Poli-301:

The Political Science Discipline

TODAY'S AGENDA

1. Finish up matching
2. Difference-in-difference

Housekeeping

Thanksgiving week

No class, but a problem set that I'll move to that week

Will update syllabus to reflect this

Last time

Some people/places got some treatment, others didn't

Want to find people/places that are otherwise similar, but didn't get treatment

We can do this by matching on people on background characteristics

CEM by hand

Name	treatment	age	grade	height	smoker	income
Lucion	1	22	86.6	71.7	0.47	3.53E+04
Leaton	1	28	85.2	68.2	0.63	4.66E+04
Tnaya	0	20	85.5	63.3	0.35	2.63E+04
Anaila	1	26	90.7	68.4	0.57	1.24E+04
Kelden	1	28	84.4	74.8	0.52	9.52E+04
Roben	0	21	80	71.3	0.3	4.21E+04
Matviy	0	18	82.4	64.4	0.38	1.91E+04
Shamia	0	24	88.9	68	0.23	5.62E+04

Step 1: separate out T and C

```
df_treat = df %>% filter(treatment == 1)
```

Table 1

Name	treatment	age	grade	height	smoker	income	sex
Jalyce	1	22	80.6	74.1	0.62	2.01E+04	Male
Yomar	1	24	89.4	75.4	0.44	5.94E+04	Female
Ayslee	1	26	84.2	72	0.65	1.05E+05	Female
Anniece	1	26	81.2	72.4	0.43	1.35E+05	Female
Andrine	1	27	86.9	73	0.56	4.08E+04	Female

```
df_treat = df %>% filter(treatment == 0)
```

Table 1-1

Name	treatment	age	grade	height	smoker	income	sex
Jozalynn	0	22	87.4	67.4	0.25	2.39E+04	Female
Erminio	0	19	90.4	70.7	0.1	2.57E+04	Female
Ivani	0	19	84.2	66.6	0.24	4.17E+04	Male
Mykhel	0	19	87.3	68.1	0.38	3.46E+04	Male
Salama	0	20	83.7	70.3	0.11	7.61E+03	Female

Step 2: coarsen variables

Unlikely that two people have the exact same income down to the cents

Need to "coarsen" variables so we can get a match

"Coarser" variables = more matches, worse quality

```
df = df %>%
mutate(age_coarse = cut_interval(age, 3),
       grade_coarse = cut_interval(grade, 3),
       height_coarse = cut_interval(height, 3),
       smoker_coarse = cut_interval(smoker, 3))
```

Step 2: coarse

Name	treatment	age_coarse	grade_coarse	height_coarse	smoker_coarse	sex
Dustin	1	(21.3,24.7]	[72.6,81.7]	(65.7,74.1]	(0.233,0.467]	Female
Claryce	1	(24.7,28]	(81.7,90.9]	(65.7,74.1]	(0.233,0.467]	Female
Kennell	1	(24.7,28]	(81.7,90.9]	(65.7,74.1]	(0.467,0.7]	Female
Cytlalli	1	(24.7,28]	(81.7,90.9]	(74.1,82.4]	(0.233,0.467]	Female
Nilah	1	(24.7,28]	(81.7,90.9]	(65.7,74.1]	(0.233,0.467]	Female

sex	age_coarse	grade_coarse	height_coarse	smoker_coarse	untreated_income
Male	(21.3,24.7]	[72.6,81.7]	(65.7,74.1]	(0.233,0.467]	2.73E+04
Female	[18,21.3]	(90.9,100]	(74.1,82.4]	(0.233,0.467]	2.36E+04
Male	(21.3,24.7]	(81.7,90.9]	(74.1,82.4]	[0,0.233]	2.1E+04
Female	[18,21.3]	(81.7,90.9]	[57.4,65.7]	[0,0.233]	4.55E+04
Female	[18,21.3]	(90.9,100]	(65.7,74.1]	(0.233,0.467]	4.5E+04

Step 3: match

We combine the two data frames using join

```
matched = inner_join(df_treat, df_control)
```

Original = 500
Treatment = 244
Control = 256
Matched = 30

In R

~~matchit(treatment ~ w1 + w2 + w3, data = df, method = “cem”)~~

CEM requires too much build up in R

CEM: every treatment unit can get zero matches, one match, or multiple matches

Nearest neighbor:
every treatment unit gets exactly one match

Today

Difference-in-differences

One of the most commonly used methods

A sorta combo of matching and FE

We have a treated group, and we observe it both before and after treatment

We have a control group, and we observe it for the same time period

Do women want to work more or more regularly? Evidence from a natural experiment[*]

Emma Duchini[†], Clémentine Van Effenterre[‡]

This version: October 2018
Click here for most recent version

Abstract

This paper provides causal evidence that children limit women's chances of having a regular Monday-Friday working schedule. We show that this constraint contributes to the persistence of the gender wage gap. Historically, French children in primary school have had no school on Wednesdays. In 2013, a reform reallocated some classes to Wednesday mornings. Exploiting variations in the implementation of this reform over time and across the age of the youngest child, we demonstrate that, once institutional constraints are relaxed, mothers are more likely to work on Wednesdays and to work full-time. By working longer and more regular hours, mothers are able to close 6 percent of the gender wage gap. These effects on hours and wages are driven by high-skilled women. We show that a very simple theoretical framework can rationalize these findings.

ABSTRACT

Semesters or Quarters? The Effect of the Academic Calendar on Postsecondary Student Outcomes*

We examine the impact of US colleges and universities switching from an academic quarter calendar to a semester calendar on student outcomes. Using panel data on the near universe of four-year nonprofit institutions and leveraging quasi-experimental variation in calendars across institutions and years, we show that switching from quarters to semesters negatively impacts on-time graduation rates. Event study analyses show that these negative effects persist well beyond the transition. Using detailed administrative transcript data from one large state system, we replicate this analysis at the student-level and investigate several possible mechanisms. We find shifting to a semester: (1) lowers first-year grades; (2) decreases the probability of enrolling in a full course load; and (3) delays the timing of major choice. By linking transcript data with the Unemployment Insurance system, we find minimal evidence that a semester calendar leads to increases in summer internship-type employment.

From Immigrants to Americans: Race and Assimilation during the Great Migration

Vasiliki Fouka[†] Soumyajit Mazumder[‡] Marco Tabellini[§]

June 2019

Abstract

How does the appearance of a new immigrant group affect the integration of earlier generations of migrants? We study this question in the context of the first Great Migration (1915-1930), when 1.5 million African Americans moved from the US South to northern urban centers, where 30 million Europeans had arrived since 1850. We exploit plausibly exogenous variation induced by the interaction between 1900 settlements of southern-born blacks in northern cities and state-level outmigration from the US South after 1910. Black arrivals increased both the effort exerted by immigrants to assimilate and their eventual Americanization. These average effects mask substantial heterogeneity: while initially less integrated groups (i.e. Southern and Eastern Europeans) exerted more assimilation effort, assimilation success was larger for those culturally closer to native whites (i.e. Western and Northern Europeans). Labor market outcomes do not display similar heterogeneity, suggesting that these patterns cannot be entirely explained by economic forces. Our findings are instead more consistent with a framework in which changing perceptions of outgroup distance among native whites lowered the barriers to the assimilation of white immigrants.

Pokemon Go!

POKéMON GO
SAFARI ZONE
KØBENHAVN
STOCKHOLM
AMSTERDAM
PARIS
PRAHA
OBERHAUSEN
0:00:01
BARCELONA

Pokemon Go!

Effects of Pokemon go

Pokemon Go involves a ton of walking to find/catch Pokemon

Did Pokemon Go! make people more active?

Seems kinda obvious, but how can we know?

Before vs. after

Average daily miles walked in US among 14-18 year olds

Before: .8 miles

After: 1.2 miles

Does this show a causal effect?

Treatment vs. control

Average daily miles walked in US among 14-18 year olds

Don’t have Pokemon Go: .7 miles

Have Pokemon Go: .9 miles

Does this show a causal effect?

PROBLEMS

Compare only treat and control

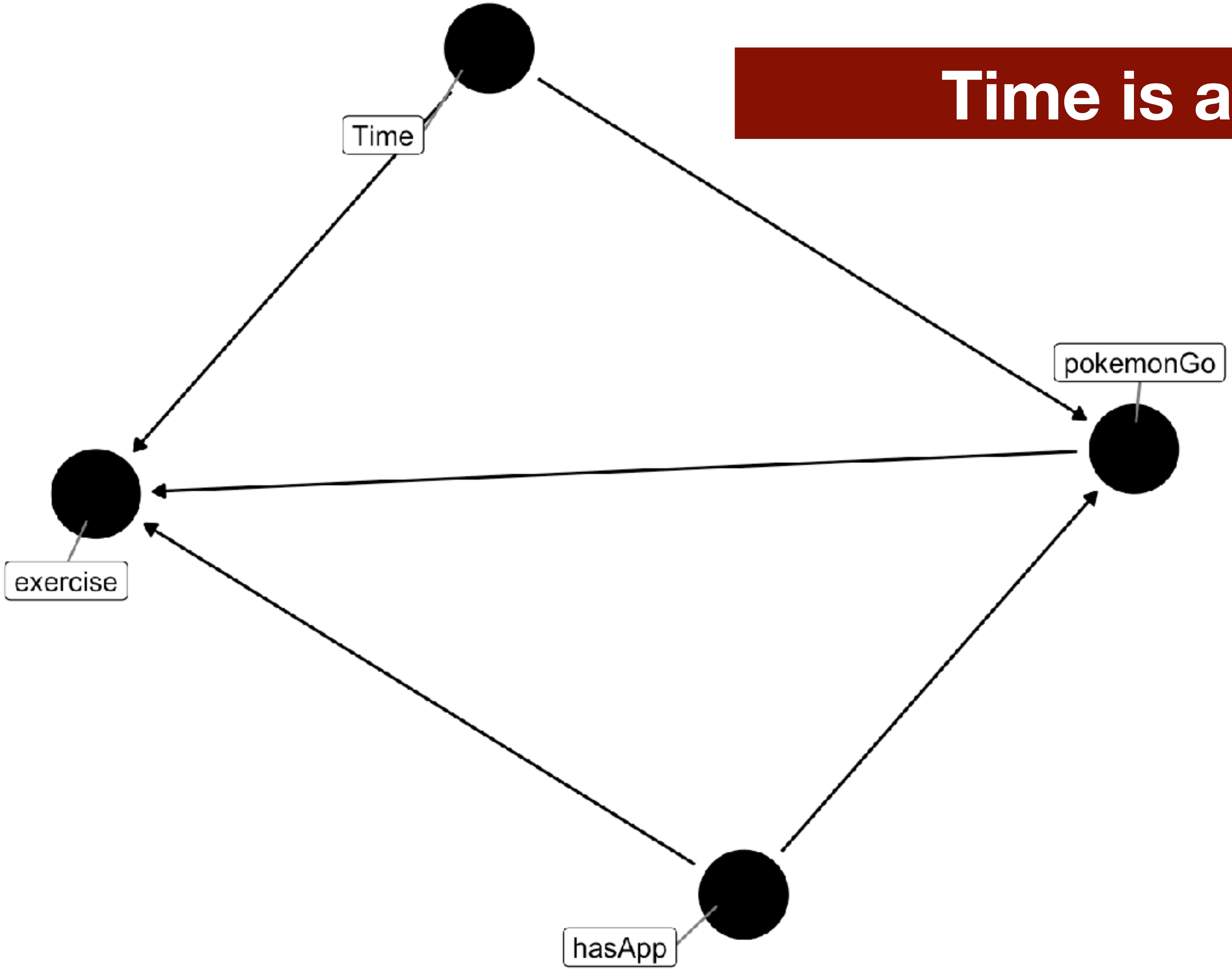
Backdoors: having the app vs. not having the app

Compare only before and after

Did growth happen *because* of app or was it just naturally?

THE DAG

Time is a problem!



SOLUTION

Again, how do we know that pre- and post-app changes in exercise aren't just natural growth?

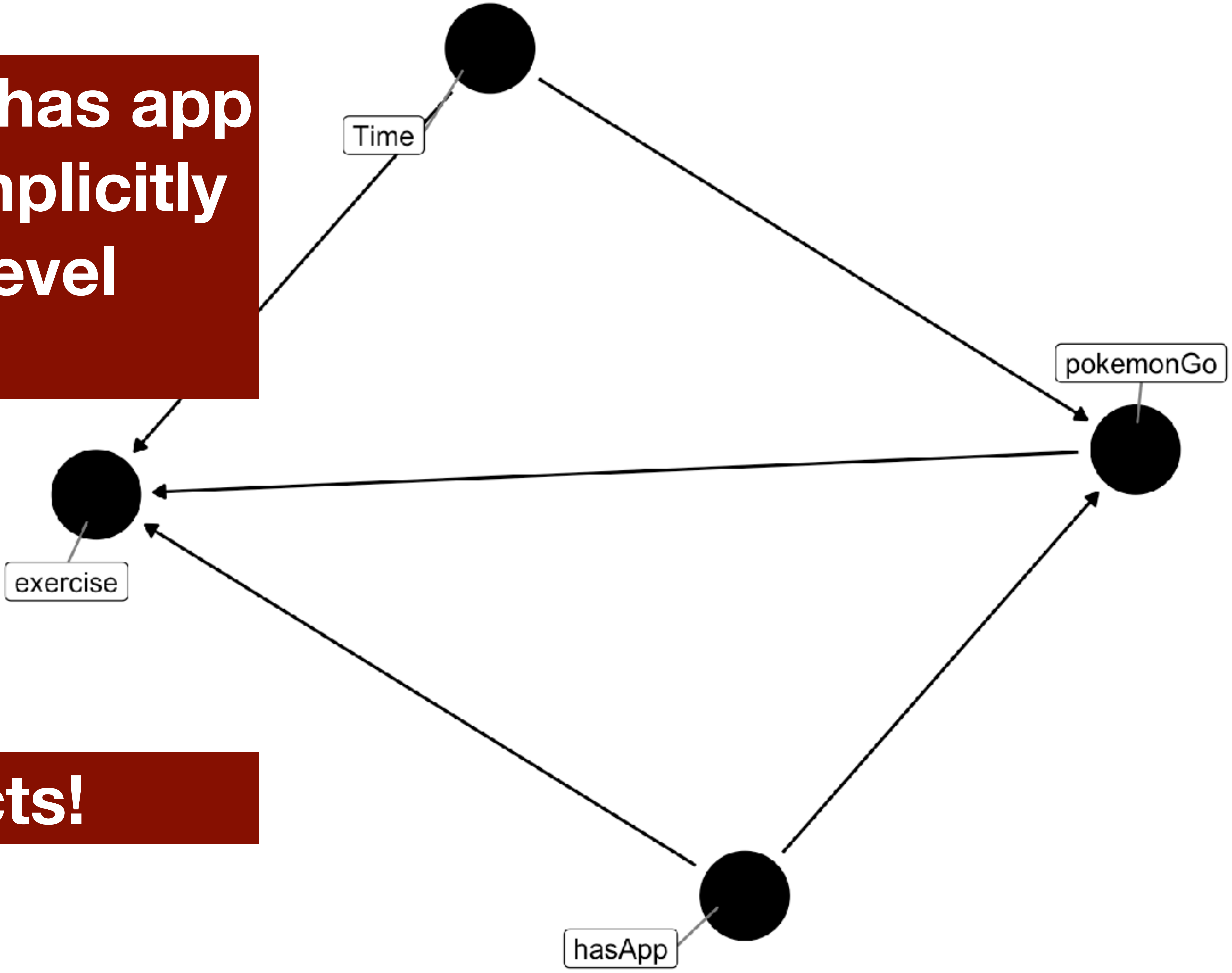
Any changes that are the result of time (just natural growth) should *also* show up in the control units (kids without the app)

So we compare pre-to-post changes among treated units and control units

SOLUTION

In comparing each group (has app vs. not) to itself, we are implicitly controlling for group-level differences

This is like fixed effects!



	Pre-mean	Post-mean	Difference (post-pre)
Treatment	A (not yet treated)	B (treated)	B - A
Control	C (never treated)	D (never treated)	D - C

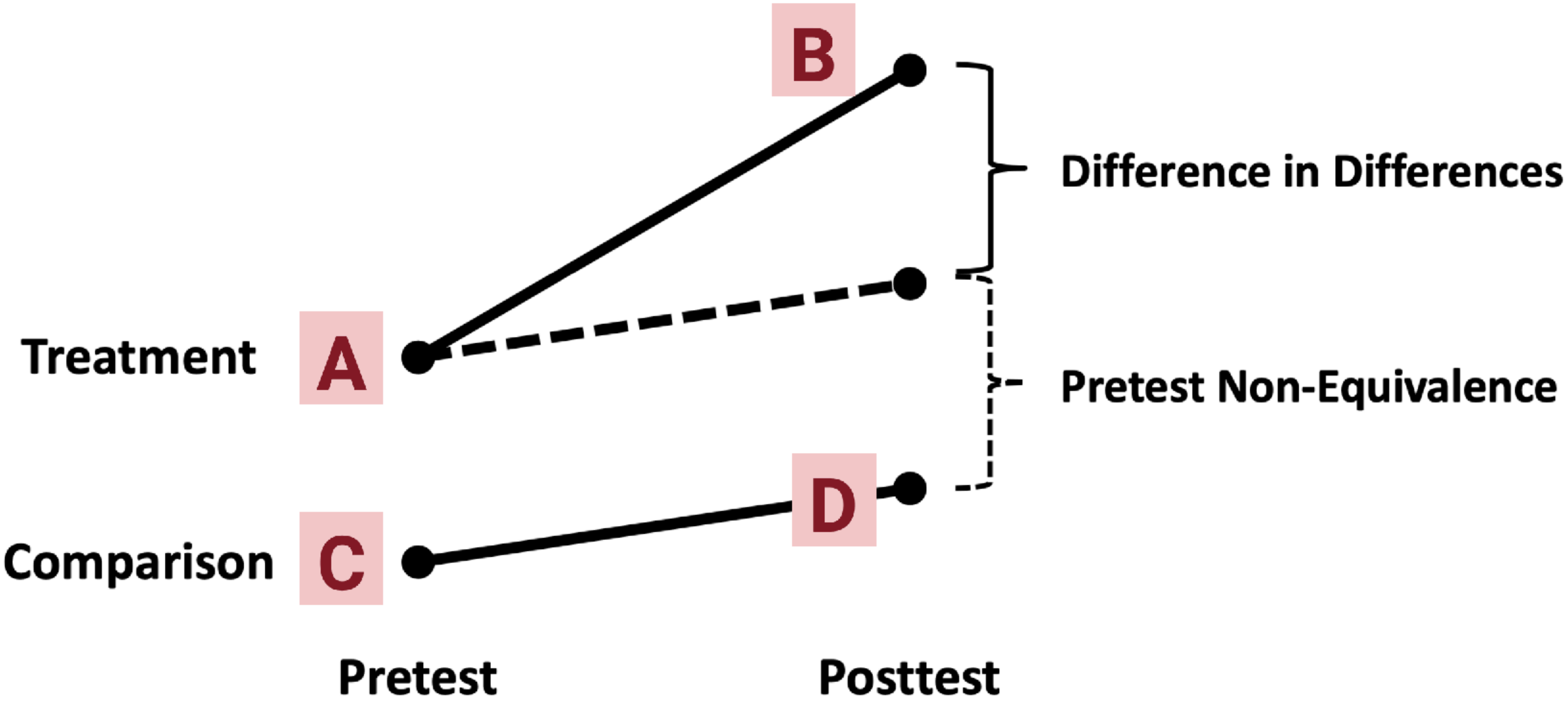
Growth

	Pre-mean	Post-mean
Treatment	A (not yet treated)	B (treated)
Control	C (never treated)	D (never treated)
Difference (T vs. C)	A - C	B - D

Within-group differences

	Pre-mean	Post-mean	Difference (post-pre)
Treatment	A (not yet treated)	B (treated)	B - A
Control	C (never treated)	D (never treated)	D - C
Difference (T vs. C)	A - C	B - D	(B - A) - (D - C)

Growth in treatment - growth in control = DiD!


B
Difference in Differences
Treatment
A
Pretest Non-Equivalence
D
Comparison
C
Pretest
Posttest

Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study

Katherine B Howe,[1,2] Christian Suharlim,[3] Peter Ueda,[4,5] Daniel Howe, Ichiro Kawachi,[2] Eric B Rimm[1,6,7]

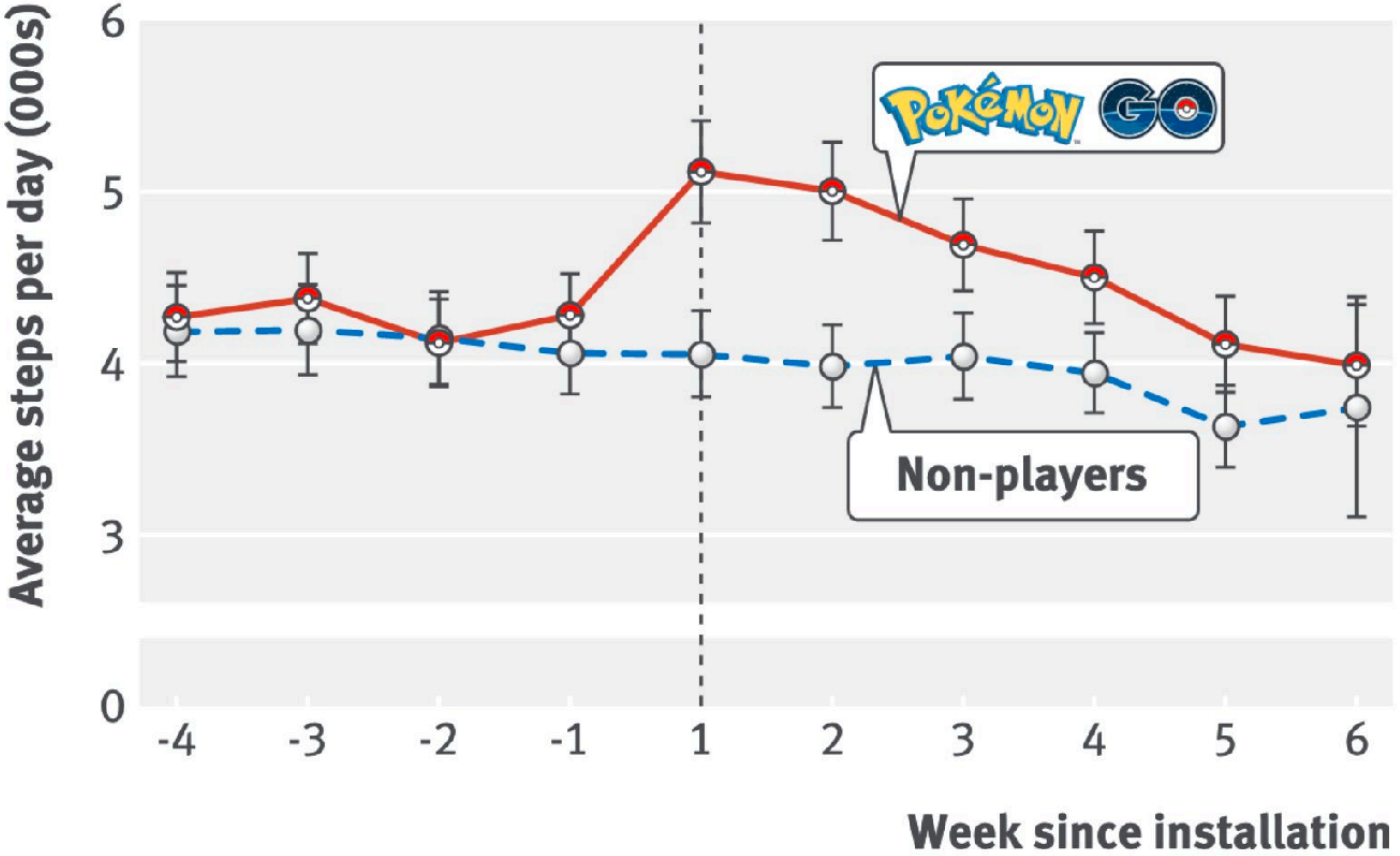


Fig 1 | Average number of daily steps and 95% confidence intervals by week before and after installation of Pokémon GO (median 8 July 2016)

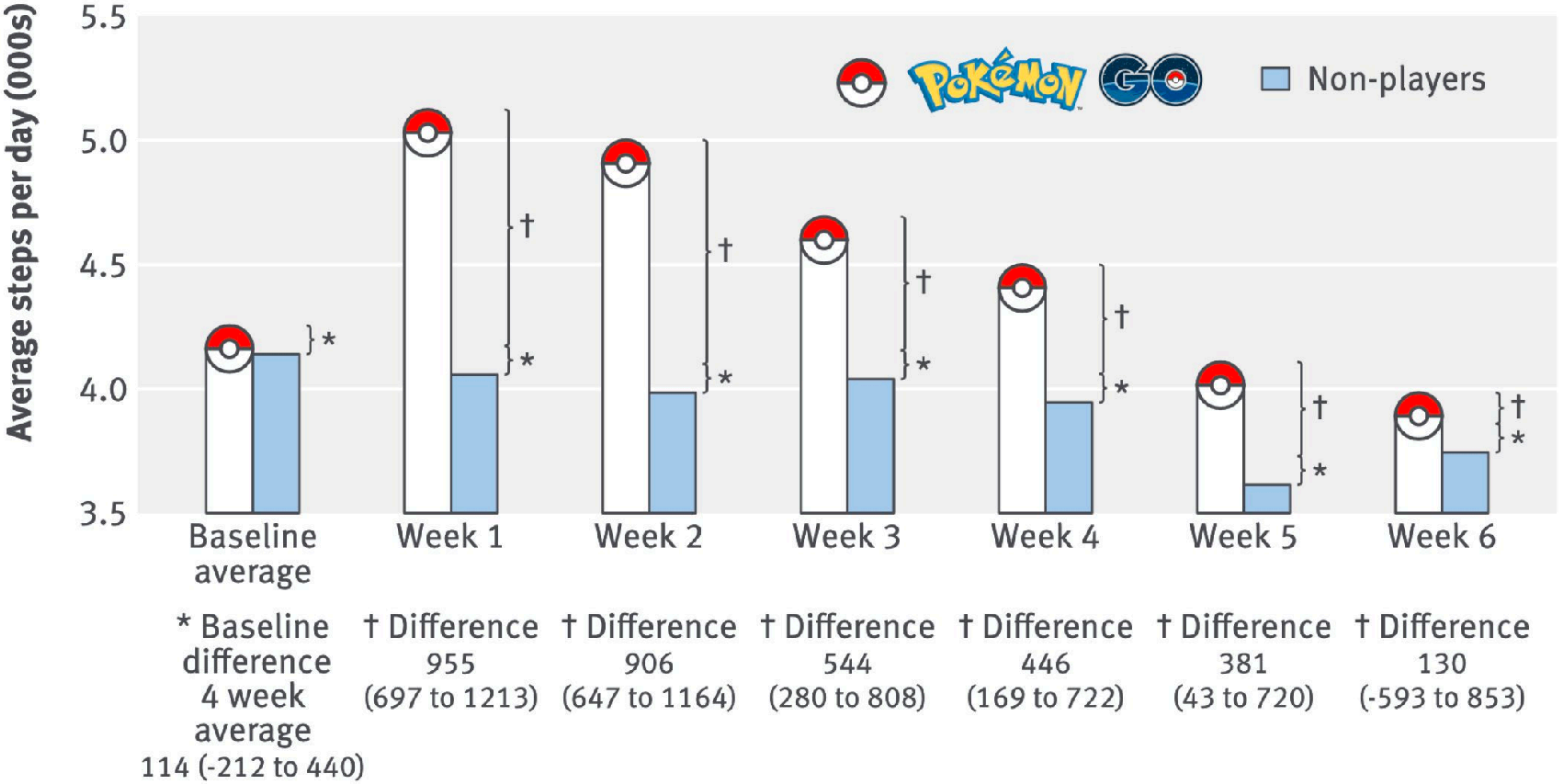


Fig 2 | Average daily number of steps before and after installation of Pokémon GO (median 8 July 2016). Confidence intervals are estimated using a difference in difference regression model (see supplementary table 1 for the full model)

How does this work?

Kids who never get app, before app comes out	Average exercise for untreated group
Kids who never get app, after app comes out	Average exercise for untreated group + time trend
Kids who get app, before app comes out	Average exercise for treated group
Kids who get app, after app comes out	Average exercise for treated group + time + treatment

Before-After Difference for Untreated

(Untreated Group + Time) -
(Untreated Group) = Time

Before-After Difference for Treated

(Treated Group + Time + Treatment) -
(Treated Group) = Time + Treatment

Difference-in-differences

Before-After Diff for Treated - B-A Diff for Untreated = (Time + Treatment) -
(Time) = Treatment

EXAMPLE

Does immigration reduce wages?

**More people looking for work —>
more competition —>
lower wages**

Classic econ paper by David Card looking at effect of Mariel boat lift on local economy of Miami

~125k(!) Cubans arrive in Miami between 15 April and 31 October 1980

MARIEL BOATLIFT

Why don't we just compare wages before and after the boat lift?

What to compare Miami to?

Atlanta, Houston, Los Angeles, and Tampa-St. Petersburg

DATA

state	hourslw	lwage	year	after	miami	class	person
93	40	2.37	82	TRUE	FALSE	1	Pariz
93	35	1.95	83	TRUE	FALSE	1	Vimala
93	40	1.11	81	TRUE	FALSE	1	Fardosa
93	40	1.16	84	TRUE	FALSE	1	Virgiline
58	40	1.53	82	TRUE	FALSE	1	Hopie
59	40	2.24	79	FALSE	TRUE	2	Bethann
59	48	1.75	81	TRUE	TRUE	1	Haliey
59	42	2.14	79	FALSE	TRUE	1	Olawale
59			83	TRUE	TRUE	1	Barnetta
59	45		80	FALSE	TRUE	1	Krysteen

After = after treatment (1981)

Miami = city is Miami

```
#Then we can do our difference in difference!
means <- df %>% group_by(after,miami) %>%
  summarize(lwage = mean(lwage, na.rm = TRUE),unemp=mean(unemp))
means %>%
  huxtable(add_colnames = TRUE)
```

Table 1

after	miami	lwage	unemp
FALSE	FALSE	1.88	0.0619
TRUE	FALSE	1.84	0.0794
FALSE	TRUE	1.74	0.0547
TRUE	TRUE	1.72	0.0854

How did wages change in Miami before and after the boatlift?

after	miami	lwage	unemp
FALSE	FALSE	1.88	0.0619
TRUE	FALSE	1.84	0.0794
FALSE	TRUE	1.74	0.0547
TRUE	TRUE	1.72	0.0854

1.72 - 1.74 = -.02

Does this mean Boatlift —> dropped wages?

THE CONTROL CITIES

How did wages in the control cities?

after	miami	lwage	unemp
FALSE	FALSE	1.88	0.0619
TRUE	FALSE	1.84	0.0794
FALSE	TRUE	1.74	0.0547
TRUE	TRUE	1.72	0.0854

1.84 - 1.88 = -.04

Wages were dropping everywhere!

What’s the diff-in-diff here?

after	miami	lwage	unemp
FALSE	FALSE	1.88	0.0619
TRUE	FALSE	1.84	0.0794
FALSE	TRUE	1.74	0.0547
TRUE	TRUE	1.72	0.0854

-.02 - -.-04 = .02

Wages actually improved slightly

What’s happening here?

We want to know *what would have happened to Miami had Mariel not taken place*

We could compare Miami against itself (pre-post), but change might just be general trend!

So instead we compare the *change* in Miami against the *change* in other cities

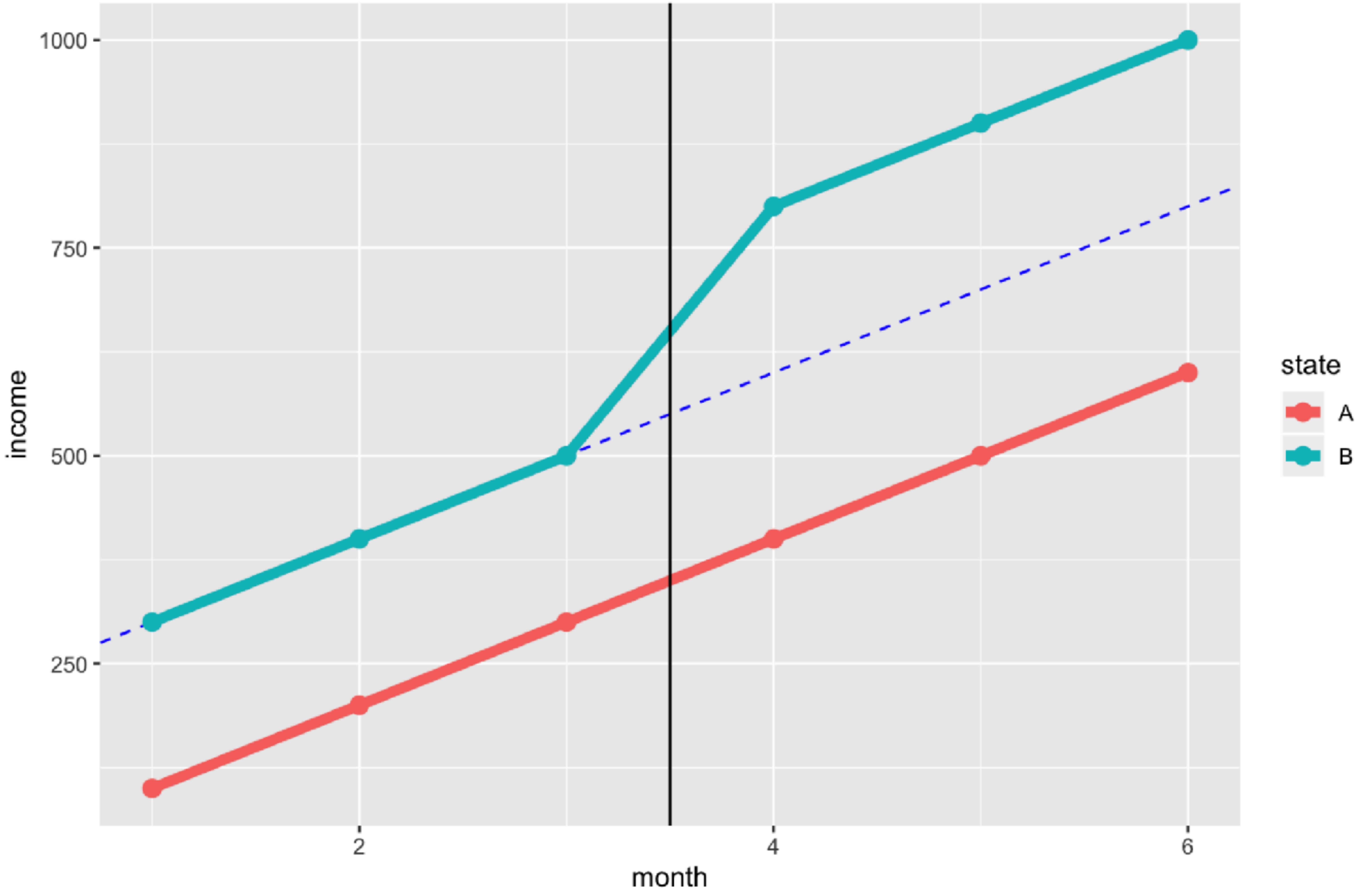
Implicitly: Miami’s wages would look like (other cities) had Mariel not happened

ASSUMPTIONS

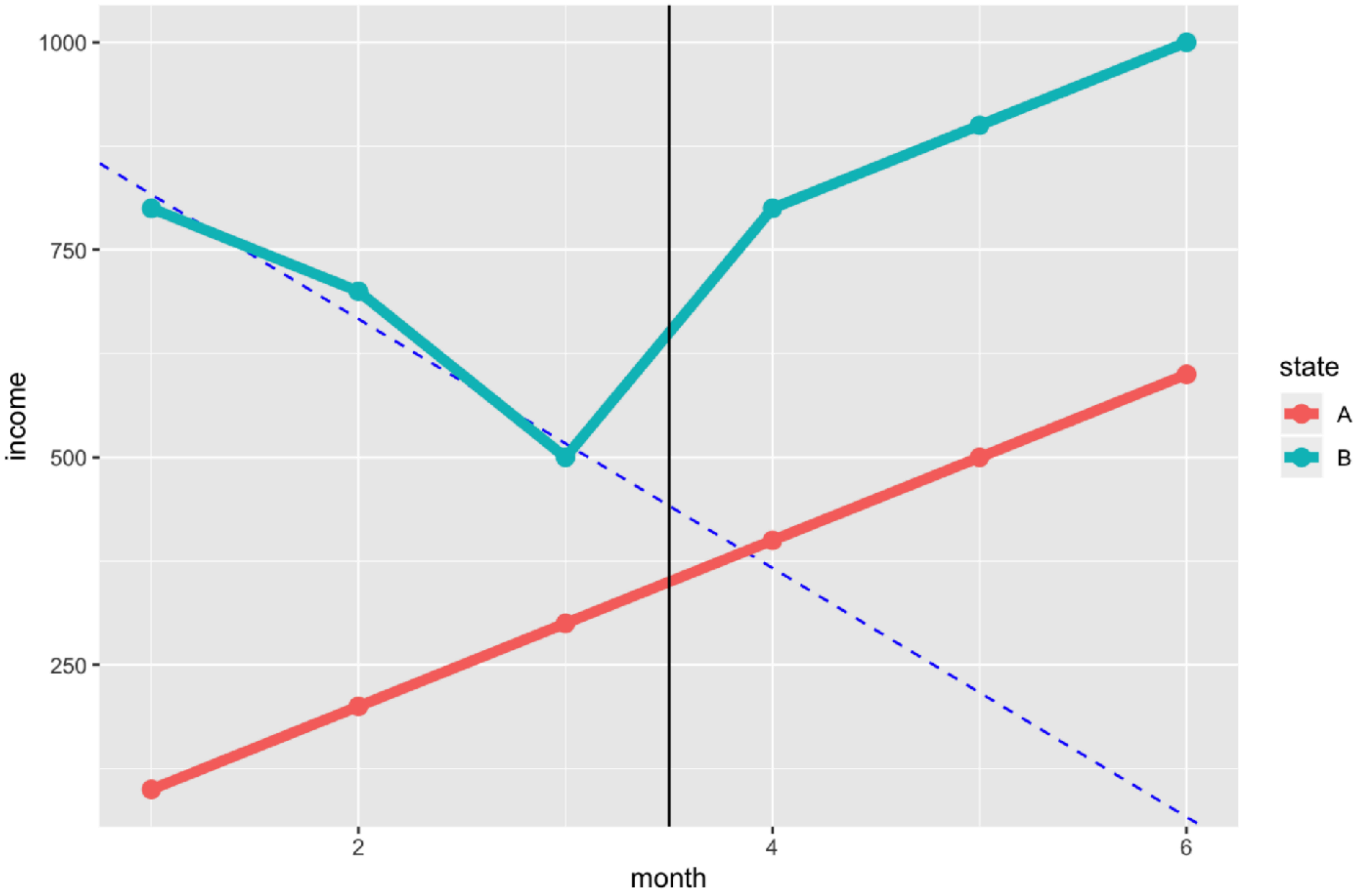
Parallel trends

Had treatment not happened (e.g., had Mariel not happened), treatment (Miami) and control group (other cities) would look the same

One way to check this is to look at years up to treatment



1000
750
500
250
income
2
4
6
month
state
A
B



1000
750
500
250
income
2
4
6
month
state
A
B

ASSUMPTIONS

What do we compare Miami to?

Are LA, Houston, Atlanta, and Tampa basically the same as Miami?

More complicated diff-in-diff approaches use matching to select groups that are trending similarly before the treatment kicks in

Next week

How do this via regression

Interactions in regression

Regression discontinuity