

Mapecto de la distribución del ingreso en la Ciudad de Buenos Aires

Felipe González

17 de julio de 2015

Introducción

El propósito de este trabajo es mapear la distribución del ingreso per cápita familiar en la Ciudad de Buenos Aires. El análisis de la distribución del ingreso en su dimensión espacial aporta una perspectiva que enriquece el análisis y puede utilizarse más adelante como insumo para otros análisis (como puede ser por ejemplo la relación entre esta distribución y la topografía de las redes de servicios públicos provistos por un municipio, o su recaudación impositiva segregada por barrios). Se selecciona el ingreso per cápita familiar por considerarse al más acertado para esos análisis ulteriores propuestos en la medida en que es el fundamento de las decisiones de económicas del hogar. De aquí en más cuando en el trabajo se hable de ingreso, se habla del ingreso per cápita familiar (*ipcf*).

Para la construcción del mapa de ingreso en la Ciudad de Buenos Aires, se utilizará el Censo Nacional de Población, Hogares y Viviendas de 2010 (**Censo** de aquí en más), dado que es la única fuente estadística que ofrece información desagregada a un nivel geográfico lo suficientemente específico como para realizar mapas que sean significativos. Lamentablemente, el Censo no pregunta por ingreso de manera directa, con lo cual el desafío radica en proponer un proceso que permita a partir de las variables existentes en el **Censo**, llegar a un proxy del ingreso del hogar.

Por su lado, la Encuesta Anual de Hogares Urbanos (**EAHU**), con diseño muestral pero de muy amplia cobertura, sí pregunta por los diferentes ingresos del hogar. Sin embargo no ofrece niveles de desagregación geográficos útiles, más allá de grandes aglomerados urbanos. Es por ello que el uso complementario de ambas puede llevar a alcanzar el objetivo deseado. Mediante la **EAHU** se puede observar las posibles relaciones existentes entre determinadas variables equivalentes en **Censo** y en **EAHU** a seleccionar y el ingreso del hogar.

Con este objetivo principal como norte, el trabajo presenta los siguiente objetivos secundarios:

- Analizar las variables disponibles en el **Censo 2010** y la **EAHU** del tercer trimestre del 2010 y comparar las variables que ambas tienen en común y las diferencias que pueda llegar a haber hacia el interior de cada una
- Determinar, entre las variables que comparten el Censo y la EAHU, cuál de ellas (ya sea en modo individual o como combinatoria de las mismas en un índice o indicador) presenta la mayor relación con el ingreso
- Obtener del **Censo** los valores para esa variable (individual o compuesta) a nivel de radio censal
- Mapear esa variable en mapa de la Ciudad Autónoma de Buenos Aires
- Esbozar un primer análisis del mapa del ingreso de la Ciudad

El ingreso en la Ciudad de Buenos Aires

La Ciudad de Buenos Aires es el distrito nacional con el Producto Bruto Geográfico (PBG) per cápita más alto del país. Realizar esta comparación es algo dificultosa ya que no todos los distritos del país informan su PBG para todos los años ni con los mismos precios. La Dirección de Estadísticas e Investigaciones Económicas

de Mendoza realizó un trabajo comparativo¹ que sirve como fuente para presentar datos comparados para 2010. Estos resultados se resumen en el **Cuadro 1**.

CUADRO 1. Producto bruto regional per capita por provincias

Provincia	Población	PBG	PBG per capita
Ciudad de Buenos Aires	2890151	136078	0.047
Neuquén	551266	13148	0.024
Santa Fe	3194537	32593	0.010
Chubut	509108	5023	0.010
Buenos Aires	15625084	153620	0.010
Córdoba	3308876	30265	0.009
Mendoza	1738929	15514	0.009
La Rioja	333642	2602	0.008
San Juan	681055	5176	0.008
Entre Ríos	1235994	9249	0.007
Corrientes	992595	5112	0.005

PBG a precios de mercado en millones de pesos de 1993.

La magnitud del PBG de la Ciudad es a su vez comparable con el de países desarrollados como Israel, España o Grecia². En la comparación de ciudades y utilizando paridad de poder de compra (ppp)³, una comparación más justa, Buenos Aires (23.606 USDppp) se puede ubicar junto a ciudades de la región como Santiago de Chile (23.929 USDppp) o Sao Paulo en Brasil (20.650 USDppp) pero también por encima de grandes ciudades de países desarrollados como Porto en Portugal (21.674 USDppp) o Napoli en Italia (19.451 USDppp). Incluso no se encuentra muy lejos de ciudades como Lille en Francia (30.259 USDppp) o Newcastle en Inglaterra (29.978 USDppp).

A la hora de analizar la distribución del ingreso con otros grandes aglomerados urbanos del país se observa que la Ciudad de Buenos Aires es el décimo aglomerado urbano en el ranking de distribución del ingreso de acuerdo al índice de Gini. Esta posición mejora si no se toman en consideración los agrupamientos de aglomerados de toda una provincia que realiza la EAHU. En ese caso, Buenos Aires sería superada en distribución equitativa del ingreso solo por 6 aglomerados urbanos.

Aglomerados urbanos de la EAHU según índice de Gini del Ingreso per cápita familiar % latex
table generated in R 3.1.2 by xtable 1.7-4 package % Tue Aug 4 16:45:55 2015

subdominio	gini
bahía blanca - cerri	0.369
santa rosa - toay	0.383
la rioja	0.389
resto de santa cruz	0.394
resto de jujuy	0.406
resto de tucumán	0.408
santiago del estero - la banda	0.408
río cuarto	0.408
gran córdoba	0.410
ciudad de buenos aires	0.412

Adentrándose en el análisis de la variable ingreso hacia el interior de la Ciudad, tal cual la registra la **EAHU**, se puede observar una variable de distribución claramente asimétrica con la mayoría de los valores registrados

¹Fuente: <http://www.deie.mendoza.gov.ar/publicaciones/mendozaencontexto/pbg/pbgmc.xls>

²World Economic Outlook Database-April 2015

³<http://www.brookings.edu/research/reports2/2015/01/22-global-metro-monitor>

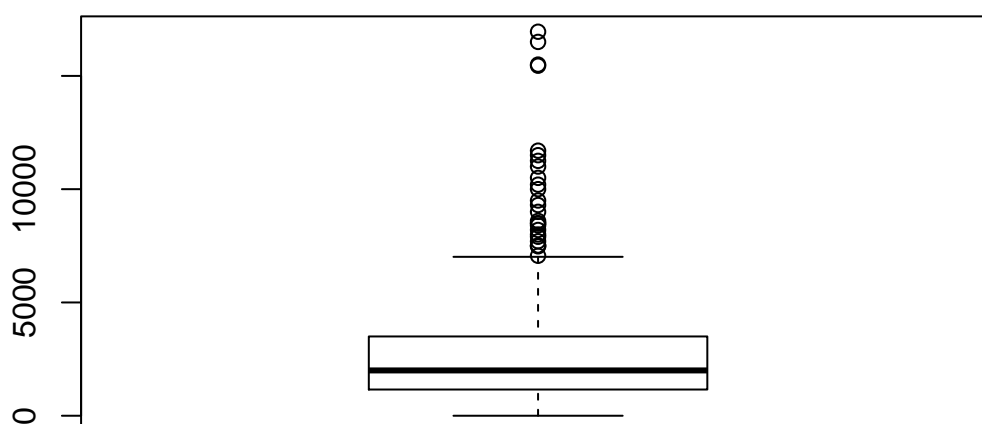
en los valores inferiores de la variable. Al observar la distribución por cuartiles, se observa que la mitad de los hogares tienen un ingreso menor a los 2 mil pesos mensuales y solo el 25% superior supera los \$3.500.

Cuartiles de ingreso per cápita familiar - Ciudad de Buenos Aires 2010

```
##          0.25      0.5 0.75
## ipcf 1157.021 1999.873 3500
```

A la hora de observar su distribución con el fin de detectar casos extremos que puedan impactar en el análisis, se puede observar que hay algunos casos que exigen mayor atención a la hora de incluirlos en el análisis puesto que pueden perturbar el mismo. Se puede afirmar a priori, pero con cierta confianza y que se verificará más adelante en el trabajo, que aquellos hogares con ingresos superiores a 15 mil pesos per cápita constituyen casos extremos.

Ingreso per cápita familiar - Ciudad de Buenos Aires 2010



Variables comunes al Censo 2010 y EAHU

El siguiente cuadro resume las variables que tanto **Censo** como **EAHU** miden y que serán utilizadas en este trabajo para intentar realizar una estimación de la distribución espacial del ingreso.

CENSO	EAHU
Hogares	Hogares
V01 - Tipo de vivienda	IV1 - Tipo de vivienda
H16 - Total de ambientes	IV2 - Total de ambientes
H05 - Material predominante de los pisos	IV3 - Material predominante de pisos interiores
H06 - Material de la cubierta exterior del techo	IV4 - Material de la cubierta exterior del techo
H07 - Revestimiento interior o cielorraso del techo	IV5 - Revestimiento interior o cielorraso del techo
H08 - Tenencia de agua	IV6 - Tenencia de agua
H09 - Procedencia del agua para beber y cocinar	IV7 - Fuente de agua
H10 - Tiene baño o letrina	IV8 - Tiene baño o letrina
H11 - Descarga para limpieza del inodoro	IV10 - Descarga para limpieza del inodoro
H12 - Desagüe del inodoro	IV11 - Desagüe del inodoro
H14 - Combustible usado principalmente para cocinar	II8 - Combustible utilizado para cocinar
PROP - Tenencia de la vivienda y propiedad del terreno	II7 - Régimen de tenencia
Personas	Personas
P01 - Relación de parentesco	CH03 - Relación de Parentesco

CENSO	EAHU
P02 - Sexo	CH04 - Sexo
P05 - Pais de nacimiento	CH15 - Pais de nacimiento
P03 - Edad	CH06 - Edad
P08 - Asiste o asistió a un establecimiento educativo	CH10 - Asiste o asistió a un establecimiento educativo
P09 - Nivel educativo que cursa o cursó	CH12 - Nivel educativo que cursa o cursó
P10 - Completó ese nivel	CH13 - Completó ese nivel
P11A- Último grado que aprobó en ese nivel	CH14 - Último grado que aprobó en ese nivel
CONDACT - Condición de actividad	ESTADO - Condición de actividad

En la medida en que Buenos Aires presenta el mayor nivel de ingreso per cápita del país, así como también conforma un espacio urbano consolidado al ser una de las ciudades cuyo proceso de urbanización data de los tiempos de la colonia, es de esperar que algunas variables vinculadas a aspectos de la pobreza estructural no tengan tanta incidencia.

Un ejemplo de ello es la fuente de procedencia del agua, variable que muestra que todos los hogares tienen agua corriente. E incluso son una minoría (el 0.16%) los que no tienen baño o solo el 1.43% el que tiene agua por fuera de la vivienda. En este sentido, no es esperable que estas variables sirvan como proxy para una variable con mucha variabilidad como el ingreso en la Ciudad de Buenos Aires.

Por otro lado, las variables que hacen a los materiales de la vivienda o al mecanismo de disposición de desechos en el baño, presentan mayor variabilidad y puede ser que estén relacionadas con el ingreso. Sin embargo, este set de variables es de esperar que solamente sirva para distinguir positivamente sectores de bajos ingresos, más no así para permitir la georeferenciación de los sectores de altos ingresos. Mientras que la ausencia de baño con desagüe o de calidad suficiente en el techo o piso de la vivienda podría ser un claro indicador de un hogar con bajos ingresos, hacia el interior de los hogares con presencia de ambas condiciones puede existir una dispersión de ingresos significativa. Es decir, mientras que la ausencia de estas condiciones básicas serían suficientes para considerar un hogar como de bajos ingresos, su presencia es necesaria más no suficiente para considerar a un hogar como de altos ingresos.

La hipótesis de este trabajo es que, de las variables disponibles tanto en el **Censo** como en la **EAHU**, son aquellas vinculadas a los años de escolaridad, a la condición de actividad y a la conformación del hogar las que mayor relación pueden presentar con el ingreso de modo que nos permitan poder determinar caracterizar a un hogar como de altos ingresos. Sin embargo, es de esperar que, mientras que este conjunto de variables pueda servir para la identificación del extremo superior de la escala de ingresos, no sea tan sensible con respecto al segmento inferior.

Es por ello que se presume que es necesario el análisis conjunto y complementario de ambos sets de variables, unas para identificar positivamente los sectores de más altos ingresos y el otro para identificar positivamente los sectores de menores ingresos.

Variables candidatas como proxy del ingreso

A partir de la **EAHU** se puede determinar el sentido y la fuerza de la relación entre las variables seleccionadas, que también se pueden encontrar en el **Censo**, y el ingreso per cápita familiar, que es el tipo de ingreso seleccionado en este trabajo. Para ello se testea cada variable seleccionada como regresor en un modelo de regresión lineal simple contra el ingreso, y se registran sus parámetros de interés. Por un lado se toma en cuenta aquellas variables que se presumen guardar alguna relación con el ingreso como conformación del hogar o características del jefe de hogar y/o cónyuge. Las variables vinculadas a las condiciones edilicias se toman no aisladamente sino como componentes del indicador NBI. Otras, dada su escasa variabilidad en el contexto urbano de la Ciudad ya mencionado, se omiten directamente.

Variables individuales presentes en EAHU y Censo

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Tue Aug 4 16:45:56 2015

variable	beta	pbeta	r	pValue
jefaturaFemenina	-271.915	0.086	0.002	0.086
jefaturaCompleta	-329.059	0.035	0.004	0.035
habitanteUnico	1009.444	0.000	0.042	0.000
hogaresMult	-462.108	0.191	0.001	0.191
nbi1	-1499.873	0.000	0.014	0.000
nbi2	-1023.610	0.072	0.003	0.072
nbi3	-1447.250	0.465	-0.001	0.465
nbi4	-2147.354	0.362	-0.000	0.362
nbi5	-2180.444	0.009	0.007	0.009
nbi	-1459.029	0.000	0.023	0.000
aniosEscolaridad	221.765	0.000	0.136	0.000
propietario	325.377	0.042	0.004	0.042
jefeNacidoArg	1178.657	0.000	0.030	0.000

De esta tabla se puede extraer como conclusión que, de todos los modelos lineales simples, los años de escolaridad en el hogar (construido a partir de los años de escolaridad del jefe y/o cónyuge) es la variable que explica la mayor variabilidad del ingreso y, al mismo tiempo, la mayor significatividad estadística al aportar el menor *p-value*. El sentido de la relación, como es de esperar, es positivo de modo que a mayor cantidad de años de escolaridad del jefe y/o cónyuge del hogar, es de esperar un ingreso mayor en ese hogar. También es necesario considerar la presencia en el hogar de algún indicador de NBI, que hacen referencia a las condiciones habitacionales y de infraestructura de los hogares, como una variable relacionada en sentido negativo con el ingreso del hogar (aunque en magnitudes menos significativas que *Años de escolaridad*, así como también con mayores, aunque aceptales, niveles de error).

Por último, también aparece como significativa la nacionalidad del jefe de hogar, dado que se espera una relación negativa (aunque de intensidad poco significativa) entre la condición de inmigrante del jefe y el ingreso del hogar.

Indicadores construidos a partir de variables presentes en EAHU y Censo

A priori, se puede comenzar a fundamentar algunas presunciones tomadas al inicio del trabajo. En primer lugar, el grado de desarrollo urbano de la Ciudad de Buenos Aires, hace que aquellos indicadores que la bibliografía asume como negativamente vinculados con el ingreso (como los resumidos en los diferentes componentes del NBI) no tengan suficiente significatividad económica ni estadística. De modo que si bien puedan ser útiles para poder identificar y georeferenciar nichos de pobreza estructural, no sirven para identificar el otro extremo de la distribución del ingreso.

Complementariamente, los años de escolaridad se presentan como la variable más significativa para una identificación positiva de los sectores de altos ingresos. Pero un modelo de regresión lineal simple de ingreso sobre esta variable, explica muy poco de la variabilidad de los ingresos. Es por eso que es necesario la utilización de otra serie de indicadores construidos a partir de las variables simples que permitan un mejor análisis.

La búsqueda de indicadores para la estimación del ingreso en unidades geográficas pequeñas, complementando encuestas de ingresos y censos, llevó al INDEC a la elaboración de un indicador denominado **Índice de Privación Material de los Hogares (IPMH)**, compuesto a su vez de 2 sub-indicadores:

- *Condiciones Habitacionales (CONDHAB)*
 - Material predominante de los pisos interiores

- Material predominante de la cubierta externa de los techos
- Presencia de revestimiento o Cielorraso en la parte interior de los techos
- Tenencia de instalación de baño con arrastre de en el inodoro

CONDHAB reagrupa los tres indicadores detallados, adoptando como criterio para la aproximación del nivel de calidad de los mismos, la durabilidad y aislamiento que proporcionan los distintos materiales, entendiendo que algunos son más aptos que otros y que ofrecen distintas condiciones de habitabilidad. El producto final es una variable ordinal con categorías: *Suficiente*, *Parcialmente insuficiente* e *Insuficiente*. Finalmente para construir *CONDHAB* se toma el indicador “Tenencia de instalación de baño con arrastre de en el inodoro” que se considera *Insuficiente* si el hogar no dispone de instalación de baño con descarga de agua en el inodoro. Finalmente se considera que la vivienda presenta condiciones habitacionales *INSUFICIENTES* cuando sus materiales constructivos o sus instalaciones sanitarias se encuentran por debajo de la norma.

- *Capacidad Económica (CAPECO)*

- Condición de ocupación de los miembros del hogar
- Años de escolaridad de los miembros del hogar
- Tamaño familiar en Adultos Equivalentes

CAPECO surge a partir de la modelización en una única ecuación los incrementos porcentuales en los ingresos laborales debidos a un incremento marginal en los años de escolaridad a partir de una función continua con pendientes diferenciales por nivel educativo. El trabajo del INDEC incluye las variables de sexo, edad y lugar de residencia como explicativas de los diferenciales de ingreso a través de un conjunto de 35 dummies que reflejan todas las combinaciones posibles entre las tres variables, asumiendo que existen relaciones entre éstas. Esta última alternativa fue la seleccionada por considerarse que las diferencias de los ingresos entre sexos, por ejemplo, no son uniformes a través de las distintas regiones geográficas. De acuerdo a la condición de ocupación de cada miembro del hogar se imputa un valor a la variable CP, lo mismo con los años de escolaridad que con determinados ajustes conforman la variable Valor de los Años de Escolaridad (VAE). Finalmente se reconstruye para cada hogar el tamaño familiar de acuerdo a la metodología de adultos equivalentes.

En el **Apéndice Metodológico** se proveen las tablas necesarias para calcular CAPECO así como el código fuente para obtener a nivel de radio censal utilizando el software REDATAM desarrollado por CEPAL, que es el que INDEC utiliza para procesamiento de datos censales.

El desarrollo de CAPECO en el presente trabajo incluye una simplificación (en tanto que solo se toma la Ciudad de Buenos Aires) y una diferencia con el desarrollado por el INDEC en base a los datos del Censo 2001. En éste, se incluyó en el cuestionario ampliado una pregunta sobre percepción de pensión y/o jubilación. En el Censo de 2010 esta pregunta fue incluida en el cuestionario ampliado que se aplicó solo a una muestra de la población y, por lo tanto, no permite obtener resultados de esta variable para cada hogar. Dado que, según el cuestionario ampliado, en la Ciudad de Buenos Aires el 91,5%⁴ de la población en edad de jubilarse percibe jubilación o pensión, se decidió en este trabajo imputar como jubilados o pensionados al conjunto de la población en edad de jubilarse. Esto tiene por efecto una sobreestimación del indicador CAPECO para la Ciudad. Queda para trabajos posteriores la recalibración del modelo sin incluir esta variable de modo que pueda llevarse adelante una estimación de CAPECO de acuerdo a los datos del Censo 2010.

Finalmente, la conformación de CAPECO para cada hogar está dada por la siguiente ecuación:

$$CAPECO = \frac{\sum_{i=1}^n (CP_i * VAE_i)}{\sum_{i=1}^n Aeq_i}$$

⁴http://www.censo2010.indec.gov.ar/cuadrosDefinitivos/Total_pais/P15-Total_pais.xls

Donde: n : total de integrantes del hogar CP : condición de percepción (asume distintos valores según la condición de actividad, la edad, el sexo y el lugar de residencia) VAE : valor de los años de escolaridad formal invertidos en el mercado laboral. Aeq : valor en unidades de adulto equivalente de cada integrante de hogar (varía de acuerdo al sexo y la edad, siguiendo una tabla de necesidades calóricas y nutricionales)

Dados estos indicadores podemos observar su desempeño como regresores en un modelo de regresión simple contra el ingreso. En este trabajo se tomó la decisión de mantener *CONDHAB* como una variable con 3 categorías de modo que mantenga un nivel de medición ordinal y no perder información. Es por ello que el parámetro de *CONDHAB* en el modelo debe ser

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Tue Aug 4 16:45:58 2015

variable	beta	pbeta	r	pValue
capeco	266.866	0.000	0.399	0.000
condhab	453.182	0.085	0.002	0.085

Se observa un notable mejor desempeño de **CAPECO** como regresor (da cuenta de un 39.91% de variabilidad del ingreso con un nivel de confianza más que significativo), mientras que **CONDHAB** es menos competitiva que NBI (39.91). De todos modos, es necesario recordar que es lógico esperar que aquellas variables relacionadas con sectores de bajos ingresos tengan mala performance como regresor en un modelo simple, dado que no permite explicar la variación de ingresos hacia el interior de los sector de ingresos medios y altos.

Este abordaje encuentra aquí sus limitaciones. Permite obtener una medida de performance de **CAPECO** como regresor en un modelo lineal simple, que nos permite avanzar en el mapeo. Deja la incógnita sobre qué indicador es el mejor a la hora de georeferenciar a los sectores de bajos ingresos. Se avanzará sobre este punto una vez realizados los mapas correspondientes. Antes se procederá a ofrecer los resultados del modelo de regresión simple del **Ingreso per cápita familiar del hogar (ipcf)** en función del **CAPECO** del mismo.

Modelo de regresión simple con CAPECO como regresor

El modelo queda definido entonces por la siguiente ecuación:

$$ipcf_i = \beta_0 + \beta_1 CAPECO_i + \epsilon_i$$

Corriendo la regresión sobre la base EAHU 2010 se obtienen los siguientes estimadores:

$$ipcf_i = 348.2 + 266.9 CAPECO_i$$

El siguiente cuadro presenta el resumen de los parámetros de interés del modelo

% latex table generated in R 3.1.2 by xtable 1.7-4 package % Tue Aug 4 16:45:58 2015

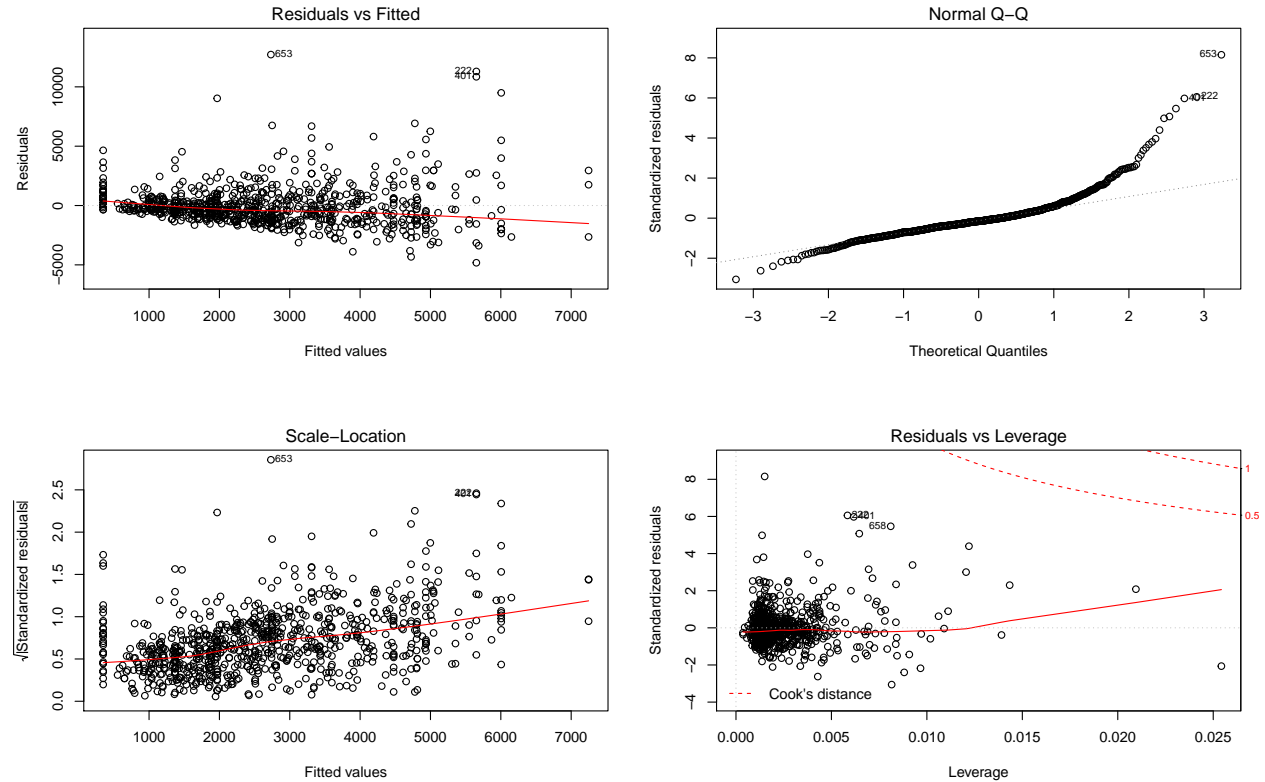
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	348.1600	115.9769	3.00	0.0028
capeco	266.8662	11.5025	23.20	0.0000

Del modelo se desprende que el valor esperado del ingreso familiar es 348.2 cuando CAPECO es 0 (ya sea por ausencia de años de escolaridad o por ausencia de miembros del hogar ocupados y/o jubilados)⁵ y que ese ingreso aumenta en 266.9 pesos por cada aumento de una unidad de CAPECO (asumiendo que todas las otras variables omitidas en el modelo, permanezcan constantes).

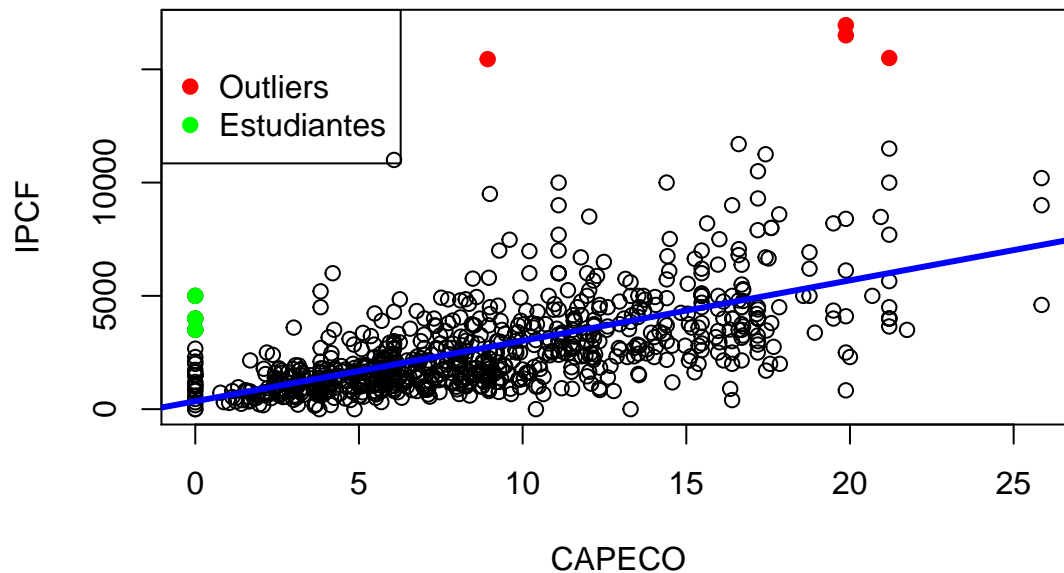
⁵Es importante recordar que no siempre la ordenada al origen tiene sentido en los modelos de regresión. En este caso es evidente que existen hogares con ingreso nulo.

Para analizar la bondad de ajuste del modelo se puede comenzar por ver que la media de los residuos es practicamente 0 ($-9.212745e-14$). A su vez se pueden observar los diferentes análisis de residuos para detectar tendencias subyacentes, casos extremos y/o muy excesivamente influyentes

Análisis de residuos del modelo



Se puede observar la incidencia de algunos outliers que fueron mencionados previamente al comienzo del trabajo relacionados a casos extremos por ingresos muy altos. También se registra un potencial problema con **CAPECO** en relación a hogares conformados por estudiantes que reciben algún tipo de dieta de los padres y por lo tanto son considerados inactivos (por ende CAPECO toma valor 0) pero que tienen un ingreso medio o alto. Esto explica parte de la variabilidad de ingreso cuando CAPECO toma valor 0. De todos modos no afecta significativamente el desempeño del modelo.



Al eliminar los casos extremos se observa una leve mejoría en el desempeño del modelo, más que nada un leve aumento en R^2 que alcanza el 42.35%

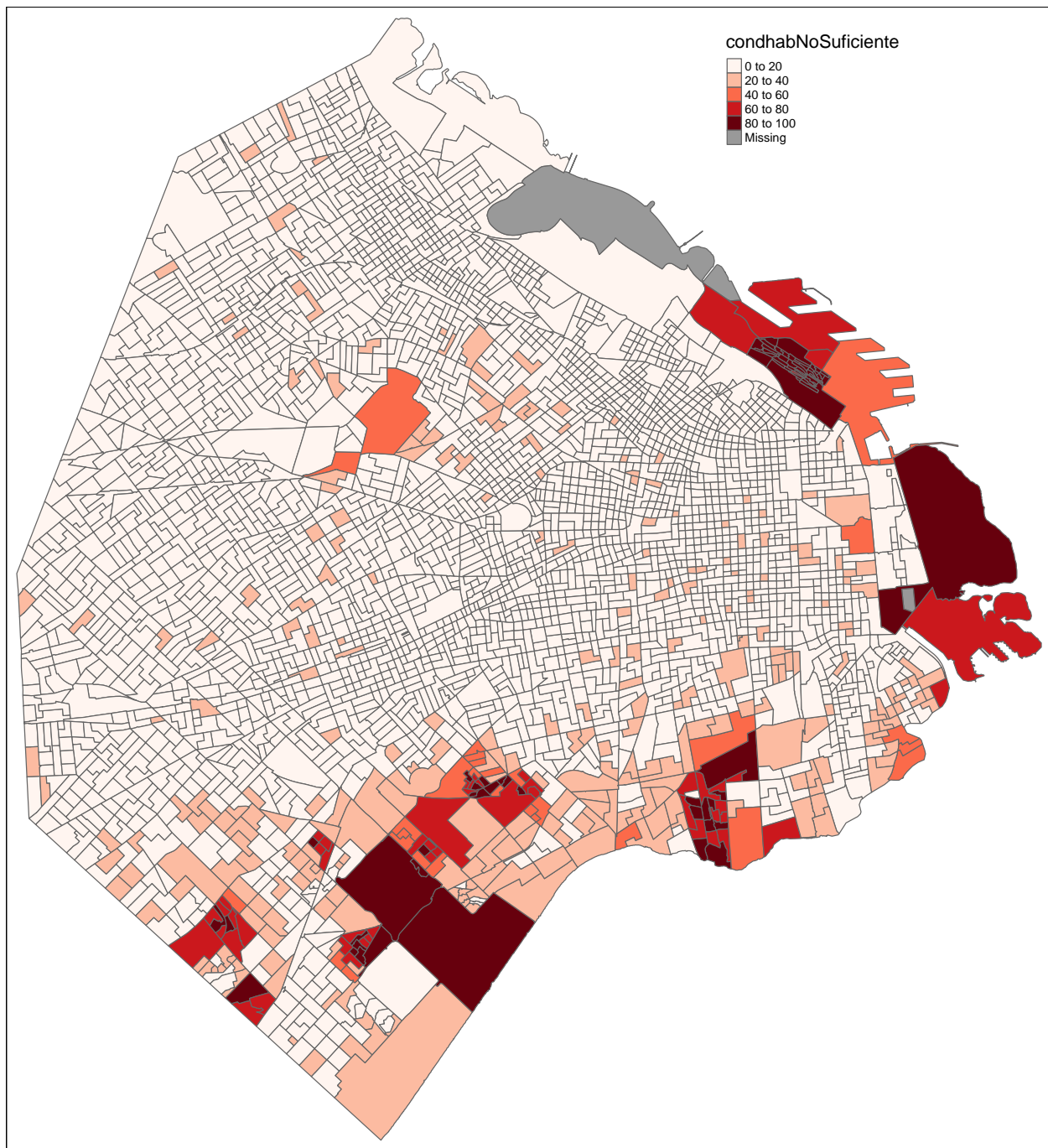
Cuadro resumen modelo sin outliers

```
##
## Call:
## lm(formula = ipcf ~ capeco, data = hogaresSinOut, weights = pondera)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -190258  -33006   -9681   19609  340232
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    426.16     103.62   4.113 4.31e-05 ***
## capeco         251.41      10.33  24.340 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 58410 on 804 degrees of freedom
## Multiple R-squared:  0.4242, Adjusted R-squared:  0.4235
## F-statistic: 592.4 on 1 and 804 DF,  p-value: < 2.2e-16
```

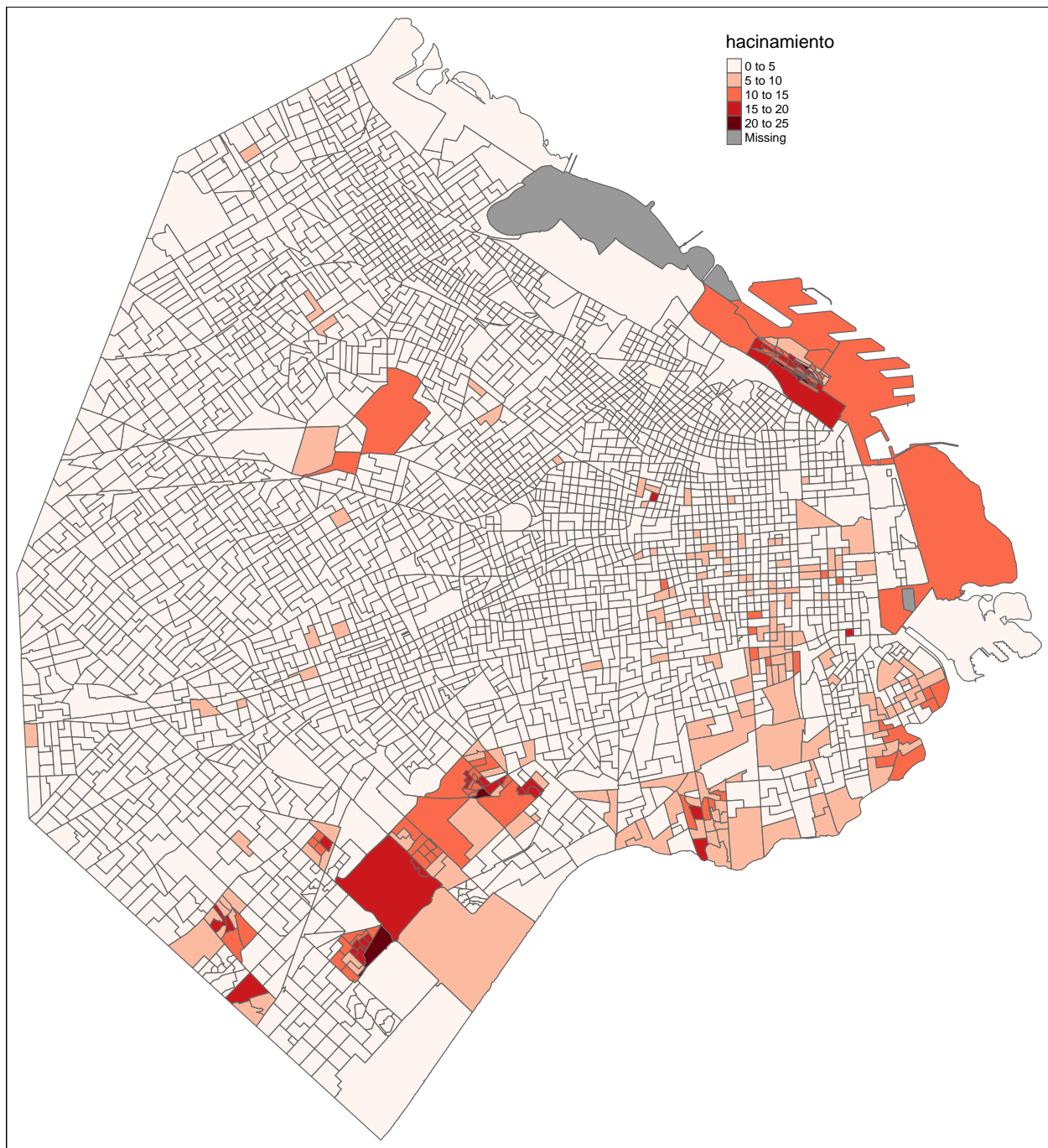
Variable en el Censo 2010: análisis y distribución

Mapa del ingreso en la Ciudad de Buenos Aires

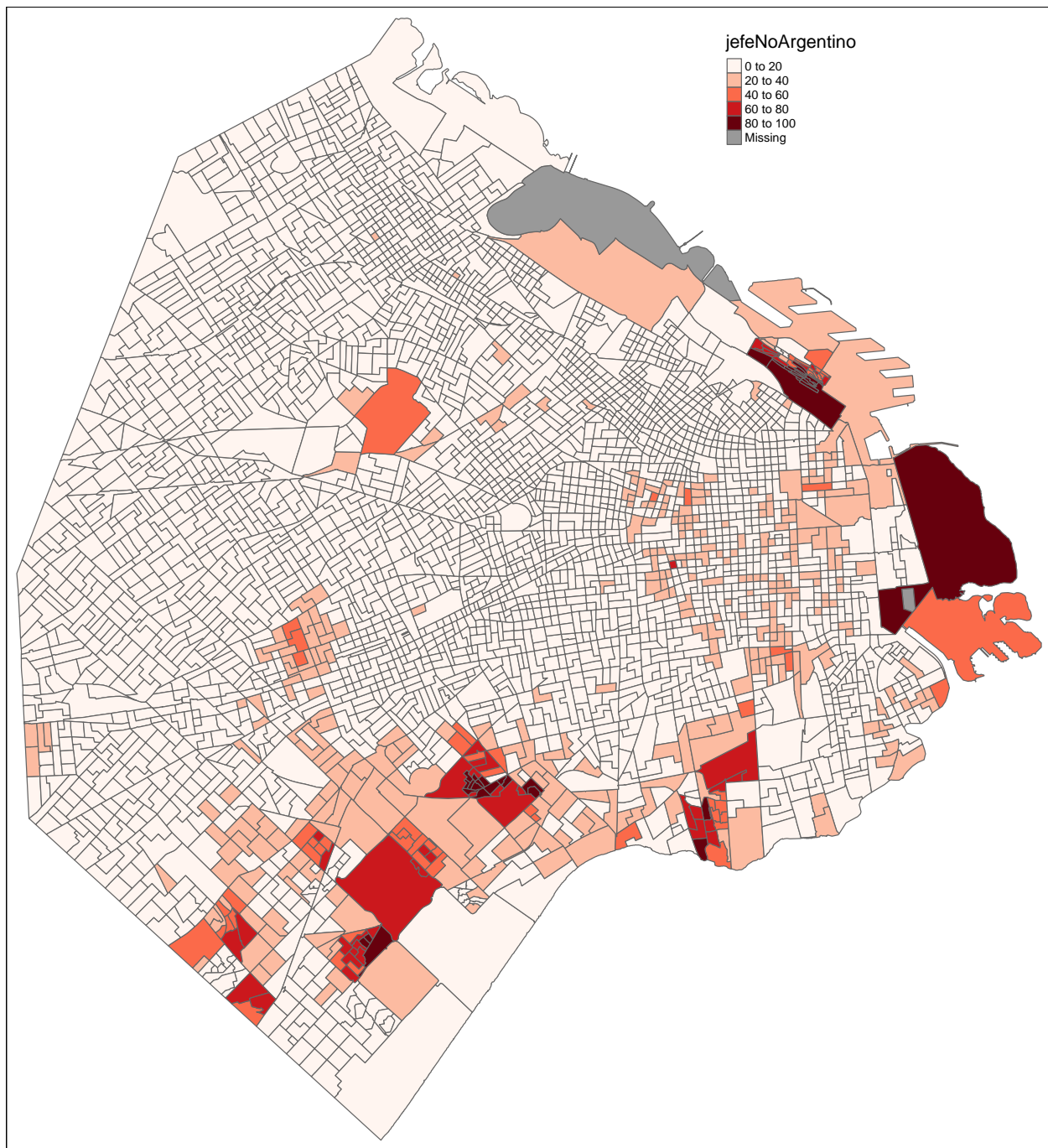
Mapa Condiciones Habitacionales CONDHAB



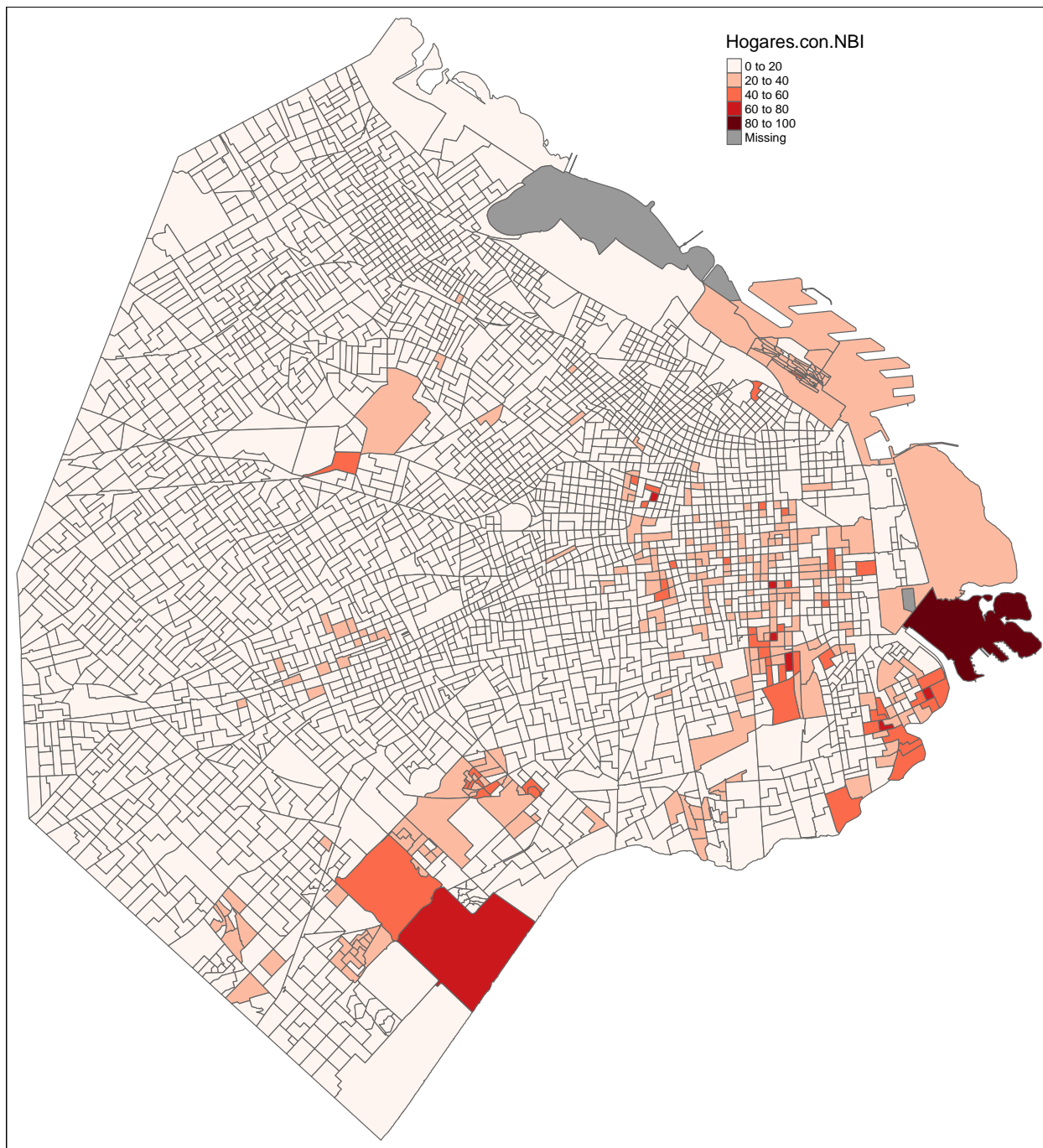
Mapa Hacinaamiento



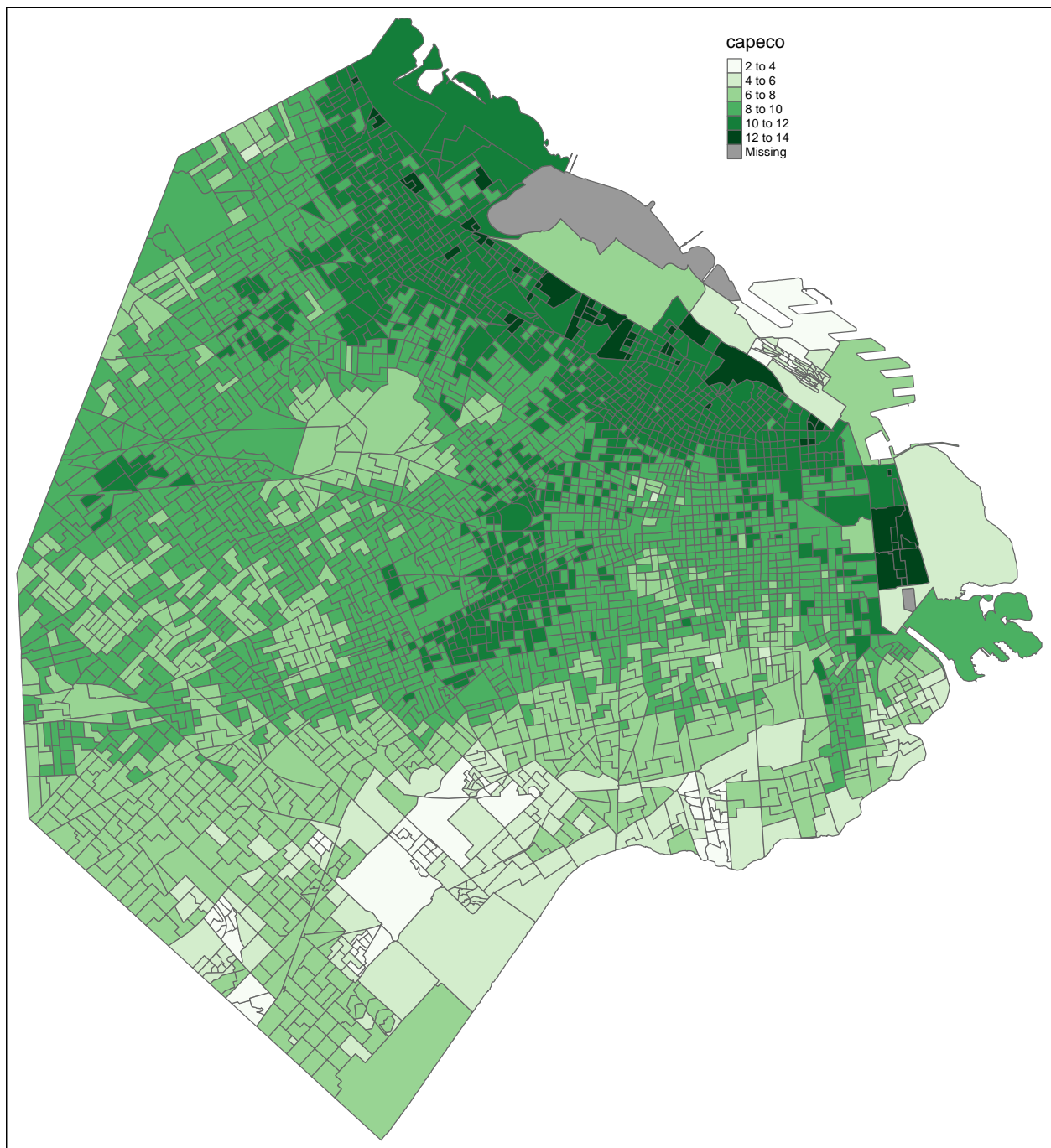
Mapa Jefe de hogar inmigrante



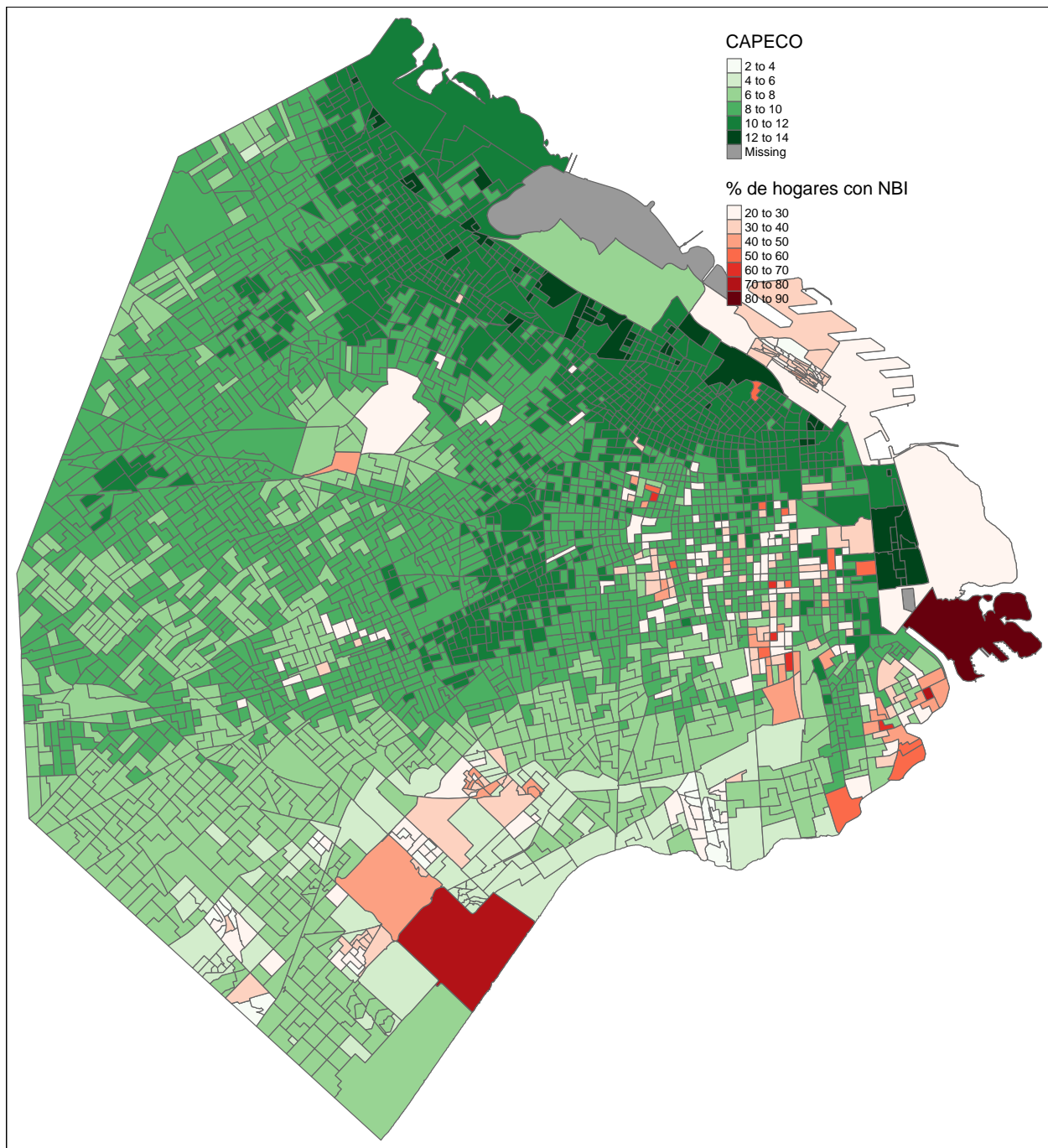
Mapa NBI



Mapa CAPECO



Mapa combinado de CAPECO y NBI



Análisis del mapa del ingreso