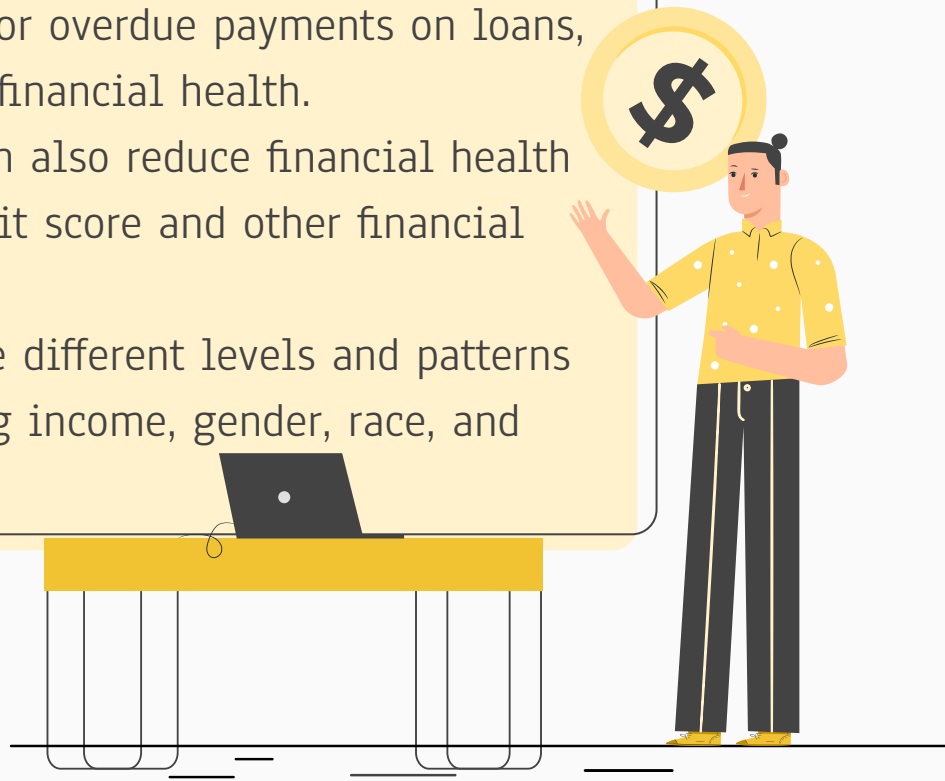
AN INVESTIGATION OF DEBT DELINQUENCY ACROSS THE U.S.

By: *Mason Kellerman, Aly Tan*, Sophia Yi, Diandian Shi, Calvin Windmiller Kolster

INTRODUCTION

- **Debt delinquency rate**, or overdue payments on loans, is an indicator of overall financial health.
- High debt delinquency can also reduce financial health in the future through credit score and other financial penalties.
- Different populations have different levels and patterns of debt delinquency, along income, gender, race, and location



DATA

SOURCE

A publicly accessible dataset from **Opportunity Insights**, containing county-level financial information on parental income, race, and adult credit outcomes across the United States.

WHAT THE DATA REPRESENTS

Each row represents a **“cohort”**: a combination of county × race × parental income. The dataset follows children born 1978–1985 and measures their credit outcomes in adulthood (2020).

WHY THIS DATA IS USEFUL

It provides diverse picture of financial health and credit access, and allow for intergenerational and developmental connections

DATA SET VARIABLES

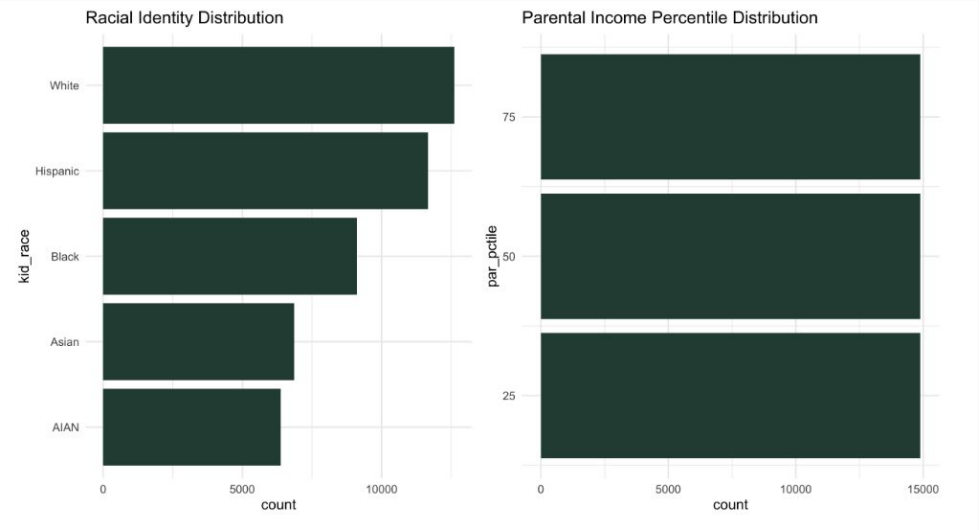
Table 1: Codebook for Opportunity Insights Credit Access Data

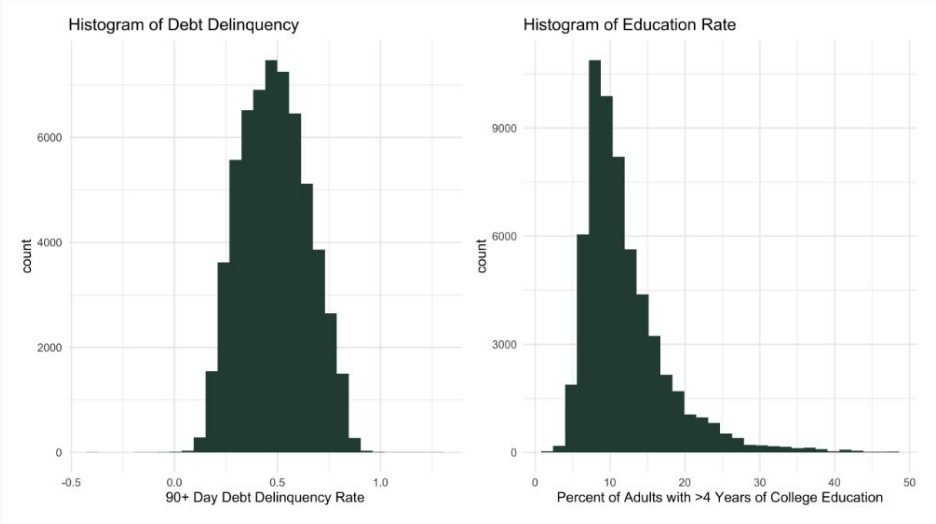
Variable	Variable Name	Data Type	Description
Parent's State (FIPS code)	`par_state`	categorical	The FIPS code for the Parent's State (or State-Like Entity)
Parent's State (Abbreviation)	`par_state_abbr`	categorical	The USPS Abbreviation for the Parent's State (or State-Like Entity)
Parent's County (FIPS code)	`par_county`	categorical	The FIPS code for the Parent's County
Parent's County Name	`par_county_name`	categorical	The Name of the Parent's County
Parent's Income Percentile	`par_pctile`	categorical	The National Income Percentile of the Parent. (`25`, `50`, `75`). Average is coded as (`-9`)
Child's Race	`kid_race`	categorical	The race of the child. (`AIAN` (American Indian and Alaska Native), `Asian`, `Black`, `Hispanic`, `White`)
Credit Score	`credit_score`	numeric	The average credit score in 2020, measured by Vantage 4.0
Student Loan Balance	`student_loan_balance`	numeric	The average balance of student loans held in 2020
Mortgage Balance	`mortgage_balance`	numeric	The average balance of mortgages held in 2020
Auto Loan Balance	`auto_loan_balance`	numeric	The average balance of auto loans held in 2020
Credit Card Balance	`credit_card_balance`	numeric	The average credit card balance in 2020
Debt Delinquency Rate	`debt_delinquency`	numeric	The rate of individuals with a 90+ day delinquency between 2016–2020, as a percentage
Debt Delinquency Controlling for Income	`debt_delinquency_income_controlled`	numeric	The average residual from a regression of `debt_delinquency` on 2016 household income rank

Key Variables:

- par_pctile (Parent's Income Percentile)
- kid_race (Child's Race)
- debt_delinquency (Debt Delinquency Rate)
- debt_delinquency_income_controlled
- credit_score

SUMMARY STATISTICS



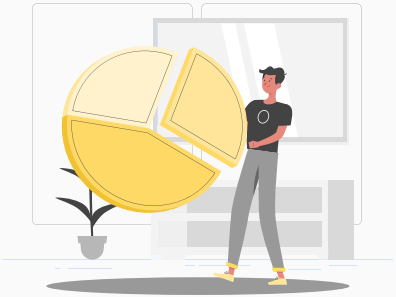


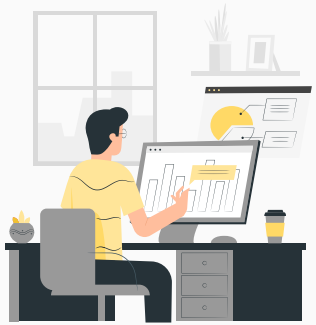
Left: the count of children in each racial identity group

Right: the count of observations in each parental income percentile group

Distributions of debt delinquency and community education levels

RESEARCH QUESTIONS





RESEARCH QUESTION 1

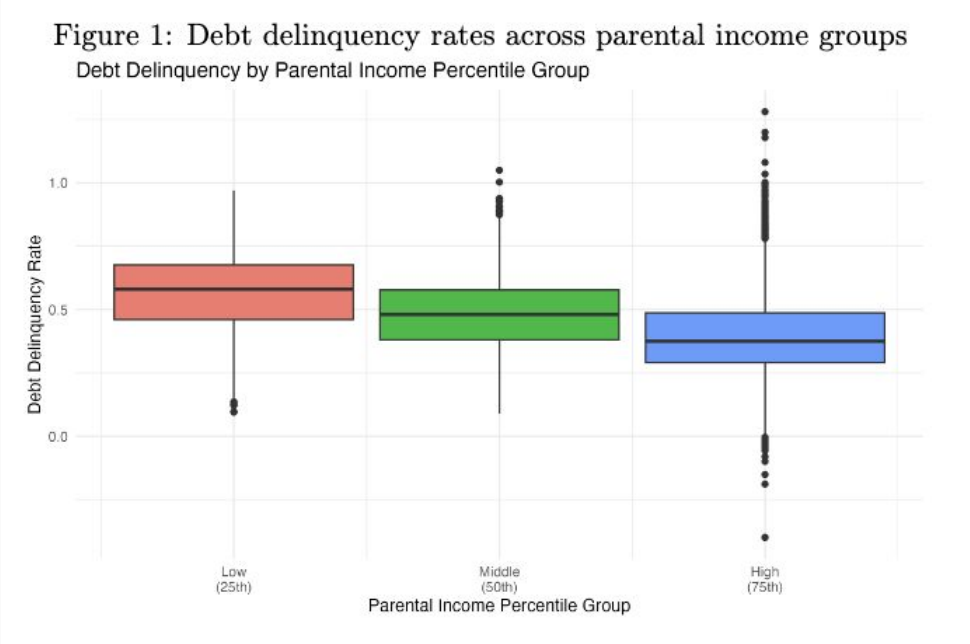
RQ1

Do children from different family incomes have different debt delinquency rates in adulthood?

Children from different parental income groups may experience different levels of debt delinquency in adulthood. This research question tests whether individuals from low-, middle-, and high-income families have significantly different 90+ day delinquency rates, in order to understand how childhood economic background influences adult financial distress.

RESEARCH QUESTION 1

(par_pctile → debt_delinquency)



The boxplot shows that debt delinquency rates decrease as parental income increases. Individuals from low-income families have the highest delinquency rates, middle-income groups fall in the middle, and those from high-income families have the lowest rates. This pattern suggests a clear negative relationship between childhood income level and adult financial delinquency.

Statistical Testing

A Kruskal–Wallis test was used to compare delinquency rates across income groups because the data are not normally distributed. The test results showed a highly significant difference ($p < 2.2e\text{-}16$), indicating that adult debt delinquency varies systematically by childhood income level.

Figure 2: Outcome of Kruskal–Wallis test

```
        Kruskal-Wallis rank sum test

data:  debt_delinquency by factor(par_pctile)
Kruskal-Wallis chi-squared = 7560.2, df = 2, p-value < 2.2e-16
```

Post-hoc Dunn Test

A post-hoc Dunn test confirmed that all three income groups differ significantly from one another. Individuals from the lowest income group have the highest adult debt delinquency, those from middle-income families have lower delinquency, and those from the highest-income families have the lowest.

Figure 3: Outcome of Post-hoc Dunn Test

Comparison <chr>	Z <dbl>	P.unadj <dbl>	P.adj <dbl>
25 - 50	42.33636	0	0
25 - 75	86.93968	0	0
50 - 75	44.60332	0	0

Result

These results reinforce a strong negative relationship between childhood family income and adult financial distress: growing up in a lower-income household is associated with a greater risk of debt delinquency later in life.

RQ2

Does socioeconomic environment during childhood affect debt delinquency rates in adulthood

It is well established that childhood environment plays an important role in economic outcomes later in life. This research question investigates how three indicators of socioeconomic environment (parental income, racial identity, and education rate) affect debt delinquency rate in adulthood. We also assess which of these environmental factors is most important for financial health in adulthood.

`(par_pctile, kid_race, college_4y → debt_delinquency)`

n.b. Excluding Puerto Rico

Model 1: Multivariate linear regression

We regress debt delinquency on each of the three measures of socioeconomic environment: parental income, race, and county education rate. The categorical variables are one-hot encoded, with the default level as the overall/pooled average.

$$\begin{aligned}\texttt{debt\_deliquency} = &\beta_0 + \beta_1(\texttt{par\_pctile}_{\text{25th}}) + \beta_2(\texttt{par\_pctile}_{\text{50th}}) + \beta_3(\texttt{par\_pctile}_{\text{75th}}) + \\ &\beta_4(\texttt{kid\_race}_{\text{AIAN}}) + \beta_5(\texttt{kid\_race}_{\text{Asian}}) + \beta_6(\texttt{kid\_race}_{\text{Black}}) + \\ &\beta_7(\texttt{kid\_race}_{\text{Hispanic}}) + \beta_8(\texttt{kid\_race}_{\text{White}}) + \beta_9(\texttt{college\_4y}) + \epsilon\end{aligned}$$

RESEARCH QUESTION 2

```
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      4.957e-01  1.380e-03 359.221  < 2e-16 ***
par_pctile25     7.613e-02  1.145e-03  66.498  < 2e-16 ***
par_pctile50    -6.999e-03  1.145e-03  -6.114 9.81e-10 ***
par_pctile75    -9.012e-02  1.145e-03 -78.725  < 2e-16 ***
kid_raceAIAN     1.452e-01  1.530e-03  94.901  < 2e-16 ***
kid_raceAsian   -1.225e-01  1.484e-03 -82.554  < 2e-16 ***
kid_raceBlack    2.462e-01  1.350e-03 182.460  < 2e-16 ***
kid_raceHispanic 7.919e-02  1.259e-03  62.883  < 2e-16 ***
kid_raceWhite   -3.604e-02  1.233e-03 -29.227  < 2e-16 ***
college_4y      -4.448e-03  7.161e-05 -62.122  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09819 on 58842 degrees of freedom
  (696 observations deleted due to missingness)
Multiple R-squared:  0.6337,    Adjusted R-squared:  0.6337
F-statistic: 1.131e+04 on 9 and 58842 DF,  p-value: < 2.2e-16
```

All coefficients are significant

63.37% variation in debt delinquency explained by childhood socioeconomic environment

Model 1: Coefficient interpretations (intercept)

$$\hat{\beta}_0 = 0.4957$$

Expected debt delinquency rate for **average** cohort with **0% education rate** is ***49.57%***

Model 1: Coefficient interpretations (parental income)

$$\hat{\beta}_1 = 0.07613$$

25th percentile parental income group has expected debt delinquency rate **7.61% higher**

$$\hat{\beta}_2 = -0.006999$$

50th percentile parental income group has expected debt delinquency rate **0.70% lower**

$$\hat{\beta}_3 = -0.09012$$

75th percentile parental income group has expected debt delinquency rate **9.01% lower**

Lower parental income → higher debt delinquency rates

RESEARCH QUESTION 2

Model 1: Coefficient interpretations (race)

$$\hat{\beta}_7 = 0.07919$$

Hispanic cohorts have expected debt delinquency rate **7.92% higher**

$$\hat{\beta}_8 = -0.03604$$

White cohorts have expected debt delinquency rate **3.60% lower**

Different racial identities *do* affect debt delinquency rates, in different ways

RESEARCH QUESTION 2

Model 1: Coefficient interpretations (education)

$$\hat{\beta}_9 = -0.004448$$

1 percentage point increase in 4y county college education rate associated with **0.44% lower** debt delinquency rates

More educated communities have lower debt delinquency rates

Model 1: Findings

- Childhood socioeconomic environment ***does*** affect adult debt delinquency rates
  - Higher parental income, higher education rates → less debt delinquency
  - Race also significant factor

Model 2: Random forest

We fit a random forest on the same variables: we predict debt delinquency from parental income percentile group, race, and education rate.

- Able to address non-linear data; no model assumptions
- Greater predictive power; variable importance policy implications

RESEARCH QUESTION 2

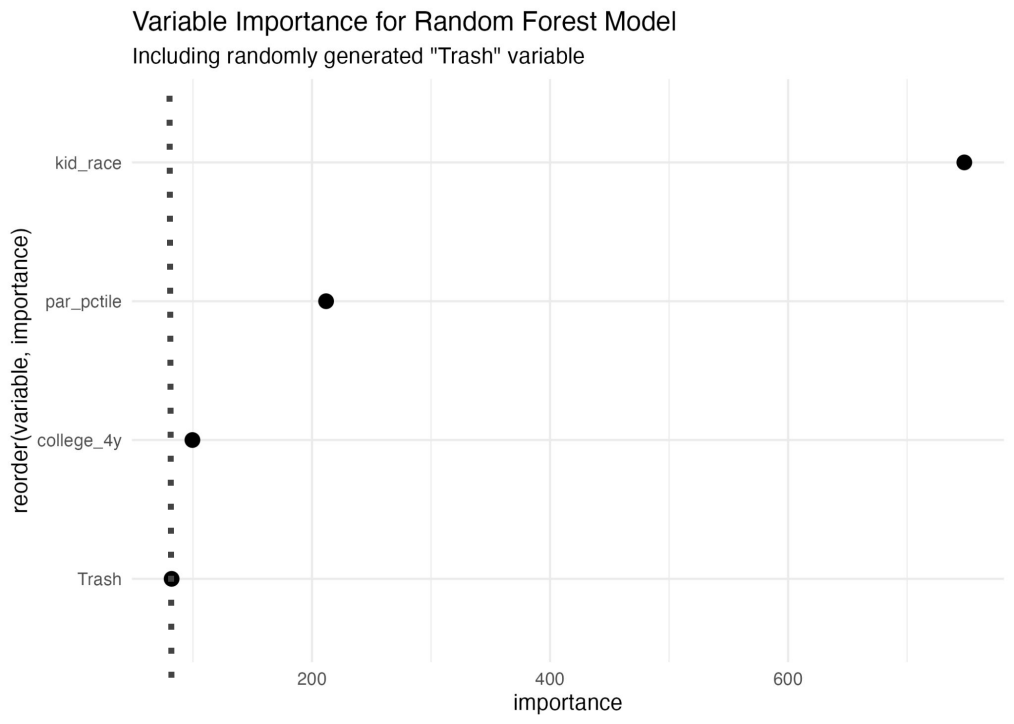
```
Call:
 randomForest(formula = debt_delinquency ~ par_pctile + kid_race +
college_4y, data = creditaccess_cty_clean, mtry = 2, importance = TRUE,
na.action = na.roughfix)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 2

          Mean of squared residuals: 0.009209166
                    % Var explained: 64.98
```

imputation

R^2

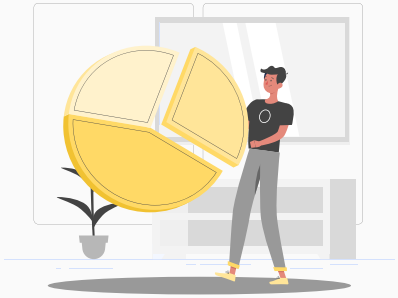
RESEARCH QUESTION 2

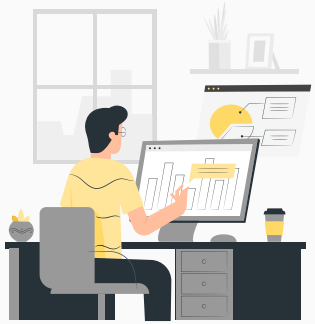


➔ **Race is most important predictor**

➔ **Followed by parental income percentile**

➔ **More important than random noise**





CONCLUSIONS, IMPLICATIONS, NEXT STEPS

Conclusions

- Negative relationship between parental income and debt delinquency
  - Children have better financial outcomes when parents are wealthier
- Adulthood debt delinquency rates can be partially predicted by socioeconomic environment during childhood
  - By parental income, race, and local education rates
  - Race, then income, appear to be most important predictors

Implications

- Financial health is strongly associated with environmental factors during childhood...
- Therefore, fiscal welfare policies that address financial disparities across racial identities
  - Recognize systemic inequities in credit system
  - Targeted support for underserved communities
- Financial outcomes from non-financial inputs
  - Environmental effects

Next steps

- Broader measure of socioeconomic environment (unemployment rate, diversity, distressed communities index, etc.
- Individual level data
  - Survey, or longitudinal study?
- Test effectiveness of targeted fiscal education programs

THANK YOU!