STA 235H - Randomized Controlled Trials II

Fall 2022

McCombs School of Business, UT Austin

Some announcements

- Homework 3 is due this Friday.
- Answer Key for Homework 2 will be posted on Wednesday.
- Remember Idea + Data Pitch is due next week: Don't leave it to the last minute.
  - Idea needs to be approved.

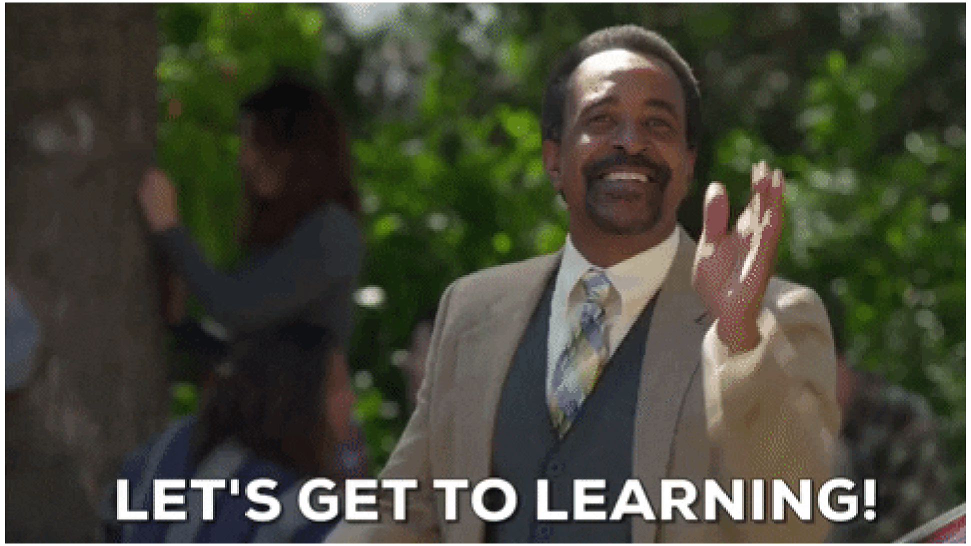
Catch-up Session this Friday

Last week

- Continued discussing the potential outcomes framework.
- Introduced the ignorability assumption:
    - Potential outcomes $Y(0)$ and $Y(1)$ are independent of the treatment Z
- Introduced Randomized Controlled Trial

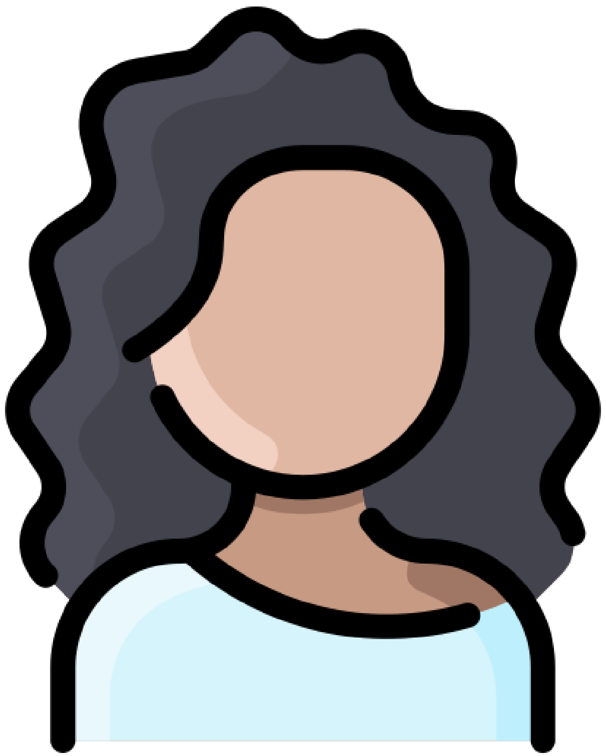
Today

- Continue with Randomized Controlled Trials:
  - Design
  - Limitations
- What about selection on observables?:
  - Types of bias
  - Matching



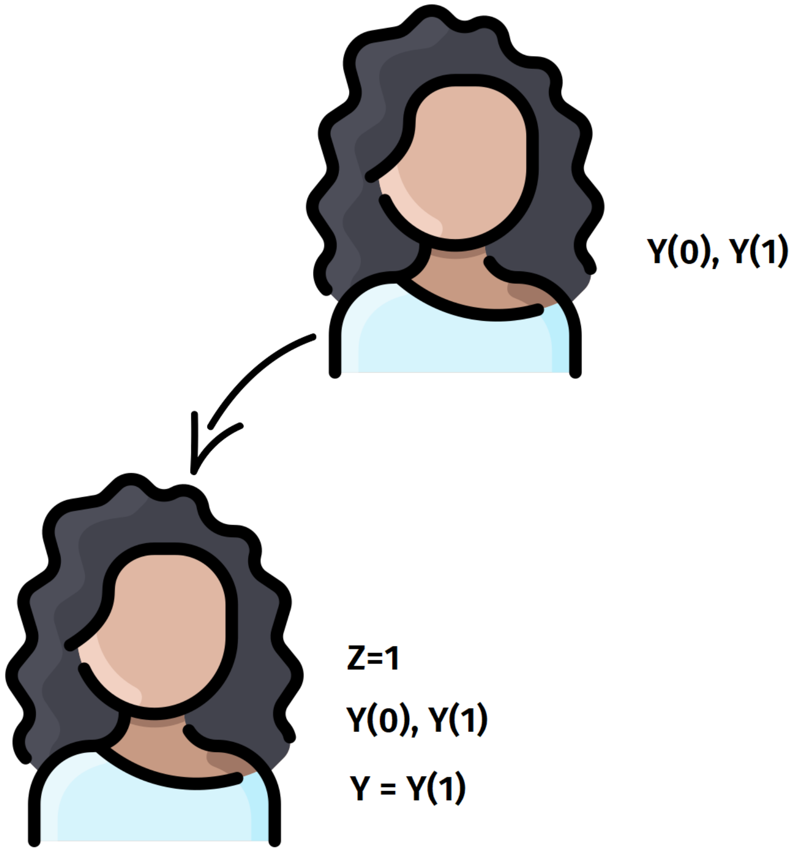
Reviewing the Ignorability Assumption

Imagine we have an individual...

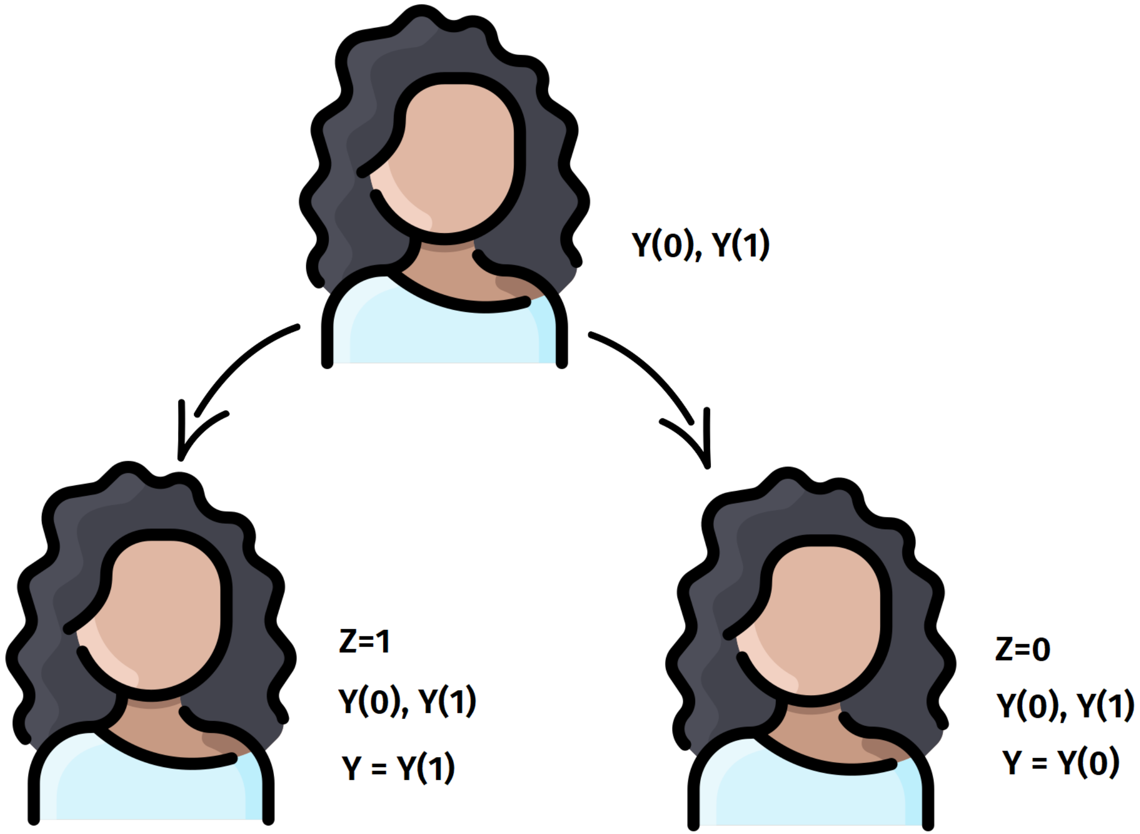


$Y(0)$, $Y(1)$

That person could choose to take the treatment...



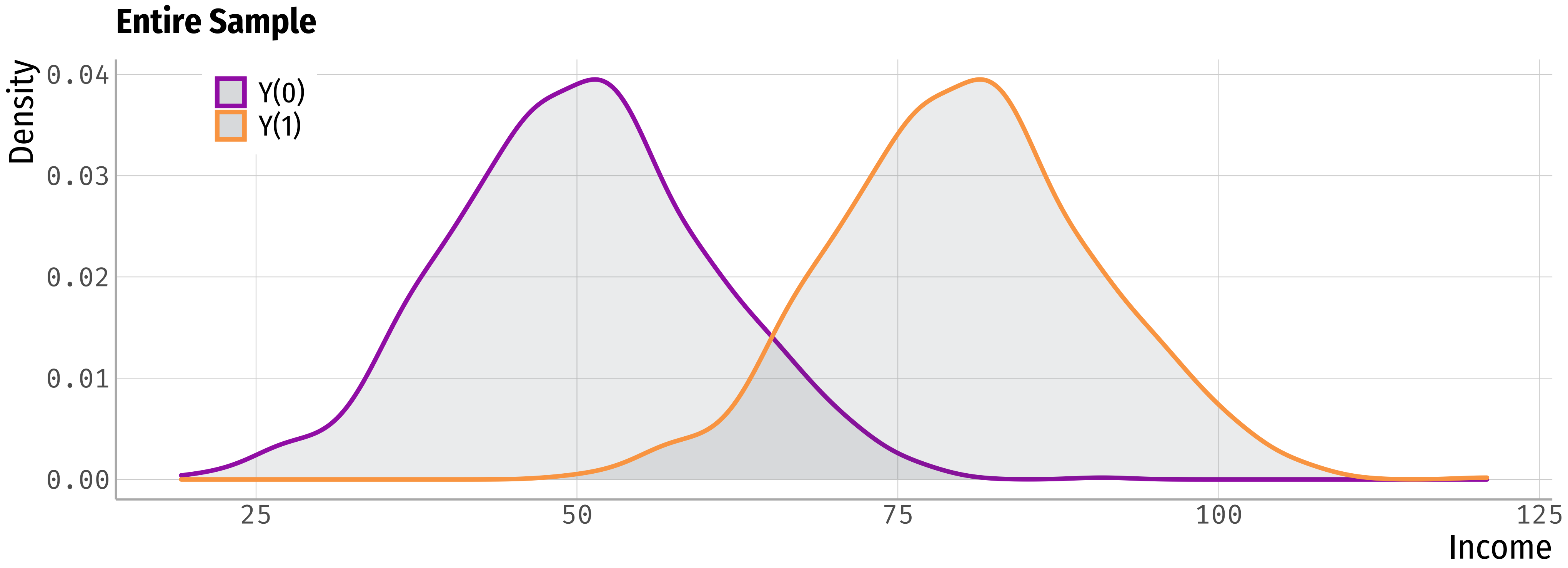
... or not take the treatment



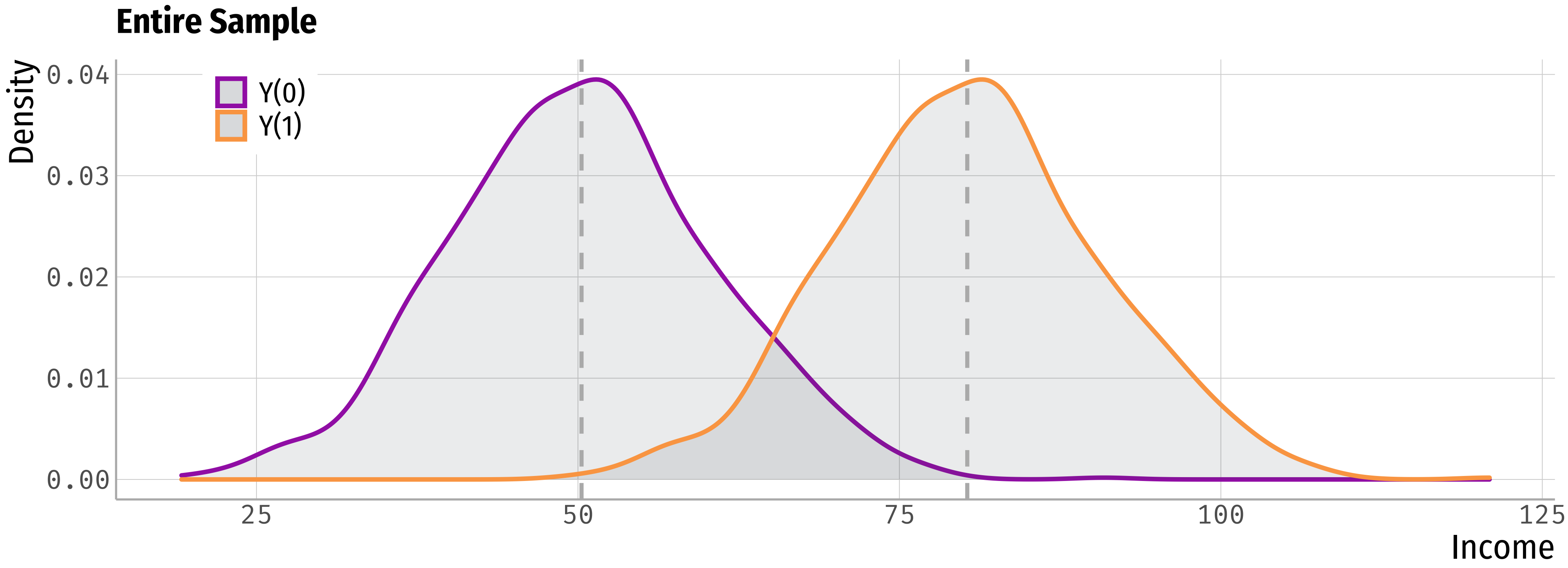
- Note that in each scenario, she still has **both potential outcomes**!

What is the effect of going to college on future income?

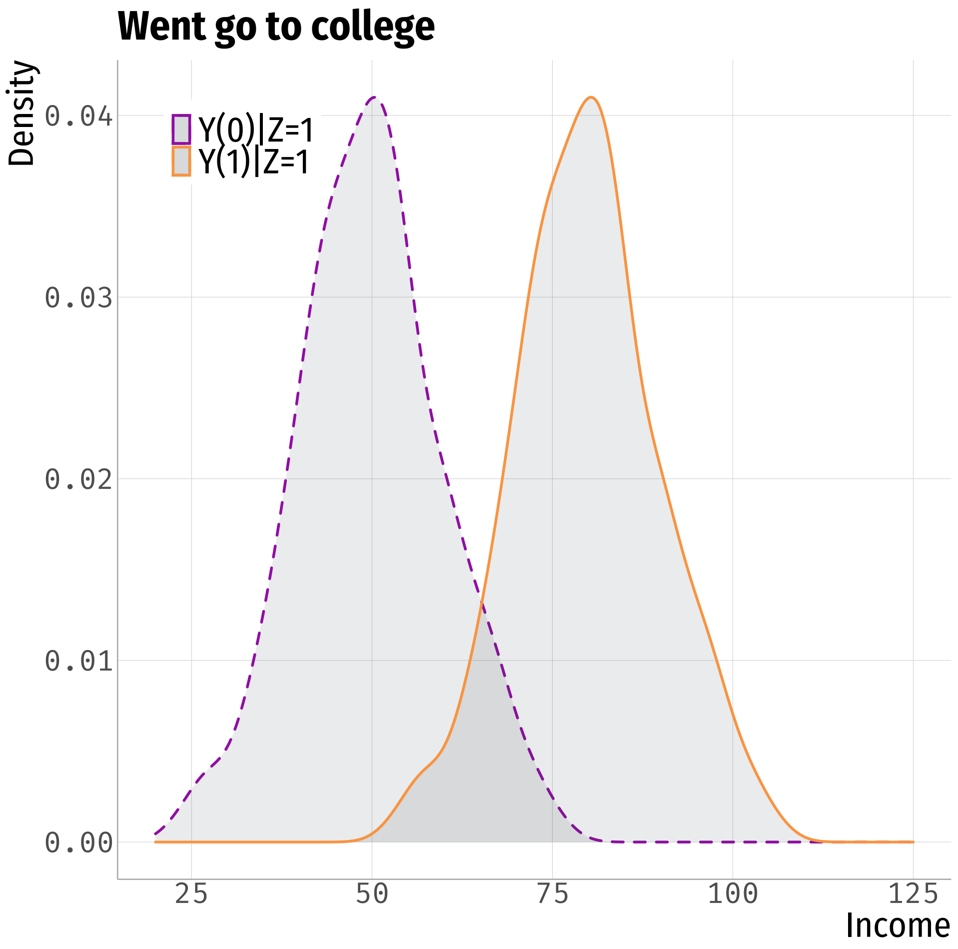
Let's see the distributions of people

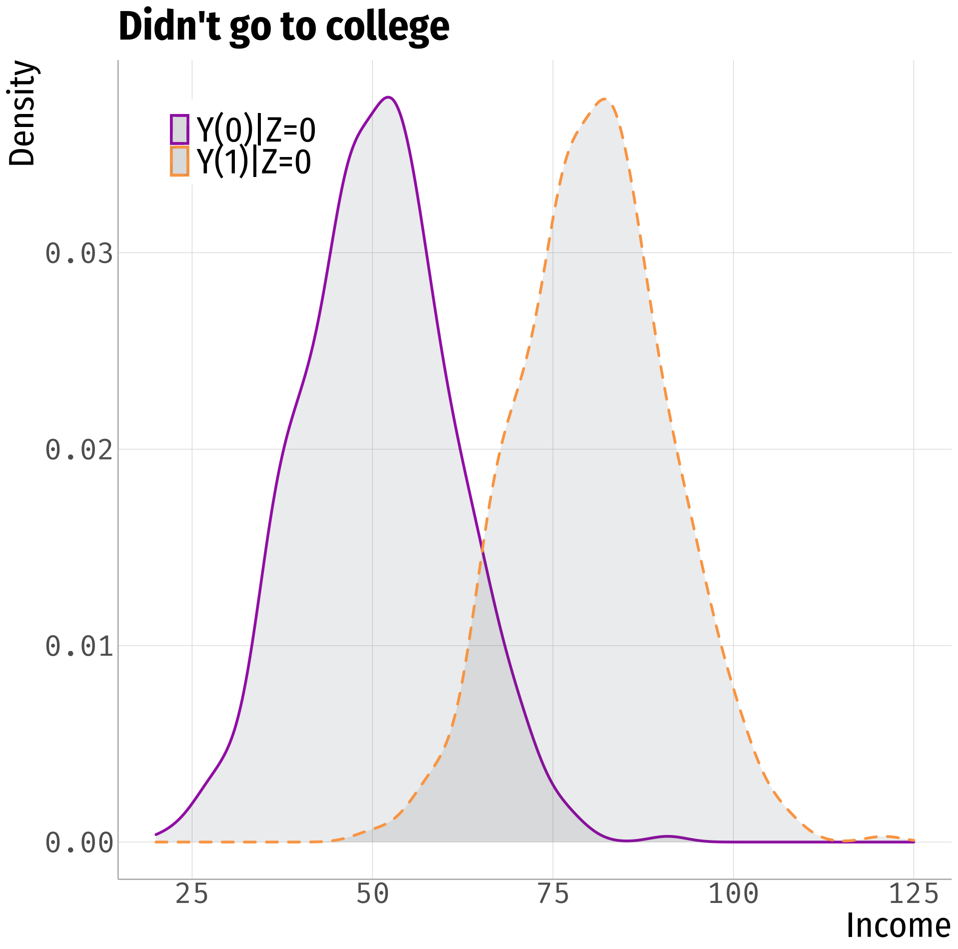


Let's see the distributions of people



We can only observe one distribution per group!

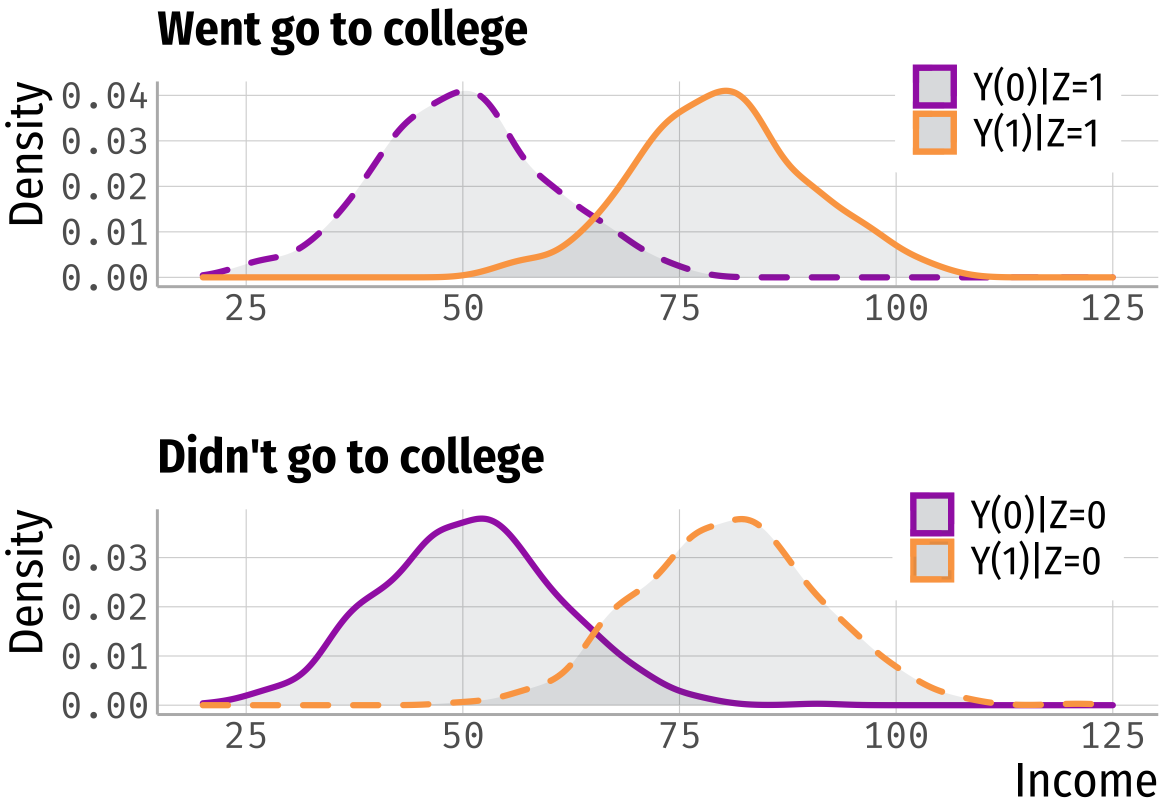




Under Ignorability Assumption

$$Y(0), Y(1) \perp\!\!\!\perp Z$$

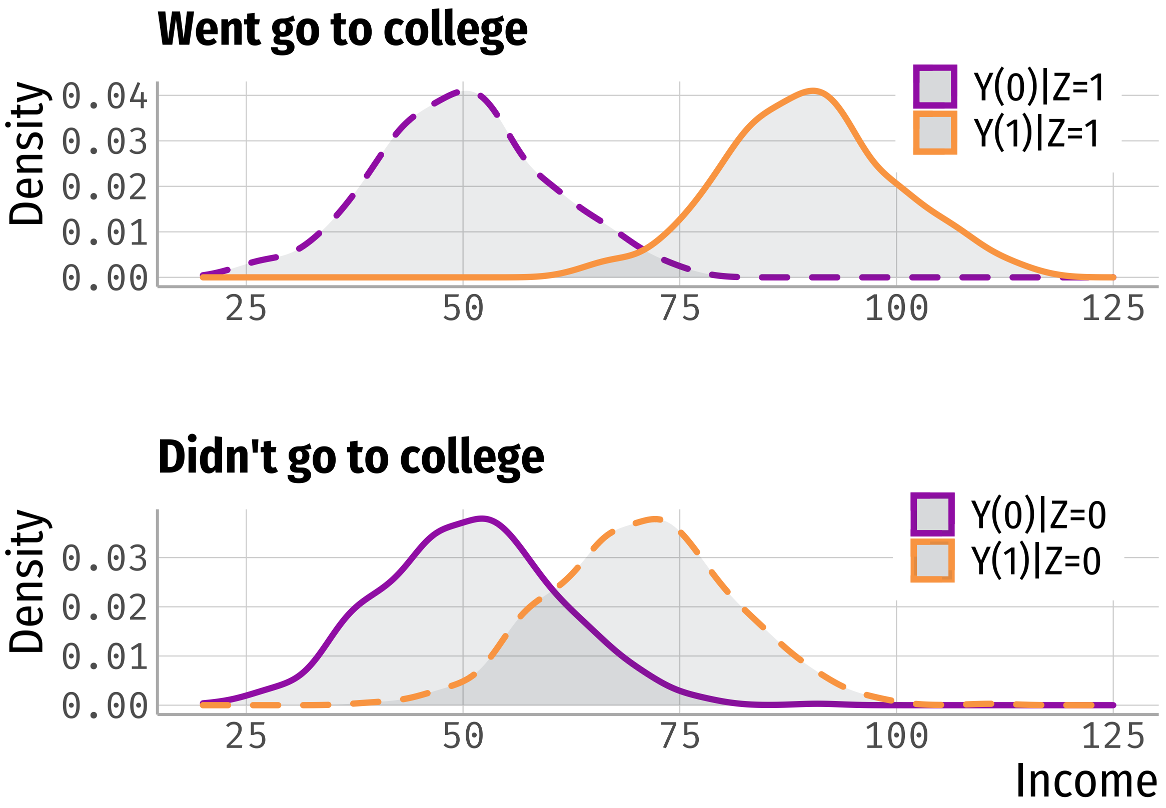
$$Income(0), Income(1) \perp\!\!\!\perp College$$



What about if the ignorability assumption doesn't hold?

$$Y(0), Y(1) \not\perp Z$$

E.g. Individuals that can take more advantage from college (in terms of income) are more likely to go.



Randomized Controlled Trials

- RCTs make the **ignorability assumption hold by design**

How?

Examples of RCTs

LinkedIn Ran Social Experiments on 20 Million Users Over Five Years

A study that looked back at those tests found that relatively weak social connections were more helpful in finding jobs than stronger social ties.

Give this article 163

Researchers examined changes that LinkedIn had made to its "People You May Know" algorithm to test what sociologists call the "strength of weak ties." Sundry Photography/Alamy

The NYT Open Team Follow

Mar 4, 2021 · 8 min read · Listen

Save

How We Rearchitected Mobile A/B Testing at The New York Times

We use A/B tests to make decisions about the products and features we release, but our mobile test allocation wasn't separating users properly and we had to figure out how to fix it.

What does randomization buy us?

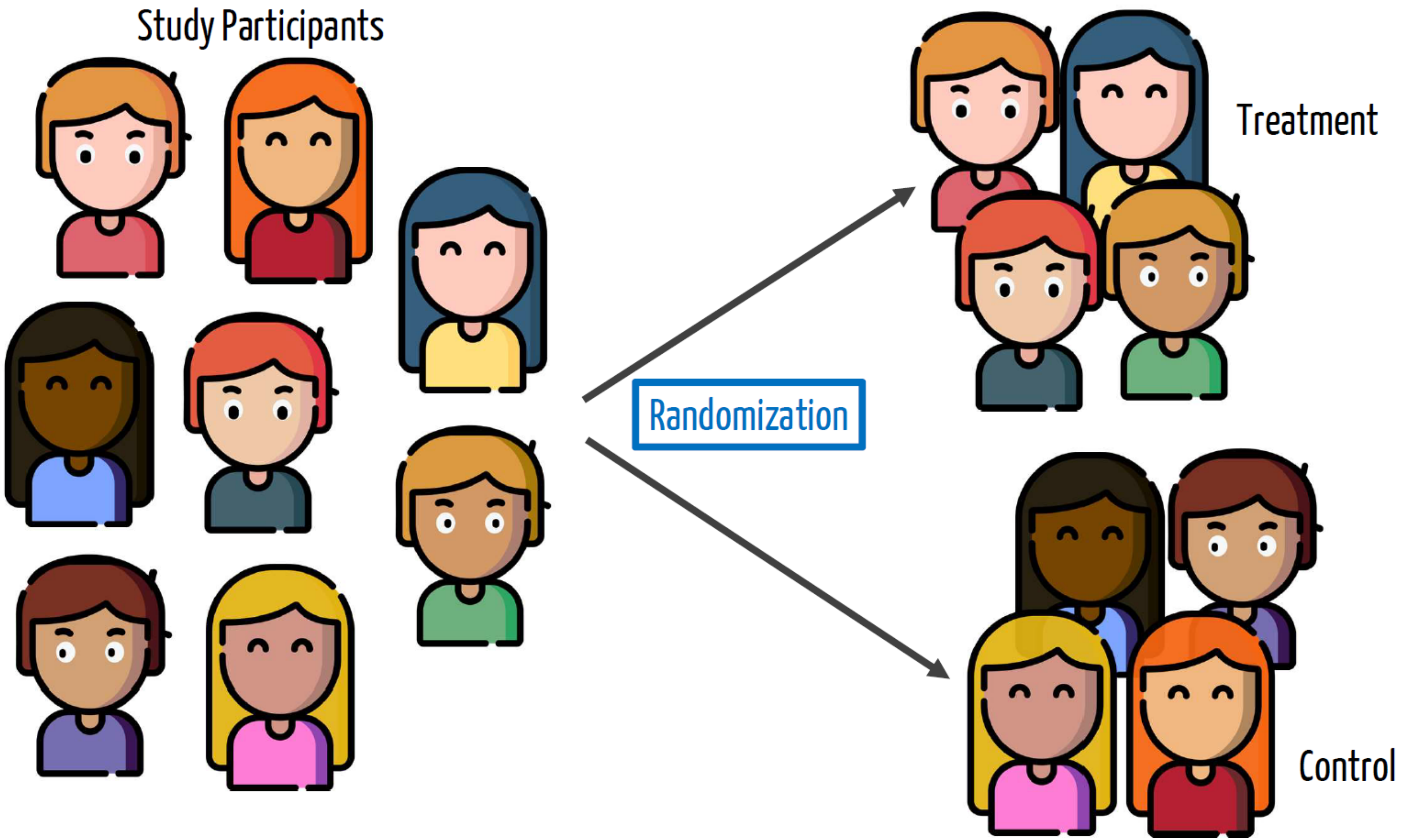
- Groups will be comparable **in expectation** in terms of their observable and unobservable characteristics.
- Are **unbalances** a problem?
  - Check magnitude and how many significant differences we have!
- Always conduct **robustness checks**!
  - Run simple regression (e.g. Y ~ Z) and then add covariates.

Can we ensure balance?

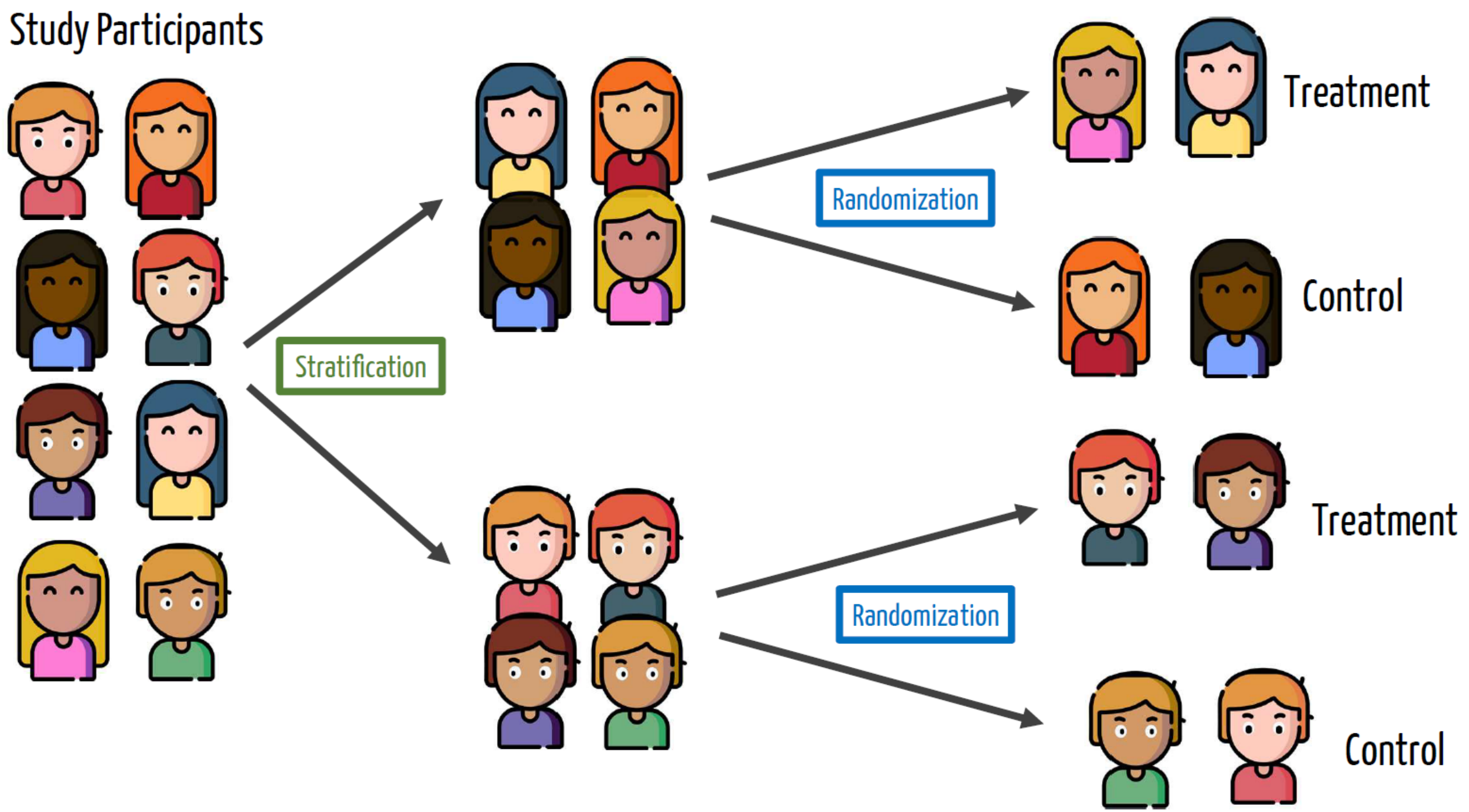
- In RCTs, in general, you can't ensure balance on all covariates.
- But you can **stratify**!
- Stratification means dividing your data into different stratas or groups S, based on one or more covariates.
- Then, you randomize in the same way *within strata*.

Randomization of treatment

Randomization of treatment with no strata



Randomization of treatment with strata



Get Out The Vote

- "Get out the Vote" Large-Scale Mobilization experiment (Arceneaux, Gerber, and Green, 2006)
    - "Households containing one or two registered voters where randomly assigned to treatment or control groups"
    - Treatment: GOTV phone calls
    - Stratified RCT: Two states divided into competitive and noncompetitive

Checking for balance

Balance Table by Stratum

	Non-competitive		Competitive		Non-competitive		Competitive	
	Treat	Control	Treat	Control	Treat	Control	Treat	Control
female2	0.552	0.546	0.541	0.535	0.549	0.545	0.543	0.541
fem_miss	0	0	0	0	0.026	0.025	0.022	0.021
age	52.157	51.977	50.81	50.862	55.795	55.782	53.481	53.464
newreg	0.117	0.116	0.133	0.134	0.048	0.049	0.048	0.046
persons	1.496	1.497	1.513	1.518	1.539	1.538	1.529	1.533
vote98	0.231	0.227	0.258	0.259	0.572	0.574	0.599	0.594
vote00	0.564	0.567	0.595	0.593	0.734	0.732	0.781	0.78

Let's go to R

Estimating the effect

- Depending on the design, usually you can **compare group means** or fit a **simple regression**

$$\frac{1}{N_T}\sum_{i \in T} Y_i - \frac{1}{N_C}\sum_{i \in C} Y_i$$

$$Y_i = \beta_0 + \beta_1 Z_i + \varepsilon_i$$

How do we incorporate stratification here?

Estimating the effect

- In stratified RCTs, we need to consider the strata!
- Run a regression with *fixed effects by strata*

$$Y_i = \beta_0 + \beta_1 Z_i + \gamma_s + \varepsilon_i$$

Estimating the effect

```r
library(estimatr)

d_s1 <- d_s1 %>% mutate(strata = interaction(state, competiv))

summary(lm_robust(vote02 ~ treat_real + strata, data = d_s1))
```

```
##
## Call:
## lm_robust(formula = vote02 ~ treat_real + strata, data = d_s1)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error t value   Pr(>|t|)  CI Lower CI Upper     DF
## (Intercept) 0.510811   0.001036 493.086  0.000e+00  0.508780  0.51284 381057
## treat_real  0.006854   0.004626   1.482  1.385e-01 -0.002213  0.01592 381057
## strata1.1   0.086655   0.003658  23.690 5.666e-124  0.079485  0.09382 381057
## strata0.2   0.048892   0.002191  22.314 3.196e-110  0.044598  0.05319 381057
## strata1.2   0.145496   0.002192  66.382  0.000e+00  0.141200  0.14979 381057
##
## Multiple R-squared:  0.01166 ,    Adjusted R-squared:  0.01165
## F-statistic:  1181 on 4 and 381057 DF,  p-value: < 2.2e-16
```

Estimating the effect

- One important thing to note in the previous analysis is that **assignment to treatment $\neq$ contact**

```
d_s1 %>% count(treat_real, contact)
```

```
##   treat_real contact      n
## 1          0       0 369068
## 2          1       0   6980
## 3          1       1   5014
```

Does this break the ignorability assumption?

When we assume…

Other potential issues to have in mind

Generalizability of our estimated effects

- Where did we get our sample for our study from? Is it representative of a larger population?

Spillover effects

- Can an individual in the control group be affected by the treatment?

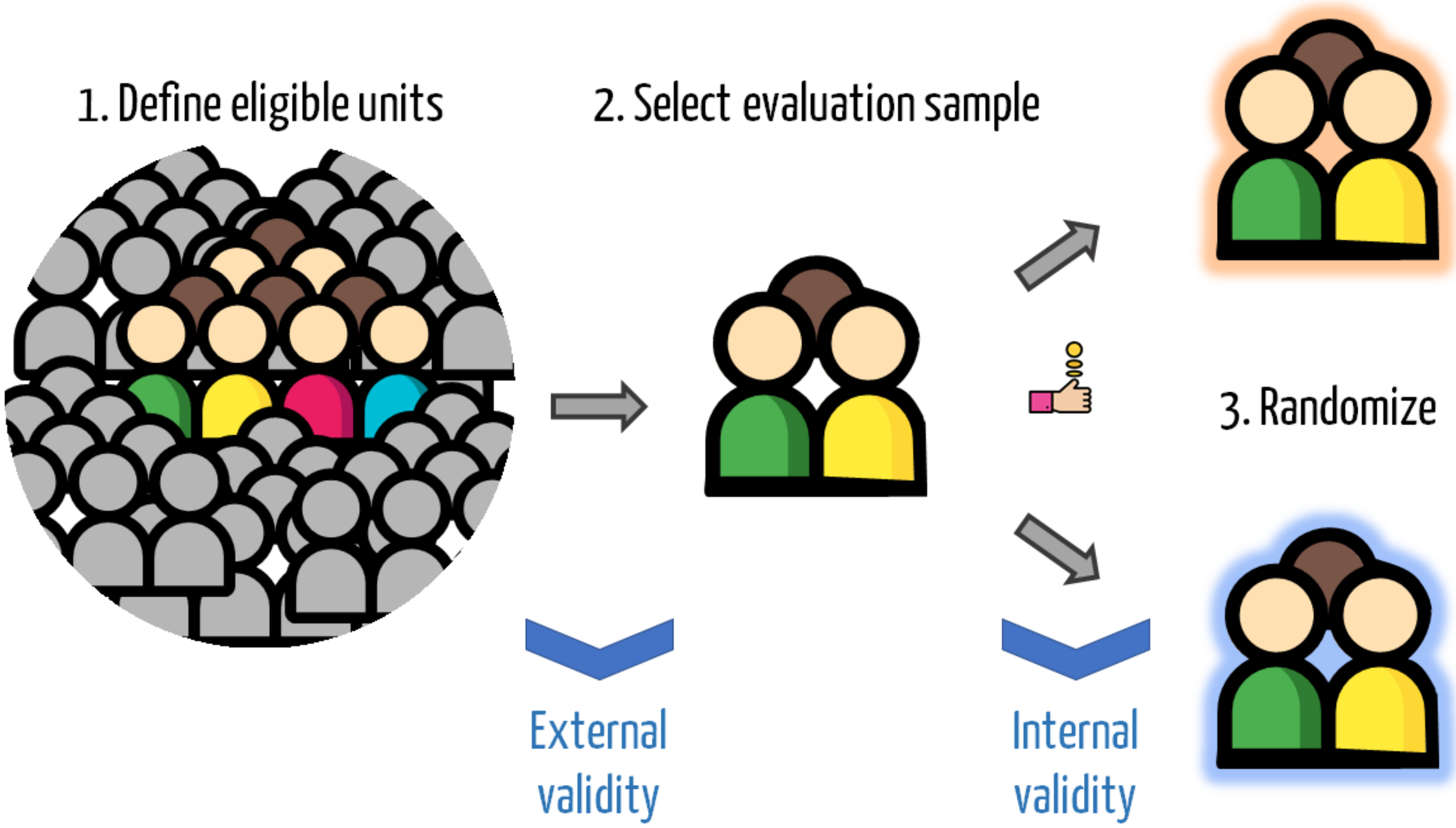
General equilibrium effects

- What happens if we scale up an intervention? Will the effect be the same?

Generalizibility of RCTs

- External Validity vs Internal Validity:
  - External validity: "The extent to which results can be generalized to other contexts or populations."
  - Internal validity: "[T]he extent to which the observed results represent the truth in the population we are studying."

External vs Internal Validity



- Many times, RCTs use **convenience samples**

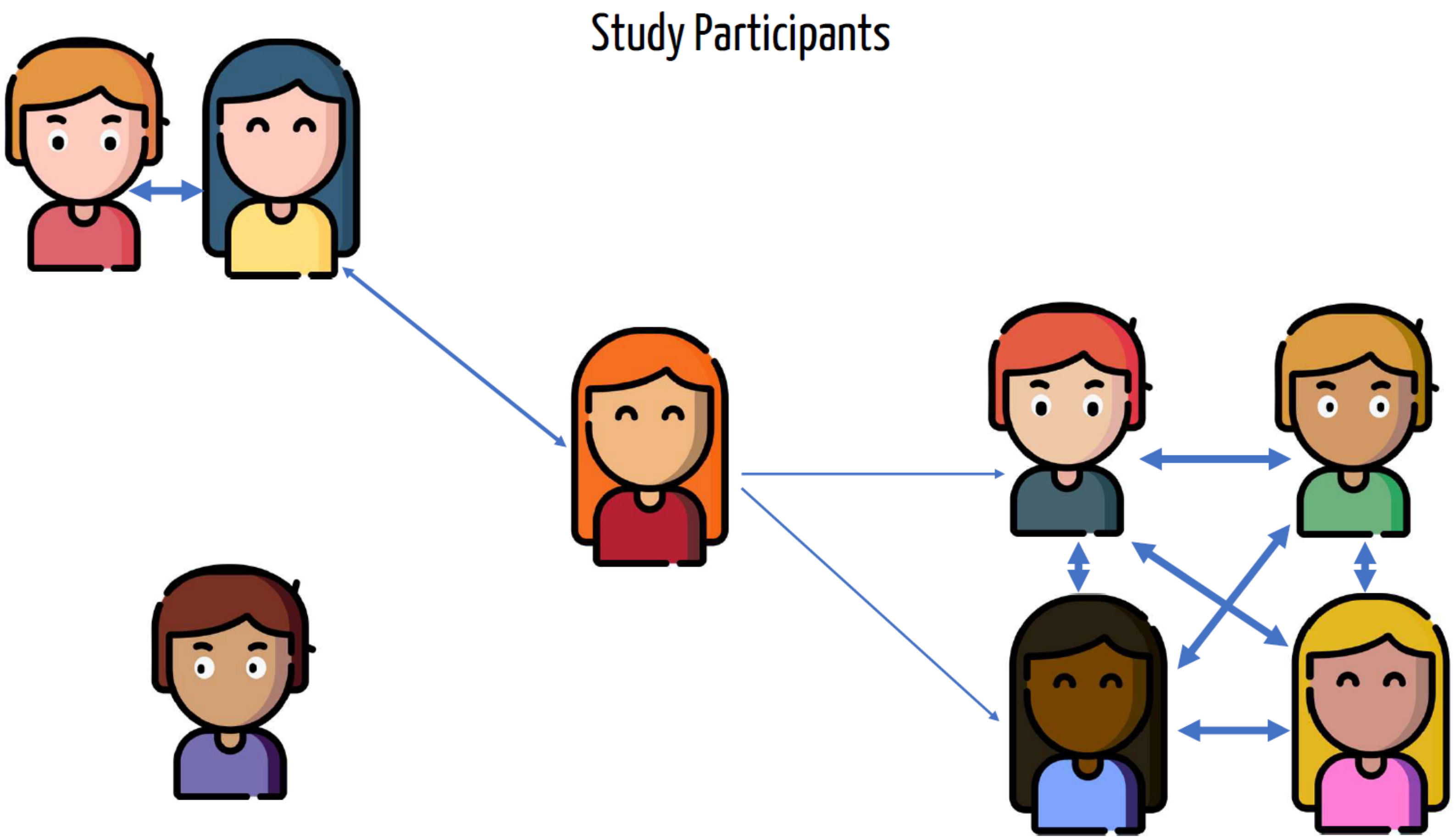
SUTVA: No interference

- Aside from ignorability, RCTs rely on the Stable Unit Treatment Value Assumption (SUTVA)

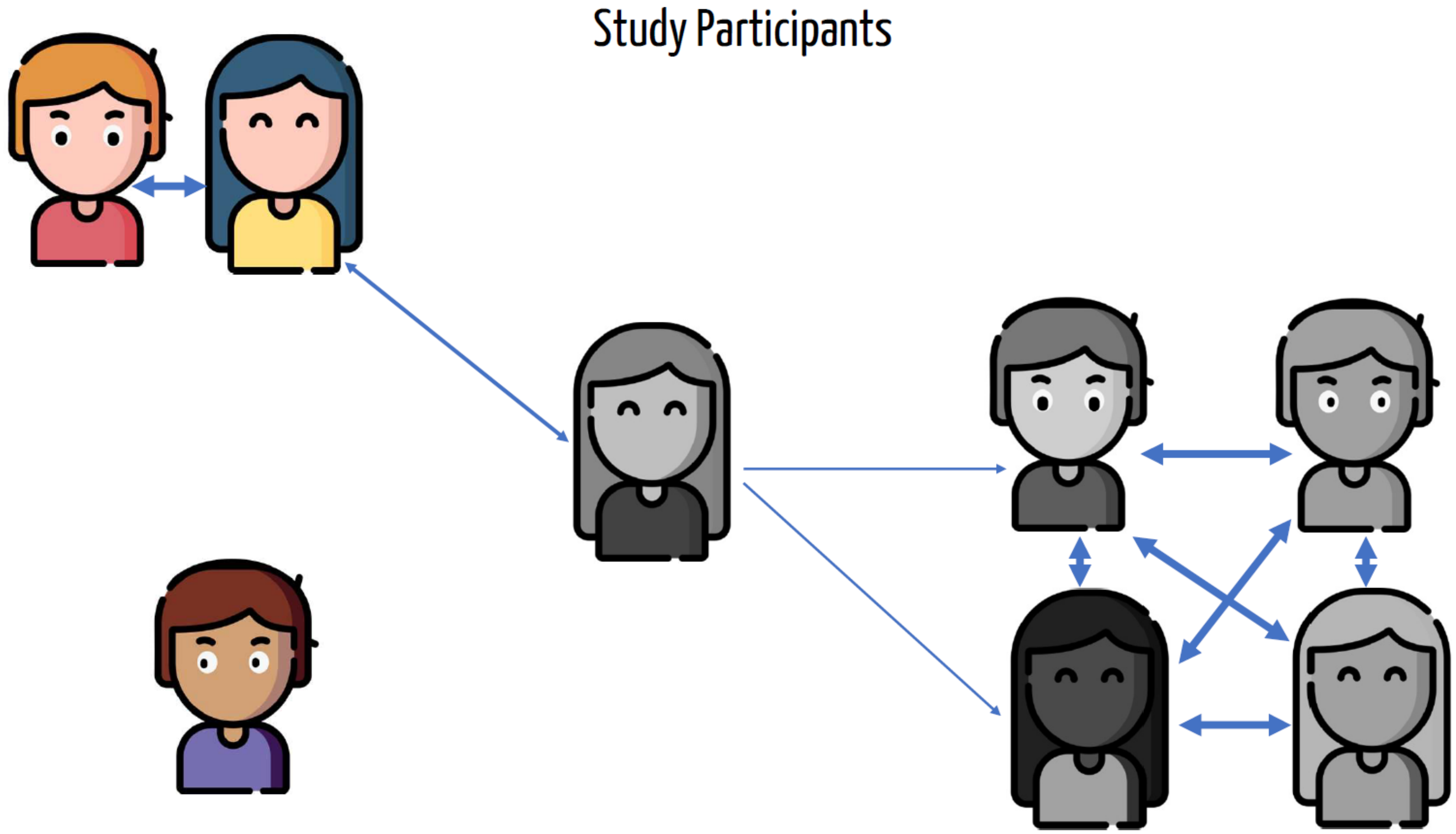
"The treatment applied to one unit does not affect the outcome for other units"

- No spillovers
- No general equilibrium effects

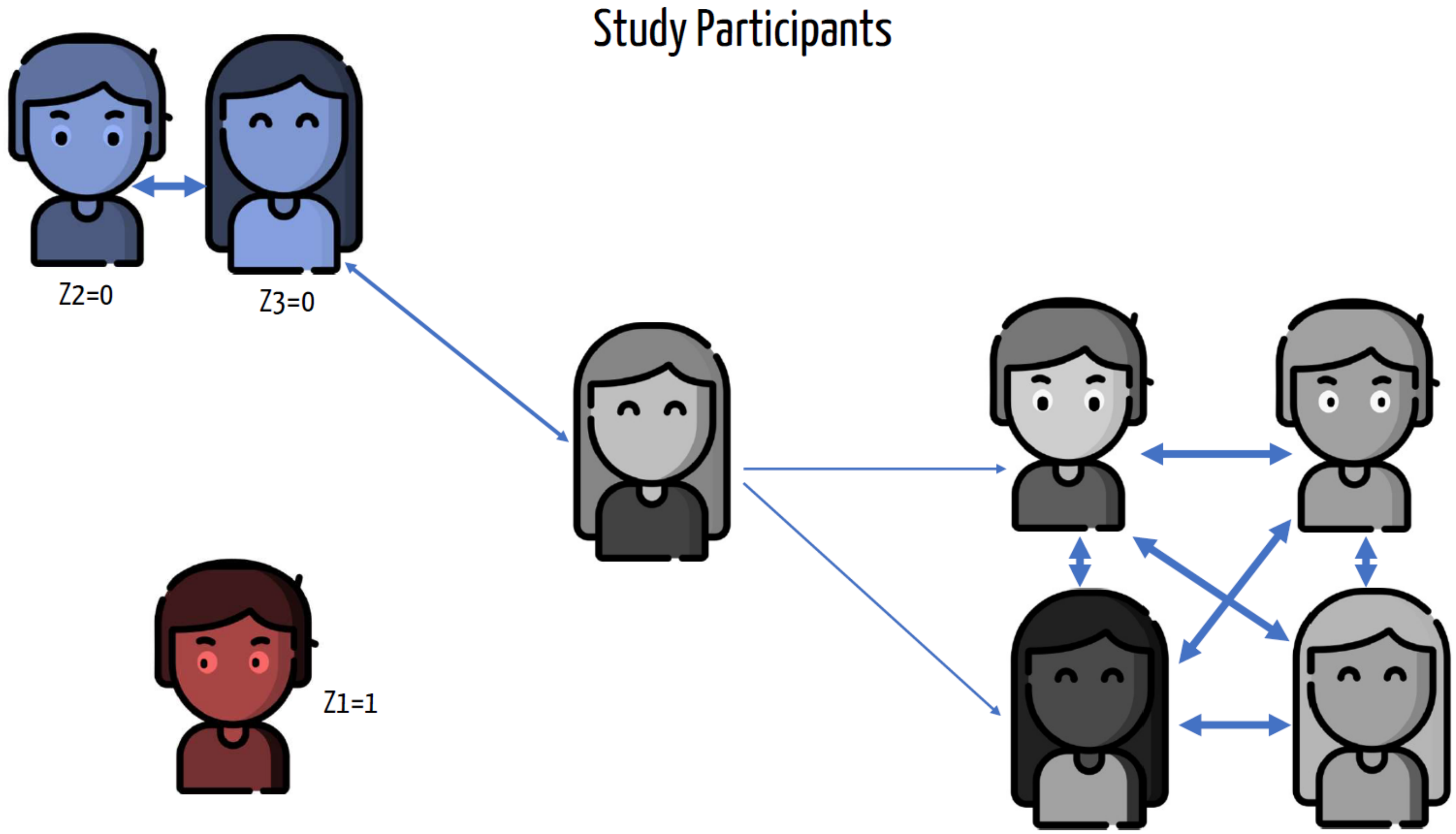
Network effects



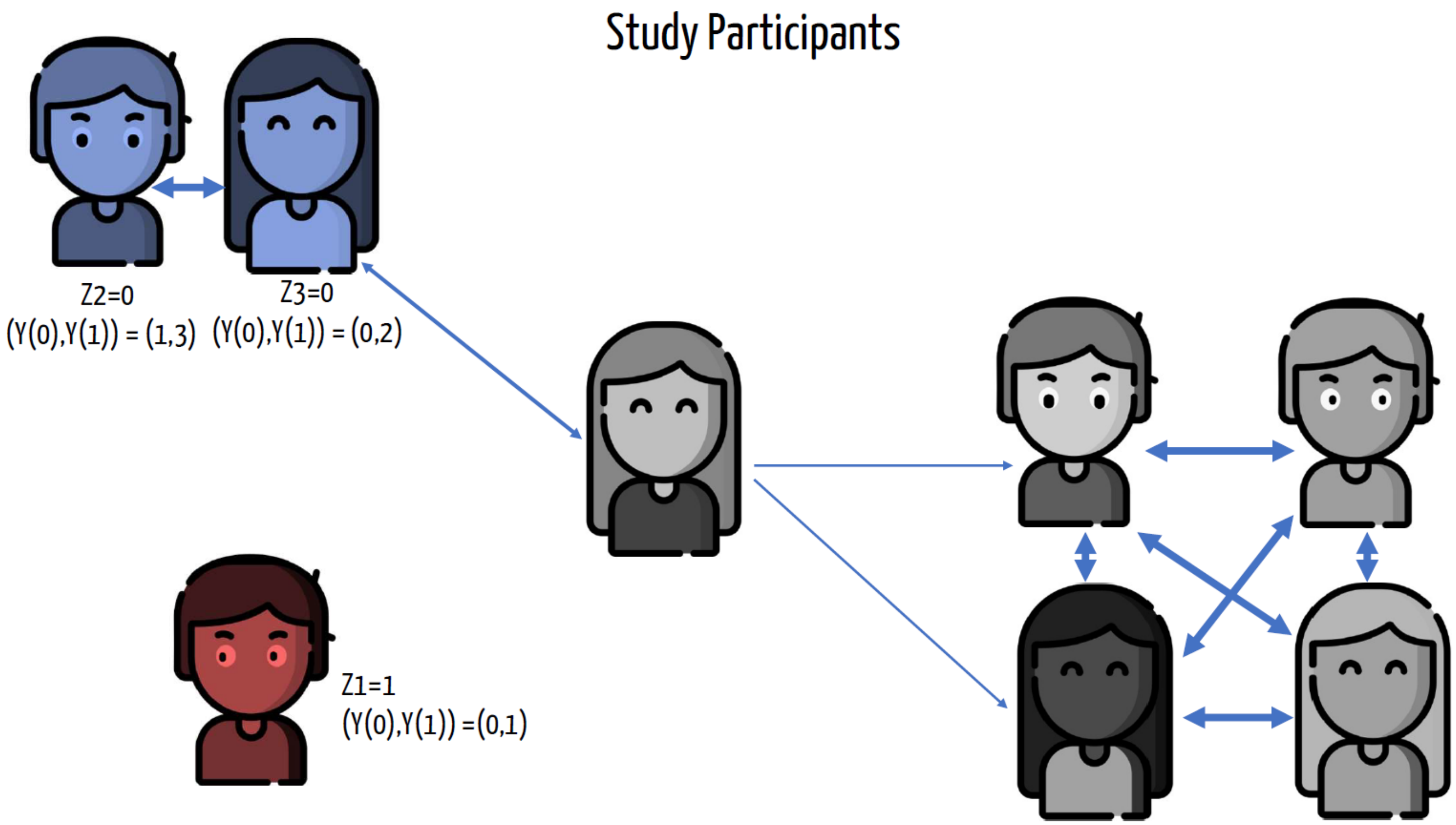
Network effects



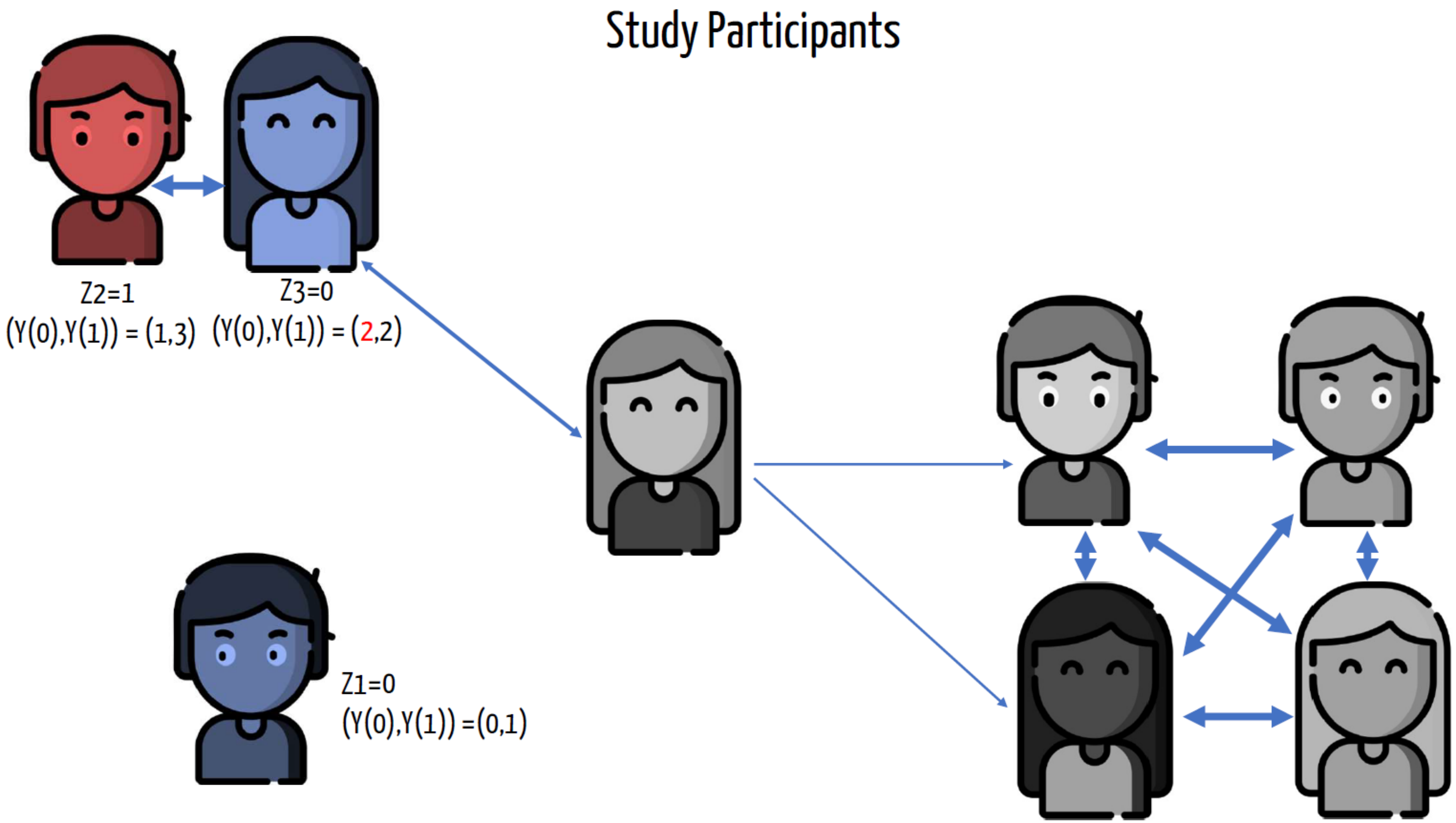
Network effects



Network effects



Network effects



Network effects

Can we do something about this?

1. Randomize at a higher level (e.g. neighborhood, school, etc. instead of at the individual level)
    - Note that you will have to cluster your standard errors
2. Model the network!

General Equilibrium Effects

- Usually arise when you **scale up** a program or intervention.
- Imagine you want to test the effect of providing information about employment and expected income to students to see whether it affect their choice of university and/or major.

What could happen if you offer it to everyone?

Wrapping things up

- Randomized controlled trials are great… but not for everything!
- Randomization buys us no systematic selection on observables or unobservables
  - But things can go wrong, too!

Check your assumptions and look out for potential issues!