Forming Area Sampling Units with Geospatial Sorts, with Application to the Program for the International Assessment of Adult Competencies (PIAAC)

Benjamin Schneider[1], Tom Krenzke[1], Wendy Van de Kerckhove[1]
Ying Chen[1], Jennifer Kali[1]
[1]Westat Inc., 1600 Research Boulevard, Rockville, MD 20850

Abstract
Multistage area sample designs are often employed to reduce the cost of in-person surveys, where sampling units are formed by combining geographic areas such as counties or Census blocks. Sampling error is reduced by forming sampling units with low between-unit variance in outcomes of interest, and costs are typically reduced by forming sampling units which are compact, complete, and based on contiguous geographic units. Kali et al. (2021) presented an algorithm for grouping Census blocks into sampling units in a way that balances these factors. Census units are geospatially sorted using space-filling curves, and then sampling units are formed by adding Census units to a sampling unit until the sampling unit attains a minimum measure of size. The space-filling curves promote compactness but often yield split sampling units, which increases data collection costs. To reduce splits, in this paper, we introduce graph-based geospatial sorts based on Census units' adjacency relationships. We evaluate the geospatial sorts' performance in forming primary and secondary sampling units based on counties and Census blocks for PIAAC.

Key Words: segments, clusters, in-person surveys, Traveling Salesman

1. Introduction

1.1 Multistage Area Samples

When conducting in-person surveys, area probability samples are often used to reduce costs by reducing interviewer travel time and concentrating data collection effort in a sample of areas. Multistage sampling allows survey designers to benefit from area sampling at multiple geographic levels. The elements sampled at each stage of sampling are referred to as "sampling units." A common example of a multistage area sample design is to begin by sampling "primary sampling units" (PSUs) formed from counties or groups of counties, then within selected PSUs sample "secondary sampling units" (called SSUs or segments) formed from groups of Census blocks, and then finally select dwelling units within selected SSUs. Primary and secondary sampling units are often built from well-defined geographic areas for which reliable published statistics are available to use for design (e.g., stratification) or estimation (e.g., calibration). In the United States, the underlying geographic areas used to form PSUs are often counties or county-equivalent units, which are administrative, state-defined areas with a variety of official statistics available from sources such as the American Community Survey. For SSUs, the underlying geographic areas are typically Census blocks, which are small areas defined by the U.S. Census Bureau and generally the finest geographic level for which official statistics are published.

1.2 Geographic Characteristics of Sampling Units

When fielding an in-person survey, costs are typically reduced by forming sampling units that are geographically compact, contiguous, and have geographic integrity (i.e., an absence of holes), as these features typically simplify logistical planning and reduce data collectors' travel time. For these reasons, Valliant, Dever, and Kreuter (2018) suggest that PSUs formed from counties should have a maximum area limit and should not have "extreme length" (that is, square shapes should be preferred over other rectangular shapes). Kali et al. (2021) propose the use of seven geographic metrics that survey designers can use to compare alternative groupings of Census blocks into secondary sampling units in terms of compactness, contiguity, and integrity. The five compactness metrics include the Polsby-Popper measure, the Reock score, the Convex Hull score, the Length-Width difference, and the perimeter. With the exception of the perimeter metric, these compactness metrics in essence measure the difference between the sampling unit's shape and an idealized unit with a similar size but a more compact shape (e.g., a circle). Integrity is measured by counting the number of holes in a sampling unit, and contiguity is measured by counting the number of splits in a sampling unit. Figure 1 provides an example of three PSUs formed from the counties of Massachusetts, where the first PSU (labeled "1") is a split PSU formed from two non-contiguous sets of counties. Holes and splits are challenging for data collectors as they complicate data collectors' ability to identify a sampling unit's boundaries and can increase travel time required to complete a caseload within a sampling unit.

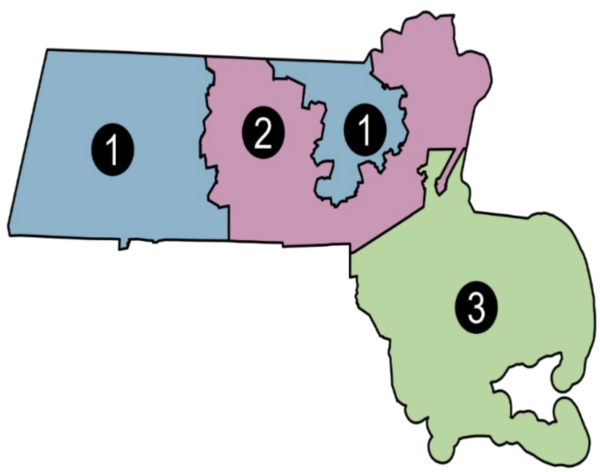


Figure 1: Example of a split sampling unit formed from counties in Massachusetts

1.3 Trade-offs in Costs and Survey Errors

It is a common finding in social surveys that nearby geographic areas tend to have more similar population characteristics compared to distant areas. For this reason, forming sampling units solely based on optimizing geographic characteristics such as compactness can increase sampling error for the survey, as sampling variances are inflated when the between-unit variance in population characteristics increases, which is typically associated with a decrease in the within-unit variance in population characteristics (Kish 1965). All else equal, both costs and survey errors tend to be reduced by forming sampling units with approximately equal population sizes (or more generally, measures of size, such as the number of housing units). This facilitates the assignment of equal interviewer workloads and in turn reduces non-sampling errors that arise in the form of interviewer effects (Groves 2004). In practice, however, the goal of obtaining equal population sizes for sampling units can conflict with the goal of forming compact sampling units, particularly in regions with varying levels of population density. For example, to form a PSU in rural Georgia with a population equal to that of a single county in the populous Atlanta region in the same state, it might be necessary to group together a large, sprawling collection of rural counties. As a strategy to balance these factors when forming PSUs, Valliant et al. (2018) suggest simultaneously imposing a minimum population size requirement and a maximum area requirement. It is important to note that there are other sample design aspects that may help to improve upon attaining a balance relating to cost and error reduction. For example, a one-PSU-per-stratum design would use the formed PSUs and then form close-to-equal size strata that reduce the between-PSU variance while achieving close-to-equal size workloads and a reduction in sampling variances for estimated totals.

1.4 Methods for Forming Sampling Units

1.4.1 Forming sampling units based on sorts

Kali et al. (2021) described a general approach to forming sampling units from lower-level geographic units using what could be described as a "sort-and-accumulate" method. At the outset, the survey designer establishes a minimum measure of size that must be attained by each sampling unit. Then, a list of geographic units is sorted into an ordered list based on a characteristic such as its numeric ID assigned by the government. Next, the survey designer assigns each county to a sampling unit by working their way down the list. The first geographic unit on the list is added to the first sampling unit, and successive geographic units on the list are added to the first sampling unit until that sampling unit attains some specified minimum measure of size (e.g., a population of 10,000 persons). Once the first sampling unit attains the minimum measure of size, geographic units are no longer added to it, and the second sampling unit is then formed in the same manner by proceeding further down the list until it too attains the minimum measure of size. This process proceeds until every geographic unit has been assigned to a sampling unit, and the last sampling unit in the list can be collapsed with a preceding sampling unit if it fails to attain the minimum measure of size. Table 1 provides an example of this process applied to forming PSUs with a minimum measure size of 15,000 based on a list of counties sorted by their numeric ID.

Table 1: Example of Grouping Counties into PSUs after Sorting

County ID	*Measure of size*	*PSU assignment*
001	10,000	1
002	4,000	1
003	6,000	1
004	10,000	2
005	9,000	2

1.4.2 Geospatial sort methods

In the example of Table 1, geographic units were sorted based on a numeric identifier code assigned by the Federal Government. This approach is only useful for sampling unit formation in limited cases since the connection between ID codes and geographic location can be strong for some types of units but tenuous for others. In the case of Census blocks defined by the U.S. Census Bureau, block IDs assigned to a pair of adjacent blocks tend to be more similar than block IDs assigned to a pair of non-adjacent blocks. For counties, the three-digit Federal information processing standard (FIPS) code identifiers assigned by the Federal Government are generally based on counties' names' rather than their locations, which makes them inappropriate for use as a geospatial sort.

Rather than sorting geographic units based on their ID codes, Kali et al. (2021) proposed forming sampling units by first sorting units using space-filling curves, which are frequently used as tools to sort points in two-dimensional (or greater) spaces into a one-dimensional list in such a way that nearby points in the higher-dimensional space appear at similar points in the one-dimensional list. This proposal was inspired by the use of Peano curves (Peano 1890) by Garrett and Harter (1995) to sort pre-defined sampling units for the purposes of conducting systematic sampling. Kali et al. (2021) evaluated the use of three space-filling curves—the Peano curve, the Hilbert curve (Hilbert 1891), and geohash—for the purpose of forming block-based segments within counties. Their evaluation found that sorting based on the space-filling curves tended to produce substantial improvements in compactness, contiguity, and integrity relative to sorting based on block ID. Of the space-filling curves considered, the Hilbert curve was found to produce the best sets of segments.

Kali et al. (2021) further proposed using not just a single sort method to form sampling units for the entire frame, but instead using a "hybrid" approach, wherein all of the available sort methods are attempted for a given hard boundary (a state, for example), and the "best" method is ultimately used for that hard boundary. To determine the "best" method, each of the four sets of sampling units (based on the four sorting methods) are ranked along each metric (e.g., Reock score), resulting in a summary of each set's relative performance with respect to each metric. From these rankings, a weighted average across the metrics is computed, yielding an overall summary index ranking the four alternative sets of sampling units from best to worst. Crucially, the use of a weighted average allows the survey designer to balance different geographic characteristics against one another and against anticipated between-unit variance based on the specific needs of the survey. For instance, in surveys where traditional in-field listing operations must be conducted to construct a frame of housing units within an SSU, the designer can assign greater weight to the splits or holes metrics in SSU formation to avoid costly errors caused by confusion over irregularly-shaped sampling units. Because sampling units are formed independently in each hard boundary, the hybrid algorithm identifies and adopts the best sort method for each hard boundary, for example, determining that in one PSU in California the best set of segments are those created using the Hilbert sort but in another California PSU the best set of segments are those created using the Peano sort method.

2. Introducing Graph-based Sort Methods

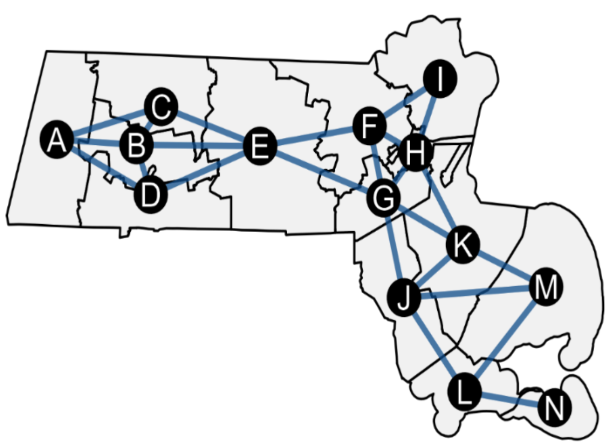
The method of sorting geographic areas based on space-filling curves takes into account only the coordinates of each geographic area's centroid (i.e., its geographic center) and thus only avoids grouping non-adjacent geographic units together in a sort key as a fortunate side-effect of attempting to group geographic units with nearby geographic centers. In applications, the sort methods were found to produce sufficiently many splits to warrant manual review and editing of sampling units as part of the frame creation process, even when the hybrid method was used to choose the best sort method for each hard boundary. This observation motivated the development of two graph-based geospatial sorts that take into account Census units' adjacency relationships (represented as an adjacency graph) and directly attempt to avoid grouping non-adjacent geographic units together in a sort key. The first method, which we term "the traveling salesman problem (TSP) method," uses optimization methods from graph theory to form a geospatial sort key that minimizes the number of non-adjacent geographic units placed next to each other. The second method, which we refer to as the "sorted neighbors method," uses a greedy optimization algorithm to place geographic units with nearby centroids into similar positions in the resulting sort key.

2.1 TSP Method

The first graph-based sort method we propose is based on the following observation about the formation of PSUs from counties within a state, although it applies to other geographic levels as well. Our goal with a geospatial sort of counties is to draw a sort path that passes through each county in a state exactly once, ideally reducing the number of connections in the path between non-adjacent counties (i.e., "jumps"). This goal can be reframed as a special case of a classical optimization problem in graph theory called the ""TSP. In the general formulation of the TSP, the goal is to "tour" each county in a state by visiting each county exactly once, returning to the starting point, and making this tour as short as possible with respect to some measure of distance. The TSP can be tailored to our purpose by choosing a distance measure that represents the number of "jumps" required to connect a pair of geographic units. Additionally, the classical TSP requirement of ending the tour at its starting point (i.e., identifying a Hamiltonian cycle on the graph) can be easily circumvented using the "dummy method" described by Hahsler and Hornik (2007).

2.1.1 Graph-based geodesic distance

An appropriate distance measure for this instance of the TSP is based on an adjacency graph, which encodes information about the adjacency relationships among geographic units by representing each unit as a node, with an edge connecting two nodes if and only if the respective counties are adjacent (i.e., are contiguous based on their borders). Based on this graph, we can define a geodesic (i.e., "shortest-path") distance measure, which for any pair of geographic units is defined as the minimum number of edges on the adjacency graph that must be traversed in order to connect their respective nodes. Figure 2 provides an example adjacency graph summarizing the 13 counties of Massachusetts and displays the corresponding geodesic distance matrix. In the first row of the distance matrix, we can see that the distance of county A from itself is 0. Because counties A and B are adjacent, their distance is 1. Counties A and E have a distance of 2 because, while they are not adjacent, they are connected by a path on the graph with two edges, A-C and C-E.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	0	1	1	1	2	3	3	4	4	4	4	5	5	6
B	1	0	1	1	1	2	2	3	3	3	3	4	4	5
C	1	1	0	2	1	2	2	3	3	3	3	4	4	5
D	1	1	2	0	1	2	2	3	3	3	3	4	4	5
E	2	1	1	1	0	1	1	2	2	2	2	3	3	4
F	3	2	2	2	1	0	1	1	1	2	2	3	3	4
G	3	2	2	2	1	1	0	1	2	1	1	2	2	3
H	4	3	3	3	2	1	1	0	1	2	1	3	2	4
I	4	3	3	3	2	1	2	1	0	3	2	4	3	5
J	4	3	3	3	2	2	1	2	3	0	1	1	1	2
K	4	3	3	3	2	2	1	1	2	1	0	2	1	3
L	5	4	4	4	3	3	2	3	4	1	2	0	1	1
M	5	4	4	4	3	3	2	2	3	1	1	1	0	2
N	6	5	5	5	4	4	3	4	5	2	3	1	2	0

Figure 2: Adjacency graph and corresponding geodesic distance matrix

2.1.2 Solutions to the TSP

The TSP is a famous example of an NP-hard problem: there is no efficient algorithm for determining an optimal solution except in special cases. Because the TSP admits any valid distance metric, it is applicable to a vast range of applied problems and has important applications in tasks such as vehicle routing or biostatistical clustering of genetic sequencing data. As such, there is a rich literature describing heuristic solution algorithms that are not guaranteed to be optimal but which have proven satisfactory in many cases and in some cases have proven bounds on their departure from optimality. Rosenkrantz et al. (1977) describe two classes of heuristic algorithms—insertion methods and k-optimal exchanges—and discuss their relation to the optimal tour of cases of the TSP. Insertion methods, on the one hand, construct a tour step-by-step, whereas k-optimal exchange algorithms improve an existing tour, such as a randomly-selected tour or even a tour constructed using an insertion method or other heuristic algorithm. For our applications, we constructed tours using the "farthest insertion" algorithm and refined the resulting tours using the "2-opt exchange" algorithm.

The "farthest insertion" algorithm gradually builds a tour step-by-step, at each step adding one additional node to the tour, always inserting each additional node into the subtour at the position that least increases the subtour's length. The new node added at each step is always the node not yet on the tour that is farthest from any of the nodes already on the tour. Figure 3 illustrates the first four steps of the farthest insertion algorithm applied to the counties of Massachusetts. At Step 1, the tour is initiated by connecting the two most distant counties—nodes A and N on the adjacency graph. At Step 2, node I is added to the tour between nodes A and N because node I was the unadded node at that step that was most distant from nodes A and N. Steps 3 and 4 similarly add nodes E and J to the tour based on their distance from the other nodes previously added to the tour.

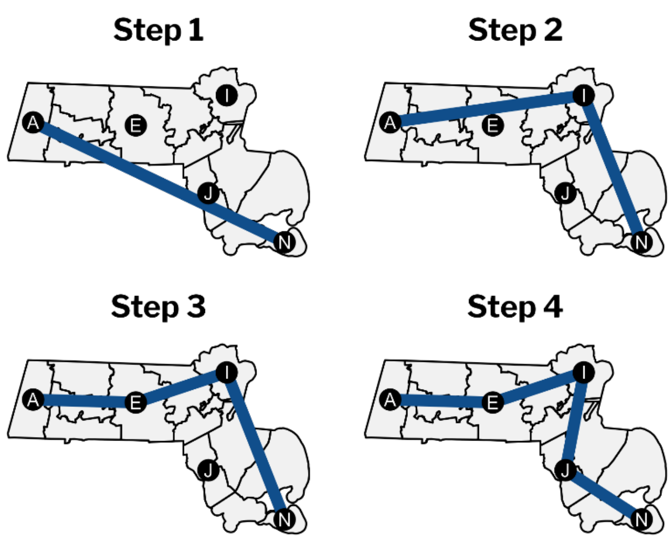


Figure 3: First four steps of the farthest insertion algorithm applied to Massachusetts counties

The "2-opt exchange" algorithm proposed by Croes (1958) iteratively improves an existing tour by exchanging two connections among a set of four points on the tour whenever the total length of the tour could be reduced by the exchange. Figure 4 provides an example where the total length of the tour is reduced by exchanging the connections A-B and C-D with the connections A-C and B-D.

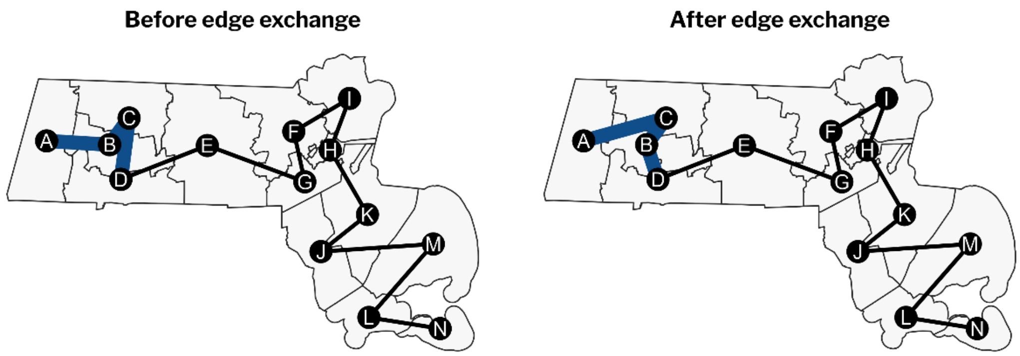


Figure 4: Illustration of edge exchange in 2-opt exchange algorithm

2.2 Sorted Neighbors (SN) Method

In addition to sorting geographic units with the TSP method based on the established heuristic algorithms, we also devised an additional simple sort method we refer to as the "SN method." Similar to the traveling salesman method, the SN method is not based on space-filling curves but is instead a heuristic algorithm applied to a discrete list of Census units. Unlike the space-filling curve methods or the TSP sort method, the SN algorithm takes into account both coordinate-based distance (derived from centroids) and adjacency relationships among geographic units. The use of both types of information was expected to potentially reduce splits relative to the sort methods based on space-filling curves while potentially improving compactness relative to the TSP method. The SN method iteratively builds a sort key by adding an additional geographic unit to the end of the key at each iteration, until every geographic unit has been added to the sort key. The SN method is a greedy algorithm: each time a geographic unit is added to the sort key, it is because it is the nearest adjacent unit to (i.e., "neighbor") of the previously-added unit (or simply the nearest unit, if there are no neighbors), pulled from the list of all the remaining neighbors not yet added to the sort key.

The SN method of sorting geographic units proceeds as follows:

Step 1. *Rank the geographic units highest to lowest based on their centroid's longitude coordinate. If there are ties for the rank based on longitude, rank the tied units based on their latitude coordinates.*

Step 2. *Initialize the sort key with a single geographic unit, denoted geographic unit i, taken by selecting the geographic unit with the highest rank from Step 1.*

Step 3. *Form a list of candidate geographic units to potentially add to the sort key. If geographic unit i has any adjacent units not already on the sort key, then these adjacent units form the list of candidate units. Otherwise, the list of candidate units is the list of all geographic units not already on the sort key.*

Step 4. *From the list of candidate geographic units, select the geographic unit with the highest rank from Step 1. Add this geographic unit to the sort key.*

Step 5. *Set the most recently-added geographic unit as geographic unit i.*

Step 6. *Repeat Steps 3 through 5 until every geographic unit has been added to the sort key.*

It is important to note that the outcome of the SN method is not invariant to the ranking method used in Step 1 of the algorithm. For our applications, we ranked the units first by their centroid's longitude coordinate and next by their centroid's latitude coordinate. However, it may be advantageous to first rotate the coordinate system using principal components analysis, since differences among units in terms of their first principal component would more closely correspond to Euclidean distance compared to differences in an arbitrarily rotated coordinate reference system.

3. Application to the PIAAC Survey of Adult Skills

PIAAC measures adult information-processing skills (e.g., literacy) using an in-person assessment by interviewers at survey respondents' homes. For U.S. data collection in 2022, the survey design employed multistage area sampling to reduce costs of field operations, which includes in-person interviews as well as compiling lists of housing units in areas deemed to be at risk of undercoverage on an address-based sampling frame. Area sampling units were formed for the first two stages of sampling. In the first stage, PSUs were formed from clusters of counties within states. Next, within PSUs selected in the first stage of sampling, SSUs were formed from clusters of Census blocks within Census tracts.

3.1 Sampling Unit Formation Method

Five alternative sets of PSUs were formed for each hard boundary: three sets of PSUs were formed using the space-filling curves (Peano, Hilbert, and geohash), and two sets of PSUs were formed using the graph-based geospatial sort methods (TSP and SN). For actual survey operations, only the three sets of PSUs based on space-filling curves were considered, but for our present discussion, we present the results from comparing all five geospatial sort methods. PSUs were formed using states as hard boundaries and a minimum measure of size of 15,000 non-institutionalized persons between the ages of 15 and 74. The hybrid algorithm of Kali et al. (2021) was used to select the best set of sampling units in each hard boundary. As a proxy of between-unit variance in the primary outcomes of interest—adult literacy and numeracy rates in 2022-2023—we calculated between-unit variance in small area estimates of the population age 16 to 74 below Level 1 in literacy based on the PIAAC Cycle I rounds of 2012, 2014, and 2017 (see Krenzke et al. 2020 for details of the small area estimation methodology). To balance cost-related geographic factors against the anticipated between-unit variance, the summary index proposed by Kali et al. (2021) was computed using weights of 1/2 for the splits metric, 1/4 for the donuts metric, 1/8 for the between-units variance metric, and 1/40 for each of the five compactness metrics.

In the SSU formation process, six alternative sets of SSUs were formed using sorts based on the Census block IDs, the three space-filling curves, and the two graph-based geospatial sort methods. All six sets were compared for the actual survey design. SSUs were formed using tracts as hard boundaries, based on a minimum measure of size of 120 housing units. Since small area estimates of literacy rates were not available at the sub-county level and few reliable official statistics are published at the block level, the between-units variance was calculated based on estimated counts for the 18+ Hispanic or

Latino population and Black or African American Alone population, both of which were identified in the PIAAC Cycle I small area modeling as useful predictors of small areas' literacy and numeracy rates (Krenzke et al. 2020). The hybrid sort algorithm was also used in the SSU formation process, where the summary index was computed with an importance weight of 13/20 assigned to the splits metric, a weight of 3/20 assigned to the donuts metric, and weights of 1/30 assigned to each of the compactness metrics and to the between-units variance metric. Compared to the PSU formation process, larger weights were assigned to the integrity and contiguity metrics for SSUs.

3.2 Results

3.2.1 Average values for geographic metrics

To evaluate the performance of each sort method as well as the hybrid method, we summarized average values of the seven geographic metrics described by Kali et al. (2021), calculated separately for sampling units formed by each sort method. Figure 5 displays the values of each of the seven geographic metrics for the average PSU (or county-based sampling unit). The metrics at left in yellow (splits, donuts, perimeter, and length – width) should ideally be small, while the metrics at right in blue (Reock, Polsby-Popper, and Convex Hull scores) should ideally be large. For each metric, we display six bars, one for each of the five geospatial sort methods (Hilbert, Peano, geohash, SN, and TSP), and one for the hybrid sampling unit formation method. We observed large differences among the various sort methods in terms of the average number of splits produced. Sampling units formed using the Peano and geohash sort methods had the largest number of splits on average (0.2). Sampling units formed with the Hilbert sort had significantly fewer splits on average (0.12), and the new adjacency-based sort methods reduced the splits still further (0.08 on average for both the TSP and SN methods). For other metrics, the Hilbert and TSP methods consistently yielded better results than the other individual sort methods, although to a much smaller extent than for the splits metric. The hybrid method produced slightly fewer splits than the graph-based methods, but effectively improved upon other geographic metrics it was tasked with balancing.

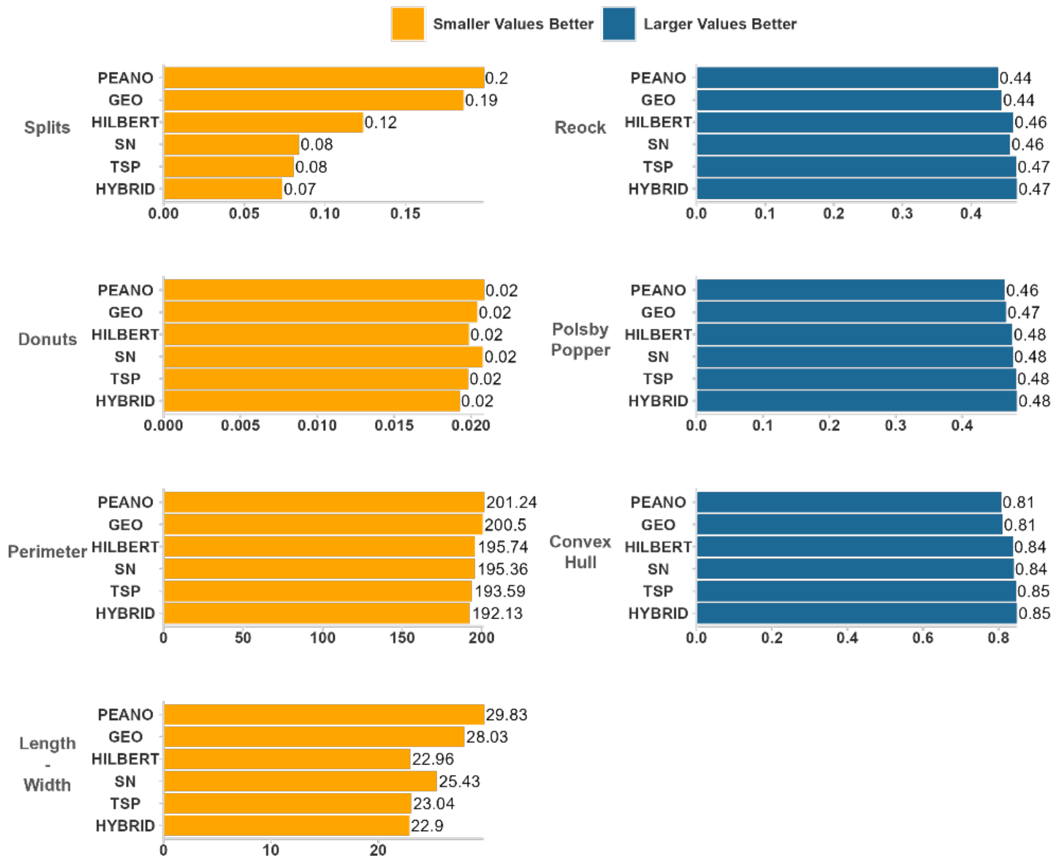


Figure 5: Geographic metrics for the average PSU in PIAAC, by sort method

Figure 6 similarly summarizes the same performance metrics for the block-based SSUs formed for PIAAC. As was the case with PSUs, the TSP method produced the fewest splits on average (0.48). The SN method outperformed the Peano and geohash methods (which formed 0.69 and 0.67 splits on average, respectively), but was itself outperformed by the Hilbert method (forming 0.59 and 0.53 splits on average, respectively). Unlike with PSUs, the hybrid method did much better than any single one of the sorting methods in terms of reducing split SSUs, even though the hybrid algorithm was required to balance performance across several other metrics (all of which were also improved or at least not worsened). For other metrics, the Hilbert and TSP methods both were generally the most effective sorts (although the perimeter metric was the one exception where the TSP method performed worse than other sort methods), and the SN method tended to perform worse than the other geospatial sort methods. In general, using block IDs tended to perform significantly worse than using the geospatial sort methods in terms of reducing splits and donuts.

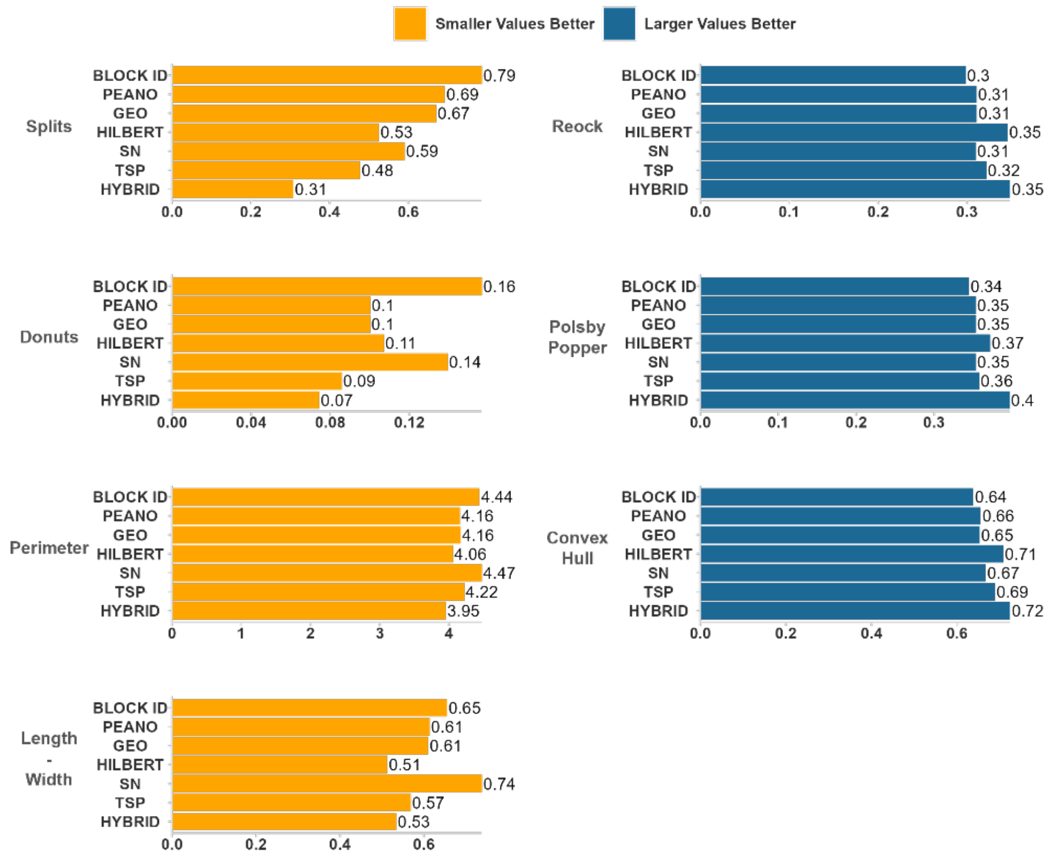


Figure 6: Geographic metrics for the average SSU in PIAAC, by sort method

3.2.2 Prevalence of splits among sampling units

Figure 7 summarizes the prevalence of splits among PSUs and SSUs formed with each sort method. The left side of the graphic displays the prevalence of splits among PSUs formed from counties, while the right side of the graphic displays the prevalence of splits among SSUs formed from Census blocks. For PSUs, the Hilbert curve yielded substantially fewer splits than the two other geospatial sorts based on space-filling curves: the Hilbert sort method led to splits in 8.3% of PSUs, while the Peano sort method led to splits in 15.7% of PSUs and the geohash sort method led to splits in 14.2% of PSUs. The new graph-based sort methods yielded further improvements that were quite substantial: some 4.2% of PSUs formed by the TSP method had splits, and 4.5% of PSUs formed by the SN method exhibited splits. Unsurprisingly, the hybrid method was substantially better than using any single geospatial sort method, resulting in 3.6% of PSUs containing a split.

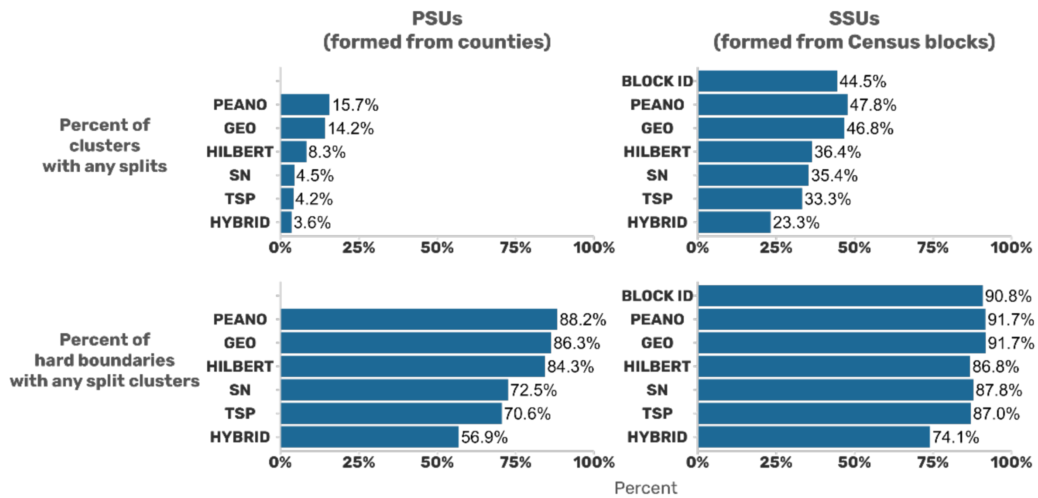


Figure 7: Prevalence of splits in PIAAC 2022 sampling units formed by each sort method

For secondary sampling units based on Census blocks, summarized in the right half of Figure 7, we see somewhat similar results. For SSU formation, the Hilbert curve method yielded splits in 36.4% of SSUs, which is a substantially lower prevalence rate than that observed in SSUs formed using the geohash or Peano curve methods (46.8% and 47.8%, respectively). The Hilbert curve method also performed much better than simply sorting blocks using their Census block ID (which resulted in splits in 44.5% of SSUs). The two adjacency-based sort methods—especially the TSP method—were nominally more effective than the Hilbert curves, with the TSP and SN methods yielding splits in 33.3% and 35.4% of SSUs, respectively. In short, our overall finding is that for the block-based SSUs, the TSP, SN, and Hilbert curve methods all performed similarly in terms of avoiding splits, and were generally much more effective than any of the other geospatial sorts considered here.

It is worth noting that no sampling unit formation method would be able to completely eliminate splits in this context without splitting a county across multiple PSUs or splitting a block across multiple SSUs. Some 1.6% of the counties included on the PIAAC 2022 sampling frame already contained splits prior to being grouped into PSUs, due to occasional instances of counties divided across multiple islands or other separated components. In addition, 0.005% of blocks in PSUs sampled for PIAAC 2022 contained splits prior to being grouped into SSUs. These prevalence rates of splits among geographic units represent an approximate lower bound on the rate of splits that could be attained by any sampling unit formation method, although this lower bound does not reflect practical constraints on sampling unit measures of size needed to control sampling variances and balance interviewer workload assignments.

4. Summary

The hybrid method developed by Kali et al. (2021) for sampling unit formation readily extends from the block level to the county level. The graph-based sort methods reduce splits in sampling units, relative to the sort methods based on space-filling curves. In PIAAC 2022, the new graph-based sort methods were especially effective at avoiding splits in the formation of county-based PSUs. At both the county and block levels, the TSP sort method in particular was able to reduce splits and donuts with little-to-no cost in terms of compactness, relative to using sorts based on space-filling curves or simply using Census block IDs. With the SN sort method, splits in PSUs and SSUs were also reduced relative to sorts based on space-filling curves or Census block IDs. Compared to the sorts based on space-filling curves, the SN method did not result in fewer donuts and generally produced less compact sampling units. Of the individual geospatial sort methods evaluated, the TSP and Hilbert methods were found to yield the best sampling units in terms of geographic characteristics. Compared to sampling units formed from any single geospatial sort method, the hybrid sampling unit formation method of Kali et al. (2021) produced substantially better geographic characteristics on average in terms of every geographic metric considered.

Acknowledgments

We would like to thank our colleagues at Westat for contributing to the development of this algorithm. In particular, we would like to thank Jeri Mulrow, Matthew Airola, and Lin Li for their valuable contributions in developing the software used to implement the methods described in this paper.

References

Croes, G. A. (1958), "A Method for Solving Traveling-Salesman Problems," *Operations Research*, 6, 791-812, DOI: 10.1287/opre.6.6.791.

Groves, R. M. (2004), "The Interviewer as a Source of Survey Measurement Error," in *Survey Errors and Survey Costs* (2nd ed.), Chapter 8, New York: Wiley-Interscience.

Hahsler, M., and Hornik, K. (2007), "TSP – Infrastructure for the Traveling Salesperson Problem," *Journal of Statistical Software*, 23, 2, 1-21, DOI: 10.18637/jss.v023.i0.

Hilbert, D. (1891), "Über die Stetige Abbildung Einer Linie auf Ein Flächenstück," *Mathematische Annalen,* 38, 459-460.

Kali, J., Krenzke, T., Chen, J., Green, J. and Chen, Y. (2021), "Evaluation of Methods to Form Segments from Census Blocks in Area Sample Designs," in Joint Statistical Meetings *Proceedings of the Survey Research Methods Section.*

Krenzke, T., Mohadjer, L., Li, J., Erciulescu, A., Fay, R., Ren, W., Van de Kerckhove, W., Li, L., and Rao, J. N. K. (2020). "Program for the International Assessment of Adult Competencies (PIAAC): State and County Estimation Methodology Report (NCES 2020-225)," U.S. Department of Education, National Center for Education Statistics, Washington, DC: U.S. Government Printing Office.

Peano, G. (1890), "Sur une Courbe, qui Remplit Toute une Aire Plane," *Mathematische Annalen,* 36, 1, 157–160, DOI: 10.1007/BF01199438.

Rosenkrantz, D.J., Stearns, R.E., Lewis, P.M. (1977), "An Analysis of Several Heuristics for the Traveling Salesman Problem," SIAM Journal on Computing, 6, 3, 563-581

Valliant, R., Dever, J., and Kreuter, F. (2018), "Rules for Defining PSUs," in *Practical Tools for Designing and Weighting Survey Samples* (2nd ed.), Chapter 10, Cham, Switzerland: Springer, DOI: 10.1007/978-3-319-93632-1.