

Analysis of Income Segregation and Intergenerational Mobility Across Colleges in the United States

**Connor Kennedy
Haoran Hu
Rebecca Netson**

Background



- Replicating the results of ‘Mobility Report Cards’ from 2017 which used data from the federal government to investigate long term outcomes of students from low-income families after attending selective Universities.
- Students from low income families that attended selective schools had excellent long term outcomes
- However, students from low income families have low access to selective schools
- Rates of upward mobility - the fraction of students who come from families in the bottom income quintile and reach the top quintile differs across colleges because access varies significantly

Results from Mobility Report Cards

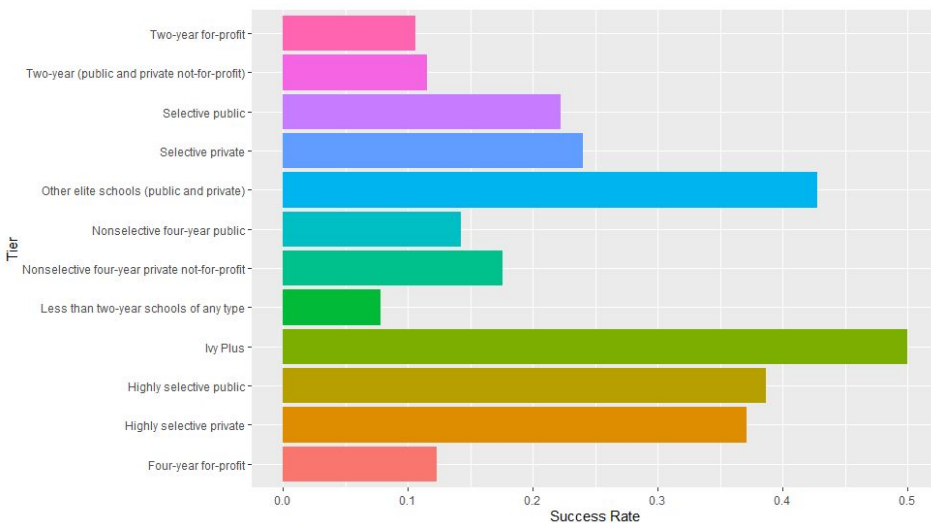
TABLE II
Key Statistics by College Tier

College Tier:	Share of Parents From:			Median Parent Income (\$)	Median Child Earnings (\$)	Within-College Rank-Rank Slope	Success Rate		Mobility Rate		Trend in Access		Num. of Colleges (80-82 cohorts)	Num. of Students (80-82 cohorts)
	Bottom 20% (%)	Bottom 60% (%)	Top 1% (%)				Top 20% (%)	Top 1% (%)	Top 20% (%)	Top 1% (%)	Bottom 20% (pp)	Bottom 60% (pp)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Ivy Plus	3.8	18.2	14.5	171,000	82,500	0.086	58.0	12.78	2.18	0.48	0.65	0.86	12	52,724
Other elite colleges	4.3	21.4	10.0	141,900	65,400	0.060	50.6	5.80	2.20	0.25	-0.46	-3.11	62	183,973
Highly selective public	5.5	29.0	2.5	107,300	53,600	0.099	40.7	2.67	2.22	0.15	-0.05	-1.71	26	393,548
Highly selective private	4.1	23.9	7.0	124,700	56,500	0.057	42.3	3.33	1.73	0.14	-0.30	-4.89	66	134,098
Selective public	8.4	39.8	1.3	87,100	41,600	0.102	23.3	0.70	1.95	0.06	-0.07	-1.89	364	1,944,082
Selective private	7.1	37.4	2.4	90,700	44,400	0.080	27.0	1.00	1.91	0.07	0.13	-2.85	446	486,852
Nonsel. 4-year public	17.0	59.5	0.6	61,200	29,800	0.085	13.5	0.19	2.30	0.03	-0.06	0.94	72	257,854
Nonsel. 4-yr. priv. non-prof.	10.7	45.2	2.0	80,500	29,000	0.079	13.6	0.42	1.45	0.04	3.43	5.54	52	55,947
2-year non-profit	14.6	55.4	0.5	66,900	29,800	0.110	12.3	0.18	1.80	0.03	1.82	3.68	604	2,021,451
Four-year for-profit	21.1	66.8	0.5	51,500	28,900	0.095	12.2	0.15	2.57	0.03	4.70	8.85	60	126,025
Two-year for-profit	20.6	67.3	0.3	51,500	31,300	0.092	13.1	0.17	2.71	0.04	5.47	9.63	37	42,313
Less than two-year colleges	20.9	65.7	N/A	53,000	18,800	0.096	7.7	0.19	1.60	0.04	2.66	8.27	14	10,032
All colleges	10.8	45.0	1.7	80,500	38,100	0.090	18.0	0.59	1.95	0.06	2.15	3.65	1,815	5,708,899

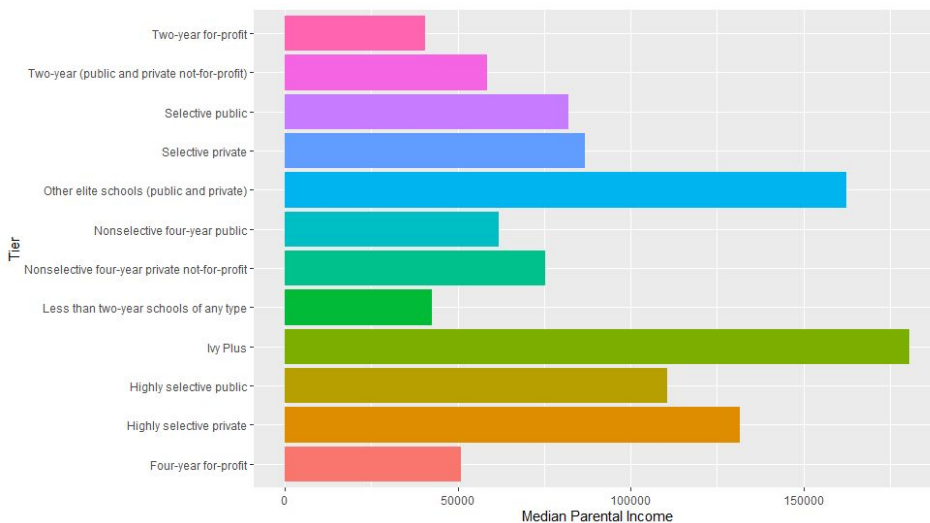
Exploratory Analysis Plots



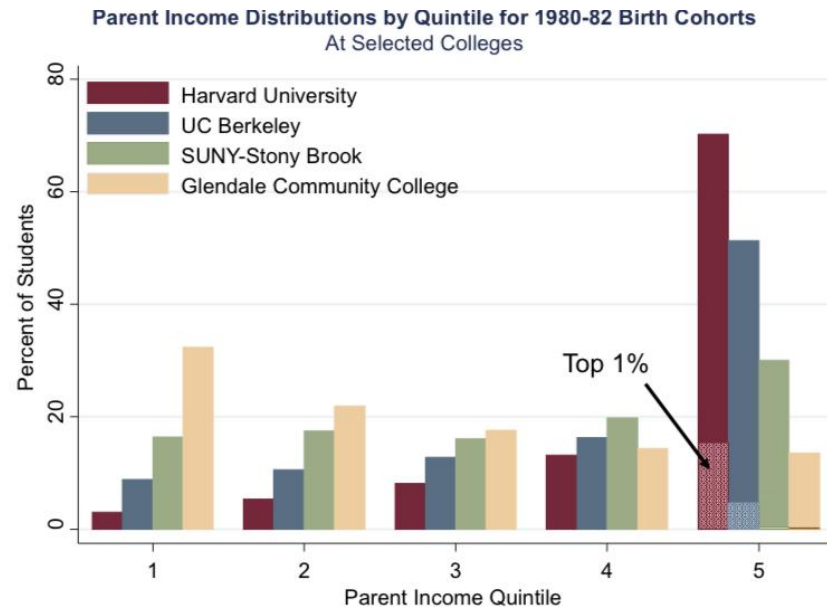
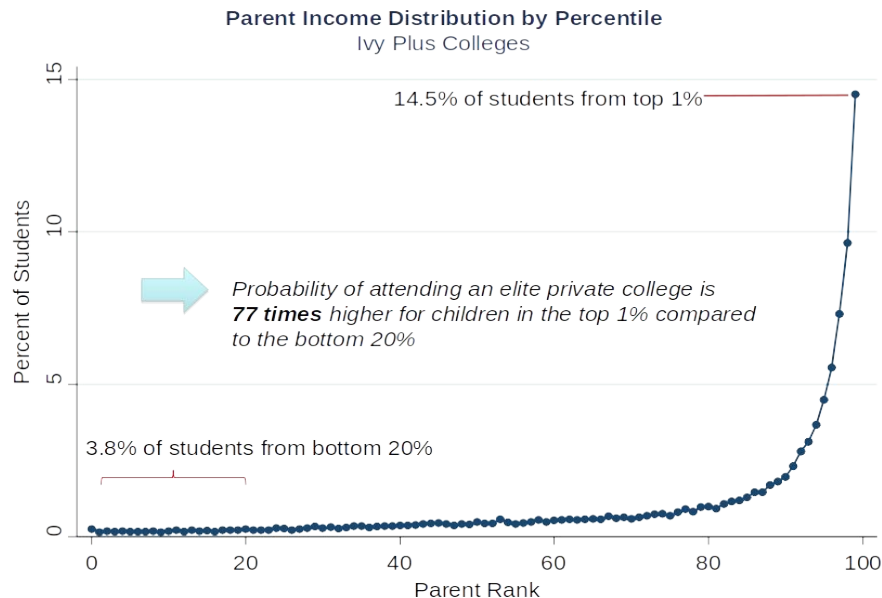
Success rate by tiers



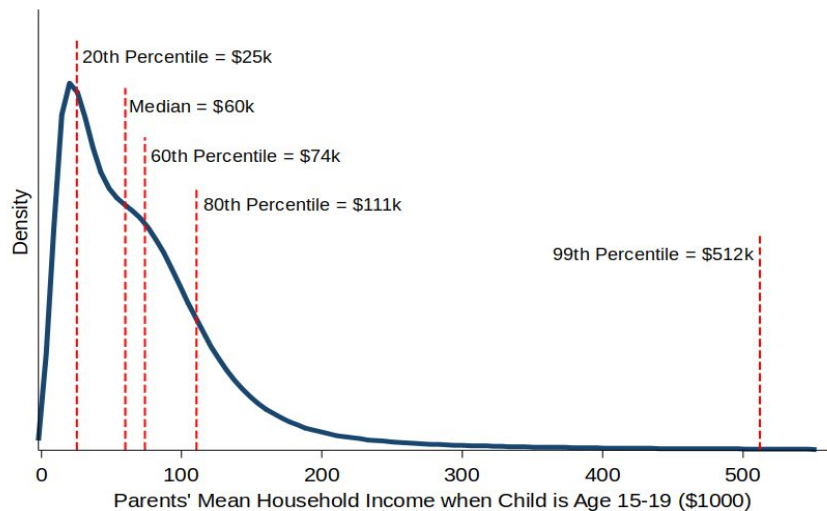
Median parents income by tiers



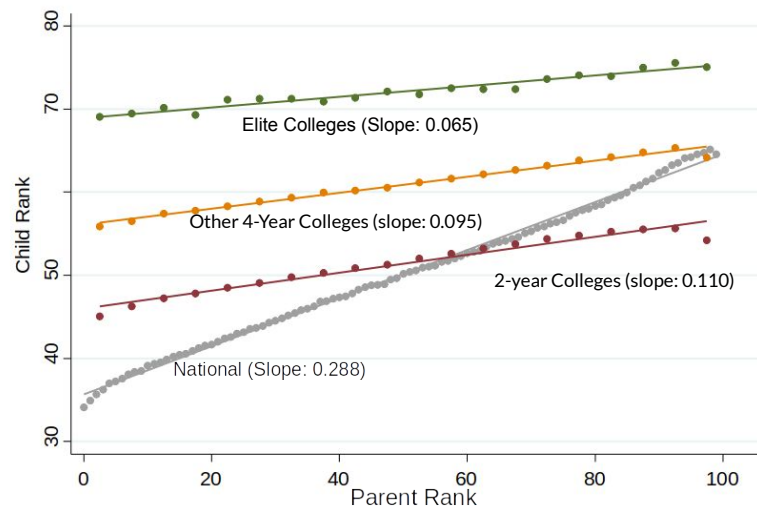
Background Plots




Background Plots



Mean Child Rank at Age 34 vs. Parent Income Rank



Goal

- 
- Assess the variables given in the code book to best explain how different factors affect a college's mobility rate.
 - Mobility rate: fraction of students who come from a family in the bottom fifth of the income distribution and end up in the top quintile of the income distribution

$$\text{mobility rate} = \text{access} * \text{success rate}$$

$$P(\text{Child in Q5 and Parent in Q1}) = P(\text{Parent in Q1}) * P(\text{Child in Q5} \mid \text{Parent in Q1})$$

Access: fraction of students who come from families in the bottom fifth

Success rate: fraction of students who come from families in the bottom fifth reached the top

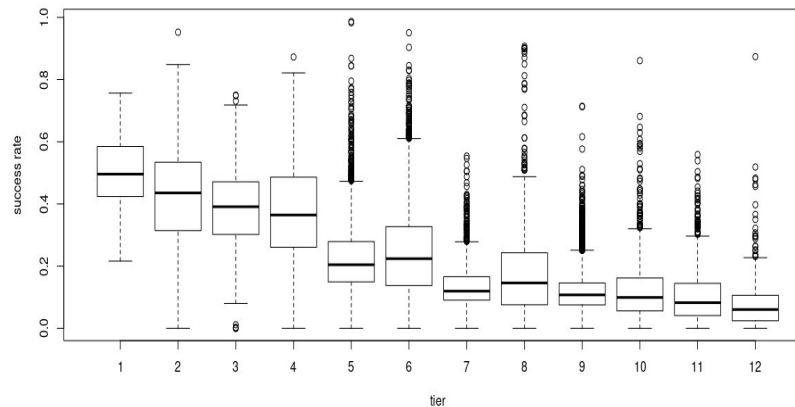
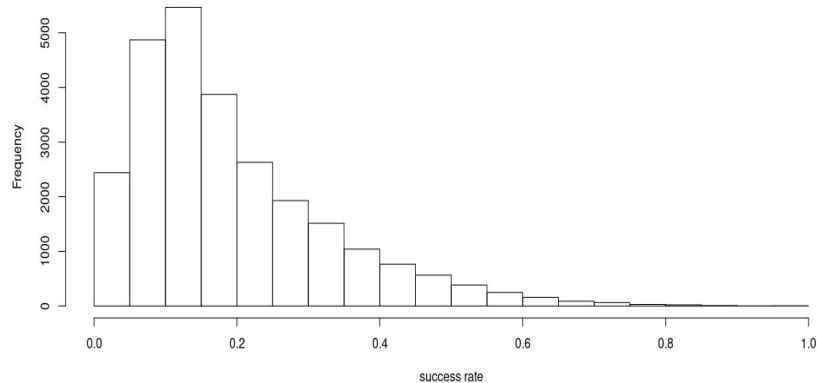


- Baseline Longitudinal Estimates of Child and Parent Income Distributions and College and Childs Cohort
- Cohort is indexed by the child's birth from 1980-1991
- Assign colleges based on college mostly attended between the ages of 19-22
- Earnings are measured between ages of 32 and 34
- Outcome variables:
 - **Joint Probability:** delineated by 'mr_kq5_pq1': parents in the bottom quintile and child in top quintile of the income distribution *Mobility Rate
 - **Conditional Probability:** delineated by 'kq5_cond_parq1 ': Probability of kid in quintile 5 conditional on parent in quintile 1. *Success Rate

Investigation



Histogram of Success Rate



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

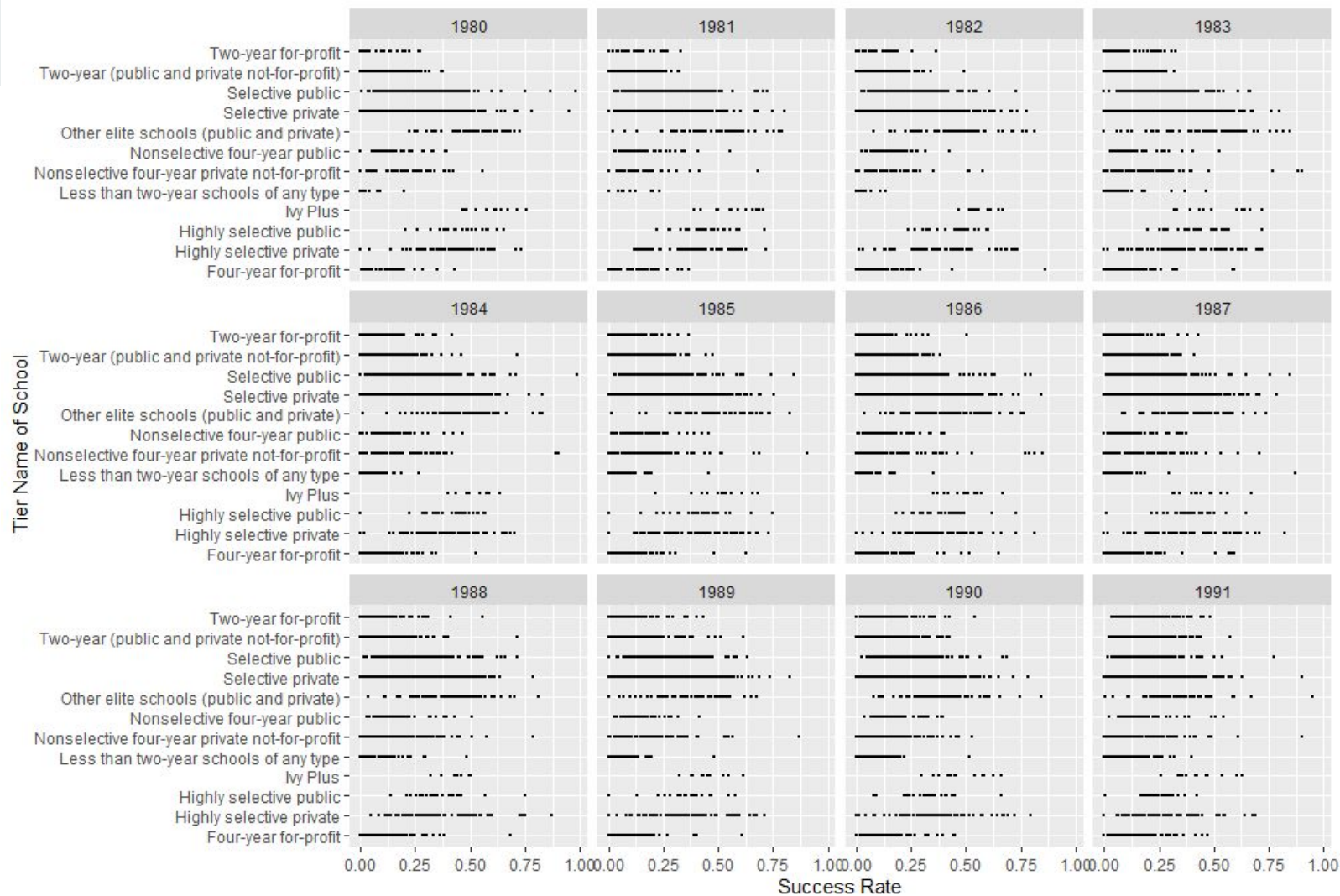
8 = Nonselective 4-year private not-for-profit

9 = Two-year (public and private not-for-profit)

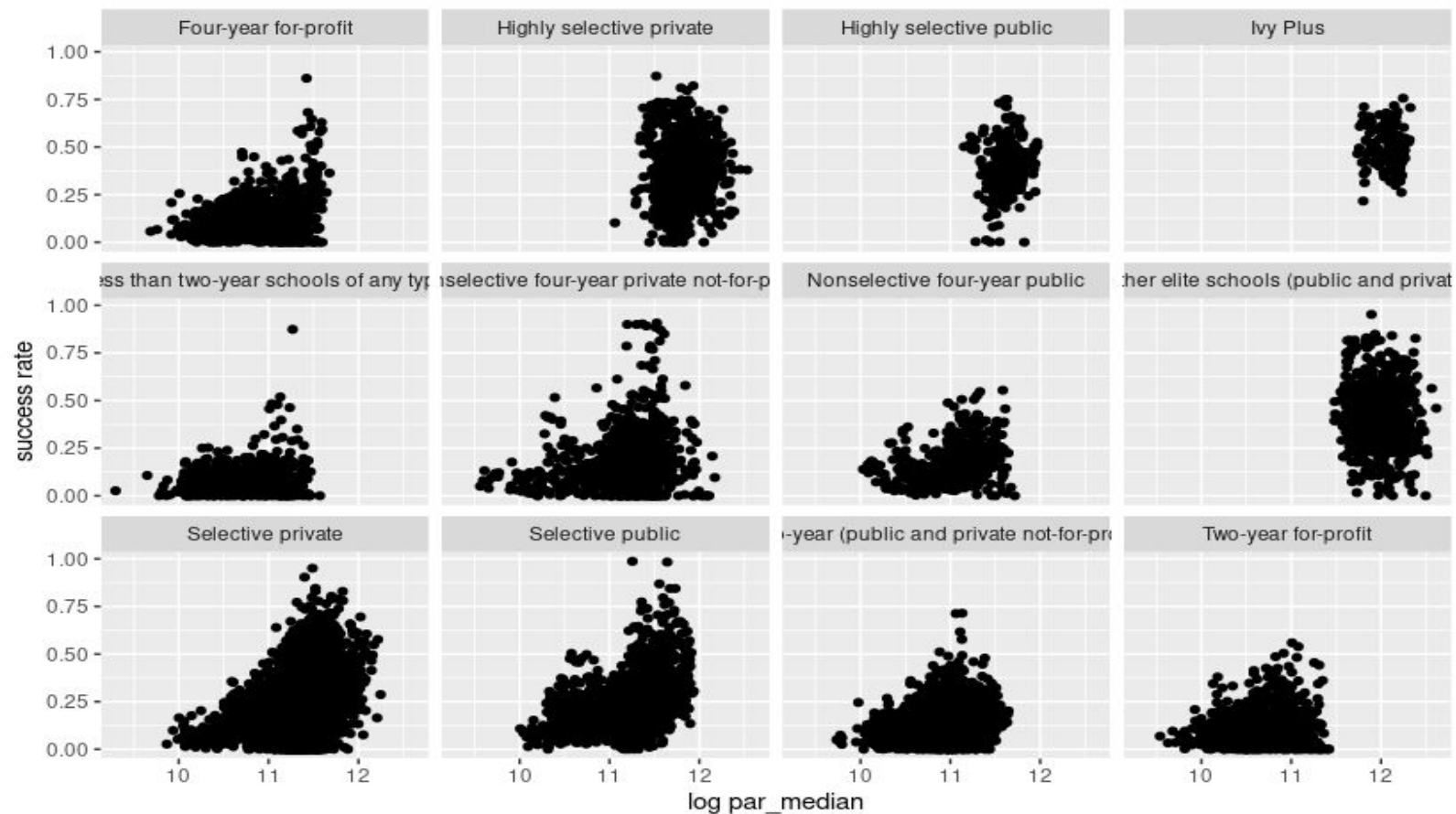
10 = Four-year for-profit

11 = Two-year for-profit

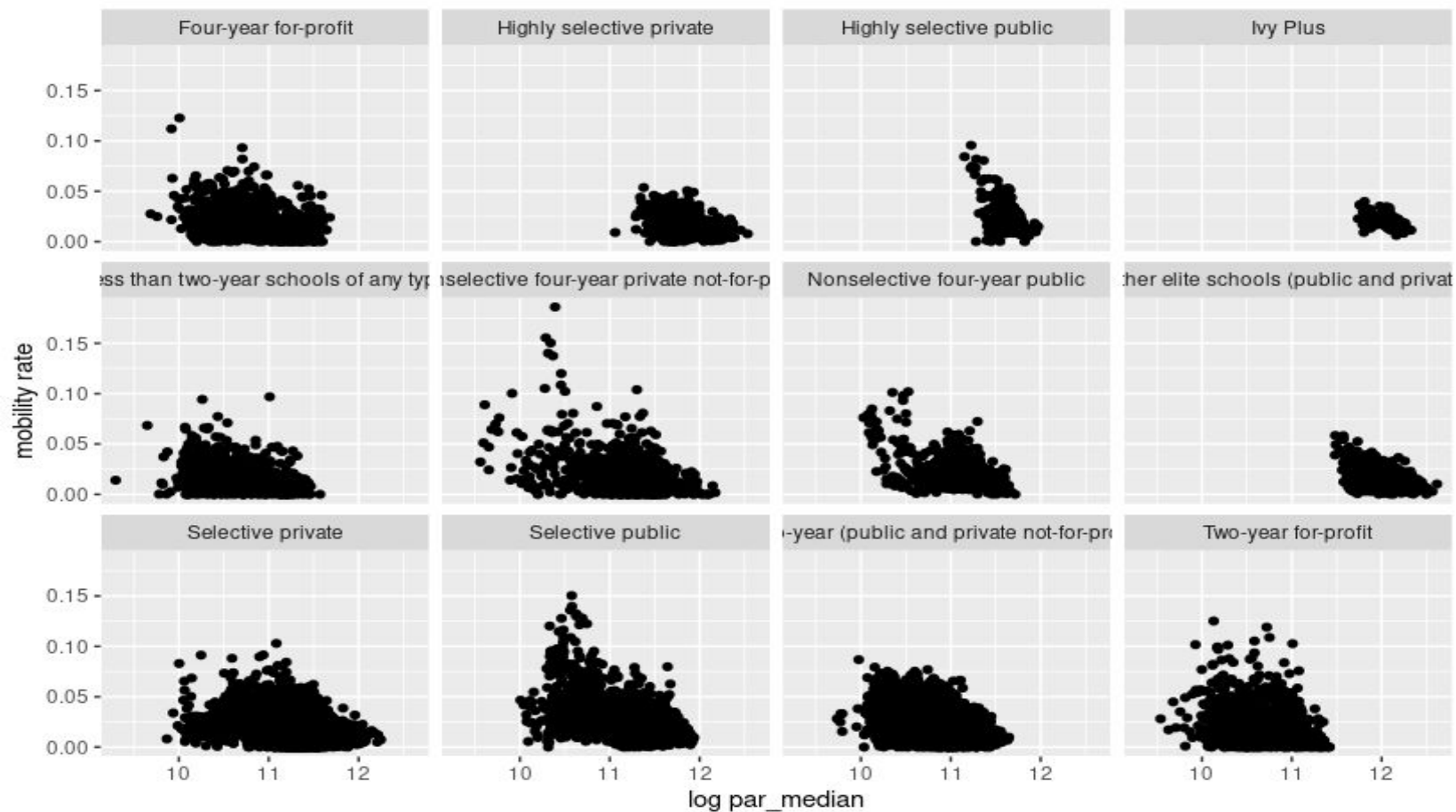
12 = Less than two year schools of any type



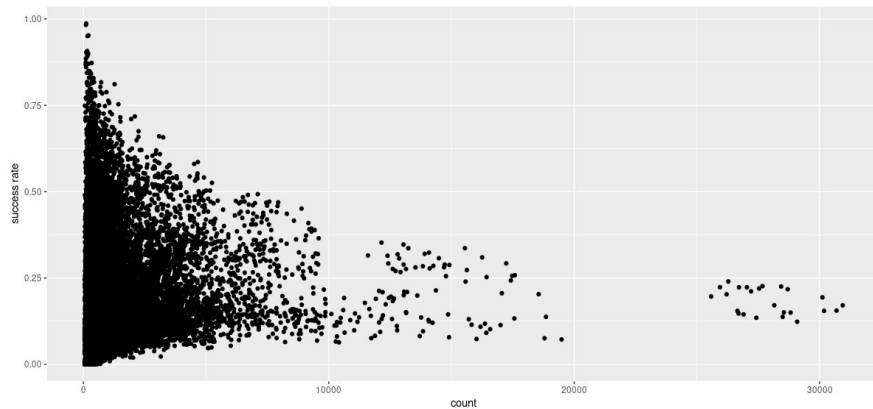
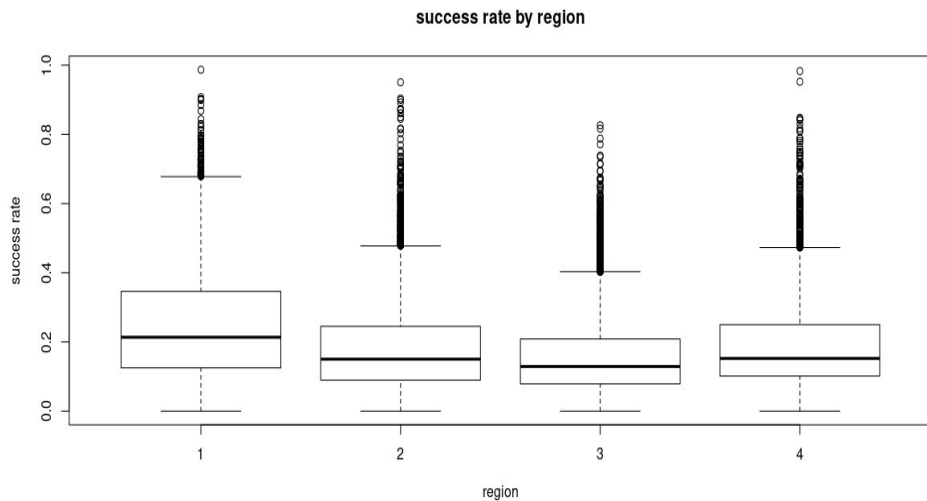
Investigation



Investigation



Investigation



1 = Northeast, 2 = Midwest, 3 = South, 4 = West

Questions




- 1.) How does the median parent income vary with each tier of school?
 - a.) Ivy Plus schools have the highest median parent income and the smallest share of parents that are from the bottom 20%. The median income of parents across all schools is \$80,000
- 2.) The effect of Ivy Plus Schools on the success rate
- 3.) The relationship between the mobility rate and success rate
 - a.) Ivy plus have a positive effect on success rate but have a negative effect on mobility rate

How we chose our model:



- Chose to model response variable(s) as Binomial
- The 'super_opeid' referred to the cohort that attended college and we excluded those who did not attend college between the ages of 19-22 or had insufficient data (3,483 observations were excluded)
- 26,097 Observations were used
- Log the parental income
- Log the count (average number of kids per cohort)
- Random effect: (1|super_opeid), college index
- Random effect: (1|cohort), birth year

Code Book and R Chunk

- 
- Tier: selectivity and type combination of school (1:Ivy Plus - 12: Less than two year schools of any type)
 - Region: 1: Northeast, 2 :Midwest, 3: South , 4: West
 - Interaction effect of parent median income and tier
 - Centered female, median parent income and count terms
 - No intercept to show effect of each tier
 - Optimizer helps reduce model fit from 30+ minutes each to under 1 minute
 - Small reduction in accuracy for estimated coefficients

```
joint_probability_model <- glmer(mr_kq5_pq1 ~ - 1 + log(par_median) * factor(tier) + female + log(count) + factor(region) + (1|super_opeid) + (1|cohort), family = "binomial", weights = count, data.new, nAGQ = 0, control=glmerCtrl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```

```
conditional_probability_model <- glmer(kq5_cond_parq1 ~ - 1 + log(par_median) * factor(tier) + female + log(count) + factor(region) + (1|super_opeid) + (1|cohort), family = "binomial", weights = pq1_count, data.new, nAGQ = 0, control=glmerCtrl(optimizer="bobyqa", optCtrl=list(maxfun=2e5)))
```



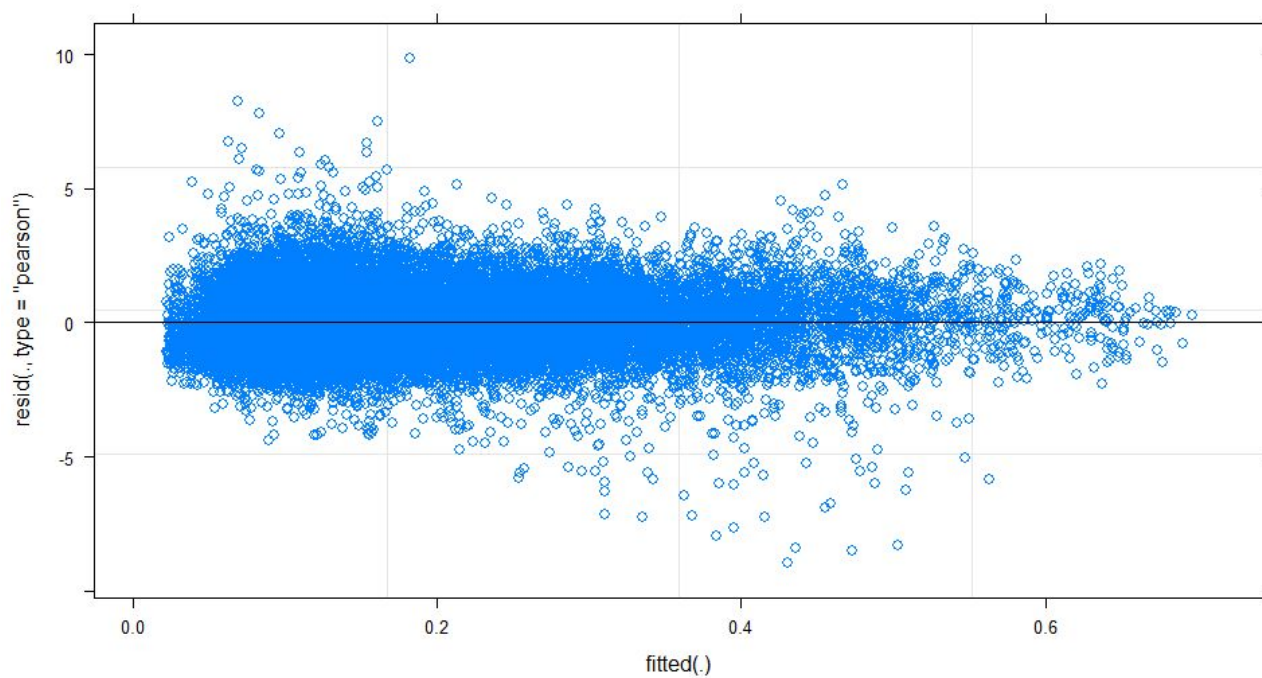
```

[5]      count1 + factor(region) + (1 | super_oid) + (1 | cohort)"
[6] "  Data: data.new"
[7] "Weights: pql_count"
[8] "Control: "
[9] "glmerControl(optimizer = \"bobyqa\", optCtrl = list(maxfun = 1e+05))"
[10] ""
[11] "      AIC      BIC    logLik deviance df.resid "
[12] "145322.6 145575.6 -72630.3 145260.6    25902 "
[13] ""
[14] "Scaled residuals: "
[15] "      Min       1Q   Median       3Q      Max "
[16] "-8.9337 -0.7851 -0.0639  0.7087  9.8356 "
[17] ""
[18] "Random effects:"
[19] " Groups      Name      Variance Std.Dev."
[20] "  super_oid (Intercept) 0.175621 0.41907 "
[21] "  cohort      (Intercept) 0.003689 0.06074 "
[22] "Number of obs: 25933, groups: super_oid, 2463; cohort, 12"
[23] ""
[24] "Fixed effects:"
[25] "      Estimate Std. Error z value Pr(>|z|) "
[26] "factor(tier)1 -0.064405  0.408724  -0.158  0.87479 "
[27] "factor(tier)2  0.378772  0.121814   3.109  0.00187 ** "
[28] "factor(tier)3 -0.459973  0.110098  -4.178  2.94e-05 ***"
[29] "factor(tier)4 -0.435201  0.108797  -4.000  6.33e-05 ***"
[30] "factor(tier)5 -1.205265  0.033896 -35.558 < 2e-16 ***"
[31] "factor(tier)6 -1.119069  0.031726 -35.273 < 2e-16 ***"
[32] "factor(tier)7 -1.752207  0.057644 -30.397 < 2e-16 ***"
[33] "factor(tier)8 -1.474342  0.053797 -27.406 < 2e-16 ***"
[34] "factor(tier)9 -2.015434  0.030731 -65.583 < 2e-16 ***"
[35] "factor(tier)10 -1.892487  0.055002 -34.408 < 2e-16 ***"
[36] "factor(tier)11 -2.379271  0.066764 -35.637 < 2e-16 ***"
[37] "factor(tier)12 -2.200077  0.084766 -25.955 < 2e-16 ***"
[38] "par_median1  0.003010  0.181237   0.017  0.98675 "
[39] "female1 -0.168703  0.007197 -23.441 < 2e-16 ***"
[40] "count1  0.023585  0.008232   2.865  0.00417 ** "
[41] "factor(region)2 -0.269214  0.026921 -10.000 < 2e-16 ***"
[42] "factor(region)3 -0.354857  0.024735 -14.346 < 2e-16 ***"
[43] "factor(region)4 -0.067329  0.029177  -2.308  0.02102 * "
[44] "factor(tier)2:par_median1 -0.295346  0.189122  -1.562  0.11837 "
[45] "factor(tier)3:par_median1  0.056356  0.192345   0.293  0.76953 "
[46] "factor(tier)4:par_median1  0.049123  0.190979   0.257  0.79701 "
[47] "factor(tier)5:par_median1  0.300341  0.182153   1.649  0.09918 . "
[48] "factor(tier)6:par_median1  0.369073  0.182394   2.023  0.04302 * "
[49] "factor(tier)7:par_median1 -0.190195  0.184442  -1.031  0.30245 "
[50] "factor(tier)8:par_median1  0.117594  0.185358   0.634  0.52581 "
[51] "factor(tier)9:par_median1 -0.199411  0.181701  -1.097  0.27244 "
[52] "factor(tier)10:par_median1 -0.007657  0.183001  -0.042  0.96662 "
[53] "factor(tier)11:par_median1 -0.281693  0.184352  -1.528  0.12651 "
[54] "factor(tier)12:par_median1 -0.004830  0.187167  -0.026  0.97941 "
[55] "----"
[56] "Signif. codes:  0. '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

```

- Insignificant fixed effect for Tier 1
- Centered covariates help interpretability
 - Tier effect becomes more meaningful
- Small variance for random cohort effect
- Tier * income interaction significant for ANOVA
 - P-values aren't perfect

Conditional Probability: Success Rate

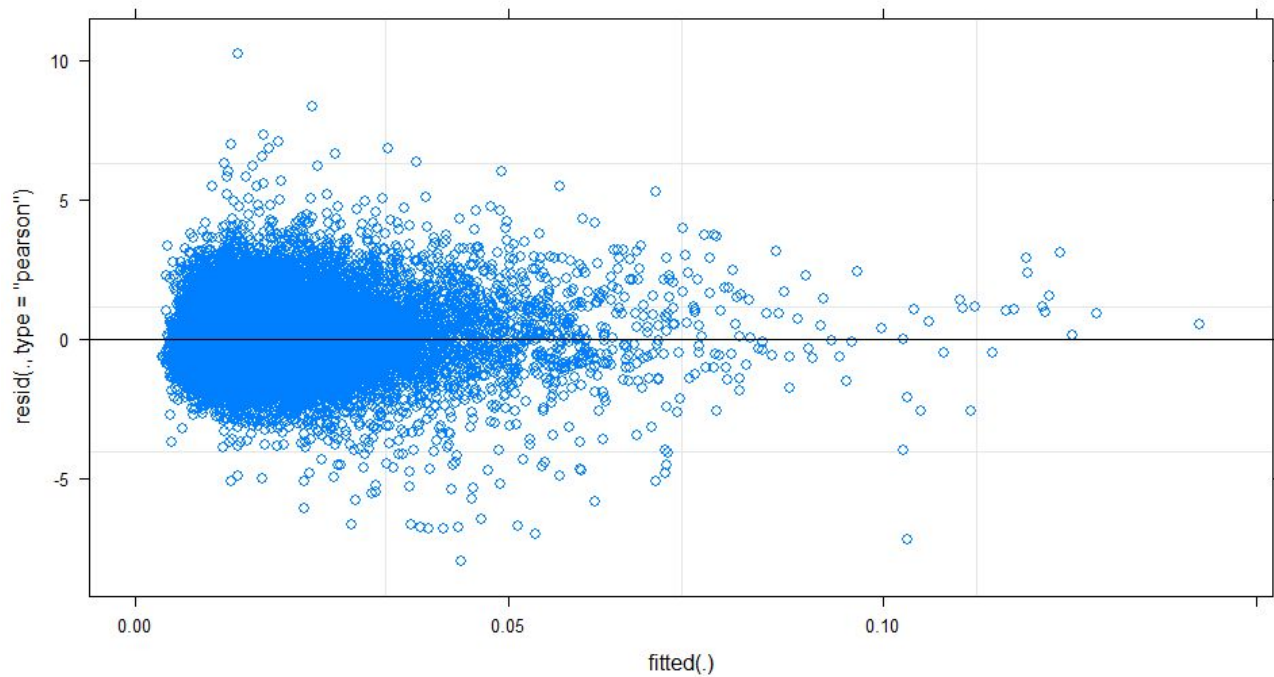


Joint Probability: Mobility Rate

```
[1] "Generalized linear mixed model fit by maximum likelihood (Adaptive Gauss-Hermite Quadrature,"
[2] "  nAGQ = 0) [glmerMod]"
[3] " Family: binomial ( logit )"
[4] "Formula: mr_kq5_pq1 ~ -1 + factor(tier) * par_mediana1 + female1 + count1 + "
[5] "  factor(region) + (1 | super_opeid) + (1 | cohort)"
[6] "  Data: data.new"
[7] "Weights: count"
[8] "Control: glmerControl(optimizer = \"bobyqa\", optCtrl = list(maxfun = 1e+05))"
[9] ""
[10] "      AIC      BIC    logLik deviance df.resid "
[11] "149253.1 149506.1 -74595.5 149191.1    25902 "
[12] ""
[13] "Scaled residuals: "
[14] "      Min       1Q   Median       3Q      Max "
[15] "-7.9487 -0.7908 -0.0993  0.6706 10.2244 "
[16] ""
[17] "Random effects:"
[18] " Groups      Name      Variance Std.Dev."
[19] " super_opeid (Intercept) 0.135341 0.36789 "
[20] " cohort      (Intercept) 0.005237 0.07237 "
[21] "Number of obs: 25933, groups: super_opeid, 2463; cohort, 12"
[22] ""
[23] "Fixed effects:"
[24] "      Estimate Std. Error z value Pr(>|z|) "
[25] "factor(tier)1 -2.770664    0.304076  -9.112 < 2e-16 ***"
[26] "factor(tier)2 -3.129037    0.096834 -32.313 < 2e-16 ***"
[27] "factor(tier)3 -3.159163    0.093577 -33.760 < 2e-16 ***"
[28] "factor(tier)4 -3.664046    0.088845 -41.241 < 2e-16 ***"
[29] "factor(tier)5 -3.752579    0.032973 -113.809 < 2e-16 ***"
[30] "factor(tier)6 -3.802330    0.031259 -121.640 < 2e-16 ***"
[31] "factor(tier)7 -4.028366    0.052789 -76.311 < 2e-16 ***"
[32] "factor(tier)8 -3.967785    0.049520 -80.125 < 2e-16 ***"
[33] "factor(tier)9 -4.228119    0.030548 -138.409 < 2e-16 ***"
[34] "factor(tier)10 -4.076030    0.050809 -80.223 < 2e-16 ***"
[35] "factor(tier)11 -4.400929    0.061587 -71.459 < 2e-16 ***"
[36] "factor(tier)12 -4.365627    0.077839 -56.086 < 2e-16 ***"
[37] "par_mediana1 -0.549649    0.131905  -4.167 3.09e-05 ***"
[38] "female1 -0.120285    0.006412  -18.750 < 2e-16 ***"
```

- Significant negative fixed effect for Tier 1
`fun = 1e+05)) "`
 - Change is consistent with paper
- Centered covariates help interpretability
- Small variance for random cohort effect
- Tier * income interaction significant for ANOVA
- Consistent covariates for each model

Joint Probability: Mobility Rate



Comparison to 'Mobility Rate' paper



- Ivy plus schools have low access to students from low income families
- High female percentage leads to low success rate
- Schools in Northeast region have better success rate than other regions

More on Intergenerational Mobility



- Racial disparities in income mobility persist, especially for men. Black and white boys who grow up on the same city neighborhood and attend the same school still have very different outcomes in adulthood.
- What extent are children's opportunities for upward economic mobility shaped by the neighborhoods in which they grow up? The area in which a child grows up has significant causal effects on their prospects for upward mobility.
- <https://opportunityinsights.org/>
- <https://www.nytimes.com/interactive/2018/03/27/upshot/make-your-own-mobility-animation.html>

References



Chetty, R., Friedman, J., Turner, E. and Yagan, D., 2020. *Income Segregation and Intergenerational Mobility Across Colleges in the United States*. [online] Opportunityinsights.org. Available at: <https://opportunityinsights.org/wp-content/uploads/2020/02/coll_mrc_qje_paper.pdf>.

Badger, Emily, et al. “Income Mobility Charts for Girls, Asian-Americans and Other Groups. Or Make Your Own.” *The Upshot*, The New York Times, 27 March 2018, <https://www.nytimes.com/interactive/2018/03/27/upshot/make-your-own-mobility-animation.html>. Accessed April 2021.