U.S. Refugee Resettlement Datasets (Beta versions)

The U.S. Refugee Resettlement Datasets cover the universe of refugees recorded as entering the U.S. over the 1975-2018 period. The data originate from the Office of Refugee Resettlement (ORR, 1975-2008) and the Bureau of Population, Refugees and Migration (PRM, 2009-2018) and comprise information on more than 3 million refugees.

Two datasets are provided to cater to different user needs:

First, the **ORR Individual Data File (1975-2008)** contains individual-level data for over 2.5 million individuals.

Secondly, we merged our digitised ORR records with publicly available data from the PRM to expand the information on refugees at the city level to 2018. As such the **(ORR & PRM) Geo-coded Data File (1975-2018)** provides the geo-locations of the universe of refugees that entered the U.S. between 1975 and 2018. The dataset is geo-coded based on city name or city code. Approximately 3% of the data did not provide sufficient enough information for geocoding and were excluded from.

Please note that several months of data were missing in the original files for the year 1990.

If you use the data please cite:

"Dreher, A., Langlotz, S., Matzat, J., Parsons, C. and Mayda, A. (2020). Immigration, Political Ideologies and the Polarization of American Politics. CEPR Discussion Paper 15587."

Variables:

In both datasets	
year	Year of arrival
citizenship_stable	Stable unit over time based on the *country of citizenship* variable from ORR (1975-2008) and the *origin* variable from PRM (2009-2018).
city	City name (original)
city_code	City code (original)
county_fips	3-digit county Federal Information Processing Standard (FIPS) codes (original)
state_fips	2-digit state Federal Information Processing Standard (FIPS) codes (original)
fips	5-digit Federal Information Processing Standard (FIPS) codes (original)
countyfp10	3-digit 2010 county Federal Information Processing Standard (FIPS) codes
statefp10	2-digit 2010 state Federal Information Processing Standard (FIPS) codes
geo_level	Level of geocoding
longitude	Longitude (World Geodetic System 1984, WGS 84)
latitude	Latitude (World Geodetic System 1984, WGS 84)
id_merging	ID that can be used for merging the ORR individual data file to the ORR & PRM geo-coded data file
Only in ORR & PRM Geo-coded Data File (1975-2018)	
refugees	Number of refugees
city_id	Place ID
city_nat_id	Place – nationality ID
city_nat_year_id	Place – nationality – year ID

Only in ORR Individual Data File (1975-2008)	
birth_country_name	Country of birth
citizenship_country_name	Country of citizenship
arrival_date	Date of arrival
birth_date	Date of birth
birth_year	Year of birth
sex	Sex
marital_status	Marital status
family_Status	Family status
education	Education
english_reading	English: reading
english_writing	English: writing
English_speaking	English: speaking

Cautionary note:

Please note that we did not modify city_name, city_code and city id's, which need further work to uniquely identify cities. First, we geocoded the location as precisely as possible. As a result, we are sometimes able to identify not only cities but also neighbourhoods. Second, we used the Google Maps API, the Open Street Maps API, as well as the data science toolkit (http://www.datasciencetoolkit.org/) to yield the best results possible. As a result, we are potentially identifying slightly different places within the same city. While city id's identify places, they might identify several places within the same city. The city and city_code variable have not been cleaned and do contain spelling mistakes.

Acknowledgments:

We do want to thank Clara Coelho and Tobias Hellmundt for excellent research assistance and Vasil Yasenov for sharing the PRM refugee data with us.

Please let us know if you find any remaining errors: johannes.matzat@uni-goettingen.de.