Visualizing Global International Migration Flow Data: From Tables to Stories

Abstract for a paper to be presented at the European Population Conference, Budapest, 2014.

Nikola D. Sander, Ramon Bauer and Guy J. Abel

Wittgenstein Centre for Demography and Global Human Capital, Wohllebengasse 12-14, 1040 Vienna, Austria

SHORT ABSTRACT

There is a long tradition in demography to display research findings using text and tables. While this approach allows data to be shown in a precise manner and provides an efficient means to look up values for a particular country or year, tables fail to highlight patterns and trends, and to tell the story behind the data. This is especially true for very complex datasets, such as migration flow tables. Despite all the progress that has been made regarding geographic information systems (GIS), migration flows are typically visualised using maps with arrows indicating the size and direction of flows. Such maps are low in complexity and visual appearance, allowing only a small number of flows to be shown. We present a new way to visualize complex migration flow data in an efficient and, indeed, beautiful manner. We draw on new estimates of 5-year bilateral flows between 196 countries from 1990 through 2010 that are estimated from the United Nations sequential stock tables by the third author. Our visualisation is open source and thus can readily be used to uncover the key patterns and trends underlying any kind of flow data, including trade and remittances flows. Specifically, we present three new methodologies for visualising flow data: First, we present circular plots created using Circos, software based on a Perl script and frequently used in genetics. Second, we show how these circular plots can be produced using R. And third, we use the Java Script library d3.js to interactively visualise global migration flows on the web, providing visually highly appealing graphics that invite the user to intuitively explore our new set of estimated global migration flows. Our novel representation has the potential to transform the way we evaluate temporal and spatial patterns in flow data.

EXTENDED ABSTRACT

Data visualization is not a new field. For centuries, we have relied on visualization to depict what words and tables cannot. In the 19th century, Charles Joseph Minard created a range of seminal visualizations of population movement, including Napoleon's March to Moscow from 1869 and The Emigrants of the World from 1858. Minard's migration flow map shown in Figure 1 depicts numbers and destinations of emigrations from Europe, Africa, China, and South Asia for the year 1858. The thickness of the flows indicated the volume of movement, with one millimetre equalling 1,500 people. Especially noteworthy is the color-coding system used by Minard, which aims to ease the identification of migrant's origins as indicated by the legend in the top-right corner.

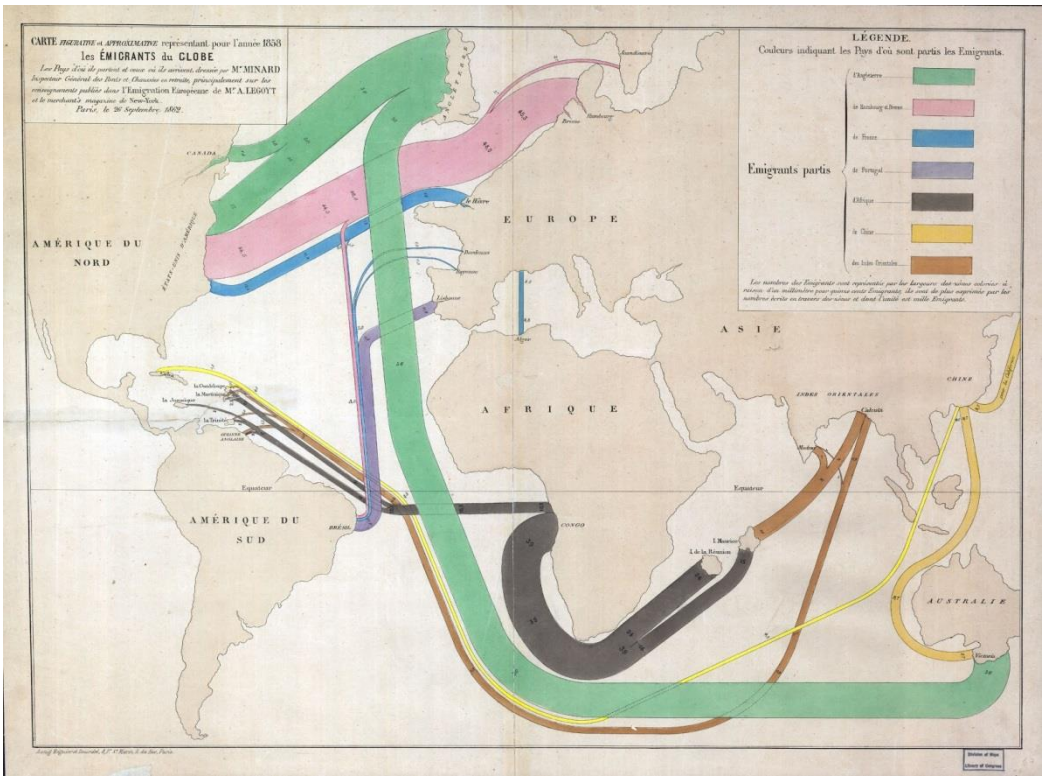

Figure 1. The Emigrants of the World, 1858. By Charles Joseph Minard

Since the seminal works by Minard, the visualisation of migration flows has not improved significantly, and Minard's color-coding scheme has not become widespread, despite substantial advances in the field of geographical information systems (GIS). Figure 2 gives an example of a typical flow map used in the geographic and demographic literature in recent decades. Fassmann and Münz (1992) chart the origins of German immigrants. Such maps suffer from lack of complexity and visual appearance, allowing only a small number of flows to be shown in a map where the exact identification of origins and destinations can be difficult, especially for small countries.

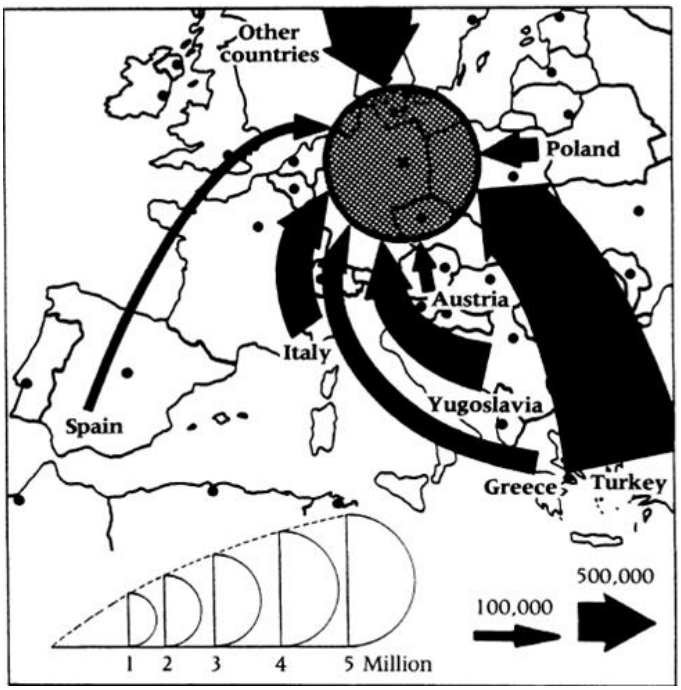

Figure 2. Foreign population in Germany, 1990. Source: Fassmann & Münz (1992)

This paper aims to fill that gap by introducing our new way to visualize migration flow data in an efficient and, indeed, beautiful manner. Our circular plots allow for the visual quantification of directional migration flows and the identification of spatial patterns. We use Minard's colour-coding scheme to indicate the origin and destination of migrants, and the width of the flows corresponds to their size. The circular plot in Figure 3 shows new global bilateral migration flows estimated from sequential United Nations stock tables. A more detailed discussion of these estimates is given in Abel (2013) and Abel and Sander (in review). Estimated quinquennial flows between 196 countries are aggregated into a more managable set of world regions, which are arranged in a circular layout. Each origin/destination region is assigned a unique colour, which is used to indicate the directionality of lows: As in Minard's graphics, flows have the same colour as their origin. The direction of the flows is also shown be the gap between flows and origin/destination region: the smaller gap denotes the origin; the larger gap denotes the destination. Tick marks show each region's (or country's) gross migration in millions. The circular plot in Figure 3 depicts the attractiveness of North America as migrant destination, the large-scale emigration from South and South-East Asia to North America and the oil-rich Gulf States, the dominance of within-region flows in Africa and the Former Soviet Union, the the high degree of connectedness of Europe with most other world regions through migration.

We first developed the circular plots using Circos, software based on a Perl script that was developed by Martin Krzywinski (2009) for the visualisation of genomes and chromosomes. We adapted the configuration files in Circos to effectively visualise directional migration flow data between up to 50 regions. In recent months, we have been working on replicating the circular plots using R.

Our novel graphics have the potential to transform the way we visualise, evaluate and present complex (migration) flow data.

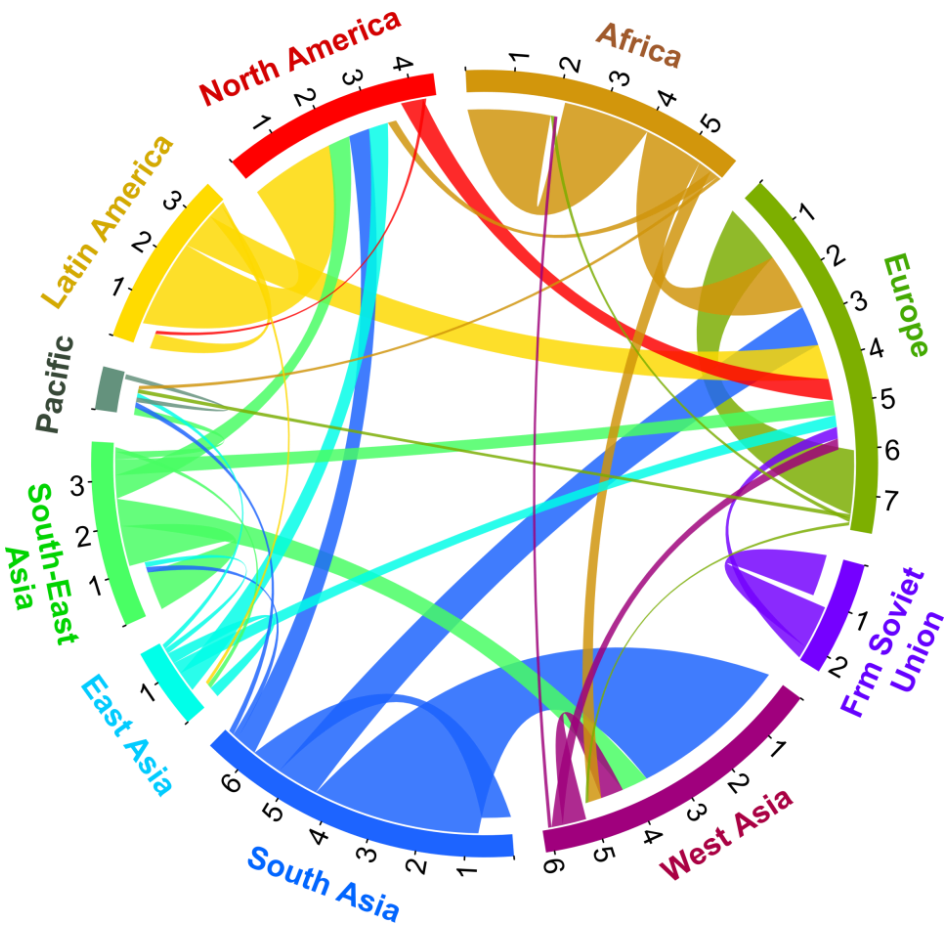

Figure 3. A circular plot showing international migration flows between world regions among males in 2005-10. Origins and destinations are represented by the circles segments. Each region is assigned a colour. Flows have the same colour as their origin and the width indicates their size. The direction of the flow is also shown by the gap between flow and region: the smaller gap denotes the origin; the larger gap denotes the destination. Only the largest 75% of all flows are shown. Tick marks show each country's gross migration in millions. (Data source: Estimates by Guy Abel)

The circular plots created using Circos and R are static and well suited for print media. In collaboration with Null2 in Berlin, we have developed a dynamic, interactive version of our circular plots that will soon be available under www.global-migration.info. Figure 4shows a screen shot of our interactive online visualisation of global migration flows for four periods from 1990-95 to 2005-10. People can explore migration at regional and country levels and can select individual regions/countries to highlight the migration flows they are most interested in. Tooltips indicate the numbers of migrants in each flow and the total gross migration in each region/country.

Figure 3. A screen shot of the interactive online visualisation of global international migration flows using the Java Script library D3. The user can switch between region and country view by clicking on the region's segment and between time periods by clicking on the blue buttons. The website will be publicly available at www.global-migration.info upon publication of the corresponding journal paper.

References

Abel, Guy J. 2013. Estimating global migration flow tables using place of birth data. Demographic Research, 28, 505–546.

Fassmann, Heinz, & Munz, Rainer. 1992. Patterns and Trends of International Migration in Western Europe. Population and Development Review, 18(3), 457.

Krzywinski, Martin, Schein, Jacqueline, Birol, ˙Inanc¸, Connors, Joseph, Gascoyne, Randy, Horsman, Doug, Jones, Steven J., & Marra, Marco A. 2009. Circos: An information aesthetic for comparative genomics. Genome Research, 19(9), 1639–1645.