

Using Big Data To Solve Economic and Social Problems

Raj Chetty

Photo Credit: Florida Atlantic University

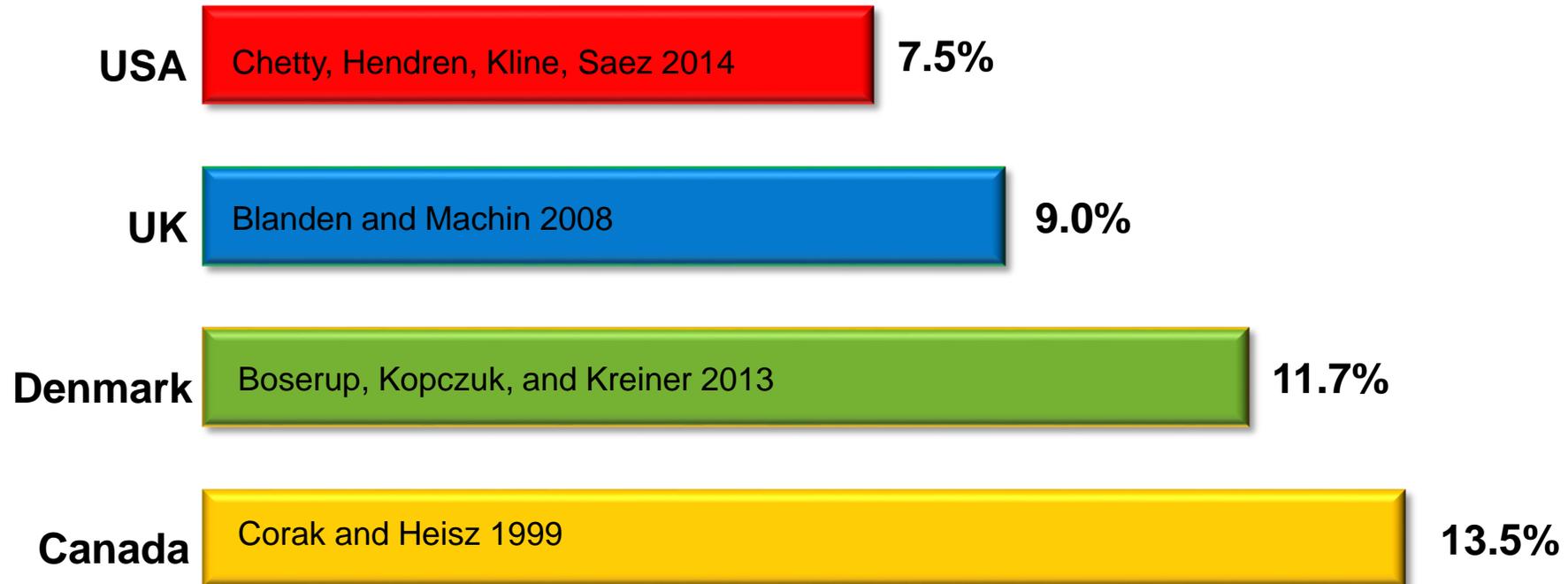


The American Dream?

- Chance that a child born to parents in the bottom fifth of the income distribution reaches the top fifth:

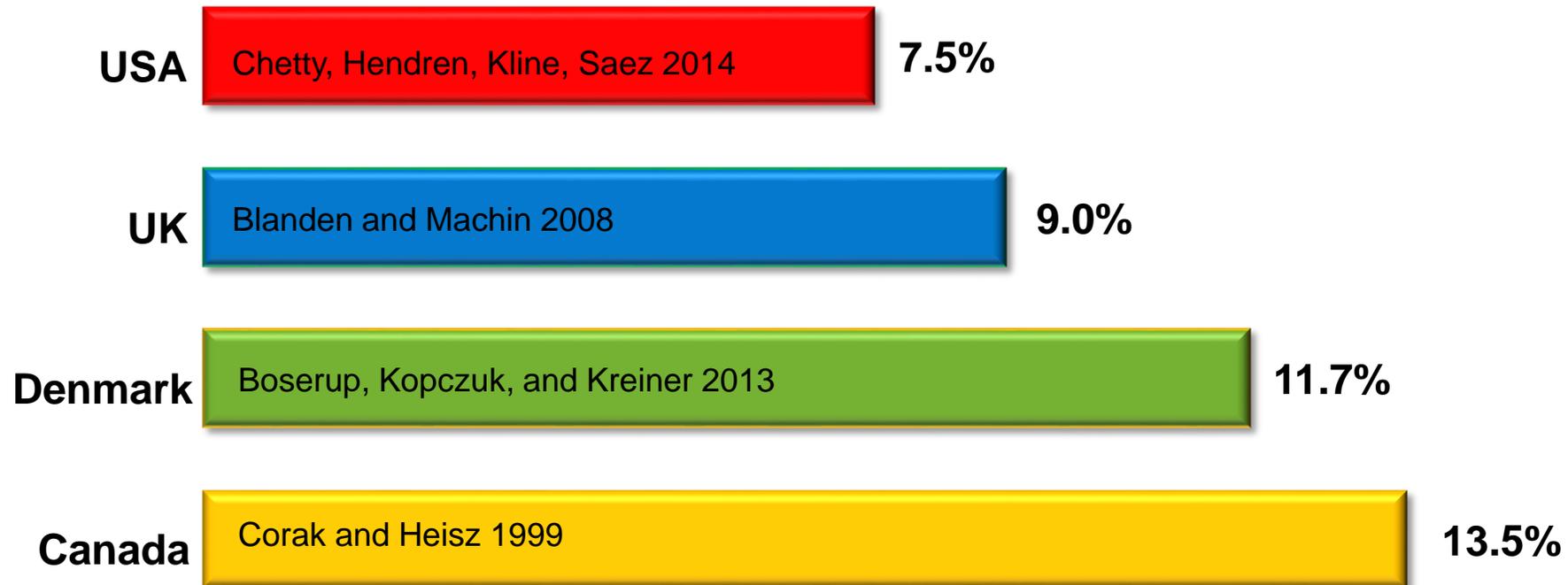
The American Dream?

- Chance that a child born to parents in the bottom fifth of the income distribution reaches the top fifth:



The American Dream?

- Chance that a child born to parents in the bottom fifth of the income distribution reaches the top fifth:



→ Chances of achieving the “American Dream” are almost two times higher in Canada than in the U.S.

Why is Upward Mobility Lower in America?

- Central policy question: why are children's chances of escaping poverty so low in America?
 - And what can we do to improve their odds...?
- Difficult to answer this question based solely on country-level data
 - Numerous differences across countries makes it hard to test between alternative explanations
 - Problem: only a handful of data points

Theoretical Social Science

- Until recently, social scientists have had limited data to study policy questions like this
- Social science has therefore been a *theoretical* field
 - Develop mathematical models (economics) or qualitative theories (sociology)
 - Use these theories to explain patterns and make policy recommendations, e.g. to improve upward mobility

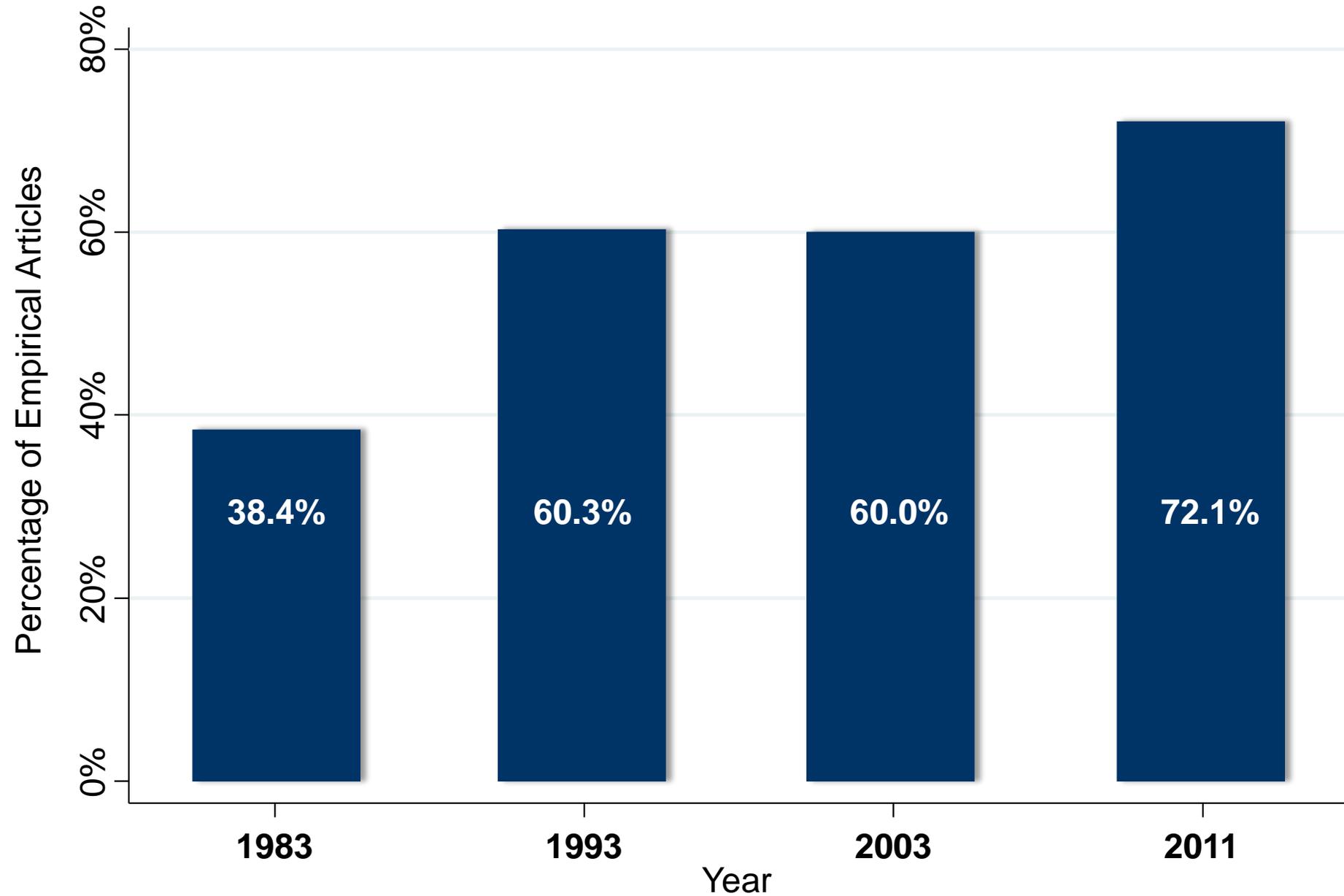
Theoretical Social Science

- Problem: theories untested → five economists often have five different answers to a given question
- Leads to a politicization of questions that in principle have scientific answers
 - Example: is Obamacare reducing job growth in America?

The Rise of Data and Empirical Evidence

- Today, social science is becoming a more empirical field thanks to the growing availability of data
 - Test and improve theories using real-world data
 - Analogous to natural sciences

Empirical (Data-Based) Articles in Leading Economics Journals, 1983-2011



Source: Hamermesh (JEL 2013)

Social Science in the Age of Big Data

- Recent availability of “big data” has accelerated this trend
 - Large datasets are starting to transform social science, as they have transformed business
- Examples:
 - Government data: tax records, Medicare
 - Corporate data: Facebook, retailer data
 - Unstructured data: Twitter, newspapers

Why is Big Data Transforming Social Science?

1. Greater reliability than surveys
2. Ability to measure new variables (e.g., emotions)
3. Universal coverage → can “zoom in” to subgroups
4. Large samples → can approximate scientific experiments

Why This Course?

- Silicon Valley has been very successful in solving private-sector problems using technology and big data
- Goal of this course: show how same skills can be used to address important social and economic problems
 - We need more talent in this area given pressing challenges such as rising inequality and global warming
- To achieve this goal, provide an introduction to a broad range of topics, methods, and real-world applications

Overview of Topics

1. Equality of Opportunity
2. Education
3. Health
4. Environment
5. Criminal Justice and Discrimination
6. Political Polarization

Overview of Methods

1. Descriptive Data Analysis
2. Experiments
3. Quasi-Experiments
4. Machine Learning
5. Stata programming

Methods: Two Types of “Big Data”

- Big data can be classified into two types
 - “Long” data: many observations relative to variables (e.g., tax records)

Clipboard: Paste, Cut, Copy, Format Painter

Font: Calibri, 11, Bold, Italic, Underline, Text Color, Background Color, Font Color

Alignment: Wrap Text, Merge & Center, Horizontal, Vertical

Number: General, Currency, Percentage, Decimals, Thousands Separator

Styles: Normal, Bad, Good, Neutral, Calculation, Check Cell, Explanatory..., Input, Linked Cell, Note

Cells: Insert, Delete, Format

Editing: AutoSum, Fill, Clear, Sort & Filter, Find & Select

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	person_id	income	years of education	gender																							
2	101	\$ 8,825.23	12	F																							
3	102	\$38,356.11	14	M																							
4	103	\$ 8,641.73	13	F																							
5	104	\$10,024.09	13	M																							
6	105	\$79,923.36	12	M																							
7	106	\$57,007.00	14	M																							
8	107	\$59,494.84	15	F																							
9	108	\$92,150.41	13	M																							
10	109	\$75,373.30	13	F																							
11	110	\$15,680.30	13	M																							
12	111	\$46,593.41	13	F																							
13	112	\$71,386.71	15	M																							
14	113	\$72,674.96	11	M																							
15	114	\$58,535.12	12	M																							
16	115	\$11,968.91	12	F																							
17	116	\$99,265.27	14	M																							
18	117	\$46,181.11	11	F																							
19	118	\$74,175.59	15	M																							
20	119	\$73,409.86	11	F																							
21	120	\$65,784.26	14	M																							
22	121	\$ 3,532.26	14	M																							
23	122	\$33,836.95	15	M																							
24	123	\$56,806.58	13	F																							
25	124	\$68,478.31	13	M																							
26	125	\$60,566.22	15	F																							
27	126	\$98,447.41	13	F																							
28	127	\$79,397.90	11	F																							
29	128	\$17,594.75	12	F																							
30	129	\$84,667.93	13	M																							
31	130	\$87,953.71	13	M																							
32	131	\$68,423.74	14	F																							
33	132	\$51,357.62	13	M																							
34	133	\$82,233.86	12	F																							
35	134	\$92,901.91	14	M																							
36	135	\$75,153.35	13	M																							
37	136	\$29,740.94	15	M																							
38	137	\$ 795.36	13	F																							
39	138	\$27,283.46	12	M																							
40	139	\$ 1,137.37	12	F																							
41	140	\$61,127.80	13	M																							
42	141	\$33,153.06	12	F																							
43	142	\$19,774.73	15	M																							
44	143	\$55,925.97	13	M																							
45	144	\$75,598.81	15	M																							

Methods: Two Types of “Big Data”

- Big data can be classified into two types
 - “Long” data: many observations relative to variables (e.g., tax records)
 - “Wide” data: few observations relative to variables (e.g. Amazon clicks, newspapers)

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	years of education	gender	ad_click1	ad_click2	ad_click3	ad_click4	ad_click5	ad_click6	ad_click7	ad_click8	ad_click9	ad_click10	ad_click11	ad_click12	ad_click13	ad_click14	ad_click15	ad_click16	ad_click17	ad_click18	ad_click19	ad_click20	ad_click21	ad_click22	ad_click23	ad_click24	ad_click25
2	12	F	0	1	1	1	0	1	0	0	0	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	
3	14	M	0	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0
4	12	F	0	0	1	0	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	0	1	0	1	1	1
5	12	M	1	0	0	0	0	0	1	1	0	1	1	0	1	1	0	1	0	1	0	0	1	1	0	1	1
6	12	M	0	0	0	0	0	0	1	1	1	0	1	0	1	0	0	1	1	0	0	1	0	1	1	1	0
7	14	M	0	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1	1
8	11	F	1	1	0	1	0	1	0	1	0	1	1	1	1	0	0	0	0	0	1	1	0	0	0	1	0
9	15	M	1	0	0	1	1	1	0	1	1	1	0	1	1	1	0	0	1	1	0	1	1	1	0	1	0
10	14	F	1	1	0	1	0	1	1	0	0	1	1	0	1	0	1	1	1	0	0	1	1	1	1	0	1
11	15	M	0	0	1	0	1	0	1	1	0	1	0	0	0	1	0	0	1	1	1	0	1	0	1	1	1
12																											
13																											
14																											
15																											
16																											
17																											
18																											
19																											
20																											
21																											
22																											
23																											
24																											
25																											
26																											
27																											
28																											
29																											
30																											
31																											
32																											
33																											
34																											
35																											
36																											
37																											
38																											
39																											
40																											
41																											
42																											
43																											
44																											
45																											

Methods: Two Types of “Big Data”

- Statistics/computer science has focused on “wide” data
 - Main application: *prediction*
 - Example: predicting income to target ads
- Social science has focused on “long” data
 - Main application: *identifying causal effects*
 - Example: effects of improving schools on income

Lecture 1: Equality of Opportunity

1. Local Area Differences in Upward Mobility within America
2. Geographical Variation: Causal Effects of Places or Sorting?
3. Characteristics of Low vs. High Mobility Areas

- Lecture 1 is based primarily on two papers:

Chetty, Hendren, Kline, Saez. “Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the U.S.” QJE 2014

Chetty and Hendren. “The Effects of Neighborhoods on Children’s Long-Term Outcomes” 2017a, b

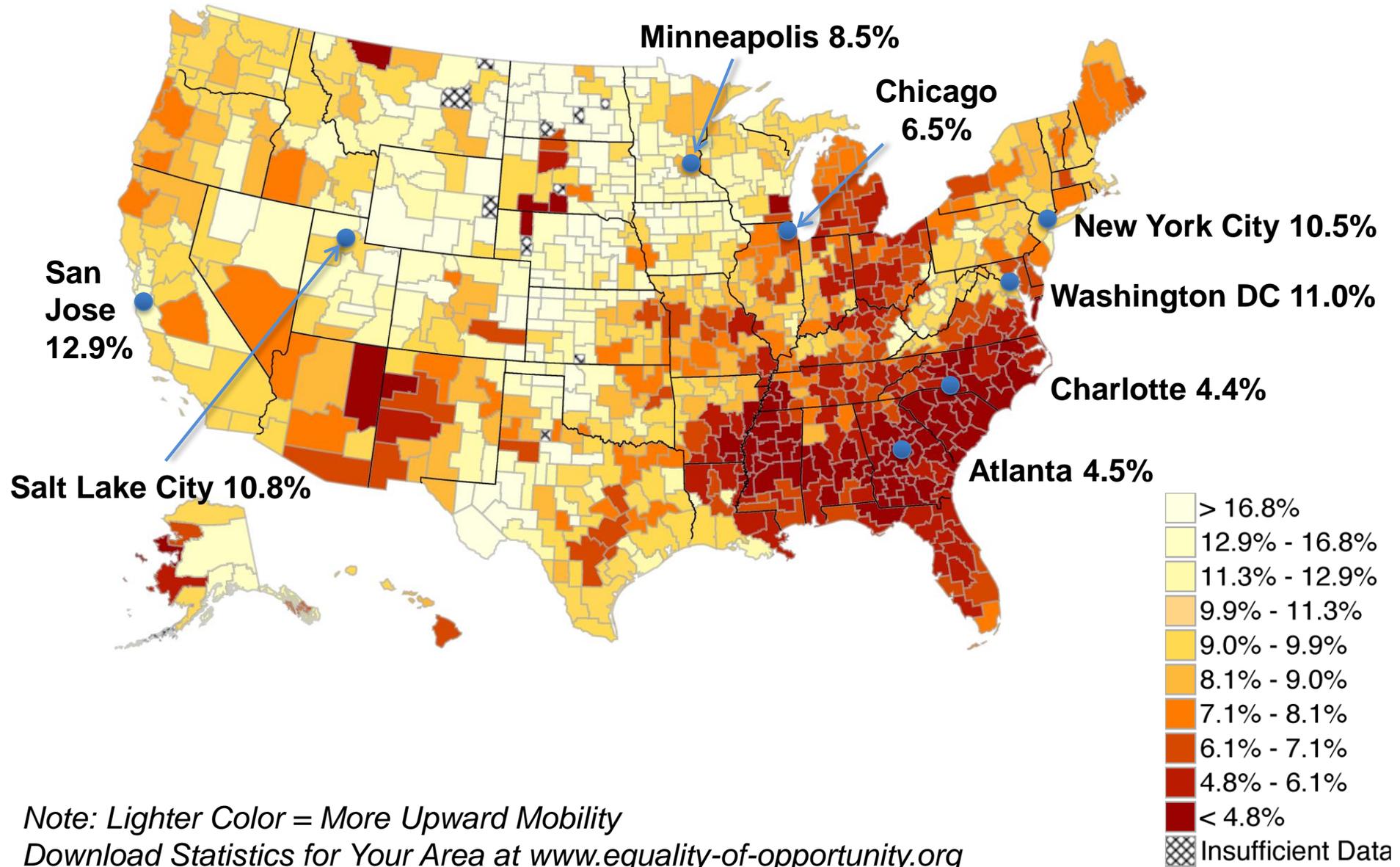
Part 1
Local Area Variation

Differences in Opportunity Across Local Areas

- Chetty et al. (2014) use “big data” to measure upward mobility for every metro and rural area in the U.S.
 - De-identified tax records on all children born in America between 1980-1982 (10 million children)
- Classify children into locations based on where they grew up
- Rank children in *national* income distribution (not local distribution) when computing rates of upward mobility

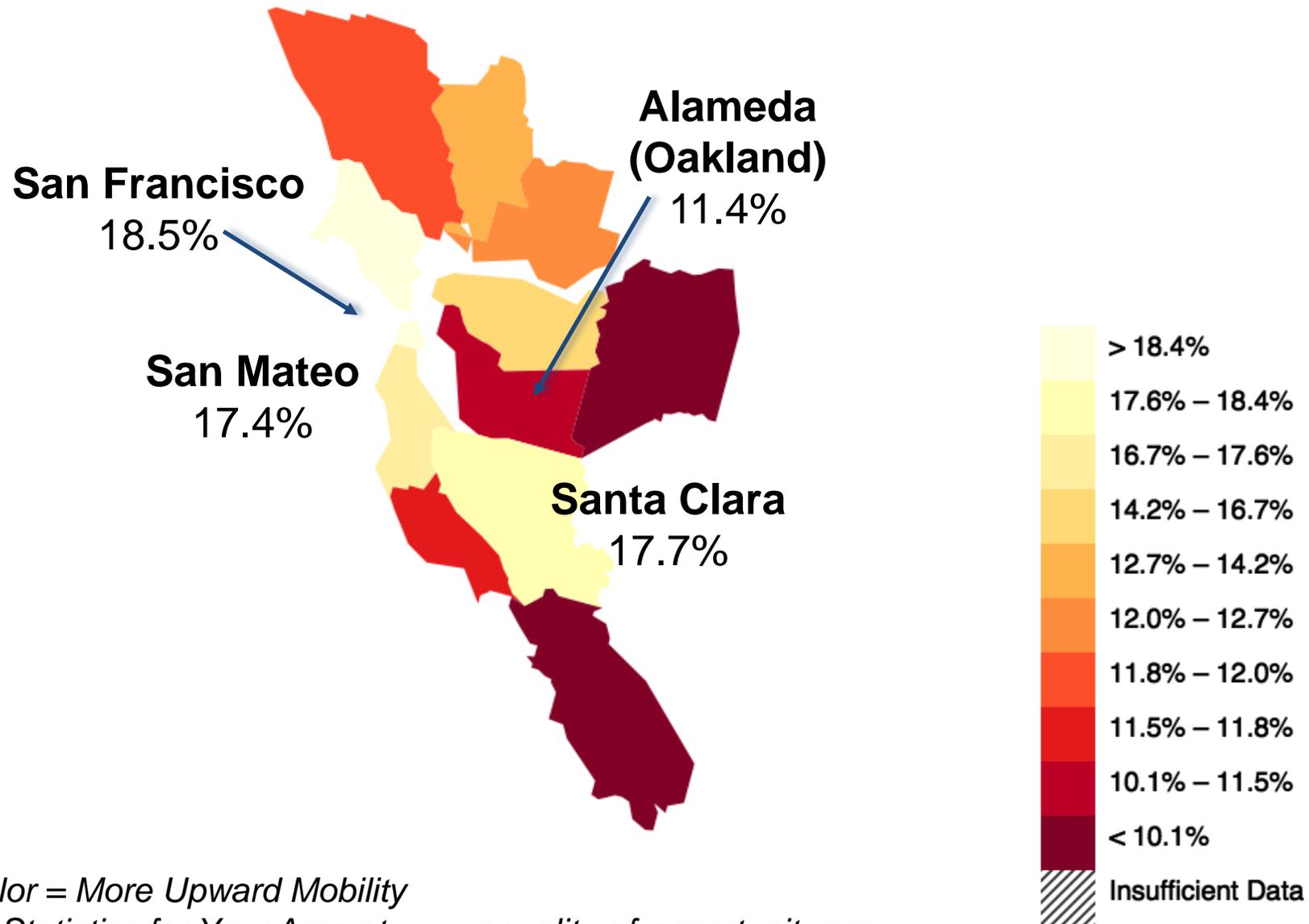
The Geography of Upward Mobility in the United States

Chances of Reaching the Top Fifth Starting from the Bottom Fifth by Metro Area



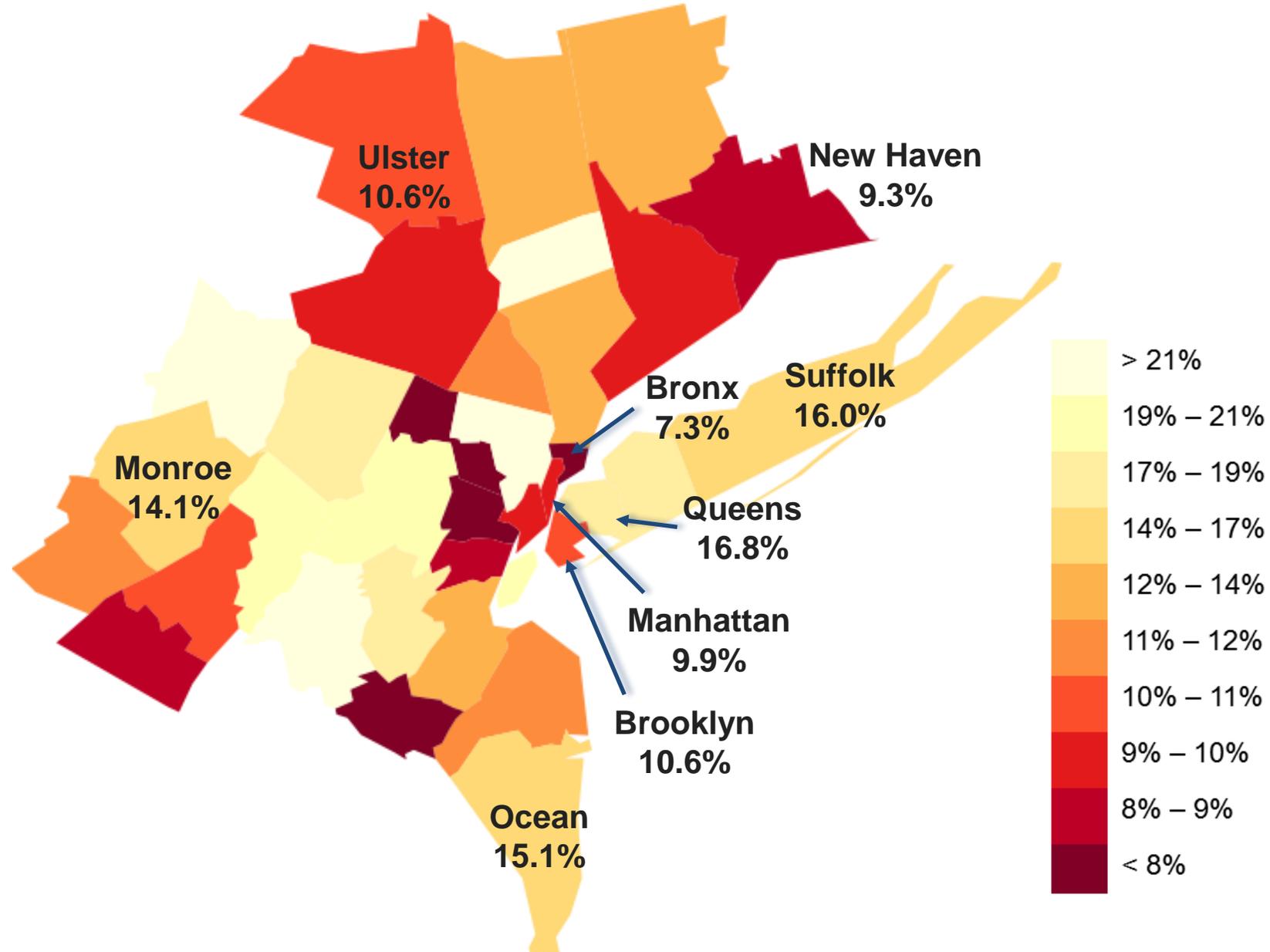
The Geography of Upward Mobility in the Bay Area

Chances of Reaching the Top Fifth Starting from the Bottom Fifth by County



The Geography of Upward Mobility in the New York Area

Chances of Reaching the Top Fifth Starting from the Bottom Fifth by County



Part 2
Causal Effects of Neighborhoods

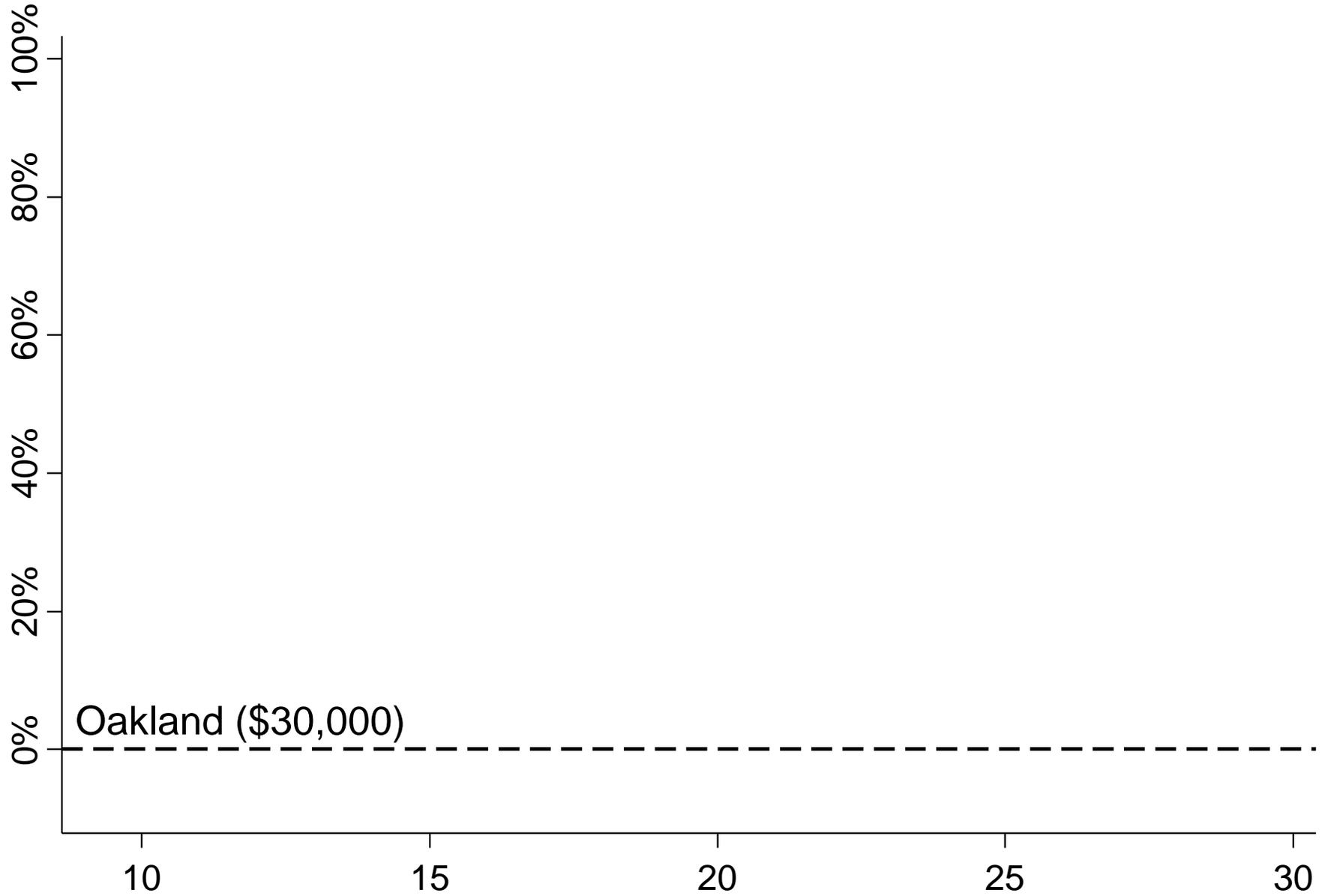
Causal Effects of Neighborhoods vs. Sorting

- Two very different explanations for variation in children's outcomes across areas:
 1. Sorting: different people live in different places
 2. Causal effects: places have a *causal* effect on upward mobility for a given person

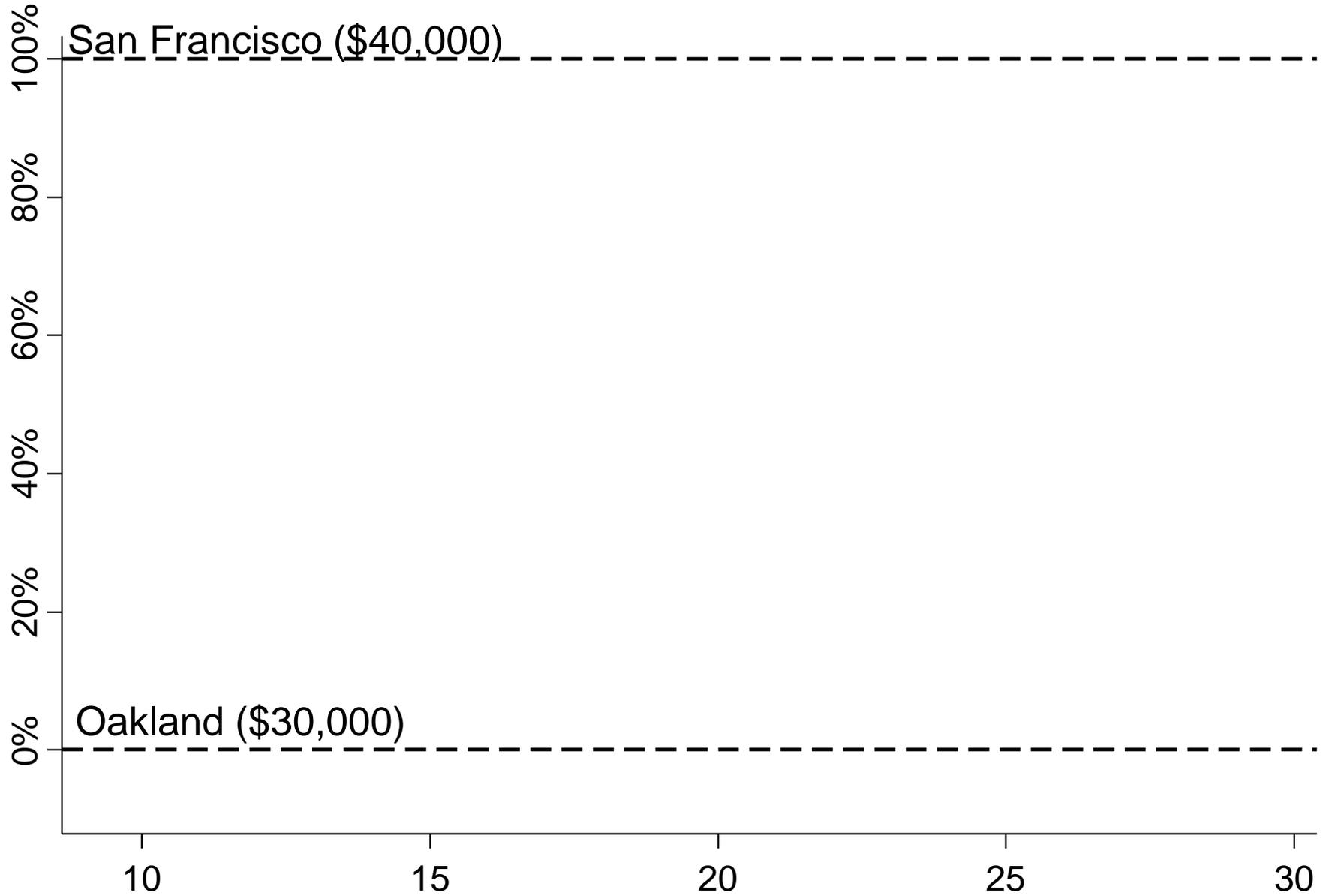
Identifying Causal Effects of Neighborhoods

- Ideal experiment: randomly assign children to neighborhoods and compare outcomes in adulthood
- We approximate this experiment using a quasi-experimental design
 - Study 7 million families who move across counties in observational data
 - Key idea: exploit variation in *age of child* when family moves to identify causal effects of environment

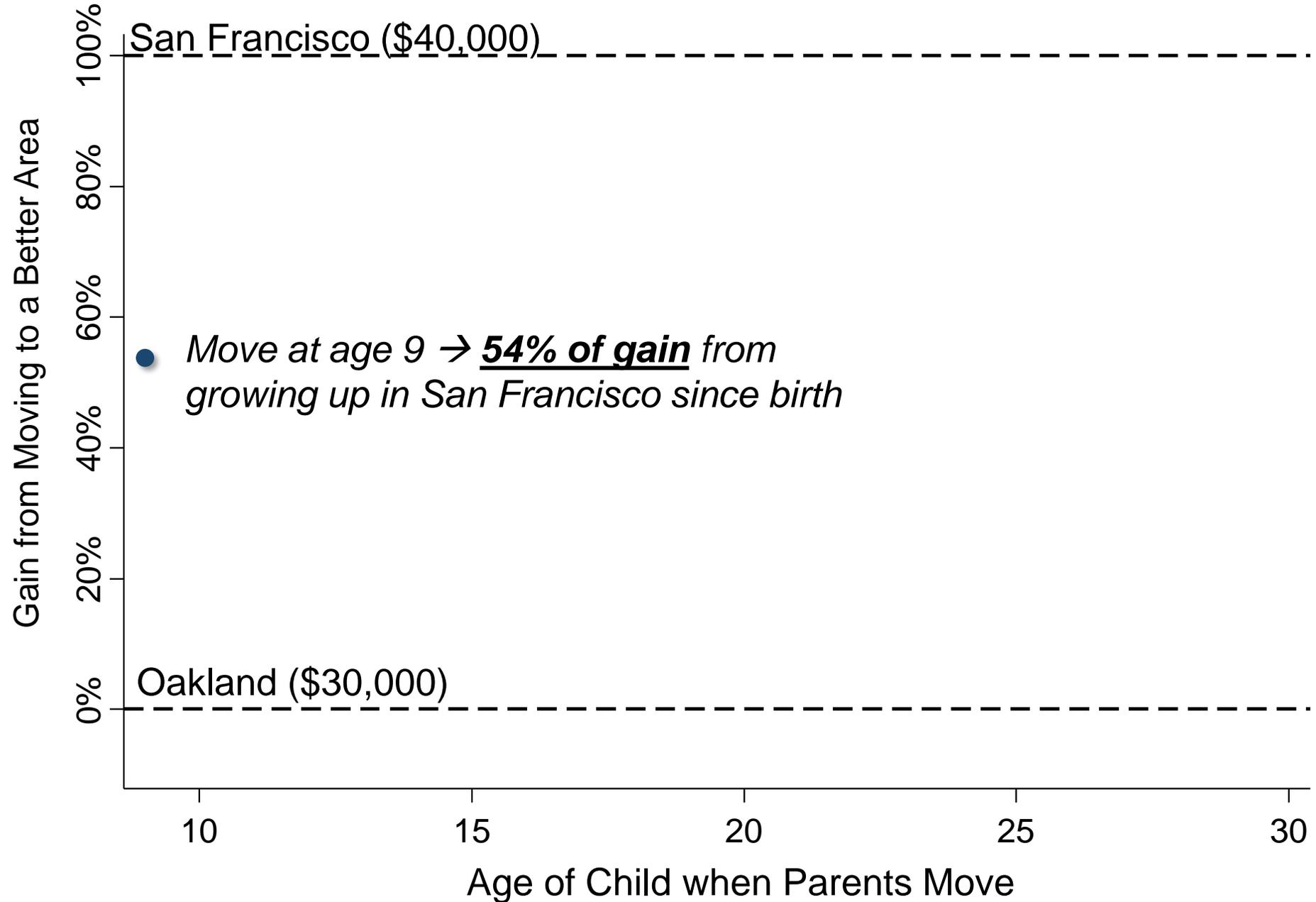
Earnings Gain from Moving to a Better Neighborhood



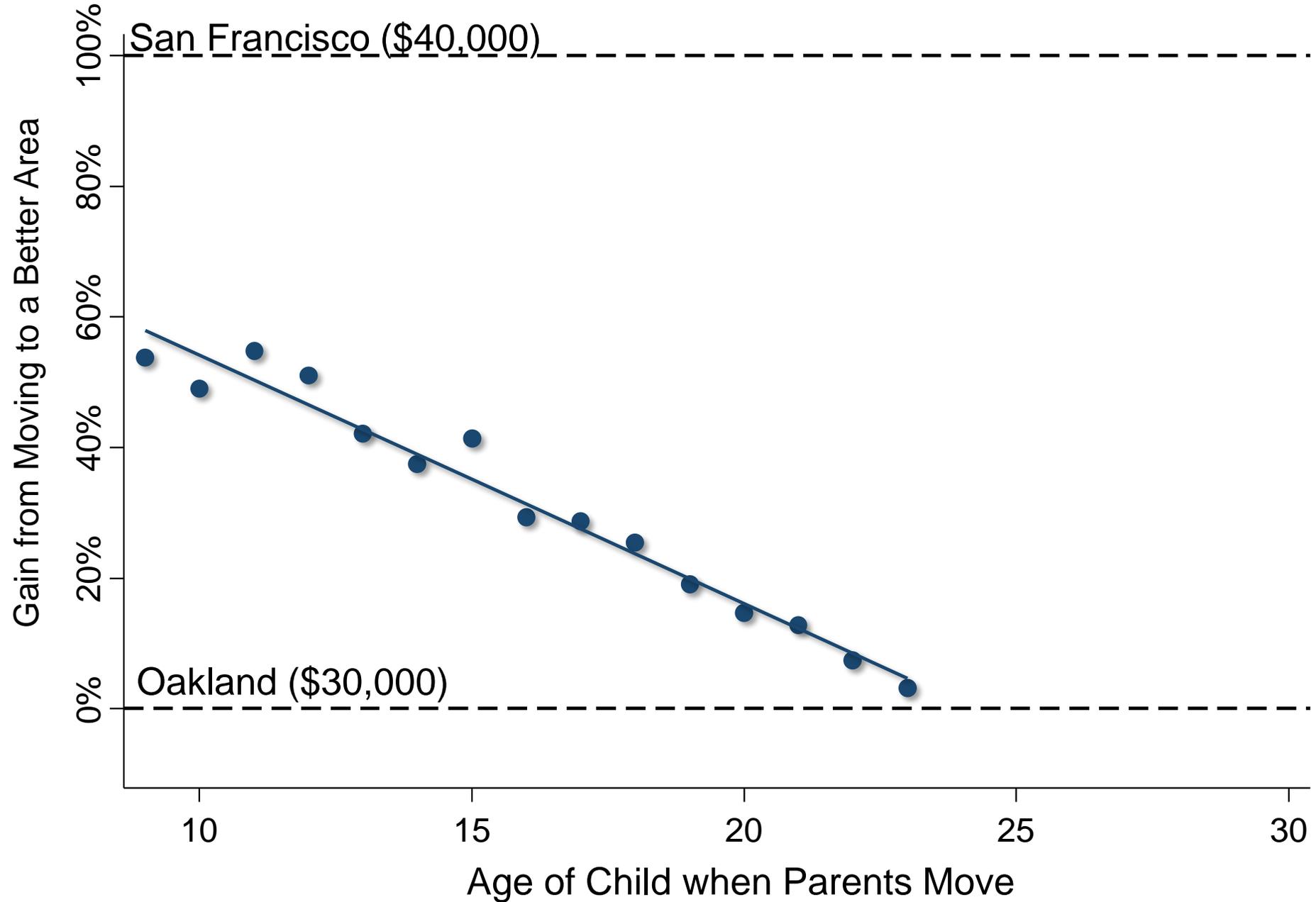
Earnings Gain from Moving to a Better Neighborhood



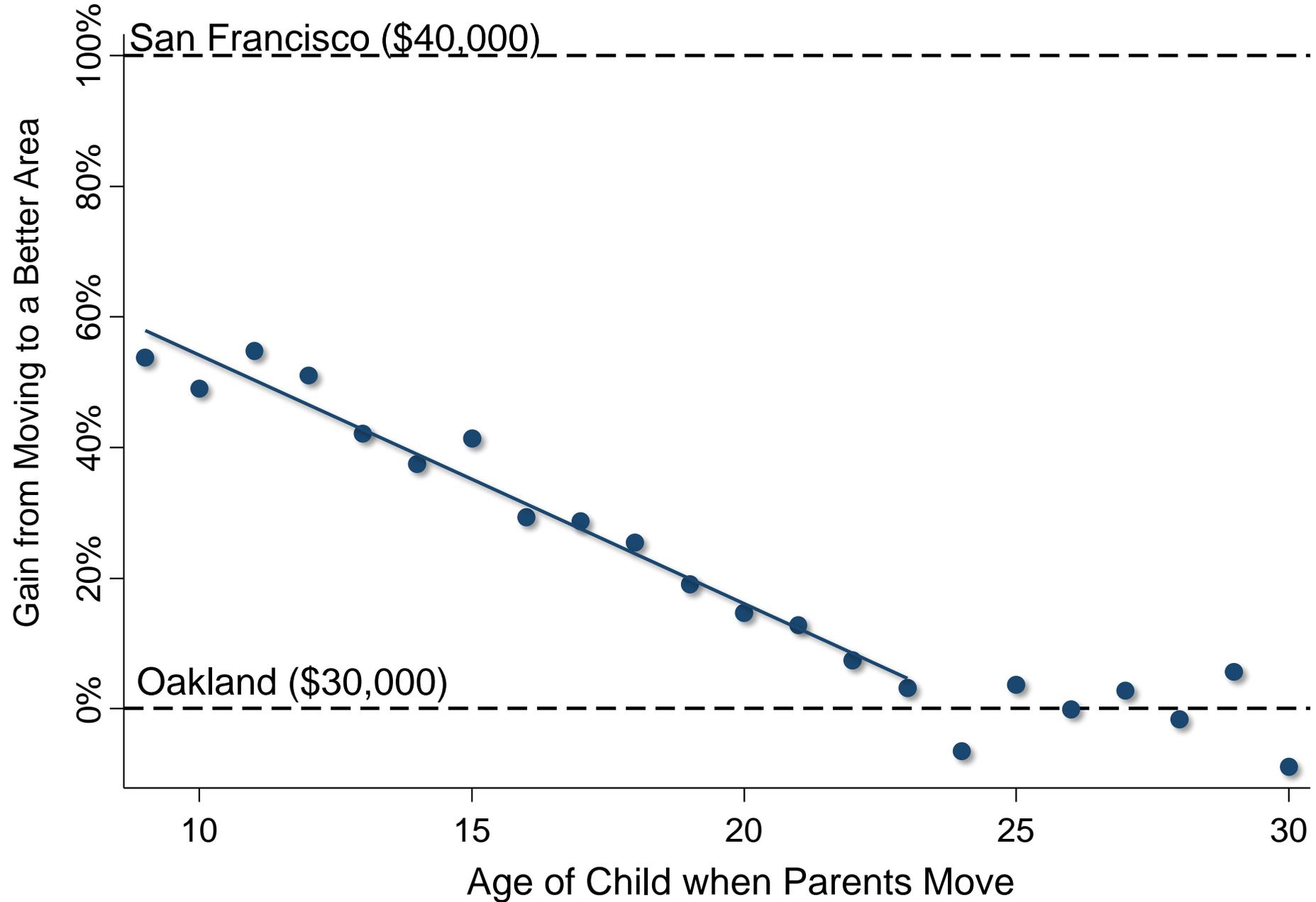
Earnings Gain from Moving to a Better Neighborhood



Earnings Gain from Moving to a Better Neighborhood



Earnings Gain from Moving to a Better Neighborhood



Identifying Causal Effects of Neighborhoods

- Key assumption: *timing* of moves to a better/worse area unrelated to other determinants of child's outcomes
- This assumption might not hold for two reasons:
 1. Parents who move to good areas when their children are young might be different from those who move later
 2. Moving may be related to other factors (e.g., change in parents' job) that affect children directly

Identifying Causal Effects of Neighborhoods

- Two approaches to evaluating validity of this assumption:
 1. Compare siblings' outcomes to control for family effects

Identifying Causal Effects of Neighborhoods

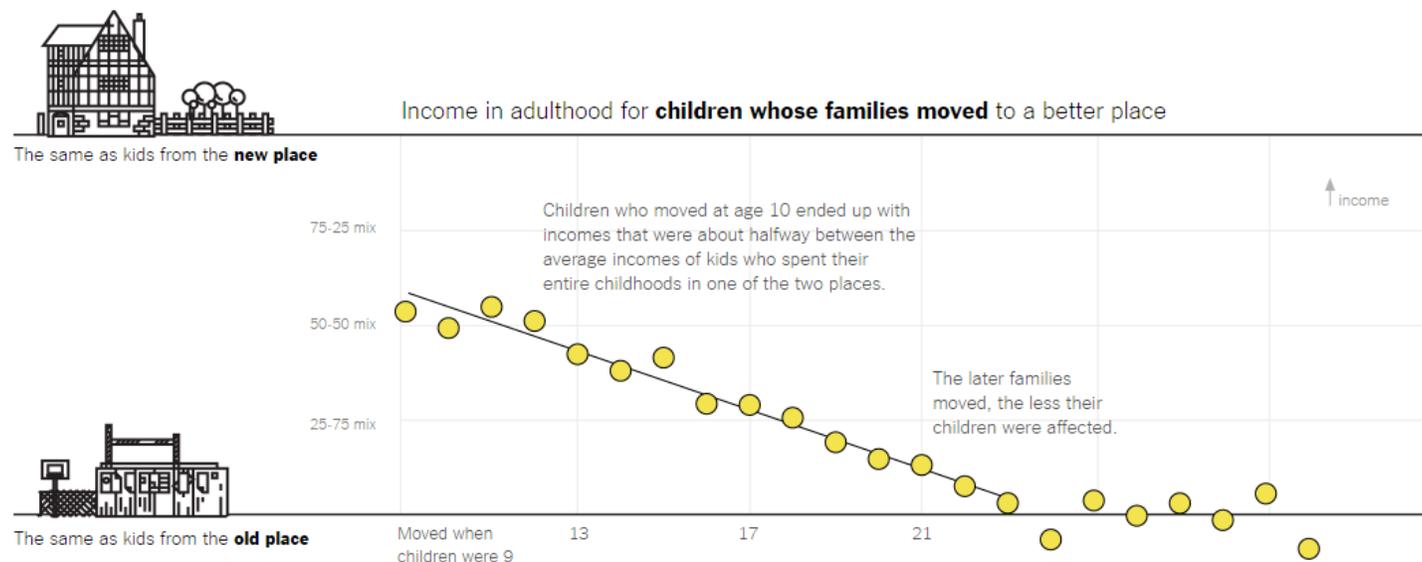
- Two approaches to evaluating validity of this assumption:
 1. Compare siblings' outcomes to control for family effects
 2. Use differences in neighborhood effects across subgroups to implement “placebo” tests
 - Ex: some places (e.g., low-crime areas) have better outcomes for boys than girls
 - Move to a place where boys have high earnings → son improves in proportion to exposure but daughter does not

Causal Effects of Neighborhoods: Summary

- Key lesson of this section: 70-80% of the variation in children's outcomes across areas is due to *causal effects*
- This result has refocused public discussion on improving upward mobility in America to a local level

An Atlas of Upward Mobility Shows Paths Out of Poverty

By DAVID LEONHARDT, AMANDA COX and CLAIRE CAIN MILLER MAY 4, 2015

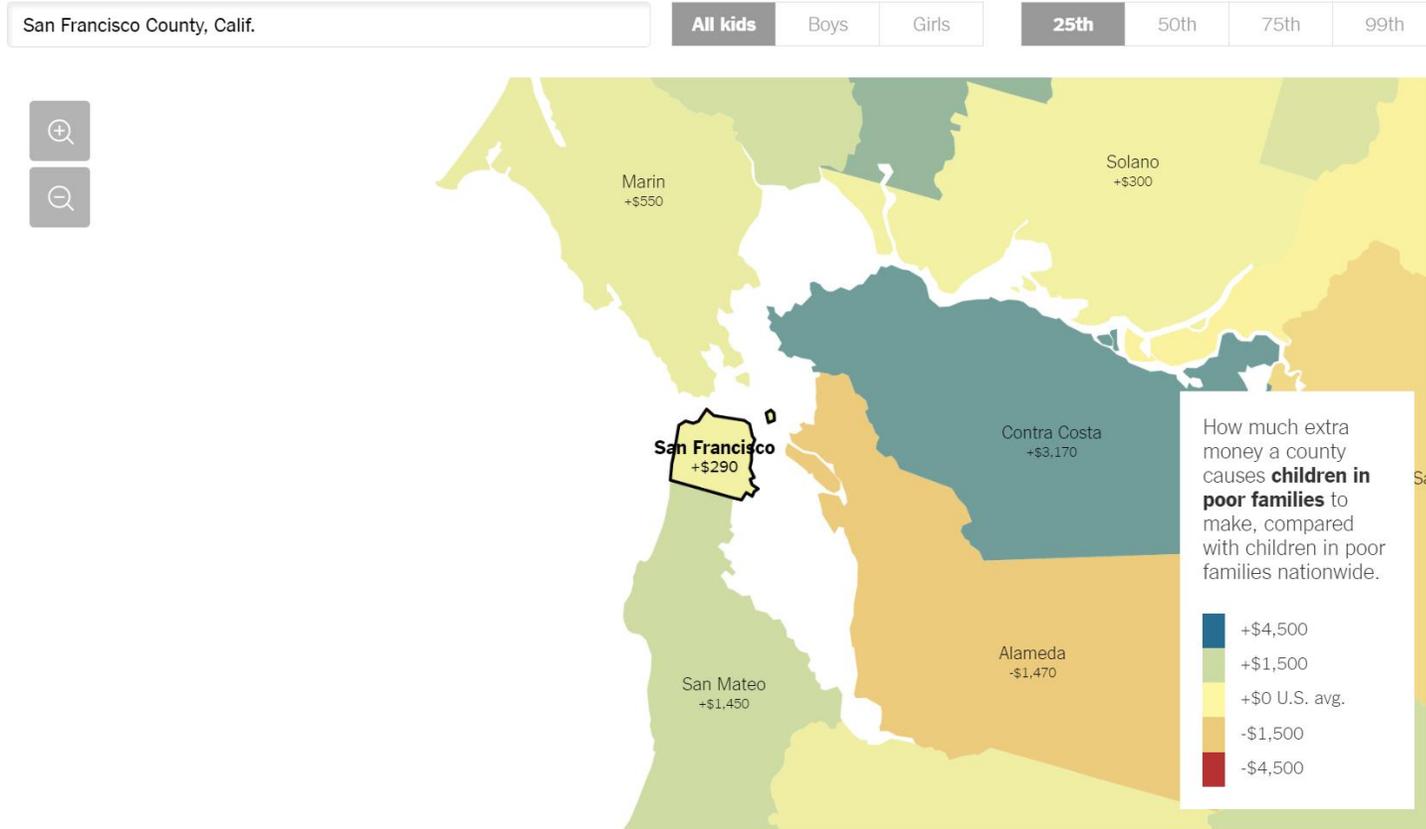


In the wake of the Los Angeles riots more than 20 years ago, Congress created an anti-poverty experiment called [Moving to Opportunity](#). It gave vouchers to help poor families move to better neighborhoods and awarded them on a random basis, so researchers could study the effects.

The results were deeply disappointing. Parents who received the vouchers did not seem to earn more in later years than otherwise similar adults, and children did not seem to do better in school. The program's apparent failure has haunted social scientists and policy makers, making poverty seem all

The Best and Worst Places to Grow Up: How Your Area Compares

Children who grow up in some places go on to earn much more than they would if they grew up elsewhere. MAY 4, 2015 | [RELATED ARTICLE](#)



San Francisco County is about average for income mobility for children in poor families. It is better than about 42 percent of counties.

[Location matters](#) – enormously. If you're poor and live in the San Francisco area, it's better to be in Contra Costa County than in San Francisco County or Alameda County. Not only that, the younger you are when you move to Contra Costa, the better you will do on average. Children who move at earlier ages are less likely to become single parents, more likely to go to college and more likely to earn more.



CHAPTER 1
Context

- A Wake-Up Call for Charlotte-Mecklenburg
- The Study That Caught Our Attention
- A Note About The American Dream
- The Charlotte-Mecklenburg Opportunity Task Force & Our Work
- Path Forward: Conclusions & Recommendations

A Wake-Up Call for Charlotte-Mecklenburg

Land of opportunity? Not by a long shot

Charlotte is nation's worst big city for climbing out of poverty

The Charlotte Observer

Over the last several decades, Charlotte-Mecklenburg has transformed from a small southern town to one of the country's largest and most dynamic communities. We continue to attract people—nearly 50 a day— who move here to take advantage of our strong business climate, favorable weather and

geographic location, and our reputation as a great place to live and raise a family. Accolades from the outside regularly tell us how tall we stand among other communities. As recently as February 7, 2017, U.S. News and World Report ranked us as the 14th best place to live in the country.ⁱ

Yet, in 2013 when the headline broke about the Harvard University/UC Berkeley study that ranked Charlotte-Mecklenburg 50th out of 50 in upward mobilityⁱⁱ for children born into our lowest income quintile, many in our community responded with disbelief. **How, on the one hand, can we be such a vital and opportunity-rich community, and on the other, be ranked dead last in the odds that our lowest income children and youth will be able to move up the economic ladder as they become adults?**

Part 3
Characteristics of High-Mobility Areas

Why Does Upward Mobility Differ Across Areas?

- Why do some places produce much better outcomes for disadvantaged children than others?
- Begin by characterizing the features of areas with high rates of upward mobility

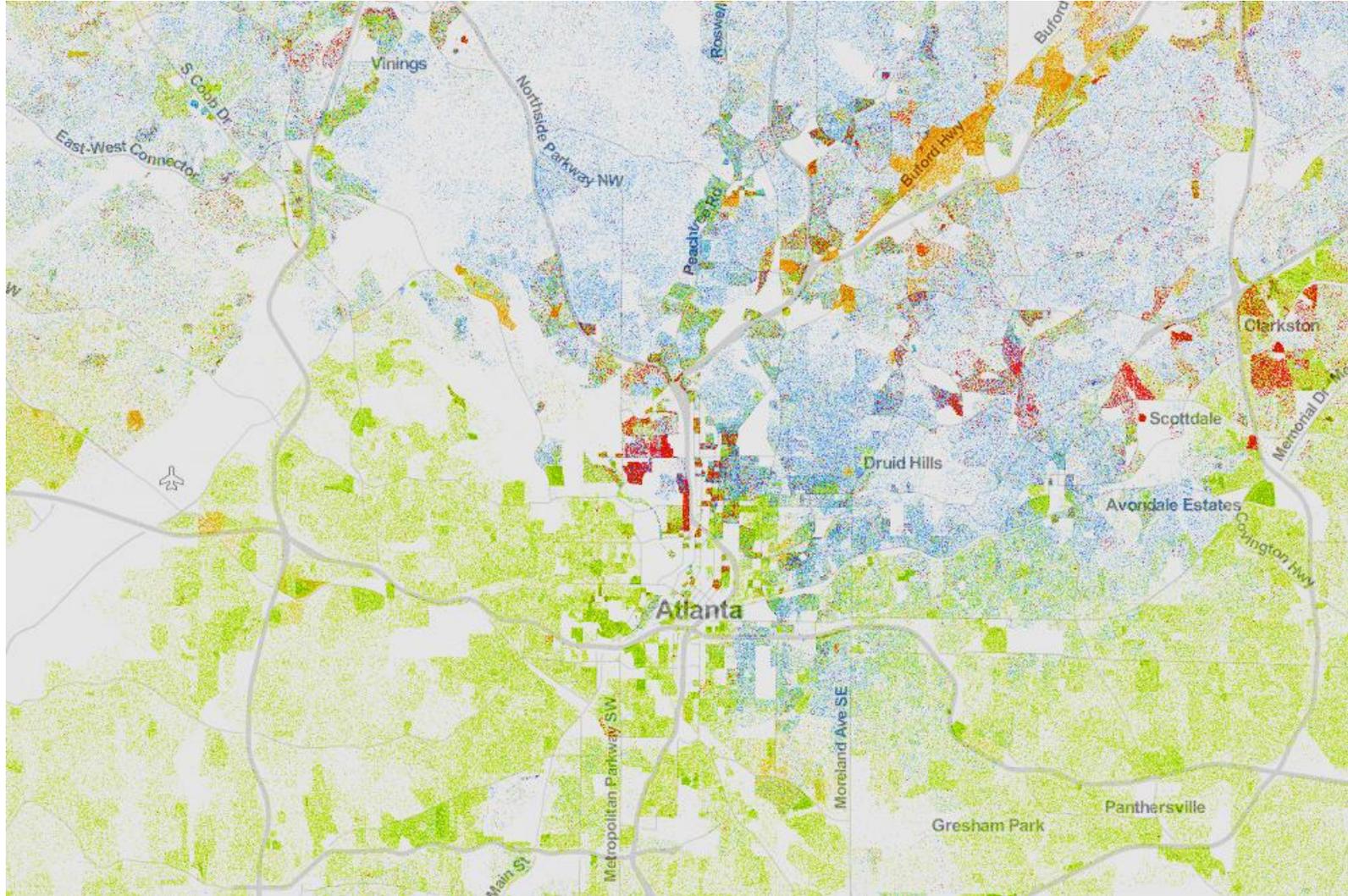
Five Strongest Correlates of Upward Mobility

1. Segregation

- Greater racial and income segregation associated with lower levels of mobility

Racial Segregation in Atlanta

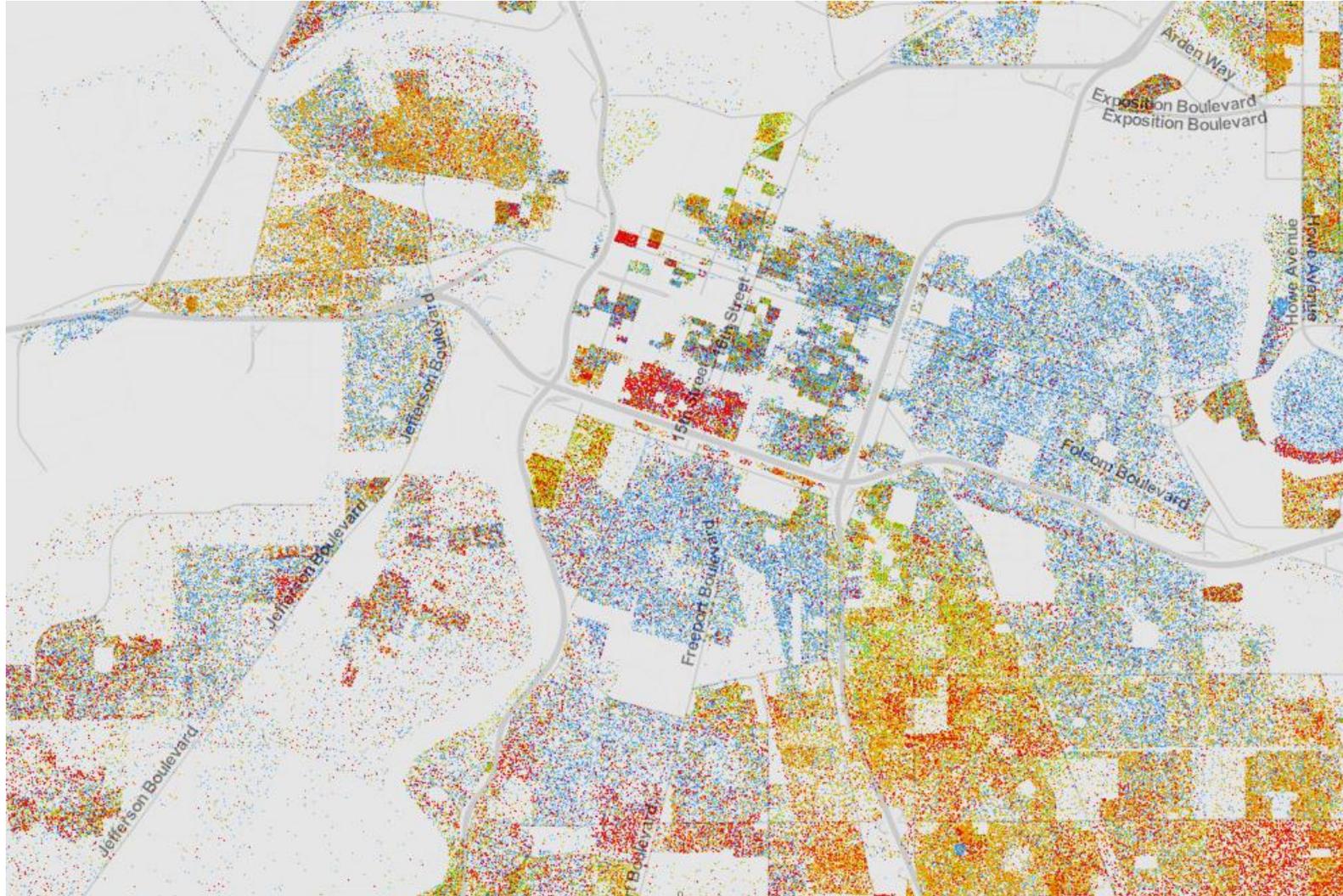
Whites (blue), Blacks (green), Asians (red), Hispanics (orange)



Source: Cable (2013) based on Census 2010 data

Racial Segregation in Sacramento

Whites (blue), Blacks (green), Asians (red), Hispanics (orange)



Source: Cable (2013) based on Census 2010 data

Five Strongest Correlates of Upward Mobility

1. Segregation

2. Income Inequality

- Places with smaller middle class have much less mobility

Five Strongest Correlates of Upward Mobility

1. Segregation
2. Income Inequality
3. School Quality
 - Higher expenditure, smaller classes, higher test scores correlated with more mobility

Five Strongest Correlates of Upward Mobility

1. Segregation
2. Income Inequality
3. School Quality
4. Family Structure
 - Areas with more single parents have much lower mobility
 - Strong correlation even for kids whose *own* parents are married

Five Strongest Correlates of Upward Mobility

1. Segregation
2. Income Inequality
3. School Quality
4. Family Structure
5. Social Capital
 - “It takes a village to raise a child”
 - Putnam (1995): “Bowling Alone”