Regression discontinuity I

Session 10

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

Arbitrary cutoffs and causal inference

Drawing lines and measuring gaps

Main RDD concerns

Arbitrary cutoffs and causal inference

Quasi-experiments again

Instead of using carefully adjusted DAGs, we can use *context* to isolate/identify the pathway between treatment and outcome in observational data

Diff-in-diff was one kind of quasi-experiment

Treatment/control + before/after

Regression discontinuity designs (RDD) are another

Arbitrary rules determine access to programs

Rules to access programs

Lots of policies and programs are based on arbitrary rules and thresholds

If you're above the threshold, you're in the program; if you're below, you're not (or vice versa)

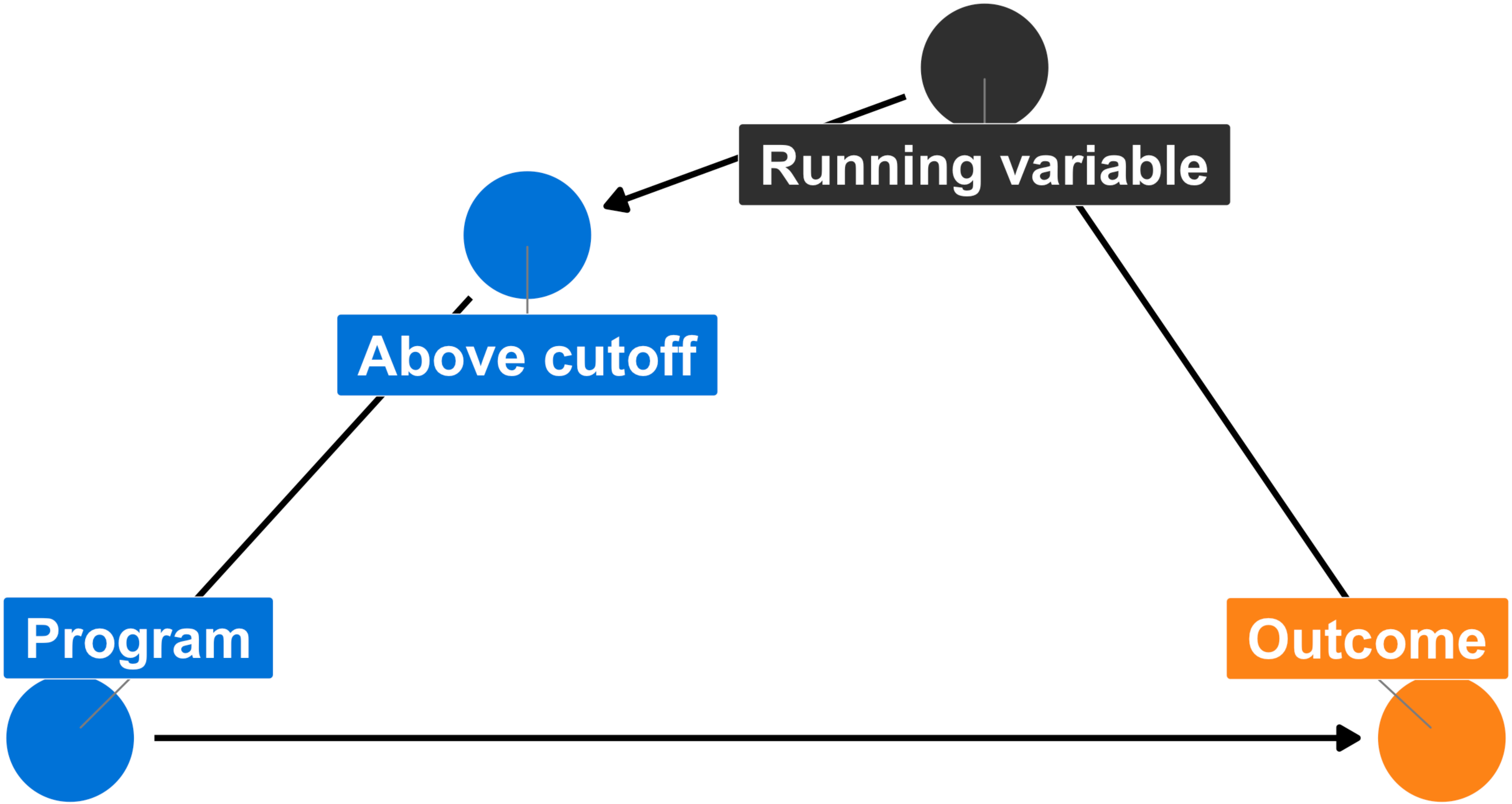
Key terms

Running / forcing variable

Index or measure that determines eligibility

Cutoff / cutpoint / threshold

Number that formally assigns access to program


Running variable
Above cutoff
Program
Outcome

Discontinuities everywhere!

Size	Annual	Monthly	138%	150%	200%
1	$12,760	$1,063	$17,609	$19,140	$25,520
2	$17,240	$1,437	$23,791	$25,860	$34,480
3	$21,720	$1,810	$29,974	$32,580	$43,440
4	$26,200	$2,183	$36,156	$39,300	$52,400
5	$30,680	$2,557	$42,338	$46,020	$61,360
6	$35,160	$2,930	$48,521	$52,740	$70,320
7	$39,640	$3,303	$54,703	$59,460	$79,280
8	$44,120	$3,677	$60,886	$66,180	$88,240

Medicaid
138%*

ACA subsidies
138–400%*

CHIP
200%

SNAP/Free lunch
130%

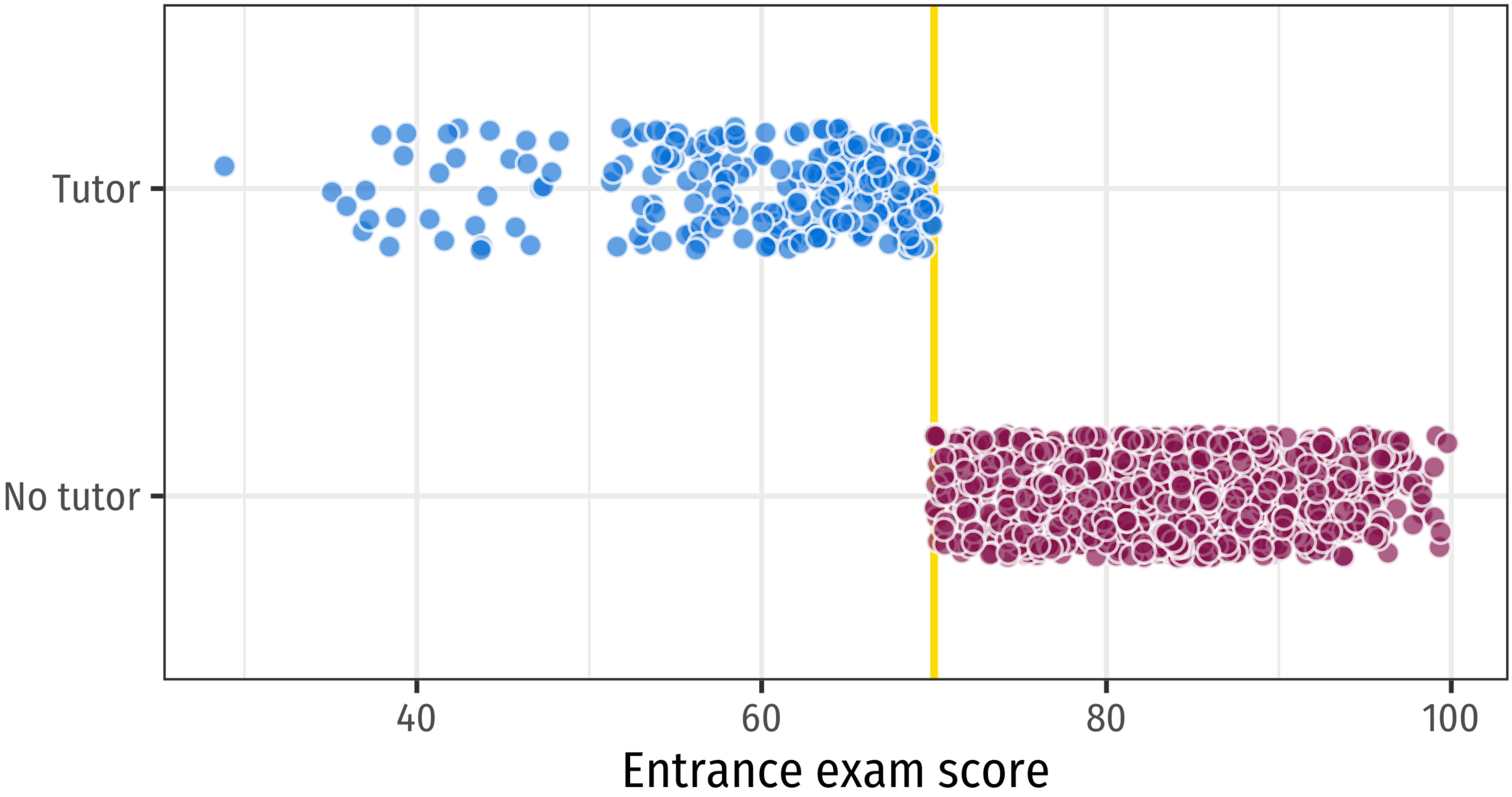
Reduced lunch
130–185%

Hypothetical tutoring program

Students take an entrance exam

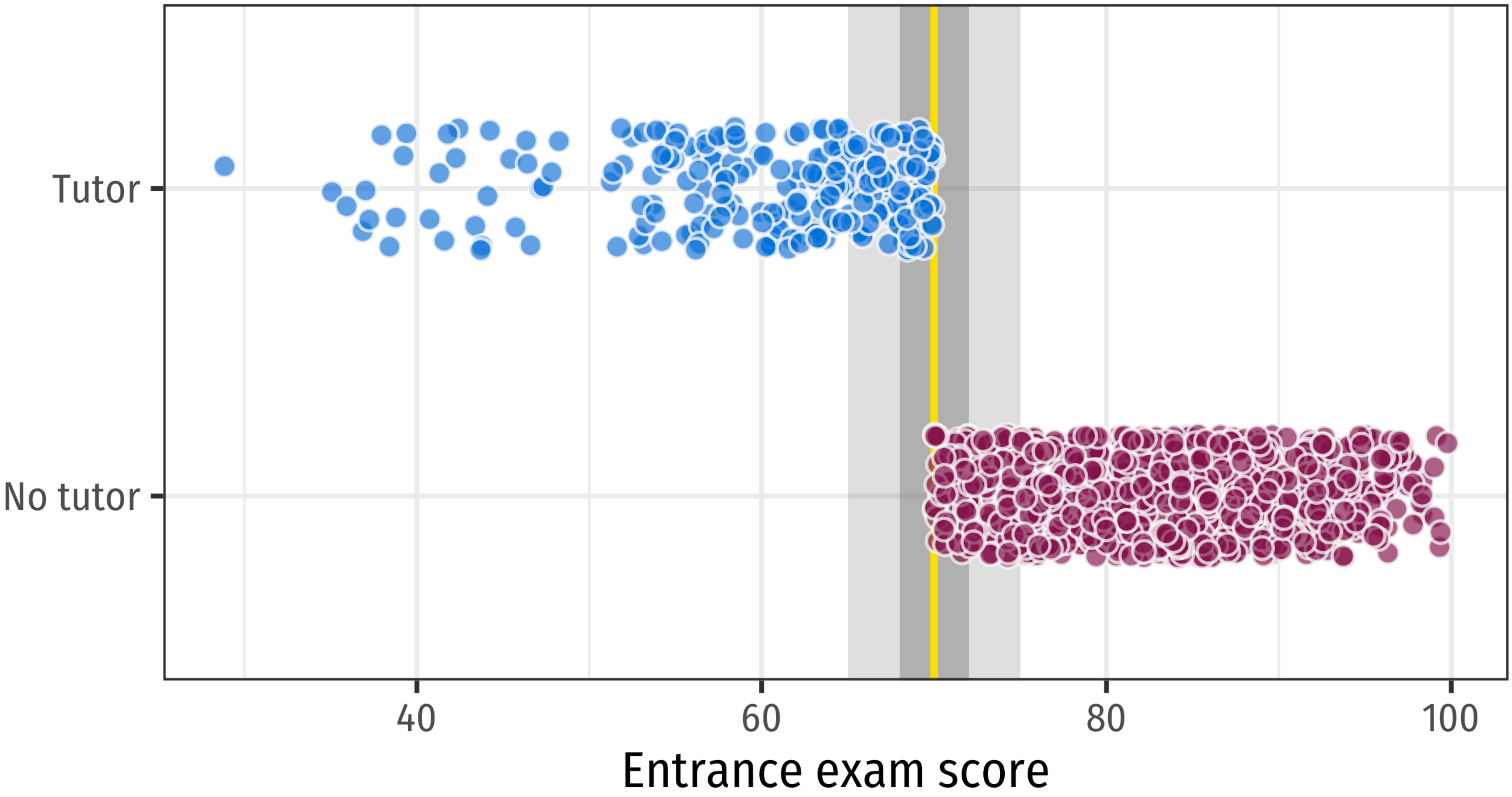
Those who score 70 or lower get a free tutor for the year

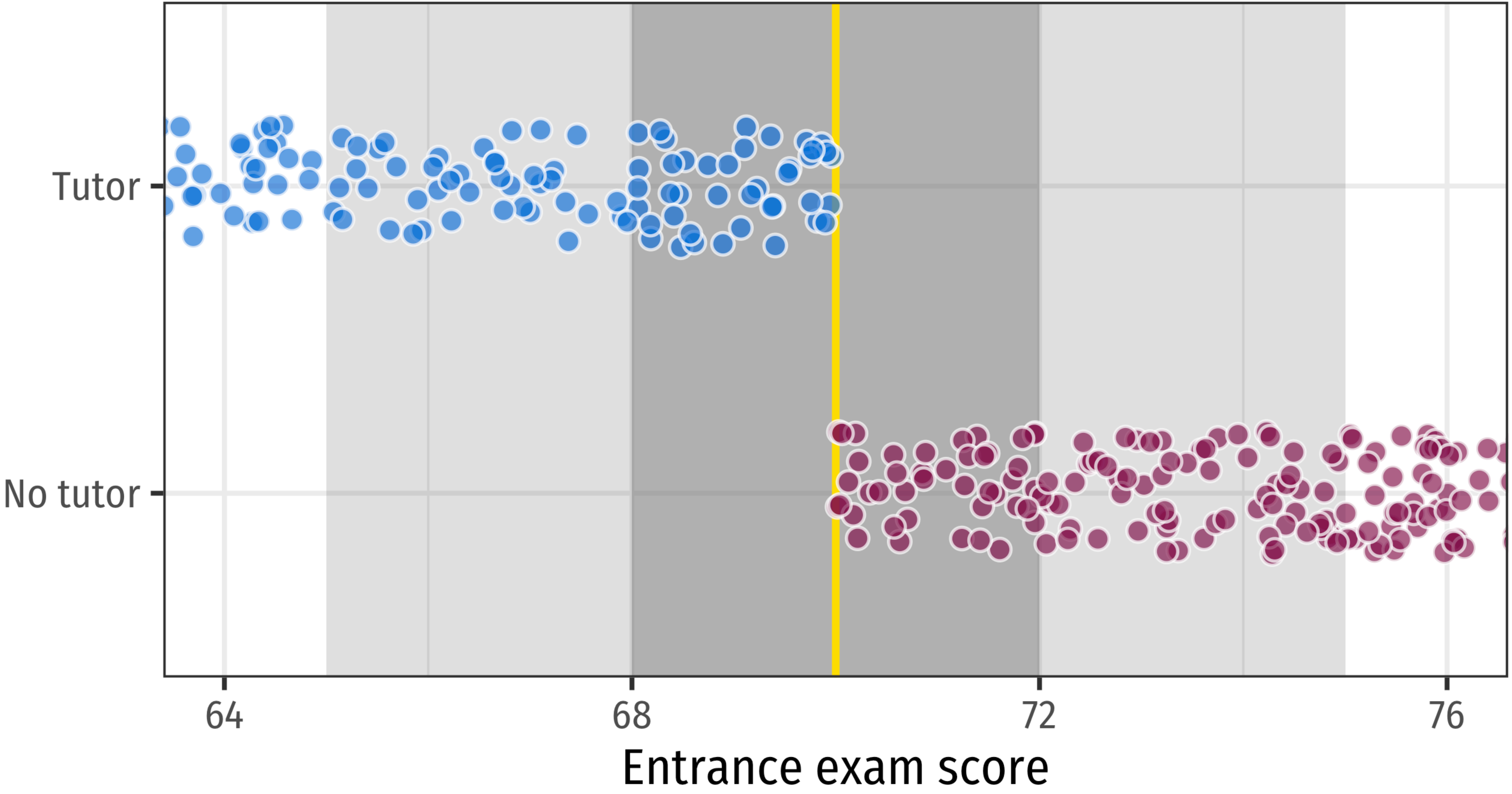
Students then take an exit exam at the end of the year


Tutor
No tutor
40
60
80
100
Entrance exam score

Causal inference intuition

The people right before and right after the threshold are essentially the same


Tutor
No tutor
40
60
80
100
Entrance exam score



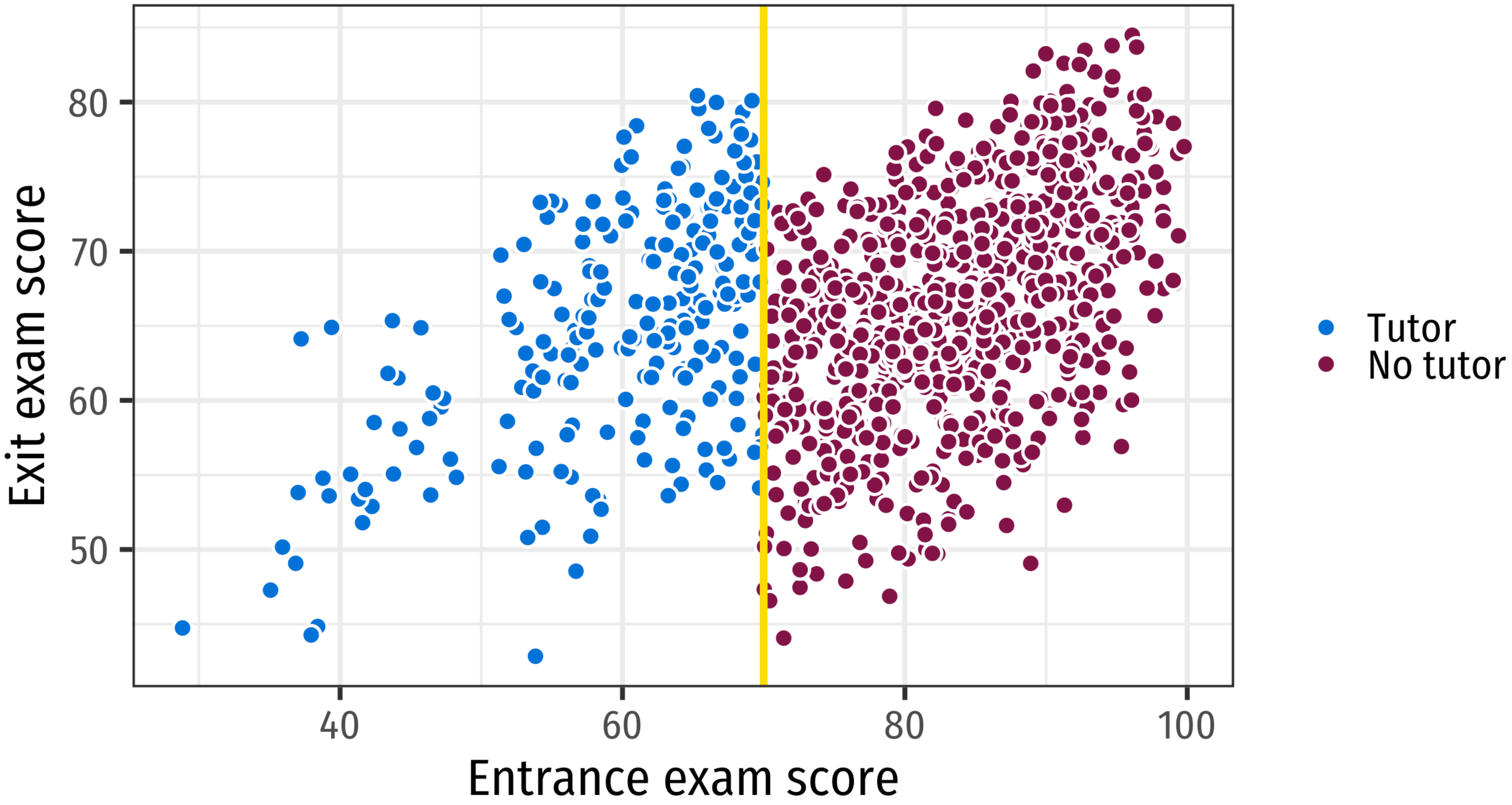
Tutor
No tutor
64
68
72
76
Entrance exam score

Causal inference intuition

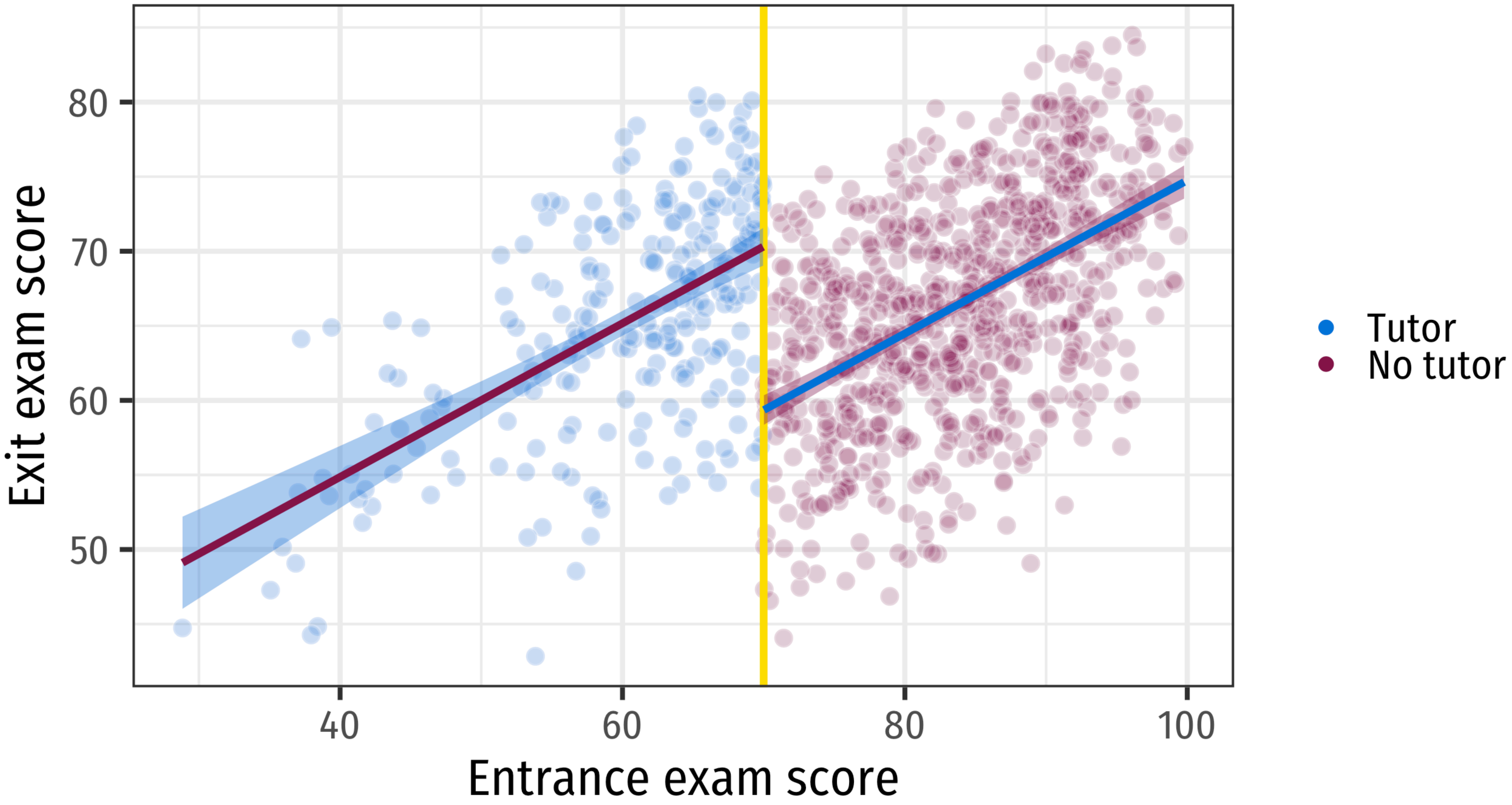
The people right before and right after the threshold are essentially the same

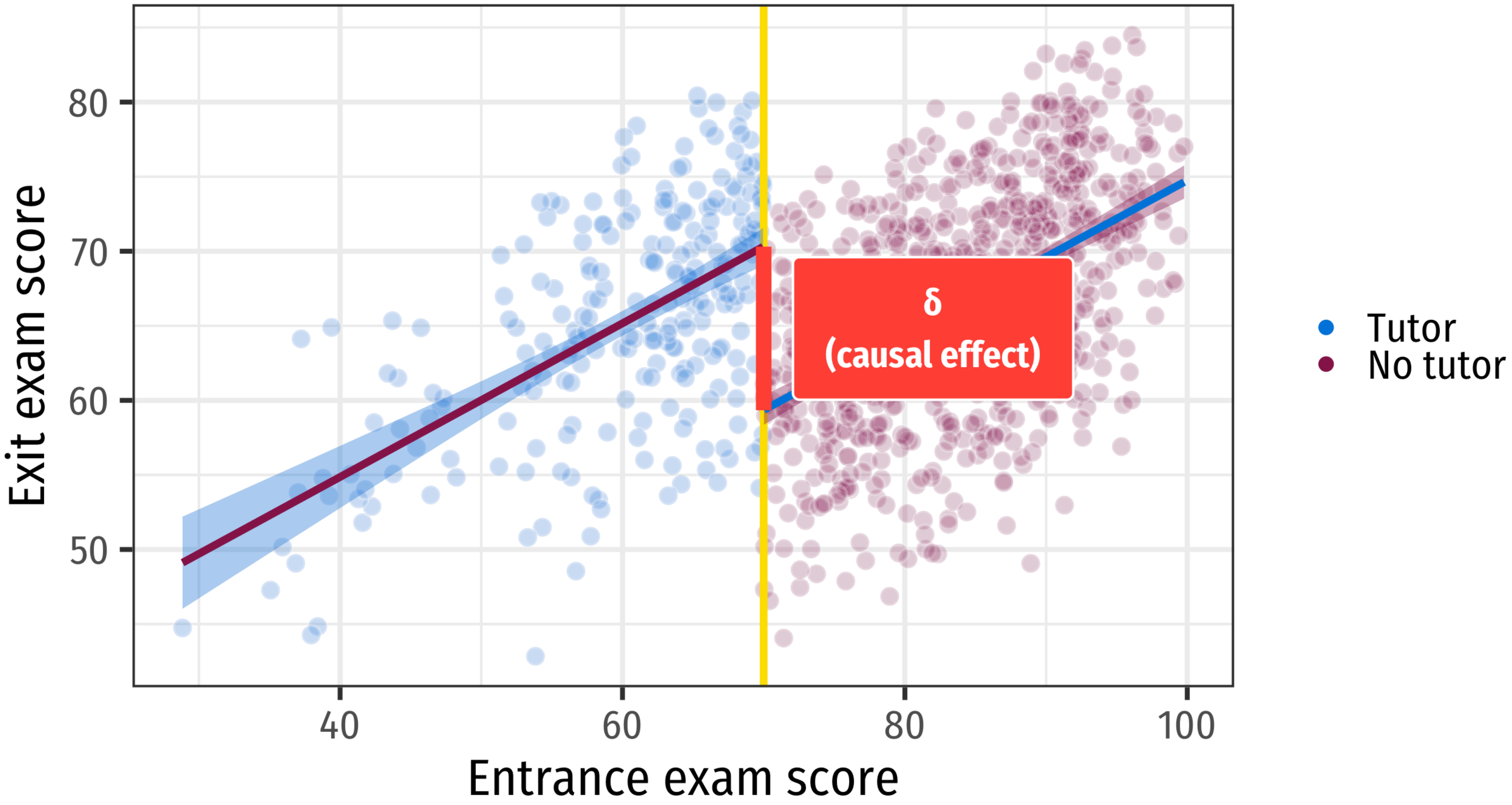
Pseudo treatment and control groups!

Compare outcomes for those right before/after, calculate difference

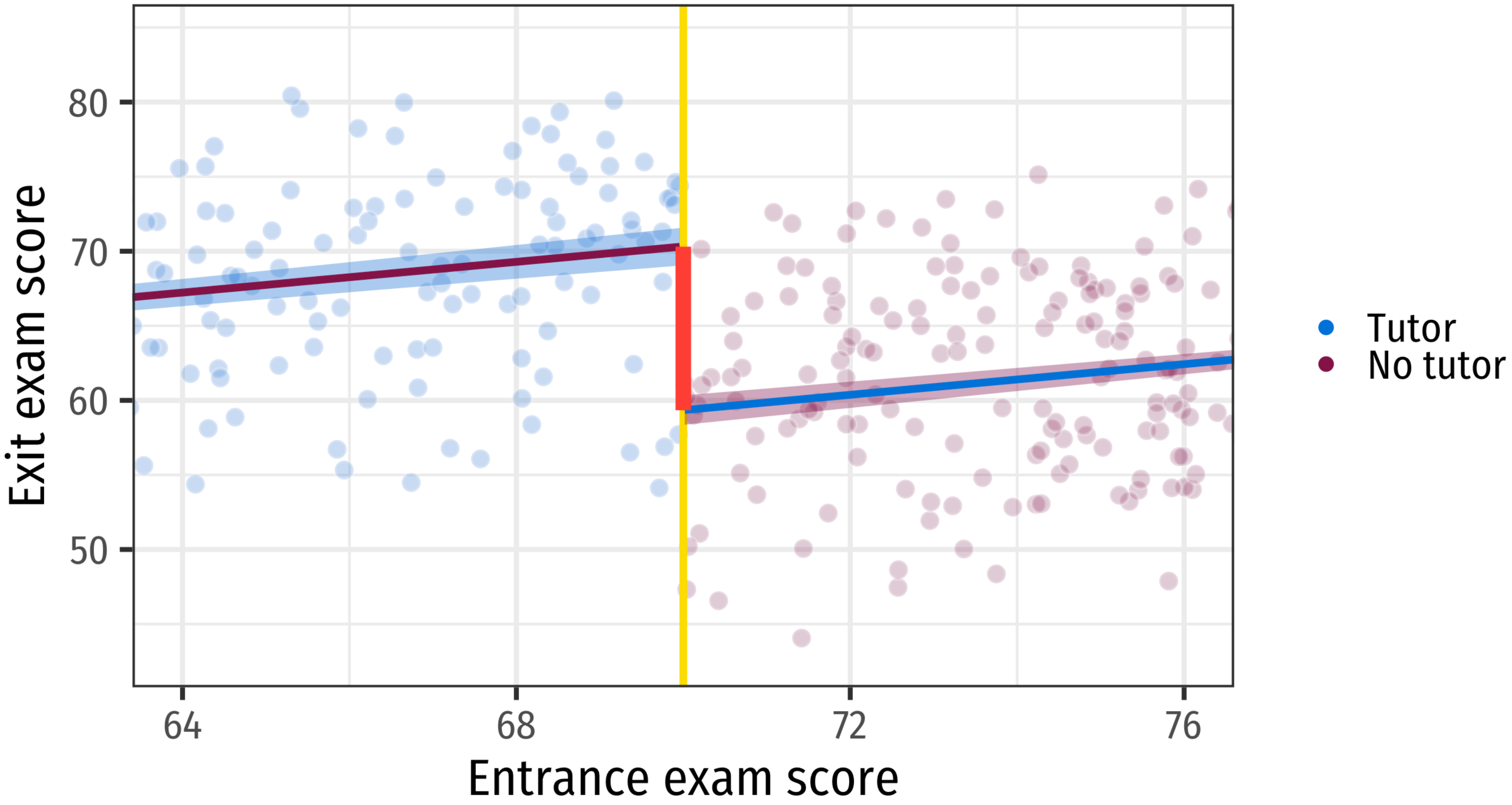


80
70
60
50
Exit exam score
40
60
80
100
Entrance exam score
Tutor
No tutor


80
70
60
50
Exit exam score
40
60
80
100
Entrance exam score
Tutor
No tutor

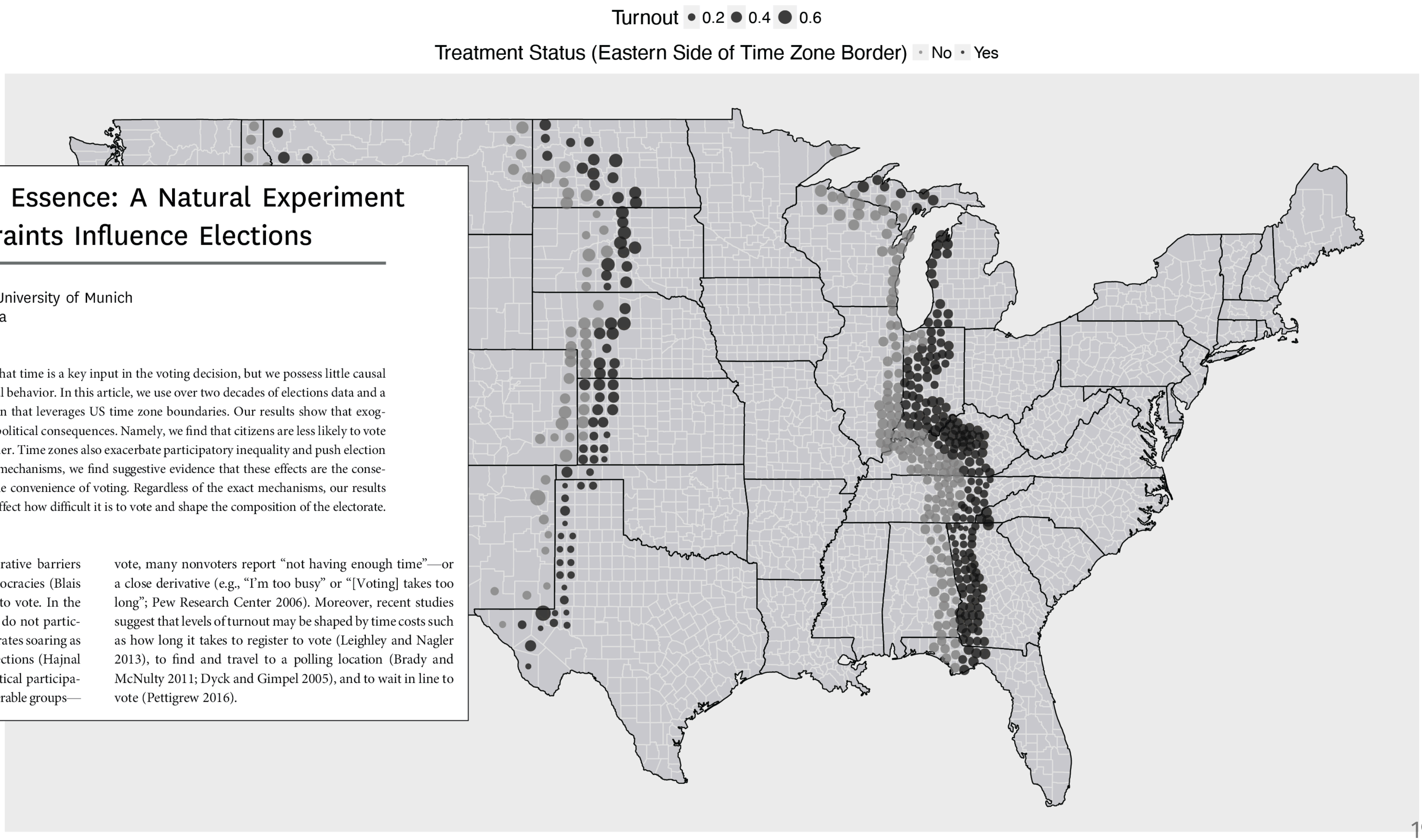


80
70
60
50
40
60
80
100
Exit exam score
Entrance exam score
δ
(causal effect)
Tutor
No tutor



80
70
60
50
Exit exam score
64
68
72
76
Entrance exam score
Tutor
No tutor

Geographic discontinuities



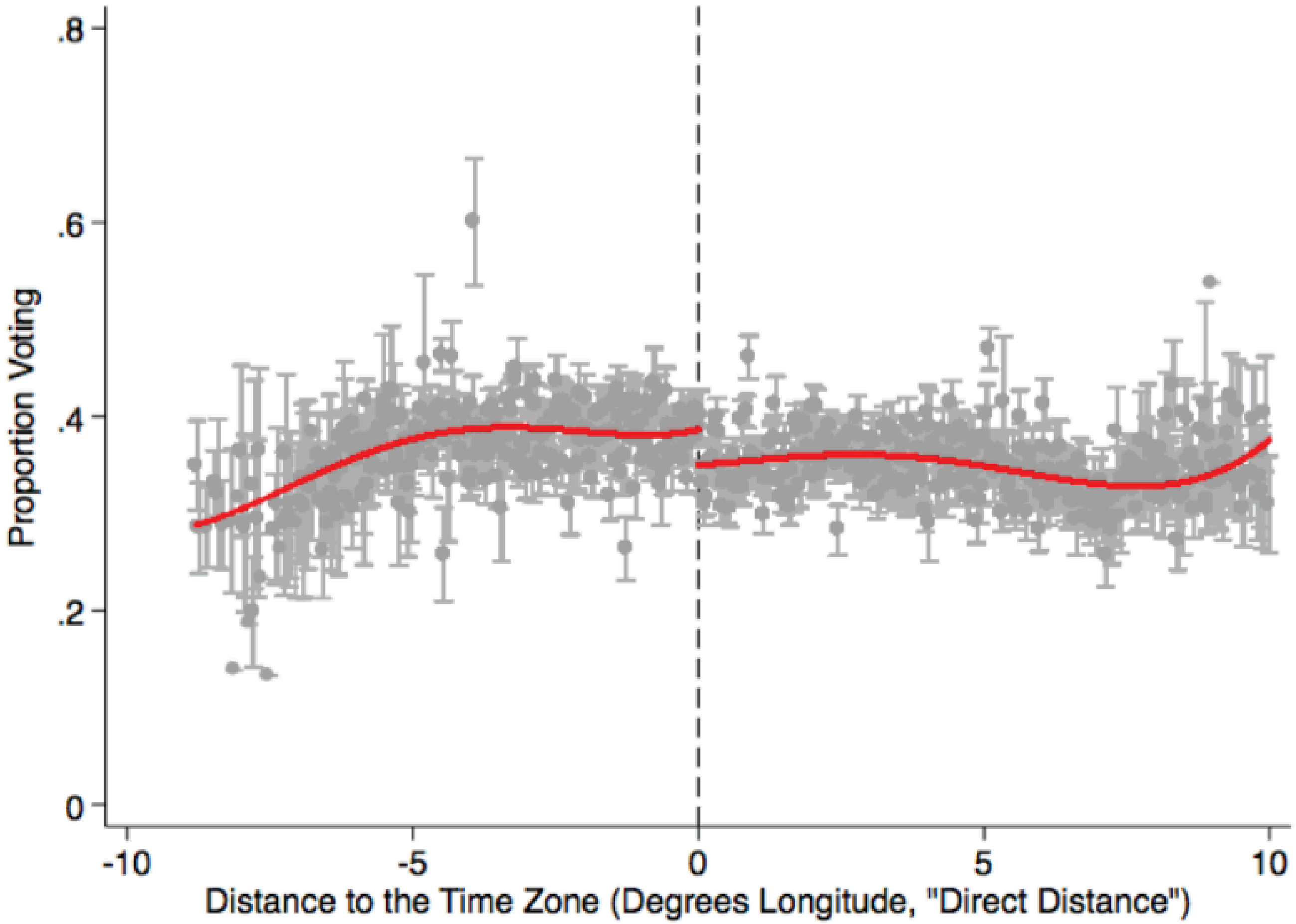
When Time Is of the Essence: A Natural Experiment on How Time Constraints Influence Elections

Jerome Schafer, Ludwig Maximilian University of Munich
John B. Holbein, University of Virginia

Foundational theories of voter turnout suggest that time is a key input in the voting decision, but we possess little causal evidence about how this resource affects electoral behavior. In this article, we use over two decades of elections data and a novel geographic regression discontinuity design that leverages US time zone boundaries. Our results show that exogenous shifts in time allocations have significant political consequences. Namely, we find that citizens are less likely to vote if they live on the eastern side of a time zone border. Time zones also exacerbate participatory inequality and push election results toward Republicans. Exploring potential mechanisms, we find suggestive evidence that these effects are the consequence of insufficient sleep and moderated by the convenience of voting. Regardless of the exact mechanisms, our results indicate that local differences in daily schedules affect how difficult it is to vote and shape the composition of the electorate.

Although in recent years the administrative barriers to voting have declined in many democracies (Blais 2010), many eligible citizens still fail to vote. In the United States, about 40% of registered voters do not participate in presidential elections, with abstention rates soaring as high as 60% in midterms and 70% in local elections (Hajnal and Trounstine 2016). Moreover, rates of political participation have remained stubbornly low among vulnerable groups—vote, many nonvoters report "not having enough time"—or a close derivative (e.g., "I'm too busy" or "[Voting] takes too long"; Pew Research Center 2006). Moreover, recent studies suggest that levels of turnout may be shaped by time costs such as how long it takes to register to vote (Leighley and Nagler 2013), to find and travel to a polling location (Brady and McNulty 2011; Dyck and Gimpel 2005), and to wait in line to vote (Pettigrew 2016).

Geographic discontinuities



Lower turnout in counties on the eastern side of the boundary

Election schedules cause fluctuations in turnout

Time discontinuities

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

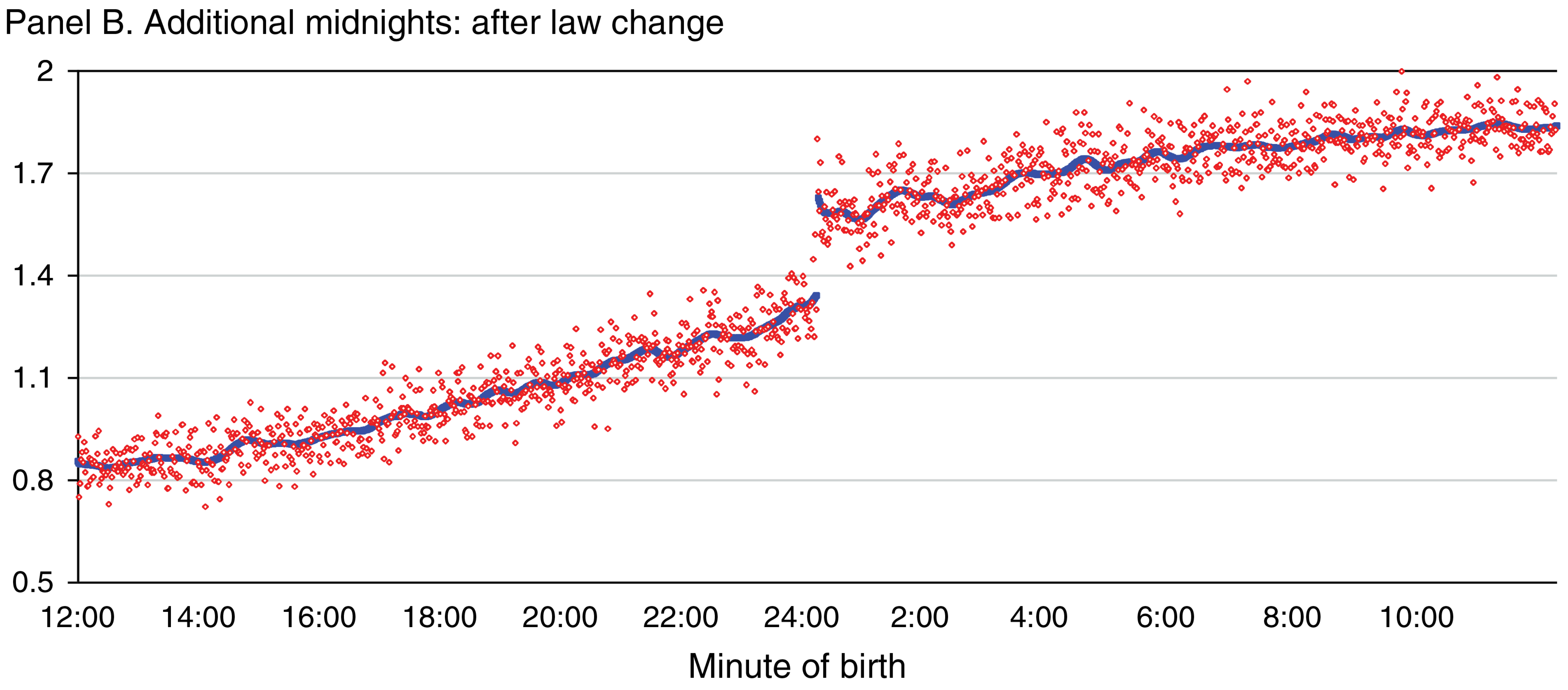
By Douglas Almond and Joseph J. Doyle Jr.*

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. (*JEL* D82, G22, I12, I18, J13)

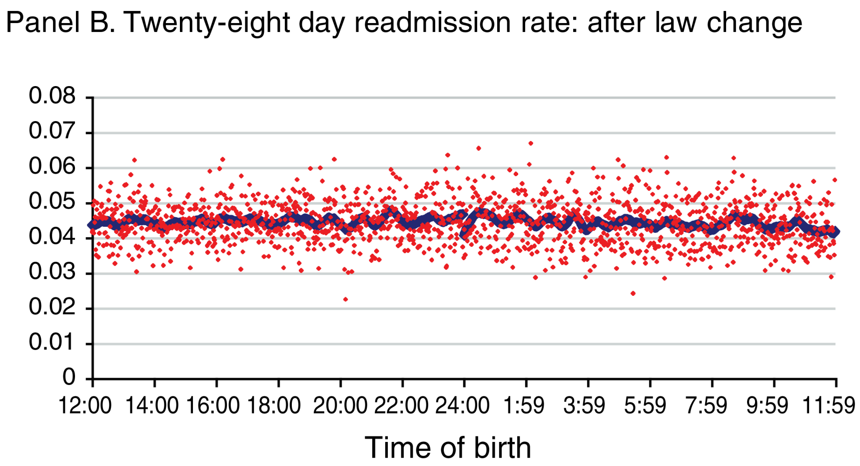
California requires that insurance cover two days of post-partum hospitalization

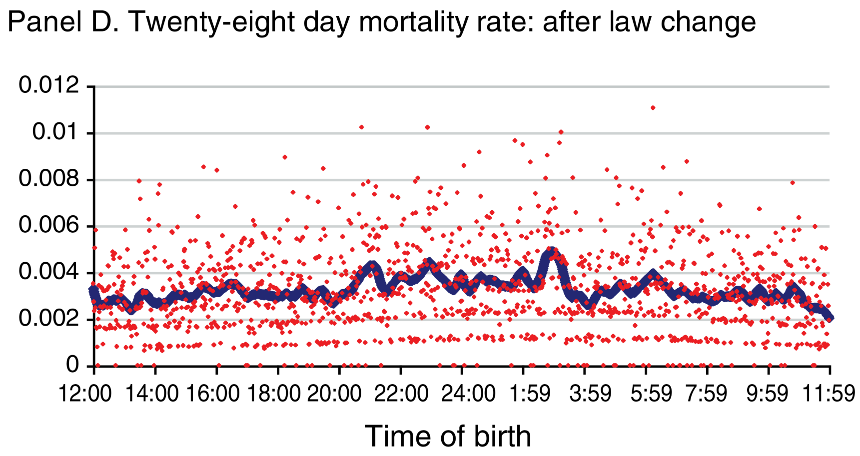
Does extra time in the hospital improve health outcomes?

Time discontinuities



Time discontinuities





...but delivering at 12:01 AM has no effect on readmission rates or mortality rates

Test score discontinuities

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.

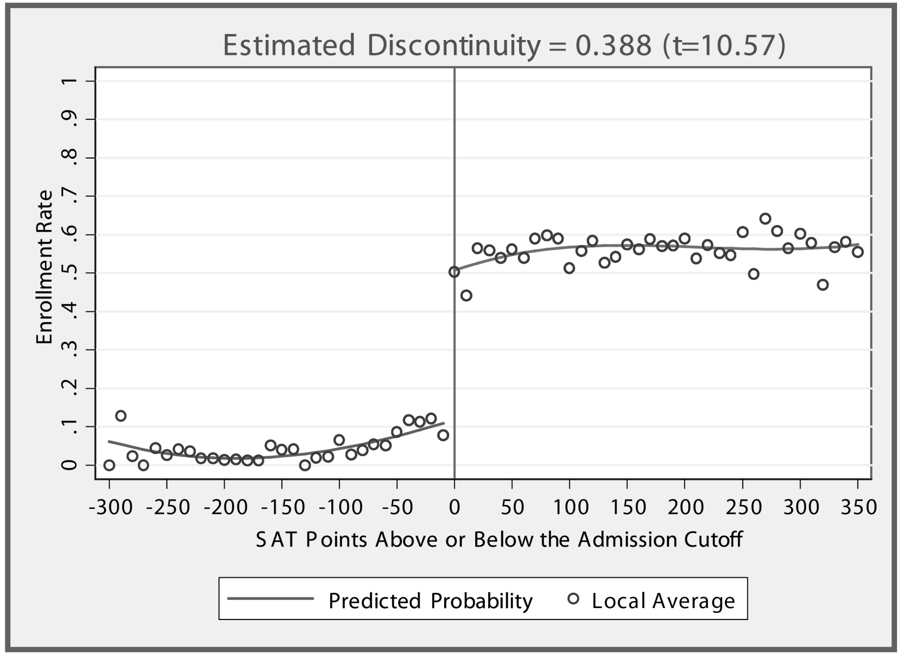
Convincingly estimating the economic returns to college quality requires overcoming the selection bias arising from the fact that attendance at more selective universities is likely correlated with unobserved characteristics that them-leges but chose to attend less selective institutions. They find that 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.

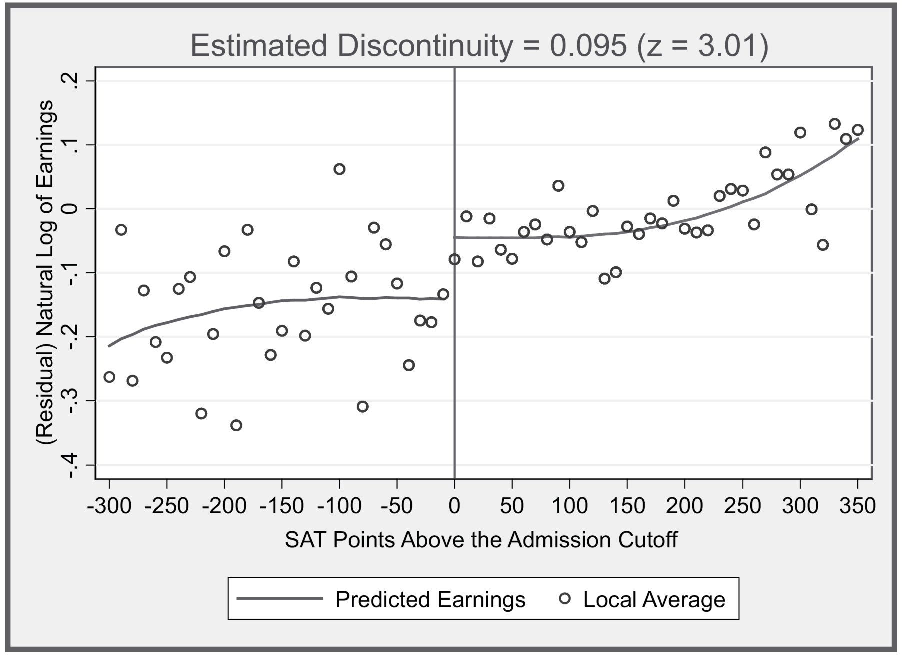
This paper uses a different strategy in that it identifies the effect of school selectivity on earnings by comparing the earnings of those just below the cutoff for admission to the flagship state university to those of applicants who were barely above the cutoff for admission. To do so, I combined confidential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

Does going to the main state university (e.g. UGA) make you earn more money?

SAT scores are an arbitrary cutoff for accessing the university

Test score discontinuities





Cutoff seems rule-based

Earnings are slightly higher

RDDs are all the rage

People love these things!

They're intuitive, compelling, and highly graphical

ABSTRACT

Methods Matter: P-Hacking and Causal Inference in Economics*

The economics 'credibility revolution' has promoted the identification of causal relationships using difference-in-differences (DID), instrumental variables (IV), randomized control trials (RCT) and regression discontinuity design (RDD) methods. The extent to which a reader should trust claims about the statistical significance of results proves very sensitive to method. Applying multiple methods to 13,440 hypothesis tests reported in 25 top economics journals in 2015, we show that selective publication and p-hacking is a substantial problem in research employing DID and (in particular) IV. RCT and RDD are much less problematic. Almost 25% of claims of marginally significant results in IV papers are misleading.

JEL Classification: A11, B41, C13, C44

Keywords: research methods, causal inference, p-curves, p-hacking, publication bias

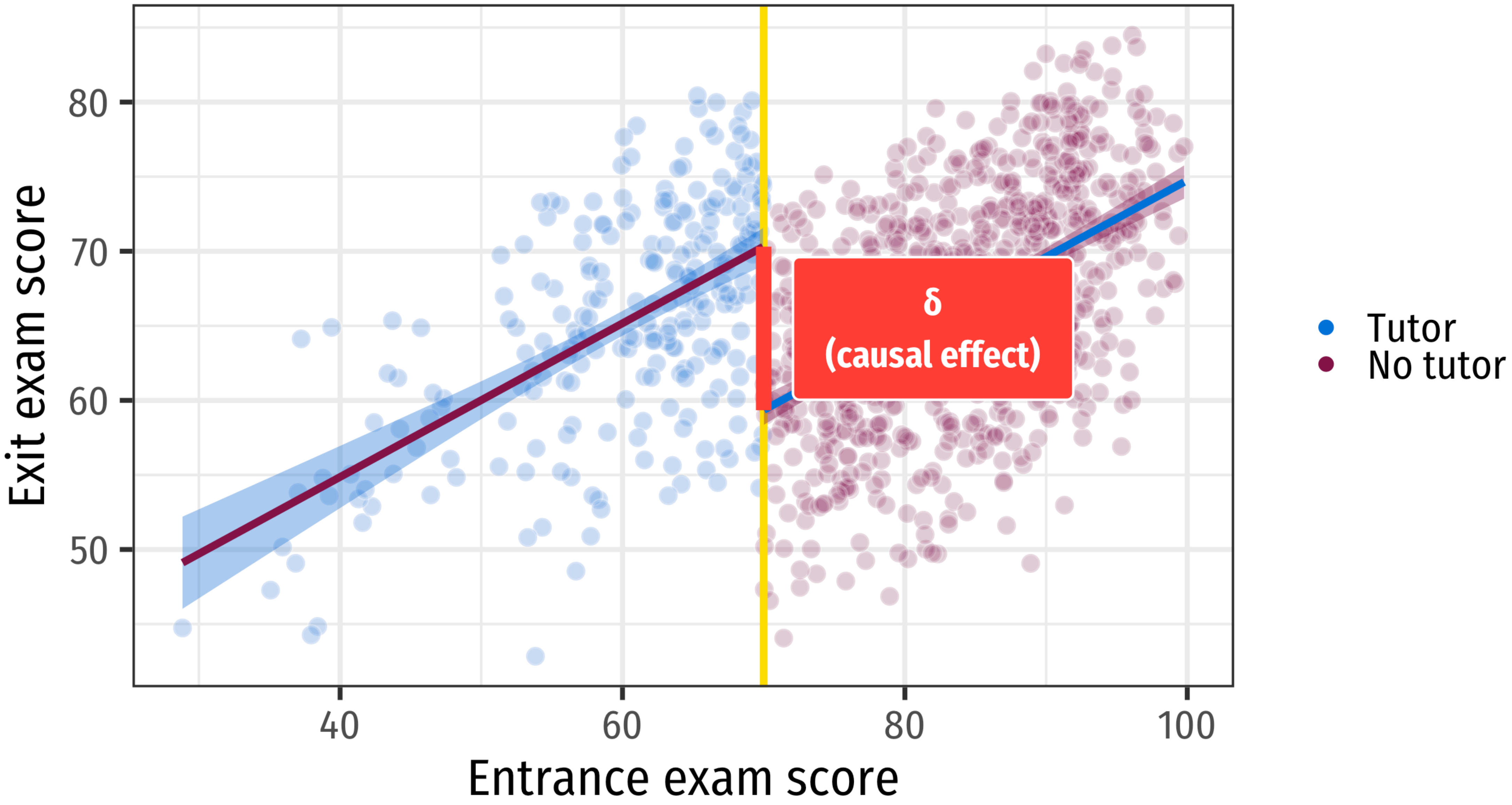
RDD less susceptible to p-hacking and selective publication than DID or IV

Drawing lines and measuring gaps

Main goal of RD

Measure the gap in outcome for people on both sides of the cutpoint

Gap = δ = local average treatment effect (LATE)


80
70
60
50
Exit exam score
δ
(causal effect)
40
60
80
100
Entrance exam score
Tutor
No tutor

Drawing lines

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

The type of lines you choose can change the estimate of δ—sometimes by a lot!

There's no one right way to draw lines!

Line-drawing considerations

Parametric vs. non-parametric lines

Measuring the gap

Bandwidths

Kernels

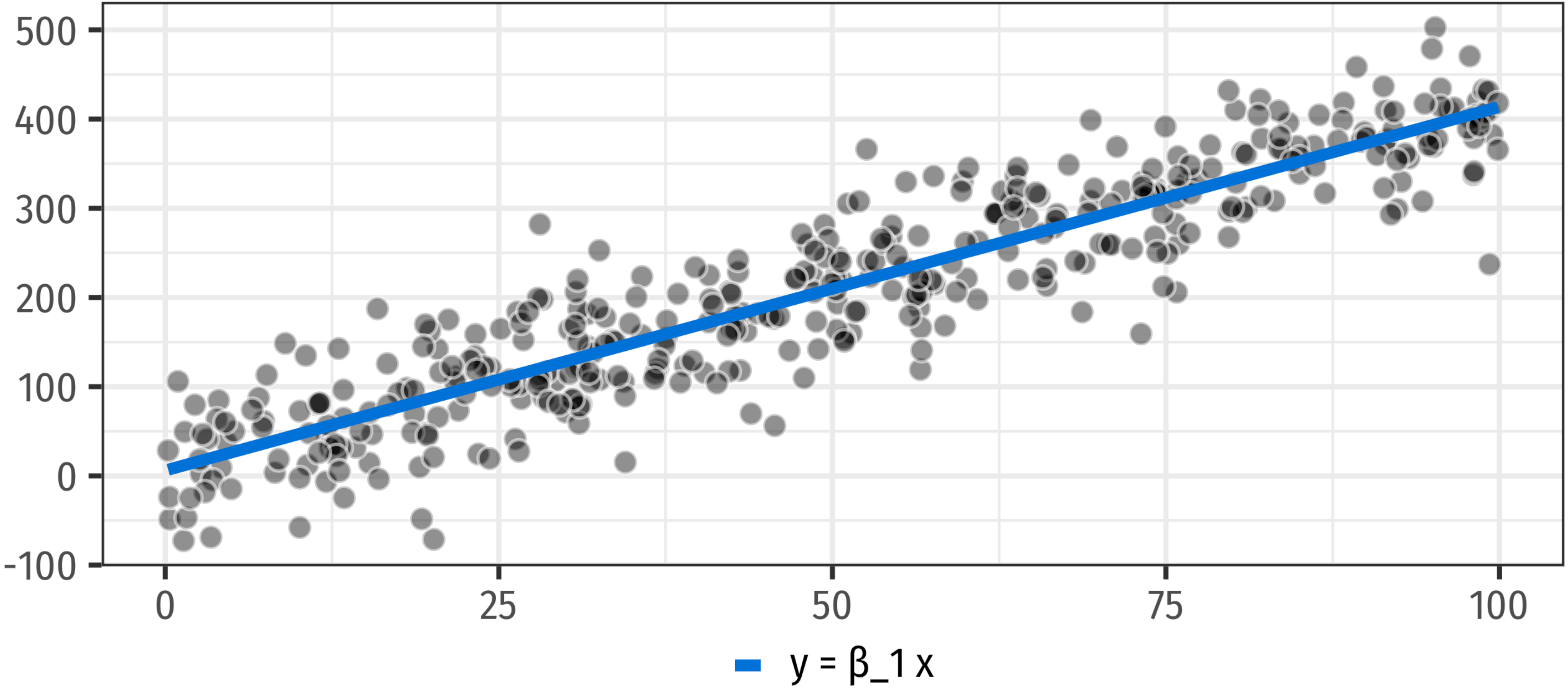
Parametric lines

Formulas with *parameters*

$$y = mx + b$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$y = 10 + 4x$$



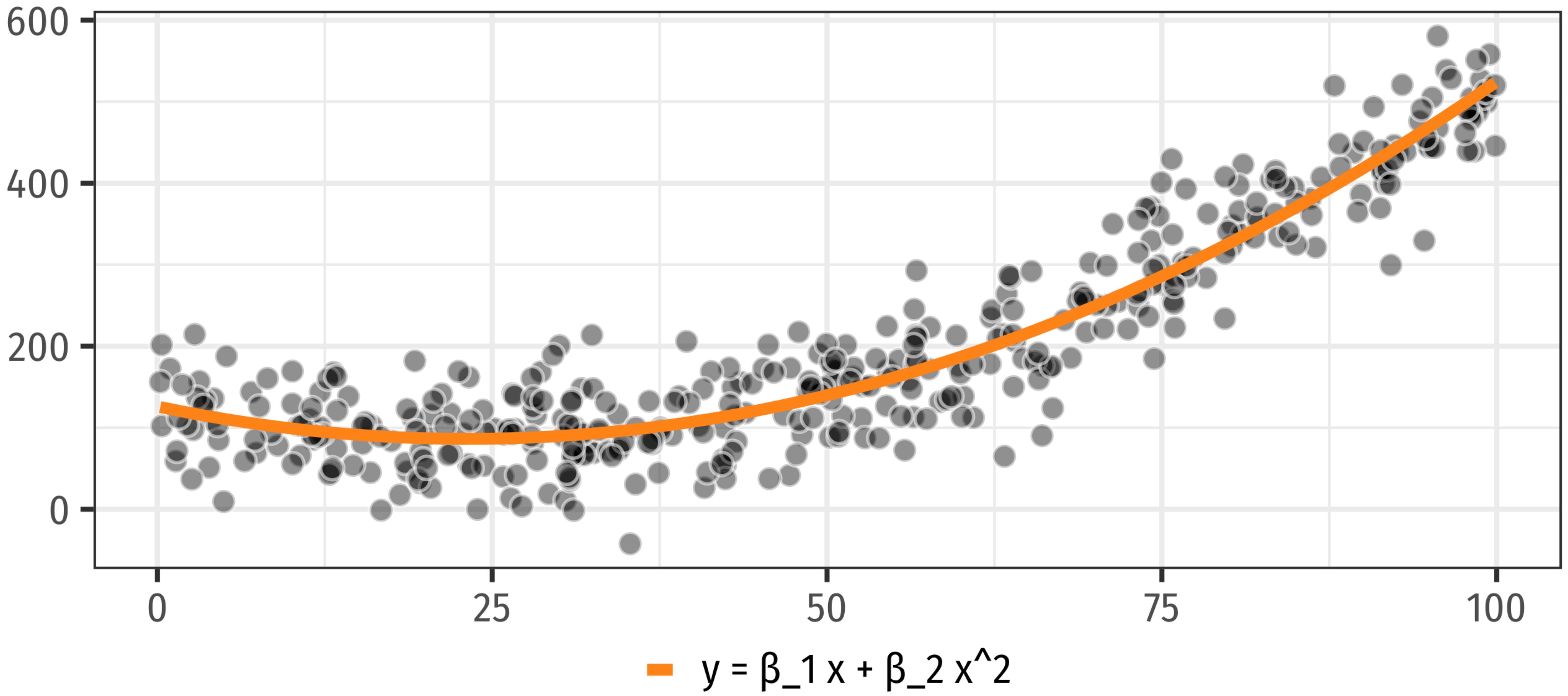
Parametric lines

Not just for straight lines!
Make curvy with exponents or trigonometry

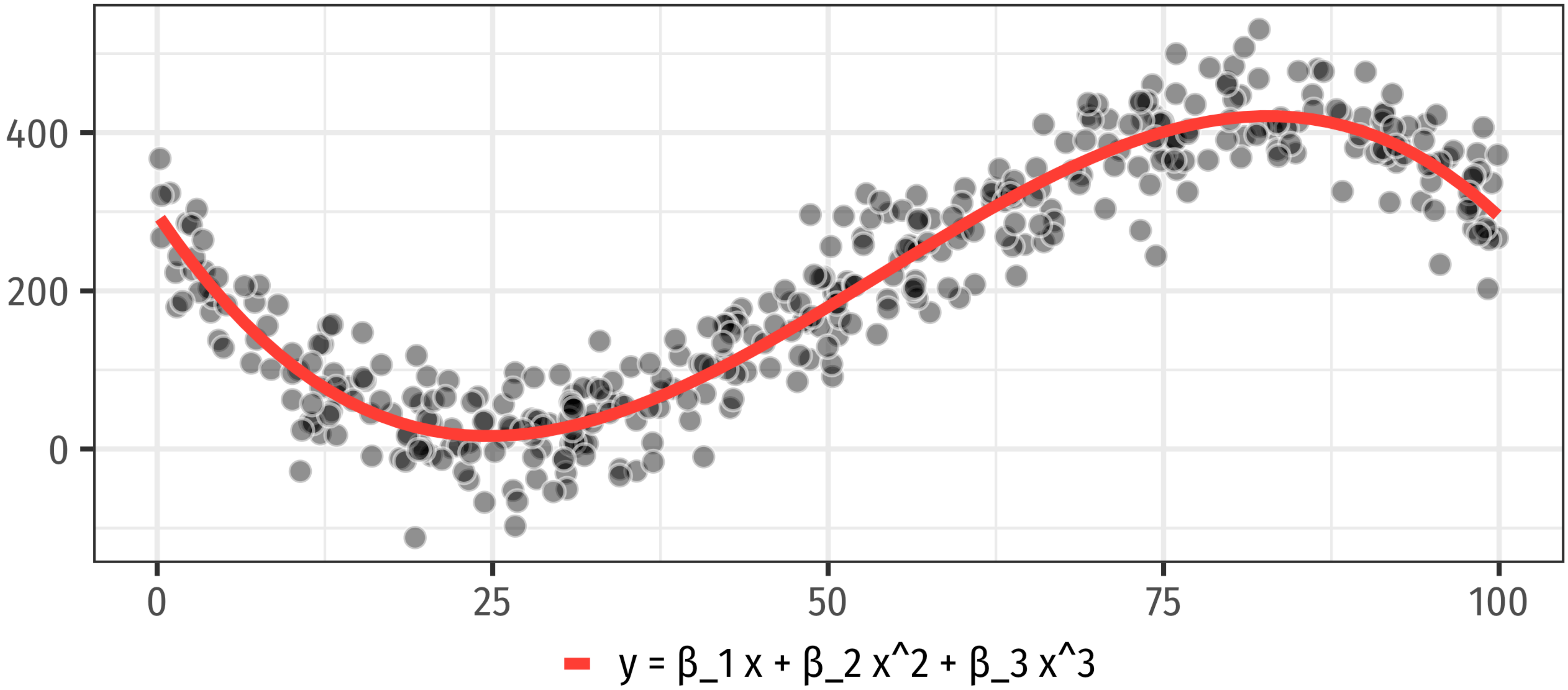
$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^7$$

$$y = \beta_0 + \beta_1 x + \beta_2 \sin(x)$$

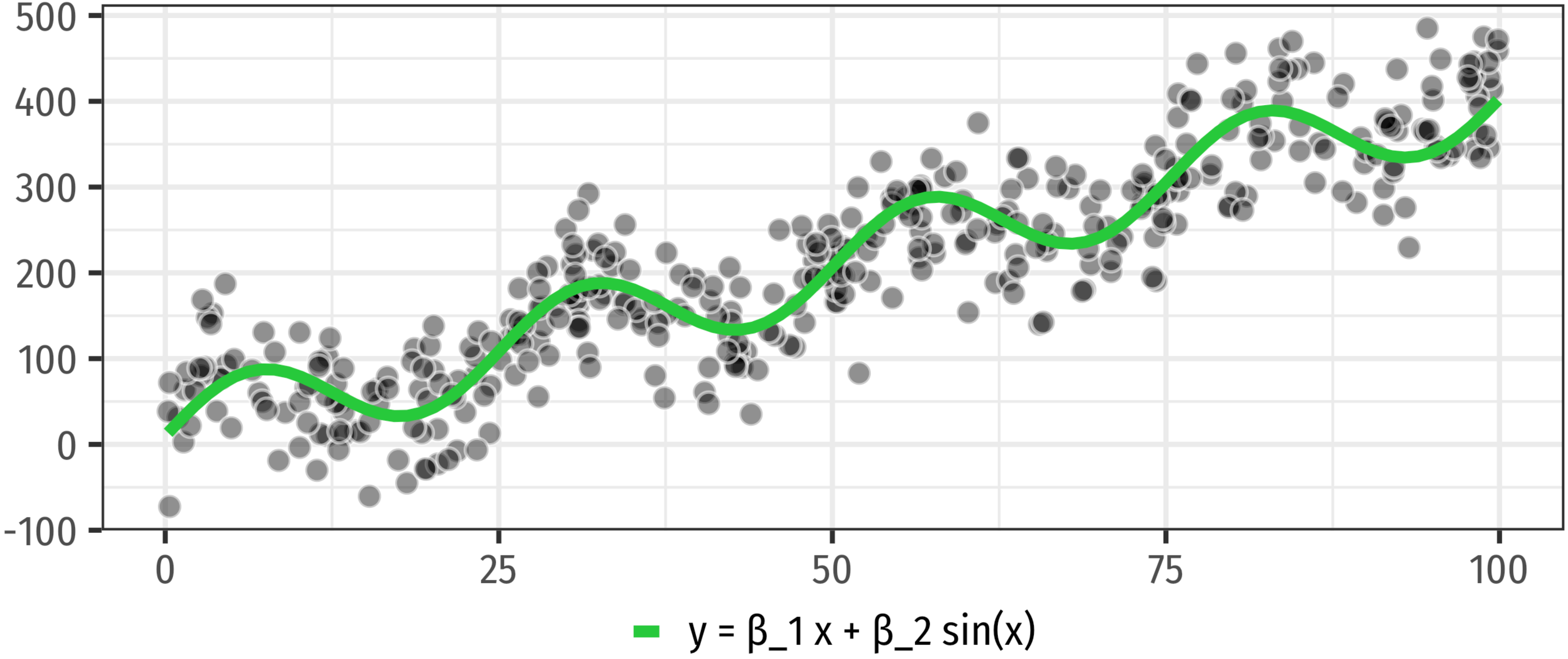
$$y = 120 - 3x + 0.07x^2$$



$$y = 300 - 25x + 0.65x^2 - 0.004x^3$$



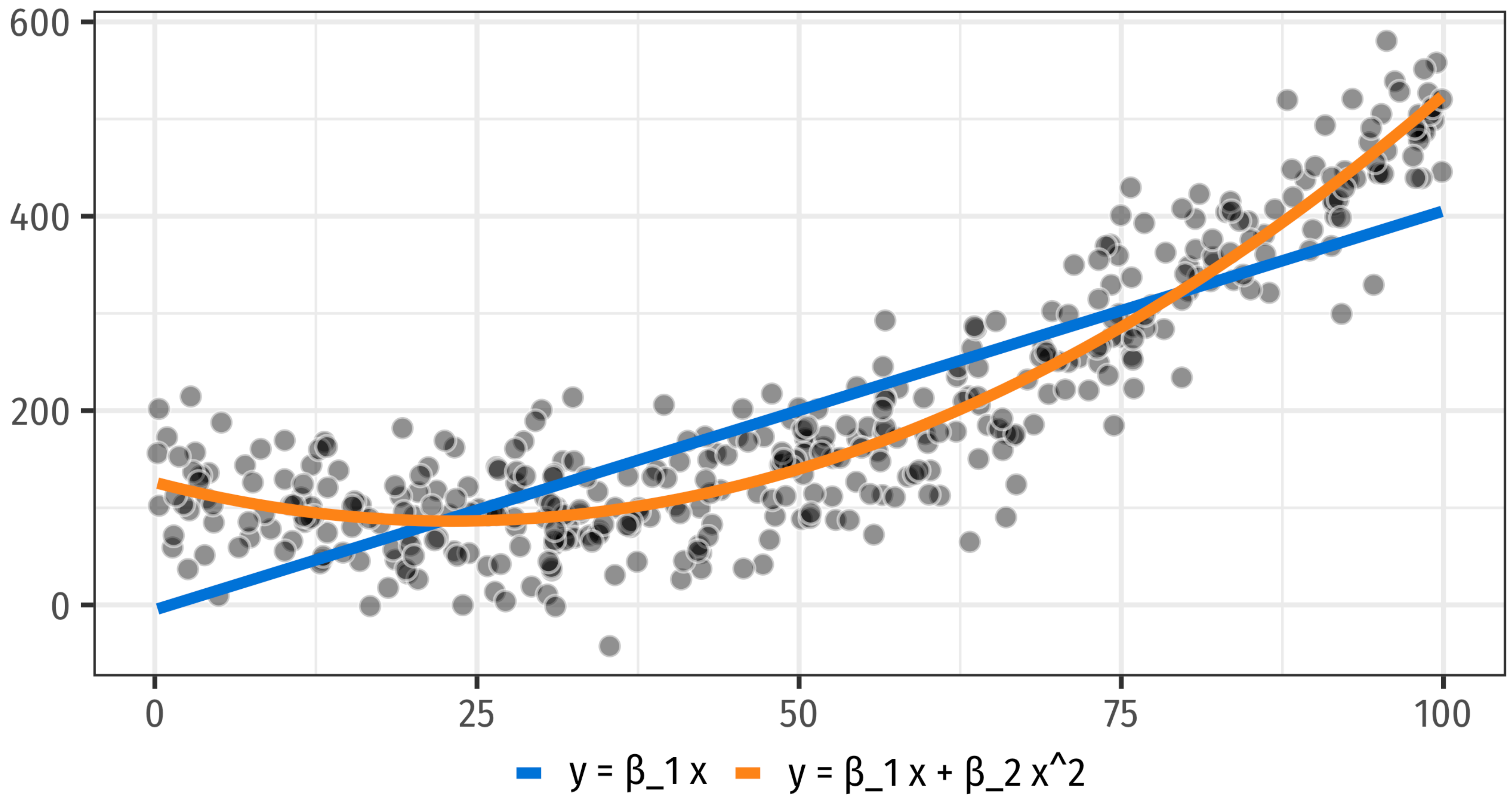
$$y = 10 + 4x + 50 \times \sin(\frac{x}{4})$$

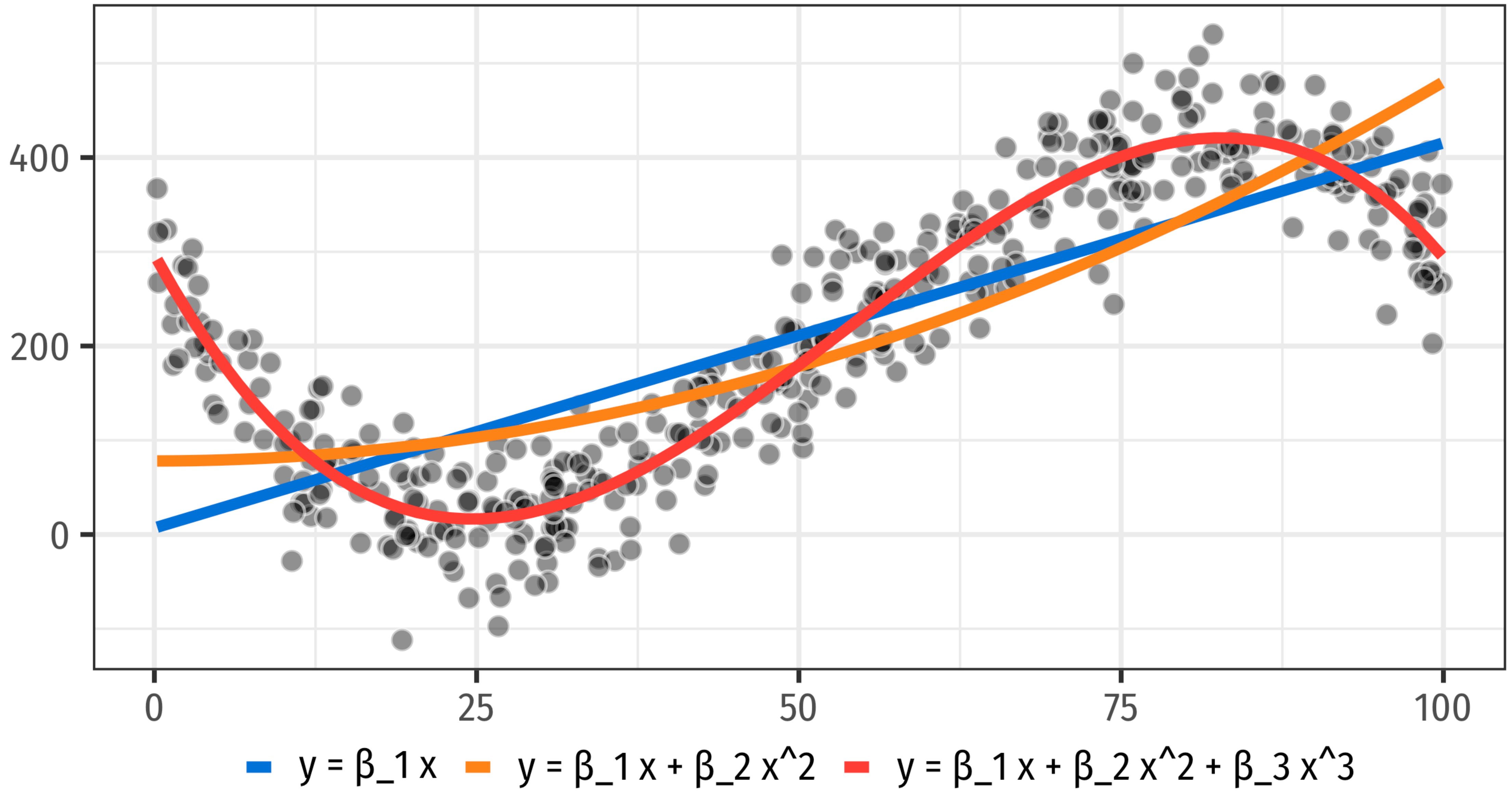


Parametric lines

It's important to get the parameters right!

Line should fit the data pretty well


600
400
200
0
0
25
50
75
100
y = β_1 x
y = β_1 x + β_2 x^2


400
200
0
0
25
50
75
100
y = β_1 x
y = β_1 x + β_2 x^2
y = β_1 x + β_2 x^2 + β_3 x^3

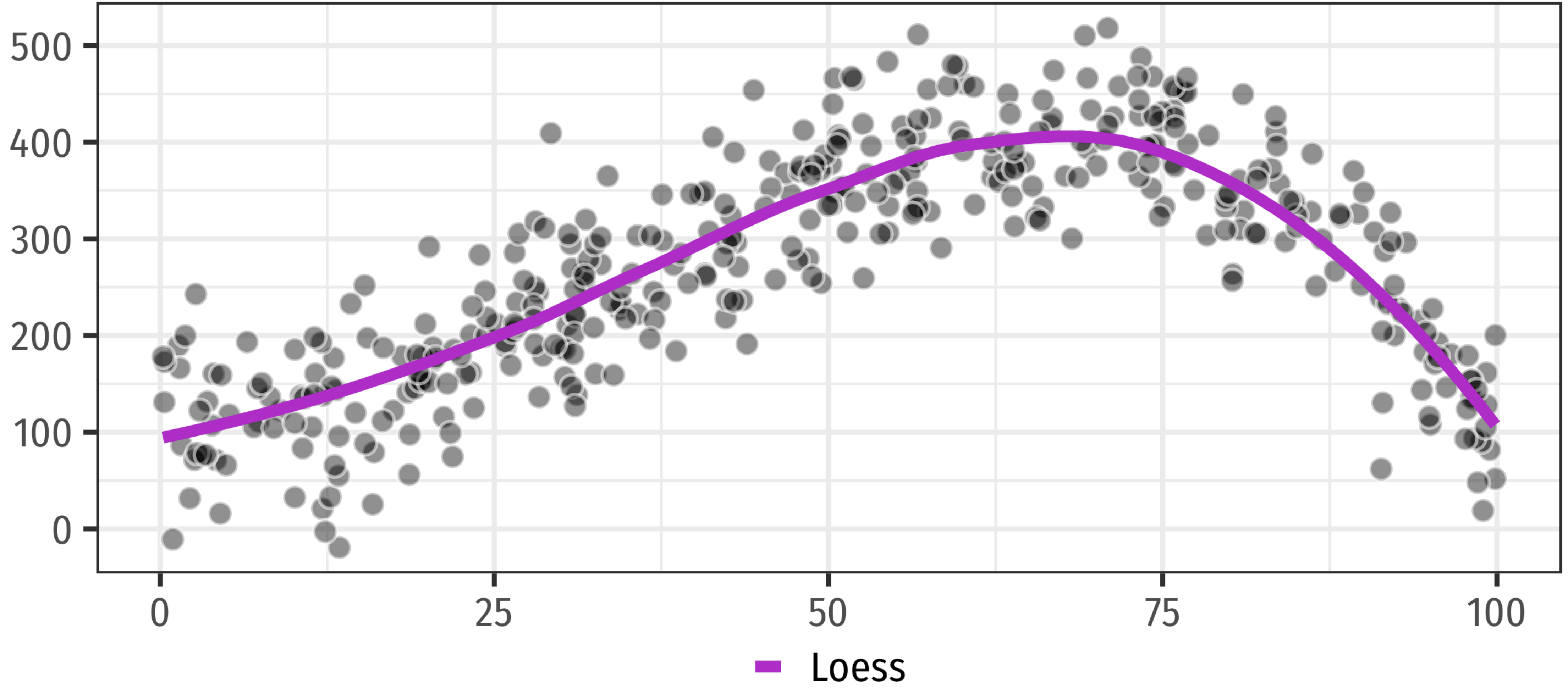
Nonparametric lines

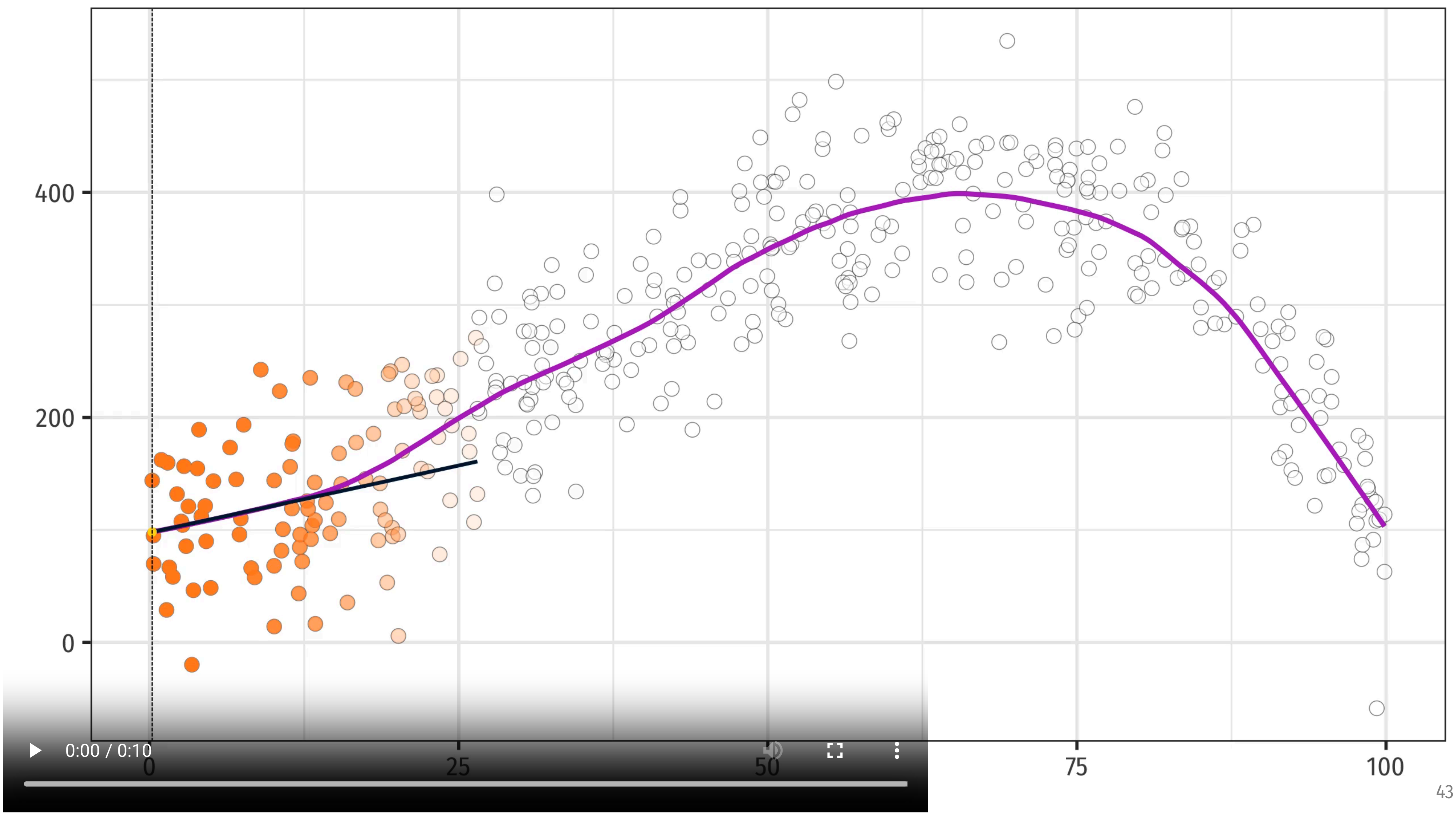
Lines without parameters

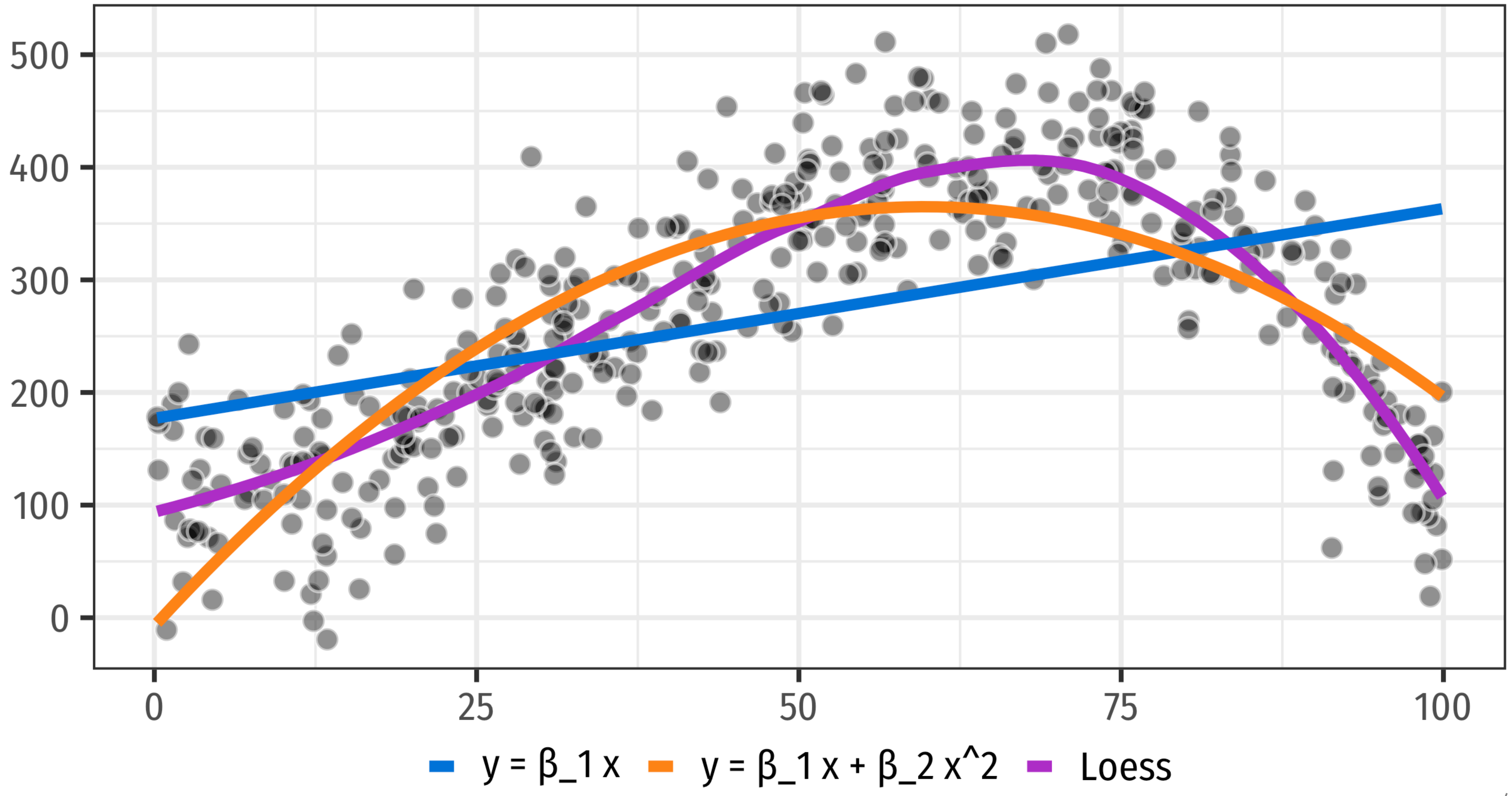
**Use the data to find the best line,
often with windows and moving averages**

**Locally estimated/weighted scatterplot smoothing
(LOESS/LOWESS)
is a common method (but not the only one!)**

$$y = \text{who knows?}$$

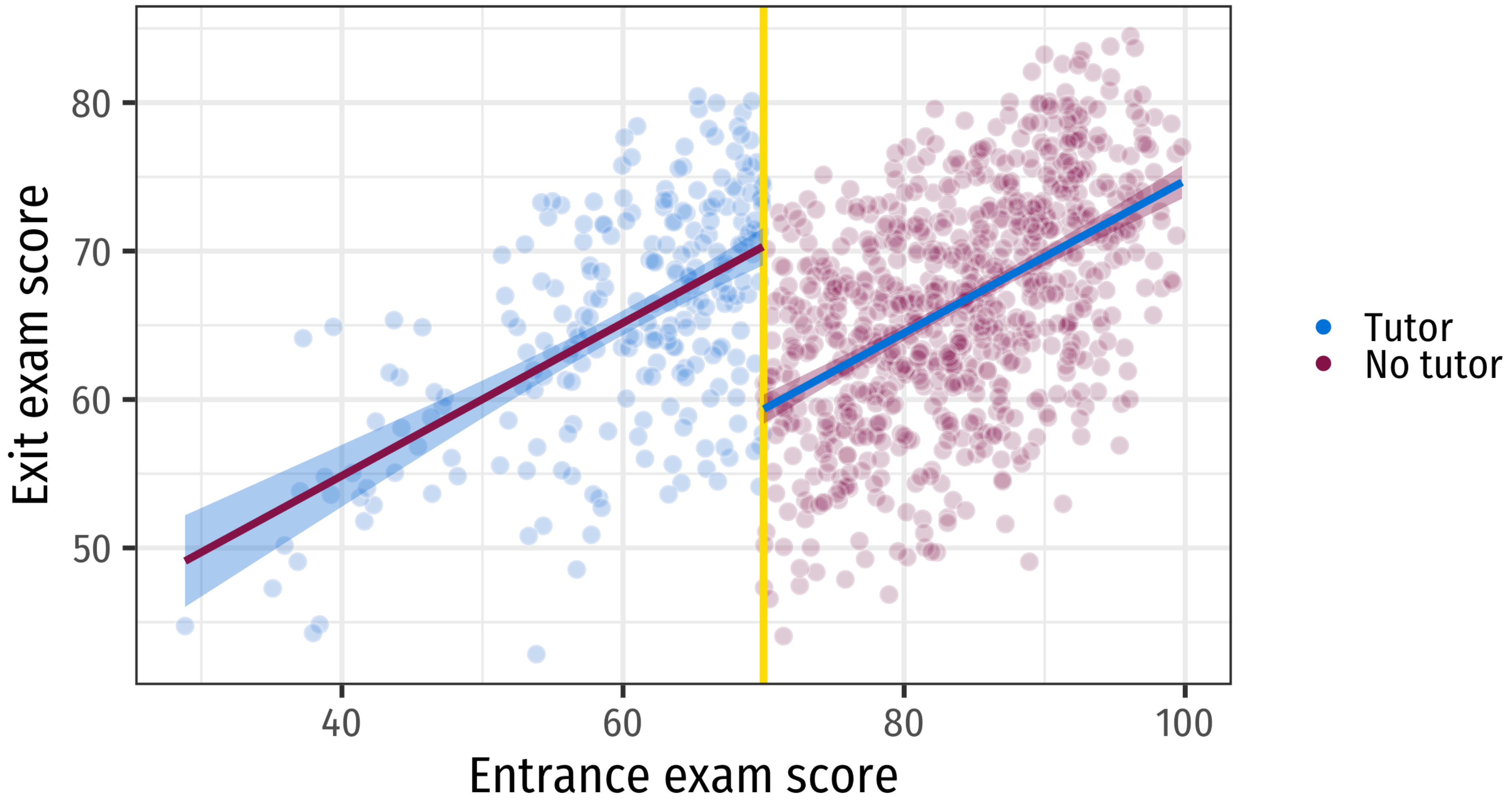



400
200
0
0
25
50
75
100
0:00 / 0:10



500
400
300
200
100
0
0
25
50
75
100
y = β_1 x
y = β_1 x + β_2 x^2
Loess

Measuring gap with parametric lines



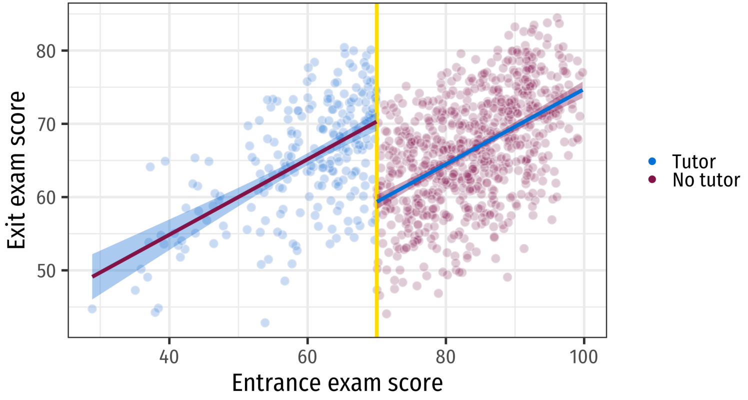
Measuring gap with parametric lines

Easiest way: center the running variable around the threshold

id	exit_exam	entrance_exam	entrance_centered	tutoring
1	78	92	22	FALSE
2	58	73	3	FALSE
3	62	54	-16	TRUE
4	67	98	28	FALSE
5	54	70	0	TRUE

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

Measuring gap with parametric lines



.right-code[

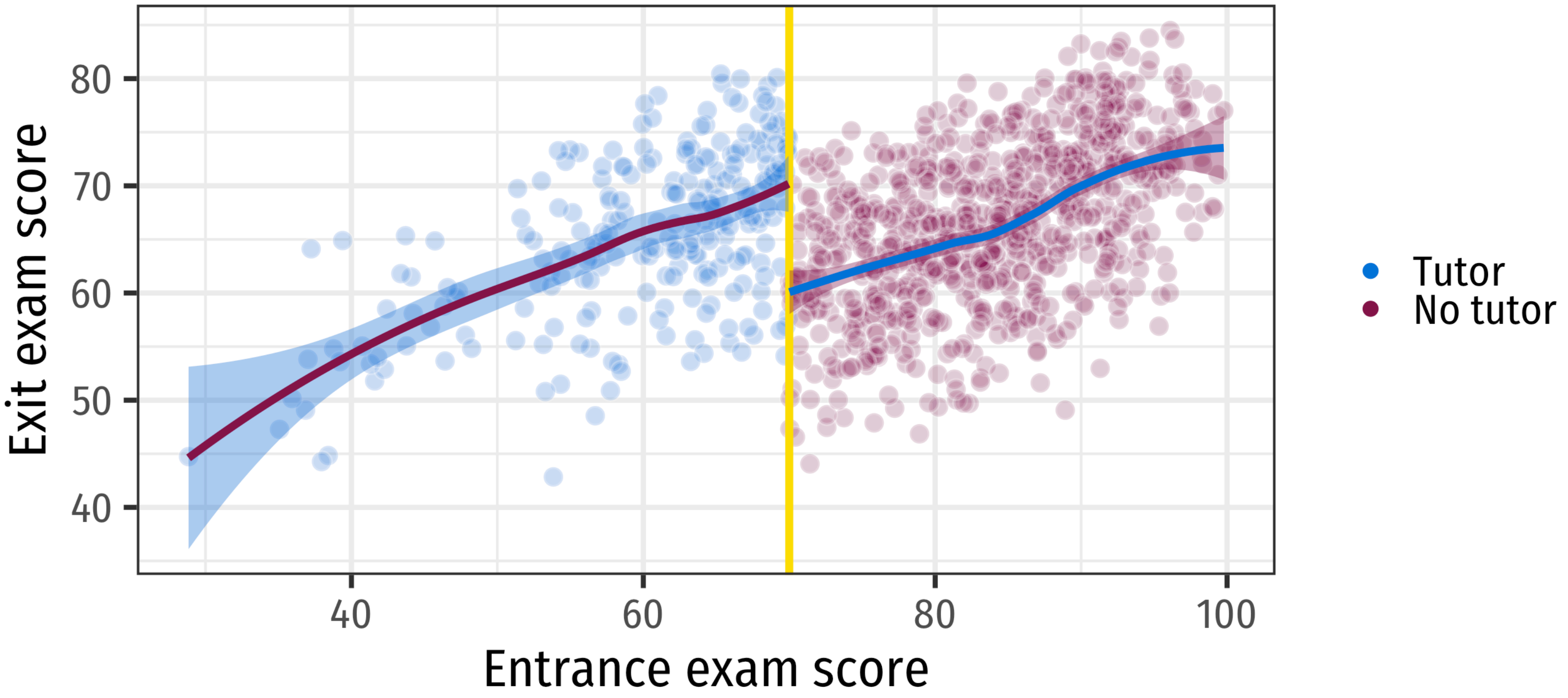
```
program_data <- tutoring |>
  mutate(entrance_centered =
           entrance_exam - 70)

model1 <- lm(exit_exam ~
               entrance_centered + tutoring,
             data = program_data)
```

```
tidy(model1)
```

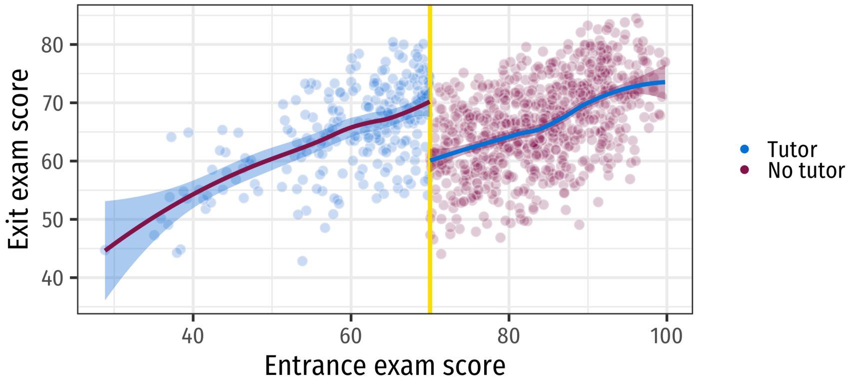
```
## [38;5;246m# A tibble: 3 × 3 [39m
##   term                  estimate std
```

Measuring gap with nonparametric lines



Can't use regression; use `rdrobust` R package

Measuring gap with nonparametric lines




```
rdrobust(y = tutoring$exit_exam, x = tutoring$entrance_exam, c = 70)
```

```
## =====================================================================
##                    Point    Robust Inference
##                 Estimate         z     P>|z|      [ 95% C.I. ]
## ---------------------------------------------------------------------
##      RD Effect    -9.992    -4.992     0.000   [-14.244 , -6.212]
## =====================================================================
```

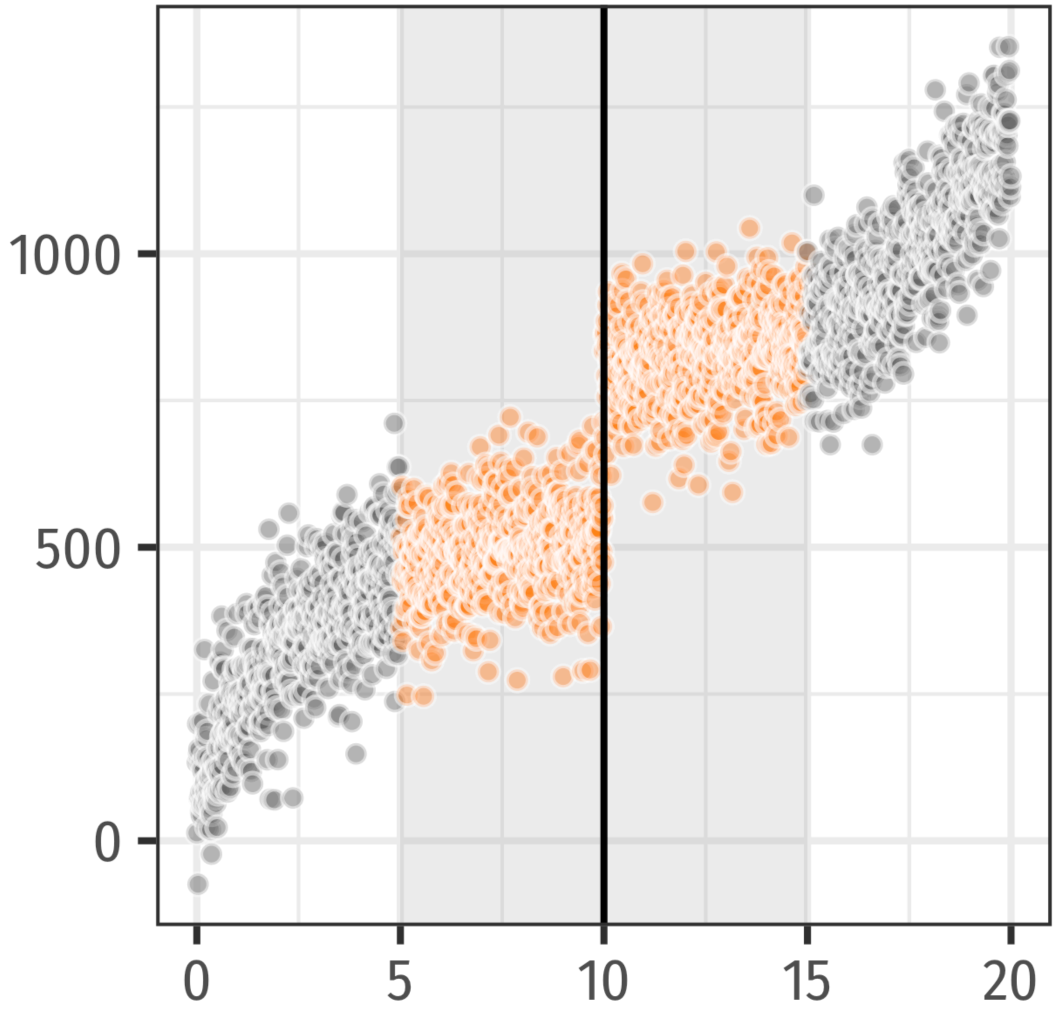
Bandwidths

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

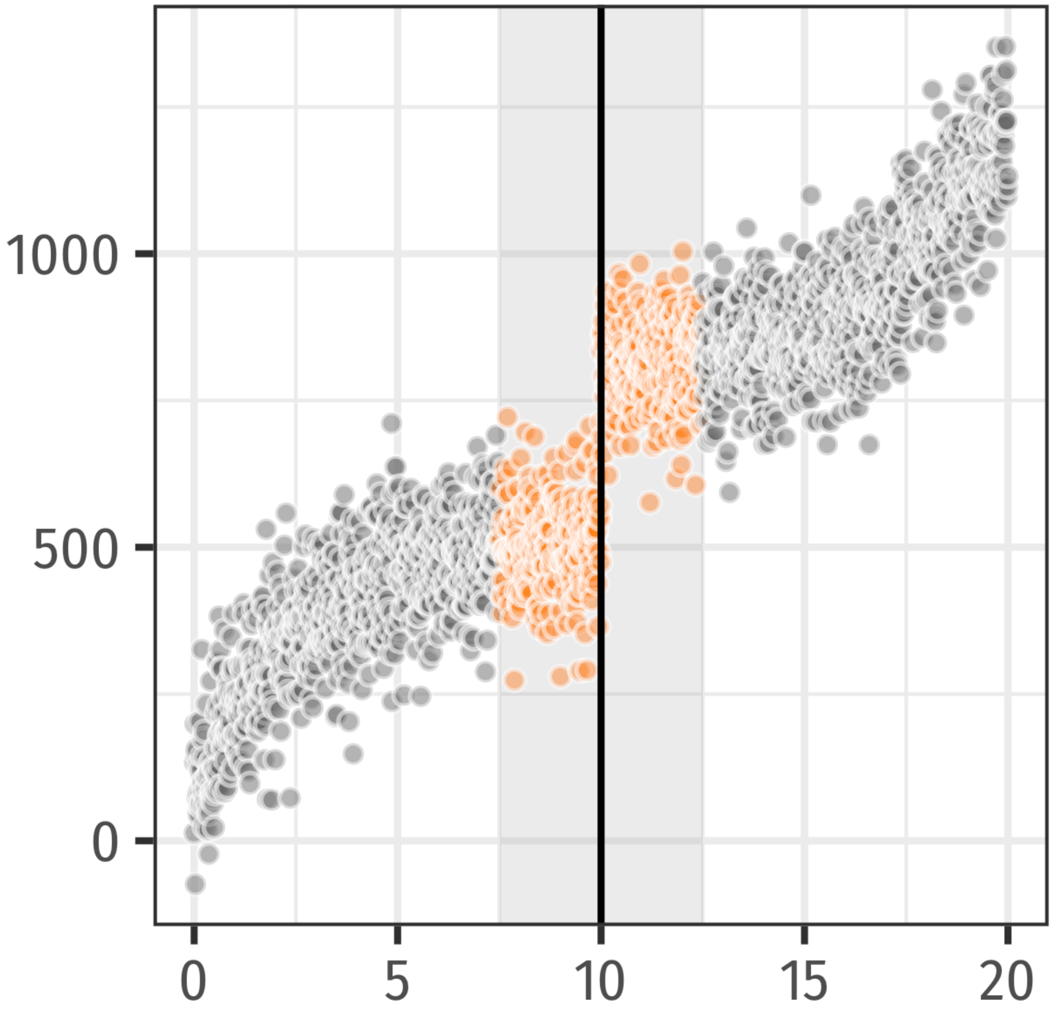
Observations far away don't matter because they're not comparable

Bandwidth = window around cutoff

Bandwidth = 5



Bandwidth = 2.5



Bandwidths

Algorithms exist to choose optimal width

Also use common sense

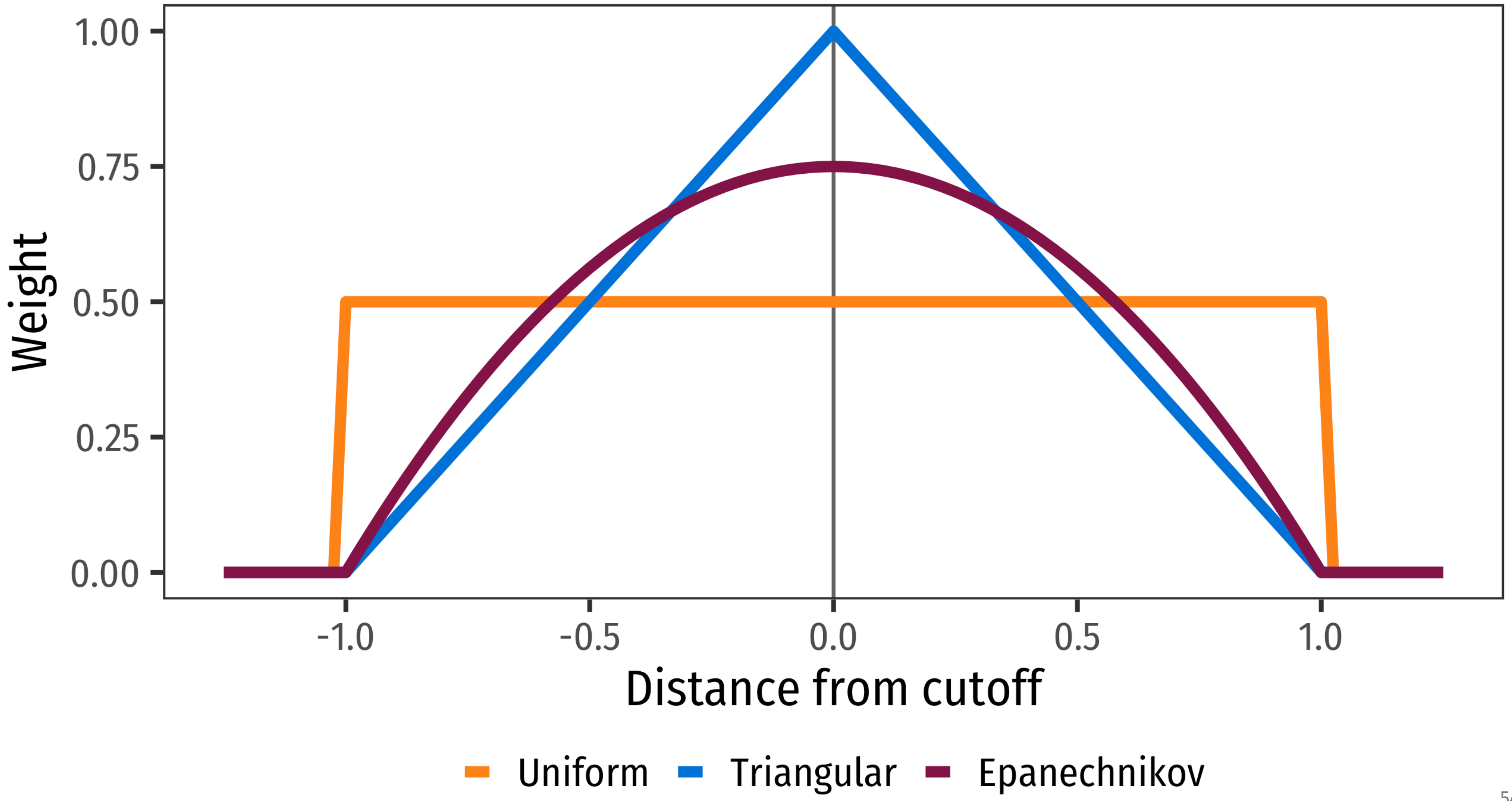
Maybe ±5 for the entrance exam?

For robustness, check what happens if you double and halve the bandwidth

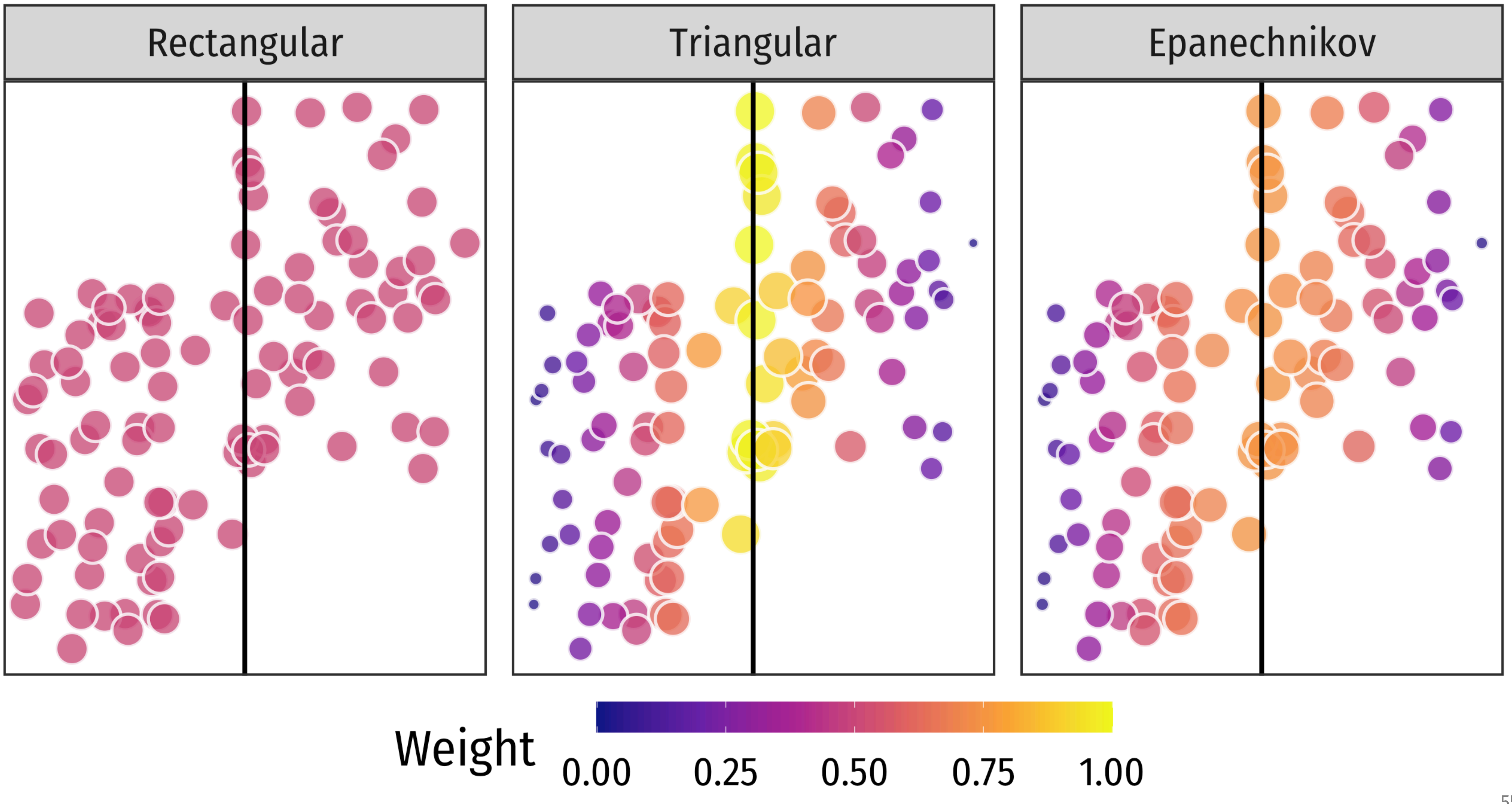
Kernels

Because we care the most about observations right by the cutoff, give more distant ones less weight

Kernel = method for assigning importance to observations based on distance to the cutoff



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


Rectangular
Triangular
Epanechnikov
Weight
0.00
0.25
0.50
0.75
1.00

Try everything!

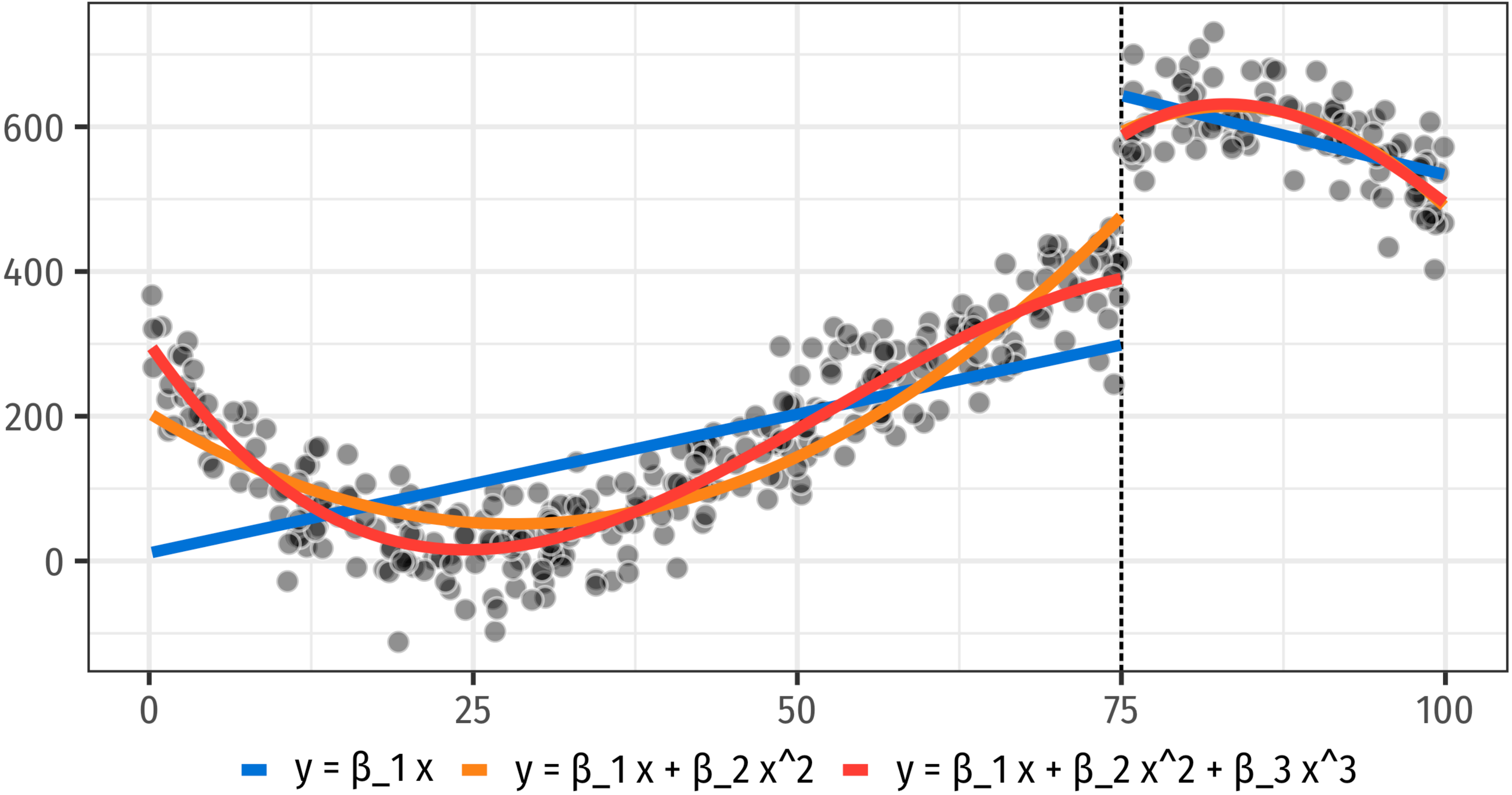
Your estimate of δ depends on all these:

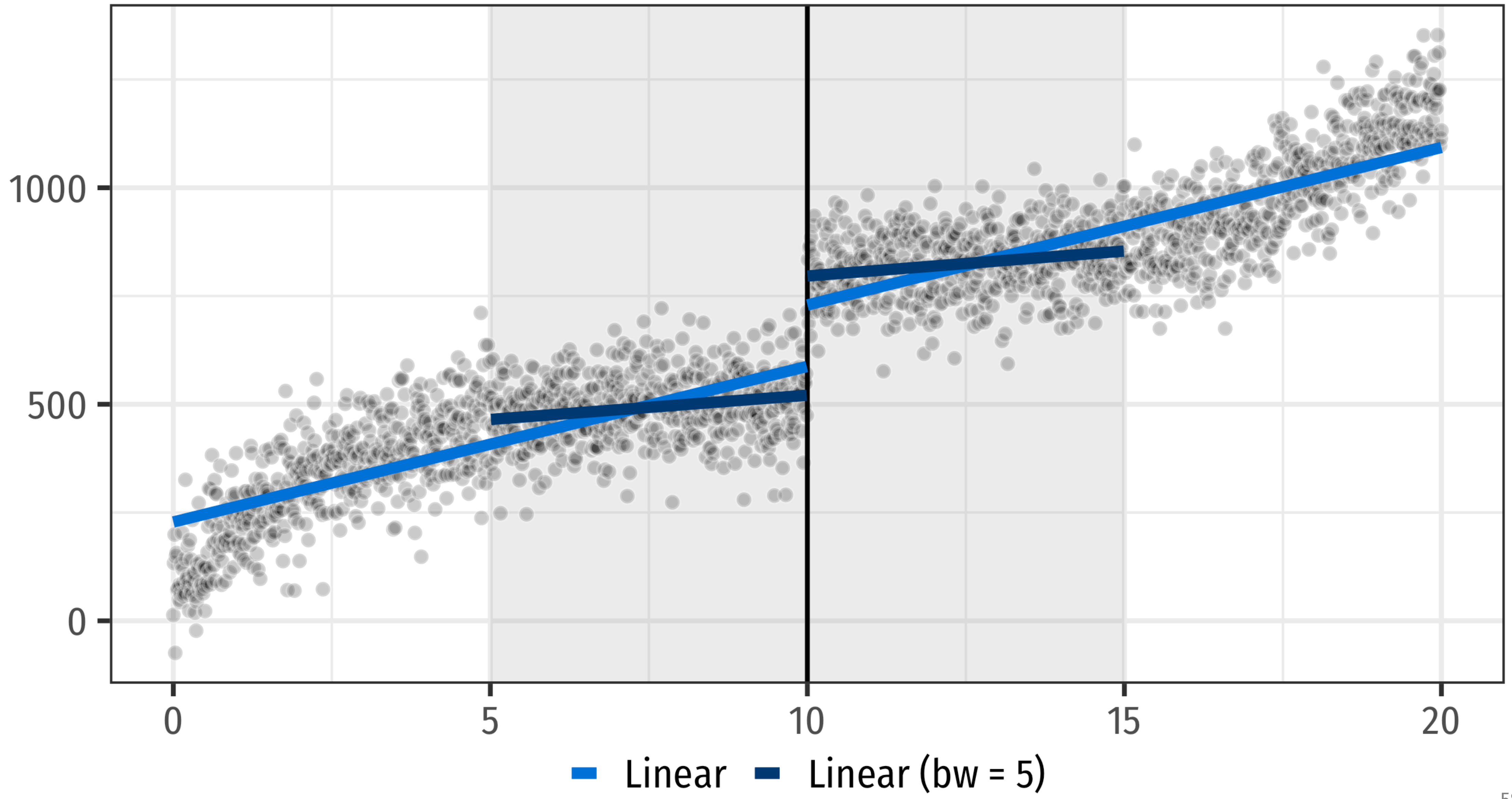
Line type (parametric vs. nonparametric)

Bandwidth (wide vs. narrow)

Kernel weighting

Try lots of different combinations!

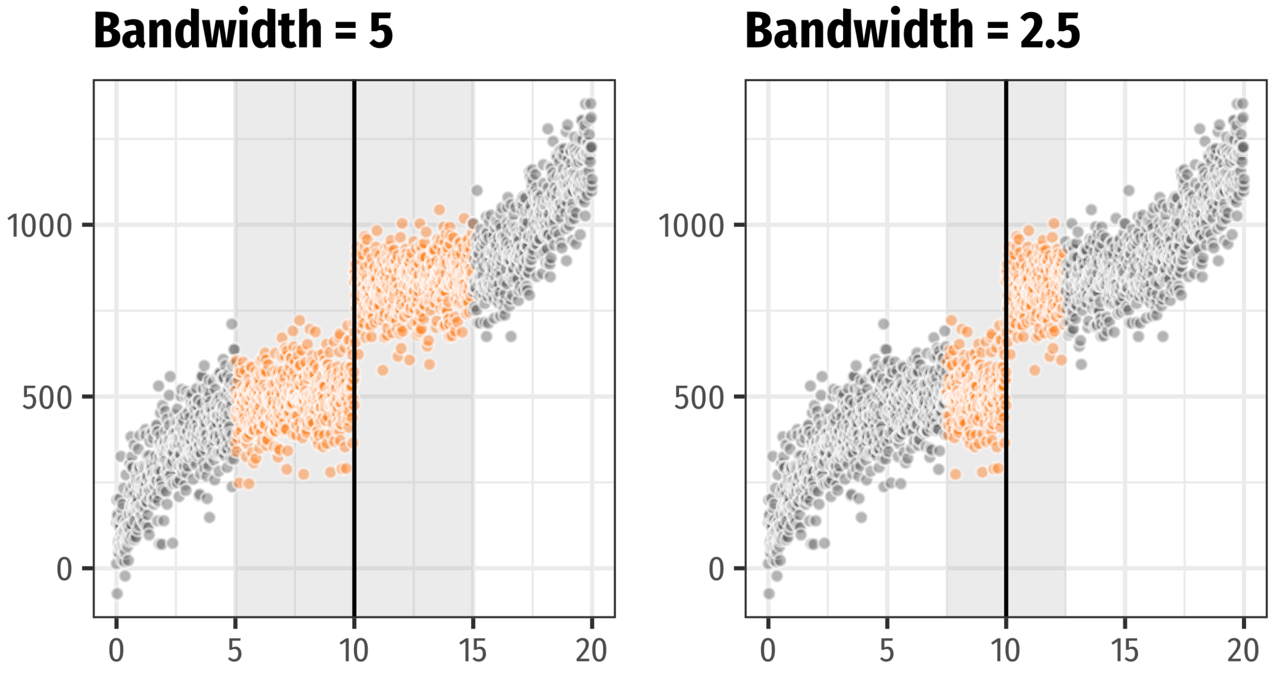

600
400
200
0
0
25
50
75
100
y = β_1 x
y = β_1 x + β_2 x^2
y = β_1 x + β_2 x^2 + β_3 x^3


1000
500
0
0
5
10
15
20
Linear
Linear (bw = 5)

Main RDD concerns

It's greedy!

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



It's limited in scope!

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

Local Average Treatment Effect (LATE)

It's limited in scope!

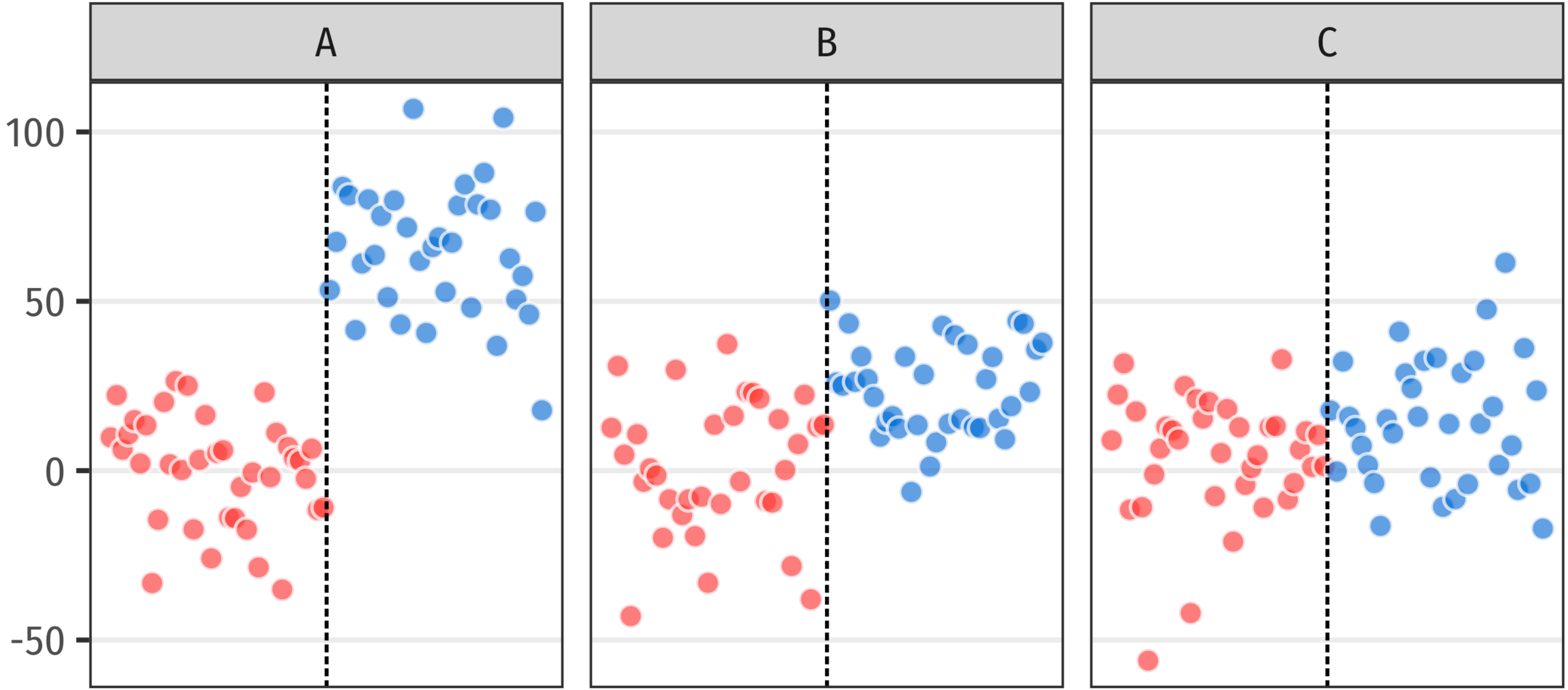
You can't make population-level claims with a LATE

(But can you really do that with RCTs or diff-in-diff?)

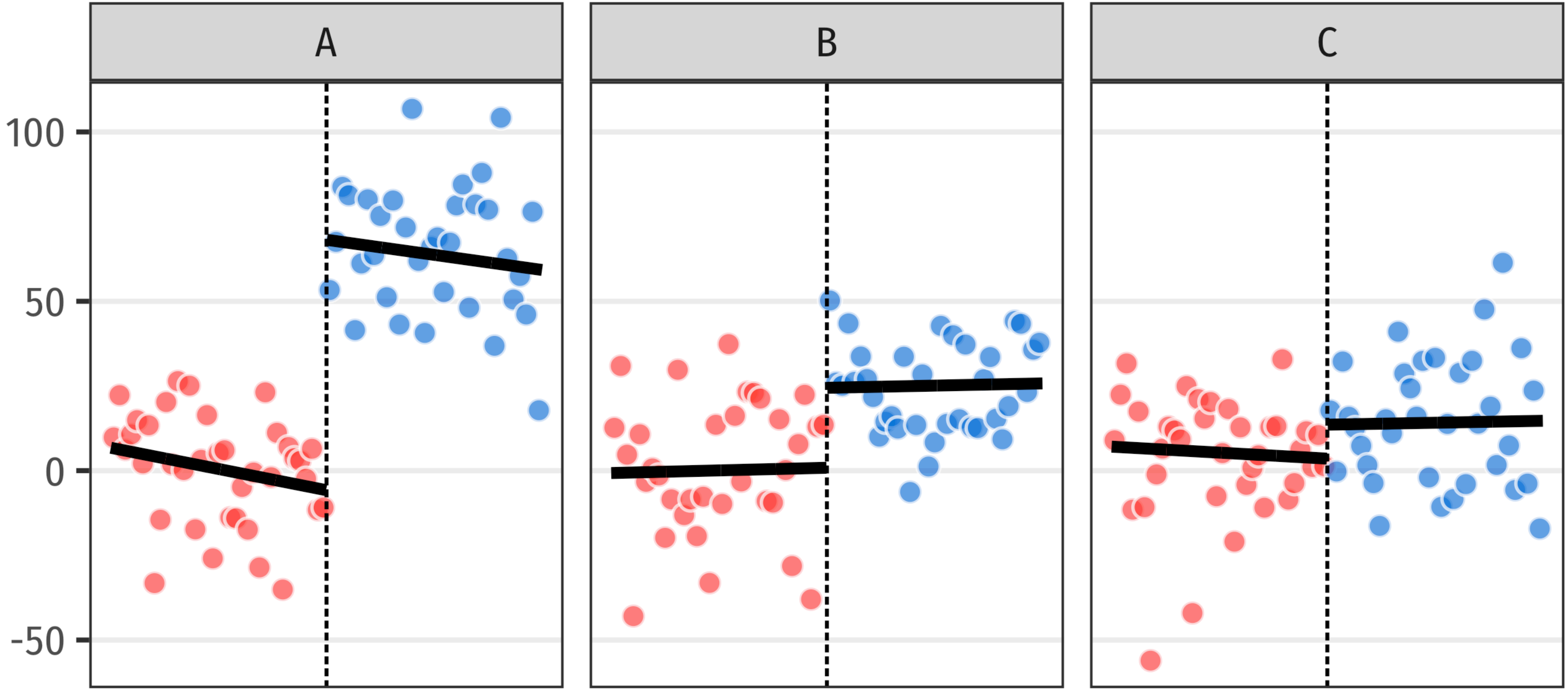
"The realistic conclusion to draw is that all quantitative empirical results that we encounter are 'local'"

Angrist and Pischke, *Mostly Harmless Econometrics*, pp. 23–24

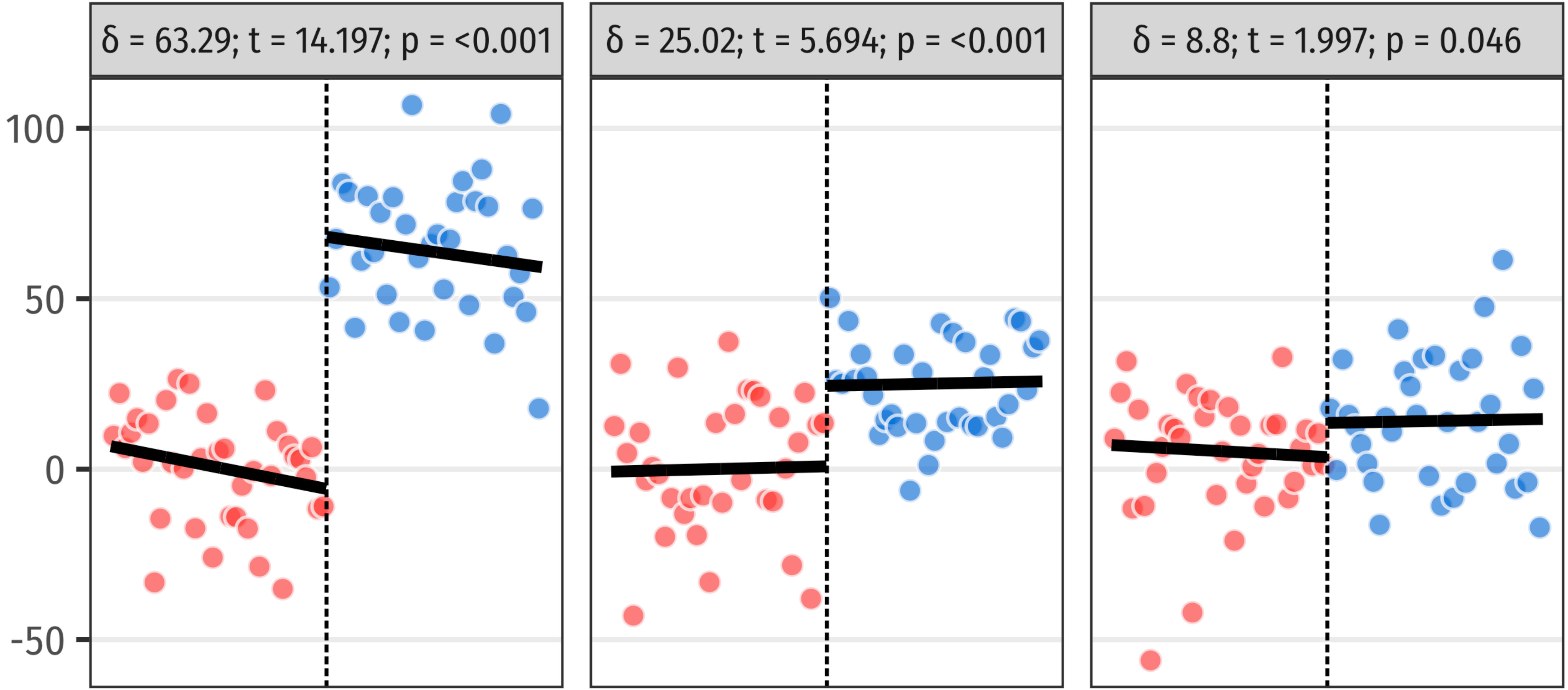
Graphics are neat!



Which gaps are significant?



All of them!

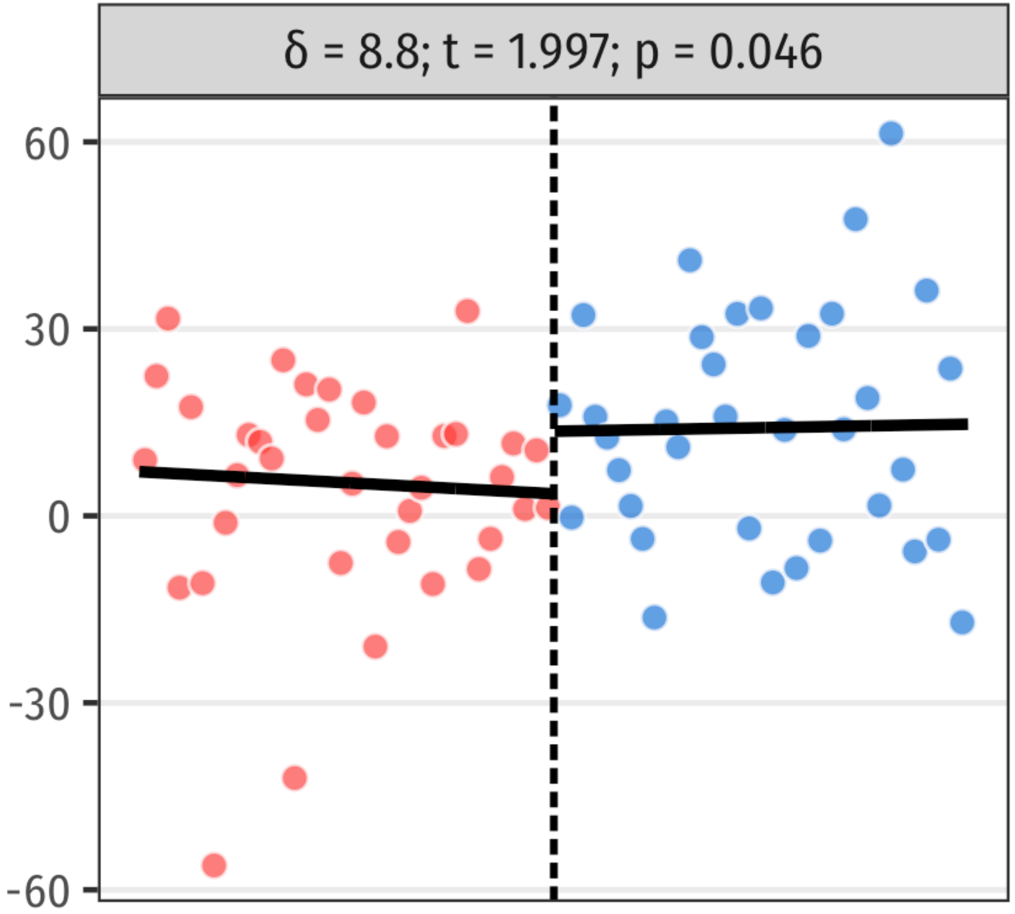


δ = 63.29; t = 14.197; p = <0.001
δ = 25.02; t = 5.694; p = <0.001
δ = 8.8; t = 1.997; p = 0.046
100
50
0
-50

Don't rely *only* on graphics

Super clear breaks are uncommon

Make graphs, but also find the actual δ value



Manipulation!

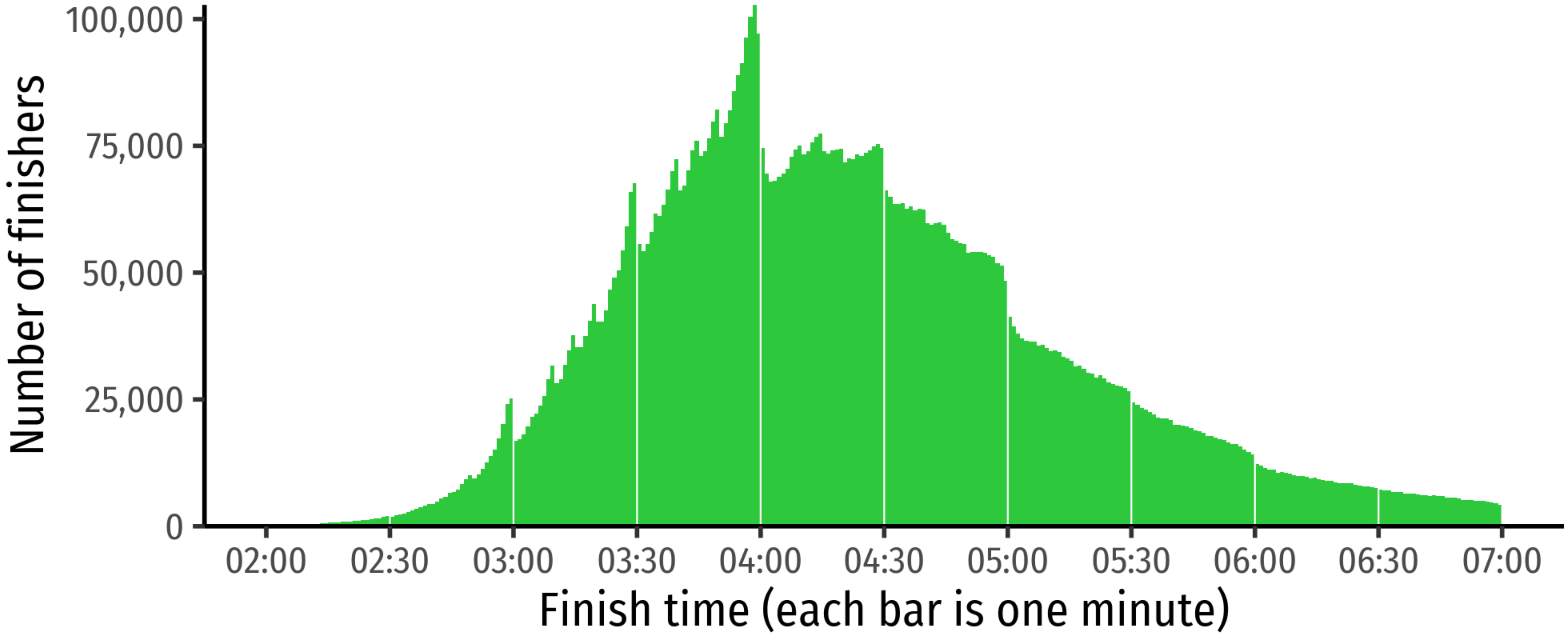
People might know about the cutoff and change their behavior

People might fudge numbers or work to cross the threshold to get in/out of program

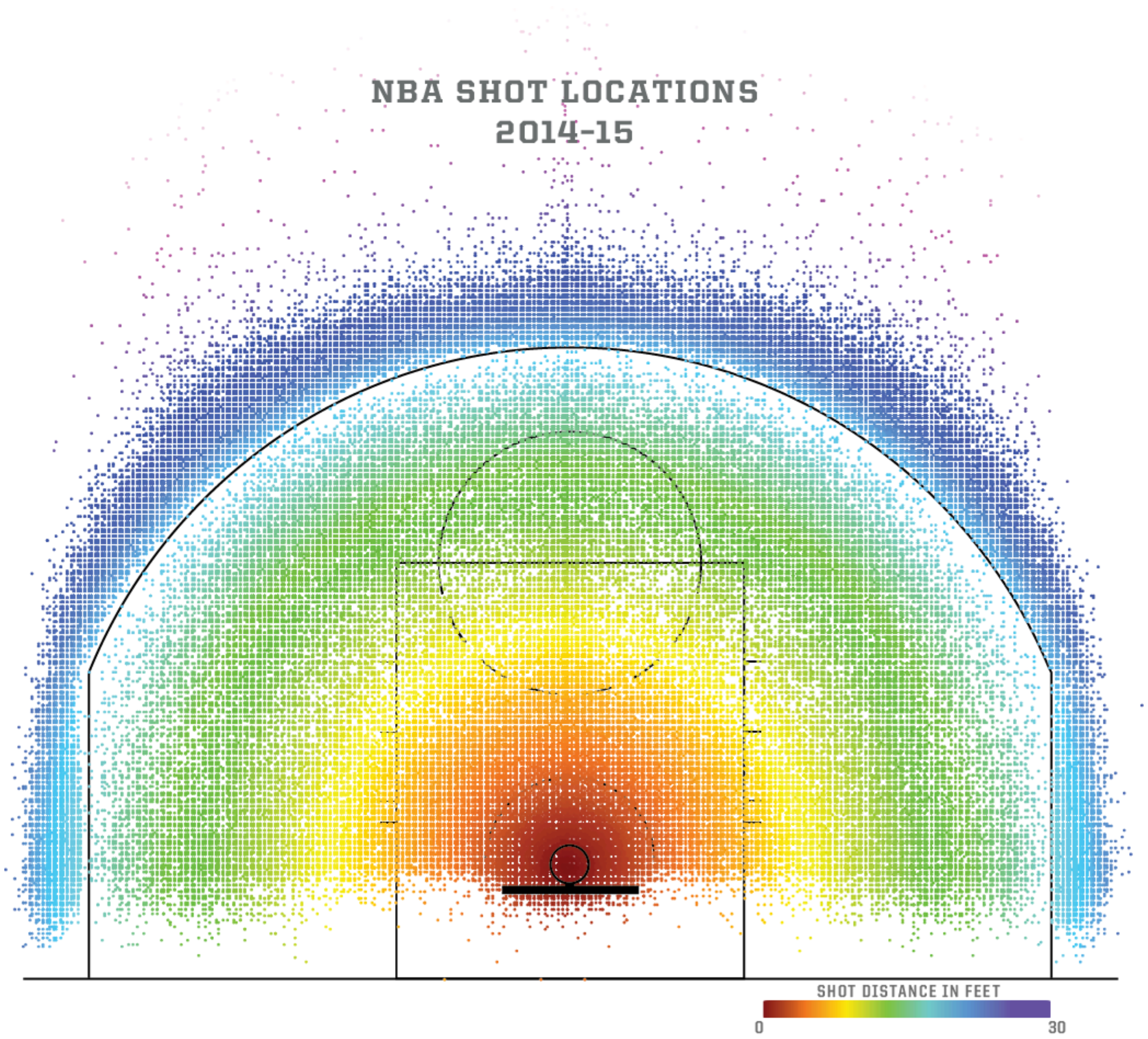
If so, those right next to the cutoff are no longer comparable treatment/control groups

Distribution of marathon finishing times

N = 9,589,053



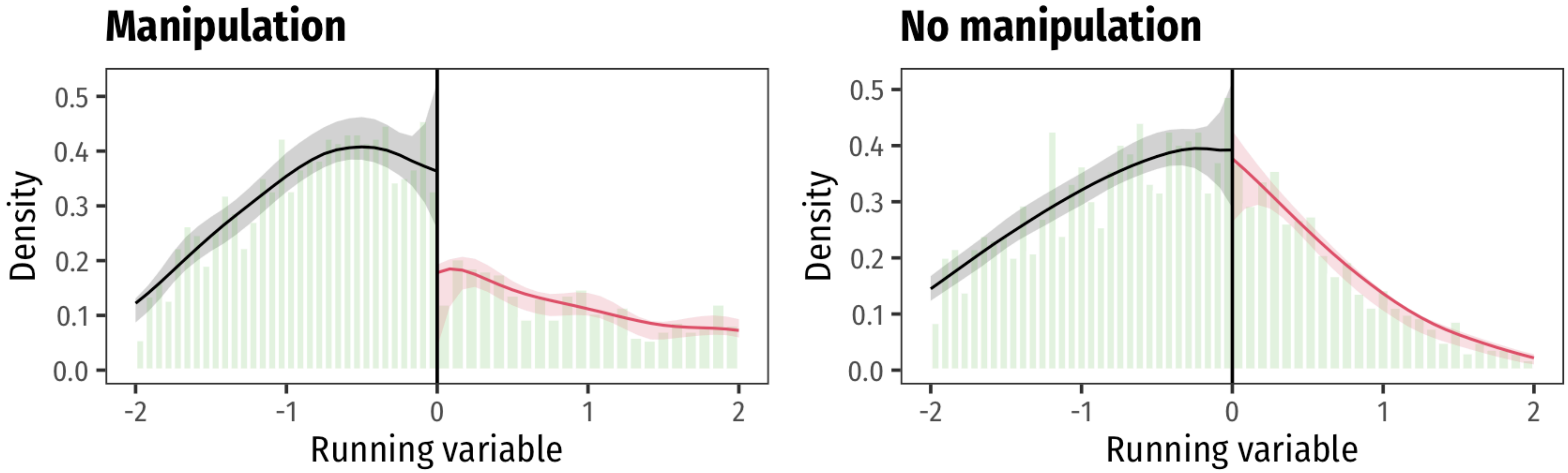
Eric J. Allen, Patricia M. Dechow, Devin G. Pope, George Wu (2017)
Reference-Dependent Preferences: Evidence from Marathon Runners.
Management Science 63(6):1657-1672. https://doi.org/10.1287/mnsc.2015.2417


NBA SHOT LOCATIONS
2014-15
SHOT DISTANCE IN FEET
0
30

Manipulation!

Check with a McCrary density test

`rddensity::rdplotdensity()` in R



Noncompliance!

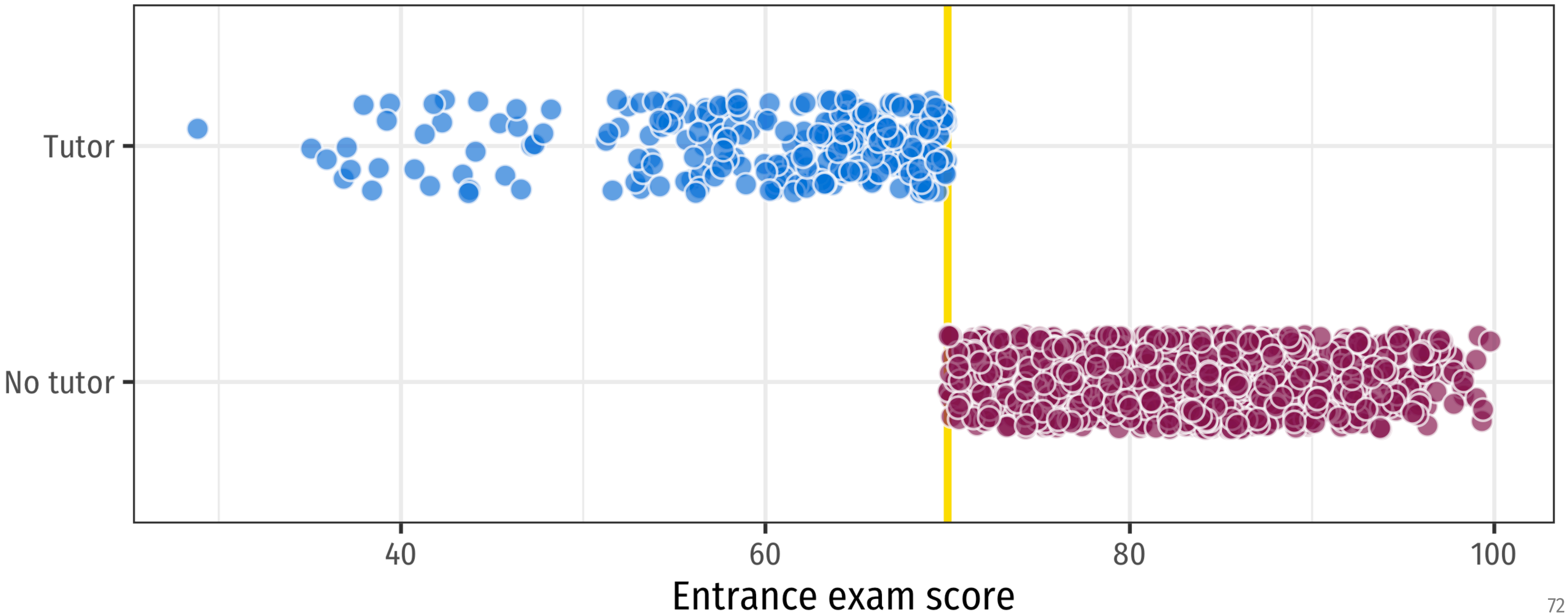
People on the margin of the cutoff might end up in/out of the program

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

Sharp vs. fuzzy discontinuities

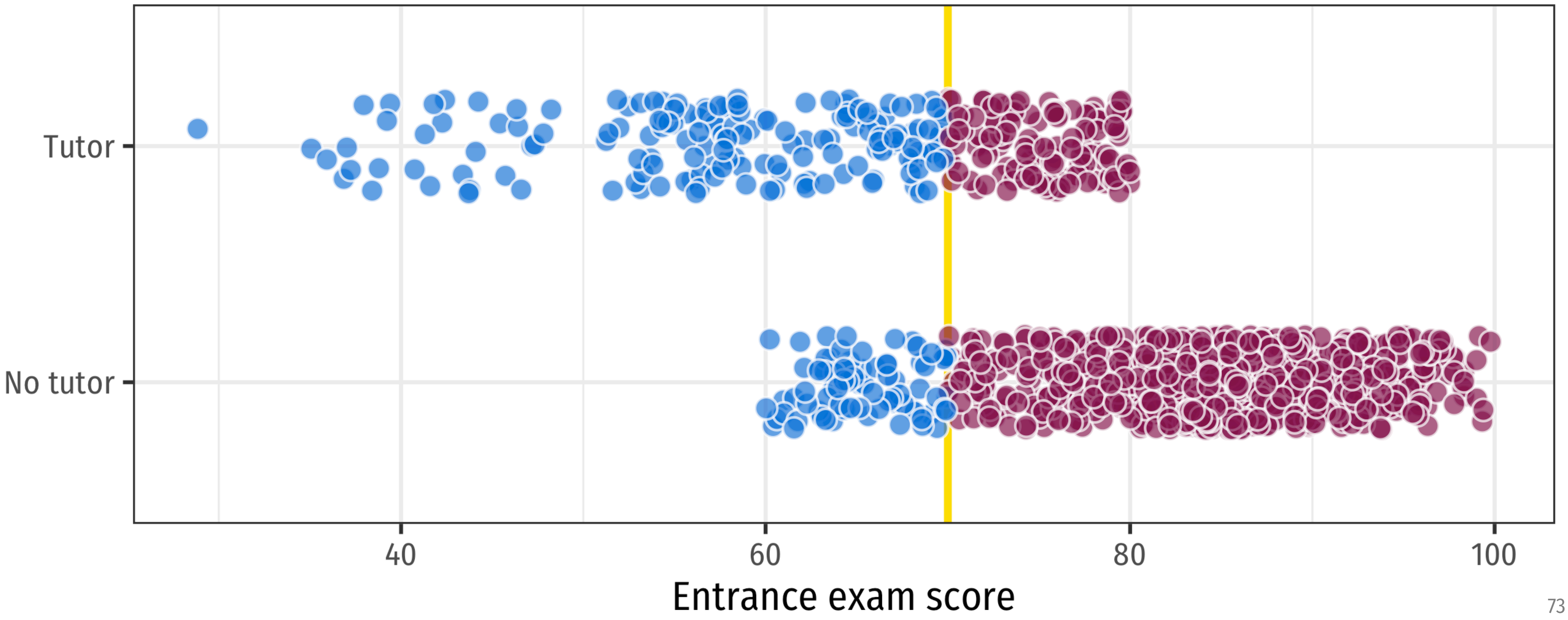
Sharp discontinuity

Perfect compliance



Fuzzy discontinuity

Imperfect compliance



Fuzzy discontinuities

Address noncompliance with instrumental variables (more on this later!)

Use an instrument for which side of the cutoff people should be on

Effect is only for compliers near the cutoff (complier LATE; doubly local effect)