ANLY 511: Homework #5

Brody Vogel

October 7th

Preparation

```
set.seed(1234)
```

Problem 33

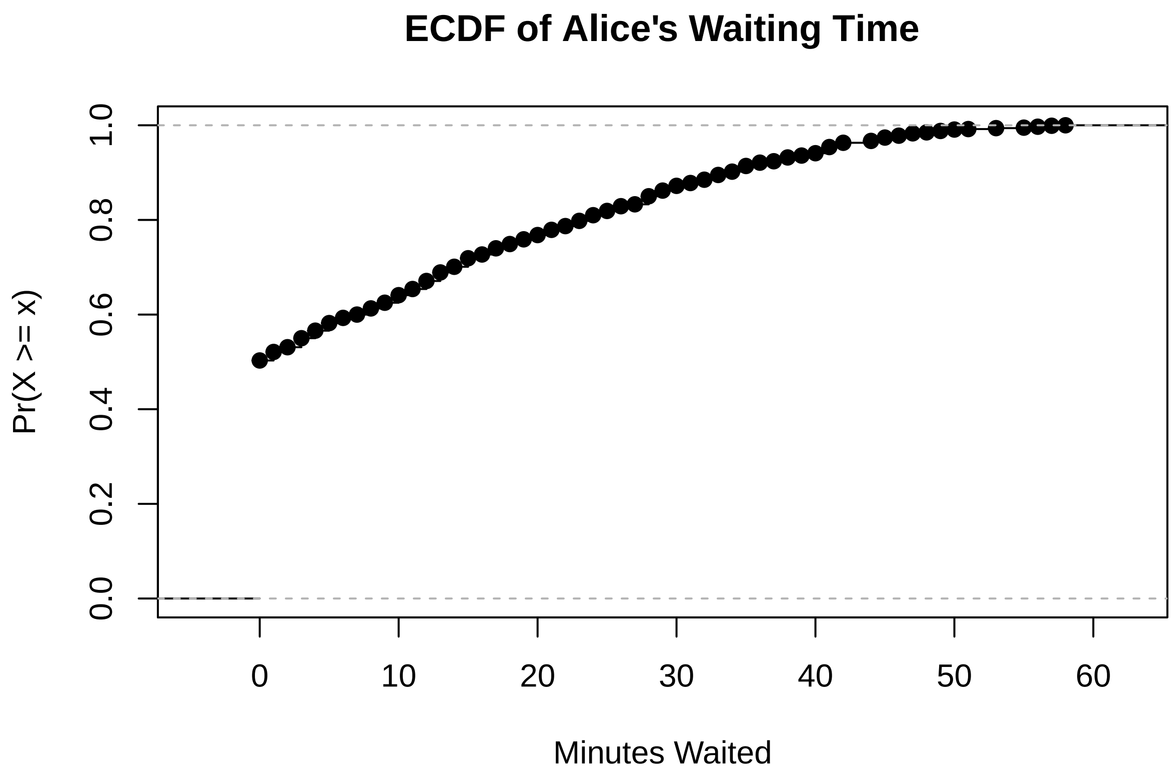
Alice's average waiting time, according to my simulation, is ≈ 10 minutes.

```
AliceWaits <- c()
while (length(AliceWaits) < 1000) {
  Alice <- round(runif(1, 0, 60), 0)
  Bob <- round(runif(1, 0, 60), 0)
  AliceWaits <- c(AliceWaits, max(Bob-Alice, 0))
}

mean(AliceWaits)
```

```
## [1] 10.422
```

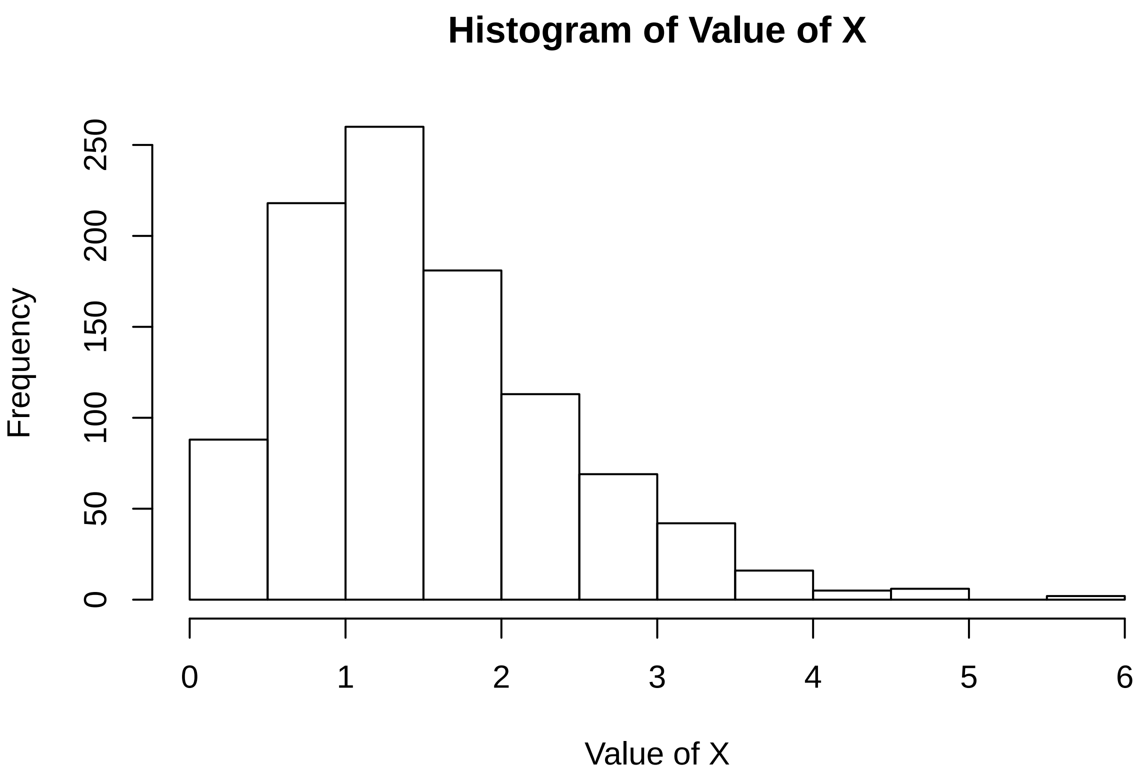
```
plot(ecdf(AliceWaits), xlab = 'Minutes Waited', ylab = 'Pr(X >= x)', main = "ECDF of Alice's Waiting Ti
```



Problem 34

```
Xs <- c()
while (length(Xs) < 1000) {
  x <- rexp(1, 1)
  y <- rpois(1, x)
  if (y == 2) {
    Xs <- c(Xs, x)
  }
}

hist(Xs, xlab = 'Value of X', main = 'Histogram of Value of X')
```



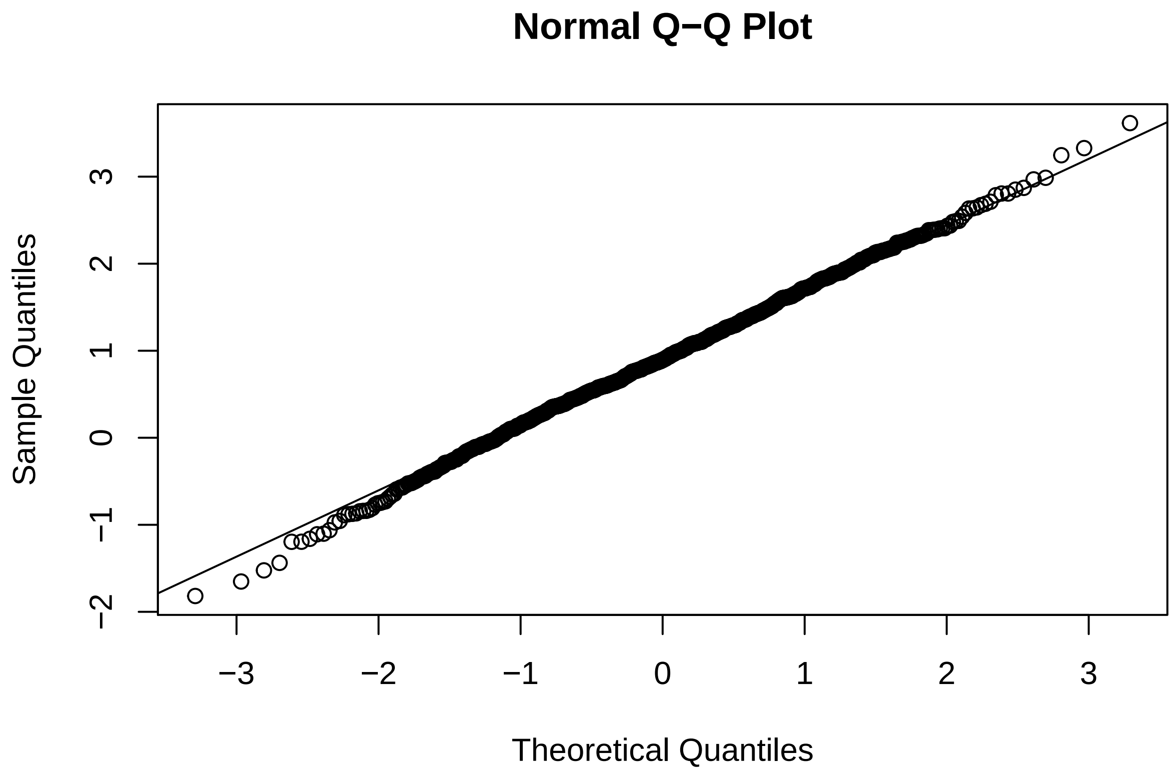
Problem 35

The approximate standard deviation of Z is $\approx .8021$, and the approximate mean is $\approx .9125$.

I think Z is normally distributed.The qqplot strays very little from the imposed normal line, and the histogram appears to have a (rougly) normal shape. Since X and Y both come from standard normal distributions, if their sum is ≥ 1, the values of X and Y that compose those sums should be normally distributed, too.

```
Zs <- c()
while (length(Zs) < 1000) {
  X <- rnorm(1)
  Y <- rnorm(1)
  if (X + Y >= 1){
    Zs <- c(Zs, Y)
  }
}
```

```r
qqnorm(Zs)
qqline(Zs)
```



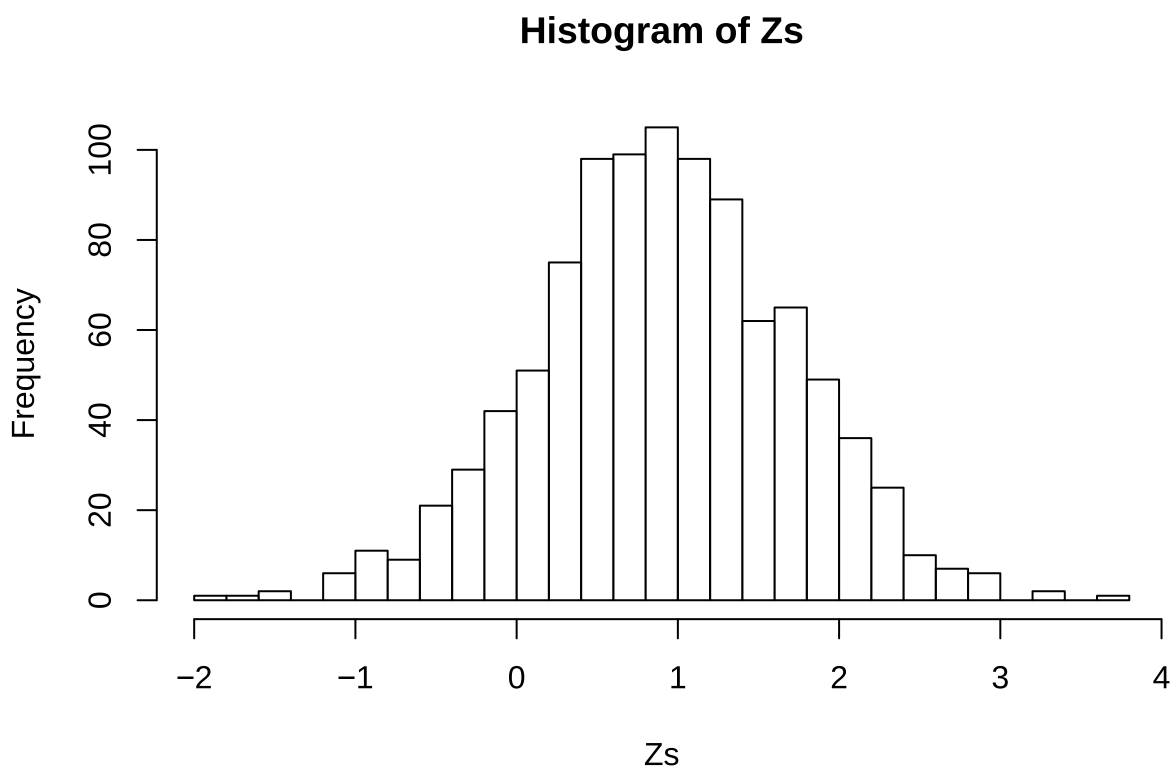

```r
mean(Zs)
```

```
## [1] 0.9124384
```

```r
sd(Zs)
```

```
## [1] 0.793504
```

```r
hist(Zs, breaks = 20)
```



Problem 36

$Y = aX + b + Z$

Because expectation is linear, $E[Y|X] = aE[X|X] + E[b|X] + E[Z|X]$.

$E[Y|X] = aE[X] + b + 0$, because expectation is linear, the expectation of a constant is just that constant, and it is given that $E[Z] = 0$.

So, if X is known as in the case of $(Y|X)$, $E[Y|X] = aX + b$.

Problem 37

a.) If we assume a point has been accepted (i.e., A has occurred), $X = Y = x$, and so $P(X = x) = P(Y = x|A)$.

b.) $P(A|Y = x) = P(U < \frac{l(x)}{M}) = \frac{l(x)}{M}$, because the probability of $\frac{l(x)}{M}$ being larger than a sample from a U(0,1) distribution is just $\frac{l(x)}{M}$.

c.) $P(A) = \sum_{x \in R} P(A|Y = x) \times P(Y = x) = \sum_{x \in R} \frac{l(x)}{M} \times \frac{1}{N} = \sum_{x \in R} \frac{cp(x)}{M} \times \frac{1}{N} = \frac{c}{MN}$, because it is given that $\sum_{x \in R} p(x) = 1$.

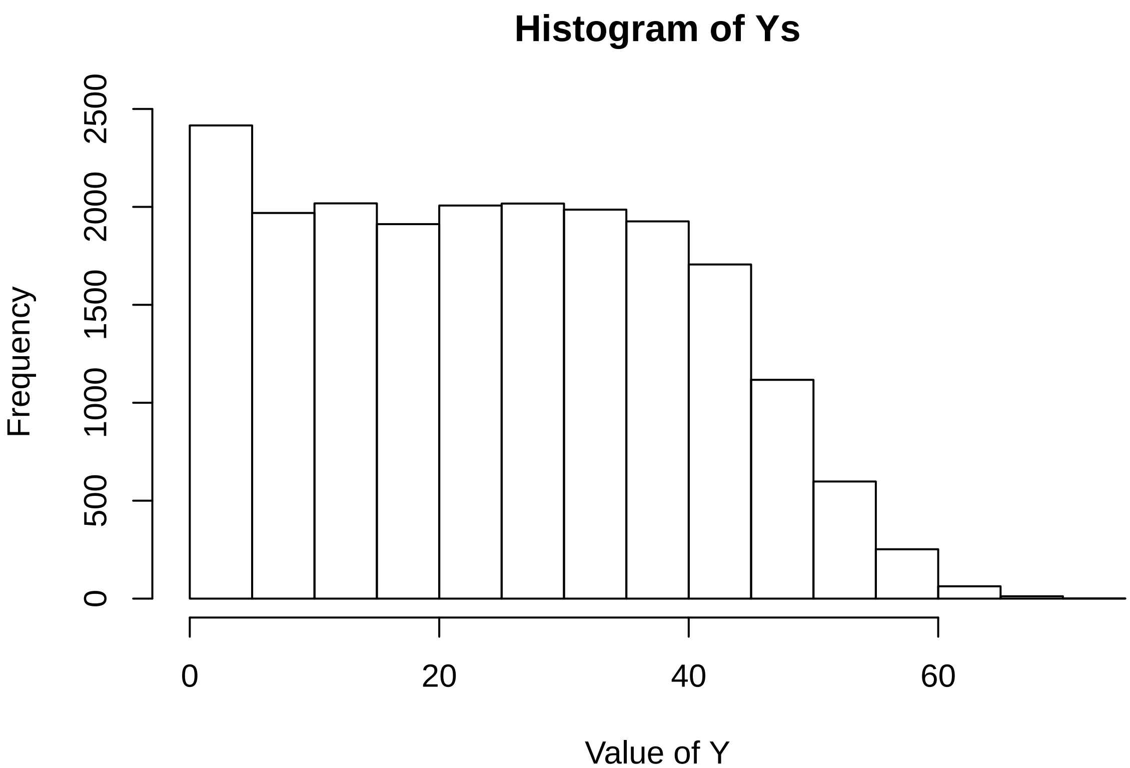
d.) $P(Y = x|A) = P(A|Y = x) \times \frac{P(Y=x)}{P(A)} = \frac{l(x)}{M} \times \frac{\frac{1}{N}}{\frac{c}{MN}} = \frac{l(x)}{M} \times \frac{M}{c} = \frac{l(x)}{c}$, by the results in a, b, and c.

So $P(X = x) = \frac{l(x)}{c}$.

Problem 38

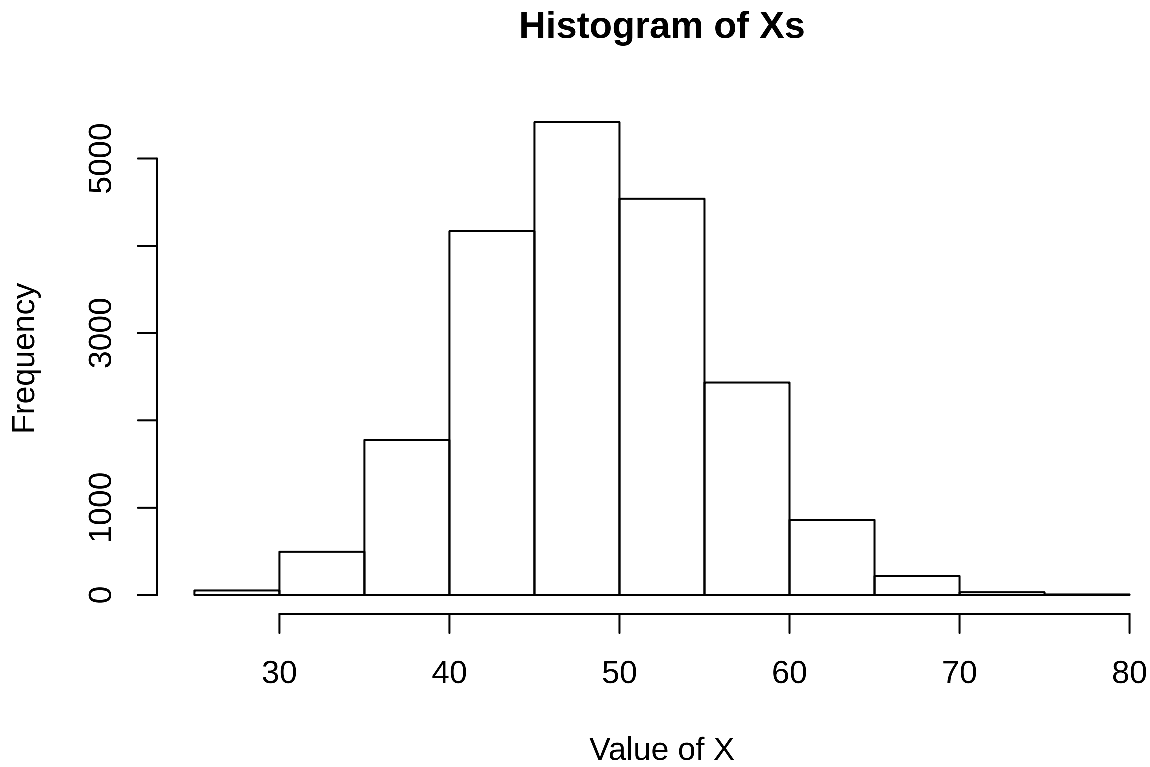
a.) Because $E[X] = xp = \lambda U$, and because U ~ U(0,1), it makes sense that we get a flat-ish histogram of $\lambda \times x \in U(0, 1)$ values.

```r
Ys <- c()
while (length(Ys) < 20000) {
  X <- rpois(1, lambda = 50)
  U <- runif(1)
  Y <- rbinom(1, X, U)
  Ys <- c(Ys, Y)
}
hist(Ys, xlab = "Value of Y")
```



b.) Because $E[X] = \lambda = 50$, and because U is evenly distributed across [0,1], it makes sense that the value of X that usually leads to a Y value of 25 is roughly normally distributed around the expected, or expected to most often occur, value of X, 50.

```r
Xs <- c()
while (length(Xs) < 20000) {
  X <- rpois(1, lambda = 50)
  U <- runif(1)
  Y <- rbinom(1, X, U)
  if (Y == 25) {
    Xs <- c(Xs, X)
  }
}
hist(Xs, xlab = 'Value of X')
```



Problem 39

1.) $E[Y|X=1]=\mu_1$; $E[Y|X=2]=\mu_2$; $E[Y]=\mu_1 w_1+\mu_2 w_2$, by the definition of expected value.

2.) The variance of a random variable is its second central moment, i.e., $Var(X)=E[X^2]-\mu^2$. And so $E[X^2]=Var(X)+\mu^2$. So, applying this, $E[Y^2|X=1]=\theta_1^2+\mu_1^2$ and $E[Y^2|X=2]=\theta_2^2+\mu_2^2$.

$Var(Y)=E[Var(Y|X)]+Var(E[Y|X])$, by the law of total variance.

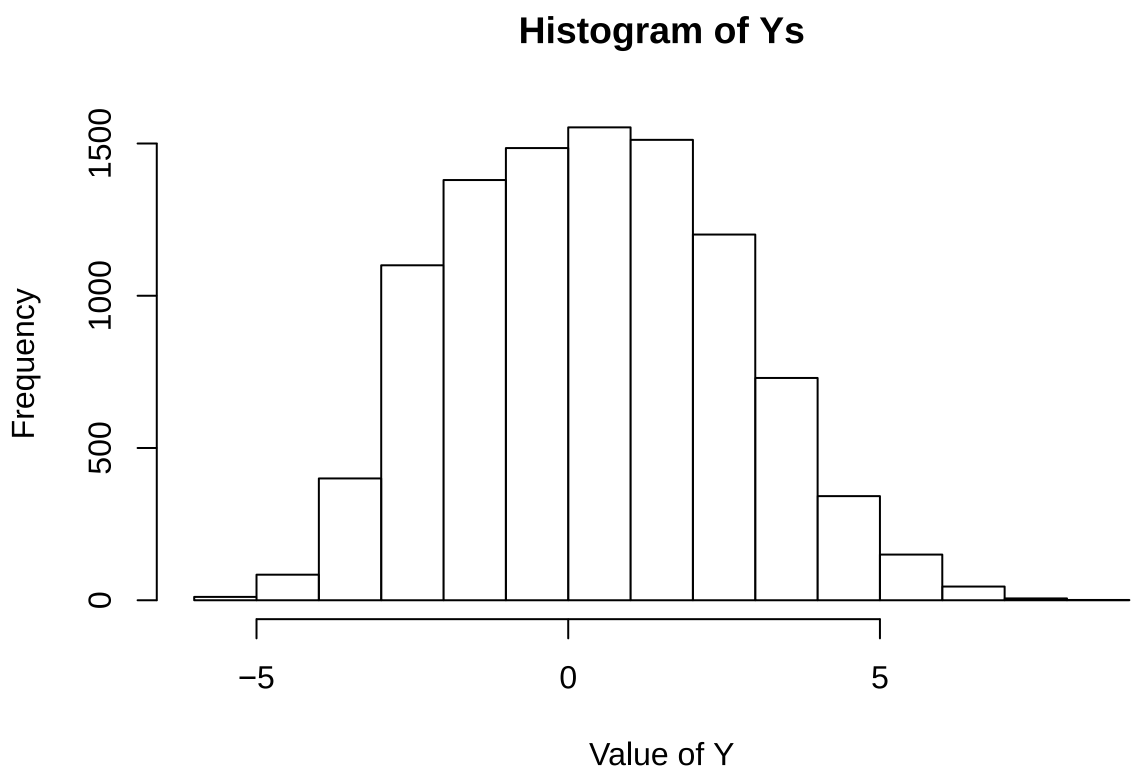
$E[Y|X=1]=\mu_1$; $E[Y|X=2]=\mu_2$, and $Var(Y|X=1)=\theta_1^2+\mu_1^2$; $Var(Y|X=2)=\theta_2^2+\mu_2^2$.

So $E[Var(Y|X)]=w_1(\theta_1^2+\mu_1^2)+w_2(\theta_2^2+\mu_2^2)$, and $Var(E[Y|X])=w_1\mu_1^2+w_2\mu_2^2-(w_1\mu_1+w_2\mu_2)^2$, by the definition of the variance of a conditional expected value, i.e., $E[Var(Y|X)]=E[(E[Y|X])^2]-(E[E[Y|X]])^2$.

And then $Var(Y)=w_1(\theta_1^2+\mu_1^2)+w_2(\theta_2^2+\mu_2^2)+w_1\mu_1^2+w_2\mu_2^2-(w_1\mu_1+w_2\mu_2)^2$

3.)

```
Ys <- c()
while (length(Ys) < 10000) {
  w = runif(1)
  if (w > .8) {
    Ys <- c(Ys, rnorm(1, -2, 1))
  }
  else {
    Ys <- c(Ys, rnorm(1, 1, 2))
  }
}
hist(Ys, xlab = 'Value of Y')
```



Problem 40

Because the sample size is large enough (≥ 30), the distribution of means will be roughly normal. Thus, we can simulate the likelihood of getting a sample mean > 51 to be $\approx .034$.

Or, standardizing the random variable $\bar{x}$, we can use the Z-table:

$Z = \frac{51-48}{\frac{9}{\sqrt{30}}} \approx 1.8257$,

which also returns: $P(\mu > 51) \approx .034$.

```
means <- c()
for (x in 1:1000) {
  means <- c(means, mean(rnorm(30, mean = 48, sd = 9)))
}

1 - pnorm(51, mean(means), sd(means))
```

```
## [1] 0.03777215
```

```
### The question says to use dnorm(), so I'll include that, too, although I'm not sure
### what it means

dnorm(51, mean(means), sd(means))
```

```
## [1] 0.04986356
```