

# Studying Social Inequality with Data Science

INFO 3370 / 5371  
Spring 2023

**Challenge: PSID Income Prediction**

# Learning goals for today

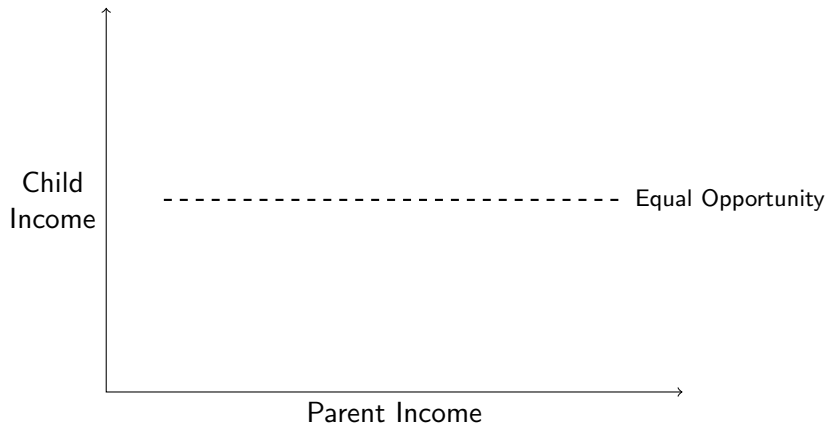
By the end of class, you will be able to

- ▶ conceptualize economic mobility in a machine learning framework

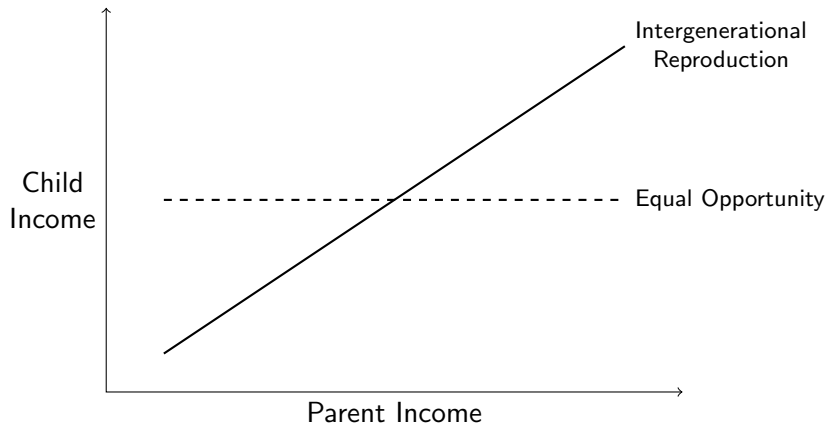
# Generalizing economic mobility



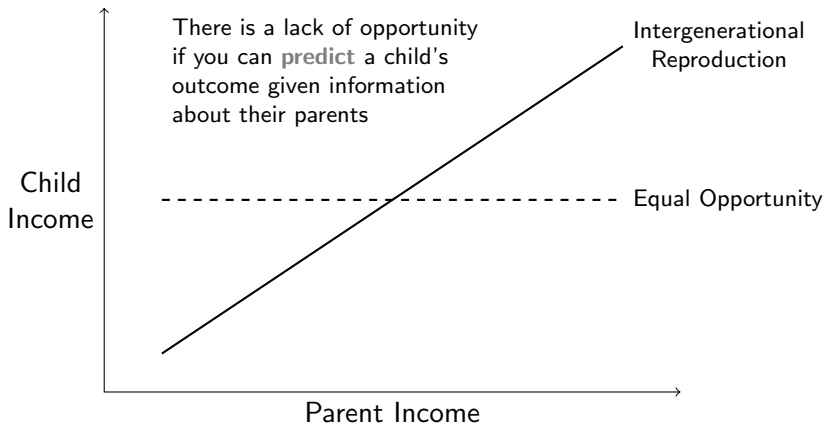
# Generalizing economic mobility



# Generalizing economic mobility



# Generalizing economic mobility



Why stop at parent income?

Why stop at parent income?

What other variables capture the luck of birth?



	Respondent Education	Parent Education	Grandparent Education	Sex	Race	Grandparent Income	Parent Income
Case 1							
Case 2							
Case 3							
Case 4							
Case 5							

Respondent Income

--	--	--	--	--

Respondent Education

Parent Education

Grandparent Education

Sex

Race

Grandparent Income


Parent Income

Case 1								
Case 2								
Case 3								
Case 4								
Case 5								

Respondent Income

--	--	--	--	--

Learn a  
prediction  
function



Respondent Education

Parent Education

Grandparent Education

Sex

Race

Grandparent Income

Parent Income


Case 1

Case 2

Case 3

Case 4

Case 5

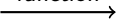
Respondent Income


?

?

?

Learn a  
prediction  
function



Respondent Education

Parent Education

Grandparent Education

Sex

Race

Grandparent Income

Parent Income

Case 1

Case 2

Case 3

Case 4

Case 5

Case 6

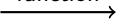
Case 7

Case 8

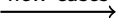


Respondent Income


Learn a  
prediction  
function



Predict for  
new cases



?
?
?

Respondent Education

Parent Education

Grandparent Education

Sex

Race

Grandparent Income

Parent Income

Case 1

Case 2

Case 3

Case 4

Case 5


Case 6

Case 7

Case 8


Respondent Income


Learn a  
prediction  
function  
→

Predict for  
new cases  
→

?
?
?

Respondent Education

Parent Education

Grandparent Education

Sex

Race

Grandparent Income

Parent Income

Case 1  
Case 2  
Case 3  
Case 4  
Case 5


Case 6  
Case 7  
Case 8


Learning Set

Holdout Set

# The challenge

How well can you predict respondent incomes?

Get started [here](#)

# Learning goals for today

By the end of class, you will be able to

- ▶ conceptualize economic mobility in a machine learning framework