In lecture 16, we looked at precipitation amounts in Madison County (at Morrisville station). We found that the Weibull distribution had a good fit to the monthly precipitation amounts.

We found that the MLEs for the Weibull distribution were

$$\hat{a} = 2.1871$$
$$\hat{\sigma} = 3.9683$$

and

$$-\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = 2166.496$$

is the realized negative log-likelihood. Note this means that the log-likelihood is

$$\mathcal{L}(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = -2166.496,$$

and the usual likelihood is

$$L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x}) = e^{[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x})]} \approx= e^{-2166.496},$$

which `R` cannot differentiate from 0.

1. Someone asked "why Weibull?" in class. That is, why wouldn't we use another right-skewed distribution like the Gamma (see Lecture 15), or the Log-Normal (see Lecture 17).

   (a) Compute the MLEs for these data using a Gamma distribution.
   Solution: The computed MLEs for these data using the gamma distribution turned out to be

   $$\hat{\alpha} = 4.17, \hat{\beta} = 1.19$$

   (b) Compute the MLEs for these data using the Log-Normal distribution.
   Solution: The computed MLEs for these data using the gamma distribution turned out to be

   $$\hat{\mu} = 1.131, \hat{\sigma} = 0.533$$

   (c) Compute the likelihood ratio to compare the Weibull and the Gamma distribution. Which has a better fit according to the likelhiood ratio?
   Solution: Since the likelihood ratio is greater than 1 that means that the Weibull distribution is a better fit for the data than the Gamma distribution.

   $$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\alpha},\hat{\beta}\}|\mathbf{x})\right]} = 1.0071$$

   (d) Compute the likelihood ratio to compare the Weibull and the Log-Normal distribution. Which has a better fit according to the likelihood ratio?
   Solution: Since the likelihood ratio is less than 1 that means that the Log-Normal distribution is a better fit for the data than the Weibull distribution.

   $$Q = \frac{L(\{\hat{a}, \hat{\sigma}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{[\mathcal{L}(\{\hat{a},\hat{\sigma}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\mu},\hat{\sigma}\}|\mathbf{x})]} = 0.9828$$

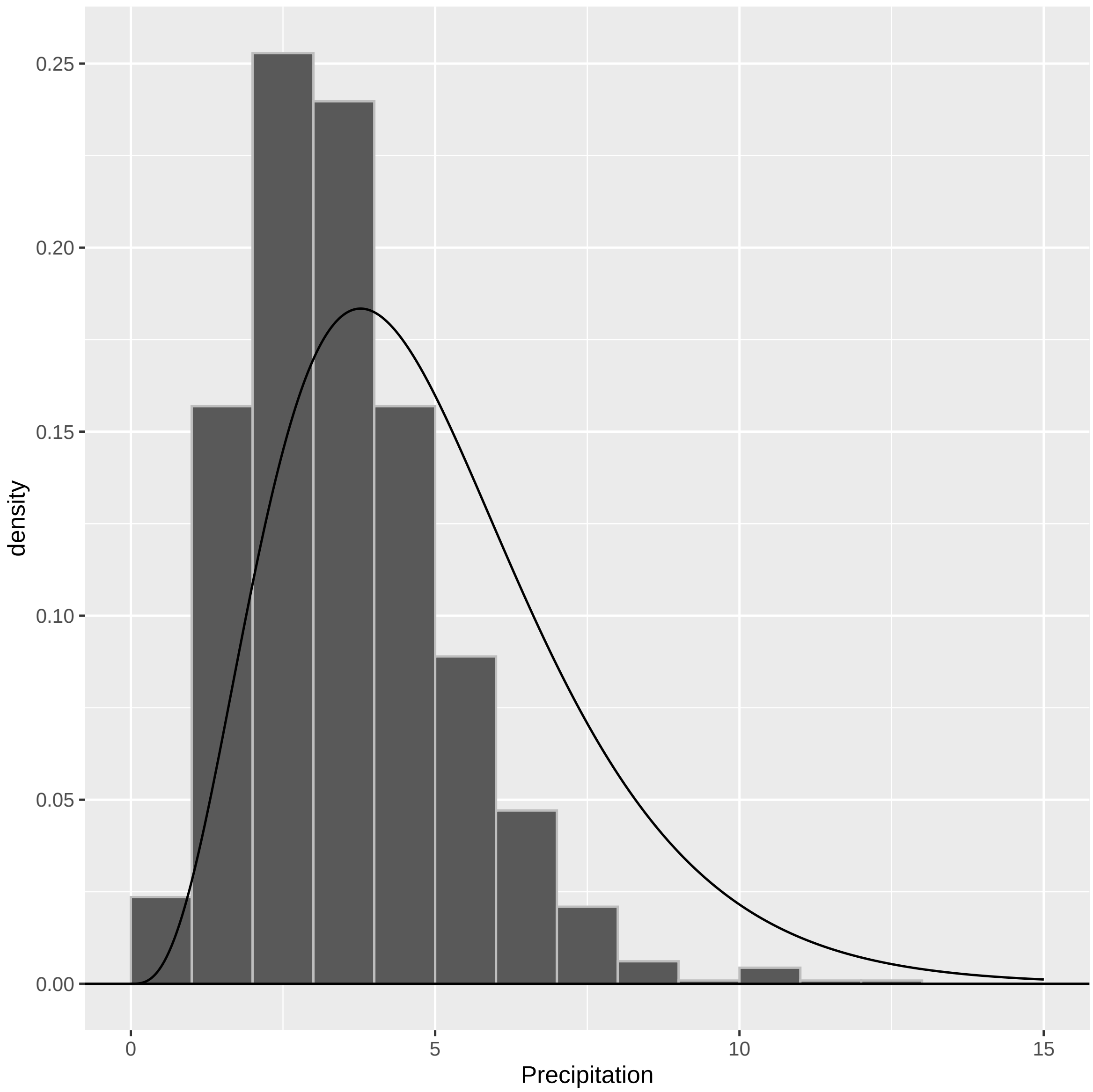
   (e) Compute the likelihood ratio to compare the Gamma and the Log-Normal distribution. Which has a better fit according to the likelhiood ratio?
   Solution: Since the likelihood ratio is less than 1 that means that the Log-Normal distribution is a better fit for the data than the Gamma distribution.

   $$Q = \frac{L(\{\hat{\alpha}, \hat{\beta}\}|\mathbf{x})}{L(\{\hat{\mu}, \hat{\sigma}\}|\mathbf{x})} = e^{\left[\mathcal{L}(\{\hat{\alpha},\hat{\beta}\}|\mathbf{x}) - \mathcal{L}(\{\hat{\mu},\hat{\sigma}\}|\mathbf{x})\right]} = 0.9759$$

2. **CODE:**

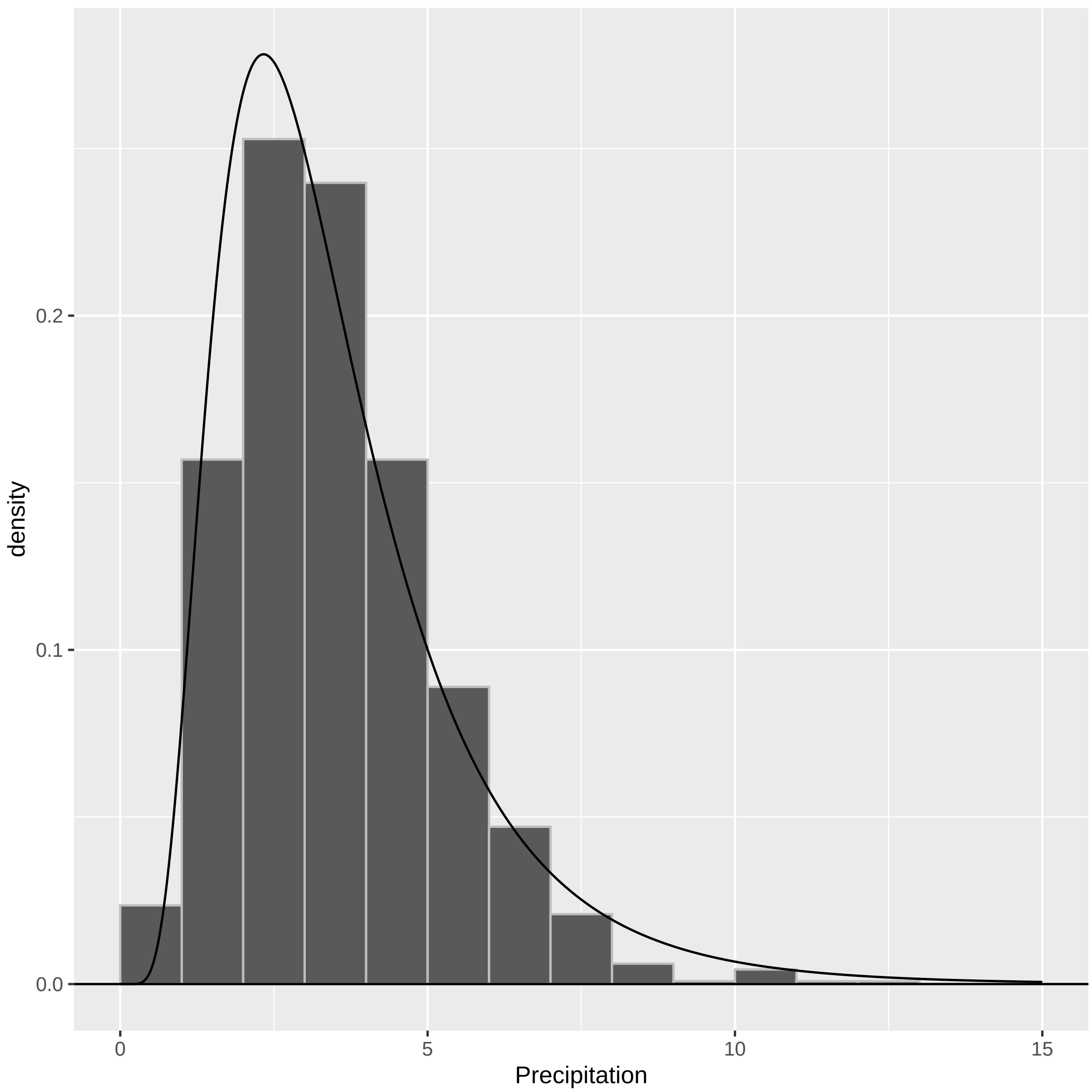
```r
library(tidyverse)
library(nleqslv)
library(patchwork)

dat = read_csv("agacis.csv")

dat.long <- dat |>
  dplyr::select(-Annual) |>                    # Remove annual column
  pivot_longer(cols = c(Jan, Feb, Mar, Apr,    # pivot the column data into one col
                        May, Jun, Jul, Aug,
                        Sep, Oct, Nov, Dec),
               values_to = "Precipitation",    # store the values in Precipitation
               names_to = "Month") |>          # store the months in Month
  mutate(Precipitation = case_when(Precipitation == "M" ~ NA_character_,
                                   TRUE                 ~ Precipitation))|>
  mutate(Precipitation = as.numeric(Precipitation))

### Weibull
llweibull <- function(par, data, neg=F){
  # a <- par[1]
  # sigma <- par[2]
  a <- exp(par[1]) # go from (-inf,inf) to (0,inf)
  sigma <- exp(par[2]) # go from (-inf,inf) to (0,inf)

  ll <- sum(log(dweibull(x=data, shape=a, scale=sigma)), na.rm=T)

  return(ifelse(neg, -ll, ll))
}

weibulls = optim(fn = llweibull,
              par = c(1,1),
              data = dat.long$Precipitation,
              neg=T)

weibull.a = weibulls$par[1]
weibull.sigma = weibulls$par[2]

### Part A
gamma.MLE = function(data, par, neg=F){
  a = par[1]
  b = par[2]

  loglik <- sum(log(dgamma(x=data, shape = a, rate = b)), na.rm = T)

  return(ifelse(neg, -loglik, loglik))
}

gammas = optim(par = c(1, 1),
              fn = gamma.MLE,
              data=dat.long$Precipitation,
              neg=T)

gamma.a = gammas$par[1]
```

```r
gamma.b = gammas$par[2]

dat.gamma <- tibble(x = seq(0,15,length.out=1000)) |>
  mutate(pdf.mle = dgamma(x=x, shape=gammas$par[1], scale=gammas$par[2]))

ggplot() +
  geom_histogram(data = dat.long, aes(x = Precipitation, y = after_stat(density)),
  breaks=seq(0, 15, 1),
  color="grey")+
  geom_hline(yintercept = 0)+
  geom_line(data = dat.gamma, aes(x = x, y = pdf.mle))
```




```r
  ###### PART B
lnorm.MLE = function(data, par, neg=F){
  mu = par[1]
```

```r
  sigma = par[2]

  loglik <- sum(log(dlnorm(x=data, meanlog = mu, sdlog = sigma)), na.rm = T)

  return(ifelse(neg, -loglik, loglik))
}

lnorms = optim(par = c(1, 1),
               fn = lnorm.MLE,
               data=dat.long$Precipitation,
               neg=T)

lnorm.mu = lnorms$par[1]
lnorm.sd = lnorms$par[2]

dat.lnorm <- tibble(x = seq(0,15,length.out=1000)) |>
  mutate(pdf.mle = dlnorm(x=x, meanlog=lnorms$par[1], sdlog=lnorms$par[2]))



ggplot() +
  geom_histogram(data = dat.long, aes(x = Precipitation, y = after_stat(density)),
                 breaks=seq(0, 15, 1),
                 color="grey")+
  geom_hline(yintercept = 0)+
  geom_line(data = dat.lnorm, aes(x = x, y = pdf.mle))
```




```
#### Part C

(weibull.gamma = llweibull(dat.long$Precipitation, par = weibulls$par)
                /gamma.MLE(dat.long$Precipitation, par = gammas$par))

[1] 1.007134

#### Part D

(weibull.lnorm = llweibull(dat.long$Precipitation, par = weibulls$par)
  /lnorm.MLE(dat.long$Precipitation, par = lnorms$par))

[1] 0.982894

#### Part E
```

```
(gamma.lnorm = gamma.MLE(dat.long$Precipitation, par = gammas$par)
  /lnorm.MLE(dat.long$Precipitation, par = lnorms$par))

[1] 0.9759315
```