Stat 341 – Homework 03

Luca Seazzu

April 11, 2023

N1

```
library(dagitty)
library(CalvinBayes)
```

```
## Loading required package: dplyr
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Loading required package: ggformula
```

```
## Loading required package: ggstance
```

```
##
## Attaching package: 'ggstance'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     GeomErrorbarh, geom_errorbarh
```

```
## Loading required package: scales
```

```
## Loading required package: ggridges
```

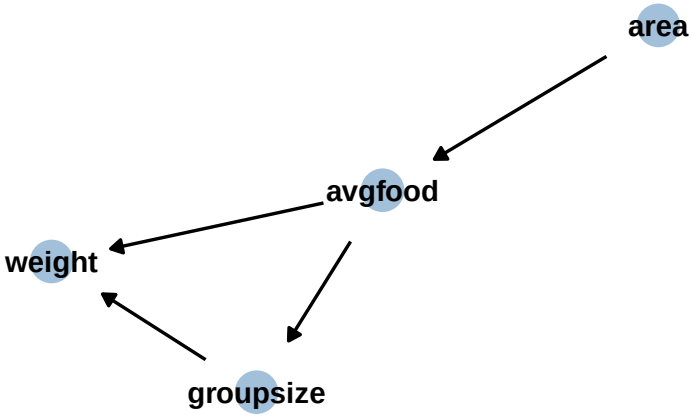
```
##
## Attaching package: 'ggridges'
```

```
## The following objects are masked from 'package:tidybayes':
##
##     scale_point_color_continuous, scale_point_color_discrete,
##     scale_point_colour_continuous, scale_point_colour_discrete,
##     scale_point_fill_continuous, scale_point_fill_discrete,
##     scale_point_size_continuous
```

```
##
## New to ggformula?  Try the tutorials:
##  learnr::run_tutorial("introduction", package = "ggformula")
##  learnr::run_tutorial("refining", package = "ggformula")
```

```
##
## Attaching package: 'CalvinBayes'
## The following object is masked from 'package:bayesplot':
##
##     rhat
## The following objects are masked from 'package:tidybayes':
##
##     hdi, rstudent_t
## The following object is masked from 'package:datasets':
##
##     HairEyeColor
```

```
# create the DAG object
causal_diagram_1 <- dagitty("dag{
  area -> avgfood -> weight;
  avgfood -> groupsize -> weight
}")
# plot it
gg_dag(causal_diagram_1)
```



N2

A. avgfood is a mediator variable between area and weight. groupsize is a mediator variable between avgfood and weight.

B. I would include both avgfood and groupsize variables since I don't think either are collider variables so I think they are helpful.

N3

A. How does age and sex of a countries' current population project their future population?

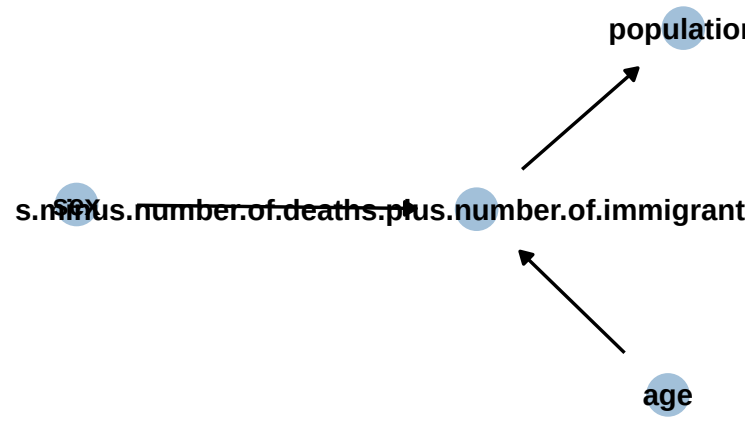
B. age and sex are the predicting variables for population, the response variable.

C. current population, number of births, number of deaths, number of immigrants, and number of emigrants.

D.

```
# create the DAG object
causal_diagram_3 <- dagitty("dag{
```

```
  age -> number.of.births.minus.number.of.deaths.plus.number.of.immigrants.minus.number.of.emigrants ->
  sex ->  number.of.births.minus.number.of.deaths.plus.number.of.immigrants.minus.number.of.emigrants ->
}")
# plot it
gg_dag(causal_diagram_3)
```



E. the calculations for population based on births, deaths, immigrants, and emigrants is a mediator variable

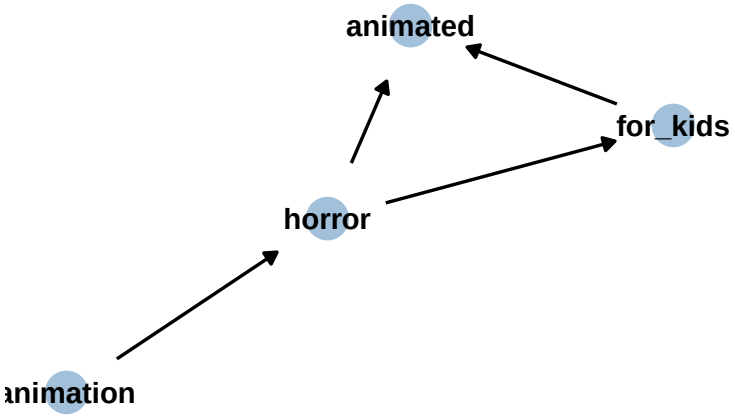
F. Seems like the researchers model was based around either age or sex and then modified by their population calculation and then modeled so the causal diagram should be consistent.

N4

A. After fitting a model in homework 2 to look at if a movie is animated or not, I would think that if the genre is or isn't horror and if the movie was for kids would both be variables relevant to the proportion I was estimating. I think more kids movies and movies that aren't horror movies would be relevant to if the movie was animated or not.

B.

```
# create the DAG object
causal_diagram_2 <- dagitty("dag{
  animation -> horror -> for_kids -> animated;
  horror -> animated
}")
# plot it
gg_dag(causal_diagram_2)
```



C. for_kids is a mediator variable and horror is a confounder variable.

D. I think both for_kids and horror should have been included in the model fitted in homework 02.