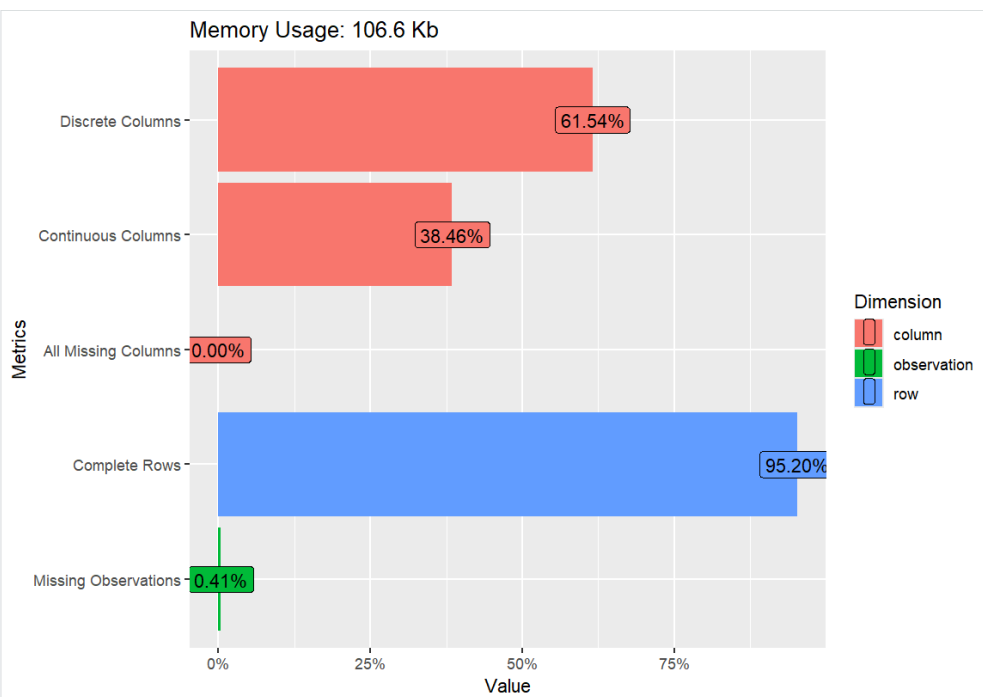
ACTIVIDAD UTILIZANDO R

El dataset seleccionado contiene información de 1000 usuarios, con información sobre su edad, estado civil y más. También incluye si compraron una bicicleta o no.

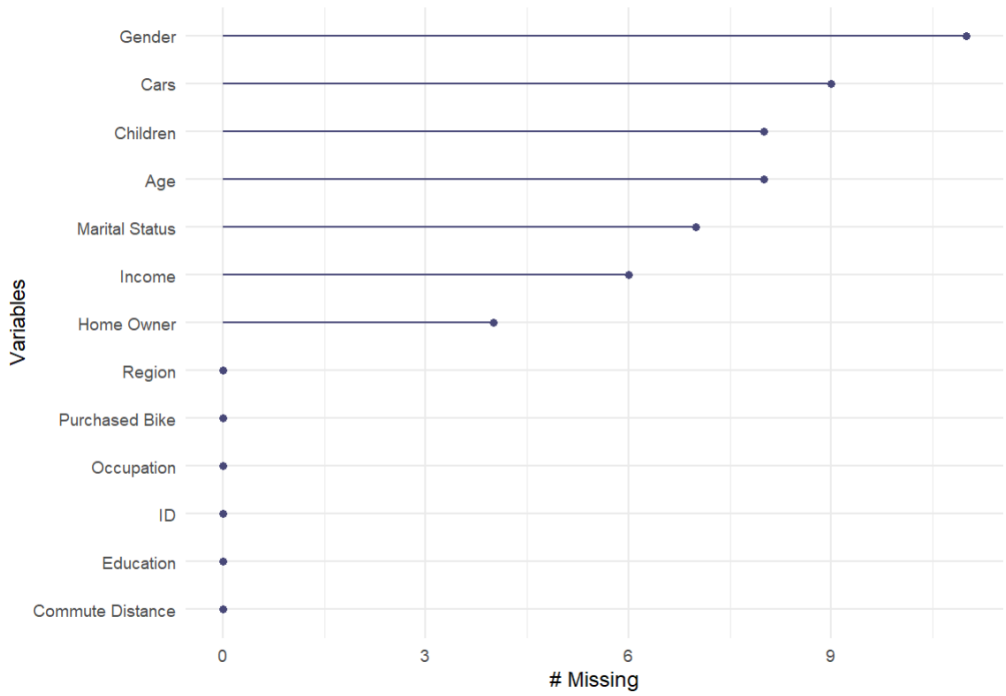
Usar skimr para resumen detallado

```
── Data Summary ────────────────────────
                           Values
Name                       df
Number of rows             1000
Number of columns          13
_______________________
Column type frequency:
  character                8
  numeric                  5
________________________
Group variables            None

── Variable type: character ──────────────────────────────────────────────────────────────────────
  skim_variable    n_missing complete_rate min max empty n_unique whitespace
1 Marital Status           7         0.993   6   7     0        2          0
2 Gender                  11         0.989   4   6     0        2          0
3 Education                0         1       9  19     0        5          0
4 Occupation               0         1       6  14     0        5          0
5 Home Owner               4         0.996   2   3     0        2          0
6 Commute Distance         0         1       9  10     0        5          0
7 Region                   0         1       6  13     0        3          0
8 Purchased Bike           0         1       2   3     0        2          0

── Variable type: numeric ────────────────────────────────────────────────────────────────────────
  skim_variable n_missing complete_rate     mean       sd    p0    p25   p50    p75   p100 hist
1 ID                    0         1     19966.    5347.   11000 15291. 19744 24471.  29447 ▇▇▇▇▇▇▇▇
2 Income                6         0.994 56268.   31068.   10000 30000  60000 70000  170000 ▇▅▂▂▁▁▁▁
3 Children              8         0.992     1.91     1.63     0     0      2     3       5 ▇▂▂▂▂▂▂▂
4 Cars                  9         0.991     1.46     1.12     0     1      1     2       4 ▇▅▅▂▂▂▂▂
5 Age                   8         0.992    44.2     11.4     25    35     43    52      89 ▇▅▅▃▂▂▂▂
>
```

Overview



Visualización de valores faltantes



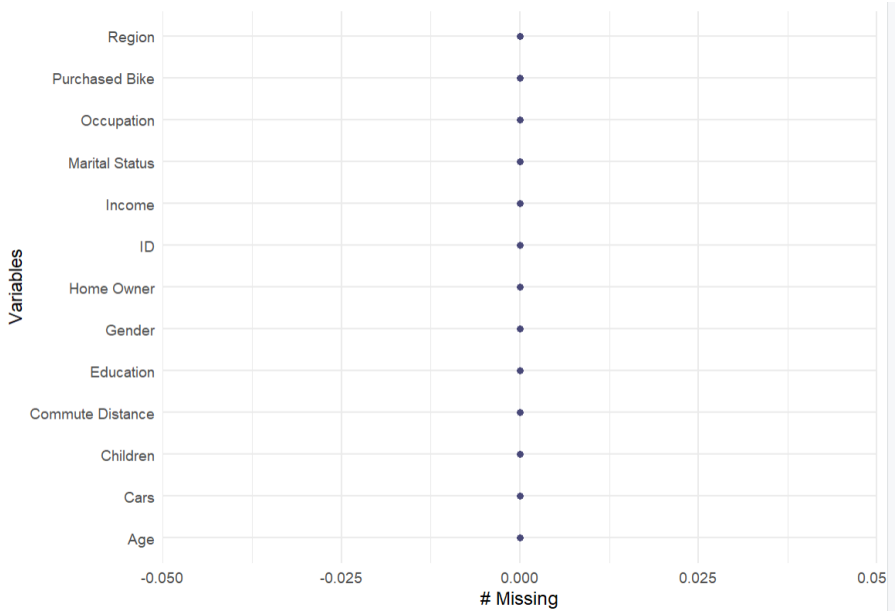
Imputar valores faltantes utilizando la mediana para las variables numéricas

```
#Imputar valores faltantes con la mediana en variables numericas
df$Income[is.na(df$Income)] <- median(df$Income, na.rm = TRUE)
df$Children[is.na(df$Children)] <- median(df$Children, na.rm = TRUE)
df$Cars[is.na(df$Cars)] <- median(df$Cars, na.rm = TRUE)
df$Age[is.na(df$Age)] <- median(df$Age, na.rm = TRUE)
```

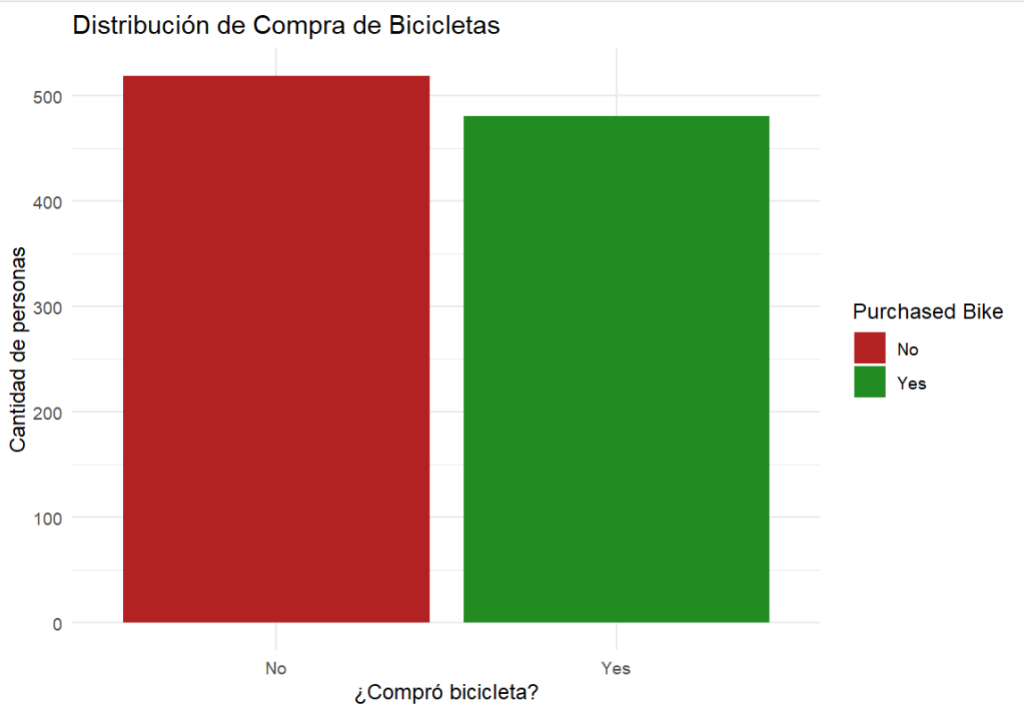
Imputar valores faltantes utilizando el modo para las variables categóricas

```
# Función para obtener el valor más frecuente (modo)
get_mode <- function(x) {
  ux <- na.omit(unique(x))
  ux[which.max(tabulate(match(x, ux)))]
}

# Imputar valores faltantes en columnas categóricas con el modo
df <- df %>%
  mutate(across(where(is.character), ~ ifelse(is.na(.), get_mode(.), .)))
```

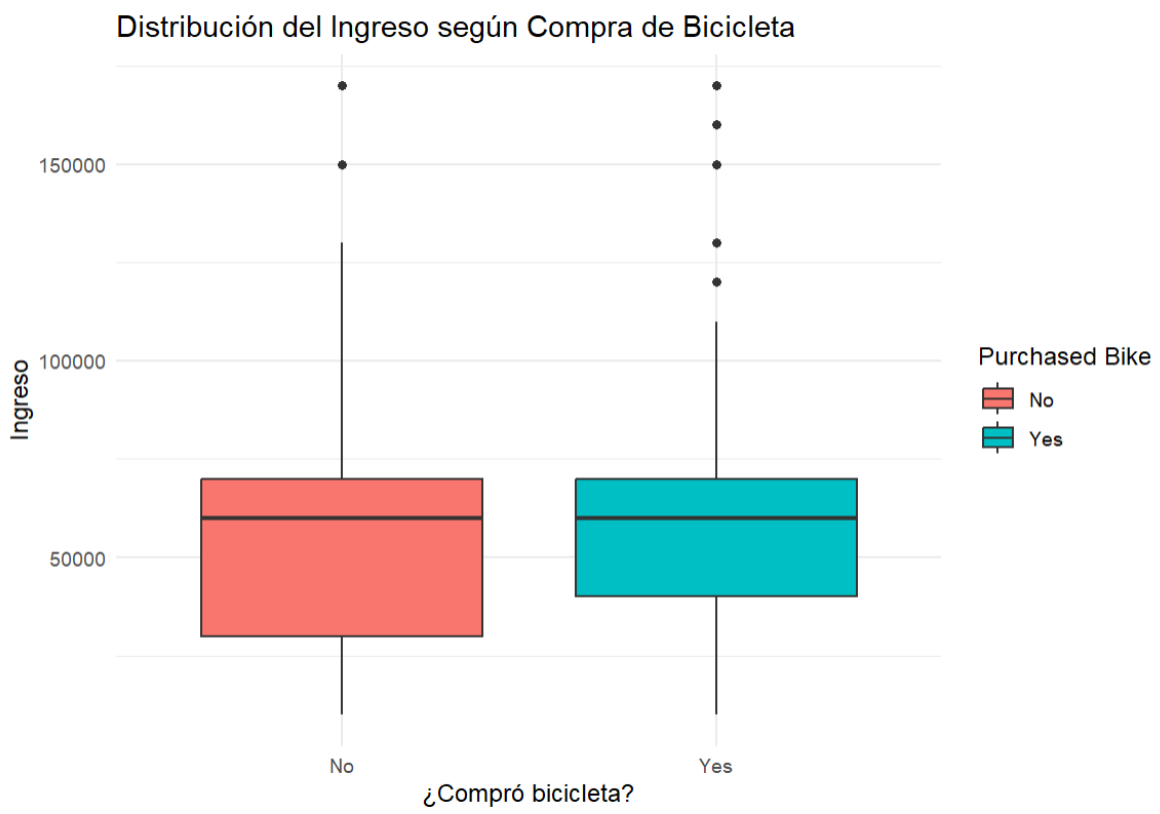
Visualización de valores faltantes después de las imputaciones



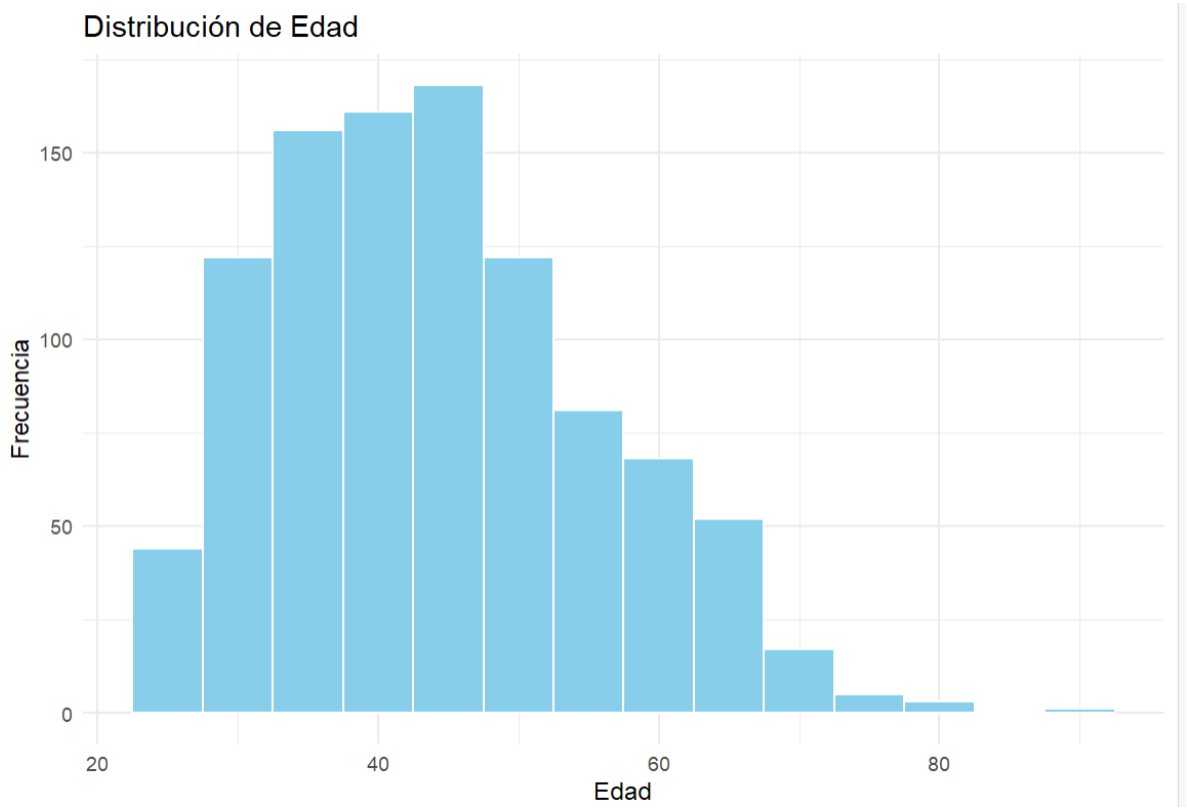
Distribución de las variables "Purchased Bike"

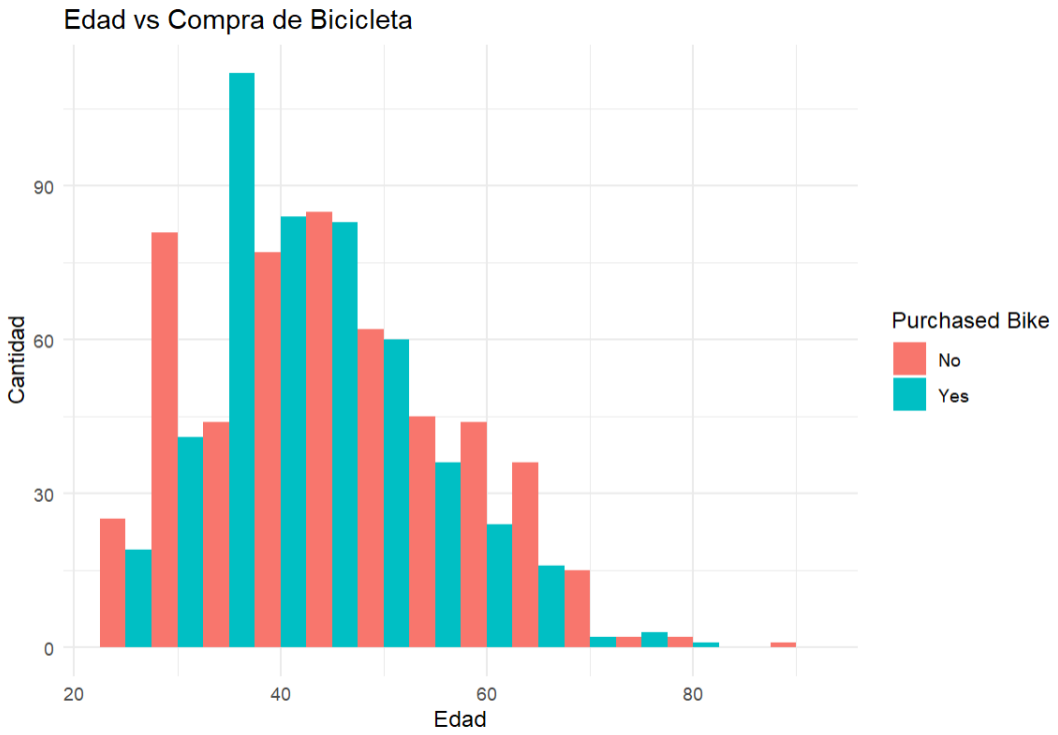


Boxplot de ingresos por compra de bicicleta



Histograma de Edades





Conclusiones

Compras de bicicleta: La mayoría no compró bicicleta.

Ingresos: Hay una gran variabilidad. Quienes compraron bicicleta tienden a tener ingresos más altos (según el boxplot).

Edad: Las compras se concentran entre los 30 y 50 años.

Las bicicletas parecen ser más populares entre adultos de mediana edad, con ingresos medios, pocas responsabilidades familiares (0-2 hijos) y que poseen 1 o 2 autos. por lo cual la compra de bicicleta es para ejercicio, recreación y no para sustituir su transporte principal.

Para el siguiente paso se puede tomar en cuenta las variables más relevantes, realizar un mejor tratamiento con los datos nulos y crear un modelo predictivo para saber si un cliente comprará o no.