Taller\_2

John Esteban

2024-03-25

## Warning: package 'rvest' was built under R version 4.2.3

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4  
## ✔ tibble 3.1.7 ✔ dplyr 1.0.9  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0  
## ✔ readr 2.1.2 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ readr::guess\_encoding() masks rvest::guess\_encoding()  
## ✖ dplyr::lag() masks stats::lag()

## Warning: package 'devtools' was built under R version 4.2.3

## Loading required package: usethis

## Warning: package 'usethis' was built under R version 4.2.3

## Warning: package 'gridExtra' was built under R version 4.2.3

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

## Warning: package 'kableExtra' was built under R version 4.2.3

##   
## Attaching package: 'kableExtra'

## The following object is masked from 'package:dplyr':  
##   
## group\_rows

## Warning: package 'png' was built under R version 4.2.2

## ℹ SHA-1 hash of file is "08930743b60b5cdcfadeb8abaa79e488be87522f"

## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.  
## ℹ Please use `all\_of()` or `any\_of()` instead.  
## # Was:  
## data %>% select(variables\_hogares)  
##   
## # Now:  
## data %>% select(all\_of(variables\_hogares))  
##   
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.  
## ℹ Please use `all\_of()` or `any\_of()` instead.  
## # Was:  
## data %>% select(variables\_personas)  
##   
## # Now:  
## data %>% select(all\_of(variables\_personas))  
##   
## See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
## generated.

## ℹ SHA-1 hash of file is "3ff8c4f235a8505f47a4ebea967cbede1a557c4c"

# Introduccion

La pobreza es uno de los temas centrales de investigación en la economía del bienestar y en las agendas políticas de todos los gobiernos, especialmente en los países en vías de desarrollo. Sin embargo, no hay un consenso en su definición y en su forma de ser medida. A pesar de esto, se puede entender que ser pobre es no disponer de los recursos para obtener los medios mínimos de subsistencia, una definición general. El Banco Mundial define la pobreza como incapacidades para llevar una vida plena, desagregándolo en buena alimentación, calidad de la vivienda, buena salud, buena educación y tener un trabajo digno. También tenemos la definición de uno de los economistas que más influencia tiene en el estudio de la pobreza, Amartya Sen. Con la teoría de las capacidades, este autor dice que la pobreza no es simplemente la falta de ingreso, sino la falta de capacidades básicas y la libertad para obtenerlo, definición muy parecida a la propuesta del Banco Mundial. En resumen, los avances del estudio teórico de la pobreza han reposado en que es un fenómeno multidimensional y no solamente carencia de ingresos.

De la misma forma que las definiciones, la forma de medir la pobreza es variada, donde se tiene en cuenta la multidimensionalidad de la pobreza que tanto se menciona en la etapa de definición. En primer lugar, se tiene el NBI (Necesidades Básicas Insatisfechas) que señala la insuficiencia que tiene un hogar en una de las siguientes necesidades básicas: Vivienda con Materiales adecuados, servicios públicos de acueducto y alcantarillado, nivel bajo de hacinamiento, bajo grado de dependencia. También se encuentra el método de la línea de pobreza, que calcula el costo de una canasta básica de bienes y servicios, luego calcula los ingresos del hogar y se compara con esta línea, si no alcanza este mínimo entonces se considera un hogar pobre. Por último, también se trabaja con el Índice de Condiciones de vida, que comprende variables que miden la calidad de la vivienda, el capital humano actual y potencial, el acceso a la calidad de los servicios y las condiciones del hogar. En este trabajo se va a trabajar con el indicador de línea de pobreza.

El objetivo de este artículo es encontrar los determinantes económicos y sociales que más inciden en clasificar a una persona como pobre o no pobre a través de la medida de línea de pobreza. Para esto se utilizan los datos del DANE que vienen de la Misión de Empalme de las Series de Empleo, Pobreza y Desigualdad. En esta base de datos se encuentra a nivel de hogar y de personas y recoge toda la información sobre las condiciones de empleo de las personas, además de todas las características socioeconómicas generales y que son importantes para determinar la pobreza de un hogar. Por tanto, con estas variables se puede estimar en términos de probabilidades las características de la población menos favorecida.

El estudio de los determinantes de la pobreza es importante porque muestra los temas en los que es necesario focalizar las políticas públicas. En Colombia se han realizado varios estudios que buscan encontrar los determinantes de la pobreza con datos de tipo transversal. Nunes & Ramírez (2002) utilizan un modelo de respuesta cualitativa para encontrar las características socioeconómicas de los hogares más pobres a través de la estimación de probabilidades. Estos autores encuentran que el nivel educativo, la cantidad de los miembros del hogar, la tasa de ocupación y la ubicación del hogar son variables que afectan la probabilidad de ser o no hogares pobres.

En la misma dirección, Chaves (2010) encuentra con un modelo logit que la educación del jefe de hogar, la posesión de activos, el tamaño del hogar, el tipo de vivienda y el género del jefe de hogar son determinantes de que este se encuentre o no en situación de pobreza. De esta forma, la mayoría de los estudios de los determinantes de la pobreza tienen una forma similar de trabajar, se establecen las variables de control, se crea la variable de respuesta binaria a través de la línea de pobreza y se procede con la estimación de los parámetros con modelos probit o logit (Nunes & Ramírez, 2002), (Nunes, Ramírez, Cuesta, 2005), (Chaves, 2010), (Marrugo et al., 2015).

Un factor común que utilizan todos los estudios es trabajar y estimar la pobreza de un hogar a través de las condiciones del jefe de hogar, esto debido a que es el responsable sobre todo económico de todos los integrantes. Por esta razón, es importante estudiar la posición del jefe de hogar en el mercado laboral, ya que muestra la capacidad de los ingresos y calidad de vida. Investigar sobre su posición en el mercado y si trabaja en condiciones de informalidad puede ser un determinante de la condición de pobreza. Esto debido a que trabajar en condiciones de informalidad, por sus características precarias y bajos salarios, puede explicar la pobreza en un hogar (Torres et al., 2022). También la edad del jefe de hogar puede servir como determinante de la pobreza porque, desde la teoría, se considera que hay edades donde es tiene más incidencia el desempleo como en la población joven y más adulta (Klose, 2012).

# Datos

## Seleccion de los datos

La seleccion de las variables se hacen tomando como referencia los estudios analizados para este proceso y todo se trabaja a nivel de hogar como lo recomienda Chaves(2010), (Torres, et al,2022), (Klose,2012)

* p6050: Si es jefe de hogar
* p620: Sexo del jefe de hogar
* p6100: Tipo de regimen
* p6210: Educacion
* p6240: Empleado
* p6430: Posicion laboral del jefe de hogar
* Personas por habitacion
* Tipo de vivienda: propia o no
* Edad del jefe de hogar
* Zona: Rural o Urbana
* Tipo de trabajo

## Limpieza de los datos.

Todo el analisis exploratorio de datos y la modelación se hace para las personas que han reportado que son jefes de hogar, por esto la limpieza mas importante es crear la variable jefe, que establece que si es jefe de hogar tiene el valor de 1 y 0 en los otros caso, por esto al momento de traer las variables del nivel de personas al de hogar se hace tomando como referencia si es jefe de hogar,

train\_personas$jefe<-ifelse(test = train\_personas$P6050==1,1,0)  
test\_personas$jefe<-ifelse(test = test\_personas$P6050==1,1,0)

Para automatizar el proceso de traer las caracteristicas que pueden explicar la pobreza monetaria de un hogar se crea la función crear\_, que tiene el siguiente funcionamiento. Primero se guarda la función segun seal nombre de la variable a traer, esta funcion recibe como parametro la base de datos sobre la cual se va a construir la variable, el primer filtro que se aplica es que en la base de datos de personas debe estar filtrada solo para los jefes de hogar, del dataframe se toma la variable de interes junto con el identificador unico de personas que crea la base de datos y por ultimo se crea el join con la base de datos de hogares por medio de este identificador. La función llamada traer variable garantiza que no se repitan los id y que tenenmos registros unicos en toda la muestra.

crear\_sexo\_jefe<-function(df){  
 aux<-df %>% filter(jefe==1)  
 aux2<-data.frame(sexo\_jefe=aux$P6020,id=aux$id)  
 df<-left\_join(df,aux2,by="id")  
 return(df)  
}  
train\_personas<-crear\_sexo\_jefe(train\_personas)  
train\_hogares<-traer\_variable(train\_hogares,train\_personas,"sexo\_jefe")  
train\_hogares$sexo\_jefe<-as.factor(train\_hogares$sexo\_jefe)  
  
test\_personas<-crear\_sexo\_jefe(test\_personas)  
test\_hogares<-traer\_variable(test\_hogares,test\_personas,"sexo\_jefe")  
test\_hogares$sexo\_jefe<-as.factor(test\_hogares$sexo\_jefe)

Este proceso se hace para cada una de las variables , en este dosumento solo se presenta el caso para el sexo del jefe del hogar, en el script Creación de variables de interes esta el proceso para cada una de las variables.

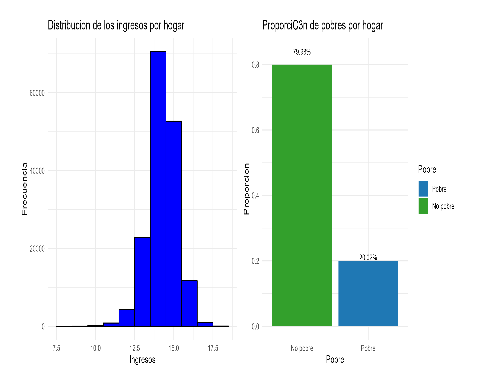
Este proceso se hace para las siguientes variables:

## Analisis exploratorio de los datos

El analisis exploratorio se datos se divide en dos partes: Analisis univariado y analisis multivariado de datos. En el analisis univariado el proposito es analizar como se distribuyen de forma individual las variables que vamos a utilizar en el proceso de modelacion. Luego con el analisis multivariado nuestro objetivo es ver como se relacionan las variable independientes con el ingreso de los hogares.

### Analisis univariado

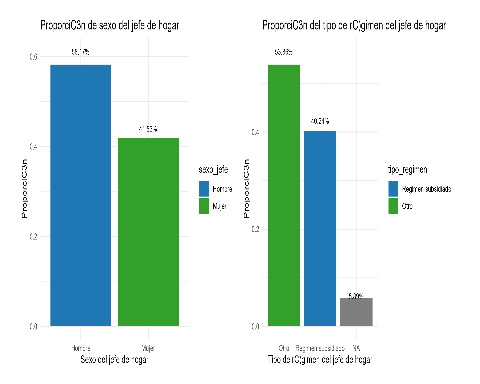
ingresos\_pobreza\_img <-readPNG('ingresos\_pobreza.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])



Estas son las dos variables objetivo que se van a explicar en este informe. En la primera gráfica se presentan los ingresos agregados del hogar, que se utilizan para determinar si un hogar es pobre utilizando la línea de pobreza, mientras que la variable ‘Pobre’ está categorizada como 0 cuando el hogar no es pobre y 1 cuando lo es. Los datos muestran que aproximadamente el 80% de los hogares no están en condición de pobreza, lo que implica un desafío en términos del proceso de modelado debido a que hay un claro desbalance de clases.

### Variables explicativas

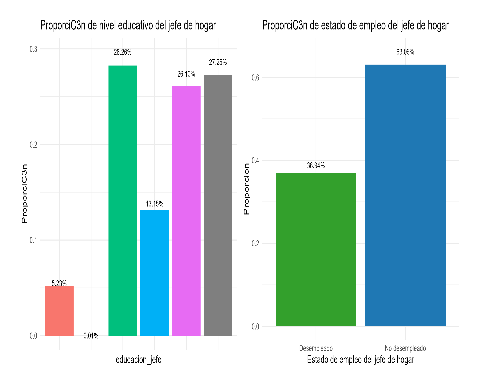
ingresos\_pobreza\_img <-readPNG('sexo\_tipo.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])



En este caso, observamos que la representación de jefes de hogar masculinos es ligeramente mayor que la de las mujeres que encabezan el hogar. Según la literatura económica, cuando la mujer es la jefa de hogar, existe un mayor riesgo de pobreza en el hogar. Esta situación puede atribuirse a diversos factores, como la brecha salarial de género y la responsabilidad desproporcionada de cuidado y trabajo doméstico no remunerado.

Respecto a los jefes de hogar afiliados al régimen subsidiado, representan aproximadamente el 42% del total. Esta variable es crucial de analizar, ya que el acceso a este mecanismo por parte de los colombianos para obtener servicios cuando no tienen capacidad de pago, sirve como una medida indirecta de la condición de pobreza de un hogar.

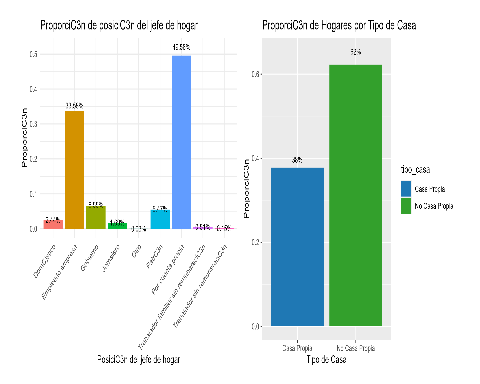
ingresos\_pobreza\_img <-readPNG('eduacion\_empleo.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])



Para revisar la educación de los jefes de hogar, primero se filtran aquellos que no saben o no responden, y se calcula la proporción con respecto al total de personas que reportaron su nivel educativo. En este análisis, observamos que la mayoría de los jefes de hogar tienen al menos aprobada la educación básica secundaria. Esta variable es una de las más importantes para predecir el nivel de ingresos de un hogar. Como se mencionó en la introducción de este informe, a medida que el jefe de hogar obtiene un mayor nivel de educación formal, la probabilidad de ser pobre disminuye, ya que una mayor educación se traduce en mayores niveles de ingresos. Sin embargo, es preocupante la gran cantidad de personas que no reportan su último nivel educativo alcanzado.

El estado de empleo es fundamental para determinar si un hogar está en condición de pobreza. Esto se explica porque si el jefe de hogar está desempleado, significa que el hogar no está percibiendo ingresos por parte de la persona encargada del sustento familiar, lo cual aumenta la probabilidad de ser pobre. En este caso, observamos que aproximadamente el 37% de los jefes de hogar no están empleados, una cifra considerable que puede explicar la situación del hogar.

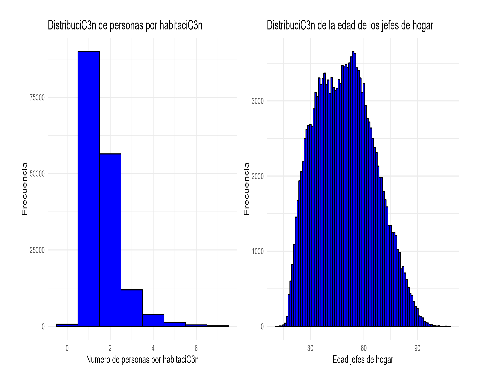
ingresos\_pobreza\_img <-readPNG('casa\_posicion.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])



La posición del jefe de hogar en el empleo es importante para determinar si el hogar corre riesgo de estar en situación de pobreza. Se espera que los jefes de hogar que trabajen sin remuneración tengan un mayor riesgo de pobreza. Sin embargo, en nuestro análisis no se observa que esta sea la situación predominante, ya que al combinar estas dos categorías apenas llegan al 1%.

Tener casa propia, en principio, puede ayudar a explicar la situación de pobreza del hogar. Esto se debe a que contar con una vivienda propia permite disponer de una mayor cantidad de ingresos que pueden destinarse a cubrir otras necesidades. La propiedad de una casa se considera una inversión y un seguro en momentos de dificultades económicas. Se supone que las personas que no tienen casa propia podrían ser más vulnerables relativamente a ser pobres. En nuestra muestra, observamos que la mayoría de los jefes de hogar no cuentan con vivienda propia y deben pagar arriendo por su alojamiento. Es fundamental profundizar en esta información para comprender mejor la dinámica de la pobreza en los hogares.

ingresos\_pobreza\_img <-readPNG('personas\_edad.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])



El número de personas por habitación también se considera importante para predecir la pobreza, ya que se espera que a mayor número de personas por habitación, exista un hacinamiento que aumente el riesgo de pobreza. En nuestro análisis, la mayoría de los datos se concentran en 2 personas por habitación, lo cual es una situación normal.

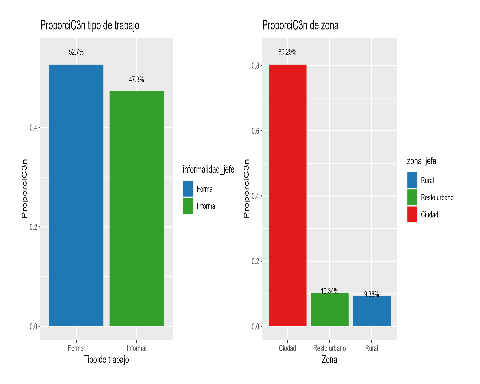
Es fundamental incorporar la edad del jefe del hogar en nuestro análisis, ya que está intrínsecamente ligada a la teoría del ciclo de vida en economía, la cual postula que las personas atraviesan diferentes etapas en su vida, cada una con características económicas específicas que pueden influir en su vulnerabilidad a la pobreza.

En el contexto colombiano, esta teoría cobra especial relevancia, ya que tanto los jóvenes como los ancianos son considerados grupos de población vulnerables. Por un lado, los jóvenes tienden a enfrentar mayores dificultades económicas debido a su entrada reciente al mercado laboral y su menor experiencia, lo que puede traducirse en ingresos más bajos y una mayor vulnerabilidad financiera. Este fenómeno se ve exacerbado por las altas tasas de desempleo que suelen afectar a este grupo demográfico. Por lo tanto, es plausible suponer que los jefes de hogar más jóvenes enfrenten desafíos adicionales para alcanzar la estabilidad financiera, lo que podría aumentar su probabilidad de experimentar situaciones de pobreza.

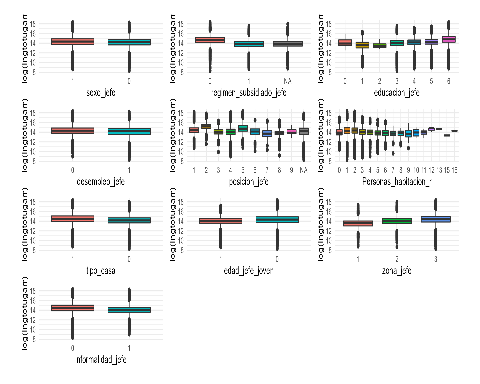
Por otro lado, la teoría del ciclo de vida también reconoce que los ancianos, especialmente después de su retiro, pueden enfrentar dificultades económicas debido a ingresos fijos y limitados, como pensiones o jubilaciones, que pueden no ser suficientes para cubrir todas sus necesidades. A pesar de haber contribuido al mercado laboral durante años, las limitaciones en los ingresos de jubilación pueden dejar a los ancianos en una situación económica precaria, lo que aumenta la probabilidad de que los hogares encabezados por personas mayores estén en situación de pobreza.

Si bien en nuestra muestra de datos observamos una distribución uniforme de las edades de los jefes de hogar, la teoría del ciclo de vida nos permite anticipar que los hogares liderados por personas más jóvenes y más mayores podrían enfrentar mayores dificultades económicas, cada uno por razones asociadas a su etapa de vida.

ingresos\_pobreza\_img <-readPNG('zona\_informal.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])

 La distinción entre el tipo de empleo a través de la informalidad laboral es un factor crucial para determinar las condiciones de pobreza de un hogar. El trabajo informal se caracteriza por salarios bajos y condiciones laborales precarias, lo que puede aumentar significativamente el riesgo de pobreza para quienes se desempeñan en este sector. En nuestros datos, observamos que aproximadamente el 47% de las personas trabajan en condiciones de informalidad, una cifra que concuerda con los datos nacionales y subraya la magnitud del desafío que representa la informalidad laboral en términos de pobreza.

Además, la ubicación geográfica del hogar también desempeña un papel crucial, ya que las viviendas ubicadas en áreas más alejadas del centro económico suelen considerarse más vulnerables. Esta disparidad se explica en parte por la teoría del centro-periferia, que destaca las diferencias en el desarrollo económico entre las áreas urbanas centrales y las regiones periféricas. En el contexto urbano, el centro se caracteriza por un mayor desarrollo económico e industrial, mientras que las áreas rurales tienden a tener un menor desarrollo, con actividades económicas centradas en la agricultura y la extracción de recursos, que generalmente se asocian con salarios más bajos. Aunque en nuestra muestra hay una representación limitada del área rural, este análisis nos brinda una aproximación útil para evaluar la probabilidad de pobreza en función de la ubicación del hogar y su relación con la teoría del centro-periferia.

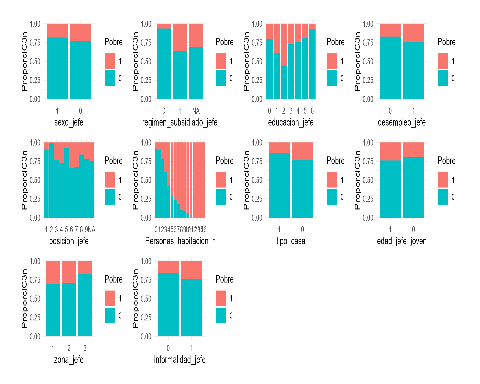


En la gráfica anterior, se observan las variables seleccionadas en relación con el logaritmo de los ingresos del hogar. Este enfoque se utiliza para normalizar los datos y facilitar la comparación entre las variables de interés. En términos generales, se puede apreciar que las personas afiliadas al régimen subsidiado tienen ingresos ligeramente inferiores en comparación con aquellos que son cotizantes. Del mismo modo, se evidencia que las personas desempleadas tienden a tener ingresos menores en el hogar. En cuanto al sexo del jefe de hogar, las diferencias en los ingresos no son tan marcadas. Sin embargo, se observan diferencias significativas en los ingresos agregados según el nivel educativo del jefe de hogar y su posición laboral.

Además, se aprecia que a medida que aumenta el número de personas por habitación, la media de ingresos del hogar tiende a disminuir, lo cual coincide con lo reportado en la literatura. Respecto a las variables tipo\_casa y edad\_jefe\_joven, los resultados son coherentes con lo esperado, ya que los jefes de hogar que poseen vivienda propia y son adultos suelen tener mayores ingresos, lo que reduce la probabilidad de estar en situación de pobreza.

El tema de la informalidad laboral también muestra diferencias significativas en los ingresos de las personas. En general, aquellos que trabajan en condiciones de informalidad tienen ingresos más bajos, lo que aumenta su vulnerabilidad económica y, por ende, la probabilidad de estar en situación de pobreza.

ingresos\_pobreza\_img <-readPNG('cualitativas.png')  
par(mar = rep(0, 4))  
plot.new()  
plot.window(xlim = c(0, dim(ingresos\_pobreza\_img)[2]), ylim = c(0, dim(ingresos\_pobreza\_img)[1]))  
rasterImage(ingresos\_pobreza\_img, 0, 0, dim(ingresos\_pobreza\_img)[2], dim(ingresos\_pobreza\_img)[1])

 Cuando se analiza en relación con la variable de pobreza, se observa un patrón muy similar al analizar los ingresos del hogar. Sin embargo, se destaca que la tenencia de vivienda propia marca una diferencia más significativa que cuando se analiza únicamente por medio de los ingresos. Esto sugiere que la posesión de vivienda propia puede ser un factor más determinante en la reducción de la probabilidad de pobreza, independientemente de los niveles de ingresos del hogar.

# Estimacion de los modelos

# Conclusiones