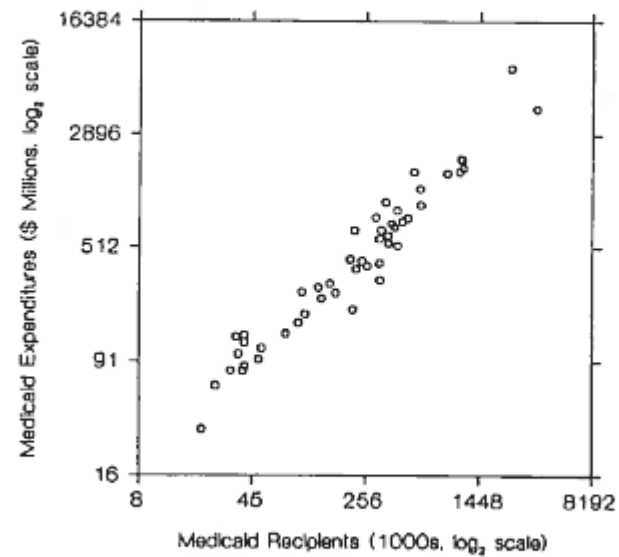
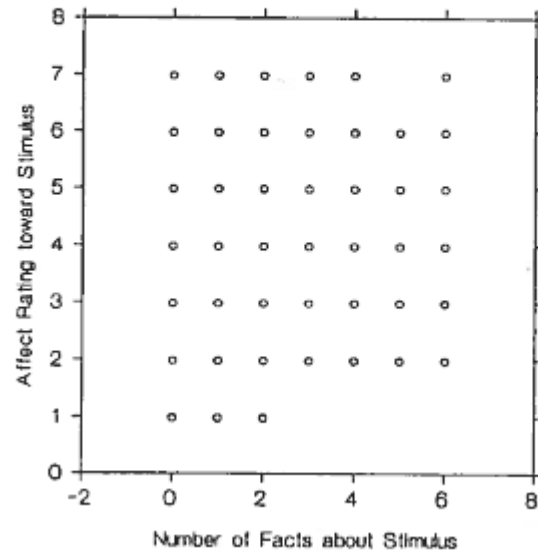


Figure 3.1. Relationship Between Medicaid Population Sizes and Medicaid Program Expenditures in the United States, 1986
SOURCE: U.S. Department of Health and Human Services.

A. Scatterplot Using Original, Discrete Data Values



B. Scatterplot Using Jittered Data Values

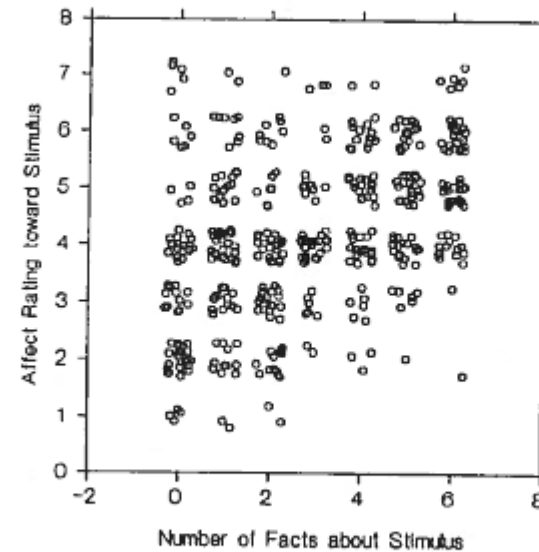


Figure 3.2. Relationship Between Information About a Stimulus Object and Affect Toward That Stimulus

SOURCE: Responses obtained from 470 experiment participants.

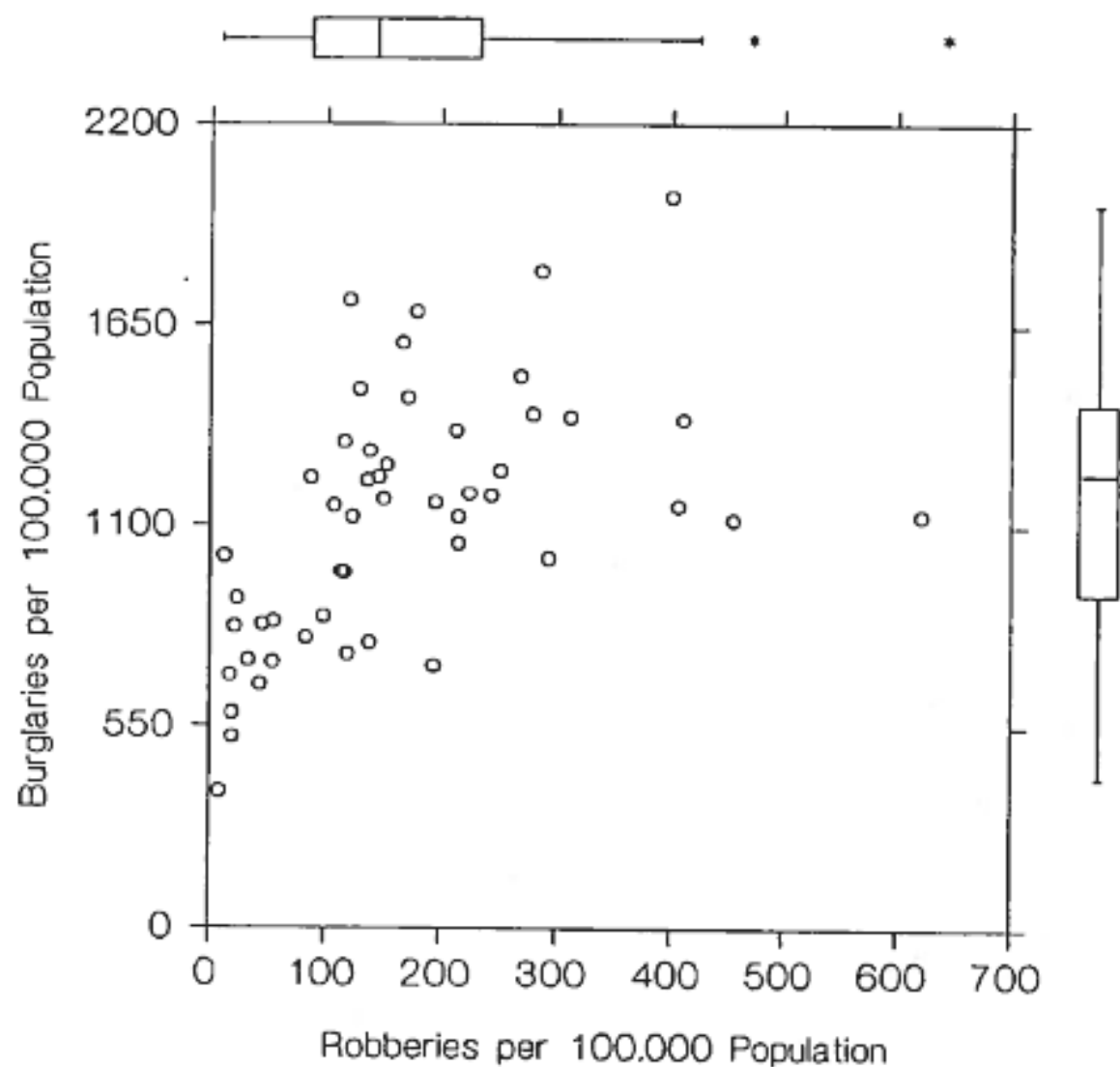
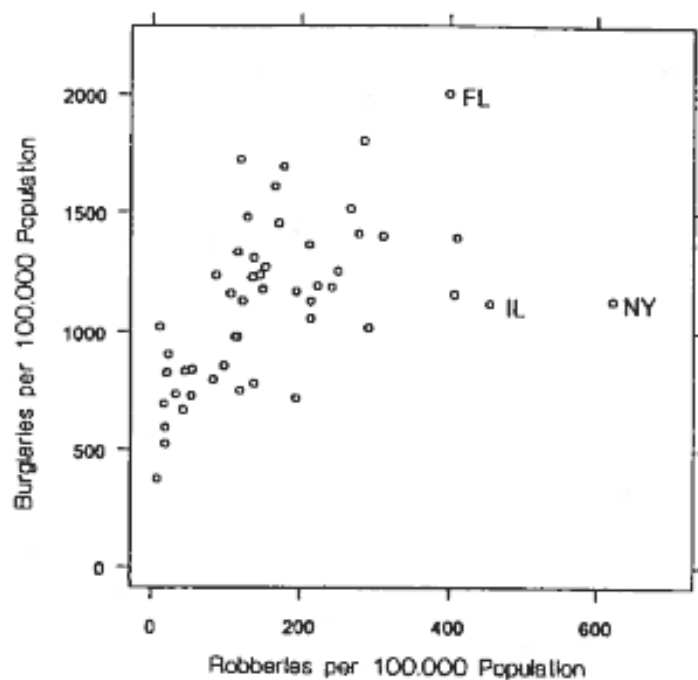


Figure 3.3. Robbery Rates Versus Burglary Rates in the United States, 1991
SOURCE: *Statistical Abstract of the United States, 1992.*

A. Only Unusual Data Points are Labelled



B. All Data Points are Labelled

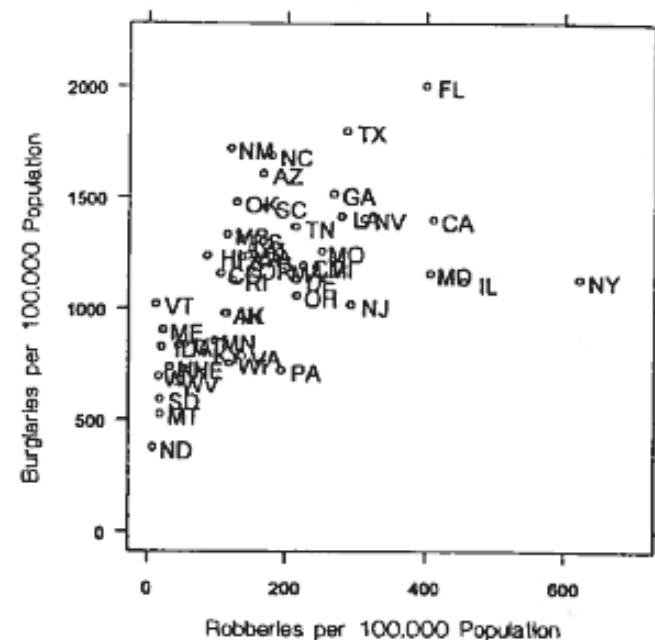
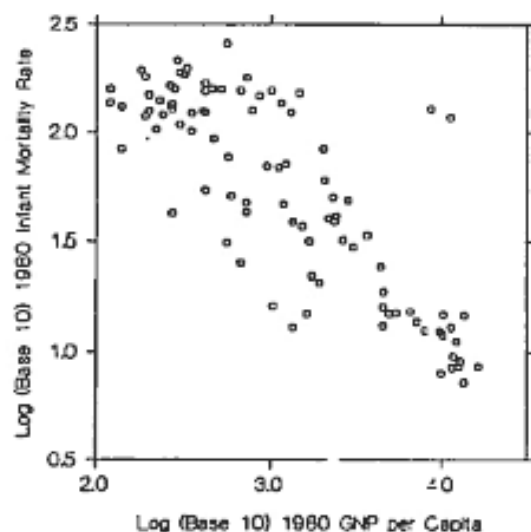
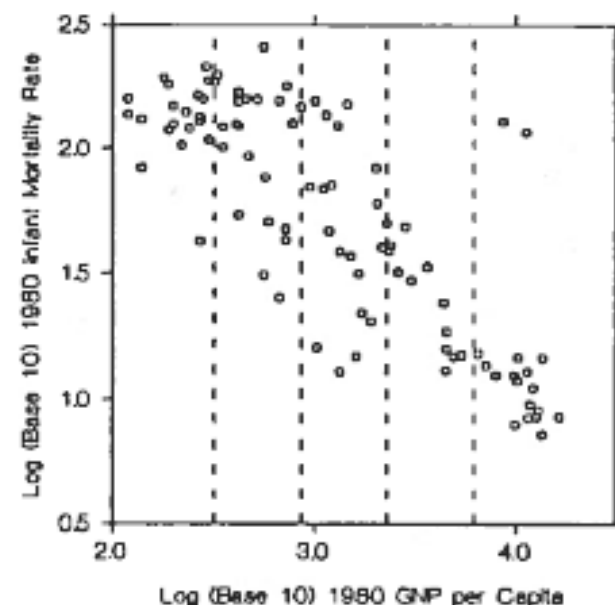


Figure 3.4. Robbery Rates Versus Burglary Rates in the United States, 1991
SOURCE: *Statistical Abstract of the United States*, 1992.

A. Basic Scatterplot



B. Slicing Intervals



C. Box Plots from Sliced Scatterplot

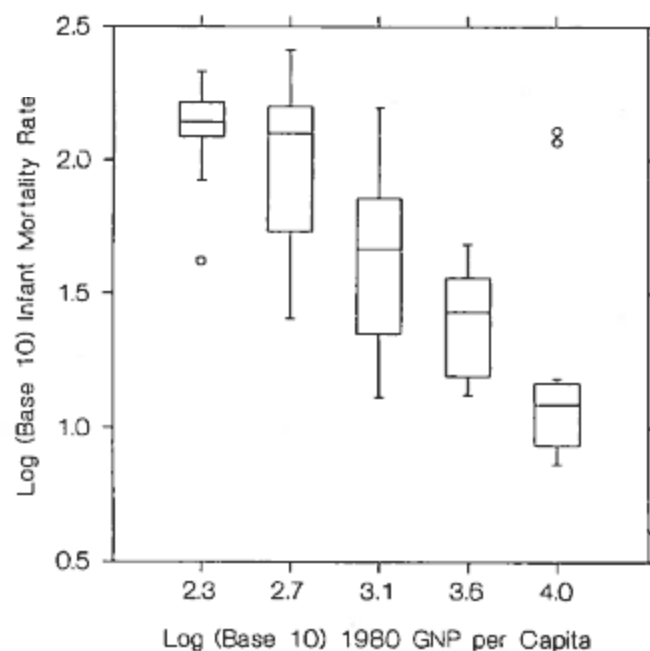
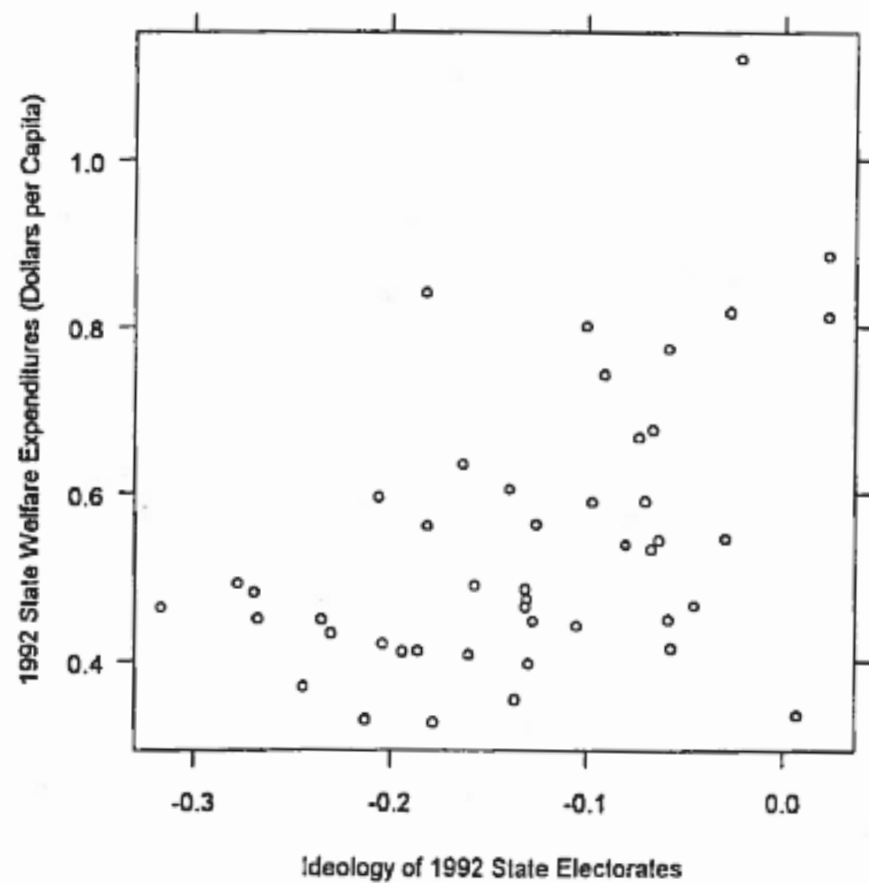


Figure 3.5. Slicing a Scatterplot of the Relationship Between 1980 GNP per Capita and Infant Mortality Rates

SOURCE: Friendly (1991).

NOTE: GNP = gross national product.

A. Basic Scatterplot



B. Scatterplot with Loess Smooth Curve

