The Role of Geography in Linguistic Change: Existential Agreement in European Spanish

Daniel Ezra Johnson, post-doctoral researcher, CRTT (Lyon 2)

A Multilingual Sociolinguistic Variable

- *Il y a une vache dans mon jardin.*
- *Il y a **des vaches** dans mon jardin.*
- **Il y **ont des vaches** dans mon jardin.*

- *Hay **vacas** en mi jardín.*
- *??**Hayn vacas** en mi jardín.*
- *Había **vacas** en mi jardín.*
- *?**Habían vacas** en mi jardín.*

- *There was a cow in my yard.*
- *There **were cows** in my yard.*
- *?There was **cows** in my yard.*
- *There's **cows** in my yard.*

Why study Existential Agreement?

- A (slow) ongoing change in opposite directions. English: rise of non-standard singular agreement. **Spanish**: rise of non-standard **plural agreement**.

- Studies show different **rates** of agreement across geographical varieties, but similar **constraints** on the variation (factors favoring or disfavoring it).

- **Variationist** studies model and measure these factors, but usually don't try to explain them.

- Can we understand variation better with **cognitivist** principles, looking at the use of **constructions** and considering **meaning** in accounting for variation?

- What about regional variation within countries? Is the change spreading? If so, how? And why?

Existential Agreement in Spain

Jeroen Claes's study (2017)

- 5500 **geocoded** tweets, 2031 tokens from *COSER* spoken corpus of rural speakers born < 1940s.

- Was able to make a **diachronic** comparison, but the comparability of the two data sets is unclear.

- Fit a **logistic mixed-effects model** to estimate the effect of Twitter vs *COSER* and other factors.

- Large effect: **verb tense**, **region**; smaller: **polarity**. Regions diverging over time, even as agreement spreads (**?**) west from Valencia and Catalonia.

- Origin of change seen as **contact with Catalan** (Valencian). The story in Catalan is complicated.

- Used GAM to map a **continuous** geographic effect.

A study of Existential Agreement on Twitter

(work with Jeroen Claes)

- Collected tweets with *haber* or *there.* R package *streamR*, August 2016. >20 million, >2 for Spain.

- We plan to look at and compare Spain, Argentina, United States, and United Kingdom / Ireland.

- Today: Spain (except Canary Islands, Ceuta, Melilla). Spain has less plural existential agreement than some Latin American countries (Claes 2015 studied Dominican Republic, Puerto Rico and Venezuela).

- After filtering – most are *hay* – and geocoding based on the user's *location* field: 26807 tokens.

- Mapped location to 1609 *municipios* (of 8028 total). Madrid: 6207 tweets, Barcelona: 2905, 680: only 1.

- GAM model with verb tense, polarity, and their interaction (not found significant by Claes 2015), and geography (latitude/longitude of each token).

Logistic Regression

and

Generalized Additive Models

- logistic regression
- estimates prob. of plural agreement for each token (even though obs. is binary).
- log-odds of probability as linear function $log(p / (1-p)) = x_0 + ax_1 + bx_2 + bx_3 \ldots$

- *mgcv* R package: GAM adds smoothing splines, here for geography (lat./long.).
- 'isotropic smoothing term on the sphere' *gam(plural ~ negation * tense + s(latitude, longitude, bs = 'sos')*
- Simon Wood http://www.biostatistiikanseura.org/GAM/Lectures/ smooth-toolbox.pdf

Linguistic Factors

Logistic Regression Model

| tense : polarity	log-odds	# plural	# singular	% plural
past subjunctive : positive	+ 0.156	165	1039	13.7 %
past subjunctive : negative	+ 0.200	57	376	13.2 %
imperfect : positive	0.000	580	4362	**11.7 %**
imperfect : negative	− 0.364	105	1134	**8.5 %**
conditional : positive	− 0.235	63	572	9.9 %
conditional : negative	− 0.377	37	375	8.9 %
present subjunctive : positive	− 0.352	535	5770	8.5 %
present subjunctive : negative	− 0.359	153	1569	8.9 %
future: positive	− 0.171	609	5458	**10.0 %**
future: negative	− 0.847	78	1383	**5.3 %**
preterite : positive	− 0.990	87	1907	4.4 %
preterite : negative	− 0.893	20	373	5.1 %

Linguistic Factors

The Effects of Tense and Mood
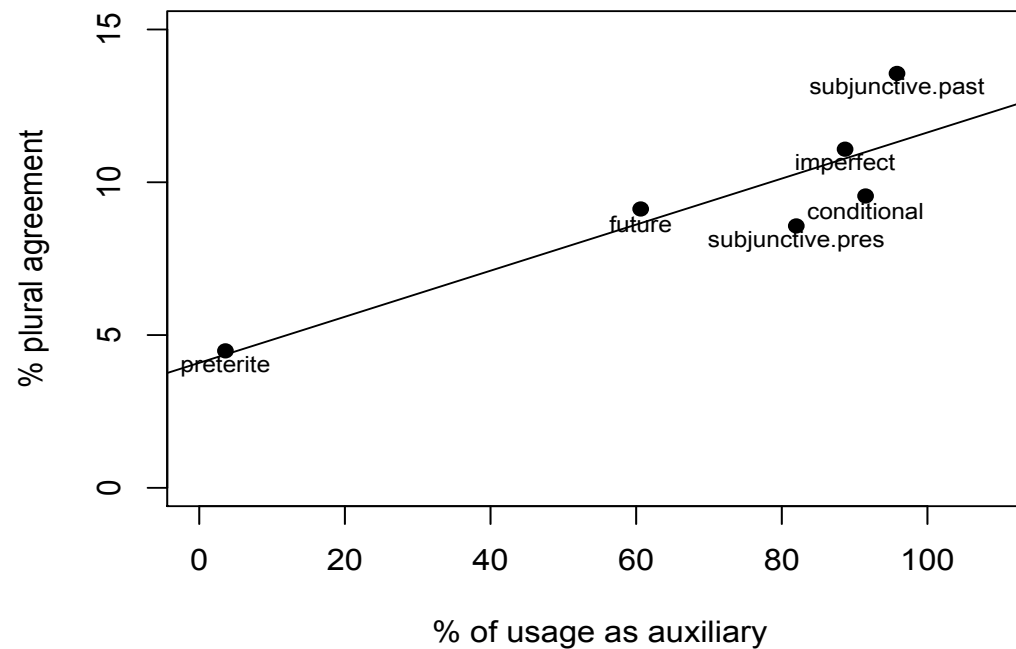
- Fairly large effect of tense/mood (esp. preterite).

- Can largely explain it with *statistical preemption.*

- Idea is: agreement spreads from auxiliary uses (perfect tenses) where it is mandatory:

  *ya lo **había** hecho , pero muchas graciassss...*
  *nunca me **habían** confundido tannnnnto*

- Forms used more often in auxiliary constructions (vs. existential) show more plural agreement.

Linguistic Factors

The Effect – or Non-Effect – of Negative Polarity
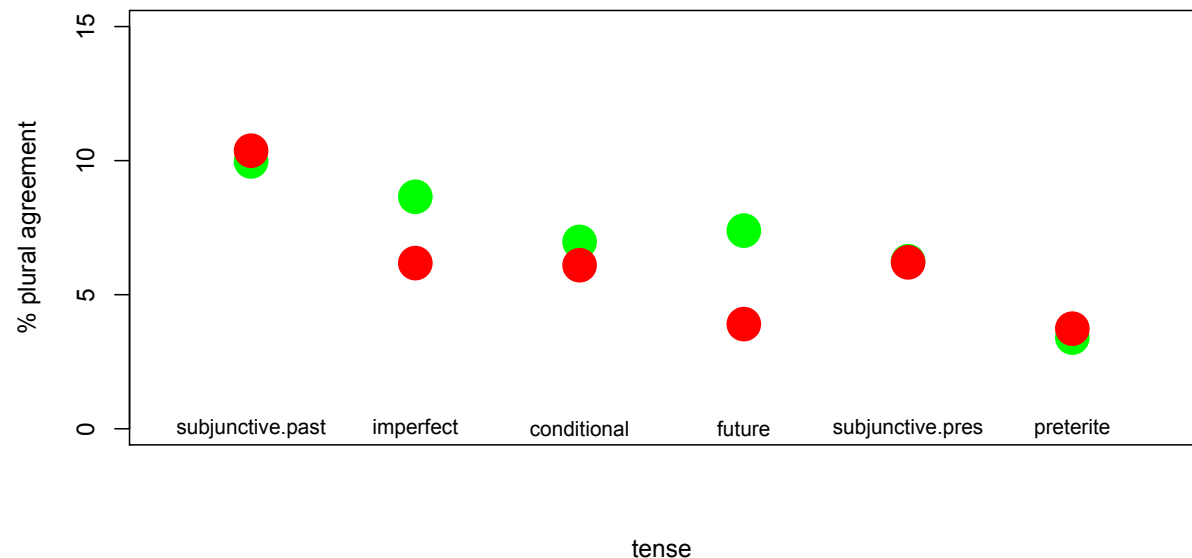
- A polarity effect exists only in **imperfect**, **future**.

- Known: also exists in **present** (Claes, LatAm).

- Idea: in the negative, a **type** not an **instance**.

- **Subjunctive**: both positive and negative: **irrealis**.

- **Conditional**: both **hypothetical** / **counterfactual**.

- **Preterite**: 'subject' is always **definite** and usually **discourse-old** and/or **qualified/modified**

Autonomous Communities and Provinces of Spain

Languages and Dialects of Spain

Existential Agreement
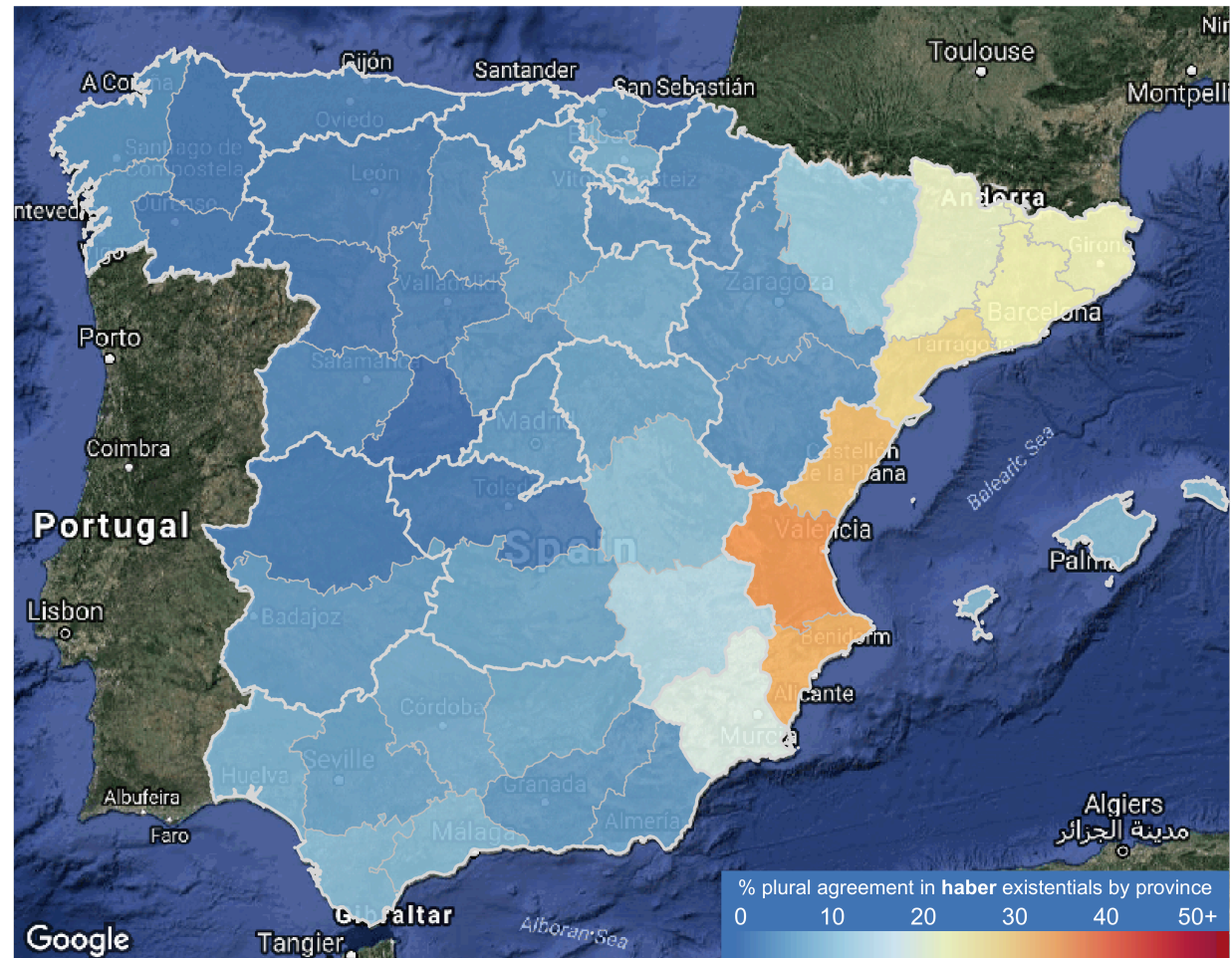
Raw results: 1 or more tweets

% plural agreement in **haber** existentials (n >= 1)
0   10   20   30   40   50+
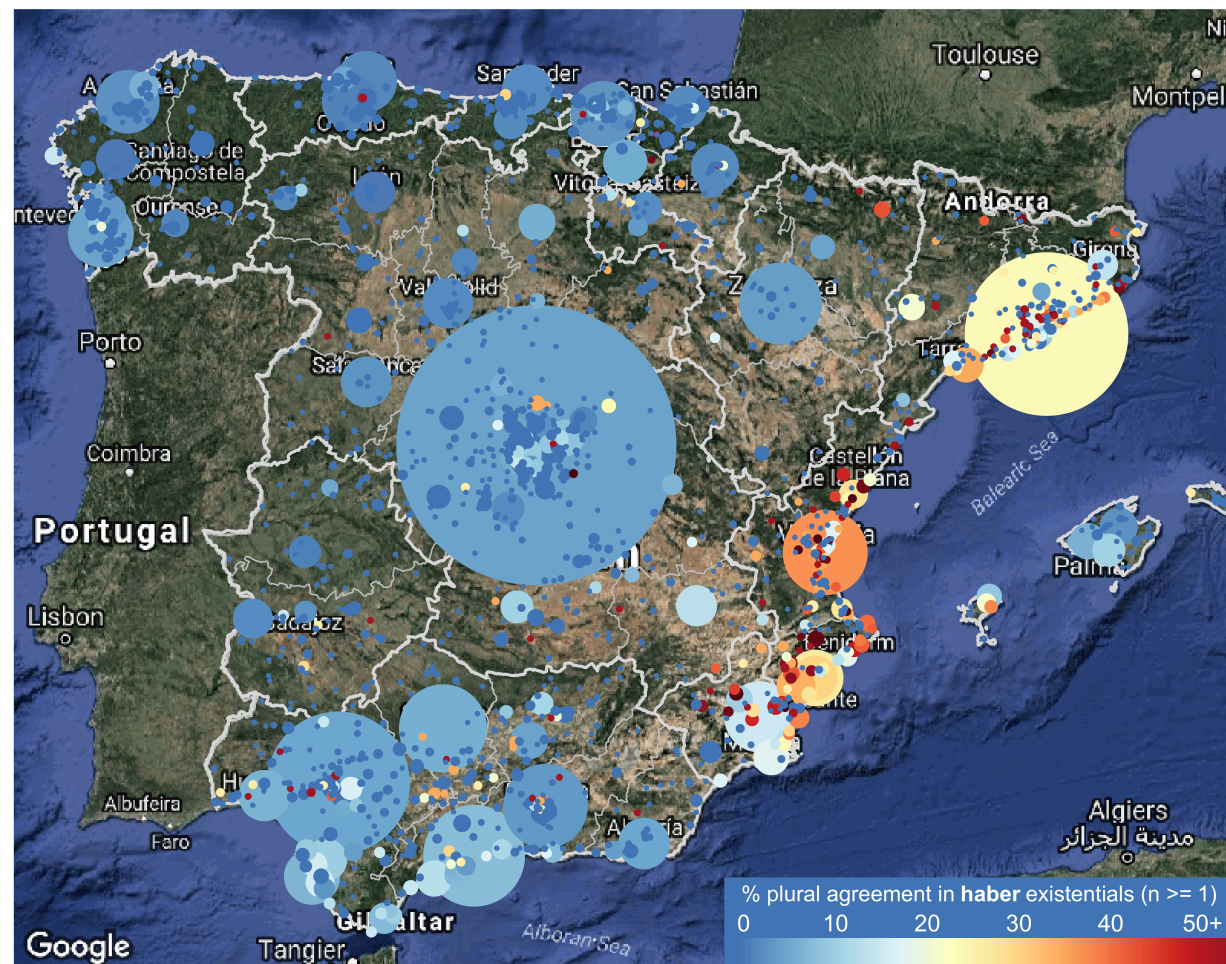
Existential Agreement:

Raw results: 25 or more tweets
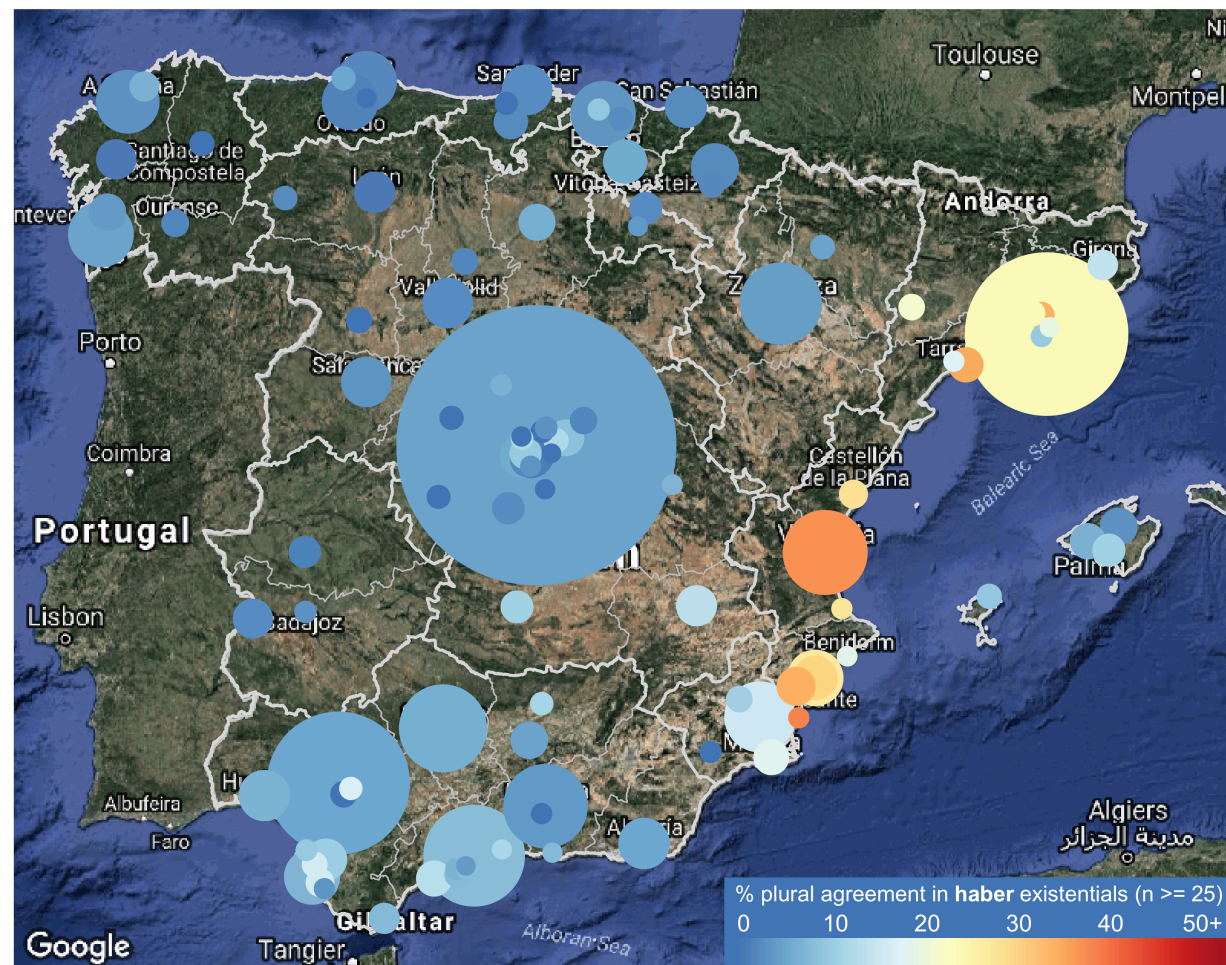
Existential Agreement

Results by Province

% plural agreement in **haber** existentials by province

0 10 20 30 40 50+

Existential Agreement

Results of a Generalized Additive Model

% plural agreement in **haber** existentials (GAM)

Existential Agreement

Raw results: 1 or more tweets

% plural agreement in **haber** existentials (n >= 1)
0   10   20   30   40   50+
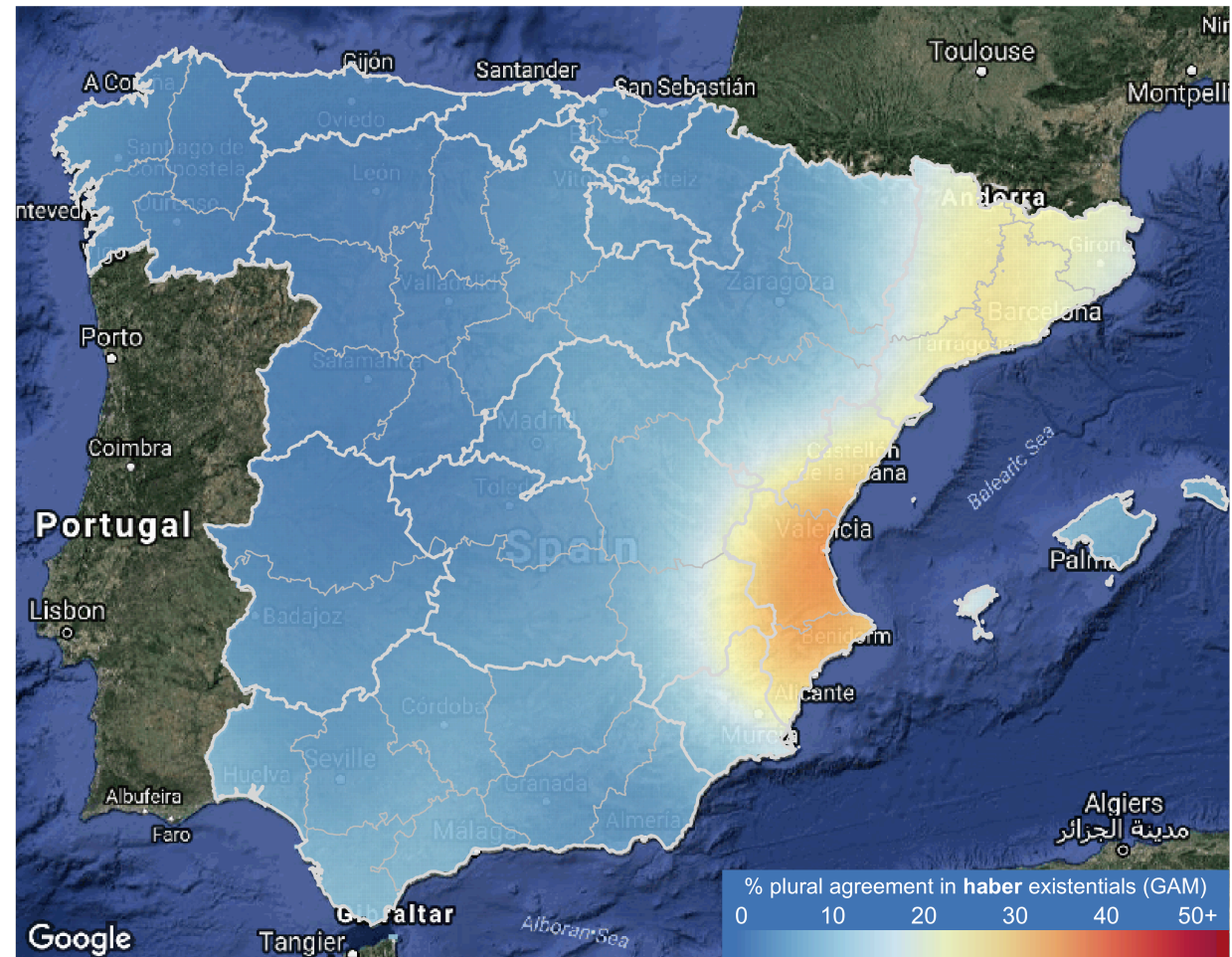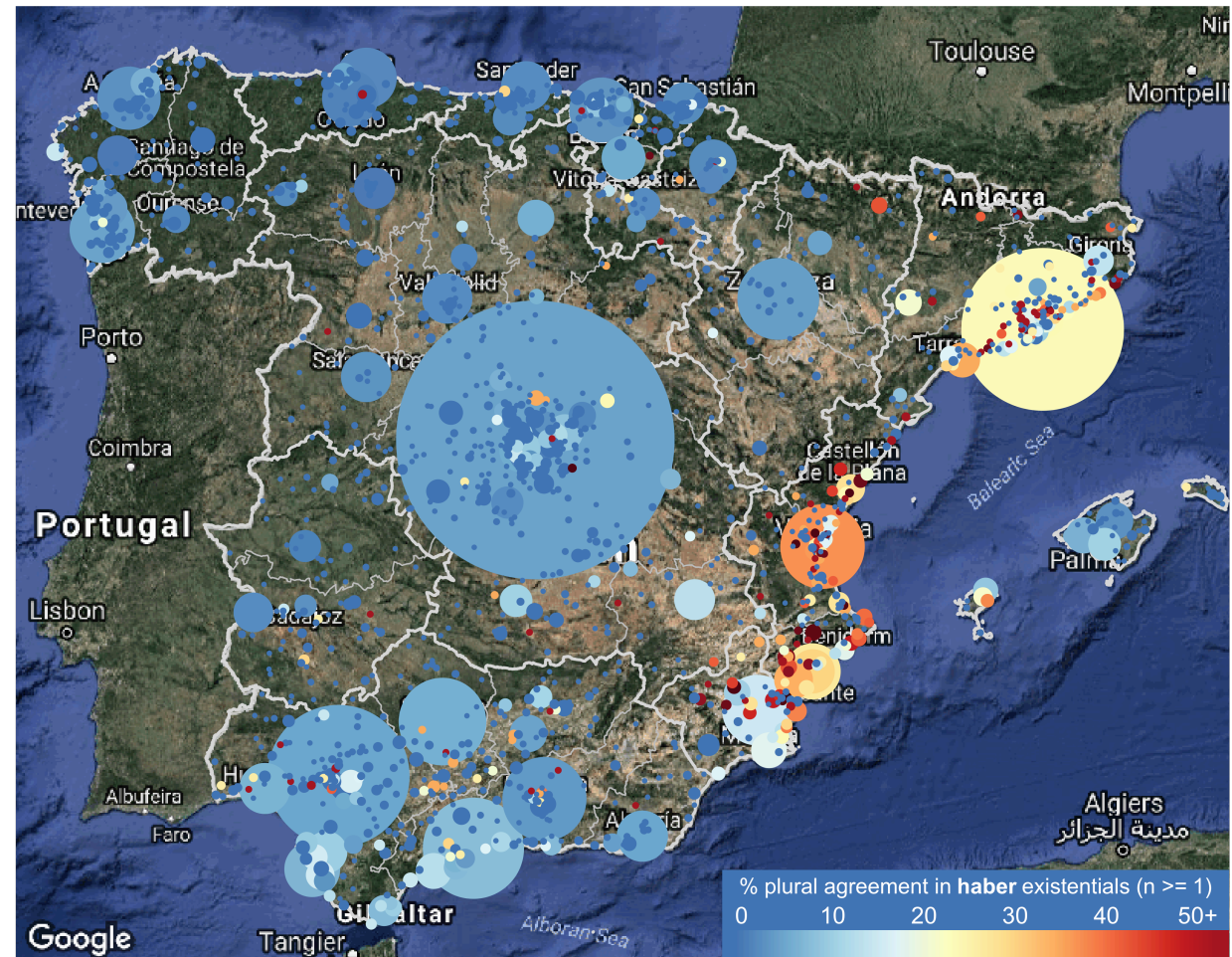
Migration within Spain

Destination of those who left Catalonia or the Valencian Community

% of those who left Catalonia or Valencian C.
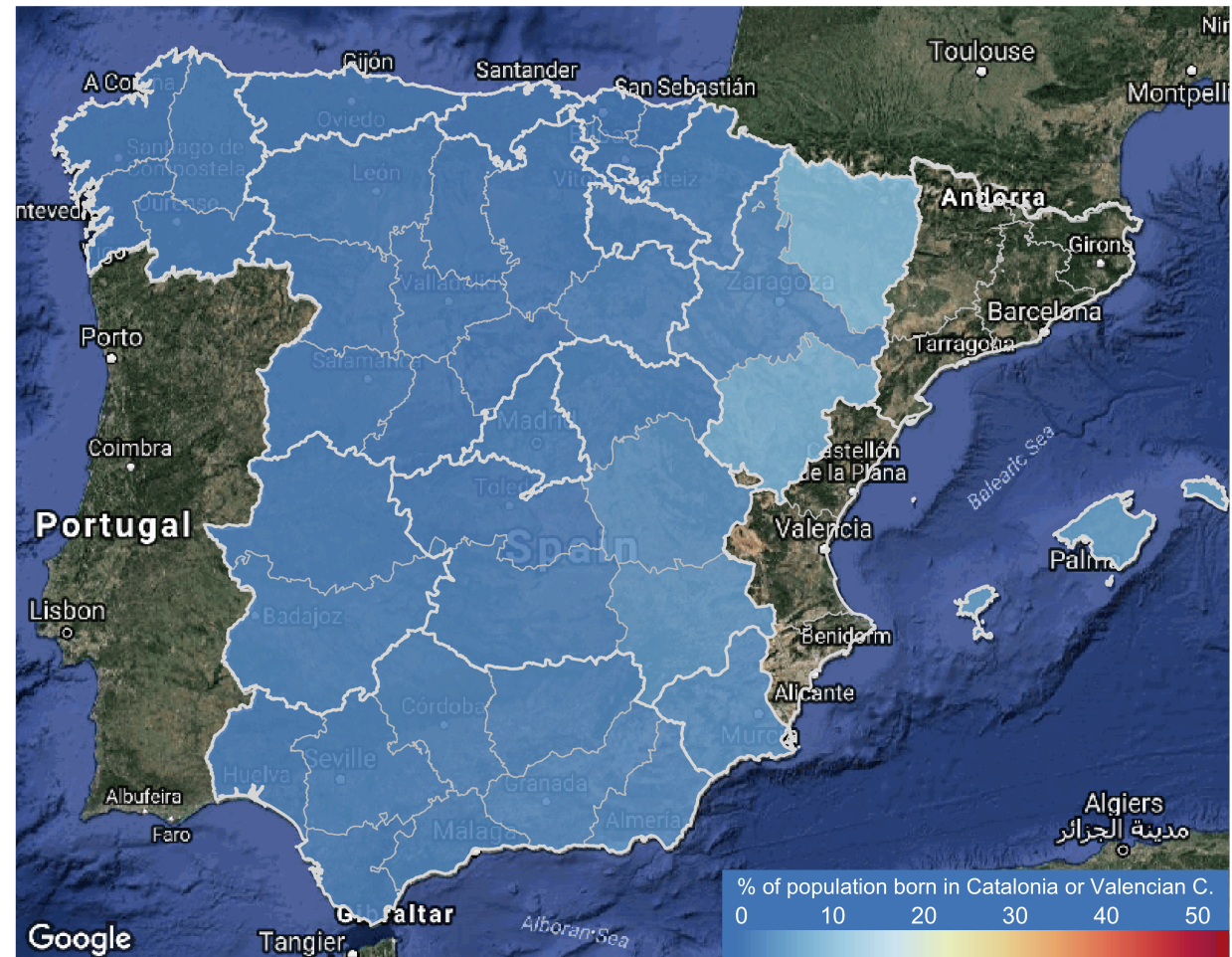0   10   20   30   40   50
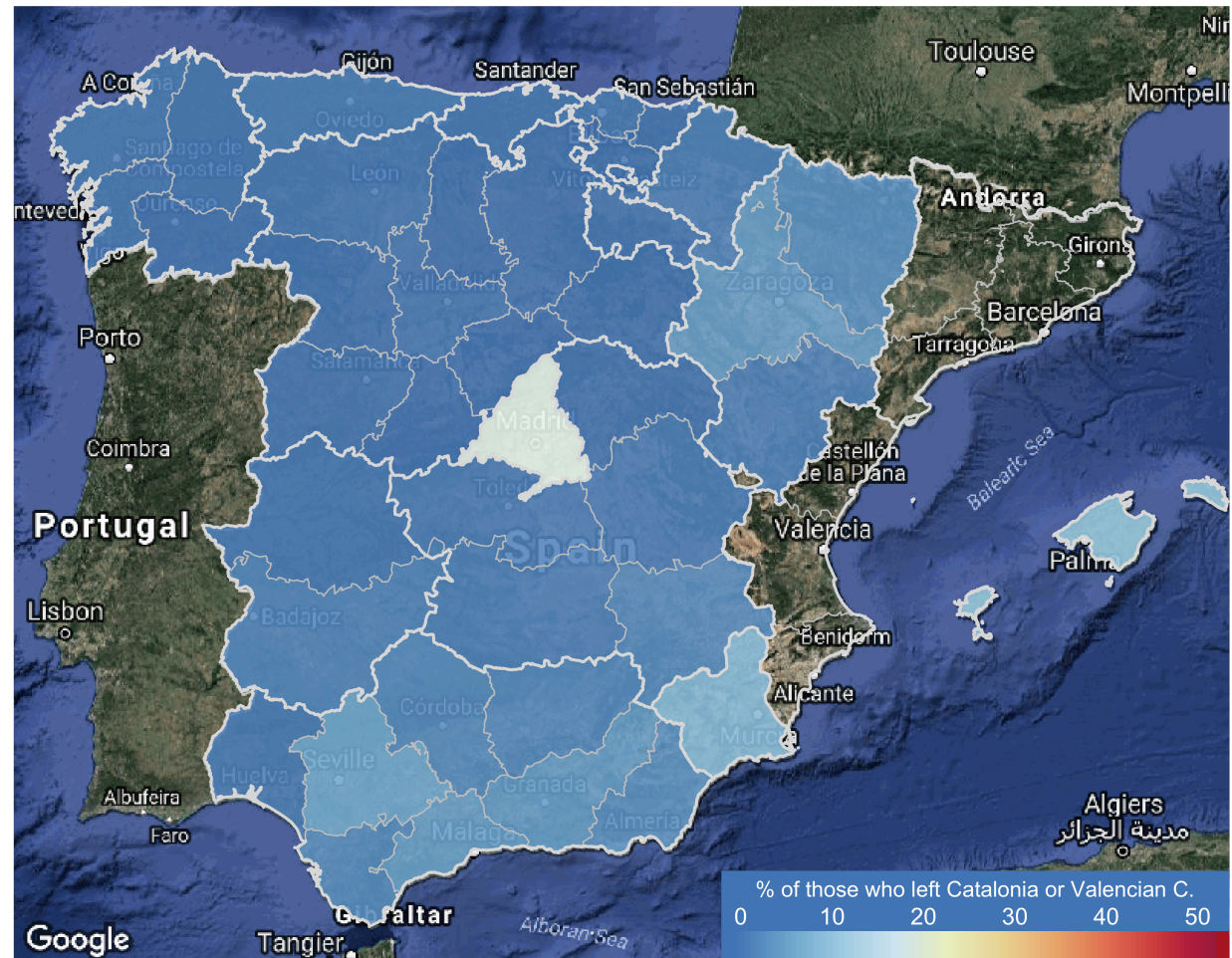
Migration into Spain

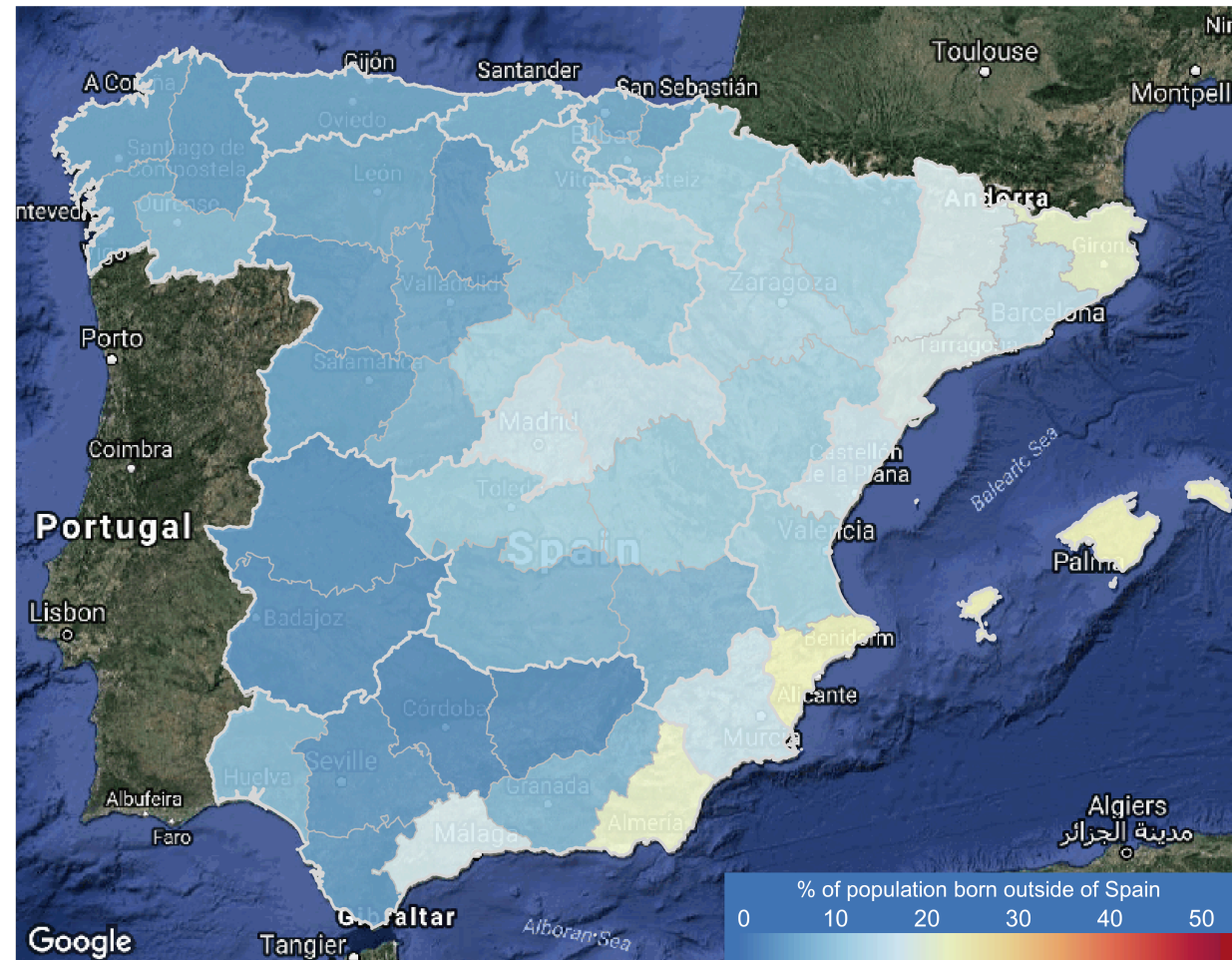Destination of those born outside Spain
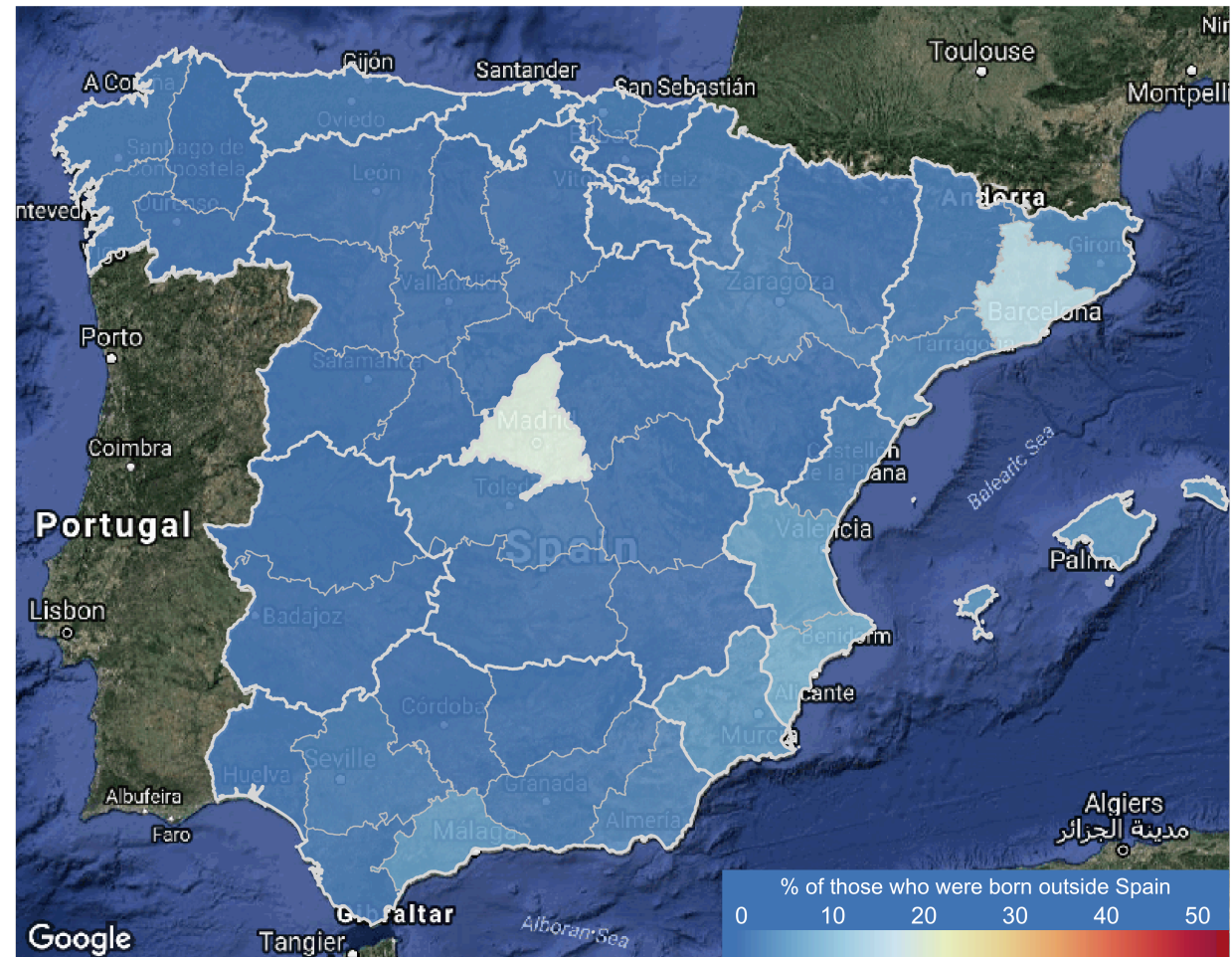
Migration within Spain

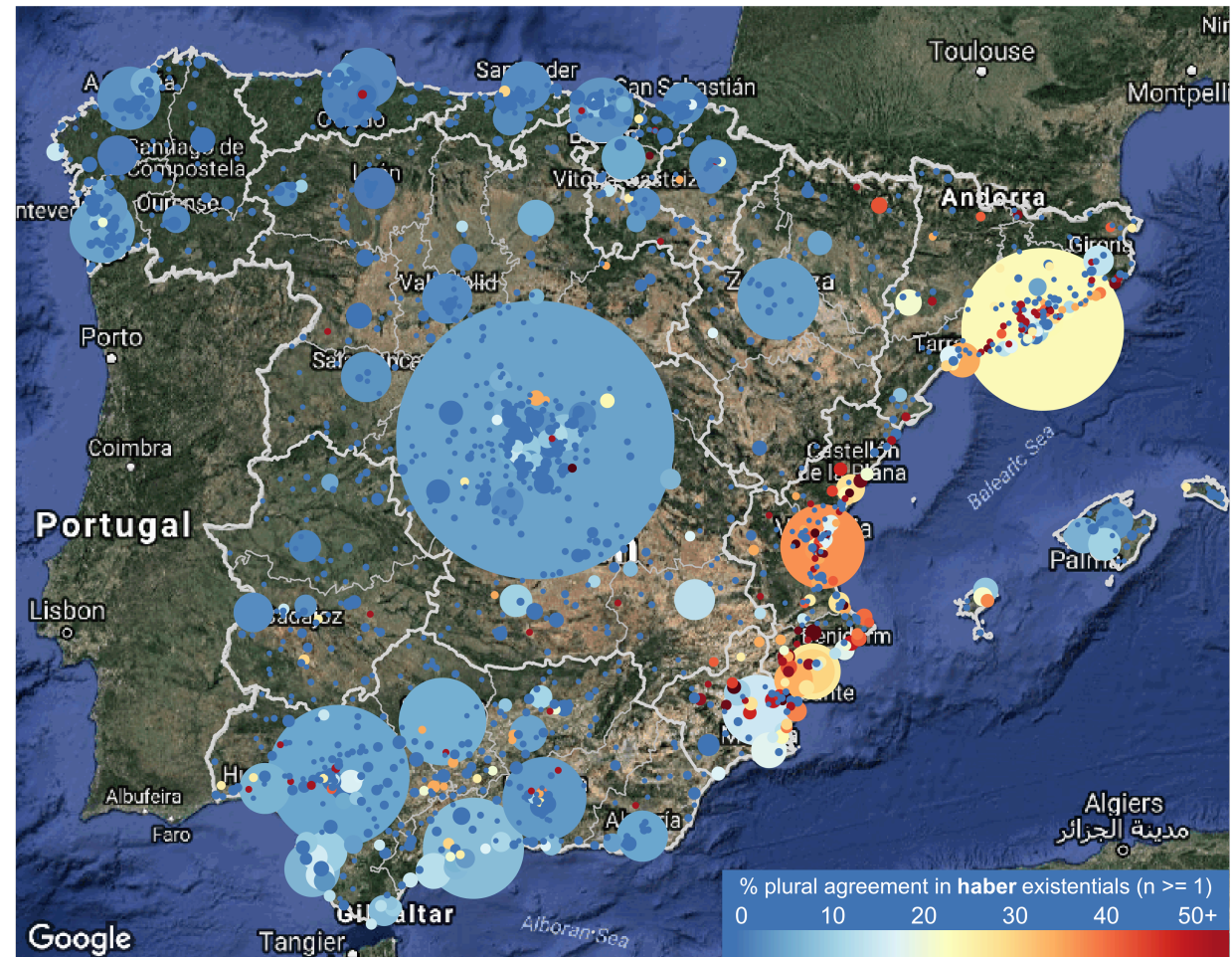## Percentage of population born in Catalonia or the Valencian Community

Migration into Spain

Percentage of population born outside Spain

Existential Agreement

Raw results: 1 or more tweets

Posts and Retweets by Province

- collected 52 **retweets** (13-14/03/17) where I could geocode locations of both original poster and retweeter
- what provinces sent **from**?
- what provinces sent **to**?
- **all 52** were **posted** and **retweeted** in the **same province**
- Barcelona: **29**; Madrid: 7 Asturias, Baleares, Vizcaya: 3 Alicante: 2; Cantabria, Girona, Lleida, Segovia, Valencia: 1 each.

Merci!

R

Twitter

Google Maps

Jeroen Claes (KU Leuven)

Verónique Lacoste (supervisor, Lyon 2)

Instituto Nacional de Estadísticas de España

Gaston Sanchez (regular expression for retweets)

Cynthia Brewer, Mark Harrower and Penn State (Color Brewer)

Pablo Barbera (*streamR*), Jeff Gentry (twitterR) , Simon Wood (*mgcv*)