Anexo 2

Proceso de análisis de la encuesta de Fundación Secretariado Gitano realizado con R

# Introducción

El análisis que se presenta en este anexo parte de los datos de la Encuesta realizada para el informe “Estudio comparado sobre la situación de la población gitana en España en relación al empleo y la pobreza 2018”. Madrid." (Fundación Secretariado Gitano, 2019). La Encuesta utiliza un tamaño muestral de 1.500 personas, lo que permite inferir resultadoscon un margen de error del +2,5%. En este anexo no entraremos en las especificidades metodológicas de diseño de encuesta, dado que para esta información nos remitimos al capítulo metodológico del mencionado informe.

Sólo destacaremos algunos aspectos por los cuales hemos optado por utilizar los microdatos de esta encuesta para la consecución de los objetivos de investigación.

El **objetivo inicial de la encuesta** es cuantificar y analizar la situación socio-laboral del pueblo gitano en el estado español. La encuesta reúne indicadores extraídos de la Encuesta de Población Activa (EPA) con el objetivo de poder comparar la muestra con el conjunto de la población.

La justificación para la elección de la base de datos de la encuesta de la Fundación Secretariado Gitano puede sintetizarse en los siguientes motivos. En primer lugar, a partir de la fase inicial de investigación, se detectó que una parte muy significativa de la población implicada en los casos parte del análisis cualitativo pertenecían al pueblo gitano. Ante esta particularidad, se puso en evidencia la necesidad de realizar una aproximación cuantitativa específica hacia este colectivo que ha sido posible gracias al acceso a los microdatos resultado de esta encuesta. En segundo lugar, la posibilidad de profundizar el análisis cuantitativo en relación a la situación de precariedad energética de un colectivo minorizado, nos permite construir conocimiento y metodologías concretas que puedan ser aplicables a otros contextos.

## Limitaciones del análisis cuantitativo

Entre las limitaciones del presente análisis cabe mencionar las siguientes. A diferencia del análisis realizado sobre la encuesta EINSFOESSA (análisis detallado en el Anexo 1) la encuesta de la Fundación Secretariado Gitano no incluye las variables necesarias para un análisis completo de los indicadores de pobreza y vulnerabilidad energética oficiales utilizados por el gobierno español. Por este motivo, el análisis que se presenta sólo puede tener un alcance parcial, limitado a los indicadores consensuales de autopercepción de temperatura adecuada en los hogares y existencia de retrasos en el pago de facturas de suministros básicos.

# Preparación de los datos a partir de la encuesta

En primer lugar, importaremos la base de datos que utilizaremos en el marco de nuestro análisis:

En segundo lugar, debemos explorar las variables de la base de datos utilizada:

Una vez se ha estudiado la estructura de la base de datos, debemos transformar las variables en factores (categóricas) o en números (numéricas) y explorar los niveles factoriales de las mismas. Cómo vemos, hay 317 variables pero no es estrictamente cierto ya que muchas de ellas son categorias de variables existentes separadas como variables categóricas con opción sí/no, o 1-0. Para que el software R no entienda que esos 1 y 0 son números, sino categorías, deben transformarse en factores.

Ahora bien, no podemos optar por una transformación masiva de todas las variables ya que pueden haber diversas de ellas que sean númericas. Para ello, como primer paso del análisis deberemos identificarlas.

Primero convertiremos todas las variables de “character” a factores, posteriormente detectaremos cuales eran númericas y deben reconvertirse. Para ello, hemos creado una función propia que nos permita convertir los datos factoriales mixtos detectados a datos numéricos.

# Creamos una tabla 'tibble' para facilitar la manipulación de los datos   
pobgit2 <- as.tibble (pobgit)  
  
  
# Renombramos algunas de las variables que usaremos con mayor asiduidad   
pobgit2 <- pobgit2 %>% rename(  
 temp\_adecuada = P51\_3,   
 retrasos = P51\_5  
 )  
  
  
# Creamos una función de limpieza de datos factoriales mixtos a numéricos   
  
limp\_factor\_mixto <- function(X){  
 var\_num <- as.character(X)  
 pobgit2 %>% mutate(var\_num = case\_when(var\_num %in% c("No", "NS", "NC") ~ "0",   
 TRUE ~ var\_num))  
 var\_num <- as.numeric(var\_num)  
 var\_num[is.na(var\_num)] <- 0  
 return(var\_num)  
}

Una vez hemos creado la función de limpieza de datos, procederemos a aplicarla al conjunto de datos detectados como problemáticos. para hacerlo, dividiremos el conjunto de datos en dimensiones de interés para nuestro análisis que se refieren a características socioeconómicas de la muestra que nos proporcionaran información valiosa. A continuación, mostramos el proceso de análisis de datos para cada una de las dimensiones seleccionadas.

## Variables y datos referentes a ingresos del hogar

La primera dimensión se refiere a todas las variables y datos disponibles a ingresos del hogar. El primer paso será aplicar la función de limpieza de datos a todos aquellos datos que hemos detectado como problemáticas en el análisis preliminar:

A continuación, realizaremos las operaciones de tratamiento de datos necesarias para obtener una variable que nos indique los ingresos totales del hogar, a partir de los datos parciales sobre ingresos de la base de datos:

# Resumen de la variable numérica creada sobre ingresos mensuales en el hogar   
str(pobgit2$A9\_num)

## num [1:1492] 400 426 600 0 200 2800 1000 1000 300 0 ...

# Creamos una función para calcular los ingresos totales del hogar de acuerdo con los ingresos por conceptos y totales   
  
pobgit2 <- pobgit2 %>% mutate (ing\_hogar\_conc = select(., A8\_1\_num:A8\_7\_num) %>% replace(is.na(.), 0) %>% rowSums())  
  
summary(pobgit2$ing\_hogar\_conc)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0 400.0 700.0 784.8 1027.0 5700.0

Creamos una fórmula con un doble objetivo: (1) detectar qué importe entre los ingresos totales o por conceptos es más alto y seleccionarlo, y (2) detectar si en la variable A9 no constan ingresos, y en ese caso, acudir a las variables anteriores aunque sean más bajos

pobgit2$ing\_hogar\_total <- ifelse(pobgit2$ing\_hogar\_conc > pobgit2$A9\_num, pobgit2$ing\_hogar\_conc, pobgit2$A9\_num)  
  
  
pobgit2 %>% select(A9\_num, ing\_hogar\_conc, ing\_hogar\_total) %>% print() # Comprobamos que el resultado final es aquel que esperamos comparando los datos de la variable A9, la variable creada de ingresos por conceptos, y finalmente nuestra nueva variable del total de ingresos del hogar

## # A tibble: 1,492 x 3  
## A9\_num ing\_hogar\_conc ing\_hogar\_total  
## <dbl> <dbl> <dbl>  
## 1 400 400 400  
## 2 426 426 426  
## 3 600 600 600  
## 4 0 752 752  
## 5 200 200 200  
## 6 2800 2800 2800  
## 7 1000 1000 1000  
## 8 1000 1000 1000  
## 9 300 300 300  
## 10 0 450 450  
## # ... with 1,482 more rows

## Variables y datos referentes a la edad

En primer lugar, crearemos una función propia de limpieza para los datos disponibles sobre la edad de los miembros de la unidad familiar. En este caso, nuestro objetivo será mantener los datos perdidos (NA) en aras de diferenciar el tamaño de cada unas de las unidades familiares incluídas en la base de datos.

#Creamos función para limpiar las variables sobre edad, en las que queremos que se mantengan las NA  
  
limp\_factor\_mixto2 <- function(X){  
 var\_z <- as.character(X)  
 pobgit2 %>% mutate(var\_z = case\_when(var\_z %in% c("No", "NS", "NC") ~ "0",   
 TRUE ~ var\_z))  
 var\_z <- as.numeric(var\_z)  
 #var\_z[is.na(var\_z)] <- 0  
 return(var\_z)  
}  
  
#Limpiar las variables referentes a la edad   
pobgit2$A1\_04\_num <- limp\_factor\_mixto2(pobgit2$A1\_04)  
pobgit2$A2\_04\_num <- limp\_factor\_mixto2(pobgit2$A2\_04)  
pobgit2$A3\_04\_num <- limp\_factor\_mixto2(pobgit2$A3\_04)  
pobgit2$A4\_04\_num <- limp\_factor\_mixto2(pobgit2$A4\_04)  
pobgit2$A5\_04\_num <- limp\_factor\_mixto2(pobgit2$A5\_04)  
pobgit2$A6\_04\_num <- limp\_factor\_mixto2(pobgit2$A6\_04)  
pobgit2$A7\_04\_num <- limp\_factor\_mixto2(pobgit2$A7\_04)  
pobgit2$A8\_04\_num <- limp\_factor\_mixto2(pobgit2$A8\_04)  
pobgit2$A9\_04\_num <- limp\_factor\_mixto2(pobgit2$A9\_04)  
pobgit2$A10\_04\_num <- limp\_factor\_mixto2(pobgit2$A10\_04)  
  
str(pobgit2$A1\_04\_num) #Comprobamos una de las variables numéricas creadas

## num [1:1492] 42 26 45 56 44 50 21 56 NA 39 ...

## Cálculo de la renta equivalente por unidad de consumo

Con el objetivo de obtener datos comparables, nuestro objetivo será obtener datos a nivel de población. Esta tarea presenta cierta dificultad, en tanto la base de datos utiliza la unidad familiar como unidad de análisis. Para facilitar la conversión y comparabilidad de los datos hemos optado por calcular el ingreso equivalente de todos los miembros de la unidad familiar.

El primer paso para realizar este cálculo será calcular los ingresos de la unidad familiar utilizando como referencia la escala de equivalencia “OCDE Modificada”. Esta escala de equivalencia nos permite asignar un valor, de forma ponderada, a cada miembro de la unidad familiar en función de la edad del mismo. De esta manera, al calcular los ingresos del hogar pueden tenerse en cuenta aspectos como las economías de escala en el consumo y el crecimiento no lineal de necesidades de consumo en unidades familiares de distintas dimensiones y características.

Así, el primer paso será calcular el taño equivalente del hogar a partir de las variables de A1\_04 a A7\_04, utilizando la siguiente escala de valores: - Primer adulto = 1 - Segundo adulto y siguientes = 0.5 - Niños y niñas menores de 14 años = 0.3

#Creamos la función para calcular el peso equivalente en la escala OECD de cada variable   
  
peso\_variable\_oecd <- function(x){  
 y <- ifelse(x >= 16, "0.5", ifelse(x < 16, "0.3", "0"))  
 y <- as.numeric(y)  
}  
  
  
#Calculaamos el peso de cada miembro de la unidad familiar en cada caso   
pobgit2$A1\_04\_oecd <- peso\_variable\_oecd(pobgit2$A1\_04\_num)  
pobgit2$A2\_04\_oecd <- peso\_variable\_oecd(pobgit2$A2\_04\_num)  
pobgit2$A3\_04\_oecd <- peso\_variable\_oecd(pobgit2$A3\_04\_num)  
pobgit2$A4\_04\_oecd <- peso\_variable\_oecd(pobgit2$A4\_04\_num)  
pobgit2$A5\_04\_oecd <- peso\_variable\_oecd(pobgit2$A5\_04\_num)  
pobgit2$A6\_04\_oecd <- peso\_variable\_oecd(pobgit2$A6\_04\_num)  
pobgit2$A7\_04\_oecd <- peso\_variable\_oecd(pobgit2$A7\_04\_num)  
pobgit2$A8\_04\_oecd <- peso\_variable\_oecd(pobgit2$A8\_04\_num)  
pobgit2$A9\_04\_oecd <- peso\_variable\_oecd(pobgit2$A9\_04\_num)  
pobgit2$A10\_04\_oecd <- peso\_variable\_oecd(pobgit2$A10\_04\_num)  
  
#Sumamos los pesos de cada hogar (fila/observación)  
pobgit2 <- pobgit2 %>% mutate(hogar\_oecd = select(., A1\_04\_oecd:A10\_04\_oecd) %>% replace(is.na(.), 0) %>% rowSums())  
pobgit2$hogar\_oecd <- pobgit2$hogar\_oecd + 1  
  
#Dividimos los ingresos totales del hogar por el tamaño equivalente del hogar  
pobgit2$ing\_equivalente <- pobgit2$ing\_hogar\_total/pobgit2$hogar\_oecd  
  
summary(pobgit2$ing\_equivalente) # A través de la nueva variable 'ing\_equivalente' obtenemos los ingresos equivalentes de cada hogar

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0 200.0 332.8 382.1 506.9 2280.0

Finalmente, crearemos una última variable relativa a ingresos en qué obtendremos los quintiles de la variable de ingresos equivalentes. Esta nueva variable nos será útil para comparar datos entre la muestra de población perteneciente al pueblo gitano y los datos disponibles de población general.

# Calcular quintiles de ingresos equivalentes   
pobgit2$quintiles\_ingresos <- quantcut(pobgit2$ing\_equivalente, q=5, na.rm=TRUE)  
levels(pobgit2$quintiles\_ingresos)

## [1] "[0,167]" "(167,281]" "(281,400]" "(400,560]"   
## [5] "(560,2.28e+03]"

pobgit2$quintiles\_ingresos <- revalue(pobgit2$quintiles\_ingresos, c("[0,167]"= "Quintil 1", "(167,281]" = "Quintil 2", "(281,400]" = "Quintil 3", "(400,560]" = "Quintil 4", "(560,2.28e+03]" = "Quintil 5"))  
  
summary(pobgit2$quintiles\_ingresos)

## Quintil 1 Quintil 2 Quintil 3 Quintil 4 Quintil 5   
## 301 296 332 264 299

# Herramientas metodológicas: funciones propias y preparación de código por defecto

En este apartado se definen las herramientas de estadística descriptiva básica que nos servirán para el análisis de datos posterior. Para ellos hemos creado funciones propias para la creación y construcción de tablas de frecuencia, tablas de contingencia y visualización gráfica de datos.

## Construcción de tablas de frecuencia para una variable (descriptiva univariante)

En primer lugar, definimos la función que se utilizará para la creación de tablas de frecuencia de los indicadores de pobreza energética (tablas de frecuencia para una sola variable) y que se utilizarán a lo largo de la tesis:

#Función para crear tablas de indicadores (una variable)  
  
tabla\_indicador <- function(x,t){  
   
 # Creamos tabla y convertimos a objeto flextable   
 z <- tabyl(x, show\_na = FALSE)   
 z <- qflextable(z)  
   
 # Definimos las propiedades generales de la tabla   
 set\_table\_properties(z, width = 1, layout = "autofit")  
   
 # Definimos y creamos objetos de texto con formato   
 # Formato general  
 def\_par <- fp\_par(text.align = "center", padding = 5)   
 # Formato texto   
 def\_text <- fp\_text(font.size = 9,   
 italic = FALSE,   
 font.family = "Arial")  
 #Formato de la cabecera   
 def\_text\_header <- fp\_text(font.size = 10,   
 italic = FALSE,   
 font.family = "Arial",  
 color="dodgerblue4",  
 bold = TRUE)  
 # Creamos el título   
 z <- add\_header\_lines(x = z,   
 values = t)  
  
 # Aplicamos los estilos de texto diseñados anteriormente   
 z <- style(x = z, pr\_p = def\_par, pr\_t = def\_text, part = "all")   
 z <- style(x = z, pr\_t = def\_text\_header, part = "header")  
 z <- set\_header\_labels(x = z,   
 n = "Total",   
 percent = "Porcentaje")  
 z <- fontsize(z, i=2, size = 9, part = "header")  
 z <- bold(z, j =1, bold = TRUE, part = "body")  
 z <- color(z, j = 1, color= "dodgerblue4", part = "body")  
   
 # Bordes   
 z <- border\_remove(z) # Primero eliminamos los bordes  
  
 big\_border = fp\_border(color ="dodgerblue4",   
 style = "dotted",   
 width = 2)  
 std\_border = fp\_border(color="dodgerblue4", # color  
 style = "solid", # estilo   
 width = 1) #anchura borde  
 z <- hline(z, i =1, border = big\_border, part = "header")  
 z <- hline(z, i =2, border = std\_border, part = "header")  
 z <- hline\_bottom(z, part="body", border = std\_border )  
   
 # Imprimimos la tabla  
 z  
}  
  
#tabla\_indicador(x = pobgit\_temp$temp\_adecuada, #variable con marcador $   
# t = "Temperatura adecuada") # Título de la tabla (nombre del indicador)

## Construcción de tablas de contingencia para dos o más variables (descriptiva multivariante)

A continuación se define la función utilizada para la creación de tablas de contingencia presentadas en esta tesis doctoral. Crearemos una función propia para la creación de tablas de contingencia con los datos que iremos obteniendo. La creación de esta función facilita y simplifica la ejecución del código posterior.

crear\_tabla2 <- function(var1,var2,title,varname,numcol){  
 # Creamos la tabla de frecuencias   
 z <- proc\_freq(pobgit2,  
 row = var1, # Variable 1  
 col = var2, # Variable 2  
 include.row\_percent = TRUE,  
 include.column\_percent = TRUE,  
 include.table\_percent = FALSE,  
 include.column\_total = FALSE,  
 include.row\_total = FALSE,  
 include.header\_row = FALSE)   
   
 # Empezamos a definir las características de formato de la tabla   
 set\_table\_properties(z, width = 1, layout = "autofit")  
   
 # Creamos objetos específicos de definición de formato de texto   
 def\_par <- fp\_par(text.align = "center", padding = 5)  
 def\_text <- fp\_text(font.size = 8, italic = FALSE, font.family = "Arial")  
 def\_text\_header <- fp\_text(font.size = 8, italic = FALSE, bold = FALSE, font.family = "Arial", color="dodgerblue4", underlined = FALSE)  
  
 # Establecemos las cabeceras de la tabla  
 z <- set\_header\_labels(x = z,   
 label = "")  
 z <- add\_header\_row( x = z,   
 values = c(" ","Cálculo", varname),  
 colwidths = c(1,1,numcol))  
 z <- add\_header\_lines(x = z,   
 values = title)  
   
 # Aplicamos los diseños da toda la tabla   
 z <- style(x = z,   
 pr\_p = def\_par,   
 pr\_t = def\_text,   
 part = "all")   
 # Aplicamos los diseños de texto a las cabeceras de la tabla de la tabla  
 z <- style(x = z,   
 pr\_t = def\_text\_header,   
 part = "header")  
 z <- fontsize(z, i = 2, size = 8, part = "header")  
 z <- bold(z, i = 2, part = "header")  
 z <- fontsize(z, i = 1, size = 10, part = "header")  
 z <- bold(z, i = 1, part = "header")  
 #z <- bg(z, i = 2,bg = "aliceblue", part = "header")  
 z <- bg(z, i = 3,bg = "aliceblue", part = "header")  
   
 # Modificamos características específicas del cuerpo de la tabla   
 z <- bg(z, j = 1,bg = "dodgerblue4", part = "body")  
 z <- color(z, j=1, color = "white", part = "body")  
 z <- bold(z, j = 1, part = "body")  
 z <- italic(z, j = 2, part = "body")  
   
 #Definimos los bordes de la tabla   
 z <- border\_remove(z) # Primero eliminamos los bordes  
  
 big\_border = fp\_border(color ="dodgerblue4",   
 style = "dotted",   
 width = 2)  
 std\_border = fp\_border(color="dodgerblue4", # color  
 style = "solid", # estilo   
 width = 1) #anchura borde  
 z <- hline(z, i =1, border = big\_border, part = "header")  
 z <- hline(z, i =2, border = std\_border, part = "header")  
 z <- hline\_bottom(z, part="header", border = std\_border )  
 z <- hline(z, i =3, border = std\_border, part = "body")  
 z <- hline(z, i =6, border = std\_border, part = "body")  
 z <- hline\_bottom(z, part="body", border = std\_border )  
   
 # Refinamos los ajustes y los espaciados de la tabla y cabecera   
 z <- line\_spacing(z, i =2, space = 0.7, part = "header")  
 z <- padding(z, i =3, padding = 2, part = "header")  
  
 # Imprimimos la tabla   
 z  
}  
  
  
  
  
#crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 # var2 ="tamano\_hogarf", # variable independiente  
 # title = "Temperatura inadecuada en el hogar y tamaño del hogar", # título de la tabla   
 # varname = "Tamaño del hogar", # Variable independiente  
 # numcol = 5) # Número de valores de la variable independiente

## Graficación de tablas de contingencia para dos o más variables (descriptiva multivariante)

También definiremos una función propia para el diseño de gráficos de barra con el objetivo de representar las tablas de contingencia más relevantes que creemos a lo largo de esta sección cuantitativa de la tesis:

crear\_grafico\_temp <- function(x,y,title,xtitle){  
 df <- pobgit2 %>%   
 drop\_na({{x}}) %>%  
 drop\_na({{y}}) %>%  
 group\_by({{y}},{{x}}) %>%   
 tally() %>%   
 complete({{x}}, fill = list(n = 0)) %>%   
 mutate(percentage = n / sum(n) \* 100)  
   
 ggplot(df, aes({{x}}, percentage, fill = {{y}})) +   
 geom\_bar(stat = 'identity', position = 'dodge') +  
 scale\_y\_continuous("Porcentaje de población", expand = c(0,0))+   
 scale\_x\_discrete(xtitle)+   
 scale\_fill\_manual("Temperatura\nadecuada\nen el hogar", values = c("#15607a",   
 "#18a1cd",  
 "cyan")) +  
 ggtitle(title) +  
 theme\_classic(base\_size=10) +   
 theme(plot.title = element\_text(face = "bold",   
 size = 10,   
 color = "#00344c",   
 vjust = 3,   
 hjust = 0.5),  
 axis.text.x = element\_text(#angle = 30,   
 hjust = 0.5,   
 vjust = 2,   
 colour = "#00344c",   
 size = rel(0.9)),  
 axis.title.x = element\_text(face = "bold",   
 size = 9),  
 axis.title.y = element\_text(face = "bold",   
 size = 9,   
 vjust = 3),  
 axis.text.y = element\_text(size=8,   
 hjust = 0),  
 axis.line = element\_blank(),  
 axis.ticks.x = element\_blank(),   
 legend.title = element\_text(colour = "#00344c", size = 8, face = "bold"))   
   
}  
  
  
crear\_grafico\_ret <- function(x,y,title,xtitle){  
 df <- pobgit2 %>%   
 drop\_na({{x}}) %>%  
 drop\_na({{y}}) %>%  
 group\_by({{y}},{{x}}) %>%   
 tally() %>%   
 complete({{x}}, fill = list(n = 0)) %>%   
 mutate(percentage = n / sum(n) \* 100)  
   
 ggplot(df, aes({{x}}, percentage, fill = {{y}})) +   
 geom\_bar(stat = 'identity', position = 'dodge') +  
 scale\_y\_continuous("Porcentaje de población", expand = c(0,0))+   
 scale\_x\_discrete(xtitle)+   
 scale\_fill\_manual("Retraso en el\npago de facturas", values = c("#15607a",   
 "#18a1cd",  
 "cyan")) +  
 ggtitle(title) +  
 theme\_classic(base\_size=10) +   
 theme(plot.title = element\_text(face = "bold",   
 size = 10,   
 color = "#00344c",   
 vjust = 3,   
 hjust = 0.5),  
 axis.text.x = element\_text(#angle = 30,   
 hjust = 0.5,   
 vjust = 2,   
 colour = "#00344c",   
 size = rel(0.9)),  
 axis.title.x = element\_text(face = "bold",   
 size = 9),  
 axis.title.y = element\_text(face = "bold",   
 size = 9,   
 vjust = 3),  
 axis.text.y = element\_text(size=8,   
 hjust = 0),  
 axis.line = element\_blank(),  
 axis.ticks.x = element\_blank(),   
 legend.title = element\_text(colour = "#00344c", size = 8, face = "bold"))   
   
}  
  
#crear\_grafico\_temp(x = miembro\_mayor, #variable independiente  
 # y = temp\_adecuada, # variable dependiente   
 # title = "Pobreza energética y personas mayores", # título del gráfico  
 # xtitle = "Hogares con un miembro de más de 65 años") # título axis X

# Indicadores primarios de pobreza y vulnerabilidad energética

En este apartado calcularemos los principales indicadores primarios de pobreza y vulnerabilidad energética a partir de los datos disponibles. Cómo hemos mencionado anteriormente, los microdatos de la encuesta desarrollada por la Fundación Secretariado Gitano sólo nos permite construir dos de los cuatro indicadores oficiales de pobreza energética, de tipo consensual: la autopercepción de temperatura inadecuada en el hogar y el retraso en el pago de facturas de suministros básicos.

## Indicador consensual: temperatura adecuada en el hogar

El primer indicador analizado es el indicador consensual de **Temperatura adecuada en el hogar**. Este indicador se construye a partir de la variable temp\_adecuada que se identifica como P51\_3 en la encuesta, y que se extrae de la pregunta:

Dígame, por favor, si su hogar puede permitirse: (…) 51.3. Mantener la vivienda con una temperatura adecuada

A partir de dicha pregunta, obtenemos datos que nos permiten construir el indicador de temperatura adecuada en el hogar tanto a nivel de hogares como a nivel poblacional. En primer lugar, debemos tener en cuenta que la encuesta con la que trabajamos considera cada observación como un hogar. Es decir, la unidad de análisis de la encuesta es la unidad familiar. Con el objetivo de poder traducir esta información en datos poblaciones, que nos permitan su comparabilidad, hemos optado por el método utilizado por el Observatorio Europeo de Pobreza Energética (EPOV).

De acuerdo con la guía metodológica de EPOV, para calcular los datos poblacionales a partir de datos basados en hogares, se atribuirá el valor de la variable en cuestión a cada uno de los miembros del hogar, operación que nos permite calcular el indicador a nivel poblacional.

#Creamos una nueva variable que incluya el tamaño del hogar (miembros) en formato numérico  
pobgit2$P50\_num <- as.numeric(as.character(pobgit2$P50))   
  
  
summary(pobgit2$P50\_num)#Resumen numérico del tamaño de los hogares de la muestra

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1.000 3.000 4.000 4.032 5.000 15.000

#Eliminamos los valores perdidos de la variable de temperatura adecuada en el hogar   
levels(pobgit2$temp\_adecuada)

## [1] "Sí" "No" "NS/NC"

pobgit2 <- pobgit2 %>% replace\_with\_na(replace = list(temp\_adecuada = "NS/NC"))  
pobgit2$temp\_adecuada <- fct\_drop(pobgit2$temp\_adecuada)  
  
  
tabla\_indicador(x = pobgit2$temp\_adecuada,   
 t = "Temperatura adecuada (% hogares)")

| **Temperatura adecuada (% hogares)** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **Sí** | 1,036 | 0.6962366 |
| **No** | 452 | 0.3037634 |

A partir de esta primera operación, podemos ver que 452 hogares de la muestra no pueden permitirse mantener su hogar a una temperatura adecuada. En este punto, procederemos a aplicar el método propuesto por el EPOV para obtener los datos poblacionales equivalentes.

# Creamos una base de datos duplicada con el total de población   
pobgit\_pob <- pobgit2 %>% map\_df(., rep, .$P50\_num)  
  
#Creamos una tabla de frecuencia con la variable de temperatura adecuada a nivel poblacional   
tabla\_indicador(x = pobgit\_pob$temp\_adecuada, #variable con marcador $   
 t = "Temperatura adecuada (% población)") # Título de la tabla (nombre del indicador)

| **Temperatura adecuada (% población)** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **Sí** | 4,202 | 0.6999833 |
| **No** | 1,801 | 0.3000167 |

## Indicador consensual de retrasos en el pago de facturas de suministros

El segundo indicador, también de tipo consensual, será el de **Retrasos en el pago de facturas**. A partir de la variable retrasos que se identifica con la variable P51\_5 de la base de datos extraída de la encuesta:

51.5. ¿Ha habido retrasos en el pago de gastos relacionados con la vivienda principal (hipoteca o alquiler, recibos de gas, comunidad…) o en compras a plazos en los últimos 12 meses?

Tal como puede apreciarse, debe tenerse en cuenta que este indicador no sólo incluye retrasos en el pago de suministros básicos, sino también de gastos vinculados con la vivienda principal en general, incluyendo hipoteca, alquiler, gastos de la comunidad u otros. En consecuencia, deberemos ser cautas al analizar sus resultados y ponerlos en contexto al compararlos con otros indicadores.

De los resultados obtenidos para la población, extraemos los siguientes resultados:

#Eliminamos los valores perdidos de la variable de retrasos en el pago de gastos relacionados con la vivienda principal  
levels(pobgit2$retrasos)

## [1] "Sí" "No" "NS/NC"

#En este caso, vemos que hay 49 casos que han expresado NSNC ante esta pregunta, esto es el 3% de la muestra.   
prop.table(table(pobgit2$retrasos))\*100

##   
## Sí No NS/NC   
## 50.804290 45.911528 3.284182

tabla\_indicador(x = pobgit2$retrasos,   
 t = "Retrasos en el pago (% hogares)")

| **Retrasos en el pago (% hogares)** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **Sí** | 758 | 0.50804290 |
| **No** | 685 | 0.45911528 |
| **NS/NC** | 49 | 0.03284182 |

En este punto, se procede a convertir los valores NSNC en valores perdidos, no excluyendo las observaciones de la muestra, pero no tenienda en cuenta ese 3% de las observaciones únicamente en relación a este indicador.

# Substituímos los valores NS/NA por valores perdidos (NA)  
pobgit2 <- pobgit2 %>% replace\_with\_na(replace = list(retrasos = "NS/NC"))  
pobgit2$retrasos <- fct\_drop(pobgit2$retrasos) #Eliminamos la categoría NS/NC de la variable que ahora está vacía  
  
tabla\_indicador(x = pobgit2$retrasos,   
 t = "Retrasos en pagos (% hogares)")

| **Retrasos en pagos (% hogares)** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **Sí** | 758 | 0.5252945 |
| **No** | 685 | 0.4747055 |

A continuación, reproducimos los cálculos para obtener el indicador en términos poblacionales, a partir de la base de datos que hemos creado anteriormente.

# Eliminaremos los valores perdidos en esta base de datos poblacional   
pobgit\_pob <- pobgit\_pob %>% replace\_with\_na(replace = list(retrasos = "NS/NC"))  
pobgit\_pob$retrasos <- fct\_drop(pobgit\_pob$retrasos)  
  
#Creamos una tabla de frecuencia con la variable de temperatura adecuada a nivel poblacional   
tabla\_indicador(x = pobgit\_pob$retrasos, #variable con marcador $   
 t = "Retrasos en pagos (% población)") # Título de la tabla (nombre del indicador)

| **Retrasos en pagos (% población)** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **Sí** | 3,128 | 0.539962 |
| **No** | 2,665 | 0.460038 |

## Superposición de indicadores

A continuación, crearemos una nueva variable para el análisis de la coincidencia de afectación de los distintos indicadores de vulnerabilidad energética sobre esta misma población. En primer lugar creamos una variable que nos indique si un sujeto está afectado por al menos uno de los dos indicadores:

#Creamos una nueva variable llamada 'PE' que nos indicará si un individuo está afectado por al menos un indicador:   
  
pobgit\_pob <- pobgit\_pob %>% mutate(PE = case\_when(  
 retrasos == "Sí" ~ "PE",   
 temp\_adecuada == "No" ~ "PE",   
 TRUE ~ "NO PE"  
))  
  
pobgit\_pob$PE <- as.factor(pobgit\_pob$PE) #Transformamos a factor para su interpretación   
  
ind\_PE <- 100\*prop.table(table(pobgit\_pob$PE))  
pandoc.table(ind\_PE)

##   
## ---------------  
## NO PE PE   
## ------- -------  
## 39.67 60.33   
## ---------------

tabla\_indicador(x = pobgit\_pob$PE,   
 t = "Población gitana afectada por alguno de los indicadores PE")

| **Población gitana afectada por alguno de los indicadores PE** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **NO PE** | 2,386 | 0.396675 |
| **PE** | 3,629 | 0.603325 |

A partir de los resultados podemos ver que un 60,3% de la muestra está afectada por alguno de los indicadores consensuales de pobreza energética, con la debida mención que el indicador sobre retrasos en los pagos es más amplio que el indicador incluído en la Estrategia Nacional contra la Pobreza Energética.

A continuación nos interesará evaluar qué proporción de la población de la muestra está afectada por dos o más indicadores:

# Recodificamos las variables afectadas para convertirlas en variables 'dummy'   
df <- dummy\_cols(pobgit\_pob,   
 select\_columns = c("retrasos", "temp\_adecuada"),   
 ignore\_na = TRUE)   
  
df <- df %>% mutate(PE\_Suma = retrasos\_Sí + temp\_adecuada\_No) #Creamos una variable para saber por cuantos de los indicadores una persona está afectada   
  
df$PE\_Suma <- as.factor(df$PE\_Suma) # Convertimos la nueva variable a factor para su mejor interpretación   
  
ind\_PESum <- 100\*prop.table(table(df$PE\_Suma)) # Construimos el nuevo indicador de superposición de indicadores  
pandoc.table(ind\_PESum) #Ejecutamos el nuevo indicador

##   
## ----------------------  
## 0 1 2   
## ------ ------- -------  
## 38.4 39.15 22.45   
## ----------------------

También podemos obtener estos datos en una tabla:

tabla\_indicador(x = df$PE\_Suma,   
 t = "Población afectada por 0, 1 y 2 indicadores")

| **Población afectada por 0, 1 y 2 indicadores** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **0** | 2,224 | 0.3840442 |
| **1** | 2,267 | 0.3914695 |
| **2** | 1,300 | 0.2244863 |

## Indicador de pobreza y vulnerabilidad energética

En primer lugar, calcularemos la redundancia entre las dos variables seleccionadas:

Este método se basa en la construcción de un indicador multidimensional, a partir de la previa selección de indicadores, teniendo en cuenta el cálculo de redundancia entre las variables seleccionadas.

# CREAMOS UNA FUNCIÓN PARA CALCULAR LA REDUNDANCIA   
  
redundancia2 <- function(var1,var2){  
 var1 <- enquo(var1)  
 var2 <- enquo(var2)  
   
 #Total población  
 total <- nrow(pobgit\_pob)  
   
 #Porcentaje de población afectada por el indicador 1  
 ind\_1\_p <- nrow(pobgit\_pob %>% filter(!!var1 == "Sí"))  
 ind\_1\_p <- (ind\_1\_p/total)\*100  
   
 #Porcentaje de población afectada por el indicador 2  
 ind\_2\_p <- nrow(pobgit\_pob %>% filter(!!var2 == "Sí"))  
 ind\_2\_p <- (ind\_2\_p/total)\*100  
   
 # Porcentaje de población afectada por ambos indicadores   
 ind\_mix <- nrow(pobgit\_pob %>% filter(!!var1 == "Sí" & !!var2 == "Sí"))  
 ind\_mix <- (ind\_mix/total)\*100  
   
 dos <- c(ind\_1\_p, ind\_2\_p)  
   
 #Cálculo de la redundancia  
 red <- ind\_mix/min(dos)  
 red  
}

A partir de esta función, calcularemos la redundancia entre las variables seleccionadas para el indicador multidimensional que construiremos:

#Redundancia entre temperatura inadecuada y retrasos - 0.56  
redundancia2(var1 = temp\_adecuada,   
 var2 = retrasos)  
  
````  
El resultado del cálculo de redundancia nos indica que las variables están relacionadas en un 58%, por lo que se trata de un porcentaje muy elevado. En este caso, dadas las dificultades de interpretación de un indicador combinado, en tanto la variable de retrasos en los pagos de gastos relacionados con la vivienda habitual es más amplio que el indicador de retrasos en el pago de suministros básicos, optaremos por una caracterización por separado de los datos disponibles.   
  
# Caracterización de los hogares en situación de precariedad energética   
Hasta el momento nos hemos centrado en la construcción de los indicadores oficiales de pobreza energética. Ahora bien, también nos interesará ir más allá de los indicadores y caracterizar esta población. Para ello, combinaremos las variables de caracterización propuestas por el gobierno español en su Estrategia Nacional contra la Pobreza Energética, con el objetivo de facilitar la comparación de datos.   
  
A diferencia de la encuesta EINSFOESSA, la base de datos de análisis de la FSG no dispone de la misma amplitud. Por este motivo, limitaremos la caracterización a las dimensiones en las cuales podemos obtener información a partir de las variables existentes o de nuevas variables que podamos crear.   
  
También debe mencionarse que, en el caso de la base de datos de la Fundación Secretariado Gitano, la caracterización se realizará en base a las características de los hogares ya que no se dispone de suficiente información para poder aplicar esta caracterización a toda la población de la muestra.   
  
A continuación especificamos las dimensiones en torno a las cuales hemos agrupado las variables de caracterización analizadas:   
1. Dimensión 1: Características familiares   
2. Dimensión 2: características económicas  
4. Dimensión 3: Características de la vivienda   
3. Dimensión 4: Características del entorno urbano   
5. Dimensión 5: Impactos sobre la salud y relaciones sociales   
6. Dimensión 6: Vulnerabilidad energética y grupos minoritarios   
  
  
## Dimensión 1: Características familiares   
  
### Vulnerabilidad Energética según tamaño del hogar  
  
Para analizar la distribución de la población en situación de vulnerabilidad y precariedad energética según el tamaño del hogar, primero necesitaremos crear una variable de tamaño del hogar que nos permite su comparabilidad:

## Error: attempt to use zero-length variable name

# Renombramos la variable que nos da informacion sobre el tamaño del hogar   
pobgit2 <- pobgit2 %>% rename(tamano\_hogar = P50)  
pobgit2 <- pobgit2 %>% rename(tamano\_hogar\_num = P50\_num)  
  
levels(pobgit2$tamano\_hogar)

## [1] "1" "10" "11" "12" "15" "2" "3" "4" "5" "6" "7" "8" "9" "NC"

#Creamos una variable funcional para el análisis  
  
pobgit2$tamano\_hogarf <- fct\_collapse(pobgit2$tamano\_hogar,   
 "1 miembro" = c("1"),   
 "2 miembros" = c("2"),   
 "3 miembros" = c("3"),   
 "4 miembros" = c("4"),   
 "5 miembros o más" = c("5", "6", "7", "8", "9", "10", "11", "12", "15"),  
 other\_level = NULL)  
  
#Eliminamos el factor level 'NC'   
pobgit2$tamano\_hogarf <- fct\_drop(pobgit2$tamano\_hogarf)  
  
levels(pobgit2$tamano\_hogarf) # Verificamos el número de niveles iniciales de la variable recodificada

## [1] "1 miembro" "5 miembros o más" "2 miembros" "3 miembros"   
## [5] "4 miembros"

#Rordenamos los niveles del factor   
pobgit2$tamano\_hogarf <- factor(pobgit2$tamano\_hogarf, levels = c("1 miembro","2 miembros","3 miembros","4 miembros","5 miembros o más"))

Una vez disponemos de una variable operativa sobre el tamaño del hogar, podremos crear tablas de contingencia para poner en relación las variables en juego.

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="tamano\_hogarf", # variable independiente  
 title = "Temperatura inadecuada en el hogar y tamaño del hogar", # título de la tabla   
 varname = "Tamaño del hogar", # Variable independiente  
 numcol = 5) # Número de valores de la variable independiente

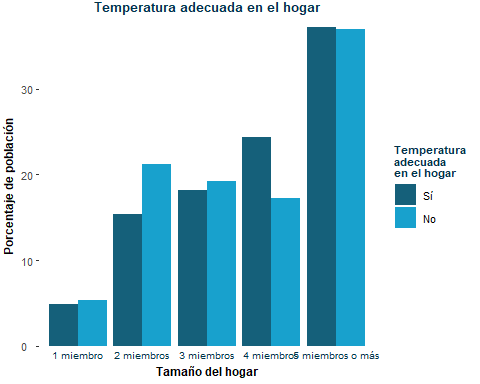
| **Temperatura inadecuada en el hogar y tamaño del hogar** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cálculo** | **Tamaño del hogar** | | | | |
| temp\_adecuada |  | 1 miembro | 2 miembros | 3 miembros | 4 miembros | 5 miembros o más |
| **Sí** | *Frequency* | 50 | 159 | 188 | 253 | 386 |
| *Row Pct* | 4.83% | 15.35% | 18.15% | 24.42% | 37.26% |
| *Col Pct* | 67.57% | 62.35% | 68.36% | 76.44% | 69.8% |
| **No** | *Frequency* | 24 | 96 | 87 | 78 | 167 |
| *Row Pct* | 5.31% | 21.24% | 19.25% | 17.26% | 36.95% |
| *Col Pct* | 32.43% | 37.65% | 31.64% | 23.56% | 30.2% |

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="tamano\_hogarf", # variable independiente  
 title = "Retrasos en el pago de facturas del hogar y tamaño del hogar", # título de la tabla   
 varname = "Tamaño del hogar", # Variable independiente  
 numcol = 5) # Número de valores de la variable independiente

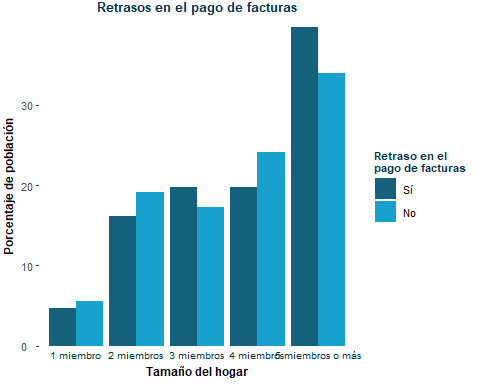
| **Retrasos en el pago de facturas del hogar y tamaño del hogar** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cálculo** | **Tamaño del hogar** | | | | |
| retrasos |  | 1 miembro | 2 miembros | 3 miembros | 4 miembros | 5 miembros o más |
| **Sí** | *Frequency* | 35 | 122 | 150 | 150 | 301 |
| *Row Pct* | 4.62% | 16.09% | 19.79% | 19.79% | 39.71% |
| *Col Pct* | 47.95% | 48.22% | 55.97% | 47.62% | 56.37% |
| **No** | *Frequency* | 38 | 131 | 118 | 165 | 233 |
| *Row Pct* | 5.55% | 19.12% | 17.23% | 24.09% | 34.01% |
| *Col Pct* | 52.05% | 51.78% | 44.03% | 52.38% | 43.63% |

Para finalizar, graficaremos la relación entre estos indicadores y la variable de tamaño del hogar en los siguientes gráficos:

crear\_grafico\_temp(x = tamano\_hogarf, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura adecuada en el hogar", # título del gráfico  
 xtitle = "Tamaño del hogar") # título axis X



crear\_grafico\_ret(x = tamano\_hogarf, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en el pago de facturas", # título del gráfico  
 xtitle = "Tamaño del hogar") # título axis X



### Hogares con (al menos) una persona de más de 65 años

# Recodificamos el nombre de la variable edad de la persona que contesta la encuesta  
pobgit2 <- pobgit2 %>% rename(edad = P1)  
  
#Convertimos a  
  
pobgit2 <- pobgit2 %>% mutate(miembro\_mayor = case\_when(  
 edad >= 65 ~ 'Sí',   
 A1\_04\_num >= 65 ~ 'Sí',  
 A2\_04\_num >= 65 ~ 'Sí',  
 A3\_04\_num >= 65 ~ 'Sí',  
 A4\_04\_num >= 65 ~ 'Sí',  
 A5\_04\_num >= 65 ~ 'Sí',  
 A6\_04\_num >= 65 ~ 'Sí',  
 A7\_04\_num >= 65 ~ 'Sí',  
 A8\_04\_num >= 65 ~ 'Sí',  
 A9\_04\_num >= 65 ~ 'Sí',  
 A10\_04\_num >= 65 ~ 'Sí',  
 TRUE ~ 'No'  
))  
  
str(pobgit2$miembro\_mayor)

## chr [1:1492] "No" "No" "No" "No" "No" "No" "No" "No" "Sí" "No" "No" "No" ...

pobgit2$miembro\_mayor <- as.factor(pobgit2$miembro\_mayor)  
  
fct\_count(pobgit2$miembro\_mayor) #263 hogares tienen un miembro de m?s de 65 a?os

## # A tibble: 2 x 2  
## f n  
## <fct> <int>  
## 1 No 1229  
## 2 Sí 263

Una vez hemos identificado los hogares en que, al menos, uno de los miembros es mayor de 65 años, crearemos tablas de contingencia para comparar esta variable con los indicadores de vulnerabilidad energética identificados. En primer lugar, crearemos una tabla de contingencia para el indicador de temperatura inadecuada:

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="miembro\_mayor", # variable independiente  
 title = "Temperatura inadecuada en el hogar y y unidades familiares con una persona mayor", # título de la tabla   
 varname = "Unidad familiar con un miembro de más de 65 años", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

| **Temperatura inadecuada en el hogar y y unidades familiares con una persona mayor** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidad familiar con un miembro de más de 65 años** | |
| temp\_adecuada |  | No | Sí |
| **Sí** | *Frequency* | 866 | 170 |
| *Row Pct* | 83.59% | 16.41% |
| *Col Pct* | 70.58% | 65.13% |
| **No** | *Frequency* | 361 | 91 |
| *Row Pct* | 79.87% | 20.13% |
| *Col Pct* | 29.42% | 34.87% |

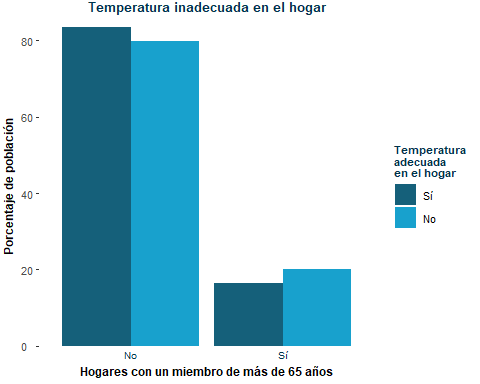
Ahora, creamos una tabla de contingencia para el indicador de retraso en el pago de facturas:

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="miembro\_mayor", # variable independiente  
 title = "Retrasos en el pago de facturas del hogar y unidades familiares con una persona mayor", # título de la tabla   
 varname = "Unidad familiar com un miembro de más de 65 años", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

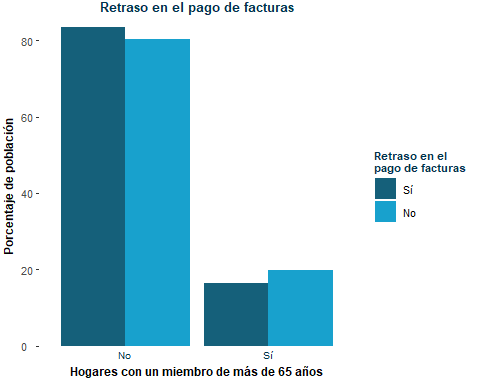
| **Retrasos en el pago de facturas del hogar y unidades familiares con una persona mayor** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidad familiar com un miembro de más de 65 años** | |
| retrasos |  | No | Sí |
| **Sí** | *Frequency* | 633 | 125 |
| *Row Pct* | 83.51% | 16.49% |
| *Col Pct* | 53.51% | 48.08% |
| **No** | *Frequency* | 550 | 135 |
| *Row Pct* | 80.29% | 19.71% |
| *Col Pct* | 46.49% | 51.92% |

Una vez tenemos las tablas de datos, crearemos los gráficos que no ayudarán a visualizar la distribución de datos:

crear\_grafico\_temp(x = miembro\_mayor, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura inadecuada en el hogar", # título del gráfico  
 xtitle = "Hogares con un miembro de más de 65 años") # título axis X



crear\_grafico\_ret(x = miembro\_mayor, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retraso en el pago de facturas", # título del gráfico  
 xtitle = "Hogares con un miembro de más de 65 años") # título axis X



### Hogares con dos adultos (sin niños) y uno de ellos de más de 65 años

Para el cálculo de estos indicadores, primero crearemos una nueva variable de referencia para identificar esta tipología específica de hogares:

#Creamos una nueva variable 'pareja\_mayor' para identificar las unidades familiares con esta estructura familiar   
pobgit2 <- mutate(pobgit2, pareja\_mayor = case\_when(  
 edad >= 65 & A1\_04\_num >= 18 & tamano\_hogarf == "2 miembros" ~ "Sí",   
 A1\_04\_num >= 65 & edad >= 18 & tamano\_hogarf == "2 miembros" ~ "Sí",  
 TRUE ~ "No"))  
  
pobgit2$pareja\_mayor <- as.factor(pobgit2$pareja\_mayor)  
  
summary(pobgit2$pareja\_mayor)

## No Sí   
## 1411 81

#Total poblacián de adultos con almenos uno de ellos de m?s de 65 años  
pobgit2 %>% filter(pareja\_mayor == "Sí")%>% summarise (sum(tamano\_hogar\_num)) #162 personas

## # A tibble: 1 x 1  
## `sum(tamano\_hogar\_num)`  
## <dbl>  
## 1 162

Una vez hemos identificado los hogares compuestos por dos miembros en que, al menos, uno de los miembros es mayor de 65 años, crearemos tablas de contingencia para comparar esta variable con los indicadores de vulnerabilidad energética identificados. En primer lugar, crearemos una tabla de contingencia para el indicador de temperatura inadecuada:

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="pareja\_mayor", # variable independiente  
 title = "Temperatura inadecuada en el hogar y parejas mayores", # título de la tabla   
 varname = "Unidades familiares de parejas con, al menos, un miembro mayor de 65 años", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

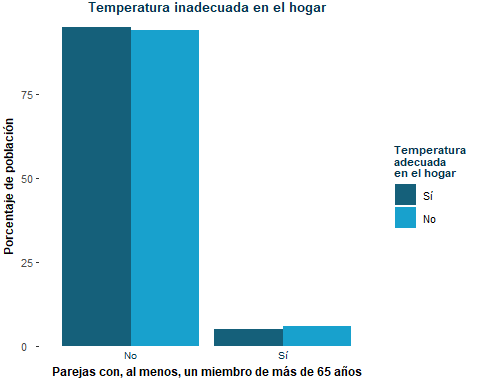
| **Temperatura inadecuada en el hogar y parejas mayores** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidades familiares de parejas con, al menos, un miembro mayor de 65 años** | |
| temp\_adecuada |  | No | Sí |
| **Sí** | *Frequency* | 984 | 52 |
| *Row Pct* | 94.98% | 5.02% |
| *Col Pct* | 69.84% | 65.82% |
| **No** | *Frequency* | 425 | 27 |
| *Row Pct* | 94.03% | 5.97% |
| *Col Pct* | 30.16% | 34.18% |

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="pareja\_mayor", # variable independiente  
 title = "Retrasos en el pago de facturas del hogar y parejas mayores", # título de la tabla   
 varname = "Unidades familiares de parejas con, al menos, un miembro mayor de 65 años", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

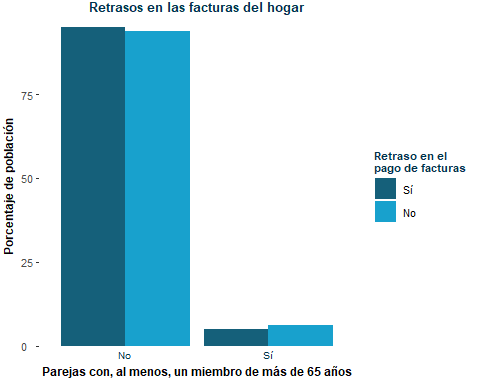
| **Retrasos en el pago de facturas del hogar y parejas mayores** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidades familiares de parejas con, al menos, un miembro mayor de 65 años** | |
| retrasos |  | No | Sí |
| **Sí** | *Frequency* | 721 | 37 |
| *Row Pct* | 95.12% | 4.88% |
| *Col Pct* | 52.9% | 46.25% |
| **No** | *Frequency* | 642 | 43 |
| *Row Pct* | 93.72% | 6.28% |
| *Col Pct* | 47.1% | 53.75% |

Una vez tenemos las tablas de datos, crearemos los gráficos que no ayudarán a visualizar la distribución de datos:

crear\_grafico\_temp(x = pareja\_mayor, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura inadecuada en el hogar", # título del gráfico  
 xtitle = "Parejas con, al menos, un miembro de más de 65 años") # título axis X



crear\_grafico\_ret(x = pareja\_mayor, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en las facturas del hogar", # título del gráfico  
 xtitle = "Parejas con, al menos, un miembro de más de 65 años") # título axis X

 ### Hogares Monoparentales y Monomarentales

En primer lugar, en relación a los hogares monomarentales y monoparentales, lo que haremos será distinguir entre dos variables distintas: \* Hogares monoparentales, a partir de la variable monoparental para hogares con niños y un solo adulto. En este caso, se consideran tanto los hogares en que el único adulto es un hombre o una mujer. \* Hogares monomarentales, en que el único adulto es una mujer

# Hogares monoparentales  
  
pobgit2 <- pobgit2 %>% mutate(monoparental = case\_when(  
 tamano\_hogar\_num == 1 ~ 'No',  
 A1\_04\_num > 18 ~ 'No',  
 A2\_04\_num > 18 ~ 'No',  
 A3\_04\_num > 18 ~ 'No',  
 A4\_04\_num > 18 ~ 'No',  
 A5\_04\_num > 18 ~ 'No',  
 A6\_04\_num > 18 ~ 'No',  
 A7\_04\_num > 18 ~ 'No',  
 A8\_04\_num > 18 ~ 'No',  
 A9\_04\_num > 18 ~ 'No',  
 A10\_04\_num > 18 ~ 'No',  
 TRUE ~ 'Sí'  
))  
  
pobgit2$monoparental <- as.factor(pobgit2$monoparental)  
summary(pobgit2$monoparental)

## No Sí   
## 1422 70

pobgit2 %>% filter(monoparental == "Sí") %>% summarise (sum(tamano\_hogar\_num)) #población de 261

## # A tibble: 1 x 1  
## `sum(tamano\_hogar\_num)`  
## <dbl>  
## 1 261

# Hogares monomarentales   
  
pobgit2 <- pobgit2 %>% mutate(monomarental = case\_when(  
 tamano\_hogar\_num == 1 ~ 'No',  
 A1\_04\_num > 18 ~ 'No',  
 A2\_04\_num > 18 ~ 'No',  
 A3\_04\_num > 18 ~ 'No',  
 A4\_04\_num > 18 ~ 'No',  
 A5\_04\_num > 18 ~ 'No',  
 A6\_04\_num > 18 ~ 'No',  
 A7\_04\_num > 18 ~ 'No',  
 A8\_04\_num > 18 ~ 'No',  
 A9\_04\_num > 18 ~ 'No',  
 A10\_04\_num > 18 ~ 'No',  
 P2 == "Hombre" ~ 'No',  
 TRUE ~ 'Sí'  
))  
  
pobgit2$monomarental <- as.factor(pobgit2$monomarental)  
summary(pobgit2$monomarental)

## No Sí   
## 1447 45

pobgit2 %>% filter(monomarental == "Sí") %>% summarise (sum(tamano\_hogar\_num)) #población de 154

## # A tibble: 1 x 1  
## `sum(tamano\_hogar\_num)`  
## <dbl>  
## 1 154

Una vez tenemos creadas las dos variables, podremos crear tablas de contingencia entre las variables.

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="monoparental", # variable independiente  
 title = "Temperatura inadecuada en el hogar y familias monoparentales", # título de la tabla   
 varname = "Unidades familiares monoparentales (ambos sexos)", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

| **Temperatura inadecuada en el hogar y familias monoparentales** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidades familiares monoparentales (ambos sexos)** | |
| temp\_adecuada |  | No | Sí |
| **Sí** | *Frequency* | 1000 | 36 |
| *Row Pct* | 96.53% | 3.47% |
| *Col Pct* | 70.52% | 51.43% |
| **No** | *Frequency* | 418 | 34 |
| *Row Pct* | 92.48% | 7.52% |
| *Col Pct* | 29.48% | 48.57% |

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="monomarental", # variable independiente  
 title = "Temperatura inadecuada en el hogar y familias monomarentales", # título de la tabla   
 varname = "Unidades familiares monomarentales", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

| **Temperatura inadecuada en el hogar y familias monomarentales** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidades familiares monomarentales** | |
| temp\_adecuada |  | No | Sí |
| **Sí** | *Frequency* | 1013 | 23 |
| *Row Pct* | 97.78% | 2.22% |
| *Col Pct* | 70.2% | 51.11% |
| **No** | *Frequency* | 430 | 22 |
| *Row Pct* | 95.13% | 4.87% |
| *Col Pct* | 29.8% | 48.89% |

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="monoparental", # variable independiente  
 title = "Retrasos en las facturas del hogar y familias monoparentales", # título de la tabla   
 varname = "Unidades familiares monoparentales (ambos sexos)", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

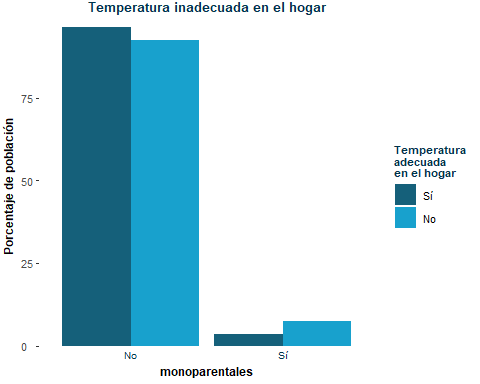
| **Retrasos en las facturas del hogar y familias monoparentales** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidades familiares monoparentales (ambos sexos)** | |
| retrasos |  | No | Sí |
| **Sí** | *Frequency* | 716 | 42 |
| *Row Pct* | 94.46% | 5.54% |
| *Col Pct* | 52.07% | 61.76% |
| **No** | *Frequency* | 659 | 26 |
| *Row Pct* | 96.2% | 3.8% |
| *Col Pct* | 47.93% | 38.24% |

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="monomarental", # variable independiente  
 title = "Retrasos en las facturas del hogar y familias monomarentales", # título de la tabla   
 varname = "Unidades familiares monomarentales", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

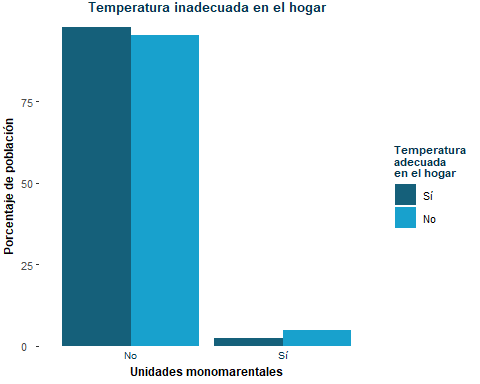
| **Retrasos en las facturas del hogar y familias monomarentales** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Unidades familiares monomarentales** | |
| retrasos |  | No | Sí |
| **Sí** | *Frequency* | 729 | 29 |
| *Row Pct* | 96.17% | 3.83% |
| *Col Pct* | 52.15% | 64.44% |
| **No** | *Frequency* | 669 | 16 |
| *Row Pct* | 97.66% | 2.34% |
| *Col Pct* | 47.85% | 35.56% |

Una vez disponemos de las tablas de contingencia, graficaremos las relaciones entre las variables en juego.

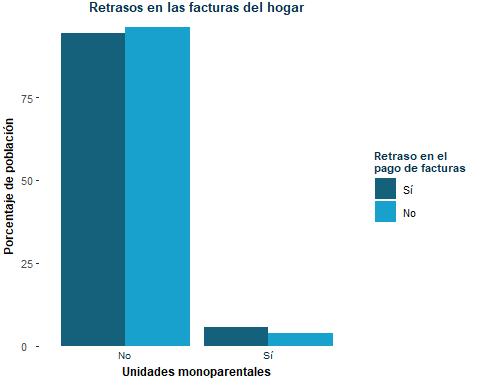
crear\_grafico\_temp(x = monoparental, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura inadecuada en el hogar", # título del gráfico  
 xtitle = "monoparentales") # título axis X



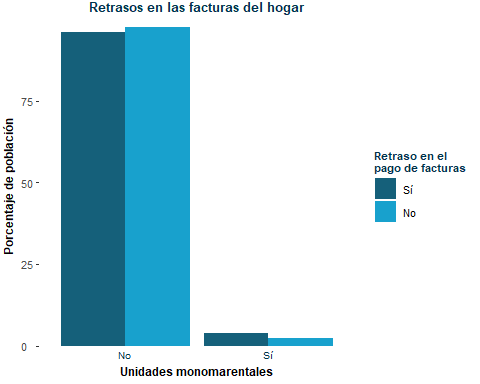
crear\_grafico\_temp(x = monomarental, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura inadecuada en el hogar", # título del gráfico  
 xtitle = "Unidades monomarentales") # título axis X



crear\_grafico\_ret(x = monoparental, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en las facturas del hogar", # título del gráfico  
 xtitle = "Unidades monoparentales") # título axis X



crear\_grafico\_ret(x = monomarental, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en las facturas del hogar", # título del gráfico  
 xtitle = "Unidades monomarentales") # título axis X



## Dimensión 2: Características socioeconómicas e indicadores relativos a ingresos familiares

### Pobreza Energética según quintil de renda

En el apartado de preparación de datos, hemos creado una variable específica sobre los ingresos equivalentes del hogar divididos en quintiles. En esta sección relacionaremos esta variable con los principales indicadores de pobreza y vulnerabilidad energética.

summary(pobgit2$quintiles\_ingresos)

## Quintil 1 Quintil 2 Quintil 3 Quintil 4 Quintil 5   
## 301 296 332 264 299

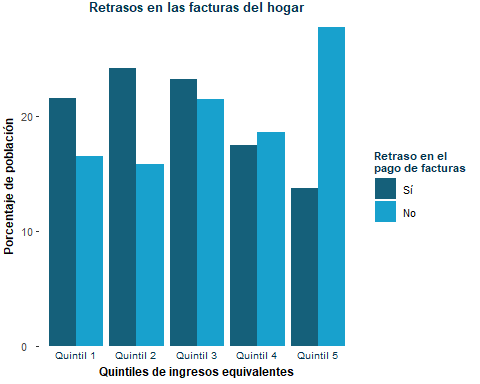
crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="quintiles\_ingresos", # variable independiente  
 title = "Retrasos en las facturas del hogar según quintiles de ingresos", # título de la tabla   
 varname = "Quintiles de ingresos equivalentes", # Variable independiente  
 numcol = 5) # Número de valores de la variable independiente

| **Retrasos en las facturas del hogar según quintiles de ingresos** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cálculo** | **Quintiles de ingresos equivalentes** | | | | |
| retrasos |  | Quintil 1 | Quintil 2 | Quintil 3 | Quintil 4 | Quintil 5 |
| **Sí** | *Frequency* | 163 | 183 | 176 | 132 | 104 |
| *Row Pct* | 21.5% | 24.14% | 23.22% | 17.41% | 13.72% |
| *Col Pct* | 59.06% | 62.89% | 54.49% | 50.97% | 35.37% |
| **No** | *Frequency* | 113 | 108 | 147 | 127 | 190 |
| *Row Pct* | 16.5% | 15.77% | 21.46% | 18.54% | 27.74% |
| *Col Pct* | 40.94% | 37.11% | 45.51% | 49.03% | 64.63% |

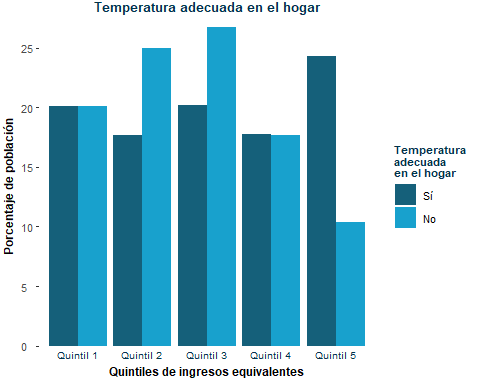
crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="quintiles\_ingresos", # variable independiente  
 title = "Temperatura adecuada en el hogar según quintiles de ingresos", # título de la tabla   
 varname = "Quintiles de ingresos equivalentes", # Variable independiente  
 numcol = 5) # Número de valores de la variable independiente

| **Temperatura adecuada en el hogar según quintiles de ingresos** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cálculo** | **Quintiles de ingresos equivalentes** | | | | |
| temp\_adecuada |  | Quintil 1 | Quintil 2 | Quintil 3 | Quintil 4 | Quintil 5 |
| **Sí** | *Frequency* | 208 | 183 | 209 | 184 | 252 |
| *Row Pct* | 20.08% | 17.66% | 20.17% | 17.76% | 24.32% |
| *Col Pct* | 69.57% | 61.82% | 63.33% | 69.7% | 84.28% |
| **No** | *Frequency* | 91 | 113 | 121 | 80 | 47 |
| *Row Pct* | 20.13% | 25% | 26.77% | 17.7% | 10.4% |
| *Col Pct* | 30.43% | 38.18% | 36.67% | 30.3% | 15.72% |

crear\_grafico\_ret(x = quintiles\_ingresos, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en las facturas del hogar", # título del gráfico  
 xtitle = "Quintiles de ingresos equivalentes") # título axis X



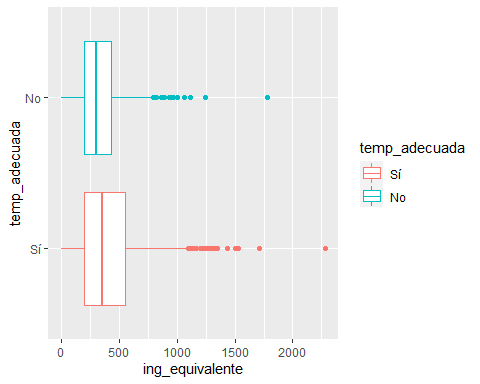
crear\_grafico\_temp(x = quintiles\_ingresos, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura adecuada en el hogar", # título del gráfico  
 xtitle = "Quintiles de ingresos equivalentes") # título axis X



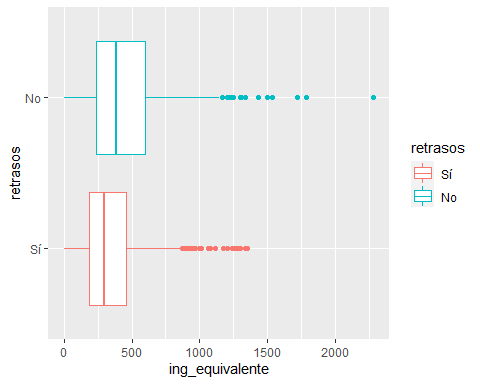
### Pobreza energética y nivel de ingresos equivalentes

A continuación, veremos gráficos representativos entre la relación entre temperatura adecuada en el hogar y los ingresos equivalentes de la unidad familiar.

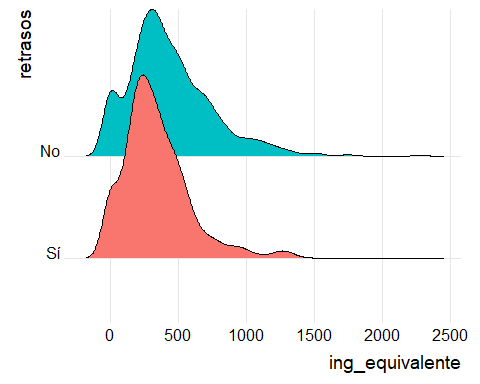
g1 <- pobgit2 %>%   
 filter(!is.na(temp\_adecuada)) %>%  
 ggplot(aes(ing\_equivalente, temp\_adecuada, color = temp\_adecuada)) + geom\_boxplot()  
g1



g2 <- pobgit2 %>%   
 filter(!is.na(retrasos)) %>%  
 ggplot(aes(ing\_equivalente, retrasos, color = retrasos)) + geom\_boxplot()  
g2



#install.packages("ggridges")  
library(ggridges)  
  
g2 <- pobgit2 %>%   
 filter(!is.na(retrasos)) %>%  
ggplot(aes(x = ing\_equivalente, y = retrasos, fill = retrasos)) +  
 geom\_density\_ridges() +  
 theme\_ridges() +   
 theme(legend.position = "none")  
  
g2



### Pobreza energética en función de la situación de ocupación de la persona de referencia del hogar

Un segundo indicador relevante en relación con la dimensión económica es el impacto de pobreza energética en función de la situación de ocupación de la persona de referencia del hogar. Para analizar esta relación, usaremos la variable ppalocupacion que crearemos a partir de la variable P5.

str(pobgit2$P5) # Nos indica quién es la persona de referencia

## Factor w/ 15 levels "Es usted la persona de referencia",..: 1 1 1 2 1 3 2 2 1 2 ...

#En caso que la persona de referencia sea quien responde -- P13  
  
summary(pobgit2$P13) # Actividad principal si quien contesta la encuesta es la persona de referencia

## Trabajador/a por cuenta ajena (Contrato fijo)   
## 56   
## Trabajador/a por cuenta ajena (Contrato temporal)   
## 117   
## Trabajador/a por cuenta ajena sin contrato   
## 35   
## Trabajador/a por cuenta propia   
## 148   
## Trabajador/a en un negocio familiar   
## 42   
## Baja por enfermedad o maternidad/paternidad   
## 7   
## Parado/a   
## 439   
## Jubilado o retirado del trabajo   
## 147   
## Incapacitado/a permanente   
## 55   
## Estudiando/realizando cursos de formación   
## 48   
## Trabajo doméstico no remunerado   
## 382   
## Realizando sin remuneración trabajos sociales, actividades benéficas   
## 3   
## Otra situación   
## 10   
## NS/NC   
## 3

#En caso que la persona de referencia sea otra  
# A2 - preguntas sobre la relación con la persona entrevistada   
# A6 - actividad principal de la persona en cuestión   
  
pobgit2$P5\_car <- as.character(pobgit2$P5)  
pobgit2$P13\_car <- as.character(pobgit2$P13)  
  
  
pobgit2 <- pobgit2 %>% mutate(actp\_yo = case\_when(  
 P5 == "Es usted la persona de referencia" ~ "Principal",   
 TRUE ~ "No Principal"  
))  
  
# Cambiaremos las etiquetas de valores en P5 para que coincidan con las existentes en las variables de familiares   
  
pobgit2$P5 <- revalue(pobgit2$P5, c("Es usted la persona de referencia" = "Es usted la persona de referencia",   
 "Es su hijo/a" = "Hijo/a, hijastro/a",   
 "Es su nieto/a" = "Nieto/a, nieto político/a (o pareja de los mismos)",   
 "Es otro pariente" = "Otro pariente",   
 "Es su cónyuge/pareja" = "Cónyuge o pareja",   
 "Es su suegro/a" = "Suegro/a o pareja de los mismos",   
 "Es su abuelo/a" = "Abuelo/a",   
 "Otros no parientes" = "Sin parentesco con usted",   
 "Es su padre/madre" = "Padre/Madre",   
 "Es su cuñado/a" = "Cuñado/a",   
 "Es su sobrino/a" = "Sobrino/a",   
 "Es su hermano/a" = "Hermano/a, hermanastro/a",   
 "Es su yerno/nuera" = "Nuera/Yerno (o pareja del hijo/a, hijastro/a)",   
 "Es su primo/a" = "Primo/a",   
 "NC" = "NC"  
 ))  
  
#Añadimos un nivel a los factores a comparar, para no obtener errores posteriormente por no coincidencia   
pobgit2$A1\_02 <- fct\_expand(pobgit2$A1\_02,"Es usted la persona de referencia" )  
pobgit2$A2\_02 <- fct\_expand(pobgit2$A2\_02,"Es usted la persona de referencia" )  
pobgit2$A3\_02 <- fct\_expand(pobgit2$A3\_02,"Es usted la persona de referencia" )  
pobgit2$A4\_02 <- fct\_expand(pobgit2$A4\_02,"Es usted la persona de referencia" )  
pobgit2$A5\_02 <- fct\_expand(pobgit2$A5\_02,"Es usted la persona de referencia" )  
pobgit2$A6\_02 <- fct\_expand(pobgit2$A6\_02,"Es usted la persona de referencia" )  
pobgit2$A7\_02 <- fct\_expand(pobgit2$A7\_02,"Es usted la persona de referencia" )  
pobgit2$A8\_02 <- fct\_expand(pobgit2$A8\_02,"Es usted la persona de referencia" )  
pobgit2$A9\_02 <- fct\_expand(pobgit2$A9\_02,"Es usted la persona de referencia" )  
pobgit2$A10\_02 <- fct\_expand(pobgit2$A10\_02,"Es usted la persona de referencia" )  
  
summary(pobgit2$A1\_02)

## Cónyuge o pareja   
## 950   
## Padre/Madre   
## 310   
## Hermano/a, hermanastro/a   
## 16   
## Hijo/a, hijastro/a   
## 106   
## Suegro/a o pareja de los mismos   
## 9   
## Cuñado/a   
## 1   
## Nuera/Yerno (o pareja del hijo/a, hijastro/a)   
## 1   
## Nieto/a, nieto políticoo/a (o pareja de los mismos)   
## 5   
## Abuelo/a   
## 9   
## Sobrino/a   
## 3   
## Primo/a   
## 3   
## Otro pariente   
## 2   
## Sin parentesco con usted   
## 3   
## NC   
## 0   
## Es usted la persona de referencia   
## 0   
## NA's   
## 74

levels(pobgit2$P5)

## [1] "Es usted la persona de referencia"   
## [2] "Cónyuge o pareja"   
## [3] "Padre/Madre"   
## [4] "Hermano/a, hermanastro/a"   
## [5] "Hijo/a, hijastro/a"   
## [6] "Suegro/a o pareja de los mismos"   
## [7] "Cuñado/a"   
## [8] "Nuera/Yerno (o pareja del hijo/a, hijastro/a)"   
## [9] "Nieto/a, nieto político/a (o pareja de los mismos)"  
## [10] "Abuelo/a"   
## [11] "Sobrino/a"   
## [12] "Primo/a"   
## [13] "Otro pariente"   
## [14] "Sin parentesco con usted"   
## [15] "NC"

levels(pobgit2$A9\_02)

## [1] "Cónyuge o pareja"   
## [2] "Padre/Madre"   
## [3] "Hermano/a, hermanastro/a"   
## [4] "Hijo/a, hijastro/a"   
## [5] "Suegro/a o pareja de los mismos"   
## [6] "Cuñado/a"   
## [7] "Nuera/Yerno (o pareja del hijo/a, hijastro/a)"   
## [8] "Nieto/a, nieto políticoo/a (o pareja de los mismos)"  
## [9] "Abuelo/a"   
## [10] "Sobrino/a"   
## [11] "Primo/a"   
## [12] "Otro pariente"   
## [13] "Sin parentesco con usted"   
## [14] "NC"   
## [15] "Es usted la persona de referencia"

str(pobgit2$P13)

## Factor w/ 14 levels "Trabajador/a por cuenta ajena (Contrato fijo)",..: 8 7 7 7 7 11 11 8 8 2 ...

# Convertimos las variables a character para poder manipularlas con mayor precisión posteriormente  
  
pobgit2 <- pobgit2 %>% mutate\_at(c(  
 "P13",   
 "P5",   
 "A1\_06",   
 "A2\_06",  
 "A3\_06",  
 "A4\_06",  
 "A5\_06",  
 "A6\_06",  
 "A7\_06",  
 "A8\_06",  
 "A9\_06",  
 "A10\_06",  
 "A1\_02",  
 "A2\_02",  
 "A3\_02",  
 "A4\_02",  
 "A5\_02",  
 "A6\_02",  
   
 "A7\_02",  
 "A8\_02",  
 "A9\_02",  
 "A10\_02"), as.character)  
  
head(pobgit2$P13)

## [1] "Jubilado o retirado del trabajo" "Parado/a"   
## [3] "Parado/a" "Parado/a"   
## [5] "Parado/a" "Trabajo doméstico no remunerado"

#Creamos una nueva variable con la actividad principal de la persona de referencia en el hogar (act\_p)  
pobgit2 <- pobgit2 %>% mutate(act\_p = ifelse(P5 == "Es usted la persona de referencia", P13, ifelse(  
 P5 == A1\_02, A1\_06, ifelse(  
 P5 == A2\_02, A2\_06, ifelse(  
 P5 == A3\_02, A3\_06, ifelse(  
 P5 == A4\_02, A4\_06, ifelse(  
 P5 == A5\_02, A5\_06, ifelse(  
 P5 == A6\_02, A6\_06, ifelse(  
 P5 == A7\_02, A7\_06, ifelse(  
 P5 == A8\_02, A8\_06, ifelse(  
 P5 == A9\_02, A9\_06, ifelse(  
 P5 == A10\_02, A10\_06, NA))))))))))))  
  
class(pobgit2$act\_p)

## [1] "character"

#Comprobamos cuantos NA hemos obtenido   
sum(is.na(pobgit2$act\_p)) #Hay 10 NA, por lo tanto no son significativos en relaci?n al total

## [1] 10

pobgit2$act\_p <- as.factor(pobgit2$act\_p)  
levels(pobgit2$act\_p) # Ahora mismo hay 22 levels que debemos recodificar a ocupados, parados, jubilados, inactivos

## [1] "Baja por enfermedad o maternidad/paternidad"   
## [2] "Estudiando/realizando cursos de formación"   
## [3] "Incapacitado/a permanente"   
## [4] "Jubilado o retirado del trabajo"   
## [5] "Jubilado/a o retirado/a del trabajo"   
## [6] "Labores del hogar"   
## [7] "NC"   
## [8] "NS/NC"   
## [9] "Otra situación"   
## [10] "Parado/a"   
## [11] "Parado/a, buscando primer empleo"   
## [12] "Parado/a, que trabajó anteriormente"   
## [13] "Realizando sin remuneración trabajos sociales, actividades benéficas"  
## [14] "Trabajador/a en un negocio familiar"   
## [15] "Trabajador/a por cuenta ajena (Contrato fijo)"   
## [16] "Trabajador/a por cuenta ajena (Contrato temporal)"   
## [17] "Trabajador/a por cuenta ajena (Fijo)"   
## [18] "Trabajador/a por cuenta ajena sin contrato"   
## [19] "Trabajador/a por cuenta propia"   
## [20] "Trabajando en un negocio familiar"   
## [21] "Trabajando por cuenta propia"   
## [22] "Trabajo doméstico no remunerado"

# Inactivos incluira las situaciones de inactividad reconocidas por el INE que no son jubilados: estudiantes, pensionistas no jubilados, incapacitados, labores del hogar...   
  
pobgit2$act\_p <- fct\_explicit\_na(pobgit2$act\_p, na\_level = "Missing")  
  
  
pobgit2$act\_p <- fct\_collapse(pobgit2$act\_p,  
 ocupada =c("Baja por enfermedad o maternidad/paternidad","Trabajador/a en un negocio familiar","Trabajador/a por cuenta ajena (Contrato fijo)","Trabajador/a por cuenta ajena (Contrato temporal)","Trabajador/a por cuenta ajena (Fijo)","Trabajador/a por cuenta ajena sin contrato","Trabajador/a por cuenta propia","Trabajando en un negocio familiar","Trabajando por cuenta propia"),  
 parada =c("Parado/a","Parado/a, buscando primer empleo","Parado/a, que trabajó anteriormente"),  
 jubilada =c("Jubilado o retirado del trabajo","Jubilado/a o retirado/a del trabajo"),  
 inactiva =c("Estudiando/realizando cursos de formación","Incapacitado/a permanente","Labores del hogar","Otra situación","Realizando sin remuneración trabajos sociales, actividades benéficas","Trabajo doméstico no remunerado"),  
 other\_level = "Valores perdidos")

Una vez hemos creado una variable simplificada, creamos una tabla de frecuencia de la nueva variable para conocer mejor su distribución:

tabla\_indicador(x = pobgit2$act\_p,   
 t = "Actividad de la persona principal en la unidad familiar")

| **Actividad de la persona principal en la unidad familiar** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **ocupada** | 583 | 0.39075067 |
| **inactiva** | 287 | 0.19235925 |
| **jubilada** | 213 | 0.14276139 |
| **parada** | 391 | 0.26206434 |
| **Valores perdidos** | 18 | 0.01206434 |

A continuación, pondremos en relación esta variable con los dos indicadores consensuales de pobreza energética:

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="act\_p", # variable independiente  
 title = "Temperatura inadecuada en el hogar y actividad de la persona de referencia de la unidad familiar", # título de la tabla   
 varname = "Actividad de la persona de referencia", # Variable independiente  
 numcol = 5) # Número de valores de la variable independiente

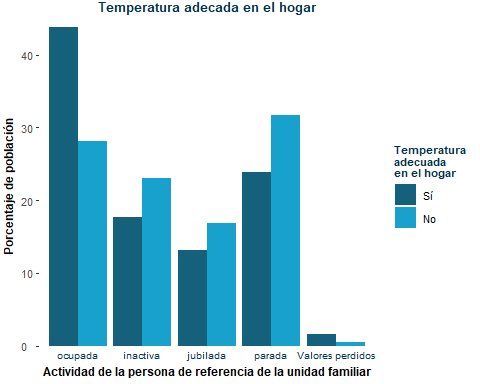
| **Temperatura inadecuada en el hogar y actividad de la persona de referencia de la unidad familiar** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cálculo** | **Actividad de la persona de referencia** | | | | |
| temp\_adecuada |  | ocupada | inactiva | jubilada | parada | Valores perdidos |
| **Sí** | *Frequency* | 454 | 183 | 136 | 247 | 16 |
| *Row Pct* | 43.82% | 17.66% | 13.13% | 23.84% | 1.54% |
| *Col Pct* | 78.14% | 63.76% | 64.15% | 63.33% | 88.89% |
| **No** | *Frequency* | 127 | 104 | 76 | 143 | 2 |
| *Row Pct* | 28.1% | 23.01% | 16.81% | 31.64% | 0.44% |
| *Col Pct* | 21.86% | 36.24% | 35.85% | 36.67% | 11.11% |

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="act\_p", # variable independiente  
 title = "Retrasos en el pago de facturas del hogar y actividad de la persona de referencia de la unidad familiar", # título de la tabla   
 varname = "Actividad de la persona de referencia", # Variable independiente  
 numcol = 5) # Número de valores de la variable independiente

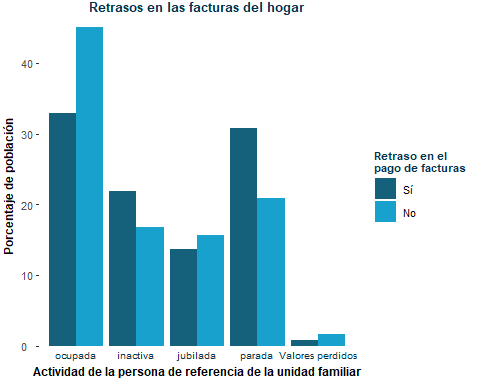
| **Retrasos en el pago de facturas del hogar y actividad de la persona de referencia de la unidad familiar** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Cálculo** | **Actividad de la persona de referencia** | | | | |
| retrasos |  | ocupada | inactiva | jubilada | parada | Valores perdidos |
| **Sí** | *Frequency* | 249 | 166 | 104 | 233 | 6 |
| *Row Pct* | 32.85% | 21.9% | 13.72% | 30.74% | 0.79% |
| *Col Pct* | 44.62% | 59.07% | 49.29% | 61.97% | 35.29% |
| **No** | *Frequency* | 309 | 115 | 107 | 143 | 11 |
| *Row Pct* | 45.11% | 16.79% | 15.62% | 20.88% | 1.61% |
| *Col Pct* | 55.38% | 40.93% | 50.71% | 38.03% | 64.71% |

Finalmente creamos los gráficos que representan los datos mostrados en las dos tablas de contingencia anteriores.

crear\_grafico\_temp(x = act\_p, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura adecada en el hogar", # título del gráfico  
 xtitle = "Actividad de la persona de referencia de la unidad familiar") # título axis X



crear\_grafico\_ret(x = act\_p, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en las facturas del hogar", # título del gráfico  
 xtitle = "Actividad de la persona de referencia de la unidad familiar") # título axis X



### Pobreza energética y miembros de la unidad familiar en situación de desocupación

Para crear una variable que nos indique la existencia de una persona parada/ en situación de desocupación, utilizaremos las siguientes variables de la base de datos inicial: \* Variable P13 que nos indica la situación laboral de la persona que contesta la encuesta \* El conjunto de variables AX\_06 que nos indican la situación laboral de los miembros de la unidad familiar

pobgit2 <- pobgit2 %>% mutate\_at(c(  
 "P13",   
 "A1\_06",   
 "A2\_06",  
 "A3\_06",  
 "A4\_06",  
 "A5\_06",  
 "A6\_06",  
 "A7\_06",  
 "A8\_06",  
 "A9\_06",  
 "A10\_06"), as.character)  
  
  
  
pobgit2 <- pobgit2 %>% mutate(parado = case\_when(  
 P13 == "Parado/a" ~ "Sí",   
 A1\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A1\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",   
 A2\_06 == "Parado/a, buscando primer empleo"~ "Sí",   
 A2\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A3\_06 == "Parado/a, buscando primer empleo"~ "Sí",   
 A3\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A4\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A4\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A5\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A5\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A6\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A6\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A7\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A7\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A8\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A8\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A9\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A9\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 A10\_06 == "Parado/a, buscando primer empleo" ~ "Sí",   
 A10\_06 == "Parado/a, que trabajó anteriormente" ~ "Sí",  
 TRUE ~ "No"   
))  
  
  
pobgit2$parado <- as.factor(pobgit2$parado)  
  
tabla\_indicador(x = pobgit2$parado,  
 t = "Unidades familiares con algun miembro desocupado")

| **Unidades familiares con algun miembro desocupado** | | |
| --- | --- | --- |
| **x** | **Total** | **Porcentaje** |
| **No** | 671 | 0.4497319 |
| **Sí** | 821 | 0.5502681 |

Una vez hemos creado la variable para identificar las unidades familiares con algun miembro en situaci´n de desocupación, pondremos en relación esta variable con nuestros indicadores de referencia de pobreza energética.

crear\_tabla2(var1 ="temp\_adecuada", #variable dependiente  
 var2 ="parado", # variable independiente  
 title = "Temperatura inadecuada en el hogar y hogares con alguna persona parada", # título de la tabla   
 varname = "Hogares con algun miembro desocupado", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

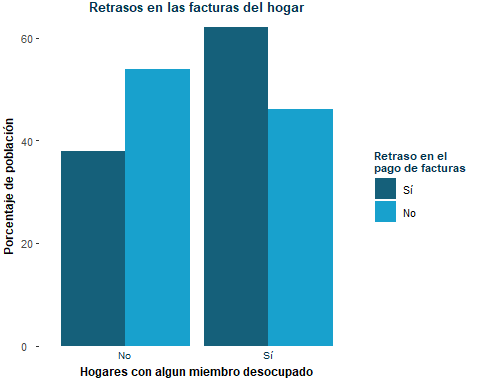
| **Temperatura inadecuada en el hogar y hogares con alguna persona parada** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Hogares con algun miembro desocupado** | |
| temp\_adecuada |  | No | Sí |
| **Sí** | *Frequency* | 488 | 548 |
| *Row Pct* | 47.1% | 52.9% |
| *Col Pct* | 72.94% | 66.91% |
| **No** | *Frequency* | 181 | 271 |
| *Row Pct* | 40.04% | 59.96% |
| *Col Pct* | 27.06% | 33.09% |

crear\_tabla2(var1 ="retrasos", #variable dependiente  
 var2 ="parado", # variable independiente  
 title = "Retrasos en el pago de facturas del hogar y hogares con alguna persona parada", # título de la tabla   
 varname = "Hogares con algun miembro desocupado", # Variable independiente  
 numcol = 2) # Número de valores de la variable independiente

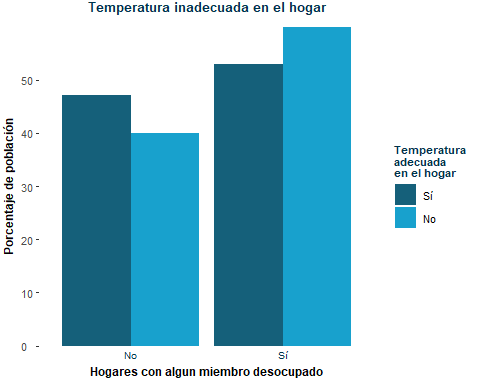
| **Retrasos en el pago de facturas del hogar y hogares con alguna persona parada** | | | |
| --- | --- | --- | --- |
|  | **Cálculo** | **Hogares con algun miembro desocupado** | |
| retrasos |  | No | Sí |
| **Sí** | *Frequency* | 287 | 471 |
| *Row Pct* | 37.86% | 62.14% |
| *Col Pct* | 43.75% | 59.85% |
| **No** | *Frequency* | 369 | 316 |
| *Row Pct* | 53.87% | 46.13% |
| *Col Pct* | 56.25% | 40.15% |

Una vez creadas las tablas, construiremos los gráficos para representar las relaciones expuestas en las tablas anteriores.

crear\_grafico\_ret(x = parado, #variable independiente  
 y = retrasos, # variable dependiente   
 title = "Retrasos en las facturas del hogar", # título del gráfico  
 xtitle = "Hogares con algun miembro desocupado") # título axis X



crear\_grafico\_temp(x = parado, #variable independiente  
 y = temp\_adecuada, # variable dependiente   
 title = "Temperatura inadecuada en el hogar", # título del gráfico  
 xtitle = "Hogares con algun miembro desocupado") # título axis X



## Dimensión 3: Características de la vivienda y del entorno