# Taller de Visualización en R: ggplot 2 y otros

Andrés Blanco y Santiago Costas

# Instalando y Cargando las librerías

El primer paso siempre que comencemos una sesión de R es cargar las librerías que utilizaremos. En caso que no las tengan instaladas, primero hay que hacerlo.

```
## Correr solo si no tienen instaladas las librerías
install.packages(c('tidyverse', 'sciplot', 'patchwork', 'viridis', 'RColorBrewer'))
## en caso de que quieran correr solo alquno de ellos, el codigo es iqual
## solo con el paquete de interes por ejemplo, install.packages('tidyverse')
library(tidyverse); library(sciplot); library(patchwork); library(viridis); library(RColorBrewer)
## -- Attaching packages ------ tidyverse 1.3.1 --
## v ggplot2 3.3.6
                     v purrr
                              0.3.4
## v tibble 3.1.8
                     v dplyr
                              1.0.9
## v tidyr
           1.2.0
                     v stringr 1.4.0
## v readr
           2.1.2
                   v forcats 0.5.1
## -- Conflicts ----- tidyverse conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()
                   masks stats::lag()
## Loading required package: viridisLite
```

# Cargando la base de datos

En este caso usamos una función del paquete tidyverse <code>read\_csv2</code> la cual le agrega otros atributos a la tabla. Utilizando el comando <code>head</code> podemos ver las primaras filas y las columnas (variables) con su tipo de objeto (que es muy importante saber las diferencias pero que hoy no lo analizaremos en este taller). La funciones de R base más usadas son <code>read.csv0</code> <code>read.csv2</code>

```
data <- read_csv2("producción soja.csv")

## i Using "','" as decimal and "'.'" as grouping mark. Use `read_delim()` for more control.

## Rows: 112 Columns: 5

## -- Column specification -------

## Delimiter: ";"

## chr (1): Area

## dbl (4): Year, sup, Yield, Production

##

## i Use `spec()` to retrieve the full column specification for this data.

## i Specify the column types or set `show_col_types = FALSE` to quiet this message.</pre>
```

Las unidades están en metros cuadrados, si las queremos en hectáreas:

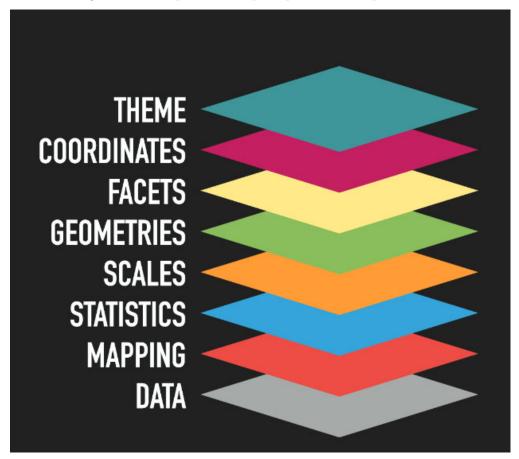
```
data$Yield <- data$Yield/10000
```

#### head(data)

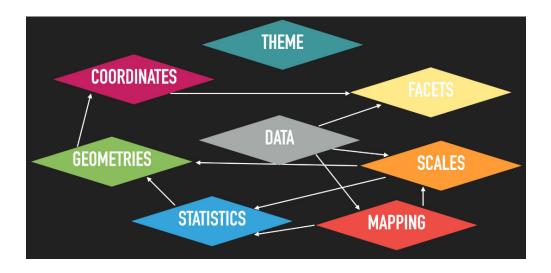
```
## # A tibble: 6 x 5
                       sup Yield Production
     Area
                Year
##
     <chr>>
               <dbl> <dbl> <dbl>
                                      <dbl>
## 1 Argentina 1961
                       980 0.976
                                        957
## 2 Argentina 1962 9649 1.16
                                      11220
## 3 Argentina 1963 19302 0.980
                                      18920
## 4 Argentina 1964 12220 1.15
                                      14000
## 5 Argentina 1965 16422 1.04
                                      17000
## 6 Argentina 1966 15689 1.16
                                      18200
```

# 1. Primeros Gráficos

Con los datos cargados hacemos los primeros gráficos con estética estándar. Es importante entender que ggplot2 funciona siguiendo la teoría de *grammar of graphics*, cuyo objetivo es **generalizar** la forma de crearlos. Los gráficos se componen de 'capas' que se van solapando.

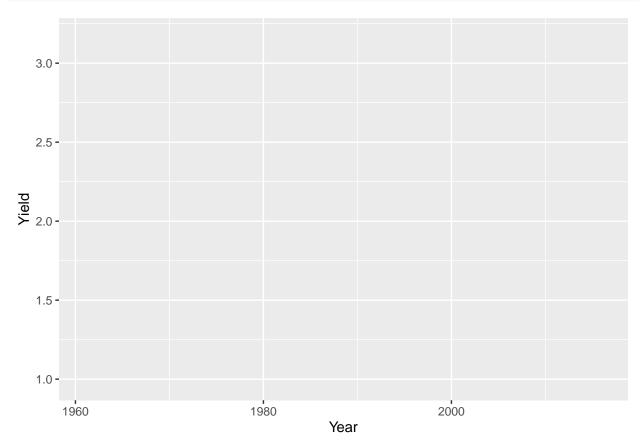


Sacada del taller de Thomas Lin Pedersen ggplot2 workshop - Pedersen



# 1.1. Datos y Mapping

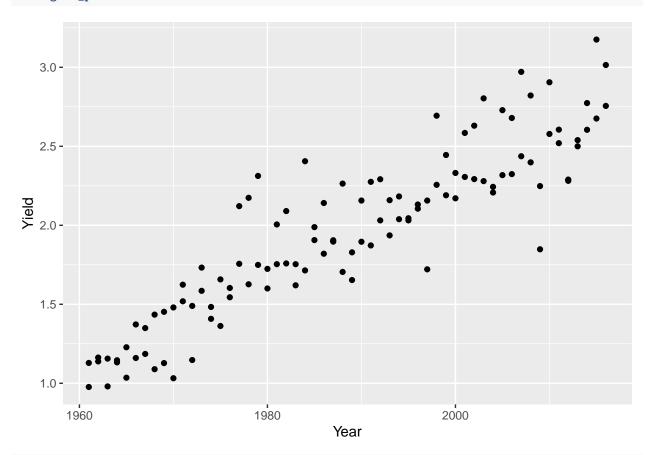
En este punto le indicamos qué datos vamos a usar y el mapping, que conecta tus datos con su significado. El software no sabe por defecto la naturaleza de tus datos, por lo que es imprescindible indicarle cuál es la x cuál es la y o demás variables/factores que pueden estar involucrados. Es decir conecta los datos y las variables con las propiedades gráficas. Si solo especificamos esto, el resultado es la nada misma.



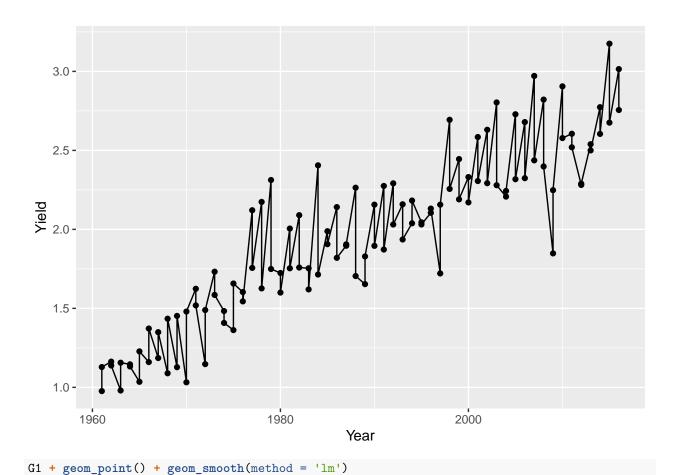
# 1.2. Geom

Los geom toman los valores escalados que vienen del mapping y los interpretan de una forma particular. Los mismos valores pueden pueden tener una gran lista diferente de posibles representaciones geométricas (puntos, líneas, barras, cajas) y los geom son como especificas estas formas. Acá es donde se destaca más el funcionamiento de 'capas' ya que estas son las verdaderas capas gráficas junto con otras que veremos mas adelante.

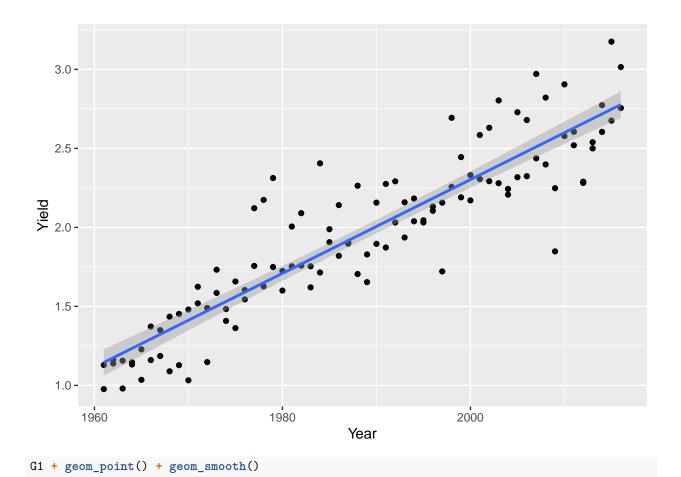
## G1 + geom\_point()



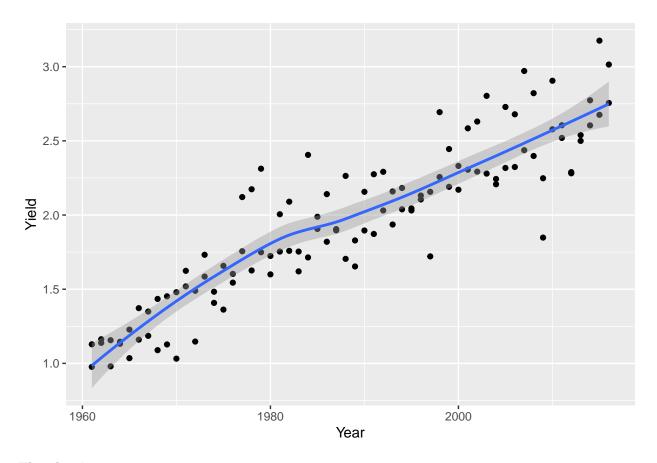
G1 + geom\_point() + geom\_line()



## `geom\_smooth()` using formula 'y ~ x'

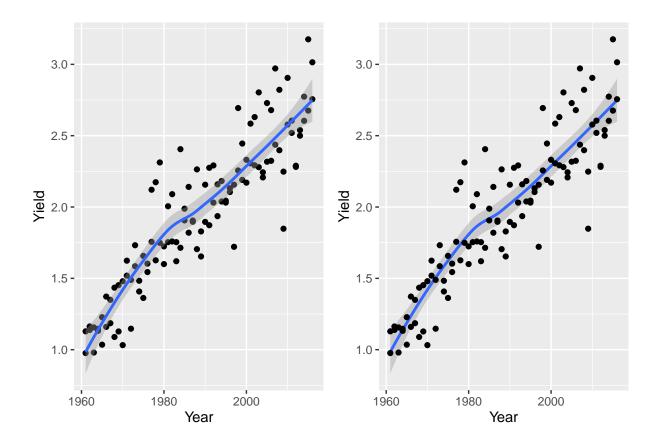


##  $geom_smooth()$  using method = 'loess' and formula 'y ~ x'



# El orden importa

```
G1 + geom_point() + geom_smooth() + G1 + geom_smooth() + geom_point()
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



## 1.3. Scales

Transcribe los rangos de las variables a los rangos de la propiedades gráficas. Por ejemplo, si queremos clasificar por un factor y que esto sea representado por el color, la *escala* le dice que el valor "argentina" por ejemplo va a ser "azul" y mundo será "rojo". También aplica en variables continuas , por ejemplo en un scatterplot conecta los números con la posición. En los pasos anteriores se definieron escalas por defecto al usar mapping y geom. Ahora vamos a especificarla, esto modificará la escala gráfica a mostrar en los ejes.

#### Escalas continuas

```
G1 <- G1 + geom_point() ## Simplemente guardamos el objeto para lo que viene

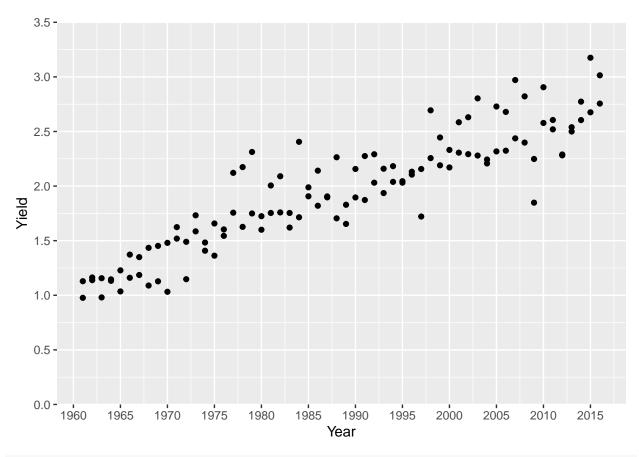
G1 +

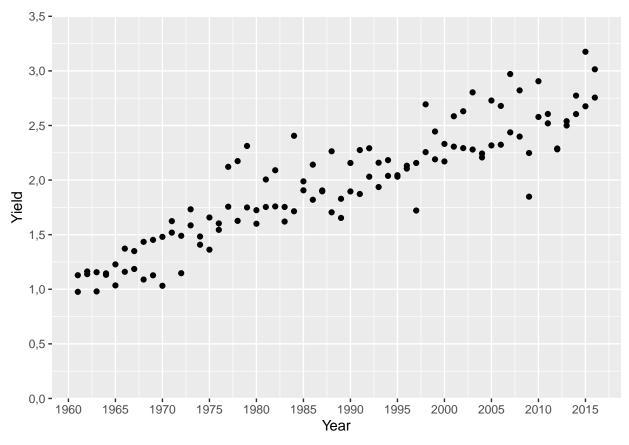
scale_y_continuous(expand = c(0,0),

limits = c(0, 3.5),

breaks = seq(0, 3.5, by = .5))+

scale_x_continuous(breaks = seq(1960,2016, by=5))
```

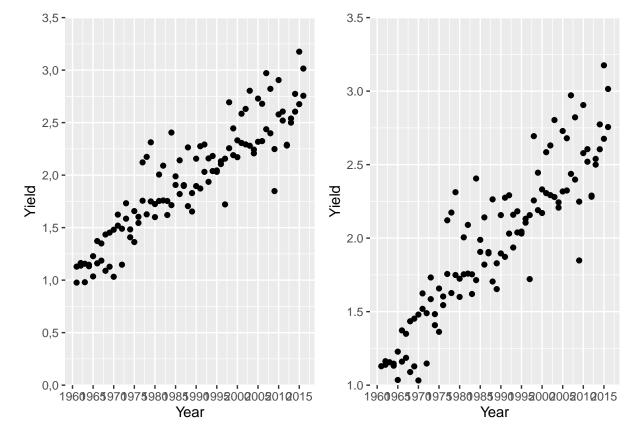


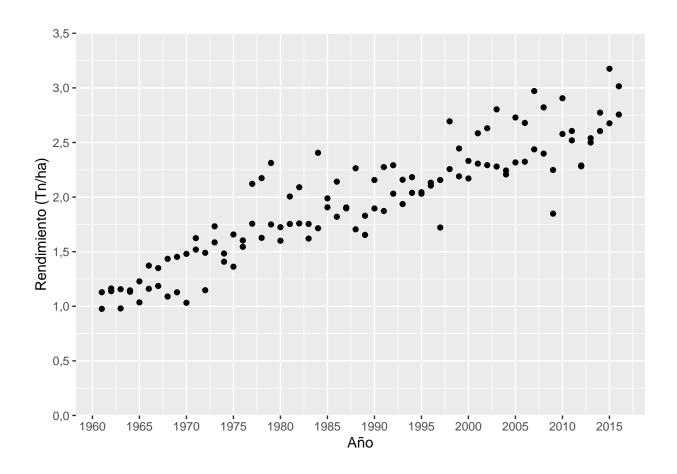


## Scale for 'y' is already present. Adding another scale for 'y', which will ## replace the existing scale.

## Scale for 'x' is already present. Adding another scale for 'x', which will ## replace the existing scale.

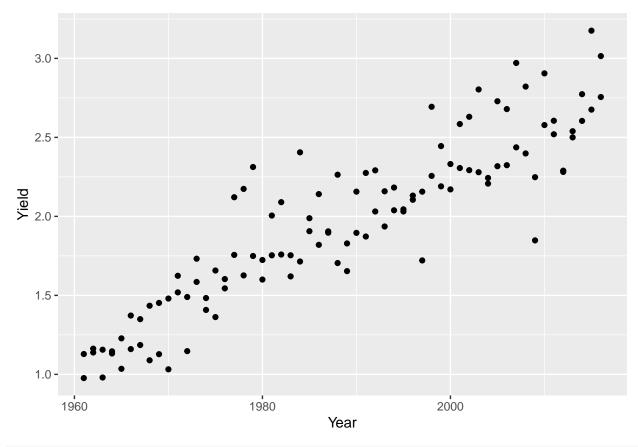
## Warning: Removed 2 rows containing missing values (geom\_point).

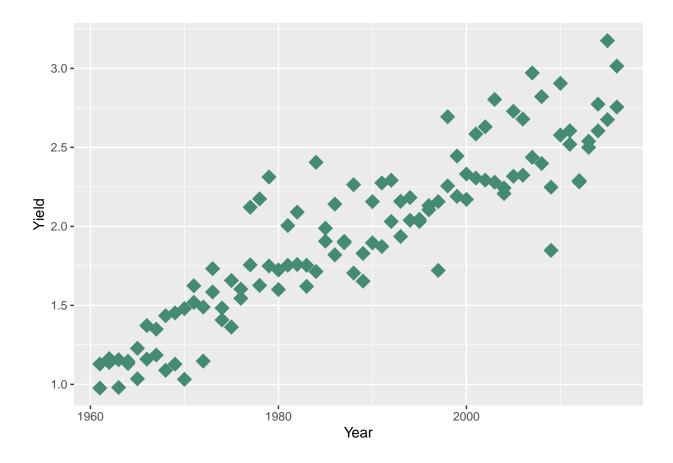




# Modificando las escalas de color, forma y tamaño los geoms

Las variables de clasificación siguen generando colores, formas y tamaños por defecto.





# 2. Gráficos con 2 variables explicativas

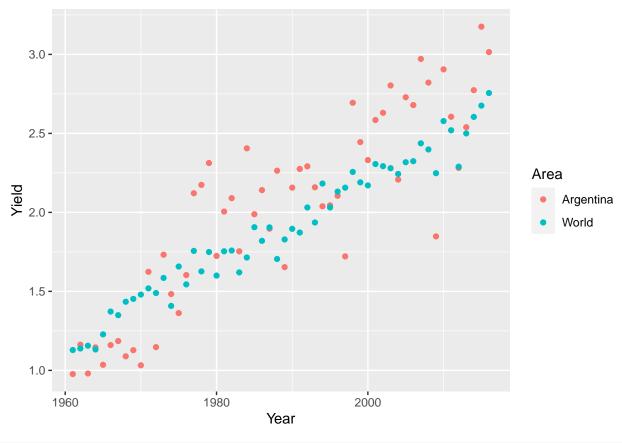
Ahora vamos a empezar a jugar con las diferentes posibilidades utilizando una variable de clasificación y modificando propiedades gráficas que antes dejamos por defecto.

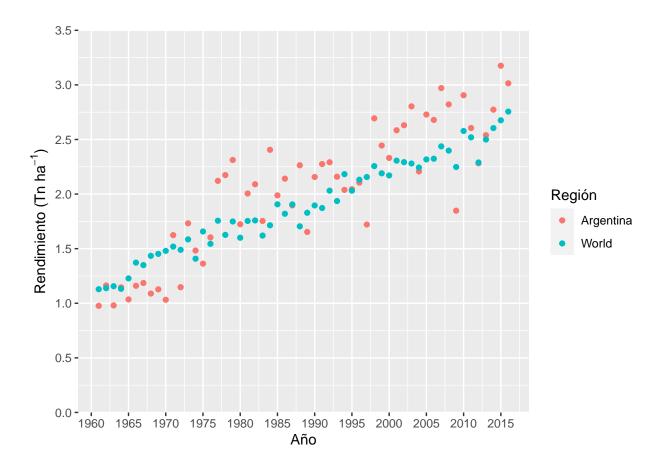
## 2.1. Clasificando por factores

En esta sección incluiremos una variable más en el mapping la cual va funcionar como factor de clasificación. Todo lo que incluyamos como clasificación en el mapping va a generar una leyenda.

#### Color

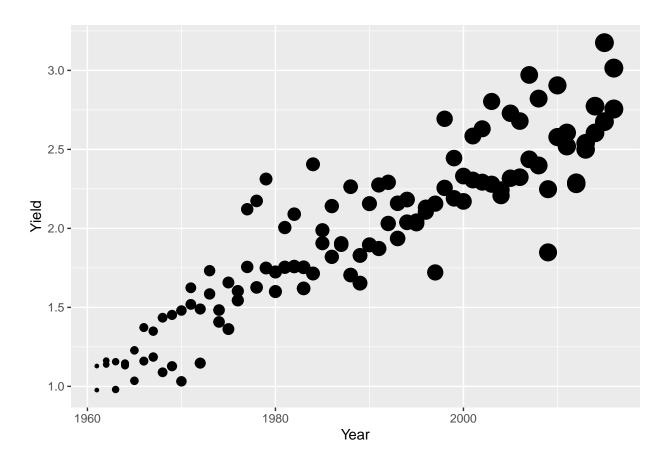
Especificaremos que el color va a estar determinado de acuerdo con el el  $\'{a}rea$ , aunque todavía la escala de estos colores será por defecto.



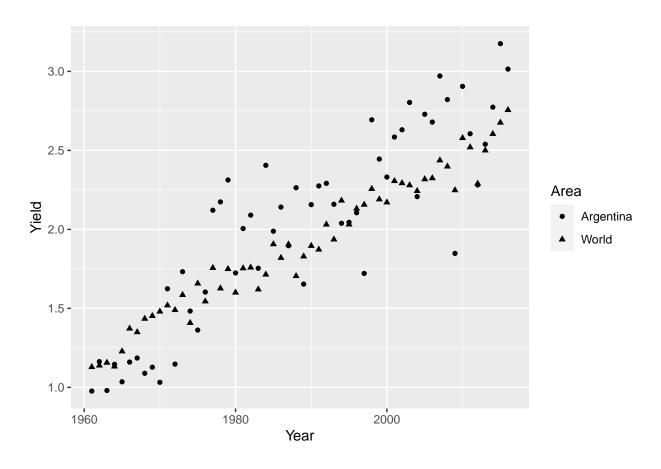


## Tamaño

Ahora haremos lo mismo, pero en vez del color, modificaremos el tamaño de acuerdo con el año. Como nos interesa la tendencia y no el valor exacto, eliminamos la leyenda.

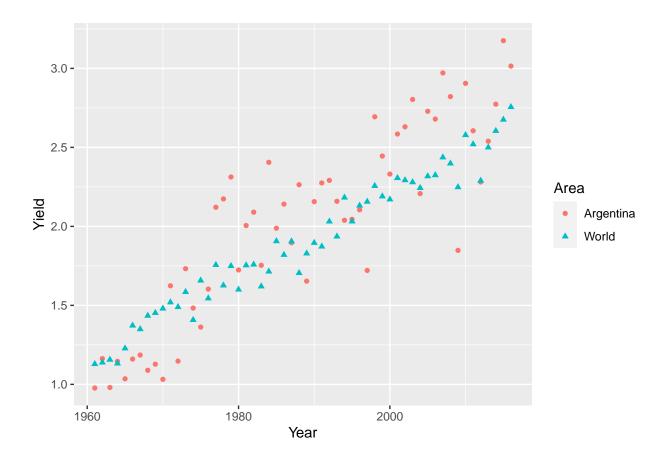


# Forma



# Combinando

Se pueden combinar más de un parámetro de clasificación. Si el factor es el mismo, la leyenda se unificará.

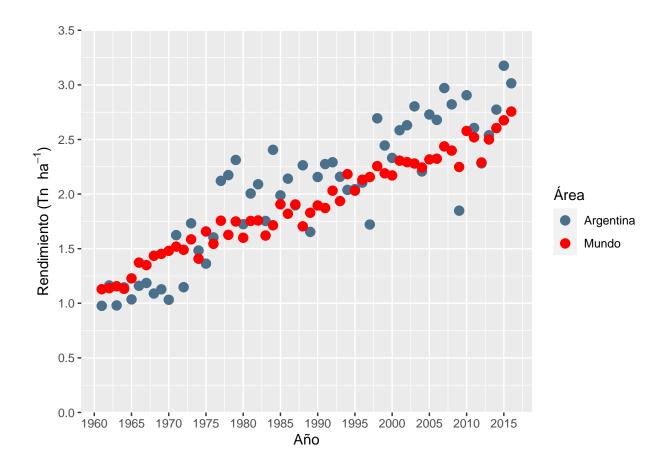


### 2.2. Modificando las escalas correspondientes a las variables de clasificación

Ahora vamos a modificar las escalas continuas (como hicimos antes), pero también escalas discretas. Recordemos que si usamos variables dentro del aes, estas serán de clasificación y generarán su propia leyenda (pueden ser variables continuas como año).

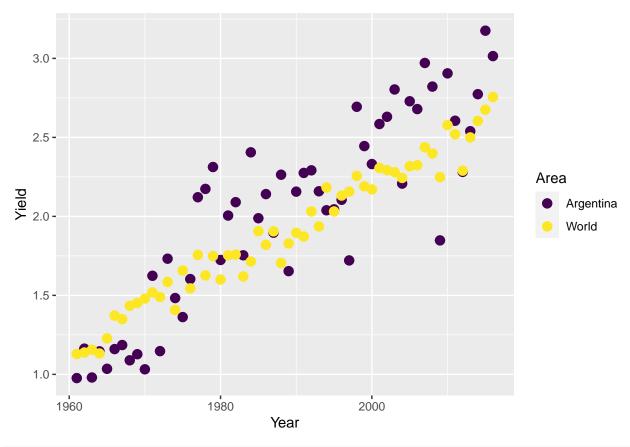
#### Ajuste manual

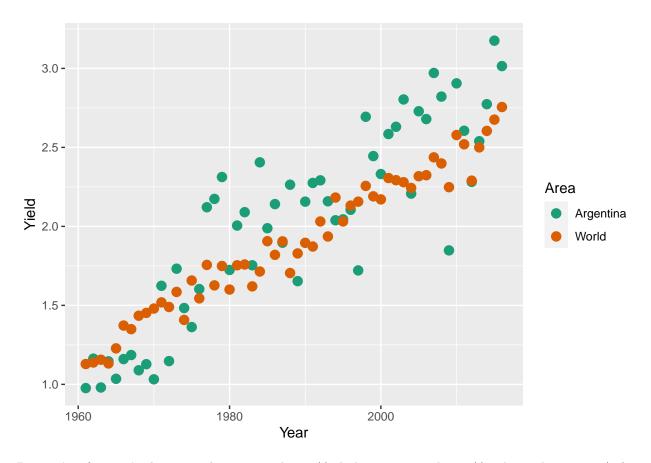
Usando scale\_color\_manual vamos a modificar manualmente la escala correspondiente a la variable de clasificación que determina el color. Este es un ejemplo para color, pero también podría ser shape, fill, etc (con scale\_shape\_manual , scale\_fill\_manual , etc). Este ajuste manual siempre estará asociado a que en el mapping haya una variable que determine el color, por lo que deberemos asignar tantos colores como niveles de la variable participen.



# Utilizando paletas de color de forma automática

También se puede modificar los colores con paletas predeterminadas. Existen paletas desarrolladas con diferentes objetivos, por ejemplo que puedan ser vistas por daltónicxs. En este caso usamos **Viridis** que es para daltónicxs y **RcolorBrewer** 

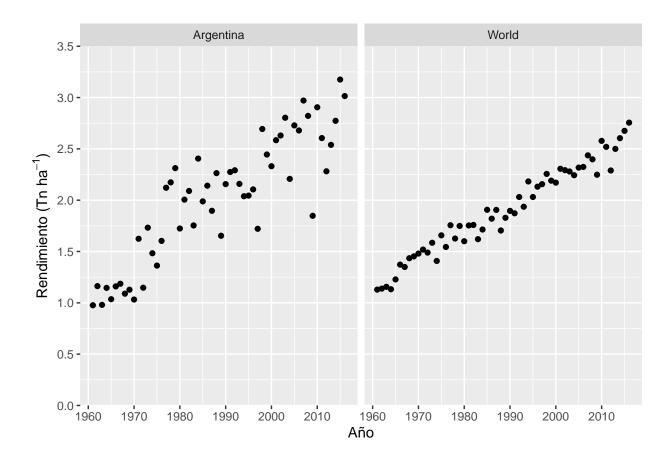




Para más información de estas paletas visitar <br/> https://colorbrewer2.org y https://ggplot2.tidyverse.org/refere <br/> nce/scale\_viridis.html

#### 2.3. Facets

Esta es una capa opcional que funciona haciendo un subset de los datos de acuerdo a una variable de clasificación. Con esto realizará varios gráficos uno por cada nivel de la variable, dividiendo la ventana gráfica y compartiendo la misma escala entre ellos.



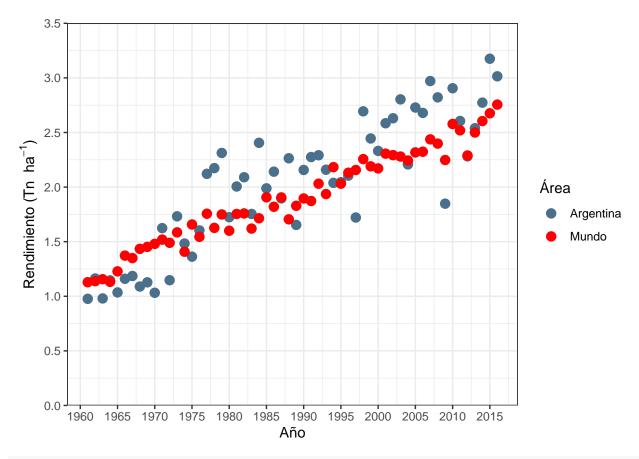
## 2.4. Tema o Theme

Define la estética general del gráfico, son todas las partes del gráfico que no tienen que ver con los datos. El color del background, los ejes, lineas, etc.

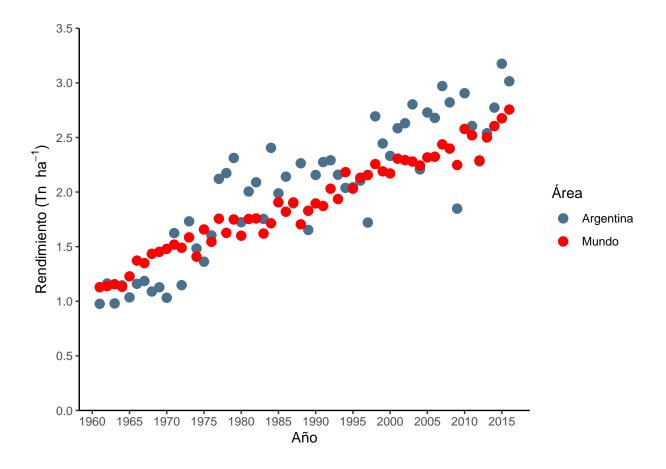
#### Temas predeterminados

Existen algunos **themes** predeterminados que determinan la estética del gráfico. Un ejemplo es el que viene por defecto que venimos utilizando theme\_grey(), pero existen otros como:theme\_bw() theme\_light() theme\_dark theme\_void, etc.

G3 + theme\_bw()

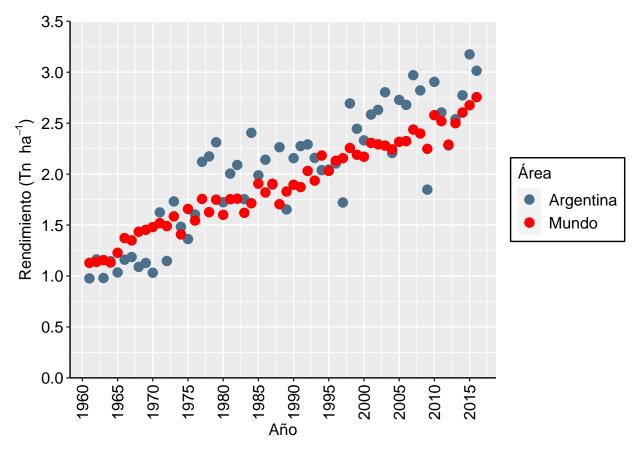


G3 + theme\_classic()



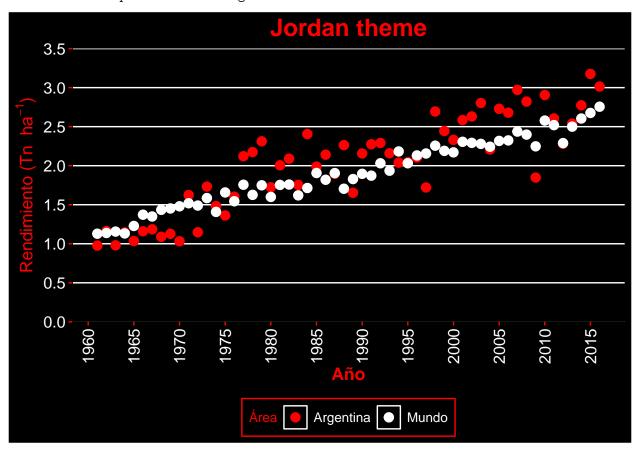
#### Edición manual del tema

Acá es importante saber que se puede modificar cada ítem gráfico que queramos, desde la posición de la leyenda con sus títulos, hasta el color de las lineas del fondo. Lo que se les ocurra. Un consejo, una vez que obtengan una estética estándar propia que les cierre, guárdenlo en una sesión de R, luego simplemente pegan el objeto cada vez que hagan un gráfico, y se olvidan de la estética. Para nosotros es uno de los mejores usos de ggplot2.



```
## un ejemplo freaky para quienes nos gusta el basquet
MICHAEL <-
            theme(
    legend.position = "bottom", legend.title = element_text(colour = "red", size = 10),
    legend.background = element_rect(fill = "black", colour = 'red'),
    legend.key = element_rect(fill = "black", colour = "white"),
    legend.text = element_text(colour = "white", size = 10),
    plot.background = element_rect(fill = "black", colour = "white"),
    panel.background = element_rect(fill = "black"),
  # panel.background = element_rect(fill = "white"),
    axis.text = element_text(colour = "white"),
    plot.title = element_text(colour = "red", face = "bold", size = 18, vjust = 1,
                              hjust = 0.5),
    axis.title = element_text(colour = "red", face = "bold", size = 13),
    panel.grid.major.y = element_line(colour = "white"),
    panel.grid.minor.y = element_blank(),
    panel.grid.major.x = element_blank(),
    panel.grid.minor.x = element_blank(),
    strip.text = element_text(colour = "white"),
    strip.background = element_rect(fill = "white"),
    axis.ticks = element_line(colour = "red")
  )
G3 + scale_color_manual(values = c('red','white'), labels = c('Argentina', 'Mundo')) + ggtitle('Jordan
 MICHAEL
```

## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.

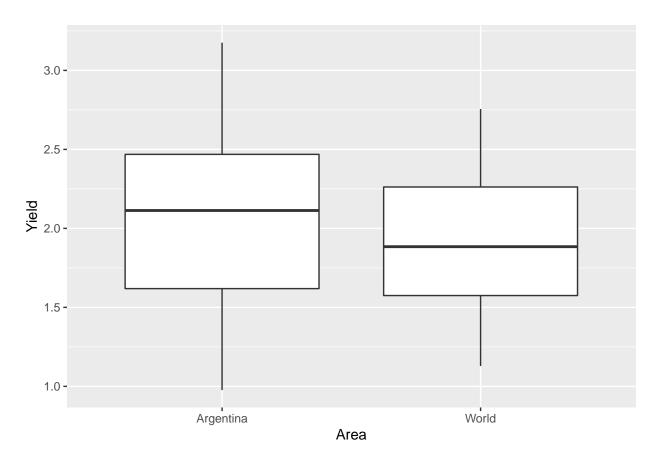


# 3. Otros gráficos con los mismos datos

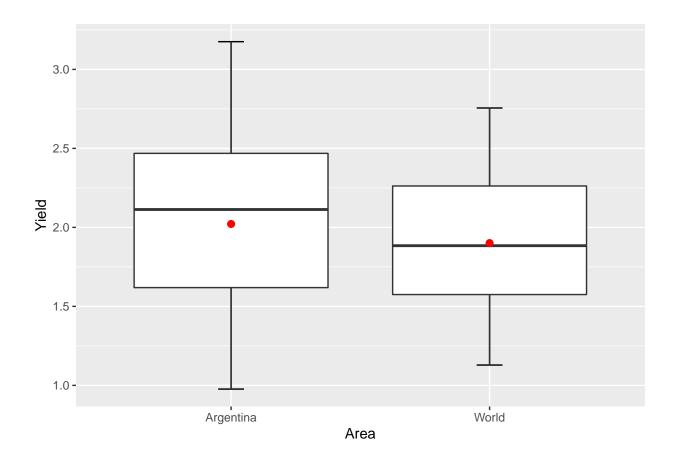
Vamos a explorar la idea de la generalización de la forma de hacer gráficos. Con exactamente los mismos datos, las representaciones pueden ser diferentes. Esto lo hacemos con los <code>geom\_</code>

## **Boxplot**

Los geom como el boxplot incluyen una capa de *statistics* que está implícita. Muchas veces los datos crudos no representan lo que queremos graficar. En estos casos, la capa de *stats* transforma los datos de cierta forma que nos interese (usando estadística). En este caso calcula los cuartiles para dibujar las cajas.

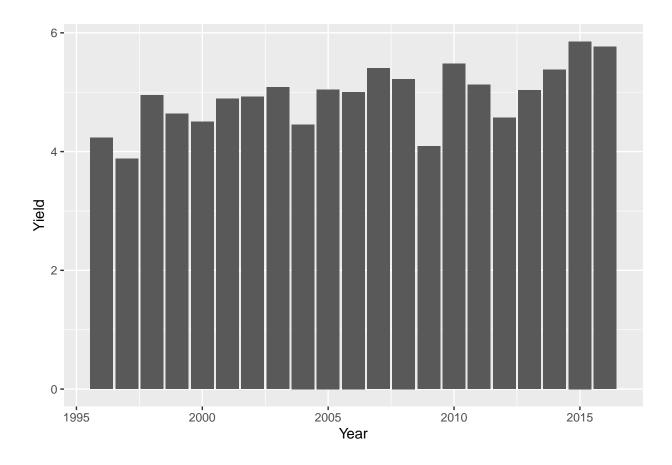


Podemos agregar el extremo y la media (el orden importa)



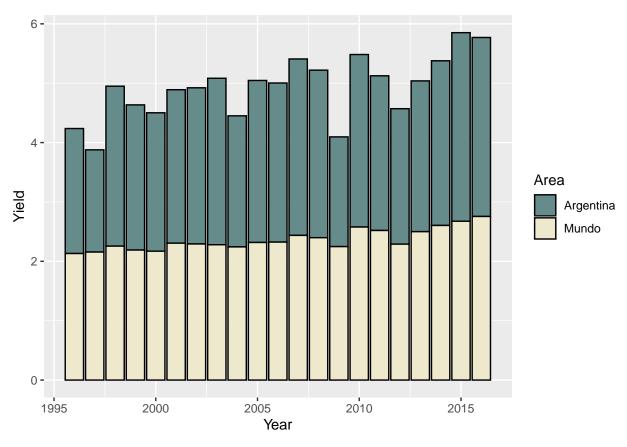
# Columnas

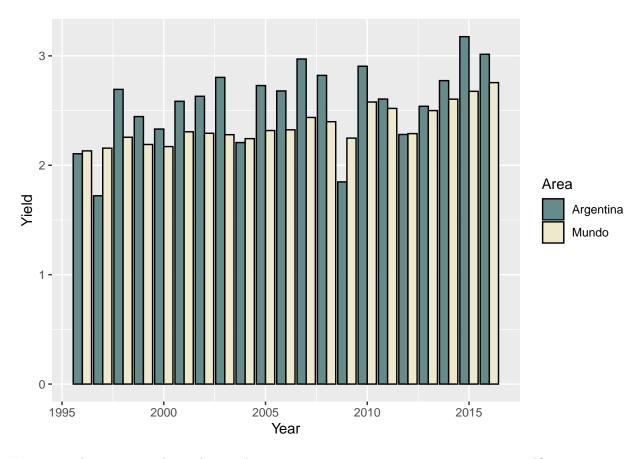
Otro ejemplo de representación, esta vez con geom\_col



## Clasificando por variables

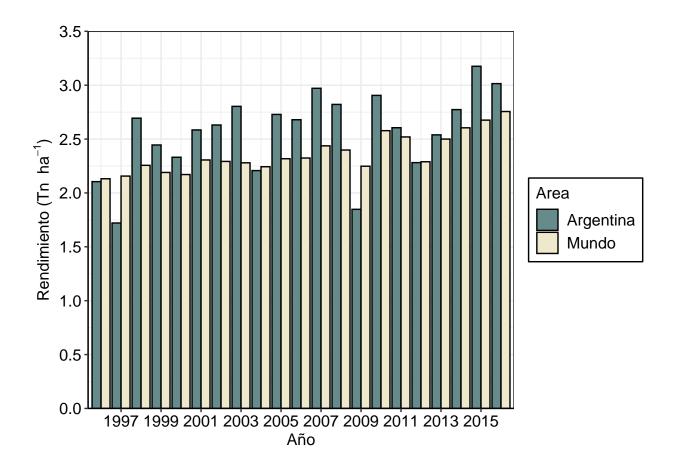
Cuando clasificamos por color, por defecto las barras se muestran apiladas, si las queremos al lado debemos cambiar el position





Incorporando cuestiones de escala y estéticas vistas anteriormente, mejoramos nuestro gráfico:

```
G5 <- G5 +
scale_y_continuous(expand = c(0,0), limits = c(0, 3.5), breaks = seq(0, 5, by = .5))+
scale_x_continuous(expand = c(0,0.2), breaks = seq(1995,2016, by=2)) +
labs(x= "Año",
        y = expression(paste("Rendimiento (Tn"," ", ha^-1, ")")))+
theme_bw()+
theme(text = element_text(size=12, colour = "black"),
        axis.text = element_text(size=12, colour = "black"),
        legend.text = element_text(size=12, colour = "black"),
        legend.background=element_rect(colour="black"),
        axis.line = element_line(colour = "black")); G5</pre>
```



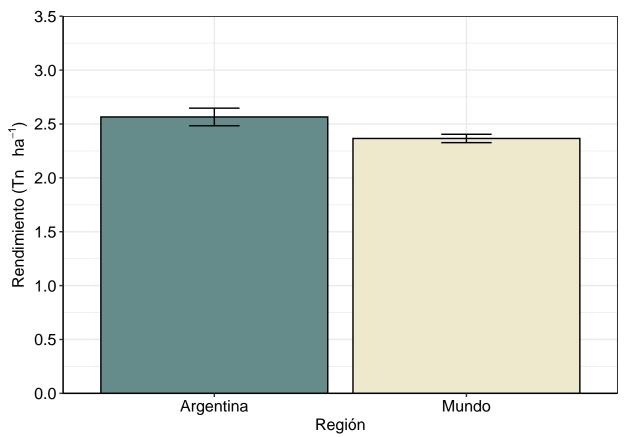
# 4. Un poquito más de ggplot2 y un poquito más allá dentro de tidyverse

Ahora vamos a hacer un poquito de manejo de tables con dplyr (tidyverse) y aplicar un poco lo que hemos visto de ggplot2

#### 4.1. Transformando los datos para graficar medias y desvíos

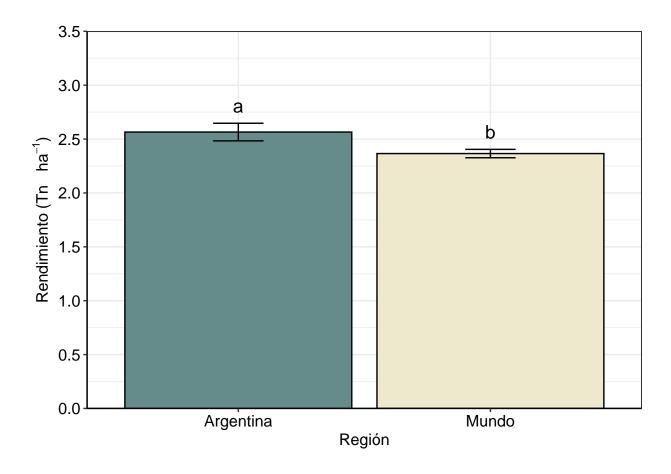
Esta función probablemente se pueda hacer dentro de una capa con los stats que hemos visto, pero la forma que encontramos es utilizando **dplyr** manejando la tabla original. Al usar pipes %>%y no crear un nuevo objeto, nos permite graficarlo directamente. Tener en cuenta que la función se pertenece al paquete sciplot.

```
legend.text = element_text(size=12, colour = "black"),
legend.background=element_rect(colour="black"),
axis.line = element_line(colour = "black")); G6
```



# 4.2. Agregando texto

Acá no utilizamos el mapping global que definimos antes, sino que usamos "otros" datos. Lo haremos con geom\_text



# 4.3 Combinando representaciones (geoms)

Una buena forma de visualizar algunas variables es superponiendo diferentes tipos de gráficos. Para ello vamos a integrar algunas cosas que venimos trabajando.

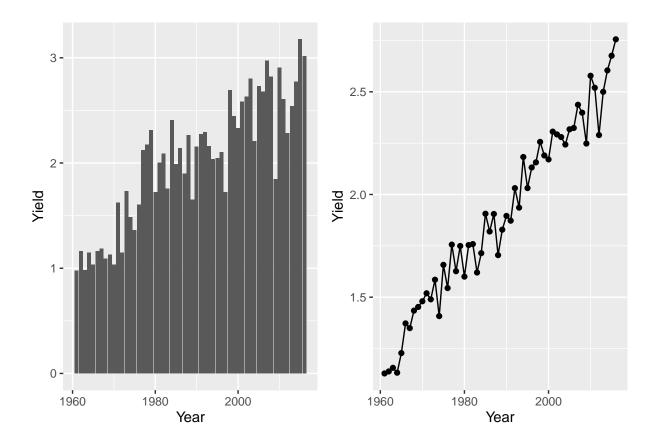
Supongamos que queremos comparar el rendimiento en Argentina y el mundo por años en los últimos 20 años. Para esto ya habíamos generado el gráfico ¿G5?, veamos otra opción.

Generamos una base de datos para cada región usando dplyr

```
data_argentina <- data %>% filter(data$Area == "Argentina")
data_mundo <- data %>% filter(data$Area != "Argentina")
```

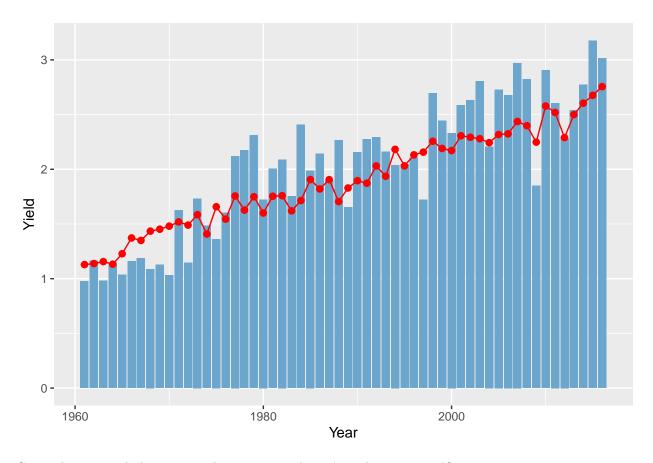
Luego vamos a superponer un gráfico base de columnas de Argentina y uno de puntos y de lineas del mundo

```
G7 <- ggplot(data_argentina, aes(Year, Yield)) +
  geom_col()
G8 <- ggplot(data_mundo, aes(Year, Yield)) +
  geom_line() +
  geom_point()
G7 + G8</pre>
```

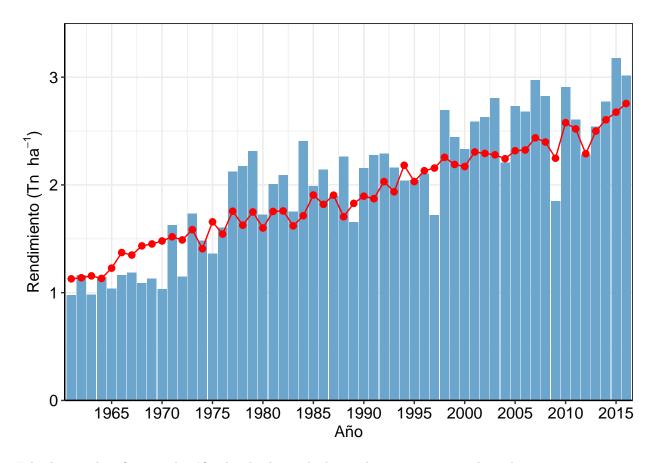


 $\xi$ Cómo los combinamos? Cambiando los **datos** y el **mapping** dentro de los geoms, es decir no utilizando el global para uno de ellos. Veamos

```
G7 <- ggplot(data = data_argentina, mapping = aes(Year, Yield)) + ## funcion global
geom_col(fill="skyblue3") + ## para el geom_col utilizamos los datos globales
geom_line(data = data_mundo, aes(Year, Yield), col= "red") + ## cambiamos los datos
geom_point(data = data_mundo, aes(Year, Yield), col= "red", size=2); G7 ## igual que antes
```

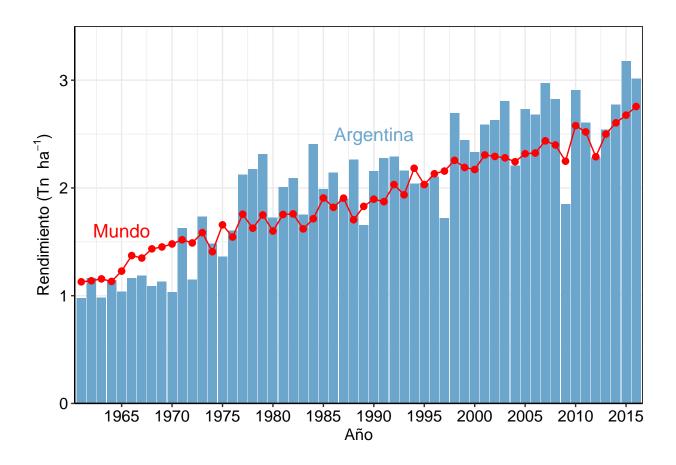


Copiando y pegando lo previo podemos mejorar la estética de nuestro gráfico



Falta lograr identificar en el gráfico las dos bases de datos. Agregamos texto aclarando

```
G7 <- G7 +
   annotate("text", x=1990, y= 2.5, label = "Argentina", col="skyblue3", size=5)+
   annotate("text", x=1965, y= 1.6, label = "Mundo", col="red", size=5); G7</pre>
```

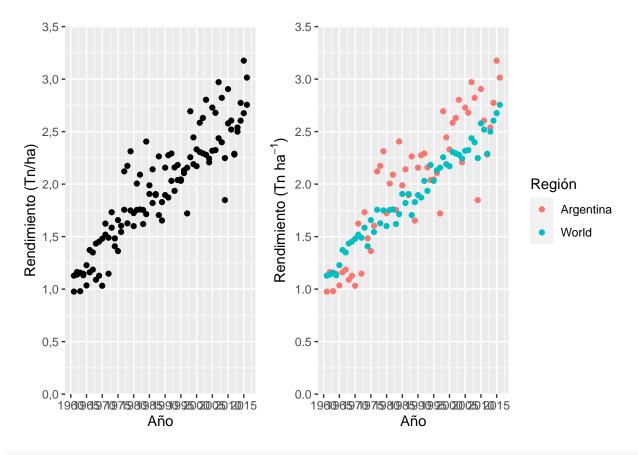


# 5. Dividiendo la pantalla gráfica para mostrar más de un gráfico

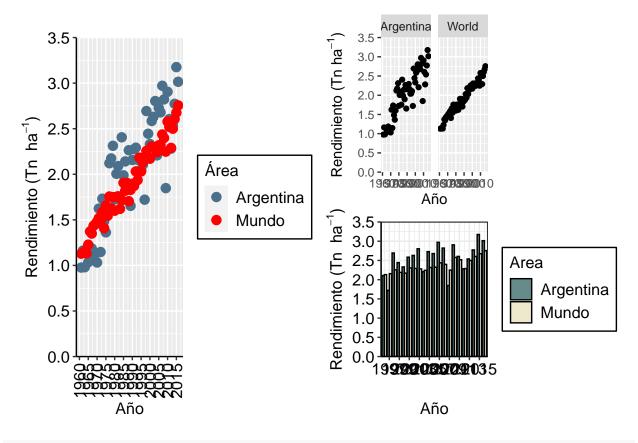
Podemos combinar diferentes gráficos en uno gracias al paquete  $\mathbf{patchwork}$  que ya hemos utilizado pero ahora exploraremos su potencial y sus funciones.

# 5.1. Algunos ejemplos

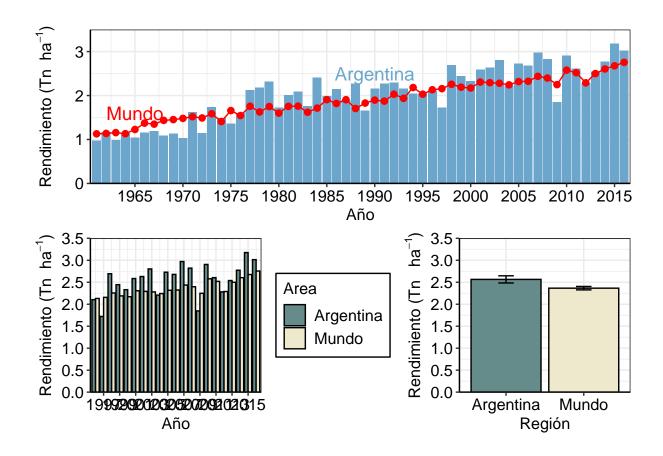
G1 + G2 ## en dos columnas



G3 + G4 / G5 # grafico G3 en 1 columna y 2 filas, G4 y G5 en 1 fila y 1 columna

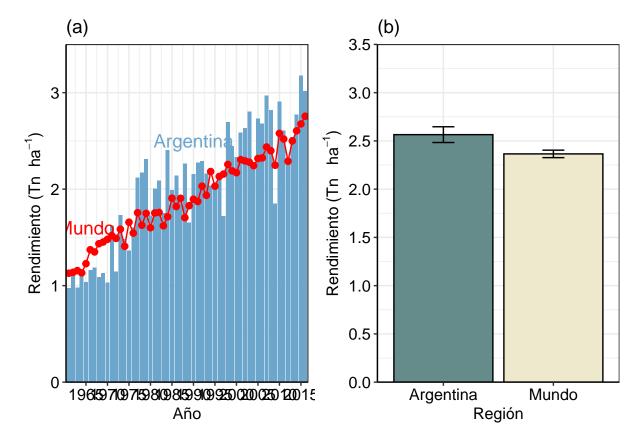


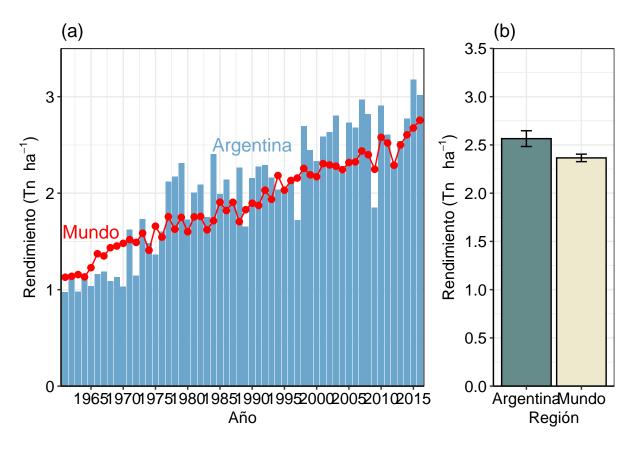
 $\mathsf{G7}$  /  $(\mathsf{G5} + \mathsf{G6})$  #  $\mathsf{G7}$  en 1 fila y 2 columnas,  $\mathsf{G5}$  y  $\mathsf{G6}$  en 1 fila y 1 columna

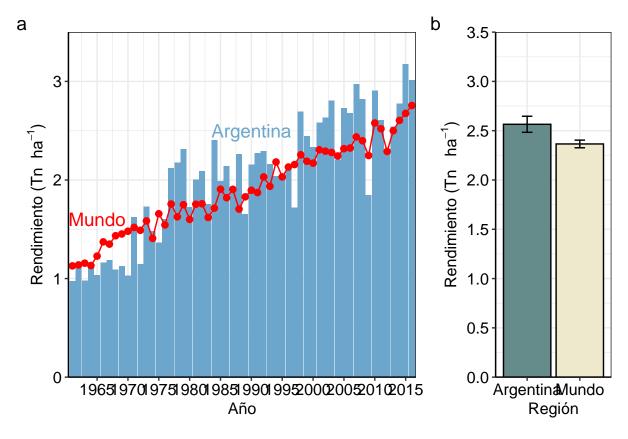


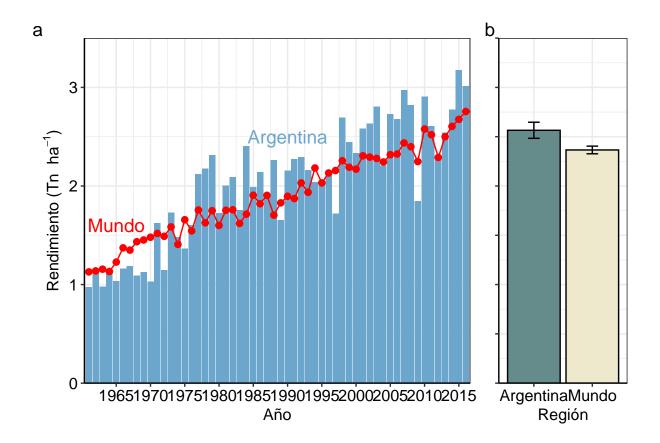
## 5.2. Dividiendo la pantalla gráfica y agregando referencia a los gráficos

Esto lo haremos utiliizando la función wrap\_plots también del paquete patchwork. La ventana gráfica funciona como una combinación de filas y columnas que marcan la posición. En la representación de un solo gráfico hay 1 sola columna y 1 una sola fila que ocupa toda la ventana. En este caso vamos a usar dos gráficos distribuidos en dos columnas y una sola fila (uno al lado del otro). Si quisiéramos uno arriba de otro sería en 2 filas y una sola columna.









# 7. Exportando los gráficos

Los gráficos pueden ser exportados en varios formatos. Exploraremos algunos de ellos. Con cualquiera de las funciones que vamos a ver el archivo se escribirá en el misma carpeta que el directorio de trabajo (working directory), a no ser que le agreguemos ruta al archivo.

## 7.1. Utilizando ggplot2

Para exportar en diferentes formatos es siempre la misma función ggsave, lo que va a determinar el tipo de archivo es la extensión .pdf, .png, .jpg, etc. (la función no soporta svg). También se puede controlar el tamaño y los dpi, entre otras cosas. La función guarda el último gráfico corrido, por defecto.

```
ggsave("Gráfico 8.png") # exporta tal cual se ve en la ventana Plots
ggsave("Gráfico 8.pdf", width = 10, height = 8, dpi= 600) # esta es el que usaremos en inkscape
```

#### 7.2. Utilizando R base

Aquí la función determina el formato (además del nombre), también se puede controlar tamaño (y la unidad de medida del mismo), pixeles, etc. Las funciones llevan el nombre de la extensión por ejemplo: svg(), pdf(), png(), etc. Aquí la diferencia es que es necesario poner el código del gráfico para crearlo (en este caso como ya lo tenemos guardado en un objeto, simplemente se llama al objeto G8) y cerrarlo con dev.off()

```
svg("Gráfico 8.svg", width = 12, height = 10)
G8
dev.off()
```

## pdf

## 2

# 8. Links recomendados para seguir explorando

Este taller pretende ser una mera aproximación al uso de R como motor gráfico mediante el paquete ggplot2. Esperamos despierte el interés en sus participantes para que a partir de aquí exploren el universo de posibilidades que se les ofrece. Para guiar esta búsqueda les recomendamos los siguientes enlaces que poseen ejemplos e información interesante:

#### Cheetsheets de Tidyverse

Son resúmenes con casi todas las funciones de cada paquete y una breve explicación de como funciona.

## Referencia de ggplot2

Explicaciones básicas y ejemplos de todas las capas posibles con ejemplos y combinaciones. Es la página que se usa de referencia cuando no te acordás algo. Recomendamos explorarla es muy interesante.

## Ejemplos de gráficos

La página esta organizada en secciones según objetivos a graficar (distribuciones, correlaciones, mapas, etc), dentro de ellas los diferentes posibles gráficos. Los gráficos estan hechos en ggplot2 pero también con R base. ¡Es buenísima!.

### Workshop ggplot2 parte 1

## Workshop ggplot2 parte 2

Taller de ggplot2 dado por Thomas Lin Pedersen uno de los desarolladores de Rstudio. Muy recomendable, explora cada función de ggplot2. En total es un poco más de 4 horas, no tiene desperdicio.

#### **Patchwork**

Describe los diferentes usos de Patchwork. Es la referencia para usarlo.

# R Graphics for Cookbook

Guía práctica que provee diferentes "recetas" para ayudar a generar gráficos de calidad rápidamente. Utiliza ggplot2 como base.