

SOBREVIVIENTES AL HUNDIMIENTO DEL TITANIC

PROYECTO INDIVIDUAL / GRUPO 02
ALESSANDRO UMAÑA VEGA C37963

CONTENIDO

1. INTRODUCCIÓN
2. BASE DE DATOS
3. VALORES FALTANTES
4. GRÁFICOS
5. CORRELACIÓN
6. CONCLUSIÓN Y
RECOMENDACIONES

INTRODUCCIÓN

- Técnicas de manipulación, limpieza e imputación de datos.
- Análisis descriptivo y estadístidico
- Visualizaciones

BASE DE DATOS

- Titanic
- 891 entradas
- Variables de, clase, edad, tarifa, cantidad de hermanos, cantidad de padres o hijos, supervivencia, nombre, número de ticket y cabina.
- Valores faltantes en edad.

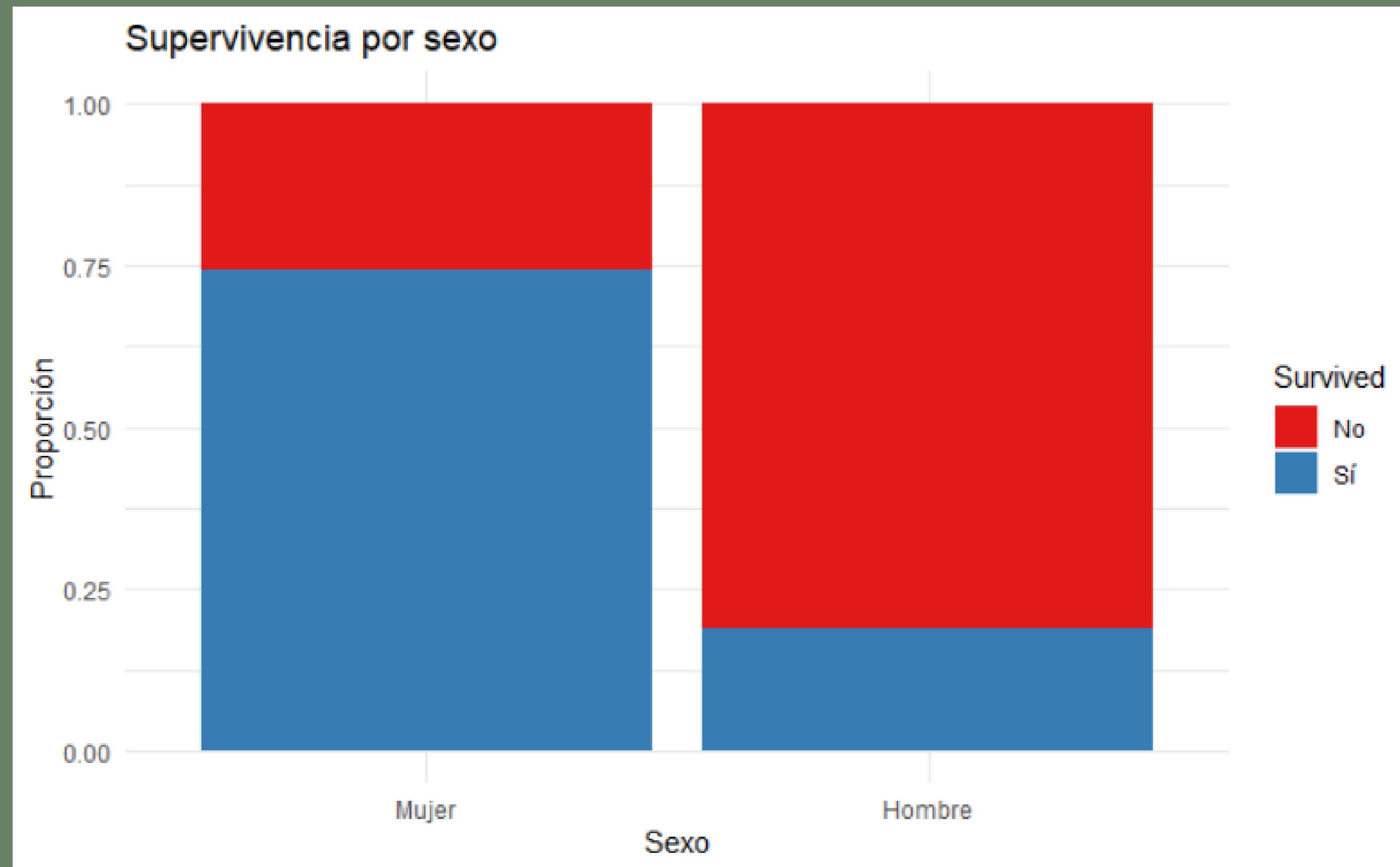
VALORES FALTANTES

- 177 NA (20% de las entradas totales).
- Imputar usando mediana por clases con un ciclo for y usando if.
- MAntener estructura socioeconómica.

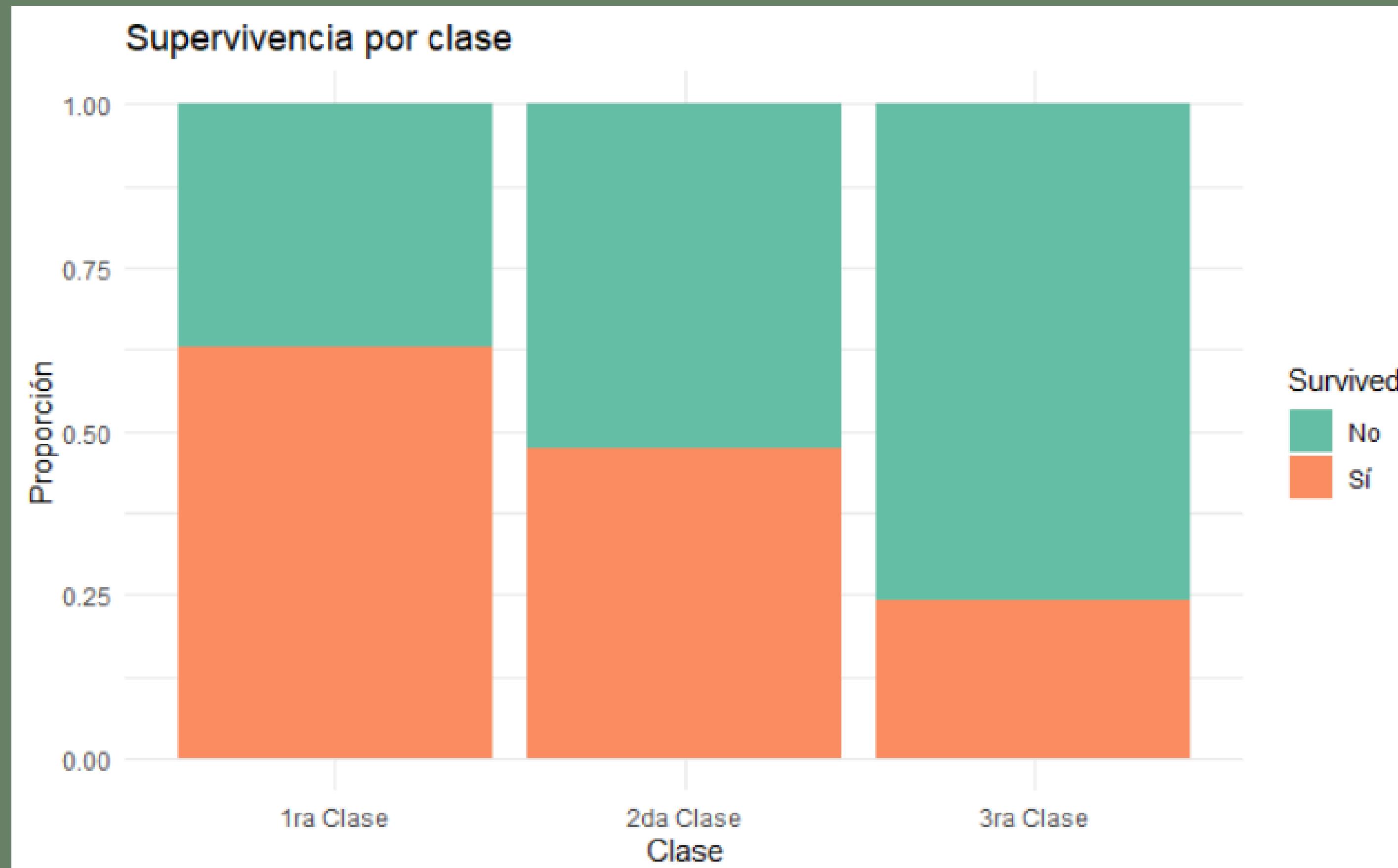
| | False | True |
|--------|-------|------|
| Mujer | 261 | 53 |
| Hombre | 453 | 124 |

GRÁFICOS

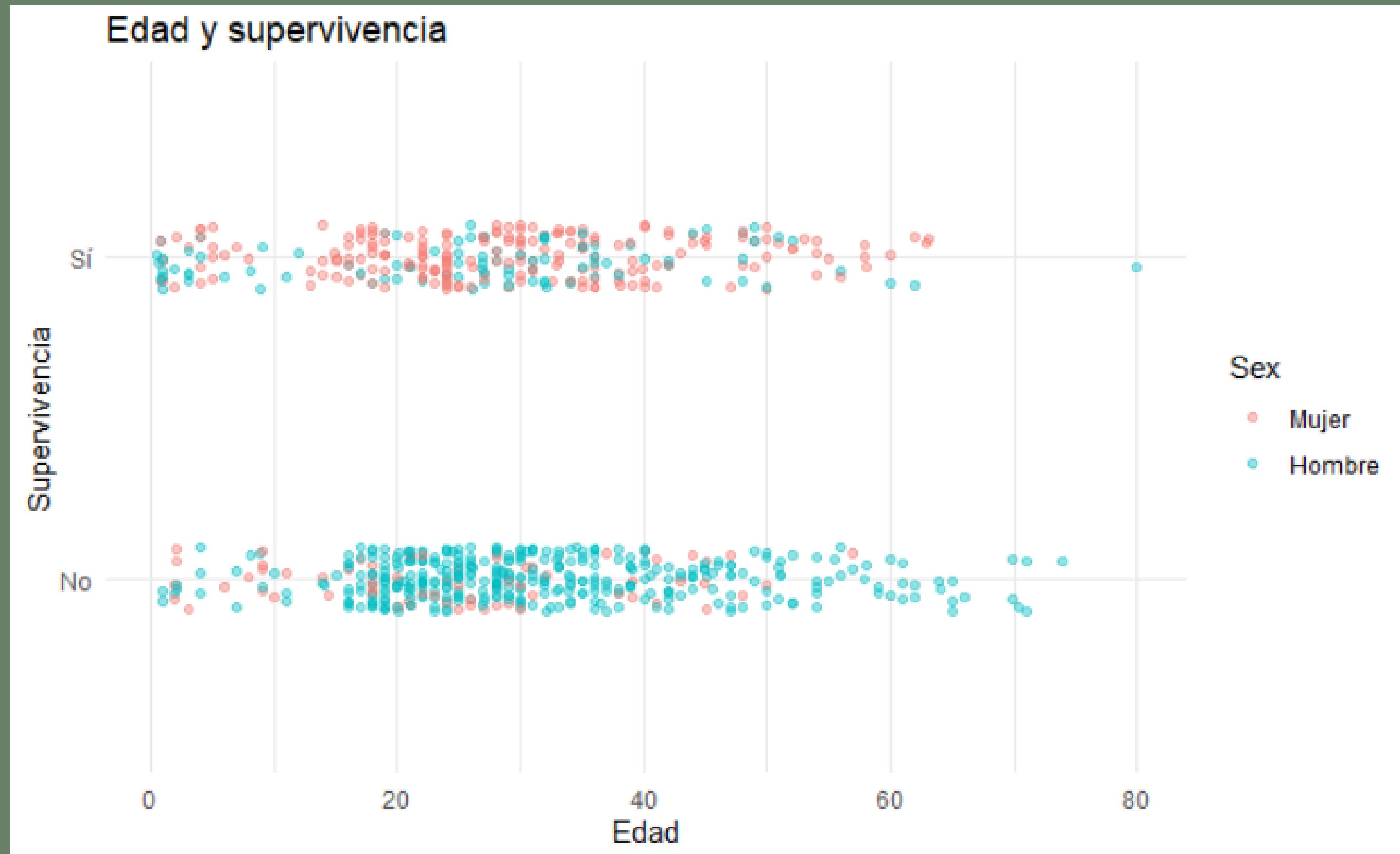
SUPERVIVENCIA POR SEXO



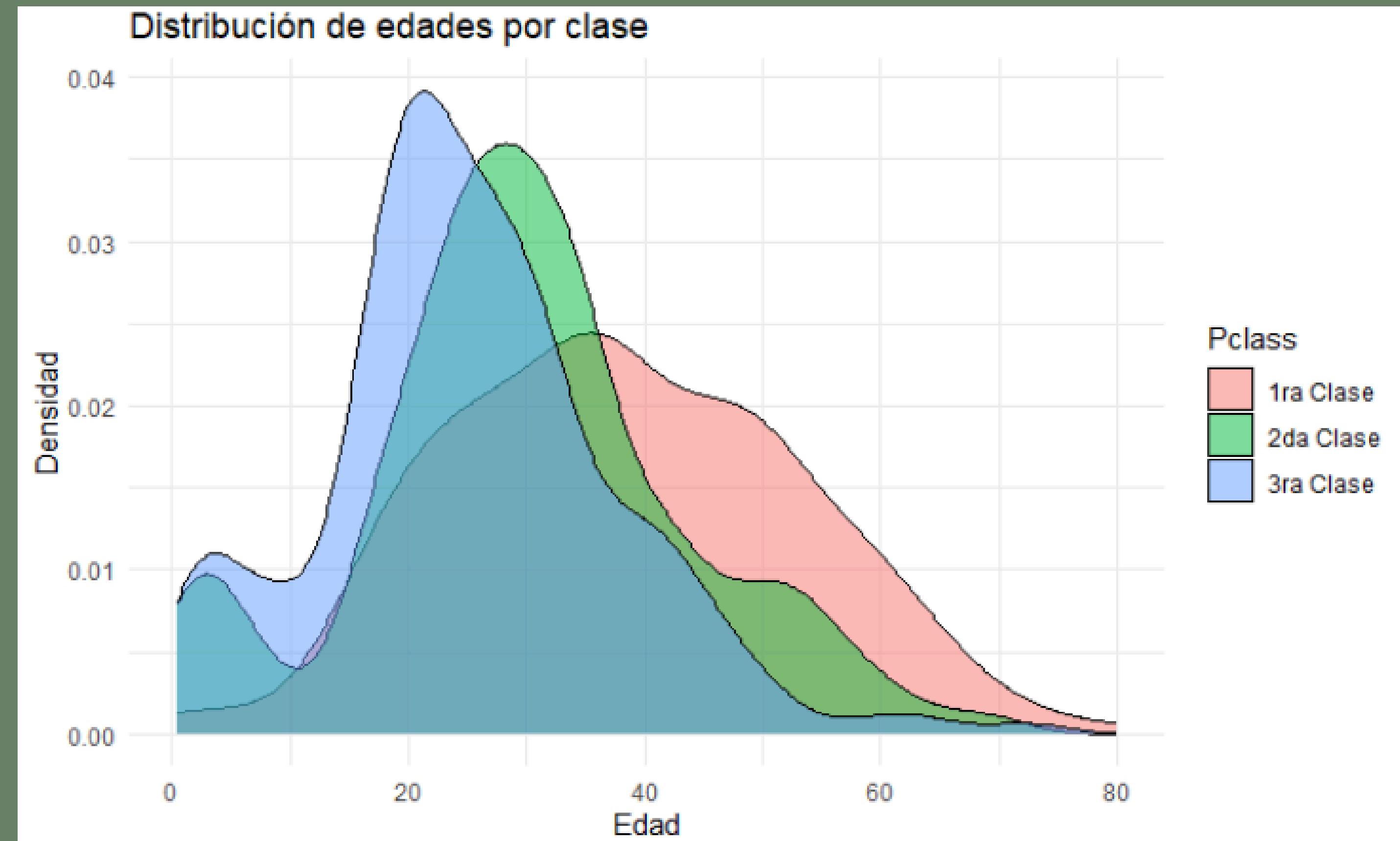
SUPERVIVENCIA POR CLASE



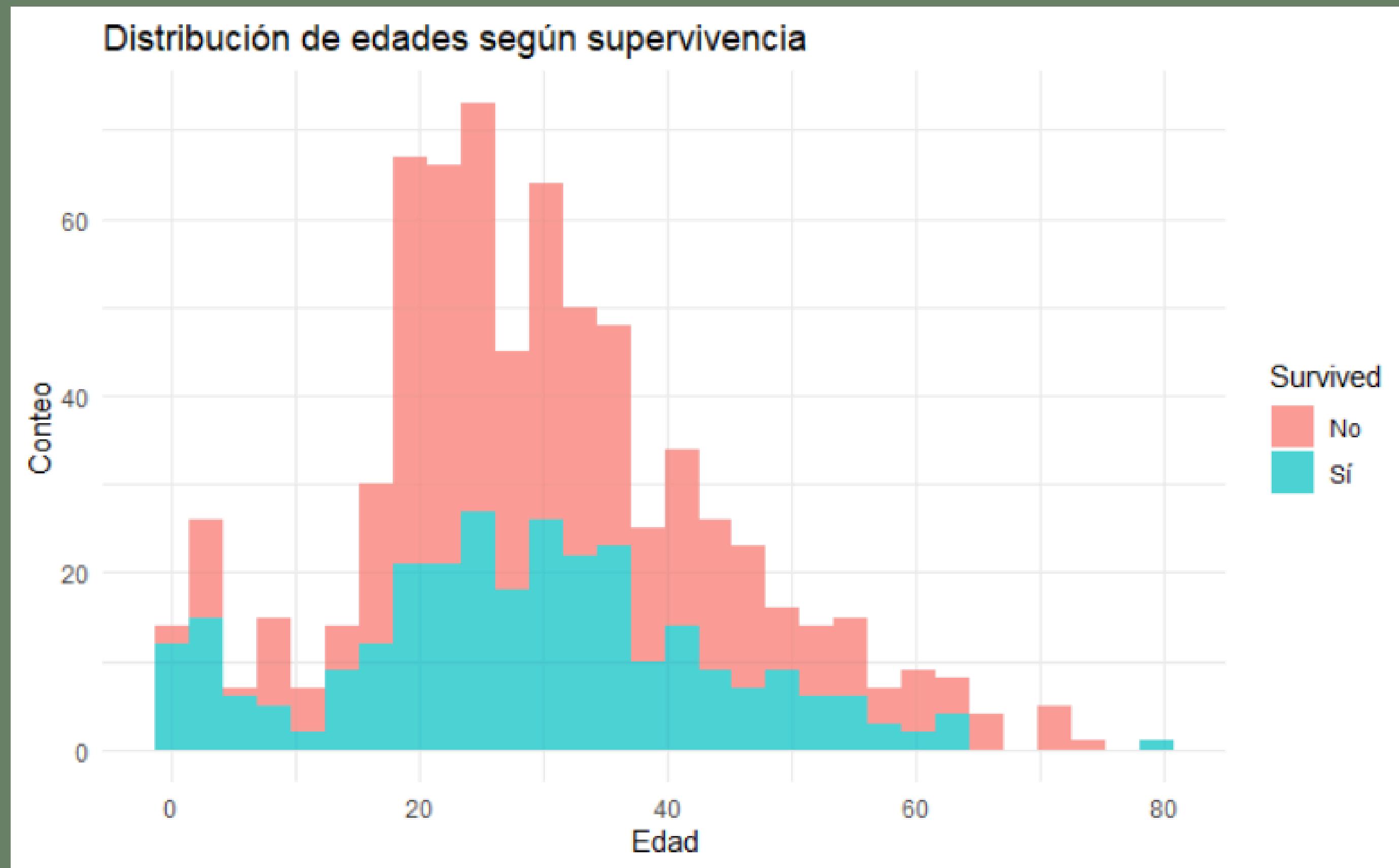
EDAD Y SUPERVIVENCIA



DISTRIBUCIÓN DE EDADES POR CLASE



DISTRIBUCIÓN DE EDADES SEGÚN SUPERVIVENCIA



CORRELACIÓN

| | Survived | Pclass | Age | SibSp | Parch | Fare |
|-----------------|-----------------|---------------|--------------|--------------|--------------|-------------|
| Survived | 1.00000000 | - 0.33848104 | - 0.04725461 | - 0.03532250 | 0.08162941 | 0.2573065 |
| Pclass | - 0.33848104 | 1.00000000 | - 0.40848699 | 0.08308136 | 0.01844267 | - 0.5494996 |
| Age | - 0.04725461 | - 0.40848699 | 1.00000000 | - 0.24352572 | - 0.17109485 | 0.1237837 |
| SibSp | - 0.03532250 | 0.08308136 | - 0.24352572 | 1.00000000 | 0.41483770 | 0.1596510 |
| Parch | 0.08162941 | 0.01844267 | - 0.17109485 | 0.41483770 | 1.00000000 | 0.2162249 |
| Fare | 0.2573065 | - 0.5494996 | 0.1237837 | 0.1596510 | 0.2162249 | 1.00000000 |

CORRELACIÓN

- Clase ↔ Tarifa (correlación fuerte)
- Clase ↔ Supervivencia
- Edad ↔ Clase
- SibSp ↔ Parch
- Importancia del estatus socioeconómico

CONCLUSIÓN Y RECOMENDACIONES

- Factores socioeconómicos
- Clase, sexo y tarifa como predictores
- Imputación adecuada
- Modelos predictivos
- Limitaciones del dataset

i G R A C I A S !