

German Willkommenskultur - Where and Why?

HSoG, Fall 2015, class taught by Christopher Gandrud

Lisa Schmid and Christopher Cosler

November 13, 2015

Word count: 1785

1 Introduction and Research Question

Germany is experiencing an unprecedented inflow of refugees. Parallel to the rise in the number of asylum seekers, the number of voluntary organizations (refugee initiatives) that offer help to the people arriving seemed to be increasing over the summer of 2015 as well and more and more citizens volunteered in these organizations (Spiegel Online, 2015; Karakayali and Kleist 2015). Politicians, social activists and researchers uniformly claim that this kind of help is an important building block in managing the challenge of welcoming and supporting the asylum seekers. Foreign media is astonished by this German “Willkommenskultur” (New York Post 2015).

Surprisingly though, these organizations are unevenly distributed throughout the country. In our paper, we try to answer the following research question: What determines the existence of refugee initiatives?

Our research proceeds as follows: We extract the location of existing refugee initiatives from a Google map, customized by ProAsyl (Pro Asyl 2015). We then link each initiative to its respective administrative district (data from Bundesamt für Kartographie und Geodäsie 2015) and a range of variables describing characteristics of the district. The data for the independent variables stems from Statistisches Bundesamt (2015) and Statistische Ämter des Bundes und der Länder (2015).

Data collection, analysis and presentation are implemented in R (R Core Team 2015).

2 Short Summary of Hypotheses

What are the determinants of political participation? According to Böhnke (2011), socio-demographic (age, gender, education, family status), economic (income, employment status), and personality factors play a role when trying to predict political participation. In this paper we will test the effect of age, gender, education, unemployment, GDP, population density, and number of refugees. In addition, we will control for East and West German districts and classical political participation, measured in voter turnout at the last German federal election in 2013.

3 Description of Data Sources

We retrieve our data from three different sources:

1. *ProAsyl map*¹
2. *German Federal Agency for Cartography and Geodesy*²

¹<https://www.google.com/maps/d/viewer?mid=zc6TdvfelKuY.kUvriXoSREXw>

²http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=19&gdz_user_id=0

3. *German Federal Statistical Office*³
4. *German Regional Statistical Offices*⁴

3.1 Dependent Variable

3.1.1 List and Location of German Refugee Initiatives

The number of refugee initiatives per district is scraped from a Google map customized by ProAsyl (Pro Asyl 2015). We exchanged emails with ProAsyl about their methodology. According to their answers, the entries in the map are based on (1) information from the local refugee councils (Flüchtlingsräte), (2) initiatives already known by ProAsyl, (3) initiatives which contacted ProAsyl after ProAsyl had promoted their map via social media and German news outlets. Each initiative is checked for whether their purpose really is to help refugees and what the possibilities of volunteering are. The map is updated more or less once a week; inactive initiatives are deleted.

In the R code, we download the data as a KML file using the link provided at the Google Map interface and the `GET` command from the `httr` package (Wickham 2015). The link can be found in the Gather code. The KML file can easily be converted to characters, which allows us to search for the relevant data marked by the respective tags (we used `strapply` from the `gsubfn` package to extract the data (Grothendieck 2014)). The result was a table with the coordinates, name, link and description of each initiative. The number of initiatives is constantly changing over time but the code is written in a way that it can handle the varying numbers.

3.1.2 Shapefiles for German Districts

The shapefiles for the German districts are retrieved from the German Federal Agency for Cartography and Geodesy Bundesamt für Kartographie und Geodäsie⁵ (see print of map in Figure 1). We manually downloaded the map in UTM32 format as an automatic download is not possible. If you want to reproduce this step, simply store the file in a sub folder called “Data/Shapefile”. The Gather code unpacks the zip and imports the correct data (using the `readOGR` function from the `rgdal` package (Bivand, Keitt, and Rowlingson 2015)). The shape file is in UTM32 format and has to be converted to WGS84 format which we did with the `sptransform` function from the `sp` package (Pebesma and Bivand 2015).

To identify in which district the respective initiatives are located, the coordinates are first converted into spatial coordinates with the `coordinates` function and then located in the respective districts with the `over` method, both from the `sp` package as well (Pebesma and Bivand 2015). The result is a table where for each district the number of initiatives is recorded.

3.2 Independent Variables

The independent variables were gathered from two sources: the German Federal Statistical Office and the Regional Statistical Offices. As the API is not free of charge for both offices, the data had to be downloaded manually from the two webpages. All data sets can be found in the sub-folder “Data/Independent Variables” on GitHub.⁶ No URLs leading to the pages from where to download the data sets can be provided, as they expire after a few days. Instead you find the specific table number, assigned to the data by the two platforms, in the Gather code. All data sets were downloaded as csv-files. No changes were made to the files before importing them into R.

³<https://www-genesis.destatis.de/genesis/online>

⁴<https://www.regionalstatistik.de/genesis/online>

⁵http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=19&gdz_user_id=0

⁶https://github.com/ChristopherCosler/CSSR_DataAnalysis

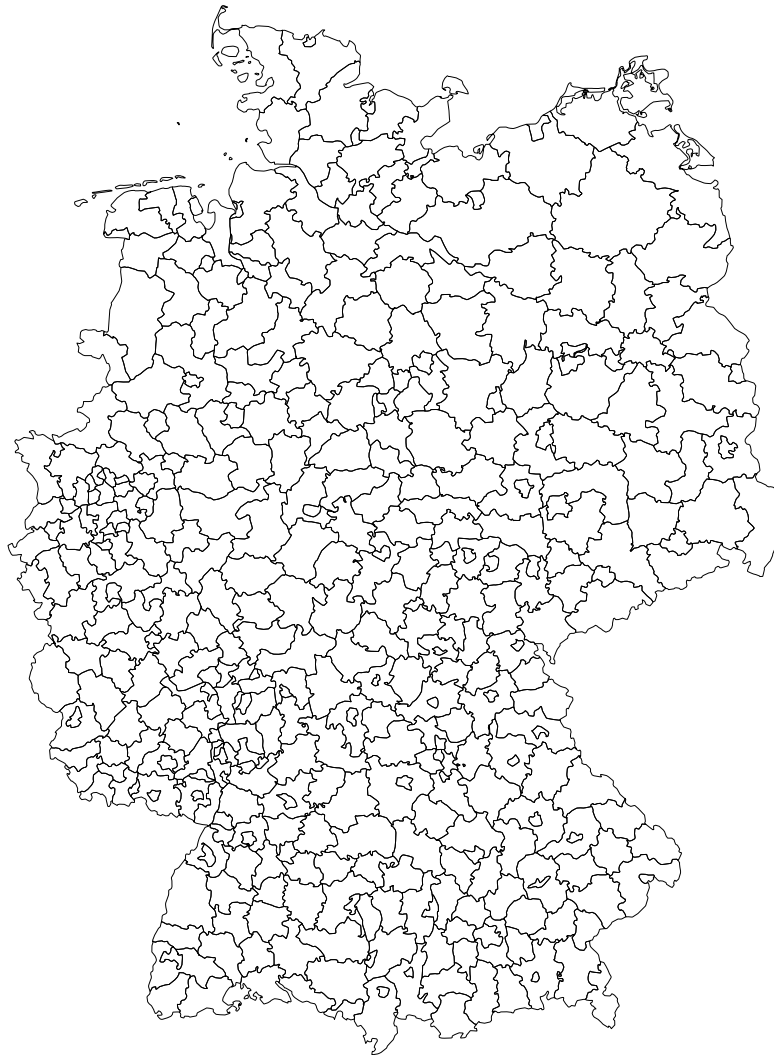


Figure 1: Administrative Districts (Landkreise) in Germany



Figure 2: Administrative Districts (Landkreise) in Germany and Refugee Initiatives (Dots)

3.2.1 Old Age Dependency Ratio

For testing our age hypothesis, we used age distribution data from Statistisches Bundesamt (2015)⁷. From this data set we computed the Old Age Dependency Ratio. The Old Age Dependency Ratio is computed by dividing all persons aged 65 and above by all persons aged 15 to 65. When the number is above 1, it means that there are an over-proportionate number of older persons in the population; when the number is below 1, it means that there is an over-proportionate number of younger persons in the population.

3.2.2 Gender Ratio

For testing our gender hypothesis, we used gender distribution data from Statistisches Bundesamt (2015)⁸. We computed the gender ratio by dividing the male population in a district by the female population. A gender ratio higher [lower] than 1 means there is an over-proportionate [under-proportionate] number of males in the district.

3.2.3 Population with Abitur

For our education hypothesis, we gathered data about the population with Abitur (high school degree) in a district from Statistische Ämter des Bundes und der Länder (2015)⁹. The variable describes the percentage of persons with a high school degree from all persons having left school.

3.2.4 GDP per Capita

The GDP per capita data stems from Statistische Ämter des Bundes und der Länder (2015)¹⁰. It is measured in 1000 Euros.

3.2.5 Unemployment Rate

We gathered the unemployment rate from Statistische Ämter des Bundes und der Länder (2015)¹¹. The employment rate is the share of registered unemployed in the labor force and therefore includes percentages.

3.2.6 Population Density

The data on the population density stems from Statistische Ämter des Bundes und der Länder (2015)¹². It is measured as inhabitants per square kilometer.

3.2.7 Refugee Ratio

The latest data on the distribution of refugees in the German districts dates from 12/31/2013 (Statistische Ämter des Bundes und der Länder 2015)¹³. However, as data from (Bundesamt für Migration und Flüchtlinge 2015) shows, the number of asylum seekers had nearly doubled from 2012 to 2013, already portraying a clear trend. We compute the number of refugees per 1000 citizens.

⁷Table 12411-0017, downloaded on 11/7/2015, specified time period: 12/31/2014.

⁸Table 12411-0017, downloaded on 11/7/2015, specified time period: 12/31/2014.

⁹Table AI003-2, downloaded on 11/8/2015, specified time period: 2013.

¹⁰Table AI-N-10, downloaded on 11/8/2015, specified time period: 2012.

¹¹Table 659-71-4, downloaded on 11/8/2015, specified time period: 2014.

¹²Table AI002-1, downloaded on 11/8/2015, specified time period: 2014.

¹³Table 661-31-4, downloaded on 11/8/2015, specified time period: 12/31/2013.

3.2.8 East West Dummy

We create a dummy variable distinguishing West from East German districts, with 1 being East and 0 being West German.

3.2.9 Voter Turnout

In order to control for “classical political participation” we include data on the voter turnout (in percent) at the last German federal election in 2013 from Statistische Ämter des Bundes und der Länder (2015)¹⁴.

After having imported the data sets into R, we cleaned them and created a series of new, more useful variables for our purpose. In a next step we merged all data sets (both independent and dependent variable) and provide some descriptive and inferential statistics in the next section.

4 Analysis

4.1 Descriptive Statistics

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Number of Refugee Initiatives per District	402	1.7	3.0	0	40
Old Age Dependency Ratio	402	0.3	0.05	0.2	0.5
Gender Ratio	402	1.0	0.03	0.9	1.0
Population with Abitur (percentage)	398	32.0	9.9	10.8	59.1
GDP per Capita in 1000 Euros	400	30.7	12.3	15.0	105.1
Unemployment Rate	400	6.3	2.9	1.4	15.4
Population Density	400	504.0	650.6	36.5	4,530.6
Refugee Ratio	400	2.5	1.1	0.2	9.1
East-West Dummy	402	0.2	0.4	0	1
Voter Turnout (percentage)	400	70.7	4.3	57.8	79.1

4.2 Shiny App

In addition to the descriptive statistics, we coded a Shiny web application (RStudio, Inc. 2015) and uploaded it to R Studio’s [shinyapps.io](https://www.shinyapps.io)¹⁵. The web app uses the `leaflet` package (Cheng and Xie 2015) to plot the initiatives on an openstreetmap and allows for coloring the districts according to the independent variables. The app however is not yet finished, most notably a legend is missing for the colors and the map fully reloads when the variables are changed. We plan on improving this for the final project. The map can be found at <https://christophercosler.shinyapps.io/RefugeeApp> and the code is on Git Hub in the Shiny folder.

4.3 Inferential Statistics

We already calculated some inferential statistics, but this should only be seen as first steps towards the final paper. As discussed above, the number of refugee initiatives per district can be seen as count data. Modeling count variables is a common task, not only social sciences (Zeileis, Kleiber, and Jackman 2007).

¹⁴Table 252-01-4, downloaded on 11/13/2015, specified time period: 9/22/2013.

¹⁵<https://www.shinyapps.io>

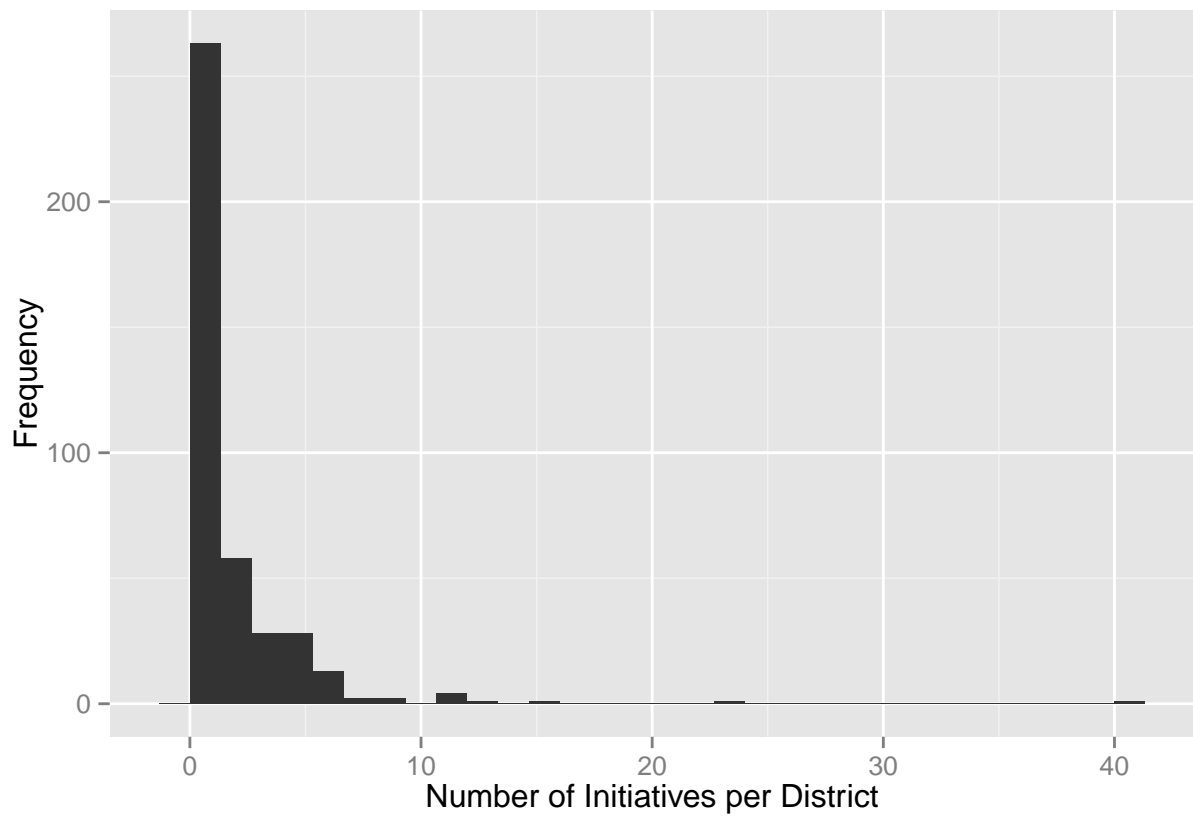


Figure 3: Histogram of Number of Initiatives per District

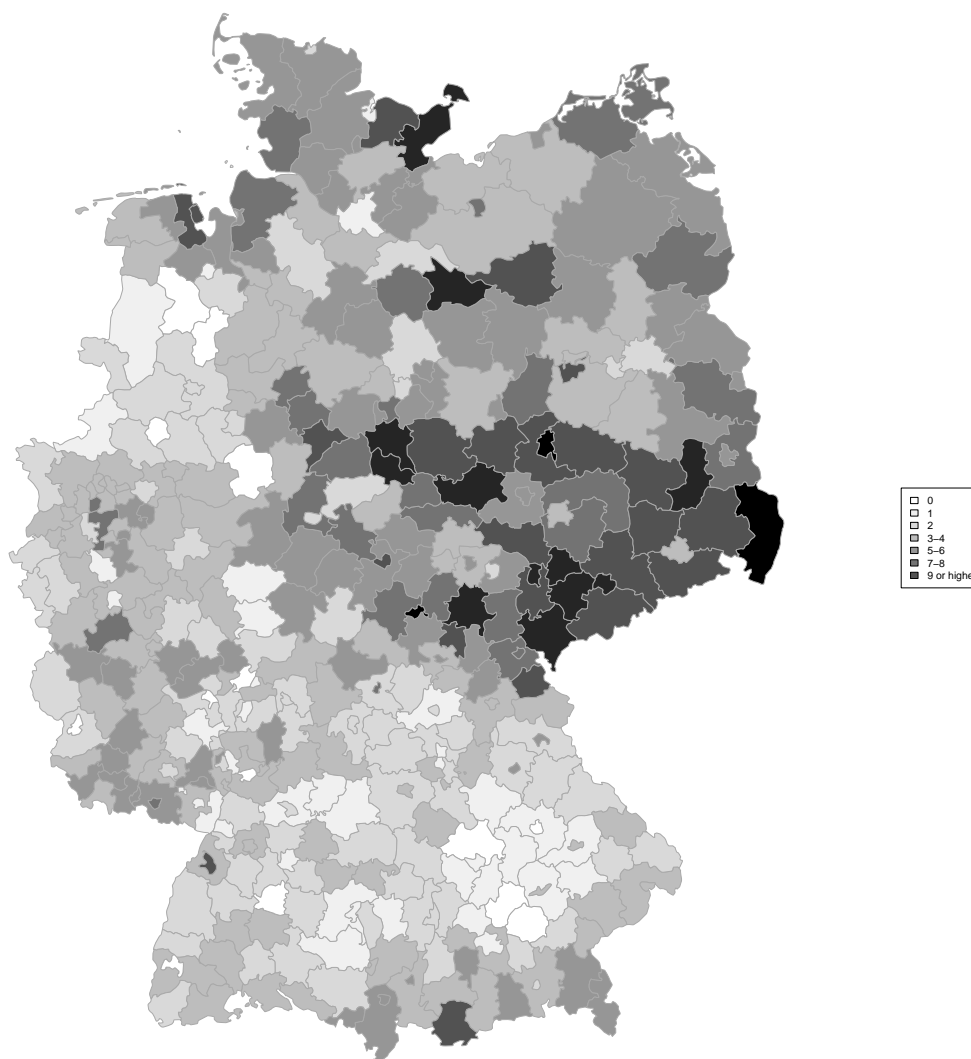


Figure 4: Old Age Dependency Ratio

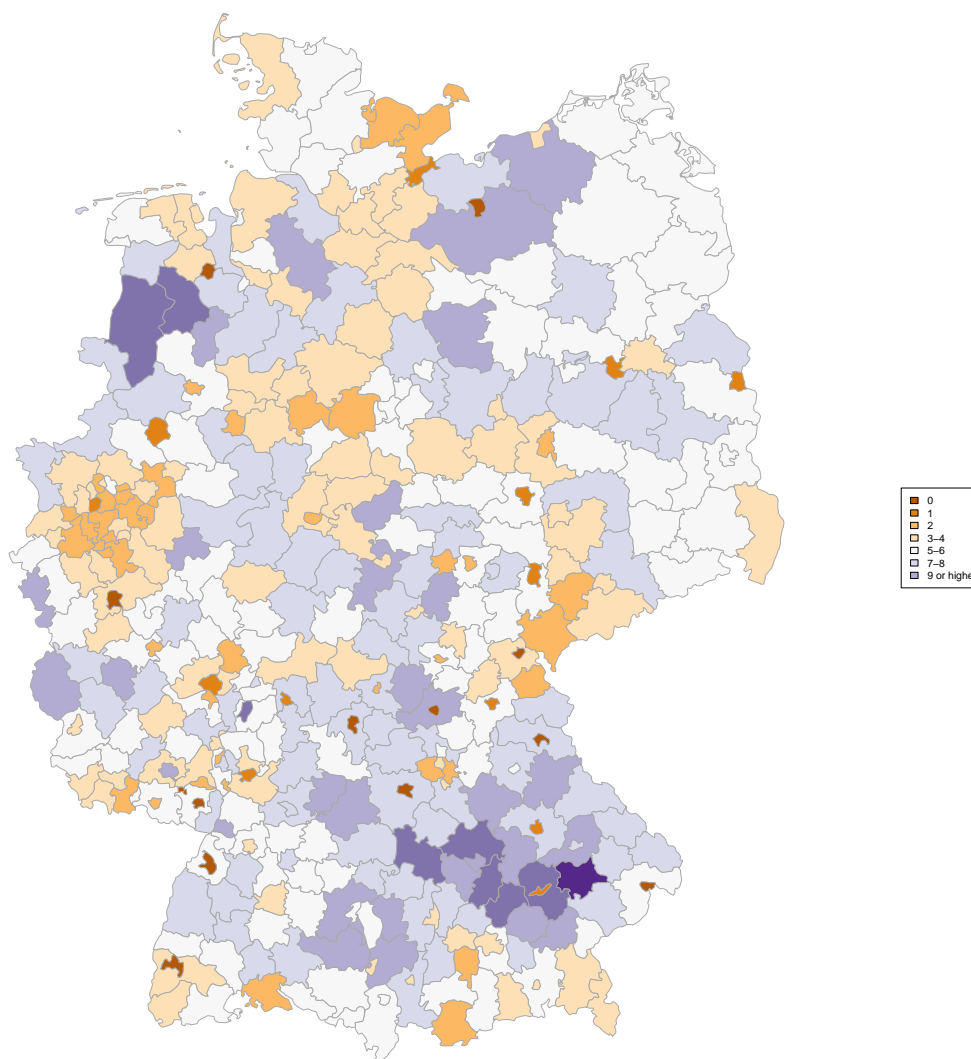


Figure 5: Gender Ratio

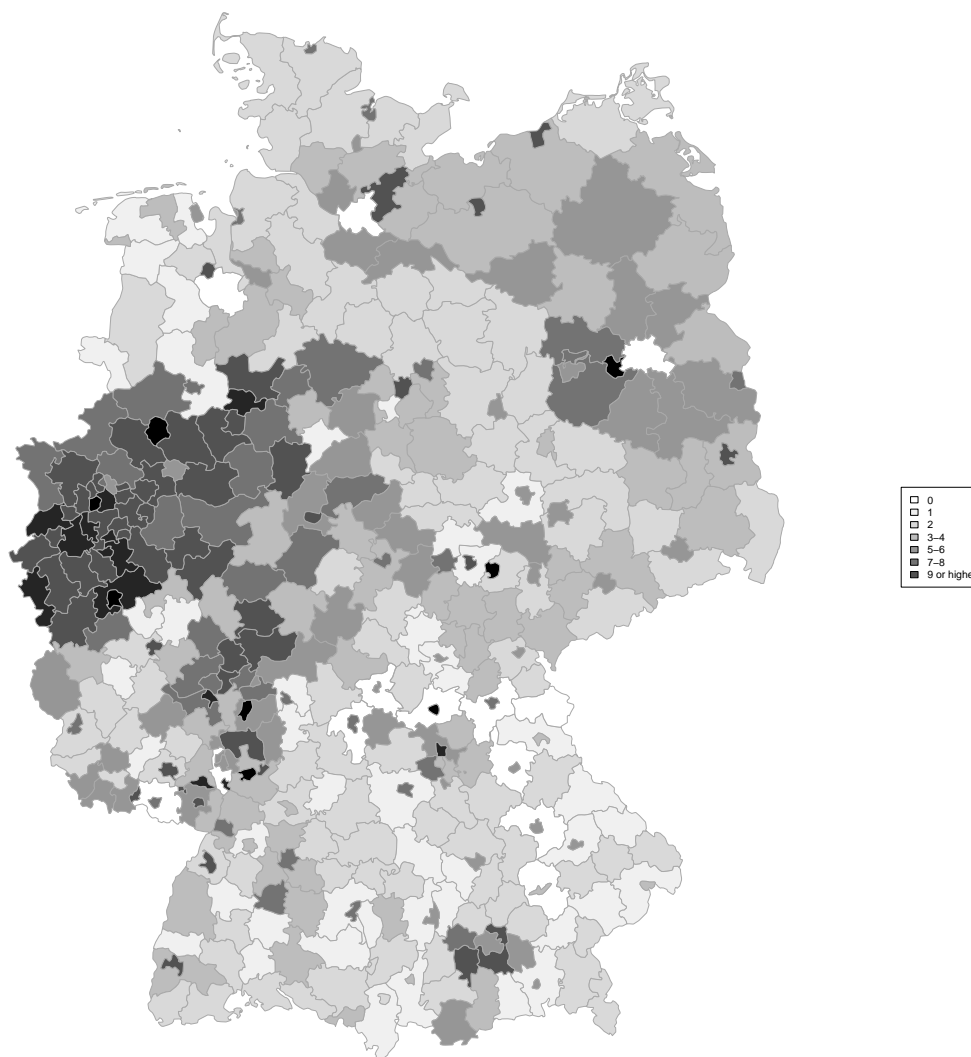


Figure 6: Population with Abitur (percentage)

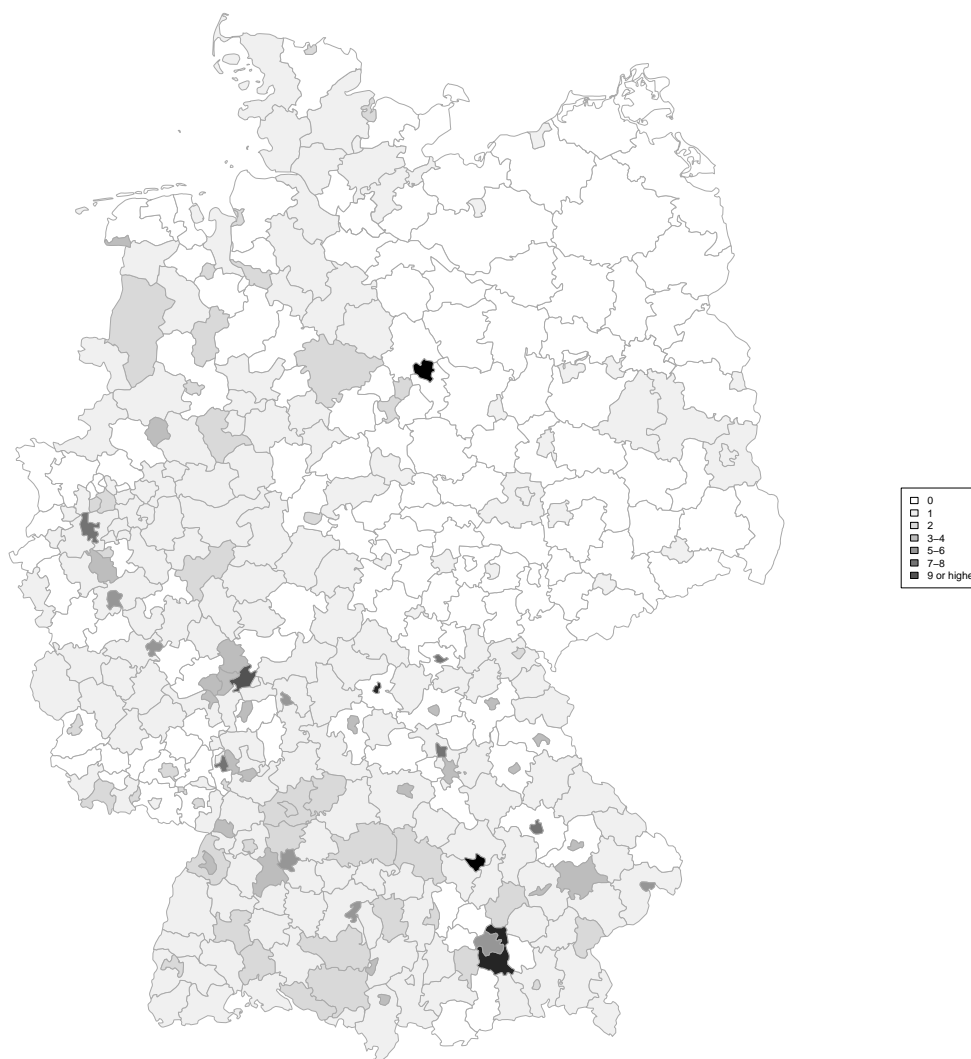


Figure 7: GDP per capita in 1000 Euros

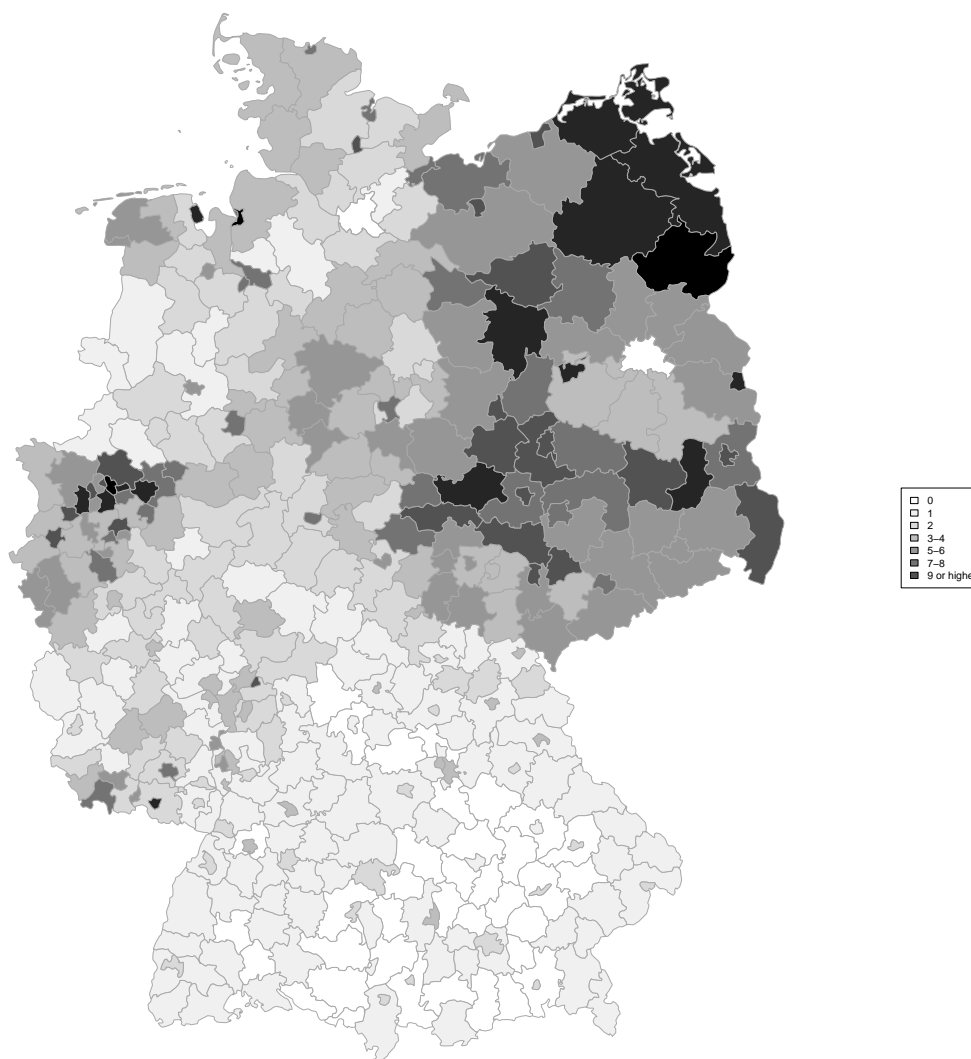


Figure 8: Unemployment Rate

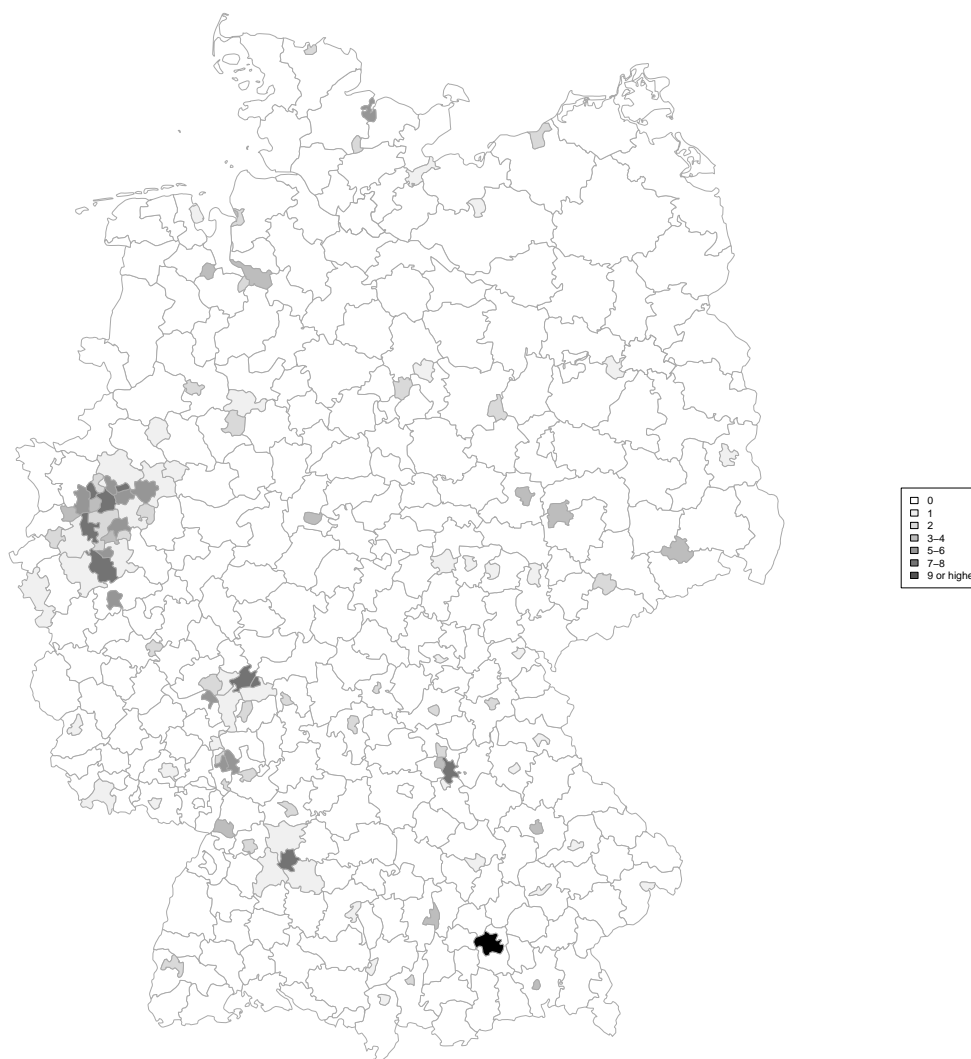


Figure 9: Population Density

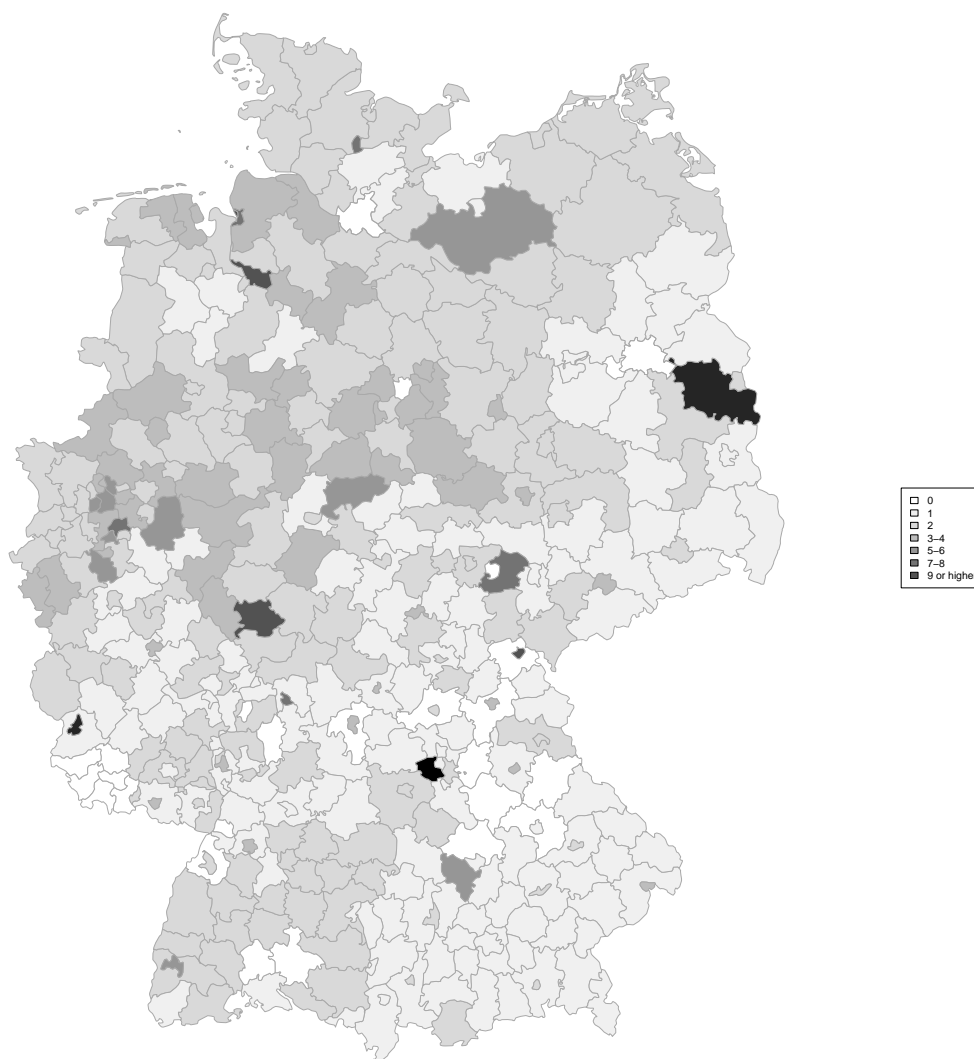


Figure 10: Refugee Ratio

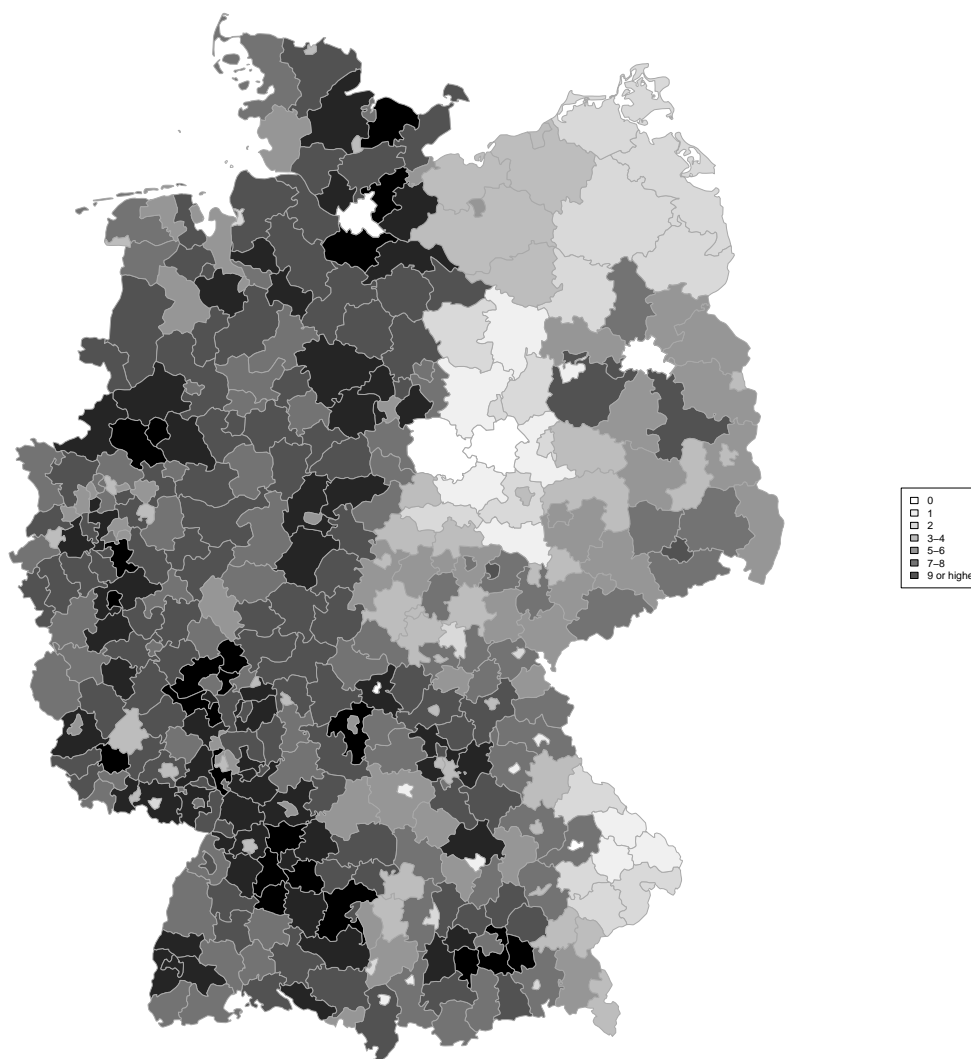


Figure 11: Voter turnout

Table 2: Poisson, Quasi-Poisson and Negative Binomial Regression Estimates for the number of Refugee Initiatives

	<i>Dependent variable:</i>		
	Number of refugee initiatives per district		
	Poisson	Quasi-Poisson	Negative Binomial
	(1)	(2)	(3)
Old Age Dependency Ratio	−5.85*** (1.24)	−5.85*** (1.66)	−7.51*** (1.74)
Gender Ratio	−3.55* (1.89)	−3.55 (2.54)	−5.56** (2.71)
Population with Abitur (percentage)	−0.01** (0.01)	−0.01 (0.01)	−0.01 (0.01)
GDP per Capita in 1000 Euros	0.01*** (0.004)	0.01*** (0.01)	0.01** (0.01)
Unemployment Rate	0.02 (0.03)	0.02 (0.04)	0.03 (0.04)
Population Density	0.0005*** (0.0001)	0.0005*** (0.0001)	0.0004*** (0.0001)
Number of Refugee Initiatives per District	0.06 (0.04)	0.06 (0.05)	0.06 (0.05)
East-West Dummy	0.94*** (0.14)	0.94*** (0.19)	1.12*** (0.19)
Voter Turnout (percentage)	0.11*** (0.01)	0.11*** (0.02)	0.12*** (0.02)
Constant	−3.12 (2.30)	−3.12 (3.10)	−1.68 (3.31)
Observations	398	398	398
Log Likelihood	−647.08		−613.99
θ			2.45*** (0.48)
Akaike Inf. Crit.	1,314.17		1,247.98

Note:

*p<0.1; **p<0.05; ***p<0.01

The classical method is a Poisson regression, which is limited though if over-dispersion or excess zeros are present (Cameron and Trivedi 2013). To deal with over-dispersion, sandwich covariances, quasi-Poisson models and especially negative binomial regressions are appropriate. If the data contains excess zeros, hurdle models or zero-inflated models should be used. There is however no reason for assuming to different data generating process in the given scenario, therefore models that account for excess zeros should not be chosen.

In the following, first a Poisson and a quasi-Poisson regression will be implemented using the `gml` function from the `stats` package (R Core Team 2015), afterwards a negative binomial regression with the `MASS` package (Ripley 2015).

As a first attempt to capture the relationship between the dependent variables and the predictors, we fit a standard Poisson regression model, the simplest distribution used for modeling count data (Zeileis, Kleiber, and Jackman 2007). The Age Dependency ratio, Gender Ratio, Education, GDP per capita, Population Density, the East-West dummy and the turnout are highly significant. These results might however be too optimistic, as over-dispersion is likely to be present. If we implement a test for overdispersion using `dispersiontest` from the (Kleiber and Zeileis 2015) package, based on (Cameron and Trivedi 2013), the result is $c = 0.8$, with $c > 0$ meaning that overdispersion is present, at a significance level smaller than 0.001.

As there are strong signs of over-dispersion, a quasi-Poisson regression might be a more appropriate method. The results however are almost identical, the only difference is that the gender ration is not significant at a level of 0.05 anymore. As a more formal way to account for over-dispersion in the data (Cameron and Trivedi 2013), a negative binomial regression is calculated. The regressor only slightly change, but the gender ratio is now highly significant again. The results will be interpreted in more depth in the actual research paper including predictions to meaningfully interpret the estimates, but we can already draw the conclusion that the type of model chosen does not strongly influence the results and the only unemployment and population density are potentially significant.

Bibliography

Bivand, Roger, Tim Keitt, and Barry Rowlingson. 2015. *Rgdal: Bindings for the Geospatial Data Abstraction Library*. <http://CRAN.R-project.org/package=rgdal>.

Böhnke, Petra. 2011. “Ungleiche Verteilung politischer Partizipation.” *Aus Politik Und Zeitgeschichte*, January, 18–25.

Bundesamt für Kartographie und Geodäsie. 2015. “Verwaltungsgebiete.” http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=19&gdz_user_id=0.

Bundesamt für Migration und Flüchtlinge. 2015. *Aktuelle Zahlen zu Asyl*. <https://www.bamf.de/SharedDocs/Anlagen/DE/Downloads/Infothek/Statistik/Asyl/statistik-anlage-teil-4-aktuelle-zahlen-zu-asyl.html>.

Cameron, A Colin, and Pravin K Trivedi. 2013. *Regression Analysis of Count Data*. Vol. 53. Cambridge university press.

Cheng, Joe, and Yihui Xie. 2015. *Leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. <http://CRAN.R-project.org/package=leaflet>.

Grothendieck, G. 2014. *Gsubfn: Utilities for Strings and Function Arguments*. <http://CRAN.R-project.org/package=gsubfn>.

Karakayali, Serhat, and J. Olaf Kleist. 2015. *EFA-Studie Strukturen und Motive der ehrenamtlichen Flüchtlingsarbeit (EFA) in Deutschland*. Berliner Institut für empirische Integrations- und Migrationsforschung.

Kleiber, Christian, and Achim Zeileis. 2015. *AER: Applied Econometrics with R*. <http://CRAN.R-project.org/package=AER>.

New York Post. 2015. “Refugees arrive to warm welcomes in Germany, Austria.” <http://nypost.com/2015/09/06/refugees-arrive-to-warm-welcomes-in-germany-austria/>.

- Pebesma, Edzer, and Roger Bivand. 2015. *Sp: Classes and Methods for Spatial Data*. <http://CRAN.R-project.org/package=sp>.
- Pro Asyl. 2015. “Mitmachen.” <http://www.proasyl.de/de/ueber-uns/foerderverein/mitmachen/>.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ripley, Brian. 2015. *MASS: Support Functions and Datasets for Venables and Ripley’s MASS*. <http://CRAN.R-project.org/package=MASS>.
- RStudio, Inc. 2015. *Shiny: Web Application Framework for R*. <http://CRAN.R-project.org/package=shiny>.
- Spiegel Online,. 2015. “Umfrage: Fast die Hälfte der Deutschen beteiligt sich an Flüchtlingshilfe.” <http://www.spiegel.de/wirtschaft/soziales/umfrage-44-prozent-der-deutschen-beteiligen-sich-an-fluechtlingshilfe-a-1056902.html>.
- Statistische Ämter des Bundes und der Länder. 2015. “Regionaldatenbank Deutschland.” <https://www.regionalstatistik.de/genesis/online>.
- Statistisches Bundesamt. 2015. “GENESIS-Online Datenbank.” <https://www-genesis.destatis.de/genesis/online>.
- Wickham, Hadley. 2015. *Httr: Tools for Working with URLs and HTTP*. <http://CRAN.R-project.org/package=httr>.
- Zeileis, Achim, Christian Kleiber, and Simon Jackman. 2007. “Regression Models for Count Data in R.” Department of Statistics; Mathematics, WU Vienna University of Economics; Business.