



Cienfi

centro de
investigación
en economía
y finanzas



Visualizar datos con ggplot2

Por: Maria Fernanda Largo Liévano  y Cristian Álvarez
Diseño por: Laura Barrera Valdés



Este documento contiene información del libro “Empezando a visualizar datos con R y ggplot2” (2022) de Julio César Alonso y Maria Fernanda Largo <https://www.icesi.edu.co/editorial/empezando-visualizar/>

Índice

01	Partes de una visualización	01
02	Primeros pasos con ggplot2	01
03	Geometrías de distribución	02
04	Geometrías de evolución	03
05	Geometrías de relación	04
06	Recomendaciones	05

01 Partes de una visualización

Componentes

Datos

Información a graficar (es una capa obligatoria).

Aesthetics

Se declaran las variables que se mapean en los ejes.

Geometría

Tipo de visualización que se desea.

Facets

Descomponer en subgráficos.

Escalas

Manejar las escalas de los ejes.

Coordenadas

Define el espacio en el que se manejan los ejes.

Tema

Define la apariencia de la gráfica.

Tema

Coordenadas

Escalas

Estadísticas

Facets

Geometría

Aesthetics

Datos



```
ggplot(gapminder, aes(x = gdpPercap, y = lifeExp)) +  
  geom_point() +  
  facet_grid(~continent) +  
  stat_smooth(method = "lm", col = "red") +  
  scale_x_log10() +  
  theme_minimal()
```

02 Primeros pasos con ggplot2

Paso 1

Instalar la librería (si no se tiene) y cargarla.

```
install.packages("ggplot2")  
library(ggplot2)
```

Paso 2

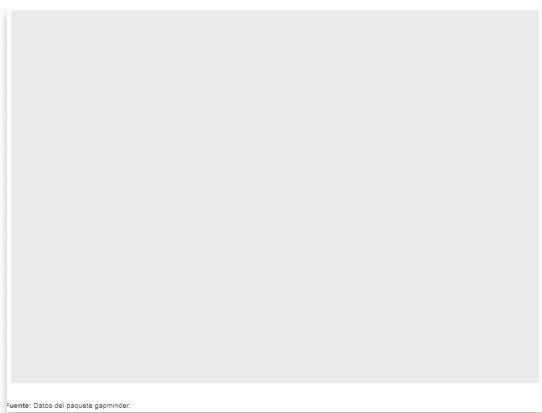
Agregar los datos

```
library(gapminder)  
data("gapminder")
```

Paso 3

Mapear las variables en los ejes del gráfico (aesthetics)

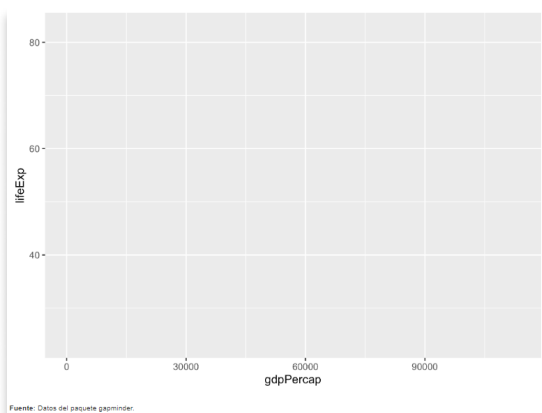
```
ggplot(data=gapminder)
```



Paso 4

Definir el tipo de gráfico (geometría)

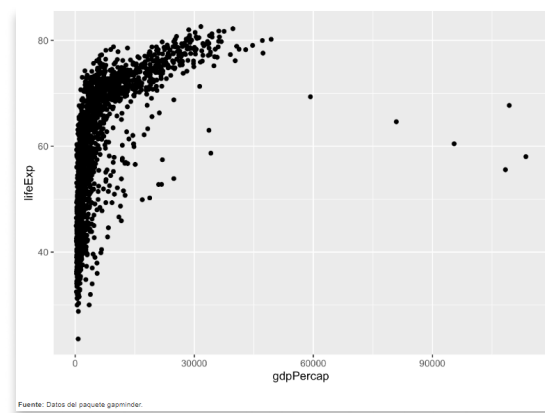
```
ggplot(data=gapminder, aes(x=gdpPercap, y=lifeExp))  
ggplot(data=gapminder, aes(x=gdpPercap, y=lifeExp))+  
  geom_point()
```



Paso 5

Jugar con las facetas, escala y coordenadas, si es necesario

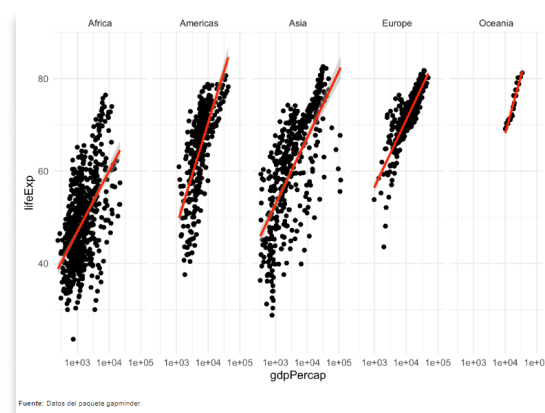
```
ggplot(data=gapminder, aes(x=gdpPercap, y=lifeExp))+  
  geom_point()+  
  facet_grid(~continent)+  
  scale_x_log10()+  
  stat_smooth(method="lm", col="red")
```



Paso 6

Definir el tema si es necesario.

```
ggplot(data=gapminder, aes(x=gdpPercap, y=lifeExp))+  
  geom_point()+  
  facet_grid(~continent)+  
  scale_x_log10()+  
  stat_smooth(method="lm", col="red")+  
  theme_minimal()
```



03

Geometrías de distribución

¿Qué quiero mostrar?

La **distribución** de una **variable** en el **tiempo**:

Histograma

Variable numérica

Gráfico de barras

Variable categórica con varios factores

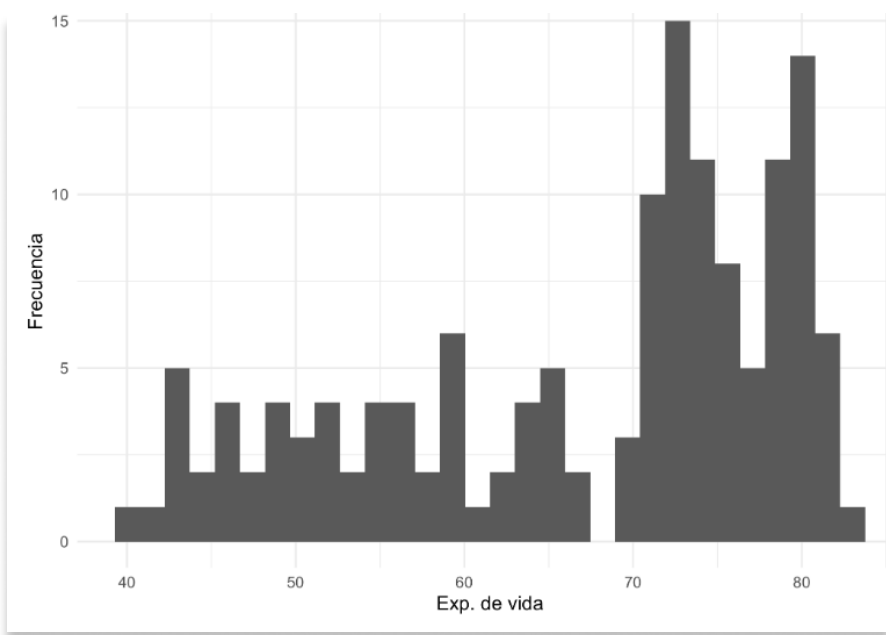
Boxplot

Variable numérica y una categórica

¿Cómo lo muestro?

Histograma

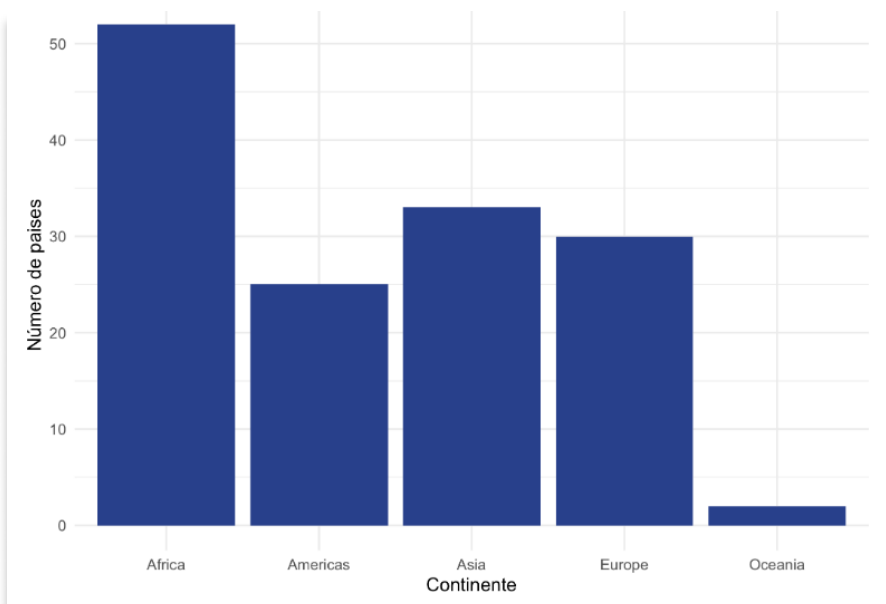
geom_histogram()



```
gapminder%>%  
filter(year==2007)%>%  
ggplot(aes(x=lifeExp))+  
geom_histogram(fill="lightblue")+  
labs(y="Frecuencia", x="Exp. de vida")+  
theme_minimal()
```

Gráfico de barras

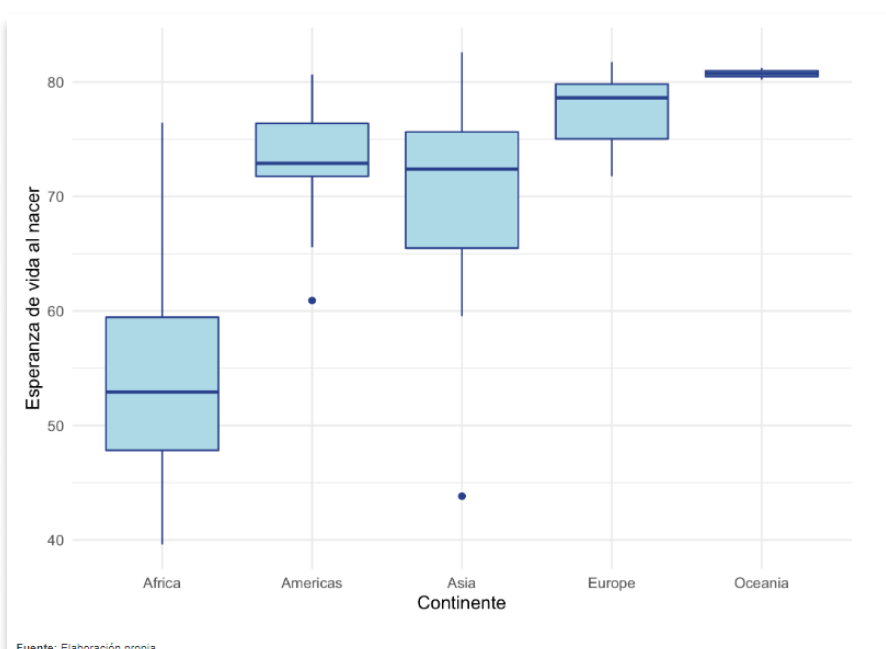
geom_bar()



```
gapminder%>%  
filter(year==2007)%>%  
ggplot(aes(x=continent))+  
geom_bar(fill="royalblue4")+  
labs(y="Frecuencia", x="Continente")+  
theme_minimal()
```

Boxplot

geom_boxplot



```
gapminder%>%  
filter(year==2007)%>%  
ggplot(aes(x=continent, y=100*(..count..)/sum(..count..)))+  
geom_bar(fill="royalblue4")+  
labs(y="%", x="Continente")+  
theme_minimal()
```

04 Geometrías de evolución

¿Qué quiero mostrar?

La **evolución** de una **variable** en el **tiempo**:

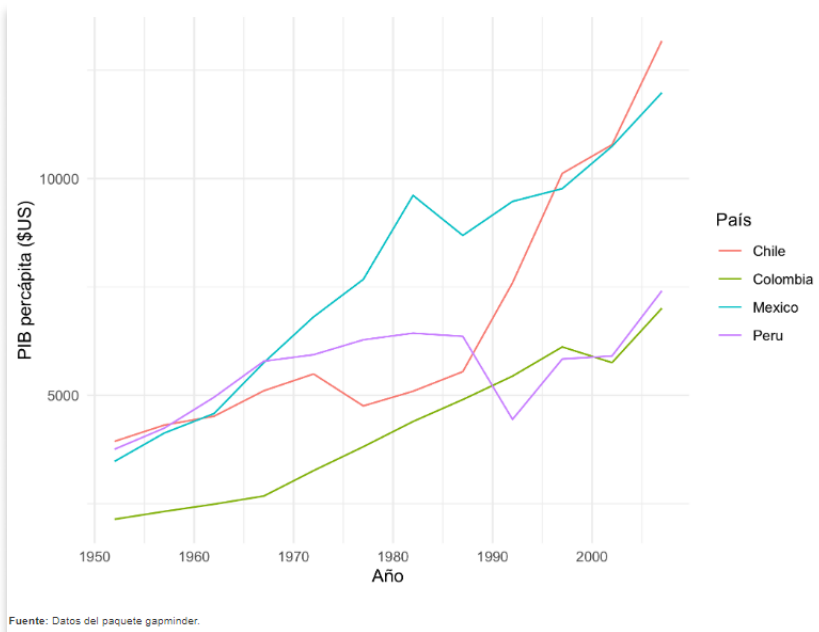
Gráfico de líneas
Variables numéricas

Gráfico de columnas apiladas
Variable categórica con varios factores

Gráfico de barras agrupadas
Variable categórica con varios factores

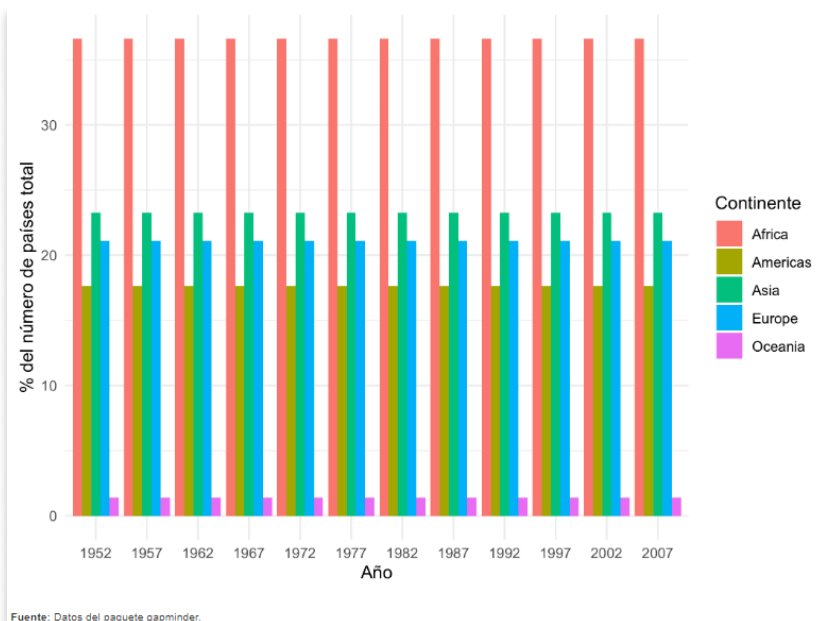
¿Cómo lo muestro?

Gráfico de líneas
geom_line()



```
gapminder %>%  
  filter(country %in% c("Colombia", "Peru", "Chile", "Mexico")) %>%  
  ggplot( aes(x = year, y = gdpPerCap, color = country)) +  
  geom_line() +  
  labs( y="PIB per cápita ($US)",  
        x="Año", color = "País") +  
  theme_minimal()
```

Gráfico de columnas apiladas
geom_bar()

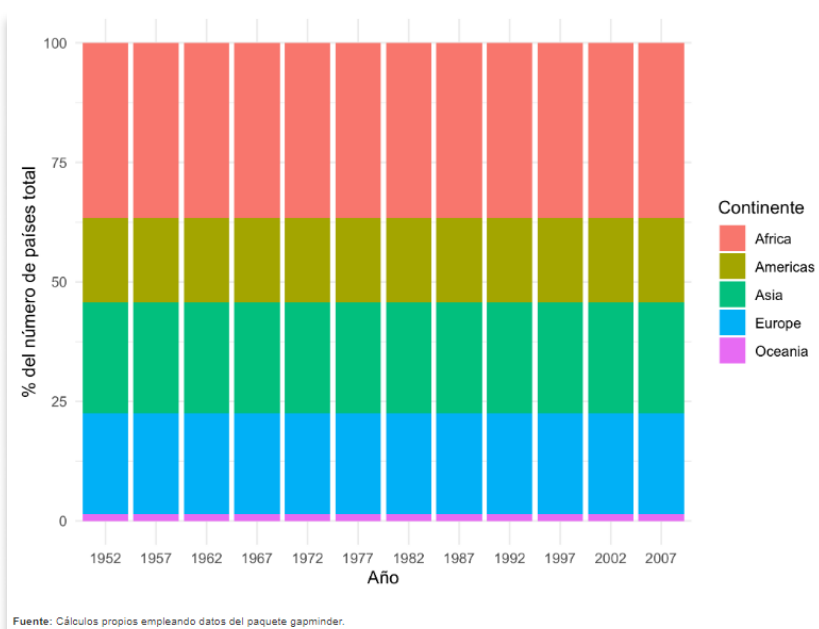


```
d1 <- gapminder %>%  
  group_by(year, continent) %>%  
  summarise( paises = n()) %>%  
  mutate( Prop_paises = 100 * paises / sum(paises))  
ggplot(d1, aes(x = as.character(year), y = Prop_paises,  
              fill = continent)) +  
  geom_bar(position = "dodge", stat = "identity") +  
  labs(y = "% del número de países total",  
       x = "Año", fill = "Continente") +  
  theme_minimal()
```

NOTA

Primero creamos una variable con el número de países por continente

Gráfico de barras agrupadas
geom_bar



```
ggplot(d1, aes(x = as.character(year), y = Prop_paises,  
              fill = continent)) +  
  geom_bar(position = "stack", stat = "identity") +  
  labs(y = "% del número de países total",  
       x = "Año", fill = "Continente") +  
  theme_minimal()
```

05 Geometrías de relación

¿Qué quiero mostrar?

La **relación** entre **variables**:

Diagrama de dispersión

Dos variables numéricas

Gráfico de burbujas

Tres variables numéricas

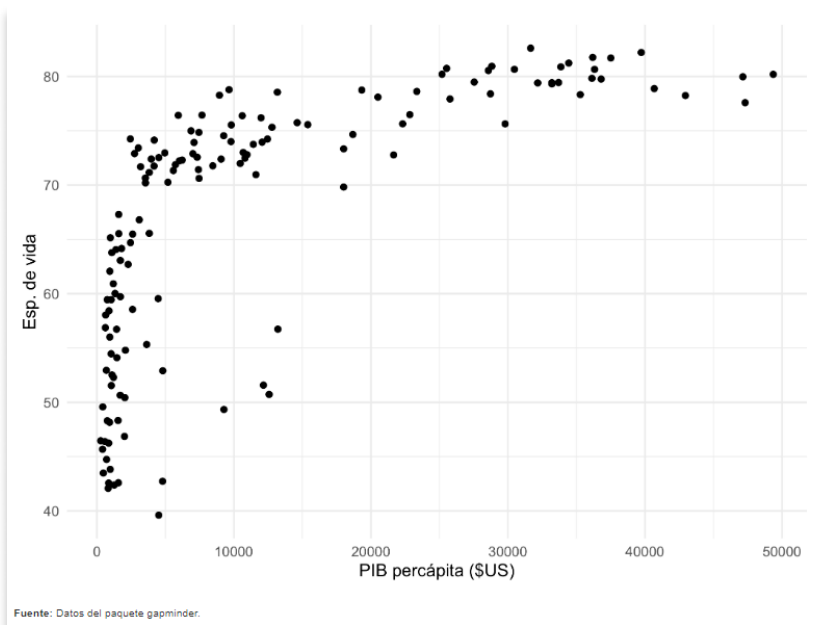
Diagrama de dispersión y variable cuantitativa

Tres variables numéricas y una categórica

¿Cómo lo muestro?

Diagrama de dispersión

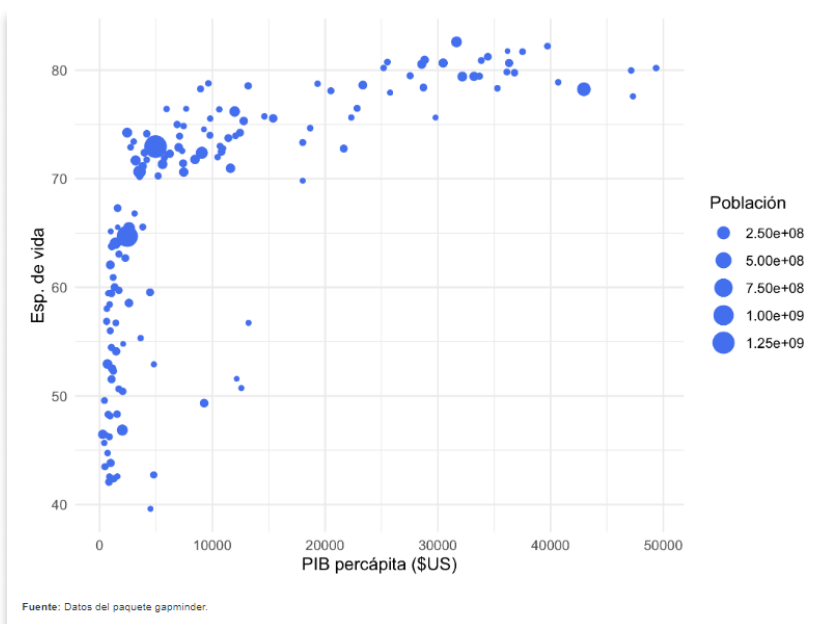
geom_point()



```
gapminder %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp)) +
  geom_point() +
  labs(x = "PIB per cápita", y = "Esp. de vida") +
  theme_minimal()
```

Gráfico de burbujas

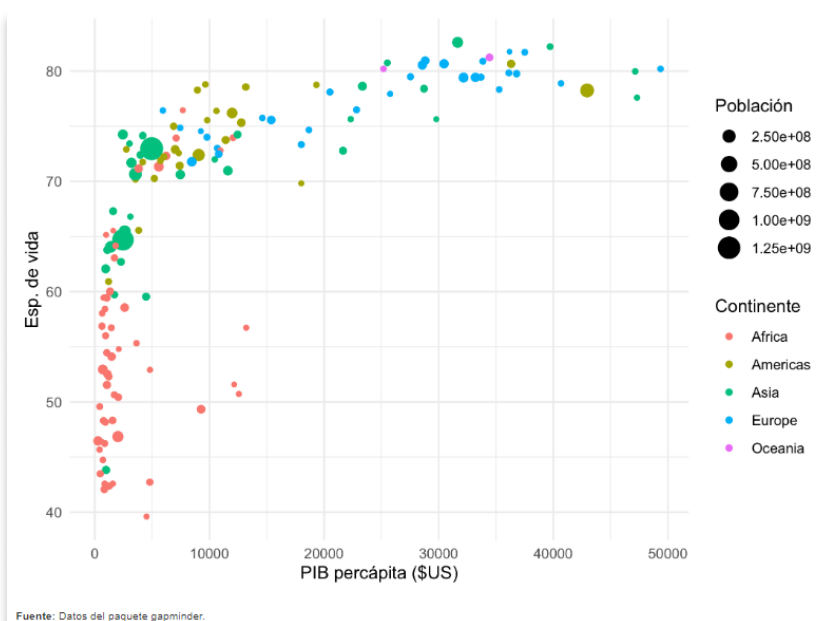
geom_point()



```
gapminder %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
             size = pop)) +
  geom_point(col = "royalblue2") +
  labs(x = "PIB per cápita ($US)",
       y = "Esp. de vida",
       size = "Población") +
  theme_minimal()
```

Diagrama de dispersión y variable cuantitativa

geom_point()

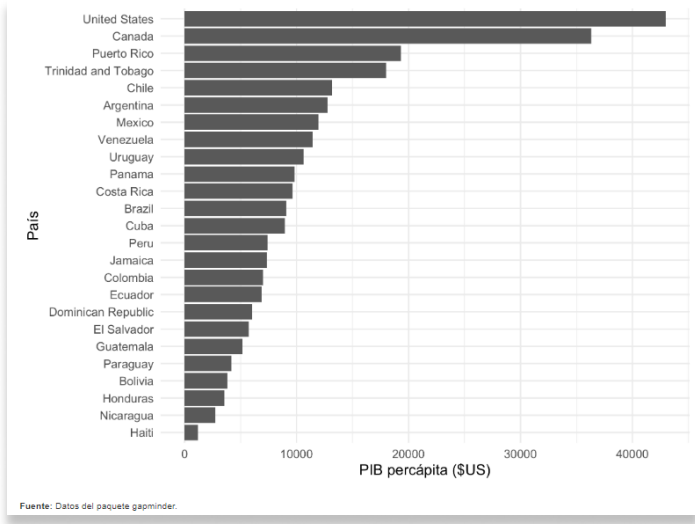


```
gapminder %>%
  filter(year == 2007) %>%
  ggplot(aes(x = gdpPercap, y = lifeExp,
             size = pop, col = continent)) +
  geom_point() +
  labs(x = "PIB per cápita ($US)",
       y = "Esp. de vida",
       size = "Población",
       col = "Continente") +
  theme_minimal()
```

06 Recomendaciones

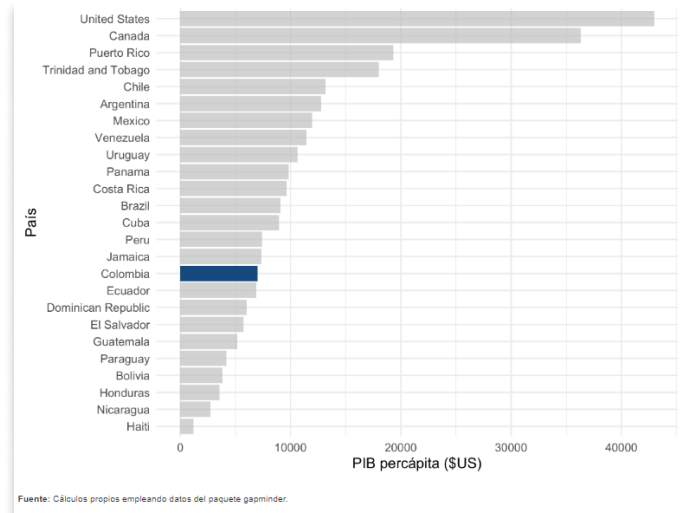
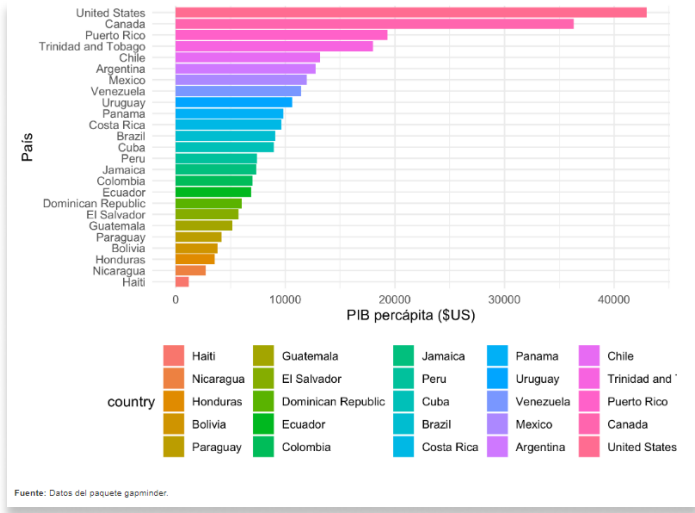
Ordena los datos

Esto mejora la recepción del mensaje.



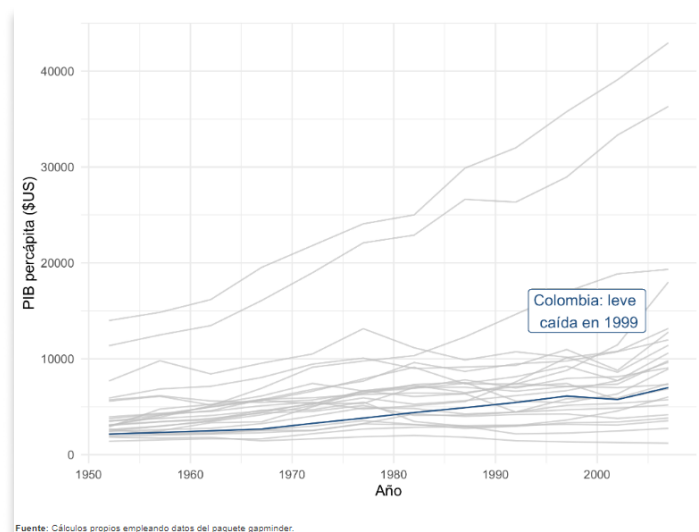
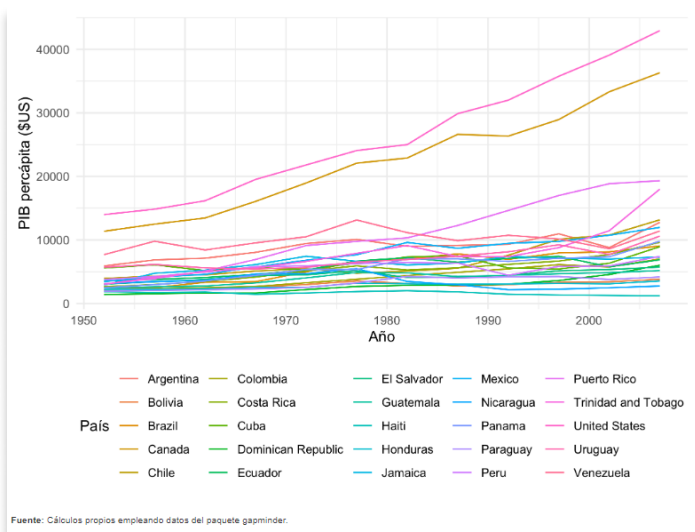
Los colores deben comunicar algo

Muchos colores pueden desviar la atención del objetivo del mensaje.



Evitar gráficos con muchas líneas (spaguetti)

No son muy claros.



No usar el gráfico de torta

El ojo humano no es bueno midiendo ángulos. La misma información la puede mostrar un gráfico de barras.

