Introduction to multiparameter models

Stat 340: Bayesian Statistics

1. Multiparameter models

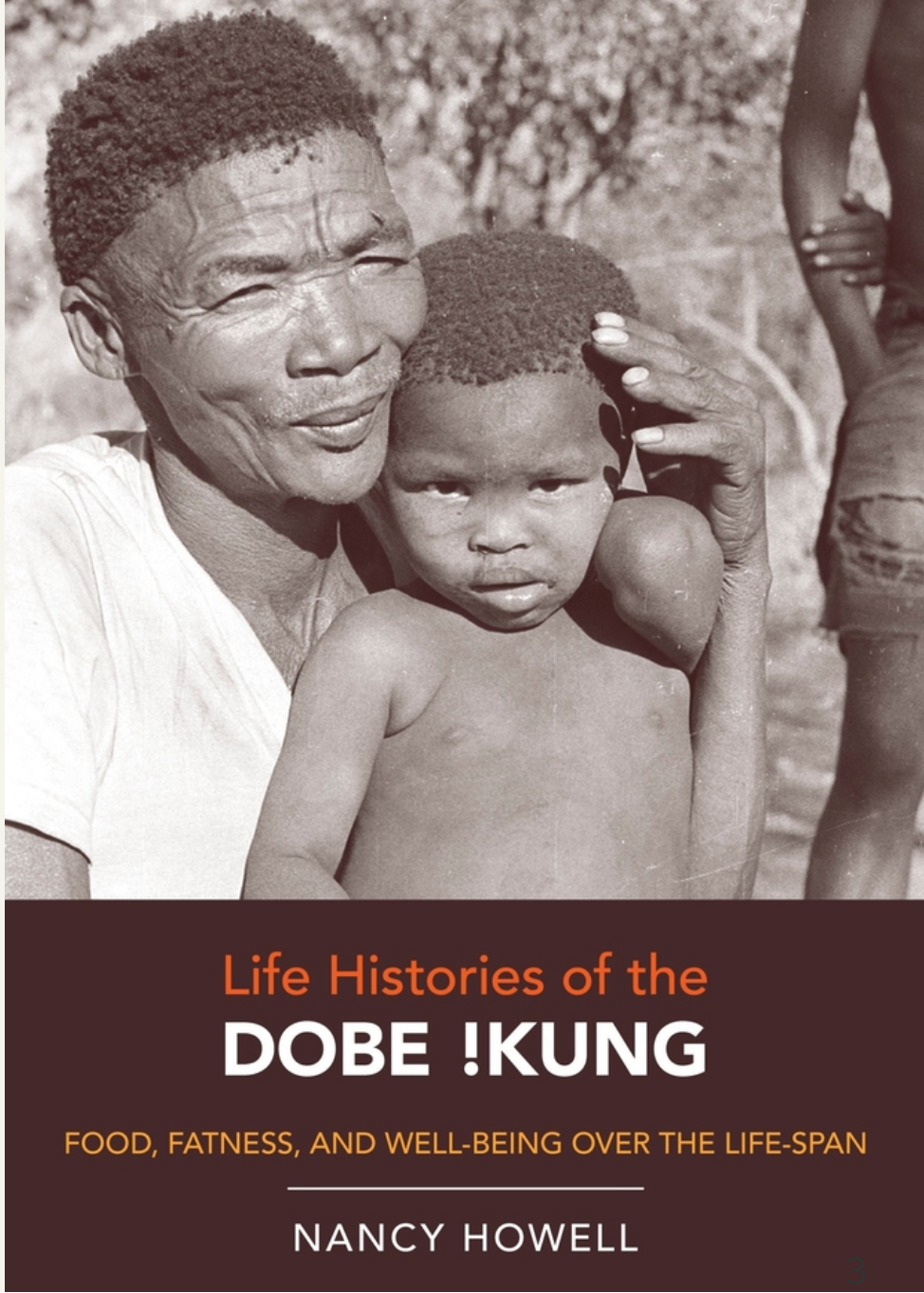
2. Grid approximation

(Problem topics 6 & 7)

Example

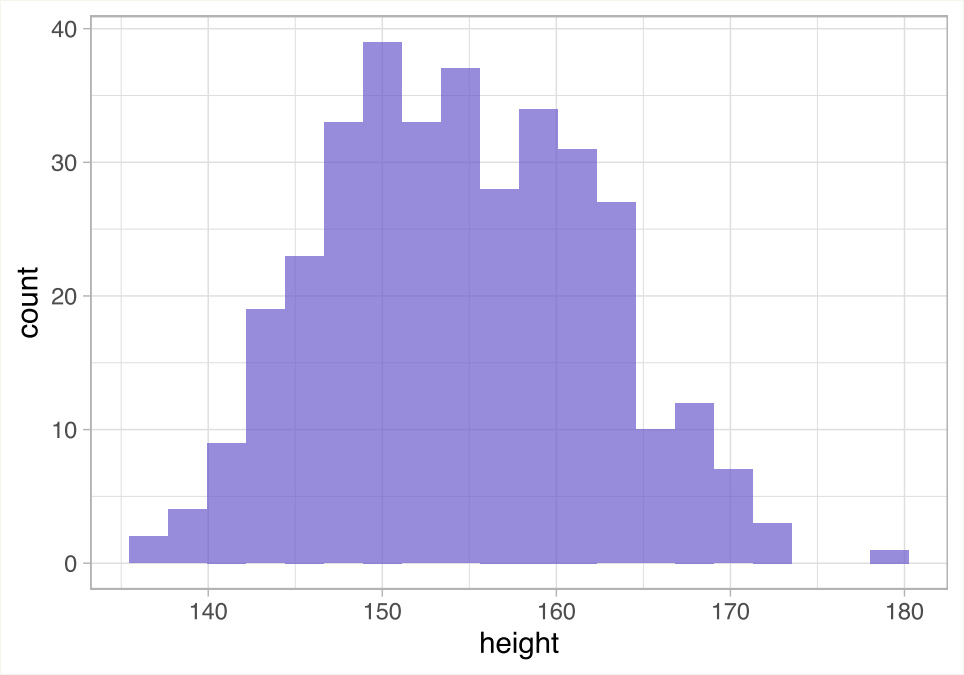
- Partial census data for the Dobe area !Kung San, a foraging population
- Compiled from Nancy Howell's interviews

```
##              mean         sd     5.5%     94.5%     histogra
## height 154.59709  7.7423321 142.8750 167.00500
## weight  44.99049  6.4567081  35.1375  55.76588
## age     41.13849 15.9678551  20.0000  70.00000
## male     0.46875  0.4997328   0.0000   1.00000
```



Example

Suppose interest is in analyzing the average height of an adult



Anthropologists would be interested in more complex relationships, but we have to start somewhere.

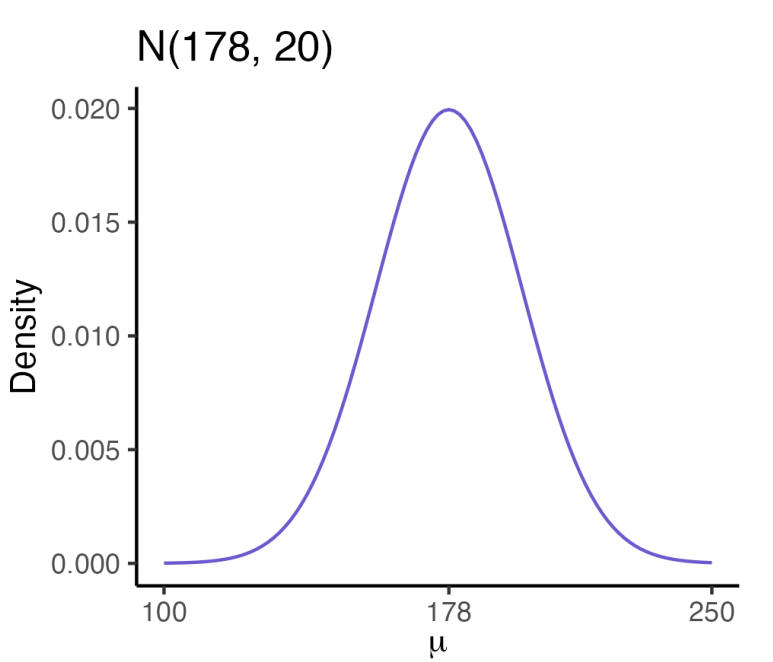
Informative analysis

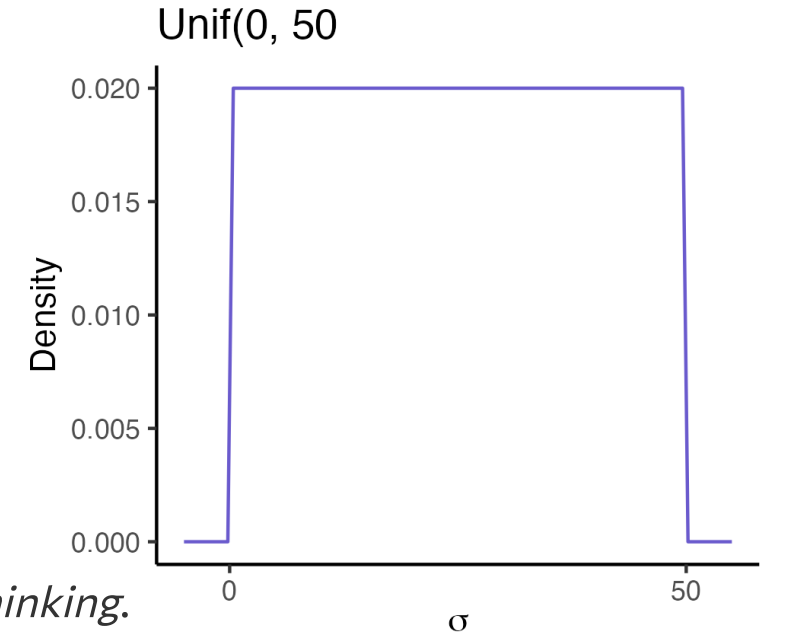
NORMAL MODEL

$$y_i \sim \mathcal{N}(\mu, \sigma)$$

$$\mu \sim \mathcal{N}(178,\ 20)$$

$$\sigma \sim \text{Unif}(0, 50)$$





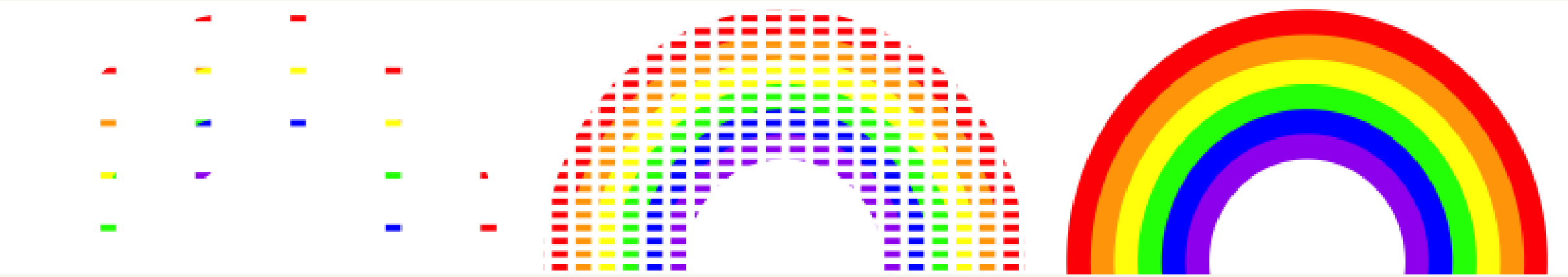
This analysis was discussed by McElreath in *Statistical Rethinking*.

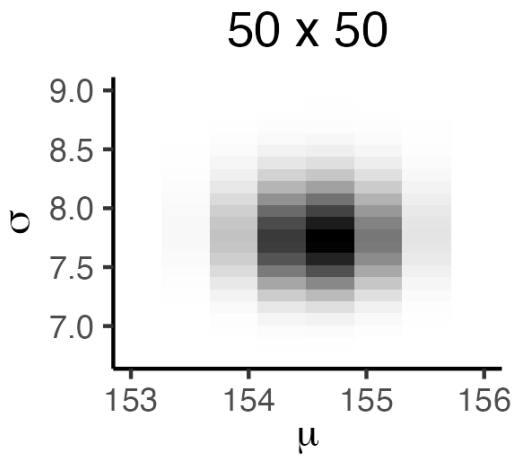
Prior predictive distribution

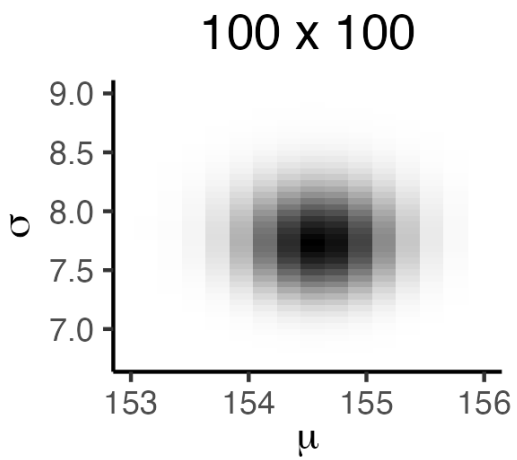
What do these priors imply about the height, before we see data?

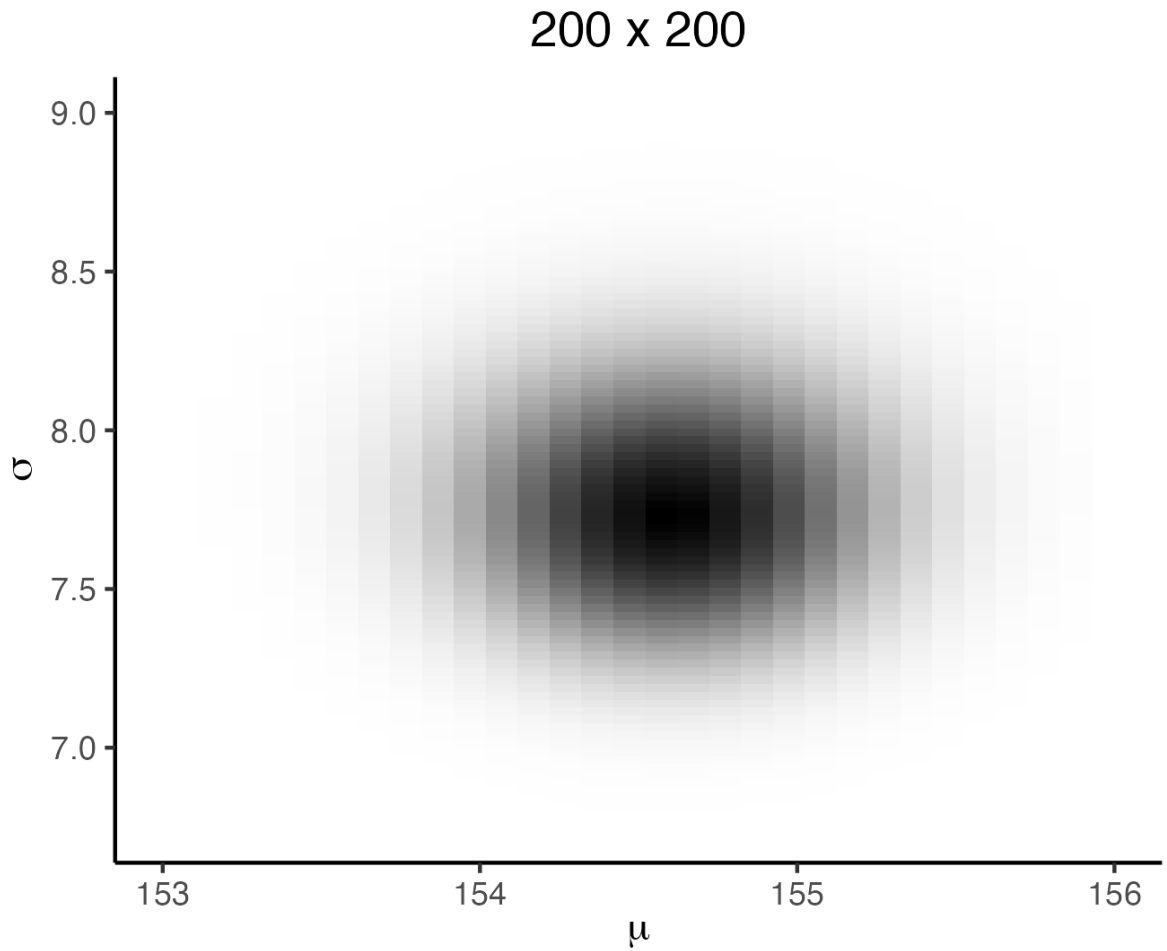
```
sample_mu    <- rnorm(1e4, 178, 20)
sample_sigma <- runif(1e4, 0, 50)
sim_heights  <- rnorm(1e4, sample_mu, sample_sigma)
```

We can *approximate* the joint posterior using a grid




50 x 50
9.0
8.5
8.0
7.5
7.0
σ
153
154
155
156
μ


100 x 100
9.0
8.5
8.0
7.5
7.0
σ
153
154
155
156
μ


200 x 200
9.0
8.5
8.0
7.5
7.0
σ
153
154
155
156
μ

Create a grid over the coordinate plane

```
param_grid <- expand.grid(
  mu = seq(from = 118, to = 238, length.out = 1000),
  sigma = seq(from = 0, to = 50, length.out = 1000)
)
```

Create a vectorized log-likelihood function

```
# log likelihood function
log_lik_norm <- function(x, mu, sigma) {
  sum(dnorm(x, mean = mu, sd = sigma, log = TRUE))
}

# Vectorize so we can pass in all mu and sigma at once
log_lik_norm <- Vectorize(log_lik_norm, vectorize.args = c("mu", "sigma"))
```

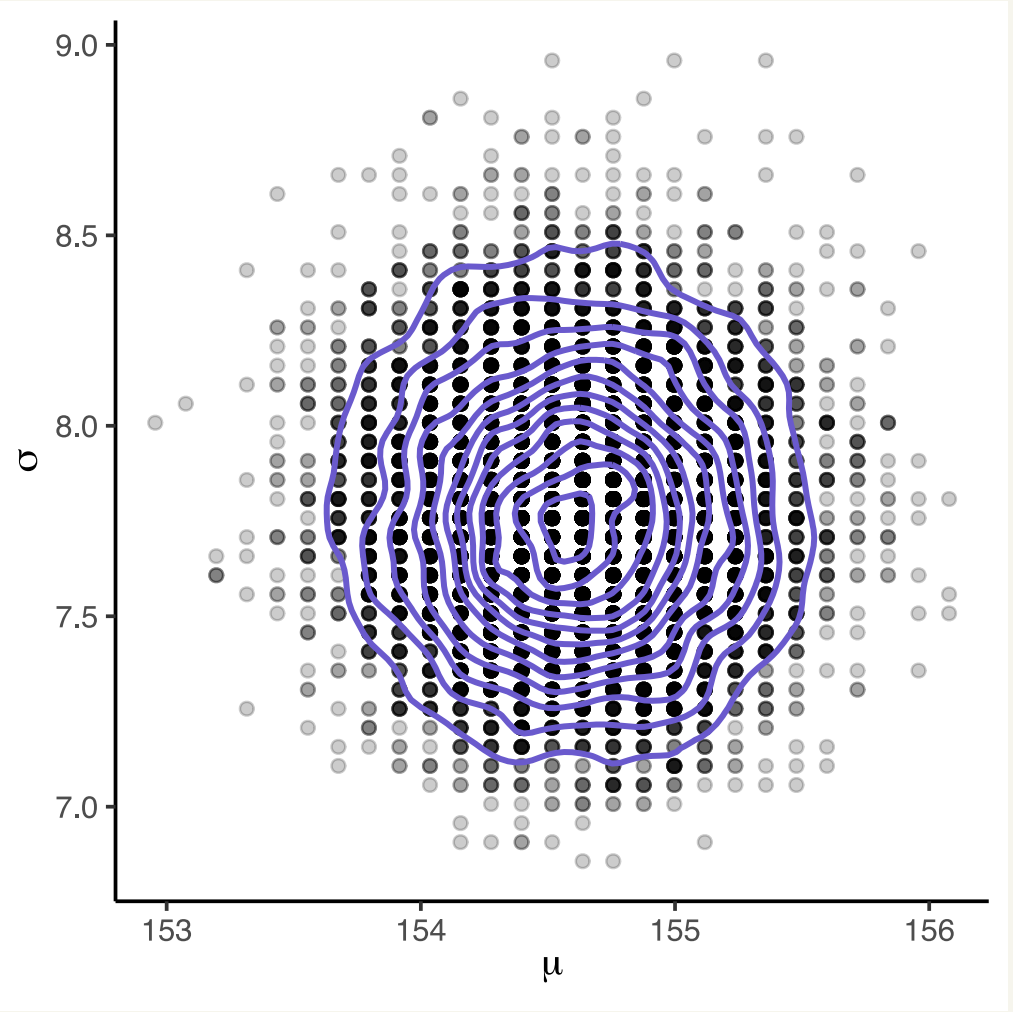
Evaluate log prior, log-likelihood on the grid, then derive the posterior

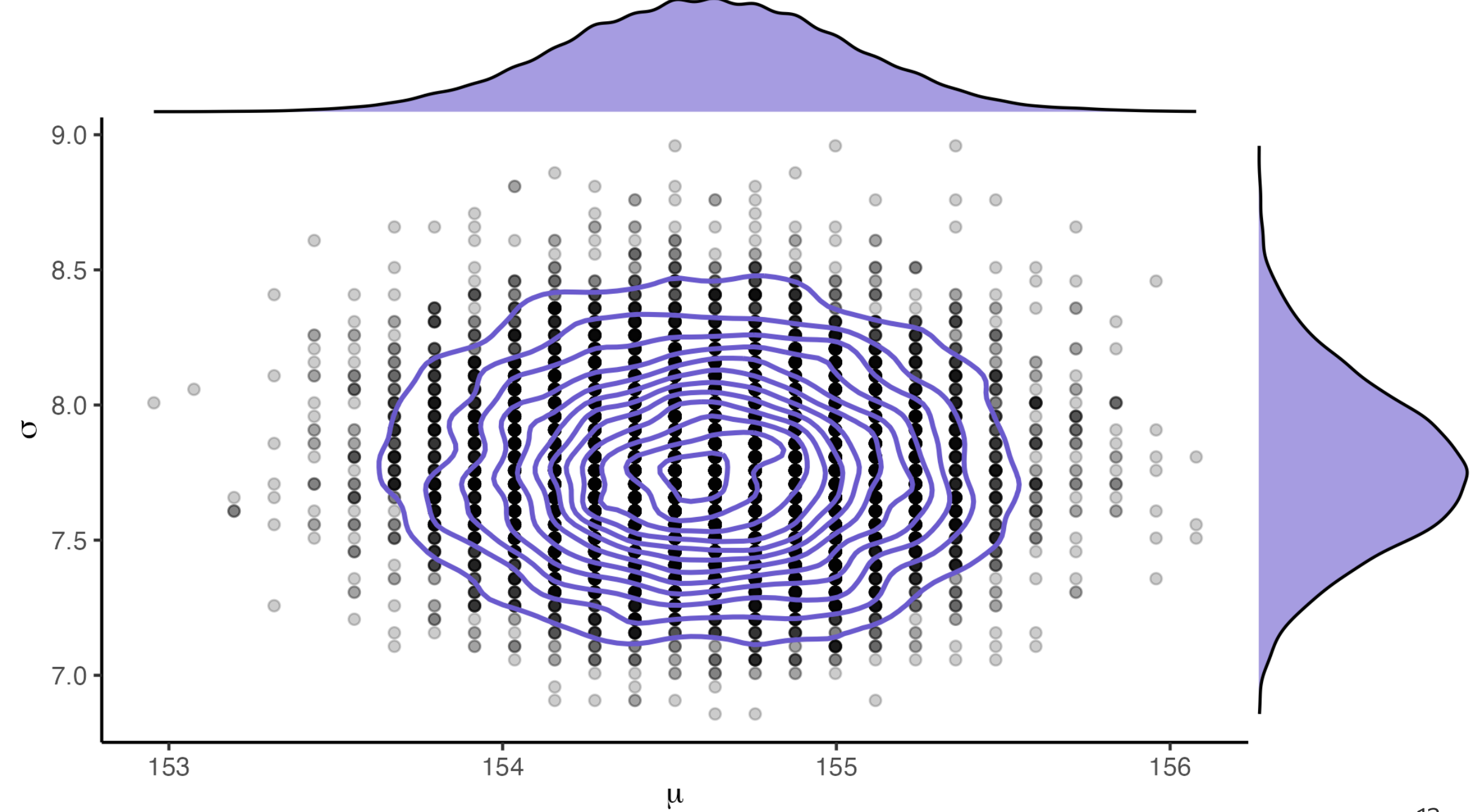
```
posterior <- param_grid %>%
  mutate(
    log_prior = dnorm(mu, 178, 20, log = TRUE) +
      dunif(sigma, 0, 50, log = TRUE),
    log_lik = log_lik_norm(adults$height, mu = mu, sigma = sigma),
    log_post = log_prior + log_lik,
    unstd_post = exp(log_post - max(log_post)),
    post = unstd_post / sum(unstd_post)
  )
```

```
## # A tibble: 1,000,000 × 7
##      mu sigma log_prior log_lik log_post unstd_post  post
##   <dbl> <dbl>     <dbl>   <dbl>    <dbl>      <dbl> <dbl>
## 1  118      0     -12.3    -Inf     -Inf          0     0
## 2  118.     0     -12.3    -Inf     -Inf          0     0
## 3  118.     0     -12.3    -Inf     -Inf          0     0
## 4  118.     0     -12.3    -Inf     -Inf          0     0
## # … with 999,996 more rows
```

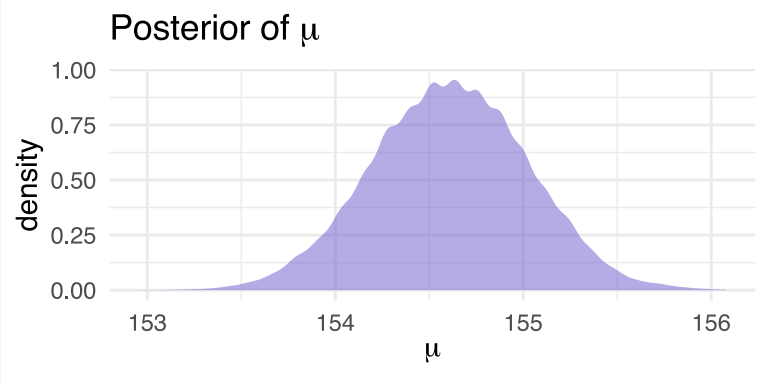
Sample from your grid posterior

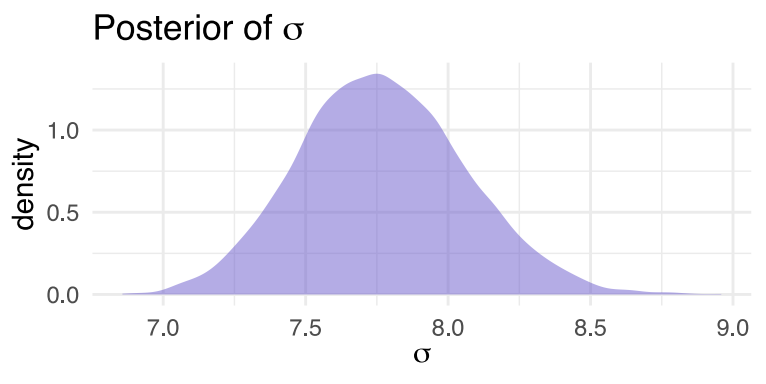
```
# dplyr needs to be loaded
posterior_draws <- sample_n(
  posterior,
  size = 1e4,
  replace = TRUE,
  weight = post
)
```




9.0
8.5
8.0
7.5
7.0
σ
153
154
155
156
μ

Approximate marginal posteriors






```
quantile(posterior_draws$mu,
         probs = c(0.05, 0.95))
```

```
##        5%       95%
## 153.9159 155.2372
```

```
quantile(posterior_draws$sigma,
         probs = c(0.05, 0.95))
```

```
##        5%       95%
## 7.307307 8.258258
```