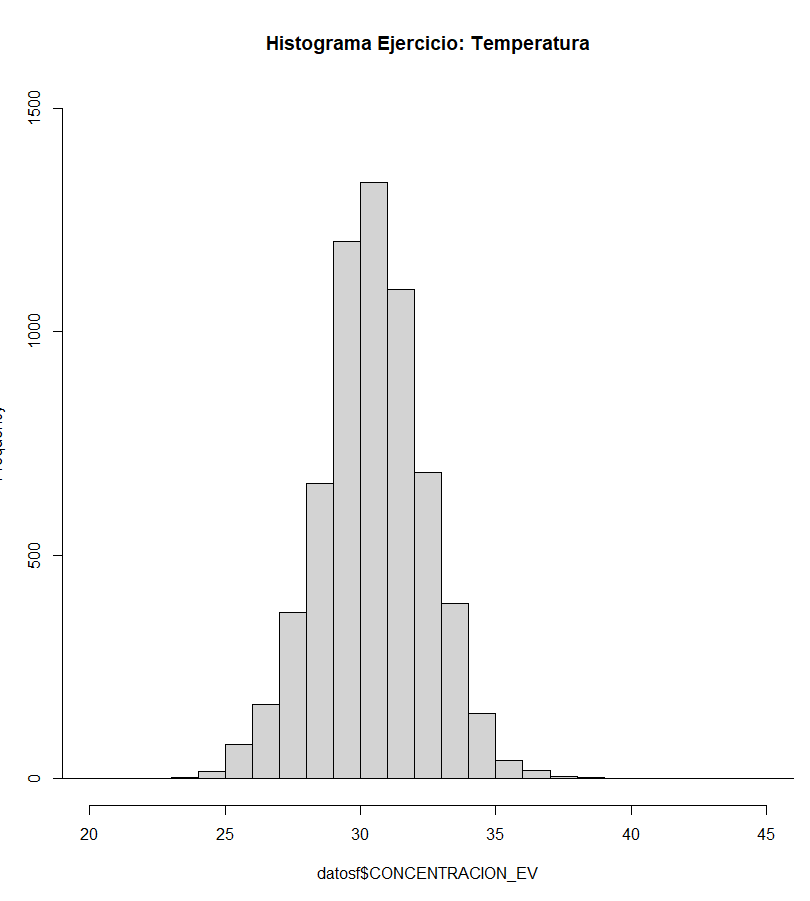
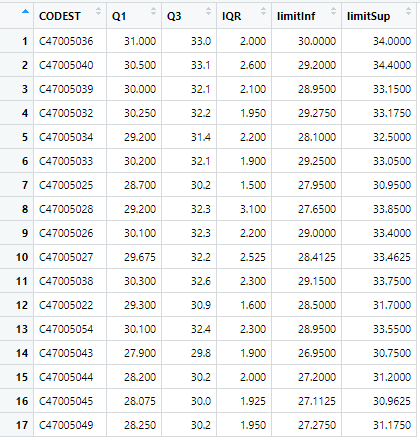
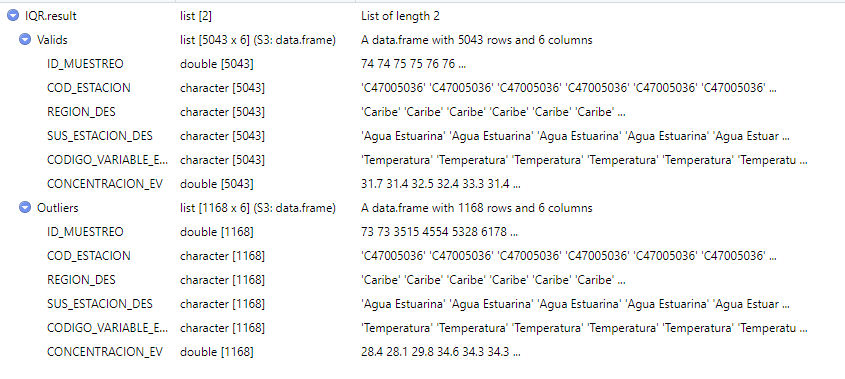
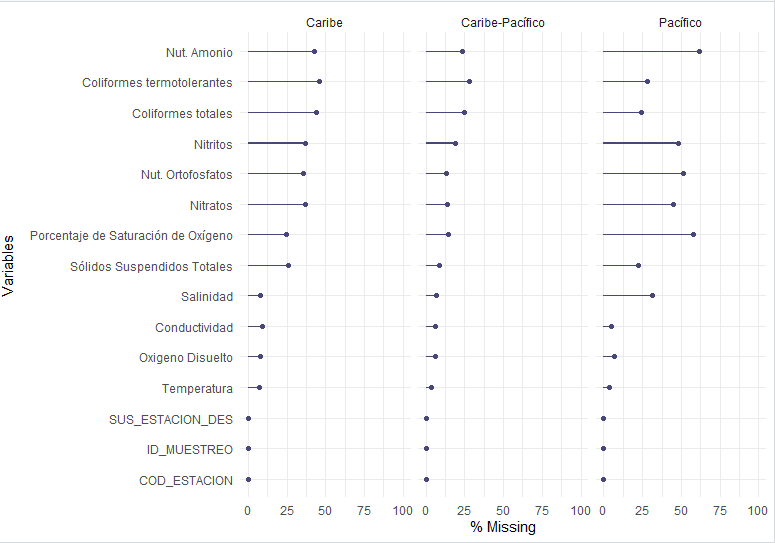
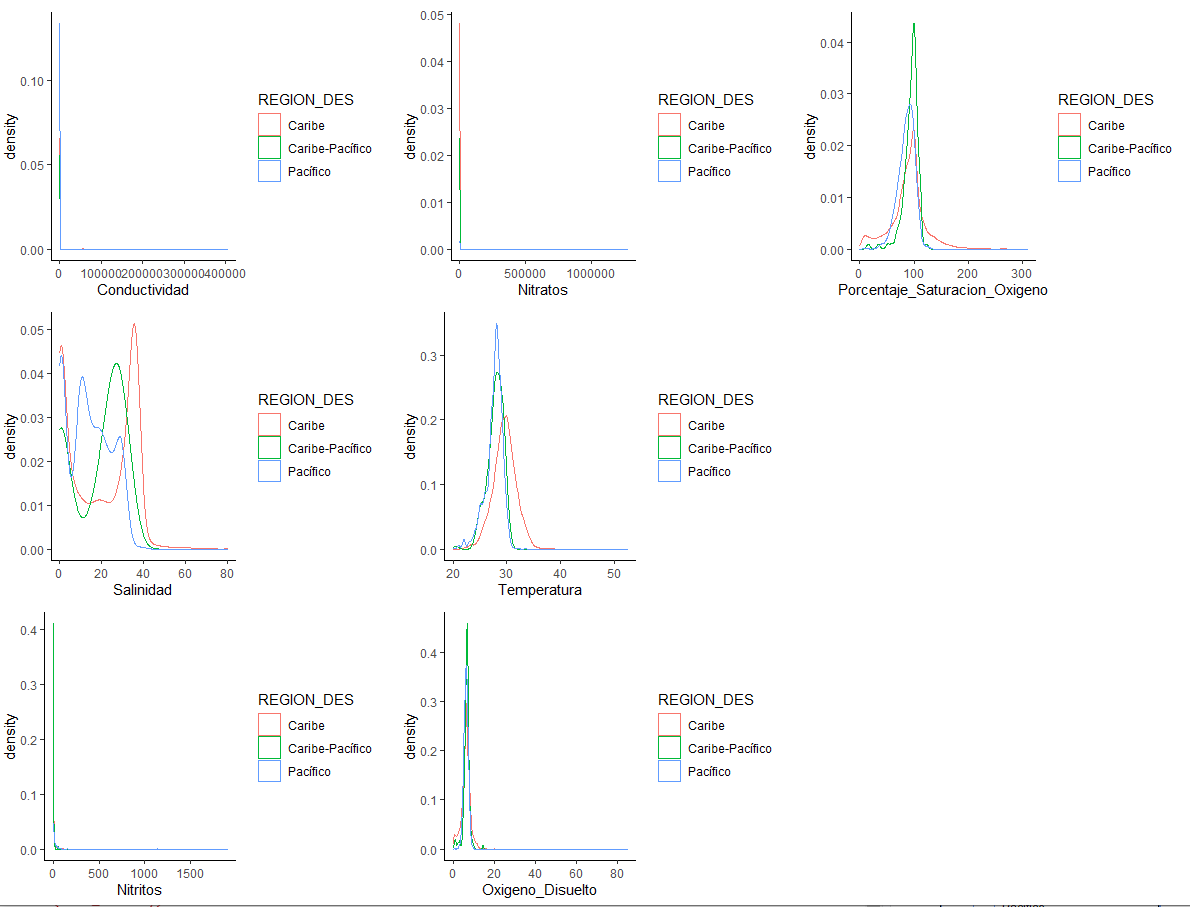
## MARIANA SILVERA

Se utilizan los datos provistos en el csv del ejercicio.

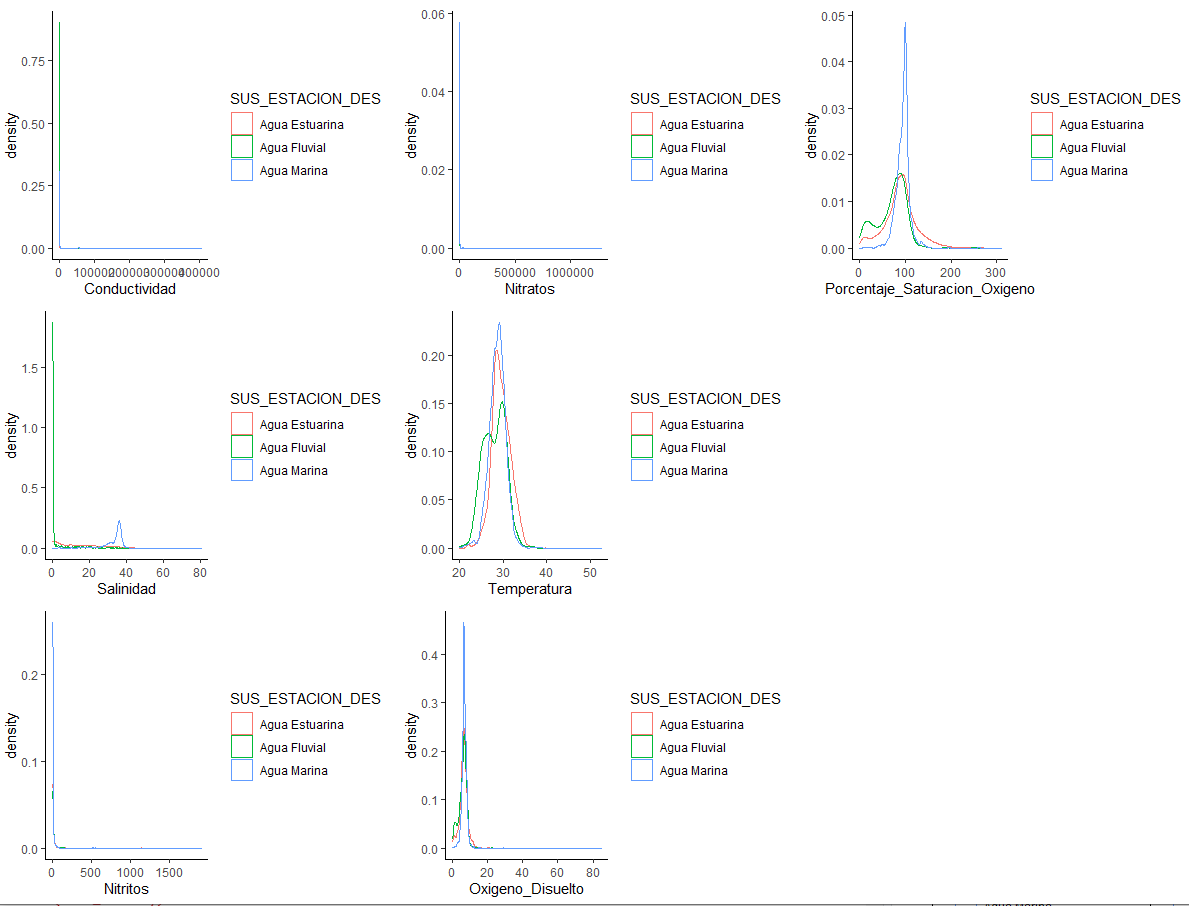
## Detección de Atípicos:

1. Selecciono los datos de Caribe y Agua Estuarina ya que son los más frecuentes en el dataset provisto.
2. Me quedo con la variable temperatura, que tmb tiene bastantes datos.
3. Elijo las estaciones que tienen más de 300 registros cada una.
4. Valido la normalidad de los datos seleccionados:  
     
   Está OK para usar los métodos paramétricos.
5. Utilizo el método IQR para la detección de atípicos:  
   En mi caso tengo 17 estaciones en mi dataset, para los cuales se calculan los cuartiles y valores mínimo y máximo:  
     
   Y se obtiene el resultado de los valores válidos y los atípicos:  
   

Imputación:

1. Lo primero que tengo que hacer es trasponer los datos para que queden en formato columna.   
   Esto lo hago con la función “pivot\_wider”
2. Analizo la distribución de los datos faltantes x región:  
   
3. Evalúo si es necesario segmentar los datos antes de imputar. Lo analizo por región:  
   

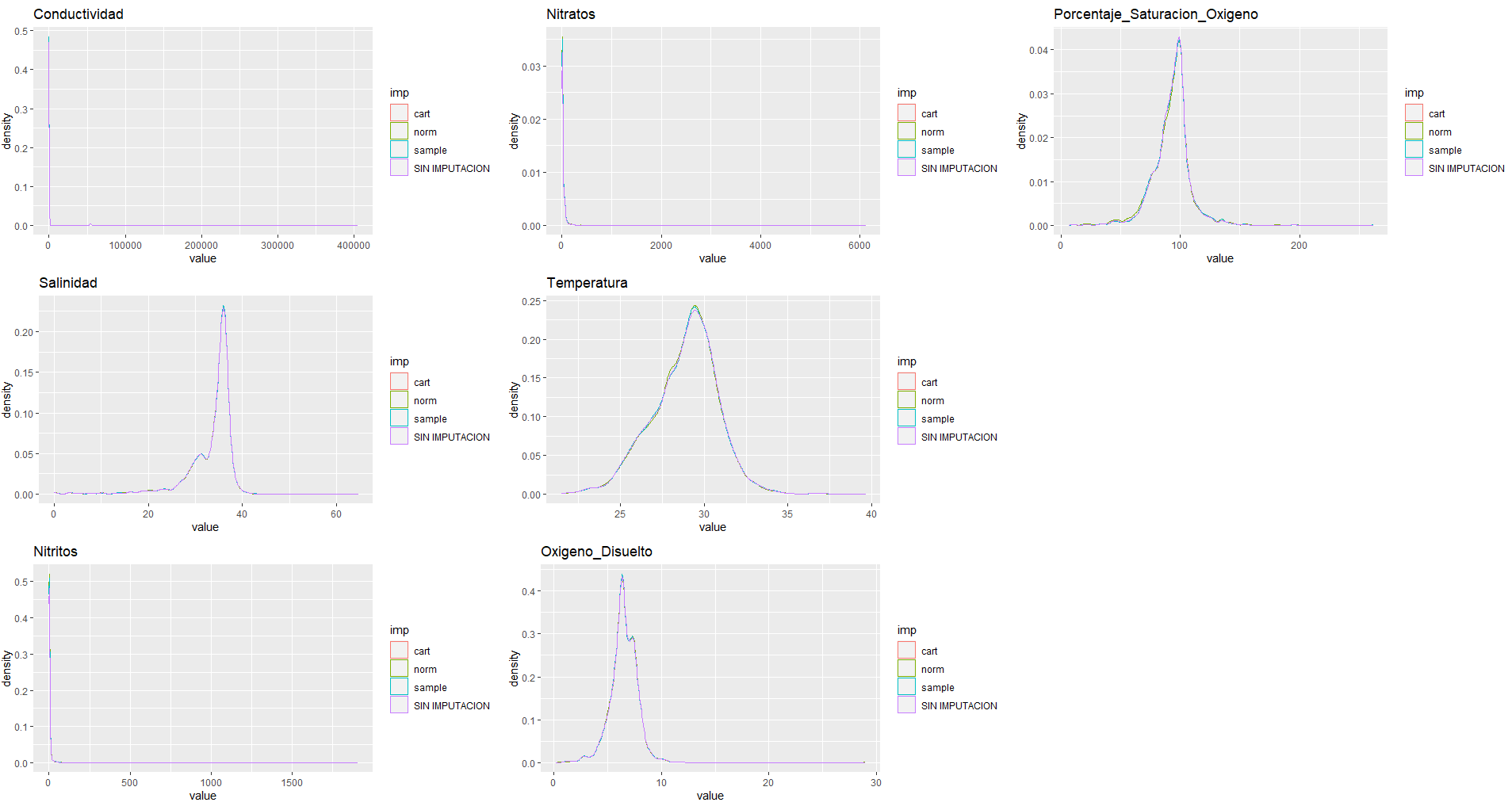
Se ve que por lo menos para la salinidad, Temperatura y la saturación de Oxígeno se recomienda segmentar por región, ya que sus distribuciones son diferentes.



Lo mismo ocurre si se analiza por sustrato, por lo cual también se recomienda segregar por este campo.

Por lo tanto la segregación la haremos por ambos campos. Para el ejercicio me quedo con el subset de datos: Caribe + Agua Marina

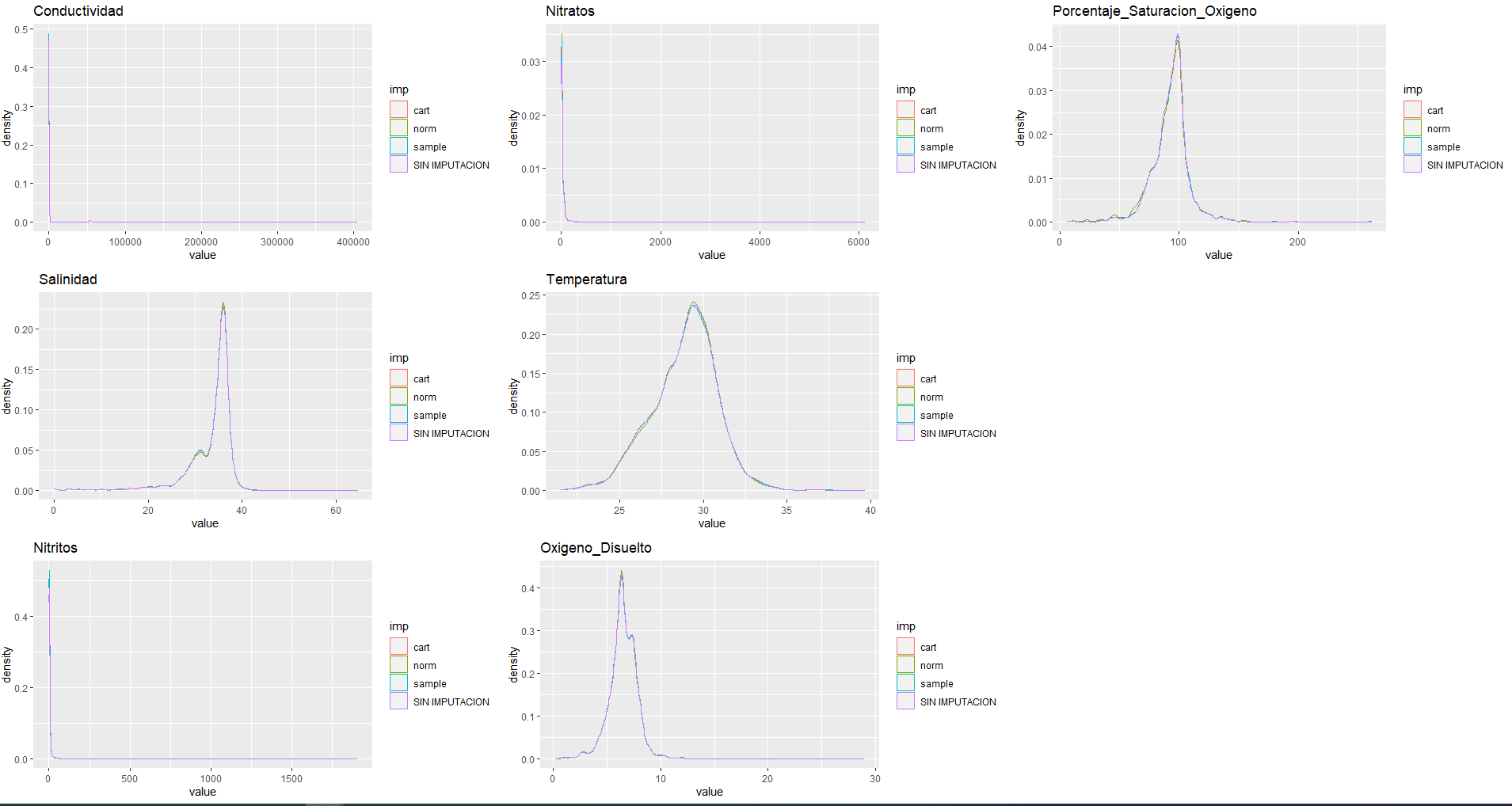
Ejecuto los 3 métodos de imputación por MICE seleccionados: sample, cart, norm.nob y comparo gráficamente los 3 resultados versus los datos sin imputación:



Los 3 métodos analizados, parecen ajustarse bien (tienen distribuciones similares).

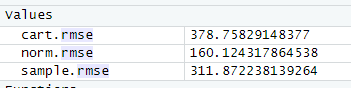
## Comparación de Imputaciones:

Se realizaron las CURVAS DE DENSIDAD PARA LA VISUALIZACION DE LAS IMPUTACIONES y se ve que las mismas se ajustan muy bien a los datos:



## Resultado del análisis de sensibilidad:

Se ejecuta el análisis de sensibilidad con la comparación por RMSE obteniendo los siguientes resultados para cada método:



Los valores obtenidos con el método RMSE parecen ser bastante altos, en comparación con el ejemplo visto en el curso.