

E.D.A

Daniel Escobar David

5/4/2021

Analisis Exploratorio

En el siguiente informe, se realiza una exploración inicial de los datos para encontrar las posibles relaciones que puedan existir entre la variable respuesta *Hijos* y las demás covariables que representan las características del hogar.

Recodificación y Contextualización De Los Datos

En la construcción de la base de datos encontramos que hay multiples tipos de variables que se pueden utilizar tales como: variables numericas, variables ordinales y variables nominales.

Después de cargada la base, se encuentra que el tipo de cada variables no necesariamente es el correcto. Esto lo podemos observar en la siguiente tabla:

| Variable | Tipo | Variable | Tipo |
|----------------------|-----------|------------------------|---------|
| Codigo | character | Trabajo.conyuge | integer |
| Hijos | integer | Ingresos.totales | integer |
| Sexo.Jefe | integer | Uso.tecnologia.Jefe | integer |
| Sexo.Conyuge | integer | Uso.tecnologia.conyuge | integer |
| Tipo.Union | integer | Edu.Jefe | integer |
| Edad.Jefe | integer | Edu.Conyuge | integer |
| Edad.Conyuge | integer | Afiliacion.salud | integer |
| Region | integer | N.cuartos | integer |
| Cantidad.de.personas | integer | N.cuartos.dormir | integer |
| Estrato.vivienda | integer | Valor.Electricidad | integer |
| Trabajo.Jefe | integer | N.usos.Lavadora | integer |

A continuación se recodifica cada variable en su tipo correcto. Los resultados son los siguientes:

| Variable | Tipo | Variable | Tipo |
|----------------------|-----------|------------------------|--------|
| Codigo | character | Sexo.Jefe | factor |
| Hijos | integer | Sexo.Conyuge | factor |
| Edad.Jefe | integer | Tipo.Union | factor |
| Edad.Conyuge | integer | Region | factor |
| Cantidad.de.personas | integer | Trabajo.Jefe | factor |
| Estrato.vivienda | integer | Trabajo.conyuge | factor |
| Ingresos.totales | integer | Uso.tecnologia.Jefe | factor |
| N.cuartos | integer | Uso.tecnologia.conyuge | factor |
| N.cuartos.dormir | integer | Edu.Jefe | factor |
| Valor.Electricidad | integer | Edu.Conyuge | factor |
| N.usos.Lavadora | integer | Afiliacion.salud | factor |

Finalmente los datos se encuentran bien codificados.

% NA

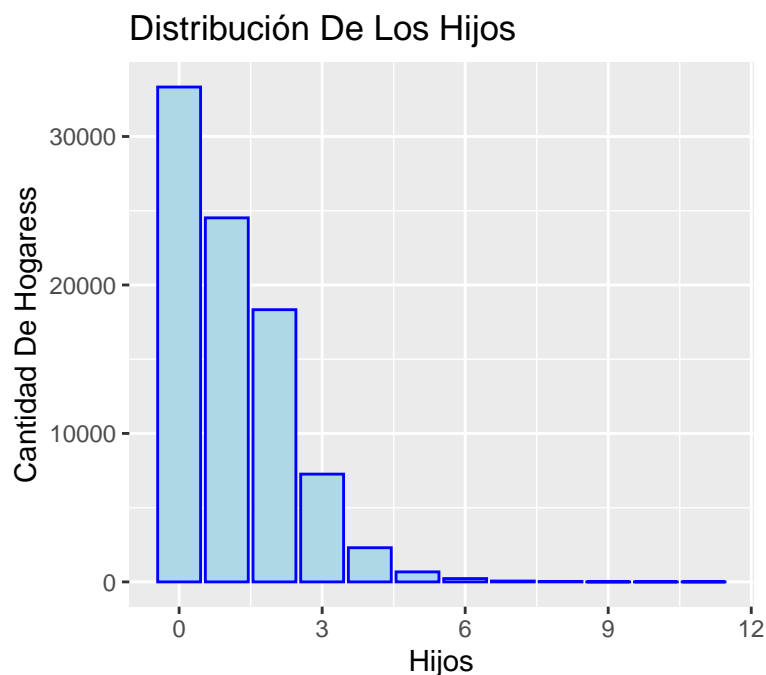
Ahora, se procede a encontrar el porcentaje de valores *NA* en cada variable, ya que esto puede ser bastante significativo en procesos posteriores como el ajustar un modelo o seleccionar variables de importancias. La tabla resultante es la siguiente:

| Variable | Tipo | Variable | Tipo |
|----------------------|------------------|------------------------|------------------|
| Codigo | 0 | Sexo.Jefe | 0 |
| Hijos | 0 | Sexo.Conyuge | 43.4001198930185 |
| Edad.Jefe | 0 | Tipo.Union | 0 |
| Edad.Conyuge | 43.4001198930185 | Region | 0 |
| Cantidad.de.personas | 0 | Trabajo.Jefe | 0 |
| Estrato.vivienda | 0 | Trabajo.conyuge | 43.4001198930185 |
| Ingresos.totales | 0 | Uso.tecnologia.Jefe | 0 |
| N.cuartos | 0 | Uso.tecnologia.conyuge | 43.4001198930185 |
| N.cuartos.dormir | 0 | Edu.Jefe | 0 |
| Valor.Electricidad | 0 | Edu.Conyuge | 43.4001198930185 |
| N.usos.Lavadora | 0 | Afiliacion.salud | 0 |

Las variables que representan información del cónyuge tienen un 43.4% de valores faltantes. Lo anterior se puede deber a que muy probablemente sean hogares en los que el jefe del hogar sea soltero, viudo, etc. Por tal motivo se considera que los análisis realizados posteriormente serán solo enfocados a hogares en los cuales se encuentren solo el Jefe del hogar, es decir que el jefe del hogar no viva en pareja.

Distribución Variable Respuesta

Inicialmente, observamos como esta distribuida la variable respuesta ya que esto puede ser de vital importancia en los análisis posteriores.

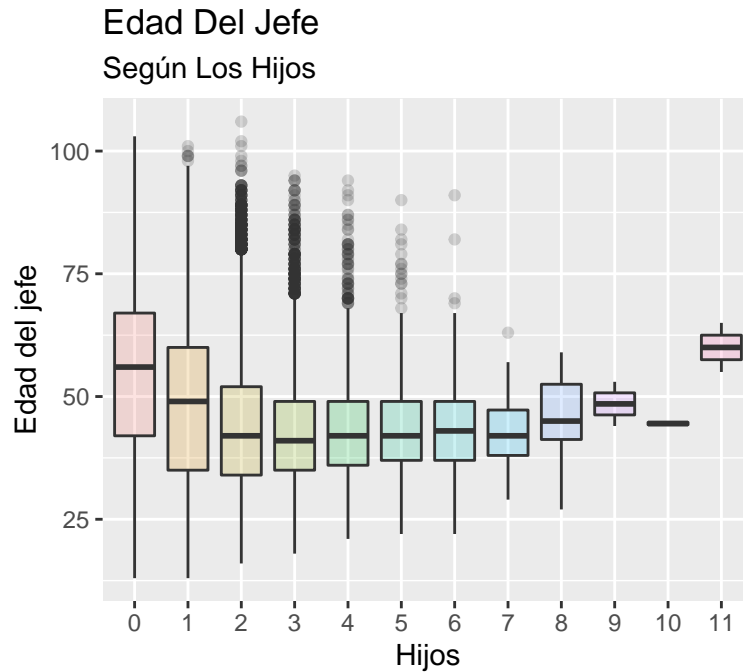


La cantidad de los hijos en los hogares que más predomina es 0, es decir que no hay hijos, después de esto la cantidad de hijos en los hogares decae. A partir de esto, se infiere que es más frecuente no encontrar hijos o encontrar al menos uno, y que es menos frecuente encontrar más de uno.

Relaciones Variable Respuesta y Covariables

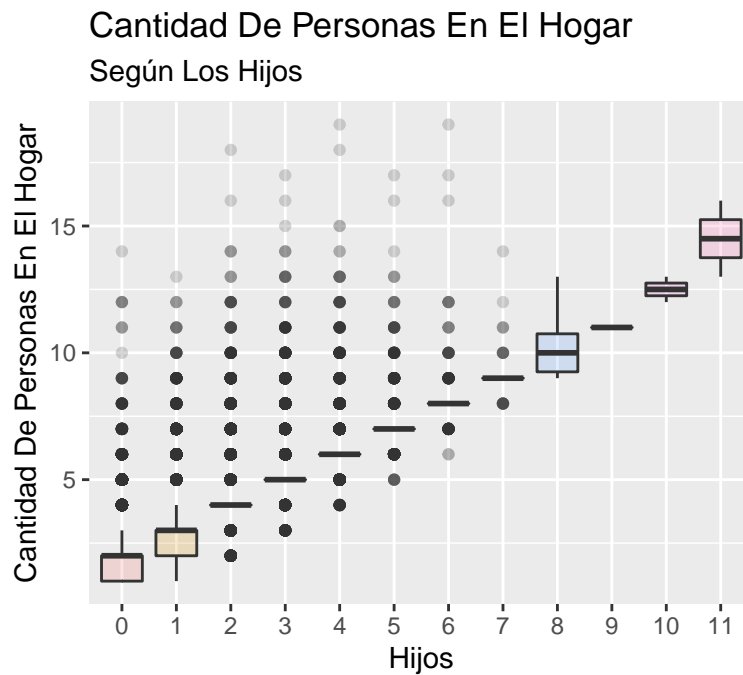
Ahora se procede a analizar las posibles relaciones que existan entre la variable respuesta y las covariables, esto ayudara a determinar que variables podrán contribuir a explicar la variabilidad de las variable respuesta.

Edad Del Jefe



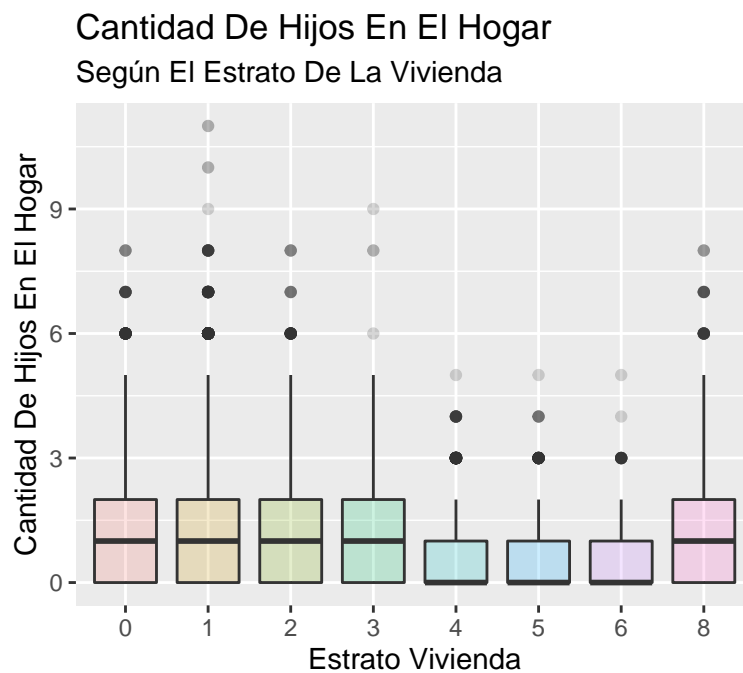
Cuando en el hogar no hay hijos o hay a lo más 1, la edad del jefe del hogar parece variar bastante, mientras que para los hogares donde hay más de 1 hijos, en general, la edades no presentan mayores variaciones. De lo anterior se puede decir, que la edad del jefe puede influir al momento de determinar la cantidad de hijos que se encuentra en el hogar.

Cantidad De Personas



En general es de esperar que entre mayor sea la cantidad de personas en el hogar, mayor sea el numero de hijos en esta. De este modo, parece que la cantidad de personas en el hogar influye positivamente en el numero de hijos.

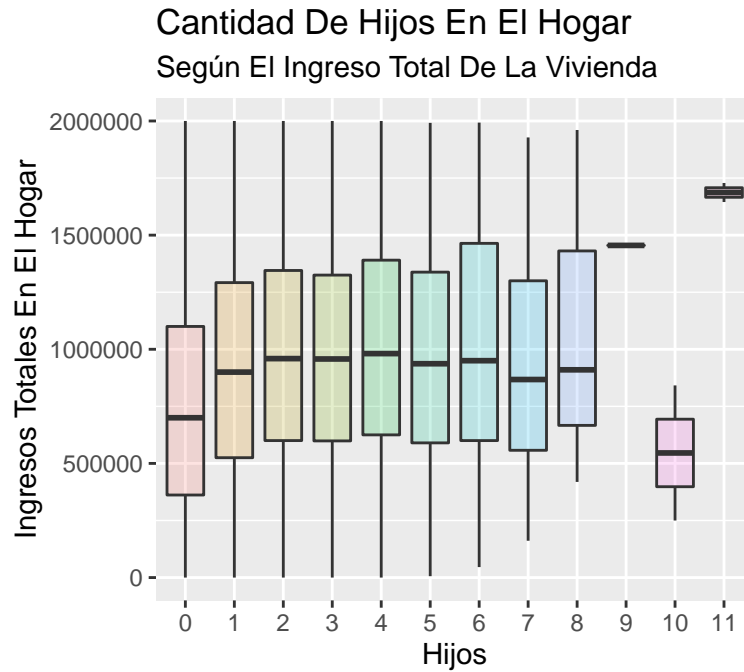
Estrato De La Vivienda



En general no parece que el estrato de la vivienda influya mucho en la cantidad de hijos que se pueda encontrar en un hogar, sin embargo, en los hogares de estrato 4 y 5 parece encontrarse una cantidad menor a

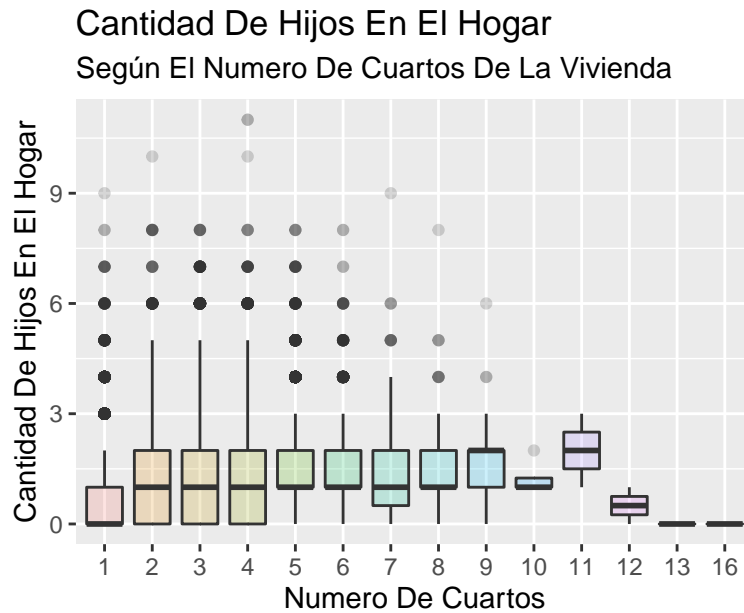
la de los hogares de los demás estratos. Por lo anterior no podemos decir completamente que el estrato de la vivienda sea influyente en el numero de hijos del hogar.

Ingresos Totales



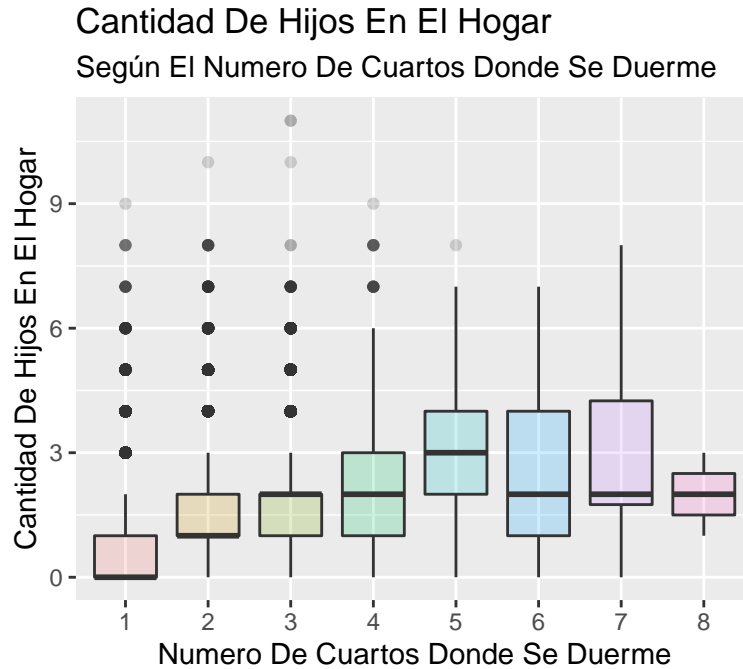
Los ingresos totales de los hogares parecen variar mucho y sin embargo no parecen cambiar mucho a medida que se tengan más o menos hijos en el hogar. Por lo anterior, los ingresos totales del hogar no influyen de manera directa en la cantidad de hijos que se pueda encontrar en el.

Numero De Cuartos



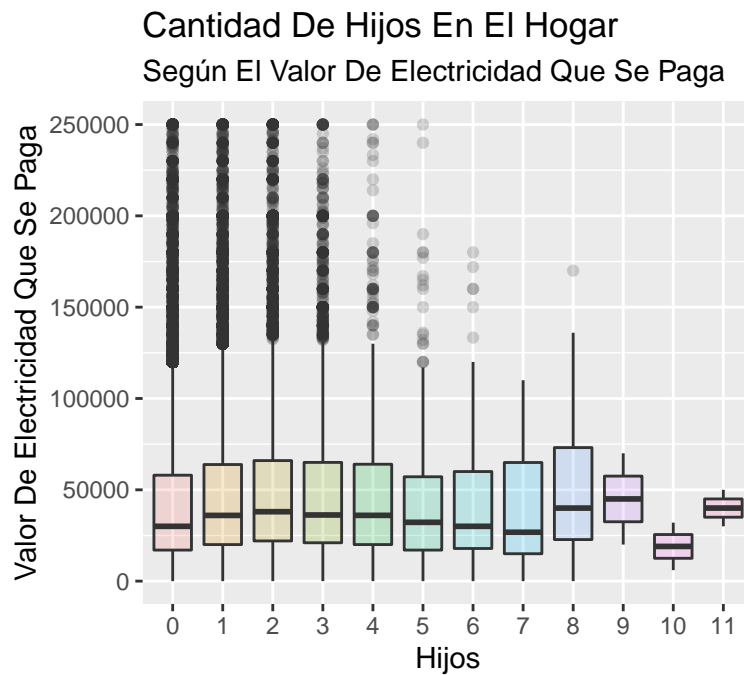
No necesariamente si hay más cuartos en un hogar se van a encontrar más hijos en este, pero posiblemente si exista algún tipo de variación. Por lo tanto no podemos asegurar con certeza que en numero de cuartos en el hogar influya en el numero de hijos que se encuentren en el.

Numero De Cuartos



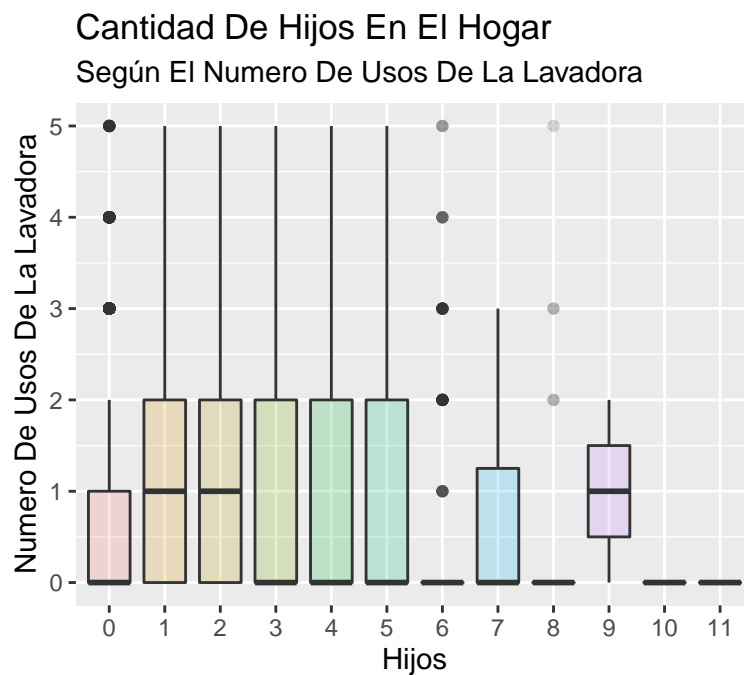
Sutilmente se muestra un patron, aunque no muy fuerte, de que a medida en que el numero de cuartos donde se duerme en el hogar aumenta, tambien lo hace la cantidad de hijos que se encuentran en el. De este modo se infiere que muy probablenmte el numero de cuartos donde se duerme en el hogar ayude a explicar la variabilidad de la cantidad de hijos que se encuentren en el.

Valor Electricidad



No se observan patrones ni negativos ni positivos en relación al cambio de los hijos que se encuentren en el hogar. De este modo el valor que se paga de electricidad no influye directamente con el numero de hijos que se encuentre en el hogar.

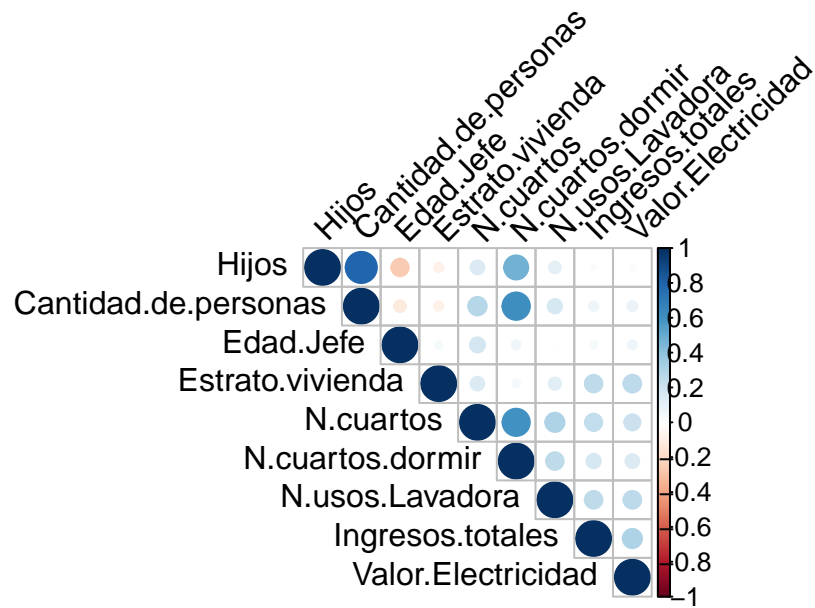
Numero Usos Lavadora



En general no se observan patrones muy claros en que existan algun tipo de relación entre las variables. Por lo tanto, decimos que el numero de usos de la lavadora no influye en la cantidad de hijos que se puedan

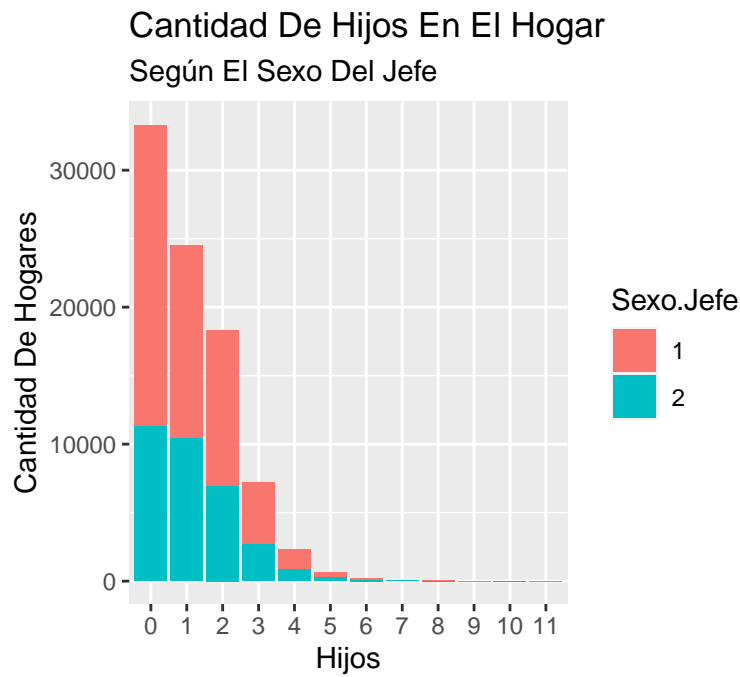
encontrar en un hogar.

Correlaciones



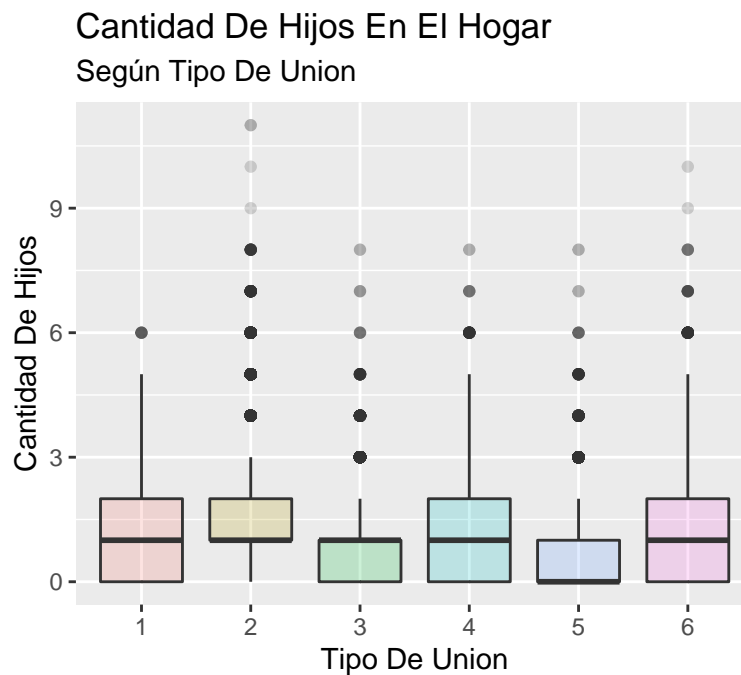
Las variables que tiene relaciones lineales más fuertes con la variable *Hijos*, son: *Cantidad.de.personas*, *N.cuartos.dormir* y *Edad.Jefe*. Por lo anterior estas 3 variable son candidatas a ser consideradas dentro de un modelo inicial.

Sexo Del Jefe



Sutilmente se observa una diferencia entre los hijos del hogar cuando el jefe del hogar en hombre o mujer. Por lo anterior podemos decir que el sexo del jefe del hogar puede ayudar a explicar la variabilidad del numero de hijos en un hogar.

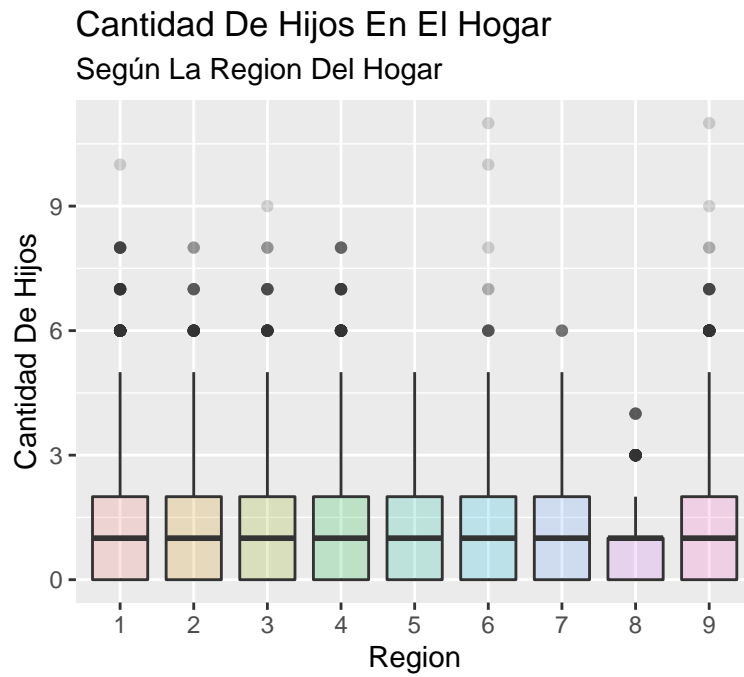
Tipo De Union



A priori parece que según el tipo de unión la cantidad de de hijos que se encuentre en el hogar puede variar. Dicho esto, es posible que el tipo de unión sea una variable que no ayude a explicar la cantidad de hijos en

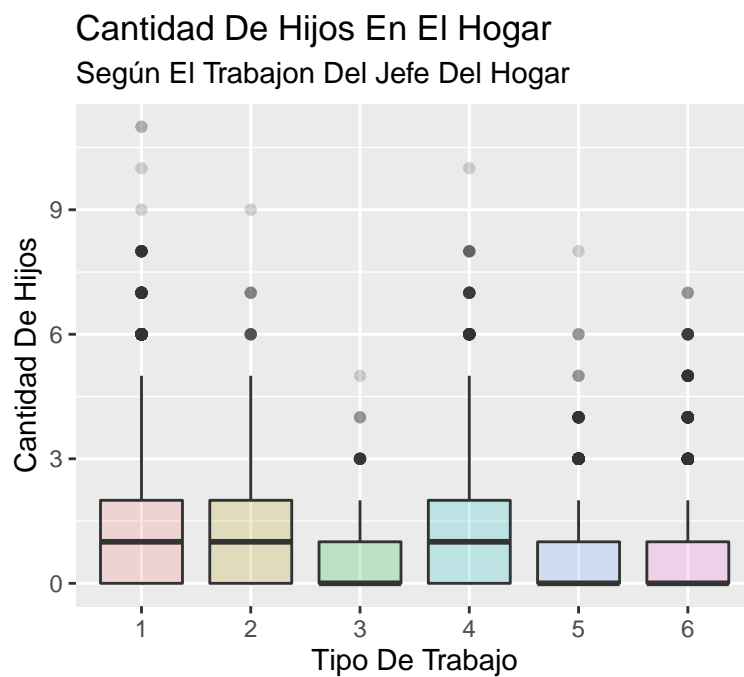
un hogar.

Region



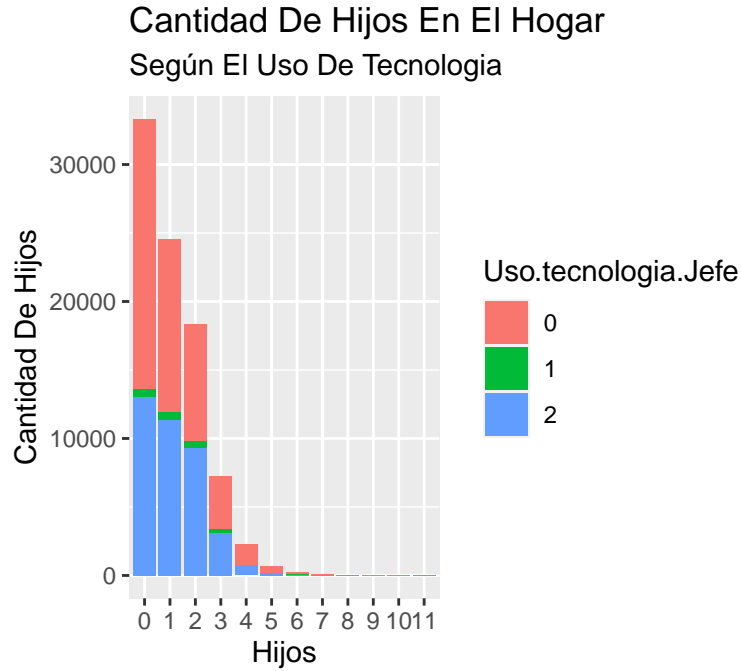
Se observa claramente que la cantidad de hijos en el hogar es independiente de la region en la cual se encuentre la vivienda. Se concluye que la region no es una variable que influya en la cantidad de hijos que se encuentren en un hogar.

Trabajo Jefe



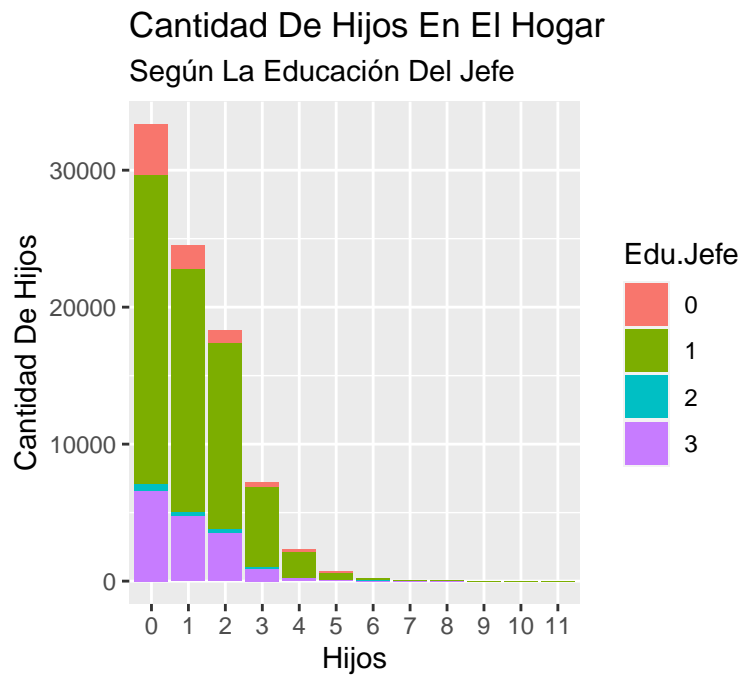
No parece haber una relación muy clara entre el tipo de trabajo del jefe del hogar y la cantidad de hijos que se encuentren en el hogar. Por lo anterior, no es recomendable sugerir que el tipo de trabajo del jefe de hogar influya en la cantidad de hijos en el.

Uso De Tecnologia



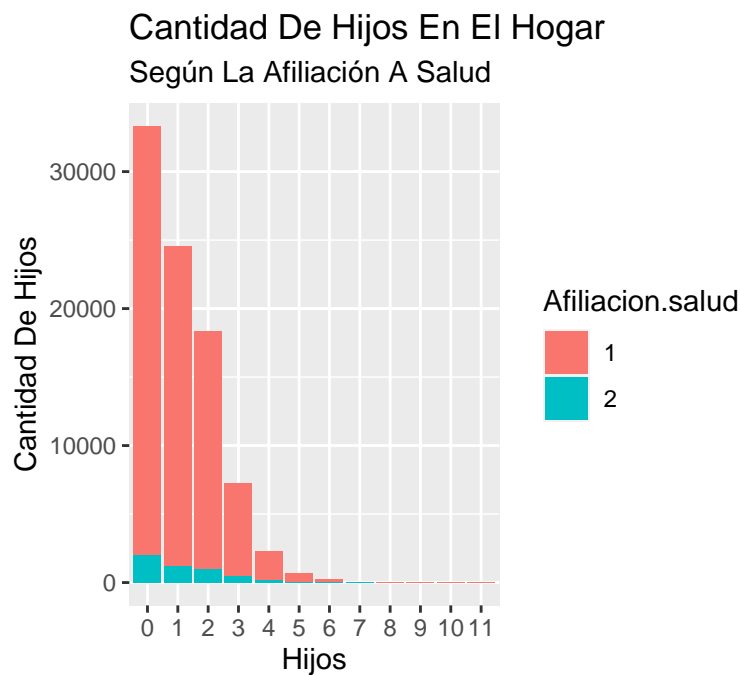
Claramente se observa que el numero de hijos cambia dependiendo el uso de tecnologia que tenga el jefe del hogar. Por lo tanto, es muy probable que la variable *Uso.tecnologia.Jefe* ayude a explicar la variabilidad de la cantidad de hijos en un hogar.

Educación Jefe



Es notable el cambio en la cantidad de hogares en los que se encuentran ciertas cantidades de los hijos según la educación que tenga el jefe del hogar. Dicho esto, es importante considerar la variable *Edu.Jefe* para poder explicar el número de hijos que se encuentren en el hogar.

Afiliación Salud



No se observa que la afiliación a salud determine o segmente fuertemente la cantidad de hijos que se encuentren en un hogar. Por lo tanto la variable *Afiliacion.salud* no ayuda a determinar la cantidad de hijos que se

encuentren en un hogar,

Conclusiones Analisis Exploratorio

Por el análisis anterior, se consideran las siguientes variables para ajustar un modelo que prediga el numero de hijos que se encuentran en un hogar:

- Edad.Jefe
- Cantidad.de.personas
- N.cuartos.dormir
- Valor.Electricidad
- Tipo.Union
- Trabajo.Jefe
- Uso.tecnologia.Jefe
- Sexo.Jefe
- Edu.Jefe

Estas variables muy probablemente nos ayuden a ajustar un modelo predictivo que funcione correctamente.