The Population Center of Canada—Just North of Toronto?!?

Mark P. Kumler
Michael F. Goodchild
University of California, Santa Barbara

The concept of a center of population is often used in geographical studies. There are, however, several ways to determine a "center point" and at least as many ways to describe them. For very large areas the problem is complicated by the spherical shape of the earth. In this investigation three such centers are defined and determined for the population of Canada: the median center, the mean center (also known as the centroid, or center of gravity), and the point of minimum aggregate travel.

Background

In the first known publication of the center of a population (Hilgard 1872), the medial lines and the centers of gravity were determined for the population of the United States for the previous four censuses. Hilgard presented the medial lines and centers of gravity, and speculated that the ultimate population center of the United States would be "not far from St. Louis" (where, remarkably, it lies today—about 60 kilometers southwest of St. Louis). His medial lines appear to have been drawn parallel to the eastern coast; they divided the population in two—half living east of the line and half living west. Hilgard's centers of gravity were determined by "calculating the positions [for] the center of each state . . . with regard to the relative density of population in their different parts, and that all cities having over 50,000 inhabitants have been treated as separate centers . . ." (Hilgard 1872, 218). He ignored the spherical shape of the earth.

The popularity of center points flourished in the 1920s and 1930s. In the United States, the Bureau of the Census (1923) determined and published the movement of the center of population for every census since 1790, and the movements of other agricultural and manufacturing centers since 1850. In Canada, H. E. M. Kensit published the movements of the centers of population, industry, and developed water, and predicted that in a few years the center of population would be, "paradoxically, in the middle of Lake Superior" (Kensit 1934, 269).

The study of center points was most fervent in the Soviet Union, where a group of geographers in Leningrad established the Mendeleev Centrographical Laboratory in 1925 (Neft 1966; Porter 1963; Sviatlovsky and Eells 1937). Named after the famous Russian chemist D. I. Mendeleev, who had discussed the concept of the center of Russia in one of his works in economic geography, the lab sought to establish centrography as a science by developing "laws of the distribution of phenomena based on the relationships and migrations of their 'centers of gravity'" (Poulsen 1959, 326). The lab met its demise in 1934, after recommending that commercial grain planting be limited to the traditional bread belt of European Russia; the recommendation clashed with a decision recently made by the Communist Party, and the laboratory was discredited and liquidated (Poulsen 1959).

It was also around this time that confusion first arose between two of the points considered in this study: the mean center and the point of minimum aggregate travel. In an attempt to popularize an understanding of the mean center of population, the U.S. Bureau of the Census added the following phrase to its otherwise correct description of the point: "If all the people in the United States were to be assembled at one place, the center of population would be the point which they could reach with minimum aggregate travel, assuming that they all traveled in direct lines from their residences to the meeting place" (Bureau of the Census 1923, 5; or see Eells 1930, for an exhaustive review of the confusion). The point just described is the point of minimum aggregate travel; this is *not* the same as the mean center.

Definitions

There have been many attempts to establish conventions for the terms and definitions of center points. Until such definitions are widely accepted it is necessary to carefully define what one means by the terms *median center*, *mean center*, and *point of minimum aggregate travel*.

Median Center

The median center is defined here to be the point that divides the set of points into two equal halves. In a one-dimensional distribution there will be an equal number of points on either side of the median point. In two dimensions the median center is the intersection of two perpendicular lines, each of which divides the distribution into two halves. The orientation of these lines is often arbitrary, and different orientations can yield different median points (Hayford 1902). On a sphere the lines become circles, either great or small. One could consider the intersection of two perpendicular great circles, in which case the solution is sensitive to the orientation of the circles, or one could use the established graticule of latitude (small circles) and longitude (great circles). The latitude-longitude median center is used here; it divides the population such that half lives east of the point and half lives west, while half lives north of the point and half lives south.

Mean Center

The mean center, also known as the centroid or the center of gravity, is the arithmetic mean, or average, of the locations of the points. It is the point on which the distribution would balance if represented by weighted points on a weightless line, plane, or sphere. Mathematically, the mean center is the point that minimizes the sum of the squared distances between it and all other points.

It should be noted that the mean center of points on a sphere is internal to the sphere. While this point can be easily projected back onto the surface of the sphere, for the purpose of referencing the location of the point, the true mean center lies beneath the surface.

Point of Minimum Aggregate Travel

The point of minimum aggregate travel (or point of mat, or mat point) is the point that minimizes the sum of the distances (unsquared) between it and all other points. For a population, the mat point is the point that would result in the least total distance traveled if all people were to travel along straight lines to a single location. For spherical problems, distances are measured along great circle routes.

Several researchers have taken to calling this point the median (Gini and Galvani 1929; Neft 1966; Porter 1964; Scates 1933), on the basis that the median center described above is sensitive to the orientation of the axes, and that in one-dimensional problems the median center is also the point of minimum aggregate travel. While this argument has many merits, the more descriptive and unambiguous term—point of minimum aggregate travel—is used here, to avoid further confusion.

Data

The data for this study were provided by the Geocartographics Division of Statistics Canada. The population counts and approximate centroid locations for all census enumeration areas were extracted from the 1976, 1981, and 1986 Geography Tape Files The centroid locations were identified in degrees and minutes of latitude and longitude, and were converted to decimal degrees or radians for input to the computer programs.

Algorithms

Median Center

The simplest way to determine the median of a two-dimensional population distribution is to order the points by their coordinates in the x- and y-directions. The median in each dimension is then easily found by stepping through the ordered points and maintaining an accumulated population total. As soon as the accumulated total is greater than or equal to one half of the total population, the median in that dimension has been found. The process is repeated for the other dimension, and the intersection of the two medial lines is the median center. If the locations are specified in latitude and longitude, this algorithm generates the lat-long median center—the point north of which lives one half of the population and east of which lives half of the population.

If the number of observations is extremely large, and sorting or storing them is impossible or difficult, shortcuts can be taken. One could divide the range of coordinates in each dimension into a large number of small sub-ranges, or buckets, and then accumulate population counts only for each bucket. The bucket that contained the median value would be easily determined, and then the observations could be scanned again, storing and sorting only those that fell within the median bucket. This procedure requires two passes through the data values.

A different shortcut, used in this investigation, relies on the knowledge of the approximate location of the median center (perhaps from the previous census). It is essentially a "three-bucket" approach, in which a very thin bucket collects the observations near the median, and the populations at all other observations are simply accumulated into large buckets on either side. The locations and populations of the observations falling in the thin center bucket can be stored and sorted, and the median point easily determined.

Mean Center

Being the simple average location of the observations, the mean center is quite easy to compute. In one- or two-dimensional problems the coordinates (perhaps weighted by population) are simply summed and divided by the number of observations (or people). If the points are located on a sphere, and the locations are given in latitude and longitude, the locations must first be converted into their equivalent three-dimensional coordinates. This is easily done with the formulae:

$$x = \cos(\text{lat}) \cdot \cos(\text{lon})$$

$$y = \sin(\text{lat})$$

$$z = \cos(\text{lat}) \cdot \sin(\text{lon}).$$

The averages in the x-, y-, and z-dimensions are easily determined, and can then be converted back to latitude and longitude.

Point of Minimum Aggregate Travel

There are no known algebraic solutions to finding the point of minimum aggregate travel; the best solutions to date are iterative, and converge, often slowly, upon the final value. E. Weiszfeld (1937) developed the first heuristic for finding the mat point; many years later it was re-discovered independently by W. Meihle (1958), H. Kuhn and R. Kuenne (1962), and L. Cooper (1963). For an exhaustive review of this measure, including some interesting graphical and analog solutions, see Brian R. Rizzo (1982).

The Weiszfeld procedure begins with the selection of an arbitrary point as the initial seed location. The distances and directions from that seed to all other (possibly weighted) points are determined, and the trial point is displaced in the indicated direction. The procedure is then repeated until it stabilizes on a value, or some displacement threshold is reached. The procedure is not entirely robust; it will fail if the trial point ever coincides with one of the weighted points. This situation, known as the vertex iterate problem, results from an undefined division by zero. Although this problem did not arise in this investigation, perhaps because of the many digits of spurious precision that are automatically carried by most computers, it can usually be circumvented by either "nudging" the trial point away from the problem point, or, more simply, by re-starting the procedure with a different seed location. The algorithm is presented in numerical form in Figure 1.

Results

The mean and median center and the point of minimum aggregate travel were computed for the population of Canada for each of the past three censuses—1976, 1981, and 1986.

do

$$\hat{x} = \frac{\sum \dfrac{x_i}{[(\hat{x}-x_i)^2+(\hat{y}-y_i)^2]^{1/2}} \cdot pop_i}{\sum \dfrac{pop_i}{[(\hat{x}-x_i)^2+(\hat{y}-y_i)^2]^{1/2}}}$$

$$\hat{y} = \frac{\sum \dfrac{y_i}{[(\hat{x}-x_i)^2+(\hat{y}-y_i)^2]^{1/2}} \, pop_i}{\sum \dfrac{pop_i}{[(\hat{x}-x_i)^2+(\hat{y}-y_i)^2]^{1/2}}}$$

loop until stable

Figure 1. The Weiszfeld procedure.

Year	Median center	Mean center	Point of mat
1976	45°37′N 79°24′W	48°15′54″N 84°21′25″W	43°49′36″N 79°23′29″W
1981	45°41′N 79°26′W	48°25′43″N 85°00′30″W	43°50′31″N 79°25′46″W
1986	45°39′N 79°26′W	48°25′09″N 85°10′32″W	43°50′35″N 79°26′23″W

The median center is near Burk's Falls, Ontario, just east of the Georgian Bay. The point moved north and west between 1976 and 1981, but only south between 1981 and 1986; it appears to be fairly well anchored by the two largest cities in Canada, at the approximate latitude of Montreal and the approximate longitude of Toronto (Figure 2). The mean center falls north of Lake Superior, approximately 55 kilometers northwest of Wawa, Ontario and 20 kilometers southeast of White River. The true mean center was 156.0 kilometers beneath the surface in 1976, 161.7 kilometers in 1981, and 162.3 kilometers in 1986. The point moved westward between the three censuses, but at a decreasing rate; it crossed the TransCanada Highway sometime in 1984. The point of minimum aggregate travel lies north of Lake Ontario, about 25 kilometers north of Toronto, and 5 kilometers south-southwest of Richmond Hill. It has also continued to move westward, at a decreasing rate; between 1981 and 1986 it moved just over two meters per day.

Different Center Points (Pros and Cons)

All of the points defined above have properties that make them useful for describing some aspect of the distribution of a population. They also have their disadvantages.

The median center is easy to determine and is conceptually simple. It does, however, have two important flaws: it is sensitive to the orientation of the axes—different axes yield different points—and it is insensitive to distances from the median point. Large movements of population can occur within any quadrant without affecting the location of the median center, while the movement of a single person from one quadrant to an adjacent quadrant will change the location of the point.

The mean center is also easy to determine, and the weighted-points-on-a-weightless-plane concept is easy to understand, if somewhat abstract. It is not sensitive to the orientation of the axes and it is sensitive to distances. The mathematical property, however, that it minimizes the sum of the squared distances, is the source of the problem with this measure: the points, or people, are effectively weighted proportionally to their distance from the center—more distant people have greater influence on the location of the mean center than people nearby. This property is clearly unsatisfactory for describing the distribution of a population.

Conceptually, the point of minimum aggregate travel is the most appealing. The property of minimizing total distance is easy to understand and appreciate. Historically, the location of the point has been difficult to determine, but with modern computing power it is certainly feasible. Each iteration of the Weiszfeld procedure required approximately 6 minutes on an IBM PS/2, and the procedure always stabilized to one ten-thousandth of a degree (one third of a second of latitude or longitude) within 15 iterations. The mat point does have one flaw—it is insensitive to radial movement: If a person moves 1,000 kilometers directly toward or away from the mat point, the point will not move; if that same person, however, moves only a few kilometers in any other direction, the point will move accordingly.

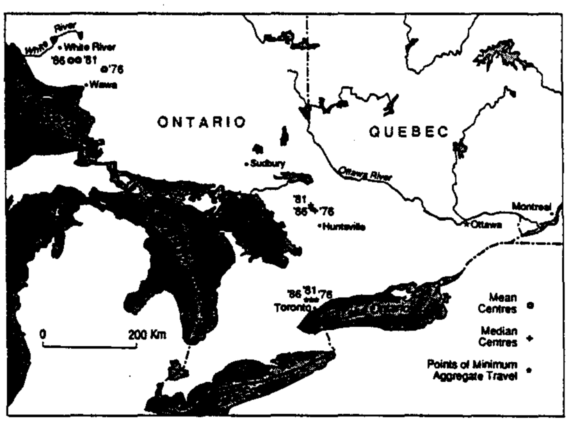

Figure 2. The population centers of Canada in 1976, 1981, and 1986.

Conclusions

Unfortunately, there is no ideal measure of the center of a population. The median center is insensitive to the distances of points from the center; the mean center puts undue weight on the distant points; and the point of minimum aggregate travel is insensitive to radial movements of the individual points. Of these shortcomings, the insensitivity to radial movement is the least severe, and the point of minimum aggregate travel is recommended as the best measure of the center of a population. The mat point can be determined for a large number of observations with a reasonable amount of computing time, and the characteristics of the point are easily understood. The point of minimum aggregate travel should be regarded as the center of a population.

Acknowledgments

The authors thank Statistics Canada for cooperation and financial support, and the University of California, Santa Barbara, Cartographic Laboratory for preparing the map.

References

Bureau of the Census. 1923. Center of population and median lines and centers of area, agriculture, manufactures, and cotton. In *Fourteenth Census of the United States, 1920.* Washington, DC: U.S. Government Printing Office.

Cooper, L. 1963. Location-allocation problems. *Operations Research* 11:331-343.

Eells, Walter C. 1930. A mistaken conception of the center of population. *Journal of the American Statistical Association*, New Series, 25(169):33-40.

Gini, C., and Galvani, L. 1929. Di talune estensioni del concetti di media a caratteri qualitativi [Of some extensions of the conceptions of average to qualitative characters]. *Metron* 8:3-209.

Hayford, John F. 1902. What is the center of an area, or the center of a population? *Journal of the American Statistical Association*, New Series, 58:47-48.

Hilgard, J. E. 1872. The advance of population in the United States. *Scribner's Monthly* 4:214-218.

Kensit, H. E. M. 1934. The centre of population moves west. *Canadian Geographic Journal* 9:262-269.

Kuhn, H., and Kuenne, R. 1962. An efficient algorithm for the numerical solution of the generalized Weber problem. *Journal of Regional Science* 4:21-34.

Meihle, W. 1958. Link length minimization in networks. *Operations Research* 6:232-243.

Neft, David S. 1966. *Statistical Analysis for Areal Distributions.* Philadelphia: Regional Science Research Institute.

Porter, Philip W. 1963. What is the point of minimum aggregate travel? *Annals of the Association of American Geographers* 53:224-232.

———. 1964. A comment on "The elusive point of minimum travel." *Annals of the Association of American Geographers* 54:403-406.

Poulsen, Thomas M. 1959. Centrography in Russian geography. *Annals of the Association of American Geographers* 49:326-327.

Rizzo, Brian R. 1982. *The Point of Minimum Aggregate Travel: Graphical, Analog, and Numerical Solutions.* M.A. Thesis, Department of Geography, University of Western Ontario.

Scates, Douglas E. 1933. Locating the median of the population in the United States. *Metron* 11:49-65.

Sviatlovsky, E. E., and Eells, Walter C. 1937. The centrographical method and regional analysis. *Geographical Review* 27:240-254.

Weiszfeld, E. 1937. Sur le point pour lequel le somme des distances de *n* points donnés est minimum. *Tohoku Mathematical Journal* 43:355-386.