Gabmap

Doing dialect analysis on the web

Therese Leinonen, Peter Kleiweg, Charlotte Gooskens, John Nerbonne

university of groningen
faculty of arts

What is Gabmap?

- Gabmap is a web application for dialectometrics and cartography
- allows you to make mappings and statistical analyses of your dialect data
- Gabmap can analyze phonetic transcriptions, numeric dialect data (e.g, formant frequencies), and categorical data (e.g., syntactic or morphological variables)
- overviews of the number of places, number of linguistic variables, number of different symbols used in the transcriptions etc.
- distribution maps
- measurement of linguistic distances
- statistical tools for analyzing and mapping dialect continuum (multidimensional scaling) and identifying dialect groups (cluster analysis)
- all maps and figures can be downloaded as image files

Visit www.gabmap.nl to:

- try a demo
- create a user account and start using Gabmap
- watch an introduction video
- follow a step-by-step tutorial with exercises
- read our FAQs
- find out about up-coming events

Example of Gabmap workflow with dialect data from Pennsylvania:

1) collect and upload data

START
by collecting dialect data

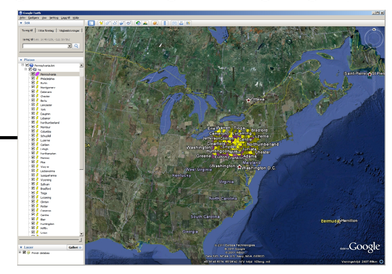
dialect data
transcriptions / numeric data / categorical data

format: tab-separated table

geographic data
site coordinates and borders

format: .kml or .kmz

Use Google Earth to get the geographic data:



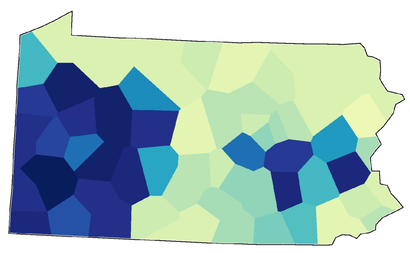
2) data inspection

index map

```
1     Philadelphia
2     Bucks
3     Montgomery
4     Delaware
5     Chester
6     Berks
7     Lancaster
8     York
9     Dauphin
10    Lebanon
[...]
```

data overviews

distribution / value maps



3) measurement of linguistic distances

string data (=transcriptions)

numeric data categorical data

Philadelphia — Schuylkill

k	ə	n	æ	ə	ɬ	
k	ə	n	æ		l	
				l	l	2

string alignments

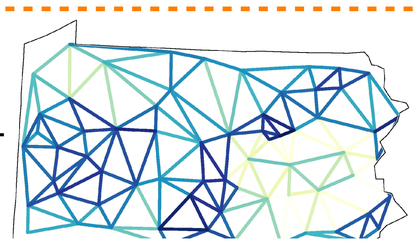
item distances

aggregation

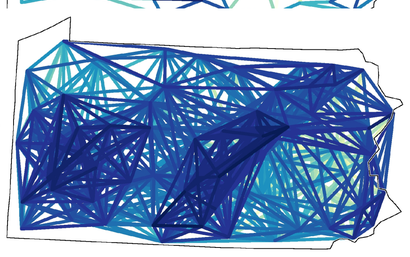
distance matrix

4) linguistic differences

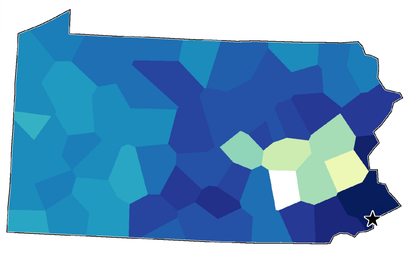
difference maps

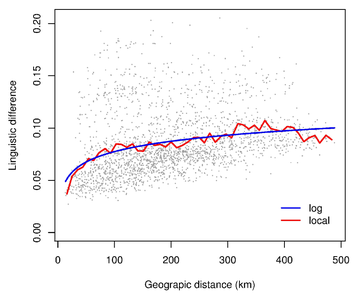


comparisons with geographic distances

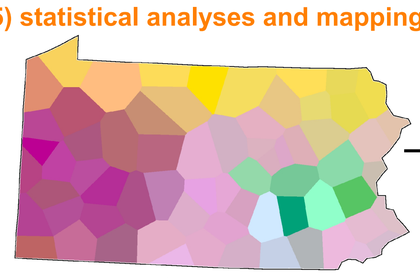


reference point maps

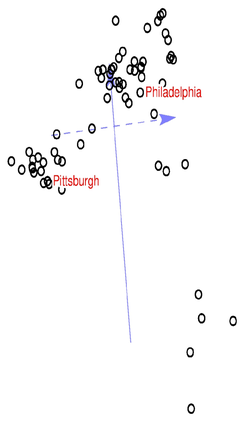




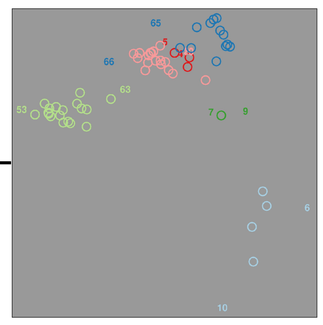
5) statistical analyses and mappings

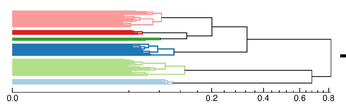


multi-dimensional scaling

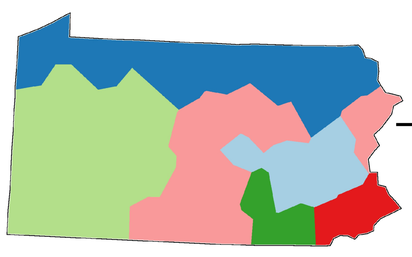


cluster validation

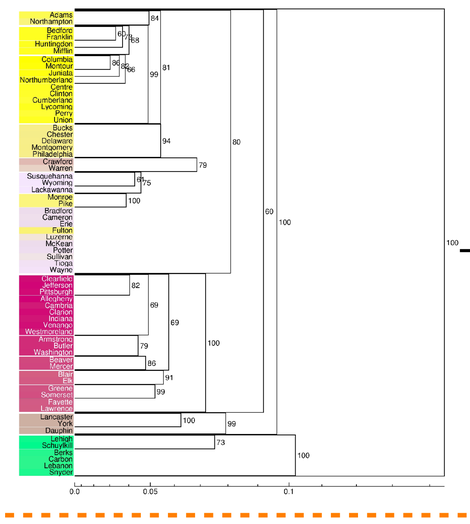


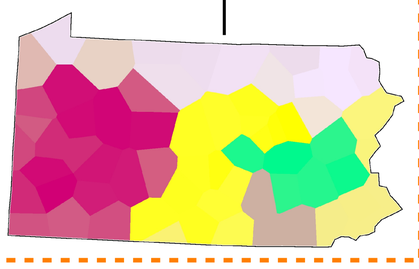


discrete clustering



fuzzy clustering





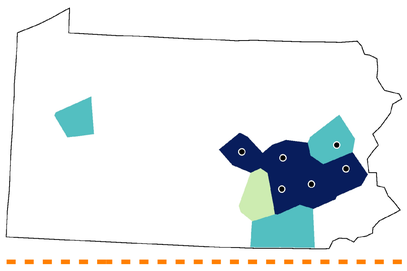
6) data mining

cluster determinants

Current item: Philadelphia
Importance (weighted): 0.78
Distinctiveness (weighted): 0.69
Representativeness (weighted): 0.87
Patterns with forms:

- 0.71 - 0.74 - 0.67 - fɪlədɛlfjə (11)
- 0.46 - 0.72 - 0.20 - fɪlədɛlfɪə (2)
- 0.24 - 0.34 - 0.13 - fɪlədɛlfɪ (3)

distribution maps



Reference:
John Nerbonne, Rinke Colen, Charlotte Gooskens, Peter Kleiweg, and Therese Leinonen (2011). Gabmap — A Web Application for Dialectology. Submitted to *Dialectologia*.