Title
PATTERNS OF UNITED STATES MORTALITY FOR TEN SELECTED CAUSES OF DEATH

Author
Selvin, S.

Publication Date
1980-10-29

LBL-10627
UC-48

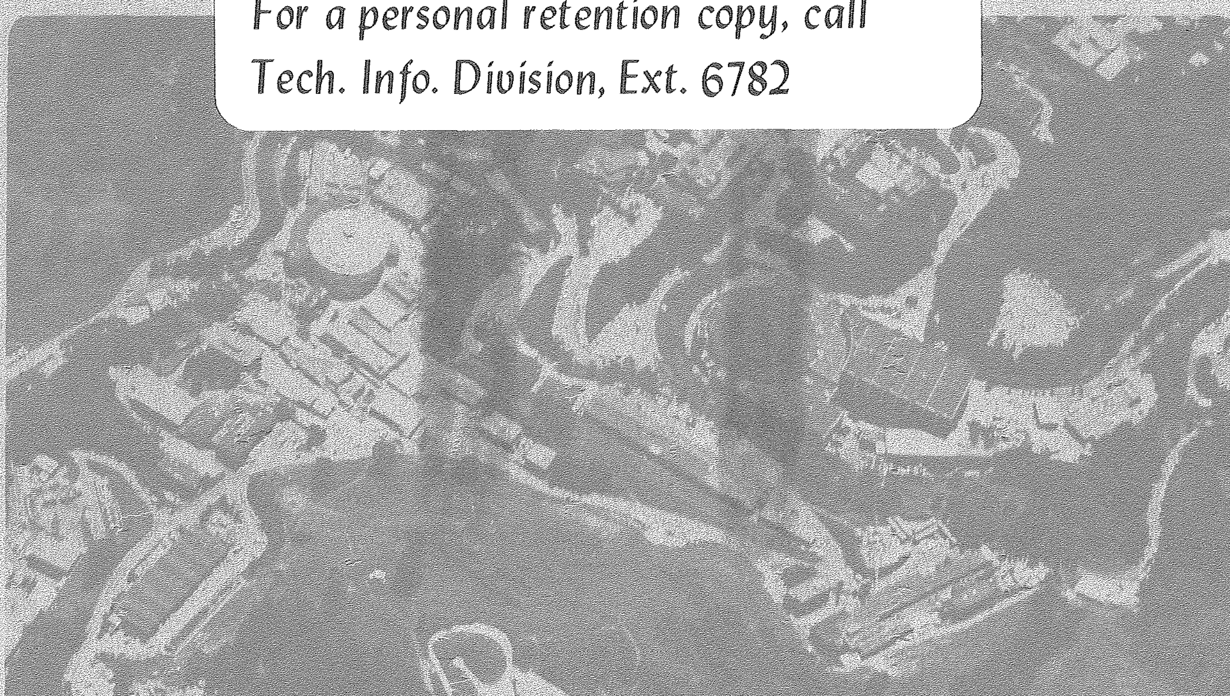
Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

Physics, Computer Science & Mathematics Division

PATTERNS OF UNITED STATES MORTALITY FOR TEN SELECTED CAUSES OF DEATH

S. Selvin, S.T. Sacks, and D.W. Merrill

November 1980



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

LBL-10627

PATTERNS OF UNITED STATES MORTALITY

FOR TEN SELECTED CAUSES OF DEATH

S. Selvin and S.T. Sacks
School of Public Health
University of California
Berkeley, CA 94720

D. W. Merrill
Computer Science and Applied Mathematics Department
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

November 6, 1980

The work described in this report was funded by the Office of Health and Environmental Research, Assistant Secretary for Environment of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098. Closely related work which contributed to the present project was funded by the U.S. Environmental Protection Agency, the U.S. Department of Employment and Training Administration, and the Electric Power Research Institute.

INTRODUCTION

Income, ethnicity, education and occupation are examples of socio-economic factors associated with the occurrence of disease, whether an investigation focuses on an individual or on an aggregation of individuals. In this study, data aggregated to the county level are used to explore two issues -- geographic variation and geographic covariation of ten selected causes of death in the United States. The counties of the United States are characterized by 15 "socio-economic" variables and age-adjusted mortality rates for the ten selected causes of death. The observed variation among the U.S. counties, as measured by the socio-economic variables, is first assessed (principal component analysis), then the geographic variation and covariation are described for each cause of death (correlation coefficients) and, finally, the covariation among causes of death is analyzed after adjusting for the influences of the measured sources of county variation (partial correlation coefficients).

MATERIAL AND METHODS

The 3,082 counties of the United States differ enormously regardless of how they are compared. (The number of U.S. counties varies depending on the definition of a county. For example, some independent cities in the state of Virginia are often considered as separate counties.) A list of the 15 socio-economic variables used to characterize the U.S. counties is given in Table 1 along with national high and low values for each variable. These 15 variables were extracted from two sources: the 1970 U.S. Census of Population (fourth count) and the United States County and City Data Book (Ref. 1). Because these variables overlap and are interrelated, the contribution of each variable to the variation among counties is difficult to assess directly. The total population of a county, for example, includes the 1970 white male population, which produces a correlation = 0.993 between white male population and total population. Similarly, the percentage of professionals residing in a county is related and, therefore, correlated with the county median educational level (correlation = 0.609). One way to measure the contribution of each variable to the total observed variation is by partitioning the variability using the method of principal components. Each principal component (linear combination) quantifies a specific stochastically independent dimension of the overall variation and, furthermore, the contribution of each individual variable can be gauged by comparing the weights (loadings) within the principal components.

Figure 1 presents the weights that make up the first four principal components derived from the 15 county characteristics. These four components explain 71.8% of the total variation measured among the 3,082 U.S. counties. The first principal component is dominated by population counts and shows county population size to be the major contributor (33.1%) to the total observed socio-economic variation among counties. The second most important contributor is educational level as measured by the median education in both male and female county populations (16.2%). The third principal component measures the independent contribution of a county's age structure as reflected by the percentage of persons over age 65 (11.8%). The fourth principal component (10.7%) is dominated by the percentage of persons employed in the work force. Figures 2-5 show the geographic variation of these principal component indices.

Figure 2 is easily interpreted, since the first principal component represents basically the population concentrations across the U.S. Figures 3,4, and 5 are less easily interpreted since they represent the independent (orthogonal) contributions of "education", "age" and "employment". That is, the influences of these four variables are depicted so that the overlapping and confounding effects of the other "socio-economic" variables summarized by the principal components are statistically removed. For example, the extremes of the distribution of the principal component representing "educational" level independent of population size are concentrated in the rural areas of the nation. Similarly, the independent contribution of the "age" principal component is highest in the midwest, Oklahoma, Texas and Florida.

The mortality experience of each county is based on age-adjusted rates from ten causes (Ref. 2): seven types of cancer (esophagus, stomach, intestine, rectum, lung, breast, and leukemia); ischemic heart disease; influenza; and suicide. These ten causes of death are among the major sources of mortality in the United States and undoubtedly have different etiologies. The mortality data, obtained from the National Center for Health Statistics (NCHS), cover the period 1968 to 1972. In 1972 the NCHS coded only half the death certificates, so that the period 1968-1972 actually covers only 4.5 years of deaths. Age-specific mortality rates were calculated using 4.5 times the U.S. 1970 census population counts, and age-adjusted rates were calculated with the total 1970 U.S. age distribution as a standard population. The study population was restricted to white males and females.

A simple comparison among age-adjusted rates is impractical since mortality rates are statistically difficult to characterize and are highly influenced by county population size (small counties being subject to large variation). For these two reasons, the mortality risk associated with each disease is measured in terms of standard deviations above and below the U.S. age-adjusted death rate. That is, a score represented by Z is produced for each county and cause of death, where

$$Z = \frac{\text{(observed deaths)} - \text{(expected deaths)}}{\text{(square root of observed deaths)}}$$

(If no deaths were observed, the denominator is the square root of the expected number of deaths. See Ref. 3 for more discussion.)

For each cause of death, this standardization yields a value whose distribution is fairly symmetric and, at the same time, has a generally stable variance (Ref. 3). The interpretation of the product-moment correlation coefficient used in subsequent analyses is improved when variables analyzed have symmetric distributions with equal variance. More importantly, comparison of Z-scores takes into account the influence of differing county size. Since county size is a major component of the standard error of an age-adjusted mortality rate (see Refs. 3 and 4), standardized scores adjust for unequal population counts and can be directly compared. For example, a small county in Georgia (Gilmer; male population = 4,397) has a high age-adjusted mortality rate for leukemia (31.2 deaths per 100,000 compared to the national rate of 9.30). However, the standardized score (Z) for this county equals 1.75 with 49 counties having higher values of Z, indicating that the elevated county rate is largely a function of the small population of the county (i.e., 49 counties have higher values of Z). Alternatively, large Z values must be attributed to factors other than small population. For example, a small county in Colorado (Lake; male population = 4,275) has the highest Z value (4.00) among all U.S. counties, indicating that chance variation in leukemia mortality occurrence is not an adequate explanation for the high rate among males in this county.

Tables 2A (males) and 2B (females) lists the five top counties, ranked by standardized Z-score, for each cause of death among white males and females, along with average annual age-adjusted mortality rates (per 100,000 population). For most causes of death, the large urban or suburban counties are ranked highest. Two exceptions exist: breast cancer among males is high in St. Helena county in Louisiana (population = 2,120) and, as mentioned, leukemia in males in Lake County, Colorado (population = 4,275) is the highest in the nation. A striking pattern is observed for suicides among females: each of the top five counties is a heavily populated county in California (population > 0.5 million females).

Geographic patterns of mortality were summarized by performing a regression analysis relating the longitude and latitude of the population centroid of each county in the nation (independent variables) to the Z-score for each disease (dependent variable). The results are summarized in Figure 6. The squared multiple correlation coefficients measure the fit of the regression planes to the Z-scores. These values indicate the degree to which latitude and longitude simultaneously predict the disease risk in a given county. The simple correlation coefficients are related to the directional slopes of the regression plane, and indicate the strength of north-south or east-west mortality trends. As in all summaries, local variation is smoothed so that specific high or low scores will not generally have explicit influences.

Cancer of the stomach, intestine and rectum show a moderate correlation with latitude ($r \geq 0.10$) or a northerly trend for both male and female scores. Ischemic heart disease shows a strong easterly trend for both sexes ($r = -0.29$ for males and $r = -0.31$ for females). Lung cancer among males has a large negative correlation with latitude and a large negative correlation with longitude, indicating a predominance of the disease in the southeastern counties of the United States. A similar but weaker trend is observed among females. Leukemia mortality, in both males and females, is not associated with either longitude or latitude.

Another aspect of geographic variation is the covariance of disease frequencies among the 3,082 U.S. counties. An often-used measure of covariation is the Pearson product-moment correlation coefficient. Figures 7 and 8 give the values of all possible pairwise comparisons among the ten standardized mortality scores (45 correlations for males and 45 correlations for females). These correlation coefficients are not rigorously tested for statistical significance; employing 90 non-independent product-moment correlations as an exploratory tool makes it virtually impossible to establish exact significance levels. The correlation coefficients given in Figures 7 and 8 are subject to sampling variation but, for the most part, cannot be considered random deviations from zero (i.e., the standard error for each coefficient is roughly 0.02).

The rates of cancer of the gastro-intestinal tract among males (esophagus, stomach, intestine, and rectum) are highly correlated for all six possible pairings ($r \geq 0.21$). This pattern is repeated, but less strongly, among females. A strong negative correlation ($r \leq -0.11$) is observed between suicide and these four cancers in males but not in females. Breast cancer in males (a rare occurrence) is not associated with any of the other nine diseases. Breast cancer in females, however, is associated with stomach, intestinal, and rectal cancers ($r \geq 0.14$). Heart disease in females shows a strong negative association ($r = -0.26$) with suicide and a positive association with rectal and intestinal cancers ($r = 0.19$ and $r = 0.13$, respectively). Leukemia is remarkable for its lack of association with any other cause of death in both males and females.

The fundamental purpose of assessing the association among diseases is to determine the extent to which common factors are directly or indirectly involved in the etiology. Many of these common factors are known and measured explicitly or implicitly by the 15 socio-economic variables. For example, an individual's income and educational level are factors associated with the rate of mortality and are reflected in a number of ways by the 15 measured county characteristics.

One method of establishing the degree to which several factors are associated with disease occurrence is to employ a multiple regression analysis. A summary of the linear association among all 15 "socio-economic" measurements and each cause-of-death score is the squared multiple correlation coefficient (R squared; Figure 9). Leukemia shows the lowest values of R squared, which are still large enough to be statistically significant (at the 0.001 level) for both males and females. Furthermore, these estimated relationships are used to adjust the cause-of-death scores for the influence of the 15 county characteristics. The resulting residual scores are thus "free" of the linear effects of the socio-economic variables. Correlations among these residual values indicate the covariation of mortality scores, with the influence from a series of known sources of variation statistically removed.

The results of applying this partial correlation approach are shown in Figures 10 and 11. A comparison of Figures 7-8 with Figures 10-11 shows the change brought about by removing the linear influence of the 15 socio-economic variables on the pairwise associations. The association among the four gastro-intestinal sites in males is reduced by about 30% but remains strong ($r \geq 0.13$). The adjusted correlations for the same cancers among females show only a moderate reduction. The effect of statistically adjusting the correlations between suicide and gastro-intestinal cancer in males for the 15 socio-economic variables is relatively small.

In females, the only cancers strongly affected by adjustment are breast cancer and intestinal cancer, which show a reduction of about 30%. The other associations are only slightly influenced by adjustment for the 15 "socio-economic" variables. For example, the association between heart disease and rectal cancer ($r = 0.19$) is not affected and the association between ischemic heart disease and suicide is only slightly reduced ($r = -0.26$ versus $r = -0.24$ adjusted).

DISCUSSION

Robinson (Ref. 5) noted that inferences drawn from aggregated data do not necessarily reflect the behavior of the individuals included in the aggregation. His work concerned the use of correlation coefficients to investigate relationships between two variables resulting from aggregation of individuals (ecologic correlation). The general pitfalls of drawing inferences from aggregated data have been explored by others (Refs. 6 and 7) and are referred to as "ecologic fallacies". Many investigators believe the "ecologic fallacy" problem is so severe that inferences made from ecologic data are flawed and do not lead to useful conclusions. The potential for misleading analyses of aggregated data is reduced, however, when interest is focused strictly on the behavior of the group and not on the individual (Ref. 8). The present investigation, which focused entirely on ecologic data, contains conclusions which must be tempered by noting that the phenomena studied are properties only of the group (county) behavior. Although the exact role of ecologic analyses in the study of disease patterns continues to be debated (Ref. 7), agreement exists that ecologic analysis does increase knowledge of mortality occurrence and can serve to detect risk factors associated with specific diseases.

Mortality rates (an ecologic variable) derived from death certificates and census counts and aggregated into county groups form the only available nationwide measure of most causes of death. Mortality rates are, however, subject to several biases. Unlike incidence rates, mortality rates provide a poor reflection of the risk of diseases characterized by rapidly changing or low case-fatality ratios. A measure whose denominator is derived from one source (U.S. census counts) and numerator from another source (National Center for Health Statistics) is particularly susceptible to biasing influences. Assigning an individual's county of residence according to the county reported on the death certificate is a long-recognized problem. For example, bias might occur if the high mortality rates observed in Nantucket county, Massachusetts, reflected deaths among a large summer population which was not present on census day (April 1, 1970). Lack of information on the migration of individuals from one county to another also adds uncertainty to mortality rates. Misclassification occurs because the exact cause of death is often equivocal. Errors in the census counts potentially affect mortality rates as well: studies of the 1960 U.S. census showed that younger groups tend to be undercounted (1% to 4.5%) and older groups overcounted. The ten causes of death in the present data set, however, involve large numbers of deaths over a relatively short period of time (4.5 years), have fairly stable mortality rates, and represent less equivocal diagnostic categories (except, perhaps, suicide). For these reasons it is unlikely that the reported mortality rates and derived scores are misleading.

In studying geographic variation, a cornerstone for many epidemiologic observations, one must often rely on a series of published maps showing the patterns of U.S. county-level mortality (Ref. 9). Such county-level maps can be difficult to interpret because of the large variations in county size, particularly in the western part of the country. Large counties with extreme mortality rates have disproportionate visual impact, and mapping techniques rarely indicate the numbers of persons affected or adjust for the influence of random fluctuations.

Correlation coefficients applied to standardized age-adjusted mortality scores provide statistical summaries of geographic trends. The results, as in all summaries, are expressed in general patterns at the expense of detail. The use of longitude and latitude as a Cartesian coordinate system, although not accounting for the curvature of the earth, introduces no meaningful systematic bias in summarizing east-west or north-south trends.

The observed correlation with latitude and longitude of scores derived from age-adjusted rates reflects a clustering of lung-cancer cases in males ($r = -0.29$ and $r = -0.19$ respectively) in the southeastern counties of the U.S. This same clustering is seen in the Atlas of Cancer Mortality (Ref. 9) for age-adjusted rates for the period 1950-1969. The tendency of gastro-intestinal cancer deaths (except cancer of the esophagus) for both males and females to concentrate in the eastern portion of the U.S. is not as easily recognized from maps. Heart disease in both sexes also shows a tendency to cluster in the eastern U.S. The similarities in the geographic patterns of gastro-intestinal cancers and heart disease for both males and females suggest the existence of common factors associated with the occurrence of these two diseases. When several causes of death exhibit geographic clustering, it is reasonable to postulate common environmental agents as major contributors to their etiologies. However, when differences in cause-specific mortality among geographic regions are observed, other important factors must be considered. The groups compared are likely to differ in a variety of ways that affect disease occurrence. Ethnic make-up, dietary habits, and general occupational patterns are a few of the factors that complicate the interpretation of geographic comparisons.

In contrast, suicide and breast cancer show two distinct patterns for males and females. Breast cancer among females is found with high frequency in northern counties, whereas the distribution among males is more or less random. Similarly, the high rates of female suicide are concentrated, as noted, in heavily populated western counties, but no particular geographic trend is observed in male suicides. This observation reinforces indications that (expectedly) breast cancer and (surprisingly) suicide involve largely different mechanisms which are sex-related.

When measurements of disease risk vary together, it is typically hypothesized that such variation is due to common factors. Conversely, if disease rates do not vary together, it is likely that different causal mechanisms are important. From this point of view, an interesting phenomenon is the covariation of esophagus, stomach, intestinal, and rectal cancers. These four gastro-intestinal cancer sites are highly correlated with each other. This association was noted by Winkelstein et al. (Ref. 10) in a nine-area study of cancer incidence data, as well as by Hoover et al (Ref. 11). A complete summary of cancer correlation studies, for both the U.S. and internationally, is contained in Winkelstein (Ref. 10). Furthermore, these four cancers remain strongly associated when adjusted for the 15 socio-economic variables. Although the variation in gastro-intestinal cancer deaths cannot be explained by a single mechanism, the fact that these cancers show a high degree of covariation across a large number of counties, and remain correlated when adjusted for "socio-economic" influences, suggests that a general environmental component is implicated in the etiology of these cancer sites.

The reduction noted when partial correlations are compared to product-moment correlations is to some extent an artifact of employing Z-scores. An age-specific mortality rate is essentially independent of population size whereas a Z-score is not. Therefore the observed reduction partially results from statistically adjusting the measure of association for the influence of population size.

The frequency of leukemia is not associated with any of the nine other causes of death, nor with any specific geographic regions, and is associated to only a small degree with the 15 socio-economic variables. The obvious implication is that the causal mechanisms for leukemia differ considerably from those underlying the nine other causes of deaths. Etiologic factors for leukemia are either not significantly associated with the 15 socio-economic variables, or operate on an individual level not detectable in aggregated data (e.g. genetic, dietary, personal, or occupational factors). Another possible explanation of the observed lack of association is that age-adjusted leukemia rates combine childhood leukemia (acute lymphatic) and adult leukemia into one summary rate, thus combining the rates of different disease entities and producing a spurious lack of association. A more complete study of leukemia is under way in which leukemia deaths are classified by each of four explicit International Classification of Disease codes (204 to 207).

The strong negative correlation between male suicide and the gastro-intestinal cancers was unexpected. This relationship could possibly result from the ethnic heterogeneity in the United States white population. Many geographic areas are fairly homogenous with respect to their ethnic composition, and it may be that those ethnic groups with low suicide rates have high gastro-intestinal mortality rates or vice versa. This situation would produce negative correlations between suicide and gastro-intestinal cancer rates due to the associations with a third variable, i.e. ethnicity. This hypothesis is purely conjectural and can only be investigated in data where ethnicity is recorded. This type of problem demonstrates the ever-present possibility that observed correlations between two variables arise from associations induced by relationships with a third, unmeasured, source.

ACKNOWLEDGMENTS

The work described here is part of the ongoing PAREP (Populations at Risk to Environmental Pollution) project, a collaboration between the LBL Energy and Environment (E & E) Division, the LBL Computer Science and Applied Mathematics (CSAM) Department and the University of California (at Berkeley) School of Public Health (SPH). Craig Hollowell (E & E), Donald M. Austin (CSAM), and Warren Winklestein (SPH) have acted as project managers. Past and present funding was obtained from the Environmental Protection Agency (Bill Nelson), the Department of Energy (Walter Weyzen and John Viren), and the Electric Power Research Institute (Ron Wyzga).

This analysis made extensive use of SEEDIS, the Socio-Economic Environmental Demographic Information System being developed by CSAM under an interagency agreement between the Department of Energy and the Department of Labor, Employment and Training Administration.

The socio-economic data from the 1970 census and the City-County Data Book were obtained through the efforts of Fred Gey, Bob Healey, and Bruce Burkhart. Other LBL personnel - primarily Barbara Levine, Linda Kwok and Simcha Knif - made significant efforts toward the completion of the data files used in this analysis.

The 1968-1972 age-adjusted mortality data used in this analysis were originally obtained from the National Center for Health Statistics, tabulated by Herbert Sauer of the University of Missouri, and provided to Lawrence Berkeley Laboratory by Larry Milask of the Council for Environmental Quality.

REFERENCES

1. U.S. Department of Commerce, Bureau of the Census, County and City Data Book (Washington, D.C.: Government Printing Office, 1970).

2. Age-adjusted mortality data for 1968-1972, tabulated by H. Sauer, University of Missouri. Computer tapes provided by U.S. Council for Environmental Quality.

3. S. Selvin, S. Sacks, and D.W. Merrill, Standardization of Age-adjusted Mortality Rates, Lawrence Berkeley Laboratory Report LBL-10323 (February 1980).

4. C.L. Chiang, Standard Error of the Age-adjusted Death Rate, vol. 47, no. 9 (Washington, D.C.: U.S. Department of Health, Education and Welfare, 1961).

5. W.S. Robinson, "Ecological Correlations and the Behavior of Individuals," Amer. Soc. Rev. 15 (1950):351-357.

6. S.W. Kasl, "Mortality and the Business Cycle: Some Questions about Research Strategies When Utilizing Macro-social and Ecological Data," Amer. J. of Pub. Health 69 (1979):784-788.

7. K. Stavraky, "The Role of Ecologic Analysis in Studies of the Etiology of Disease: A Discussion with Reference to Large Bowel Cancer," J. Chron. Dis. 29 (1976):435-444.

8. L.A. Goodman, "Ecologic Regressions and the Behavior of Individuals," Amer. Soc. Rev. 18 (1953):663-664.

9. T.J. Mason, F.W. McKay, R. Hoover, W.J. Blot, and J.F. Fraumeni, Atlas of Cancer Mortality 1950-1969 (Washington, D.C.: U.S. Department of Health, Education and Welfare, 1975).

10. W. Winkelstein, S.T. Sacks, V.L. Ernster, and S. Selvin, "Correlations of Incidence Rates for Selected Cancers in the Nine Areas of the Third National Cancer Survey," Amer. J. Epidem. 105 (1977):407-419.

11. R. Hoover, T.J. Mason, and F.W. McKay, "Geographic Patterns of Cancer Mortality in the United States," in Persons at High Risk of Cancer, edited by J.F. Fraumeni (New York: Academic Press, 1975).

Figure 10. The "socio-economic-adjusted" correlation coefficients for the ten selected causes of death for males.

Figure 11. The "socio-economic-adjusted" correlation coefficients for the ten selected causes of death for females.

Table 1. The ranges of the 15 "socio-economic" variables.

Variable	Highest Value	Lowest Value
1. pop. 1970(a)	7,036,463 (Los Angeles, CA)	73 (Loving, TX)
2. pop. white male 1970	2,918,916 (Los Angeles, CA)	40 (Loving, TX)
3. pop. white female 1970	3,111,115 (Los Angeles, CA)	33 (Loving, TX)
4. pop. density/sq. mile	66,923 (New York, NY)	<1.0 (many)
5. % urban(b)	100.0 (many)	0.0 (many)
6. % foreign	66.8 (Maverick, TX)	0.0 (many)
7. % black	81.1 (Macon, AL)	0.0 (many)
8. % age 65+ male	39.5 (Charlotte, FL)	0.0 (Daggett, UT)
9. % age 65+ female	33.4 (Manatee, FL)	0.0 (Loving, TX)
10. % employed(c)	96.2 (King, TX)	12.9 (Pulaski, MO)
11. % professional(d)	63.0 (Los Alamos, NM)	0.0 (2 counties)
12. med. educ. male, yrs	15.8 (Los Alamos, NM)	5.3 (Kenedy, TX)
13. med. educ. female, yrs	14.0 (Pitkin, CO)	5.1 (Zavala, TX)
14. av. fam. income 1969	18,611 (Hinsdale, CO)	2,467 (Owsley, KY)
15. % non-resident 1965(e)	33.7 (Monroe, KY)	0.0 (2 counties)

(a) Covers all races for the entire United States. All other variables, except for percent black, refer to the individuals who reported their race as white in the U.S. Census.
(b) Percentage of persons residing within the urban portions of a county, as defined by the U.S. Census.
(c) Percentage of the work force that is employed.
(d) Percentage of the employed population $\geq$ 16 years of age, that is employed in professional occupations.
(e) As of 1970, the percentage of county residents who resided in a different county five years earlier.

Table 2A. The five top ranked counties by Z-score for 10 selected causes of death along with the age-adjusted mortality rates per 100,000 (white males -- 1968-1972). (ICDA = International Classification of Diseases and Accidents).

Cancer of the Esophagus
(ICDA = 150): U.S. Rate = 4.22

			Rate	Score
1.	Cuyahoga	OH	7.34	6.29
2.	Cook	IL	5.78	6.23
3.	Wayne	MI	6.15	5.06
4.	Prince George	MD	6.25	4.96
5.	Suffolk	MA	8.25	4.92

Cancer of the Stomach
(ICDA = 151): US rate = 10.24

			Rate	Score
1.	Cook	IL	13.09	7.57
2.	Middlesex	NJ	18.61	6.82
3.	Wayne	MI	14.04	6.59
4.	Cuyahoga	OH	14.74	6.41
5.	New York	NY	15.10	6.00

Cancer of the Intestine
(ICDA = 152-3): US rate = 19.20

			Rate	Score
1.	Philadelphia	PA	28.11	8.78
2.	Nassau	NY	27.32	8.43
3.	Cook	IL	23.38	8.27
4.	New York	NY	26.74	6.97
5.	Middlesex	MA	25.09	6.38

Cancer of the Rectum
(ICDA = 160-3): US rate = 6.61

			Rate	Score
1.	Cook	IL	9.30	8.51
2.	Hudson	NJ	13.92	6.70
3.	Erie	NY	11.00	6.18
4.	Allegheny	PA	9.54	5.31
5.	New York	NY	9.85	4.96

Cancer of the Lung
(ICDA = 160-3): US rate = 6.61

			Rate	Score
1.	Baltimore	MD	100.11	12.27
2.	Duval	FL	98.97	11.33
3.	Harris	TX	78.40	10.41
4.	Philadelphia	PA	78.38	9.82
5.	New Orleans	LA	98.06	9.59

Cancer of the Breast
(ICDA = 174): US rate = .29

			Rate	Score
1.	Suffolk	NY	.77	2.66
2.	Prince George	MD	.88	2.23
3.	Wayne	MI	.52	2.08
4.	St. Helena	LA	41.30	1.97
5.	Muscogee	GA	1.85	1.88

Leukemia
(ICDA = 204-7): US rate = 9.30

			Rate	Score
1.	Lake	CO	100.73	4.00
2.	Franklin	OH	12.47	3.57
3.	Tarrant	TX	12.55	3.43
4.	Taylor	TX	17.73	2.84
5.	Genesee	MI	12.74	2.81

Acute Ischemic Heart Disease
(ICDA = 410-1): US rate = 272.75

			Rate	Score
1.	Cook	IL	318.96	24.88
2.	Erie	NY	347.73	18.74
3.	Davidson	TN	372.29	14.33
4.	Greenville	SC	405.94	13.76
5.	Genesee	MI	362.59	13.73

Influenza and Pheumonia
(ICDA = 470-86): US rate = 39.34

			Rate	Score
1.	Suffolk	MA	92.38	19.83
2.	Middlesex	MA	59.07	13.98
3.	New York	NY	55.71	10.52
4.	Denver	CO	65.81	10.20
5.	Fulton	GA	68.18	9.85

Suicide
(ICDA = E950-9): US rate = 18.30

			Rate	Score
1.	Los Angeles	CA	25.85	17.01
2.	San Francisco	CA	43.76	12.77
3.	Harris	TX	24.80	7.21
4.	Denver	CO	30.13	6.74
5.	Alameda	CA	25.30	6.06

Table 2B. The five top ranked counties by Z-score for 10 selected causes of death along with the age-adjusted mortality rates per 100,000 (white females -- 1968-1972). (ICDA = International Classification of Diseases and Accidents).

Cancer of the Esophagus
(ICDA = 150): US rate = 1.22

			Rate	Score
1.	San Francisco	CA	2.99	3.75
2.	Los Angeles	CA	1.49	3.57
3.	New York	NY	1.90	2.90
4.	Nassau	NY	1.79	2.80
5.	Cumberland	ME	2.88	2.20

Cancer of the Stomach
(ICDA = 151): US rate = 5.04

			Rate	Score
1.	Cook	IL	7.23	8.09
2.	Wayne	MI	7.06	5.07
3.	New York	NY	7.71	4.94
4.	Nassau	NY	7.27	4.63
5.	Hudson	NJ	8.84	4.56

Cancer of the Intestine
(ICDA = 152-3): US rate = 16.62

			Rate	Score
1.	Hamilton	OH	23.38	5.98
2.	Philadelphia	PA	21.46	5.76
3.	Nassau	NY	20.95	5.31
4.	Lake	IN	23.57	4.49
5.	Bergen	NJ	21.08	4.37

Cancer of the Rectum
(ICDA = 154): US rate = 3.84

			Rate	Score
1.	Hudson	NJ	7.63	4.91
2.	Cook	IL	4.83	4.49
3.	Bergen	NJ	5.99	3.96
4.	New York	NY	5.48	3.61
5.	Lake	OH	8.89	3.56

Cancer of the Lung
(ICDA = 160-3): US rate = 11.72

			Rate	Score
1.	Los Angeles	CA	14.58	8.85
2.	Harris	TX	16.73	6.88
3.	Nassau	NY	15.86	5.82
4.	Dade	FL	16.10	5.49
5.	Orange	CA	15.57	5.49

Cancer of the Breast
(ICDA = 174): US rate = 26.46

			Rate	Score
1.	Nassau	NY	38.62	10.97
2.	Cook	IL	31.58	9.06
3.	Westchester	NY	36.05	6.96
4.	New York	NY	34.09	6.72
5.	Milwaukee	WI	33.96	6.04

Leukemia
(ICDA = 204-7): US rate = 5.69

			Rate	Score
1.	Sacramento	CA	8.70	3.69
2.	Davis	UT	11.10	2.38
3.	Woodbury	IA	10.56	2.32
4.	Syandot	OH	19.87	2.26
5.	Wood	OH	10.84	2.24

Acute Ischemic Heart Disease
(ICDA = 410-1): US rate = 113.87

			Rate	Score
1.	Cook	IL	145.23	25.88
2.	Philadelphia	PA	154.10	17.84
3.	Erie	NY	157.79	16.99
4.	Wayne	MI	136.14	12.76
5.	Luzerne	PA	165.90	11.50

Influenza and Pneumonia
(ICDA = 470-86): US rate = 23.153

			Rate	Score
1.	Suffolk	MA	43.10	11.75
2.	New York	NY	36.94	11.67
3.	Nassau	NY	32.01	8.78
4.	Middlesex	MA	31.01	8.00
5.	Erie	NY	31.17	6.98

<u>Suicide</u>
(ICDA = E950-9): US rate = 6.76

			Rate	Score
1.	Los Angeles	CA	16.00	27.34
2.	San Francisco	CA	27.74	13.84
3.	Orange	CA	12.93	9.67
4.	San Deigo	CA	13.30	9.37
5.	Santa Clara	CA	13.87	9.16

Figure 1. First Four Principal Components
Summarizing the Influence of the
15 County 'Socio-Economic' Variables

	Principal Component I	II	III	IV
Pop 1970	.362	-.007	.011	-.001
Pop White Male	.367	-.007	.012	.000
Pop White Female	.367	-.007	.011	-.001
Pop Dens/Sq Mile	-.093	.015	-.024	-.002
Pct Urban	-.070	-.120	.076	.111
Pct Foreign	-.068	-.075	-.027	-.003
Pct Black	-.025	.082	.007	-.020
Pct Age 65+ Male	.014	.005	.517	.046
Pct Age 65+ Female	.012	-.003	.589	.047
Pct Employed	-.004	-.092	.058	1.105
Pct Professional	.042	-.144	.068	-.106
Med Educ Male	-.009	.570	.006	-.057
Med Educ Female	-.004	.663	-.007	-.073
Av Fam Income 1969	-.038	-.049	.065	.087
Pct Non-Res 1965	-.041	.106	-.008	-.125

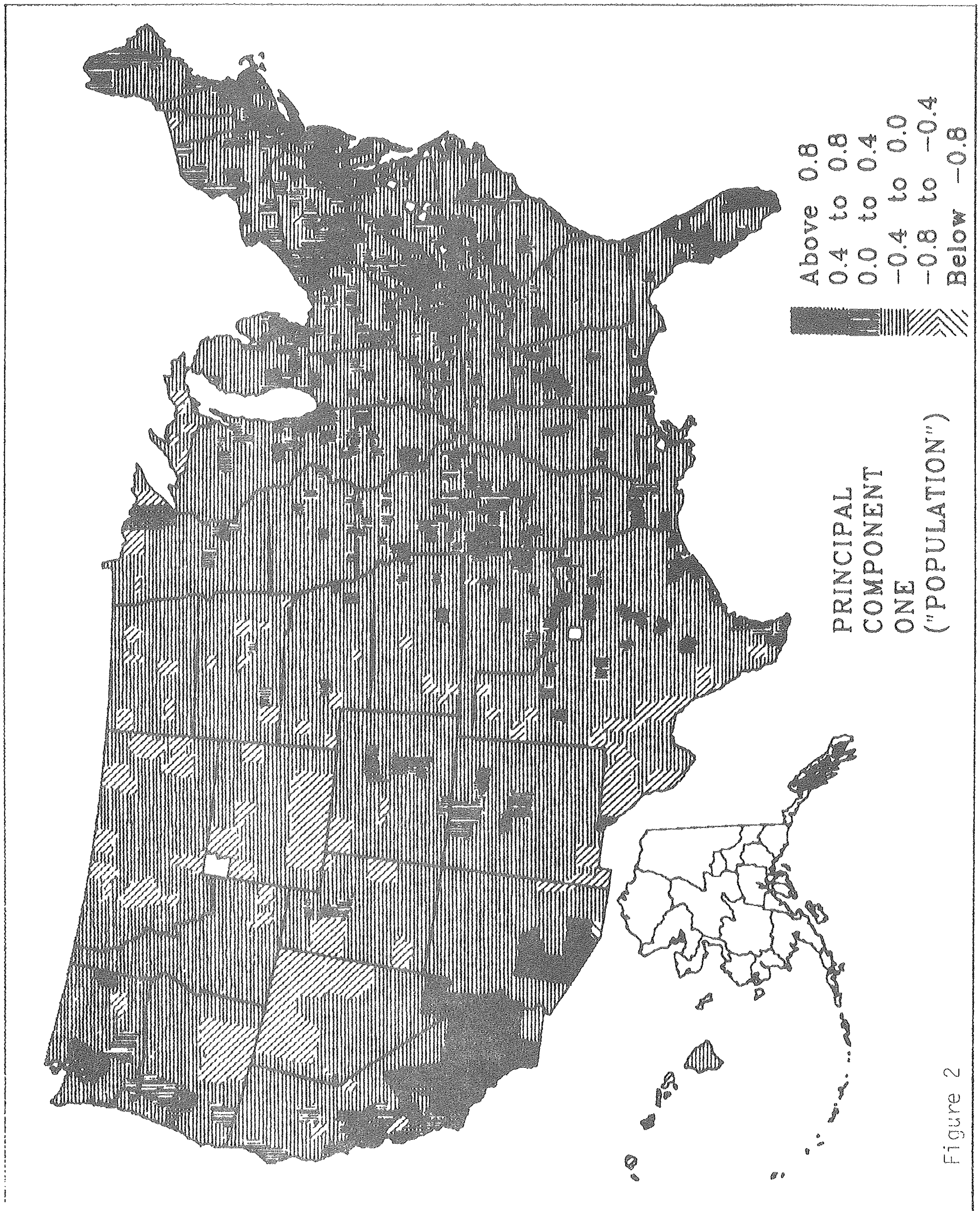

PRINCIPAL
COMPONENT
ONE
("POPULATION")
Above 0.8
0.4 to 0.8
0.0 to 0.4
−0.4 to 0.0
−0.8 to −0.4
Below −0.8

Figure 2

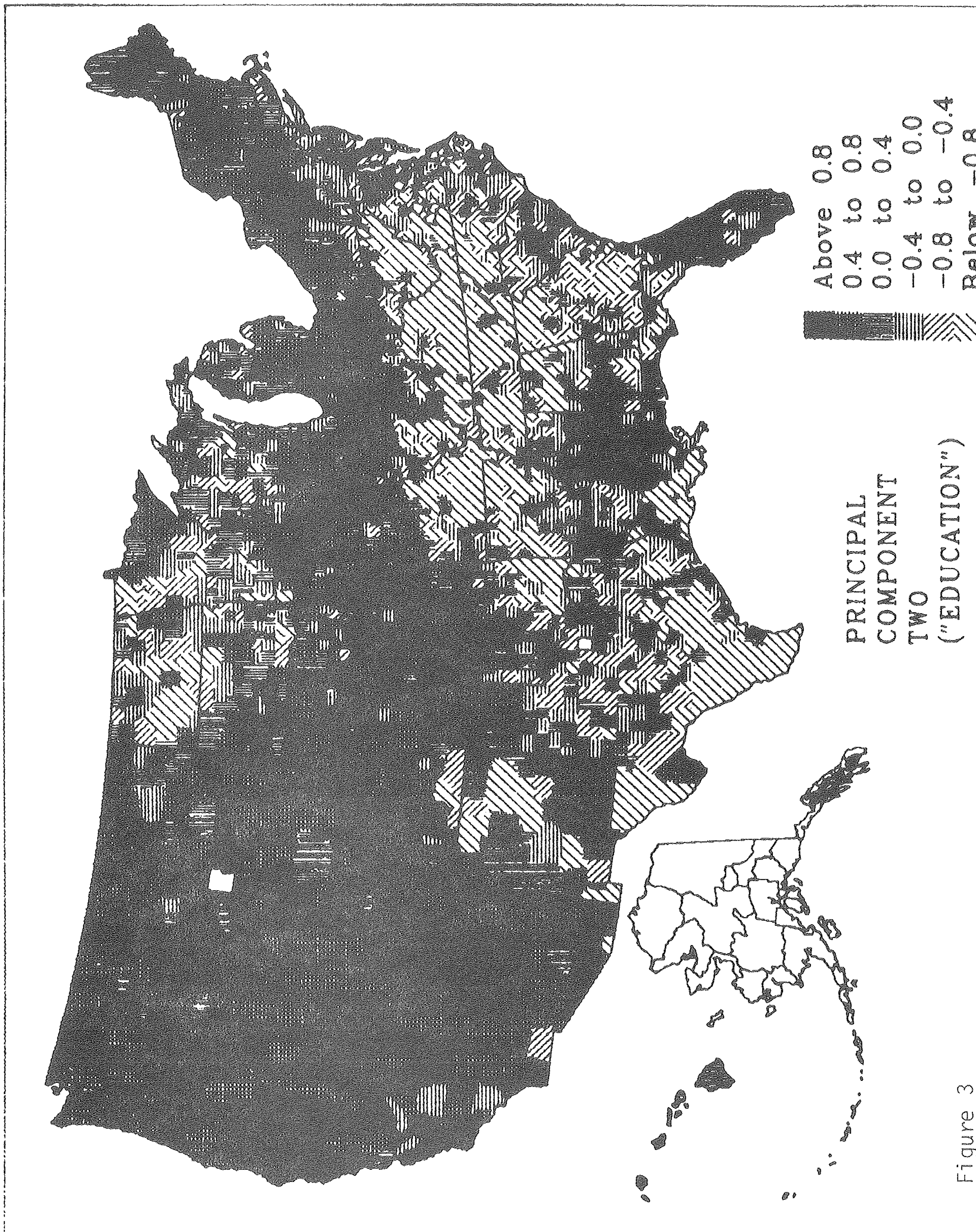

PRINCIPAL
COMPONENT
TWO
("EDUCATION")
Above 0.8
0.4 to 0.8
0.0 to 0.4
−0.4 to 0.0
−0.8 to −0.4
Below −0.8

Figure 3

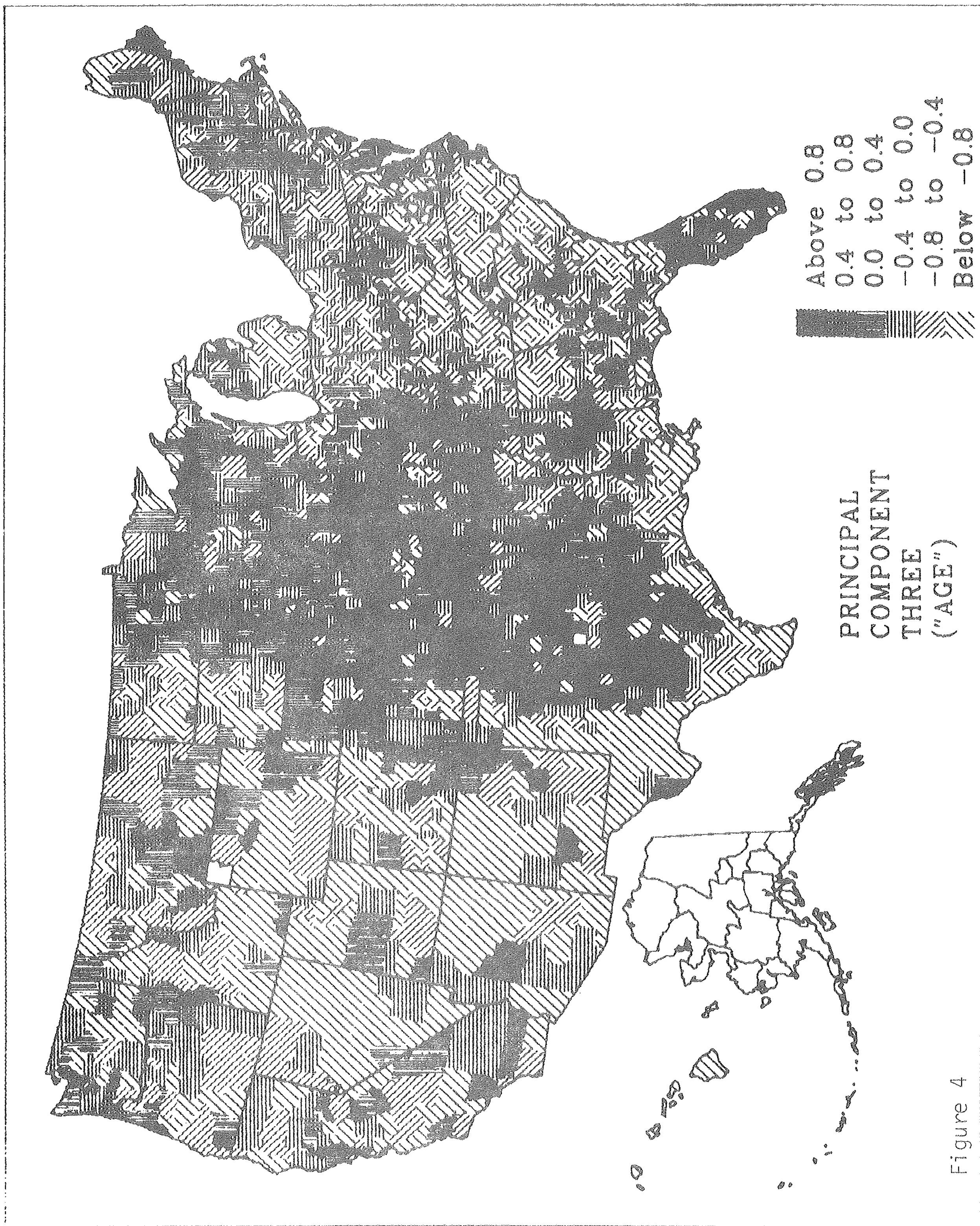

PRINCIPAL
COMPONENT
THREE
("AGE")
Above 0.8
0.4 to 0.8
0.0 to 0.4
−0.4 to 0.0
−0.8 to −0.4
Below −0.8

Figure 4

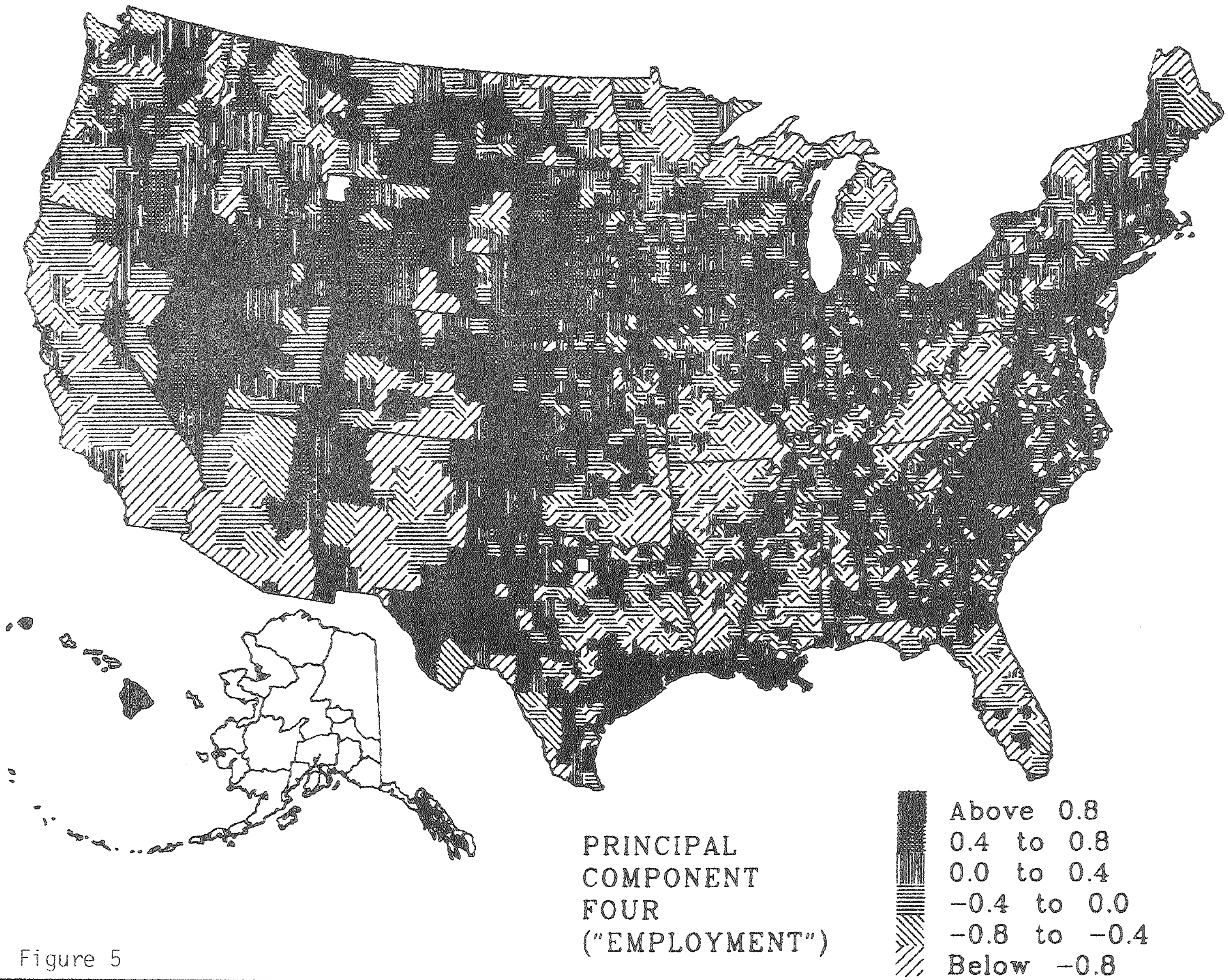

PRINCIPAL
COMPONENT
FOUR
("EMPLOYMENT")
Above 0.8
0.4 to 0.8
0.0 to 0.4
-0.4 to 0.0
-0.8 to -0.4
Below -0.8

Figure 5

Figure 6. Summary of Regression Analysis of Z-score (Dependent Variables) and Latitude and Longitude (Independent Variables) for the 10 Selected Causes of Death

	R Squared	Correlation With North Latitude	West Longitude
Males			
Esophageal Cancer	.1	-.02	-.04
Stomach Cancer	2.1	.14	.08
Intestinal Cancer	4.3	.14	-.08
Rectum Cancer	5.4	.19	-.06
Lung Cancer	9.1	-.29	-.19
Breast Cancer	1.0	.04	-.07
Leukemia	.4	.03	.06
Ischemic Heart Disease	10.1	-.01	-.29
Influenza and Pneumonia	1.5	-.11	-.09
Suicide	1.3	-.05	.07
Females			
Esophageal Cancer	.1	-.03	-.02
Stomach Cancer	1.2	.10	.05
Intestinal Cancer	8.7	.24	-.12
Rectum Cancer	4.2	.18	.08
Lung Cancer	1.9	-.10	-.08
Breast Cancer	9.3	.28	.08
Leukemia	.5	.00	.07
Ischemic Heart Disease	9.9	.01	-.31
Influenza and Pneumonia	.7	-.07	-.05
Suicide	5.5	-.07	.21

Figure 7.
Correlation Coefficients
White Males

	STOM	INTES	RECT	LUNG	BRST	LEUK	HEART	FLU	SUIC
ESOPH	.23	.24	.32	.17	-.05	.02	-.03	.09	-.11
STOM		.21	.31	-.04	-.04	.05	-.04	.04	-.17
INTES			.33	.11	.01	.06	.01	-.02	-.21
RECT				.05	.01	.03	.04	.05	-.21
LUNG					-.01	.01	.10	.13	.02
BRST						-.03	.08	-.01	-.03
LEUK							.03	.04	.02
HEART								.11	-.07
FLU									-.03

Figure 8.
Correlation Coefficients
White Females

	STOM	INTES	RECT	LUNG	BRST	LEUK	HEART	FLU	SUIC
ESOPH	.14	.01	.08	.11	.08	.01	.00	.06	.07
STOM		.12	.25	.08	.14	.02	.10	.10	.04
INTES			.29	.08	.31	.02	.13	.04	.10
RECT				.09	.26	.03	.19	.03	-.06
LUNG					.11	.04	-.06	.04	.22
BRST						.00	.02	.03	.03
LEUK							.01	-.01	.07
HEART								.13	-.26
FLU									-.09

Figure 9. Squared Multiple Correlation Coefficients for Multiple Regression Analysis of Z-score versus the 15 "Socio-Economic" Variables

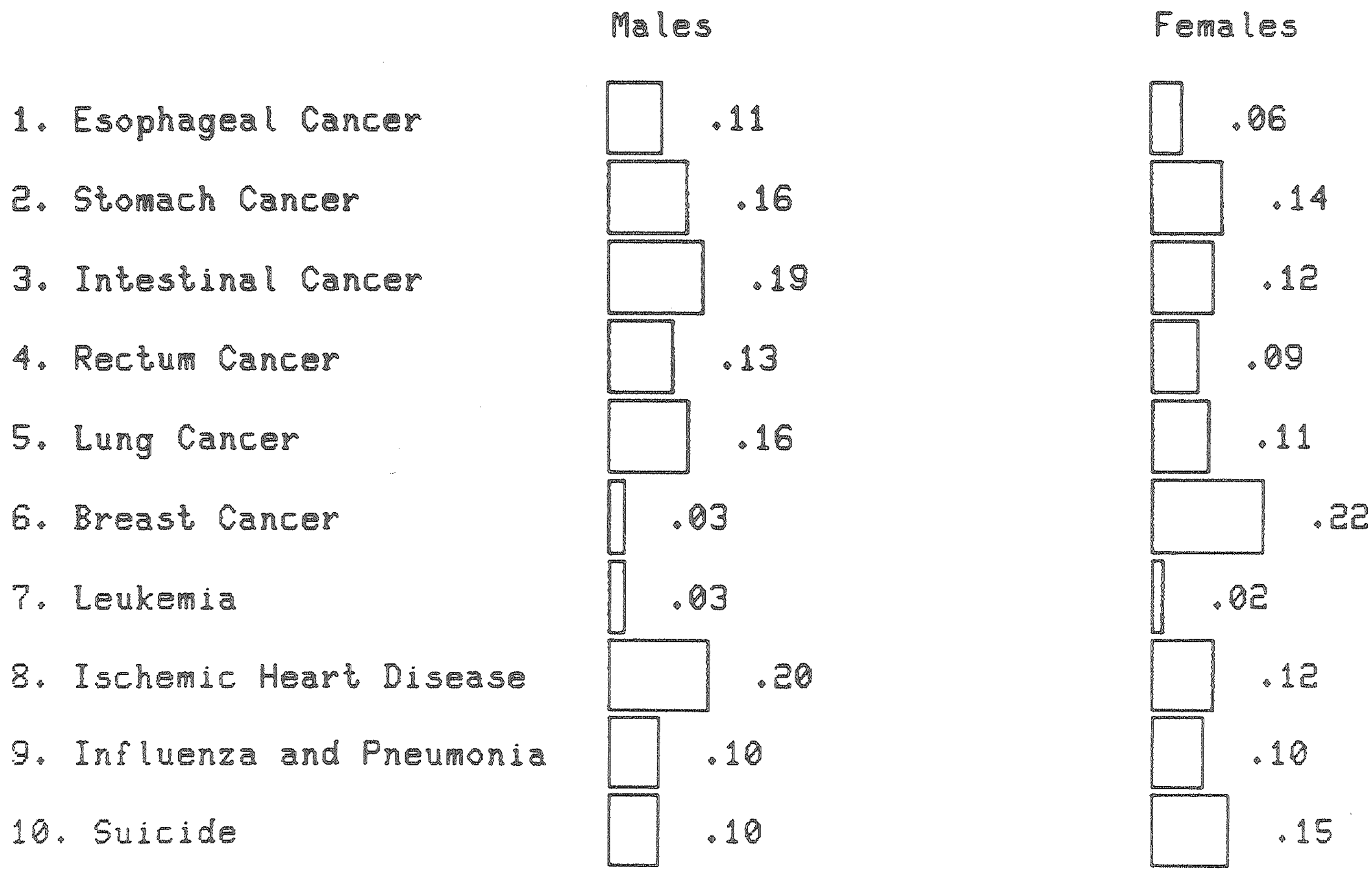


Figure 10.
Adjusted Correlation Coefficients
White Males

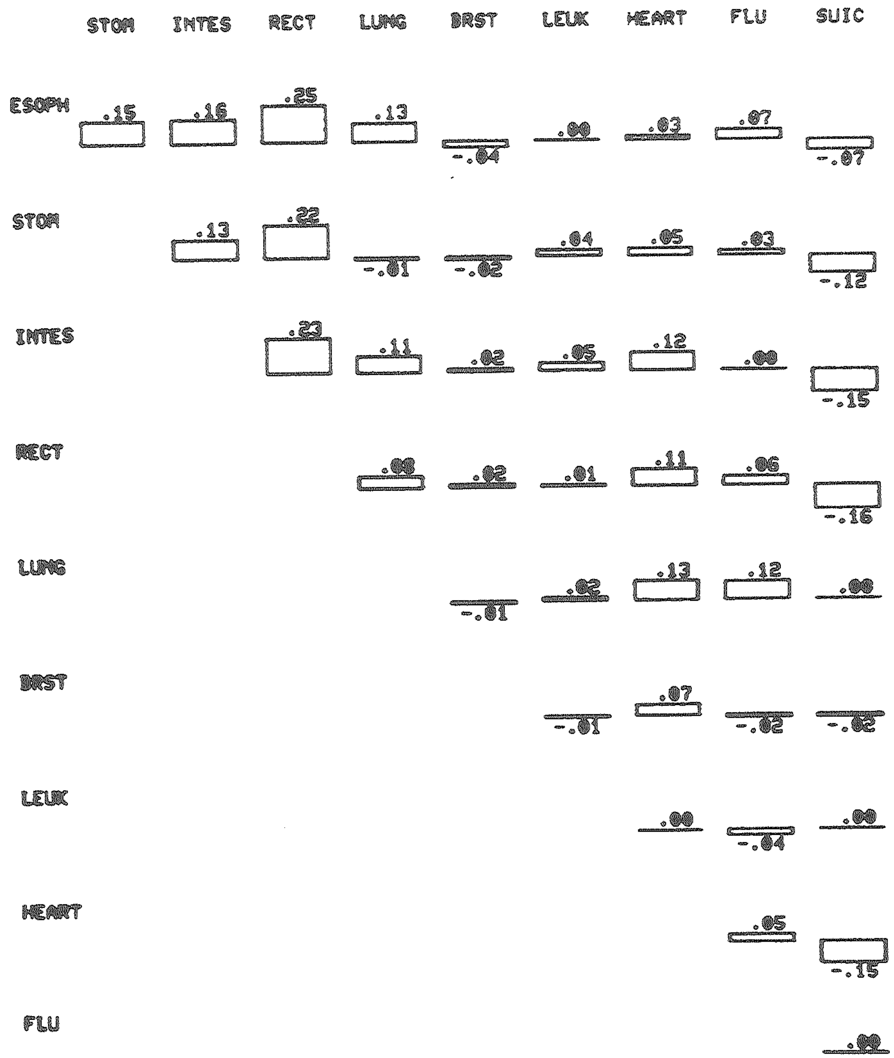


Figure 11.
Adjusted Correlation Coefficients
White Females