EXERCICI 2. ¿ES VENENOSA ESTA SETA?

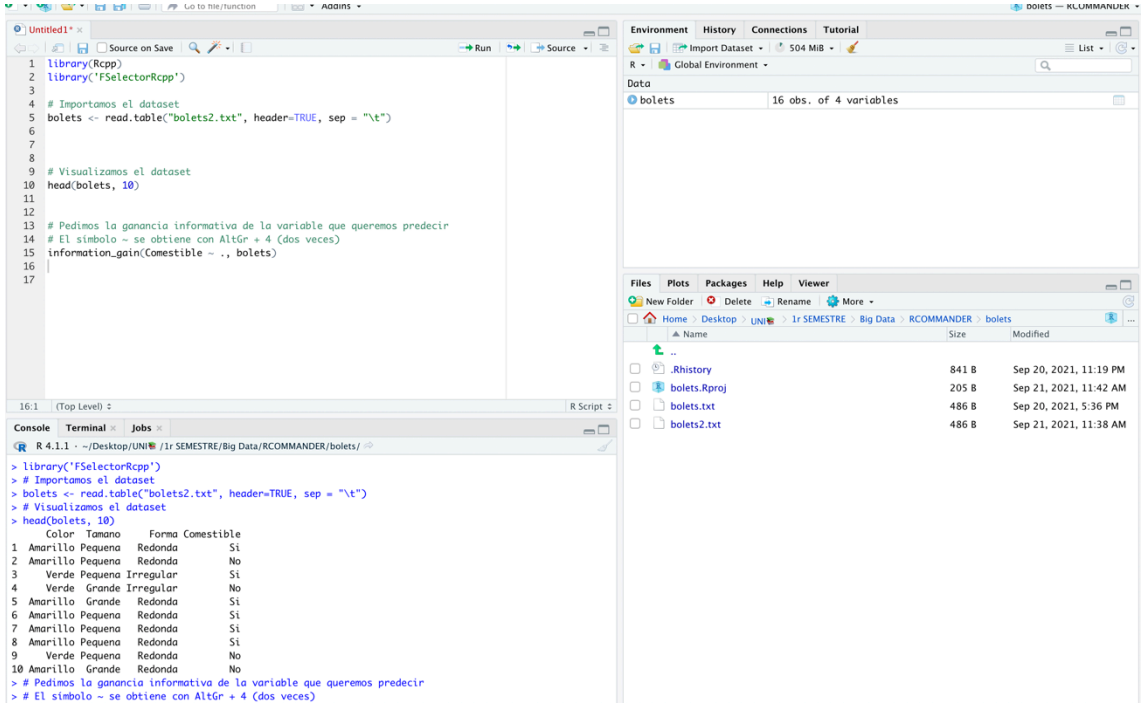
Hemos recogido información sobre 16 variedades de setas. Para cada una de ellas conocemos el «color», el «tamaño» y la «forma», así como si es una variedad «comestible» o «venenosa». El objetivo de la actividad es determinar qué características de las setas («color», «tamaño» o «forma») aportan más información sobre su carácter comestible o venenoso. Utiliza el paquete FSelectorRcpp para resolver el ejercicio.

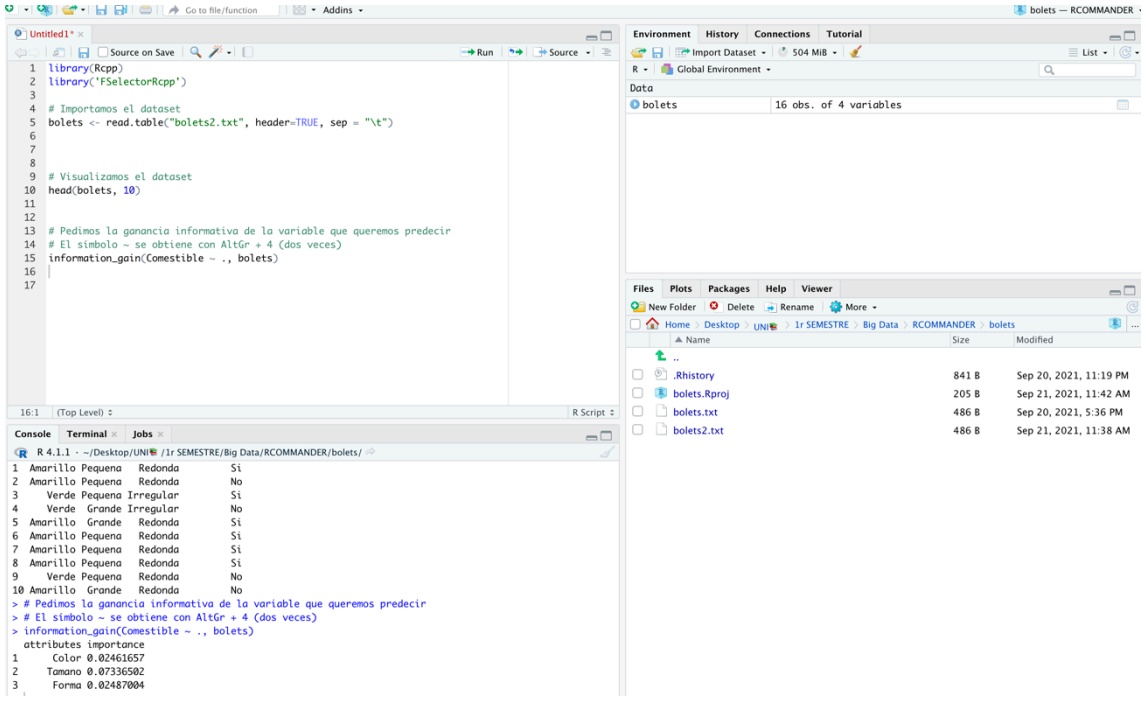
Table 2.2: Datos de una muestra de 16 setas

Color	Tamaño	Forma	Comestible
Amarillo	Pequeña	Redonda	Sí
Amarillo	Pequeña	Redonda	No
Verde	Pequeña	Irregular	Sí
Verde	Grande	Irregular	No
Amarillo	Grande	Redonda	Sí
Amarillo	Pequeña	Redonda	Sí
Amarillo	Pequeña	Redonda	Sí
Amarillo	Pequeña	Redonda	Sí
Verde	Pequeña	Redonda	No
Amarillo	Grande	Redonda	No
Amarillo	Grande	Redonda	Sí
Amarillo	Grande	Redonda	No
Amarillo	Grande	Redonda	No
Amarillo	Grande	Redonda	No
Amarillo	Pequeña	Irregular	Sí
Amarillo	Grande	Irregular	Sí

Es posible realizar esta misma actividad con un dataset más completo que contiene información sobre 8124 variedades de setas disponible en: http://archive.ics.uci.edu/ml/datasets/Mushroom. Cada seta está representada en una fila con datos en 23 columnas. La primera columna indica si es comestible (e, edible) o venenosa (p, poisonous). Las 22 variables restantes indican características como la forma, la textura, el color, el olor, etc. Todas las variables son categóricas discretas (adoptan un único valor). Por ejemplo, la segunda columna indica la forma del sombrero (cap shape) y puede adoptar seis valores (bell=b, conical=c, convex=x, flat=f, knobbed=k, sunken=s). El objetivo de la actividad

es determinar cuál de las 22 características es la que aporta más información sobre el carácter comestible o venenoso de una seta.





La variable tamaño/medida de las setas, aporta más información sobre el carácter comestible o venenoso de esta, con el valor más alto del 0.07336502, seguido de otras variables que no aportan tanta información como el color y la forma que tienen valores más pequeños (0.02).