

Storytelling with R

OR: How to turn a finding into a story

Newton's (inferred) law of people reading stuff:

Most people don't want to read/consume things that feel like work. But you can make decisions that make even complex topics and findings less work.

So what makes a good story?

Here are some things journalists think about when they are working on a narrative:

- A story has **characters** - people seem to sympathize most with people as characters, but they can also be an entity like a corporation, or a system like a school district.
- **Protagonists and antagonists** - sometimes this means good guys and bad guys, but obviously sometimes that is a ridiculous oversimplification. More generally, this is the entities whose experiences will serve to advance the story.
- A story has a **setting**, making it richer by grounding it in a place.
- **Beginning, middle, end** - Here's the setup. But wait! It is more complicated than that. But then the story is resolved in some way. The same information, structured differently, can tell a much more interesting story.

How does data fit in?

Some common ways:

1. Using data, we have established a pattern, and it is telling us which way the story should go. We want characters who **illustrate this pattern**.
2. Using data, we have identified **outliers**, and we want to give them the attention they deserve. In the investigative journalism world, we usually don't mean *attention* in a good way. In investigative journalism this is often more specifically bad actors. #FFS
3. Data contributes insights into the mechanisms by which something happens. We've learned something new about **how something works**. Caution: specifics will nearly always be more readable than generalizations.
4. On *extremely rare occasions* the process (data acquisition, methods) is interesting

Is your work relevant?

Our society is a cesspool (s***hole?) of inequality, corruption and dysfunction.

Does your work highlight someone or something that needs to change? Is someone or something to blame? Who has the power to change the situation?

How does your work intersect with this?

Give some thought to units

Can you base your inputs and outputs in tangible things?

- number of people vs. infections per respirator hour
- Jobs vs. FTEs
- Dollar\$ vs. percent change

Methods

Data analysts often don't realize how much the design of an analysis determines what they will be able to say later.

One example: Odds v. Risk

Can you run it both ways?

Sometimes it is a successful strategy to analyze the most robust way, and also a simpler more-explainable way. This way you can be more accessible without sacrificing rigor.

Help people make sense of it

Give people a framework and context to make sense of it all.

Often people don't want dashboard-type data without a framework for understanding. That doesn't mean being overbearing or non-transparent.

Case study #1: Surgeon Scorecard

Surgeon Scorecard

by Sisi Wei, Olga Pierce and Marshall Allen, ProPublica, Updated July 15, 2015

Guided by experts, ProPublica calculated death and complication rates for surgeons performing one of eight elective procedures in Medicare, carefully adjusting for differences in patient health, age and hospital quality. Use this database to know more about a surgeon before your operation.

READ OUR STORY

Making the Cut: Why Choosing the Right Surgeon Matters Even More Than You Know

METHODOLOGY

Read how we calculated complications and the key questions we considered.

EDITOR'S NOTE

Why ProPublica is naming surgeons and what experts are saying about it

Find Near Me

Find a Surgeon

Find a Hospital

Surgeons and Hospitals Near My Location

ex: 155 6th Avenue, New York, New York

Knee Replacement



Use My Location

Case study #1: Surgeon Scorecard

PROCEDURE SELECTED

Knee Replacement Performance at Hospitals Within 25 Miles

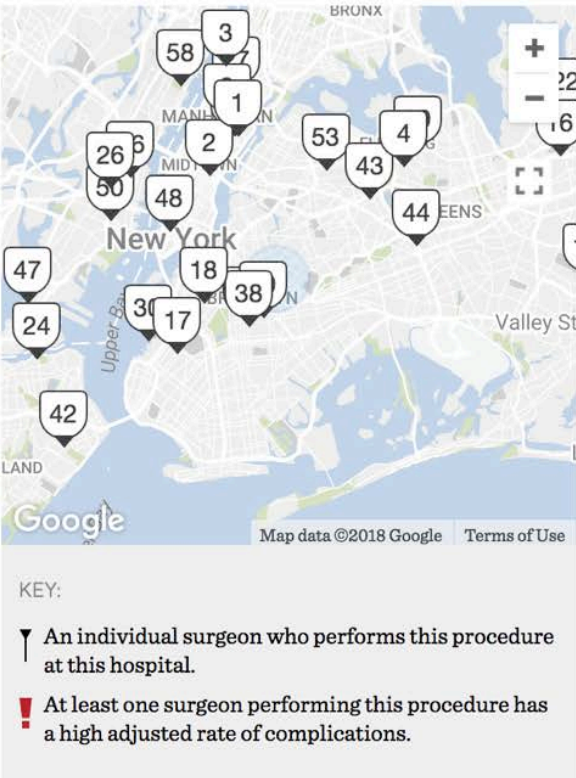
Replace diseased knee joint with an artificial knee. The most common reason for a knee replacement is osteoarthritis, which is a breakdown of the cartilage in the joint. [More information »](#)
















61
nearby hospitals performed this procedure on Medicare patients

3 (5%)
have at least one surgeon with a low adjusted complication rate

! 26 (43%)
have at least one surgeon with a high adjusted complication rate

Sorted by the surgeon with the lowest adjusted rate of complications at each hospital, along with a measure representing the combined performance of surgeons and hospitals for these procedures.



1		! HOSPITAL FOR SPECIAL SURGERY, NEW YORK
2		! NYU HOSPITALS CENTER, NEW YORK
3		ST LUKE'S ROOSEVELT HOSPITAL, NEW YORK
4		! NEW YORK HOSPITAL MEDICAL CENTER OF QUEENS, FLUSHING
5		! HOLY NAME MEDICAL CENTER, TEANECK
6		GLEN COVE HOSPITAL, GLEN COVE
7		VALLEY HOSPITAL, RIDGEWOOD
8		SOUND SHORE MEDICAL CENTER OF WESTSCHESTER, NEW RO...
9		LENOX HILL HOSPITAL, NEW YORK
10		! ENGLEWOOD HOSPITAL AND MEDICAL CENTER, ENGLEWOOD
11		! RARITAN BAY MEDICAL CENTER, PERTH AMBOY
12		! RIVERVIEW MEDICAL CENTER, RED BANK
13		! ST JOHN'S RIVERSIDE HOSPITAL, YONKERS
14		! ST JOSEPH'S MEDICAL CENTER, YONKERS
15		NEW YORK-PRESBYTERIAN HOSPITAL, NEW YORK

Case study #1: Surgeon Scorecard

Lumbar Spinal Fusion, Anterior Column, Posterior Approach

Fusing two or more vertebrae in the lower back; performed on the front of the spine, incision in the back (ICD-9-CM code 81.08)

The fusing of two or more vertebrae in the lower back, performed on the front portion of the spine. One of the most common reasons is the narrowing of the space between the vertebrae, which puts pressure on the spinal cord and nerves, causing pain. It can also be done because of disc degeneration or a condition where one bone in the back slides forward over the bone below it. [More information](#)

JAMES RONZO »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE
297 times	1-10	Redacted	<div><div>LowMediumHigh Adjusted Rate of Complications</div><div>2.8%</div></div>

FRANK BONO »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE
417 times	17	4.1%	<div><div>LowMediumHigh Adjusted Rate of Complications</div><div>4.3%</div></div>

CONSTANTINE TOUMBIS »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE
246 times	27	11%	<div><div>LowMediumHigh Adjusted Rate of Complications</div><div>7.5%</div></div>

Case study #1: Surgeon Scorecard



Photo color altered to be less visceral.

While in medical school, Toumbis slashed a friend's throat with a large knife.

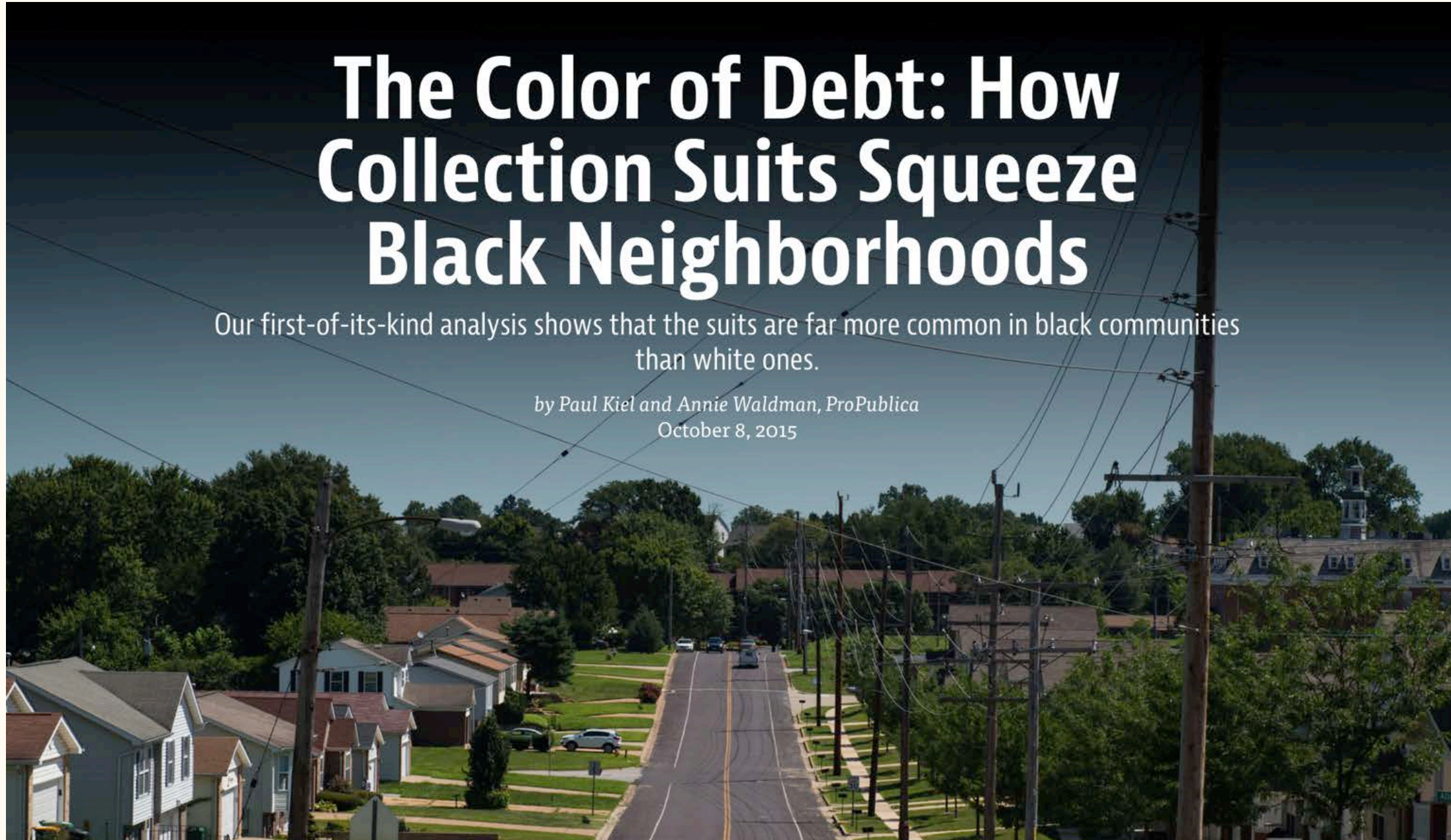
Though he was required to disclose this in the Florida licensure process, he failed to do so. In some cases, he omitted it altogether. In another, changed the story, saying he was the victim of the attack.

Case study #2: Color of Debt

The Color of Debt: How Collection Suits Squeeze Black Neighborhoods

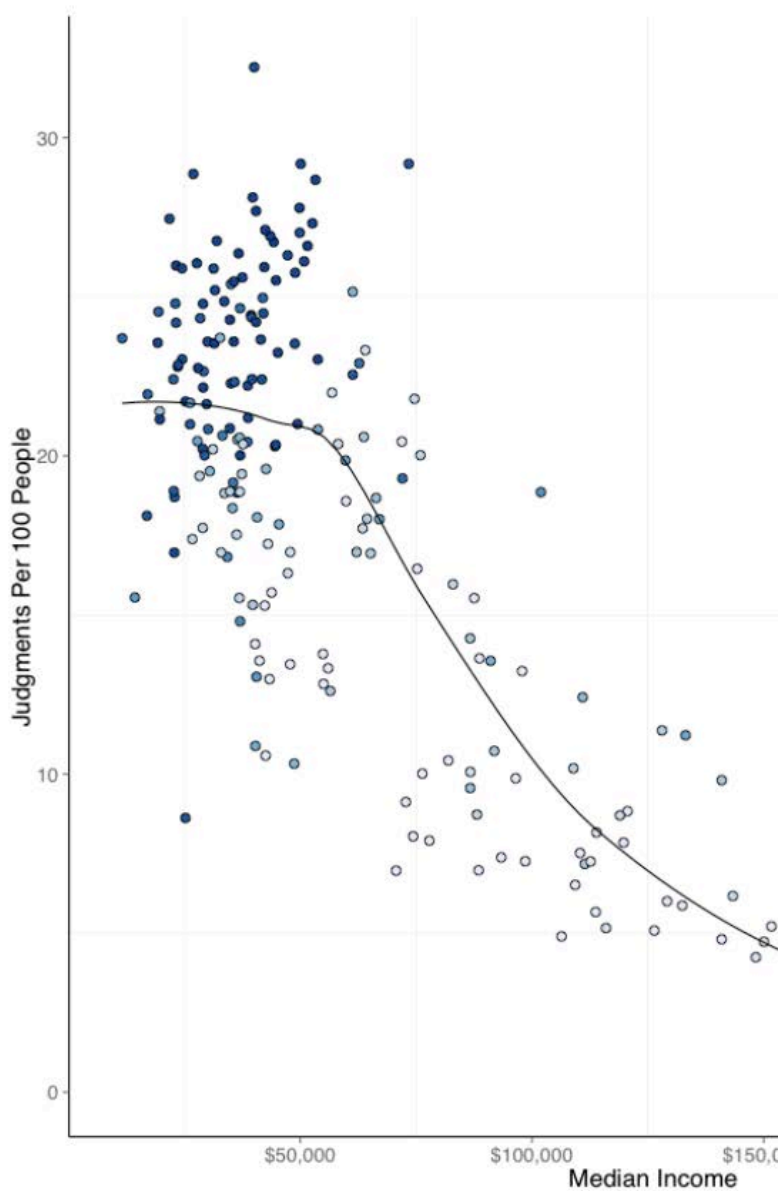
Our first-of-its-kind analysis shows that the suits are far more common in black communities than white ones.

by Paul Kiel and Annie Waldman, ProPublica
October 8, 2015

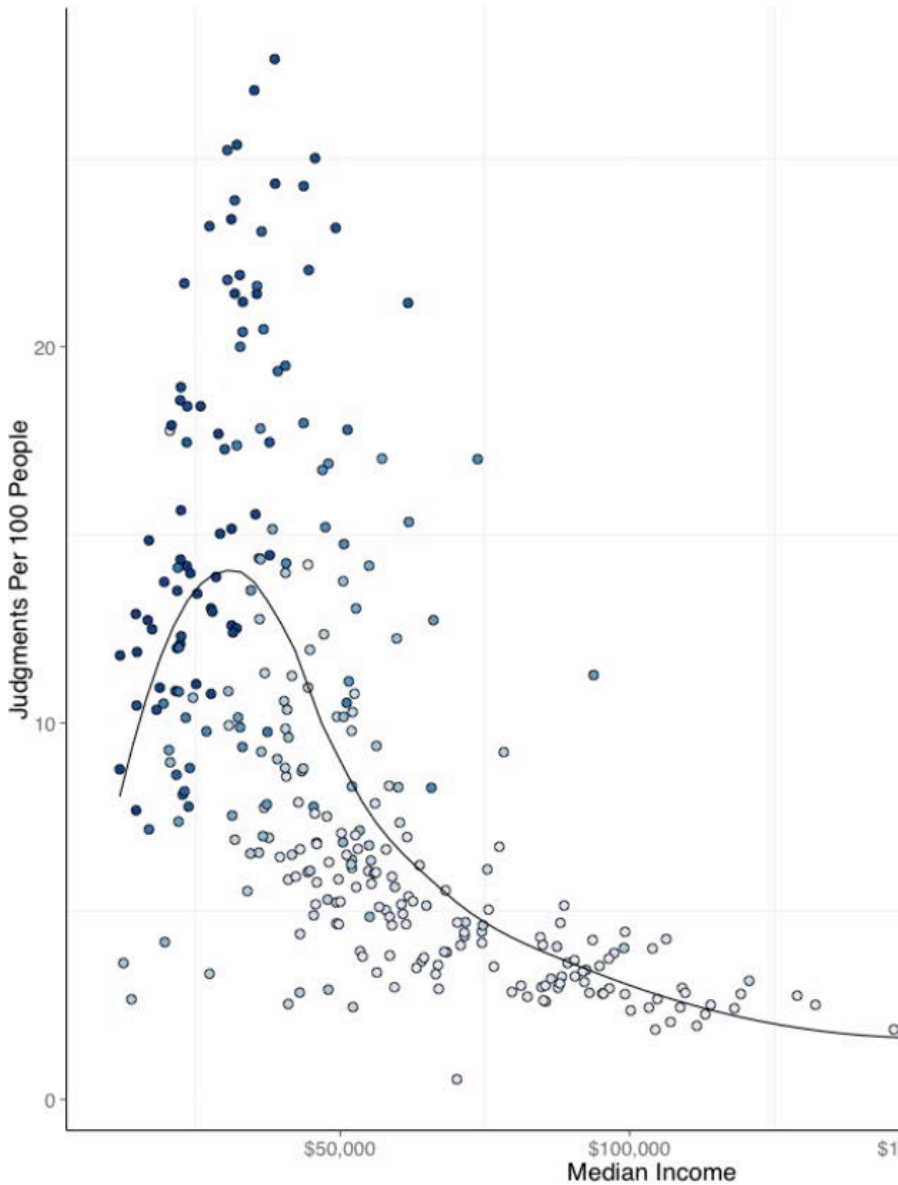


Case study #2: Color of Debt

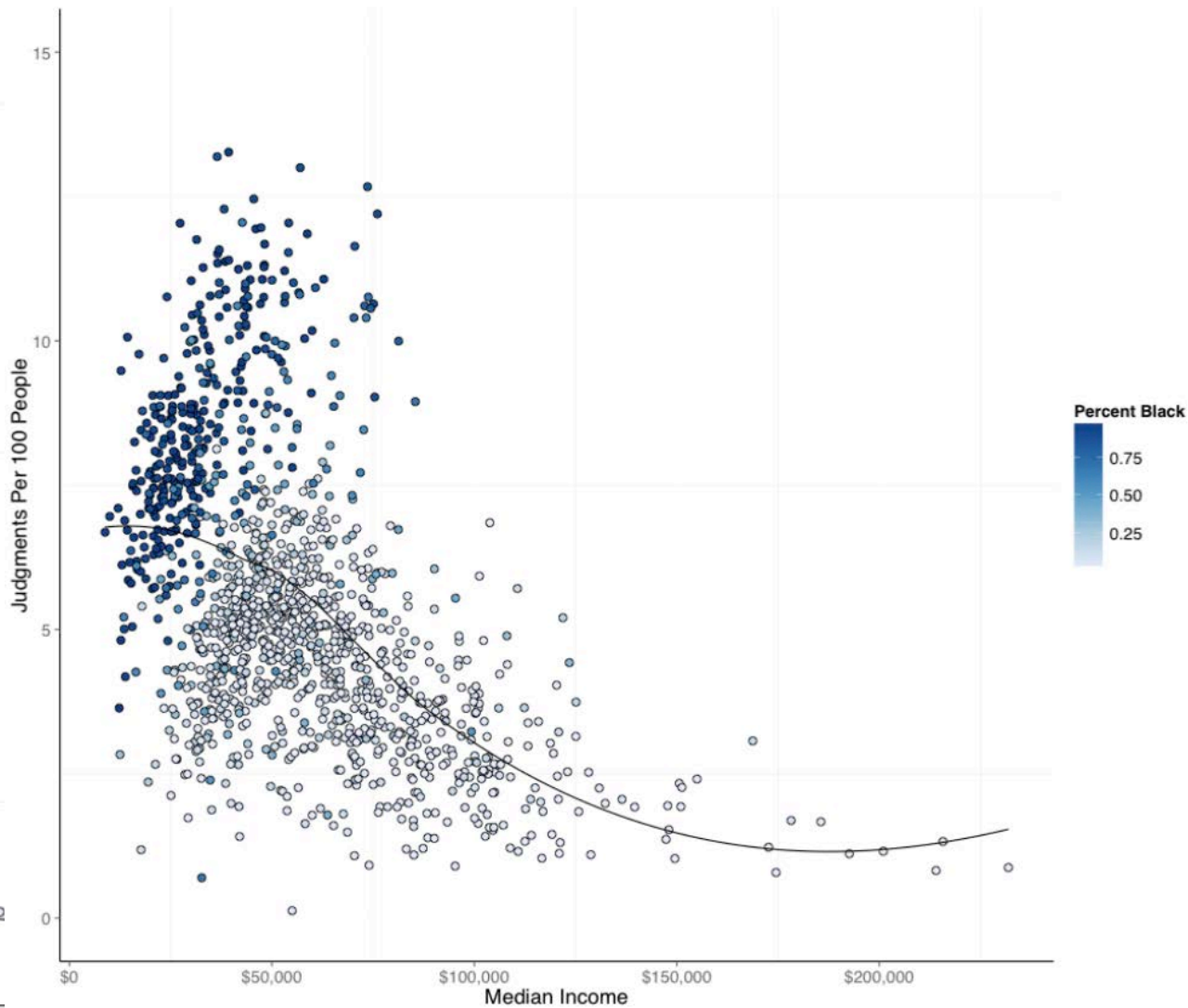
3.2.3. Analysis of Essex County, New Jersey



3.2.1. Analysis of St. Louis County and City, Missouri



3.2.2. Analysis of Cook County, Illinois



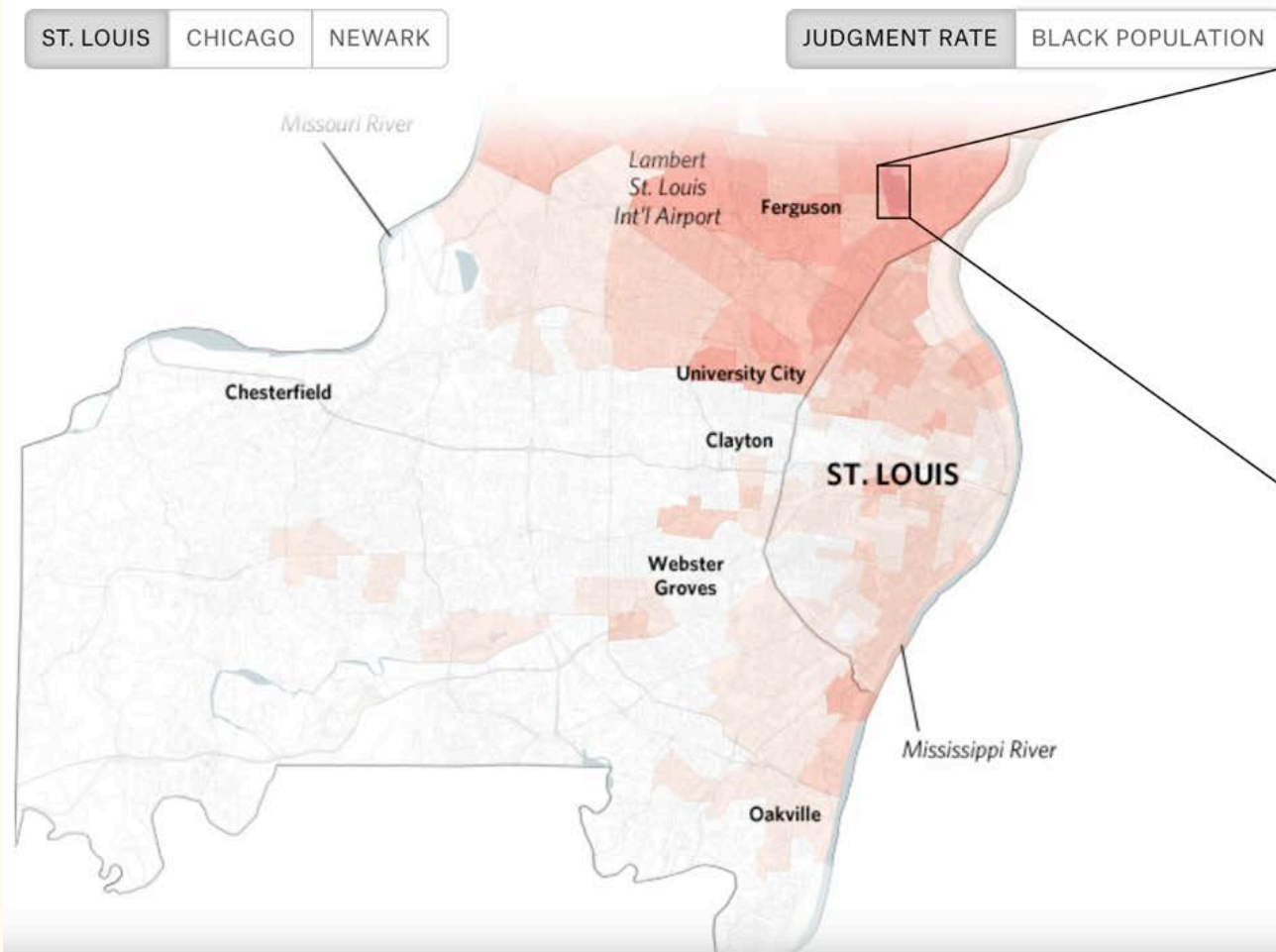
Case study #2: Color of Debt

The Color of Debt

The Black Neighborhoods Where Collection Suits Hit Hardest

by Al Shaw, Annie Waldman and Paul Kiel, ProPublica, October 8, 2015

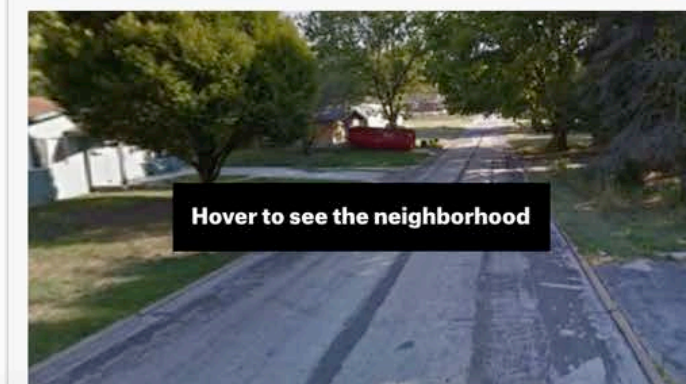
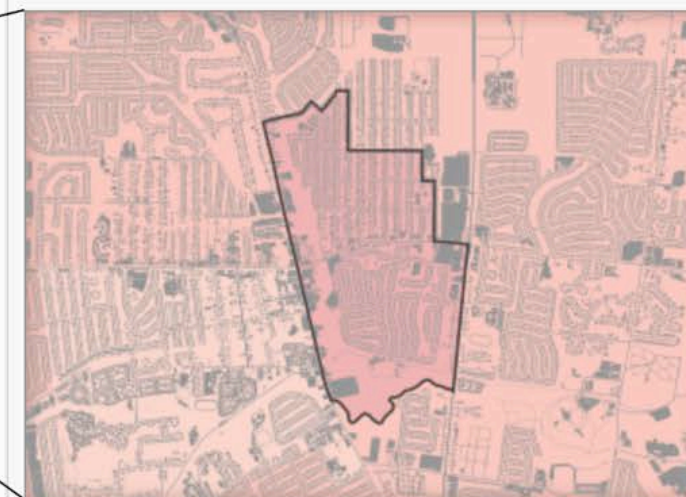
In the St. Louis area, the neighborhoods hardest hit by debt collection lawsuits are in the suburbs of north St. Louis County, an area that since 1960 has gone from nearly all white to mostly black. Here's a closer look at five of those neighborhoods. *Note: Data includes suits filed between 2008-2012.* | [Related story »](#)



Castle Point & Moline Acres

Median Household Income	<u>Judgments Per 100 Residents</u>	Black Pop.
\$38,674	27.6	92.1%

Castle Point and Moline Acres are the neighborhoods hardest hit by collection suits across St. Louis city and county. Debt buyers filed the most suits here, accounting for more than 23 percent of the judgments.



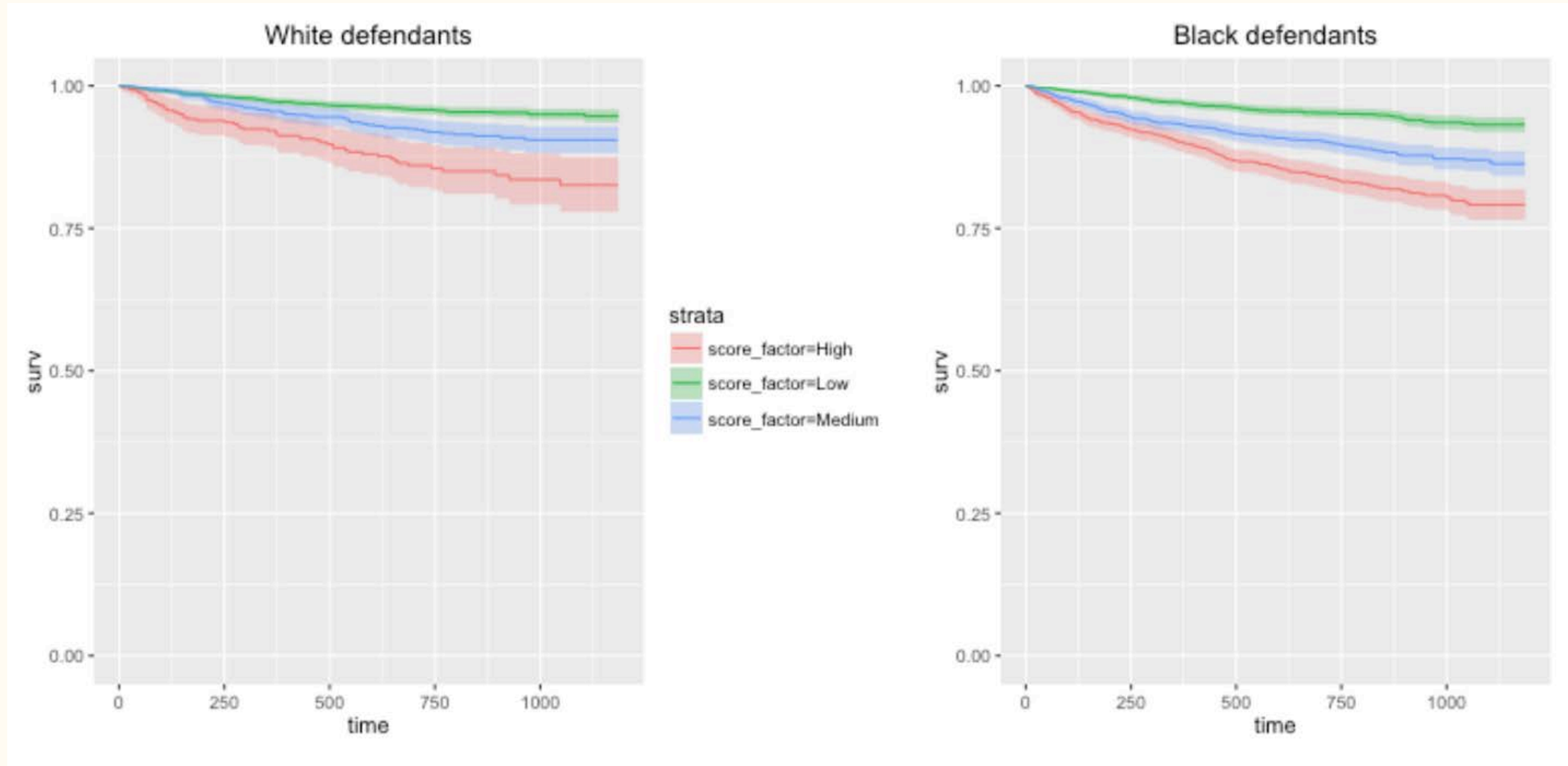
Case study #2: Color of Debt



Case study #3: Machine Bias



Case study #3: Machine Bias

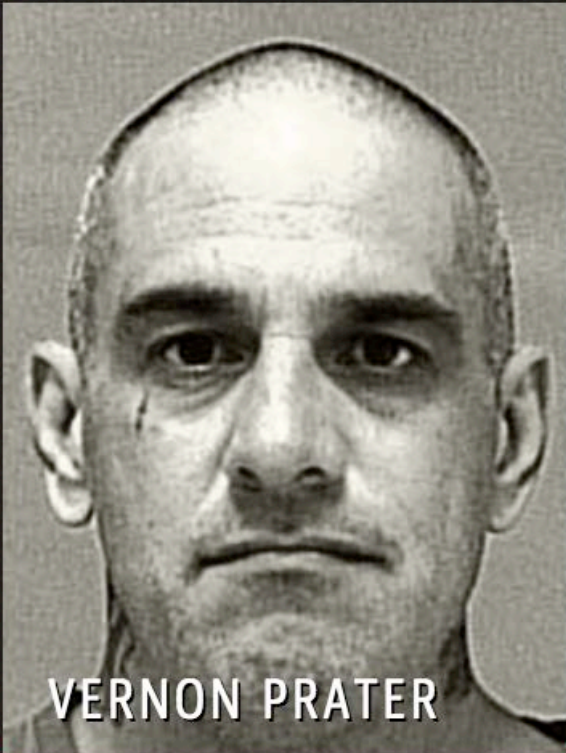


Case study #3: Machine Bias

	All Defendants			Black defendants			White defendants	
	Low	High		Low	High		Low	High
Survived	4121	1597	Survived	1692	1043	Survived	1679	380
Recidivated	347	389	Recidivated	170	273	Recidivated	129	77
FP rate: 27.93			FP rate: 38.14			FP rate: 18.46		
FN rate: 47.15			FN rate: 38.37			FN rate: 62.62		
PPV: 0.20			PPV: 0.21			PPV: 0.17		
NPV: 0.92			NPV: 0.91			NPV: 0.93		
LR+: 1.89			LR+: 1.62			LR+: 2.03		
LR-: 0.65			LR-: 0.62			LR-: 0.77		

Case study #3: Machine Bias

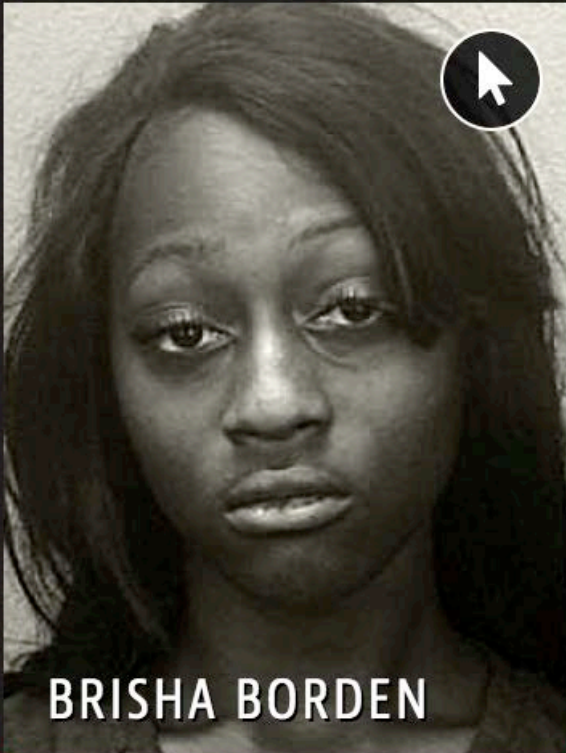
Two Petty Theft Arrests



VERNON PRATER

LOW RISK

3



BRISHA BORDEN

HIGH RISK

8

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

VERNON PRATER

Prior Offenses

2 armed robberies, 1 attempted armed robbery

Subsequent Offenses

1 grand theft

LOW RISK

3

BRISHA BORDEN

Prior Offenses

4 juvenile misdemeanors

Subsequent Offenses

None

HIGH RISK

8