Homework 4

Yuki Joyama

```r
# libraries
library(tidyverse)
library(ggplot2)
library(dagitty)

# setup plot theme
theme_set(
  theme_bw() +
    theme(legend.position = "top")
  )
```

```r
# import data
df = read_csv("./data/hw4_data.csv")
```

1. DAG

```r
g = dagitty('dag {
age [pos="-1.842,-0.039"]
black [pos="1.458,-0.317"]
educ [pos="-0.698,-0.114"]
hispan [pos="0.269,-0.760"]
married [pos="-2.461,-0.426"]
nodegree [pos="-1.567,-0.659"]
re74 [pos="-1.219,0.594"]
re75 [pos="0.011,0.971"]
re78 [outcome,pos="0.893,0.468"]
treat [exposure,pos="-2.558,0.500"]
age -> educ
age -> married
age -> re74
age -> re75
age -> re78
age -> treat
black -> educ
black -> nodegree
black -> re74
black -> re75
black -> re78
black -> treat
educ -> re74
educ -> re75
educ -> re78
educ -> treat
```

```
hispan -> educ
hispan -> nodegree
hispan -> re74
hispan -> re75
hispan -> re78
hispan -> treat
married -> treat
nodegree -> educ
nodegree -> re74
nodegree -> re75
nodegree -> re78
nodegree -> treat
re74 -> re75
re74 -> re78
re75 -> re78
treat -> re74
treat -> re75
treat -> re78
}')
plot(g)
```
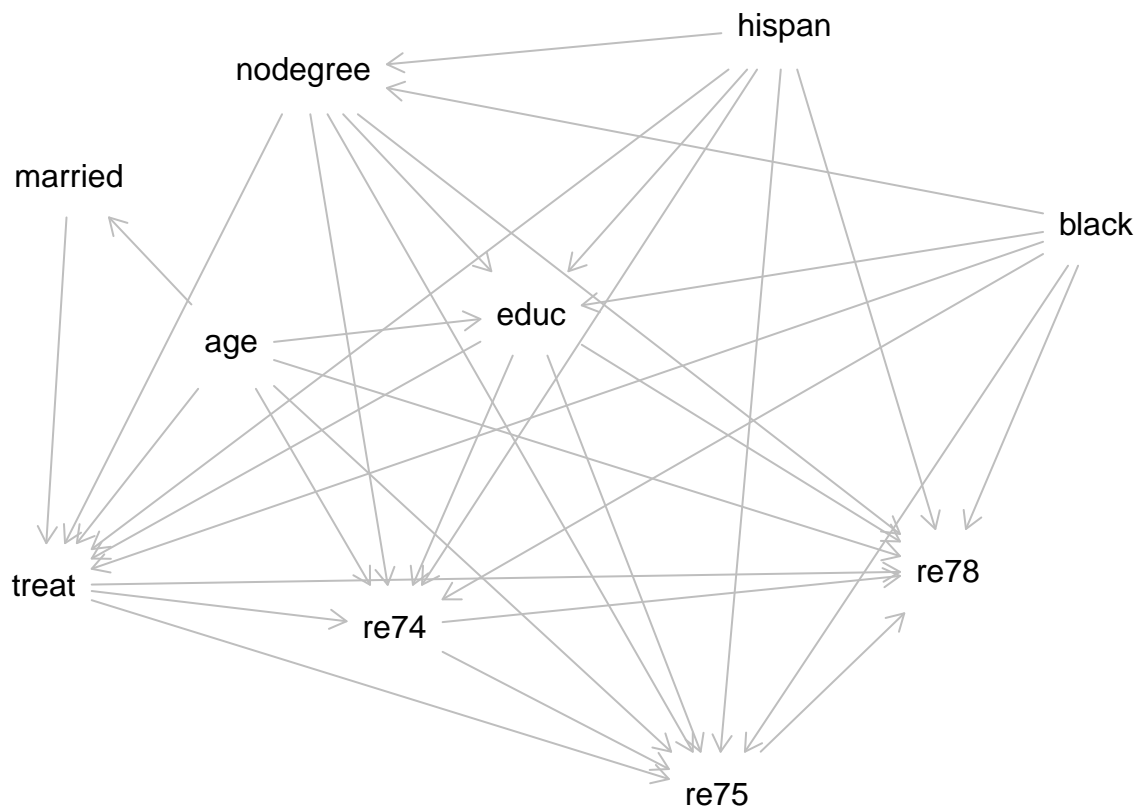

Note of the variables in the DAG:

treat: treatment assignment (job training); exposure

- Suppose that the job training was completed before 1974, **treat** is likely to influence **re74**, **re75** and **re78**

age: age in years

- `age` may affect `married`, `treat`, `educ` and all the income status (`re74`, `re75`, `re78`)

`educ`: education in years
- `educ` may affect all the income status and `treat`

`black`, `hispan`: indicators for African American and hispanic
- Both of the ethnicity indicator may affect all the income status, `nodegree`, `treat`, and `educ`

`married`: indicator for married
- `married` could influence `treat`

`nodegree`: indicator for highschool degree
- `nodegree` may affect `treat`, `educ`, and all the income status

`re74`: income in 1974
- `re74` can influence `re75` and `re78`

`re75`: income in 1975
- `re75` can influence `re78`

`re78`: income in 1978; outcome