Exercise on Mixture Models

STA 325 (exercise from BMLR)

```
library(gridExtra)
library(knitr)
library(kableExtra)
library(tidyverse)
```

1. **Mixture of two normal distributions**

Sometimes, a value may be best modeled by a mixture of two normal distributions. We would have 5 parameters in this case— $\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha$, where $0 < \alpha < 1$ is a mixing parameter determining the probability that an observation comes from the first distribution. We would then have $f(y) = \alpha\ f_1(y) + (1 - \alpha)\ f_2(y)$ (where $f_i(y)$ is the pdf of the normal distribution with μ_i, σ_i).

One phenomenon which could be modeled this way would be the waiting times between eruptions of Old Faithful geyser in Yellowstone National Park. The data can be accessed in R through `faithful`, and a histogram of wait times can be found the figure below. The MLEs of our 5 parameters would be the combination of values that produces the maximum probability of our observed data. We will need to approximate the MLE's using the EM algorithm. Find a combination of $\mu_1, \sigma_1, \mu_2, \sigma_2, \alpha$ for this distribution such that the logged likelihood is above -1050. (The command `dnorm(x, mean, sd)`, which outputs $f(y)$ assuming $Y \sim \text{N}(\mu, \sigma)$, will be helpful in calculating likelihoods.)

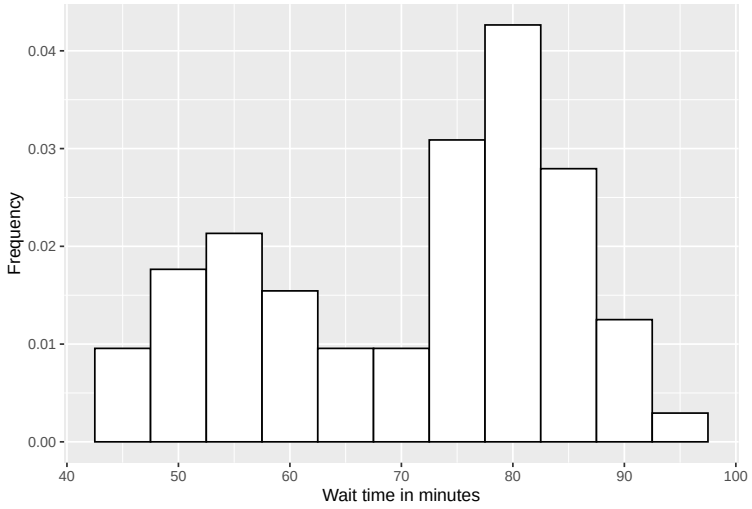

Figure 1: Waiting time between eruptions of Old Faithful.

Hint: Use the `normalmixEM()` from the `packagemixtools` package to estimate the MLE's.