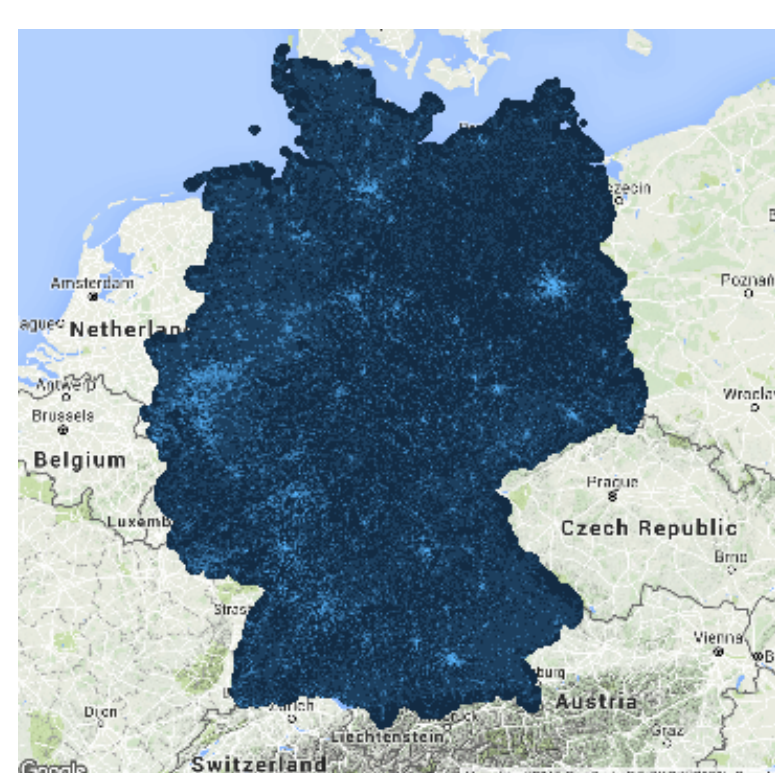# Usage of Linked Social Science and Spatial Data in Survey Statistics

Motivation/Intro

Attempts to link social with spatial data.
- New official data with spatial perspective (INSPIRE)
- Growing volume of collaborative data (OSM)
- Geocode own survey data (new BKG service)

  ⇒ Application examples in this poster

Data

zensus2011
- INSPIRE-compliant service:
- Download for Communities, 1 km^2 & 100 m^2 Raster

Methods/API's


Methods to access, transform, combine, analyse and visualize data e.g. R-package `raster`

Challenges

- Large volumes of data
- Different definitions of grid cells
- Combination of raster data with other data types
- Most adequate raster for analysis/visualisation?
- Is all necessary information available?

Survey Design

Spatial information can be used for planning survey/sampling designs
Many requests for e.g. telephone Surveys where spatial information plays a role
- Smart cities - use of transport services
- Survey on security perceptions
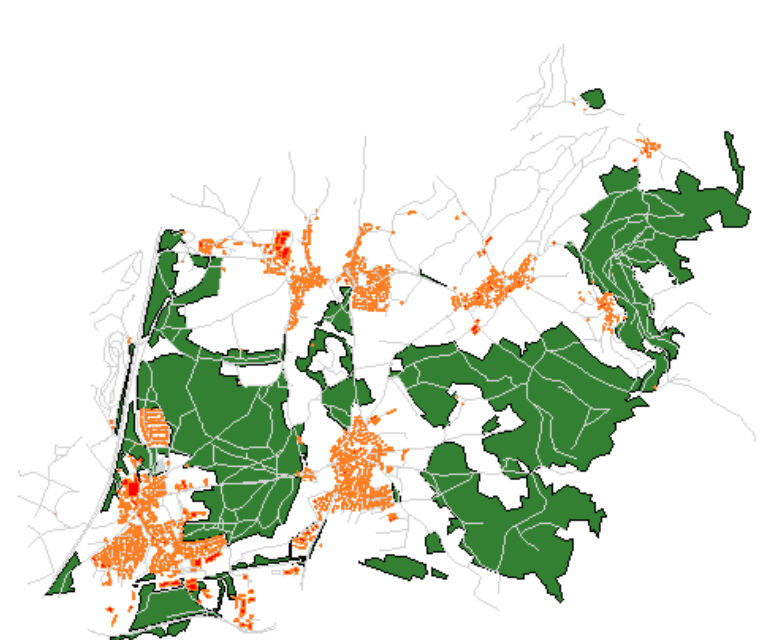
OpenStreetMap
- Collaborative project
- Free access to data
+ further web-data (e.g. street register)

Use of additional Spatial information:
- Stratification
- Allocation
- Simulation

R-packages to access OpenStreetMap data (`osmar`)

- Use of new methods where register based samples impossible
- Volunteered geographic information
  ⇒ heterogeneous data
  ⇒ variable quality
  ⇒ completeness?

Data Processing

- Spatial Imputation/Weighting
- Analysis of representativeness relies on auxiliary variables that are observed respondents and non-respondents.
- Imputation of e.g. interviewer observations in order to receive unbiased estimates.

Panel gesis
- Probability-based panel data drawn from registers
- CAPI Interviews
- Aprox. n=4900

Bundesamt für Kartographie und Geodäsie
- Geocoding as condition for the analysis of distances and spatial links

Use of distances between observations:
- Extension of the nearest neighbor imputation
- Distance function is a weighted combination of geographical distances and a similarity measure
- Representativeness indicators like the R-Indicator across time.

- Spatial information can make valuable contribution to the reduction of total survey error.
- Disclosure control
- Similarity measure based upon geographically aggregated auxiliary variables

Analysis

Spatial information can make valuable contribution to the analysis of social data:
- Link social science with spatial data
- Analyse effects of spatial proximity and cluster effects

Survey on linguistic & cultural competencies of children in pre-school years (Rhein-Neckar)
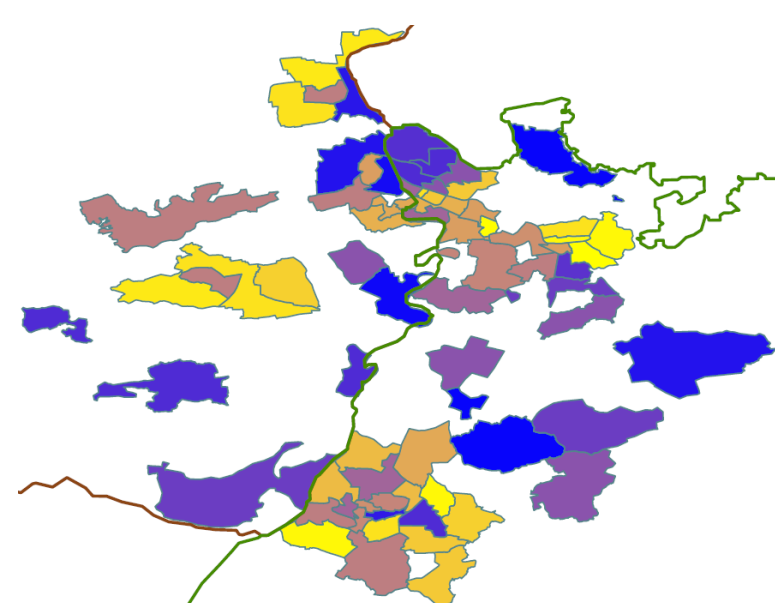
- CAPI with parents, tests with children (K-ABC)
- N=1283
- 50% turk. migrant background

Use of (linked) context information:
- Logistic regression with cluster-robust standard errors
- Multilevel regression
- Small area estimation

- Proximity is not always most important
- Investigation of cluster effects
- Expansion to other projects/datasets e.g. regional/geographical differences in the perception of...
  - ... measures to promote climatic change
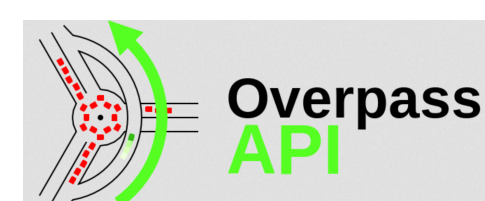 - ... big infrastructure projects

Visualisation

- The map is the most conventional way to visualize areal data.
- Maps help people to understand complex phenomena.
- Differences between regions are better understood with spatial visualization.

Perception of Energetic Change
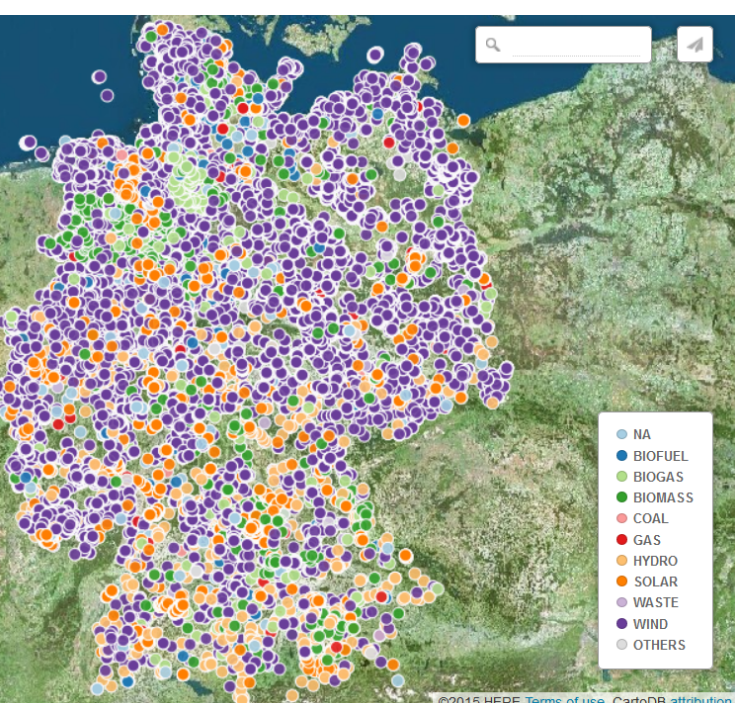
Source	#
wind	22879
solar	5619
hydro	599

Use of (context) information and cartograms:

Overpass API
PostgreSQL
CARTODB
- R-package geosmdata2 using the Overpass API
- PostgreSQL, PostGis
- Tools for visualisation e.g. R-packages sp, ggmap

- Target: draw maps clear and simple
- Ideally these maps help to understand coherence's better.
- Data basis might be incomplete

Summary

More spatial data which can be used to
- ameliorate data collection
- develop adequate weighting schemes
- analyse the spatial context
- improve understanding with visualisation

- Official data
- Collaborative data
- Survey data
- Linked data
- Synthetic data

Adequate methods are necessary for
- Data access
- Data process, transformation and editing
- Data analysis
- Visualisation

- Linked spatial data sets have great potential
- Simulations can be used to test the implementation of appropriate survey designs

Further Informations

Jan-Philipp Kolb
GESIS, Leibniz Institute for Social Sciences B2,1 D-68159 Mannheim