causal inference is not a statistical problem

Davi

2023-06-15

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.2     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dagitty)
library(ggdag)
```

```
##
## Attaching package: 'ggdag'
##
## The following object is masked from 'package:stats':
##
##     filter
```

Variáveis exogenas

```
e_x = rnorm(1000)
e_y = rnorm(1000)
e_z = rnorm(1000)
```

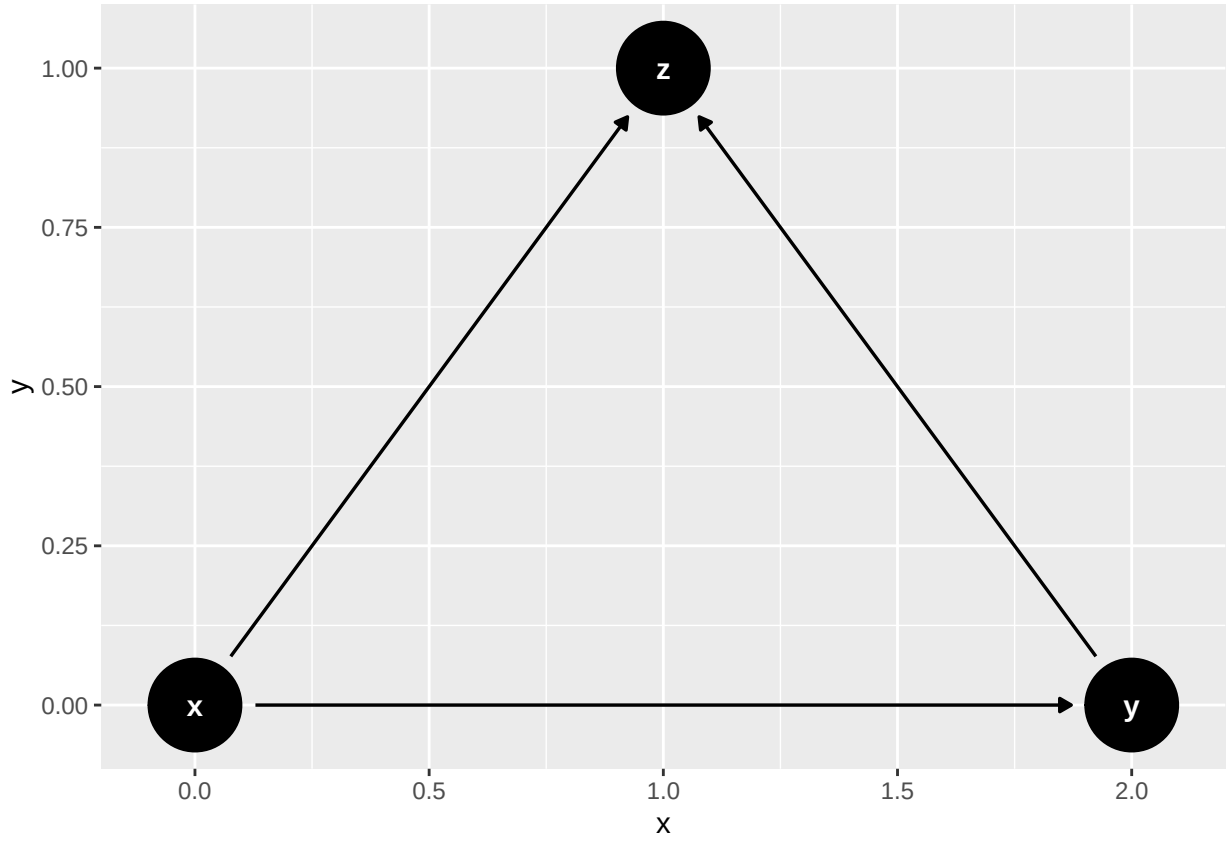
Collider

```
#SCM
x = rnorm(1000)
y = x + e_y
z = .45 * x + .77 * y + e_z

df_collider = data.frame(x, y, z)

#Causal Graph
coords = data.frame(matrix(c('x', 0, 0,
                             'z', 1, 1,
                             'y', 2, 0), nrow = 3, byrow = T))
colnames(coords) = c('name', 'x', 'y')
dag = dagify(z ~ y, z ~ x,
```

```
            y ~ x,
            exposure = 'x',
            outcome = 'y',
            coords = coords)
ggdag(dag)
```




```
v = round(c(cor(x, z), coef(lm(y ~ x, df_collider))[2], coef(lm(y ~ x + z, df_collider))[2]), digits = 
cat('Correlação X e Y:', v[1], '\nATE Y ~ X:', v[2], '\nATE Y ~ X | Z:', v[3])
```

```
## Correlação X e Y: 0.7
## ATE Y ~ X: 1
## ATE Y ~ X | Z: 0.4
```

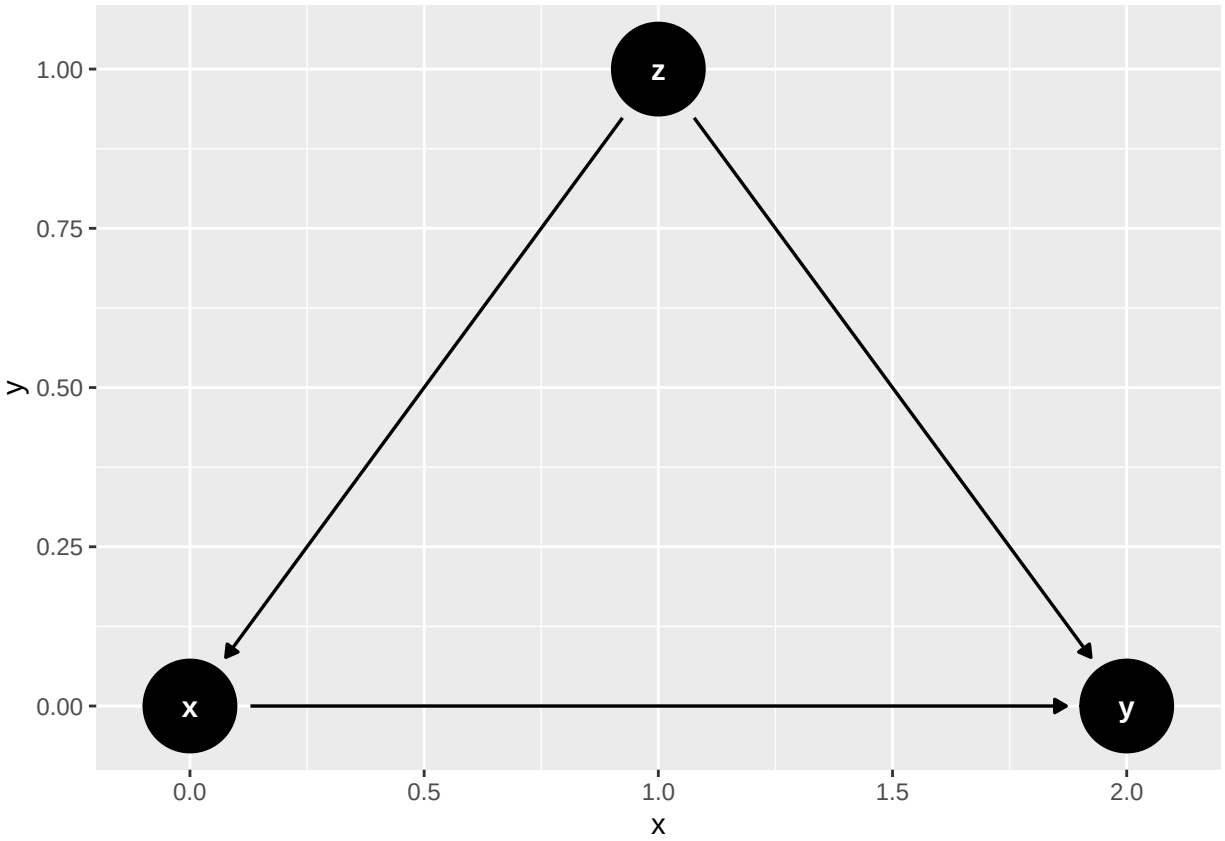
Confounder

```
#SCM
z = rnorm(1000)
x = z + e_x
y = .5 * x + z + e_z

df_confounder = data.frame(x, y, z)

#Causal Graph
coords = data.frame(matrix(c('x', 0, 0,
                             'z', 1, 1,
                             'y', 2, 0), nrow = 3, byrow = T))
```

```r
colnames(coords) = c('name', 'x', 'y')
dag = dagify(y ~ z, y ~ x,
             x ~ z,
             exposure = 'x',
             outcome = 'y',
             coords = coords)
ggdag(dag)
```




```r
v = round(c(cor(x, z), coef(lm(y ~ x, df_confounder))[2], coef(lm(y ~ x + z, df_confounder))[2]), digits

cat('Correlação X e Y:', v[1], '\nATE Y ~ X:', v[2], '\nATE Y ~ X | Z:', v[3])
```

```
## Correlação X e Y: 0.7
## ATE Y ~ X: 1
## ATE Y ~ X | Z: 0.5
```

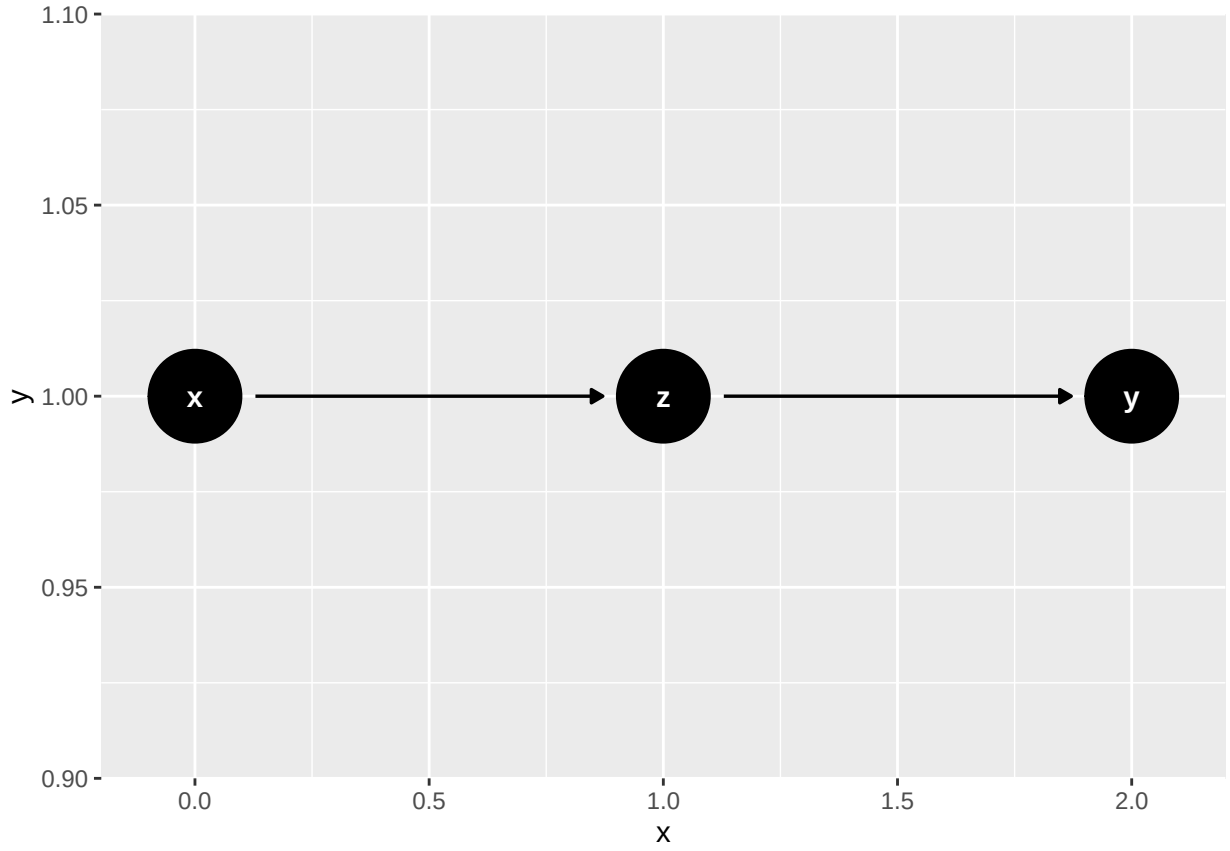
Mediator

```r
#SCM
x = rnorm(1000)
z = x + e_z
y = z + e_y

df_mediator = data.frame(x, y, z)

#Causal Graph
```

```
coords = data.frame(matrix(c('x', 0, 1,
                             'z', 1, 1,
                             'y', 2, 1), nrow = 3, byrow = T))
colnames(coords) = c('name', 'x', 'y')
dag = dagify(y ~ z,
             z ~ x,
             exposure = 'x',
             outcome = 'y',
             coords = coords)
ggdag(dag)
```




```
v = round(c(cor(x, z), coef(lm(y ~ x, df_mediator))[2], coef(lm(y ~ x + z, df_mediator))[2]), digits =

cat('Correlação X e Y:', v[1], '\nATE Y ~ X:', v[2], '\nATE Y ~ X | Z:', v[3])
```

```
## Correlação X e Y: 0.7
## ATE Y ~ X: 1
## ATE Y ~ X | Z: 0
```

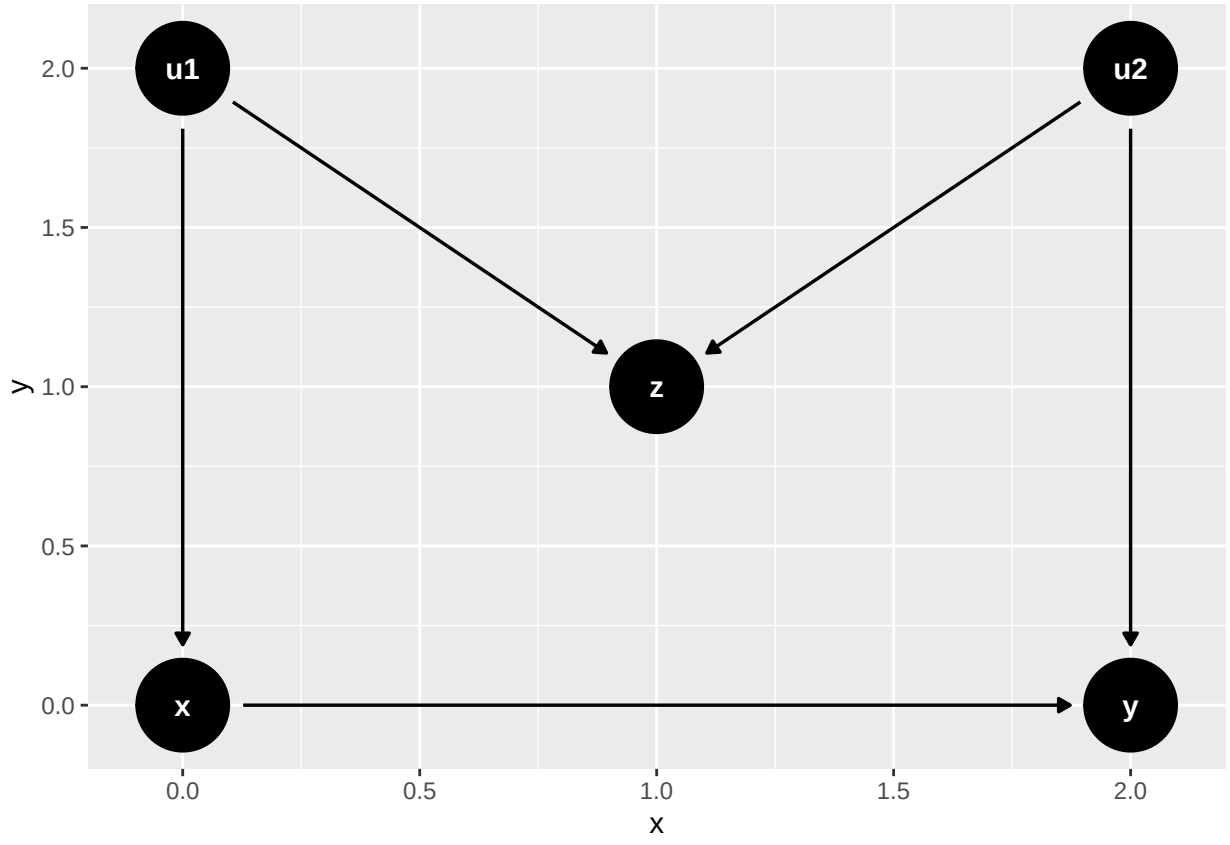
M-Bias

```
#SCM
u1 = rnorm(1000)
u2 = rnorm(1000)
z = 8 * u1 + u2 + e_z
x = u1 + e_x
```

```
y = x + u2 + e_y

df_mbias = data.frame(x, y, z)

#Causal Graph
coords = data.frame(matrix(c('x', 0, 0,
                             'z', 1, 1,
                             'y', 2, 0,
                             'u1', 0, 2,
                             'u2', 2, 2), nrow = 5, byrow = T))
colnames(coords) = c('name', 'x', 'y')
dag = dagify(z ~ u1, z ~ u2,
             x ~ u1,
             y ~ x, y ~ u2,
             exposure = 'x',
             outcome = 'y',
             coords = coords)
ggdag(dag)
```



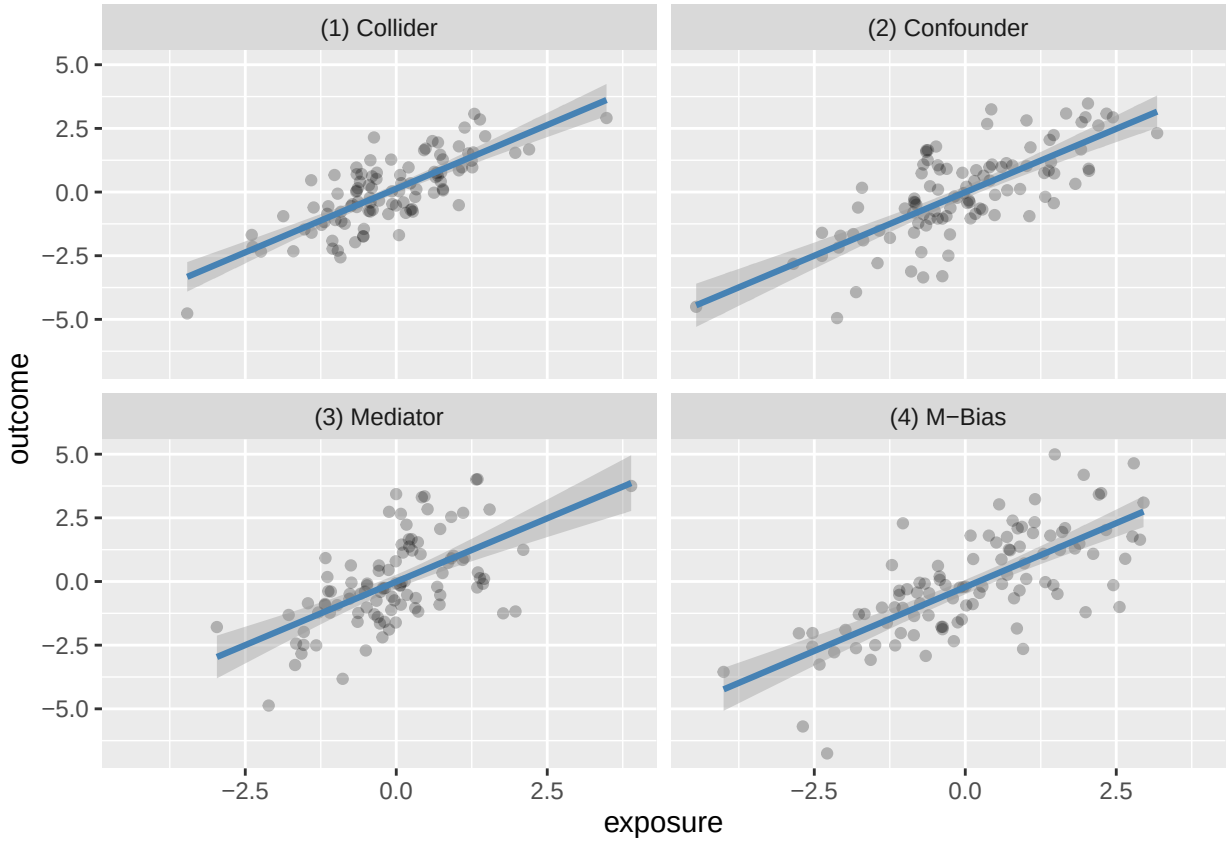

```
v = round(c(cor(x, z), coef(lm(y ~ x, df_mbias))[2], coef(lm(y ~ x + z, df_mbias))[2]), digits = 1)
cat('Correlação X e Y:', v[1], '\nATE Y ~ X:', v[2], '\nATE Y ~ X | Z:', v[3])
```

```
## Correlação X e Y: 0.7
## ATE Y ~ X: 1
## ATE Y ~ X | Z: 0.9
```

Gráficos do artigo

```
library(quartets)

ggplot(causal_quartet, aes(x = exposure, y = outcome)) +
geom_point(alpha = 0.25) +
geom_smooth(
method = "lm",
formula = "y ~ x",
linewidth = 1.1,
color = "steelblue") +
facet_wrap(~dataset)
```



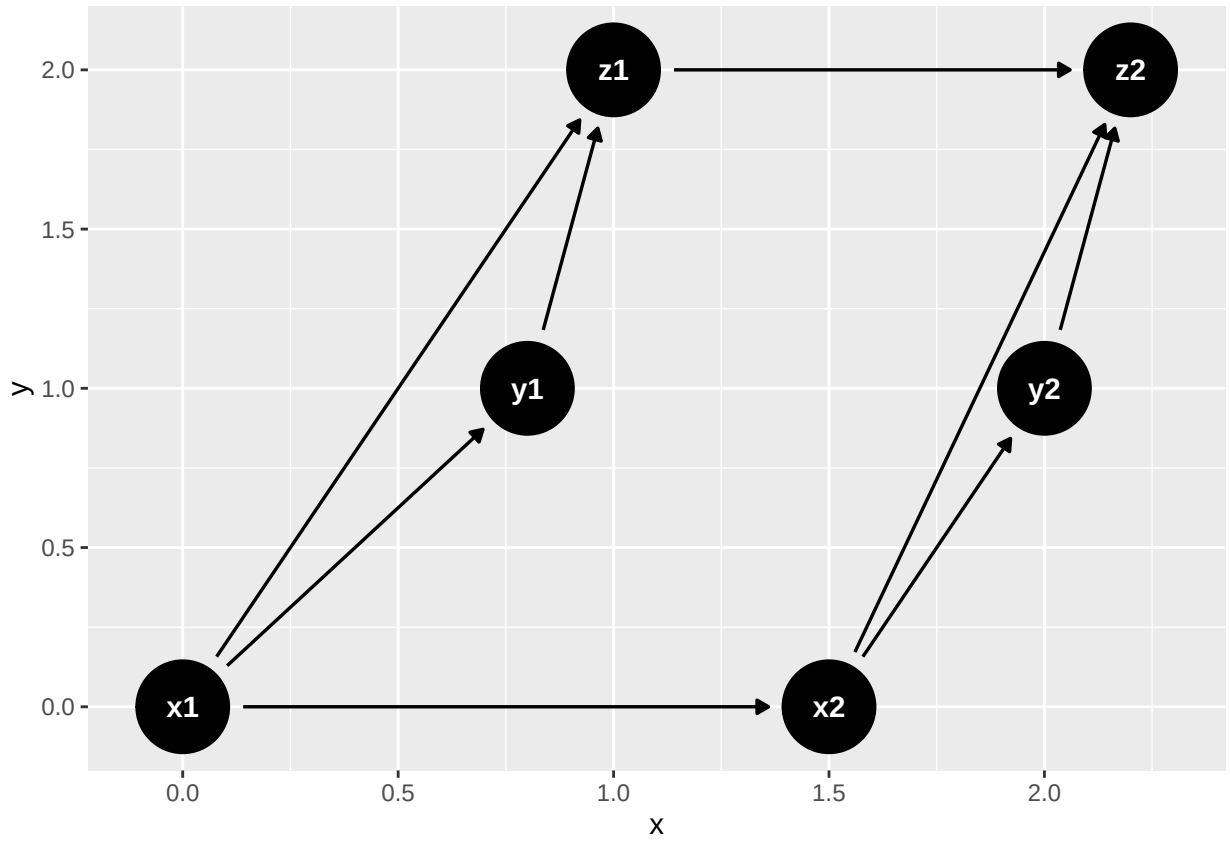

```
## Table 2
```

Time ordered DAGs

Collider

```
coords = data.frame(matrix(c('x1', 0, 0,
                             'z1', 1, 2,
                             'y1', .8, 1,
                             'x2', 1.5, 0,
                             'z2', 2.2, 2,
                             'y2', 2, 1), nrow = 6, byrow = T))
colnames(coords) = c('name', 'x', 'y')
```

```
dag = dagify(z2 ~ y2, z2 ~ x2, z2 ~ z1,
             y2 ~ x2,
             x2 ~ x1,
             z1 ~ y1, z1 ~ x1,
             y1 ~ x1,
             exposure = 'x2',
             outcome = 'y2',
             coords = coords)
ggdag(dag)
```




```
ATE = round(c(coef(lm(outcome_followup ~ exposure_baseline, causal_collider_time))[2],
coef(lm(outcome_followup ~ exposure_baseline + covariate_baseline, causal_collider_time))[2]),
digits = 2)

cat('ATE sem controlar por Z:', ATE[1], '\nATE controlando por Z:', ATE[2])
```

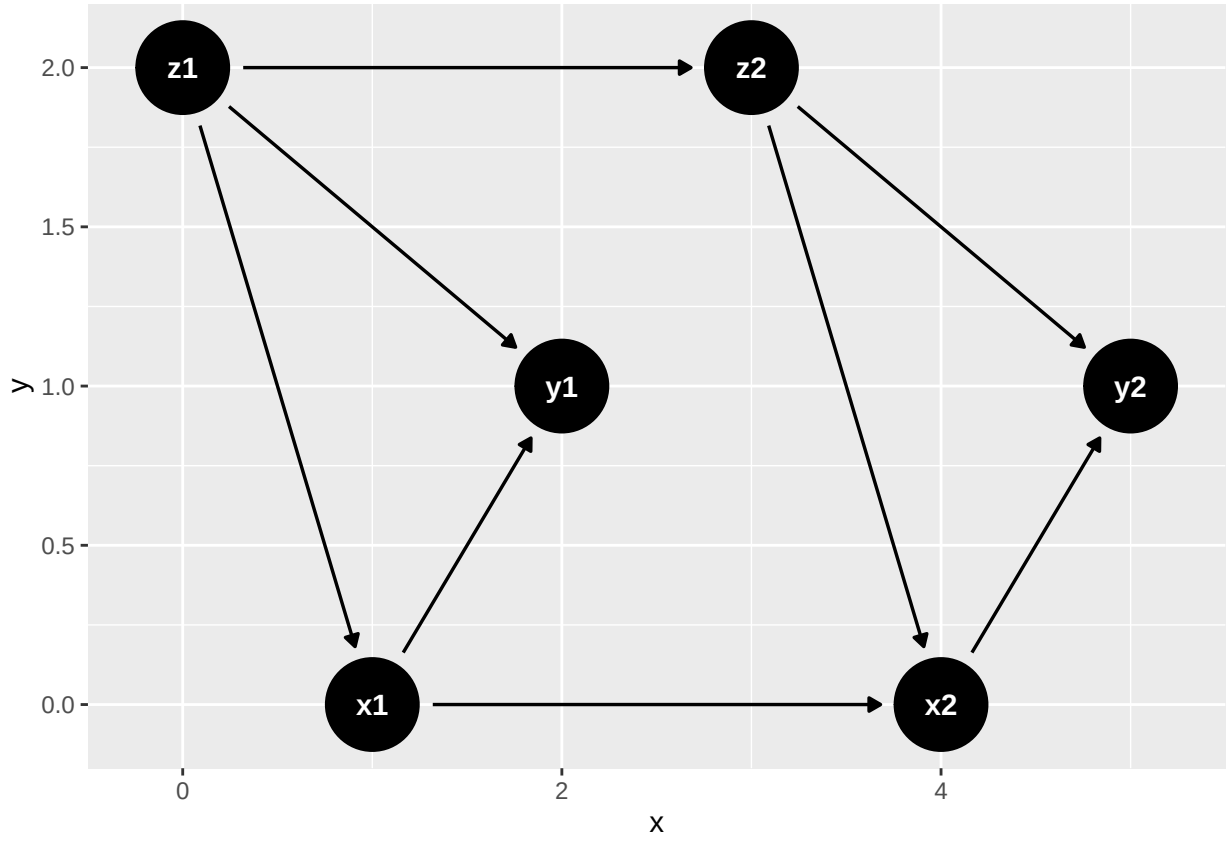
```
## ATE sem controlar por Z: 1
## ATE controlando por Z: 1
```

Confounder

```
coords = data.frame(matrix(c('x1', 1, 0,
                             'z1', 0, 2,
                             'y1', 2, 1,
                             'x2', 4, 0,
                             'z2', 3, 2,
                             'y2', 5, 1), nrow = 6, byrow = T))
```

```
colnames(coords) = c('name', 'x', 'y')

dag = dagify(y2 ~ x2, y2 ~ z2,
             x2 ~ z2, x2 ~ x1,
             y1 ~ z1, y1 ~ x1,
             z2 ~ z1,
             x1 ~ z1,
             exposure = 'x2',
             outcome = 'y2',
             coords = coords)
ggdag(dag)
```




```
ATE = round(c(coef(lm(outcome_followup ~ exposure_baseline, causal_confounding_time))[2],
coef(lm(outcome_followup ~ exposure_baseline + covariate_baseline, causal_confounding_time))[2]),
digits = 2)

cat('ATE sem controlar por Z:', ATE[1], '\nATE controlando por Z:', ATE[2])
```

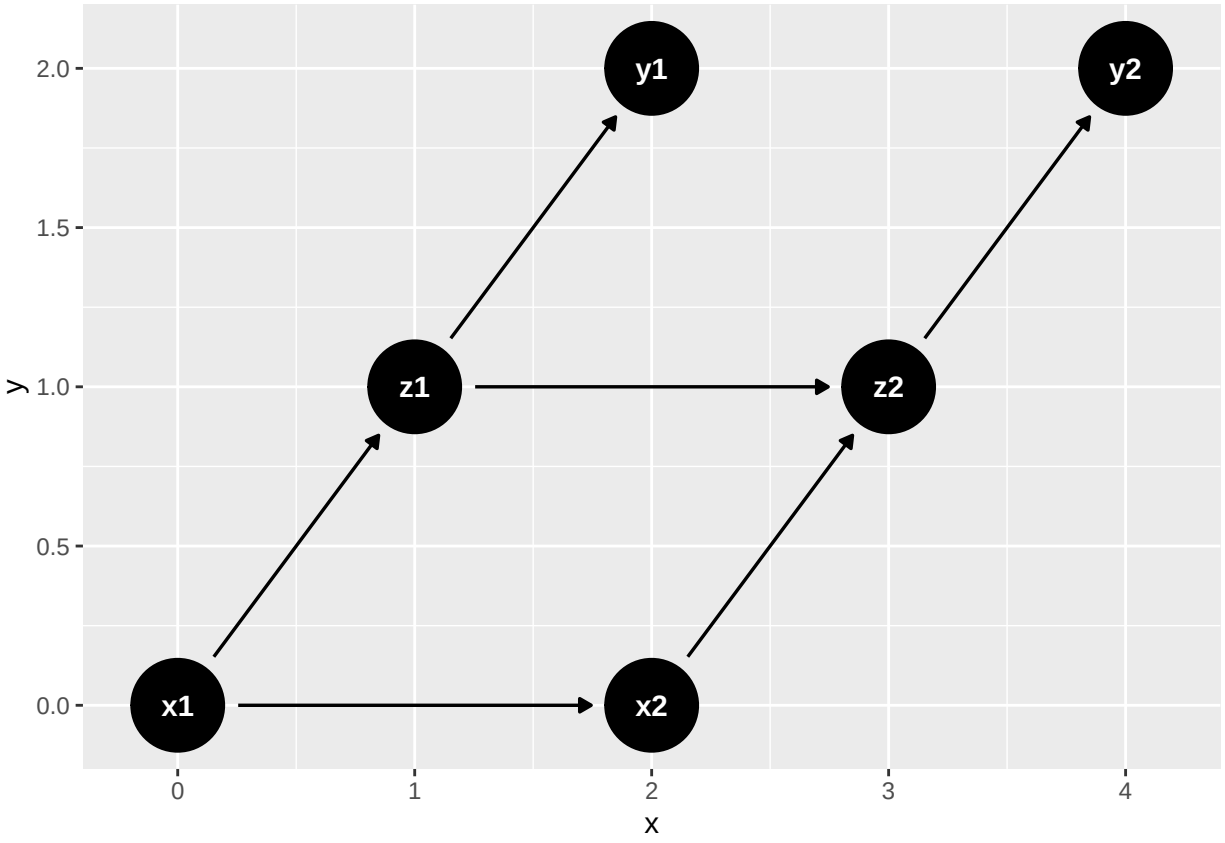
```
## ATE sem controlar por Z: 1
## ATE controlando por Z: 0.5
```

Mediator

```
coords = data.frame(matrix(c('x1', 0, 0,
                             'z1', 1, 1,
                             'y1', 2, 2,
                             'x2', 2, 0,
```

```
                              'z2', 3, 1,
                              'y2', 4, 2), nrow = 6, byrow = T))
colnames(coords) = c('name', 'x', 'y')

dag = dagify(y2 ~ z2,
             z2 ~ z1, z2 ~ x2,
             x2 ~ x1,
             y1 ~ z1,
             z1 ~ x1,
             exposure = 'x2',
             outcome = 'y2',
             coords = coords)
ggdag(dag)
```




```
ATE = round(c(coef(lm(outcome_followup ~ exposure_baseline, causal_mediator_time))[2],
coef(lm(outcome_followup ~ exposure_baseline + covariate_baseline, causal_mediator_time))[2]),
digits = 2)

cat('ATE sem controlar por Z:', ATE[1], '\nATE controlando por Z:', ATE[2])
```

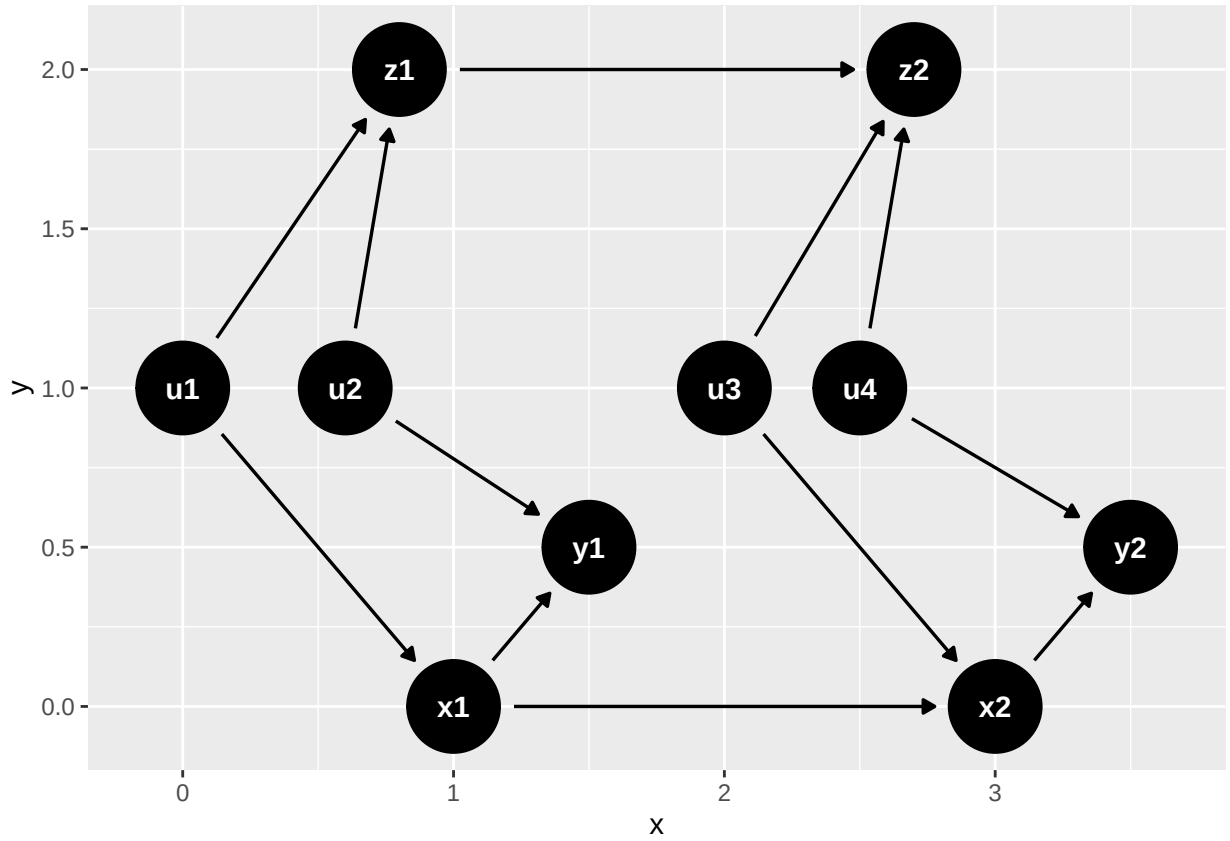
```
## ATE sem controlar por Z: 1
## ATE controlando por Z: 1
```

M-Bias

```
coords = data.frame(matrix(c('x1', 1, 0,
                             'z1', .8, 2,
```

```
                              'y1', 1.5, .5,
                              'x2', 3, 0,
                              'z2', 2.7, 2,
                              'y2', 3.5, .5,
                              'u1', 0, 1,
                              'u2', .6, 1,
                              'u3', 2, 1,
                              'u4', 2.5, 1), nrow = 10, byrow = T))
colnames(coords) = c('name', 'x', 'y')

dag = dagify(z2 ~ u4, z2 ~ u3, z2 ~ z1,
             y2 ~ u4, y2 ~ x2,
             x2 ~ x1, x2 ~ u3,
             z1 ~ u2, z1 ~ u1,
             y1 ~ x1, y1 ~ u2,
             x1 ~ u1,
             exposure = 'x2',
             outcome = 'y2',
             coords = coords)
ggdag(dag)
```




```
ATE = round(c(coef(lm(outcome_followup ~ exposure_baseline, causal_m_bias_time))[2],
coef(lm(outcome_followup ~ exposure_baseline + covariate_baseline, causal_m_bias_time))[2]),
digits = 2)

cat('ATE sem controlar por Z:', ATE[1], '\nATE controlando por Z:', ATE[2])
```

```
## ATE sem controlar por Z: 1
## ATE controlando por Z: 0.88
```