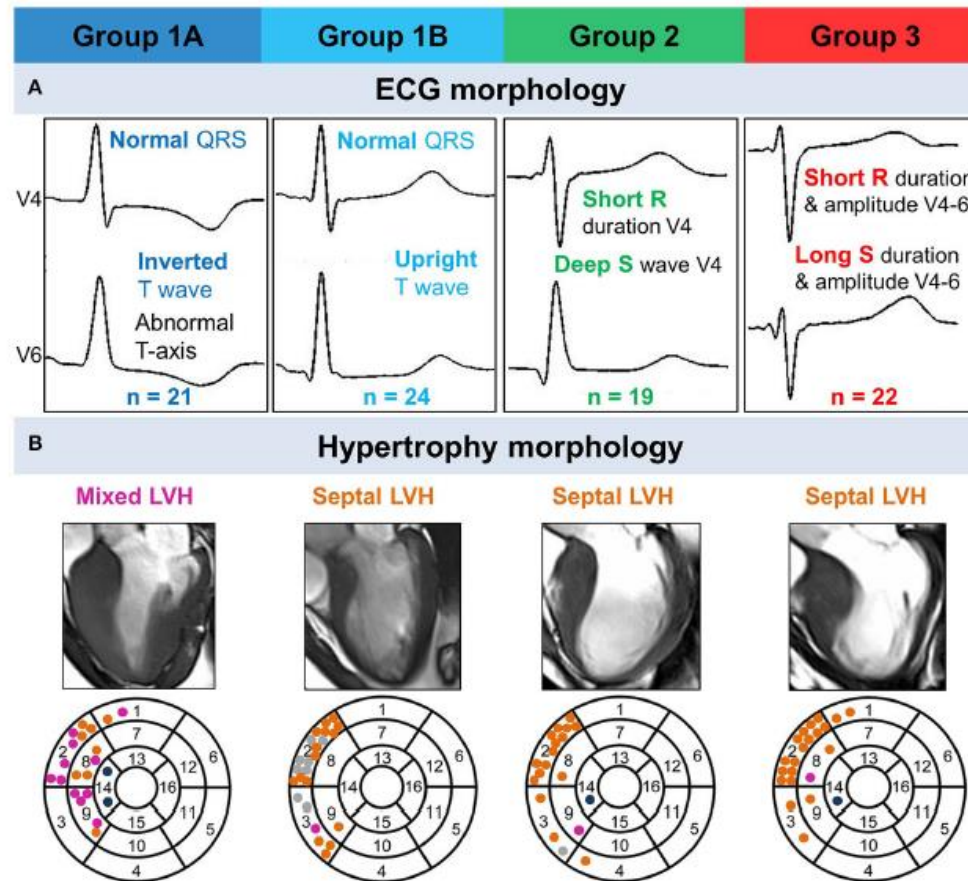


# Diagnóstico de hipertrofia ventricular izquierda mediante ML



# HVI y ML. Datos

---

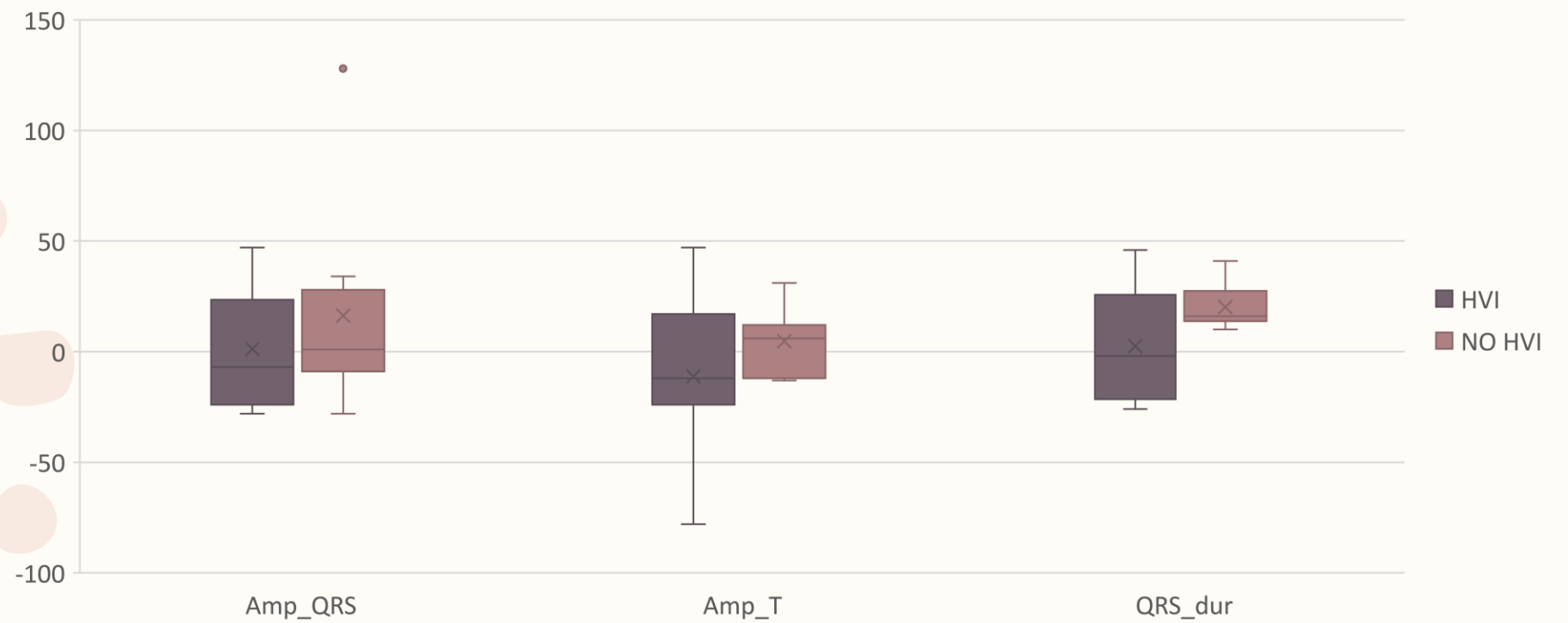
- 45 pacientes. Cada paciente es una estructura "Pt" en matlab con varios campos. Los campos de interés son:
- Pt.ECG: las 12 derivaciones de ECG de cada paciente
- Pt.fm: la frec de muestreo
- Pt.labels: la etiqueta de si es hipertrofico o no (0: sano, 1: hipertrofico)
- Pt.fiduciales: los puntos fiduciales del ECG (les puede servir para analizar complejo QRS por un lado y onda T por el otro). Descartar los registros que tengan menos de 10 puntos fiduciales por cada onda.

# Consigna HVI y ML

---

1. Preprocesar los datos mediante filtros adecuados para el ECG
2. Detectar los complejos QRS de los registros
3. Extracción de características en complejos QRS y ondas T (sugerencia: pueden ser las amplitudes y las duraciones de cada onda)
4. Separación de las características mediante algoritmo de clustering Kmeans (lo vimos en clase) con  $K = 2$  (un grupo hipertrófico y otro no)
5. Estadística de resultados sobre cada grupo. Para esto hagan graficos de cajas con las medias y duraciones de cada onda en cada grupo, y test de hipotesis (t-test por ej.) para saber si son significativamente diferentes, así se podrá apreciar si el clustering logró separar las características efectivamente o no.
6. Clasificación de los grupos mediante un algoritmo supervisado. Armado de la matriz de características y labels. Permutación de las mismas, separación en 70-30 para entrenamiento-testeo. Obtención de la matriz de confusión.

# Gráficos de cajas



# Matriz de confusión (Aprendizaje supervisado)

Matriz de confusión		Estimado por el modelo			
		Negativo (N)	Positivo (P)		
Real	Negativo	a: (TN)	b: (FP)		
	Positivo	c: (FN)	d: (TP)	<b>Precisión ("precision")</b> Porcentaje predicciones positivas correctas:	$d/(b+d)$
		<b>Sensibilidad, exhaustividad ("Recall")</b> Porcentaje casos positivos detectados	<b>Especificidad ("Specificity")</b> Porcentaje casos negativos detectados	<b>Exactitud ("accuracy")</b> Porcentaje de predicciones correctas <i>(No sirve en datasets poco equilibrados)</i>	
		$d/(d+c)$	$a/(a+b)$	$(a+d)/(a+b+c+d)$	

# Matriz de confusión (Aprendizaje supervisado)

---

	HVI_estim	NO HVI_estim
HVI_real		
NO HVI_real		

Presentar sensibilidad,  
especificidad,  
Exactitud y precisión