Regression Discontinuity

EDLD 650: Week 4

David D. Liebowitz

Agenda

1. Roadmap and Goals (9:00–9:10)

   - Thoughts on DARE #1

2. Discussion Questions (9:10–10:20)

   - Murnane and Willett
   - Angrist and Lavy
   - Dee and Penner

3. Break (10:20–10:30)
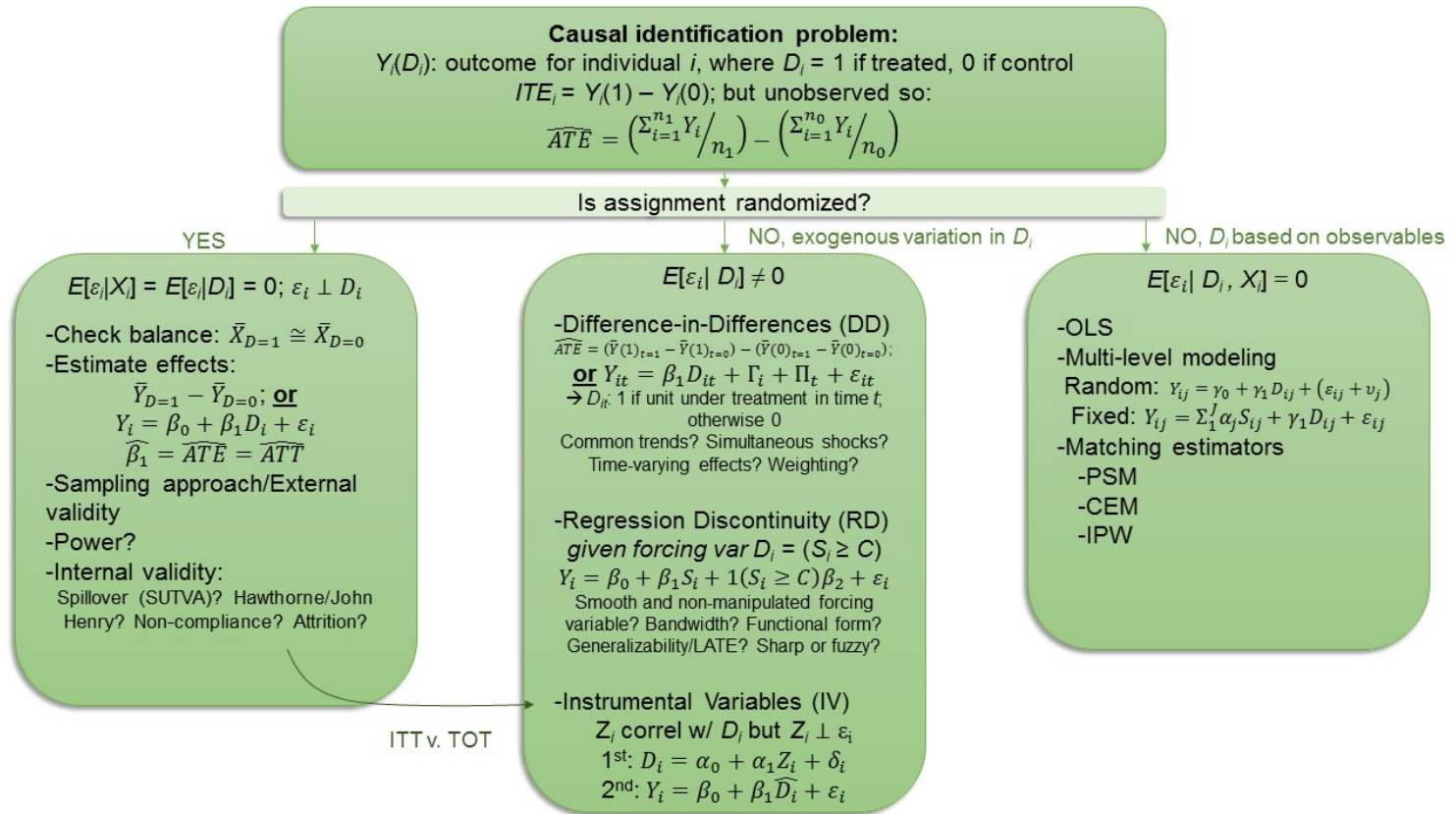
4. Applied regression discontinuity (10:30–11:40)

5. Wrap-up (11:40–11:50)

   - DARE #2 prep

DARE #1: Last words

- You *all* did a good job; many of you have stellar skills in writing functions and/or familiarity with the `tidyverse`
- All DARE exemplars will be substantively consistent in sign/magnitude with paper. Sometimes identical.
  - If you see your results are different, you know that misalignment exists; **that's okay!**
  - Try to solve it, but if you can't write up what you have, note and interpret the differences
- Need to make transition to drafting for public audiences
- Non-causal and causal estimates shouldn't appear side-by-side in tables (w/o very good reason) -- beware Table 2 fallacy!!!
- Prepare for challenge of assignment without model answers -- think hard about model development

Roadmap

Causal identification problem:
$Y_i(D_i)$: outcome for individual i, where $D_i = 1$ if treated, 0 if control
$ITE_i = Y_i(1) - Y_i(0)$; but unobserved so:

$$\widehat{ATE} = \left(\frac{\Sigma_{i=1}^{n_1} Y_i}{n_1}\right) - \left(\frac{\Sigma_{i=1}^{n_0} Y_i}{n_0}\right)$$

Is assignment randomized?

YES

NO, exogenous variation in D_i

NO, D_i based on observables

$E[\varepsilon_i | X_i] = E[\varepsilon_i | D_i] = 0$; $\varepsilon_i \perp D_i$

-Check balance: $\bar{X}_{D=1} \cong \bar{X}_{D=0}$
-Estimate effects:
$\bar{Y}_{D=1} - \bar{Y}_{D=0}$; **or**
$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$
$\widehat{\beta_1} = \widehat{ATE} = \widehat{ATT}$
-Sampling approach/External validity
-Power?
-Internal validity:
  Spillover (SUTVA)? Hawthorne/John Henry? Non-compliance? Attrition?

ITT v. TOT

$E[\varepsilon_i | D_i] \neq 0$

-Difference-in-Differences (DD)
$\widehat{ATE} = (\bar{Y}(1)_{t=1} - \bar{Y}(1)_{t=0}) - (\bar{Y}(0)_{t=1} - \bar{Y}(0)_{t=0})$;
or $Y_{it} = \beta_1 D_{it} + \Gamma_i + \Pi_t + \varepsilon_{it}$
→ D_{it}: 1 if unit under treatment in time t, otherwise 0
Common trends? Simultaneous shocks? Time-varying effects? Weighting?

-Regression Discontinuity (RD)
given forcing var $D_i = (S_i \geq C)$
$Y_i = \beta_0 + \beta_1 S_i + 1(S_i \geq C)\beta_2 + \varepsilon_i$
Smooth and non-manipulated forcing variable? Bandwidth? Functional form? Generalizability/LATE? Sharp or fuzzy?

-Instrumental Variables (IV)
Z_i correl w/ D_i but $Z_i \perp \varepsilon_i$
1st: $D_i = \alpha_0 + \alpha_1 Z_i + \delta_i$
2nd: $Y_i = \beta_0 + \beta_1 \widehat{D_i} + \varepsilon_i$

$E[\varepsilon_i | D_i, X_i] = 0$

-OLS
-Multi-level modeling
  Random: $Y_{ij} = \gamma_0 + \gamma_1 D_{ij} + (\varepsilon_{ij} + v_j)$
  Fixed: $Y_{ij} = \Sigma_1^J \alpha_j S_{ij} + \gamma_1 D_{ij} + \varepsilon_{ij}$
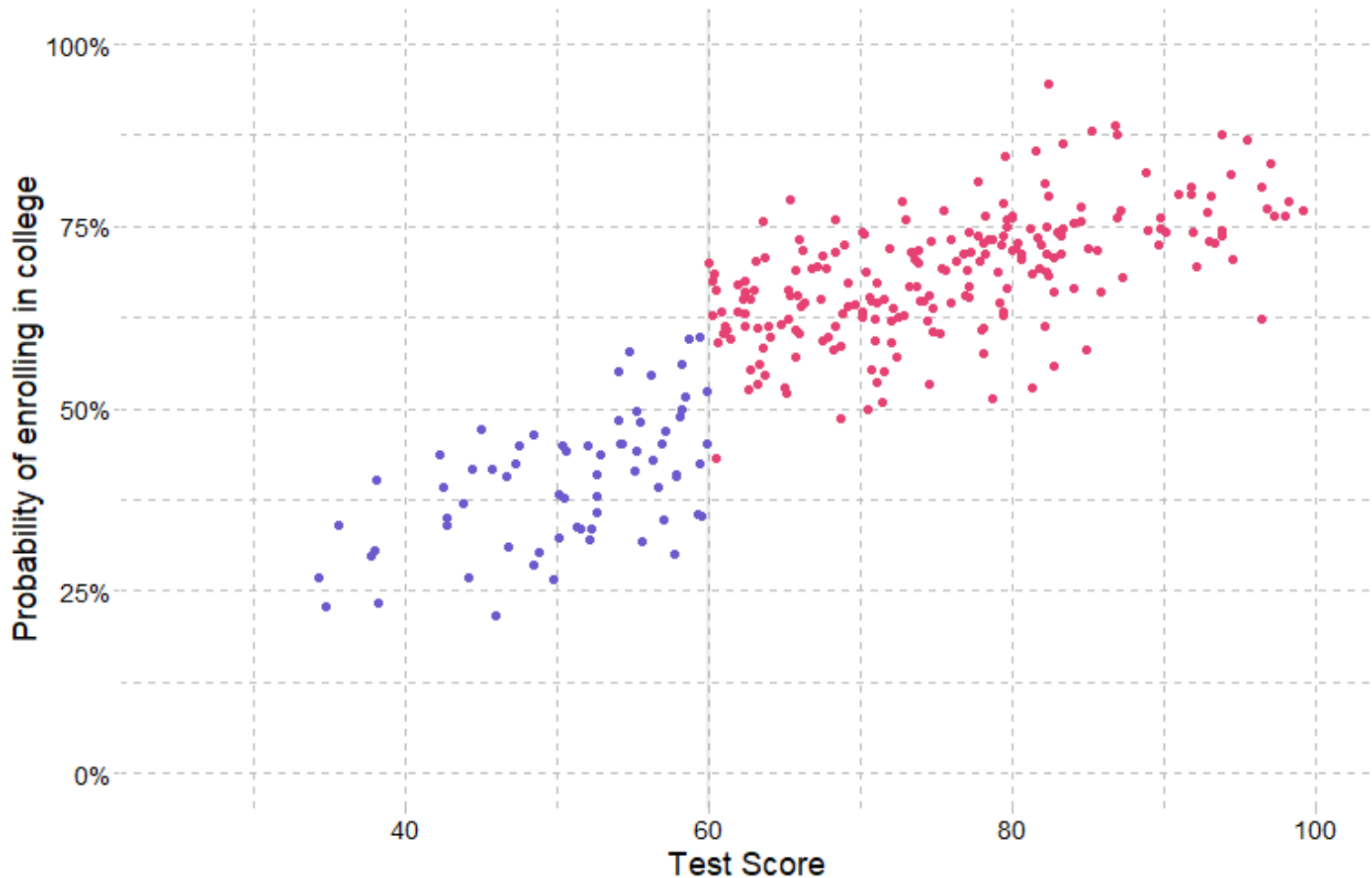-Matching estimators
  -PSM
  -CEM
  -IPW

Goals

1. Describe conceptual approach to regression discontinuity analysis

2. Assess validity of RD assumptions in applied context

3. Conduct and interpret RD analysis in simplified data
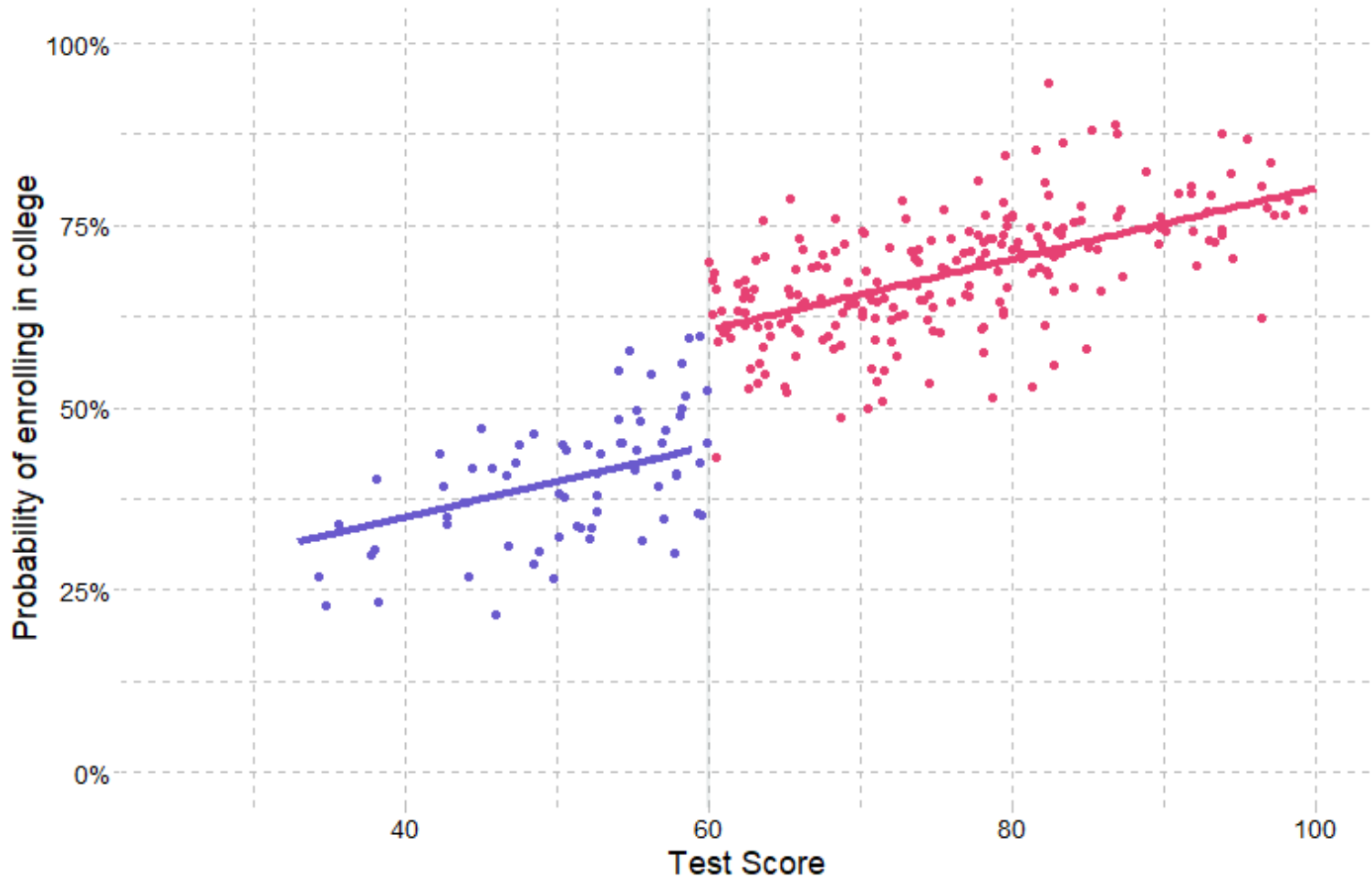
So random...

Break

Recall the basic set up of RD

Failing a graduation test

Failing a graduation test

Failing a graduation test

The basic set up in regression

Given a continuous forcing variable S_i such that individuals receive a treatment (D_i) when $S_i \geq$ a cutoff (C):

$$Y_i = \beta_0 + \beta_1 S_i + 1(S_i \geq C)\beta_2 + \varepsilon_i$$

Can you explain what is happening in this regression?

What about applied in a specific context?

$$p(COLL_i = 1) = \beta_0 + \beta_1 TESTSCORE_i + 1(TESTSCORE_i \geq 60)\beta_2 + \varepsilon_i$$

> This equation estimates a linear probability model, in which whether individuals attend college or not (expressed as a dichomotous indicator taking on the values of 0 or 1), is regressed on a linear measure of individual i's test score ($TESTSCORE_i$) and a indicator variable that takes the value of 1 if individual i scored 60 or higher on the test. β_2 is the causal parameter of interests and represents the discontinuous jump in the probability (p.p.) of attending college (adjusting for test score) of scoring just above the pass score.

Let's practice!

Read in modified Angrist & Lavy data and look at its characteristics:

```
maimonides <- read_dta(here("data/ch9_angrist.dta"))
d <- select(maimonides,
            read, size, intended_classize, observed_classize)
summary(d)
```

```
#>      read              size        intended_classize observed_classize
#>  Min.   :34.80   Min.   :  8.00   Min.   : 8.00      Min.   : 8.00
#>  1st Qu.:69.86   1st Qu.: 50.00   1st Qu.:27.00      1st Qu.:26.00
#>  Median :75.38   Median : 72.00   Median :31.67      Median :31.00
#>  Mean   :74.38   Mean   : 77.74   Mean   :30.96      Mean   :29.94
#>  3rd Qu.:79.84   3rd Qu.:100.00   3rd Qu.:35.67      3rd Qu.:35.00
#>  Max.   :93.86   Max.   :226.00   Max.   :40.00      Max.   :44.00
```

```
sapply(d, sd, na.rm=TRUE)
```

```
#>          read             size intended_classize observed_classize
#>      7.678460        38.810731          6.107924          6.545885
```

Variation in the treatment?
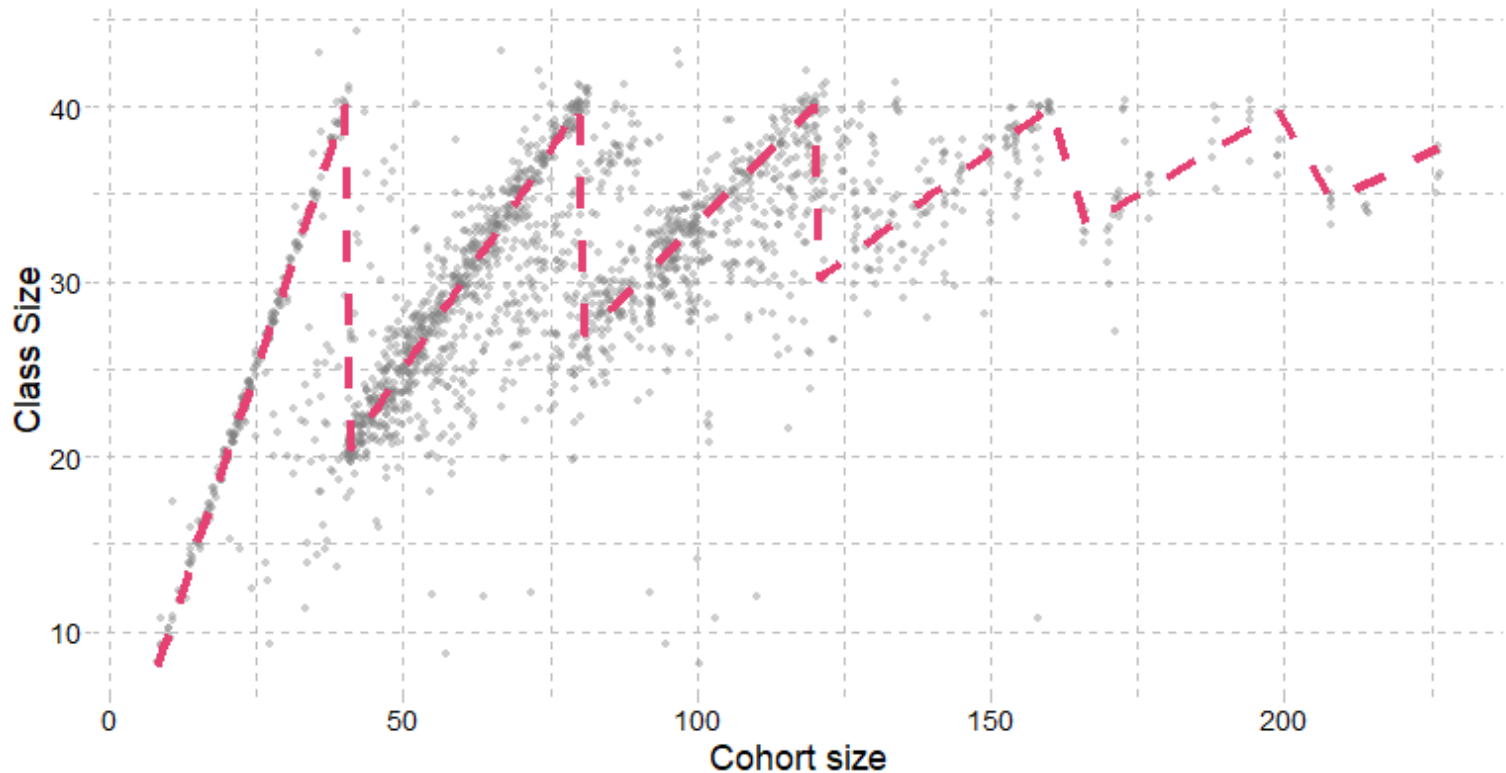
Note that we are plotting the receipt of treatment (actual class size) against the forcing variable (cohort size). What assumption are we testing?

Maimonides Rule Redux

Angrist, J. Lavy, V. Leder–Luis, J. & Shany, A. (2019). Maimonides' rule redux.
American Economic Review: Insights, 1(3), 1–16.
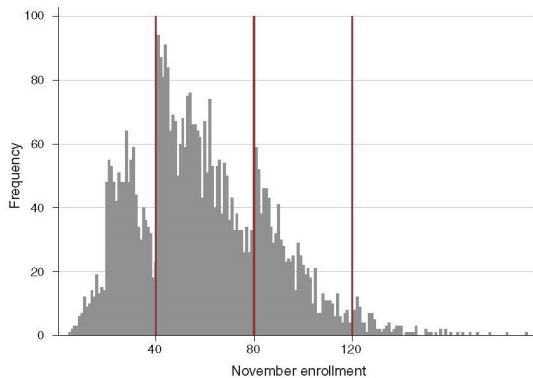
FIGURE 1. THE FIFTH-GRADE ENROLLMENT DISTRIBUTION REPORTED IN NOVEMBER (2002–2011)

Notes: This figure plots the distribution of fifth-grade enrollment as reported by school headmasters in November. Reference lines indicate Maimonides' rule cutoffs at which an additional class is added.

TABLE 1—CLASS SIZE EFFECTS ESTIMATED USING NOVEMBER ENROLLMENT INSTRUMENTS (2002–2011)

	Language			
	OLS (1)	2SLS (2)	2SLS (3)	2SLS (4)
Class size	0.0091 (0.0173)	−0.0220 (0.0314)	−0.0299 (0.0329)	−0.0288 (0.0322)
SES index	−0.4268 (0.0602)	−0.4312 (0.0602)	−0.4281 (0.0603)	−0.4286 (0.0603)
November enrollment	0.0025 (0.0037)	0.0052 (0.0043)	0.0226 (0.0144)	
Enrollment squared/100			−0.0103 (0.0081)	
Piecewise linear trend				0.0147 (0.0094)
Observations	225,108			

Notes: This table reports OLS and 2SLS estimates of equation (3). The endogenous variable is June class size; Maimonides' rule is constructed using November enrollment. Standard errors reported in parentheses are clustered at the school and year level. The dependent variable is a math or language test score. Additional covariates include student characteristics (a gender dummy, parents' years of schooling, number of siblings, a born-in-Israel indicator, and ethnic-origin indicators), year fixed effects, an indicator for religious schools, the school socioeconomic index, and interactions of the socioeconomic index with dummies for the 2002–2003 and 2008–2011 periods. The reported SES coefficient is for 2004–2007.

What does the picture on the left tell you about class size in Israel from 2002–2011?

What does the picture on the right tell you about the effects of class size in Israel from 2002–2011?
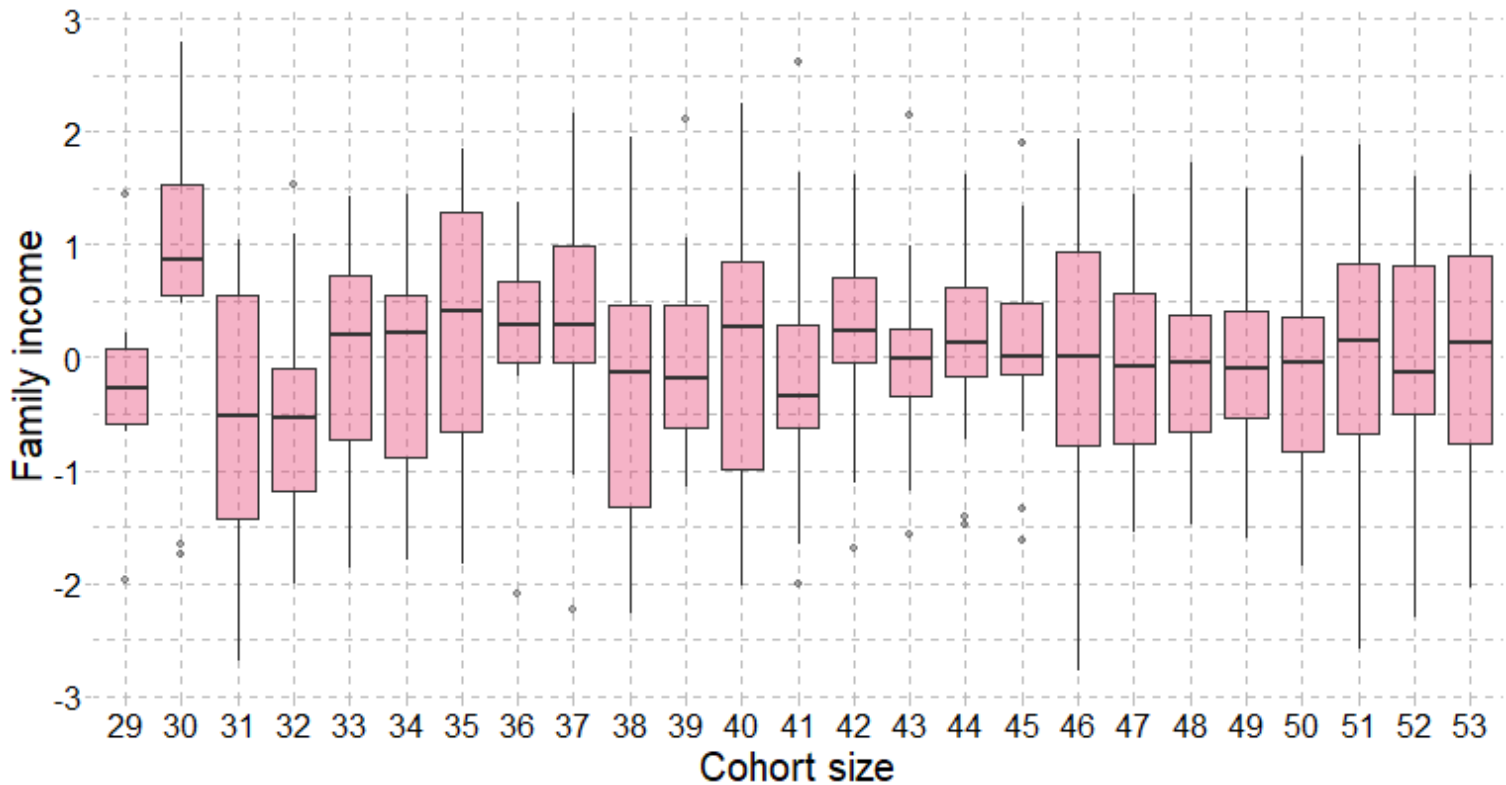
Are RD assumptions met?

```
bunch <- ggplot() +
    geom_histogram(data=d, aes(size), fill=blue, binwidth = 1)
```
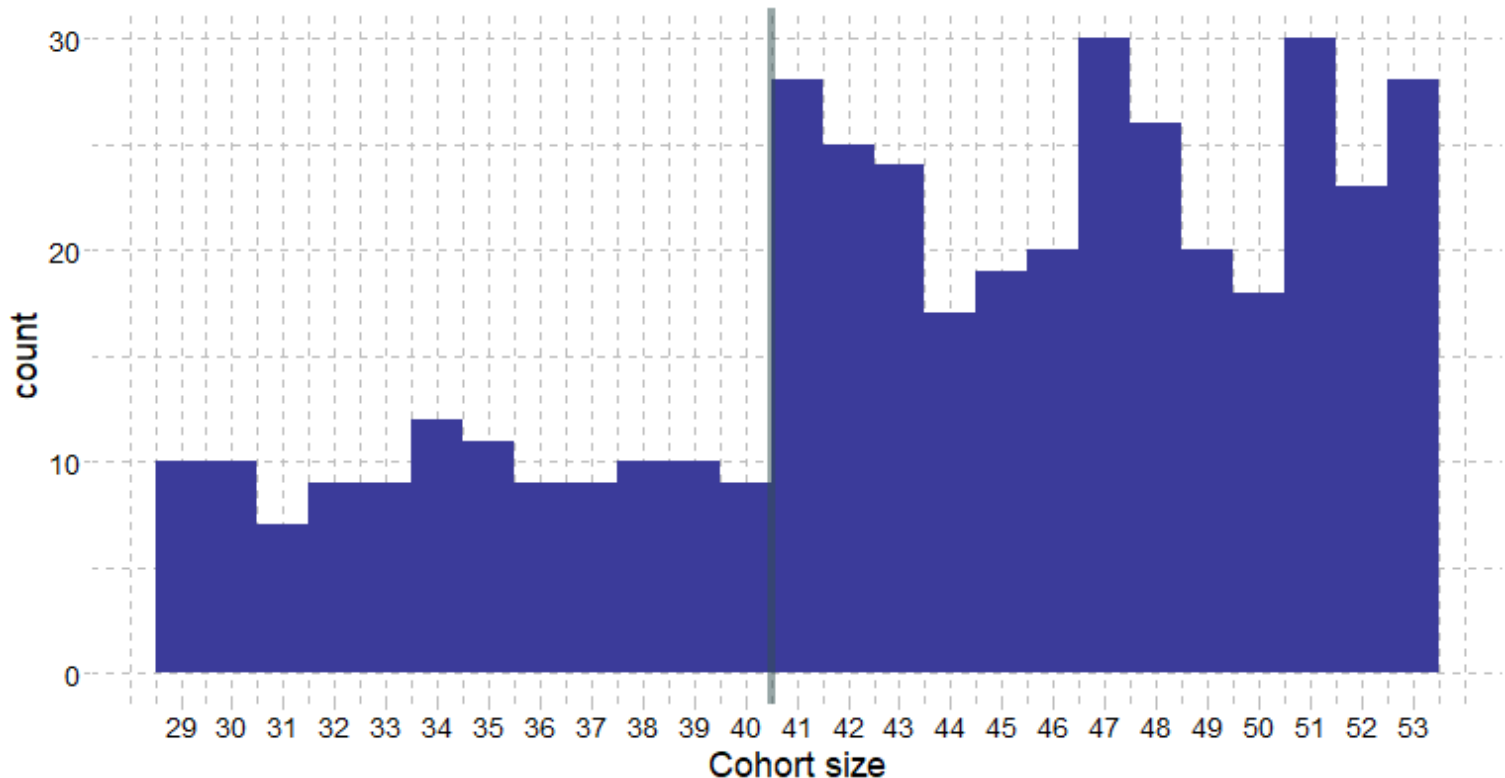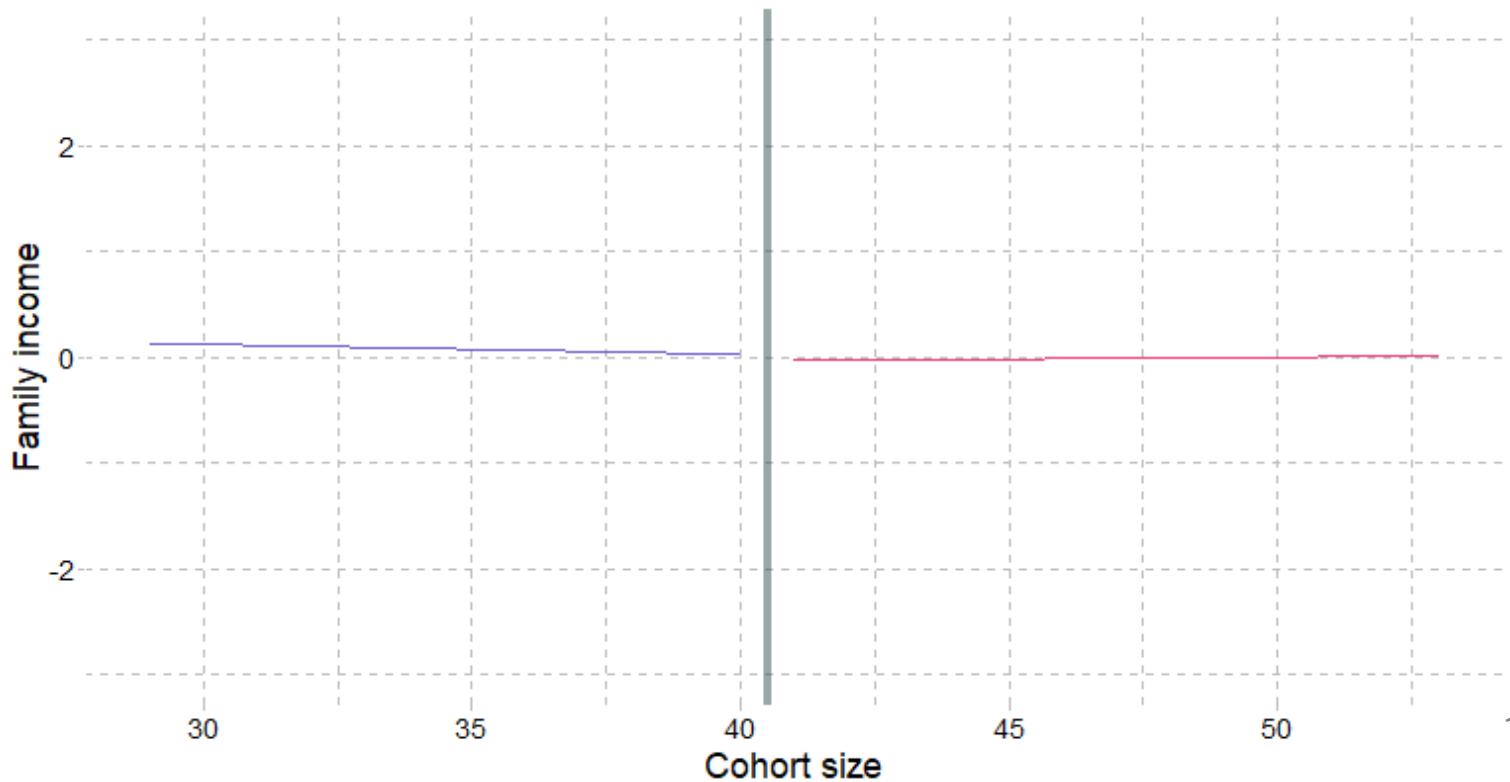
Are RD assumptions met?

```
sort <- ggplot() +
  geom_boxplot(data=d, aes(x=as.factor(size), y=ses),
  fill=red_pink, alpha=0.4)
```
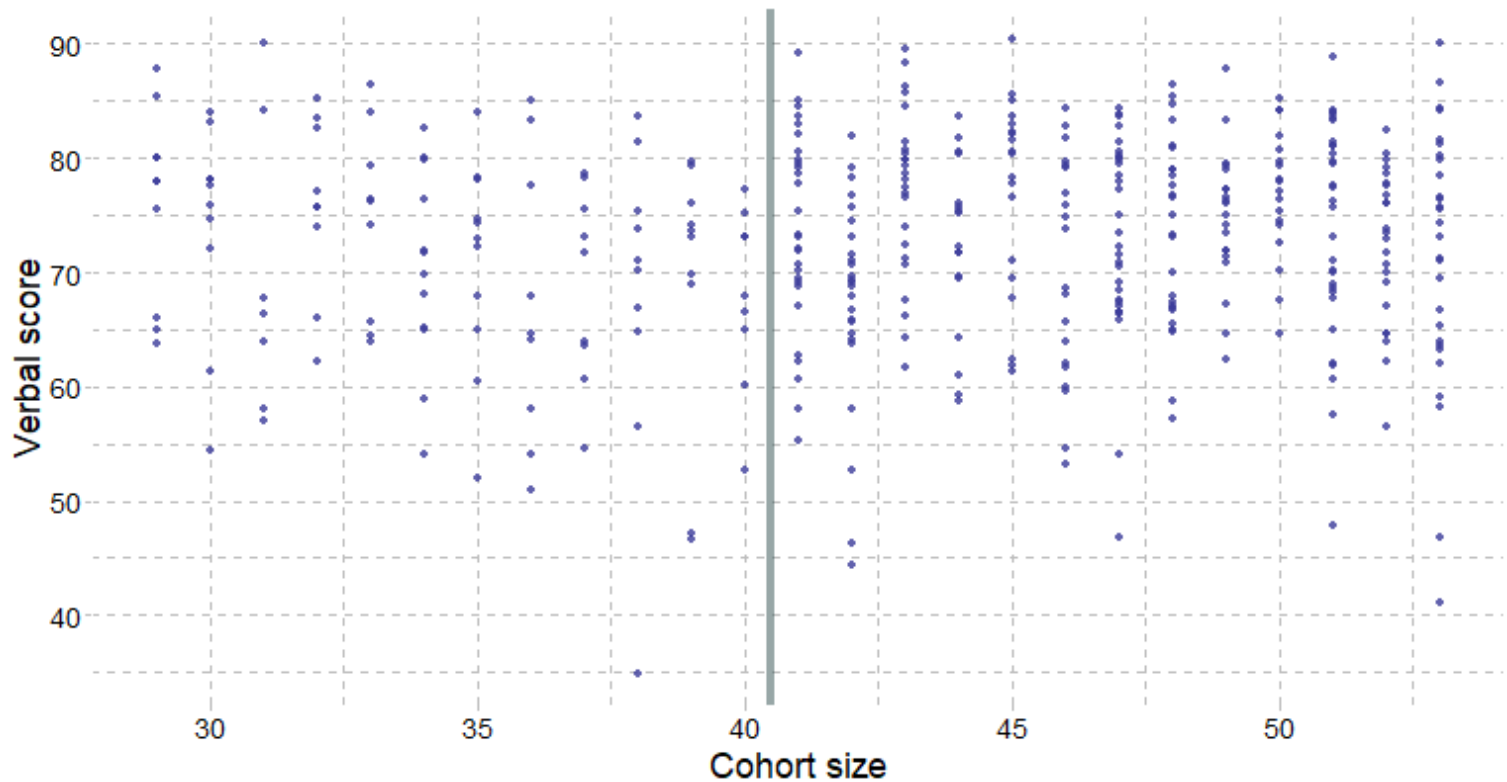
Are RD assumptions met?

```
quantile <- ggplot() +
  geom_quantile(data=filter(d, size<41), aes(size, ses), quantiles=0.
  geom_quantile(data=filter(d, size>=41), aes(size, ses), quantiles=0
```

Let's see if there's an effect

Refresh on how RD works

$$READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + 1(COHORTSIZE_i \geq 41)\beta_2 + \varepsilon_i$$

Could also write this as:

$$READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 SMALLCLASS_i + \varepsilon_i$$

Can you explain the identification strategy as you would in your methods section (using *secular trend, forcing variable, equal in expectation, projecting across the discontinuity, ITT*)?

Refresh on how RD works

We estimate the effects of class size on individual i's reading score. Specifically, we regress their test score outcome on whether the size of their grade cohort predicts that they will be assigned to a small class. We account for the secular relationship between test scores and cohort size by adjusting our estimates for the linear relationship between cohort size and test scores.

Our identification strategy relies on the assumption that cohorts that differ in size by only a few students are equal in expectation prior to the exogenous assignment to a small class size $(D_i = 1)$. Our modeling approach depends on our ability to project a smooth relationship between reading scores and cohort size across the discontinuity and then estimate the discontinuous effect of being quasi–randomly assigned to learn in smaller classes. Given that compliance with Maimonides Rule is imperfect, our approach models Intent–to–Treat estimates. Specifically, what is the effect on reading scores of being assigned by rule to a smaller class size?
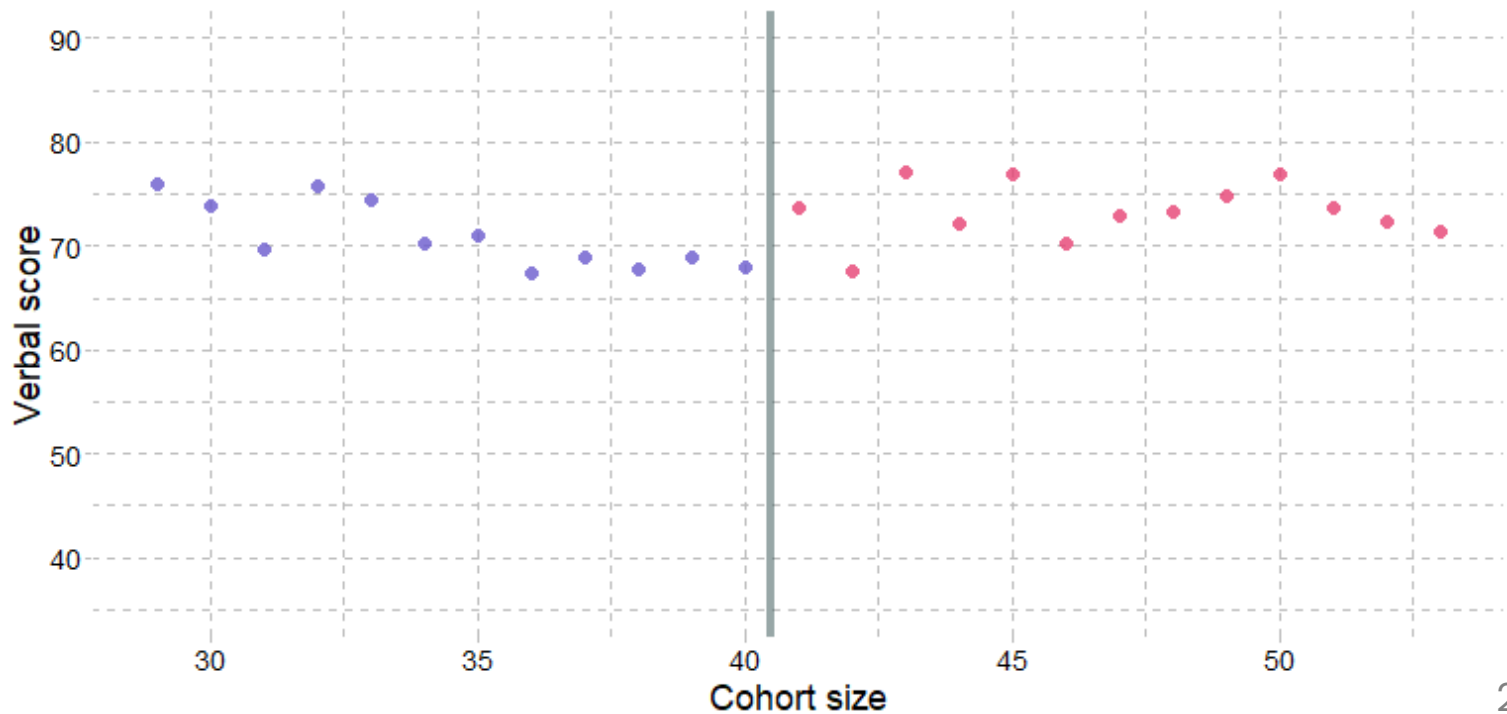
Let's see if there's an effect

```
d <- d %>% mutate(small = ifelse(size >= 41,TRUE,FALSE))

fx2 <- ggplot() +
geom_point(data=d, aes(x=size, y=read, color=small), alpha=0.8, shape
```
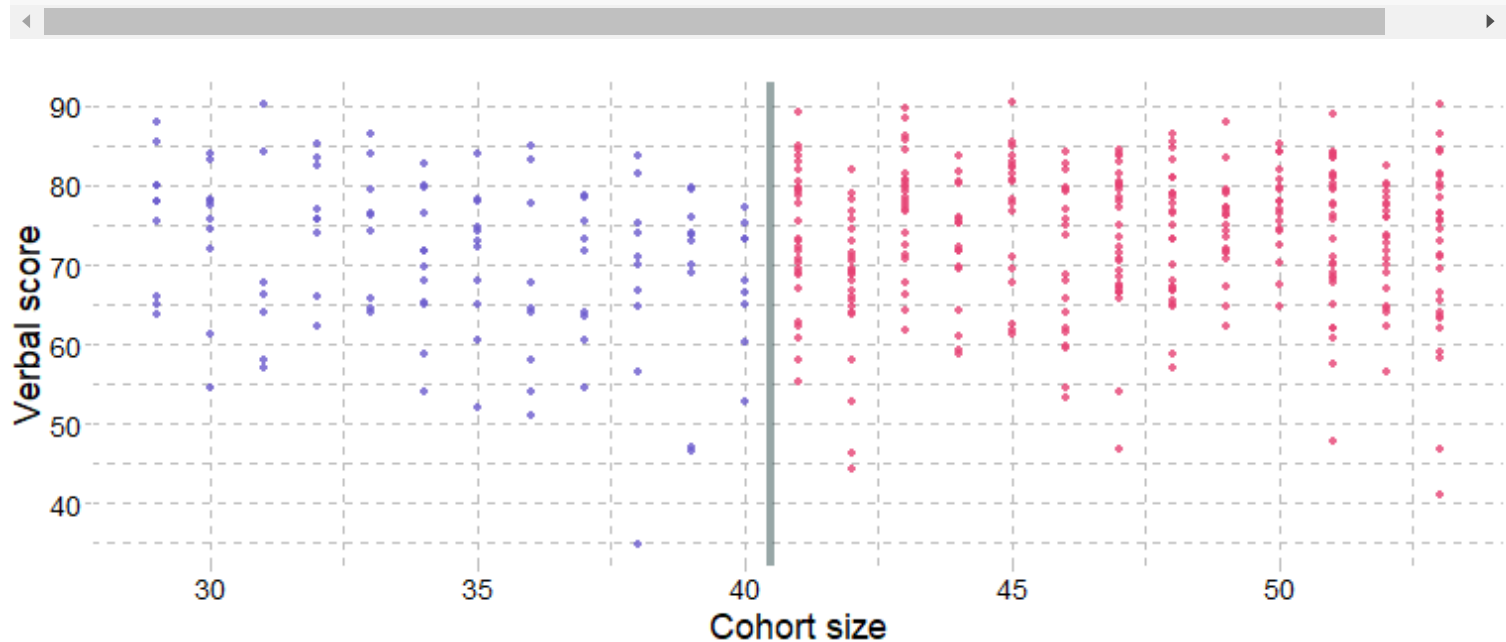

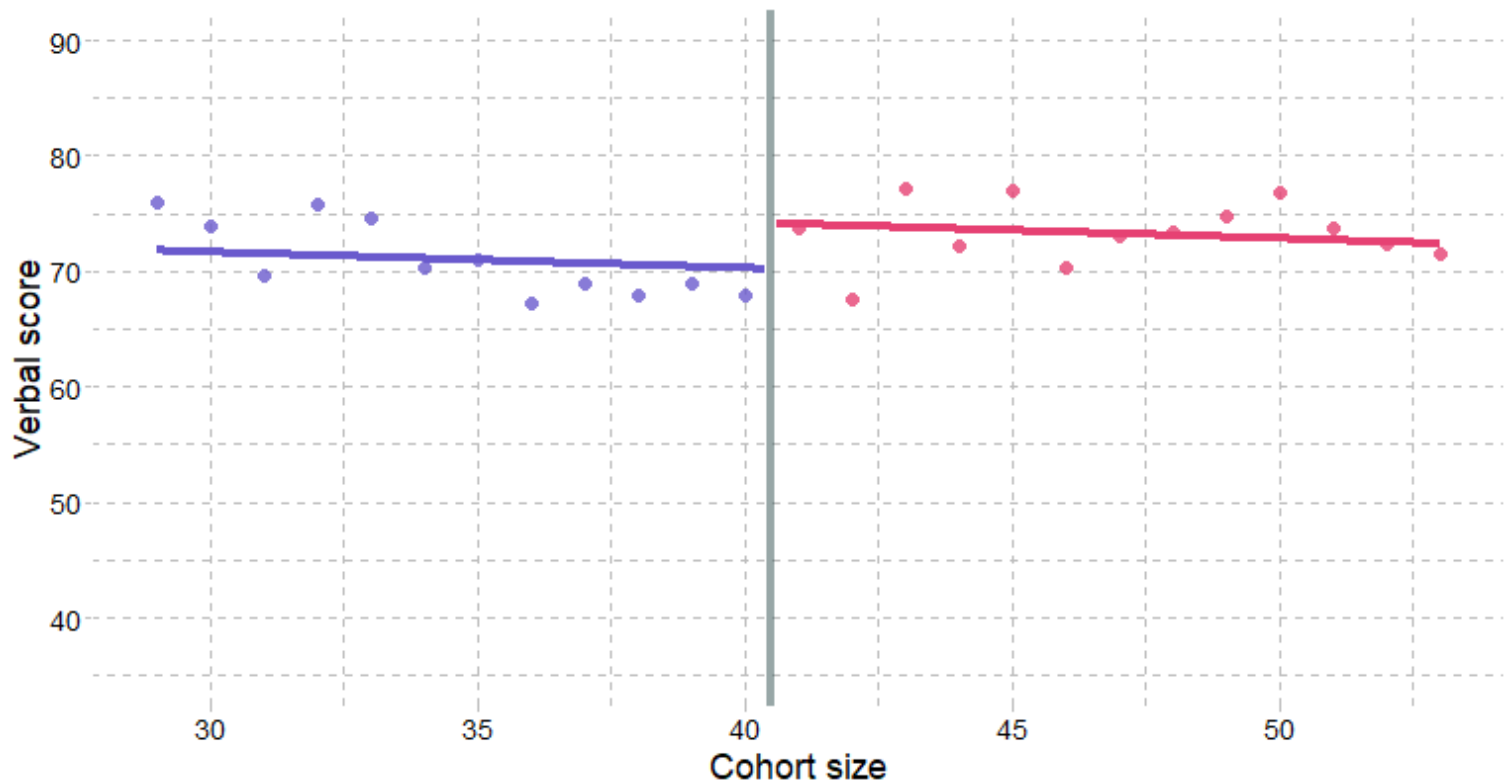Scatter prevents visual detection of discontinuity...thus, the value of bin scatter.

Let's see if there's an effect

```
bin<- d %>% group_by(size) %>% summarise(across(c("read", "small"),

binned_plot <- ggplot() +
  geom_point(data=bin, aes(x=size, y=read, color=as.factor(small)),
```
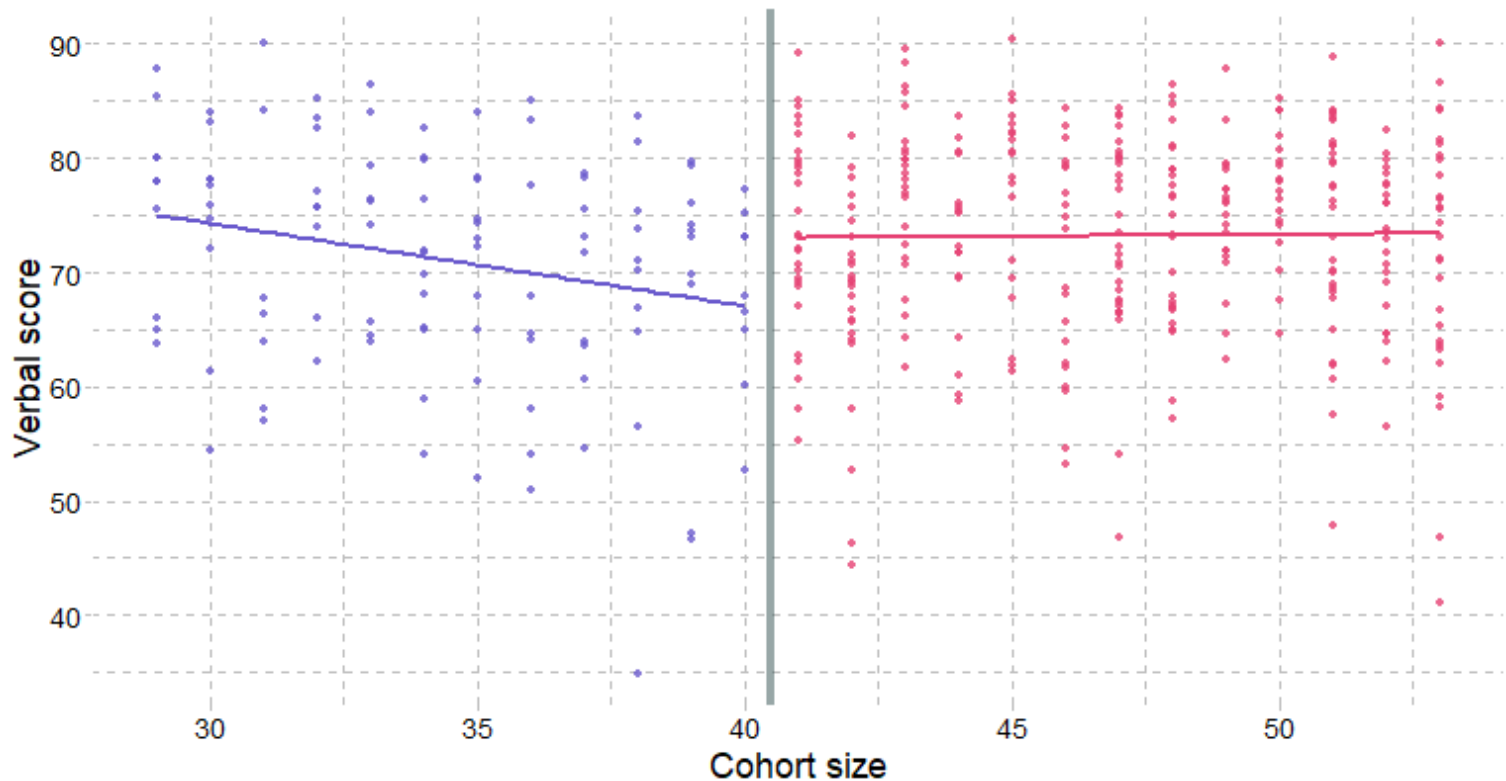
Let's see if there's an effect
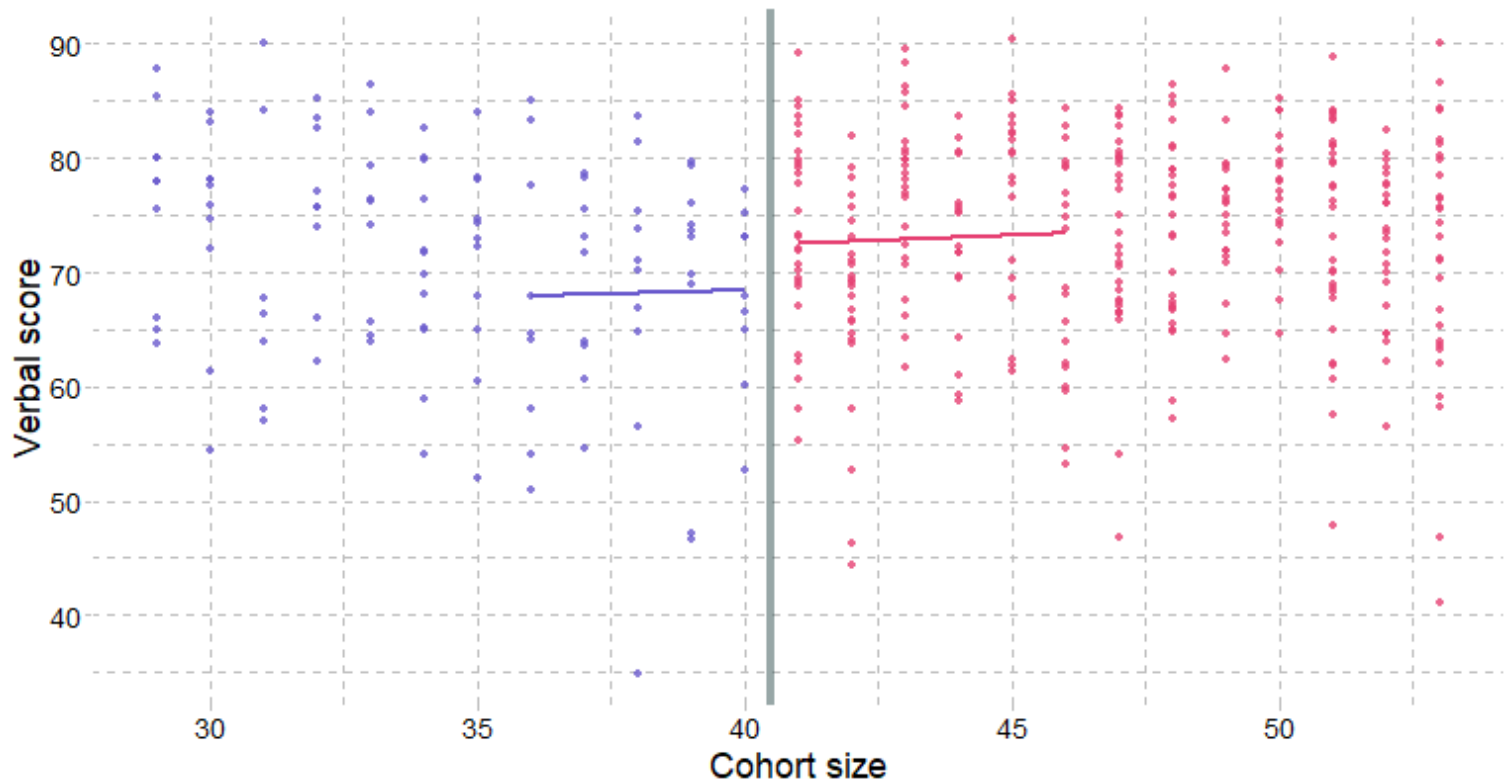
Fitted lines:

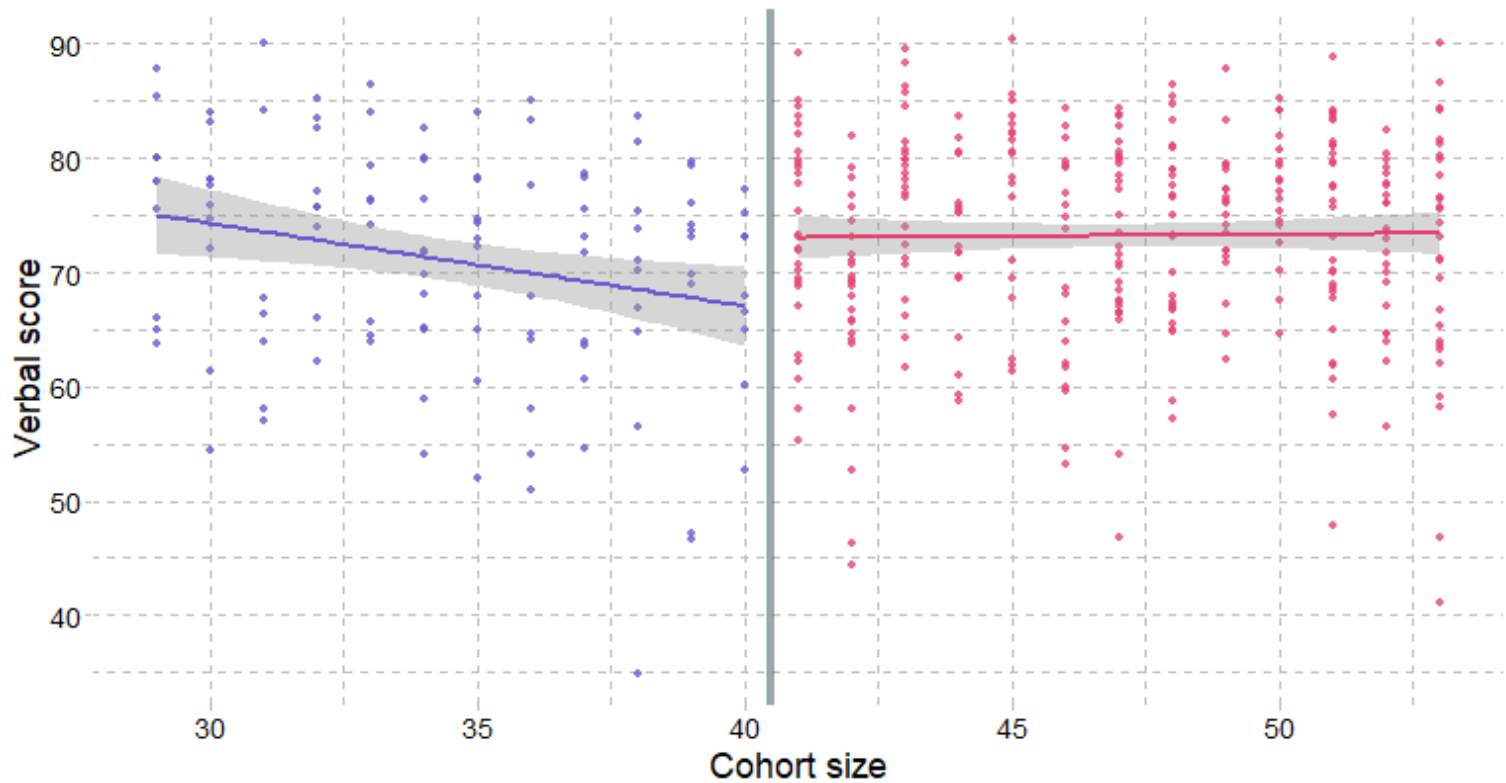# Let's see if there's an effect

Different slopes:

Let's see if there's an effect
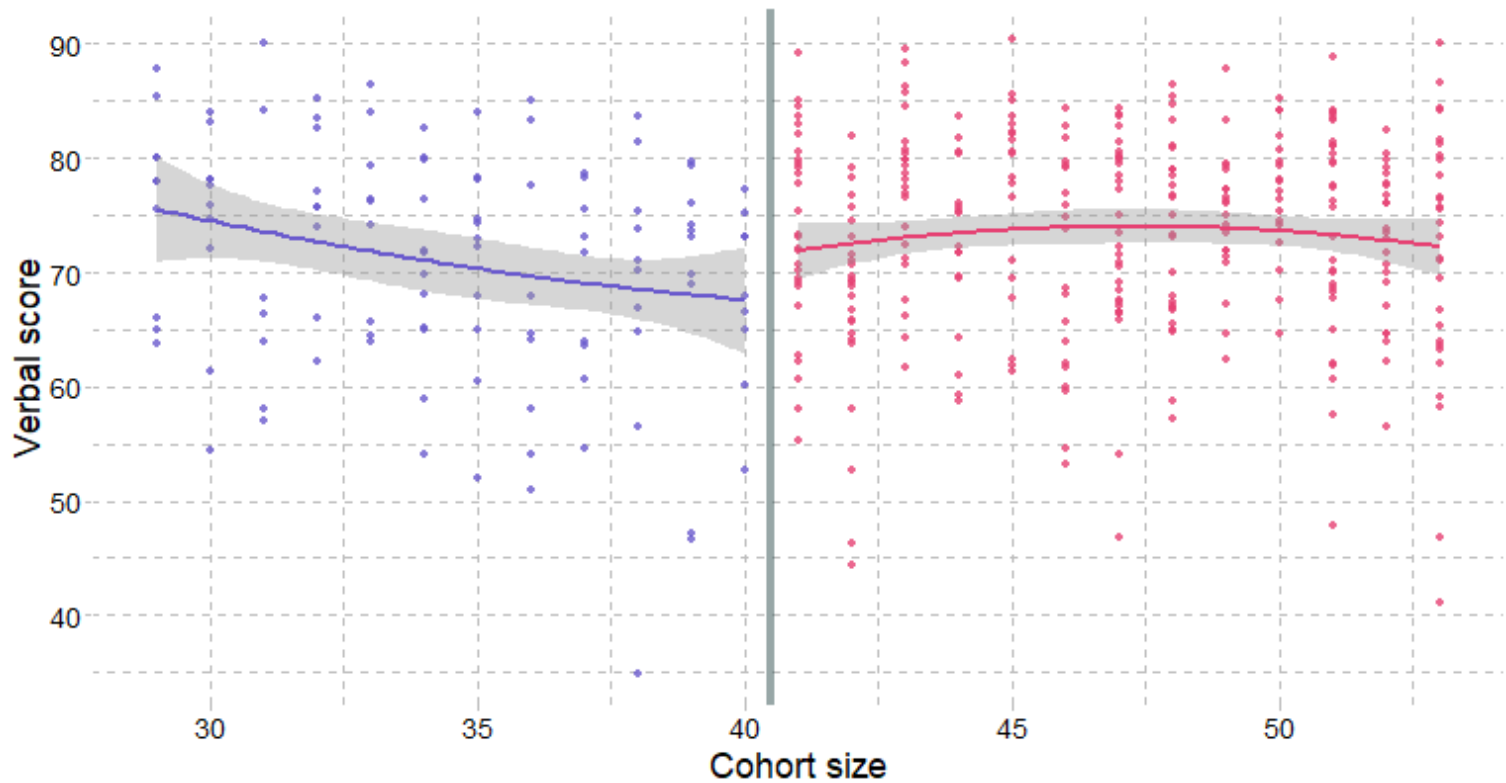
Change the bandwidth:

Let's see if there's an effect

Formal-ish:

But it could be non-linear

Formal-ish:

Regression RD

Let's fit three different intent–to–treat (ITT) RD models, each of which assumes a different functional form for the forcing variable:

Linear trend, same slope

$$(1) READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 SMALLCLASS_i + \epsilon_i$$

Linear trend, different slope

$$(2) READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 SMALLCLASS_i + \\ \beta_3 COHORTSIZE \times SMALLCLASS_i + \epsilon_i$$

Quadratic trend, same slope

$$(3) READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 \text{COHORTSIZE}_i^2 + \\ \beta_3 SMALLCLASS_i + \epsilon_i$$

Results

	(1)	(2)	(3)
Intercept	75.825***	96.046***	68.334***
	(4.202)	(8.621)	(1.516)
Intended size	–0.139	–0.725**	–33.987+
	(0.119)	(0.248)	(17.666)
Intended small class	3.953*	–24.346*	5.894**
	(1.800)	(10.708)	(1.990)
Size x Small		0.757**	
		(0.282)	
(Intended size)^2			22.716*
			(10.154)
Num.Obs.	423	423	423
R2	0.015	0.031	0.026
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001			

Can you explain these results?

Models 1 and 3 seem sensibly connected to the graphical evidence, but Model 2 suggest that the effect of an offer of small class size is negative and an estimated whopping –24.3 points lower reading scores. What gives?

Recall the three models we fit:

$$(1) READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 SMALLCLASS_i + \epsilon_i$$

$$(2) READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 SMALLCLASS_i + \\ \beta_3 COHORTSIZE \times SMALLCLASS_i + \epsilon_i$$

$$(3) READSCORE_i = \beta_0 + \beta_1 COHORTSIZE_i + \beta_2 \text{COHORTSIZE}_i^2 + \\ \beta_3 SMALLCLASS_i + \epsilon_i$$

We need to project the fitted values that our regression results predict *at the discontinuity*. The most straightforward way is to plug in the values for grade cohorts that are just under and over the threshold for being split in two by Maimonides' Rule using the estimated coefficients from the table on the previous slide. Take Eq. 2 and try doing this for cohorts of 40 and 41, respectively.

Can you explain these results?

Big class, grade cohort = 40

$$READ\hat{S}CORE_i = 96.046 + (-0.725)(40) + (-24.346)(0) + (0.757)(40)(0)$$
$$96.046 + (-29) + 0 + 0$$
$$67.046$$

Small Class, grade cohort=41

$$READ\hat{S}CORE_i = 96.046 + (-0.725)(41) + (-24.346)(1) + (0.757)(41)(1)$$
$$96.046 + (-29.725) + (-24.346) + 31.037$$
$$73.012$$

So the predicted effect of being assigned to receive a smaller class when we allow the slopes to vary around the discontinuity is $73.012 - 67.046 = 5.9766$, or slightly larger than either the linear, constant slope or the quadratic specifications.

Note: this is all implicitly solved for when you re-center the forcing variable at 0.

Can you explain these results?

Now that we've harmonized our findings, can you explain these results in technically accurate and substantively clear ways?

> We estimate an effect of being assigned to a small class of between roughly 4 and 6 scale score points, depending on our assumptions about the nature of the underlying secular relationship between cohort size and reading performance. At the lower bound, these represent effects of around one half of a standard deviation (SD) unit. At the upper bound, these effects are as large as three-quarters of a standard deviation in the full sample. These estimates are Local Average Treatment Effects (LATE), specific to being a member of a cohort whose size is just above or below the threshold for being divided into a smaller class.

Extensions

1. Bandwidth variation (bias v. variance tradeoff)
    - Manual
    - Cross-fold validation (leave-one-out)
    - Imbens-Kalyanaraman (2009) Optimal Bandwidth Calculation
2. Higher-order polynomials
3. Non-parametric estimates
    - Local-linear approaches (LOESS)
    - Kernel (how to value points closest to cutoff)
    - Machine learning
4. Binning for visualizations
5. Diff-in-RD
6. Packages
    - R: `rddapp`, `rdd`, `rddtools`, `rdrobust`

Just for fun...
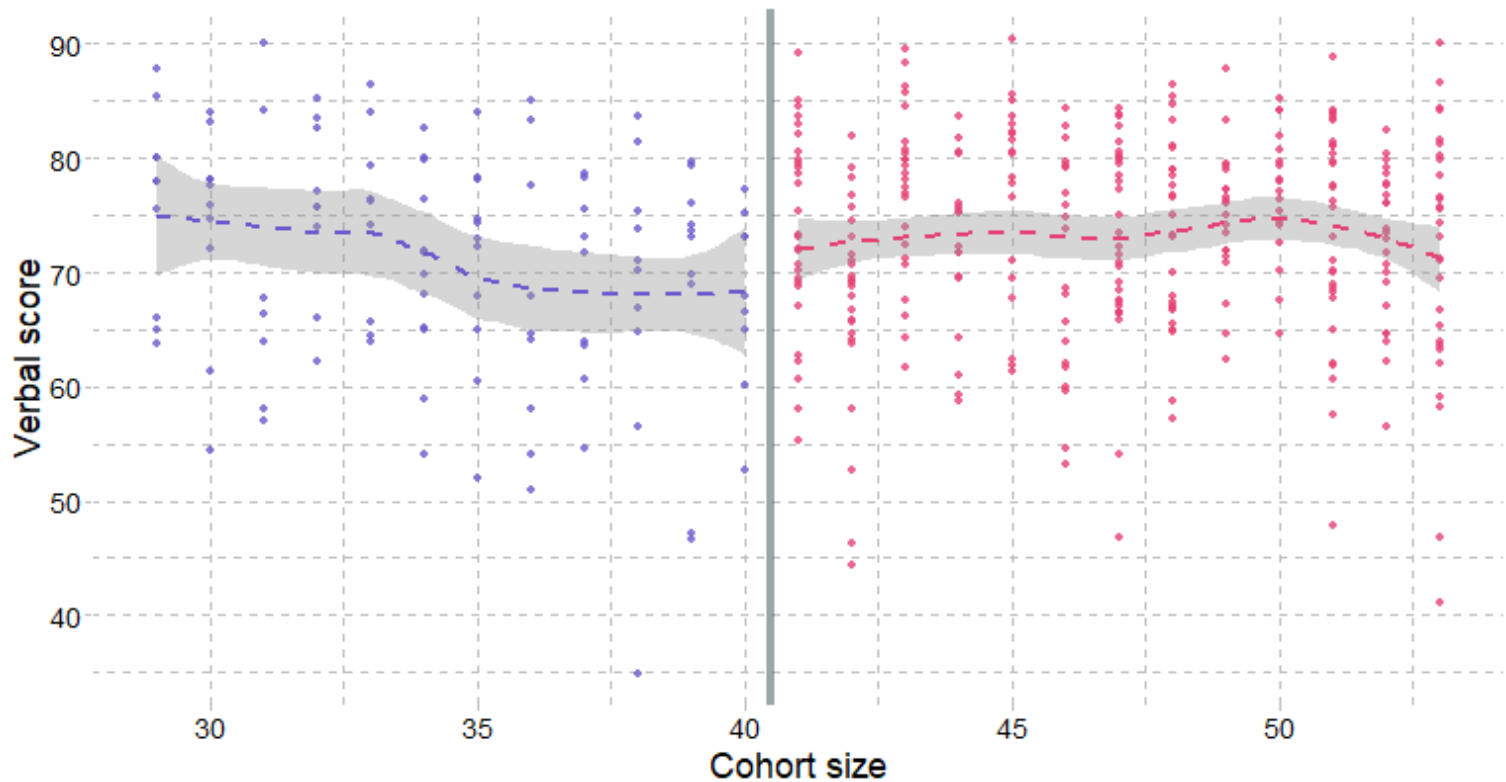
Wrap-up

Goals

1. Describe conceptual approach to regression discontinuity analysis

2. Assess validity of RD assumptions in applied context

3. Conduct RD analysis in simplified data

To-Dos

Week 5: Regression Discontinuity II

Readings:

- Holden (2016)

DARE #2

- Due 2/4, 11:59pm

Project proposal

- Due 2/2, 11:59pm

Feedback

Midterm Student Experience Survey