## Instructions distances

### Main output file

The main outcome of interest is an R code that produces csv file (comma delimited) with three columns. 1) "iso_o": the identity of the region of origin (e.g., Alabama, AUS, etc). 2) "iso_d": the identity of the region of destination. and 3) "dist": the distance in km between the two regions. Each region is either a U.S state or a country. There are 50 states and 37 countries (including the rest of the world, or RoW). The final dataset has $87 \times 87$ observaciones (each region combined with each other region). Note that the distance between regions 1 and 2 is must be the same as the distance between regions 2 and 1 (please check that in the final data).

### Input files

1. "JPV_JIG\Trade\0-Raw_Data\Fips\us_states_coordinates_counties.xlsx": here you find the state code, the county identity, the longitude and latitude of the county, and the total population.

2. "JPV_JIG\Trade\1-Intermediate_Processed_Data\country_coordinates.dta": similar data as above but for countries.

3. "JPV_JIG\Trade\0-Raw_Data\Fips\state_codes.txt": state code and state names.

### Computation of distances

Define $dist_{ij}$ as the distance in km between region $i$ and $j$. The idea is to calculate the distance between two regions ($dist_{ij}$) based on bilateral distances between the cities and counties of those two regions, those inter-city/county distances being weighted by the share of the city in the overall country's population in 2010.

We use population for 2010 and coordinates data for all U.S. counties, and all cities around the world with more than 300,000 inhabitants (which are in the databases I shared). Coordinates are important to calculate the physical bilateral distances in kms between each county $r$ in state $i$ and county $s$ in state $j$ ($d_{rs}$), and define $dist_{ij}$ as:

$$dist_{ij} = \left( \sum_{r \in i} \sum_{s \in j} \left( \frac{pop_r}{pop_i} \right) \left( \frac{pop_s}{pop_j} \right) d_{rs}^{\theta} \right)^{1/\theta}$$

where $r \in i$ means all counties (cities) $r$ in state (country) $i$ and $pop_i$ is the population of place $i$. We set $\theta = -1$.