



FACULTAD DE MATEMÁTICAS  
PONTIFICIA UNIVERSIDAD  
CATÓLICA DE CHILE

# EYP1113 - Probabilidad y Estadística

Taller R - 03: Gráficos en R

Ricardo Aravena C. - Cristian Capetillo C. - Ingrid Guevara R.  
Bladimir Morales T. - Ricardo Olea O. - Daniel Saavedra M.

Facultad de Matemáticas  
Departamento de Estadística  
Pontificia Universidad Católica de Chile

Segundo Semestre 2024

# Contenido I

Gráficos en R

Aplicación ilustrativa

Aplicación

# Gráficos en R

En R existen funciones básicas para visualizar el comportamiento de los datos:

- ▶ `hist()` Histograma
- ▶ `plot()` Dispersión
- ▶ `lines()` Agrega un plot sobre un gráfico ya existente
- ▶ `boxplot()` Diagrama de cajas
- ▶ `vioplot()` Diagrama de violín, librería "vioplot"
- ▶ `pie()` Torta\*
- ▶ `barplot()` Diagrama de barras
- ▶ `corrplot()` Gráficos de correlación, librería "corrplot"\*
- ▶ `faces()` Caras de Chernoff, librería "aplpack"\*
- ▶ `wordcloud()` Nube de palabras, librería "wordcloud"\*
- ▶ `treemap()` Treemap, librería "treemap"\*

Por otra parte la librería "ggplot2" permite realizar múltiples gráficas que visualmente son muy atractivas.



# Aplicación ilustrativa

No cabe duda que Michael Jackson ha dejado un legado importante en la música. Es esperable que el rey del pop nos pueda dar un indicio de cómo lograr una carrera exitosa.

Con el ánimo de entender qué características diferencian a Michael Jackson con el resto de la muestra un análisis descriptivo tanto de la muestra como de nuestro artista de interés resulta pertinente.

La base de datos contiene 23 canciones de las 263 que Michael compuso a lo largo de su vida. Más aún, la base no es rica en cuanto a artistas, sus canciones y el género al cual pertenecen éstas. Así, este análisis resulta en tan solo una aplicación ilustrativa.



# Aplicación ilustrativa

Con lo anterior, cabe preguntarse, ¿cómo es el comportamiento general de la popularidad y la duración de las canciones?, o estadísticamente hablando, ¿qué característica tiene la distribución de tales variables en la muestra total?

Hecho el análisis general, ¿se diferencia tal distribución si sólo nos concentramos en las canciones de Michael Jackson?, ¿en qué se diferencia?



# Aplicación ilustrativa

Las preguntas anteriores pueden ser resueltas mediante medidas descriptivas discutidas la clase anterior. Promedio, cuantiles o medidas de dispersión son algunas de las opciones más comunes.

Notaremos que sin embargo, tantas medidas pueden resultar difíciles de interpretar y en consecuencia, de comunicar. Una opción, quizás menos exacta pero mucho más comprensible, es el análisis gráfico de las variables.

Comprobemos esto en R.

# Aplicación

El abalón es un molusco marino de la familia Haliotidae, conocido por su concha en forma de oreja, que es iridiscente y muy valorada tanto en joyería como en artesanía. En términos biológicos, el abalón es un tipo de caracol marino, y su carne es considerada un manjar en muchas culturas, especialmente en Asia.

La base de datos `ABALON.xlsx` contiene información correspondiente a la recolección de abalones en 4 centros distintos: Caldera, Coquimbo, Puerto Montt y Chiloé. En este taller se busca realizar un análisis descriptivo y gráfico utilizando funciones y librerías de R.



# Aplicación

Se tienen las siguientes variables: largo, diámetro, alto, peso total, peso del cuerpo, peso de la concha, y número de anillos. Se desea analizar lo siguiente:

- a) Realizar un análisis de la distribución del largo de los abalones para todas las observaciones. ¿Cómo representaría gráficamente esta distribución? En el gráfico, incluya la media, mediana, y los cuartiles 1 y 3. Interprete los resultados. ¿Qué puede inferir sobre la asimetría y curtosis de la distribución?
- b) Si se desea comparar las distribuciones del largo de los abalones entre los centros de Caldera y Coquimbo en un mismo gráfico utilizando histogramas, ¿qué conclusiones se pueden obtener de esta comparación?



# Aplicación

- c) Si se desea comparar la variabilidad de los centros de Caldera, Coquimbo y Chiloé, ¿qué se puede decir respecto a la desviación estándar de cada uno? Si graficamos un boxplot para los tres centros, ¿qué interpretación se puede dar?
- d) Se quiere analizar el grado de asociación entre el largo y el peso total de los abalones. Una posible medida estadística para esto es la correlación. ¿Cuál es el valor de este coeficiente? Si se grafica un diagrama de dispersión de ambas variables, ¿el grado de asociación reflejado por el coeficiente de correlación tiene sentido? ¿Por qué?
- e) A partir del gráfico de dispersión del punto anterior, diferencie por color los centros de Coquimbo, Caldera y Chiloé.

