

# Kleinräumige extrapolation von Umfragedaten

Kai Husmann, Alexander Lange

Georg-August-Universität Göttingen

15. Juli 2016

# Inhaltsverzeichnis

## Einleitung

## Deskriptive Statistik

Datensatz

Räumliche Effekte

## Methodik

Model

Modellwahl

Modellvalidierung

## Ergbnisse

GAM Schätzung

Prediction

# Einleitung

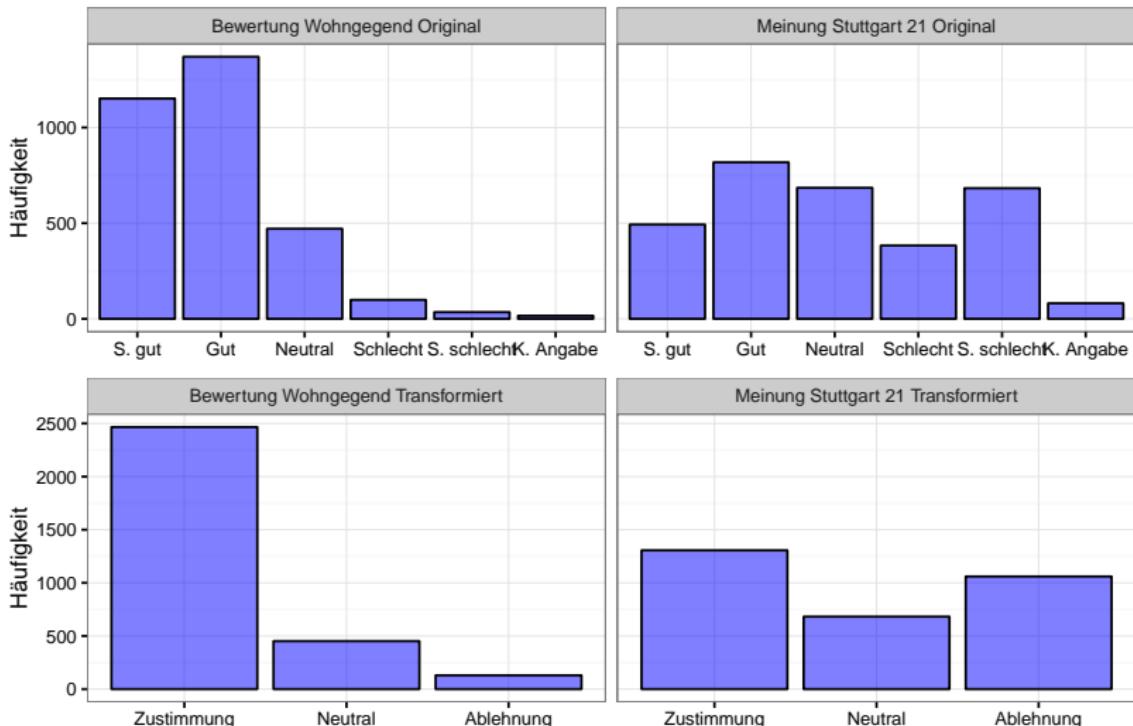
- Daten zur Bevölkerungsstruktur Stuttgarts
- Nur Informationen aus der Stichprobe enthalten, keine Grundgesamtheit aus dieser Erhebung verfügbar
- Hochrechnung der Informationen mithilfe der Grundgesamtheiten aus anderen Erhebungen (Bürgerumfrage und Zensus)
- Brücksichtigung von unterschiedlichen räumlichen Effekten
- Ziel: Extrapolation der Informationen aus der Stichprobe und Möglichkeit der Validierung finden

# Datensatz

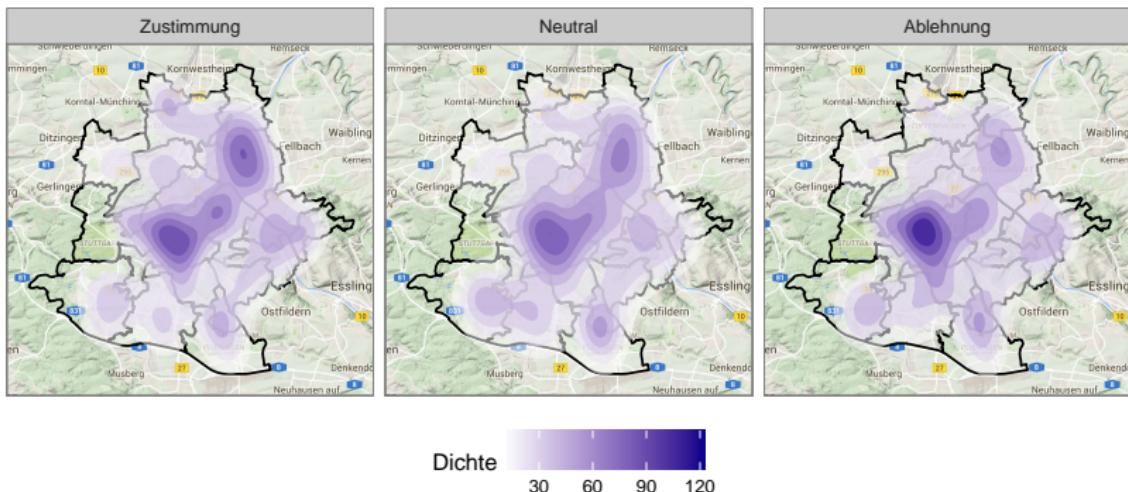
Anzahl Beobachtungen Stichprobe: 3.143

Variable	Anzahl Klassen	Modellierung
Bewertung Wohngegend	6	Geordnet Kategorial
Meinung Stuttgart 21	6	Geordnet Kategorial
Personenanzahl im Haushalt	5	Nicht Parametrisch
Monatliches Netto Haushaltseinkommen	6	Nicht Parametrisch
Altersklasse Befragter	6	Nicht Parametrisch
Geschlecht	2	Parametrisch
Familienstand	4	Parametrisch
Nationalität	2	Parametrisch
Stadtbezirk	23	Diskret Räumlich
Stadtteil	142	Diskret Räumlich
Gauß-Krüger		Stetig Räumlich

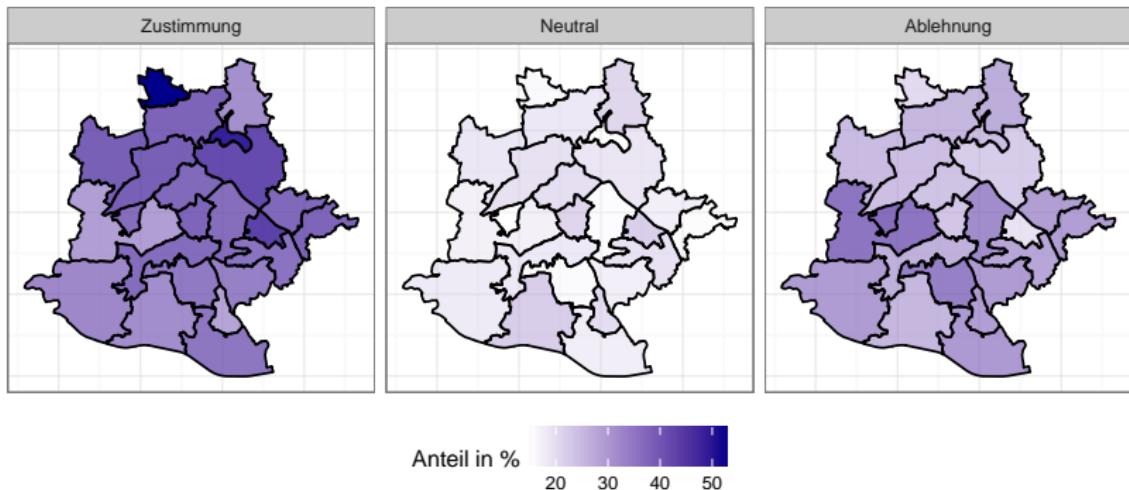
# Response Variablen



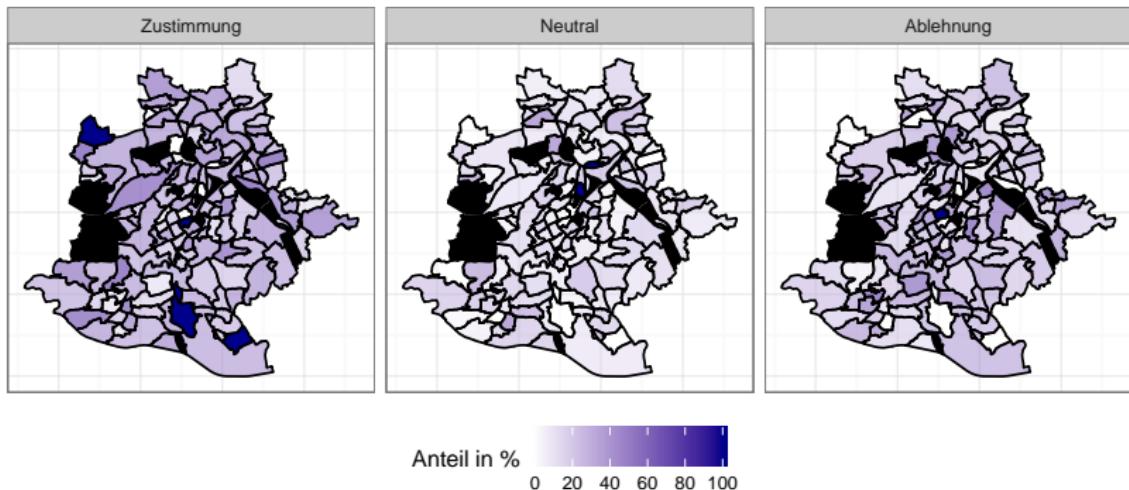
# Gauss-Krüger Informationen



# Diskrete Informationen zu Stadtbezirken



# Diskrete Informationen zu Stadtteilen



# Modellierung

## Generalisierte additive Modelle

- Einbindung von parametrischen Effekten durch linearen Schätzer
- Einbindung von nicht-parametrischen Effekten über P-Splines
- Stetige räumliche Gauss-Krüger Effekte mit Tensorprodukt-Spline
- Diskrete räumliche Effekte durch Gauss-Markov Zufallsfeld

$$P(y_i = r) = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + f(\mathbf{z}) + \text{räumlich} = \begin{cases} f(s_x, s_y) \\ f(s_{\text{Stadtteil}}) \\ f(s_{\text{Stadtbezirk}}) \end{cases} + \varepsilon$$

Insgesamt 6 Modelle zu schätzen

# Modellwahl

## AIC-Basierte schrittweise Modellwahl

- Vorab Einschätzen, ob Variablen parametrisch oder nichtparametrisch eingehen sollen
- Potenziell nichtparametrische Variablen werden auch parametrisch getestet
- Räumliche Variablen gehen als feste Modellbestandteile ein
- Sinnvoll erscheinende Wechselwirkungen werden händisch untersucht

# Modellvalidierung

## Kreuzvalidierung

- Es liegt keine Grundgesamtheit vor
- Das beste Modell wird mit einer Zufallsauswahl neu parameterisiert
- Alle Datensätze werden vorhergesagt
- Interpretation der Confusion Matrix

# GAM Zusammenfassung: Stetiger räumlicher Effekt

<i>Abhängige Variable:</i>	
Meinung zu Stuttgart 21	
Geschlecht - Weiblich	0.524*** (0.070)
Nationalität - Nicht Deutsch	-0.445*** (0.110)
Familienstand - ledig	0.090 (0.159)
Familienstand - verheiratet	-0.176 (0.151)
Familienstand - verwitwet	-0.356* (0.207)
Personenzahl im Haushalt	-0.340*** (0.060)
 <i>s(X, Y)</i>	*
<i>s</i> (Personenzahl im Haushalt, Altersklasse)	***
<i>s</i> (Altersklasse)	**
 Beobachtungen	3,062
Log Likelihood	-3,186.494
AIC	6379

*Legende:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# GAM Zusammenfassung: Diskreter räumlicher Effekt

Tabelle:

<i>Abhängige Variable:</i>	
Meinung zu Stuttgart 21	
Geschlecht - Weiblich	0.526*** (0.070)
Nationalität - Nicht Deutsch	-0.448*** (0.110)
Familienstand - ledig	0.100 (0.159)
Familienstand - verheiratet	-0.165 (0.151)
Familienstand - verwitwet	-0.344* (0.208)
Personenzahl im Haushalt	-0.344*** (0.060)
s(Stadtbezirk)	*
s(Personenzahl im Haushalt, Altersklasse)	***
s(Altersklasse)	**
Beobachtungen	3,062
Log Likelihood	-3,185.948
AIC	6378

Legende:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

# Prediction

