

# VISUALIZACIÓN XEOGRÁFICA DE DATOS DE ONOMÁSTICA CON R

María José Ginzo Villamayor

Departamento de Estatística e Investigación Operativa  
Universidade de Santiago de Compostela



DEPARTAMENTO DE ESTATÍSTICA  
E INVESTIGACIÓN OPERATIVA



## Motivación

### Shiny

Que é Shiny?

Representación de nomes con Shiny

Representación de apelidos con Shiny

Unha mirada máis profunda en certos apelidos

Motivación e obxectivos

Modelos espaciais

Representación con Shiny

## └ Motivación

## └ Nomes

## Estatísticas oficiais

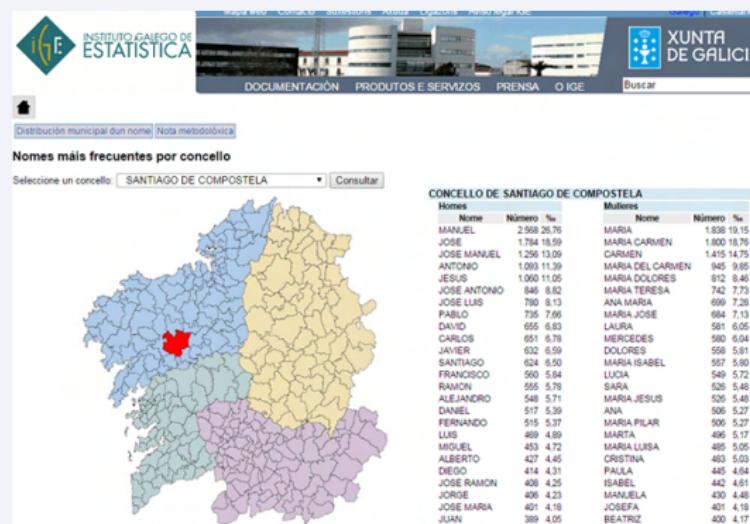


Figura: Nomes más frecuentes por concello  
(<http://www.ige.eu>)

## Visualizador interactivo Estados Unidos

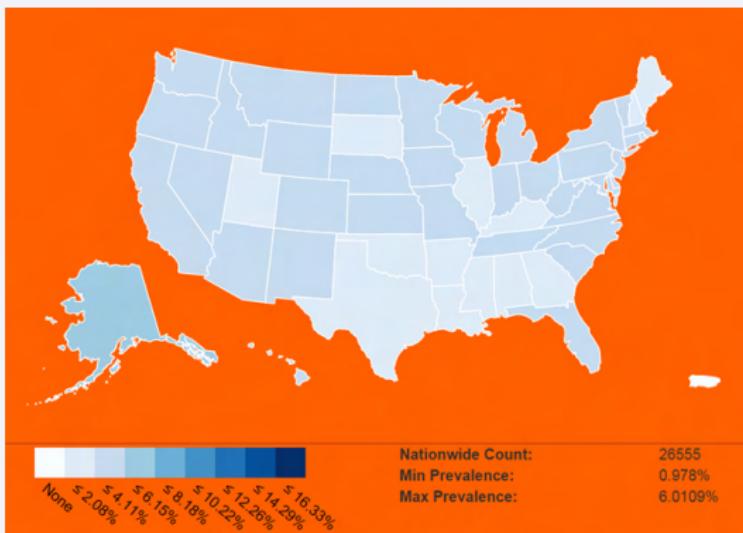


Figura: Mapa interactivo, distribución do nome Margaret a través do tempo ([http://zatonovo.com/dataviz/baby\\_names](http://zatonovo.com/dataviz/baby_names))

## Visualizador de nomes e tendencias de nomes



Figura: Mapa interactivo, tendencia do nome Rafael (<http://www.babynamewizard.com/voyager>)

## Visualizador interactivo sobre tendencias de nomes

### England & Wales Baby Names

Baby name trends in England & Wales since 1996

**How to use:** Type names in the box below, then explore the graph to see trends, annual rankings and numbers of babies born. Or [see some naming trends](#).

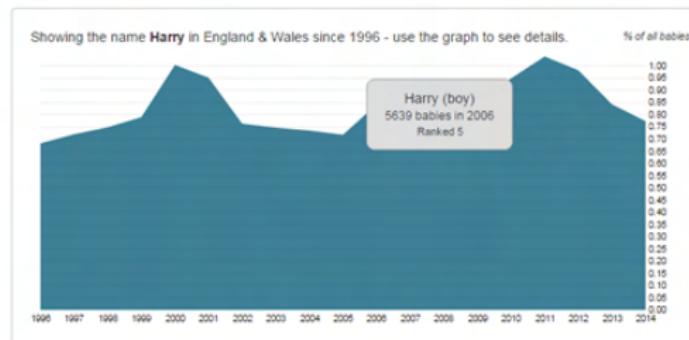


Figura: Mapa interactivo, tendencia dun nome concreto  
[\(http://names.darkgreener.com/\)](http://names.darkgreener.com/)

## Recursos Instituto da Lingua Galega

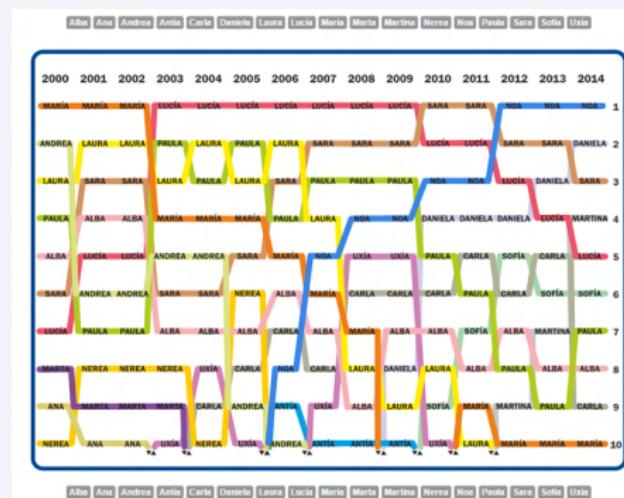
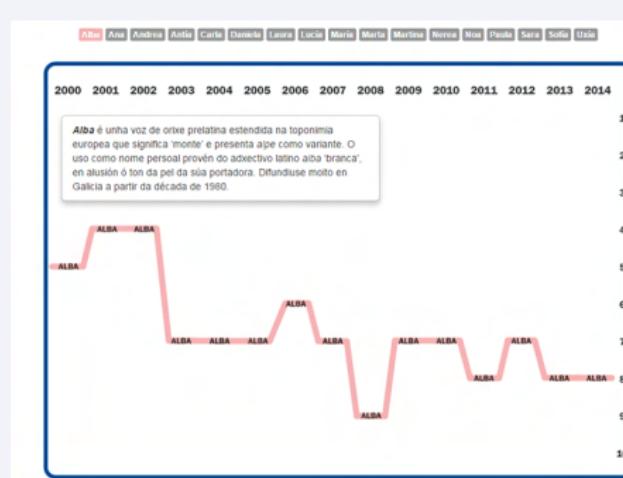


Figura: Ranking dos 10 nomes más populares en Galicia (2000-2014) ([http://ilg.usc.es/nomes\\_galicia/](http://ilg.usc.es/nomes_galicia/))

## Recursos Instituto da Lingua Galega



**Figura:** O gráfico permite coñecer a evolución da frecuencia dos nomes; ademais, ofrécese información sobre as súas orixes etimolóxicas e o seu significado  
[\(http://ilg.usc.es/nomes\\_galicia/\)](http://ilg.usc.es/nomes_galicia/)

- └ Motivación
- └ Apellidos

## Estatísticas oficiais



Figura: Apelidos más frecuentes por concello  
(<http://www.ige.eu>)

└ Motivación

└ Apellidos

## Estatísticas oficiais



Figura: Apellidos por provincia de nacemento  
(<http://www.ine.es/>)

- └ Motivación
- └ Apellidos

# Estatísticas oficiais

The screenshot shows the official statistics page from the U.S. Census Bureau. The header includes the U.S. Department of Commerce logo, the Census Bureau logo, and navigation links for People, Business, Geography, Data, Research, and Newsroom. Below the header, a breadcrumb trail shows the path: Census.gov > People and Households > Genealogy Data > Frequently Occurring Surnames from Census 2000. The main title is "Genealogy". A sidebar on the left lists sections such as Genealogy Data Main, Age Search Service, Frequently Occurring Surnames from Census 2000, Frequently Occurring Surnames from Census 1990, and Historical Data. The main content area displays a table titled "Genealogy Data: Frequently Occurring Surnames from Census 2000". The table has two columns: "NAME" and "NUMBER OF OCCURRENCES". The data is as follows:

NAME	NUMBER OF OCCURRENCES
Smith	2,376,206
Johnson	1,857,160
Williams	1,534,042
Brown	1,380,145
Jones	1,362,755
Miller	1,127,803
Davis	1,072,335
Garcia	858,289
Rodriguez	804,240
Wilson	783,051

Figura: Apellidos más frecuentes en Estados Unidos, 2000  
(Census Bureau) (<http://www.census.gov/genealogy/www/data/2000surnames/index.html>)

## Recursos web



Figura: Apellidos más frecuentes de Estados Unidos  
(<http://ngm.nationalgeographic.com/>)



Willians, A.R. (2011) What's in a Surname? *National Geographic magazine*.

- └ Motivación
- └ Apellidos

### Recursos web

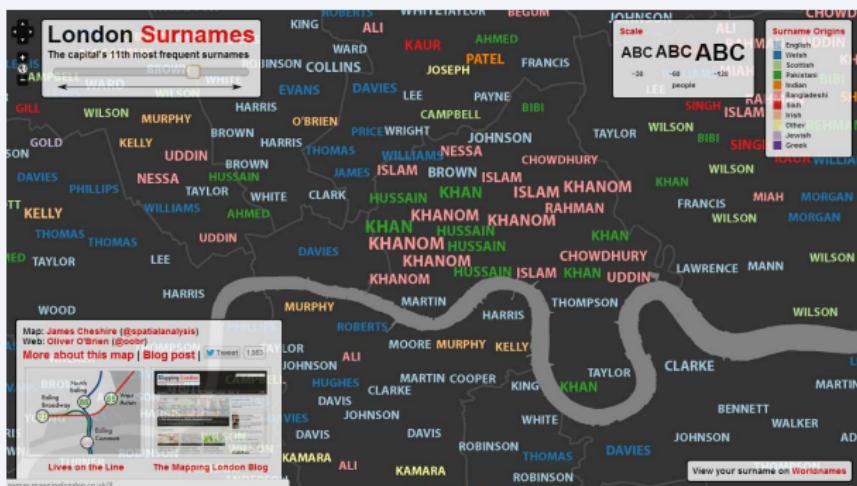
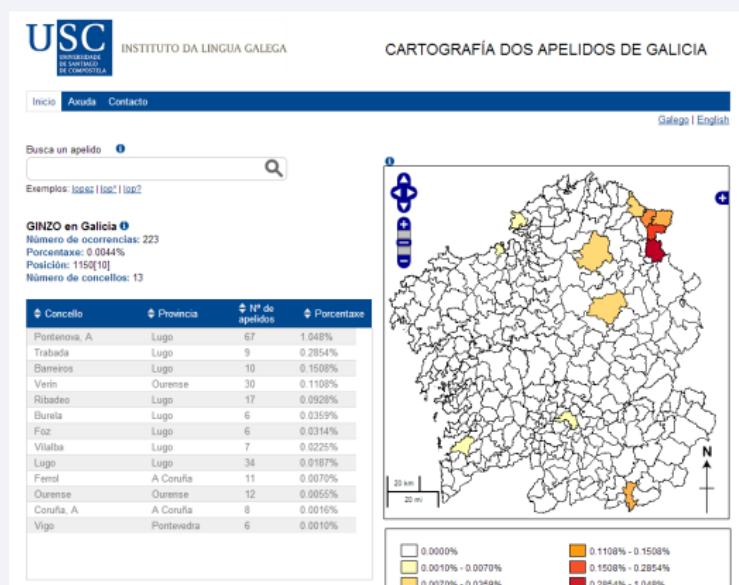


Figura: Apellidos más frecuentes en Londres  
(<http://names.mappinglondon.co.uk/>)

## └ Motivación

## └ Apellidos

## Recursos Instituto da Lingua Galega

Figura: Apellido Ginzo en Galicia (<http://ilg.usc.es/>)

└ Motivación

└ Apellidos

## Visualizador Geogen

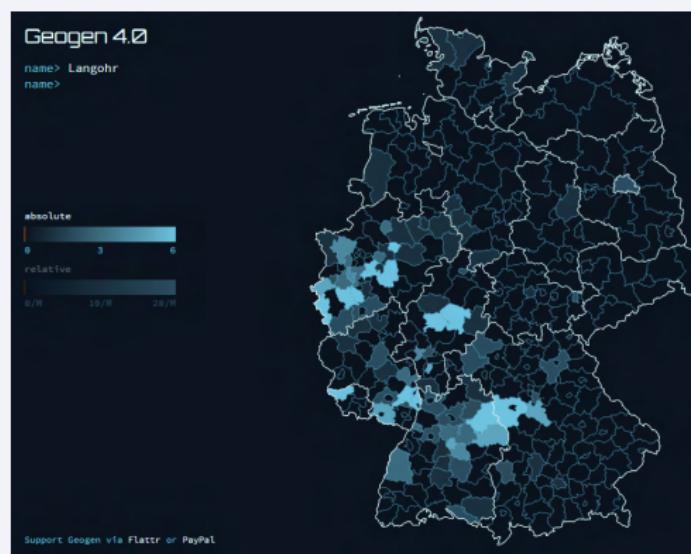


Figura: Apellido Langohr en Alemania  
(<http://geogen.stoepel.net/>)



## Motivación

- ▶ R permite facer excelentes representacións gráficas.
- ▶ Shiny permite crear aplicacións web interactivas con R.
- ▶ Non se requieren coñecementos de HTML, JavaScript ou CSS.
- ▶ Shiny é un paquete en R creado en 2012 por Rstudio (GPLv3) para desenvolver aplicacións Web utilizando R.

└ Shiny

└ Que é Shiny?

## How to Start Shiny



 Get inspired  
(gallery)

 Get started  
(tutorial)

 Go deeper  
(articles)

- ▶ Tutorial
- ▶ Artículos:
  - ▶ Partes básicas dunha aplicación Shiny
  - ▶ Como construír unha aplicación Shiny
  - ▶ Como executar unha aplicación Shiny
  - ▶ Como obter axuda
  - ▶ ...

## Estrutura dunha aplicación Shiny

- ▶ As aplicaciones con Shiny teñen dúas componentes:
  1. unha secuencia de comandos de interfaz de usuario e
  2. unha secuencia de comandos de servidor.

## Estrutura dunha aplicación Shiny

- ▶ As aplicaciones con Shiny teñen dúas componentes:
  1. **unha secuencia de comandos de interfaz de usuario:** a interfaz de usuario (`ui`) controla o deseño e o aspecto da aplicación: `ui.R` e
  2. **unha secuencia de comandos de servidor** o script `server.R` ten as instrucións que o equipo necesita para construír a aplicación.

## Estrutura dunha aplicación Shiny

- ▶ As aplicaciones con Shiny teñen dúas componentes:
  1. **unha secuencia de comandos de interfaz de usuario:** a interfaz de usuario (`ui`) controla o deseño e o aspecto da aplicación: `ui.R` e
  2. **unha secuencia de comandos de servidor** o script `server.R` ten as instrucións que o equipo necesita para construír a aplicación.

## Estrutura dunha aplicación Shiny

- ▶ As aplicaciones con Shiny teñen dúas componentes:
  1. **unha secuencia de comandos de interfaz de usuario:** a interfaz de usuario (`ui`) controla o deseño e o aspecto da aplicación: `ui.R` e
  2. **unha secuencia de comandos de servidor** o script `server.R` ten as instrucións que o equipo necesita para construír a aplicación.

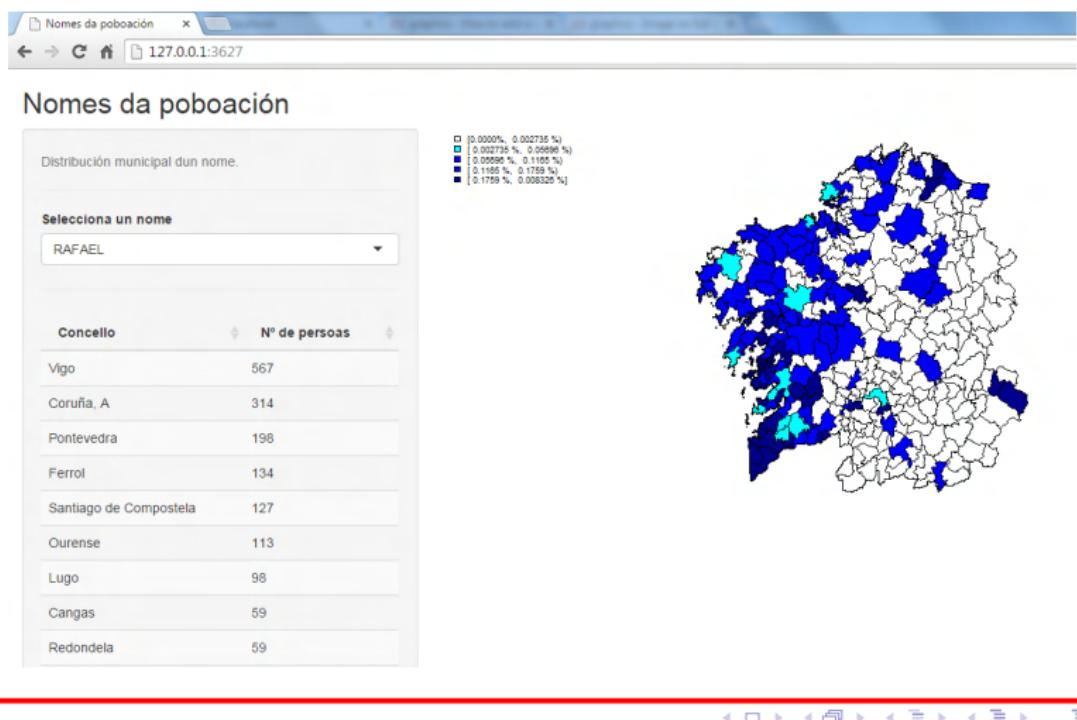
## Estrutura dunha aplicación Shiny

- ▶ As aplicaciones con Shiny teñen dúas componentes:
  1. **unha secuencia de comandos de interfaz de usuario:** a interfaz de usuario (`ui`) controla o deseño e o aspecto da aplicación: `ui.R` e
  2. **unha secuencia de comandos de servidor** o script `server.R` ten as instrucións que o equipo necesita para construír a aplicación.

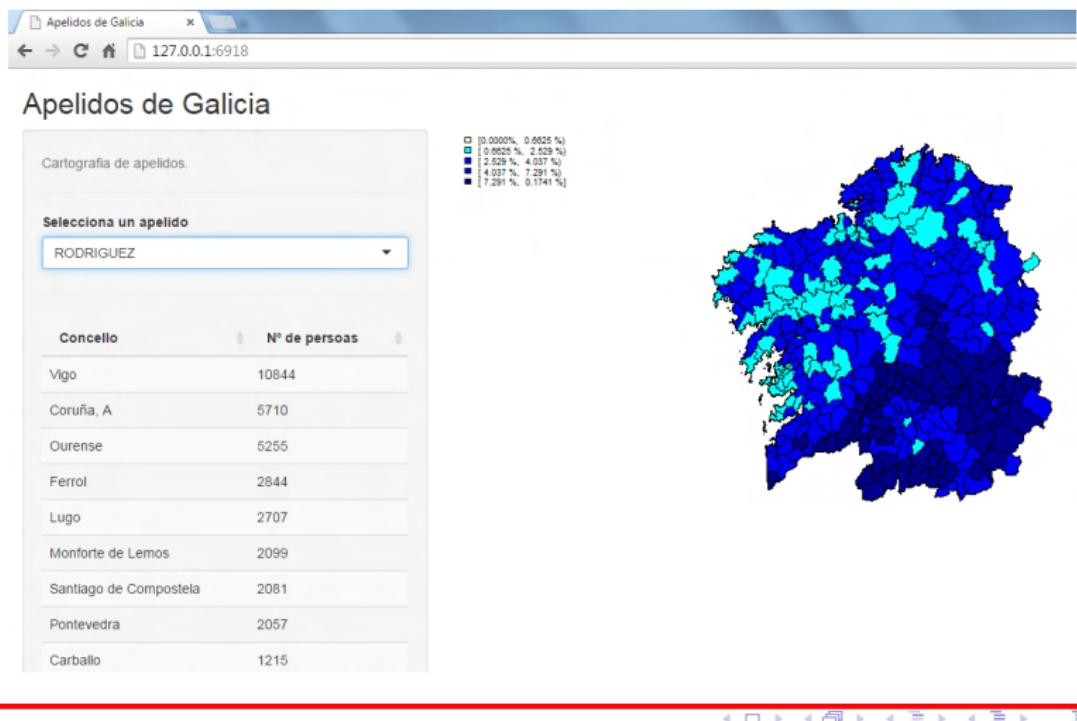
## Executar unha aplicación Shiny

- ▶ Cada aplicación Shiny ten a mesma estrutura: como mínimo dous scripts de R gardados no mesmo directorio: `ui.R` e `server.R`.
- ▶ Cada aplicación terá o seu propio directorio único.
- ▶ Executar unha aplicación Shiny, proporcionando o nome do directorio á función `runApp`.
- ▶ Si a aplicación está no directorio `my_app`, empregaríase o seguinte código:  
`> library(shiny)`  
`> runApp("my_app")`

## Representación con Shiny



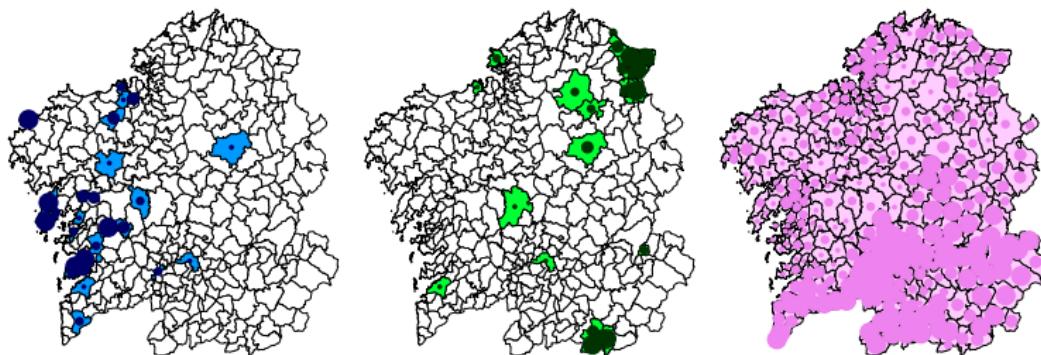
## Representación con Shiny



└ Unha mirada máis profunda en certos apelidos

└ Motivación e obxectivos

Para un apellido concreto, o obxectivo é atopar unha estrutura subxacente que explica a variación espacial (e posiblemente temporal) do apellido tendo en conta unha estrutura de veciñanza.



Localizacions en Galicia dos apelidos Crujeiras, Ginzo e Rodríguez. O tamaño do punto é proporcional o número de persoas que levan ese apellido ponderado pola poboación do concello.

- ▶ As diferentes liñas de investigación no contexto da onomástica non soen ter en conta a dimensión espacial e espazo-temporal da evolución dos apelidos.
- ▶ Fixando rexións administrativas, por exemplo, concellos, pode facerse uso de métodos espaciais e espazo-temporais para a análise de datos de conteo que permitan modelar o patrón subxacente á evolución dos apelidos.
- ▶ Estes métodos serán útiles para modelar patróns de evolución dos apelidos.
- ▶ Para analizar o patrón espacial dos apelidos en Galicia axustouse o modelo proposto por Besag et al. (1991) (modelo BYM), adaptado a este contexto.
- ▶ Para o axuste de modelos xerárquicos neste contexto faise uso da metodoloxía INLA (Integrated Nested Laplace Approximation) proposta por Rue et al. (2009).
- ▶ Datos cedidos polo Instituto Galego de Estatística.



Besag, J., York, J. e Mollié, A. (1991) Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, **43**, 1–59.



Rue, H., Martino, S. e Chopin, N (2009) Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, **71**, 319–392.



Rue, H. e Held, L. (2005) *Gaussian Markov random fields: theory and applications*. Chapman and Hall, CRC Press, London.

## Modelos Bayesianos espaciais

$Y_i$  = número de persoas con ese apellido no concello  $i$ ,  $i = 1, \dots, n$

este proceso de conteo será modelado a través dun modelo Poisson–LogNormal (Banerjee et al. (2004)),

$$Y_i | \eta_i \sim \text{Pois}(E_i \exp(\eta_i)),$$

onde  $E_i$  é a poboación en risco,  $\eta_i$  un predictor lineal e as variables  $Y_i$  son independentes.

$$\eta_i = \mu + f_s(s_i) + f_u(s_i),$$

onde  $s_i$  é o centroide de cada concello,  $f_s$  e  $f_u$  son os efectos estruturados e non estruturados.  $f_s$ , é un campo aleatorio Gaussiano de Markov (Rue e Held (2005)).  $f_u$ , ruido branco que representa a “sobredispersion” dos concellos.



Banerjee S., Carlin, B. E. e Gelfand, A. (2004) *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability



Rue, H. e Held, L. (2005) *Gaussian Markov random fields: theory and applications*. Chapman and Hall, CRC Press, London.

## Apelidos Crujeiras, Ginzo e Rodríguez (Padrón 2011)

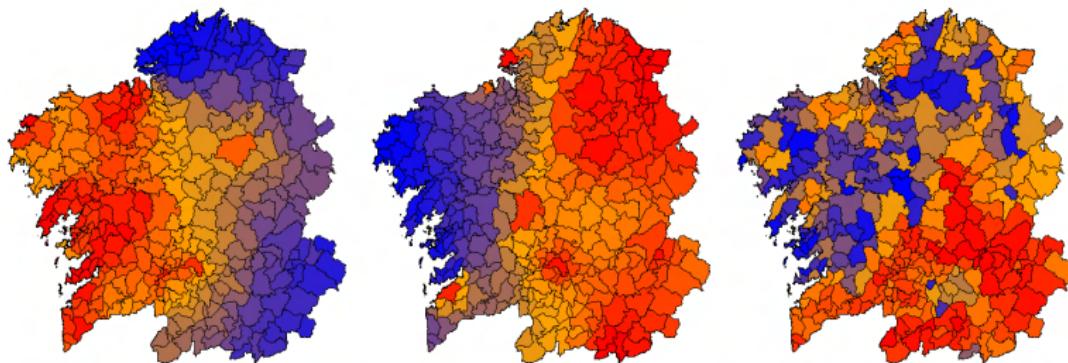


Figura: Media a posteriori da parte estruturada do campo latente.

└ Unha mirada máis profunda en certos apelidos

└ Representación con Shiny

## Representación con Shiny



# VISUALIZACIÓN XEOGRÁFICA DE DATOS DE ONOMÁSTICA CON R

María José Ginzo Villamayor

Departamento de Estatística e Investigación Operativa  
Universidade de Santiago de Compostela



DEPARTAMENTO DE ESTATÍSTICA  
E INVESTIGACIÓN OPERATIVA

