

New approaches to mapping in linguistics

GIS Mapping for Linguistic Research

Jennifer Cramer

University of Kentucky

Mapping in linguistics

According to Lameli (2010), cartographic quality is improving in linguistics, thanks to GIS

The 1990s and early 2000s had linguists still trying to do atlases in CorelDraw...

With web mapping, “both the computational basis and collaborations within and across philologies seem set to become more relevant than ever” (Lameli 2010: 587)

He even argues that linguists will do more collaboration, more data linking

Which will take linguistics to a “new analytical sphere”

But how do we handle these kinds of cross-connectivities?

We have to work together, use the same data source/types, and be wary of interpretive traps!

Mapping in linguistics

Discussion time! Think about your own research programs...

What are some of the best maps you've seen in linguistics research?

What are some of the worst?

What do you think makes a map better/worse than others?

Atlases over time

Atlas	Data processing/display
Wenker (1881) – Sprachatlas des Deutschen Reichs	Pen and ink
Gilliéron (1902-1910) – Atlas linguistique de la France	Pen and ink
Wenker et al. (1927-1956) – Deutscher Sprachatlas	Pen and ink
Kurath et al. (1939-1943) – Linguistic Atlas of New England	Pen and ink
Orton & Wright (1975) – A Word Geography of England	Pen and ink for initial map compilation; use of a cartographer for final map versions
Orton, Sanderson, & Widdowson (1978) – The Linguistic Atlas of England	
Pederson et al. (1986-1992) – The Linguistic Atlas of the Gulf States	Modified computer tools (Electronic Matrix maps)
Kretzschmar et al. (1993) – Linguistic Atlas of the Middle and South Atlantic States	GIS
Viereck & Ramisch (1991-1997) – The Computer Developed Linguistic Atlas of England	Modified mapping software, Postscript, plus various dialectometry tools
Labov, Ash, & Boberg (2006) – The Atlas of North American English	MapInfo (GIS), with additional capabilities added by team at Marburg
Lameli et al. (2010) – Digital Wenker Atlas	GIS-based online mapping

Dialect maps

We know that space and spatial distribution of linguistic forms is important

Dialectologists have been the primary movers on this front

Indeed, much of the literature in dialectology is devoted to such topics

But dialectologists have been challenged recently on their understanding of space, saying that, for dialectologists, space is:

“carefully controlled out of the study” (Britain 2009: 143); or

“[turned into a] homogenised, historically-, socio-, economically-, and institutionally blind blank canvas” (Britain 2010: 87)

Ouch!

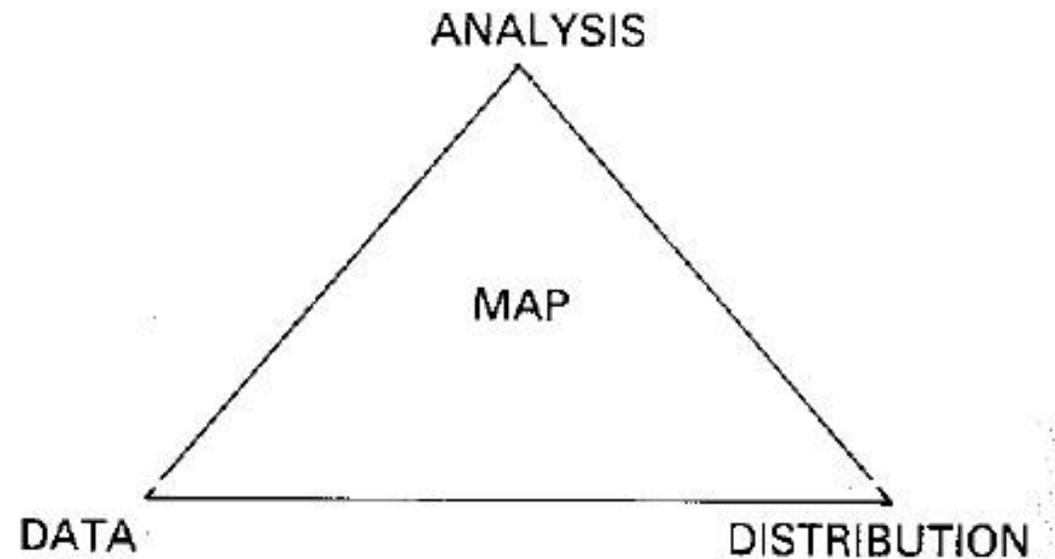
Recall...

Kirk, Sanderson, and Widdowson (1985) say we need the triangle:

So dialect maps must:

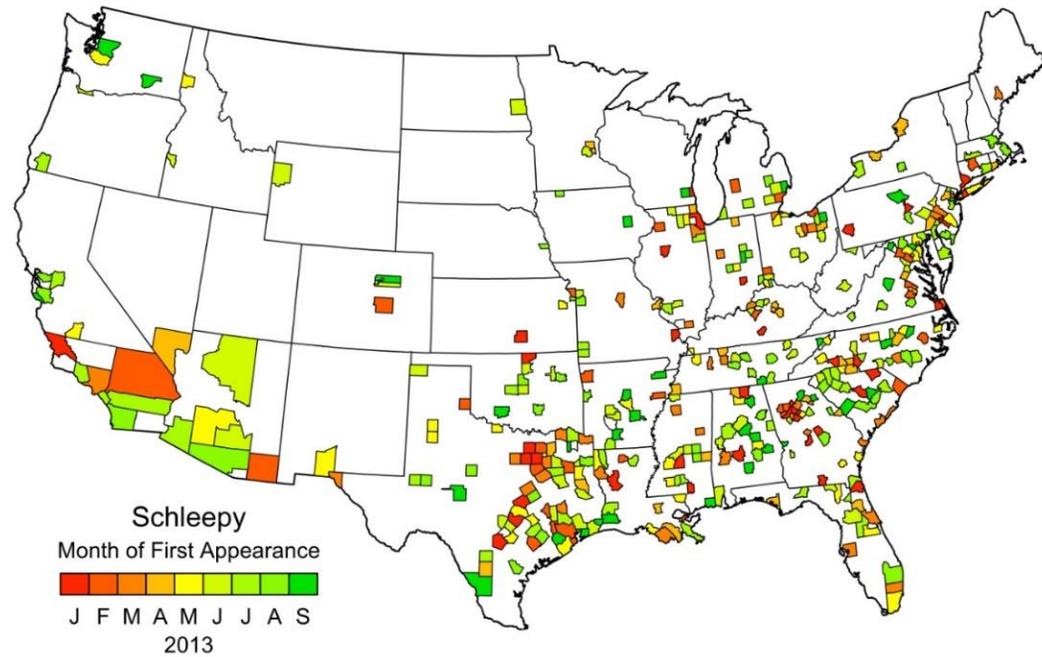
1. Show data
2. Show analysis
3. Show distribution [but also]
4. Combine datasets (linguistic and not)
5. Disseminate data

Figure 1.2: The Mapping Triangle: Properties

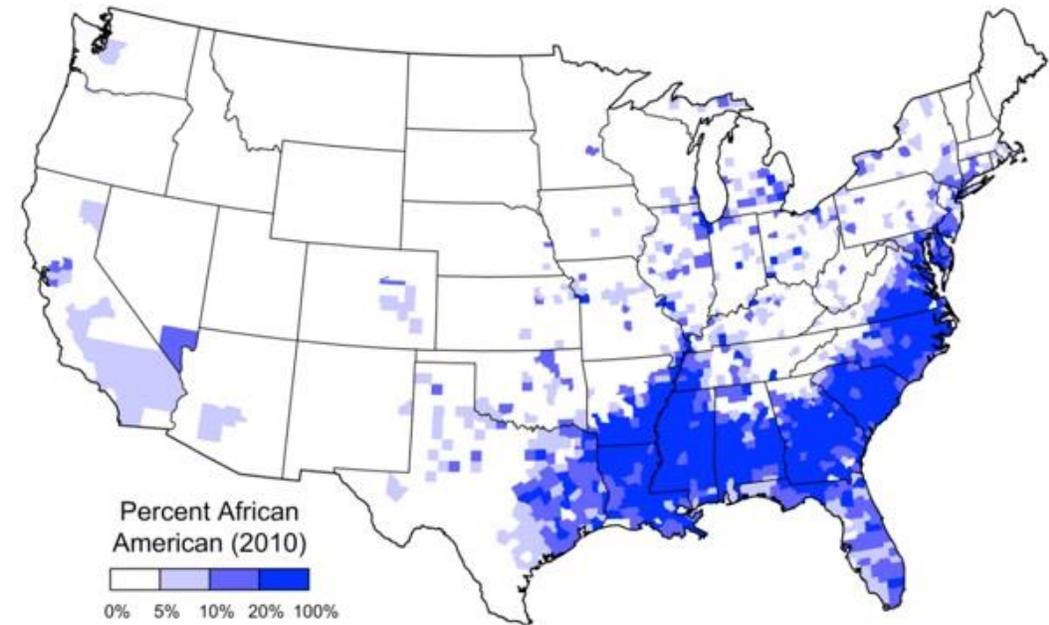


For example...

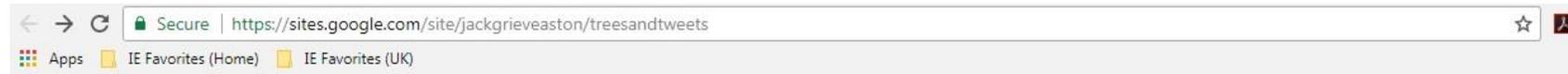
SHOW DATA



COMBINE DATASETS



Disseminate Data



Research Blog

<https://sites.google.com/view/grievejw>

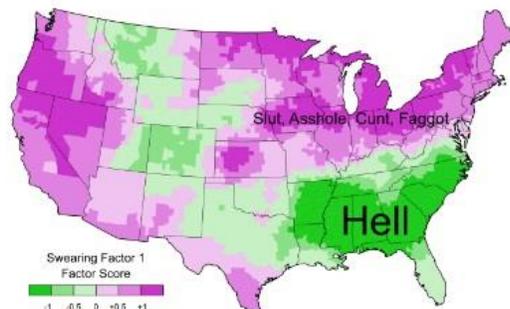
Ok, got it

January 5, 2016

A couple papers I'm involved with were just accepted/published. The [first paper](#) (with Yuan Huang, Diansheng Guo, and Alice Kasakoff, all from the University of South Carolina) replicates my [2011](#) study of American letters to the editor on the Twitter data finding similar patterns of regional variation and the [second paper](#) (with Martijn Weiling and Gosse Bouma from Groningen, Josef Fruehwald from Edinburgh, John Coleman from Oxford, and Mark Liberman from Pennsylvania) looks at the rise of the hesitation marker *um* over *uh* in Germanic languages, identifying a surprising cross-linguistic change in progress.

October 28, 2015

I also aggregated the swearing data. Basically I took about 40 of the smoothed swear word maps, like the ones I posted below, and used a factor analysis to find and map common patterns of regional variation. I extracted three factors that are mapped below, which represent the three most common patterns of regional variation in the data set. I've also annotated the maps to show the swear words in order are most strongly associated with these three patterns. Finally I combined the three factor maps to create a single overall map of American swearing regions using RGB mapping. For more information on the method see a number of my recent dialect studies ([2011](#), [2013](#), [2015](#)).



Working with electronic tools

As you saw with Google My Maps, there are stages to getting the right output:

1. Code tokens, as per usual
 - On Tuesday, data had already been coded
2. Add geospatial information
 - On Tuesday, geospatial info was included in the form of city/state
3. Project onto map
 - On Tuesday, that meant uploading your file to Google My Maps

Georeferencing

That “Add geospatial information” step is called **georeferencing**

“Georeferencing describes the process of locating an entity in ‘real world’ coordinates.” (EPA 2009)

For example, you could georeference your house by

Determining its latitude and longitude coordinates

Or maybe by using a zip code

In GIS, you can georeference many things, including images

All you need to do is find a “series of fixed reference points for which geographic coordinates are known.” (Heywood, Cornelieus, and Carver 2006: 143)

Building a map

To get from data to map, you need:

- Georeferenced linguistic data (like our Excel sheet)

- Base map data (we'll discuss this for Tuesday)

- Other relevant data we want to map (like info from the census about population)

- Our favorite mapping tool!

And a very important fact about all of this (as we saw on Tuesday) is data management, organization, and planning!

Data organization

There are lots of things going into the mapping tool that you need to keep track of

Most tools will simply be a way of organizing the data to display it

Which means you need to know where the right pieces of data are when the tool is in need of them!

You will want to have

- A logical folder/directory structure

- Filenames that make sense to you

- Files that are the same in the same places

- A backup!

File types

Working with new tools likely means working with different file formats

In mapping tools, you may encounter

CSV – like Excel files, but saved in “comma separated value” format

Shapefiles – these are files formatted to work with ESRI tools, but they often work with other GIS tools, like QGIS (<https://en.wikipedia.org/wiki/Shapefile>)

These will often come packaged as a ZIP file, and you (often) need to unzip it first!

In the ZIP file, you'll see .shp, .shx, .dbf, .prj files

GeoJSON – shapefiles are proprietary (ESRI); this is quickly becoming the (open-source) alternative (<https://en.wikipedia.org/wiki/GeoJSON>)

KML – specifies set of features for online geographic display; stands for “keyhole markup language” and is common in Google Earth ([https://en.wikipedia.org/wiki/Keyhole Markup Language](https://en.wikipedia.org/wiki/Keyhole_Markup_Language))

Guiding Principle

For geographers, there is a law called the **First Law of Geography** (or **Tobler's Law**) which states that “everything is related to everything else, but near things are more related than distant things” (Tobler 1970)

Based on the idea of **friction of distance** – distance hinders interaction

How can we understand this in linguistics?

What can it tell us about language families? Language isolates?

What can it tell us about language contact?

[https://en.wikipedia.org/wiki/Tobler%27s first law of geography](https://en.wikipedia.org/wiki/Tobler%27s_first_law_of_geography)

What can we do in GIS?

Data processing = spatial statistics = formal techniques for studying the topographic/geographic properties of data

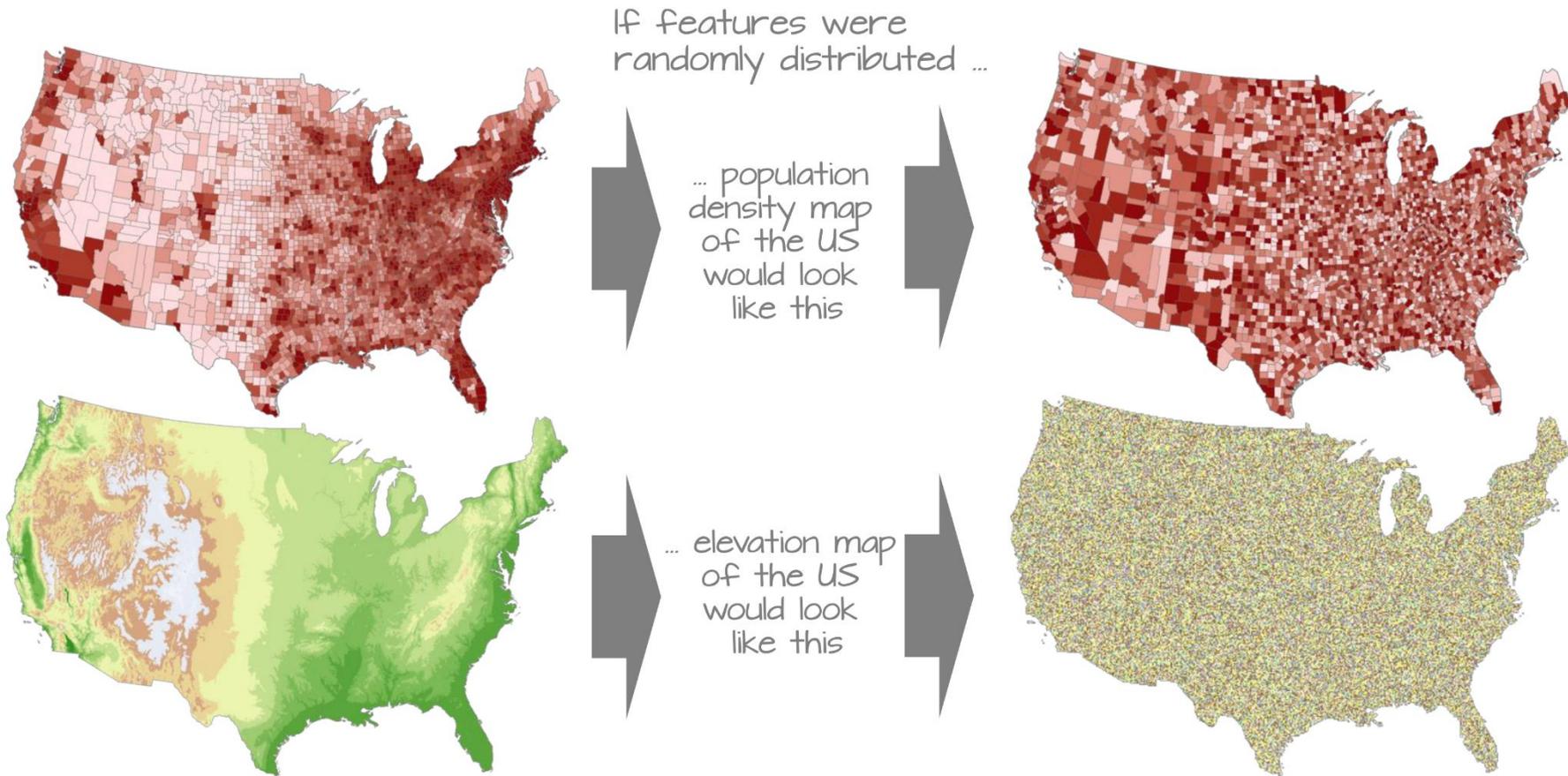
For example, we might want to see if a geographical pattern in our data exhibits “complete spatial randomness” (CSR)

If the pattern is NOT spatially random, we can find out if it is more uniform or more clustered than CSR

Can consider the study locations either with respect to the study area (whole country) or each other

Good for binary, point-pattern data – like lexical items in dialect study

What can we do in GIS?



What can we do in GIS?

Maybe we want to calculate distances between different locations

- Like if you wanted to know the most direct route from point A to point B

- Be aware of the physical boundaries or conduits of transportation...

- Distance might mean miles, but we can also talk about “linguistic distance”

Dialectometry, the study of dialects plus computational and quantitative methods (Nerbonne and Kretzschmar 2013), calculates distance between pairs of places based on the number of different responses on a dialect survey

- Because dialect data are too complex to be studied one feature at a time!

- Jack Grieve’s work is a fantastic example of combining GIS with dialectometry

Back to linguistics!

Lameli's (2010) claim that things are improving with GIS is not fully accepted

Hoch and Hayes (2010) say linguists are not doing enough!

Clearly Lameli read my dissertation, and they didn't...

But I will also note that Hoch is a med guy and Hayes is a geographer PLUS they seem to have limited their search results to "geolinguistics"

Geolinguistics, also known as "language geography", is "an interdisciplinary field that often incorporates language maps depicting spatial patterns of language location or the results of processes that lead to language change." (Hoch and Hayes 2010: 23)

Geolinguistics

They claim that GIS is being used for atlases (as we saw before) but not for much else

I, too, have found that linguists shy away from GIS

And even those who do employ the tools will seldom discuss clearly the analytical and processing components that GIS bring to their projects

Their goal in the article is to show that “Geolinguistics is poised to adapt GIS and the fundamentals of geography and cartography to address both well-developed and new questions within the field.” (24)

Geolinguistics

Geolinguists began moving to computerized linguistics data in the 1970s

Thomas (1980) provided an early example of GIS

He used it to measure spatial autocorrelation in computerized data from a linguistic survey

Described how he placed numerical values representing Welsh word usage in appropriate regions on a base map of Wales

He did so by using a grid over mapped survey sites to discover site clusters

But he said we needed something more advanced than this

Geolinguistics

The 1980s brought personal computing

At the same time, GIS products were finally moving from hard copy to digital

Unfortunately, the timing on this caused some geolinguists pause

Nonetheless, some dialectologists/geolinguists of the time went forward with GIS

Pederson's work with LAGS took some early steps toward integrating GIS, though maps were hard to read

Kretzschmar's work with LAMSAS revealed certain problems with the ways we had been mapping that GIS can readily handle, like the modified areal unit problem

Geolinguistics

Kretzschmar and LAMSAS (<http://www.lap.uga.edu/>)

From Lee and Kretzschmar 1993, via Kretzschmar 2003

Map shows “connectedness” of communities where
a particular feature was elicited

Feature = “improving weather”

“Each point where the term was found is plotted as the
center of a set of connections reaching out toward
neighboring communities. When nearby places use the
feature, the map gives the appearance of a connected
Network” (Kretzschmar 2003: 146)

FIGURE 16
LAMSAS Join-Count Analysis, *fair*
(Kretzschmar 1997, 403, fig. 6)



Geolinguistics

Kretzschmar and LAMSAS (<http://www.lap.uga.edu/>)

Connectedness can be measured

Neighboring places with same condition, or opposite

If not random, can show significant clustering = more connected

In this map, significant clusters found in the South

Shows that proximity must play a role in the distribution of linguistic features

FIGURE 16
LAMSAS Join-Count Analysis, *fair*
(Kretzschmar 1997, 403, fig. 6)



Beyond geolinguistics

Some language maps today are utilizing GIS in good ways!

Ethnologue (<https://www.ethnologue.com/map/SVHN>)

Doesn't look like UK has a subscription to this yet...

Ethnologue uses World Language Mapping System (WLMS) = vector data resource = GIS-ready

MLA Language Map (<http://arcmapping.mla.org/mla/default.aspx>)

Some even have real-world applications

Like Williams and Van der Merwe's (1996) work in Cape Town that employs GIS to discuss language policy

Or McGuirk's (2004) work in Miami that uses GIS to understand bilingualism

We will see more of this on "What else can I do?" day

ArcGIS demo

Now for something completely different...

First things first!

Know your tool, know your resources!

I can't emphasize this enough – if you are going to learn a new tool, you're going to need help

Google is a great resource, but usually the tool you are using has good resources, too

For ArcGIS Desktop, I recommend

<http://www.esri.com/news/arcuser/newtogis.html>

The help menu (built-in and web-based)

GIS Stack Exchange (<https://gis.stackexchange.com>)

A good resource manual/textbook – I like “GIS Research Methods” by Steinberg and Steinberg

ArcMap

Today, I'm going to show off some of the features of ArcMap (10.3.1)

“ArcMap is where you display and explore GIS datasets for your study area, where you assign symbols, and where you create map layouts for printing or publication. ArcMap is also the application you use to create and edit datasets. ArcMap represents geographic information as a collection of layers and other elements in a map....ArcMap is the primary application used in ArcGIS and is used to perform a wide range of common GIS tasks as well as specialized, user-specific tasks” (ESRI's website)

I'm actually going to walk you through one of ESRI's examples dealing with deforestation in the Amazon

If you have a license to ArcMap (and already have it on your laptop), you can follow along at <http://learn.arcgis.com/en/projects/get-started-with-arcmap/>

If you don't have a license, check with your institution

But most of the stuff you can do in ArcGIS you can do in QGIS...

That's all for today!

Some reminders

Do your discussion board assignment in Canvas before next class

We'll do our next hands-on activity on Tuesday

If you are a laptop bringer, please make sure you do the following:

Create a free public account on ArcGIS online at <http://www.arcgis.com/home/index.html>

Download QGIS at <http://www.qgis.org/en/site/> (should be available to all types)

Enjoy the weekend! Go see some cool Kentucky stuff!