

Week 7 OCT 6 - OCT 12 Lab 5 International Migration

Lysa Vanible

October 08, 2021

#Introduction

In this lab, we will examine trends in international migration to the USA. Net international migration is the difference between the people coming to the USA and the people leaving the USA. We calculate net international migration to the US for all the years from 1933 to 2014 using what is known as the population balancing equation. Once again are going to be using data from the mortality.org so you will need credentials for the mortality.org account you created in lab 2.

We encountered the balancing equation in week 3 when we discussed population growth. In its most basic form, it says that $\text{End Population(Pend)} = \text{Start Population(Pstart)} + \text{Births(B)} - \text{Deaths(D)} + \text{Net Migration(M)}$

In other words, to find the population size at the end of a period (Pend) you should take the population size at the start of the period(Pstart), add the number of babies born during the period(Births), subtract the number of deaths(Deaths) during the period and add the difference between people entering the country and people leaving the country during the period(Net Migration). The result of that calculation will be the population size at the end of the period(Pend).

In our case, we are going to use it to compute the net migration to the USA. This requires that we reshuffle the items in the equation a bit so it looks like this

$$M = \text{Pend} - \text{Pstart} - B + D.$$

The equation now means that to find net international migration to the US during any period, we take End Population, subtract Start Population and Births and add Deaths. Whatever comes out will be the net international migration to the USA during the period. We are going to use this equation to find out net international migration to the USA for the years between 1933 and 2020 in this lab.

- 1). First replace “Write your name here” in the YAML header above with your name before you start. MAKE SURE NOT TO MODIFY ANY OTHER PART OF THE HEADER.
- 2). Again the same general rules we used in the previous labs hold for this lab. Do not delete or modify anything unless it was written by you. All R code must be entered in the grey boxes/code chunks.
- 3). To run R code as you work your way through the assignment, highlight the code and click on the “Run” command in the R studio menu and then on “Run Selected Lines” or click on the green button at the top right of a particular code chunk to run all the code within it.
- 4). When you get to the end of the assignment you will knit the document to produce a pdf document using the Knit command in the Rstudio menu.
- 5). Continue reading from Part A.

```
knitr::opts_chunk$set(error = TRUE)
```

PART A EXTRACT USA POPULATION DATA FROM 1933 to 2014

We'll start by loading our libraries.

```
library(HMDHFDplus)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.3      v purrr 0.3.4
## v tibble 3.1.2       v dplyr 1.0.6
## v tidyr 1.1.3        v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()      masks stats::lag()
```

As with previous labs, we'll import the population of the USA from the Human Mortality Database for years 1933 to 2014 into RStudio. Enter in your username and password information for mortality.org as before and run the code below.

```
population <- readHMDweb(CNTRY = 'USA', item = 'Population',
  username = 'lysa.vanible@lc.cuny.edu', password = 'Oceansky2004')
total_pop_by_year <- aggregate(x=population[4:9], by=list(Year=population$Year), FUN=sum)
```

In the dataset “population” containing the data extracted from the mortality.org, there are two populations listed for each age and each year. Total1 is the population on January 1st of the year, and Total2 is the population on December 31st of the same year. We'll want both of these totals for the population balancing equation. These will serve as our start and end populations for each year. The first line of the code below extracts three columns from the dataset “population” containing year, Total1 and Total 2 and places them in a new dataset called “total_pop_by_year”. The third of fourth lines of code below rename the January 1st population, Pop_T1, and Dec 31st population, Pop_T2, and places them in a third dataset that we call migration_by_year.

```
total_pop_by_year <- dplyr::select(total_pop_by_year, Year, Total1, Total2)
migration_by_year <- total_pop_by_year
migration_by_year <- dplyr::rename(migration_by_year, Pop_T1=Total1)
migration_by_year <- dplyr::rename(migration_by_year, Pop_T2=Total2)
```

PART B EXTRACT USA BIRTHS DATA FROM 1933 to 2014

You may recognize this chunk from the fertility lab. Here we are importing and formatting the data on births for each year 1933 to 2014. Remember that we need data on births for the balancing equation. Enter in your username and password information for mortality.org and run the code below.

```
births_by_year <- readHMDweb(CNTRY = 'USA', item = 'Births',
  username = 'lysa.vanible@lc.cuny.edu', password = 'Oceansky2004')
births_by_year <- dplyr::select(births_by_year, Year, Total)
births_by_year <- dplyr::rename(births_by_year, Births=Total)
```

PART C EXTRACT USA DEATHS DATA FROM 1933 to 2014

We also need data on deaths so we import and format the data on deaths for each year 1933 to 2014. Enter in your username and password information for mortality.org as before and run the code below.

```
deathcount <- readHMDweb(CNTRY='USA', item='Deaths_1x1',
  username='lysa.vanible@lc.cuny.edu', password='Oceansky2004')
deaths_by_year <- aggregate(x=deathcount[3:5], by=list(Year=deathcount$Year), FUN=sum)
deaths_by_year <- dplyr::select(deaths_by_year, Year, Total)
deaths_by_year <- dplyr::rename(deaths_by_year, Deaths=Total)
```

PART D MERGE USA POPULATION, BIRTHS AND DEATHS DATA FROM 1933 to 2014

Right now we have three separate datasets. They are “migration_by_year”, which has information on start and end populations, “births_by_year”, which has information on births for each year between 1933 and 2014, and “deaths_by_year” which has information on deaths for the same period. Let’s add the variables on births and deaths by year into the “migration_by_year” dataset so they are all in the same place. The codes below create two new variables in the migration_by_year dataset and assigns the births and deaths to them. Run the code.

```
migration_by_year$Births <- births_by_year$Births
migration_by_year$Deaths <- deaths_by_year$Deaths
```

PART E MERGE USA POPULATION, BIRTHS AND DEATHS DATA FROM 1933 to 2014

Now it’s time to calculate net migration. Remember that we already said that to find net migration using the balancing equation, we have to subtract the starting population from the ending population and then subtract births and add deaths. The formula looks like this:

$$M = P_{end} - P_{start} - B + D.$$

We are going to do this in the code below.

On the left hand side of the code below, we tell R that we want to calculate a new variable called Net_Migration, and we would like it to be in the migration_by_year dataset so we refer to it as migration_by_year\$Net_Migration. On the right hand side, we need to tell R to take the ending population (migration_by_year\$Pop_T2), subtract the starting population (migration_by_year\$Pop_T1), subtract births (migration_by_year\$Births), and add deaths (migration_by_year\$Deaths). R will know to do this for each year between 1933 and 2014.

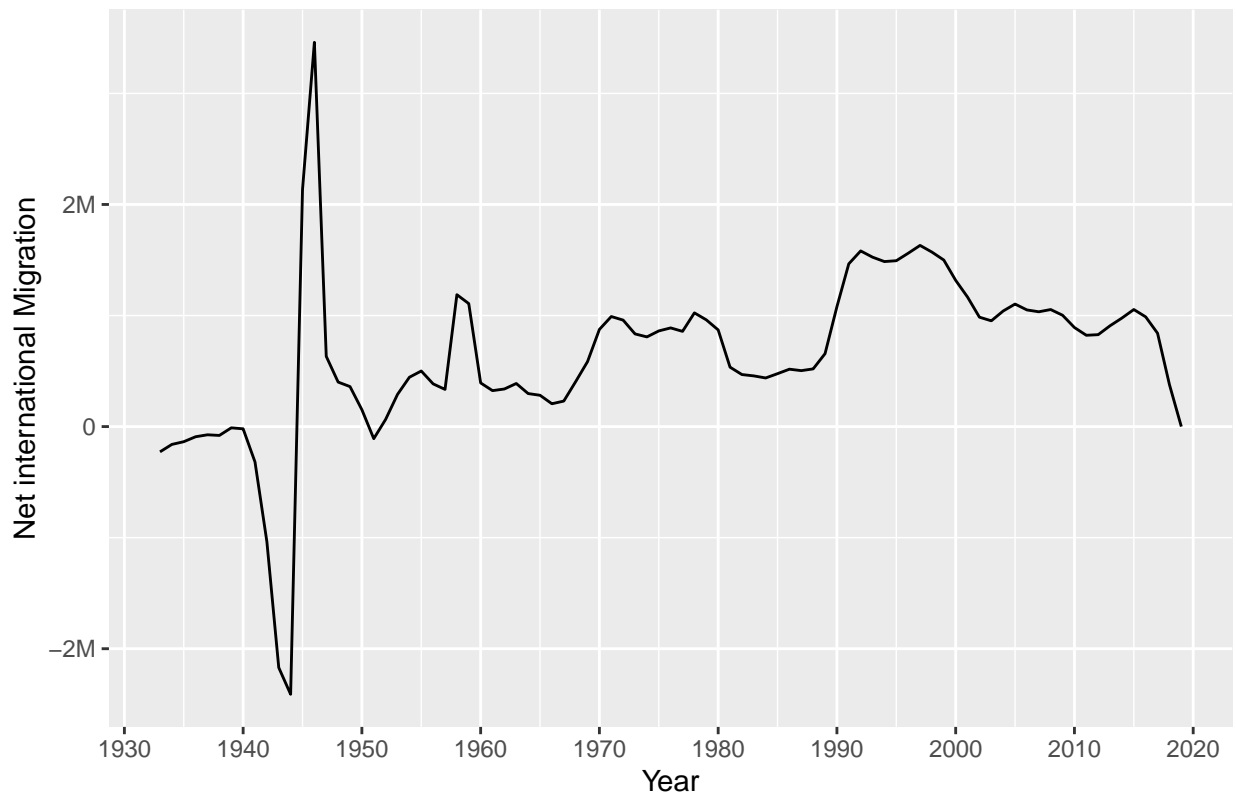
```
migration_by_year$Net_Migration <- migration_by_year$Pop_T2 - migration_by_year$Pop_T1 - migration_by_year$Births + migration_by_year$Deaths
```

PART F GRAPH US NET MIGRATION FROM 1933 to 2014

If you were successful in PART E, then you have net international migration counts from 1933 to 2014. Now graph the net international migration variable by year so they are easy to process. Go ahead and create the graph by running the code below.

```
ggplot(migration_by_year) +
  geom_line(mapping=aes(x=Year, y=Net_Migration))+
  scale_y_continuous(breaks=c(-2000000, 0, 2000000, 4000000), labels=c("-2M", "0", "2M", "4M"))+
  scale_x_continuous(breaks = scales::pretty_breaks(n = 10)) +
  ggtitle("US Net International Migration 1933-2014 (in Millions)") +
  ylab(label="Net international Migration")+
  theme(plot.title = element_text(hjust = 0.5))
```

US Net International Migration 1933–2014 (in Millions)



PART G CONCLUSION

Great work!!!! You are almost at the end of your fifth lab. You will now prepare to knit the lab. Click on the “Knit” command at the top of this window to turn this document with your code and all the results into a pdf document. When the knit is complete, Rstudio will open the pdf document. Inspect it very carefully to make sure that you have not missed any part of the assignment. If you have missed any tasks or see errors, please go back and fix them and knit the document again. Once the pdf document is knitted and you are sure you have completed all the tasks in lab 5 prepare to submit it to the week 7 folder of SOC339’s blackboard site. To do that there are still a few more things for you to do.

- 1). First save your rmarkdown file (the file you have been working on in Rstudio) by clicking on the FILE command in the Rstudio menu and then on SAVE. You will see the rmarkdown document in the FILES folder of the window to the bottom right of your screen. It will be named Lab 5 International Migration.Rmd.
- 2). To save the pdf file, locate it in the FILES folder of the window to the bottom right of your screen. It will be named Lab 5 International Migration.pdf. Click the little box to the left of it. Then click on the MORE tab near the top of the FILES window. Then on EXPORT and then when prompted by RSTUDIO click on DOWNLOAD. RSTUDIO will save the lab 5 International Migration.pdf to the DOWNLOAD folder of your computer. Locate and open the pdf document and use it to answer the following question.
- 3). Now go to the week 7 FOLDER in blackboard and follow the instructions inside that folder to upload the Lab 5 International Migration.pdf and complete the lab 5 assignment.