

Control 3, respuestas

Web Scraping y acceso a datos desde la web

Cristián Ayala

Ponderación 20% de la nota final del curso

Formato Desarrollar esta tarea con [Quarto](#) o [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 Colombia](#)
- [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 = '37i9dQZEVXbNG2KDCFcK0F',
                 Argentina = '37i9dQZEVXbKPTKrnFPDOG',
                 Chile = '37i9dQZEVXbLJ0paT1JkgZ',
                 Colombia = '37i9dQZEVXb0a2lmxNORXQ',
                 Perú = '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 |>
    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)
}

l_tracks <- map2(l_listas_id, names(l_listas_id),
               f_get_playlist)

```

```

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

```

```

df_tracks <- l_tracks |>
  list_rbind()

```

Orden de listas para que “Global” quede primero. Variable `nombre_lista` como factor.

```

df_tracks <- df_tracks |>
  mutate(nombre_lista = fct_relevel(nombre_lista, 'Global'))

```

Número de seguidores de cada lista es una constate que se repite en cada canción. Confirmamos que cada lista tiene 50 canciones.

```

count(df_tracks,
      nombre_lista, followers, name = 'n_canciones')

```

	nombre_lista	followers	n_canciones
1	Global	1424083	50

2	Argentina	78193	50
3	Chile	23658	50
4	Colombia	822682	50
5	Perú	14698	50

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

```
# tbl-cap: 'Canciones por lista'

f_tabla <- function(x){
  tt(x,
    theme = 'bootstrap') |>
  format_tt(escape = TRUE) |>
  style_tt(fontsize = .7) |>
  style_tt(i = 0, # Header de la tabla.
    fontsize = .8)
}

df_tracks |>
  summarise(canciones = str_c(track.name,
                              collapse = ' - '),
    .by = nombre_lista) |>
  f_tabla() |>
  style_tt(aligned = 'l',
    tabularray_inner = "column{2} = {wd=12cm}")
```

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'),
    track.artists = map_chr(track.artists,
      \(t) paste0(t, collapse = ', '))) |>
  arrange(-track.popularity)

n_tracks_populares <- nrow(df_tracks_populares)
```

La Tabla 2 muestra todas las canciones que están presentes en al menos 4 de listas analizadas. En total son 10 canciones.

Tabla 1

nombre_lista	canciones
Global	Please Please Please - Espresso - BIRDS OF A FEATHER - Not Like Us - Gata Only - MILLION DOLLAR BABY - Beautiful Things - Too Sweet - A Bar Song (Topsy) - I Had Some Help (Feat. Morgan Wallen) - Good Luck, Babe! - Houdini - i like the way you kiss me - LUNCH - REAL GANGSTA LOVE - Si Antes Te Hubiera Conocido - End of Beginning - Lose Control - we can't be friends (wait for your love) - Fortnite (feat. Post Malone) - Stargazing - CHIIHIRO - Cruel Summer - As It Was - BAND4BAND (feat. Lil Baby) - One Of The Girls (with JENNIE, Lily Rose Depp) - Austin (Boots Stop Workin') - Santa - LUNA - Sweater Weather - Belong Together - Locked out of Heaven - I Wanna Be Yours - I Don't Wanna Wait - Yellow - The Night We Met - Pour Me A Drink (Feat. Blake Shelton) - I Can Do It With a Broken Heart - Slow It Down - Feather - Like That - Riptide - Nonsense - Like Crazy - SI NO ES CONTIGO - Stick Season - us. (feat. Taylor Swift) - Si No Quieres No - Saturn - FE!N (feat. Playboi Carti)
Argentina	REAL GANGSTA LOVE - Perdonarte, ¿Para Qué? - Hoy - Ojos Verdes - Tu Foto - En Vivo - Un Besito Más - HOLA PERDIDA REMIX - IMAN (Two Of Us) - Piel - HOY - LUNA - Santa - SAN TURRONA - BESAME (feat. Tiago PZK, Khea & Neo Pisteá) - Remix - Gata Only - TATTOO - SI NO ES CONTIGO - APAGA EL CEL - QUE ME FALTE TODO - La_Playlist.mpeg - FLECHAZO EN EL CENTRO - ROZE DJ TAO Turreo Sessions #22 - Tumba - SOLTERO - HOLA PERDIDA - Mil Preguntas - BBY BOO - REMIX - Mil Preguntas - 30 Grados - SINVERGÜENZA - con Angela Torres - offline - Loquita - Remix - Tal Para Cual - Cierre Rkt - La_Original.mp3 - UH LA LA LA - Por Mil Noches - Si Antes Te Hubiera Conocido - Una Foto Remix (feat. Emilia) - SORRY 4 THAT MUCH - Exclusive.mp3 - M.A.I - Bad Boy - Please Please Please - La Morocha - CEO - La Bachata - Una Foto - EL NEGRO TECLA Omar Algo Anda Mal #9 - Rara Vez
Chile	SI NO ES CONTIGO - REAL GANGSTA LOVE - No Ponga Excusas - Gata Only - MUJER FINA - Si Antes Te Hubiera Conocido - APAGA EL CEL - Gitana (feat. Bryartz) - Enrolar - Fantastic Four Remix - SOLO DIME - Avisame - Daytona - Que Hay Amor - LUNA - Bastardo (feat. Ñengo Flow & Jere Klein) - miao - Coa Xorete - Santa - REELS - Escarlata - Remix - Rapido (with Cris Mj) - Piel - Ese Guille - Diabólica - Es un Secreto - BBY BOO - REMIX - ADIVINO - SORRY 4 THAT MUCH - Minnie <3 - Please Please Please - PRINCESITA DE ... - La Bachata - ME GUSTA - LA BANDA - YUMMY! - ¿Cuál Es Su Nombre? - Que Brille Sola Mami - Partyson - AMARGURA - BLANKITA COMO NIEVE - Grabándote:) - La Receta - Espresso - Perdonarte, ¿Para Qué? - offline - QLONA - MA DIAMANTE - 2012 - Tas Texteandome - QUICKIE
Colombia	Mirame - Si Antes Te Hubiera Conocido - SORRY 4 THAT MUCH - LUNA - Chismofilia - REAL GANGSTA LOVE - Ohnana - Brickell - AMARGURA - Contentoso - La Bachata - Razones - El Amor Y La Felicidad - PERRO NEGRO - Santa - Orion - Gata Only - QLONA - GATITAS SANDUNGUERAS VOL.1 - Feliz Cumpleaños Ferxxo - Mi Palabra Contra la de Ellos - En Vivo - CUAL ES ESA - EL RITMO QUE NOS UNE - feat. Selección Colombia - Según Quién - offline - Oe Bebé - Una Vida Pasada - WYA REMIX BLACK AND YELLOW (feat. iZaak & Pirlo) - MI EX TENÍA RAZÓN - Primera Cita - CLASSY 101 - SI SABE FERXXO - ADIVINO - Señal De Vida - La Diabla - Es un Secreto - Normal - Monastery - Borro Cassette - La Ex - En Vivo - Llorarás - Dile - Ferxxo 100 - Pa' Que Retozen - Sicaria - Dile a Tu Amiga - LA FALDA - WYA REMIX RED - Hey Mor - PROVENZA
Perú	REAL GANGSTA LOVE - LUNA - Santa - Gata Only - SORRY 4 THAT MUCH - Si Antes Te Hubiera Conocido - SI NO ES CONTIGO - PARTIDO EN DOS - La Bachata - Piel - offline - AMARGURA - Please Please Please - Una Vida Pasada - Rara Vez - PERRO NEGRO - WYA REMIX RED - ADIVINO - QLONA - Feliz Cumpleaños Ferxxo - CLASSY 101 - Borro Cassette - Mamichula - con Nicki Nicole - BBY BOO - REMIX - Espresso - MI EX TENÍA RAZÓN - Una Foto Remix (feat. Emilia) - GATITAS SANDUNGUERAS VOL.1 - Si Esta Casa Hablara - La Curiosidad - Es un Secreto - Amanece - Desesperados - Orion - Según Quién - Brickell - LA FALDA - APAGA EL CEL - Columbia - Perdonarte, ¿Para Qué? - Tiroteo - Remix - Ride Or Die, Pt. 2 (with Villano Antillano & Tokischa) - LA CANCIÓN - Normal - Like Crazy - La Muralla Verde - Cúrame - BIRDS OF A FEATHER - Miénteme - Daytona

Tabla 2: Canciones repetidas en las cuatro listas estudiadas

track.name	track.artists	track.album.name	track.popularity	n
Please Please Please	Sabrina Carpenter	Please Please Please	97	4
Gata Only	FloyyMenor, Cris Mj	Gata Only	96	5
LUNA	Feid, ATL Jacob	FERXXOCALIPSIS	90	5
REAL GANGSTA LOVE	Trueno	EL ÚLTIMO BAILE	90	5
SI NO ES CONTIGO	Cris Mj	SI NO ES CONTIGO	88	4
Si Antes Te Hubiera Conocido	KAROL G	Si Antes Te Hubiera Conocido	87	5
SORRY 4 THAT MUCH	Feid	SORRY 4 THAT MUCH	86	4
La Bachata	Manuel Turizo	2000	85	4
offline	Young Miko, Feid	att.	85	4
Santa	Rvssian, Rauw Alejandro, Ayra Starr	Santa	84	5

```
df_tracks_populares |>
  f_tabla() |>
  style_tt(aligned = 'lllrr',
           tabularray_inner = "column{1-3} = {wd=3.8cm}")
```

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(Artista = name, nombre_lista) |>
  add_count(Artista,
            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, 'Artista'][[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 **Feid**.

Tabla 3: Artistas repetidos en las cuatro listas estudiadas

Artista	Total	Global	Argentina	Chile	Colombia	Perú
Feid	30	1	3	3	14	9
Cris Mj	18	2	2	10	1	3
KAROL G	14	1	1	3	5	4
FloyyMenor	9	1	2	3	1	2
Jere Klein	9	0	0	9	0	0
Sabrina Carpenter	9	4	1	2	0	2
Emilia	8	0	5	1	0	2
Rauw Alejandro	8	1	1	1	1	4
Kidd Voodoo	7	0	0	7	0	0
Anuel AA	6	0	1	1	2	2

```
df_artists |>
  head(10) |>
  f_tabla() |>
  style_tt(align = 'lrrrrrr')
```

2.4 Análisis de las canciones

- 4) Revisando lo visto en [clases](#), comparar las canciones de las listas respecto de sus [características acústicas](#): *danceability*, *energy*, *instrumentalness* y *valence*¹.

```
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 |>
  summarise(value = mean(value),
            .by = c(nombre_lista, audio_feature))
```

¹Estos conceptos son definidos por Spotify.

Tabla 4: Características de audio según listas

nombre_lista	danceability	energy	instrumentalness	valence
Global	0.664	0.651	0.032	0.565
Argentina	0.726	0.701	0.005	0.674
Chile	0.758	0.700	0.012	0.631
Colombia	0.783	0.691	0.000	0.629
Perú	0.775	0.692	0.006	0.612

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() |>
  style_tt(align = 'lrrrr')
```

El gráfico en el que se compara la media de atributos para las 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 = 'linerrange',
              fun.data = rango) +
  geom_tile(aes(fill = nombre_lista),
           height = 0.5,
           width = 0.015,
```

```

    alpha = 0.7) +
scale_x_continuous('Puntaje',
  limits = c(0,1),
  labels = ~scales::number(., accuracy = 0.1)) +
scale_fill_manual('Listas',
  values = c(Global = 'gray10',
    Argentina = '#74ACDF',
    Chile = '#E53935',
    Colombia = '#FCD116',
    Perú = '#B71C1C')) +
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

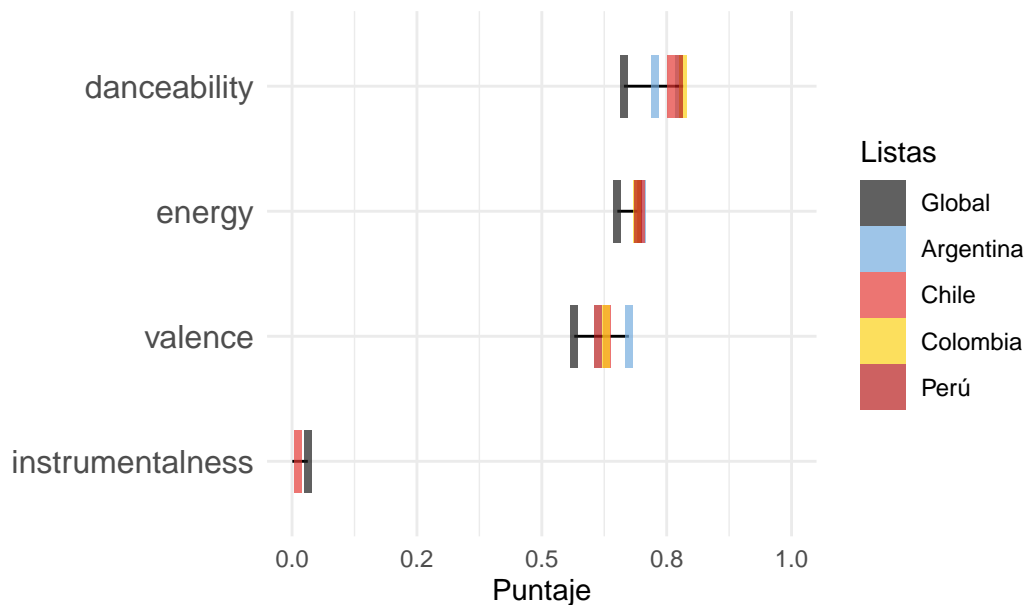


Figura 1: Comparación en la media de las características de audio según listas

¿Hay diferencias en el gusto musical entre los 4 países y el mundo?

- Los cuatro países se parecen bastante y están cerca de la lista global.
- La lista global se distingue al estar siempre por debajo de las listas de los países latinos evaluados.
- Dentro de latinoamérica, Argentina tiene un perfil algo diferente al resto.

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

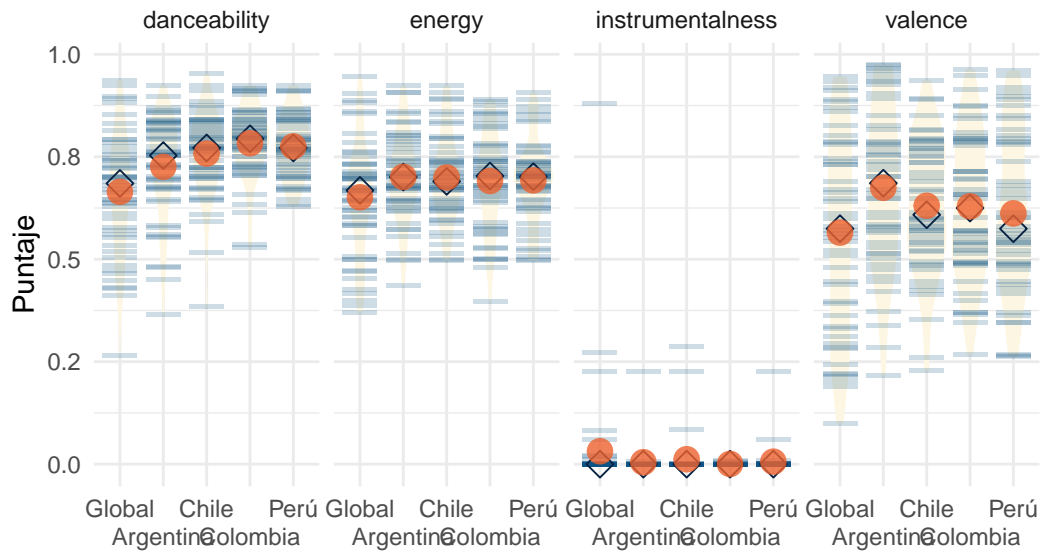


Figura 2: Distribución de canciones según atributos sónicos y listas