REGRESSION DISCONTINUITY I

PMAP 8521: Program Evaluation for Public Service
November 4, 2019

Fill out your reading report on iCollege!

PLAN FOR TODAY

Jumps and cutoffs

Measuring the size of the discontinuity

Main RDD concerns

RDD with R

JUMPS & CUTOFFS

Think of five social programs that use eligibility cutoffs to determine who can access the program

How is eligibility measured? What's the cutoff?

Federal/state/local governments; school districts; nonprofits; etc.

KEY TERMS

Running/forcing variable

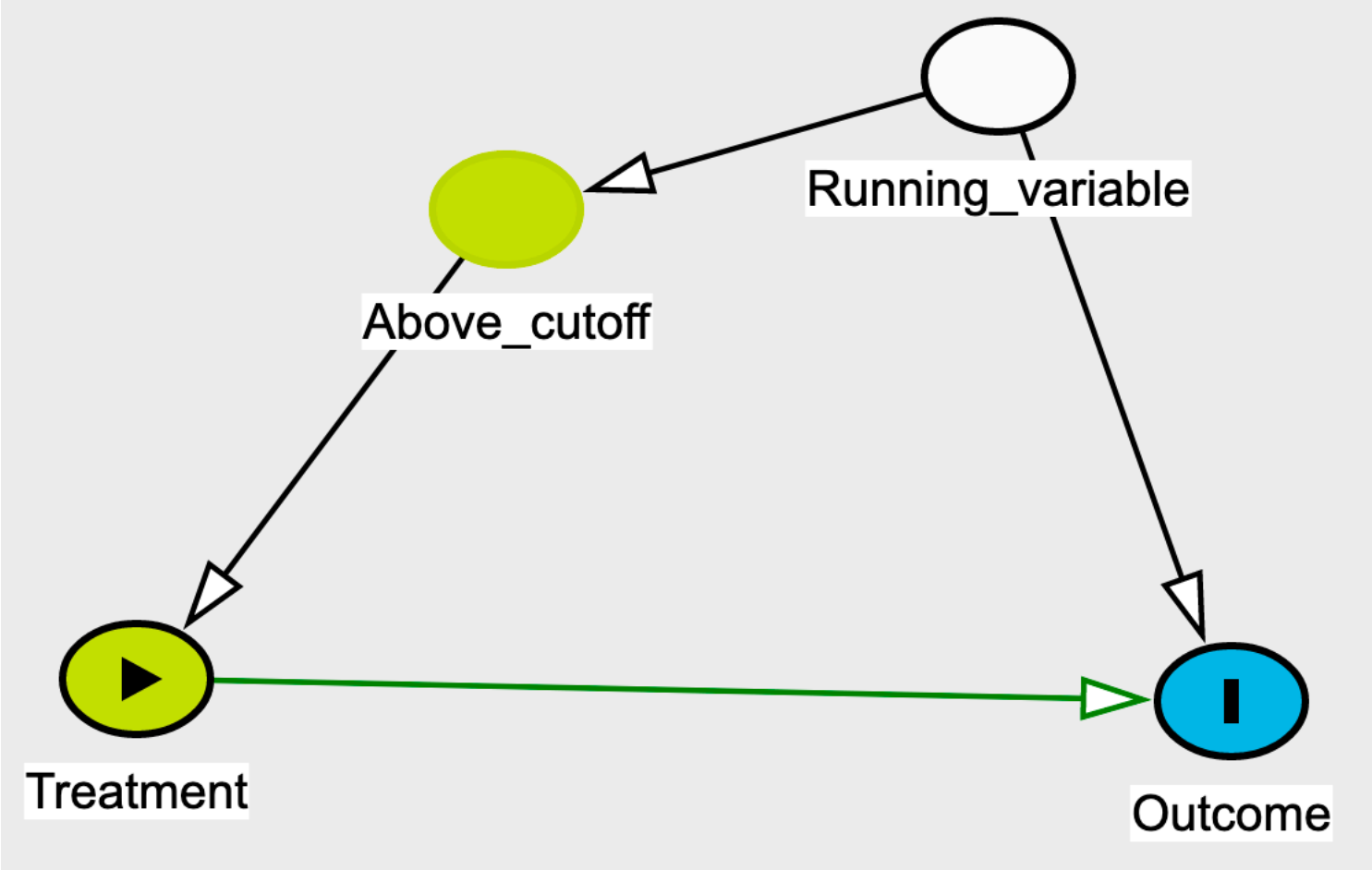
Index or measure that determines eligibility

Cutoff/cutpoint/threshold

Number that formally assigns access to program

Outcome

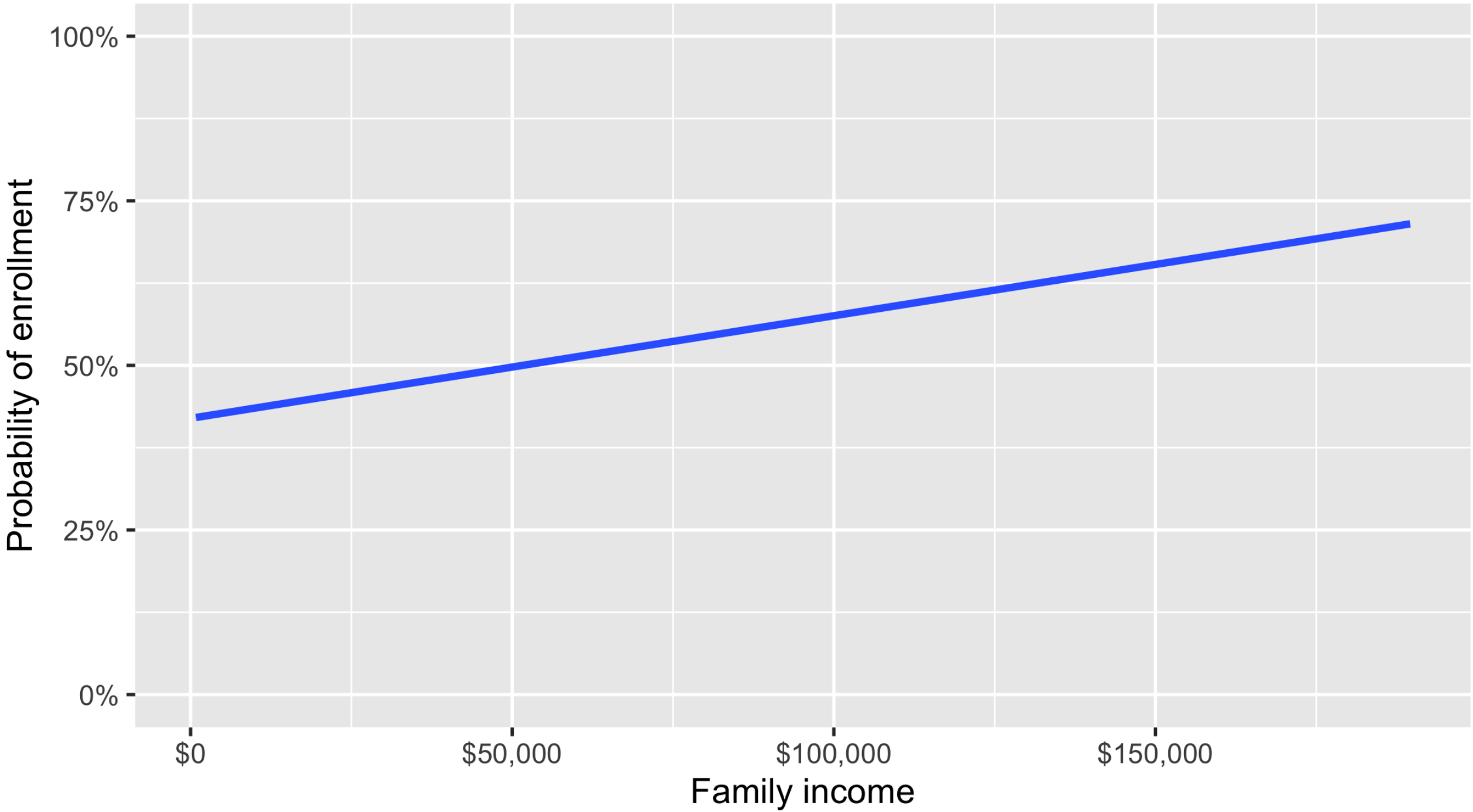
The thing you want to see the causal effect on

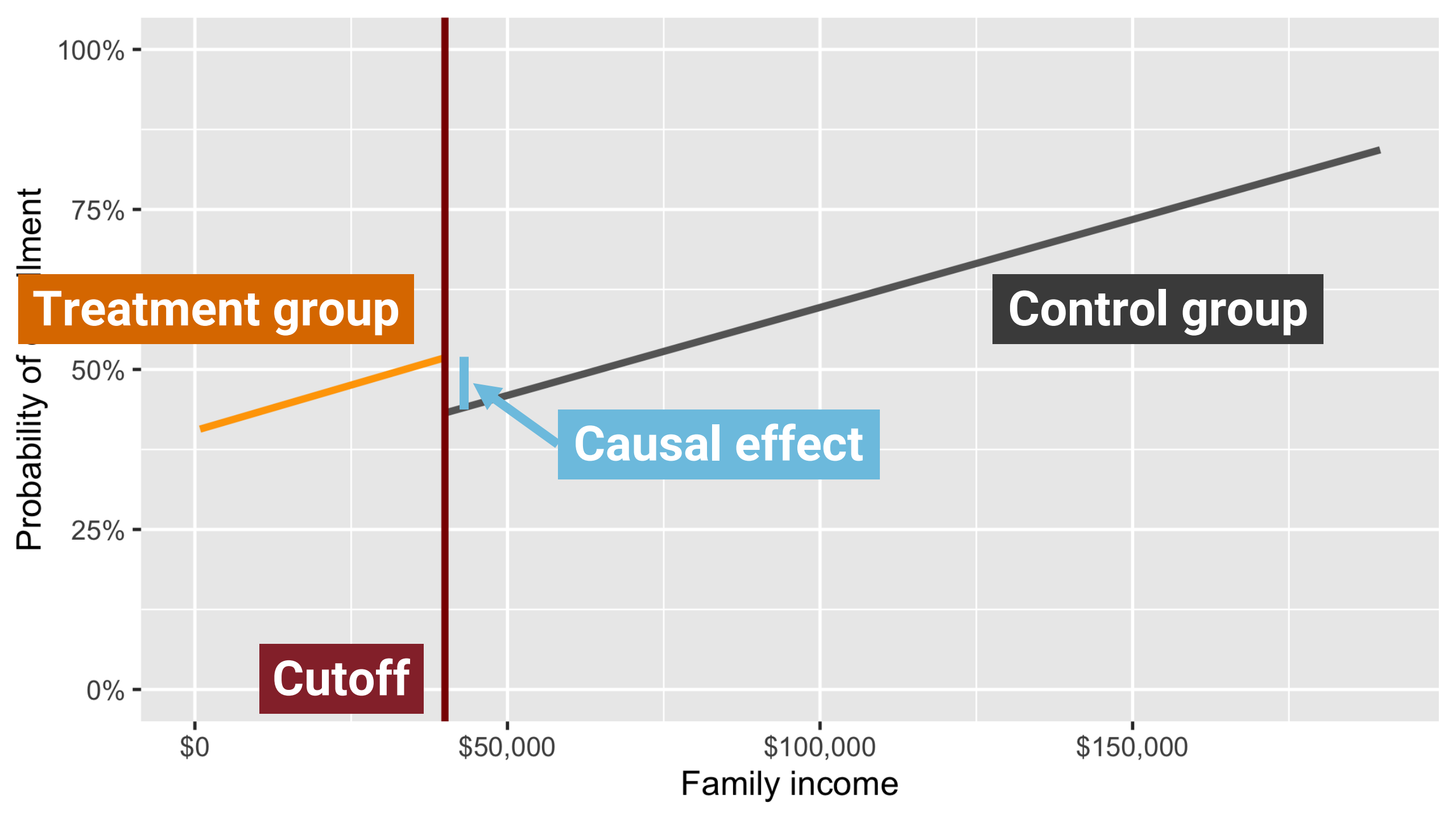

Running_variable
Above_cutoff
Treatment
Outcome

MAIN INTUITION

People right before and right after the cutoff are essentially the same

This mimics the idea of treatment and control groups


100%
75%
50%
25%
0%
Probability of enrollment
$0
$50,000
$100,000
$150,000
Family income


100%
75%
50%
25%
0%
Probability of
llment
Treatment group
Control group
Causal effect
Cutoff
$0
$50,000
$100,000
$150,000
Family income

After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays*

By Douglas Almond† and Joseph J. Doyle Jr.‡

Estimates of moral hazard in health insurance markets can be confounded by adverse selection. This paper considers a plausibly exogenous source of variation in insurance coverage for childbirth in California. We find that additional health insurance coverage induces substantial extensions in length of hospital stay for mother and newborn. However, remaining in the hospital longer has no effect on readmissions or mortality, and the estimates are precise. Our results suggest that for uncomplicated births, minimum insurance mandates incur substantial costs without detectable health benefits.

Figure 3A: Additional Midnights: Before Law Change

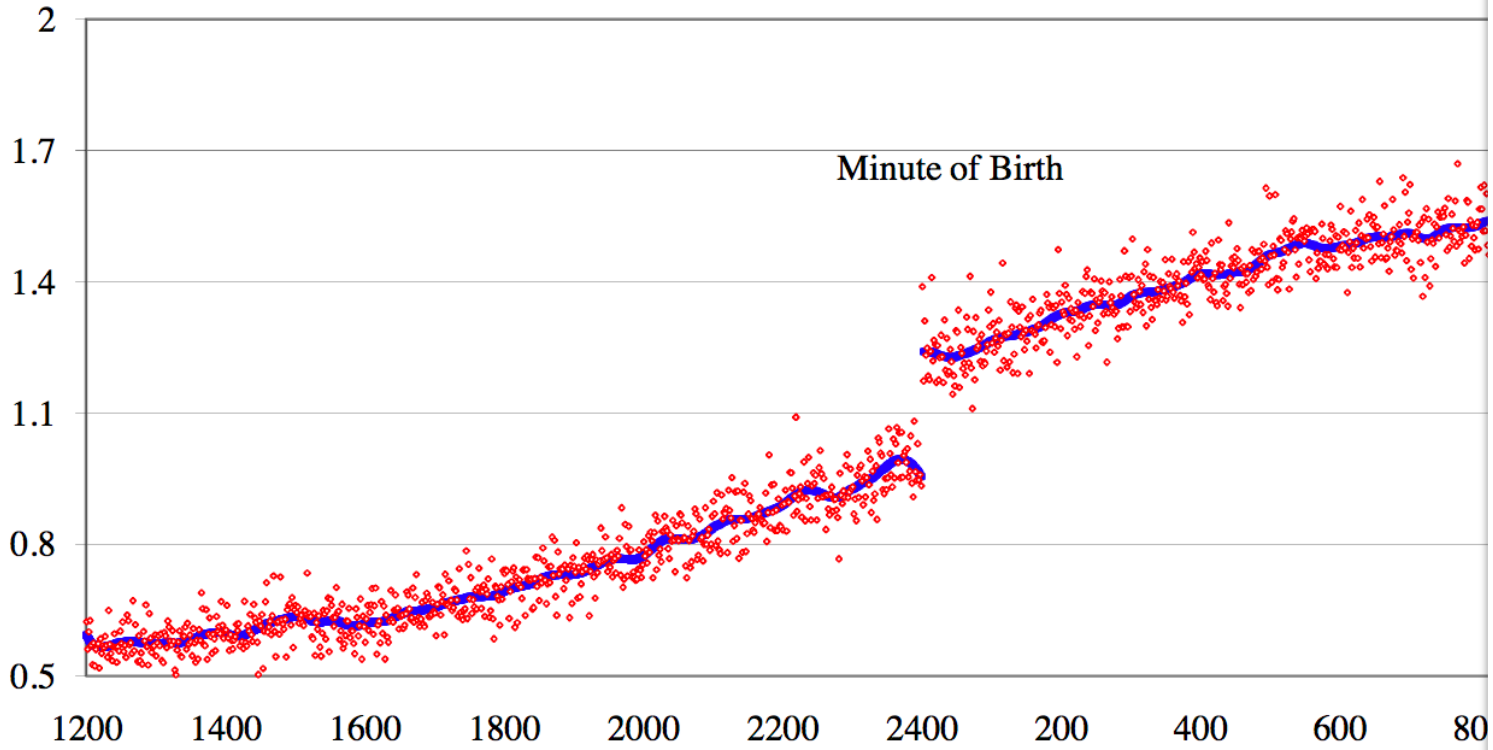


Figure 4A: 28-Day Readmission Rate: Before Law Change

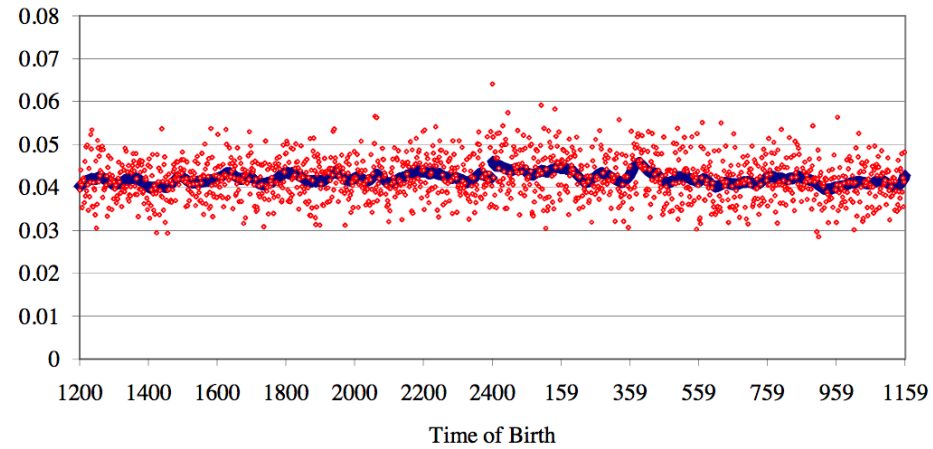
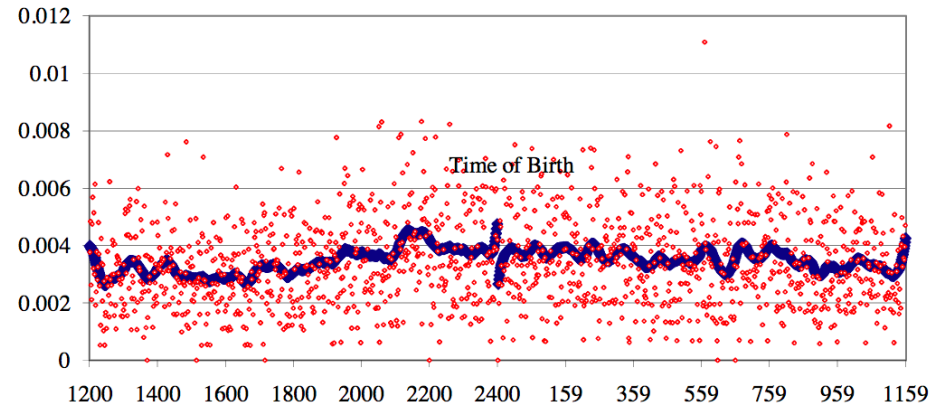


Figure 4C: 28-Day Mortality Rate: Before Law Change



THE EFFECT OF ATTENDING THE FLAGSHIP STATE UNIVERSITY ON EARNINGS: A DISCONTINUITY-BASED APPROACH

Mark Hoekstra*

Abstract—This paper examines the effect of attending the flagship state university on the earnings of 28 to 33 year olds by combining confidential admissions records from a large state university with earnings data collected through the state's unemployment insurance program. To distinguish the effect of attending the flagship state university from the effects of confounding factors correlated with the university's admission decision or the applicant's enrollment decision, I exploit a large discontinuity in the probability of enrollment at the admission cutoff. The results indicate that attending the most selective state university causes earnings to be approximately 20% higher for white men.

I. Introduction

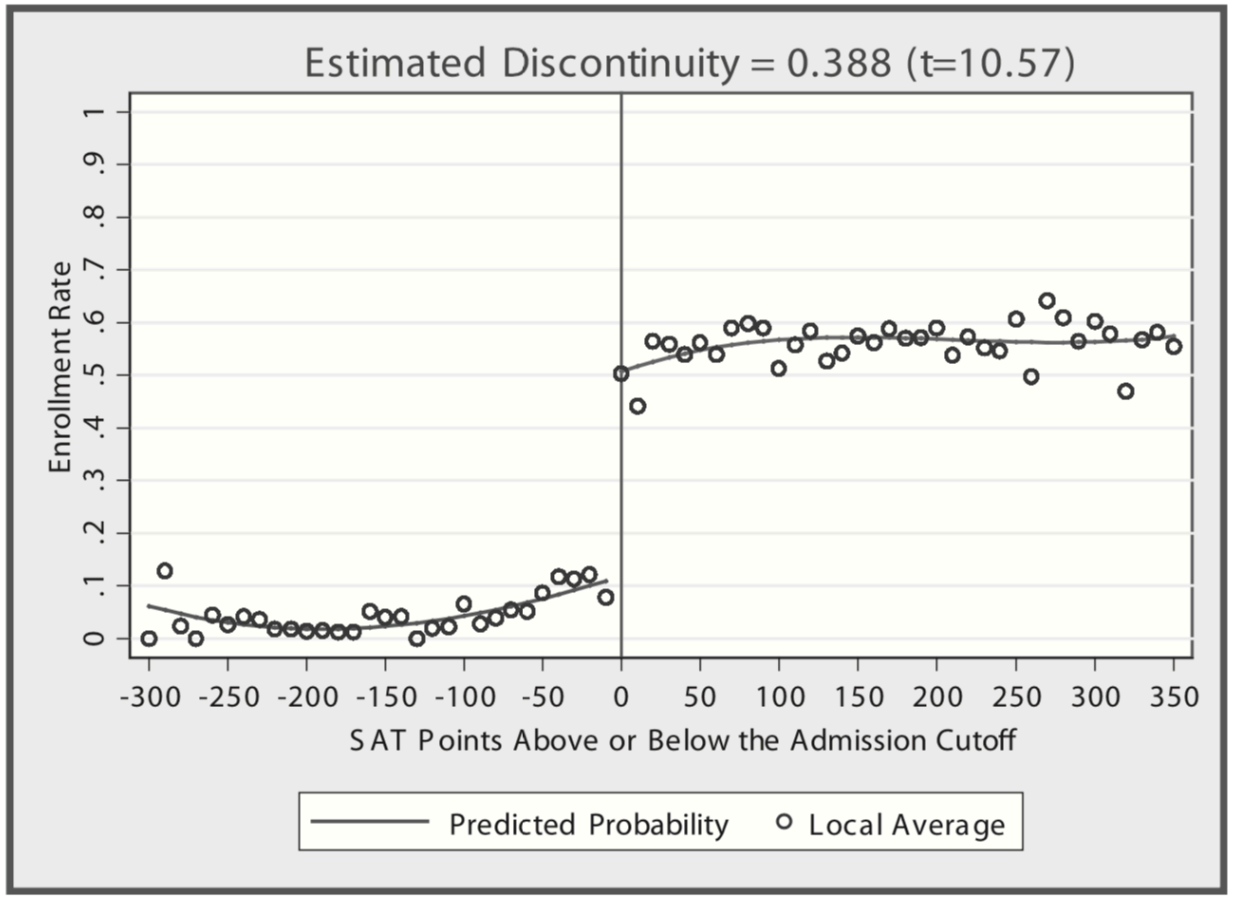
WHILE there has been considerable study of the effect of educational attainment on earnings, less is known regarding the economic returns to college quality. This paper examines the economic returns to college quality in the context of attending the most selective public state university. It does so using an intuitive regression discontinuity design that compares the earnings of 28 to 33 year olds who were barely admitted to the flagship to those of individuals who were barely rejected.

Convincingly estimating the economic returns to college quality requires overcoming the selection bias arising from

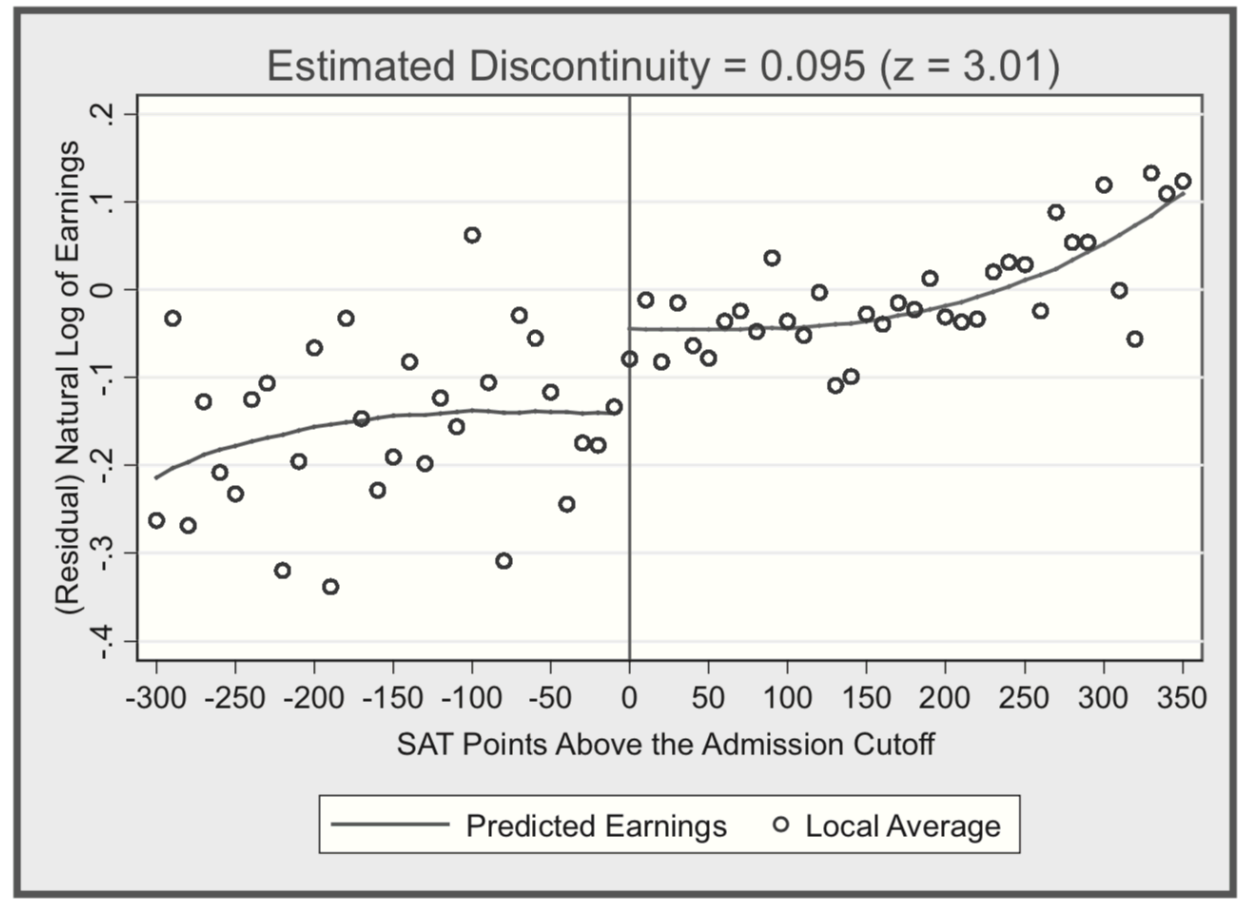
leges but chose to attend less selective institutions. They find that attending more selective colleges has a positive effect on earnings only for students from low-income families. Brewer, Eide, and Ehrenberg (1999) estimate the payoff by explicitly modeling high school students' choice of college type and find significant returns to attending an elite private institution for all students. Behrman, Rozenzweig, and Taubman (1996) identify the effect by comparing female twin pairs and find evidence of a positive payoff from attending Ph.D.-granting private universities with well-paid senior faculty. Using a similar approach, Lindahl and Regner (2005) use Swedish sibling data and show that cross-sectional estimates of the selective college wage premium are twice the within-family estimates.

This paper uses a different strategy in that it identifies the effect of school selectivity on earnings by comparing the earnings of those just below the cutoff for admission to the flagship state university to those of applicants who were barely above the cutoff for admission. To do so, I combined confidential administrative records from a large flagship state university with earnings records collected by the state

FIGURE 1.—FRACTION ENROLLED AT THE FLAGSHIP STATE UNIVERSITY



RAL LOG OF ANNUAL EARNINGS FOR WHITE MEN TEN TO FIFTEEN YEARS AFTER HIGH SCHOOL GRADUATION (F
POLYNOMIAL OF ADJUSTED SAT SCORE)



MEASURING THE SIZE OF THE DISCONTINUITY

The size of the discontinuity depends on how you draw the trend lines on each side of the cutoff

There’s no one right way to draw lines!

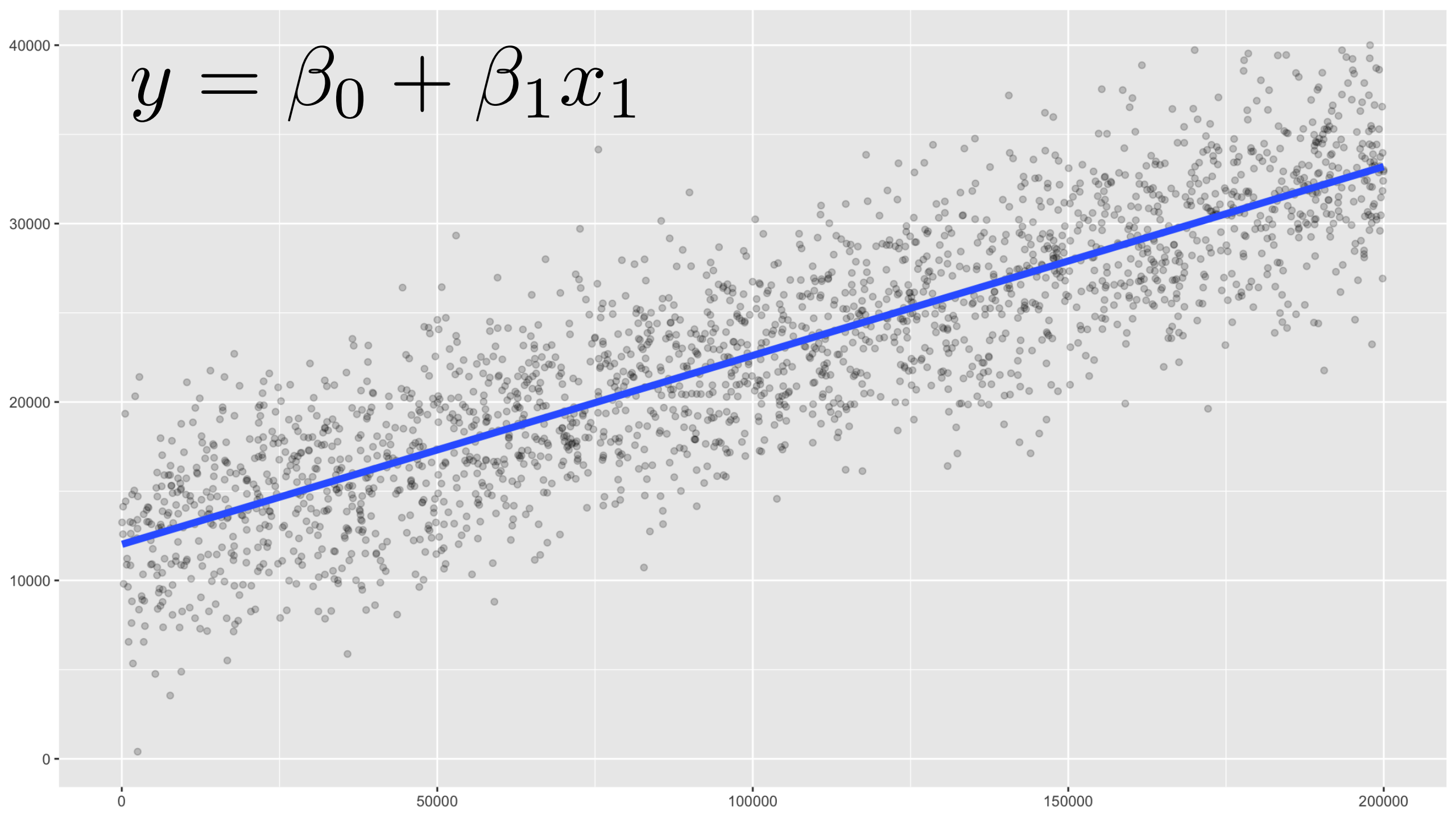
Parametric **Nonparametric**

Bandwidths **Kernels**

PARAMETRIC LINES

Formulas with *parameters*

$$y = mx + b \qquad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$


$y = \beta_0 + \beta_1 x_1$
40000
30000
20000
10000
0
0
50000
100000
150000
200000

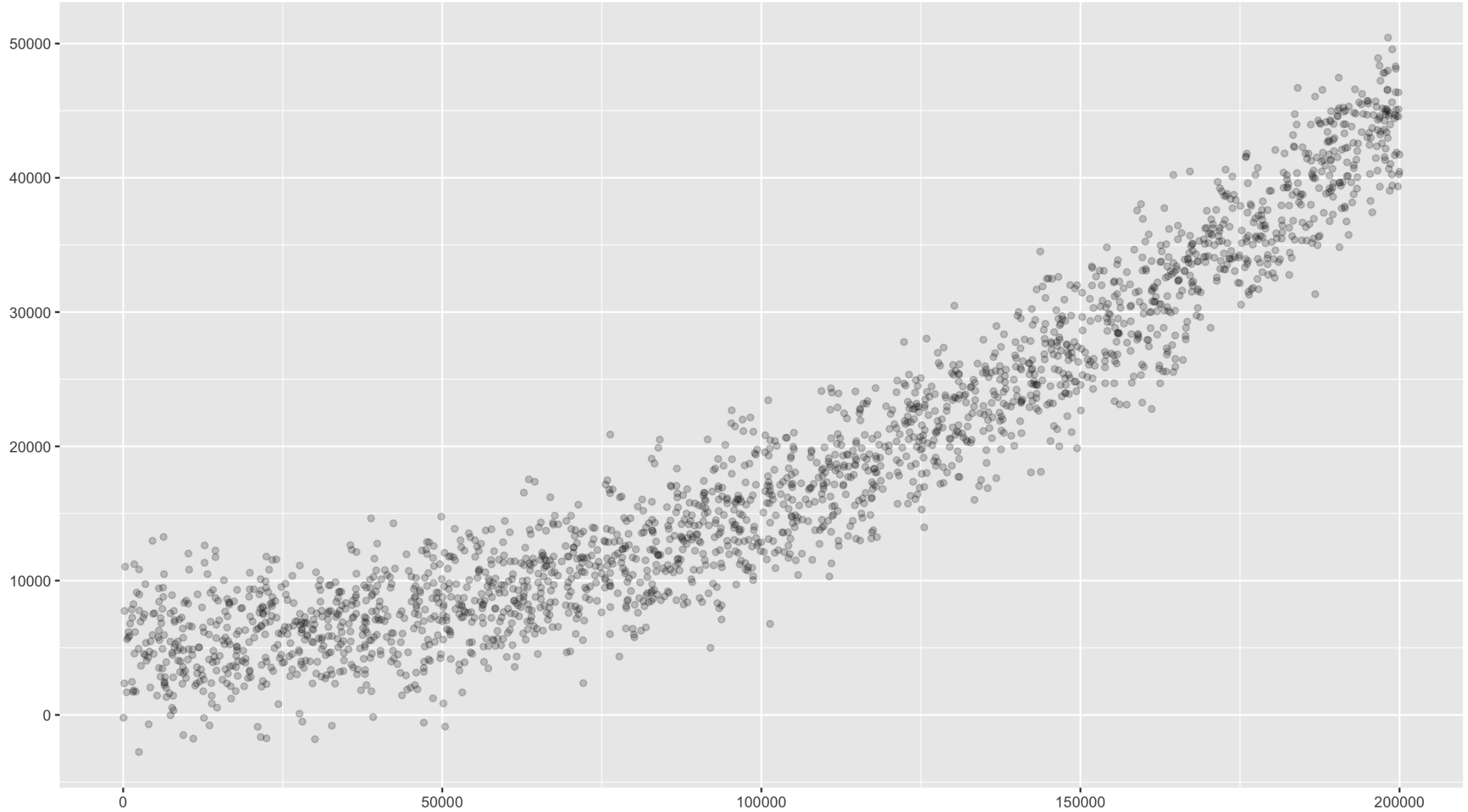
PARAMETRIC LINES

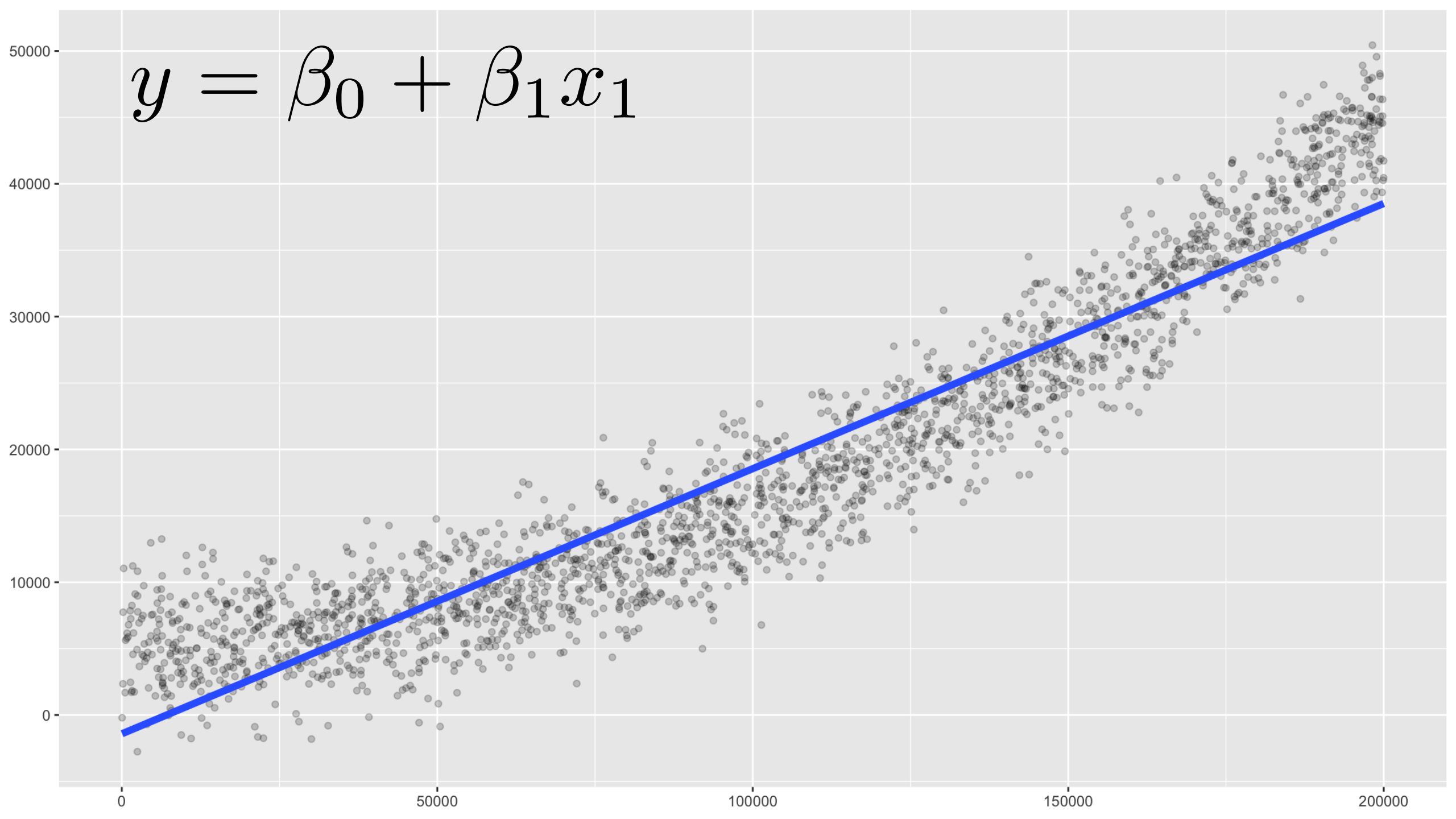
Formulas with *parameters*

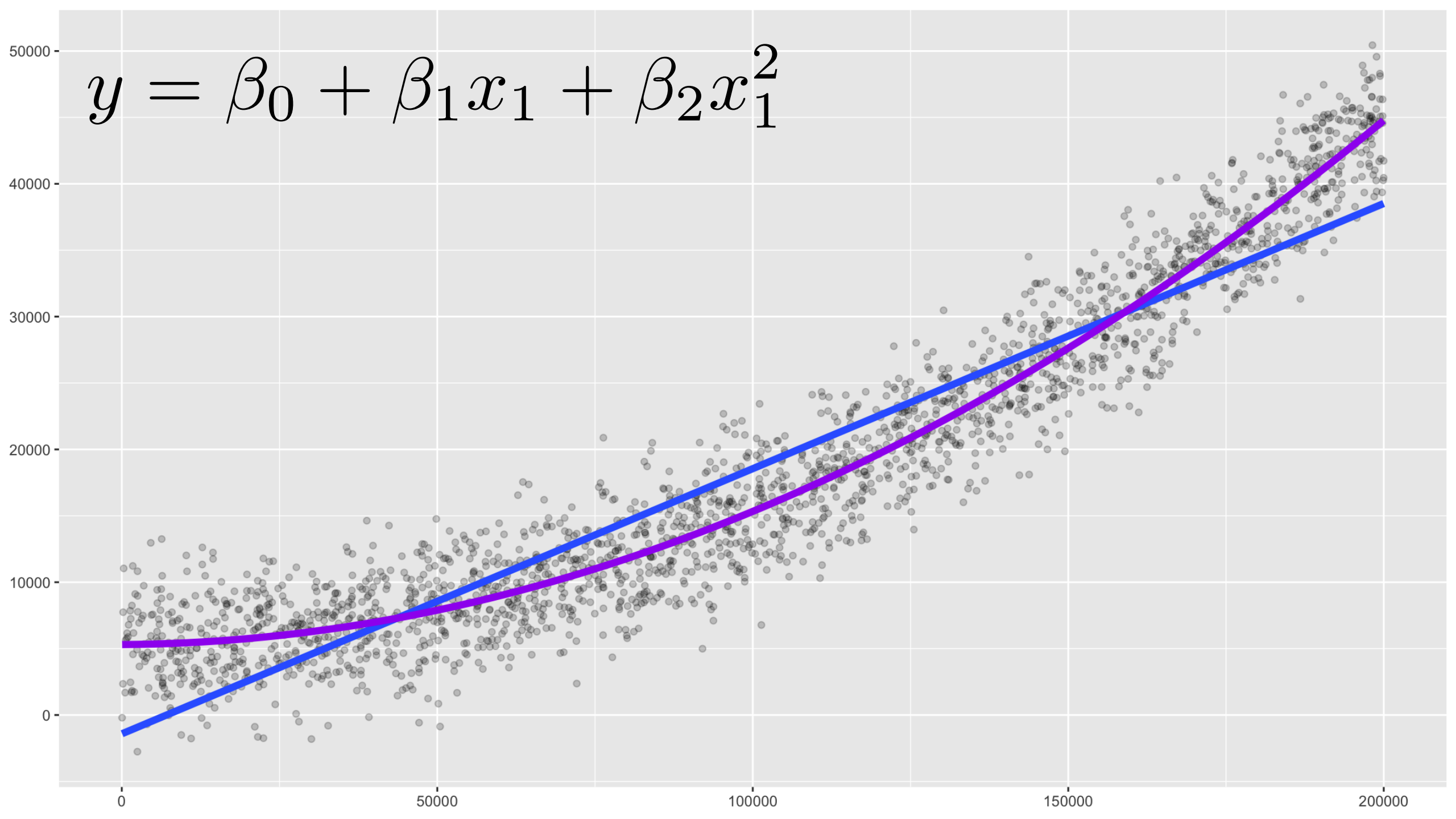
$$y = mx + b \qquad y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

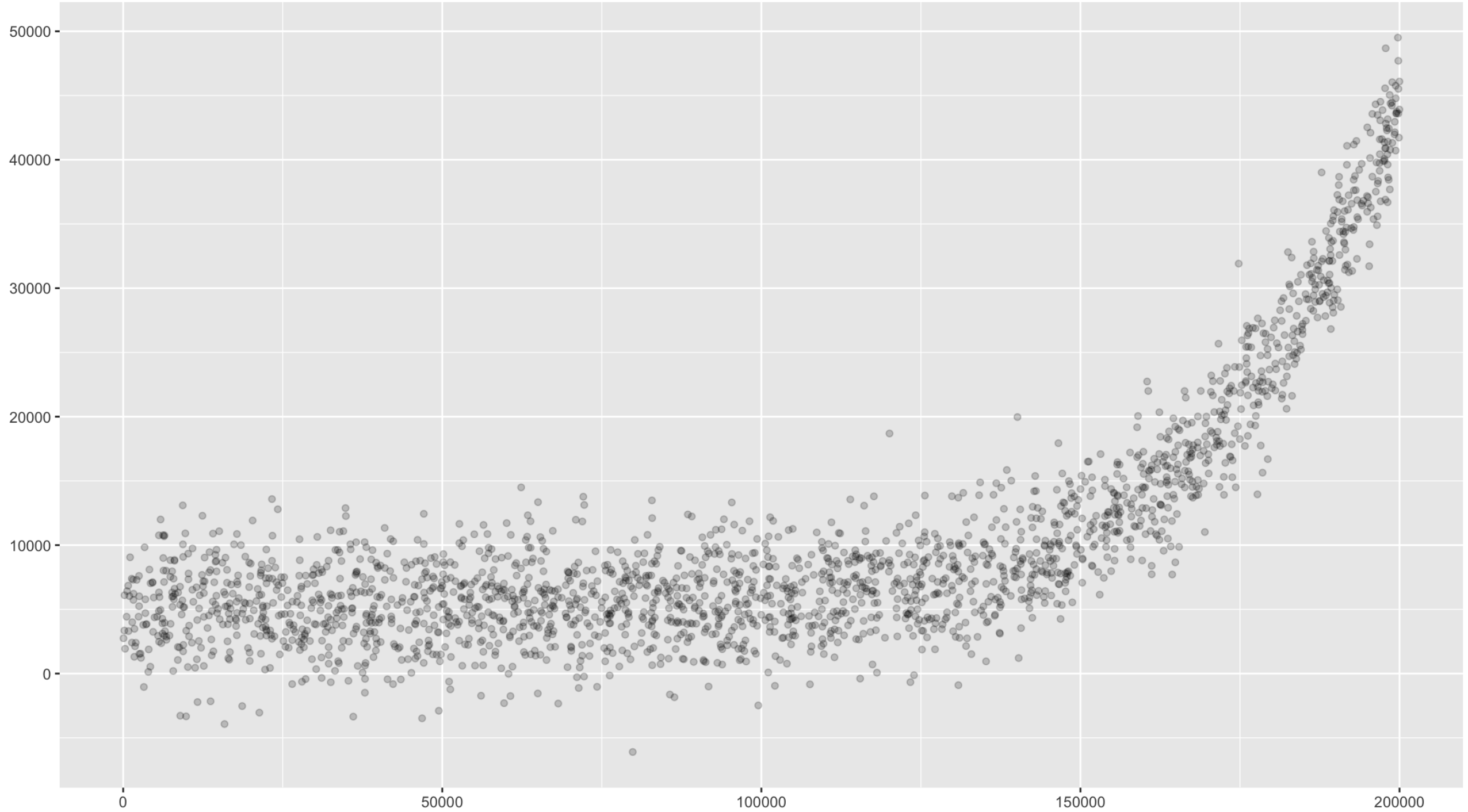
Not just for straight lines!
Make curvy with exponents

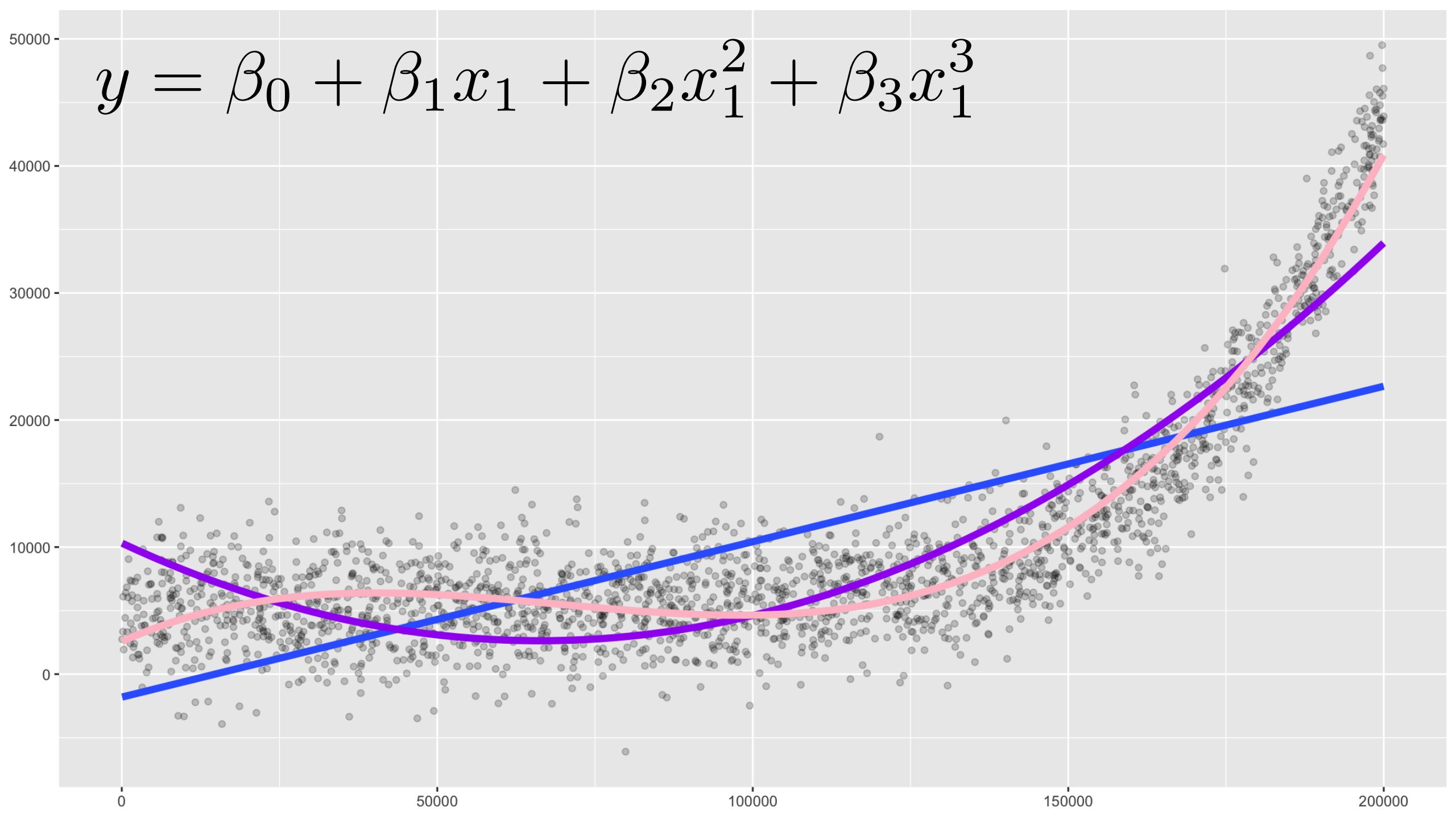
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$$


50000
40000
30000
20000
10000
0
0
50000
100000
150000
200000


$y = \beta_0 + \beta_1 x_1$
50000
40000
30000
20000
10000
0
0
50000
100000
150000
200000


$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2$
50000
40000
30000
20000
10000
0
0
50000
100000
150000
200000


50000
40000
30000
20000
10000
0
0
50000
100000
150000
200000

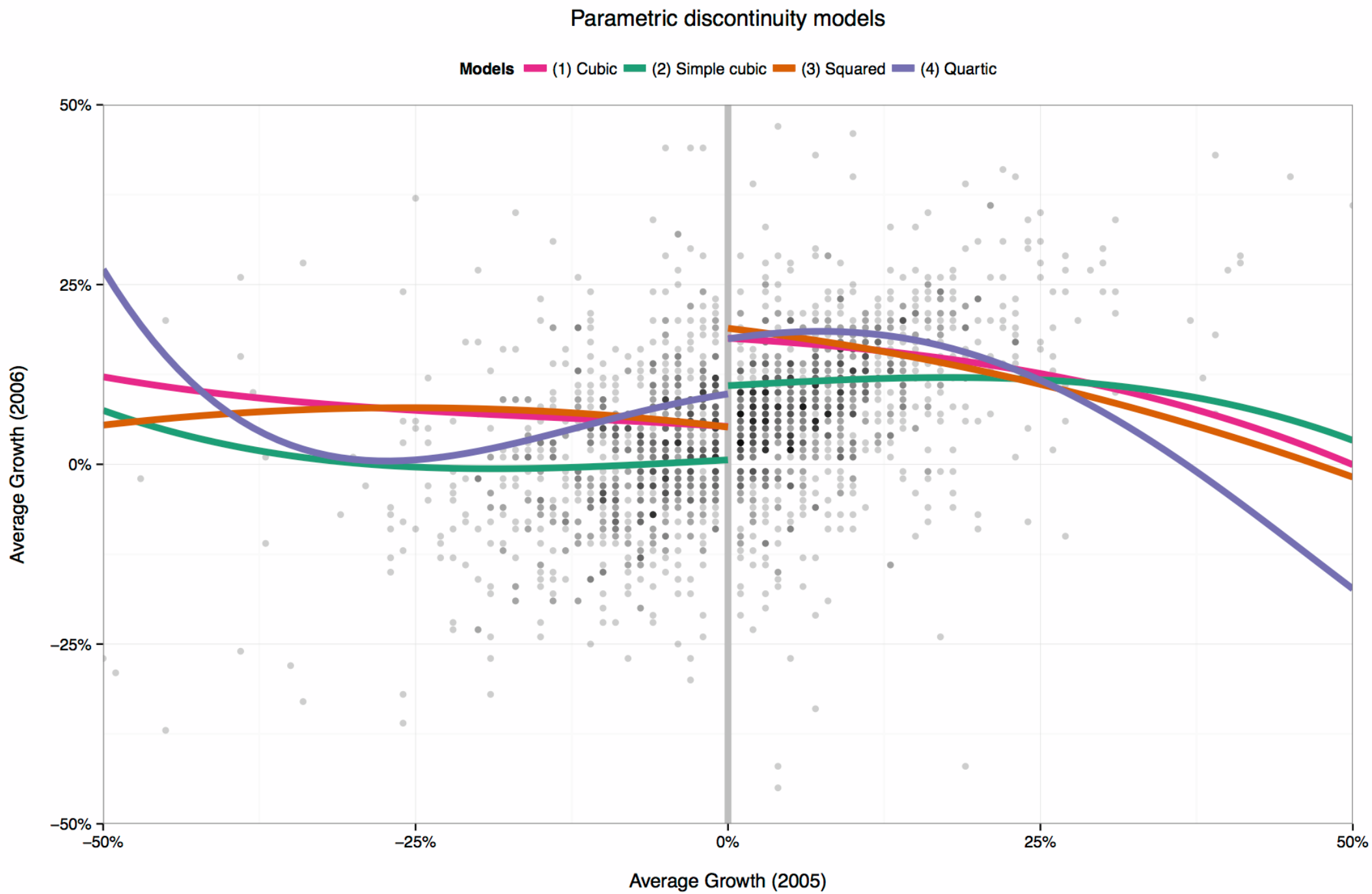

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_1^3$
50000
40000
30000
20000
10000
0
0
50000
100000
150000
200000

MEASURING THE GAP

$$y = \beta_0 + \beta_1 \text{Running variable (centered)} + \beta_2 \text{Indicator for treatment}$$

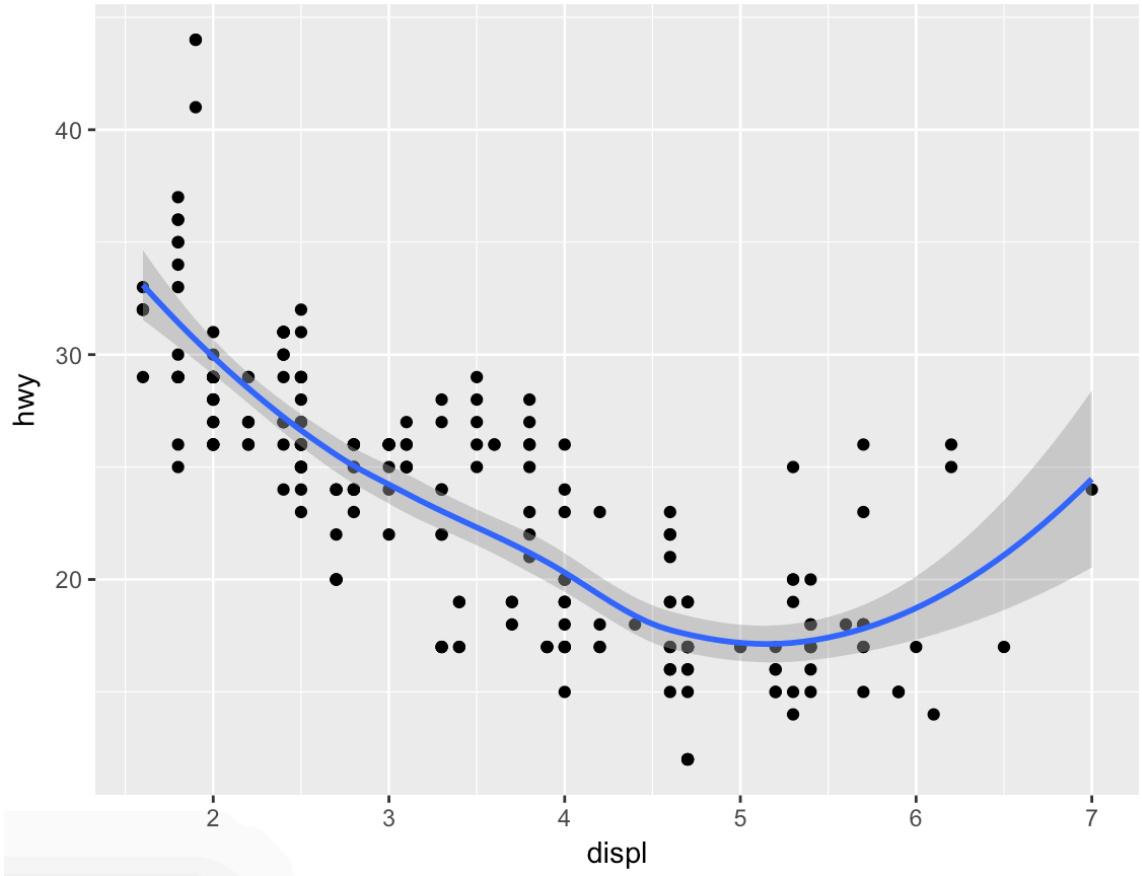
ID	outcome	running_var	running_var_centered	treatment
1	90.0	64	-6	FALSE
2	85.7	70	0	TRUE
3	85.8	73	3	TRUE
4	85.7	60	-10	FALSE
5	84.4	71	1	TRUE

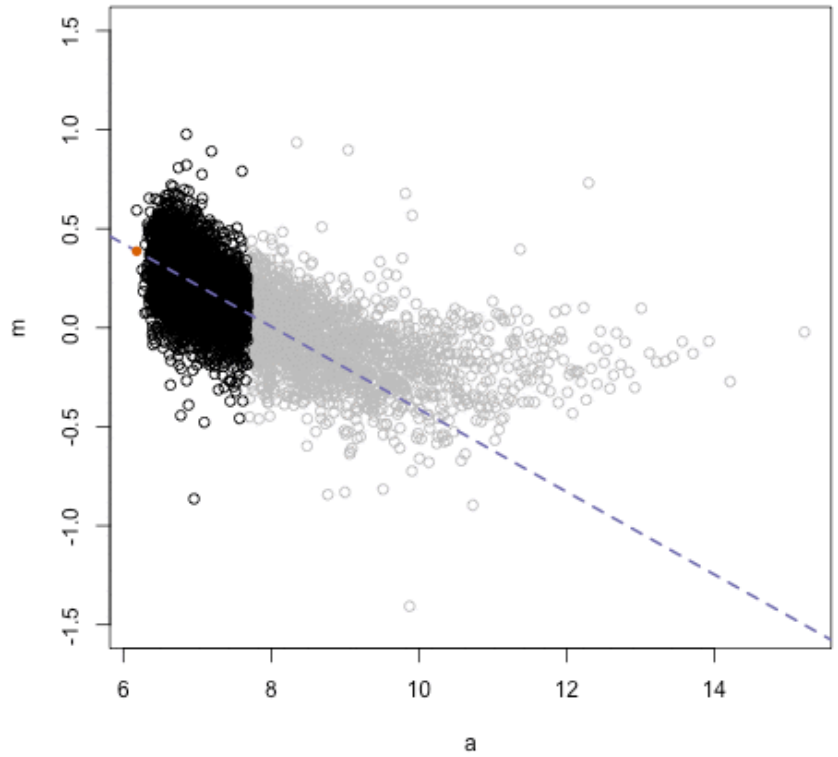
```
  term                 estimate std.error statistic p.value
  <chr>                   <dbl>     <dbl>     <dbl>   <dbl>
1 (Intercept)            93.7       3.24      28.9  0.00120
2 running_var_centered    0.726     0.386      1.88 0.201
3 treatmentTRUE          -9.33      3.82      -2.44 0.135
```

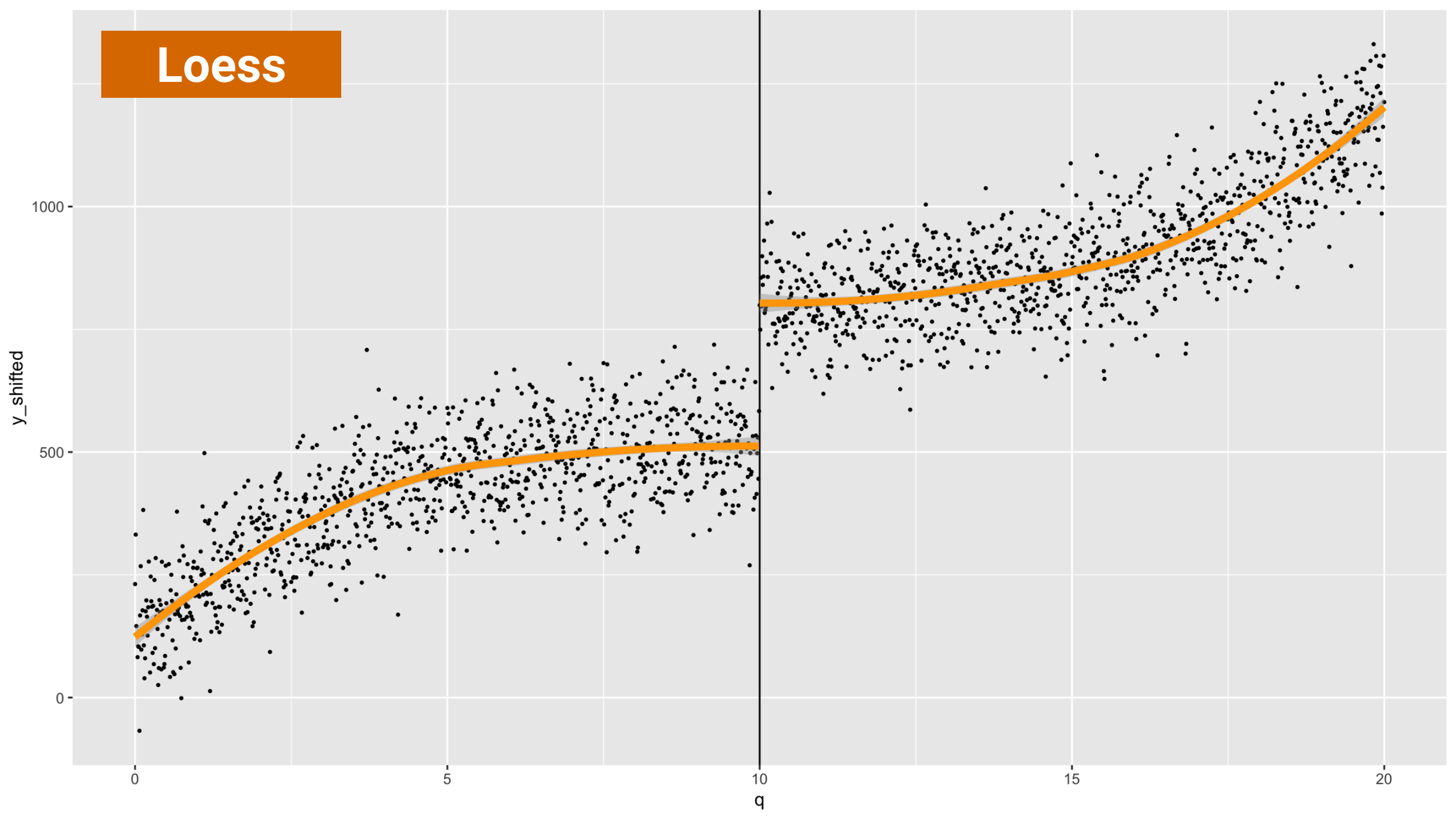

Parametric discontinuity models
Models
(1) Cubic
(2) Simple cubic
(3) Squared
(4) Quartic
50%
25%
0%
−25%
−50%
Average Growth (2006)
−50%
−25%
0%
25%
50%
Average Growth (2005)

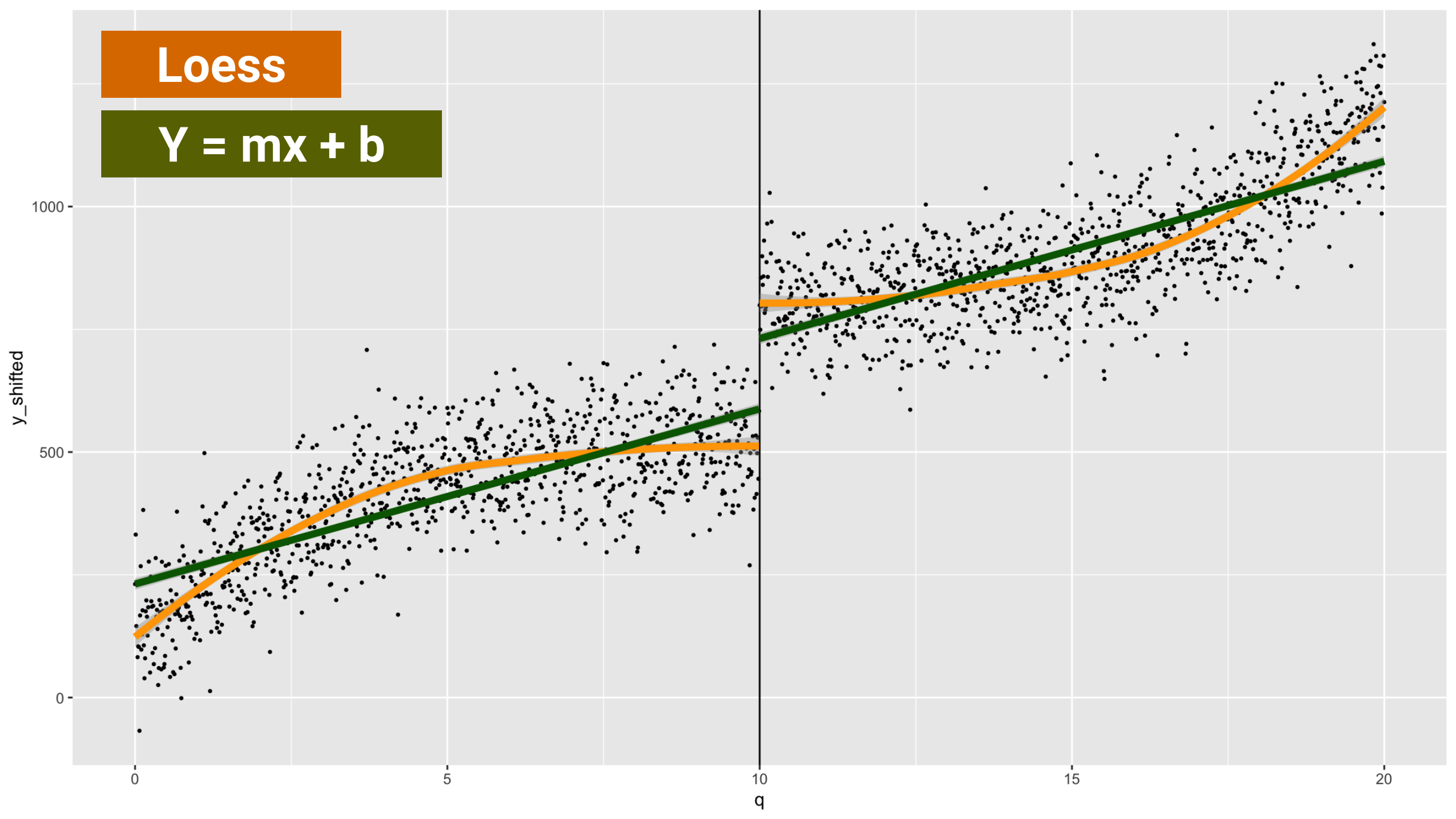
NONPARAMETRIC LINES

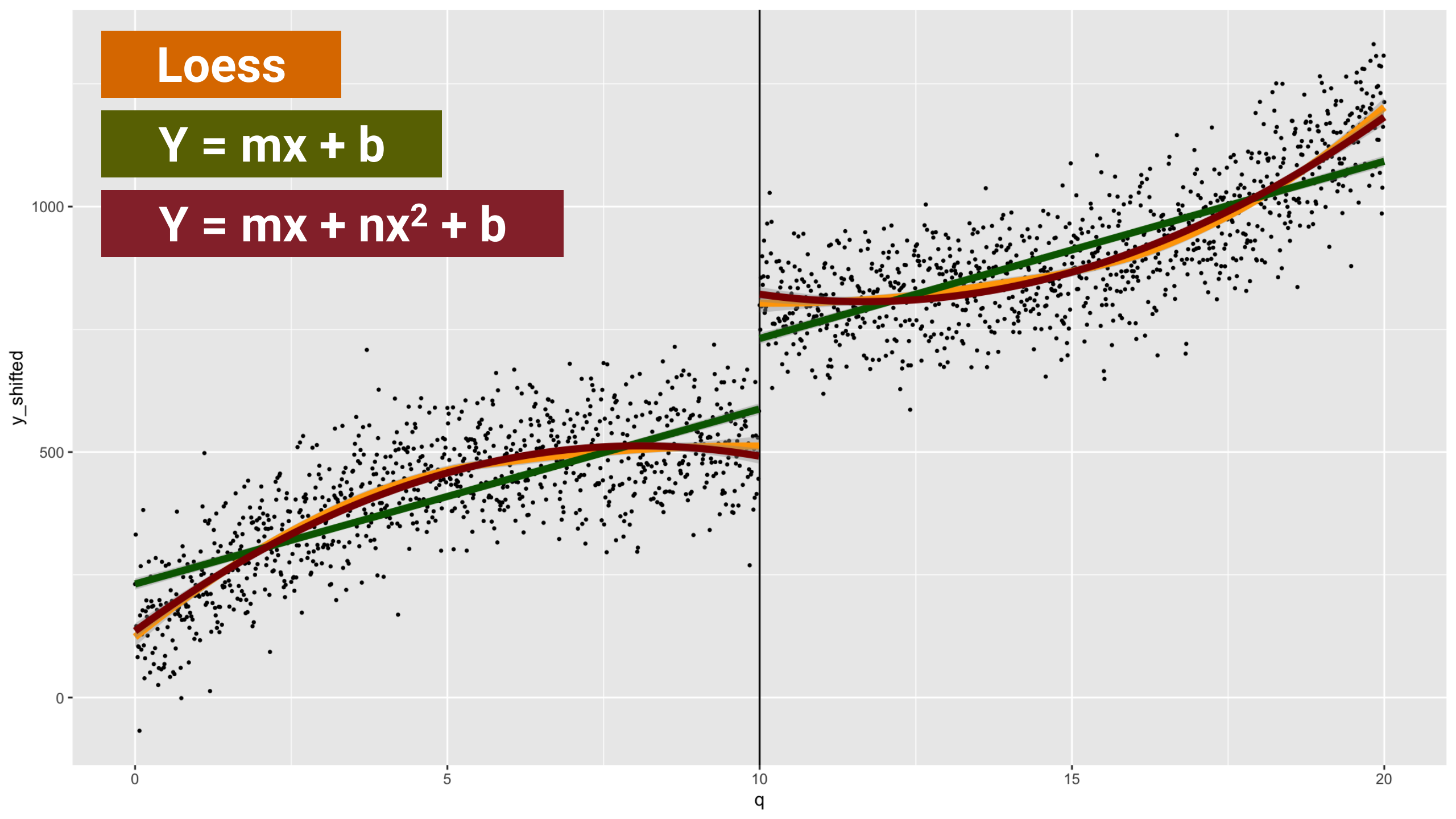
Lines without parameters






Loess
1000
500
0
y_shifted
0
5
10
15
20
q


Loess
Y = mx + b
1000
500
0
y_shifted
0
5
10
15
20
q

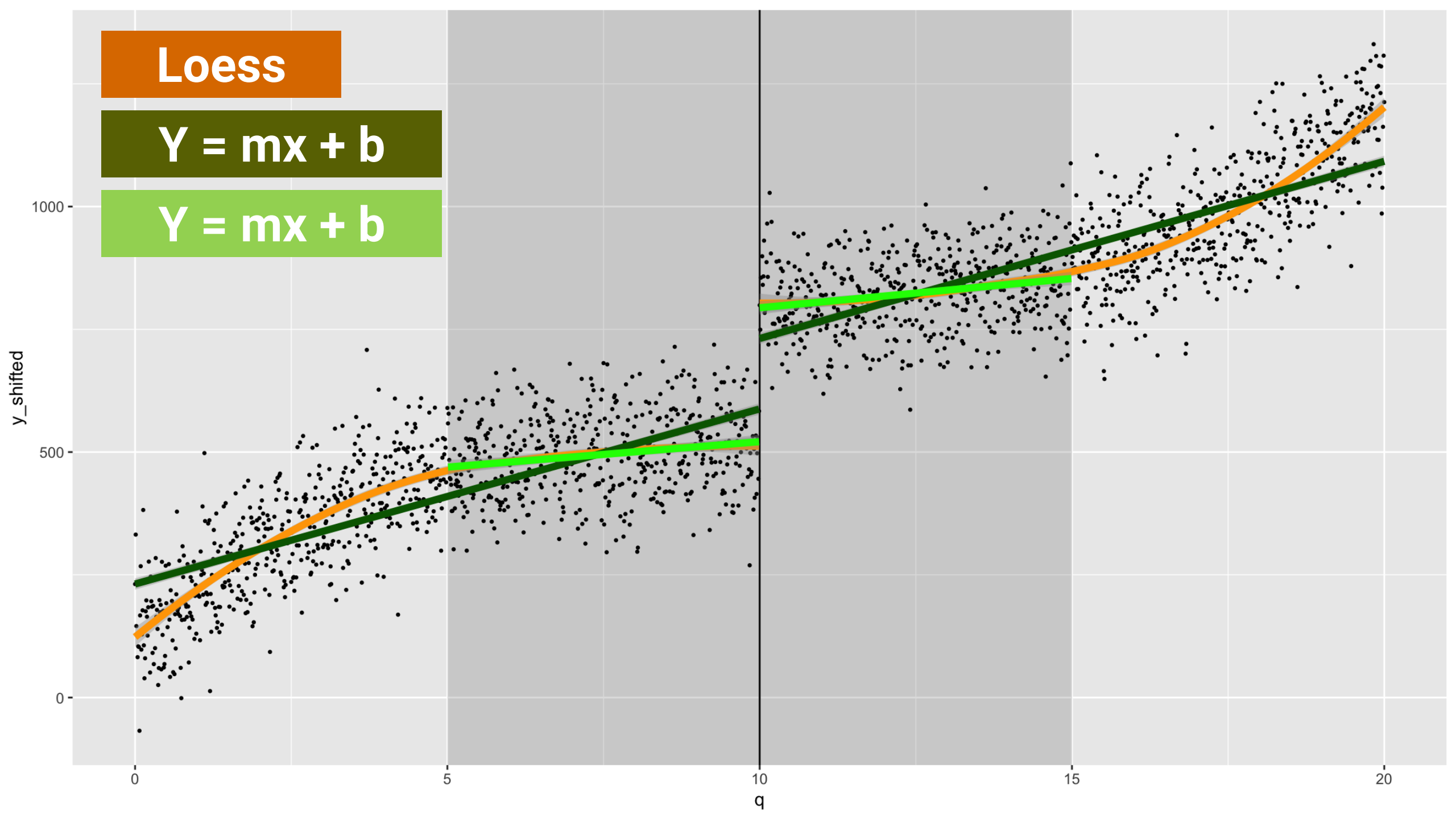

Loess
Y = mx + b
Y = mx + nx2 + b
1000
500
0
y_shifted
0
5
10
15
20
q

BANDWIDTHS

All you really care about is the area right around the cutoff

Observations far away from cutoff don't really matter

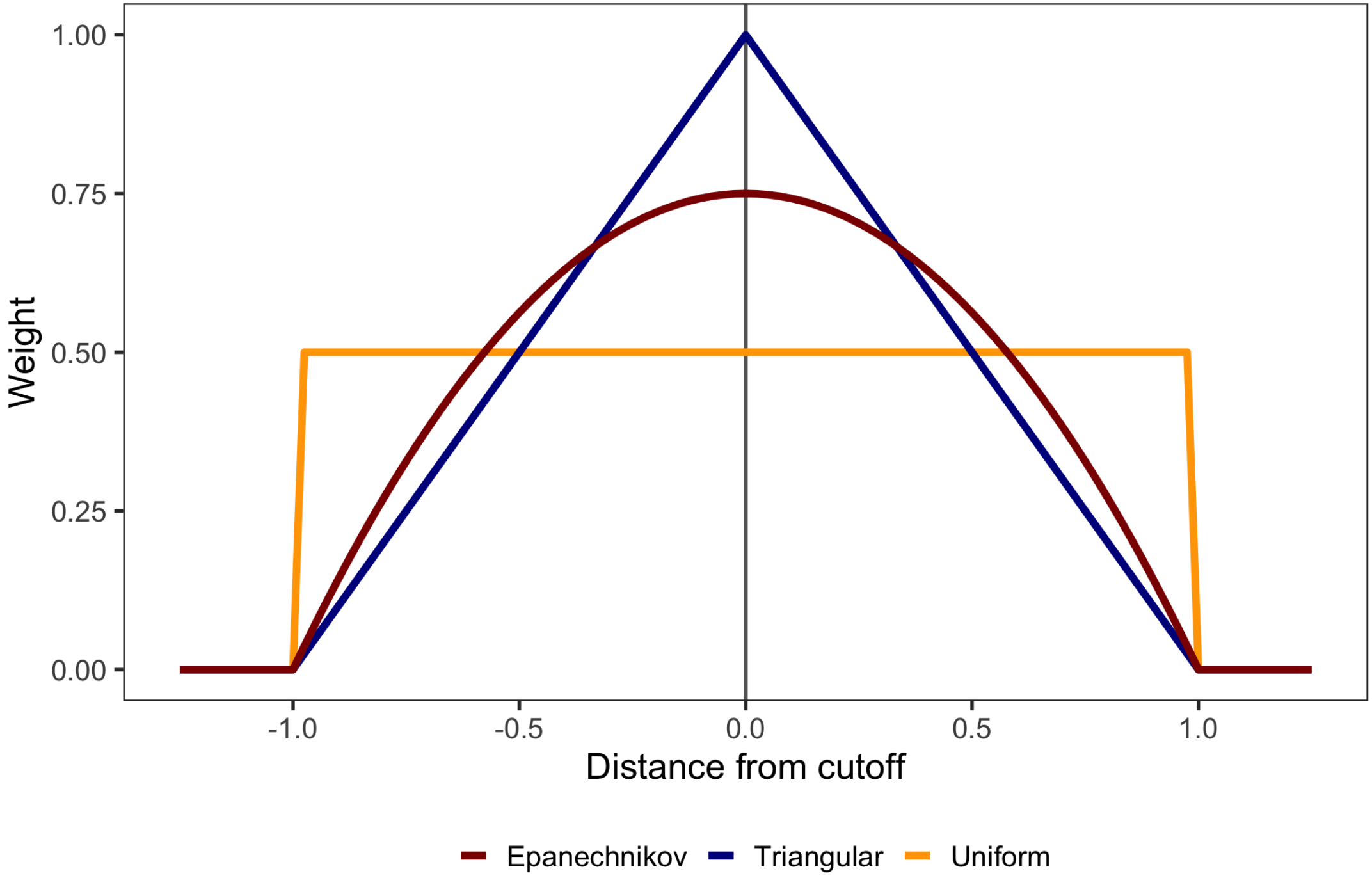
Bandwidth = window around cutoff where you focus your analysis


Loess
Y = mx + b
Y = mx + b
1000
500
0
y_shifted
0
5
10
15
20
q

KERNELS

You care the most about observations right by the cutoff, so give them extra weight

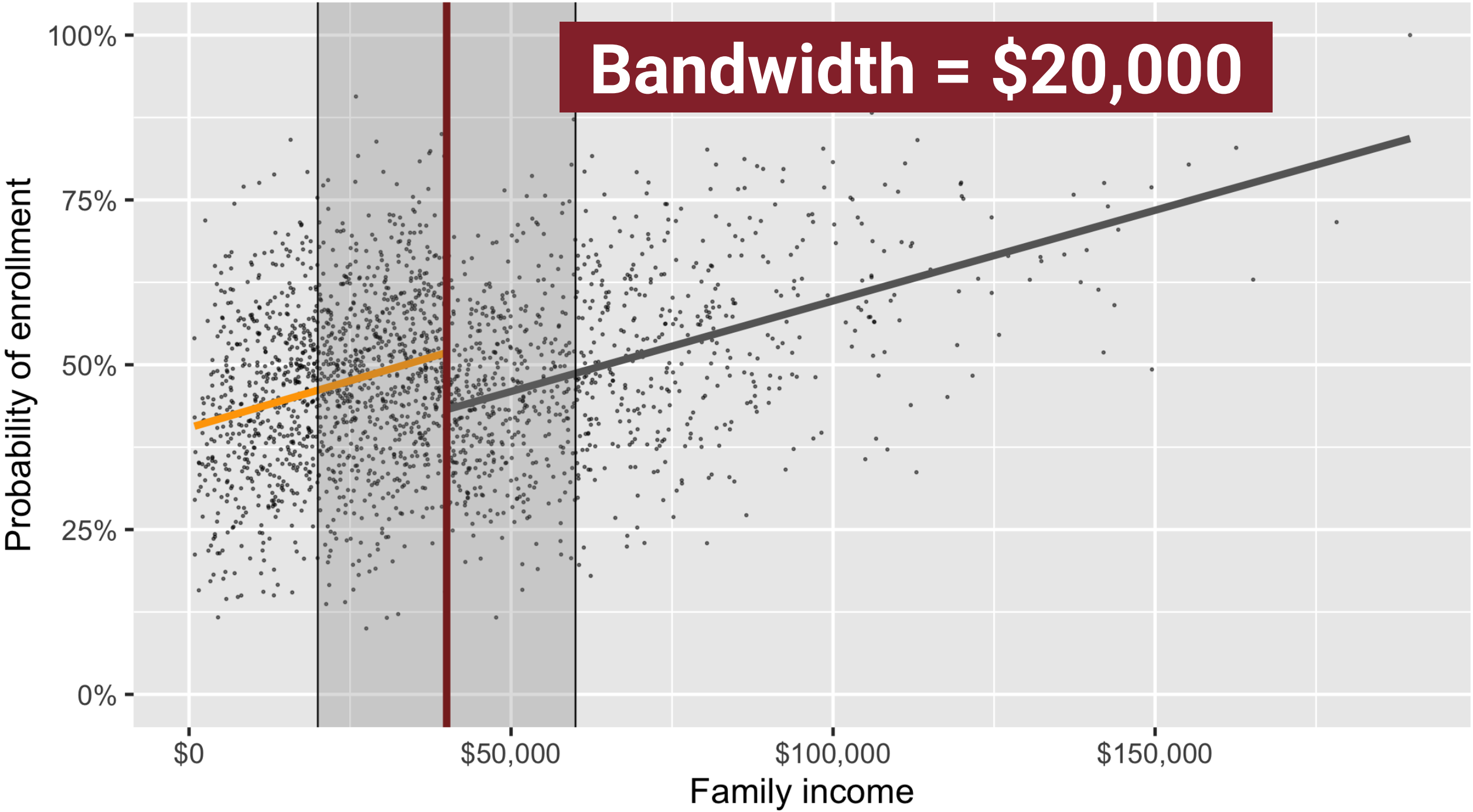
Kernel = method for assigning importance to values by distance to cutoff

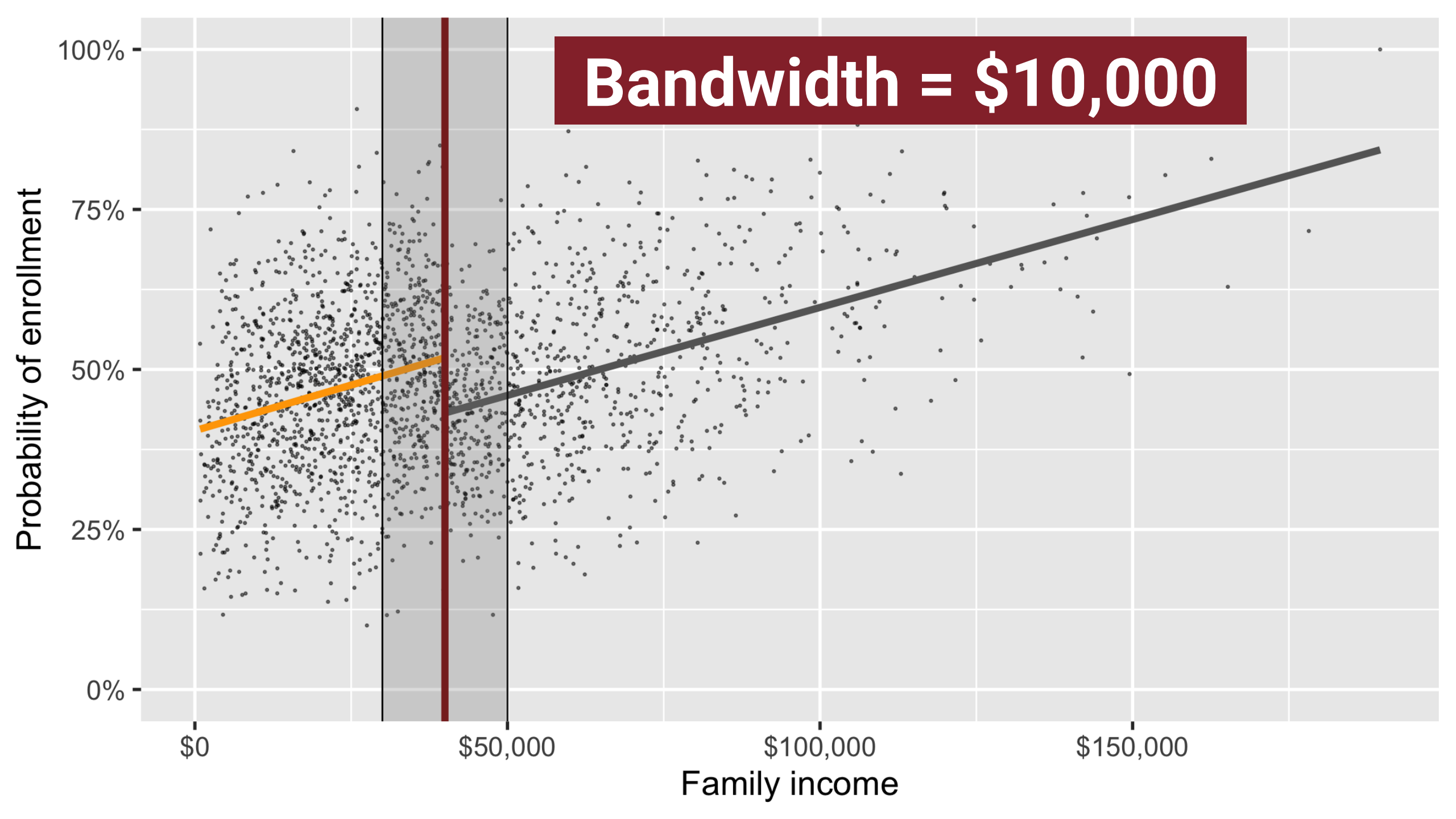

1.00
0.75
0.50
0.25
0.00
Weight
-1.0
-0.5
0.0
0.5
1.0
Distance from cutoff
Epanechnikov
Triangular
Uniform

MAIN RDD CONCERNS

GREEDY METHOD

You need lots of data, since you're throwing lots of it away


Bandwidth = $20,000
100%
75%
50%
25%
0%
Probability of enrollment
$0
$50,000
$100,000
$150,000
Family income


Bandwidth = $10,000
100%
75%
50%
25%
0%
Probability of enrollment
$0
$50,000
$100,000
$150,000
Family income

LATE VS. ATE

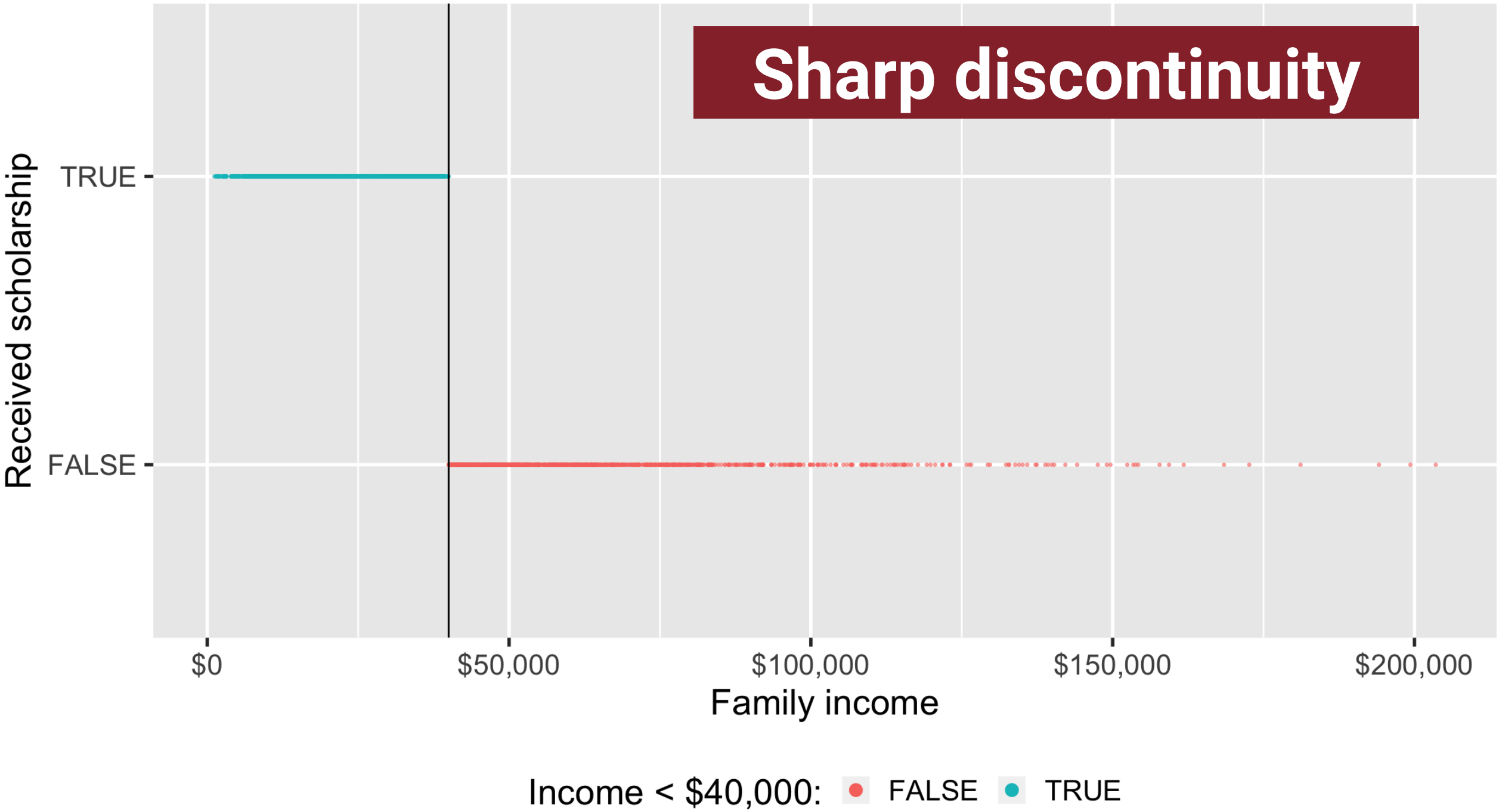
You're only measuring the ATE for people in the bandwidth

Local Average Treatment Effect (LATE)

NONCOMPLIANCE

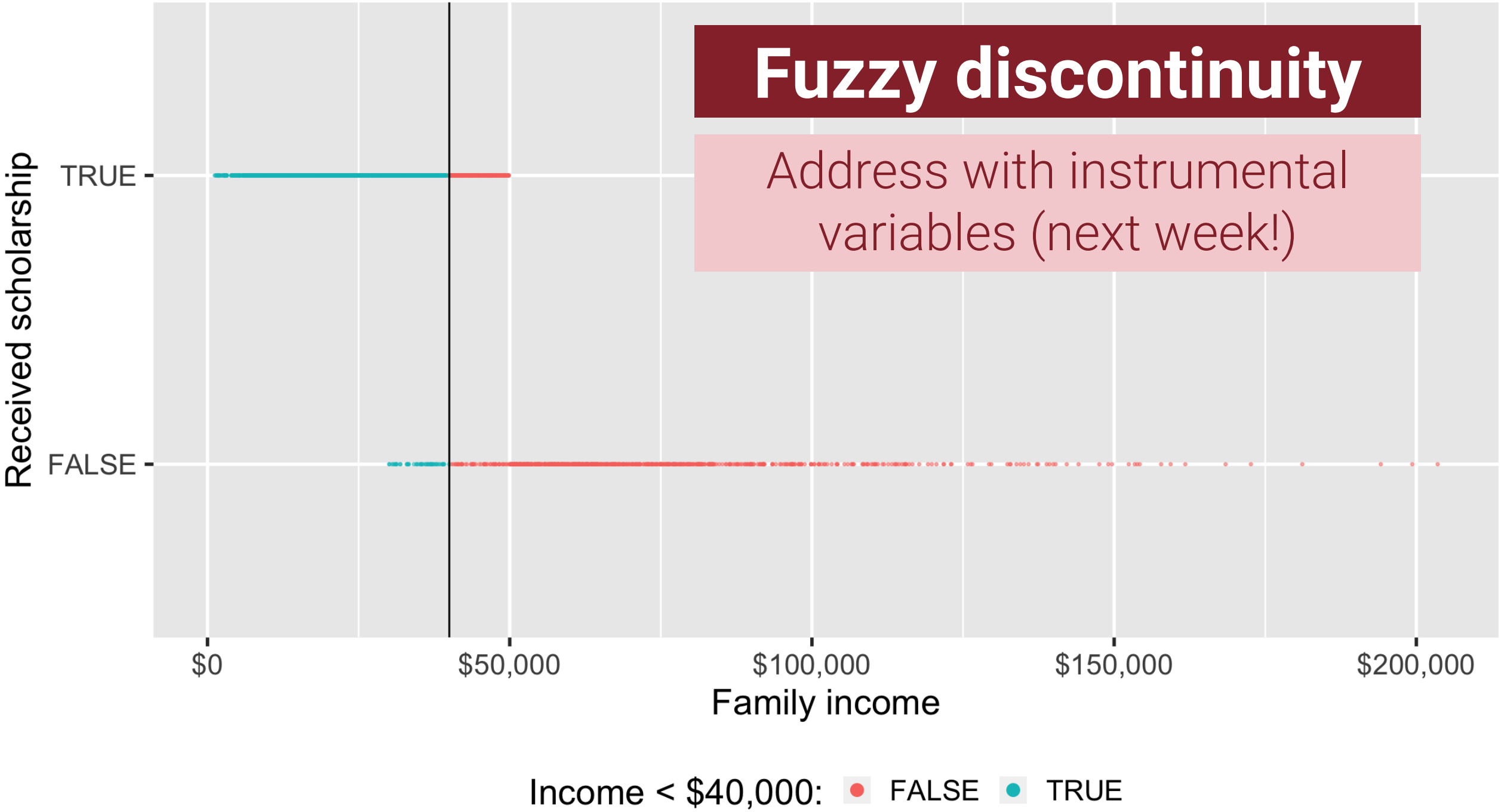
People on the margin of the discontinuity might end up in/out of the program

The ACA, Medicaid, and 138% of the poverty line


Sharp discontinuity
Received scholarship
TRUE
FALSE
$0
$50,000
$100,000
$150,000
$200,000
Family income
Income < $40,000:
FALSE
TRUE

Fuzzy discontinuity

Address with instrumental variables (next week!)



RDD WITH R

1: Is assignment to treatment rule-based?

If not, stop!

2: Is design fuzzy or sharp?

Either is fine; sharp is easier.

3: Is there a discontinuity in running variable at cutpoint?

Hopefully not.

4: Is there a discontinuity in outcome variable at cutpoint in running variable?

Hopefully.

5: How big is the gap?

Measure parametrically and nonparametrically.