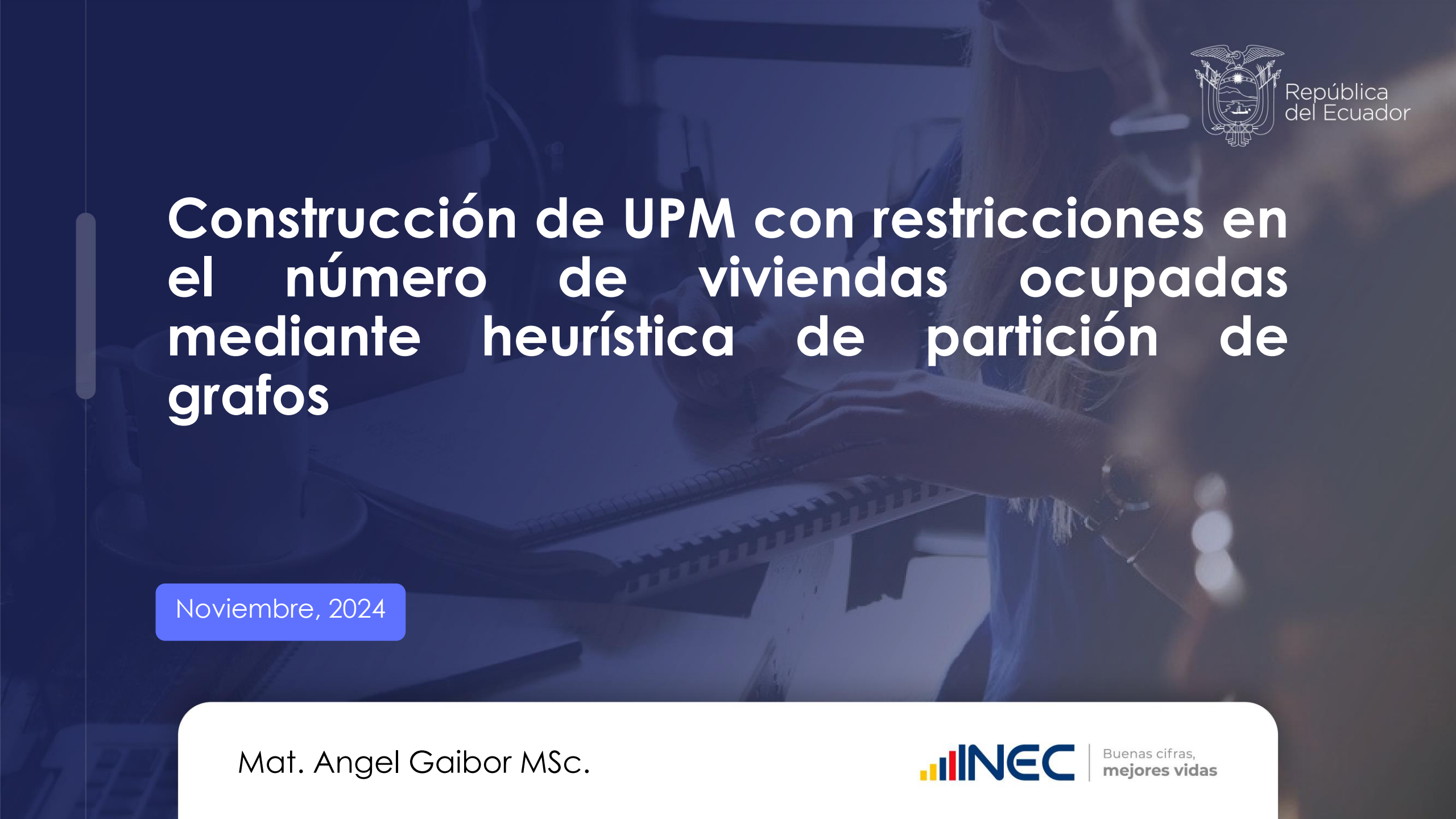


República del Ecuador

Construcción de UPM con restricciones en el número de viviendas ocupadas mediante heurística de partición de grafos

Noviembre, 2024

Mat. Angel Gaibor MSc.

INEC Buenas cifras, mejores vidas

www.ecuadorencifras.gob.ec

Buenas cifras, mejores vidas

01 **Introducción**
¿Qué es un conglomerado?
Características deseables

02 **psuR**
Flujo
De la Geografía a la Matemática

03 **Formulación del problema**
Programación entera
Formulación formal
Un ejemplo sobre Teselia

04 **Resultados y futuros desafíos**
Conglomeración sobre Ambato
Futuros desafíos

Buenas cifras,
mejores vidas

01.

Introducción

Unidad Primaria de Muestreo (UPM)

Buenas cifras, mejores vidas

¿Qué es un conglomerado?

Según Valliant, Dever y Kreuter (2013), los conglomerados son áreas geográficas, establecimientos, escuelas, o cualquier otro tipo de agregación de unidades utilizados en diseños muestrales multietápicos, pues son particularmente prácticos y costo-eficientes en comparación al muestreo aleatorio simple.

Las encuestas dirigidas a hogares suelen considerar al menos dos etapas de selección, seleccionando conglomerados en la primera etapa, los cuales pasan a denominarse Unidades Primarias de Muestreo.

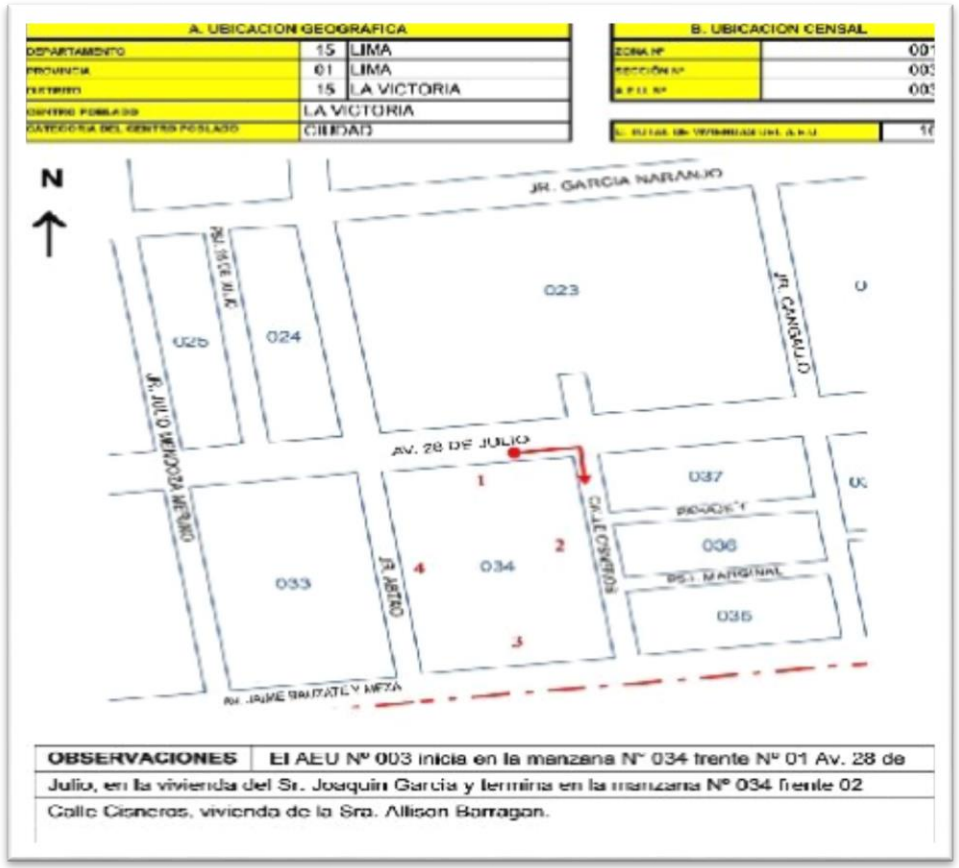
La selección de las unidades dentro de cada conglomerado (unidades secundarias de muestreo) varía dependiendo de la metodología de la encuesta, pudiendo ser de enumeración completa, aleatoria, diferenciada, entre otras.

Unidad Primaria de Muestreo (UPM)

Buenas cifras, mejores vidas

Características

- Es un área geográfica contigua de fácil identificación y recorrido.
- Tienen un número de viviendas dentro de un rango determinado (INEC más de 60 viviendas).
- Está conformada por polígonos (manzanas o sectores dispersos) de la misma parroquia censal.
- Las UPM se utilizan para todas las encuestas dirigidas a hogares en el periodo intercensal.



Ejemplo de cartografía censal

www.ecuadorencifras.gob.ec

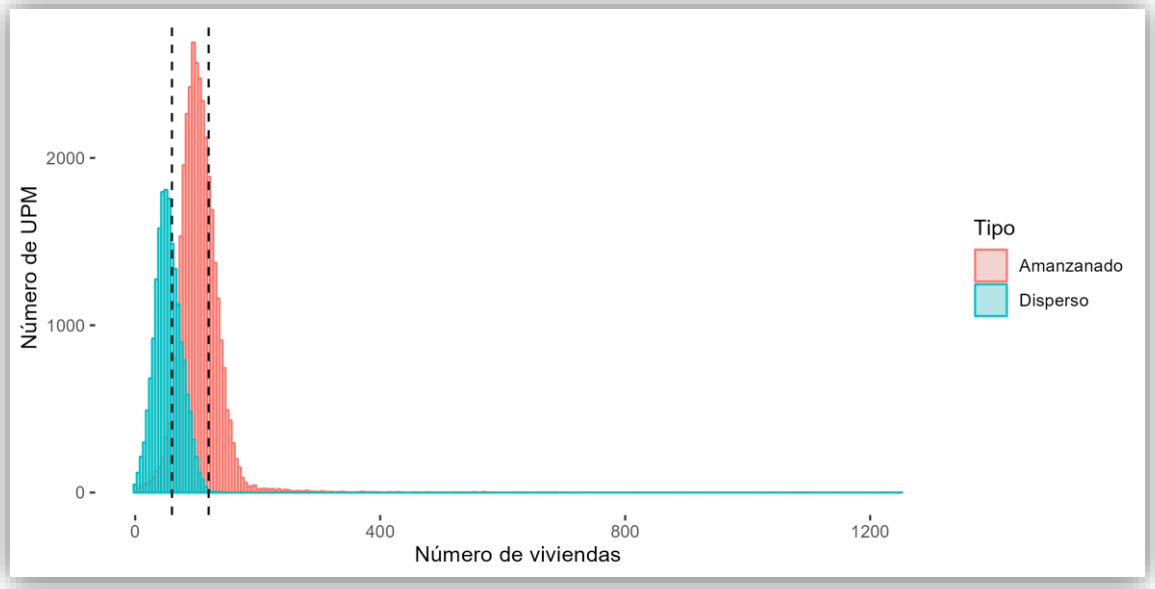
Unidad Primaria de Muestreo (UPM)

Buenas cifras, mejores vidas

Sectores censales

- Son delimitaciones geográficas construidas con fines operativas para distribución de carga en el censo.

- Se utilizaban como unidades primarias de muestreo hasta el año 2017.

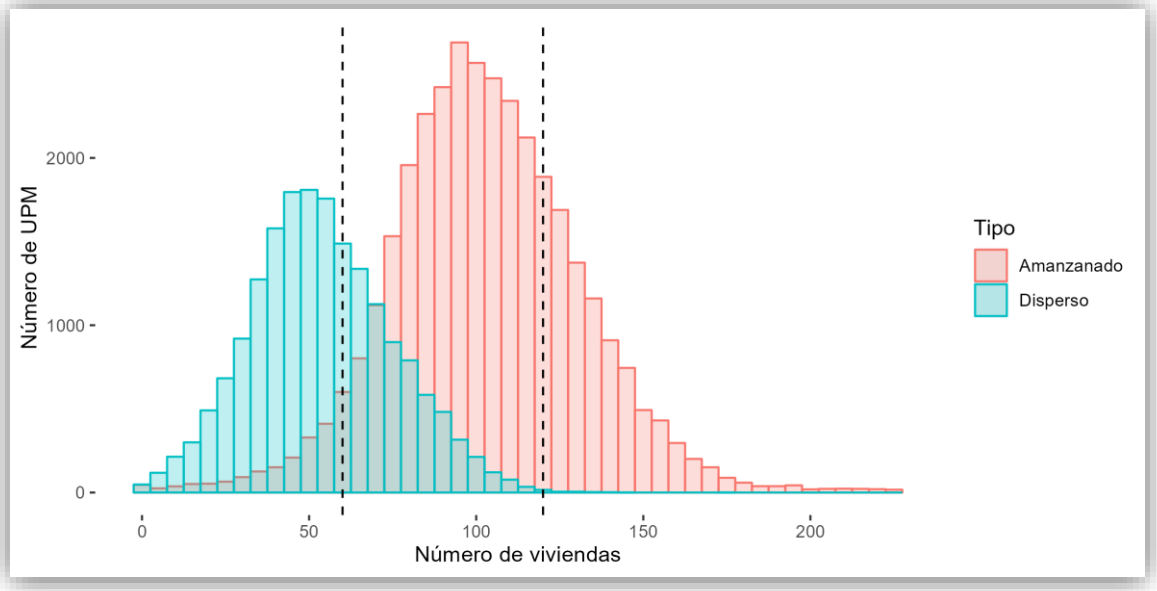


Unidad Primaria de Muestreo (UPM)

Sectores censales

- Son delimitaciones geográficas construidas con fines operativas para distribución de carga en el censo.

- Se utilizaban como unidades primarias de muestreo hasta el año 2017.

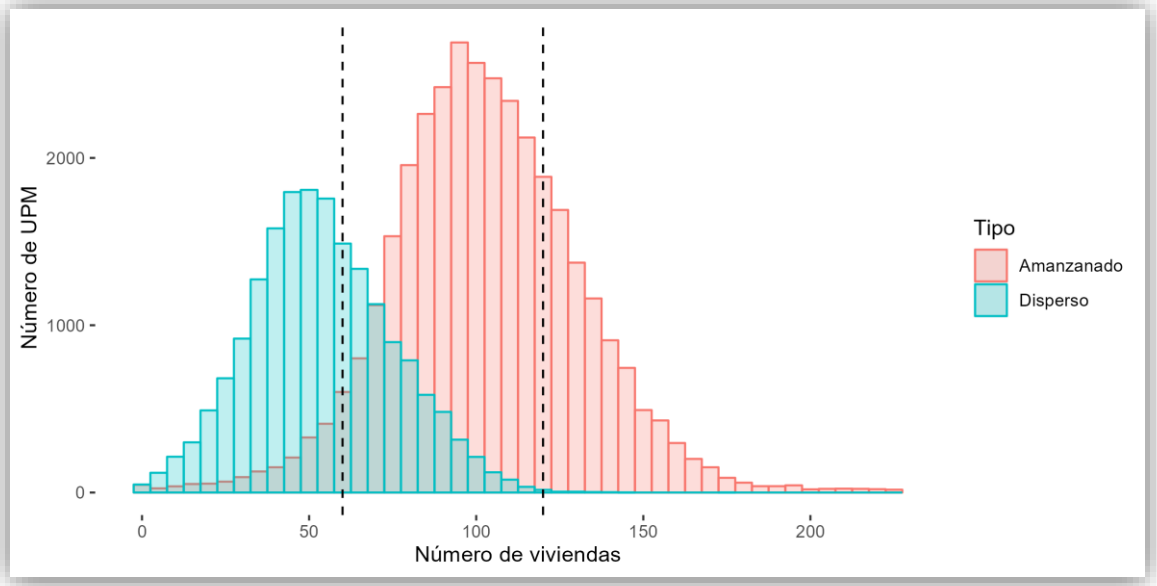


Unidad Primaria de Muestreo (UPM)

Sectores censales

- Son delimitaciones geográficas construidas con fines operativas para distribución de carga en el censo.

- Se utilizaban como unidades primarias de muestreo hasta el año 2017.



Menos de 60
13.436
(25,4%)

+

Entre 60 y 120
30.637
(57,8%)

+

Más de 120
8.897
(16,8%)

=

52.960
Sectores censales

www.ecuadorencifras.gob.ec

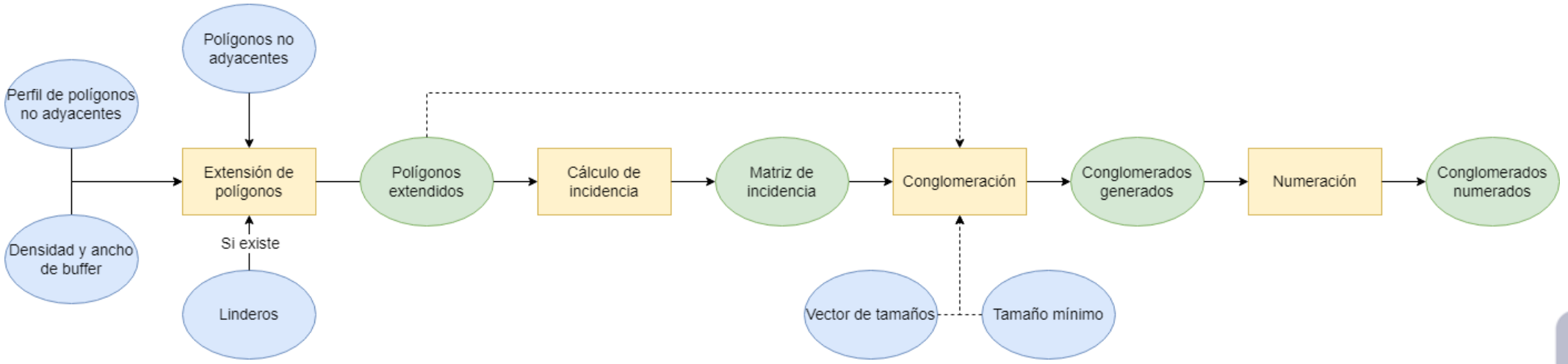
Buenas cifras,
mejores vidas

02.

psuR

psuR

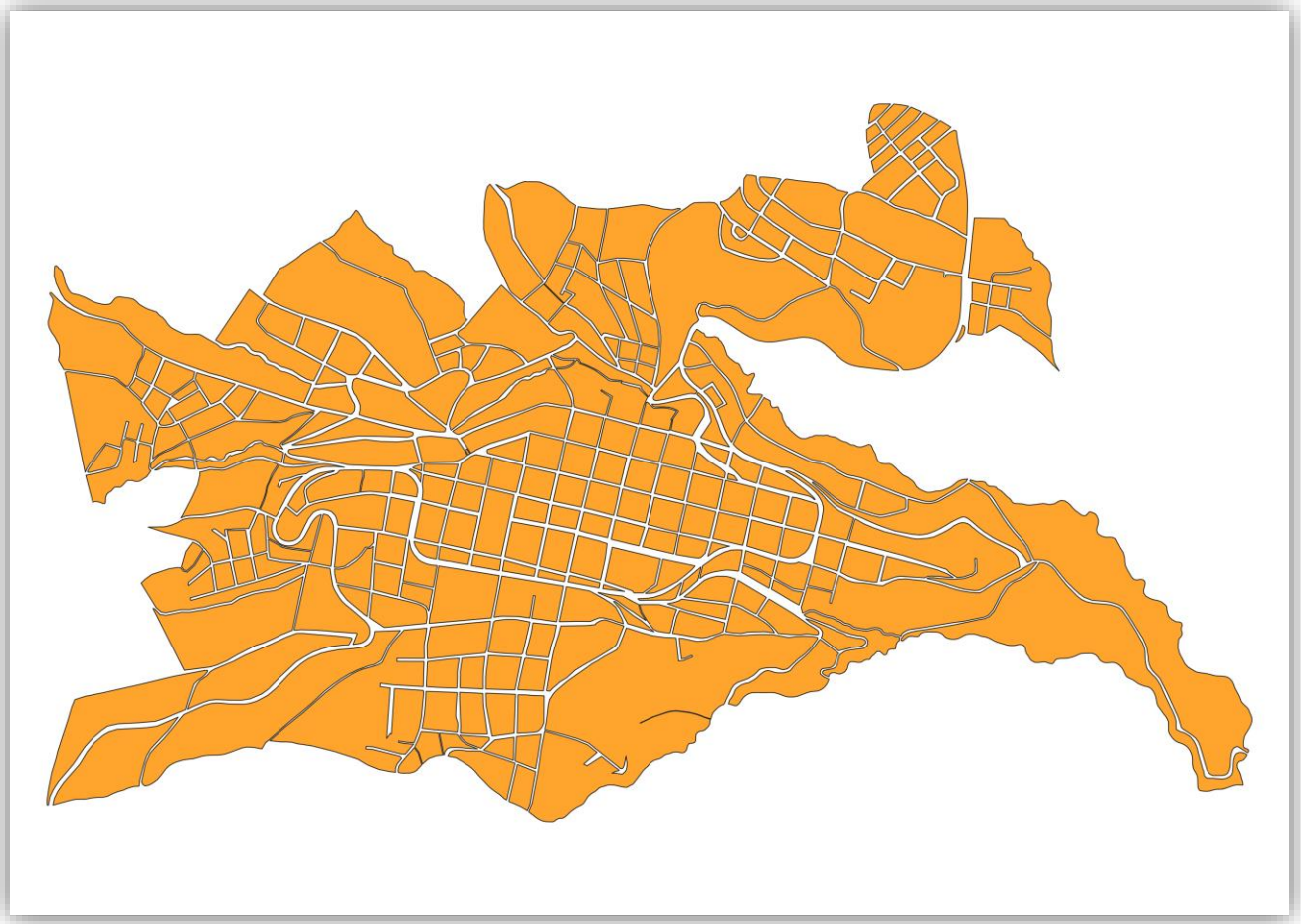
El algoritmo se desarrolló en la Dirección de Infraestructura Estadística y Muestreo y, gracias al apoyo de la CEPAL, se logró crear el paquete en R que abarca todas las funciones que conforman el algoritmo.

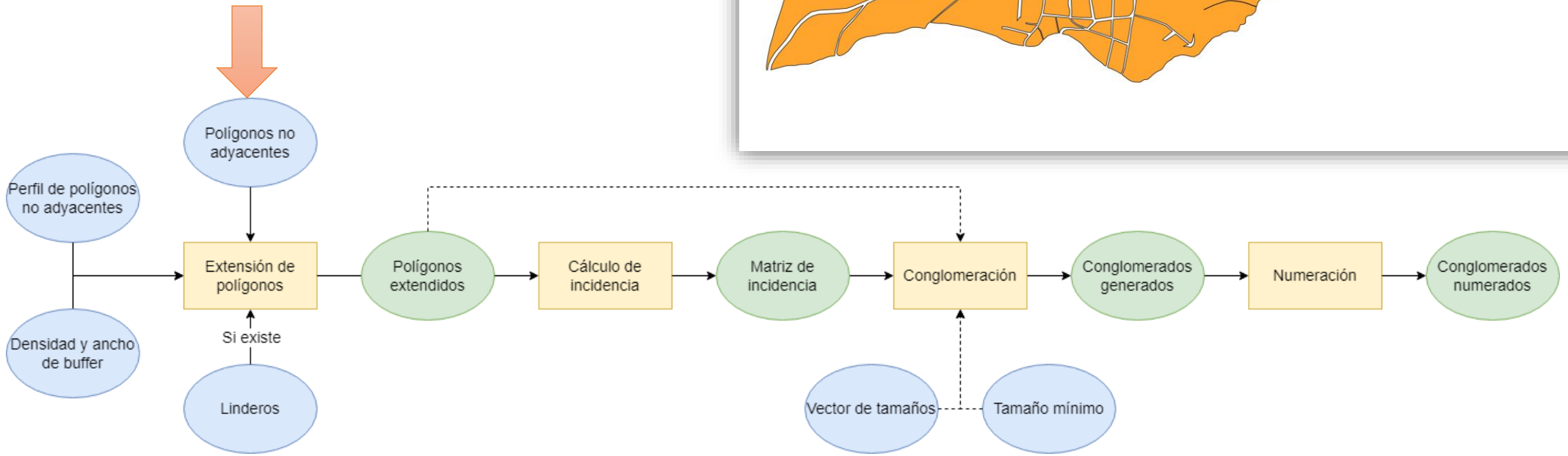


www.ecuadorencifras.gob.ec

psuR

Polígonos no adyacentes generalmente hace referencia a manzanas.

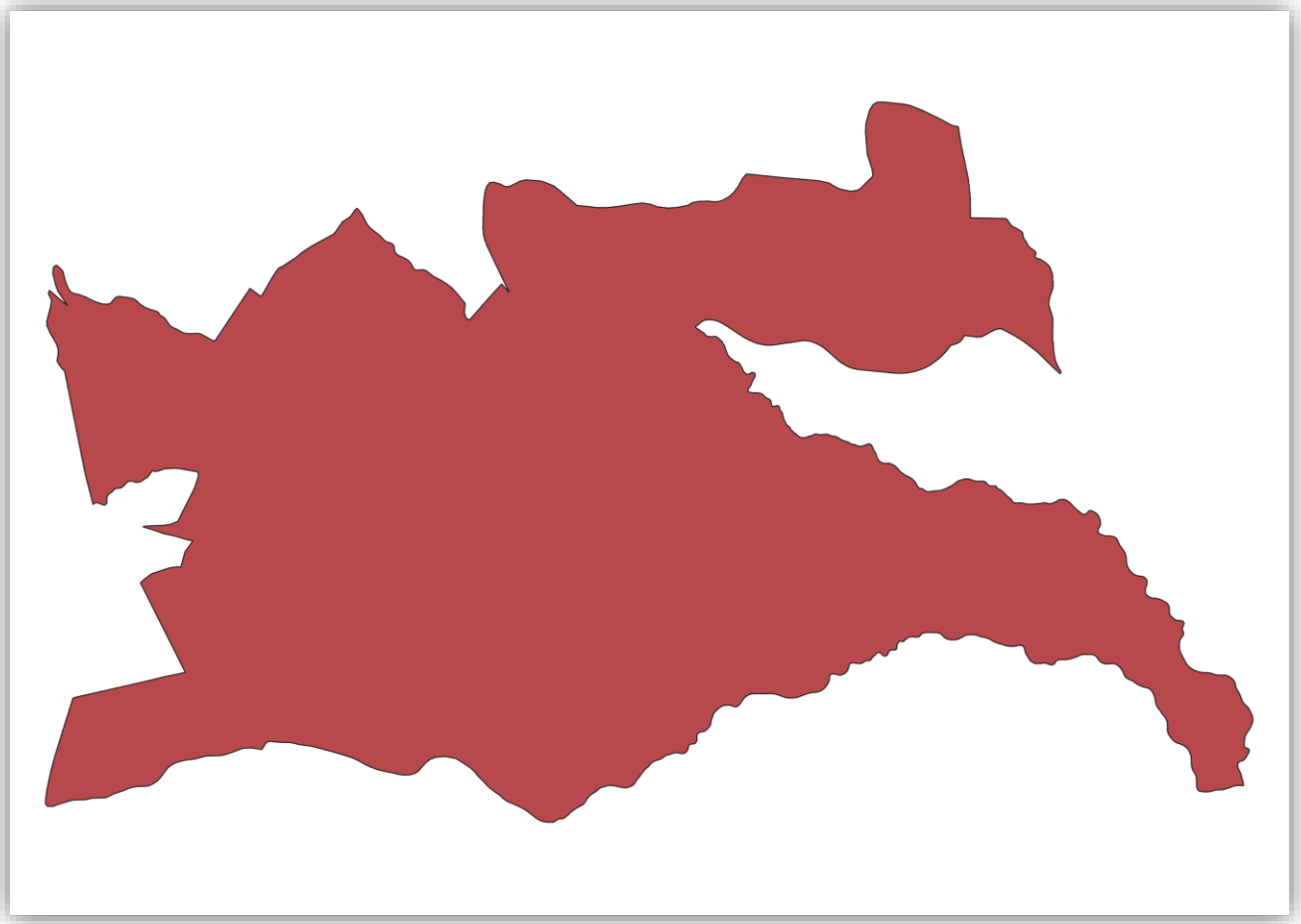


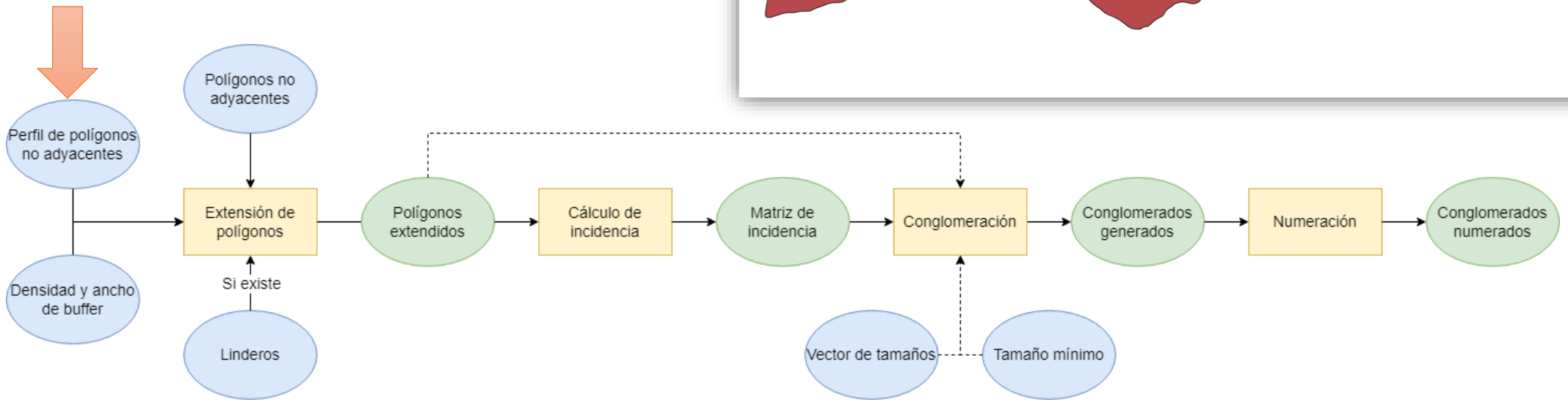


www.ecuadorencifras.gob.ec

psuR

Es un polígono que delimita la zona de extensión a trabajar.

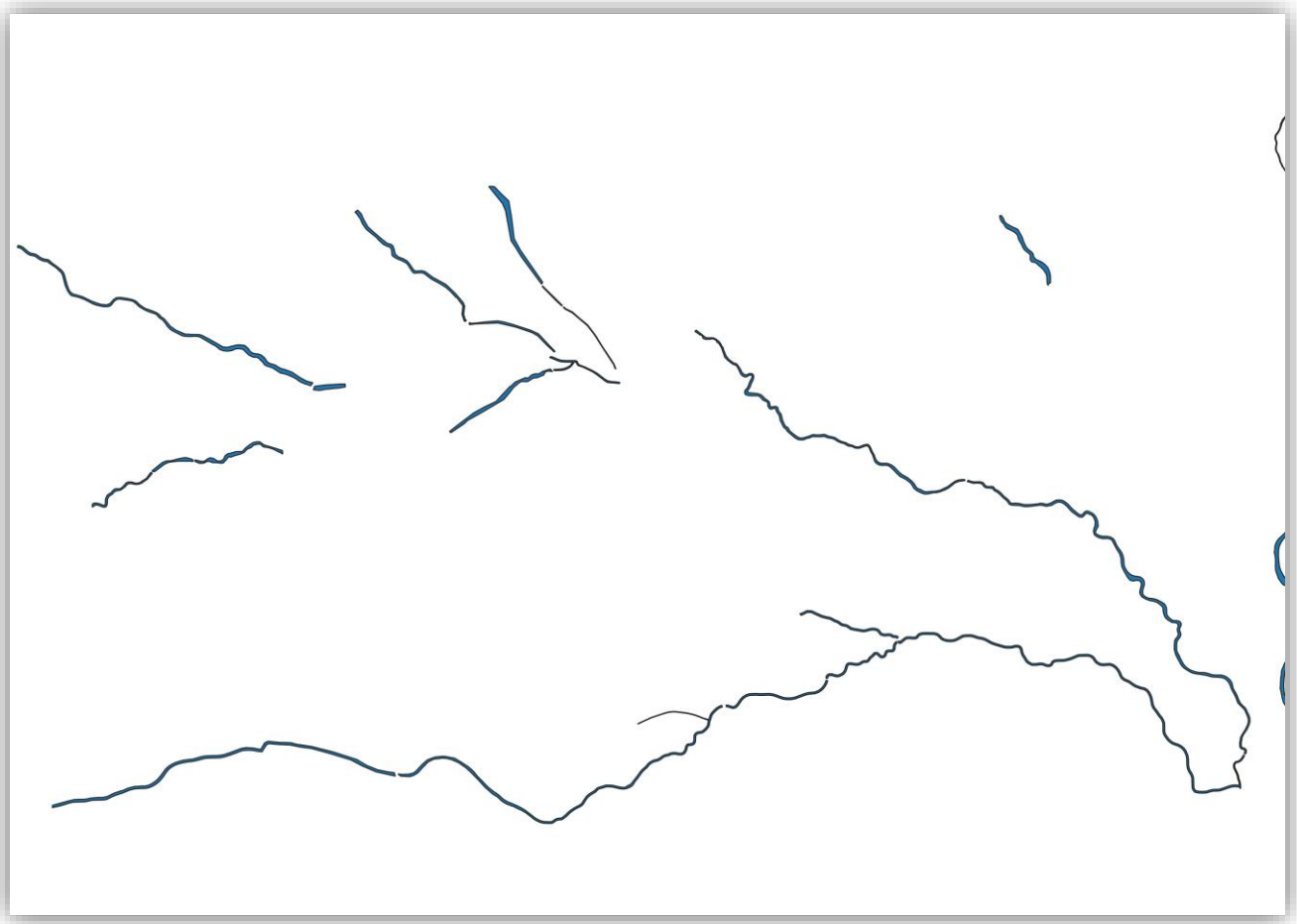


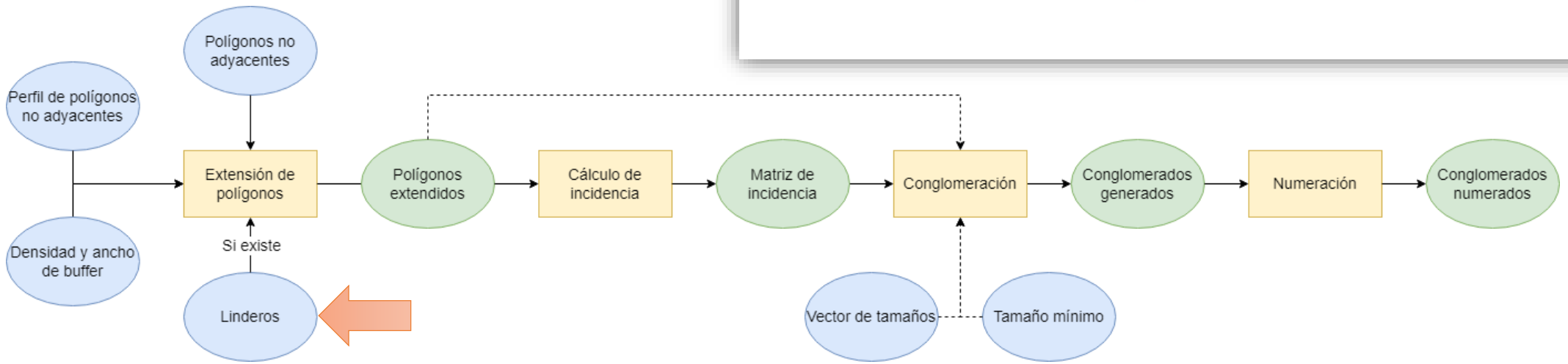


www.ecuadorencifras.gob.ec

psuR

Es un multipolígono que representa los ríos o linderos, sirve para evitar uniones entre manzanas separadas por los mismos, pueden o no estar contenidos en el perfil.



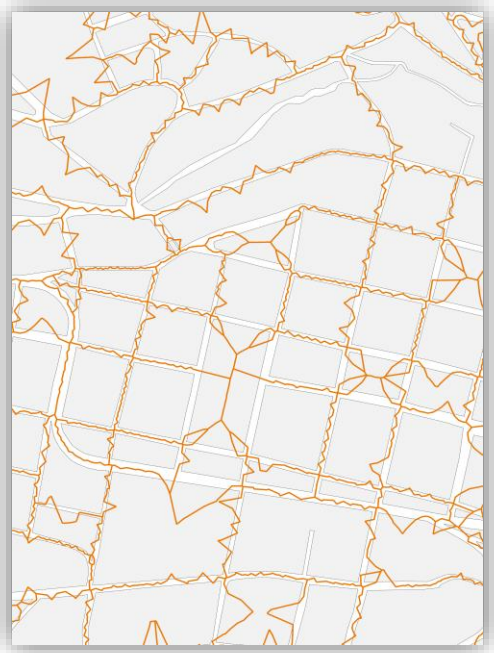


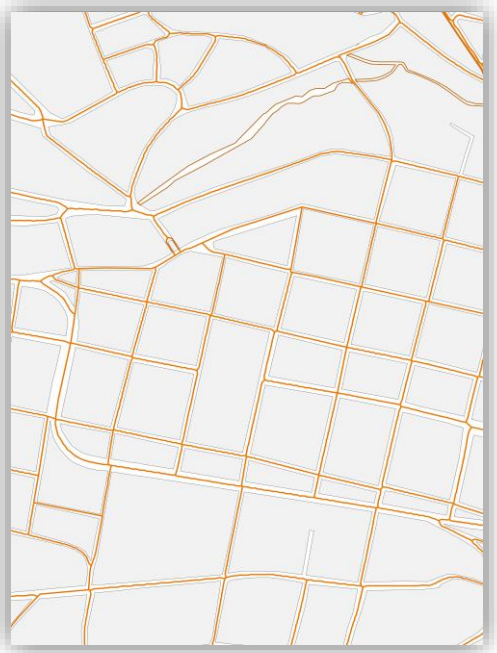
www.ecuadorencifras.gob.ec

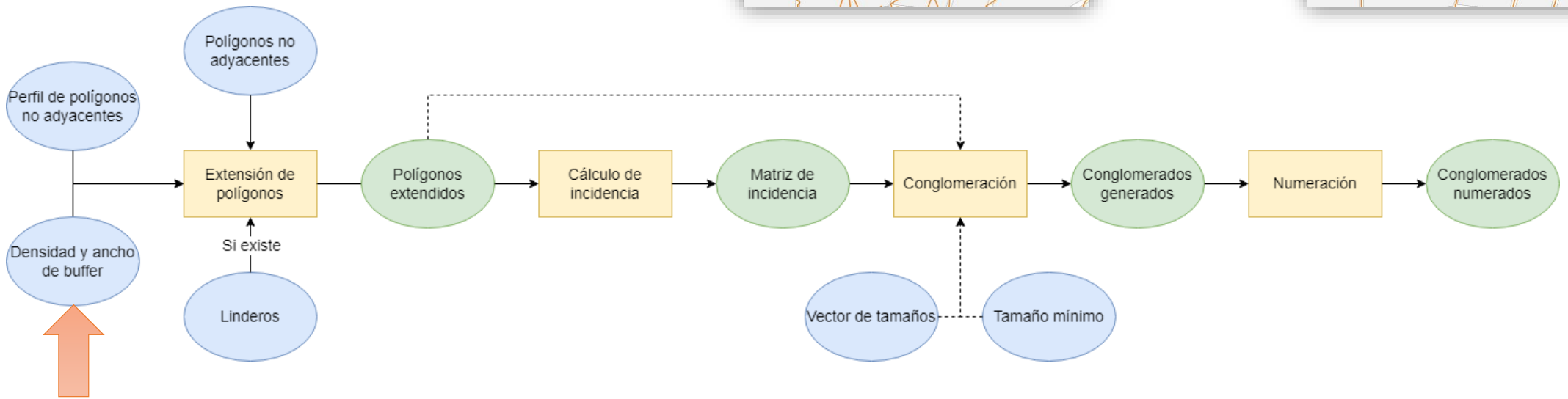
psuR

La densidad y el ancho del buffer, son valores escalares que determinan la calidad de la extensión del polígono.

A mayor calidad mayor cantidad de procesamiento requerida.



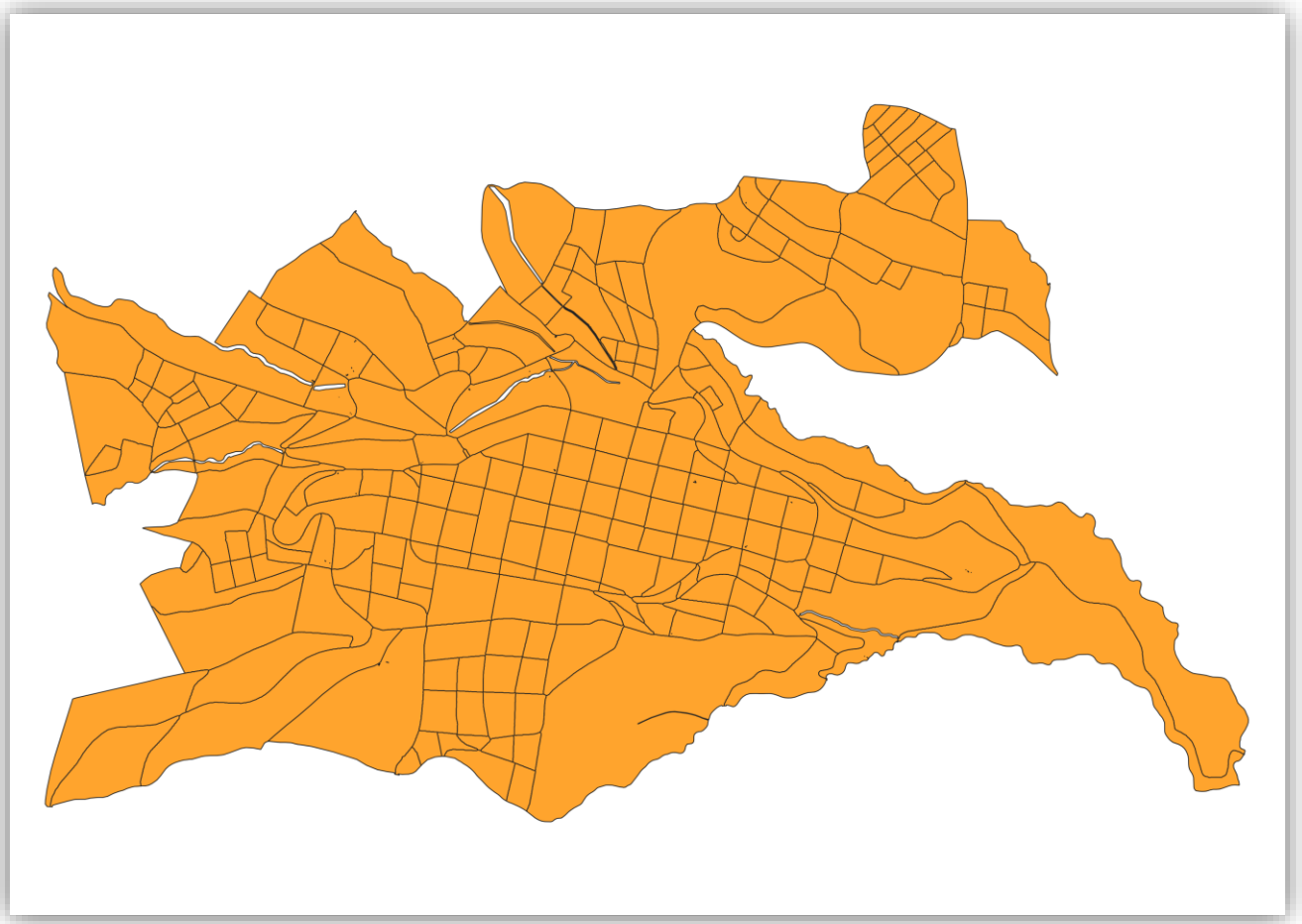


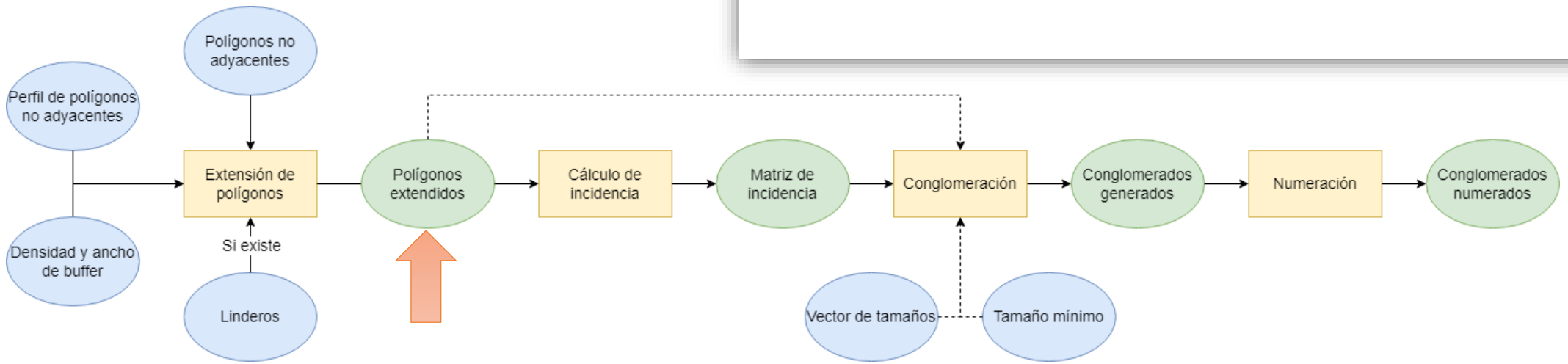


www.ecuadorencifras.gob.ec

psuR

Al extender las manzanas se puede obtener su relación utilizando herramientas geográficas.

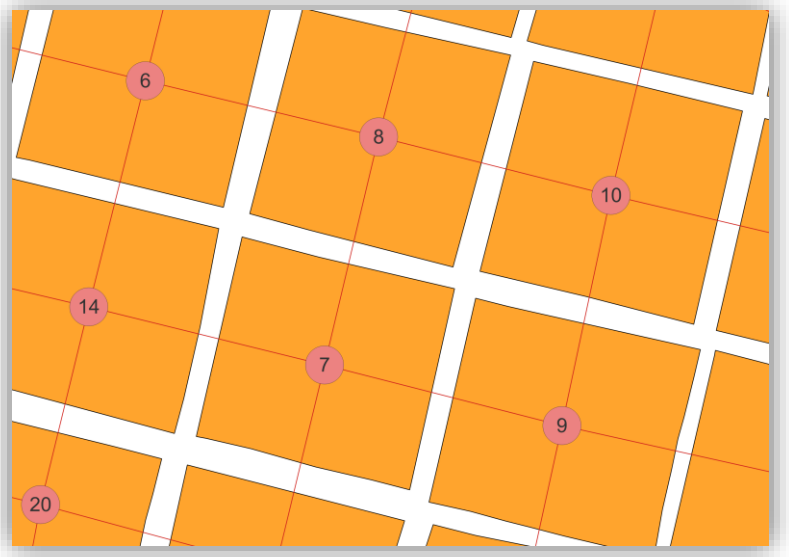




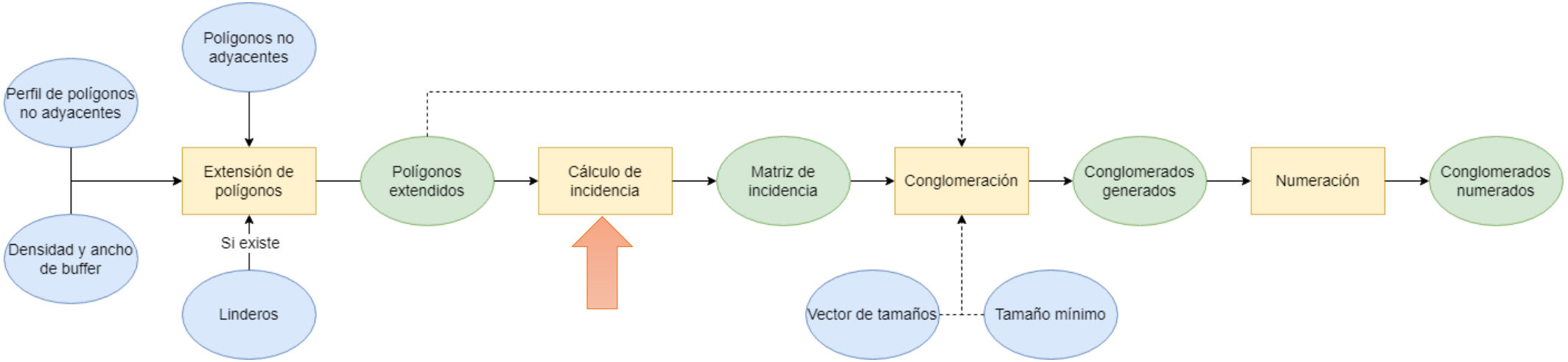
www.ecuadorencifras.gob.ec

psuR

La relación binaria que representa la matriz de incidencia es "polígono A comparte frontera con el polígono B".



	6	7	8	9	10	14	20
6	0	0	1	0	0	1	0
7	0	0	1	0	0	1	0
8	1	1	0	0	1	0	0
9	0	0	0	0	1	0	0
10	0	0	1	1	0	0	0
14	1	1	0	0	0	0	1
20	0	0	0	0	0	1	0



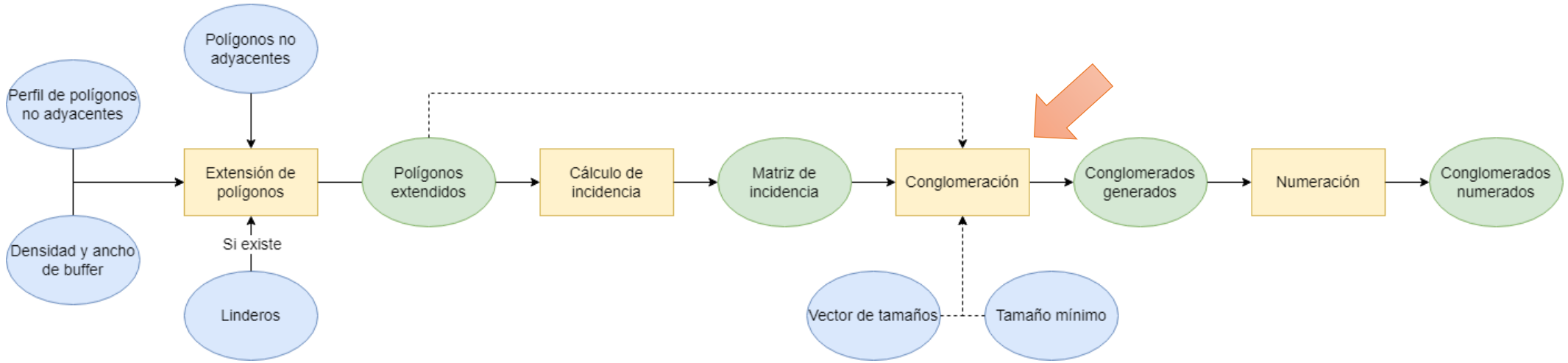
www.ecuadorencifras.gob.ec

psuR

El algoritmo de conglomeración recibe como insumos una matriz M_{nxn}, y un vector de tamaño n dando como resultado un vector de tamaño n que identifica a las upm generadas.

	6	7	8	9	10	14	20
6	0	0	1	0	0	1	0
7	0	0	1	0	0	1	0
8	1	1	0	0	1	0	0
9	0	0	0	0	1	0	0
10	0	0	1	1	0	0	0
14	1	1	0	0	0	0	1
20	0	0	0	0	0	1	0

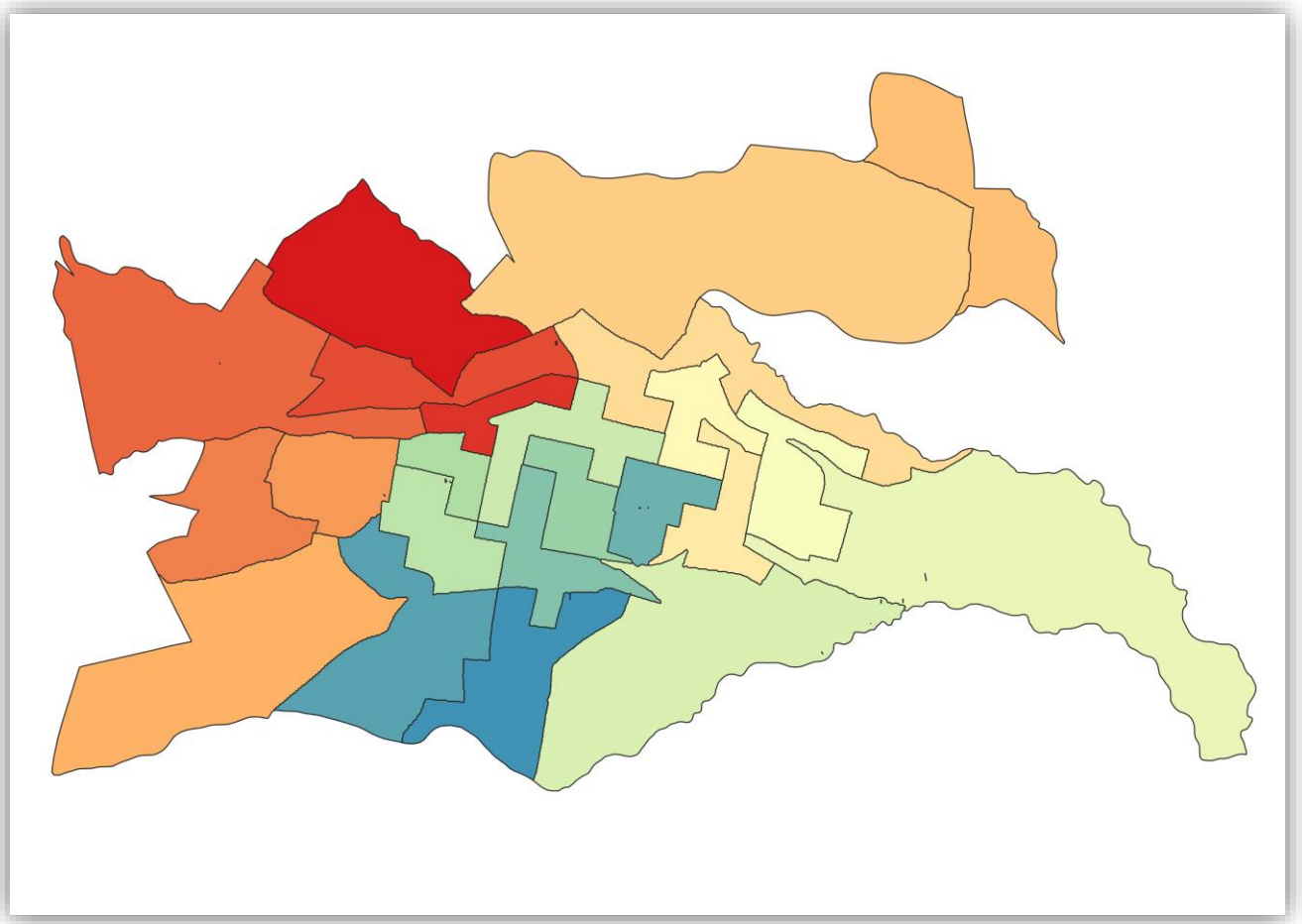
6	35
7	64
8	0
9	12
10	0
14	48
20	29

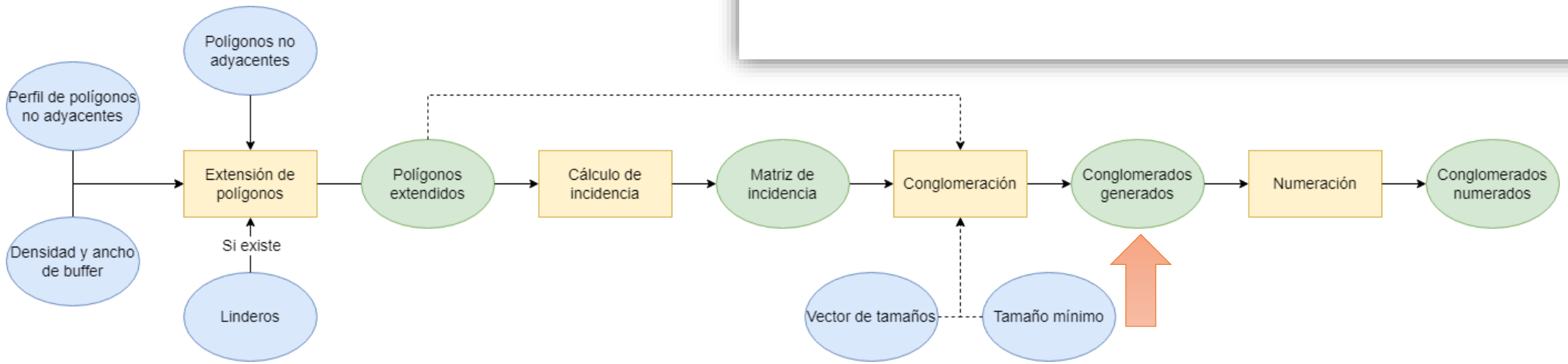


www.ecuadorencifras.gob.ec

psuR

A partir de las manzanas extendidas se obtiene de forma automática la delimitación geográfica de las UPM creadas.

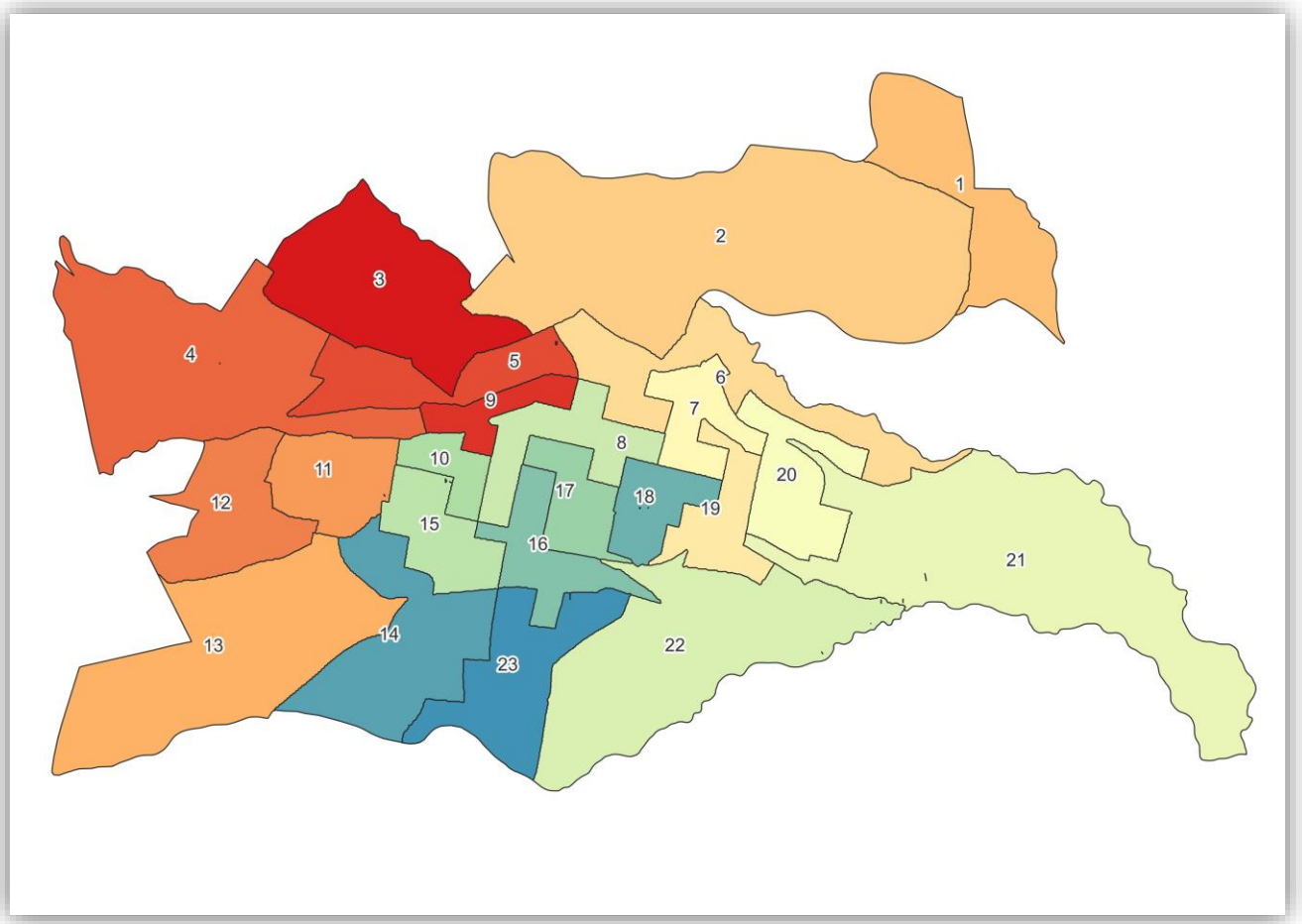


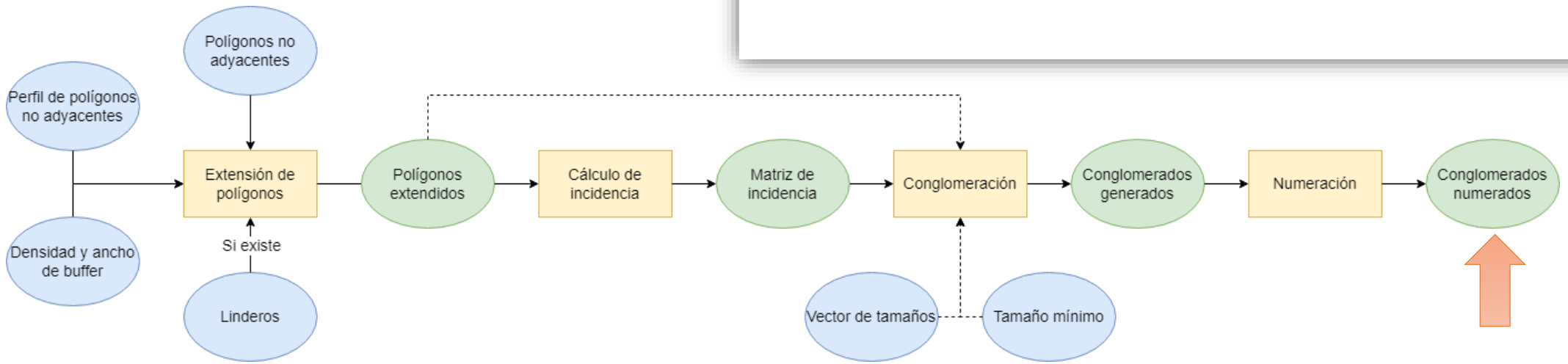


www.ecuadorencifras.gob.ec

psuR

El algoritmo numera a las UPM en forma serpenteante empezando por aquella que está más al norte.





www.ecuadorencifras.gob.ec

Buenas cifras,
mejores vidas

04.

Resultados

Resultados

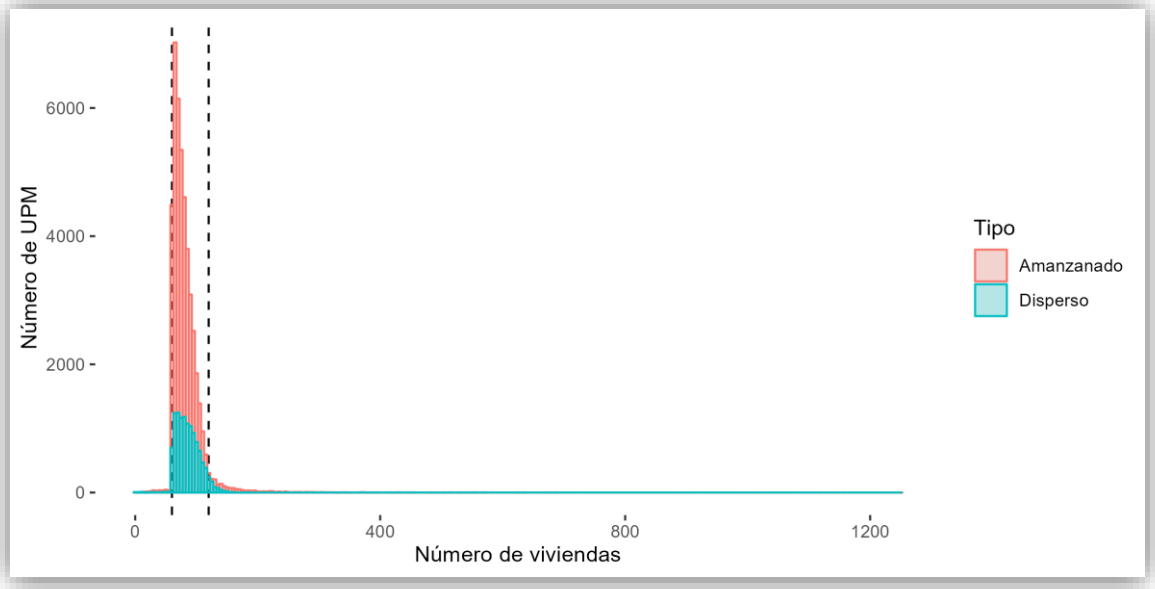
Conglomerados generados

www.ecuadorencifras.gob.ec

Unidad Primaria de Muestreo (UPM)

Resultado del algoritmo

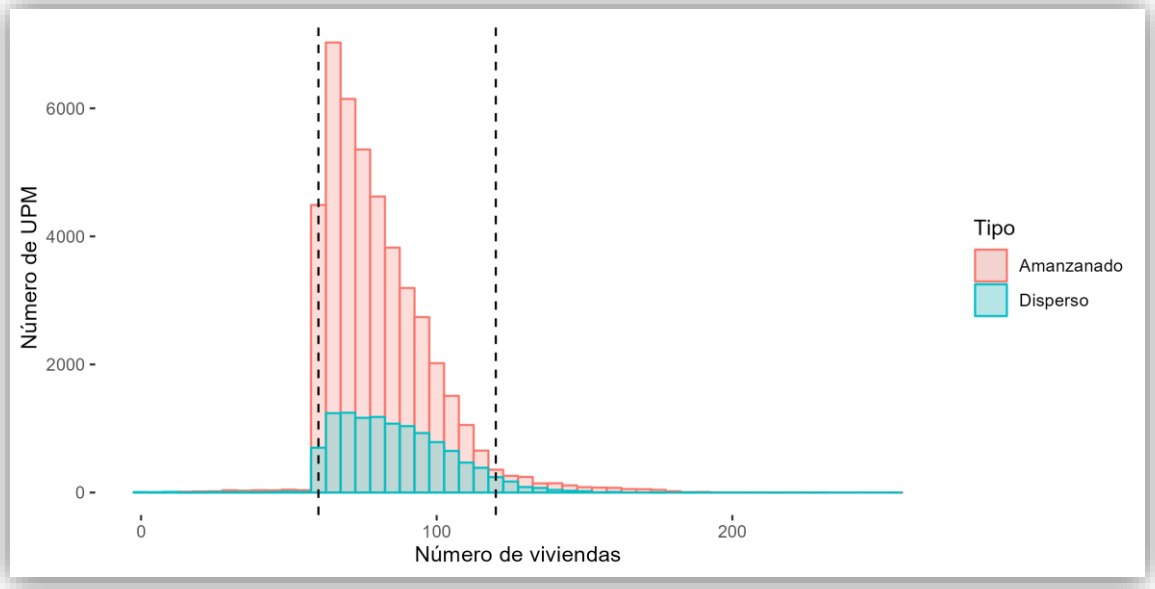
- Son delimitaciones geográficas que cumplen las mismas características geográficas que los sectores censales.
- Se utilizan como unidades primarias de muestreo desde el año 2017.



Unidad Primaria de Muestreo (UPM)

Resultado del algoritmo

- Son delimitaciones geográficas que cumplen las mismas características geográficas que los sectores censales.
- Se utilizan como unidades primarias de muestreo desde el año 2017.

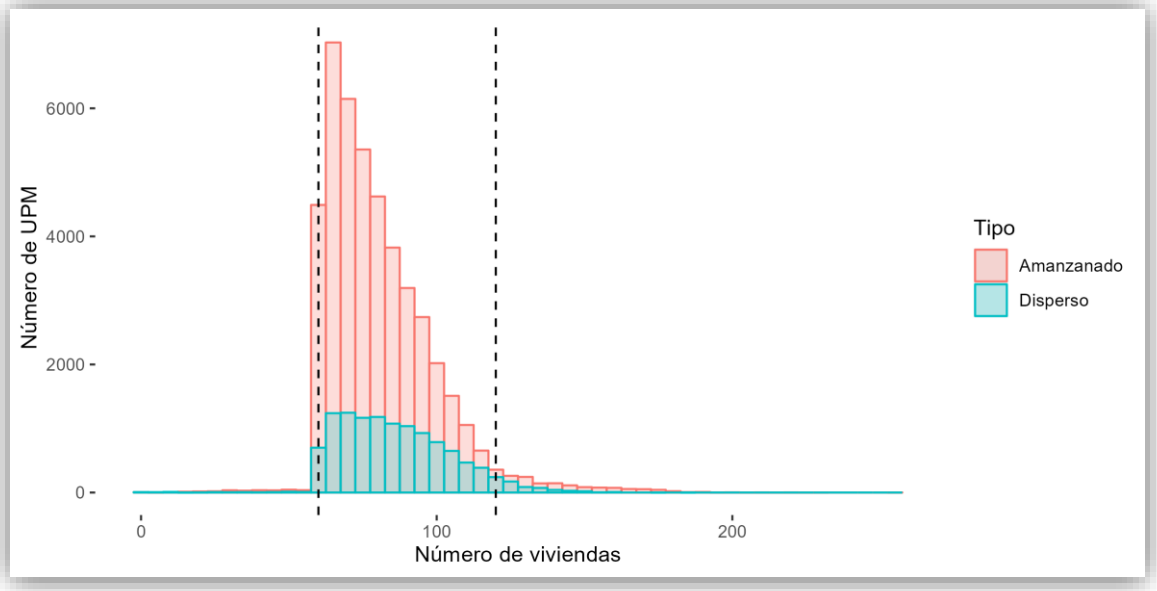


www.ecuadorencifras.gob.ec

Unidad Primaria de Muestreo (UPM)

Resultado del algoritmo

- Son delimitaciones geográficas que cumplen las mismas características geográficas que los sectores censales.
- Se utilizan como unidades primarias de muestreo desde el año 2017.



Menos de 60
377
(0,7%)

+

Entre 60 y 120
53.874
(95,8%)

+

Más de 120
1.988
(3,5%)

=

56.239
UPM

www.ecuadorencifras.gob.ec

Buenas cifras,
mejores vidas

03.

Formulación del problema

¿Qué es Programación Entera?

Buenas cifras, mejores vidas

Los modelos de programación entera son una extensión de los modelos lineales en los que al menos una variable asociada toma valores restringidos a los números enteros.

Un pequeño ejemplo:

¿Cuántos aguacates puede comprar Juanito con 1 USD, si cada uno cuesta 0,40 ctvs.?

Programación lineal
X = número de aguacates D = total de dinero P = precio unitario
$X = {}^{D}/_{P}$ $= {}^{1}/_{0,40}$ $= 2,5$

Programación entera

www.ecuadorencifras.gob.ec

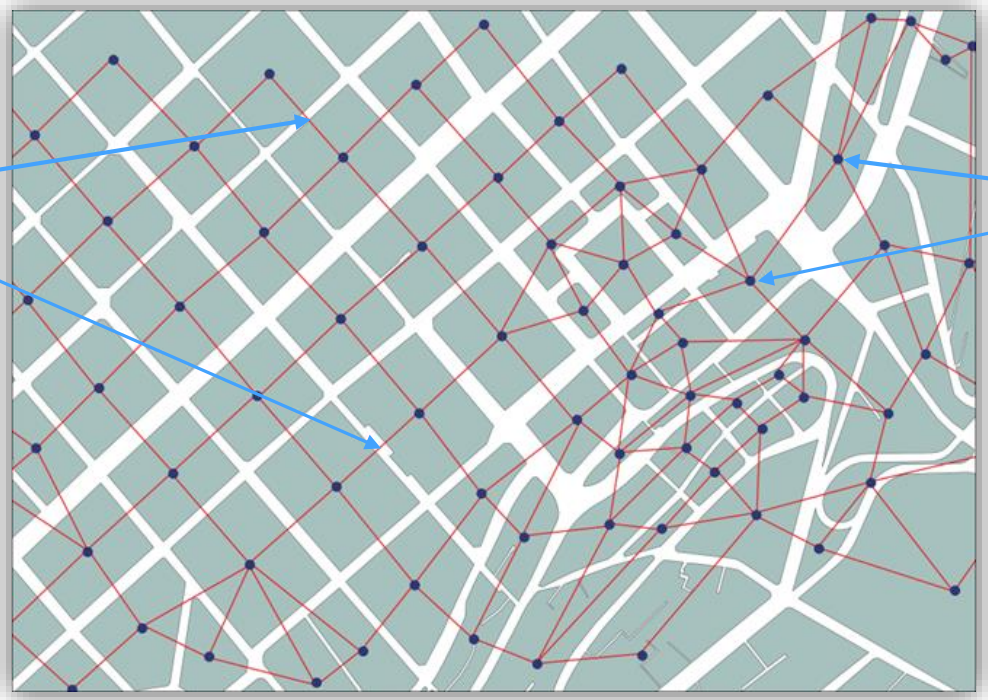
De la Geografía a la Matemática

Nodos (manzanas) y aristas (frontera en común)

En Matemática, un grafo es un conjunto de objetos llamados **nodos** unidos por enlaces llamados **aristas**, que permiten representar relaciones binarias entre elementos de un conjunto.

Para la generación de UPM, el grafo estaría determinado por:

Aristas:
Determinado por la existencia de una frontera en común entre dos manzanas.



Nodo:
Es el centroide de cada una de las manzanas en la zona de estudio

Formulación formal

Problema de partición de grafos

Sea $G = (V, E)$ un grafo conexo no dirigido, donde $V = \{1, 2, \dots, n\}$ es el conjunto de nodos y $E = \{e_{ij} = (i, j) : i, j \in V\}$ el de aristas. Sea $W = \{w_1, w_2, \dots, w_n\}$ el conjunto de pesos de los nodos. Sea S_L el límite inferior. La formulación general del problema de partición de grafos con restricción de peso es:

$$z = \min \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} \qquad (1)$$

$$s.a$$

$$S_L \leq \sum_{i=1}^{n} w_i x_{ij} \qquad \forall j \in V \qquad (2)$$

$$x_{ij} = 1 \quad si\ y\ solo\ si \quad \exists i_1, i_2, \dots, i_p\ tales\ que\ e_{ii_1} \cdot x_{ii_1} = e_{i_1 i_2} \cdot x_{i_1 i_2} = \cdots = e_{i_p j} \cdot x_{i_p j} = 1 \qquad (3)$$

$$x_{ij} \in \{0,1\} \qquad \forall i, j \in V \qquad (4)$$

www.ecuadorencifras.gob.ec

Formulación formal

Problema de partición de grafos

Sea $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ un grafo conexo no dirigido, donde $\mathrm{V} = \{1, 2, \ldots, \mathrm{n}\}$ es el conjunto de nodos y $\mathrm{E} = \{e_{ij} = (i, \mathrm{j}) : \mathrm{i}, \mathrm{j} \in V\}$ el de aristas. Sea $\mathrm{W} = \{w_1, w_2, \ldots, w_n\}$ el conjunto de pesos de los nodos. Sea S_L el límite inferior. La formulación general del problema de partición de grafos con restricción de peso es:

$$z = \min \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} \qquad (1)$$

$$s.a$$

$$S_L \leq \sum_{i=1}^{n} w_i x_{ij} \qquad \forall j \in V \qquad (2)$$

$$x_{ij} = 1 \quad si\ y\ solo\ si \quad \exists i_1, i_2, \ldots, i_p\ tales\ que\ e_{ii_1} \cdot x_{ii_1} = e_{i_1 i_2} \cdot x_{i_1 i_2} = \cdots = e_{i_p j} \cdot x_{i_p j} = 1 \qquad (3)$$

$$x_{ij} \in \{0,1\} \qquad \forall i, j \in V \qquad (4)$$

Función objetivo: Minimizar el número de uniones maximizando el número de UPM.

www.ecuadorencifras.gob.ec

Formulación formal

Problema de partición de grafos

Sea $\mathrm{G} = (\mathrm{V}, \mathrm{E})$ un grafo conexo no dirigido, donde $\mathrm{V} = \{1, 2, \ldots, \mathrm{n}\}$ es el conjunto de nodos y $\mathrm{E} = \{e_{ij} = (i, \mathrm{j}) : \mathrm{i}, \mathrm{j} \in V\}$ el de aristas. Sea $\mathrm{W} = \{w_1, w_2, \ldots, w_n\}$ el conjunto de pesos de los nodos. Sea S_L el límite inferior. La formulación general del problema de partición de grafos con restricción de peso es:

$$z = \min \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} \qquad (1)$$

$$s.a$$

$$S_L \leq \sum_{i=1}^{n} w_i x_{ij} \qquad \forall j \in V \qquad (2)$$

$$x_{ij} = 1 \quad si\ y\ solo\ si \quad \exists i_1, i_2, \ldots, i_p\ tales\ que\ e_{ii_1} \cdot x_{ii_1} = e_{i_1 i_2} \cdot x_{i_1 i_2} = \cdots = e_{i_p j} \cdot x_{i_p j} = 1 \qquad (3)$$

$$x_{ij} \in \{0,1\} \qquad \forall i, j \in V \qquad (4)$$

Capacidad: Controla el tamaño de las UPM (mayor o igual a S_L).

www.ecuadorencifras.gob.ec

Formulación formal

Problema de partición de grafos

Sea $G = (V, E)$ un grafo conexo no dirigido, donde $V = \{1, 2, \dots, n\}$ es el conjunto de nodos y $E = \{e_{ij} = (i, j) : i, j \in V\}$ el de aristas. Sea $W = \{w_1, w_2, \dots, w_n\}$ el conjunto de pesos de los nodos. Sea S_L el límite inferior. La formulación general del problema de partición de grafos con restricción de peso es:

$$z = \min \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} \qquad (1)$$

$$s.a$$

$$S_L \leq \sum_{i=1}^{n} w_i x_{ij} \qquad \forall j \in V \qquad (2)$$

$$x_{ij} = 1 \quad si\ y\ solo\ si \quad \exists i_1, i_2, \dots, i_p\ tales\ que\ e_{ii_1} \cdot x_{ii_1} = e_{i_1 i_2} \cdot x_{i_1 i_2} = \cdots = e_{i_p j} \cdot x_{i_p j} = 1 \qquad (3)$$

$$x_{ij} \in \{0,1\} \qquad \forall i, j \in V \qquad (4)$$

Conexidad: Las UPM están formadas por polígonos que comparten frontera.

www.ecuadorencifras.gob.ec

Un ejemplo sobre Teselia

Un ejemplo sobre Teselia

Buenas cifras, mejores vidas

Planilandia, novela escrita por Edwin A. Abbott, es una aventura conmovedora de matemáticas puras, una fantasía de espacios extraños poblados por figuras geométricas que piensan y hablan y tienen todas las emociones humanas. No es ningún relato intrascendente de ciencia-ficción.

Teselia, ciudad ficticia inspirada por el universo de Planilandia, es una ciudad rectangular de 10x16 manzanas, sobre la cual se aplicará el algoritmo de conglomeración.

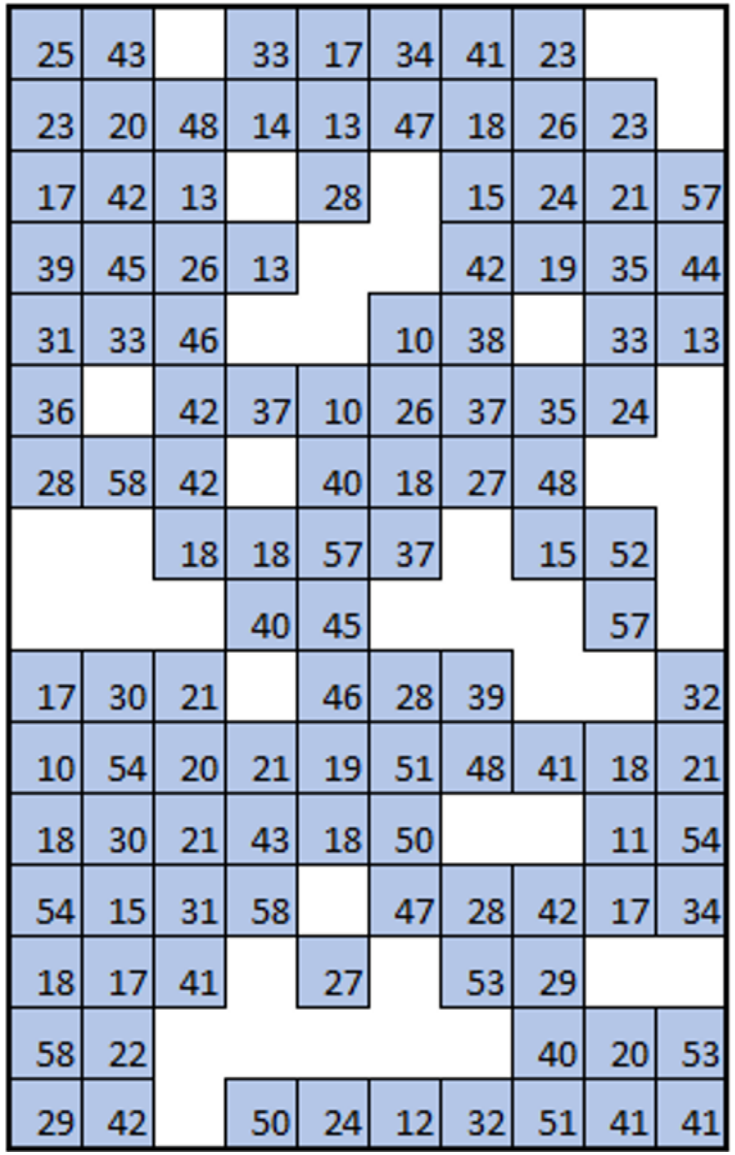
A continuación se presenta el desarrollo del mismo, paso a paso, junto con su resultado final, la generación de conglomerados que cumplen las restricciones de tamaño, peso y pertenencia.

25	43	68	33	17	34	41	23	65	69
23	20	48	14	13	47	18	26	23	64
17	42	13	67	28	65	15	24	21	57
39	45	26	13	66	68	42	19	35	44
31	33	46	63	63	10	38	78	33	13
36	78	42	37	10	26	37	35	24	66
28	58	42	67	40	18	27	48	68	79
71	77	18	18	57	37	74	15	52	76
61	71	79	40	45	63	74	61	57	67
17	30	21	75	46	28	39	78	61	32
10	54	20	21	19	51	48	41	18	21
18	30	21	43	18	50	60	60	11	54
54	15	31	58	69	47	28	42	17	34
18	17	41	80	27	69	53	29	64	65
58	22	71	64	75	79	74	40	20	53
29	42	78	50	24	12	32	51	41	41

Bienvenidos a Teselia!

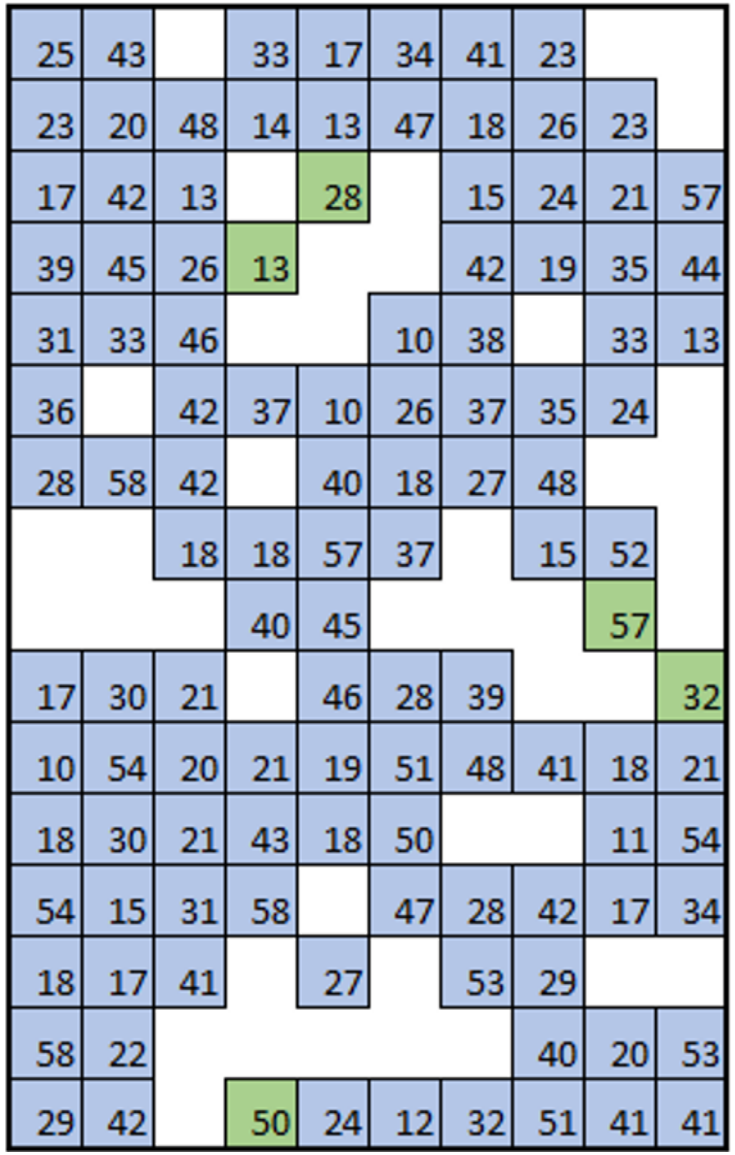
25	43	68	33	17	34	41	23	65	69
23	20	48	14	13	47	18	26	23	64
17	42	13	67	28	65	15	24	21	57
39	45	26	13	66	68	42	19	35	44
31	33	46	63	63	10	38	78	33	13
36	78	42	37	10	26	37	35	24	66
28	58	42	67	40	18	27	48	68	79
71	77	18	18	57	37	74	15	52	76
61	71	79	40	45	63	74	61	57	67
17	30	21	75	46	28	39	78	61	32
10	54	20	21	19	51	48	41	18	21
18	30	21	43	18	50	60	60	11	54
54	15	31	58	69	47	28	42	17	34
18	17	41	80	27	69	53	29	64	65
58	22	71	64	75	79	74	40	20	53
29	42	78	50	24	12	32	51	41	41

Al fijar un tamaño mínimo de 60, se identifican las manzanas que superan ese umbral.



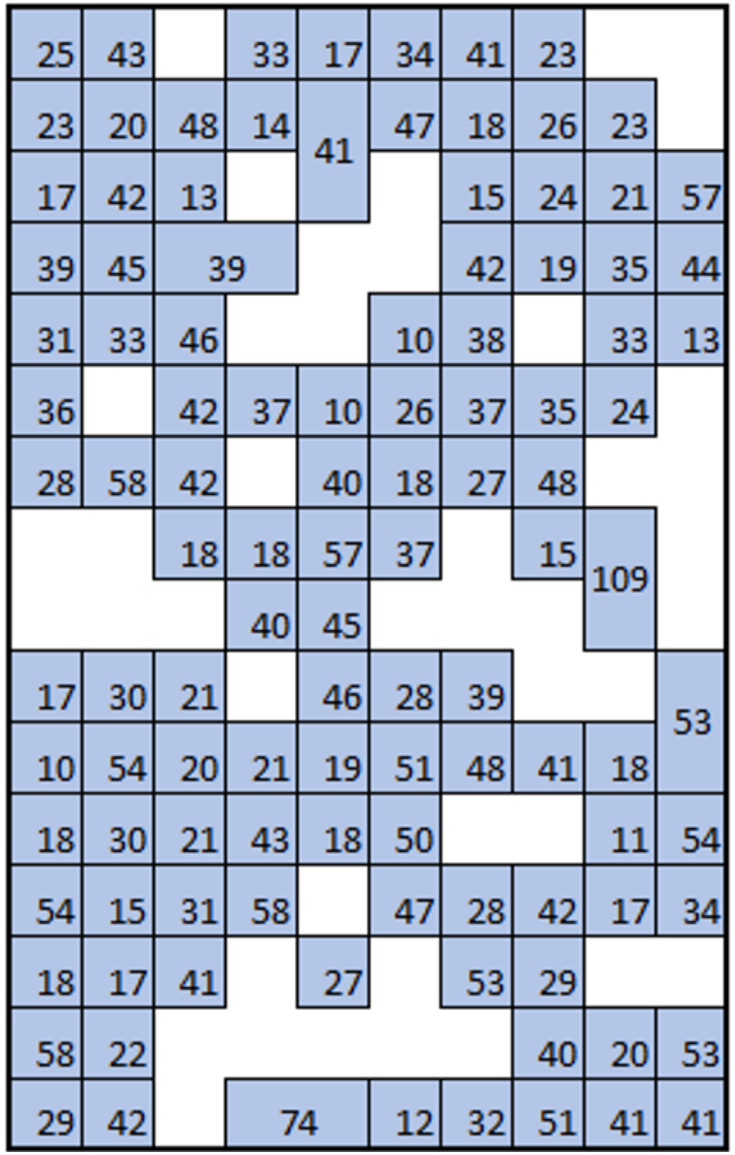
Y al ser conglomerados por si mismo… se excluyen!

www.ecuadorencifras.gob.ec



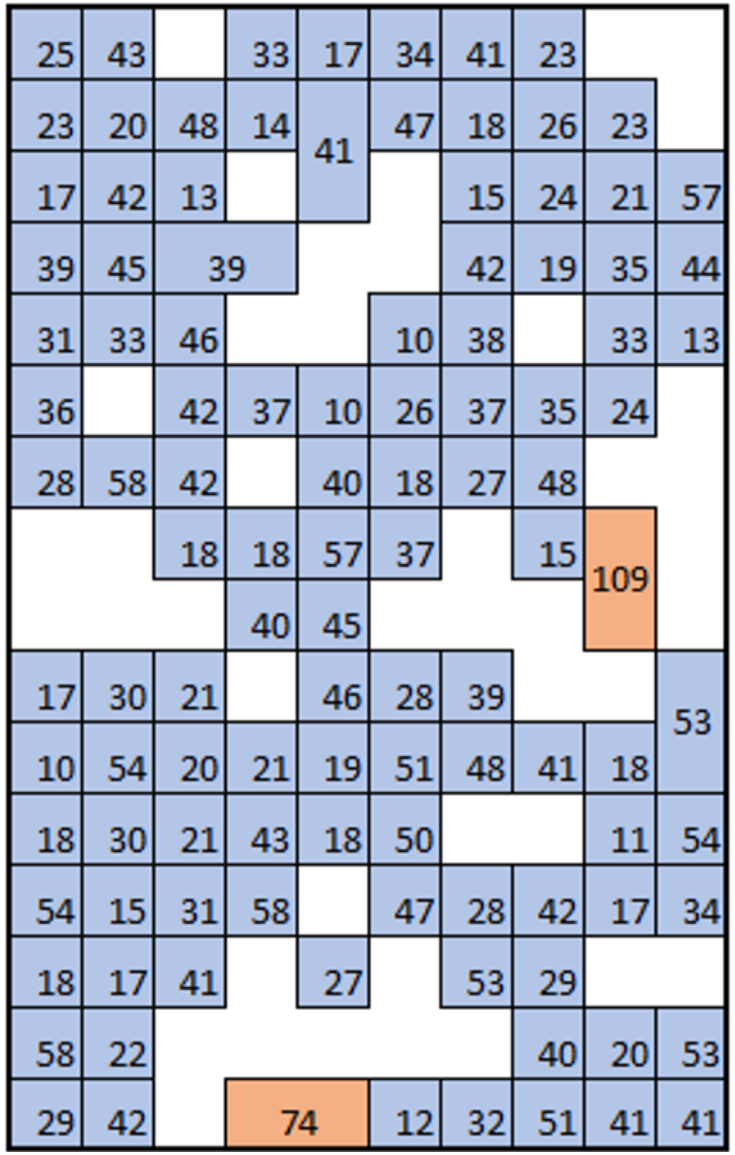
Luego, se determinan las manzanas que tienen el menor número de opciones (incidencias) para juntarse con otras.

www.ecuadorencifras.gob.ec



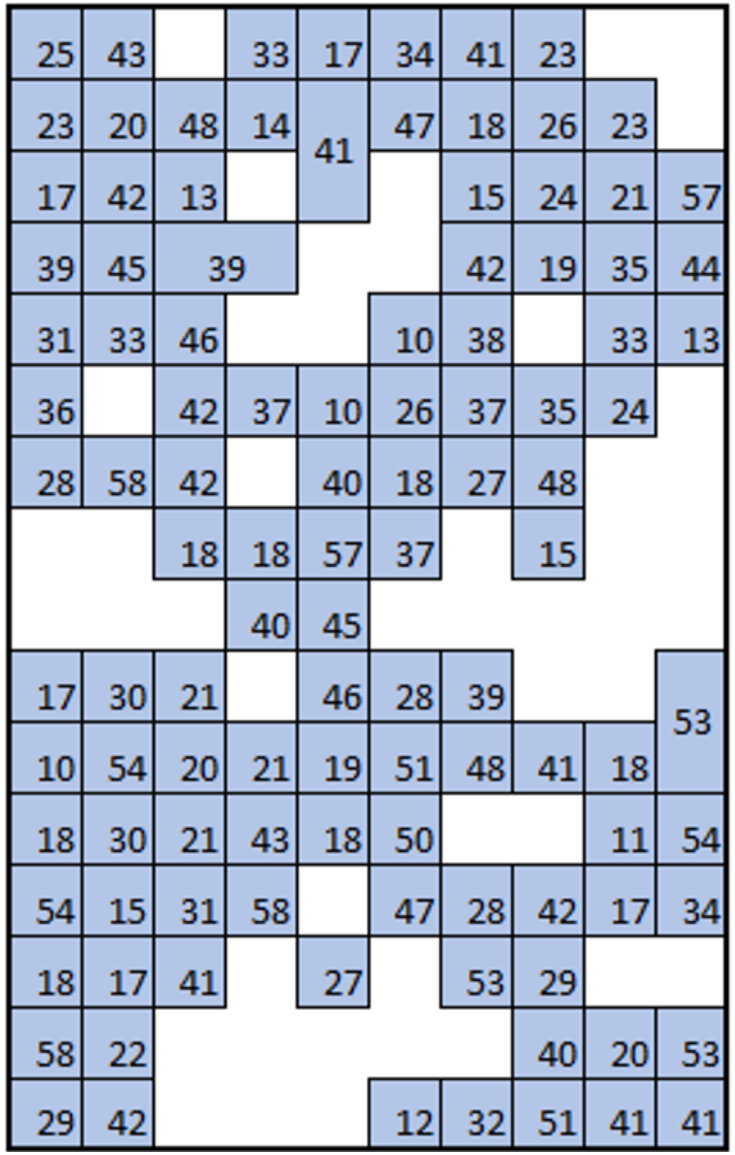
Y se juntan (tanto incidencias como peso), formando un solo conjunto.

www.ecuadorencifras.gob.ec



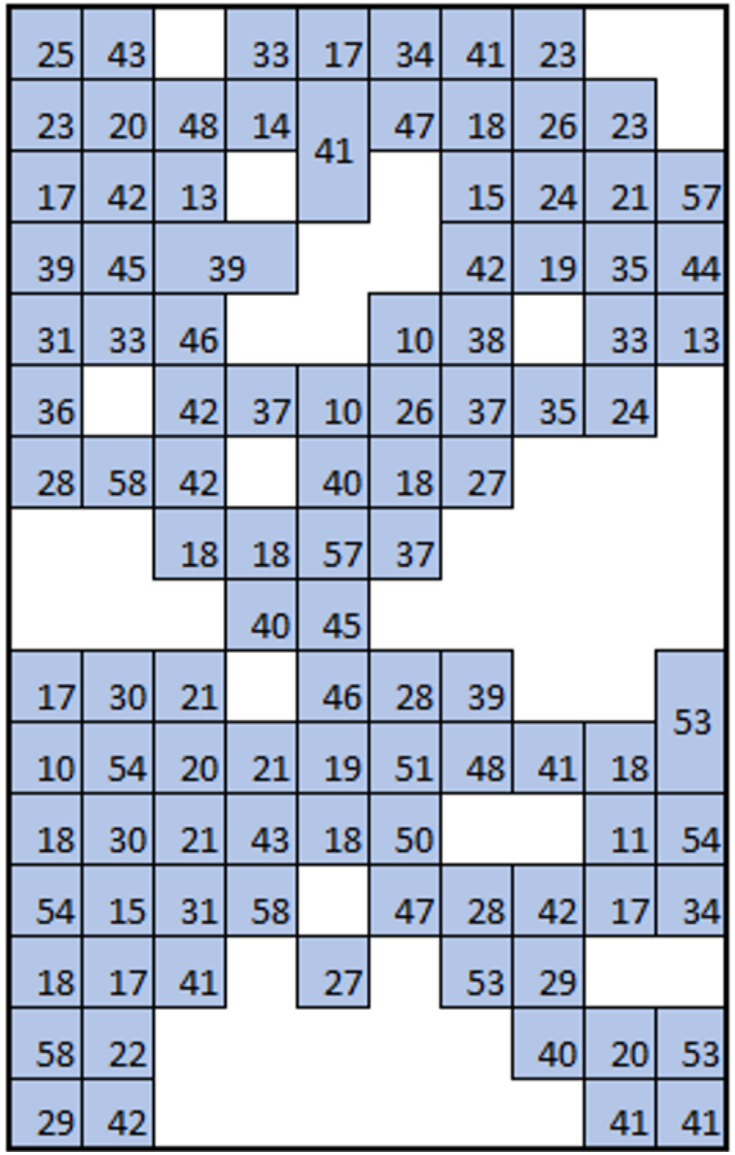
Y, si superan el límite de tamaño, conglomerados nuevos han sido generados.

www.ecuadorencifras.gob.ec



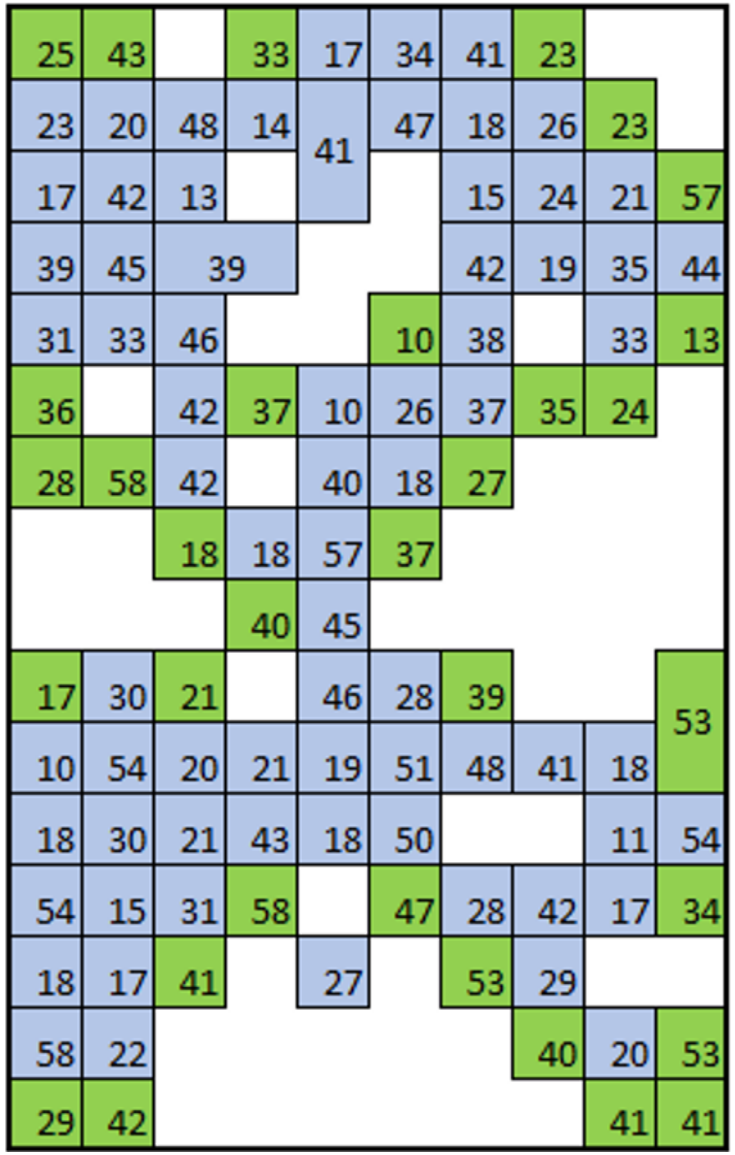
Y se excluyen!

www.ecuadorencifras.gob.ec



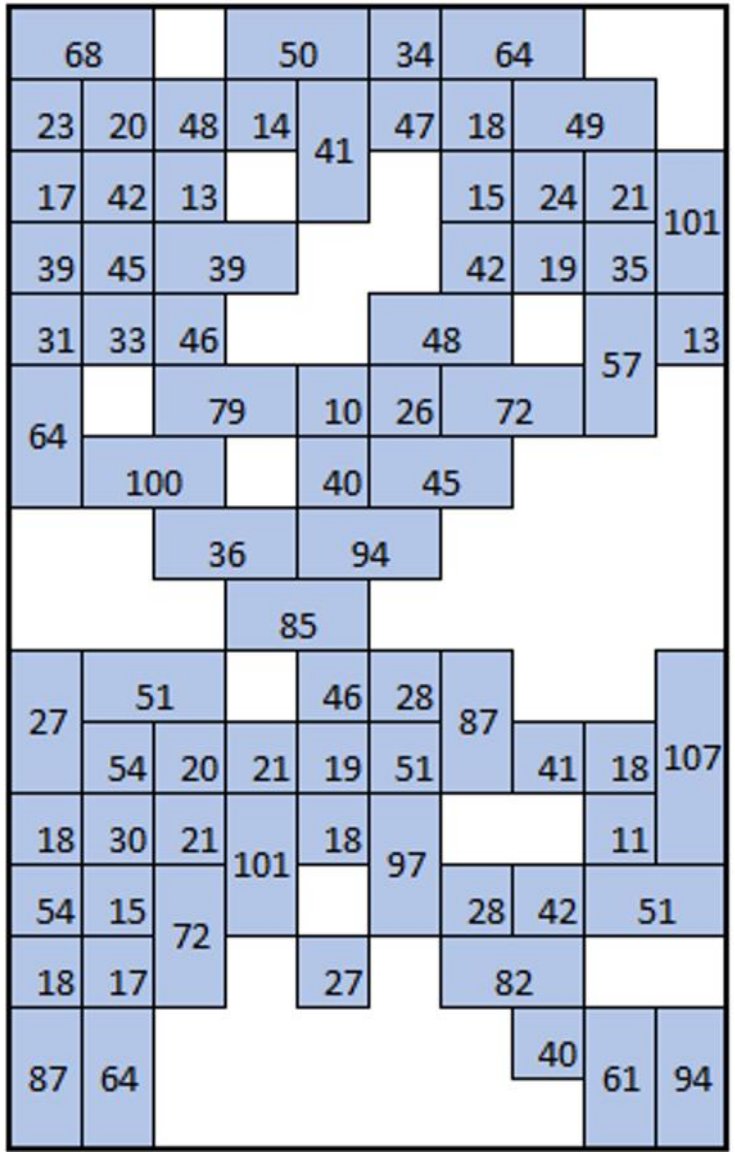
Y así, empezamos este proceso recursivo.

www.ecuadorencifras.gob.ec



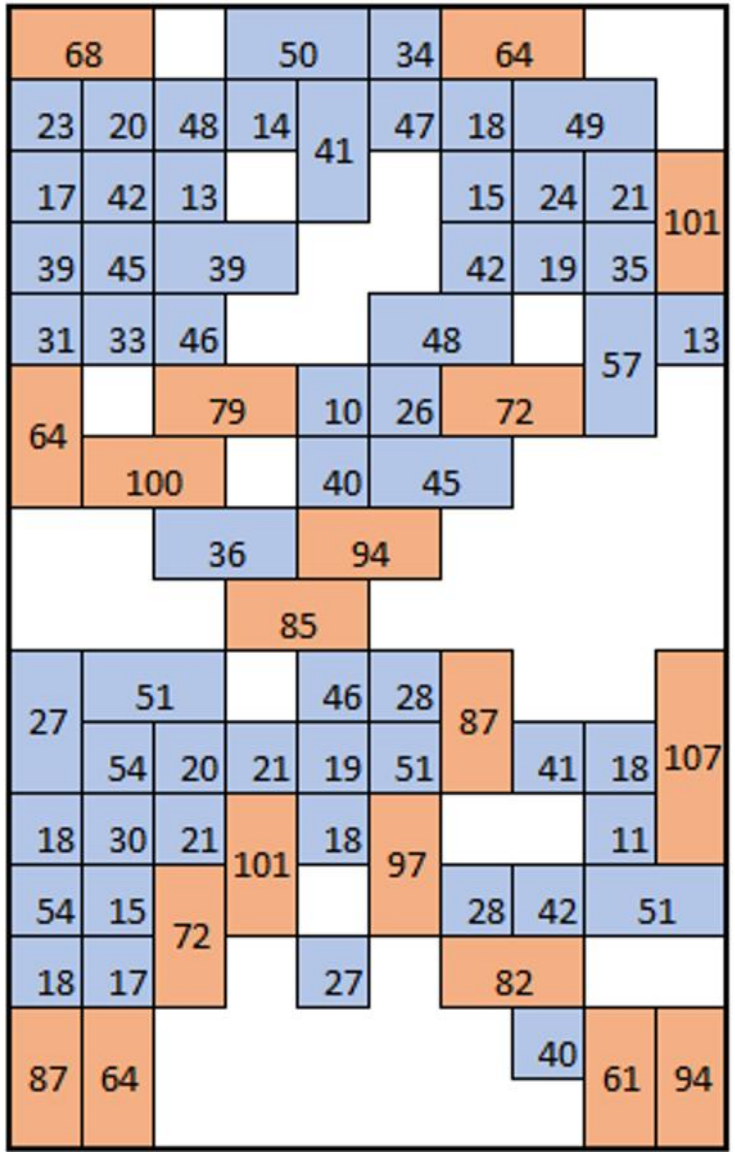
Identificando nuevamente las manzanas con el menor número de incidencias.

www.ecuadorencifras.gob.ec



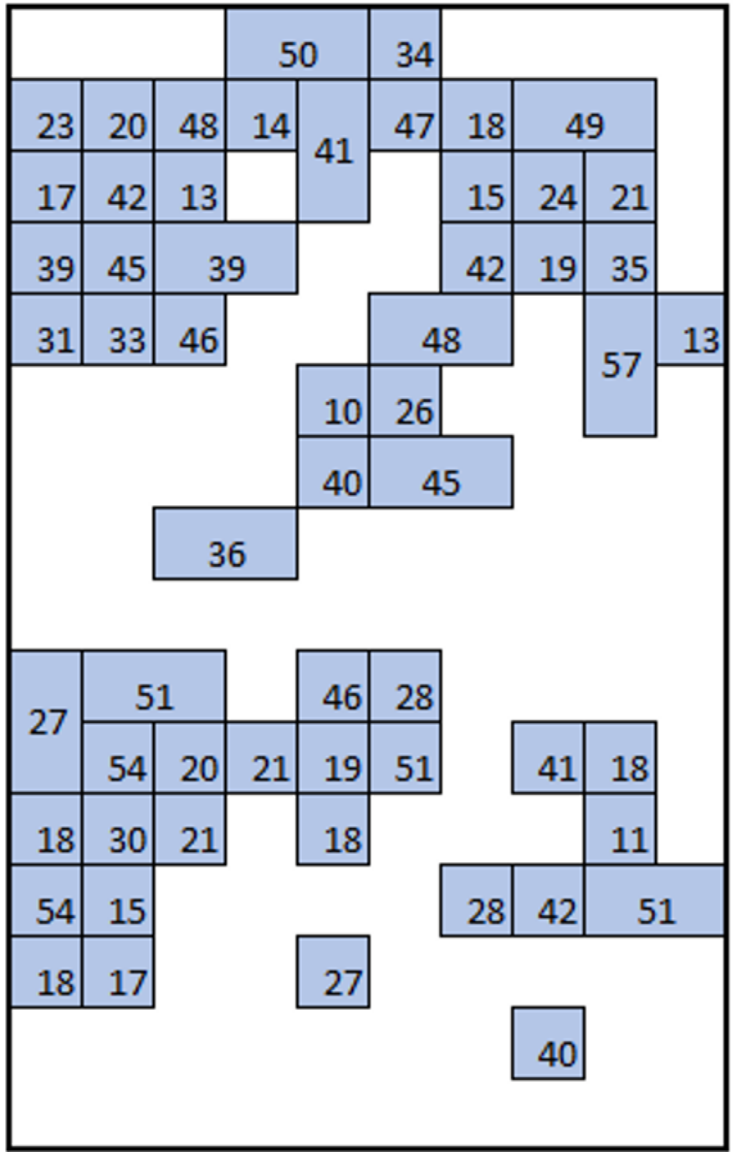
Para juntarlas con las manzanas que comparten frontera.

www.ecuadorencifras.gob.ec



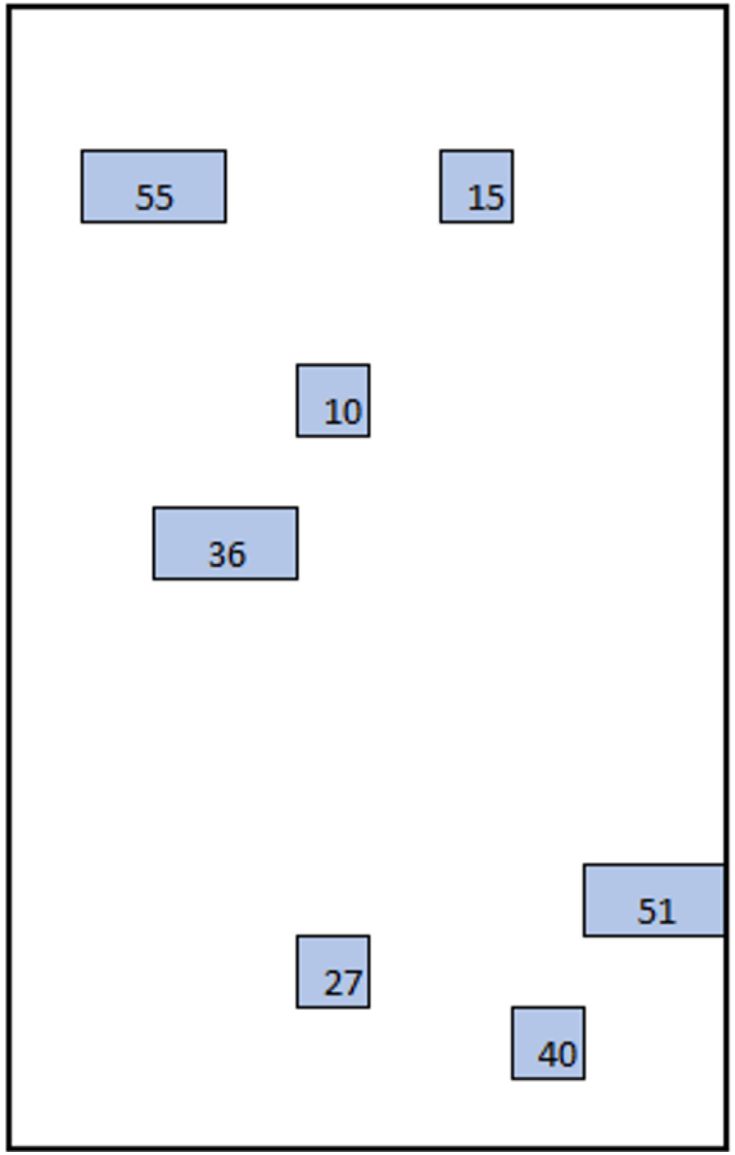
Y evaluando si superan la restricción de tamaño.

www.ecuadorencifras.gob.ec



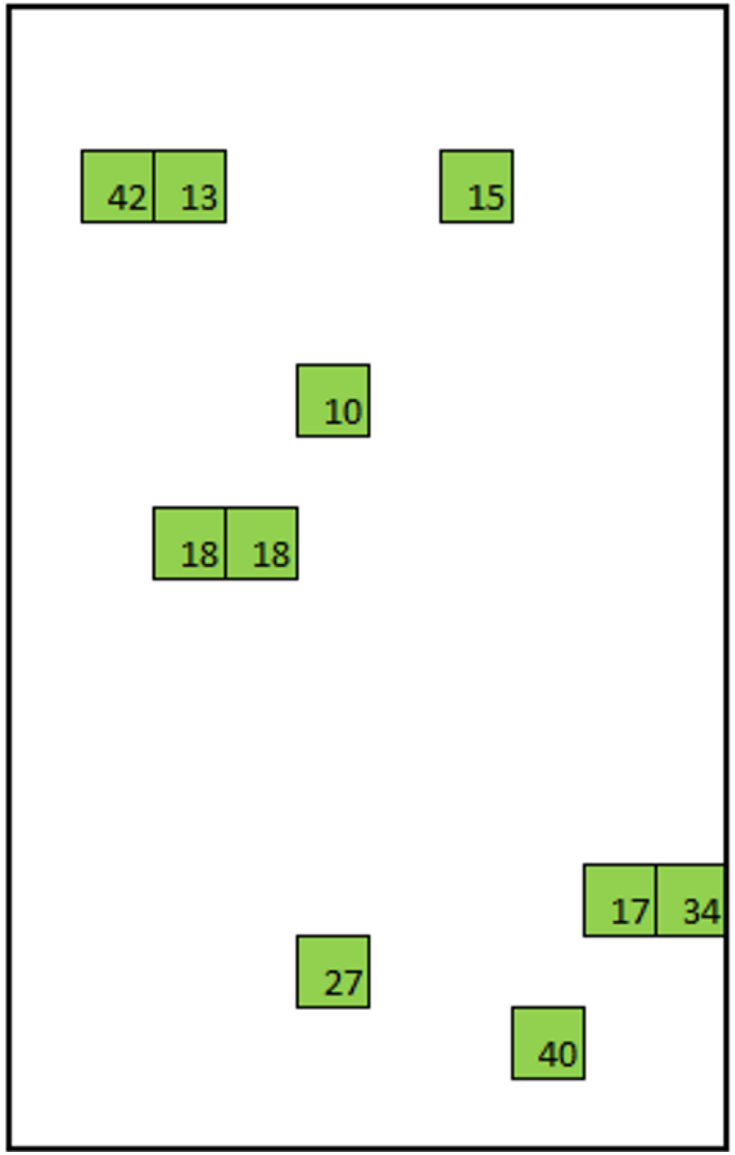
Para ser excluidas si son conglomerados.

www.ecuadorencifras.gob.ec



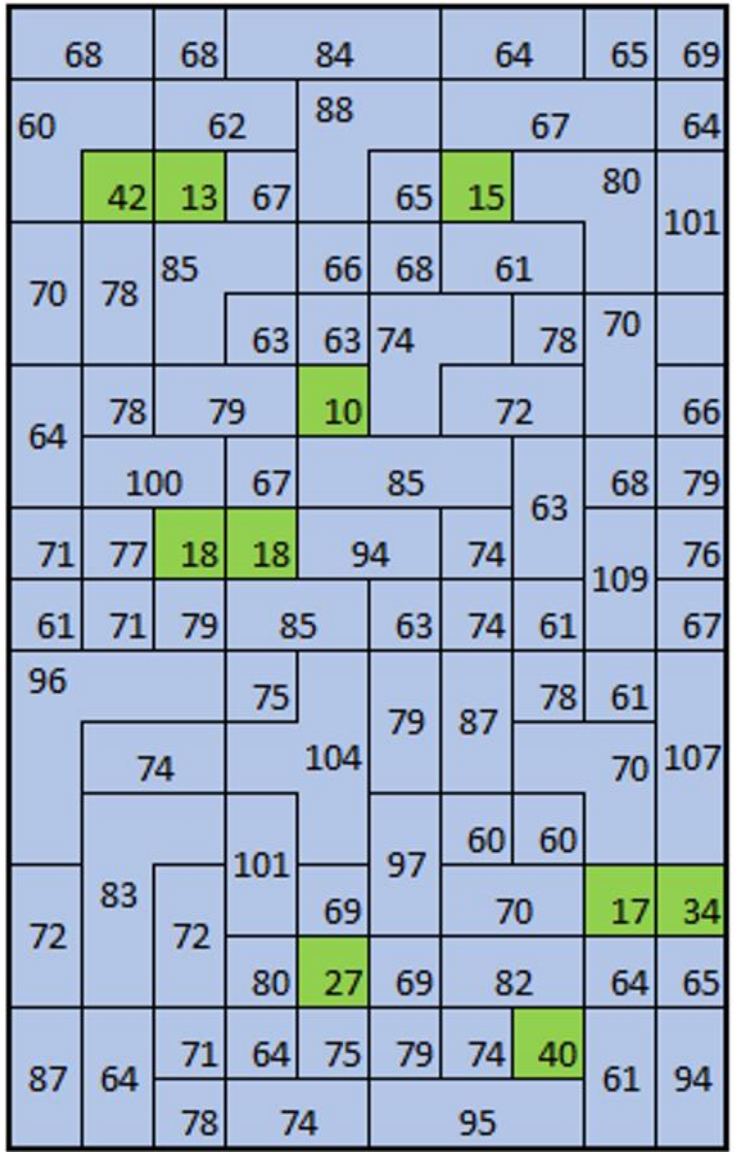
Al final, o todas las manzanas pertenecen a un conglomerado (proceso terminado) o existen conjuntos aislados que no cumplen la restricción de tamaño.

www.ecuadorencifras.gob.ec



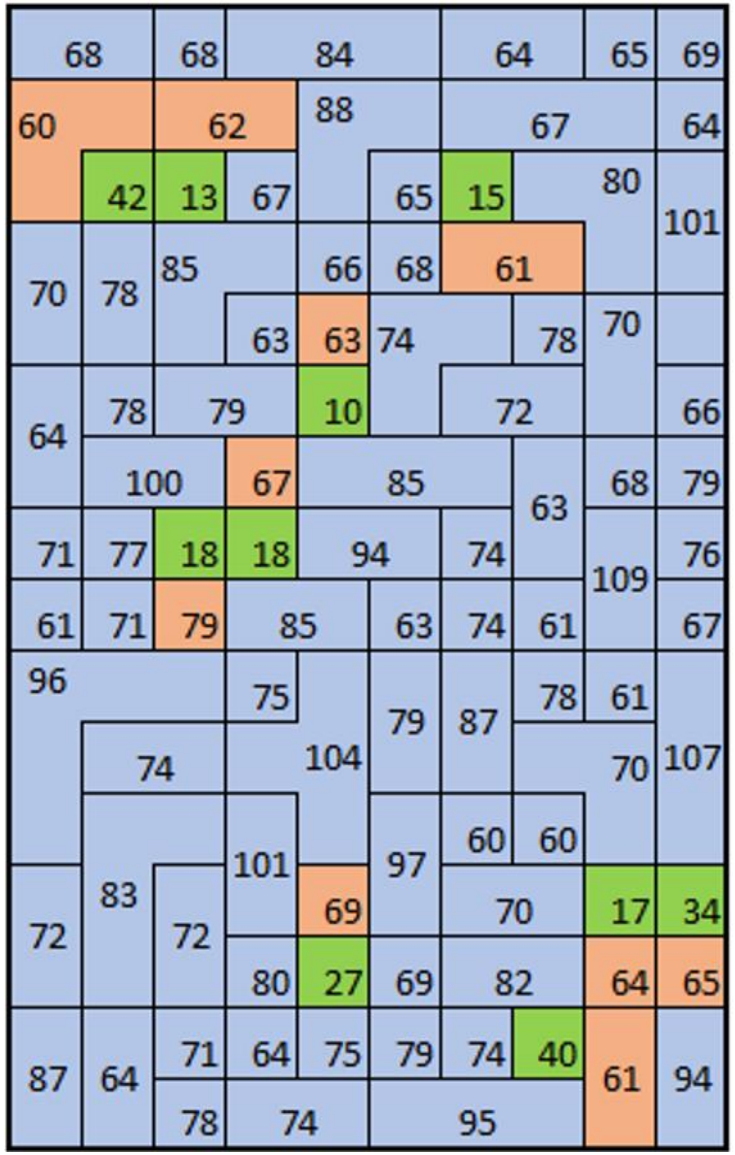
Para el tratamiento de aislados, recuperamos sus nodos y tamaños primigenios.

www.ecuadorencifras.gob.ec



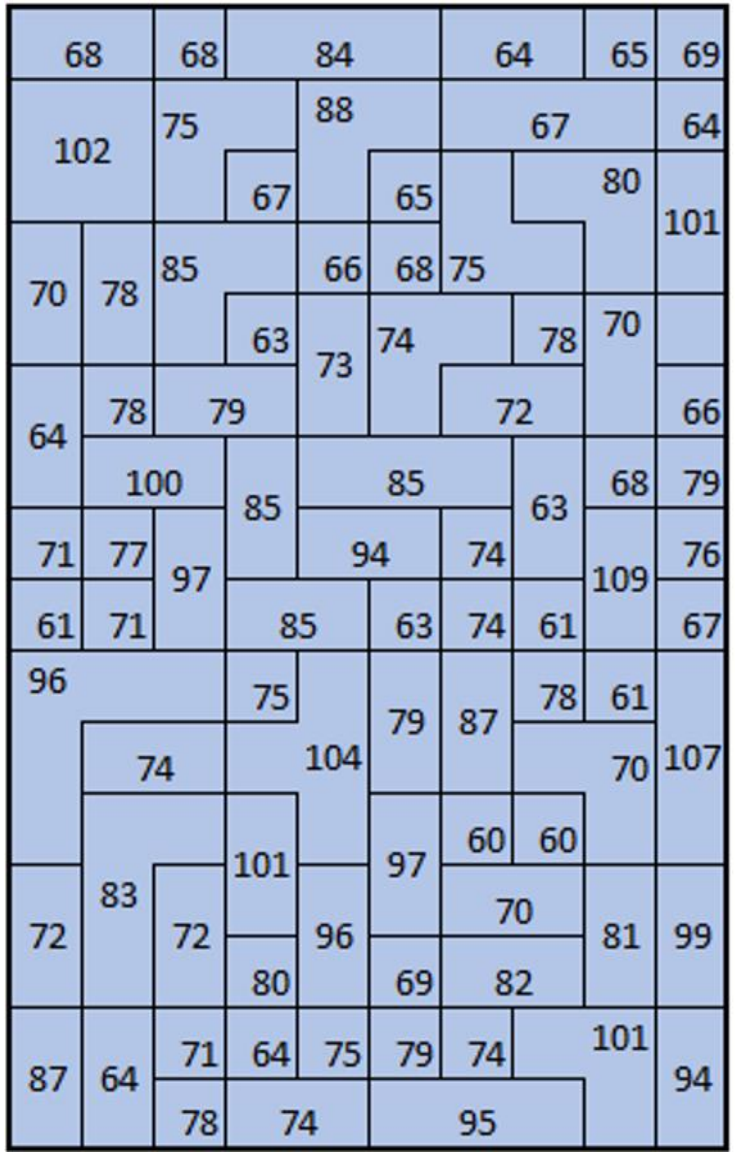
Luego, generamos las incidencias entre las manzanas aisladas y los conglomerados creados.

www.ecuadorencifras.gob.ec



Con esto, se junta la manzana aislada con el conglomerado de menor tamaño que tiene incidencia.

www.ecuadorencifras.gob.ec



Terminando así el ejercicio de conglomeración sobre Teselia!

www.ecuadorencifras.gob.ec

Buenas cifras, mejores vidas

Muchas gracias!

(Espacio de preguntas)

Mas información a:
angel_gaibor@inec.gob.ec javier_nunez@inec.gob.ec

Buenas cifras,
mejores vidas

@ecuadorencifras @ecuadorencifras @InecEcuador t.me/ecuadorencifras INEC/Ecuador INECEcuador INEC Ecuador