

```
<!--Ejercicios y visualizaciones con R base-->
```

# Visualización de DATOS {

```
<Por="Nazaret Basaldella"/>
```

}

```
#Actividad_2
```

```
#Docente_Francesc_Busquet
```

```
#04/06/2025
```

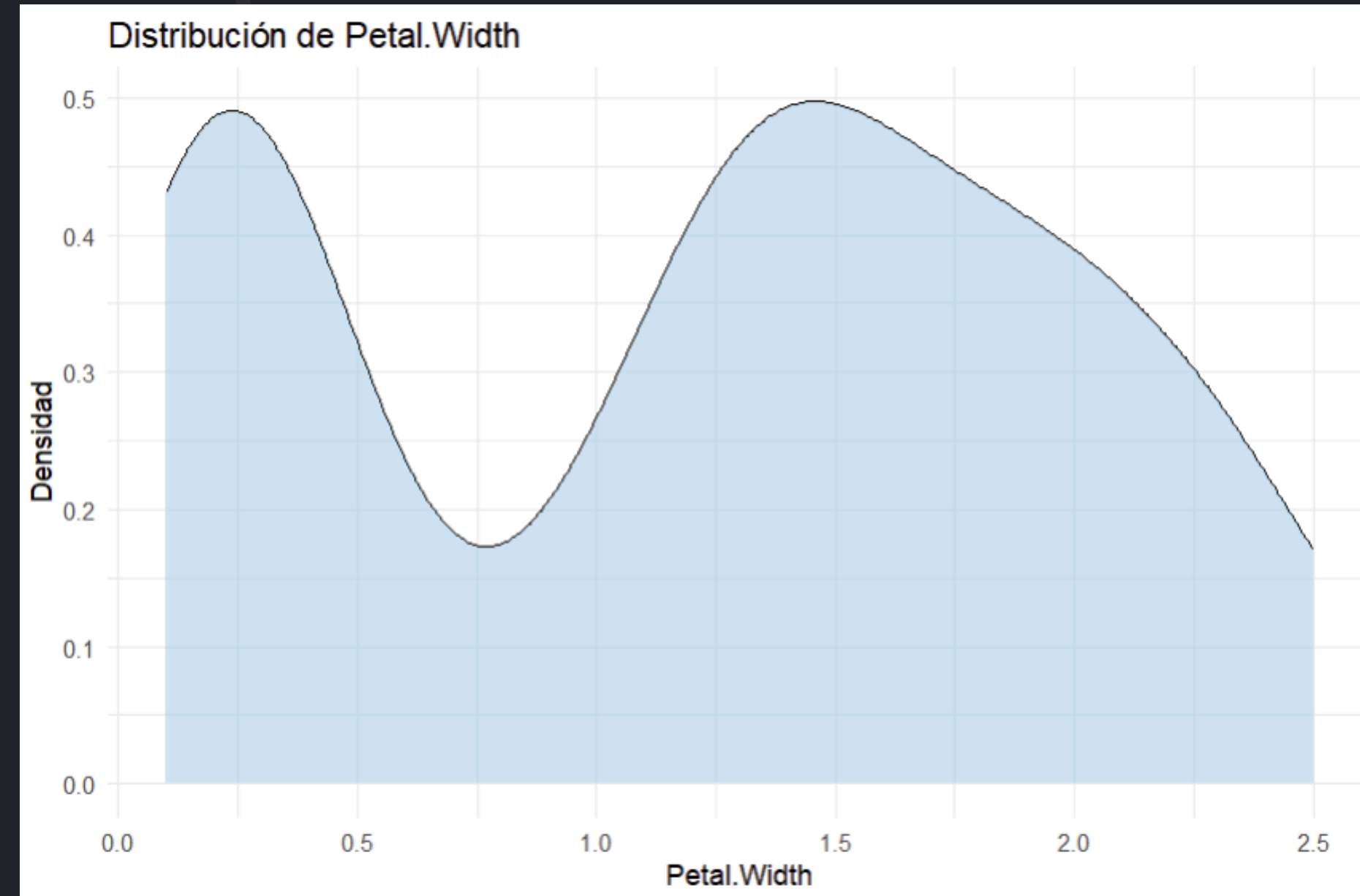


# Punto 1 {

Este código:

- Carga el conjunto de datos iris.
- Utiliza ggplot y geom\_density() para mostrar la distribución de cada variable.
- Itera sobre todas las variables relevantes para no repetir código.

```
1 # Cargamos las librerías necesarias
2 library(tidyverse)
3
4 # Cargamos el dataset iris
5 data(iris)
6
7 # Creamos una lista con los nombres de las variables numéricas
8 variables <- c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width")
9
10 # Para cada variable, generamos una gráfica de densidad
11 for (var in variables) {
12   ggplot(iris, aes_string(x = var)) +
13     geom_density(fill = "skyblue", alpha = 0.5) +
14     labs(title = paste("Distribución de", var), x = var, y = "Densidad") +
15     theme_minimal() -> p
16   print(p)
17 }
18 |
```



}

## #Punto 2 {

Explicación:

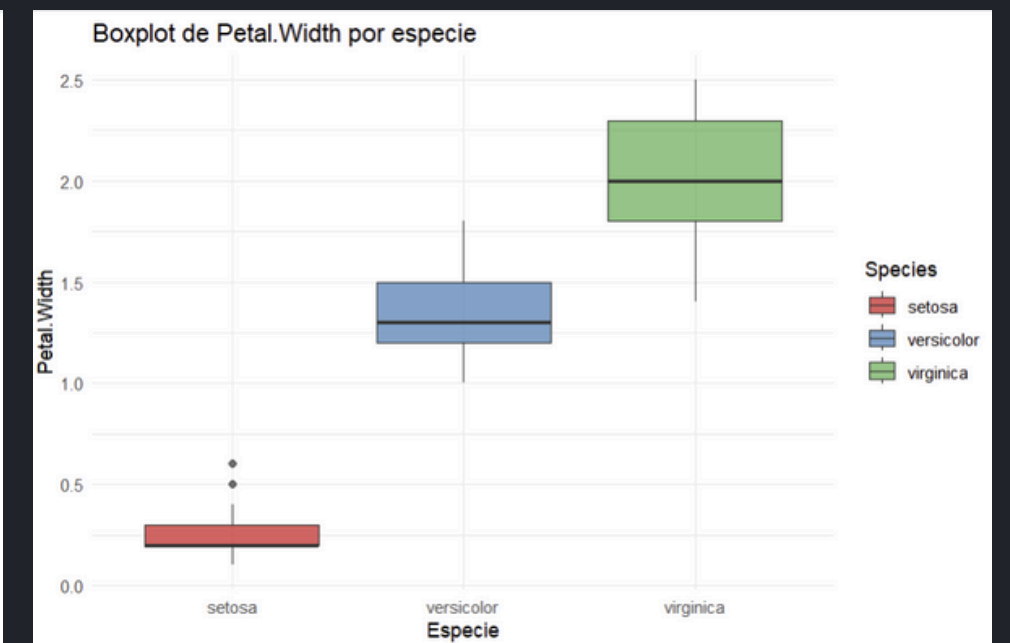
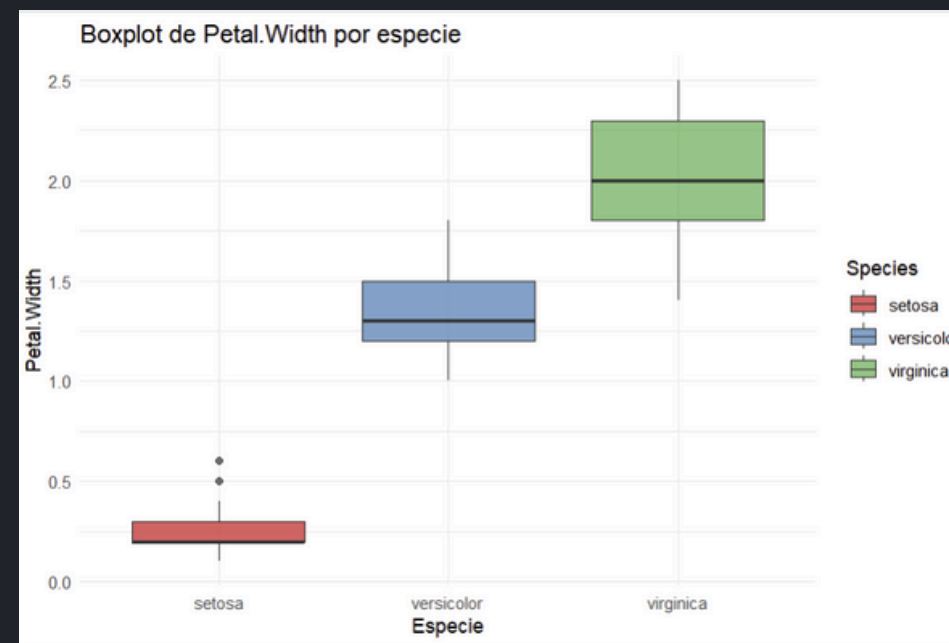
-geom\_density(): muestra las curvas de densidad para cada variable, diferenciadas por especie.

-geom\_boxplot(): muestra los boxplots para cada variable por especie, lo que permite comparar la distribución, medianas, rangos, etc.

-scale\_fill\_brewer(palette = "Set1"): añade colores diferenciados por especie para mayor claridad.

```
# 1. Distribuciones por especie usando geom_density()
for (var in variables) {
  ggplot(iris, aes_string(x = var, fill = "Species")) +
    geom_density(alpha = 0.5) +
    labs(title = paste("Distribución de", var, "por especie"),
         x = var, y = "Densidad") +
    theme_minimal() +
    scale_fill_brewer(palette = "Set1") -> p
  print(p)
}
```

```
# 2. Distribuciones por especie usando geom_boxplot()
for (var in variables) {
  ggplot(iris, aes_string(x = "Species", y = var, fill = "Species")) +
    geom_boxplot(alpha = 0.7) +
    labs(title = paste("Boxplot de", var, "por especie"),
         x = "Especie", y = var) +
    theme_minimal() +
    scale_fill_brewer(palette = "Set1") -> p
  print(p)
}
```



Conclusión:

Se observan diferencias significativas entre las especies:

- Setosa destaca por tener pétalos pequeños (longitud y ancho).
- Versicolor tiene valores intermedios.
- Virginica tiene pétalos más grandes (longitud y ancho).
- Los boxplots muestran claramente la diferencia en rango y valores medios entre especies.
- Las curvas de densidad también muestran picos diferentes y en distintas posiciones para cada especie.

## #Punto 3 {

### Conclusión:

-Sepal.Length y Sepal.Width tienen medias y medianas relativamente cercanas, lo que indica que las distribuciones no son muy sesgadas. La desviación estándar es baja (especialmente para Sepal.Width), por lo que la dispersión es moderada.

-Petal.Length y Petal.Width muestran una desviación estándar alta en comparación con sus medias, lo que sugiere que estas variables son más dispersas. Es posible que esto se deba a las diferencias marcadas entre las especies, como vimos en el punto 2.

-Conclusión general: Los sépalos muestran menor variabilidad y una distribución más simétrica, mientras que los pétalos presentan más dispersión, reflejando posiblemente la diferenciación entre especies. La mediana de Petal.Length (4.35) es mayor que su media (3.76), lo que sugiere una posible asimetría a la izquierda (valores más bajos concentrados).

```
#Punto3
# Calculamos media, mediana y desviación estándar de cada variable
iris %>%
  summarise(
    Sepal.Length_media = mean(Sepal.Length),
    Sepal.Length_mediana = median(Sepal.Length),
    Sepal.Length_sd = sd(Sepal.Length),
    Sepal.Width_media = mean(Sepal.Width),
    Sepal.Width_mediana = median(Sepal.Width),
    Sepal.Width_sd = sd(Sepal.Width),
    Petal.Length_media = mean(Petal.Length),
    Petal.Length_mediana = median(Petal.Length),
    Petal.Length_sd = sd(Petal.Length),
    Petal.Width_media = mean(Petal.Width),
    Petal.Width_mediana = median(Petal.Width),
    Petal.Width_sd = sd(Petal.Width)
  )
```

Sepal.Length_media	Sepal.Length_mediana	Sepal.Length_sd	Sepal.Width_media	Sepal.Width_mediana
5.843333	5.8	0.8280661	3.057333	3
Sepal.Width_sd	Petal.Length_media	Petal.Length_mediana	Petal.Length_sd	Petal.Width_media
0.4358663	3.758	4.35	1.765298	1.199333
Petal.Width_mediana	Petal.Width_sd			
1.3	0.7622377			

## #Punto 4 {

```
#Punto 4  
iris %>%  
  group_by(Species) %>%  
  summarise(count = n())
```

```
# A tibble: 3 x 2  
  Species      count  
  <fct>      <int>  
1 setosa         50  
2 versicolor    50  
3 virginica      50  
>
```

## #Punto 5 {

### CONCLUSIÓN

- Setosa tiene menor longitud de sépalo y ancho intermedio del sépalo.
- Versicolor y Virginica tienen sépalos más largos y anchos que Setosa, especialmente Virginica, que destaca por tener sépalos y pétalos más grandes.
- La variabilidad (desviación estándar) es mayor para Virginica, lo que indica una mayor dispersión en las medidas.

```
64 #Punto5
65 # Calculamos media, mediana y desviación típica por especie
66 iris %>%
67   group_by(Species) %>%
68   summarise(
69     Sepal.Length_media = mean(Sepal.Length),
70     Sepal.Length_mediana = median(Sepal.Length),
71     Sepal.Length_sd = sd(Sepal.Length),
72     Sepal.Width_media = mean(Sepal.Width),
73     Sepal.Width_mediana = median(Sepal.Width),
74     Sepal.Width_sd = sd(Sepal.Width),
75     Petal.Length_media = mean(Petal.Length),
76     Petal.Length_mediana = median(Petal.Length),
77     Petal.Length_sd = sd(Petal.Length),
78     Petal.Width_media = mean(Petal.Width),
79     Petal.Width_mediana = median(Petal.Width),
80     Petal.Width_sd = sd(Petal.Width)
81   )
```

```
# A tibble: 3 × 13
  Species      Sepal.Length_media Sepal.Length_mediana Sepal.Length_sd Sepal.Width_media
  <fct>          <dbl>          <dbl>          <dbl>          <dbl>
1 setosa         5.01             5             0.352          3.43
2 versicolor     5.94             5.9           0.516          2.77
3 virginica      6.59             6.5           0.636          2.97
# i 8 more variables: Sepal.Width_mediana <dbl>, Sepal.Width_sd <dbl>,
#   Petal.Length_media <dbl>, Petal.Length_mediana <dbl>, Petal.Length_sd <dbl>,
#   Petal.Width_media <dbl>, Petal.Width_mediana <dbl>, Petal.Width_sd <dbl>
>
```

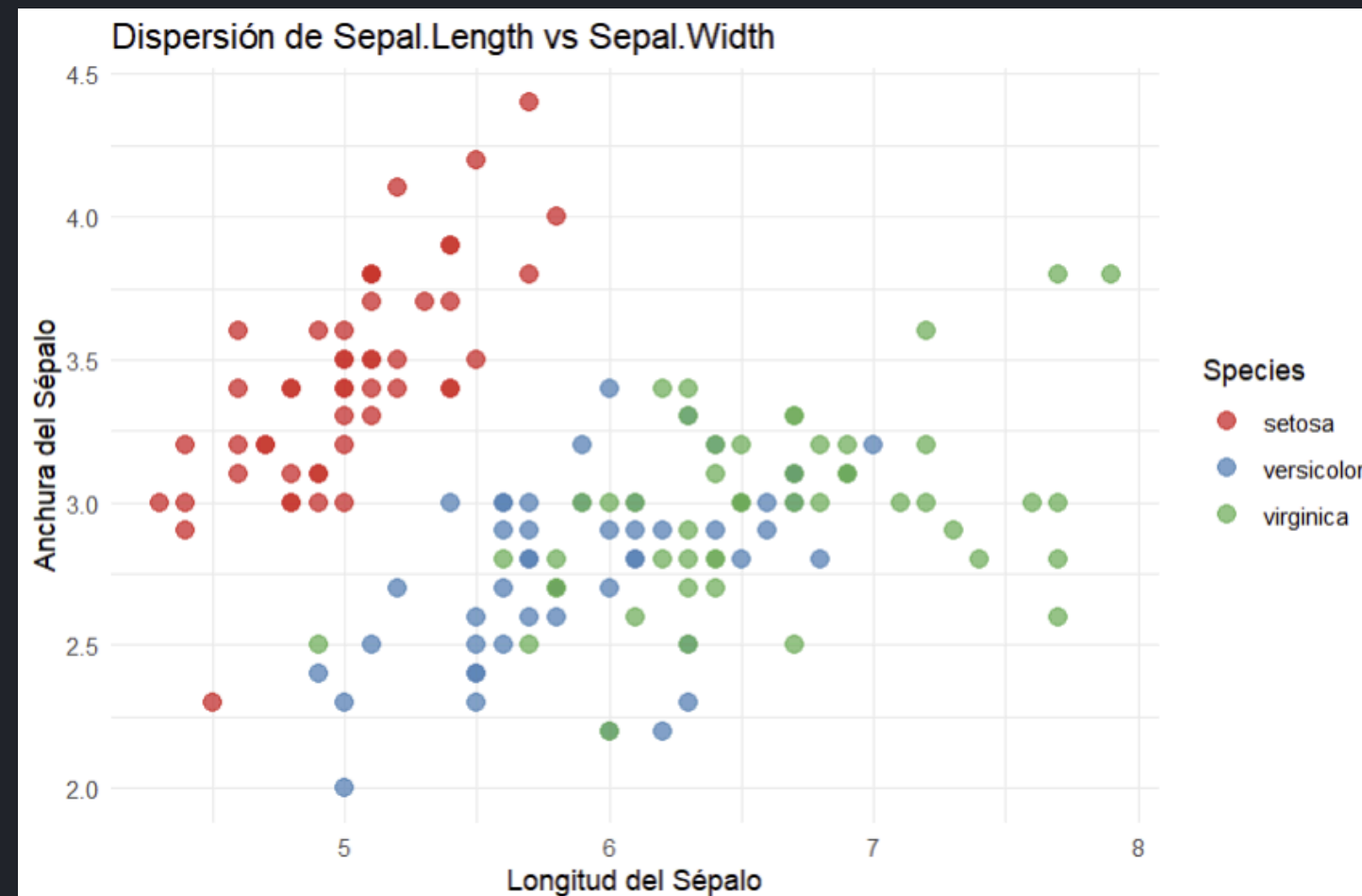


## #Punto 6 {

### CONCLUSIÓN

- Setosa se distingue claramente de las otras especies, concentrándose en la zona de sépalos más anchos (Sepal.Width) y más cortos (Sepal.Length).
- Versicolor y Virginica se solapan más, pero Virginica tiene tendencia a sépalos más largos y un poco más anchos.
- Este gráfico muestra que estas dos medidas permiten distinguir bien a Setosa, pero no tanto a Versicolor y Virginica.

```
#Punto 6
# Gráfico de dispersión Sepal.Length vs Sepal.Width coloreado por Species
ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(title = "Dispersión de Sepal.Length vs Sepal.Width",
       x = "Longitud del Sépalo",
       y = "Anchura del Sépalo") +
  theme_minimal() +
  scale_color_brewer(palette = "Set1")
```

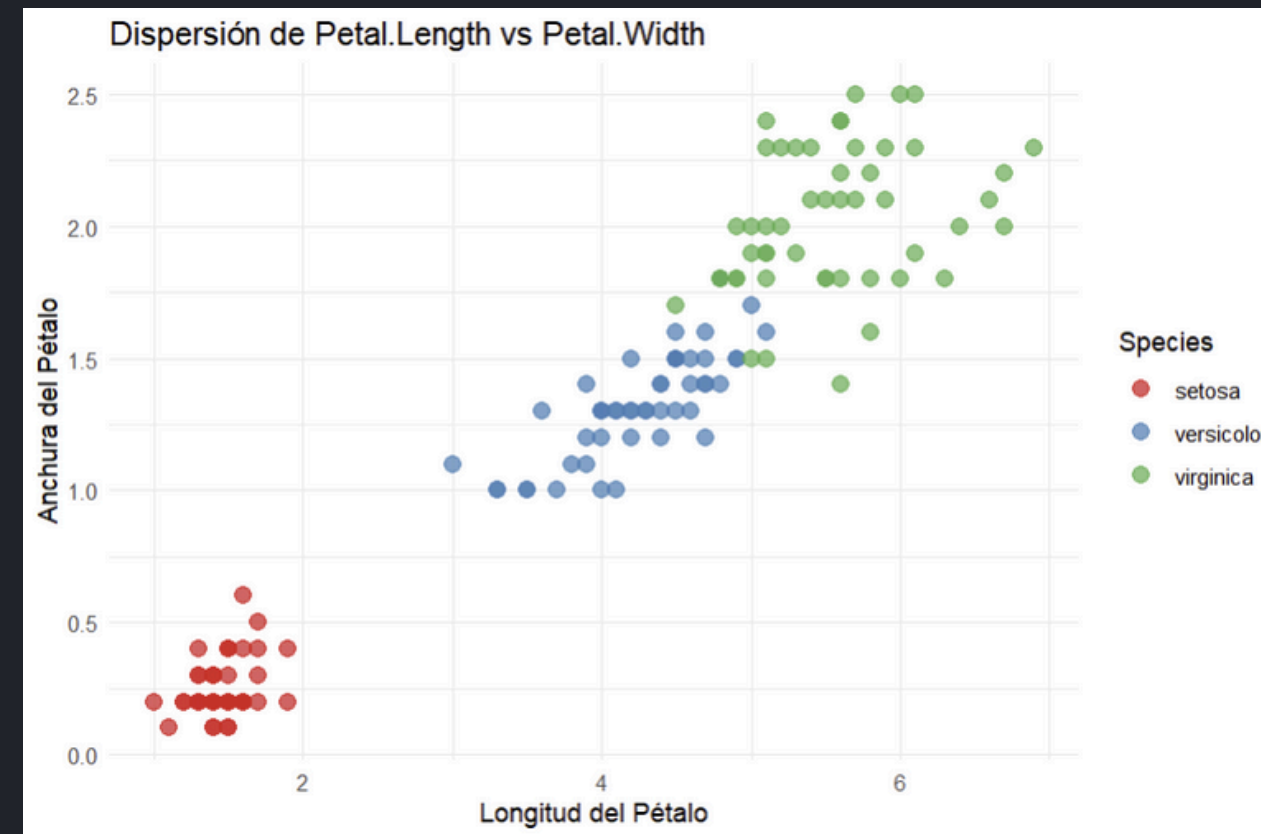


## #Punto 7 {

### CONCLUSIÓN

- Setosa (rojo) se distingue claramente por tener pétalos cortos y estrechos.
- Versicolor (azul) y Virginica (verde) muestran una relación lineal positiva entre la longitud y el ancho de los pétalos.
- Virginica tiene los pétalos más largos y anchos, mientras que Versicolor está en un rango intermedio.
- Las tres especies están claramente separadas, lo que indica que estas variables son muy útiles para la clasificación.

```
#Punto 7
# Gráfico de dispersión Petal.Length vs Petal.Width coloreado por Species
ggplot(iris, aes(x = Petal.Length, y = Petal.Width, color = Species)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(title = "Dispersión de Petal.Length vs Petal.Width",
       x = "Longitud del Pétalo",
       y = "Anchura del Pétalo") +
  theme_minimal() +
  scale_color_brewer(palette = "Set1")
```





## #Punto 8 {

### CONCLUSIÓN

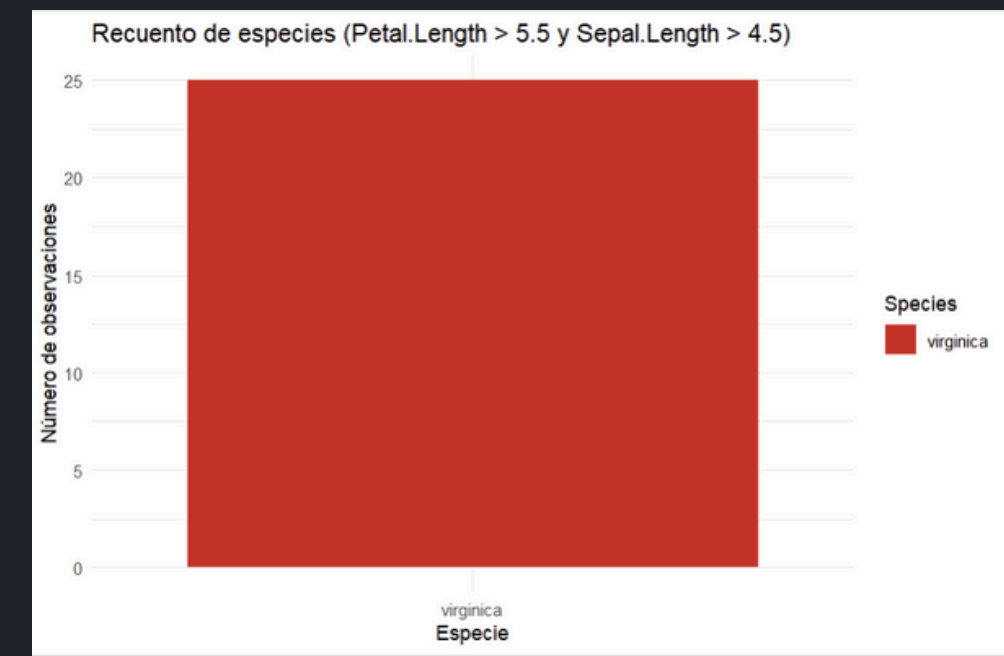
- Solo Virginica tiene observaciones que cumplen la condición  $\text{Petal.Length} > 5.5$  y  $\text{Sepal.Length} > 4.5$ .
- Por eso, el gráfico muestra solo una barra alta para Virginica, y no aparecen barras para Setosa ni Versicolor.

```
#Punto 8
# Crear un subconjunto con las condiciones indicadas
iris_sub <- iris %>%
  filter(Petal.Length > 5.5, Sepal.Length > 4.5)

# Contar el número de observaciones por especie en el subconjunto
iris_sub %>%
  count(Species)

# Crear un diagrama de barras con los recuentos
ggplot(iris_sub, aes(x = Species, fill = Species)) +
  geom_bar() +
  labs(title = "Recuento de especies (Petal.Length > 5.5 y Sepal.Length > 4.5)",
       x = "Especie",
       y = "Número de observaciones") +
  theme_minimal() +
  scale_fill_brewer(palette = "Set1")
```

Species	n
1 virginica	25



## #Punto 9 {

### CONCLUSIÓN

- Todas las observaciones del subconjunto pertenecen a la especie Virginica, lo que confirma que solo esta especie tiene pétalos largos y sépalos grandes.
- La distribución de los puntos muestra que los pétalos de Virginica tienen un rango amplio tanto en longitud como en anchura.
- La relación entre Petal.Length y Petal.Width sigue siendo positiva, con una ligera dispersión.

#Punto 9

```
# Gráfico de dispersión Petal.Length vs Petal.Width para el subconjunto
ggplot(iris_sub, aes(x = Petal.Length, y = Petal.Width, color = Species)) +
  geom_point(size = 3, alpha = 0.7) +
  labs(title = "Dispersión de Petal.Length vs Petal.Width (Subconjunto)",
       x = "Longitud del Pétalo",
       y = "Anchura del Pétalo") +
  theme_minimal() +
  scale_color_brewer(palette = "Set1")
```



# #Punto 10 {

## CONCLUSIÓN GENERAL

- Durante este análisis exploramos el conjunto de datos iris, usando R y tidyverse para visualizar y entender sus variables. Las distribuciones de las medidas (longitud y anchura del sépalo y pétalo) mostraron que las especies presentan diferencias notables, especialmente en las medidas del pétalo.
- Los gráficos de densidad y boxplots confirmaron que setosa tiene pétalos más pequeños, mientras que virginica destaca por pétalos más grandes y mayor variabilidad. En comparación, versicolor ocupa un punto intermedio. Las estadísticas descriptivas (media, mediana y desviación estándar) respaldaron estas diferencias.
- En los gráficos de dispersión, vimos que setosa se distingue claramente de las otras dos especies, mientras que versicolor y virginica están más cerca, pero con diferencias visibles en tamaño. Al filtrar los datos para incluir solo observaciones con  $\text{Petal.Length} > 5.5$  y  $\text{Sepal.Length} > 4.5$ , solo quedaron muestras de virginica, lo cual muestra que esta especie es la que presenta mayores medidas.
- En resumen, las visualizaciones y cálculos nos ayudaron a ver claramente cómo se diferencian las especies. Se pudo identificar que las medidas del pétalo son especialmente útiles para distinguirlas. La especie setosa destaca por sus medidas pequeñas y virginica por las más grandes, mientras que versicolor queda en un punto intermedio.

