

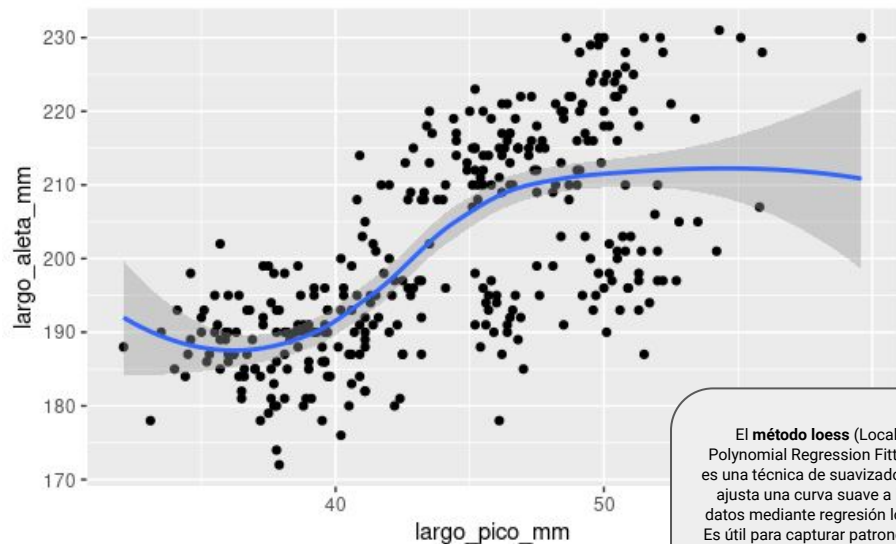
Clase 6

unab

VISUALIZACIÓN DE LA INFORMACIÓN 2024

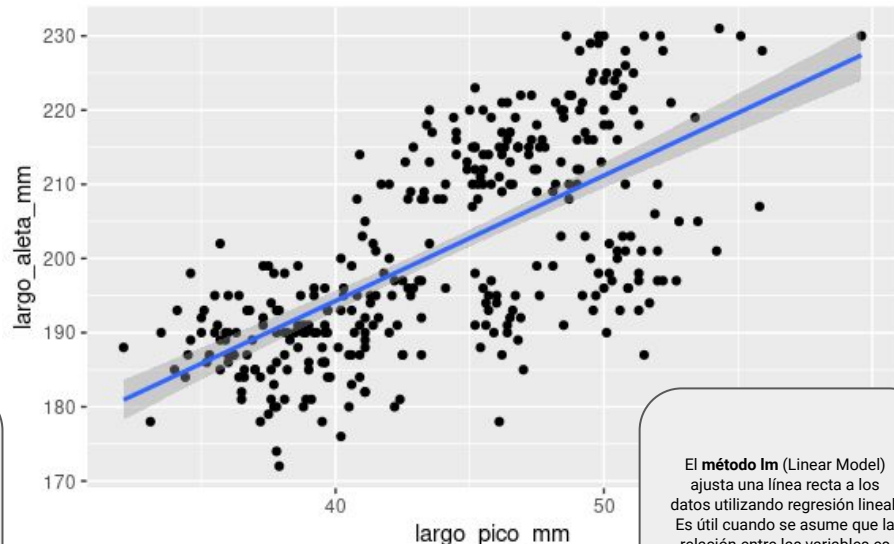
Visualizando tendencias

geom_smooth()



Método loess
(default)

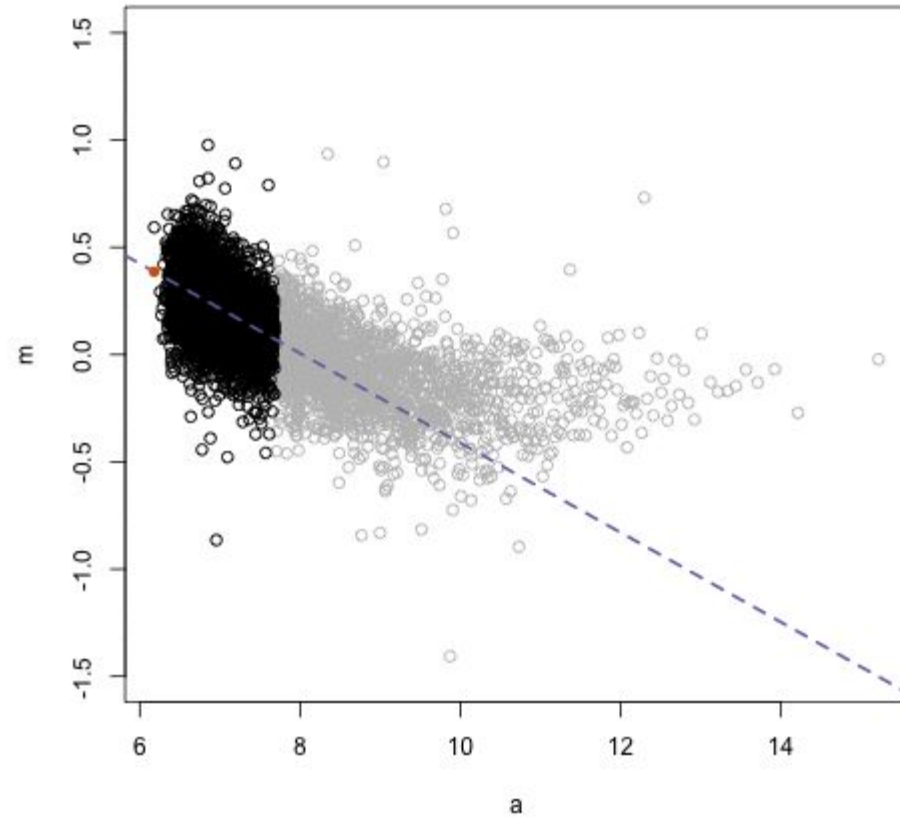
El método loess (Local Polynomial Regression Fitting) es una técnica de suavizado que ajusta una curva suave a los datos mediante regresión local. Es útil para capturar patrones no lineales en los datos sin asumir una forma específica de la relación entre las variables.



Método lm
(method = "lm")

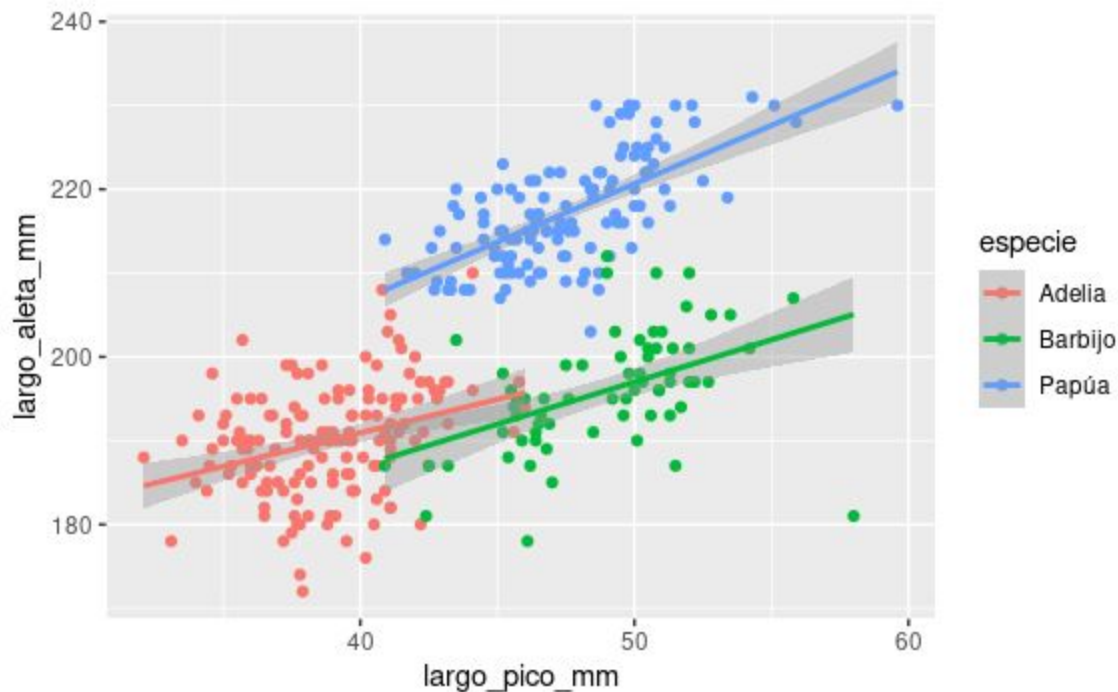
El método lm (Linear Model) ajusta una línea recta a los datos utilizando regresión lineal. Es útil cuando se asume que la relación entre las variables es lineal.

El tipo de ajuste por defecto depende del tamaño de la muestra



<https://simplystatistics.org/posts/2014-02-13-loess-explained-in-a-gif/>

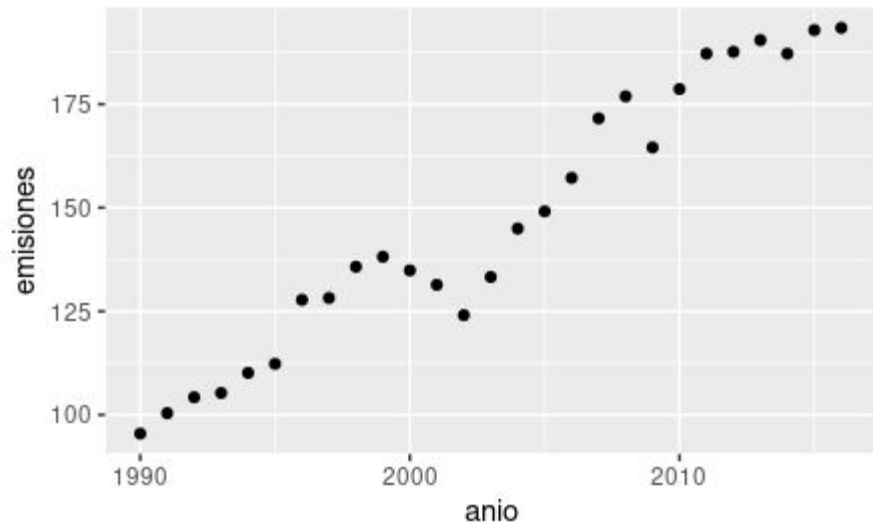
más de dos variables



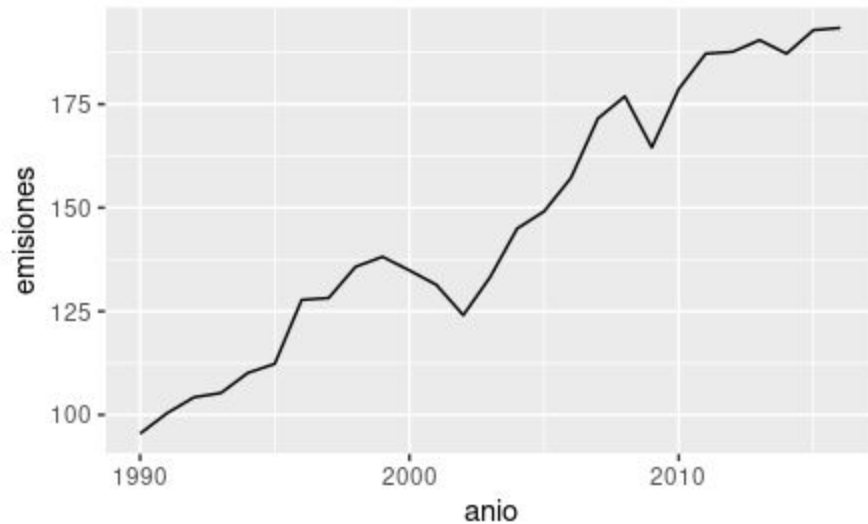
El ajuste se puede aplicar a cada grupo

Visualizar series temporales

geom_line()

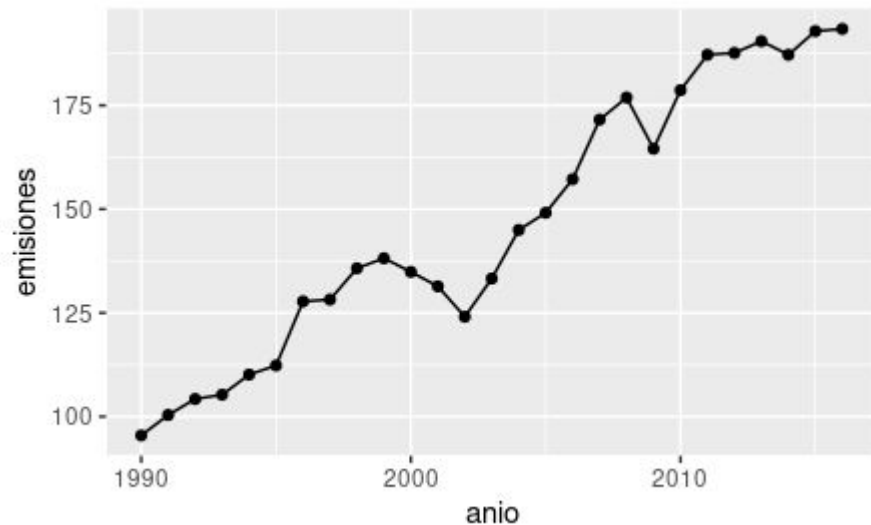


geom_point()



geom_line()

redundancia

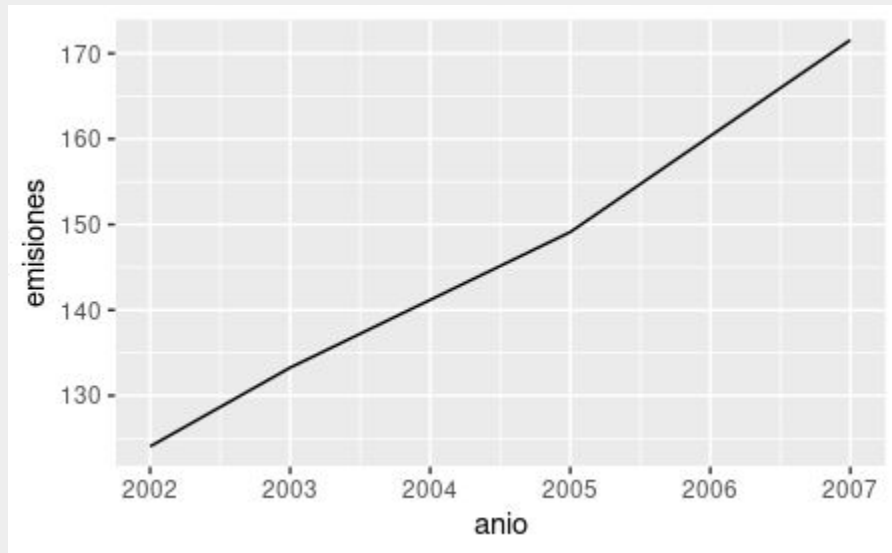


¿Es realmente necesario?

```
geom_point() +  
geom_line()
```

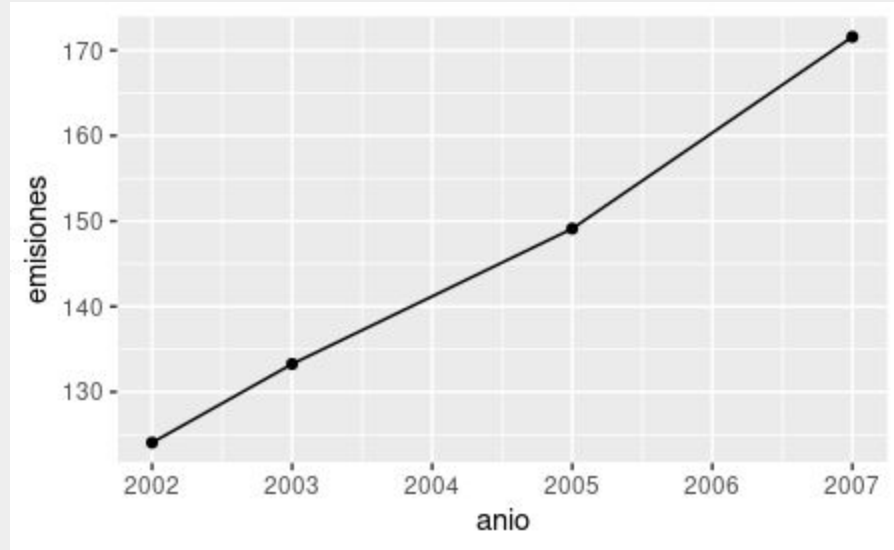

Que opinan?

¿Cuántos datos (información de emisiones en cada año) se usan en este gráfico?

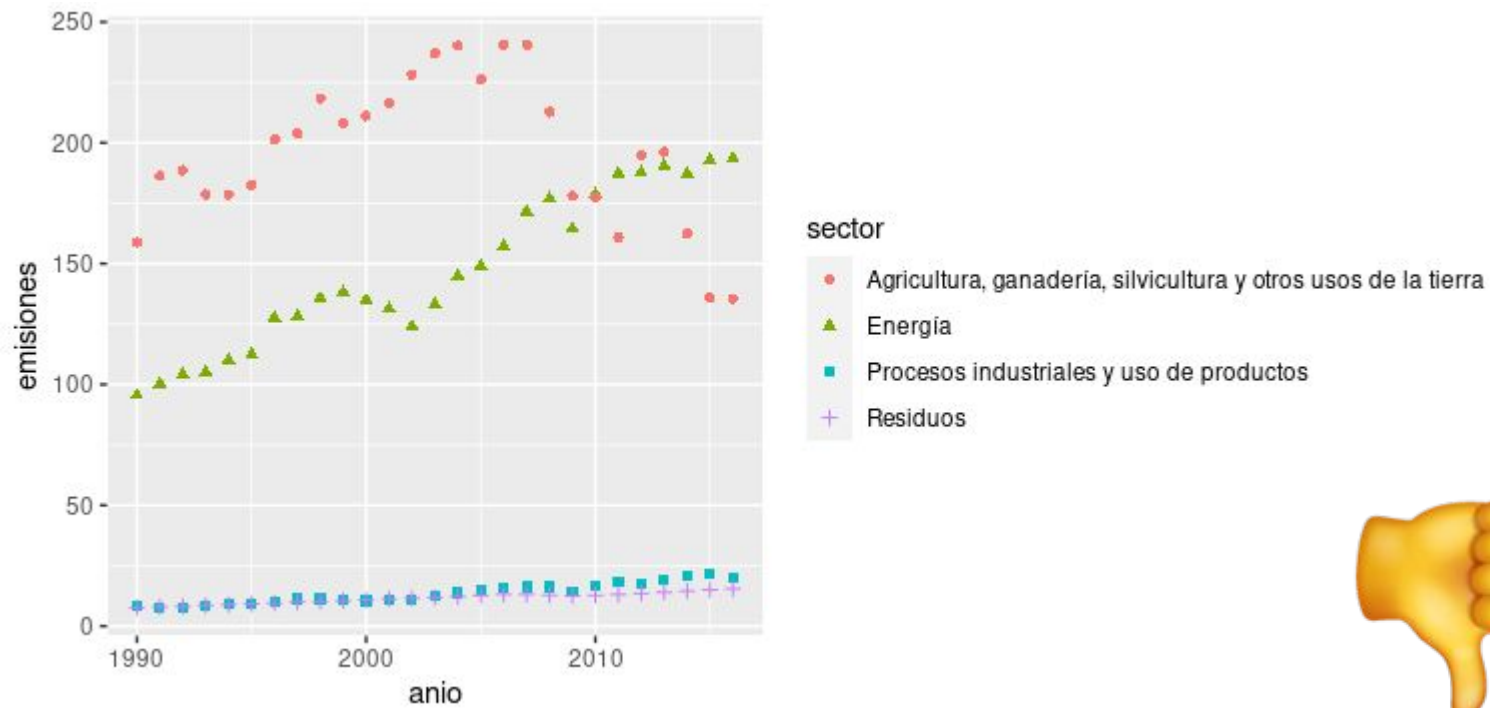


Que opinan?

¿Cuántos datos (información de emisiones en cada año) se usan en este gráfico?

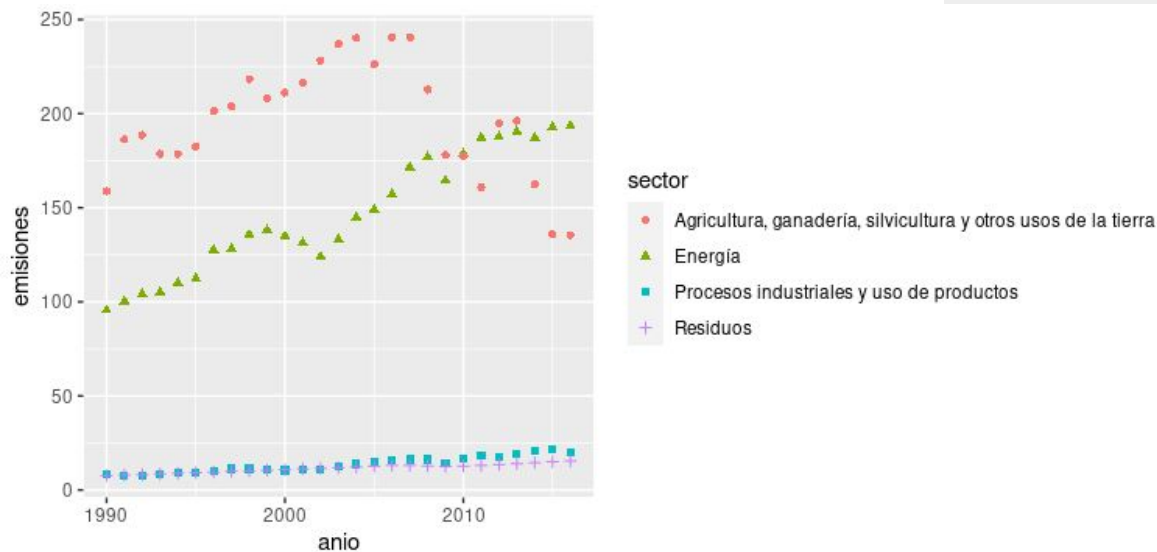


Varias series temporales

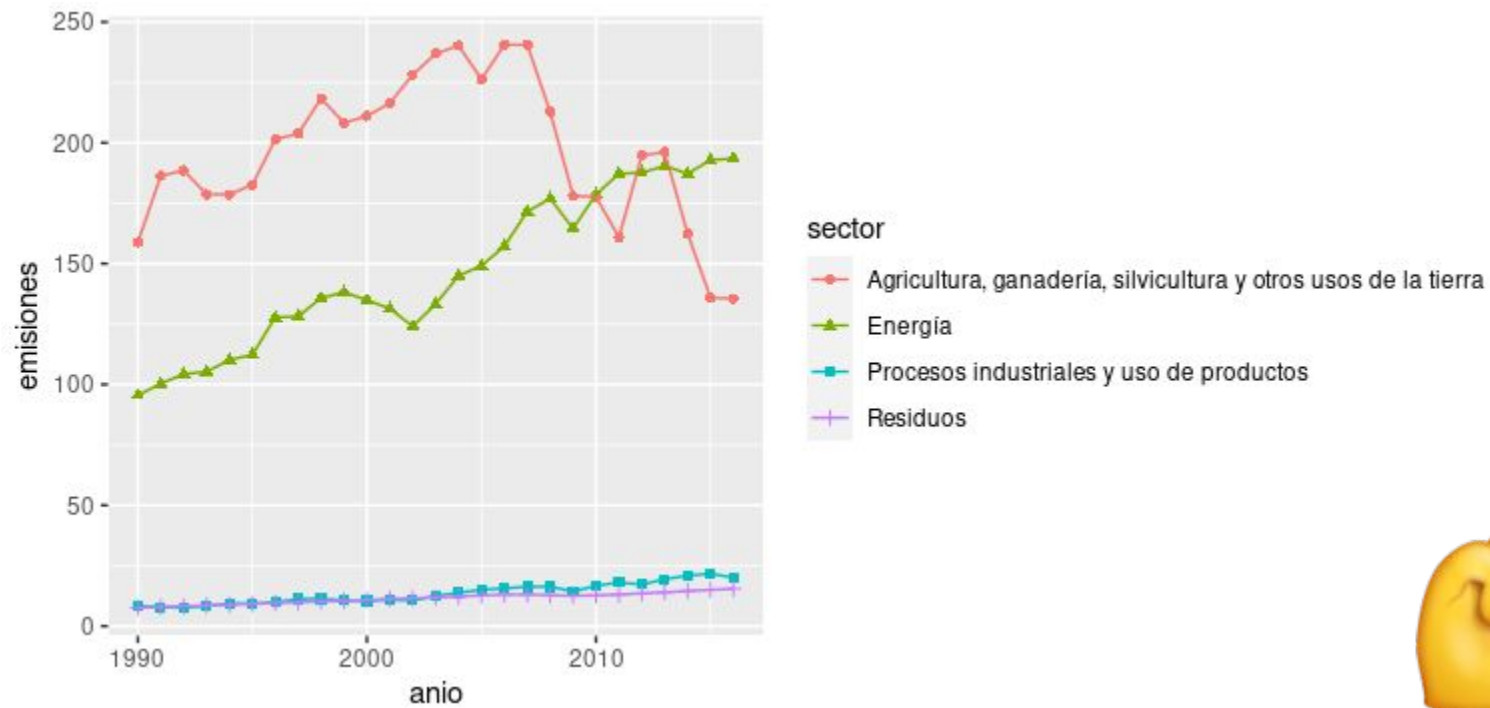


Que opinan?

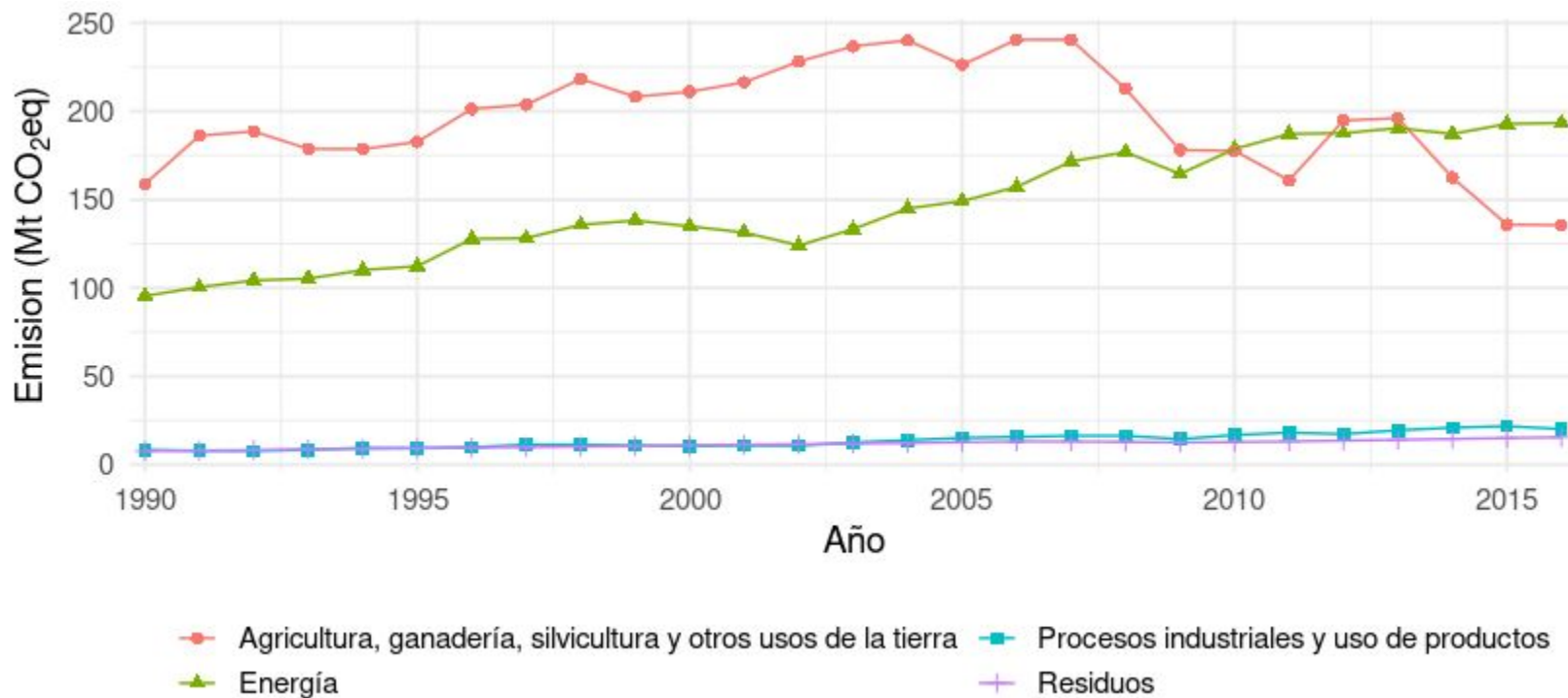
1. ¿Qué falta para generar el gráfico?
2. ¿Qué cosas podríamos mejorarle?



Lineas y puntos



Combinando canales



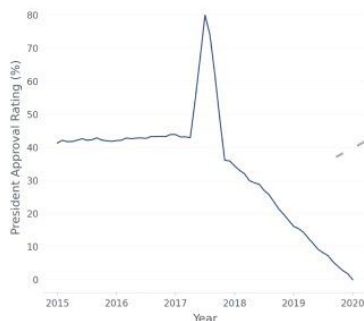
Pausa

10 minutos

**No se
desconecten
pero
retirensen de las
pantallas.**

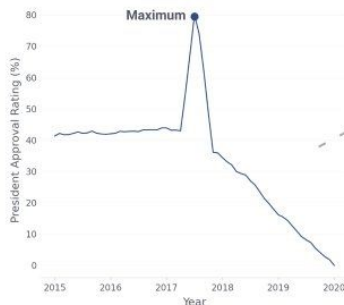
Texto en los gráficos

Pero, ¿cuánto texto?



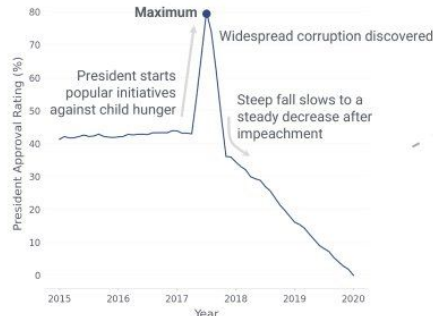
Sin texto, excepto por los ejes

President approval rating over 5 years (2015-2020)



título y 1 anotación

President approval rating over 5 years (2015-2020)



título y muchas anotaciones,
cuenta una historia

The presidential approval rating held steady from 2015 to 2017 at around 42%. In 2017, the president introduced a popular initiative combating child hunger. This led to a quick increase in approval rating, doubling to around 80%. However, the positive sentiment did not last. Due to widespread corruption, the approval rating fell back to around 37% by 2018, and has continued to decline steadily to almost 0% following an impeachment.

solo texto

Análisis de Stokes

De acuerdo a Stokes et al:

- Las personas prefieren gráficos con **muchas anotaciones** por sobre gráficos con poco texto o sólo texto.
- Las **anotaciones con información estadística y relacional o de contexto** generan que las personas recuerden más esa información.
- La **ubicación** de cada tipo de información* **es importante**.

*niveles semánticos

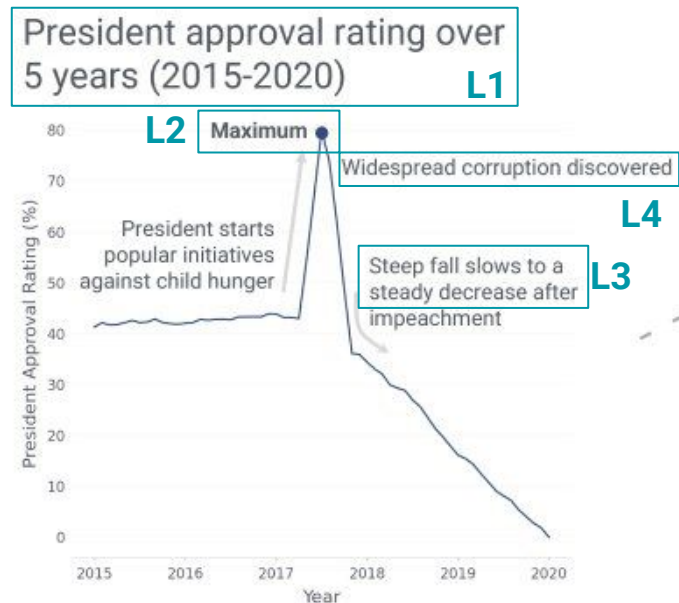
Niveles semánticos → tipos de información

Nivel 1 (L1): aspectos elementales del gráfico

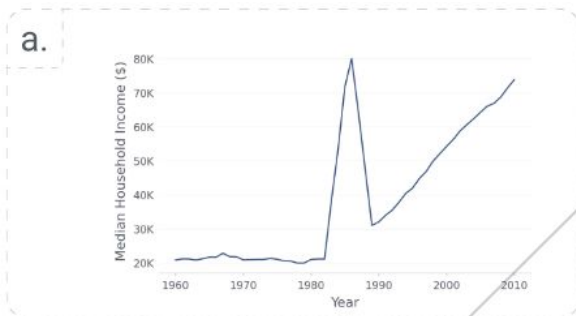
Nivel 2 (L2): aspectos estadísticos o relacionales

Nivel 3 (L3): aspectos perceptuales o cognitivos

Nivel 4 (L4): aspectos externos al gráfico



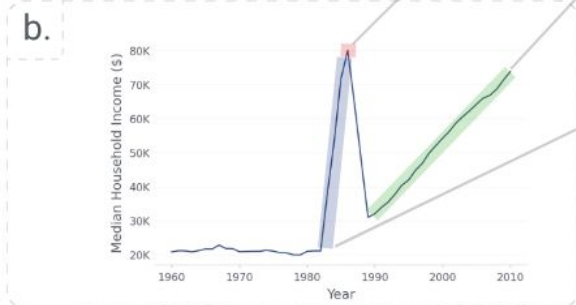
Niveles semánticos - Datos adicionales



c.

L2: Máximo
L3: Pico luego de aumento repentino
L4: Se rompe la burbuja tecnológica

L2: El ingreso en 2010 es mayor que en 1990
L3: Luego de caer, el ingreso familiar aumenta
L4: Recuperación de la economía gracias a estímulos



L2: El ingreso es mayor en 1982 que en 1980
L3: Aumento rápido y repentino
L4: Boom tecnológico

d.

Position: Title

Ingreso familiar medio a lo largo del tiempo

La línea muestra el cambio del ingreso familiar medio cada 10 años entre 1960 y 2010

Position: Trend

Position: Axis
Años 1960-2010, en décadas

L1

Consejos generales

1. Las anotaciones deben incluir texto relevante
2. “relevante” o “mejor” depende del mensaje que se quiere dar
3. Para transmitir el mensaje, la ubicación de las anotaciones dependen del tipo de contenido
4. Considerar incluir una versión “sólo texto”.

La anotación debe cumplir:

- **Que sea legible:** que no se superponga con otras anotaciones
- **Usar lenguaje visual:** flechas para tendencias, puntos para puntos específicos, intervalos para rangos.
- **Concordancia espacial:** la anotación debe estar cerca del punto de referencia

Gráfico estándar

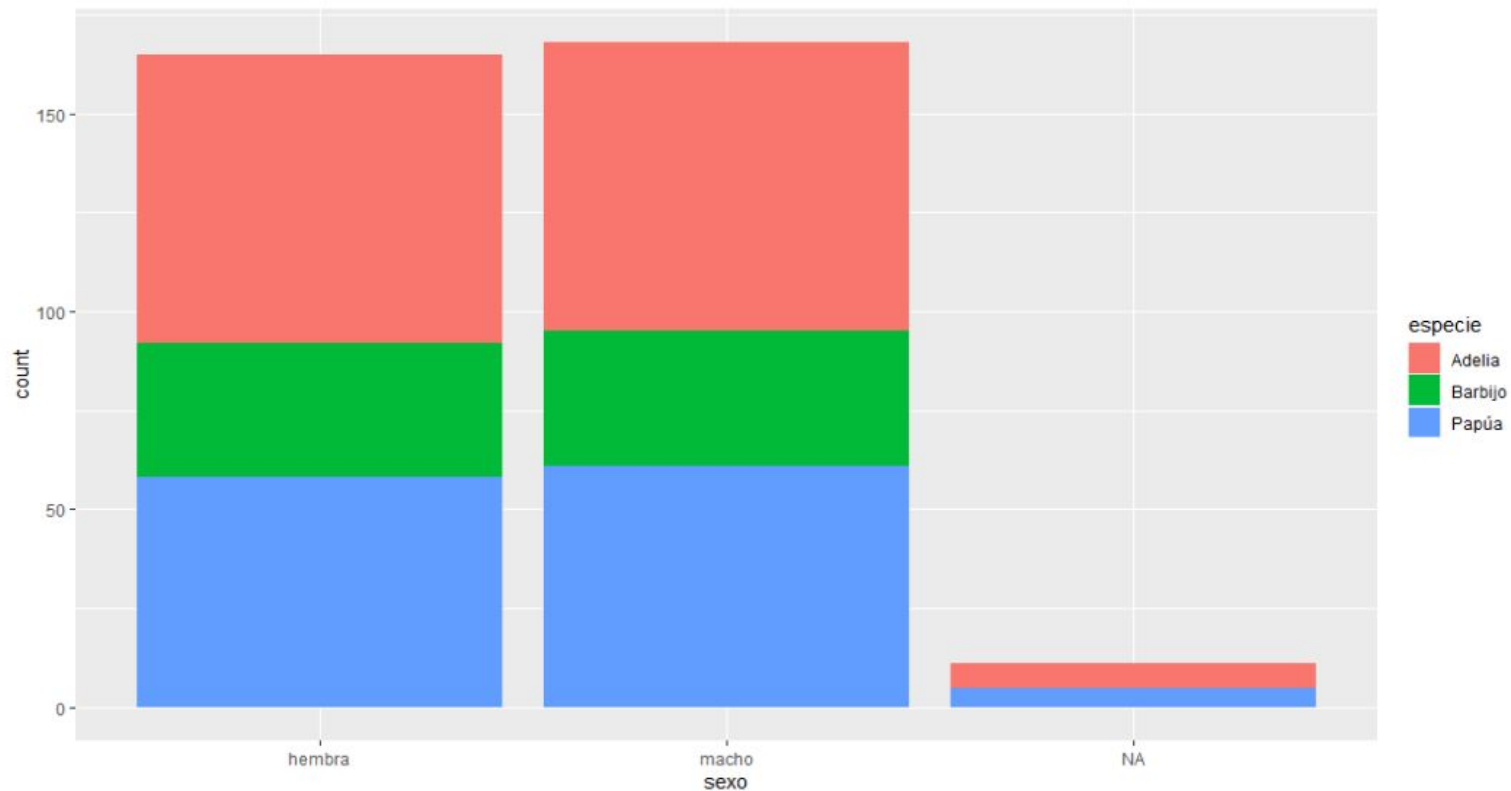
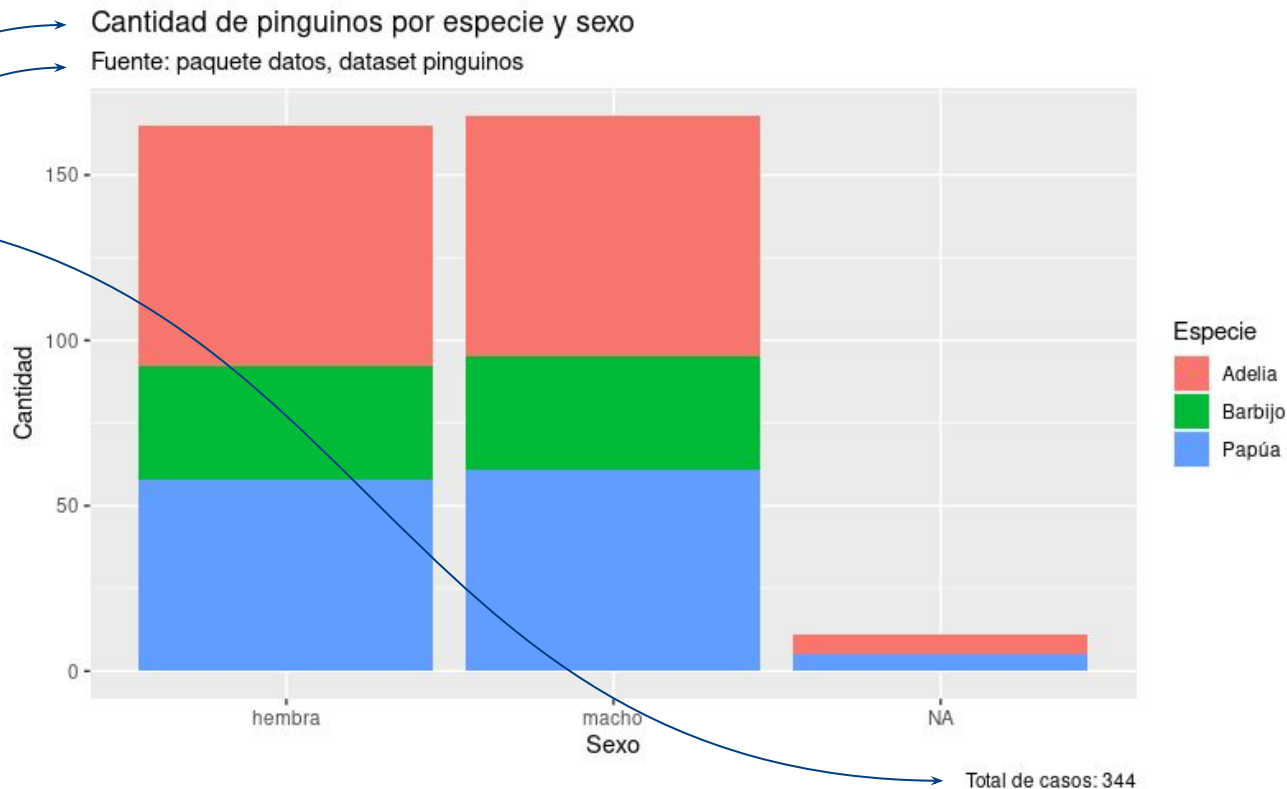
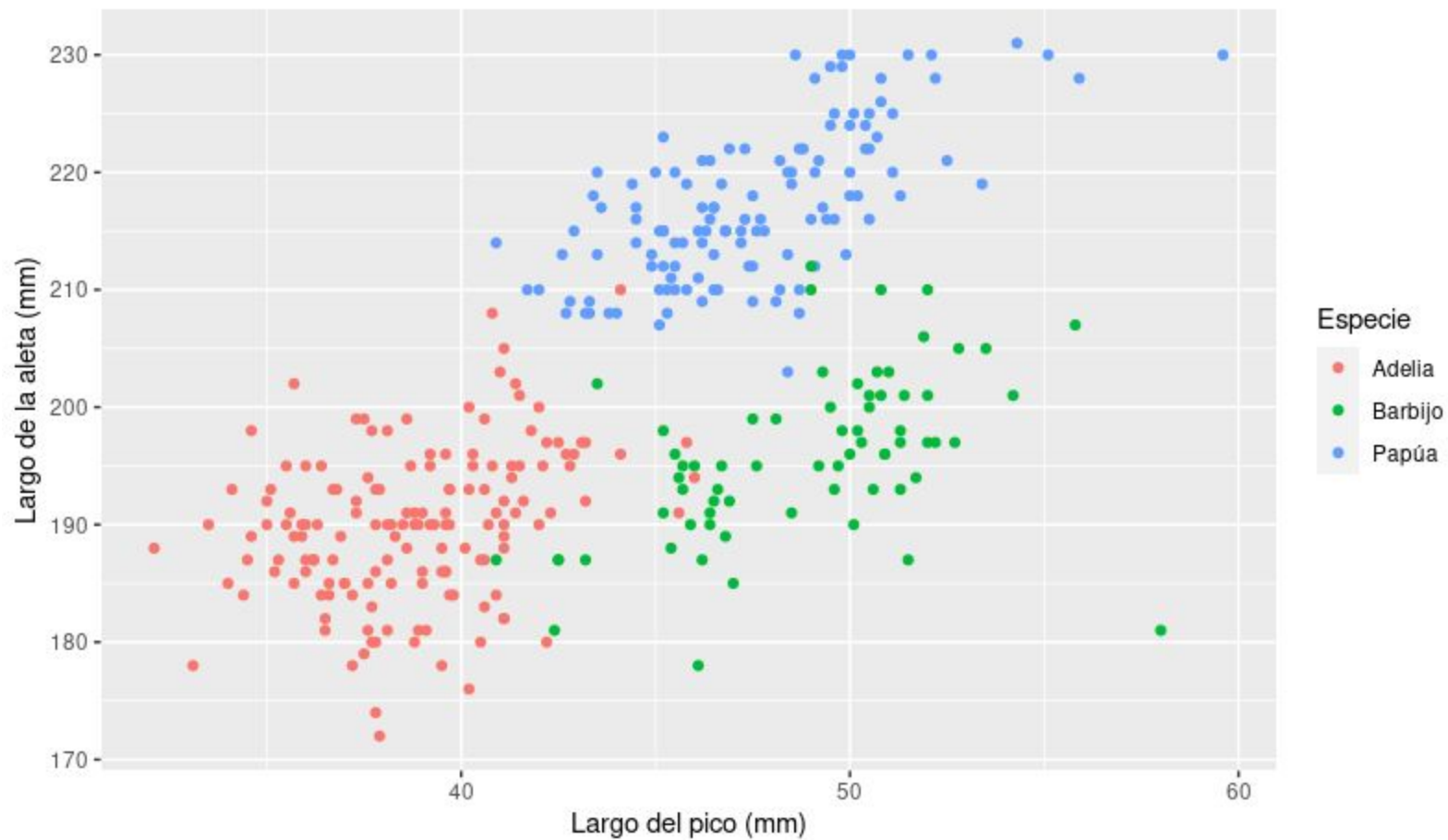


Gráfico con anotaciones generales

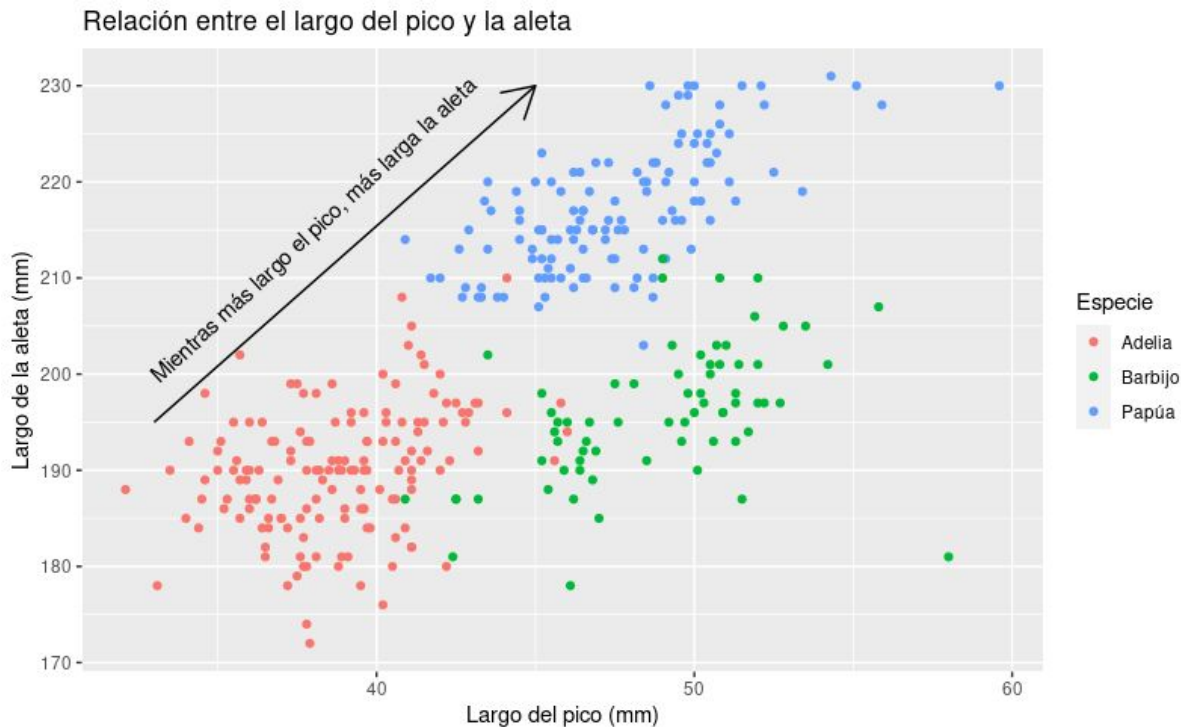
```
labs(  
  title = ...,  
  subtitle = ...,  
  caption = ...,  
  x = ...,  
  y = ...,  
  fill = ...  
)
```



Relación entre el largo del pico y la aleta

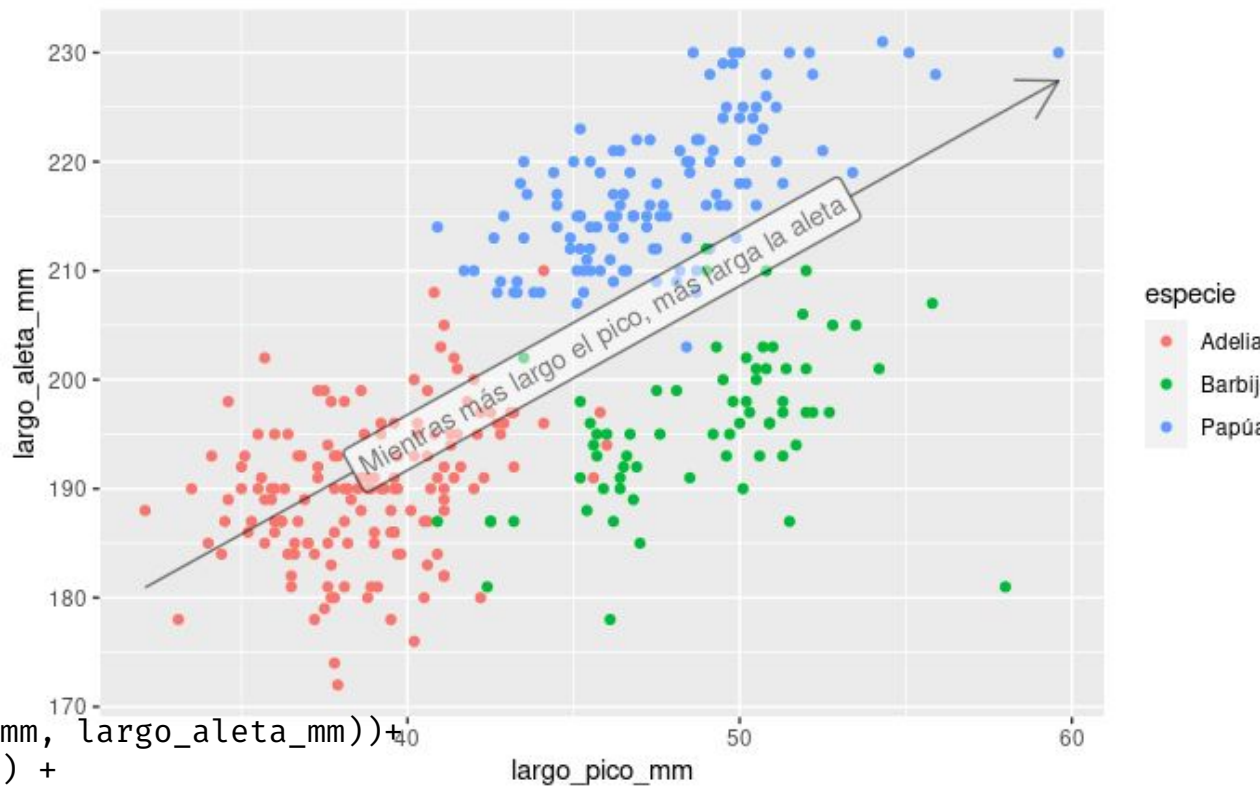


annotate()



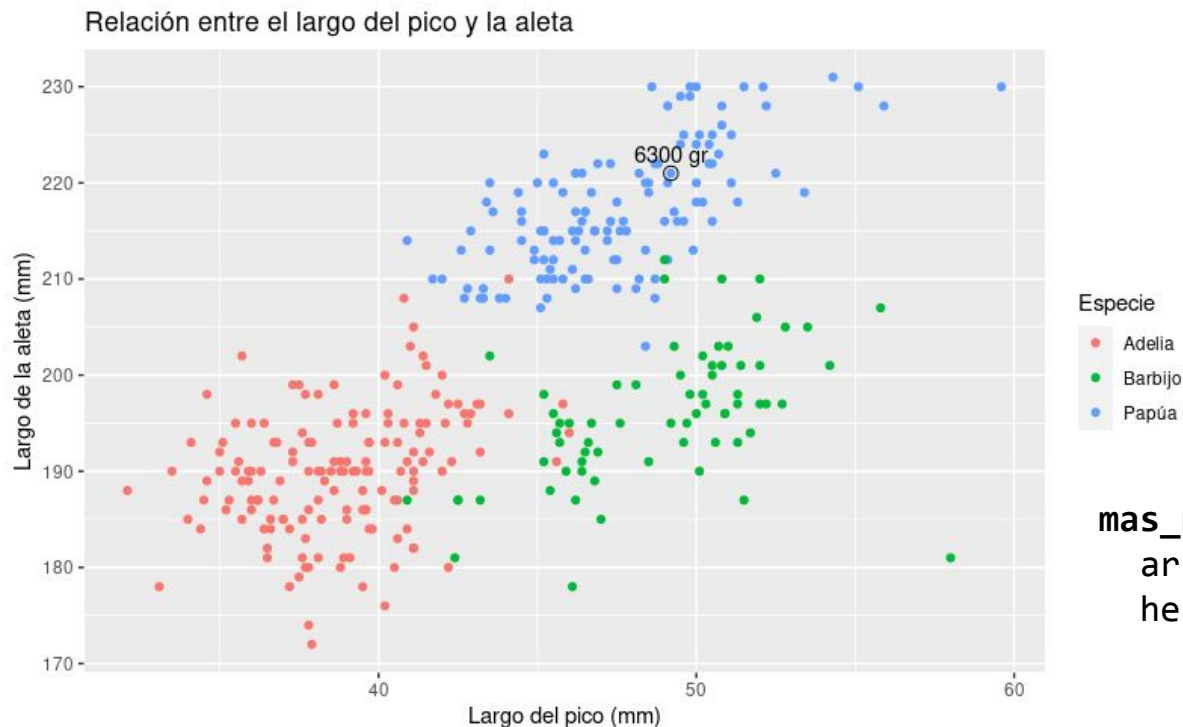
```
annotate("text", x = 38, y = 215, angle = 43,  
        label = "Mientras más largo el pico, más larga la aleta") +  
annotate("segment", x = 33, xend = 45, y = 195, yend = 230, arrow = arrow())
```

geom_textpath::geom_labelsmooth()



```
ggplot(pinguinos, aes(largo_pico_mm, largo_aleta_mm)) +  
  geom_point(aes(color = especie)) +  
  geom_labelsmooth(method = "lm",  
    label = "Mientras más largo el pico, más larga la aleta", alpha = 0.5,  
    arrow = arrow())
```

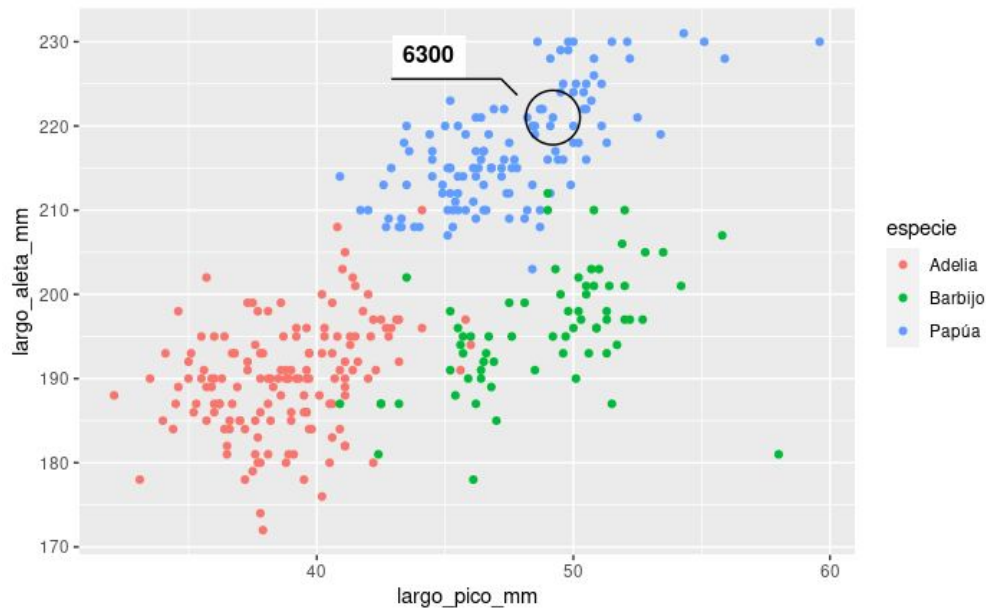
geom_text()



```
mas_pesado <- pinguinos %>%  
  arrange(desc(masa_corporal_g)) %>%  
  head(n = 1)
```

```
geom_text(data = mas_pesado, aes(label = paste0(masa_corporal_g, " gr")), nudge_y = 2) +  
geom_point(data = mas_pesado, shape = 21, size = 3, fill = NA)
```

ggforce::geom_mark_circle()



```
pinguinos %>%  
  filter(!is.na(masa_corporal_g)) ▷  
  ggplot(aes(largo_pico_mm, largo_aleta_mm))+  
  geom_point(aes(color=especie))+  
  geom_mark_circle(aes(filter = masa_corporal_g == max(masa_corporal_g, na.rm = TRUE),  
                    label = masa_corporal_g))
```

otras funciones útiles

`geom_curve()`

`geom_label()`

`geom_rect()`

`geom_vline()`

`geom_hline()`

`ggforce::geom_mark_ellipse()`

Pausa

10 minutos

**No se
desconecten
pero
retirensen de las
pantallas.**

Escalas

Generalidades

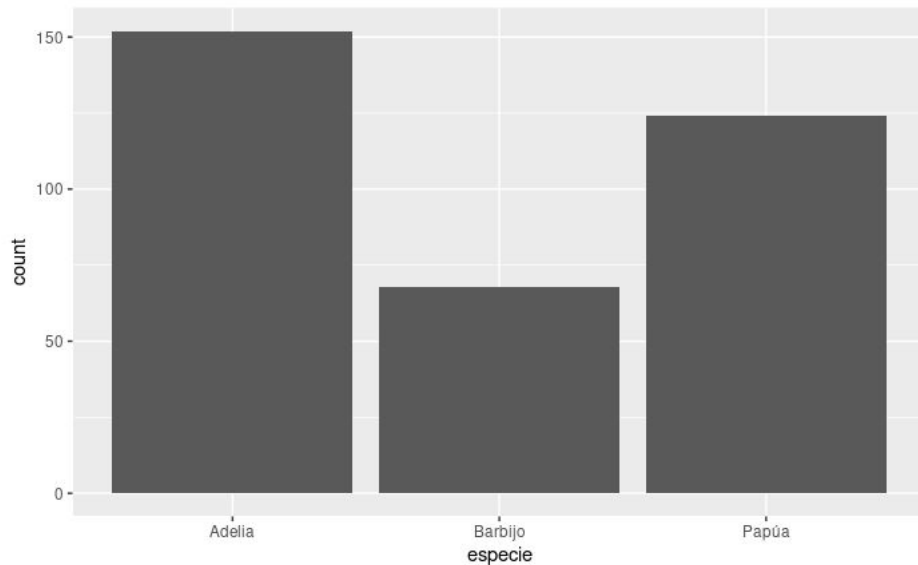
- Se agregan como nuevas capas,
- Puede haber tantas escalas como elementos en el gráfico
- Permiten customizar los elementos del gráfico
- Son funciones de {ggplot2} que siguen (general) un patrón:

```
scale_<elemento>_<tipo de dato>()
```

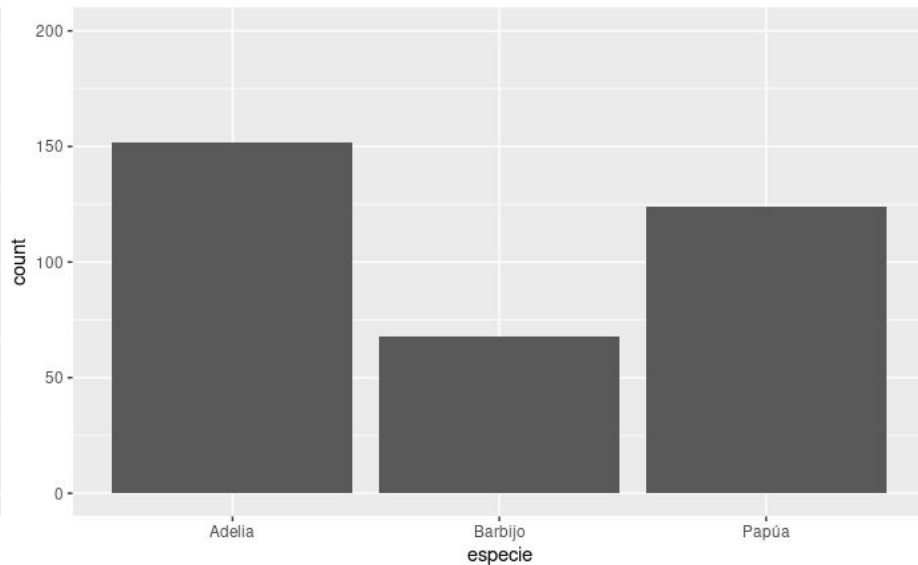
```
ej:  scale_x_continuous()
```

→ **Dependen del elemento y el tipo de dato usado**

Escalas para los ejes

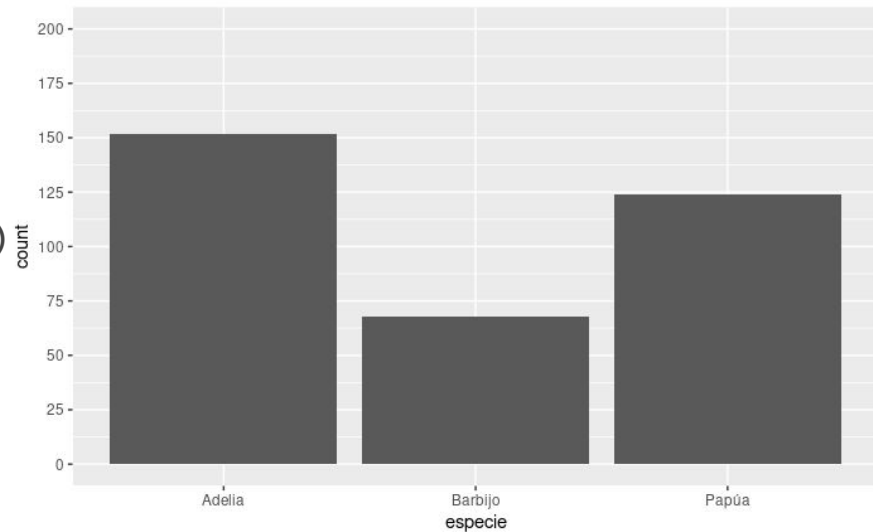


```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete() +  
  scale_y_continuous()
```

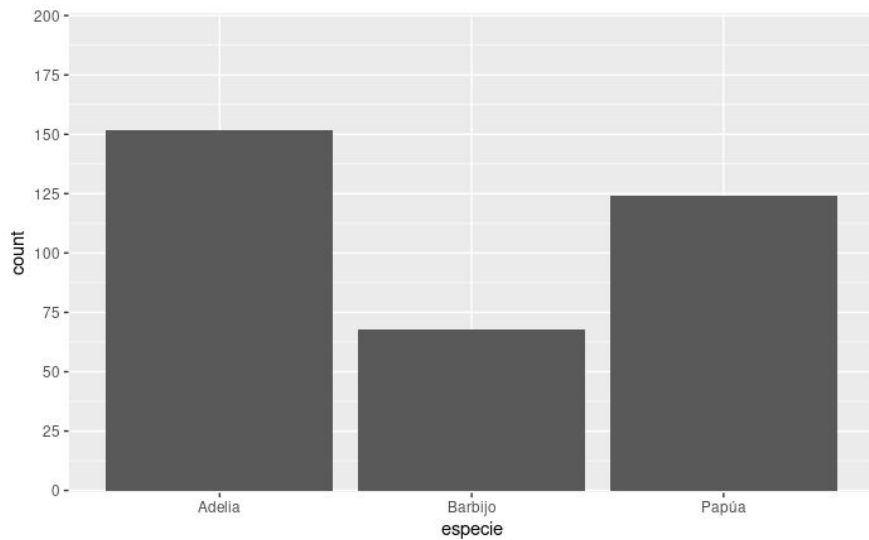


```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete() +  
  scale_y_continuous(limits = c(0, 200))
```

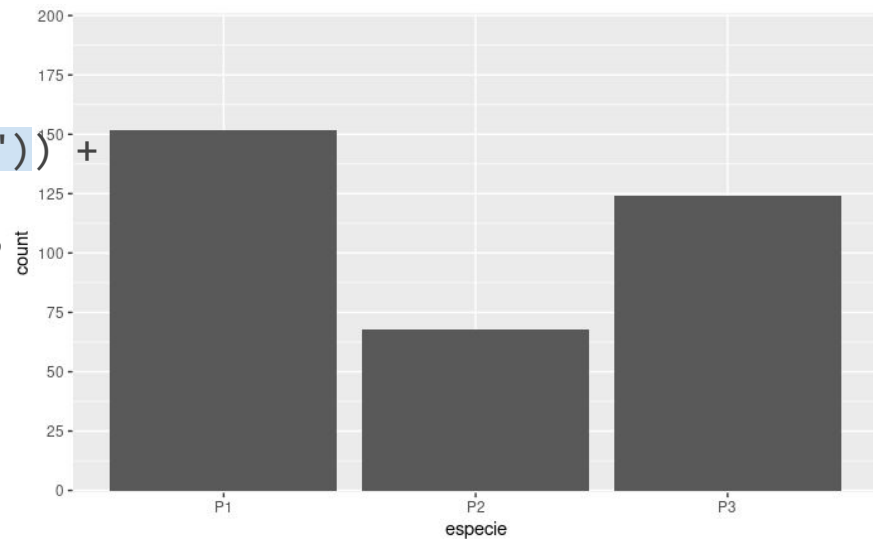
```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete() +  
  scale_y_continuous(limits = c(0, 200),  
                     breaks = seq(0, 200, 25))
```



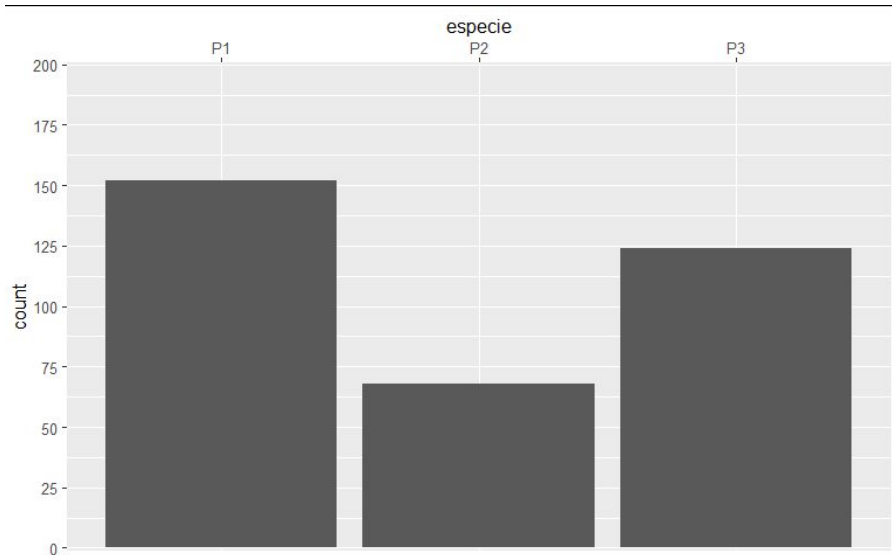
```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete() +  
  scale_y_continuous(limits = c(0, 200),  
                     breaks = seq(0, 200, 25),  
                     expand = c(0, 1))
```



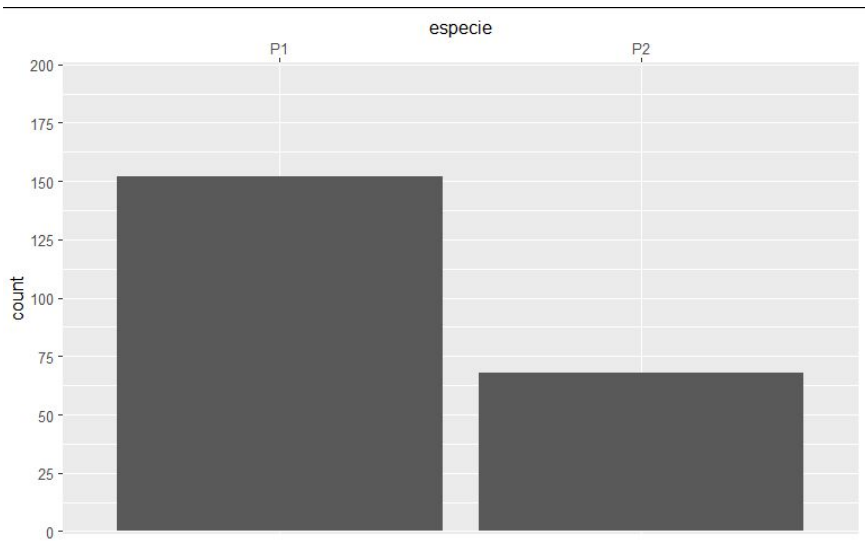
```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete(labels = c("P1", "P2", "P3")) +  
  scale_y_continuous(limits = c(0, 200),  
                     breaks = seq(0, 200, 25),  
                     expand = c(0, 1))
```



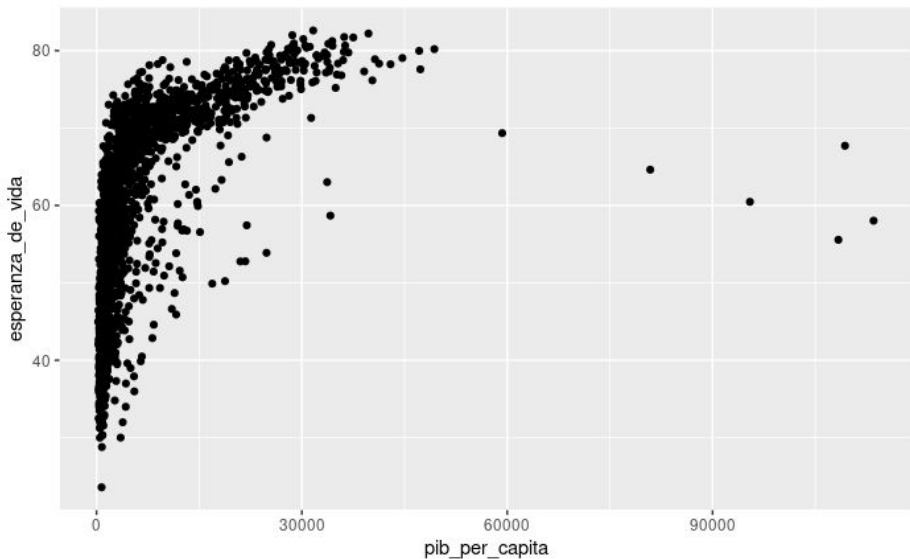
```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete(labels = c("P1", "P2", "P3"),  
                    position = "top") +  
  scale_y_continuous(limits = c(0, 200),  
                     breaks = seq(0, 200, 25),  
                     expand = c(0, 1))
```



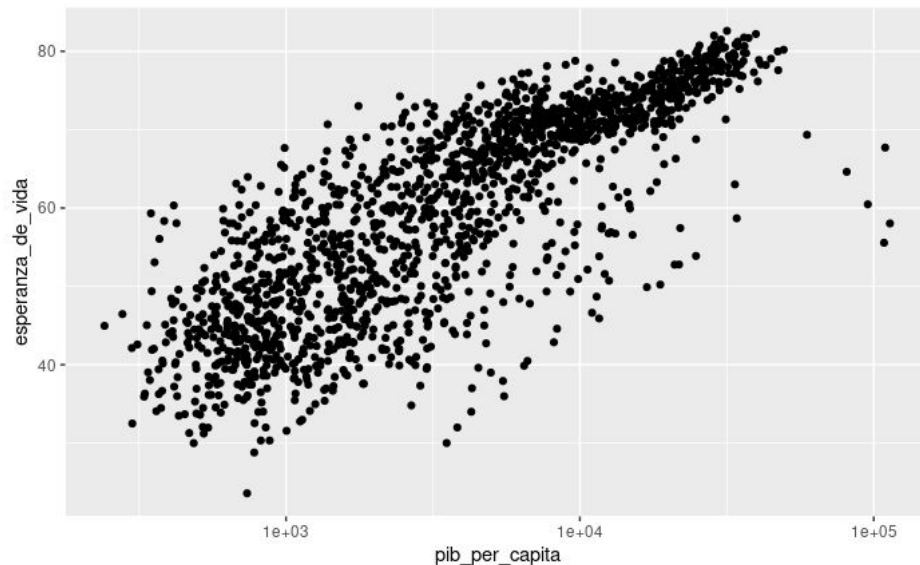
```
ggplot(pinguinos, aes(especie)) +  
  geom_bar() +  
  scale_x_discrete(labels = c("P1", "P2", "P3"),  
                    position = "top",  
                    limits = c("Adelia", "Barbijo")) +  
  scale_y_continuous(limits = c(0, 200),  
                     breaks = seq(0, 200, 25),  
                     expand = c(0, 1))
```



Escalas para los ejes con una transformación

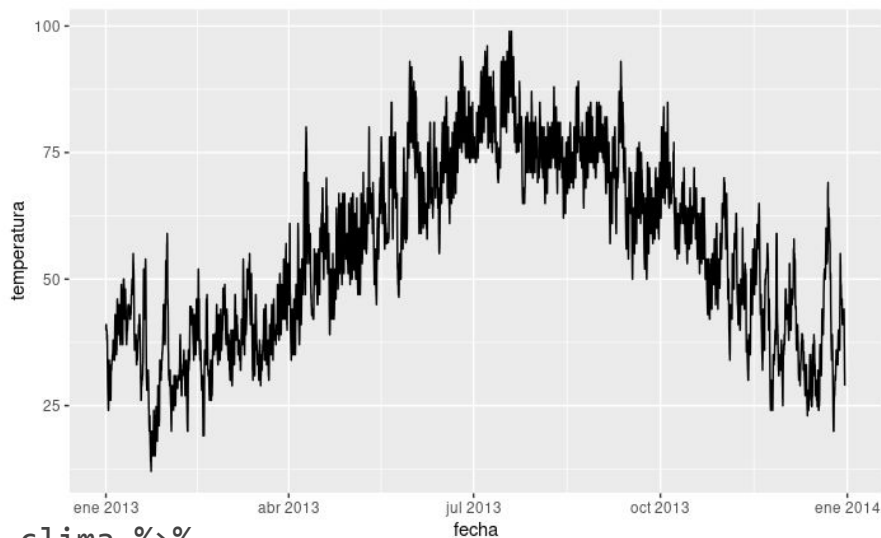


```
ggplot(países,  
  aes(pib_per_capita, esperanza_de_vida)) +  
  geom_point() +  
  scale_x_continuous()
```



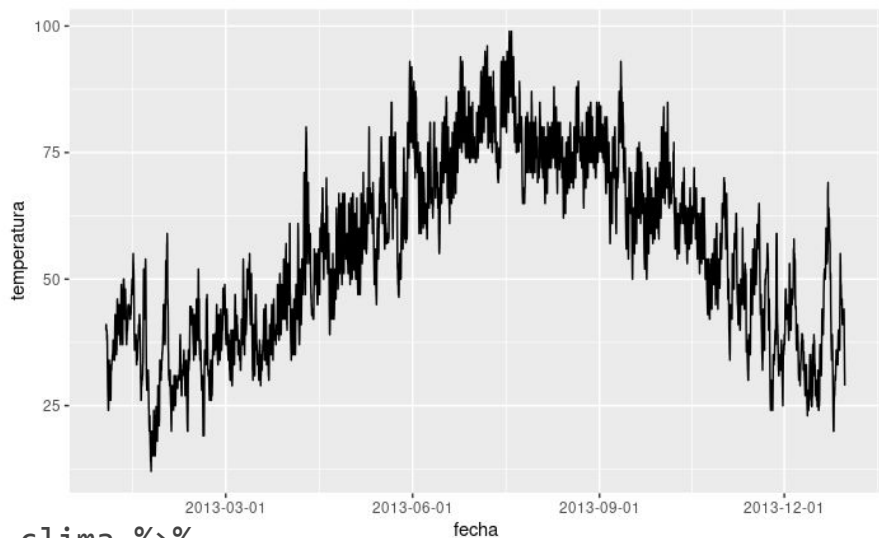
```
ggplot(países,  
  aes(pib_per_capita, esperanza_de_vida)) +  
  geom_point() +  
  scale_x_log10()
```

Escalas para fechas y horas



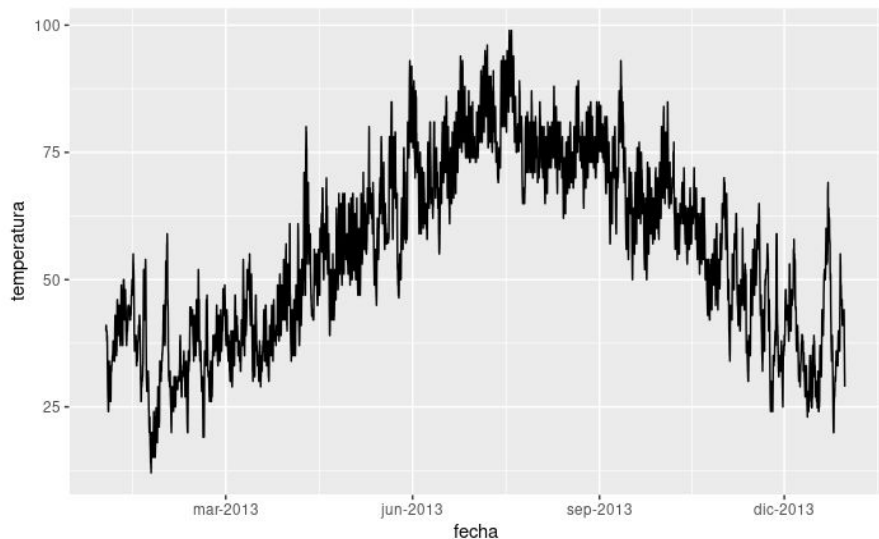
```
clima %>%  
  mutate(fecha = lubridate::make_datetime(anio, mes, dia, hora)) %>%  
  filter(origen == "LGA" ) %>%  
  ggplot(aes(fecha, temperatura)) +  
  geom_line()
```


Escalas para fechas y horas



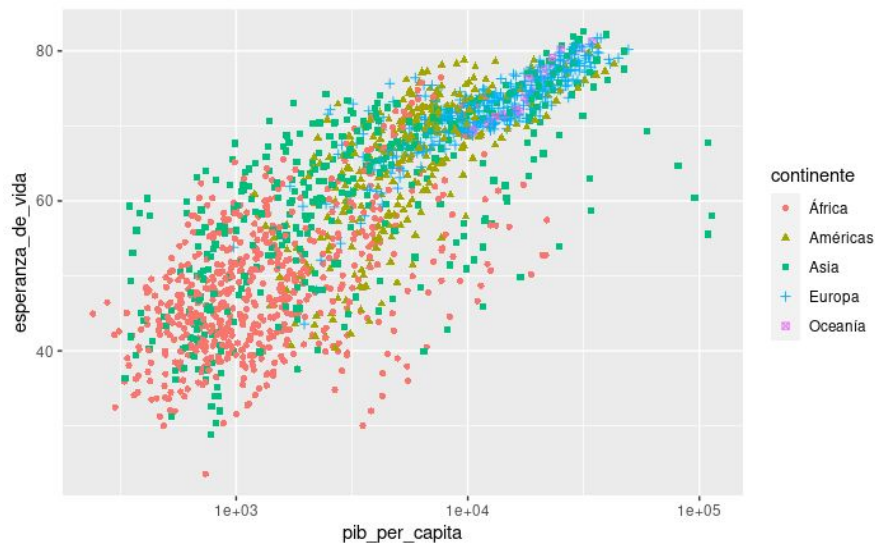
```
clima %>%  
  mutate(fecha = lubridate::make_datetime(anio, mes, dia, hora)) %>%  
  filter(origen == "LGA" ) %>%  
  ggplot(aes(fecha, temperatura)) +  
  geom_line() +  
  scale_x_datetime(date_breaks = "3 months")
```

Escalas para fechas y horas



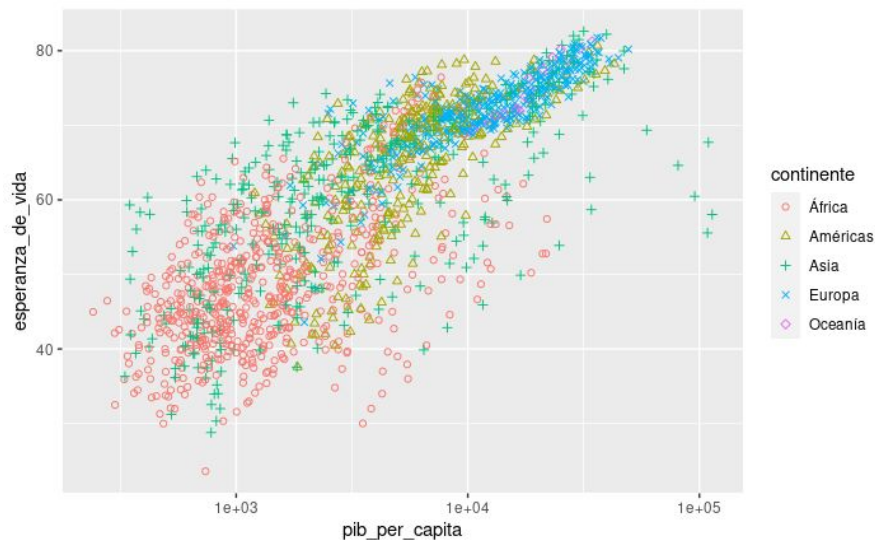
```
mutate(fecha = lubridate::make_datetime(ano, mes, dia, hora)) %>%  
filter(origen == "LGA" ) %>%  
ggplot(aes(fecha, temperatura)) +  
geom_line() +  
scale_x_datetime(date_breaks = "3 months",  
                  date_labels = "%b-%Y")
```

Escalas para shapes/linetype



```
ggplot(países, aes(pib_per_capita, esperanza_de_vida)) +  
  geom_point(aes(color = continente, shape = continente)) +  
  scale_x_log10()
```

Escalas (manuales) para shapes/linetype



```
ggplot(países, aes(pib_per_capita, esperanza_de_vida)) +  
  geom_point(aes(color = continente, shape = continente)) +  
  scale_x_log10() +  
  scale_shape_manual(values = c(1, 2, 3, 4, 5))
```

Nos vemos la semana que viene

Estén atentos al campus