HW9

1.

(a) If the two estimates pf the mean are fairly similar, it is probably safe to use the pooled estimate. Otherwise, the researcher should use the unpooled estimate.

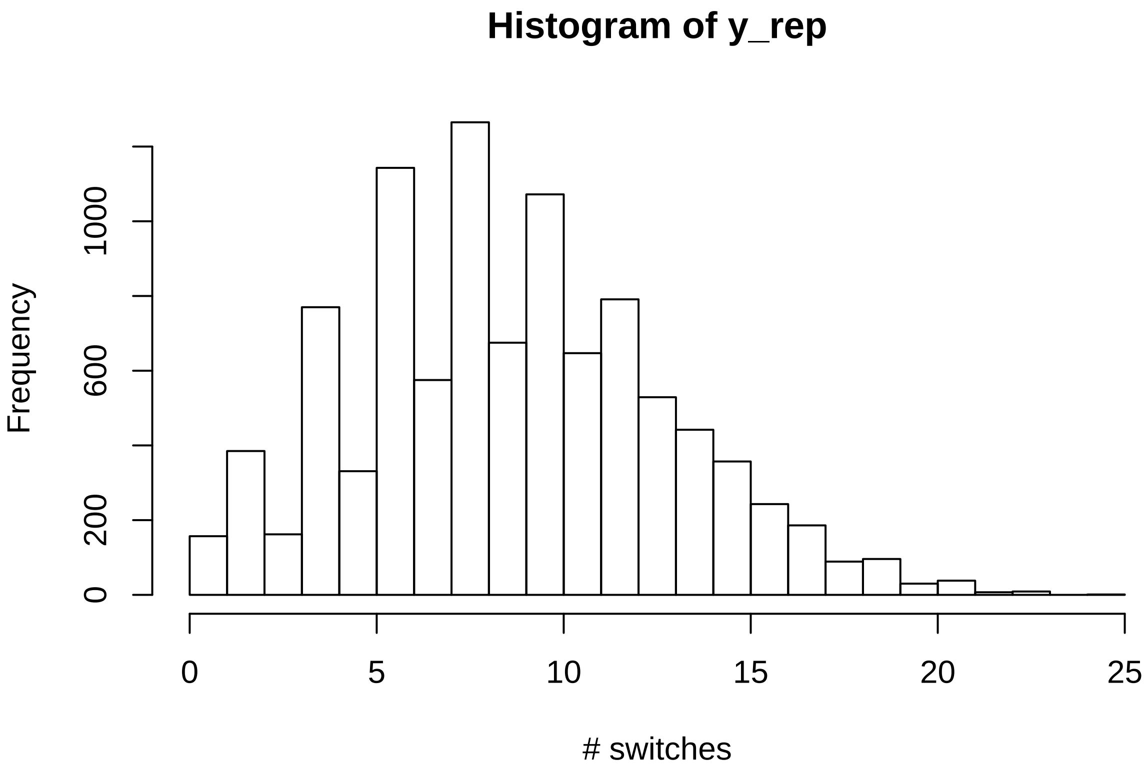
b. $data_j \sim N(\mu_j, \sigma_j)$
$\mu_j \sim N(0, \tau_j)$

2.

(a) Since we only have one experiment and the parameter θ is only dependent on n and the number of zeros, the new data generation process doesn't change inference under this assumed model.
(b) Under this data model we actually see more switches, presumably due to there often being more than 20 trials before 13 zeros are reached. This gives further evidence that a better model would include some autocorrelation

```
samples <- 10000
y_rep <- vector(mode = "numeric", length = samples)
for(s in 1:samples) {
  theta <- rbeta(1, 8, 14)
  nzeros <- 0
  nswitches <- 0
  n <- 0
  prev <- 0
  while(nzeros < 13) {
    samp <- rbinom(1, 1, theta)
    if(samp == 0){
      nzeros <- nzeros + 1
    }
    if(samp != prev & n > 0)
    {
      nswitches <- nswitches + 1
    }
    prev <- samp
    n <- n + 1
  }
  y_rep[s] <- nswitches
}

hist(y_rep, breaks = 20, freq = TRUE, xlab = "# switches")
```



3.

(a) The numerator and denominator of the Bayes factor are essentially the normalization constant a product of gaussians. Thus $p(H_1|y) = \prod_j \int N(\theta_j|y_j, \sigma_j^2) N(\theta_j|0, A^2) = \prod_j \frac{1}{\sqrt{2\pi\sigma_j^2 + A^2}} e^{-\frac{y_i^2}{2(\sigma_j^2+A^2)}}$ and $p(H_2|y) = \int \prod_j (N(\theta|y_j, \sigma_j^2)) N(\theta_j|0, A^2) = \frac{1}{2\pi^4} \sqrt{\frac{1}{\sum_j(\sigma_j^2)+A^2}} e^{-\frac{1}{2}(\sum_j \frac{y_j^2}{\sigma_j^2} - (\frac{y_j}{\sigma_j^2})^2 \frac{A^2 \prod_j \sigma_j^2}{\sum_j (\sigma_j^2)+A^2})}$.

(b) As A goes to infinity, the Bayes factor will go to infinity since the A in exponent term of $p(H_2|y)$ will make the whole term get very large while $p(H_1|y)$ will not. So regardless of the data, the pooled model will win. This also makes intuitive sense since the probability of choosing just one reasonable θ is higher than the probability of choosing n reasonable θs under the noninformative prior.

(c) As n gets really large, we can basically interpret this as A getting smaller in the prior and the prior mean getting close to $\bar{y}$ since getting more data is equivalent to having a stronger prior. Thus, for each additional data point y_j, $p(H_2|y_j, y) = \prod_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(y_i-\bar{y})^2}{2(\sigma_j^2)}}$. In contrast, $p(H_1|y) = \frac{1}{\sqrt{2\pi\sigma_j^2+A^2}} e^{-\frac{y_i^2}{2(\sigma_j^2+A^2)}}$ since each data point is essentially independent. These will be the same when $\bar{y} = 0$ and A equals the sampling variance of y. Otherwise, the pooled model will be preferred under the Bayes factor calculation.