

Data Sources for “context.dta”

Title of the paper: Hate Crimes and Gender Imbalances: Fears over Mate Competition and Violence against Refugees

Authors:

Rafaela Dancygier, Princeton University, rdancygi@princeton.edu

Naoki Egami, Columbia University, naoki.egami@columbia.edu

Amaney Jamal, Princeton University, ajamal@princeton.edu

Ramona Rischke, Humboldt-Universität zu Berlin, ramona.rischke@hu-berlin.de

Summary

context.dta is a dataset used to produce contextual analyses in the paper (Figure 1) and in Appendix C (Table and Figures C1 --- C8, and C10).

In this document, we provide details of data sources. We used “merge_context.R” and “produce_context_data.do” to merge data sources we described below into “context.dta”.

For our codebook of our main analysis data, “context.dta”, please see “codebook_context.pdf”.

Details about Data Sources

- **Mut Gegen Rechte Gewalt**
 - This data is provided in a joint effort by the Amadeu Antonio Foundation and PRO ASYL. The data is available here: <https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle> and was retrieved on May 15, 2018.
 - This data source is used to produce: “Physical_muni”, “Physical_muni_bin”, “Hate_all_muni”, “Hate_all_muni_bin” in “context.dta”
 - The original raw data has been processed in the following manner:
 - The aforementioned variables are generated based on: “hc_type”, “hc_year”
 - The German variable names in the original raw data file are: “Art” (hc_type) and “Datum” (date that provides hc_year).
 - AGS codes (i.e. municipality identifier) were merged to the data based on GPS coordinates (“Latitude”, “Longitude” in the original raw data)
 - The AGS “reference codes” and the associated geographical map (so-called “shapefile”) come from the Federal Agency for Cartography and Geodesy, 2016, GeoBasis - DE / BKG 2016 www.bkg.bund.de, downloaded: 10.03.2017.

- For each municipality, we count the number of each type of hate crimes, which produces "Arson_muni", "Demonstration_muni", "Other_muni", "Physical_muni", "Murder_muni".
 - This original dataset is called, "hate.dta" in "merge_context.R"
- **Statistische Ämter des Bundes und der Länder**
 - County and municipality level control variables are sourced from the Federal Statistical Office of Germany (Destatis) and Statistical Offices of Federal States, respectively. These statistics are made available on a joint platform: regionalstatistik.de and have been retrieved between March 2017 and July 2019. Different tables can be accessed entering the table numbers below in the searching tool and subsequently specifying the relevant years.
 - This data source is used to produce: "pop_25_44_muni_gendergap_2015", "pop_15_44_muni_gendergap_2015", "population_muni_2015", "log_population_muni_2015", "popdens_muni_2015", "log_popdens_muni_2015", "d_pop1511_muni", "vote_afd_2013_muni", "pc_hidegree_all2011", "d_manuf1115", "pc_manufacturing_2015", "unemp_gendergap_2015" in "context.dta"
 - The original raw data has been processed in the following manner:
 - In order to facilitate data handling of original raw data (Excel format), some tables had to be reformatted manually before importing to R or Stata (e.g. to deal with nested tables and merged cells, renaming and translating variables for import, removing footnotes etc.).
 - The variables related to population by gender and age at the level of municipalities were produced from the tables "**12411-02-03-5**".
 - This data set is called, "pop_gemeinde_2008_2018.dta" in "merge_context.R".
 - The German variable names in the original raw data files are as follows:
 - "Insgesamt Altersgruppen (unter 3 bis 75 u.m.) ... :
 - ... 15 bis unter 18 Jahre" (pop_mf_15_17).
 - ... 18 bis unter 20 Jahre" (pop_mf_18_19).
 - ... 20 bis unter 25 Jahre" (pop_mf_20_24).
 - ... 25 bis unter 30 Jahre" (pop_mf_25_29).
 - ... 30 bis unter 35 Jahre" (pop_mf_30_34).
 - ... 35 bis unter 40 Jahre" (pop_mf_35_39).
 - ... 40 bis unter 45 Jahre" (pop_mf_40_44).
 - ... 45 bis unter 50 Jahre" (pop_mf_45_49).
 - ... 50 bis unter 55 Jahre" (pop_mf_50_54).
 - ... 55 bis unter 60 Jahre" (pop_mf_55_59).
 - ... 60 bis unter 65 Jahre" (pop_mf_60_64).
 - ... Insgesamt" (pop_mf_total).
 - "männlich Altersgruppen (unter 3 bis 75 u.m.) ... :
 - ... 15 bis unter 18 Jahre" (pop_m_15_17).
 - ... 18 bis unter 20 Jahre" (pop_m_18_19).
 - ... 20 bis unter 25 Jahre" (pop_m_20_24).
 - ... 25 bis unter 30 Jahre" (pop_m_25_29).
 - ... 30 bis unter 35 Jahre" (pop_m_30_34).
 - ... 35 bis unter 40 Jahre" (pop_m_35_39).

- ... 40 bis unter 45 Jahre" (pop_m_40_44).
 - ... 45 bis unter 50 Jahre" (pop_m_45_49).
 - ... 50 bis unter 55 Jahre" (pop_m_50_54).
 - ... 55 bis unter 60 Jahre" (pop_m_55_59).
 - ... 60 bis unter 65 Jahre" (pop_m_60_64).
 - "weiblich Altersgruppen (unter 3 bis 75 u.m.) ... :
 - ... 15 bis unter 18 Jahre" (pop_f_15_17).
 - ... 18 bis unter 20 Jahre" (pop_f_18_19).
 - ... 20 bis unter 25 Jahre" (pop_f_20_24).
 - ... 25 bis unter 30 Jahre" (pop_f_25_29).
 - ... 30 bis unter 35 Jahre" (pop_f_30_34).
 - ... 35 bis unter 40 Jahre" (pop_f_35_39).
 - ... 40 bis unter 45 Jahre" (pop_f_40_44).
 - ... 45 bis unter 50 Jahre" (pop_f_45_49).
 - ... 50 bis unter 55 Jahre" (pop_f_50_54).
 - ... 55 bis unter 60 Jahre" (pop_f_55_59).
 - ... 60 bis unter 65 Jahre" (pop_f_60_64).
 - AGS codes (i.e. municipality identifier) were provided in the data (original variable name "Gemeinden")
- The variable related to population on the level of counties (pop_mf_total_total) was imported from the tables **"12411-03-03-4"**.
 - This data set is called, "pop_kreise_2015_2017.dta" in "merge_context.R".
 - The German variable names in the original raw data files are as follows:
 - In cross-table: [Bevölkerungsstand (Anzahl)] „Insgesamt“ [i.e. both foreign and german] in „Insgesamt“ [i.e. both male and female] and „Insgesamt“ [i.e. all ages] (pop_mf_total_total)
 - AGS codes (i.e. county identifier) were provided in the data (original variable name "Kreise und kreisfreie Städte")
- The variable on surface areas to produce population density comes from the table **"11111-01-01-5"**.
 - This data set is called, "area_mun.dta" in "merge_context.R".
 - The German variable names in the original raw data files are as follows:
 - "Gebietsfläche qkm" (area_sqk)
 - AGS codes (i.e. municipality identifier) were provided in the data (original variable name "Gemeinden")
- The variable related to voting is produced from the table **"14111-01-03-5"**.
 - This dataset is called, "voting.dta" in "merge_context.R".
 - The German variable names in the original raw data files are as follows:
 - "Gültige Zweitstimmen Parteien AfD - Anzahl" (AfD)
 - "Gültige Zweitstimmen - Anzahl" (valid.second.votes)
 - AGS codes (i.e. municipality identifier) were provided in the data (original variable name "Wahltag Gemeinden")
- The variable related to education is produced from the table **"12111-09-01-4"**.
 - This data set is called, "education.dta" in "produce_context_data.do".
 - The German variable names in the original raw data files are as follows:
 - In cross-table: „Allgemeine/fachgebundene Hochschulreife (Abitur)“ in „Höchster Schulabschluss“ and „Bevölkerung (ab 15 Jahren)“ (pop15_high_degree2011)

- In cross-table: „Insgesamt“ in „Höchster Schulabschluss“ and „Bevölkerung (ab 15 Jahren)“ (pop15_total)
 - AGS codes (i.e. county identifier) were provided in the data (original variable name “Kreise und kreisfreie Städte”)
- The variable “unemp_gendergap” is produced from the table “**13211-02-05-4**”.
 - This data set is called, “unemployment.dta” in “produce_context_data.do”.
 - The German variable names in the original raw data files are as follows:
 - Arbeitslosenquote bez. auf alle zivile Erwerbsp. – Männer - Prozent“ (unemp_men)
 - „Arbeitslosenquote bez. auf alle zivile Erwerbsp. – Frauen - Prozent“ (unemp_female)
 - AGS codes (i.e. county identifier) were provided in the data (original variable name “Kreise und kreisfreie Städte”)
- The variable related to producing variables related to employment sectors are produced from table “**13312-01-05-4**”.
 - This data set is called, “sectors.dta” in “produce_context_data.do”.
 - The German variable names in the original raw data files are as follows:
 - “Erwerbstätige im Jahresdurchschnitt – Insgesamt - 1000” (no_employed)
 - “Erwerbstätige im Jahresdurchschnitt – Produzierendes Gewerbe ohne Baugewerbe (B-E) - 1000” (no_manufacturing)
 - AGS codes (i.e. county identifier) were provided in the data (original variable name “Kreise und kreisfreie Städte”)
- **Statistik der Bundesagentur für Arbeit**
 - This data is provided by the Federal Employment Agency (“Bundesagentur für Arbeit”) and was retrieved on March 8 and March 9, 2017. The data is made available on the platform: regionalstatistik.de. The variables were produced from the tables “13211-01-03-5”. These tables can be accessed entering the table numbers in the searching tool and subsequently specifying the relevant years.
 - This data set is called, “unempl_gemeinde_2008_2017.dta” in “merge_context.R”.
 - This data source provides us with unemployment figures to produce “log_unemp_all_muni_2015”,
 - The original raw data has been processed in the following manner:
 - In order to facilitate data handling of original raw data (Excel format), some tables had to be reformatted manually before importing to Stata (e.g. to deal with nested tables and merged cells, renaming and translating variables for import, removing footnotes etc)
 - The variable “unempl_all_total” is imported from the table “**13211-01-03-5**”.
 - The German variable names in the original raw data files are as follows:
 - „Arbeitslose insgesamt“ (unempl_all_total)
 - AGS codes (i.e. municipality identifier) were provided in the data (original variable name “Gemeinde”)
- **Statistisches Bundesamt (Destatis)**
 - The refugee data used has been purchased from the Federal Statistical Office of Germany (Destatis) and provided on November 11, 2017 and June 28, 2018. We

have the permission from the Federal Statistical Office of Germany to distribute the data.

- Data sets generated from this source are called: “refugees_2008_2017.dta” in “merge_context.R” and “refugee_gender.dta” in “produce_context_data.do”.
- This data source is used to produce “log_ref_inflow_1514”, “pop_ref_2014”, “log_pop_ref_2014”, “pc_ref_male”.
- The original raw data has been processed in the following manner:
 - In order to facilitate data handling of original raw data (Excel format), some tables had to be reformatted manually before importing to Stata (e.g. to deal with nested tables and merged cells, renaming and translating variables for import, removing footnotes etc)
 - The German variable names in the original raw data files are as follows:
 - [Ausländische Bevölkerung] “darunter Schutzsuchende” (pop_ref / all_totref)
 - „Alter der Schutzsuchenden von ... bis unter ... Jahren – männlich...”
 - ... unter 3 Jahre” (all_male0_3).
 - ... 3 – 6” (all_male3_6).
 - ... 6 – 15” (all_male6_15).
 - ... 15-18” (all_male15_18).
 - ... 18 – 25” (all_male18_25).
 - ... 25 – 30” (all_male25_30).
 - ... 30 – 40” (all_male30_40).
 - ... 40 – 50” (all_male40_50).
 - ... 50 – 65” (all_male50_65).
 - ... 65 – 75” (all_male65_75).
 - ... 75 und mehr” (all_male75_up).
 - AGS codes (i.e. county identifier) were provided as names of individual Excel data sheets)

- **Bundeskriminalamt**

- This data is provided by the German Federal Office of Criminal Investigations and was retrieved on Oct 2, 2017. The data can be accessed selecting a specific table from this link:
<https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/PKS2015/BKATabellen/bkaTabellenLaenderKreiseStaedteFaelle.html?nn=51356>.
- The name of the table specifically to select is “[Grundtabelle - Kreise - ausgewählte Straftaten/-gruppen - excel \(xlsx, 2MB\)](#)”.
- This dataset is called, “crime.dta” in “merge_context.R”
- This data source is used to produce “violence_percap_2015”, “log_violence_percap_2015”
- The German variable names in the original raw data files are as follows:
 - In cross table: “Gewaltkriminalität” in „Straftat“ and then “erfasste Fälle” (violence_num_cases)

- AGS codes (i.e. county identifier) were provided in the data (original variable name “Gemeindeschlüssel”)

References:

- Amadeu Antonio Foundation and PRO ASYL. 2018. “Mut Gegen Rechte Gewalt.” <https://www.mut-gegen-rechte-gewalt.de/service/chronik-vorfaelle> (retrieved on May 15, 2018)
- Statistische Ämter des Bundes und der Länder. “Regionaldatenbank Deutschland.” 2017 and 2018. <https://www.regionalstatistik.de/genesis/online/> (retrieved between March 2017 and July 2018)
- Bundesagentur für Arbeit. 2017. “Statistik der Bundesagentur für Arbeit.” <https://www.regionalstatistik.de/genesis/online/> (retrieved on March 8 and March 9, 2017)
- Statistisches Bundesamt. 2018. “Ausländische Bevölkerung und Schutzsuchende in Deutschland” <https://www.destatis.de/> (Data provided on June 28, 2018)
- Bundeskriminalamt. 2017. “BKA (Länder-, Kreis- und Städte) Übersicht Falltabellen.” <https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/PKS2015/BKATabellen/bkaTabellenLaenderKreiseStaedteFaelle.html?nn=51356> (retrieved on Oct 2, 2017)