

# **From Rungs to Riches**

## **-College and Mobility**

### **Objectives**

Social mobility is the ability to move up the social ladder from a low socio-economic status (SES) to a higher socio-economic status. It has been observed that it has been declining in the United States in the last few decades. Studies show that each consecutive generation's likelihood of earning more than their parents and moving into a higher social class is decreasing. This trend raises critical questions about how to improve the situation and make social mobility more feasible, especially for individuals from lower social backgrounds. This was the project's motivation, and to analyze that, we picked the roles that colleges play in helping children of low SES parents reach higher SES groups. Below are the key questions the project tries to address.

- What is the relationship between college education and intergenerational mobility?

To answer this broad question and to understand whether the region or a specific tier of college aids intergenerational mobility, we are trying to answer the questions below.

- How does college admission affect lower socio-economic backgrounds?
- Does the region of the college play a role in the relationship?
- How does the tier of colleges have an impact on the relationship?

### **Data description**

The data was sourced from Opportunity Insights contributed by the paper, “Income Segregation and Intergenerational Mobility Across Colleges in the United States.” The data consists of parent’s and children’s income distributions by college, separately by birth cohort. The important variables and their descriptions are below,

## Categorical Variables:

1. Region: 4 different regions in the US (United States)



- Midwest
- Northeast
- South
- West

2. Tier of the Colleges:

- 1 = Ivy Plus
- 2 = Other elite schools (public and private)
- 3 = Highly selective public
- 4 = Highly selective private
- 5 = Selective public
- 6 = Selective private
- 7 = Nonselective 4-year public
- 8 = Nonselective 4-year private not-for-profit
- 9 = Two-year (public and private not-for-profit)
- 10 = Four-year for-profit
- 11 = Two-year for-profit
- 12 = Less than two-year schools of any type
- 13 = Attending college with insufficient data
- 14 = Not in college between the years of 19-22

3. Cohort: 1980-91

The children in the 1980-91 period would be now at the peak of their earning ability, so the cohort of 1980-91 would give the best picture to analyze intergenerational social mobility with respect to college education as they would be usually at their highest level of earning capacity.

## Numeric Variables:

1. Fraction of parents in a specific income quintile (1-5)
  - Fraction of parents in an income quintile. 1 is the bottom quintile and 5 is the top.
2. Fraction of children in a specific income quintile (1-5)

- Fraction of parents in the top percentile. For instance, par\_toppt1pc refers to parents in the top 0.1% of the income distribution.

### 3. Mobility Rate

- Mobility rate of a college is calculated by calculating the joint probability of parents in the bottom range and children in the top range of the income distribution

## Answers to our research questions

### Relationship between different categorical variables:

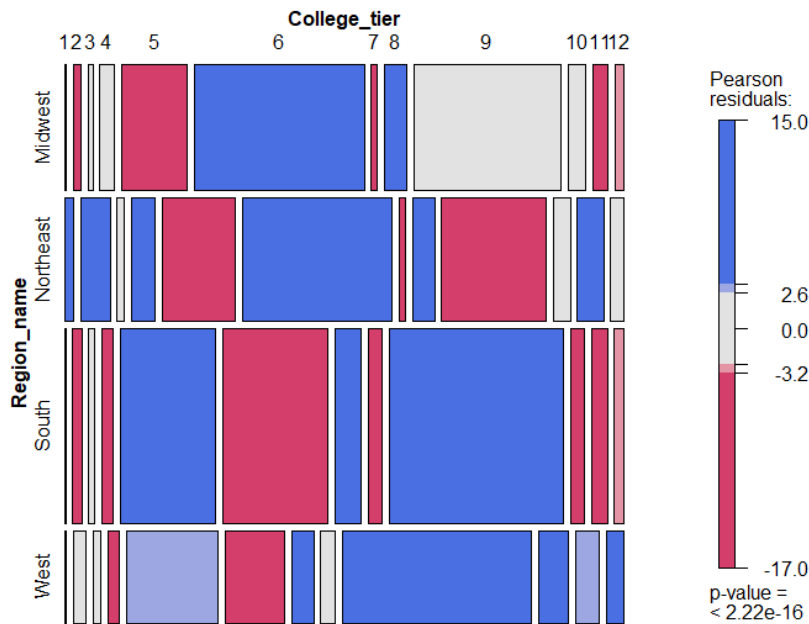


Fig 1 – Mosaic Plot of Region Name and College Tier

From the graph above, we can observe a higher correlation between Region Name and the College Tier. The size of each box segment explains the relative quantity of a specific tier of college in each region. From Fig 1, we can see that the region does affect the distribution among different tiers of colleges. For instance, Ivy Plus colleges in the Northeast region dominate the distribution and have a positive association between them. Whereas regions like the South and Midwest do not show much variation in the data.

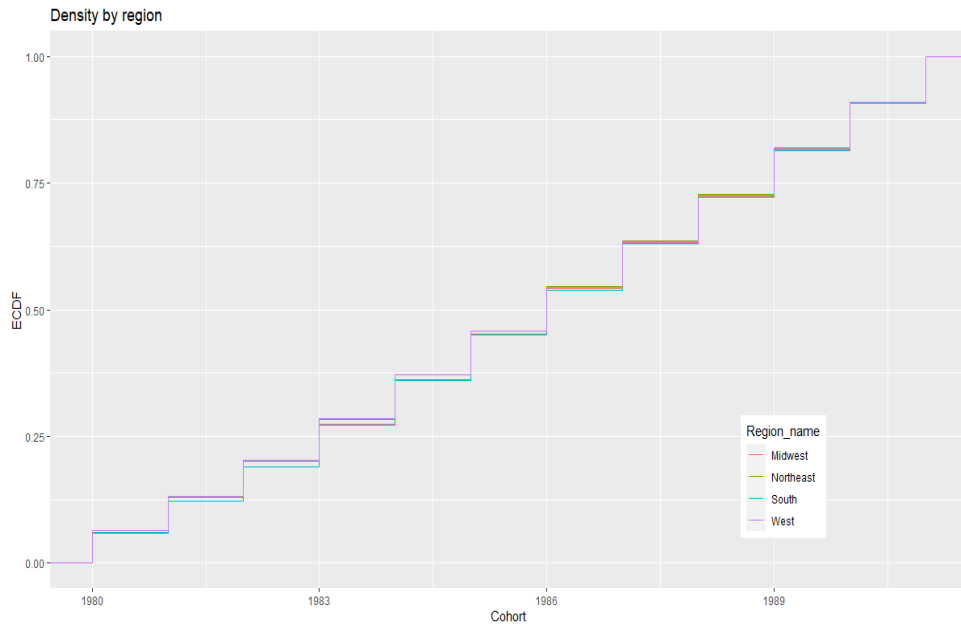


Fig 2: ECDF plot of the region of colleges over the different cohort

To observe if the cohort affects the region distribution, we plotted an ECDF (Empirical Cumulative Distribution Function). ECDF helps us to visualize the distribution of the parameters in a non-parametric way. On our x-axis, we have cohort whereas on the y-axis we have density distribution. From our graph above, we can observe that for all regions, the density is approximately the same and does not change much throughout the years. Hence, we do not observe a significant relationship between the region of the colleges and the cohort.

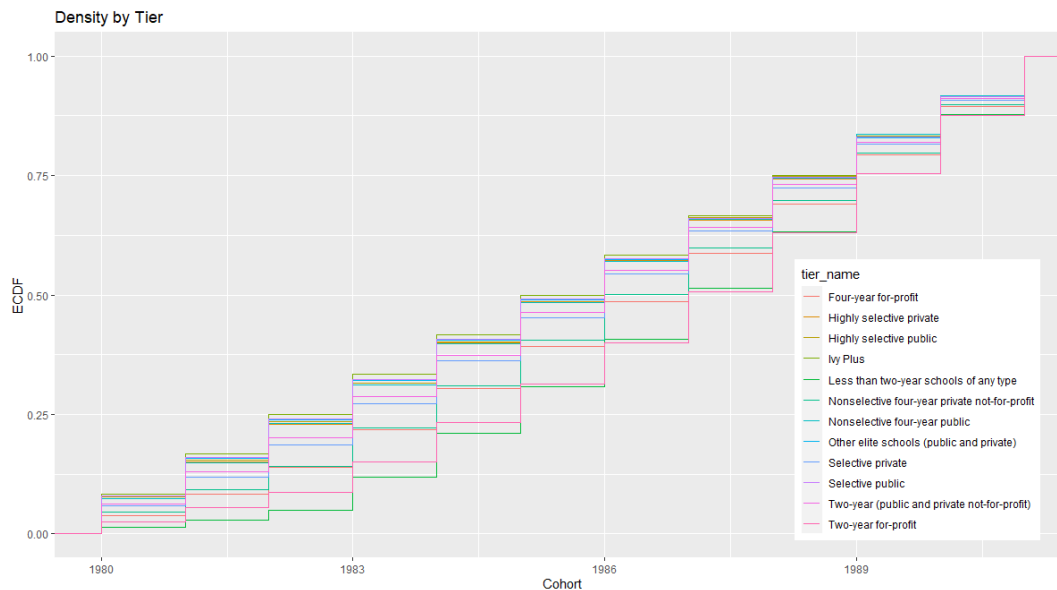


Fig 3 – ECDF of Tier of colleges over the different cohort

To observe if the cohort affects the Tier, we again used the ECDF plot. With cohort on the x-axis and density distribution on the y-axis, we can observe that the distribution is not consistent over the years. Different Tier of colleges shows different distribution in their density over various cohort. For example, the Ivy Plus Tier distribution is completely different from the distribution of

other non-selective colleges from 1980-1991. This clearly shows that there is an underlying relationship between cohort and tier of colleges.

### **Relationship between different Numerical variables (considering those that affect mobility):**

As mobility depends on the lower quintile of the parent's income and on the upper quintile of the child's income



Fig 4 – PCA Plot with the axis of variation shown only for the main variables

To observe the relationship between the numeric variables, we used Principal Component Analysis (PCA). PCA helps to reduce the dimensionality of the data and capture the association between variables. We used PCA on our entire data set of numeric variables but displayed the variable axes only for mobility rate, parent's income in quintile 1 and quintile 2, and children's income in quintile 4 and quintile 5. From the graph above we can observe that the mobility rate is highly associated with the parent's income in quintile 1 rather than being associated with the children's income in quintile 5. This means that the mobility rate is more dependent on the parent's income for the given dataset.

## Model fitting for the relationship between the deciding variables:

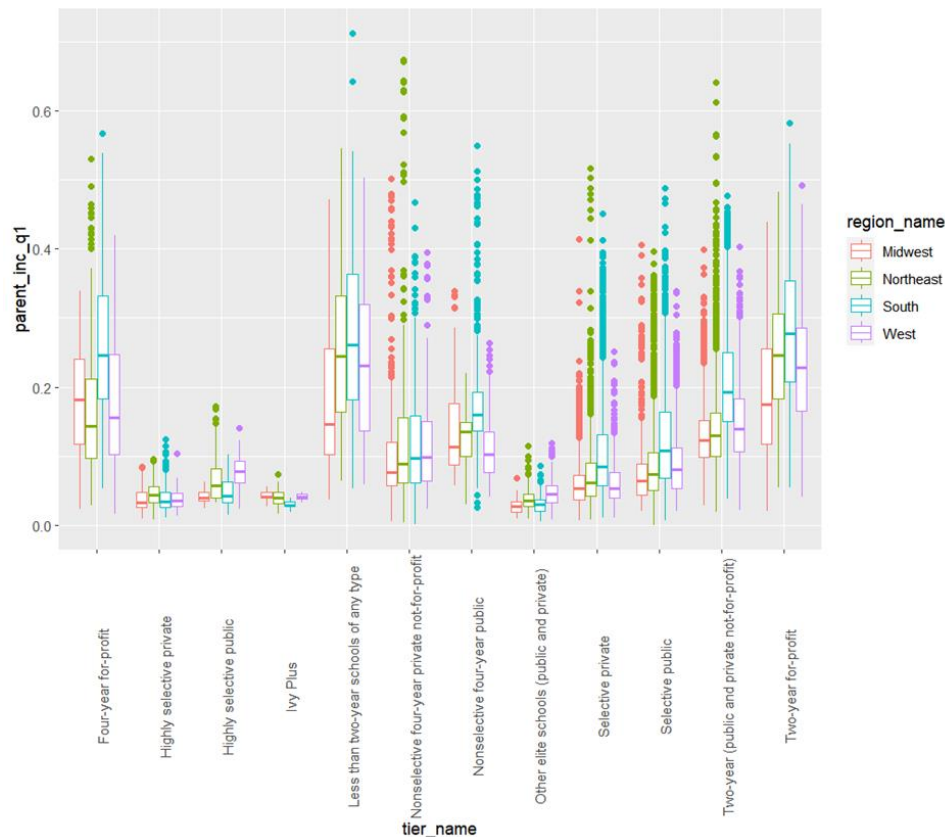


Fig 5 – Box plots between parents' Q1 income and the Tier of colleges for each region.

The Above graph has Tier name in the x-axis and Parents' Q1 income in the y-axis within each tier of the college individual boxplot representation done to indicate different regions. Since the mobility rate has a higher interaction with the parent's income in quintile 1, we generated a box plot of the parent's income in quintile 1 with respect to tier and colored by region. This graph shows the enrollment status of the children whose parents are in the Q1 income range. As you can observe from the graph above, parents in low-income distribution groups do not have high enrollment in Ivy League or highly selective colleges. Instead, colleges like selective public and selective private contribute to enrollment.

### Effect of type of colleges:

Based on the above model we understand that the spread of parent's Q1 income enrolled in specific types of colleges is higher than that of Highly selective, Ivy Plus, Other elite colleges. With an incredibly low spread and an overall median near 0, shows that IVY Plus has a very low enrolment of children whose parents are in this category. The Other elite colleges, the highly selective private, and highly selective public also have overall low enrolment, but the spread is larger in the specified order respectively. The Spread shows the difference in enrollment rate of children whose parents are in the Q1 quartile income with the same type of colleges. IQR value indicates that nearly 50 percent of the data points spread to a specific type of college. Higher IQR indicates a large difference even within 50 percent of data points. A higher Median value indicates the overall average enrolment of the specific type of college in a specific region. Overall, the median is very

high in less than two-year schools, two years for profit, and then next highest for Four-year for-profit colleges. There is also a huge spread in these colleges indicating not all colleges of this specific type have similar enrolment. The variance of points between colleges of the same type seems to be very high in selective and non-selective colleges and also in two-year public and private nonprofit colleges.

Effect of regions:

Here we could notice two types of trends, In the Ivy Plus, highly selective, and other elite colleges the Northeast region has an upper median (average) than the South region. But other than this type, the South has slightly higher in some college types and way higher in other types of colleges. So, in other types of colleges, the trend indicates that the South has a good enrollment of children of this type of parents.

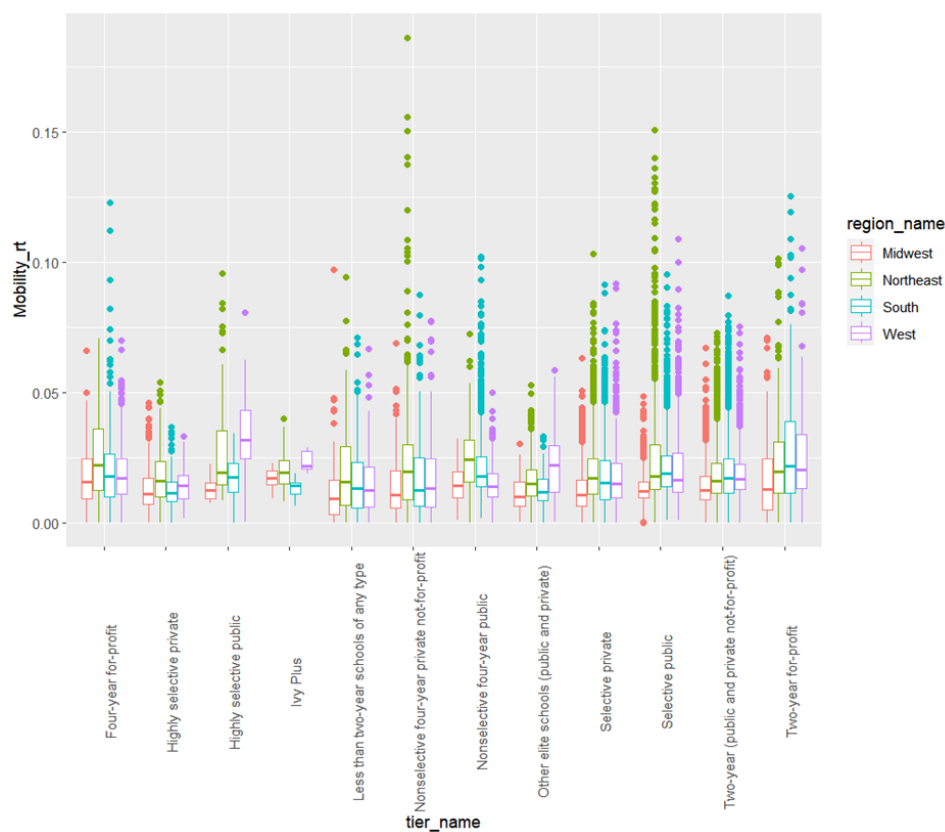


Fig 6 – Box plots between Mobility rate and Tier of College for each region

Now to understand the mobility rate distribution over different colleges and also over the different regions for a given college, we have done a boxplot representing the variation of mobility rate for each tier of the college colored by region it belongs to. So, the x-axis represents the tier of the college and y-axis represents the mobility rate and different colors represent the respective region it belongs to. Here you can observe that colleges like Ivy League, other elite schools, or highly selective public and private universities do have a similar or very less difference in the median of the distribution compared to other tiers of colleges. This indicates that these types of colleges have on average similar contributions to the mobility rate while having less enrollment as compared to

other-tier universities. This explains that if a child with a low quintile income parent enrolled in this specific tier of college, they move up in ladder with high probability compared to other tiers of colleges.

#### Effect of Tier of colleges:

The Spread of boxplots is not that much different as compared to the previous plot. Though we could see some considerable difference in the boxplot's size, the difference is lesser compared to the previous plot. Here the median value is similar across the college types. The selective and nonselective four-year colleges show huge variance, indicating that some colleges of these categories do help way better than the other colleges of the same category

#### Effect of region of colleges:

Overall, the northeast region has a higher mean or sometimes equal average compared to the South which mostly is followed by the west and finally by Midwest. There is some change in order at specific colleges but overall, the trend seems to be like this.

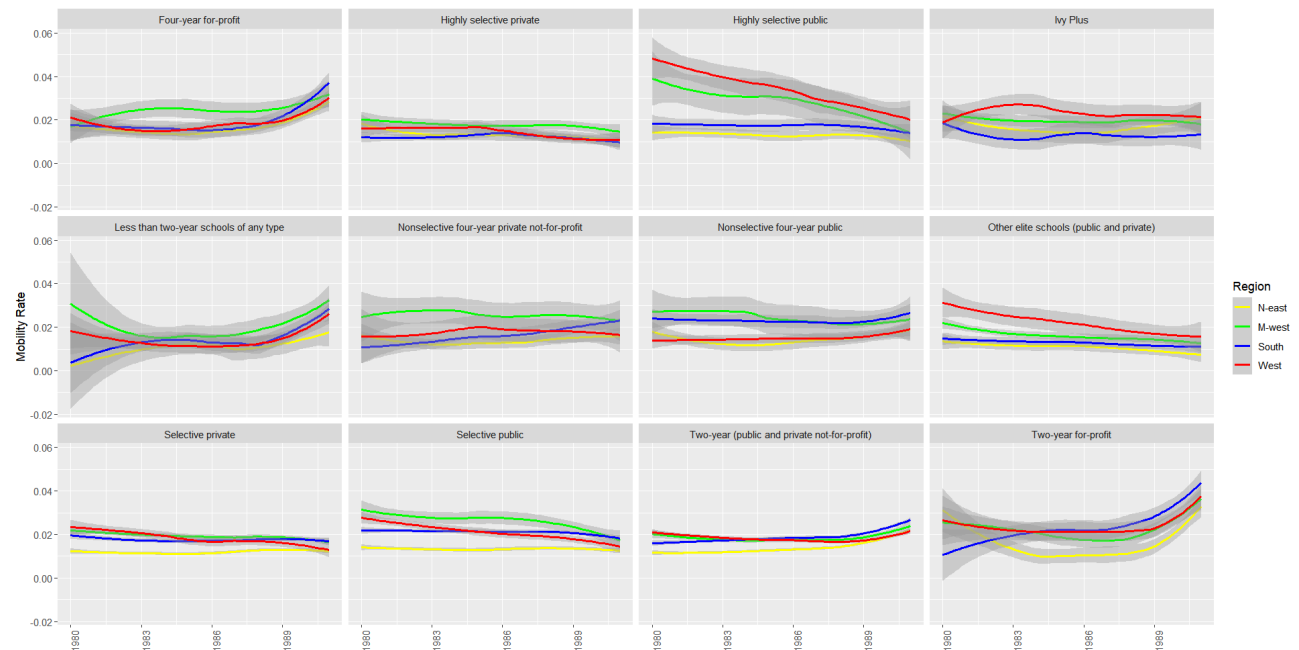


Fig 7 – Cohort vs Mobility rate faceted by college tier.

The above graph represents the relationship between cohort and mobility rate faceted by the college tier. We did a LOESS (locally estimated scatterplot smoothing) fit and from the observation, this model best represents the relationship than other models we tried.



### Effect of Cohort on Type of Colleges:

Overall, the cohort has a more or less flat curve which means there is no significant effect on the type of college except in a few cases. In the Highly selective public type, we could see a downward trend in west and mid-west region colleges indicating over the period the mobility rate has slightly reduced in these region's Highly selective colleges. Also, in two years of for-profit colleges, we could see a slight upper trend in the latest years indicating there is an increase in the mobility rate of this type of colleges in all the regions. Other than this we could not find any significant interpretations.

### **Conclusion**

The tier of college and region of colleges play important role in how college affects intergenerational social mobility. The type of college can influence the mobility rate of children from low socio-economic backgrounds, with Ivy League and other elite schools having similar mobility rates despite having lower enrollment. The region of the college also plays a role, with the northeast region having a higher mean or average mobility rate compared to the south, west, and Midwest regions. The tier of the college can also impact the relationship, with selective public and private colleges contributing more to the enrollment of children from low socio-economic backgrounds than Ivy League and other elite schools.

### **Work left to do/ limitations**

The mobility rate calculated here is the joint probability of the parent in the first quintile and their children in the fifth. This only gives a very high-level variable to work upon. Critical questions like, is the mobility rate equal across varied gender, race, and ethnicity are unanswered with this data. While we found some relationship between the region and mobility rate, there could be other confounders that can influence the mobility apart from the college data, which the data set does not have. Also, the area of education in today's world is a crucial factor in deciding a person's earning capacity, but we lack the department-wise data for each college, which can help us in exploring other influencing factors in terms of college education and intergenerational mobility.