Stat 341 – Homework 3

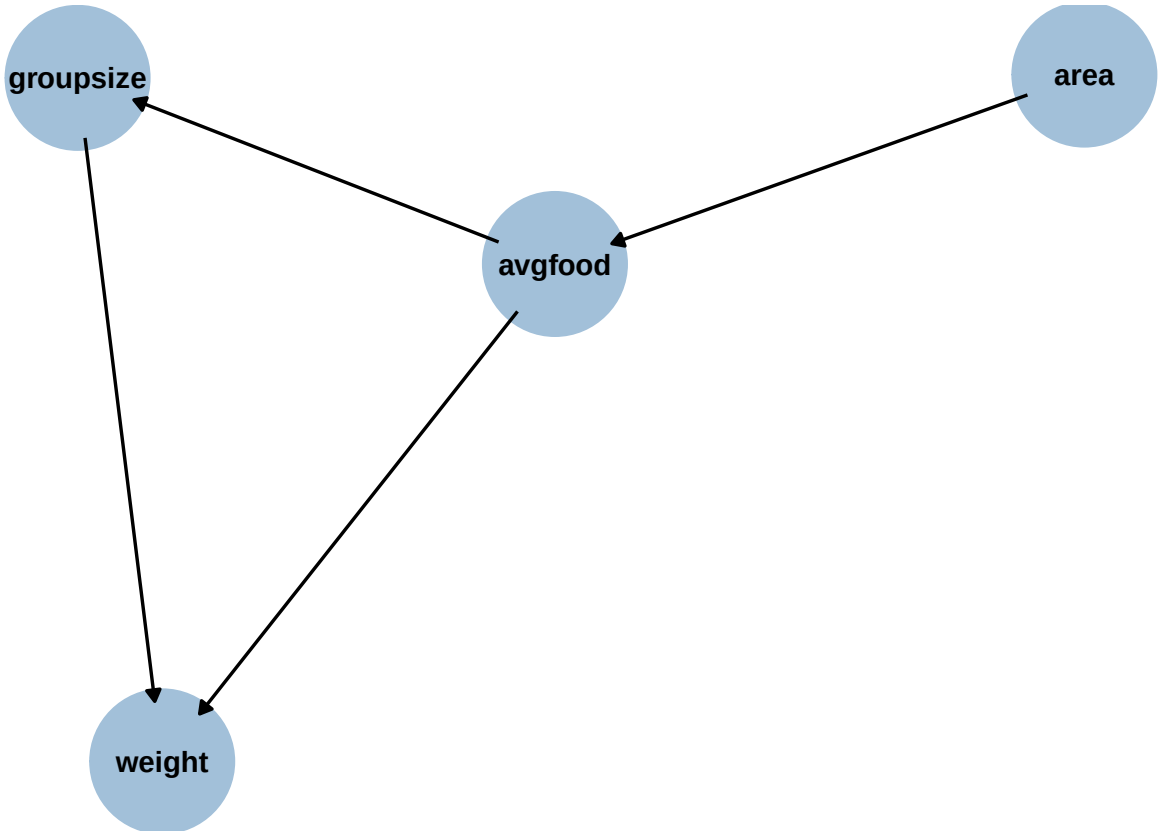
Gloria Grace

February 01, 2023

N1

```
# create the DAG object
causal_diagram <- dagitty("dag{
  area -> avgfood
  avgfood -> groupsize
  avgfood -> weight;
  groupsize -> weight;

}")
# plot it
gg_dag(causal_diagram,
       size = 25)
```



N2

A

avgfood: Mediator

groupsize: Moderator

B

Based on my answer for part A, if I were going to devise a model to estimate the effect of area on weight, I would include both avgfood and groupsize in the model.

N3

A

Article 4: "How many cases of HIV contaminated transmission caused by contaminated injections?"

B

Predictor: Contaminated Injections

Response: HIV-contaminated transmission

C

They included

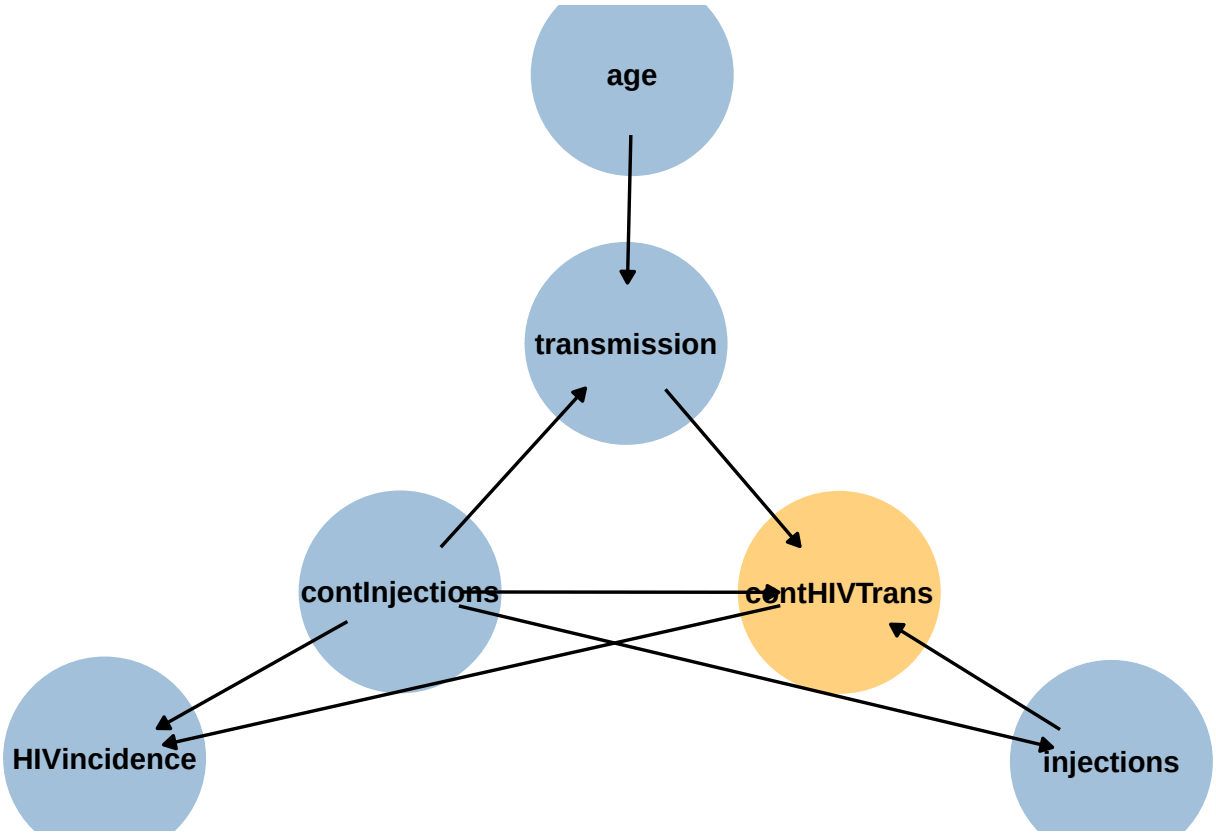
- age groups ()
- type of transmissions such as unsafe injections, unsafe transfusions, and mother-to-child transmission,
- HIV incidence,
- Type of equipment (needlestick, IDU, Deep wound, and romanian).

They also include HIV prevalence, Injection rates, and HIV infection status for the graph but not specifically to answer the question.

D (REVISE PLS)

```
HIV_dag <- dagitty("dag{
  contInjections -> injections
  HIVincidence <- contInjections
  injections -> contHIVTrans
  age -> transmission
  contInjections ->contHIVTrans
  transmission -> contHIVTrans
  contInjections -> transmission
  HIVincidence <- contHIVTrans

}")
gg_dag(HIV_dag,
       size = 35,
       highlight = 'contHIVTrans')
```



E

Injections:Mediator Transmission: Mediator Age: Moderator HIVincidence: Collider

F

Based on my answer for E, I would have not include HIVincidence because based on the DAG, HIV is the cause of both of the response and predictor variable. The contaminated Injection can cause the HIV incidence and contaminated HIV transmission could also result in HIV incidence.

N4

A

The model that I chose was movielens. Looking back to the proportion that I was estimating, the other factors that might influence the proportion of movies that are comedies would be the duration of the movie and the age restriction of the movie.

I think these two variables might be relevant because action movies tend to have long duration whereas comedy I would guess have shorter length. Also, for age restriction, there are a lot of movies that are comedies with family content and there are with higher age requirement so I think this variable might be relevant.

I am not really sure that you meant to add variables that are not include in the data or not, but if only restricted to only use the data and the variables in it, I would say the two variables could be the horror and for_kids.

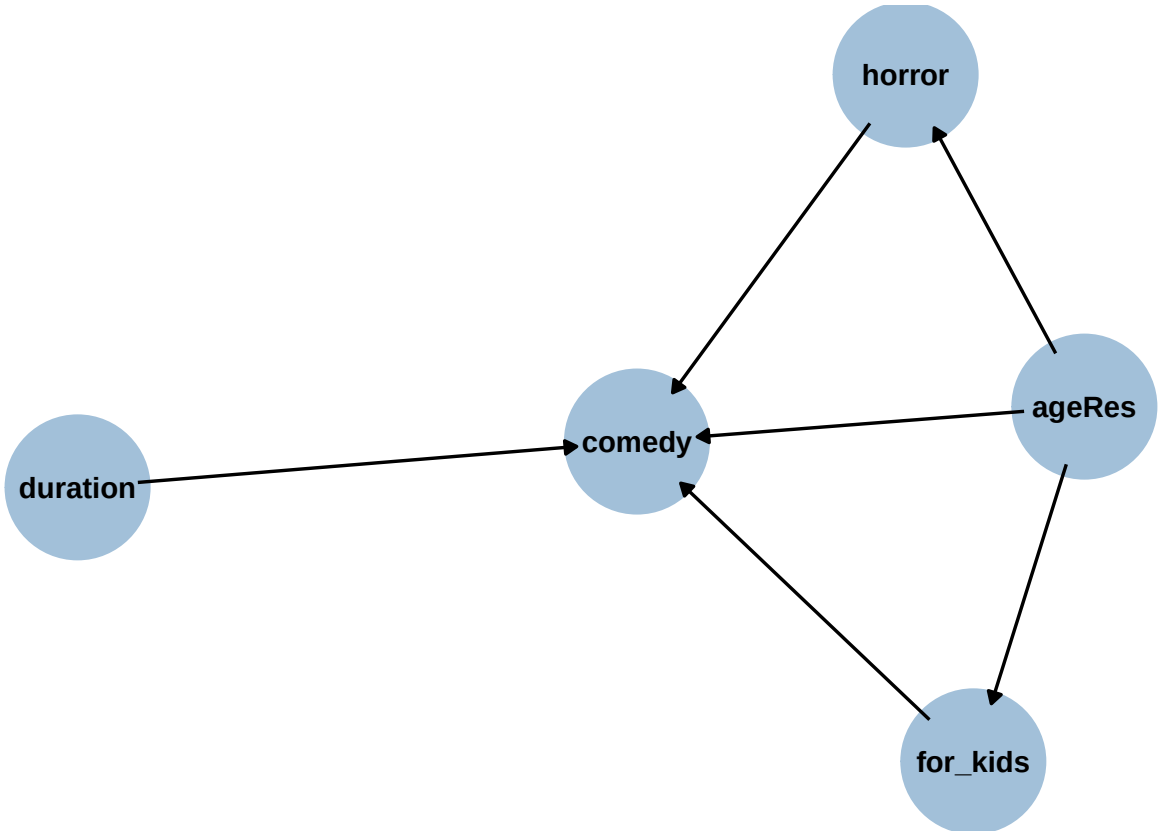
For these two variables, I think because most of horror movies is pretty intense, the chance of it being comedy is lower. For the for_kids variable, usually if it is a family friendly movie, the proportion of movie being

comedy would increase.

B

```
# create the DAG object
movie_dag <- dagitty("dag{
  duration -> comedy
  ageRes -> for_kids
  for_kids -> comedy;
  horror -> comedy;
  ageRes -> horror
  ageRes -> comedy

}")
# plot it
gg_dag(movie_dag,
       size = 25)
```



C

for_kids: Mediator Horror: Mediator Duration: Precision Covariate

D

Based on C, the two variables that I came up with for A, I should have include them in the model that I fitted.