Opportunities through the use of Open-Street-Map data in social sciences

UseR Los Angeles

Jan-Philipp Kolb
Jan-Philipp.Kolb@gesis.org

University of California, Los Angeles

03. July 2014

Outline

Outline

Outline

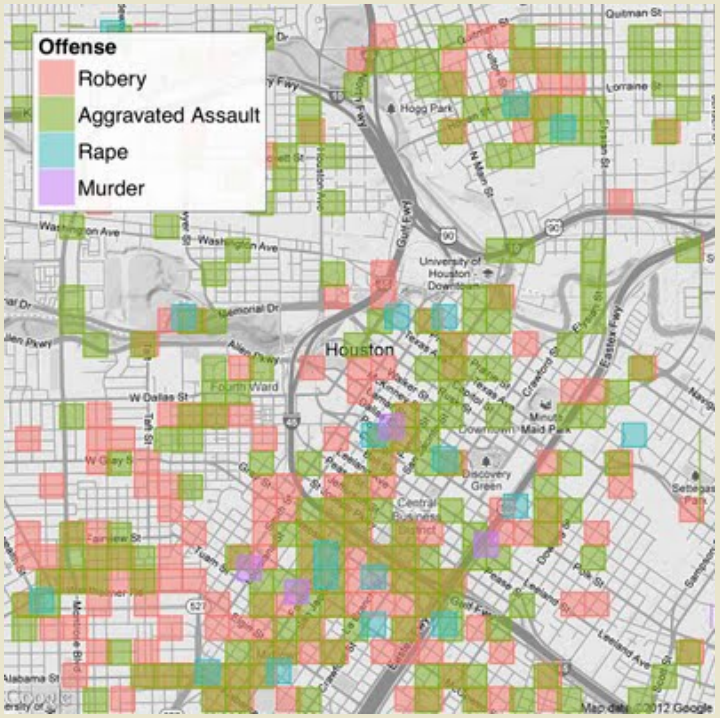
ggmap : Interfacing ggplot2 and RgoogleMaps

David Kahle
Assistant Professor
Department of Statistical Science
BAYLOR UNIVERSITY

Hadley Wickham
Assistant Professor / Dobelman Family Junior Chair
Department of Statistics
RICE

useR!
June 13, 2012

Example from David Kahle- Houston Crime Map

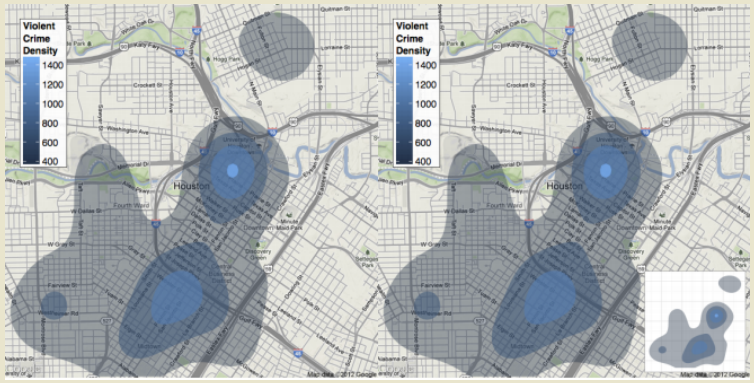


Source: `https://sites.google.com/site/davidkahle/ggmap`

R-packages - Download geo-information

R-Journal: June 2013

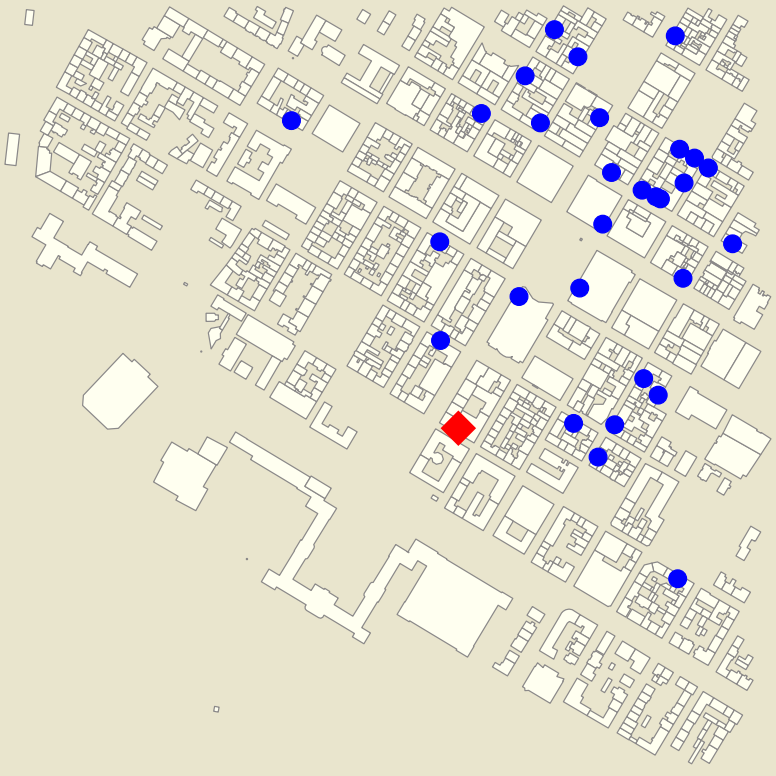
Manuel J. A. Eugster and Thomas Schlesinger:
`osmar` - OpenStreetMap and R



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

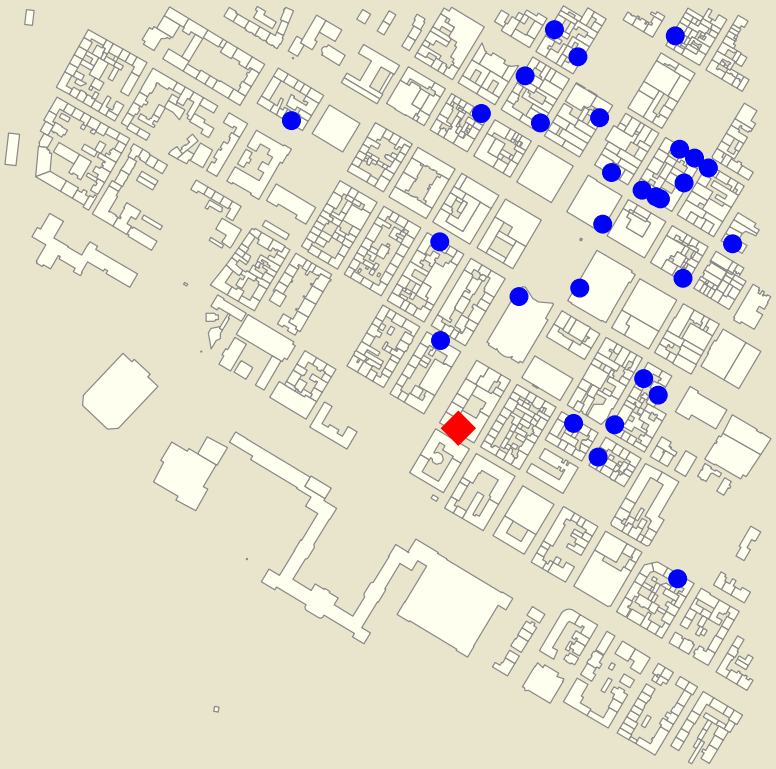
How to use this?

Information from R-package `osmar`



How to use this?

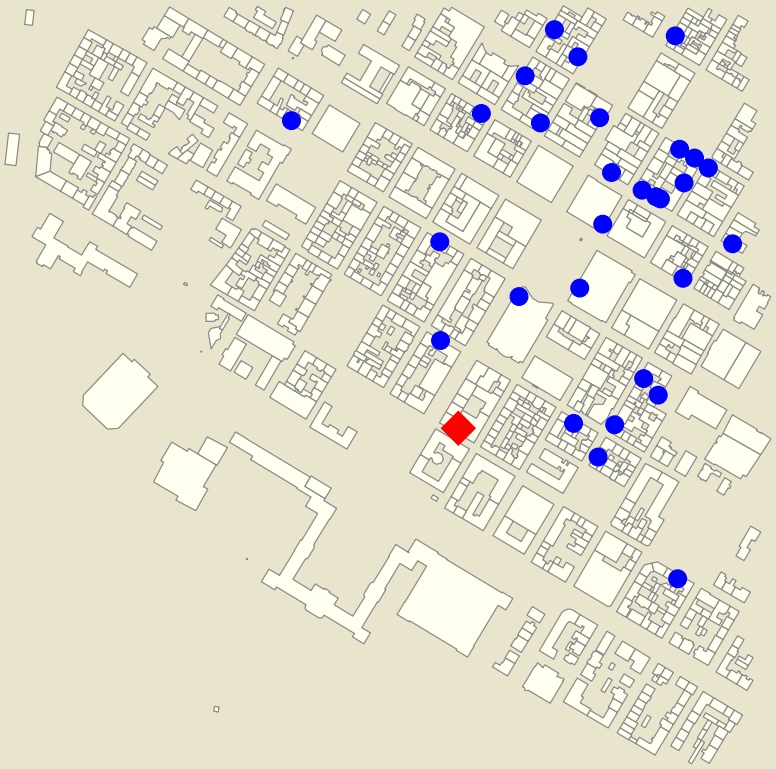
Information from R-package `osmar`



My office (GESIS Mannheim)

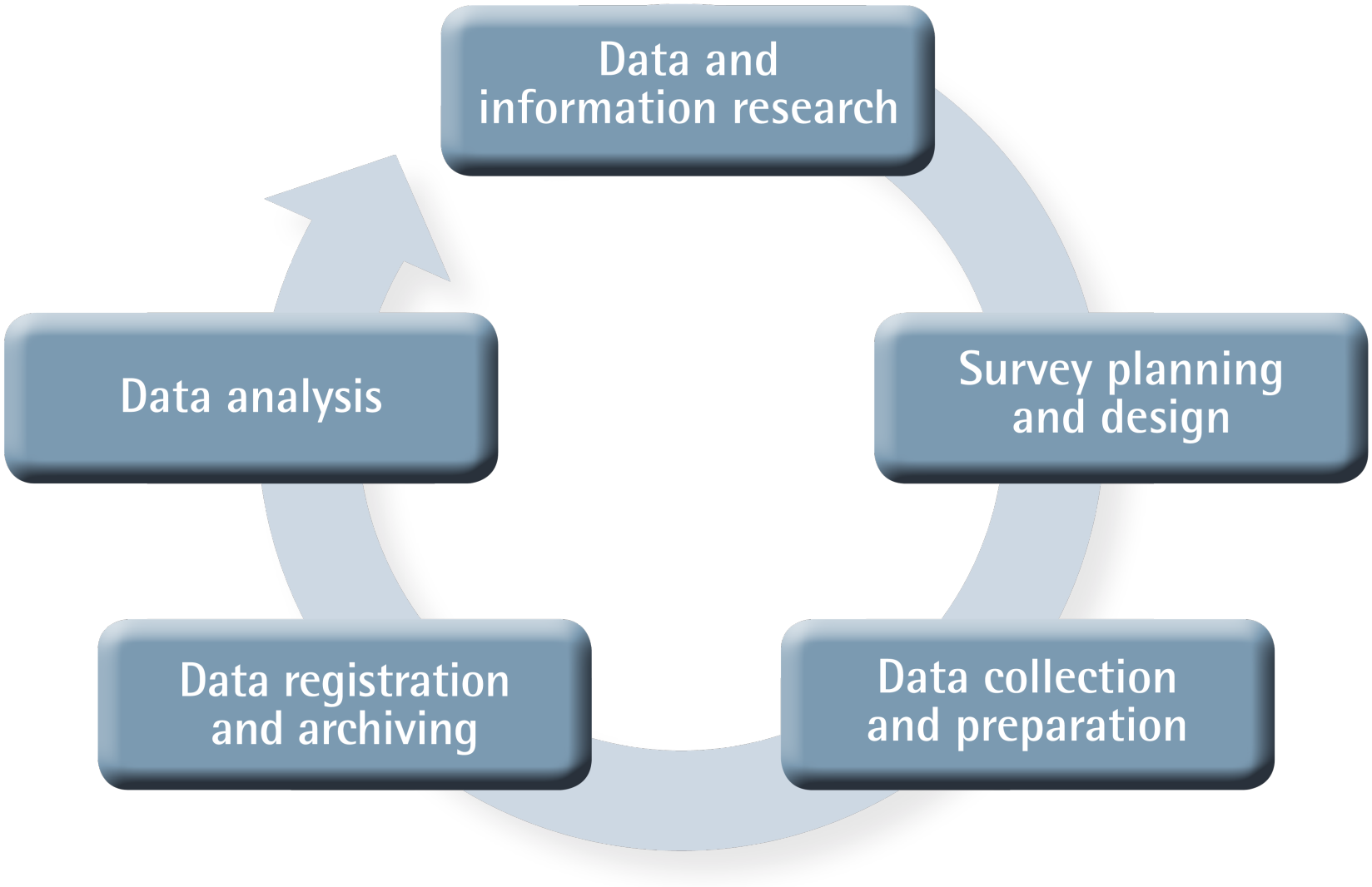
How to use this?

Information from R-package `osmar`



My office (GESIS Mannheim)
Fast Food restaurants

Presentation of GESIS



Tasks GESIS-team statistics

Consulting and research on ...

- Planning of survey designs
- Development of sample designs for face-to-face, written, and telephone-assisted surveys
- Data analysis and visualisation

Outline

Geo use case 1 - Visualisation

Package maptools from Bivand et al.

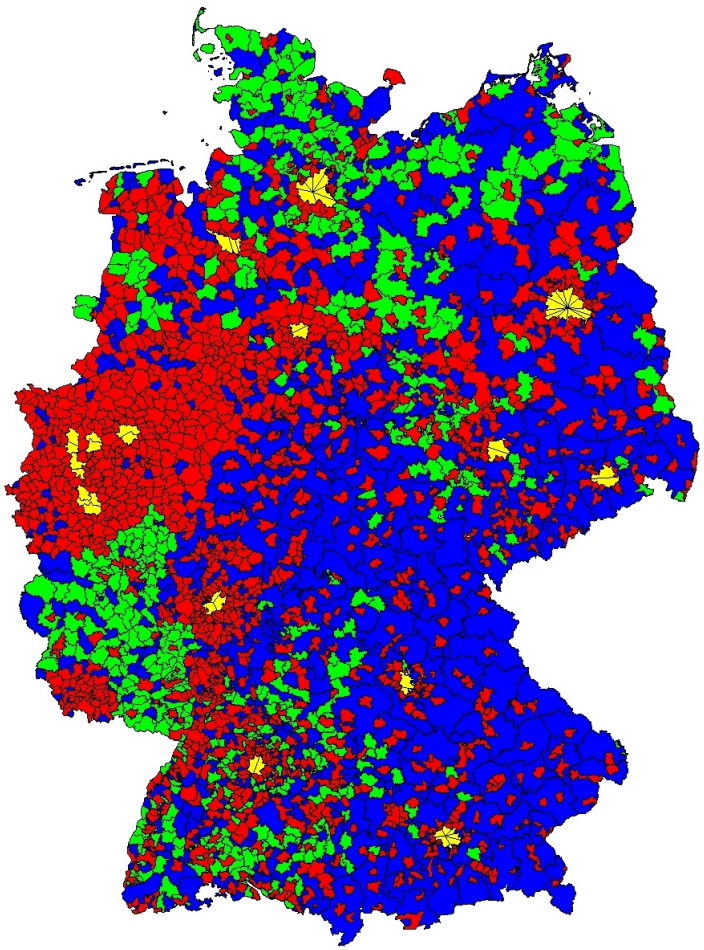
Source: www.gadm.org

Global Administrative Areas
Boundaries without limits

```
B<-readShapePoly("BRA_adm1.shp"
plot(B,col="green")
```

Geo use case 1 - Visualisation

German register-based Census 2011



SMPtypes

- Less than 10.000 inhabitants
- Less than 10.000 inhabitants
- More than 10.000 inhabitants
- More than 400.000 inhabitants

Source: Münnich, Gabler et al. (2012): Stichprobenoptimierung und Schätzung im Zensus 2011

Geo use case 2 - Stratification/Allocation

Relevant info for ...

...Sampling-point determination/Allocation

- Administrative area (district/municipality)
- Number of inhabitants in municipality

... Stratification

- Adress size

Geo use case 2 - Stratification/Allocation

Relevant info for ...

...Sampling-point determination/Allocation

- Administrative area (district/municipality)
- Number of inhabitants in municipality

... Stratification

- Adress size

$\Rightarrow$ Regional informations

Geo use case 2 - Stratification/Allocation

Package osmar from Schlesinger/Eugster

```
src<-osmsource_api()
bb<-center_bbox(8.4,49.4,
1000,1000)
ua<-get_osm(bb,source=src)
MA<-getBuildingShapes(
ua=ua,what="building")
plot(MA,col="orange")
```

Zip code area 68161

On-site information OpenStreetMap

Building size

Geo use case 3 - Analysis GeoData

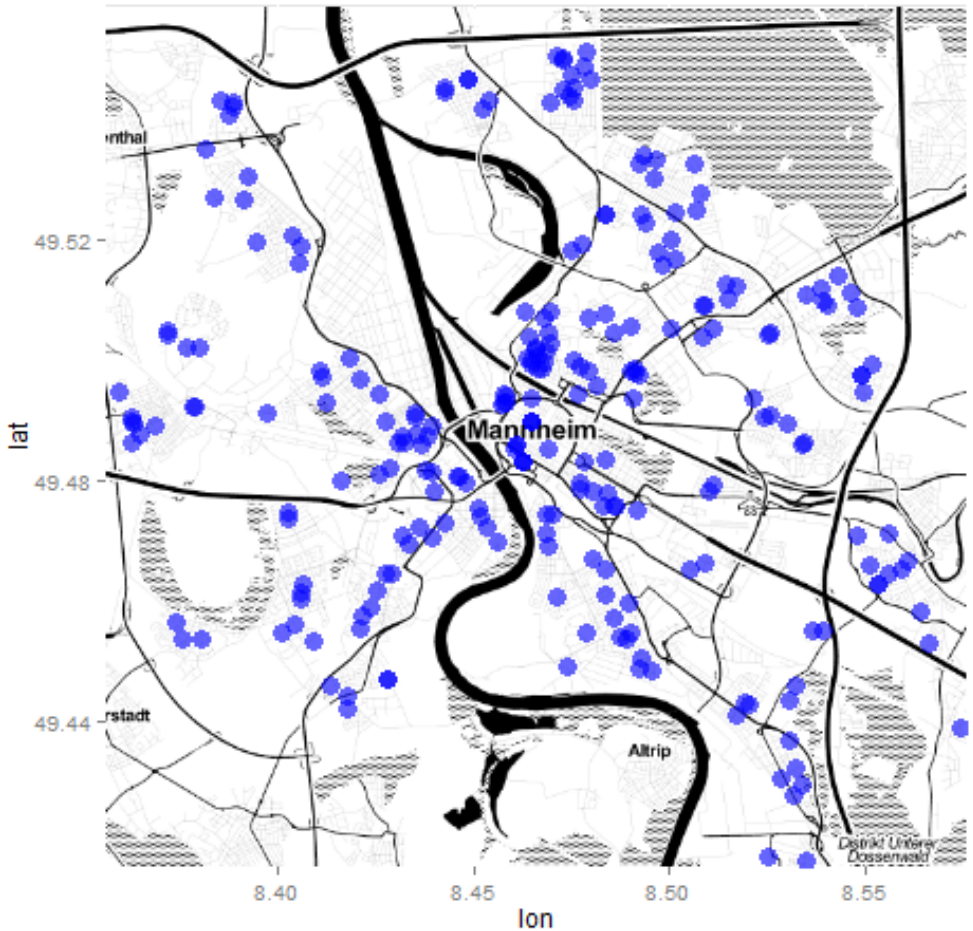
Georeferencing with R-package ggmap (Kahle/Wickham)

```
> geocode("UCLA")
Information from URL :
http://maps.googleapis.com/maps/api/geocode/json?
address=UCLA&sensor=false
Google Maps API Terms of Service :
http://developers.google.com/maps/terms
        lon      lat
1 -118.4452 34.06892
>
```

Measuring of distance with R-Package ggmap (Kahle/Wickham)

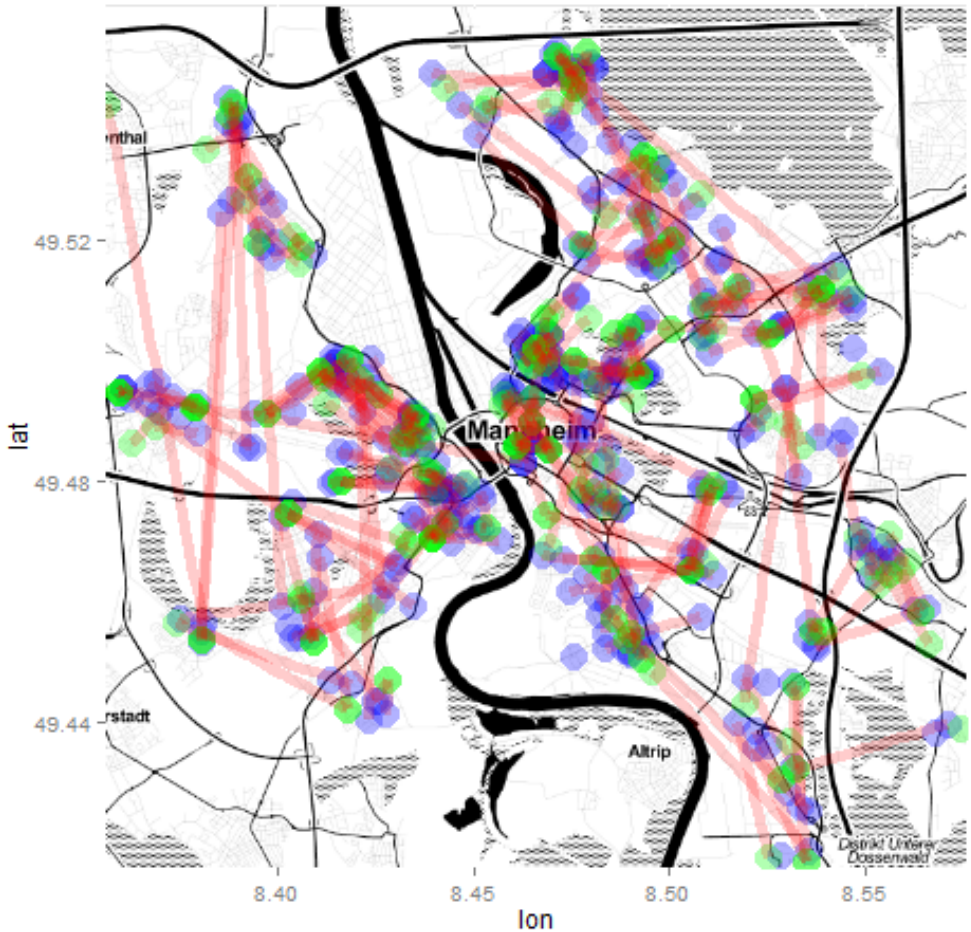
```
> mapdist("Ney York","UCLA")
Information from URL :
http://maps.googleapis.com/maps/api/distancematrix/json?
origins=Ney+York&destinations=UCLA&mode=driving&sensor=false
Google Maps API Terms of Service :
http://developers.google.com/maps/terms
      from   to       m       km    miles seconds
1 Ney York UCLA 4517037 4517.037 2806.887  145775
   minutes    hours
1 2429.583 40.49306
```

Geo use case 3 - Analysis GeoData



Geocoded place of living

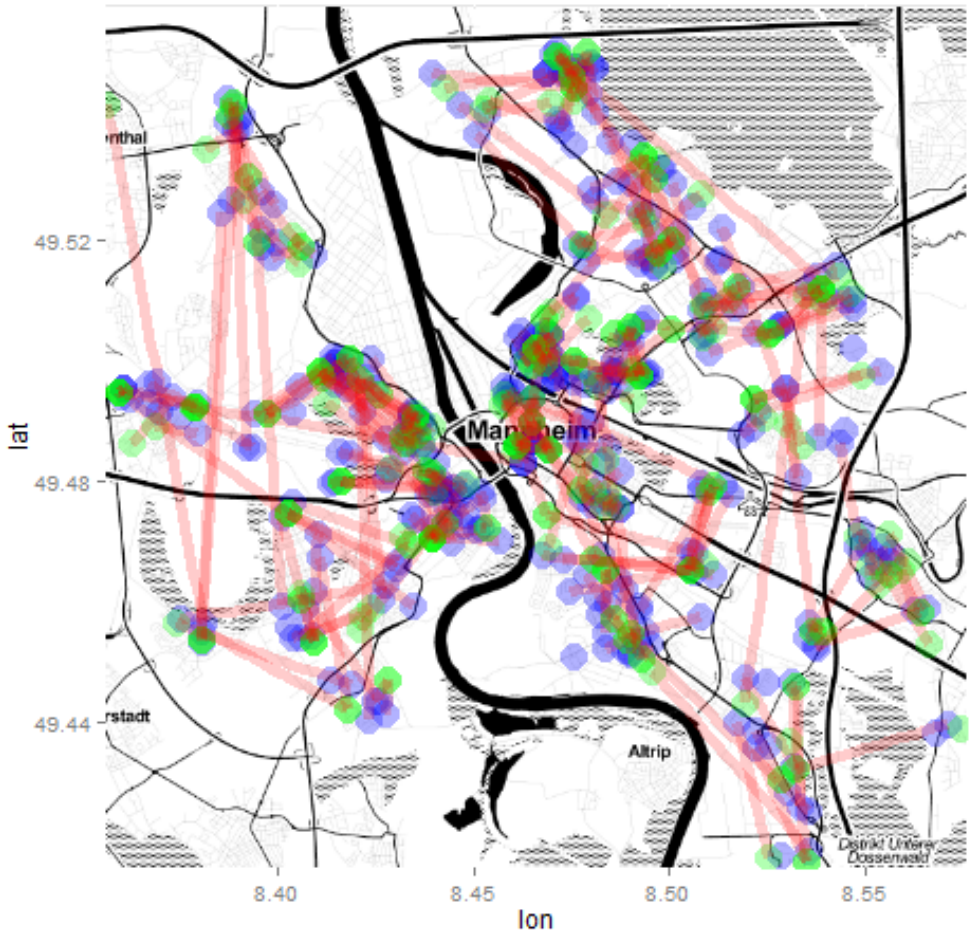
Geo use case 3 - Analysis GeoData



Geocoded place of living

Geocoded Kindergarden

Geo use case 3 - Analysis GeoData



Geocoded place of living

Geocoded Kindergarden

Distance

Geo use case 3 - Analysis GeoData

Informations from data

- How much Kindergarden are in the sourrounding?
- Who is financing the kindergardens nearby?

Geo use case 3 - Analysis GeoData

Informations from data

- How much Kindergarden are in the sourrounding?
- Who is financing the kindergardens nearby?

OpenStreetMap

- How is the structure of public transport in zip code area?
- How is the building structure in the surrounding?

Preliminary conclusions

Examples: Usage of GeoData

- Visualisation with GeoData
- Better planning of survey design with GeoData
- Analysis of Nonresponse Bias
- Integration of GeoData in Surveys
- Influence of climate data on response

Outline

Challenges

Spatial structures

- NUTS regions
- Zip code areas
- City block

Different projections available

→ Transformation?

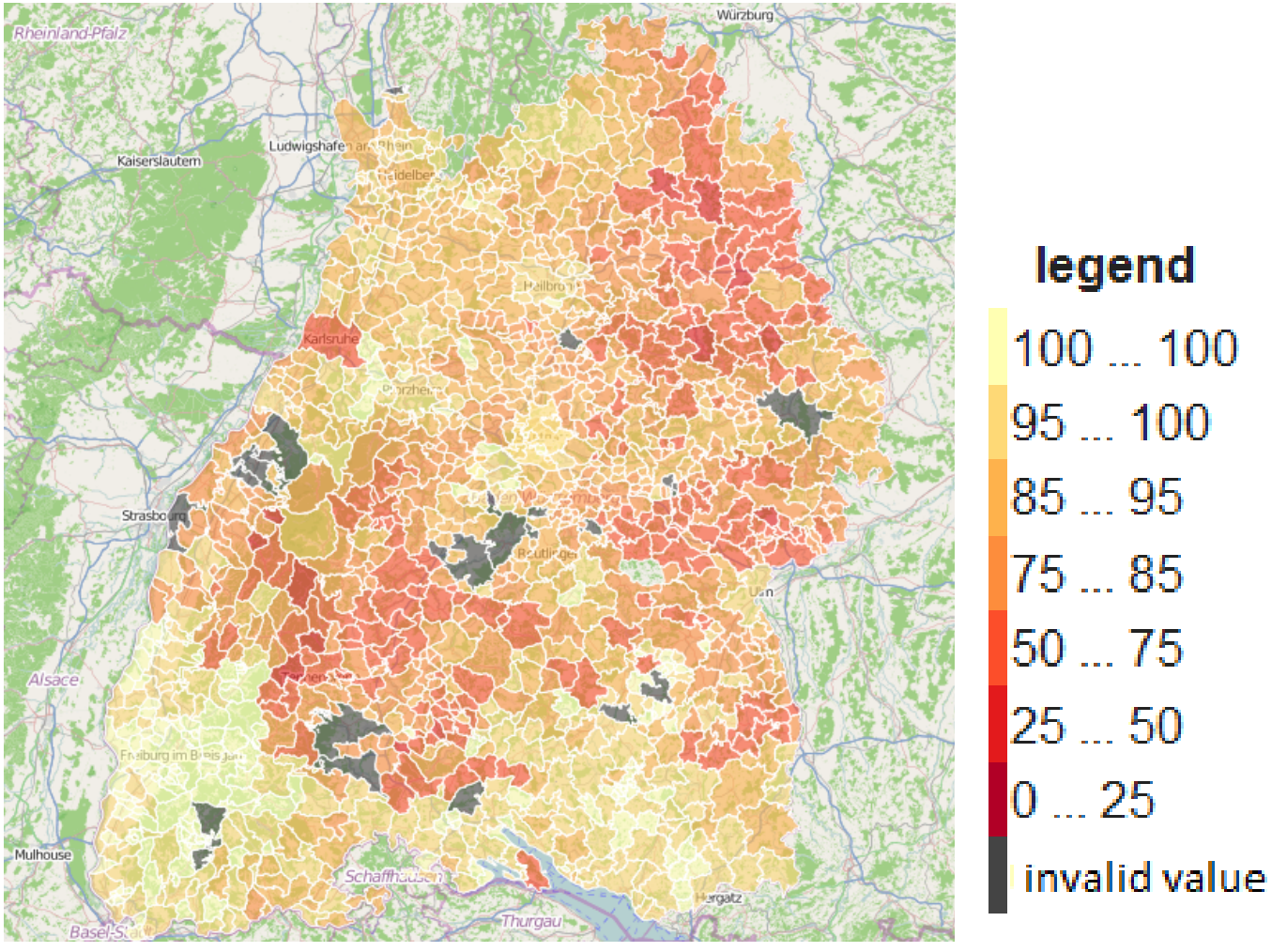
Scaling

- Download either via `osmar` or `http://download.geofabrik.de/`
- Information either only for small grids or for too big entities

Disclosure

- Data protection is very important in this coherence

Completeness of OSM-data



http://regio-osm.de/listofstreets/kartenansicht.html?file=DE_BW

Outline

Outlook

The combination of GeoData and R offers fascinating possibilities for survey statistics:

- Integration of GeoData in planning of survey-designs
- Process control - e.g. in random route methods
- Analysis of nonresponse bias
- Usage of on-site-information for analysis
- Visualisation with GeoData
- Publication of informations about the souroundings?

Fields of research

- What is the definition of the surrounding?
- How complete is the data?
- How to deal with data-disclosure?

Thank you for your attention!

Please contact me!

Jan-Philipp.Kolb@gesis.org

https://github.com/Japhilko/GeoData