

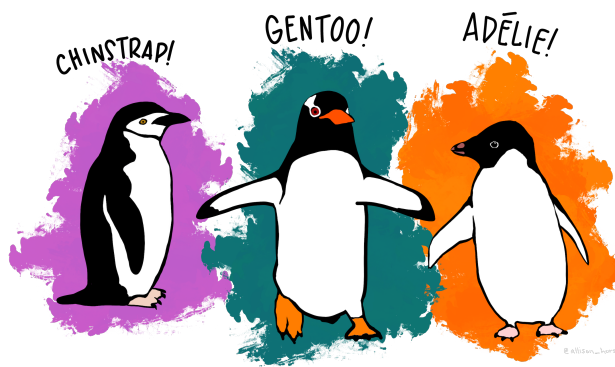
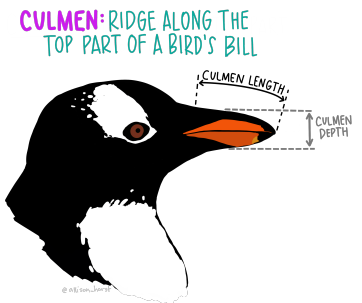
TP2 - Reconocimiento de Patrones

Elio Campitelli

1 Datos

En este TP voy a estar usando la base de datos penguins del paquete de R **palmerpenguins** (KB, TD, and WR 2014). Los datos recolectados por la Dra. Kristen Gorman en la Estación Palmer, consisten en mediciones de la longitud del culmen¹, alto del culmen y la masa corporal de 342 pingüinos de las especies *Pygoscelis adeliae* (Pingüino de Adelaida), *Pygoscelis papua* (Pingüino Juanito), y *Pygoscelis antarcticus* (Pingüino barbijo).

¹ El culmen es la parte superior del pico de las aves.



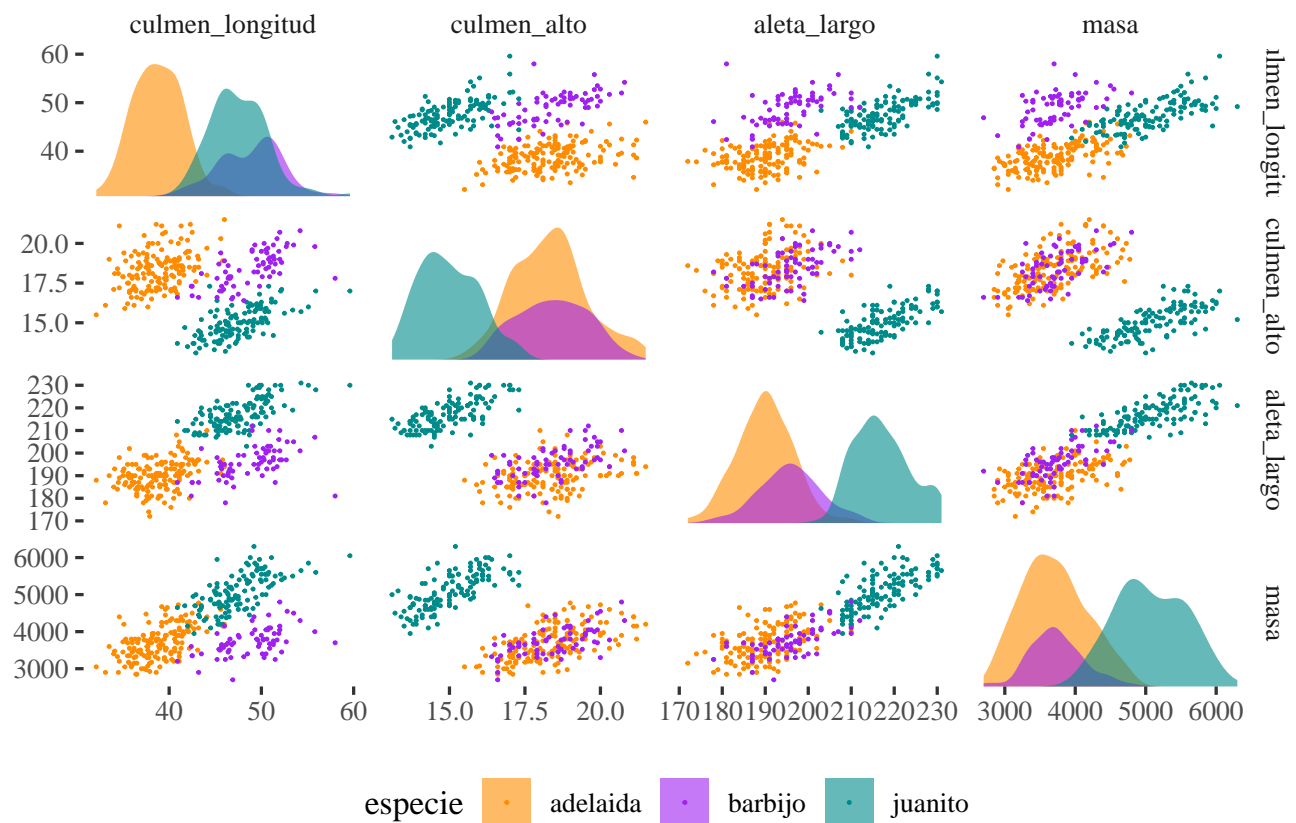
En la Tabla 1 se muestran las primeras 3 mediciones para cada especie.

Especie	Longitud de culmen [mm]	Alto de culmen [mm]	Longitud de la aleta [mm]	Masa corporal [g]
adelaida	39.1	18.7	181	3750
adelaida	39.5	17.4	186	3800
adelaida	40.3	18.0	195	3250
juanito	46.1	13.2	211	4500
juanito	50.0	16.3	230	5700
juanito	48.7	14.1	210	4450
barbijo	46.5	17.9	192	3500
barbijo	50.0	19.5	196	3900
barbijo	51.3	19.2	193	3650

Table 1: Primeras 3 entradas de cada especie en los datos utilizados

La variable categórica a predecir va a ser la especie, y las posibles variables predictoras son las dimensiones del culmen, la longitud de

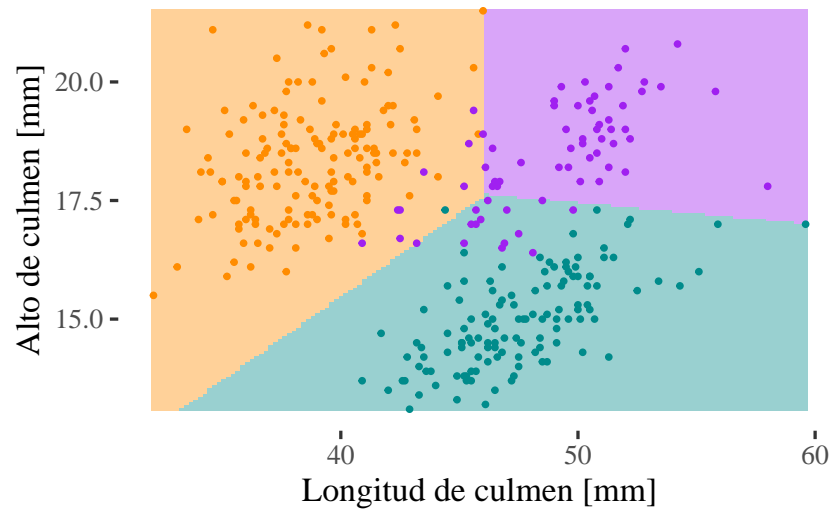
la aleta y la masa corporal. Es decir, en principio es un espacio de dimensión 4. Como la idea es trabajar en \mathbb{R}^2 , conviene explorar qué grado de separación permite cada combinación de dos variables. Esto se muestra en la Figura 1 donde se grafican scatterplots para todas las combinaciones de dos variables con la especie representada con color. Mirando las densidades de probabilidad (gráficos en la diagonal) se puede ver que la longitud del culmen separa bastante bien entre pingüino de adelaida y el resto mientras que las otras variables separan bien al pingüino juanito. Por lo tanto, las combinaciones que incluyen la longitud del culmen (gráficos en la primera columna) separan bastante bien entre las tres especies, mientras que el resto de las combinaciones tienen algún grado de mezcla entre pingüino de adelaida y pingüino juanito.

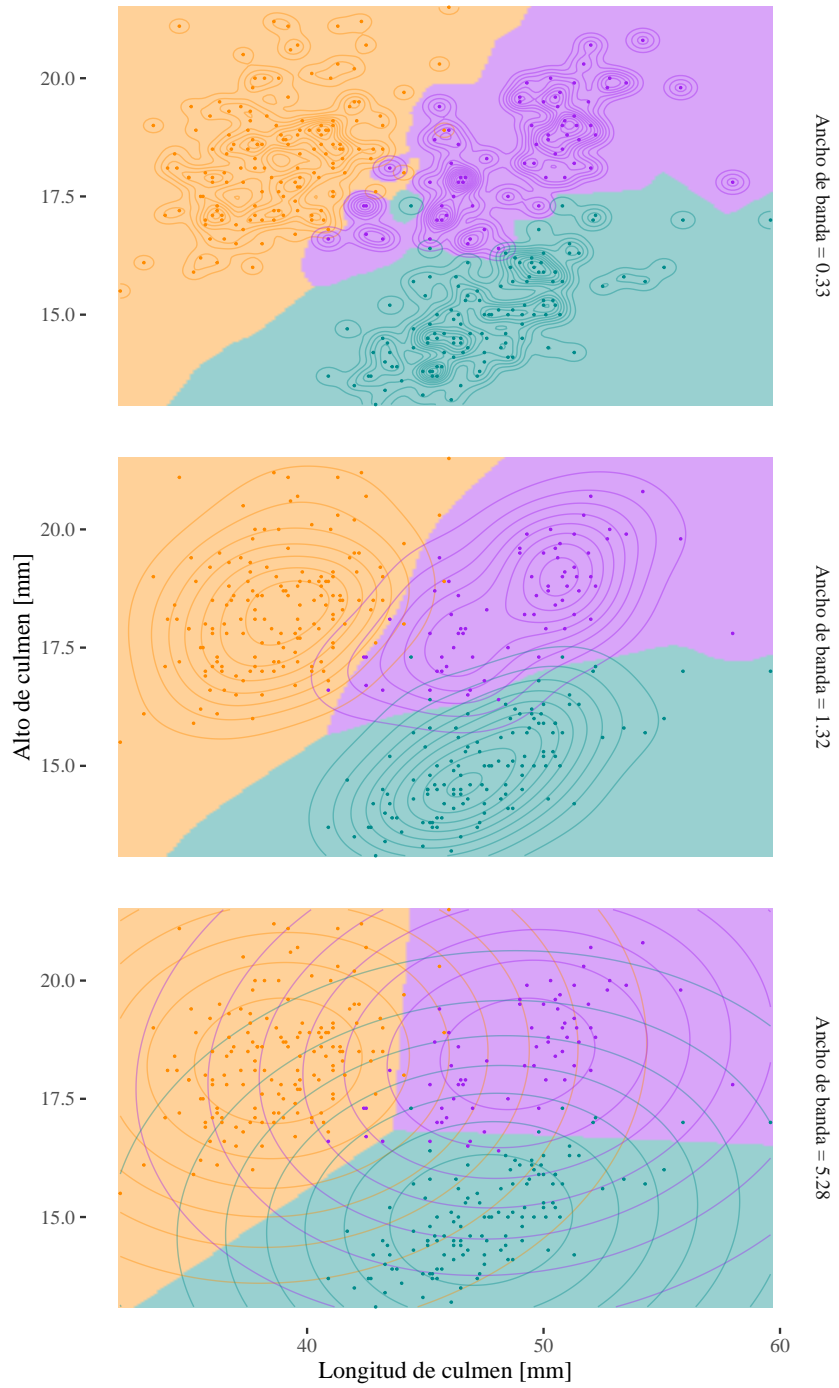


En lo que sigue, voy a usar los datos de Longitud de culmen y de alto del culmen para separar entre las tres especies.

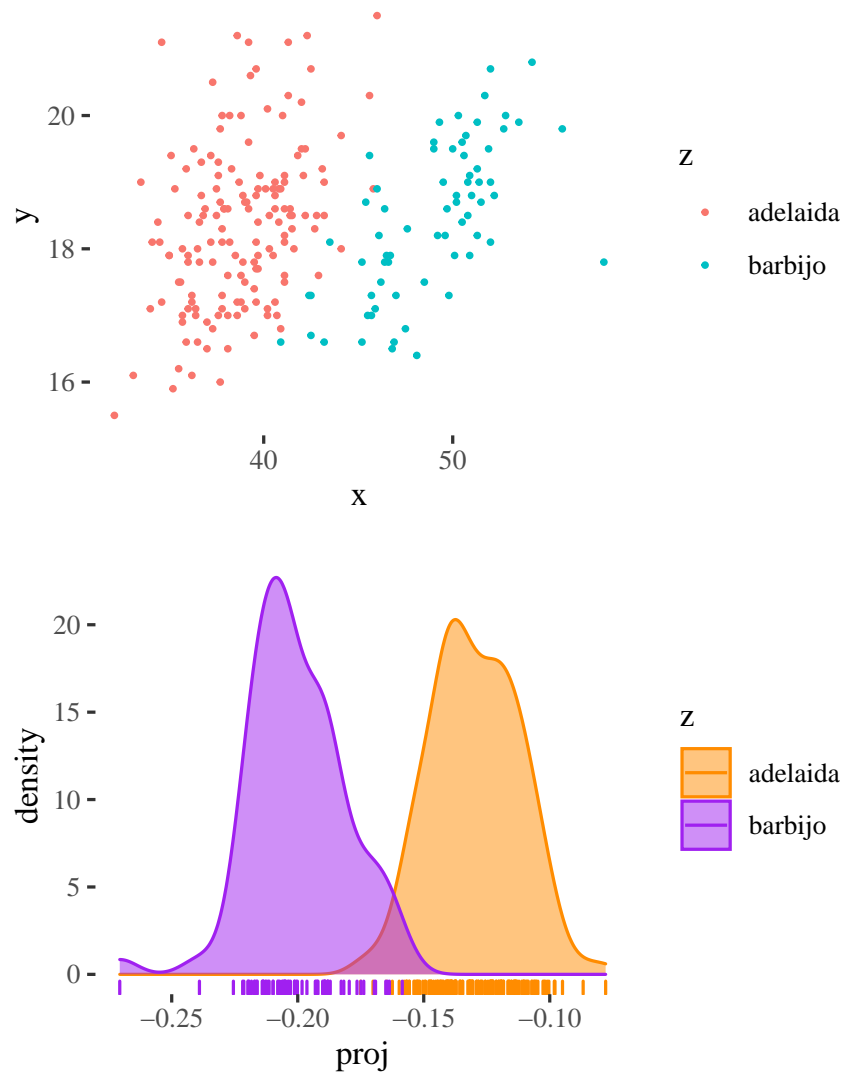
Figure 1: Scatterplot de todas las combinaciones de variables posibles en \mathbb{R}^2 . En la diagonal, estimaciones de densidad de cada variable separadas por especies.

2 Clasificador cuadrático

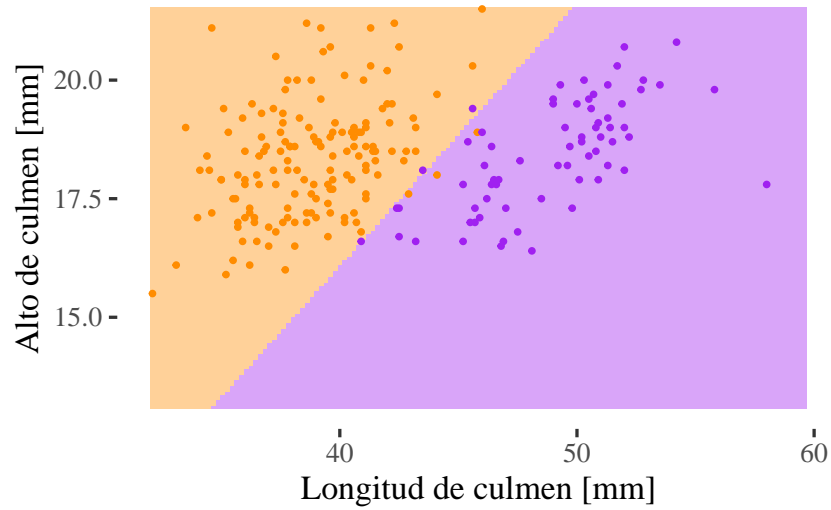




2.1 Fisher



2.2 Logística



Referencias

KB, Gorman, Williams TD, and Fraser WR. 2014. "Ecological Sexual Dimorphism and Environmental Variability Within a Community of Antarctic Penguins (Genus *Pygoscelis*)."
PLoS ONE 9(3) (e90081): -13.
<https://doi.org/10.1371/journal.pone.0090081>.