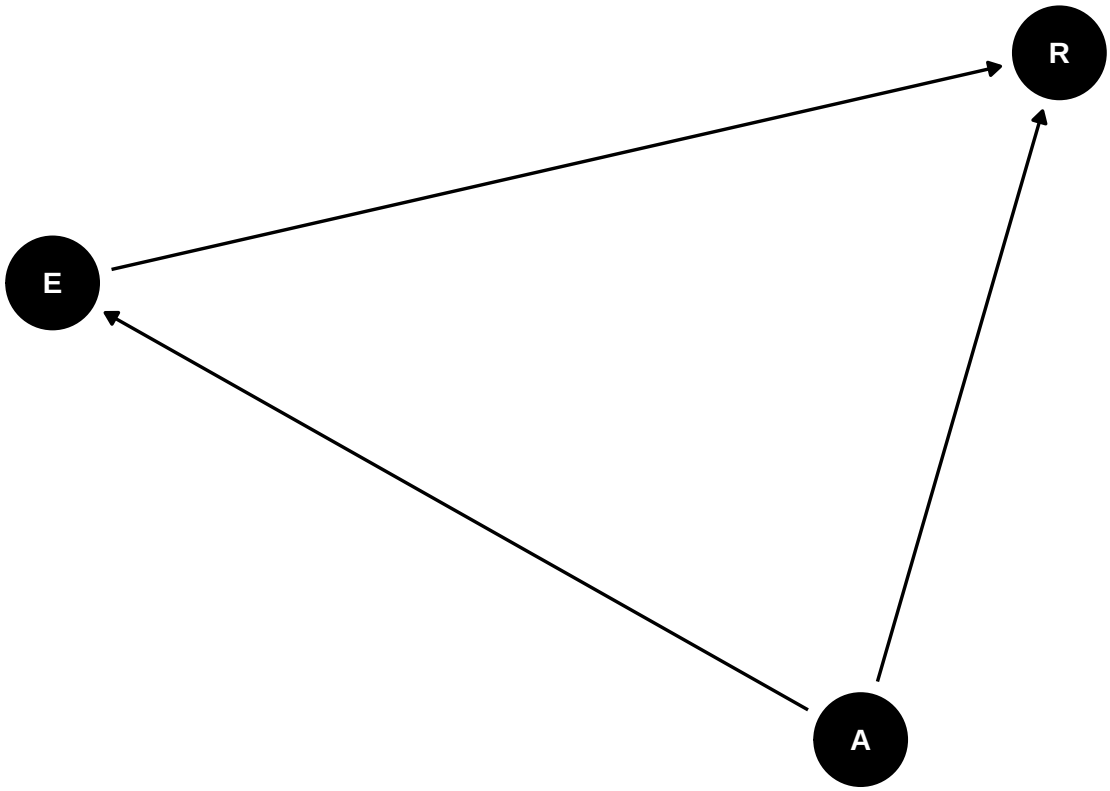
Homework 6

Question 1

In the Trolley data—data(Trolley)—we saw how education level (modeled as an ordered category) is associated with responses. Is this association causal? One plausible confound is that education is also associated with age, through a causal process: People are older when they finish school than when they begin it. Reconsider the Trolley data in this light. Draw a DAG that represents hypothetical or models do you need to evaluate the causal influence of education on responses? Fit these models to the trolley data. What do you conclude about the causal relationships among these three variables?

```
dag <- dagitty("dag{
                   A -> E
                   A -> R
                   E -> R
                }")
ggdag(dag) +
  theme_dag()
```



```
data(Trolley)

d <- Trolley

head(d)
```

```
##    case response order     id age male           edu action intention contact
## 1 cfaqu        4     2 96;434  14    0 Middle School      0         0       1
## 2 cfbur        3    31 96;434  14    0 Middle School      0         0       1
## 3 cfrub        4    16 96;434  14    0 Middle School      0         0       1
## 4 cibox        3    32 96;434  14    0 Middle School      0         1       1
## 5 cibur        3     4 96;434  14    0 Middle School      0         1       1
## 6 cispe        3     9 96;434  14    0 Middle School      0         1       1
##   story action2
## 1   aqu       1
## 2   bur       1
## 3   rub       1
## 4   box       1
## 5   bur       1
## 6   spe       1
```

```
data {

  int N; // Number of individuals
  int K; // Number of careers

  vector[N] A;
  vector[N] C;
  vector[N] I;
  vector[N] IA;
  vector[N] IC;

  int E[N];
  vector[N] AGE;

  int<lower=1,upper=K> response[N];
}

parameters {
  real bA;
  real bC;
  real bI;
  real bIA;
  real bIC;

  real bAGE;
  real bE;

  ordered[(K - 1)] q;
  simplex[7] delta_education;
}

model{
```

```
  vector[N] u;
  vector[8] education_effect;

  delta_education ~ dirichlet([2, 2, 2, 2, 2, 2, 2]');
  education_effect = cumulative_sum(append_row([0]', delta_education));

  bE ~ normal(0, 1);

  bA ~ normal(0, 1);
  bC ~ normal(0, 1);
  bI ~ normal(0, 1);

  bIA ~ normal(0, 1);
  bIC ~ normal(0, 1);
  bAGE ~ normal(0, 1);

  u = bE * education_effect[E] + bA * A + bC * C + bI * I + bIA * IA + bIC * IC + bAGE * AGE;

  response ~ ordered_logistic(u, q);

}
```

```
edu_levels <- c(6, 1, 8, 4, 7, 2, 5, 3)
d$edu_new <- edu_levels[d$edu]

dat_list1 <- list(
  response=d$response,
  A=d$action,
  I=d$intention,
  C=d$contact,
  IA=d$intention * d$action,
  IC=d$intention * d$contact,
  E=as.integer(d$edu_new),
  AGE=standardize(d$age),
  K=7,
  N=9930
)
```

Question 2

Consider one more variable in the Trolley data: Gender. Suppose that gender might influence education as well as response directly. Draw the DAG now that includes response, education, age, and gender. Using only the DAG, is it possible that the inferences from Problem 1 are confounded by gender? If so, define any additional models you need to infer the causal influence of education on response. What do you conclude?

```
library(dagitty)


dag2 <- dagitty("dag{
E -> R <- A
A -> E
G -> E
G -> R
```

```
}")

adjustmentSets( dag2 , exposure="E" , outcome="R" , effect="total" )
```

```
## { A, G }
```

```
data {

  int N; // Number of individuals
  int K; // Number of careers

  vector[N] A;
  vector[N] C;
  vector[N] I;
  vector[N] IA;
  vector[N] IC;

  vector[N] F;
  int E[N];
  vector[N] AGE;

  int<lower=1,upper=K> response[N];
}

parameters {
  real bA;
  real bC;
  real bI;
  real bIA;
  real bIC;

  real bAGE;
  real bE;
  real bF;

  ordered[(K - 1)] q;
  simplex[7] delta_education;
}

model{
  vector[N] u;
  vector[8] education_effect;

  delta_education ~ dirichlet([2, 2, 2, 2, 2, 2, 2]');
  education_effect = cumulative_sum(append_row([0]', delta_education));

  bE ~ normal(0, 1);
  bF ~ normal(0, 1);

  bA ~ normal(0, 1);
  bC ~ normal(0, 1);
  bI ~ normal(0, 1);
```

```
  bIA ~ normal(0, 1);
  bIC ~ normal(0, 1);
  bAGE ~ normal(0, 1);

  u = bE * education_effect[E] + bA * A + bC * C + bI * I + bIA * IA + bIC * IC + bAGE * AGE + bF * F;

  response ~ ordered_logistic(u, q);

}
```

```
edu_levels <- c(6, 1, 8, 4, 7, 2, 5, 3)
d$edu_new <- edu_levels[d$edu]

dat_list2 <- list(
  response=d$response,
  A=d$action,
  I=d$intention,
  C=d$contact,
  IA=d$intention * d$action,
  IC=d$intention * d$contact,
  E=as.integer(d$edu_new),
  F=1-d$male,
  AGE=standardize(d$age),
  K=7,
  N=9930
)
```

```
model_q2_1_fit <- sampling(model_q2_1, data=dat_list2, cores=4)
```