

# Practica 3

Silvia Pineda

## Instrucciones (leer antes de empezar)

Modifica dentro del documento Practica3.qmd tus datos personales (nombre y DNI) ubicados en la cabecera del archivo.

Asegúrate, ANTES de seguir editando el documento, que el archivo .qmd se renderiza correctamente y se genera el .html correspondiente en tu carpeta local de tu ordenador.

Los chunks (cajas de código) creados están o vacíos o incompletos, de ahí que la mayoría tengan la opción `#| eval: false`. Una vez que edites lo que consideres, debes ir cambiando cada chunk a `#| eval: true` (o quitarlo directamente) para que se ejecuten.

## ENUNCIADO DE LA PRÁCTICA

Para esta práctica vais a usar la base de datos **Titanic.Rdata**

En el dataset de Titanic las variables disponibles son:

- PassengerId: identificador único del pasajero.
- Survived: si el pasajero sobrevivió al naufragio, codificada como 0 (no) y 1 (si). Esta es la variable respuesta que interesa predecir.
- Pclass: clase a la que pertenecía el pasajero: 1, 2 o 3.
- Name: nombre del pasajero.
- Sex: sexo del pasajero.
- Age: edad del pasajero.
- SibSp: número de hermanos, hermanas, hermanastros o hermanastras en el barco.
- Parch: número de padres e hijos en el barco.
- Ticket: identificador del billete.

- Fare: precio pagado por el billete.
- Cabin: identificador del camarote asignado al pasajero.
- Embarked: puerto en el que embarcó el pasajero.

## CARGA DE DATOS y LIBRERÍAS

```
load("Titanic.Rdata")
```

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
v dplyr      1.1.4      v readr      2.1.5
```

```
v forcats    1.0.0      v stringr    1.5.1
```

```
v ggplot2    3.5.0      v tibble     3.2.1
```

```
v lubridate  1.9.3      v tidyr      1.3.1
```

```
v purrr      1.0.2
```

```
-- 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
```

```
library(ggplot2)
```

```
library(naniar)
```

```
library(mice)
```

Attaching package: 'mice'

The following object is masked from 'package:stats':

filter

The following objects are masked from 'package:base':

cbind, rbind

```
library(VIM)
```

```
Loading required package: colorspace
Loading required package: grid
VIM is ready to use.
```

Suggestions and bug-reports can be submitted at: <https://github.com/statistikat/VIM/issues>

```
Attaching package: 'VIM'
```

```
The following object is masked from 'package:datasets':
```

```
sleep
```

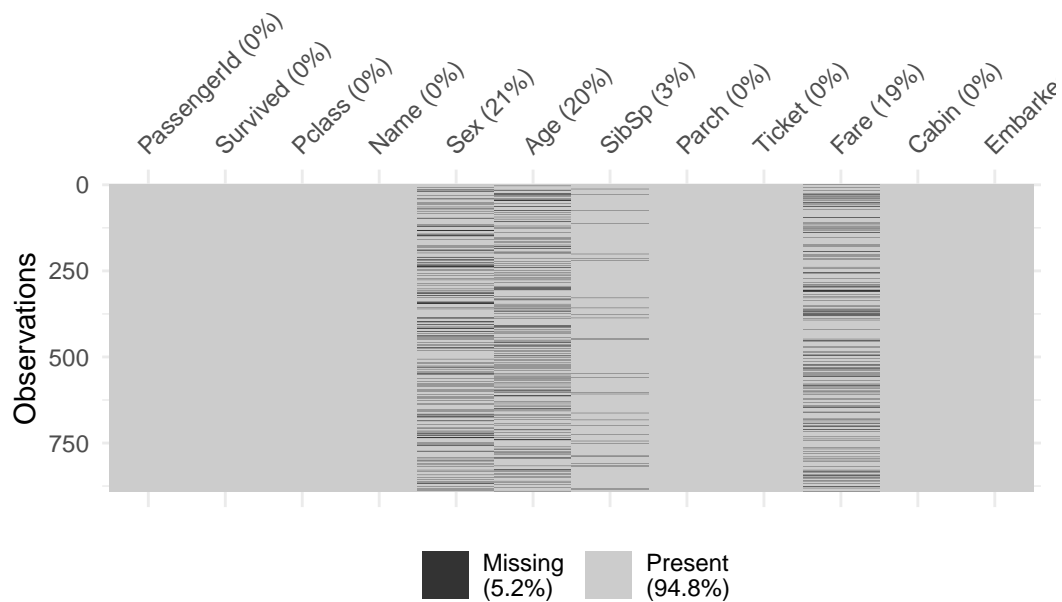
### **1. Identifica las variables cuantitativas y cualitativas.**

**V. Cualitativas:** Survived, Pclass, Sex, Embarked, PassengerID, Name, Ticket y Cabin

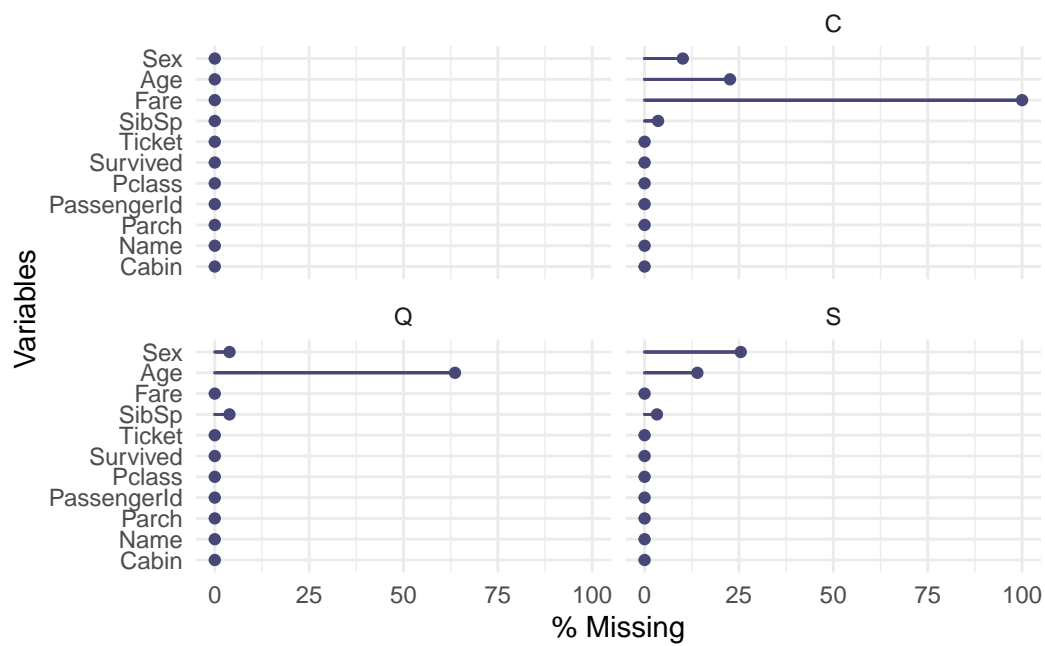
**V. Cuantitativas:** Age, SibSp, Parch, Fare

### **2. Visualiza y Cuantifica los datos missing. Después decide a qué tipo de datos missing (MCAR, MAR o MNAR) pertenecen los missing de cada variable.**

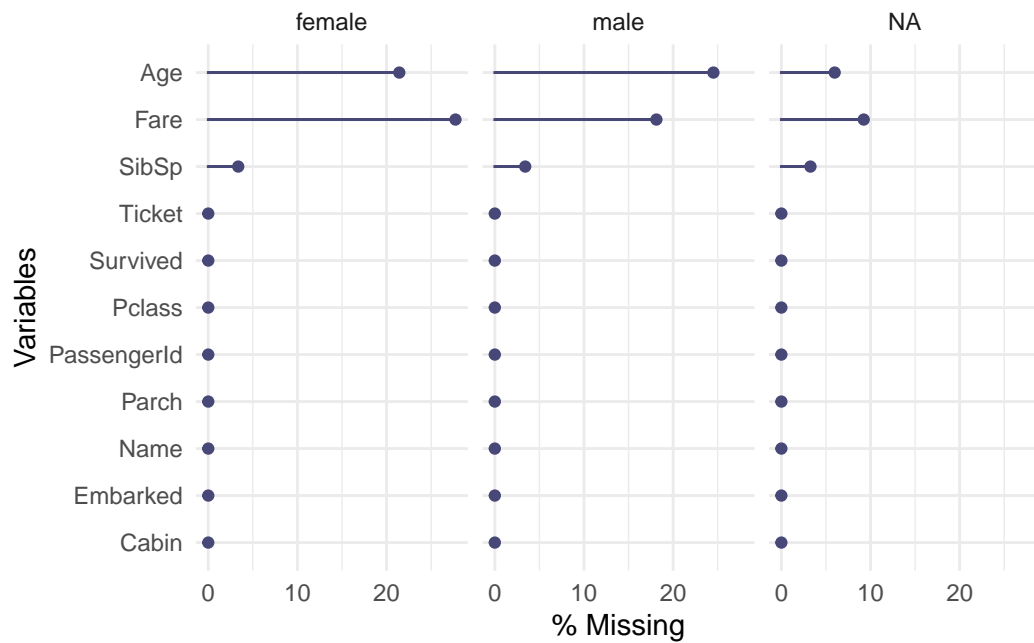
```
vis_miss(data)
```



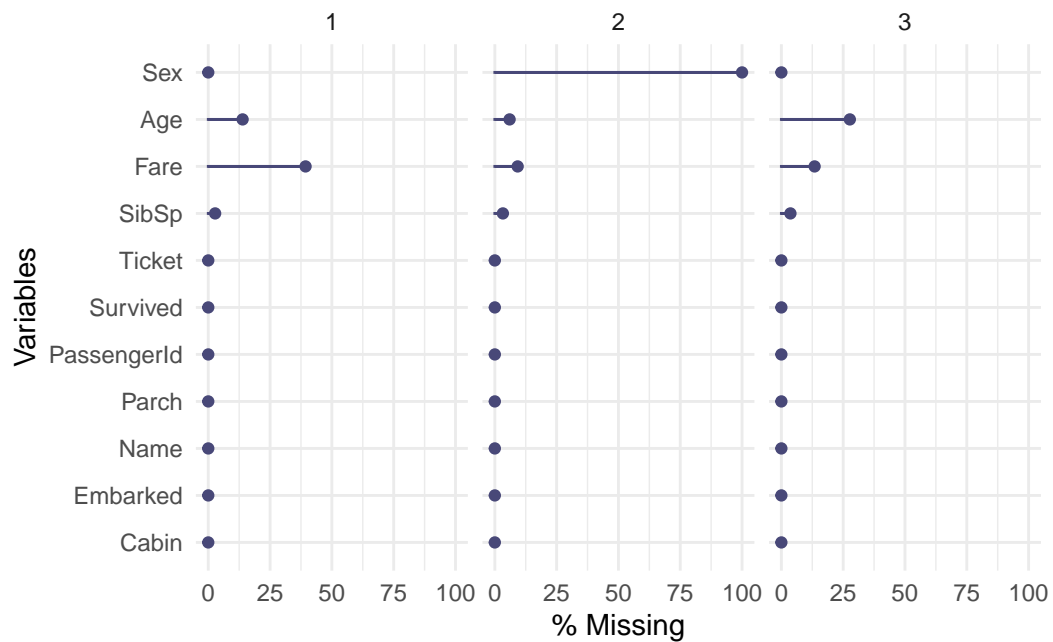
```
##Mirar la asociación de los datos missing con las V. Cualitativas
gg_miss_var(data, show_pct = TRUE, facet=Embarked)
```



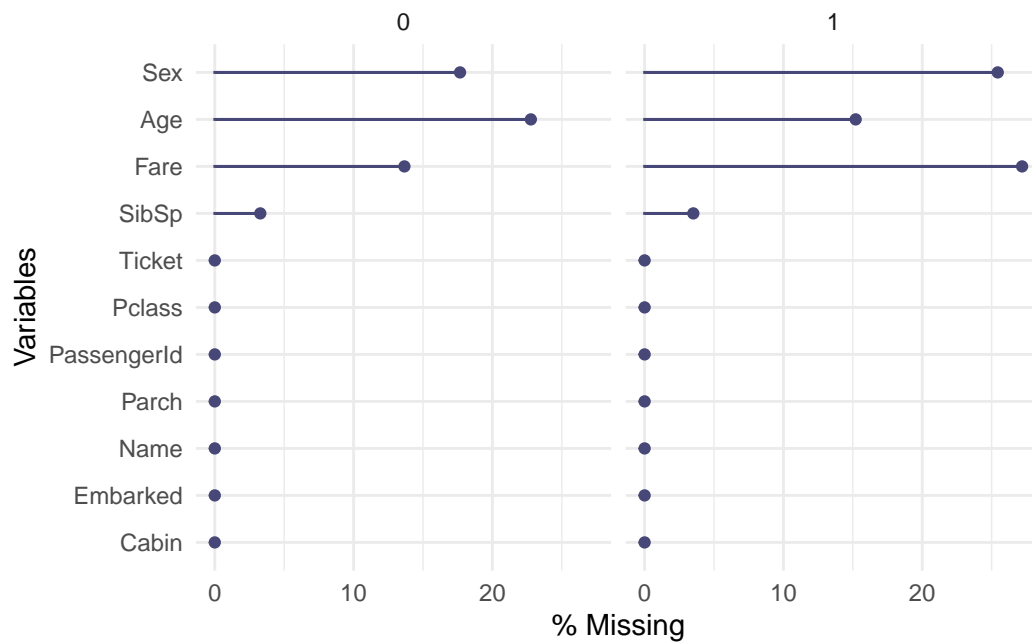
```
gg_miss_var(data, show_pct = TRUE, facet=Sex)
```



```
gg_miss_var(data, show_pct = TRUE, facet=Pclass)
```

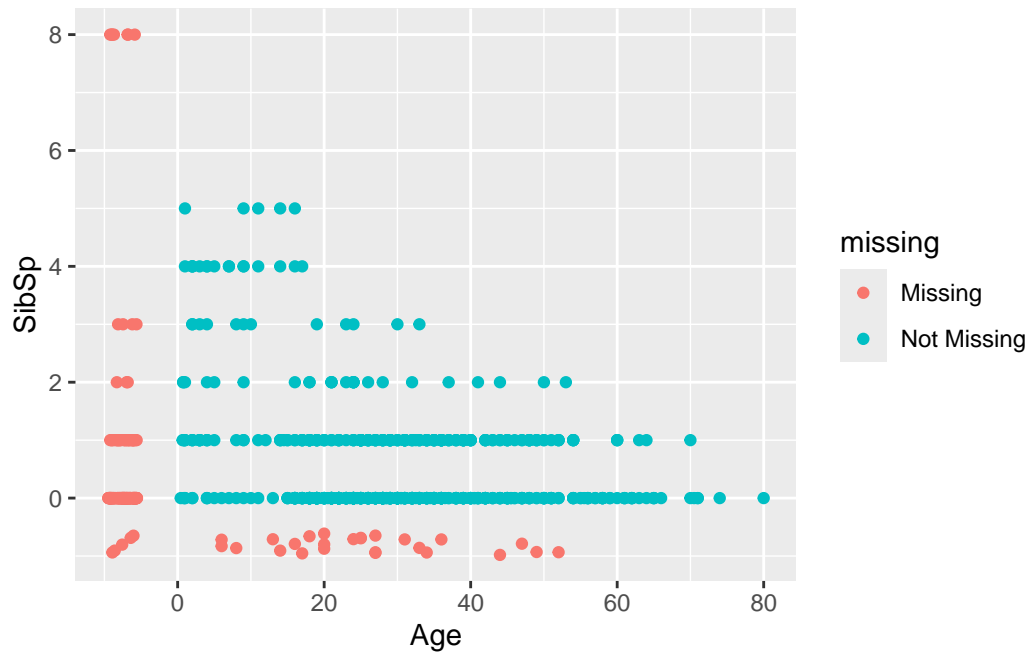


```
gg_miss_var(data, show_pct = TRUE, facet=Survived)
```

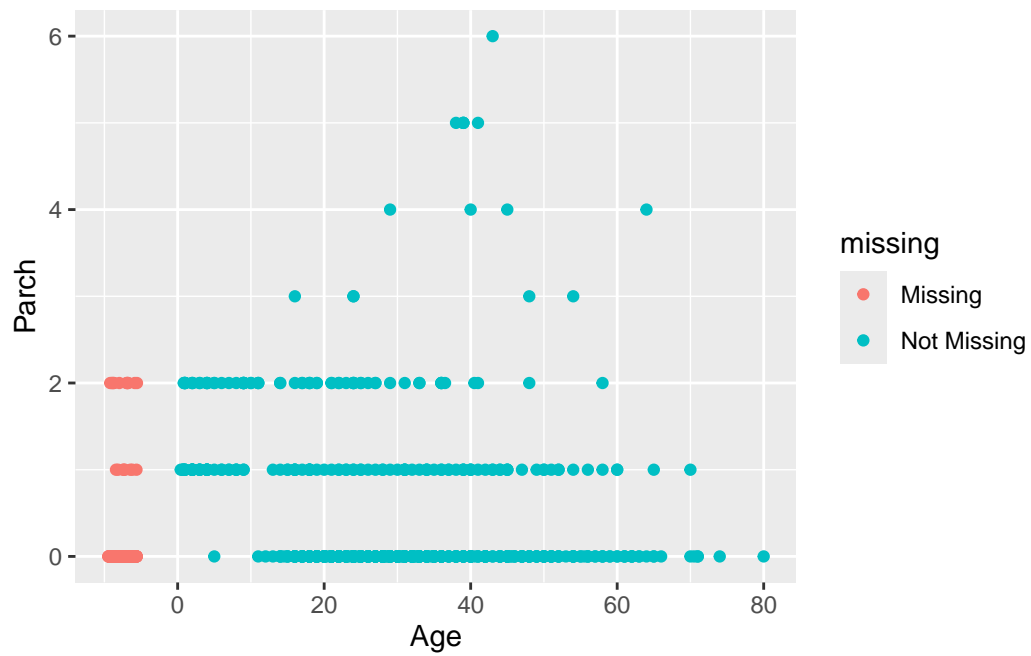


```
##Mirar la asociación de los datos missing con las V.Cuantitativas
#V.Cuantis con missing: Age, Fare, Sibsp
#V.Cauntis sin missing: Parch

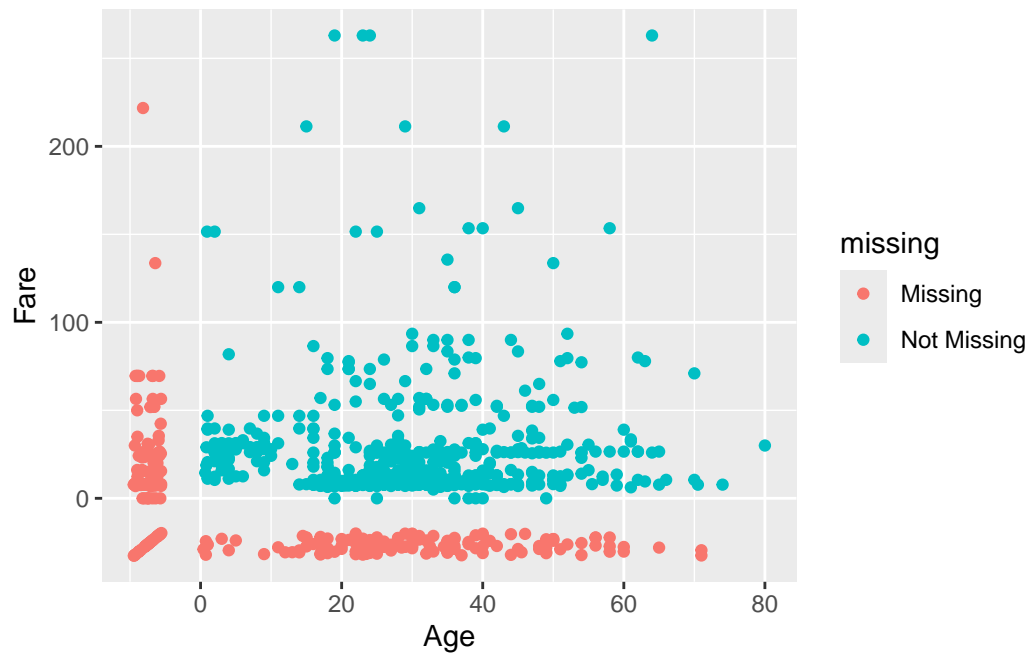
ggplot(data = data, aes (x = Age, y = SibSp)) + geom_miss_point()
```



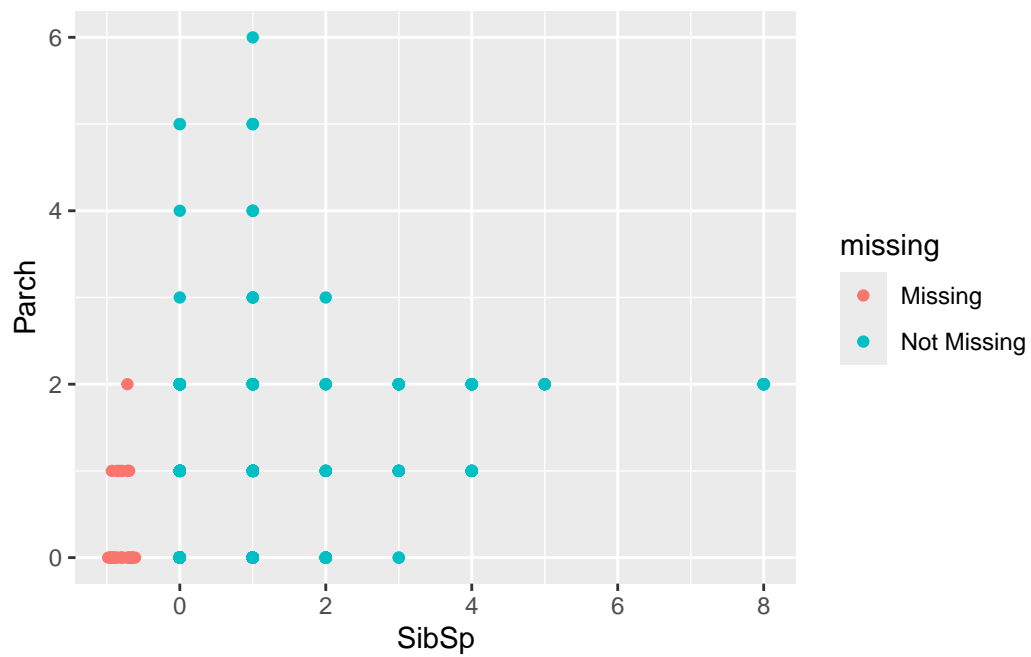
```
ggplot(data = data, aes (x = Age, y = Parch)) + geom_miss_point()
```



```
ggplot(data = data, aes (x = Age, y = Fare)) + geom_miss_point()
```

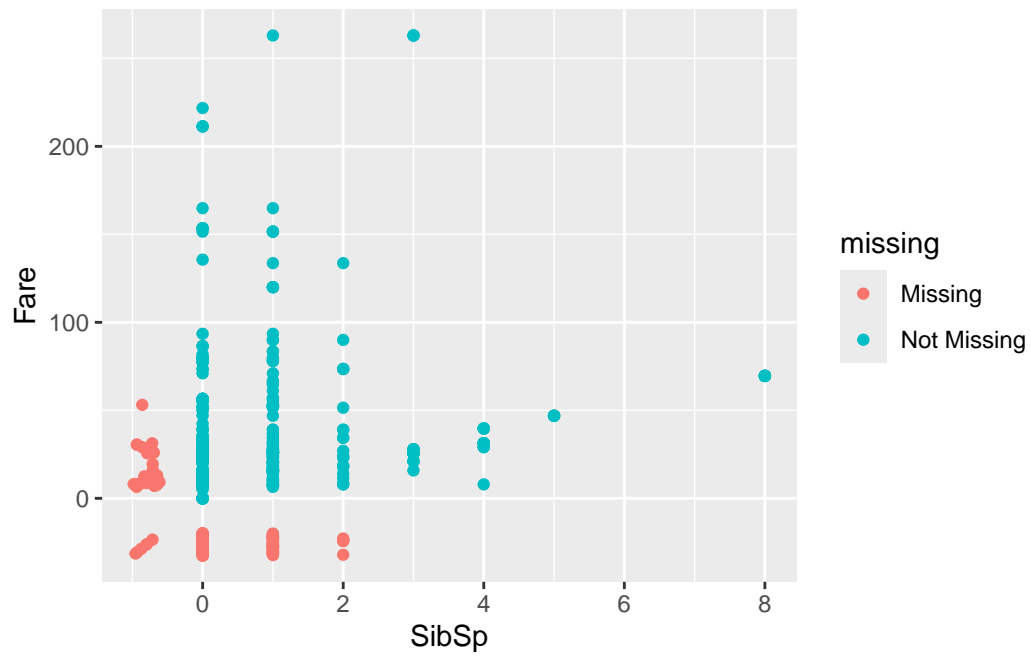


```
ggplot(data = data, aes (x = SibSp, y = Parch)) + geom_miss_point()
```

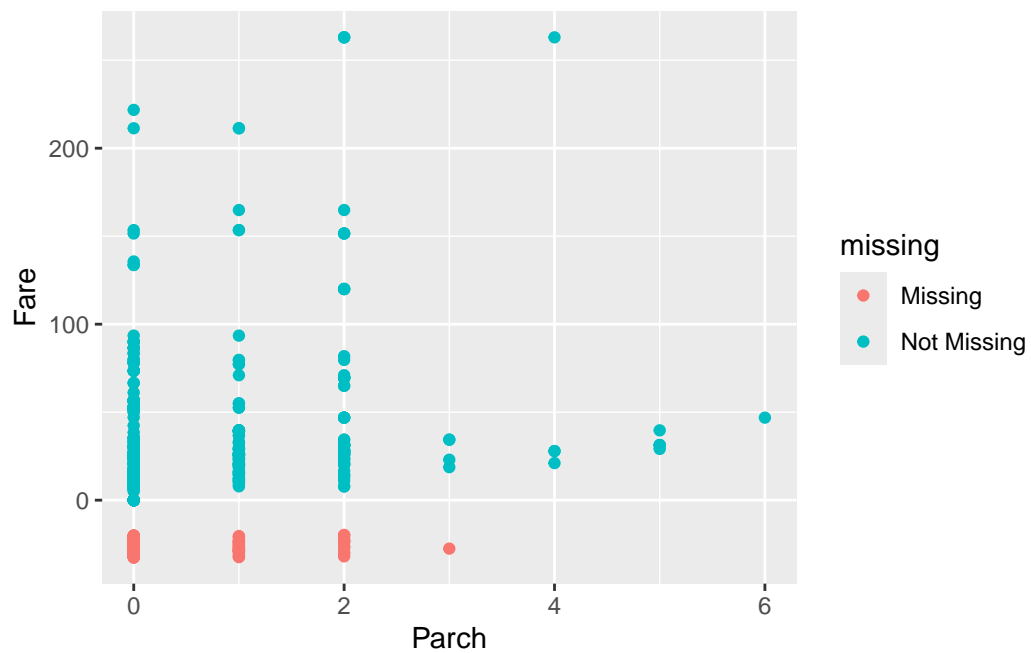




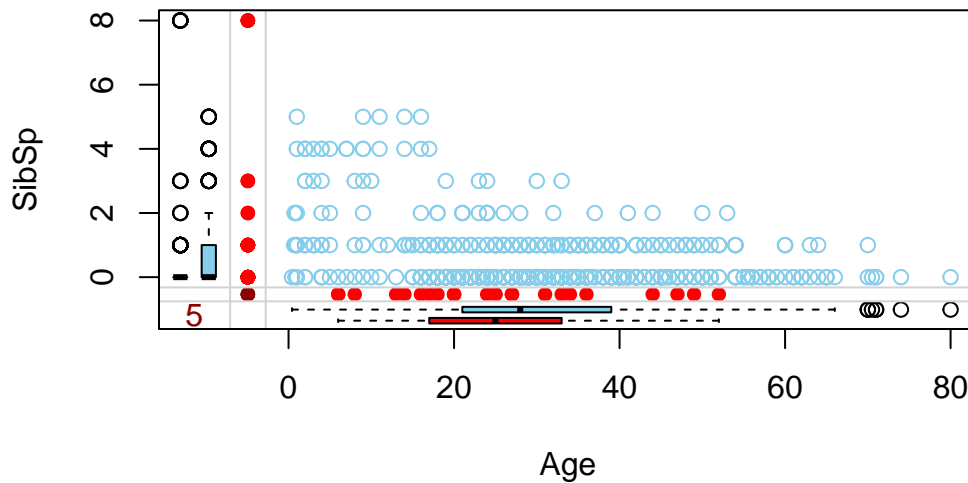
```
ggplot(data = data, aes (x = SibSp, y = Fare)) + geom_miss_point()
```



```
ggplot(data = data, aes (x = Parch, y = Fare)) + geom_miss_point()
```



```
library(VIM)
marginplot(data[c("Age", "SibSp")])
```



Variables con datos missing:

Sex (21%), Age (20%), SibSp (3%), Fare (19%)

Todos los missing de Fare corresponden a Embarked = C (MAR)

Todos los missing de Sex corresponden a Pclass = 2 (MAR)

Age y SibSp no parecen tener asociación con nada por tanto son (MCAR)

### 3. Realiza una imputación simple de la siguiente forma:

1. Imputa Sex, Age y Sibsp por la media/moda según consideres. ¿Crees que es correcta la imputación?
2. Para las variables cuantitativas Fare y Age imputa por un modelo de regresión. Ten en cuenta que la variable *Age* está asociada con *Pclass* que tiene todos los datos completos y la variable Fare está asociada con *Survived*. ¿Crees que son correctas las imputaciones?
3. Recuerda comprobar las imputaciones gráficamente y comenta si te parece que se ha imputado bien o mal o regular cada variable.

```
##Sex
prop.table(table(data$Sex))
```

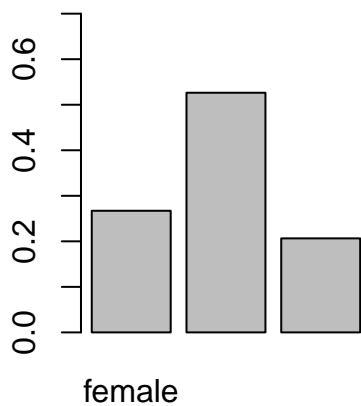
```
      female      male
0.3366337 0.6633663
```

```
data$Sex_moda<-data$Sex
data$Sex_moda[is.na(data$Sex_moda)]<-"male"
prop.table(table(data$Sex_moda))
```

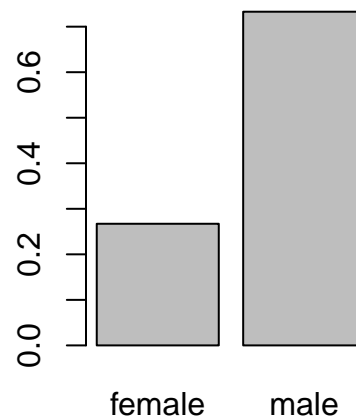
```
      female      male
0.2671156 0.7328844
```

```
par(mfrow = c(1, 2)) # Organizar las gráficas en una fila de 2 columnas
barplot(prop.table(table(data$Sex, useNA = "always")), main = "Antes de la imputación",ylim=
barplot(prop.table(table(data$Sex_moda)), main = "Después de la imputación",ylim=c(0.0,0.7))
```

**Antes de la imputación**



**Después de la imputación**



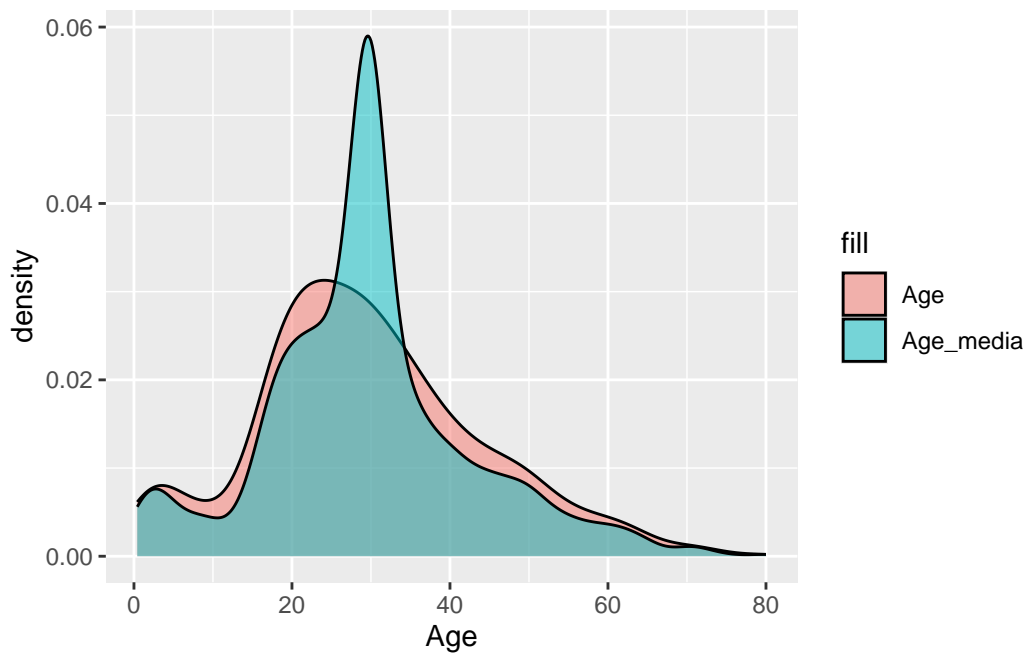
```
#Age
data$Age_media<-data$Age
mean(data$Age, na.rm = TRUE)
```

```
[1] 29.69912
```

```
data$Age_media[is.na(data$Age_media)] <- mean(data$Age_media, na.rm = TRUE)

##Gráfico para representar las diferencias
ggplot(data, aes(x = Age, fill = "Age")) +
  geom_density(alpha = 0.5) +
  geom_density(aes(x = Age_media, fill = "Age_media"), alpha = 0.5)
```

Warning: Removed 177 rows containing non-finite outside the scale range (`stat\_density()`).



```
#SibSp
prop.table(table(data$SibSp))
```

```

      0          1          2          3          4          5
0.678281069 0.238095238 0.032520325 0.018583043 0.018583043 0.005807201
      8
0.008130081

```

```

data$SibSp_moda<-data$SibSp
data$SibSp_moda[is.na(data$SibSp_moda)]<-"0"
prop.table(table(data$SibSp_moda))

```

```

      0          1          2          3          4          5
0.689113356 0.230078563 0.031425365 0.017957351 0.017957351 0.005611672
      8
0.007856341

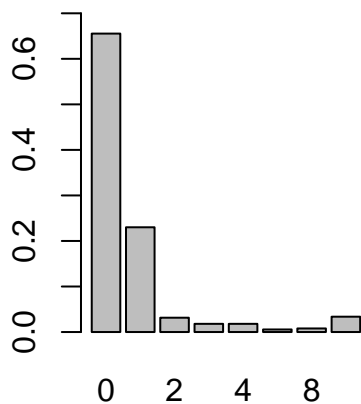
```

```

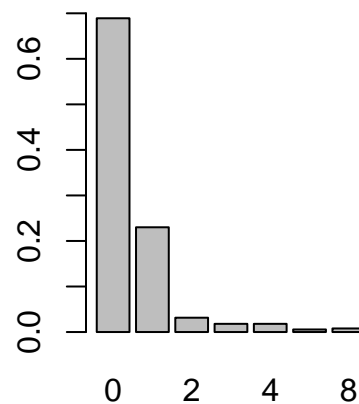
par(mfrow = c(1, 2)) # Organizar las gráficas en una fila de 2 columnas
barplot(prop.table(table(data$SibSp, useNA = "always")), main = "Antes de la imputación",ylim=c(0.0,0.7))
barplot(prop.table(table(data$SibSp_moda)), main = "Después de la imputación",ylim=c(0.0,0.7))

```

**Antes de la imputación**



**Después de la imputación**



Sex y age son dos variables que al imputar por la moda/media se desconfigura por completo la imputación ya que el % de datos missing es muy elevado.

SibSp es una variable cuantitativa discreta que tiene mucho sentido imputarla por la moda por la media nos genera valores no válidos. Y en este caso el % es pequeño por tanto podemos asumir el error.

```
###AGE
#Ajustar un modelo de regresión lineal de Age ~ Pclass
model1 <- lm(Age ~ Pclass, data = data)
summary(model1)
```

Call:

```
lm(formula = Age ~ Pclass, data = data)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-37.313	-7.878	-0.878	7.859	48.859

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	38.2334	0.9897	38.633	< 2e-16 ***
Pclass2	-8.3558	1.4257	-5.861	7.04e-09 ***
Pclass3	-13.0928	1.2217	-10.717	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.5 on 711 degrees of freedom

(177 observations deleted due to missingness)

Multiple R-squared: 0.1391, Adjusted R-squared: 0.1367

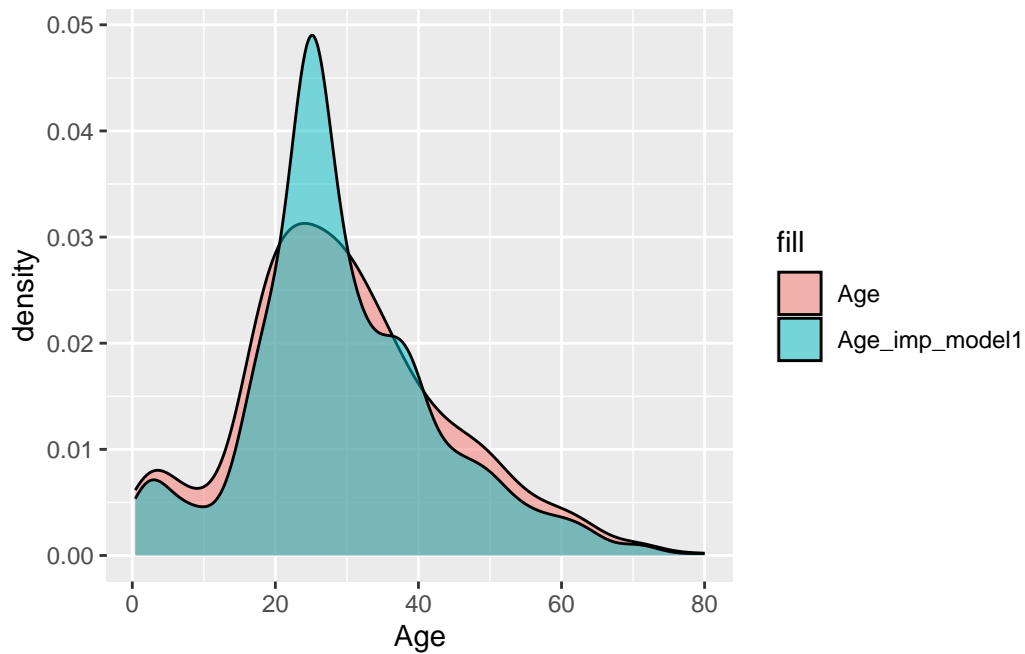
F-statistic: 57.44 on 2 and 711 DF, p-value: < 2.2e-16

```
#Predecir los valores solo para las observaciones faltantes
predictions <- predict(model1,newdata = data [is.na(data$Age),])

##Crear una nueva variable de linelist con la temperatura imputada
data$Age_imp_model1 <- data$Age
data$Age_imp_model1[is.na(data$Age)]<- predictions

#Hacer un gráfico para comparar las observaciones
ggplot(data, aes(x = Age, fill = "Age")) +
  geom_density(alpha = 0.5) +
  geom_density(aes(x = Age_imp_model1, fill = "Age_imp_model1"), alpha = 0.5)
```

Warning: Removed 177 rows containing non-finite outside the scale range  
(`stat\_density()`).

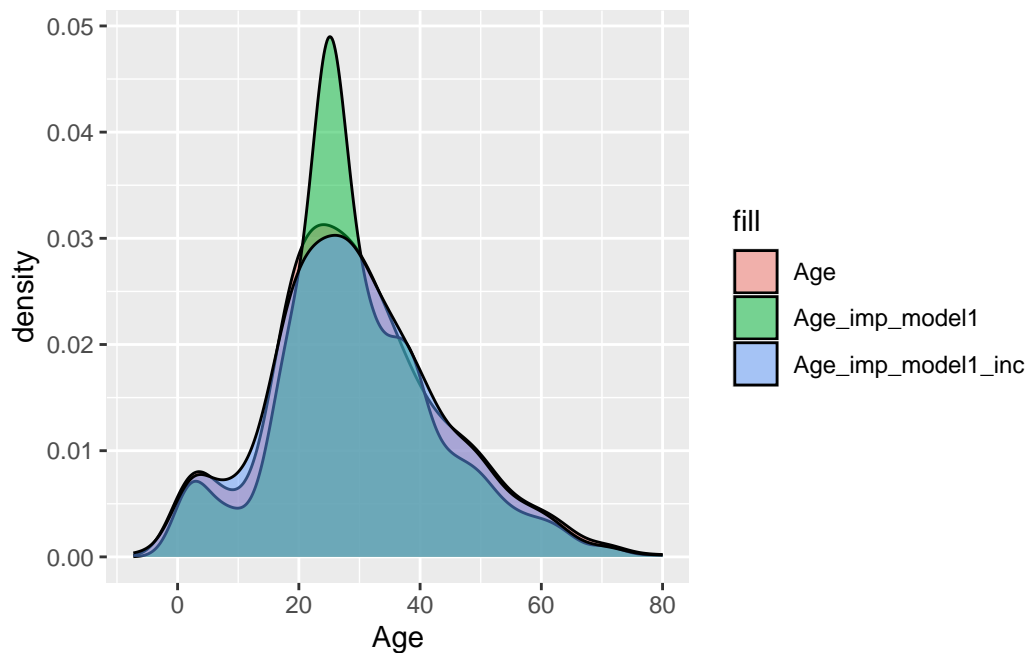


```
set.seed(3)
x<-rnorm(sum(is.na(data$Age)), 0, sd = 13.5)

data$Age_imp_model1_inc <- data$Age
data$Age_imp_model1_inc[is.na(data$Age)]<- predictions + x

#Hacer un gráfico para comparar las observaciones con la media y la regresión
#Hacer un gráfico para comparar las observaciones
ggplot(data, aes(x = Age, fill = "Age")) +
  geom_density(alpha = 0.5) +
  geom_density(aes(x = Age_imp_model1, fill = "Age_imp_model1"), alpha = 0.5) +
  geom_density(aes(x = Age_imp_model1_inc, fill = "Age_imp_model1_inc"), alpha = 0.5)
```

Warning: Removed 177 rows containing non-finite outside the scale range  
(`stat\_density()`).



```
###FARE
#Ajustar un modelo de regresión lineal de Fare ~ Survived
model1 <- lm(Fare ~ Survived, data = data)
summary(model1)
```

Call:

```
lm(formula = Fare ~ Survived, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-36.696	-12.530	-10.696	5.457	242.991

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	20.009	1.546	12.941	< 2e-16 ***
Survived1	16.686	2.635	6.333	4.22e-10 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 33.66 on 721 degrees of freedom  
(168 observations deleted due to missingness)



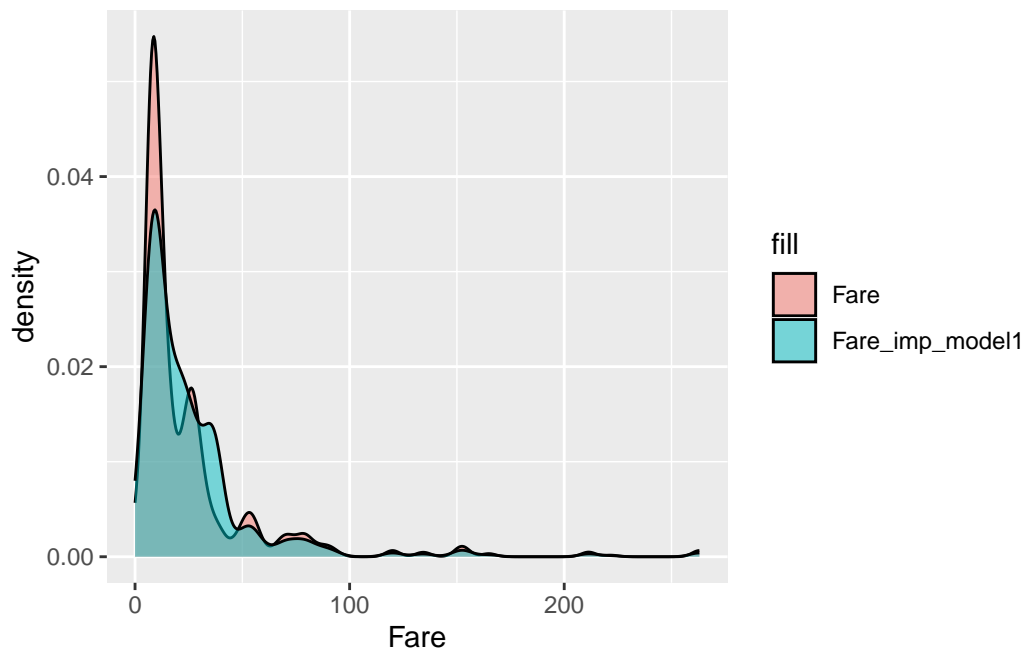
Multiple R-squared: 0.0527, Adjusted R-squared: 0.05139  
F-statistic: 40.11 on 1 and 721 DF, p-value: 4.221e-10

```
#Predecir los valores solo para las observaciones faltantes
predictions <- predict(model1,newdata = data [is.na(data$Fare),])

##Crear una nueva variable de linelist con la temperatura imputada
data$Fare_imp_model1 <- data$Fare
data$Fare_imp_model1[is.na(data$Fare)]<- predictions

#Hacer un gráfico para comparar las observaciones
ggplot(data, aes(x = Fare, fill = "Fare")) +
  geom_density(alpha = 0.5) +
  geom_density(aes(x = Fare_imp_model1, fill = "Fare_imp_model1"), alpha = 0.5)
```

Warning: Removed 168 rows containing non-finite outside the scale range  
(`stat\_density()`).



```
set.seed(3)
x<-rnorm(sum(is.na(data$Fare)), 0, sd = 33.66)
```

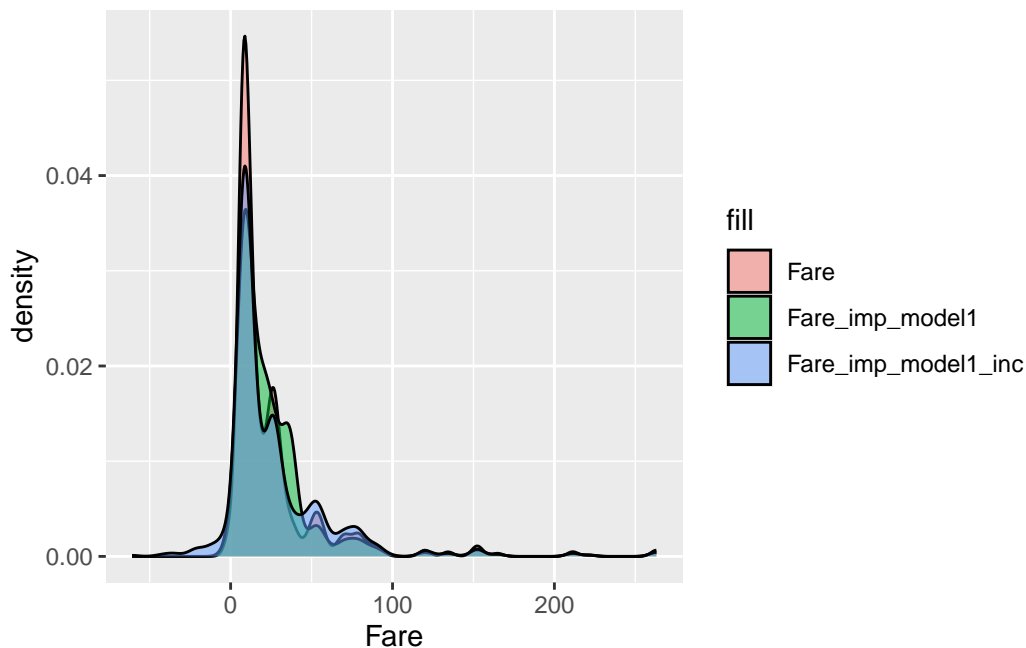
```

data$Fare_imp_model1_inc <- data$Fare
data$Fare_imp_model1_inc[is.na(data$Fare)]<- predictions + x

#Hacer un gráfico para comparar las observaciones con la media y la regresión
#Hacer un gráfico para comparar las observaciones
ggplot(data, aes(x = Fare, fill = "Fare")) +
  geom_density(alpha = 0.5) +
  geom_density(aes(x = Fare_imp_model1, fill = "Fare_imp_model1"), alpha = 0.5) +
  geom_density(aes(x = Fare_imp_model1_inc, fill = "Fare_imp_model1_inc"), alpha = 0.5)

```

Warning: Removed 168 rows containing non-finite outside the scale range (`stat\_density()`).



En ambos modelos de regresión aunque las variables estás muy asociadas, el R2 es muy bajito indicativo de que no es un buen modelo de regresión, por eso cuando se imputa no es demasiado buena, pero con el añadido de la incertidumbre se soluciona, aunque con esos valores tan altos, es un valor bastante al azar.

**4. Realiza una imputación múltiple con `mice()`, para ello ten en cuenta que seguramente no tenga sentido usar todas las variables de la base de datos y tampoco las que has imputado previamente. Después contesta a las siguientes preguntas:**

- 1.¿Con qué métodos has imputado cada variable en la función `mice`?
- 2.Una vez comprobada la imputación múltiple mediante gráficos ¿Qué te parecen las imputaciones?
- 3.¿Cuál es la proporción de la variable `*Sex*` cuando se imputa por la moda y cuando se imputa por la media?

```
data_mice<-data[,c(2,3,5,6,7,8,10,12)]
impData <- mice(data_mice,m=5,maxit=50,seed=500)
```

```
iter imp variable
1 1 Sex Age SibSp Fare
1 2 Sex Age SibSp Fare
1 3 Sex Age SibSp Fare
1 4 Sex Age SibSp Fare
1 5 Sex Age SibSp Fare
2 1 Sex Age SibSp Fare
2 2 Sex Age SibSp Fare
2 3 Sex Age SibSp Fare
2 4 Sex Age SibSp Fare
2 5 Sex Age SibSp Fare
3 1 Sex Age SibSp Fare
3 2 Sex Age SibSp Fare
3 3 Sex Age SibSp Fare
3 4 Sex Age SibSp Fare
3 5 Sex Age SibSp Fare
4 1 Sex Age SibSp Fare
4 2 Sex Age SibSp Fare
4 3 Sex Age SibSp Fare
4 4 Sex Age SibSp Fare
4 5 Sex Age SibSp Fare
5 1 Sex Age SibSp Fare
5 2 Sex Age SibSp Fare
5 3 Sex Age SibSp Fare
5 4 Sex Age SibSp Fare
5 5 Sex Age SibSp Fare
6 1 Sex Age SibSp Fare
```

6	2	Sex	Age	SibSp	Fare
6	3	Sex	Age	SibSp	Fare
6	4	Sex	Age	SibSp	Fare
6	5	Sex	Age	SibSp	Fare
7	1	Sex	Age	SibSp	Fare
7	2	Sex	Age	SibSp	Fare
7	3	Sex	Age	SibSp	Fare
7	4	Sex	Age	SibSp	Fare
7	5	Sex	Age	SibSp	Fare
8	1	Sex	Age	SibSp	Fare
8	2	Sex	Age	SibSp	Fare
8	3	Sex	Age	SibSp	Fare
8	4	Sex	Age	SibSp	Fare
8	5	Sex	Age	SibSp	Fare
9	1	Sex	Age	SibSp	Fare
9	2	Sex	Age	SibSp	Fare
9	3	Sex	Age	SibSp	Fare
9	4	Sex	Age	SibSp	Fare
9	5	Sex	Age	SibSp	Fare
10	1	Sex	Age	SibSp	Fare
10	2	Sex	Age	SibSp	Fare
10	3	Sex	Age	SibSp	Fare
10	4	Sex	Age	SibSp	Fare
10	5	Sex	Age	SibSp	Fare
11	1	Sex	Age	SibSp	Fare
11	2	Sex	Age	SibSp	Fare
11	3	Sex	Age	SibSp	Fare
11	4	Sex	Age	SibSp	Fare
11	5	Sex	Age	SibSp	Fare
12	1	Sex	Age	SibSp	Fare
12	2	Sex	Age	SibSp	Fare
12	3	Sex	Age	SibSp	Fare
12	4	Sex	Age	SibSp	Fare
12	5	Sex	Age	SibSp	Fare
13	1	Sex	Age	SibSp	Fare
13	2	Sex	Age	SibSp	Fare
13	3	Sex	Age	SibSp	Fare
13	4	Sex	Age	SibSp	Fare
13	5	Sex	Age	SibSp	Fare
14	1	Sex	Age	SibSp	Fare
14	2	Sex	Age	SibSp	Fare
14	3	Sex	Age	SibSp	Fare
14	4	Sex	Age	SibSp	Fare

14	5	Sex	Age	SibSp	Fare
15	1	Sex	Age	SibSp	Fare
15	2	Sex	Age	SibSp	Fare
15	3	Sex	Age	SibSp	Fare
15	4	Sex	Age	SibSp	Fare
15	5	Sex	Age	SibSp	Fare
16	1	Sex	Age	SibSp	Fare
16	2	Sex	Age	SibSp	Fare
16	3	Sex	Age	SibSp	Fare
16	4	Sex	Age	SibSp	Fare
16	5	Sex	Age	SibSp	Fare
17	1	Sex	Age	SibSp	Fare
17	2	Sex	Age	SibSp	Fare
17	3	Sex	Age	SibSp	Fare
17	4	Sex	Age	SibSp	Fare
17	5	Sex	Age	SibSp	Fare
18	1	Sex	Age	SibSp	Fare
18	2	Sex	Age	SibSp	Fare
18	3	Sex	Age	SibSp	Fare
18	4	Sex	Age	SibSp	Fare
18	5	Sex	Age	SibSp	Fare
19	1	Sex	Age	SibSp	Fare
19	2	Sex	Age	SibSp	Fare
19	3	Sex	Age	SibSp	Fare
19	4	Sex	Age	SibSp	Fare
19	5	Sex	Age	SibSp	Fare
20	1	Sex	Age	SibSp	Fare
20	2	Sex	Age	SibSp	Fare
20	3	Sex	Age	SibSp	Fare
20	4	Sex	Age	SibSp	Fare
20	5	Sex	Age	SibSp	Fare
21	1	Sex	Age	SibSp	Fare
21	2	Sex	Age	SibSp	Fare
21	3	Sex	Age	SibSp	Fare
21	4	Sex	Age	SibSp	Fare
21	5	Sex	Age	SibSp	Fare
22	1	Sex	Age	SibSp	Fare
22	2	Sex	Age	SibSp	Fare
22	3	Sex	Age	SibSp	Fare
22	4	Sex	Age	SibSp	Fare
22	5	Sex	Age	SibSp	Fare
23	1	Sex	Age	SibSp	Fare
23	2	Sex	Age	SibSp	Fare

23	3	Sex	Age	SibSp	Fare
23	4	Sex	Age	SibSp	Fare
23	5	Sex	Age	SibSp	Fare
24	1	Sex	Age	SibSp	Fare
24	2	Sex	Age	SibSp	Fare
24	3	Sex	Age	SibSp	Fare
24	4	Sex	Age	SibSp	Fare
24	5	Sex	Age	SibSp	Fare
25	1	Sex	Age	SibSp	Fare
25	2	Sex	Age	SibSp	Fare
25	3	Sex	Age	SibSp	Fare
25	4	Sex	Age	SibSp	Fare
25	5	Sex	Age	SibSp	Fare
26	1	Sex	Age	SibSp	Fare
26	2	Sex	Age	SibSp	Fare
26	3	Sex	Age	SibSp	Fare
26	4	Sex	Age	SibSp	Fare
26	5	Sex	Age	SibSp	Fare
27	1	Sex	Age	SibSp	Fare
27	2	Sex	Age	SibSp	Fare
27	3	Sex	Age	SibSp	Fare
27	4	Sex	Age	SibSp	Fare
27	5	Sex	Age	SibSp	Fare
28	1	Sex	Age	SibSp	Fare
28	2	Sex	Age	SibSp	Fare
28	3	Sex	Age	SibSp	Fare
28	4	Sex	Age	SibSp	Fare
28	5	Sex	Age	SibSp	Fare
29	1	Sex	Age	SibSp	Fare
29	2	Sex	Age	SibSp	Fare
29	3	Sex	Age	SibSp	Fare
29	4	Sex	Age	SibSp	Fare
29	5	Sex	Age	SibSp	Fare
30	1	Sex	Age	SibSp	Fare
30	2	Sex	Age	SibSp	Fare
30	3	Sex	Age	SibSp	Fare
30	4	Sex	Age	SibSp	Fare
30	5	Sex	Age	SibSp	Fare
31	1	Sex	Age	SibSp	Fare
31	2	Sex	Age	SibSp	Fare
31	3	Sex	Age	SibSp	Fare
31	4	Sex	Age	SibSp	Fare
31	5	Sex	Age	SibSp	Fare

32	1	Sex	Age	SibSp	Fare
32	2	Sex	Age	SibSp	Fare
32	3	Sex	Age	SibSp	Fare
32	4	Sex	Age	SibSp	Fare
32	5	Sex	Age	SibSp	Fare
33	1	Sex	Age	SibSp	Fare
33	2	Sex	Age	SibSp	Fare
33	3	Sex	Age	SibSp	Fare
33	4	Sex	Age	SibSp	Fare
33	5	Sex	Age	SibSp	Fare
34	1	Sex	Age	SibSp	Fare
34	2	Sex	Age	SibSp	Fare
34	3	Sex	Age	SibSp	Fare
34	4	Sex	Age	SibSp	Fare
34	5	Sex	Age	SibSp	Fare
35	1	Sex	Age	SibSp	Fare
35	2	Sex	Age	SibSp	Fare
35	3	Sex	Age	SibSp	Fare
35	4	Sex	Age	SibSp	Fare
35	5	Sex	Age	SibSp	Fare
36	1	Sex	Age	SibSp	Fare
36	2	Sex	Age	SibSp	Fare
36	3	Sex	Age	SibSp	Fare
36	4	Sex	Age	SibSp	Fare
36	5	Sex	Age	SibSp	Fare
37	1	Sex	Age	SibSp	Fare
37	2	Sex	Age	SibSp	Fare
37	3	Sex	Age	SibSp	Fare
37	4	Sex	Age	SibSp	Fare
37	5	Sex	Age	SibSp	Fare
38	1	Sex	Age	SibSp	Fare
38	2	Sex	Age	SibSp	Fare
38	3	Sex	Age	SibSp	Fare
38	4	Sex	Age	SibSp	Fare
38	5	Sex	Age	SibSp	Fare
39	1	Sex	Age	SibSp	Fare
39	2	Sex	Age	SibSp	Fare
39	3	Sex	Age	SibSp	Fare
39	4	Sex	Age	SibSp	Fare
39	5	Sex	Age	SibSp	Fare
40	1	Sex	Age	SibSp	Fare
40	2	Sex	Age	SibSp	Fare
40	3	Sex	Age	SibSp	Fare

40	4	Sex	Age	SibSp	Fare
40	5	Sex	Age	SibSp	Fare
41	1	Sex	Age	SibSp	Fare
41	2	Sex	Age	SibSp	Fare
41	3	Sex	Age	SibSp	Fare
41	4	Sex	Age	SibSp	Fare
41	5	Sex	Age	SibSp	Fare
42	1	Sex	Age	SibSp	Fare
42	2	Sex	Age	SibSp	Fare
42	3	Sex	Age	SibSp	Fare
42	4	Sex	Age	SibSp	Fare
42	5	Sex	Age	SibSp	Fare
43	1	Sex	Age	SibSp	Fare
43	2	Sex	Age	SibSp	Fare
43	3	Sex	Age	SibSp	Fare
43	4	Sex	Age	SibSp	Fare
43	5	Sex	Age	SibSp	Fare
44	1	Sex	Age	SibSp	Fare
44	2	Sex	Age	SibSp	Fare
44	3	Sex	Age	SibSp	Fare
44	4	Sex	Age	SibSp	Fare
44	5	Sex	Age	SibSp	Fare
45	1	Sex	Age	SibSp	Fare
45	2	Sex	Age	SibSp	Fare
45	3	Sex	Age	SibSp	Fare
45	4	Sex	Age	SibSp	Fare
45	5	Sex	Age	SibSp	Fare
46	1	Sex	Age	SibSp	Fare
46	2	Sex	Age	SibSp	Fare
46	3	Sex	Age	SibSp	Fare
46	4	Sex	Age	SibSp	Fare
46	5	Sex	Age	SibSp	Fare
47	1	Sex	Age	SibSp	Fare
47	2	Sex	Age	SibSp	Fare
47	3	Sex	Age	SibSp	Fare
47	4	Sex	Age	SibSp	Fare
47	5	Sex	Age	SibSp	Fare
48	1	Sex	Age	SibSp	Fare
48	2	Sex	Age	SibSp	Fare
48	3	Sex	Age	SibSp	Fare
48	4	Sex	Age	SibSp	Fare
48	5	Sex	Age	SibSp	Fare
49	1	Sex	Age	SibSp	Fare



```

49  2  Sex  Age  SibSp  Fare
49  3  Sex  Age  SibSp  Fare
49  4  Sex  Age  SibSp  Fare
49  5  Sex  Age  SibSp  Fare
50  1  Sex  Age  SibSp  Fare
50  2  Sex  Age  SibSp  Fare
50  3  Sex  Age  SibSp  Fare
50  4  Sex  Age  SibSp  Fare
50  5  Sex  Age  SibSp  Fare

```

Warning: Number of logged events: 500

`impData`

Class: mids

Number of multiple imputations: 5

Imputation methods:

```

Survived  Pclass      Sex      Age      SibSp      Parch      Fare Embarked
      ""      "" "logreg"  "pmm"  "pmm"      ""  "pmm"      ""

```

PredictorMatrix:

```

      Survived Pclass Sex Age SibSp Parch Fare Embarked
Survived      0      1  1  1      1      1  1      1
Pclass        1      0  1  1      1      1  1      1
Sex           1      1  0  1      1      1  1      1
Age           1      1  1  0      1      1  1      1
SibSp         1      1  1  1      0      1  1      1
Parch         1      1  1  1      1      0  1      1

```

Number of logged events: 500

```

  it im dep  meth      out
1  1  1  Sex logreg  Pclass2
2  1  1  Fare   pmm EmbarkedC
3  1  2  Sex logreg  Pclass2
4  1  2  Fare   pmm EmbarkedC
5  1  3  Sex logreg  Pclass2
6  1  3  Fare   pmm EmbarkedC

```

#The logged events significa que hay un predictor constante en todos los valores

#(o tal vez constante en todos los valores faltantes/no faltantes)

#Lo que te dice es que ha excluido esa variable o categoría para el estudio

##Métodos

`summary(impData)`

```

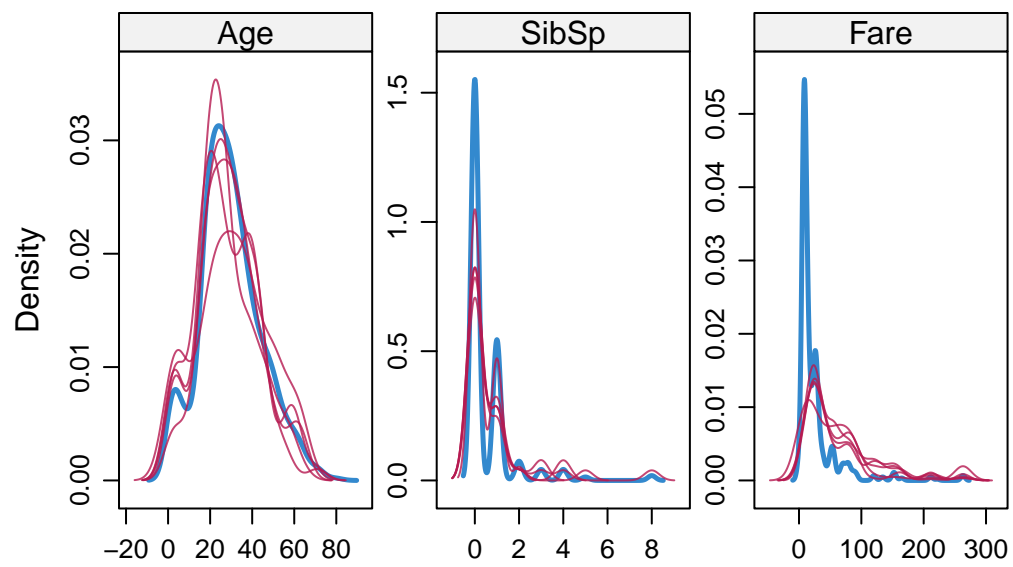
Class: mids
Number of multiple imputations: 5
Imputation methods:
Survived  Pclass    Sex      Age      SibSp    Parch    Fare Embarked
      ""      "" "logreg"  "pmm"    "pmm"      ""    "pmm"      ""
PredictorMatrix:
      Survived Pclass Sex Age SibSp Parch Fare Embarked
Survived      0      1  1  1      1      1  1      1
Pclass        1      0  1  1      1      1  1      1
Sex            1      1  0  1      1      1  1      1
Age            1      1  1  0      1      1  1      1
SibSp          1      1  1  1      0      1  1      1
Parch          1      1  1  1      1      0  1      1
Number of logged events: 500
  it im dep  meth      out
1  1  1  1 Sex logreg  Pclass2
2  1  1  1 Fare   pmm EmbarkedC
3  1  2  2 Sex logreg  Pclass2
4  1  2  2 Fare   pmm EmbarkedC
5  1  3  3 Sex logreg  Pclass2
6  1  3  3 Fare   pmm EmbarkedC

```

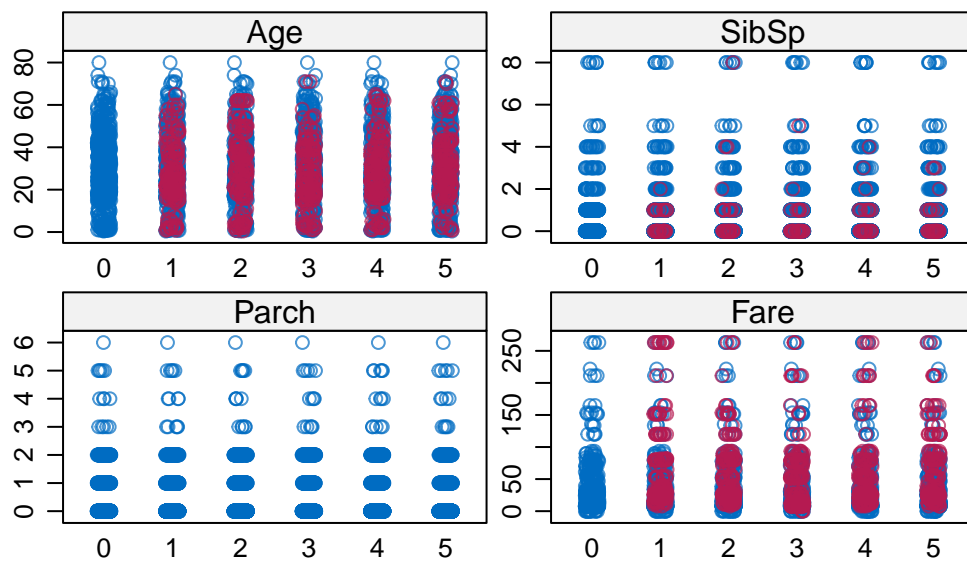
```

##Gráficos de comprobación
densityplot(impData)

```

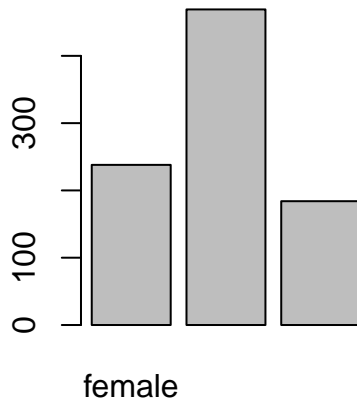


```
stripplot(impData)
```

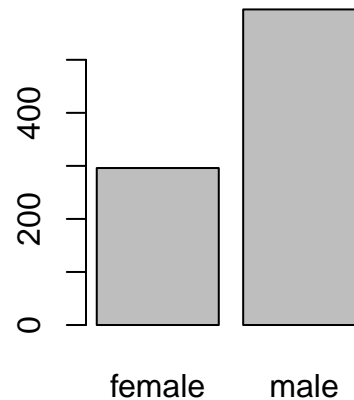


```
completeData<-complete(impData,1)
par(mfrow = c(1, 2)) # Organizar las gráficas en una 60+fila de 2 columnas
barplot(table(data$Sex, useNA = "ifany"), main = "Antes de la imputación")
barplot(table(completeData$Sex), main = "Después de la imputación")
```

**Antes de la imputación**



**Después de la imputación**



```
prop.table(table(completeData$Sex))
```

```
female    male
0.332211 0.667789
```

```
#¿Cómo es la proporción de los datos missing?
prop.table(table(data$Sex,useNA="always"))
```

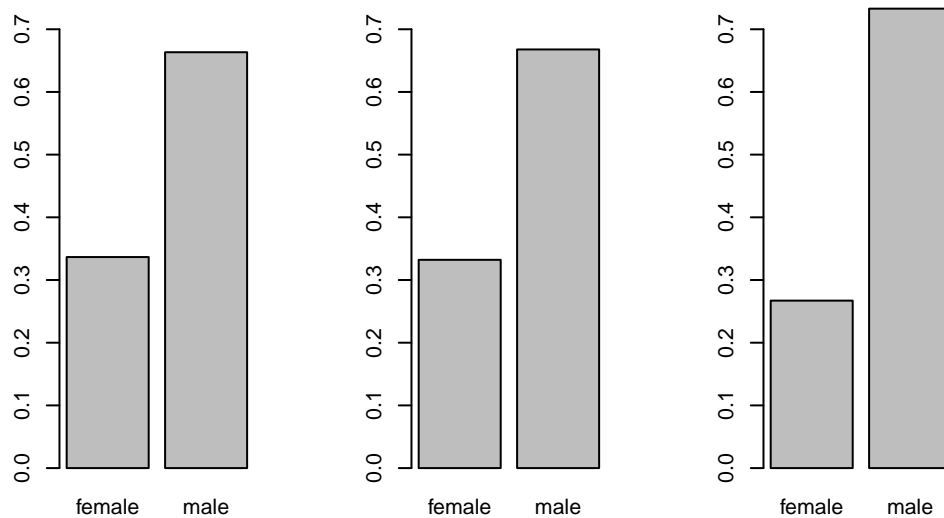
```
female    male    <NA>
0.2671156 0.5263749 0.2065095
```

```
data$Sex_moda<-data$Sex
data$Sex_moda[is.na(data$Sex_moda)]<-"male"
prop.table(table(data$Sex_moda))
```

female	male
0.2671156	0.7328844

```
par(mfrow = c(1, 3)) # Organizar las gráficas en una fila de 2 columnas
barplot(prop.table(table(data$Sex)), main = "Antes de la imputación",ylim = c(0.0,0.7))
barplot(prop.table(table(completeData$Sex)), main = "Después de la imputación con mice",ylim
barplot(prop.table(table(data$Sex_moda)), main = "Después de la imputación con la moda",ylim
```

**Antes de la imputación   después de la imputación con mice   después de la imputación con la moda**



Sex se ha imputado con logreg

Age, SibSp y Fare se ha imputado con pmm

La variable Age se imputa bastante bien, la de SibSp logra dibujar todos los picos, pero en la variable Fare no se consigui imputar bien debido a la distribución bastante rara que tiene de que la mayoría de valores son muy bajos. Quizás fuera necesario una transformación o usar otro método de imputación.

La variable sex se imputa mucho mejor con mice que con la moda