

Assignment 4

Natali Tkcvitishvili

2024-07-05

```
library(ggplot2)
library(tidyverse)
library(rethinking)
library(scales)
```

```
heart <- read.csv("heart.csv")
head(heart)
```

```
##   male age education currentSmoker cigsPerDay BPMeds prevalentStroke
## 1    1  39         4             0          0      0                0
## 2    0  46         2             0          0      0                0
## 3    1  48         1             1         20      0                0
## 4    0  61         3             1         30      0                0
## 5    0  46         3             1         23      0                0
## 6    0  43         2             0          0      0                0
##   prevalentHyp diabetes totChol sysBP diaBP BMI heartRate glucose TenYearCHD
## 1             0         0    195 106.0   70 26.97         80        77         0
## 2             0         0    250 121.0   81 28.73         95        76         0
## 3             0         0    245 127.5   80 25.34         75        70         0
## 4             1         0    225 150.0   95 28.58         65       103         1
## 5             0         0    285 130.0   84 23.10         85        85         0
## 6             1         0    228 180.0  110 30.30         77        99         0
```

```
summary(heart)
```

```
##           male           age           education    currentSmoker
## Min.      :0.0000   Min.    :32.00   Min.      :1.000   Min.      :0.0000
## 1st Qu.:0.0000   1st Qu.:42.00   1st Qu.:1.000   1st Qu.:0.0000
## Median :0.0000   Median :49.00   Median :2.000   Median :0.0000
## Mean      :0.4292   Mean     :49.58   Mean     :1.979   Mean     :0.4941
## 3rd Qu.:1.0000   3rd Qu.:56.00   3rd Qu.:3.000   3rd Qu.:1.0000
## Max.      :1.0000   Max.      :70.00   Max.      :4.000   Max.      :1.0000
##                                     NA's      :105
##           cigsPerDay           BPMeds           prevalentStroke    prevalentHyp
## Min.      : 0.000   Min.      :0.000000   Min.      :0.000000   Min.      :0.0000
## 1st Qu.: 0.000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.0000
## Median : 0.000   Median :0.000000   Median :0.000000   Median :0.0000
## Mean      : 9.003   Mean     :0.02963   Mean     :0.005899   Mean     :0.3105
## 3rd Qu.:20.000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:1.0000
## Max.      :70.000   Max.      :1.00000   Max.      :1.000000   Max.      :1.0000
```

```
## NA's :29      NA's :53
## diabetes      totChol      sysBP      diaBP
## Min. :0.00000  Min. :107.0  Min. : 83.5  Min. : 48.00
## 1st Qu.:0.00000 1st Qu.:206.0  1st Qu.:117.0 1st Qu.: 75.00
## Median :0.00000 Median :234.0  Median :128.0 Median : 82.00
## Mean :0.02572 Mean :236.7  Mean :132.4 Mean : 82.89
## 3rd Qu.:0.00000 3rd Qu.:263.0  3rd Qu.:144.0 3rd Qu.: 89.88
## Max. :1.00000 Max. :696.0  Max. :295.0 Max. :142.50
##      NA's :50
## BMI      heartRate      glucose      TenYearCHD
## Min. :15.54  Min. : 44.00  Min. : 40.00  Min. :0.000
## 1st Qu.:23.07 1st Qu.: 68.00 1st Qu.: 71.00 1st Qu.:0.000
## Median :25.40 Median : 75.00 Median : 78.00 Median :0.000
## Mean :25.80 Mean : 75.88 Mean : 81.97 Mean :0.152
## 3rd Qu.:28.04 3rd Qu.: 83.00 3rd Qu.: 87.00 3rd Qu.:0.000
## Max. :56.80 Max. :143.00 Max. :394.00 Max. :1.000
## NA's :19      NA's :1      NA's :388
```

Task Set 1

Task 1.1

Run a Bayesian logistic regression model to estimate the risk of men and women to develop a coronary heart disease (TenYearCHD). Provide a summary of the posterior distributions.
What is the average probability of men and women to develop the disease?

```
# gender: 1 is female, 2 is male

heart_list <- list(
  n = nrow(heart),
  gender = heart$male + 1, # re-index as ulam doesn't work with 0
  CHD = heart$TenYearCHD,
  diabetes = heart$diabetes + 1, # for task 2
  age = heart$age,
  avg_age = mean(heart$age)
  # age_women = heart$age[heart$male == 0], # for task 3
  # age_men = heart$age[heart$male == 1]
)

disease_gender <- ulam(
  alist(
    CHD ~ dbinom(n, p), # binomial distribution - success or not
    logit(p) <- a[gender],
    a[gender] ~ dnorm(0, 1)
  ), data = heart_list, chains = 4, cores = 4
)
```

```
## Running MCMC with 4 parallel chains, with 1 thread(s) per chain...
##
## Chain 1 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 2 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 3 Iteration: 1 / 1000 [ 0%] (Warmup)
```

```

## Chain 4 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 1 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 2 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 3 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 4 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 2 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 3 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 1 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 4 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 2 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 3 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 4 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 1 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 3 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 2 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 4 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 1 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 3 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 3 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 2 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 2 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 4 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 1 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 1 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 3 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 2 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 4 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 1 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 4 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 3 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 1 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 4 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 3 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 1 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 2 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 3 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 4 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 1 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 2 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 4 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 2 finished in 4.3 seconds.
## Chain 4 finished in 4.2 seconds.
## Chain 3 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 3 finished in 4.3 seconds.
## Chain 1 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 1 finished in 4.5 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 4.3 seconds.
## Total execution time: 4.7 seconds.

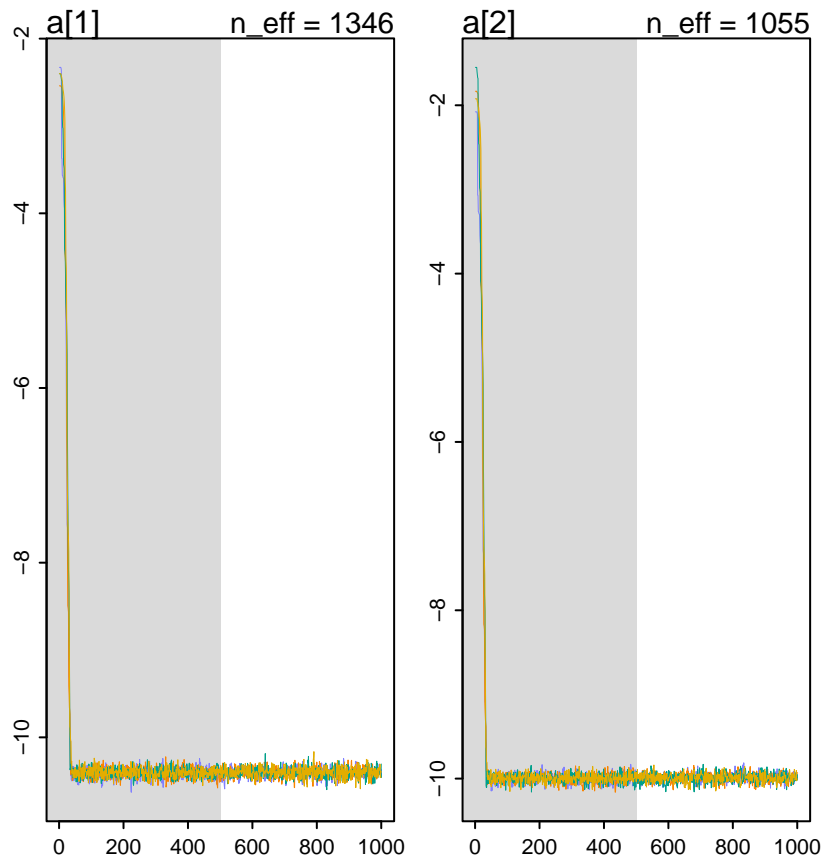
```

```
traceplot(disease_gender) # looks stable
precis(disease_gender, depth = 2)
```

```
##           mean          sd      5.5%      94.5%      rhat ess_bulk
## a[1] -10.403960 0.05803583 -10.49921 -10.309989 1.004920 1345.724
## a[2]  -9.992069 0.05281433 -10.07821  -9.905612 1.003959 1054.851
```

```
inv_logit(coef(disease_gender))
```

```
##           a[1]          a[2]
## 3.031129e-05 4.575932e-05
```



Average probabilities to develop the disease for: * women (gender = 1) \sim 0.003% * men (gender = 2) \sim 0.005%

Task 1.2

For the model of Task 1.1, visualize the posterior distribution of gender-differences to assess the credibility of the gender difference.

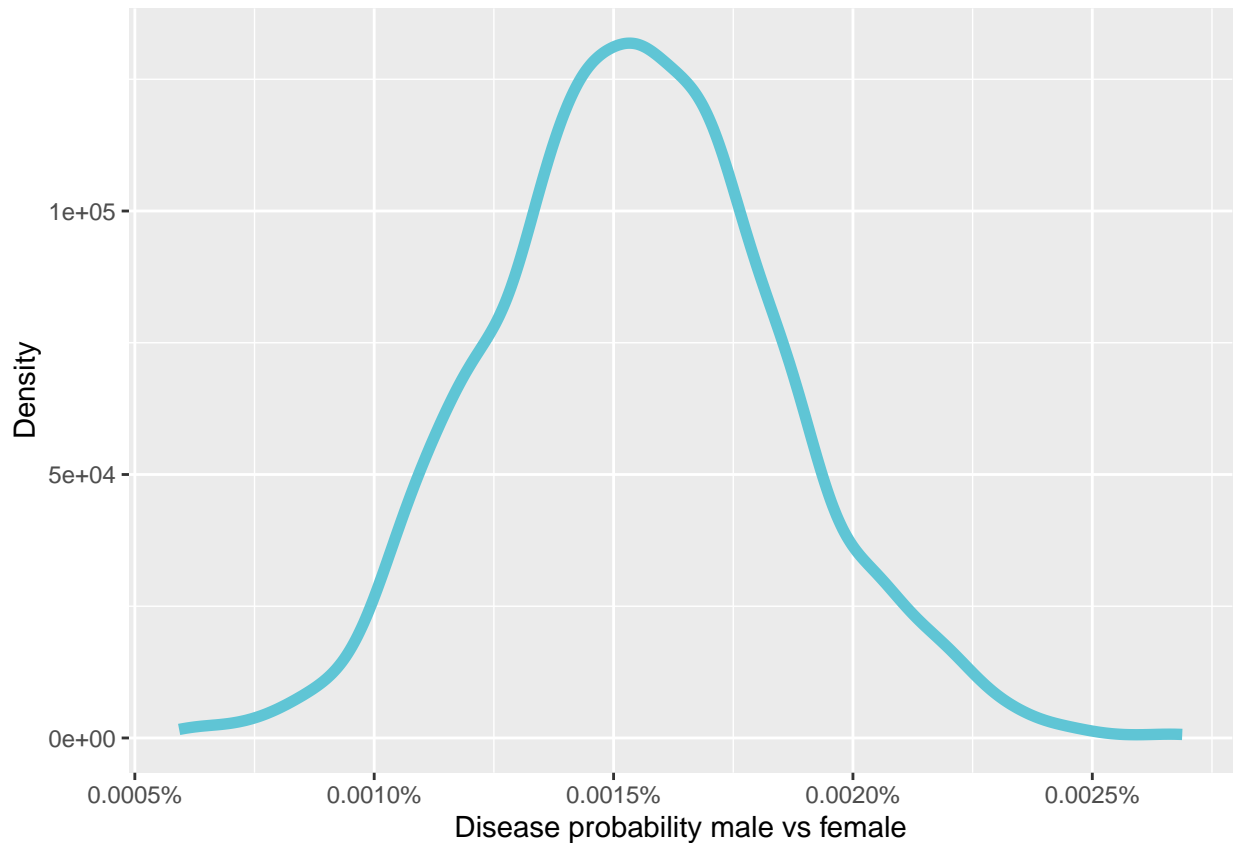
```
set.seed(1111)
disease_gender_samples <- extract.samples(disease_gender)
posterior <- tibble(p_female = inv_logit(disease_gender_samples$a[, 1]),
```

```

p_male = inv_logit(disease_gender_samples$a[, 2]),
diff = p_male - p_female)

ggplot(posterior, aes(diff)) +
  geom_density(color = "#5fc5d5", linewidth = 2) +
  # converting to % scale for easier reading
  scale_x_continuous(labels = percent_format(accuracy = 0.0001)) +
  labs(x = "Disease probability male vs female",
       y = "Density")

```



Most of the probability mass is concentrated between 0.001% and 0.002% difference, therefore, we can assume that there is a high probability that genders differ in chances of developing a coronary heart disease. But there might be other confounders, for example, men in general smoke more than women, I suppose smoking increases the risk of heart disease more than just being a man

Task Set 2

Task 2.1

Run a Bayesian logistic regression model to estimate the risk of men and women with and without diabetes to develop a coronary heart disease (TenYearCHD). Provide a summary of the posterior distributions. Does the effect of diabetes differ between men and women?

```
# diabetes: 1 without, 2 with

disease_diabetes <- ulam(
  alist(
    CHD ~ dbinom(n, p), # binomial distribution - success or not
    logit(p) <- a[gender, diabetes],
    matrix[gender, diabetes]:a ~ dnorm(0, 1)
  ), data = heart_list, chains = 4, cores = 4
)

## Running MCMC with 4 parallel chains, with 1 thread(s) per chain...
##
## Chain 1 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 2 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 3 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 4 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 2 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 1 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 4 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 3 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 1 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 2 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 4 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 3 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 1 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 2 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 4 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 2 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 3 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 1 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 4 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 3 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 2 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 2 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 4 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 1 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 1 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 2 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 3 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 3 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 1 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 2 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 3 Iteration: 600 / 1000 [ 60%] (Sampling)
```

```

## Chain 4 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 1 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 3 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 1 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 4 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 2 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 3 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 4 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 1 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 2 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 2 finished in 5.5 seconds.
## Chain 3 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 4 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 4 finished in 5.8 seconds.
## Chain 1 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 1 finished in 5.9 seconds.
## Chain 3 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 3 finished in 6.0 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 5.8 seconds.
## Total execution time: 6.2 seconds.

```

```

traceplot(disease_diabetes) # looks stable
precis(disease_diabetes, depth = 3)

```

```

##           mean          sd      5.5%      94.5%      rhat ess_bulk
## a[1,1] -10.438463 0.05783098 -10.531933 -10.347078 1.001065 2306.559
## a[2,1] -10.029450 0.05404840 -10.115811  -9.942229 1.002935 1656.506
## a[1,2]  -9.113110 0.19121990  -9.431072  -8.821439 1.002143 2046.694
## a[2,2]  -8.890144 0.17956229  -9.175418  -8.602505 1.002372 2622.412

```

```

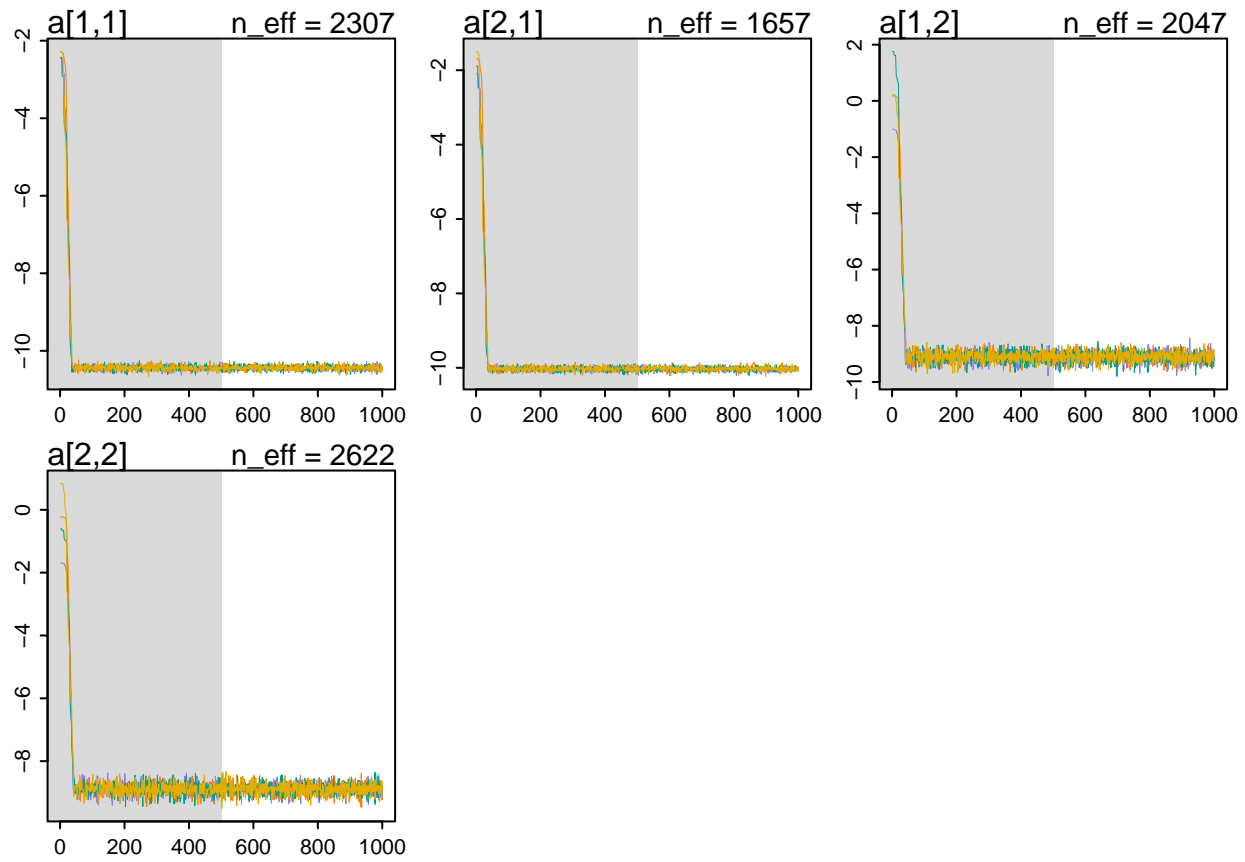
# renaming columns to make analysis easier
coef_diabetes <- inv_logit(coef(disease_diabetes))
names(coef_diabetes) <- c("female_without", "male_without", "female_with", "male_with")
coef_diabetes

```

```

## female_without  male_without  female_with  male_with
## 2.928332e-05 4.408045e-05 1.101993e-04 1.377209e-04

```



Among people who have diabetes the difference is not as big as among those who do not have it: * women with diabetes: $\sim 0.011\%$ * men with diabetes: $\sim 0.014\%$

- women without diabetes: $\sim 0.0029\%$
- men without diabetes: $\sim 0.0044\%$

Task 2.2

For the model of Task 2.1, visualize the posterior distributions of each group in one plot to better assess the credibility of the group differences.

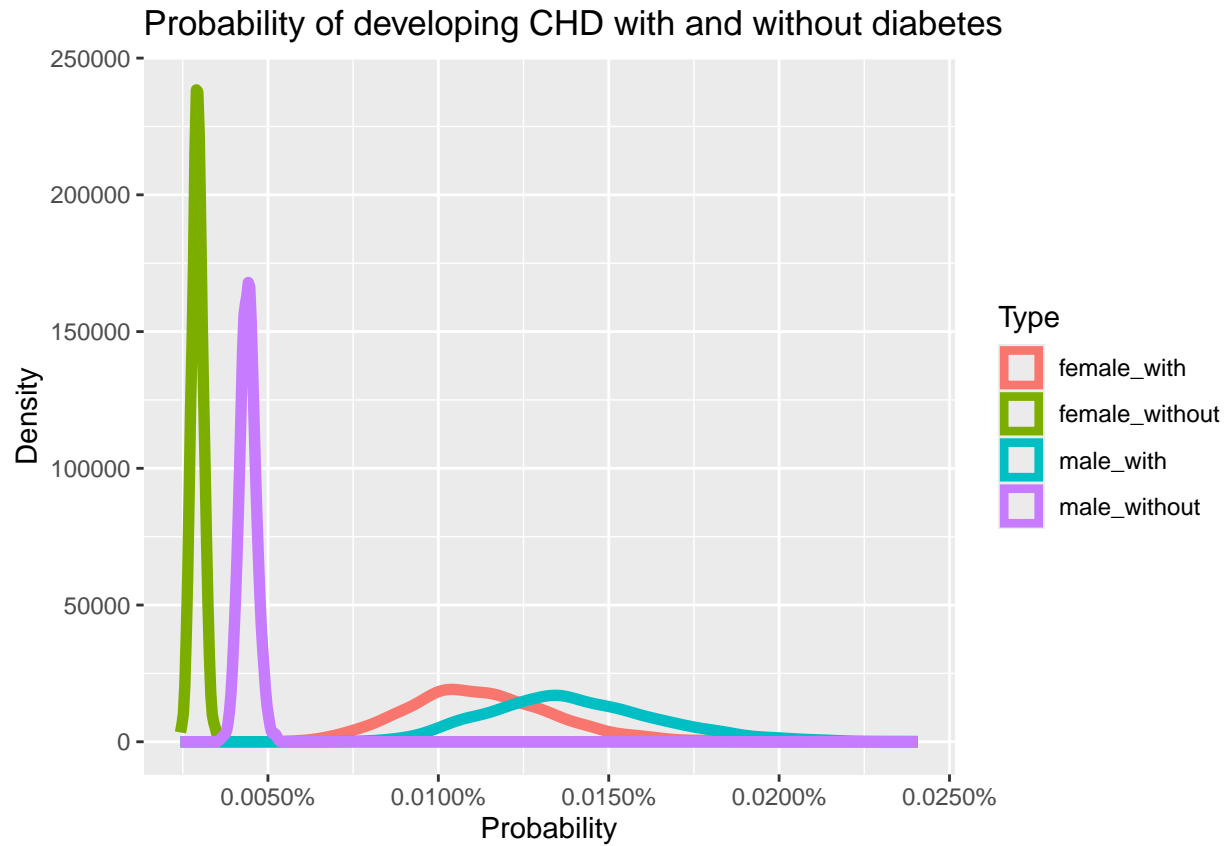
```
set.seed(1111)
disease_diabetes_samples <- data.frame(extract.samples(disease_diabetes))
names(disease_diabetes_samples) <- c("female_without", "male_without", "female_with", "male_with")

posterior_diabetes <- data.frame(
  female_without = inv_logit(disease_diabetes_samples$female_without),
  male_without = inv_logit(disease_diabetes_samples$male_without),
  female_with = inv_logit(disease_diabetes_samples$female_with),
  male_with = inv_logit(disease_diabetes_samples$male_with)
) %>%
  pivot_longer(cols = female_without:male_with, names_to = "Type", values_to = "Probability")

ggplot(posterior_diabetes, aes(Probability, color = Type)) +
```



```
geom_density(alpha = .3, linewidth = 2) +
scale_x_continuous(labels = percent_format(accuracy = 0.0001)) +
labs(x = "Probability",
     y = "Density",
     title = "Probability of developing CHD with and without diabetes"
)
```



According to the graph, men are under higher risk than women to develop coronary heart disease both with and without diabetes. Besides, both genders with diabetes have significantly higher chances to develop CHD than those without.

Task Set 3

Task 3.1

Run a Bayesian logistic regression model to estimate the effect of age on the risk of developing a coronary heart disease (TenYearCHD), separately for women and men. Ensure that the regression intercept represents the risk of women and men with average age. Provide a summary of the posterior distributions.

```
women_list <- list(
  n = nrow(heart[heart$male == 0, ]),
  CHD = heart$TenYearCHD[heart$male == 0],
  age_women = heart$age[heart$male == 0],
  avg_age_women = mean(heart$age[heart$male == 0])
)

disease_age_women <- ulam(
  alist(
    CHD ~ dbinom(n, p),
    logit(p) <- a_1 + b_1 * (age_women - avg_age_women), # standardizing around the mean
    a_1 ~ dnorm(10, 5),
    b_1 ~ dnorm(1, 0.1)
  ), data = women_list, chains = 4, cores = 4
)
```

```
## Running MCMC with 4 parallel chains, with 1 thread(s) per chain...
```

```
##
## Chain 1 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 2 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 3 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 4 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 1 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 2 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 1 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 4 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 2 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 1 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 3 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 2 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 4 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 1 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 3 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 2 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 1 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 1 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 3 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 1 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 2 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 2 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 3 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 1 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 600 / 1000 [ 60%] (Sampling)
```

```

## Chain 4 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 4 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 1 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 2 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 3 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 3 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 1 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 4 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 3 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 1 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 4 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 1 finished in 2.8 seconds.
## Chain 3 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 4 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 2 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 2 finished in 4.8 seconds.
## Chain 3 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 4 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 4 finished in 5.2 seconds.
## Chain 3 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 3 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 3 finished in 6.5 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 4.9 seconds.
## Total execution time: 6.7 seconds.

```

```

traceplot(disease_age_women)
precis(disease_age_women, depth = 2)

```

```

##           mean           sd      5.5%      94.5%      rhat ess_bulk
## a_1 -10.09787817 0.072974118 -10.21910000 -9.98763175 1.007780 814.0449
## b_1  0.07685664 0.007113709  0.06563522  0.08821551 1.008188 825.7073

```

```

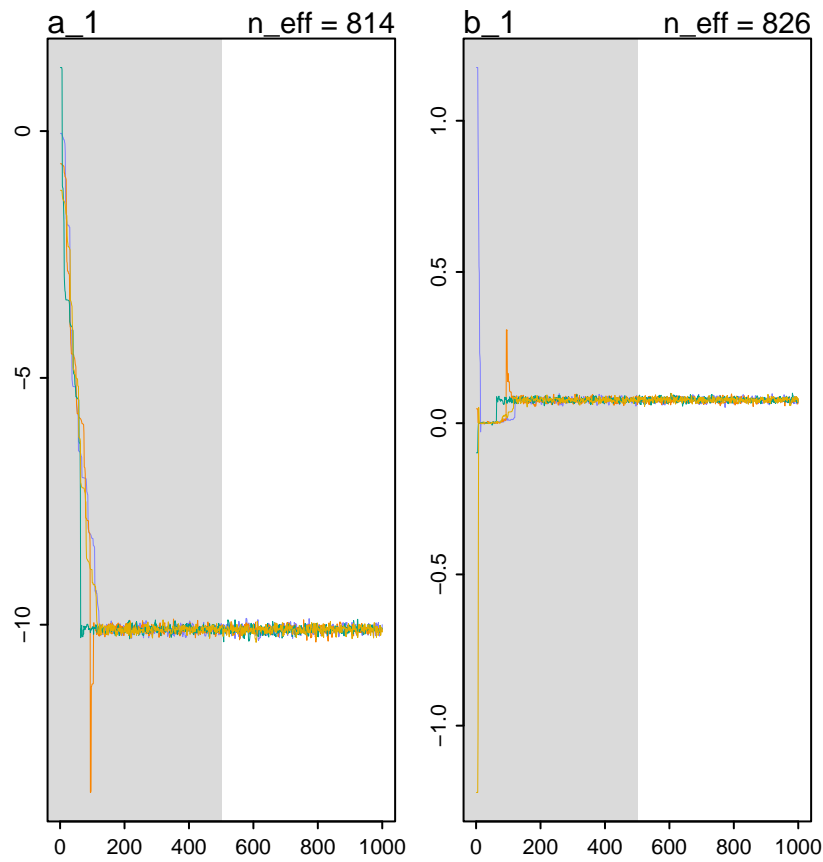
inv_logit(coef(disease_age_women))

```

```

##           a_1           b_1
## 4.116512e-05 5.192047e-01

```



```
men_list <- list(
  n = nrow(heart[heart$male == 1, ]),
  CHD = heart$TenYearCHD[heart$male == 1],
  age_men = heart$age[heart$male == 1],
  avg_age_men = mean(heart$age[heart$male == 1])
)

disease_age_men <- ulam(
  alist(
    CHD ~ dbinom(n, p),
    logit(p) <- a_1 + b_1 * (age_men - avg_age_men), # standardizing around the mean
    a_1 ~ dnorm(10, 5),
    b_1 ~ dnorm(1, 0.1)
  ), data = men_list, chains = 4, cores = 4
)
```

```
## Running MCMC with 4 parallel chains, with 1 thread(s) per chain...
##
## Chain 1 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 2 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 3 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 4 Iteration: 1 / 1000 [ 0%] (Warmup)
## Chain 1 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 2 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 4 Iteration: 100 / 1000 [ 10%] (Warmup)
## Chain 3 Iteration: 100 / 1000 [ 10%] (Warmup)
```

```

## Chain 1 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 4 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 2 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 3 Iteration: 200 / 1000 [ 20%] (Warmup)
## Chain 1 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 4 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 2 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 3 Iteration: 300 / 1000 [ 30%] (Warmup)
## Chain 1 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 4 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 2 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 3 Iteration: 400 / 1000 [ 40%] (Warmup)
## Chain 1 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 1 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 4 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 1 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 2 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 2 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 3 Iteration: 500 / 1000 [ 50%] (Warmup)
## Chain 3 Iteration: 501 / 1000 [ 50%] (Sampling)
## Chain 4 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 1 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 3 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 4 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 2 Iteration: 600 / 1000 [ 60%] (Sampling)
## Chain 3 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 4 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 1 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 2 Iteration: 700 / 1000 [ 70%] (Sampling)
## Chain 3 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 1 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 4 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 2 Iteration: 800 / 1000 [ 80%] (Sampling)
## Chain 3 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 4 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 1 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 1 finished in 8.5 seconds.
## Chain 4 finished in 8.2 seconds.
## Chain 3 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 3 finished in 8.6 seconds.
## Chain 2 Iteration: 900 / 1000 [ 90%] (Sampling)
## Chain 2 Iteration: 1000 / 1000 [100%] (Sampling)
## Chain 2 finished in 9.3 seconds.
##
## All 4 chains finished successfully.
## Mean chain execution time: 8.7 seconds.
## Total execution time: 9.6 seconds.

```

```

traceplot(disease_age_men)
precis(disease_age_men, depth = 2)

```

```

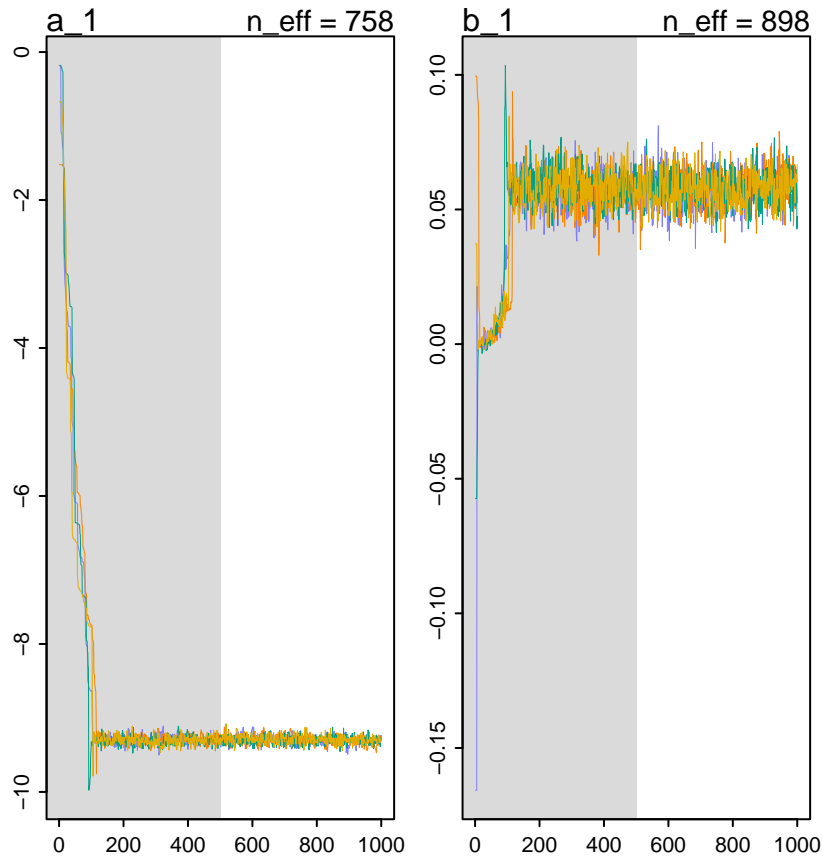
##          mean          sd    5.5%    94.5%    rhat ess_bulk
## a_1 -9.29630447 0.058828259 -9.38963 -9.20380845 1.009806 758.1149

```

```
## b_1 0.05751204 0.006338585 0.04729 0.06769243 1.005534 897.8784
```

```
inv_logit(coef(disease_age_men))
```

```
##          a_1          b_1
## 9.17543e-05 5.14374e-01
```



```
## Task 3.2
```

For the model of Task 3.1, visualize the posterior distribution of differences in the age effect between women and men. Does age increase the risk of developing the disease and does this effect differ between women and men?

```
set.seed(1111)
women_age_samples <- data.frame(extract.samples(disease_age_women))
head(women_age_samples)
```

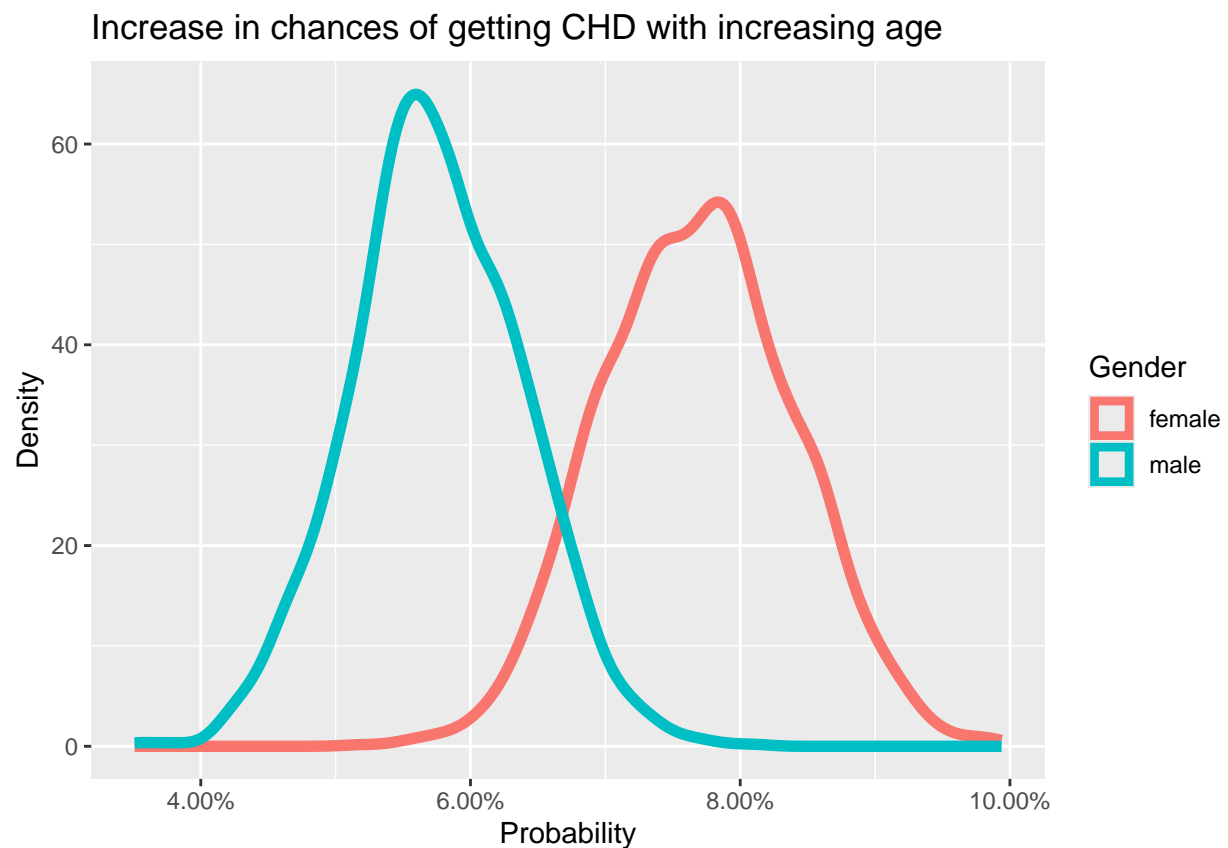
```
##          a_1          b_1
## 1 -9.98749 0.0652675
## 2 -10.11360 0.0781495
## 3 -10.18160 0.0842182
## 4 -10.16150 0.0780057
## 5 -10.07700 0.0766004
## 6 -10.10560 0.0753352
```

```
men_age_samples <- data.frame(extract.samples(disease_age_men))
head(men_age_samples)
```

```
##      a_1      b_1
## 1 -9.37573 0.0590561
## 2 -9.39396 0.0636202
## 3 -9.36557 0.0637937
## 4 -9.34592 0.0592983
## 5 -9.30485 0.0636519
## 6 -9.30645 0.0565415
```

```
posterior_age <- data.frame(
  female = women_age_samples$b_1,
  male = men_age_samples$b_1
) %>%
  pivot_longer(cols = female:male, names_to = "Gender", values_to = "Coef")

ggplot(posterior_age, aes(Coef, color = Gender)) +
  geom_density(alpha = .3, linewidth = 2) +
  scale_x_continuous(labels = percent_format(accuracy = 0.01)) +
  labs(x = "Probability",
       y = "Density",
       title = "Increase in chances of getting CHD with increasing age"
  )
```



With increasing age, women tend to have higher risk of developing CHD than men That, honestly, doesn't sound right, but I don't know how to change the model :(