Quantitative modelling of race, ethnicity and migration issues

By Carlos Vargas-Silva, University of Oxford

19/03/2021

A recent seminar series at the Centre on Migration, Policy and Society (COMPAS) looked at the links between issues of race, ethnicity and migration. As the Editor-in-Chief of the Migration Studies journal I often have to ask myself if a submitted paper is about migration (i.e. interest of the journal) or about related topics such as race and/or ethnicity (and therefore a better fit for another journal). I always struggle with this type of decision and there is a major degree of subjectivity involved.

One of the sessions of the seminar series, in which I participated, was about approaching these questions from different methodological and disciplinary perspectives. I have long rescinded the idea of pursuing knowledge from a given unique disciplinary perspective, yet most of my research is quantitative in nature. Hence, I decided to explore these issues from the perspective of quantitative modelling. By this, I mean the use of statistical tools, such as regression analysis, to understand questions related to race, ethnicity and migration. For example, the analysis of a dataset to understand the importance of migration status and ethnicity in explaining pay gaps across different groups.

One necessary aspect of quantitative modelling is the use of adequate datasets for the analysis. That is, it is necessary to codify categories related to race, ethnicity and migration in order to analyse the data. Many key variables could be included in the analysis, but I will focus on six different variables as presented in Table 1 below: location, birthplace, passport, race, ethnicity and language. My main objective is to explain issues related to the variability of each category across time and space.

Table 1 — The six categories

Location	Birthplace	Passport	Race	Ethnic	Language

For this purpose, I need a useful case study. Look at Image 1. Take away the hair and put some more weight in and you will realise that you are looking at a younger version of myself. I will use my own experience responding to all six categories during my migratory experience as an example of the complexities that lie in these variables.

Image 1 — The case study (around 4 years of age)

Photo: Copyright Carlos Vargas-Silva.

The default is White

I was born in Puerto Rico, a small island in the Caribbean and the kind of place "Where everybody knows your name". In order to provide some perspective in terms of size and population density, I have put Puerto Rico next to Ireland on the same size scale (Figure 1). When I was born, the population of Puerto Rico was 3.1 million, while the population of Ireland (Republic) was 3.4 million. However, as it is clear from Figure 1, Puerto Rico is a lot smaller. Today, the population of Puerto Rico is still around 3.1 million, while the population of Ireland has increased to 4.9 million. This difference reflects the high level of emigration from Puerto Rico and the diverging economic fortunes of the two places. To date, there are more Puerto Ricans outside of Puerto Rico than within the country.

Figure 1 — Puerto Rico: geographical location and comparison with Ireland
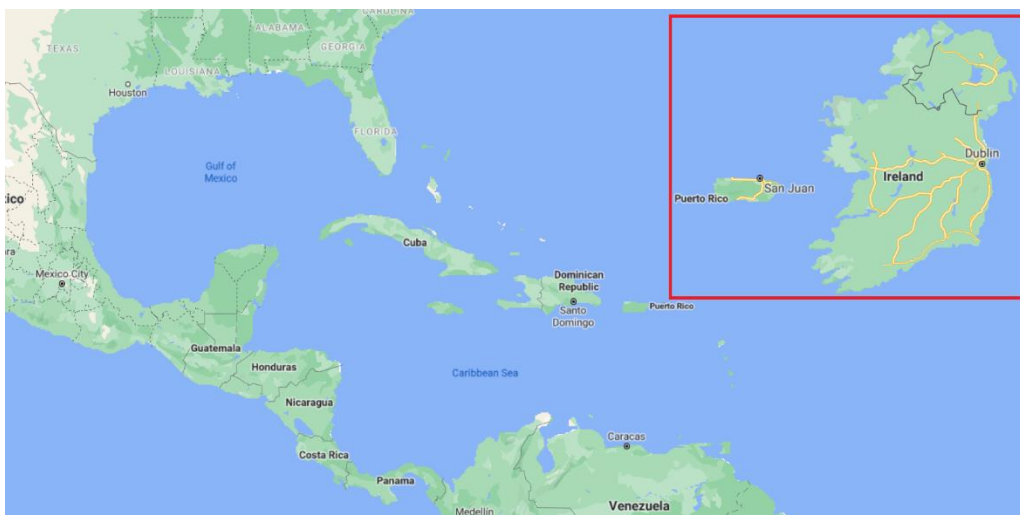
Christopher Columbus arrived to Puerto Rico for the first time, at least officially, in 1493 (second voyage). The island was a Spanish colony until 1898. That year the United States invaded and took possession of the island as part of the Spanish American war. Since then, the island has belonged to the United States (i.e. "belongs to", but it is "not part of"). In 1917, Puerto Ricans were granted US citizenship, but even after more than 100 years as a USA non-incorporated territory (i.e. possession or the equivalent of a colony), Spanish remains the main language in the island.

With this historical explanation, it is relatively easy to fill out four of the cells in Table 1: Location = USA, Birthplace = USA, Passport = USA, Language = Spanish.

The concept of race is more complicated. Most people visiting Puerto Rico would probably think that it has a very diverse population, which is the result of the mix between the pre-colonisation habitants of the island (i.e. indigenous groups such as Tainos), the colonizing power (Spain) and the slaves brought from Africa to work in the sugar cane plantations.

However, in the Census typically over 80% of respondents will describe themselves as White (Figure 2). The majority of Puerto Ricans recognise that they are the result of mixing of three races (Spanish, Taino, African), but think of the Spanish one as the central one. That is, the mixing of the three races did not lead to a mixed race result and over 90% described themselves as belonging to one race in the Census. While in the USA, famously, one drop of Black blood often makes people Black, in Puerto Rico the one-drop rule is reversed. One drop of European blood makes people White.

Figure 2 — Racial classification response in the Census for Puerto Rico
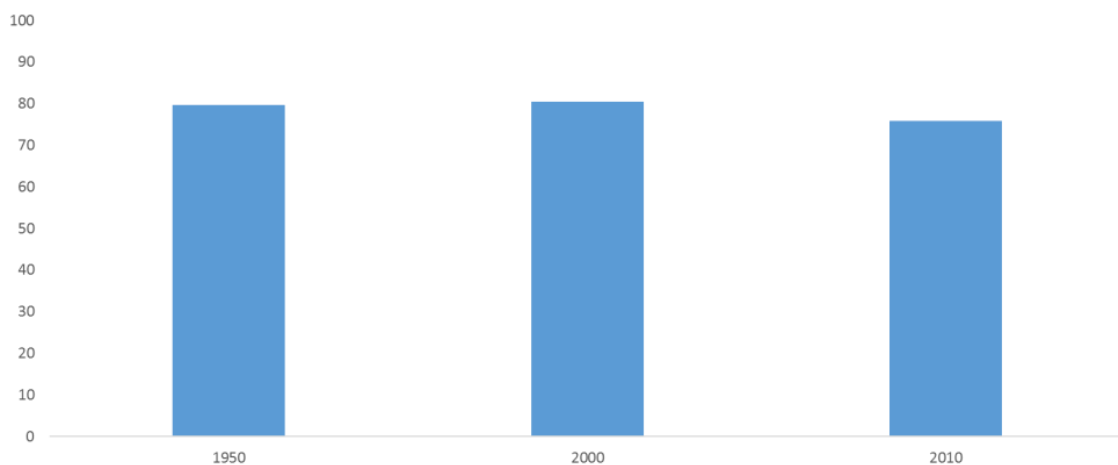
Take for instance, Image 2, in which it is possible to observe that my brother has a much lighter skin than I do. These kinds of variations in skin tone make little difference; the default option is to be White. In school, out of a class of 30 students, there were only 3 or 4 kids that we would consider to be Black. In order to be Black, the person had to be "pure Black", that is, no mixing could be reflected in the physical appearance of the person. This is uncommon in an island in which mixing is common.

Image 2 — The case study (around 7 years of age) and brother

Photo: Copyright Carlos Vargas-Silva.

This racial hierarchy is also reflected in all aspects of the culture, including classroom teaching. Image 3 shows the emblem of the Institute of Puerto Rican Culture, the main organisation for cultural issues in the island. It is clear that in the emblem, approved in 1956, the Spanish conquistador is at the centre, giving it a central role in terms of culture while on the sides there is an African and a *Taino* (both semi-naked).

Image 3 — Emblem of the Institute of Puerto Rican Culture

However, the importance and association with different racial categories has been changing over time. While the demographic composition of the island has not changed, the share of those identifying as *White alone* has decreased and it currently stands at 64%. This is a substantial change from 2000. I have not done any study on this issue, as my understanding of race in Puerto Rico comes from lived experiences rather than academic analysis. However, I suspect that racial identification in the island has been likely influenced by the greater visibility of racial issues across the world, particularly in the United States.

Table 2 — Racial classification response in Puerto Rico for the American Community Survey (2019)

	Total	Share
White alone	2,054,764	64.34
Black or African American alone	395,439	12.38
American Indian and Alaska Native alone	3,082	0.10
Asian alone	6,573	0.21
Native Hawaiian and Other Pacific Islander alone	183	0.01
Some other race alone	506,563	15.86
Two or more races	227,090	7.11

This means that we can fill one more cell in Table 1. Growing up I self-identified as White.

Finally, I never heard of the term ethnicity before leaving Puerto Rico. As such, we can complete the first round of categories for our six variables as:

Table 3 — The six categories (Round 1)

Location	Birthplace	Passport	Race	Ethnic	Language
USA	USA	USA	White	????	Spanish

You are a Latino

My first migration experience was to the United States. Soon after arriving, I had to visit the dentist. Going to the dentist has been always a traumatic experience for me, but this one was a bit more traumatic than usual.

The secretary of the dentist was a friendly older woman. I explained to her that I had just arrived to the country and that I was just learning my way around and getting used to how things work. She provided me with a form to fill out in order to include my details in the system. That form included a question about race/ethnicity, along the lines of Image 4. As explained above, then I had never doubted my racial identity, hence, I checked the *White* box. After returning the form, she "corrected" the form and

explained that I was a *Latino*. She further explained that this is a racial/ethnic category that includes anyone coming from, or with ancestry from, Mexico or any country to the south of Mexico.

Image 4 — The form at the dentist

Race /Ethnicity
- White ☐
- Black ☐
- Latino ☐
- Asian ☐
- Other ☐

Losing a previous racial identity can be shocking, but it can also have benefits. I quickly learned that there were scholarships for Latinos and other programmes to support those with that racial and ethnic identity or its equivalent categories (e.g. Hispanics). In addition, while I personally never fully identified with the category, it did affect my social relationships in the United States. I felt closer to Mexicans, Cubans, Colombians and other groups that fall under that same category. I even married a "Latina".

Yet, it is still a very broad category. It is, therefore, no surprise than in the recent US election many "Latinos" supported Trump, while he was insulting and demeaning other "Latinos". Some would feel offended by his comments while others would not. The umbrella is just quite big.

Hence, from the six categories, location, birthplace and passport have not changed, and the racial and ethnic categories have combined under the Latino umbrella. The one category left is language. Spanish is a strong second language in the United States. It is possible to access services, buy groceries, watch television and listen to the radio in Spanish. In some cities such as Miami, Spanish is by far the dominant language. Therefore, my main language remained the same.

Table 4 — The six categories (Round 2)

Location	Birthplace	Passport	Race	Ethnic	Language
USA	USA	USA	White	????	Spanish
USA	USA	USA	Latino		Spanish

The Caribbean

My second migratory experience was to the United Kingdom. Similar to the case of the United States, soon after arriving I received a form in which I had to identify my ethnic group. The categories in the form looked similar to those in Image 5. This form was very confusing. First, there was no Latino category. There were several variants of the White category, but after my experience in the United States, I was clear that those were not for me. However, one category sounded like quite a good fit: Caribbean.

When I say a good fit in this case, I really mean it. I was born and raised in the Caribbean. Unlike the Latino category, which was never a great fit, this one felt natural. My lived experiences in Puerto Rico are probably closer to someone from Jamaica, than to someone from Uruguay, even with the

differences in language and colonial history. I was filling this form online, hence there was no one to look at me and "correct" my choice as it happened at the dentist office in the USA.

Image 5 — The form in the UK

Ethnic group
- White British ☐
- White Irish ☐
- White Other ☐
- Indian ☐
- Pakistani ☐
- Bangladeshi ☐
- Chinese ☐
- African ☐
- Caribbean ☐
- Other ☐

Hence, going back to our six variables now Location = UK, Birthplace = USA, Passport = USA, and Race/Ethnic = Caribbean.

Defining my language was one of the most interesting experiences in the UK. As part of the migration process, it is necessary to show that you can speak English at a certain level. For this purpose, countries are divided into two types. First, there are countries from which individuals need to take an exam in order to show that they can speak English. Second, there are countries from which individuals are assumed native English speakers. In light of my US passport, I was assumed an English native speaker, but anyone hearing my accent would quickly realise otherwise.

Table 5 — The six categories (Round 3)

Location	Birthplace	Passport	Race	Ethnic	Language
USA	USA	USA	White	????	Spanish
USA	USA	USA	Latino		Spanish
UK	USA	USA	????	Caribbean	English

You are not Black

The joy of finding The Caribbean category was short- lived. I interpreted The Caribbean category as a reference to a place, regardless of other issues related to race or colonial history. Others, who identified with this category, explained that this category was mainly a reference for individuals of Black race coming from former British colonies in the Caribbean. As such, my lived experiences did not fit the category well. In fact, they suggested that I was White.

I contested this latter point by explaining my experience in the United States and that I did not fit the White category. However, the White category in the UK is different. It includes several different types of White individuals, such as White British, White Irish and the "very inclusive" category of White Other. Hence, after not being White for a long time, I returned to being White, or better to say: White Other.

In addition to my inclusion in the White Other category, I was placed socially in the same group with Southern Europeans. Again and again, this was described as my natural social group by others. For instance, at Oxford people would make jokes about people from Spain or from Southern Europe in general in order to make fun of me. They were trying to be nice and friendly, there was no bad intention, but I had no connection with the jokes. However, as it happened in the USA with other Latino groups, over time I did start to identify socially with individuals from Southern Europe.

The last variable that changed during this period was my passport. In Image 6, you can see the day all four of us became UK citizens. Even though my son was born in Oxford, being born in the UK does not make you a British citizen. It depends on the status of your parents at the time of your birth. My son slept through the whole process, and we always joke about whether you can have a valid citizenship if you sleep through the ceremony. My daughter, on the other hand, was very excited about everything. She was just a bit disappointed that the Queen was not there in person (there was a very large picture!).

Image 6 — The case study (in his 30s) with family

Photo: Copyright Carlos Vargas-Silva.

Of course, while we still have our previous passport, for statistical purposes in the UK, we are UK nationals. As such, we can complete the six variables as:

Table 6 — The six categories (Rounds 4 and 5)

Location	Birthplace	Passport	Race	Ethnic	Language
USA	USA	USA	White	????	Spanish
USA	USA	USA	Latino		Spanish
UK	USA	USA	????	Caribbean	English
UK	USA	USA	White - other		English
UK	USA	UK	White - other		English

There is a question of whether the categories reflected in Table 6, refer to self-perception or to the perception of others. In my case, these two factors are strongly linked. I stopped thinking about myself as White and, later on, Caribbean, because I was told that I belonged to another category. Obviously, each person is different and for some individuals the perception of others could have less of an impact on their own self-perceptions. In my case, when arriving at a new place, the information provided by others was very important.

The second generation

At this stage, five rounds of categories are perhaps enough for any one person. However, it is important to highlight that these issues change across generations, while not necessarily going away. For instance, I have heard my daughter, not much older now than I was in Image 2, describe herself as mixed race as she perceives that my wife and I are from different racial groups. However, my wife and I see ourselves as part of the same racial group. Obviously, my daughter is much more ingrained in the UK racial perceptions than we are.

Lessons for research

In order to consider the implications of the discussion above for research, it would be good to start by considering the green horizontal arrow in Table 7. This arrow refers to research that is cross-sectional in nature and that can be estimated relatively easy with most existing datasets. That is, at some point in time, in a given location, it is possible to estimate the impact of the other five elements in the Table. For instance, what is the impact in the United States of having Spanish as a first language on earnings?

It is also possible to estimate relatively easy the relationship presented by the vertical green arrow. For instance, it is possible to estimate for someone in the US, what is the implication for earnings of being born abroad? (or in even in a particular country)

It is much more difficult to estimate factors related to the red vertical arrows. There are at least, two difficulties here. First, the racial and ethnic categories depend on the location of an individual. Moving locations means that the previous categories might not even exist (e.g. Latino in the UK) or mean something different (e.g. White in Puerto Rico versus White in the United States). Second, even in the same place the racial category of the individual could change. This could happen, for instance, because there is new information (e.g. you do not fit the Caribbean category) or because the way the data is collected changes over time (e.g. change in census racial/ethnic categories).

Table 7 — Implications for research

Location	Birthplace	Passport	Race	Ethnic	Language
USA	USA	USA	White	????	Spanish
USA	USA	USA	Latino		Spanish
UK	USA	USA	????	Caribbean	English
UK	USA	USA	White - other		English
UK	USA	UK	White - other		English

Implications of factors that do not change

Implications over time of factors that do change

Implications of the situation at any point in time

One big question is what to do in order to better answer questions such as those represented by the red lines. There are three factors that can help in answering those questions:

1. Collecting longitudinal data on migrants, as they move across countries, including data on location.

2. Ask information on race and ethnicity on all rounds of data collection. If any of those categories changes across rounds, ask why.

3. Ask information on all nationalities, and all languages spoken on all rounds of data collection. For languages data on fluency is important.

Collecting data in this way will not provide us with perfect information about the implications of changing racial and ethnic categories across time, but can go a long way in the direction.