

# gesis

Leibniz-Institut  
für Sozialwissenschaften



## Regionale Kontextdaten aus dem Mikrozensus

Klaus Pforr

Mannheim, 28.11.2018

# Warum räumliche Kontextdaten?

- Kontexteffekte: wesentlicher Aspekt vieler soziologischer Fragestellungen (Makro  $\Leftrightarrow$  Mikro)  
*(z.B. KZfSS Sonderheft Friedrichs/Nonnenmacher (2014))*
- Räumliche Kontexte *(vgl. Horr (2016))*
  - ▶ Eigene Teildisziplin Stadtsoziologie
  - ▶ Räumliche Nahumgebung
    - Exogene Ressourcen/Rahmenbedingungen
    - Selbstselektion in Nachbarschaft
- Beispiele: Lokale Partnermärkte, Economic Voting

# Notwendige Datenqualität

- Messung der Kontextmerkmale am Individuum
  - ▶ Subjektiv mit den anderen Individualmerkmalen vs. Objektive Daten zugespielt
- Längsschnittdaten zur Kontrolle von Selektion
- Möglichst präzise Messung der Kontexteffekte zur Reduktion von attenuation bias

## Daten mit Anspielmöglichkeit

- Fast alle Scientific-Use-Files
  - ▶ Zuspielung auf Bundeslandebene (NUTS1)
- ALLBUS, pairfam, SOEP, NEPS
  - ▶ Zuspielung auf Gemeindeebene (und tiefer)
  - ▶ Fast immer vor Ort, mittlerweile auch per remote access
- Bei neuen Stichproben Zuspielung auf Adressebene möglich

## Kontextdaten-Quellen

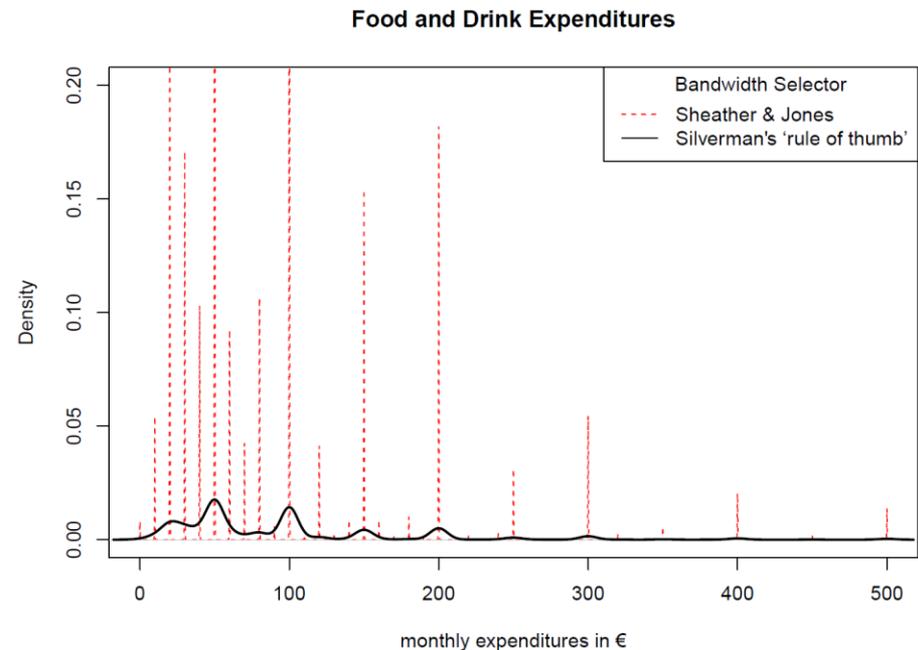
- Umfragedaten → Inhalte interessant, aber N zu klein
  - MICROM → Erstellung intransparent
  - Amtliche Aggregatdaten → Ebenen zu hoch oder Inhalte beschränkt
  - Amtliche Mikrodaten → Inhalte interessant, aber ohne weiteres Ebenen auch zu hoch
- ➔ Schätzung von kleinräumigeren Verteilungen ausgehend von amtlichen Mikrodaten

# Schätzung von räumlichen Kontexten

- Daten liegen in regionalen Einheiten zusammengefasst vor
- Schätzung der darunterliegenden, realen, kontinuierlichen, räumlichen Verteilung mit Kerndichteschätzer
  - ▶ Theoretisch punktgenaue räumliche Information
  - ▶ Schätzgenauigkeit abhängig von gegebenen Informationen und Schätzverfahren
  - ▶ Kernelheaping (Groß & Rendtel 2016)

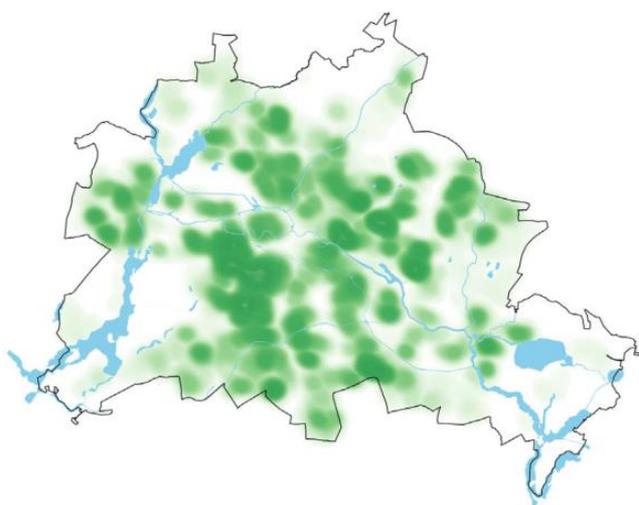
# Kernelheaping (1)

- Übertragung eines Verfahrens zur Schätzung der kontinuierlichen Variablen aus klassierten Variablen (Groß & Rendtel 2016)
  - ▶ Modellierung als Messfehlerproblem
  - ▶ Bandweite wird im Modell durch Plug-In-Schätzer (Wand & Jones 1994) und SEM-Algorithmus

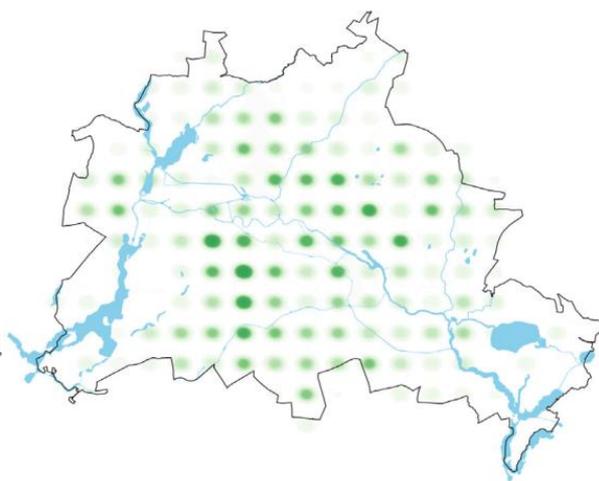


## Kernelheaping (2)

- Anwendung auf räumlich zusammengefasste Variablen
  - ▶ Aus Groß et al. 2016; Daten der amtl. Statistik Berlin-Brandenburg; Anteil Alter 60+



Original & nicht  
öffentlich



Zusammengefasst auf  
2,5 km x 2,5 km &  
veröffentlichbar



Geschätzt

# Anwendung auf Mikrozensus (1)

- Prototypische Anwendung auf **MZ Regionalfile 2000**
- MZ 2000 leicht vergrößert auf MZKR-Ebene

Amtliche Kreise



MZKR



Kreise in der Umgebung



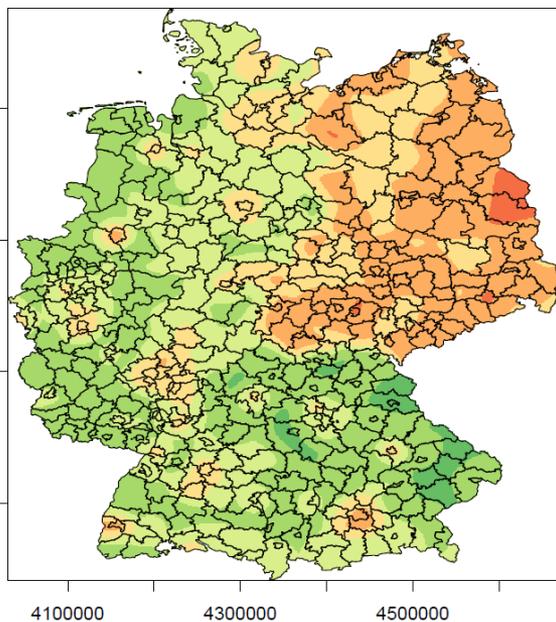
## Anwendung auf Mikrozensus (2)

- Schätzung von räumlichen Verteilung von Anteilen von beliebigen kategorialen Variablen
- Genauigkeit
  - ▶ 500x500-Zellenraster (1285 m x 1738 m)
- Variablen
  - ▶ Höchster Schulabschluss
  - ▶ Nationalität
  - ▶ Netto-HH-Einkommensquartile
  - ▶ Anteil gleichgeschlechtlicher Kohabitierender
  - ▶ Bildungsmobilität

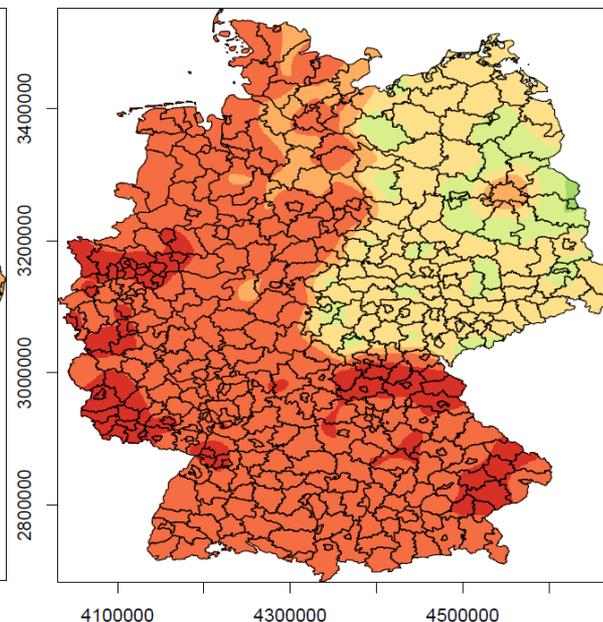
# Anwendung auf Mikrozensus (3)

- Höchster Schulabschluss

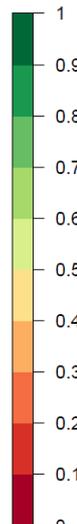
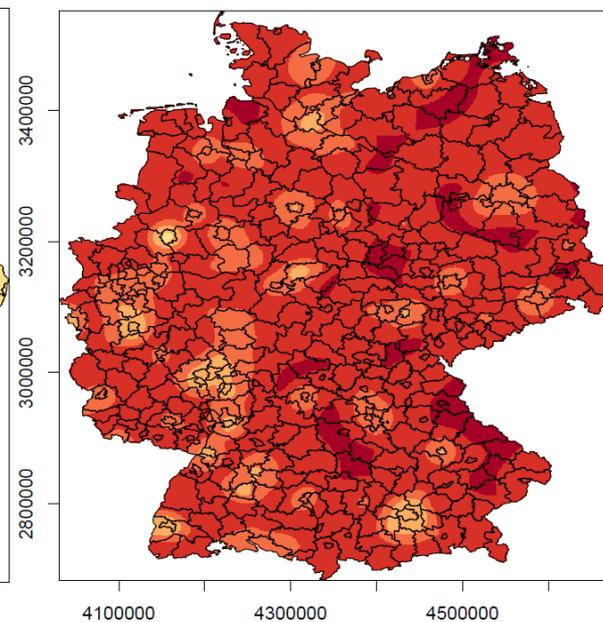
Anteil kein Abschluss/  
Hauptschule



Anteil Realschule/POS



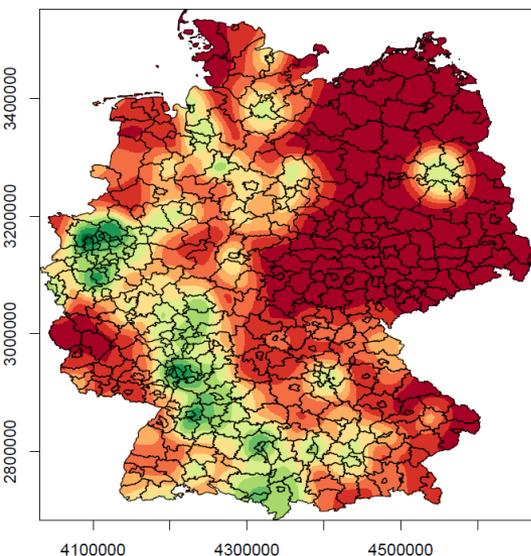
Anteil Abi/Fachabi



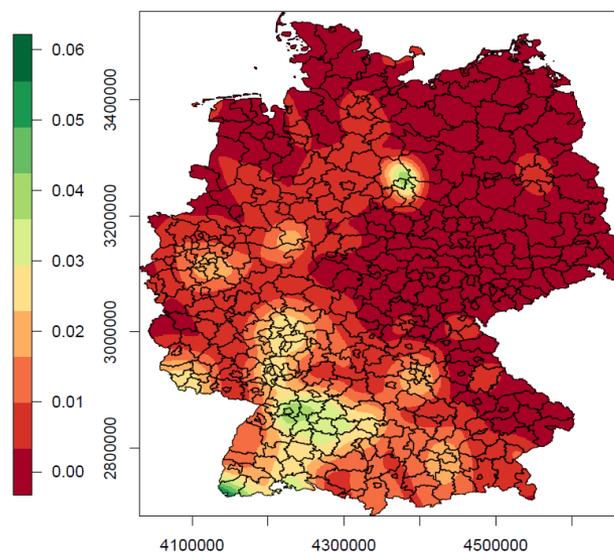
# Anwendung auf Mikrozensus (4)

- Nationalität

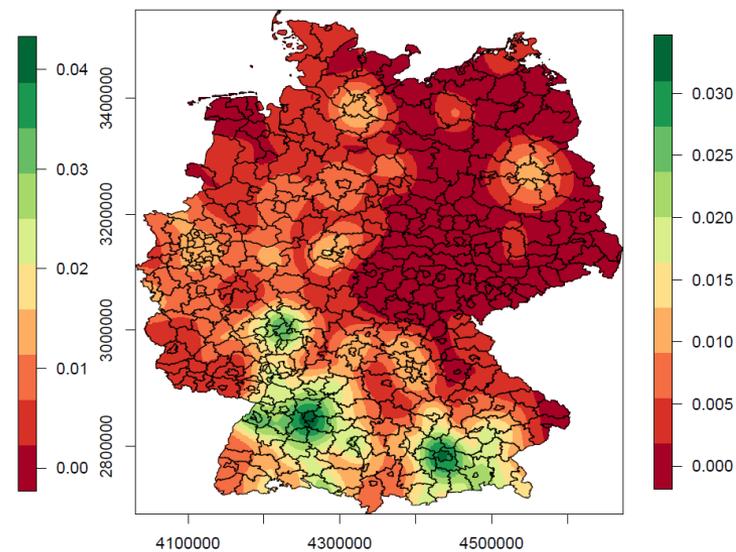
## Anteil Türken



## Anteil Italiener



## Anteil Ex-Jugos



# Anwendung auf Mikrozensus (5)

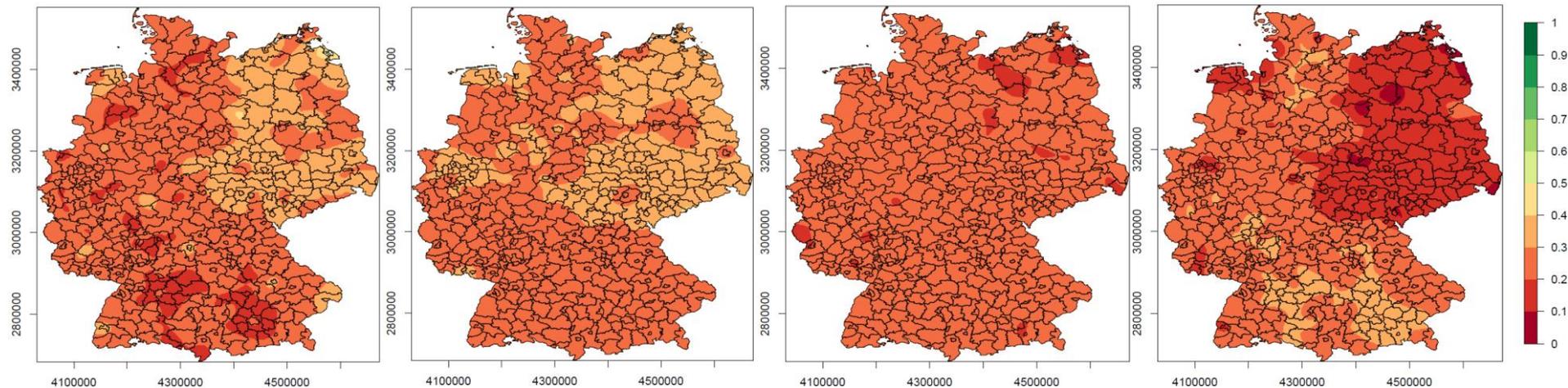
- HH-Netto-Einkommensquartile

Anteil HH-NettoEink 1.Quartil

Anteil HH-NettoEink 2.Quartil

Anteil HH-NettoEink 3.Quartil

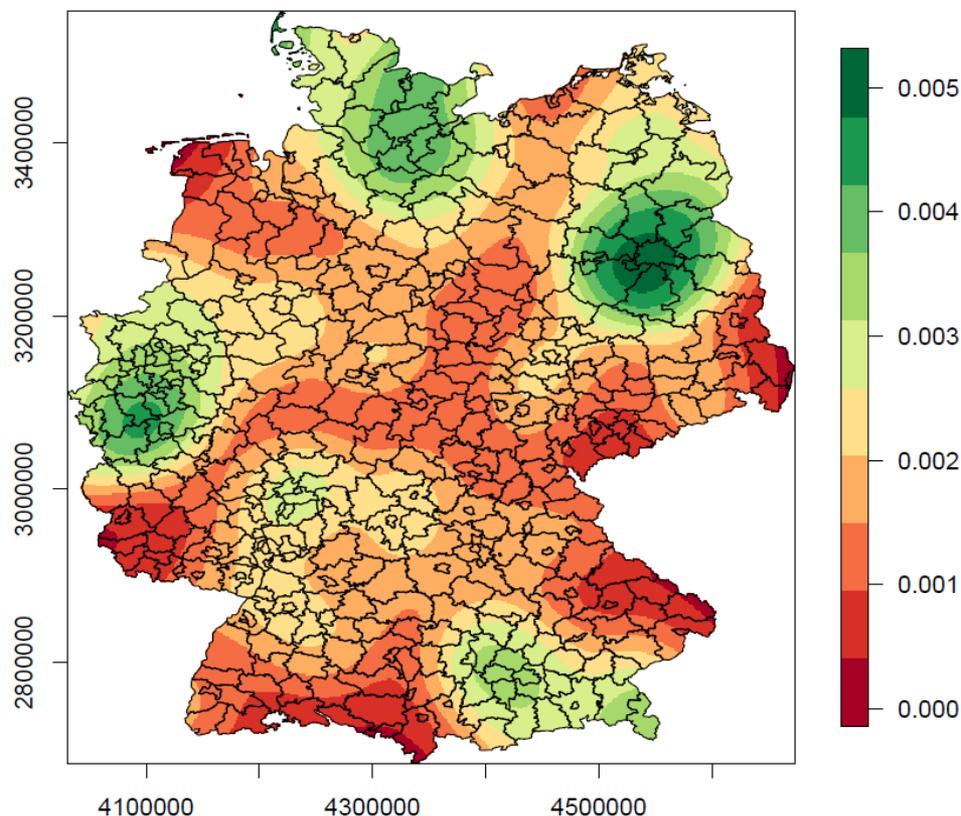
Anteil HH-NettoEink 4.Quartil



## Anwendung auf Mikrozensus (7)

- Anteil Gleichgeschlechtliche von allen Kohabitierenden Paaren

Anteil Gleichgeschlechtliche Kohabs

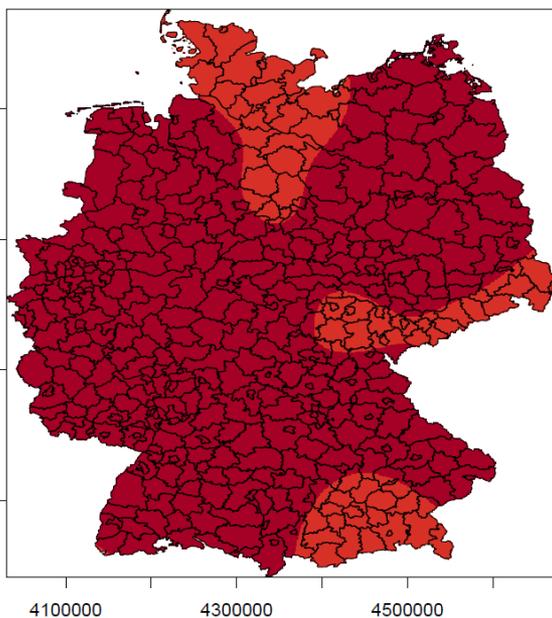


# Anwendung auf Mikrozensus (8)

## ■ Bildungsmobilität

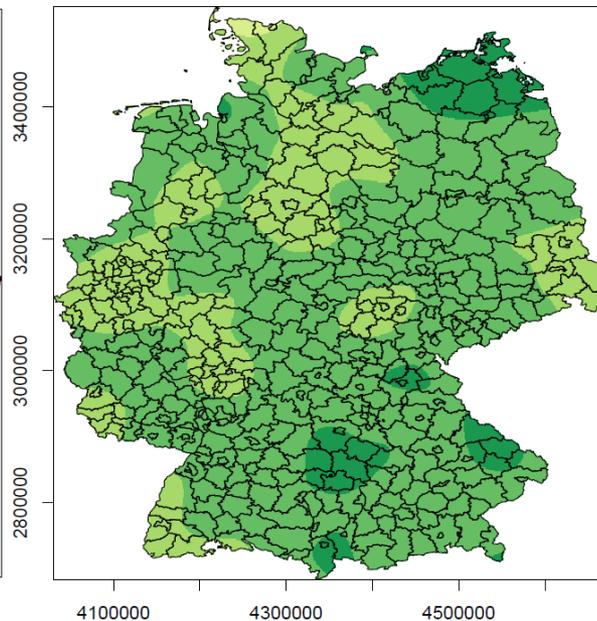
### Absteiger

Anteil Kinder nicht in gymn.  
Oberstufe von Eltern mit  
Abitur



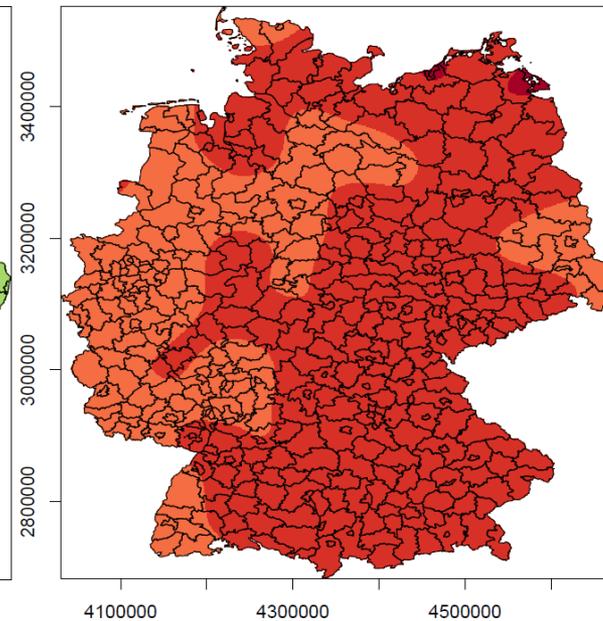
### Reproduktion

Anteil Kinder in gymn.  
Oberstufe von Eltern mit  
Abitur oder beide nicht



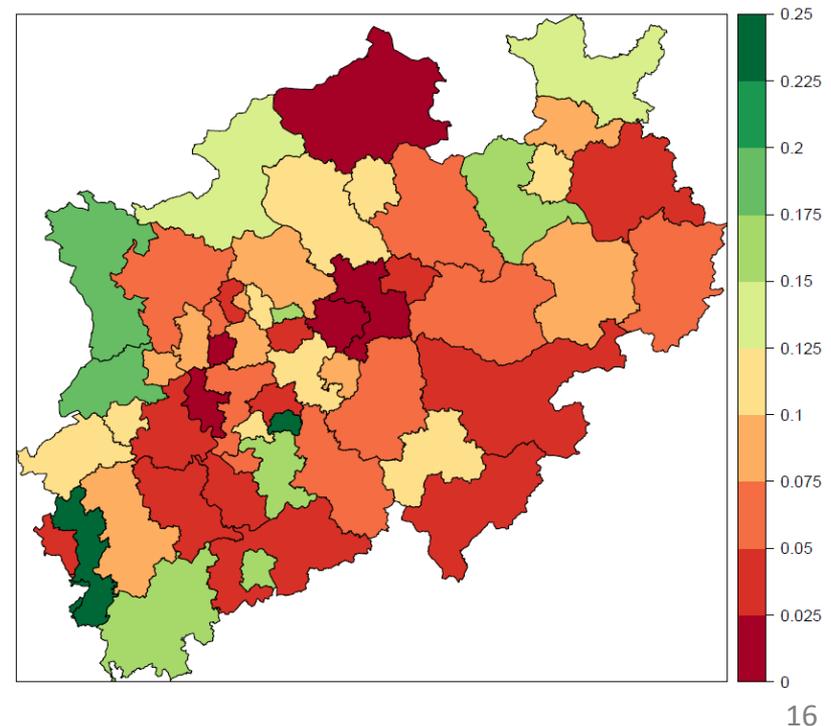
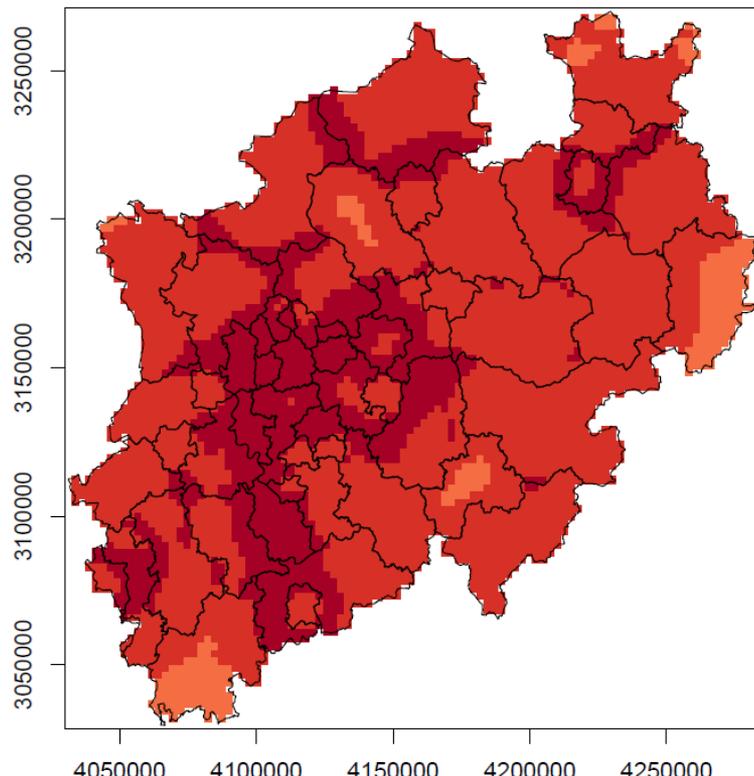
### Aufsteiger

Anteil Kinder in gymn.  
Oberstufe von Eltern ohne  
Abitur



# Anwendung auf Mikrozensus (9)

- Stichprobenvarianz?
  - ▶ 50 Bootstrap-Samples der Geschlechtsverteilung in NRW
  - ▶ 100x100 Zellen, 5 burn-in, 15 estimation samples, 64 Chains
  - ▶ Differenz zwischen 95. Perzentil und 5. Perzentil



## Fazit

- Verfahren bildet großflächige räumliche Verteilung ab
- Varianz der räumliche Punktschätzer zufriedenstellend
- Schätzung vermutlich abhängig von gegebenen räumlichen Einheiten

## Nächste Schritte

- Anwendung auf Mikrozensus Onsite
  - ▶ Ausgangsniveau Gemeindeebene
  - ▶ Small-Area-Schätzung für fehlende Gemeinden
  - ▶ Verbesserter Bootstrap
- Inhaltliche Anwendungen
- Verbesserung der Schätzung
  - ▶ Räumliche Info über Baubestand
  - ▶ Einbeziehung des Messzeitpunkts
  - ▶ Berücksichtigung von multivariater Korrelation

Vielen Dank für Ihre Aufmerksamkeit!

gesis

Leibniz-Institut  
für Sozialwissenschaften

Mitglied der  
*Leibniz*  
Leibniz-Gemeinschaft

# SEM-Algorithmus

1. Erste Schätzung der räumlichen Verteilung mit großer Bandweite
2. Berechne geschätzte Anteile auf Knotenpunkten des feinen Zielgrids
3. Ziehe Pseudo-Stichprobe aus geschätzter Dichteverteilung und geschätzten Anteilen auf Zielgrid
4. Schätze neue Bandweite nach Wand und Jones (1994) und schätze neue räumliche Verteilung
5. Wiederhole (2)-(4) mit B burn-in-Iterationen und N Schätziterationen
6. Geschätzte Anteile auf Grid-Knotenpunkten aus Mittelwert über N Schätziterationen

# Literatur

- Friedrichs, J., & Nonnenmacher, A. (2014). Die Analyse sozialer Kontexte. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(1), 1-16.
- Groß, M., & Rendtel, U. (2016). Kernel density estimation for heaped data. *Journal of Survey Statistics and Methodology*, 4(3), 339-361.
- Groß, M., Rendtel, U., Schmid, T., Schmon, S., & Tzavidis, N. (2017). Estimating the density of ethnic minorities and aged people in Berlin: multivariate kernel density estimation applied to sensitive georeferenced administrative data protected via measurement error. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 161-183.
- Horr, A. (2016). Nachbarschaftseffekte. In *Ethnische Ungleichheiten im Bildungsverlauf* (pp. 397-430). Springer VS, Wiesbaden.
- Wand, M. P., & Jones, M. C. (1994). Multivariate plug-in bandwidth selection. *Computational Statistics*, 9(2), 97-116.