

Bioinformática y análisis de datos ómicos

TEMA 5: REPRESENTACIÓN GRÁFICA DE DATOS DE ORIGEN BIOLÓGICO



Ignacio Varela Egocheaga

DEPARTAMENTO DE BIOLOGÍA MOLECULAR

Este material se publica bajo la siguiente licencia:

[Creative Commons BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Tema 5: Representación gráfica de datos de origen biológico

5.1 Sintaxis básica del paquete ggplot2

5.2 Elementos geométricos

5.3 Elementos estéticos del gráfico

5.4 Transformaciones y escalas

5.5 Estética avanzada

Tema 5: Representación gráfica de datos de origen biológico

5.1 Sintaxis básica del paquete ggplot2

5.2 Elementos geométricos

5.3 Elementos estéticos del gráfico

5.4 Transformaciones y escalas

5.5 Estética avanzada

Sintaxis básica de un gráfico ggplot2

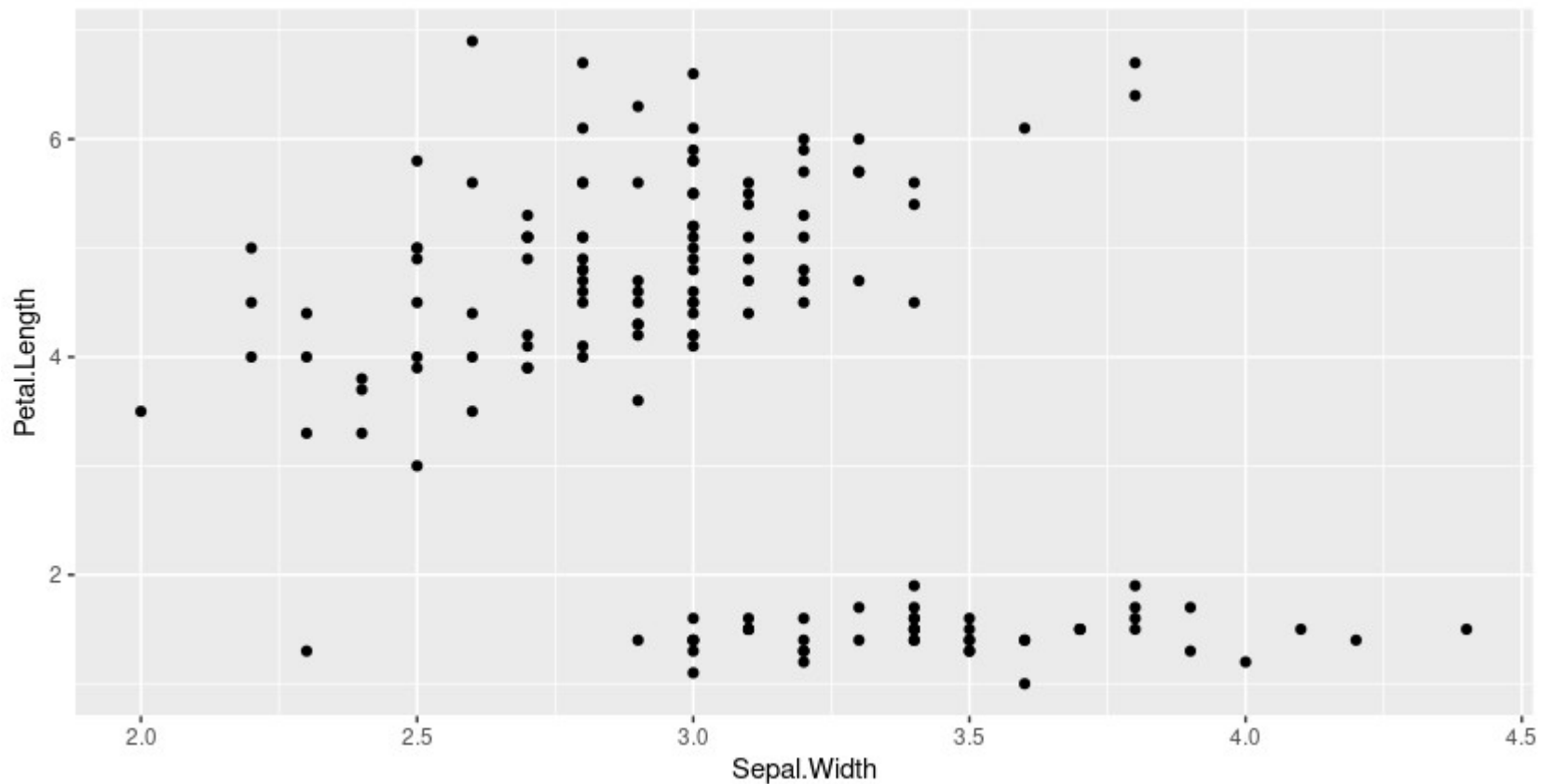
ggplot(data) +	←	Orden principal y origen datos
geom_	←	Tipo de gráfico
aes()	←	Variables para representar
stat_	←	Transformaciones estadísticas
labs()	←	Títulos de ejes
lims()	←	Escalas de los ejes
theme()	←	Estética avanzada
.....		

Gráfico sencillo de nube de puntos

Iris dataframe

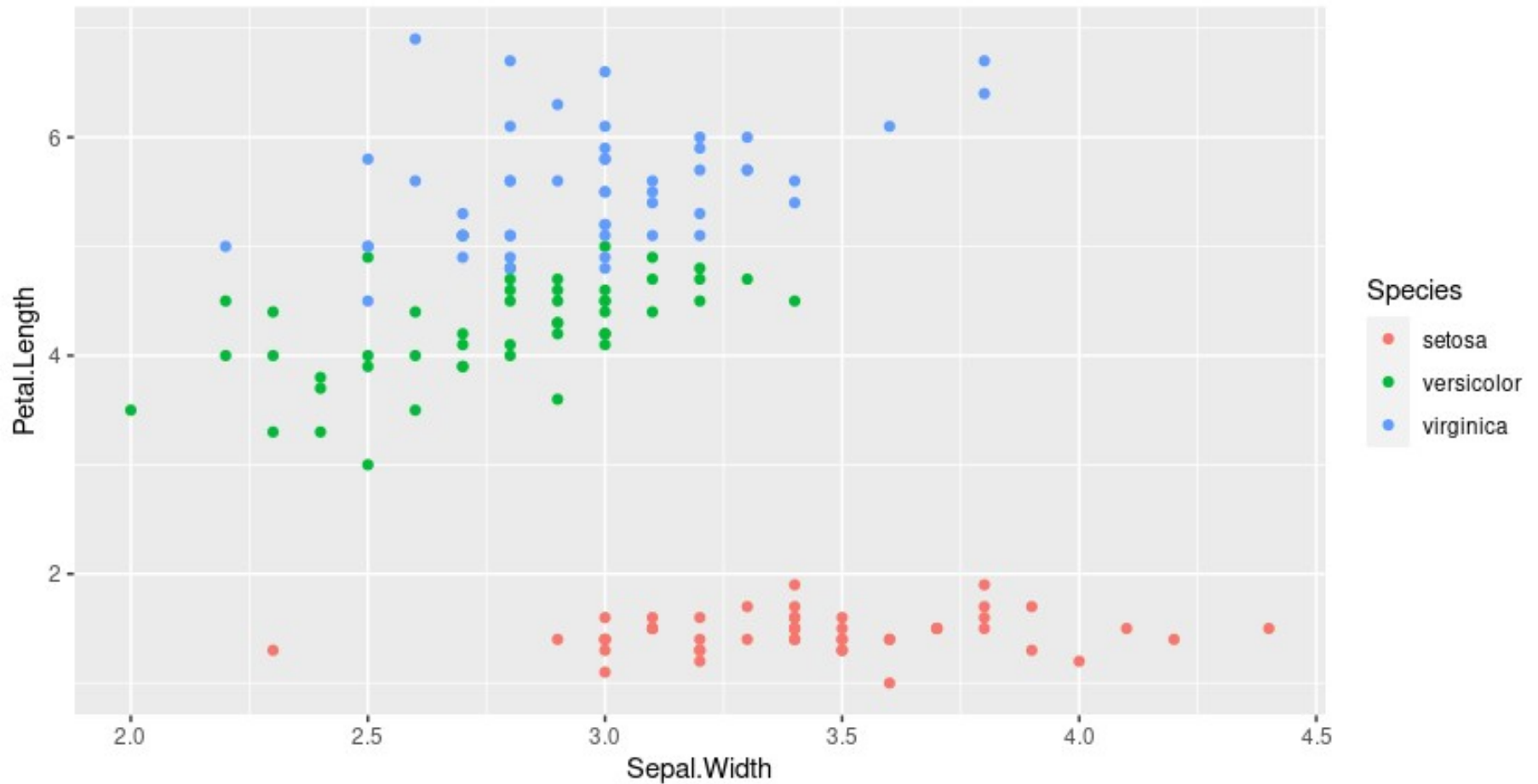
	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5.0	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length)) + geom_point()
```



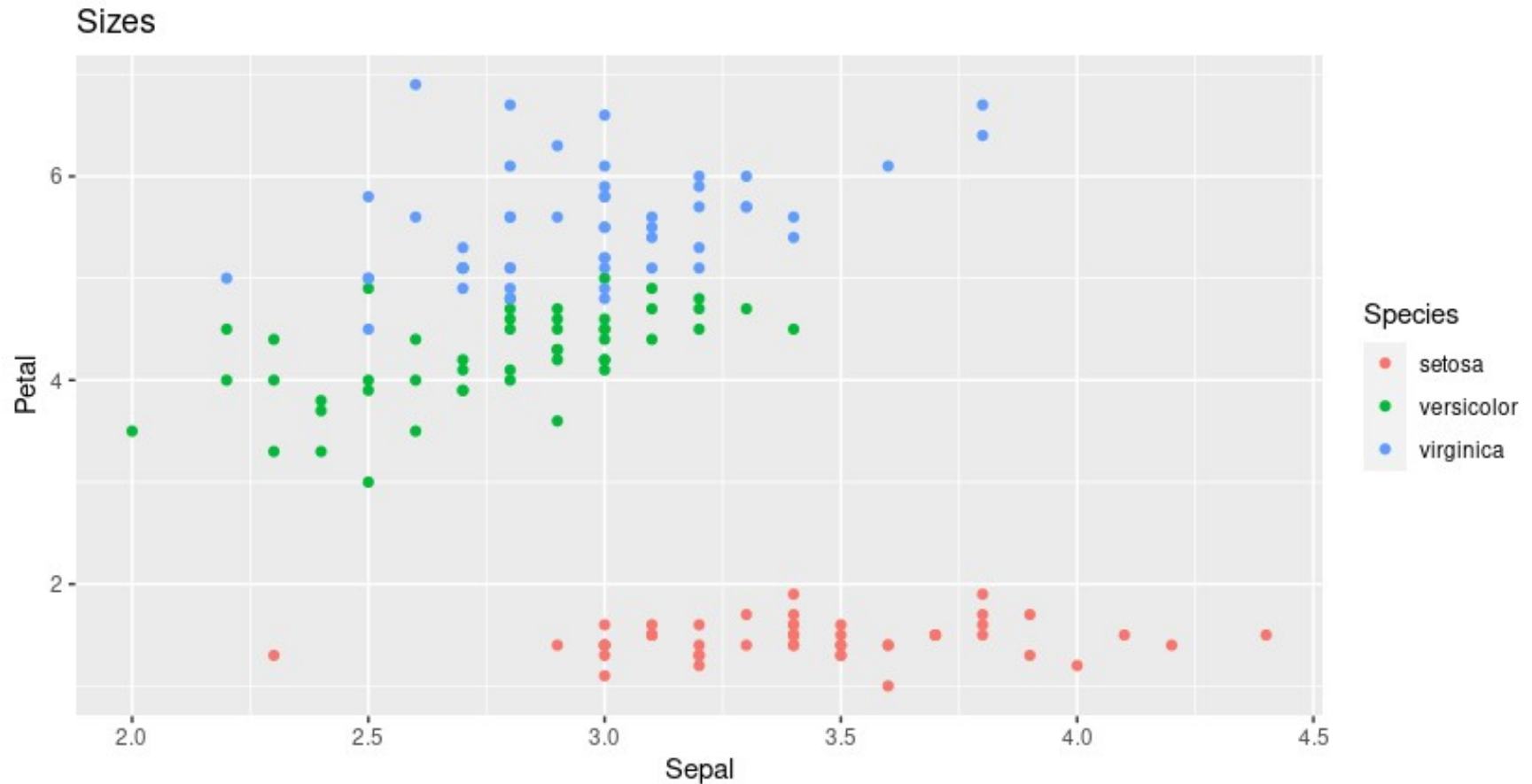
Representar más de 2 variables

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()
```



Adición de más capas

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point() +  
labs((x="Sepal", y = "Petal", title = "Sizes"))
```



Tema 5: Representación gráfica de datos de origen biológico

5.1 Sintaxis básica del paquete ggplot2

5.2 Elementos geométricos

5.3 Elementos estéticos del gráfico

5.4 Transformaciones y escalas

5.5 Estética avanzada

Tipos de elementos geométricos

geom_abline	55
geom_bar	58
geom_bin_2d	62
geom_blank	64
geom_boxplot	65
geom_contour	69
geom_count	74
geom_crossbar	77
geom_density	80
geom_density_2d	84
geom_dotplot	88
geom_errorbarh	92
geom_freqpoly	94
geom_function	99
geom_hex	102
geom_jitter	104
geom_label	106
geom_map	111
geom_path	114
geom_point	118
geom_polygon	120
geom_qq_line	123
geom_quantile	127
geom_raster	129
geom_ribbon	132
geom_rug	135
geom_segment	138
geom_smooth	141
geom_spoke	145
geom_violin	147
get_alt_text	151

Diagrama tipo boxplot

```
ggplot(data=iris,aes(y=Petal.Width, x=Species, fill=Species)) + geom_boxplot()
```

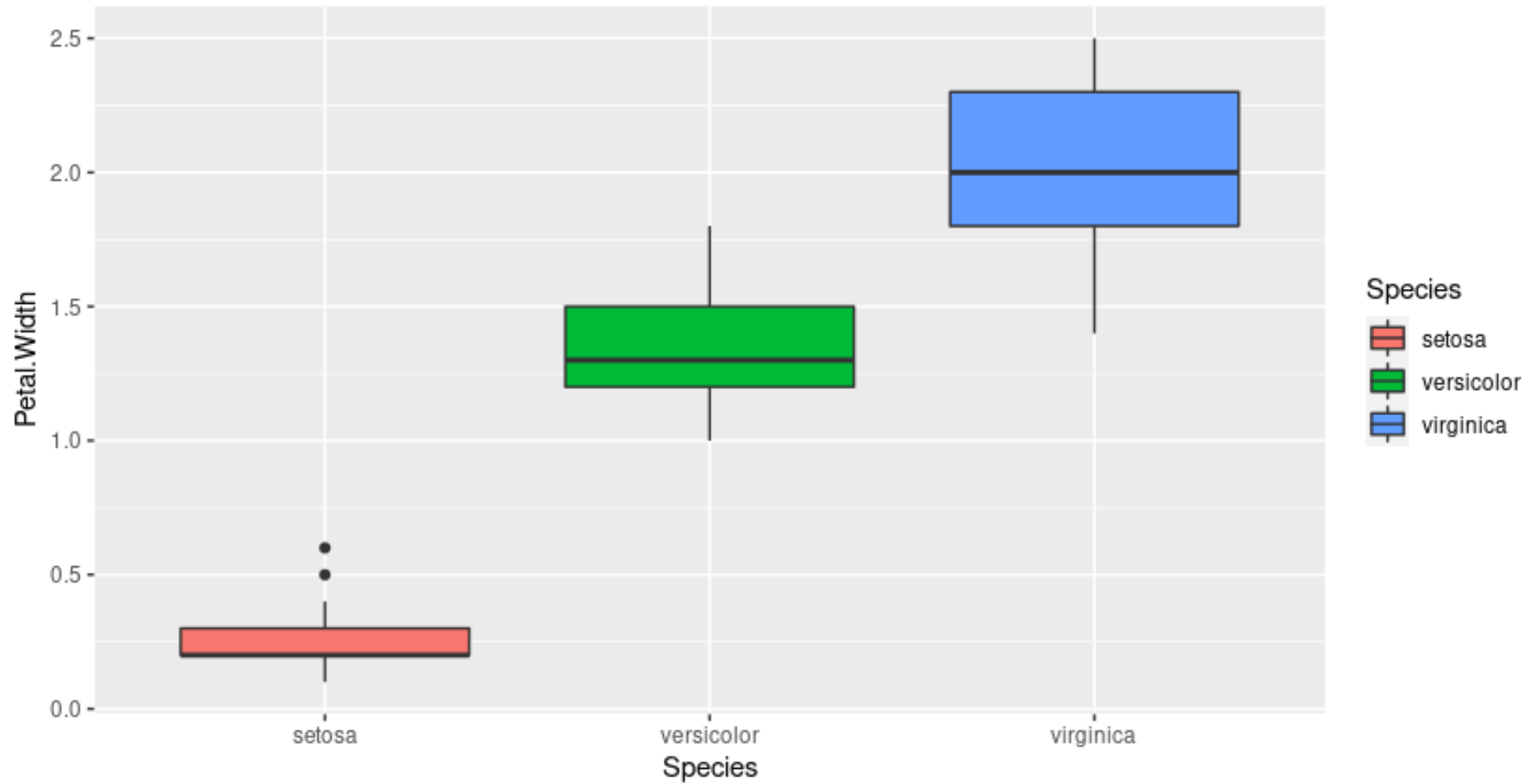
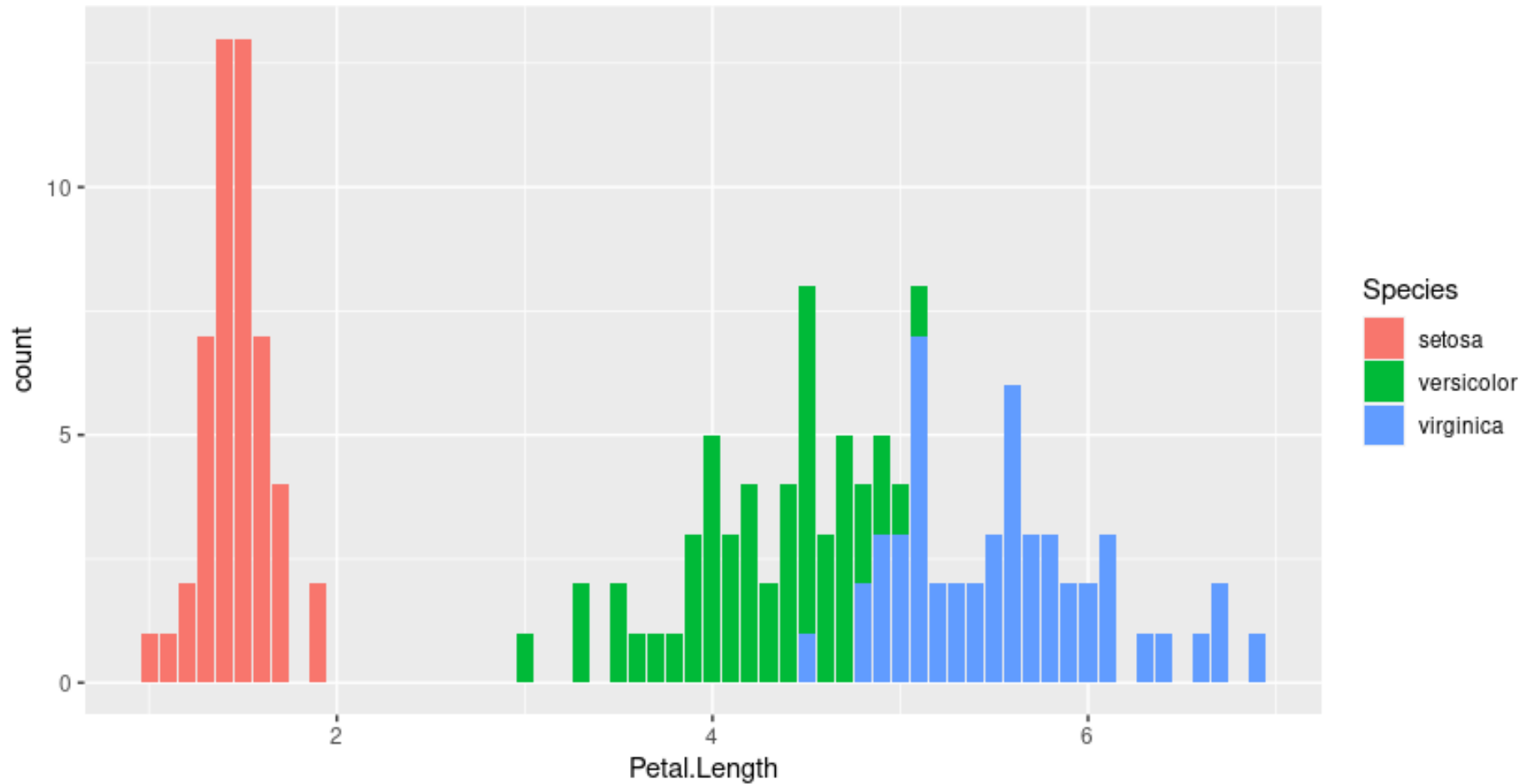


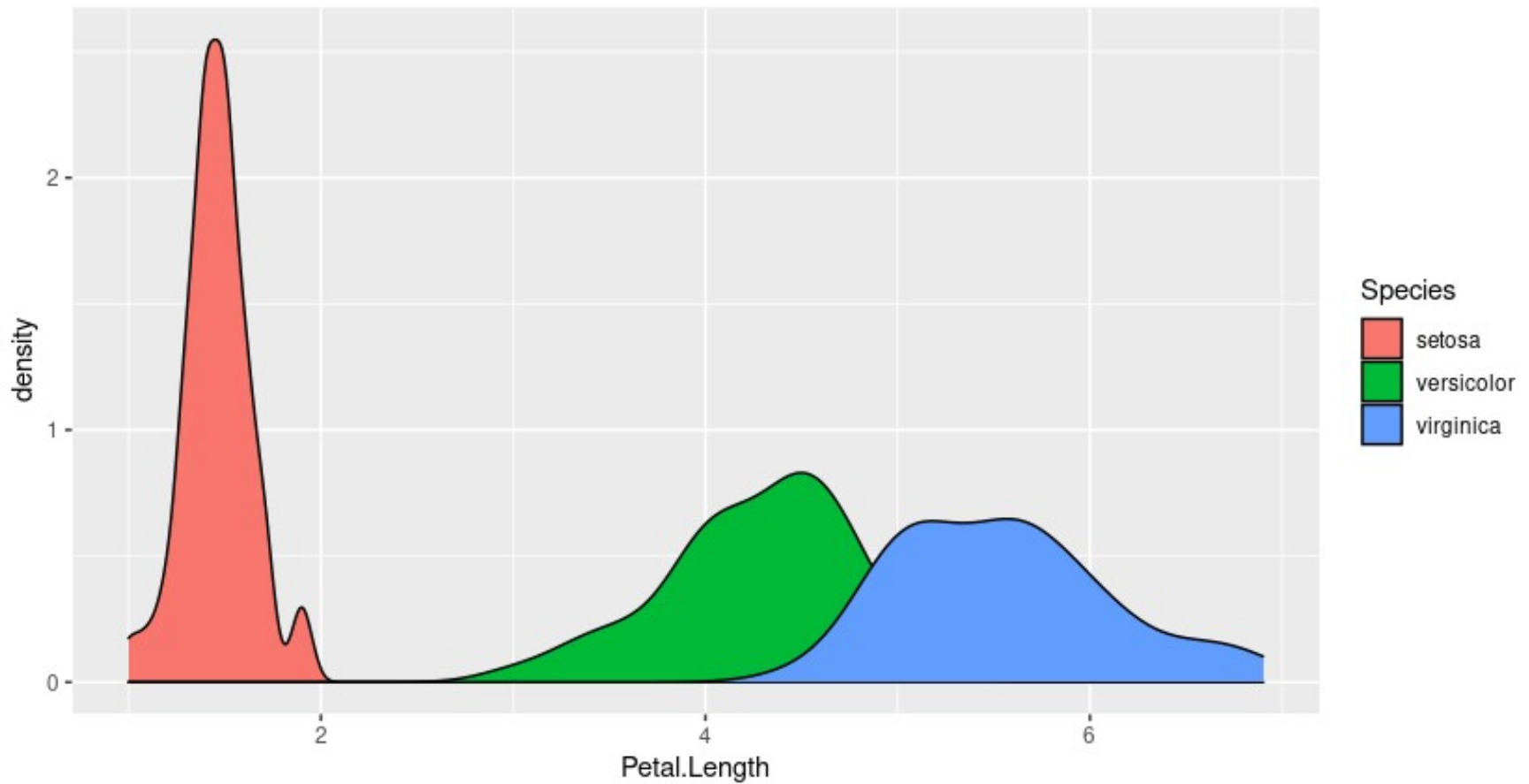
Diagrama de barras

```
ggplot(data=iris,aes( x=Petal.Length, fill=Species)) + geom_bar()
```



Línea suavizada de frecuencia

```
ggplot(data=iris,aes( x=Petal.Length, fill=Species)) + geom_density()
```

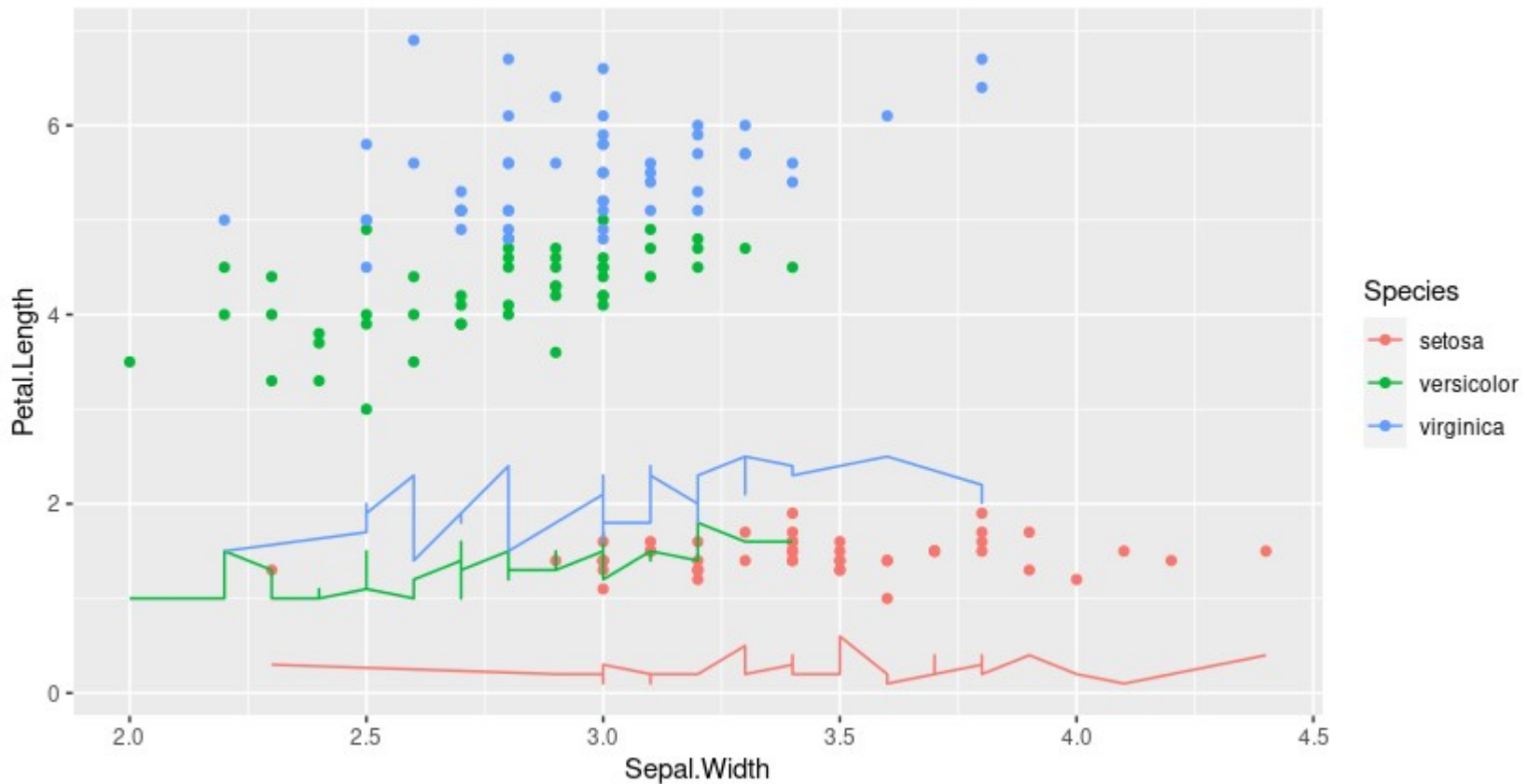


Combinar varias geometrías

Afecta a todo el gráfico

Afecta sólo a esta geometría

```
ggplot(data=iris,aes(x=Sepal.Width, color=Species)) + geom_point(aes(y=Petal.Length))  
+ geom_line(aes(y=Petal.Width))
```



Tema 5: Representación gráfica de datos de origen biológico

5.1 Sintaxis básica del paquete ggplot2

5.2 Elementos geométricos

5.3 Elementos estéticos del gráfico

5.4 Transformaciones y escalas

5.5 Estética avanzada

Estilo de puntos y líneas

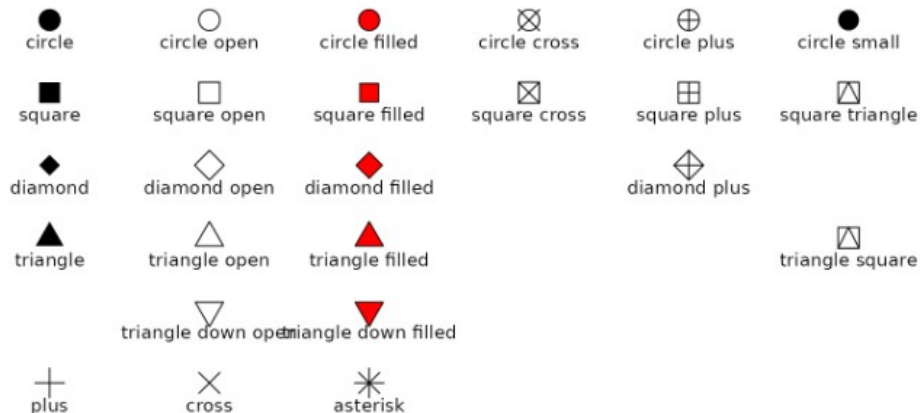
Siempre fuera del bloque aes()

Colour, size y fill

657 colores con nombres (comprobar con `colours()`)

Tb se pueden especificar con los tres componentes RGB en hexagesimal y el último componente para la transparencia
`#RRGGBBAA`

shape



Tipos de línea (linetype)

solid

dashed

dotted

dotdash

longdash

twodash

Modificar estética del gráfico

```
ggplot(data=iris,aes(x=Sepal.Width, colour=Species)) + geom_point(aes(y=Petal.Length),  
  shape="diamond", size=3) + geom_line(aes(y=Petal.Width), linetype="dashed")
```



Estética del texto

family (fuente)

library(extrafont) para más fuentes

Si no:

sans

serif

mono

fontface

bold.italic

italic

bold

plain

size

Por defecto en mm pero se puede especificar en puntos con **.pt**

Justification

hjust o vjust

- top = 1, middle = 0.5, bottom = 0
- left = 0, center = 0.5, right = 1

Tema 5: Representación gráfica de datos de origen biológico

5.1 Sintaxis básica del paquete ggplot2

5.2 Elementos geométricos

5.3 Elementos estéticos del gráfico

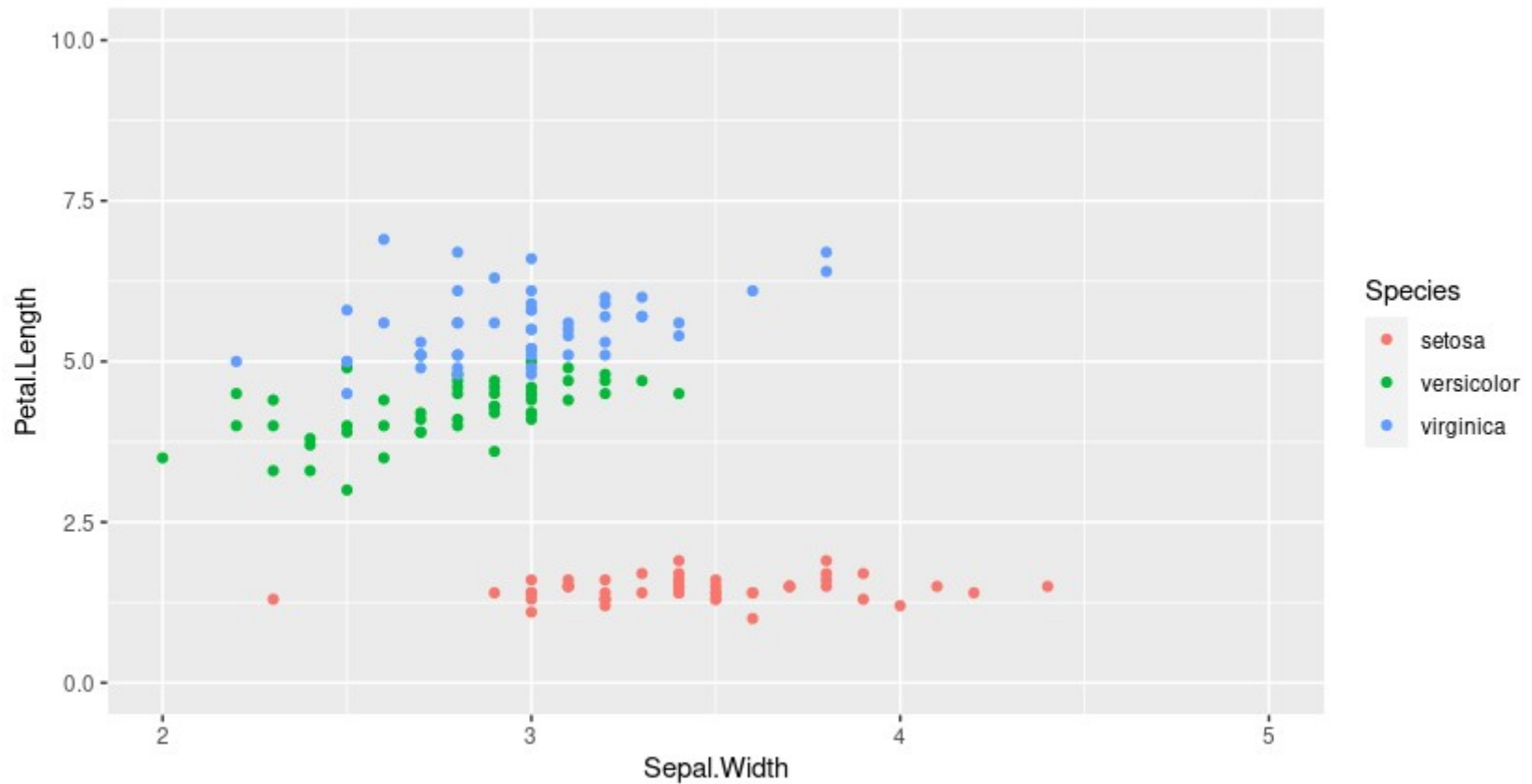
5.4 Transformaciones y escalas

5.5 Estética avanzada

Transformación de los ejes

Límite de los ejes con xlim(start,end) y ylim(start,end)

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point() + xlim(2,5) + ylim(0,10)
```



Transformación de los ejes

Generación de ejes en escalas no lineales

La función general `coord_trans(x="log10",y="log10")` permite realizar muchas transformaciones en los ejes como `log2`, `log10`, `sqrt`, etc.

Existen funciones específicas para las transformaciones más habituales

`scale_x_log10()` o `scale_y_log10()`

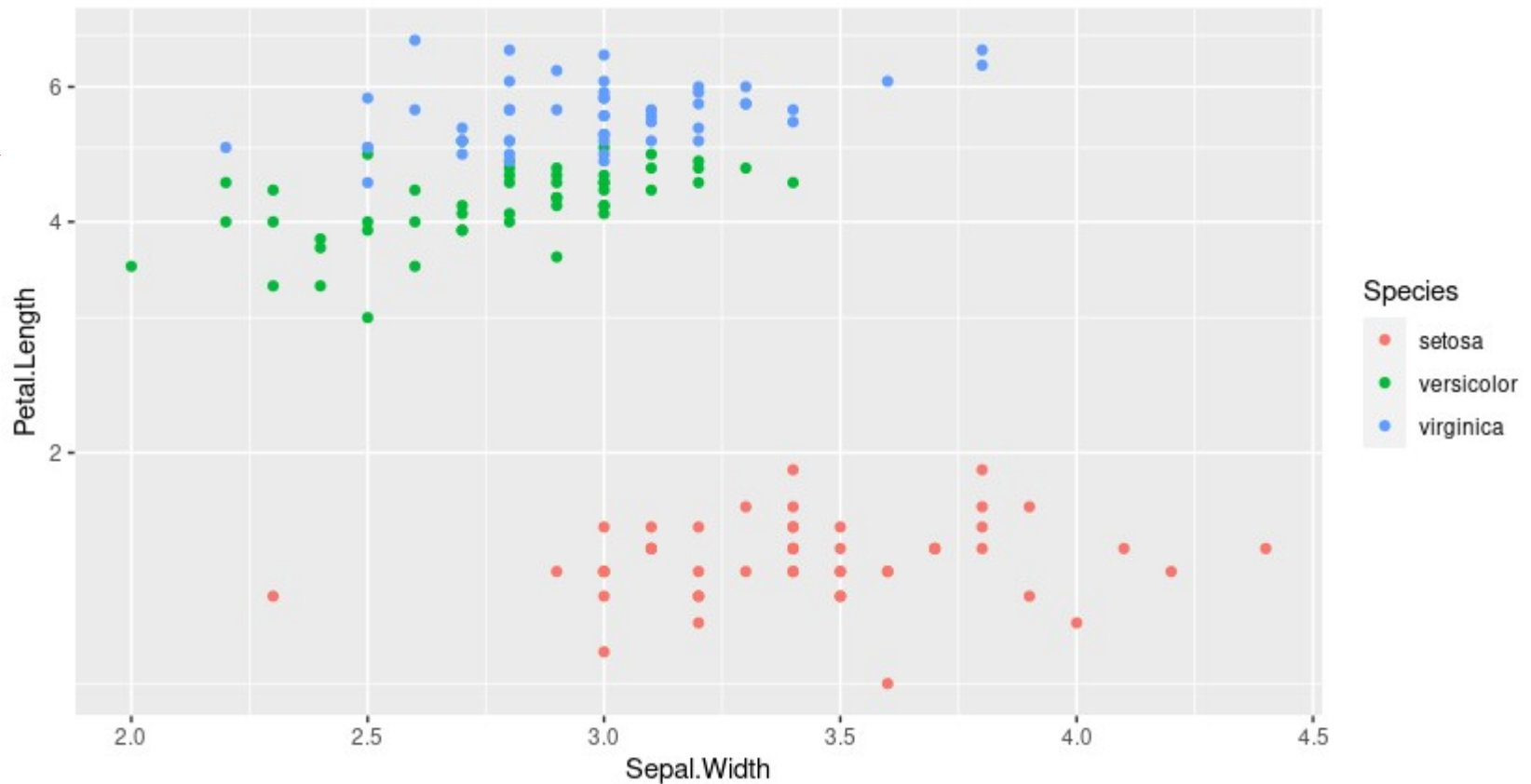
`scale_x_sqrt()` o `scale_y_sqrt()`

`scale_x_reverse()` o `scale_y_reverse()` ← invierte el orden de la escala

Transformación de los ejes

Generación de ejes en escalas no lineales

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()  
+ coord_trans(y="log10")
```



Tema 5: Representación gráfica de datos de origen biológico

5.1 Sintaxis básica del paquete ggplot2

5.2 Elementos geométricos

5.3 Elementos estéticos del gráfico

5.4 Transformaciones y escalas

5.5 Estética avanzada

Etiquetado de ejes, título y subtítulo

La función general `labs()` permite rotular ejes, título y subtítulo del gráfico. Contiene los siguientes elementos:

`x` ← eje X.

`y` ← eje Y.

`colour` ← Leyenda de colores si la hubiere

`shape` ← Leyenda de formas si la hubiere

`title` ← Título del gráfico

`subtitle` ← Subtítulo

`caption` ← Texto que se muestra en la parte inferior derecha del gráfico

`tag` ← Texto que se muestra en la parte superior izquierda del gráfico

Existen también funciones específicas para las tareas más habituales:

`xlab()`

`ylab()`

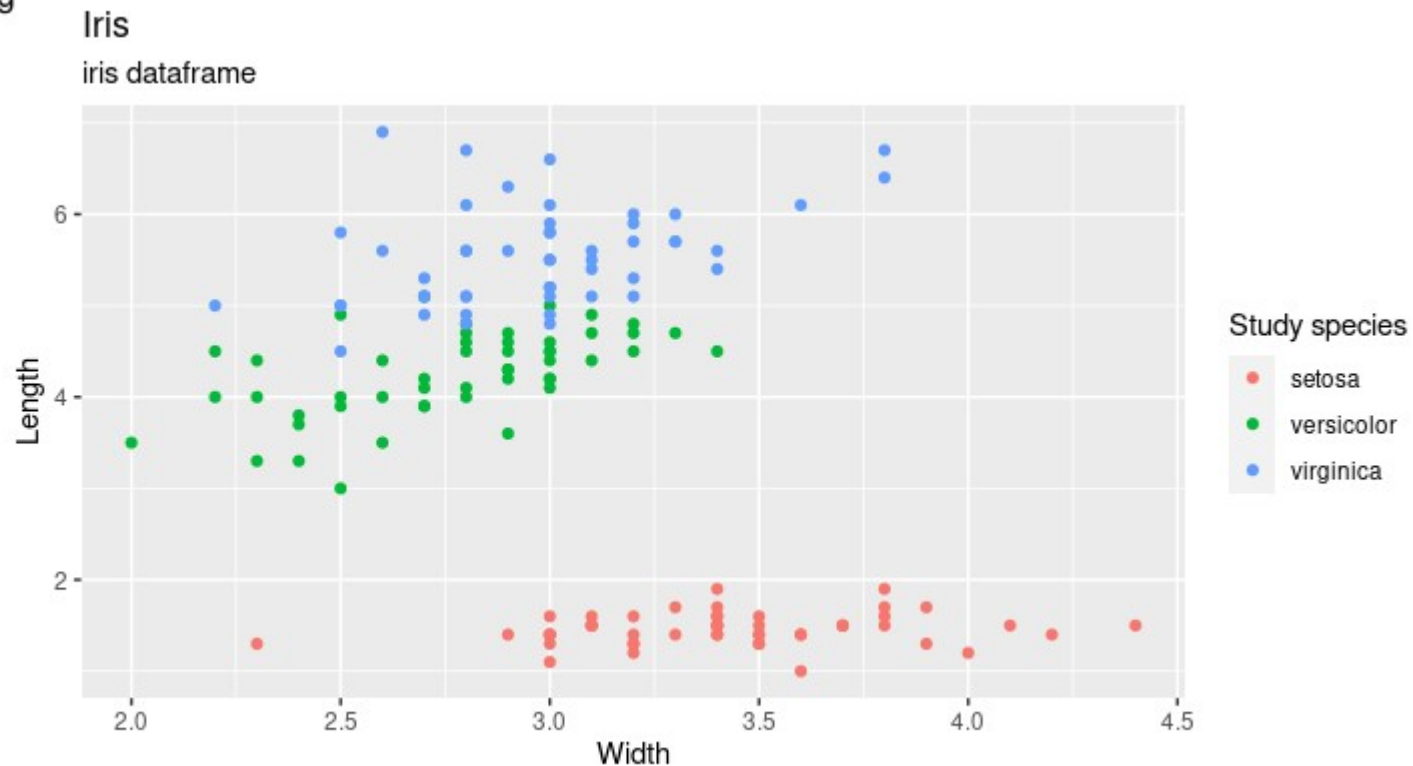
`ggtitle()`

Etiquetado de ejes, título y subtítulo

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()  
+ labs(x="Width", y="Length", title="Iris",subtitle = "iris dataframe", colour="Study species",  
caption="This is the caption", tag="This is the tag")
```

```
g ← ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()  
g + labs(x="Width", y="Length", title="Iris",subtitle = "iris dataframe", colour="Study species",  
caption="This is the caption", tag="This is the tag")
```

This is the tag



This is the caption

Modificado avanzado de la estética del gráfico

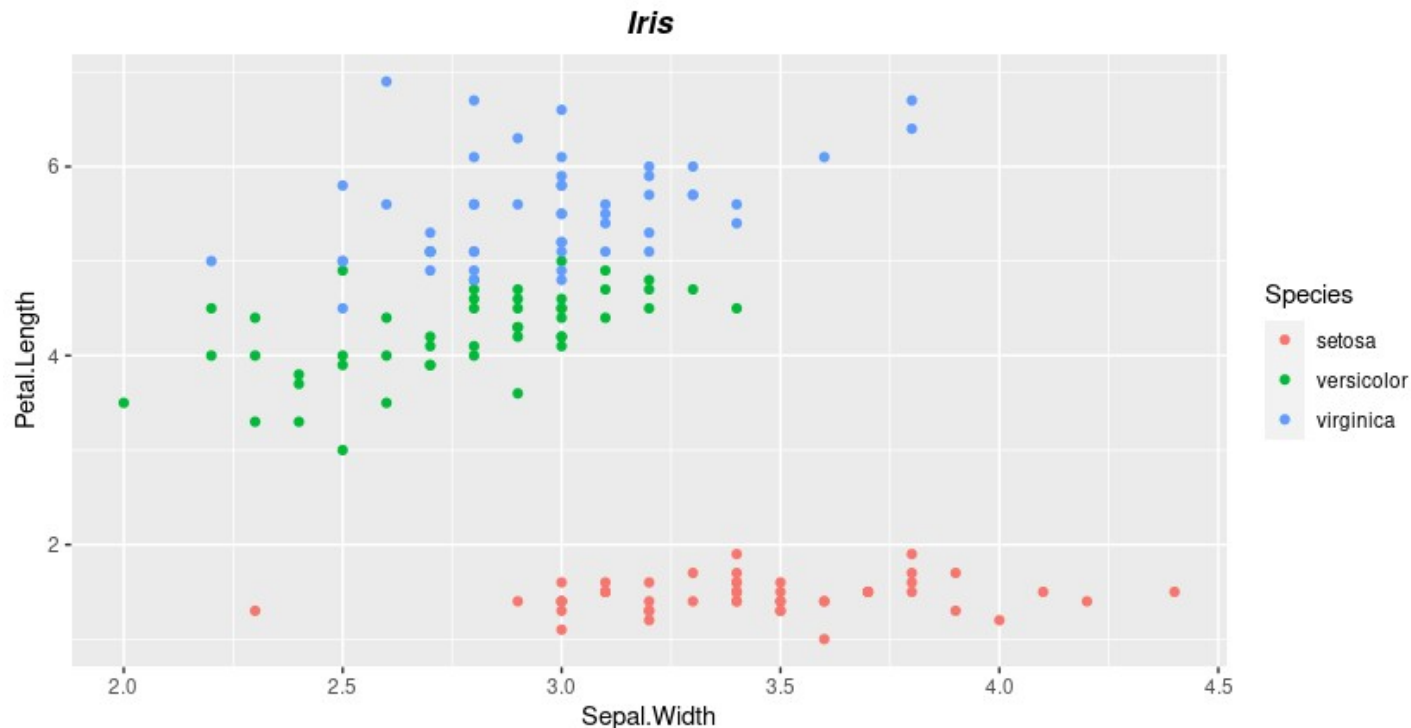
La función general `theme()` permite cambiar una gran cantidad de elementos estéticos del gráfico, para una lista exhaustiva: <https://ggplot2.tidyverse.org/reference/theme.html>

Dentro de cada elemento(`plot.title`,`axis.title`,etc.) puedes cambiar la línea, el texto con (`element_text`, `element_line`, `element_rect`,)

Para cambiar el texto de los ejes:

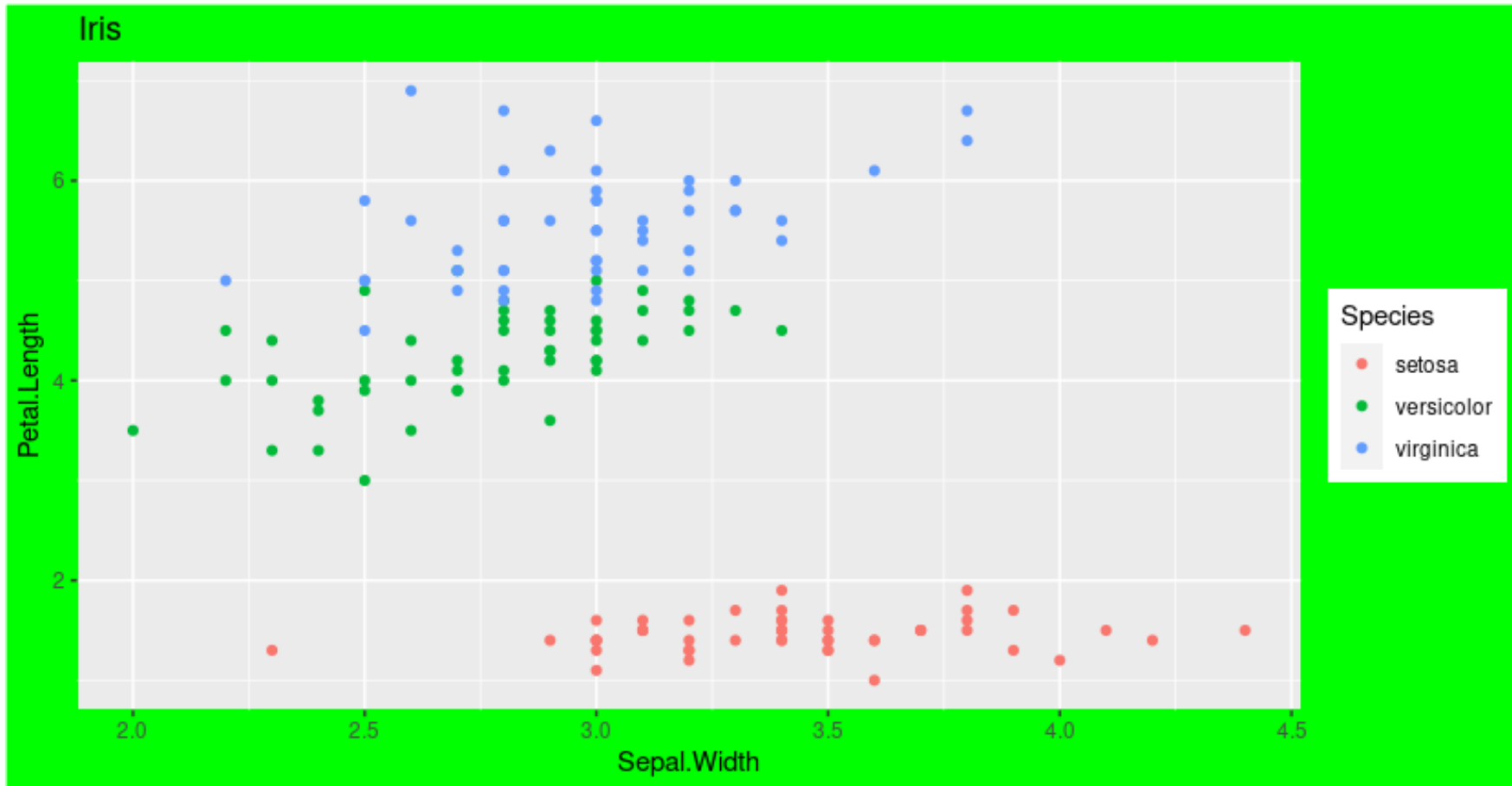
`plot.title`, `axis.title.x` o `axis.title.y` = `element.text(family, face, colour, size, hjust,vjust,)`

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()  
+ labs(title="Iris") + theme(plot.title=element_text(face="bold.italic", hjust=0.5))
```



Fondos

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()  
+ labs(title="Iris") + theme(plot.background=element_rect(fill=green))
```



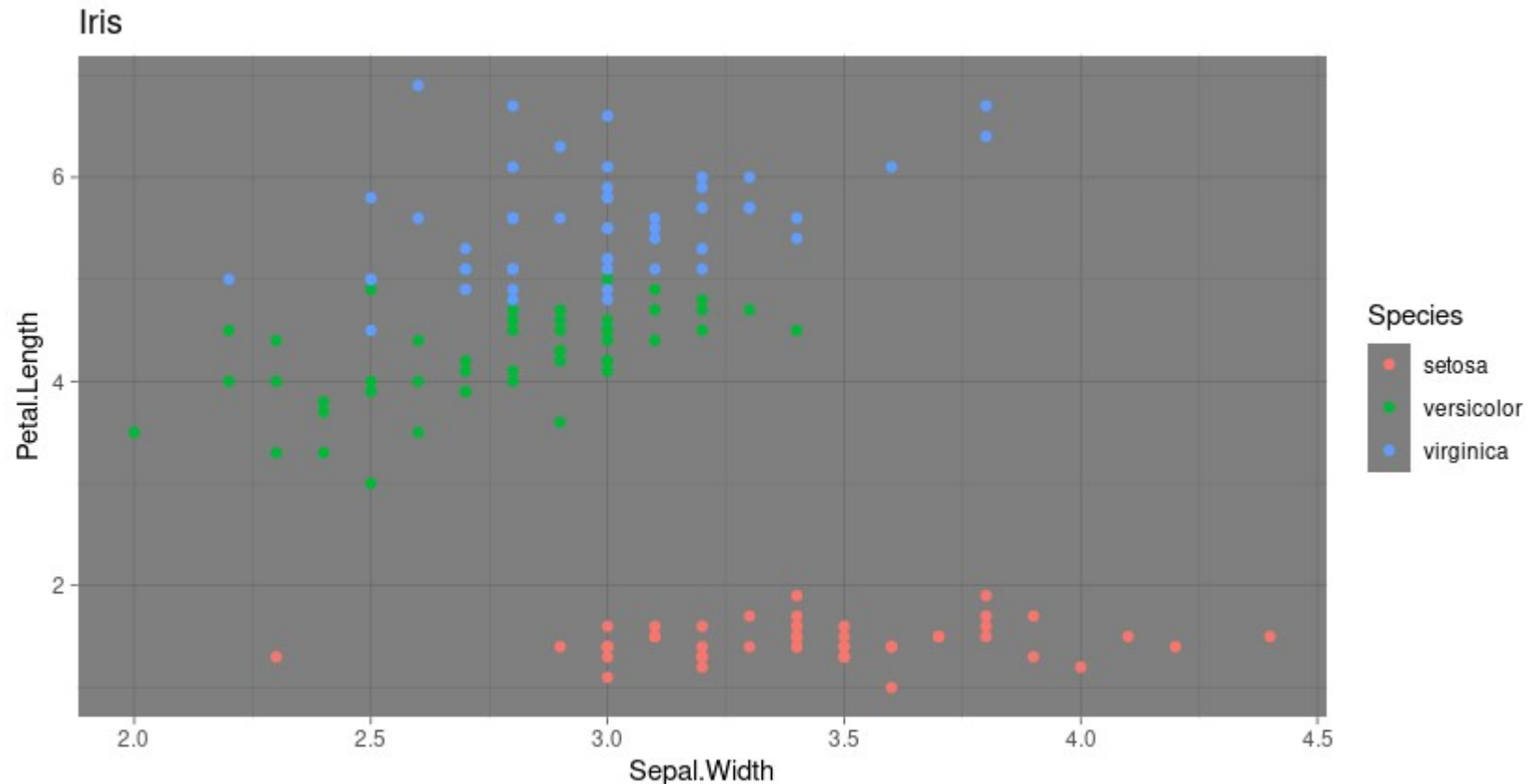
Temas prediseñados

Existen ya algunos temas prediseñados que se pueden consultar aquí:

<https://ggplot2.tidyverse.org/reference/ggtheme.html>

Algunos ejemplos son `theme_grey()` o `theme_dark()`

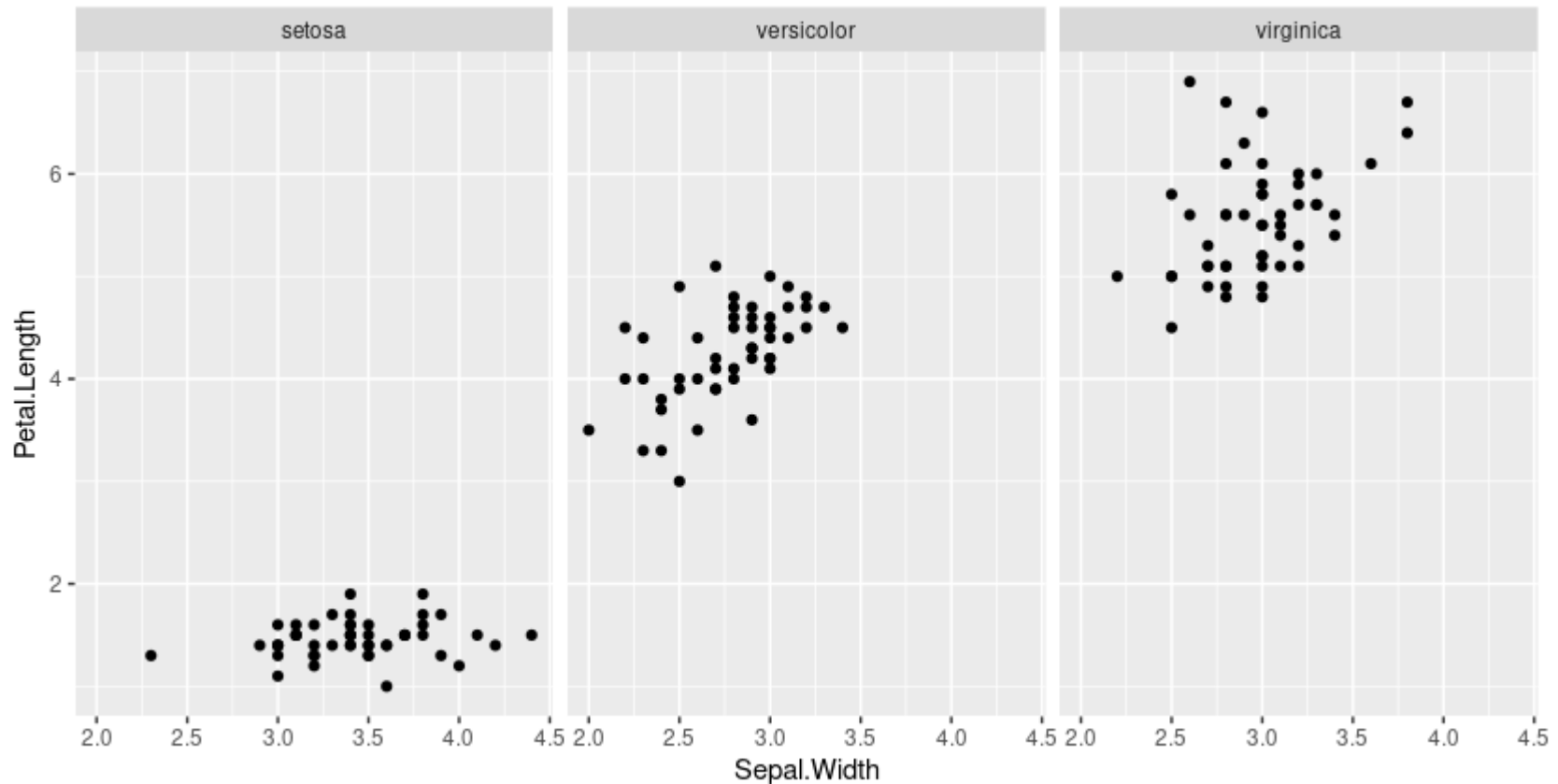
```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length, color=Species)) + geom_point()  
+ labs(title="Iris") + theme_dark()
```



Múltiples gráficos

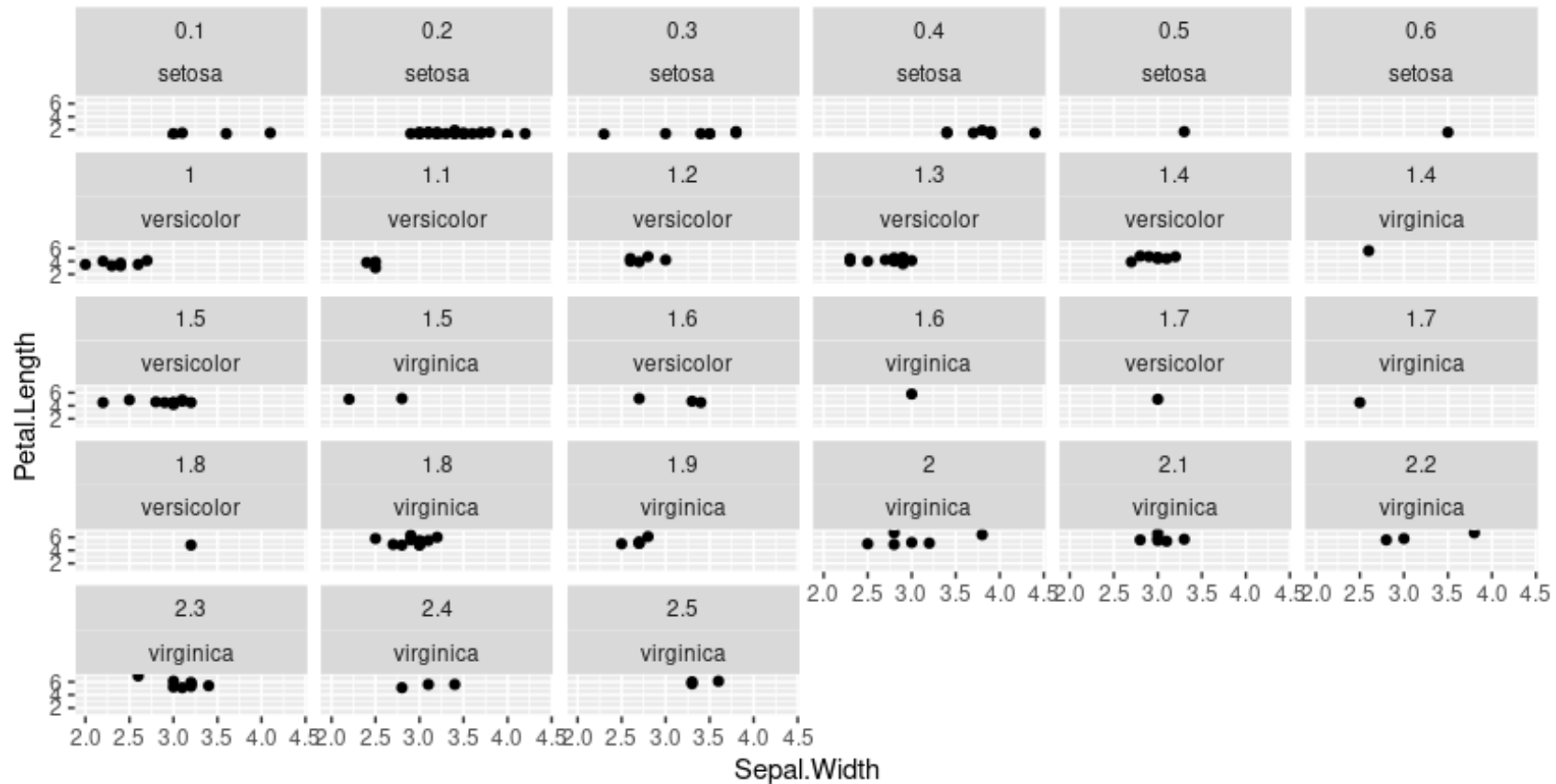
Con los comandos `facet_grid()` y `facet_wrap()` se pueden mostrar diversos gráficos en la misma pantalla. Grid los muestra en cuadrantes mientras que Wrap los muestra uno al lado del otro. Se puede dividir el gráfico de acuerdo a una o a dos variables

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length)) + geom_point() + facet_grid(. ~ Species)
```



Múltiples gráficos

```
ggplot(data=iris,aes(x=Sepal.Width, y=Petal.Length)) + geom_point() + facet_wrap(Petal.Width ~ Species)
```



<http://www.sthda.com/english/wiki/ggplot2-facet-split-a-plot-into-a-matrix-of-panels>

Bioinformática y análisis de datos ómicos

