

## SOEP Survey Papers

Series D – Variable Descriptions and Coding

SOEP – The German Socio-Economic Panel study at DIW Berlin

2017

# Coding of Free-Text Answers on the Most Important Political Problem in Germany – Methods of Coding and Documentation of Results

Tabea Naujoks, Wiebke Nestler, Martin Brümmer

Running since 1984, the German Socio-Economic Panel study (SOEP) is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin.

The aim of the SOEP Survey Papers Series is to thoroughly document the survey's data collection and data processing. The SOEP Survey Papers is comprised of the following series:

**Series A** – Survey Instruments (Erhebungsinstrumente)

**Series B** – Survey Reports (Methodenberichte)

**Series C** – Data Documentation (Datendokumentationen)

**Series D** – Variable Descriptions and Coding

**Series E** – SOEPmonitors

**Series F** – SOEP Newsletters

**Series G** – General Issues and Teaching Materials

The SOEP Survey Papers are available at <http://www.diw.de/soepsurveypapers>

**Editors:**

Dr. Jan Goebel, DIW Berlin

Prof. Dr. Martin Kroh, DIW Berlin and Humboldt Universität Berlin

Prof. Dr. Carsten Schröder, DIW Berlin and Freie Universität Berlin

Prof. Dr. Jürgen Schupp, DIW Berlin and Freie Universität Berlin

Please cite this paper as follows:

Tabea Naujoks, Wiebke Nestler, Martin Brümmer. 2017. Coding of Free-Text Answers on the Most Important Political Problem in Germany - Methods of Coding and Documentation of Results. SOEP Survey Papers 476: Series D. Berlin: DIW/SOEP



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

© 2017 by SOEP

ISSN: 2193-5580 (online)

DIW Berlin

German Socio-Economic Panel (SOEP)

Mohrenstr. 58

10117 Berlin

Germany

[soeppapers@diw.de](mailto:soeppapers@diw.de)

# **Coding of Free-Text Answers on the Most Important Political Problem in Germany – Methods of Coding and Documentation of Results**

**Tabea Naujoks, Wiebke Nestler, Martin Brümmer**

Berlin, 2017

## 0 Introduction

What is the most important political problem? In both 2015 and 2017, survey respondents (more than 1,000 individuals answered this open-ended question. This paper documents the subsequent coding process of the resulting free-text answers. This documentation might be helpful for coding the answers in similar surveys in the future.

In the first part, the data used will be described. In the second part, the coding scheme as well as its development will be presented. The coding process as well as the coding tool used will form the subject of the third section. Results of the coding in the form of descriptive frequencies will be presented in the fourth part of the paper. In addition, various possibilities of evaluation are discussed and compared. In the concluding discussion section, we will once again consider the specific features of the coding process which should be taken into account for further usage of the resulting data set.

## 1 Data

### 1.1 Data collection

In January 2017, a representative telephone survey on the relevance of selected political objectives in 1,016 eligible voters was carried out by Kantar Public.<sup>1</sup> The main objective of the survey was to assess the importance of different policy objectives. In addition to a list of predefined policy objectives, respondents were also asked what they considered to be ‘the most important political problem’ at the moment. The interviewees were able to name keywords on the telephone, which were recorded by the interviewers.

In addition, a small part of the respondents of the German Socio-Economic Panel Study (SOEP) who participated in a citizens’ dialogue with the chancellor (or belonged to the control group) answered the same question in 2015, both before as well as after the dialogue. All in all, 289 answers from 216 unique respondents are available.<sup>2</sup>

The data set discussed here pools all text answers from 2015 and 2017, resulting in a sample of 1,305 responses.

### 1.2 Data characteristics

At the first survey in 2015, an average of 21.1 characters (range: 4-97) and 2.3 words (range: 1-17) were recorded; at the follow-up 2015 an average of 15.7 characters (range: 4-55) and 1.5 words (range: 1-9). In the 2017 survey, an average of 27.1 characters (range: 3-214) and 3.9 words (range: 1-38) were used.<sup>3</sup>

These differences in the number of recorded characters and words likely reflect differences in the willingness of the telephone interviewers 2015 and 2017 to record more or less words.

\* These two authors contributed equally to the work.

<sup>1</sup> Giesselmann, M., Brümmer, M., Kroh, M., Siegel, N. A., & Wagner, G. G. (2017). Fluchtwanderung ganz oben auf der Liste der dringenden politischen Prioritäten. *Wirtschaftsdienst*, 97(3), 192-200.

<sup>2</sup> Wagner, G. G., Bruemmer, M., Glemser, A., Rohrer, J. M., & Schupp, J. (2017). Dimensions of Quality of Life in Germany: Measured by Plain Text Responses in a Representative Survey (SOEP). SOEPPaper No. 893, Berlin.

<sup>3</sup> Giesselmann, M., Brümmer, M., Kroh, M., Siegel, N. A., & Wagner, G. G. (2017). Fluchtwanderung ganz oben auf der Liste der dringenden politischen Prioritäten. *Wirtschaftsdienst*, 97(3), 192-200.

## 2 Development of the coding scheme

The coding scheme was developed prior to the coding process by Tabea Naujoks and Wiebke Nestler. A list of categories was collected based on theoretical considerations and summarized in a smaller number of supercategories. This first coding scheme was examined using a small subsample of the respondents' answers and was afterwards modified accordingly. The modification process involved the renaming and exchange of categories, supercategories, as well as the change of descriptions of single categories wherever the existing coding scheme was not able to represent the respondents' answers adequately.

For a summarising list of the final supercategories, categories and their descriptions please refer to Table 1.

Supercategory	Category	Description
A	1 refugees ambiguous	Statements that contain the words 'refugees', 'foreigners', '(im)migrants' or 'asylum seeker' and no clear value judgement. This also includes statements, which do not refer without doubt in a positive or negative way to refugees.
	2 refugees negative	Statements which show that the interviewed person considers the existence of refugees in Germany a problem. The respondent e.g. demands a stricter control of refugees, a limitation of refugees or their deportation.
	3 refugees positive	Statements are classified as 'refugees positive' if respondents e.g. criticise the treatment of refugees or demand a better support for refugees.
B	4 security	Statements in this category deal with crime, desire for security or directly mention inner security.
	5 peace	Respondents mention the desire for peace.
	6 terrorism	This category summarizes statements concerning terrorism, assassinations and assaults.
C	7 party political dysfunction	Statements that consider the upcoming election, the party system, democratic deficit, political division of society, citizens' communication among themselves or individual politicians a problem are included in this category.
	8 swing to the right	This category contains statements which consider the existence of right-wing parties (AfD, NPD), right-wing populism, xenophobia and right-wing mindset in general a political problem.
	9 swing to the left	In this category, statements on the subject of left-wing extremism are
D	10 environment	Topics such as energy transformation, environmental protection are included in the 'environment' category.
E	11 economy	Statements on the stabilization of the economy, inflation or government debt are summarized in this category.
F	12 foreign relations	Statements concerning relationships with other countries and foreign policy.
	13 EU	Statements concerning the EU and present EU relevant problems.
G	14 social justice	Statements on social (in)justice, inequality and the distribution of goods and income / assets in Germany as well as tax questions are summarized. Problems such as poverty, housing shortages and inequalities on the labor market are also coded as 'social justice'.
	15 education / children / family	Statements on the education system, school or family policy or childcare.
	16 age / pension	Problems such as old-age provision, pensions (equality) and age poverty are summarized here.
H	17 health	Statements on health care.
	18 internet and data protection	Statements that concern data protection or the internet.
I	19 other	Statements that are unidentifiable, that can not be clearly assigned to one category or that mention a different political problem belong to this category.
	20 no answer	Respondents that do not answer the question, state that they are not able to answer the question or deny answering the question.
J	21 too many problems	Respondents that state that there are too many problems to decide on one single problem.
	22 no problems	Respondents that state that there are no problems and that everything is good the way it is.

In the following paragraphs, we would like to address some central decisions of the final version of the coding scheme and its categories that were made during the development phase of the scheme.

First of all, a relatively broad definition of the term 'refugees' provides the basis for the category 'refugees ambiguous', 'refugees negative' and 'refugees positive'. Individuals who fall into these categories are the following: asylum seekers, foreigners, immigrants, refugees, immigrants and migrants. Since a

differentiated use of those more precise and accurate terms could not be noticed and guaranteed in the respondents' answers, a differentiation in the coding scheme did not seem appropriate.

The supercategory 'refugees' was structured in three different subcategories, one of them covering all negative comments on refugees in Germany ('refugees negative') and one reflecting a positive attitude towards refugees in Germany ('refugees positive'). A third category was introduced to summarize all comments that were not clearly identifiable as either negative or positive ('refugees ambiguous'). This third category was chosen whenever the short answer given by a respondent did not allow a clear assessment of the interviewees attitude towards refugees. This also included statements that are connoted with a clear evaluative judgement (like 'refugees flood') in everyday use.

For statements like the example 'refugees flood' we decided against a coding based on the everyday connotation of the given answer, since one could not tell whether the connotation was intended by the respondent or not and was therefore largely in the eye of the coder and dependent on her sensitivity to handling of language in this context. Additionally, one could not tell if the possibly negative evaluative judgement implied by the statement referred to the refugees themselves or the situation in general. That is why we tried to minimize the room for interpretation of the coders by suggesting a conservative way of coding for answers addressing the topic of migration and refugees whenever there was doubt about the evaluative judgement contained in a statement. This approach was chosen to reduce the bias of codings, leading to an ambiguous category that contains a broad variety of statements. Therefore, we would like to point out, that although this third category was formerly named and conceptualized as a neutral category, it should not be treated as neutral.

### **3 Coding instruction and process of coding**

#### **3.1 Coding Instructions and Process of Coding**

The respondents' answers were coded by three independent coders, Laura Lükemann (Universität Bielefeld), Tabea Naujoks (Freie Universität Berlin) and Wiebke Nestler (Universität Leipzig).

Although the interviewees were asked for *the* most important political problem, some of them reported more than one political problem. To capture the problems reported as granular as possible, we decided to code up to five entries of the mentioned political problems (no respondent reported more than five entries). The coders had the instruction to name the number of problems, and afterwards to separately encode the different entries. If an interviewee mentioned several entries, coders were instructed to separately assign them to the corresponding categories.

The following example is meant to clarify this process of coding: The entry 'Refugee crisis, terrorism, contact with Turkey' covers three political problems. The first problem 'refugee crisis' is assigned to the first category 'refugees ambiguous', the second entry 'terrorism' is assigned to category 6, 'terrorism', and the last entry 'contact with turkey' is assigned to 'foreign relations' (category 12). It may be argued that if a person mentions the refugee crisis and terrorism as the most important political problems, the person considers the existence of refugees in Germany as a problem, and the entry would hence belong in the second category ('refugees negative'). Since a combined examination of several entries is harder to control and reproduce, we decided upon a more transparent, conservative and verifiable procedure.

### 3.2 Development of the coding tool

In order to ease the coding process and allow for quick and consistent coding, the coding instructions were implemented by Martin Bruemmer in a coding tool using the Python Tk library.<sup>4</sup> The tool loads comma-separated values (CSV) files that contain respondents' IDs and their respective answer. It then displays each answer to the coder, allowing them to choose the number of topics mentioned in the answer (Figure 1 A). For each topic identified, the tool then shows all categories available for coding (Figure 1 B). After categories have been selected for each topic identified in the answer, the next answer is presented. Coding is finished when all responses in the loaded CSV file have been coded. The results are then saved to another CSV file named after the coder and the file coded. In addition to the respondents' IDs and answers, the result CSVs contain the number of topics identified as well as one column for each topic containing the number of the category selected. The source code is available on github<sup>5</sup>.

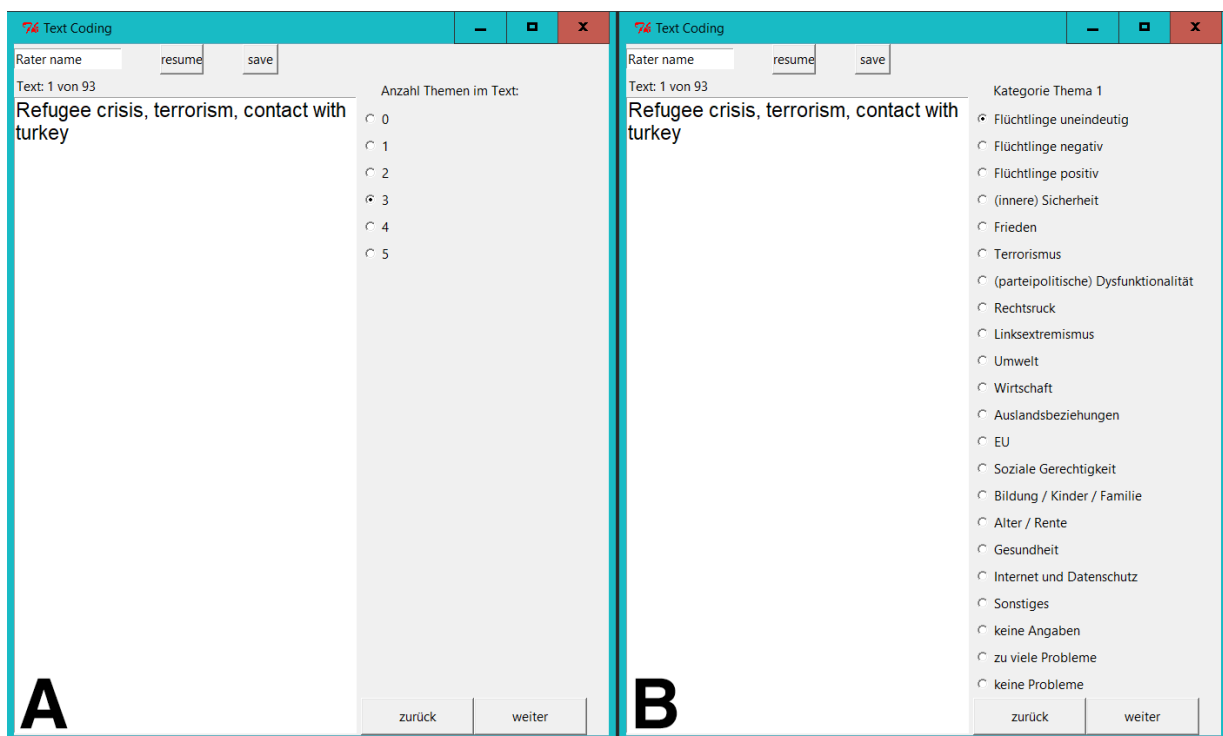


Figure 1. User interface of the coding tool

The tool was built for Windows 10 and Windows 7 using py2exe<sup>6</sup> and for Mac OS X using py2app<sup>7</sup>. These libraries only work when the built itself is run on the operating system it is built for. Thus, to supply all coders with version compatible with their systems, the tool had to be built three times on three different systems. This lack in cross-platform compatibility constitutes a major drawback of using Python to develop user interfaces. Electron<sup>8</sup> apps are currently the state-of-the-art in local cross-platform applications but come with a relatively

<sup>4</sup> <https://docs.python.org/3/library/tk.html>

<sup>5</sup> <https://github.com/der-bruemmer/political-text-codingtool>

<sup>6</sup> <http://www.py2exe.org/>

<sup>7</sup> <https://pypi.python.org/pypi/py2app/>

<sup>8</sup> <https://electron.atom.io/>

large overhead in development time and system resources required. The online survey framework formR<sup>9</sup> provides an easily configurable alternative, allowing ratings to be collected in a web application. However, for the purpose of this study, responses were deemed too sensitive for the web-based approach.

### 3.3 Harmonization and Indices of Coding

The codings of the three coders were merged together and anonymized by an independent person.

As shown in table 2, 78 % of the codings of the first entry coincide. In 20 % of the codings, two coders agreed while the third coder diverged. In those cases, the majority principle was used to harmonize the coding. If there were three different codes (2 %), one of the three codes was chosen randomly.

	absolute frequency	relative frequency
three identical codes	1014	77.70
two identical codes	266	20.38
Code 1 = Code 2	144	11.03
Code 2 = Code 3	76	5.82
Code 1 = Code 3	46	3.52
no identical codes	25	1.92
total	1305	100

Two measures of “inter-rater reliability”, the percentage of agreement and Krippendorff’s alpha, were used to investigate the agreement between the coders regarding the number of topics mentioned in one statement and the categorization of the entries themselves. Both measures were calculated for the entirety of codes (i.e. combining 2015 and 2017). A separate consideration of the years was not implemented, since there were no substantial reasons to do so for the analysis of the inter-rater reliability. The agreement on the categorization of the single entries was calculated using only those cases in which all of the three coders agreed on the topic count ( $n = 1,203$ ), since it did not seem appropriate to calculate the agreement on the second entry if, for example, one could not be sure if both coders coded the same piece of information. Furthermore, eight cases with no codings resulting from errors in the reading of data were removed, leaving a total of  $n = 1,195$ .

The percentage of agreement was computed by calculating the average pairwise agreement among all possible coder pairs across all observations. The coders agreed to 95 % on the number of topics that were mentioned in one statement ( $n = 1,305$ ). The agreement was 88 % for the first entry ( $n = 1195$ ), 81 % for the second entry ( $n = 102$ ) and 79 % for the third entry ( $n = 13$ ). After that, a significant drop of the percentage of agreement was observable, with only 22 % for the fourth entry ( $n = 3$ ) and 33 % for the fifth entry ( $n = 1$ ).

Since the percentage of agreement does not take into account the possibility of random agreement between coders, Krippendorff’s alpha was additionally calculated. Ratings showed an agreement of  $\alpha = .753$  ( $n = 1,305$ ) on the number of topics mentioned in one statement. The agreement amounted to  $\alpha = .847$  ( $n = 1,195$ ) for the first entry,  $\alpha = .787$  ( $n = 102$ ) for the second entry and  $\alpha = .77$  ( $n = 13$ ) for the third entry. Inter-rater reliability dropped significantly below the acceptable level of .70 for the fourth ( $\alpha = .07$ ,  $n = 3$ ) and fifth rating ( $\alpha = -.25$ ,  $n = 1$ ).

Both percentage of agreement<sup>10</sup> and Krippendorff’s alpha<sup>11</sup> showed an acceptable reliability for the number of topics and for the rating of the first three entries. Since the fourth and fifth entry are not to be

<sup>9</sup> Arslan, R.C., & Tata, C.S. (2017). *formr.org survey software (Version v0.16.5)*. Zenodo. <http://doi.org/10.5281/zenodo.398836>

considered separately in further analyses, the drop in reliability does not seem problematic for the use of the coded data.

## 4 Results

A frequency analysis was performed to answer the question what people actually considered to be the most important political problem in 2015 and 2017. Although the interviewees were only asked to name the single most important topic, up to five entries were given and thus coded in the coding process, obtaining the highest possible information content. This allowed different ways of combining the data for the frequency analysis. In this chapter, we would like to illustrate the three separate methods of data combination chosen and afterwards present their results. The results are discussed in the following chapter with regard to the question if those different possibilities of data combination actually lead to different conclusions.

The first analysis of frequencies was performed using only the ratings for the first entry. Although this option does not take into account the fact that people mentioned more than one topic and therefore has the lowest information content in regard to the actual material, it is the option that is indicated by the phrasing of the question itself. The absolute and relative frequencies were calculated for the harmonized codings of 2015 and 2017, using only those cases in which all coders agreed on the same topic count ( $n = 1,203$ ).

Since the focus was to compare different methods of frequency analysis, the subsamples for 2015 and 2017 were not considered separately. A comparison of the frequencies in 2015 and 2017, although not the main objective of this paper, is an interesting question for further research.

For an explanation of the decision to remove those cases where the coders disagreed on the topic count please refer to the previous chapter. Frequencies were calculated for the single categories as well as for the supercategories. Table 3 shows the frequencies of the harmonized codings for the first entry.

<sup>10</sup> Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, California: Sage Publications.

<sup>11</sup> Krippendorff, K. (2004a). *Content analysis: An introduction to its methodology*. Thousand Oaks, California: Sage.

Table 3. *Frequencies for first entry only*

supercategory	category	first entry only			
		absolute frequency		relative frequency (in %)	
refugees	refugees ambiguous	573	530	47.63	44.06
	refugees negative		23		1.91
	refugees positive		20		1.66
security and terrorism	security	137	94	11.39	7.81
	peace		18		1.50
	terrorism		25		2.08
polity and political culture in Germany	party political dysfunction	151	112	12.55	9.31
	swing to the right		39		3.24
	swing to the left		0		0
environment	environment	15	15	1.25	1.25
economy	economy	23	23	1.91	1.91
foreign relations and EU	foreign relations	35	20	2.91	1.66
	EU		15		1.25
social system	social justice	149	96	12.39	7.98
	education / children / family		28		2.33
	age / pension		21		1.75
	health		4		0.33
internet and data protection	internet and data protection	5	5	0.42	0.42
other	other	60	60	4.99	4.99
no answer	no answer	47	31	3.91	2.58
	too many problems		4		0.33
	no problems		12		1.00
	NA	8	8	0.67	0.67

In order to preserve all information and to take into account the fact that respondents mentioned more than one topic, the second and third analysis of frequencies were performed using all the available coded topics. Therefore, the answers for up to five topics were analysed, unweighted (second analysis) or weighted by the inverse of the number of topics mentioned by the respective respondent (third analysis). As in the previous analysis, absolute and relative frequencies were calculated for the combined samples of 2015 and 2017 only for those cases in which the coders agreed on the topic count ( $n = 1,203$ ).

For the second analysis, each of the mentioned topics was counted as one answer. A respondent who gave five answers thereby had a higher impact on the results. Thus, in the third analysis, all answers were weighted, making sure that no respondent was overrepresented in the analysis. In order to do so, each answer was provided with a weight from 1 to 5, depending on the number of topics mentioned in the answer given by the interviewee, according to the formula  $\frac{5}{n_{topic}}$ . The resulting absolute frequencies for each category were divided by 5 in the end. The results for analysis 2 and 3 are presented in table 4.

Table 4. *Frequencies for all five entries*

supercategory	category	all five entries (unweighted)				all five entries (weighted)			
		absolute frequency		relative frequency (in %)		absolute frequency		relative frequency (in %)	
refugees	refugees ambiguous	592	548	44.78	41.45	555.37	513.87	46.17	42.72
	refugees negative		23		1.74		21.50		1.79
	refugees positive		21		1.59		20		1.66
security and terrorism	security	168	110	12.71	8.32	143.17	97	11.90	8.06
	peace		20		1.51		18		1.50
	terrorism		38		2.87		28.17		2.34
polity and political culture in Germany	party political dysfunction	160	118	12.10	8.93	154.09	113.75	12.81	9.46
	swing to the right		42		3.18		40.34		3.35
	swing to the left		0		0		0		0
environment	environment	19	19	1.44	1.44	15.17	15.17	1.26	1.26
economy	economy	31	31	2.34	2.34	23.28	23.28	1.94	1.94
foreign relations and EU	foreign relations	43	26	3.25	1.97	34.57	19.40	2.87	1.61
	EU		17		1.29		15.17		1.26
social system	social justice	185	116	13.99	8.77	158.37	100.45	13.16	8.35
	education / children / family		37		2.80		31.92		2.65
	age / pension		27		2.04		22		1.83
	health		5		0.38		4		0.33
internet and data protection	internet and data protection	6	6	0.45	0.45	4.83	4.83	0.40	0.40
other	other	63	63	4.77	4.77	59.17	59.17	4.92	4.92
no answer	no answer	47	31	3.56	2.34	47	31	3.91	2.58
	too many problems		4		0.30		4		0.33
	no problems		12		0.91		12		1.00
NA	NA	8	8	0.61	0.61	47	8	0.67	0.67

## 5 Discussion

In this concluding section, we would like to discuss the particularities of the data, the coding process and the resulting codings that should be kept in mind when using the data for further analysis.

Initially we would like to emphasise once more that a conservative way of coding ambiguous answers was chosen for the separate categories of the supercategory ‘refugees’. Every statement that did not contain a clear evaluative judgement of refugees themselves was coded as ‘refugees ambiguous’. Therefore, the category ‘refugees ambiguous’ is by no means interpretable as neutral.

Another critical point is the asymmetrical contribution of the case numbers in 2015 and 2017. In 2015 only 289 entries exist, whereas 2017 1016 entries are available. As consequence of the low sample size, analyses restricted to answers from 2015 do not seem particularly fruitful. Rather, the focus should be either on 2017 or on the combination of 2015 and 2017.

Although respondents were asked for the most important political problem in Germany, some interviewees reported more than one political problem. Table 4 shows some slight differences in the frequencies of the categories between the first entry only and all five entries. We suggest that substantive consideration should come into play when deciding how to handle multiple replies by single respondents. If the aim is to provide all information given by the respondent, including all answers might be more appropriate.

Whether or not replies should be weighted by the number of topics provided depends on whether one prefers to assign equal weight to each respondent or actually assumed that respondents who gave more comprehensive replies had “more to communicate”. In any case, robustness checks should be applied to ensure that conclusions do not hinge on arbitrary decisions regarding the handling of multiple replies.

## References

- Arslan, R.C., & Tata, C.S. (2017). *formr.org survey software (Version v0.16.5)*. Zenodo.
- Giesselmann, M., Brümmer, M., Kroh, M., Siegel, N. A., & Wagner, G. G. (2017). Fluchtzuwanderung ganz oben auf der Liste der dringenden politischen Prioritäten. *Wirtschaftsdienst*, 97(3), 192-200.
- Krippendorff, K. (2004a). *Content analysis: An introduction to its methodology*. Thousand Oaks, California: Sage.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, California: Sage Publications.
- Wagner, G. G., Bruemmer, M., Glemser, A., Rohrer, J. M., & Schupp, J. (2017). Dimensions of Quality of Life in Germany: Measured by Plain Text Responses in a Representative Survey (SOEP). SOEPpaper No. 893, Berlin.

<https://docs.python.org/3/library/tk.html>

<https://github.com/der-bruemmer/political-text-codingtool>

<http://www.py2exe.org/>

<https://pypi.python.org/pypi/py2app/>

<https://electron.atom.io/>