

Migration between Countries

Daniel Kaplan

CVC Workshop, July 2014

The World Bank maintains [data on migration between countries](#), based on censuses and other information. These have been translated to an R format as `MigrationFlows` in the `DCFdevel` package.

Here's a small sample of the data:

Referring to the sample, you can see that in year 2000 there were 24,118 males who moved from Poland to Kazakhstan. In that year, there were 2289 females who moved from Indonesia to Australia.

Basic Questions

Structure of data

- How many variables are there?
- Is `MigrationFlows` in “long” or “narrow” format with respect to the years?
- How many cases are there?
- How many countries are there?
- Construct a table that combines the females and males for each country pair.
- How many migrants originated (that is, emigrants) in each country in 2000?
 - Imagine what the table containing the result will look like. How many rows? Is the meaning of a row the same or different in the result than in `MigrationFlows`. Which variables will you use in from `MigrationFlows` to construct the table with the results? What will be the variables in the result table?
 - Hint: `sum()` to add numbers together, `group_by()`, `summarize()`
Use `head()` or `sample_n()` to display just a few rows of your result.
- Which 5 countries had the largest number of emigrants? (Hint: `arrange()`, `desc()`)
- Compute the fraction of each origin country's year-2000 emigration that goes to each destination country.
 - In the result table, what will be the cases? Are they the same as the original table? What will be the variables? Which ones are the same as in the original table.
 - Explain why `mutate()` and not `summarize()` was used here.
 - Is the `select()` necessary to finding the `outPercent` variable?
- For each origin country, what is the largest destination?
- For each destination country, what is the largest origin?
- For each destination country, what are the **two** largest origin countries?

Bringing in other data

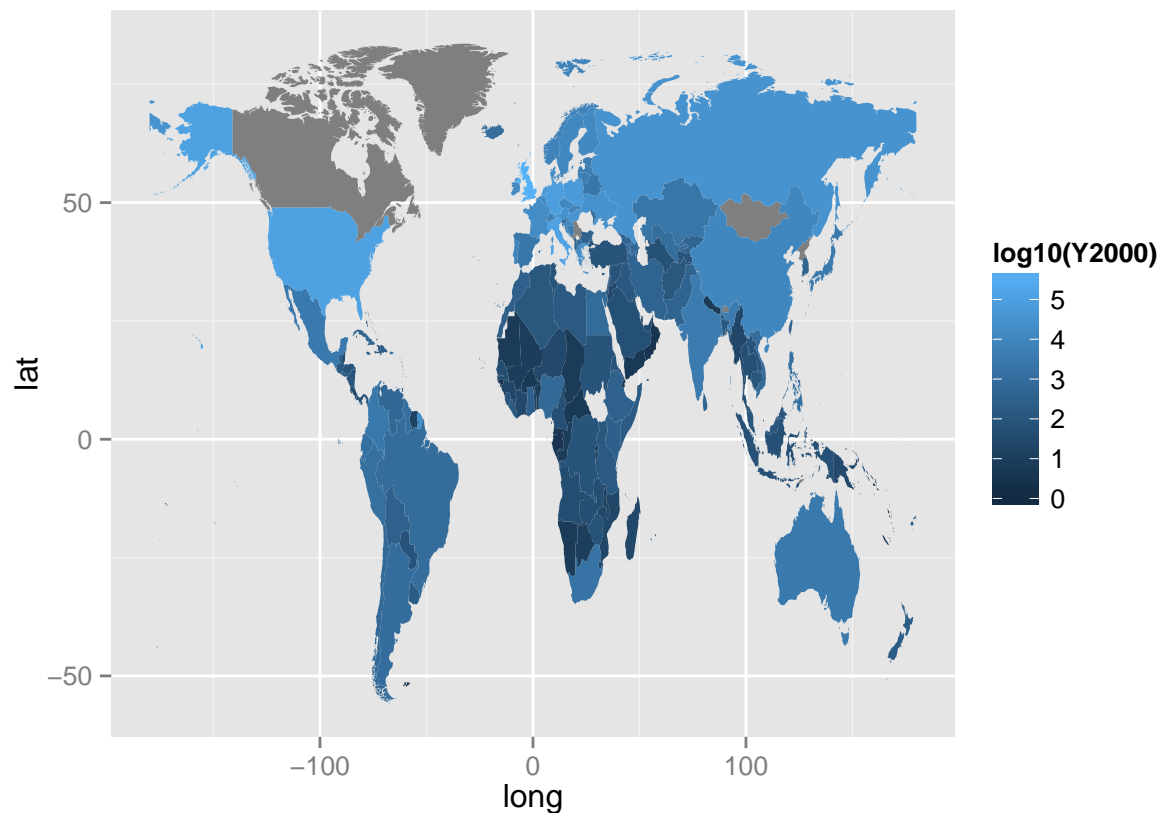
- Compute the per-capita emigration rate for each pair of countries. Use the population of the originating country to set the rate. Use the `CIAdata()` function to access population data and create a column `origincode` to make it easy to join the population data with the `MigrationFlows` data:

```
pop <- CIAdata(2119)
pop <- mutate(pop, origincode=toISO3(country))
```

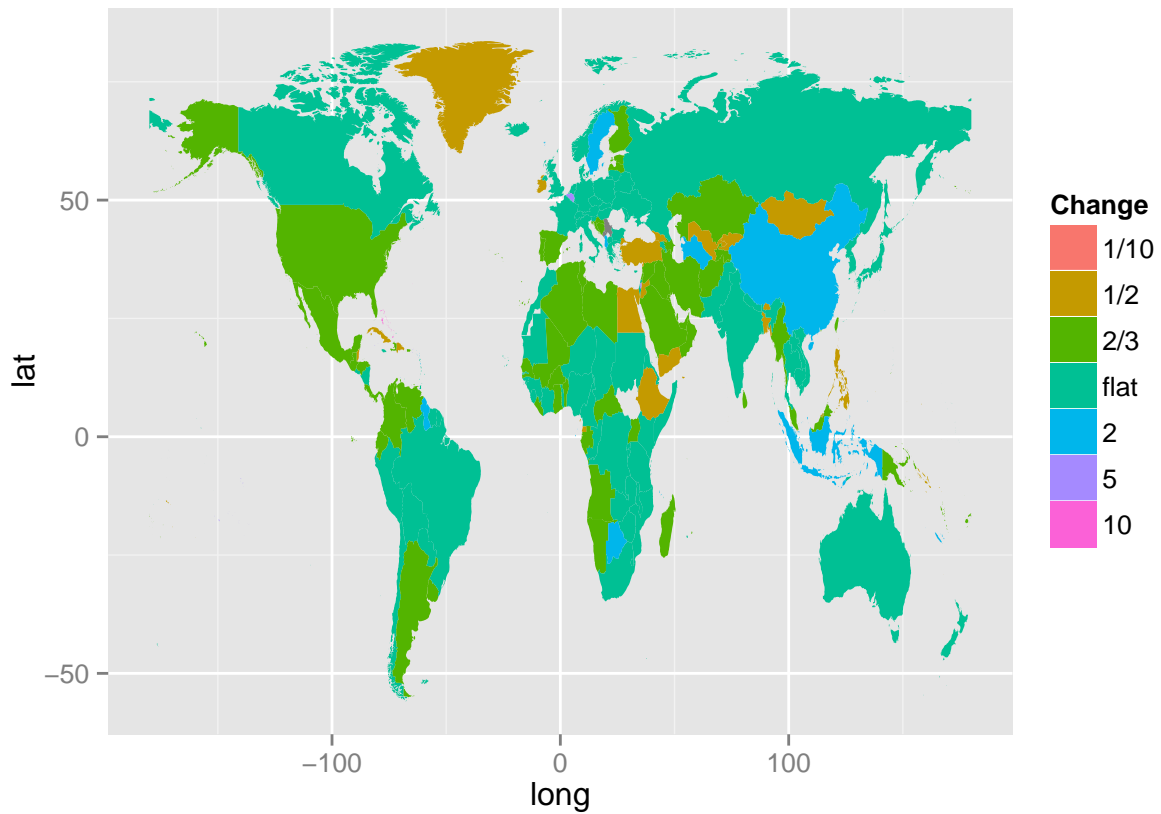
Then use a join to add the population data to the `MigrationFlows` data.

Maps

- Make a map of the countries whose people flow into a country of your choice. Here's the result for Canada. (Note: the logarithm of Y2000 is shown.)



- Construct a data table indicating for each country pair whether emigration increased or not between 1990 and 2000.
- Make a map showing the global pattern of the increase or decrease in emigration from each country.

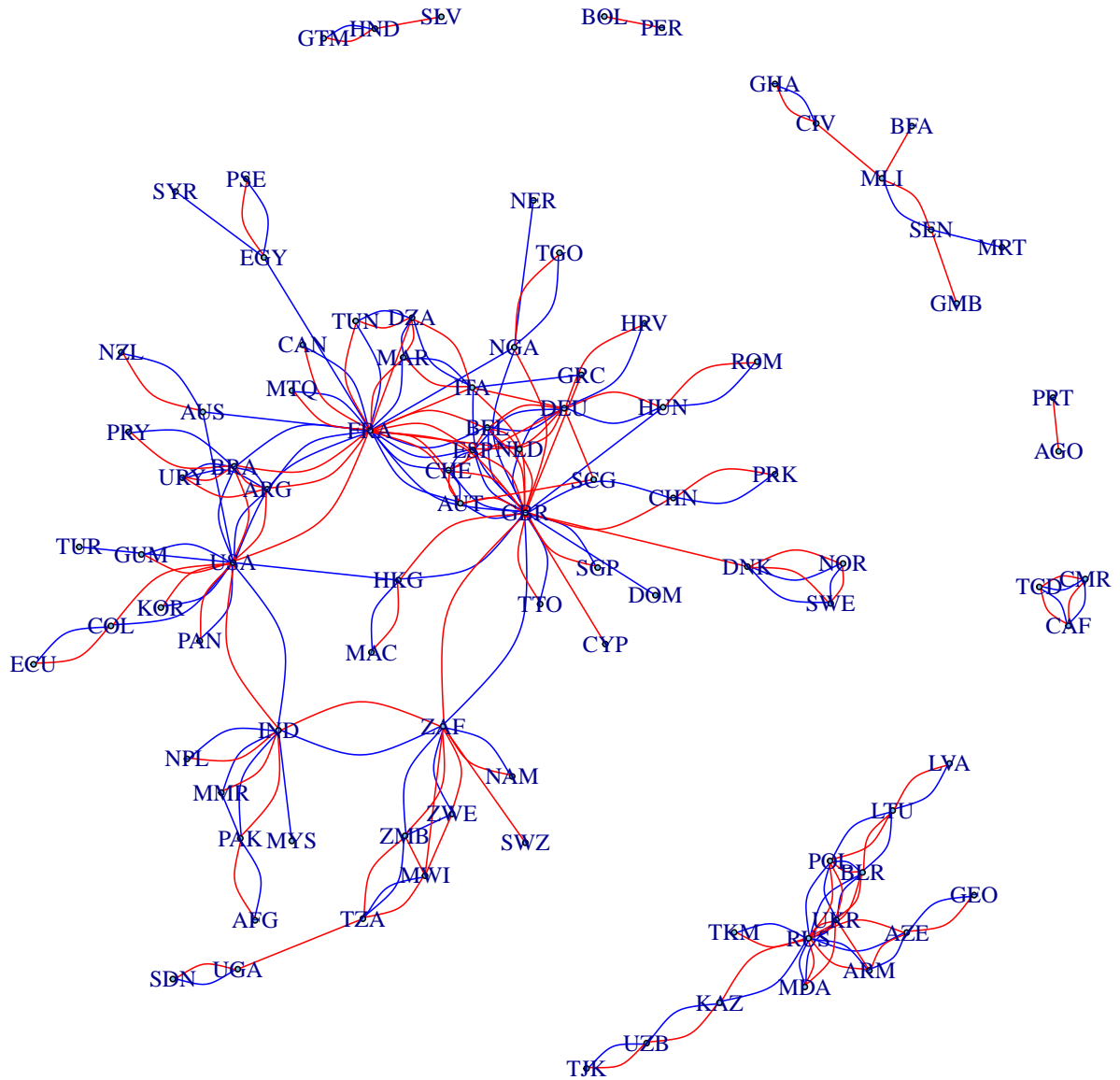


Networks

- Pull out the major country pairs with more than 1,000,000 migrants, male and female, over the 1960 to 2000 period.
- Imagine what the output data table will look like. Will the variables be different? Will the meaning of the cases be different?
- Make a directed network graph showing the flows among the major pairs.

On your own . . .

- For each pair of countries, find the ratio of immigration/emigration between them in year 2000.
- Select a couple of dozen pairs with the ratio most similar to 1. Make a network diagram showing the relationship.



Countries where the size of the exchange is balanced to within a factor of 3.3. Red is for females; blue for males.

- Which country pairs have a strongly male- or female-oriented emigration?