Assignment 2 - Methods 4

Frederik Laursen Nielsen (FN), Halfdan Nordahl Fundal (HF) & Mate Schuszter (MS)

2025-04-06

```
library(dagitty)
library(rethinking)
library(ggplot2)
```

Second assignment

The second assignment uses chapter 3, 5 and 6. The focus of the assignment is getting an understanding of causality.

Chapter 3: Causal Confussion

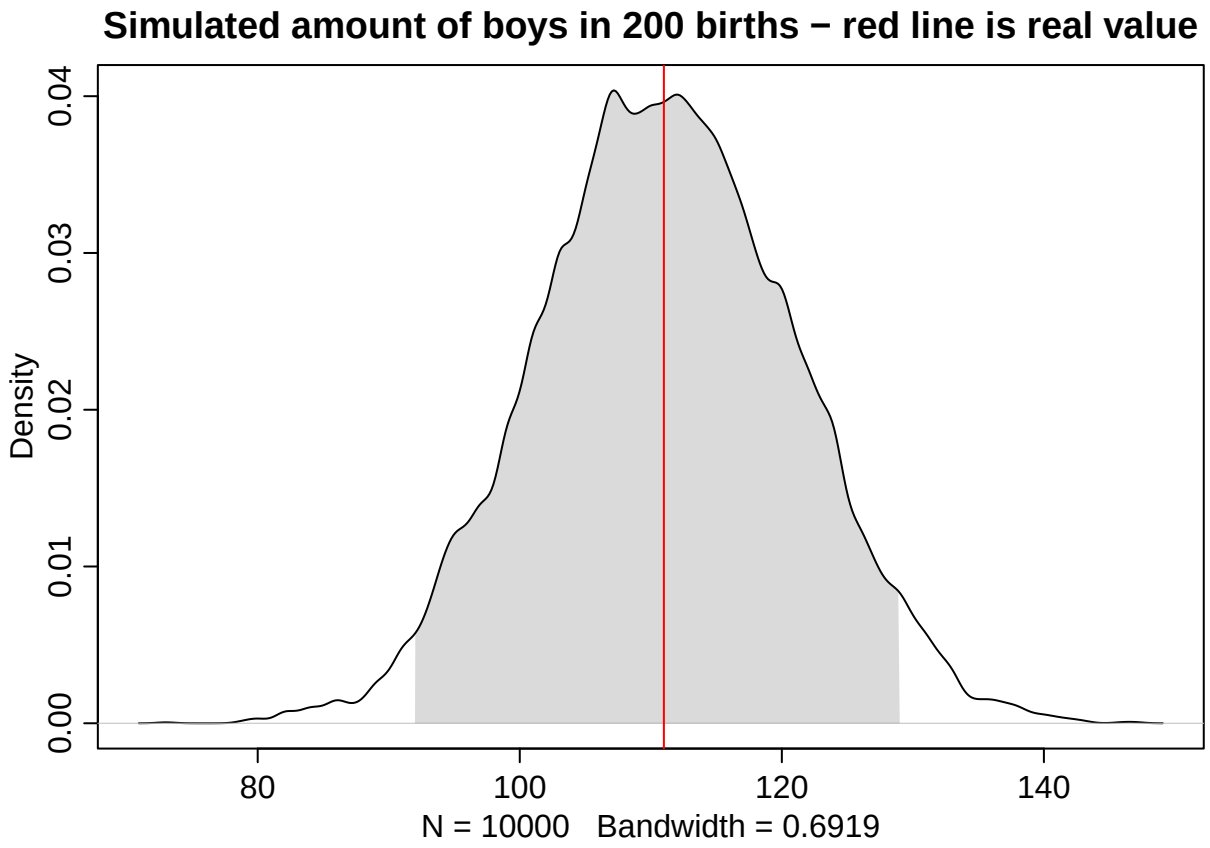
Section author: Frederik Laursen Nielsen (FN)

Reminder: We are tying to estimate the probability of giving birth to a boy I have pasted a working solution to questions 6.1-6.3 so you can continue from here:)

3H3 (FN) Use rbinom to simulate 10,000 replicates of 200 births. You should end up with 10,000 numbers, each one a count of boys out of 200 births. Compare the distribution of predicted numbers of boys to the actual count in the data (111 boys out of 200 births).

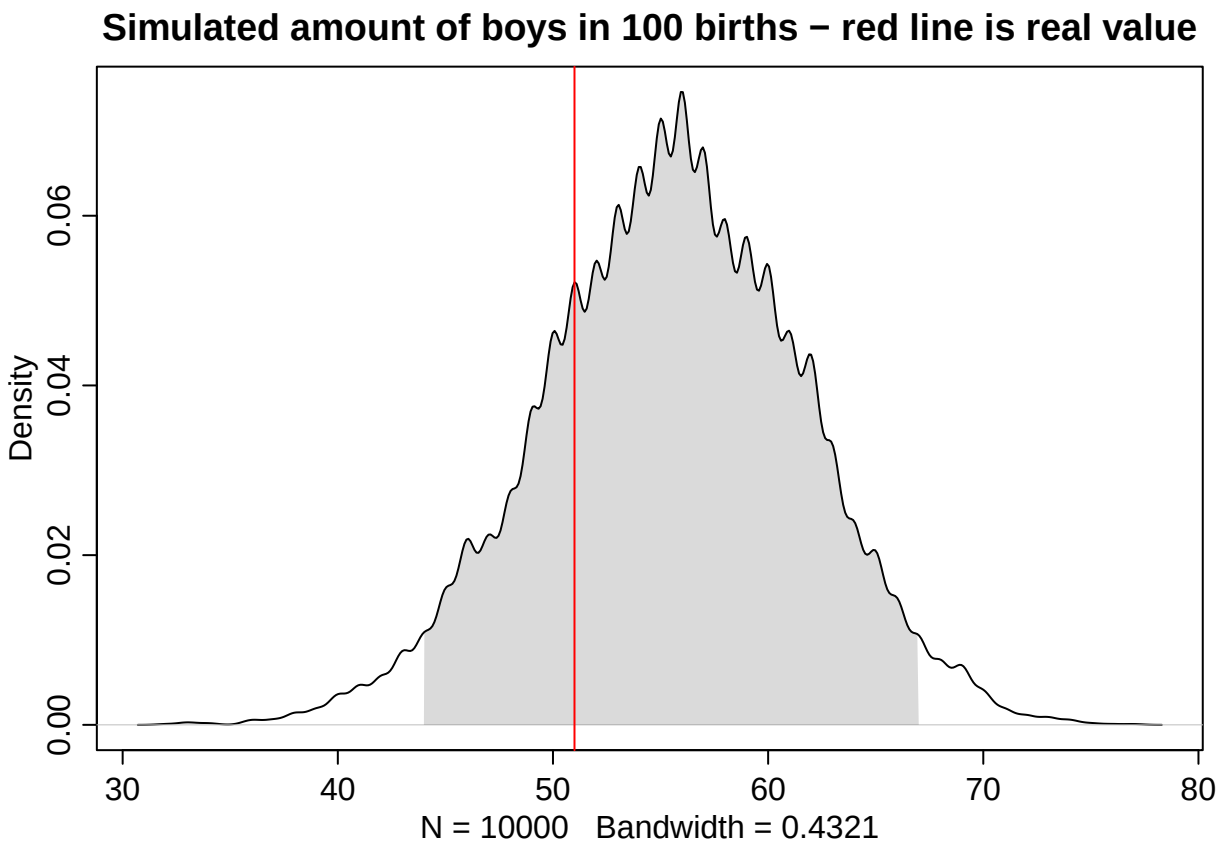
```
# 3H1
# Find the posterior probability of giving birth to a boy:
data(homeworkch3)
set.seed(1)
W <- sum(birth1) + sum(birth2)
N <- length(birth1) + length(birth2)
p_grid <-seq(from =0, to = 1, len =1000)
prob_p <- rep(1,1000)
prob_data <- dbinom(W,N,prob=p_grid)
posterior <-prob_data * prob_p
posterior <- posterior / sum(posterior)

# 3H2
# Sample probabilities from posterior distribution:
samples <- sample (p_grid, prob = posterior, size =1e4, replace =TRUE)


# 3H3
# Simulate births using sampled probabilities as simulation input, and check if they allign with real v
```

```
simulated_births <- rbinom(n = 1e4, size = N, prob = samples)
rethinking::dens(simulated_births,show.HPDI = 0.95)
abline(v=W, col="red")
title("Simulated amount of boys in 200 births - red line is real value")
```



3H4. (FN) Now compare 10,000 counts of boys from 100 simulated first borns only to the number of boys in the first births, birth1. How does the model look in this light?

```
simulated_births2 <- rbinom(n = 1e4, size = 100, prob = samples)

rethinking::dens(simulated_births2,show.HPDI = 0.95)
abline(v=sum(birth1), col="red")
title("Simulated amount of boys in 100 births - red line is real value")
```

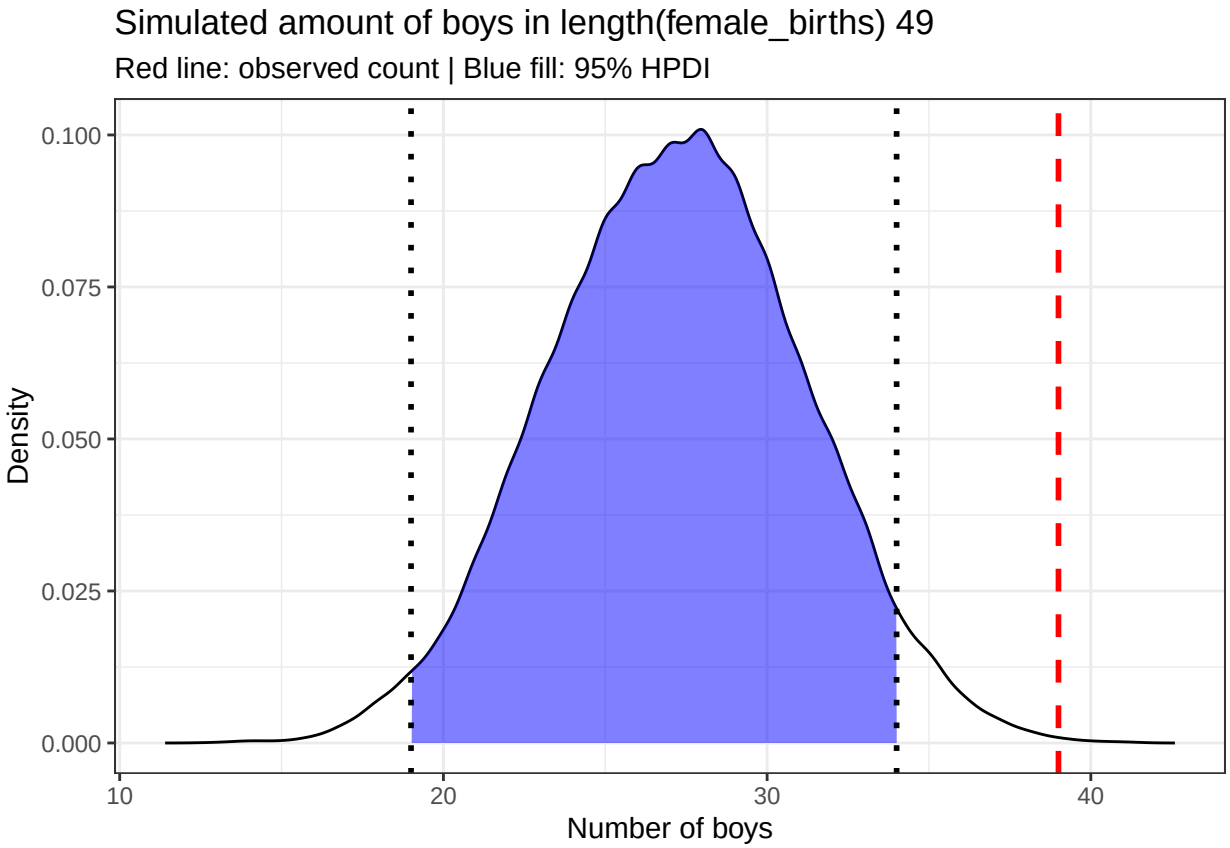


Answer: In this light it becomes apparent that the model is slightly inaccurate as the real value now lies further away from the mean value of the simulated distribution.

3H5. (FN) The model assumes that sex of first and second births are independent. To check this assumption, focus now on second births that followed female first borns. Compare 10,000 simulated counts of boys to only those second births that followed girls. To do this correctly, you need to count the number of first borns who were girls and simulate that many births, 10,000 times. Compare the counts of boys in your simulations to the actual observed count of boys following girls. How does the model look in this light? Any guesses what is going on in these data?

```
set.seed(1)
female_births <- which(birth1 == 0)
followed_female <- birth2[female_births]
simulated_3H5 <- rbinom(n = 1e4, size = length(female_births), prob = samples)

# Compute HPDI
hpdi_values <- HPDI(simulated_3H5, prob = 0.95)

df <- data.frame(simulated_3H5)

dens <- density(simulated_3H5)
dens_df <- data.frame(x = dens$x, y = dens$y)

# Filter density data
dens_df_hpdi <- subset(dens_df, x >= hpdi_values[1] & x <= hpdi_values[2])

ggplot() +
```

```r
  geom_line(data = dens_df, aes(x = x, y = y), color = "black") +  # Density line
  geom_area(data = dens_df_hpdi, aes(x = x, y = y), fill = "blue", alpha = 0.5) +  # Fill only HPDI reg
  geom_vline(xintercept = sum(followed_female), color = "red", linetype = "dashed", lwd=1) +
  geom_vline(xintercept = hpdi_values[1], color = "black", linetype = "dotted", lwd=1) +
  geom_vline(xintercept = hpdi_values[2], color = "black", linetype = "dotted", lwd=1) +
  labs(
    title = "Simulated amount of boys in length(female_births) 49",
    subtitle = "Red line: observed count | Blue fill: 95% HPDI",
    x = "Number of boys",
    y = "Density"
  ) +
  theme_bw()
```



Answer: Seemingly there are more boy births (39) following first born females in the real data set than the simulated amount would have us believe. As can be seen from the plot, the actual number of boys following female births lie well outside the 95% HPDI

Chapter 5: Spurrious Correlations

Section author: Máté Schuszter (MS) Start of by checking out all the spurrious correlations that exists in the world. Some of these can be seen on this wonderfull website: https://www.tylervigen.com/spurious/random All the medium questions are only asking you to explain a solution with words, but feel free to simulate the data and prove the concepts.

5M1. (MS) Invent your own example of a spurious correlation. An outcome variable should be correlated

with both predictor variables. But when both predictors are entered in the same model, the correlation between the outcome and one of the predictors should mostly vanish (or at least be greatly reduced).

5M1 Answer One can imagine the number of art galleries in an area (outcome) to be correlated with the rate of children going to private high schools in the same area (predictor 1) and median household income in the same area (predictor 2). However it is to be expected, that when both are included in the model, private schooling would have little effect remaining on the number of art galleries.

5M2. (MS) Invent your own example of a masked relationship. An outcome variable should be correlated with both predictor variables, but in opposite directions. And the two predictor variables should be correlated with one another.

5M1 Answer An example for a relationship of such kind could be the following. Life expectancy (outcome) is positively correlated with strength training (predictor 1) and negatively correlated with steroid use (predictor 2), while we might also expect a correlation between strength training and steroid use.

5M3. (MS) It is sometimes observed that the best predictor of fire risk is the presence of firefighters—States and localities with many firefighters also have more fires. Presumably firefighters do not cause fires. Nevertheless, this is not a spurious correlation. Instead fires cause firefighters. Consider the same reversal of causal inference in the context of the divorce and marriage data. How might a high divorce rate cause a higher marriage rate? Can you think of a way to evaluate this relationship, using multiple regression

5M3 Answer First, a divorce allows one individual to remarry, thus contributing to the marriage rate twice, so partially divorces could "cause" some marriages. Second, considering a confounding variable, like median age at marriage, might permit a more detailed understanding of this relationship. Median age at marriage has a positive relationship with marriage rate, since more of the population is alive, and able to get married at a younger age, however also potentially negatively correlated with divorce rate, for a number of reasons, like uneducated choice of partner or the simple fact that there is more time to divorce for a couple for whatever reason.

5M5. (MS) One way to reason through multiple causation hypotheses is to imagine detailed mechanisms through which predictor variables may influence outcomes. For example, it is sometimes argued that the price of gasoline (predictor variable) is positively associated with lower obesity rates (outcome variable). However, there are at least two important mechanisms by which the price of gas could reduce obesity. First, it could lead to less driving and therefore more exercise. Second, it could lead to less driving, which leads to less eating out, which leads to less consumption of huge restaurant meals. Can you outline one or more multiple regressions that address these two mechanisms? Assume you can have any predictor data you need.

5M5 Answer Higher gasoline prices might also be associated with a general decline / recession in the country's economy, which is potentially also associated with high inflation, increasing prices on all items, food included. This can lead individuals to spend less on going out (even if it does not involve the car), to spend less money on unhealthy food delicacies and in extreme cases on food in general.

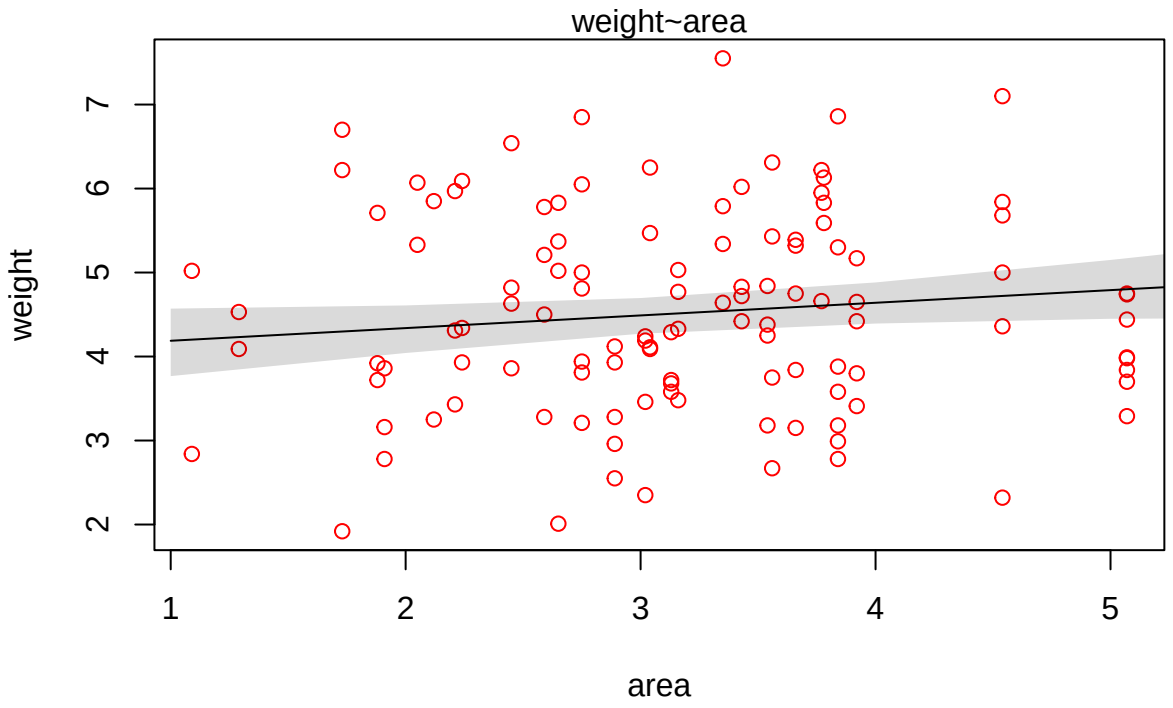
Chapter 5: Foxes and Pack Sizes

All five exercises below use the same data, data(foxes) (part of rethinking).84 The urban fox (Vulpes vulpes) is a successful exploiter of human habitat. Since urban foxes move in packs and defend territories, data on habitat quality and population density is also included. The data frame has five columns: (1) group: Number of the social group the individual fox belongs to (2) avgfood: The average amount of food available in the territory (3) groupsize: The number of foxes in the social group (4) area: Size of the territory (5) weight: Body weight of the individual fox

5H1. (MS) Fit two bivariate Gaussian regressions, using quap: (1) body weight as a linear function of territory size (area), and (2) body weight as a linear function of groupsize. Plot the results of these regressions, displaying the MAP regression line and the 95% interval of the mean. Is either variable important for predicting fox body weight?

```r
#loading data
pacman::p_load(rethinking)
data(foxes)
```

```r
#model 1. weight ~ area

#fitting model
m_1 <- quap(
  alist(
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b * (area), #the linear model
    a ~ dnorm(3, 1), #prior for alpha
    b ~ dlnorm(0, 1), #prior for beta, a positive relationship is to be expected, thus the prior does n
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)

#getting predictions from the model
#a sequence of area sizes to get predictions
sim_area <- seq(from = 1, to = 10, by = 1)

#getting the mean prediction and the uncertainty around it for the regression line
mu_1 <- link(m_1, data=data.frame(area = sim_area))
mu_mean <- apply(mu_1, 2, mean)
mu_PI <- apply(mu_1, 2, PI, prob = 0.95)


#plotting
#raw data
plot( weight ~ area, data=foxes, col="red" )
mtext("weight~area" )

#MAP regression line
lines(sim_area, mu_mean)

#PI for regression line
shade(mu_PI, sim_area)
```




```
#model 2 weight ~ groupsize

#average group size for foxes
avg_gsize <- mean(foxes$groupsize)

avg_gsize
```

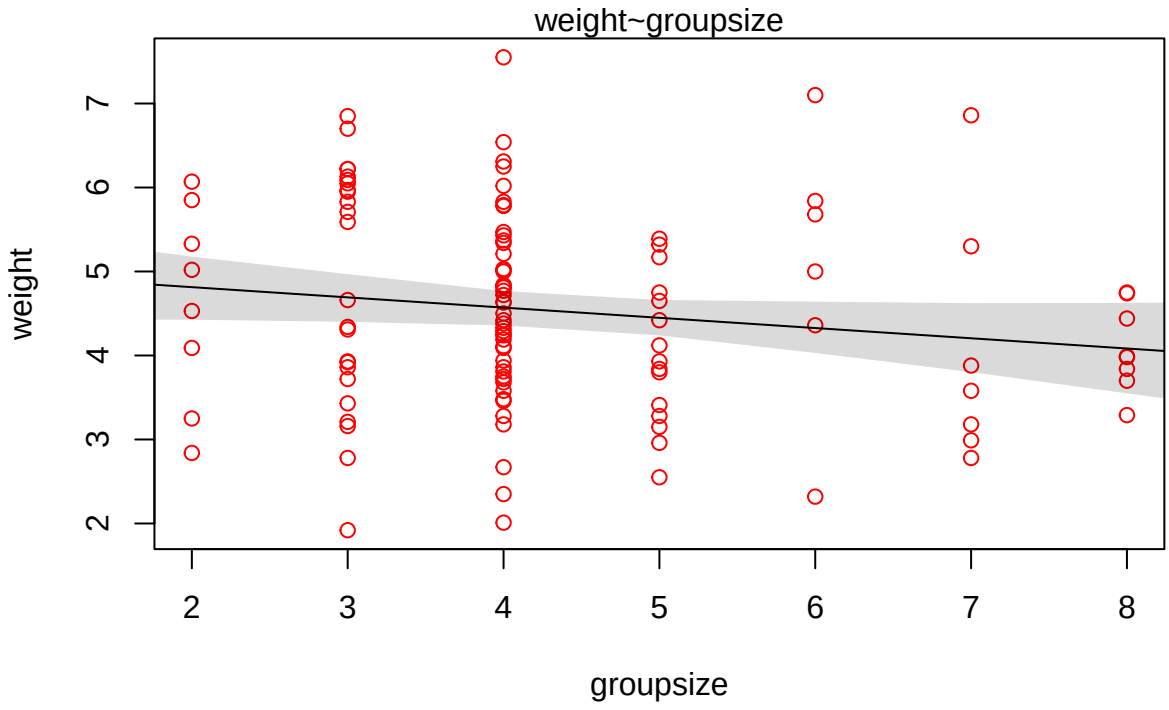
```
## [1] 4.344828
```

```
max(foxes$groupsize)
```

```
## [1] 8
```

```
#fitting model
m_2 <- quap(
  alist(
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b * (groupsize),  #the linear model
    a ~ dnorm(5, 2), #prior for alpha
    b ~ dnorm(0, 1), #prior for beta
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)
```

```r
#getting predictions from the model
#a sequence to get predictions
sim_gsize <- seq(from = 1, to = 20, by = 1)

#getting the mean prediction and the uncertainty around it for the regression line
mu_2 <- link(m_2, data=data.frame(groupsize = sim_gsize))
mu_mean2 <- apply(mu_2, 2, mean)
mu_PI2 <- apply(mu_2, 2, PI, prob = 0.95)


#plotting
#raw data
plot( weight ~ groupsize, data=foxes, col="red" )
mtext("weight~groupsize" )

#MAP regression line
lines(sim_gsize, mu_mean2)

#PI for regression line
shade(mu_PI2, sim_gsize)
```



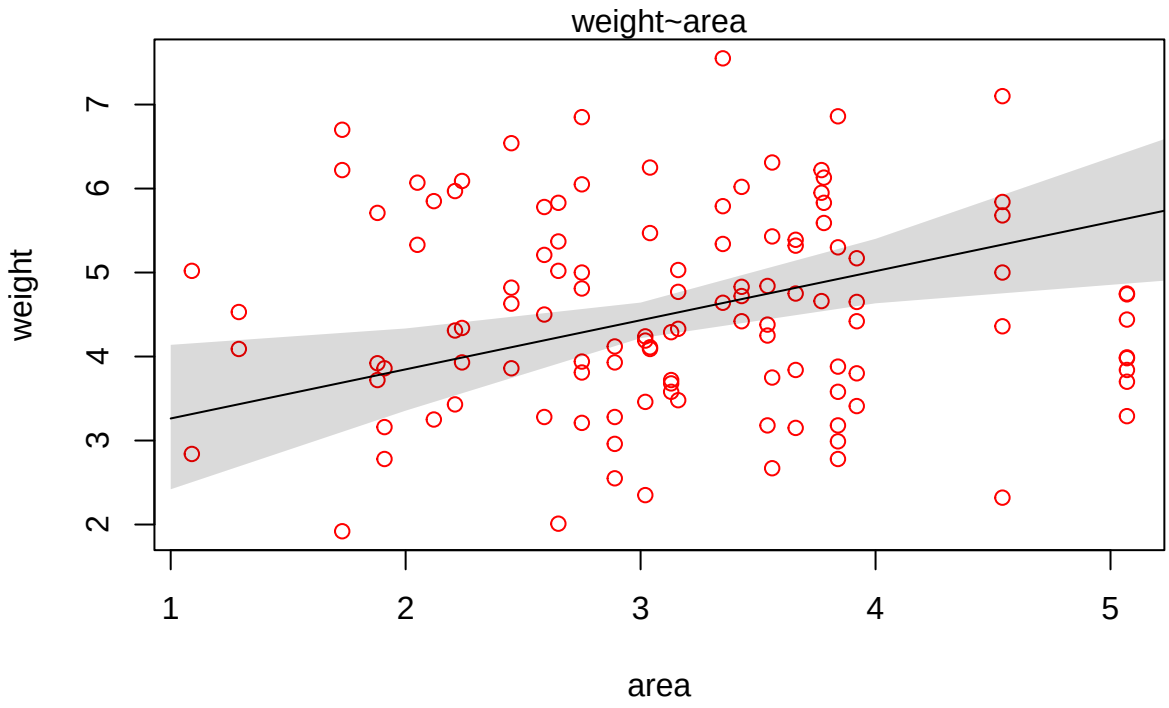
5H1 Answer: There seems to be a slight positive relationship between area size and weight and a negative relationship between groupsize and weight. Visual inspection does not suggest that either one is an important predictor of fox body weight.

5H2. (MS) Now fit a multiple linear regression with weight as the outcome and both area and groupsize as

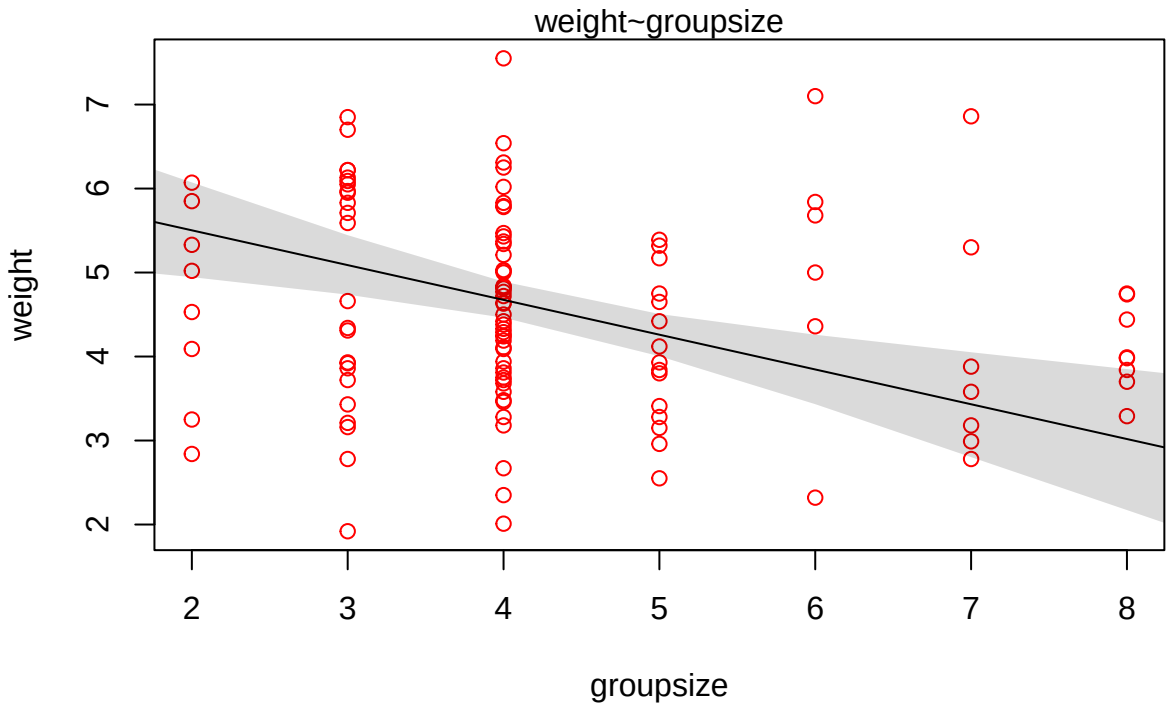
predictor variables. Plot the predictions of the model for each predictor, holding the other predictor constant at its mean. What does this model say about the importance of each variable? Why do you get different results than you got in the exercise just above?

```
#model 3 weight ~ groupsize + area

#fitting model
m_3 <- quap(
  alist(
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b_g * groupsize + b_a * area, #the linear model
    a ~ dnorm(5, 2), #prior for alpha
    b_g ~ dnorm(0, 1), #prior for coefficient for groupsize
    b_a ~ dnorm(0, 1), #prior for coefficient for area
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)
```

```
#getting predictions from the model

#getting the mean prediction and the uncertainty around it for the regression line
mu_32 <- link(m_3, data=data.frame(area = sim_area, groupsize = mean(foxes$groupsize)))
mu_mean32 <- apply(mu_32, 2, mean)
mu_PI32 <- apply(mu_32, 2, PI, prob = 0.95)


#plotting
#raw data
plot( weight ~ area, data=foxes, col="red" )
mtext("weight~area" )

#MAP regression line
lines(sim_area, mu_mean32)

#PI for regression line
shade(mu_PI32, sim_area)
```




```
#predictions

#a sequence to get predictions for groupsize
sim_gsize <- seq(from = 1, to = 20, by = 1)

#getting the mean prediction and the uncertainty around it for the regression line
mu_3 <- link(m_3, data=data.frame(groupsize = sim_gsize, area = mean(foxes$area)))
mu_mean3 <- apply(mu_3, 2, mean)
mu_PI3 <- apply(mu_3, 2, PI, prob = 0.95)


#plotting
#raw data
plot( weight ~ groupsize, data=foxes, col="red" )
mtext("weight~groupsize" )

#MAP regression line
lines(sim_gsize, mu_mean3)

#PI for regression line
shade(mu_PI3, sim_gsize)
```



5H2 Answer: The model including both predictors seems to unveil a masked relationship. Groupsize is negatively correlated with weight, while area is positively correlated with weight. However the two are also correlated, at bigger areas there are usually groups with more foxes. These two opposite effects make it hard to understand relationships of one with weight if the other is not controlled for.

5H3. (MS) Finally, consider the avgfood variable. Fit two more multiple regressions: (1) body weight as an additive function of avgfood and groupsize, and (2) body weight as an additive function of all three variables, avgfood and groupsize and area. Compare the results of these models to the previous models you've fit, in the first two exercises. (a) Is avgfood or area a better predictor of body weight? If you had to choose one or the other to include in a model, which would it be? Support your assessment with any tables or plots you choose. (b) When both avgfood or area are in the same model, their effects are reduced (closer to zero) and their standard errors are larger than when they are included in separate models. Can you explain this result?

5H3 Answer: Which predictor to include depends largely on our research question. The DAG of the following question however sheds light to the causal relations between them. If we include area as a predictor, avgfood acts as a pipe, while if we include avgfood the effect of area is implicit in it. Considering that area has potentially little other causal paths to having an influence on weight, other than through avgfood, I would choose food as a predictor. But if we are interested in investigating these other paths, area should be taken into account and more work is necessary.

Since avgfood and area are correlated, including both of them in the model will cause their effects to be smaller and less certain.

```
#model 4 weight ~ groupsize + avgfood

m_4 <- quap(
  alist(
```

```
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b_g * groupsize + b_f * avgfood, #the linear model
    a ~ dnorm(5, 2), #prior for alpha
    b_g ~ dnorm(0, 1), #prior for coefficient for groupsize
    b_f ~ dnorm(0, 1), #prior for coefficient for food
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)

#model 5 weight ~ groupsize + area + avgfood

m_5 <- quap(
  alist(
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b_g * groupsize + b_a * area + b_f * avgfood, #the linear model
    a ~ dnorm(5, 2), #prior for alpha
    b_g ~ dnorm(0, 1), #prior for coefficient for groupsize
    b_a ~ dnorm(0, 1), #prior for coefficient for area
    b_f ~ dnorm(0, 1), #prior for coefficient for food
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)

precis(m_4)
```

```
##             mean         sd       5.5%     94.5%
## a      4.7021116 0.36524441  4.1183804  5.285843
## b_g   -0.2969023 0.11359656 -0.4784515 -0.115353
## b_f    1.4877264 0.77942080  0.2420614  2.733391
## sigma  1.1343680 0.07539488  1.0138724  1.254864
```

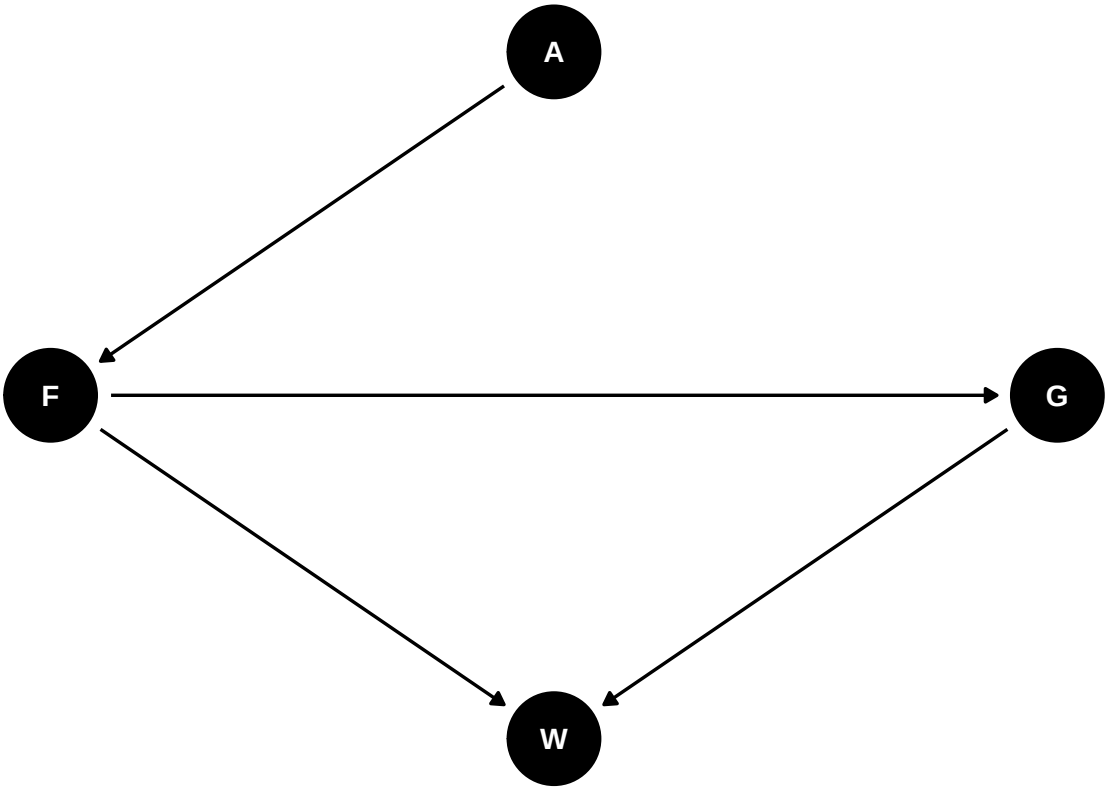
```
precis(m_5)
```

```
##             mean         sd       5.5%     94.5%
## a      4.3695774 0.38180012  3.7593870  4.979768
## b_g   -0.4711177 0.13125675 -0.6808913 -0.261344
## b_a    0.5088306 0.20746329  0.1772642  0.840397
## b_f    0.7930120 0.82306935 -0.5224118  2.108436
## sigma  1.1111223 0.07324251  0.9940666  1.228178
```

Defining our theory with explicit DAGs Assume this DAG as an causal explanation of fox weight:

```
pacman::p_load(dagitty,
               ggdag)
dag <- dagitty('dag {
A[pos="1.000,0.500"]
F[pos="0.000,0.000"]
G[pos="2.000,0.000"]
W[pos="1.000,-0.500"]
A -> F
F -> G
F -> W
```

```
G -> W
}')

# Plot the DAG
ggdag(dag, layout = "circle")+
  theme_dag()
```



where A is area, F is avgfood,G is groupsize, and W is weight.

Using what you know about DAGs from chapter 5 and 6, solve the following three questions: (MS)

1) Estimate the total causal influence of A on F. What effect would increasing the area of a territory have on the amount of food inside of it?

```
data("foxes")

#model
model61 <- quap(
  alist(
    avgfood ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b * area, #the linear model
    a ~ dnorm(3, 1), #prior for alpha
    b ~ dnorm(0, 1), #prior for coefficient for area
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
```

```
data = foxes)

#estimate and plots
print(precis(model61))
```

```
##             mean          sd       5.5%     94.5%
## a     0.15660492 0.030710037 0.10752435 0.2056855
## b     0.18785142 0.009302987 0.17298345 0.2027194
## sigma 0.09265418 0.006081016 0.08293554 0.1023728
```

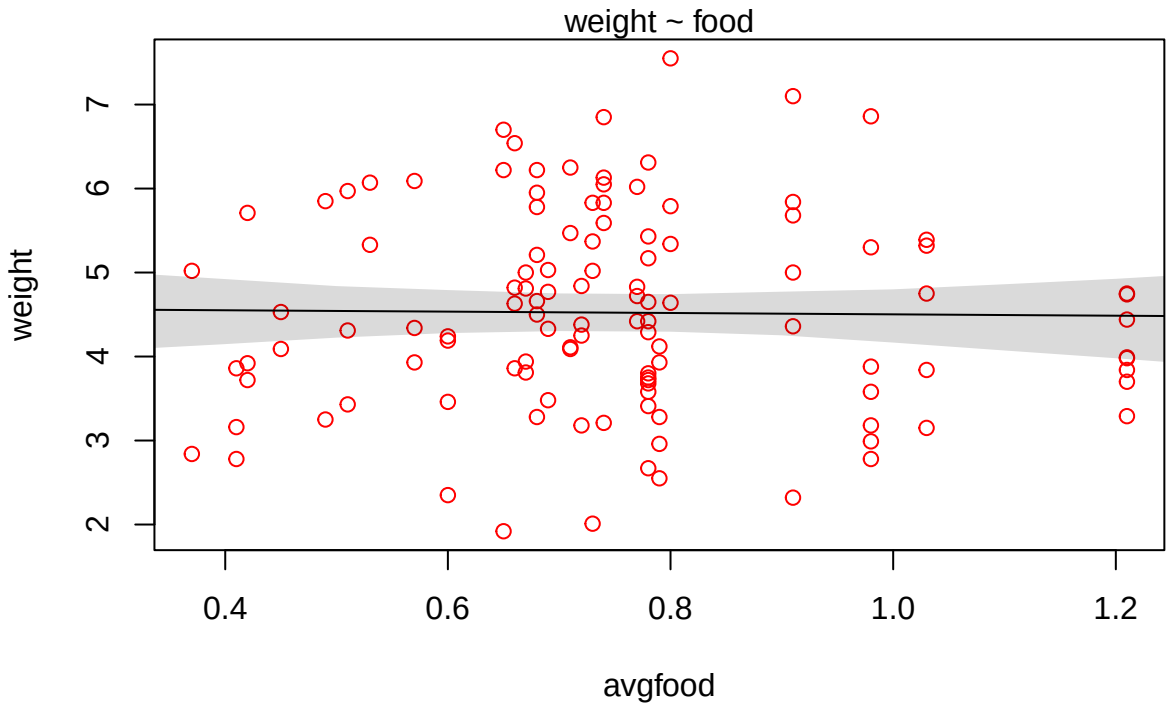
Answer: Observing the dag, there is one causal pathway from area size to food. The model suggests a clear positive relationship. The posterior estimate for beta, the effect of area on food has a mean of 0.19, with a sd of 0.01. This estimate would suggest that a one unit increase in area would result in a 0.19 unit increase in (average) food.

2) Infer the **total** causal effect of adding food F to a territory on the weight W of foxes. Can you calculate the causal effect by simulating an intervention on food?

```
#model
model62 <- quap(
  alist(
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b * avgfood, #the linear model
    a ~ dnorm(4, 2), #prior for alpha
    b ~ dnorm(0, 1), #prior for coefficient for food, we expect to see a positive relationship
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)

precis(model62)
```

```
##             mean        sd       5.5%    94.5%
## a      4.5929376 0.3737882  3.9955519 5.190323
## b     -0.0863178 0.4769416 -0.8485626 0.675927
## sigma  1.1785943 0.0773837  1.0549202 1.302268
```

```
#the simulated (average) food for estimates
sim_food <- seq(from = 0, to = 2, by = 0.1)

#getting the mean prediction and the uncertainty around it for the regression line
mu_62 <- link(model62, data=data.frame(avgfood = sim_food))
mu_mean62 <- apply(mu_62, 2, mean)
mu_PI62 <- apply(mu_62, 2, PI, prob = 0.95)

#raw data
plot( weight ~ avgfood, data=foxes, col="red" )
mtext("weight ~ food" )

#MAP regression line
lines(sim_food, mu_mean62)

#PI for regression line
shade(mu_PI62, sim_food)
```



Answer: There are two paths leading from F to W. To infer the total causal effect we have to consider both. Therefore I do not include any other controls in the model.

The regression line appears flat, there seems to be no relationship between food and weight. This is of course counter intuitive, rather we should consider the confounding variable, groupsize.

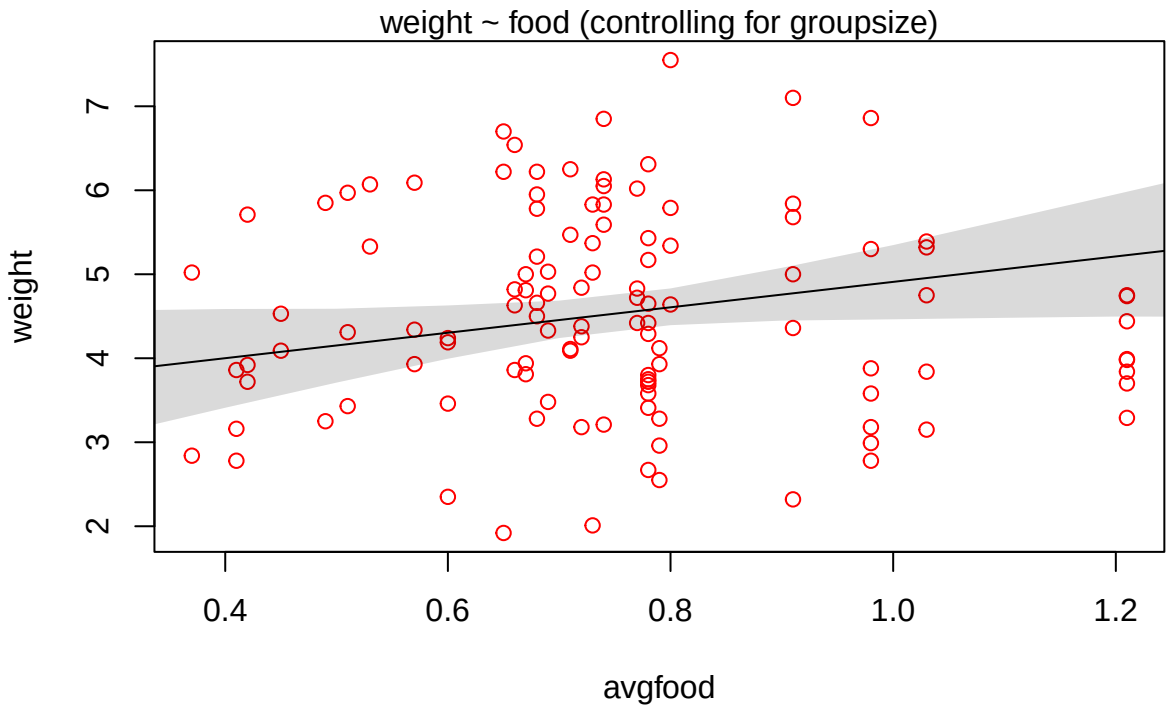
3) Infer the **direct** causal effect of adding food F to a territory on the weight W of foxes. In light of your estimates from this problem and the previous one, what do you think is going on with these foxes?

```
#model
model63 <- quap(
  alist(
    weight ~ dnorm(mu, sigma), #the likelihood
    mu <- a + b_f * avgfood + b_g *groupsize, #the linear model
    a ~ dnorm(4, 2), #prior for alpha
    b_f ~ dnorm(0, 1), #prior for coefficient for food, we expect to see a positive relationship
    b_g ~ dnorm(0, 1), #prior for groupsize
    sigma ~ dunif(0, 50) #uniform prior for sigma 0-50
  ) ,
data = foxes)

precis(model63)
```

```
##            mean         sd       5.5%      94.5%
## a      4.6685496 0.36482787  4.0854842  5.2516150
## b_f    1.5255026 0.77898911  0.2805276  2.7704777
```

```
## b_g   -0.2963297 0.11361702 -0.4779117 -0.1147478
## sigma  1.1338170 0.07533291  1.0134204  1.2542135
```

```r
#the simulated (average) food for estimates
sim_food <- seq(from = 0, to = 2, by = 0.1)

#getting the mean prediction and the uncertainty around it for the regression line
mu_63 <- link(model63, data=data.frame(avgfood = sim_food, groupsize = mean(foxes$groupsize)))
mu_mean63 <- apply(mu_63, 2, mean)
mu_PI63 <- apply(mu_63, 2, PI, prob = 0.95)

#raw data
plot( weight ~ avgfood, data=foxes, col="red" )
mtext("weight ~ food (controlling for groupsize)" )

#MAP regression line
lines(sim_food, mu_mean63)

#PI for regression line
shade(mu_PI63, sim_food)
```



Answer: To get the direct affect, we have to close the pipe going through groupsize. We do this by controlling for groupsize, including it in the model.

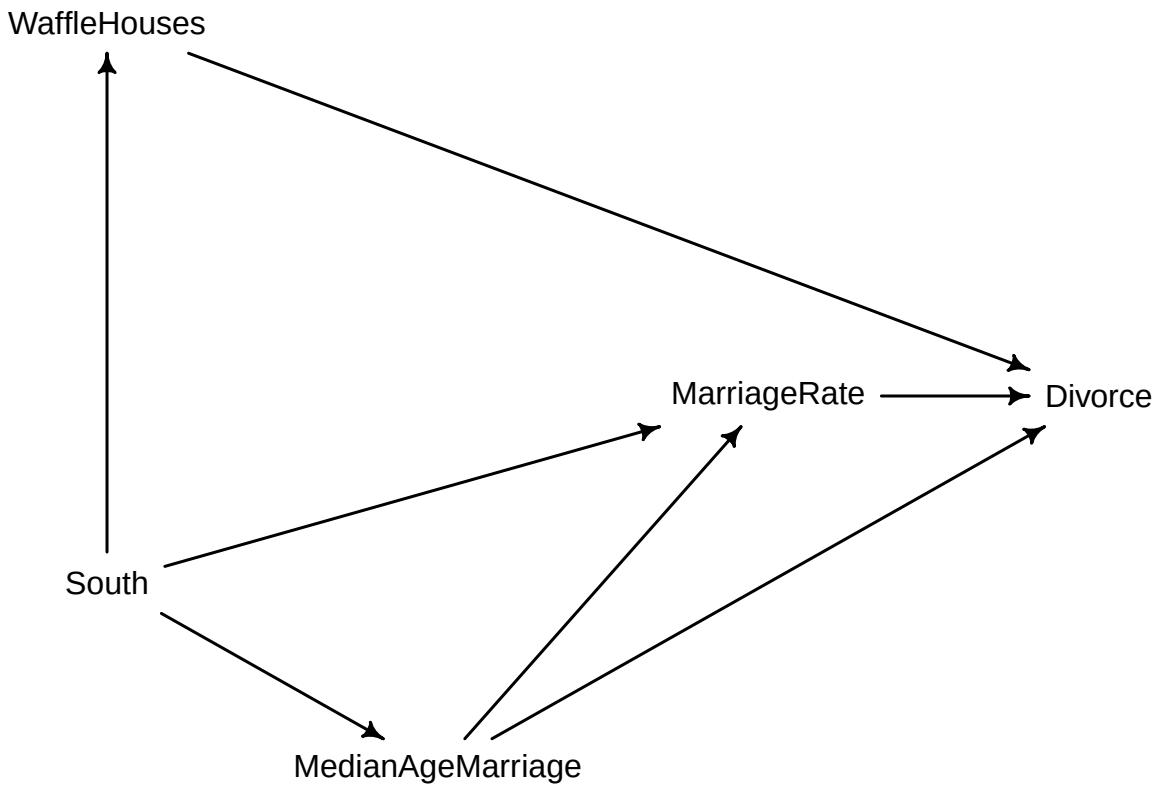
This model provides a better understanding of the situation. While (average) food has an estimated positive effect on weight (M=1.52, SD=0.78), groupsize on the other hand has a negative effect (M=-0.30, SD=0.11).

A possible explanation is that in bigger groups each fox gets a smaller share of the available food, therefore the negative relationship. Looking at groups of the same size however, the expected association between more food and higher weight can be observed.

Chapter 6: Investigating the Waffles and Divorces (HF)

Section author: Halfdan Fundal (HF) **6H1**. (HF) Use the Waffle House data, data(WaffleDivorce), to find the total causal influence of number of Waffle Houses on divorce rate. Justify your model or models with a causal graph.

```
# New DAG with full variable names and new layout
waffle_dag <- dagitty('dag {
  Divorce [outcome, pos="2,1"]
  MedianAgeMarriage [pos="0,2"]
  MarriageRate [pos="1,1"]
  South [pos="-1,1.5"]
  WaffleHouses [exposure, pos="-1,0"]

  MedianAgeMarriage -> Divorce
  MedianAgeMarriage -> MarriageRate
  MarriageRate -> Divorce
  South -> MedianAgeMarriage
  South -> MarriageRate
  South -> WaffleHouses
  WaffleHouses -> Divorce
}')

drawdag(waffle_dag)
```



First we need to identify and block all backdoor paths. There are currently three backdoors from WaffleHouses to Divorce, that introduces the spurious associations: WaffleHouses ← South → MedianAgeMarriage → Divorce WaffleHouses ← South → Marriage → Divorce Wafflehouses ← South → MedianAgeMarriageM → Marriage → Divorce Since South is a cause of both WH and the downstream variables, conditioning on S, closes this backdoor and isolates the effect of Wafflehouses on Divoce rate.

```
data(WaffleDivorce)

WaffleDivorce$WaffleHouses_standardized <- scale(WaffleDivorce$WaffleHouses)
WaffleDivorce$Divorce_standardized <- scale(WaffleDivorce$Divorce)

m6H1 <- quap(
  alist(
    Divorce_standardized ~ dnorm(mu, sigma),
    mu <- a + bW * WaffleHouses_standardized + bS * South,
    a ~ dnorm(0, 0.2),
    bW ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = WaffleDivorce
)

precis(m6H1)
```

```
##             mean         sd        5.5%     94.5%
```

```
## a     -0.07725526 0.12384317 -0.27518057 0.1206700
## bW     0.12311737 0.15331646 -0.12191194 0.3681467
## bS     0.39435930 0.29088871 -0.07053703 0.8592556
## sigma  0.92660409 0.09197977  0.77960265 1.0736055
```

6H1 Answer: Causal effect of wafflehouses on divorce rate is very low with a mean around 0.1 and 89% interval lying in the range of -0.12 and 0.37. This indicates no causal effect between Wafflehouses and Divorce rate after blocking backdoors.

6H2. (HF) Build a series of models to test the implied conditional independencies of the causal graph you used in the previous problem. If any of the tests fail, how do you think the graph needs to be amended? Does the graph need more or fewer arrows? Feel free to nominate variables that aren't int he data.

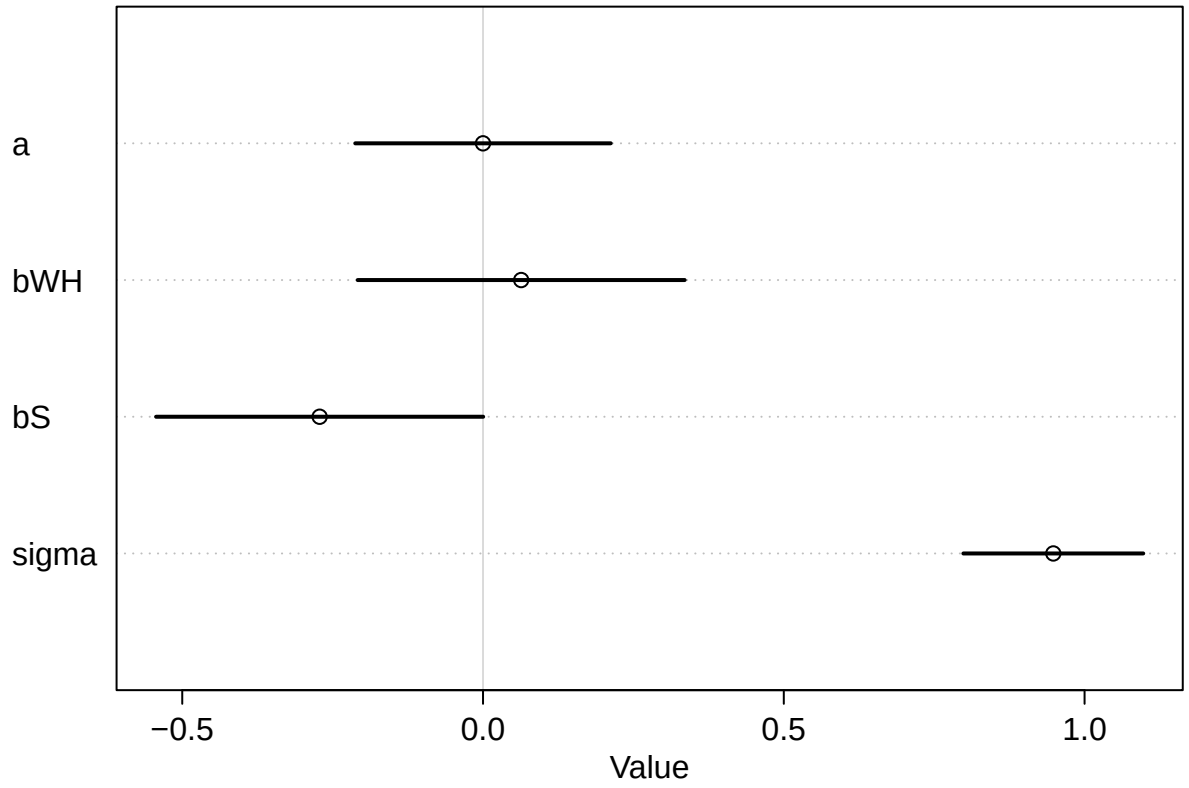
```
impliedConditionalIndependencies(waffle_dag)
```

```
## Dvrc _||_ Soth | MrrR, MdAM, WffH
## MrrR _||_ WffH | Soth
## MdAM _||_ WffH | Soth
```

```
data <- WaffleDivorce
data$AgeStd <- scale(data$MedianAgeMarriage)
data$MarriageStd <- scale(data$Marriage)
data$DivorceStd <- scale(data$Divorce)
data$WaffleStd <- scale(data$WaffleHouses)
data$SouthStd <- scale(data$South)
```

Median Age at Marriage $\perp\!\!\!\perp$ Waffle Houses | South?

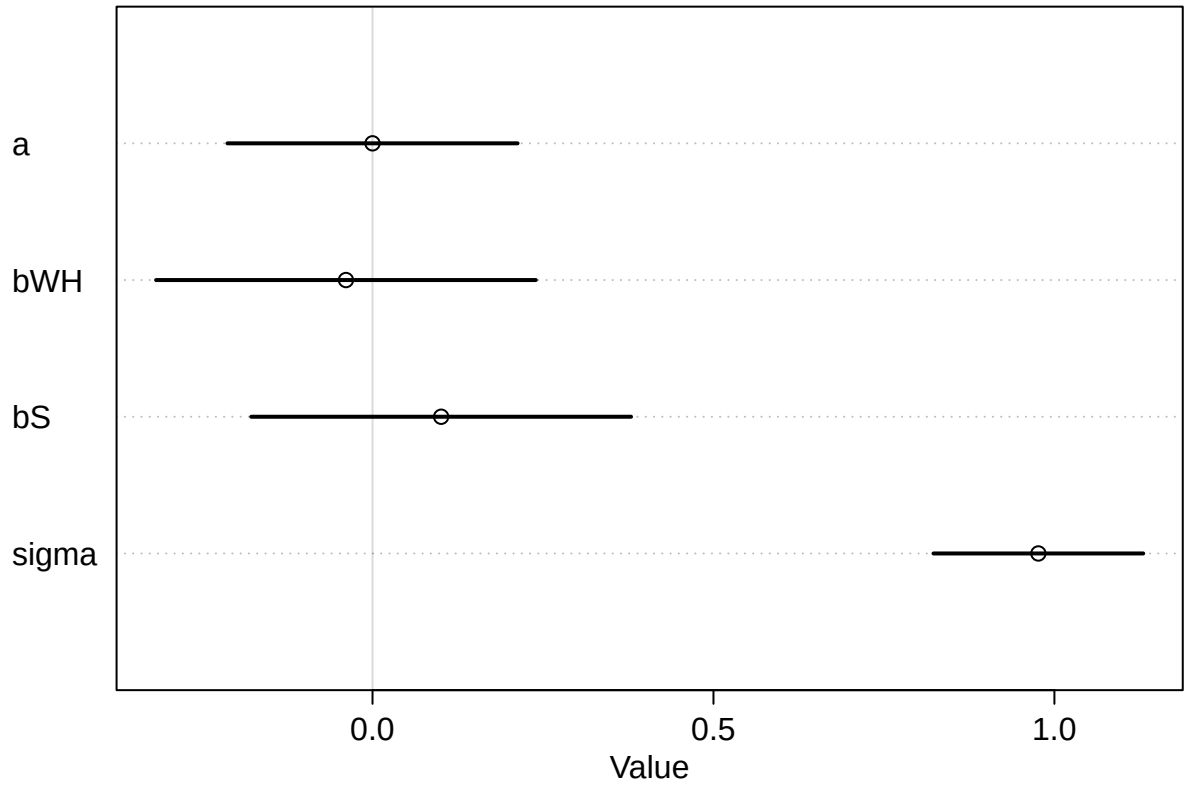
```
mod_age <- quap(
  alist(
    AgeStd ~ dnorm(mu, sigma),
    mu <- a + bWH * WaffleStd + bS * SouthStd,
    a ~ dnorm(0, 1),
    bWH ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = data
)

plot(precis(mod_age))
```



After conditioning on South, the association between WH and MAM becomes none. This confirms/supports the first of the DAG's implied conditional independencies

Marriage Rate ⫫ Waffle Houses | South?

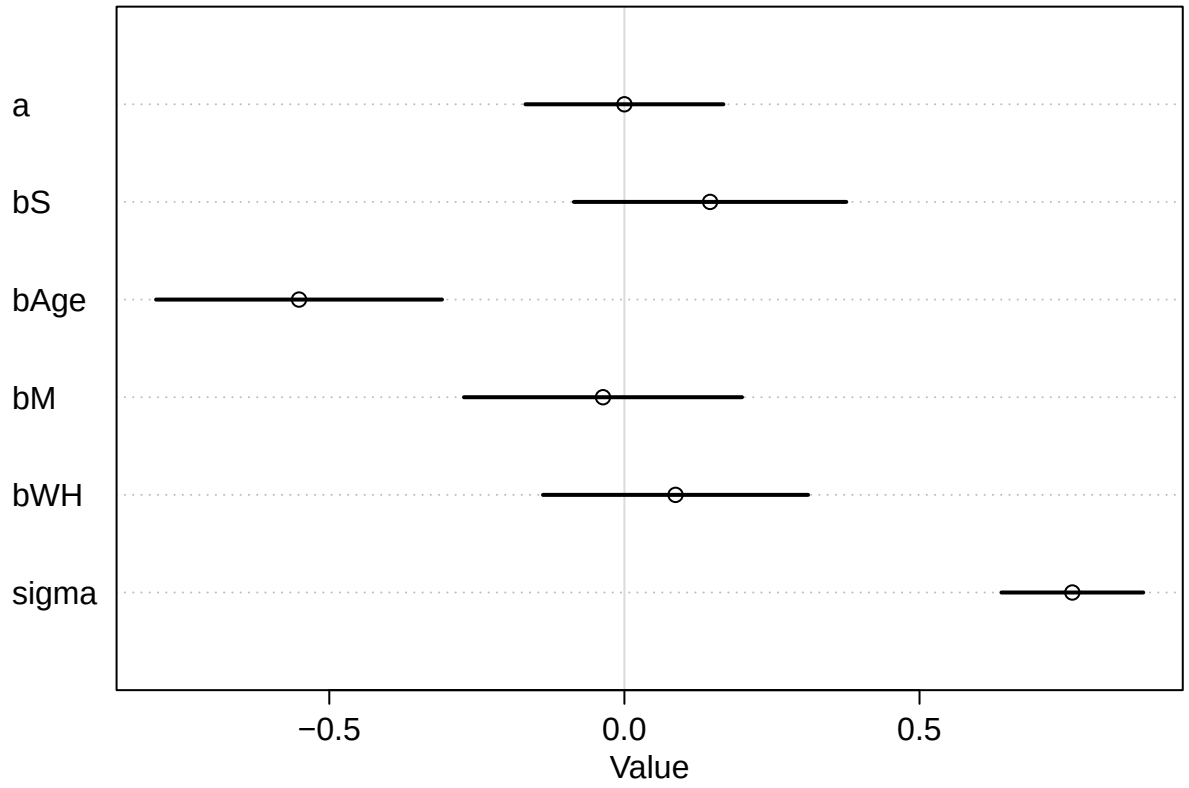
```
mod_mrate <- quap(
  alist(
    MarriageStd ~ dnorm(mu, sigma),
    mu <- a + bWH * WaffleStd + bS * SouthStd,
    a ~ dnorm(0, 0.5),
    bWH ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = data
)

plot(precis(mod_mrate))
```



Again the coefficient for WH is close to zero and supports with the assumption that Marriage Rate and Waffle Houses are independent after conditioning on South.

Divorce ⫫ South | Age, Marriage, Waffle Houses?

```
mod_divorce <- quap(
  alist(
    DivorceStd ~ dnorm(mu, sigma),
    mu <- a + bS * SouthStd + bAge * AgeStd + bM * MarriageStd + bWH * WaffleStd,
    a ~ dnorm(0, 0.5),
    bS ~ dnorm(0, 0.5),
    bAge ~ dnorm(0, 0.5),
    bM ~ dnorm(0, 0.5),
    bWH ~ dnorm(0, 0.5),
    sigma ~ dexp(1)
  ),
  data = data
)

plot(precis(mod_divorce))
```



6H2 Answer: After conditioning on MedianAgeMarriageRate, MarriageRate and WaffleHouses, the coefficient for South becomes very small, and indicates a conditional independence between South and Divorce rate, when conditioning on the mediators.