Austrian Journal of Statistics
March 2016, Volume 45, 45–54.
http://www.ajs.or.at/
doi:10.17713/ajs.v45i1.88

Geovisualisation: Possibilities with R

Jan-Philipp Kolb
GESIS

Abstract

The processing of information related to a geographic location has long been difficult due to the lack of (pertinent) data sources and computational power. However, the recent developments of web-based technologies like OpenStreetMap (OSM) and Google Maps change this fundamentally. With R it is possible to process a large amount of data and produce appropriate visualisations. The challenge is to find the necessary spatial information, like appropriate polygons and data corresponding to these polygons. In this paper ways are presented to access this information via internet and to combine and visualise these information.

Keywords: spatial visualisation, choropleths, polygons, geocoding, R.

1. Introduction

Geographical visualisation facilitate the understanding of social phenomena. Often we compare the information depicted on the map with the perception of the real situation in our working or living environment. To enforce this process it is good to have a map which is as accurate as possible. In the past, it was often complicated to produce highly detailed maps with publicly available data (e.g. scientific or public use files). Mostly, complications were related with the disclosure control of data that included personal characteristics.

In the past few years, a massive amount of information has been uploaded on the internet and its volume is still growing. This information has an ubiquitous nature. In the course of this development, much information related to geographic dispositions has been published.

Primarily, this was caused by the introduction of new web based services and collaborative mapping. For the past 10 years, OpenStreetMap has offered crowd-sourced information (see Haklay 2010 and Neis, Zielstra, and Zipf 2011) and in February 2005 Google Maps was introduced, which offers a fast-loading, tiled map display and a deep user interface (c.f. Gibson and Erle 2006).

As a result, a huge amount of spatial information is now freely accessible, for example through application programming interfaces (APIs). Thus, there is a big analytic potential but the information is often unstructured or only semi-structured (e.g. web documents, news archives). In addition, this information is often very heterogeneous and not intended for geographic purposes but contains geographic information implicitly (Web 2.0). Often there is little or no metadata available.

In the following sections, examples will be presented which combine polygons and the data available from APIs to produce choropleth maps. Additionally, simple methods will be described to combine this geodata. To do so it is initially necessary to find appropriate sources for polygons. These are presented in the next section, that will explain how to access spatial information available on the internet.

Recently, numerous R packages have been published, that allow data processing and visualisation of geographical information. A short overview of the scope of these packages and the opportunities R offers for spatial information is part of Section 3. In Section 4, possibilities to visualise this information will be described. Hence, two examples of application will be used to highlight the potential of information gained from OSM for visualising spatial data. The examples will be presented in the form of choropleth maps. The paper will conclude with a closing discussion in Section 5.

2. Information access

A choropleth map shows distributions by area (Pitzl 2004). These kind of maps are often used to visualise spatial information in the social sciences. The information is then often linked to an administrative entity. Administrative entities are organized in an hierarchical manner. To create a choropleth map it has to be clarified on which level the displayed information is available. The easiest example are values related to different countries. In this example the most simple way to visualise spatial information is to use the R package **maps** (Becker, Wilks, Brownrigg, and Minka 2013) which also contains country data. Alternatively it is possible to use the R package **choroplethr** (Lamstein and Johnson 2015) or the new **tmap** package (Tennekes 2015) especially for creating thematic maps such as choropleths. Both packages contain polygons on country level.

If the information is available below the country level, it is advisable to use the R package **maptools** (Lewin-Koh, Bivand, Pebesma, Archer, Baddeley, Bibiko, Dray, Forrest, Friendly, Giraudoux *et al.* 2011). Here, the information of areas is organised in polygons (Leipzig and Li 2011). If the administrative entity is not implemented in R packages, it is necessary to download the information from an external source. For example, a world borders dataset is available on thematicmapping.org. To import shapefiles in R the package **rgdal** can be used. It has the advantage, that it can handle projection information (Keitt, Bivand, Pebesma, and Rowlingson 2011).

One of the most comprehensive sources is the Global Administrative Areas database (GADM, `www.gadm.org`), which also provides downloadable information in the format of `RData`-files which contain objects of class `SpatialPolygonsDataFrame`. The package **raster** provides the function `getData()` with which it is possible to download the data automatically.

The data can also be downloaded with the following commands:

```
con <- url("http://biogeo.ucdavis.edu/data/gadm2/R/DEU_adm3.RData")
print(load(con))
close(con)
```

For the example above, polygons for NUTS3 level[1] in Germany have been downloaded.[2] This is the most detailed information published in the GADM database and includes smaller regions and large cities. For other countries only information on NUTS1 or 2 level is available.

The availability of data necessary to visualise information that is more detailed than the NUTS3 level differs highly by country. Germany, for example, provides information on communities at geodatenzentrum.de.[3] Every country has its specific sources of information. The

[1]The Nomenclature of Territorial Units for Statistics (NUTS) is a standard for referencing the subdivisions of countries, see Eurostat (1995) for more details.

[2]For more information on how to work with shapefiles see Kennedy (2013).

[3]`www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=14&gdz_user_id=0`

US Census Bureau offers lots of information in the TIGER/Line program[4] (see for example Almquist 2010, p. 2) and this information can easily be linked with other data available from the US Census Bureau.

The number of publications which refer to this data source is an indication of how promising the program of publishing spatial information is. In Europe, the establishment of the infrastructure for spatial information (INSPIRE - website `http://inspire.ec.europa.eu`) accelerated the exchange about this topic.

Beside the possibility to plot choropleth maps for administrative areas, other areas such as electoral districts or the zip-code might be of interest as well. In Germany the *Bundesnetzagentur* offers a dataset of the prefix zones.[5]

But, as already stated, very detailed information is often not available due to reasons of disclosure control. The user has to decide on the degree of refinement necessary for her/his application. Sometimes, information is not present as data that correspondents to an administrative entity, but as address information. Then, an additional step can be interposed which is the geocoding of the information. In this situation the package **ggmap** is very useful (Kahle and Wickham 2013). The package can for example be used to geocode (function `geocode()`) points of interests (POI). It is possible to get latitude and longitude referenced to the World Geodetic System 1984 (WGS84) ellipsoid (Lovelace and Cheshire 2014, pp. 7). The query `mapdist()` provides results on the distance between two points of interest. In both cases, it is necessary to give an exact address of the location. The number of requests is limited per day and you have to pay attention to data security issues. The problem is not that important when institutions are geocoded, but might be more considerable for interviewees, especially if the questionnaire content is delicate.

Some other packages are available for geocoding, like the **geocodeHERE** package, which is a Wrapper for Nokia's HERE geocoding API (Nissen 2014). There are some possibilities to download related information from crowdsourced services like wikipedia.org. This information can be accessed using for example the **geonames** package (see for example Ceolin, Moreau, O'Hara, Schreiber, Sackley, Fokkink, van Hage, and Shadbolt 2013 or Van Hage, Van Erp, and Malaisé 2012). Other packages like the development version of the R package **tmap**[6] are using the OSM-service Nominatim (Warden 2011, p. 25).

In Germany, the block sides are often discussed as the smallest administrative area for which it would be possible to publish data. A block side is an area with the same street name which is restricted by intersections or similar geographical limitations. In fact, buildings are the smallest entities for which polygons might be used. The necessary information can be accessed via the package **osmar** (Eugster and Schlesinger 2013). In combination with the R package **sp** it is possible to transform this information into classes for points, lines, and polygons. With the **osmar** package it is possible to get information from user-generated street maps. The package uses the API provided by OpenStreetMap[7] (see for example Haklay and Weber 2008). With this API it is possible to access the offered map data that is free to use, editable, and licensed under new copyright regulations.[8]

The package **osmar** is designed to get raw OSM data. The usage of the package is described in Schlesinger (2011). There are three important steps in order to download information from OpenStreetMap via **osmar**.

- First, you have to provide the information about the API of OpenStreetMap.
- Second, the bounding box must be defined.

[4] `http://census.gov/geo/maps-data/data/tiger-line.html`

[5] `http://bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Unternehmen_Institutionen/Nummerierung/Rufnummern/ONVerzeichnisse/GISDaten_ONBGrenzen/ONBGrenzen_Basepage.html`

[6] A stable development version can be installed with: `devtools::install_github("mtennekes/tmap/pkg", ref = "45855fa")`.

[7] `http://api.openstreetmap.org/api/0.6/`

[8] `http://rstudio-pubs-static.s3.amazonaws.com/12696_9fd49fb7055c40ff9b3a3ea740e13ab3.html`

- and the third step is the download of the information.

The bounding box can be defined with `center_bbox()`, one has to specify the latitude, longitude and the size of the box as arguments. The boxsize is rendered in metres. In the example below, we have a box of 500 × 500 meters. The download time and the object size depend very much on the size of the chosen bounding box and of course on the number of objects in the bounding box.

```
library("ggmap")
library("osmar")
src <- osmsource_api()
Ma_Schloss <- geocode("Mannheim Schloss")
bb_MA_S <- center_bbox(Ma_Schloss$lon, Ma_Schloss$lat, 800, 800)
ua_MA_S <- get_osm(bb_MA_S, source = src)
```

However, the information downloadable by **osmar** is limited by boxsize. As a maximum an area of 0.25 square degrees can be queried (Eugster and Schlesinger 2013, p. 2). Another interesting option which is available for R is the usage of some services that offer the download of data extracts from OpenStreetMap. One of these services is geofabrik.de, where general information is available for free and specialised requests require a fee.

With OpenStreetMap it is possible to combine geographic information and social web contents. This is possible because more and more objects are geotagged (see for instance Scharl and Tochtermann 2009). A good example is the website flickr.com where many photos are geotagged. That means that the photos are associated with latitude and longitude and the user generated content can therefore be used for social analysis (Yee and Moodle 2008, p. 245). In this case, information is used that was not originally created for this purpose. You, DesArmo, and Joo (2013) mine for example user-generated text on Flickr to measure the happiness of US citizens. Sizov (2010) uses Flickr data to evaluate an algorithm for content management, retrieval and sharing.

The R package **twitteR** enables the access to geographic information from the twitter.com API (Gentry 2015). Kaczmirek, Mayr, Vatrapu, Bleier, Blumenberg, Gummer, Hussain, Kinder-Kurlanda, Manshaei, Thamm *et al.* (2013) show how to use Twitter data for social and political research. They monitor the campaigns for the 2013 German Bundestag elections in social media.

Also data from OpenStreetMap can be a valuable source of information. With this data it is for example possible to analyse the number of services available for children in one zip-code area. Furthermore it is possible to compute the floor area of the buildings.

3. Information processing

Numerous R packages are available to process the spatial information gained from the sources described above. A good overview of these packages is available at the CRAN Task View about Analysis of Spatial Data from Roger Bivand.[9] The book of Bivand, Pebesma, and Gómez-Rubio (2013) can also be recommended in this coherence. An overview of the implementation of spatial data analysis software tools in R is available in Bivand (2006). Most of the important packages are listed there, some will be described in the following.

The **maptools** package was already mentioned as a source for polygons. But this package is also very useful for the information processing of spatial data (Bivand 2011, p. 18). The **maptools** package has been adapted to use **sp** classes, and in combination with the **sp** package (Pebesma, Bivand, Rowlingson, and Gomez-Rubio 2013), it provides a good set of tools to process spatial data.

If data sets from different sources are combined, it is important to ensure that the same map projection is used. If that is not the case, transformation can be done for example with the

[9] `http://cran.r-project.org/web/views/Spatial.html`

function spTransform from package **sp**. See (Pebesma 2012, pp. 20) or (Rossiter 2012, pp. 9) for examples.

These and other packages enable the usage of the R language as a geographic information system (GIS). Here the package **rgeos** is very useful (Bivand and Rundel 2013) for basic topology operations.

rgdal for example can be used to bridge information from the geospatial data abstraction library (GDAL, www.gdal.org) to R (Keitt *et al.* 2011). To import shapefiles, **rgdal** has the advantage that it takes care of the projection, which is not the case for **maptools**. Alternatively the read_shape-command from **tmap** can be used.

The **raster** package is for reading, writing, manipulating, analysing and modeling of gridded spatial data (Hijmans and van Etten 2014). The package **rgrass7** provides an interface between the geographic information system GRASS 7.0 and R (Bivand 2015). The **rgeos** package is also useful to edit polygons.

In case of the example depicted in this article, the shapefiles downloaded for example from geodatenzentrum.de or other sources can be bridged to R using the readOGR command from the **rgdal** package.

```
library("rgdal")
zip <- readOGR(".", "post_pl")
```

It is then possible to plot and edit the polygons with the spplot-function from the **sp** package (Bivand *et al.* 2013, pp. 73). The data downloaded from OpenStreetMap via the package **osmar** has to be edited (for example with the **sp** package) to get such polygons. First steps to the processing of the information from OpenStreetMap are described in (Eugster and Schlesinger 2013). To convert the OpenStreetMap information to polygons, the following commands can be used:

```
ua_ids <- find(ua_MA_S, way(tags(k == "building")))
ua_ids <- find_down(ua_MA_S, way(ua_ids))
bg <- subset(ua_MA_S, ids = ua_ids)
bg_erg <- as_sp(bg, "polygons")
```

A wrapper around these functions is provided in the development version of the **tmap** package:

```
bb_schloss <- bb(q="Mannheim Schloss")
sp_schloss <- read_osm(bb_schloss, building = osm_poly("building"),
                       castle = osm_poly("historic=castle"))
qtm(sp_schloss$building, fill = "ivory", borders = "snow4") +
  qtm(sp_schloss$castle, fill = "royalblue")
```

As already described, it is only possible to download the data set for a small area. To get complete information about areas on NUTS3-level, many queries have to be realised. It is therefore advisable to choose a top-down approach. In the present case, this implies the download of information from geofabrik.de and the division into smaller entities. The information is not provided for entities more detailed than the administrative district. In the example it is the administrative district Karlsruhe. It is not advisable to plot all the information in one map. Hence, the available information must be processed. Due to the limited scope of this article the procedure will not be described. For details on this procedure please refer to the GitHub Repository http://github.com/Japhilko/GeoData.

4. Visualisation of information

Much potential is available to visualise information related to spatial data. Every type of geographical presentation is feasible in R if the conditions are satisfied. The most important condition is the availability of information necessary for the desired visualisation. In some research fields it is easier to get the data, in others it is more complicated. Topographic maps for example can be visualised with **GEOmap** (Lees 2008). In the following, visualisations for social science data are addressed.

The **sp** package offers many fascinating possibilities to visualise spatial data (e.g. Endel and Filzmoser 2012). One type of spatial visualisations that is of particular interest in the social sciences are choropleth maps. This type of maps can be used to display area values. Often values are depicted on NUTS3 level and upwards. But, the finer the scale, the more interesting the map can be. The difficulty to access such detailed information makes the combination of choropleth maps and OSM data so interesting. In the given example, the information from OpenStreetMap is used to produce a choropleth map. Polygons of semi-administrative areas (zip-code areas) are connected with the aggregated information from OpenStreetMap.

As already described, polygons for entities below the community level are available for download using the **osmar** package. The R code for this minimal working example is presented in Appendix A. As a result, we get the buildings of the city of Mannheim with its castle highlighted in blue.

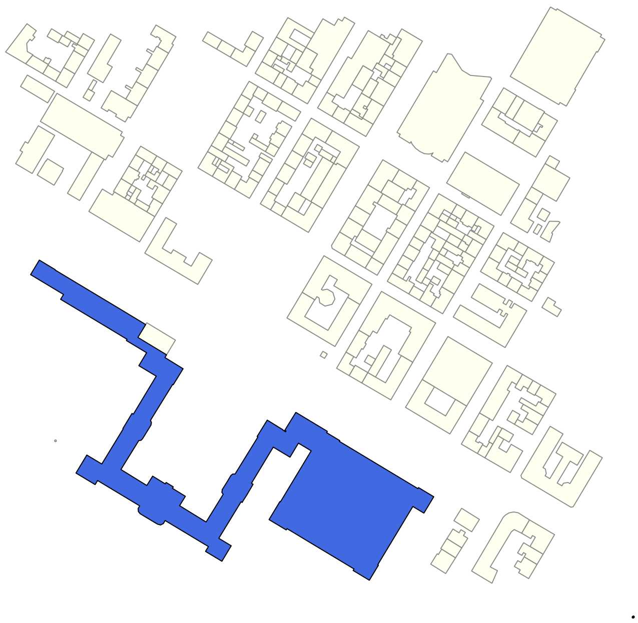

Figure 1: Buildings around the castle of Mannheim.

In a next step the user may want to combine the information gained from web-services like OpenStreetMap with polygons. In Figure 2 this is done for schools.

The OSM community did an amazing job in collecting all the information. But there are of course still white spots in regions, where not so many OSM members are active or the public authorities do not provide necessary information. In addition, one has to keep in mind that the situation is changing very fast. In Figure 1 for example the district court is missing, which is located across the castle. Haklay (2010) provides a list of criteria to evaluate geographical information.

The information downloaded from OpenStreetMap can of course also be used for other purposes than only visualising a social coherence. For example Behrisch, Bieker, Erdmann, and Krajzewicz (2011) show how to use the information for a geographic simulation model. Lovelace and Cheshire (2014) give an introduction to visualise spatial data in R.

Without a question, the analysis of geodata and context-related information offers a great potential for scientific purposes. However, the publication of geodata might be in conflict with data security. In the present case, this is a minor problem because it is hardly possible to extract any sensitive information from this choropleth map.

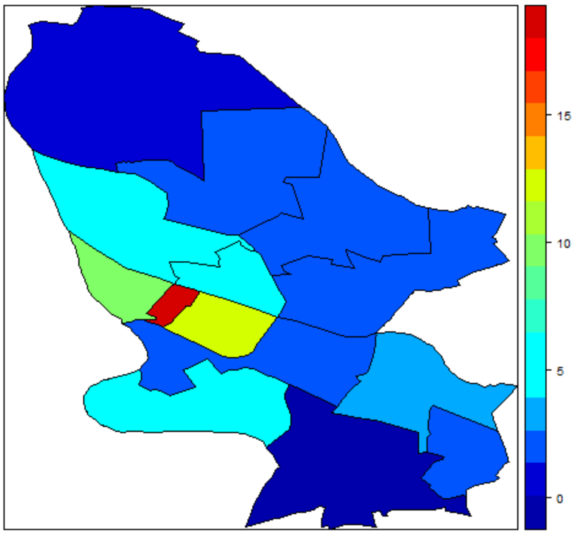

Figure 2: Number of schools per zip-code area in Mannheim.

5. Conclusion

The location of a POI is of special interest for the social sciences. But in practice, the sample size is often too small to create choropleth maps on NUTS1 level with survey-information, while hardly any other official dataset with geolocated information was available for researchers. That changed due to crowd-sourced information which is nowadays available. It is possible to access this information by using the so called API. Many services, like Google Maps, OpenStreetMap, Flickr, Twitter or Wikipedia do provide such API and the R environment offers many possibilities to use them.

Shapefiles are available for administrative areas but the level differs widely across the countries. Self-created polygons for non-administrative areas can be created using the R package **osmar**. This package uses information from OpenStreetMap which is a prime example for volunteered geographic information.

In general, one can say that the type of entities depends very much on the area of research. The challenge is to find and visualise context-related information. The next challenge is to show this information in reasonable visualisations. In this situation, the advantages of R can be used to shed light into the dark.

A. Transfer and plot data from OpenStreetMap

```
library("osmar")
library("ggmap")

src <- osmsource_api()
Ma_Schloss <- geocode("Mannheim Schloss")
bb_MA_S <- center_bbox(Ma_Schloss$lon, Ma_Schloss$lat, 800, 800)
ua_MA_S <- get_osm(bb_MA_S, source = src)

# filter buildings and convert to sp object
ua_ids <- find(ua_MA_S, way(tags(k == "building")))
ua_ids2 <- find_down(ua_MA_S, way(ua_ids))
bg <- subset(ua_MA_S, ids = ua_ids2)
bg_sp <- as_sp(bg, "polygons")

id <- ua_MA_S$ways$tags$id
vs <- ua_MA_S$ways$tags$v
```

```
id_s <- id[grepl("Schloss Mannheim", vs)]
plot(bg_sp, col = "ivory", border = "snow4")
plot(subset(bg_sp, id %in% id_s), col="royalblue", add = TRUE)
```

B. Subset POI from Geofabrik extracts

```
# Download postal code polygons from
# http://datahub.io/de/dataset/postal-codes-de

library("sp")
library("rgdal")
PLZ <- readOGR(".", "post_pl")

# Download OSM information from geofabrik.de
points <- readOGR(".", "points")

# Filter data sets
MA <- subset(PLZ, PLZORT99 == "Mannheim")
MA$PLZ99 <- droplevels(MA$PLZ99)
schools <- subset(points, type == "school")
proj4string(MA) <- proj4string(points)

# Get number of schools per ZIP-code area
tmp <- over(schools, MA)
school_freq <- data.frame(freq = tapply(tmp$PLZ99, tmp$PLZ99, length))
school_freq[is.na(school_freq)] <- 0
MA@data <- merge(MA@data, school_freq, by.x = "PLZ99", by.y = 0, all = TRUE)
spplot(MA, "freq")
```

The information from Geofabrik can be downloaded using the following link:
http://download.geofabrik.de/europe/germany/baden-wuerttemberg-latest.shp.zip

References

Almquist ZW (2010). "US Census Spatial and Demographic Data in R: The **UScensus2000** Suite of Packages." *Journal of Statistical Software*, **37**, 1–31.

Becker RA, Wilks AR, Brownrigg R, Minka TP (2013). ***maps**: Draw Geographical Maps*. URL http://CRAN.R-project.org/package=maps.

Behrisch M, Bieker L, Erdmann J, Krajzewicz D (2011). "SUMO – Simulation of Urban Mobility – An Overview." In *SIMUL 2011, The Third International Conference on Advances in System Simulation*, pp. 55–60.

Bivand R (2011). "Geocomputation and Open Source Software: Components and Software Stacks." *NHH Dept. of Economics Discussion Paper*, **23**. doi:10.2139/ssrn.1972280.

Bivand R (2015). ***rgrass7**: Interface Between GRASS 7 Geographical Information System and R*. R package version 0.1-2, URL http://CRAN.R-project.org/package=rgrass7.

Bivand R, Rundel C (2013). "**rgeos**: Interface to Geometry Engine - Open Source (GEOS)." *R package version 0.3-3*. URL http://cran.r-project.org/web/packages/rgeos/index.html.

Bivand RS (2006). "Implementing Spatial Data Analysis Software Tools in R." *Geographical Analysis*, **38**(1), 23–40.

Bivand RS, Pebesma EJ, Gómez-Rubio V (2013). *Applied Spatial Data Analysis with R*. Second edition. Springer, New York. ISBN 978-1-4614-7617-7.

Ceolin D, Moreau L, O'Hara K, Schreiber G, Sackley A, Fokkink W, van Hage WR, Shadbolt N (2013). *Reliability Analyses of Open Government Data*. Proceedings of the 9th International Workshop on Uncertainty Reasoning for the Semantic Web. URL http://eprints.soton.ac.uk/357160/1/paper6.pdf.

Endel F, Filzmoser P (2012). "R & GIS: Geospatial Plotting." In *Mathematical Modelling*, volume 7, pp. 618–623.

Eugster MJ, Schlesinger T (2013). "**osmar**: OpenStreetMap and R." *The R Journal*, **5**(1), 53–63.

Eurostat (1995). "Nomenclature of Territorial Units for Statistics." URL http://ec.europa.eu/eurostat/web/nuts/.

Gentry J (2015). ***twitteR**: R Based Twitter Client*. URL http://CRAN.R-project.org/package=twitteR.

Gibson R, Erle S (2006). *Google Maps Hacks*. O'Reilly Media, Inc. ISBN 978-0-596-10161-9.

Haklay M (2010). "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets." *Environment and Planning B, Planning & Design*, **37**(4), 682–703.

Haklay M, Weber P (2008). "OpenStreetMap: User-Generated Street Maps." *IEEE Pervasive Computing*, **7**(4), 12–18.

Hijmans RJ, van Etten J (2014). ***raster**: Geographic Data Analysis and Modeling*. URL http://CRAN.R-project.org/package=raster.

Kaczmirek L, Mayr P, Vatrapu R, Bleier A, Blumenberg M, Gummer T, Hussain A, Kinder-Kurlanda K, Manshaei K, Thamm M, *et al.* (2013). "Social media monitoring of the campaigns for the 2013 german bundestag elections on facebook and twitter." *pre-print*. URL http://arxiv.org/abs/1312.4476.

Kahle D, Wickham H (2013). "**ggmap**: Spatial Visualization with **ggplot2**." *R Journal*, **5**(1).

Keitt TH, Bivand R, Pebesma E, Rowlingson B (2011). ***rgdal**: bindings for the Geospatial Data Abstraction Library*. URL http://CRAN.R-project.org/package=rgdal.

Kennedy MD (2013). *Introducing Geographic Information Systems with ArcGIS: A Workbook Approach to Learning GIS*. John Wiley & Sons. ISBN 978-1-118-15980-4.

Lamstein A, Johnson BP (2015). ***choroplethr**: Simplify the Creation of Choropleth Maps in R*. R package version 3.1.0, URL http://CRAN.R-project.org/package=choroplethr.

Lees J (2008). ***GEOmap**: Topographic and Geologic Mapping*. URL http://CRAN.R-project.org/package=GEOmap.

Leipzig J, Li XY (2011). *Data Mashups in R*. O'Reilly Media, Inc. ISBN 978-0-596-55964-9.

Lewin-Koh NJ, Bivand R, Pebesma E, Archer E, Baddeley A, Bibiko H, Dray S, Forrest D, Friendly M, Giraudoux P, *et al.* (2011). ***maptools**: Tools for Reading and Handling Spatial Objects*. URL http://CRAN.R-project.org/package=maptools.

Lovelace R, Cheshire J (2014). "Introduction to Visualising Spatial Data in R." *Technical report*, National Centre for Research Methods. URL http://eprints.ncrm.ac.uk/3295/.

Neis P, Zielstra D, Zipf A (2011). "The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011." *Future Internet*, **4**(1), 1–21.

Nissen C (2014). ***geocodeHERE**: Wrapper for Nokia's HERE Geocoding API*. URL http://CRAN.R-project.org/package=geocodeHERE.

Pebesma E (2012). "**spacetime**: Spatio-temporal Data in R." *Journal of Statistical Software*, **51**(7), 1–30.

Pebesma E, Bivand R, Rowlingson B, Gomez-Rubio V (2013). ***sp**: Classes and Methods for Spatial Data*. URL http://CRAN.R-project.org/package=sp.

Pitzl GR (2004). *Encyclopedia of Human Geography*. Greenwood Publishing Group. ISBN 978-0313320101.

Rossiter D (2012). *Applied Geostatistics Exercise 9: R and GIS*. URL http://tw.rpi.edu/media/latest/ex9.pdf.

Scharl A, Tochtermann K (2009). *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. Springer, London. ISBN 1-84628-826-6.

Schlesinger T (2011). "OpenStreetMap in R - Freie Räumliche Daten für geostatistische Analysen." Thesis (Bachelor). URL http://epub.ub.uni-muenchen.de/12463/.

Sizov S (2010). "GeoFolk: Latent Spatial Semantics in Web 2.0 Social Media." In *Proceedings of the third ACM international conference on Web search and data mining*, pp. 281–290. ACM.

Tennekes M (2015). ***tmap**: Thematic Maps*. R package version 1.0, URL http://CRAN.R-project.org/package=tmap.

Van Hage WR, Van Erp M, Malaisé V (2012). "Linked Open Piracy: A Story about e-Science, Linked Data, and Statistics." *Journal on Data Semantics*, **1**(3), 187–201.

Warden P (2011). *Data Source Handbook*. O'Reilly Media, Inc. ISBN 978-1-4493-0314-3.

Yee R, Moodle M (2008). *Pro Web 2.0 Mashups*. Apress, Berkeley, CA; New York. ISBN 978-1590598580.

You S, DesArmo J, Joo S (2013). "Measuring happiness of US cities by mining user-generated text in Flickr.com: A pilot analysis." *Proceedings of the American Society for Information Science and Technology*, **50**(1), 1–4.

Affiliation:

Jan-Philipp Kolb
GESIS - Leibniz Institute for the Social Sciences
Survey Design and Methodology
68072 Mannheim
E-mail: Jan-Philipp.Kolb@gesis.org
URL: www.gesis.org

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 45 *Submitted:* 2014-10-31
March 2016 *Accepted:* 2015-08-26