Title
DENSITY-EQUALIZING MAP TRANSFORMATIONS IN ENVIRONMENTAL HEALTH APPLICATIONS

Author
Merrill, D.W.

Publication Date
1987-05-01

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA, BERKELEY

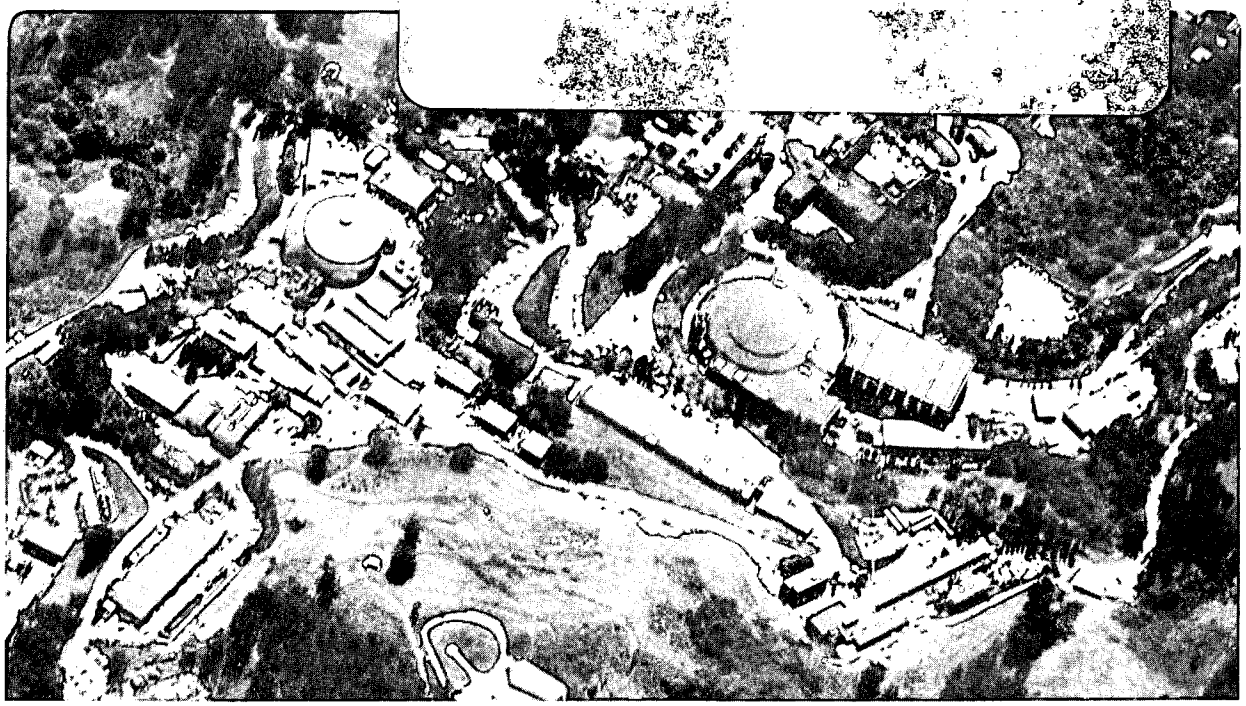
Information and Computing Sciences Division

Presented at "Piecing the Puzzle Together:
A Conference on Integrating Data for Decisionmaking,"
Washington, DC, May 27–29, 1987

DENSITY-EQUALIZING MAP TRANSFORMATIONS IN ENVIRONMENTAL HEALTH APPLICATIONS

D.W. Merrill, Jr.

May 1987



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

DISCLAIMER

LBL-23629

Density-Equalizing Map Transformations in Environmental Health Applications

Deane W. Merrill, Jr.

Computer Science Research Department
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

May, 1987

This research was supported by the Office of Health and Environmental Research, U.S. Department of Energy under contract DE-AC03-76SF00098.

DENSITY-EQUALIZING MAP TRANSFORMATIONS IN ENVIRONMENTAL HEALTH APPLICATIONS

Deane W. Merrill, Jr.
Lawrence Berkeley Laboratory, Berkeley, California 94720

ABSTRACT. In a density-equalized map, or cartogram, an ordinary geopolitical map is transformed to make the areas of individual subareas proportional to population. As a display tool, such a map is visually meaningful for population-based data, for example death rates. More important, using density-equalized projections one can analyze geographic distributions of events which are so infrequent that reliable rates cannot be calculated for individual subareas. Statistically significant non-uniformity of cases on the transformed map implies significant non-uniformity of rates on the original map. A computer algorithm for density-equalizing projections has been developed by the author and used in various environmental health applications.

KEYWORDS

computer mapping; cartogram; density-equalizing projection; disease clusters; population density; statistical analysis; spatial analysis; statistical geography; computer algorithm.

BACKGROUND

THE PROBLEM

A useful tool for the analysis of geographic data is the choropleth map, in which areas are shaded to represent different levels of some variable. For example, atlases of cancer mortality rates (Mason et al. 1975, Riggan et al. 1983) have been widely circulated and have motivated detailed studies leading to new epidemiological knowledge. A common criticism of such maps is that the maps are visually dominated by large sparsely populated counties, and many populous counties such as the boroughs of New York City are practically invisible.

Another difficulty is that rates for small-population counties are unstable; there is no convenient way to represent visually both the magnitude and the statistical significance of a rate in a particular county. Combining counties, for example into state economic areas, provides more stable rates, but the choice of county groups is arbitrary and of course leads to loss of geographic detail.

An alternate display of geographic data is the dot map, in which the position of each event (for example each cancer death) is represented by a dot. If not too numerous, the exact location of each death can be visually conveyed with no loss of geographic detail, and the number of dots in a given area provides a measure of statistical significance. For example, Figure 1 is a map showing the number of smallpox deaths in each California county during the period 1915-1924 (Gillihan 1927). (In these data the exact locations of deaths are unknown; dots are randomly plotted within their respective county boundaries.) As pointed out by Gillihan, the dots are unduly grouped around the several population centers of the state, so that one cannot visually compare rates.

Figure 1. Smallpox in California, original map

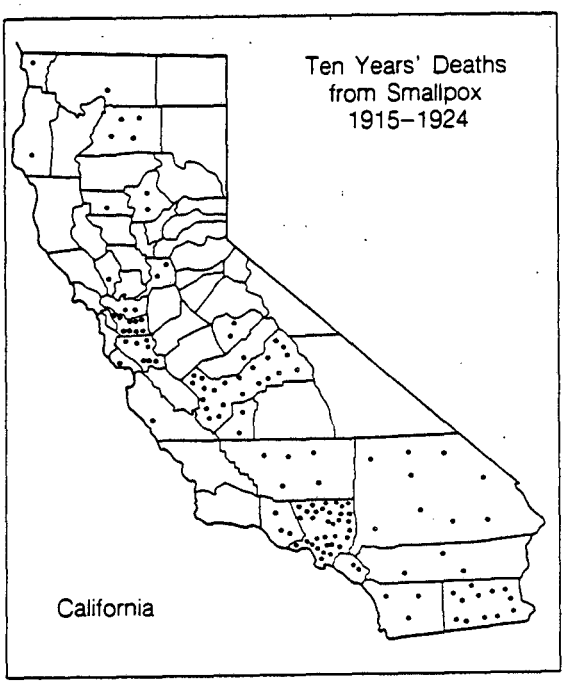


Figure 1. Smallpox deaths in California, 1915-1924, displayed on a conventional geopolitical map. The apparent clusters observed in several counties are merely the result of non-uniform population distribution. Source: adapted from Gillihan (1927).

DENSITY-EQUALIZING MAP PROJECTIONS

To the best of our knowledge, the earliest description of a population-based map is that of Karsten (1923). Figure 2 is a population-based map, produced by Gillihan (1927) by a novel method: he assembled lumps of plastocine whose weights were proportional to the county populations, then rolled them to a uniform thickness. In Figure 2, unlike Figure 1, the population density is equal over the entire map; the remaining differences in dot density reflect actual differences in the smallpox rates.

Figure 2. California, density-equalized map

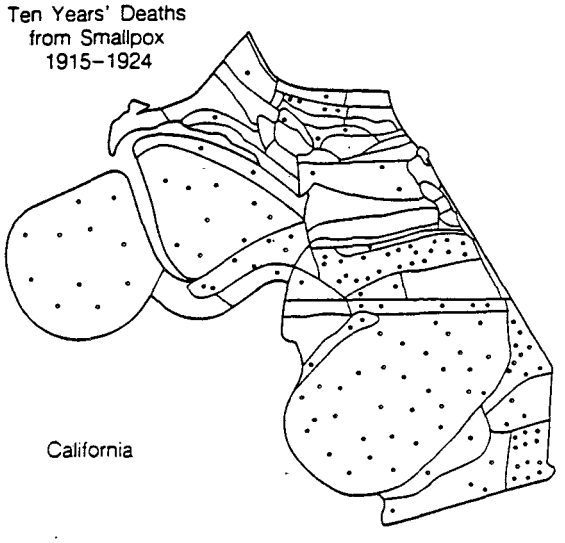


Figure 2. Smallpox deaths in California, 1915-1924, displayed on a density-equalized map. The clustering seen in Figure 1 as a result of non-uniform population distribution has been removed. Source: adapted from Gillihan (1927).

Since the 1920's, other authors have constructed cartograms by a variety of manual methods, and used them in public health applications. Tobler laid down a general framework for geographic projections (1961) and wrote a density-equalizing algorithm (1973). Recently, Lawrence Berkeley Laboratory (LBL) has implemented a different algorithm and applied it to specific epidemiological investigations (Schulman 1986, Schulman et al. 1987, Selvin et al. 1987a, Selvin et al. 1987b, Selvin et al. 1987c, Shaw 1986, Shaw et al. 1987).

COMPUTER ALGORITHM

A simplified illustration of the LBL algorithm appears in Figure 3. On the left side, the circle A and the ring-shaped region B have equal populations but different areas (A=10, B=20). On the right side, we have changed the radii so that A' and B' have equal areas (A'=20, B'=20). Note that we have changed the area but not the shape of A, and we have changed the shape but not the area of B. (B' is a "skinnier" ring than B.) The transformed radii of A' and B' are easily calculated.

Figure 3. Density-equalizing algorithm: a simple case

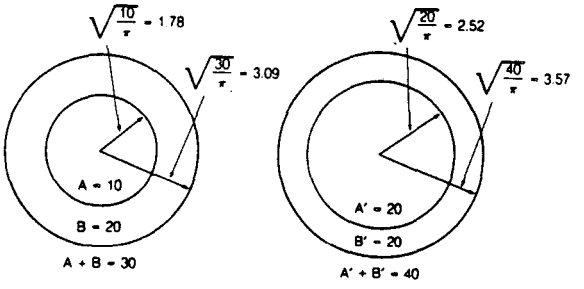


Figure 3. Illustration of the LBL density-equalizing algorithm, for a simple case. On the left side, the circle A and the ring-shaped region B have equal populations but different areas (A=10, B=20). On the right side, A' and B' have equal areas (A'=20, B'=20).

In the LBL algorithm we start with figures A,B,C... and choose an arbitrary expansion center for each. We calculate the area magnification to be applied respectively to A,B,C... Next we choose one figure A and apply the correct magnification without changing the areas of B,C,... For arbitrarily shaped figures, the algorithm is essentially the same as in Figure 3. The radii of A,B,C... and A',B',C'..., were independent of angle in the case of circles but now depend on the angle relative to A's expansion center. The magnification of A changes the area but not the shape of A, and changes the shapes but not the areas of B,C,... Next, in turn, we magnify B without changing the areas of A,C,...; then C without changing A,B,...; and so on.

As pointed out by Tobler (1961), there are infinitely many transformations that produce the correct areas A',B',C'... In the LBL algorithm, the choice of expansion centers for A,B,C..., and also the order of processing A,B,C..., are arbitrary. We observe minimum distortion if the expansion center of each figure is near its area centroid, and if figures requiring the greatest magnification (or demagnification) are processed first.

The LBL algorithm is non-iterative and faster than Tobler's interative algorithm. For figures represented as polygons, its accuracy is limited by the number of points in the map, because in general the algorithm transforms straight lines into curved lines. Unfortunately, computation time increases as the square of the number of points in the map. Implementing an efficient and accurate algorithm for small computers is a challenging problem in numerical computation, which requires still further work (Selvin et al. 1987c).

In population-based cartograms, including hand-drawn ones, transformed areas are proportional, by definition, to their respective populations. The LBL algorithm has a further important property: magnification is everywhere constant within a given subarea (Schulman 1986).

It follows that if population density is constant within each subarea of the original map, then population density is constant everywhere on the transformed map. Therefore the transformation cannot produce statistically significant false clustering, any more than "rebinning" a histogram can produce a statistically significant "peak".

We now have a powerful statistical tool: we can transform events such as cancer cases along with the subarea boundaries, then remove the subarea boundaries, and perform statistical tests to look for non-uniformity or clustering of the transformed cases. A significant non-uniformity of cases in the transformed map implies a non-uniformity of rates in the original map. Of course, interpretation and demonstration of causality may still pose difficult problems.

EXAMPLE: SAN FRANCISCO

In Figures 4 and 5 we illustrate the LBL density-equalizing algorithm as applied to the census tracts of San Francisco. Figure 4 is an ordinary geopolitical map, and Figure 5 is a transformed map in which the area of each census tract is proportional to the white male population 35 to 54 years of age. To show the correspondence between Figures 4 and 5, a few selected tracts have been shaded on both maps.

We see that Golden Gate Park (the long rectangular tract near left center in Figure 4) disappears in Figure 5 because it has no resident population. At the same time, the five shaded tracts in the upper right quadrant of Figure 4 become more important in Figure 5, because they have relatively large proportions of white males 35 to 54 years of age.

Figure 4. San Francisco, geopolitical

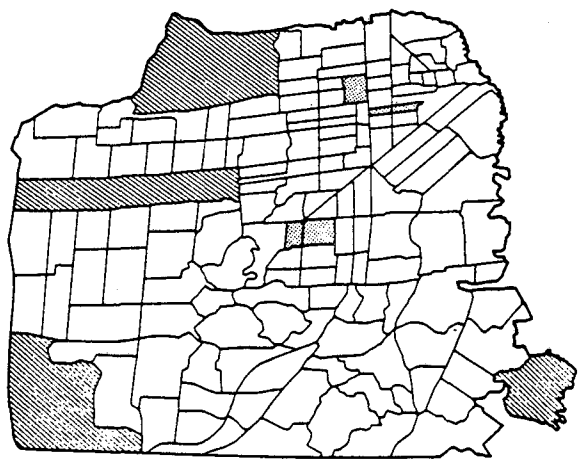


Figure 4. Geopolitical map of census tracts in San Francisco. The long rectangular tract near top left is Golden Gate Park. The five shaded tracts in the upper right quadrant have high proportions of white males 35 to 54 years of age.

Figure 5. San Francisco, density-equalized

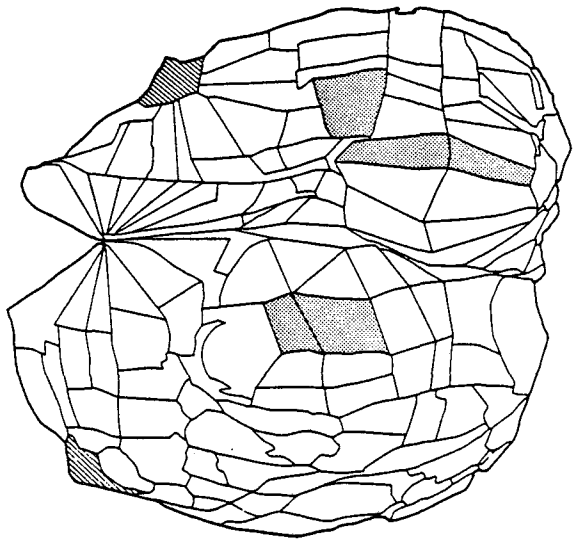


Figure 5. Density-equalized map of census tracts in San Francisco. The area of each census tract is proportional to the white male population 35 to 54 years of age.

STATISTICAL ANALYSIS

In Figure 6 we show the distribution, on density-equalized maps, of four types of cancer cases in San Francisco, from the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. For details, see Selvin et al. (1987c). Under the "null hypothesis", i.e. the absence of any geographic differences in cancer rates, one expects a random distribution of cases in each map. One can construct various test functions or metrics, for example (a) the position of the case centroid, or (b) the radius of the median case "distance" from the centroid, or (c) the "distance" from a suspected environmental hazard. Then for a given boundary (which depends on the particular transformation performed), one can calculate the expected distribution of the test metric under the null hypothesis. This leads finally to a "p-value", or the probability that the observed value of the test metric could have resulted from chance variation (Schulman 1986).

CONCLUSION

Density-equalizing map projections provide an unbiased and sensitive means of measuring geographic variation in disease rates. The method works well even when the number of cases is very small, and is well suited to the analysis of routinely collected surveillance data. Complete geographic detail is preserved, both of case numerators and of population denominators. For a given test metric, one can calculate a "p-value"; namely the probability that the observed value could have resulted from

chance variation. Statistically significant non-uniformity of cases on the transformed map implies significant non-uniformity of rates on the original map.

Figure 6. Cancer cases in San Francisco

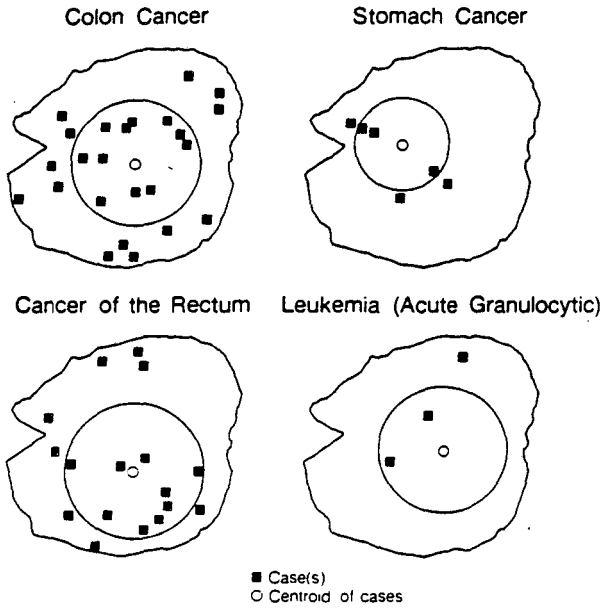


Figure 6. Sample density-equalized maps of cancer cases in San Francisco. In each map, density of the population at risk is uniform. One expects a priori a random distribution of cases (black squares). The small circle is the case "centroid"; the large circle indicates the median "distance" of cases from the centroid. Data Source: Surveillance, Epidemiology, and End Results (SEER) Program, National Cancer Institute.

ACKNOWLEDGMENTS

This research was supported by the Office of Health and Environmental Research, U.S. Department of Energy under contract number DE-AC03-76SF00098.

REFERENCES

Gillihan, A. F. 1927. Population maps. Amer. Jour. Pub. Health 17:316-319.

Karsten, K. G. 1923. Charts and Graphs, Chapter LII.

Mason, T. J., F. W. McKay, R. Hoover, W. J. Blot, and J. F. Fraumeni, Jr. 1975. Atlas of Cancer Mortality for U.S. Counties: 1950-1969. U.S. Dept. of Health, Education and Welfare (DHEW) Publication No. (NIH) 75-780.

Riggan, W. B., J. Van Bruggen, J. F. Acquavella, J. Beaubier, and T. J. Mason. 1983. U.S. Cancer Mortality Rates and Trends, 1950-1979, Volume IV. U.S. Env. Prot. Agency Publication No. EPA-600/1-83-015a.

Schulman, J. 1986. The statistical analysis of density-equalized map projections. Ph.D. Thesis. University of California, Berkeley.

Schulman, J., S. Selvin, and D. W. Merrill. 1987. Density equalized map projections: a method for analyzing clustering around a fixed point. Statistics and Medicine (in press).

Selvin, S., D. Merrill, J. Schulman, G. Shaw, W. Benson, and M. Mohr. 1987. Illustrations of a density-equalizing map projection technique. Lawrence Berkeley Laboratory Report LBL-23189.

Selvin, S., G. Shaw, J. Schulman, and D. W. Merrill. 1987. Spatial distribution of disease: three case studies. J. Nat. Cancer Inst. (in press).

Selvin, S., D. Merrill, J. Schulman, S. Sacks, L. Bedell, and L. Wong. 1987. Transformations of maps to investigate clusters of disease. Soc. Sci. Med. (submitted).

Shaw, G. 1986. A comparison of techniques used for the analysis of spatial and temporal-spatial disease clustering. Ph.D. thesis. University of California, Berkeley.

Shaw, G., S. Selvin, S. Swan, D. Merrill, and J. Schulman. 1987. An empirical comparison of three disease clustering methodologies. International Journal of Epidemiology (submitted).

Tobler, W. R. 1961. Map transformations of geographic space. Ph.D. thesis. University of Washington.

Tobler, W. R. 1973. Cartogram Programs. Ann Arbor, Cartographic Laboratory Report.

For further information, please contact:

Deane W. Merrill, Jr.
Building 50B, Room 3238
Computer Science Research Department
Lawrence Berkeley Laboratory
Berkeley CA 94720
Telephone: 415-486-5063
Electronic mail:
ARPANET: DWMerrill@lbl.arpa
BITNET: MERRILL@LBL
DIALCOM: 64:DOE1209
ASYNC: 329-1020

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720