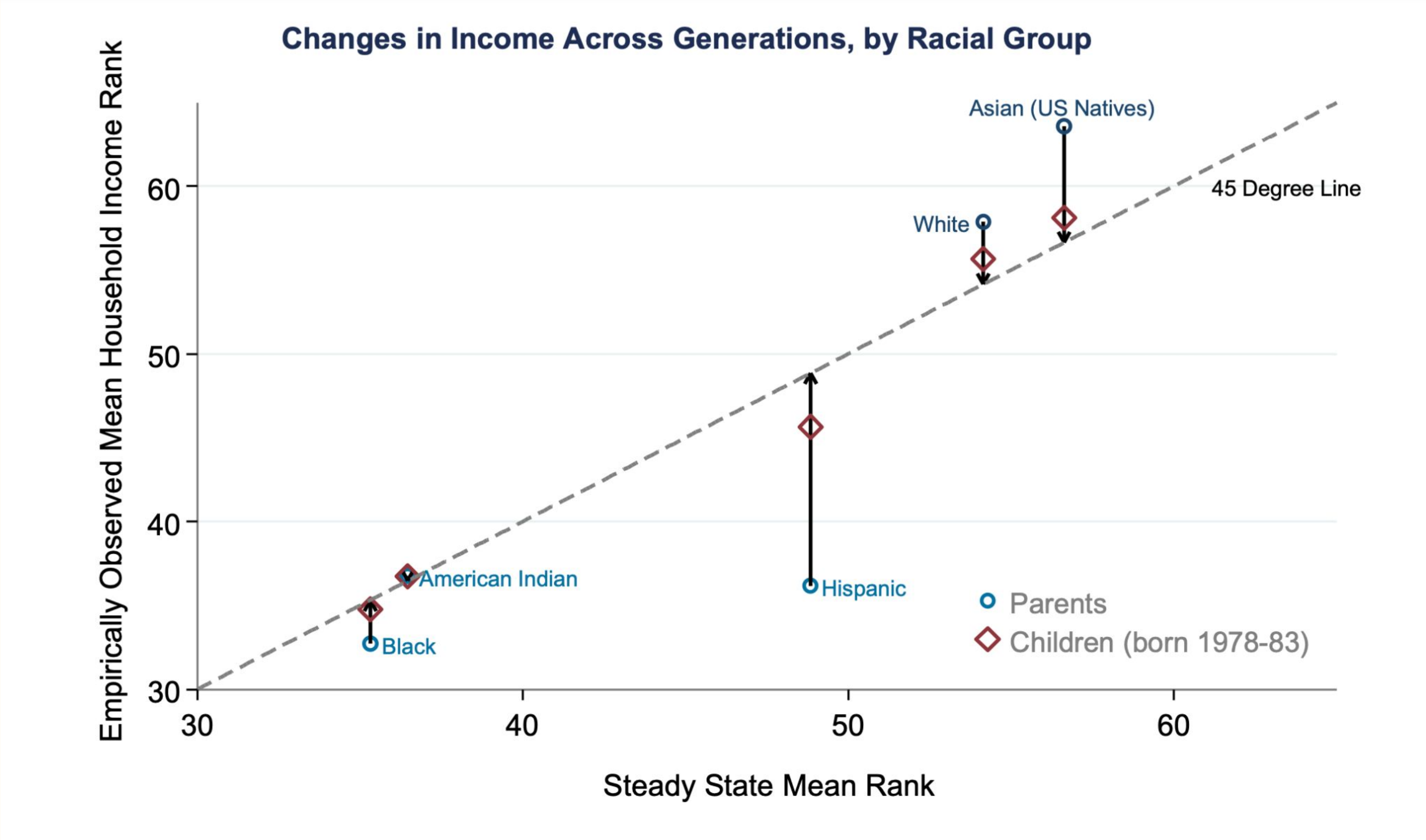
Race and Economic Opportunity in the United States: An Intergenerational Perspective

Fatima Gordon, Hannah Hansraj, Michelle Anne Austria, Michelle Maldonado, Qili Zhang, Sofia Belhouari

Project Description

Abstract

Objective: Analyze economic data to understand and address racial disparities in income mobility in the U.S.
Methods: Applied various regression techniques (e.g., Linear Regression, Lasso, Ridge, ElasticNet, Decision Tree, Random Forest, Gradient Boosting) using tools like pandas, matplotlib, and scikit-learn.
Data Exploration: Utilized libraries such as NumPy, pandas, and seaborn for data manipulation and visualization.
Key Findings: Significant income mobility disparities, especially among black men, persist beyond differences in parental income, education, or marital status. These gaps exist even within the same neighborhoods.
Implications: Highlights the necessity for targeted interventions addressing both systemic and neighborhood-specific factors to enhance economic mobility and reduce racial income disparities.
Significance: Emphasizes the importance of selecting appropriate regression models to analyze complex socio-economic issues accurately.


Changes in Income Across Generations, by Racial Group
Empirically Observed Mean Household Income Rank
Steady State Mean Rank
30
40
50
60
Asian (US Natives)
White
45 Degree Line
American Indian
Hispanic
Black
Parents
Children (born 1978-83)

Project Background

Economic mobility in the United States has long been a subject of interest among economists and policymakers. Traditional views often emphasize the role of individual effort and education in achieving economic success. However, the reality is more complex, especially when considering racial disparities. This study provides an **in-depth look at how race influences economic opportunity from an intergenerational perspective**, shedding light on persistent inequalities.

The paper highlights significant differences in economic mobility between black and white Americans.

- Black children born to parents in the bottom quintile of the income distribution are **more likely** to remain in the bottom quintile as adults compared to their white counterparts.
- This trend persists even when controlling for factors such as **parental income, education, and family structure**.

Neighborhoods play a crucial role in shaping economic outcomes.

- Black children who grow up in low-poverty, high-opportunity areas have substantially better economic outcomes than those who grow up in high-poverty, low-opportunity areas.
- However, black children are less likely than white children to grow up in such neighborhoods, even when their parents have similar incomes.

Project Background

The study also reveals that black men, in particular, face substantial barriers to economic mobility. Black women have outcomes more similar to white women, suggesting that both race and gender interact to influence economic opportunity.

Quantitative analysis is essential to understand the extent and nuances of these disparities.

By leveraging large datasets and advanced statistical techniques, the authors can control for various factors and isolate the effects of race on economic outcomes. This allows for a **more accurate and comprehensive understanding of the systemic issues at play**.

Understanding these dynamics can inform policies aimed at reducing economic disparities.

For example, targeted interventions in high-poverty neighborhoods, efforts to improve access to quality education, and policies addressing systemic discrimination in the labor market could all be informed by the findings of this research.

Ultimately, such analyses can help to **promote a more equitable society by ensuring that economic opportunity is not determined by race**.

Project Objective

The primary objective of this project is to investigate the impact of father presence in single-parent and two-parent families on the average earnings of their children, with a specific focus on variations across different racial groups within the low-income bracket. This study aims to provide a comprehensive analysis of how paternal involvement influences economic outcomes, thereby contributing valuable insights to policy-making and social interventions aimed at reducing income disparities and promoting upward mobility among disadvantaged populations.

Data Description & Cleaning

Table 1

Table 1 covers outcomes by **parent income percentile, child's gender and race for children born** in years 1978-83. There is one row for each parent income percentile. For some only it only focus black and white children. Race-specific statistics are reported by gender (male or female) or for all children (pooled).

```
df_1 = df_1.drop_duplicates()

na_cols = ["kid_college_black_female", "kid_jail_black_female", "kid_no_hs_black_female", "kid_pos_hours_black

# Fill NaN values with the mean for the selected columns
df_1[na_cols] = df_1[na_cols].fillna(df_1[na_cols].mean())

# Function to remove outliers using IQR
def remove_outliers(df_1):
    for column in df_1.columns:
        if df_1[column].dtype in ['int64', 'float64']:  # Adjust this to include/exclude certain types
            Q1 = df_1[column].quantile(0.25)
            Q3 = df_1[column].quantile(0.75)
            IQR = Q3 - Q1
            lower_bound = Q1 - 1.5 * IQR
            upper_bound = Q3 + 1.5 * IQR

            # You can choose to remove or replace the outliers
            df_1[column] = np.where((df_1[column] < lower_bound) | (df_1[column] > upper_bound), np.nan, df_1[
    return df_1

# Apply the function
df_1_clean = remove_outliers(df_1)
df_1_clean
```

count_hisp_pooled f.	count_white_pooled	density_aian_pool...	density_asian_poo...	density_black_poo...	density_hisp_pool...
5600.0 - 44900.0	69000.0 - 186600.0	0.1512 - 1.996	0.8035 - 1.328	0.0654 - 2.161	0.2137 - 1.717
34900	69000	1.996	1.328	2.161	1.335
39900	69000	1.814	1.3	2.016	1.527
40400	72000	1.755	1.3	1.961	1.546
41900	73500	1.694	1.3	1.925	1.603
42400	75000	1.755	1.3	1.889	1.623
43400	75000	1.755	1.3	1.871	1.661

Table 1

The granularity of the data seems to **cover the different results between different minorities**. For instance, categories such as percentage of children with college attendance by race and gender. It seems to be detailed but there are a few nan's located in dataset. Due to the size of the data frame being 100 rows and 69 columns. I would consider it to be granular since it takes into account different races and genders

Variable Name	Description
par_pctile	Parent household income rank.
count_pooled	Number of children born to parents at a given income percentile (rounded to the next 100s).
count_[race]_pooled	Number of children born to parents at a given income percentile by race.
density_[race]_pooled	Percentage of children born to parents at a given income percentile by race. Due to rounding, the total sum of percentages by race does not necessarily add up to exactly 100%.
kfr_pooled	Mean child family (household) income rank for all children.
kfr_[race]_[gender]	Mean child family (household) income rank by race and by gender.
kid_college_[race]_[gender]	Percentage of children with college attendance by race and gender.
kid_hours_[race]_[gender]	Mean number of weekly working hours over the past year by race and gender.
kid_jail_[race]_[gender	Percentage of children incarcerated by race and gender.

Table 2

This table reports quintile-quintile intergenerational transition matrices and marginal distributions of child and parent income by race and gender for **all children in our primary analysis sample**.

We measure parent income at the household level; we present statistics measuring children's income at both the individual and household level.

Variable Name	Description
kid_race	A string categorical variable describing the race of a given child, taking on the following values: White, Black, Hispanic, Asian, AIAN
gender	A string categorical variable describing the gender of a given child, taking on the following values: F, M, P
count	A numeric variable reporting the total count of children with native mothers in a given race-gender cell.
kir_q[quintile]	The fraction of children with individual income in a given quintile.
kfr_q[quintile]	The fraction of children with household income in a given quintile.
par_q[quintile]	The fraction of children with parent household income in a given quintile of the national distribution.
kir_q[quintile i]_cond_par_q[quintile j]	The fraction of children with parents in quintile j who are in quintile i of the national distribution of individual income.
kfr_q[quintile i]_cond_par_q[quintile j]	The fraction of children with parents in quintile j who are in quintile i of the national distribution of household income.

Table 3

This table reports quintile-quintile intergenerational transition matrices and marginal distributions of child and parent income by race and gender for **children with mothers born in the U.S**.

We measure parent income at the household level; we present statistics measuring children's income at both the individual and household level.

Variable Name	Description
kid_race	A string categorical variable describing the race of a given child, taking on the following values: White, Black, Hispanic, Asian, AIAN
gender	A string categorical variable describing the gender of a given child, taking on the following values: F, M, P
count	A numeric variable reporting the total count of children with native mothers in a given race-gender cell.
kir_q[quintile]	The fraction of children with individual income in a given quintile.
kfr_q[quintile]	The fraction of children with household income in a given quintile.
par_q[quintile]	The fraction of children with parent household income in a given quintile of the national distribution.
kir_q[quintile i]_ cond_par_q[quintile j]	The fraction of children with parents in quintile j who are in quintile i of the national distribution of individual income.
kfr_q[quintile i]_ cond_par_q[quintile j]	The fraction of children with parents in quintile j who are in quintile i of the national distribution of household income.

Table 4:

```
#Cleaning the data

#Drops the nulls
df_4.dropna()

#Columns cz and cz_name are unique and don't have any actual correlation
cleaned_df4 = df_4.drop(columns=['cz', 'cz_name'])

cleaned_df4
```

Visualize

	cz_pop2000 int64 1193 - 16393360	cz_pop_black2000 i... 0 - 2311940	cz_pop_hisp2000 i... 4 - 6602908	cz_pop_white2000 i... 521 - 6399667	kfr_pooled_p25 flo... 28.98 - 62.37	kfr_black_pooled_... 16.45 - 56.56	kfr_white_pooled_... 37.56 - 64.11
0	576081	12006	5349	551309	36.71	30.91	37.56
1	227816	4506	5351	214507	36.71	32.56	37.56
2	66708	977	411	64439	37.6	35.01	38.38
3	727600	41011	8492	661477	36.71	29.64	38.38
4	493180	85582	28599	369663	35.83	30.91	39.74
5	92753	20759	2723	68090	36.91	33.01	39.74
6	1055133	229554	42842	754019	35.4	31.01	39.74
7	90016	2874	2852	83417	37.74	31.52	39.06
8	64676	1873	2043	60200	38.7	29.9	40.02
9	354533	73565	8923	264424	35.25	31.46	38.68

741 rows, 19 cols, showing 10 rows/page   Page 1 of 75

Variable Name	Description
cz_pop2000	CZ population in 2000 Census
cz_pop_[race]2000	CZ population of race/ethnicity [race] in 2000 decennial Census where race is either of: • black (non-Hispanic black) • Hispanic • white (non-Hispanic white)
kfr_pooled_p25	Mean kid household income rank in the CZ for a child with parents at the 25th percentile rank of the national income distribution (pooled across race and gender)
kfr_[race]_pooled_p[pctile]	Mean kid household income rank in the CZ for children (pooled across gender) with • [race] black or white • born to parents in the [pctile] 25th or 75th percentile of the national distribution for parents of children in the same cohort
kir_[race]_[gender]_p[pctile]	Mean kid individual income rank in the CZ for a child with • [race] black, white, or Hispanic • [gender] male, female, or pooled • born to parents in the [pctile] 25th or 75th percentile of the national distribution for parents of children in the same cohort

Table 5

```
df_5.drop(df_5.tail(1).index,inplace=True)
df_5
df_5.drop(df_5.head(1).index, inplace=True)
df_5
```

	percentile int64 2 - 99	kid_hh_income int64 0 - 298300	kid_indiv_income i... -477 - 189300	parent_hh_income i... 3143 - 373300
1	2	0	-477	3143
2	3	0	0	4714
3	4	0	0	6106
4	5	0	0	7376
5	6	0	0	8553
6	7	0	0	9660
7	8	0	0	10720
8	9	0	0	11740
9	10	0	0	12730
10	11	72	0	13690

Variable Name	Description
Percentile	Child or parent income percentile rank
kid_hh_income	Mean child family (household) income in 2015 dollars
kid_indiv_income	Mean child individual income in 2015 dollars
parent_hh_income	Mean parent household income in 2015 dollars

Table 6a and 6b

```
df_6a.head()
```

Visualize

	country object	n_kfr_P int64	kfr_P_p25 float64	kfr_P_p25_se float...	kfr_P_p75 float64	kfr_P_p75_se float...	n_kir_F int64	k
0	ARGENTINA	1900	0.47780001	0.01094	0.6232	0.0070489999	900	
1	AUSTRIA	2000	0.4747	0.01338	0.62199998	0.0063240002	950	
2	BRAZIL	1300	0.52179998	0.0155	0.61320001	0.0087369997	600	
3	CAMBODIA	1300	0.5151	0.01016	0.6153	0.01152	650	
4	CANADA	18500	0.4639	0.0041809999	0.60640001	0.0021250001	9000	

Variable	Meaning
country	The country of origin of the second-generation immigrant children.
n_kfr_P	Number of observations for parental income.
kfr_P_(p25...)_M/F	Income rank at the 25th (or other) percentile for parental income. Some gendered.
age_inXXXX_mom_F/M	Age of the mother in XXXX (females/males).
us_yrs_before_M/F_P	Years in the U.S. before a specific milestone for the mother/father.
kfr_P_p25_se	Standard error for the 25th percentile income rank for parental income.

This data is Parametric and Non-Parametric Estimates of Income Ranks for Second Generation Immigrant Children by Parent Income, Country of Origin, and Gender and is sourced from a paper titled "Race and Economic Opportunity in the United States: An Intergenerational Perspective." They based this dataset on various sources including the Census, tax returns, and American Community Surveys from 2005 to 2015. (Appendix A: Data Construction)

The data is highly granular, providing detailed information for different percentiles (25th and 75th) of income ranks, separated by gender and country of origin. While it does not explicitly mention the time period for each data point, it includes age-related columns (e.g., age of parents in 2015), which provide temporal context for the observations.

Table 7

```python
#5 outliners
# Function to remove outliers using IQR
def remove_outliers(df_7):
    for column in df_7.columns:
        if df_7[column].dtype in ['int64', 'float64']:  # Adjust this to include/exclude certain types, to only get nu
            Q1 = df_7[column].quantile(0.25)
            Q3 = df_7[column].quantile(0.75)
            IQR = Q3 - Q1
            lower_bound = Q1 - 1.5 * IQR
            upper_bound = Q3 + 1.5 * IQR

            # You can choose to remove or replace the outliers
            df_7[column] = np.where((df_7[column] < lower_bound) | (df_7[column] > upper_bound), np.nan, df_7[column])
    return df_7

# Apply the function
df_7_clean = remove_outliers(df_7)
df_7_clean
```

	kid_race object	gender object	count float64	kid_edu1 float64	kid_edu2 float64	kid_edu3 float64	kid_edu4 float64
	White 20% Black 20% 3 others 60%	F 33.3% M 33.3% P 33.3%	6400.0 - 116000.0	0.0376 - 0.1803	0.0938 - 0.42	0.2742 - 0.4132	0.0962 - 0.6223
10	Hispanic	M	58500	0.1803	0.3251	0.3308	0.1638
11	Hispanic	P	116000	0.1538	0.2871	0.3533	0.2057
12	AIAN	F	6400	0.132	0.3109	0.4132	0.1439
13	AIAN	M	6700	0.1739	0.42	0.3098	0.0962
14	AIAN	P	13000	0.1535	0.3667	0.3604	0.1195

15 rows, 27 cols, showing 10 rows/page Page 2 of 2

```python
na_cols_df_7=['count','kid_edu3','par_edu1','kid_edu1_cond_par_edu1','kid_edu2_cond_par_edu1','kid_edu3_cond_par_edu1'
df_7_clean[na_cols_df_7]=df_7_clean[na_cols_df_7].fillna(df_7_clean[na_cols_df_7].mean())
df_7_clean[na_cols_df_7]
```

	count float64	kid_edu3 float64	par_edu1 float64	kid_edu1_cond_pa...	kid_edu2_cond_p...	kid_edu3_cond_p...	kid_edu3_cond_p...
	6400.0 - 116000.0	0.2742 - 0.4132	0.07719 - 0.235	0.09238 - 0.3126	0.1906 - 0.4749	0.227 - 0.3536	0.3235 - 0.4866
0	45300	0.3389	0.07719	0.1777	0.3496	0.3258	0.3855
1	45300	0.3267	0.0777	0.243	0.4247	0.2361	0.3851
2	45300	0.3327	0.07745	0.2112	0.3881	0.2799	0.3853
3	51500	0.3891	0.1653	0.1933	0.3247	0.3473	0.4437
4	50000	0.3195	0.1633	0.3126	0.3891	0.227	0.3759
5	101000	0.3549	0.1643	0.2516	0.3562	0.2885	0.4102
6	19000	0.3412428571	0.2239	0.2193128571	0.3590692308	0.2843	0.4009571429
7	22500	0.2977	0.235	0.1128	0.1906	0.3106	0.3489
8	41000	0.2742	0.2299	0.09238	0.3590692308	0.2989	0.3235
9	58000	0.3762	0.1650866667	0.1942	0.2929	0.3536	0.4338

15 rows, 9 cols, showing 10 rows/page Page 1 of 2

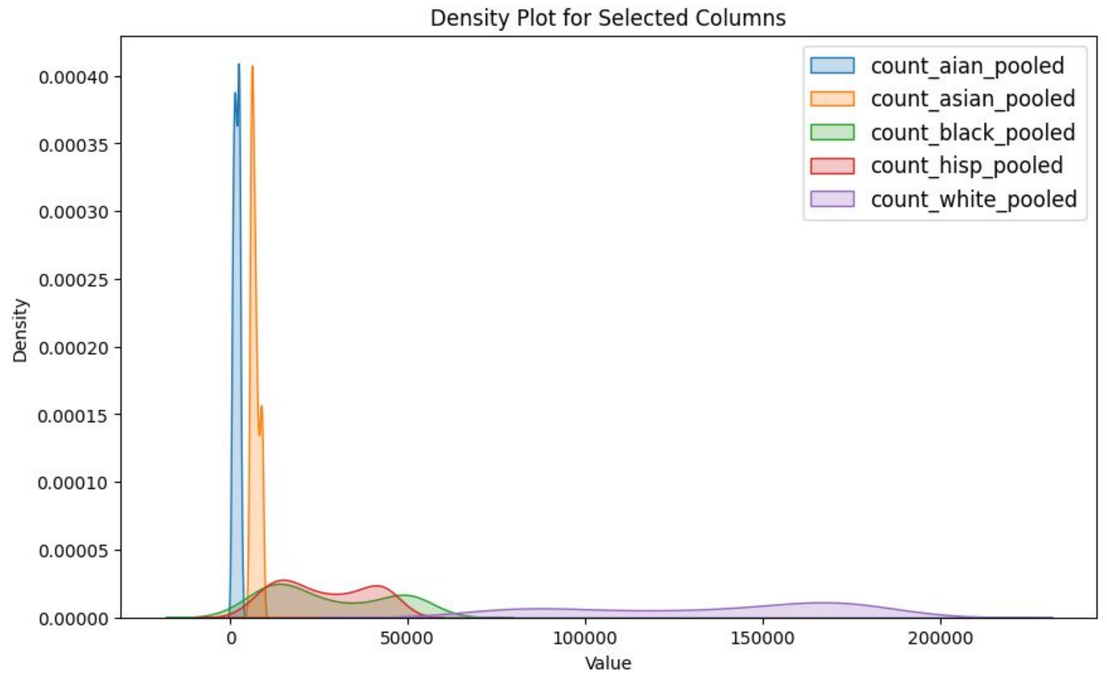
```python
df_Description=pd.DataFrame(
    [["kid_race"," describing the race of a given child"],
     ["gender",    "describing the gender of a given child,"],
     ["kid_edu[k]", "children whose educational attainment is k-level) "],
     ["par_edu[p]", "parental educational attainment is p-level "],
     ["kid_edu[k]_cond_par_edu[p]","children with parental educational attainment p whose own educational attainment 
    columns = ["Variable Name","Description"])
df_Description
```

	Variable Name obj...	Description object
0	kid_race	describing the race of a given child
1	gender	describing the gender of a given child,
2	kid_edu[k]	children whose educational attainment is k-level)
3	par_edu[p]	parental educational attainment is p-level
4	kid_edu[k]_cond_...	children with parental educational attainment p whose own educational attainment is k

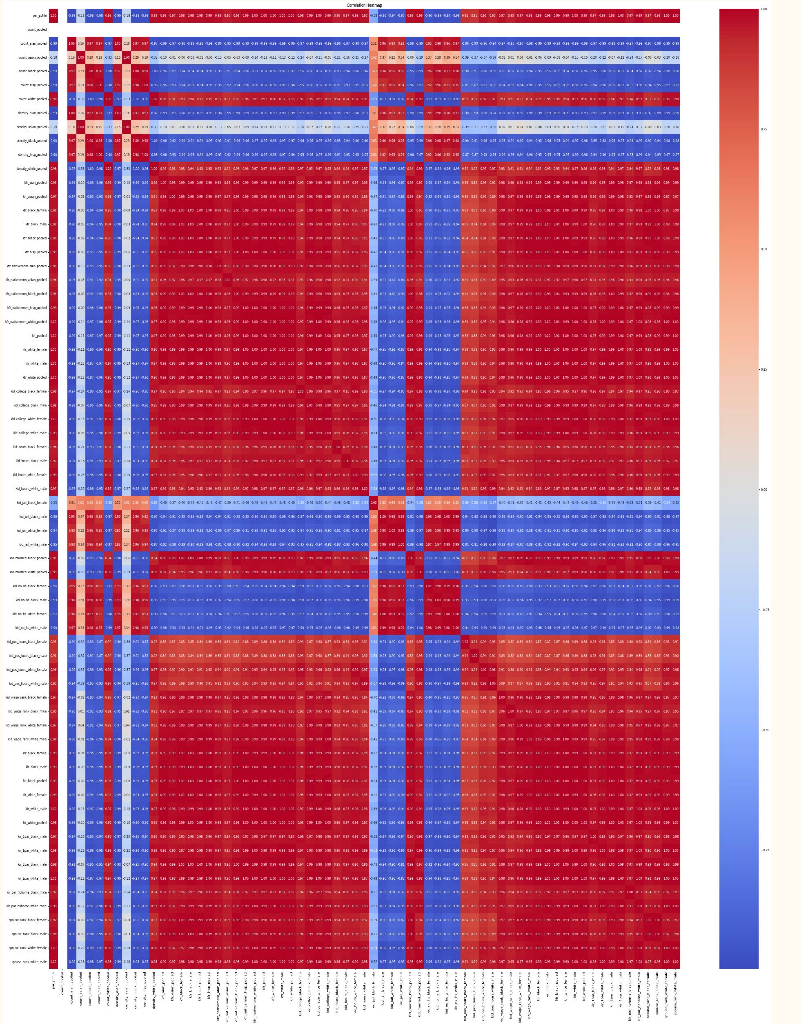
5 rows, 2 cols, showing 10 rows/page Page 1 of 1

EDA

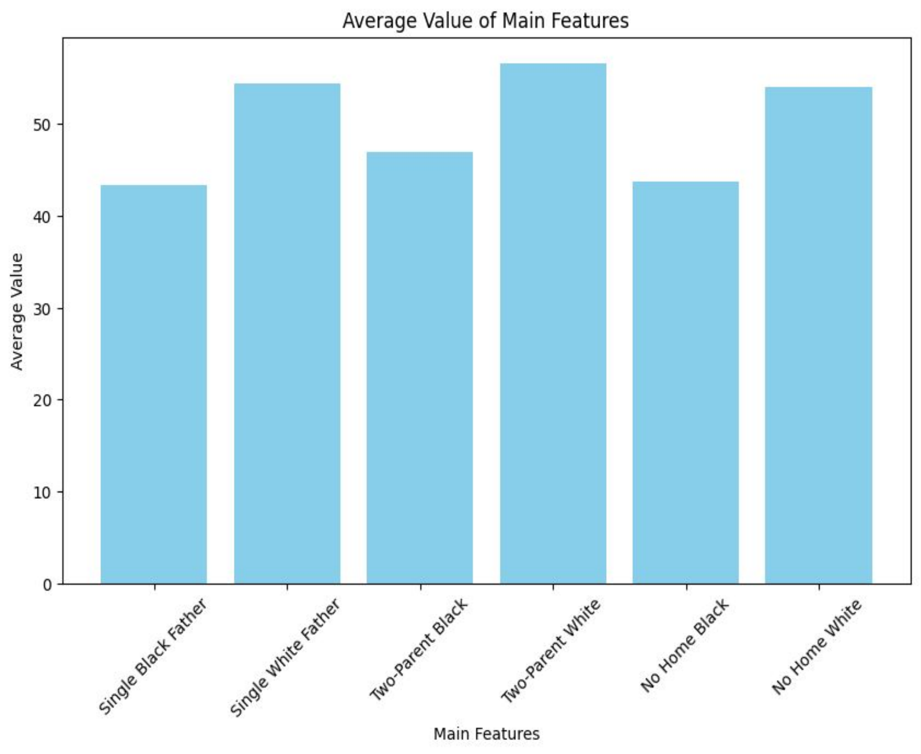
Table 1



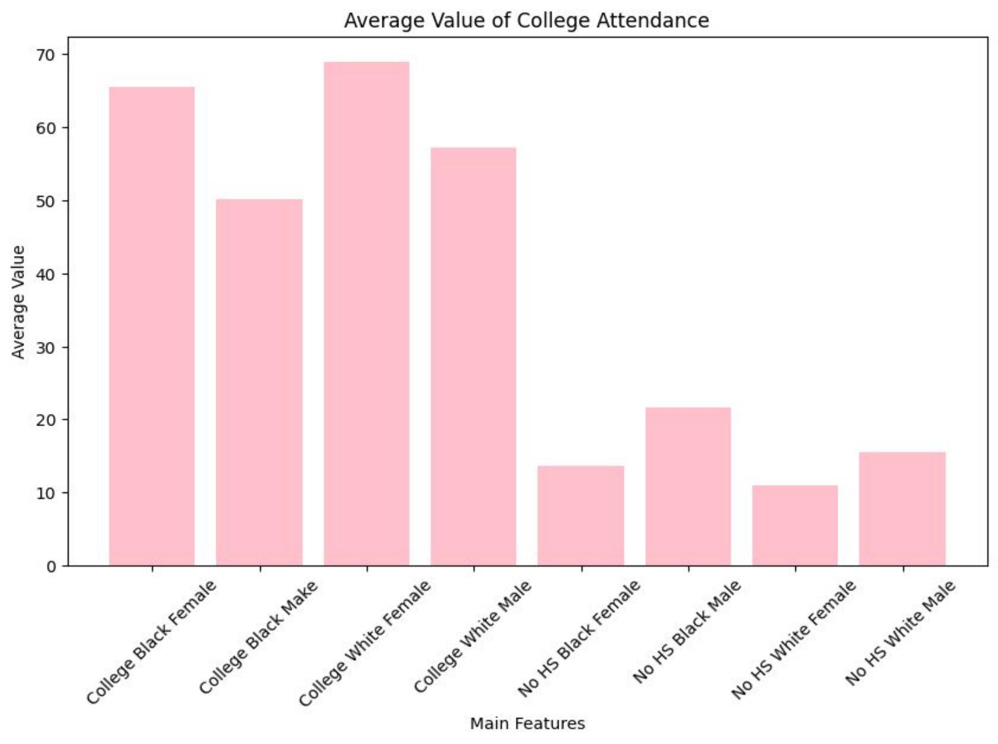
Density Plot for Selected Columns
count_aian_pooled
count_asian_pooled
count_black_pooled
count_hisp_pooled
count_white_pooled
Density
Value
0.00040
0.00035
0.00030
0.00025
0.00020
0.00015
0.00010
0.00005
0.00000
0
50000
100000
150000
200000



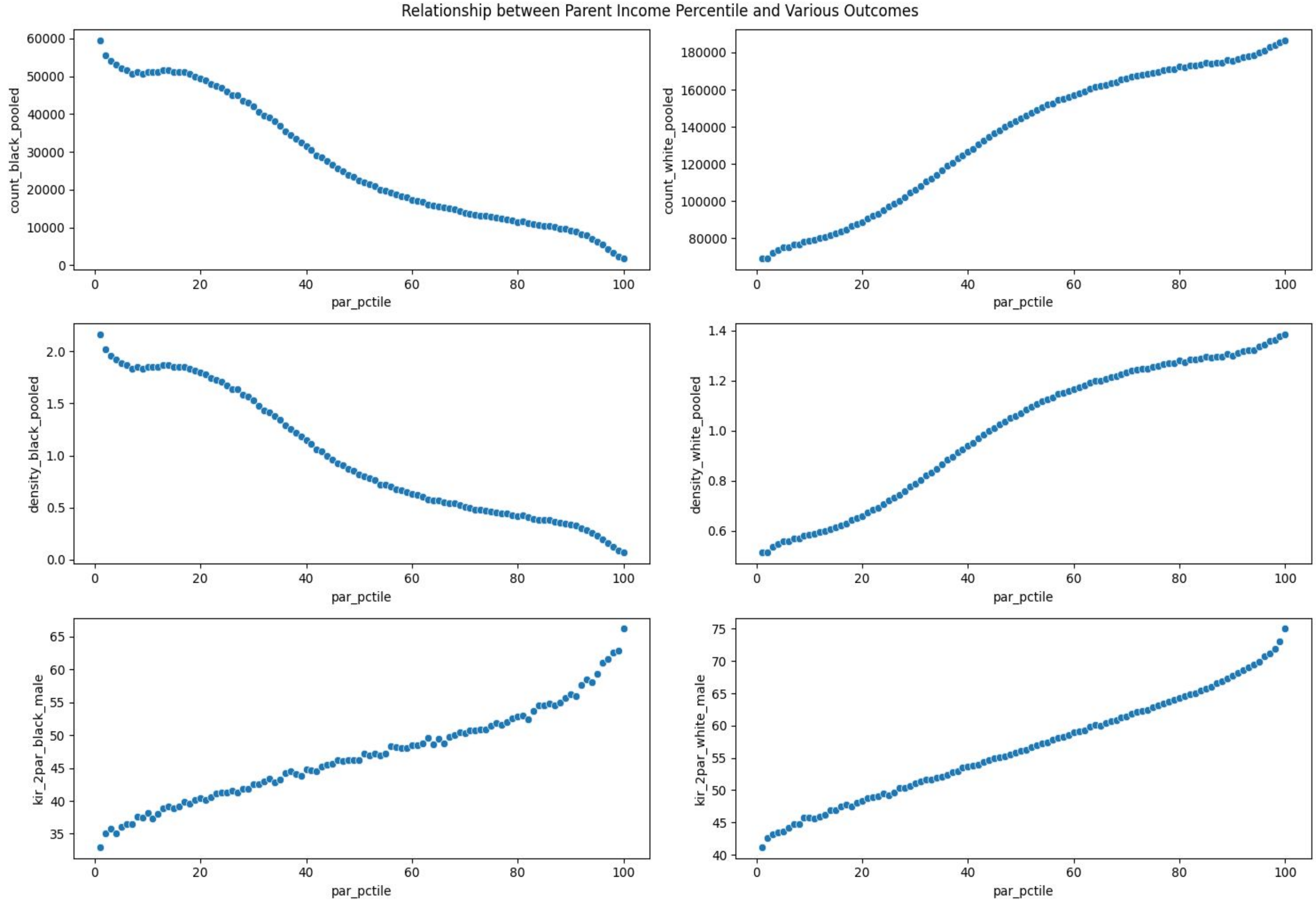
Correlation Heatmap



Mean child individual income rank for children based on

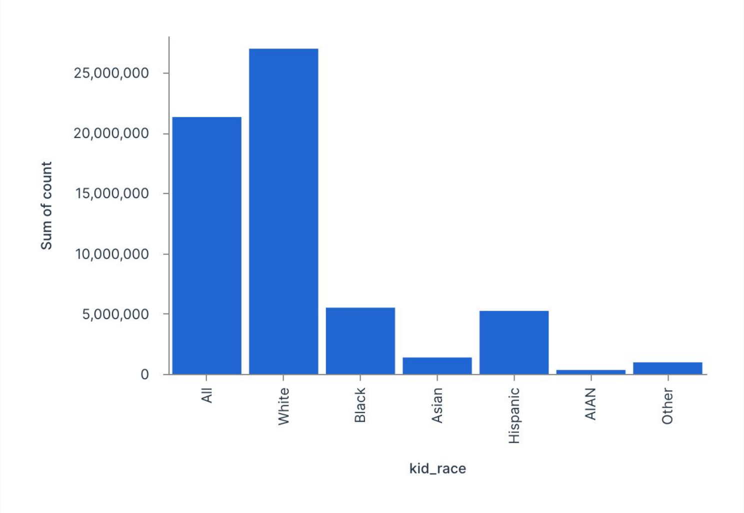


Percentage of children with college attendance by race and gender

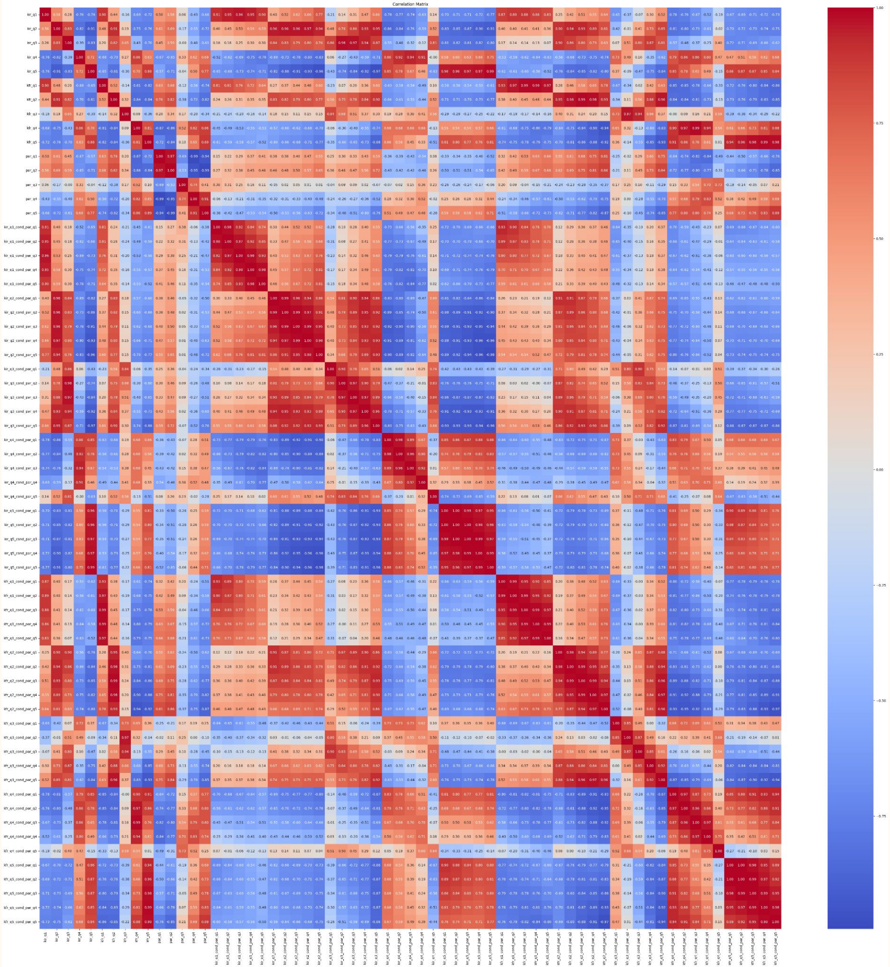


Shows the distribution of parent household income rank by race

Table 2

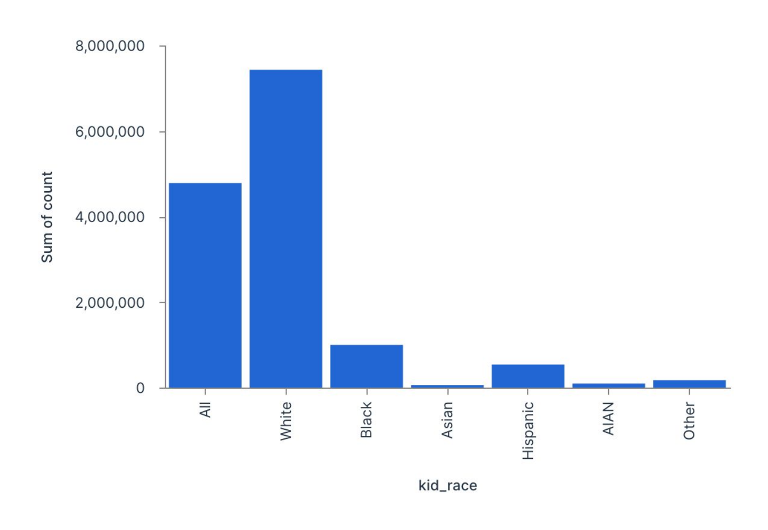


25,000,000
20,000,000
15,000,000
10,000,000
5,000,000
0
Sum of count
All
White
Black
Asian
Hispanic
AIAN
Other
kid_race

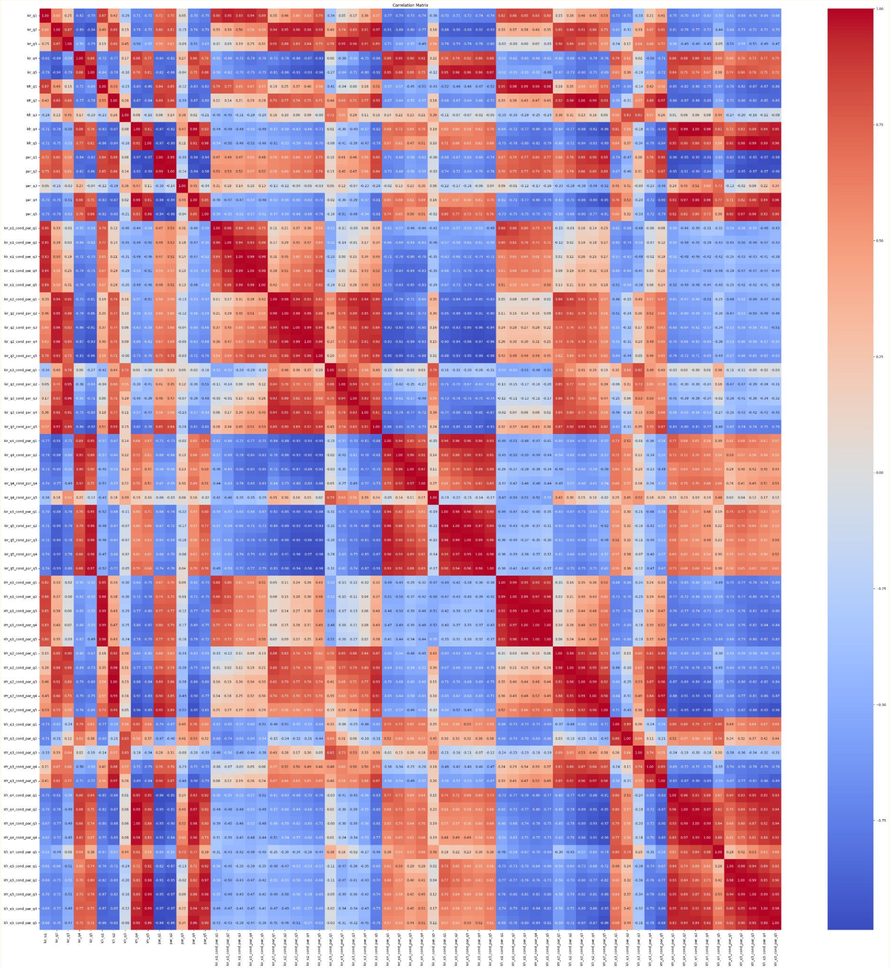


Correlation Matrix

Table 3

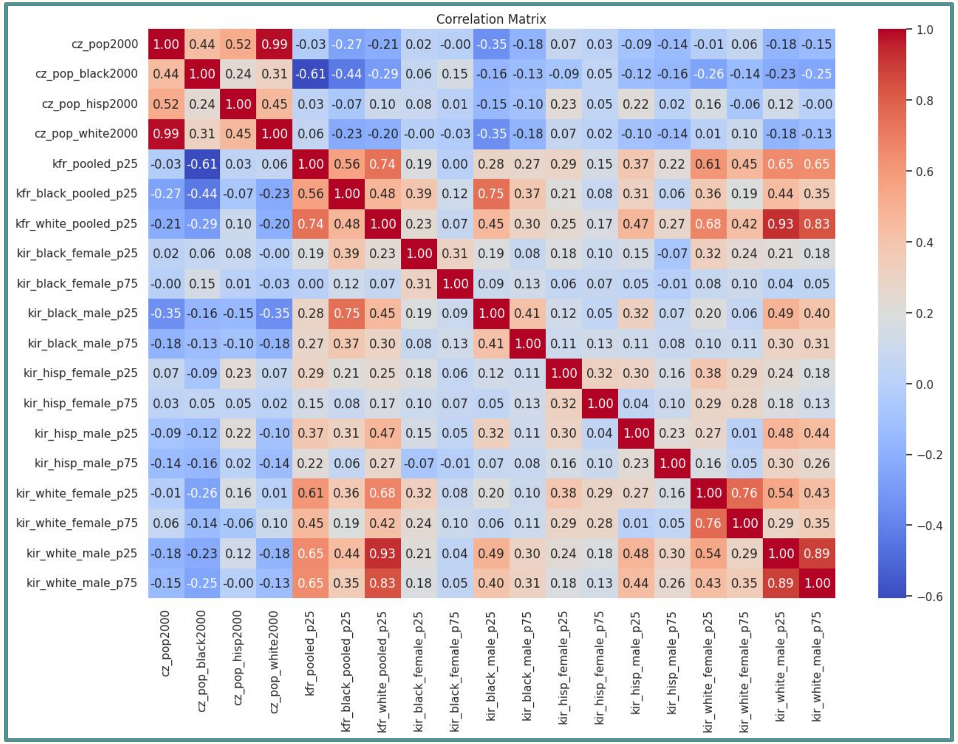


8,000,000
6,000,000
4,000,000
2,000,000
0
Sum of count
All
White
Black
Asian
Hispanic
AIAN
Other
kid_race

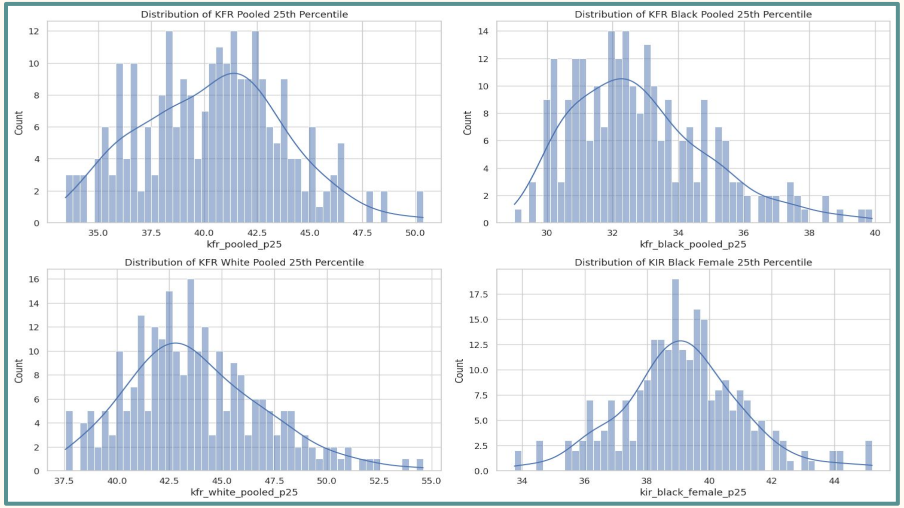


Correlation Matrix

Table 4



	cz_pop2000	cz_pop_black2000	cz_pop_hisp2000	cz_pop_white2000	kfr_pooled_p25	kfr_black_pooled_p25	kfr_white_pooled_p25	kir_black_female_p25	kir_black_female_p75	kir_black_male_p25	kir_black_male_p75	kir_hisp_female_p25	kir_hisp_female_p75	kir_hisp_male_p25	kir_hisp_male_p75	kir_white_female_p25	kir_white_female_p75	kir_white_male_p25	kir_white_male_p75
cz_pop2000	1.00	0.44	0.52	0.99	-0.03	-0.27	-0.21	0.02	-0.00	-0.35	-0.18	0.07	0.03	-0.09	-0.14	-0.01	0.06	-0.18	-0.15
cz_pop_black2000	0.44	1.00	0.24	0.31	-0.61	-0.44	-0.29	0.06	0.15	-0.16	-0.13	-0.09	0.05	-0.12	-0.16	-0.26	-0.14	-0.23	-0.25
cz_pop_hisp2000	0.52	0.24	1.00	0.45	0.03	-0.07	0.10	0.08	0.01	-0.15	-0.10	0.23	0.05	0.22	0.02	0.16	-0.06	0.12	-0.00
cz_pop_white2000	0.99	0.31	0.45	1.00	0.06	-0.23	-0.20	-0.00	-0.03	-0.35	-0.18	0.07	0.02	-0.10	-0.14	0.01	0.10	-0.18	-0.13
kfr_pooled_p25	-0.03	-0.61	0.03	0.06	1.00	0.56	0.74	0.19	0.00	0.28	0.27	0.29	0.15	0.37	0.22	0.61	0.45	0.65	0.65
kfr_black_pooled_p25	-0.27	-0.44	-0.07	-0.23	0.56	1.00	0.48	0.39	0.12	0.75	0.37	0.21	0.08	0.31	0.06	0.36	0.19	0.44	0.35
kfr_white_pooled_p25	-0.21	-0.29	0.10	-0.20	0.74	0.48	1.00	0.23	0.07	0.45	0.30	0.25	0.17	0.47	0.27	0.68	0.42	0.93	0.83
kir_black_female_p25	0.02	0.06	0.08	-0.00	0.19	0.39	0.23	1.00	0.31	0.19	0.08	0.18	0.10	0.15	-0.07	0.32	0.24	0.21	0.18
kir_black_female_p75	-0.00	0.15	0.01	-0.03	0.00	0.12	0.07	0.31	1.00	0.09	0.13	0.06	0.07	0.05	-0.01	0.08	0.10	0.04	0.05
kir_black_male_p25	-0.35	-0.16	-0.15	-0.35	0.28	0.75	0.45	0.19	0.09	1.00	0.41	0.12	0.05	0.32	0.07	0.20	0.06	0.49	0.40
kir_black_male_p75	-0.18	-0.13	-0.10	-0.18	0.27	0.37	0.30	0.08	0.13	0.41	1.00	0.11	0.13	0.11	0.08	0.10	0.11	0.30	0.31
kir_hisp_female_p25	0.07	-0.09	0.23	0.07	0.29	0.21	0.25	0.18	0.06	0.12	0.11	1.00	0.32	0.30	0.16	0.38	0.29	0.24	0.18
kir_hisp_female_p75	0.03	0.05	0.05	0.02	0.15	0.08	0.17	0.10	0.07	0.05	0.13	0.32	1.00	0.04	0.10	0.29	0.28	0.18	0.13
kir_hisp_male_p25	-0.09	-0.12	0.22	-0.10	0.37	0.31	0.47	0.15	0.05	0.32	0.11	0.30	0.04	1.00	0.23	0.27	0.01	0.48	0.44
kir_hisp_male_p75	-0.14	-0.16	0.02	-0.14	0.22	0.06	0.27	-0.07	-0.01	0.07	0.08	0.16	0.10	0.23	1.00	0.16	0.05	0.30	0.26
kir_white_female_p25	-0.01	-0.26	0.16	0.01	0.61	0.36	0.68	0.32	0.08	0.20	0.10	0.38	0.29	0.27	0.16	1.00	0.76	0.54	0.43
kir_white_female_p75	0.06	-0.14	-0.06	0.10	0.45	0.19	0.42	0.24	0.10	0.06	0.11	0.29	0.28	0.01	0.05	0.76	1.00	0.29	0.35
kir_white_male_p25	-0.18	-0.23	0.12	-0.18	0.65	0.44	0.93	0.21	0.04	0.49	0.30	0.24	0.18	0.48	0.30	0.54	0.29	1.00	0.89
kir_white_male_p75	-0.15	-0.25	-0.00	-0.13	0.65	0.35	0.83	0.18	0.05	0.40	0.31	0.18	0.13	0.44	0.26	0.43	0.35	0.89	1.00



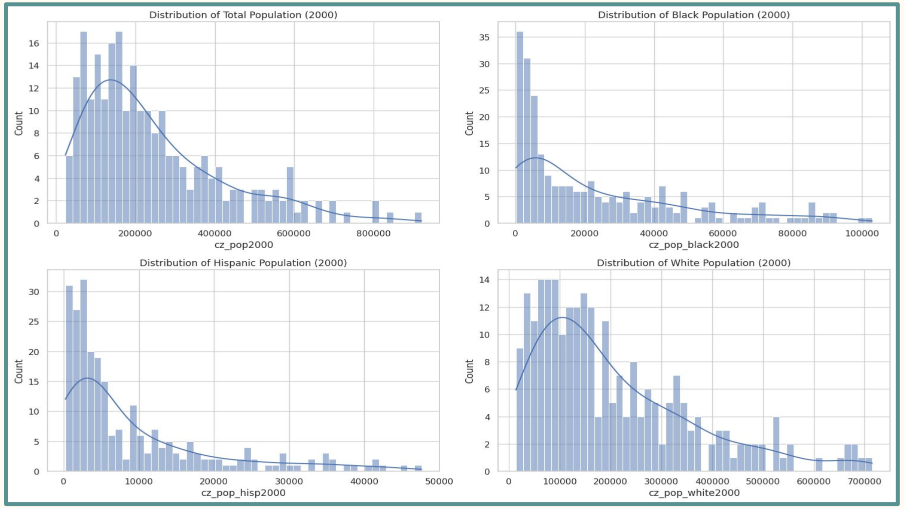
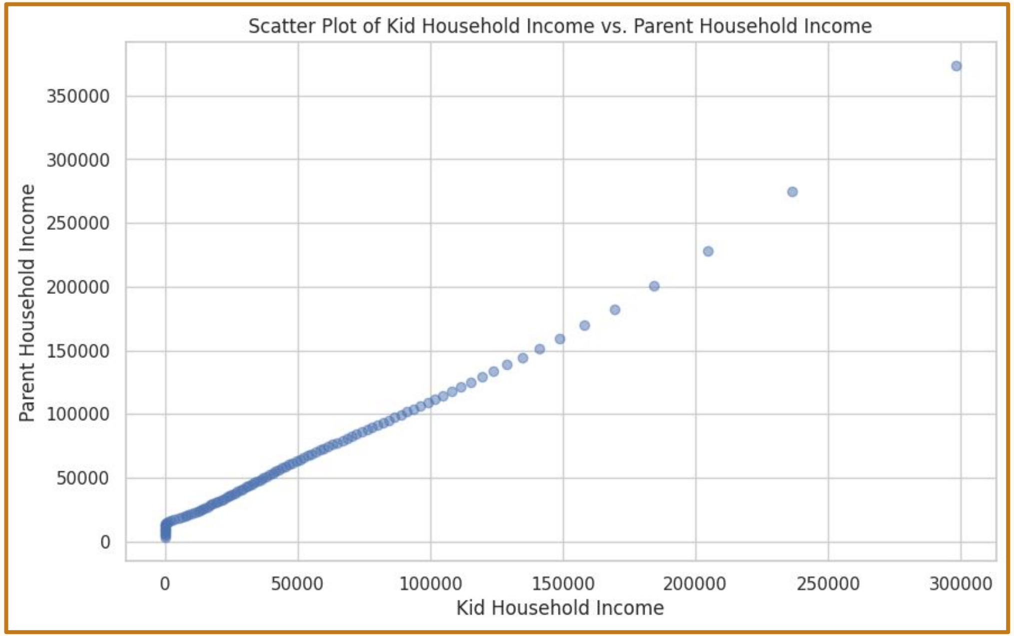
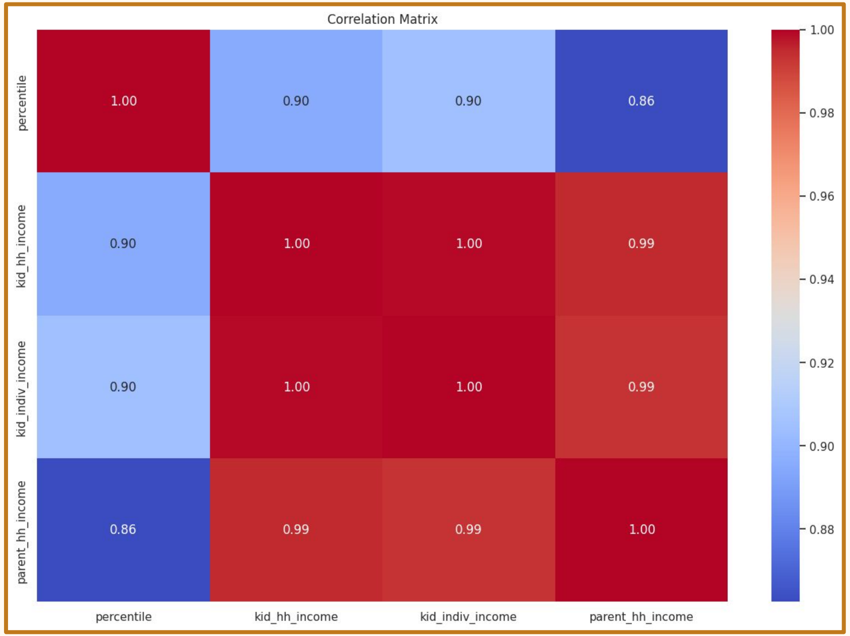


Table 5



Scatter Plot of Kid Household Income vs. Parent Household Income
Parent Household Income
Kid Household Income
0
50000
100000
150000
200000
250000
300000
350000



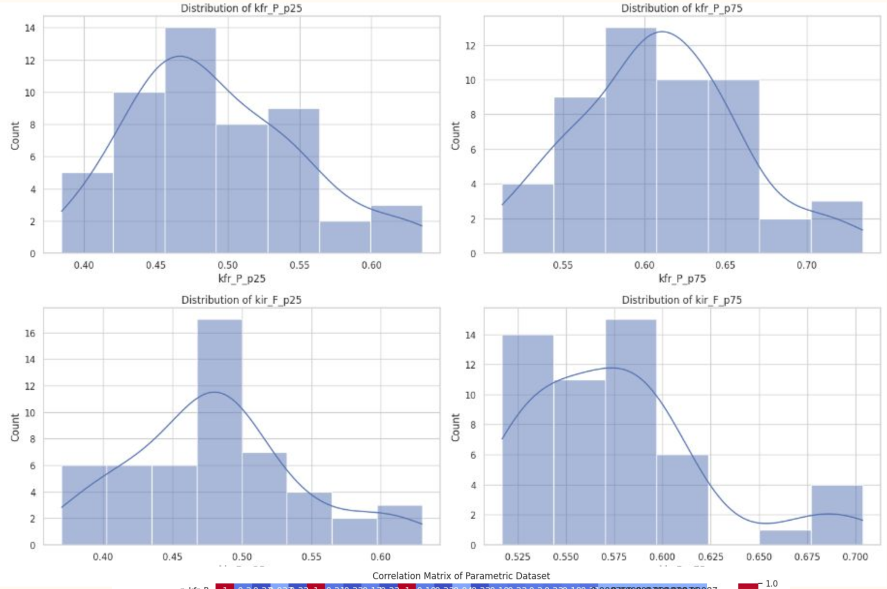
Correlation Matrix
percentile
kid_hh_income
kid_indiv_income
parent_hh_income
1.00 0.90 0.90 0.86
0.90 1.00 1.00 0.99
0.90 1.00 1.00 0.99
0.86 0.99 0.99 1.00
1.00
0.98
0.96
0.94
0.92
0.90
0.88

Table 6a

The histograms are useful for understanding the distribution characteristics of income rank across different percentiles and potentially different groups within a population. The P75 distributions appear slightly more spread out than the P25 ones, indicating more variability in the higher percentile ranks.

The heatmap was useful for identifying relationships between different socioeconomic and demographic variables. For instance, understanding how the presence of parents or their duration in the U.S. correlates with income rank was crucial for the paper's arguments on social mobility.

The kfr_P_p25 and kfr_P_p75 (parents' income rank in the commuting zone for children) show strong correlations with kir_F_p25, kir_F_p75, kir_M_p25, kir_M_p75, etc. This implies a significant relationship between the economic status of parents and their children, aligning with the intergenerational transmission of economic status.



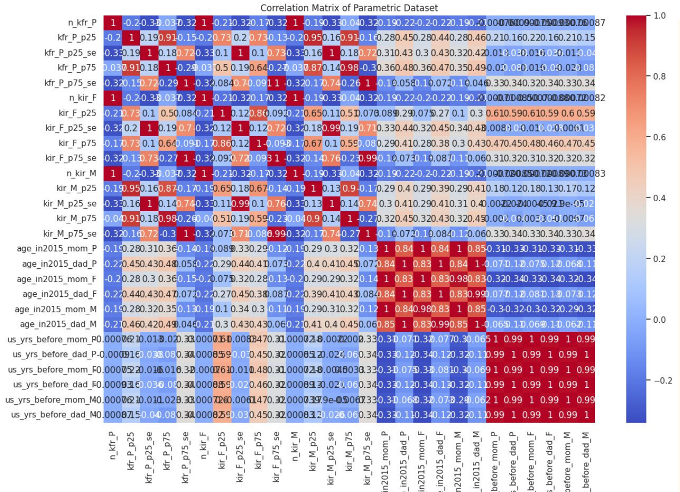
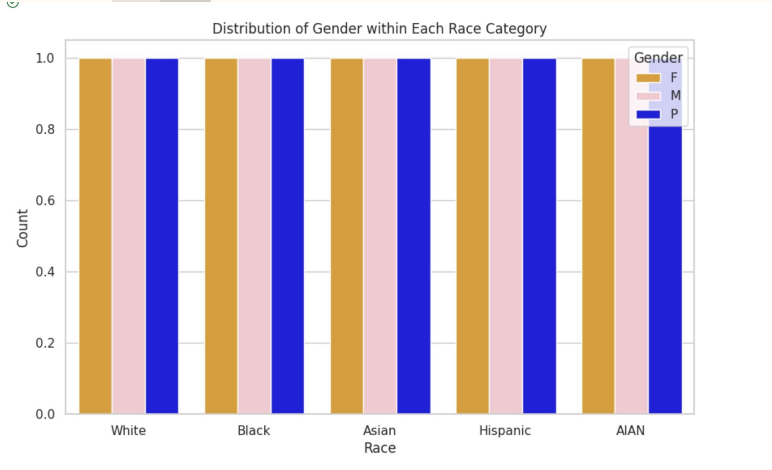
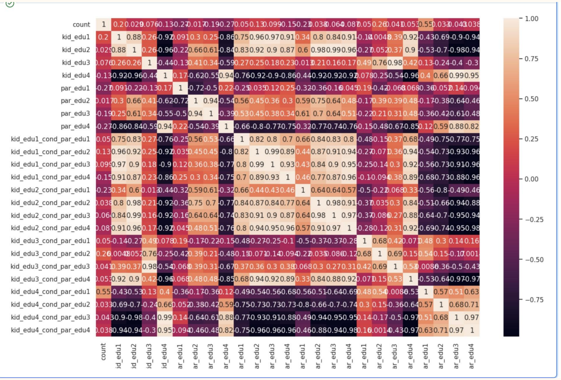


Table 7

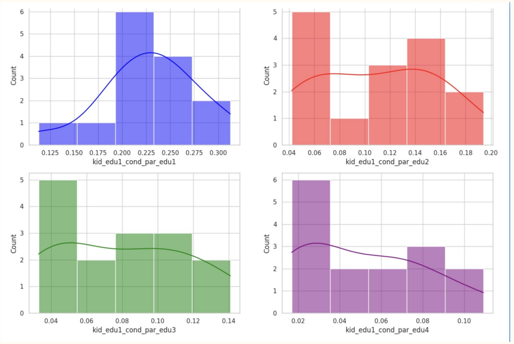
parental educational attainment[p] levels:

1. Less than high school
2. High school
3. Some college/associate degree
4. Bachelor degree or higher

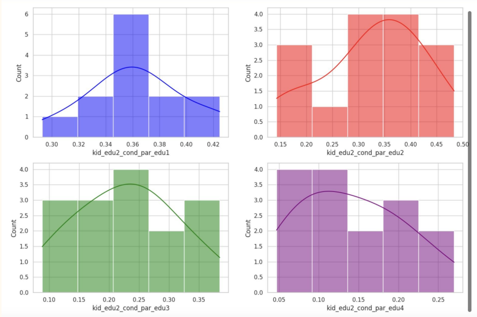




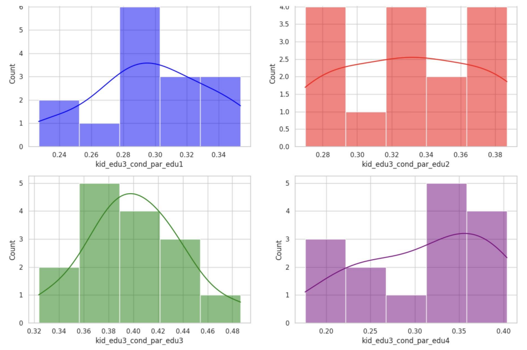
American Indian/Alaska Native (AIAN)



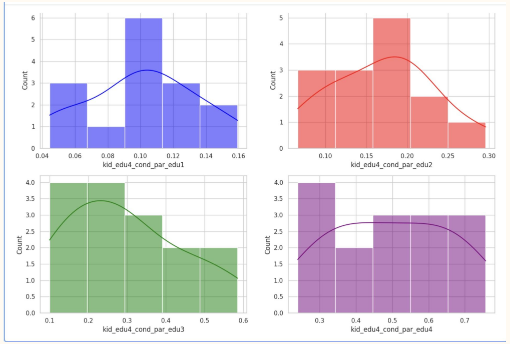
children with parental educational attainment *at each levels* whose own educational attainment is *Less than High school*



children with parental educational attainment *at each levels* whose own educational attainment is *High School*



children with parental educational attainment *at each levels* whose own educational attainment is *Some College/Associate Degree*

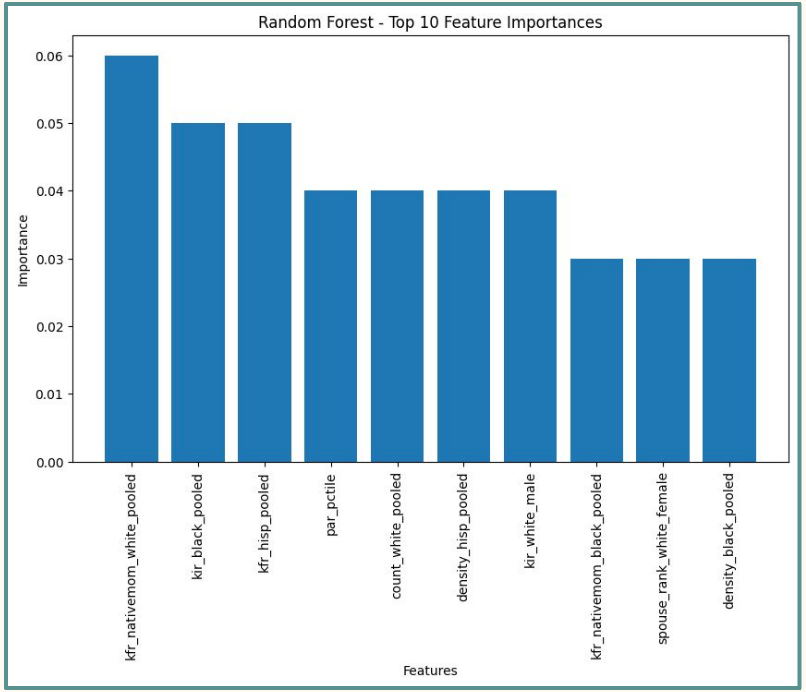


children with parental educational attainment *at each levels* whose own educational attainment is *Bachelor Degree or Higher*

Modeling

Feature Importance

Feature	Importance
kfr_nativemom_white_pooled	0.06
kir_black_pooled	0.05
kfr_hisp_pooled	0.05
par_pctile	0.04
count_white_pooled	0.04
density_hisp_pooled	0.04
kir_white_male	0.04
kfr_nativemom_black_pooled	0.03
spouse_rank_white_female	0.03
density_black_pooled	0.03



Comparison to Paper

Table II
Association Between Black Father Presence and Black Men's Upward Mobility Across Census Tracts: OLS Regression Estimates

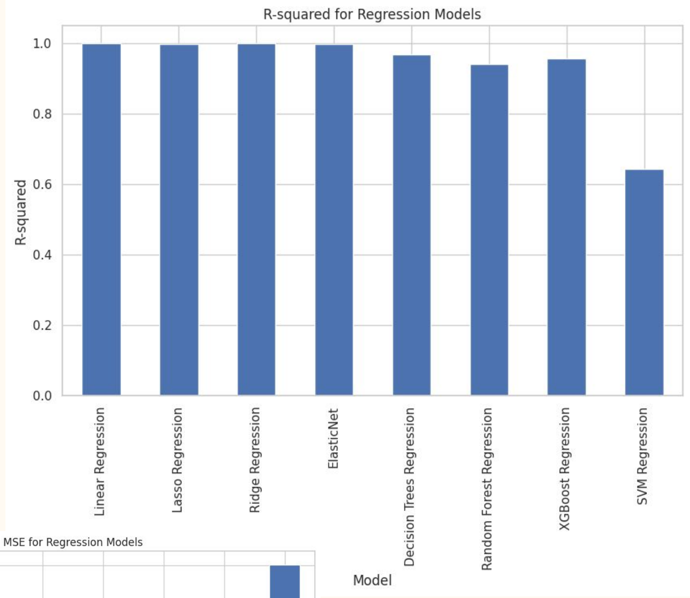
Dependent Variable: Mean Individual Income Rank for Black Men Raised in Low-Income (p = 25) Families

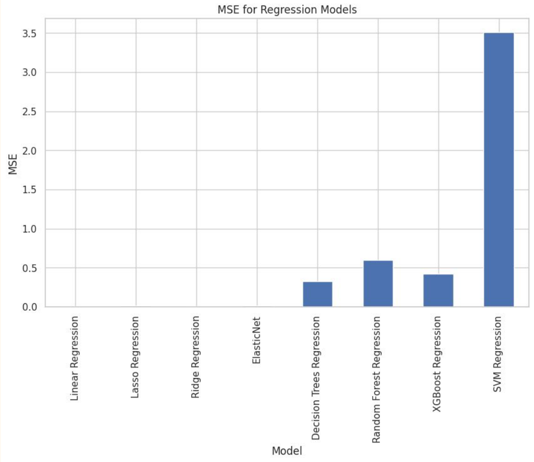
	Baseline	Black vs. White Father Presence	State FE	Children with Fathers Absent	Children with Two Parents	Black Fathers per Child	Gender Ratio	All Tracts	Adulthood Tract FE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Low-Income Black Father Presence	0.0492	0.0450	0.0501	0.0279	0.0461			0.0806	0.1052
	(0.0062)	(0.0068)	(0.0066)	(0.0108)	(0.0128)			(0.0036)	(0.0079)
Low-Income White Father Presence		0.0077							
		(0.0076)							
Low-Income Black Father Presence in 2000						0.0382	0.0387		
						(0.0043)	(0.0043)		
Low-Income Black Male Filers Per Child							-0.0011		
							(0.0011)		
Low Poverty Tracts	X	X	X	X	X	X	X		
State FE's			X						
Current Tract FE's									X
Number of Tracts	10,582	10,444	10,582	5,159	5,280	10,177	10,177	28,850	
Number of Individual Obs.									193,000

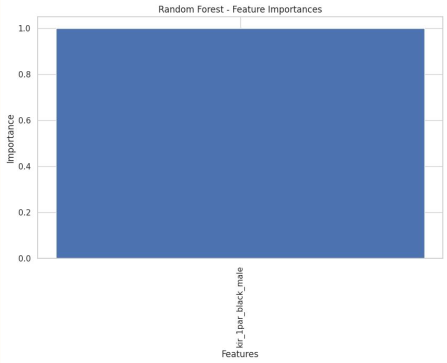
Model 1

```
Results for regression 1:
                            OLS Regression Results
==============================================================================
Dep. Variable:         kir_black_male   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
Method:                 Least Squares   F-statistic:                 3.864e+04
Date:                Thu, 27 Jun 2024   Prob (F-statistic):           1.33e-38
Time:                        20:06:24   Log-Likelihood:                 33.564
No. Observations:                  25   AIC:                            -63.13
Df Residuals:                      23   BIC:                            -60.69
Df Model:                           1
Covariance Type:            nonrobust
=====================================================================================
                        coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------
const                -0.3548      0.183     -1.936      0.065      -0.734       0.024
kir_1par_black_male   1.0233      0.005    196.571      0.000       1.013       1.034
=====================================================================================
Omnibus:                        3.648   Durbin-Watson:                   1.753
Prob(Omnibus):                  0.161   Jarque-Bera (JB):                1.997
Skew:                           0.430   Prob(JB):                        0.368
Kurtosis:                       4.085   Cond. No.                         490.
==============================================================================
```

```
Linear Regression R-squared: 0.999, MSE: 0.011
Lasso Regression R-squared: 0.998, MSE: 0.021
Ridge Regression R-squared: 0.999, MSE: 0.011
ElasticNet R-squared: 0.998, MSE: 0.021
Decision Trees Regression R-squared: 0.966, MSE: 0.332
Random Forest Regression R-squared: 0.939, MSE: 0.598
XGBoost Regression R-squared: 0.957, MSE: 0.426
SVM Regression R-squared: 0.644, MSE: 3.510
```



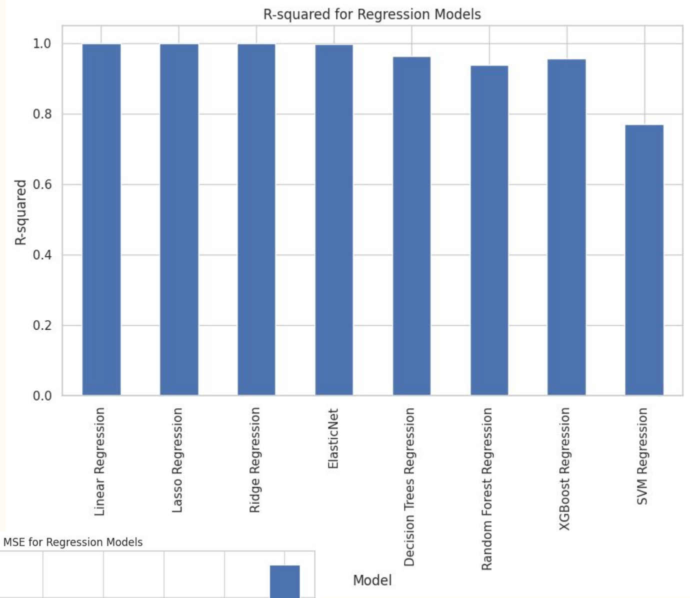


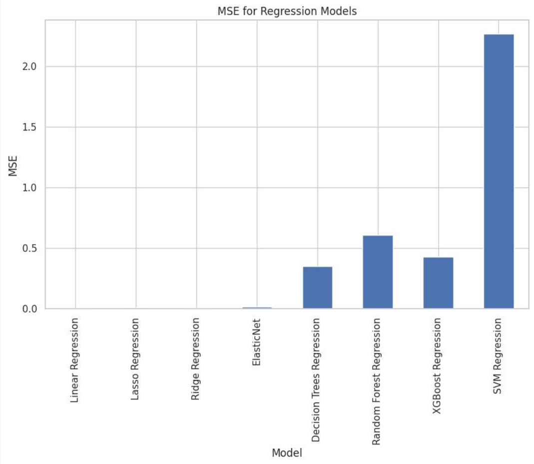


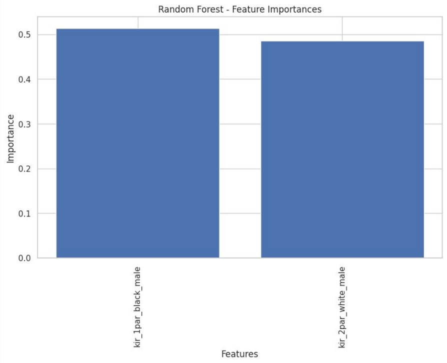
Model 2

```
Results for regression 2:
                            OLS Regression Results
==============================================================================
Dep. Variable:         kir_black_male   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 2.709e+04
Date:                Thu, 27 Jun 2024   Prob (F-statistic):           4.93e-38
Time:                        20:06:24   Log-Likelihood:                 38.340
No. Observations:                  25   AIC:                            -70.68
Df Residuals:                      22   BIC:                            -67.02
Df Model:                           2
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                  -0.8650      0.222     -3.892      0.001      -1.326      -0.404
kir_1par_black_male     0.9138      0.034     26.498      0.000       0.842       0.985
kir_1par_white_male     0.0980      0.031      3.200      0.004       0.034       0.162
==============================================================================
Omnibus:                        1.643   Durbin-Watson:                   2.184
Prob(Omnibus):                  0.440   Jarque-Bera (JB):                1.120
Skew:                           0.231   Prob(JB):                        0.571
Kurtosis:                       2.071   Cond. No.                     1.14e+03
==============================================================================
```

```
Linear Regression R-squared: 0.999, MSE: 0.006
Lasso Regression R-squared: 0.999, MSE: 0.009
Ridge Regression R-squared: 1.000, MSE: 0.004
ElasticNet R-squared: 0.998, MSE: 0.016
Decision Trees Regression R-squared: 0.964, MSE: 0.351
Random Forest Regression R-squared: 0.938, MSE: 0.611
XGBoost Regression R-squared: 0.957, MSE: 0.426
SVM Regression R-squared: 0.770, MSE: 2.265
```



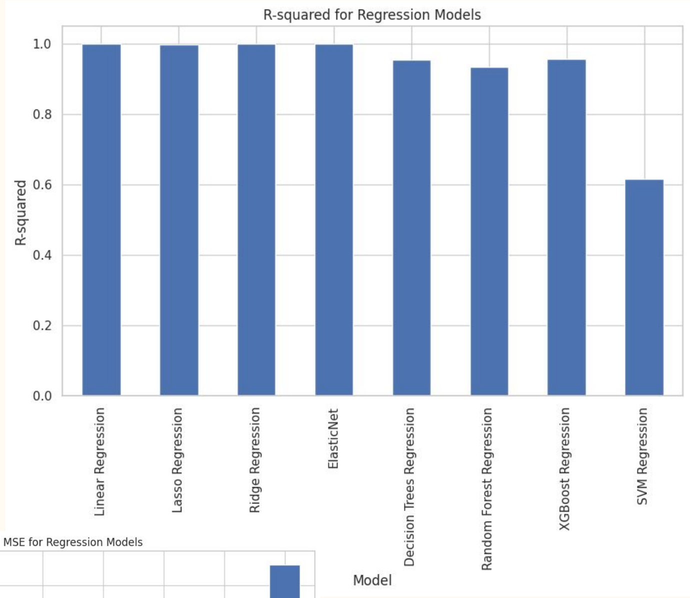


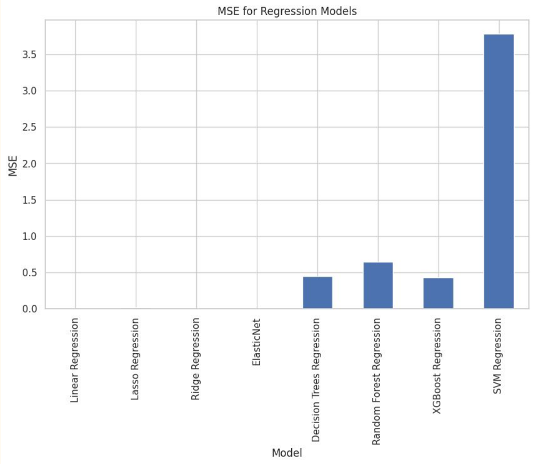


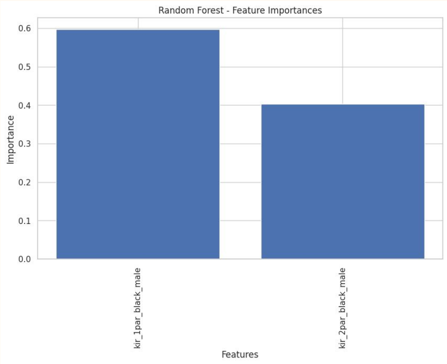
Model 3

```
Results for regression 3:
                            OLS Regression Results
==============================================================================
Dep. Variable:         kir_black_male   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 1.020e+05
Date:                Thu, 27 Jun 2024   Prob (F-statistic):           2.29e-44
Time:                        20:06:24   Log-Likelihood:                 54.911
No. Observations:                  25   AIC:                            -103.8
Df Residuals:                      22   BIC:                            -100.2
Df Model:                           2
Covariance Type:            nonrobust
=================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------
const              -1.4953      0.139    -10.720      0.000      -1.785      -1.206
kir_1par_black_male  0.9103      0.012     78.786      0.000       0.886       0.934
kir_2par_black_male  0.1335      0.013      9.968      0.000       0.106       0.161
==============================================================================
Omnibus:                        5.809   Durbin-Watson:                   1.415
Prob(Omnibus):                  0.055   Jarque-Bera (JB):                3.817
Skew:                           0.737   Prob(JB):                        0.148
Kurtosis:                       4.222   Cond. No.                     1.27e+03
==============================================================================
```

```
Linear Regression R-squared: 1.000, MSE: 0.003
Lasso Regression R-squared: 0.998, MSE: 0.020
Ridge Regression R-squared: 1.000, MSE: 0.002
ElasticNet R-squared: 1.000, MSE: 0.004
Decision Trees Regression R-squared: 0.955, MSE: 0.447
Random Forest Regression R-squared: 0.935, MSE: 0.642
XGBoost Regression R-squared: 0.956, MSE: 0.431
SVM Regression R-squared: 0.616, MSE: 3.781
```



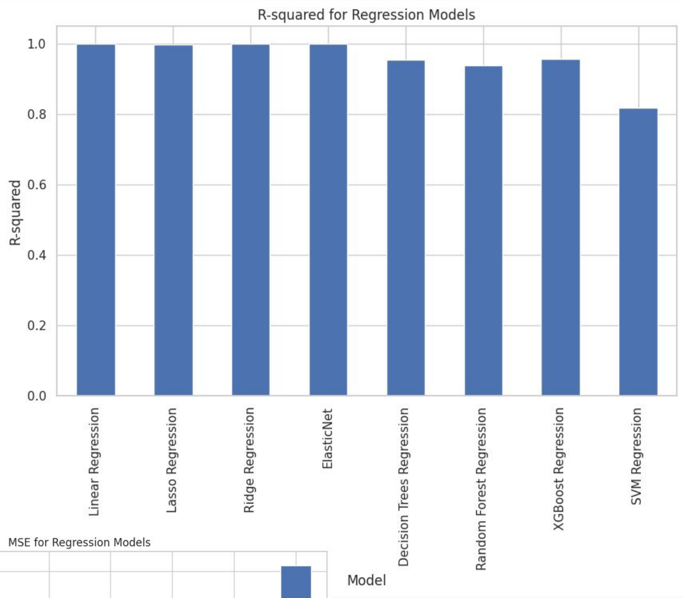


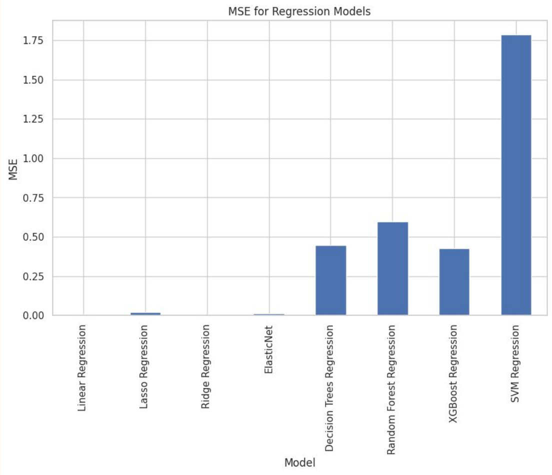


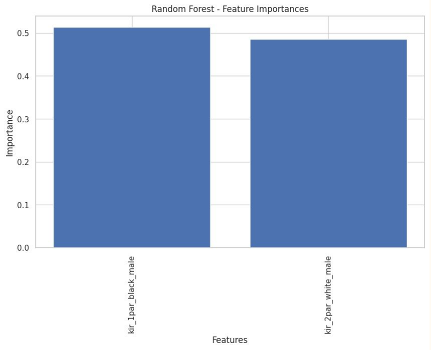
Model 4

```
Results for regression 4:
                            OLS Regression Results
==============================================================================
Dep. Variable:         kir_black_male   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 3.044e+04
Date:                Thu, 27 Jun 2024   Prob (F-statistic):           1.37e-38
Time:                        20:06:24   Log-Likelihood:                 39.799
No. Observations:                  25   AIC:                            -73.60
Df Residuals:                      22   BIC:                            -69.94
Df Model:                           2
Covariance Type:            nonrobust
==================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const               -1.8693      0.427     -4.375      0.000      -2.755      -0.983
kir_1par_black_male  0.9368      0.023     40.210      0.000       0.888       0.985
kir_2par_white_male  0.0985      0.026      3.772      0.001       0.044       0.153
==============================================================================
Omnibus:                        0.574   Durbin-Watson:                   2.064
Prob(Omnibus):                  0.751   Jarque-Bera (JB):                0.357
Skew:                           0.281   Prob(JB):                        0.836
Kurtosis:                       2.837   Cond. No.                     2.37e+03
==============================================================================
```

Linear Regression R-squared: 0.999, MSE: 0.006
Lasso Regression R-squared: 0.998, MSE: 0.021
Ridge Regression R-squared: 1.000, MSE: 0.005
ElasticNet R-squared: 0.999, MSE: 0.012
Decision Trees Regression R-squared: 0.955, MSE: 0.447
Random Forest Regression R-squared: 0.939, MSE: 0.597
XGBoost Regression R-squared: 0.957, MSE: 0.425
SVM Regression R-squared: 0.819, MSE: 1.786







Interpretations and Conclusion

Interpretation

Persistent Income Disparities:

- Significant income mobility disparities persist among black men, unaffected by differences in parental income, education, or marital status, and are evident even within the same neighborhoods.

Influence of Parental Presence:

- Analysis using a Random Forest model shows that the presence of one or two parents (`kir_1par_black_male` and `kir_2par_black_male`) significantly influences the income rank outcomes for black males, with both factors having nearly equal importance.

Caveats

Data Limitations Impact Model Accuracy:

- Challenges arose during modeling due to the unavailability of mergeable data sets from the original authors, restricting us to limited data for each predictive question. This led to difficulties in accurately modeling the income rank of black children based on parental presence, highlighting discrepancies in social mobility across different races in America due to the use of disjointed government data collected over various years and states.

Analysis of R-squared Values:

- R-squared values for different regression models, such as Linear Regression, Lasso, Ridge, and ElasticNet, approach 1, indicating high explained variability. However, lower R-squared values in the SVM Regression model suggest it explains less variability in income rank. These high values across most models may indicate overfitting, stressing the need for careful model evaluation and the inclusion of more comprehensive data to prevent misleading results.

Conclusion

Implications:

- Had our results have been reliable, specifically around the significant positive impact of parental presence on income rank, policymakers should invest in programs that support single-parent and two-parent families. This could include financial assistance, parenting support programs, and community resources aimed at strengthening family units. Understanding the factors that influence the income rank of black males is crucial for addressing economic disparities. The results highlight the importance of family structure in economic outcomes, guiding policymakers in designing effective interventions to foster socio-economic mobility and reduce income inequality.

Significance

- Understanding the factors that influence the income rank of black males is crucial for addressing economic disparities. The results highlight the importance of family structure in economic outcomes.

Thank You!