Source: Boucher, Anna and Justin Gest. 2018. *Crossroads: Comparative Immigration Regimes in a World of Demographic Change*. New York: Cambridge University Press. Page 243-267.

Methodological Appendix

This appendix outlines our data sources, methodological choices, and any assumptions employed this book, including the non-OECD sources of data and our procedures for data collection, interpretation, and analysis.

Distribution of immigrant origin

In measuring the distribution of immigrant origin across states, our data are drawn for all countries from the United Nations Population Division (2013) database POP/DB/MIG/Stock/Rev.2013. As Chapter Three introduces this concept for the first time, as a conventional operationalization does not yet exist. We opt for two different approaches:

> **Herfindahl-Hirschman Index**: A conventional measure of market concentration in economics, a Herfindahl Index is typically calculated by squaring the market share of each company competing in a particular industry, and then summing the value outputs (Hirschman 1964). Here, we consider the percentage of destination state's total migrant

stock from different source countries. Whereas a Herfindahl Index of one would represent a perfect monopoly, this would be tantamount to one country being the sole source of a destination state's migrant stock. In the formula below, $\tilde{s}_{ij}$ is the share of immigrants from origin *j* in destination country *i* as a percentage of total stock of immigrants in country *i*. Higher values suggest the greater concentration of migrants from a few countries of origin—that the destination state has a less diversified stock of immigrants. (A similar calculation is employed by Czaika and de Haas 2014.)

$$H_i = \sum_j \tilde{s}_{ij}^2$$

Gravity Weighted Diversity Measure: The Gravity Measure attempts to combine the information in gross stocks of immigrant populations with the distance separating the origin-destination pair, in order to account for the geographic concentration or dispersion of a destination state's migrant stock from original source location. This metric first obtains pairwise distance values by calculating the space between the destination and origin states' capital cities (e.g. London to Dhaka; Singapore to Kuala Lumpur). The calculation sums the share of immigrants across the different origin countries weighting each by the reciprocal of the distance between the origin and destination squared. In the formula below, G_i is the gravity measure, d_{ij} is the distance between destination *i* and origin *j*, and s_{ij} is the stock of immigrants from country *j* in country *i*. For two countries with similar stocks of immigrants but one across a much smaller set of distances then the other, the former will have a higher value—denoting higher concentration. In this way,

higher values suggest that "gravity" must have a much greater force in determining the immigration patterns.

$$G_i = \sum_j \frac{1}{d_{ij}^2} s_{ij}$$

These calculations for Gravity Weighted Diversity Measure were based on UN stock data, broken down into country-specific stock values—all for 2013, which requires no imputation. Distance values were determined from the Geographic Distance Matrix Generator (Ersts 2013), which computes all pairwise distances from a list of geographic coordinates. UN estimates themselves are based on official statistics for the foreign-born or the foreign population, classified by sex and age (UNPD 2013). The statistics utilized to estimate the international migrant stock were mostly obtained from population censuses, but also population registers and nationally representative surveys (UNPD 2013). This means that these data typically do not account for the presence of undocumented migrants—a significant factor in some destination states.

Visa Mix and Temporary Ratio

Chapter Four (Visa Mix) relied upon the OECD's categorization of flows as either "temporary" or "permanent." In order to standardize data across countries, the OECD relied upon a reclassification process. For the settler states, this division between "temporary" and "permanent" is a fairly simple exercise as these are clearly delineated visa categories. For the

continental European states, the concept of permanent immigration is less applicable, given that migrants are generally tolerated on long-term temporary visas rather than first being admitted on a permanent basis as is common in settler states. For this reason, the OECD adopts a benchmarking exercise in determining whether a visa is permanent or not, rather than relying exclusively upon each domestic definition (Fron et al. 2008). According to this approach, when it is not immediately apparent whether entrants on a visa in practice remain permanently or not, the OECD asks whether over two-thirds of entrants on that visa remained after five years. If they did, the visa was categorized as a permanent visa, even if immigration officials labeled it "temporary." This re-categorization is important not only because it standardizes the definition of "temporary" across states but also because it captures those migrants who, while nominally on a temporary visa—potentially for political reasons—are in fact long-term residents.

It is important to note here that while the OECD data represent a gold standard with regards to both Visa Mix and arguably temporary work flows, aggregating these categories together to constitute total flows for the purposes of the denominator for each country is more challenging. The reason for this is that despite best efforts, statisticians may double count permanent and temporary records because some immigrants will adjust from temporary to permanent visa status (Thomas Liebig 2012, OECD Migration Section, pers. comm.). This renders direct comparison between temporary and permanent categories difficult. Given this marginal inaccuracy, aggregate measures of total flows are not considered as a stand-alone dimension within this book. In Chapter 4, we report Visa Mix for permanent visas where possible and where this is not possible (for the Gulf countries, for Singapore in some instances and for China, which do not have permanent forms of migration), we report Visa Mix for temporary visas only. For the purposes of comparison, we combine temporary and permanent

flows as our denominator of interest to calculate percentages of Visa Mix and for the Temporary Ratio in Chapter 5. In Chapter 5, the numerator is only temporary economic immigration as the OECD does not provide consistent flow data on non-economic forms of temporary immigration.

In this section, we discuss the sources of non-OECD data utilized in the Visa Mix and Temporary Ratio (Chapters Four and Five).

Bahrain

Generally, data for Bahrain were gained from the Gulf Labour Markets and Migration (GLLM) website. We relied upon the new visa data sourced from GLMM (2014a), "Bahrain: New Visa Renewals and Terminations," available online at: http://gulfmigration.eu/bahrain-new-visas-renewals-and-terminations-by-type-of-visa-q3-2008-q1-2014 in http://gulfmigration.eu. Originally from Labour Market Regulatory Authority, Manama, Bahrain.

For the Visa Mix (Chapter Four), the categories "Employer," "Investor," and "Temporary Worker" were coded as our category of "work" while the category "dependent" was coded as our category of "accompanying family" and therefore subsumed within "work." For the purposes of Temporary Ratio (Chapter Five), all economic immigration into Bahrain was treated as temporary. We defend this characterization of all immigration into Bahrain as temporary in Chapter Five. As the Bahrain data differentiated between new visas and renewals, we only utilized new visas.

China

For China, we rely upon data from the Bureau of Entry and Exit Administration of the Ministry of Public Security, 2010. These data are only available for the calendar year of 2009 and published in 2010. The Chinese government does not reveal humanitarian data, so refugee admission is not considered.[46] Available data are aggregated into categories that reflect the OECD classifications. "Business and employment" are combined to represent "work," settlement and others to reflect "other." "Family" represents family reunification as with the OECD data. Further, as the flows in some of these categories are so small compared to the other categories and other countries, they are not reflected in Figure 4.1. The category of "international study" is removed from the Chinese data as the OECD data only considers international study for temporary immigration. Although there were considerable levels of entry through the family category in China in 2009 (80,058 persons), the overall scale of total immigration flows (21,924,427) renders the percentage for this category at close to zero. These flow data were preferred over census data as according to an official in the Shanghai Bureau of Statistics, these census data provide an underestimate (see China Bureau of Entry and Exit 2011).

Aside from international study, other forms of immigration of a temporary nature into China were available, but not that differentiated across visa categories. Therefore, these data were excluded from the Temporary Ratio (Chapter Five). Instead, we assume a one-hundred percent rate of temporary economic immigration into China. This decision is corroborated by available evidence about the scarcity of permanent residency opportunities for migrants in China. Around a quarter of international migrants stay for more than five years (25.4%) (National Bureau of Statistics 2010). In the rare event that a migrant holds a "D" Visa, applying for a

[46] The CIA World Factbook estimates that 300,697 Vietnamese, and between 30,000 to 50,000 North Korean refugees reside in China (CIA 2010). Ninety-eight percent of China's Vietnamese refugees are ethnically Chinese, and arrived in the aftermath of Vietnam's 1979 invasion of Cambodia. Despite the fact that they have lived in China for over 30 years, these migrants do not have formal citizenship (Jing 2007).

Chinese "Green Card"/"Green Booklet" allows one to attain residency that lasts up to ten years. The D-Visa system was only instituted in 2004, and is highly selective. Further, restrictions on the fulfillment of eligibility criteria prevent the majority of long-term international migrants from seeking permanent residency (Zou 2012). As of 2011, only 4,752 Chinese Green Cards have been awarded to foreigners since the program's creation in 2004 (Zhang 2011; Lu 2012).

Kuwait

Data for Kuwait are drawn from the GLLM website, in particular: Residence permits by type and purposes of permit, 2009-2012, sourced from the Ministry of the Interior (GLMM 2013b), available online at: http://gulfmigration.eu/residence-permits-by-type-purpose-of-permit-and-sex-of-holder-2009-2011-2012 in http://gulfmigration.eu. Originally this data was acquired from the Ministry of Interior in Kuwait City, Kuwait.

For Visa Mix (Chapter Four), we took the aggregate of both sexes to calculate data across the various categories. The work category was calculated as encompassing the following categories: "temporary permits," "governmental sector permits," "private sector permits," "business," "domestic help" and "dependent permits." The "study" category was omitted as we do not consider student visas in the Visa Mix. Self-residence permits were coded as "other."

For the Temporary Ratio (Chapter Five) we took: "temporary permits," "governmental sector permits," "private sector permits," "business," and "domestic help" as an aggregate for temporary economic. There is no permanent immigration into Kuwait and therefore we coded permanent economic immigration in Chapter Five as zero. We defend this decision in the chapter.

Latin American states

Visa Mix data for the Latin American states except for Mexico (which is an OECD country) are sourced from a recent data collection exercise by the Organization of American States, the International Development Bank, and the OECD. To provide consistency with the OECD data, any measures of international study or regularization collected for the Latin American states are taken out of the total measures for analysis, as these categories are not comparable across the remaining countries (OECD/IDB/OAS 2012). The report these organizations produce undertakes the same OECD standardization method for countries in that region (OECD/IDB/OAS 2012), which is necessary for Chapters Four and Five.

In the OECD/IDB/OAS report, regularization was included as a unique category for Brazil. Given its singularity, we therefore collapsed it into the "other" category. No description of the nature of international agreements was provided for Argentina, Colombia, or Peru. As such, the comparability of international agreement data across these states is unclear. However, this is the best available current data for our purposes and we therefore assumed that "international agreement" was a coherent category across these countries and included it under "Other" for the purposes of our coding for Visa Mix. No permanent Visa Mix data was available for Colombia. Supplementing the IDB sources, some additional data for the Latin American states were gained from the following sources:

For Brazil, refugee data were gained from the National Immigration Council, Ministry of Labor and Employment, available online at: maisemprego.mte.gov.br/portal/pages/trabalhador.xhtml.

Oman

For Oman, data are drawn from the GLLM website; in particular: GLLM (2014b), "Oman: Residence permits by type of permit (employment/family reunion/domestic labour) (2007-2013)," available online at: http://gulfmigration.eu/oman-residence-permits-by-type-of-permit-employment-family-reuniondomestic-labour-2007-2013 in gulfmigration.eu. Originally from Royal Oman Police, Muscat, Oman.

For Visa Mix (Chapter Four), "Employment," "Domestic Servant," and "Business" were categorized as "work," and "family joining" was categorized in the "accompanying family" group and therefore incorporated into the Work category as well. For Temporary Ratio (Chapter Five), the categories "Employment," "Domestic Servant," and "Business" together constitute total "temporary" economic migration. Like Kuwait, there is no permanent immigration into Oman and therefore we coded permanent economic immigration in Chapter Five as zero. We defend this decision in Chapter Five. It is important to note that some of these data could include visa renewals as the source data for Oman does not distinguish between new issues, renewals and transfers of sponsorships. However, we could not access data that provided further differentiation between these categories.

Saudi Arabia

Data for Saudi Arabia are drawn from the GLLM website; in particular, "Residency permits issued by purpose of entry (Saudi Arabia 1984-2011)" (2013c) and "Permits issued by type (residence/ work) (Saudi Arabia, 1984-2011)" (2013b), available online at: http://gulfmigration.eu/permits-issued-by-type-residence-work-saudi-arabia-1984-2011 in http://gulfmigration.eu. Originally from Ministry of Interior, Riyadh, Saudi Arabia.

For Visa Mix (Chapter Four), the work category includes residency permits issued to "workers" and "accompanying persons." For the Temporary Ratio (Chapter Five), the temporary economic category includes "workers." As for the other GCC countries, we defend the decision to code all entry as temporary in Chapter Five.

Singapore

We could not include Singapore in the Visa Mix (Chapter Four) as we did not have access to complete data across the different categories of the immigration mix (e.g. work, family and humanitarian). For Singapore for Chapter 5, no flow data were available and for this reason we relied upon a range of flow and converted stock to flow data, using the following formula adopted from Roberts and Camarena (2012):

$$\text{Stock} > \text{flow}$$

Estimates and methodology:

$$\text{Flow}(t) = \text{Stock}(t) - \text{Stock}(t-1)$$

In order to make these calculations, for the Temporary Ratio (Chapter Five), temporary stock data were acquired for 2007-2012 from Ministry of Manpower Singapore (2015), "Foreign workforce numbers," available online at: http://www.mom.gov.sg/statistics-

publications/others/statistics/Pages/ForeignWorkforceNumbers.aspx and converted to flow data. Temporary migration was defined as (Total Temporary Migration) minus the number of foreign workforce, which comprised of the number of economic visas according to permit type, including Employment Pass (EP), S Pass and Work Permit).

We did not use data from before 2011 for historical analysis in Chapter Five as we did not have full data for the denominator (total migration flows) prior to that year. Further, according to Singaporean national experts, data prior to this period are unreliable due to a considerable change in admissions policy in that country between 2006-2010 (Nadica Pavlovska, pers. comm.).

South Africa

For South Africa, we relied on published permanent immigration data (Statistics South Africa 2011). These data are presented with the tables on OECD countries below, as they comprise permanent immigration flows, which are analogous to counterparts across the OECD. Flow data were only available in standardizable fashion for 2011 and were drawn from the following document: Statistics South Africa (2011/2012), "Documented immigrants in South Africa." Our decision to focus upon 2012 is consistent with the analysis of Budlender (2013), which makes clear that other immigration data sources and earlier data sources in South Africa are unreliable for a number of reasons, rendering historical analysis impossible. For the reference year 2011 for the South African data, the original South African coding of "Relations" was recoded as Family; "Refugees" as Humanitarian; and "Retired," a category that appears to serve wealthy European and South Korean retirees, as "Other." "Business and Work" were coded together as Work.

Data source years

For Chapter Seven, our reference year was 2011 for all OECD data (unless missing, in which case the most recent year was used). For all other countries we used 2011 except for Mexico (2010), China (2009), South Africa (2012), and Singapore (2011-2012), as these were the most proximate available years to the reference year. For the stock and origin diversity measures in Chapter Three, we used 2013 data. For Naturalization Rates for all countries, we used 2011 unless otherwise noted. More details are provided below.

Naturalization data

As we discuss in Chapter Six, there are a number of calculations that different scholars call "Naturalization Rates." The most common measure of Naturalization Rates and the one we use is the annual naturalization flow rate. This is found by dividing the number of naturalizations that occur in a given country in a given year by the population of foreign citizens in that country at the beginning of that year.

Where:

R_Y = Naturalization Rate at time Y

N_Y = Total naturalized aliens in year Y

S_Y = Total migrant stock S calculated in year Y = Foreigners who are residents of Country 1 that are "at risk of naturalization." Not merely the foreign-born population because many foreign-born people have already naturalized, and because some foreign born are *jus sanguinis* citizens born abroad.

$$\text{Naturalization Rate } R_Y = \frac{N_Y}{S_Y} = \frac{\text{Total naturalized aliens in year } Y}{\text{Total migrant stock } S \text{ calculated in year } Y}$$

This measure of Naturalization Rates is a better reflection of current citizenship policy than stock rates alone (Reichel 2012: 3) and is used widely (Clarke et al. 1998; Bloemraad 2006; Howard 2009; Janoski 2010; Reichel 2012). In order to contextualize the *jus soli* adjustments made by Janoski, we compare with OECD (2012) data.

As explained in Chapter Six, in order to compare Janoski's adjusted naturalization rates to the raw OECD rates, we calculated the unadjusted rates for 2000-05 in countries with *jus soli* policies and found the ratio between the OECD data and the Janoski data. Our aim here was to ensure that Janoski's data did not differ substantially from the OECD data for countries without a *jus soli* policy, and to gain an intuition about the magnitude of the effect *jus soli* has in countries

that use it. We were able to complete this calculation primarily using the OECD naturalization rates, but for Canada, Australia, and France, we were forced to make estimations based on national statistics office data. The process we used for obtaining that historical data is outlined below. After conducting these comparisons, we deemed Janoski's data credible and utilized them for the *jus soli* countries in our dataset. We also used these data in our record of historic naturalization rates.

Canada only collects data on its foreign citizen stock in its quinquennial census. We therefore used the 2001 and 2006 census data as reported in OECD (2013) to estimate foreign stock for 1999, 2000, and 2002-2005 assuming linear growth over an annual period. Given that this population increased by 190,225 from 2001 to 2006 and 198,415 from 2006 to 2011 we believe that this is a reasonable approximation (ibid.). We then divided the OECD's raw citizenship acquisition data by the estimated foreign population in the prior year to obtain Naturalization Rates from 2000 to 2005, which we then averaged.

We used Australian census data from 2001 and 2006 to calculate the difference between the total population and Australian citizens, which we assume roughly corresponds to the number of foreigners. We then used the same method we applied to Canada to calculate estimated foreign stocks and Naturalization Rates from 2000 to 2005. The raw data come from the Australian Bureau of Statistics (2001; 2006).

Like Australia and Canada, France's rate is based on estimations of foreign stocks assuming linear growth. We used the foreign stock as reported in the OECD International Migration Outlook (2010) for 1999 and 2006 to interpolate foreign stocks for 2000 to 2004, which we used with the OECD's raw citizenship acquisition numbers to calculate estimated Naturalization Rates. Eurostat has published foreign stocks for France for 1 January 2003 and 1

January 2005, which we also used to calculate Naturalization Rates for 2003 and 2005. We did not use these stocks for 2004 and 2006, since that data was specifically presented on 1 January, and therefore meets Reichel's (2012) requirement that foreign stock at the beginning of the year should serve as the denominator for Naturalization Rates (see footnote 1 in Chapter Six for more detail). Using these Eurostat-based rates for 2003 and 2005 did not change the average rate from 4.3%, so we used our estimates based on OECD data for consistency. The raw data came from Eurostat (2016), "Population on 1 January by age group, sex, and citizenship," available online at: https://data.europa.eu/euodp/data/dataset/DyCiBSvR4z283JjDuwvdAQ.

After ensuring the validity of the Janoski adjustments, we assembled our Naturalization Rate dataset. Our default source was 2011 data from OECD (2014). The following exceptions should be noted: Australia, Belgium, Canada, France, Germany, Ireland, New Zealand, United Kingdom, and United States Naturalization Rates are unpublished 2011 data from Tom Janoski (2016) containing *jus soli* adjustments.

For Brazil, Colombia, and Peru we used: OECD/IDB/OAS (2012), *International Migration in the Americas: Second Report of the Continuous Reporting System on International Migration in the Americas (SICREMI)*. These countries do not publish consistent foreign stock data, forcing us to use the UNPD (2015) foreign-born stock data as the denominator for Naturalization Rates in these countries. Like for Canada and Australia, we estimated the denominator in years without data assuming linear growth between available data points. We then divided the raw acquisition of citizenship data by the prior year's foreign-born stock estimate to get an estimate of the Naturalization Rate in these countries for that year.

For Bahrain, 2010 data are based on a qualitative estimate of 60,000 naturalizations between 2001 and 2011 (which we assumed occurred at a rate of 6,000 per year) and stock data

from its national statistics office: see Baker (2011) for the qualitative estimate and stock data from GLMM (2014c), "Population estimates by nationality (Bahraini/Non-Bahraini) (mid-year estimates, 1981; 1990-2011)," available online at: http://gulfmigration.eu/population-estimates-by-nationality-bahraininon-bahraini-mid-year-estimates1981-1990-2011 in http://gulfmigration.eu. Originally from Central Informatics Organization, Manama, Bahrain.

For Kuwait, all naturalization data and foreign stock are drawn from: Kuwait Central Statistics Bureau (2007; 2010; 2012), *Annual Statistical Abstract*, available online at: https://www.csb.gov.kw/Socan_Statistic_EN.aspx?ID=18. In the 2010 and 2012 reports see Tables 19 and 29. In the 2010 report, we referred to Tables 19 and 30.

Data for China, Oman, Qatar, Saudi Arabia, Singapore, and the United Arab Emirates are all based on estimates of a 0% Naturalization Rate according to consultations with personnel from national statistics offices and news sources about policy standards. For China, the 2010 estimate is also based on government reporting on naturalized stock (National Bureau of Statistics of China 2010) and evidence that pathways to attaining Chinese citizenship are all but impossible, and attaining permanent residency is a special privilege of a select few (National Bureau of Statistics China 2010).

K-means clustering analysis

For the K-means clustering analysis we standardized our variable data and used the kmeans() package in R to calculate the clusters. We use twenty-five random initial points as cluster centers and ran one-hundred iterations of the algorithm. Because the user needs to define the number of

clusters, k, we ran the algorithm with two through nine clusters to analyze the results. There are a number of fit measures that can be analyzed to help choose clusters. First, we plotted the within-group sum of squares against number of clusters to look for an "elbow" where additional clusters do little to improve fit. This was largely unhelpful, as no clear elbow emerged.

We also used the NbClust() package in R, which provides thirty indexes to determine the optimal number of clusters in a dataset. Here we display two- to nine-cluster solutions.

Table A1: Global Cluster Solutions

2 Clusters	3 Clusters	4 Clusters	5 Clusters	6 Clusters	7 Cluster	8 Clusters	9 Clusters
Australia	Brazil	Austria	Bahrain	Finland	Brazil	Brazil	Bahrain
Bahrain	Japan	Belgium	China	Sweden	Japan	South Korea	Kuwait
China	Mexico	Denmark	Kuwait	United States	Mexico	Finland	Oman
Kuwait	South Korea	France	Oman	Brazil	South Korea	Sweden	Saudi Arabia
New Zealand	Australia	Germany	Russia	Japan	Australia	United States	Brazil
Oman	Bahrain	Ireland	Saudi Arabia	Mexico	Canada	China	South Korea
Russia	China	Italy	Singapore	South Korea	New Zealand	Japan	Finland
Saudi Arabia	Kuwait	Netherlands	Austria	Bahrain	United Kingdom	Mexico	Sweden
Singapore	New Zealand	Norway	Denmark	China	Austria	Belgium	United States
Austria	Oman	Portugal	Germany	Kuwait	Denmark	France	Russia
Belgium	Russia	Spain	Norway	Oman	Germany	Ireland	Singapore
Brazil	Saudi Arabia	Switzerland	Switzerland	Russia	Netherlands	Italy	Australia
Canada	Singapore	Brazil	Australia	Saudi Arabia	Norway	Portugal	Canada
Denmark	Austria	Japan	Canada	Singapore	Switzerland	Spain	New Zealand
Finland	Belgium	Mexico	New Zealand	Australia	Belgium	Bahrain	United Kingdom
France	Canada	South Korea	United Kingdom	Canada	France	Kuwait	China
Germany	Denmark	Australia	Belgium	New Zealand	Ireland	Oman	Japan
Ireland	Finland	Canada	Finland	United Kingdom	Italy	Saudi Arabia	Mexico
Italy	France	Finland	France	Austria	Portugal	Australia	Belgium
Japan	Germany	New Zealand	Ireland	Denmark	Spain	Canada	France
Mexico	Ireland	Sweden	Italy	Germany	China	New Zealand	Ireland
Netherlands	Italy	United Kingdom	Netherlands	Netherlands	Russia	United Kingdom	Netherlands
Norway	Netherlands	United States	Portugal	Norway	Singapore	Russia	Norway
Portugal	Norway	Bahrain	Spain	Switzerland	Bahrain	Singapore	Austria
South Korea	Portugal	China	Sweden	Belgium	Kuwait	Austria	Denmark

Spain	Spain	Kuwait	United States	France	Oman	Denmark	Germany
Sweden	Sweden	Oman	Brazil	Ireland	Saudi Arabia	Germany	Switzerland
Switzerland	Switzerland	Russia	Japan	Italy	Finland	Netherlands	Italy
United Kingdom	United Kingdom	Saudi Arabia	Mexico	Portugal	Sweden	Norway	Portugal
United States	United States	Singapore	South Korea	Spain	United States	Switzerland	Spain

Table A.2: OECD Cluster Solutions

2 Clusters	3 Clusters	4 Clusters	5 Clusters	6 Clusters	7 Clusters	8 Clusters	9 Clusters
Australia	Japan	Belgium	Australia	Australia	France	Germany	Austria
Canada	Mexico	Finland	Canada	Canada	Ireland	Switzerland	Belgium
Denmark	South Korea	France	New Zealand	New Zealand	Italy	Japan	Netherlands
Germany	Austria	Ireland	United Kingdom	United Kingdom	Portugal	Mexico	Norway
Japan	Belgium	Italy	Austria	Finland	Australia	South Korea	Italy
Mexico	Denmark	Netherlands	Denmark	Sweden	Canada	France	Spain
New Zealand	France	Portugal	Germany	United States	New Zealand	Ireland	Japan
South Korea	Germany	Spain	Netherlands	Italy	United Kingdom	Portugal	Mexico
Switzerland	Ireland	Sweden	Norway	Portugal	Austria	Australia	South Korea
United Kingdom	Italy	United States	Switzerland	Spain	Belgium	Canada	Finland
Austria	Netherlands	Japan	Finland	Austria	Netherlands	New Zealand	Sweden
Belgium	Norway	Mexico	Sweden	Denmark	Norway	United Kingdom	United States
Finland	Portugal	South Korea	United States	Germany	Denmark	Finland	Germany
France	Spain	Austria	Belgium	Switzerland	Spain	Sweden	Switzerland
Ireland	Switzerland	Denmark	France	Japan	Germany	United States	Denmark
Italy	Australia	Germany	Ireland	Mexico	Switzerland	Italy	Australia
Netherlands	Canada	Norway	Italy	South Korea	Finland	Spain	New Zealand
Norway	Finland	Switzerland	Portugal	Belgium	Sweden	Denmark	France
Portugal	New Zealand	Australia	Spain	France	United States	Austria	Ireland
Spain	Sweden	Canada	Japan	Ireland	Japan	Belgium	Portugal
Sweden	United Kingdom	New Zealand	Mexico	Netherlands	Mexico	Netherlands	Canada
United States	United States	United Kingdom	South Korea	Norway	South Korea	Norway	United Kingdom

We defend our selection of a seven (Global Cluster) and five (OECD) cluster solution in Chapter Seven.

Accounting for demand in the regime clusters

Some may argue that variation in immigration regime outcomes is a product not only of the independent variables that we identify, but also of broader demand factors that dictate migrant choices. To test for the possible role of such factors, we ran an independent algorithm that only considers cross-national variation using demand-based variables. These factors we select are among the most conventional drivers of migrant demand for admission. They include gross domestic product (GDP) per capita, unemployment rates, democracy scores, and the destination state's fragility index. These sources of these data are outlined below. Clustering solely on these desirability factors, we find demand taxonomies of little resemblance to our demographic data-driven solution. The substantial difference between clusters based on demographic data and demand factors undermine suggestions that demographic outcomes are simply expressions of pull factors (See Table A.3.).

Table A.3: Six-Cluster Solution for Demand-Related Factors Only

2 Cllusters	3 Clusters	4 Clusters	5 Clusters	6 Clusters	7 Clusters	8 Clusters	9 Clusters
Bahrain	Bahrain	Bahrain	Bahrain	Norway	Brazil	Italy	Austria
Brazil	Brazil	Brazil	Oman	Switzerland	China	New Zealand	Canada
China	China	China	Portugal	Austria	Mexico	Spain	Finland
Mexico	Mexico	Mexico	Saudi Arabia	Belgium	Russia	United Kingdom	Ireland
Oman	Oman	Oman	South Korea	Canada	Australia	Bahrain	Netherlands
Portugal	Portugal	Portugal	Spain	Finland	Denmark	Oman	Singapore
Russia	Russia	Russia	Australia	Germany	Sweden	Portugal	United States
Saudi Arabia	Saudi Arabia	Saudi Arabia	Austria	Ireland	Bahrain	Saudi Arabia	Norway
South Korea	South Korea	South Korea	Canada	Japan	Oman	South Korea	Brazil
Spain	Spain	Norway	Denmark	Kuwait	Portugal	Norway	China
Australia	Norway	Switzerland	Finland	Netherlands	Saudi Arabia	Switzerland	Mexico
Austria	Switzerland	France	Ireland	Singapore	South Korea	Brazil	Belgium
Belgium	Australia	Germany	Netherlands	United States	Belgium	China	France
Canada	Austria	Italy	Singapore	France	France	Mexico	Germany
Denmark	Belgium	Japan	Sweden	Italy	Germany	Russia	Japan
Finland	Canada	New Zealand	United States	New Zealand	Japan	Austria	Kuwait
France	Denmark	Spain	Belgium	Spain	Kuwait	Canada	Switzerland

Germany	Finland	United Kingdom	France	United Kingdom	Italy	Finland	Russia
Ireland	France	Australia	Germany	Australia	New Zealand	Ireland	Saudi Arabia
Italy	Germany	Austria	Italy	Denmark	Spain	Netherlands	Bahrain
Japan	Ireland	Belgium	Japan	Sweden	United Kingdom	Singapore	Oman
Kuwait	Italy	Canada	Kuwait	Bahrain	Austria	United States	Portugal
Netherlands	Japan	Denmark	New Zealand	Oman	Canada	Australia	South Korea
New Zealand	Kuwait	Finland	United Kingdom	Portugal	Finland	Denmark	Australia
Norway	Netherlands	Ireland	Norway	Saudi Arabia	Ireland	Sweden	Denmark
Singapore	New Zealand	Kuwait	Switzerland	South Korea	Netherlands	Belgium	Sweden
Sweden	Singapore	Netherlands	Brazil	Brazil	Singapore	France	Italy
Switzerland	Sweden	Singapore	China	China	United States	Germany	New Zealand
United Kingdom	United Kingdom	Sweden	Mexico	Mexico	Norway	Japan	Spain
United States	United States	United States	Russia	Russia	Switzerland	Kuwait	United Kingdom

Sources for demand variables considered in Chapter Seven

The following data were used to assess demand factors in Chapter 7.

To analyse GDP per capita as a possible pull factor for migrants, we used data from the United Nations (2016) in 2013 USD, from "GDP and its breakdown at current prices in US Dollars," United Nations Statistics, available online at: http://unstats.un.org/unsd/snaama/dnlList.asp

To assess the role of democratic status as a migrant pull factor, we employed Democracy Score of The Economist (2012) Intelligence Unit, "Democracy Index."

To analyze the role of political instability, we referred to Marshall and Cole (2013), "State Fragility Index and Matrix 2013," *Center for Systemic Peace*, available online at: http://www.systemicpeace.org/inscr/SFImatrix2013c.pdf.

To consider the role of unemployment rate as a pull factor for migrants, we consulted the World Bank (2013c), "Unemployment, total (% of total labor force)," available online at: http://data.worldbank.org/indicator/SL.UEM.TOTL.ZS.

Reasons for rejecting regression analysis to understand regime clustering (Chapter Eight)

As we argue in Chapter Eight, it is not possible to undertake a regression analysis for our small dataset of only thirty countries with complete data. Here, we provide a detailed defense of our decision. Because of our small dataset, Maximum Likelihood Estimates are highly unstable and often unable to converge. An ideal approach would be to use a multinomial logistic regression with cluster membership as the outcome variable that we regress on our independent variable of interest.

One alternative approach is to use a linear probability model, which is easier to estimate using ordinary least squares. Because the outcome is dichotomous, our error terms will necessarily be heteroskedastic and need to be adjusted. With this method, we dichotomize cluster membership as in cluster m_{i} (1) or not (0). Each cluster gets its own regression. We then correct standard errors and combine regressions into a single table. However, this approach raised concerns about omitted variable bias due to our small sample size. For this reason, we elected to rely upon bivariate analysis in our discussion in Chapter Eight.

Sources for correlational variables in Chapter Eight

In order to examine correlations between regime cluster placement and possible variables, we considered a series of correlations. The data sources for each of these are set out below.

Colony-colonizer dataset

We used data from the Correlates of War Project, which documents states' colonial status from 1816 to the present, to create a colony-colonizer continuum for our countries.
We used: Correlates of War 2 Project (2015), "Colonial/Dependency Contiguity Data, 1816-2002," Version 3.0, available online at: http://www.correlatesofwar.org/data-sets/colonial-dependency-contiguity.
In coding countries along a colony-colonizer continuum, we drew upon and extended previous work through the Correlates of War dataset (2015). Colony: Authors' index: 1—Major colonizer, 2—minor colonizer, 3—neither colonizer nor colonized, 4—minor colonization, 5—major colonization

Our countries are categorized as a colony if part of their current integral territory (not including overseas regions or offshore islands) is listed as being a colony (not a protectorate or occupied territory or some other status) in the Correlates of War dataset. Likewise, countries that are listed as possessing colonies that were not minor islands are coded as colonizers. We excluded colonies that were contiguous with the current mainland territory from our analysis. A few countries both had their territory colonized and colonized other countries. We made the following subjective decisions in those cases based on which role colonization played in that nation's history:

- Australia – coded as colony
- New Zealand – coded as colony
- United States – coded as colony
- Denmark – coded as minor colonizer

The following countries are coded as neither on the basis that their colonies were too insignificant to make them behave like a colonizer:

- Russia (Alaska)
- Norway (Faeroe Islands)

Start year

Under the Correlates of War dataset, the start year is 1816. However, setting this as the beginning of colonial empire is misleading as it ignores the period of colonization for many major colonies, including Australia (1788) and Mexico (1519). For this reason, we supplemented the Correlates of War starting years with web sources to recode to the true start date, where applicable, through internet sources. First priority was given to official government websites. Second priority was following citation links in Wikipedia. Third priority was given to third-party websites found through Google searches or Google books text searches. Sources for each country are available from the authors upon request.

End year

The year (before 1993) when the last part of that nation's territory gained independence from a colonizer (for colonies) or when it ceded its last colony (for colonizers)

Total time

End Year – Start Year

Categorization of major/minor

We found the median years that colonized countries were colonized and used that as a point to split the data. Every country that was colonized for more than the median was coded as a major colony and less than the median, a minor colony.

For colonizers, we adopted a different method that relied upon estimates of the maximum territory the colonizer held outside of its homeland at its peak size, even if this was before 1816. On the basis of this characterization, we divided the colonizers into two categories: 1) major colonizers and 2) minor colonizers. Major colonizers had a colonial land mass over three million square miles at the height of their colonial period. We employed this measure rather than a time dimension as it is possible that a colonizer could have been in control of a very small piece of land for a long period, which provides a misleading metric of the scope of colonial rule. Minor colonizers were those with less than three million square miles.

We then created a scale that differentiated between the following: 1) major colonizer; 2) minor colonizer; 3) neither; 4) short-term colony; 5) long-term colony. Long-term versus short term colony was determined by calculating median years of colonization. Those countries to the right of the median were coded as long-term and those to the left as short-term. This coded scale was used as the independent variable in the analysis in Chapter Eight.

Economic and demographic indices

To assess economic freedom, we used the Fraser Institute's (2013) "Economic Freedom of the World Index," available online at: http://efwdata.com/grid/WxRvYnU#/Grid .

In order to analyze the relationship between resource wealth and regime placement, we used the measurement of total natural resource rents (% of GDP): World Bank (2013b), "Total natural resource rents (% of GDP)," available online at: http://data.worldbank.org/indicator/NY.GDP.TOTL.RT.ZS. We used the ten-year average from 2004 through to 2013.

To assess the role of welfare state provisions for OECD countries, we drew upon the Combined Generosity Welfare Index in the Comparative Welfare Entitlements Dataset 1970-2011 (Scruggs et al. 2014). We used the ten-year average from 2002 through to 2011.

To assess the role of partisanship of the governing party, we drew upon data on the Left-Right party balance (gov_party) measure from the CPDS III (Armingeon et al. 2013), average of 1990-2012. We used the ten-year average from 2003 through to 2012. This variable relies upon the so-called Schmidt Index to assess Cabinet partisan composition on a scale of 0 to 5 with 0 on the right and 5 on the social democratic left.

To analyze population aging, we took the population over 65 and above as a percentage of total population from World Bank (2013a), "Population 65 and above (% of total)," available online at: http://data.worldbank.org/indicator/SP.POP.65UP.TO.ZS.