The challenges of producing small areas statistics for public policies

Combine multiple and small geographies with statistical data guaranteeing confidentiality

Claudio Stenner

Director of Geosciences

11th UN-GGIM Session - Side Event - Innovations in Integrating Geospatial and Statistical Information for the SDGs, COVID-19 and National Development – August 18, 2021

The challenges of producing small areas statistics for public policies

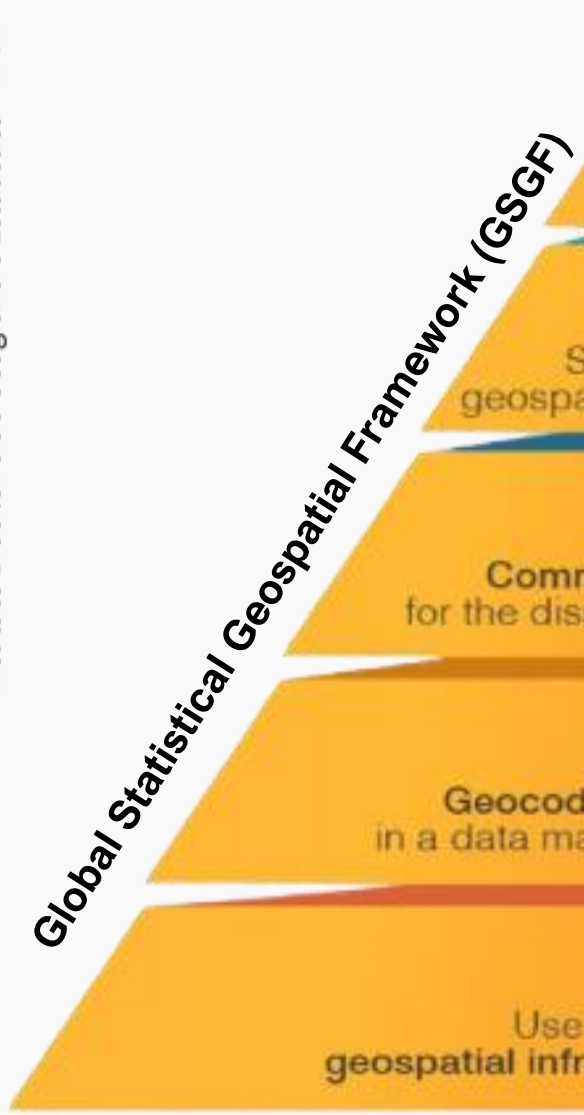
"**Leave no one behind**" is central to the success SDG implementation. To "**Leave no one behind**", it is necessary to produce and make available statistics for small areas. **But…**

There are some **challenges** to achieving this goal:

- Produce georeferenced statistics, preferably by coordinates or addresses;
- Develop methodologies to deal with positional errors;
- Have a broad set of common geographies to meet different information needs;
- Create tools that allow the user to add data to their geographic area of interest;
- Ensure statistical confidentiality;

The GSGF and its 5 principles in Brazil

Brazilian application of GSGF



The link between the main statistical system with the Geographic Framework is done. There are also a Interactive Geographic Platform and a web aplication that allow users to customize and Interact in maps with most of IBGE statistical data.

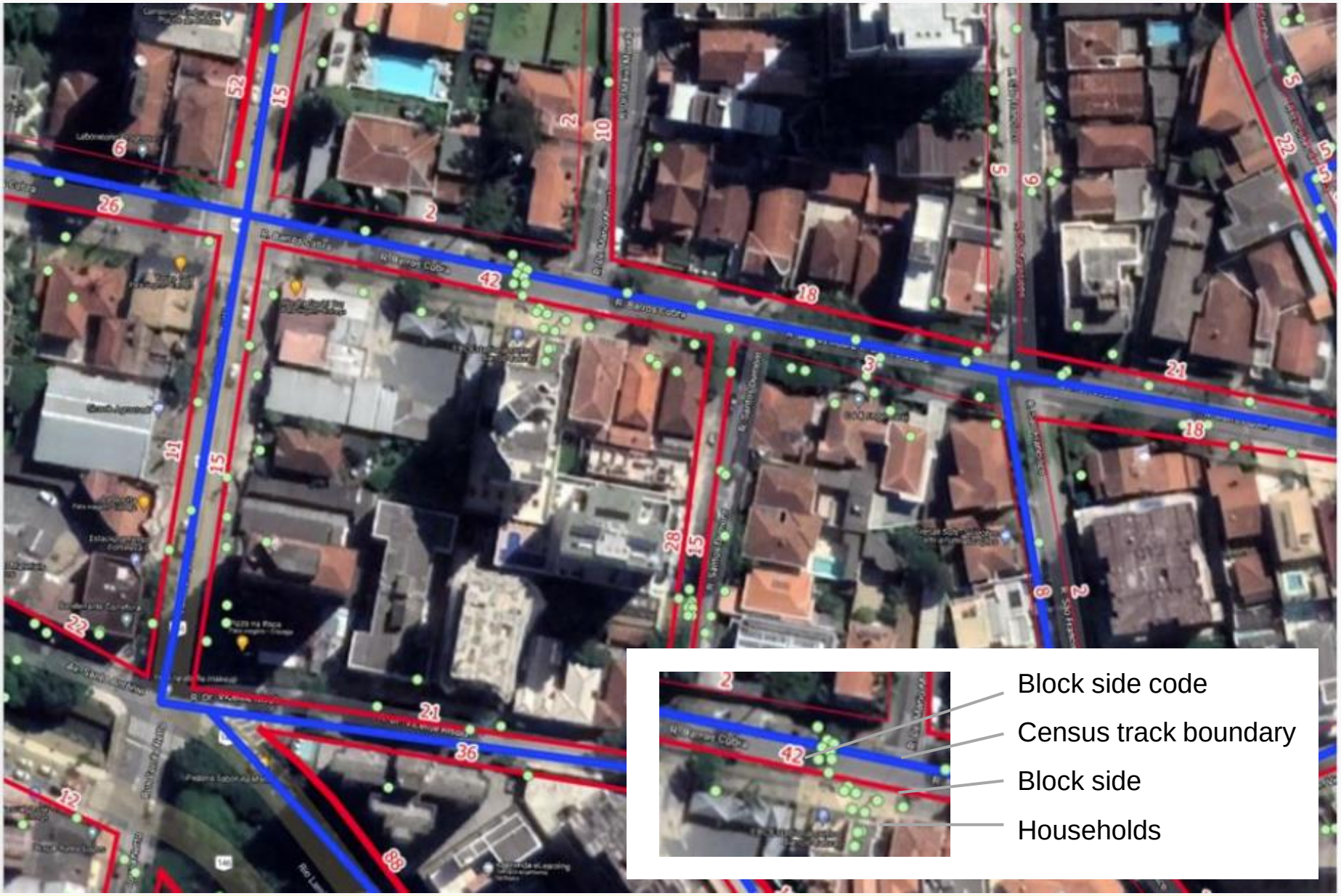
IBGE use very most of international statistical and geospatial standards, both in data and metadata

Geographic Reference Framework for the Production, Analysis and Dissemination of Statistics published. Work in progress on automating the update of the geographic framework in statistical dissemination systems

Statistical data management, geocodes and coordinates.

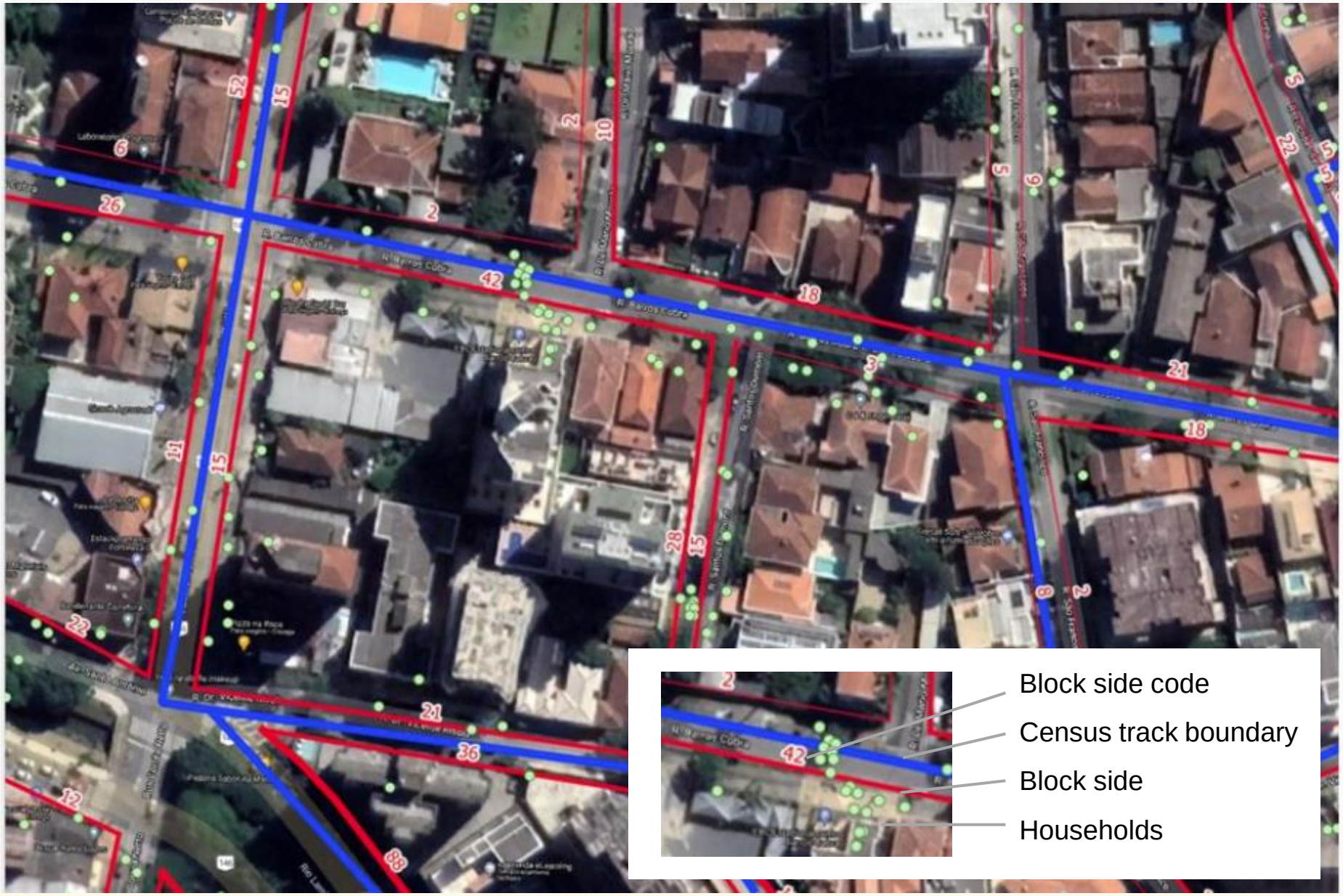
BET – Territorial Structure Database & CNEFE - National Address File for Statistical Purposes

Produce georeferenced statistics



Portable devices with GNSS make it relatively simple to collect home coordinates in the field, as shown in the image. This fact has an enormous potential for the production and dissemination of statistical information in the most varied types of geographic areas, contributing to not leaving anyone out.

Develop methodologies to deal with positional errors



Green dots represent raw GNSS household location data collected in the field. The coordinate shift is visible in the image. Later office work can correct this position by associating the households with the location of the corresponding block side. But it is probably that some coordinate errors will remain. IBGE is working on methodologies to deal with this problem.

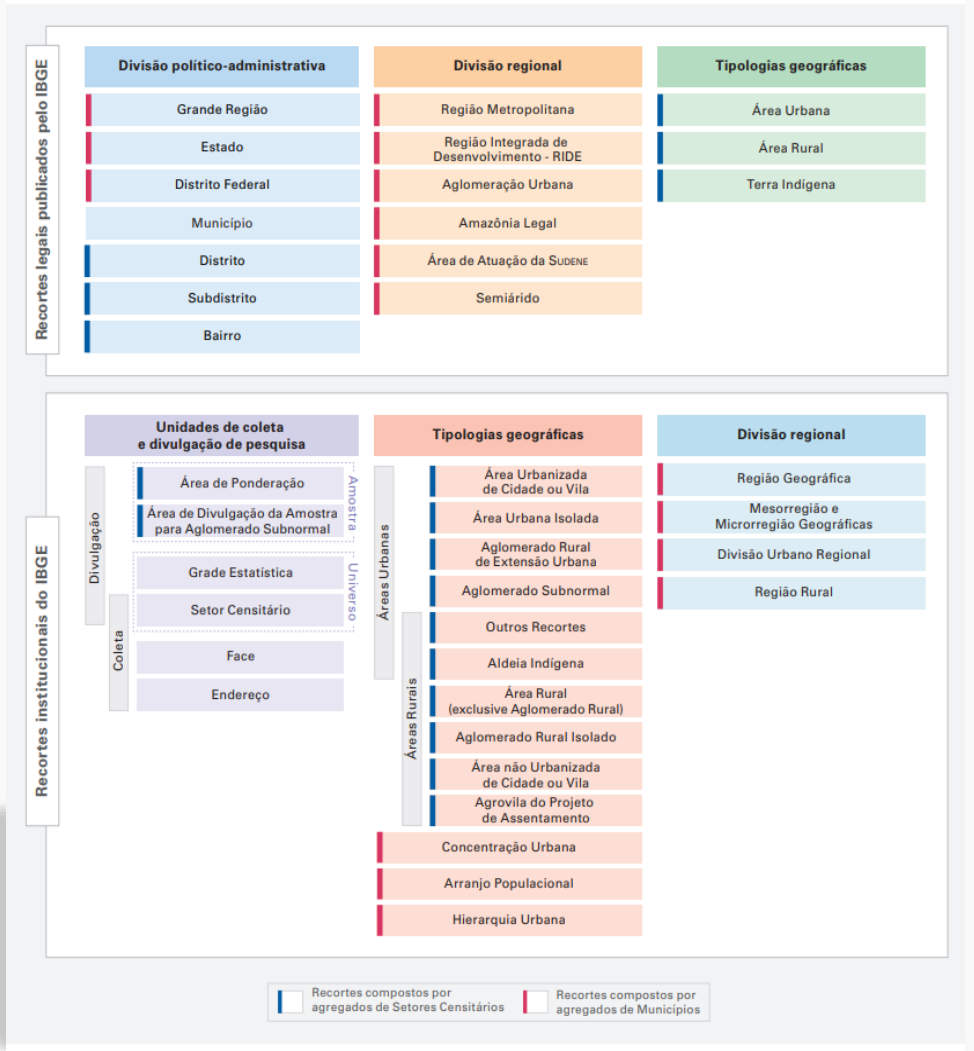
Have a broad set of common geographies

Brazilian Geographic Framework

The **Geographical Framework** is a unique and national reference for relating **statistical information** with the corresponding **geographies**. Provides users a coherent spatial framework for visualizing, analyzing and understanding statistics in a comparable and spatially integrated way. ”

Source: Adaptated from IBGE (2019), *Quadro Geográfico de Referência para Produção, Análise e Disseminação de Estatísticas*

Have a broad set of common geographies
Brazilian Geographic Framework



Have a broad set of common geographies
Brazilian Geographic Framework

- New geographies are being included in the Geographic Framework, including environmental geographies, such as biomes and watersheds
- Large area geographies do not represent new methodological challenges;
- Small area geographies can present challenges regarding statistical accuracy and put statistical confidentiality at risk, but they are very useful to the society.
- The challenge increases when considering the possibility of the user obtaining statistics totaled by custom geographies.

The GSGF and its 5 principles

How to ensure easy access to small areas statistics while preserving the confidentiality?

MAIN CHALLENGES

How to combine the statistical data to the smallest spatial unit possible without violating confidentiality and data privacy?

How to preserve confidentiality and data privacy combining different geographies?

How to guarantee the accuracy of statistics in small areas, taking into account the possibility of positioning errors?

How to disseminate small areas data and ensure the confidentiality?

Some possible paths….

- “blur” techniques to show information on maps of areas that do not meet statistical confidentiality criteria. In this case, full statistical data will not be made available;
- Introduce a random variation in the statistics of small areas, in order to prevent the informant's privacy from being exposed;
- Build a set of minimal geographic units;
- A combination of methodologies?

Have a broad set of common geographies
Build a set of Minimal Geographic Units (MGU)

WORK IN PROGRESS

- **Minimal Geographic Unit (MGU)** is the **fundamental building block** for any aggregation in larger areas. In demographic censuses in Brazil, this "brick" was, until last 2010, the census tract , that has about 300 households.

- The rules that guarantee **statistical confidentiality** are applied to each census tract.

- Brazil is divided into 450,000 census tract This geographic unit, however, proves **to be large** when statistics for very small areas are desired. The "shape" of census tract may not be compatible with other geographies.

A sample of census tract and household location

Have a broad set of common geographies
Build a set of Minimal Geographic Units (MGU)

WORK IN PROGRESS

- Each **MGU** will be subject to **confidentiality criteria**. For some basic Census data, it is enough that the **MGU** is composed of more than 5 households and more than 20 people. This criterion potentially makes it possible to divide the country into **10 millions MGU**, significantly improving the capacity to generate statistics for small areas. The production of **MGUs** can be done automatically, by geoprocessing;

- The **MGUs** preserve the **"territory semantics"**.

- The **MGUs** are **subsets** of the census tract;

- All other Census disclosure geographies will be **derived from the aggregation** of **MGUs**. As a result, it is not possible to broken statistical confidentiality by spatial algebra operations.

A sample of census tract and household location.

Have a broad set of common geographies
Build a set of Minimal Geographic Units (MGU)

- The **MGU** has a **variable geometry**, according to the analyzed variable. Basic information, without crossover, such as total population, allows for smaller MGUs. Information derived from variable crossings needs bigger MGUs to ensure confidentiality.

- **MGUs** can be the basis of web application systems that allow the user to extract statistical information with **custom geographies**.

- In some special cases is possible to combine MGU approach with other approach, like **random statistical variation.**

A sample of census tract and household location.

Final Considerations

- Increased demand for small area statistics;
- Improved geocoded statistical information;
- Possibility of WEB applications that allow access to statistical data in multiple geographies, including custom geographies;
- Needs for new methodologies that guarantee access to statistical data from small areas and preserve confidentiality;
- **Minimal Geographic Units (MGUs)** and random statistical variation are two approaches that have great potential to achieve this goal.

Thank you very much!