

Dataset Titanic

Activa los paquetes y datos que utilizaremos

```
library(tidyverse) #activa los paquetes para el análisis de datos
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.2      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag() masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(carData) #activa el paquete donde se encuentran los datos
```

```
data(TitanicSurvival) #activa los datos
```

```
head(TitanicSurvival) #muestra los primeros casos
```

```
##               survived    sex    age passengerClass
## Allen, Miss. Elisabeth Walton    yes female 29.0000      1st
## Allison, Master. Hudson Trevor    yes  male  0.9167      1st
## Allison, Miss. Helen Loraine      no female 2.0000      1st
## Allison, Mr. Hudson Joshua Crei    no  male 30.0000      1st
## Allison, Mrs. Hudson J C (Bessi    no female 25.0000      1st
## Anderson, Mr. Harry                yes  male 48.0000      1st
```

```
summary(TitanicSurvival) #resume la información que tenemos
```

```
##   survived    sex      age      passengerClass
## no :809  female:466  Min.   : 0.1667  1st:323
## yes:500  male  :843  1st Qu.:21.0000  2nd:277
##                                     Median :28.0000  3rd:709
##                                     Mean   :29.8811
##                                     3rd Qu.:39.0000
##                                     Max.   :80.0000
##                                     NA's   :263
```

¿Cuántas personas sobrevivieron a la tragedia del Titanic?

```
TitanicSurvival %>%
```

```
  count(survived) #elige la función que te permite contar (obtener la frecuencia) de la variable superv
```

```
##   survived    n
## 1         no 809
## 2         yes 500
```

Sobrevivieron **500** personas.

¿Cuántos sobrevivieron de cada sexo?

```
TitanicSurvival %>%
  filter(survived=="yes") %>%
  group_by(survived, sex) %>% #lo mismo que antes pero previamente debes elegir la función que te permite
  count(survived)

## # A tibble: 2 x 3
## # Groups:   survived, sex [2]
##   survived sex      n
##   <fct>    <fct> <int>
## 1 yes      female  339
## 2 yes      male    161
```

De los 500 supervivientes, *339* fueron *mujeres*, y *161* *hombres*.

¿Cuál es la edad media de los supervivientes y no supervivientes?

```
TitanicSurvival %>%
  group_by(survived) %>% #elige la función que te permite agrupar los datos por superviviente o no %>%
  summarise(mean_age = mean(age, na.rm = TRUE)) #elige la función que te permite resumir la edad calculada

## # A tibble: 2 x 2
##   survived mean_age
##   <fct>      <dbl>
## 1 no         30.5
## 2 yes        28.9
```

La edad media de los que sobrevivieron fue de **28.91 años**, mientras que la edad media de los que no sobrevivieron fue de **30.54 años**.

Crea categorías para identificar los sujetos menores de edad.

Crea una nueva columna llamada “age_cut” que tome los valores: “Minor” si la edad es menor que 18 o “Major” en caso contrario. ¿Cuál es el porcentaje de supervivientes en cada clase de edad?

```
TitanicSurvival %>%
  mutate(age_bracket = ifelse(age < 18, 'Minor', 'Major')) %>% #elige la función que te permita calcular la edad
  na.omit() %>%
  group_by(age_bracket, survived) %>% #elige la función que te permita agrupar los cálculos por los grupos
  summarise(n = n()) %>% # elige la función que te permita resumir el conteo de casos para cada grupo
  mutate(freq = 100* n / sum(n)) #calcula una nueva variable a partir del conteo, pasándolo a porcentaje

## ‘summarise()’ has grouped output by ‘age_bracket’. You can override using the
## ‘.groups’ argument.

## # A tibble: 4 x 4
## # Groups:   age_bracket [2]
##   age_bracket survived      n freq
##   <chr>      <fct>    <int> <dbl>
## 1 Major      no         546  61.2
## 2 Major      yes         346  38.8
## 3 Minor      no          73  47.4
## 4 Minor      yes          81  52.6
```

El porcentaje de supervivientes por edad fue de un **38,7%** de supervivientes mayores de edad, con respecto a un **52,5%** menores de edad.

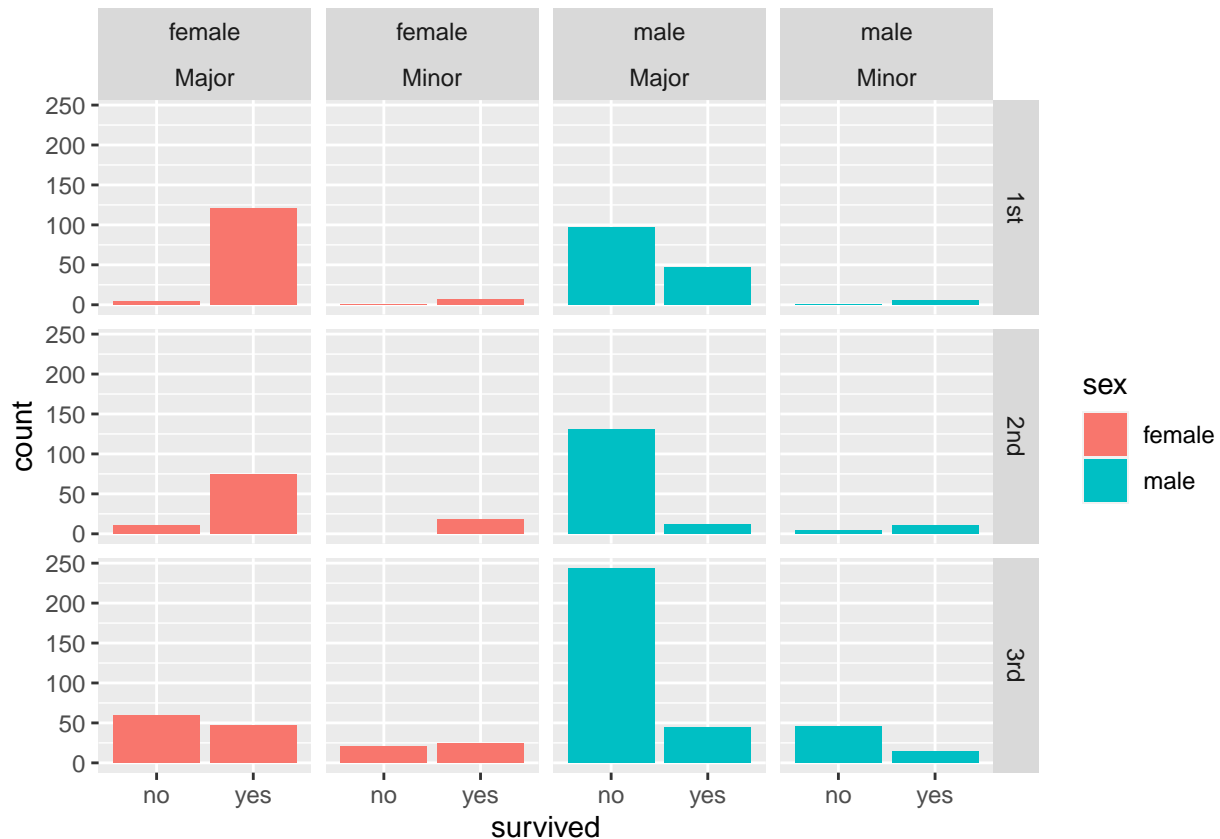
¿Se priorizó salvar a los menores de edad y a las mujeres? ¿y de clase?

Es decir, el número (o porcentaje) de menores y mujeres supervivientes varió para los pasajeros con billetes de primera, segunda o tercera clase. Utiliza gráficos para responder a esta pregunta.

```
TitanicSurvival %>%
```

```
  mutate(age_bracket = ifelse(age < 18, 'Minor', 'Major')) %>% #con la misma elección que antes  
  na.omit() %>% #elimina los NA
```

```
  ggplot(aes(x = survived, group = passengerClass, fill=sex)) + #grafica supervivientes vs clase de pa.  
  geom_bar() + #selecciona el tipo geom_* que te permite obtener barras  
  facet_grid(passengerClass ~ sex + age_bracket) #divide el gráfico en parcelas según clase de pasajero
```



Se priorizó salvar a mujeres y niños, y la diferencia es más notoria en aquellos de primera clase. En primera clase las mujeres supervivientes mayores de edad fueron muchas más que las no supervivientes, y lo mismo en segunda clase, aunque la diferencia no es tan notoria. Sin embargo en tercera clase fueron más las mujeres que no sobrevivieron. En hombres la cantidad de no supervivientes es mayor a medida que se pasa de primera a tercera clase, y en las tres clases hay un mayor número de no supervivientes que de aquellos que sí sobrevivieron.