Coğrafi Bilimler Dergisi / Turkish Journal of Geographical Sciences
2026; 24(1): 213-236 DOI: 10.33688/aucbd.1805466
Research Article / Araştırma Makalesi

GIS-Based enumeration area delineation: A trial for Türkiye with selected examples from the United States

CBS tabanlı sayım alanlarının oluşturulması: Türkiye için bir deneme ve Amerika Birleşik Devletleri'nden seçilmiş örnekler

Cansu Öztürk [*a], **Ahmet Sinan Türkyılmaz** [b]

[a] Hacettepe University Institute of Population Studies, Department of Social Research Methodology, Ankara/Türkiye.
[b] Hacettepe University Institute of Population Studies, Department of Social Research Methodology, Ankara/Türkiye.

Abstract

This study proposes a Geographic Information System (GIS)-based approach to delineate Enumeration Areas (EAs) in Türkiye. Analyses were conducted on four datasets from the United States and Türkiye to examine how the distributional properties of input variables and spatial contiguity constraints influence the performance of zone creation. International practices are reviewed to contextualize Türkiye's position and to derive lessons for best-practice adoption. The workflow tests alternative threshold values, contiguity rules, and advanced parameter settings to produce EAs that are geographically coherent, operationally workable, and statistically balanced. The findings indicate that GIS-based delineation can support the creation of clearly bounded, cost-effective, and more manageable areas. The study concludes that strengthening EA systems is not merely a technical exercise but a strategic investment in the efficiency and quality of national statistical operations.

Keywords: *Population, Enumeration area, Sampling frame, Demography, Geographic information systems*

Öz

Bu çalışma, Türkiye'de Sayım Alanlarının (SA) Coğrafi Bilgi Sistemleri (CBS) tabanlı tanımlanmasına yönelik bir yaklaşım önermektedir. Çalışmada, Amerika Birleşik Devletleri ve Türkiye'den olmak üzere dört veri seti üzerinde uygulama yapılmış; girdi değişkenlerinin dağılım özellikleri ile mekânsal bitişiklik kısıtlarının alan oluşturma performansını nasıl etkilediği incelenmiştir. Uluslararası uygulamalar, Türkiye'nin konumunu bağlamlandırmak ve iyi uygulamaların benimsenmesine yönelik çıkarımlar üretmek amacıyla gözden geçirilmiştir. Uygulamada; farklı eşik değerler, bitişiklik kuralları ve gelişmiş parametre ayarları altında, coğrafi olarak bütüncül, sahada uygulanabilir ve istatistiksel açıdan dengeli sayım alanları oluşturulması hedeflenmiştir. Bulgular, CBS tabanlı alan tanımlamanın sınırları net, maliyet etkin ve daha yönetilebilir alanlar oluşturmayı destekleyebileceğini göstermektedir. Çalışma, SA sistemlerinin güçlendirilmesinin yalnızca teknik bir süreç olmadığını; ulusal istatistik operasyonlarının verimliliği ve kalitesi açısından stratejik bir yatırım olduğunu ortaya koymaktadır.

Anahtar Kelimeler: *Nüfus, Sayım alanı, Örnekleme çerçevesi, Demografi, Coğrafi bilgi sistemleri*

*** Sorumlu Yazar/Corresponding Author:** ozturkcansu07@gmail.com
Geliş/Received: 17.10.2025 **Kabul/Accepted:** 17.02.2026 **Yayın Tarihi/Online Published**: 22.04.2026
Atıf /To Cite: Öztürk, C., Türkyılmaz, A.S. (2026). GIS-Based enumeration area delineation: A trial for Türkiye with selected examples from the United States, *Coğrafi Bilimler Dergisi/ Turkish Journal of Geographical Sciences,* 24(1), 213-236, https://doi.org/10.33688/aucbd.1805466

1. Introduction

An enumeration area (EA) is a small, clearly bounded geographic unit used as the basic building block for censuses and household surveys. It is designed so that one enumerator (or a small team) can cover all dwellings within its boundary during a defined fieldwork period. EAs are typically constructed to ensure complete coverage without overlap or gaps, to contain a manageable and broadly similar workload (often defined by a target number of households or dwellings), and to follow recognizable features such as roads, rivers, or administrative boundaries so that they can be located and worked on consistently in the field.

Census geography refers to the standardized set of geographic units and their hierarchical relationships that a national statistical office uses to plan field operations and to produce, tabulate, and disseminate census and survey statistics. It provides a consistent geographic framework by organizing space into nested units, enabling data to be aggregated and compared across places and over time. A well-defined census geography supports full territorial coverage, improves operational planning, and strengthens the comparability of published statistics. Within this framework, EAs function as the core operational building blocks: they translate the census geography structure into actionable field units for workload allocation and coverage control, while also providing small-area units that can be aggregated into higher-level reporting geographies (United Nations, 2009; United Nations, 2021).

Accurate and efficient data collection is the cornerstone of modern national statistical systems. In both censuses and sample surveys, EAs serve as the smallest operational units, providing the basic spatial framework for organizing fieldwork and ensuring systematic coverage of the target population. A well-designed EA system contributes to complete population coverage, reduces data collection costs, and improves the overall quality, comparability, and consistency of statistical outputs. As the United Nations Statistics Division (United Nations, 2009) emphasizes, EA delineation should be guided by principles such as spatial compactness, manageable workload for enumerators, and alignment with administrative or natural boundaries. These principles are essential for ensuring that enumeration can be performed efficiently and that data collection activities remain standardized across diverse regions.

Beyond the technical dimension of delineation, the importance of EAs lies in their connection to the broader logic of sample surveys. In modern statistics, surveys are indispensable for generating timely, detailed, and representative information about populations (Groves et al., 2009). Full population enumeration, while theoretically desirable, is often infeasible due to the immense financial and logistical resources required. As a result, statistical offices rely on sample surveys, which collect data from a carefully selected subset of the population and generalize the findings to the entire population with measurable accuracy (Kish, 1965; Lohr, 2010). For this reason, the design of EAs directly affects the ability to construct reliable sampling frames—comprehensive lists or databases that serve as the basis for drawing representative samples.

The rationale for conducting surveys, particularly sample-based surveys, extends beyond resource efficiency. They allow statistical agencies to generate continuous and specialized insights into economic, demographic, and social processes without waiting for the decennial census cycle. For example, household budget surveys, labor force surveys, and health surveys rely on accurate EA delineation to ensure representative sampling at national and subnational levels (Dillman, Smyth, &

Christian, 2014). In this sense, EAs serve as the operational link between the statistical concepts of representativeness and the practical execution of fieldwork.

A critical methodological element of surveys is the sampling frame. The sampling frame is the operational listing of all units in the population from which the sample is drawn, and its quality directly determines the representativeness of survey results (Bethlehem, Cobben, & Schouten, 2011). A good sampling frame must be complete, accurate, current, and free of duplication. When frames are incomplete or outdated, coverage errors arise—leading to the systematic exclusion of certain population groups and thereby biasing estimates. In the context of Türkiye, the Spatial Address Registration System (SARS) provides a promising foundation for building modern sampling frames, as it combines geographic precision with administrative reliability. By shifting from traditional, text-based enumeration lists to spatially defined digital frameworks, SARS has the potential to reduce coverage errors and enhance operational efficiency.

Türkiye, like many other countries transitioning to register-based statistical systems, faces the challenge of modernizing its EA methodology in line with changing settlement patterns, rapid urbanization, infrastructure development, and increasing availability of digital geospatial data. Traditional EA systems, based on paper maps and descriptive boundary definitions, often fail to capture the complexity of urban sprawl or the fluidity of administrative boundaries. The introduction of spatially defined EAs using Geographic Information Systems (GIS) addresses this limitation by allowing automated delineation, dynamic updating, and integration of multiple data sources. International experience demonstrate that GIS-based EA systems not only improve efficiency but also support innovations in data collection, such as handheld devices, GPS-based field monitoring, and real-time quality assurance mechanisms (Kalton, 2020).

The role of EAs becomes even more significant when viewed from the perspective of data quality dimensions. According to Groves et al. (2009), survey quality depends on minimizing total survey error, which includes coverage error, sampling error, nonresponse error, and measurement error. The design and maintenance of EAs directly affect the first two components: coverage and sampling. Spatially consistent and up-to-date EAs ensure that every household has a known and nonzero probability of selection, thereby enhancing the representativeness and accuracy of survey results. At the same time, compact and manageable EAs reduce the workload for enumerators, lowering the risk of data collection errors.

The aim of this study is to propose a GIS-based approach for delineating Enumeration Areas (EAs) for Türkiye through a trial implementation. Selected case applications from the United States illustrate how properties of input variables, dataset scale, and spatial contiguity constraints influence automated zone creation. The Türkiye application illustrates how the workflow can be configured to produce EAs that are geographically coherent, operationally workable, and statistically balanced, and it has potential relevance for future census and household survey operations in Türkiye. In addition, the analysis considers institutional prerequisites for sustainable implementation, including data governance and interagency coordination, and reviews international practices to contextualize Türkiye's position and derive practical lessons for adoption. In this context, strengthening EA systems is regarded as an area for operational and governance-oriented improvement of national statistical systems.

2. Literature Review

The global literature demonstrates a significant evolution in the methods used for delineating Enumeration Areas (EAs) moving from traditional manual cartography to advanced, GIS-supported, and even automated processes. EAs are no longer viewed only as operational units for census fieldwork but as fundamental components of national statistical systems that interact with geospatial infrastructures, administrative registers, and digital technologies. This transformation reflects the growing demand for high-quality, timely, and spatially precise data for evidence-based policymaking, economic planning, and social research (Esri, 2020; United Nations, 2009).

In the early stages of census-taking, EAs were defined primarily by hand-drawn boundaries on paper maps, often following visible physical landmarks or administrative units. While functional, these approaches were highly dependent on local knowledge and difficult to replicate consistently. Errors were frequent, especially in rural or remote areas, where natural boundaries were vague and settlement patterns were scattered. By the mid-20th century, national statistical offices (NSOs)increasingly realized the need for more standardized methods, particularly as censuses grew in scope and complexity. The growing demand for household and labor force surveys further highlighted the importance of precise EA delineation, as survey quality is directly linked to the accuracy of the underlying sampling frame (Groves et al., 2009; Kish, 1965).

In developed countries, the transition from manual to digital approaches was accelerated by the availability of robust administrative registers and technological infrastructure. The United States, for instance, institutionalized the use of Census Blocks and Census Tracts through the TIGER/Line system, which, since the 1980s, has enabled detailed mapping, consistent boundary updates, and the integration of statistical data with digital geographies (U.S. Census Bureau, 2021). Canada's Dissemination Areas (DAs), established in 2001, provide standardized units for detailed socio-economic analysis across provinces and territories, facilitating cross-regional comparisons and supporting federal funding allocations (Statistics Canada, 2022).

The United Kingdom introduced Output Areas (OAs) in 2001, which have since become the backbone of small-area statistics, supporting not only census dissemination but also a wide variety of social and economic indicators (ONS, 2021). France's IRIS units, developed in 1999, similarly illustrate how statistical geography can be adapted to urban planning and sub-municipal policymaking (INSEE, 2021). Germany's Erhebungsbezirke, established in 1987, show how alignment with administrative and infrastructural realities supports high-quality labor force and household surveys.

Australia offers another important example with its Statistical Area Level 1 (SA1) units, designed to provide consistent small-area data for policy planning across a vast and diverse territory. These units are particularly effective in supporting local planning because they can be aggregated into larger statistical areas, ensuring flexibility in both design and application. Estonia, meanwhile, represents a pioneering case of digital innovation, in which EAs are dynamically generated through the integration of population registers and GIS. This adaptive approach allows for continuous updating and significantly reduces coverage error (Statistics Estonia, 2021).

In contrast, developing countries have faced greater challenges due to weaker administrative infrastructures, limited technical capacity, and financial constraints. Somalia's pilot use of high-resolution satellite imagery to delineate EAs (Qader et al., 2019) demonstrates how geospatial innovation can partially compensate for institutional fragility. Botswana (Statistics Botswana, 2016) has also used geospatial solutions to update enumeration boundaries, underscoring the potential of digital tools in resource-limited environments.

Kenya's 2019 census illustrates how GPS-supported enumeration geography can strengthen coverage control in rapidly changing urban environments. In the Kenyan case, georeferenced EA maps were loaded onto tablets, and household location data (GPS coordinates) were used to support quality checks and completeness verification during field operations (Wanyoike, 202**1**). India, a country with a population exceeding one billion, relies on Enumeration Blocks (typically designed to cover 120–150 households) and has increasingly integrated GIS-based databases into its census mapping and fieldwork organization (Chakravorty, 2007). Indonesia is not a "billion-plus" country; yet it similarly institutionalizes GIS-enabled small-area operational units through Wilayah Kerja Statistik (Wilkerstat), supported by satellite imagery, GPS devices, and GIS for census and survey mapping activities (BPS-Statistics Indonesia, 2018; Worldometer, 2025). In South Africa, Small Area Layers (SALs) were developed by aggregating small enumeration areas to meet confidentiality and usability thresholds for dissemination, demonstrating how GIS-enabled statistical geographies can extend beyond enumeration to support broader small-area statistical products (Mokhele, Mutanga, & Ahmed, 2016; Statistics South Africa, n.d.).

Brazil faces a distinct set of challenges because rapid urban growth and the dynamism of informal settlements can outpace intercensal boundary maintenance, creating divergences between census basemaps and municipal settlement maps. For example, evidence from São Paulo shows that differences in identification and delimitation of favelas between federal census and municipal basemaps can be substantial, which may contribute to underestimation and difficulties in maintaining stable small-area boundaries over time (Pedro & Queiroz, 2019). In response to these territorial dynamics, Brazil's national statistical office (IBGE) reports a major expansion and a decentralised update of its census cartography (Territorial Base) for the 2022 Census, supported by high-resolution imagery and locally sourced operational information, which are intended to reduce the risk that newly occupied areas are missed (IBGE, 2022). Nevertheless, the official post-enumeration coverage analysis for the 2022 Census documents material coverage gaps and explicitly identifies difficulty in covering large urban centres — including "slums and urban communities" — as one of the operational challenges associated with higher error rates (IBGE, 202**2**). Collectively, these observations underscore a broader point: the sustainability of EA systems depends not only on GIS technology but also on stable institutional arrangements that ensure continuous maintenance, adequate funding, and coordinated statistical–geospatial integration across agencies. Without these, GIS tools alone cannot guarantee data quality (PARIS21 & Statistics Sweden, 2021; United Nations, 2008).

A common thread in the literature is that EAs are not only essential for census-taking but also central to sample surveys, household surveys, and labor force studies. They ensure that representative samples can be drawn with known selection probabilities, thereby minimizing bias and improving the reliability of estimates (Lohr, 2010). Moreover, EAs facilitate spatial analysis by linking demographic and geographic information, which has become increasingly important in fields such as health, education, environmental monitoring, and infrastructure planning (Esri, 2020; United Nations, 2009).

Advantages of well-designed EAs include enhanced operational efficiency, reduced fieldwork costs, minimized risks of double-counting or omission, and improved integration of demographic and geospatial data (UNECE, 2013). By providing standardized geographic units, EAs allow policymakers to compare conditions across regions, identify vulnerable populations, and allocate resources more effectively. The visualization capabilities enabled by GIS further expand the utility of EAs by supporting decision-making through maps and spatial analytics.

Nevertheless, significant limitations and challenges remain. One persistent difficulty is the resource-intensive nature of maintaining EA boundaries. Updating requires not only technical expertise but also coordination among government agencies, local authorities, and field staff. In rapidly urbanizing countries, informal settlements and migration flows often outpace official boundary revisions, generating coverage errors. Privacy and confidentiality concerns also intensify as geospatial precision increases. Balancing the demand for small-area statistics with the need to protect individual anonymity remains a key methodological and ethical challenge (UNECE, 2015).

Emerging trends suggest several promising directions. First, adaptive and dynamic EA systems are gaining attention. Estonia exemplifies how continuous integration of administrative registers and GIS technologies can create real-time updating processes, reducing coverage errors and improving efficiency (Statistics Estonia, 2021). Second, the increasing role of artificial intelligence and remote sensing in automated EA delineation offers opportunities for timelier updates, particularly in countries with limited staff capacity. Third, interoperability between statistical and geospatial systems is becoming a priority, as it allows data from multiple sectors to be linked and analyzed at the EA level.

In Türkiye, there are currently no officially defined enumeration areas. Instead, in sample surveys—particularly household-based ones—artificial clusters are created within the same settlement unit (quarter or village) when household addresses are grouped by the Turkish Statistical Institute (TurkStat). These clusters serve as the primary sampling units. During the clustering process, the occupied addresses within each settlement are sorted by their address components (such as street or avenue name, building number, and apartment number) and then divided into clusters designed to contain approximately equal numbers of units (~100 households each). The sampling frame is updated and clusters are reconstructed twice a year. Since this process is not based on geographic coordinates, the clusters and the number of units they contain are not fixed and may vary over time.

The output of the current blocking procedure, which does not account for spatial characteristics, is illustrated in Figure 1. The author prepared the figure to align with the method's output.

Figure 4. Schematic illustration of the existing non-spatial blocking method

An early academic effort discussed the feasibility of designing a dedicated census geography for Türkiye using GIS-based delineation principles. In this study, a candidate small-area hierarchy was explored through a case study in Çankaya (Ankara),with the aim of producing standardised statistical units for census dissemination and for small-area statistics beyond the constraints of existing administrative boundaries (Kırlangıçoğlu, 2005). Importantly, this design was proposed as a methodological alternative but was not adopted as an official operational system in Türkiye's routine official statistical production. Nevertheless, it remains a useful reference point because it frames EA-like units as the backbone of national statistical geography and, by contrast, clarifies the institutional trade-off observed in practice. While boundary-independent statistical areas can be analytically attractive, official statistics and policy reporting often benefit from EA designs that can be consistently related to administrative hierarchies (e.g., through nesting or robust correspondence tables) and can also satisfy operational constraints such as workload balancing.

In conclusion, the literature clearly demonstrates that EA systems are evolving from static, manually defined units into dynamic, geospatially integrated infrastructures that underpin modern statistical systems. Developed countries leverage robust registers and advanced GIS tools to maintain up-to-date, highly detailed EAs, while developing countries employ innovative, resource-constrained approaches, such as satellite imagery or grid-based models (Table 1). Across contexts, the fundamental role of EAs in ensuring representativeness, accuracy, and efficiency remains constant, even as the methods and technologies evolve. Their continued development will be critical for the future of evidence-based policymaking, social research, and national statistical capacities worldwide.

Table 1. Comparative overview of EA systems

Country	EA Unit Type	Data Source/Frame	Technology Used	Update Frequency	Advantages	Challenges
United States	Census Blocks/Tracts	TIGER/Line database	GIS, digital mapping	Every 10 years (census)	Detailed small-area statistics, integration with mapping	Costly updates, confidentiality issues
United Kingdom	Output Areas (OAs)	National census registers	GIS-enabled, statistical registers	Every 10 years	Supports local policy and deprivation indices	Static boundaries, slow updates
Canada	Dissemination Areas (DAs)	Household and address registers	GIS, geocoding	Every 5 years (census)	Granular socio-economic analysis	Coverage issues in remote areas
France	IRIS Units	Census and administrative registers	GIS-based statistical system	Every census and major update	Sub-municipal level analysis for urban planning	Resource-intensive maintenance
Germany	Erhebungsbezirke	Administrative boundaries	GIS, integrated mapping	Before major surveys	Efficient for labor force surveys	Dependent on local admin data quality
Kenya	Enumeration Areas	Census lists	GPS-enabled, GIS-based	2019 census	High geospatial accuracy	Sustainability of GPS systems
Somalia	Grid-based EAs	Satellite-derived maps	High-resolution imagery	Pilot since 2014	Innovation under constraints	Fragile institutions, limited coverage
Brazil	Custom EAs	Field-based mapping	Mixed GIS/manual	Irregular updates	Flexibility in diverse contexts	Urban sprawl, informal settlements
Estonia	Adaptive EAs	Population registers	Dynamic GIS integration	Continuously updated	Real-time adaptation, low coverage error	Requires strong digital infrastructure
South Africa	Small Area Layers	Census and admin data	GIS-based, digital mapping	Every census	Supports census and ongoing surveys	High resource needs for updates
India	Enumeration Blocks	Census operations	Manual + GIS integration	Every 10 years	Covers large rural population	Challenging in remote areas
Australia	Statistical Area Level 1 (SA1)	National address file	GIS, ABS systems	Regular census cycles	Policy-relevant small-area data	Requires large administrative effort

Source: Compiled from national statistical office reports and international guidelines.

3. Methodology

The ArcGIS Build Balanced Zones (BBZ) tool was used to delineate enumeration areas (EAs) in four case-study applications—Richmond County, Seattle, Story County, and Ankara—to examine how properties of input variables, dataset scale, and spatial contiguity constraints affect automated zoning performance. To assess the tool's operational behavior and its sensitivity to extreme or atypical conditions, the methodology first applies controlled trials to international open datasets with stable schemas and well-documented attributes. These preliminary experiments are used to validate the core workflow, identify common failure modes, and clarify how boundary definitions, barrier configurations, and settlement patterns shape the resulting zones. Building on these insights, we then apply the same workflow to the Ankara case study, which is based on a single quarter.

In many countries, GIS platforms—and frequently ArcGIS—have become core components of the production and maintenance of census geography. The U.S. Census Bureau, for example, integrates GIS workflows with the TIGER/Line database to support nationwide geographic consistency for census tracts and blocks (U.S. Census Bureau, 2019). In Canada, ArcGIS Pro is used in the management of Dissemination Areas, which provide a standardized framework for small-area census dissemination and socio-economic analysis (Statistics Canada, 2022). Similar applications are reported in Kenya, where the 2019 census employed GPS-enabled field mapping to digitize and verify enumeration areas, aiming to improve coverage control and reduce operational errors (Kenya National Bureau of Statistics, 2020). In India, census geography and mapping activities have also been supported through GIS-based approaches, including the development and operational use of census geographic databases for enumeration planning and field implementation (Chakravorty, 2007). In Australia, the Australian Bureau of Statistics maintains Statistical Area Level 1 (SA1) units as the smallest geographic units for many demographic and socio-economic outputs, which are supported by ArcGIS Pro-based spatial management workflows (ABS, 2021). In France, IRIS units represent sub-municipal geographic divisions that support detailed census dissemination and urban analysis (INSEE, 2021). Taken together, these cases indicate that ArcGIS Pro is often used not only for cartographic production, but also as part of the operational infrastructure that supports the design, maintenance, and use of statistical geographies.

Within ArcGIS Pro, one of the most relevant tools for EA delineation is the Build Balanced Zones (BBZ) function. BBZ automates the creation of compact and balanced spatial units by optimizing user-defined variables. The tool aggregates small base units, such as census blocks or polygons, into larger meaningful zones. Unlike traditional manual delineation, which is time-consuming and prone to inconsistencies, BBZ relies on optimization algorithms that minimize shape irregularities while ensuring that each unit adheres to the balancing criteria. Users can define population thresholds, numbers of households, or land area as balancing variables, and specify a target value for each EA. For instance, zones can be designed to contain roughly equal resident populations while maintaining contiguity and compactness. BBZ also allows the integration of multiple balancing variables simultaneously, making it possible to create zones that consider not only the population but also the number of households, housing units, or service facilities.In the Build Balanced Zones (BBZ) method, "spatial constraints" define how neighboring features are identified and merged as zones expand (Esri, 2023). Zones are allowed to grow only into features adjacent to at least one feature already included in the same zone, ensuring geographical contiguity and consistency. When the input features are polygons, the default constraint is Contiguity Edges Corners, while for point-based inputs, the default is Trimmed Delaunay Triangulation. The BBZ tool offers several spatial-constraint options, including Contiguity Edges Only, Contiguity Edges Corners, Trimmed Delaunay Triangulation, and Get Spatial Weights from File. By defining these spatial relationships, the BBZ algorithm ensures that zones grow logically and maintain contiguity according to the selected constraint type.

The BBZ tool also provides a set of "Zone Characteristics Criteria" that shape the internal properties of the resulting zones. The main parameters are Equal Area, Compactness, and Equal Number of Features, which respectively aim to balance surface area, geometric regularity, and the number of input features. Adjusting these criteria allows users to fine-tune the trade-offs between spatial compactness, uniformity, and statistical balance (Esri, 2023).

Complementing these criteria, the BBZ tool includes "Advanced Parameters" that control the optimization process. The most influential parameters are Population Size, Number of Generations, and Mutation Factor, which define the scope of iterations, search diversity, and variability in achieving balanced zones. Together, these parameters enable users to balance computational efficiency with spatial and statistical precision.

This tool has found applications across a wide range of sectors. In the United States, education authorities have used BBZ to delineate school districts that maintain equitable student populations. In Europe, statistical offices have integrated BBZ into automated EA creation, particularly in densely populated urban areas where demographic variations are pronounced. The healthcare sector has also adopted the tool to delineate hospital service zones in ways that balance population coverage with geographic accessibility. Transportation planners use BBZ to define traffic analysis zones that support the equitable modeling of road usage and infrastructure demands (Goodchild & Li, 2021). These cross-sectoral applications underscore the flexibility of BBZ and its potential to ensure fairness, efficiency, and reproducibility in zone design.

The advantages of BBZ are particularly evident in large-scale statistical operations such as censuses. By creating compact and balanced zones, the tool ensures a more equitable workload distribution among field staff, reduces the risk of omission or duplication in data collection, and guarantees that the delineation process remains transparent and replicable. This is especially critical for developing countries, where limited resources make efficient EA design a priority. Automating zone creation using BBZ reduces the human and financial costs associated with traditional manual methods while enhancing data quality. Furthermore, the capability to incorporate spatial and demographic variables simultaneously ensures that the resulting EAs are not only statistically sound but also geographically meaningful.

This study applies the BBZ methodology to the Turkish context. Türkiye is modernizing its statistical infrastructure and has already established key register-based components through the Spatial Address Registration System (SARS). In recent years, SARS has been routinely used by municipalities in official administrative processes and by TurkStat in statistical production, indicating a level of operational maturity rather than an early-stage transition. The availability of SARS therefore provides a practical foundation for moving from traditional text-based enumeration frameworks toward spatially defined and more systematically managed EAs. In this study, SARS is used as a primary administrative source to support EA delineation. However, to fully capitalize on this opportunity, Türkiye must adopt methodological tools that ensure efficiency, accuracy, and scalability. The integration of ArcGIS Pro and the BBZ function directly addresses this need by allowing for automated delineation of compact and balanced EAs that align with demographic realities.

In practice, the methodology involves integrating demographic data, administrative boundaries, and other relevant spatial datasets within the ArcGIS environment. Base units, such as address points or census blocks, are aggregated into larger enumeration areas using BBZ. Criteria such as population size, area thresholds, and spatial contiguity are applied to ensure that the resulting units are both operationally feasible and analytically meaningful. For example, a target of 2,500 residents per EA might be established, with adjustments made to accommodate urban density patterns or rural sparsity. In addition, by setting parameters for compactness and contiguity, the tool avoids irregular or fragmented zones that could complicate field operations.

International best practices further illustrate the potential of this approach for Türkiye. In Estonia, adaptive EAs have been developed by integrating national registers with GIS, allowing real-time updates that reflect demographic changes (Statistics Estonia, 2021). In Somalia, high-resolution satellite imagery has been combined with automated delineation techniques to create grid-based EAs under severe resource constraints (Qader et al., 2019). These cases demonstrate that GIS-enabled automation is not only a luxury of developed countries but also an achievable innovation in developing and transitioning economies. By adopting BBZ within ArcGIS Pro, Türkiye positions itself alongside countries that are leveraging advanced spatial methodologies for statistical modernization.

The benefits of this methodological approach extend beyond census operations. Spatially balanced and well-defined EAs provide a foundation for representative household surveys, socio-economic studies, and administrative monitoring. They also facilitate integration with broader policy areas such as healthcare, education, infrastructure planning, and disaster risk management. In Türkiye, where rapid urbanization and demographic shifts create ongoing challenges for statistical agencies, an adaptive and GIS-based EA system ensures that statistical operations remain both reliable and relevant.

In conclusion, this methodology builds upon global experience with ArcGIS Pro and the BBZ function to address the challenges of EA delineation in Türkiye. By automating the creation of compact, contiguous, and demographically balanced zones, the approach enhances efficiency, reduces costs, and improves the accuracy of statistical outputs. Furthermore, it aligns Türkiye's statistical practices with international standards, ensuring interoperability with global data systems. Through the integration of demographic data, administrative registers, and spatial analysis tools, this study demonstrates how GIS technologies can serve as a catalyst for the modernization of national statistical infrastructures.

4. Application and Findings

This section presents the empirical applications of the proposed workflow and summarizes the outputs obtained across the case datasets. It first reports results from datasets from the United States, which illustrate how BBZ behaves across different input-variable distributions, dataset sizes, and levels of spatial contiguity. It then presents the main application conducted in a single quarter of Ankara (Türkiye),where multiple parameter configurations were tested using integrated administrative and spatial layers. The section documents the implemented thresholds, contiguity rules, and advanced parameter settings and reports the resulting EA configurations in terms of geographic coherence and statistical balance. Figures and maps are used throughout to demonstrate the input layers, intermediate outputs, and final zones produced in each case.

4.1. Application on the United States Datasets

4.1.1. Richmond County Data

The first dataset used for testing was obtained from the North Carolina Integrated Cadastral Data Exchange project. It includes parcel boundaries together with attributes such as ownership, parcel size, and assessed land value. Figure 2 illustrates the parcel–point structure.

Figure 5. View of the parcel and point data of the Richmond County

For this application, the variable 'Land Value' was selected as the balancing criterion using the BBZ tool. In this step, the tool was used to aggregate parcels into zones based on the distribution of land values. Table 2 presents the descriptive statistics of this variable.

Table 2. Summary statistics of the Land Value variable in the Richmond County data

Fieldname	Count	Min	Max	Mean	Median	Mode	Outliers	Iqr	Q1	Q3
LANDVAL	32518	0	30718148	33871	12800	15000	3353	19855	6037	25893

This dataset was primarily used to practice using the BBZ tool and to evaluate its performance on parcel-level cadastral data.

4.1.2. City of Seattle Data

The second dataset was obtained from the City of Seattle open data portal. It contains parcel boundaries with attributes including address, land use, and the number of housing units per parcel. Figure 3 illustrates the parcels and their housing unit data.

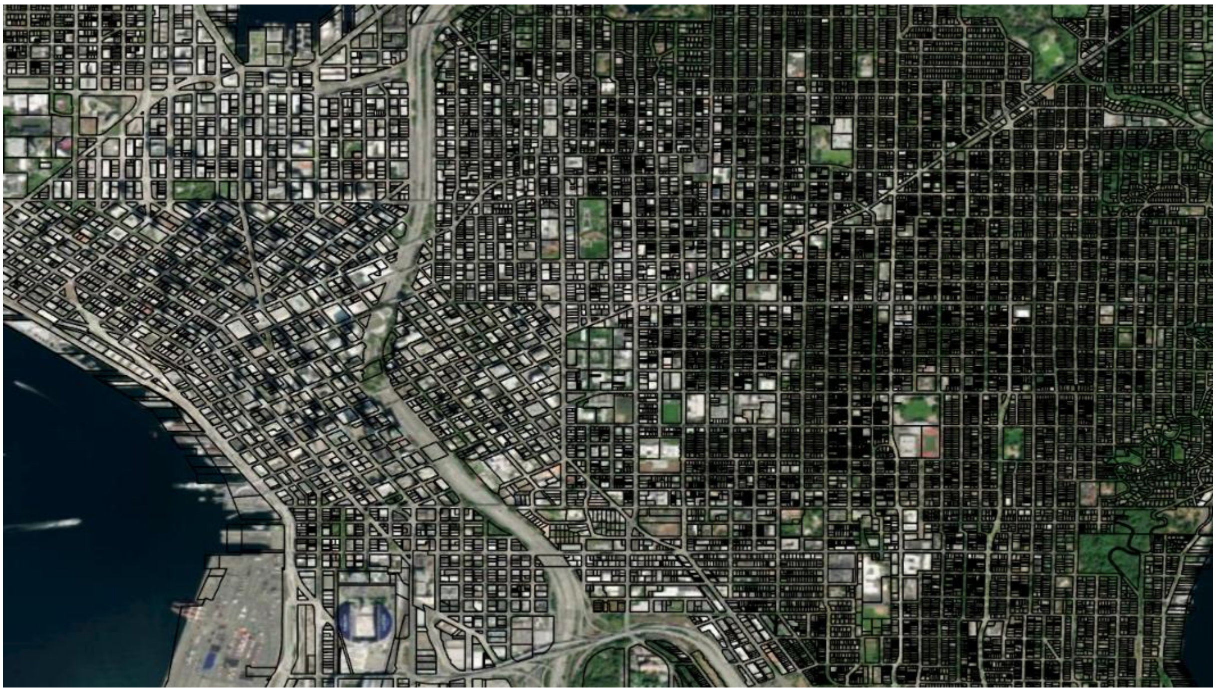

Figure 6. View of the parcel data of Seattle

In this application, the variable Number of Housing Units in the parcel (Exist_Unit) was used as the balancing criterion in the BBZ tool. The tool was applied to group parcels into zones so that each zone would have a comparable number of housing units. Table 3 provides descriptive statistics for the variable.

Table 3. Summary statistics of the Exist_Unit variable in the Seattle data

Fieldname	Count	Min	Max	Mean	Median	Outliers	Sum	Range	Iqr	Q1	Q3
EXIST_UNIT	178.897	0	707	1.8	1	35.485	330.527	707	0	1	1

This dataset was used to evaluate how the BBZ tool performs when applied to housing-related parcel attributes.

4.1.3. Story County, Iowa Data

The third dataset was obtained from publicly available cadastral records for Story County, Iowa. It comprises more than 43,000 parcels with attributes such as ownership, parcel identification numbers, assessed values, and land-use types. The spatial structure of the dataset is shown in Figures 4 and 5.

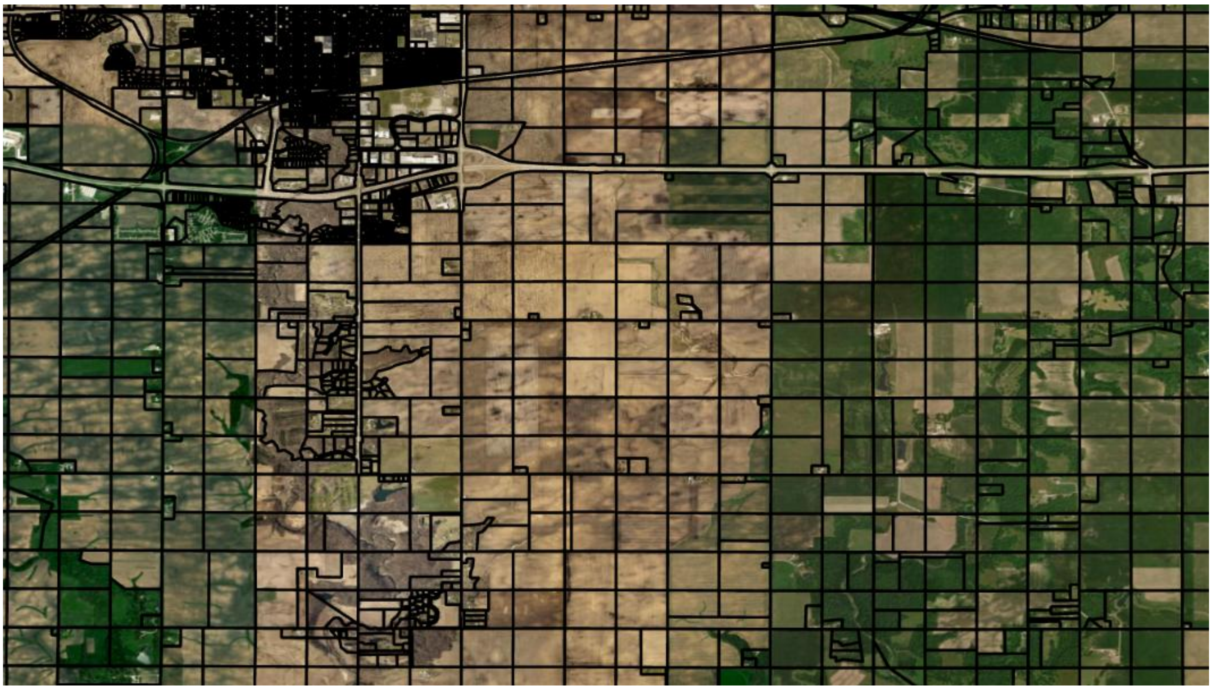

Figure 7. View-1 from Story County parcel data

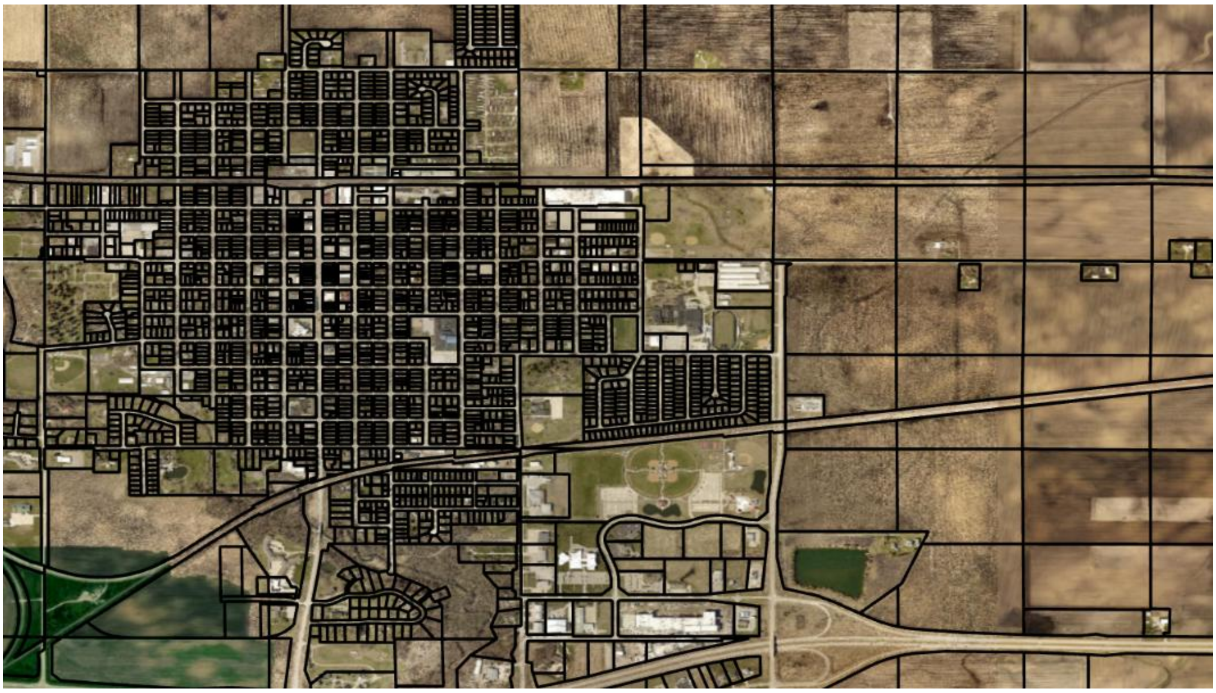

Figure 8. View-2 from Story County parcel data

For the application, the BBZ tool was run using parcel-level attributes as balancing variables. The "Homestead" and "FullDwelli" variables were used separately as target variables when creating enumeration areas from Iowa data. Thresholds were set for grouping parcels into zones, and different parameters were tested to observe how the tool handles a dataset of this size. Table 4 summarizes the descriptive statistics of the selected variables.

Table 4. Summary statistics of the Homestead and FullDwelli variables in the Story County data

Fieldname	Count	Min	Max	Mean	Median	Outliers	Sum	Iqr	Q1	Q3
FullDwelli	43.112	0	10.132.400	87.469	72.100	1.077	3.770.962.644	124.925	0	124.925
Homestead	43.112	0	366.686	1.824	0	42	78.644.608	4850	0	4.850

This dataset was primarily used to explore the use of BBZ with large-scale parcel data comprising many individual records.

4.2. Application on a Dataset from a Single Quarter in Ankara

The final dataset used in this study comes from Ankara (Türkiye). After sufficient experience had been gained from the United States cases, the analysis focused on the Ankara dataset as the primary application for Türkiye. The case study was conducted in a single quarter, consistent with TurkStat's practice of applying the blocking procedure within quarter and village boundaries. The selected quarter is a rapidly developing urban area that has expanded in recent years and is expected to continue to change in the near future. The name of the quarter is not disclosed due to confidentiality considerations, given the relatively small size of the study area.

Several spatial datasets were used in this application, including quarter boundaries, road centerlines, numbering data linking parcels to structures, building data containing information on structures and independent sections, and parcel boundaries. Figure 6 illustrates the study area and the integrated representation of these datasets. Because the datasets were obtained from different sources, minor boundary inconsistencies were observed.

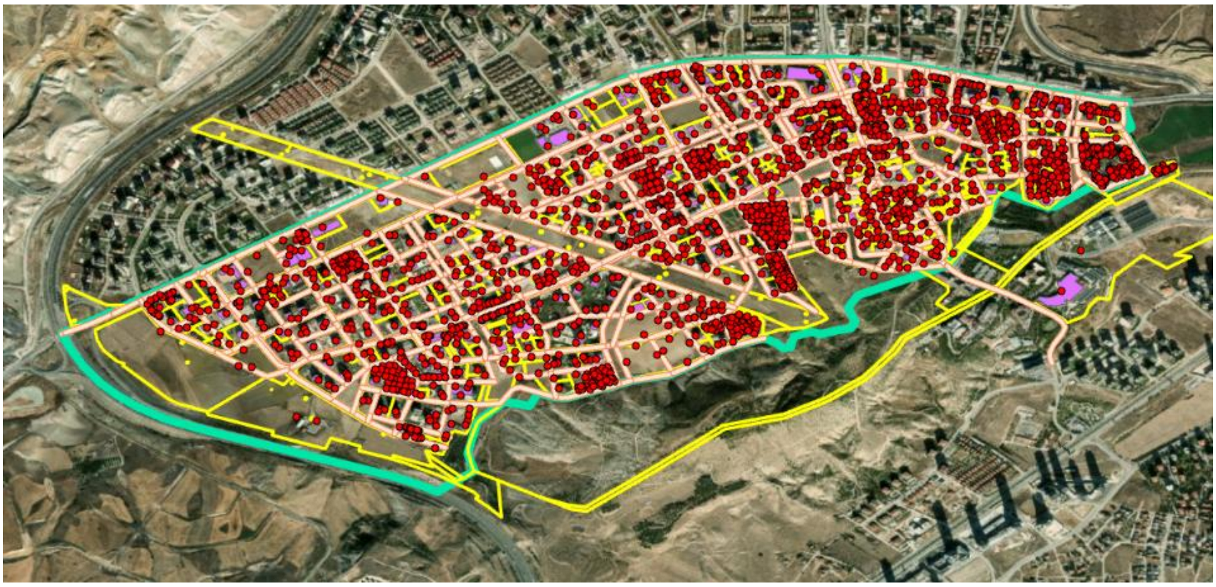

Figure 9. Spatial view of the quarter dataset

For the Ankara application, parcels were grouped into enumeration areas(EAs)using the ArcGIS Pro Build Balanced Zones(BBZ)tool. The primary balancing criterion was the Number of Independent Sections per parcel (total_bb), where independent sections consist of residential units(households), commercial units(workplaces),and public buildings. The total count (total_bb) was used as the main target variable, while the counts of residential, commercial, and public units were also available at parcel level as separate attributes. To support replication and clarify the trial structure, the BBZ configurations tested in this study are summarized as follows. The baseline specification used total_bb as the target variable with a threshold of 120 units per zone. This threshold was informed by TurkStat's household survey practice, in which clusters are typically formed within quarter boundaries, with approximately 100 occupied household addresses and roughly equal cluster sizes. In addition to the baseline, an alternative specification was tested by including household counts as a target variable with a threshold of 100 per zone. Across trials, multiple combinations of thresholds, spatial contiguity settings, zone-shape preferences, and advanced optimization parameters were explored, together with optional distance-based constraints derived from quarter borders and the road network.

Tested BBZ parameter set:

(1) Input features: parcel polygons.

(2) Primary balancing variable (target attribute): total_bb (Number of Independent Sections per parcel).

(3) Target threshold(s): total_bb = 120 (baseline); household count = 100 when used as a secondary target.

(4) Spatial contiguity constraints (adjacency definition): Contiguity (Edges Only); Contiguity (Edges and Corners); and proximity-based adjacency where applicable (e.g., triangulation-based option), with additional constraints implemented using quarter borders and road network layers.

(5) Zone Characteristics (shape/structure preferences): Compactness; Equal Area; Equal Number of Features (tested as alternative settings).

(6) Advanced Parameters (optimization controls): Population Size, Number of Generations, and Mutation Factor (tested under multiple combinations to assess sensitivity and improve convergence).

(7) Distance option (when activated): Distance to Consider, using quarter borders and road centerlines as reference layers.

When enabled, the Distance to Consider option accounts for proximity to selected reference features. For each input parcel, it calculates the distance to the nearest feature in the specified reference layer and uses this distance as an additional constraint when selecting the final zone configuration. In this application, quarter boundaries and road centerlines served as reference layers for the distance parameter.

5. Results and Discussion

The results obtained from both the United States datasets and the Ankara application provide meaningful insights into BBZ's performance and the prospects for geospatial zoning within Türkiye’s statistical system. Each dataset highlighted unique aspects of the methodology, offering lessons on variable suitability, dataset size, parameter sensitivity, and operational feasibility.

The Richmond County dataset included parcel boundaries and land value attributes. Applying the BBZ tool to this dataset revealed the disruptive effects of extreme variable distributions (Figure 7). Some parcels had exceptionally high land values while others had almost none, preventing the algorithm from generating statistically and geographically balanced zones(Figure 8).This outcome highlighted the methodological challenge posed by highly skewed variables and emphasized the importance of selecting indicators that are both representative and statistically stable. In practice, census planners are advised to use variables that directly and consistently reflect population or household characteristics rather than those influenced by irregular market dynamics (Flowerdew & Feng, 2005).

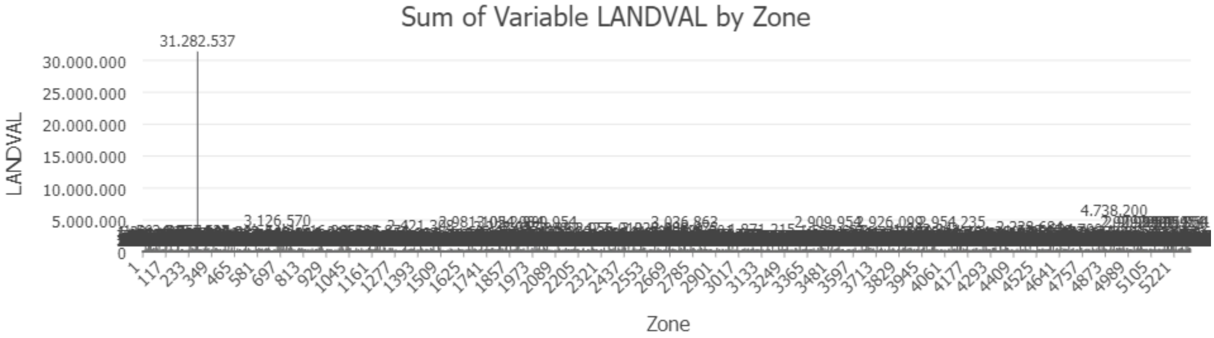

Figure 10. Sum of land value by zone

Figure 11. The result of the BBZ for Richmond County data

The second dataset, from Seattle, used the number of housing units per parcel as the balancing variable. Since most parcels contained only one housing unit, variation was extremely low, limiting the BBZ tool's ability to redistribute features and achieve balance. The Seattle dataset revealed limitations caused by insufficient variability, resulting in zones of limited operational and analytical significance (Figure 10). Various alternatives were tested, but no meaningful results were obtained, as the values were clustered around 1. This finding underscored the importance of selecting input variables that provide sufficient diversity for the algorithm to function effectively (Flowerdew & Feng, 2005). The issue is particularly relevant for Türkiye, where certain quarters may exhibit highly uniform housing patterns. Such conditions necessitate careful preprocessing or the use of alternative variables to ensure meaningful and interpretable outputs (United States Census Bureau, 2020).

The Story County case highlighted how dataset size affects the stability of algorithmic zoning. With over 43,000 parcels, the tool created thousands of fragmented and impractical zones (Figures 9 and 10). The operation of the BBZ tool was time-consuming and produced inconsistent results. Adjusting thresholds did not fully resolve the instability, reinforcing the idea that zoning algorithms require tailored parameterization for large-scale applications.

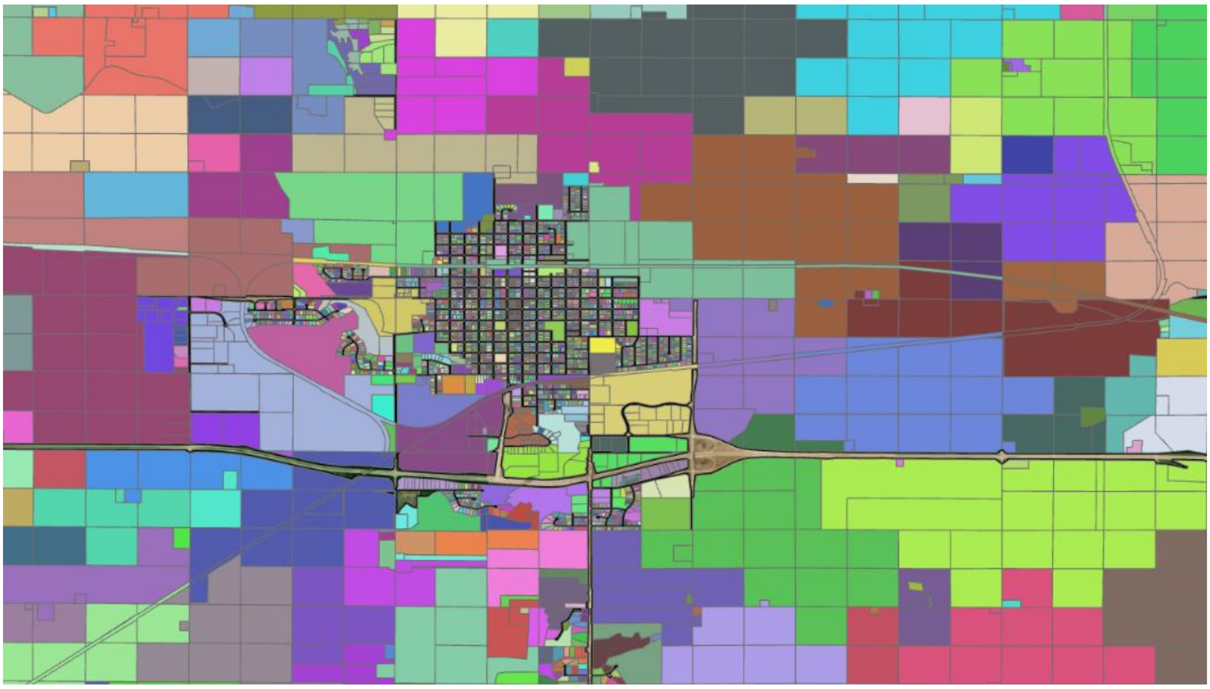

Figure 12. View-1 from BBZ result of Homestead variable

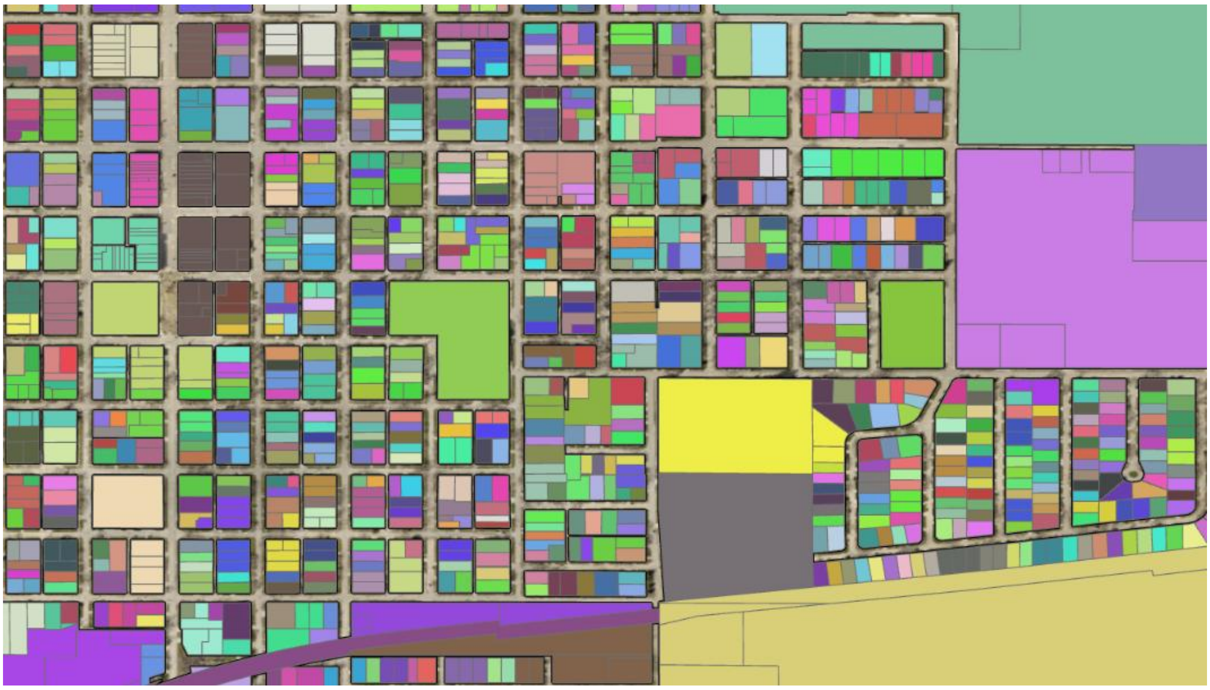

Figure 13. View-2 from BBZ result of Homestead

The influence of threshold selection was also evident in this case. When the threshold for FullDwelli was set to 100.000, approximately 20.000 zones were generated for 43.000 parcels, corresponding to nearly half of the total dataset. Increasing the threshold to 1,000,000 reduced the number of zones to about 3,000, which still represented a relatively high proportion (Figure 11).

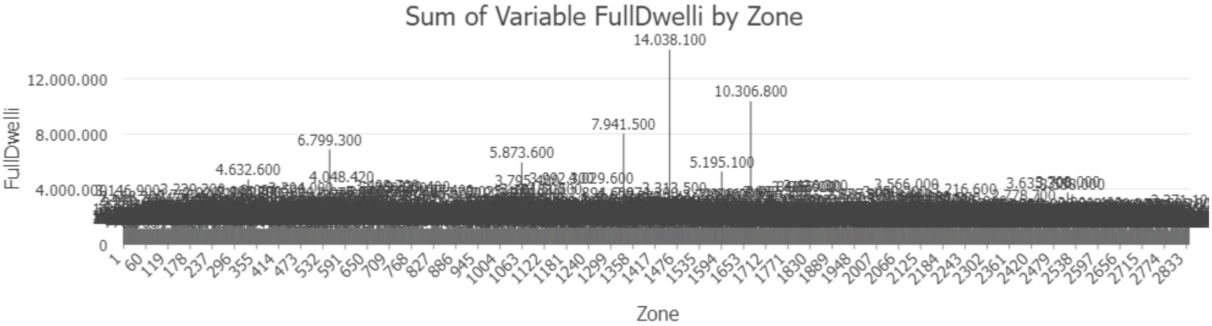

Figure 14. BBZ Result for FullDwelli when threshold is 1 million

These findings demonstrate that the configuration of threshold values critically affects computational performance and spatial aggregation, particularly when working with extensive parcel datasets. This observation aligns with broader international discussions on the Modifiable Areal Unit Problem (Openshaw, 1984).

The Quarter dataset was used to evaluate the applicability of the Build Balanced Zones tool in Türkiye’s spatial and statistical context. During data preparation, minor inconsistencies were detected between different boundary and parcel layers, which created small alignment mismatches. These technical discrepancies highlighted common challenges in spatial data management and emphasized the importance of thorough data cleaning and integration before applying automated zoning tools.

Several spatial constraint configurations were tested to determine their influence on the zoning process. Using only contiguity edges produced compact, geometrically regular zones but offered limited stability in statistical balance. Incorporating both edges and corners increased connectivity between parcels, although this occasionally led to irregular shapes. The Trimmed Delaunay Triangulation option produced proximity-based zones, yet these were not always consistent with defined statistical targets.

The analysis of Zone Characteristics Criteria revealed the trade-offs between geometric fairness and statistical adequacy. The Compactness criterion generated visually regular, contiguous zones, but made balancing the independent sections more difficult. The Equal Area and Equal Number of Features settings provided slight improvements, but did not fully eliminate imbalances, reinforcing the need to balance simultaneously in EA design (Esri, 2023).

Among the Advanced Parameters, increasing the Mutation Factor resulted in the most visible improvements, producing outcomes closer to the desired threshold. Introducing household counts as a secondary balancing variable, with a threshold of 100, enhanced demographic alignment. Incorporating distance constraints based on quarter borders and road networks further improved spatial coherence, producing zones that corresponded more closely to recognizable urban features. This configuration improved the operational feasibility of EAs, enabling field staff to more easily identify and navigate them during data collection.

Overall, the Quarter case demonstrated that the BBZ tool can be effectively adapted to Turkish datasets, given adequate data quality and parameter optimization. The tool produced zones that were both geographically coherent and statistically balanced, confirming its potential for census and survey applications (Figure 12). However, even small inconsistencies in boundaries and spatial layers significantly affected outputs, emphasizing the necessity of standardized data preparation and harmonized spatial frameworks for national-scale implementation.

Figure 15. The final BBZ result for Ankara data

The discussion of zone characteristics also confirmed the multi-objective nature of EA delineation. Compactness enhances geometric regularity but often conflicts with demographic balance, whereas equal-area options ensure visual fairness but lack statistical adequacy. These findings highlight that EA delineation cannot rely on a single optimization criterion but must balance geography, demography, and operational manageability (Esri, 2023; United Nations, 2017).

From a methodological perspective, the Ankara study demonstrated that BBZ is a flexible and scalable tool for EA generation but remains sensitive to parameter settings such as Mutation Factor and Number of Generations. Hybrid approaches that combine BBZ with manual verification or supplementary geospatial tools could offer the most practical balance between automation and expert control.

From a policy standpoint, adopting geospatial methods could strengthen Türkiye's statistical infrastructure by enabling standardized, reproducible, and auditable delineation of primary sampling units and by reducing reliance on ad hoc manual boundary production. Automated delineation, as shown in the Ankara case, offers reproducibility, transparency, and scalability—qualities that align with modernization goals set by the United Nations Statistics Division (United Nations, 2017). Nevertheless, automated tools should complement, not replace, local expertise, as contextual knowledge remains essential for designing meaningful and functional enumeration areas.

In summary, the results highlight both the opportunities and the limitations of using geospatial methods for EA delineation. The United States datasets illustrated methodological risks linked to poor variable selection or dataset imbalance, while the Ankara application demonstrated that, with proper data and parameterization, the method can yield zones that are both statistically meaningful and operationally feasible. For Türkiye, the main challenges lie in harmonizing spatial datasets across institutions and establishing clear guidelines for parameter optimization. With these measures in place, geospatial zoning can significantly enhance the efficiency and quality of survey and census operations.

6. Conclusions

This study proposed a GIS-based approach for delineating Enumeration Areas (EAs) for statistical field operations in Türkiye through a trial implementation designed to test its feasibility under different data conditions. The empirical structure combined three case applications from the United States with one main application, conducted in a single quarter of Ankara (Türkiye). Rather than claiming nationwide applicability, the Ankara-quarter application demonstrated the workflow in a realistic local setting, while the United States cases illustrated how automated zone creation behaves under contrasting input-variable distributions and dataset scales. Taken together, the cases clarify the conditions under which GIS-based delineation can support geographically coherent and operationally workable zones, and they identify limitations that must be addressed before wider institutional implementation.

The case applications in Richmond County, Seattle, and Story County provided a diagnostic foundation for understanding methodological strengths and weaknesses. Across these cases, three practical “screening conditions” emerged that should be checked before applying automated delineation in routine production. First, extreme variation and skewness in balancing variables can dominate the optimization problem and contribute to fragmented or unstable outputs, indicating that selected variables should be both representative and distributionally stable for EA purposes. Second, insufficient variability can mechanically constrain the balancing process because the workflow has limited degrees of freedom to redistribute features, thereby reducing the likelihood of operationally meaningful outputs. Third, very large datasets introduce computational and stability challenges: threshold choices can yield either an excessive number of zones or overly coarse aggregation, and long run times and fragmentation become more likely. These findings reinforce the importance of selecting suitable balancing variables, assessing distributional properties in advance, and ensuring that the intended scale of delineation matches the structure of the input data. They also are consistent with the Modifiable Areal Unit Problem, which emphasizes that scale choices and distributional properties can lead to substantially different spatial outputs (Openshaw, 1984; Flowerdew & Feng, 2005).

Building on these lessons, the Ankara quarter application illustrated how a GIS-based delineation workflow can be configured to reflect field-operational realities. By using independent sections per parcel and household counts as balancing variables and incorporating road networks and quarter borders as spatial constraints, the workflow supported the production of zones that were closer to intended workload targets and easier for field operations to interpret than zones produced by ad hoc manual delineation. This application, therefore, serves as an applied demonstration of how a Türkiye-relevant EA workflow can be parameterized around workload-oriented targets and recognizable geographic features.

At the same time, the Ankara application highlighted a set of critical contradictions that must be discussed with respect to both methodology and application. Inconsistencies between the Spatial Address Registration System and the General Directorate of Land Registry and Cadastre —particularly mismatches in quarter boundaries and parcel alignments—directly affected zone membership and adjacency. Methodologically, these cross-source inconsistencies challenge core assumptions of automated delineation, because the workflow depends on a coherent spatial topology in which parcels, boundaries, and barriers define stable neighbor relations and a consistent spatial support for the balancing variables. When boundary lines do not align across sources or when parcel or building geometries are shifted, the induced contiguity relationships may change, and the same parameter settings can yield different configurations depending on which layer is treated as authoritative. In practical terms, boundary mismatches can reassign parcels at the margins and alter zone totals, while parcel–building misalignment can weaken attribute linkage and introduce local distortions in balancing variables. Accordingly, evaluation in the Turkish context requires reporting not only zoning parameters but also the pre-delineation harmonization decisions and spatial validation steps that define the effective analytical support for the workflow.

The study further confirms that EA delineation is inherently multi-objective. Geometric preferences such as compactness or equal area can improve certain shape properties, but they must be balanced against demographic relevance and operational constraints. More workable configurations were obtained when demographic criteria were combined with geographic and field-operational constraints, consistent with international guidance emphasizing multi-criteria design for census geography (United Nations, 2017). In addition, the findings suggest that reproducibility and transparency are achievable only when workflows are standardized and data integration is treated as a prerequisite rather than a post-processing step (Cockings & Martin, 2005; United States Census Bureau, 2020). For this reason, a core practical output of the study is not only the zoning results identification of the need for clear parameter reporting—at minimum, target variables, threshold values, contiguity rules, optional constraints, and optimization settings—to support replication and diagnosis across quarters and update cycles.

From a policy and implementation perspective, the results indicate that the potential benefits of GIS-based EA delineation—standardization, reproducibility, and operational manageability—depend on institutional coordination and data governance. Establishing shared standards for dataset integration, implementing routine cross-agency validation procedures, and formalizing cooperation mechanisms are central prerequisites for sustained adoption. Such steps are consistent with broader census modernization recommendations that emphasize integrated geospatial infrastructure and documented procedures for boundary management and fieldwork geographies (United Nations Statistics Division, 2017).

The main limitation of this study is its scope. The application in Ankara was conducted over a single quarter, and the United States cases were chosen to illustrate contrasting data patterns rather than to provide national benchmarks. Future work should, therefore, extend the workflow to multiple quarters with different urban forms, test stability under data updates, and evaluate scalability using larger official datasets. A hybrid production model—automated delineation followed by structured expert review—also appears essential for balancing efficiency with operational realism. Additional research may explore dynamic updating mechanisms and complementary data sources to improve adaptability in rapidly changing urban settings (Krebs & MacQueen, 2016).

Overall, the study supports the conclusion that GIS-based EA delineation is a feasible methodological direction for Türkiye when (i) balancing variables have sufficient variability and distributional stability, (ii) contiguity and constraint rules reflect field-operational requirements, and (iii) input layers are harmonized with adequate geometric consistency. Under these conditions, the approach offers a practical foundation for enhancing the production of clearly bounded, operationally workable small-area units for future census and household survey operations, explicitly acknowledging that methodological rigor and data governance are inseparable components of reliable automated delineation.

Notes

This article is derived from the PhD thesis prepared by Cansu ÖZTÜRK at Hacettepe University Institute of Population Studies, Department of Social Research Methodology

Author Contributions

Author Name	Types of Contribution
Cansu Öztürk	Research Design, Data Collection, Analysis, Methodology, Writing, Editing
Ahmet Sinan Türkyılmaz	Research Design, Methodology, Editing

Conflict of Interest

The authors declare that there are no commercial or financial relationships that could be construed as a potential conflict of interest in the preparation and publication of this study.

Referanslar/References

ABS. (2016). Census of Population and Housing: Mesh Block Counts, Australia, 2016. *Australian Bureau of Statistics*. Retrieved from https://www.abs.gov.au

Bethlehem, J., Cobben, F., & Schouten, B. (2011). *Handbook of nonresponse in household surveys. New* Jersey: Wiley and Sons

BPS-Statistics Indonesia. (2018, July 12). BPS Builds Wilkerstat with Geospatial Technology. Retrieved from https://www.bps.go.id/en/news/2018/07/12/201/bps-builds-wilkerstat-with-geospatial-technology.html

Chakravorty, S. (2007). Geography in India's census: A GIS-based approach (ESA/STAT/AC.115/17). United Nations Expert Group Meeting on Contemporary Practices in *Census Mapping and Use of Geographical Information Systems.* Retrieved from https://unstats.un.org/unsd/demographic-social/meetings/2007/egm-census-mapping/docs/esa_stat_ac115_17.pdf

Cockings, S., & Martin, D. (2005). Zone design for environment and health studies using pre-aggregated data. *Social Science & Medicine,* 60(12), 2729–2742. https://doi.org/10.1016/j.socscimed.2004.11.005

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014*). Internet, phone, mail, and mixed-mode surveys: The tailored design method (4th ed.)*. Indianapolis: Wiley.

Esri. (n.d.). How Build Balanced Zones works. Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/learnmore-buildbalancedzones.htm

Esri. (2020). ArcGIS Pro: Spatial Statistics Tools. Redlands, CA: Environmental Systems Research Institute.

Esri. (2021). ArcGIS Pro: Geostatistical Analyst. Redlands, CA: Environmental Systems Research Institute.

Esri. (2023). Build Balanced Zones (Spatial Statistics*)*. ArcGIS Pro 3.2 Documentation. Retrieved from https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/build-balanced-zones.htm

Flowerdew, R., & Feng, Z. (2005). Enumeration districts, output areas and geographical information systems. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 168(1), 49–67.

Goodchild, M. F., & Li, W. (2021). Replication across space and time must be weak in the social and environmental sciences. *Proceedings of the National Academy of Sciences,* 118(35), e2105274118. https://doi.org/10.1073/pnas.2015759118

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology (2nd ed.).* New Jersey: Wiley.

Holt, D., Steel, D. G., Tranmer, M., & Wrigley, N. (2004). Small area estimation and the geography of poverty. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 167(3), 319–341.

INSEE. (2019). IRIS: Ilots Regroupés pour l'Information Statistique. Institut National de la Statistique et des Études

Économiques. Retrieved from https://www.insee.fr

Instituto Brasileiro de Geografia e Estatística (IBGE). (2022). Country report of Brazil: Twelfth session of the United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM). United Nations.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2024, August 22). IBGE releases coverage analysis of the Post-Enumeration Survey of the 2022 Population Census. IBGE News Agency.

Kalton, G. (2020). Developments in survey sampling. *Journal of Official Statistics,* 36(4), 855–879.

Kırlangıçoğlu, C. (2005). A new census geography for Turkey using geographic information systems: A case study on Çankaya District, Ankara (Master's thesis). Middle East Technical University, Ankara, Türkiye.

Kish, L. (1965). Survey sampling. Wiley.

KNBS. (2019). *Kenya Population and Housing Census 2019: Enumeration Areas and Sampling Frame.* Kenya National Bureau of Statistics.

Krebs, C. J., & MacQueen, J. (2016). Spatial sampling. In *Ecological methodology* (pp. 381–403). Harper & Row.

Laaribi, A., & Peters, A. (2019). GIS and the Sustainable Development Goals. United Nations Economic and Social Council Report. United Nations.

Lohr, S. L. (2010). *Sampling: Design and analysis* (2nd ed.). New York: Chapman and Hall/CRC.

Martin, D. (2011). Geography for the 2011 Census in England and Wales. *Population Trends,* 145(1), 7–15.

Mokhele, T., Mutanga, O., & Ahmed, F. (2016). Development of census output areas with AZTool in South Africa. *South African Journal of Science,* 112(7/8). https://doi.org/10.17159/sajs.2016/20150010

Openshaw, S. (1984). *The modifiable areal unit problem*. Geo Books.

ONS. (2020). *Output Area Boundaries (2020) for England and Wales*. Office for National Statistics. Retrieved from https://www.ons.gov.uk

Qader, S. H., Dierkes, C., & Schneider, C. (2019a). Integrating remote sensing and GIS for delineating urban areas. *Journal of Geospatial Information Science*, 22(2), 45–60.

Qader, S. H., Dierkes, C., & Schneider, C. (2019b). Remote sensing-based assessment of land cover change in Baghdad. *International Journal of Remote Sensing,* 40(3), 1120–1145.

PARIS21, & Statistics Sweden. (2021). Guide on geospatial data integration in official statistics. Partnership. in Statistics for Development in the 21st Century (PARIS21).

Pedro, A. A., & Queiroz, A. P. (2019). Slum: Comparing municipal and census basemaps. Habitat International, 83, 30–40. https://doi.org/10.1016/j.habitatint.2018.11.001

Statistics Canada. (2016). Census dictionary: Census year 2016. Statistics Canada. Retrieved from https://www.statcan.gc.ca

Statistics South Africa. (n.d.). Small area statistics. Retrieved from https://www.statssa.gov.za/?page_id=4086

UNECE. (2015). *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing*. United Nations Economic Commission for Europe.

United Nations, Department of Economic and Social Affairs, Statistics Division. (2008). Handbook on geographic databases and census mapping (Draft, 7 March 2008). United Nations.

United Nations. (2009). Handbook on geospatial infrastructure in support of census activities. New York: United Nations Statistics Division.

United Nations. (2017). Principles and recommendations for population and housing censuses (Rev. 3). New York: United Nations Statistics Division.

United Nations. (2021). Handbook on the management of population and housing censuses, Revision 2 (Studies in Methods, Series F No. 83/Rev.2). New York, NY: United Nations.

United States Census Bureau. (2019). Geographic Areas Reference Manual. Washington, DC: U.S. Government Printing Office.

United States Census Bureau. (2020). Census Enumeration Areas and Geographic Concepts. Washington, DC: U.S. Government Printing Office.

United States Census Bureau. (2021). Geographic Areas Reference Manual Update. Washington, DC: U.S. Government Printing Office.

Wanyoike, H. (2023). Use of geospatial information in census-taking: 2019 Census Kenya experience. United Nations Statistics Division Expert Group Meeting on the Use of Geospatial Information in Census-taking. Retrieved from https://unstats.un.org/unsd/demographic-social/meetings/2023/egm-20230523/docs/s03-03-KEN.pdf

Worldometer. (2025). Indonesia population (2025). Retrieved from https://www.worldometers.info/world-population/indonesia-population/