Exploratory Data Analysis of US Migration Policy

Amber Lee

August 2019

Introduction

This project examines the Determinants of Migration Policy (DEMIG), a data set that records over 6,500 migration policies enacted in 45 countries, including the United States. Researchers assessed each policy on 13 different variables like target group and change in restrictiveness. Policies included in DEMIG are from 1721 to 2013 at the national or international level.

This analysis was conducted from July to August 2019 as part of a policy internship at the organization Only Through Us. This was a self-led project informed by R for Data Science and An Introduction to Statistical Learning.

My guiding questions for this project are as follows:

1. How has the restrictiveness of migration policy changed over time?
2. How does migration policy treat groups of migrants differently?

Setting up

For this project, I use the haven, tidyverse, mclust, and knitr packages.

```
knitr::opts_chunk$set(echo = TRUE)
library(haven) # For reading stata file
library(tidyverse) # Primarily for dplyr, ggplot, and pipe  %>% function
library(mclust) # Gaussian finite mixture models
library(kableExtra) #Table formatting
library(knitr)

cbPalette <- c("#0072B2", "#D55E00", "#CC79A7", "#E69F00", "#56B4E9") #For ggplot
```

About the dataset

I am primarily interested by contemporary trends, so I limited the policies in this exploratory data analysis to be from 1950 onward. The variables of interest in this project are explained as follows: * change_restrict: Change in restrictiveness. This is a categorical measure of how a policy increases or decreases the rights of a migrant group, relative to the current legal framework. There are three levels: less restrictive (coded -1), no change in restrictiveness (coded 0), and more restrictive (coded +1).

- pol_area: Policy area. This categorical variable has 4 levels: 1) Border patrol and land control: policies that regulate external and internal border controls and aim to secure national territory, 2) Legal entry and stay: policies that regulate entry visa and stay permits, without distinguishing temporary and long-term permits, 3) Integration: policies that regulate post-entry rights or affect integration, including state's relations with citizens living abroad, 4) Exit: policies that regulate both forced and voluntary exit from a territory
- target_group: Target group. This is a categorical variable that specifies the migrant group targeted by a policy. Examples of levels in target_group include All migrants, Refugees and Asylum Seekers, High-skilled workers, and Family Members.

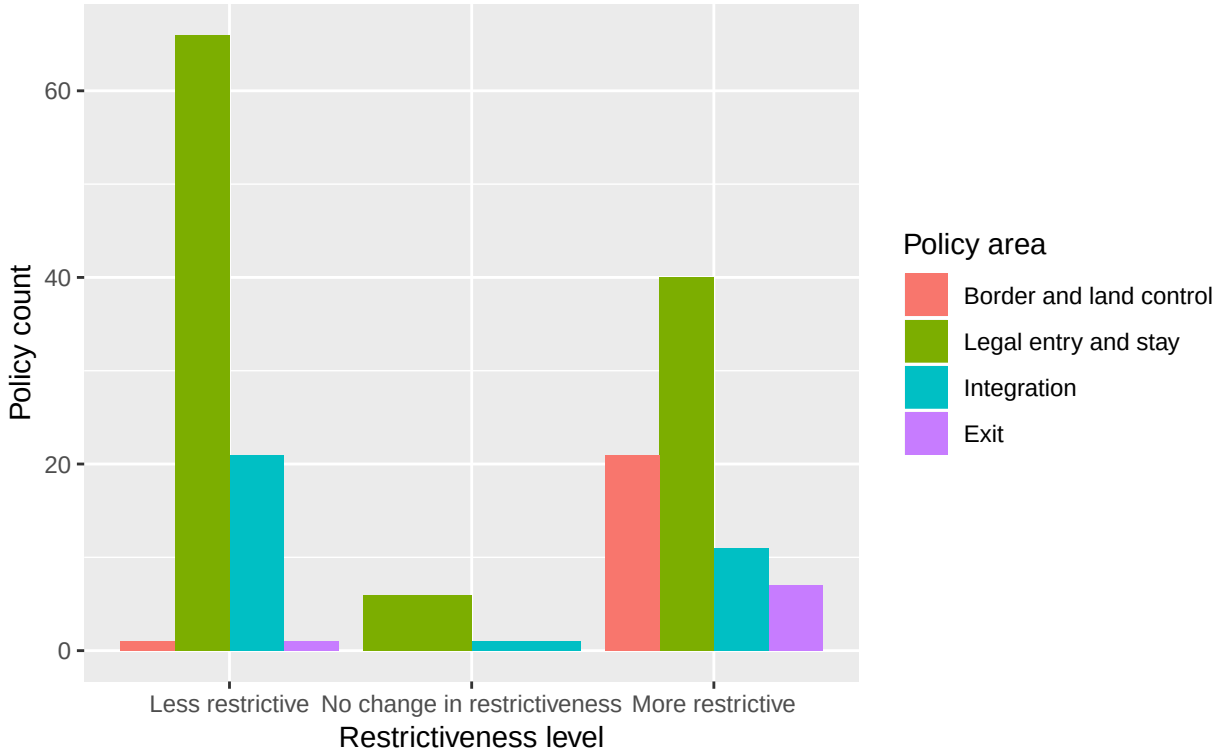
Table 1: Number of policies for policy areas by change in restrictiveness

	Less restrictive	No change in restrictiveness	More restrictive
Border and land control	1	0	21
Legal entry and stay	66	6	40
Integration	21	1	11
Exit	1	0	7

- change_level: Level of change. This is a categorical variable that assesses whether a policy is a major change, mid-level change, minor change or fine-tuning change.

Initial descriptive results

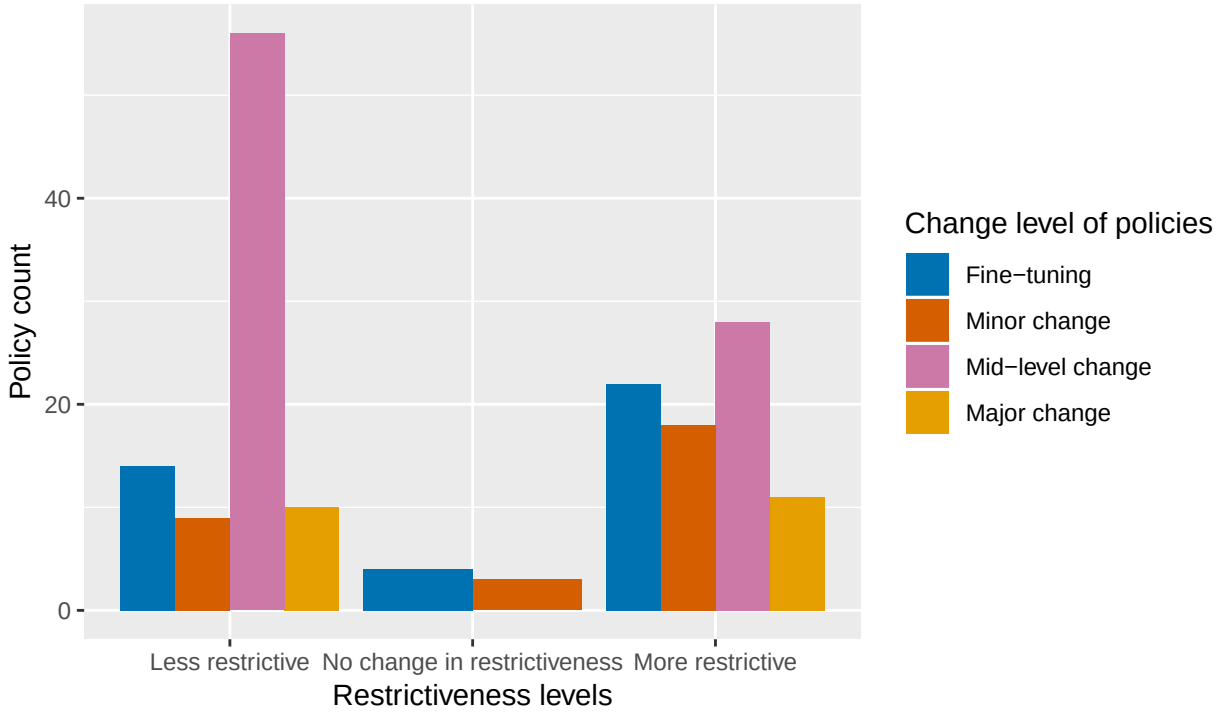
Figure 1:
Restrictiveness corresponds to different policy areas



Border and land control policies are almost exclusively more restrictive. 95% (or 21 out of 22) are more restrictive. Furthermore, legal entry and stay are primarily less restrictive. 59% (or 65 out of 112) are less restrictive.

Figure 2:
Restrictiveness levels also correspond to different change levels

Twice as many mid−level less restrictive changes than mid−level more restrictive changes



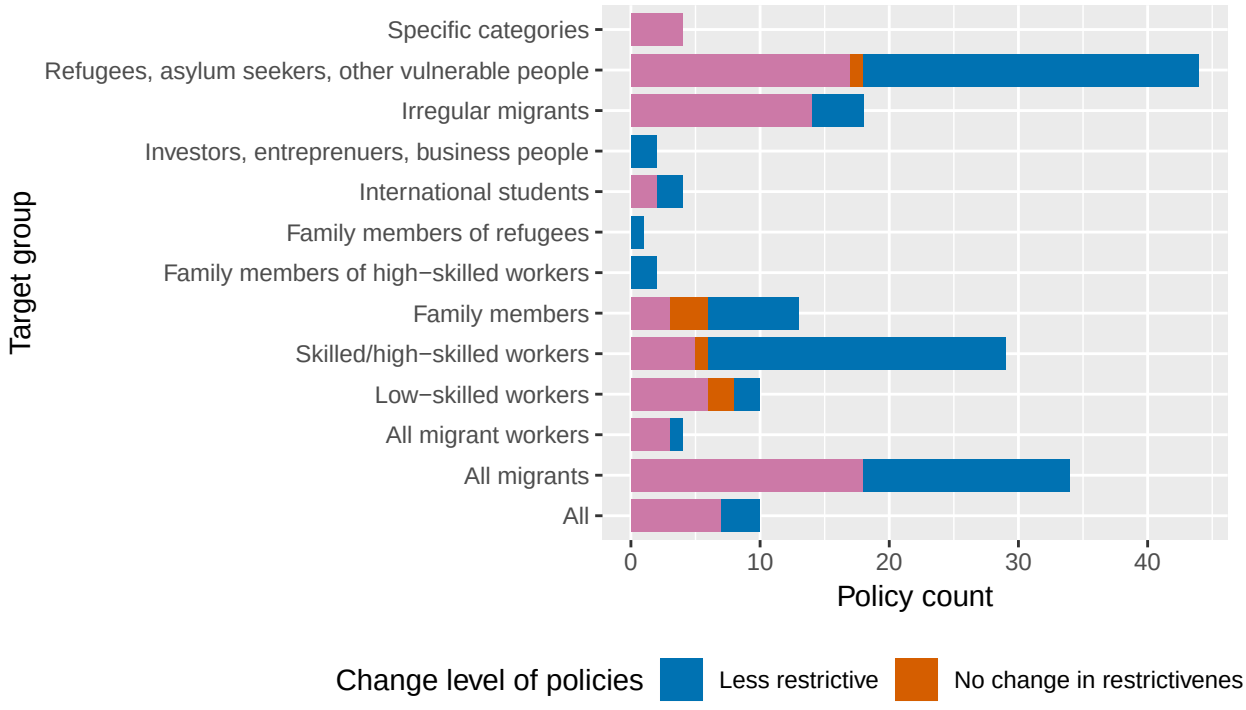
Both change in restrictiveness and change level determine the overall magnitude of restrictiveness. Thus, this chart conveys the difficulty in assessing how the direction of overall restrictiveness in migration policy. This distribution suggests that since 1950, US migration policies have been primarily less restrictive, as mid-level changes, less restrictive policies dominate.

However, another complication in determining the magnitude of overall restrictiveness is the size of the target group that is affected. For example, a mid-level less restrictive change for family members of refugees may have less of an impact than a mid-level more restrictive change for all migrant workers.

Table 2: Number of policies for target group by change in restrictiveness

	Less restrictive	No change in restrictiveness	More restrictive
All	3	0	7
All migrants	16	0	18
All migrant workers	1	0	3
Low-skilled workers	2	2	6
Skilled/high-skilled workers	23	1	5
Family members	7	3	3
Family members of high-skilled workers, investors or students	2	0	0
Family members of irregular migrants or refugees, asylum seekers and other vulnerable people	1	0	0
International students	2	0	2
Investors, entrepreneurs and business people	2	0	0
Irregular migrants	4	0	14
Refugees, asylum seekers and other vulnerable people	26	1	17
Specific categories	0	0	4

Figure 3:
Different target groups experience different restrictiveness levels



The distribution of restrictiveness varies per migrant group. The distributions for irregular migrants and low-skilled workers are predominantly more restrictive, while high-skilled workers have less restrictive policies. The distribution seems split more even for refugees and all migrants than for the three aforementioned groups. 60% (44 out of 175) of refugee policies are less restrictive.

Table 3: Key for Change in Restrictiveness

numeric_change_restrictive	number
Less restrictive	-1
No change	0
More restrictive	1

Table 4: Key for Change Level

numeric_change_level	number
Fine-tune	1
Minor	2
Mid-level	3
Major	4

Another result is that refugees are the most regulated group, with migration policies targeting refugees outnumbering those targeting all migrants. 25% of all migration policies target refugees.

Figure 4:
Restrictiveness of policy areas through time

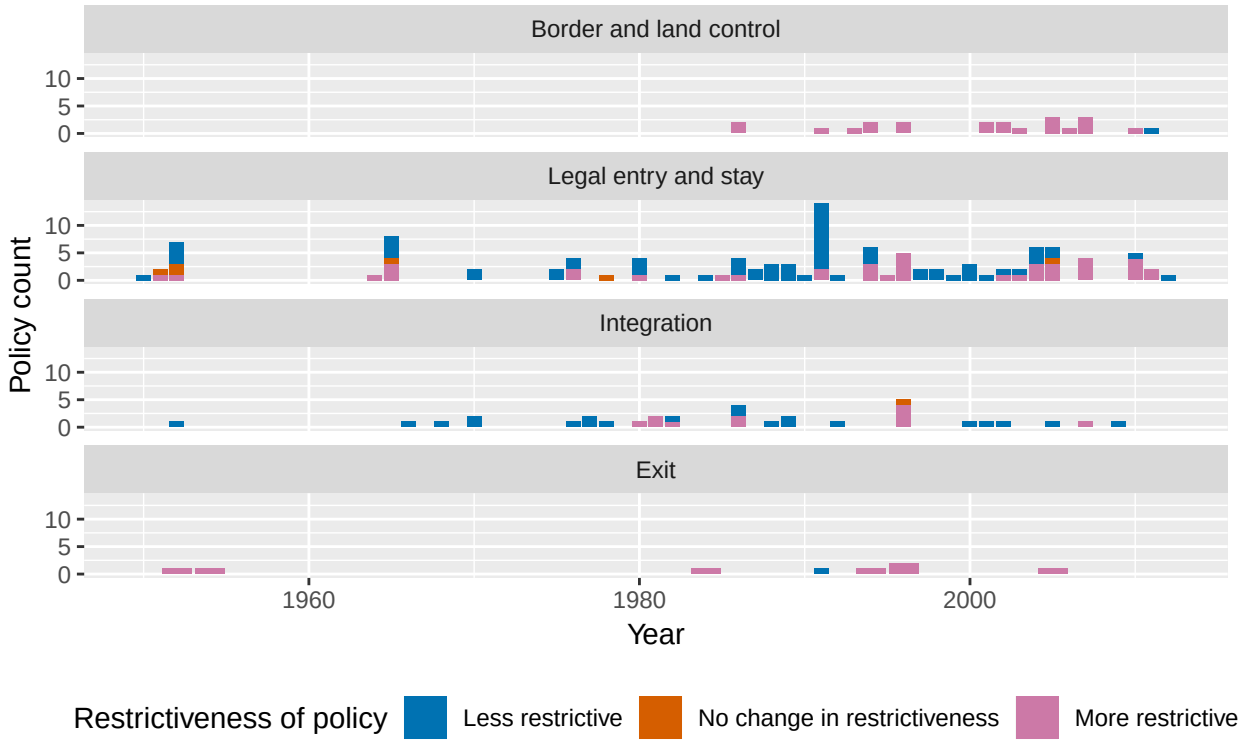


Figure 4 gives a visual representation of contemporary migration policy. Border and land control policies have been enacted steadily in throughout certain 5-year periods, while legal entry and stay policies look to be spurred once per decade and passed steadily otherwise.

Creating a new variable: mag_score

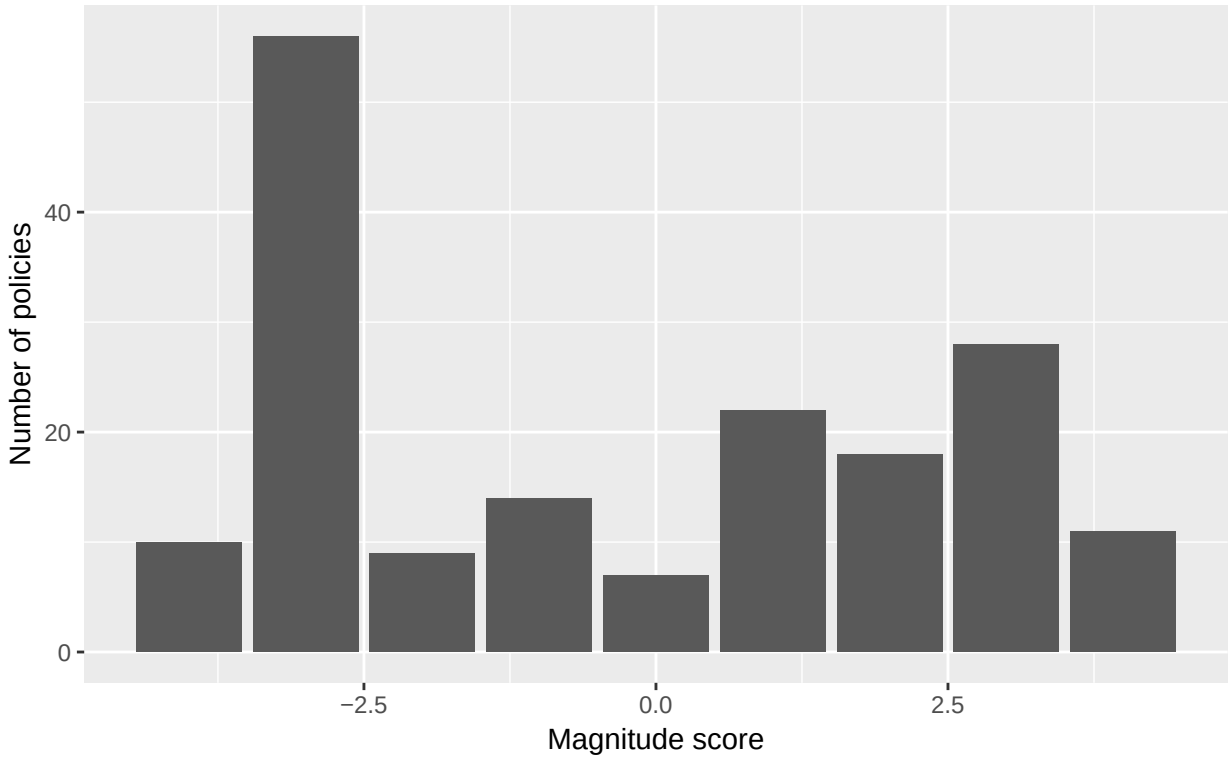
I created a new variable magnitude score that combines the variables change in restrictiveness and change level. First, I recoded each level of the two variables to be numeric as specified by the following two tables:

Table 5: Excerpt of the DEMIG data with magnitude score

year	mag_score	change_restrict	change_level
1950	-1	Less restrictive	Fine-tuning
1951	0	No change in restrictiveness	Fine-tuning
1951	3	More restrictive	Mid-level change
1952	1	More restrictive	Fine-tuning
1952	-3	Less restrictive	Mid-level change
1952	-3	Less restrictive	Mid-level change

Then, to create the magnitude score, I multiplied together the values of numeric change level and numeric change in restrictiveness. As shown in the excerpt of the cleaned DEMIG data set below, a policy with -1 magnitude score refers to a less restrictive and fine-tuning change.

Figure 5:
Distribution of magnitude scores

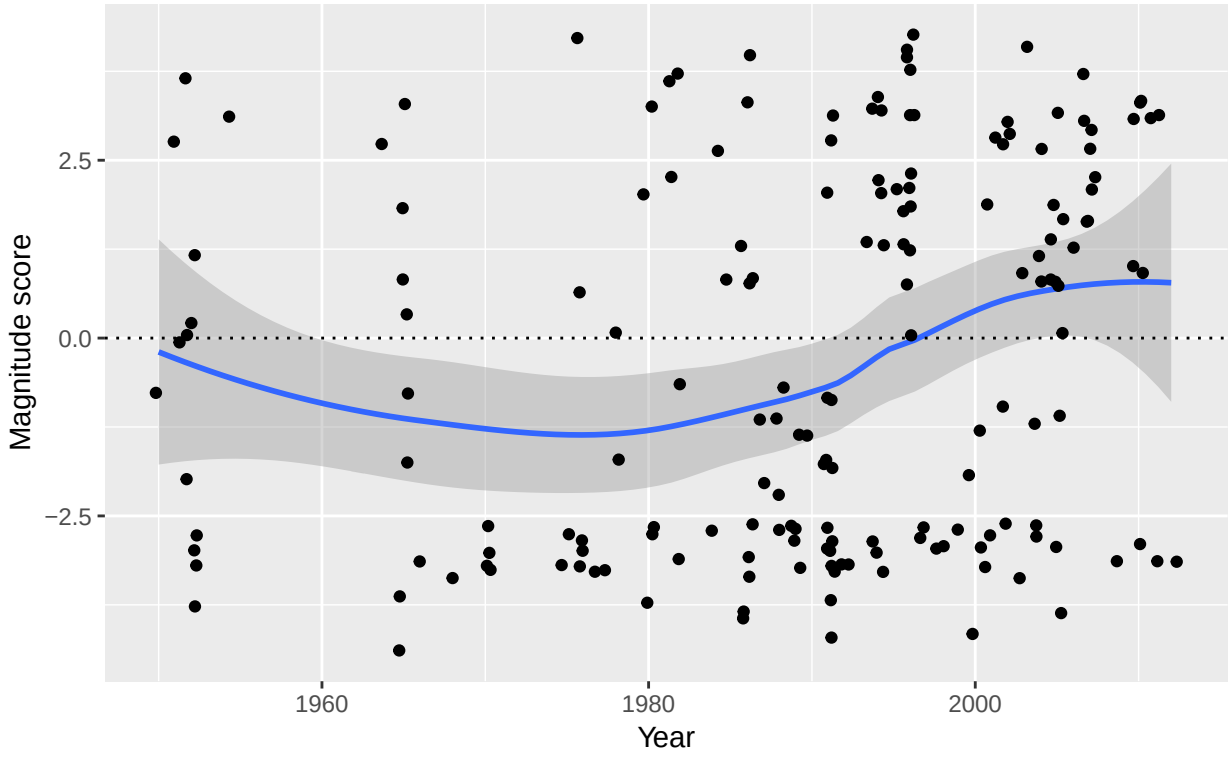


The most frequent migration policies are less restrictive mid-level changes. This is another representation of the Figure 2.

Analysis of magnitude scores

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figure 6:
Magnitude scores through time, all policies



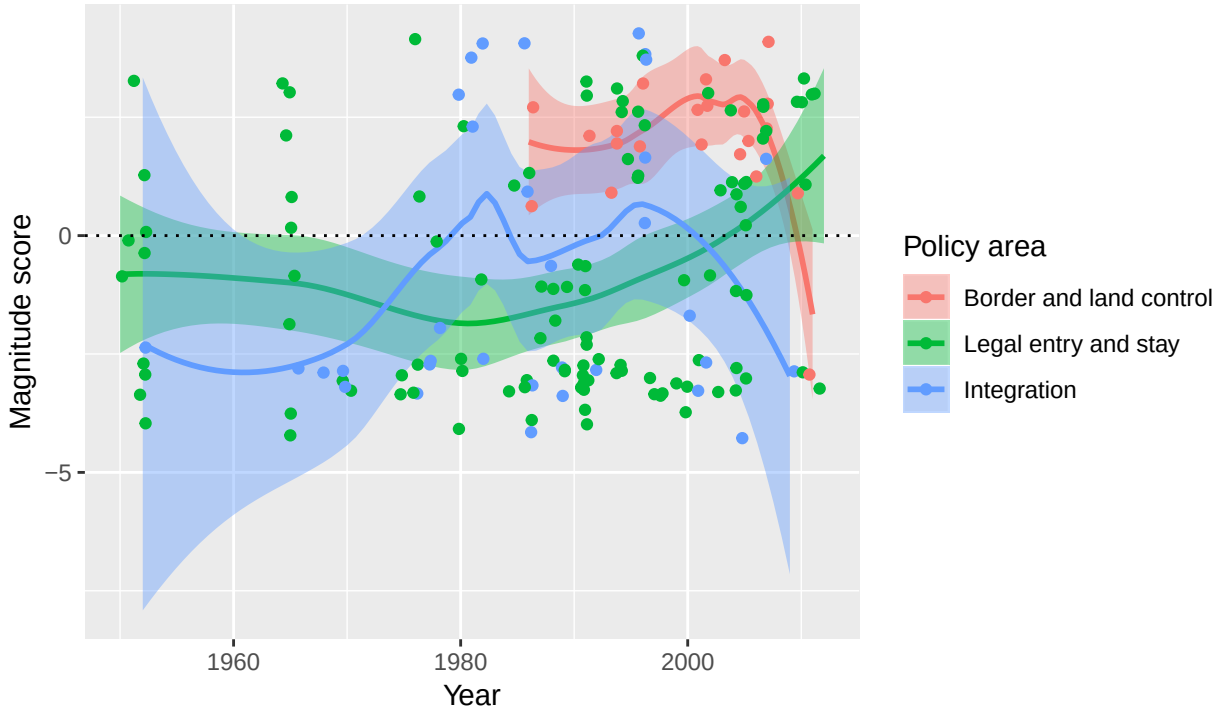
By plotting the magnitude scores per year, a general trend emerges of less restrictive policies from 1960 to 1990. After 1990, the average magnitude score becomes more positive, but whether or not policies are as a whole more restrictive is unclear.

Figure 1 conveyed that certain policy areas can be predominantly more or less restrictive. By dis-aggregating the trends per policy area, there is a more nuanced understanding of magnitude score trends:

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figure 7:
Magnitude scores through time, by policy area

Excluding exit−related policies because there are too few of them




```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figure 8:
Magnitude scores through time, by policy area

Excluding integration−related policies because of high variance or few policies

Legal entry and stay policies account for the majority of negative magnitude scores, while border and land control policies account for the majority of positive magnitude scores. While migration policy as a whole looks to be less restrictive between 1960 and 1990, legal entry and stay policies are the main drivers behind that trend. Integration policies look to have a high-variability, cyclical trend.

To extend on Figure 3, I plotted the distributions of magnitude score per target group. The target groups with interesting results were all migrants, high-skilled workers, and refugees.

Figure 9:
Distribution of magnitude scores for different target groups

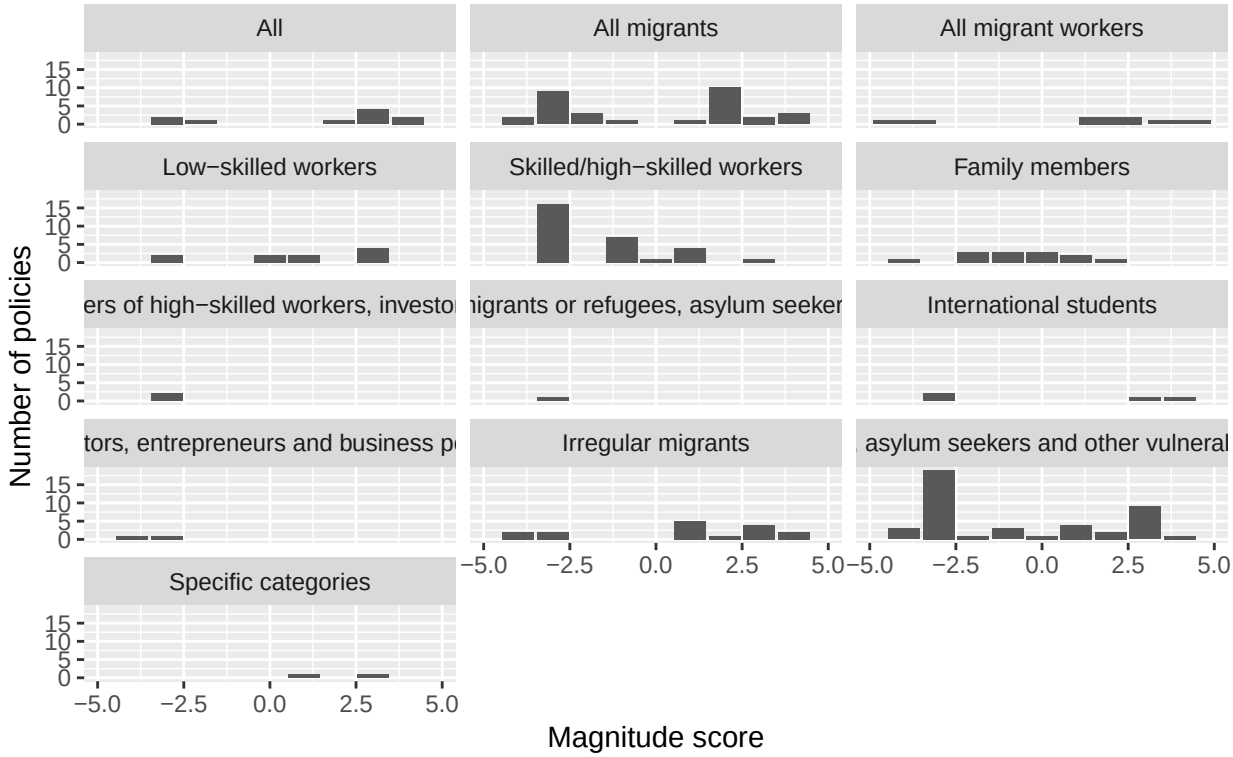
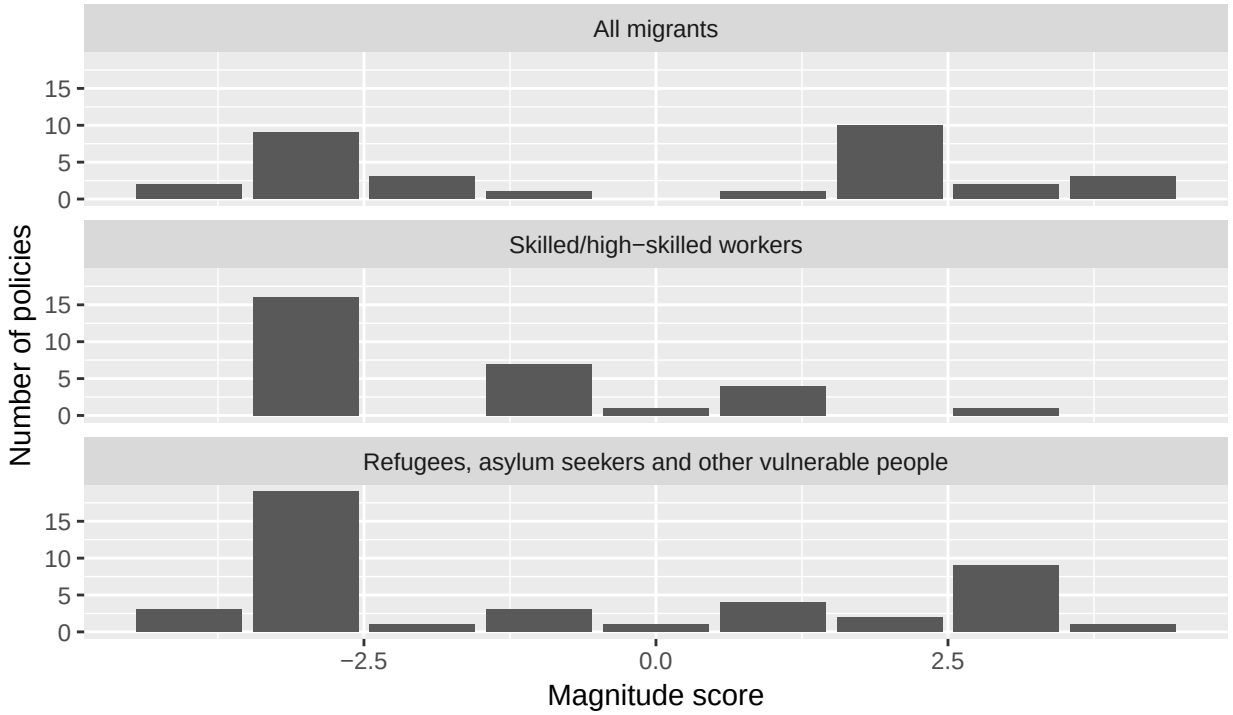


Figure 10:
Distribution of magnitude scores for selected target groups

Policies for All migrants and Refugees looks to be bi−modal



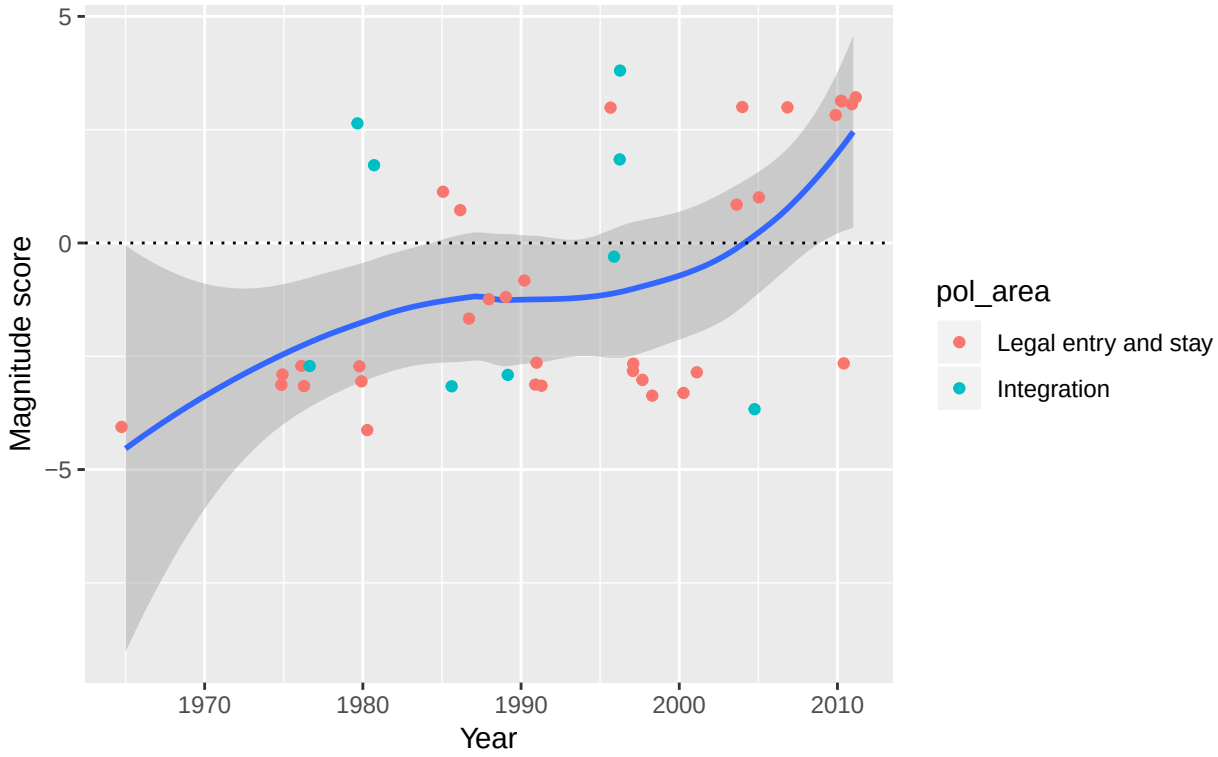
The high-skilled workers distribution looks quite intuitive: the less restrictive policies towards high-skilled workers are greater in number. Refugee- and all migrants-related policies look to be sourced from two different groups as the distributions look bi-modal.

Magnitude scores regarding refugees

One finding from Figure 3 is that refugee-related migration policy accounts for 25% of all migration policies, the largest category. Thus, I conducted the same magnitude score-time plots but for refugee-targeting policy in particular.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

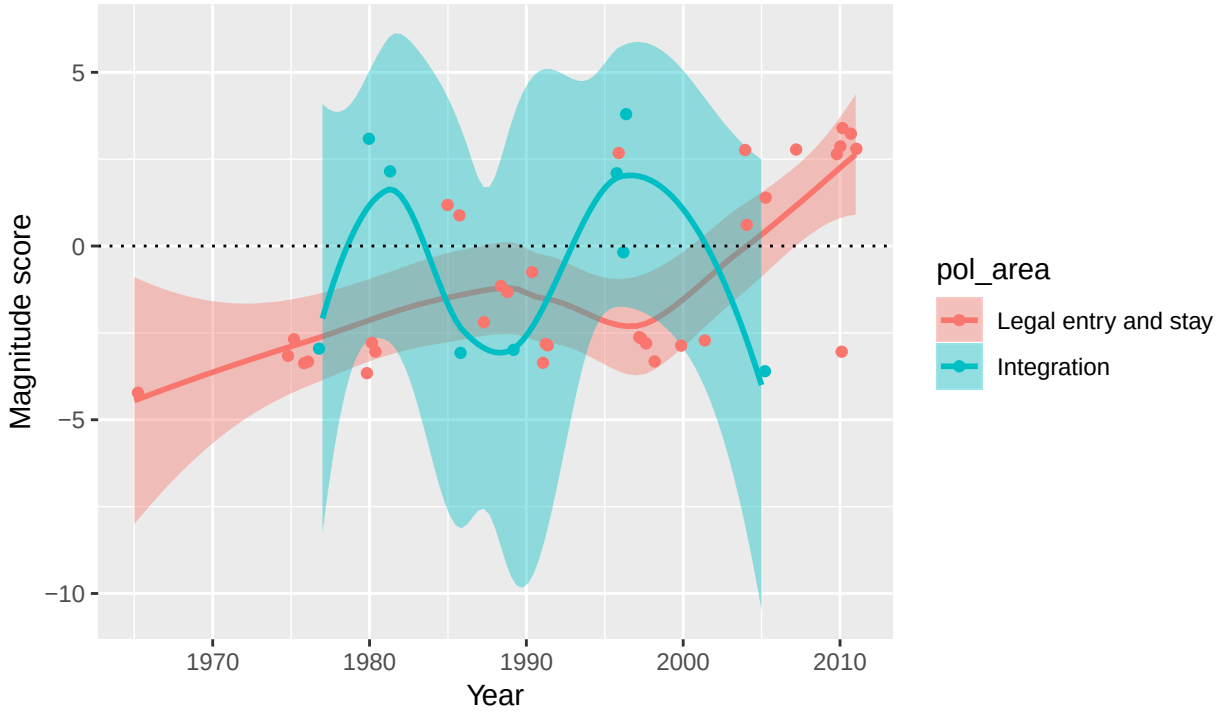
Figure 11:
Magnitude scores of refugee migration policy has become more restrictive r




```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Figure 12:
Magnitude scores of refugee migration policy by policy area

Legal entry and stay look to be driving the trend of more restrictive refugee policy

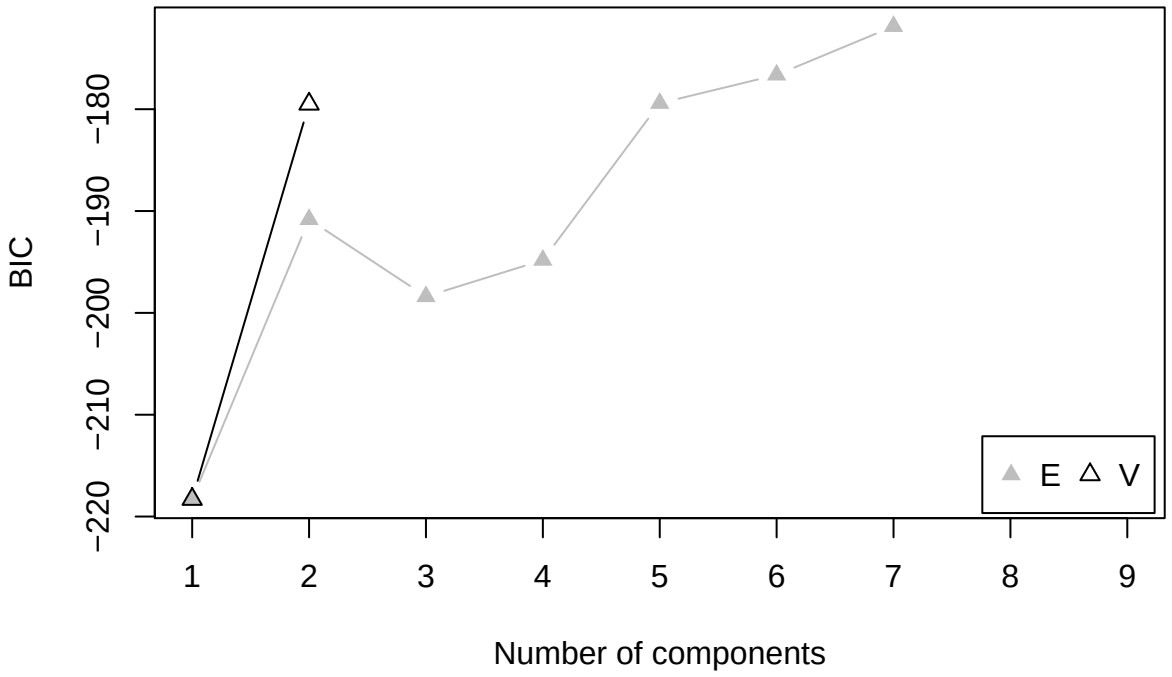


Th trends shown here look similar to those in Figure 7 and Figure 8. Refugees face more restrictive measures than they did previously. This is driven by legal entry and stay policies, which have been historically less restrictive with moderate magnitude, although since 1995, policies have become increasingly more restrictive. There are too few integration policies for a meaningful interpretation.

Univariate mixture modeling of magnitude score for refugee policy

```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust E (univariate, equal variance) model with 7 components:
##
##  log-likelihood  n df       BIC       ICL
##       -59.44322 44 14 -171.8651 -171.8959
##
## Clustering table:
##  1  2  3  4  5  6  7
##  3 19  4  1  5  2 10
```

Figure 13: BIC for equal (E) and unequal (V) variance model



The first model (Mclust model E) uses a univariate equal variance model that has seven components, dividing magnitude scores along their scores. As one would expect, this has the highest Bayesian Indicator Criteria (BIC), as indicated in the plot. However, a univariate model with unequal variances is maximized to a reasonably high BIC with two components, which matches the initial observation in Figure 10 of bimodality. A model with unequal variance is more realistic and has more meaningful implications.

```
## ----------------------------------------------------
## Gaussian finite mixture model fitted by EM algorithm
## ----------------------------------------------------
##
## Mclust V (univariate, unequal variance) model with 2 components:
##
##  log-likelihood  n df       BIC       ICL
##       -80.28589 44  5 -179.4927 -180.4484
##
## Clustering table:
##  1  2
## 22 22
##
## Mixing probabilities:
##         1         2
## 0.5032951 0.4967049
##
## Means:
##         1         2
## -3.120591  1.606287
##
## Variances:
##         1         2
## 0.1318548 2.7929895
```

The first cluster has a mean score of -3.12 and variance .13; the second cluster has a mean of 1.6 and a variance of 2.79. This suggests that refugee policy has two main sources: ardent supporters (cluster one) and non-supporters consisting of neutral bystanders to ardent opponents (cluster two). Sympathetic refugee policy tends to result in significant change at once with a magnitude score of -3. Non-sympathetic refugee policy has greater variance, including neutral changes and higher magnitude, more restrictive changes.

Conclusion

Quantifying the impacts and trends of migration policy requires attention to the various dimensions of each policy; it's not so simple to chart magnitude change. There are three main takeaways from the initial descriptive results. First, different policy areas correspond to different levels of restrictiveness. Border control policies are overwhelmingly more restrictive (95% are more restrictive), while about two-thirds of legal entry policies are less restrictive. Next, less restrictive policies are dominated by mid-level changes, while more restrictive policies are more spread out with regards to change level. Finally, target groups matter in determining the aggregate impact of migration policies – refugees are the most regulated target group, representing one quarter of total migration policies from 1950 to 2013.

Creating the variable magnitude score that combines change level and change in restrictiveness provides a more nuanced understanding of migration policy through time. Migration policy was less restrictive from 1960 to 1990. Beginning around 1980, magnitude score began to increase, but it is unclear if the average migration policy has been more restrictive (having a positive magnitude score) since then. Disaggregating by policy area provides more granular insight: legal entry and stay policies have historically been less restrictive and the main driver of the 1960 to 1990 less restrictive migration trend noted earlier. Border and land control policies, in contrast, have been historically more restrictive. Similar trends can be found when analyzing magnitude scores through time for refugee policy for legal entry and stay.

Examining magnitude score distribution per target group yields a bimodal distribution for target groups All Migrants and Refugees. Applying the univariate Gaussian finite mixture model results in two clusters of policies: one cluster of policies representing mid-level to major changes that increase legal rights of migrants, the average of which is mid-level less restrictive change (mean magnitude score of -3.6). The second cluster represents neutral to major changes that decrease the rights of migrants, the average of which is a more restrictive, minor-level change (mean magnitude score of 1.6).

Works Cited

DEMIG (2015) DEMIG POLICY, version 1.3, Online Edition. Oxford: International Migration Institute, University of Oxford. www.migrationdeterminants.eu

Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal 8/1, pp. 205-233