

Visualisierungstechniken für raumbezogene Daten

Yves Scherrer, Universität Helsinki

Alm-Workshop Geolinguistik im Alpenraum, 9. Juni 2018

Einführung

Visualisierungstechniken für raumbezogene Daten

Konkret:

- Visualisierung von raumbezogenen Daten = Kartierung
- Daten = Sprachdaten / Sprachvariationsdaten

Programm:

- Einführung in die Dialektkartografie
- Hands-on session: Kartografie mit R

Folien und Hands-on-Material verfügbar unter:

<https://www.github.com/yvesscherrer/geolinguistik>

Visualisierungstechniken für raumbezogene Daten

Konkret:

- Visualisierung von raumbezogenen Daten = Kartierung
- Daten = Sprachdaten / Sprachvariationsdaten

Programm:

- Einführung in die Dialektkartografie
- Hands-on session: Kartografie mit R

Folien und Hands-on-Material verfügbar unter:

<https://www.github.com/yvesscherrer/geolinguistik>

Visualisierungstechniken für raumbezogene Daten

Konkret:

- Visualisierung von raumbezogenen Daten = Kartierung
- Daten = Sprachdaten / Sprachvariationsdaten

Programm:

- Einführung in die Dialektkartografie
- Hands-on session: Kartografie mit R

Folien und Hands-on-Material verfügbar unter:

<https://www.github.com/yvesscherrer/geolinguistik>

Woraus besteht eine Karte?

1. Geografische Objekte: Punkte, Linien, Polygone, Zellen



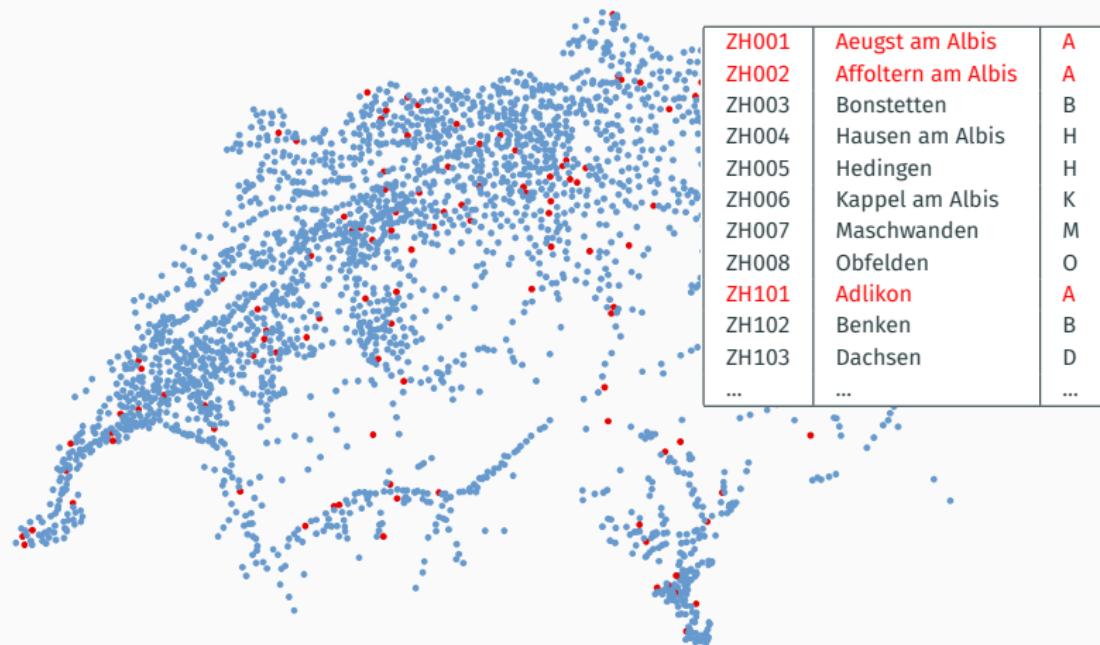
Woraus besteht eine Karte?

2. Attribute: Kategorien, Frequenzen, Proportionen, ...

ZH001	Augst am Albis	A
ZH002	Affoltern am Albis	A
ZH003	Bonstetten	B
ZH004	Hausen am Albis	H
ZH005	Hedingen	H
ZH006	Kappel am Albis	K
ZH007	Maschwanden	M
ZH008	Obfelden	O
ZH101	Adlikon	A
ZH102	Benken	B
ZH103	Dachsen	D
...

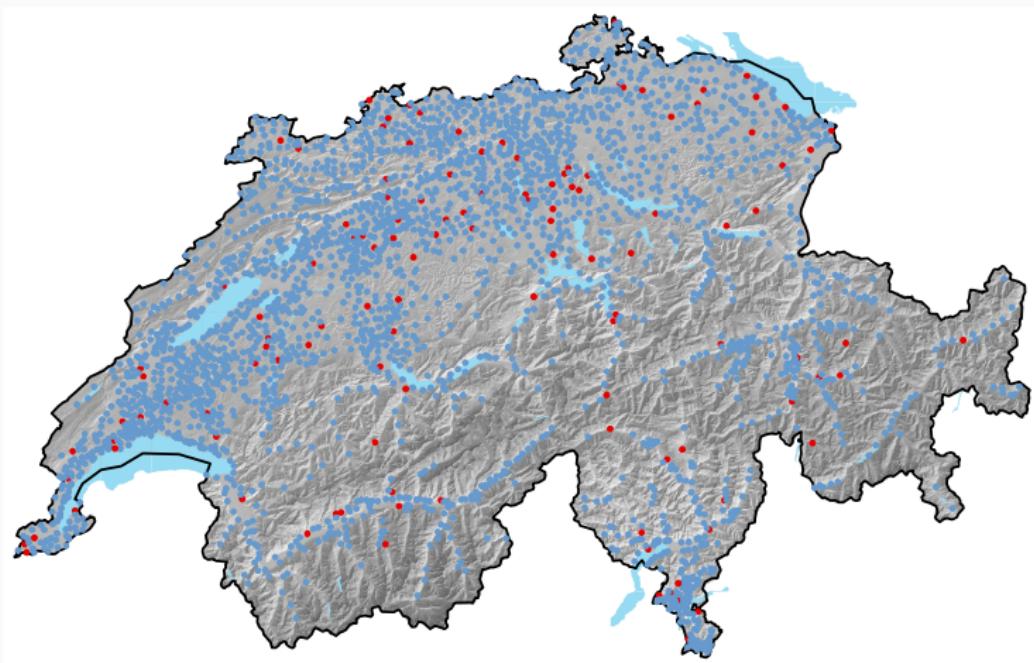
Woraus besteht eine Karte?

Kartierung: Darstellung der geografischen Objekte gemäss Attributwerten



Woraus besteht eine Karte?

3. Hintergrundelemente, Legenden



Woraus besteht eine Karte?

3. Hintergrundelemente, Legenden



Woraus besteht eine Karte?

1. **Geografische Objekte:** Punkte, Linien, Polygone, Zellen
2. **Attribute:** Kategorien, Frequenzen, Proportionen, ...
Kartierung: Darstellung der geografischen Objekte
gemäss Attributwerten
3. **Hintergrundelemente, Legenden**

Arbeitsablauf:

- Welche Attribute möchte ich darstellen?
- Welche geografischen Objekte eignen sich dafür?
- Welche Hintergrundelemente fördern das Verständnis der Karte?

Woraus besteht eine Karte?

1. **Geografische Objekte:** Punkte, Linien, Polygone, Zellen
2. **Attribute:** Kategorien, Frequenzen, Proportionen, ...
Kartierung: Darstellung der geografischen Objekte
gemäss Attributwerten
3. **Hintergrundelemente, Legenden**

Arbeitsablauf:

- Welche Attribute möchte ich darstellen?
- Welche geografischen Objekte eignen sich dafür?
- Welche Hintergrundelemente fördern das Verständnis der Karte?

Traditionelle Dialektatlas-Projekte

Traditionelle Dialektatlas-Projekte (Erste Hälfte 20. Jh.)

Typische Eigenschaften:

- Gleichmäßig verteilte Ortspunkte
- Ein Informant pro Ortspunkt, d.h. eine Antwort pro Frage
- Kategorielle Antworten (Varianten)

Visualisierungsmöglichkeiten:

- Punktsymbolkarten
- Punkt-Farbe-Karten
- Farbige Thiessen-Polygone

Beispiel: Sprachatlas der deutschen Schweiz (SDS)

<http://www.dialektkarten.ch>

Typische Eigenschaften:

- Gleichmäßig verteilte Ortspunkte
- Ein Informant pro Ortspunkt, d.h. eine Antwort pro Frage
- Kategorielle Antworten (Varianten)

Visualisierungsmöglichkeiten:

- Punktsymbolkarten
- Punkt-Farbe-Karten
- Farbige Thiessen-Polygone

Beispiel: Sprachatlas der deutschen Schweiz (SDS)

<http://www.dialektkarten.ch>

Traditionelle Dialektatlas-Projekte (Erste Hälfte 20. Jh.)

Typische Eigenschaften:

- Gleichmäßig verteilte Ortspunkte
- Ein Informant pro Ortspunkt, d.h. eine Antwort pro Frage
- Kategorielle Antworten (Varianten)

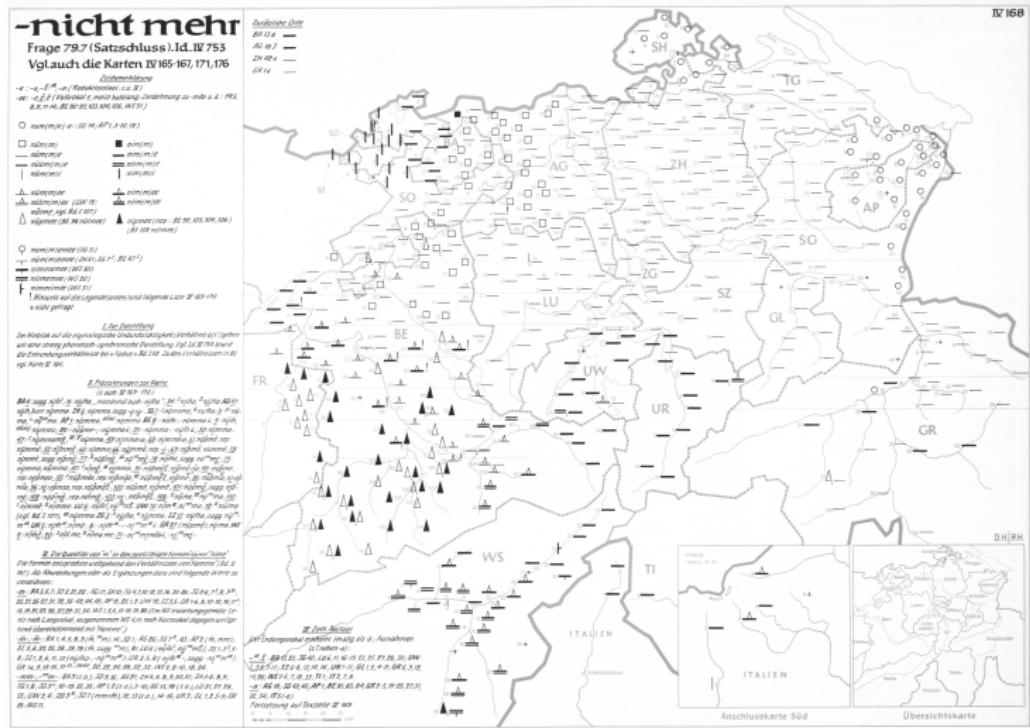
Visualisierungsmöglichkeiten:

- Punktsymbolkarten
- Punkt-Farbe-Karten
- Farbige Thiessen-Polygone

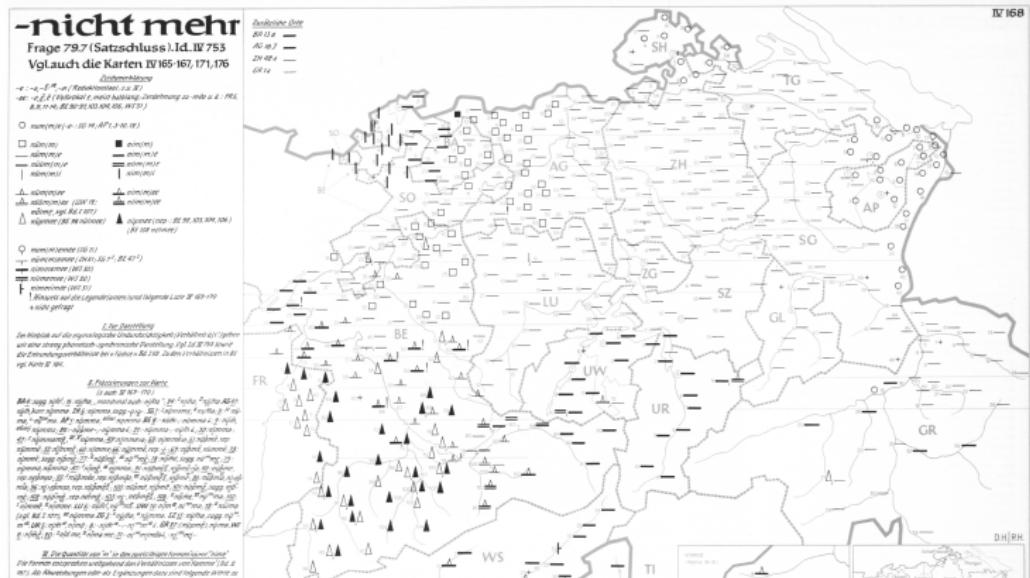
Beispiel: Sprachatlas der deutschen Schweiz (SDS)

<http://www.dialektkarten.ch>

Punktsymbolkarten



Punktsymbolkarten

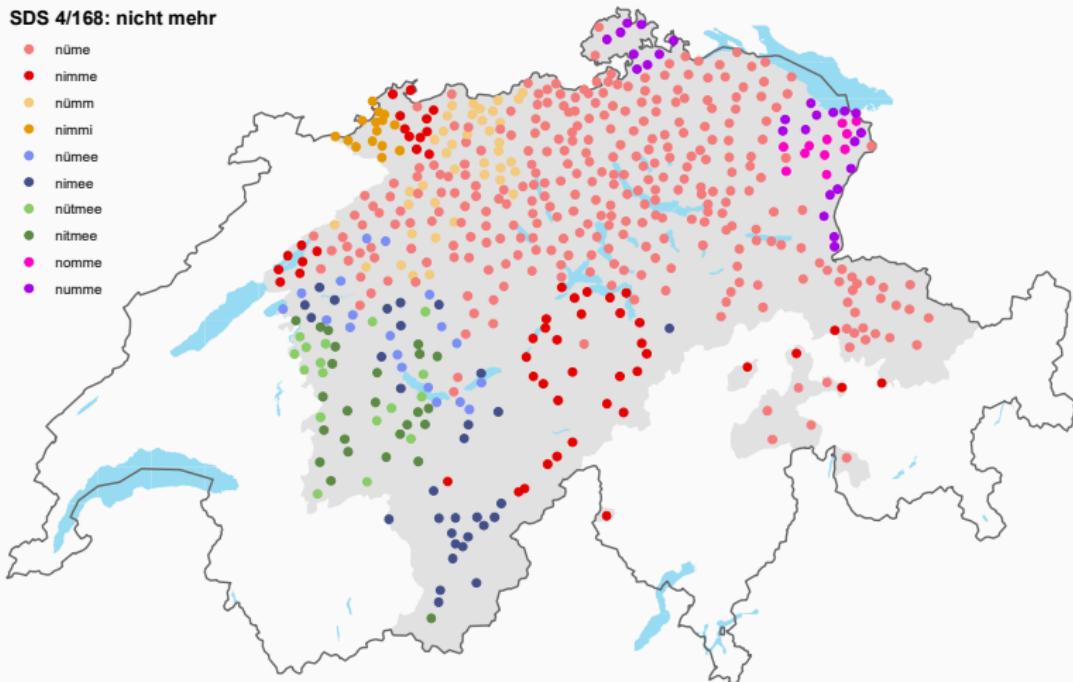


- 1 Ortspunkt = 1 Symbol (mit wenigen Ausnahmen)
 - Ähnliche Symbolformen = ähnliche Varianten

Farbpunktkarten

SDS 4/168: nicht mehr

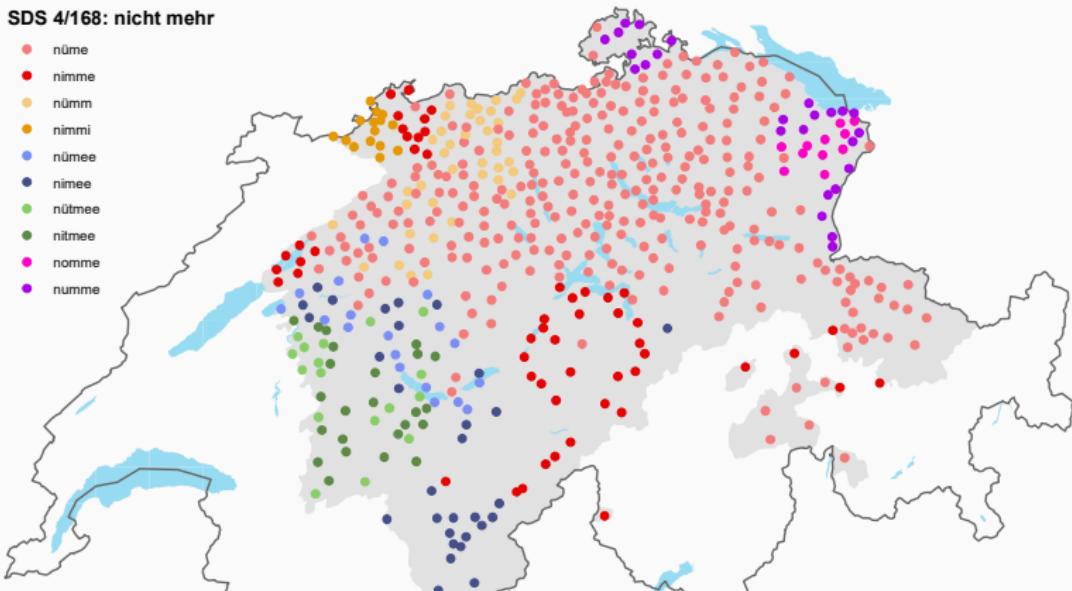
- nüme
- nimme
- nümm
- nimmi
- númee
- nímee
- nütmee
- nomme
- numme



Farbpunktkarten

SDS 4/168: nicht mehr

- nüme
- nimme
- nümm
- nimmi
- númee
- nímee
- nütmee
- nomme
- numme

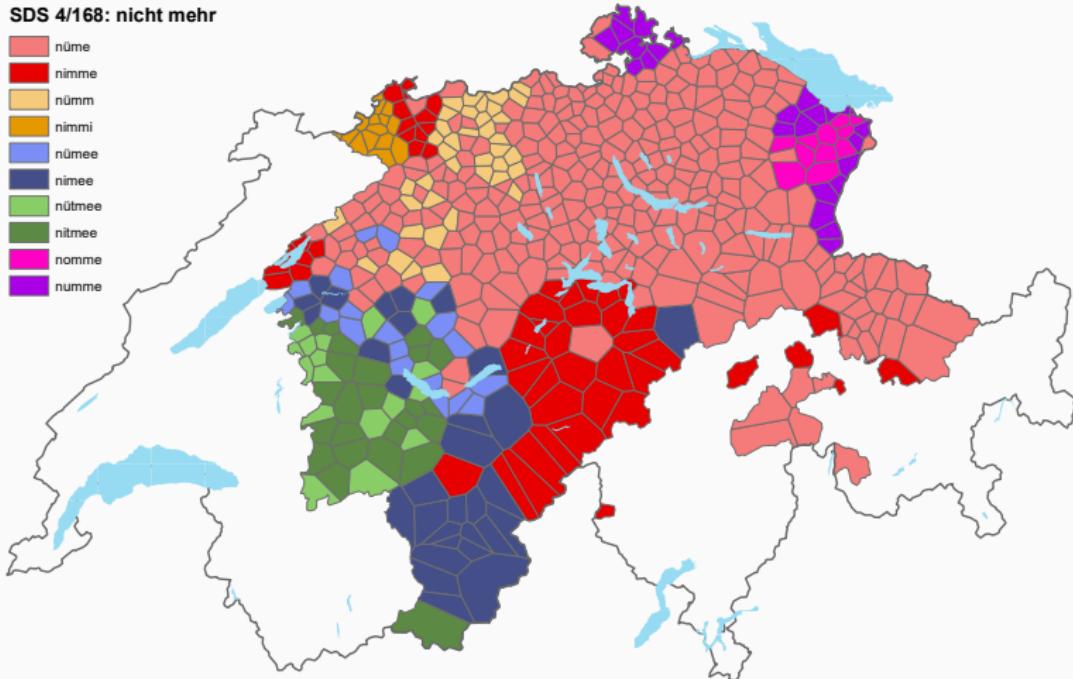


- 1 Ortspunkt = 1 Symbol
- Ähnliche Farben = ähnliche Varianten

Thiessen-Polygone

SDS 4/168: nicht mehr

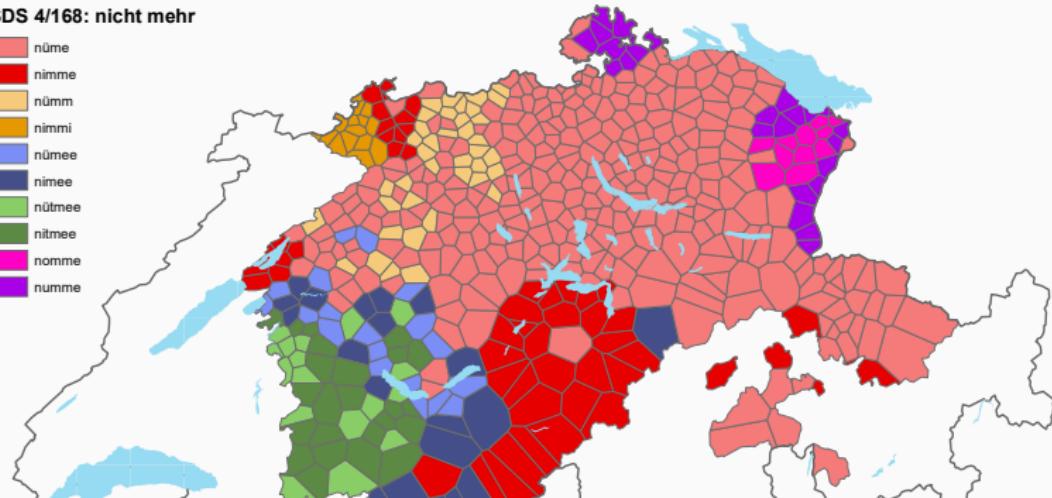
- [red] nüme
- [dark red] nimme
- [orange] nümm
- [yellow] nimmi
- [light blue] nümee
- [dark blue] nimee
- [green] nütmee
- [dark green] nütme
- [pink] nomme
- [purple] numme



Thiessen-Polygone

SDS 4/168: nicht mehr

- [red] nüme
- [dark red] nimme
- [orange] nümm
- [yellow] nimmi
- [light blue] nümee
- [dark blue] nimee
- [green] nütmee
- [dark green] nütmee
- [pink] nomme
- [purple] numme

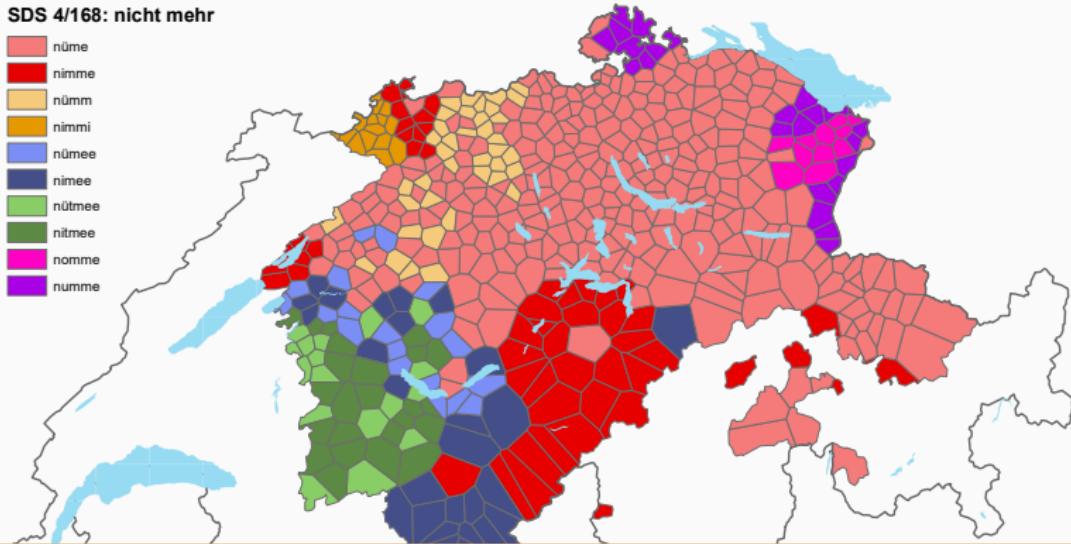


- Thiessen-Polygone = Voronoi-Diagramm
- Um jeden Punkt wird ein Polygon gelegt, so dass die Polygonkanten den Bisektoren zwischen je zwei Punkten entsprechen

Thiessen-Polygone

SDS 4/168: nicht mehr

- nüme
- nimme
- nümm
- nimm
- nüme
- nimee
- nütmee
- nütmee
- nomme
- numme

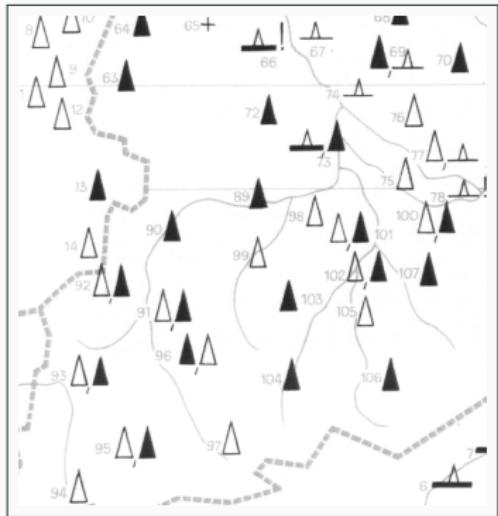


Annahmen:

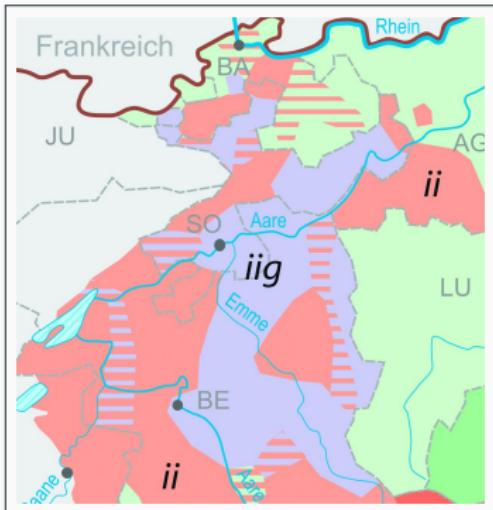
- Ortspunkte sind repräsentativ für deren Umgebung
- Attribute sind auf die gesamte Fläche anwendbar

Mehrfachantworten

Mehrere Symbole pro
Ortspunkt, durch Komma
getrennt (SDS):



Schraffur auf Polygonkarten
(Kleiner Sprachatlas der
deutschen Schweiz, KSDS):



Neuere Dialektatlas-Projekte

Typische Eigenschaften:

- Gleichmässig verteilte Ortspunkte
- Mehrere Informanten pro Ortspunkt
- Frequenzverteilung über Varianten pro Ort und Variable

Beispiel: Syntaktischer Atlas der deutschen Schweiz (SADS)

- 4 Fragebögen mit je ca. 30 Fragen zu (morpho-) syntaktischen Phänomenen
- 383 Ortspunkte, 3187 Informanten
(d.h. 3-26 Informanten pro Ortspunkt)

E. Glaser & G. Bart (2015): Dialektsyntax des Schweizerdeutschen.

In: R. Kehrein, A. Lameli, S. Rabanus (Hrsg.): *Regionale Variation des Deutschen – Projekte und Perspektiven*. Berlin: De Gruyter, 79–105.

Typische Eigenschaften:

- Gleichmässig verteilte Ortspunkte
- Mehrere Informanten pro Ortspunkt
- Frequenzverteilung über Varianten pro Ort und Variable

Beispiel: Syntaktischer Atlas der deutschen Schweiz (SADS)

- 4 Fragebögen mit je ca. 30 Fragen zu (morpho-)syntaktischen Phänomenen
- 383 Ortspunkte, 3187 Informanten
(d.h. 3-26 Informanten pro Ortspunkt)

E. Glaser & G. Bart (2015): Dialektsyntax des Schweizerdeutschen.

In: R. Kehrein, A. Lameli, S. Rabanus (Hrsg.): *Regionale Variation des Deutschen – Projekte und Perspektiven*. Berlin: De Gruyter, 79–105.

Neuere Dialektatlas-Projekte

SADS Frage I/9: Ich weiss auch nicht, ob er einmal ...

	Ebnat-Kappel	Wildhaus	Grabs	Sevelen
... heiraten will	0%	25%	50%	60%
... will heiraten	100%	75%	50%	40%

Visualisierungsmöglichkeiten:

- Dominante Variante auswählen + Punkt-/Polygon-Karte
- „Räpplikarten“ / Tortendiagramm-Karten
- Gewichtete Thiessen-Polygone
- Rasterkarten

Neuere Dialektatlas-Projekte

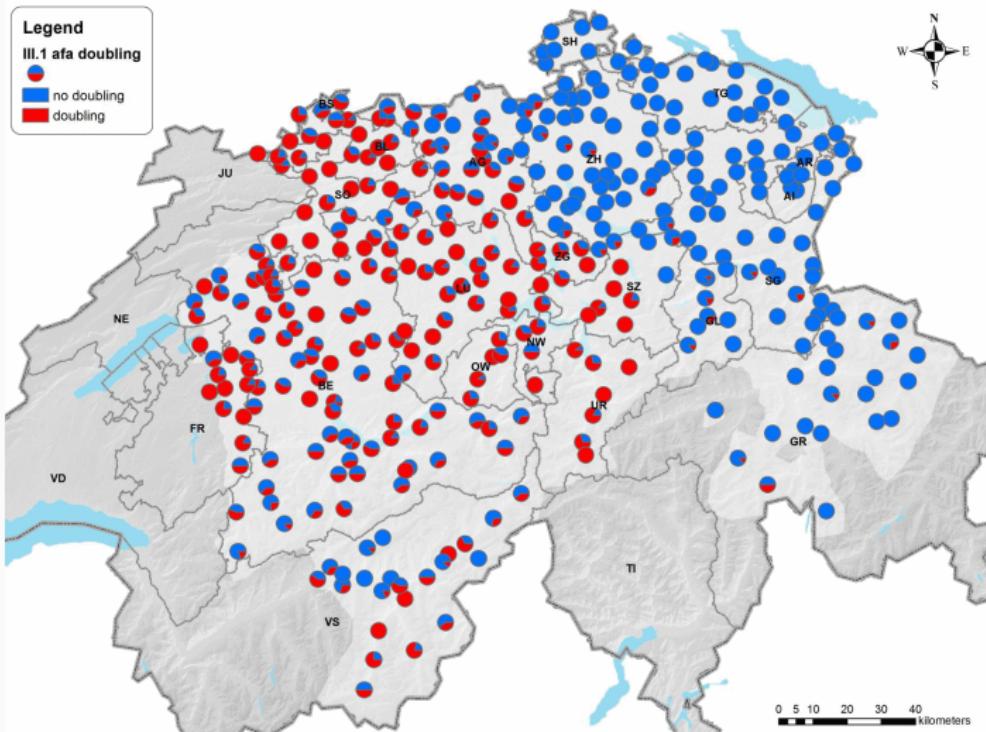
SADS Frage I/9: Ich weiss auch nicht, ob er einmal ...

	Ebnat-Kappel	Wildhaus	Grabs	Sevelen
... heiraten will	0%	25%	50%	60%
... will heiraten	100%	75%	50%	40%

Visualisierungsmöglichkeiten:

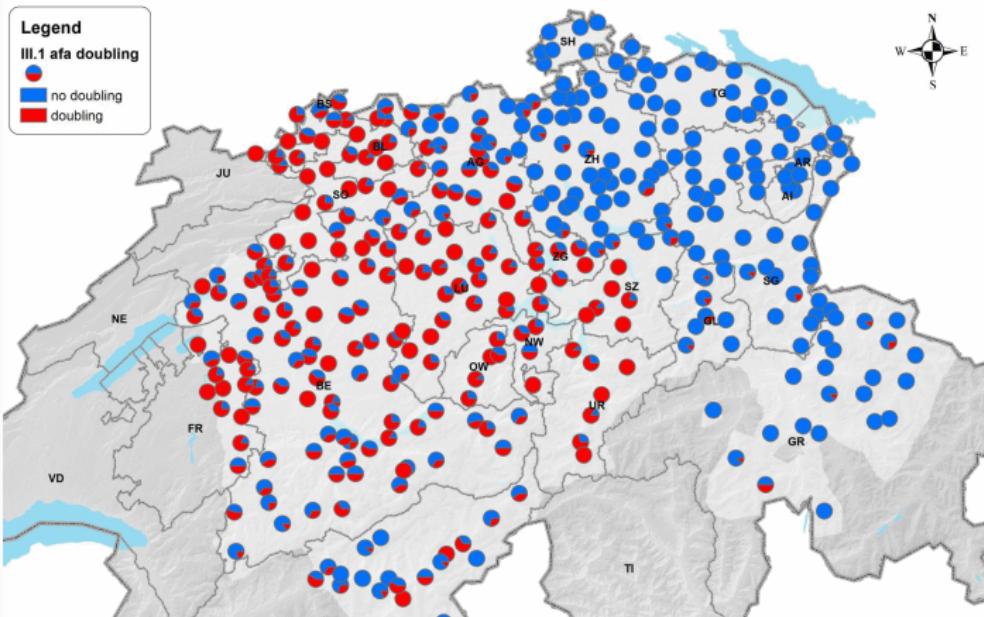
- Dominante Variante auswählen + Punkt-/Polygon-Karte
- „Räpplikarten“ / Tortendiagramm-Karten
- Gewichtete Thiessen-Polygone
- Rasterkarten

Räpplikarten / Tortendiagramm-Karten



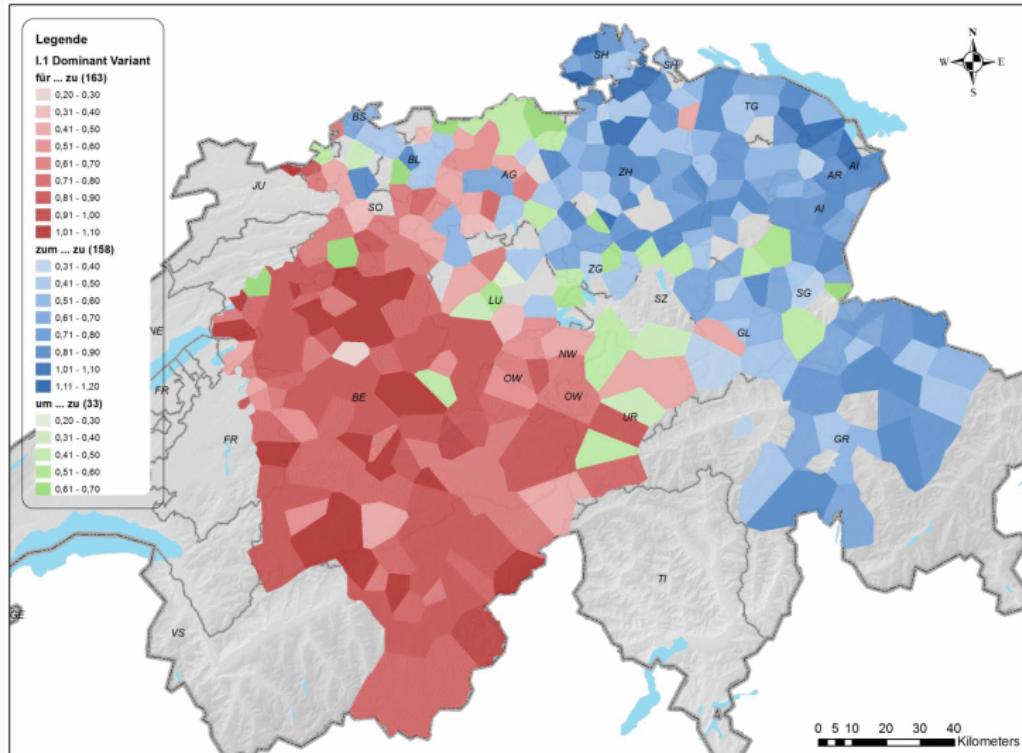
Karte: Philipp Stoeckle

Räpplikarten / Tortendiagramm-Karten

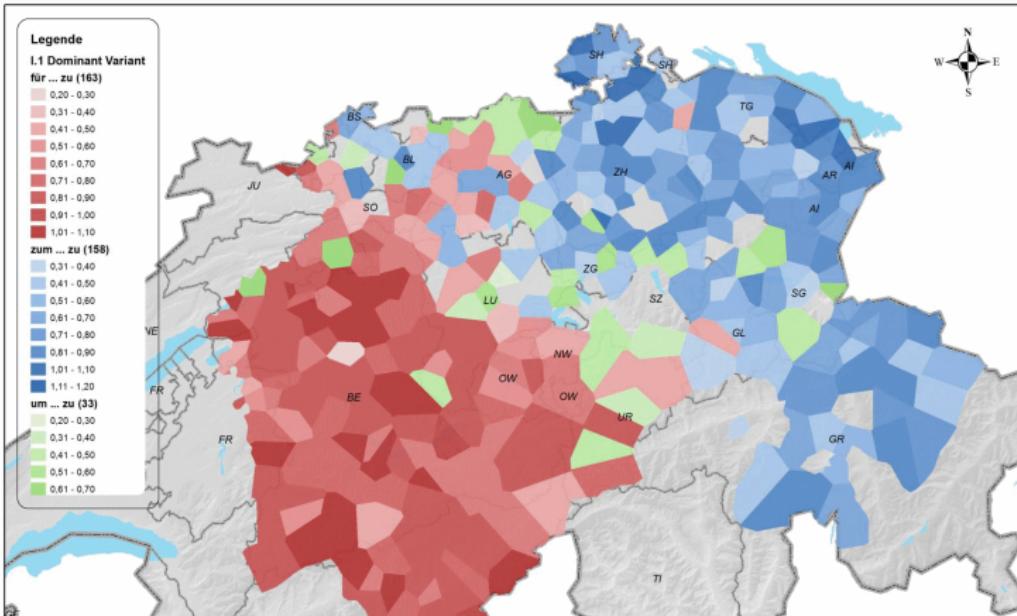


Jeder Erhebungspunkt entspricht einem Tortendiagramm, das die Proportionen der verschiedenen Varianten in verschiedenen Farben darstellt

Gewichtete Thiessen-Polygone

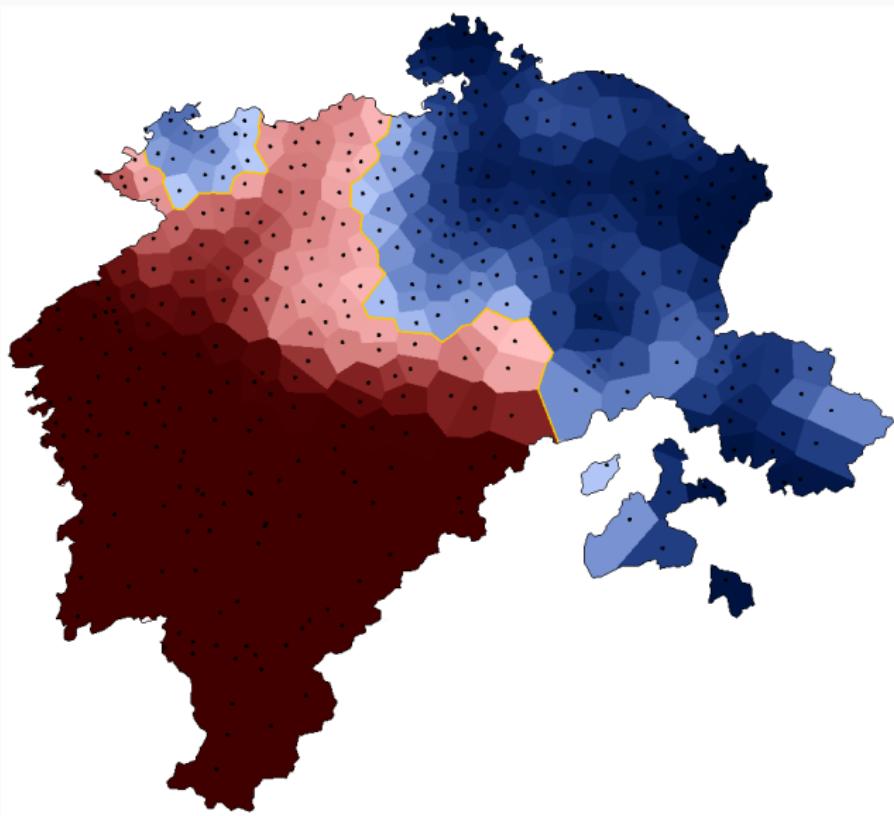


Gewichtete Thiessen-Polygone

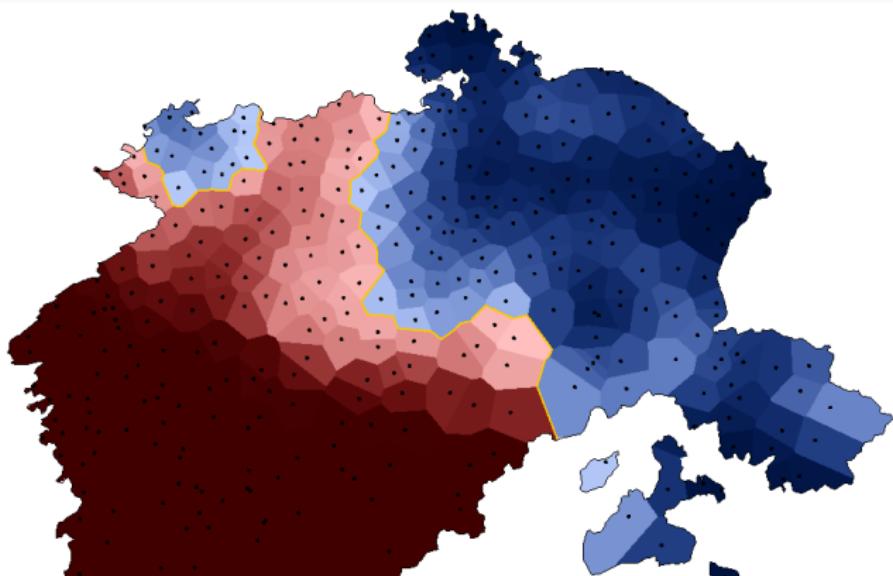


- Verschiedene Farben = verschiedene dominante Varianten
- Farbintensität = Grad der Dominanz

Gewichtete Thiessen-Polygone mit Glättung

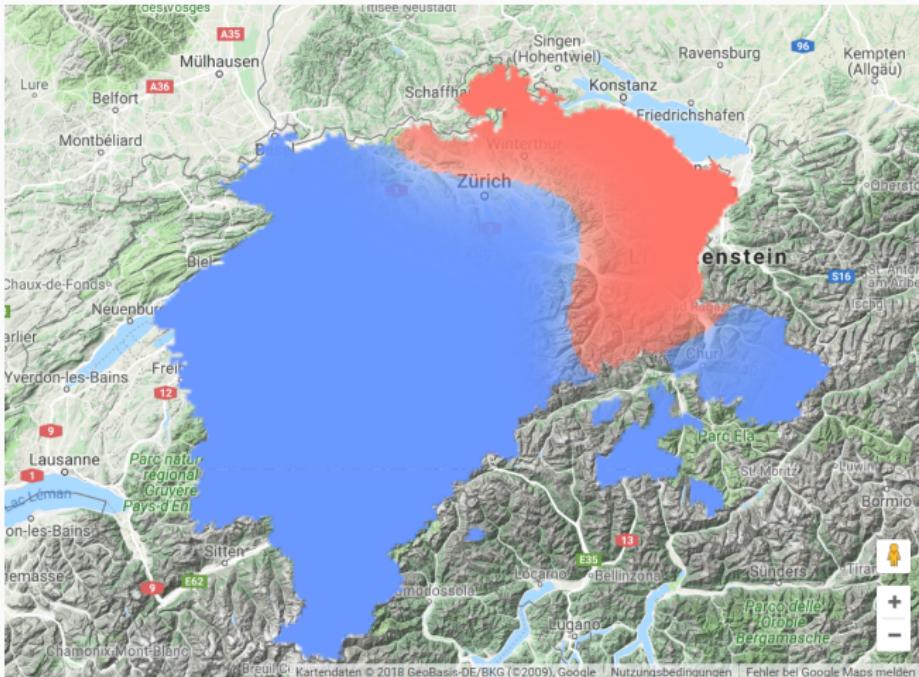


Gewichtete Thiessen-Polygone mit Glättung



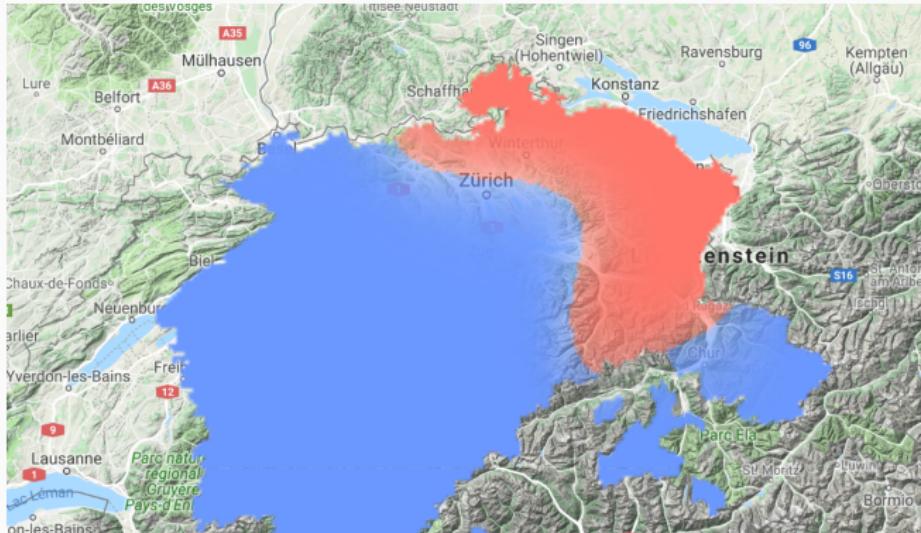
- Glättung der Frequenzen, z.B. um Outlier zu entfernen
- Hier: Kerndichteschätzung (<http://www.geoling.net>)
- Grüne Variante entfällt komplett

Rasterkarten



http://dialektkarten.ch/dominance.de.html#3262_gehen_lassen

Rasterkarten



- Keine Thiessen-Polygone, sondern regelmässiges Netz aus quadratischen Zellen (**Raster**)
- Für jede Zelle werden geglättete Dominanzwerte errechnet

Dialektometrie

Typische Eigenschaften:

- Berechnung von Kennzahlen auf der Basis von aggregierten Einzelkarten
- Ortsnetz entspricht dem Grunddatensatz
- Werteverteilung über Ortspunkte, i.d.R. Ordinaldaten

Visualisierungsmöglichkeiten:

- Gewichtete Thiessen-Polygone
- Farbverläufe (Regenbogenkarten, Heat Maps)

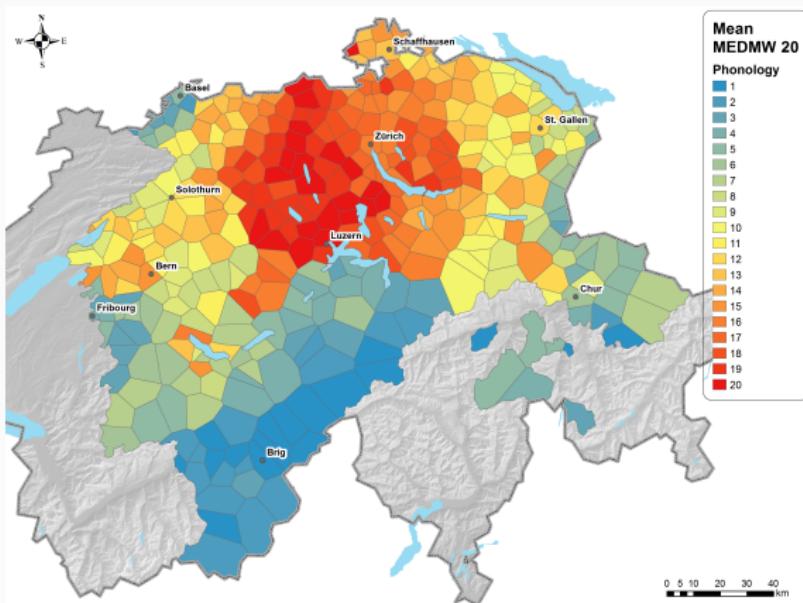
Typische Eigenschaften:

- Berechnung von Kennzahlen auf der Basis von aggregierten Einzelkarten
- Ortsnetz entspricht dem Grunddatensatz
- Werteverteilung über Ortspunkte, i.d.R. Ordinaldaten

Visualisierungsmöglichkeiten:

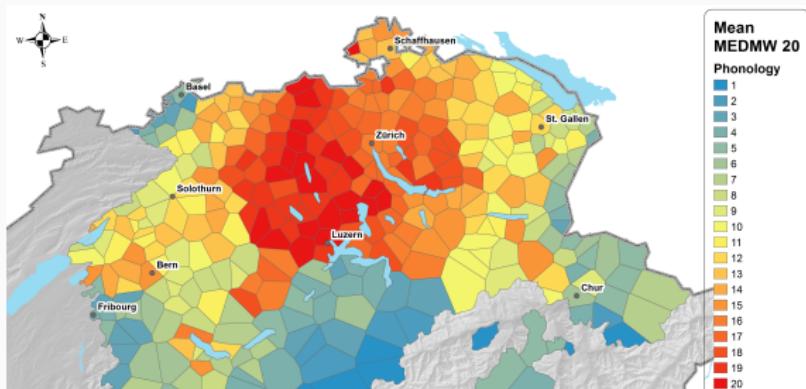
- Gewichtete Thiessen-Polygone
- Farbverläufe (Regenbogenkarten, Heat Maps)

Beispiel: Lautkarten des SDS Mittlere Distanz zu allen anderen Ortspunkten



Scherrer, Y., & Stoeckle, P. (2016). A quantitative approach to Swiss German –
Dialectometric analyses and comparisons of linguistic levels. *Dialectologia et Geolinguistica*, 24(1), 92-125.

Beispiel: Lautkarten des SDS Mittlere Distanz zu allen anderen Ortspunkten



- Rot: geringe Distanz zu anderen Punkten,
„Durchschnittsdialekte“
- Blau: grosse Distanz zu anderen Punkten,
„periphere Dialekte“

Crowdsourcing

Crowdsourcing-Projekte

Typische Eigenschaften:

- Keine effektive Steuerung der geografischen Verteilung der Teilnehmer
 - Aggregation von Teilnehmern aus naheliegenden Punkten
 - Interpolation bei grösseren Lücken
- Mehrere Informanten pro Ortspunkt
- Frequenzverteilung über Varianten pro Ort und Variable
 - Unter Umständen extreme Dominanz einer Standardvariante

Visualisierungsmöglichkeiten:

- Ausgewählte Varianten + Punkt-/Polygonkarten
- „Räpplikarten“ / Tortendiagramm-Karten
- Gewichtete Thiessen-Polygone / administrative Einheiten
- Rasterkarten

Crowdsourcing-Projekte

Typische Eigenschaften:

- Keine effektive Steuerung der geografischen Verteilung der Teilnehmer
 - Aggregation von Teilnehmern aus naheliegenden Punkten
 - Interpolation bei grösseren Lücken
- Mehrere Informanten pro Ortspunkt
- Frequenzverteilung über Varianten pro Ort und Variable
 - Unter Umständen extreme Dominanz einer Standardvariante

Visualisierungsmöglichkeiten:

- Ausgewählte Varianten + Punkt-/Polygonkarten
- „Räpplikarten“ / Tortendiagramm-Karten
- Gewichtete Thiessen-Polygone / administrative Einheiten
- Rasterkarten

Crowdsourcing-Projekte

Typische Eigenschaften:

- Keine effektive Steuerung der geografischen Verteilung der Teilnehmer
 - Aggregation von Teilnehmern aus naheliegenden Punkten
 - Interpolation bei grösseren Lücken
- Mehrere Informanten pro Ortspunkt
- Frequenzverteilung über Varianten pro Ort und Variable
 - Unter Umständen extreme Dominanz einer Standardvariante

Visualisierungsmöglichkeiten:

- Ausgewählte Varianten + Punkt-/Polygonkarten
- „Räpplikarten“ / Tortendiagramm-Karten
- Gewichtete Thiessen-Polygone / administrative Einheiten
- Rasterkarten

Cartophono: Projekt und Teilnehmer

- Online-Umfrage zur Nord-Süd-Variation bei der *loi de position* im Französischen:

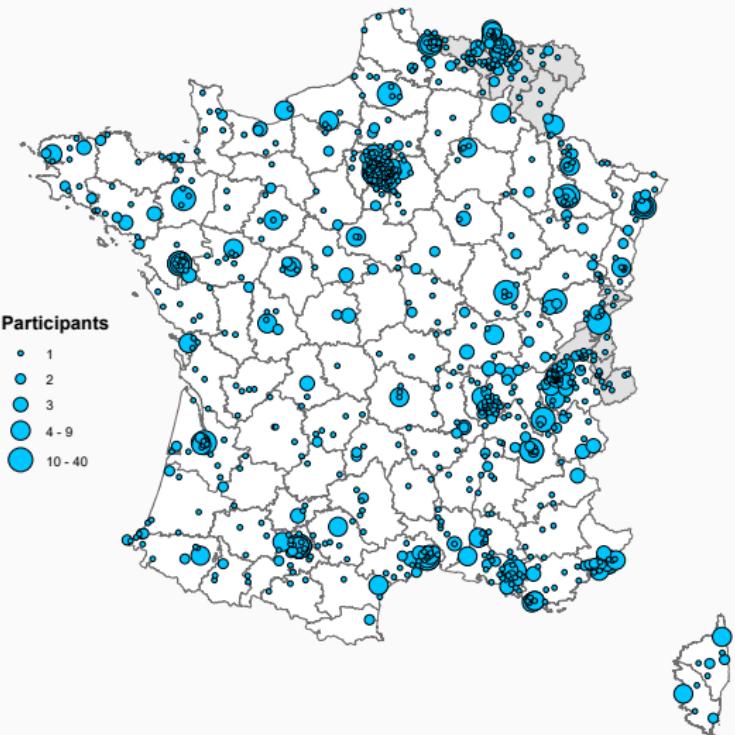
Syllable	Vowel	E	Œ	O
Closed (CVC)	Open	ɛ <i>sel</i>	œ ø <i>neutre</i>	ɔ ɔ̃ <i>chose</i>
Open (CV)	Close	ɛ e <i>lait</i>	œ <i>ceux</i>	o <i>sot</i>

- 70 Stimuli:
 - Je zwei Aussprachen werden vorgespielt, Teilnehmer sollen die ihrer eigenen Aussprache ähnlichere auswählen
- Erhebungszeitraum Oktober 2014 – August 2015:
 - 1250 TeilnehmerInnen, 72% weiblich, 40% 20-30 Jahre alt

Scherrer, Y., Boula de Mareüil, P., & Goldman, J.-P. (2015):
Crowdsourced mapping of pronunciation variants in European French.
Proceedings of the 18th International Congress of Phonetic Science.

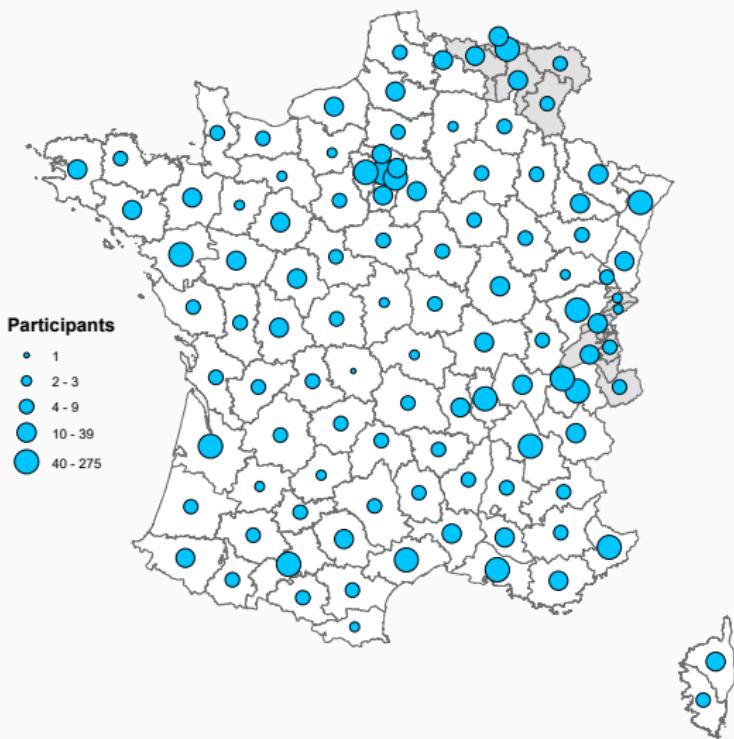
Cartophono: Verteilung der Teilnehmer

Originalangabe: Postleitzahlen



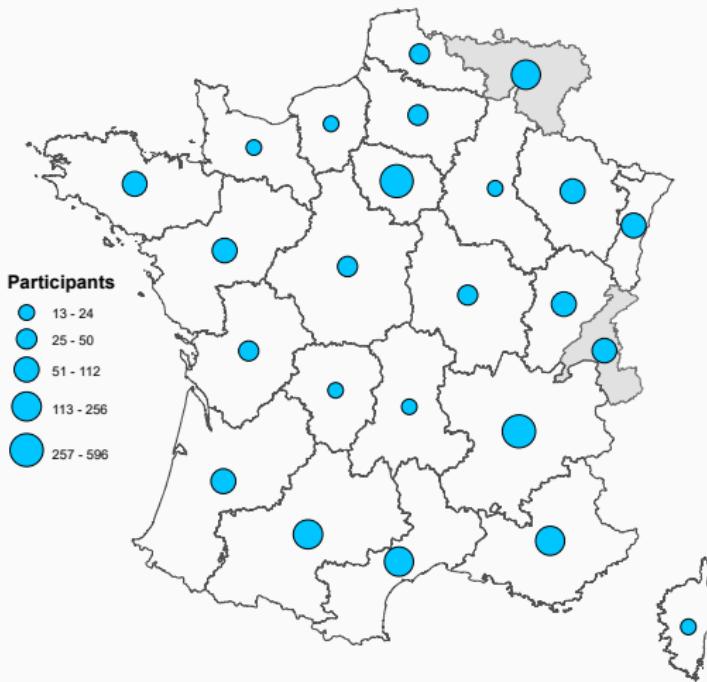
Cartophono: Verteilung der Teilnehmer

Aggregierung: Départements/Kantone/Provinzen



Cartophono: Verteilung der Teilnehmer

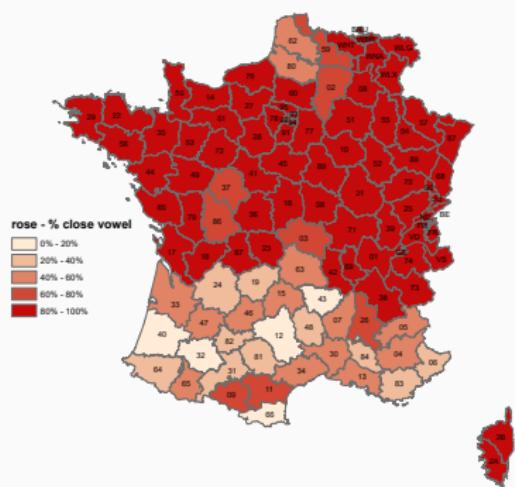
Aggregierung: Regionen/Sprachregionen



Cartophono: Darstellung der Resultate

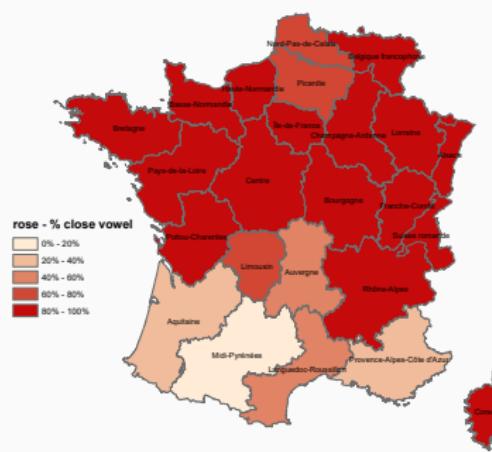
Rose

Departementsweise
Aggregierung



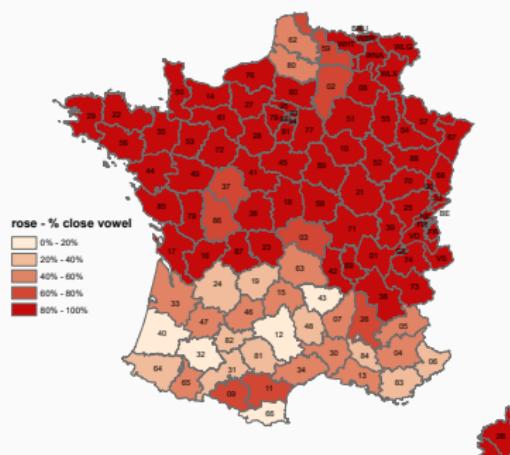
Rose

Regionsweise
Aggregierung

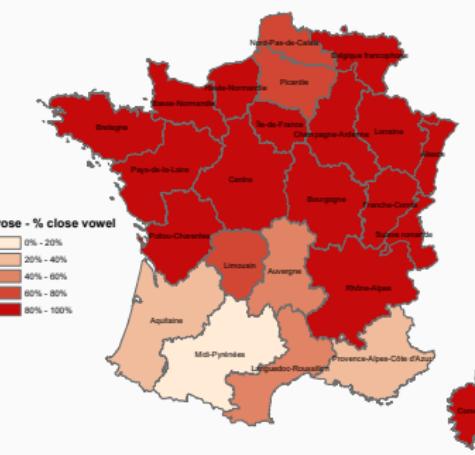


Cartophono: Darstellung der Resultate

Rose
Departementsweise
Aggregierung



Rose
Regionsweise
Aggregierung

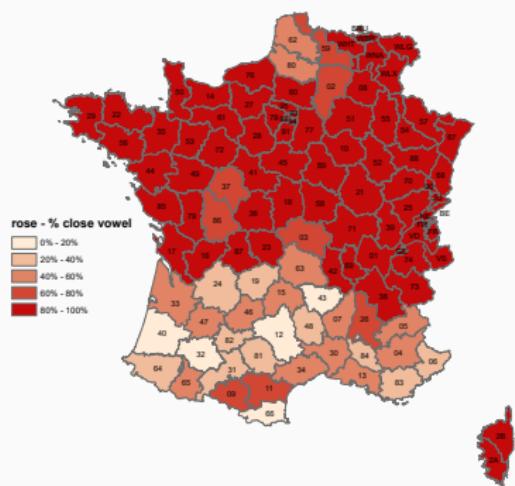


Anstatt Thiessen-Polygone werden die echten
Gebietsgrenzen verwendet

Cartophono: Darstellung der Resultate

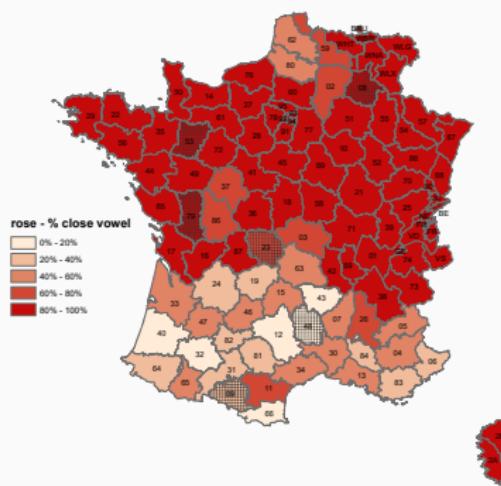
Rose

Departementsweise
Aggregierung



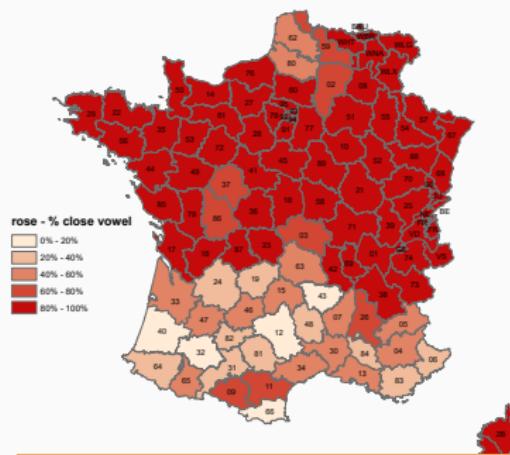
Rose

Departementsweise
Aggregierung mit Glättung

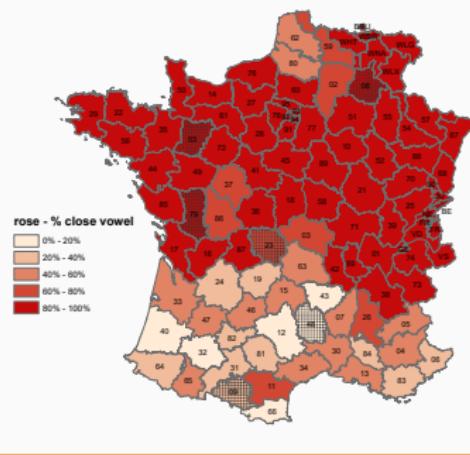


Cartophono: Darstellung der Resultate

Rose
Departementsweise
Aggregierung



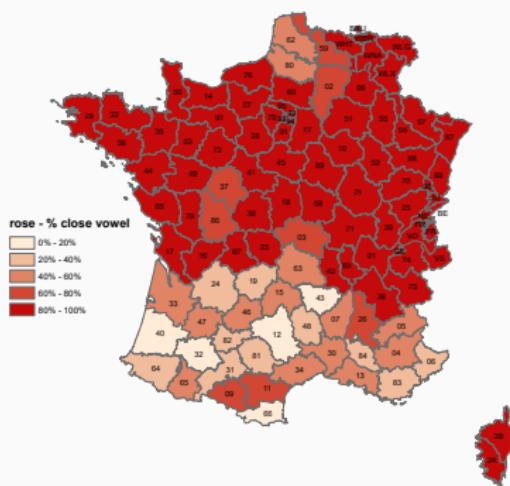
Rose
Departementsweise
Aggregierung mit Glättung



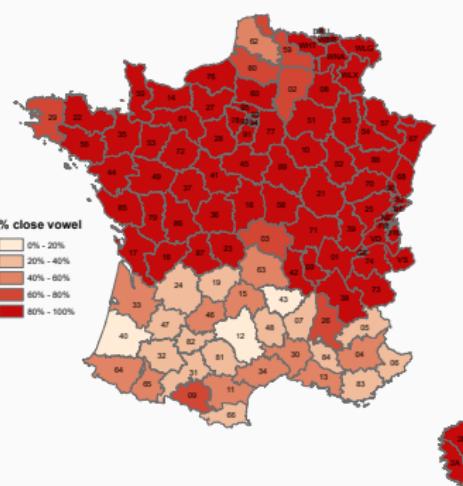
Glättung: Gebieten mit < 5 Teilnehmern wird der Durchschnittswert aller Nachbargebiete zugewiesen

Cartophono: Darstellung der Resultate

Rose
Departementsweise
Aggregierung

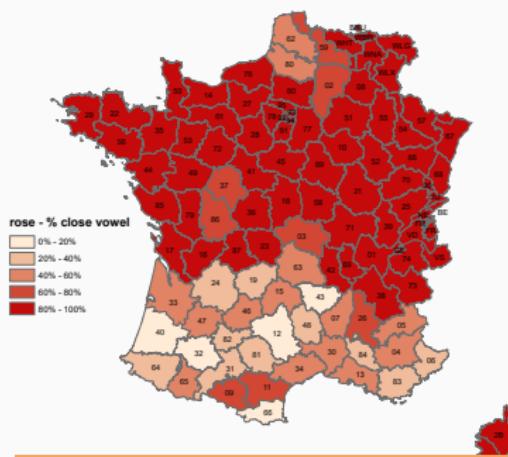


17 Stimuli mit O
Departementsweise
Aggregierung

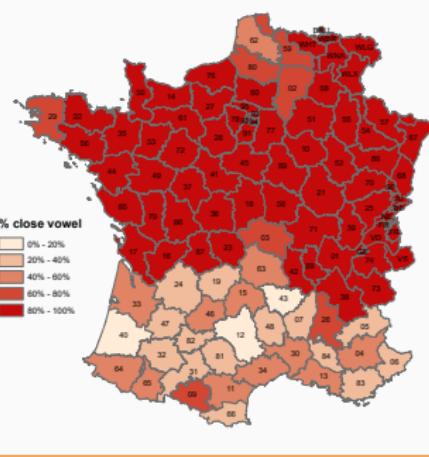


Cartophono: Darstellung der Resultate

Rose
Departementsweise
Aggregierung



17 Stimuli mit O
Departementsweise
Aggregierung



Durchschnittswerte mehrerer Stimuli ergeben
verlässlicheres Bild

Français de nos régions: Projekt und Teilnehmer

- Mehrere Online-Umfragen zur regionalen Variation im Französischen:

Name	Beginn	Gebiet	Phänomene	Fragen	Teilnehmer
Lex1	Mai 2015	F+B+CH	Lexikon/Syntax	40	10'000
Lex2	Sept. 2015	F+B+CH	Lexikon/Syntax	70	5000
Vok	Sept. 2015	F+B+CH	Vokale	45	1500
...					

- Frage (Lexik): Wie nennen Sie das Objekt auf dem Bild?
 - Cf. *Atlas der deutschen Alltagssprache*

Avanzi, M., Barbet, C., Glikman, J., Peuvergne, J. (2016): Présentation d'une enquête pour l'étude les régionalismes du français. *Actes du 5ème congrès mondial de linguistique française (CMLF)*, Tours.

Avanzi, M. (2017): *Atlas du français de nos régions*. Armand Colin.

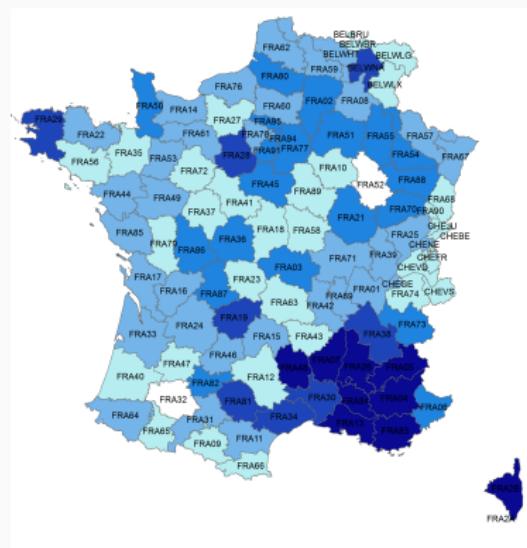
<https://francaisdenosregions.com/>

Français de nos régions: Darstellung binärer Variablen

Huitante ,80'



*Chaler ,jemanden auf dem
Gepäckträger des Fahrrads
mitnehmen'*

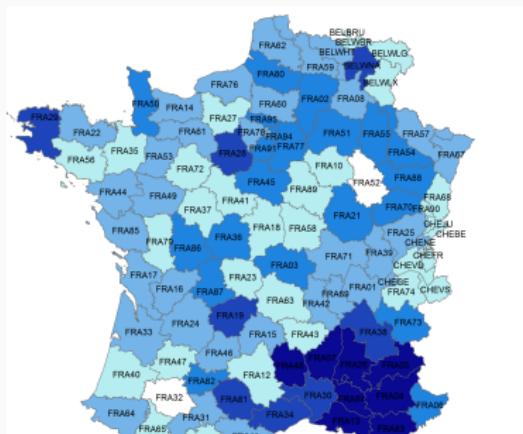


Français de nos régions: Darstellung binärer Variablen

Huitante ,80'



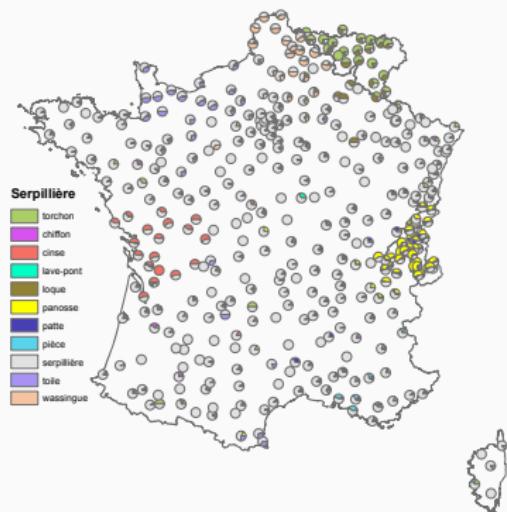
Chaler ,jemanden auf dem Gepäckträger des Fahrrads mitnehmen‘



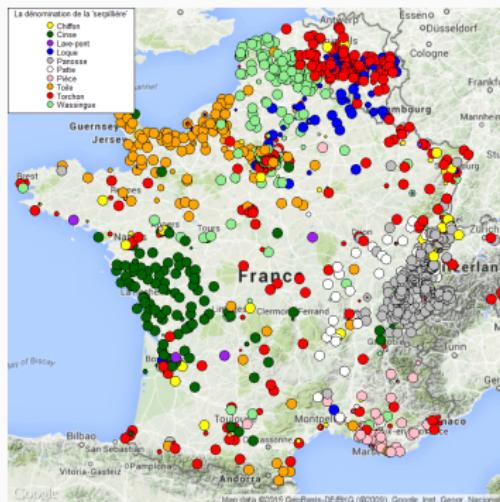
Binäre Variablen können als gewichtete Polygonkarten dargestellt werden. Aggregierung auf Departementebene

Français de nos régions: Darstellung mehrstufiger Variablen

Serpillière ,Putzlappen'
Tortendiagramme
Aggr.: Arrondissements
Standardvariante in grau



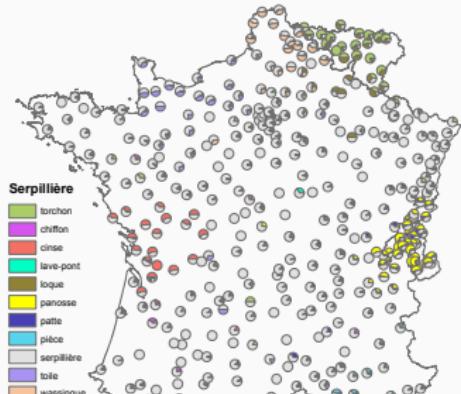
Serpillière ,Putzlappen'
Punktkarte
Aggr.: Postleitzahl
Ohne Standardvariante



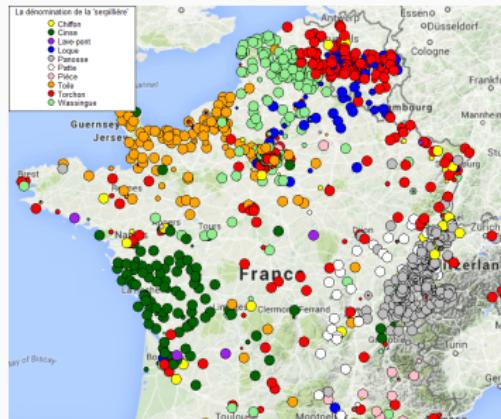
Karte: Mathieu Avanzi

Français de nos régions: Darstellung mehrstufiger Variablen

Serpillière ,Putzlappen'
Tortendiagramme
Aggr.: Arrondissements
Standardvariante in grau

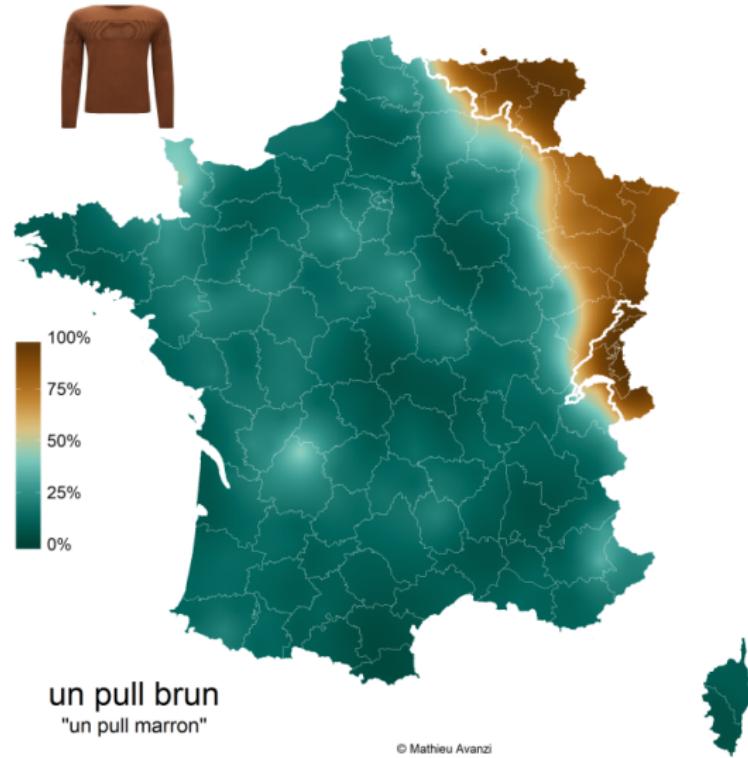


Serpillière ,Putzlappen'
Punktkarte
Aggr.: Postleitzahl
Ohne Standardvariante



Wie wichtig ist die Standardvariante?
Wie wichtig sind die Proportionen?

Français de nos régions: Rasterkarten



Dialektometrie als Datenvalidierungsmethode

Clusteranalyse
Umfrage *Lexikon 1*
Ward's Methode, 10 Cluster



Clusteranalyse
Umfrage *Vokale*
Ward's Methode, 10 Cluster



Dialektometrie als Datenvalidierungsmethode

Clusteranalyse

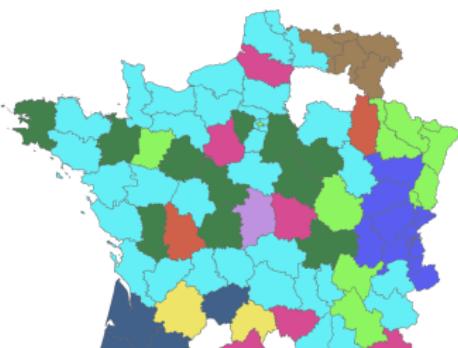
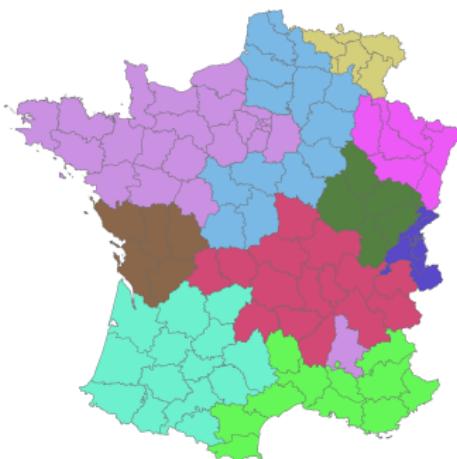
Umfrage *Lexikon 1*

Ward's Methode, 10 Cluster

Clusteranalyse

Umfrage *Vokale*

Ward's Methode, 10 Cluster



- Zu wenig Teilnehmer?
- Zu ähnliche Fragen?
- Schlecht verständliche Fragen?

Nicht-dialektologische Datensätze

Nicht-dialektologische Datensätze: ArchiMob-Korpus

Das Archimob-Projekt:

- Ein von Historikern und Filmschaffenden gegründeter Verein mit dem Ziel, die neuere Schweizer Geschichte zu erschliessen (<http://www.archimob.ch>)
- 555 auf Video aufgezeichnete Interviews mit Zeitzeugen des Zweiten Weltkriegs in der Schweiz

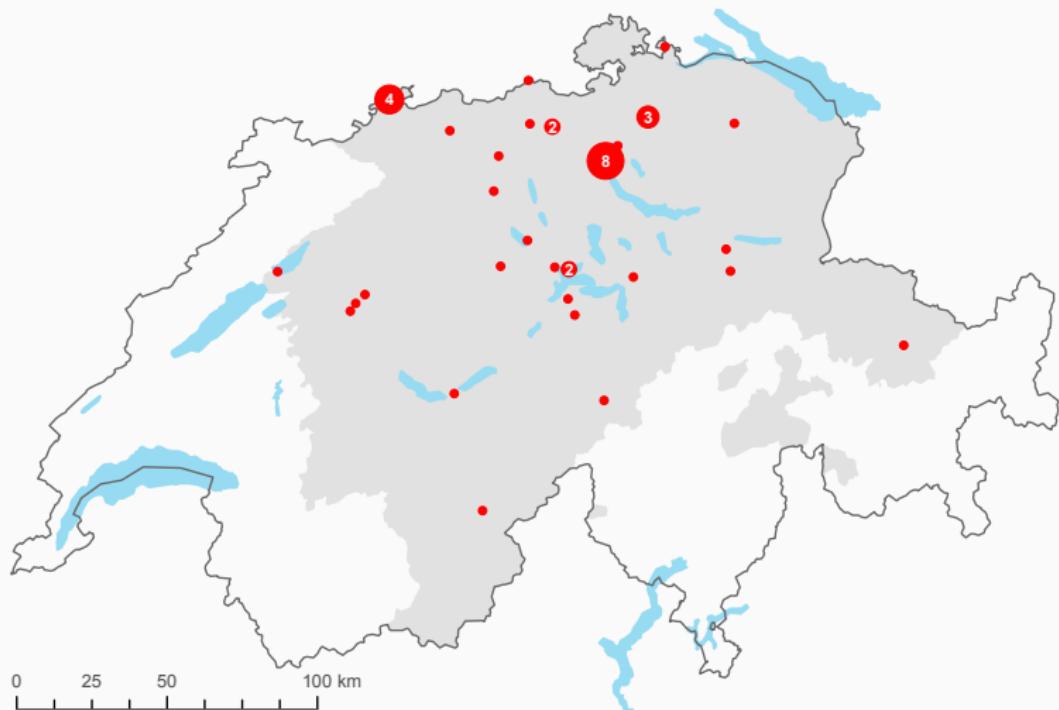
Das ArchiMob-Korpus:

- 43 ausgewählte Interviews in schweizerdeutschen Dialekten
- Transkribiert, normalisiert

Tanja Samardžić, Yves Scherrer & Elvira Glaser (2016):

ArchiMob – A Corpus of Spoken Swiss German. In *Proceedings of LREC 2016*.

Nicht-dialektologische Datensätze: ArchiMob-Korpus



Nicht-dialektologische Datensätze: ArchiMob-Korpus

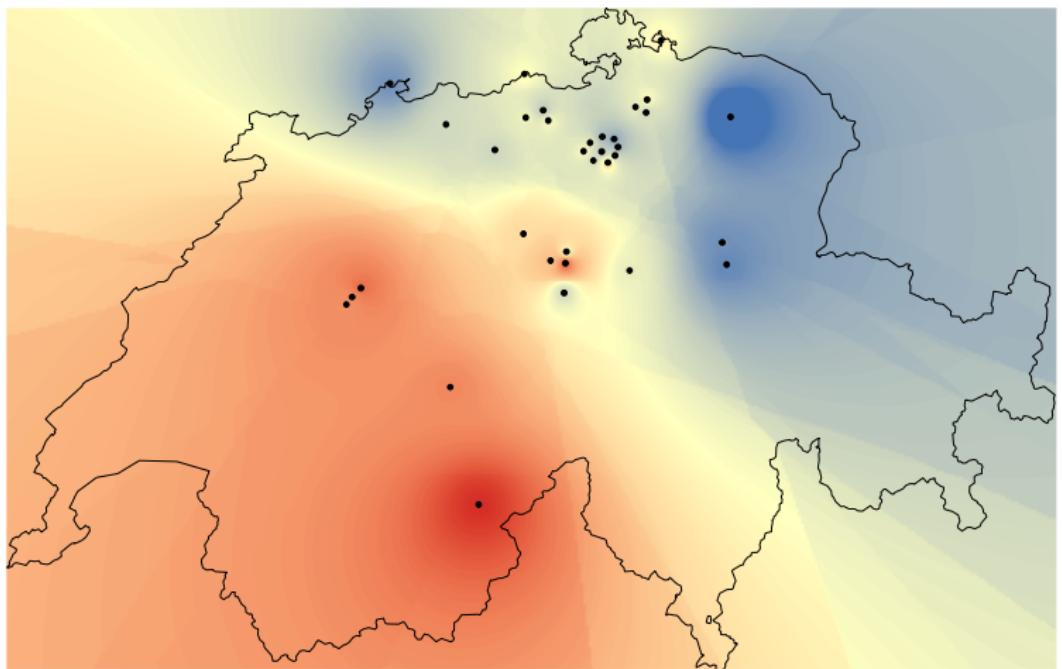


Herausforderungen für die Kartierung:

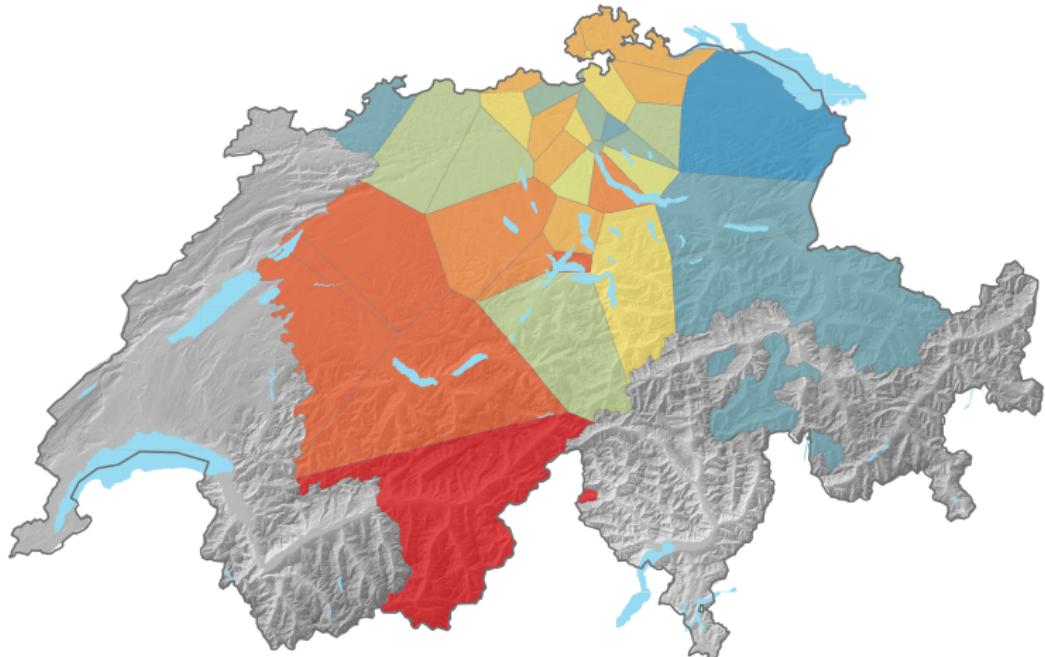
- Unregelmässige geografische Verteilung der Interviews
- Grosse Lücken zwischen den Datenpunkten

→ Polygon- und Rasterkarten machen keinen Sinn

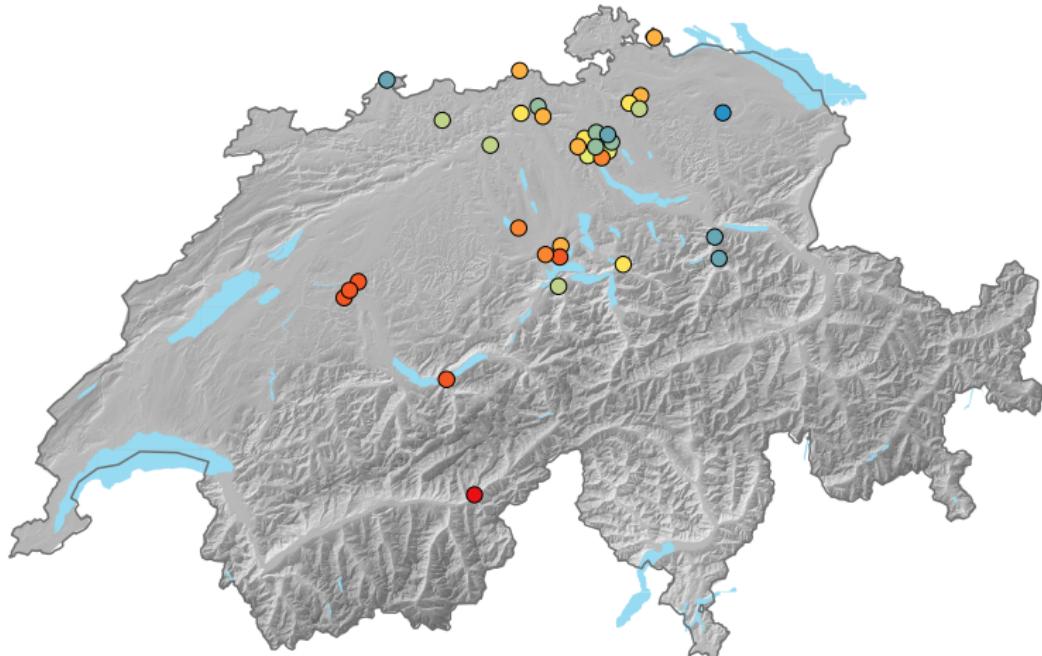
ArchiMob: Heat maps?



ArchiMob: Heat maps?

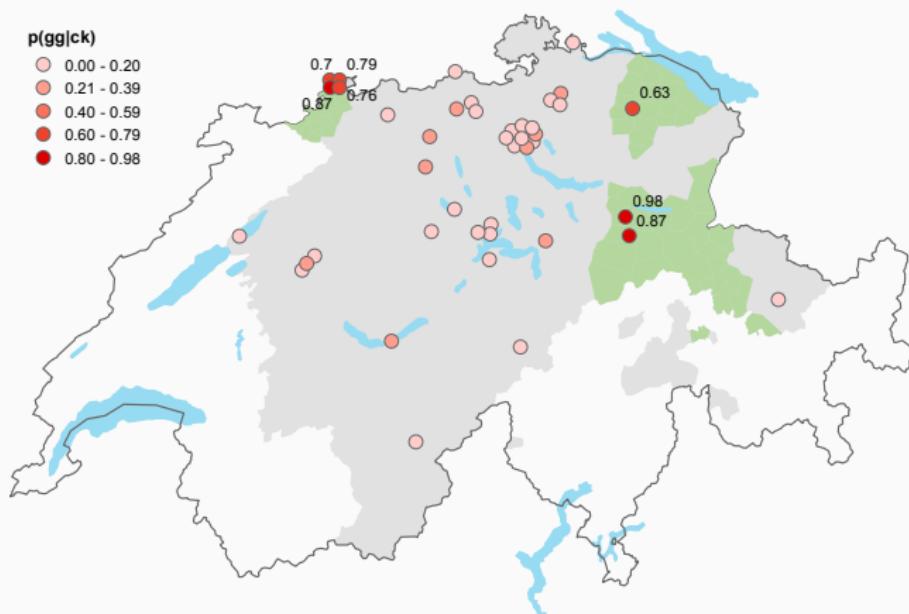


ArchiMob: Heat maps?



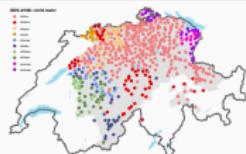
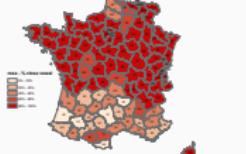
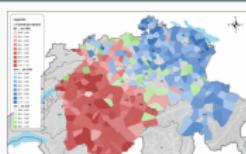
ArchiMob: Vergleichende Darstellung

- Rote Punkte: Proportionen von gg-Aussprache in ArchiMob-Interviews
- Grüne Flächen: Attestierte gg-Aussprache im SDS



Schluss

Zusammenfassung

Geogr. Objekte → Attribute ↓	Punkte	Polygone (Thies-sen / politisch)	Raster
Nominalskala			
Verhältnisskala, 1 Variable			
Verhältnisskala, Komplementäre Variablen			
Geglättete Daten	möglich	möglich	ja
Dichtes Netz	möglich	ja	ja

Zusammenfassung

- Kartografie ist eher eine Kunst als eine Wissenschaft
- Es gibt kaum strikte Regeln, die zu befolgen sind
- Die zu vermittelnde Botschaft bestimmt die Darstellung
- Die gezeigten Beispiele sind als Inspiration gedacht...

Hands-on session

Hands-on session

Ziele:

- Woher bekomme ich geografische Objekte?
- Wie verbinde ich Attribute mit geografischen Objekten?
- Woher bekomme ich Hintergrunddaten?

<https://www.github.com/yvesscherrer/geolinguistik>

Aufgaben:

- Exercise 0: benötigte R-Pakete installieren
- Exercise 1: Datensätze laden und darstellen
- Exercise 2: Geodaten und Attribute verbinden

Hands-on session

Ziele:

- Woher bekomme ich geografische Objekte?
- Wie verbinde ich Attribute mit geografischen Objekten?
- Woher bekomme ich Hintergrunddaten?

<https://www.github.com/yvesscherrer/geolinguistik>

Aufgaben:

- Exercise 0: benötigte R-Pakete installieren
- Exercise 1: Datensätze laden und darstellen
- Exercise 2: Geodaten und Attribute verbinden