

DSpace da Universidade de Santiago de Compostela

<http://dspace.usc.es/>

Instituto da Lingua Galega

---

Soraya Suárez Quintas / Laura Calaza Díaz (2015): “Dialectoloxía perceptiva: procesamento e análise de datos”. *Empanada break XIII*. Santiago de Compostela, 11 decembro 2015.

---



You are free to copy, distribute and transmit the work under the following conditions:

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Non commercial** — You may not use this work for commercial purposes.

# Dialectoloxía perceptiva: procesamento e análise de datos

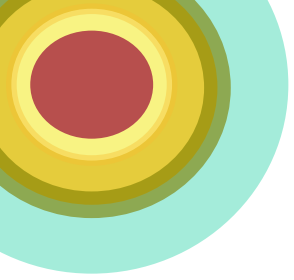
Soraya Suárez Quintas - Laura Calaza Díaz (ILG)

Esta presentación foi financiada pola Secretaría Xeral de Universidades (Xunta de Galicia – Fondo Europeo de Desenvolvemento Rexional) ó grupo de investigación Filoloxía e Lingüística Galega (USC).



INSTITUTO DA LINGUA GALEGA

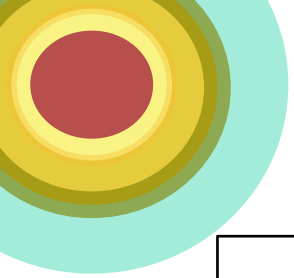




# PERCEPCIÓN

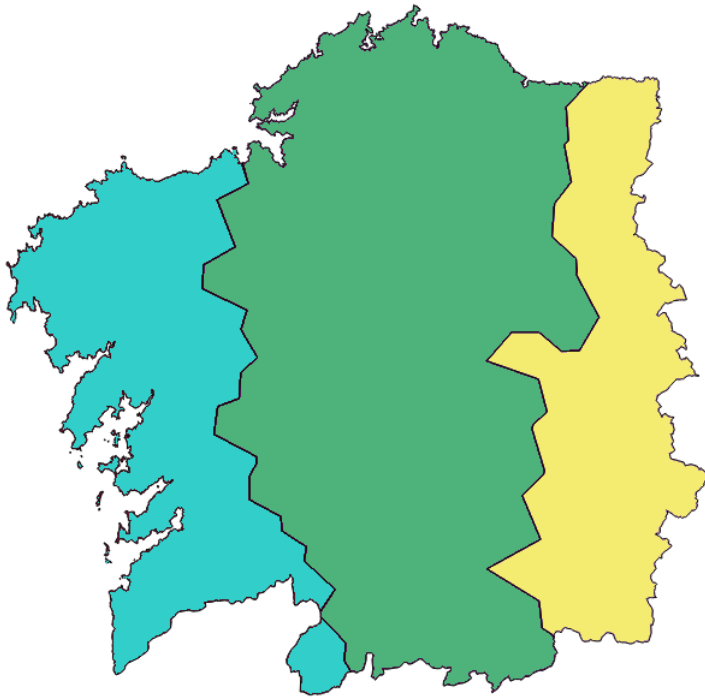
---



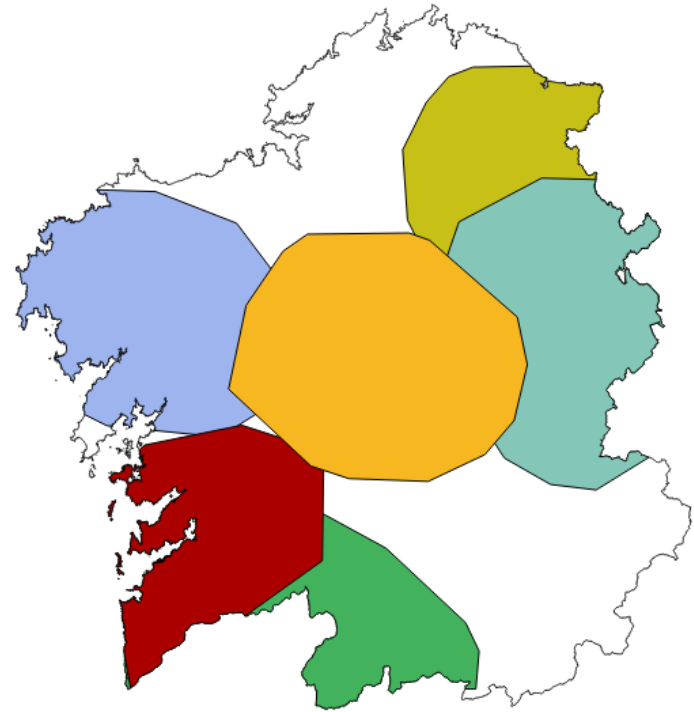


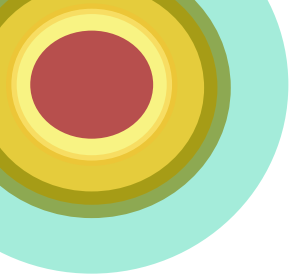
# PERCEPCIÓN

## DIALECTOLOGÍA TRADICIONAL



## DIALECTOLOGÍA PERCEPTIVA





# OBXECTIVOS

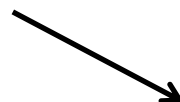


Descubrir e analizar...



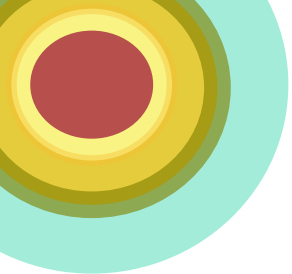
**A IDENTIFICACIÓN E  
XEORREFERENCIACIÓN DOS  
DIFERENTES DIALECTOS  
DUNHA OU VARIAS LINGUAS**

AS AVALIACIÓNS  
VALORATIVAS QUE REALIZAN  
OS FALANTES SOBRE AS  
VARIÉDADES DIALECTAIS



COMPARAR OS  
RESULTADOS COS  
DATOS DA  
DIALECTOLOXÍA  
TRADICIONAL

IDENTIFICAR OS  
FACTORES QUE  
INFLÚEN NO  
RECOÑECIMENTO DAS  
ÁREAS DIALECTAIS



# METODOLOXÍA



O  
B  
J  
E  
C  
T  
I  
V  
O  
S

**IDENTIFICACIÓN E  
XERREFERENCIACIÓN DOS  
DIFERENTES DIALECTOS DUNHA  
LINGUA**

AVALIACIÓNS VALORATIVAS QUE  
REALIZAN OS FALANTES SOBRE  
AS VARIEDADES DIALECTAIS

Deseño de mapas

Identificación  
dialectal

Gradación da  
diferenza

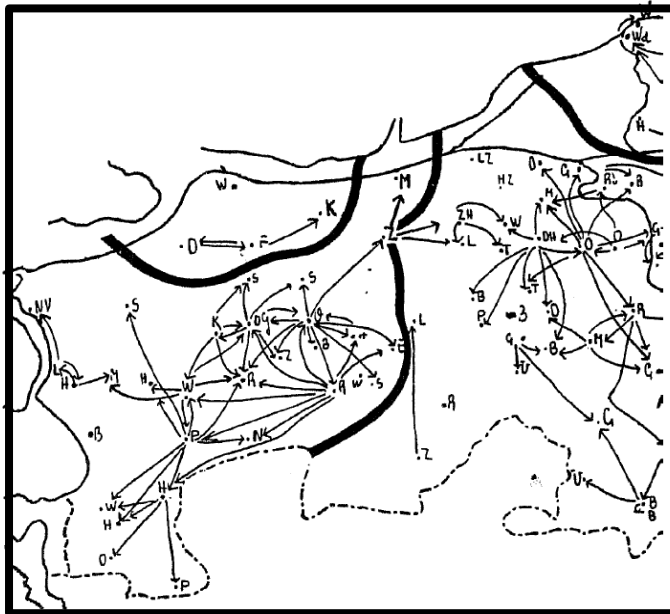
Corrección e  
agradabilidade

Recollida de datos  
cualitativos

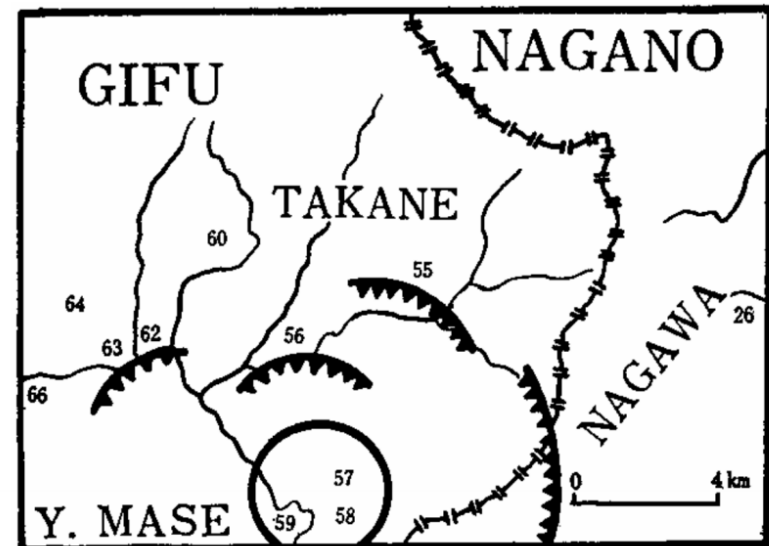
M  
E  
T  
O  
D  
O  
L  
O  
X  
Í  
A

# ESTUDOS PREVIOS

## DESEÑO DE MAPAS



Weijnen: 1946

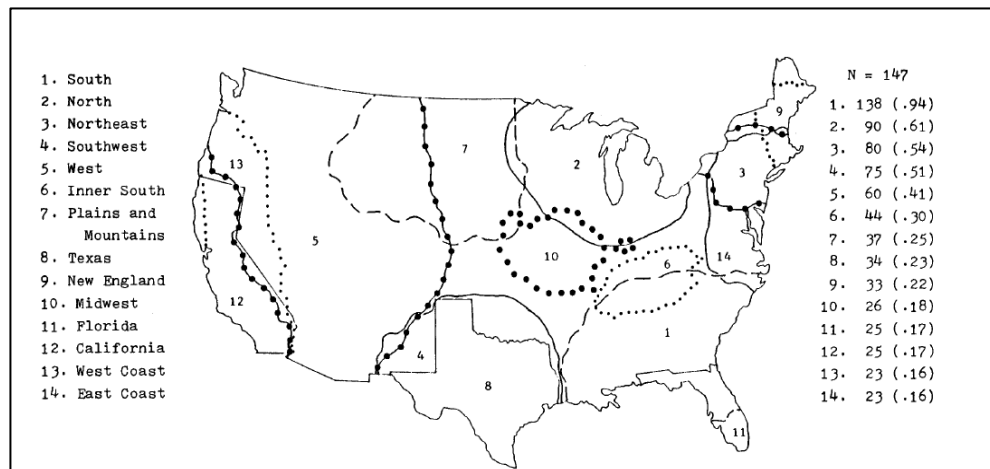
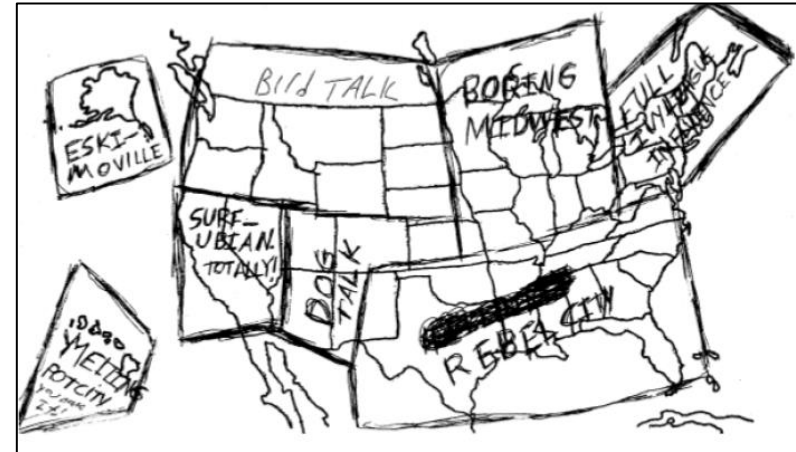
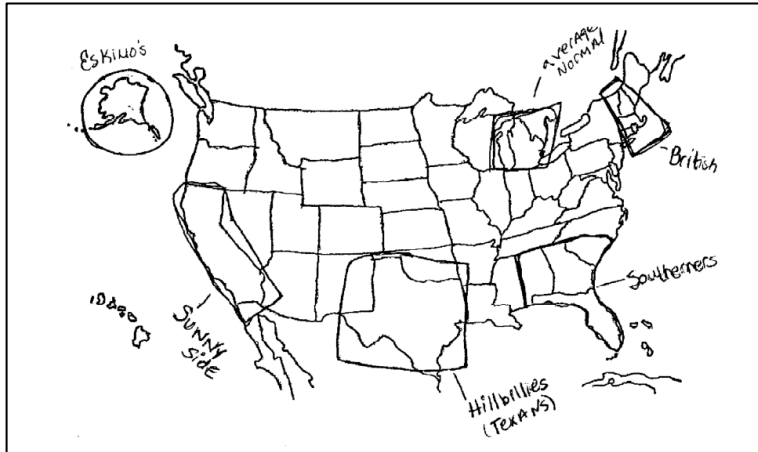


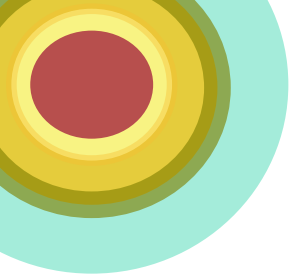
Mase: 1964



# ESTUDOS PREVIOS

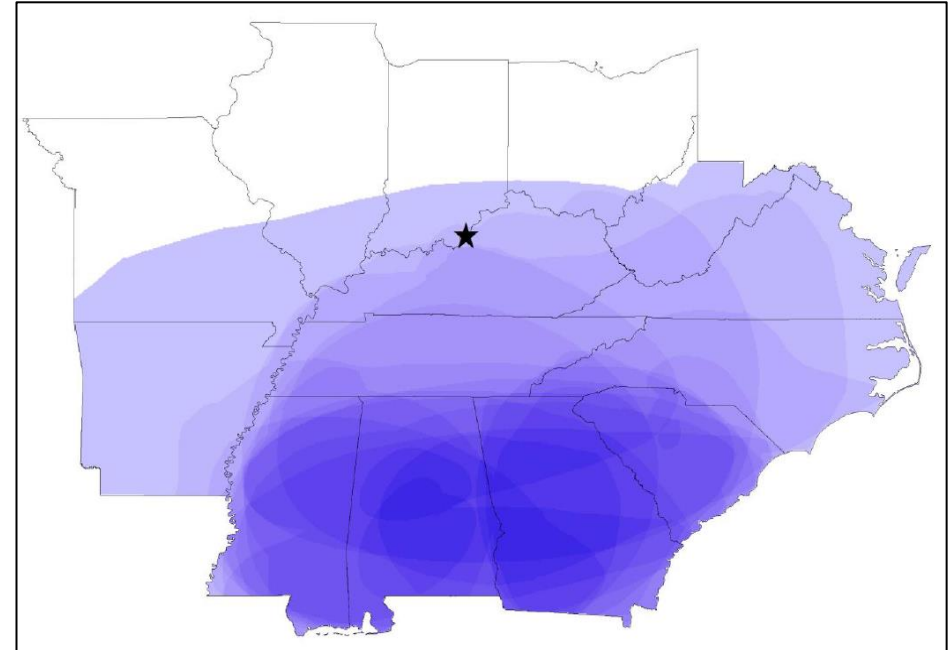
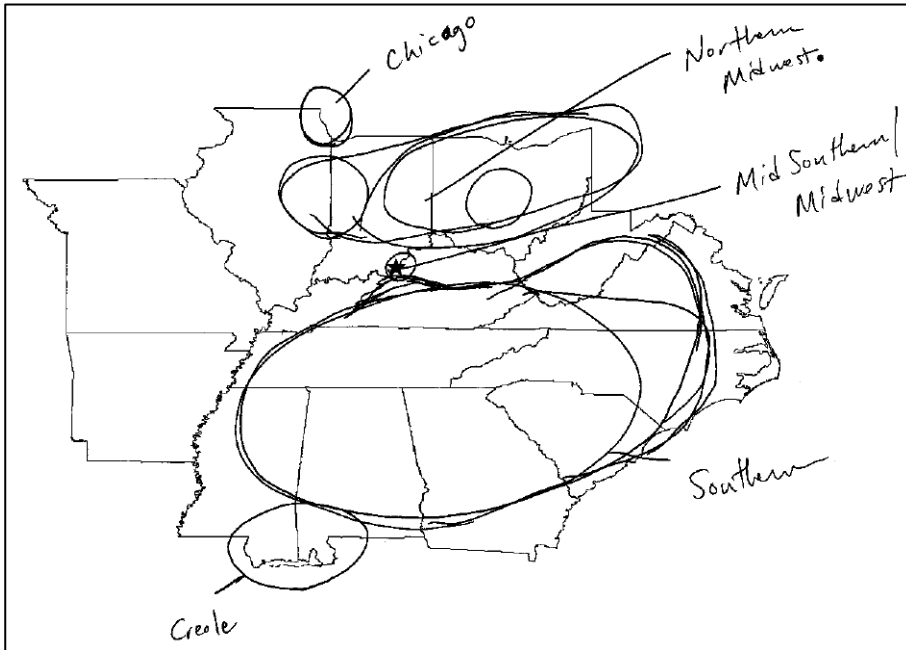
## DESEÑO DE MAPAS

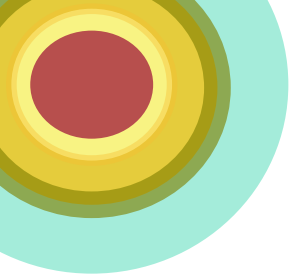




# ESTUDOS PREVIOS

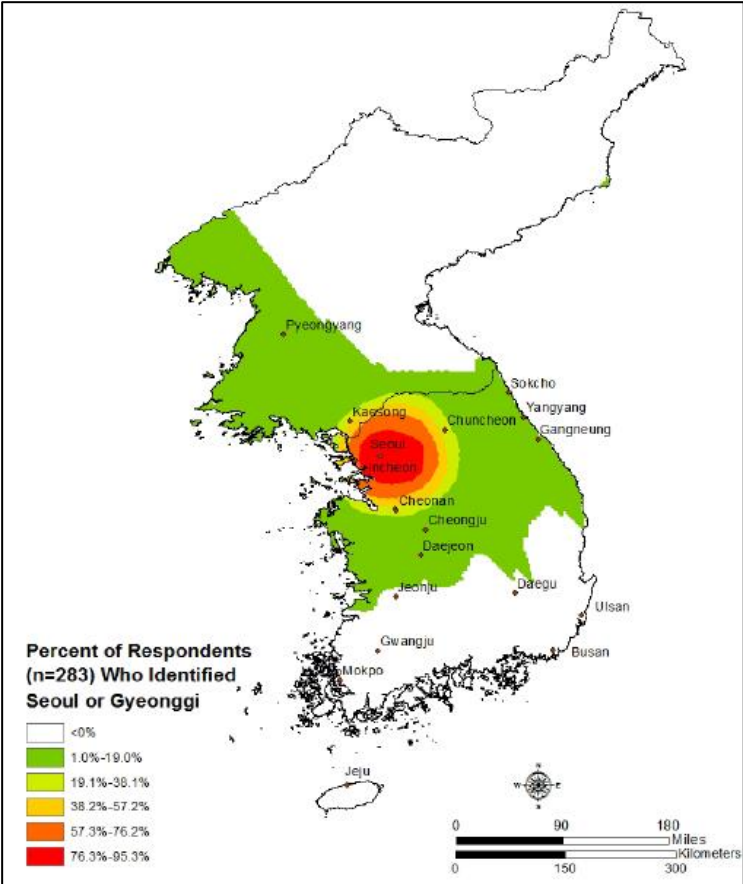
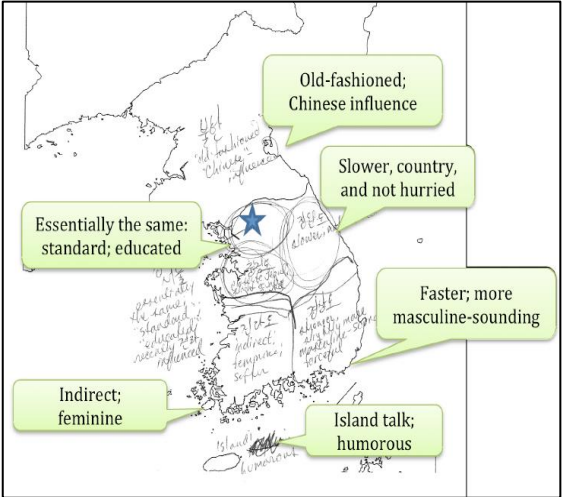
## DESEÑO DE MAPAS





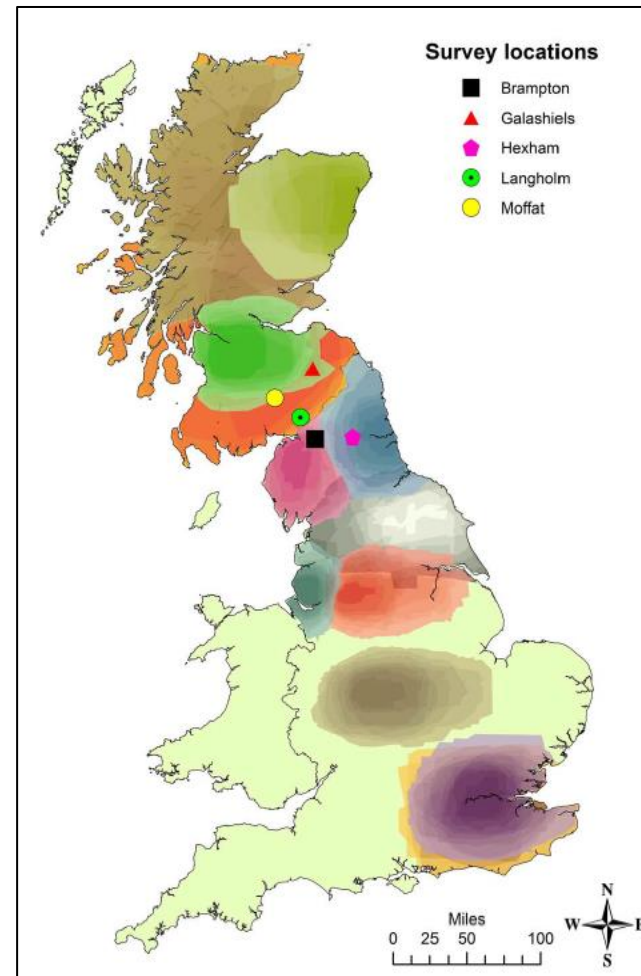
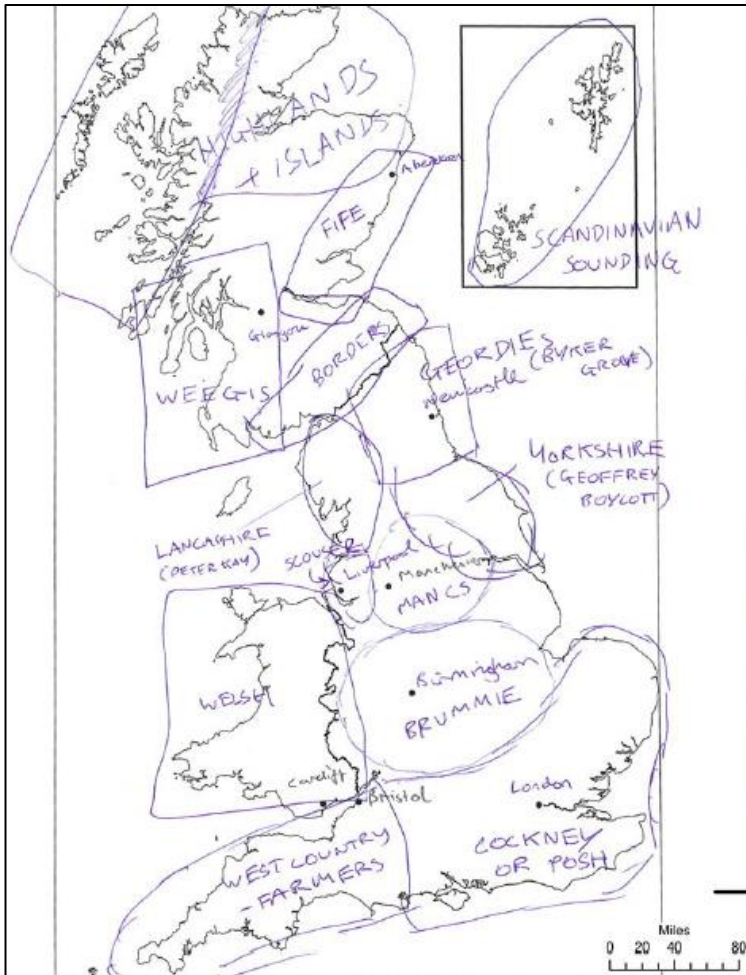
# ESTUDOS PREVIOS

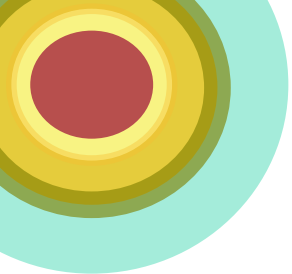
## DESEÑO DE MAPAS



# ESTUDOS PREVIOS

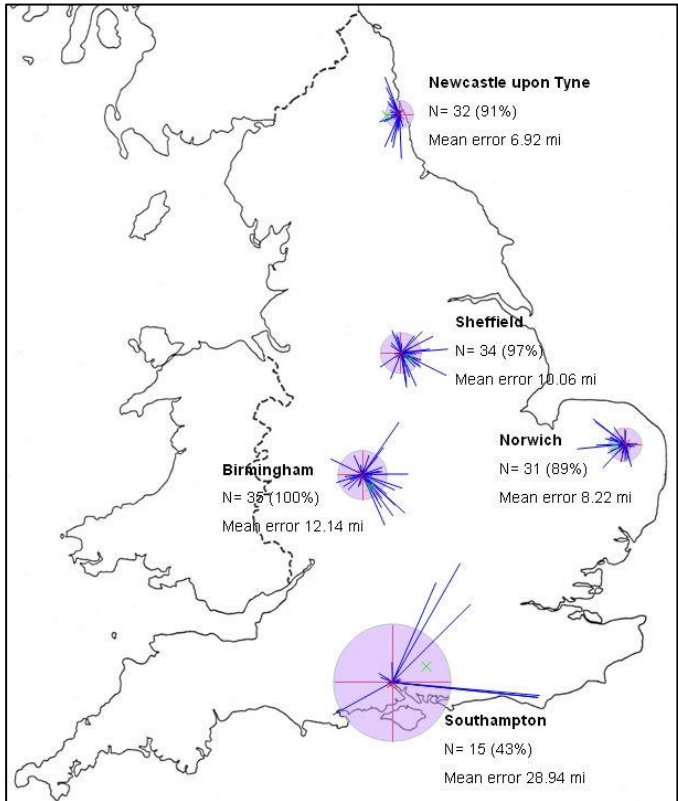
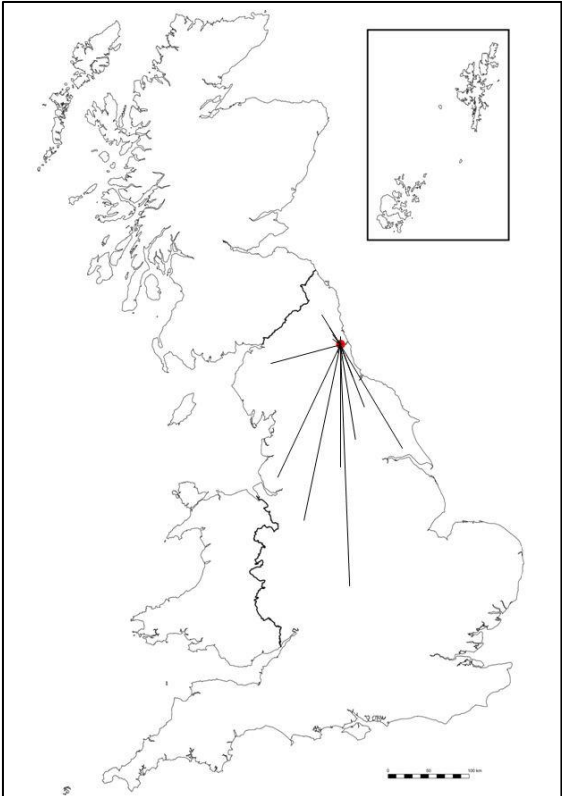
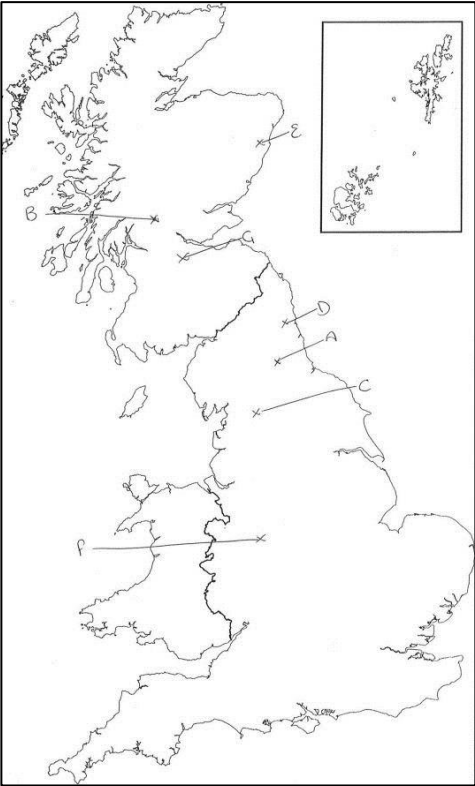
## DESEÑO DE MAPAS

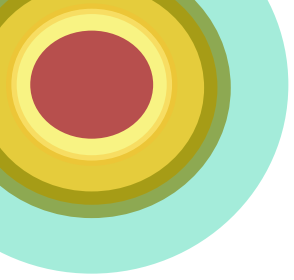




# ESTUDOS PREVIOS

## IDENTIFICACIÓN DIALECTAL





# RECOLLIDA DE DATOS

---

PROCEDENCIA ESTÍMULOS ORAIS



# RECOLLIDA DE DATOS

## APLICACIÓN INFORMÁTICA

Instituto da Lingua Galega - USC

Datos do informante

52

Ano de nacemento

Sexo  Home  Muller

Nivel de estudos

**Lugar de enquisa**

Provincia

Concello

Lugar

Outro

**Lugar de nacemento**

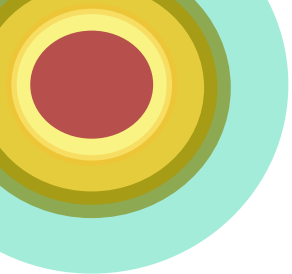
Igual ó lugar de enquisa

Provincia

Concello

Lugar

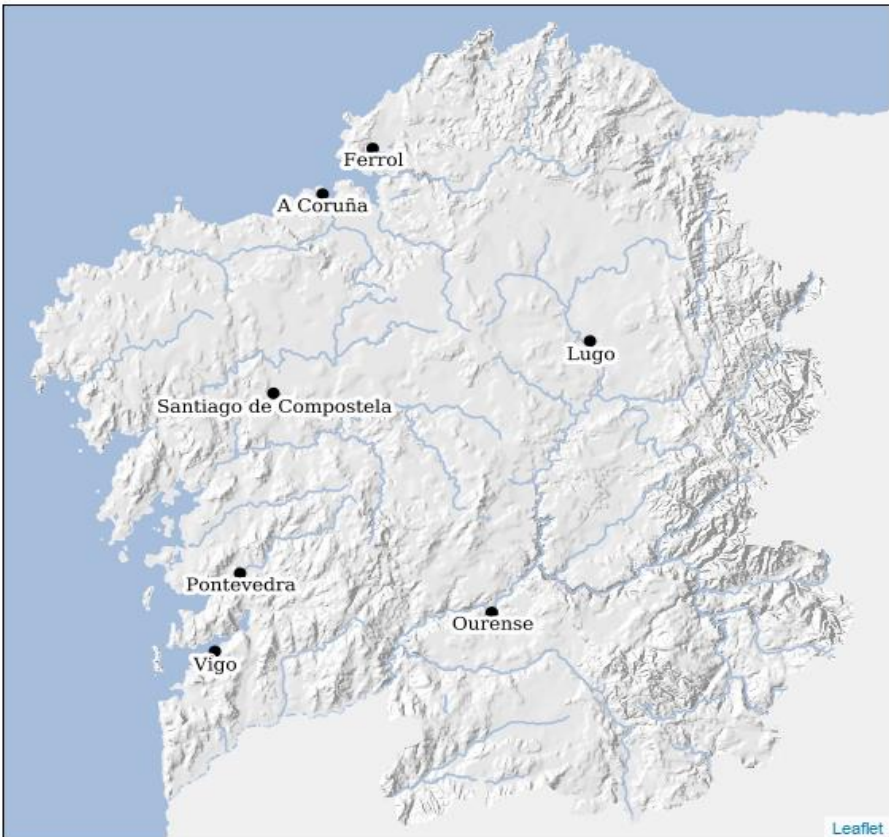
Outro



# RECOLLIDA DE DATOS

## APLICACIÓN INFORMÁTICA

Instituto da Lingua Galega - USC



A topographic map of Galicia, Spain, showing major cities and rivers. The cities labeled are Ferrol, A Coruña, Lugo, Santiago de Compostela, Pontevedra, Ourense, and Vigo. The map is credited to Leaflet.

**Audicións**

- 1 [Escoitar](#)
- 2 [Escoitar](#)
- 3 [Escoitar](#)
- 4 [Escoitar](#)
- 5 [Escoitar](#)
- 6 [Escoitar](#)
- 7 [Escoitar](#)

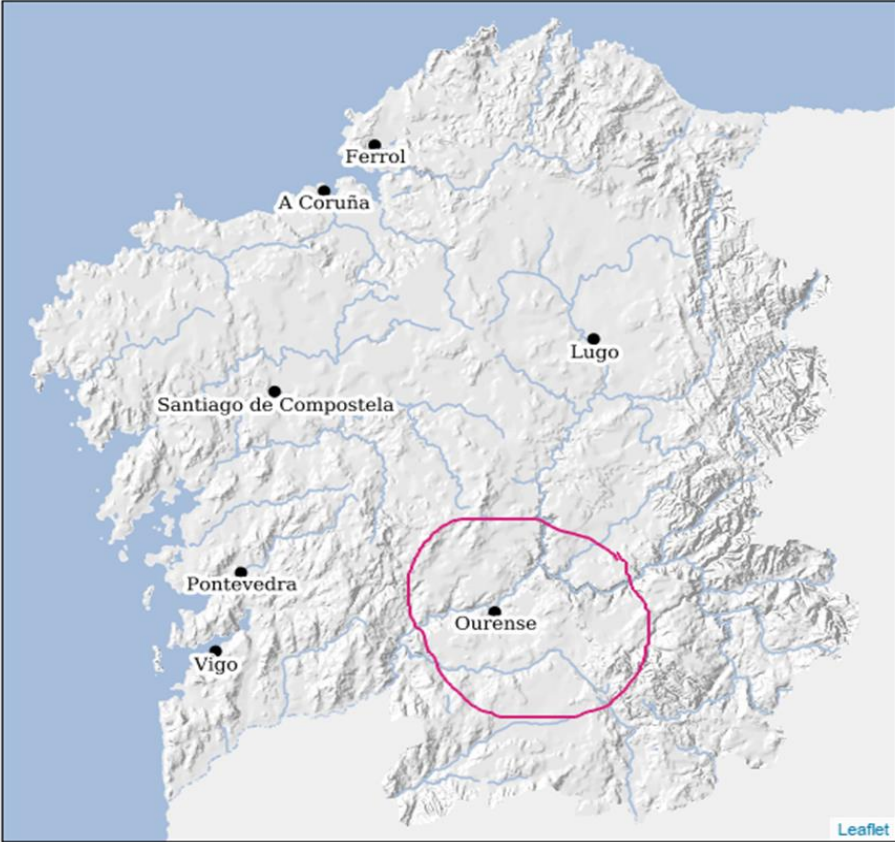
[Saír](#) [Atrás](#) [Continuar](#)



# RECOLLIDA DE DATOS

## APLICACIÓN INFORMÁTICA

Instituto da Lingua Galega - USC



A topographic map of Galicia, Spain, showing major cities: Ferrol, A Coruña, Santiago de Compostela, Pontevedra, Vigo, Lugo, and Ourense. A pink circle is drawn around the city of Ourense. The map is credited to Leaflet in the bottom right corner.

**Audicións**

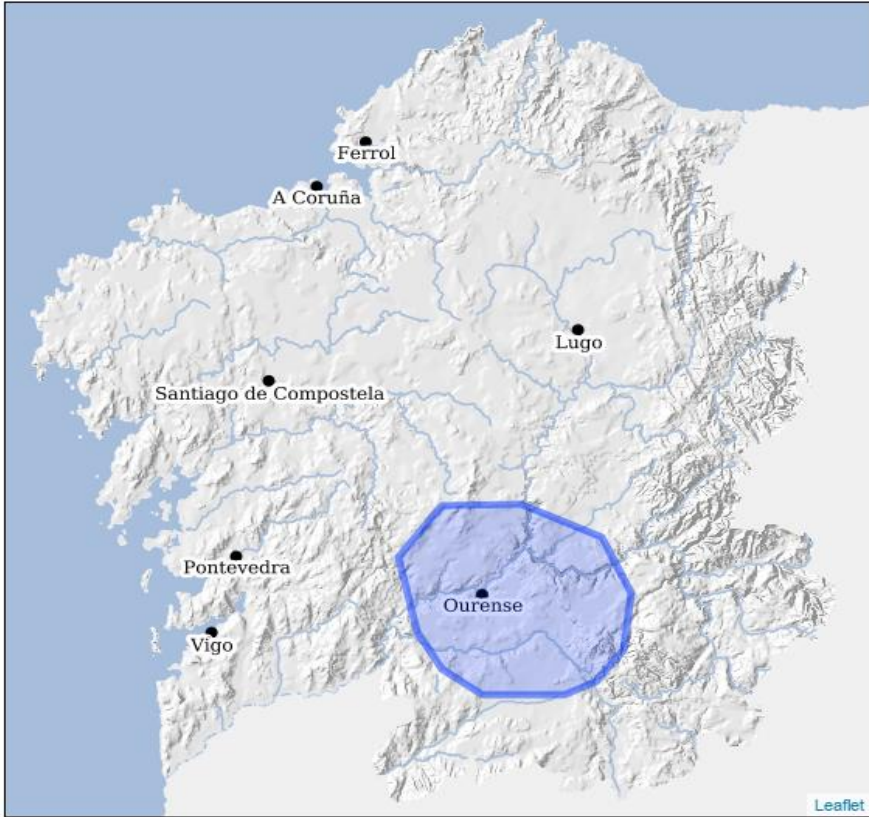
Debuxa a zona da audición no mapa

Cancelar

# RECOLLIDA DE DATOS

## APLICACIÓN INFORMÁTICA

Instituto da Lingua Galega - USC



The map shows the geographical outline of Galicia, Spain, with major cities labeled: Ferrol, A Coruña, Santiago de Compostela, Lugo, Pontevedra, Vigo, and Ourense. A blue polygon highlights the Ourense region. The map is credited to Leaflet.

**Audicións**

1

2

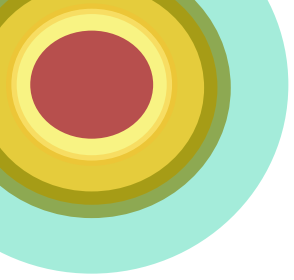
3

4

5

6

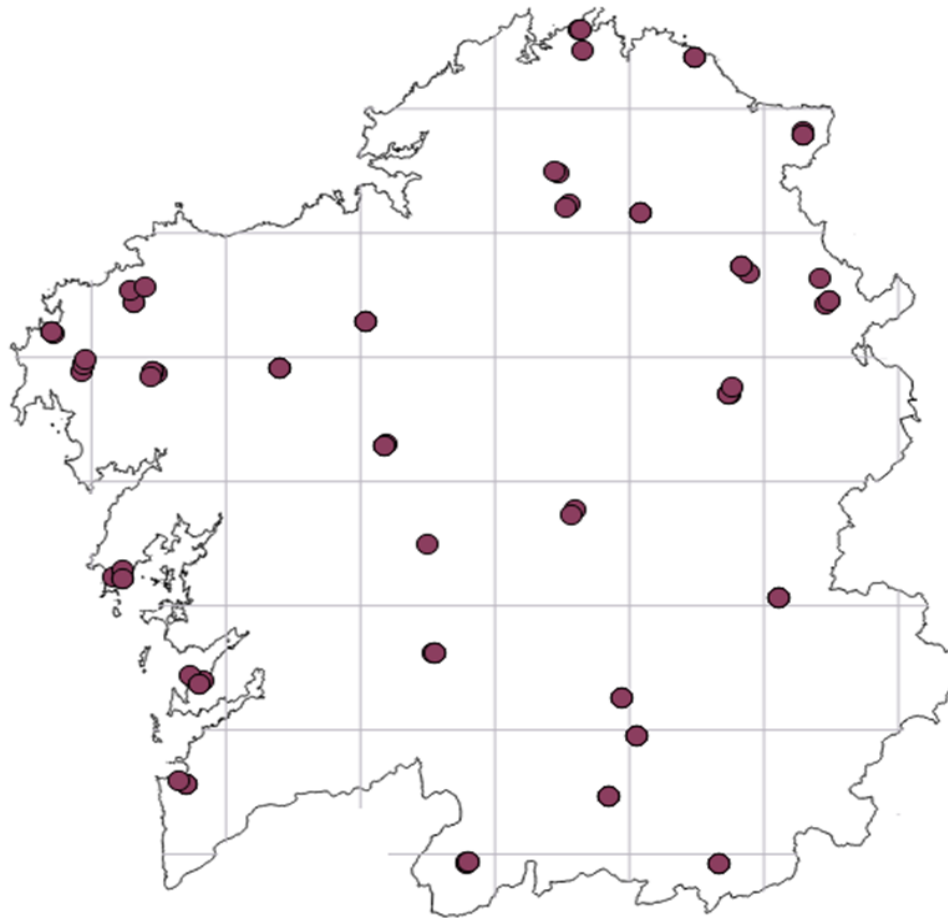
7



# RECOLLIDA DE DATOS

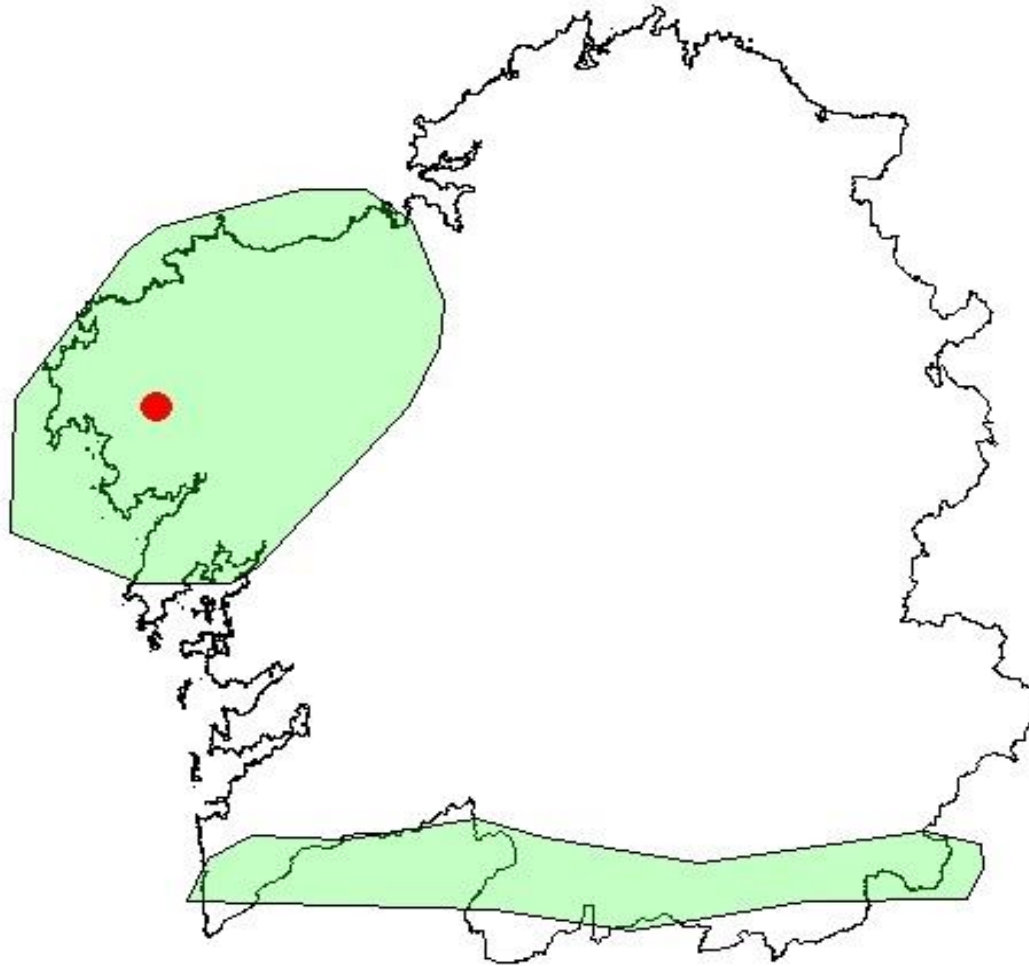
---

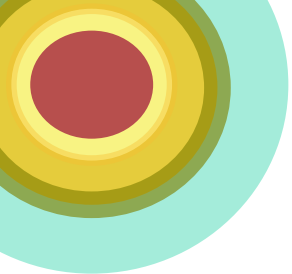
PUNTOS DE ENQUISA



# RECOLLIDA DE DATOS

DEMOSTRACIÓN 





# PROCESAMENTO DA INFORMACIÓN

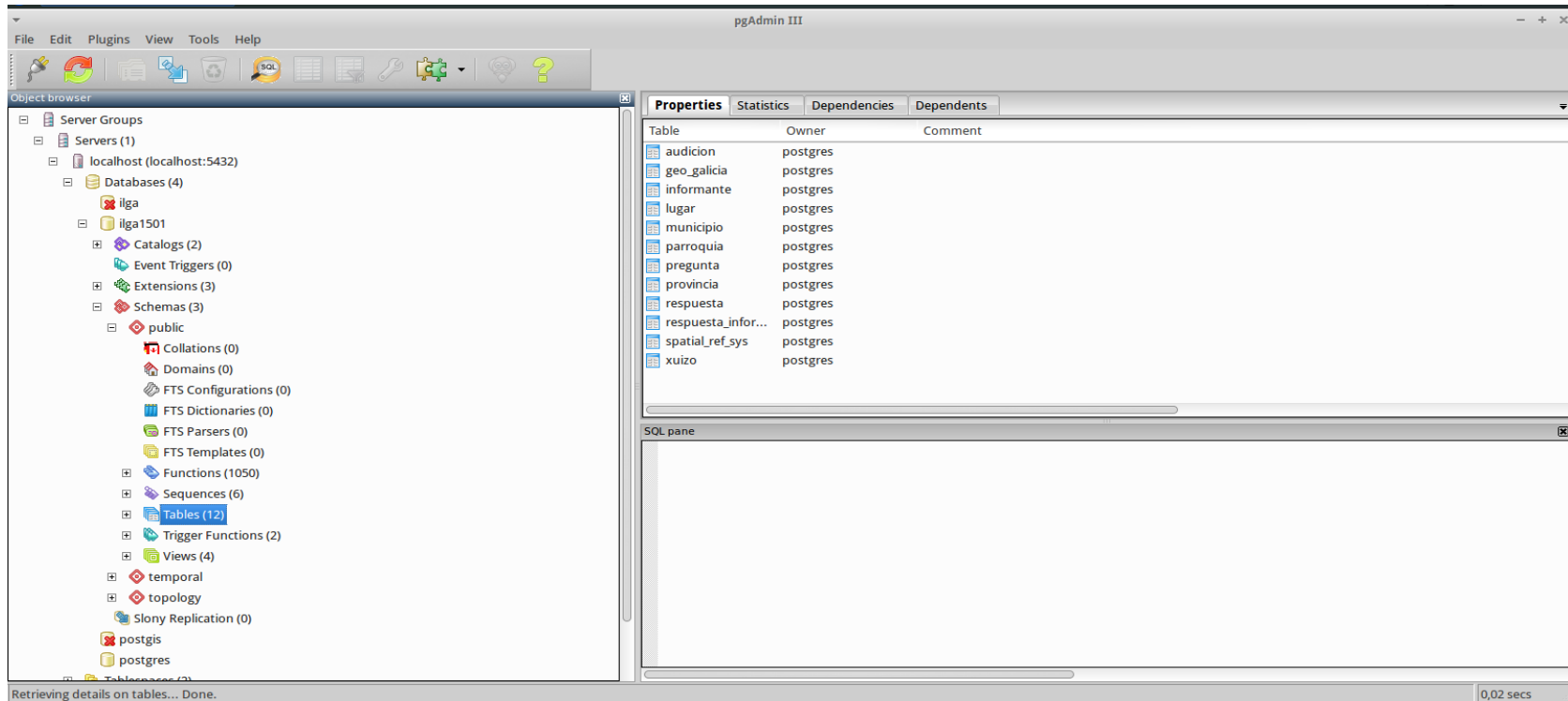
---

## BASE DE DATOS

- Localizacións das audicións
  - Datos dos informantes:
    - sexo
    - idade
    - nivel de estudos
    - lugar de nacemento
    - 7 xuízos (percepcións xeográficas dos dialectos)
  - 3 divisións do mapa de Galicia (Oriental, Central, Occidental)

# PROCESAMIENTO DA INFORMACIÓN

## POSTGRESQL (STRUCTURED QUERY LANGUAGE)



The screenshot displays the pgAdmin III interface. On the left, the 'Object browser' shows a tree view of the database structure, including 'Servers (1)', 'localhost (localhost:5432)', 'Databases (4)', and 'ilga1501'. The 'Tables (12)' folder is selected. On the right, the 'Properties' tab is active, showing a table list with columns 'Table', 'Owner', and 'Comment'. The table list contains the following entries:

Table	Owner	Comment
audicion	postgres	
geo_galicia	postgres	
informante	postgres	
lugar	postgres	
municipio	postgres	
parroquia	postgres	
pregunta	postgres	
provincia	postgres	
respuesta	postgres	
respuesta_infor...	postgres	
spatial_ref_sys	postgres	
xuizo	postgres	

The 'SQL pane' at the bottom is empty. The status bar at the bottom right indicates '0,02 secs'.

### Posibilidades:

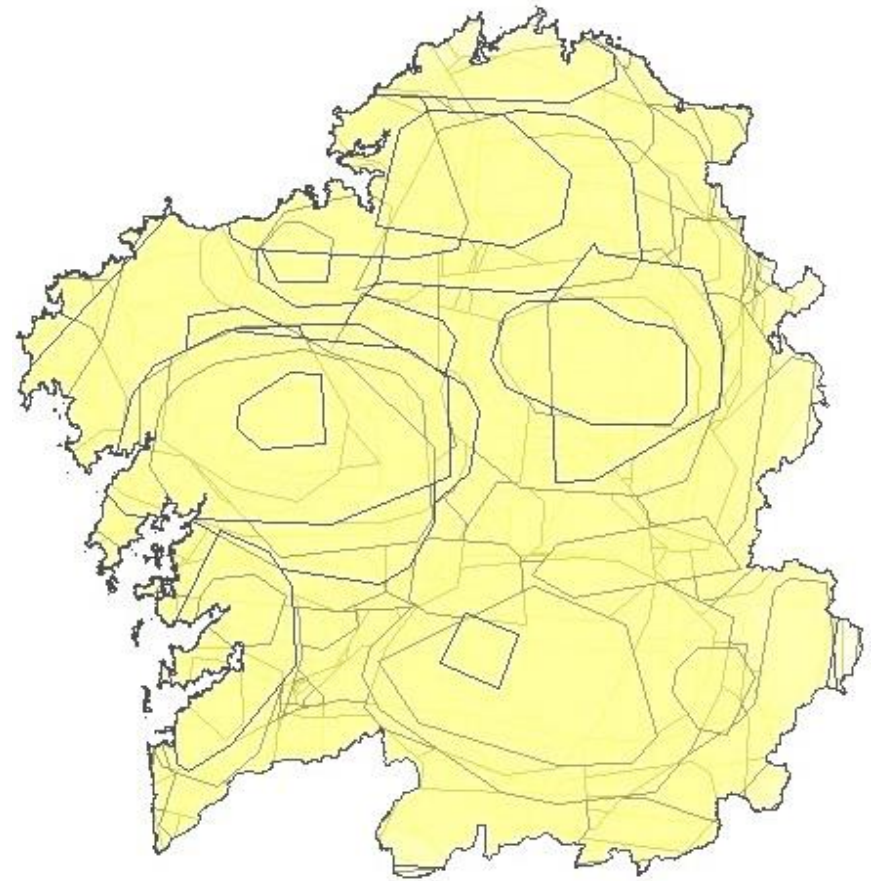
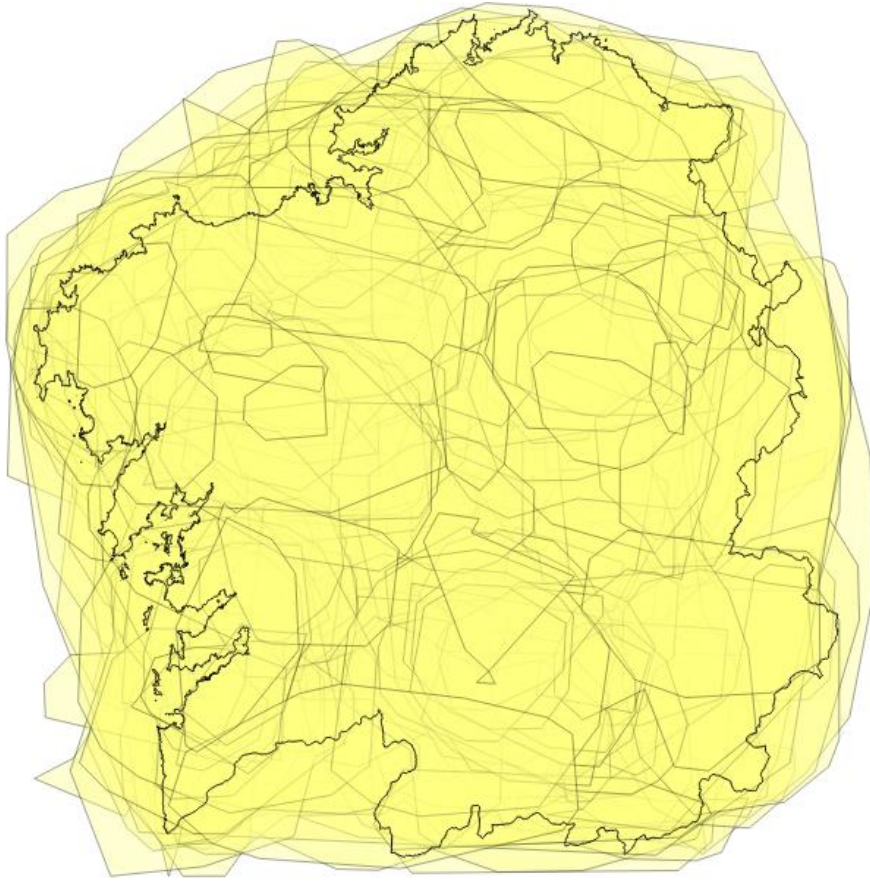
- Geographic Information System (GIS) - QGIS
- R project - Package RPostgreSQL

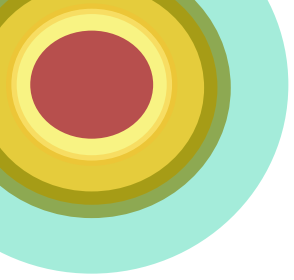


# PROCESAMIENTO DA INFORMACIÓN

---

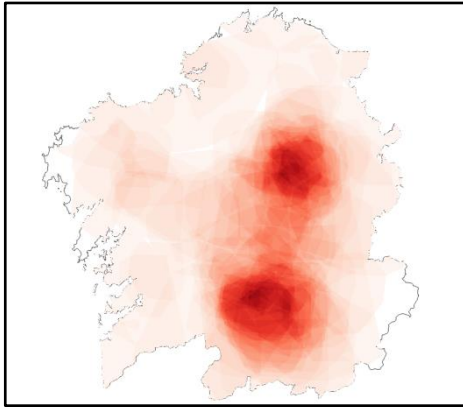
QGIS



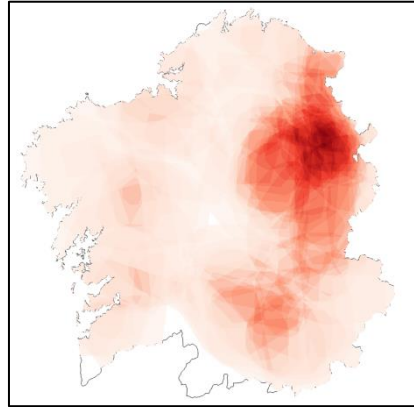


# PROCESAMENTO DA INFORMACIÓN

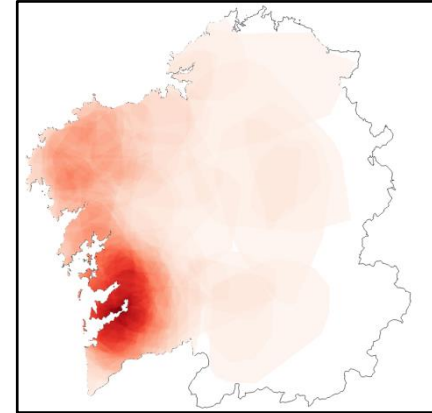
QGIS



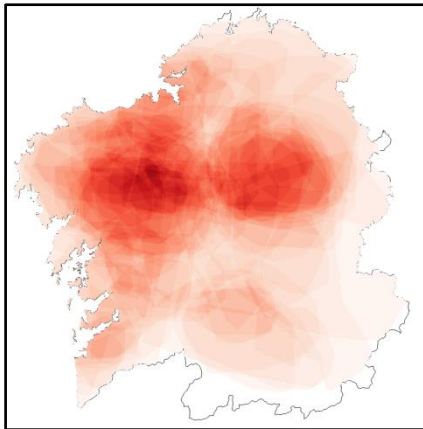
Xinzo de Limia



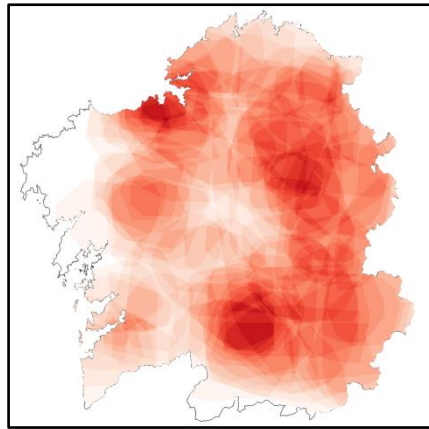
A Fonsagrada



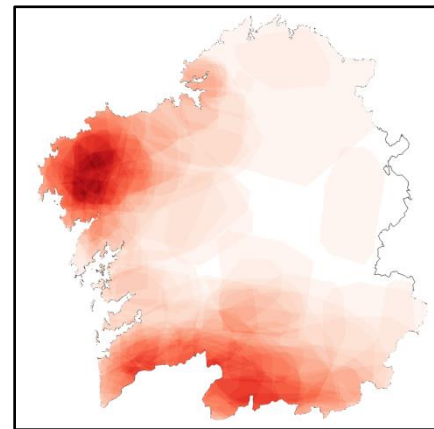
Bueu



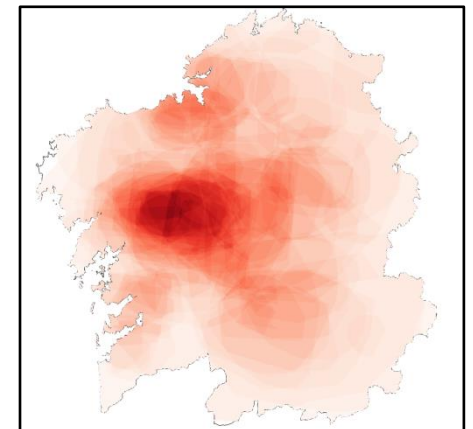
Moeche



O Barco de Valdeorras

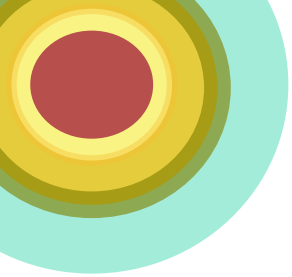


Mazaricos



Galego estándar





# PROCESAMIENTO DA INFORMACIÓN

---

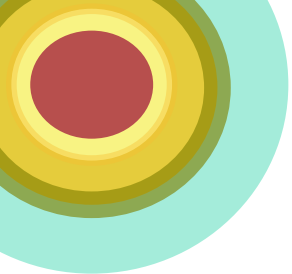
R PROJECT

Trataremos estes datos para:

- Identificar que factores inflúen no recoñecemento das variedades dialectais.
- Comparalos coa dialectoloxía tradicional.



**MEDIR DISTANCIAS ENTRE CONXUNTOS**

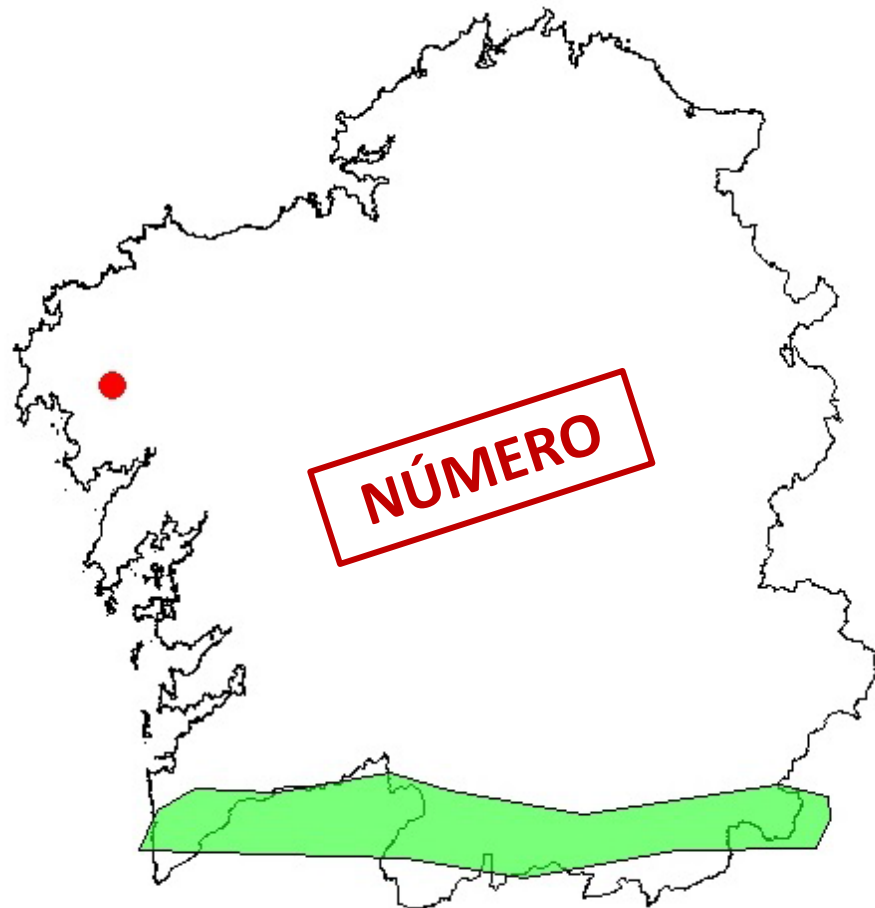


# PROCESAMIENTO DA INFORMACIÓN

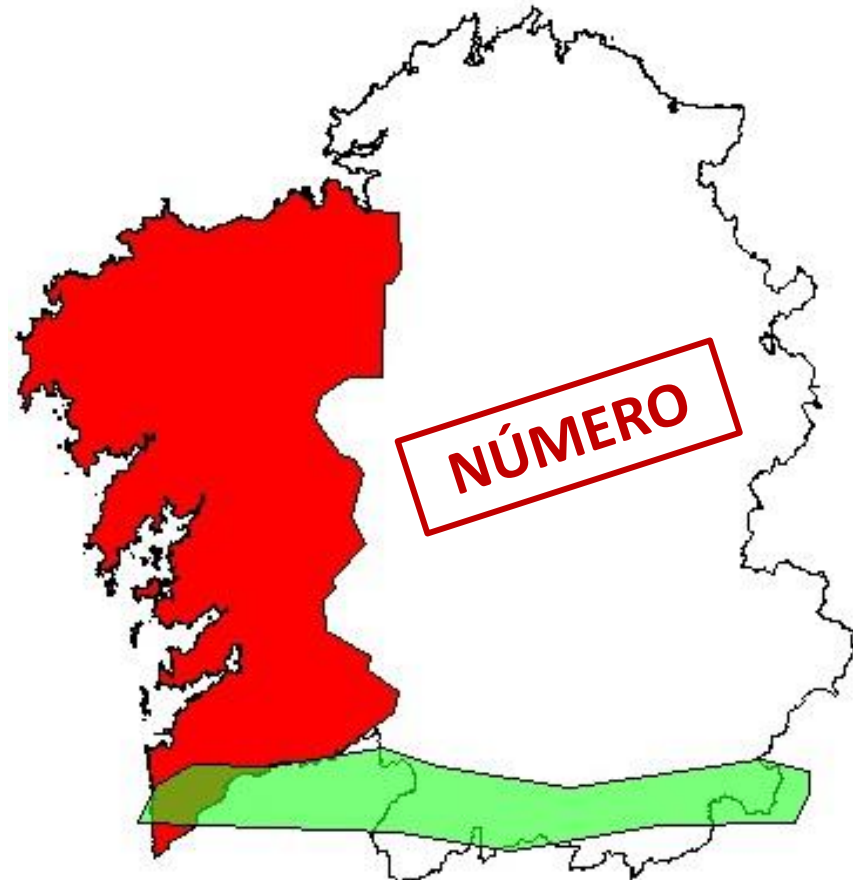
R PROJECT

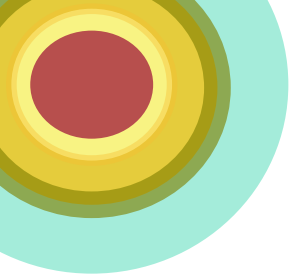
## MEDIR DISTANCIAS ENTRE CONXUNTOS

1º CASO



2º CASO





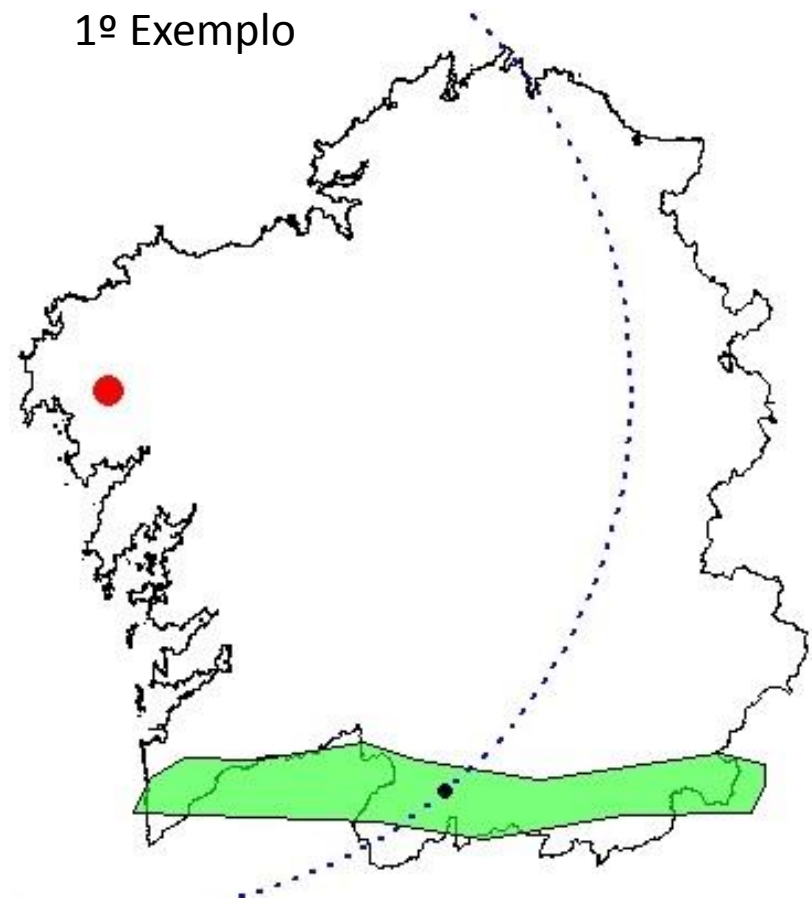
# PROCESAMIENTO DA INFORMACIÓN

R PROJECT

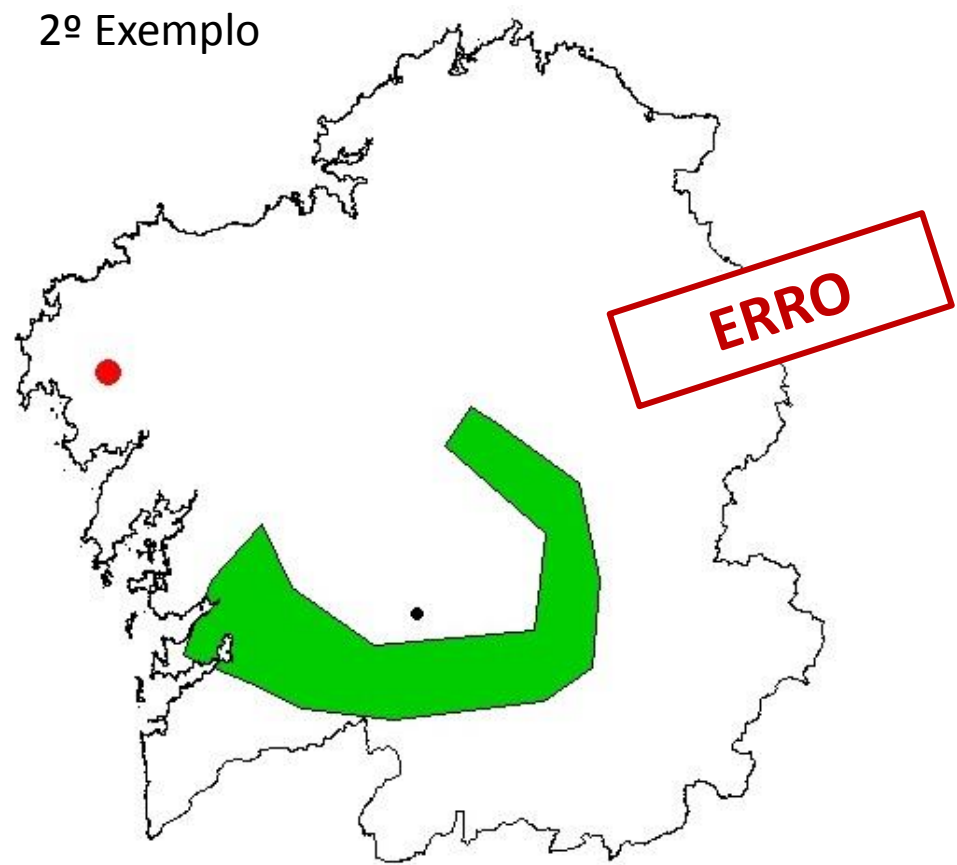
## ALTERNATIVAS DE DISTANCIAS ENTRE CONXUNTOS

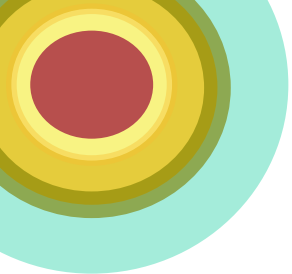
### 1. Centroide

1º Exemplo



2º Exemplo





# PROCESAMENTO DA INFORMACIÓN

R PROJECT

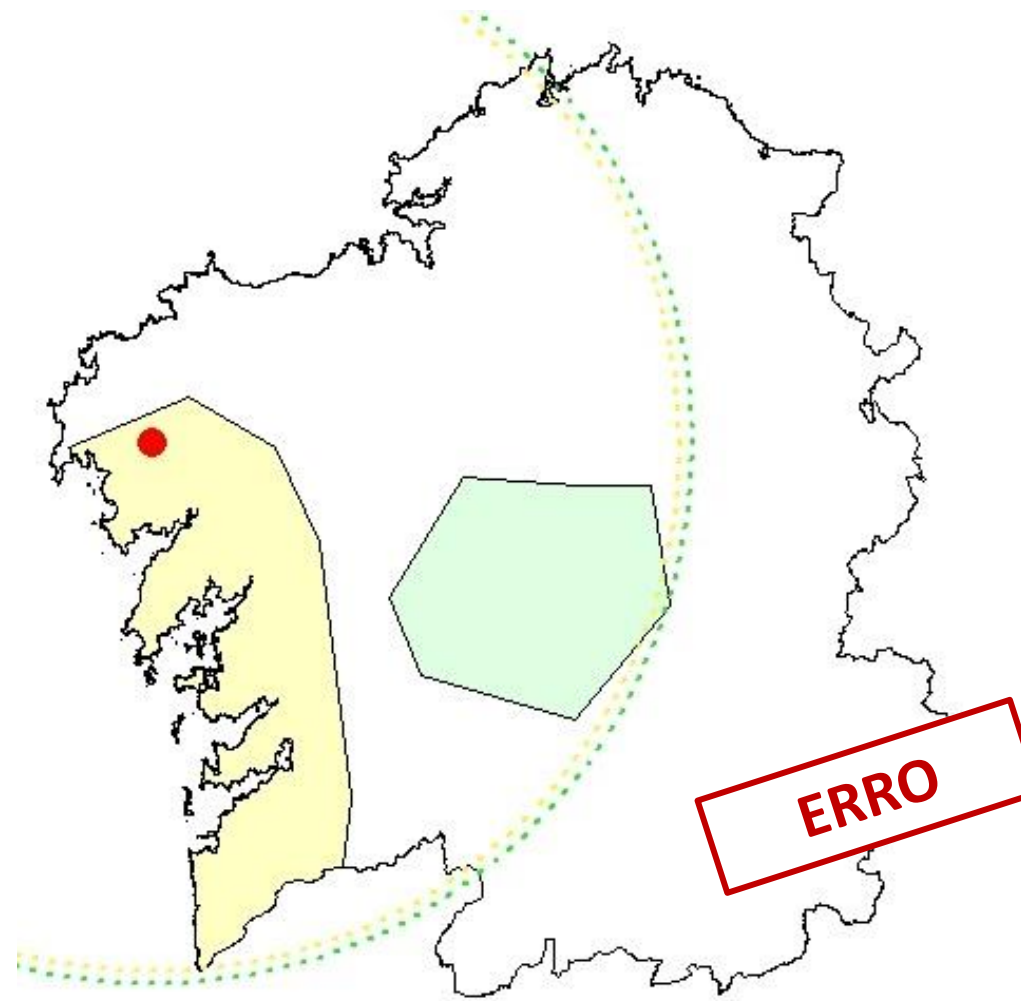
ALTERNATIVAS DE DISTANCIAS ENTRE CONXUNTOS

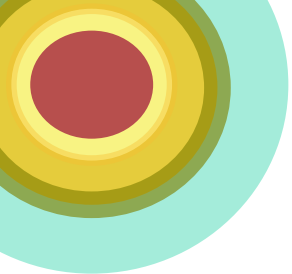
## 2. Hausdorff

Sexan os conxuntos  $A, B \subset X$ , entón

$$H(A; B) = \sup_{\{x \in A\}} |d(x, A) - d(x, B)|$$

con  $d(x, A) = \inf\{\rho(x, a), a \in A\}$ , e  $\rho(x, y)$  a distancia entre dous píxeles  $x, y \in X$ .





# PROCESAMIENTO DA INFORMACIÓN

R PROJECT

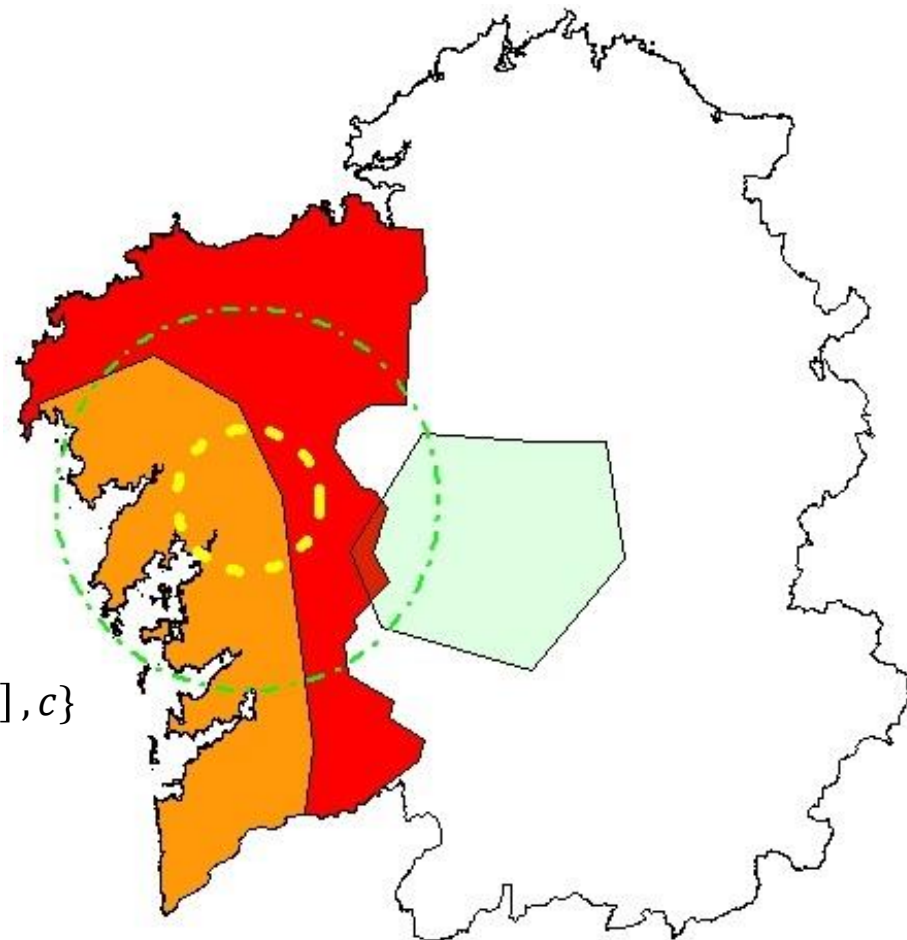
ALTERNATIVAS DE DISTANCIAS ENTRE CONXUNTOS

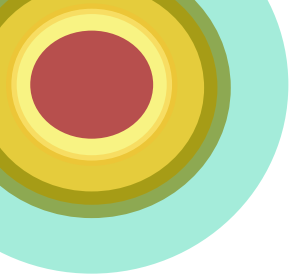
## 3. Baddeley

Sexan os conxuntos  $A, B \subset X$ , entón

$$\Delta_b(A, B) = \left[ \frac{1}{n(X)} \sum_{x \in X} |d^*(x, A) - d^*(x, B)|^p \right]^{\frac{1}{p}},$$

onde  $d^*(x, A) = \min\{d(x, A), c\} = \min\{\inf[d(x, a), a \in A], c\}$

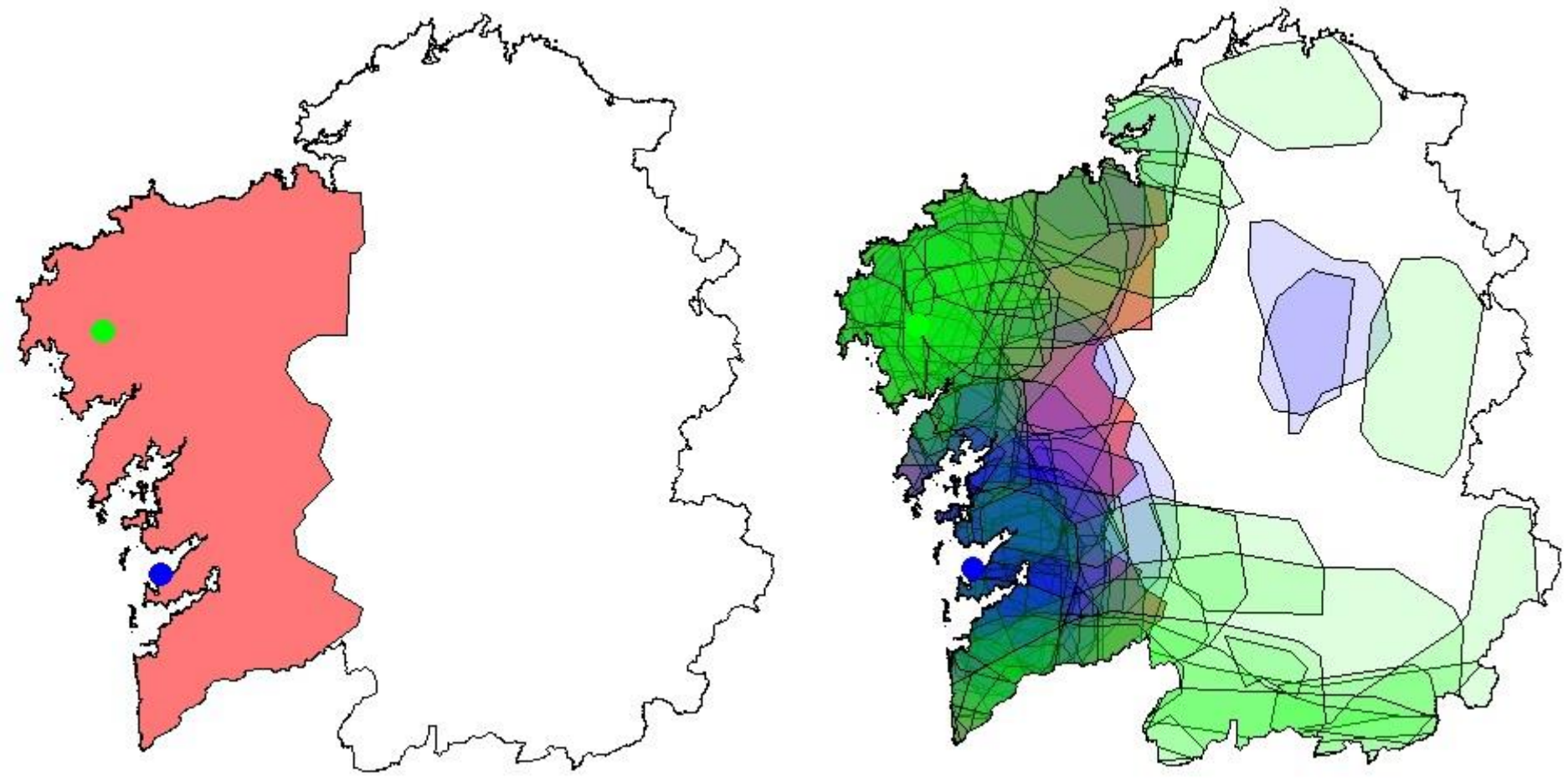


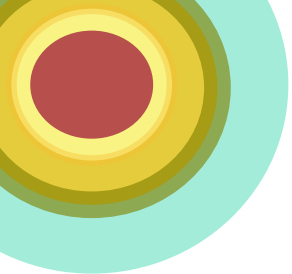


# PROCESAMENTO DA INFORMACIÓN

## RESULTADOS

### 1º OBXECTIVO: COMPARACIÓN COA DIALECTOLOXÍA TRADICIONAL

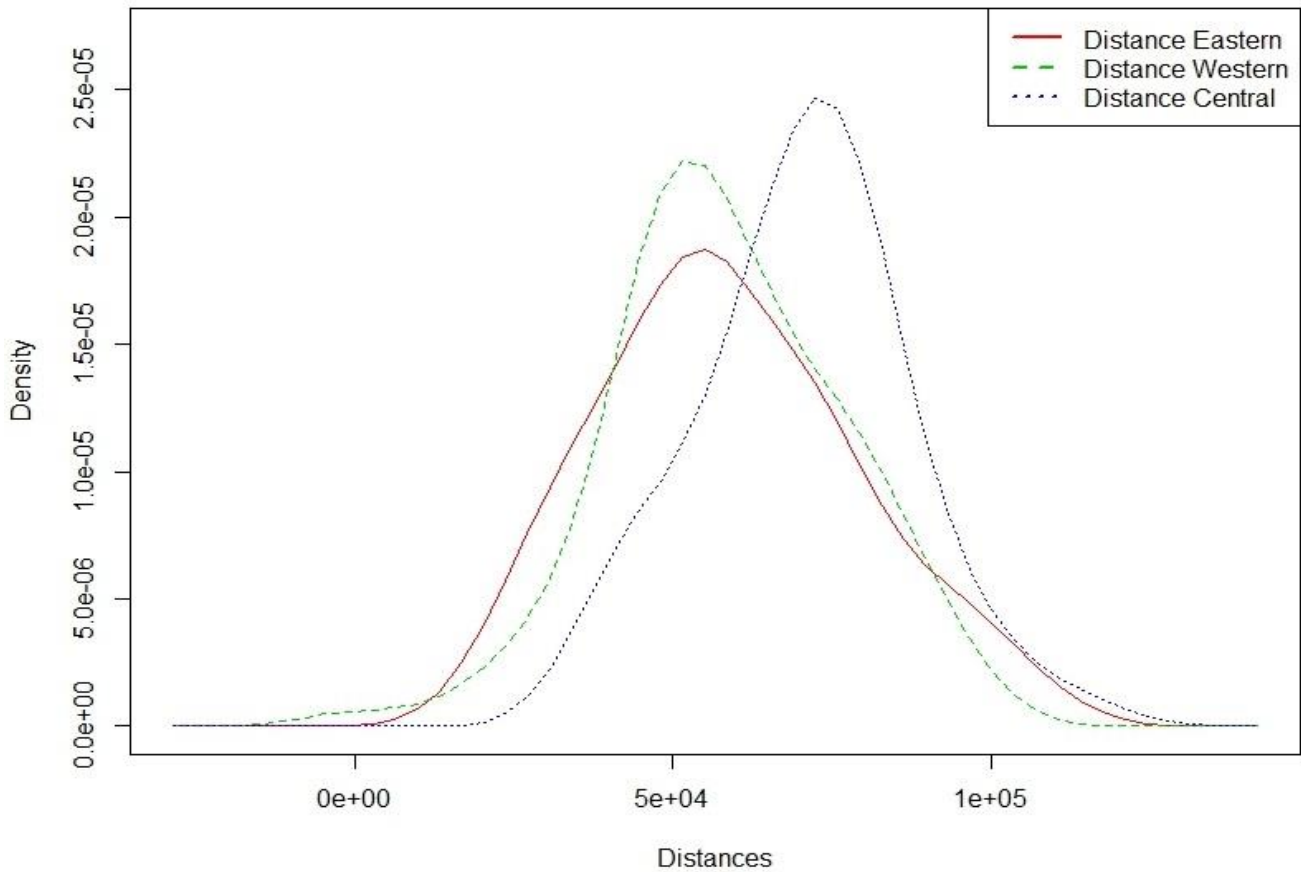




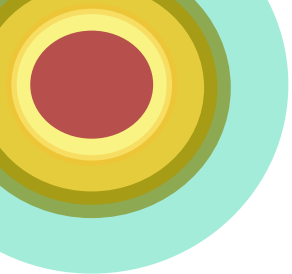
# PROCESAMIENTO DA INFORMACIÓN

## RESULTADOS

### 1º OBXECTIVO: COMPARACIÓN COA DIALECTOLOXÍA TRADICIONAL



- Test de normalidade (p.values > 0.05)  
Media bloq. Orient.: 55019:33 m  
Media bloq. Occid.: 54227:32 m  
Media bloq. Central : 71083:86 m
- Comparación das estimacións de densidades univariantes:  
Test de igualdade de distribucións  
p.value=0



# PROCESAMENTO DA INFORMACIÓN

## RESULTADOS

### 2º OBXECTIVO: IDENTIFICAR QUE FACTORES INFLÚEN NO RECOÑECIMENTO DAS VARIEDADES DIALECTAIS

#### MODELO DE REGRESIÓN (para cada audición)

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4 + \beta_5x_5 + \epsilon$$

#### Variable resposta (y):

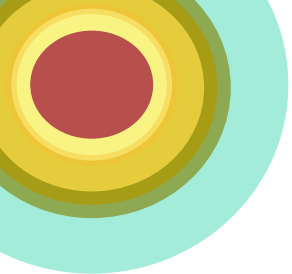
- Distancia entre a localización da audición e as seleccións feitas polos informantes (para cada audición)

#### Variables explicativas:

- Sexo ( $x_1$ )
- Idade ( $x_2$ )
- Nivel de estudos ( $x_3$ )
- Distancia entre o lugar de nacemento e o lugar de enquisa ( $x_4$ )
- Distancia entre o lugar de nacemento e o lugar de audición ( $x_5$ ) (para cada audición)

#### Erro ou residuo( $\epsilon$ )





# PROCESAMENTO DA INFORMACIÓN

## RESULTADOS

### 2º OBXECTIVO: IDENTIFICAR QUE FACTORES INFLÚEN NO RECOÑECIMENTO DAS VARIEDADES DIALECTAIS

#### MODELO DE REGRESIÓN MULTIVARIANTE (xeneralización)

$$Y = X\beta + \epsilon$$

Variable resposta ( $Y$ ):

$$Y = \sum_{\{i=1\}}^{\{n\}} w_i Y_{ik},$$

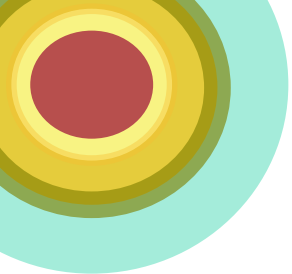
con  $w_i = \frac{1}{n}$ ,  $\forall i = 1, \dots, 6, k = 1 \dots, n$

(distancia media ponderada das distancias entre as localizacións das audicións e as seleccións feitas polos informantes)

Variables explicativas ( $X$ ):

- Sexo
- Idade
- Nivel de estudos
- Distancia entre o lugar de nacemento e o lugar de enquisa

Erro ou residuo( $\epsilon$ )



# PROCESAMIENTO DA INFORMACIÓN

## RESULTADOS

### 2º OBXECTIVO: IDENTIFICAR QUE FACTORES INFLÚEN NO RECOÑECIMENTO DAS VARIEDADES DIALECTAIS

#### MODELO DE REGRESIÓN MULTIVARIANTE (xeneralización)

lm(formula = baddeleydista ~ estudos, data = basepromedio)

Residuals:

Min	1Q	Median	3Q	Max
-21386.6	-5132.4	299.7	5512.6	23109.4

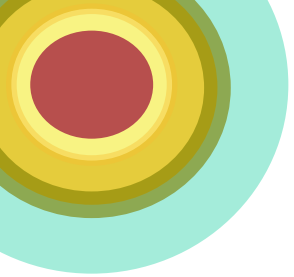
Coefficients:

	Estimate	Std. Error	t	value Pr(> t )
(Intercept)	59691	2693	22.168	< 2e-16 ***
estudosSECUNDARIOS	-5323	3517	-1.514	0.13682
estudosUNIVERSITARIOS	-11372	3376	-3.369	0.00151 **

Residual standard error: 9328 on 47 degrees of freedom

Multiple R-squared: 0.201, Adjusted R-squared: 0.167

F-statistic: 5.911 on 2 and 47 DF, p-value: 0.005132



# CONCLUSIÓN

---

## ANÁLISE INICIAL

- Os bloques occidental e oriental son os mellor identificados.
- A variable “Estudos” (estudos universitarios fronte a estudos primarios) é a que resulta significativa no recoñecemento das diferentes áreas dialectais do galego.



