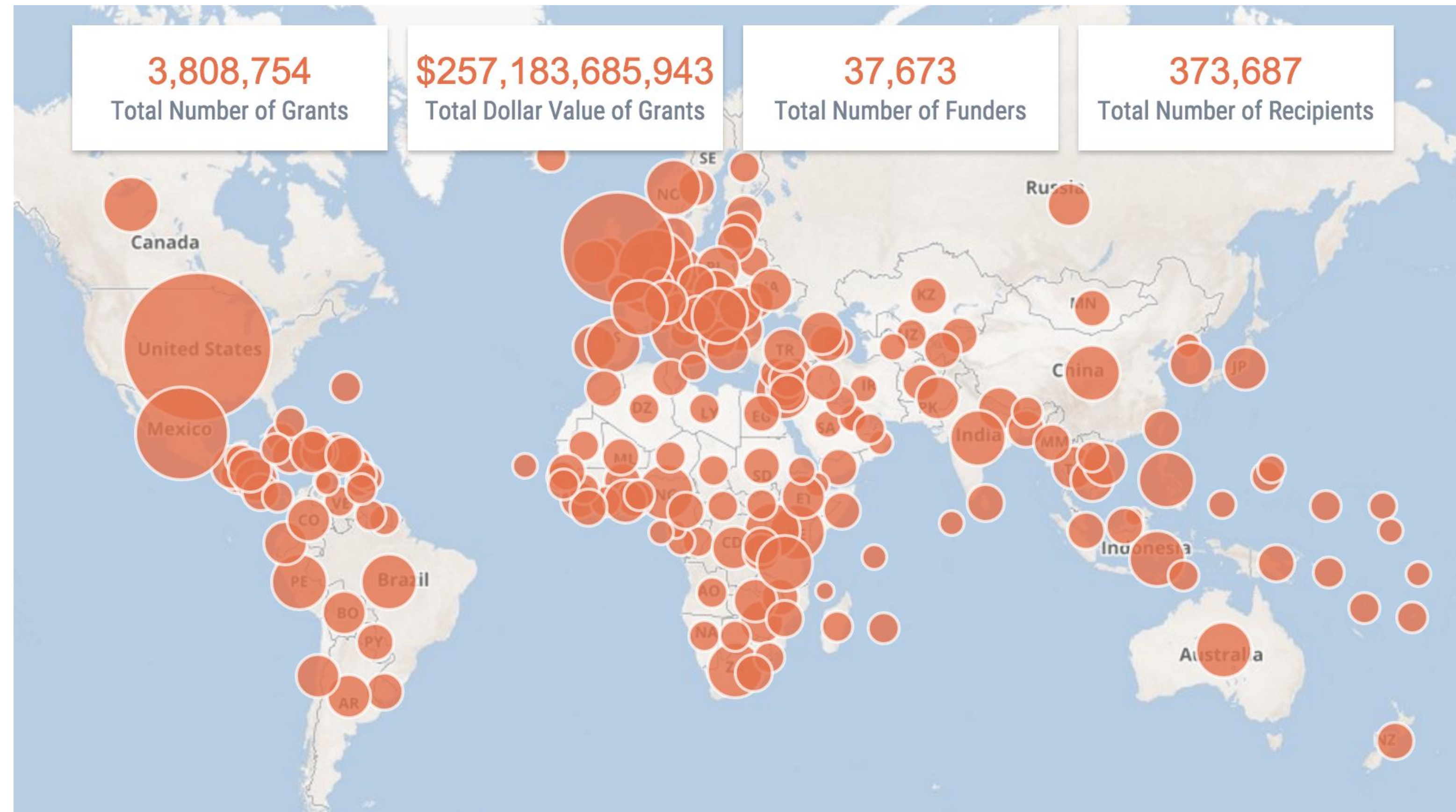


Foundation Center

Nonprofit organization based in New York

Leading source of information about philanthropy worldwide

- Information on more than **140,000** global foundations, corporate donors, and grantmaking public charities
- Database of **3.8 million** grants to non-profits around the world
- 2 million sessions per month on our web properties, like Foundation Maps, Foundation Directory Online, Philanthropy News Digest, and Grantspace



Problem statement

Using the following data sets:

- **news from Common Crawl**
- **the institutions we have in our database (with unique identifiers [e.g., BRIDGE IDs])**
- **the Geonames geographic gazetteer of place names**

...provide the following output:

- **Articles related to philanthropy (e.g., grantmaking, charitable giving, etc.);**
- **Articles that relate to philanthropic institutions we have in our database**
- **new philanthropic institutions that *are not or might not* be in our database**
- **philanthropic activity occurring in a given country (like India), if any place in India is mentioned (using the Geonames hierarchy)**

Details

We want to collect, clean, and curate recent data on philanthropy from news sources. We're looking for the following:



Current (2015-2016) data about philanthropy in the U.S.

Recent data (2010 to present) about philanthropy outside the U.S.

Requirements / limitations

For the data to be used in our products, it must meet the following requirements:

Where possible, organization names need to be disambiguated and matched with organizations already in our database (140,000 foundations, and 374,000 nonprofits). In our experience, the same organization can appear in datasets with many alternative names. (E.g. “Franklin Center for the Arts” vs. “Franklin Arts Center”; or “U.C.L.A.” vs. “U. C. Los Angeles”, etc.). We need to assess the likelihood of matches, if no direct match is found.

Locations of activity need to be disambiguated and coded to known locations in the Geonames gazetteer (<http://geonames.org>). In the text, “This project will fund a homeless shelter in Springfield”, we need to identify that Springfield is a location, and then we need to figure out which Springfield this refers to.

Collaboration

Foundation Center staff.

We have data scientists, geographers, programmers, and data analysts available and happy to answer questions on a regular basis.

Auto-classification.

We have a machine learning auto-classifier that can be used to sort out which articles might be related to philanthropy.

We also have some starter code for disambiguating non-profits and for disambiguating Geonames locations — but these aren't very accurate nor precise.

Contact: Jake Garcia, jkg@foundationcenter.org