Cluster Analysis of the Census Sections in the Basque Country

2010

Eustat
EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADÍSTICA
www.eustat.es

Organismo Autónomo del
EUSKO JAURLARITZA
GOBIERNO VASCO

Presentation

In this report we offer the results of the Cluster Analysis of the Census Sections of the Basque Country for 2010. The administrative sections of our territory are grouped in a series of homogenous clusters or classes, with their corresponding description.

Similar analyses have previously been carried out in Eustat. In this edition emphasis has been put on representing on the territory the statistical synthesis that means the clustering or grouping of the sections, according to their population and housing characteristics.

In order to visualise it, a layer has been created that is represented in a geographic system and contains the information obtained from the analysis. This allows knowledge to be generated from statistical data.

In this investigation, individual, family and housing characteristics are studied, based on Sociodemographic Database 2010 statistical operation. As in previous studies (1991, 1996, 2001 y 2006), the purpose of the analysis is to group the census sections that bring together common features that distinguish them from other clusters.

These analyses help in terms of the description and understanding of our overall population in our geographic areas. They are also applied to other fields, such as town planning, or statistical methodology and sample design stratification, for example.

I hope that this publication is useful for all those interested in this field of statistics.

Vitoria-Gasteiz, February 2013

JOSU IRADI ARRIETA

General Director

Contents

Chapter
1

Introduction

The object of this study is to develop a methodology for creating a typology of the census sections of the Basque Country and to use it in accordance with current data. This methodology will be used to obtain a grouping or classification of the sections (that is, a typology).

The sections are described in terms of various characteristics and groupings are made by considering sections with similar characteristics. In this way, sections of the same cluster will have similar descriptions, whereas sections belonging to different clusters will differ in certain characteristics.

The groups of sections in the typology developed in the study are defined in two ways:

- *By their extension*, providing a list of sections belonging to a given cluster, or even indicating, on a geographical map, which sections they are and where they are.

- *By their content*, revealing the important characteristics in the description of each cluster. The description of a cluster corresponds to the approximate description of the sections it includes.

When such a result is achieved it provides new information, as new relations between sections are established and a synthetic perspective of the provinces of the Basque Country is developed. This new perspective will subsequently serve other functions, such as for gaining a more exhaustive knowledge of the socio-economic situation in society, for designing strategic sampling or for the planning of cities and municipalities.

Chapter
2

Methodology

12 characteristics or variables describe each section, of which the majority are qualitative and some quantitative. The variables will be defined in the following sections, along with their corresponding modalities.

In order to complete the description of a section, the Database of Socio Demographic statistics (BSD) of the year 2010 is to be taken as a reference. The aim of the BSD is to collect, integrate and produce detailed individual information on issues of population, activity, level of education, level of Basque and housing on an annual basis, in order to update the information on the Population Register. The analysis has taken into consideration the occupied family dwellings of each section, and the main characteristics of their residents. The description of each section is based on the frequency of the characteristics of its constituent dwellings and residents.

There are a number of techniques for conducting an integrated study of the sections based on multiple variables. This study has used two techniques in succession:

- *Multiple factor analysis,* for converting the information derived from multiple variables into a reduced number of new or synthetic variables (in other words, factors).

- *Automatic classification* or cluster, for developing a typology of sections by means of the factors obtained.

Factor analysis is essential for initiating automatic classification. Through this analysis, the information is summarised by creating new uncorrelated variables, random "noisy" information is excluded by selecting primary information, and computational complexity is reduced for classification.

A sections typology is defined as a consequence of classification. The main characteristics of each cluster reveal an interpretation of the condition of the sections. A characteristic of a cluster is considered a main characteristic when the modalities of said characteristic differ considerably within the cluster, in comparison with other clusters.

In the description of a cluster, the importance of each characteristic has been determined according to *Pearson's* X^2 statistical test. To describe a cluster, the graphical behaviour of the cluster's characteristics are first represented, showing the order as mentioned and important aspects.

Although they are not shown in this document, each section is described by means of graphs. These graphs indicate the cluster each section belongs to, as well as the ordering of characteristics according to their significance in relation to the rest, as was performed with the clusters.

All of these calculations and graphs were obtained using the ***R*** programming language.

As a result of all of the above, a geographical map of the sections was obtained (See http://www.eustat.es/indic/idioma_i/indicadoresgraficos.asp). This is the most intuitive result. In the map of the Basque Country, the sections of each cluster are shown by colour. The definition of clusters allows for a complete socio-economic understanding of

the area. The colours were selected according to the location of the clusters in relation to the two principal factors, in such a way that similarities in colour coincide with similarities in the clusters' corresponding profiles (this will explained in more detail later). The software tool *Google Earth* was used to visualise the projection of the clusters on the map.

Data

A total of 2,167,381 residents are included in the file produced with the information from the BSD of 2010 and, according to the definition of the Basque Country, they are grouped into 1,731 sections. A code has been assigned to each section, so that when broken down the sections can be situated on a map.

In fact, each resident belongs to a sex, age group, level of education and level of Basque, receives an income and lives in a type of dwelling. The description of each section represents the frequency distribution of its residents, in the case of qualitative variables, and their average, in the case of quantitative variables. Certain quantitative variables (such as age) have been qualified or categorised and have therefore been described by means of a frequency distribution in each section.

The observed variables are as follows:

i. *Place of birth (**place_birth**)*
ii. *Age and sex (**age**)*
iii. *Level of education (**level_educ**)*
iv. *Relationship to activity by sex (**activity**)*
v. *Profession (**profession**)*
vi. *Surface area of the dwelling (**area_dwelling**)*
vii. *Family size (**fam_size**)*
viii. *Year of construction of the dwelling (**year_constr_dwelling**)*
ix. *Level of Basque (**level_basque**)*
x. *Mother tongue (**mother_tongue**)*
xi. *Family income (**fam_income**)*
xii. *Personal income (**per_income**)*

The following sections details their modalities. All variables are qualitative, with the exception of the last two.

Given that these two variables (*Family income* and *Personal income*) are lineally correlated and were not active variables in the analysis, as they lack the required level of desegregation, a new qualitative variable has been defined based on the variable *Family Income* that facilitates the final description of the clusters. For this reason, the range of the variable has been divided into six fractions, in such a way that similar frequencies have been imposed on them.

The modalities of the original variables are as follows:

i. Place of birth (***place_birth***): This refers to the place of birth of the surveyed individual.

basque_c: Born in the Basque Country.
e_country: Born elsewhere in the country (excluding the Basque Country).
outside: Born outside the country.

ii. ***Age and sex (age):*** This refers to the age, in number of full years, of a person of a specific sex on a specific date. In other words, the age that a person of a specific sex "reached" on their last birthday.

a_16_m: Men under the age of 16.
a16_24_m: Men between the ages of 16 and 24.
a25_44_m: Men between the ages of 25 and 44.
a45_64_m: Men between the ages of 45 and 64.
a_64_m: Men aged 65 and over.
a_16_w: Women under the age of 16.
a16_24_w: Women between the ages of 16 and 24.
a25_44_w: Women between the ages of 25 and 44.
a45_64_w: Women between the ages of 45 and 64.
a_64_w: Women aged 65 and over.

iii. ***Education level (level_educ):*** This is the highest level of completed education attained by an individual, possibly certified by a qualification.

prima: Primary and no formal education.
vocation: Vocational training.
secon: Secondary:
med-high: Medium-Higher.
higher: Higher.

iv. ***Relationship to activity by sex (activity).*** Activity carried out by a person of a specific sex at a time of reference.

employ_m: Employed men.
employ_w: Employed women.
unempl_m: Unemployed men.
unempl_w: Unemployed women.
inact_m: Inactive men.
inact_w: Inactive women.

v. ***Profession (profession):*** This refers to the job held at that date by an employed individual, without regard to the branch of economic activity and the professional status (such as businessperson, employee, etc.) in which the individual should be classified.

tec_prof: Technical professional.
sup_tec: Support technician.
adm_wor: Administrative worker.
shopk_wt: Shopkeeper, waiter.
far_fish: Farmer, fisherman.
q_worker: Qualified worker.
mach_ope: Machine operator.
uns_work: Unskilled worker.

vi. ***Useable floor space of the dwelling (area_dwelling).*** This refers exclusively to the dwelling and corresponds to the number of "useable" square metres of space intended for such a use and limited by the inner face of its exterior enclosure or of the wall that separates it from other dwellings or premises of the same building. That is, not including the communal areas of the building, the thickness of hollow walls, etc.

a_60: Dwellings of less than 60 m^2.
a61_90: Dwellings of between 61 and 90 m^2.
a91_120: Dwellings of between 91 and 120 m^2.
a121_150: Dwellings of between 121 and 150 m^2.
a151_180: Dwellings of between 151 and 180 m^2.
a_181: Dwellings of more than 180 m^2.

vii. ***Family size (fam_size).*** The family is defined as a group of people, generally linked through a blood or in-law relationship, independently of the degree, that live together, normally occupying the totality of the dwelling. The family includes domestic workers who stay overnight in the dwelling and family lodgers.

fams_1: Families of 1 individual.
fams_2: Families of 2 individuals.
fams_3: Families of 3-4 individuals.
fams_4: Family of 5 or more individuals.

viii. ***Year of construction of the dwelling*** (***year_constr_dwelling***). This refers to the year in which the dwelling was built, or the age of the dwelling.

c_1900: Dwellings from 1900 and earlier.
c1901_40: Dwellings from between 1901 and 1940.
c1941_50: Dwellings from between 1941 and 1950.
c1951_60: Dwellings from between 1951 and 1960.
c1961_70: Dwellings from between 1961 and 1970.
c1971_80: Dwellings from between 1971 and 1980.
c1981_90: Dwellings from between 1981 and 1990.
c1991_00: Dwellings from between 1991 and 2000.
c_2001: Dwellings from 2001 and later.

ix. ***Level of Basque (level_basque)*** This variable was obtained by assigning a typology of "none", "with difficulty" and "good" to the basic indicators "understands" and "speaks" of the Population Census in relation to knowledge of Basque.

basque: Basque speakers: Individuals who understand and speak Basque well.
quasi_basq: Quasi-Basque speakers: People who understand Basque well or with difficulty.
n_basque: Non Basque speakers: Individuals who do not speak or understand Basque.

x. ***Mother tongue (leng_mat).*** This refers to the first language learned during childhood up to the age of 3.

mt_basque: Individuals whose mother tongue is Basque.
mt_spanish: individuals whose mother tongue is Spanish.
mt_both: Individuals who were exposed to Basque and Spanish during early childhood almost simultaneously.
mt_other: Individuals who learned neither Basque nor Spanish as a mother tongue.

xi. ***Family income (fam_income)*** (Quantitative). The total family income is produced as an aggregation of total personal income of all adult members of one family who receive some type of income during the year considered.

xii. ***Personal income (per_income)*** (Quantitative). Result of the aggregation, for each individual aged 18 and over, of his or her income from work, income derived from economic activities, transfers or social provisions and those received for capital gains, both moveable and immoveable.

As mentioned earlier, the variable "*Family income*" was qualified or categorised into 6 fractions. Consequently, one more variable was defined, the thirteenth.

xiii. ***Income (income)*** (Qualitative)

very_low: less than quantile 1/6 (16.67%).
low: between quantiles 1/6 (16.67%) and 2/6 (33.33%).
med_low: between quantiles 2/6 (33.33%) and 3/6 (50.00%).
med_high: between quantiles 3/6 (50.00%) and 4/6 (66.67%).
high: between quantiles 4/6 (66.67%) and 5/6 (83.33%).
very_high: more than quantile 5/6 (83.33%).

The number of residents (*'total'*) is very important in the description of each section and in the method of analysis. It is not a variable in itself, but it must be taken into account in the analysis. Furthermore, when the clusters are created at the end of the study, the number of residents of each cluster helps to describe the area, even though it is not represented graphically.

By way of example, consider the description of section **0100101001** below:

```
    section total basque_c e_country outside a_16_m a_16_f
0100101001  2733   1963     536       234    293    304
 a16_24_m a16_24_f a25_44_m a25_44_f a45_64_m a45_64_f
   80        89      609      557      301      256
 a_64_m a_64_f prima vocation secon med_high higher employ_m
   123    121   1498   450     385    185      215     818
 employ_f unempl_m unempl_f inact_m inact_f uns_worker
   595      49       66       539     666       642
 mach_ope q_worker far_fish shopk_wt adm_wor sup_tec
   163      301       31      507      207      396
 tec_prof s_60 s61_90 s91_120 s121_150 s151_180 s_181
   486     99   822     803     520      290     199
 fams_1 fams_2 fams_3_4 fam_5_ c_1900 c1901_40 c1941_50
   254    642    1525    312     88      46      30
 c1951_60 c1961_70 c1971_80 c1981_90 c1991_00 c_2001 basque
   59        262      369     204      454     936     771
 quasi_basq n_basque mt_basque mt_spanish mt_both lm_other
    696       1266      116       2451      107      59
 very_low low med_low med_high high very_high fam_income
    0     0       0     2733    0       0       37259
 per_income
    18763
```

As shown, frequencies are represented in all of the columns, except in the last two (that is, *Family income* and *Personal income*). The original data table contains 1,731 lines.

Multiple Factor Analysis

Mutiple Factor Analysis (MFA), developed by *B. Escofier* and *J. Pagès* (1992), allows for a simultaneous analysis of tables in which a same group of individuals is described by means of various groups of variables. Groups of variables can arise from the joint use of variables of a different nature, quantitative or qualitative, from the use of tables derived from others in three dimensions or from the use of a same group of variables measured at distinct time periods.

In some cases the variables are naturally defined in groups and instead of assigning the same weight or importance to the variables (a technique used in the majority of cases where no more information is available for any analysis), they are assigned to the groups of variables in order to subsequently complete the analysis. First of all, each group of variables is analysed and the results of each analysis give rise to a new analysis, in which the weight or importance within the initial results is clearly controlled. This is precisely the methodology proposed by Escofier and Pagès, and integrated in the *Factominer* package in the aforementioned ***R*** computing environment.

According to this study's methodology, the modalities of a qualitative variable form a group of variables and, given that they are frequencies, they are analysed with the appropriate techniques (such as *correspondence analysis*, for example).

Consequently, multiple factor analysis was performed on a table with dimensions of 1,731x60, in which the lines (sections) have their own weight and the columns (modalities) are grouped together by variables.

7 is the number of remaining factors. This decision was made taking into account the fact that the first 7 factors explain more than 70% of the initial variance (in the quantity of information). The automatic classification for completing the typology was performed with these 7 factors.

Automatic classification

A number of steps have been taken towards grouping the sections into clusters.

The characteristics of the sections are the starting point, and correspond to the first 7 factors or dimensions obtained via factor analysis. The dimensions retain their own variance; that is, they are non-standardised values. The first factor has therefore more importance in the classification process than the seventh.

The usual Euclidean system was used for classification.

Generally, the first step involves performing a *hierarchical classification* following the *Ward method*, to obtain a partition of the initial group of sections and to create the aforementioned clusters. The *reassignment algorithm* is used to improve the quality of the partition (the *Ward* method tends to produce compact clusters). Finally, taking the centres of the produced clusters as nuclei, a new partition is created by applying the so-called *k-means* algorithm.

In this particular case, after performing the *hierarchical classification* using the *Ward method*, 15 clusters were obtained, to which the *reassignment algorithm* was subsequently applied. To be consistent, the decision was made to redefine the 15 clusters as 12, fusing certain clusters with others. Finally, the *k-means* algorithm was applied, taking the 12 clusters as nuclei and creating a new partition.

The clusters produced are as follows:

Cluster 1/12. A predominance of dwellings from 1951-1970, of under 91 m^2 and with a significant population aged 65 and over.

Cluster 2/12. A predominance of dwellings from 1971-1980, of between 61-90 m^2 and with a significant population of between the ages of 45 and 64.

Cluster 3/12. A predominance of dwellings of between 91-150 m^2 and of professionals or executives, and a significant proportion of dwellings from before 1961 and of people aged 65 and over.

Cluster 4/12. A predominance of dwellings of more than 90 m^2 and of professionals or executives, and a significant proportion of dwellings built between 1900-1940 and 1981-2000, and of people between the ages of 45 and 64 and between 16 and 24.

Cluster 5/12. A predominance of dwellings from before 1961, and with a significant proportion of dwellings under 61 m^2 and of people between the ages of 25 and 44.

Cluster 6/12. A predominance of dwellings over 90 m^2 built in 1990 and later or before 1900, and with a significant proportion of farmers and fishermen.

Cluster 7/12. A predominance of dwellings from between 1981 and 2000, between 61-120 m^2, in which there reside couples between the ages of 45 and 64 with children under 25.

Cluster 8/12. A predominance of dwellings between 61-90 m^2 and with a significant proportion built after 1990 in which there reside couples between the ages of 25 and 44 with children under 16.

Cluster 9/12. A predominance of dwellings built after 2000, between 61-90 m^2, in which there reside individuals between the ages of 25 and 44 and with a significant proportion under the age of 16 and degree holders.

Cluster 10/12. A predominance of Basque speakers in dwellings of between 61-90 m^2, and with a significant proportion of dwellings built between 1951-1970 and of individuals aged 65 and over.

Cluster 11/12. A predominance of Basque speakers, and a significant proportion of dwellings between 91-120 m^2, built after 1991 or before 1900 and of individuals under 25.

Cluster 12/12. A predominance of Basque speakers in dwellings more than 90 m^2, and a significant proportion of dwellings built after 1991 or before 1900 and of technical professionals, farmers and fishermen.

The results are shown graphically in the following image (Image 1):

- Projections of the sections in the main plane (formed by first two dimensions) according to the number of corresponding residents, and coloured by cluster.

- Projections of the centres (averages) of the 12 clusters in the main plan and in three-dimensional space, projected according to the size of the clusters.

-

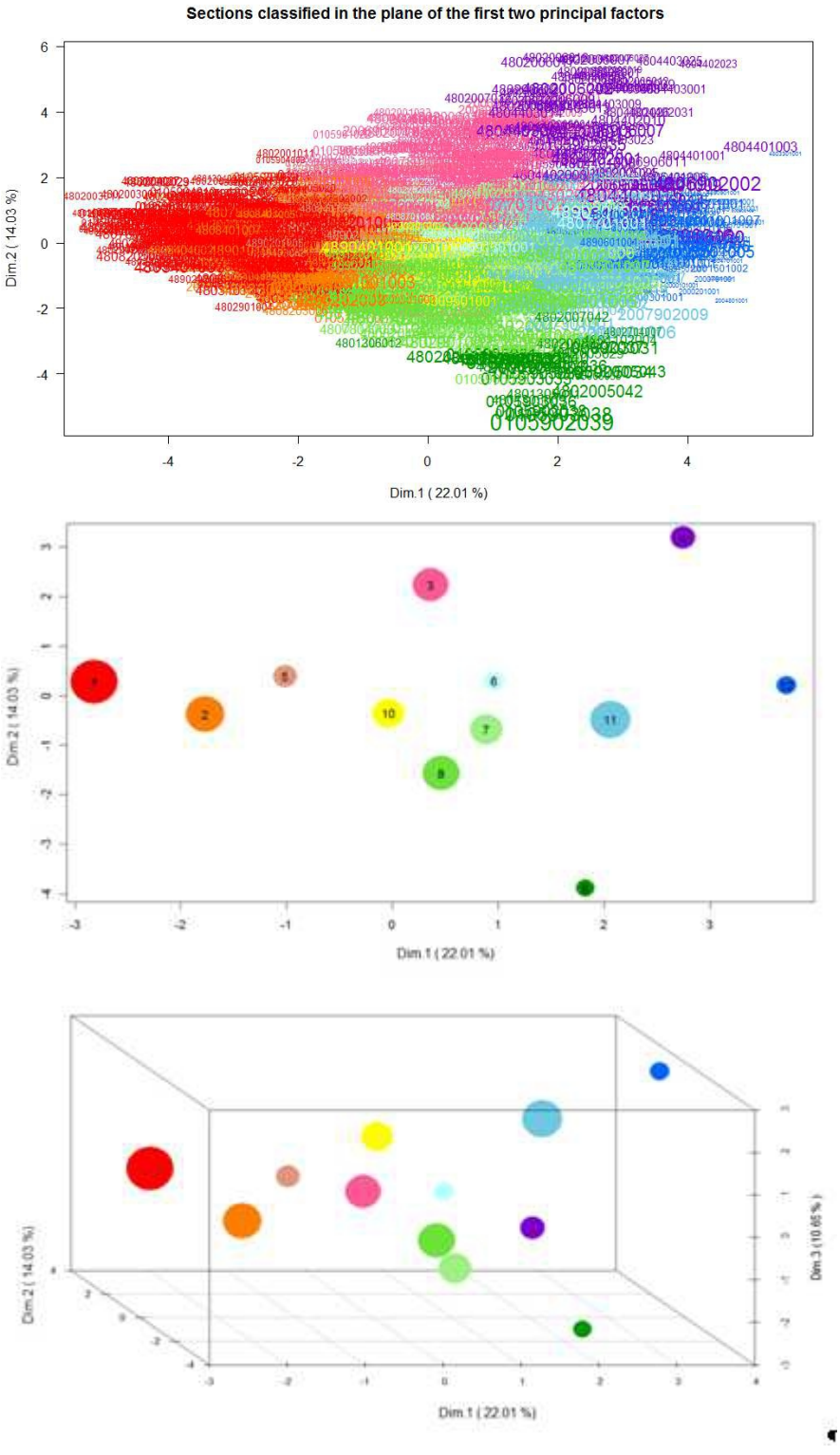


Image 1. Projection of the clusters in the space of two and three principal factors.

Chapter
3

Typology of the sections. Description of the clusters or classes.

The above classification produces a typology of the sections. The description of these clusters must be understood in order to be useful, and three complementary methods are offered to achieve this aim:

- Brief description of the 12 clusters based on the above graphical description.

- Graphical descriptions of the 12 clusters according to the original variables. On the graph of each variable, the cluster sections and remaining sections were compared. The bar representing the frequency of each modality has been given a more intense colour when the frequency in the sections belonging to the cluster is greater than that of those that don't belong to the cluster, to make it easier to understand. In this respect, the order of variables has been determined according to *Pearson's* X^2 *statistical test*, used to measure homogeneity.

- File kmz (compressed version of a kml) that represents the geographical location of the classified sections on a coloured map of the Basque Country, visualised using Google Earth. The colours used are associated with the RGB colour triangle. Given that the location of clusters is approximately triangular in the plane of the two principal factors (the main plane), it has been attempted to make it similar to the RGB colour triangle (http://en.wikipedia.org/wiki/SRGB). The blue-red difference has been associated with the first factor.

CLUSTER 1/12.

A predominance of dwellings from 1951-1970, of under 91 m^2 and with a significant population aged 65 and over.

This cluster groups 355 sections 390,692 residents (18.03% of the population).

There is a predominance of:

- Dwellings from 1951-1970.
- Below average family income.
- Dwellings under 91 m^2.
- Spanish speakers.
- Primary degrees.
- Inactive.

Particularly significant:

- Born elsewhere in the country (excluding the Autonomous Region of the Basque Country).
- Population aged 65 and over.
- Families of 1 or 2 individuals.

CLUSTER 2/12.

A predominance of dwellings from 1971-1980, of between 61-90 m^2 and with a significant population of between 45-64 years of age.

This cluster groups 203 sections 253,669 residents (11.70% of the population).

There is a predominance of:

- Dwellings from 1971-1980.
- Below average family income.
- Dwellings of between 61-90 m^2.
- Spanish speakers.
- Primary qualifications.
- Families of 2-4 individuals.

Particularly significant:

- Born elsewhere in the country.
- Skilled and unskilled workers, shopkeepers and waiters.
- Individuals between the ages of 45 and 64.

CLUSTER 3/12.

A predominance of dwellings of between 91-150 m^2 and of professionals or executives, and a significant proportion of dwellings from before 1961 and of population aged 65 and over.

This cluster groups 194 sections 229,199 residents (10.57% of the population).

There is a predominance of:

- Above average family income.
- Professionals or executives, technicians and administrative employees.

Particularly significant:

- Dwellings built before 1961.
- Secondary, medium and higher education qualifications.
- Dwellings of between 91-150 m^2.
- Population aged 65 and over, and women between the ages of 45 and 64.
- Families of 1 or 2 individuals.
- Inactive women.

CLUSTER 4/12.

A predominance of dwellings of more than 90 m^2 and of professionals or executives, and a significant proportion of dwellings built between 1900-1940 and 1981-2000, and of individuals between the ages of 45 and 64 and between 16 and 24.

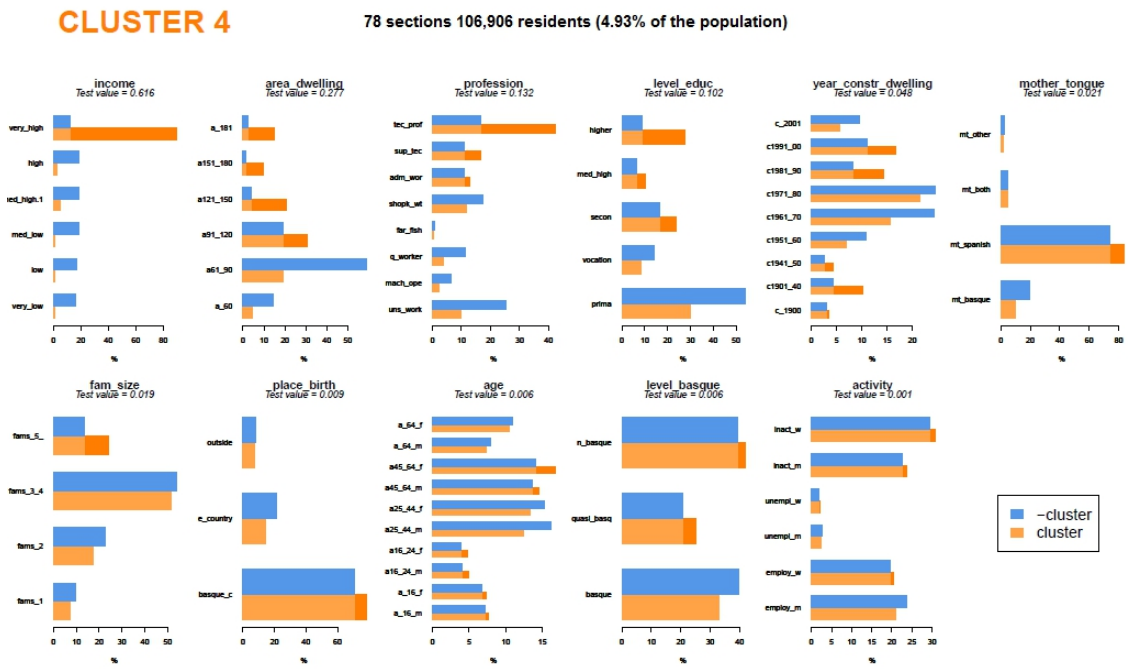
This cluster groups 78 sections 106,906 residents (4.93% of the population).

There is a predominance of:

- Very high family income.
- Dwellings over 90 m^2.
- Professionals or executives and support technicians.
- Secondary, upper-medium and higher education qualifications.
- Born in the Basque Country.

Particularly significant:

- Dwelling built between 1900-1940 and 1981-2000.
- Families of 5 or more members.
- Individuals between the ages of 45-64 and 16-24.
- Quasi-Basque speakers.



CLUSTER 5/12.

A predominance of dwellings from before 1961, and with a significant proportion of dwellings under 61 m^2 and of individuals between the ages of 25 and 44.

This cluster groups 87 sections 98,911 residents (4.56% of the population).

There is a predominance of:

- Below average family income.
- Dwellings from before 1961.
- Primary qualifications.

Particularly significant:

- Dwellings under 61 m^2.
- Born abroad.
- Families of 1 individual and of 5 or more.
- Spanish speakers.
- Unskilled workers.
- Individuals between the ages of 25 and 44.
- Unemployed.

CLUSTER 6/12.

A predominance of dwellings over 90 m^2 built in 1990 and later or before 1900, and with a significant proportion of farmers and fishermen.

This cluster groups 77 sections 75,051 residents (3.46% of the population).

There is a predominance of:

- Dwellings of more than 90 m^2, built after 1990 or before 1990.
- Born in the Basque Country.
- Primary qualifications.

Particularly significant:

- Spanish speakers.
- Families of 5 or more members.
- Unskilled workers, farmers and fishermen.
- Men of 45-64 and individuals under 16 years of age.

CLUSTER 7/12.

A predominance of dwellings from between 1981 and 2000, between 61-120 m^2, in which there reside couples between the ages of 45 and 64 with children under 25.

This cluster groups 122 sections 185,554 residents (8.56% of the population).

There is a predominance of:

- Dwellings built between 1981-2000 and of 61-120 m^2.
- Couples of 45-64 years of age with children under 25.
- Families of 3-4 individuals.

Particularly significant:

- Above average family income.
- Quasi-Basque speakers.
- Employed.

CLUSTER 8/12.

A predominance of dwellings between 61-90 m^2 and with a significant proportion built after 1990 in which there reside couples between the ages of 25 and 44 with children under 16.

This cluster groups 151 sections 244,354 residents (11.27% of the population).

There is a predominance of:

- Dwellings of between 61-90 m^2.
- Families of 3-4 individuals.
- Born in the Basque Country.

Particularly significant:

- Dwellings built in 1990 and later
- Below average family income.
- Couples of 25-44 years of age with children under 16.
- Quasi-Basque speakers.
- Employed.

CLUSTER 9/12.

A predominance of dwellings built after 2000, between 61-90 m^2, in which there reside individuals between the ages of 25 and 44 and with a significant proportion under the age of 16 and degree holders.

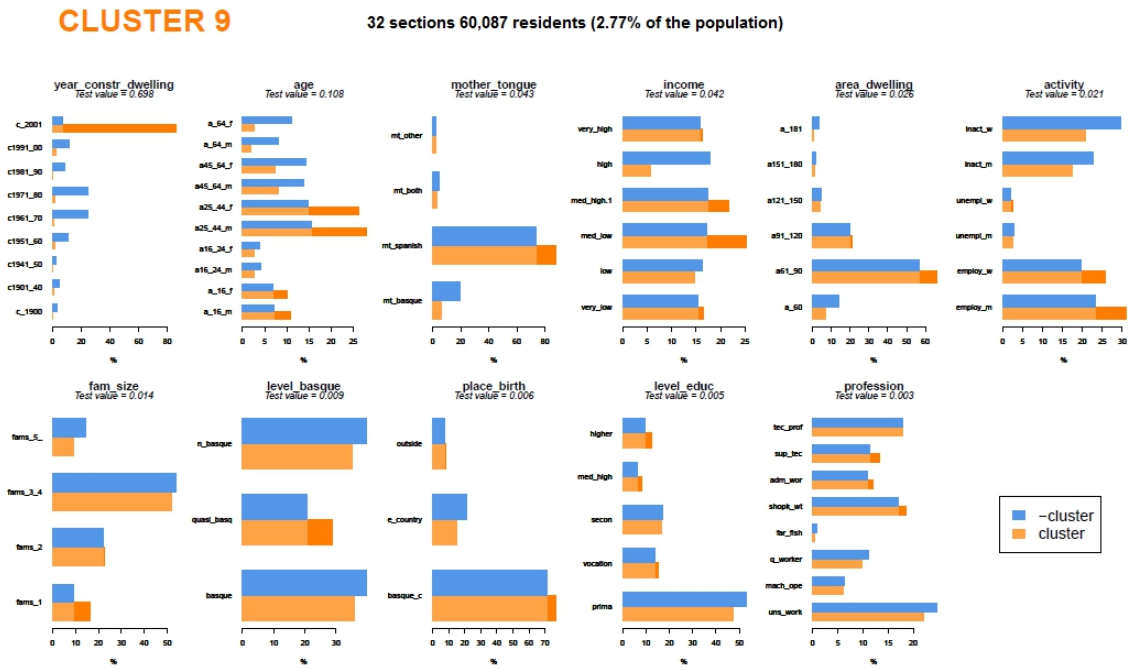
This cluster groups 32 sections 60,087 residents (2.77% of the population).

There is a predominance of:

- Dwellings built in or after 2000.
- Individuals between the ages of 25 and 44.
- Dwellings of between 61-90 m^2.
- Employed.
- Born in the Basque Country.

Particularly significant:

- Individuals under the age of 16.
- Average family income.
- Families of 1 individual.
- Quasi-Basque speakers.
- Degree holders.



CLUSTER 10/12.

A predominance of Basque speakers in dwellings of between 61-90 m^2, and with a significant proportion of dwellings built between 1951-1970 and of population aged 65 and over.

This cluster groups 143 sections 173,666 residents (8.01% of the population).

There is a predominance of:

- Above average family income.
- Basque speakers.
- Dwellings of between 61-90 m^2.
- Primary qualifications.
- Born in the Basque Country.

Particularly significant:

- Basque as mother tongue.
- Dwellings built between 1951 and 1970.
- Population aged 65 and over.
- Inactive.

CLUSTER 11/12.

A predominance of Basque speakers, and a significant proportion of dwellings between 91-120 m^2, built in or after 1991 or before 1900 and of individuals under 25.

This cluster groups 195 sections 281,496 residents (12.99% of the population).

There is a predominance of:

- Basque as mother tongue.
- Above average family income.
- Basque speakers:
- Born in the Basque Country
- Families of 3-4 individuals.
- Primary and professional qualifications.

Particularly significant:

- Dwellings of between 91-120 m^2, built in or after 1990 or before 1990.
- People under the age of 25.

CLUSTER 12/12.

A predominance of Basque speakers in dwellings more than 90 m^2, and a significant proportion of dwellings built after 1991 or before 1900 and of technical professionals, farmers and fishermen.

This cluster groups 94 sections 67,796 residents (3.13% of the population).

There is a predominance of:

- Basque as mother tongue. Basque speakers:
- Dwellings over 90 m^2.
- Above average family income.
- Born in the Basque Country.
- Primary and professional qualifications.

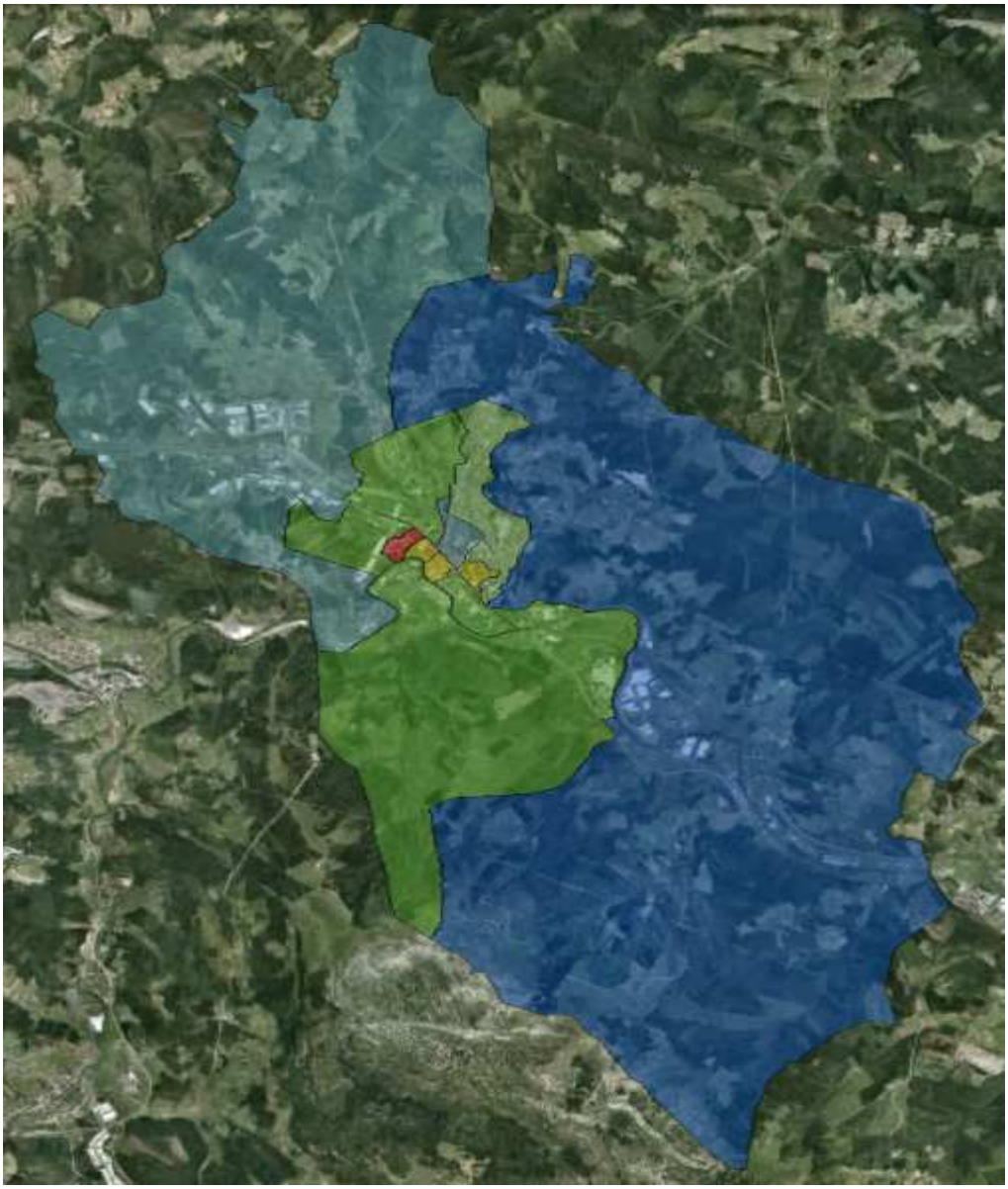
Particularly significant:

- Dwellings from 1990 or later or before 1900.
- Families of 5 or more members.
- Technical professionals, farmers and fishermen, machine operators.
- Individuals under the age of 16 and men between the ages of 45 and 64.
- Inactive men.

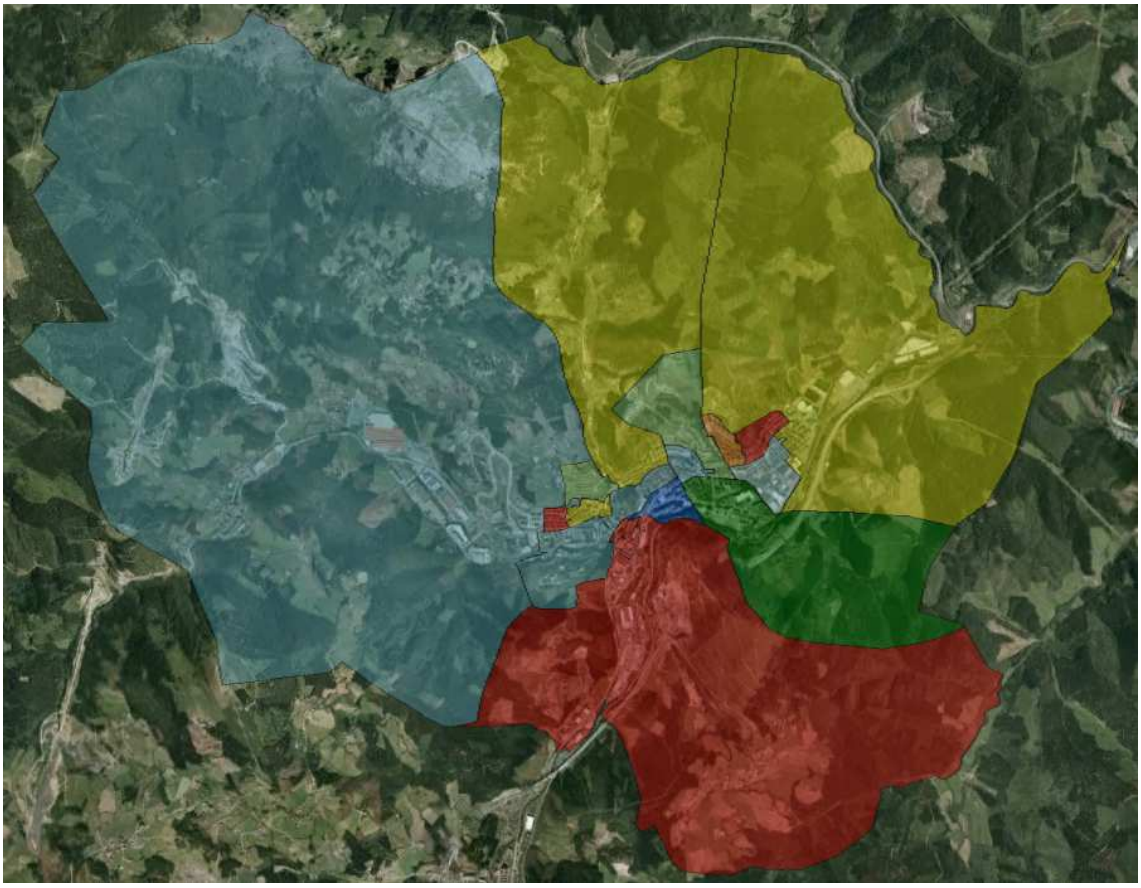
Chapter
4

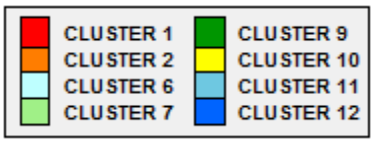
Municipalities of 15,000 or more residents.

Amorebieta-Etxano

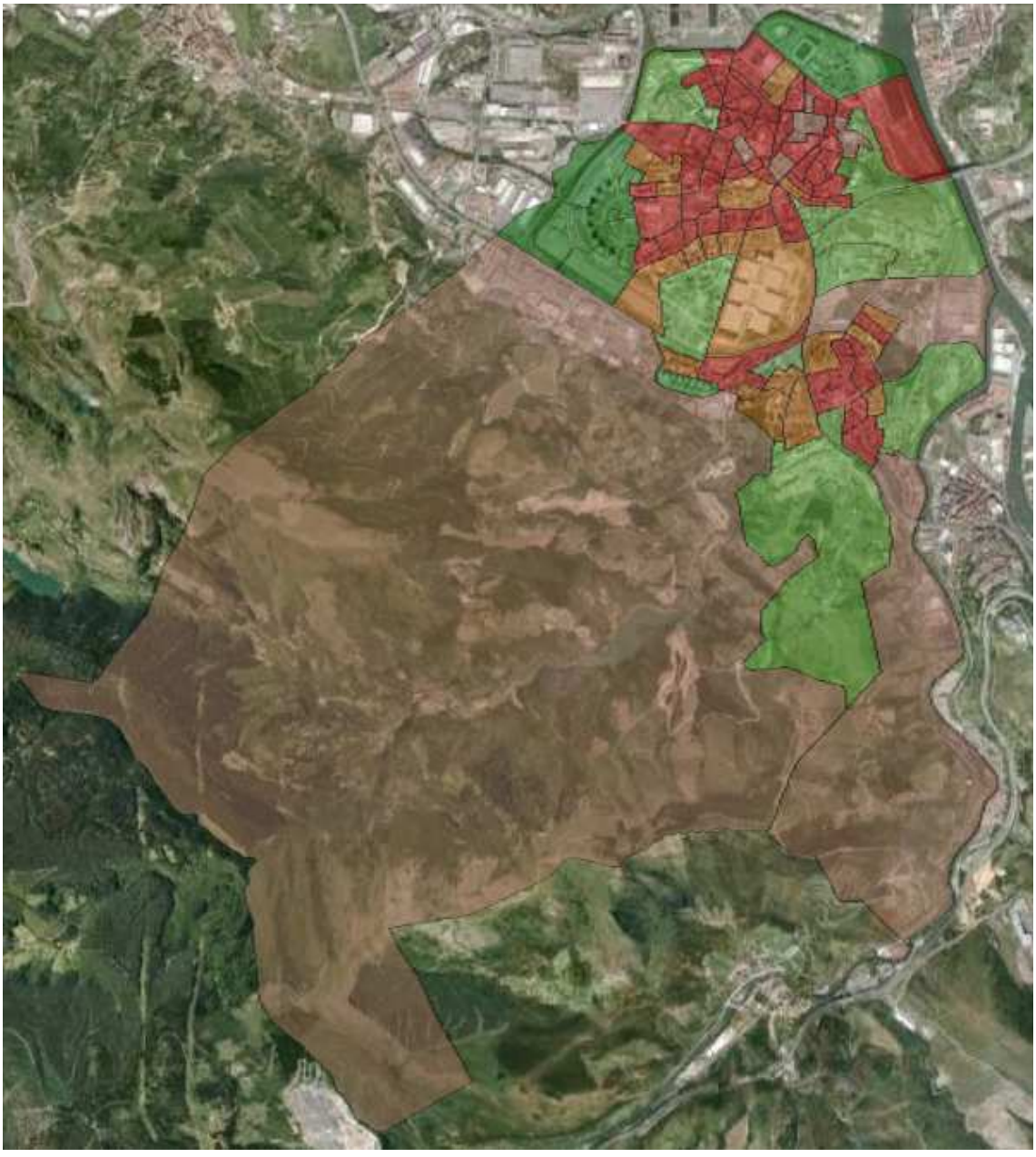


Arrasate/Mondragón

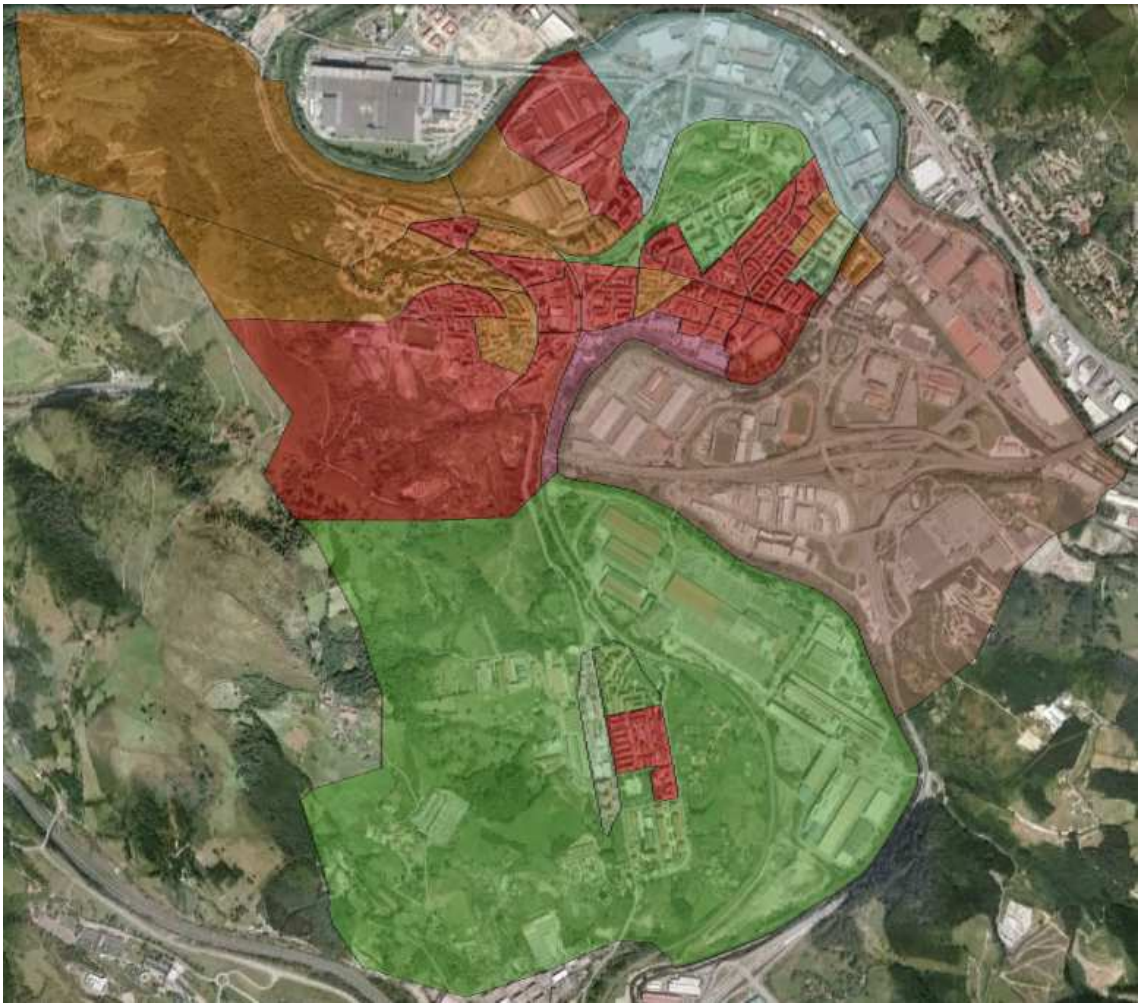




Barakaldo

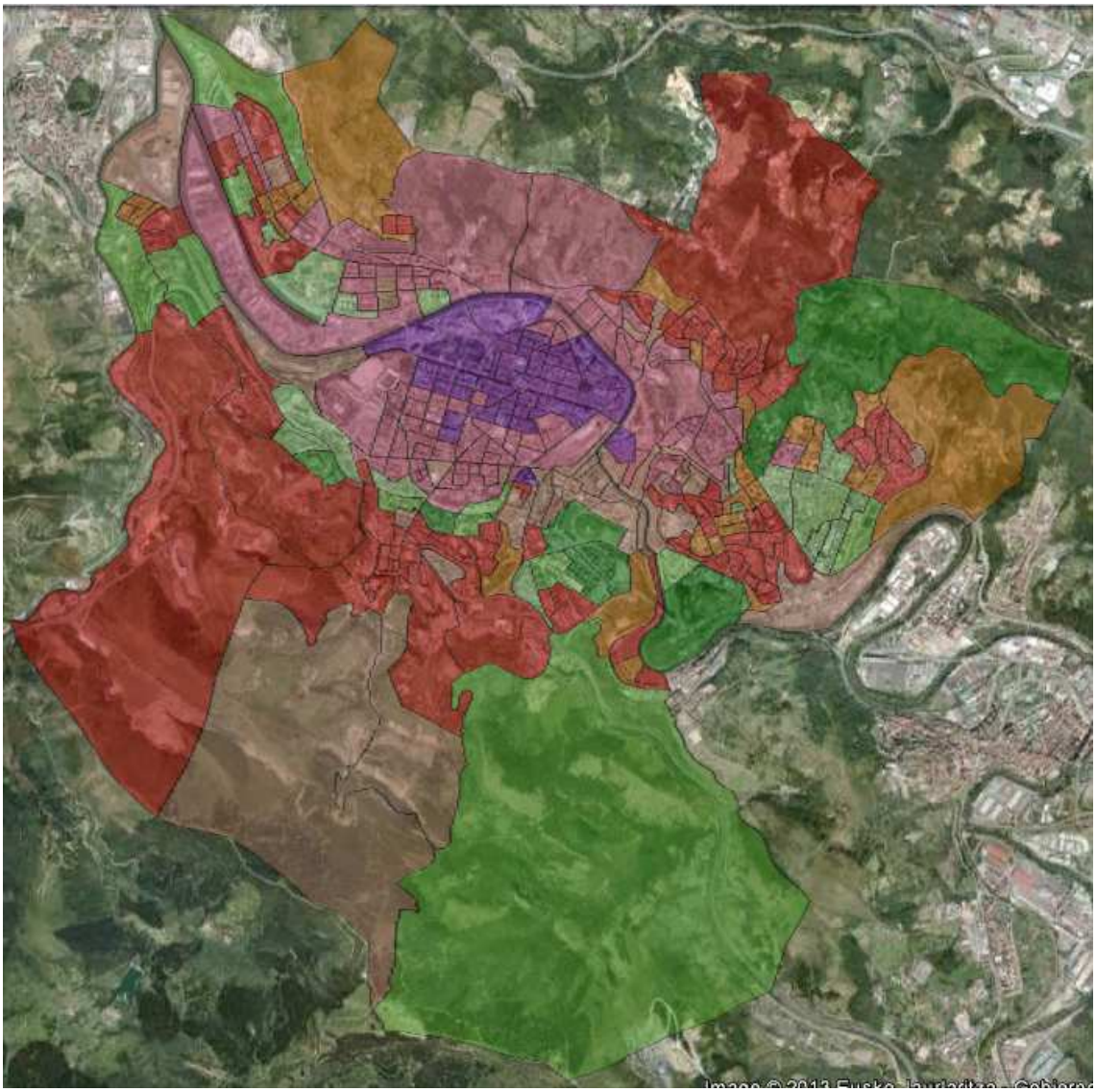


Basauri

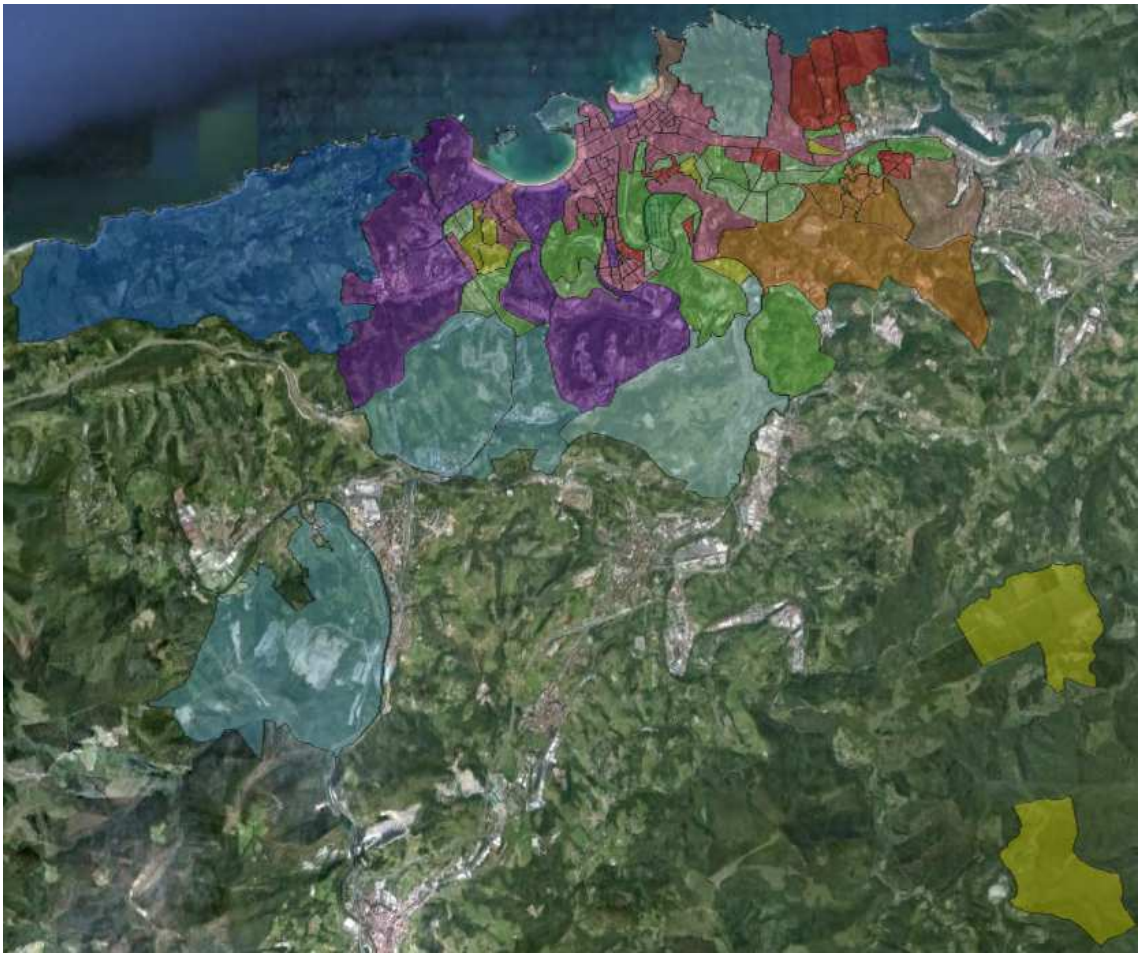

CLUSTER 1
CLUSTER 2
CLUSTER 3
CLUSTER 5
CLUSTER 6
CLUSTER 7
CLUSTER 8

Bermeo

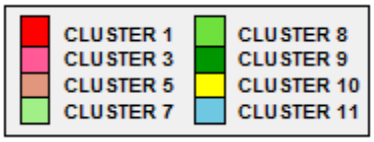
Bilbao


CLUSTER 1
CLUSTER 2
CLUSTER 3
CLUSTER 4
CLUSTER 5
CLUSTER 7
CLUSTER 8
CLUSTER 9

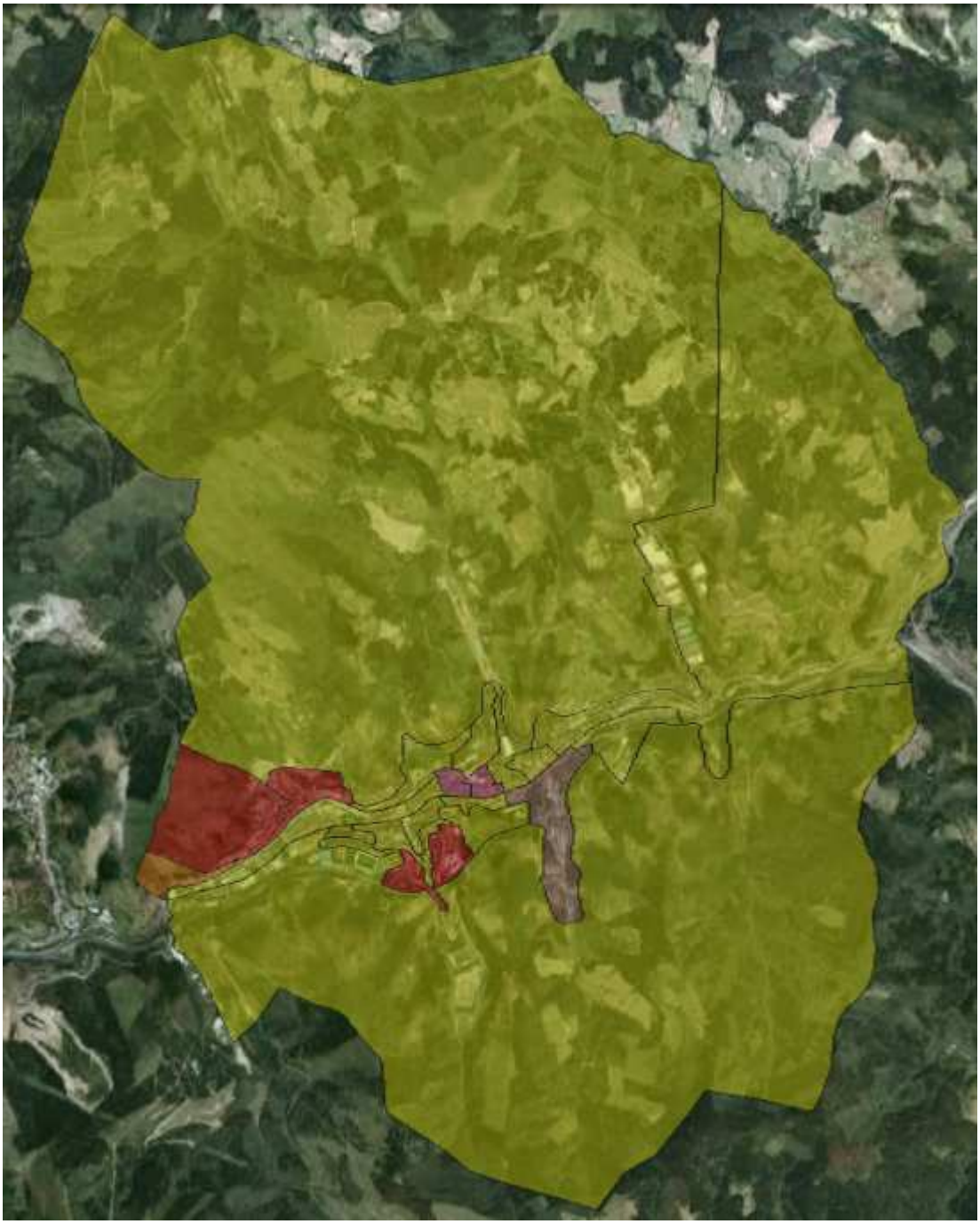
Donostia-San Sebastián



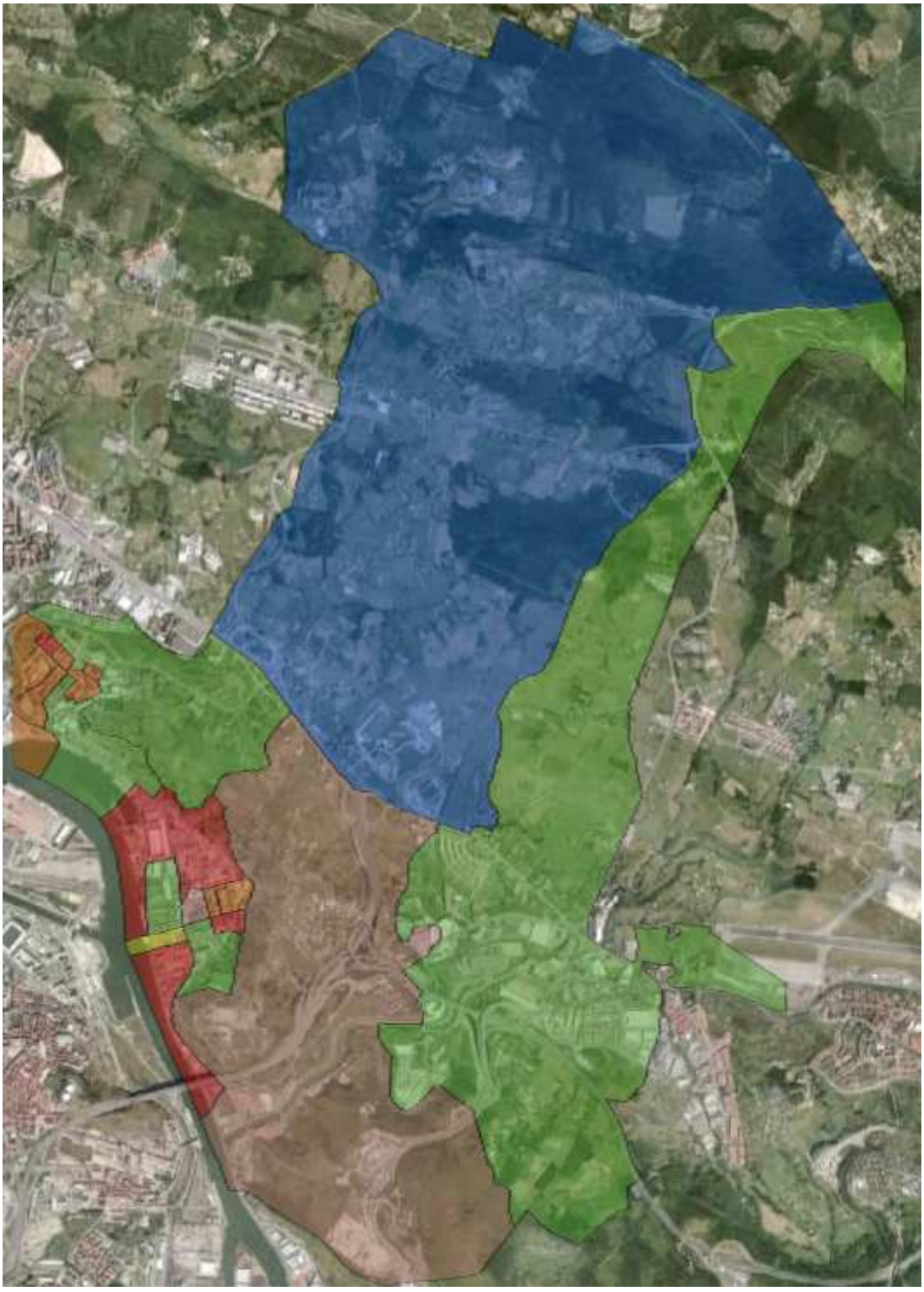
Durango


CLUSTER 1
CLUSTER 3
CLUSTER 5
CLUSTER 7
CLUSTER 8
CLUSTER 9
CLUSTER 10
CLUSTER 11

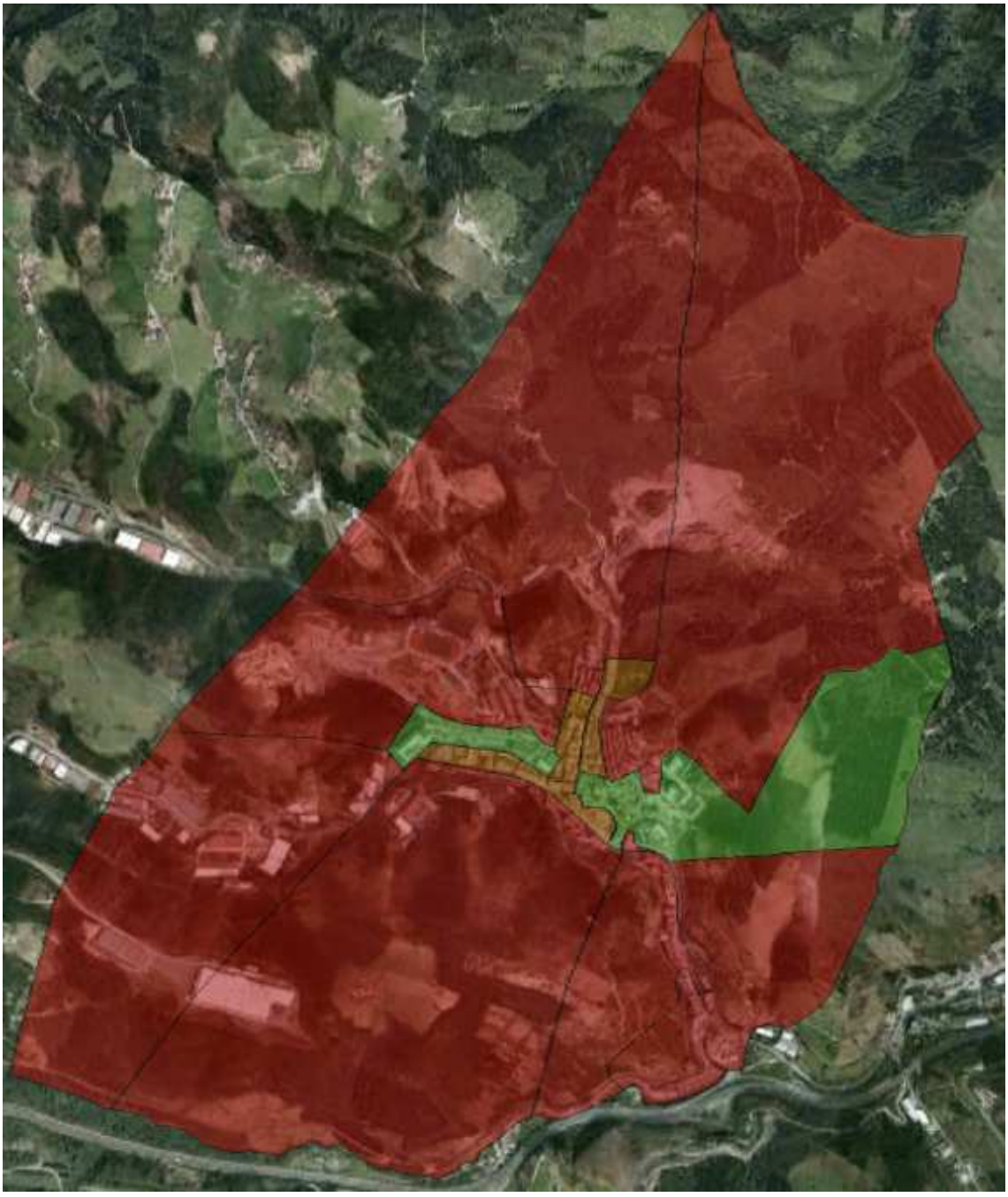
Eibar



Erandio


CLUSTER 1
CLUSTER 2
CLUSTER 5
CLUSTER 8
CLUSTER 10
CLUSTER 12

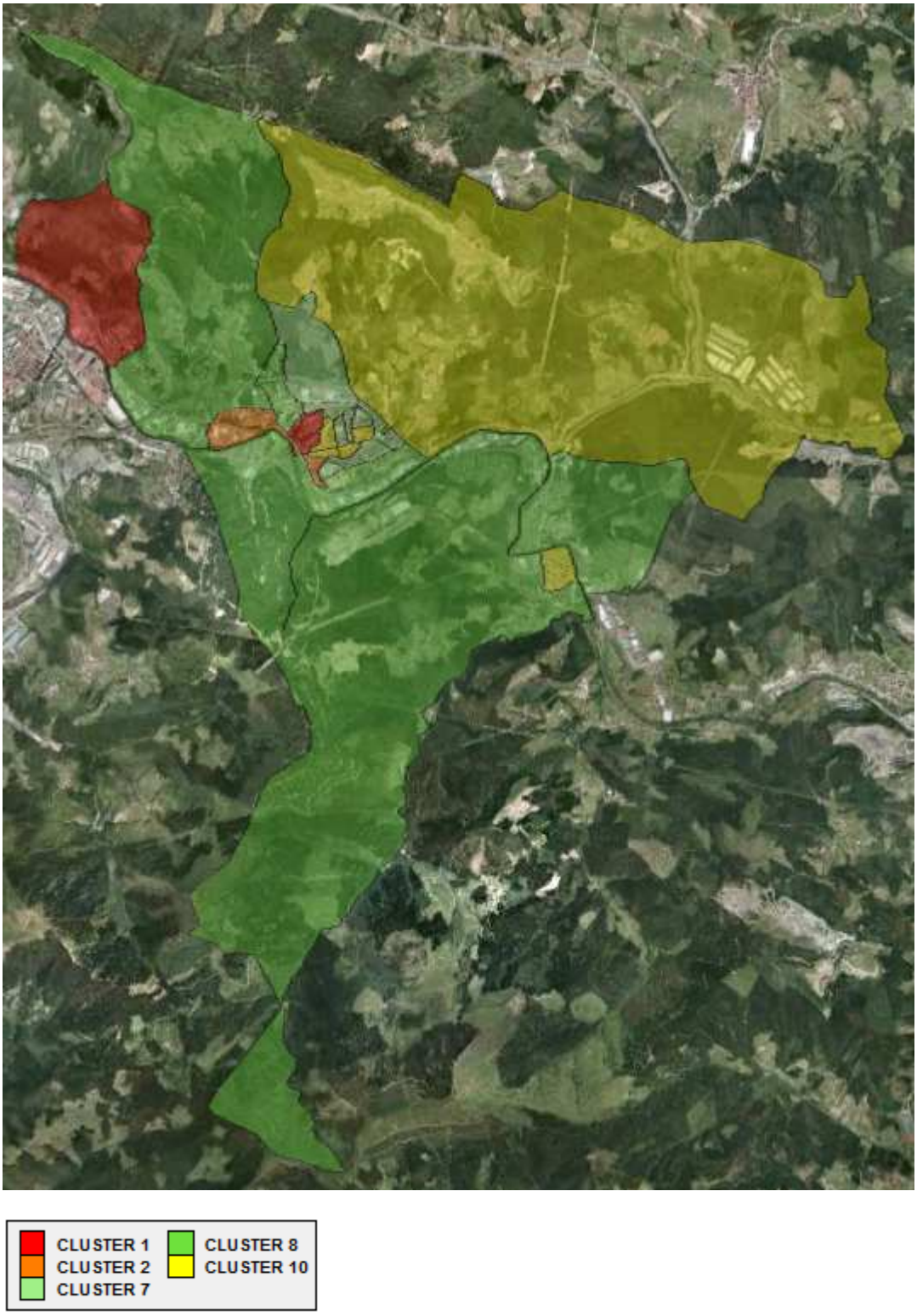
Ermua



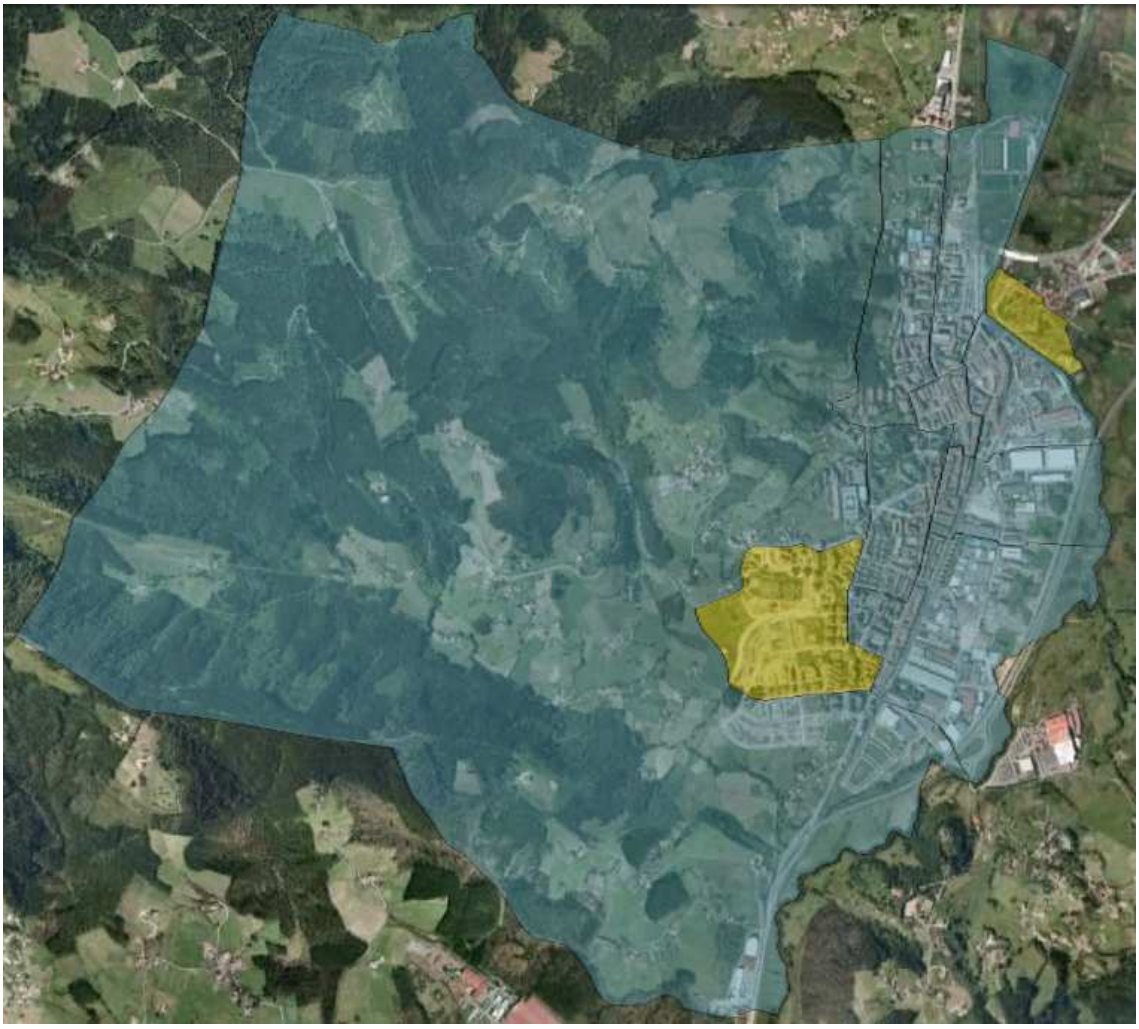
Errenteria


CLUSTER 1
CLUSTER 2
CLUSTER 7
CLUSTER 8
CLUSTER 10
CLUSTER 12

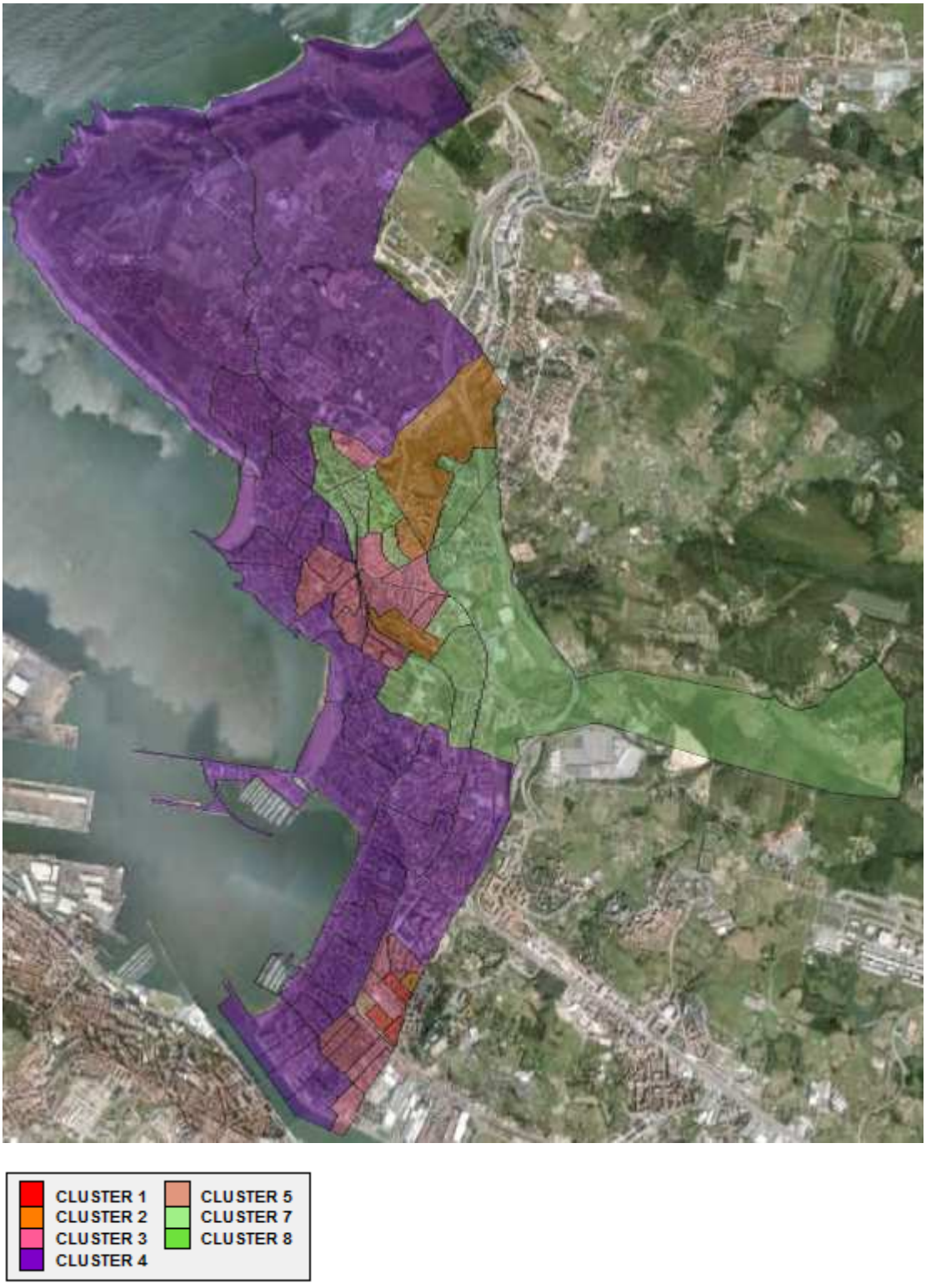
Galdakao



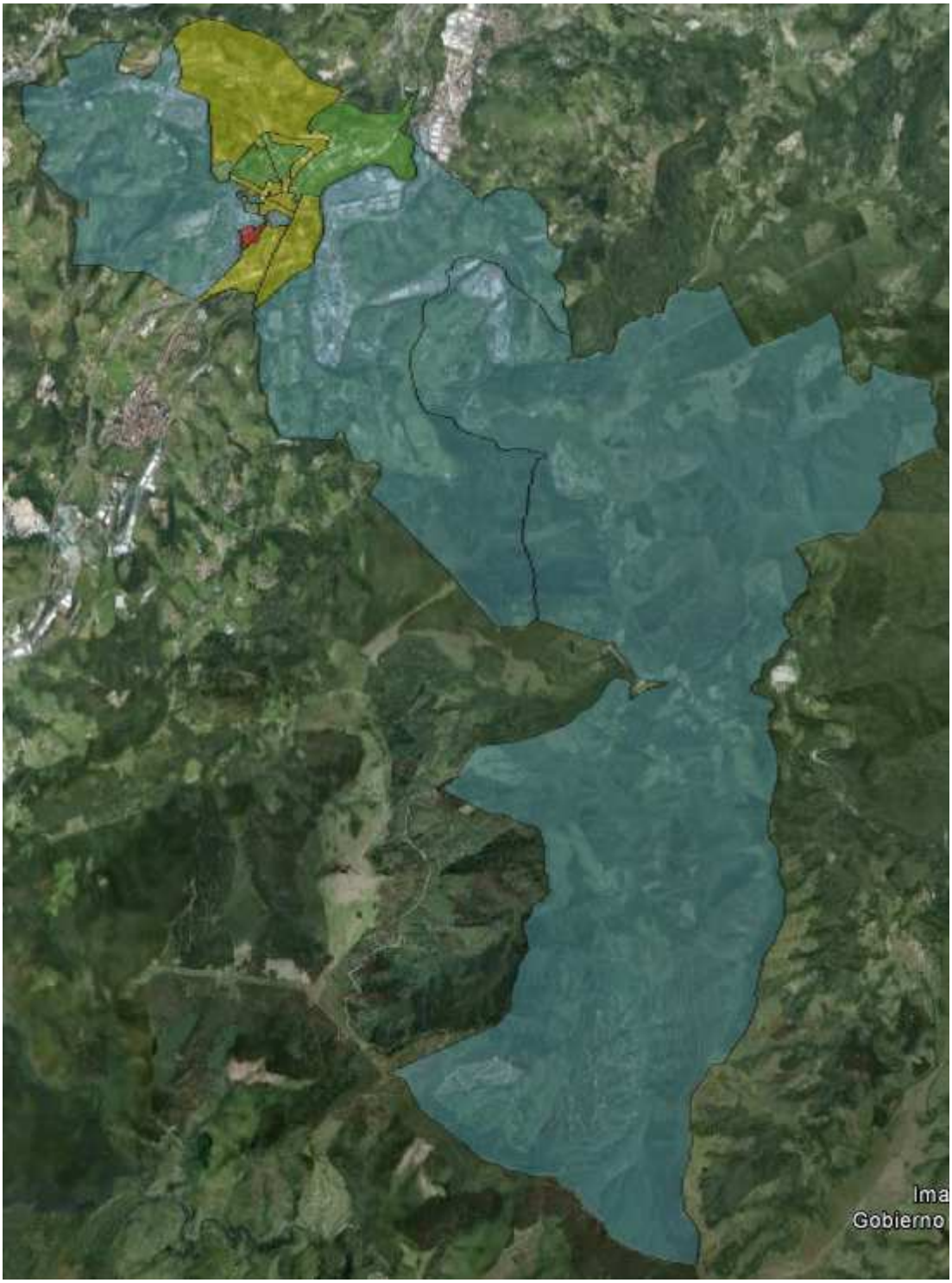
Gernika-Lumo



Getxo


CLUSTER 1
CLUSTER 2
CLUSTER 3
CLUSTER 4
CLUSTER 5
CLUSTER 7
CLUSTER 8

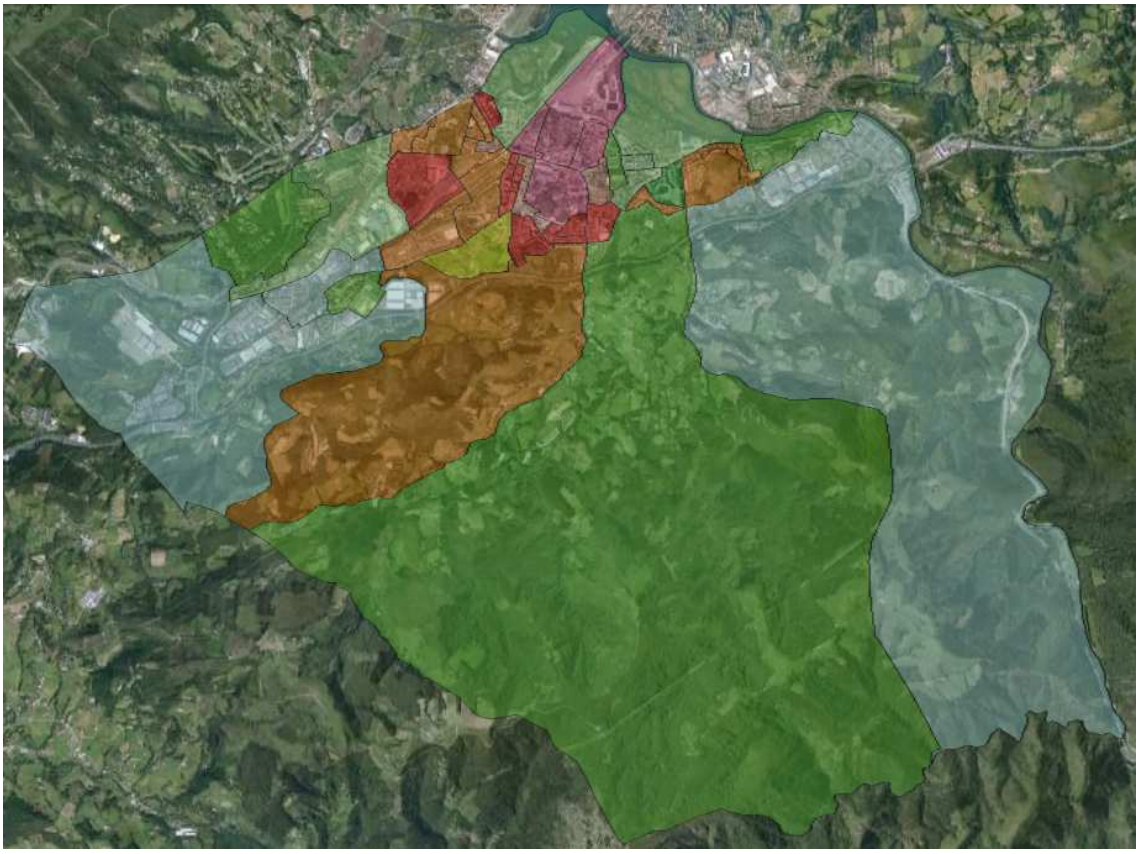
Hernani



CLUSTER 1
CLUSTER 7
CLUSTER 8
CLUSTER 10
CLUSTER 11

Hondarribia

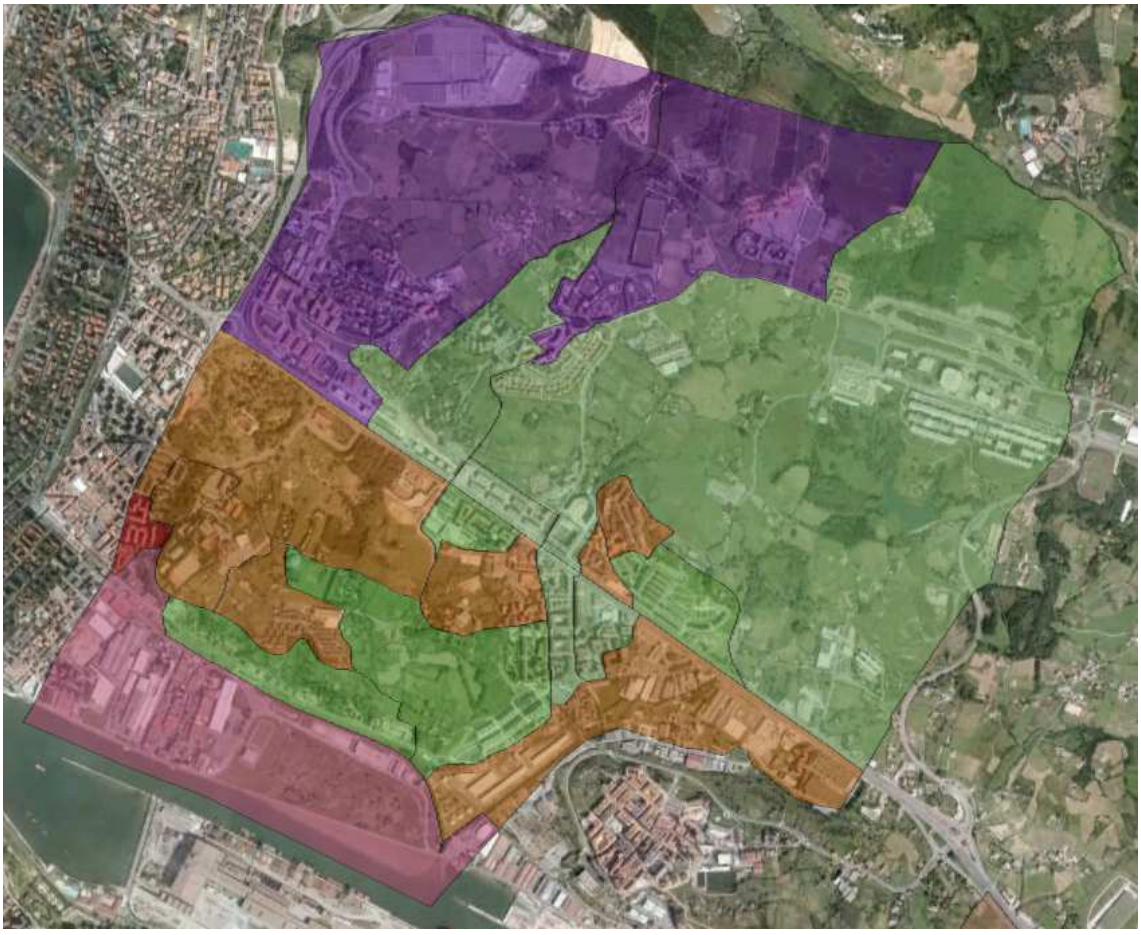
Irun


CLUSTER 1
CLUSTER 2
CLUSTER 3
CLUSTER 5
CLUSTER 6
CLUSTER 7
CLUSTER 8
CLUSTER 9
CLUSTER 10

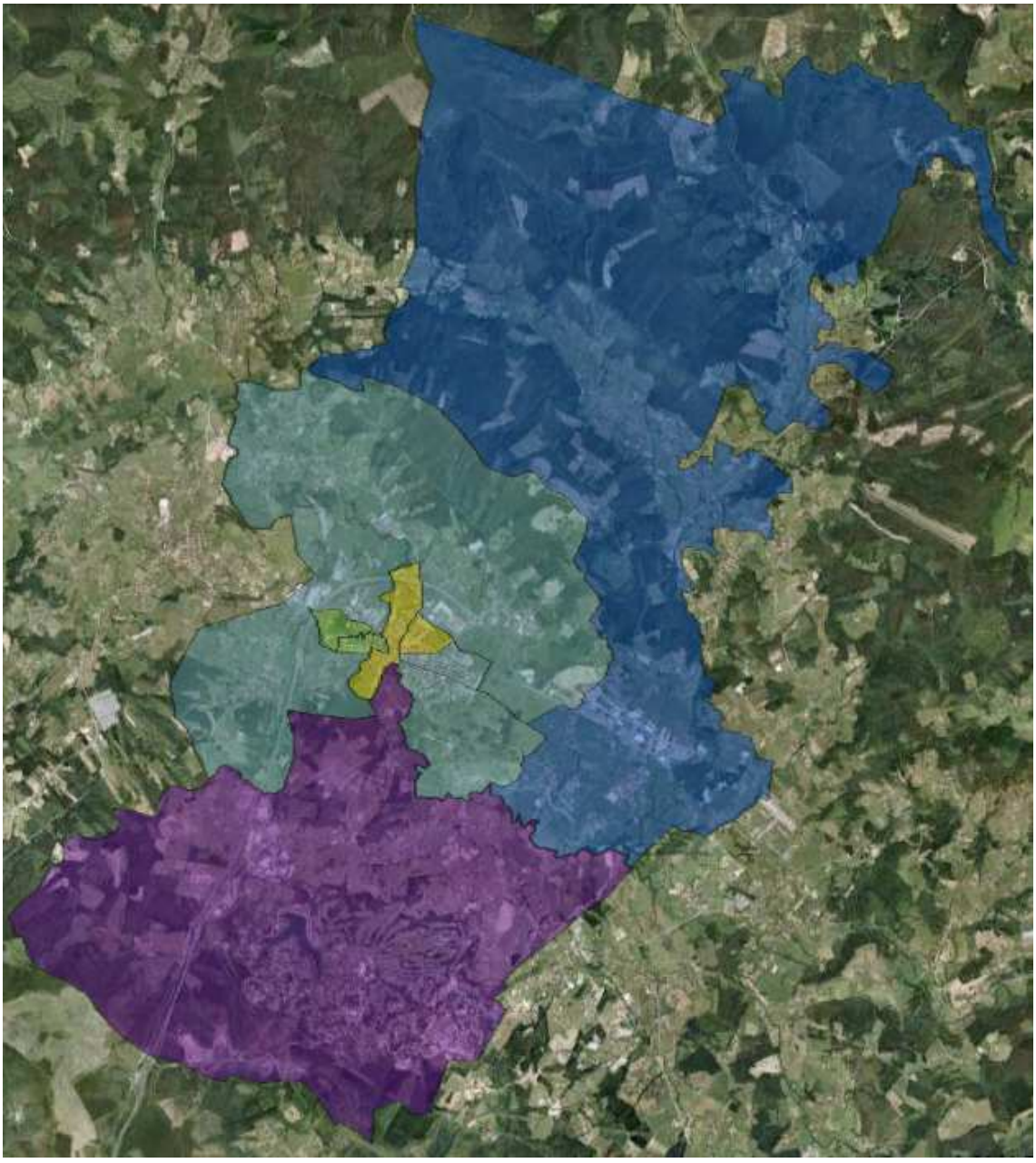
Lasarte-Oria

Laudio/Llodio

Leioa



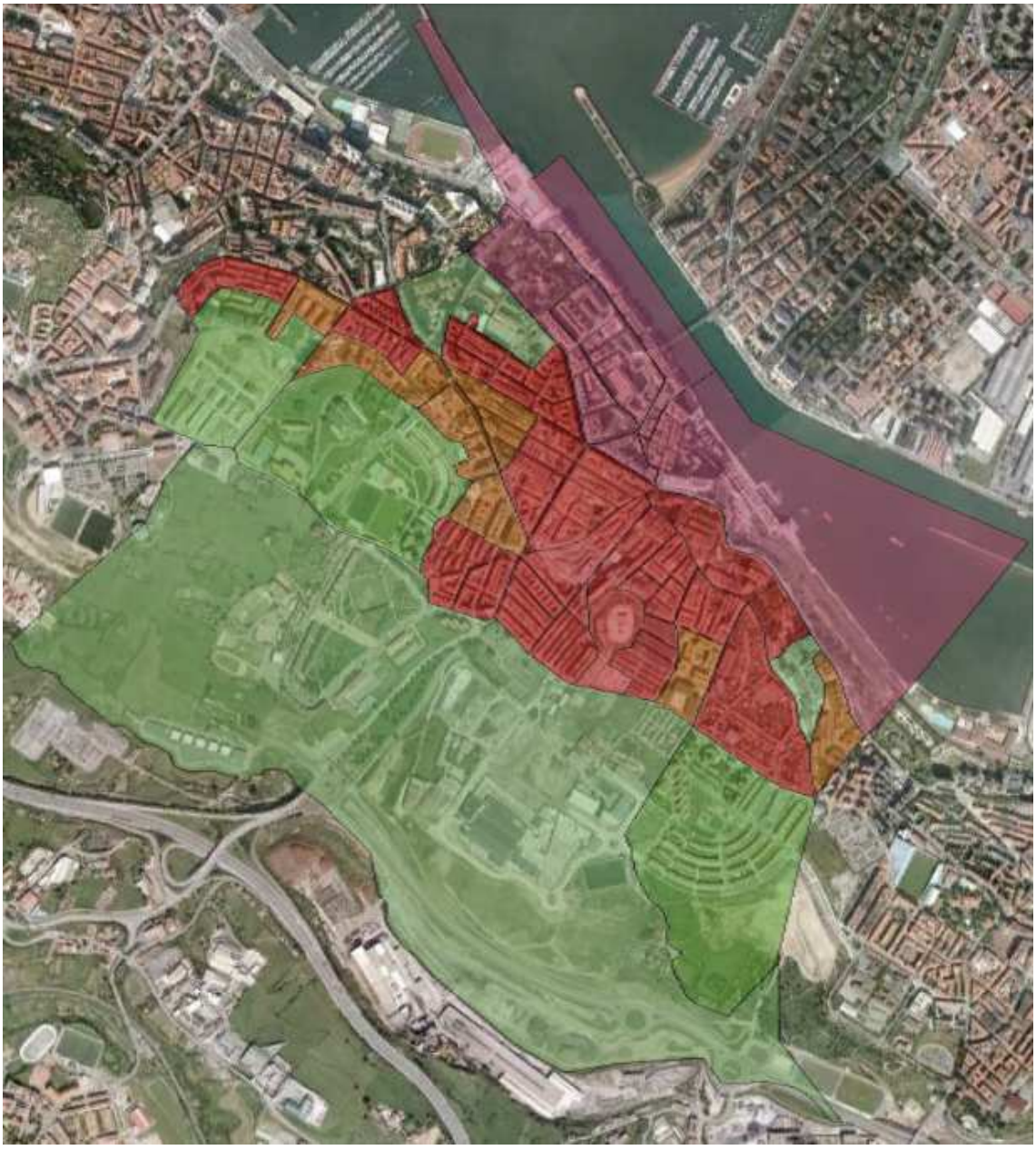
Mungia



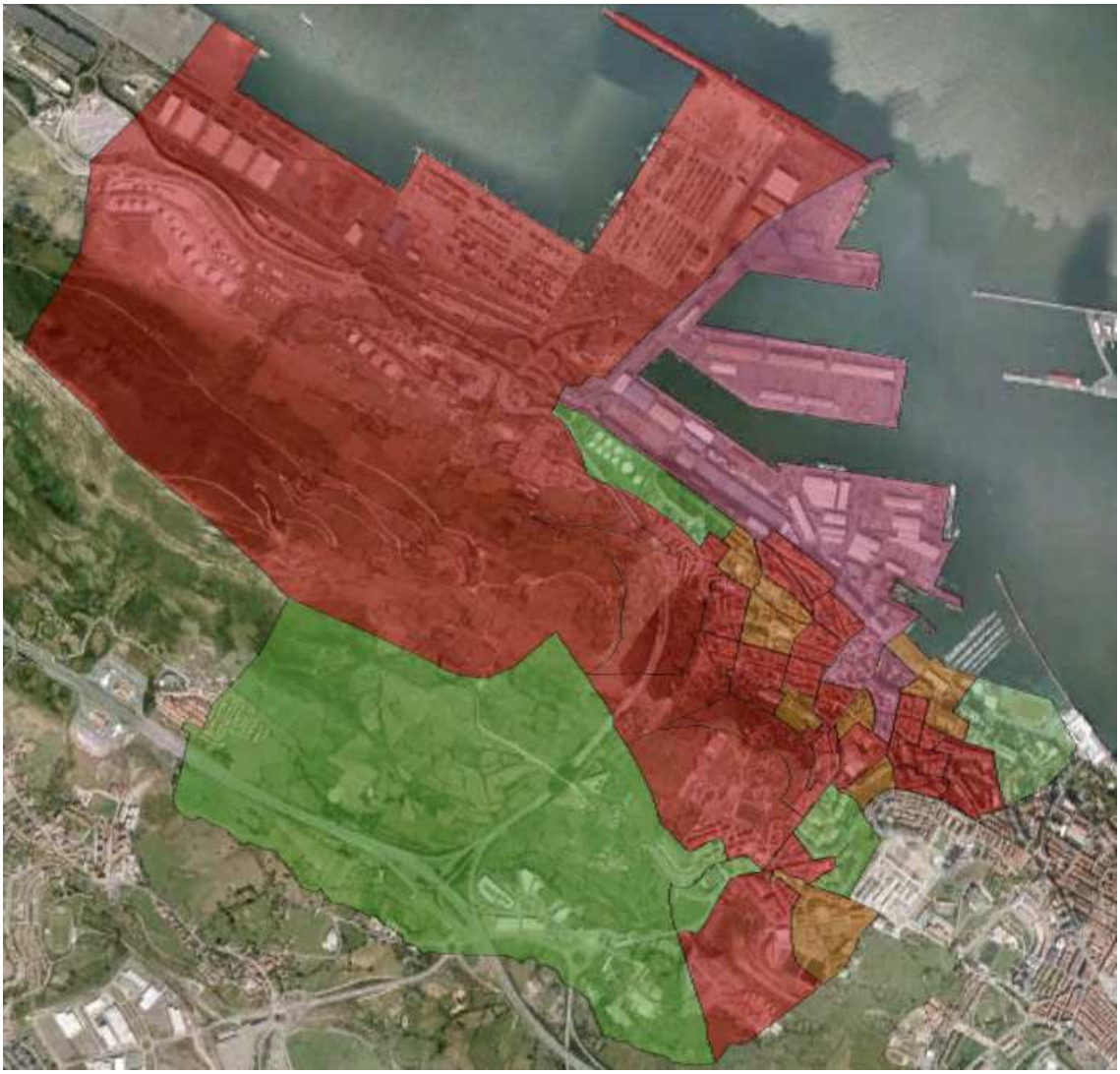
Pasaia

CLUSTER 1
CLUSTER 2
CLUSTER 5
CLUSTER 8
CLUSTER 10

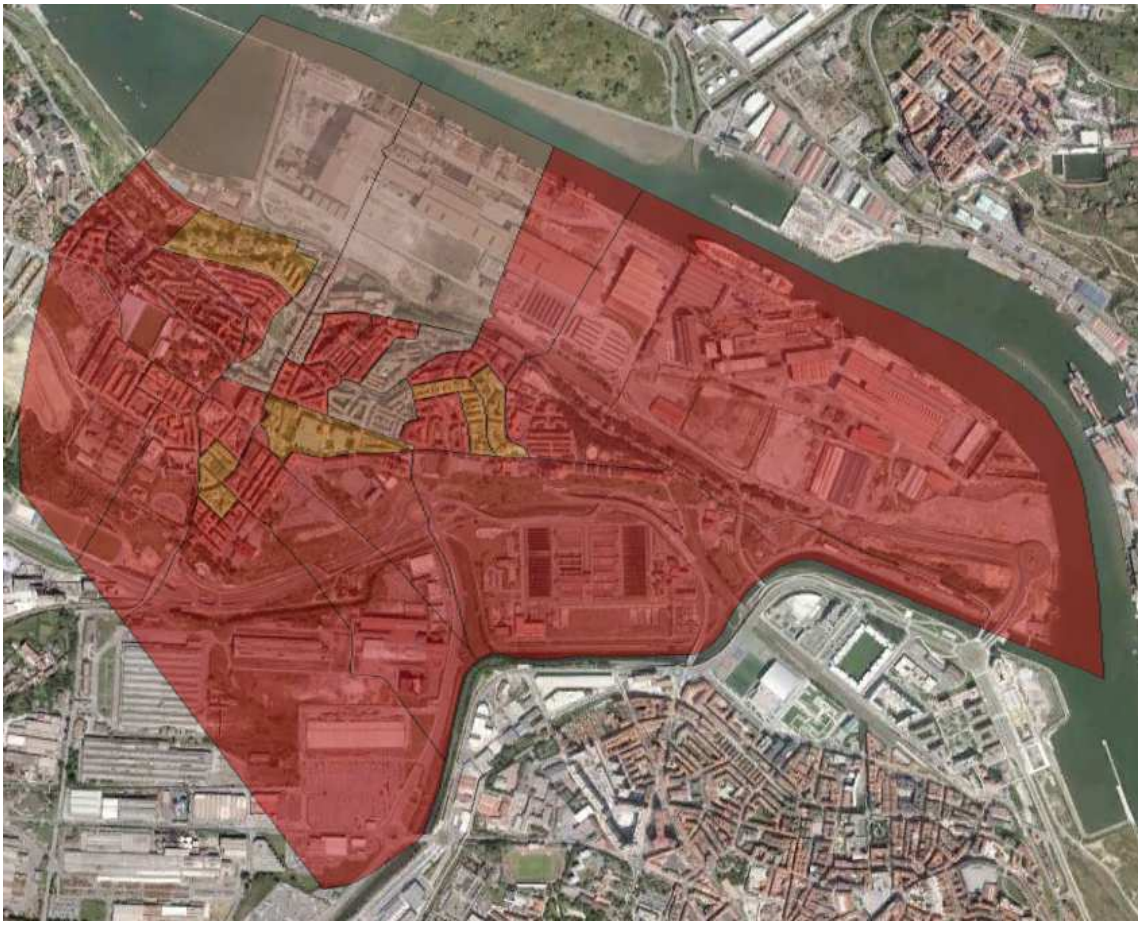
Portugalete



Santurtzi



Sestao



Tolosa

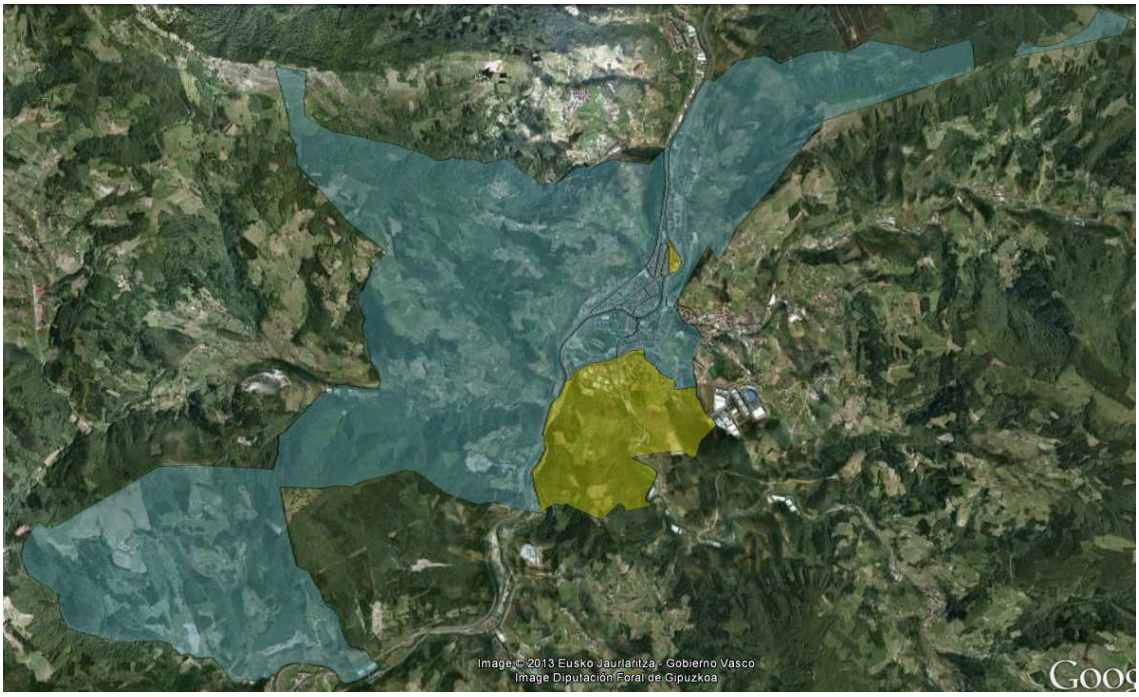
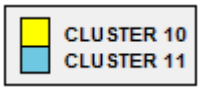
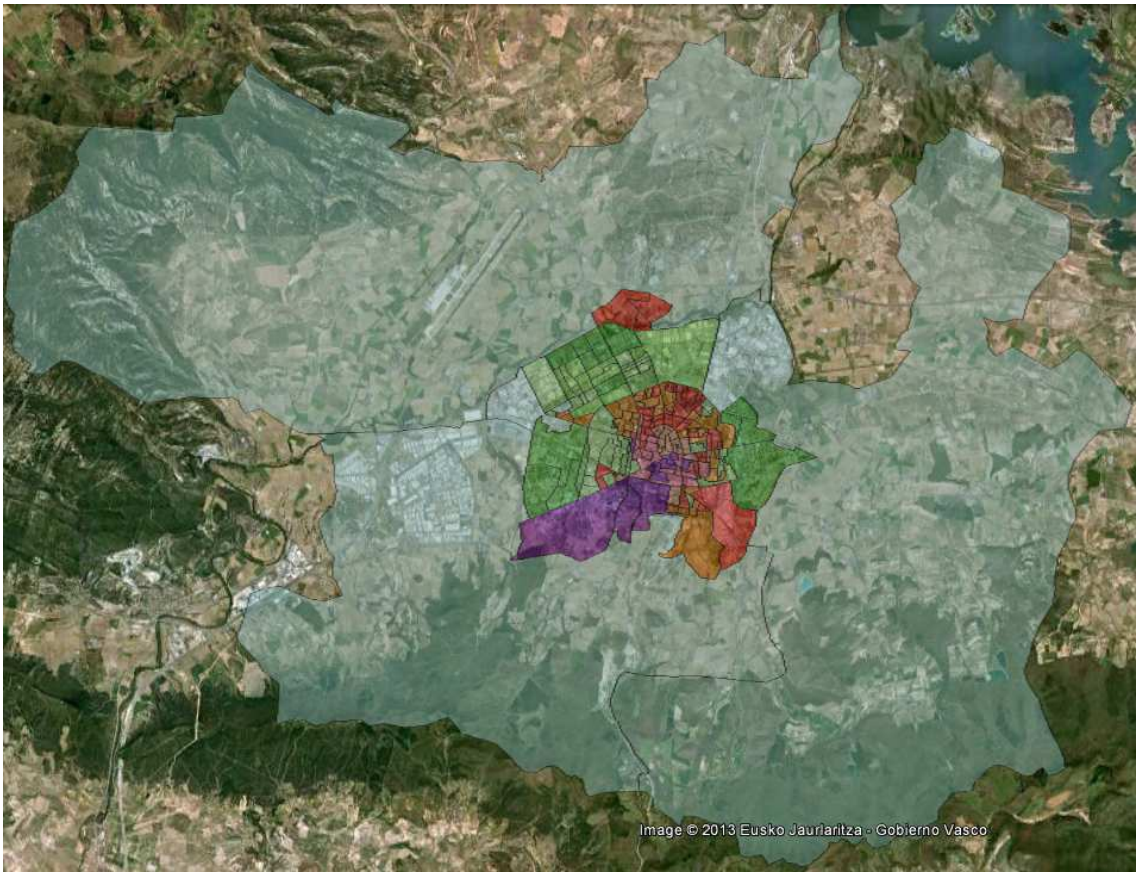
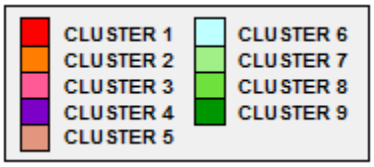


Image © 2013 Eusko Jaurlaritza - Gobierno Vasco
Image Diputación Foral de Gipuzkoa

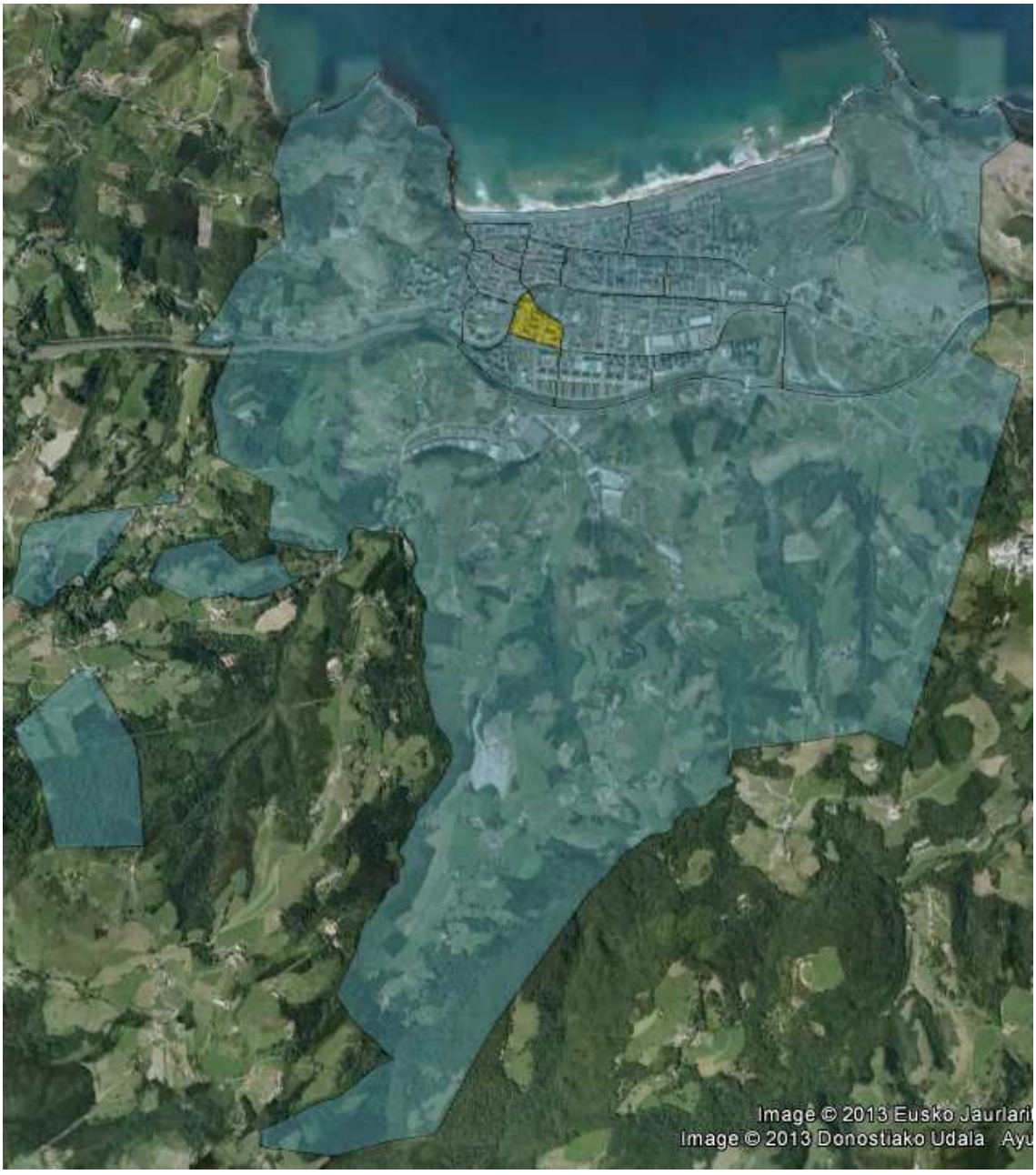

CLUSTER 10
CLUSTER 11

Vitoria-Gasteiz





Zarautz



CLUSTER 10
CLUSTER 11

Chapter
5

References

[1] http://www.r-project.org

[2] ESCOFIER, B. AND PAGÈS, J. (1992)

Multiple and simple factor analysis. Objectives, methods and interpretation. Euskal Herriko Unibertsitatea, Argitalpen zerbitzua.

[3] LÊ, S.; JOSSE, J. AND HUSSON, F. (March 2008)

FactoMineR: An R Package for Multivariate Analysis. Journal of Statistical Software, Volume 25, Issue 1.

[4] HUSSON, F.; JOSSE, J.; LÊ, S. AND MAZET, J. (February 14, 2012)

FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R. Version 1.18. http://cran.es.r-project.org