

Nutzung von Geoinformationen in der Survey Statistik

Erste deutschsprachige Survey-Statistiker Tagung

Jan-Philipp Kolb

Zeltingen-Rachtig

09. April 2014



Gliederung

- 1 Eingrenzung
- 2 Anwendungsbeispiel
- 3 Nutzung von Umgebungs-
informationen
- 4 Fazit

- 1 Eingrenzung
- 2 Anwendungsbeispiel
- 3 Nutzung von Umgebungsinformationen
- 4 Fazit

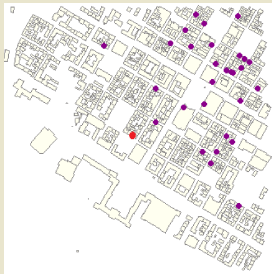
- R bietet zahlreiche Möglichkeiten geographische Informationen zu gewinnen und zu visualisieren
- Wie können diese Möglichkeiten bestmöglich für die Survey-Statistik verwendet werden?

Geoinformationen



- R bietet zahlreiche Möglichkeiten geographische Informationen zu gewinnen und zu visualisieren
- Wie können diese Möglichkeiten bestmöglich für die Survey-Statistik verwendet werden?

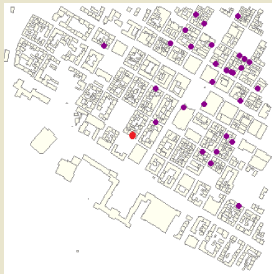
Geoinformationen



GESIS Mannheim

- R bietet zahlreiche Möglichkeiten geographische Informationen zu gewinnen und zu visualisieren
- Wie können diese Möglichkeiten bestmöglich für die Survey-Statistik verwendet werden?

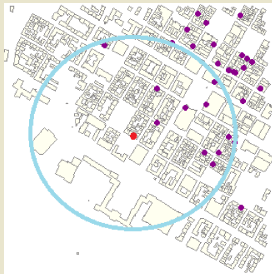
Geoinformationen



GESIS Mannheim
Fast Food Mannheim

- R bietet zahlreiche Möglichkeiten geographische Informationen zu gewinnen und zu visualisieren
- Wie können diese Möglichkeiten bestmöglich für die Survey-Statistik verwendet werden?

Geoinformationen

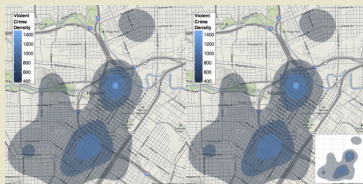


GESIS Mannheim
Fast Food Mannheim
Umkreis

R-Journal: Juni 2013



**Manuel J. A. Eugster and
Thomas Schlesinger:**
osmar - OpenStreetMap and R



**David Kahle and Hadley
Wickham:** ggmap - Spatial
Visualization with ggplot2

Georeferenzierung mit R-Paket ggmap

```
> geocode("Zeltingen")
Information from URL :
http://maps.googleapis.com/maps/api/geocode/json?
address=Zeltingen&sensor=false
Google Maps API Terms of Service :
http://developers.google.com/maps/terms
      lon      lat
1 7.013036 49.95489
```

Entfernungsmessung mit R-Paket ggmap

```
> mapdist("Trier", "Zeltingen")
Information from URL :
http://maps.googleapis.com/maps/api/distancematrix/json?
origins=Trier&destinations=Zeltingen&mode=driving&sensor=false
Google Maps API Terms of Service :
http://developers.google.com/maps/terms
      from      to      m      km      miles seconds minutes      hours
1 Trier zeltingen 48467 48.467 30.11739    2323 38.71667 0.6452778
```

- 1 Eingrenzung
- 2 Anwendungsbeispiel
- 3 Nutzung von Umgebungsinformationen
- 4 Fazit

Projekt: Entwicklung und Lebensumfeld von Vorschulkindern



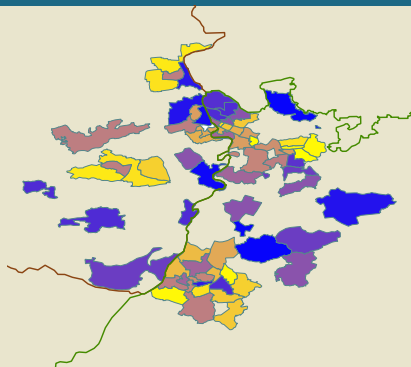
- Daten zu Kindergärten und Familien im Rhein-Neckar-Raum
- 2 Bundesländer, 32 Gemeinden, 73 PLZ-Bereiche

Projekt: Entwicklung und Lebensumfeld von Vorschulkindern

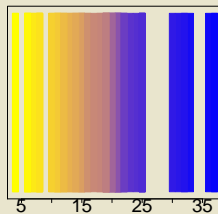


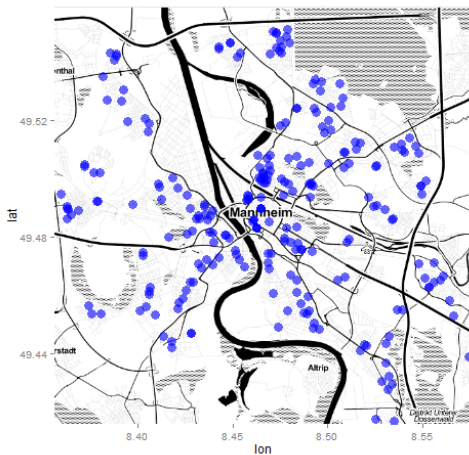
- Daten zu Kindergärten und Familien im Rhein-Neckar-Raum
- 2 Bundesländer, 32 Gemeinden, 73 PLZ-Bereiche

Projekt: Entwicklung und Lebensumfeld von Vorschulkindern

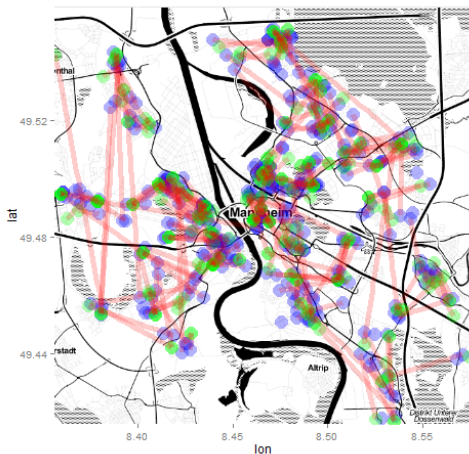


Farbschlüssel



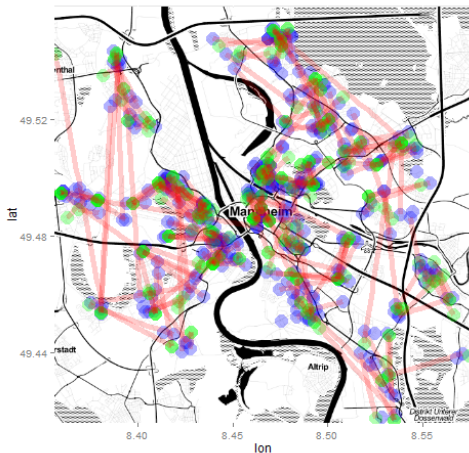


Geokodierter Wohnort



Geokodierter Wohnort

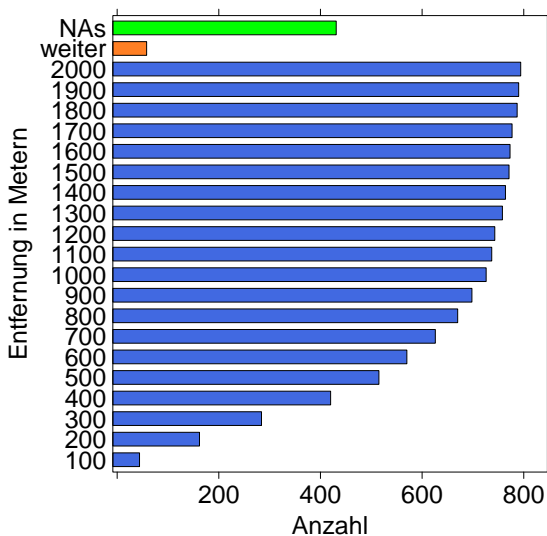
Geokodierter Kindergarten



Geokodierter Wohnort

Geokodierter Kindergarten

Entfernung



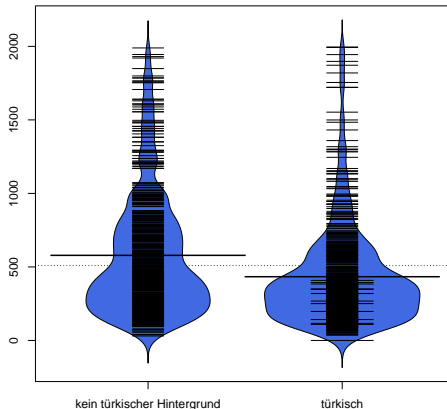
1. Schritt: lineare Regression

Abhängige Variable:

Entfernung -
Wohnort zum Kindergarten

Unabhängige Variablen

Merkmale der Familie
(z.B. Migrationshintergrund)



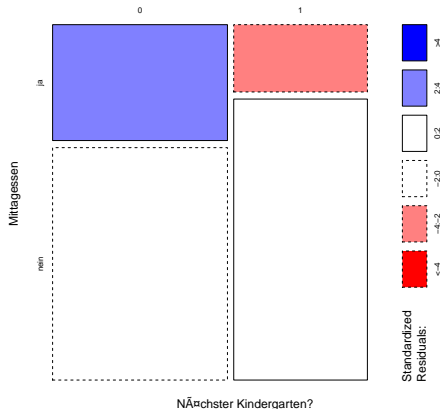
2. Ansatz: logistische Regression

Abhängige Variable:

Wurde der nächste
Kindergarten gewählt

Unabhängige Variablen

Merkmale der Familie
(z.B. Migrationshintergrund)



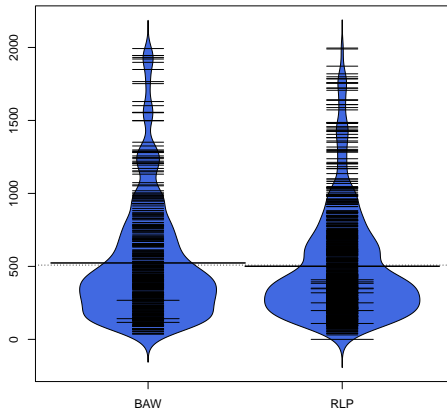
3. Ansatz: lineare Regression

Abhängige Variable:

Differenz
zwischen gewähltem und
nächsten Kindergarten

Unabhängige Variablen

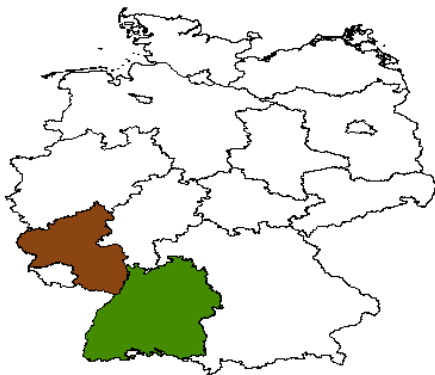
Bundesland

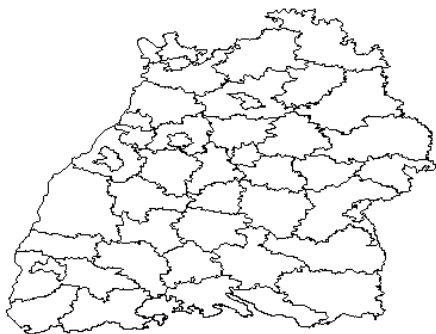


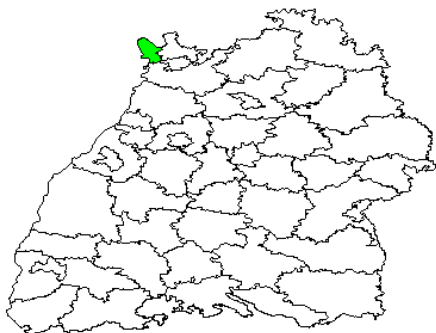
- 1 Eingrenzung
- 2 Anwendungsbeispiel
- 3 Nutzung von Umgebungsinformationen
- 4 Fazit



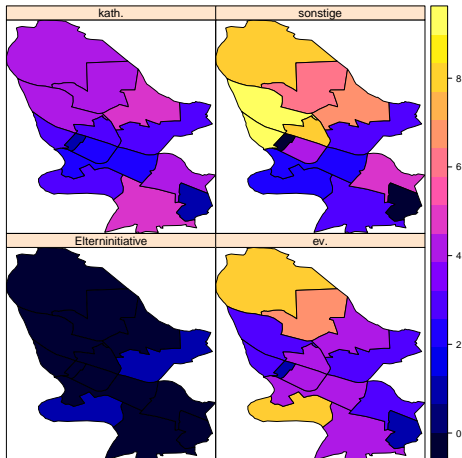








Träger des Kindergartens



Informationen aus dem Datensatz

- Wieviele Kindergärten befinden sich in der Umgebung?
- Welchen Träger haben die Kindergärten in der Umgebung?

Informationen aus dem Datensatz

- Wieviele Kindergärten befinden sich in der Umgebung?
- Welchen Träger haben die Kindergärten in der Umgebung?



OpenStreetMap

- Wie ist die ÖPNV-Struktur im PLZ-Bereich?
- Wie ist die Gebäudestruktur in der Umgebung?

PLZ-Bereich 68239



PLZ-Bereich 68161



Gebäudegröße



osmar object

Bounding box:

	lat	lon
min	49.4621	8.475615
max	49.4835	8.518027

Key-Value contingency table:

	Key	Value	Freq
1	railway	level_crossing	23
2	railway	crossing	10
3	highway	bus_stop	4
4	barrier	lift_gate	4
5	railway	tram_stop	4
6	addr:postcode	68165	3
7	barrier	bollard	3
8	addr:country	DE	3
9	addr:street	Fahrlachstraße	3
10	addr:suburb	Fahrlach	3

- Studie befindet sich in frühem Stadium - noch nicht alle Daten sind aufbereitet
- Eltern scheinen vor allem den nächsten Kindergarten zu wählen
- Gewünschte Informationen bekommt man, gewünschte Ergebnisse leider nicht immer ;-)

Die Kombination aus Open-Street-Map und R bietet faszinierende Möglichkeiten

- Non-Response Bias Analyse
- Random Route Verfahren
- Evtl. Nutzung von Umgebungsinformationen

Forschungsfelder

- Wie ist die Umgebung definiert?
- Wie wird mit Datenschutz umgegangen?
- Wie vollständig sind die Daten?

Vielen Dank für die Aufmerksamkeit

