



Economic and Social Council

Distr.: General
13 July 2015

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Seventeenth Meeting

Geneva, 30 September to 2 October 2015

Item 5 of the provisional agenda

Grid statistics: opportunities and challenges

Innovative products for disseminating geographical information (part II)

Note by National Statistics Institute, Spain

Summary

The possibilities raised by having geo-referenced information in the 2011 Census are vast. During the last months, we have continued to develop new products that use this information and meet the demands detected from users. This paper, which is the second part of the one presented last year, contains the details of the most recent developments.

Moreover, some aspects that were not mentioned in the first part, like the confidentiality rules implemented and the algorithm that divides the territory into cells of up to 50 meters each side depending on the population sample size are also explained in this paper.

I. Introduction

1. Achievements made in all the sectors during the recent years by the geo-referenced information is beyond doubt. It has become more and more popular to find systems that store users' coordinates and present information in maps for a more intuitive interpretation.
2. Censuses, as a statistical operation that provides very detailed information in the territory, could not pass up the opportunity to use geographical information. Some examples that demonstrate the importance of having geo-referenced information available in the Census context:
 - A supermarket chain may choose the location of a new store.
 - A local authority can decide the most suitable place to build a school or an airport.
 - It is possible to know details about the population living within a certain radius away.

II. New geographical products released

3. In the paper presented last year¹ a set of products to consult geographical information such as maps, predefined tables, microdata files, etc. related to the Census were mentioned. All of them are visible in Census 2011 microsite²
4. During this year, we continued developing new products and also took into account the opinion expressed by users to modify the functionality of the existing ones.
5. For example, rules regarding confidentiality of information were relaxed in order to provide the users with more amount of information. More details about the new confidentiality rules can be found in section 4 of this paper.

A. Geographical thematic maps

6. In the last year, we have included in our Census dissemination system the possibility of viewing 7 geographical maps at different levels.
7. Users can directly select on the map the geographical area they want to analyse. Depending on the zoom level, it is possible to see 3 different maps:
 - Cells of 100 km each side (10.000 km²)
 - Cells of 10 km each side (100 km²)
 - Cells of 1km each side (1km²)

These are the 7 maps that have been published at grid level:

- Total population
- Average age of people
- Percentage of foreign nationals

¹

http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2014/mtg1/WP22_s5_Spain.pdf

² http://www.ine.es/en/censos2011_datos/cen11_datos_inicio_en.htm

- Total dwellings
- Percentage of seasonal or empty dwellings
- Total households
- Average size of household

Figure 1
Percentage of foreign nationals per grid



B. 25 predefined 1 km² maps

8. At the beginning of 2015, the only printed publication of the Census has seen the light, a small book called “What is Spain like? 25 maps to discover it, km² by km²”.

Figure 2

Cover of the publication



9. This publication contains 25 maps that were compiled with information about the most interesting Census variables:

(a) People (14 maps): Total population, percentage of foreign nationals, predominant foreign nationality, average age, percentage of population aged 65 or over, percentage of population aged 25-34 that live with their parents, percentage of people that live in a different place from that in which they were born, average year of arrival to Spain, average level of studies, percentage of population with high level of studies, average number of children, average time taken to place of work or to place of study and percentage of population that spend 14 nights or more in another municipality or country.

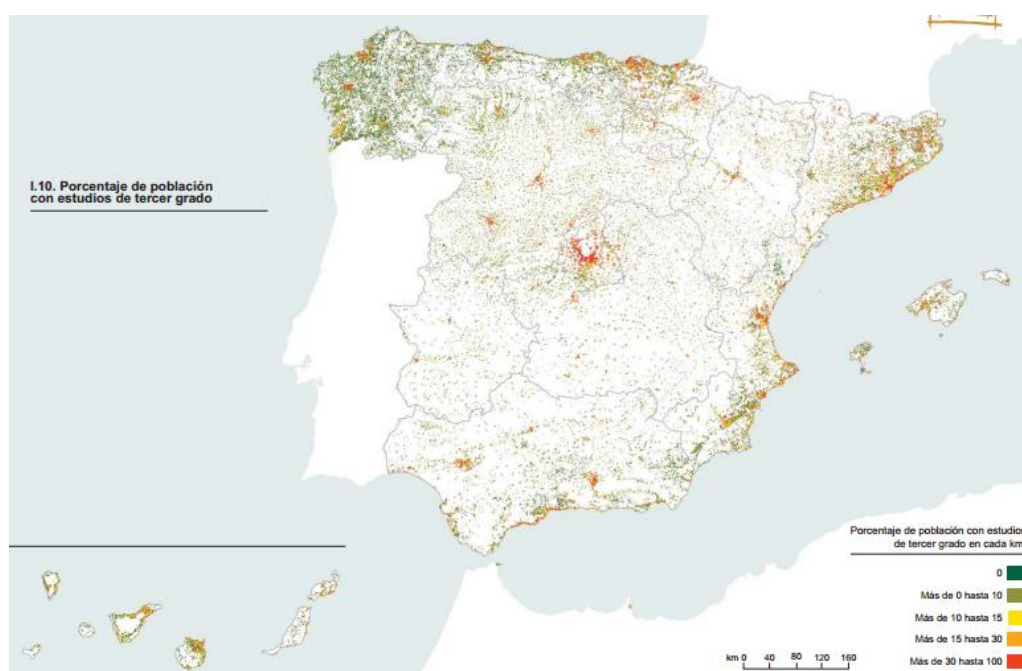
(b) Households (3 maps): Average size, percentage of single-person households, percentage of households with two or more generations

(c) Dwellings (8 maps): Percentage of secondary or empty dwellings, average useful floor space, percentage of dwellings with Internet connection, percentage of dwellings with heating system, percentage of rented dwellings, percentage of dwellings with a mortgage and average period of construction.

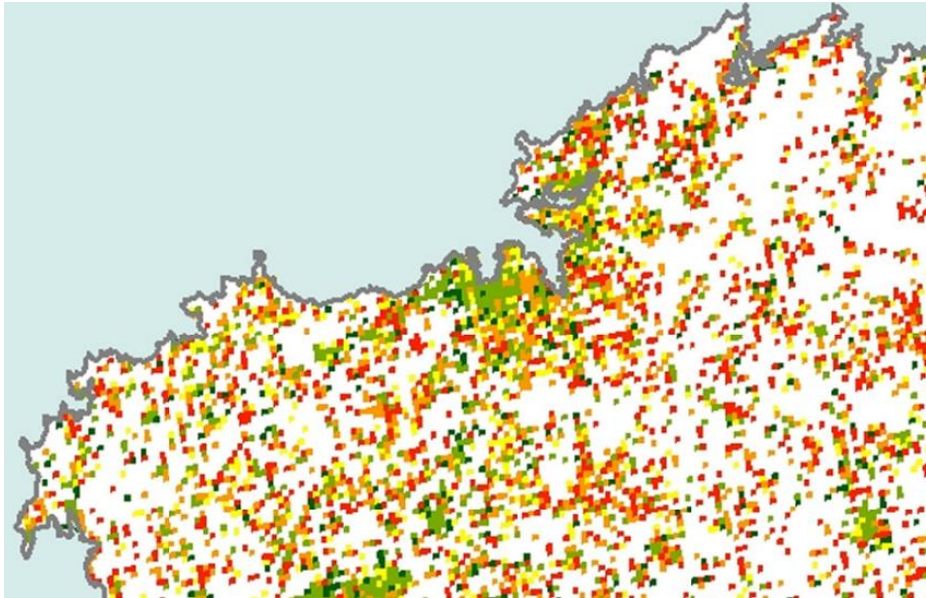
10. These maps reflect information from a different and innovative point of view (all the cells have the same size of 1 km²) and independent from the administrative one, that people are used to. In most occasions, and especially in big cities, 1km² maps contain more detailed information and show better the reality than the municipal maps.

Figure 3

An example of predefined maps (% people with tertiary education)



11. One of the most important advantages of these maps is the possibility of seeing detailed data at a glance. Furthermore, this publication is recommended for all types of users, not necessarily those ones most familiarized with Census information. Comparison of data among different parts of Spain is very simple.



III. Cells smaller than 1 km²: Algorithm of partitioning

12. The 1 km² grid used by INE-Spain for the dissemination of geographical information is the one established by Eurostat³. According to our figures, there are 79,857 cells with at least one dwelling and 63,527 of those cells have people living inside them.

13. INE's census site allows users to consult information of 100 km², 10 km² and 1 km² cells.

14. This grid can be insufficient, particularly in urban areas. It should be reminded that this geo-referenced information comes from a survey that was one of the elements of the census operation in Spain. Around 9% of the population was surveyed. In those cells with more sampling units (persons) we carried out additional partitions (of up to 50 meters each side) depending on the population sample size of the cells.

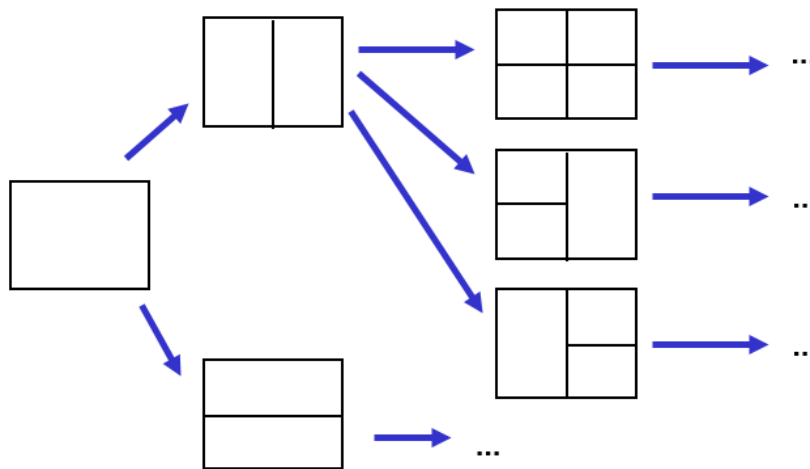
15. Users can view the partitions that have been made in the following link (clicking the option "Unlimited navigation"), <http://www.ine.es/censos2011/visor/>

16. The 1 km² cell partitioning algorithm has three distinct situations:

- Situation 1: Amount of sampling units in the 1 km² cell is below 40. In this case, no additional partition of the cell is carried out.
- Situation 2: Amount of sampling units in the 1 km² cell is equal to or greater than 40 and below 250. In this case, recursively half partitions (vertical and horizontal) that remain at least 20 sample units on both sides of the division will be applied.

³ More information about the grid established by Eurostat can be obtained in the following link: http://www.efgs.info/data/eurogrid/Grid_ETRS89_LAEA_ES_1K.zip/at_download/file

Figure 5
Cell division mechanism



- **Situation 3.** Amount of sampling units in the 1km² cell is equal to or greater than 250. In this case, first of all, the 1km² cell is divided into 400 subcells of 50 meters each side. Then the parameter “p” is calculated:

$$p = 20 * \left[\left(\sqrt{\frac{\text{sampling_units}}{20}} \right) + 1 \right]$$

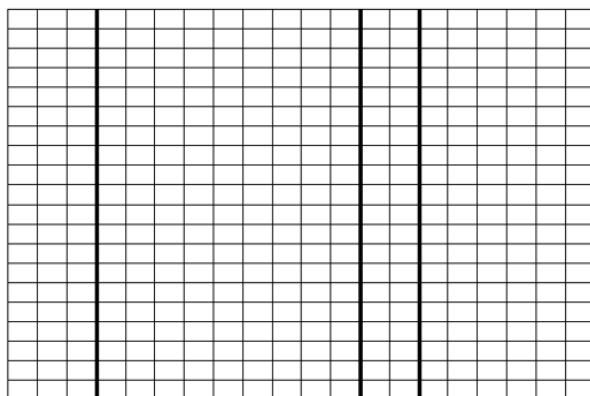
17. The 1 km² cell with the highest density in Spain⁴ contained 3,209 sampling units (52,845 inhabitants approximately) and was divided into 273 sub-cells.

18. Finally, the following three tasks are performed:

- (a) Draw as many vertical lines as possible that leave at least “p” sample units to the left and right side of each division.

Figure 6

First step (n=500, p=120). 4 divisions

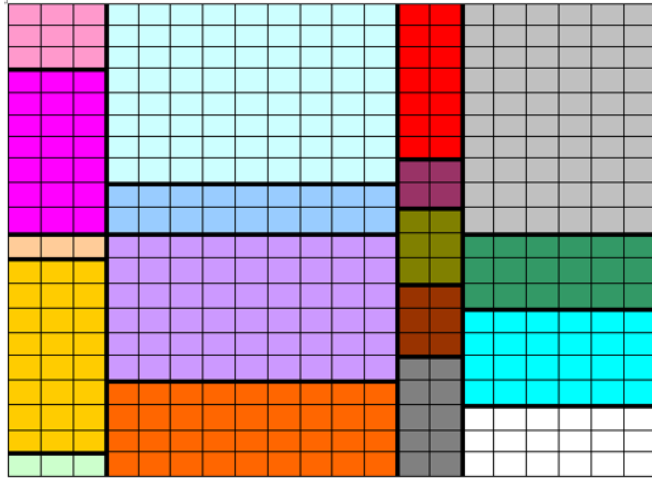


⁴ Situated in L'Hospitalet del Llobregat (Barcelona)

(b) Draw as many horizontal lines as possible that leave at least “20” sample units to the top and bottom of each division.

Figure 7

Second step. 18 divisions



(c) Draw as many vertical lines as possible that leave at least “20” sample units to the left and right side of each division.

Figure 8

Third step. 20 divisions

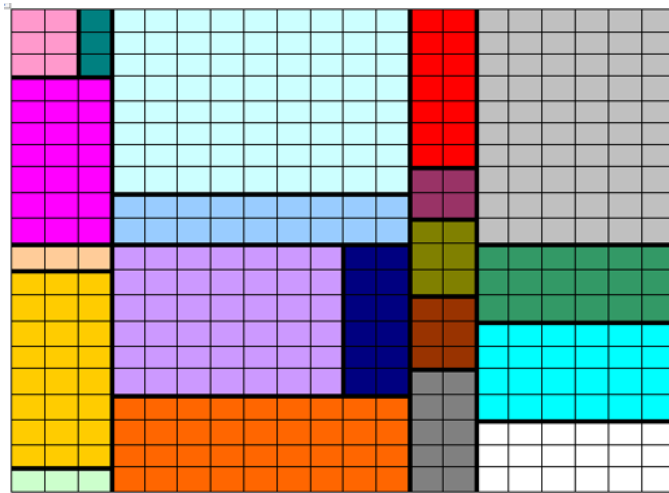


Figure 9
Example of the algorithm in the city of Barcelona



IV. Treatment of confidentiality

19. One of the most important aspects in the dissemination of geo-referenced information is the treatment that is done regarding the confidentiality of the data.

20. INE-Spain applied the same measures regarding confidentiality in geographical products than in other types of products. Information was disseminated when one of these two conditions were satisfied:

(a) When the quotient between the number of existing sampling units and the number of categories⁵ involved in the query is greater than or equal to 5, the total information of the query is disseminated.

(b) Otherwise, when there are at least five sampling units in any of the breakdowns involved in the query, the information of those breakdowns that satisfy the condition mentioned above is disseminated.

21. This can be clarified with the following example of how this rule works. Imagine that a user wants information from an area that contains 20 sampling units about legal marital status (5 categories).

(a) First of all, the quotient between 20 and 5 will be calculated. This result is 4, which is not greater than 5, so the total information of the query is not disseminated.

(b) Secondly, imagine that those 20 sampling units are distributed as follows: 7 single people, 8 married, 2 widow and 3 divorced. In this situation, information about

⁵ For some sensible variables, there is a “sensitivity factor” greater than one, and the condition that has to be satisfied is that the quotient between the number of existing sampling units and the product of the number of categories and the sensitivity factor of the variables involved in the query should be greater than or equal to 5.

single and married people will be disseminated because the amount of sampling units of those breakdowns are greater than 5.

22. Moreover, in order to try to preserve the confidentiality of the information disseminated and to prevent data disclosure by means of differences, all data is rounded to the closest integer multiple of 5 (for example, if the result of a query is 422 then the number disseminated will be 420).

23. Finally, concerning thematic maps with geographical information, they have always been developed with interval ranges inside the legend so that individual information is never unveiled.

24. All the measures carried out to protect information guarantee that, despite very detailed geographical levels are involved, such as the 1km², individual information is never revealed.

V. Conclusion

25. The variety of geo-referenced information that provide a census is very large. As opposed to administrative level, very heterogeneous in surface and where homogenous comparisons are difficult to carry out, using geo-referenced information in INE's census system it is possible to compare units that always have the same surface or to design regions for analysis in an almost unlimited way.

26. Moreover, it is also a very interesting tool to analyse the situation in borderlands inside or among countries.

27. Some institutions have taken note of the power of this information⁶. Eurostat is trying to encourage the countries to collect this information (although it cannot be included in the 2021 Census regulation).

28. For all these reasons the number of countries which plan to disseminate geographical information in the next Census is very likely to increase largely.

⁶ For example, it is planned to be used as data source in order to distribute the budget of European funds. Furthermore, other institutions such as EFGS <http://www.efgs.info/> have spent years promoting the development of grid-based statistics.