

# Control 3, respuestas

## Web Scraping y acceso a datos desde la web con R

Cristián Ayala

**Ponderación** 20% de la nota final del curso

**Formato** Desarrollar esta tarea en un Rmarkdown generando un .pdf, agregando comentarios cuando sea necesario.

### 1 Objetivo:

Interesa indagar sobre parecidos y diferencias en el gusto musical en Chile, Perú y Argentina además de ver su relación con las canciones más escuchadas a nivel global. Para ello, analizaremos las listas semanales de canciones de Spotify, usando el paquete [spotifyr](#).

Las listas de canciones más escuchadas desde donde obtendremos la información son las siguientes:

- [Top Global](#)
- [Top Argentina](#)
- [Top Chile](#)
- [Top Perú](#)

Para este análisis será necesario tener o crear una cuenta en Spotify para obtener [credenciales de desarrollador](#) y acceder a su API.

### 2 Tareas:

#### 2.1 Captura de datos

- 1) Obtener y mostrar el listado de las 50 canciones en cada lista.

Genero una lista con los id de las playlists a capturar y luego itero sobre ellas para obtener la información requerida. La función `f_get_playlist` lee la playlist y manipula la respuesta de Spotify para generar una data.frame con los tracks más el número de seguidores y el nombre de la lista.

```
l_listas_id <- c(Global      = '37i9dQZEVXbMDoHDwVN2tF',  
                 Argentina = '37i9dQZEVXbKPTKrnfPD0G',  
                 Chile      = '37i9dQZEVXbLJ0paT1JkgZ',
```

```

Parú      = '37i9dQZEVXbMGcjiWgg253')

f_get_playlist <- function(.playlist_id,
                           .names){

  # Captura de datos desde la API de Spotify.
  l_playlist <- spotifyr::get_playlist(.playlist_id)

  # Mensaje para ver el avance de la captura.
  print(l_playlist$description)

  l_playlist$tracks$items |>
    distinct() |> # Elimino tracks duplicados.
    add_column(followers = l_playlist$followers$total, # Agrego seguidores de la lista-
               nombre_lista = .names, # Agrego el nombre de la lista según l_listas_id.
               .before = 1)
}

df_tracks <- map2_dfr(l_listas_id,
                     names(l_listas_id), f_get_playlist)

[1] "Your daily update of the most played tracks right now - Global."
[1] "Your weekly update of the most played tracks right now - Argentina."
[1] "Your weekly update of the most played tracks right now - Chile."
[1] "Your weekly update of the most played tracks right now - Peru."

# Orden de listas para que "Global" quede primero.
df_tracks <- df_tracks |>
  mutate(nombre_lista = fct_relevel(nombre_lista, 'Global'))

count(df_tracks,
      nombre_lista, followers)

```

	nombre_lista	followers	n
1	Global	16881599	50
2	Argentina	36360	50
3	Chile	13323	50
4	Parú	7499	50

Listado de las canciones según lista se muestra en Tabla 1.

```

f_tabla <- function(x){
  kable(x,
        caption = NULL,
        booktabs = TRUE) |>

```

```

kable_styling(font_size = 7,
              latex_options = c('HOLD_position'))
}

df_tracks |>
  group_by(nombre_lista) |>
  summarise(canciones = list((track.name))) |>
  f_tabla() |>
  column_spec(2, width = '12cm') |>
  row_spec(1:3, hline_after = TRUE)

```

Tabla 1: Canciones por lista

nombre_lista	canciones
Global	As It Was , Glimpse of Us , Me Porto Bonito , Running Up That Hill (A Deal With God) , Tití Me Preguntó , Ojitos Lindos , Left and Right (Feat. Jung Kook of BTS) , Efecto , Moscow Mule , PROVENZA , Heat Waves , BREAK MY SOUL , Jimmy Cooks (feat. 21 Savage) , Te Felicito , About Damn Time , Party , STAY (with Justin Bieber) , Tarot , Bam Bam (feat. Ed Sheeran) , Late Night Talking , First Class , Cold Heart - PNAU Remix , I Like You (A Happier Song) (with Doja Cat) , Massive , Yet To Come , Después de la Playa , One Kiss (with Dua Lipa) , MIDDLE OF THE NIGHT , Una Noche en Medellín , MAMIII , Woman , Ghost , I Ain't Worried , Un Ratito , La Corriente , Enemy (with JID) - from the series Arcane League of Legends , Where Are You Now , Shivers , WAIT FOR U (feat. Drake & Tems) , Until I Found You , Desesperados , INDUSTRY BABY (feat. Jack Harlow) , Vegas (From the Original Motion Picture Soundtrack ELVIS) , ULTRA SOLO REMIX , Blinding Lights , Pepas , Sunroof , Neverita , Run BTS , Sticky
Argentina	Ojitos Lindos , La Triple T , Te Felicito , OJALÁ , Me Porto Bonito , Tití Me Preguntó , Moscow Mule , Antes de Perderte , Efecto , Nos Comemos (feat. Ozuna) , PROVENZA , intoxicacio , Marte , Villano Antillano: Bzrp Music Sessions, Vol. 51 , Una Noche en Medellín , Si Quieren Frontear , BANDIDO , cuatro veinte , ARRANCARMELO , Universo Paralelo , Te Mentiría Remix , Desesperados , Maléfica , Tiago PZK: Bzrp Music Sessions, Vol. 48 , Otra Noche , Fantasi , As It Was , Carne y Hueso , Party , Tarot , Entre Nosotros (Remix) [con Nicki Nicole] , La Llevo Al Cielo (Ft. Ñengo Flow) , MAMIII , Tu Turrítto , Bar , Tranquila , Paulo Londra: Bzrp Music Sessions, Vol. 23 , Tamo Chelo - Remix , Plan A , ULTRA SOLO , Te Mentiría , Felices x Siempre , Miénteme , Numero 1 , París , Esto Recién Empieza , QUEREME , Un Ratito , Jordan , El Mismo Aire
Chile	ULTRA SOLO REMIX , Me Porto Bonito , Efecto , Ojitos Lindos , Moscow Mule , ULTRA SOLO , Me Arrepentí , Tití Me Preguntó , Cochinae , Paz Mental , Una Noche en Medellín , Te Felicito , PROVENZA , Party , La Llevo Al Cielo (Ft. Ñengo Flow) , Bailando , Tarot , SEXTIME , Dime Tú , Un Ratito , Que se hace , La Corriente , Si Sako La 40 , Desesperados , Pégate , Flotao , Noche Loca , Anti Rana , Par de Veces , MAMIII , Después de la Playa , Panamera , As It Was , Yo Quiero , X ÚLTIMA VEZ , Los Dueños Del Malianteo , Yo No Soy Celoso , Neverita , En El Jacuzzi , Más de una Vez , Me Fui de Vacaciones , Aguacero , Running Up That Hill (A Deal With God) , Andrea , Los Malvekes , Phillie y Botella (Remix) , Bailemos Reggaeton , Un Coco , Flaytiando , Mi Gata
Parú	Me Porto Bonito , Efecto , Ojitos Lindos , Tití Me Preguntó , Party , PROVENZA , Moscow Mule , Te Felicito , Tarot , As It Was , MAMIII , Desesperados , Una Noche en Medellín , Después de la Playa , Un Ratito , La Corriente , ULTRA SOLO , Me Fui de Vacaciones , Aguacero , Yo No Soy Celoso , 12x3 , Fuera del mercado , Neverita , Andrea , Lo Siento BB:/ (with Bad Bunny & Julieta Venegas) , ULTRA SOLO REMIX , Tiago PZK: Bzrp Music Sessions, Vol. 48 , Envolver , Jordan , Un Coco , Dos Mil 16 , X ÚLTIMA VEZ , Yonaguni , Todo De Ti , Otro Atardecer , Villano Antillano: Bzrp Music Sessions, Vol. 51 , La Llevo Al Cielo (Ft. Ñengo Flow) , Tacones Rojos , Pantysito , Glimpse of Us , Tiroteo - Remix , Nostálgico , Plan A , Un Verano Sin Ti , Enséñame a Bailar , Medallo , Callaita , A Tu Merced , Running Up That Hill (A Deal With God) , Yet To Come

## 2.2 Canciones

2) Mostrar una tabla en que se muestren las 10 canciones más repetidas.

```
df_tracks_populares <- df_tracks |>
  count(track.name,
        track.artists, track.album.name, track.popularity) |>
  filter(n > 3) |>
  mutate(track.artists = map(track.artists, 'name')) |>
  arrange(track.artists)

n_tracks_populares <- nrow(df_tracks_populares)
```

La Tabla 2 muestra todas las canciones que están presentes en las 4 listas analizadas. En total son 14 canciones.

```
df_tracks_populares |>
  f_tabla()
```

Tabla 2: Canciones repetidas en las cuatro listas estudiadas

track.name	track.artists	track.album.name	track.popularity	n
As It Was	Harry Styles	Harry's House	94	4
Desesperados	Rauw Alejandro , Chenchó Corleone	VICE VERSA	86	4
Efecto	Bad Bunny	Un Verano Sin Ti	96	4
Moscow Mule	Bad Bunny	Un Verano Sin Ti	97	4
Tití Me Preguntó	Bad Bunny	Un Verano Sin Ti	98	4
Un Ratito	Bad Bunny	Un Verano Sin Ti	92	4
MAMIII	Becky G, KAROL G	ESQUEMAS	82	4
Me Porto Bonito	Bad Bunny , Chenchó Corleone	Un Verano Sin Ti	99	4
Ojitos Lindos	Bad Bunny , Bomba Estéreo	Un Verano Sin Ti	99	4
Party	Bad Bunny , Rauw Alejandro	Un Verano Sin Ti	96	4
PROVENZA	KAROL G	PROVENZA	96	4
Tarot	Bad Bunny , Jhay Cortez	Un Verano Sin Ti	94	4
Te Felicito	Shakira , Rauw Alejandro	Te Felicito	94	4
Una Noche en Medellín	Cris Mj	Una Noche en Medellín	93	4

## 2.3 Artistas

3) Mostrar una tabla en que se muestren los 10 artistas más mencionados en cada lista.

```
df_artists <- df_tracks |>
  unnest(track.artists) |>
  count(name, nombre_lista) |>
  add_count(name, wt = n, name = 'total',
            sort = TRUE) |>
  pivot_wider(names_from = nombre_lista,
              values_from = n,
```

```

values_fill = 0)

name_artista_mas_mencionado <- df_artists[1, 'name'][[1]]

```

La Tabla 3 muestra a los 10 artistas con mayor presencia en las listas en estudio. Pueden ser mencionados como solistas o en trabajo conjunto con otros. El cantante con mayor presencia es **Bad Bunny**.

```

df_artists |>
  head(10) |>
  f_tabla()

```

Tabla 3: Artistas repetidos en las cuatro listas estudiadas

name	total	Global	Argentina	Chile	Parú
Bad Bunny	61	11	8	17	25
Rauw Alejandro	15	3	3	3	6
Pailita	12	1	1	8	2
Chencho Corleone	11	2	3	3	3
Cris Mj	9	1	1	6	1
Standly	9	0	0	9	0
KAROL G	8	2	2	2	2
Maria Becerra	7	0	7	0	0
Polimá Westcoast	7	1	1	3	2
El Jordan 23	6	0	0	6	0

## 2.4 Análisis de las canciones

- 4) Revisando lo visto en [clases](#), comparar las canciones de Chile, Argentina, Perú y la lista global respecto de sus [características acústicas](#): *danceability*, *energy*, *instrumentalness* y *valence*<sup>1</sup>.

```

df_tracks_af <- df_tracks |>
  rowwise() |>
  mutate(af = list(get_track_audio_features(track.id)))

df_tracks_af_wide <- df_tracks_af |>
  unnest_wider(af) |>
  select(nombre_lista, track.name, danceability, energy, instrumentalness, valence)

df_tracks_af_long <- df_tracks_af_wide |>
  pivot_longer(cols = c(danceability:last_col()),
    names_to = 'audio_feature')

df_af_listas_mean <- df_tracks_af_long |>
  group_by(nombre_lista, audio_feature) |>
  summarise(value = mean(value),

```

<sup>1</sup>Estos conceptos son definidos por Spotify.

```
.groups = 'drop')
```

Obtener el promedio de esos indicadores en las 50 canciones más populares para cada lista y compárelos en un gráfico.

Primero, la Tabla 4 muestra los valores promedio de las cuatro características de audio de las 50 canciones más escuchadas en las listas de interés.

```
df_af_listas_mean |>
  pivot_wider(names_from = audio_feature,
              values_from = value,
              values_fn = ~round(., 3)) |>
  f_tabla()
```

Tabla 4: Características de audio según listas

nombre_lista	danceability	energy	instrumentalness	valence
Global	0.718	0.685	0.001	0.525
Argentina	0.753	0.678	0.001	0.567
Chile	0.792	0.682	0.001	0.609
Parú	0.746	0.690	0.001	0.525

El gráfico en el que se compara la media de atributos para las cuatro listas se muestra en la Figura 1.

```
# Asigno como theme_minimal a todos los gráficos posteriores.
# También cambio las características del título y subtítulo para que acepte markdown.
theme_set(
  theme_minimal() +
  theme(plot.title = ggtext::element_markdown(),
        plot.subtitle = ggtext::element_markdown(),
        plot.title.position = 'plot')
)

rango <- function(x) {
  data.frame(list(ymin = min(x),
                 ymax = max(x)))
}

ggplot(df_af_listas_mean,
       aes(x = value,
           y = fct_reorder(audio_feature, value, mean))) +
  stat_summary(geom = 'linerange',
              fun.data = rango) +
  geom_point(aes(colour = nombre_lista),
             size = 5, alpha = 0.7) +
  scale_x_continuous('Puntaje',
                    limits = c(0,1),
```

```

labels = ~scales::number(., accuracy = 0.1)) +
scale_colour_manual(NULL,
  values = c('#767D94', '#64D7FA', '#FBC2A3', '#C7593C')) +
labs(title = '**Comparación de medias de las características de audio**',
  y = NULL) +
theme(axis.text.y = element_text(size = rel(1.4)))

```

## Comparación de medias de las características de audio

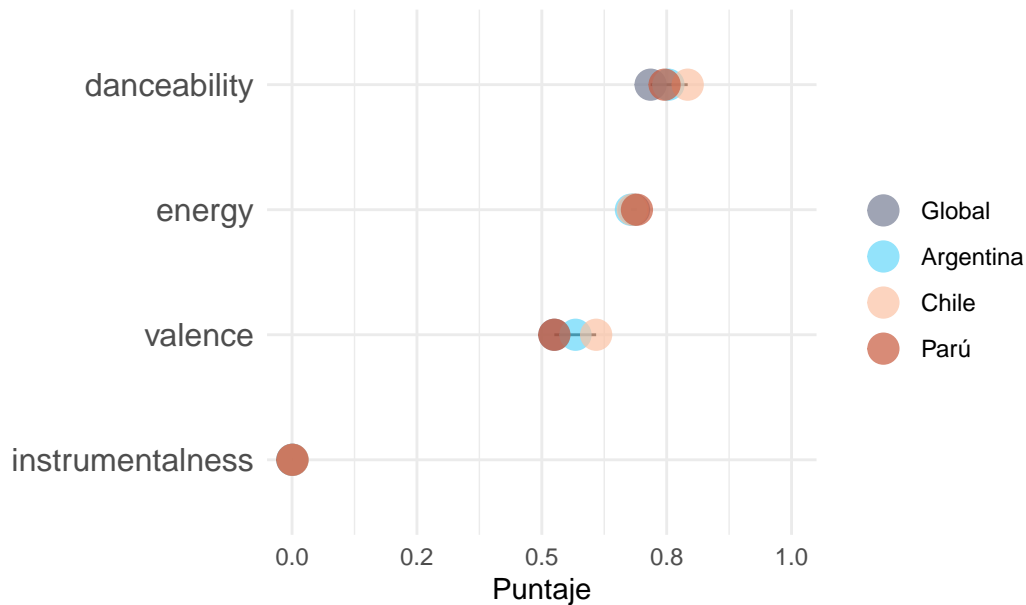


Figura 1: Comparación en la media de las características de audio según listas

¿Hay diferencias en el gusto musical entre los 3 países y el mundo?

- Los tres países se parecen bastante y están cerca de la lista global.
- En el margen, Chile tiene canciones más danzables y positivas —*valance*— que en los otros dos países.
- Las canciones más populares tienen voz. Lo instrumental está totalmente fuera.

La distribución de las canciones en estas cuatro métricas se muestra en la Figura 2.

```

library(ggtext)

colores = c("#082243", "#0e5385", "#fdf6e2", "#ed612d")

ggplot(df_tracks_af_long,
  aes(x = nombre_lista,
    y = value)) +

```

```

geom_violin(fill = colores[3],
            colour = NA) +
geom_tile(alpha = 0.2,
          width = 0.8, height = .01,
          fill = colores[2],
          colour = NA) +
stat_summary(geom = 'point',
            fun = median,
            shape = 5,
            colour = colores[1], size = 3) +
stat_summary(geom = 'point',
            fun = mean,
            alpha = .8,
            colour = colores[4], size = 4) +
facet_grid(cols = vars(audio_feature)) +
scale_x_discrete(NULL,
                guide = guide_axis(n.dodge = 2)) +
scale_y_continuous('Puntaje',
                  labels = ~scales::number(., accuracy = 0.1)) +
labs(title = 'Distribución de canciones según características de audio',
     subtitle = str_glue("Se muestra la **<span style='color:{colores[4]}>media</span>**  
y **<span style='color:{colores[1]}>mediana</span>** de cada distribución"))

```

## Distribución de canciones según características de audio

Se muestra la **media** y **mediana** de cada distribución

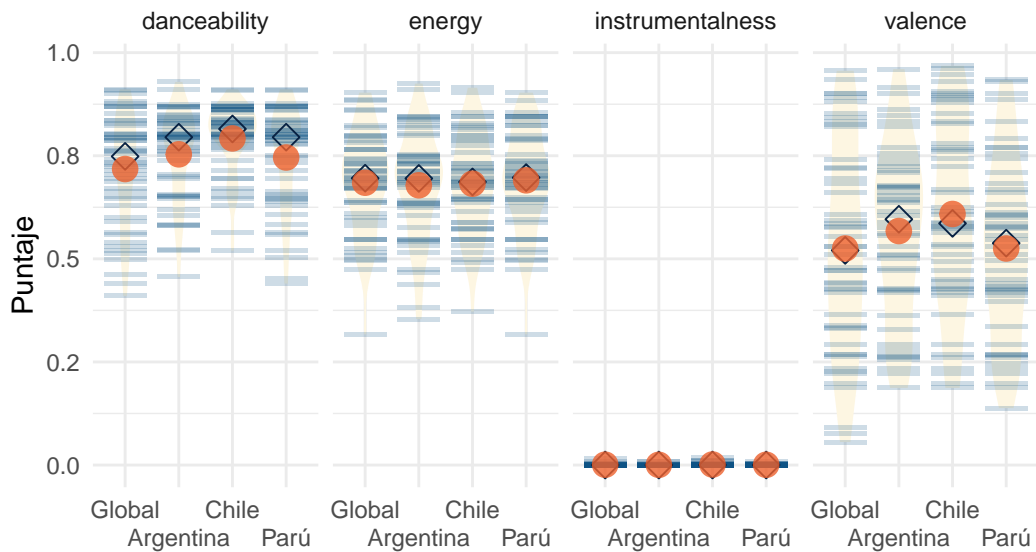


Figura 2: Distribución de canciones según atributos sónicos y listas