

Visualisierung von Zensusergebnissen mit der Statistiksoftware R – eine Kurzeinführung

Erste Ergebnisse des Zensus 2011 sind seit dem 31.05.2013 publiziert. Für die Öffentlichkeit sind ausgewählte Ergebnisse in der frei verfügbaren Zensusdatenbank¹⁾ unter <https://ergebnisse.zensus2011.de> zugänglich. Nutzerinnen und Nutzer können dort auf umfangreiches Datenmaterial für fachliche und regionalisierte Auswertungen zurückgreifen. Zum zweiten Veröffentlichungstermin im Frühjahr 2014 wird dort u.a. ein umfangreicherer Datenbestand zur Visualisierung von Zensusergebnissen, insbesondere auch der Darstellung von interaktiven Karten, angeboten.

Für eigene Datenauswertungen kann beispielsweise die Statistiksoftware R genutzt werden, mit deren Hilfe in kurzer Zeit statistische Daten visualisiert werden können. Diese Möglichkeit wird im vorliegenden Beitrag vorgestellt.

Nutzung der Statistiksoftware R

R ist eine kostenfreie Programmiersprache für statistisches Rechnen und in der Wissenschaft aufgrund der zahlreichen Möglichkeiten bei der Erstellung statistischer Grafiken weit verbreitet. R ist in Anlehnung an die Programmiersprache S²⁾ entstanden. Im kommerziellen und wissenschaftlichen Bereich setzt sich R zunehmend als statistische Standardsprache durch.³⁾

Der Funktionsumfang von R kann – neben den in der Basisversion zur Verfügung gestellten Standardpaketen – durch weitere, ebenfalls kostenlose Pakete erweitert und

- 1) Vgl. Rehm, M., Rehmer, C. Zensus: 2011 – Die Zensusdatenbank 2011, in: Statistische Monatshefte Niedersachsen, Heft 6 (2013), S. 314-318.
- 2) [http://de.wikipedia.org/wiki/S_\(Statistiksprache\)](http://de.wikipedia.org/wiki/S_(Statistiksprache)).
- 3) [http://de.wikipedia.org/wiki/R_\(Programmiersprache\)](http://de.wikipedia.org/wiki/R_(Programmiersprache)).

an die spezifischen Nutzerwünsche angepasst werden. Zentrales Archiv für diese Pakete ist das Comprehensive R Archive Network (CRAN). Dem Nutzer steht somit ein mächtiges sowie kostenloses Auswertungsprogramm zur Verfügung⁴⁾.

Umfangreiches Datenangebot zum Zensus 2011 zur Auswertung mit Statistiksoftware

Das Statistische Bundesamt hat im Rahmen der Pressekonferenz zum Zensus am 31.05.2013 ausgewählte Ergebnisse⁵⁾ für demografische Merkmale sowie Merkmale zur Wohnsituation für die regionalen Einheiten Gemeinde, Verwaltungseinheit⁶⁾, Landkreis / kreisfreie Stadt und Land im MS-Excel-Format bereitgestellt. Das Angebot richtet sich vorwiegend an Journalistinnen und Journalisten sowie die Wissenschaft.

Auswahl der Daten und der Kartengrundlage zur Weiterverarbeitung in R

Nach dem Download der ausgewählten Ergebnisse⁷⁾ sollte das die Nutzerinnen und Nutzer interessierende Tabellenblatt vom MS-Excel-Format zur besseren Weiterverarbeitung in eine plattformunabhängige CSV-Datei umgewan-

- 4) Weitere Informationen sowie eine Downloadmöglichkeit der Software befinden sich unter www.r-project.org.
- 5) https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2013/Zensus2011/zensus_pk.html#Wechsel%20zur%20%C3%9Cbersichtsseite%20der%20Pressekonferenz.
- 6) In Niedersachsen: Einheits- und Samtgemeinden.
- 7) https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2013/Zensus2011/zensus_pk.html#Wechsel%20zur%20%C3%9Cbersichtsseite%20der%20Pressekonferenz.

1. Ausgewählte Zensusdaten im CSV-Format

D7	A	B	C	D	E	F	G	H	I
1	SATZART	AGS	LAND	RB	KREIS	BEV	ERWERBST	ERWERBSL	NERWERB
2		0	Bund	Gesamt		79652370	39985940	2140860	37525570
30	40	3156	Niedersachsen	Braunschweig	Northeim	136790	63490	3810	69490
31	40	3156	Niedersachsen	Braunschweig	Osterode am Harz	76160	34240	2220	39690
32	40	3157	Niedersachsen	Braunschweig	Peine	129670	61970	3300	84400
33	40	3158	Niedersachsen	Braunschweig	Wolfenbüttel	119040	56860	2960	59220
34	20	32	Niedersachsen	Hannover	Gesamt	2079440	1017480	58280	1003690
35	40	3241	Niedersachsen	Hannover	Region Hannover	1094760	542890	32290	519580
36	40	3251	Niedersachsen	Hannover	Diepholz	210440	108960	3630	97850
37	40	3252	Niedersachsen	Hannover	Hameln-Pyrmont	148510	69970	4370	74180
38	40	3254	Niedersachsen	Hannover	Hildesheim	274830	130220	8490	136120
39	40	3255	Niedersachsen	Hannover	Holzminde	73120	32160	2330	36630
40	40	3256	Niedersachsen	Hannover	Nienburg (Weser)	121580	59120	3010	59450
41	40	3257	Niedersachsen	Hannover	Schaumburg	156200	74150	4160	77890
42	20	33	Niedersachsen	Lüneburg	Gesamt	1656020	822410	37890	795720
43	40	3351	Niedersachsen	Lüneburg	Celle	174550	84400	4810	85340
44	40	3352	Niedersachsen	Lüneburg	Cuxhaven	198100	95100	4110	98890
45	40	3353	Niedersachsen	Lüneburg	Harburg	237910	122280	5620	110010
46	40	3354	Niedersachsen	Lüneburg	Lüchow-Dannenberg	48720	21630	1040	26060
47	40	3355	Niedersachsen	Lüneburg	Lüneburg	171820	84900	4130	82790
48	40	3356	Niedersachsen	Lüneburg	Osterholz	110020	57340	2340	50340

2. Shapefile im Format WGS84 zum Download

Name	Typ	Gepack...	K...	Größe	Ve...	Datum
VG250_Verwaltungsgemeinschaften.shx	SHX-Datei	27 KB	Nein	48 KB	45%	07.05.2013 08:56
VG250_Bundeslaender.dbf	DBF-Datei	2 KB	Nein	7 KB	78%	18.04.2013 11:40
VG250_Bundeslaender.prj	PRJ-Datei	1 KB	Nein	1 KB	17%	18.04.2013 11:40
VG250_Bundeslaender.shp	SHP-Datei	2.366 KB	Nein	3.07...	24%	18.04.2013 11:40
VG250_Bundeslaender.shx	SHX-Datei	1 KB	Nein	1 KB	19%	18.04.2013 11:40
VG250_Gemeinden.dbf	DBF-Datei	731 KB	Nein	3.52...	80%	18.04.2013 11:41
VG250_Gemeinden.prj	PRJ-Datei	1 KB	Nein	1 KB	17%	18.04.2013 11:41
VG250_Gemeinden.shp	SHP-Datei	23.211...	Nein	30.4...	24%	18.04.2013 11:41
VG250_Gemeinden.shx	SHX-Datei	65 KB	Nein	96 KB	33%	18.04.2013 11:41
VG250_Kreise.dbf	DBF-Datei	28 KB	Nein	134 KB	80%	18.04.2013 11:42
VG250_Kreise.prj	PRJ-Datei	1 KB	Nein	1 KB	17%	18.04.2013 11:42
VG250_Kreise.shp	SHP-Datei	6.920 KB	Nein	9.03...	24%	18.04.2013 11:42
VG250_Kreise.shx	SHX-Datei	3 KB	Nein	4 KB	24%	18.04.2013 11:42
VG250_Verwaltungsgemeinschaften.dbf	DBF-Datei	301 KB	Nein	1.47...	80%	16.05.2013 08:26
VG250_Verwaltungsgemeinschaften.prj	PRJ-Datei	1 KB	Nein	1 KB	17%	06.05.2013 14:43
VG250_Verwaltungsgemeinschaften.shp	SHP-Datei	17.549...	Nein	22.6...	23%	07.05.2013 08:56

delt werden.⁸⁾ Für die Weiterverarbeitung in R werden, neben den bereits vorinstallierten Paketen, die zusätzlichen R Pakete „mapproj“, „rgdal“, „maps“, „RcolorBrewer“ sowie „Shapefiles“ benötigt.⁹⁾

Kartengrundlagen Zensus 2011

Das Bundesamt für Geodäsie und Kartographie hat unter www.zensus2011.de Kartengrundlagen als Shapefile¹⁰⁾ für interessierte Nutzerinnen und Nutzer bereitgestellt (siehe Abb. 2).¹¹⁾

In der Regel waren bisher die Nutzung und die Weiterverarbeitung eines solchen Kartenmaterials Personen und Institutionen mit entsprechendem wissenschaftlich-geografischem „Know-How“ sowie einer Ausstattung mit kostenpflichtiger Software für die Erstellung von Landkartengrafik vorbehalten. Die Open-Data Kampagnen der Bun-

- 8) In MS-Excel oder anderen Tabellenkalkulationsprogrammen beim Speichern über die angebotene Auswahl des Dateityps.
 9) Der Funktionsumfang der Pakete wird unter www.r-project.org näher erläutert.
 10) <http://de.wikipedia.org/wiki/Shapefile>.
 11) https://www.zensus2011.de/SharedDocs/Downloads/DE/Shapefile/VG250_1Jan2011_WGS84.zip?__blob=publicationFile&v=18.

desregierung^{12),13)} sowie die Verfügbarkeit von kostenlosen Programmen zur Weiterverarbeitung ermöglichen immer mehr Nutzern einen Zugang zu den Möglichkeiten der grafischen Visualisierung von Kartenmaterial bis zur regionalen Einheit der Gemeinde.

Programmcode für eine Verarbeitung in R

Im Folgenden wird eine Möglichkeit der Kartenerstellung in Form von R-Code anhand von einer Karte auf der Ebene Landkreis / kreisfreie Stadt dargestellt. Es wird das Format WGS84¹⁴⁾ verwendet. Mit Hilfe dieses Programmcodes, welcher schnell und flexibel an den Bedarf der Nutzerinnen und Nutzer angepasst werden kann, lässt sich innerhalb kürzester Laufzeit die in Abbildung 4 dargestellte Karte erzeugen. Alle mit dem #-Symbol eingeleiteten Zeilen dienen der Beschreibung des programmierten Schrittes und sind nicht Bestandteil des Programms im engeren Sinne.

12) <http://www.geodatenzentrum.de/docpdf/geonutzv.pdf>.

13) <https://www.govdata.de/>.

14) Im Folgenden wurde das Format WGS84 verwendet. Analog kann auch das Format UTM32 verwendet werden. Bei der Einzeichnung von Städten und Legendengenen sind dann jedoch die UTM32 Koordinaten zu verwenden.

3. Programmcode R zur Kartenerstellung

```
# Download der Pakete und deren Aufruf in R
install.packages("mapproj")
install.packages("rgdal")
install.packages("maps")
install.packages("RColorBrewer")
install.packages("shapefiles")
library(mapproj)
library(rgdal)
library(maps)
library(RColorBrewer)
library(shapefiles)

# Einlesen der Daten: alle Geodaten der .zip Datei müssen im gleichen Ordner vorhanden sein
Kreise <- readShapeSpatial("C:/Program Files/R/VG250_Kreise")
unique(Kreise$AGS)

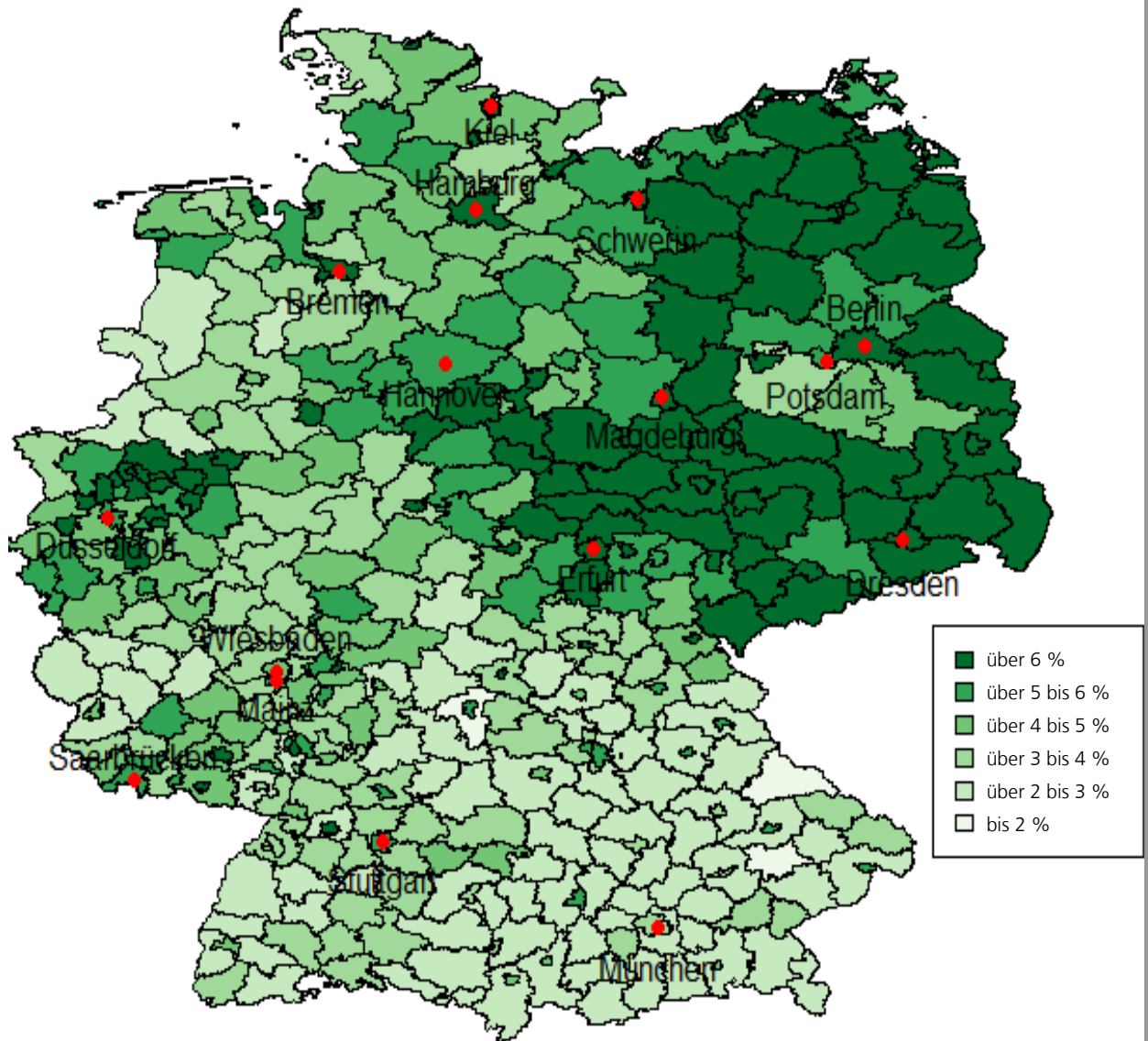
# Einlesen der Zensusdaten und filtern auf Kreisebene (Satzart=40, formatieren des AGS)
zensus=read.csv2("C:/Program Files/R/zensus.csv") #Muss vom Nutzer angepasst werden
zensus2=data.frame(zensus[zensus$SATZART==40,])
zensus2$AGS=formatC(zensus2$AGS, width = 5, format = "d", flag = "0")
names(zensus2)[2]="RS"
```

Noch: 3. Programmcode R zur Kartenerstellung

```
#Zusammenführen der Zensusdaten und der Daten für Geodäsie über den AGS
Zusammen=merge(Kreise@data,zensus2, by= "RS",all.x=TRUE)
Zusammen=Zusammen[order(Zusammen$OBJECTID), ]
# Auswahl der Kombination Anteil der Erwerbslosen an den Erwerbstätigen
plotvar <- (Zusammen$ERWERBSL)/(Zusammen$ERWERBST+ Zusammen$ERWERBSL)
nclr <- 6 # sechs Kategorien vergeben
plotclr <- brewer.pal(nclr, "Greens") # Blues, Reds...
plotvar2=c(1:439)
plotvar2[plotvar<=0.02]=1
plotvar2[plotvar<=0.03 & plotvar>0.02]=2
plotvar2[plotvar<=0.04 & plotvar>0.03]=3
plotvar2[plotvar<=0.05 & plotvar>0.04]=4
plotvar2[plotvar<=0.06 & plotvar>0.05]=5
plotvar2[plotvar>0.06]=6
colormum <- cut(plotvar2, nclr, labels=FALSE)
colorcode <- plotclr[colormum]
#####Legenden Formatieren
legende=c(
"über 6%",
"über 5 bis 6%",
"über 4 bis 5%",
"über 3 bis 4%",
"über 2 bis 3%",
"bis 2%")
##### Erstellen der Grafik
plot(Kreise,axes=FALSE,border="black",col=colorcode)
##### Legende einfügen
a=names(table(colorcode))
legend(14.004954, 50.400007, as.character(legende),
fill=a, xpd=T, inset=c(-.12, .1), cex = 0.6 )
title(main="Erwerbslosenquote Mai 2011")
# Geokoordinaten für Städte und Legenden. Diese sind auf diversen Plattformen im Internet
#abrufbar, im vorliegenden Programm von: http://www.deine-berge.de/Rechner/Koordinaten.
#Für Karten im UTM32 Format müssen die Koordinaten entsprechend umgerechnet werden.
points(13.404954, 52.520007, pch = 19, col = "red")
text(13.404954, 52.820007, "Berlin", cex = 1.0, col = "black")
points(9.181759,48.775418, pch = 19, col = "red")
text(9.181759,48.475418, "Stuttgart", cex = 1.0, col = "black")
points(11.581981,48.135125, pch = 19, col = "red")
text(11.581981,47.835125, "München", cex = 1.0, col = "black")
points(13.73726, 51.050409, pch = 19, col = "red")
text(13.73726, 50.750409, "Dresden", cex = 1.0, col = "black")
points(11.029880, 50.984768, pch = 19, col = "red")
text(11.029880, 50.784768, "Erfurt", cex = 1.0, col = "black")
points(11.628878, 52.130807, pch = 19, col = "red")
text(11.628878, 51.830807, "Magdeburg", cex = 1.0, col = "black")
points(13.072691, 52.394189, pch = 19, col = "red")
text(13.072691, 52.154189, "Potsdam", cex = 1.0, col = "black")
points(11.401250, 53.635502, pch = 19, col = "red")
text(11.401250, 53.335502, "Schwerin", cex = 1.0, col = "black")
points(9.993682, 53.551085, pch = 19, col = "red")
text(9.993682, 53.751085, "Hamburg", cex = 1.0, col = "black")
points(10.122765, 54.323293, pch = 19, col = "red")
text(10.122765, 54.153293, "Kiel", cex = 1.0, col = "black")
points(8.801694, 53.079296, pch = 19, col = "red")
text(8.801694, 52.879296, "Bremen", cex = 1.0, col = "black")
points(9.732010, 52.375892, pch = 19, col = "red")
text(9.732010, 52.175892, "Hannover", cex = 1.0, col = "black")
points(6.773456, 51.227741, pch = 19, col = "red")
text(6.773456, 51.027741, "Düsseldorf", cex = 1.0, col = "black")
points(8.243344, 50.063080, pch = 19, col = "red")
text(8.243344, 50.33080, "Wiesbaden", cex = 1.0, col = "black")
points(8.247253, 49.992862, pch = 19, col = "red")
text(8.247253, 49.792862, "Mainz", cex = 1.0, col = "black")
points(6.996933, 49.240157, pch = 19, col = "red")
text(6.996933, 49.440157, "Saarbrücken", cex = 1.0, col = "black")
```

4. Erwerbslosenquote¹⁵⁾ Landkreis / kreisfreie Stadt

– Erwerbslosenquote Mai 2011 –



15) Die Definition der Ausprägungen richtet sich nach dem Labour-Force-Konzept der International-Labour-Organisation (ILO-Konzept). Nach dieser Definition gilt als erwerbstätig, wer in dem einwöchigen Berichtszeitraum (9. bis 15. Mai 2011) mindestens eine Stunde pro Woche gegen Bezahlung gearbeitet hat und mindestens 15 Jahre alt ist. Zu den Erwerbslosen zählen alle Personen im Alter von mindestens 15 Jahren, die in dem einwöchigen Berichtszeitraum (9. bis 15. Mai 2011) nicht erwerbstätig waren, jedoch in den letzten vier Wochen vor der Befragung aktiv nach einer Tätigkeit gesucht haben (im Unterschied zu Nicht-Erwerbspersonen). Berechnung Erwerbslosenquote: $\text{Erwerbslose} / (\text{Erwerbstätige} + \text{Erwerbslose})$. Weitere Informationen finden sich im Internet unter: https://www.zensus2011.de/SharedDocs/Aktuelles/Vielfaeltiges%20Deutschland_Ergebnisse_zur_Erwerbstaetigkeit.html.