Applicant: Jarmila Ruzicka

Application Project Proposal for Data Incubator

Inspiration:

Admiration for small countries such as New Zealand, Singapore and Iceland that are progressive in nature. I would like to explore if the connectedness index has something to do with this. In this intro we will be only exploring the countries with the highest indices, the aim is to find an optimal country size (either land or population wise) that will optimize other variables that could depend on the connectedness index.

The goal is to study all the countries and find correlations between size, connectedness and what influence does it have in other sociopolitical or socioeconomic factors.

Goal:

To answer the following question: Is there an optimal community size in order to achieve maximum progress?

Is there an ideal population density or territory size that would lead a state to achieve an optimal rate of progress?

Why this goal?

Once an optimal community size is estimated, local governments can make sure that community has a governing body, such as a municipality, or that it belongs to its own county as in the US in order to ensure its fullest participation and progress.

Possible Applications:

A communal rating system for government officials that serve that community. This would be a proposed solution to the most recent national controversy of the police use of force in the US. If each community had a rating system similar to Yelp, the information shared can be used to correct issues that otherwise might go unnoticed; if the sentiment is good, they will achieve a higher rating, this could also be a way of handling immediate concerns.

Background:

Connectedness is the state of being linked with others as a result forming one body. Connectedness is powerful as it creates an infrastructure to broadcast.

$$Social\ Connectedness\ Index_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i * FB\_Users_j}$$

Here, FB_Users_i and FB_Users_j are the number of Facebook users in locations i and j, and $FB\_Connections_{i,j}$ is the total number of Facebook friendship connections between individuals in the two locations.

The scaling is such that each dataset has a maximum value of one billion and a minimum of one.

We can then study if there is a relationship between connectedness and progress within that country. We can compare how many bills are passed, how much growth is achieved, crime diminished or other factors of progress to countries of different connectedness levels.

Example of Costa Rica: Use of social media to create awareness of recycling, tree planting, to report new criminal practices, to organize events for a good cause, to petition new laws, to report missing children, to shame and be aware of criminals in the area.

## Who is the end user?

Communities and Governments.

As it would help decide how many municipalities or local governments are needed per population density, also it may help redefine what issues can be addressed by those bodies.

Method:

First study the countries with the max connectedness index within, that is both locations of the link are the same country. Then run linear regressions and on elements of progress such as: crime progression, gpd, policy making, education and find its relationship to the connectedness.
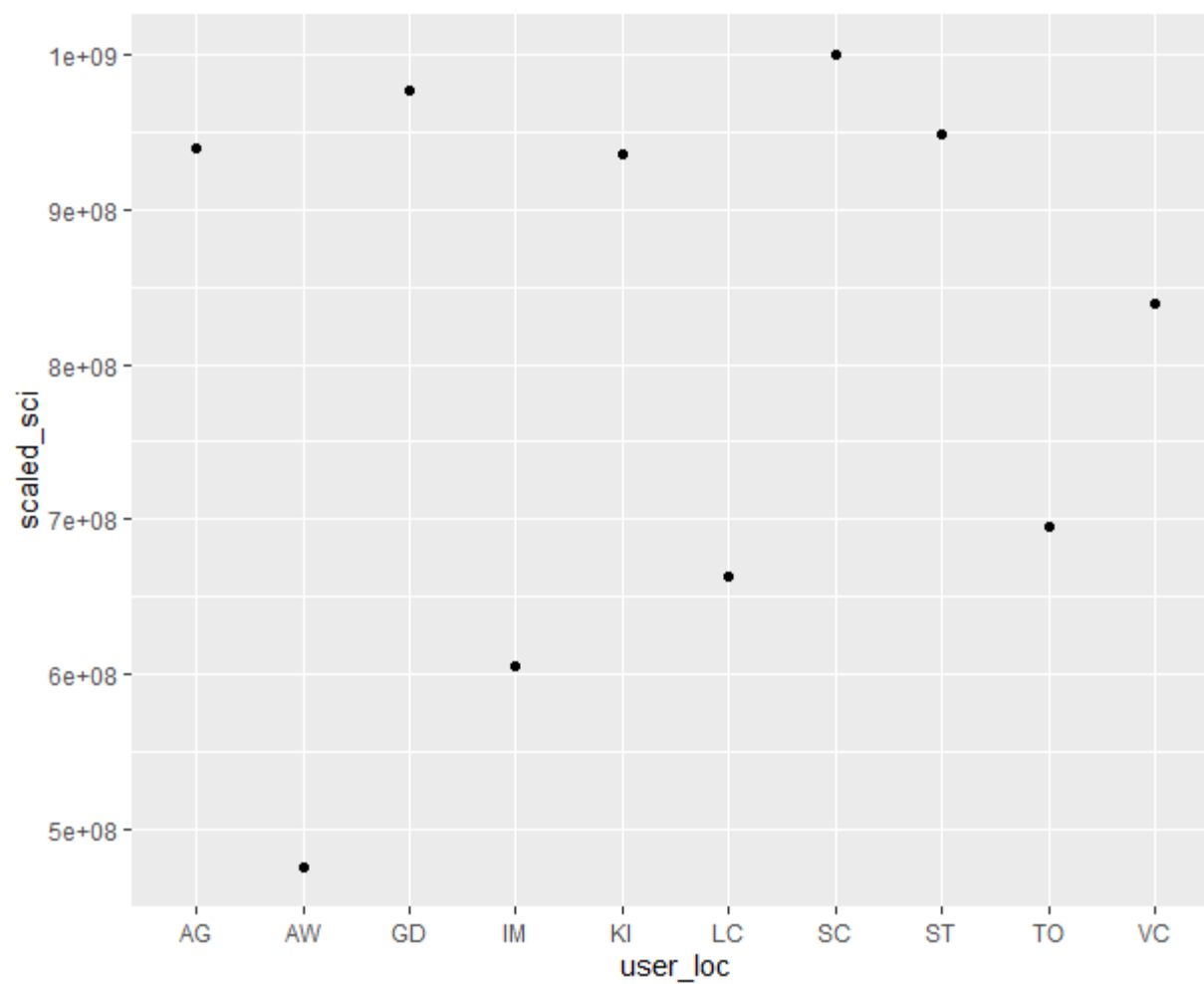
We can also study county to county as in the case within US.

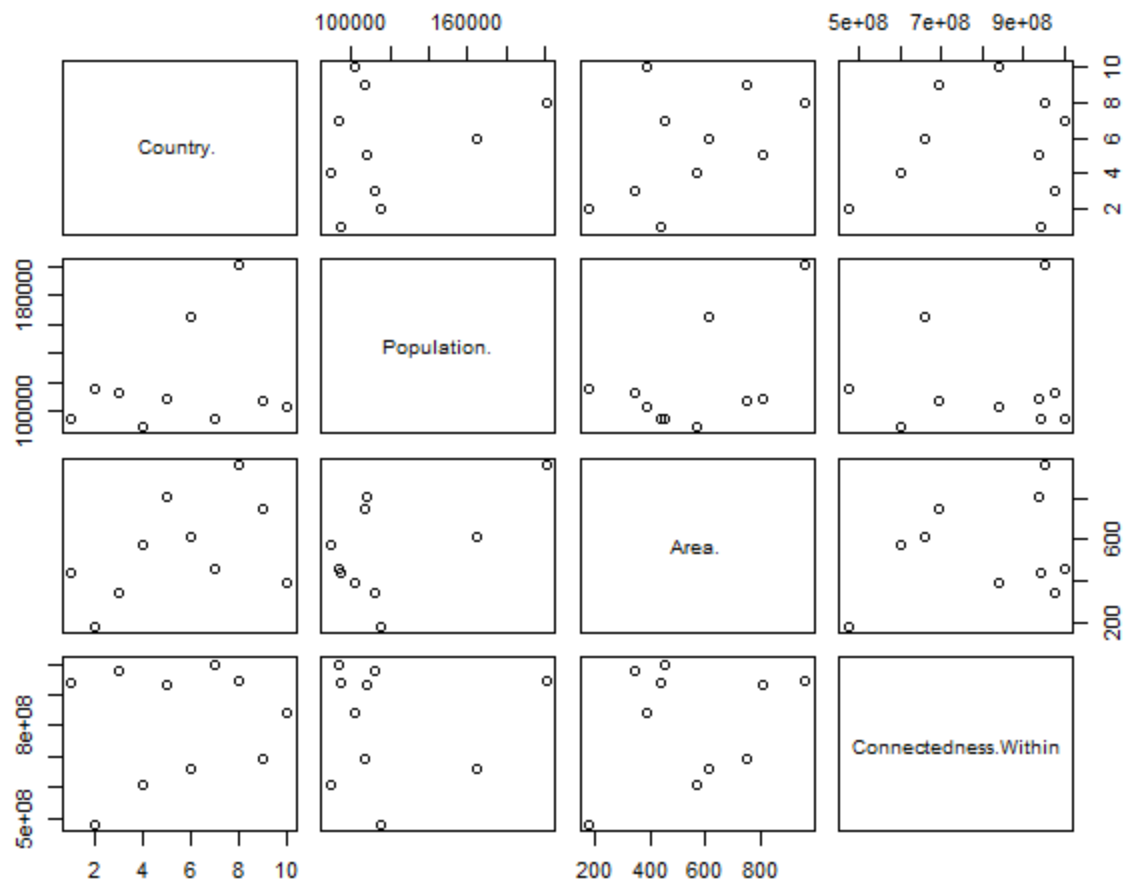Connectedness of some countries (clockwise: Costa Rica, New Zealand, Singapore, US)
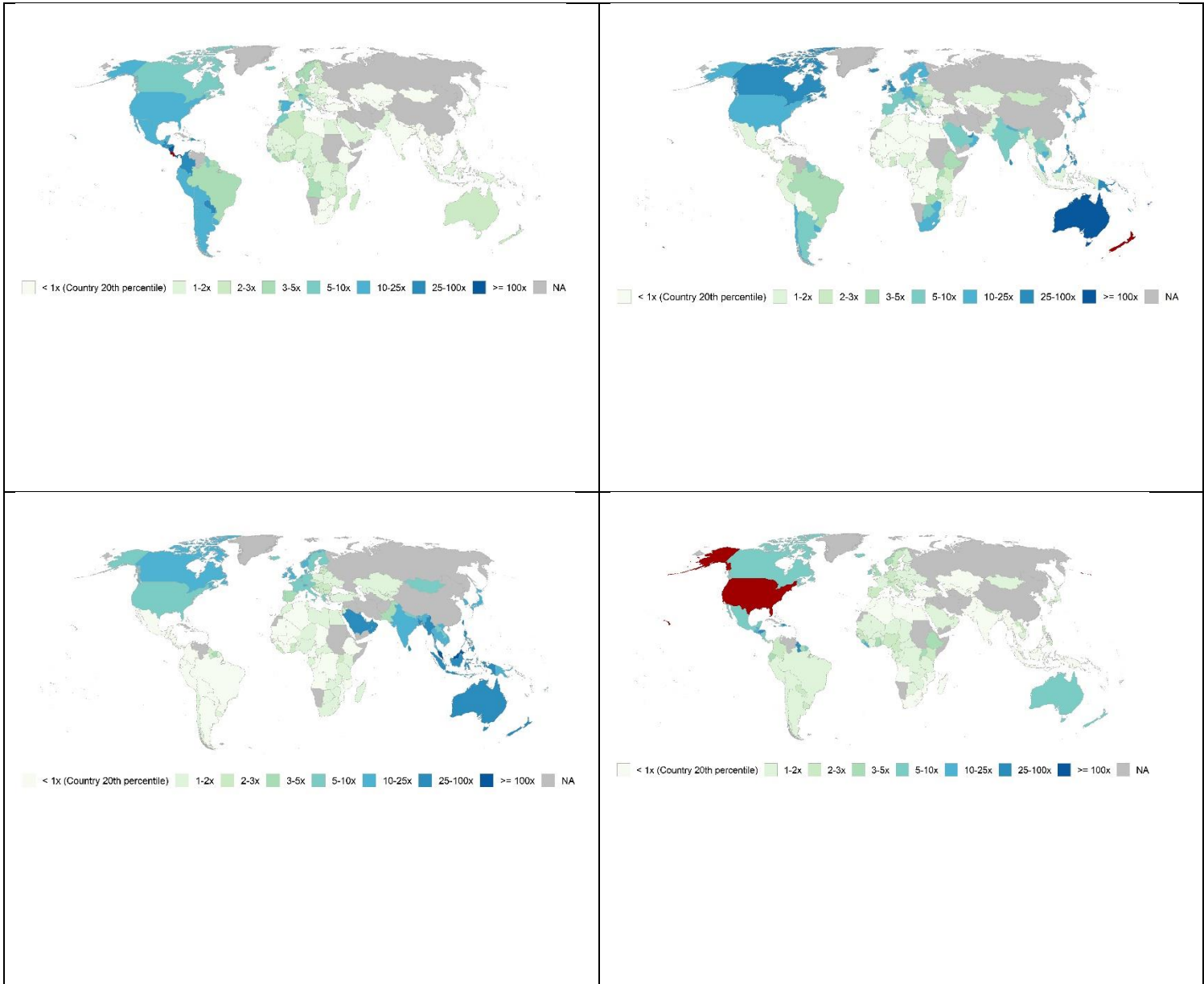
AW-AG
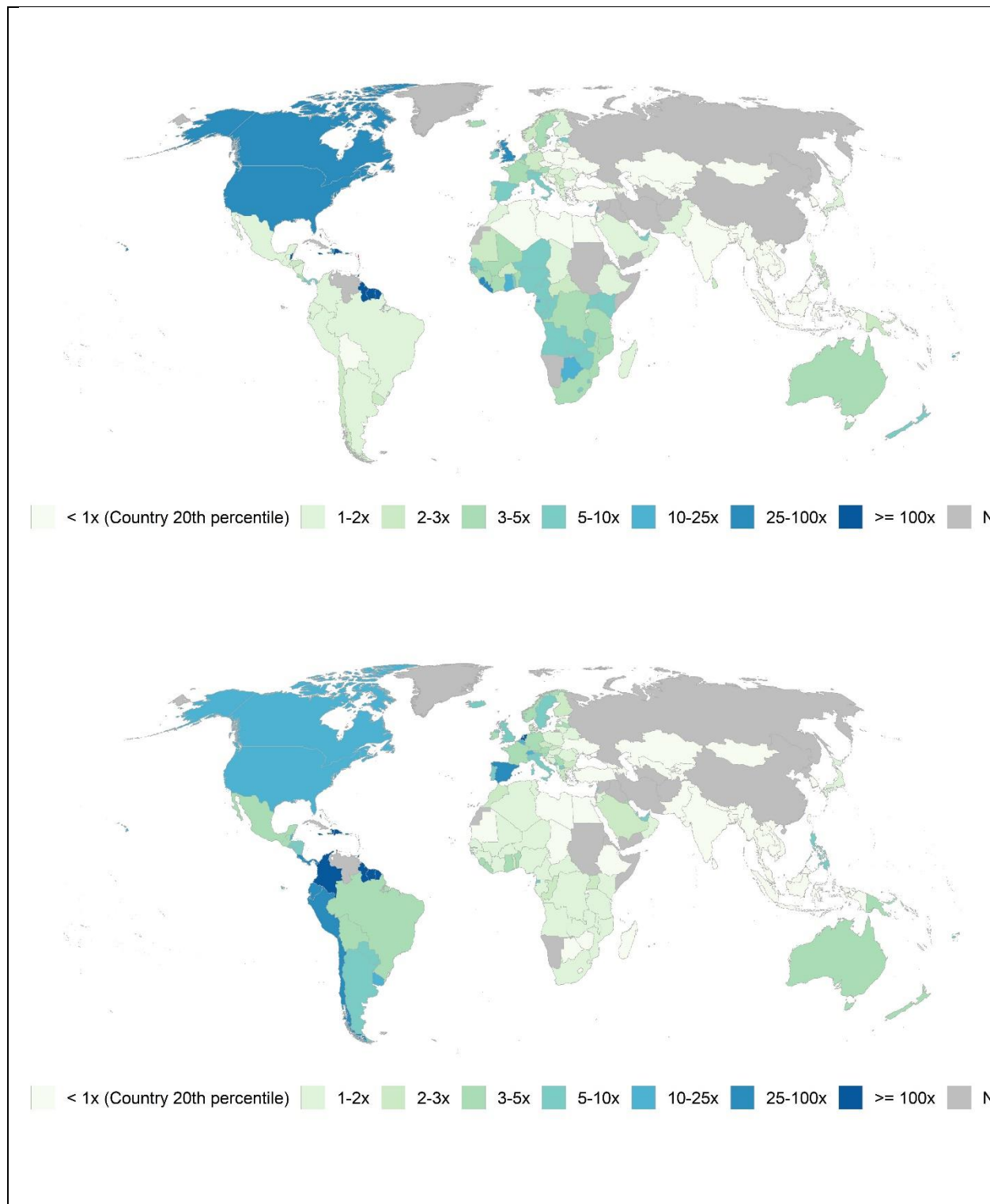
Databases:

Social Connectedness Index
Data Notes

Background and Construction of Social Connectedness Index

The *Social Connectedness Index* uses an anonymized snapshot of active Facebook users and their friendship networks to measure the intensity of social connectedness between locations. Users are assigned to locations based on their information and activity on Facebook, including the stated city on their Facebook profile, and device and connection information.

Formally, the *Social Connectedness Index* between two locations *i* and *j* is defined as:

$$Social\ Connectedness\ Index_{i,j} = \frac{FB\_Connections_{i,j}}{FB\_Users_i * FB\_Users_j}$$

Here, FB_Users_i and FB_Users_j are the number of Facebook users in locations *i* and *j*, and $FB\_Connections_{i,j}$ is the total number of Facebook friendship connections between individuals in the two locations.

The publicly available measures of the *Social Connectedness Index* are scaled within each dataset to have a maximum value of 1,000,000,000 and a minimum value of 1. As a result, the public release version of the $Social\ Connectedness\ Index_{i,j}$ measures the **relative probability of a Facebook friendship** link between a given Facebook user in location *i* and a given user in location *j*. Put differently, if this measure is twice as large, a given Facebook user in location *i* is about twice as likely to be connected with a given Facebook user in location *j*. We also add a small amount of random noise and round the *Social Connectedness Index* to the nearest integer, to ensure that no single individual or friendship link can be identified from the data.

This measure was first proposed and analyzed in Bailey, Cao, Kuchler, Stroebel, and Wong (2018), which explored the social connectedness across U.S. counties.

Data Included For Distribution

This folder includes the *Social Connectedness Index* calculated for different geographical areas as of August 2020. Each dataset includes every (symmetric) *i* to *j* and *j* to *i* location pair, including links of each location to itself.

Each dataset has three columns:

user_loc	First Location
fr_loc	Second Location
scaled_sci	Scaled SCI as described above

The datasets included contain the *Social Connectedness Index* for the following areas:

- **Countries – Countries.** Each row is a country – country pair. Countries are denoted by their ISO2 codes. Excludes certain countries, e.g., countries where Facebook is banned or countries with only few active users. Overall, we have 185 unique countries.

  This dataset was first introduced, described, and analyzed in Bailey, Gupta, Hillenbrand, Kuchler, Richmond and Stroebel (2020).

- **US Counties – US Counties.** Each row is a US county – US county pair. Counties are denoted by their 5-digit FIPS code. Excludes counties with few active users.

  This dataset was first introduced, described, and analyzed in Bailey, Cao, Kuchler, Stroebel, and Wong (2018).

- **US Counties - Countries.** Each row is a US county – country pair. Counties are denoted by their 5-digit FIPS code, countries are denoted by ISO2 code. Excludes counties and countries with few active users.

  This dataset was first introduced, described, and analyzed in Bailey, Cao, Kuchler, Stroebel, and Wong (2018).

- **GADM/NUTS – GADM/NUTS.** There are two files built on the Database of Global Administrative Areas (GADM, version 2.8) and the European Nomenclature of Territorial Units for Statistics (NUTS 2016) areas. Excludes regions with few active users.

  - **GADM1_NUTS2**: European countries are divided into their NUTS2 regions (e.g., 12 provinces in the Netherlands). Countries outside of Europe are divided into their GADM level 1 regions (e.g., states in USA). Countries with a population less than 1 million are not divided. Each row is a pair of these areas.

  - **GADM1_NUTS3_Counties**: European countries are divided into their NUTS3 regions (e.g., 40 regions in the Netherlands). The United States, Canada, and some countries in South Asia (Bangladesh, India, Nepal, Pakistan, and Sri Lanka) are divided into their GADM level 2 regions (e.g., US counties). Other countries are usually divided into their GADM level 1 region. Countries with a population less than 1 million are not divided. Each row is a pair of these areas.

  A separate set of files (gadm1_nuts2_levels and gadm1_nuts3_counties_levels) provide the levels of each of the keys in the GADM/NUTS files.

  Shape files for NUTS-level data and GADM-level data can be downloaded from:

  https://gadm.org/old_versions.html

  https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/administrative-units-statistical-units

  The GADM/NUTS-level dataset was first introduced, described, and analyzed in Bailey, Kuchler, Johnston, Russel, State, and Stroebel (2020).

Other Resources

Our academic partners on this project have put together a set of resources for individuals interested in working with these files, including code and shapefiles that easily map the social connectedness of different locations, and code that explores the relationship between the *Social Connectedness Index* and the spread of COVID-19 (see Kuchler, Russel, and Stroebel, 2020).

These files can be found at: https://github.com/social-connectedness-index/example-scripts

https://github.com/social-connectedness-index/example-scripts

References

- Bailey, M., Cao, R., Kuchler, T., Stroebel, J., & Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3), 259-80.

```
@article{bailey2018social,
  title={Social connectedness: Measurement, determinants, and effects},
  author={Bailey, Michael and Cao, Rachel and Kuchler, Theresa and Stroebel, Johannes and Wong, Arlene},
  journal={Journal of Economic Perspectives},
  volume={32},
  number={3},
  pages={259--80},
  year={2018}
}
```

- Bailey, M., Gupta, A., Hillenbrand, S., Kuchler, T., Richmond, R. J., & Stroebel, J. (2020). International trade and social connectedness. *NBER Working Paper*, (w26960)

```
@techreport{bailey2020international,
  title={International trade and social connectedness},
  author={Bailey, Michael and Gupta, Abhinav and Hillenbrand, Sebastian and Kuchler, Theresa and Richmond, Robert J and Stroebel, Johannes},
  year={2020},
  institution={National Bureau of Economic Research}
}
```

- Bailey, M., Kuchler, T., Johnston, D., Russel, D., State, B., & Stroebel, J. (2020). The Determinants of Social Connectedness in Europe. *SocInfo2020*

```
@InProceedings{bailey2020determinants,
  title={The Determinants of Social Connectedness in Europe},
  author={Bailey, Michael and Johnston, Drew and Kuchler, Theresa and Russel, Dominic and State, Bogdan and Stroebel, Johannes},
  booktitle={Social Informatics},
  publisher={Springer International Publishing},
  year={2020},
}
```

- Kuchler, T., Russel, D.,& Stroebel, J. (2020). The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook, NBER Working Paper, (w26990)

```
@techreport{kuchler2020geographic,
  title={The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook},
  author={Kuchler, Theresa and Russel, Dominic and Stroebel, Johannes},
  year={2020},
  institution={National Bureau of Economic Research}
}
```