InteRculturales
Introducción a técnicas de ciencias sociales computacionales

Centro de Estudios Interculturales e Indígenas
Sesión 2 - 2025

@matdknu | matdknu.github.io

Recapitulando... 

CIIR
Center for Intercultural
and Indigenous Research

Tidyverse

Conjunto de paquetes de R diseñados para ciencia de datos. Incluye herramientas como ggplot2, dplyr, tidyr, readr, entre otros, que comparten una sintaxis coherente y principios comunes (funciones pipe (`|>` o `%>%`) y manipulación de datos "ordenados").

```
head(starwars)
```

```
names(starwars)
```

```
##  [1] "name"       "height"     "mass"       "hair_color" "skin_color"
##  [6] "eye_color"  "birth_year" "sex"        "gender"     "homeworld"
## [11] "species"    "films"      "vehicles"   "starships"
```

Recapitulando... 

CIIR
Center for Intercultural
and Indigenous Research

Tidyverse: `Select`

Sirve para elegir columnas específicas de un data frame.

```
starwars %>%
  select(name, species, height, mass) |> head(3)
```

```
## # A tibble: 3 × 4
##   name           species height  mass
##   <chr>          <chr>    <int> <dbl>
## 1 Luke Skywalker Human      172    77
## 2 C-3PO          Droid      167    75
## 3 R2-D2          Droid       96    32
```

Recapitulando...

CIIR
Center for Intercultural
and Indigenous Research

Tidyverse: `Filter`

Permite filtrar filas según condiciones lógicas.

```
starwars %>% filter(species == "Human", mass > 80)
```

```
## # A tibble: 6 × 14
##   name      height  mass hair_color skin_color eye_color birth_year sex   gender
##   <chr>      <int> <dbl> <chr>      <chr>      <chr>           <dbl> <chr> <chr>
## 1 Darth Va…    202   136 none       white      yellow           41.9 male  mascu…
## 2 Owen Lars    178   120 brown, gr… light      blue             52   male  mascu…
## 3 Biggs Da…    183    84 black      light      brown            24   male  mascu…
## 4 Anakin S…    188    84 blond      fair       blue             41.9 male  mascu…
## 5 Qui-Gon …    193    89 brown      fair       blue             92   male  mascu…
## 6 Mace Win…    188    84 none       dark       brown            72   male  mascu…
## # ℹ 5 more variables: homeworld <chr>, species <chr>, films <list>,
## #   vehicles <list>, starships <list>
```

Recapitulando... 

CIIR
Center for Intercultural
and Indigenous Research

Tidyverse: `Mutate`

Crea nuevas columnas o modifica variables existentes.

```
starwars %>%
  mutate(imc_ficticio = mass / (height / 100)^2) %>%
  select(name, imc_ficticio) %>%
  arrange(desc(imc_ficticio)) |>
  head(3)
```

```
## # A tibble: 3 × 2
##   name                  imc_ficticio
##   <chr>                        <dbl>
## 1 Jabba Desilijic Tiure        443.
## 2 Dud Bolt                      50.9
## 3 Yoda                          39.0
```

Recapitulando...📝

CIIR
Center for Intercultural
and Indigenous Research

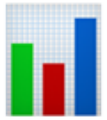
Tidyverse: `Group_by` + `summarise`

`Group_by`: Crea nuevas columnas o modifica variables existentes y `Summarise`: Genera resúmenes estadísticos por grupo (si se usó group_by()) o para todo el conjunto.

```
starwars %>% group_by(species) %>%
  summarise(
    promedio_altura = mean(height, na.rm = TRUE)
  ) %>%
  arrange(desc(promedio_altura)) |>  # Ordenar por número de personajes
  head(3)
```

```
## # A tibble: 3 × 2
##   species  promedio_altura
##   <chr>              <dbl>
## 1 Quermian             264
## 2 Wookiee              231
## 3 Kaminoan             221
```

CIIR
Center for Intercultural
and Indigenous Research

Hoy veremos... 

- ¿Qué es `ggplot2` y cómo funciona?
- Visualizaciones básicas: puntos, densidades, barras
- Transformación de datos antes de graficar
- Facetas y estética
- Guardar gráficos con `ggsave()`
- Buenas prácticas en visualización

ggplot2: un nueva forma de pensar y visualizar datos

CIIR
Center for Intercultural
and Indigenous Research

`tidyr` permite:

- "Traduce" datos en elementos visuales

Bases de datos ordenadas ("tidy")

CIIR
Center for Intercultural
and Indigenous Research

- El punto de partida de un gráfico en `ggplot` es una base de datos "tidy".
- Si los datos no existen en el formato necesario para visualizarlos, necesitamos primero "darles forma".

Bases de datos ordenadas ("tidy")

CIIR
Center for Intercultural
and Indigenous Research

```
datos <- data.frame(
  region = c("Norte", "Centro", "Sur"),
  hombres = c(50000, 70000, 60000),
  mujeres = c(52000, 68000, 63000)
)

datos
```

```
##   region hombres mujeres
## 1  Norte   50000   52000
## 2 Centro   70000   68000
## 3    Sur   60000   63000
```

Bases de datos ordenadas ("tidy")

CIIR
Center for Intercultural
and Indigenous Research

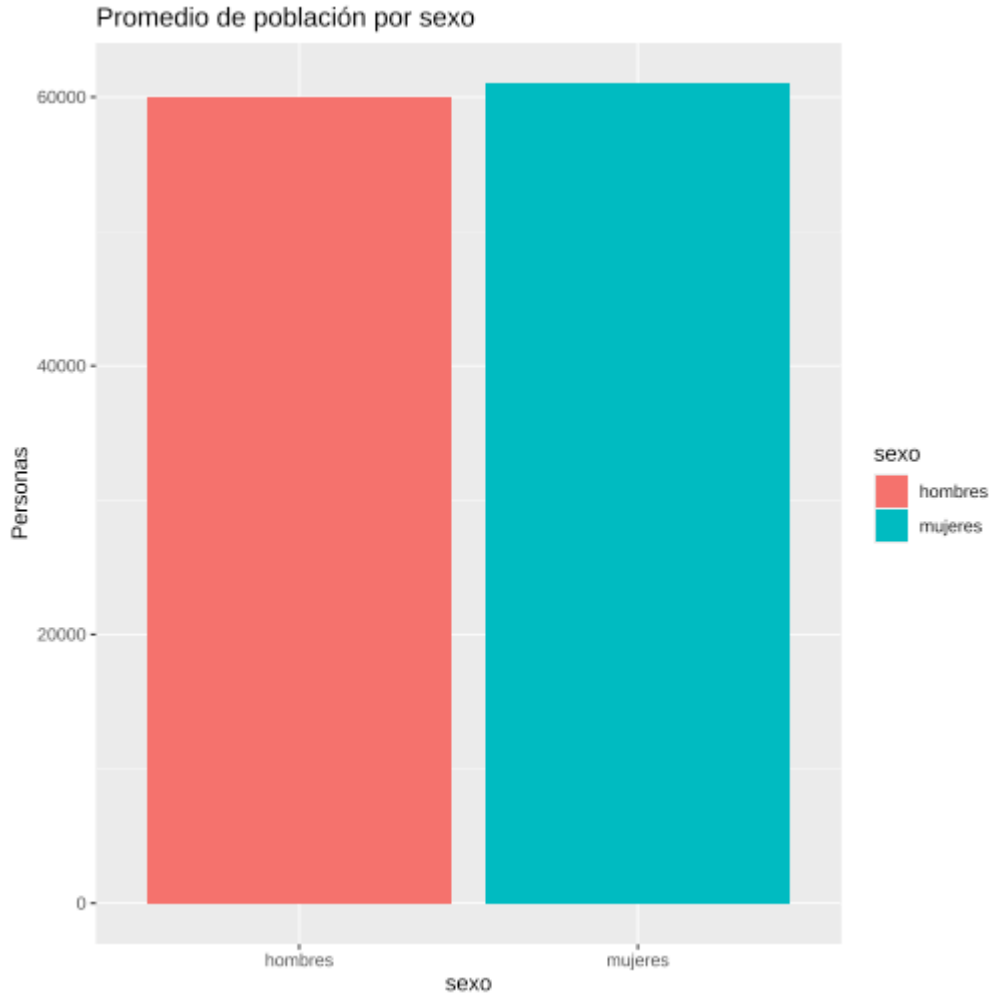
```
library(tidyverse)

resumen <- datos %>%
  summarise(
    hombres = mean(hombres),
    mujeres = mean(mujeres)
  ) %>%
  t() %>%                          # transponer
  as.data.frame() %>%
  mutate(sexo = rownames(.),    # pasar nombres de fila a columna
         promedio = V1) %>%
  select(sexo, promedio); resumen
```

```
##            sexo promedio
## hombres hombres    60000
## mujeres mujeres    61000
```

Bases de datos ordenadas ("tidy")

CIIR
Center for Intercultural
and Indigenous Research

```
library(ggplot2)

resumen <- data.frame(
  sexo = c("hombres", "mujeres"),
  promedio = c(60000, 61000)
)

g1 <- ggplot(resumen,
             aes(x = sexo,
                 y = promedio,
                 fill = sexo)) +
  geom_col() +
  labs(title = "Promedio de población
       y = "Personas")
```



Sobre Ggplot2

CIIR
Center for Intercultural
and Indigenous Research

{ggplot2} es una librería de visualización de datos bastante popular en el mundo de la ciencia de datos. Sus principales características son su atractivo, su conveniencia para la exploración de datos, un gran potencial de personalización, y un extenso ecosistema de extensiones que nos permiten generar visualizaciones prácticamente de cualquier tipo

Sistema de capas en `ggplot2`

CIIR
Center for Intercultural
and Indigenous Research

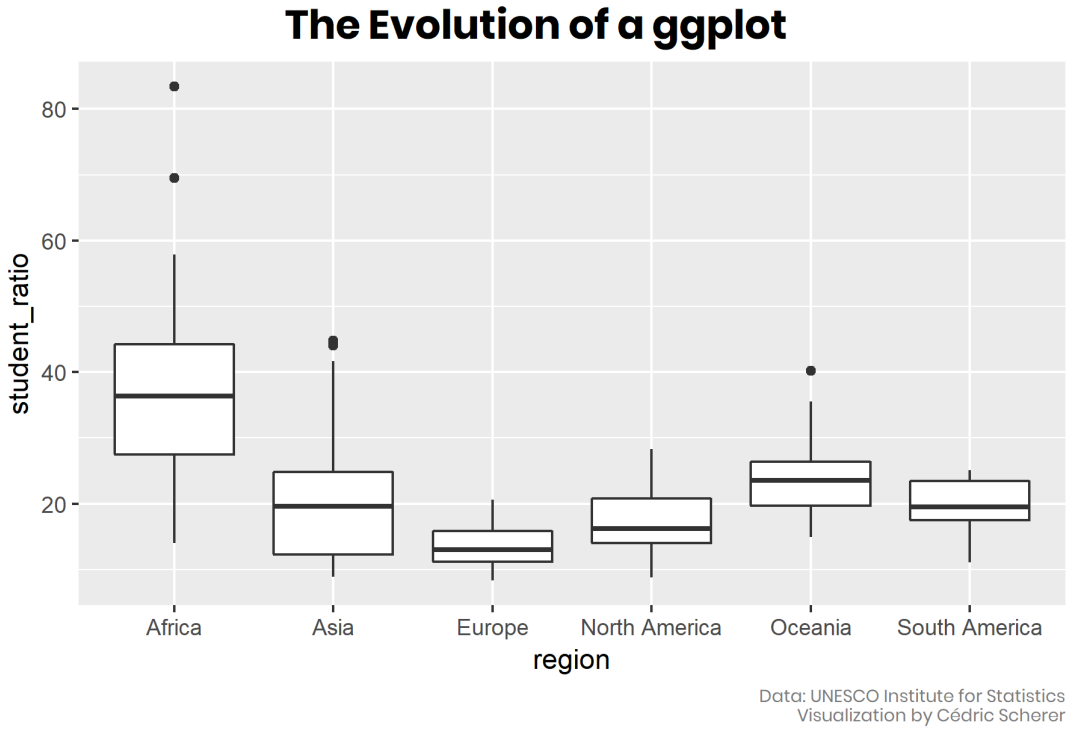
La librería **{ggplot2}** permite crear gráficos sumando capas. Cada capa cumple una función específica y se puede agregar según lo que se quiera comunicar o refinar.

Capas principales:

- **Datos**: base a graficar.
- **Estéticas (`aes()`)**: asigna variables a ejes, color, forma, etc.
- **Geometrías (`geom_`)**: define el tipo de gráfico (puntos, líneas, barras...).
- **Escalas (`scale_`)**: ajusta rangos, paletas, límites.
- **Coordenadas (`coord_`)**: define el sistema de ejes y límites espaciales.
- **Facetas (`facet_`)**: divide datos en subgráficos según una variable.
- **Temas (`theme()`)**: controla apariencia visual (texto, fondo, grillas...).

CIIR
Center for Intercultural
and Indigenous Research

📦 Recursos esenciales para trabajar con `ggplot2`



📚 Documentación oficial y Cheatsheets

CIIR
Center for Intercultural
and Indigenous Research

1. 🔗 ggplot2.tidyverse.org

- Sitio oficial del paquete `ggplot2`.
- Contiene documentación completa, funciones ordenadas por categoría, ejemplos y novedades del desarrollo.

1. 📄 Cheatsheet oficial de ggplot2 (PDF)

- Publicado por RStudio (ahora Posit).
- Muy útil para tener a mano todas las funciones esenciales y su sintaxis.
- Descargable en PDF.

🎨 Galerías de gráficos y ejemplos prácticos

CIIR
Center for Intercultural
and Indigenous Research

1. 🖼️ The R Graph Gallery

- Gran repositorio de gráficos creados con `ggplot2`.
- Incluye código, ejemplos con datos simulados y personalizaciones.
- Organizado por tipo de gráfico (líneas, barras, mapas, etc.)

1. 📊 Data to Viz

- Recurso que ayuda a elegir el tipo de gráfico según tu tipo de variable (categórica, continua, etc.)
- Cada recomendación incluye un ejemplo hecho con `ggplot2`.
- Gran puente entre teoría visual y aplicación práctica.

🛠️ Personalización de gráficos y temas visuales

🎨 ggthemes (temas pre-hechos)

- Reproduce estilos como Wall Street Journal, The Economist, FiveThirtyEight, Excel y más.
- Permite una personalización estética muy rápida.

🖌️ BBC style plot – bbplot

- Tema y funciones auxiliares creadas por la BBC.
- Genera gráficos que cumplen con sus estándares editoriales (claridad, color, tipografía).

Vamos al Ggplot2

CIIR
Center for Intercultural
and Indigenous Research

```
library(ggplot2)
library(palmerpenguins)
```

Vamos al Ggplot2: Ejemplo vacío

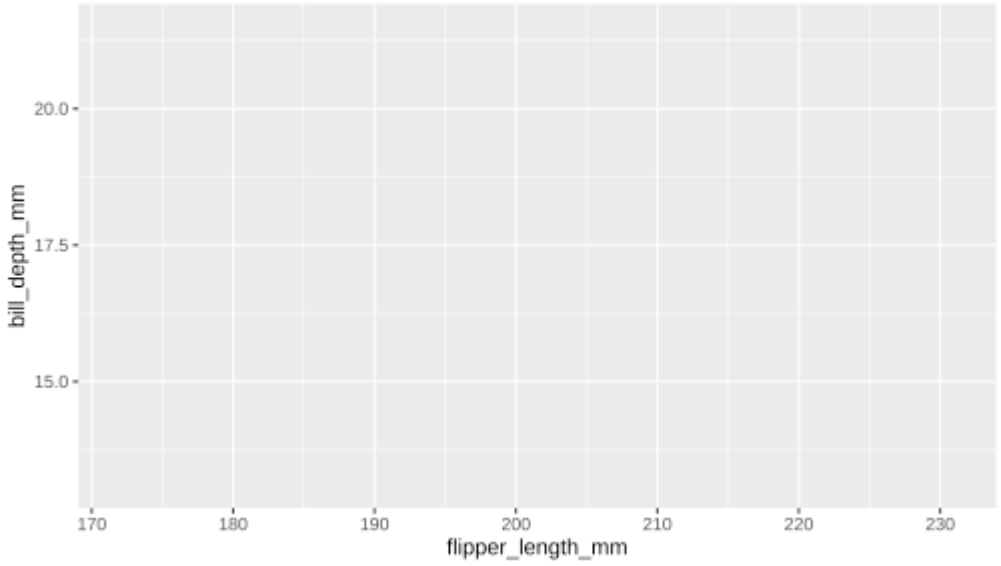
CIIR
Center for Intercultural
and Indigenous Research

```
penguins |> ggplot()
```

Vamos al Ggplot2: Pongamos lo ejes

CIIR
Center for Intercultural
and Indigenous Research

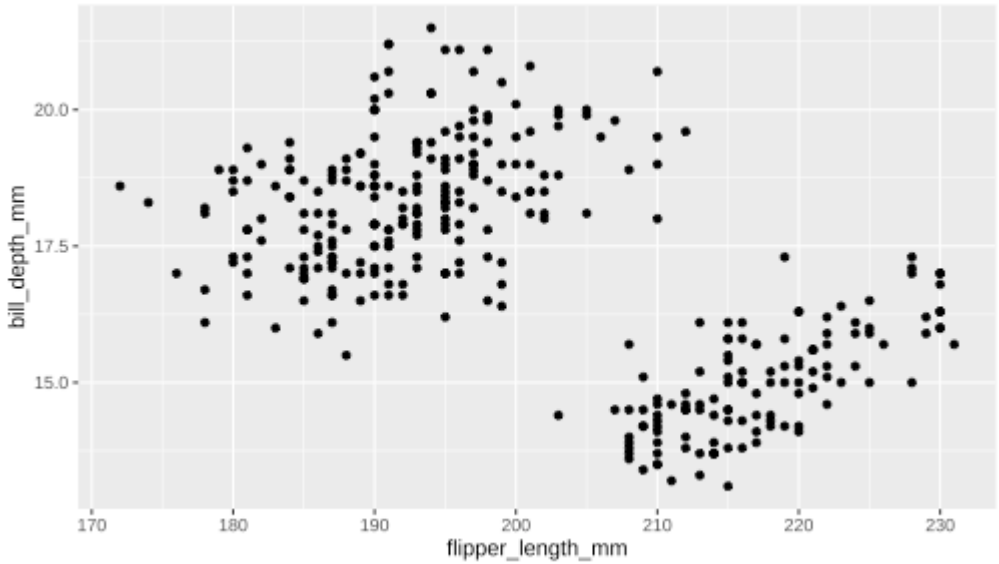
```
penguins |>
  ggplot() + # iniciar el gráfico
  # definir el mapeo de variables a características estéticas del gráfico
  aes(x = flipper_length_mm , # eje x (horizontal)
      y = bill_depth_mm)
```



Vamos al Ggplot2: Definimos los puntos

CIIR
Center for Intercultural
and Indigenous Research

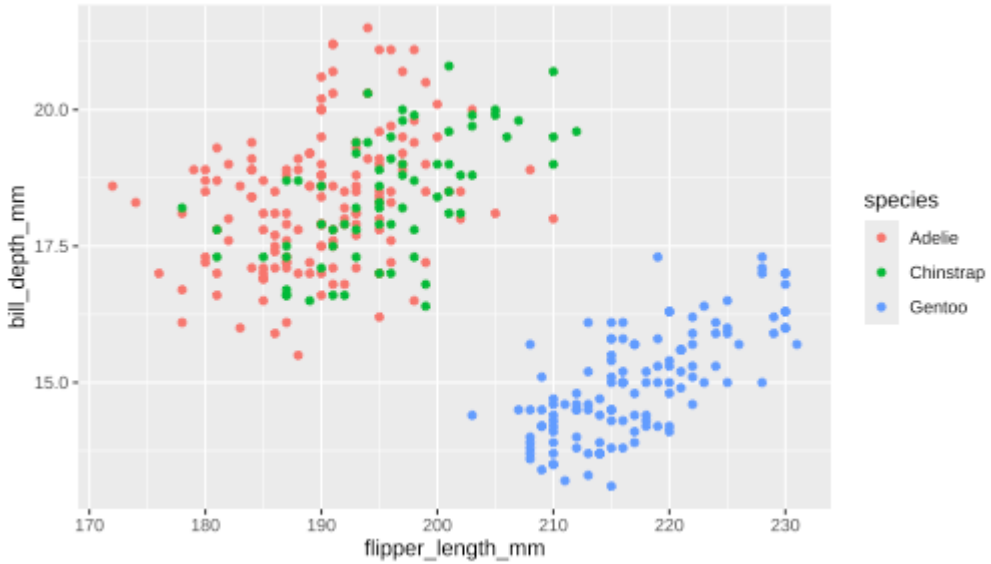
```
penguins |>
  ggplot() + # iniciar el gráfico
  # definir el mapeo de variables a características estéticas del gráfico
  aes(x = flipper_length_mm , # eje x (horizontal)
      y = bill_depth_mm) + # eje y (vertical)
  # agregar una capa de geometría
  geom_point() # geometría de puntos
```



Vamos al Ggplot2: Definimos los puntos y el color

CIIR
Center for Intercultural
and Indigenous Research

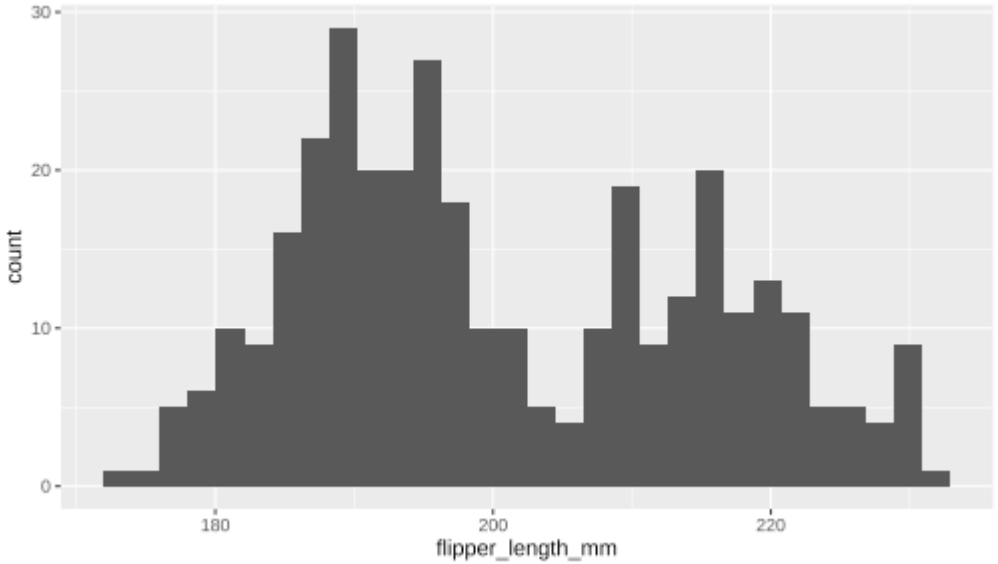
```
penguins |>
  ggplot() + # iniciar el gráfico
  # definir el mapeo de variables a características estéticas del gráfico
  aes(x = flipper_length_mm , # eje x (horizontal)
      y = bill_depth_mm,  # eje y (vertical)
      color = species) + # color por especie
  # agregar una capa de geometría
  geom_point() # geometría de puntos
```



Vamos al Ggplot2: Histograma

CIIR
Center for Intercultural
and Indigenous Research

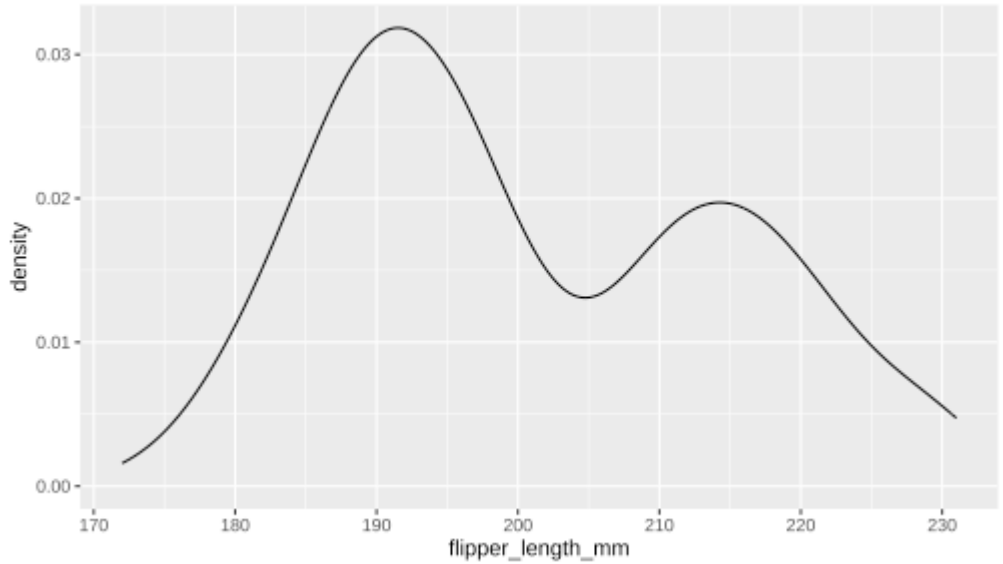
```
penguins  |> # datos
  ggplot() + # iniciar
  aes(x = flipper_length_mm) + # variable horizontal
  geom_histogram() # histograma
```



Vamos al Ggplot2: Densidad

CIIR
Center for Intercultural
and Indigenous Research

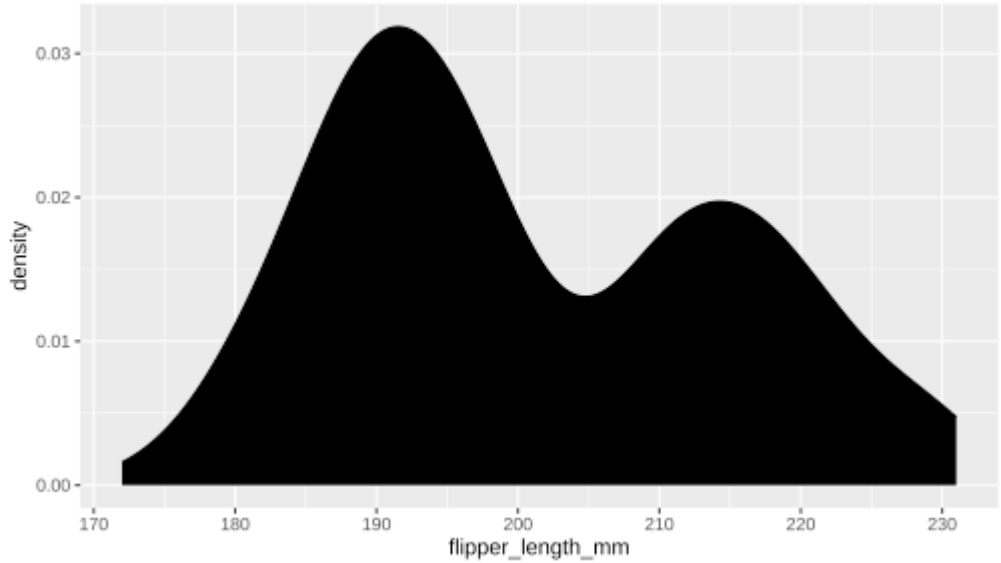
```
penguins  |> # datos
  ggplot() + # iniciar
  aes(x = flipper_length_mm) + # variable horizontal
  geom_density() # histograma
```



Vamos al Ggplot2: Densidad y fondo

CIIR
Center for Intercultural
and Indigenous Research

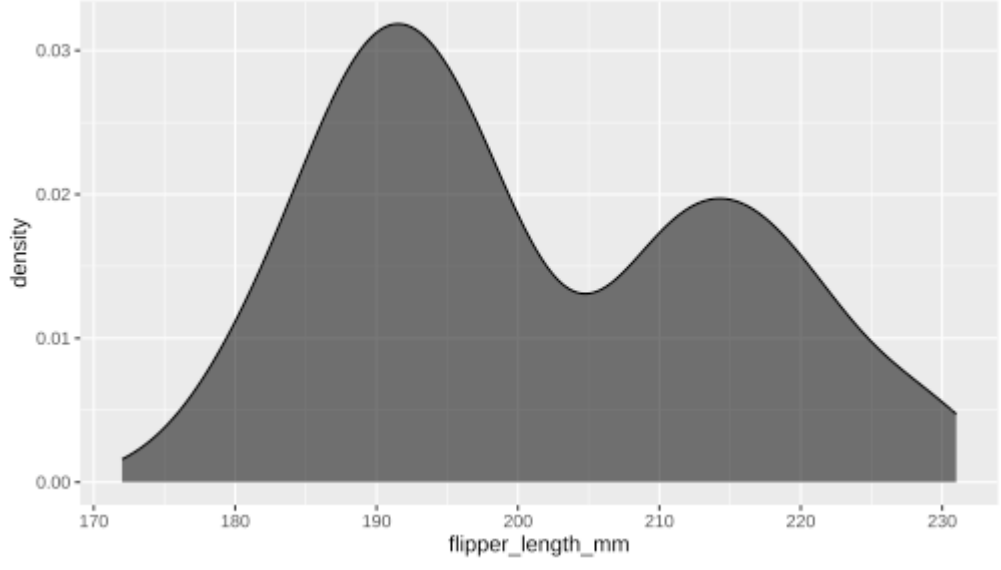
```
penguins  |> # datos
  ggplot() + # iniciar
  aes(x = flipper_length_mm) + # variable horizontal
   geom_density(fill = "black")
```



Vamos al Ggplot2: Densidad y fondo

CIIR
Center for Intercultural
and Indigenous Research

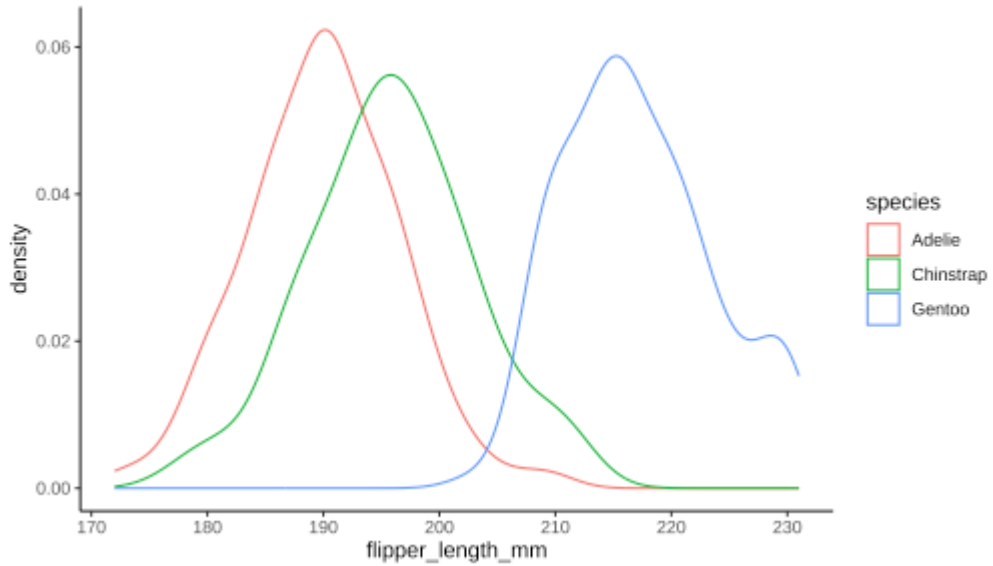
```
penguins  |> # datos
  ggplot() + # iniciar
  aes(x = flipper_length_mm) + # variable horizontal
   geom_density(fill = "black", alpha = 0.6)
```



Vamos al Ggplot2: Densidad y colores por species

CIIR
Center for Intercultural
and Indigenous Research

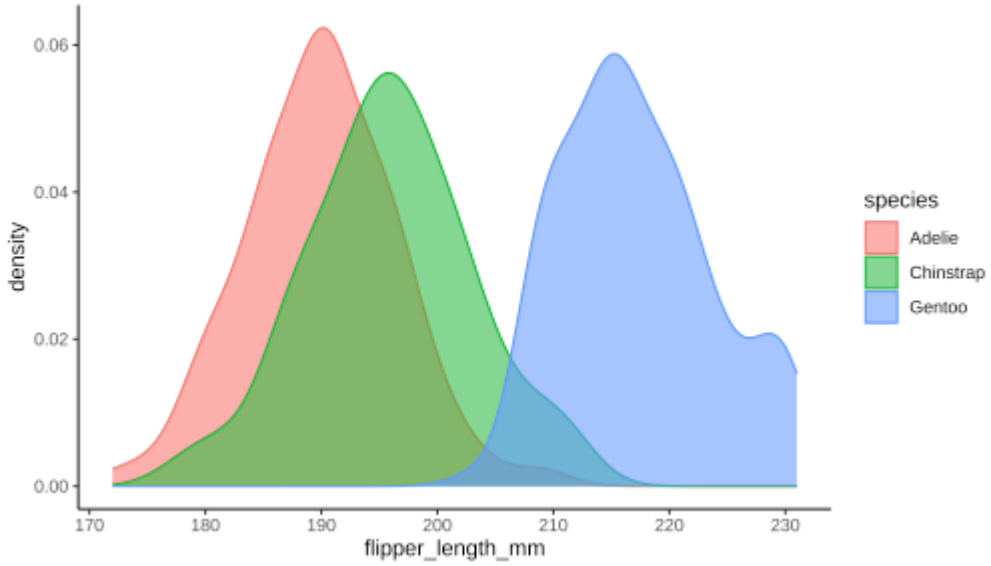
```
penguins |>
  ggplot() +
  aes(x = flipper_length_mm,
      color = species) + # bordes de la figura
  geom_density() +
  # tema
  theme_classic()
```



Vamos al Ggplot2: Densidad y colores por species

CIIR
Center for Intercultural
and Indigenous Research

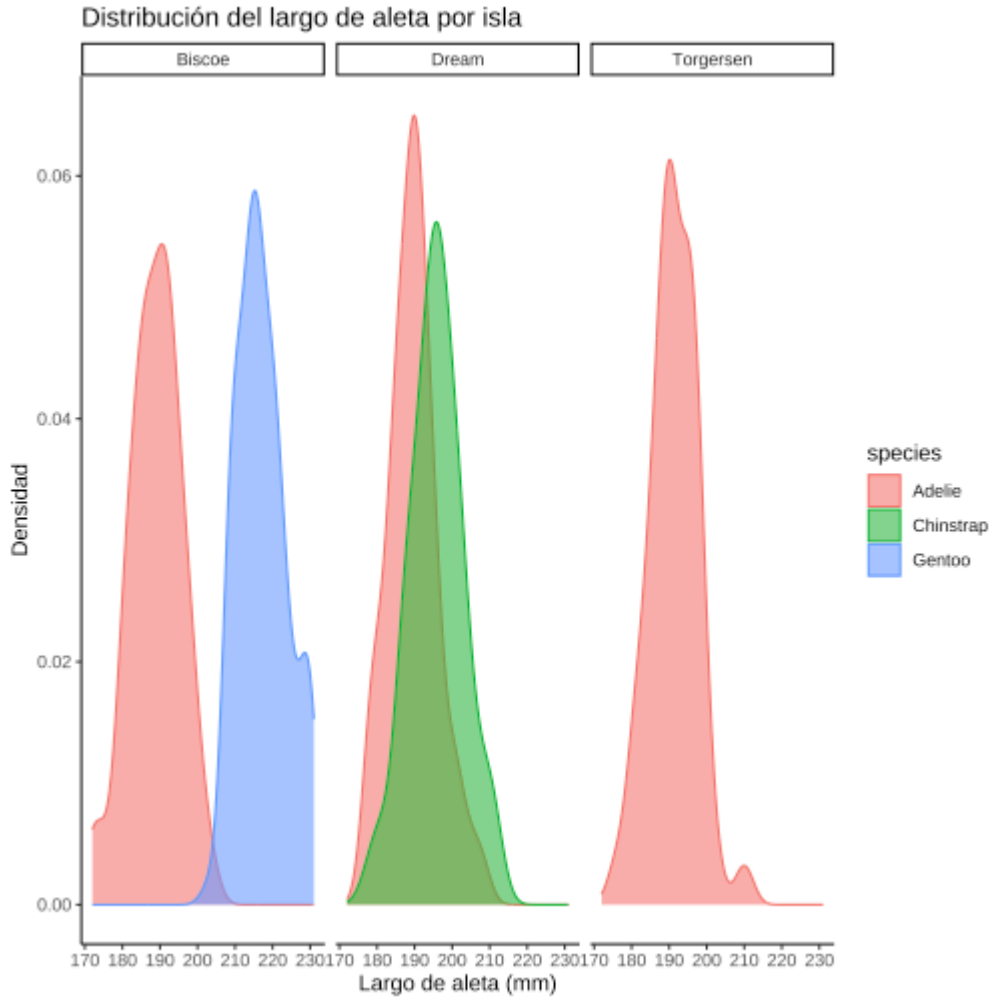
```
penguins |>
  ggplot() +
  aes(x = flipper_length_mm,
      fill = species, # relleno de la figura
      color = species) + # bordes de la figura
  geom_density(alpha = 0.6) +
  # tema
  theme_classic()
```



Vamos al Ggplot2: Densidad y colores por species

CIIR
Center for Intercultural
and Indigenous Research

```
g5 <-penguins |>
  ggplot() +
  aes(x = flipper_length_mm,
      fill = species,
      color = species) +
  geom_density(alpha = 0.6) +
  facet_wrap(~ island) +  # 👈 Facet
  theme_classic() +
  labs(
    title = "Distribución del largo d
    x = "Largo de aleta (mm)",
    y = "Densidad"
  )
```



Vamos al Ggplot2: Retomemos lo Tidy

CIIR
Center for Intercultural
and Indigenous Research

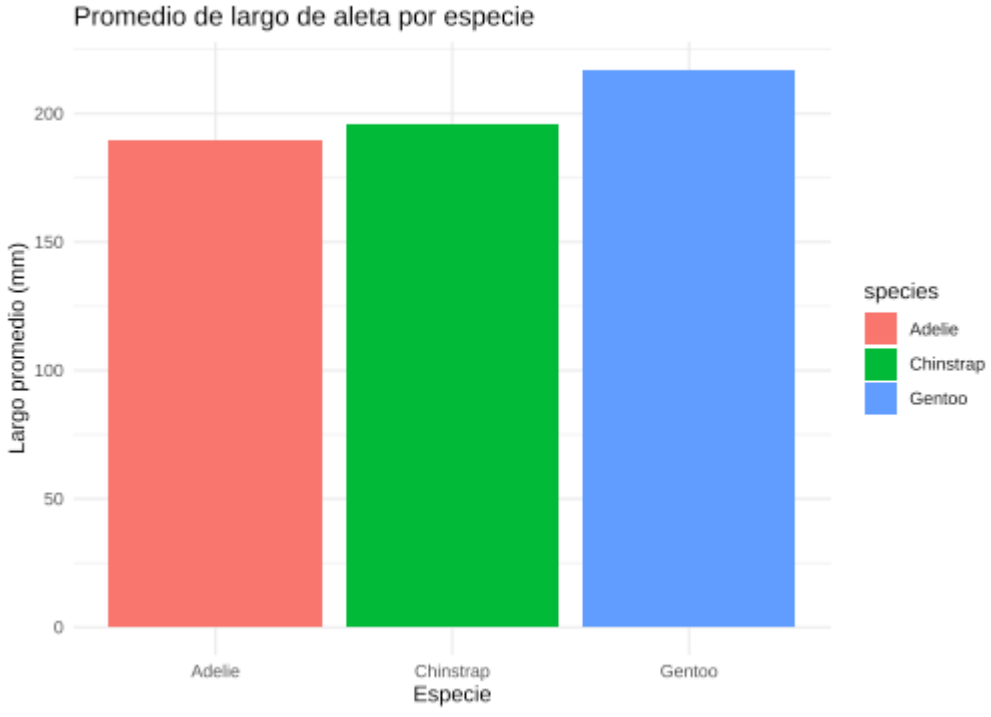
```
# Transformar: seleccionar columnas relevantes, eliminar NA, y resumir
resumen <- penguins %>%
  select(species, flipper_length_mm) %>%
  filter(!is.na(flipper_length_mm)) %>%
  group_by(species) %>%
  summarise(promedio_aleta = mean(flipper_length_mm))

resumen
```

```
## # A tibble: 3 × 2
##   species   promedio_aleta
##   <fct>              <dbl>
## 1 Adelie              190.
## 2 Chinstrap           196.
## 3 Gentoo              217.
```

Vamos al Ggplot2: Retomemos lo Tidy

CIIR
Center for Intercultural
and Indigenous Research

```
ggplot(resumen, aes(x = species, y = promedio_aleta, fill = species)) +
  geom_col() +
  labs(title = "Promedio de largo de aleta por especie",
       x = "Especie",
       y = "Largo promedio (mm)") +
  theme_minimal()
```



Vamos al Ggplot2: Hagamos algunos ajustes

CIIR
Center for Intercultural
and Indigenous Research

```
ggplot(resumen, aes(x = species, y = promedio_aleta, fill = species)) +
  geom_col() +
  geom_text(aes(label = round(promedio_aleta, 1)), vjust = -0.5) +  # 👈 etiquet
  labs(title = "Promedio de largo de aleta por especie",
       x = "Especie",
       y = "Largo promedio (mm)") +
  theme_minimal()
```

Ajustes

CIIR
Center for Intercultural
and Indigenous Research

En ggplot2, el uso de theme() permite personalizar la apariencia visual del gráfico, modificando elementos como:

Texto: tamaño, color, fuente, posición (centrado, cursiva, negrita).

1. Títulos y subtítulos.
2. Fondos.
3. Posición de leyendas.
4. Márgenes, ejes, etiquetas, bordes, etc.

Ajustes

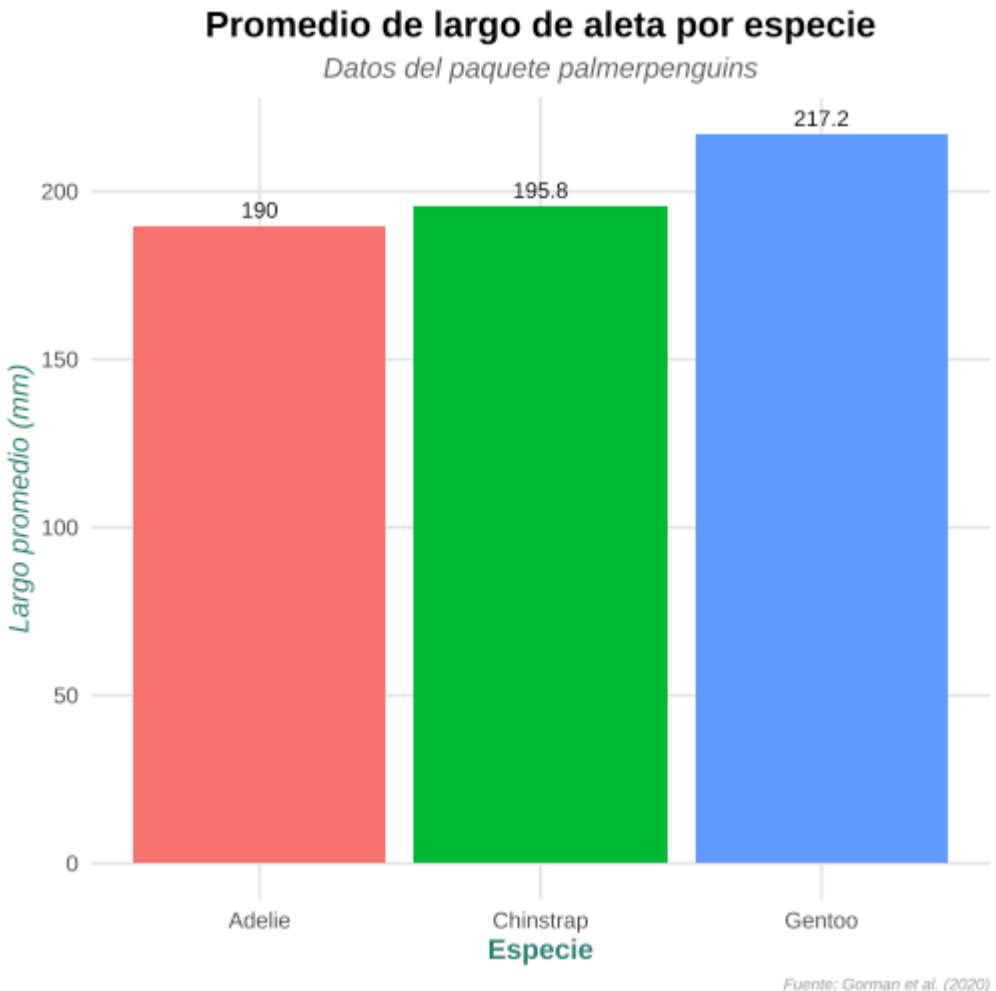
CIIR
Center for Intercultural
and Indigenous Research

```
ajustes <-resumen |>
ggplot(aes(x = species, y = promedio_aleta, fill = species)) +
  geom_col() +
  geom_text(aes(label = round(promedio_aleta, 1)), vjust = -0.5) +
  labs(
    title = "Promedio de largo de aleta por especie",
    subtitle = "Datos del paquete palmerpenguins",
    caption = "Fuente: Gorman et al. (2020)",
    x = "Especie",
    y = "Largo promedio (mm)"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 18),       # Ce
    plot.subtitle = element_text(hjust = 0.5, face = "italic", color = "gray40")
    axis.title.x = element_text(face = "bold", color = "#2c8475"),
    axis.title.y = element_text(face = "italic", color = "#2c8475"),
    legend.position = "none",   # Ocultar leyenda
    plot.caption = element_text(size = 8, hjust = 1, face = "italic", color = "g
    panel.grid.major = element_line(color = "gray90"),
    panel.grid.minor = element_blank()
  )
```

Resultados de los Ajustes

CIIR
Center for Intercultural
and Indigenous Research

```
ajustes
```



Retomemos lo Tidy

CIIR
Center for Intercultural
and Indigenous Research

```
penguins |> select(year, species, flipper_length_mm) |> head(1)
```

```
## # A tibble: 1 × 3
##    year species flipper_length_mm
##   <int> <fct>                <int>
## 1  2007 Adelie                 181
```

```
penguins %>%
  filter(!is.na(flipper_length_mm)) %>%
  group_by(species, year) %>%
  summarise(promedio_aleta = mean(flipper_length_mm))
```

Table: Promedio de largo de aleta por especie y año

species	year	media
Adelie	2007	186.5918
Adelie	2008	191.0400
Adelie	2009	192.0769

Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g6 <- tabla_pinguino |>
  ggplot(aes(x = year,
             y = promedio_aleta,
             color = species)) +
  geom_line(size = 1.2)
```

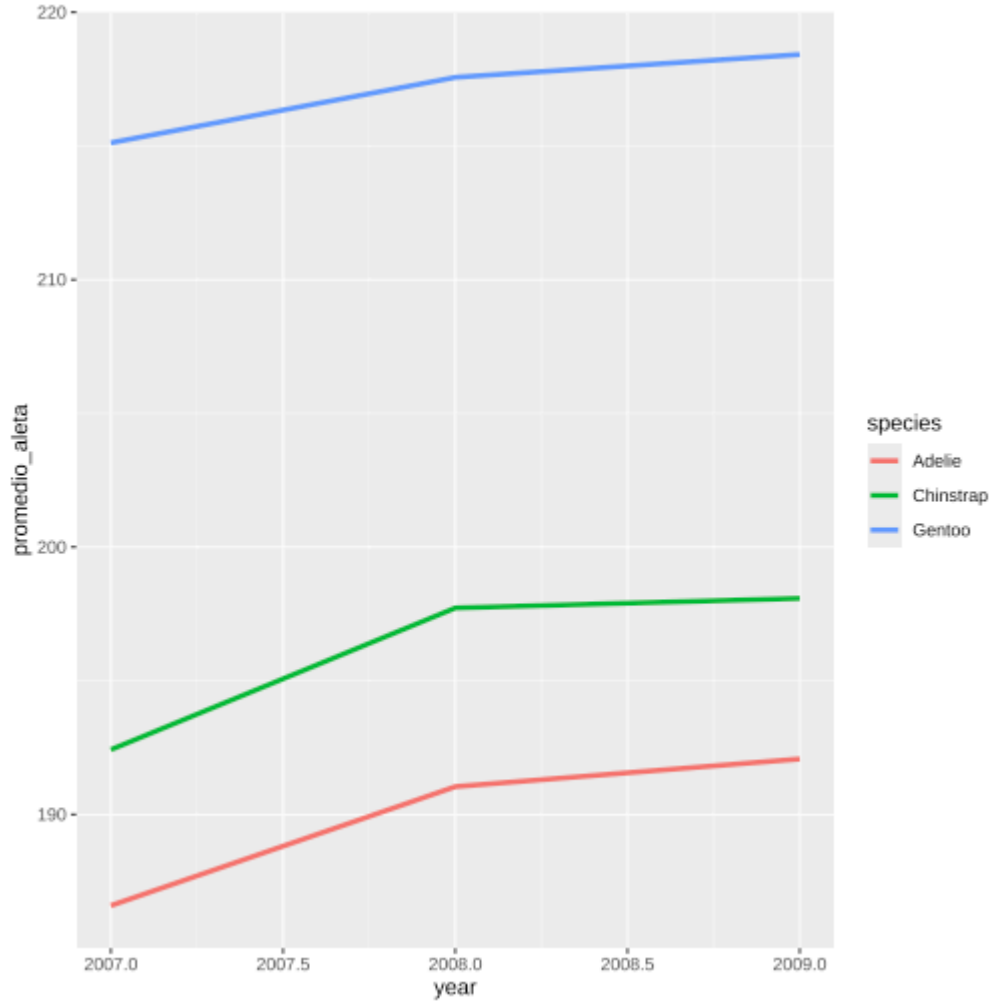


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g7 <- tabla_pinguino |>
  ggplot(aes(x = year,
             y = promedio_aleta,
             color = species)) +
  geom_line(size = 1.2) +
  geom_point(size = 2)
```

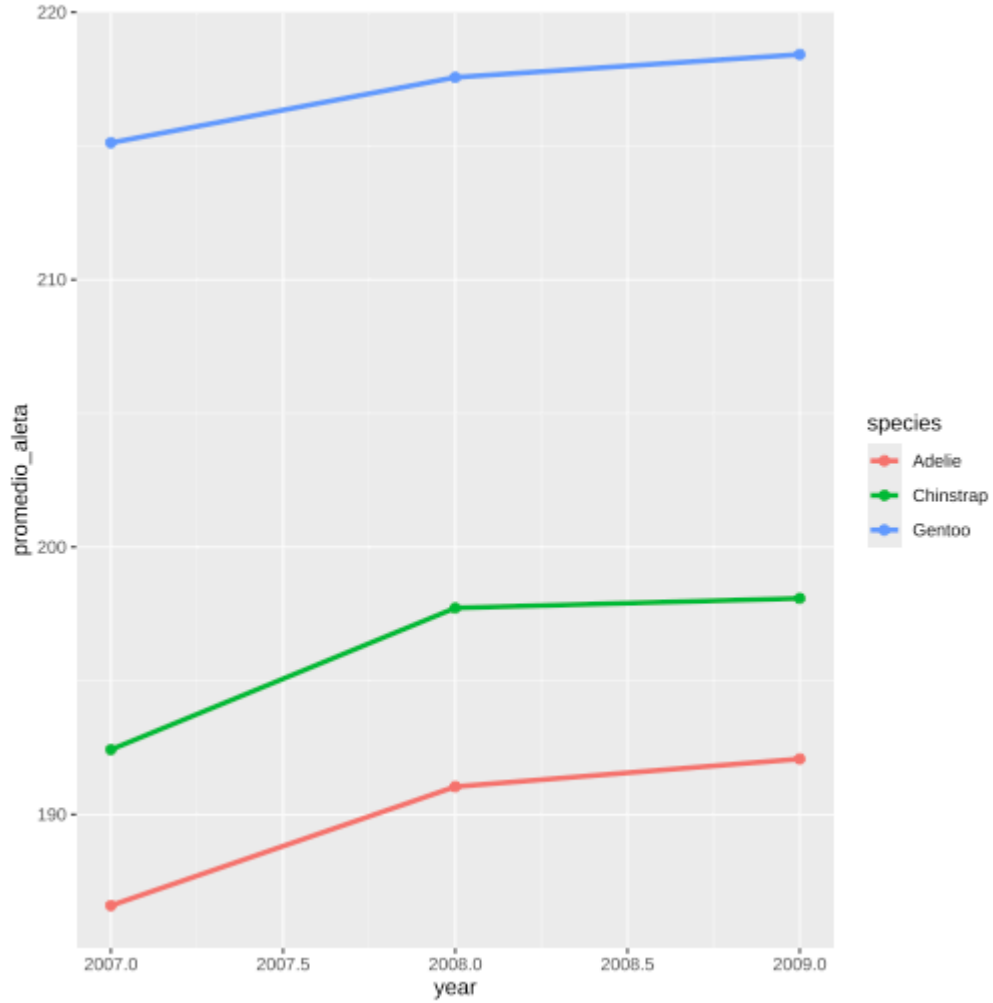


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g8 <- tabla_pinguino |>
  ggplot(aes(x = year,
             y = promedio_aleta,
             color = species)) +
  geom_line(size = 1.2) +
  geom_point(size = 2) +
  labs(
    title = "Tendencia del largo de a
    x = "Año",
    y = "Largo promedio de aleta (mm)
    color = "Especie"
  )
```

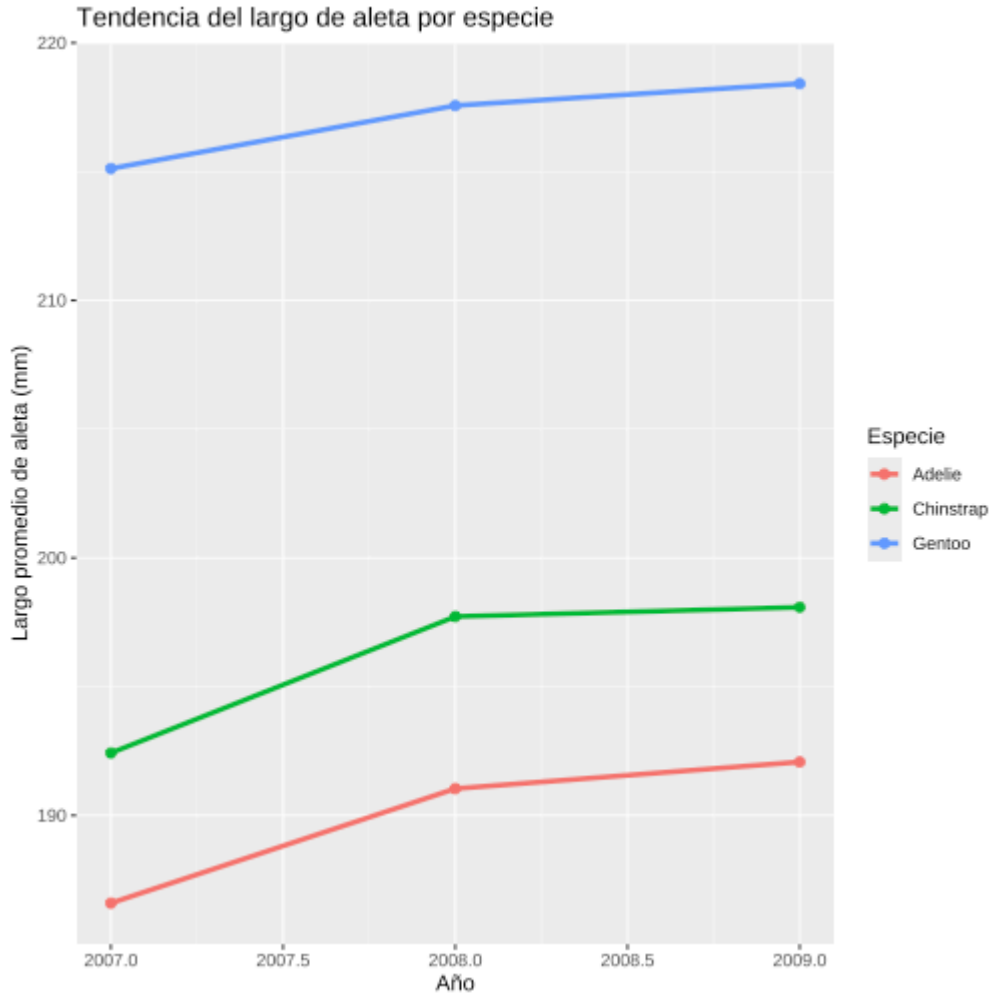


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g9 <-tabla_pinguino |>
  mutate(year = as.numeric(year)) |>
  ggplot(aes(year, # identificación a
             promedio_aleta,# identif
             color = species)) +
  geom_line(size = 1) +
  geom_point(size = 1.5) +
  scale_x_continuous(breaks =
  unique(tabla_pinguino$year)) +  # 
  labs(
    title = "Tendencia del largo de
    aleta por especie",
    x = "Año",
    y = "Largo promedio (mm)",
    color = "Especie"
  )
```

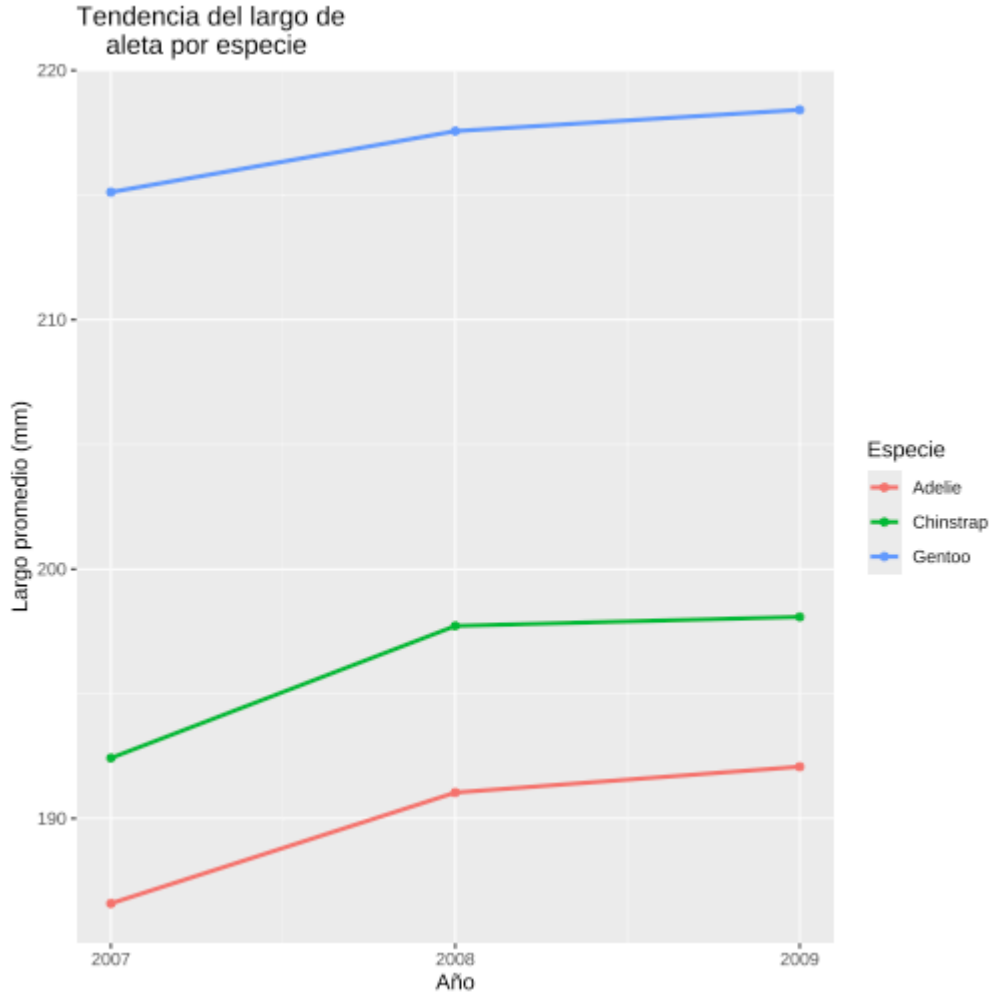


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g10 <-tabla_pinguino |>
  mutate(year = as.numeric(year)) |>
  ggplot(aes(year, # identificación a
             promedio_aleta,# identif
             color = species)) +
  geom_line(size = 1) +
  geom_point(size = 1.5) +
  scale_x_continuous(breaks =
  unique(tabla_pinguino$year)) +  # 
  labs(
    title = "Tendencia del largo de
    aleta por especie",
    x = "Año",
    y = "Largo promedio (mm)",
    color = "Especie"
  ) +
  theme_bw()
```

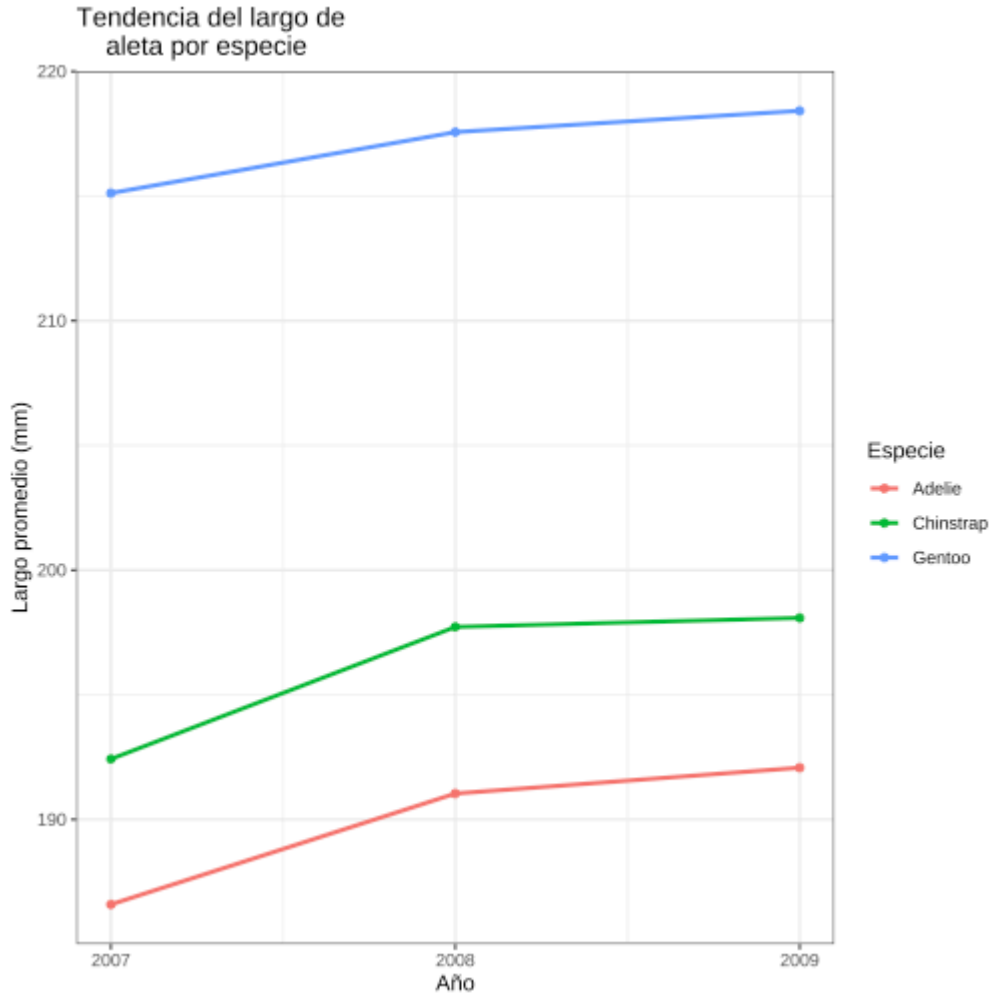


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g12 <-tabla_pinguino |>
  mutate(year = as.numeric(year)) |>
  ggplot(aes(year, # identificación a
             promedio_aleta,# identif
             color = species)) +
  geom_line(size = 1) +
  geom_point(size = 1.5) +
  scale_x_continuous(breaks =
  unique(tabla_pinguino$year)) +  # 
  labs(
    title = "Tendencia del largo de
    aleta por especie",
    x = "Año",
    y = "Largo promedio (mm)",
    color = "Especie"
  ) +
  theme_classic()
```

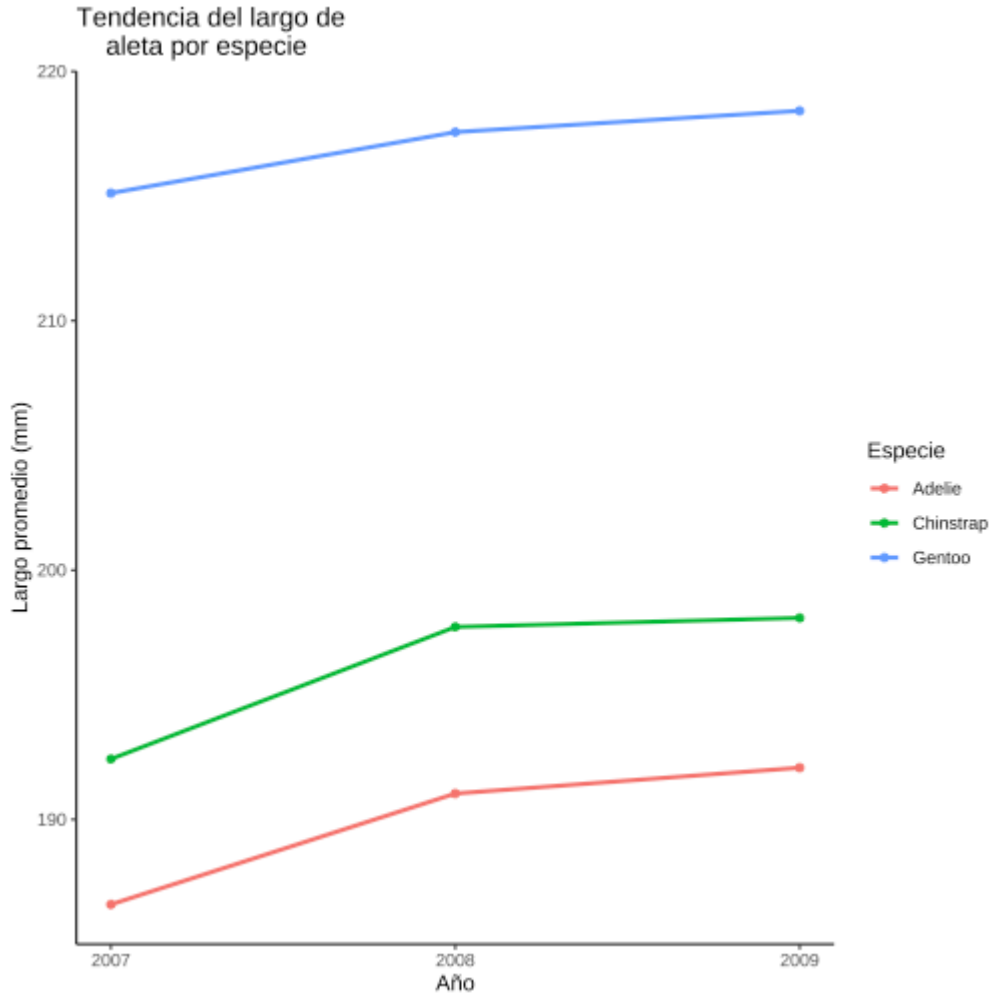


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

```
g13 <-tabla_pinguino |>
  mutate(year = as.numeric(year)) |>
  ggplot(aes(year, # identificación a
             promedio_aleta,# identif
             color = species)) +
  geom_line(size = 1) +
  geom_point(size = 1.5) +
  scale_x_continuous(breaks =
  unique(tabla_pinguino$year)) +  # 
  labs(
    title = "Tendencia del largo de
    aleta por especie",
    x = "Año",
    y = "Largo promedio (mm)",
    color = "Especie"
  ) +
  theme_minimal()
```

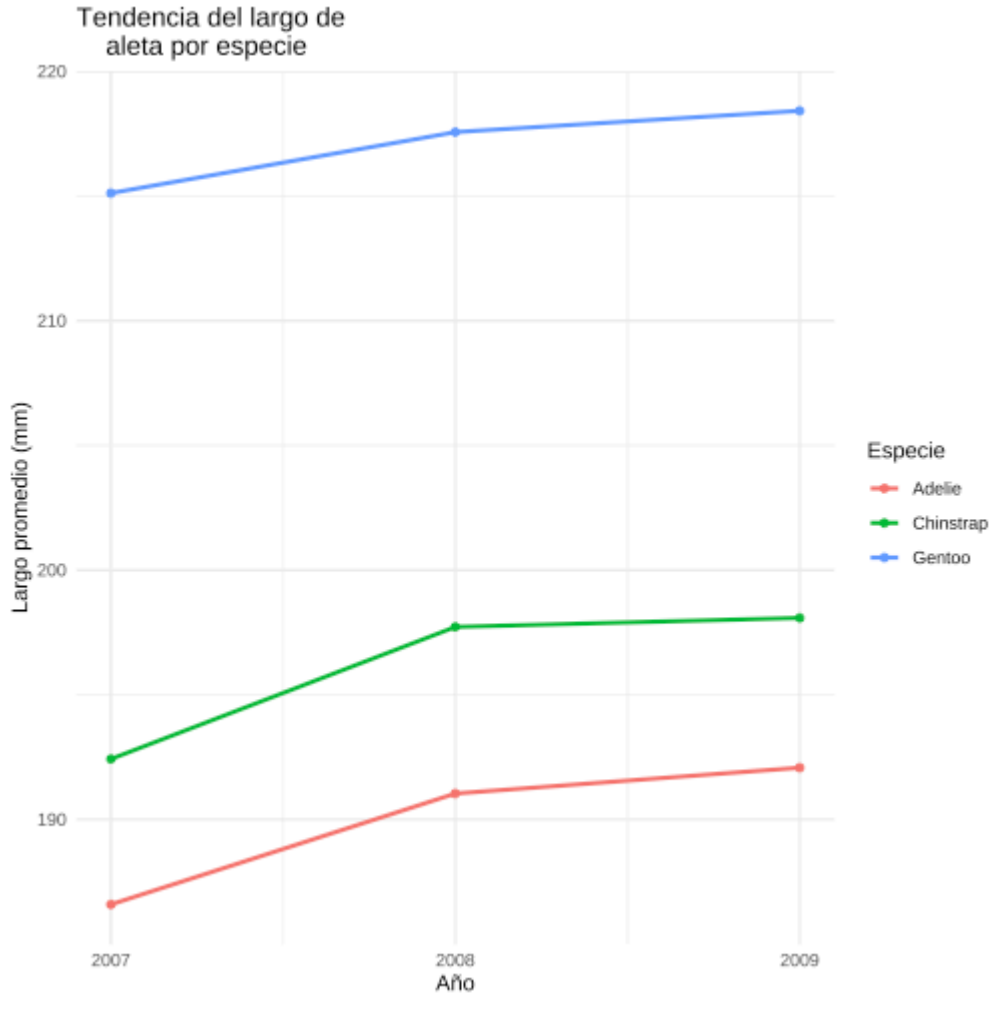
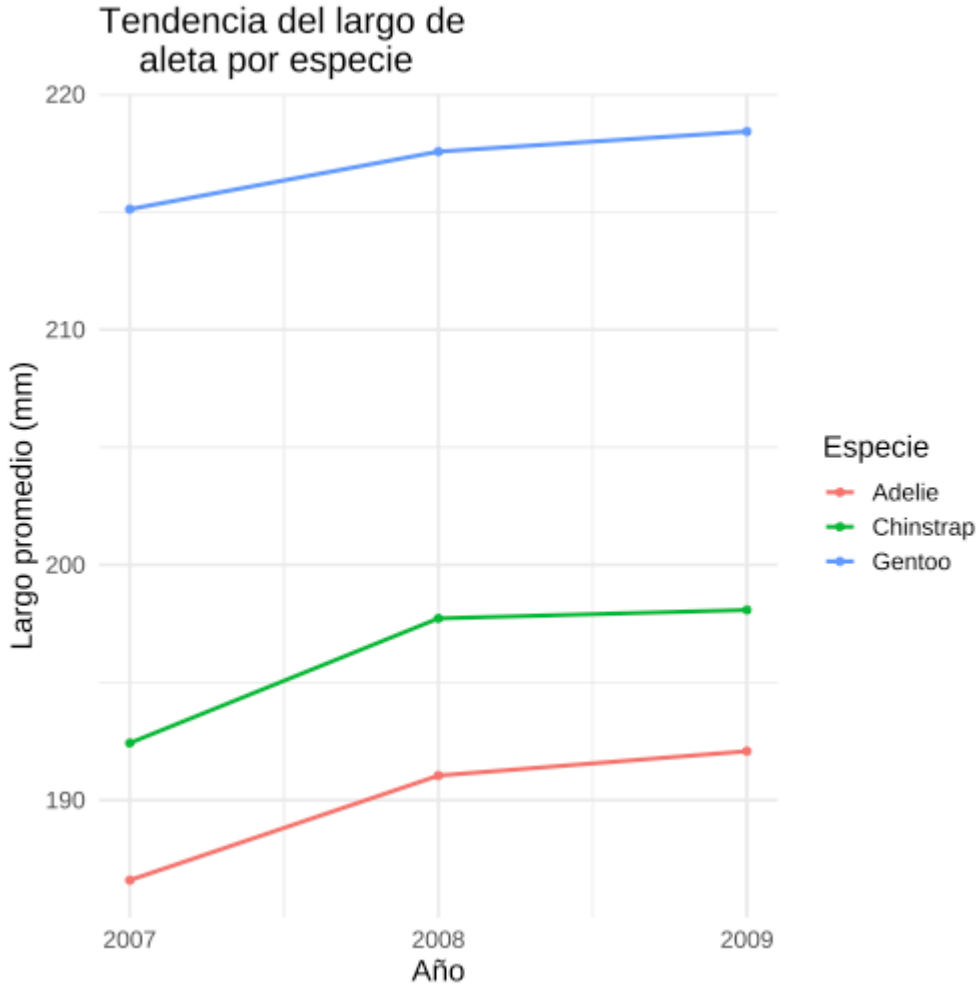


Grafico de tendencia

CIIR
Center for Intercultural
and Indigenous Research

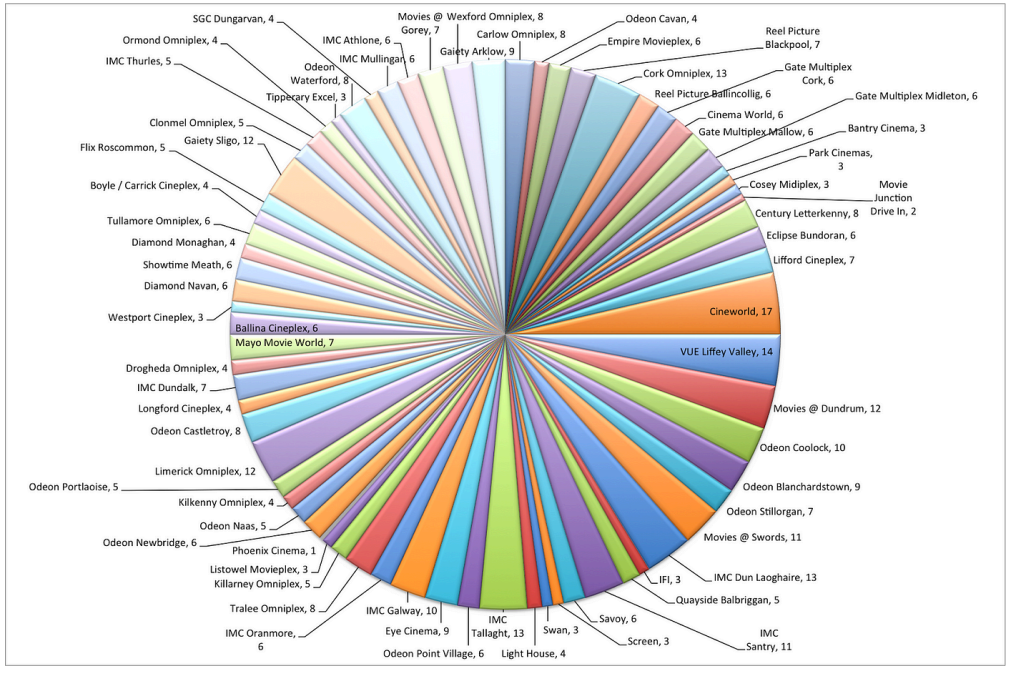
```
g14 <-tabla_pinguino |>
  mutate(year = as.numeric(year)) |>
  ggplot(aes(year, # identificación a
             promedio_aleta,# identif
             color = species)) +
  geom_line(size = 1) +
  geom_point(size = 1.5) +
  scale_x_continuous(breaks =
  unique(tabla_pinguino$year)) +  # T
  labs(
    title = "Tendencia del largo de
    aleta por especie",
    x = "Año",
    y = "Largo promedio (mm)",
    color = "Especie"
  ) +
  theme_minimal(base_size = 15)
```



La necesaria creatividad de la visualización

CIIR
Center for Intercultural and Indigenous Research

- Ojo con la sobreinformación gráfica
- Entender que cada gráfico tiene limitancias y ventajas
- No todo los datos son realmentes necesarios (Por ej: bajo 1% o NA)



La necesaria creatividad de la visualización

CIIR
Center for Intercultural
and Indigenous Research

- Definición correcta de los ejes
- Resaltar lo que nos parece interesante

Hice todo en Ggplot2, ¿y ahora me piden crear gráficos Excel?

CIIR
Center for Intercultural
and Indigenous Research

ggsave() es la función de ggplot2 que se usa para guardar gráficos como archivos en tu computador (por ejemplo, PNG, PDF, JPG, etc.).

Es muy útil cuando quieres exportar un gráfico generado con ggplot() para usarlo en:

- Informes
- Diapositivas
- Publicaciones
- Sitios web

```
ggsave("grafico_alta_calidad.png",
       plot = g1,                # tu gráfico
       width = 20,               # en centímetros
       height = 12,
       units = "cm",
       dpi = 600)                # alta resolución
```

CIIR
Center for Intercultural
and Indigenous Research

Muchas Gracias

Vamos al código!