Matias Cortes
York University

Final Report and Preliminary Data Analysis: Instructions

This document provides details on the submission of your **Final Report** (due **December 6 at 6:00pm** and worth **25%** of your overall course grade) and of your **Preliminary Data Analysis** (due **November 15 at 6:00pm** and worth **5%** of your overall course grade).

Your **Final Report** should present and discuss the results from your data analysis based on your approved topic choice. The structure of the report should be similar to the practice report that you have submitted for the in-class Stata exercises, but this final report will be more detailed and will be based on your individually chosen topic.

Your report **must include at least 3 "exhibits" (i.e. tables, graphs, or combinations of the two), and should be at least 900 words long**. You must also copy and past the code from your Stata **do-file** showing all of your data analysis (the text from the code does not count towards the word requirement). I have posted an example of a good report submitted by a student during a previous term to eClass.

In order to provide you feedback on your data work, you must submit your **Preliminary Data Analysis** through eClass by November 15 at 6:00pm**.** This submission should simply be a document that includes the three tables and/or graphs that you intend to include in your Final Report, as well as some brief bullet points describing each of them. It does not need to include any detailed writing. I will review your graphs and tables, and will provide feedback on whether you are on the right track. I will also make suggestions for improvement, if relevant.

The **Final Report** must be submitted through eClass by December 6 at 6:00pm. The **grading criteria** for your Final Report (out of a maximum of 25 points) are as follows:

- Referencing the source of the data **[1 point]**
- Describing any restrictions or manipulations imposed on the original data from IPUMS (such as observations that are dropped, variables that are generated, etc.) **[2 points]**
- Providing the number of observations that your analysis is based on (before collapsing, if relevant) **[2 points]**
- Showing results from your data analysis in clear graphs and/or tables **[8 points]**
- Providing a comprehensive and clear description and discussion of the results of the analysis that are shown in your graphs and/or tables **[8 points]**
- Overall quality of the writing **[4 points]**

Note: If you do not copy and paste the code from your do-file at the end of your submission, you automatically lose all points and receive a grade of zero.

Final Report

This report compares the educational composition and average earnings of immigrants from Italy to the educational composition and average earnings of people born in the US over time using data for the years: 1960, 1970, 1980, 1990, 2000, and 2010.

The following analysis is based on data from the American Community Survey (ACS), as obtained from IPUMS-USA (Ruggles et al, 2022).

The variables used in this report are: *"bpl", "educ" "incwage".* The variable *"bpl"* documents a person's place of birth. Values of this variable that record a person's place of birth not being from Italy or the United States are excluded from the analysis. The variable *"educ"* notes the educational attainment of the individual. Individuals are grouped into 12, ranging from no schooling to a Doctorate. No group was removed, for instance, the "no schooling" was maintained, considering the selected years of the analysis. The 12 groups were then divided initially into four categories by education level as follows: grade eleven or below, a high school diploma, some college, and a bachelor's degree or above. In the first analysis, the groups are condensed into two educational attainments: "lower education" and "higher education". People having a lower education are defined as having some college or below and people with higher education as having a bachelor's degree or above. Lastly, the variable *"incwage"* reports yearly wages and salaries of individuals, not adjusted for inflation. The missing values and not applicable observations are replaced. The educational attainment of immigrants from Italy and people born in the US are compared, with a detailed focus on the years 1960 and 2010; further, the relationship between educational attainment and yearly income is presented.

The analysis will include six graphs. The first graph will compare the educational attainment between Italian immigrants and people born in the US, outlining the greater picture. The next two graphs will compare in detail the educational attainment between Italian immigrants and people born in the US for two specific years: 1960 and 2010. The last three graphs show first a greater picture of the earnings across years, and lastly a comparison of the educational attainment and the yearly earnings between Italian immigrants and people born in the US, for the years 1960 and 2010

The sample contains 36,450,459 observations, of these 36,269,933 are people born in the US, and 180,256 are Italian immigrants. For the year 1960, the sample contains 5,610,619 observations of which 61,004 are immigrants from Italy. For the year 2010, the sample contains 2,109,609 observations of which 4,530 are immigrants from Italy.

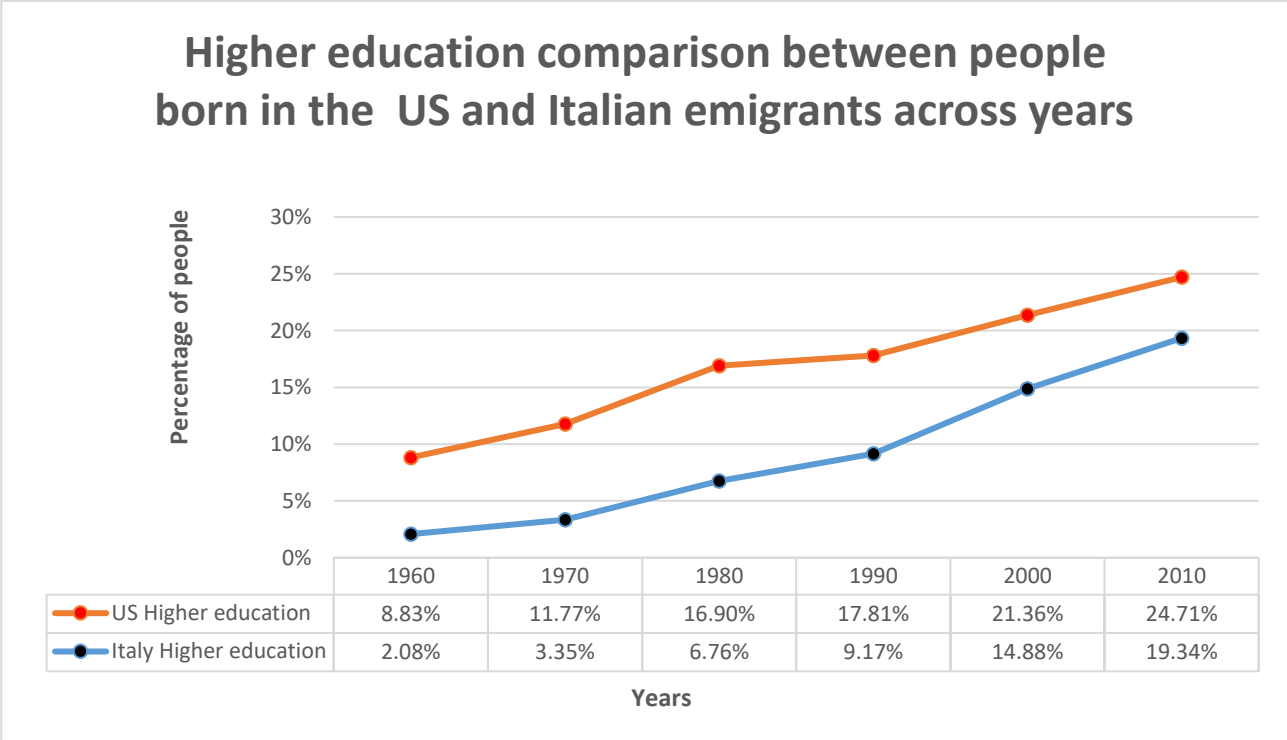
The following table summarizes the share of Italian immigrants and people born in the US for each year:

Birthplace	Census Year						
	1960	1970	1980	1990	2000	2010	Total
US	5,549,615	2,619,403	7,878,028	8,675,636	9,442,172	2,105,079	36,269,933
Italy	61,004	19,540	41,403	29,625	24,424	4,530	180,526
Total	5,610,619	2,638,943	7,919,431	8,705,261	9,466,596	2,109,609	**36,450,459**

Individuals below sixteen years of age have been excluded from the analysis. For each birthplace country, Italian immigrants and people born in the US, the number of individuals and their percentage for each educational attainment group is calculated. The analysis weights individuals using the variable *"perwt"* for all variables: *"educ"* and *"incwage".* Individuals are identified by using the

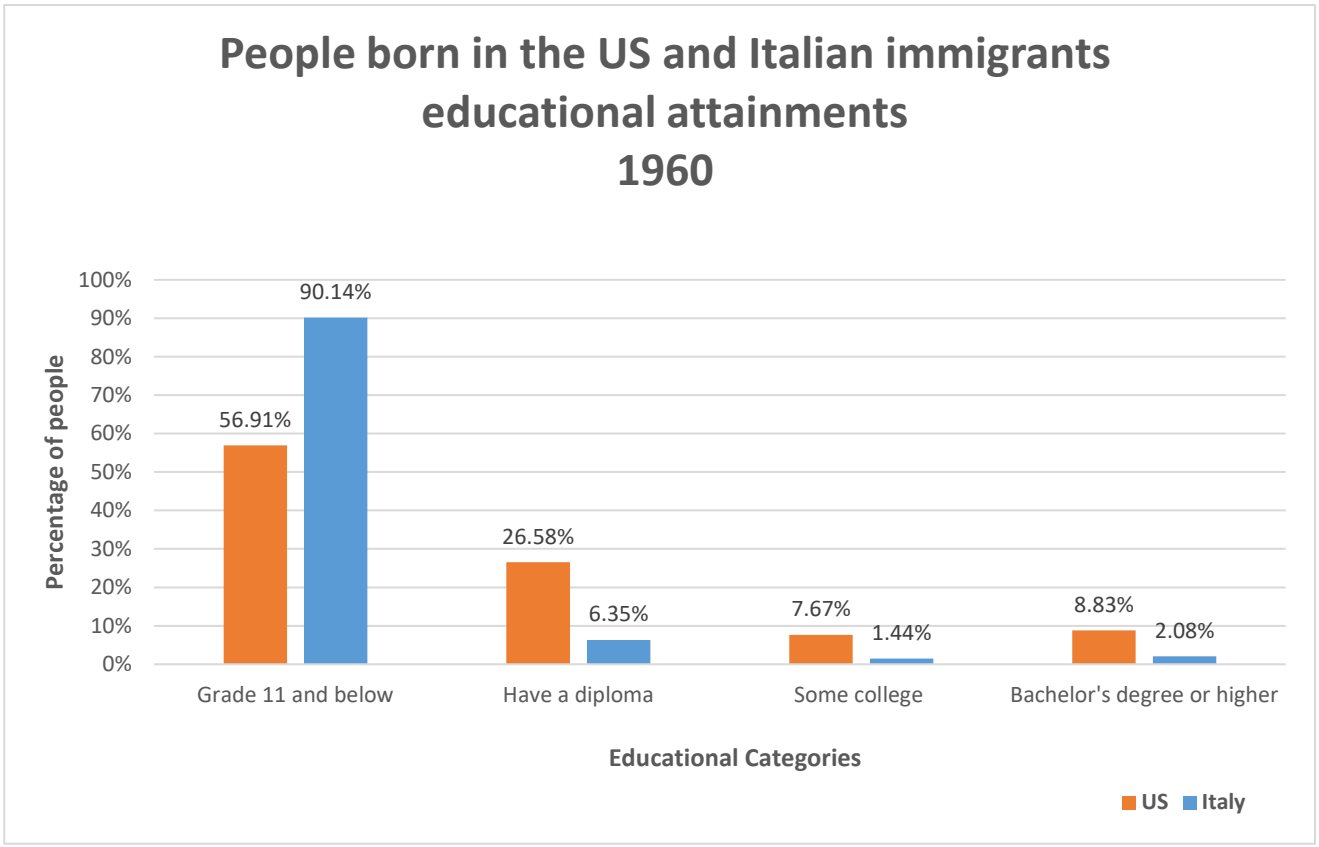
count function and grouped by year. Following the replacement of missing value earnings and not applicable earnings, the average weighted earnings are calculated and sorted by educational group, birthplace, and year.

The graph below shows the trend line of the higher educational attainment for Italian immigrants and people born in the US for the years: 1960, 1970, 1980, 1990, 2000, and 2010; as obtained from the data:



The data shows that the percentage of people obtaining a higher education is increasing over time for Italian immigrants and people born in the US. The percentage point change for the people born in the US between the years 1960 and 2010 has been 16.83%. The percentage point change for Italian immigrants since the year 1960 has been 17.26% in 2010. The largest percentage point difference between ethnicities in higher education is seen in the year 1980. The graph shows a 10.14% difference between Italian immigrants and people born in the US in 1980. The gap between birthplace countries has been narrowing in the most recent years under analysis. The narrowest percentage point difference between ethnicities, in higher educational attainment, is seen in 2010, a 5.37% difference.

The figure below compares the educational attainment of Italian immigrants to the educational attainment of people born in the US, for the year 1960 as obtained from the data:

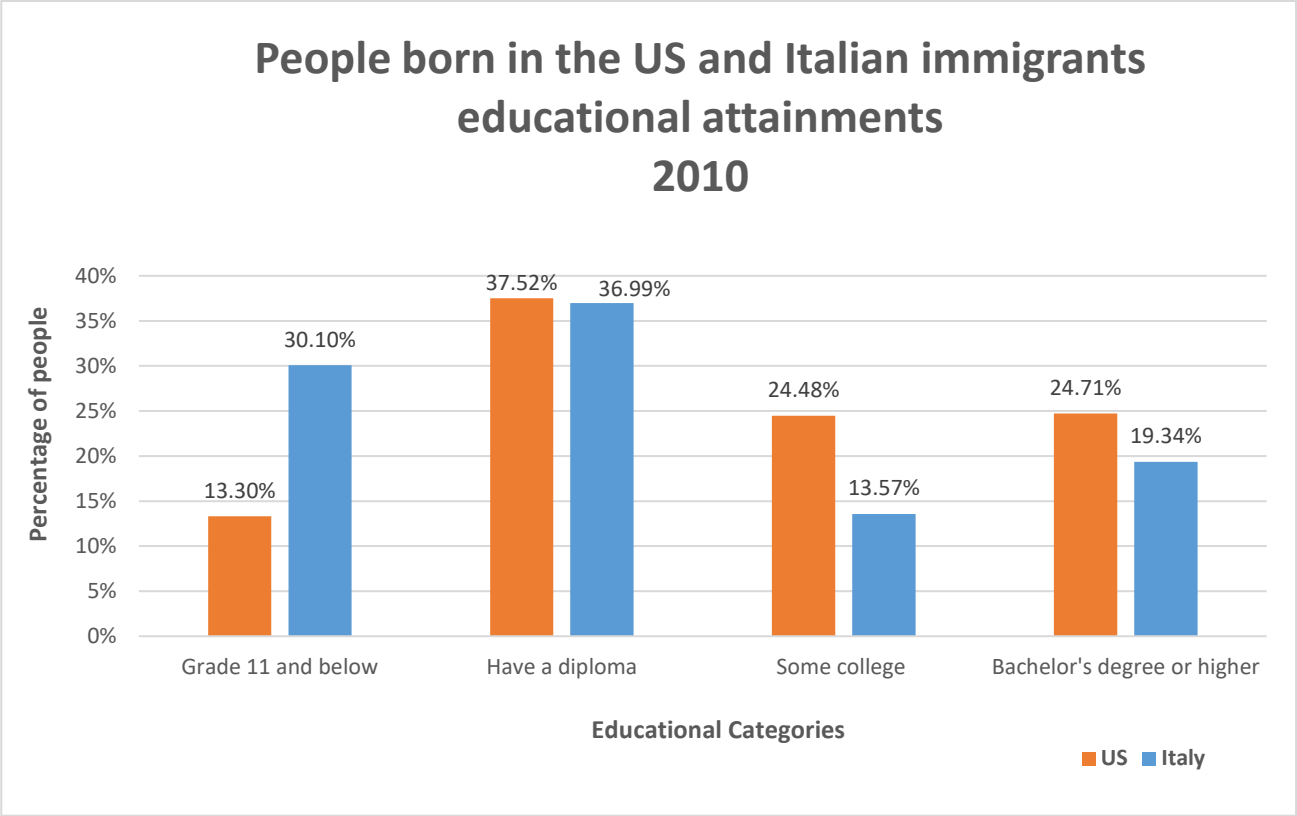


The data shows that individuals that have obtained a grade eleven or below are the majority among the four categories, for both countries. The number of people born in the US is slightly lower at 56.91% compared to Italian immigrants at 90.14%. The other three categories are significantly lower for both countries. The high school diploma category is composed of 26.58% of people born in the US and 6.35% of Italian immigrants. Individuals with some college are 7.67% for people born in the US and 1.44% for people immigrating from Italy. In the 1960's individuals born in the US and having a bachelor's degree were 8.83% of people and 2.08% were Italian immigrants having a bachelor's degree.

These results show that most people in the year 1960, from both birthplace countries, had an education lower than the eleventh high school grade. This is seen in the data primarily for Italian immigrants. Furthermore, the number of people that had an education of a high school diploma and above was less than 10% for Italian immigrants. People born in the US having a high school diploma and above, on the contrary, were 43.09%. Surprisingly, Italian immigrants were 33.23% more likely to only have an educational attainment of grade eleven or below.

It would be interesting to do some further research and analyze the job choice of Italian immigrants in the year 1960 compared to that of US citizens by birth, to uncover the job placements for the 90.14% of Italian immigrants having a grade eleven education.

The graph below compares the educational attainment between Italian immigrants and people born in the US, in the year 2010, as obtained from the data:

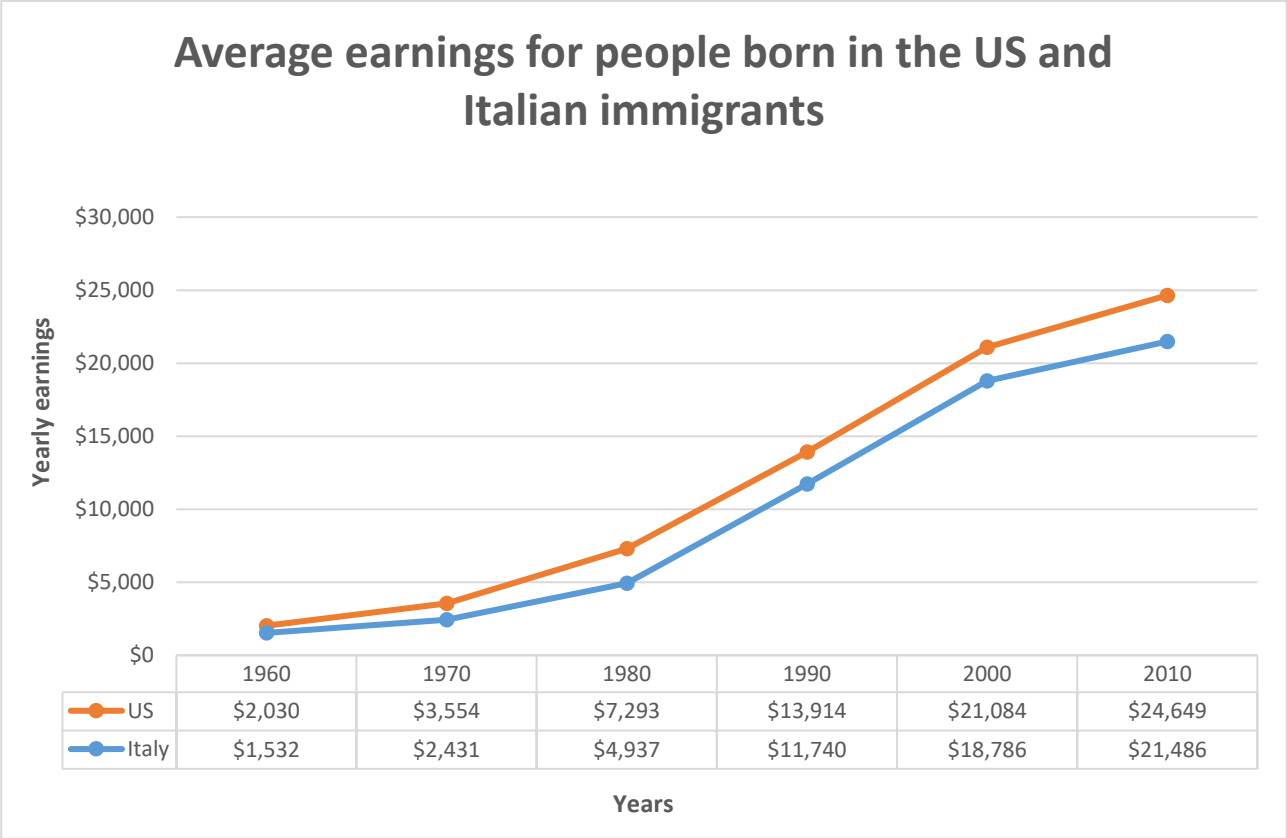


The data reveals that the percentage of individuals with a high school diploma is the highest category for both countries of birth, people born in the US are 37.52% and Italian immigrants are 36.99%. The lowest educational category for people born in the US is having a grade eleven or below, 13.30%, and the lowest category for Italian immigrants is having some college, 13.57%. People having a bachelor's degree or higher are the second highest category for people born in the US, 24.71%, and the third highest for Italian immigrants, 19.34%. The second highest category for Italian immigrants is having a grade 11 or below, 30.10%, and the third highest category for people born in the US, is having some college, 24.48%.

The results deviate from the paired trend shown in the year 1960 graph, where both people born in the US and Italian immigrants had the same highest and lowest categories. There was the same order of magnitude, from grade eleven being the largest category by percentage and a bachelor's degree category including the lowest percentage of people for both countries of birth. Now the data shows, for the year 2010, an increase in bachelor's degree attainments and a decrease in the percentage of people who have less than a grade eleven. It is interesting to notice that Italian immigrants, in 2010, with an educational attainment of grade eleven or below had a percentage point difference of about 60% lower compared to the year 1960. There were 60% fewer people having only a grade eleven or below in 2010. The percentage point change in people having a high school diploma is also interesting, compared to 1960. The people born in the US experienced about a 10% difference, in the high school diploma category; meanwhile Italian immigrants that obtained a high school diploma experienced a percentage point change of 30% compared to the year 1960. The most interesting category is the increase in bachelor's degree attainment for both countries compared to the 1960s. In 2010 bachelor's degrees attainment increased to 24.71% for people born in the US and 19.34% for Italian immigrants. The two birthplace countries have a 5.34% difference in achieving a bachelor's degree, in the year 2010.

It would be interesting to see what drives the 5.37% difference between the people born in the US and Italian immigrants in achieving a bachelor's degree in 2010, could it be differences in government support through loan support and grants? Or may it be due to preferences rather than income? Further analysis would help to uncover these questions.

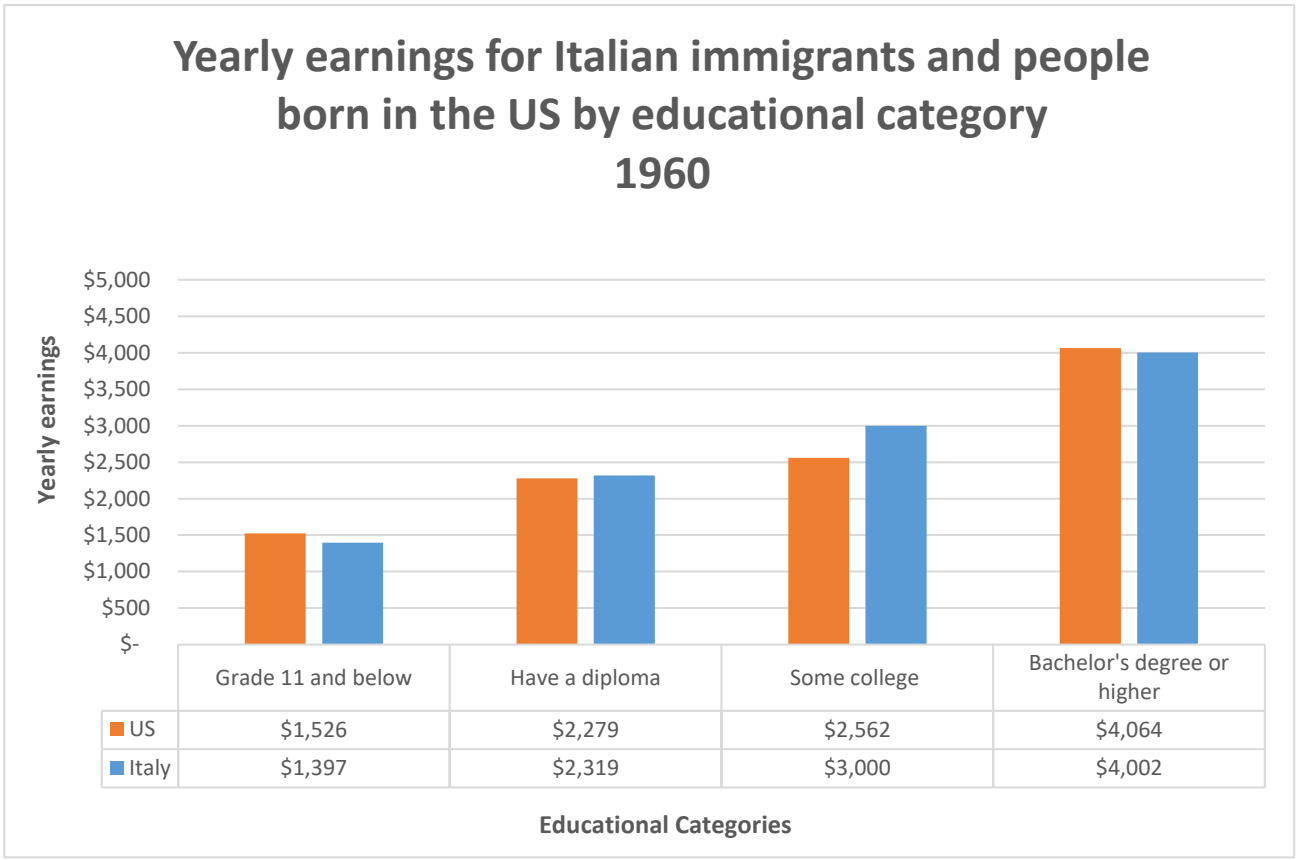
The graph compares earnings between people born in the US and Italian immigrants for the years: 1960, 1970, 1980, 1990, 2000, and 2010; as obtained from the data:



	1960	1970	1980	1990	2000	2010
US	$2,030	$3,554	$7,293	$13,914	$21,084	$24,649
Italy	$1,532	$2,431	$4,937	$11,740	$18,786	$21,486

The data shows earnings for both countries of birth have been increasing since the year 1960. The average weighted earnings for people born in the US are higher than the average earnings for Italian immigrants. The difference in average earnings between the two countries has been widening reaching a difference of $3163 in 2010. The narrowest difference between people born in the US's average earnings and Italian immigrants is seen in 1960, the difference is $498. Italian immigrants were earning 75.47% of people born in the US, in 1960; however, in 2010, Italian immigrants were earning 87.17% of people born in the US. The largest increase in average earnings for both countries is between the years 1980 and 2000, the graph shows a sharp upward trend.

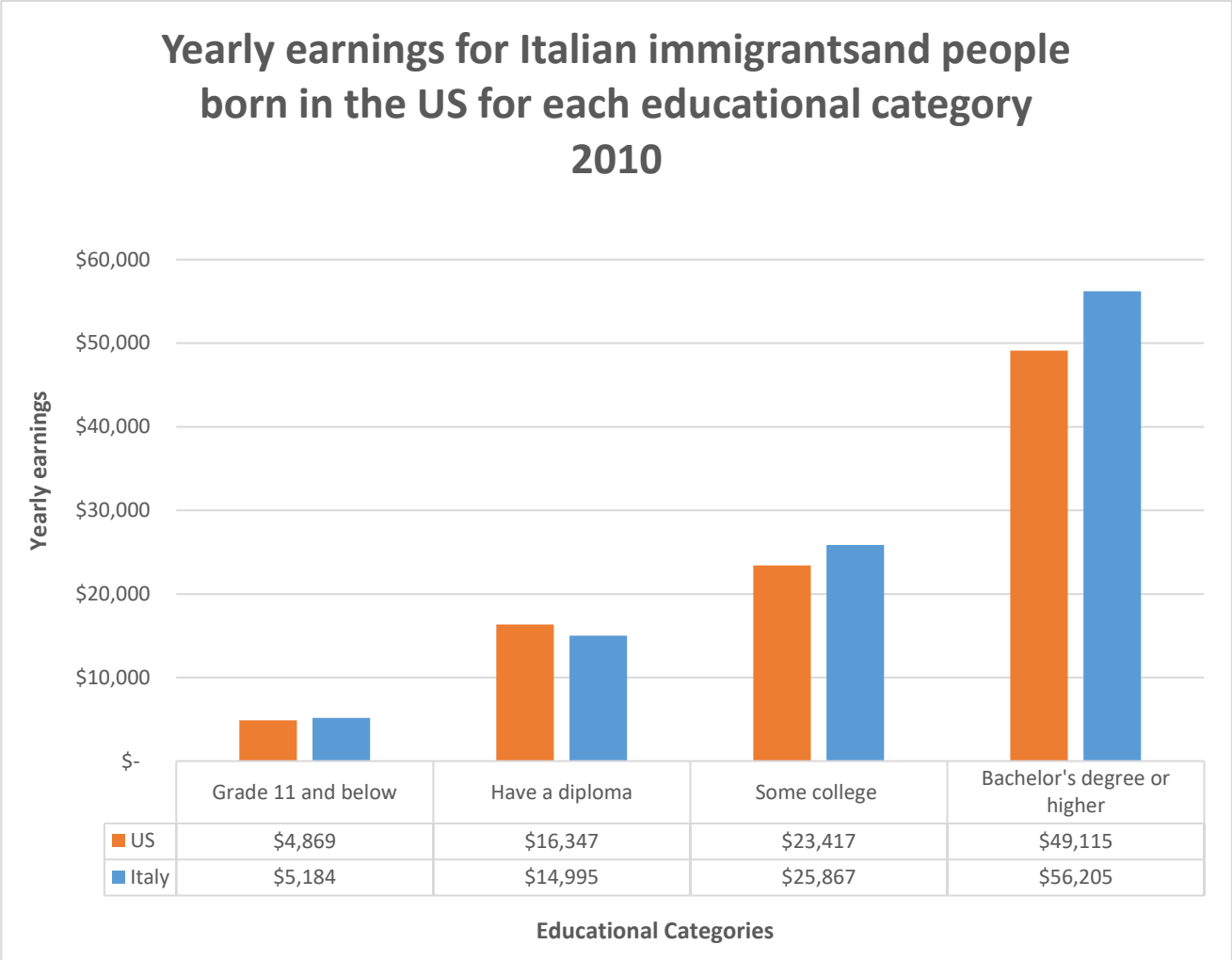
The graph below compares the average earnings for Italian immigrants and people born in the US for each educational attainment group for the year 1960, as obtained from the data:



The data clearly shows that people, of both countries of birth, had a higher yearly income with a bachelor's degree or above, in 1960. Italian immigrants with a bachelor's degree or higher earn $4002 and people born in the US earn on average $4,064, a $62 difference is evident between the two birthplace countries. People that achieved less than a grade eleven, have the lowest yearly income. Italian immigrants that have a grade eleven or below have a yearly income of $1,397 and people born in the US in the same group have a yearly income of $1,526. However, the two birthplace countries are similar in average yearly earnings if they have both obtained a high school diploma. People born in the US having a diploma have a yearly income of $2,279, slightly lower than the yearly income of Italian immigrants in the same group who earn $2,319. The largest difference between people born in the US and Italian immigrants is found in the yearly income of people who have some college. Italian immigrants with some college earn $3000 compared to people born in the US who earn a yearly income of $2,562, a $438 difference between the two countries of birth.

The yearly income difference between people having a bachelor's degree or higher and people having a grade eleven or below is $2,538 for people born in the US and $2,606 for Italian immigrants. Overall we notice that people born in the US earned more, in dollar terms, in the year 1960 than Italian immigrants, in the grade eleven category and the bachelor's degree and above category.

The graph that follows shows the relationship between education and average yearly earnings for people born in the US and Italian immigrants for the year 2010, as obtained from the data:



The data reveals that both Italian immigrants and people born in the US earned the highest yearly income in 2010 when they achieve a bachelor's degree or higher. Italian immigrants with a bachelor's degree or higher earned $56,205 and people born in the US in the same group earned $49,115. The difference between the two countries of birth is $7090. It is interesting that in 2010, Italian immigrants earned much more than people born in the US. People with a grade eleven or below earn the least. The average yearly income for people born in the US was $4,869 and $5,184 for Italian immigrants having less than a grade eleven. Italian immigrants having a diploma earned $16,347 and people born in the US in the same group earned $14,995. People born in the Us with some college earned a yearly income of $23,417 and Italian immigrants earned a yearly average income of $25,867.

It is interesting to note the difference between the earnings of people with a bachelor's degree and people with a grade eleven or less. In 2010, people born in the US, that have a grade eleven or below earn $44,245 less than people with a bachelor's degree or higher. This yearly income difference between educational attainments used to be $2,538 in the year 1960. Likewise, in 2010, Italian immigrants having an educational attainment of grade eleven or below earned $51,021 less than Italian immigrants with a bachelor's degree or higher. The difference in yearly income between the highest educational achievement and the lowest is significant considering that this difference was $2,606, only 50 years earlier. While in the year 1960, people having a bachelor's degree, born in the US earned more than Italian immigrants; in 2010 people born in the US earned

87.39% of Italian immigrants having a bachelor's degree. However, as already mentioned, Italian immigrants with a bachelor's degree and above were 5.37% less than people born in the US.

It would be intriguing to develop further the topic and uncover the underlying reasons that led to this massive change in yearly earnings across educational groups. Could it be because of the ever-more reliance on technology and the shift of our work environment from blue-collar labour to the need for white-collar labour?

In a further analysis, it may also be captivating to distinguish the employment categories in the higher educational attainment group, to uncover the differences in job placements about average earnings. How did the recent year's increase in productivity influence yearly income and preference for higher education across the sectors such as medicine, law, and computer science?

Do-file:

```
use DataProject_US_ITA.dta, clear

*exclude individuals below 16 years of age
drop if age<16

tab educ [aw=perwt]

*Group education into 4 categories
generate educ_4cat=1 if educ<=5
replace educ_4cat=2 if educ==6
replace educ_4cat=3 if educ>=7 & educ<=8
replace educ_4cat=4 if educ>=9 & educ<=11

* Verify with a two-way tabulation
tabulate educ educ_4cat [aw=perwt]

*Group education into 2 categories (lower and higher education) – (separate file re-run for only two categories)
generate educ_4cat=1 if educ<=8
replace educ_4cat=2 if educ>=9 & educ<=11

*generate variable bpl_2cat=1 (US), bpl_2cat=2 (Italy), bpl_2cat=3 and bpl_2cat=4 (everyone else)
tabulate bpl [aw=perwt]

generate bpl_2cat=1 if bpl<=120
replace bpl_2cat=2 if bpl==434
replace bpl_2cat=3 if bpl>=121 & bpl<=433
replace bpl_2cat=4 if bpl>=435

tabulate bpl bpl_2cat [aw=perwt]

*drop country of birth other than Italy and the US
drop if bpl_2cat==3
drop if bpl_2cat==4

tabulate bpl bpl_2cat [aw=perwt]

*educational attainment by birthplace and year (number and percentage of individuals)
bysort bpl_2cat year : tabulate educ_4cat [aw=perwt]

*dropped values
replace incwage=. if incwage==999999 | incwage==999998

*count by birthplace and year
tabulate bpl_2cat year
save EconDataProject_1, replace
use EconDataProject_1.dta, clear
```

***calculate average weighted earnings and group them by birthplace, education, and year**
collapse (mean) incwage [aw=perwt], by (bpl_2cat educ_4cat year)

References:

- Steven Ruggles, Sarah Flood, Ronald Goeken, Megan Schouweiler and Matthew Sobek. IPUMS USA: Version 12.0 [dataset]. Minneapolis, MN: IPUMS, 2022. https://doi.org/10.18128/D010.V12.0