



# Prediccion y Analisis del Abuso de drogas



# Objetivo:

---

- 1 Analizar características demográficas de cada individuo / observación
- 2 Analizar características rasgos de personalidad de cada individuo / observación
- 3 Conocer las drogas que cada observación está consumiendo y el nivel de consumición
- 4 Realizar una predicción y awareness del consumo de Benzodiazepinas

# El Conjunto de Datos



# ¿Tipos de droga de abuso?

---

Podemos clasificar o dividir las drogas que son abusadas por las personas en tres secciones o pequeños subgrupos totales:

## Estimulantes

- Frecuencia Cardiaca
- Respiratoria
- Actividad Cerebreal
- S.N.C

## Alucinógenas

- Percepción
- Conciencia
- Emociones
- Todas

## Depresoras

- Relajacion
- Somnolencia
- Sedacion



# Rasgos de personalidad

---

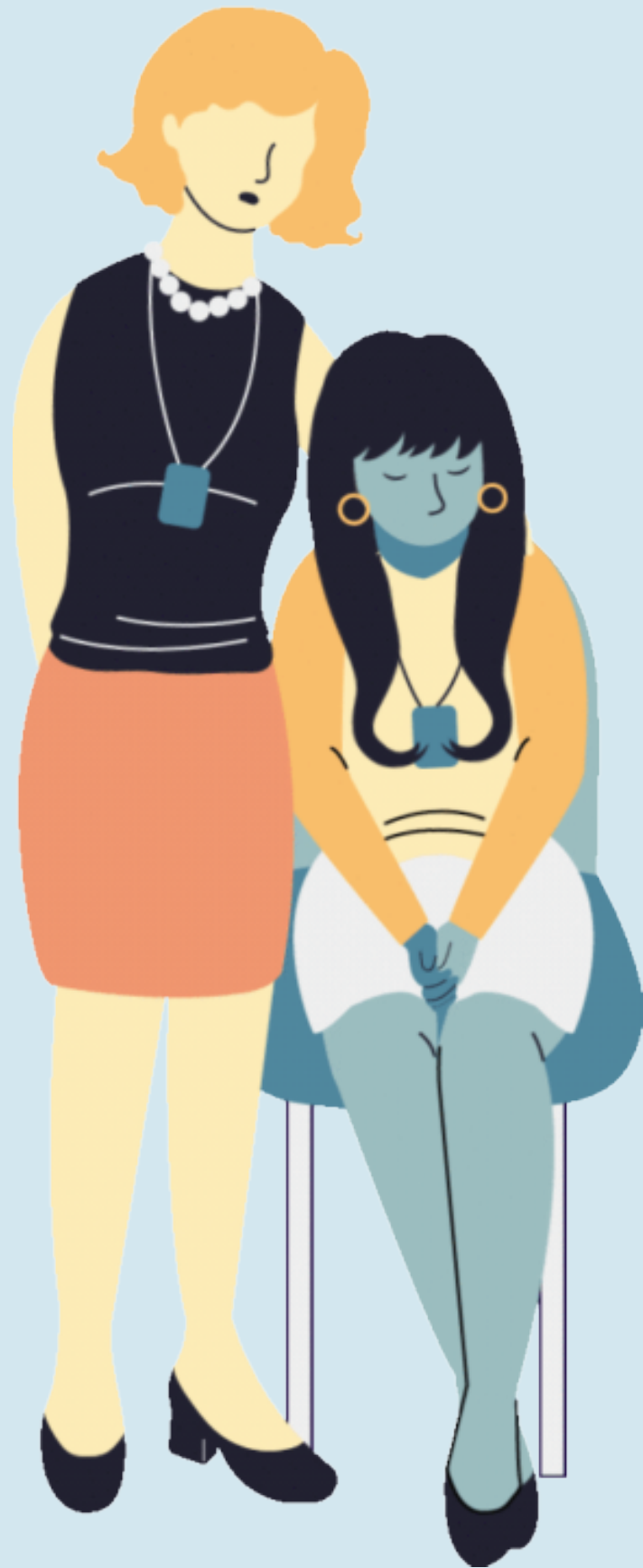
El NEO-FFI-R mide cinco rasgos principales de personalidad:

- Neuroticismo: Tendencia a experimentar emociones negativas como ansiedad, ira o depresión.
- Extroversión: Nivel de sociabilidad, energía y búsqueda de estímulos sociales.
- Apertura a la experiencia: Curiosidad intelectual, creatividad y preferencia por la variedad.
- Amabilidad: Empatía, cooperación y confianza hacia los demás.

El test NEO-FFI-R se basa en preguntas autoinformadas donde las personas califican afirmaciones sobre su comportamiento y emociones, evaluando estos cinco rasgos principales de personalidad.







# Benzodiacepinas

---

Las benzodiacepinas son medicamentos que actúan en el sistema nervioso central al potenciar el efecto del neurotransmisor GABA, lo que produce efectos sedantes, ansiolíticos y anticonvulsivos.

Son interesantes para analizar porque su mecanismo de acción ofrece una comprensión profunda de cómo se regulan los procesos cerebrales relacionados con la ansiedad, el sueño y la actividad convulsiva.

Las podemos subdividir mayormente en Xanax, Valium, Rivotril, etc.

# Verdadera razon por la cual elegi Benzos.

---

Me gusto la complejidad del nombre.

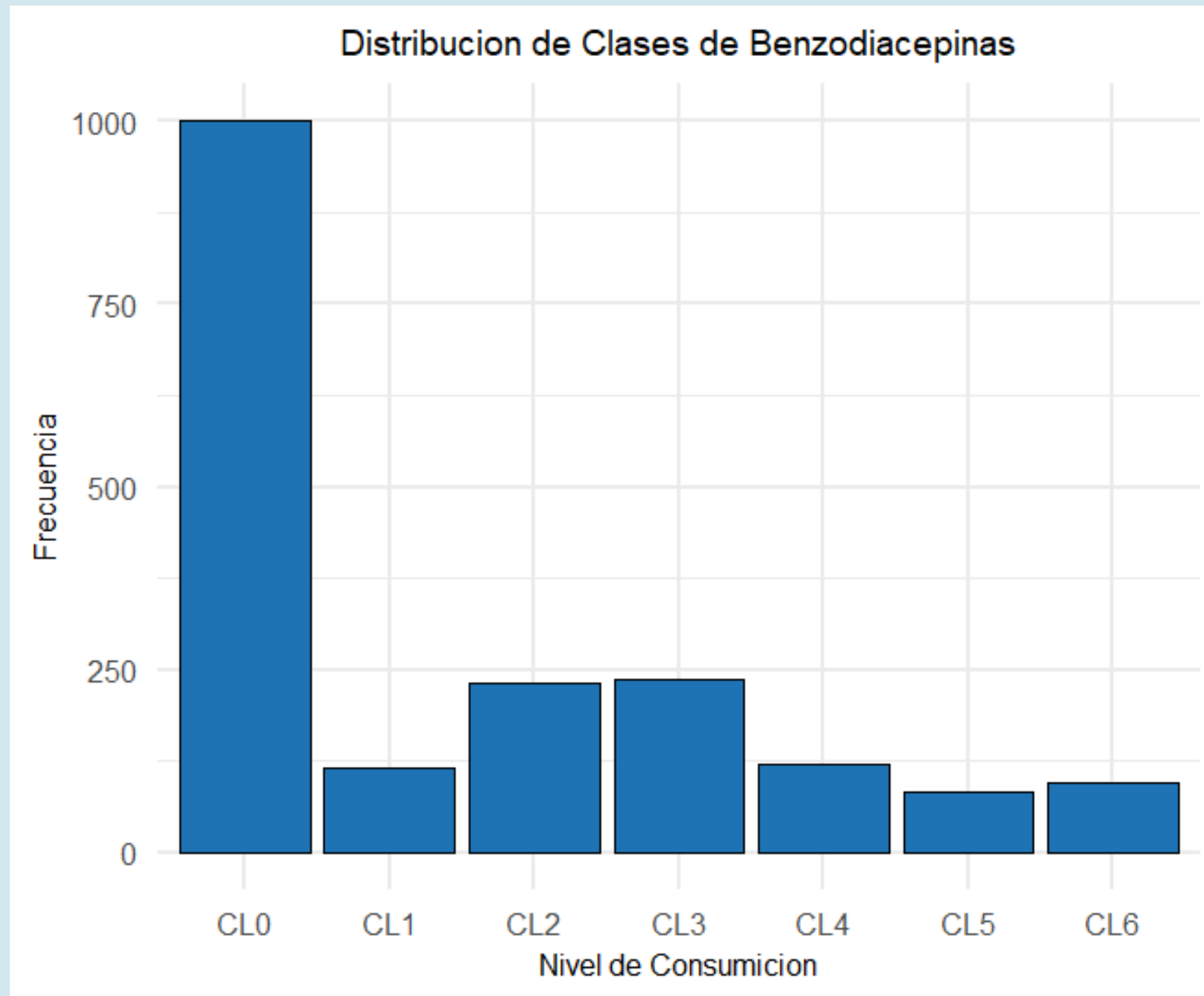


# Analisis de la variable objetivo





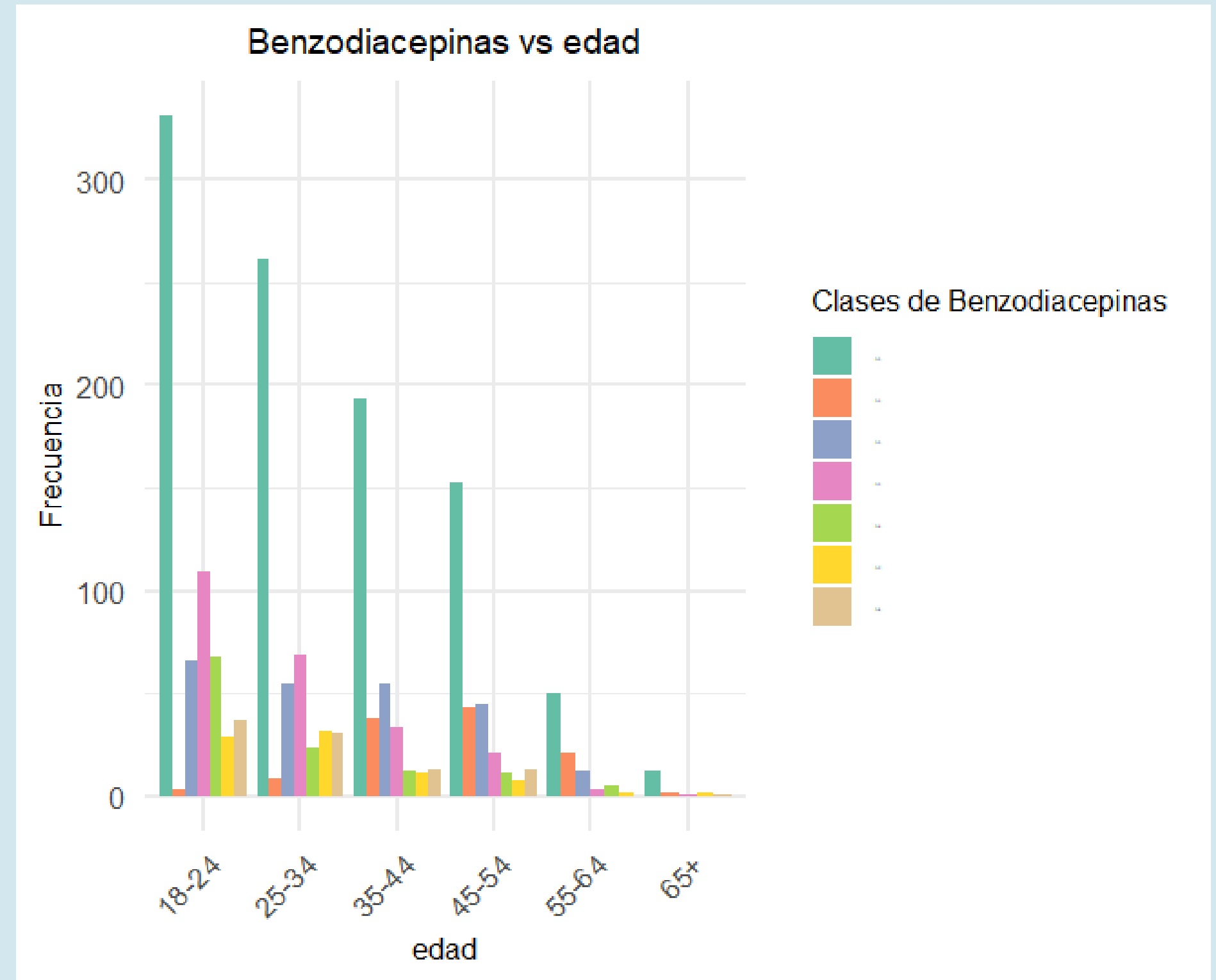
# Distribucion del consumo de benzos



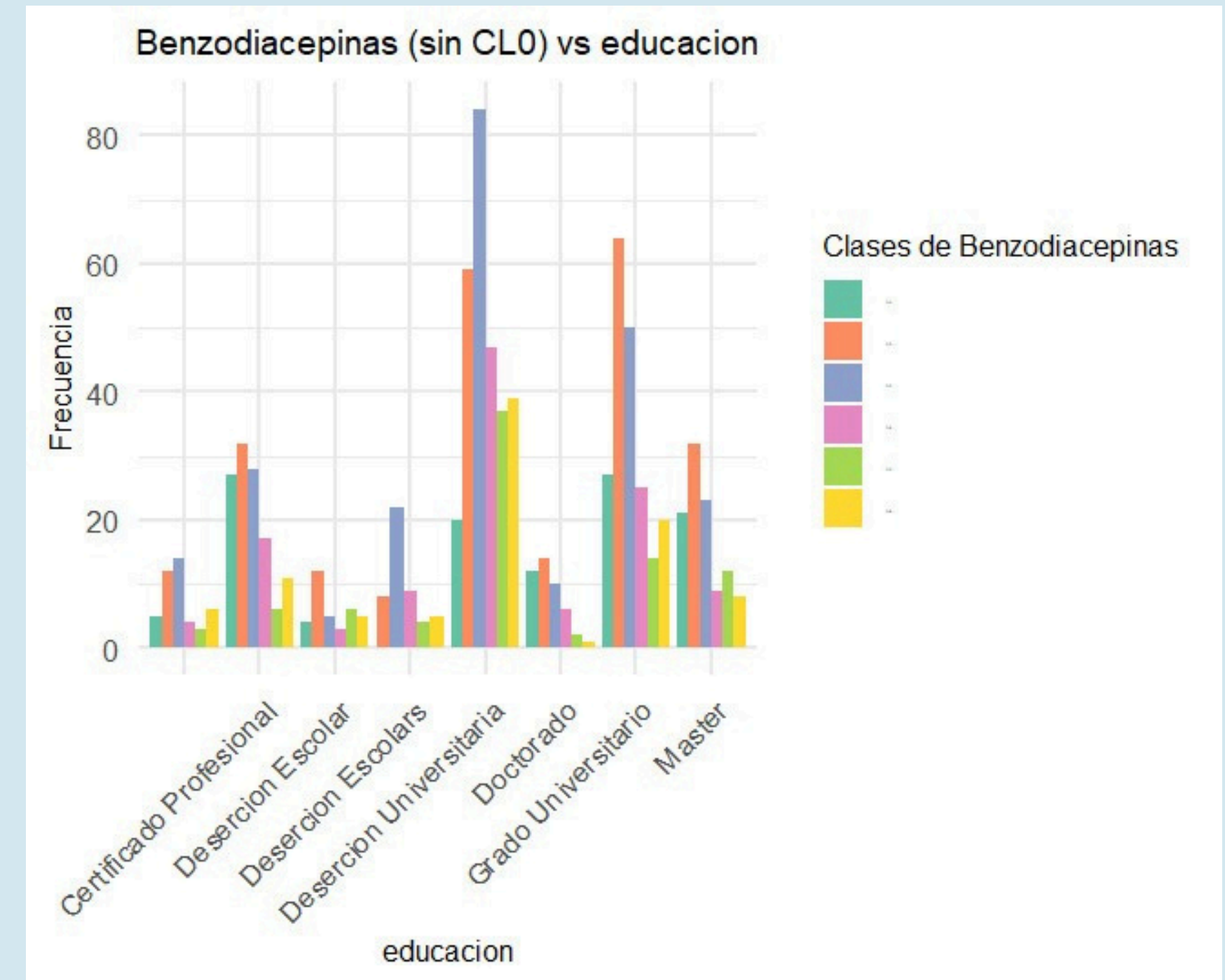
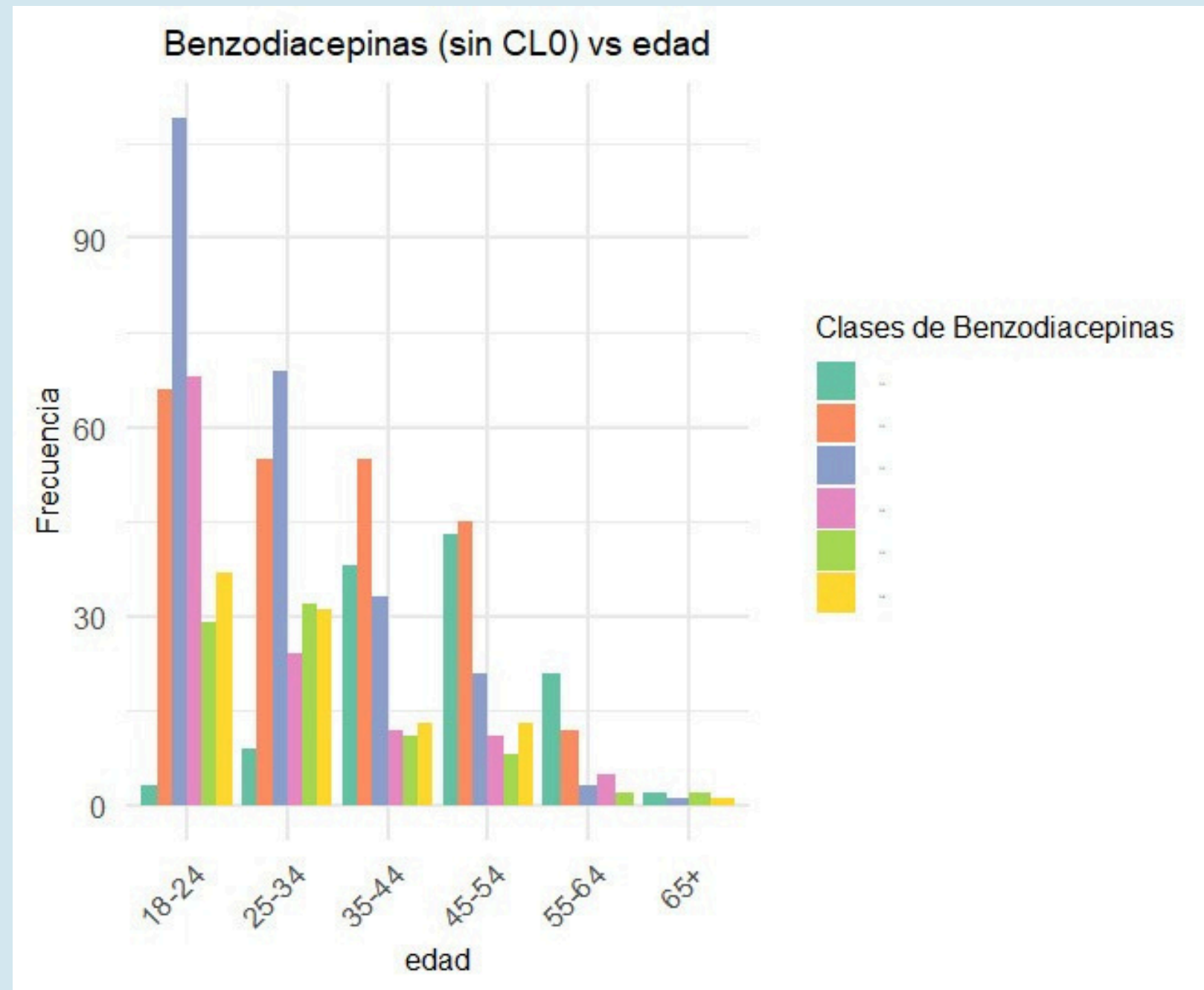
- La mayoría de los casos no consumen este tipo de drogas aproximadamente 1000 / 1880 no lo hacen.
- De los que si lo hacen se ve como:
  - CL1: Hace mas de diez años no lo consume
  - CL2: La ultima vez fue entre dos a diez años anteriores
  - CL3: La ultima vez fue el año anterior
  - CL4: Lo consumen una cada mes
  - CL5: Lo consumen unca cada semana
  - CL6: Lo consumen cada dia

# Benzos vs Edad

- Notamos un claro patrón en donde las personas de baja edad son mas propensas a consumir este tipo de drogas
- Las personas de edad media son mas propensas a volverse medianamente adictiva
- A mayor edad menos probabilidad de consumir estas drogas
- Entre las edad 45-55, por mas que no haya tanto casos notamos una igualdad entre CL2 y CL3



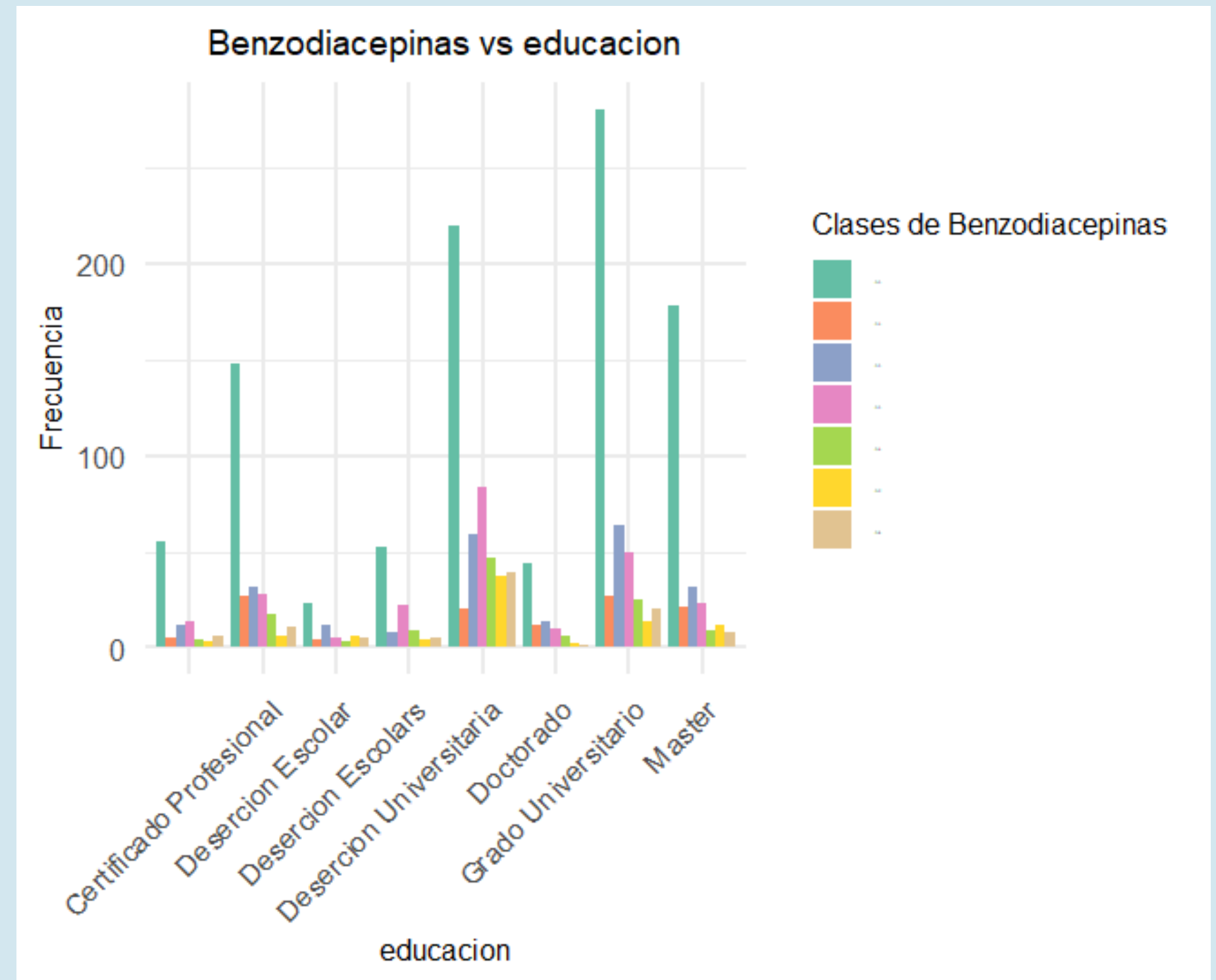
# Benzos vs Educacion



## Benzos vs Edad

# Benzos vs Educacion

- El nivel de “Grado Universitario” y “Master” presenta una mayor frecuencia general de no consumicion de benzodiazepinas
- Las deserciones educativas (escolar y universitaria) tambien tienen frecuencias relevantes, aunque significativamente menores que los niveles superiores.
- Se observa una distribucion variada en las categorias de educacion, lo que sugiere que el uso de benzodiazepinas no está limitado a un único nivel educativo.

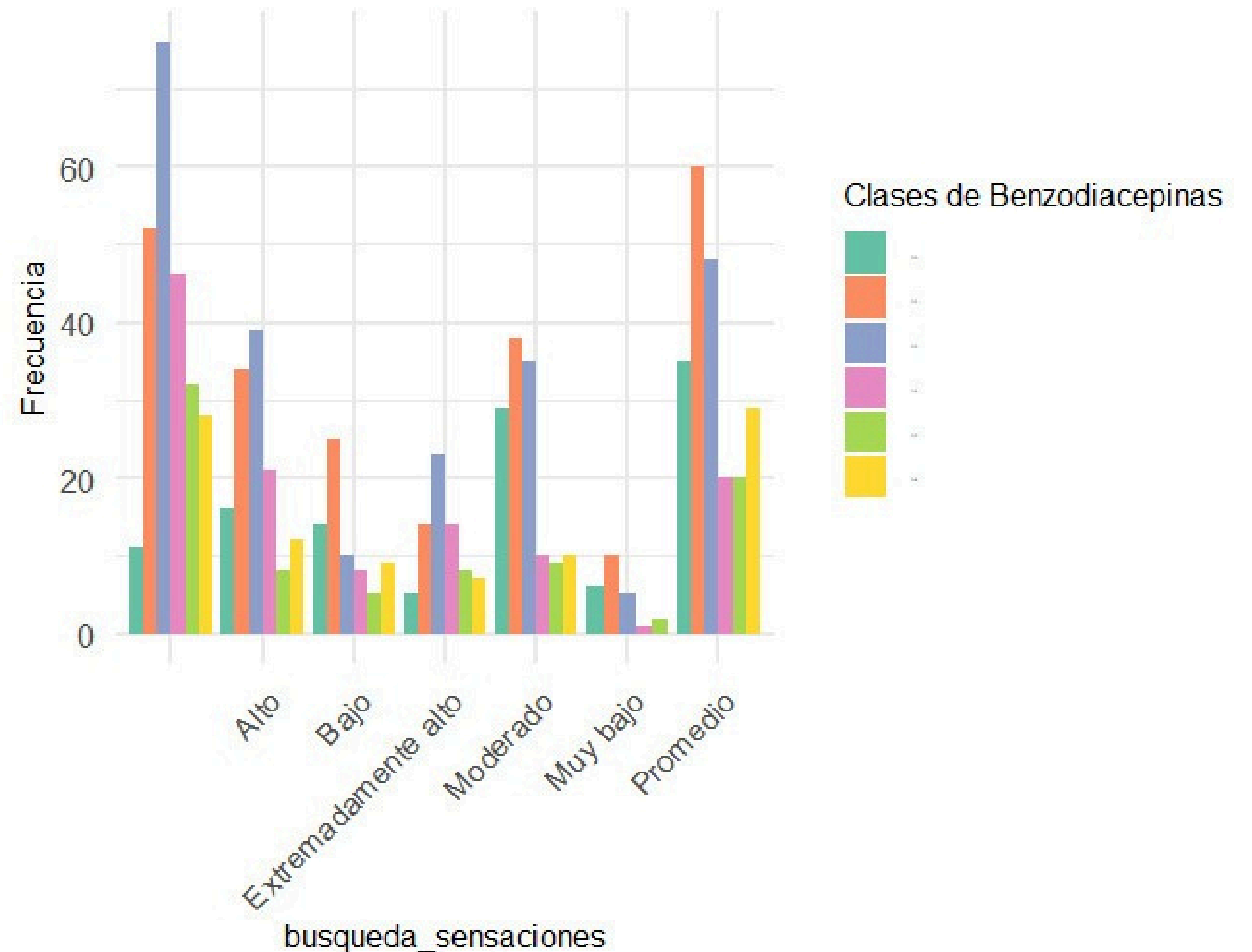


# Benzos vs SS

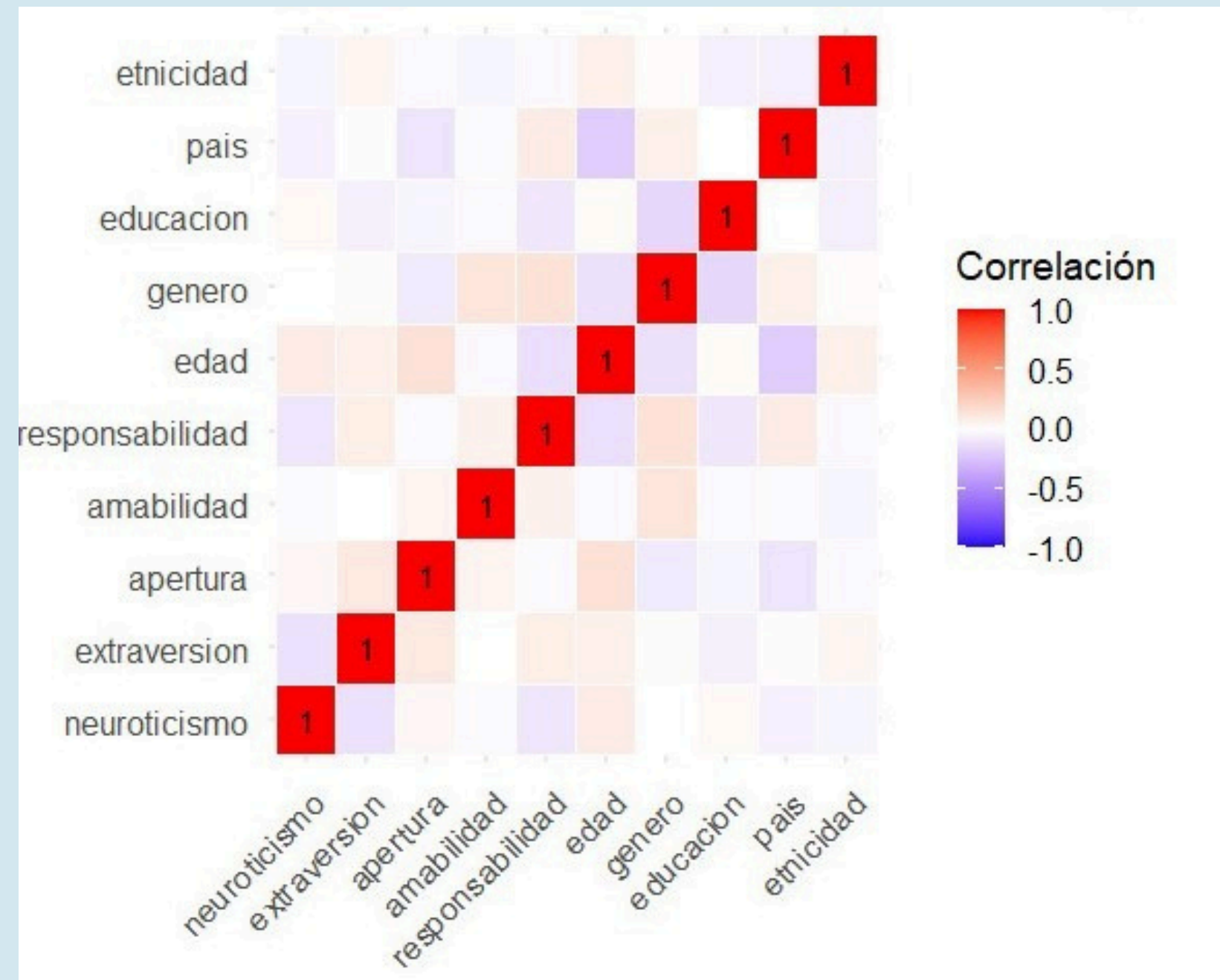
## Medida en ImpSS

- El nivel extremadamente alto presenta una frecuencia destacada, especialmente para ciertas clases de benzodiazepinas (marcadas por el color predominante).
- Se observa un aumento significativo en la frecuencia de consumo, con una distribución variada entre las clases.
- Los niveles muy bajo y bajo tienen una menor representación en comparación con otros niveles.
- La búsqueda de sensaciones podría influir en la elección y frecuencia de uso de ciertas benzodiazepinas.

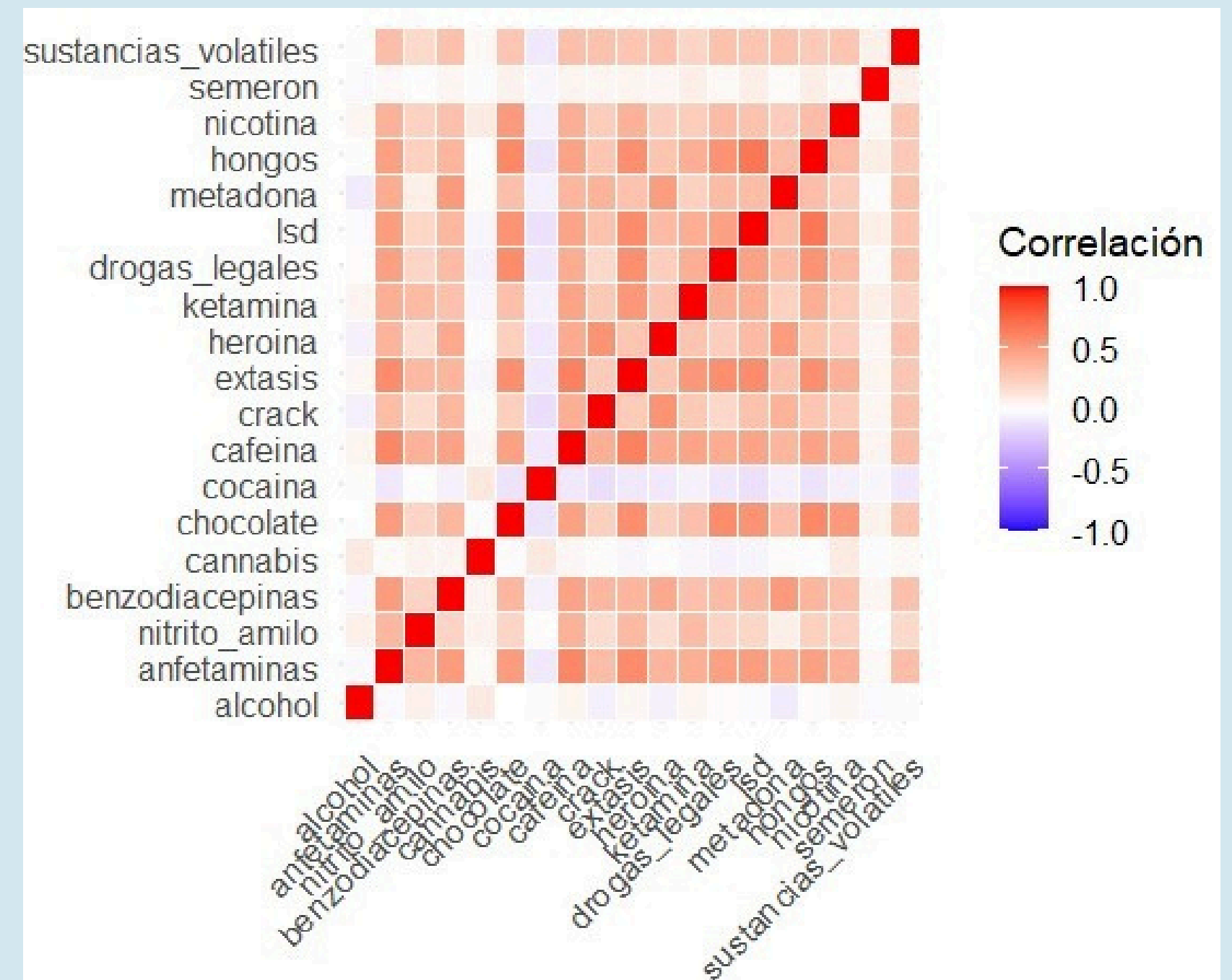
Benzodiazepinas (sin CL0) vs busqueda\_sensaciones







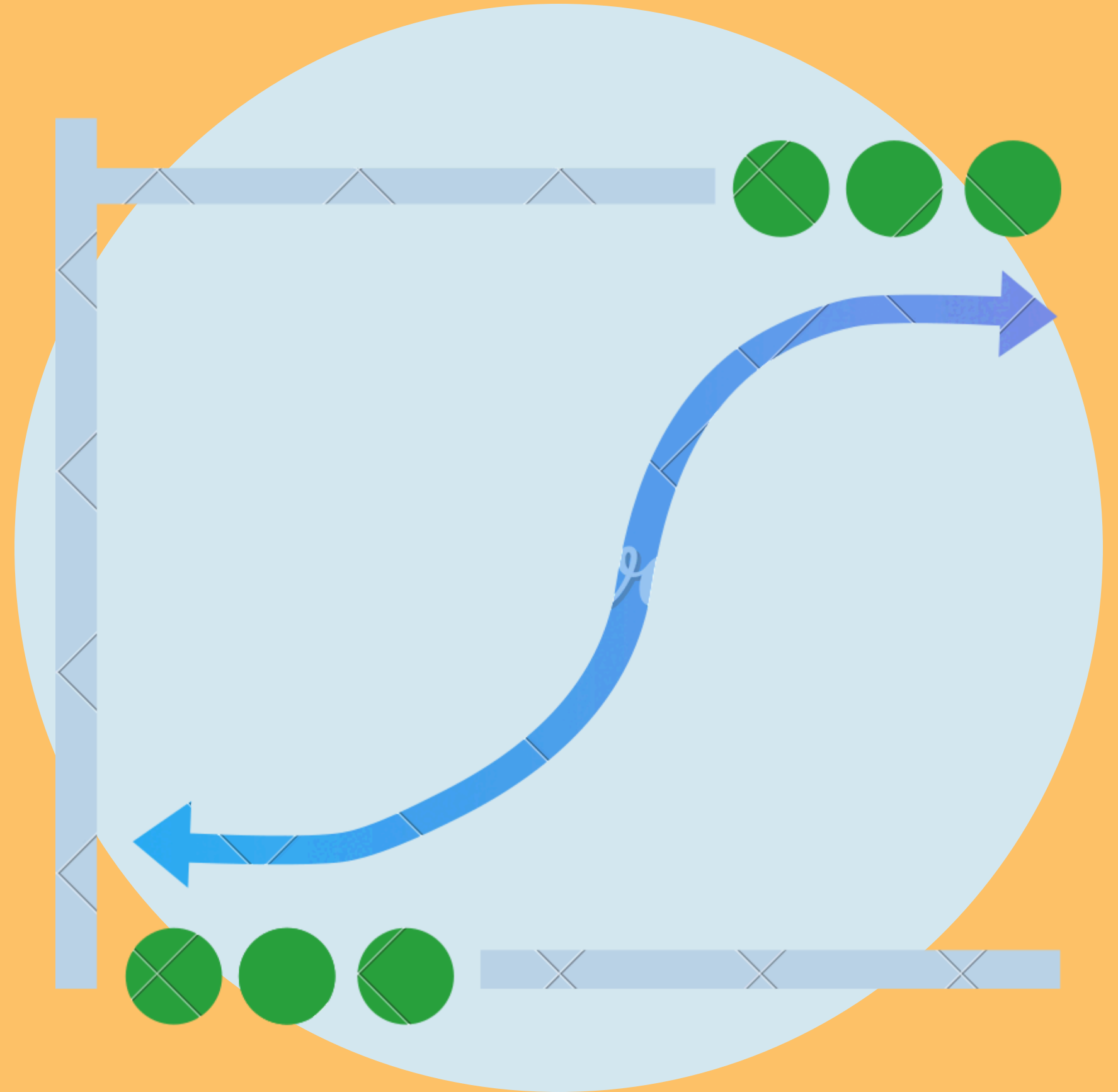
## Heatmap DC

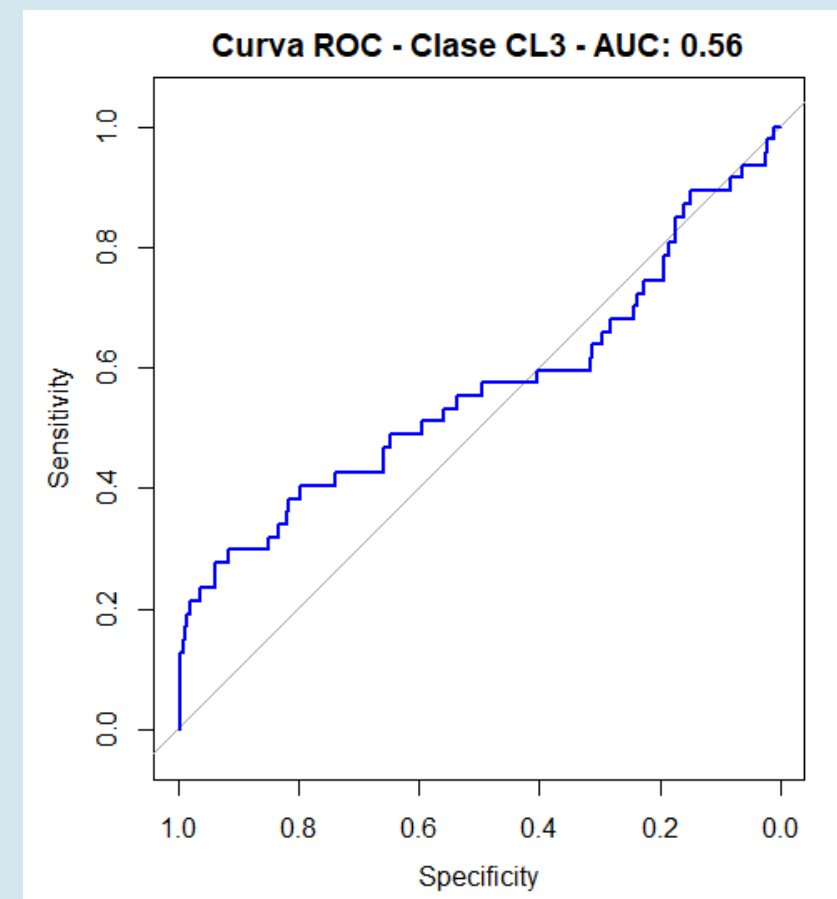
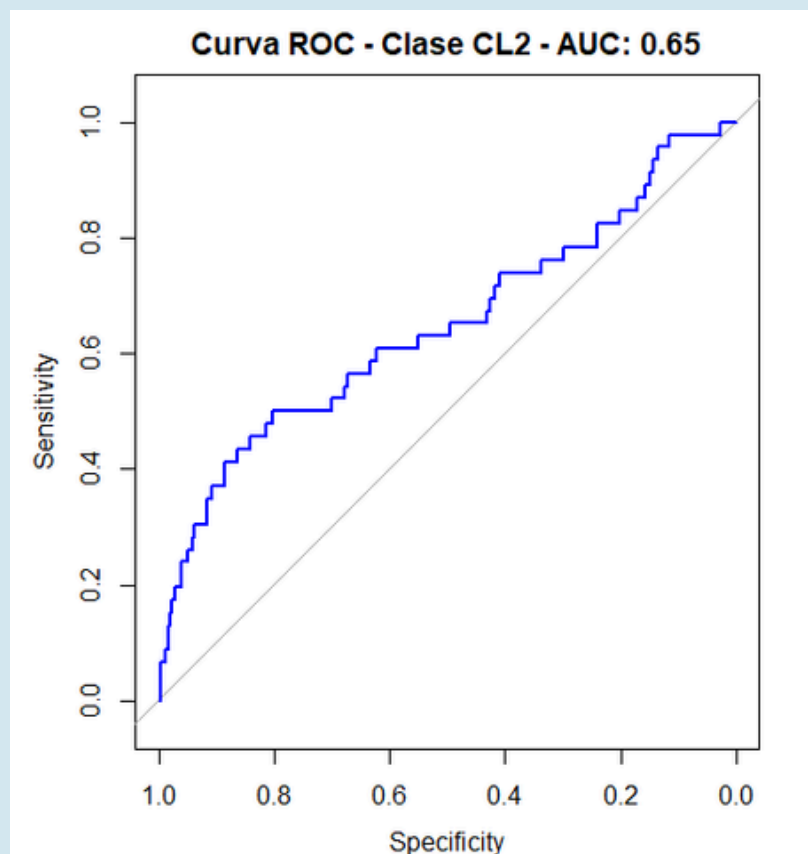
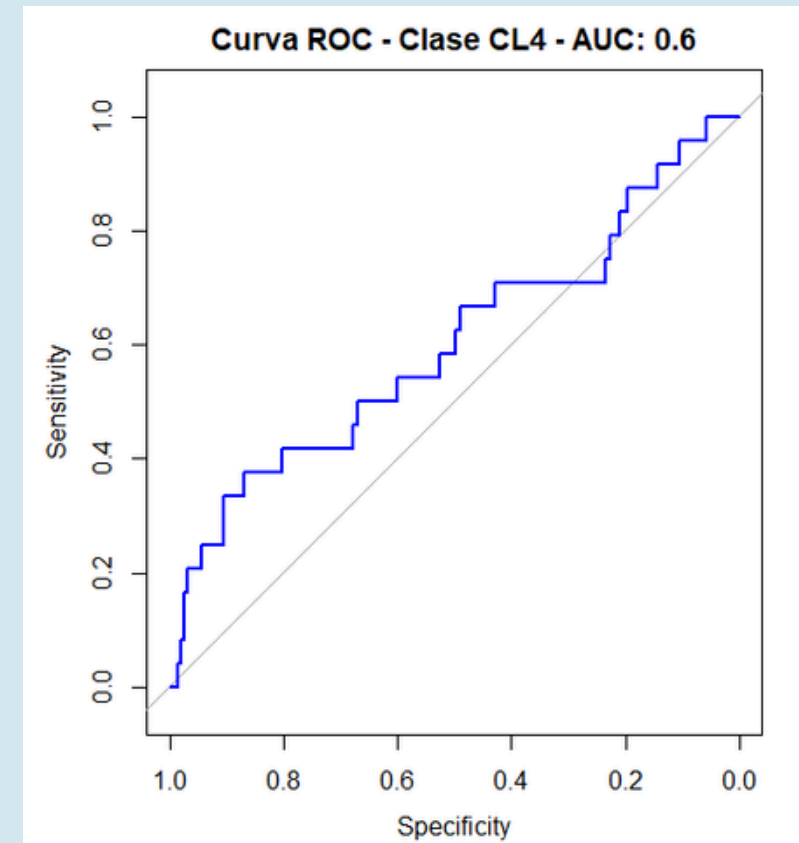
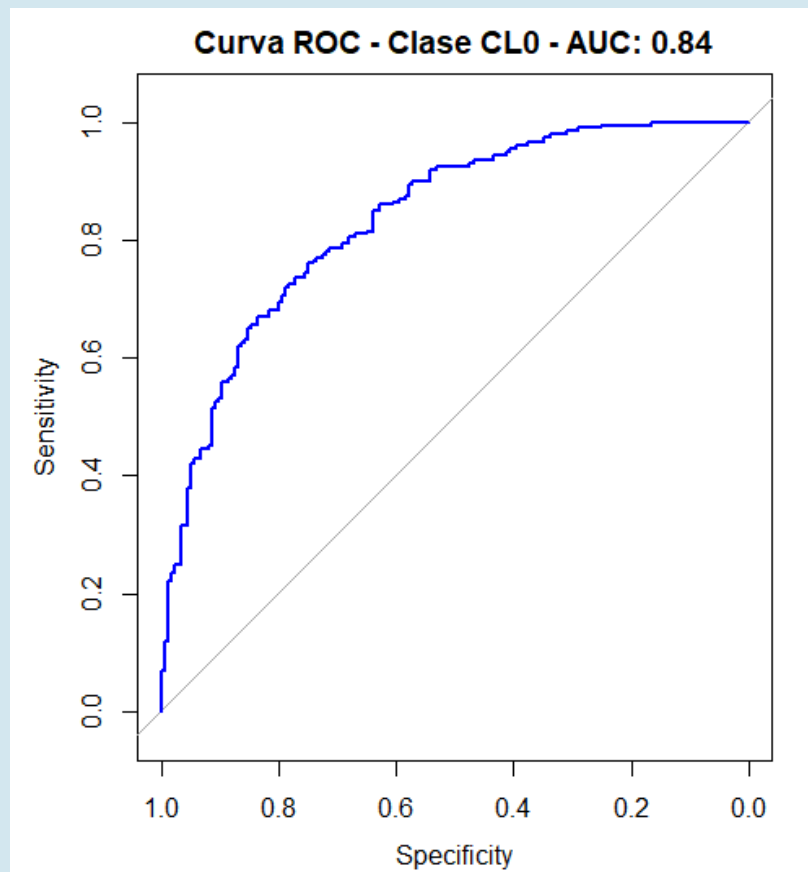


## Heatmap DyRP



# Regresión Multiclase



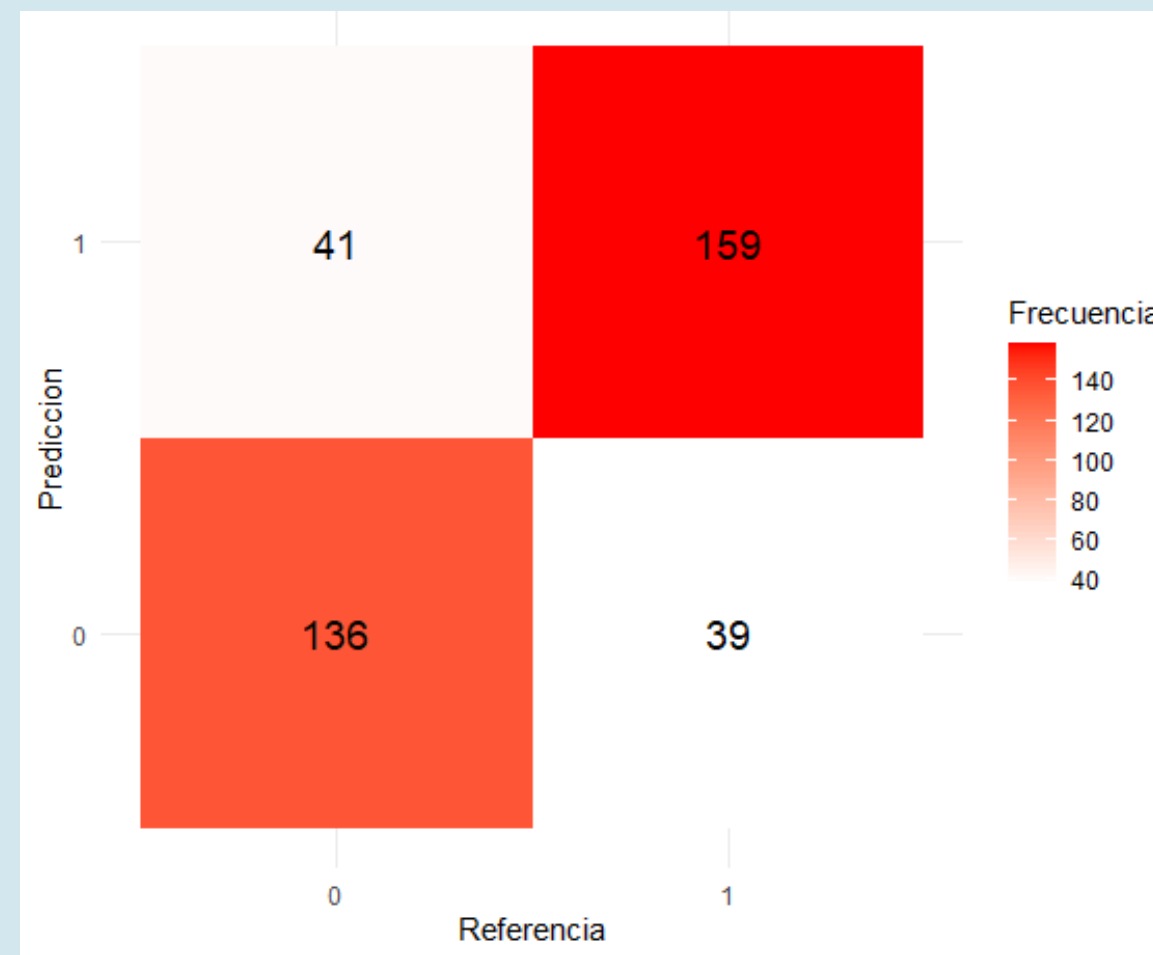


# ACCURACY: 53,16%

- Gran nivel de discriminacion con niveles de consumo bajos como CL0 Y CL1
- Nivel muy bajo de discriminacion con niveles mas altos como CL3, CL4, CL5 y CL6
- Debido al desbalanceo de datos no se puede construir un buen modelo de regresion logistica ordinal

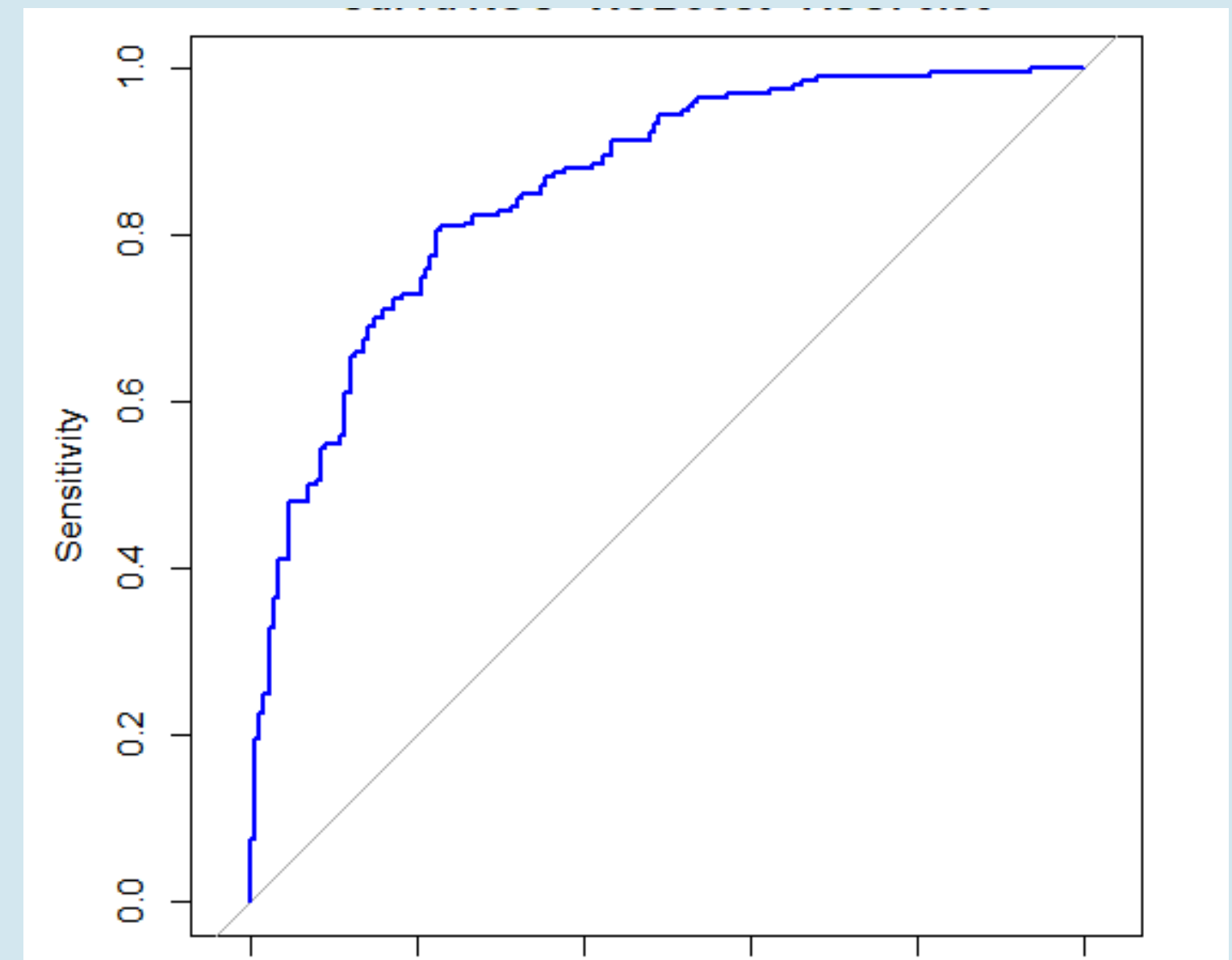
# XGBOOST



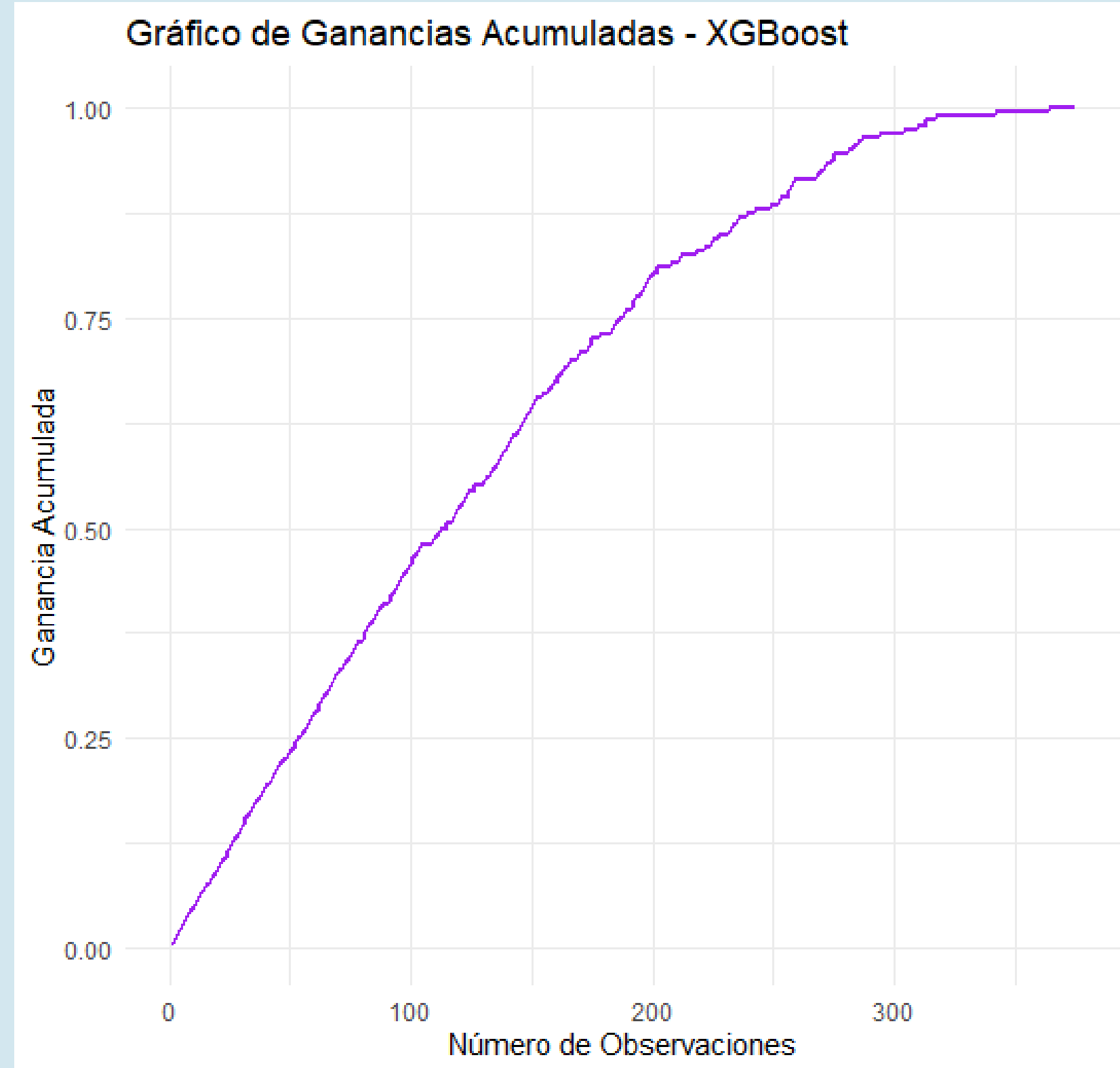


- Tiene un buen desempeño general.
- Los mayores aciertos están en las categorías correctamente clasificadas: 159 verdaderos positivos (1 y 1) y 136 verdaderos negativos (0 y 0).
- Los errores son menores, con 41 falsos negativos (1 y 0) y 39 falsos positivos (0 y 1).
- Predice con mayor precisión ambas clases, mostrando una buena capacidad de clasificación.
- Accuracy del modelo XGBoost es 78.67%

- La curva ROC del modelo XGBoost muestra un desempeño sólido con un área bajo la curva (AUC) de 0.86.
- Esto indica que el modelo tiene una buena capacidad para diferenciar entre las clases positivas y negativas.
- La curva se acerca al extremo superior izquierdo, lo que refuerza su buen rendimiento en términos de sensibilidad y especificidad.

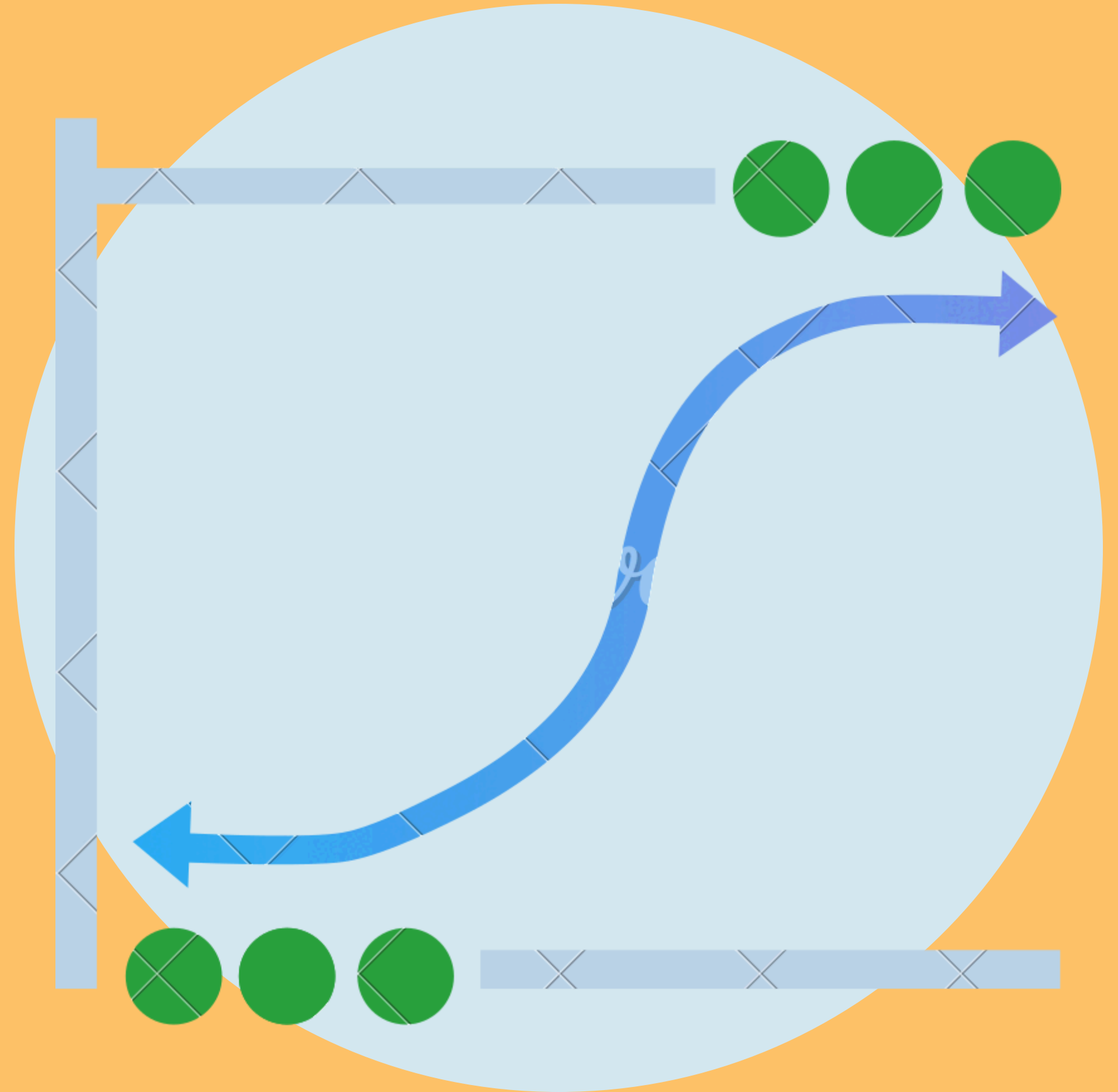


# Ganancias Acumuladas



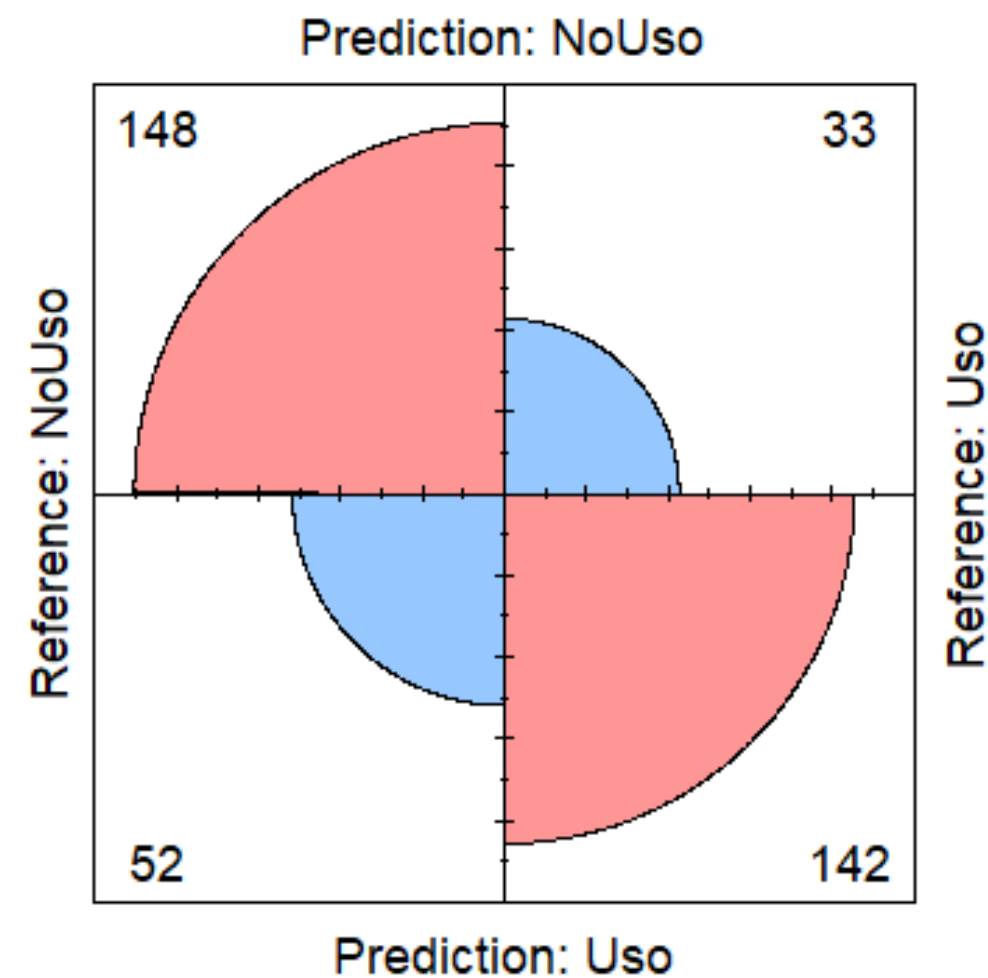
- El gráfico de ganancias acumuladas refleja cómo el modelo XGBoost identifica correctamente las observaciones relevantes a medida que aumenta el número de datos procesados.
- En las primeras observaciones, la curva crece rápidamente, lo que indica que el modelo es capaz de capturar una proporción significativa de las respuestas positivas al inicio.
- Conforme se procesan más observaciones, la pendiente de la curva se reduce, estabilizándose cerca de 1.0, lo que indica que el modelo ha capturado casi todas las ganancias posibles.
- Este comportamiento demuestra que el modelo es eficiente en priorizar correctamente las instancias más importantes al principio, siendo adecuado para problemas donde es crucial identificar rápidamente las observaciones más relevantes.

# Regression Logistica



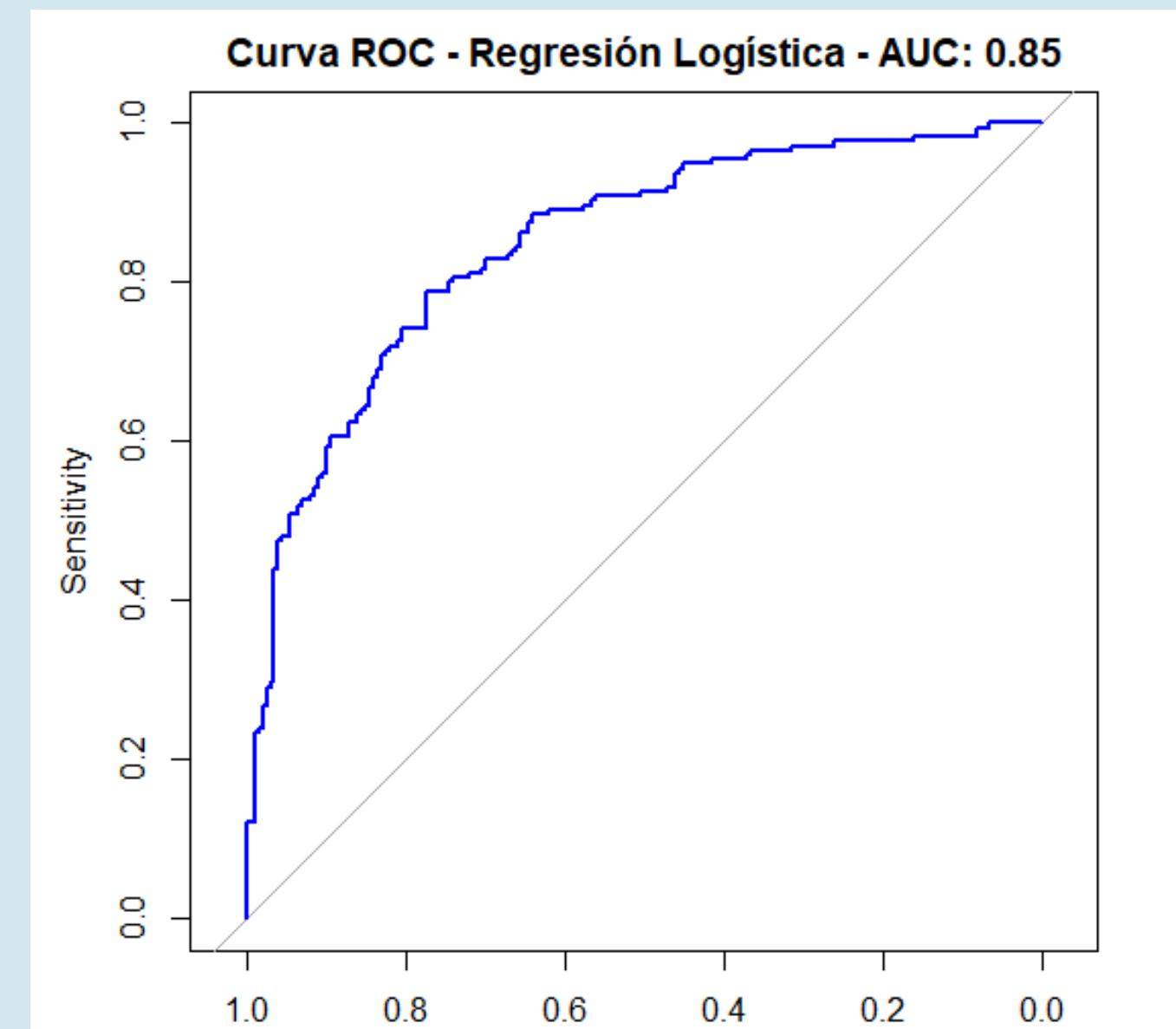


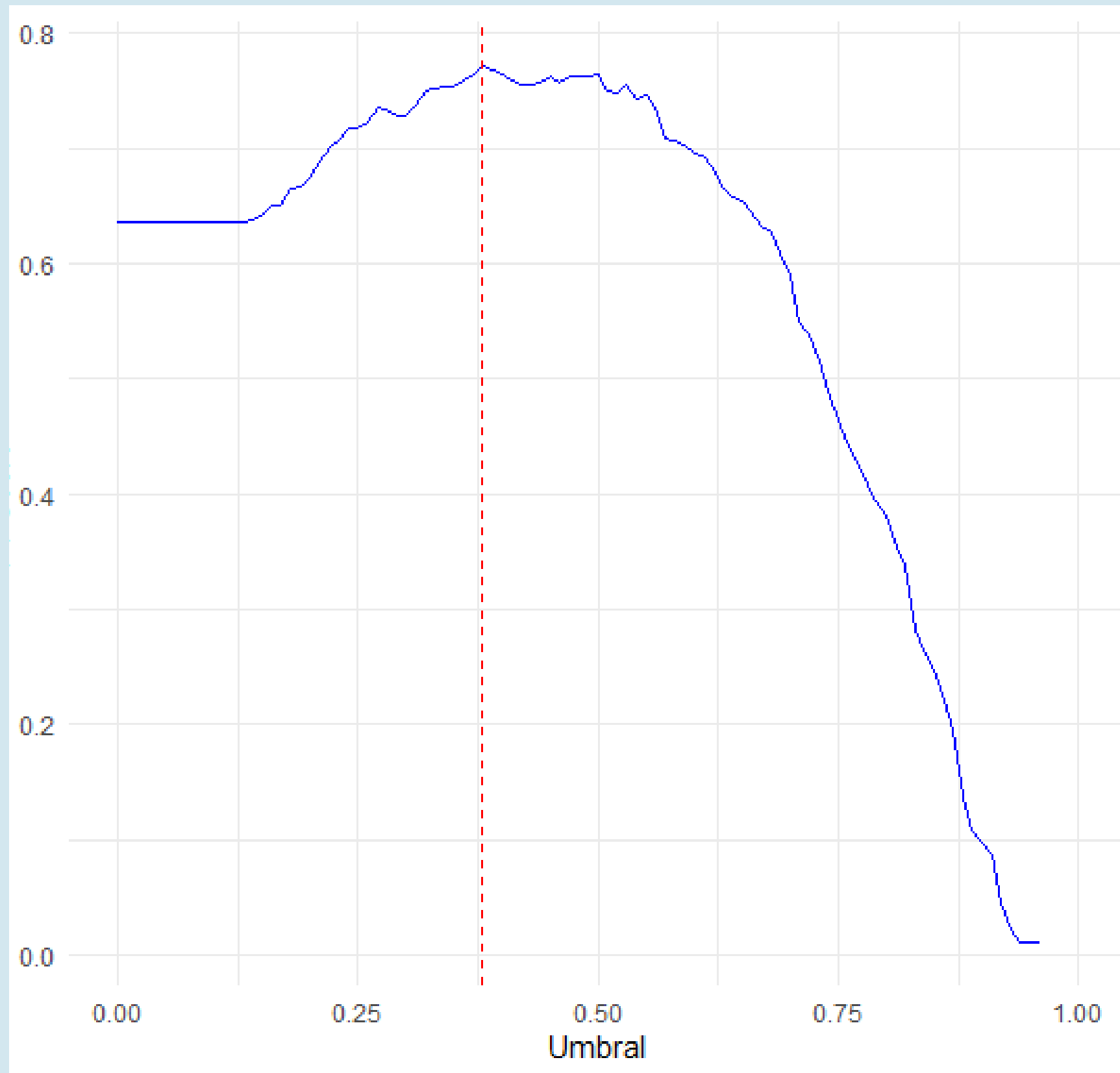
## Matriz de Confusión



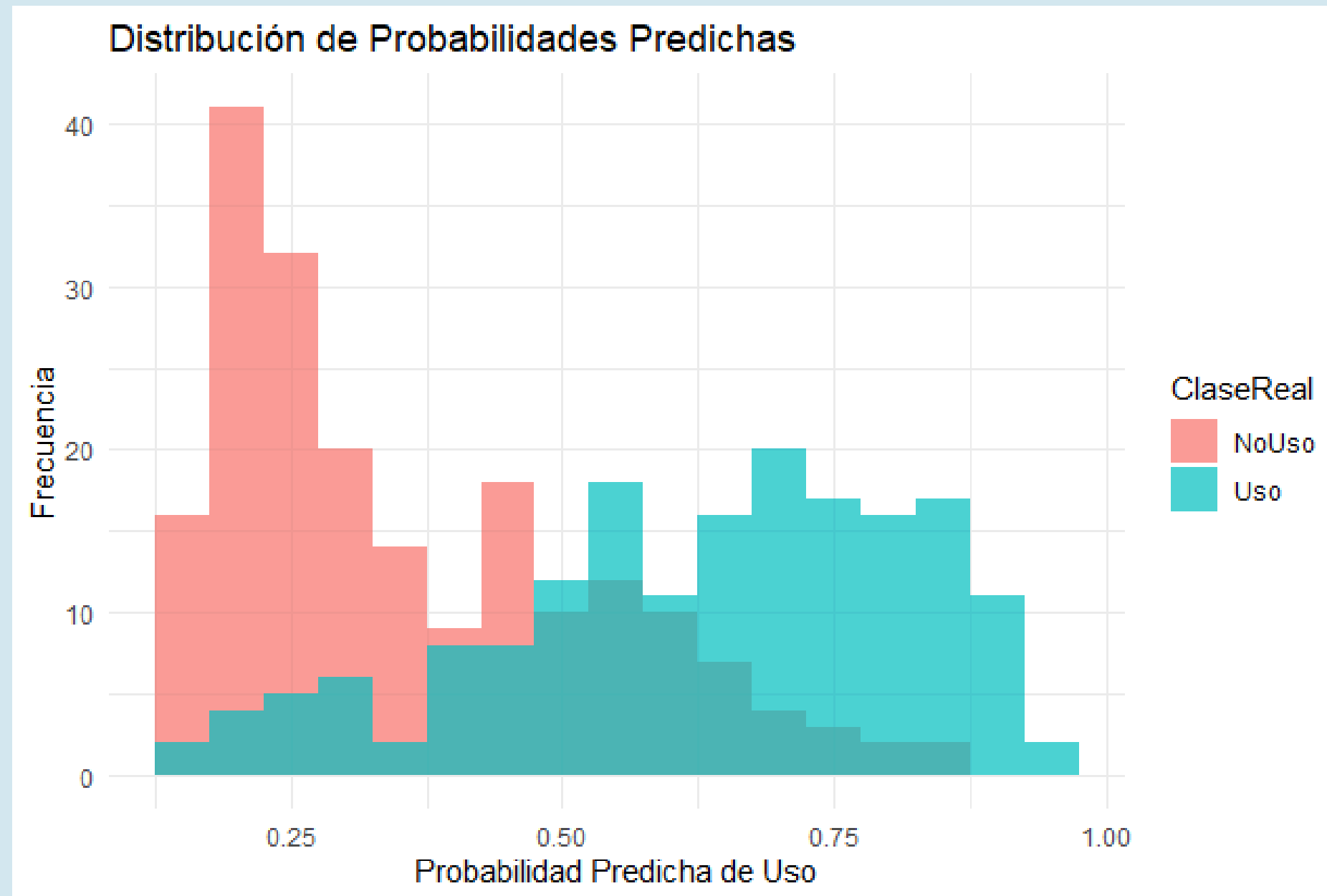
- La matriz de confusión indica un buen desempeño del modelo.
- Aciertos: 148 casos de NoUso y 142 casos de Uso fueron correctamente clasificados.
- Errores: 33 falsos negativos (predichos como NoUso pero realmente eran Uso) y 52 falsos positivos (predichos como Uso pero realmente eran NoUso).
- Los aciertos son mayores que los errores, lo que sugiere que el modelo tiene un buen balance general en la clasificación.
- Accuracy del modelo es 77.33%

- La curva ROC para la regresión logística muestra un área bajo la curva (AUC) de 0.85, indicando un buen rendimiento del modelo.
- La curva se aproxima al extremo superior izquierdo, refleja una alta capacidad del modelo para distinguir entre las clases Uso y NoUso.
- El modelo tiene un buen equilibrio entre sensibilidad y especificidad, adecuado para tareas de clasificación en este contexto.





- Muestra cómo varía el F1-Score en función del umbral de clasificación utilizado por el modelo.
- Alcanza su valor máximo cercano a 0.8 alrededor de un umbral en el rango de 0.4 a 0.5, indicado por la línea roja vertical.
- A medida que el umbral se aleja de este punto óptimo (hacia valores más altos o más bajos), el F1-Score disminuye significativamente, reflejando una reducción en el equilibrio entre precisión y recall.
- Esto sugiere que un umbral cercano a 0.5, específicamente 0.47, ofrece el mejor desempeño del modelo, maximizando la efectividad en clasificaciones correctas mientras se minimizan los errores.



- El gráfico muestra la distribución de probabilidades predichas por el modelo para las clases Uso y NoUso.
- La clase NoUso (rojo) se concentra en probabilidades más bajas, principalmente entre 0.1 y 0.4.
- La clase Uso (azul) se distribuye en probabilidades más altas, predominando entre 0.6 y 0.9.
- Existe una zona de superposición entre 0.4 y 0.6, donde las predicciones son menos claras, lo que podría generar errores de clasificación.
- En general, el modelo separa razonablemente bien las dos clases, pero podría mejorarse la discriminación en la región de superposición.

# Conclusion



Los niveles educativos más altos (como grado universitario) y ciertas búsquedas de sensaciones están fuertemente relacionados con el uso de benzodiazepinas. Estas variables podrían estar influyendo positivamente en el desempeño de los modelos.

El modelo XGBoost fue el más efectivo, con un AUC de 0.86 y un accuracy de 78.67%, superando ligeramente al modelo de regresión logística, que obtuvo un AUC de 0.85 y un accuracy de 77.33%. Esto demuestra que ambos modelos funcionan bien para la tarea, pero XGBoost tiene una ventaja en la clasificación.

Lograron una separación adecuada entre las clases Uso y NoUso. Sin embargo, se identificó una zona de incertidumbre en las probabilidades predichas entre 0.4 y 0.6, donde ambos modelos cometieron la mayoría de los errores.

XGBoost es el más eficiente para esto, mostrando mejor balance entre sensibilidad y especificidad. La regresión logística, aunque ligeramente menos precisa, sigue siendo una alternativa válida, especialmente si se prioriza interpretabilidad.

Y de la Regresión multiclase ni hablemos.

**Muchas  
Gracias!**

