



---

**Universidad de Valladolid**

Facultad de Ciencias

Trabajo Fin de Grado

Grado en Estadística

2020/2021

**Caracterización de las Secciones Censales de Valladolid en  
función de su comportamiento electoral**

Autor:

**Carlos González Duque**

Tutora:

**M. Pilar Rodríguez del Tío**



Quiero agradecer el gran apoyo que he recibido por parte de mis padres, de Laura, de Jaime, de mis compañeros y, en general, de todas aquellas personas que han aportado su granito de arena para que hayamos llegado juntos hasta este punto.

También me gustaría destacar la ayuda proporcionada por mi tutora Pilar y las contribuciones de Ángel Agúndez desde el Ayuntamiento de Valladolid durante la realización de este trabajo.



# Índice general

Índice de tablas	5
Índice de figuras	8
Resumen	10
Abstract	11
<b>1. Introducción</b>	<b>12</b>
<b>2. Obtención y tratamiento de datos</b>	<b>14</b>
2.1. Datos de resultados electorales . . . . .	14
2.2. Datos cartográficos . . . . .	15
<b>3. Análisis cluster</b>	<b>16</b>
3.1. Métodos Jerárquicos . . . . .	17
3.2. Métodos no Jerárquicos . . . . .	18
<b>4. Aplicación a resultados electorales de Valladolid</b>	<b>20</b>
4.1. Análisis no jerarquizado con partición óptima . . . . .	21
4.1.1. Elecciones 10 de junio de 1987 . . . . .	22
4.1.2. Elecciones 26 de mayo de 1991 . . . . .	24
4.1.3. Elecciones 28 de mayo de 1995 . . . . .	26
4.1.4. Elecciones 13 de junio de 1999 . . . . .	29
4.1.5. Elecciones 25 de mayo de 2003 . . . . .	31
4.1.6. Elecciones 27 de mayo de 2007 . . . . .	33
4.1.7. Elecciones 22 de mayo de 2011 . . . . .	36
4.1.8. Elecciones 24 de mayo de 2015 . . . . .	38
4.1.9. Elecciones 26 de mayo de 2019 . . . . .	40
4.1.10. Secciones próximas al perfil global de Valladolid . . . . .	43
4.2. Análisis no jerarquizado con partición en 20 clusters . . . . .	45
4.2.1. Elecciones 10 de junio de 1987 . . . . .	45
4.2.2. Elecciones 26 de mayo de 1991 . . . . .	46
4.2.3. Elecciones 28 de mayo de 1995 . . . . .	47
4.2.4. Elecciones 13 de junio de 1999 . . . . .	47
4.2.5. Elecciones 25 de mayo de 2003 . . . . .	48
4.2.6. Elecciones 27 de mayo de 2007 . . . . .	49

## ÍNDICE GENERAL

---

4.2.7. Elecciones 22 de mayo de 2011 . . . . .	49
4.2.8. Elecciones 24 de mayo de 2015 . . . . .	50
4.2.9. Elecciones 26 de mayo de 2019 . . . . .	51
4.2.10. Secciones próximas al perfil global de Valladolid . . . . .	52
4.3. Análisis jerarquizado . . . . .	54
4.4. Resumen de resultados . . . . .	56
<b>5. Conclusiones</b>	<b>60</b>
<b>Anexo</b>	<b>62</b>
Siglas de Partidos Políticos . . . . .	62
<b>Bibliografía</b>	<b>64</b>

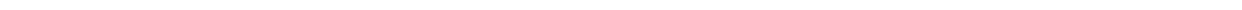
# Índice de tablas

4.1. Tabla resumen de la división para el año 1987 . . . . .	22
4.2. Distribución del voto global y de los clusters en el año 1987 . . . . .	23
4.3. Tabla de secciones representativas el año 1987 . . . . .	24
4.4. Tabla resumen de la división para el año 1991 . . . . .	25
4.5. Distribución del voto global y de los clusters en el año 1991 . . . . .	25
4.6. Tabla de secciones representativas el año 1991 . . . . .	25
4.7. Tabla resumen de la división para el año 1995 . . . . .	27
4.8. Distribución del voto global y de los clusters en el año 1995 . . . . .	27
4.9. Tabla de secciones representativas el año 1995 . . . . .	28
4.10. Tabla resumen de la división para el año 1999 . . . . .	29
4.11. Distribución del voto global y de los clusters en el año 1999 . . . . .	30
4.12. Tabla de secciones representativas del año 1999 . . . . .	31
4.13. Tabla resumen de la división para el año 2003 . . . . .	32
4.14. Distribución del voto global y de los clusters en el año 2003 . . . . .	32
4.15. Tabla de secciones representativas del año 2003 . . . . .	32
4.16. Tabla resumen de la división para el año 2007 . . . . .	34
4.17. Distribución del voto global y de los clusters en el año 2007 . . . . .	34
4.18. Tabla de secciones representativas del año 2007 . . . . .	35
4.19. Tabla resumen de la división para el año 2011 . . . . .	36
4.20. Distribución del voto global y de los clusters en el año 2011 . . . . .	37
4.21. Tabla de secciones representativas del año 2011 . . . . .	38
4.22. Tabla resumen de la división para el año 2015 . . . . .	39
4.23. Distribución del voto global y de los clusters en el año 2015 . . . . .	39
4.24. Tabla de secciones representativas del año 2015 . . . . .	39
4.25. Tabla resumen de la división para el año 2019 . . . . .	40
4.26. Distribución del voto global y de los clusters en el año 2019 . . . . .	41
4.27. Tabla de secciones representativas del año 2019 . . . . .	42
4.28. Tabla de frecuencias de secciones representativas aplicando 4.1 . . . . .	43
4.29. Tabla de secciones más representativas aplicando 4.1 . . . . .	43
4.30. Tabla de frecuencias de secciones representativas aplicando 4.1 a las cuatro últimas convocatorias . . . . .	44
4.31. Tabla de secciones más representativas aplicando 4.1 a las cuatro últimas convocatorias . . . . .	44
4.32. Tabla resumen de la división para el año 1987 aplicando 4.2 . . . . .	45
4.33. Tabla de secciones representativas del año 1987 aplicando 4.2 . . . . .	46
4.34. Tabla resumen de la división para el año 1991 aplicando 4.2 . . . . .	46

4.35. Tabla de secciones representativas del año 1991 aplicando 4.2 . . . . .	46
4.36. Tabla resumen de la división para el año 1995 aplicando 4.2 . . . . .	47
4.37. Tabla de secciones representativas del año 1995 aplicando 4.2 . . . . .	47
4.38. Tabla resumen de la división para el año 1999 aplicando 4.2 . . . . .	47
4.39. Tabla de secciones representativas del año 1999 aplicando 4.2 . . . . .	48
4.40. Tabla resumen de la división para el año 2003 aplicando 4.2 . . . . .	48
4.41. Tabla de secciones representativas del año 2003 aplicando 4.2 . . . . .	48
4.42. Tabla resumen de la división para el año 2007 aplicando 4.2 . . . . .	49
4.43. Tabla de secciones representativas del año 2007 aplicando 4.2 . . . . .	49
4.44. Tabla resumen de la división para el año 2011 aplicando 4.2 . . . . .	49
4.45. Tabla de secciones representativas del año 2011 aplicando 4.2 . . . . .	50
4.46. Tabla resumen de la división para el año 2015 aplicando 4.2 . . . . .	50
4.47. Tabla de secciones representativas del año 2015 aplicando 4.2 . . . . .	50
4.48. Tabla resumen de la división para el año 2019 aplicando 4.2 . . . . .	51
4.49. Tabla de secciones representativas del año 2019 aplicando 4.2 . . . . .	51
4.50. Tabla de frecuencias de secciones representativas aplicando 4.2 . . . . .	52
4.51. Tabla de secciones más representativas aplicando 4.2 . . . . .	52
4.52. Tabla de frecuencias de secciones representativas aplicando 4.2 a las últimas cuatro convocatorias . . . . .	53
4.53. Tabla de secciones más representativas aplicando 4.2 a las últimas cuatro convocatorias . . . . .	53
4.54. Tabla de coeficientes de agregación . . . . .	54
4.55. Tabla con secciones más representativas aplicando 4.3 . . . . .	55
4.56. Tabla de secciones más representativas en sentido global . . . . .	57
5.1. Tabla de siglas de los partidos políticos nombrados en el Trabajo . . . . .	62

# Índice de figuras

3.1. Ejemplo de análisis cluster en datos simulados . . . . .	16
4.1. TWSS en función del n <sup>o</sup> de grupos 1987 . . . . .	22
4.2. Mapa de secciones más representativas del año 1987 . . . . .	23
4.3. Silueta media en función del n <sup>o</sup> de grupos 1991 . . . . .	24
4.4. Mapa de secciones más representativas del año 1991 . . . . .	26
4.5. TWSS en función del n <sup>o</sup> de grupos 1995 . . . . .	27
4.6. Mapa de secciones más representativas del año 1995 . . . . .	28
4.7. TWSS en función del n <sup>o</sup> de grupos 1999 . . . . .	29
4.8. Mapa de secciones más representativas del año 1999 . . . . .	30
4.9. TWSS en función del n <sup>o</sup> de grupos 2003 . . . . .	31
4.10. Mapa de secciones más representativas del año 2003 . . . . .	33
4.11. TWSS y silueta media en función del n <sup>o</sup> de grupos 2007 . . . . .	34
4.12. Mapa de secciones más representativas del año 2007 . . . . .	35
4.13. TWSS en función del n <sup>o</sup> de grupos 2011 . . . . .	36
4.14. Mapa de secciones más representativas del año 2011 . . . . .	37
4.15. TWSS en función del n <sup>o</sup> de grupos 2015 . . . . .	38
4.16. Mapa de secciones más representativas del año 2015 . . . . .	40
4.17. TWSS en función del n <sup>o</sup> de grupos 2019 . . . . .	41
4.18. Mapa de secciones más representativas del año 2019 . . . . .	42
4.19. Mapa con todas las secciones más representativas . . . . .	58



---

## Resumen

En este trabajo se pretende caracterizar las secciones censales de la ciudad de Valladolid de acuerdo con la distribución de los votos en procesos electorales, así como encontrar aquellas secciones más representativas del comportamiento global de la ciudad. Para ello, se utilizan resultados de las elecciones municipales desde el año 1987 hasta el 2019, nueve convocatorias en total. La clasificación se lleva a cabo por medio de la técnica de análisis cluster implementada en el lenguaje de programación R.

### Palabras clave

Elecciones municipales, Valladolid, secciones censales, Clasificación, Análisis Cluster.

---

## Abstract

This dissertation aims to characterize the census tracts of the city of Valladolid according to the distribution of votes in electoral processes as well as to find the most representative tracts of the overall behaviour of the city. For this purpose results of the municipal elections from 1987 to 2019 are used, nine callings in total. The classification is carried out using the technique cluster analysis implemented in the R Programming Language.

### Key words

Municipal elections, Valladolid, Census tracts, Classification, Cluster analysis.

# Capítulo 1

## Introducción

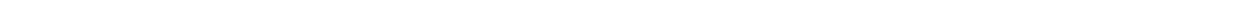
El propósito de este trabajo es estudiar las distribuciones del voto de las secciones censales de Valladolid con el objetivo de determinar cuales de ellas son las más representativas de toda la ciudad. Los datos empleados para realizar el estudio son los resultados electorales correspondientes a las nueve convocatorias municipales realizadas entre el año 1987 y 2019.

La técnica de clasificación utilizada consiste en realizar un análisis cluster sobre un conjunto de objetos formado por las secciones censales de Valladolid en cada convocatoria. El resultado obtenido está formado por una lista de grupos o clusters en los cuales las secciones que han sido asignadas a una misma agrupación tendrán distribuciones del voto parecidas. Esta técnica pertenece a la familia de los métodos de análisis no supervisado, es decir, aquellos en los que no se conoce ningún tipo de *clase* preestablecida en la cual se agrupen los elementos.

A lo largo de la memoria se muestra la aplicación de tres implementaciones diferentes del análisis cluster junto con los resultados obtenidos con cada una.

- La primera se basa en un análisis no jerarquizado utilizando un número de clusters óptimo en cada convocatoria electoral estudiada.
- Con el objetivo de que el proceso de elección de las secciones representativas del resultado total de la ciudad sea más preciso, se realiza el análisis no jerarquizado con una partición dividida en 20 clusters.
- La última de las aplicaciones empleadas en este trabajo es un análisis jerarquizado, utilizando el índice de agregación que mejor agrupación genera.

Por otro lado, para acompañar a los resultados de algunos de los análisis realizados se muestran mapas cartográficos de la ciudad en los que poder visualizar la ubicación de secciones concretas.



# Capítulo 2

## Obtención y tratamiento de datos

En la elaboración de este trabajo se han necesitado dos bases de datos diferentes. La primera formada por los resultados electorales con los que realizar los análisis cluster y la segunda relativa a los datos cartográficos empleados en la construcción de mapas. Cada una de las dos bases de datos tiene una fuente distinta.

### 2.1. Datos de resultados electorales

La obtención de este conjunto de datos se ha llevado a cabo desde la página web del Ministerio del Interior, concretamente en el área de descargas del apartado "Consulta de Resultados Electorales" [1]. La plataforma ofrece varios archivos de descarga por cada convocatoria electoral dependiendo del nivel de desagregación del conjunto de datos deseado, en este caso se ha necesitado el nivel más bajo posible (mesas electorales) para obtener información de cada una de las secciones censales.

Los nueve archivos descargados, uno por cada convocatoria estudiada, contienen gran cantidad de datos tanto demográficos como electorales de cada una de las secciones censales del país, toda la información está codificada mediante cadenas de caracteres.

Como para este trabajo solo nos interesa la información relativa a los votos de cada candidatura en todas las secciones censales de Valladolid (250 en total aproximadamente), se ha realizado, mediante programación con el lenguaje R, una decodificación y selección de las filas de interés. Una vez completado este proceso se ha conseguido obtener conjuntos en los que cada fila representa la distribución del voto entre las diferentes candidaturas de cada sección en tantos por 1. En el caso en el que en una convocatoria electoral una de las candidaturas no haya obtenido ningún voto, se eliminará la columna correspondiente a dicha candidatura. Con esta disposición de la información ya se pueden aplicar todas las técnicas descritas en el siguiente capítulo de la memoria.

### 2.2. Datos cartográficos

Por un lado, los archivos de datos espaciales necesarios para representar la división por secciones censales de Valladolid se han obtenido desde la página web del Instituto Nacional de Estadística (INE) en el apartado de "Productos y Servicios" [2]. En esta plataforma están disponibles archivos en formato *shape* con los contornos georeferenciados de las secciones censales de España para todos los años desde 2011. Por otro lado, el Ayuntamiento de Valladolid nos ha proporcionado los mapas cartográficos correspondientes de los años 2003 y 2007.

La división en secciones censales de la ciudad de Valladolid ha ido variando a lo largo del tiempo atendiendo a razones demográficas relativas a cada sección, dado que el número de electores debe de estar entre 500 y 2500. Como consecuencia, la mayoría de convocatorias desde 1987 hasta 2019 han tenido un número de secciones distinto e incluso, en algunos casos, se ha reorganizado la distribución de ciertas zonas urbanísticas. Esto nos obliga a que para representar los resultados del análisis cluster para una convocatoria electoral concreta debemos emplear el mapa correspondiente al año de dicha convocatoria.

Por este motivo, solo se podrán representar mediante mapas de forma íntegra los resultados de las convocatorias de 2003, 2007, 2011, 2015 y 2019. Para el resto de convocatorias se ha empleado el mapa cartográfico del año 2003 para visualizar los resultados restringiendo el conjunto de secciones censales a aquellas que no hayan sufrido ningún cambio a lo largo del tiempo hasta dicho año.

# Capítulo 3

## Análisis cluster

El análisis cluster es un conjunto de técnicas de clasificación no supervisada cuyo objetivo es encontrar grupos de observaciones en función de los valores que tomen las variables. El mayor interés reside en construir una partición del conjunto de datos en la que las observaciones que pertenezcan a un mismo subconjunto sean similares y al mismo tiempo las que no pertenezcan al mismo grupo sean lo más diferentes posibles [3].

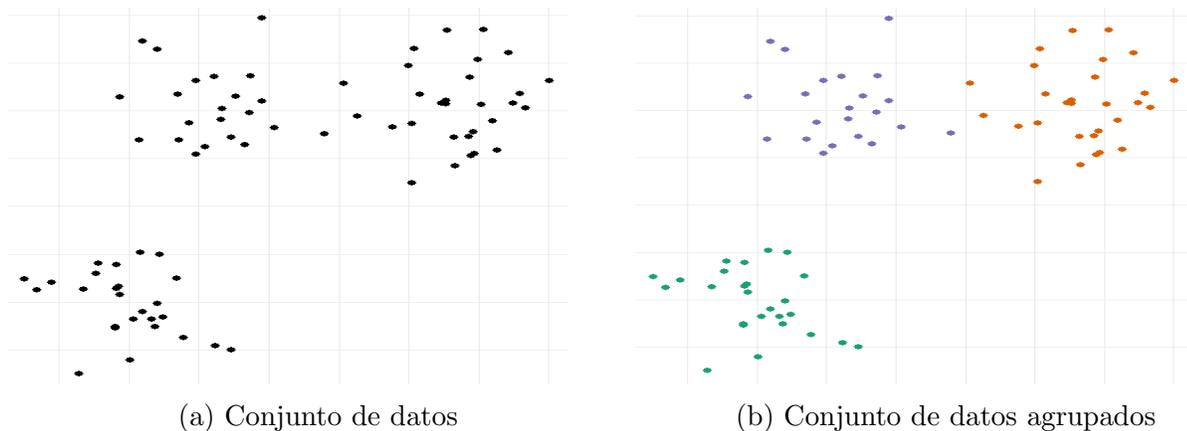


Figura 3.1: Ejemplo de análisis cluster en datos simulados

El ingrediente principal que hay que definir para aplicar un análisis cluster es la forma con la que se cuantifica la similitud-divergencia entre las observaciones. A esta medida de asociación se la denomina *índice de disimilaridad* y puede calcularse de múltiples maneras. Algunas de las más comunes son la distancia *euclídea*, la distancia de *Manhattan* o la distancia *1-coseno*. Habitualmente la estructura de datos que se emplea para implementar esta información es un matriz simétrica de dimensión igual al número de observaciones del conjunto.

### 3.1. MÉTODOS JERÁRQUICOS

---

La aplicación del análisis cluster requiere, por lo general, no sólo el cálculo de las distancias o similitudes entre los individuos iniciales, sino, también la determinación de las distancias o similitudes entre los grupos y/o entre un grupo y un individuo [4].

A esta herramienta se la conoce como *índice de agregación* y tres de los posibles métodos utilizados para calcularla son:

- *Single Linkage*: La distancia entre dos clusters equivale a la distancia entre los dos individuos más próximos.
- *Complete Linkage*: Este método considera la distancia entre dos clusters a la distancia entre los individuos más lejanos.
- *Ward*: Aplicando este índice de agregación se unen los clusters o individuos que minimicen el incremento del valor de la suma de los cuadrados total de las diferencias de cada individuo al centro de gravedad del cluster al que pertenece cada uno.

El conjunto de técnicas de análisis cluster se puede dividir en dos grupos dependiendo de la filosofía que tengan los algoritmos de clasificación utilizados, métodos jerárquicos y no jerárquicos.

## 3.1. Métodos Jerárquicos

En los métodos jerárquicos los individuos no se particionan en clusters de una sola vez, sino que se van haciendo particiones sucesivas a distintos niveles de agregación o agrupamiento [4].

Fundamentalmente el algoritmo de clasificación sigue el método ascendente jerárquico que consta de los siguientes pasos:

1. Se parte de tantos grupos como observaciones haya en el conjunto.
2. Secuencialmente, se va uniendo los dos grupos con menor índice de agregación, una agrupación por iteración.
3. Se finaliza cuando se han agrupado en un solo grupo todas las observaciones disponibles.

El número de grupos es una decisión subjetiva y depende del contexto en el que se esté realizando el análisis, así como de la finalidad del mismo. Una buena herramienta para abordar esta cuestión es el llamado *dendrograma* en donde se muestra de forma visual el método ascendente jerárquico.

### 3.2. Métodos no Jerárquicos

El conjunto de observaciones no se va particionando sucesivamente si no que se divide en el mismo número de subconjuntos desde un inicio, con la particularidad de que en cada paso una observación puede cambiar el grupo al que es asignado. Por lo tanto, el número de grupos en los que se divide el conjunto ha de ser fijado previamente y no varía en todo el proceso.

Tras decidir un criterio de agrupamiento adecuado, el objetivo a considerar es encontrar una partición en  $k$  grupos que optimice dicho criterio [5].

La mecánica de los métodos no jerárquicos se puede enumerar en los siguientes pasos:

1. Encontrar una partición inicial del conjunto en  $k$  grupos.
2. Calcular el cambio producido en el criterio de agrupamiento al mover cada objeto de su grupo a otro.
3. Realizar el cambio que conduzca a la mayor mejora en el valor del criterio de agrupación.
4. repetir los pasos 2 y 3 hasta que ningún cambio de grupo de alguna observación mejore el valor del criterio de agrupamiento [5].

Dentro de este campo, el algoritmo más utilizado es el de las *k-medias*, en el cual se asigna cada observación al grupo cuyo centroide (centro de gravedad de las observaciones del grupo) esté más cercano en el sentido de la distancia euclídea.

El criterio de agrupación que se minimiza es la suma en todos los grupos de la suma de cuadrados de la diferencia entre cada observación y el centroide de cada grupo.



# Capítulo 4

## Aplicación a resultados electorales de Valladolid

En este capítulo se aplica la técnica de análisis cluster jerárquico y no jerárquico explicada anteriormente a los resultados electorales correspondientes a las nueve convocatorias municipales celebradas entre el año 1987 y 2019 en Valladolid.

Con el objetivo de caracterizar las secciones censales de la ciudad en base a su distribución del voto se va a realizar un análisis sobre los resultados electorales de cada año de forma independiente. El objetivo principal es encontrar aquellas secciones cuya distribución de voto se comporte de forma similar a la del conjunto global de toda la ciudad en ese año electoral. Para lo cual se crea una sección artificial que contenga la distribución global del voto, de modo que las secciones que, tras aplicar el análisis, pertenezcan a su mismo grupo se considerarán las más representativas de la ciudad.

El primer aspecto que se debe abordar es establecer un índice de disimilaridad para poder cuantificar las diferencias entre las distribuciones del voto en las secciones censales. En este aspecto se ha optado por emplear la distancia de Kullback-Leibler ya que es una medida apropiada para medir diferencias entre dos distribuciones de probabilidad.

Se define como:

$$D(g, h) = \sum_{i=1}^c g(i) \ln \frac{g(i)}{h(i)}$$

Siendo  $g$  y  $h$  dos distribuciones de probabilidad de una variable aleatoria discreta que toma  $c$  valores. En nuestro caso  $c$  es el número de candidaturas que se presentaron en el año electoral. El paquete *seewave* de R nos proporciona la herramienta para la construcción de la matriz de distancias de kullback-Leibler a través de la función *kl.dist()* [6]. Puede darse el caso en el que una candidatura no haya obtenido ningún voto en una sección censal concreta, en estas circunstancias el coeficiente  $h(i)$  valdría cero y, por lo tanto, se invalidaría la división. Para atajar este problema, la distancia proporcionada por la función *kl.dist()* en esta situación es por defecto cero.

El siguiente punto a tratar es decidir el número de clusters en los que dividir el conjunto de secciones en el proceso. Es una cuestión subjetiva y existen varias alternativas que se estudiarán a continuación.

## 4.1. Análisis no jerarquizado con partición óptima

En este apartado el objetivo es elegir el número de clusters óptimo en los que realizar la agrupación. Por partición óptima se entiende una división en la que las asignaciones de las secciones sean lo más estables posibles.

Existen múltiples alternativas para tomar esta decisión, en este caso se determinará teniendo en cuenta dos criterios:

1. *Total within sum of square, TWSS*: Suma del cuadrado de las desviaciones de cada observación al representante del grupo (centroide o medoide  $C_i$ ) al que ha sido asignado. Cuanto más pequeño sea este valor más compacta será la partición ya que los clusters generados tendrán menor variabilidad.

$$T = \sum_{i=1}^k \sum_{i \in C} \|x - C_i\|^2$$

2. *Average Silhouette Width, ADW*: La anchura media de la silueta mide en promedio cómo de similares son las observaciones de un grupo en comparación con el resto de grupos. El intervalo de valores que puede tomar es  $[-1, 1]$  y la clasificación será mejor cuanto más cercano a 1 esté:

$$ASW = \frac{1}{n} \sum_{i=1}^n s(i)$$
$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Donde  $a(i)$  es la disimilaridad media entre la observación  $i$  y el resto del grupo al que pertenezca, y  $b(i)$  es la disimilaridad media entre la observación  $i$  y todas las observaciones que pertenecen al siguiente grupo más cercano [7].

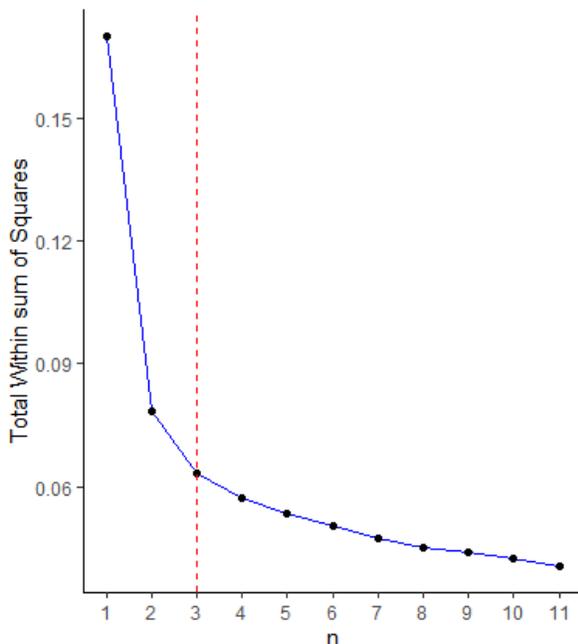
Teniendo en cuenta todos los aspectos indicados se aplicará el análisis no jerárquico con la función de R `pam()` (*Partitioning Around Medoids*) incluida en el paquete `cluster`. [8] Esta función implementa el algoritmo **k-medoids** (Kaufman and Rousseeuw 1990) [9], el cual pertenece a la familia de algoritmos k-medias. Tiene la particularidad de que no necesariamente se tiene que utilizar la distancia euclídea para obtener la distancia entre dos objetos, de esta forma se puede utilizar la distancia de Kullback-Leibler. Otra diferencia respecto al algoritmo k-medias original es que el representante de cada grupo no es el centro de gravedad del mismo si no que es aquella observación con la cual se obtiene menor suma de disimilaridades en el cluster (TWSS), a este elemento se le denomina *Medoide*. En la práctica, es un método más robusto al ser menos sensible a valores atípicos.

Por otro lado, con el propósito de conseguir una visualización más clara de los resultados del análisis realizado se incluyen mapas de la ciudad divididos por secciones censales en los cuales cada color muestra el grupo al que se ha asignado cada sección, es decir, la distribución del voto entre las candidaturas en secciones pintadas de un mismo color es similar. La implementación de los mapas se ha llevado a cabo con ayuda de los paquetes `sf` [10] y `ggplot2` [11].

4.1.1. Elecciones 10 de junio de 1987

En las elecciones del 10 de junio de 1987 se presentaron 12 candidaturas, la división censal de Valladolid constaba de 248 secciones y 12 distritos. Teniendo en cuenta la silueta media y la suma de cuadrados dentro de los grupos total (*TWSS*) se aproxima el número de grupos óptimo en los que dividir el conjunto de secciones.

En la figura 4.1 se muestra gráfica y analíticamente la evolución de la suma de cuadrados total dentro de los grupos. Se puede observar que la convergencia se alcanza cuando el conjunto se divide en 3 grupos, el valor mejora cuantas más agrupaciones se hagan pero la diferencia en el beneficio es leve por lo que se decide que el número óptimo buscado es 3.



nº grupos	TWSS
1	0.170
2	0.0784
<b>3</b>	<b>0.0631</b>
4	0.0570
5	0.0533
6	0.05018
7	0.0472
8	0.0452
9	0.0438
10	0.0424
11	0.0405

(a) Gráfico con TWSS de 1987

(b) Tabla con TWSS de 1987

Figura 4.1: TWSS en función del nº de grupos 1987

Tras aplicar el algoritmo *k-medoids* para 3 agrupaciones y empleando la distancia de Kullback-Leibler como índice de disimilaridad se obtiene la división mostrada en la tabla 4.1. En todas las convocatorias estudiadas, el cluster objetivo en donde se agrupan las secciones con distribuciones más cercanas a la del conjunto de toda la ciudad va a ser el grupo G1.

G1	G2	G3	N	Silueta media	TWSS
89	65	94	248	0.42611	0.0631

Tabla 4.1: Tabla resumen de la división para el año 1987

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

Candidatura	PSOE	AP	CDS	IU	PTE-UC
% G1	35.5	26	26.5	6.1	3.6
% G2	18	52.8	21.7	3.7	1.3
% G3	50	14.6	21.2	6.8	4.4
% Valladolid	36	28.6	23.7	5.8	3.3

Tabla 4.2: Distribución del voto global y de los clusters en el año 1987

En la tabla 4.2 se representa la distribución promedio de las secciones de cada uno de los tres clusters junto con el resultado global de Valladolid en las cinco candidaturas más votadas, se puede ver que el Partido Socialista Obrero Español (PSOE) es la primera fuerza política seguida del Centro Democrático y Social (CDS) y de Alianza Popular (AP), estos dos últimos con porcentajes de voto muy similares. La distribución del cluster G2 es muy distinta, AP obtiene más de la mitad de los votos seguido de CDS y del PSOE. Por último, la tercera agrupación está formada por secciones en las que el PSOE fue el partido más votado con una ventaja más significativa que en las secciones del grupo G1.

A continuación se representa la situación espacial de las 89 secciones cuya distribución es similar a la del comportamiento de toda la ciudad y la lista de dichas secciones desagregadas por distritos.

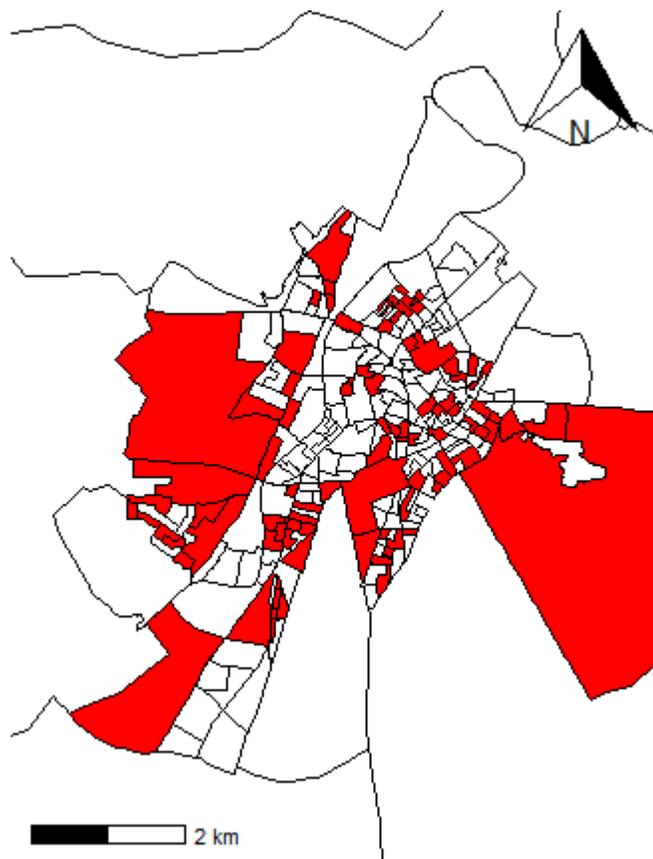


Figura 4.2: Mapa de secciones más representativas del año 1987

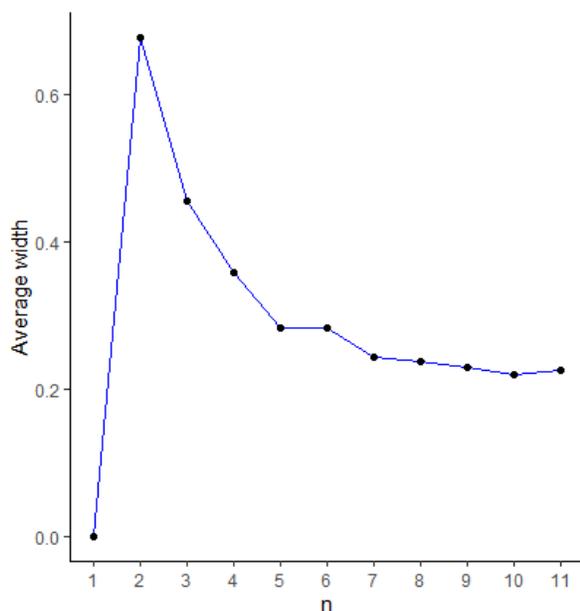
Distrito	Secciones
2	15 16 19
3	1 4 5 6 7
4	1 3 10 12 14 15 16 19 24 26 27
5	5
6	2 4 5 7 8 9 10 11 12 14 15 17 18 32 34
7	5 6 7 8 9 15
8	1 2 3 5 8 9 12 13 24 25 28 37 41
9	6 10 12
10	2 4 8 10 13 17 18 21 24 25 26 28
11	1 2 5 7 8 9 11 12 13 16 18 19 21 23 24 25 26 27 28 29

Tabla 4.3: Tabla de secciones representativas el año 1987

### 4.1.2. Elecciones 26 de mayo de 1991

En las elecciones del 26 de mayo de 1991 se presentaron 9 candidaturas, en este momento la división censal de Valladolid constaba de 249 secciones y 12 distritos.

En este caso, para establecer el número de grupos óptimo se muestra el valor de la silueta media dependiendo de dicho parámetro. El valor más adecuado es claramente 2. Dividiendo el conjunto de secciones en 2 se consigue que los dos clusters generados estén lo más alejados posible el uno del otro.



n <sup>o</sup> grupos	Silueta media
1	0
<b>2</b>	<b>0.67664</b>
3	0.4546
4	0.3579
5	0.2819
6	0.2825
7	0.2430
8	0.2373
9	0.2286
10	0.2196
11	0.2248

(a) Gráfico con silueta media de 1991

(b) Tabla con silueta media 1991

Figura 4.3: Silueta media en función del n<sup>o</sup> de grupos 1991

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

Por otro lado, la suma de cuadrados interna total obtenida con esta división toma un valor de 0.06853, lo que nos indica que la variabilidad de las secciones de un mismo grupo es baja, por lo que finalmente se opta por esta división.

G1	G2	N	Silueta media	TWSS
158	91	249	0.6766428	0.06853

Tabla 4.4: Tabla resumen de la división para el año 1991

En las divisiones del cluster G1 se agrupan los elementos con resultados más parecidos al perfil global, en estos casos el PSOE es la candidatura más votada seguido del Partido Popular (PP). El otro cluster generado agrupa secciones en las que, por lo general, el PP gana de forma contundente a costa del PSOE. En ambos casos Izquierda Unida (IU) es la tercera fuerza política. En la siguiente tabla se muestra la distribución del voto para las cinco candidaturas más votadas en los 2 clusters y el resultado de toda la ciudad en esta convocatoria.

Candidatura	PSOE	PP	IU	CDS	LV
% G1	49	28.6	12.7	5.8	1.5
% G2	23.5	60.9	7.6	4.8	1.4
% Valladolid	38.7	41.5	10.8	5.5	1.5

Tabla 4.5: Distribución del voto global y de los clusters en el año 1991

A continuación, se identifican los individuos del cluster G1 desagregados por distritos, considerados los más representativos.

Distrito	Secciones
4	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23 24 25 26 27 28 29 30 32 33 34 35 36 37
6	3 5 8 9 10 11 12 13 14 15 16 17 19 20 22 23 24 25 26 27 28 29 30 32 33 34 35 36
7	6 7 8 9 10 11 12 13 14 15 16 19 20
8	2 4 6 7 8 9 10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
10	1 2 3 4 5 6 7 8 9 14 15 16 17 19 22 23 24 25 26 27 28 29 30 31 32 35 36
11	8 11 12 13 17 18 19 20 21 22 23 24 26 27 28 29
12	1 2 3

Tabla 4.6: Tabla de secciones representativas el año 1991

En la figura 4.4 se representa la situación espacial de las 158 secciones del cluster G1.

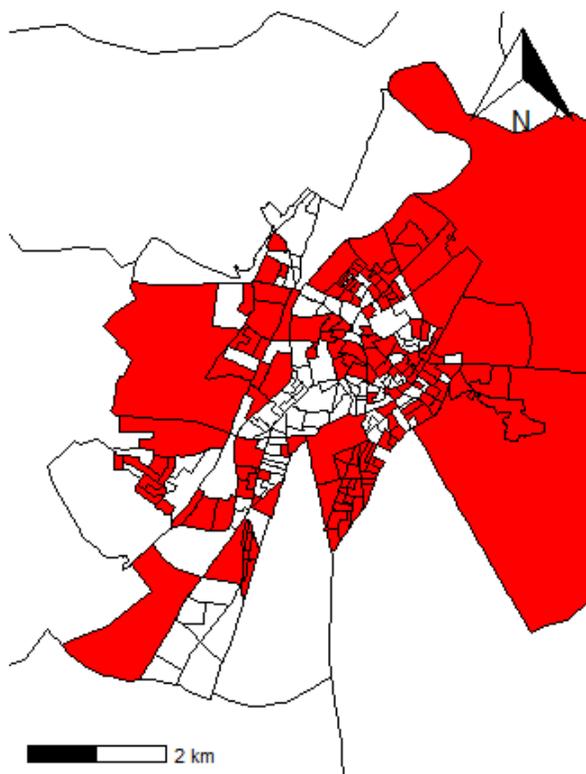
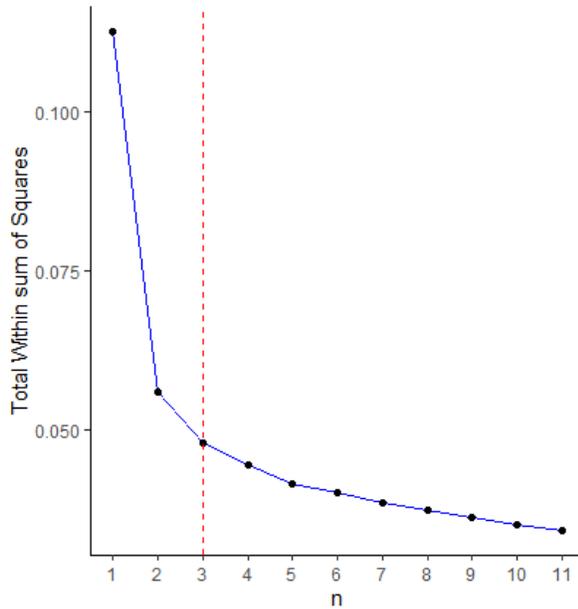


Figura 4.4: Mapa de secciones más representativas del año 1991

### 4.1.3. Elecciones 28 de mayo de 1995

En las elecciones del 28 de mayo de 1995 se presentaron 10 candidaturas, la división censal de Valladolid constaba de 257 secciones y 12 distritos.

En la figura 4.5 se representa la evolución de la suma de cuadrados total dentro de los grupos en función del número de grupos en los que se divida el conjunto de secciones. Se observa que este valor se estabiliza con una división en 3 grupos o más. La silueta media empeora cuantos más grupos se empleen por lo que el análisis cluster se realizará con 3 grupos.



nº grupos	TWSS
1	0.1127
2	0.0560
<b>3</b>	<b>0.0480</b>
4	0.0444
5	0.04149
6	0.0400
7	0.0385
8	0.0373
9	0.0361
10	0.0351
11	0.0341

(a) Gráfico con TWSS de 1995

(b) Tabla con TWSS de 1995

Figura 4.5: TWSS en función del nº de grupos 1995

La partición en 3 grupos obtenida se muestra a continuación:

G1	G2	G3	N	Silueta media	TWSS
86	69	102	257	0.4071	0.0480

Tabla 4.7: Tabla resumen de la división para el año 1995

La distribución de las secciones del primer cluster es similar al del computo de Valladolid y se caracteriza por tener al PP como partido más votado y al PSOE en segundo lugar. El cluster 2 tiene una estructura similar pero con una mayoría más amplia para el PP, alcanzando en torno al 65% de los votos. Por el contrario, en el tercer cluster generado se obtuvieron distribuciones en las que el PSOE ganó con algo menos de la mitad de los votos seguido del PP y posteriormente de IU y de Candidatura Independiente (CI).

A continuación, se muestra los porcentajes de voto promedio de las cinco candidaturas más votadas en las secciones censales de los 3 clusters junto con el resultado global de Valladolid.

Candidatura	PP	PSOE	IU	CI	AI
% G1	47.1	34.4	15.3	1	1
% G2	69	19.6	8.6	0.8	1.1
% G3	46.1	33.6	16.8	1.5	0.7
% Valladolid	48.3	34.5	14.1	1.1	0.9

Tabla 4.8: Distribución del voto global y de los clusters en el año 1995

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

En la figura 4.6 se representa la situación espacial de las secciones del cluster G1 y en la tabla 4.9 la lista de dichas secciones desagregadas por distritos.

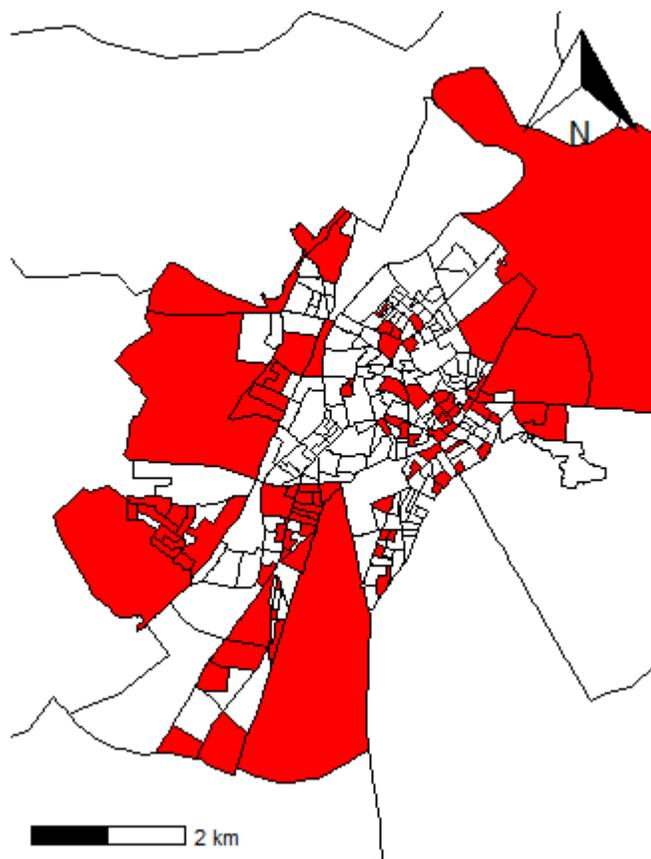


Figura 4.6: Mapa de secciones más representativas del año 1995

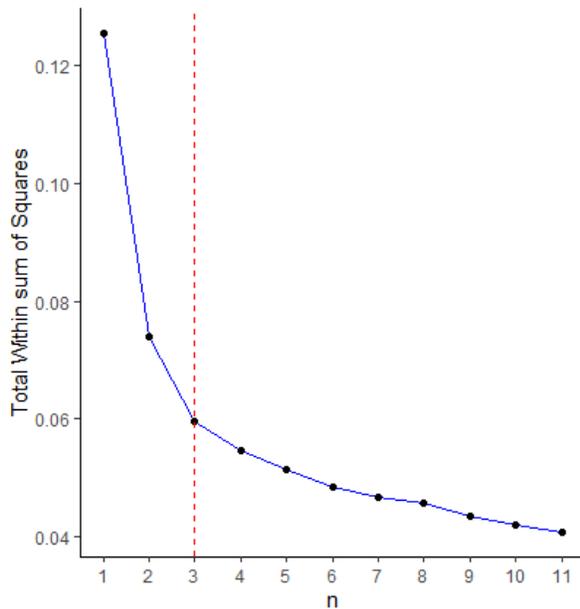
Distrito	Secciones
2	15 19
3	4 5 6
4	1 12 14 15 24 26 27 29 37
6	2 3 5 7 8 9 10 11 12 15 16 17 18 31 32 34 37
7	5 6 7 8 15
8	1 2 5 10 13 28 41
9	10 12
10	1 2 9 18 21 24 25 27 28 29 30 31 32 33 34 35 37 38 39 40 41 42
11	2 5 8 9 10 11 12 13 17 18 19 21 23 24 25 26 28 29
12	1

Tabla 4.9: Tabla de secciones representativas el año 1995

4.1.4. Elecciones 13 de junio de 1999

En las elecciones del 13 de junio de 1999 se presentaron 11 candidaturas, la división censal de Valladolid constaba de 258 secciones y 12 distritos.

Al igual que ocurría en el caso anterior, tras analizar la evolución de la suma de cuadrados interna total en función del número de clusters de la partición, a partir de 3 grupos se estabiliza este valor, por lo que también conviene aplicar el análisis con dicho valor. Con esta configuración la silueta media indica que la comodidad de las secciones en su grupo es algo peor pero sigue siendo aceptable.



nº grupos	TWSS
1	0.1255
2	0.07406
<b>3</b>	<b>0.0596</b>
4	0.0546
5	0.0513
6	0.0485
7	0.0466
8	0.0456
9	0.0434
10	0.0420
11	0.0408

(a) Gráfico con TWSS de 1999

(b) Tabla con TWSS de 1999

Figura 4.7: TWSS en función del nº de grupos 1999

El resumen de la distribución obtenida entre los 3 grupos se muestra a continuación:

G1	G2	G3	N	Silueta media	TWSS
88	58	112	258	0.3706	0.0596

Tabla 4.10: Tabla resumen de la división para el año 1999

La distribución específica de cada uno de los 3 clusters coincide en gran parte con las obtenidas en el análisis de la convocatoria del año 1995. El PP gana en los clusters G1 y G2, con mayor ventaja en el segundo caso alcanzando al rededor del 70% de los votos. En las secciones de la tercera agrupación por el contrario el primer partido en votos es el PSOE seguido del PP, también es en este último cluster donde IU obtuvo mejores porcentajes.

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

En la tabla 4.11 se muestra los porcentajes de voto promedio de las cinco candidaturas más votadas en los 3 clusters generados y del resultado de toda la ciudad.

Candidatura	PP	PSOE	IU	CI	PIE
% G1	48.5	36.7	8.1	2.5	0.5
% G2	70.6	19.3	4.8	4	0.4
% G3	33.2	49.7	9.2	4.5	1.2
% Valladolid	47.1	38.1	7.9	3.9	0.8

Tabla 4.11: Distribución del voto global y de los clusters en el año 1999

A continuación, se representa su situación espacial de las 88 secciones censales del cluster G1 sobre el mapa cartográfico de Valladolid y la lista con su correspondiente identificación desagregadas por distritos.

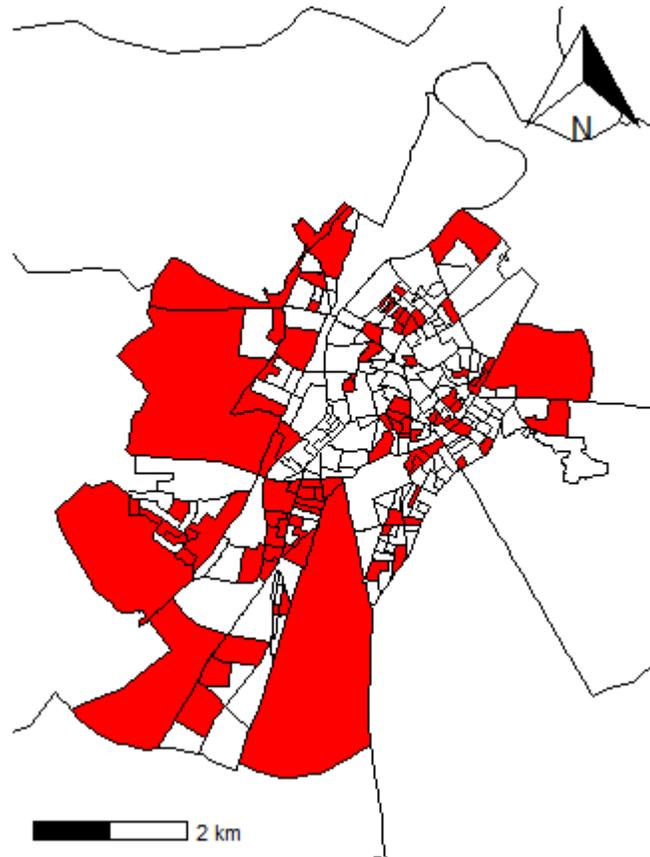


Figura 4.8: Mapa de secciones más representativas del año 1999

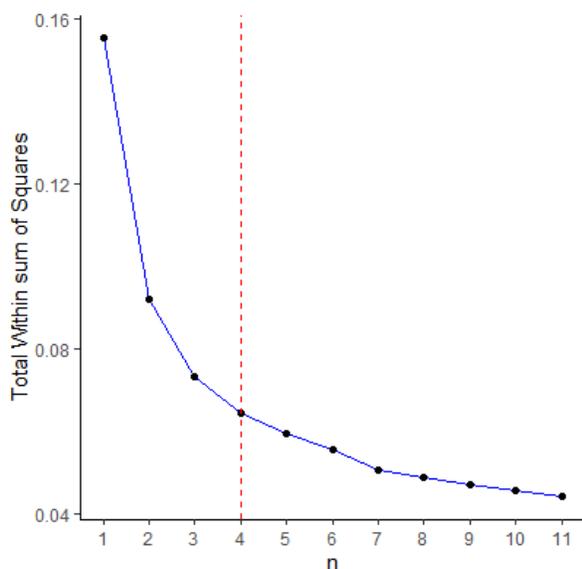
Distrito	Secciones
2	15 17 19
3	3 4 5 7
4	5 9 12 15 24 27 32 37
6	2 3 7 8 9 10 11 14 17 18
7	5 6 7 15
8	1 2 5 6 7 9 12 13 16 24 28 38 41
9	6 10 11 12
10	2 13 18 21 24 25 27 28 29 30 31 33 34 35 38 39 40 41 42 43 44
11	2 4 5 6 7 8 9 10 11 12 13 17 18 21 24 25 26 28 29 30
12	1

Tabla 4.12: Tabla de secciones representativas del año 1999

### 4.1.5. Elecciones 25 de mayo de 2003

En las elecciones del 25 de mayo de 2003 se presentaron 13 candidaturas, la división censal de Valladolid constaba de 261 secciones y 12 distritos.

En este caso la suma de cuadrados interna total empieza a converger con una partición formada por 4 grupos de secciones. Con esta división la silueta media obtenida es de 0.328096, esto implica que las asignaciones de las secciones a los grupos pierde calidad de forma leve en comparación con otras particiones con menos clusters pero se compensa al disminuir la variabilidad dentro de los grupos. Por todo ello la mejor opción es emplear 4 grupos en el análisis.



nº grupos	TWSS
1	0.1552
2	0.0922
3	0.07339
4	<b>0.0644</b>
5	0.05936
6	0.0554
7	0.0508
8	0.04874
9	0.04700
10	0.0455
11	0.0441

(a) Gráfico con TWSS de 2003

(b) Tabla con TWSS de 2003

Figura 4.9: TWSS en función del nº de grupos 2003

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

<b>G1</b>	G2	G3	G4	N	Silueta media	TWSS
<b>67</b>	71	110	13	261	0.328096	0.0644

Tabla 4.13: Tabla resumen de la división para el año 2003

En las secciones de los clusters G1 y G2 el PP fue el partido más votado, en el segundo caso se obtuvieron ventajas más importantes de dicho partido. Por otro lado, los otros dos grupos se caracterizan por tener al PSOE como candidatura más votada. La principal diferencia entre las dos últimas distribuciones comentadas es que en el cluster G4 el PSOE fue menos votado en favor de Corporación Participativa Ciudadana (CPC). A continuación, se muestra el porcentaje de voto promedio en los 4 clusters y del resultado global de Valladolid para las cinco candidaturas más votadas.

Candidatura	PP	PSOE	IU	CI	CPC
% G1	47.5	40.4	5.6	3.4	0.7
% G2	67.1	24.4	3.6	2.5	0.5
% G3	32.9	52.1	6.5	4.8	0.7
% G4	30.9	41.1	5	3.6	17.6
% Valladolid	45.7	40.8	5.4	3.7	1.8

Tabla 4.14: Distribución del voto global y de los clusters en el año 2003

A continuación, se muestra la tabla con las 70 secciones censales más representativas del año 2003.

Distrito	Secciones
2	16
3	4 5
4	1 3 12 15 18 24 37
6	2 3 7 8 9 10 11 17 18 34
7	5 6 7
8	1 2 5 6 12 13 40 41
9	6 10 12
10	2 18 21 24 25 27 28 29 30 34 37 38 39 41 42 43 44
11	5 8 11 12 13 17 18 21 25 26 28 29 31 35 36
12	1

Tabla 4.15: Tabla de secciones representativas del año 2003

## 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

En la figura 4.10 se representa a través de un mapa la situación espacial de los 4 grupos de secciones generados por el análisis cluster.

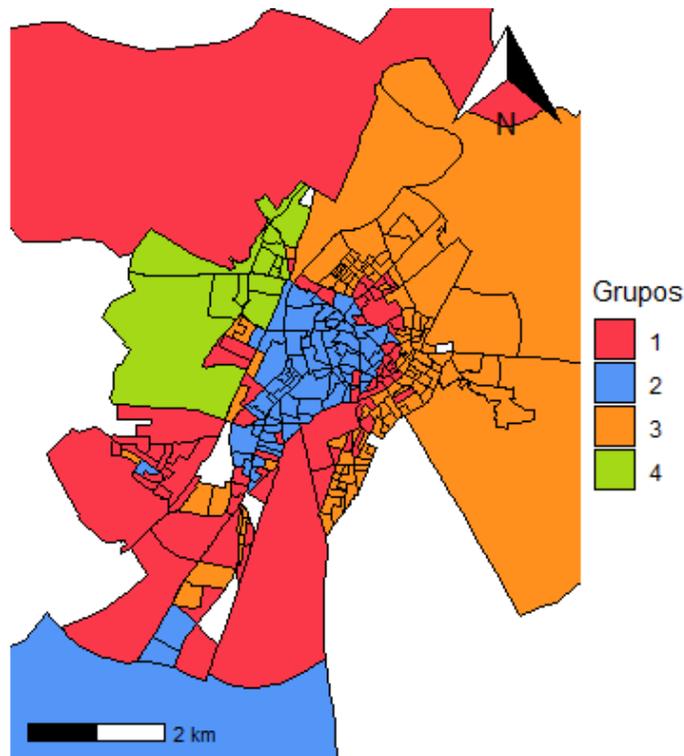
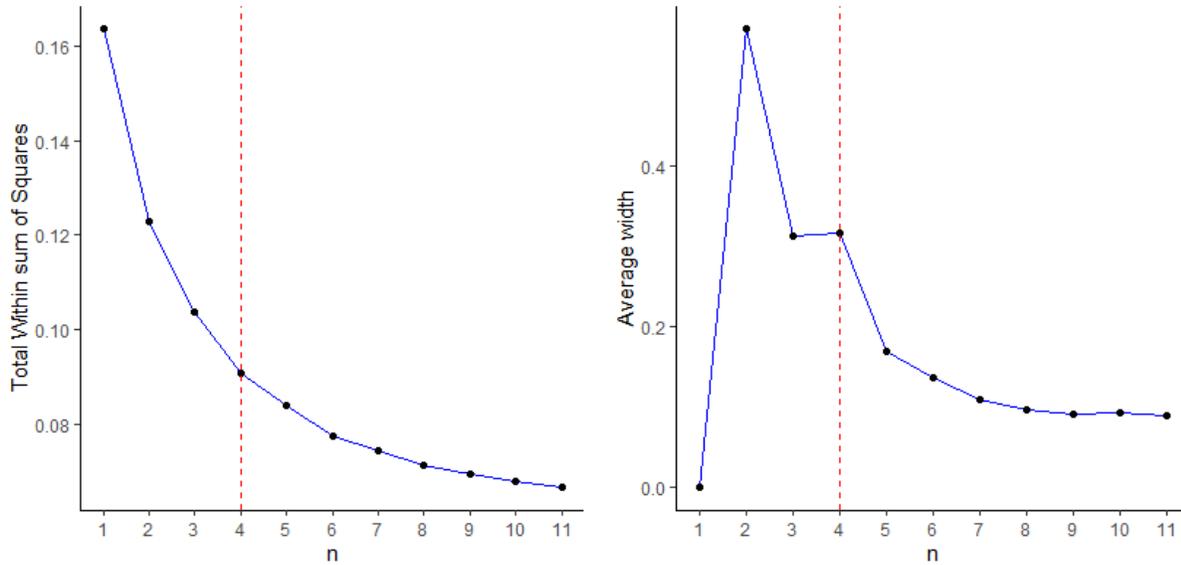


Figura 4.10: Mapa de secciones más representativas del año 2003

### 4.1.6. Elecciones 27 de mayo de 2007

En las elecciones del 27 de mayo de 2007 se presentaron 16 candidaturas, la división censal de Valladolid constaba de 268 secciones y 12 distritos.

En la figura 4.11 se muestra gráficamente la evolución de la suma de cuadrados interna total y la silueta media empleando desde un único grupo a 11. Se aprecia que, en un principio, no hay un número de clusters óptimo, pero la mejor opción puede ser realizar la división en 4 grupos. La justificación se basa en el buen equilibrio al que se llega teniendo los dos criterios en cuenta, la variabilidad dentro de los grupos es baja al mismo tiempo de que la claridad en la asignación de las secciones es moderadamente buena, incluso mejor que empleando 3 grupos.



(a) Gráfico con TWSS de 2007

(b) Gráfico de silueta media de 2007

Figura 4.11: TWSS y silueta media en función del nº de grupos 2007

G1	G2	G3	G4	N	Silueta media	TWSS
108	35	124	1	268	0.31602	0.09068

Tabla 4.16: Tabla resumen de la división para el año 2007

Al igual que ocurría en la convocatoria de 2003, los clusters G1 y G2 se caracterizan por tener una distribución en la que el PP es la primera fuerza política con la diferencia de que en las secciones del grupo G2 los votos a esta candidatura ocuparon tres cuartas partes del total. De nuevo, el cluster G3 se caracteriza por tener como primera fuerza al PSOE. La última agrupación generada está constituida por una única sección, su distribución destaca porque la candidatura más votada es el Partido Antitaurino Contra el Maltrato Animal (PACMA) seguido muy de cerca por el PP. Posiblemente se trate de un error, si bien se han revisado los datos y se ha comprobado que la información figura de ese modo en el archivo descargado.

En la tabla 4.17 se muestra el porcentaje de voto promedio de los 4 clusters para las cinco candidaturas más votadas junto con el resultado de toda la ciudad en esta convocatoria.

Candidatura	PP	PSOE	IU	CI	TC
% G1	54.7	34.8	5.9	2.3	0.5
% G2	74.3	18.4	3.9	1.4	0.3
% G3	35.5	50.8	6.6	4.6	0.6
% G4	43.5	0	7.4	2.9	0.2
% Valladolid	48.7	39.7	6	3.2	0.5

Tabla 4.17: Distribución del voto global y de los clusters en el año 2007

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

En la figura 4.12 se representa a través de un mapa la situación espacial de los 4 grupos de secciones generados por el análisis.

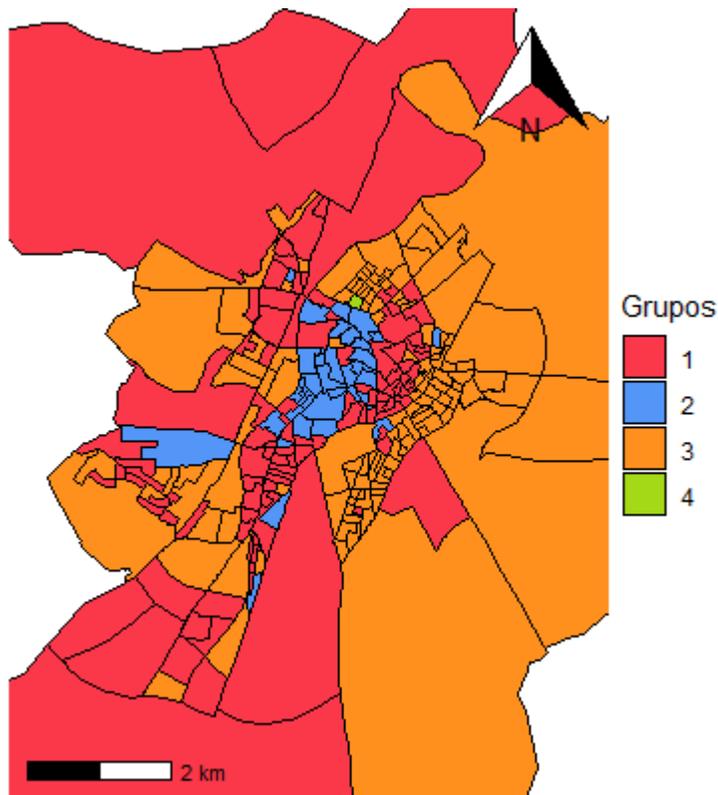


Figura 4.12: Mapa de secciones más representativas del año 2007

A continuación, se identifica las secciones asignadas al cluster G1.

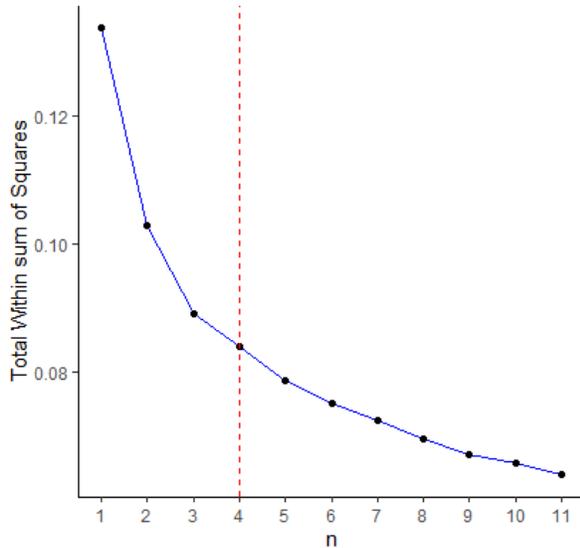
Distrito	Secciones
2	6 13 15 16 17 19
3	1 3 4 5 7
4	1 2 24 37
5	1
6	1 2 3 6 7 8 9 10 17 18 21
7	1 2 3 4 5 6 7 8 17
8	1 2 4 5 12 24 41
9	6 7 10 11 12
10	2 7 9 10 13 18 21 25 29 30 32 33 34 35 39 40 41 42 43 44 45 48 49 50 51
11	1 2 3 4 5 6 7 8 9 10 11 12 13 16 17 18 19 21 24 25 26 27 28 29 30 31 32 36 37 38 39 40
12	2 3 4

Tabla 4.18: Tabla de secciones representativas del año 2007

### 4.1.7. Elecciones 22 de mayo de 2011

En las elecciones del 22 de mayo de 2011 se presentaron 14 candidaturas, la división censal de Valladolid constaba de 270 secciones y 12 distritos.

En la figura 4.17 se muestra gráfica y analíticamente la evolución de la suma de cuadrados dentro de los grupos total. Se puede observar que la convergencia se alcanza cuando el conjunto se divide en 4 grupos, prácticamente la suma de cuadrados de las disimilaridades no se reduce cuando la partición se realiza en 2 o más grupos.



nº grupos	TWSS
1	0.1337
2	0.1029
3	0.0891
<b>4</b>	<b>0.0838</b>
5	0.0786
6	0.0751
7	0.0722
8	0.0694
9	0.0671
10	0.0657
11	0.0638

(a) Gráfico con TWSS de 2011

(b) Tabla con TWSS de 2011

Figura 4.13: TWSS en función del nº de grupos 2011

G1	G2	G3	G4	N	Silueta media	TWSS
70	67	61	72	270	0.1622	0.0838

Tabla 4.19: Tabla resumen de la división para el año 2011

Curiosamente, en los 4 clusters generados la candidatura más votada fue el PP seguida del PSOE, la diferencia principal entre las 4 distribuciones es la diferencia de porcentajes entre estos dos partidos políticos. En las secciones pertenecientes a los clusters G1 y G4 el resultado fue más igualado y la diferencia fue del 10% aproximadamente. En cambio, en el grupo G2 el Partido Popular gana las elecciones con una ventaja de 50 puntos. En los 4 clusters la tercera candidatura más votada fue IU.

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

La siguiente tabla presenta la repartición el voto promedio en los 4 clusters y en el resultado global en las cinco candidaturas más votadas.

Candidatura	PP	PSOE	IU	UPYD	CI
% G1	42.6	33.7	12.4	3.9	4.1
% G2	67.9	15.7	7.8	4.8	1.5
% G3	54.6	24.2	11.6	5.1	2.1
% G4	41.8	38.6	11.1	2.5	3.3
% Valladolid	51.8	27.7	10.8	4.2	2.7

Tabla 4.20: Distribución del voto global y de los clusters en el año 2011

En la figura 4.14 se representa el mapa de la ciudad dividido por los grupos de secciones censales generados tras realizar el análisis cluster. En rojo se representa el grupo con aquellas que tengan un comportamiento similar al del conjunto global de Valladolid.

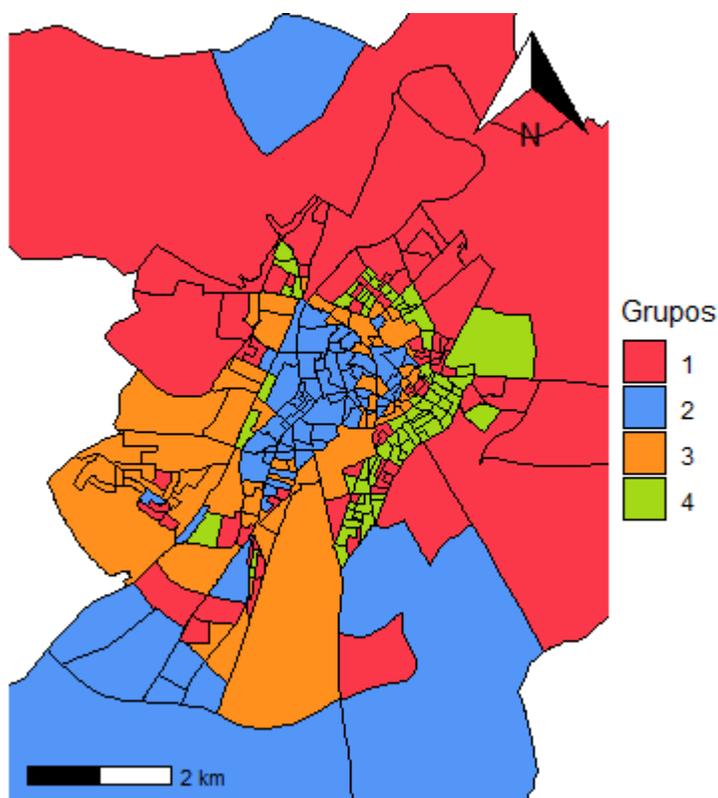


Figura 4.14: Mapa de secciones más representativas del año 2011

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

A continuación se muestra la tabla con las 70 secciones censales que se han considerado más representativas en la convocatoria del año 2011.

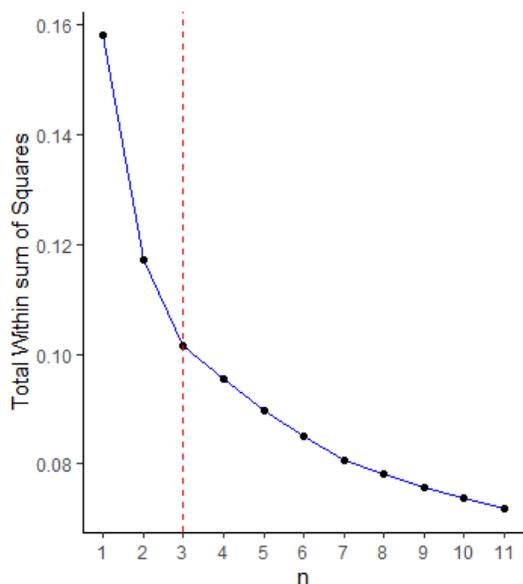
Distrito	Secciones
4	1 2 8 9 13 17 20 24 29 30 38 39 41
6	9 15 20 22 31 32 36 38
7	7 8 9 10 13 14 16 19 20
8	10 13 17 18 23 24 27 28 31 32 33 34 35 36
10	1 6 7 8 9 14 15 26 31 36 37 38 39 46 47
11	8 13 18 20 21 23 24 25 33 34
12	1

Tabla 4.21: Tabla de secciones representativas del año 2011

#### 4.1.8. Elecciones 24 de mayo de 2015

En las elecciones del 24 de mayo de 2015 se presentaron 15 candidaturas, la división censal de Valladolid constaba de 266 secciones y 12 distritos.

En la figura 4.17 se muestra gráfica y analíticamente la evolución de la suma de cuadrados dentro de los grupos total. La convergencia se hace más fuerte a partir de la división del conjunto en 3 grupos. Al igual que ocurría en anteriores casos la variabilidad dentro de los grupos disminuye cuantos más clusters se empleen pero la comodidad de las secciones en su grupo asignado empeora. Por este motivo el número de grupos elegido es 3.



(a) Gráfico con TWSS de 2015

n <sup>o</sup> grupos	TWSS
1	0.1578
2	0.1170
<b>3</b>	<b>0.1015</b>
4	0.0953
5	0.0897
6	0.0849
7	0.0807
8	0.0781
9	0.0758
10	0.0737
11	0.0718

(b) Tabla con TWSS de 2015

Figura 4.15: TWSS en función del n<sup>o</sup> de grupos 2015

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

G1	G2	G3	N	Silueta media	TWSS
112	64	90	266	0.2867	0.1015

Tabla 4.22: Tabla resumen de la división para el año 2015

En los resultados del año 2015 se aprecia la aparición de nuevos partidos con bastante relevancia en el panorama electoral como Valladolid Toma La Palabra (VTLP), Ciudadanos (C's) y Sí Se Puede Valladolid (SÍVA). La distribución del cluster G1 se caracteriza por su diversidad, el PP gana las elecciones pero con poca ventaja respecto al PSOE y al resto de partidos nombrados. Como ocurría en convocatorias anteriores, en las secciones del cluster G2 se observa un perfil más conservador ya que el PP obtiene una mayoría más abultada y los porcentajes de voto de partidos con carácter progresista son reducidos. Por último, el tercer cluster generado tienen como primera fuerza política al PSOE seguido del PP y posteriormente de VTLP, SÍVA y C's.

La siguiente tabla muestra el porcentaje de voto promedio entre las cinco candidaturas más votadas en los 3 clusters generados y en el resultado global de Valladolid.

Candidatura	PP	PSOE	VTLP	SÍVA	C's
% G1	34.1	22.8	15.4	11	8
% G2	53.3	12.5	11	5.7	9.7
% G3	26.5	35.4	12.4	12.3	4.7
% Valladolid	36.3	23.6	13.6	10.1	7.8

Tabla 4.23: Distribución del voto global y de los clusters en el año 2015

A continuación, podemos identificar las 112 secciones que se han considerado las más representativas en la convocatoria del año 2015.

Distrito	Secciones
2	15 19
3	4 5
4	1 2 3 13 14 17 29 32 34 37 40 41
6	2 3 5 7 9 10 11 12 15 18 31 32 37
7	5 6 7 8 9 11 14 15 19 22
8	1 2 4 5 9 24 35 41
9	6 10 11 12
10	1 2 4 6 7 8 9 14 16 18 21 24 27 28 29 30 32 33 34 35 37 38 39 40 42 43 44 45 46 47 48 50 51 52 53 54 55 56
11	2 5 9 10 12 17 18 20 21 25 27 28 29 33 34 35 36 37 38 40 42 43
12	1

Tabla 4.24: Tabla de secciones representativas del año 2015

## 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

En la figura 4.16 se representa el mapa de la ciudad dividido por los grupos de secciones censales generados tras realizar el análisis cluster. En rojo se representa el grupo con aquellas que tengan un comportamiento similar al del conjunto global de Valladolid.

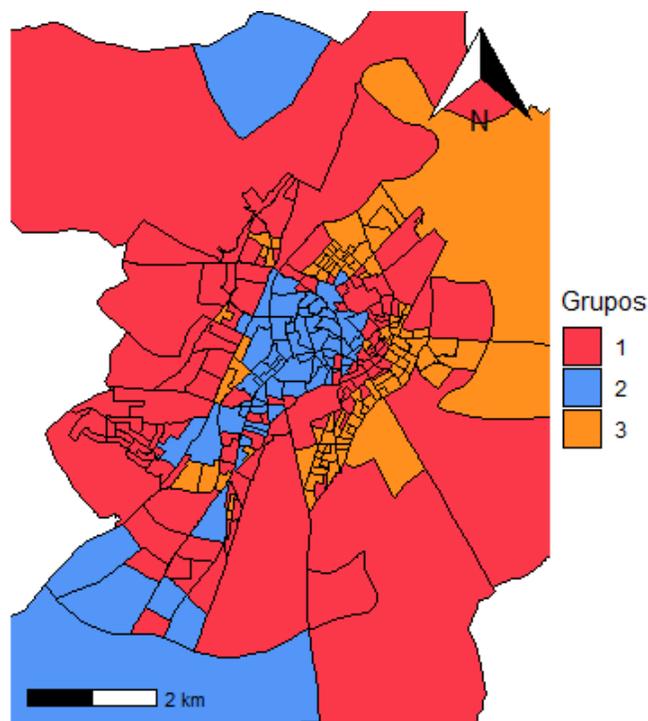


Figura 4.16: Mapa de secciones más representativas del año 2015

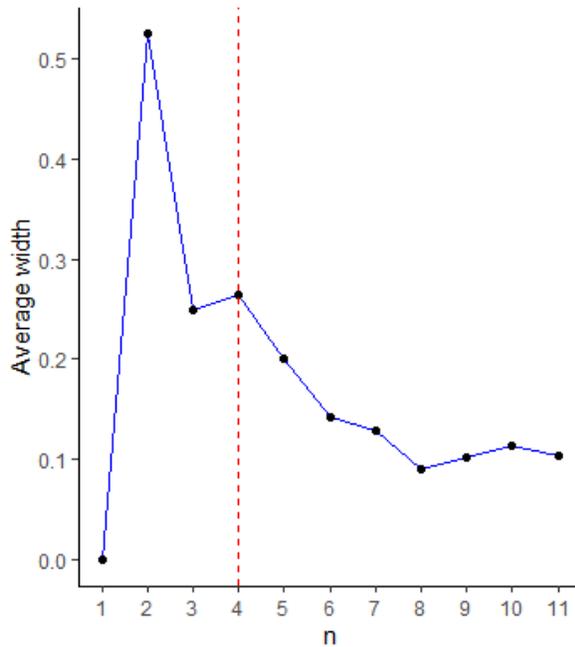
### 4.1.9. Elecciones 26 de mayo de 2019

En las elecciones del 26 de mayo de 2019 se presentaron 14 candidaturas, la división censal de Valladolid constaba de 258 secciones y 12 distritos.

En la figura 4.17 se muestra gráfica y analíticamente la evolución de la suma de cuadrados dentro de los grupos total. La convergencia se hace más fuerte a partir de la división del conjunto en 3 grupos. Al igual que ocurría en anteriores casos la variabilidad dentro de los grupos disminuye cuantos más clusters se empleen pero la comodidad de las secciones en su grupo asignado empeora. Por este motivo el número de grupos elegido es 4.

G1	G2	G3	G4	N	Silueta media	TWSS
64	47	89	58	258	0.26455	0.07401

Tabla 4.25: Tabla resumen de la división para el año 2019



(a) Gráfico con TWSS de 2019

nº grupos	Silueta media
1	0
2	0.5239
3	0.2486
<b>4</b>	<b>0.26455</b>
5	0.20105
6	0.14261
7	0.1284
8	0.0906
9	0.10195
10	0.11392
11	0.10342

(b) Tabla con TWSS de 2019

Figura 4.17: TWSS en función del nº de grupos 2019

En cuanto a las distribuciones particulares de las agrupaciones generadas podemos decir que en la distribución del cluster G1 los resultados entre el PSOE y el PP son muy próximos, los siguientes dos partidos más votados son C's y VTLP, también con diferencias no significativas en cuanto a porcentaje de votos entre ambas. En el segundo cluster el PP obtuvo más de la mitad de los votos y por el contrario PSOE y C's obtuvieron sus peores porcentajes. En cambio, en esta agrupación VOX alcanza su mejor resultado alcanzando más del 11 %. En los otros dos clusters el PSOE es la primera fuerza política obteniendo una ventaja significativa frente al segundo partido más votado, el PP. Cabe destacar que C's se mantiene en el tercer puesto en las 4 distribuciones.

La tabla 4.26 muestra el porcentaje de voto promedio entre las cinco candidaturas más votadas en los 4 clusters junto con el resultado de toda la ciudad.

Candidatura	PSOE	PP	C's	VTLP	VOX
% G1	31	33	14.3	11.4	6.8
% G2	19.9	47.3	12.1	8.5	9.2
% G3	43.1	23.5	12.1	11	5.5
% G4	48.1	22.3	10	9.4	4.7
% Valladolid	35.9	30.3	12.7	10.5	6.4

Tabla 4.26: Distribución del voto global y de los clusters en el año 2019

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

En la figura 4.18 se representa el mapa de la ciudad dividido por los grupos de secciones censales generados tras realizar el análisis cluster. En rojo se representa el grupo objetivo en el cual se encuentran aquellas secciones que tengan un comportamiento similar al del conjunto global de Valladolid.

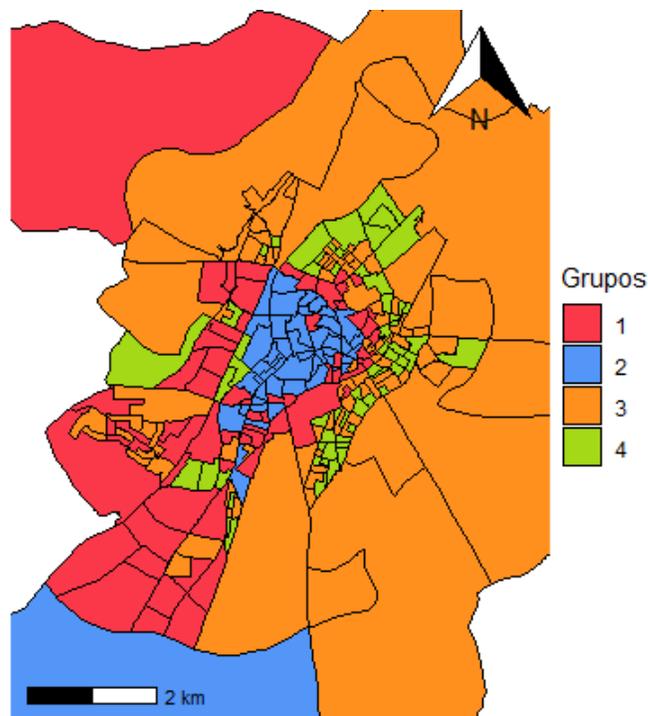


Figura 4.18: Mapa de secciones más representativas del año 2019

En la siguiente tabla se muestra el conjunto de secciones censales asignadas al cluster objetivo G1 desagregadas por los distritos de la ciudad.

Distrito	Secciones
2	15 17 19
3	4 5 7
4	2 37
5	1
6	3 7 10 18
7	4 5 6
8	1 5 12
9	6 7 10 11
10	7 13 18 21 29 30 32 33 34 35 40 41 43 44 48 49 52 54 55 56
11	2 4 5 6 9 12 17 21 26 28 29 31 32 35 36 37 38 39 41 43 43
12	4

Tabla 4.27: Tabla de secciones representativas del año 2019

### 4.1.10. Secciones próximas al perfil global de Valladolid

Una vez realizado el análisis de cada una de las nueve elecciones se ha obtenido una selección de las secciones censales con el comportamiento más similar al resultado de toda la ciudad en cuanto a la distribución del voto en cada convocatoria.

Llegado este punto nos planteamos identificar aquellos elementos que pertenecen a la intersección de las nueve listas creadas. En este sentido solo existe una sección que cumpla la anterior característica, se encuentra en el **distrito 11 sección 21**.

Para completar el análisis se ha calculado la frecuencia de ocasiones en las que aparece cada sección como representativa a lo largo de las nueve convocatorias. La tabla 4.28 muestra información acerca de la distribución del número de secciones representativas según el número de convocatorias en las que lo son. De esta forma podemos saber que la mayoría de secciones son representativas en una o dos ocasiones en total, o que existen 8 secciones que aparecen como representativas en todas las convocatorias salvo en una.

Nº convocatorias	1	2	3	4	5	6	7	8	9
Nº secciones	64	68	28	21	19	23	24	8	1

Tabla 4.28: Tabla de frecuencias de secciones representativas aplicando 4.1

Consideramos como conjunto de secciones más importantes para el estudio el formado tanto por aquellas que aparecen en 8 ocasiones como la sección indicada anteriormente que es representativa en todas las convocatorias electorales. En la tabla 4.29 de detalla este conjunto además de los años concretos en donde cada una de ellas ha aparecido en la lista.

Distrito	Secciones	Nº convocatorias	Años representativa
11	21	9	1987, 1991, 1995, 1999, 2003, 2007, 2011, 2015, 2019
6	9	8	1987, 1991, 1995, 1999, 2003, 2007, 2011, 2015
6	10	8	1987, 1991, 1995, 1999, 2003, 2007, 2015, 2019
7	6	8	1987, 1991, 1995, 1999, 2003, 2007, 2015, 2019
7	7	8	1987, 1991, 1995, 1999, 2003, 2007, 2011, 2015
11	12	8	1987, 1991, 1995, 1999, 2003, 2007, 2015, 2019
11	18	8	1987, 1991, 1995, 1999, 2003, 2007, 2011, 2015
11	28	8	1987, 1991, 1995, 1999, 2003, 2007, 2015, 2019
11	29	8	1987, 1991, 1995, 1999, 2003, 2007, 2015, 2019

Tabla 4.29: Tabla de secciones más representativas aplicando 4.1

Cabe destacar que de entre las secciones que son representativas en 8 elecciones solo 5 son representativas en las dos últimas convocatorias (años 2015 y 2019) y que en ningún caso lo son en las últimas tres (años 2011, 2015 y 2019).

#### 4.1. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN ÓPTIMA

---

En este sentido y con el interés fijado en futuras elecciones se incide en las secciones más importantes de las últimas convocatorias, para lo cual se realiza la última parte del análisis utilizando los resultados electorales de los años 2007, 2011, 2015 y 2019.

En la tabla 4.30 se representa la nueva disposición del número de secciones representativas en función del número de ocasiones que lo son, en este caso como máximo pueden aparecer en 4 convocatorias.

n <sup>o</sup> convocatorias	1	2	3	4
n <sup>o</sup> secciones	77	62	47	3

Tabla 4.30: Tabla de frecuencias de secciones representativas aplicando 4.1 a las cuatro últimas convocatorias

Se observa que tan solo existen 3 secciones censales con una distribución similar al perfil global en las últimas 4 convocatorias de elecciones municipales (tabla 4.31). En cambio, el conjunto de individuos representativos en 3 de las 4 ocasiones asciende a 47 integrantes.

Distrito	Secciones	N <sup>o</sup> ocasiones	Años representativa
4	2	4	2007, 2011, 2015, 2019
10	7	4	2007, 2011, 2015, 2019
11	21	4	2007, 2011, 2015, 2019

Tabla 4.31: Tabla de secciones más representativas aplicando 4.1 a las cuatro últimas convocatorias

Exceptuando la sección 21 del distrito 11 que ya conocíamos por ser la única que es representativa en todas la convocatorias desde 1987 hasta 2019, se observan dos nuevos individuos a tener en cuenta de cara a futuras elecciones. Dado que las secciones 2 y 7 de los distritos 4 y 10 respectivamente han sido representativas en las 4 anteriores elecciones municipales es probable que también lo sean en las próximas convocatorias ya que la evolución demográfica de la ciudad en los últimos años es estable.

## 4.2. Análisis no jerarquizado con partición en 20 clusters

En este caso, se realiza de nuevo un análisis no jerarquizado con la particularidad de que el número de grupos utilizado está prefijado en 20 y en todas las convocatorias electorales estudiadas será el mismo. El principal propósito de agrupar las secciones en 20 divisiones es establecer un conjunto más reducido de secciones censales representativas en cada año. De esta forma el perfil de aquellas que pertenezcan al grupo objetivo será más preciso en cuanto a la cercanía respecto a la distribución del resultado de toda la ciudad.

Para realizar el procedimiento se va a utilizar la función de R *pam* del mismo modo que en la sección previa. También se implementa la matriz de distancias de Kullback-Leibler como índice de disimilaridad entre las secciones censales.

Los resultados obtenidos aplicando el algoritmo descrito sobre los resultados de las elecciones municipales disputadas entre el año 1987 y 2019 se exponen en las siguientes subsecciones. En cada uno de los casos la sección artificial con los resultados globales va a pertenecer al **grupo 1** (en color rojo), por lo tanto la secciones censales más representativas van a ser aquellas que estén incluidas en dicho grupo.

### 4.2.1. Elecciones 10 de junio de 1987

Agrupando en 20 clusters las secciones censales presentes en el año 1987 se obtiene la siguiente partición:

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>19</b>	13	4	8	26	13	4	13	18	5
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
27	29	11	12	23	1	1	11	9	1

Tabla 4.32: Tabla resumen de la división para el año 1987 aplicando 4.2

Se puede observar que 3 de los grupos están formados por una única observación y que existen 19 secciones con un comportamiento parecido al resultado general ya que es el número de individuos del grupo 1. En la tabla 4.33 se identifican estas 19 secciones censales.

Distrito	Secciones
3	1 6
4	24
6	9 10 11 17
7	6
8	24 25 28
10	26
11	2 11 13 23 26 28 29

Tabla 4.33: Tabla de secciones representativas del año 1987 aplicando 4.2

### 4.2.2. Elecciones 26 de mayo de 1991

El análisis cluster aplicado sobre el conjunto de secciones censales del año 1991 genera la siguiente partición.

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>20</b>	15	8	15	14	5	20	7	16	23
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
17	18	14	10	7	7	16	13	2	2

Tabla 4.34: Tabla resumen de la división para el año 1991 aplicando 4.2

En este caso la distribución es levemente más uniforme en los 20 grupos en comparación con el resultado de las elecciones de 1987. El grupo 1 está formado por 20 secciones, las cuales se muestran en detalle en la tabla 4.35.

Distrito	Secciones
4	27
6	3 9
7	5 6
8	9 13 28 41
9	10
10	25 33
11	2 5 8 12 13 23 25 28

Tabla 4.35: Tabla de secciones representativas del año 1991 aplicando 4.2

### 4.2.3. Elecciones 28 de mayo de 1995

En la siguiente tabla se puede ver la distribución de los clusters generados en el análisis para los resultados del año 1995.

<b>G1</b> 27	G2 10	G3 17	G4 24	G5 11	G6 8	G7 12	G8 5	G9 15	G10 29
G11 14	G12 18	G13 20	G14 11	G15 12	G16 2	G17 10	G18 9	G19 2	G20 1

Tabla 4.36: Tabla resumen de la división para el año 1995 aplicando 4.2

El grupo 1 esta formado por 27 secciones censales, las cuales se detallan en la tabla 4.37.

Distrito	Secciones
3	6
4	15 37
6	8 17
7	5
8	2 5 13 28
10	18 24 27 28 32 33 34 35 38
11	5 9 19 24 25 26 28 29

Tabla 4.37: Tabla de secciones representativas del año 1995 aplicando 4.2

### 4.2.4. Elecciones 13 de junio de 1999

Para los resultados electorales del año 1999 se obtiene la siguiente partición del conjunto de secciones totales.

<b>G1</b> 35	G2 16	G3 11	G4 17	G5 12	G6 13	G7 12	G8 13	G9 21	G10 22
G11 24	G12 9	G13 16	G14 11	G15 9	G16 3	G17 3	G18 9	G19 1	G20 1

Tabla 4.38: Tabla resumen de la división para el año 1999 aplicando 4.2

El grupo 1 es el más numeroso de todos los clusters generados, las siguientes 35 secciones tienen una distribución del voto análoga al de toda la ciudad.

## 4.2. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN EN 20 CLUSTERS

Distrito	Secciones
3	5
4	29 37
6	2 3 8 10 11
7	5 6 7
8	2 5 24 41
9	10
10	2 21 27 28 29 33 34 35 39 41 42 43
11	8 12 13 18 21 25 26

Tabla 4.39: Tabla de secciones representativas del año 1999 aplicando 4.2

### 4.2.5. Elecciones 25 de mayo de 2003

Agrupando en 20 clusters las secciones censales presentes en el año 2003 se obtiene la siguiente partición:

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>16</b>	13	21	16	16	15	17	23	11	24
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
25	28	8	5	1	3	14	3	1	1

Tabla 4.40: Tabla resumen de la división para el año 2003 aplicando 4.2

Las secciones se reparten de forma desigual, existen clusters formados desde un solo integrante hasta 28 en esta partición. Por otro lado, hay 16 secciones con distribuciones representativas, se muestran detalladamente a continuación.

Distrito	Secciones
3	7
4	37
6	2 3 7 18
7	5 6
8	5 12
10	21 34 43
11	2 26 36

Tabla 4.41: Tabla de secciones representativas del año 2003 aplicando 4.2

### 4.2.6. Elecciones 27 de mayo de 2007

Tras aplicar el análisis cluster con 20 grupos a los resultados del año 2007 se genera la siguiente distribución de las secciones.

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>25</b>	16	14	24	9	27	16	14	13	16
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
20	9	11	6	4	12	16	12	3	1

Tabla 4.42: Tabla resumen de la división para el año 2007 aplicando 4.2

El grupo 1 esta formado por 25 individuos, se corresponden con las secciones más representativas de Valladolid en este año electoral.

Distrito	Secciones
4	2 15 24 37
6	3 9 10
8	2 4 41
10	7 34 35 42 45
11	8 9 12 13 17 18 19 25 33
12	3

Tabla 4.43: Tabla de secciones representativas del año 2007 aplicando 4.2

### 4.2.7. Elecciones 22 de mayo de 2011

En la siguiente tabla se puede ver la distribución de los clusters generados en el análisis para los resultados del año 2011.

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>31</b>	14	9	13	15	18	18	23	26	15
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
19	16	6	13	1	11	11	7	2	1

Tabla 4.44: Tabla resumen de la división para el año 2011 aplicando 4.2

El cluster con las secciones cuyas distribuciones son más representativas es el más numeroso y está integrado por las siguientes 31 secciones.

Distrito	Secciones
2	15
6	9 10 11 17 31
8	1 2 5 41
9	11
10	1 7 21 27 33 35 38 39 42 43 46
11	8 10 17 18 25 26 27 33 36

Tabla 4.45: Tabla de secciones representativas del año 2011 aplicando 4.2

### 4.2.8. Elecciones 24 de mayo de 2015

En la siguiente tabla se representa la distribución de los clusters generados en el análisis para los resultados electorales del 2015.

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>31</b>	14	4	12	9	13	12	28	16	21
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
20	22	14	15	8	19	2	4	1	1

Tabla 4.46: Tabla resumen de la división para el año 2015 aplicando 4.2

El grupo 1 esta formado por 31 secciones censales, lo que le convierte en el más poblado de la partición. En la tabla 4.47 se identifican dichas secciones.

Distrito	Secciones
2	15
3	4
4	2 3 34 37
6	3 7 9 10 32 37
7	5 6
8	5 24 41
9	10
10	1 8 18 27 32 33 35
11	10 17 18 25 28 40

Tabla 4.47: Tabla de secciones representativas del año 2015 aplicando 4.2

### 4.2.9. Elecciones 26 de mayo de 2019

Por último se representa la distribución del análisis cluster con 20 grupos para este año.

<b>G1</b>	G2	G3	G4	G5	G6	G7	G8	G9	G10
<b>30</b>	10	12	9	26	14	18	16	20	23
G11	G12	G13	G14	G15	G16	G17	G18	G19	G20
1	6	18	7	13	16	7	10	1	1

Tabla 4.48: Tabla resumen de la división para el año 2019 aplicando 4.2

Nuevamente el cluster formado por más secciones es el que contiene la sección artificial y por tanto al que pertenecen las secciones con las distribución del voto más parecida al resultado global. Esta compuesto por las siguientes 30 secciones.

Distrito	Secciones
2	15
3	5
4	2 14 16 37 40
6	5 7 10
8	2 4
9	11
10	1 25 27 28 32 33 42 43 50 54
11	9 10 12 13 26 29 38

Tabla 4.49: Tabla de secciones representativas del año 2019 aplicando 4.2

### 4.2.10. Secciones próximas al perfil global de Valladolid

Al igual que ocurriría con el análisis cluster aplicado sobre la partición óptima, para conocer las secciones censales que pueden considerarse más representativas del perfil global de Valladolid se realiza la intersección de las listas generadas anteriormente en cada una de las elecciones estudiadas.

En este caso, al realizar la partición con un número de clusters mayor (20) las listas de secciones más representativas en cada año son más reducidas ya que existen más grupos en los que poder distribuirse. Como consecuencia la intersección de estas listas es vacía, ninguna sección censal tiene una distribución del voto similar al perfil global en las 9 nueve elecciones.

En vistas de los resultados obtenidos se ha calculado la frecuencias de ocasiones en las que aparece cada sección como representativa a lo largo de las nueve convocatorias. La tabla 4.50 muestra información acerca de la distribución del número de secciones representativas según el número de ocasiones en las que lo son. De esta forma podemos saber que 37 secciones son representativas en solo una de las nueve elecciones, o que no hay ninguna sección que aparezca como representativa en más de 6 ocasiones.

n <sup>o</sup> convocatorias	1	2	3	4	5	6	7	8	9
n <sup>o</sup> secciones	37	19	17	7	10	5	0	0	0

Tabla 4.50: Tabla de frecuencias de secciones representativas aplicando 4.2

Las secciones con más interés son aquellas que aparecen un mayor número de años electorales en estas listas, en la tabla 4.51 se detalla las secciones cuya distribución del voto ha sido similar al de toda la ciudad en más ocasiones y los años concretos en donde cada una de ellas ha sido representativa.

Distrito	Secciones	N <sup>o</sup> convocatorias	Años representativa
4	37	6	1995, 1999, 2003, 2007, 2015, 2019
<b>6</b>	<b>10</b>	6	1987, 1999, 2007, 2011, 2015, 2019
10	33	6	1991, 1995, 1999, 2011, 2015, 2019
11	25	6	1991, 1995, 1999, 2007, 2011, 2015
11	26	6	1987, 1995, 1999, 2003, 2011, 2019

Tabla 4.51: Tabla de secciones más representativas aplicando 4.2

Es importante observar la evolución de las convocatorias en las que las anteriores secciones son representativas. En este aspecto es destacable que la **sección 10 del distrito 6** se mantiene en el cluster objetivo desde la convocatoria del año 2007 hasta la última, celebrada en 2019. Del mismo modo la sección 33 del distrito 10 es representativa en las últimas 3 convocatorias.

## 4.2. ANÁLISIS NO JERARQUIZADO CON PARTICIÓN EN 20 CLUSTERS

Al igual que en la sección 4.1, se va a repetir el análisis del perfil global de Valladolid restringiendo los resultados electorales a las 4 últimas convocatorias, correspondientes a los años 2007, 2011, 2015 y 2019. En la tabla 4.52 se representa la nueva disposición del número de secciones representativas en función del número de veces en las que lo son, en este caso como máximo pueden aparecer en 4 convocatorias.

n <sup>o</sup> convocatorias	1	2	3	4
n <sup>o</sup> secciones	40	14	15	1

Tabla 4.52: Tabla de frecuencias de secciones representativas aplicando 4.2 a las últimas cuatro convocatorias

Los resultados confirman que la sección 10 del distrito 6 es la única que se mantiene representativa en las últimas 4 ocasiones. La tabla 4.53 identifica dicha sección junto con los 15 individuos que aparecen en el cluster objetivo en 3 de las 4 ocasiones estudiadas.

Distrito	Secciones	N <sup>o</sup> convocatorias	Años representativa
6	10	4	2007, 2011, 2015, 2019
<b>2</b>	<b>15</b>	3	2011, 2015, 2019
4	2	3	2007, 2015, 2019
4	37	3	2007, 2015, 2019
6	9	3	2007, 2011, 2015
8	2	3	2007, 2011, 2019
8	41	3	2007, 2011, 2015
<b>10</b>	<b>1</b>	3	2011, 2015, 2019
<b>10</b>	<b>27</b>	3	2011, 2015, 2019
<b>10</b>	<b>33</b>	3	2011, 2015, 2019
10	35	3	2007, 2011, 2015
10	42	3	2007, 2011, 2019
<b>11</b>	<b>10</b>	3	2011, 2015, 2019
11	17	3	2007, 2011, 2015
11	18	3	2007, 2011, 2015
11	25	3	2007, 2011, 2015

Tabla 4.53: Tabla de secciones más representativas aplicando 4.2 a las últimas cuatro convocatorias

De entre las secciones reunidas en la tabla anterior que son representativas en 3 convocatorias, es importante destacar aquellas 5 cuya distribución del voto se ha mantenido similar al resultado global de la ciudad en las últimas 3 ocasiones. Estos sectores de Valladolid se ubican en la sección 15 del distrito 2, la sección 10 del distrito 11 y por último dentro del distrito 10, las secciones 1, 27 y 33.

### 4.3. Análisis jerarquizado

Como se explicó en la sección 3.1, Los análisis cluster jerárquicos son métodos iterativos de agrupamientos sucesivos que no requieren establecer previamente el número de grupos en los que particionar el conjunto de objetos de interés, sin embargo, es necesario aportar el índice de agregación que se desea utilizar en el agrupamiento. En este caso, al igual que ocurría en los anteriores análisis realizados, el conjunto de objetos a agrupar está formado por las secciones censales de Valladolid y una sección artificial que almacena los resultados globales de toda la ciudad.

Para conocer la estructura de grupos que existe en los resultados electorales de cada una de las 9 elecciones estudiadas se ha empleado el algoritmo de agrupamiento jerarquizado implementado en la función *agnes()* del paquete *cluster* [8], incluyendo la matriz de distancias de Kullback-Leibler como matriz de disimilaridades.

De entre los posibles métodos de agrupamiento (ó índices de agregación) interesa aquel que genere una partición más firme. Una forma de cuantificar esta característica es utilizar el coeficiente de agregación (*ca*). Sabiendo que *n* es el número de observaciones que se quieren agrupar **ca** se define como:

$$ca = \frac{1}{n} \sum_{i=1}^n 1 - m(i)$$

En donde, para cada observación *i*, *m(i)* es igual a la disimilaridad con el primer cluster al que es añadida, dividida entre la disimilaridad del agrupamiento en el último paso del algoritmo. De esta forma se ha calculado la media del coeficiente de agregación obtenido en las 9 elecciones por medio de los métodos *single linkage*, *complete linkage*, *average linkage* y *Ward*. Los resultados obtenidos son los siguientes.

Método	single linkage	complete linkage	average linkage	<b>Ward</b>
ca	0.94040	0.8817	0.9544	<b>0.9843</b>

Tabla 4.54: Tabla de coeficientes de agregación

Como se puede observar en la tabla 4.54, el método de agregación que mayor **ca** produce es el llamado *Ward*, cuyo criterio es minimizar el total de la varianza interna de los clusters. En cada paso del algoritmo se fusiona la pareja de clusters con menor distancia entre grupos [12].

Una vez aplicado el procedimiento de agrupamiento descrito, el objetivo es encontrar aquellas secciones con comportamientos similares al resultado global de Valladolid. Para alcanzar este propósito se necesita identificar las secciones que se encuentran en el cluster al que se ha anexionado la sección artificial en el momento concreto de dicha anexión.

### 4.3. ANÁLISIS JERARQUIZADO

---

El proceso se aplica a cada conjunto de resultados electorales por separado y posteriormente se realiza la intersección de las 9 listas de secciones generadas para tener una visión más general y fiable de las secciones representativas.

En la tabla 4.55 se detallan los resultados del análisis anterior utilizando la función de *R agnes*, mostrando cada una de las listas de secciones representativas de cada año evaluado.

Año	Distrito	Sección
1987	6	2
1991	11	13
1995	<b>8</b>	<b>2</b>
1999	<b>11</b>	<b>18</b>
2003	10	31
2007	6	10
	10	35
	<b>11</b>	<b>18</b>
2011	<b>8</b>	<b>2</b>
	10	27
	<b>10</b>	<b>33</b>
	10	38
2015	11	40
2019	<b>10</b>	<b>33</b>

Tabla 4.55: Tabla con secciones más representativas aplicando 4.3

Como se puede ver, en la mayoría de convocatorias se considera una sola sección censal representativa, esto ocurre porque dicha sección es el único integrante del cluster en el momento en que se fusiona la sección con los resultados globales de esa convocatoria electoral.

Solo existen tres secciones que aparecen en más de una convocatoria. Todas ellas son representativas en dos convocatorias electorales, la primera se corresponde con el **distrito 8 sección 2** (años 1995 y 2011), la segunda con el **distrito 10 sección 33** (años 2011 y 2019) y la tercera con el **distrito 11 sección 18** (años 1999 y 2007). Estos tres elementos también se obtuvieron aplicando la técnica de análisis cluster no jerarquizado con una partición en 20 grupos (sección 4.2). En dicho análisis se obtuvo que la segunda sección descrita tuvo un comportamiento análogo al resultado de la ciudad en 6 ocasiones diferentes, del mismo modo que la sección 10 del distrito 6. Para las otras dos secciones repetidas en la anterior tabla se obtuvo que su distribución se había mantenido representativa en la mayoría de las convocatorias celebradas desde el año 2007.

## 4.4. Resumen de resultados

En este último apartado se resumen los resultados obtenidos aplicando las tres implementaciones del análisis cluster en la clasificación de las secciones censales de Valladolid en base a la distribución del voto que hayan tenido en las últimas 9 convocatorias municipales.

El número óptimo de agrupaciones en las convocatorias estudiadas atendiendo a criterios dictados por el coeficiente de silueta y el total de variabilidad interna de los grupos oscila entre 2 y 4 clusters en todos los casos. A pesar de que las candidaturas hayan ido cambiando a lo largo del tiempo, el perfil de las distribuciones de los clusters generados se mantiene en la misma línea. Generalmente, el grupo con las secciones más representativas tiene resultados más ajustados entre las dos primeras candidaturas. Un segundo cluster suele estar constituido por secciones con un perfil más conservador obteniendo mayor ventaja los partidos cercanos a esta ideología. En el tercer cluster se agrupan secciones en las que los partidos más a la izquierda en el espectro ideológico cosechan sus mejores resultados.

En cuanto a las secciones más representativas, al utilizar la partición óptima en el análisis no jerarquizado destaca la **sección 21 del distrito 11** situada en la zona de **Las Villas-Valparaíso**, es el único elemento que se mantiene representativo desde el año 1987 en todas las ocasiones. En menor medida también llaman la atención diferentes secciones que tienen distribuciones representativas en todo el periodo salvo en una de las convocatorias, en este grupo se incluyen, por ejemplo, la **sección 10 del distrito 6 (Plaza Circular)** o la **sección 18 del distrito 11 (La Rubia)**.

Por regla general se ha observado que, con esta técnica, el conjunto de secciones más representativas se mantiene más o menos estable hasta la convocatoria del año 2007, a partir de la cual la distribución de varios individuos significativos hasta ese momento dejan de ser similares al perfil de la ciudad. En este sentido, las secciones **sección 2 del distrito 4 (Delicias-Canterac)** y la **sección 7 del distrito 10 (Girón)** tienen un alto valor de cara a futuras elecciones municipales ya que en ellas la distribución del voto lleva siendo representativa desde 2007.

Por otro lado, en el contexto del análisis no jerarquizado empleando 20 clusters se han obtenido 5 secciones con una mayor relevancia (Tabla 4.42), entre las cuales destaca la anteriormente nombrada **sección 10 del distrito 6 (Plaza Circular)** por mantenerse representativa desde el año 2007, es el único elemento que cumple dicha característica.

Con la técnica de análisis jerarquizado aplicada se generan grupos de secciones muy reducidos, sin embargo, existen 3 elementos que sobresalen del resto siendo representativos en dos convocatorias electorales: **sección 2 del distrito 8 (Hospital)**, **sección 18 del distrito 11 (La Rubia)** y la **sección 33 del distrito 10 (Huerta del Rey-Feria de muestras)**. El último elemento es el más relevante ya que las dos convocatorias en donde fue representativa de la ciudad son más recientes en comparación con las de las otras dos secciones.

#### 4.4. RESUMEN DE RESULTADOS

Como se puede observar, se han encontrado coincidencias en los resultados aplicando las diferentes técnicas, concretamente se ha detectado una convergencia significativa entre el análisis no jerarquizado con 20 clusters y el jerarquizado. En la tabla 4.56 se presentan las secciones censales que se han considerado candidatas a representar a la distribución del voto en la ciudad teniendo en cuenta los tres procedimientos.

Distrito	Sección	Zona	Tablas				
			4.29	4.31	4.51	4.53	4.55
4	2	Delicias-Canterac	-	✓	-	✓	-
6	10	Plaza Circular	✓	-	✓	✓	✓
10	7	Girón	-	✓	-	-	-
10	33	Huerta del Rey-Feria de Muestras	-	-	✓	✓	✓
11	18	La Rubia	✓	-	-	✓	✓
11	21	Las Villas-Valparaíso	✓	✓	-	-	-

Tabla 4.56: Tabla de secciones más representativas en sentido global

A continuación, en la figura 4.19 se representa la ubicación de las secciones de mayor interés desagregadas en los siguientes grupos.

- *Grupo 1*: En este grupo se encuentran las secciones que, tras aplicar el análisis no jerarquizado con partición óptima, han tenido comportamientos representativos en más de 7 convocatorias desde el año 1987 (Tabla 4.29).
- *Grupo 2*: Teniendo en cuenta los resultados del análisis no jerarquizado con partición en 20 clusters, este grupo está constituido tanto por secciones que hayan tenido comportamientos representativos en más de 6 convocatorias desde el año 1987 (Tabla 4.51) como por los elementos que se hayan mantenido representativos en las últimas 3 convocatorias (marcados en la tabla 4.53).
- *Grupo 3*: Este grupo está constituido por aquellas secciones que han tenido distribuciones similares al perfil global en alguna convocatoria desde 1987 aplicando el análisis jerarquizado descrito en el apartado 4.3 (tabla 4.55).
- *Grupo 4*: En este último conjunto se agrupan las secciones censales que se han considerado más representativas de Valladolid observando los resultados de las tres técnicas utilizadas (tabla 4.56).

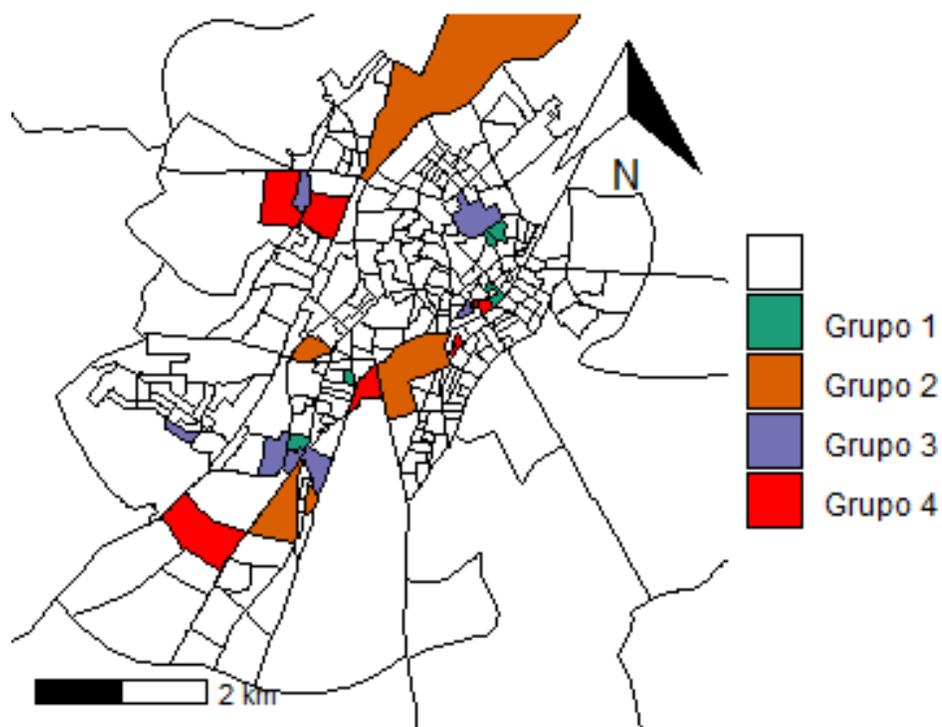


Figura 4.19: Mapa con todas las secciones más representativas



# Capítulo 5

## Conclusiones

Con ayuda del software R, se han estudiado los resultados electorales de las convocatorias municipales celebradas en Valladolid desde 1987 hasta 2019 con el objetivo de caracterizar las secciones censales de la ciudad mediante tres métodos de análisis cluster diferentes, dos procedimientos no jerárquicos implementados por el algoritmo K-Medoids y un procedimiento jerárquico.

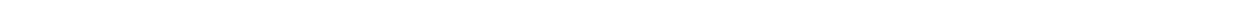
Por lo general, en el proceso de caracterización de secciones, el cual se ha llevado a cabo aplicando un método no jerárquico con el número de clusters óptimo en cada convocatoria, destacan tres tipos de grupos cuya distribución del voto se ha mantenido a lo largo del tiempo. El grupo de secciones más próximas al perfil de toda la ciudad es habitualmente el más ajustado, los otros dos grupos destacados suelen tener una candidatura dominante y un mayor peso de partidos con una ideología más radical.

Considerando los resultados de los tres métodos empleados, finalmente se ha seleccionado un conjunto de 6 secciones censales que pueden considerarse las más representativas de la ciudad. Las zonas concretas de Valladolid en donde se ubican dichas secciones son muy diversas, Delicias-Canterac, Plaza Circular, Barrio Girón, Huerta del Rey-Feria de Muestras, La Rubia y por último, la zona de Las Villas-Valparaíso.

Adicionalmente, se ha probado con otras métricas diferentes a la de Kullback-Leibler como medida de disimilaridad entre las distribuciones del voto de las secciones y se ha concluido que no hay diferencias significativas en los resultados de los análisis realizados.

Los resultados obtenidos en este trabajo podrán ser utilizados para la realización de sondeos electorales ó estudios sociopolíticos de la ciudad.

Por último, apuntar una posible continuación de este trabajo. En este Trabajo de fin de Grado se utiliza el análisis cluster como técnica de clasificación de secciones censales de la ciudad de Valladolid de acuerdo con los resultados obtenidos en elecciones municipales anteriores, podría ser de interés realizar un estudio similar aplicado a resultados en otro tipo de convocatorias electorales como por ejemplo, a elecciones generales o autonómicas y comparar los resultados del análisis con los obtenidos en este trabajo.

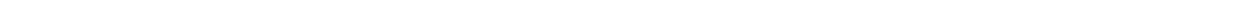


# Anexo

## Siglas de Partidos Políticos

Partido Político	Siglas
Alianza Independiente	AI
Alianza Popular	AP
Candidatura Independiente	CI
Centro Democrático y Social	CDS
Ciudadanos-Partido de la Ciudadanía	C's
Corporación Participativa Ciudadana	CPC
Izquierda Castellana	IzCa
Izquierda Unida	IU
Los Verdes	LV
Partido Antitaurino Contra el Maltrato Animal	PACMA
Partido de los Trabajadores de España-Unidad Comunista	PTE-UC
Partido Popular	PP
Partido Socialista Obrero Español	PSOE
Plataforma de los Independientes de España	PIE
Sí se puede Valladolid	SÍVA
Tierra Comunera	TC
Unión Progreso Y Democracia	UPYD
Valladolid Toma La Palabra	VTLP
VOX	VOX

Tabla 5.1: Tabla de siglas de los partidos políticos nombrados en el Trabajo



# Bibliografía

- [1] Ministerio del Interior. Consulta de resultados electorales - Área de descargas. <http://www.infoelectoral.mir.es/infoelectoral/min/areaDescarga.html?method=search>, 2013. [Online; Último acceso: 09/06/2021].
- [2] INE. Productos y servicios - cartografía digitalizada de secciones censales. [https://www.ine.es/ss/Satellite?L=es\\_ES&c=Page&cid=1259952026632&p=1259952026632&pagename=ProductosYServicios%2FPYSLayout](https://www.ine.es/ss/Satellite?L=es_ES&c=Page&cid=1259952026632&p=1259952026632&pagename=ProductosYServicios%2FPYSLayout). [Online; Último acceso: 10/06/2021].
- [3] James G., Witten D., Hastie T., and Tibshirani R. (2017). *An Introduction to Statistical Learning with Applications in R*, chapter 10.3. Springer.
- [4] Universitat de València. Contenedor hipermedia de estadística aplicada a las ciencias económicas y sociales. Introducción al análisis cluster. <https://www.uv.es/ceaces/multivari/cluster/CLUSTER.htm>. [Online; Último acceso: 22/06/2021].
- [5] Everitt B., Landau S., Leese M., and Stahl D. (2011). *Cluster Analysis*, pages 121–126. Wiley.
- [6] Sueur J., Aubin T., and Simonis C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18:213–226.
- [7] Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.
- [8] Maechler M., Rousseeuw P., Struyf A., Hubert M., and Hornik K. (2021). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.1.
- [9] Kaufman Peter L. and Rousseeuw J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- [10] E. Pebesma. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1):439–446.
- [11] H. Wickham. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- [12] kassambara A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning, Volumen 1 de Multivariate Analysis*. STHDA.