Matching

PSCI 2301: Quantitative Political Science II

Prof. Brenton Kenkel

brenton.kenkel@gmail.com

Vanderbilt University

February 11, 2025

Today's agenda

1. Motivating question: Does holding office increase wealth?
2. Covariates and conditional independence
3. Estimating treatment effects by matching

Motivating question: Political office and personal wealth

Does holding political office increase wealth?

Population: British candidates for Parliament elected 1950–1970

→ So this doesn't answer whether election would make a regular Briton richer

Outcome: Total wealth at death

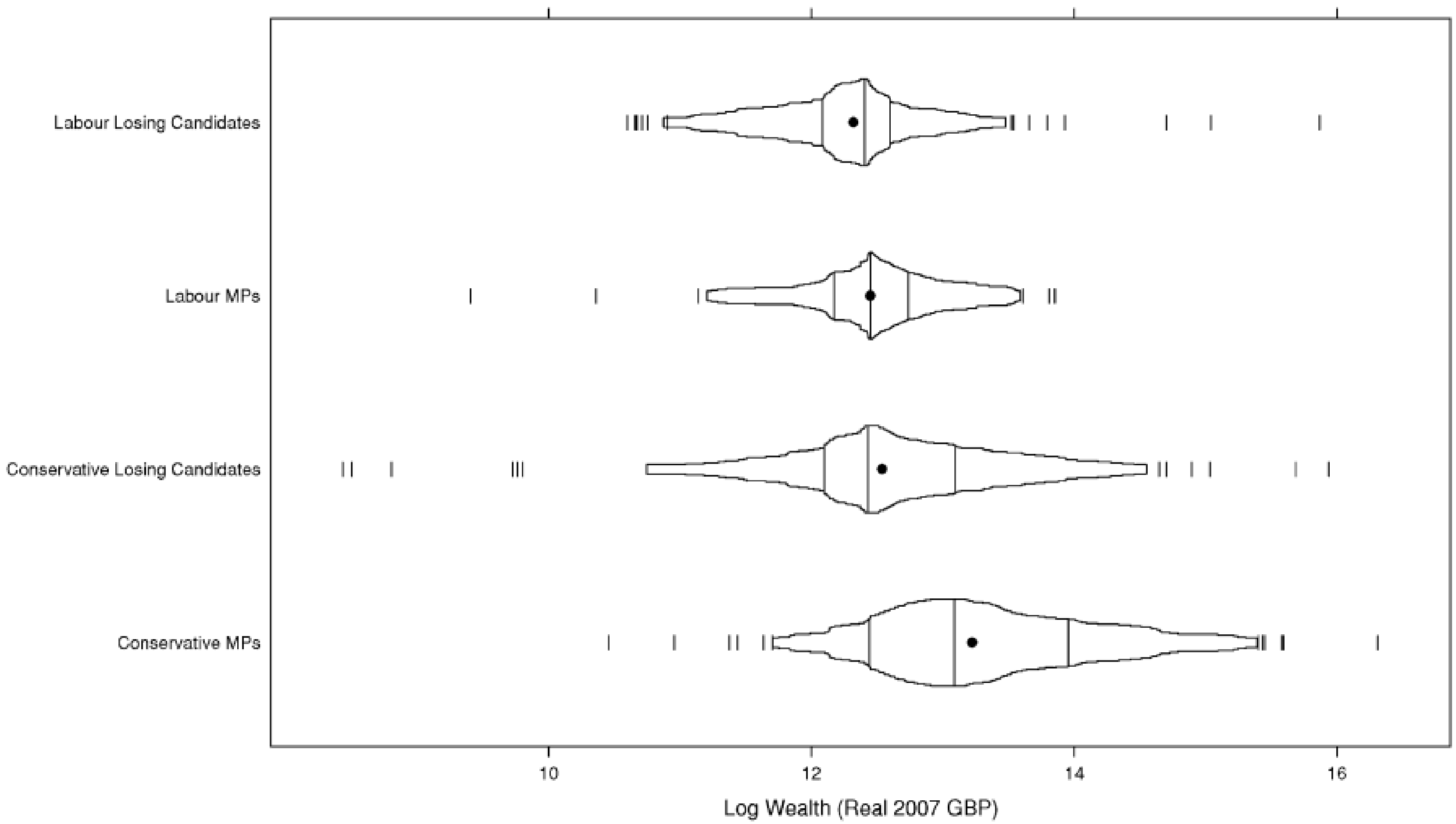
Treatment: Being elected to Parliament

Comparison: Not being elected to Parliament

Correlational evidence suggests positive effect, esp. for Conservatives

Correlational evidence

FIGURE 2. Distributions of (Log) Wealth at Death by Party for Winning and Losing Candidates to House of Commons 1950–1970

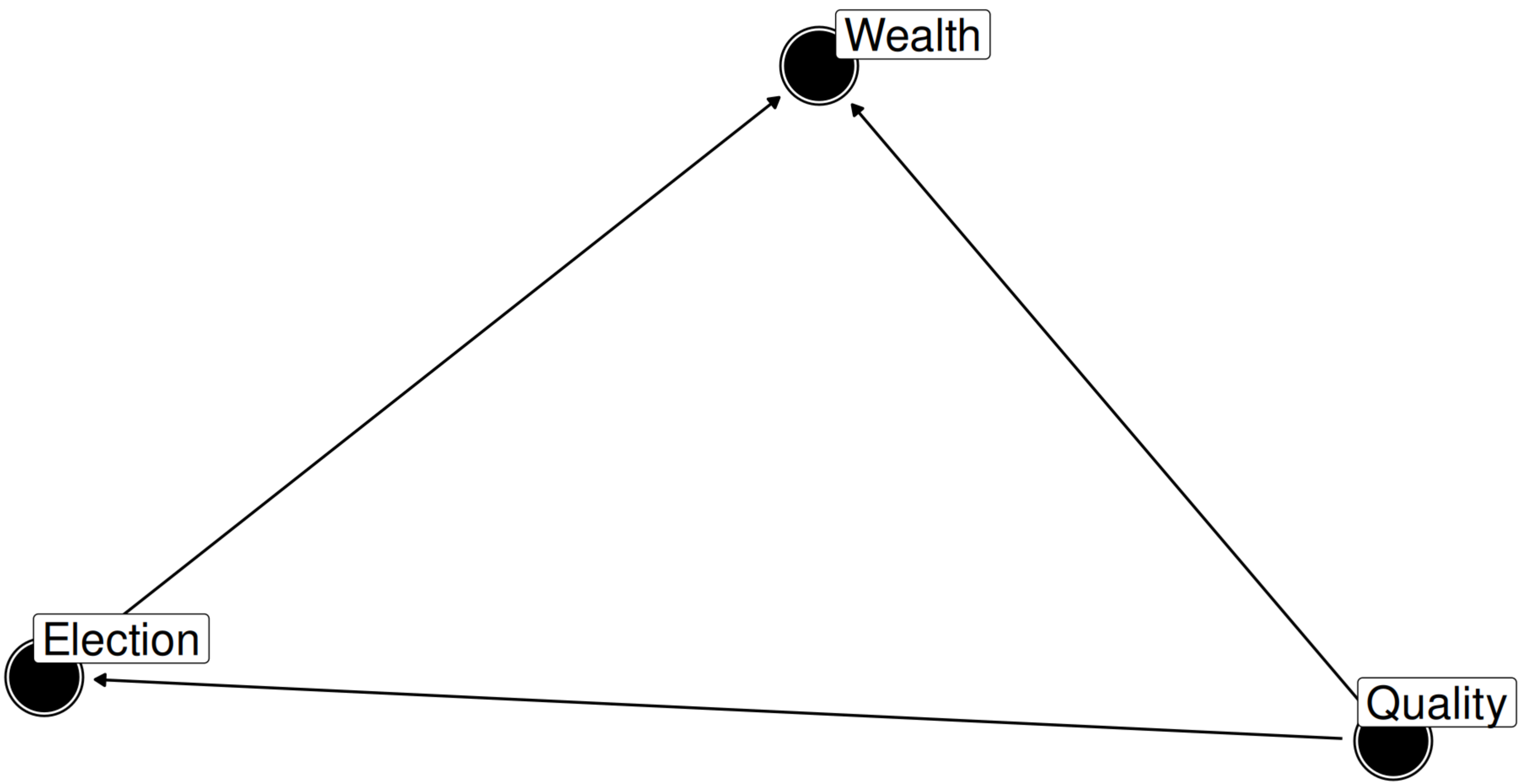


Note: Box percentile plots. Box shows empirical distribution function from .05 to .95 quantile; vertical lines indicate the .25, .5, and .75 quantile, respectively. Observations outside the .05–.95 quantile range are marked by vertical whiskers. The dot indicates the mean.

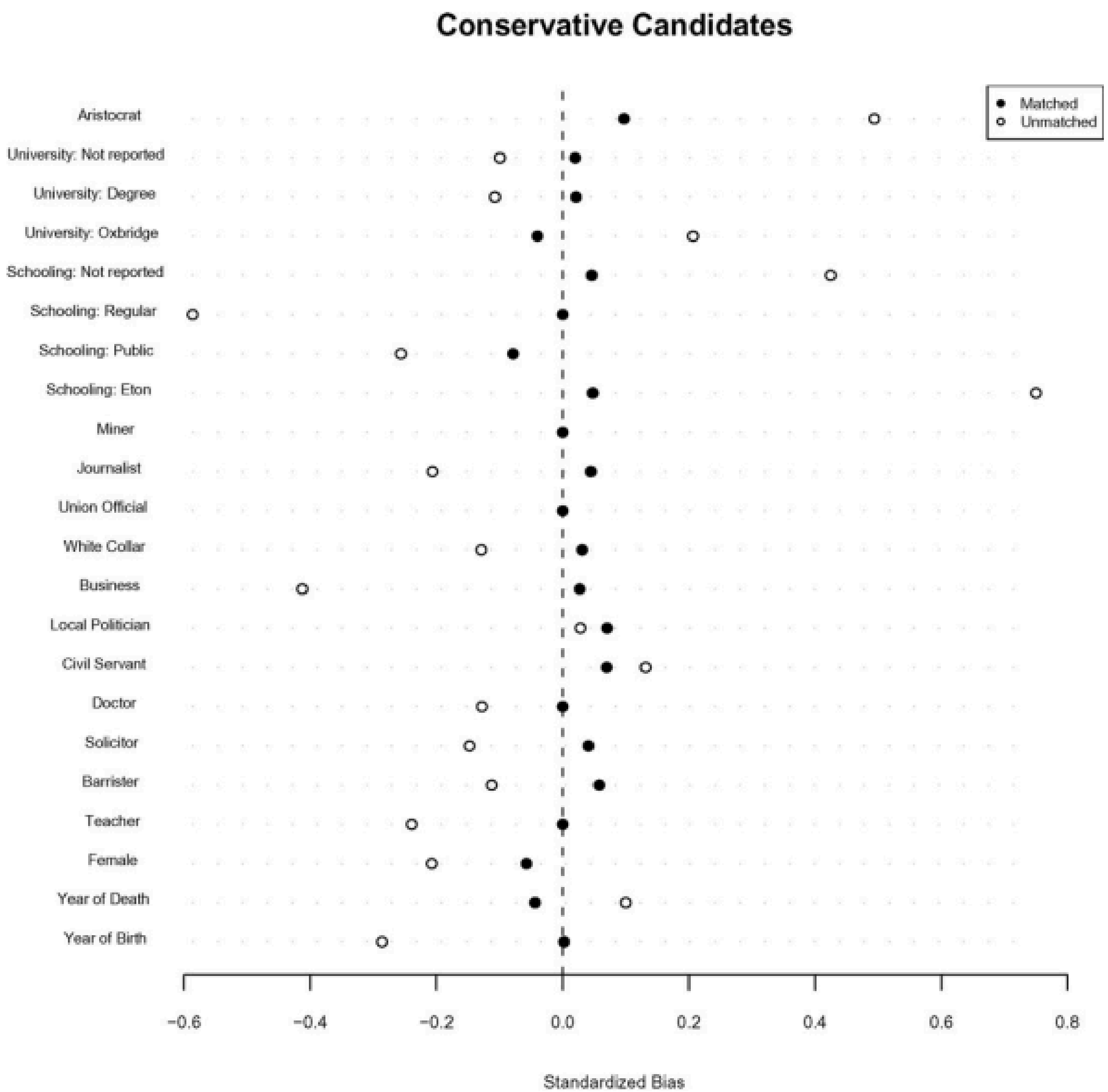
Problems for causal inference

Election is not a randomly assigned treatment

Those elected probably aren't representative of all candidates



Lots of confounding variables



Covariates and conditional independence

A confounding variable

Imagine every MP fell into one of two categories…

"Dashing" (Tony Benn)

"Plain" (George Brown)

Dashing and plain MPs

Confounding variable affects **assignment to treatment**

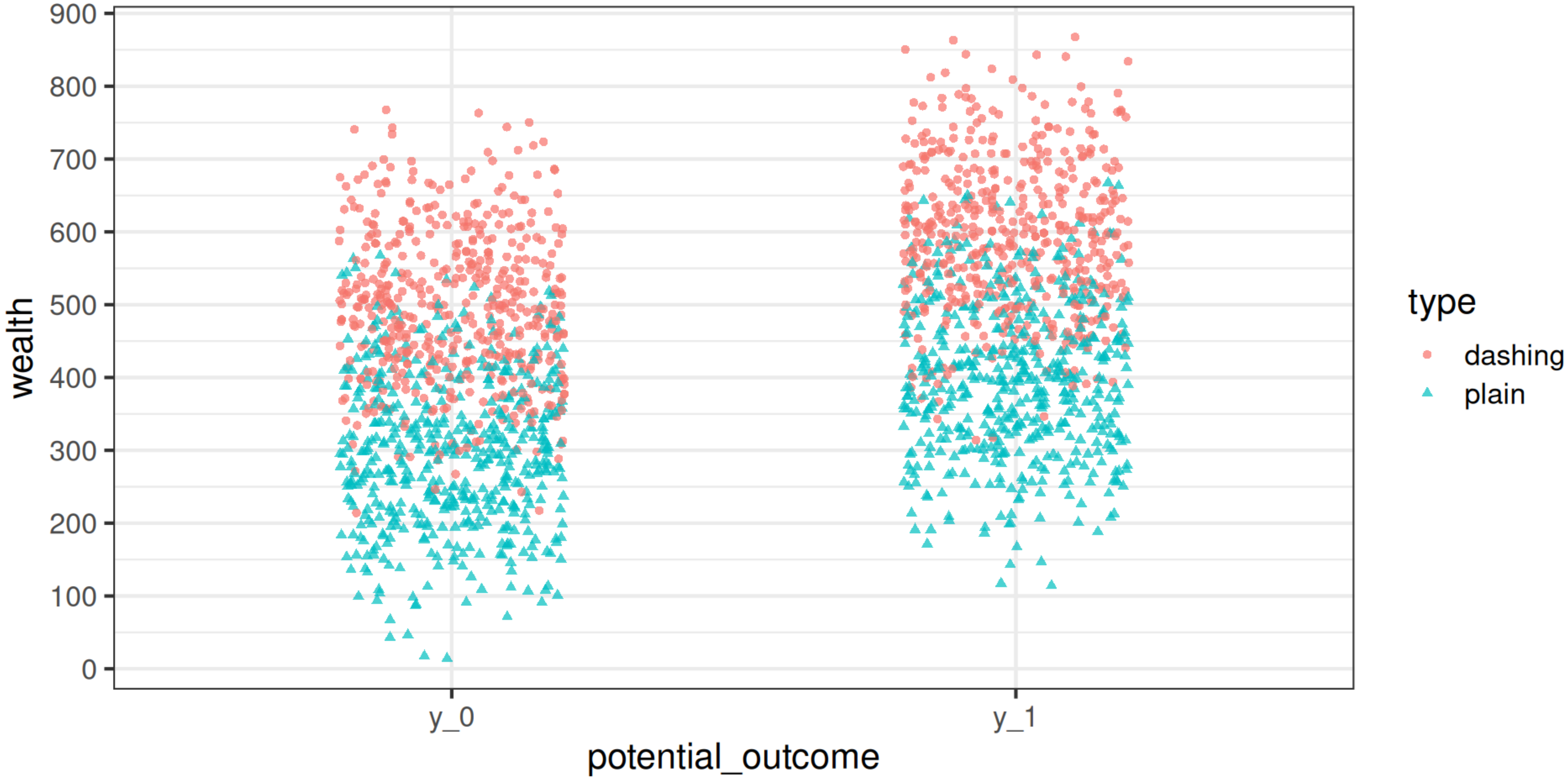
- Assume candidate population is split evenly between dashing and plain
- …but 80% of dashing ones are elected, versus 20% of plains

Confounding variable also affects **potential outcomes**

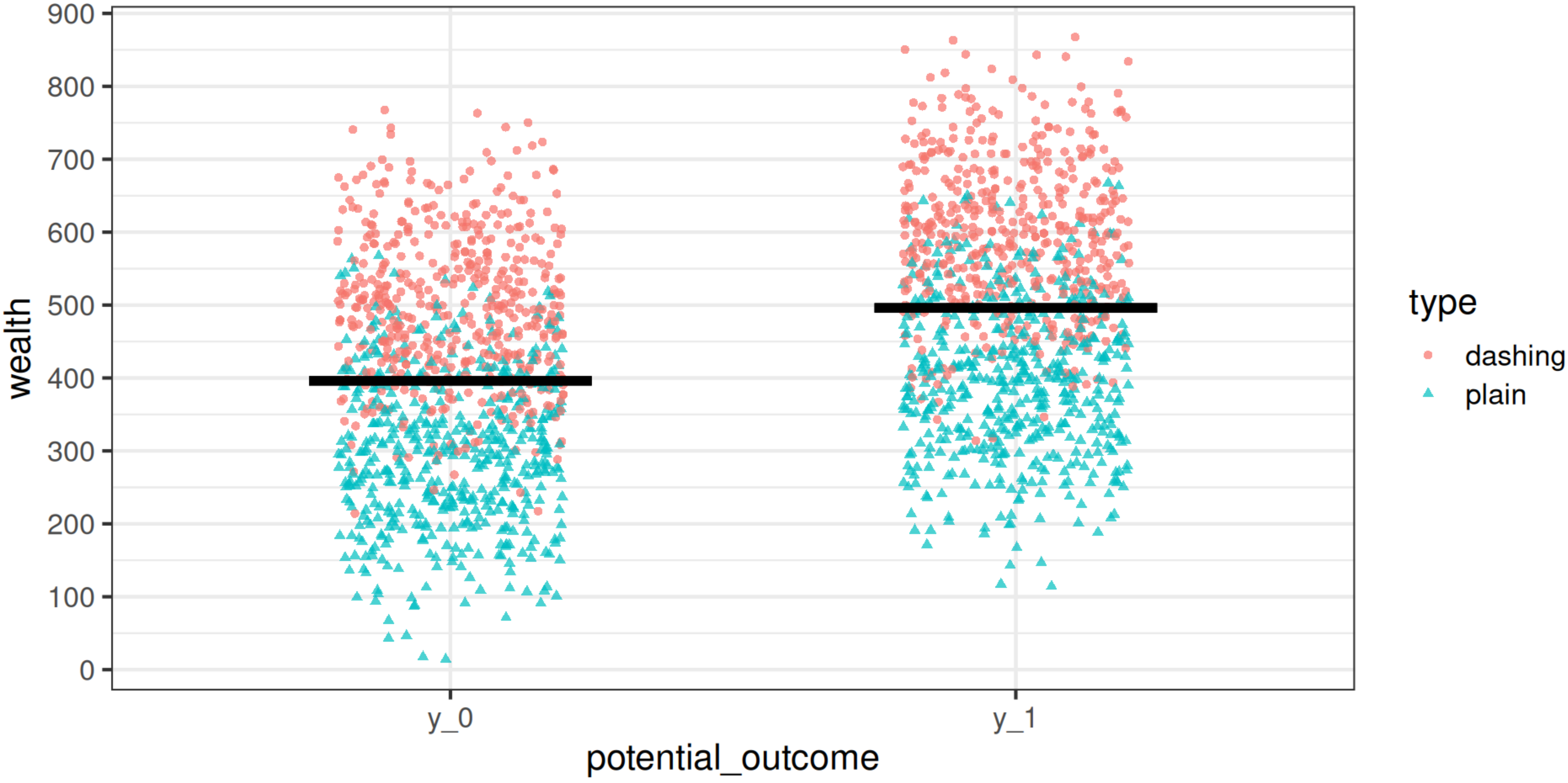
- Assume election raises wealth by £100k for all candidates
- …but baseline wealth of dashing candidates is £200k higher

If we just compare elected to not-elected, will we get the right answer?

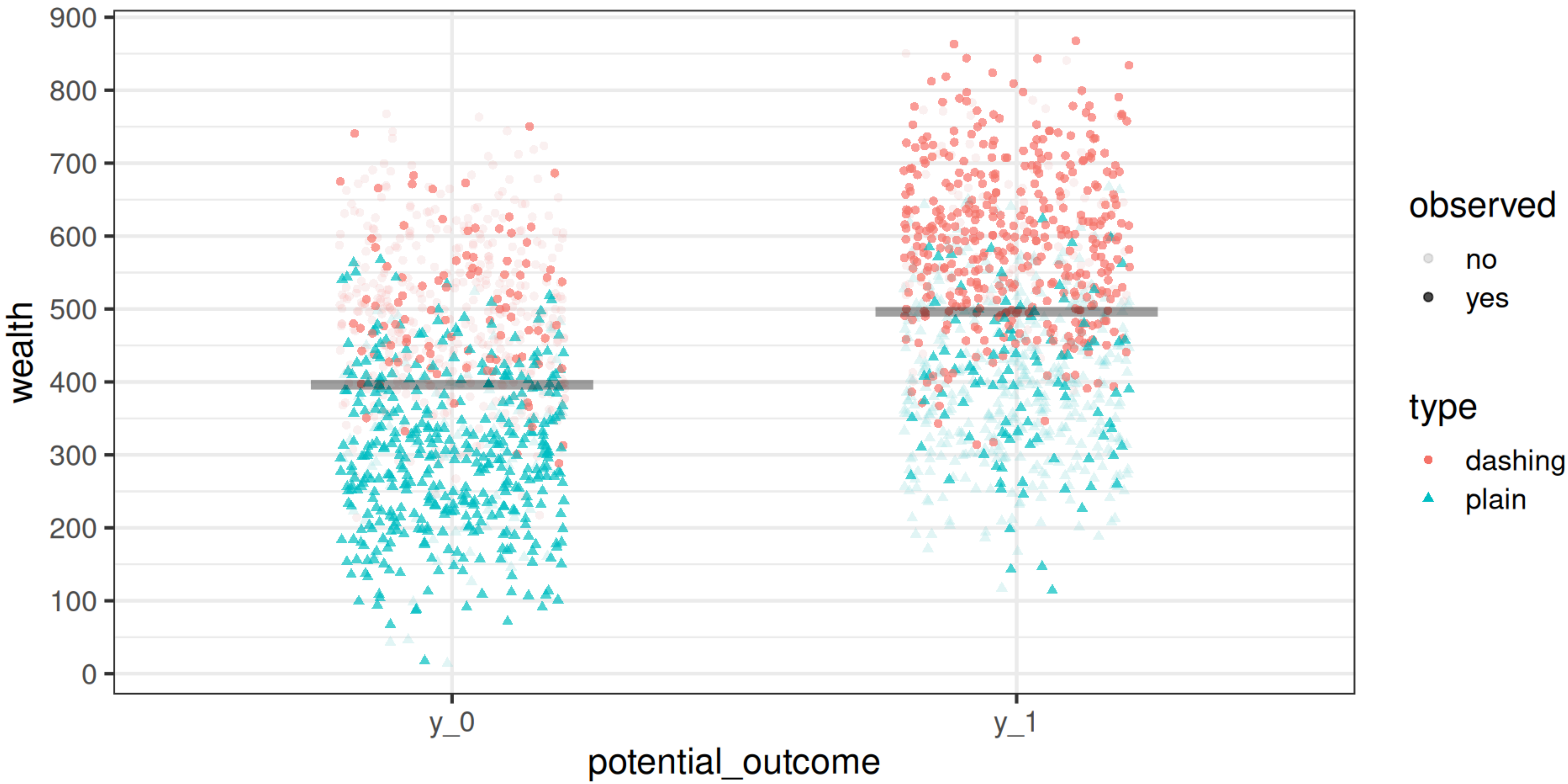
Hypothetical MP data



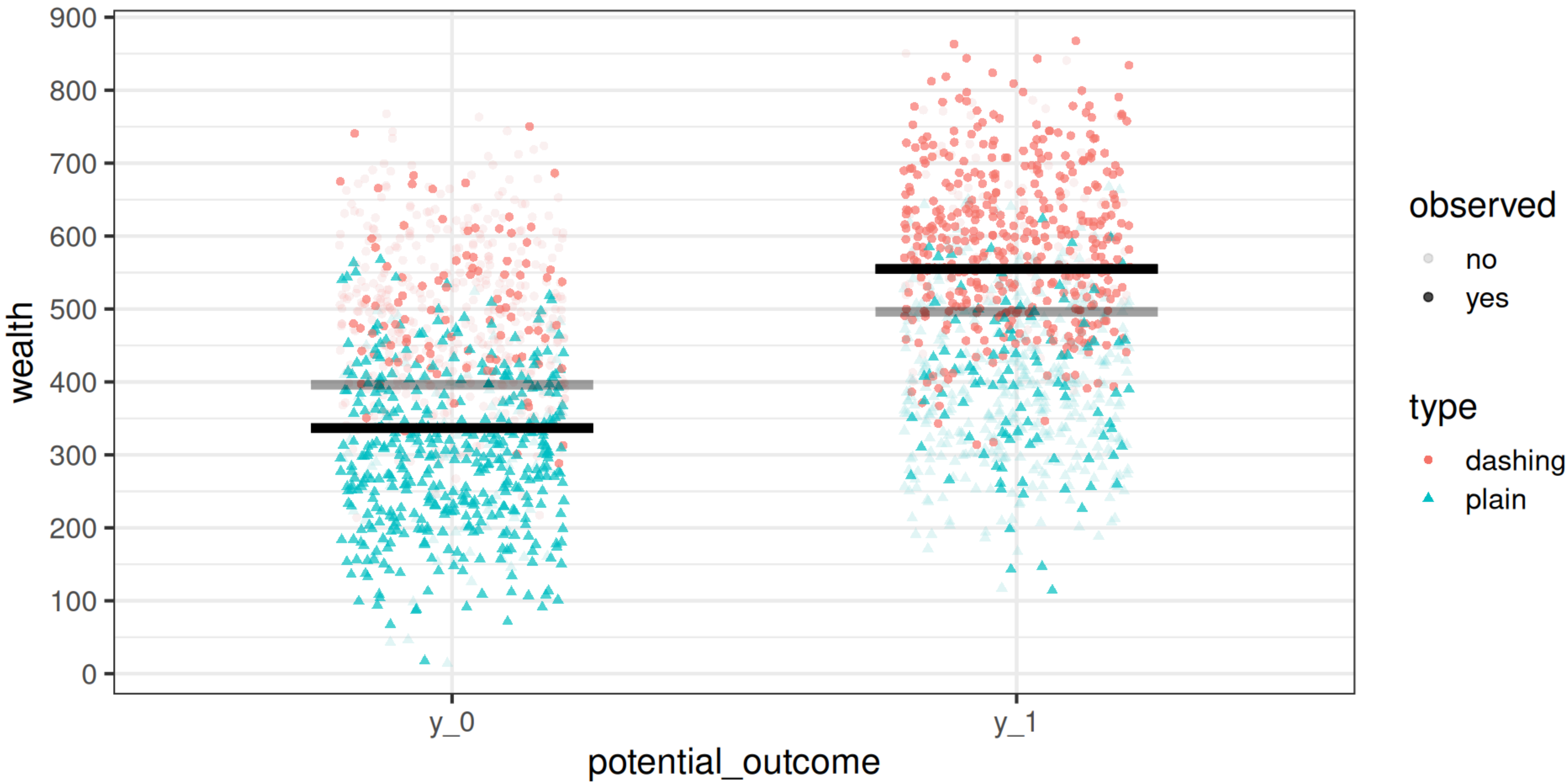
Hypothetical MP data



Hypothetical MP data



Hypothetical MP data



Confounding and independence failure

Independence condition: $\mathbb{C}[Y_{1i}, D_i] = 0$, $\mathbb{C}[Y_{0i}, D_i] = 0$

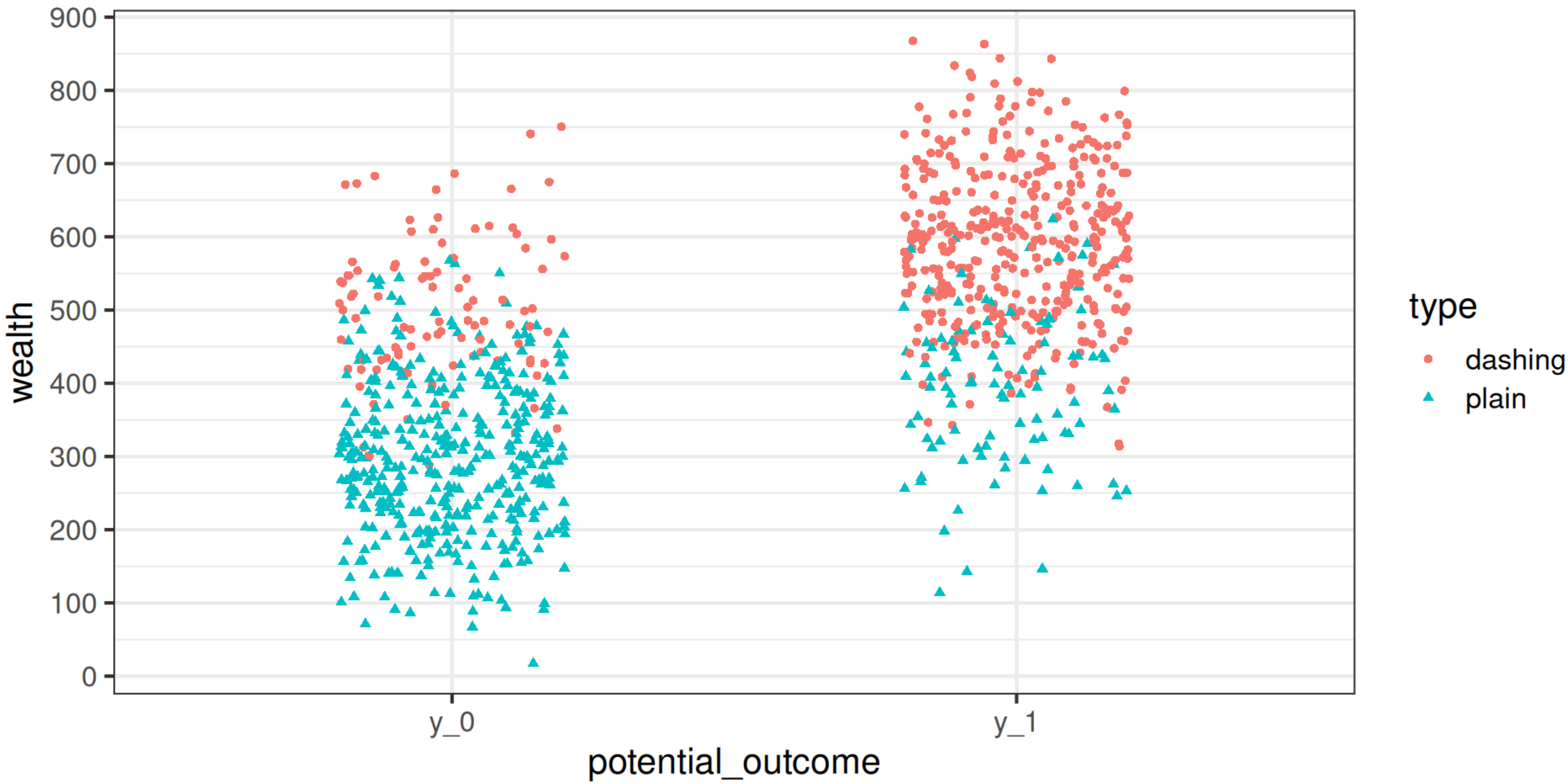
If independence holds:

$$\mathbb{E}[Y_{1i} \mid D_i = 1] = \mathbb{E}[Y_{1i}]$$
$$\mathbb{E}[Y_{0i} \mid D_i = 0] = \mathbb{E}[Y_{0i}]$$

But because dashingness $\rightsquigarrow$ election chances, we don't have that here

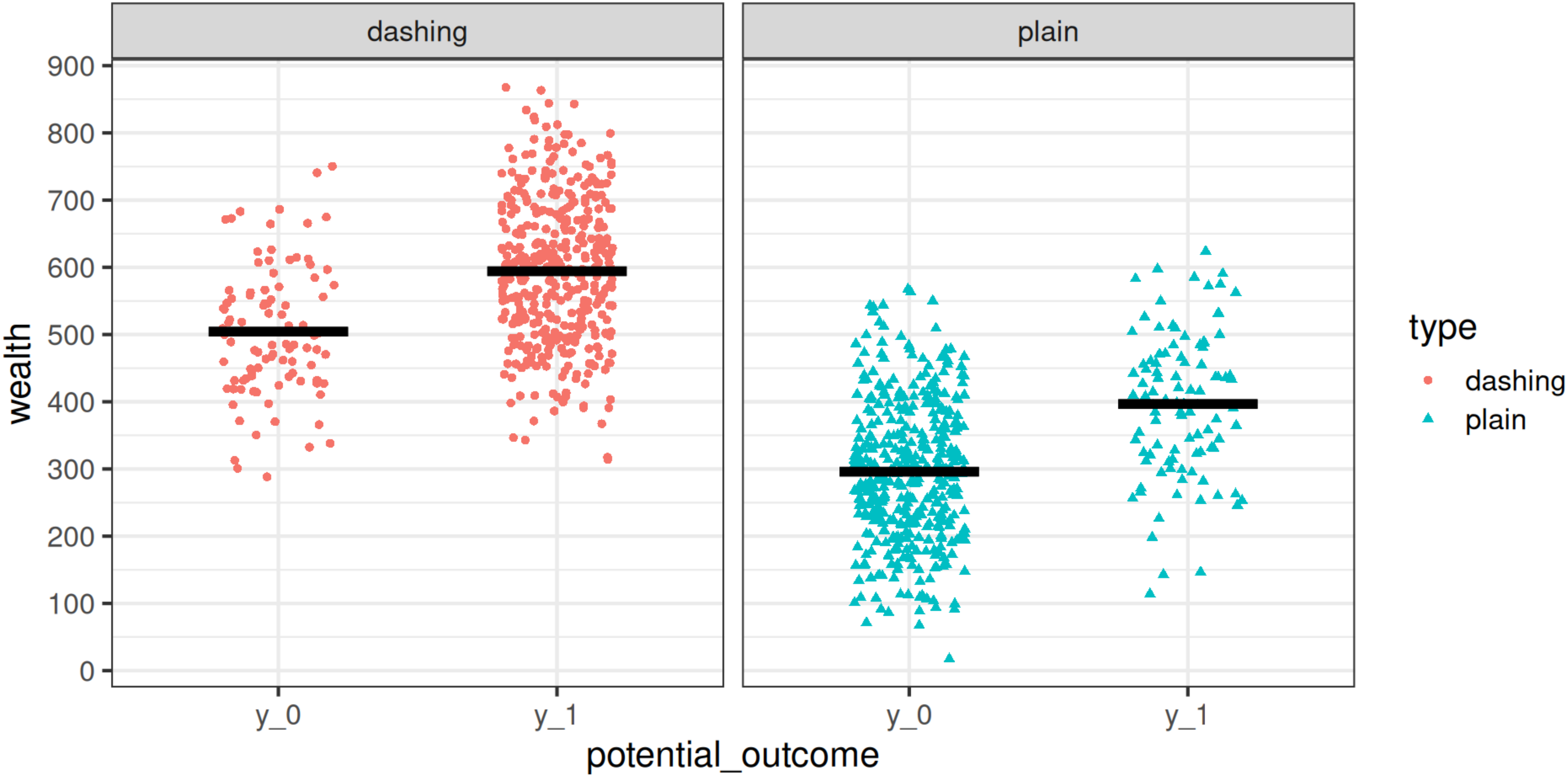
$$\mathbb{E}[Y_{1i}] = \tfrac{1}{2}\,\mathbb{E}[Y_{1i} \mid \textbf{dashing}] + \tfrac{1}{2}\,\mathbb{E}[Y_{1i} \mid \textbf{plain}]$$
$$\mathbb{E}[Y_{1i} \mid D_i = 1] = \tfrac{3}{4}\,\mathbb{E}[Y_{1i} \mid \textbf{dashing}] + \tfrac{1}{4}\,\mathbb{E}[Y_{1i} \mid \textbf{plain}]$$

Analyzing within subgroups



Analyzing within subgroups

Analyzing within subgroups



Conditional independence

The conditional independence condition

Let X_i be an observable variable (or collection of variables). Treatment assignment and potential outcomes are conditionally independent given X_i if

$$\mathbb{C}[Y_{1i}, D_i \mid X_i = x] = 0,$$
$$\mathbb{C}[Y_{0i}, D_i \mid X_i = x] = 0$$

for every possible variable value x.

Interpretation:

- Treatment assignment at random within subgroups defined by X_i values
- OK if X_i affects treatment assignment
- OK if other vars affect potential outcomes, but not treatment assignment
- Not OK if there are unobserved confounders

Estimation under conditional independence

Law of iterated expectation

Suppose X_i has M possible values $x_1, \ldots, x_M$ The population mean of any variable Z_i is a weighted average of means within subgroups defined by values of X_i:

$$\mathbb{E}[Z_i] = \sum_{m=1}^{M} \underbrace{\mathbb{E}[Z_i \mid X_i = x_m]}_{\text{subgroup avg}} \cdot \underbrace{\Pr(X_i = x_m)}_{\text{subgroup size}}$$

Suggests how to estimate avg effects if conditional independence holds:

- Calculate $\mathbf{avg}[Y_i \mid D_i = 1, X_i = x] - \mathbf{avg}[Y_i \mid D_i = 0, X_i = x]$ within each subgroup
- Take weighted average, weighting by proportion of obs in each group

Estimation under conditional independence

```r
df_sim_agg <- df_sim |>
  group_by(type) |>
  summarize(
    n = n(),
    avg_treat = mean(wealth[elected == "yes"]),
    avg_control = mean(wealth[elected == "no"]),
    diff = avg_treat - avg_control
  )
df_sim_agg
```

```
# A tibble: 2 × 5
  type          n avg_treat avg_control  diff
  <chr>     <int>     <dbl>       <dbl> <dbl>
1 dashing     500      594.        504.  89.8
2 plain       500      397.        296. 101.
```

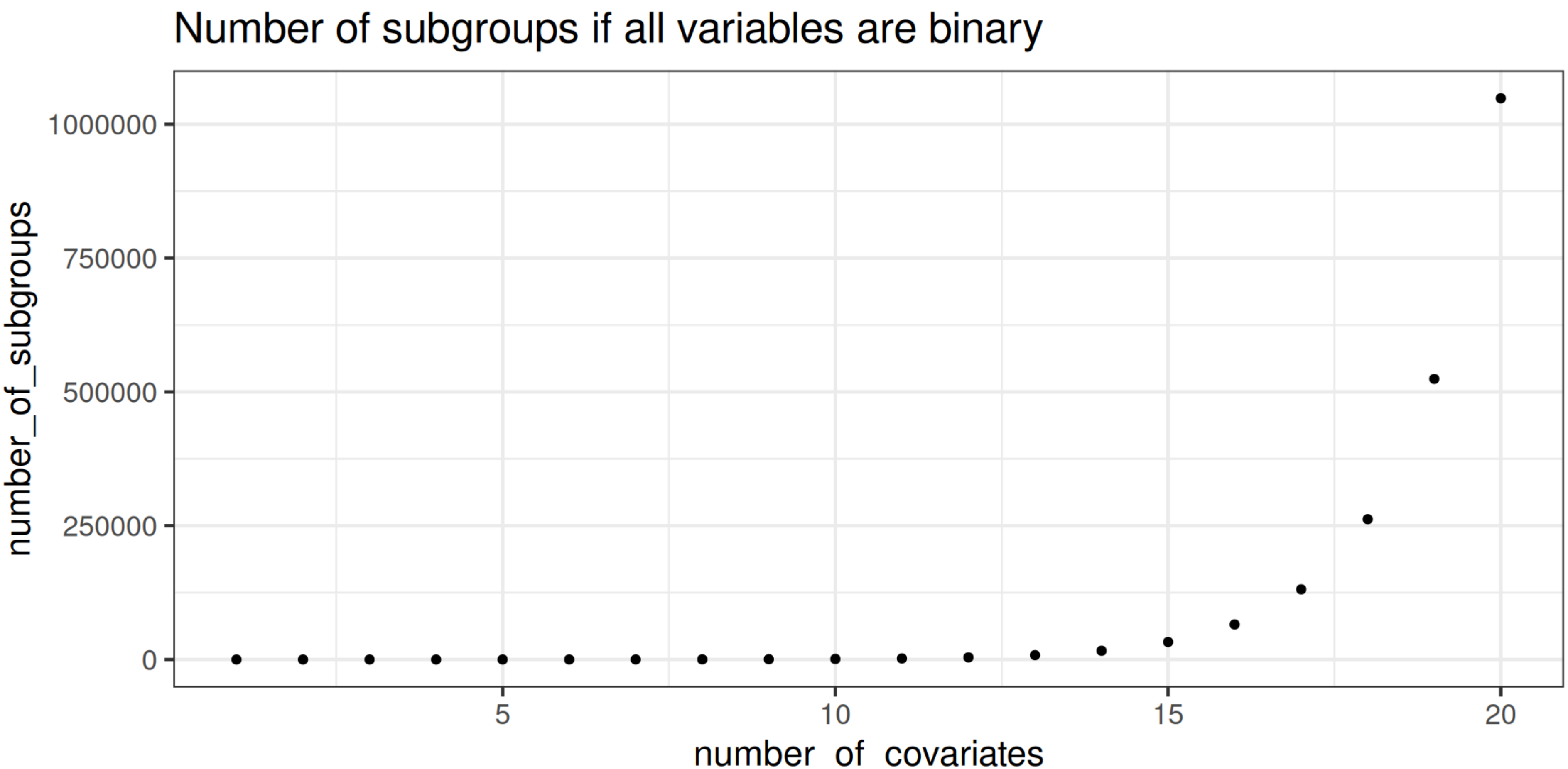
```r
df_sim_agg |>
  summarize(ate = weighted.mean(diff, n))
```

```
# A tibble: 1 × 1
    ate
  <dbl>
1  95.2
```

The curse of dimensionality

In reality, being dashing versus plain isn't only confounding variable

Curse of dimensionality: Number of subgroups grows exponentially



Solving the curse of dimensionality

Many confounders $\rightsquigarrow$ can't take averages within subgroups

- Very small samples within each group $\rightsquigarrow$ high standard errors
- Many subgroups won't have both control and treatment

Matching estimators solve this by finding closest matches

Loose idea — Pair each treated obs with the control obs with closest X_i

- Idea is to make treatment and control groups as balanced as possible
- Measures of "closest" differ
- Some control obs will not end up being used for estimation!

Wrapping up

What we did today

1. Looked at the question of how officeholding affects wealth
2. Few confounders $\rightsquigarrow$ Weighted average across subgroups
3. Many confounders $\rightsquigarrow$ Matching

Next time:

- More details on matching
- Work through Eggers & Hainmueller ourselves