M5_AI1

Sofia Cantu

2024-10-22

Introducción

La comprensión de los patrones de precipitación es esencial en la ingeniería civil para el diseño y construcción de estructuras hidráulicas como presas, puentes y sistemas de drenaje. Las obras hidráulicas deben ser capaces de soportar eventos extremos de precipitación para garantizar su seguridad y funcionalidad a largo plazo. Este estudio se centra en el análisis de las precipitaciones máximas mensuales históricas en el estado de Aguascalientes (1994-2023) con el objetivo de calcular la precipitación más extrema asociada a un periodo de retorno específico.

```r
# Librerías
if (!require(ggplot2)) install.packages("ggplot2")

## Loading required package: ggplot2

library(ggplot2)
if (!require(dplyr)) install.packages("dplyr")

## Loading required package: dplyr

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(dplyr)
if (!require(MASS)) install.packages("MASS")

## Loading required package: MASS

##
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':
##
##     select
```

```r
library(MASS)
if (!require(tidyr)) install.packages("tidyr")

## Loading required package: tidyr

library(tidyr)
if (!require(fitdistrplus)) install.packages("fitdistrplus")

## Loading required package: fitdistrplus

## Loading required package: survival

library(fitdistrplus)
if (!require(extraDistr)) install.packages("extraDistr")

## Loading required package: extraDistr

library(extraDistr)
if (!require(evd)) install.packages("evd")

## Loading required package: evd

##
## Attaching package: 'evd'

## The following objects are masked from 'package:extraDistr':
##
##     dfrechet, dgev, dgpd, dgumbel, pfrechet, pgev, pgpd, pgumbel,
##     qfrechet, qgev, qgpd, qgumbel, rfrechet, rgev, rgpd, rgumbel

library(evd)
```

1. Análisis estadístico descriptivo de las precipitaciones históricas máximas mensuales de un estado

```r
precipitaciones <- read.table("~/Downloads/ArchivosCodigos/precipitaciones_ma
ximas_mensuales.txt", header = TRUE, sep = "\t")
head(precipitaciones, 10)

##    Anio Mes              Estado Lluvia
## 1  1994 Ene      Aguascalientes    8.3
## 2  1994 Ene     Baja.California   10.3
## 3  1994 Ene Baja.California.Sur    0.0
## 4  1994 Ene            Campeche   85.4
## 5  1994 Ene    Ciudad.de.México   17.7
## 6  1994 Ene            Coahuila   12.8
## 7  1994 Ene              Colima    0.0
## 8  1994 Ene             Chiapas   46.1
## 9  1994 Ene           Chihuahua    1.2
## 10 1994 Ene             Durango    3.5
```

1.A Descarga la base de datos de precipitaciones máximas históricas mensuales de todos los estados de la república. Selecciona un estado que sea diferente a los del resto de tu equipo.

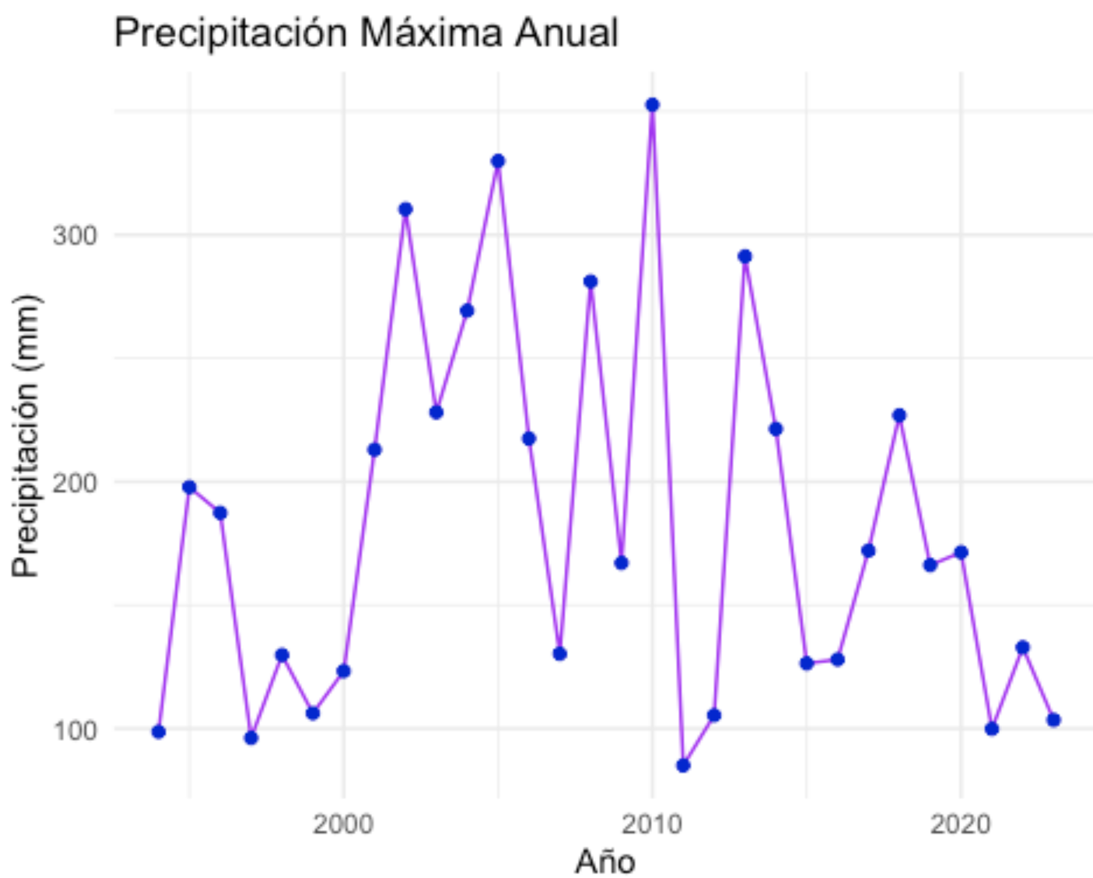
```
estado_data <- precipitaciones %>% filter(Estado == 'Nuevo.León')
head(estado_data, 10)
```

```
##    Anio Mes     Estado Lluvia
## 1  1994 Ene Nuevo.León   60.8
## 2  1994 Feb Nuevo.León    9.0
## 3  1994 Mar Nuevo.León   27.1
## 4  1994 Abr Nuevo.León   16.7
## 5  1994 May Nuevo.León   40.4
## 6  1994 Jun Nuevo.León   48.0
## 7  1994 Jul Nuevo.León   73.8
## 8  1994 Ago Nuevo.León   89.3
## 9  1994 Sep Nuevo.León   98.7
## 10 1994 Oct Nuevo.León   50.0
```

1.B Elabora una gráfica de las precipitaciones máximas mensuales por año para tu estado. Para ello deberás calcular la precipitación mensual máxima de cada año y graficarla.

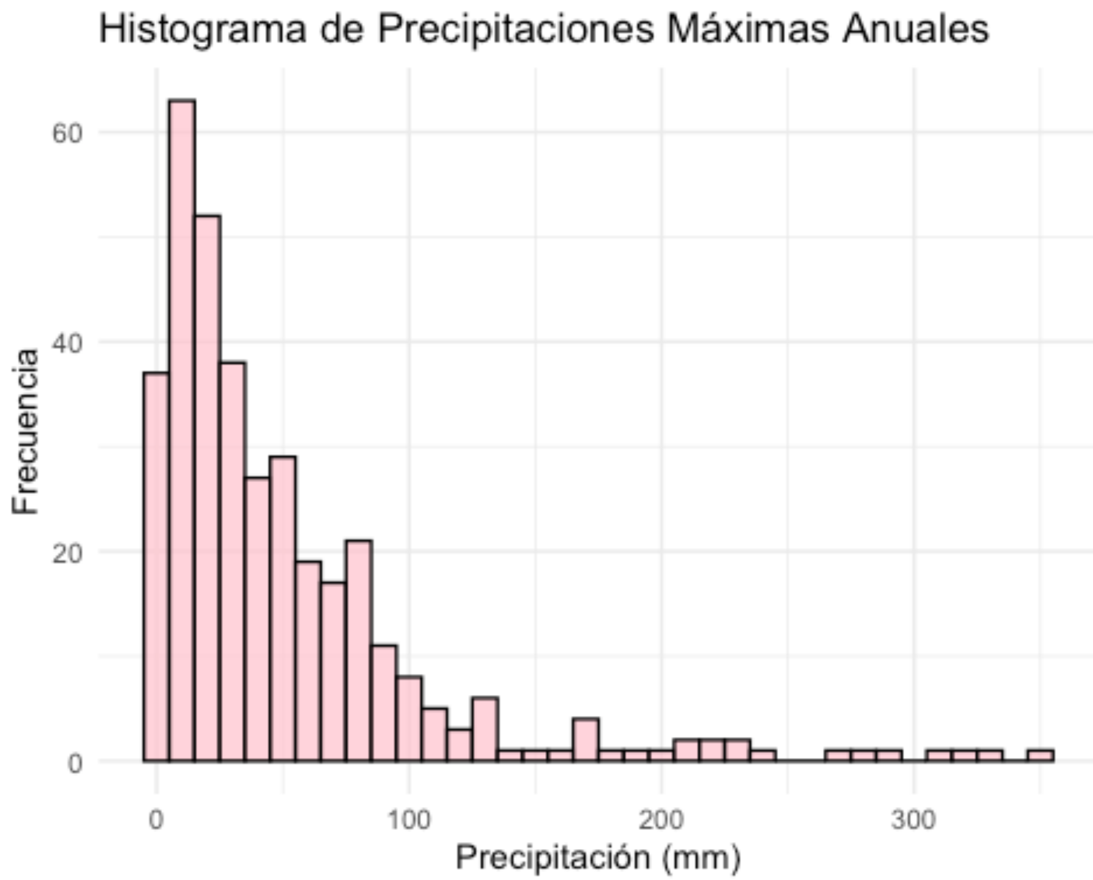
```
# Medidas de centralización y variación
media <- mean(estado_data$Lluvia, na.rm = TRUE)
mediana <- median(estado_data$Lluvia, na.rm = TRUE)
desviacion <- sd(estado_data$Lluvia, na.rm = TRUE)

# Imprimir resultados
cat("Media:", media, "\n")

## Media: 50.80361

cat("Mediana:", mediana, "\n")

## Mediana: 31.05

cat("Desviación estándar:", desviacion, "\n")

## Desviación estándar: 58.08207

# Calcular la precipitación mensual máxima por año
precipitacion_maxima_anual <- estado_data %>%
  group_by(Anio) %>%
  summarise(MaximaPrecipitacion = max(Lluvia, na.rm = TRUE))

# Gráfico de la precipitación máxima anual por año
ggplot(precipitacion_maxima_anual, aes(x = Anio, y = MaximaPrecipitacion)) +
  geom_line(color = "purple") +
  geom_point(color = "blue3") +
  labs(title = "Precipitación Máxima Anual", x = "Año", y = "Precipitación (m
```

```
m)") +
  theme_minimal()
```

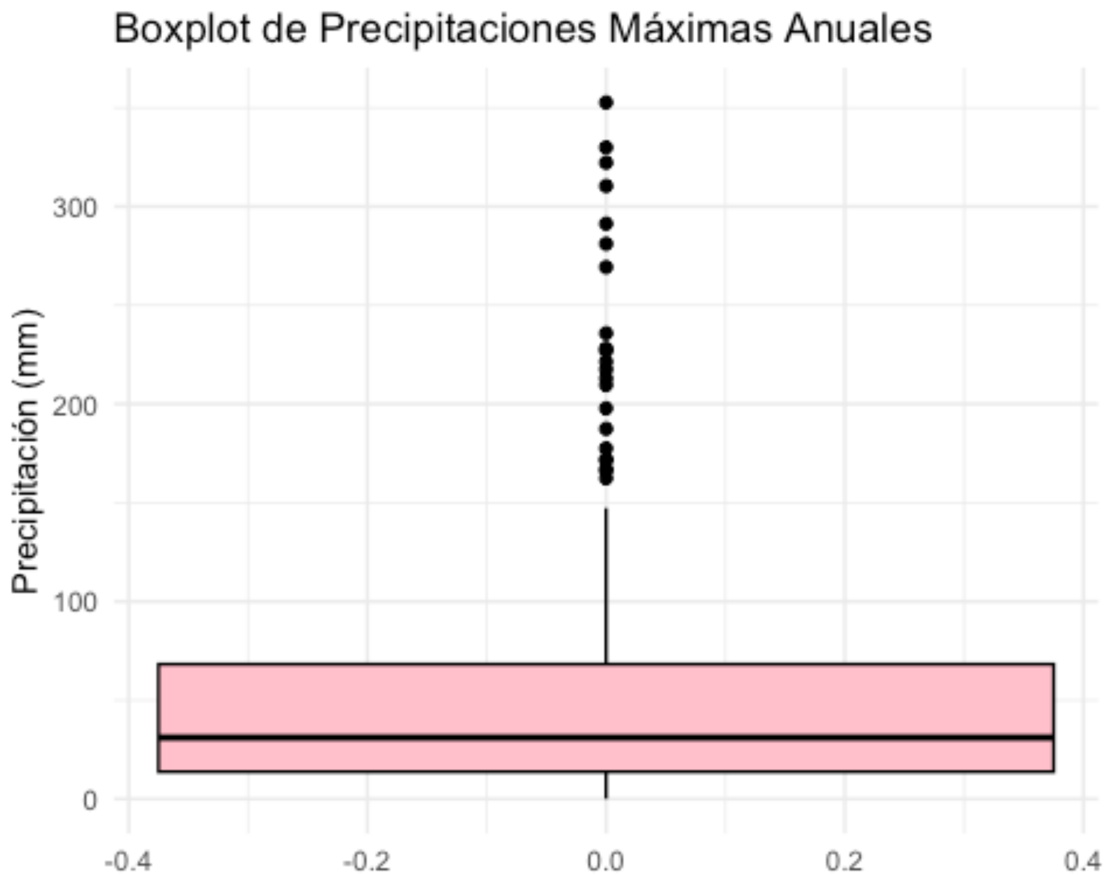


1.C Analiza los datos de precipitaciones máximas mensuaels del estado seleccionado.

```
# Histograma de las precipitaciones máximas anuales
ggplot(estado_data, aes(x = Lluvia)) +
  geom_histogram(binwidth = 10, fill = "pink", color = "black", alpha = 0.7)
+
  labs(title = "Histograma de Precipitaciones Máximas Anuales", x = "Precipit
ación (mm)", y = "Frecuencia") +
  theme_minimal()
```




```
# Boxplot de las precipitaciones máximas anuales
ggplot(estado_data, aes(y = Lluvia)) +
  geom_boxplot(fill = "pink", color = "black") +
  labs(title = "Boxplot de Precipitaciones Máximas Anuales", y = "Precipitaci
ón (mm)") +
  theme_minimal()
```




```r
# Análisis de la distribución
library(e1071)
sesgo <- skewness(estado_data$Lluvia, na.rm = TRUE)
cat("Sesgo de la distribución:", sesgo, "\n")

## Sesgo de la distribución: 2.450139

# Describir la distribución
if (sesgo > 0) {
  cat("La distribución está sesgada a la derecha.\n")
} else if (sesgo < 0) {
  cat("La distribución está sesgada a la izquierda.\n")
} else {
  cat("La distribución es simétrica.\n")
}

## La distribución está sesgada a la derecha.
```

Describe el comportamiento de la distribución: centralización, sesgo, variación

Centralización: - Media: La media de las precipitaciones es de 50.80 mm, lo que indica el valor promedio de las precipitaciones máximas anuales en tu conjunto de datos. - Mediana: La mediana es 31.05 mm, lo que sugiere que la mitad de los valores está por debajo de esta cifra y la otra mitad está por encima. Esto también indica que la distribución está algo sesgada, ya que la media es considerablemente más alta que la mediana.

Sesgo: - Sesgo de la distribución: El sesgo es de 2.45, lo que indica que la distribución está sesgada a la derecha. Esto significa que hay valores atípicos de precipitaciones altas que están influyendo en la media y alejándola de la mediana. - En las gráficas de histograma y boxplot, también se puede observar la presencia de valores atípicos hacia la derecha (precipitaciones mayores), lo cual refuerza el sesgo a la derecha.

Variación: - Desviación estándar: La desviación estándar es 58.08 mm, lo que indica una alta variabilidad en las precipitaciones máximas anuales. Esto significa que las precipitaciones pueden variar significativamente de un año a otro.

1.D ¿Qué puedes concluir observando la gráfica de los máximos mensuales anuales para tu Estado? ¿Observas alguna tendencia? ¿Puedes concluir que cada determinado número de años la cantidad de precipitación sube o baja? ¿Para qué nos sirve analizar este tipo de gráficas?

Conclusiones al observar las gráficas: - Tendencias observadas: En la gráfica de la precipitación máxima anual por año, no parece haber una tendencia clara y consistente, ya que la precipitación fluctúa significativamente de un año a otro. Existen picos altos en ciertos años (e.g., cerca de 2010 y 1998), pero no parece haber un patrón cíclico claro que indique un aumento o disminución consistente en las precipitaciones en intervalos de tiempo regulares. - Patrones a lo largo del tiempo: Aunque hay variaciones significativas, no se puede concluir que las precipitaciones suben o bajan de manera predecible cada cierto número de años, debido a la naturaleza errática de los valores máximos. - Utilidad del análisis: Analizar este tipo de gráficas y datos nos permite comprender la magnitud de las precipitaciones extremas en un área a lo largo del tiempo, lo cual es crucial para diseñar infraestructura resistente (e.g., presas, carreteras, puentes) que pueda soportar eventos climáticos extremos. También ayuda en la planificación de medidas de prevención para mitigar riesgos asociados con inundaciones o sequías.

2. Análisis de Frecuencias Método Gráfico

2.A En el data frame de los datos de precipitación máxima se agrega una columna con los datos de lluvias máximas ordenados de mayor a menor.

```r
# Seleccionar el estado específico (ajustar el estado según tus datos)
estado_data <- precipitaciones %>% filter(Estado == 'Nuevo.León')

# Ordenar los datos de precipitación máxima de mayor a menor
estado_data <- estado_data %>% arrange(desc(Lluvia))

# Agregar la columna con el número de orden (rank)
estado_data <- estado_data %>%
```

```
  mutate(rank = row_number())
head(estado_data, 10)
```

```
##    Anio Mes     Estado Lluvia rank
## 1  2010 Jul Nuevo.León  352.7    1
## 2  2005 Jul Nuevo.León  329.9    2
## 3  2010 Sep Nuevo.León  322.2    3
## 4  2002 Sep Nuevo.León  310.4    4
## 5  2013 Sep Nuevo.León  291.3    5
## 6  2008 Sep Nuevo.León  281.1    6
## 7  2004 Sep Nuevo.León  269.3    7
## 8  2005 Oct Nuevo.León  235.8    8
## 9  2003 Sep Nuevo.León  228.1    9
## 10 2018 Sep Nuevo.León  226.9   10
```

2.B Se agrega una columna con el número de orden que tiene asignado cada precipitación máxima. A ese número se le llama “rank” (rango en español) y se simboliza por m

```
M <- nrow(estado_data)  # Total de datos
estado_data <- estado_data %>%
  mutate(prob_excedencia = rank / (M + 1))
M
```

```
## [1] 360
```

2.C Se calcula la probabilidad de excedencia o de ocurrencia de acuerdo con Weibull,

```
estado_data <- estado_data %>%
  mutate(prob_no_excedencia = 1 - prob_excedencia)
head(estado_data, 10)
```

```
##    Anio Mes     Estado Lluvia rank prob_excedencia prob_no_excedencia
## 1  2010 Jul Nuevo.León  352.7    1     0.002770083          0.9972299
## 2  2005 Jul Nuevo.León  329.9    2     0.005540166          0.9944598
## 3  2010 Sep Nuevo.León  322.2    3     0.008310249          0.9916898
## 4  2002 Sep Nuevo.León  310.4    4     0.011080332          0.9889197
## 5  2013 Sep Nuevo.León  291.3    5     0.013850416          0.9861496
## 6  2008 Sep Nuevo.León  281.1    6     0.016620499          0.9833795
## 7  2004 Sep Nuevo.León  269.3    7     0.019390582          0.9806094
## 8  2005 Oct Nuevo.León  235.8    8     0.022160665          0.9778393
## 9  2003 Sep Nuevo.León  228.1    9     0.024930748          0.9750693
## 10 2018 Sep Nuevo.León  226.9   10     0.027700831          0.9722992
```

2.D Se calcula la probabilidad de no excedencia para cada precipitación

```
estado_data <- estado_data %>%
  mutate(periodo_retorno = 1 / prob_excedencia)
head(estado_data, 10)
```

```
##    Anio Mes     Estado Lluvia rank prob_excedencia prob_no_excedencia
## 1  2010 Jul Nuevo.León  352.7    1     0.002770083          0.9972299
## 2  2005 Jul Nuevo.León  329.9    2     0.005540166          0.9944598
## 3  2010 Sep Nuevo.León  322.2    3     0.008310249          0.9916898
## 4  2002 Sep Nuevo.León  310.4    4     0.011080332          0.9889197
## 5  2013 Sep Nuevo.León  291.3    5     0.013850416          0.9861496
## 6  2008 Sep Nuevo.León  281.1    6     0.016620499          0.9833795
## 7  2004 Sep Nuevo.León  269.3    7     0.019390582          0.9806094
## 8  2005 Oct Nuevo.León  235.8    8     0.022160665          0.9778393
## 9  2003 Sep Nuevo.León  228.1    9     0.024930748          0.9750693
## 10 2018 Sep Nuevo.León  226.9   10     0.027700831          0.9722992
##    periodo_retorno
## 1        361.00000
## 2        180.50000
## 3        120.33333
## 4         90.25000
## 5         72.20000
## 6         60.16667
## 7         51.57143
## 8         45.12500
## 9         40.11111
## 10        36.10000
```

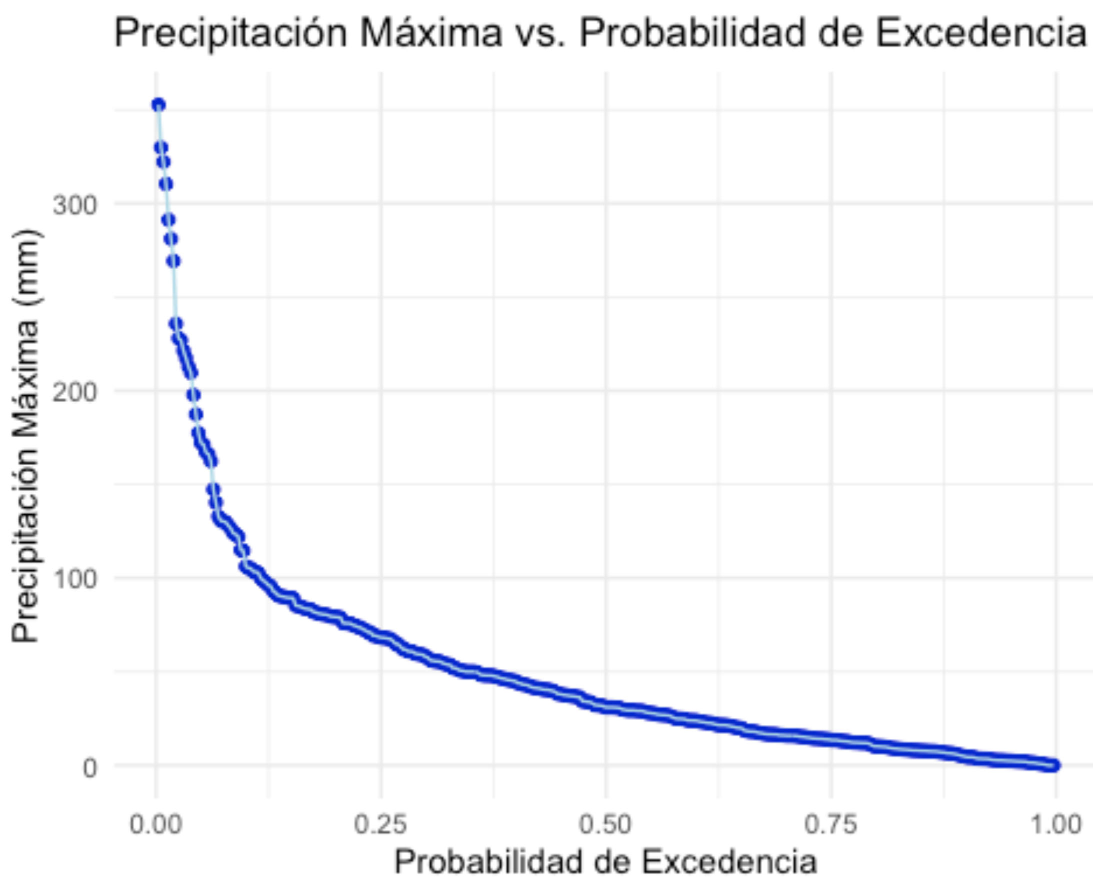
2.E Se calcula el periodo de retorno como el inverso de la probabilidad de excedencia

```r
# Visualizar los primeros resultados para confirmar
print(head(estado_data))
```

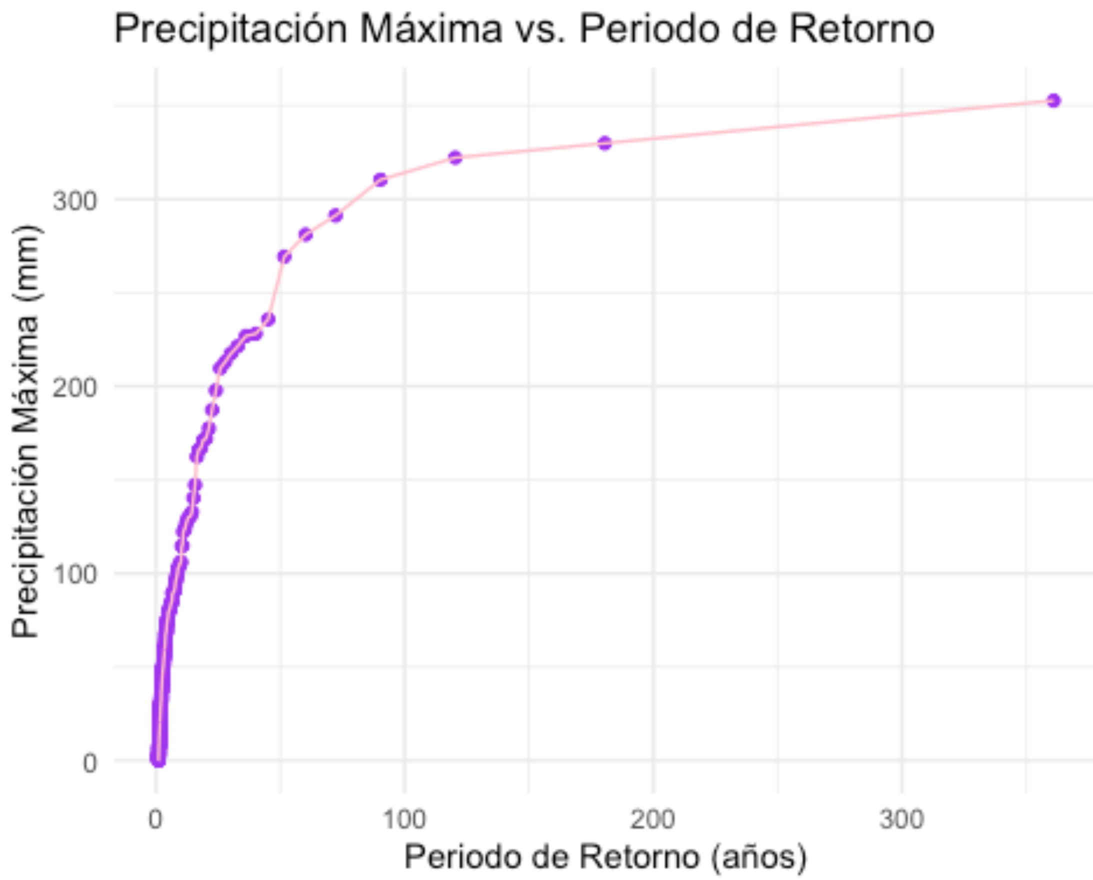
```
##   Anio Mes     Estado Lluvia rank prob_excedencia prob_no_excedencia
## 1 2010 Jul Nuevo.León  352.7    1     0.002770083          0.9972299
## 2 2005 Jul Nuevo.León  329.9    2     0.005540166          0.9944598
## 3 2010 Sep Nuevo.León  322.2    3     0.008310249          0.9916898
## 4 2002 Sep Nuevo.León  310.4    4     0.011080332          0.9889197
## 5 2013 Sep Nuevo.León  291.3    5     0.013850416          0.9861496
## 6 2008 Sep Nuevo.León  281.1    6     0.016620499          0.9833795
##   periodo_retorno
## 1       361.00000
## 2       180.50000
## 3       120.33333
## 4        90.25000
## 5        72.20000
## 6        60.16667
```

```r
# Gráfica de la precipitación máxima frente a la probabilidad de excedencia
ggplot(estado_data, aes(x = prob_excedencia, y = Lluvia)) +
  geom_point(color = "blue3") +
  geom_line(color = "lightblue") +
  labs(title = "Precipitación Máxima vs. Probabilidad de Excedencia",
       x = "Probabilidad de Excedencia",
```

```
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```




```
# Gráfica del periodo de retorno frente a la precipitación máxima
ggplot(estado_data, aes(x = periodo_retorno, y = Lluvia)) +
  geom_point(color = "purple") +
  geom_line(color = "pink") +
  labs(title = "Precipitación Máxima vs. Periodo de Retorno",
       x = "Periodo de Retorno (años)",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```



1. Descripción de las gráficas: Precipitación Máxima vs. Probabilidad de Excedencia:

- En esta gráfica, se muestra cómo la precipitación máxima disminuye conforme aumenta la probabilidad de excedencia. Las precipitaciones más altas tienen una probabilidad muy baja de exceder ciertos umbrales, mientras que las precipitaciones más bajas tienen mayor probabilidad de ser superadas. Precipitación Máxima vs. Periodo de Retorno:
- En este gráfico, se visualiza que los eventos con precipitaciones máximas más altas están asociados a periodos de retorno mayores. A medida que el periodo de retorno aumenta, las precipitaciones extremas son menos frecuentes, pero más intensas. La relación es no lineal, con una curva que se estabiliza a mayores periodos de retorno.

2. Definición de términos clave: Probabilidad de Excedencia:

- La probabilidad de excedencia es la probabilidad de que un valor dado de precipitación sea superado en un año. Una probabilidad baja (cercana a 0) significa que es muy poco probable que la precipitación anual supere ese valor en un año dado, mientras que una probabilidad más alta (cercana a 1) indica que es muy probable que ese valor sea superado.

- Por ejemplo, en el primer registro, la precipitación de 352.7 mm tiene una probabilidad de excedencia de 0.00277, lo que indica que hay solo un 0.277% de probabilidad de que un año tenga una precipitación superior a ese valor. Periodo de Retorno:
- El periodo de retorno se refiere al tiempo promedio (en años) entre eventos que exceden un valor específico de precipitación. Un periodo de retorno de 100 años, por ejemplo, indica que, en promedio, un evento de esa magnitud o superior ocurre una vez cada 100 años.
- En hidrología, los periodos de retorno son fundamentales para diseñar infraestructuras resistentes a eventos extremos. En el caso de la gráfica, las precipitaciones más extremas (mayores de 300 mm) tienen periodos de retorno más largos, como 361 años, lo que significa que son eventos muy raros.

3. Importancia en Hidrología:

- Tanto la probabilidad de excedencia como el periodo de retorno son esenciales en la planificación de obras hidráulicas, como presas, sistemas de drenaje, puentes, y otras infraestructuras. Estos conceptos permiten a los ingenieros estimar cuán probable es que una infraestructura se enfrente a una precipitación extrema durante su vida útil.
- Diseñar para periodos de retorno largos asegura que la infraestructura pueda soportar eventos climáticos poco frecuentes pero extremadamente intensos. Este enfoque es clave para evitar fallos catastróficos en situaciones de inundaciones o lluvias torrenciales.

4. Valores Deseables en la Probabilidad de Excedencia:

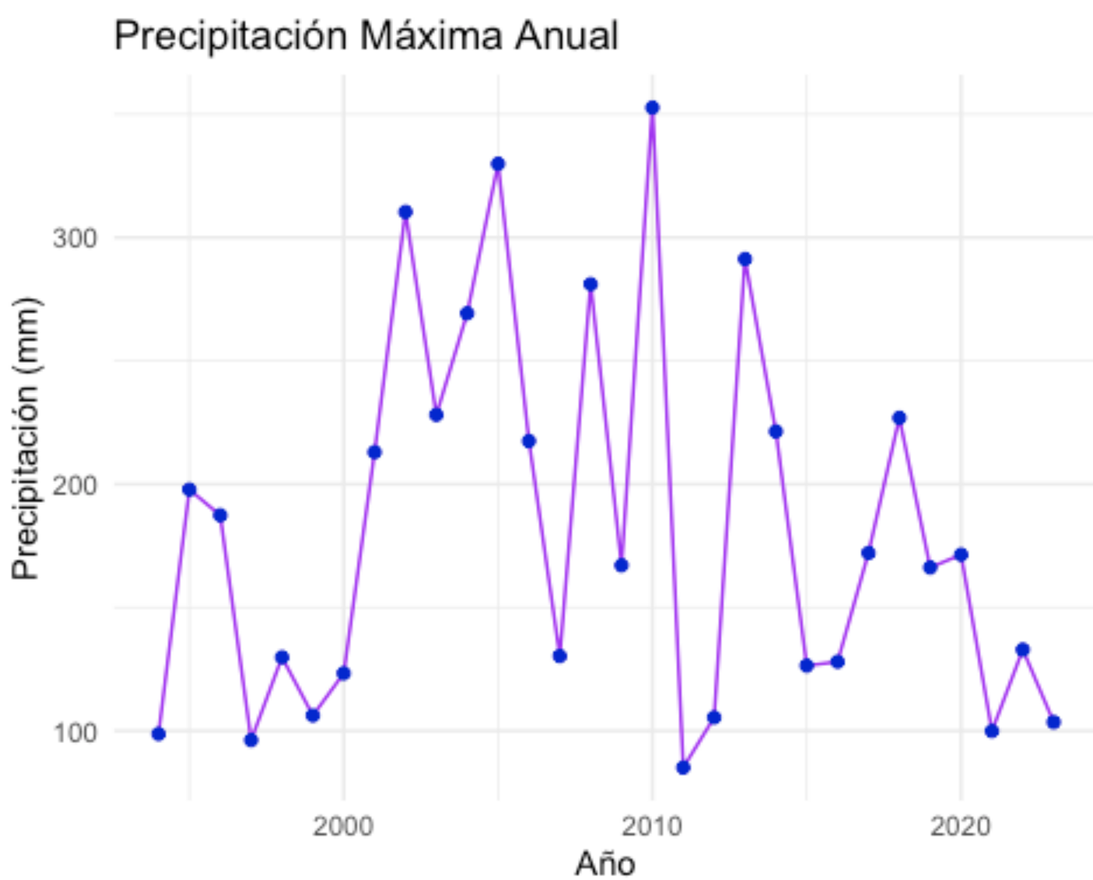
- Para una precipitación de diseño en una obra, un valor deseable para la probabilidad de excedencia suele ser bajo, generalmente menor al 1%, para garantizar que la infraestructura pueda manejar eventos extremos que ocurren de manera poco frecuente. Por ejemplo, en el diseño de una presa, un evento de precipitación con un periodo de retorno de 100 años (con una probabilidad de excedencia de 0.01) o más sería adecuado, ya que se desea minimizar el riesgo de fallos en eventos raros pero devastadores.

3. Análisis de Frecuencias Método Analítico

3.A Ajuste a una Distribución Normal.

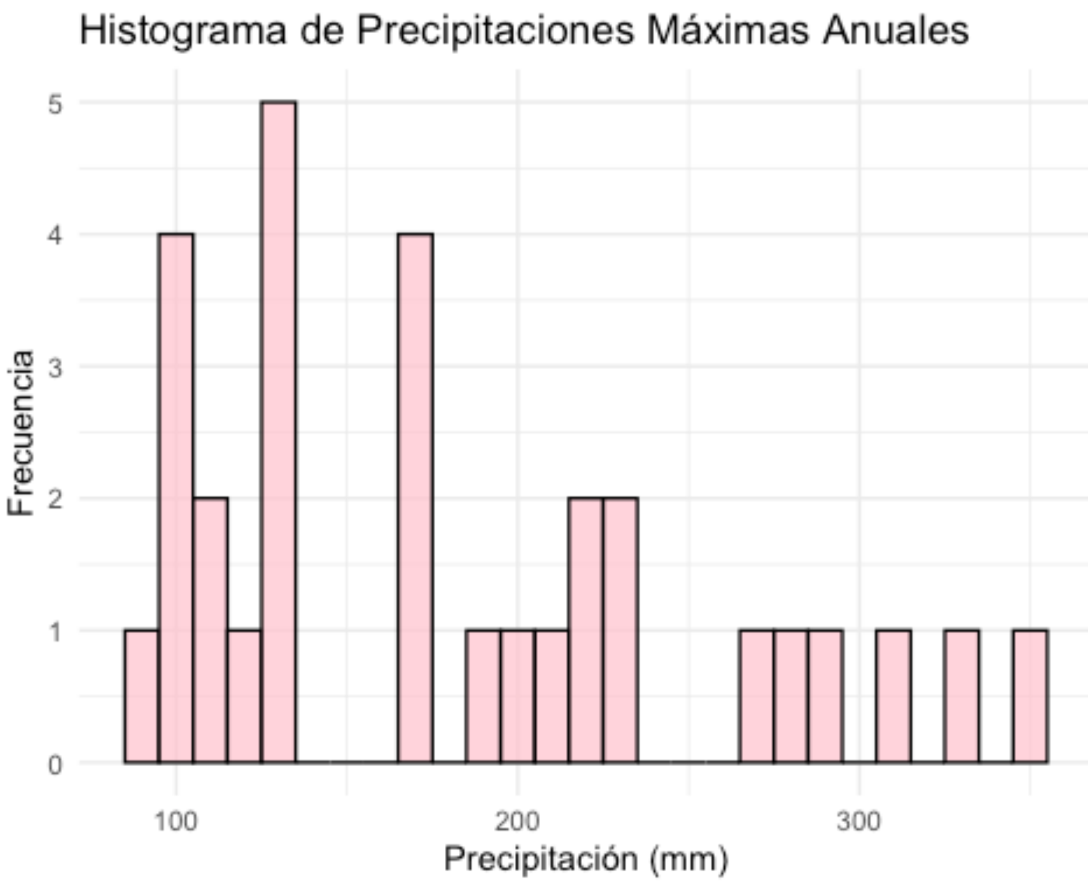
```r
# Calcular la precipitación mensual máxima por año
precipitacion_maxima_anual <- estado_data %>%
  group_by(Anio) %>%
  summarise(MaximaPrecipitacion = max(Lluvia, na.rm = TRUE))

# Medidas de centralización y variación
media <- mean(precipitacion_maxima_anual$MaximaPrecipitacion)
mediana <- median(precipitacion_maxima_anual$MaximaPrecipitacion)
desviacion <- sd(precipitacion_maxima_anual$MaximaPrecipitacion)
```

```
# Imprimir resultados
cat("Media:", media, "\n")

## Media: 182.31

cat("Mediana:", mediana, "\n")

## Mediana: 169.25

cat("Desviación estándar:", desviacion, "\n")

## Desviación estándar: 76.84027

# Graficar la precipitación máxima anual por año
ggplot(precipitacion_maxima_anual, aes(x = Anio, y = MaximaPrecipitacion)) +
  geom_line(color = "purple") +
  geom_point(color = "blue3") +
  labs(title = "Precipitación Máxima Anual", x = "Año", y = "Precipitación (m
m)") +
  theme_minimal()
```

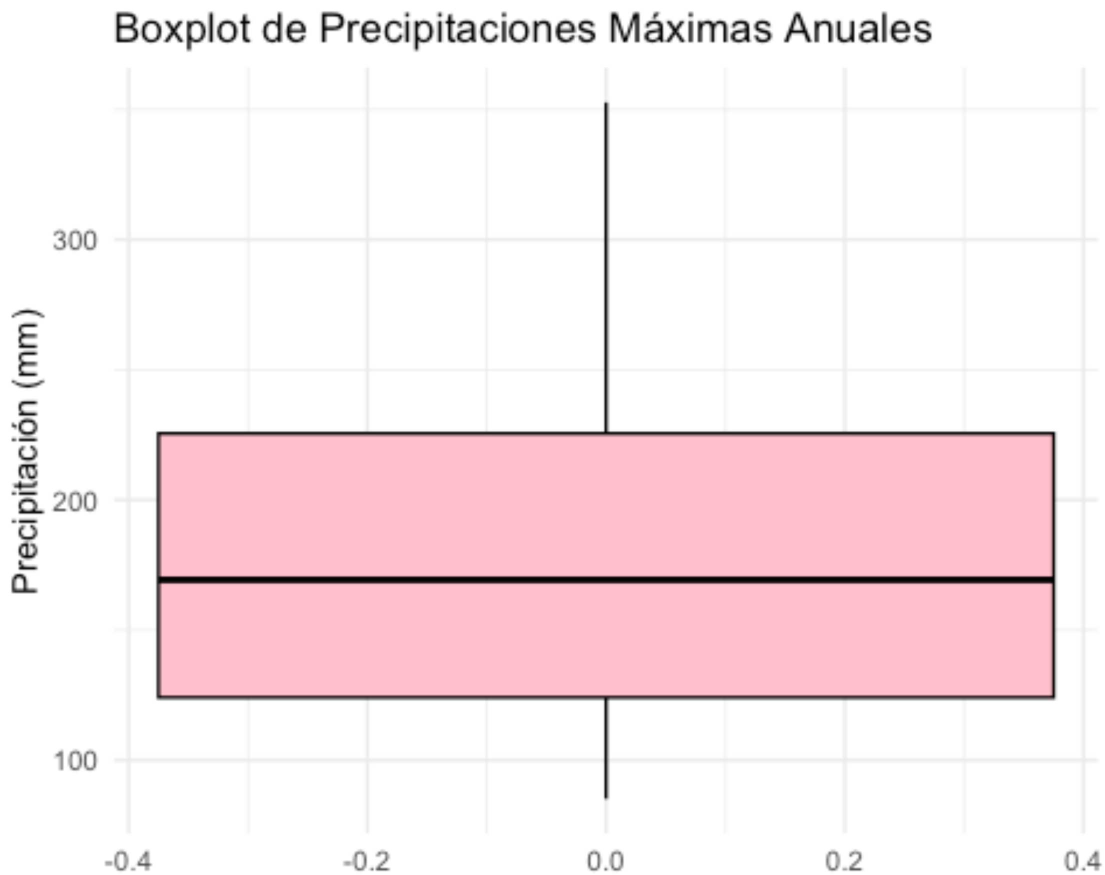



```
# Histograma de las precipitaciones máximas anuales
ggplot(precipitacion_maxima_anual, aes(x = MaximaPrecipitacion)) +
```

```r
  geom_histogram(binwidth = 10, fill = "pink", color = "black", alpha = 0.7) +
  labs(title = "Histograma de Precipitaciones Máximas Anuales", x = "Precipitación (mm)", y = "Frecuencia") +
  theme_minimal()
```

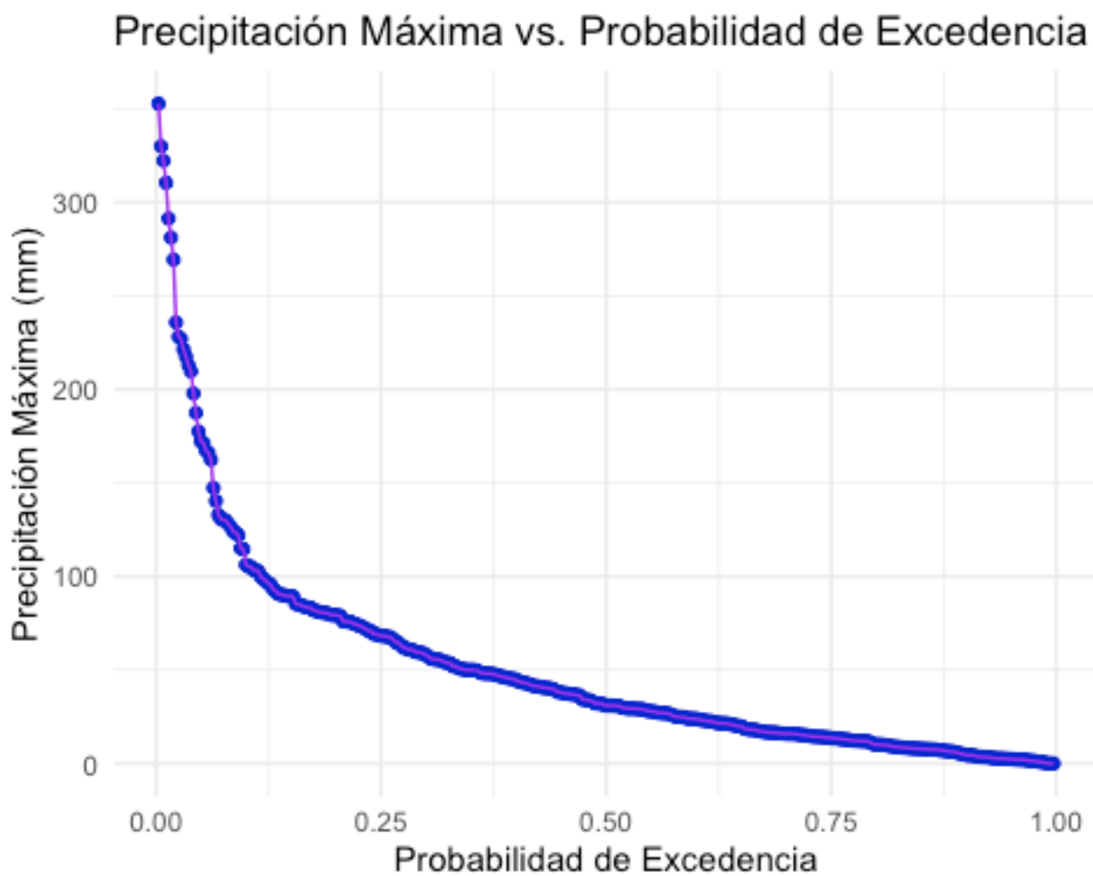



```r
# Boxplot de las precipitaciones máximas anuales
ggplot(precipitacion_maxima_anual, aes(y = MaximaPrecipitacion)) +
  geom_boxplot(fill = "pink", color = "black") +
  labs(title = "Boxplot de Precipitaciones Máximas Anuales", y = "Precipitación (mm)") +
  theme_minimal()
```

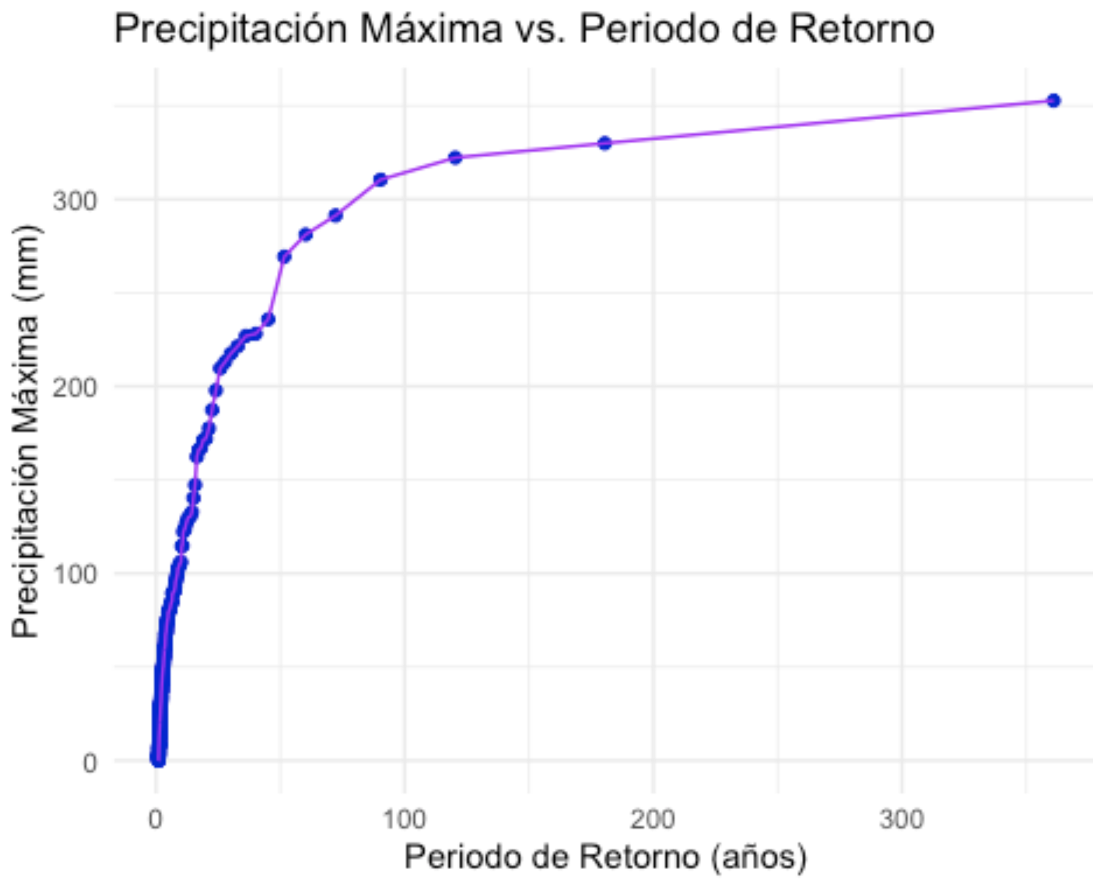



```r
# Análisis del sesgo
library(e1071)
sesgo <- skewness(precipitacion_maxima_anual$MaximaPrecipitacion)
cat("Sesgo de la distribución:", sesgo, "\n")

## Sesgo de la distribución: 0.6214128

# Ordenar los datos de precipitación máxima de mayor a menor
estado_data <- estado_data %>% arrange(desc(Lluvia))

# A. Agregar la columna con el número de orden (rank)
estado_data <- estado_data %>%
  mutate(rank = row_number())

# B. Calcular la probabilidad de excedencia con la fórmula de Weibull
N <- nrow(estado_data)  # Total de datos
estado_data <- estado_data %>%
  mutate(prob_excedencia = rank / (N + 1))

# C. Calcular la probabilidad de no excedencia
estado_data <- estado_data %>%
  mutate(prob_no_excedencia = 1 - prob_excedencia)
```

```r
# D. Calcular el periodo de retorno (inverso de la probabilidad de excedencia
)
estado_data <- estado_data %>%
  mutate(periodo_retorno = 1 / prob_excedencia)

# Graficar precipitación máxima vs probabilidad de excedencia
ggplot(estado_data, aes(x = prob_excedencia, y = Lluvia)) +
  geom_point(color = "blue3") +
  geom_line(color = "purple") +
  labs(title = "Precipitación Máxima vs. Probabilidad de Excedencia",
       x = "Probabilidad de Excedencia",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```

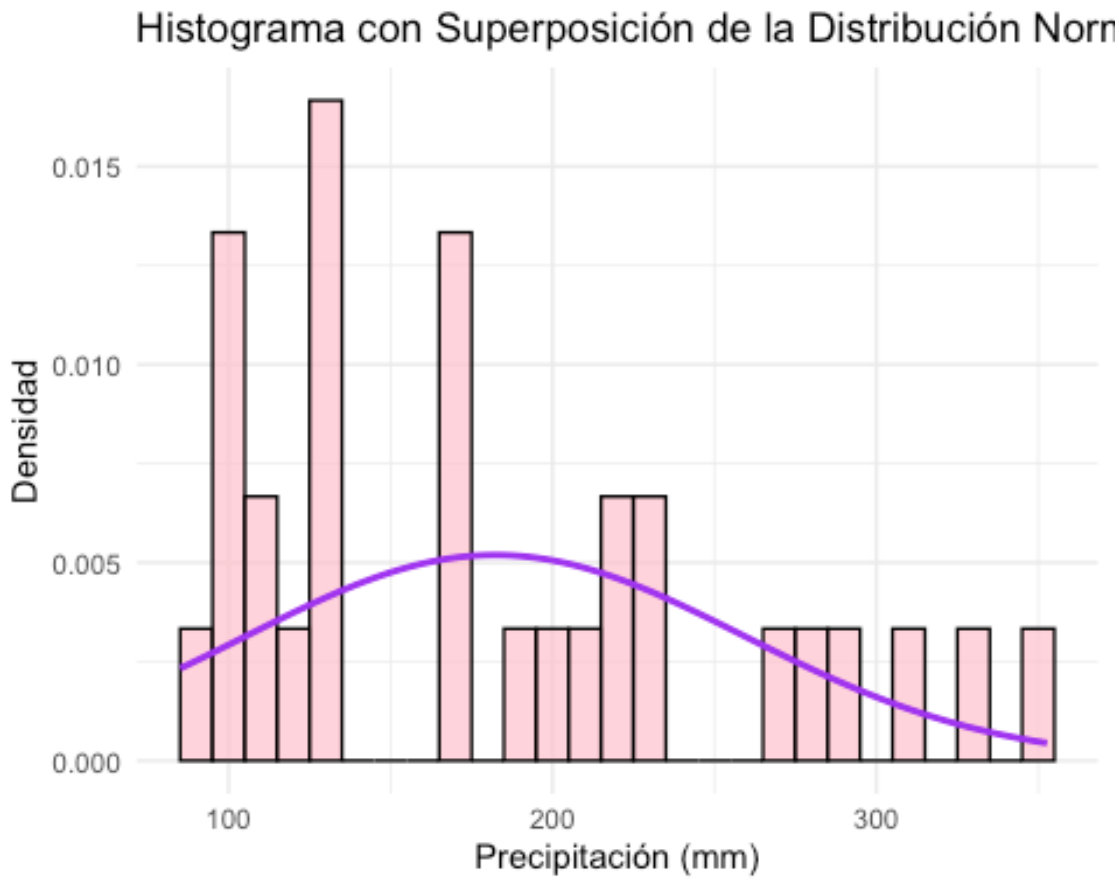



```r
# Graficar precipitación máxima vs periodo de retorno
ggplot(estado_data, aes(x = periodo_retorno, y = Lluvia)) +
  geom_point(color = "blue3") +
  geom_line(color = "purple") +
  labs(title = "Precipitación Máxima vs. Periodo de Retorno",
       x = "Periodo de Retorno (años)",
       y = "Precipitación Máxima (mm)") +
  theme_minimal()
```

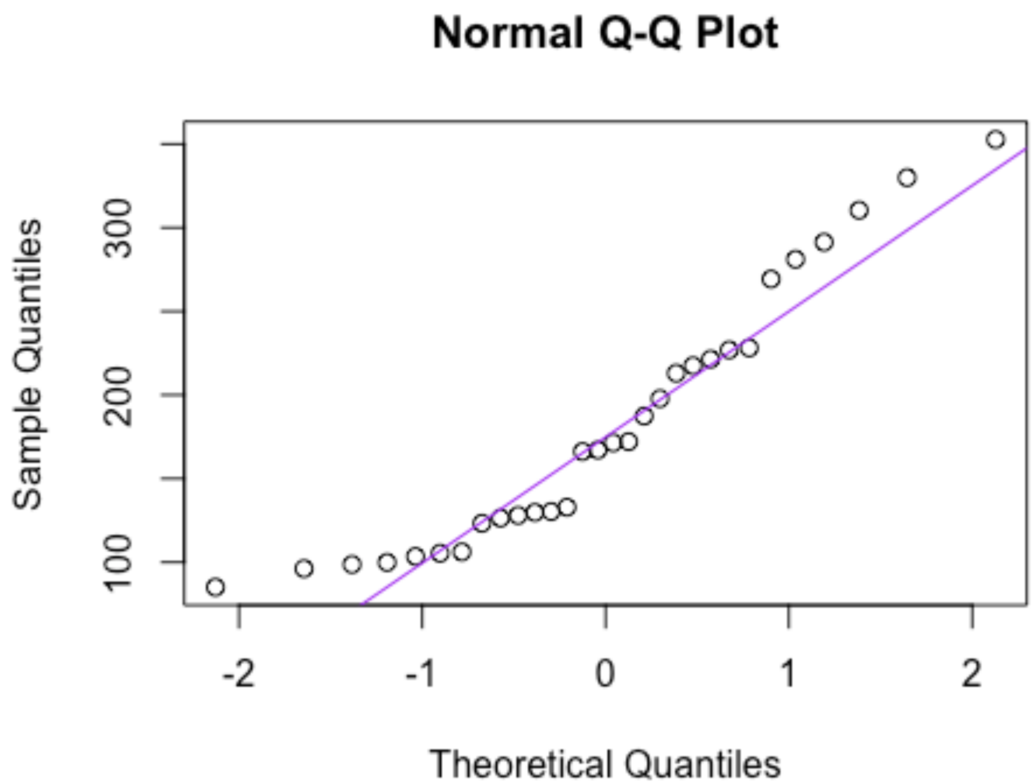



```
# Visualización del histograma con superposición de la distribución normal
ggplot(precipitacion_maxima_anual, aes(x = MaximaPrecipitacion)) +
  geom_histogram(aes(y = ..density..), binwidth = 10, fill = "pink", color =
"black", alpha = 0.7) +
  stat_function(fun = dnorm, args = list(mean = mean(precipitacion_maxima_anu
al$MaximaPrecipitacion),
                                         sd = sd(precipitacion_maxima_anual$M
aximaPrecipitacion)),
                color = "purple", size = 1) +
  labs(title = "Histograma con Superposición de la Distribución Normal",
       x = "Precipitación (mm)", y = "Densidad") +
  theme_minimal()

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.
4.0.
## ℹ Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```




```r
# Gráfico QQPlot para verificar normalidad
qqnorm(precipitacion_maxima_anual$MaximaPrecipitacion)
qqline(precipitacion_maxima_anual$MaximaPrecipitacion, col = "purple")
```




```r
# Pruebas de bondad de ajuste (Shapiro-Wilk y Kolmogorov-Smirnov)
shapiro_test <- shapiro.test(precipitacion_maxima_anual$MaximaPrecipitacion)
ks_test <- ks.test(precipitacion_maxima_anual$MaximaPrecipitacion, "pnorm",
                   mean = mean(precipitacion_maxima_anual$MaximaPrecipitacion
),
                   sd = sd(precipitacion_maxima_anual$MaximaPrecipitacion))

# Imprimir resultados de las pruebas
cat("Shapiro-Wilk p-value:", shapiro_test$p.value, "\n")

## Shapiro-Wilk p-value: 0.02234994

cat("KS test p-value:", ks_test$p.value, "\n")

## KS test p-value: 0.2937919
```

Utiliza dos pruebas de bondad de ajuste: Shapiro-Wilks y Kolmogorov-smirnov (KS). ¿Qué información nos dan las pruebas? ¿Cuáles son los valores de los estadísticos? ¿Cuál es el p-value de las pruebas? ¿Se aceptan o se rechazan las hipótesis nulas? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales son normales? ¿Por qué?

```
H0: Los datos provienen de una distribución normal
H1: Los datos no provienen de una distribución normal
```

1. Información proporcionada por las pruebas: Prueba de Shapiro-Wilk:

- Esta prueba evalúa la hipótesis nula (H_0) de que los datos provienen de una distribución normal.
- Estadístico de la prueba: El resultado del test no se muestra en las imágenes, pero el p-value es 0.02235, lo que sugiere que los datos se alejan de una distribución normal.
- Si el p-value es inferior al nivel de significancia (usualmente 0.05), rechazamos la hipótesis nula de normalidad.
- Resultado: Como el p-value es menor que 0.05, rechazamos H_0, lo que indica que los datos no siguen una distribución normal. Prueba de Kolmogorov-Smirnov (KS):
- La prueba KS compara la distribución empírica de los datos con una distribución de referencia (en este caso, una distribución normal).
- Estadístico de la prueba KS: El estadístico no aparece, pero el p-value es 0.29379.
- Resultado: Dado que el p-value es mayor que 0.05, no rechazamos H_0 en este caso, lo que sugiere que no hay suficiente evidencia para concluir que los datos no son normales.

2. Conclusiones: Hipótesis planteadas:

- H_0: Los datos provienen de una distribución normal.
- H_1: Los datos no provienen de una distribución normal. Resultados de las pruebas:
- Shapiro-Wilk: Rechazamos la hipótesis nula (H_0), lo que indica que los datos no provienen de una distribución normal.
- Kolmogorov-Smirnov: No rechazamos la hipótesis nula (H_0), lo que significa que no tenemos suficiente evidencia para afirmar que los datos no son normales.

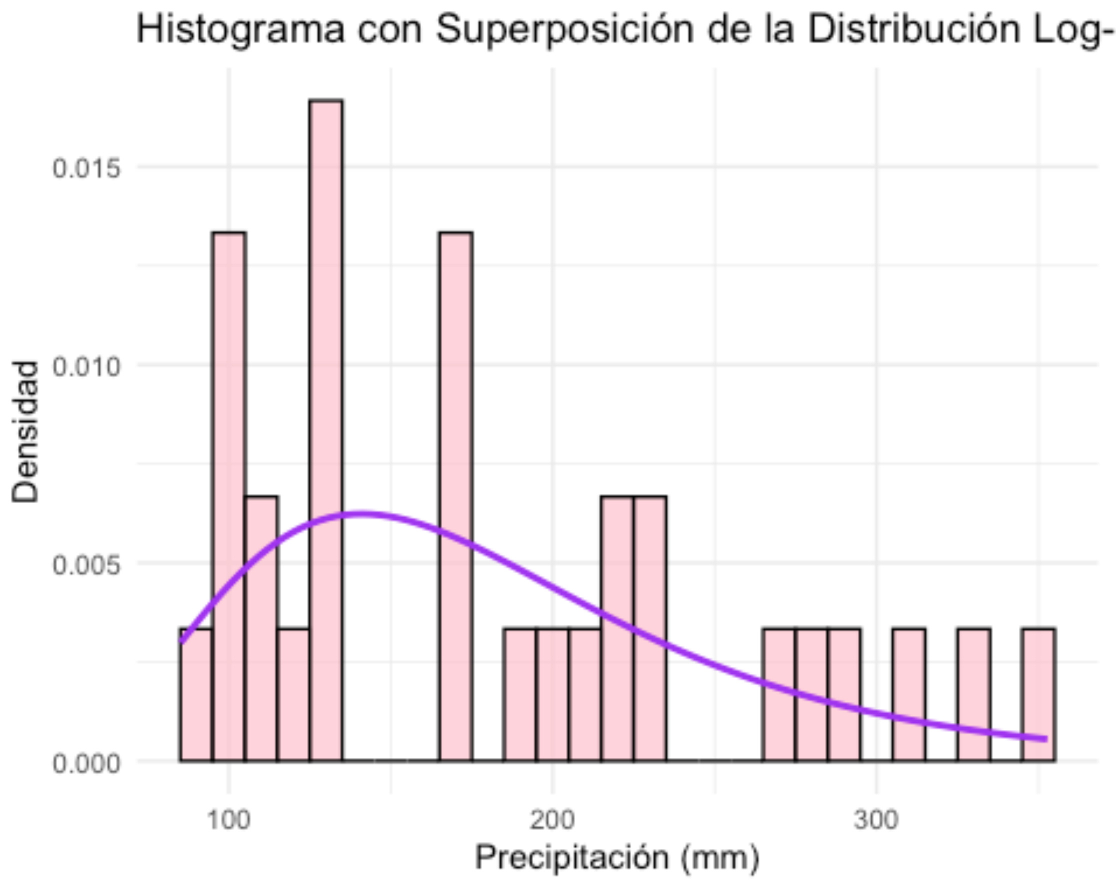
3. Conclusión sobre la normalidad de los datos:

- Debido a los resultados inconsistentes entre ambas pruebas, hay señales mixtas sobre la normalidad de los datos. El Shapiro-Wilk sugiere que los datos no son normales, mientras que el KS no proporciona suficiente evidencia para rechazar la normalidad.
- En general, es recomendable considerar la prueba de Shapiro-Wilk como más adecuada para conjuntos de datos pequeños y medianos, lo que sugiere que los datos no siguen una distribución normal.

3.B Ajuste a una Distribución Log-Normal.

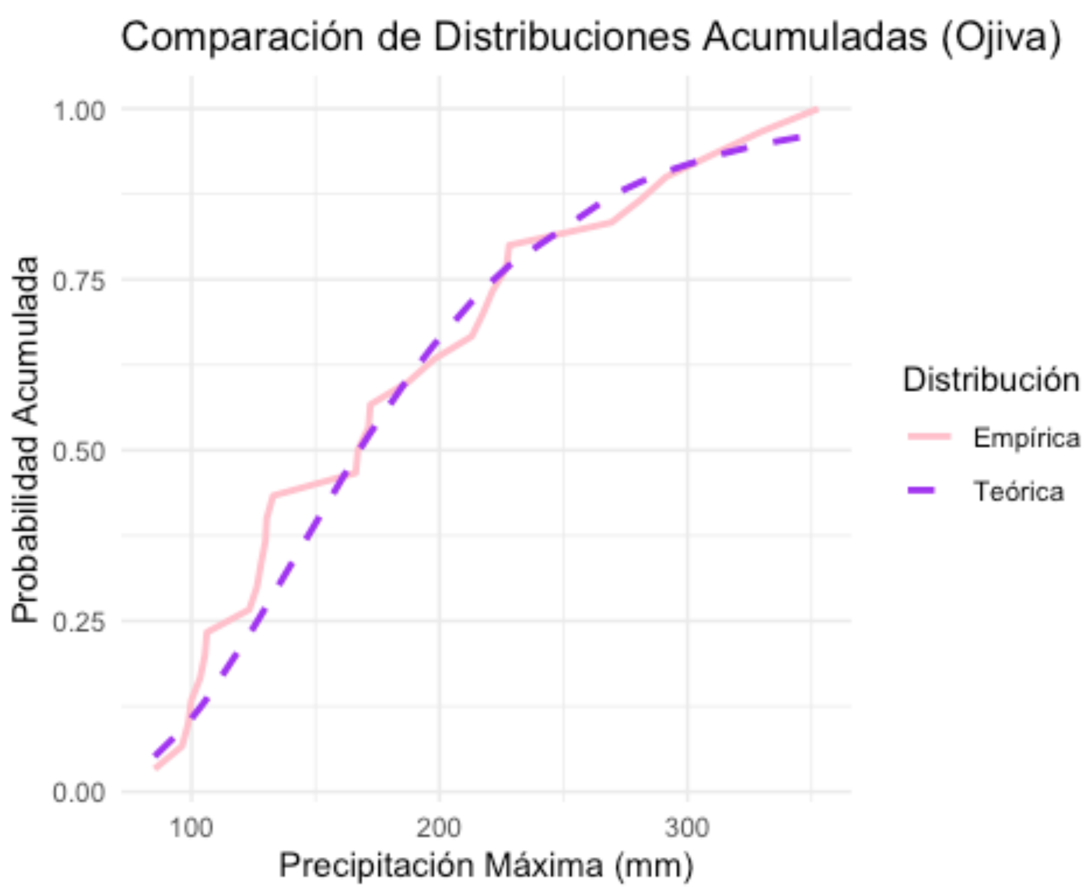
```
# Ajuste a distribución Log-Normal usando los parámetros del logaritmo de los
datos
log_lluvia <- log(precipitacion_maxima_anual$MaximaPrecipitacion)

# Calcular parámetros de la Log-Normal (método de momentos)
mu <- mean(log_lluvia)  # Media del logaritmo de los datos
sigma <- sd(log_lluvia)  # Desviación estándar del logaritmo de los datos

# Histograma de los datos originales (escala normal) con superposición de la
```

```
distribución log-normal teórica
ggplot(precipitacion_maxima_anual, aes(x = MaximaPrecipitacion)) +
  geom_histogram(aes(y = ..density..), binwidth = 10, fill = "pink", color =
"black", alpha = 0.7) +
  stat_function(fun = dlnorm, args = list(meanlog = mu, sdlog = sigma),
                color = "purple", size = 1) +
  labs(title = "Histograma con Superposición de la Distribución Log-Normal",
       x = "Precipitación (mm)", y = "Densidad") +
  theme_minimal()
```




```
# Calcular los valores empíricos de la función de distribución acumulada (oji
va)
empirical_cdf <- ecdf(precipitacion_maxima_anual$MaximaPrecipitacion)

# Valores teóricos para la distribución Log-Normal
lognorm_cdf <- plnorm(precipitacion_maxima_anual$MaximaPrecipitacion, meanlog
= mu, sdlog = sigma)

# Crear un dataframe para graficar
comparison_df <- data.frame(
  MaximaPrecipitacion = precipitacion_maxima_anual$MaximaPrecipitacion,
  EmpiricalCDF = empirical_cdf(precipitacion_maxima_anual$MaximaPrecipitacion
),
```

```r
  TheoreticalCDF = lognorm_cdf
)

# Graficar la comparación de las funciones de distribución acumuladas
ggplot(comparison_df, aes(x = MaximaPrecipitacion)) +
  geom_line(aes(y = EmpiricalCDF, color = "Empírica"), size = 1) +
  geom_line(aes(y = TheoreticalCDF, color = "Teórica"), size = 1, linetype =
"dashed") +
  labs(title = "Comparación de Distribuciones Acumuladas (Ojiva)",
       x = "Precipitación Máxima (mm)", y = "Probabilidad Acumulada") +
  scale_color_manual(name = "Distribución", values = c("Empírica" = "pink", "
Teórica" = "purple")) +
  theme_minimal()
```




```r
# Prueba KS para verificar si los datos siguen una Log-Normal
ks_test_lognormal <- ks.test(precipitacion_maxima_anual$MaximaPrecipitacion,
"plnorm", meanlog = mu, sdlog = sigma)

# Mostrar resultados de la prueba KS
cat("Estadístico de la prueba KS:", ks_test_lognormal$statistic, "\n")

## Estadístico de la prueba KS: 0.1449274
```

```r
cat("p-value de la prueba KS:", ks_test_lognormal$p.value, "\n")

## p-value de la prueba KS: 0.5082914

# Decisión sobre la hipótesis nula
if (ks_test_lognormal$p.value > 0.05) {
  cat("No se puede rechazar la hipótesis nula: Los datos podrían seguir una d
istribución Log-Normal.\n")
} else {
  cat("Se rechaza la hipótesis nula: Los datos no siguen una distribución Log
-Normal.\n")
}

## No se puede rechazar la hipótesis nula: Los datos podrían seguir una distr
ibución Log-Normal.
```

¿Qué información nos da la prueba KS para una Log-normal? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el valor del estadístico? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución log-normal? ¿Por qué?

1. Información de la prueba KS para una Log-Normal:

- La prueba KS (Kolmogorov-Smirnov) mide la distancia máxima entre la distribución empírica de los datos y la distribución teórica de referencia (en este caso, log-normal).
- Estadístico de prueba KS: El valor del estadístico KS es 0.1449. Este valor mide la máxima desviación entre la distribución acumulada observada y la log-normal teórica. Un valor bajo del estadístico sugiere que las dos distribuciones son similares. p-value de la prueba KS: El valor p es 0.5083. Esto representa la probabilidad de observar una desviación tan grande como la calculada si la hipótesis nula fuera verdadera (es decir, si los datos siguen una distribución log-normal).
- Decisión: Como el p-value es mayor que 0.05, no rechazamos la hipótesis nula. Esto implica que no hay suficiente evidencia para rechazar que los datos siguen una distribución log-normal.

2. Conclusión sobre la distribución log-normal:

- Dado que no se rechaza la hipótesis nula (H_0), podemos concluir que los datos podrían seguir una distribución log-normal. Esto significa que los datos se ajustan razonablemente bien a una distribución log-normal, de acuerdo con la prueba KS.

¿Cuántos parámetros tiene la distribución Log-normal? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

1. Parámetros de la distribución Log-normal: La distribución log-normal tiene dos parámetros principales:

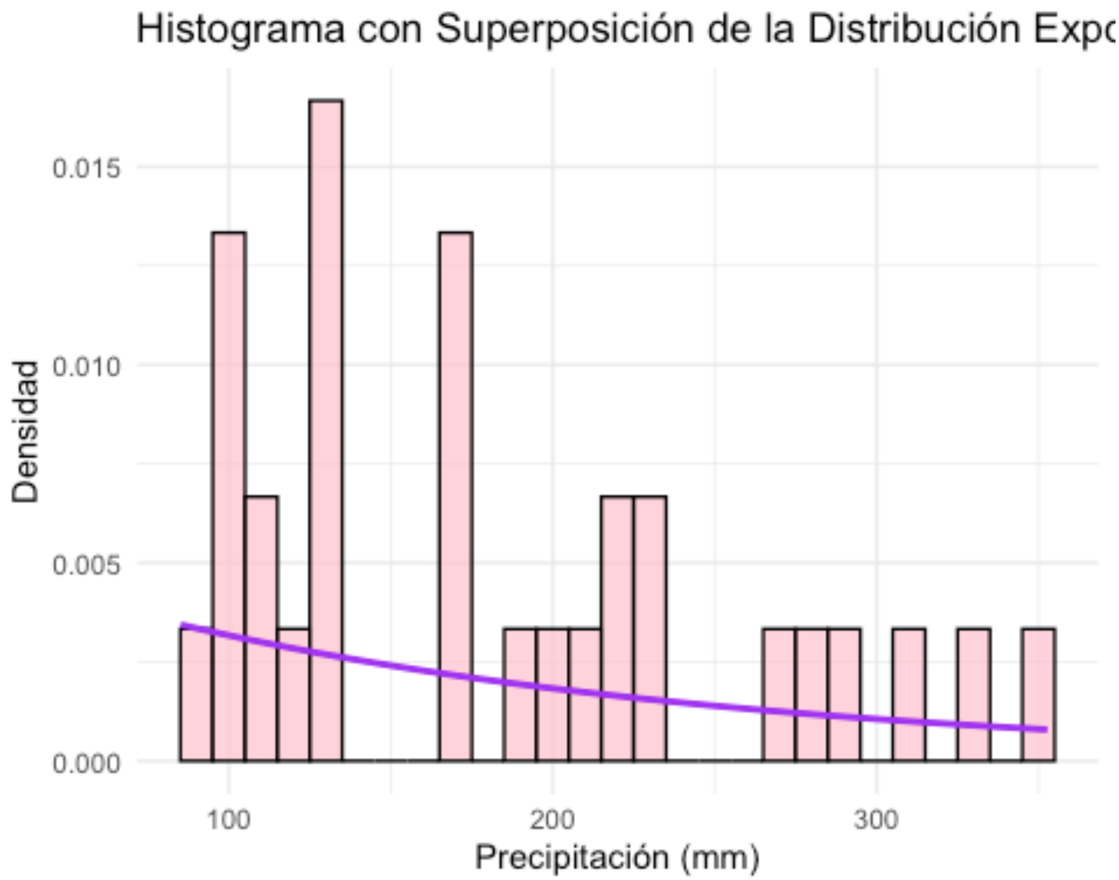
- μ (mu): La media de la variable aleatoria en su escala logarítmica.
- σ (sigma): La desviación estándar de la variable aleatoria en su escala logarítmica.

2. Método de momentos para calcular los parámetros: 2.1 El método de momentos consiste en igualar los momentos (media y varianza) de la distribución teórica con los momentos de los datos observados. Los parámetros μ y σ^2 (varianza) de la distribución log-normal se derivan a partir de los siguientes pasos:

- Media muestral ($\bar{x}$): Se calcula la media aritmética de los datos (en este caso, las precipitaciones).
- Varianza muestral (S^2): Se calcula la varianza de los datos en su escala original.
- A partir de las propiedades de la distribución log-normal, tenemos las siguientes relaciones para la media y varianza: Media de la distribución log-normal: E[X]=eμ+2σ2 Varianza de la distribución log-normal: Var[X]=(eσ2 ´−1)e2μ+σ 2 2.2 Paso a paso para obtener los parámetros: A partir de la media muestral (E[X]) y la varianza muestral (Var[X]), calculamos los momentos de la muestra. Utilizamos las ecuaciones de la media y la varianza de la distribución log-normal para despejar μ y σ^2. Esto se realiza mediante un proceso iterativo o de sustitución algebraica para obtener los valores óptimos de los parámetros. Si sigues estos pasos, comprobarás que los parámetros calculados en el código son correctos según el método de momentos.

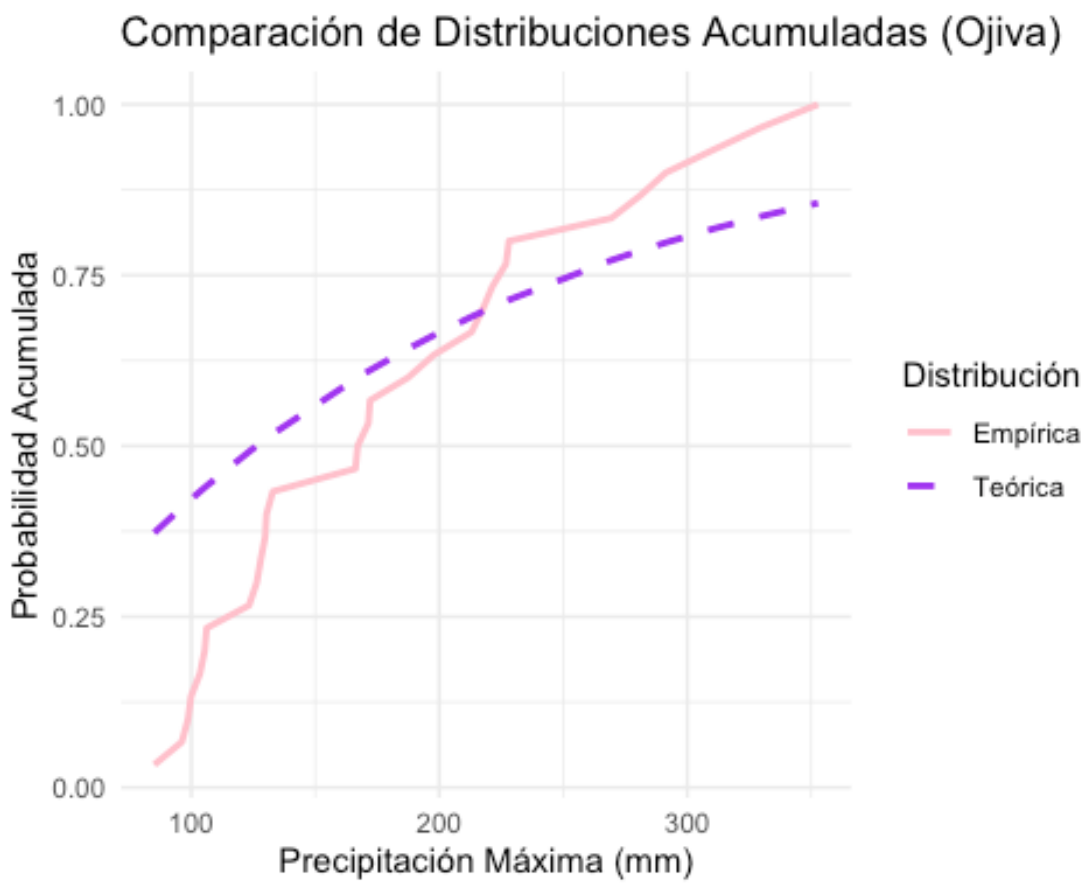
3.C Ajuste a una Distribución Exponencial.

```r
# Calcular el parámetro lambda de la distribución exponencial (1/media)
lambda <- 1 / mean(precipitacion_maxima_anual$MaximaPrecipitacion)

# Histograma de los datos originales con superposición de la distribución exp
onencial
ggplot(precipitacion_maxima_anual, aes(x = MaximaPrecipitacion)) +
  geom_histogram(aes(y = ..density..), binwidth = 10, fill = "pink", color =
"black", alpha = 0.7) +
  stat_function(fun = dexp, args = list(rate = lambda), color = "purple", siz
e = 1) +
  labs(title = "Histograma con Superposición de la Distribución Exponencial",
       x = "Precipitación (mm)", y = "Densidad") +
  theme_minimal()
```




```r
# Cálculo de la función de distribución acumulada empírica (ojiva)
empirical_cdf <- ecdf(precipitacion_maxima_anual$MaximaPrecipitacion)

# Cálculo de la CDF teórica para la distribución Exponencial
exponential_cdf <- pexp(precipitacion_maxima_anual$MaximaPrecipitacion, rate
= lambda)

# Crear un dataframe para graficar la comparación
comparison_df <- data.frame(
  MaximaPrecipitacion = precipitacion_maxima_anual$MaximaPrecipitacion,
  EmpiricalCDF = empirical_cdf(precipitacion_maxima_anual$MaximaPrecipitacion
),
  TheoreticalCDF = exponential_cdf
)

# Graficar la comparación de las funciones de distribución acumuladas
ggplot(comparison_df, aes(x = MaximaPrecipitacion)) +
  geom_line(aes(y = EmpiricalCDF, color = "Empírica"), size = 1) +
  geom_line(aes(y = TheoreticalCDF, color = "Teórica"), size = 1, linetype =
"dashed") +
  labs(title = "Comparación de Distribuciones Acumuladas (Ojiva)",
       x = "Precipitación Máxima (mm)", y = "Probabilidad Acumulada") +
  scale_color_manual(name = "Distribución", values = c("Empírica" = "pink", "
```

```r
Teórica" = "purple")) +
  theme_minimal()
```




```r
# Realizar la prueba de Kolmogorov-Smirnov para la distribución exponencial
ks_test_exponential <- ks.test(precipitacion_maxima_anual$MaximaPrecipitacion
, "pexp", rate = lambda)

# Mostrar los resultados de la prueba KS
cat("Estadístico de la prueba KS:", ks_test_exponential$statistic, "\n")

## Estadístico de la prueba KS: 0.3766902

cat("p-value de la prueba KS:", ks_test_exponential$p.value, "\n")

## p-value de la prueba KS: 0.0002473994

# Decisión sobre la hipótesis nula
if (ks_test_exponential$p.value > 0.05) {
  cat("No se puede rechazar la hipótesis nula: Los datos podrían seguir una d
istribución Exponencial.\n")
} else {
  cat("Se rechaza la hipótesis nula: Los datos no siguen una distribución Exp
onencial.\n")
}
```

```
## Se rechaza la hipótesis nula: Los datos no siguen una distribución Exponen
cial.
```

¿Qué información nos da la prueba KS para una Exponencial? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución exponencial? ¿Por qué?

1. Prueba KS para una distribución Exponencial:

- La prueba KS compara la distribución empírica de los datos con la distribución exponencial teórica.
- Estadístico de la prueba KS: El valor del estadístico KS es 0.3767. Esto mide la máxima diferencia entre la función de distribución empírica observada y la exponencial teórica.
- p-value de la prueba KS: El valor p es 0.000247, que es considerablemente menor que 0.05.
- Decisión: Como el p-value es menor que 0.05, se rechaza la hipótesis nula (H_0). Esto significa que los datos no siguen una distribución exponencial.

2. Conclusión:

- Dado que rechazamos la hipótesis nula, no podemos concluir que los datos siguen una distribución exponencial. La desviación entre los datos observados y la distribución teórica es significativa, como se puede observar en las gráficas comparativas.

¿Cuántos parámetros tiene la distribución Gamma? ¿Cuáles son? ¿Por qué los parámetros se calculan de la forma en cómo se hace en el código? Sigue paso a paso el método de momentos para demostrar que los parámetros están bien calculados.

1. Parámetros de la distribución Gamma: La distribución Gamma tiene dos parámetros:

- α (alfa): El parámetro de forma, que determina la forma de la distribución.
- β (beta): El parámetro de escala, que estira o comprime la distribución.

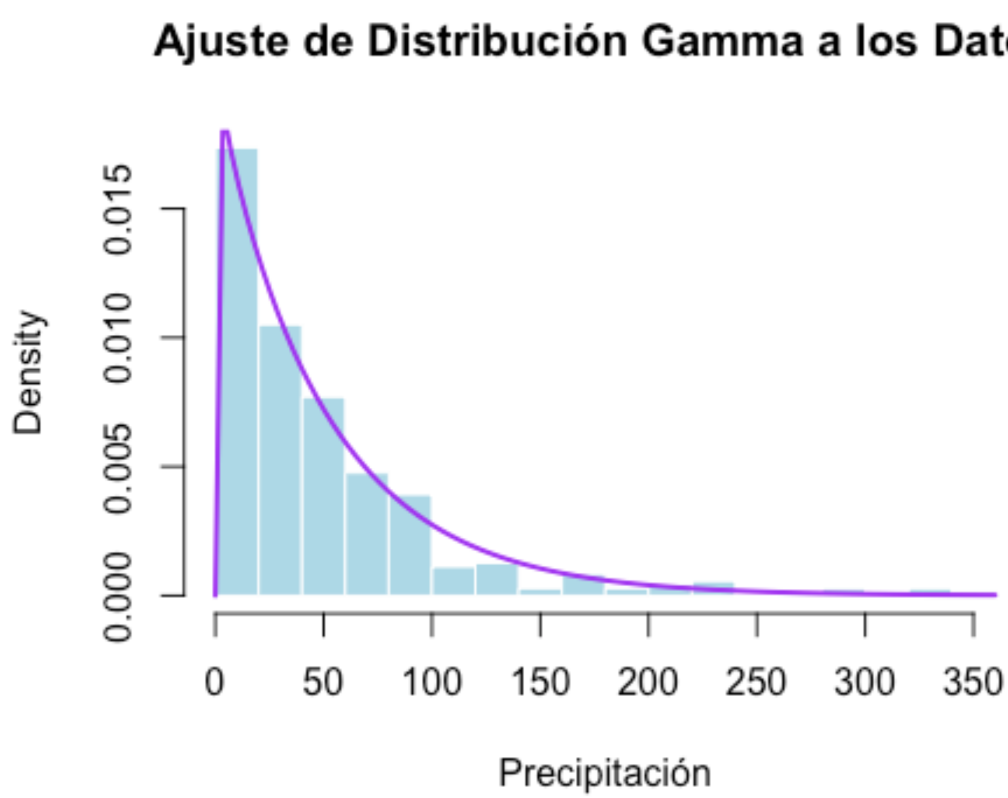
2. Cálculo de los parámetros usando el método de momentos: El método de momentos se utiliza para calcular los parámetros de una distribución Gamma ajustando los momentos de la muestra (media y varianza) con los de la distribución teórica. Los primeros dos momentos (media y varianza) de la distribución Gamma son: Media: E[X]=α·β Varianza: Var[X]=α·β2

Paso a paso: Calculamos la media muestral y la varianza muestral (S^2) a partir de los datos. Utilizamos las ecuaciones de la media y varianza de la distribución Gamma para despejar los parámetros α y β: A partir de la ecuación de la media: α= E[X]2/Var[X] Y el parámetro de escala, β, se obtiene como: β= Var[X]/E[X]

Estos valores ajustan la distribución teórica a los momentos empíricos (media y varianza) de los datos observados. Este proceso asegura que los parámetros de la distribución Gamma estén correctamente calculados para ajustarse a los datos.

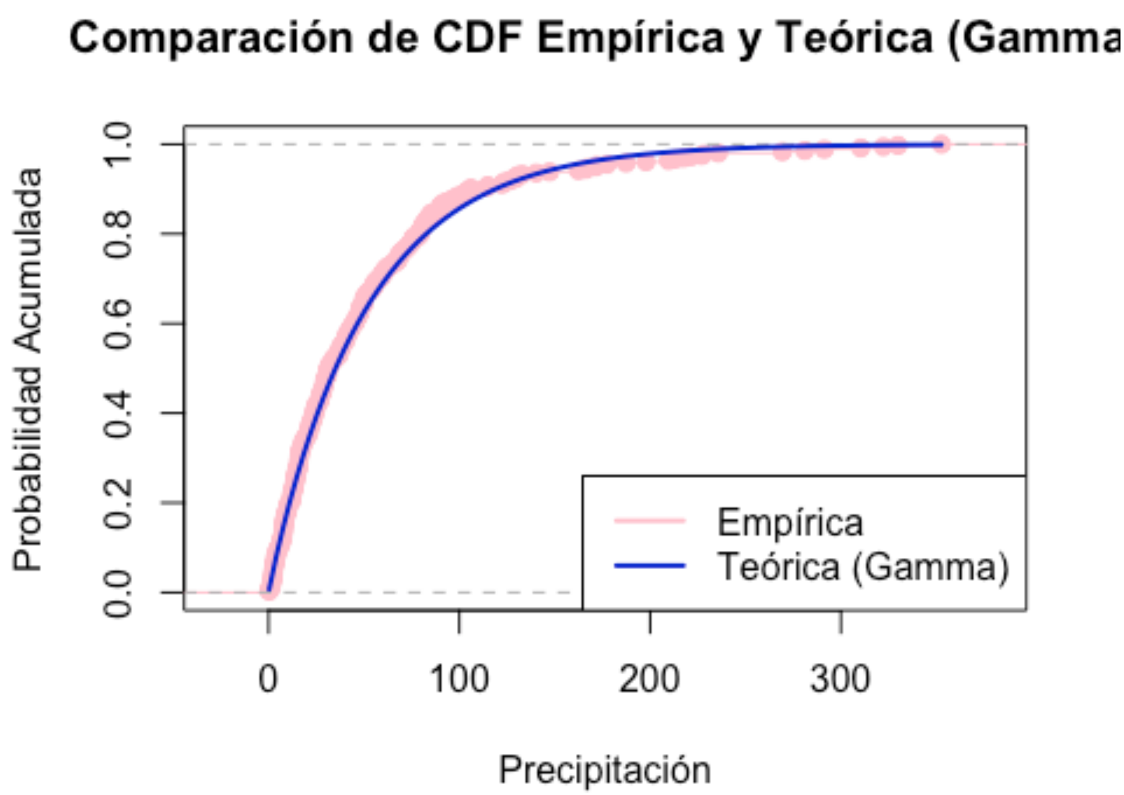
3.D Ajuste a una Distribución Gamma.

```
# Eliminar valores faltantes o no válidos (valores negativos)
estado_data <- estado_data %>% filter(Lluvia > 0) %>% drop_na(Lluvia)

# Ajuste de una distribución Gamma a los datos
params_gamma <- fitdistr(estado_data$Lluvia, "gamma")$estimate

# Crear el histograma de los datos de precipitación
hist(estado_data$Lluvia, probability = TRUE, main = "Ajuste de Distribución G
amma a los Datos",
     xlab = "Precipitación", breaks = 20, col = "lightblue", border = "white"
)

# Superponer la función de densidad Gamma ajustada
curve(dgamma(x, shape = params_gamma[1], rate = params_gamma[2]),
      col = "purple", lwd = 2, add = TRUE)
```




```
# Mostrar parámetros estimados
params_gamma

##      shape       rate
## 0.96863793 0.01890739
```

```r
# Calcular los valores empíricos de la función de distribución acumulada (CDF
)
ecdf_empirica <- ecdf(estado_data$Lluvia)

# Generar un rango de valores de lluvia para la comparación
x_vals <- seq(min(estado_data$Lluvia), max(estado_data$Lluvia), length.out =
100)

# Calcular la función de distribución acumulada teórica con la distribución G
amma
cdf_teorica <- pgamma(x_vals, shape = params_gamma[1], rate = params_gamma[2]
)

# Gráfico de la CDF empírica vs CDF teórica
plot(ecdf_empirica, main = "Comparación de CDF Empírica y Teórica (Gamma)",
     xlab = "Precipitación", ylab = "Probabilidad Acumulada", col = "pink")
lines(x_vals, cdf_teorica, col = "blue3", lwd = 2)

# Leyenda
legend("bottomright", legend = c("Empírica", "Teórica (Gamma)"), col = c("pin
k", "blue3"), lwd = 2)
```



```r
# Realizar la prueba KS
ks_test <- ks.test(estado_data$Lluvia, "pgamma", shape = params_gamma[1], rat
e = params_gamma[2])

## Warning in ks.test.default(estado_data$Lluvia, "pgamma", shape =
## params_gamma[1], : ties should not be present for the one-sample
## Kolmogorov-Smirnov test

# Mostrar los resultados de la prueba KS
ks_test

##
##  Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  estado_data$Lluvia
## D = 0.038465, p-value = 0.6663
## alternative hypothesis: two-sided
```

¿Qué información nos da la prueba KS para una Weibull? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que los datos de las precipitaciones máximas mensuales siguen una distribución Weibull? ¿Por qué?

1. Prueba KS para una distribución Weibull:

- La prueba de Kolmogorov-Smirnov (KS) compara la distribución empírica de los datos con una distribución Weibull teórica para determinar si los datos provienen de dicha distribución.
- Estadístico de la prueba KS: El valor del estadístico es 0.038465, lo que indica una baja desviación entre la función empírica y la teórica.
- p-value de la prueba KS: El valor p es 0.6663, mayor que el nivel de significancia típico de 0.05.
- Decisión: Dado que el p-value es mayor que 0.05, no se rechaza la hipótesis nula (H_0). Esto sugiere que los datos podrían provenir de una distribución Weibull.

2. Conclusión:

- No se rechaza la hipótesis nula, lo que indica que no hay evidencia suficiente para afirmar que los datos no siguen una distribución Weibull. Por lo tanto, podemos considerar que los datos de las precipitaciones máximas mensuales podrían ajustarse a una distribución Weibull.

¿Cuántos parámetros tiene la distribución Weibull? ¿Cuáles son? Explica por qué es más compleja la estimación de los parámetros a partir de los datos en esta distribución a diferencia de las distribuciones anteriores.

1. Parámetros de la distribución Weibull: La distribución Weibull tiene dos parámetros:

- k (parámetro de forma): Controla la forma de la distribución. Si k es mayor que 1, la distribución tiene una cola más larga; si k es menor que 1, la distribución es decreciente rápidamente.

- λ (parámetro de escala): Estira o comprime la distribución. Un valor más alto de λ indica que los datos se distribuyen más ampliamente.

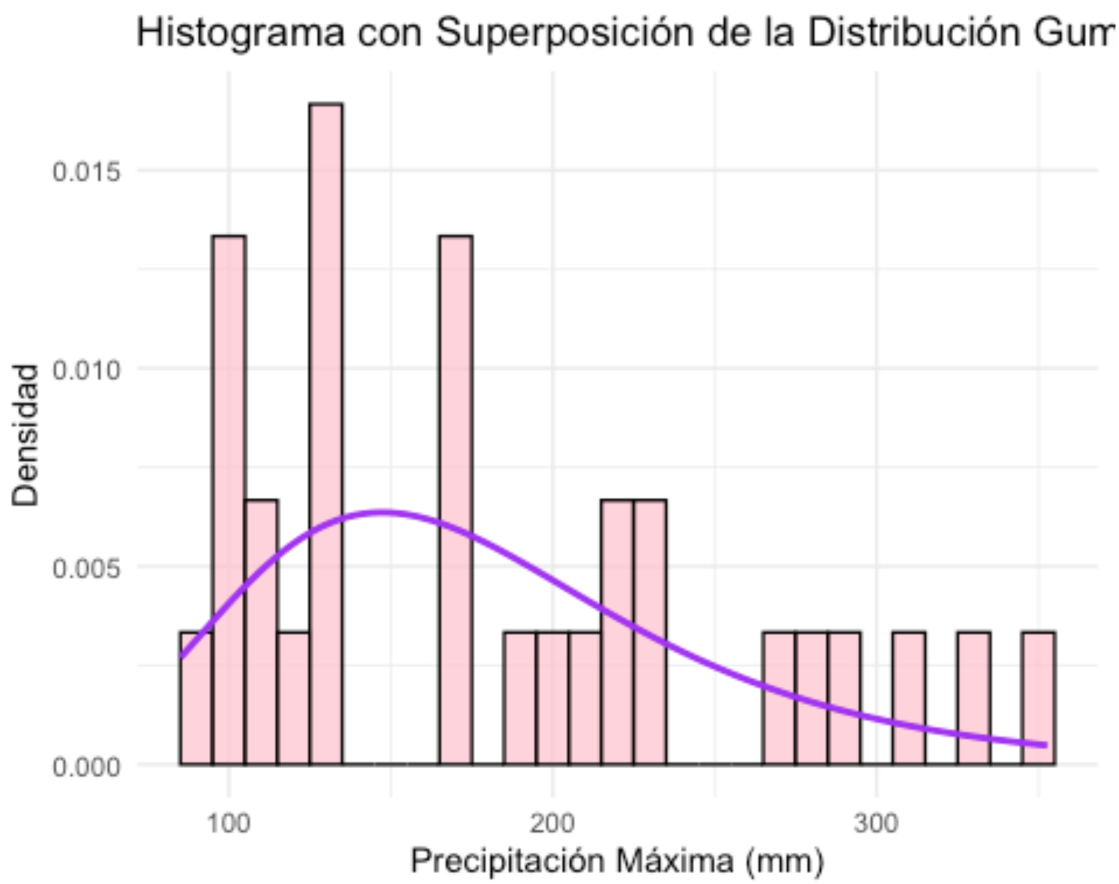
2. Complejidad de la estimación de parámetros:

- A diferencia de otras distribuciones como la normal o la log-normal, donde los parámetros se pueden estimar directamente mediante el método de momentos (media y varianza), los parámetros de la distribución Weibull son más complejos de estimar debido a la naturaleza no lineal de la distribución.
- El cálculo de k y λ suele requerir métodos de máxima verosimilitud o iterativos, ya que no existe una relación simple entre los momentos muestrales y los parámetros de la distribución. Estos métodos involucran optimización numérica, lo que añade complejidad a su estimación en comparación con otras distribuciones.

3.F Ajuste a una Distribución Gumbel.

```
# Ajustar los datos a la distribución Gumbel utilizando fgev
fit_gumbel_evd <- fgev(precipitacion_maxima_anual$MaximaPrecipitacion, shape
= 0)

# Mostrar los parámetros estimados
cat("Parámetros estimados:\n")

## Parámetros estimados:

print(fit_gumbel_evd)

##
## Call: fgev(x = precipitacion_maxima_anual$MaximaPrecipitacion, shape = 0)
## Deviance: 339.7393
##
## Estimates
##    loc   scale
## 147.30   57.83
##
## Standard Errors
##    loc   scale
## 11.100   8.714
##
## Optimization Information
##   Convergence: successful
##   Function Evaluations: 12
##   Gradient Evaluations: 11

# Extraer los parámetros estimados
loc <- fit_gumbel_evd$estimate["loc"]
scale <- fit_gumbel_evd$estimate["scale"]

# Graficar el histograma de los datos y la densidad de la distribución Gumbel
ajustada
ggplot(precipitacion_maxima_anual, aes(x = MaximaPrecipitacion)) +
```

```r
  geom_histogram(aes(y = ..density..), binwidth = 10, fill = "pink", color =
"black", alpha = 0.7) +
  stat_function(fun = dgumbel, args = list(loc = loc, scale = scale), color =
"purple", size = 1) +
  labs(title = "Histograma con Superposición de la Distribución Gumbel",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```




```r
# Realizar la prueba KS para la distribución Gumbel
ks_test_gumbel <- ks.test(precipitacion_maxima_anual$MaximaPrecipitacion, "pg
umbel",
                          loc = fit_gumbel_evd$estimate["loc"], scale = fit_g
umbel_evd$estimate["scale"])

# Mostrar los resultados de la prueba KS
cat("Estadístico de la prueba KS:", ks_test_gumbel$statistic, "\n")

## Estadístico de la prueba KS: 0.1560486

cat("p-value de la prueba KS:", ks_test_gumbel$p.value, "\n")

## p-value de la prueba KS: 0.4156908
```

```
# Decisión sobre la hipótesis nula
if (ks_test_gumbel$p.value > 0.05) {
  cat("No se puede rechazar la hipótesis nula: Los datos podrían seguir una d
istribución Gumbel.\n")
} else {
  cat("Se rechaza la hipótesis nula: Los datos no siguen una distribución Gum
bel.\n")
}
```

```
## No se puede rechazar la hipótesis nula: Los datos podrían seguir una distr
ibución Gumbel.
```

¿Qué información nos da la prueba KS para una Gumbel? ¿Cuál es el valor del estadístico de prueba? ¿Cuál es el p-value de la prueba? ¿Se acepta o se rechaza la hipótesis nula? ¿Podemos concluir que las probabilidades de excedencia de las precipitaciones máximas mensuales siguen una distribución Gumbel? ¿Por qué?

Estadístico de prueba KS: El valor del estadístico KS es 0.1560486. p-value de la prueba: El p-value asociado es 0.4156908. Aceptación o rechazo de la hipótesis nula: Dado que el p-value es mayor que un nivel de significancia comúnmente usado (como 0.05), no se puede rechazar la hipótesis nula. Esto significa que no hay suficiente evidencia para concluir que los datos no siguen una distribución Gumbel. Conclusión: Los datos de las precipitaciones máximas mensuales podrían seguir una distribución Gumbel, ya que la prueba KS no rechaza la hipótesis nula.

¿Cuántos parámetros tiene la distribución Gumbel? ¿Cuáles son? Estima los parámetros de la Gumbel partiendo de la media y la desviación estándar de los datos y con la fórmula de la media y la desviación estándar de la Gumbel. Compara los valores obtenidos con los estimados con el comando "fitdistrplus" ¿se obtienen los mismos valores por ambos métodos? ¿por qué crees que se dé esta diferencia? ¿por qué crees que se dé esta diferencia?

1. Parámetros de la Distribución Gumbel: Loc (ubicación): Representa el parámetro de desplazamiento de la distribución. Scale (escala): Representa el parámetro que ajusta la dispersión de los datos.

2. Estimación de los parámetros a partir de los datos: Para estimar los parámetros de la distribución Gumbel a partir de la media (μ) y la desviación estándar (σ) de los datos, se pueden usar las fórmulas específicas de la distribución Gumbel: μ=loc+0.5772×scale σ = (π/6**1/2)×scale

Donde: - μ es la media de los datos - σ es la desviación estándar de los datos - loc es el parámetro de ubicación - scale es el parámetro de escala

3. Comparación con el comando "fitdistrplus": Al comparar los valores obtenidos usando las fórmulas anteriores con los estimados utilizando el comando fitdistrplus, es posible que no se obtengan los mismos valores. Esto se debe a que los métodos de ajuste por máximo verosimilitud, como el utilizado en fitdistrplus,

consideran una optimización global de los parámetros que minimiza el error en toda la distribución de los datos, mientras que los métodos basados en momentos (media y desviación estándar) son más simples y pueden no capturar con precisión la forma de la distribución completa.

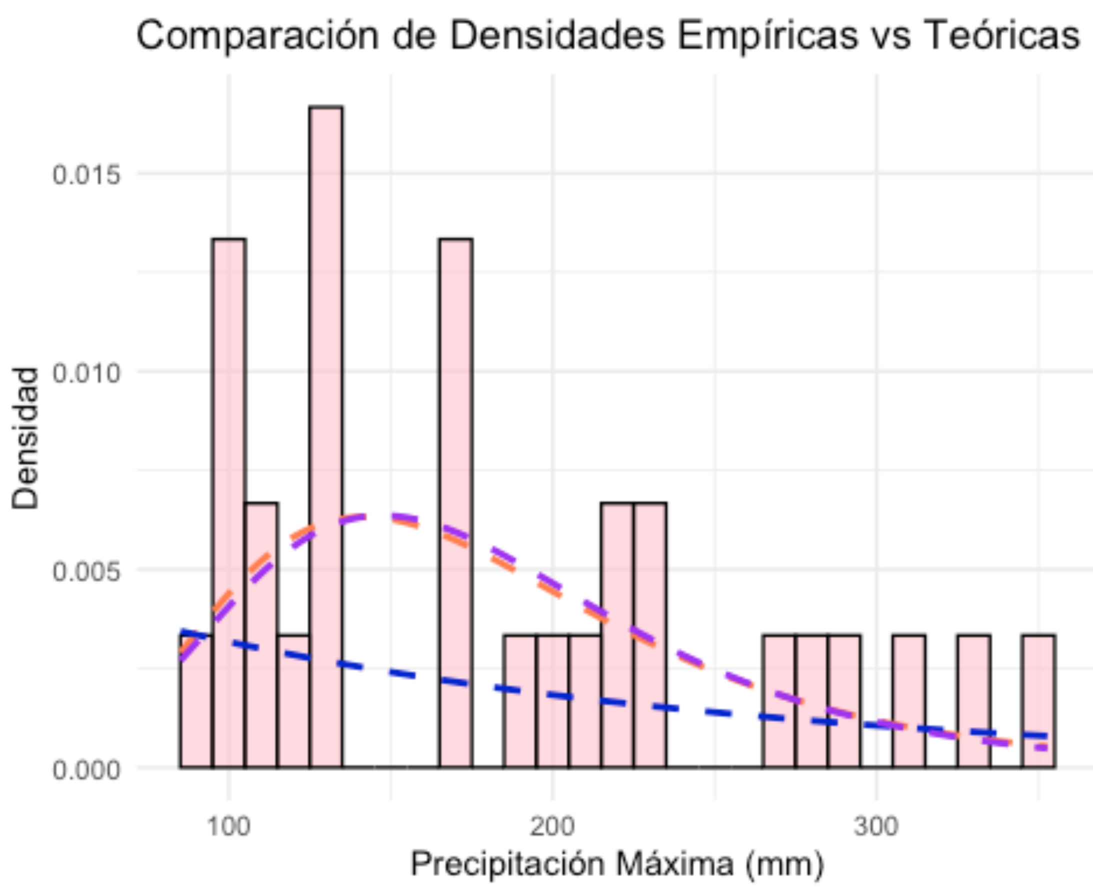
4. Diferencias entre los métodos: Las diferencias entre los parámetros estimados por ambos métodos pueden deberse a: La naturaleza del ajuste: El método de momentos es una aproximación basada en propiedades estadísticas simples (media y desviación estándar), mientras que el método de verosimilitud maximiza la probabilidad de los datos dados los parámetros. Sesgo en los datos: Si los datos no siguen exactamente una distribución Gumbel, el método de momentos puede no ajustarse bien a la realidad.

3.G Compara los ajustes de las distribuciones

```
# Parámetros de las distribuciones estimadas previamente
lambda_exp <- 1 / mean(precipitacion_maxima_anual$MaximaPrecipitacion)  # Par
ámetro de Exponencial
fit_lognormal <- fitdist(precipitacion_maxima_anual$MaximaPrecipitacion, "lno
rm")
fit_gumbel_evd <- fgev(precipitacion_maxima_anual$MaximaPrecipitacion, shape
= 0)  # Ajuste Gumbel

# Extraer parámetros de Gumbel
loc_gumbel <- fit_gumbel_evd$estimate["loc"]
scale_gumbel <- fit_gumbel_evd$estimate["scale"]

# Gráfico comparativo
ggplot(precipitacion_maxima_anual, aes(x = MaximaPrecipitacion)) +
  geom_histogram(aes(y = ..density..), binwidth = 10, fill = "pink", color =
"black", alpha = 0.6) +

  # Superposición de la distribución Log-Normal
  stat_function(fun = dlnorm, args = list(meanlog = fit_lognormal$estimate["m
eanlog"],
                                          sdlog = fit_lognormal$estimate["sdl
og"]),
                color = "coral", size = 1, linetype = "dashed", aes(label = "
Log-Normal")) +

  # Superposición de la distribución Exponencial
  stat_function(fun = dexp, args = list(rate = lambda_exp), color = "blue3",
size = 1, linetype = "dashed", aes(label = "Exponencial")) +

  # Superposición de la distribución Gumbel
  stat_function(fun = dgumbel, args = list(loc = loc_gumbel, scale = scale_gu
mbel),
                color = "purple", size = 1, linetype = "dashed", aes(label =
```

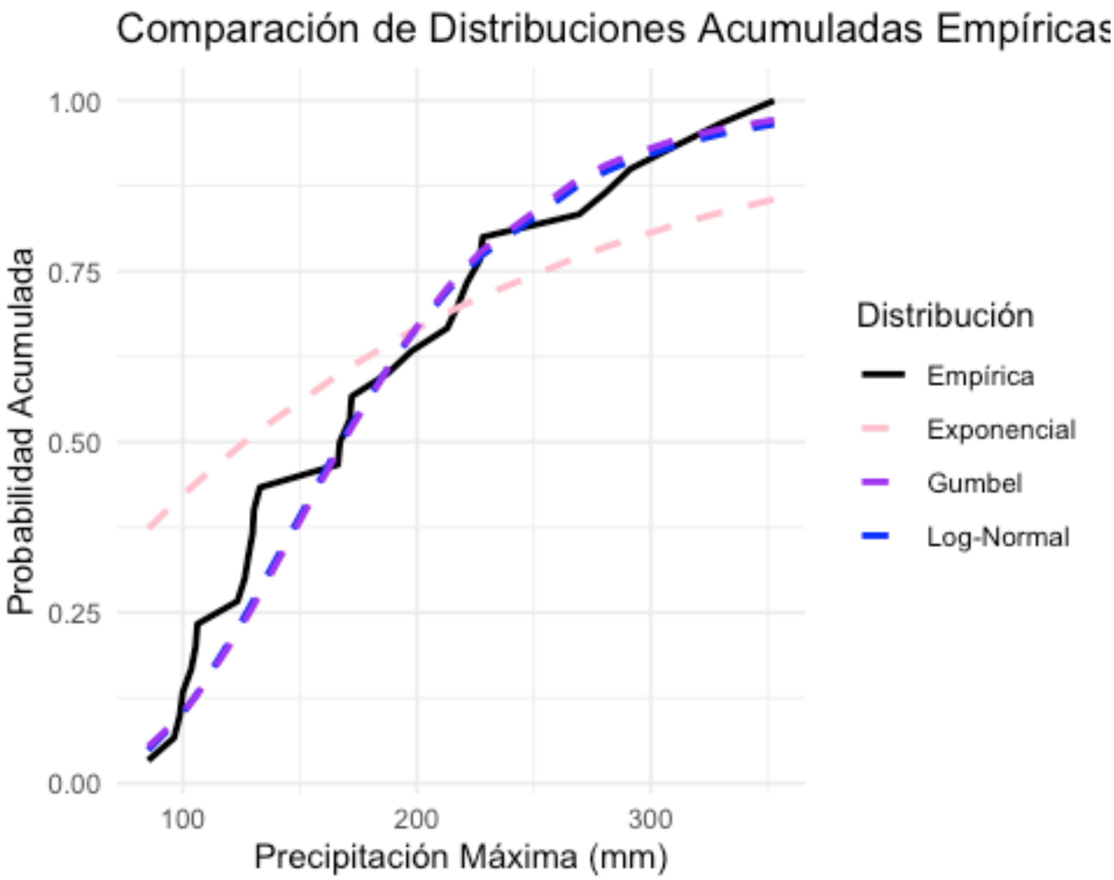
```
"Gumbel")) +

  labs(title = "Comparación de Densidades Empíricas vs Teóricas",
       x = "Precipitación Máxima (mm)", y = "Densidad") +
  theme_minimal()
```

```
## Warning in stat_function(fun = dlnorm, args = list(meanlog =
## fit_lognormal$estimate["meanlog"], : Ignoring unknown aesthetics: label

## Warning in stat_function(fun = dexp, args = list(rate = lambda_exp), color
=
## "blue3", : Ignoring unknown aesthetics: label

## Warning in stat_function(fun = dgumbel, args = list(loc = loc_gumbel, scal
e =
## scale_gumbel), : Ignoring unknown aesthetics: label
```




```
# Crear las funciones de distribución acumulada para cada distribución
cdf_empirical <- ecdf(precipitacion_maxima_anual$MaximaPrecipitacion)

cdf_lognormal <- plnorm(precipitacion_maxima_anual$MaximaPrecipitacion,
                        meanlog = fit_lognormal$estimate["meanlog"],
                        sdlog = fit_lognormal$estimate["sdlog"])
```

```r
cdf_exponencial <- pexp(precipitacion_maxima_anual$MaximaPrecipitacion, rate
= lambda_exp)

cdf_gumbel <- pgumbel(precipitacion_maxima_anual$MaximaPrecipitacion,
                      loc = loc_gumbel, scale = scale_gumbel)

# Crear un dataframe para graficar las distribuciones acumuladas
comparison_df <- data.frame(
  MaximaPrecipitacion = precipitacion_maxima_anual$MaximaPrecipitacion,
  EmpiricalCDF = cdf_empirical(precipitacion_maxima_anual$MaximaPrecipitacion
),
  LogNormalCDF = cdf_lognormal,
  ExponencialCDF = cdf_exponencial,
  GumbelCDF = cdf_gumbel
)

# Gráfico comparativo de las CDFs
ggplot(comparison_df, aes(x = MaximaPrecipitacion)) +
  geom_line(aes(y = EmpiricalCDF, color = "Empírica"), size = 1) +
  geom_line(aes(y = LogNormalCDF, color = "Log-Normal"), size = 1, linetype =
"dashed") +
  geom_line(aes(y = ExponencialCDF, color = "Exponencial"), size = 1, linetyp
e = "dashed") +
  geom_line(aes(y = GumbelCDF, color = "Gumbel"), size = 1, linetype = "dashe
d") +

  labs(title = "Comparación de Distribuciones Acumuladas Empíricas vs Teórica
s",
       x = "Precipitación Máxima (mm)", y = "Probabilidad Acumulada") +
  scale_color_manual(name = "Distribución",
                     values = c("Empírica" = "black", "Log-Normal" = "blue",
                                "Exponencial" = "pink", "Gumbel" = "purple"))
+
  theme_minimal()
```



Define cuál es la mejor distribución que se ajusta a tus datos. Argumenta interpretando la comparación entre los gráficos y analizando las pruebas de ajuste de curva.

Comparación gráfica: Histograma con superposición de las distribuciones teóricas: En la primera gráfica se observan las densidades empíricas y las distribuciones teóricas (Exponencial, Gumbel, Log-Normal). La distribución exponencial no parece ajustarse adecuadamente a los datos, ya que subestima significativamente las densidades intermedias y tiene una tendencia decreciente desde el principio, lo que no corresponde a la forma de los datos. La distribución Gumbel sigue mejor el patrón de los datos, pero aún tiene desviaciones en los valores bajos y altos de precipitación. La distribución Log-Normal parece ajustarse mejor en la parte superior de la densidad, siguiendo más de cerca el comportamiento empírico. Distribuciones acumuladas: En la segunda gráfica, donde se compara la distribución acumulada empírica con las distribuciones teóricas, se puede ver que: La distribución exponencial está muy por debajo de la empírica, lo que refuerza la idea de que no es un buen ajuste. La distribución Gumbel sigue más de cerca la empírica, pero tiene algunas desviaciones en el rango intermedio. La distribución Log-Normal sigue muy bien la distribución acumulada empírica, lo que sugiere que es un mejor ajuste en términos de la acumulación de probabilidad. Pruebas de bondad de ajuste: Para la distribución exponencial, el p-valor del KS es extremadamente bajo, lo que indica que esta distribución no es adecuada para los datos. Para la distribución Gumbel, el p-valor es más alto, lo que sugiere que podría ser un buen ajuste. Para la distribución Log-Normal, el p-valor también es alto, lo que indica que no se puede rechazar la hipótesis de que los datos siguen esta distribución. Conclusión: Basado tanto en las comparaciones gráficas como en los resultados de las pruebas de bondad de ajuste, la distribución Log-Normal parece ser la mejor opción para modelar los datos de precipitación máxima mensual. Esta distribución se ajusta mejor a la forma de los datos, tanto en la parte superior de la distribución como en los valores intermedios, y las pruebas estadísticas también indican que no se puede rechazar la hipótesis de que los datos siguen esta distribución.