



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE GRADO

Grado en Estadística

**Clasificación de las provincias de España en base a la
estructura de su comercio exterior**

Autor: Adrián Poncela Gómez

Tutores:

Jesús María Rodríguez Rodríguez

María Teresa González Arteaga

Resumen:

En este TFG se realizará una clasificación de las provincias de España en base a la estructura de su comercio exterior. Para ello se utilizarán los datos descargados de la página web de la Agencia Tributaria pertenecientes al año 2017. Estos datos brutos muestran cada una de las exportaciones e importaciones que se realizaron en cada provincia durante dicho año.

Para lograr la caracterización de las provincias españolas es necesario realizar un proceso de tratamiento de los datos brutos para poder aplicar técnicas multivariantes. A lo largo de este proyecto serán utilizadas las técnicas de “Análisis en componentes principales” y “Análisis Clúster” mediante el lenguaje estadístico R.

Abstract:

The present dissertation includes a classification of the provinces of Spain based on the structure of their foreign trade. It will be used the data downloaded from the “Agencia Tributaria” website belonging to the year 2017. These raw data show each of the exports and imports that were made in each province during that year

To achieve the characterization of the Spanish provinces it is necessary to carry out a process of treatment of the raw data to be able to apply multivariate techniques. It will be used in this dissertation the techniques of “Principal component analysis” and “Cluster Analysis” through the statistical language R.

Índice general

1	Introducción	12
1.1	Objetivos.....	13
1.2	Estructura de la memoria	13
2	Herramientas utilizadas.....	15
2.1	Lenguaje R.....	15
2.2	RStudio.....	16
2.3	Paquetes utilizados.....	17
3	Obtención, descripción y preparación de los datos.....	18
3.1	Obtención de los datos	18
3.2	Descripción de los datos	21
3.3	Preparación de los datos	28
4	Análisis exploratorio de los datos	34
4.1	Análisis de casos peculiares con secciones de mercancías ausentes	34
4.2	Análisis exploratorio de las variables	35
4.2.1	Valor monetario total de exportaciones/importaciones en euros	35
4.2.2	Valor monetario de exportaciones/importaciones en euros por habitante.....	42
4.3	Correlación de las variables	51
5	Análisis de componentes principales.....	53
5.1	Metodología.....	53
5.2	Resultados.....	54
5.2.1	ACP del valor monetario total de exportaciones en euros	54
5.2.2	ACP del valor monetario total de importaciones en euros.....	56
5.2.3	ACP del Valor monetario de exportaciones en euros por habitante	59
5.2.4	ACP del valor monetario de exportaciones en euros por habitante.....	60
6	Clasificación de las provincias	63
6.1	Metodología.....	63
6.2	Resultados.....	64
6.2.1	Análisis clúster del valor monetario total de exportaciones en euros.....	64
6.2.2	Análisis clúster del valor monetario total de importaciones en euros	68
6.2.3	Análisis clúster del valor monetario de exportaciones en euros por habitante	72
6.2.4	Análisis clúster del valor monetario de importaciones en euros por habitante.....	76
6.2.5	Análisis clúster del porcentaje del valor de exportaciones de cada sección por provincia .	80
6.2.6	Análisis clúster del porcentaje del valor de importaciones de cada sección por provincia .	84
7	Conclusiones.....	88
8	Referencias.....	90

9	<i>Anexos</i>	92
9.1	Lista de capítulos de mercancías	92

Índice de figuras

Ilustración 1. Logo lenguaje R y RStudio	15
Ilustración 2. Descarga del software R	15
Ilustración 3. Ventanas de RStudio [7]	16
Ilustración 4. Agencia Tributaria principal	18
Ilustración 5. Agencia Tributaria memorias y estadísticas	18
Ilustración 6. Agencia Tributaria estadísticas	19
Ilustración 7. Agencia Tributaria comercio exterior	19
Ilustración 8. Agencia Tributaria descarga	19
Ilustración 9. Agencia Tributaria descarga de ficheros de datos	19
Ilustración 10. Agencia Tributaria anuales territoriales	20
Ilustración 11. Agencia Tributaria años	20
Ilustración 12. Agencia Tributaria conjunto de datos	20
Ilustración 13. Ejemplo fichero de datos	21
Ilustración 14. Nomenclatura posición estadística	24
Ilustración 15. Código R leer ficheros comprimidos	28
Ilustración 16. Código R cargar archivos auxiliares	29
Ilustración 17. Función para leer los datos	30
Ilustración 18. Ejemplo resultado tras función	30
Ilustración 19. Código R creación matrices de datos	31
Ilustración 20. Ejemplo matriz de datos de Importación	31
Ilustración 21. Código R contar NAs	34
Ilustración 22. Código R sustituir NAs	35
Ilustración 23. Gráfico barras exportaciones provincias	35
Ilustración 24. Mapa de exportaciones totales	36
Ilustración 25. Gráfico barras importaciones provincias	37
Ilustración 26. Mapa de importaciones totales	37
Ilustración 27. Gráfico barras exportaciones secciones	38
Ilustración 28. Gráfico barras importaciones secciones	39
Ilustración 29. Gráfico barras exportaciones por habitante	48
Ilustración 30. Gráfico barras importaciones por habitante	49
Ilustración 31. Gráfico barras exportaciones por habitante para cada sección	49
Ilustración 32. Gráfico barras importaciones por habitante para cada sección	50
Ilustración 33. Correlaciones exportaciones	51
Ilustración 34. Correlaciones importaciones	51
Ilustración 35. Correlación exportaciones por habitante	51
Ilustración 36. Correlación importaciones por habitante	51
Ilustración 37. Correlaciones exportaciones sección/provincia	52
Ilustración 38. Correlaciones importaciones sección/provincia	52
Ilustración 39. Summary ACP exportaciones	54
Ilustración 40. Scree plot ACP exportaciones	55
Ilustración 41. Biplot ACP exportaciones	56
Ilustración 42. Summary ACP importaciones	56
Ilustración 43. Scree plot ACP importaciones	57
Ilustración 44. Biplot ACP importaciones	58
Ilustración 45. Summary ACP exportaciones por habitante	59
Ilustración 46. Scree plot ACP exportaciones por habitante	59
Ilustración 47. Biplot ACP exportaciones por habitante	60
Ilustración 48. Summary ACP importaciones por habitante	60

Ilustración 49. Scree plot ACP importaciones por habitante	61
Ilustración 50. Biplot ACP importaciones por habitante	62
Ilustración 51. Clúster jerárquico método simple	63
Ilustración 52. Clúster jerárquico método completo	64
Ilustración 53. Clúster jerárquico método promedio	64
Ilustración 54. Dendograma Ward exportaciones	65
Ilustración 55. Grupos Ward exportaciones.....	65
Ilustración 56. Mapa Ward exportaciones	65
Ilustración 57. Grupos k-means exportaciones.....	66
Ilustración 58. Mapa k-means exportaciones	66
Ilustración 59. Centroides k-means exportaciones	67
Ilustración 60. Dendograma Ward importaciones	69
Ilustración 61. Grupos Ward importaciones	69
Ilustración 62. Mapa Ward importaciones.....	69
Ilustración 63. Grupos k-means importaciones	70
Ilustración 64. Mapa k-means importaciones.....	70
Ilustración 65. Centroides k-means importaciones.....	71
Ilustración 66. Dendograma Ward exportaciones por habitante	73
Ilustración 67. Grupos Ward exportaciones por habitante.....	73
Ilustración 68. Mapa Ward exportaciones por habitante	73
Ilustración 69. Grupos k-means exportaciones por habitante.....	74
Ilustración 70. Mapa k-means exportaciones por habitante	74
Ilustración 71. Centroides k-means exportaciones por habitante	75
Ilustración 72. Dendograma Ward importaciones por habitante	77
Ilustración 73. Grupos Ward importaciones por habitante	77
Ilustración 74. Mapa Ward importaciones por habitante.....	77
Ilustración 75. Grupos k-means importaciones por habitante	78
Ilustración 76. Mapa k-means importaciones por habitante	78
Ilustración 77. Centroides k-means importaciones por habitante.....	78
Ilustración 78. Dendograma Ward exportaciones porcentaje sección	80
Ilustración 79. Grupos Ward exportaciones Ward porcentaje sección.....	81
Ilustración 80. Mapa Ward exportaciones porcentaje sección	81
Ilustración 81. Porcentaje sección Ward clúster exportaciones	81
Ilustración 82. Grupos k-means exportaciones porcentaje sección.....	82
Ilustración 83. Mapa k-means exportaciones porcentaje sección	82
Ilustración 84. Porcentaje sección k-means clústeres exportaciones	82
Ilustración 85. Porcentaje sección k-means centroides exportaciones	83
Ilustración 86. Dendograma Ward importaciones porcentaje sección	84
Ilustración 87. Grupos Ward importaciones porcentaje sección	85
Ilustración 88. Mapa Ward importaciones porcentaje sección	85
Ilustración 89. Porcentaje sección Ward clústeres importaciones	85
Ilustración 90. Grupos k-means importaciones porcentaje sección	86
Ilustración 91. Mapa k-means importaciones porcentaje sección.....	86
Ilustración 92. Porcentaje sección k-means clústeres importaciones.....	86
Ilustración 93. Porcentaje sección k-means centroides importaciones.....	87

Índice de tablas

Tabla 1. Campos ficheros de datos	21
Tabla 2. Longitudes de los campos.....	22
Tabla 3. Códigos provincias (INE)	23
Tabla 4. Ejemplos códigos países	24
Tabla 5. Nombres de los capítulos de mercancías	25
Tabla 6. Nombres de las secciones en que se agrupan los capítulos de mercancías	25
Tabla 7. Capítulos por sección de mercancía	26
Tabla 8. Población provincias 2017	31
Tabla 9. Secciones punteras	39
Tabla 10. Coeficientes análisis clúster exportaciones	65
Tabla 11. Coeficientes análisis clúster Importaciones.....	68
Tabla 12. Coeficientes análisis clúster exportaciones por habitante	72
Tabla 13. Coeficientes análisis clúster importaciones por habitante.....	76
Tabla 14. Coeficientes análisis clúster exportaciones porcentaje sección	80
Tabla 15. Coeficientes análisis clúster importaciones porcentaje sección.....	84

1 Introducción

Actualmente, nuestra sociedad se encuentra en la denominada “Era de la Información”, en una etapa también conocida como “Era Digital” o “Era Informática”. Así se denomina al periodo de la historia de la humanidad que va ligado a las tecnologías de la información y la comunicación (TIC). El comienzo de este periodo se asocia con la revolución digital. Dentro de esta era el denominado “Big data” tiene una importancia fundamental [1].

Big Data hace referencia al tratamiento y análisis de grandes cantidades de datos, con el objetivo de extraer valor de la información. Esta gestión y análisis de datos no pueden ser tratadas por técnicas convencionales, sino por otras técnicas más avanzadas entre las que se encuentran las técnicas multivariantes, el Machine Learning o análisis predictivo.

El Big data ha abierto la puerta hacia un nuevo enfoque de entendimiento y de toma de decisiones. Se puede explicar mediante las tradicionales 3 Vs: volumen, velocidad y variedad.

En este TFG, se va a realizar un análisis y tratamiento de información relacionada con los datos que proporciona la Agencia Tributaria acerca del comercio exterior de las provincias de España.

El comercio exterior es aquel que se refiere al conjunto de transacciones de naturaleza comercial y financiera, que implica el intercambio de bienes y servicios entre un país en particular con otros países o naciones. El objetivo del comercio exterior es satisfacer la demanda de los consumidores de determinados productos [2].

La importación ocurre cuando, para satisfacer la demanda interna de un país, ciertos productos deben comprarse fuera, bien porque son escasos o no existen en el país, bien porque su elaboración es más barata o de mejor calidad en otro país.

La exportación, por su lado, se da cuando un país logra producir determinado producto con gran valor o calidad, o con un mayor margen de rentabilidad que le permite venderlo a otros países.

El comercio exterior es fundamental para la vitalidad de la economía y para cubrir la demanda de los respectivos mercados. Además, mejora la competitividad, fortalece la pequeña y mediana industria y promueve las cadenas productivas, todo lo cual deriva en bienestar y calidad de vida para los ciudadanos.

Asimismo, el comercio exterior es una fuente de ingreso de divisas extranjeras al país, lo cual supone, a su vez, la generación de riquezas.

El comercio exterior español tuvo en 2017 un año extraordinario, con cifras récord tanto en los registros de exportación como en los de importación. El comercio exterior es muy relevante en la economía española y con un gran crecimiento desde la gran crisis económica del año 2008 [3].

Relacionado con este TFG, se van a utilizar los datos del comercio exterior del año 2017, que a pesar de ser un número considerable de datos no es algo que se pueda considerar intratable

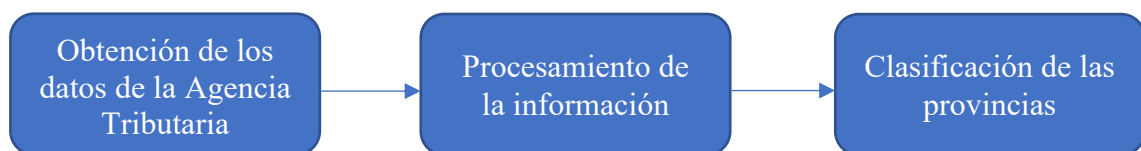
con los medios actuales. Se podría hacer un estudio de la evolución a lo largo de varios años, lo que aumentaría la complejidad significativamente.

1.1 Objetivos

Este TFG tiene dos objetivos principales:

- Los datos de la “Agencia Tributaria” se proporcionan mediante varios ficheros que necesitan un proceso de tratamiento para poder extraer la información relevante que contienen. Por tanto, el primer objetivo es realizar un procesamiento a los datos obtenidos con la finalidad de que sean tratables.
- Tras el procesamiento de los datos originales brutos, el segundo objetivo es clasificar las provincias españolas en base a la estructura de su comercio exterior, mediante la aplicación de técnicas multivariantes como son el “Análisis de Componentes Principales” y el “Análisis Clúster”.

En el siguiente esquema se resume el desarrollo de este TFG por etapas:



1.2 Estructura de la memoria

Esta memoria del TFG está estructurada por capítulos. A continuación, se explica brevemente qué contenidos se tratan en cada uno de ellos, así como las secciones que contienen.

- **Capítulo 1. Introducción:** En este capítulo se muestra una contextualización para este TFG, así como una breve exposición de los objetivos.
- **Capítulo 2. Herramientas utilizadas:** Se introducen las diferentes herramientas, así como los lenguajes de programación y software usados.
- **Capítulo 3. Obtención, descripción y preparación de los datos:** Se aborda el tratamiento de los datos brutos que se ha realizado tras su obtención. Así como la lectura y una descripción de éstos.
- **Capítulo 4. Análisis exploratorio de los datos:** Se realiza un estudio descriptivo de los datos mostrando algunos gráficos, así como mapas para tener una idea general de la información que puede ser obtenida.
- **Capítulo 5. Análisis de componentes principales:** Se aplica el método de componentes principales a los datos agregados en los capítulos anteriores.
- **Capítulo 6. Clasificación de las provincias:** Caracterización de las provincias de España con la aplicación de análisis clúster.

- **Capítulo 7. Conclusiones:** En este último capítulo se incluyen las conclusiones obtenidas, así como los objetivos alcanzados. También se comentan las posibles líneas de trabajo futuro.
- **Capítulo 8. Referencias:** Documentación utilizada y consultada a lo largo de la realización de la memoria de este TFG.
- **Capítulo 9. Anexos.**

2 Herramientas utilizadas

En el presente TFG, se ha utilizado el lenguaje de programación estadístico R, con la ayuda del IDE RStudio que se procede a introducir brevemente en este capítulo.



Ilustración 1. Logo lenguaje R y RStudio

2.1 Lenguaje R

R es un entorno y lenguaje de programación con un enfoque al análisis estadístico. Se trata de uno de los lenguajes de programación más utilizados en investigación científica y en ciencia de datos, siendo además muy popular en el campo de la minería de datos, investigación biomédica, la bioinformática y las matemáticas financieras. Esto, es debido a la posibilidad de extender sus capacidades mediante la carga de diferentes bibliotecas o paquetes con funcionalidades de cálculo y creación de gráficas. Así, proporciona un amplísimo abanico de herramientas de ajuste de modelos estadísticos y gráficas.

Una destacable característica del R es que se trata de software libre, y por tanto incluido en la filosofía Open Source. Además, cuenta con algunas ventajas reseñables para su utilización en el análisis de datos como son las siguientes:

- Lenguaje de programación orientado a objetos.
- Disponible para todos los sistemas operativos.
- Contiene técnicas estadísticas tanto básicas como avanzadas.
- Potente herramienta gráfica.
- Compatible con otros lenguajes.
- Manual con información de ayuda sobre las funciones y paquetes que contiene.

Según IEEE, el lenguaje R fue el quinto lenguaje de programación más utilizado por detrás de Python, Java, C y C++ [4].



Ilustración 2. Descarga del software R

2.2 RStudio

RStudio es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la gestión del espacio de trabajo [5] [6].

RStudio está dividido en cuatro ventanas. En la Ilustración 3 se muestra una imagen dónde:

- 1) La ventana superior izquierda es un editor adaptado para escribir código de R.
- 2) La ventana inferior izquierda corresponde a la consola para utilizar R de forma interactiva.
- 3) La ventana superior derecha muestra un listado de los objetos (ficheros, vectores, matrices, etc.) que se encuentran en el área de trabajo, con una pestaña para el historial de órdenes.
- 4) La ventana inferior derecha contiene varias pestañas con diferentes utilidades donde se destaca aquella en la que se representan los gráficos, la ayuda, y el manejo de paquetes.

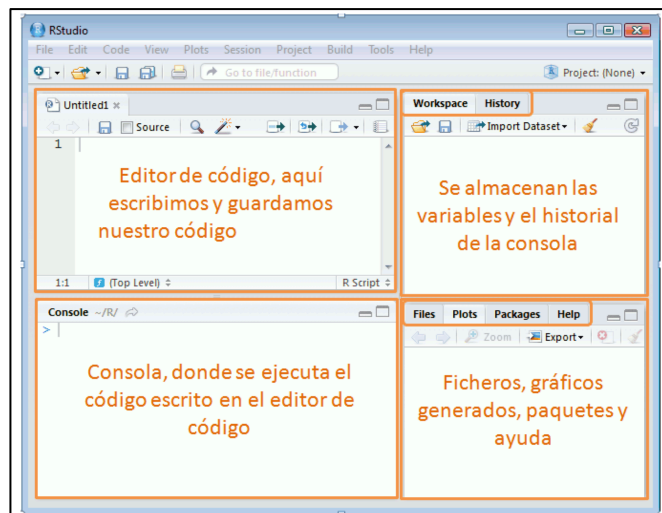


Ilustración 3. Ventanas de RStudio [7]

Algunas ventajas de RStudio:

- **Ayuda de los comandos de R.** Una vez escrito un comando en la consola o en el editor, si se sitúa el cursor entre los paréntesis y se pulsa el tabulador, se abrirá una ventana con los principales argumentos de ese comando y una ayuda sobre la función de cada argumento.
- **Visualización de los datos en el área de trabajo.** Al hacer click en alguno de los ficheros(data.frames) del listado, se visualizará en la ventana superior izquierda.
- **R + knitr + Markdown.** Permite la creación de páginas web que incluyen el código R. También se pueden insertar fórmulas escritas con LaTeX.

2.3 Paquetes utilizados

Varios paquetes han sido requeridos a lo largo del desarrollo de este TFG para la aplicación de técnicas o para la representación de gráficos. Todos ellos se citan en la sección de referencias.

3 Obtención, descripción y preparación de los datos

En las secciones mostradas seguidamente se explicará paso a paso cómo se obtienen los datos, así como su lectura y descripción de estos.

3.1 Obtención de los datos

Para la realización de este proyecto, se va a utilizar un conjunto de datos público, situado en la página web de la Agencia Tributaria: <https://www.agenciatributaria.es> [8]

A continuación, se muestran una serie de imágenes a modo de manual para poder encontrar y descargar dicho conjunto de datos.

- 1) En el menú principal hay que pulsar en el apartado que se corresponde con “Agencia Tributaria”.



Ilustración 4. Agencia Tributaria principal

- 2) Dentro del apartado, pulsar en “Memorias y estadísticas”.

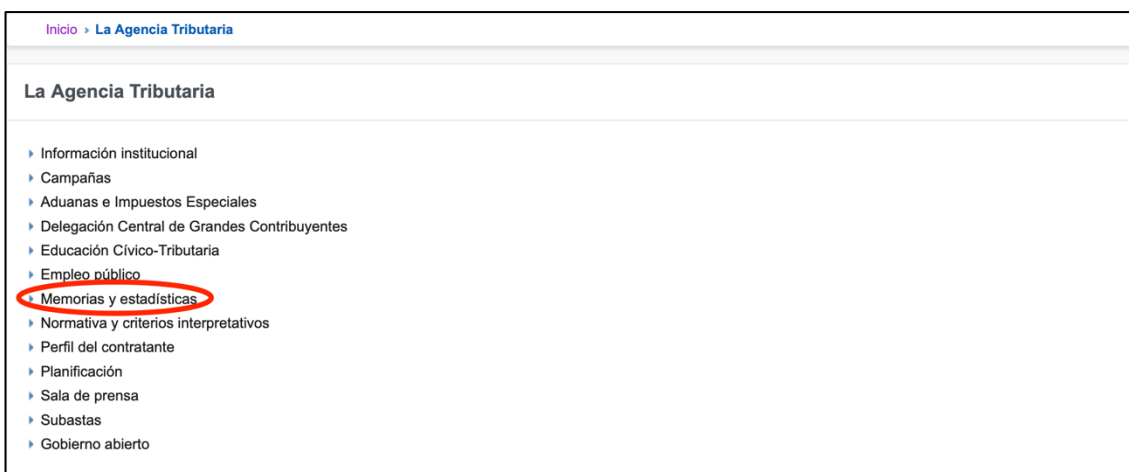


Ilustración 5. Agencia Tributaria memorias y estadísticas

3) Pulsar en el apartado de “Estadísticas”.



Ilustración 6. Agencia Tributaria estadísticas

4) Pulsar en el apartado de “Comercio exterior”.

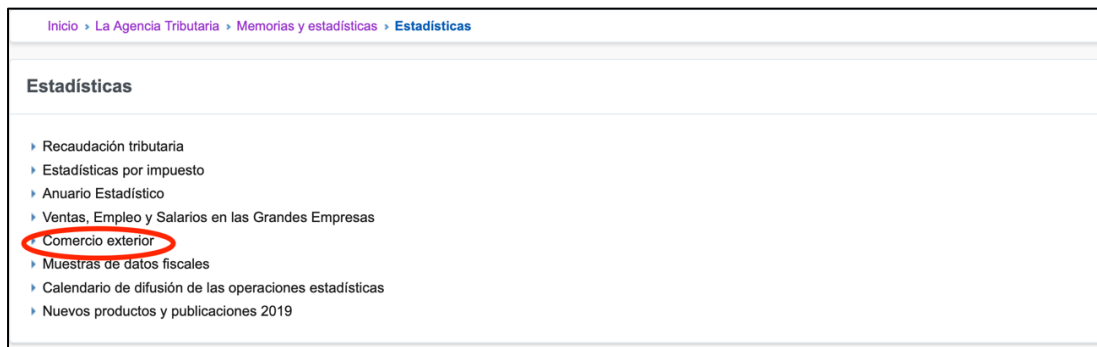


Ilustración 7. Agencia Tributaria comercio exterior

5) Pulsar sobre la imagen que indica “Descarga” de los datos estadísticos.



Ilustración 8. Agencia Tributaria descarga

6) Pulsar sobre “Descarga de ficheros de datos”.

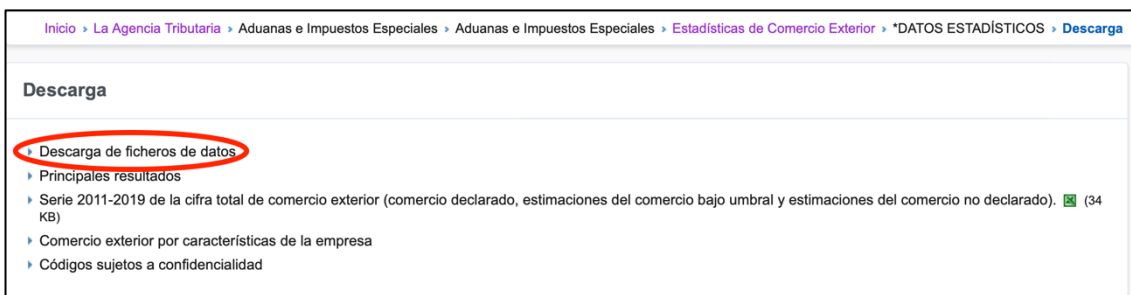


Ilustración 9. Agencia Tributaria descarga de ficheros de datos

7) Para este TFG, se han utilizado los datos por “Anuales territoriales”.

Inicio > La Agencia Tributaria > Aduanas e Impuestos Especiales > Aduanas e Impuestos Especiales > Estadísticas de Comercio Exterior > *DATOS ESTADÍSTICOS > Descarga >

Descarga de ficheros de datos

NOTA:
Si descargó los datos definitivos de máxima desagregación mensuales del año 2016 entre las fechas 01/12/2017 y 10/12/2017, deberá volver a descargarse los ya que se han detectado errores en los mismos.

- > Mensuales máxima desagregación
- > Mensuales territoriales
- > Mensuales resumen
- > Anuales máxima desagregación
- > **Anuales territoriales**
- > Anuales resumen

Ilustración 10. Agencia Tributaria anuales territoriales

8) Para este TFG, se han utilizado los datos pertenecientes al año “2017”.

Descarga de datos anuales territoriales en euros (céntimos)

Ofrece para cada año, flujo de Importación y Exportación, provincia de la aduana, provincia origen/destino, y código de la Nomenclatura Combinada, los siguientes datos tanto para la Importación/Introducción como para la Exportación/Expedición:

- > Peso en Kgrs.
- > Valor estadístico en céntimos de euro
- > Unidades suplementarias

La información suministrada de Comercio Exterior tiene el siguiente formato:

Año	Flujo	Aduana	Provincia	País	Posición estadística	Modo transporte	Peso en KG	Unidades	Valor Céntimos Euro	Domicilio fiscal
2	1	2	2	2	8	1	12	12	12	2

Longitud del campo en caracteres.

- > **2017**
- > 2016
- > 2015
- > 2014
- > 2013
- > 2012
- > 2011
- > 2010

Ilustración 11. Agencia Tributaria años

9) Por último, se encuentran 11 archivos, de los cuáles diez son ficheros .zip que contienen los datos y un .txt. Se descargan y guardan en el directorio de trabajo los diez ficheros comprimidos que tienen un tamaño total de 33.2 MB. No se va a utilizar el fichero denominado “Estimaciones.txt” por lo que no se es necesario descargarlo.

Los ficheros aparecen nombrados mediante un código denominado “Posición Estadística” que indica el producto que es importado/exportado. Este código se explicará mas adelante.

2017

- > 2017 01 al 24 (3595 KB)
- > 2017 25 al 38 (2485 KB)
- > 2017 39 al 43 (2651 KB)
- > 2017 44 al 49 (1415 KB)
- > 2017 50 al 67 (7322 KB)
- > 2017 68 al 71 (1122 KB)
- > 2017 72 al 83 (2902 KB)
- > 2017 86 al 92 (2207 KB)
- > 2017 84 al 85 (6574 KB)
- > 2017 93 al 99 (2185 KB)
- > Estimaciones .txt (4 KB)

Ilustración 12. Agencia Tributaria conjunto de datos

3.2 Descripción de los datos

Al descomprimir los archivos se puede ver la gran cantidad de datos disponibles, pasando de 33.2 MB, con los ficheros comprimidos a un tamaño total de 225.2 MB.

La siguiente imagen (Ilustración 13), muestra un ejemplo de uno de los ficheros de datos abierto con un editor de texto.

AA	F	AD	PR	PA	POSICION	T	PESO (KG.)	UNIDADES	V.ESTADISTI	DF
17	E	01	01	AE	09012100	4	00000000423	00000000000	00000762692	47
17	E	01	01	AE	09019090	4	00000000007	00000000000	00000011627	48
17	E	01	01	AG	22041096	4	00000000008	00000000006	00000001390	50
17	E	01	01	AT	21050010	3	000000150262	00000000000	00004675858	01
17	E	01	01	AT	21069092	3	00000000436	00000000000	00000705114	01
17	E	01	01	AT	22042136	3	000000008810	000000008600	000002600323	01
17	E	01	01	AT	22042138	3	000000010143	000000010143	000004534357	01
17	E	01	01	AT	22042176	3	000000306089	000000289118	000180641087	01
17	E	01	01	AT	22042178	3	000000002316	00000002316	000001836336	01
17	E	01	01	AT	22042179	3	000000000225	00000000025	000000063438	01
17	E	01	01	AT	22042180	3	000000001953	000000001778	000000659112	01
17	E	01	01	AT	22042182	3	000000000063	000000000063	000000034980	01
17	E	01	01	AT	22042184	3	000000019336	000000019336	0000002506071	01
17	E	01	01	AT	22042278	3	000000000138	000000000090	000000128215	01
17	E	01	01	AU	22042184	4	000000000001	00000000001	000000000800	01
17	E	01	01	BB	22042176	3	000000000810	000000000810	0000000606420	01
17	E	01	01	BE	07095910	3	000000002682	000000000000	000003191730	01
17	E	01	01	BE	07095930	3	000000000152	000000000000	000000449750	01
17	E	01	01	BE	07095990	3	000000003807	000000000000	000006712417	01

Ilustración 13. Ejemplo fichero de datos

En estos ficheros se dispone de once columnas con distinto significado. En la Tabla 1 se incluye tanto el nombre como la descripción.

Tabla 1. Campos ficheros de datos

Nombre	Nombre Completo	Descripción
AA	Año	17, que se corresponde con el año 2017.
F	Flujo	Exportación (E) o Importación (I)
AD	Aduana	Código de la Aduana
PR	Provincia	Provincia a dónde se exporta/importa la mercancía.
PA	País	País de dónde se exporta/importa la mercancía.
POSICION	Posición Estadística	Nomenclatura para indicar el tipo de mercancía.
T	Modo transporte	Código del modo de transporte.
PESO (KG.)	Peso en Kg	Peso del transporte.
UNIDADES	Unidades	Número de unidades del transporte.
V.ESTADISTI	Valor Céntimos Euro	Valor total en céntimos de euro de lo que ha sido exportado/importado
DF	Domicilio Fiscal	Código del Domicilio Fiscal

Los ficheros descargados de la página Web de la “Agencia Tributaria” tienen un formato predeterminado en el que cada una de las columnas tiene una longitud fija. Esto facilita la lectura de dichos ficheros de datos con un software estadístico. La longitud de cada campo se muestra en la Tabla 2:

Tabla 2. Longitudes de los campos

Nombre	Longitud del campo
AA	2 caracteres
F	1 carácter
AD	2 caracteres
PR	2 caracteres
PA	2 caracteres
POSICION	8 caracteres
T	1 carácter
PESO (KG.)	12 caracteres
UNIDADES	12 caracteres
V.ESTADISTI	12 caracteres
DF	2 caracteres

En este estudio se utilizan las siguientes variables:

- **Año**
- **Flujo**
- **Provincia**
- **País**
- **Posición Estadística**
- **Valor (Monetario)**

Cada fila muestra un transporte de mercancía que se realizó en el año 2017, que puede tratarse tanto de una Exportación como de una Importación, que afecta a una provincia determinada y a un país en concreto, por un valor monetario fijo. La mercancía es del tipo que indica su “posición estadística”.

A continuación, se describe el significado de las variables con detalle:

- **Año**

Tiene un valor fijo que es 17, correspondiente con el año 2017.

- **Flujo**

Esta variable puede tomar dos valores, dependiendo del tipo de envío.

- “E” para Exportaciones
- “I” para Importaciones

- **Provincia**

Cada una de las provincias aparece identificada con el código INE correspondiente. Estos códigos se muestran en la Tabla 3:

Tabla 3. Códigos provincias (INE)

CODIGO	PROVINCIA	CODIGO	PROVINCIA
1	Álava	27	Lugo
2	Albacete	28	Madrid
3	Alicante	29	Málaga
4	Almería	30	Murcia
5	Ávila	31	Navarra
6	Badajoz	32	Ourense
7	Baleares	33	Asturias
8	Barcelona	34	Palencia
9	Burgos	35	Las Palmas
10	Cáceres	36	Pontevedra
11	Cádiz	37	Salamanca
12	Castellón	38	Tenerife
13	Ciudad Real	39	Cantabria
14	Córdoba	40	Segovia
15	A Coruña	41	Sevilla
16	Cuenca	42	Soria
17	Girona	43	Tarragona
18	Granada	44	Teruel
19	Guadalajara	45	Toledo
20	Guipúzcoa	46	Valencia
21	Huelva	47	Valladolid
22	Huesca	48	Vizcaya
23	Jaén	49	Zamora
24	León	50	Zaragoza
25	Lleida	51	Ceuta
26	La Rioja	52	Melilla

- País

Los países, aparecen identificados por un código alfabético de 2 letras que se corresponde con el estándar internacional de normalización ISO 3166. Sabiendo dicho código, se puede determinar que país está implicado en el transporte de las mercancías.

En la tabla 4 se incluyen algunos ejemplos de países con su código correspondiente.

Tabla 4. Ejemplos códigos países

PAÍS	CÓDIGO
Dominica	DM
República Dominicana	DO
Argelia	DZ
Ecuador	EC
España	ES

- **Posición Estadística**

La variable “Posición Estadística” es una cifra de 8 caracteres para la cuál existe una nomenclatura que permite identificar qué tipo de mercancía referencia.

Un documento Excel que se encuentra en la página web de la “Agencia Tributaria” indica el significado de cada una de las cifras de la “Posición Estadística”. En la Ilustración 14 se muestra un ejemplo:

CODE	LEVEL	DASHES	ES
I	1		SECCIÓN I - ANIMALES VIVOS Y PRODUCTOS DEL REINO ANIMAL
01	2		CAPÍTULO 1 - ANIMALES VIVOS
0101	3		Caballos, asnos, mulos y burdéganos, vivos
	4	-	- Caballos
0101 21 00	5	--	-- Reproductores de raza pura
0101 29	5	--	-- Los demás
0101 29 10	6	---	--- Que se destinen al matadero
0101 29 90	6	---	--- Los demás
0101 30 00	4	-	- Asnos
0101 90 00	4	-	- Los demás
0102	3		Animales vivos de la especie bovina
	4	-	- Bovinos domésticos
0102 21	5	--	-- Reproductores de raza pura
0102 21 10	6	---	--- Terneras (que no hayan parido nunca)
0102 21 30	6	---	--- Vacas
0102 21 90	6	---	--- Los demás
0102 29	5	--	-- Los demás
0102 29 05	6	---	--- De los subgéneros Bibos o Poephagus

Ilustración 14. Nomenclatura posición estadística

En los códigos de la Ilustración 14 se puede ver cómo las 2 primeras cifras se corresponden con el capítulo de mercancías. Existen un total de 98 capítulos entre los que se divide la mercancía.

En la ilustración se puede observar como el primer capítulo de mercancías se corresponde con “Animales Vivos”. De la tercera cifra en adelante, se va especificando aún más el tipo exacto de la mercancía.

Para el presente TFG, sólo se tendrán en cuenta las 2 primeras cifras de la posición estadística, es decir, todos los posibles envíos de mercancía quedarán clasificados en 98 capítulos.

En la Tabla 5, se puede apreciar el nombre de algunos de los capítulos de mercancías. La tabla completa que incluye el nombre de cada uno de los 98 capítulos de mercancías se encuentra en el primer anexo de esta memoria.

Tabla 5. Nombres de los capítulos de mercancías

Capítulo de mercancía
CAPÍTULO 1 – ANIMALES VIVOS
CAPÍTULO 2 – CARNE Y DESPOJOS COMESTIBLES
CAPÍTULO 3 – PESCADOS Y CRUSTÁCEOS, MOLUSCOS Y DEMÁS INVERTEBRADOS ACUÁTICOS
CAPÍTULO 96 – MANUFACTURAS DIVERSAS
CAPÍTULO 97 – OBJETOS DE ARTE O COLECCIÓN Y ANTIGÜEDADES
CAPÍTULO 98 – CONJUNTOS INDUSTRIALES

El objetivo principal de este trabajo es la clasificación de las provincias. Uno de los principales aspectos para caracterizar una provincia es el tipo de mercancía que bien exporta o importa. 98 capítulos son demasiados, por lo que se ha decidido utilizar una división que realiza la “Agencia Tributaria” en secciones, agrupando en 21 secciones de mercancías los 98 capítulos de mercancías.

Las 21 secciones de mercancías son las que se muestran en la Tabla 6.

Tabla 6. Nombres de las secciones en que se agrupan los capítulos de mercancías

Sección de mercancía
SECCIÓN 1 – ANIMALES VIVOS Y PRODUCTOS DEL REINO ANIMAL
SECCIÓN 2 – PRODUCTOS DEL REINO VEGETAL
SECCIÓN 3 – GRASAS Y ACEITES ANIMALES O VEGETALES; PRODUCTOS DE SU DESDOBLAMIENTO; GRASAS ALIMENTICIAS ELABORADAS; CERAS DE ORIGEN ANIMAL O VEGETAL
SECCIÓN 4 – PRODUCTOS DE LAS INDUSTRIAS ALIMENTARIAS; BEBIDAS, LÍQUIDOS ALCOHÓLICOS Y VINAGRE; TABACO Y SUCEDÁNEOS DEL TABACO ELABORADOS
SECCIÓN 5 – PRODUCTOS MINERALES
SECCIÓN 6 – PRODUCTOS DE LAS INDUSTRIAS QUÍMICAS O DE LAS INDUSTRIAS CONEXAS
SECCIÓN 7 – PLÁSTICO Y SUS MANUFACTURAS; CAUCHO Y SUS MANUFACTURAS
SECCIÓN 8 – PIELES, CUEROS, PELETERÍA Y MANUFACTURAS DE ESTAS MATERIAS; ARTÍCULOS DE TALABARTERÍA O GUARNICIONERÍA; ARTÍCULOS DE VIAJE, BOLSOS DE MANO (CARTERAS) Y CONTINENTES SIMILARES; MANUFACTURAS DE TRIPA
SECCIÓN 9 – MADERA, CARBÓN VEGETAL Y MANUFACTURAS DE MADERA; CORCHO Y SUS MANUFACTURAS; MANUFACTURAS DE ESPARTERÍA O CESTERÍA

SECCIÓN 10 – PASTA DE MADERA O DE LAS DEMÁS MATERIAS FIBROSAS CELULÓSICAS; PAPEL O CARTÓN PARA RECICLAR (DESPERDICIOS Y DESECHOS); PAPEL O CARTÓN Y SUS APLICACIONES
SECCIÓN 11 – MATERIAS TEXTILES Y SUS MANUFACTURAS
SECCIÓN 12 – CALZADO, SOMBREROS Y DEMÁS TOCADOS, PARAGUAS, QUITASOLES, BASTONES, LÁTIGOS, FUSTAS Y SUS PARTES; PLUMAS PREPARADAS Y ARTÍCULOS DE PLUMAS; FLORES ARTIFICIALES; MANUFACTURAS DE CABELLO
SECCIÓN 13 – MANUFACTURAS DE PIEDRA, YESO FRAGUABLE, CEMENTO, AMIANTO (ASBESTO), MICA O MATERIAS ANÁLOGAS; PRODUCTOS CERÁMICOS; VIDRIO Y SUS MANUFACTURAS
SECCIÓN 14 – PERLAS FINAS (NATURALES) O CULTIVADAS, PIEDRAS PRECIOSAS O SEMIPRECIOSAS, METALES PRECIOSOS, CHAPADOS DE METAL PRECIOSO (PLAQUÉ) Y MANUFACTURAS DE ESTAS MATERIAS; BISUTERÍA; MONEDAS
SECCIÓN 15 – METALES COMUNES Y MANUFACTURAS DE ESTOS METALES
SECCIÓN 16 – MÁQUINAS Y APARATOS, MATERIAL ELÉCTRICO Y SUS PARTES; APARATOS DE GRABACIÓN O REPRODUCCIÓN DE SONIDO, APARATOS DE GRABACIÓN O REPRODUCCIÓN DE IMÁGENES Y SONIDO EN TELEVISIÓN, Y LAS PARTES Y ACCESORIOS DE ESTOS APARATOS
SECCIÓN 17 - MATERIAL DE TRANSPORTE
SECCIÓN 18 - INSTRUMENTOS Y APARATOS DE ÓPTICA, FOTOGRAFÍA O CINEMATOGRAFÍA, DE MEDIDA, CONTROL O PRECISIÓN; INSTRUMENTOS Y APARATOS MEDICOQUIRÚRGICOS; APARATOS DE RELOJERÍA; INSTRUMENTOS MUSICALES; PARTES Y ACCESORIOS DE ESTOS INSTRUMENTOS O APARATOS
SECCIÓN 19 - ARMAS, MUNICIONES, Y SUS PARTES Y ACCESORIOS
SECCIÓN 20 - MERCANCÍAS Y PRODUCTOS DIVERSOS
SECCIÓN 21 - OBJETOS DE ARTE O COLECCIÓN Y ANTIGÜEDADES

En la Tabla 7 se muestran qué capítulos de mercancías pertenecen a cada una de las secciones anteriores especificadas por la Agencia Tributaria.

Tabla 7. Capítulos por sección de mercancía

Sección	Capítulos de mercancía
1	1,2,3,4,5
2	6,7,8,9,10,11,12,13,14
3	15
4	16,17,18,19,20,21,22,23,24
5	25,26,27
6	28,29,30,31,32,33,34,35,36,37,38
7	39,40
8	41,42,43
9	44,45,46

10	47,48,49
11	50,51,52,53,54,55,56,57,58,59,60,61,62,63
12	64,65,66,67
13	68,69,70
14	71
15	72,73,74,75,76,77,78,79,80,81,82,83
16	84,85
17	86,87,88,89
18	90,91,92
19	93
20	95,95,96
21	97,98

En resumen, se ha pasado de una variable que en un principio se denominaba “Posición Estadística” que consistía en un número de 8 cifras a una variable denominada “Sección” que contiene valores desde el 1 hasta el 21.

- **Valor (Monetario)**

El valor es la cantidad monetaria que cuesta cada uno de los envíos en céntimos de Euro.

Por último, para finalizar la descripción de los datos, se muestran tres ejemplos de filas de los archivos de datos, descargados de la página web de la Agencia Tributaria, con el objetivo de que se entienda el significado de estos:

17 E 01 15 CN 64035999 4 000000000020 000000000014 000000547928 15

- **Año:** 2017
- **Flujo:** Exportación
- **Provincia:** A Coruña
- **País:** China
- **Posición:** 64035999
 - *Capítulo:* 64 (CALZADO, POLAINAS Y ARTÍCULOS ANÁLOGOS; PARTES DE ESTOS ARTÍCULOS)
 - *Sección:* 12 (CALZADO, SOMBREROS Y DEMÁS TOCADOS, PARAGUAS, QUITASOLES, BASTONES, LÁTIGOS, FUSTAS Y SUS PARTES; PLUMAS PREPARADAS Y ARTÍCULOS DE PLUMAS; FLORES ARTIFICIALES; MANUFACTURAS DE CABELLO)
- **Valor:** 547928 céntimos

17 E 48 26 BR 90158020 4 000000000165 000000000000 000005517051 26

- **Año:** 2017
- **Flujo:** Exportación
- **Provincia:** La Rioja
- **País:** Brasil
- **Posición:** 90158020
 - *Capítulo:* 90 (INSTRUMENTOS Y APARATOS DE ÓPTICA, FOTOGRAFÍA O CINEMATOGRAFÍA, DE MEDIDA, CONTROL O PRECISIÓN; INSTRUMENTOS Y APARATOS MEDICOQUIRÚRGICOS; PARTES Y ACCESORIOS DE ESTOS INSTRUMENTOS O APARATOS)
 - *Sección:* 18 (INSTRUMENTOS Y APARATOS DE ÓPTICA, FOTOGRAFÍA O CINEMATOGRAFÍA, DE MEDIDA, CONTROL O PRECISIÓN; INSTRUMENTOS Y APARATOS MEDICOQUIRÚRGICOS; APARATOS DE RELOJERÍA; INSTRUMENTOS MUSICALES; PARTES Y ACCESORIOS DE ESTOS INSTRUMENTOS O APARATOS)
- **Valor:** 5517051 céntimos

17 I 28 05 DE 84219910 3 000000000132 000000000000 000001031750 28

- **Año:** 2017
- **Flujo:** Importación
- **Provincia:** Ávila
- **País:** Alemania
- **Posición:** 84219910
 - *Capítulo:* 84 (REACTORES NUCLEARES, CALDERAS, MÁQUINAS, APARATOS Y ARTEFACTOS MECÁNICOS; PARTES DE ESTAS MÁQUINAS O APARATOS)
 - *Sección:* 16 (MÁQUINAS Y APARATOS, MATERIAL ELÉCTRICO Y SUS PARTES; APARATOS DE GRABACIÓN O REPRODUCCIÓN DE SONIDO, APARATOS DE GRABACIÓN O REPRODUCCIÓN DE IMÁGENES Y SONIDO EN TELEVISIÓN, Y LAS PARTES Y ACCESORIOS DE ESTOS APARATOS)
- **Valor:** 1031750 céntimos

3.3 Preparación de los datos

El objetivo final del preprocesado de los datos, es la obtención de varias matrices a las que se puedan aplicar técnicas multivariantes. La matriz que se pretende construir está formada por 52 filas que se corresponden con las provincias y 21 columnas que se corresponden con las secciones de mercancías. Cada uno de los valores de la matriz se corresponderá con la suma de todos los envíos de una provincia para una determinada sección.

Se procede a hacer una distinción entre importaciones y exportaciones, por lo que se crearán dos matrices, para cada uno de los dos flujos posibles. Se procede con los siguientes pasos.

- Lectura de los 10 ficheros .zip (Ilustración 15)

```
# Lectura de los ficheros .zip (Directorio actual)
ficheros=dir(".", "zip$", full.names=TRUE)
```

Ilustración 15. Código R leer ficheros comprimidos

- Se utilizan dos archivos auxiliares en formato CSV, creados “exprofeso”, donde se incluyen los nombres de cada una de las provincias indexadas con el código del INE. Además, se incluyen el nombre de los capítulos de mercancías, así como a qué sección pertenece cada uno. (Ilustración 16)

```
# Lectura de tablas auxiliares con los nombres de los
# capítulos, secciones y provincias (Directorio actual).

capitulos=read.csv2("CapitulosUTF8.csv", colClasses="character", row.names=1)
provincias=read.csv2("todasprovUTF8.csv", colClasses="character", row.names=1)
```

Ilustración 16. Código R cargar archivos auxiliares

- Se crea una función, mostrada en la Ilustración 17, para leer la información necesaria de cada uno de los ficheros. En esta función tenemos como entrada los ficheros de datos descritos en la sección anterior.

Un ejemplo de cómo aparecen los datos en los ficheros es el siguiente:

17 E 48 26 BR 90158020 4 000000000165 000000000000 000005517051 26

A partir de millones de filas como la anterior, se pretende crear dos matrices de 52 filas y 21 columnas.

- Para cada uno de los 10 ficheros, se crea un data.frame con 6 variables:
 - Año
 - Flujo
 - Provincia
 - País
 - Sección
 - Valor
- Los valores se obtienen de la siguiente forma:
 - Año. Primeras dos cifras de cada observación.
 - Flujo. Carácter que se encuentra en la posición cuarta.
 - Provincia. Valor que se encuentra en la posición 9 y 10. Utilizando el archivo auxiliar y el índice se obtiene el nombre de la provincia a partir del número de dos cifras leído.
 - País. Caracteres que se encuentran en la posición 12 y 13.
 - Sección. Valor que se encuentra en la posición 15 y 16. Utilizando el archivo auxiliar y el índice se obtiene el nombre de la sección a la que pertenece el envío.
 - Valor. Cifras que se encuentran en las posiciones 52-63.
- A medida que se completa la lectura de un fichero y se empieza el siguiente, en el mismo data.frame se siguen almacenando a continuación de las anteriores, las nuevas observaciones leídas.
- Tras finalizar la lectura de los 10 ficheros, en el data.frame hay una entrada por cada una de las observaciones.
- Aquellas observaciones que son iguales, es decir, mismo flujo, mismo país, misma provincia y sección, se agrupan sumando su valor, reduciéndose el número de filas del data.frame.

```

# Función para leer los datos que se necesitan del fichero.

LeerUno=function(fichero){
  nombre=split(basename(fichero),".")[[1]][1]
  con=unz(fichero,nombre)
  lineas=readLines(con)[-1] # Se elimina la cabecera
  close(con)
  datos=data.frame(Año=substr(lineas,1,2),
                  Flujo=substr(lineas,4,4),
                  Provincia=provincias[as.numeric(substr(lineas,9,10)),"PROVINCIA"],
                  Pais=substr(lineas,12,13),
                  #Capitulo=capitulos[as.numeric(substr(lineas,15,16)),"Capitulo"],
                  Seccion=capitulos[as.numeric(substr(lineas,15,16)),"Seccion"],
                  Valor=as.numeric(substr(lineas,52,63)), # En céntimos
                  stringsAsFactors=FALSE)
  datosRes=aggregate(datos$Valor,datos[,-(which(colnames(datos)== "Valor"))],FUN=sum)
  colnames(datosRes)=gsub("x","Valor",colnames(datosRes))
  invisible(datosRes)
}
datos=do.call(rbind,lapply(ficheros,LeerUno))

```

Ilustración 17. Función para leer los datos

Una vez que se ha ejecutado la función de la Ilustración 17, el resultado obtenido es el `data.frame` que se muestra en la Ilustración 18.

Año	Flujo	Provincia	Pais	Seccion	Valor
17	E	Cantabria	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	404692756
17	I	Cantabria	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	261879048
17	E	Girona	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	6203945
17	E	Guadalajara	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	3418255
17	E	Guipúzcoa	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	179232907
17	I	Guipúzcoa	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	59231451
17	E	Jaén	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	1547963
17	E	Lleida	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	1785000
17	E	Murcia	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	7540310
17	E	Pontevedra	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	132720058
17	E	Sevilla	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	7241760
17	I	Sevilla	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	32570587
17	E	Valencia	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	25967848
17	E	Vizcaya	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	51203903
17	E	Zaragoza	AL	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	5810718
17	E	Barcelona	AM	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	5451000
17	E	Jaén	AM	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	964412
17	E	Lugo	AM	SECCIÓN 1 - ANIMALES VIVOS Y PRODUCTOS DEL REI...	3364556

Ilustración 18. Ejemplo resultado tras función

En la imagen anterior se puede apreciar el valor monetario ya agrupado para cada flujo, provincia, país y sección.

A partir de este `data.frame` se crean dos matrices de datos:

- 1) Exportaciones de provincias por secciones.
- 2) Importaciones de provincias por secciones.

- Creación de matrices para exportaciones e importaciones. (Ilustración 19)

```
# Creación de una tabla de Provincias x Sección, para Exportaciones e Importaciones.

exporProvXSeccion=apply(datos[datos$Flujo=="E", "Valor"],
                        datos[datos$Flujo=="E", c("Provincia", "Seccion")],
                        sum)
imporProvXSeccion=apply(datos[datos$Flujo=="I", "Valor"],
                        datos[datos$Flujo=="I", c("Provincia", "Seccion")],
                        sum)
```

Ilustración 19. Código R creación matrices de datos

En la Ilustración 20 se presenta un ejemplo de una parte de una de las matrices obtenidas tras la realización del preprocesado de los datos.

	S1	S2	S3	S4	S5	S6
A Coruña	46292729999	18813337211	20230414700	43894108588	262107897657	1.224772e+10
Álava	4264932682	2635712449	1196203261	8327953287	2586993831	1.554021e+10
Albacete	7840237524	3265860562	92244946	1885512044	523021853	6.132259e+09
Alicante	15500342009	38050793644	301660729	21266150551	1805733933	2.096548e+10
Almería	4897561978	36298806070	655213863	315074840	117912458956	9.548824e+09
Asturias	13098393112	7342968410	853881446	5160578968	242605551315	2.449553e+10
Ávila	593615451	3943444155	24122192	724945404	63541435	3.447036e+09
Badajoz	3618650858	6020028366	2782242764	5681512074	692492022	6.867407e+09
Baleares	5980925590	3254363172	52901143	4609066435	29141869173	2.764203e+10
Barcelona	171688601517	301276101312	55337019118	275567346313	239111577482	1.146050e+12
Burgos	16201340990	4272644622	1060468718	13127016951	1125965875	6.779369e+10
Cáceres	4587368358	850669065	114543195	899054296	954219009	1.084572e+09

Ilustración 20. Ejemplo matriz de datos de Importación

Con la construcción de estas matrices, se ha conseguido agrupar la información inicial que contaba con aproximadamente 3 millones de observaciones en dos matrices de 52 filas por 21 columnas.

El volumen de exportaciones e importaciones depende de la población de la provincia. No es lo mismo Madrid con 5 millones de habitantes que Melilla que no llega a 100.000 habitantes. También afectan otros factores como la estructura industrial y comercial de cada provincia.

Se han obtenido los datos del INE (Tabla 8) correspondientes al número de habitantes por provincia para el año 2017 y se han obtenido dos nuevas matrices utilizando las dos anteriores de provincias por secciones [9].

Tabla 8. Población provincias 2017

	2017
Total	46.572.132
02 Albacete	390.032
03 Alicante/Alacant	1.825.332
04 Almería	706.672
01 Araba/Álava	326.574

33 Asturias	1.034.960
05 Ávila	160.700
06 Badajoz	679.884
07 Balears, Illes	1.115.999
08 Barcelona	5.576.037
48 Bizkaia	1.148.302
09 Burgos	358.171
10 Cáceres	400.036
11 Cádiz	1.239.435
39 Cantabria	580.295
12 Castellón/Castelló	575.470
13 Ciudad Real	502.578
14 Córdoba	788.219
15 Coruña, A	1.120.294
16 Cuenca	198.718
20 Gipuzkoa	719.282
17 Girona	755.716
18 Granada	912.938
19 Guadalajara	253.310
21 Huelva	518.930
22 Huesca	219.702
23 Jaén	643.484
24 León	468.316
25 Lleida	432.384
27 Lugo	333.634
28 Madrid	6.507.184
29 Málaga	1.630.615
30 Murcia	1.470.273
31 Navarra	643.234
32 Ourense	311.680
34 Palencia	163.390
35 Palmas, Las	1.100.480
36 Pontevedra	942.731
26 Rioja, La	315.381
37 Salamanca	333.603
38 Santa Cruz de Tenerife	1.007.641
40 Segovia	154.184
41 Sevilla	1.939.527
42 Soria	88.903
43 Tarragona	791.693
44 Teruel	135.562
45 Toledo	686.841
46 Valencia/València	2.540.707
47 Valladolid	521.130
49 Zamora	177.404
50 Zaragoza	953.486
51 Ceuta	84.959
52 Melilla	86.120

- Con los datos de la Tabla 8, se ha dividido el valor monetario calculado en las dos matrices anteriores entre la población para cada provincia. De esta forma, se obtienen dos nuevas matrices que contienen tanto para las exportaciones e importaciones el valor monetario por habitante.
- Además, utilizando las dos matrices de datos originales se han construido otras dos nuevas matrices calculando el porcentaje que supone cada sección para las provincias. Para ello se ha dividido cada valor de las filas de la matriz entre el total de dicha fila, es decir entre el total de la provincia.

En definitiva, los 3 conjuntos de datos que se utilizan, en adelante se denominarán:

- 1) Valor monetario total de exportaciones/importaciones en euros.
- 2) Valor monetario de exportaciones/importaciones en euros por habitante.
- 3) Porcentaje del valor de exportaciones/importaciones de cada sección por provincia.

4 Análisis exploratorio de los datos

Antes de aplicar cualquier técnica de análisis multivariante es preciso realizar un análisis exploratorio previo de los datos disponibles. Se examinarán las variables individuales y las relaciones entre ellas y la presencia de datos anómalos o atípicos.

4.1 Análisis de casos peculiares con secciones de mercancías ausentes

Se utilizan las siguientes órdenes (Ilustración 21) para averiguar si existe algún valor ausente en las matrices.

```
> sum(is.na(exporProvXSeccion))  
[1] 15  
> sum(is.na(imporProvXSeccion))  
[1] 0
```

Ilustración 21. Código R contar NAs

Se comprueba que la matriz de importaciones no tiene ningún valor ausente, sin embargo, para la matriz de exportaciones hay 15 valores ausentes (NAs).

Los valores ausentes (NAs) se han producido al intentar sumar valores que no existían, es decir, hay provincias que no exportan ciertas secciones de mercancías. Los 15 valores ausentes respecto de los 315 valores totales de la matriz, representa un 1.37%. Los valores ausentes se corresponden con las siguientes provincias y secciones:

- 1) Sección 3 en Melilla.
- 2) Sección 13 en Ceuta.
- 3) Sección 14 en Melilla.
- 4) Sección 14 en Palencia.
- 5) Sección 19 en Ávila.
- 6) Sección 19 en Burgos.
- 7) Sección 19 en Cáceres.
- 8) Sección 19 en Huesca.
- 9) Sección 19 en La Rioja.
- 10) Sección 19 en Ourense.
- 11) Sección 19 en Palencia.
- 12) Sección 19 en Segovia.
- 13) Sección 19 en Soria.
- 14) Sección 19 en Teruel.
- 15) Sección 19 en Ceuta.

También se comprueba respecto de los valores ausentes lo siguiente:

- 4/15 se corresponden con ciudades las ciudades autónomas de Ceuta y Melilla.
- 11/15 se corresponden con la Sección 19 (Armas, Municiones y sus partes y accesorios).

- 5/10 son provincias de Castilla y León.

Analizando estos aspectos, se puede llegar a la conclusión de que no se trata de un error al obtener o codificar los datos, sino que al tratarse la mayoría de ellos de una sección de mercancía tan específica como es la de Armas y Municiones y más aún siendo exportaciones, se puede entender que estas provincias no fabrican armas, por lo que no las exportan.

En cuanto al resto de los valores ausentes, tampoco sería descabellado pensar que esas provincias no exporten ningún tipo de mercancía de ese tipo. Se tratan de provincias que no son las más grandes en cuanto a población ni potencia económica.

En conclusión, los valores ausentes se toman como valores que son zeros, debido a que esas provincias no exportaron esa determinada mercancía. (Ilustración 22)

```
# Se sustituyen los valores nulos (NAs) por cero
# Se considera que esa provincia no ha exportado la sección determinada

exporProvXSeccion[is.na(exporProvXSeccion)] <- 0
importProvXSeccion[is.na(importProvXSeccion)] <- 0
```

Ilustración 22. Código R sustituir NAs

4.2 Análisis exploratorio de las variables

Se muestran una serie de gráficos de barras que permiten clasificar las provincias en función de la cantidad de dinero que importan o exportan, según corresponda.

Se realizará el análisis para cada uno de los conjuntos de datos agregados que se describen en el capítulo anterior de esta memoria de TFG [10].

4.2.1 Valor monetario total de exportaciones/importaciones en euros

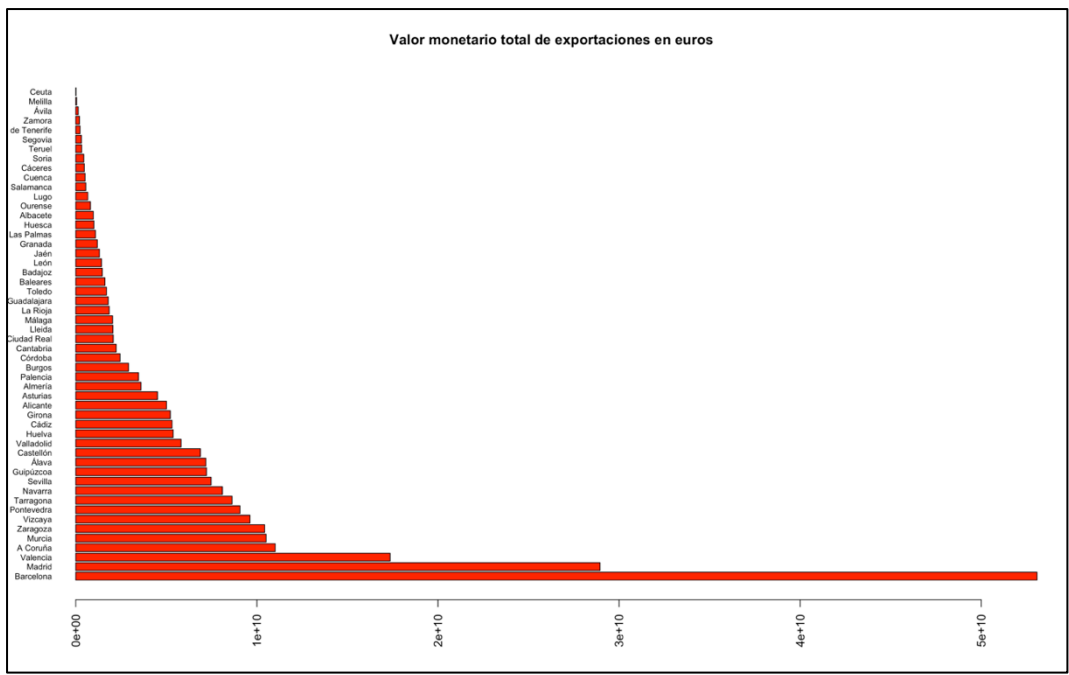


Ilustración 23. Gráfico barras exportaciones provincias

Provincias con mayor valor de exportaciones:

- 1) Barcelona
- 2) Madrid
- 3) Valencia
- 4) A Coruña
- 5) Murcia

Provincias con menor valor de exportaciones:

- 1) Ceuta
- 2) Melilla
- 3) Ávila
- 4) Zamora
- 5) Santa Cruz de Tenerife

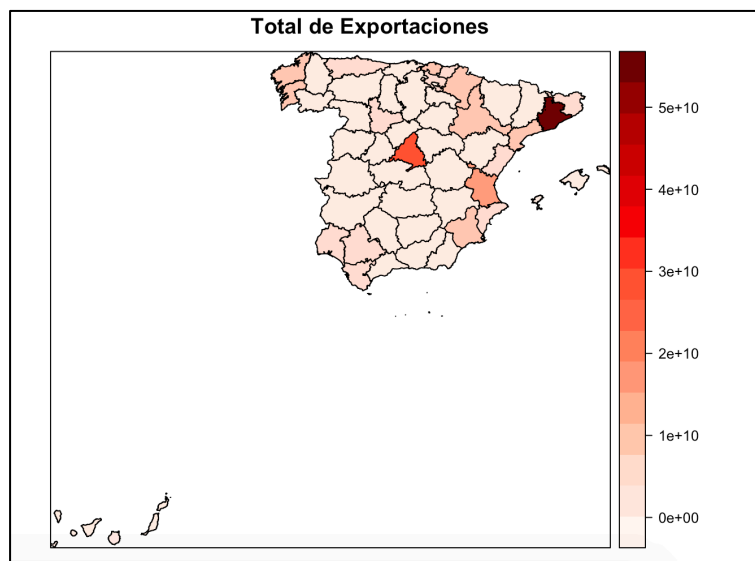


Ilustración 24. Mapa de exportaciones totales

Se puede ver que, las provincias que más exportan después de Madrid, la capital de España, son provincias costeras con puertos. Por tanto, un aspecto importante para tener en cuenta a la hora de clasificar las provincias que más exportan es la presencia del mar en su frontera.

En cuanto a las provincias con menos exportaciones, hay que decir que tanto Ceuta, Melilla, Ávila y Zamora son cuatro de las provincias menos pobladas de España, excepto Santa Cruz de Tenerife no lo es, ya que es la decimocuarta provincia más poblada de España. Influye la población al considerar las provincias que menos exportan, pero no es el principal factor. Para comprender las razones de esto habría que investigar otros elementos como por ejemplo el impulso de la industria y otros sectores económicos en la provincia.

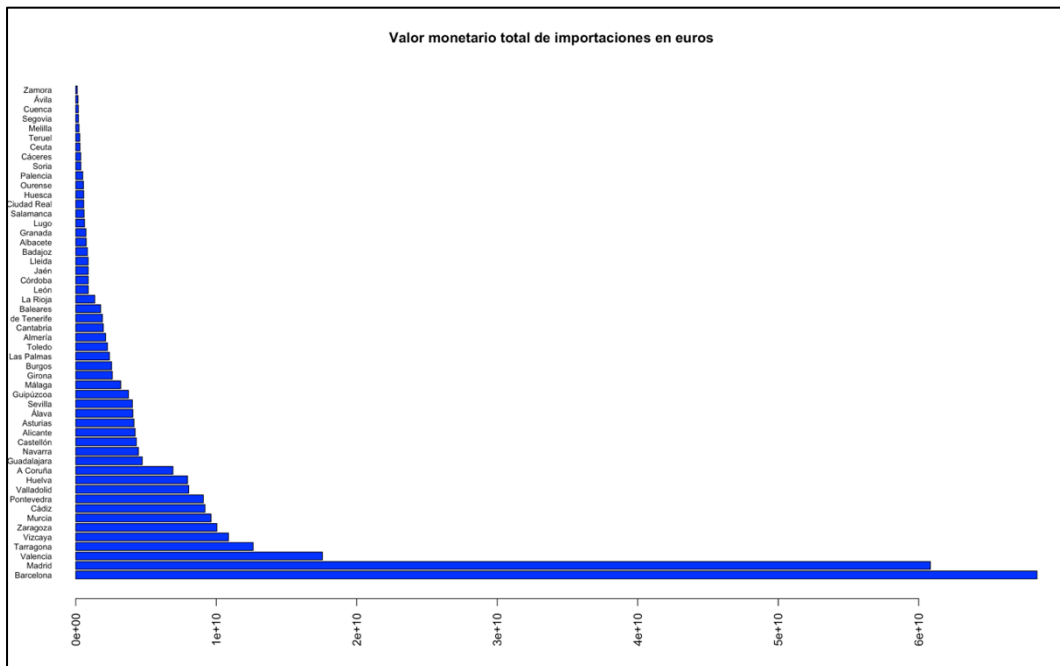


Ilustración 25. Gráfico barras importaciones provincias

Provincias con mayor valor de importaciones:

- 1) Barcelona
- 2) Madrid
- 3) Valencia
- 4) Tarragona
- 5) Vizcaya

Provincias con menor valor de importaciones:

- 1) Zamora
- 2) Ávila
- 3) Cuenca
- 4) Segovia
- 5) Melilla

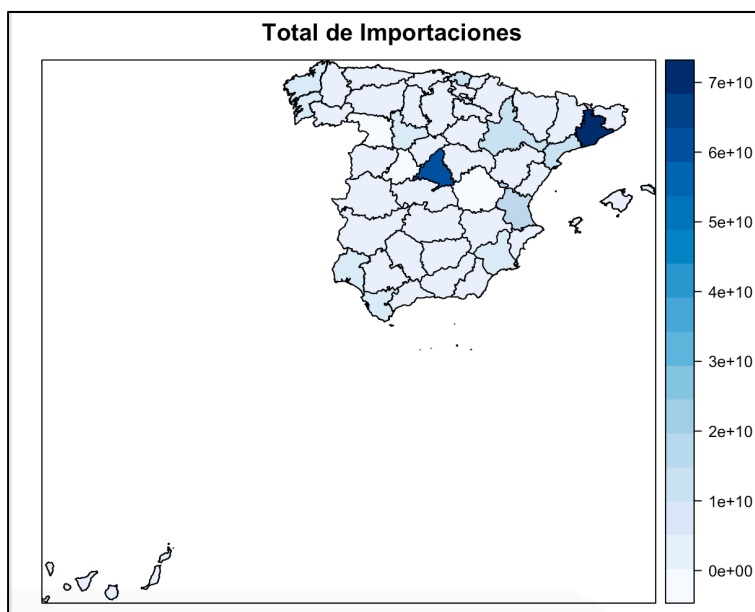


Ilustración 26. Mapa de importaciones totales

Las provincias con valores más altos de importaciones siguen siendo provincias costeras, a excepción de Madrid. En lo más bajo de la clasificación, aparecen provincias que están entre las menos pobladas del país.

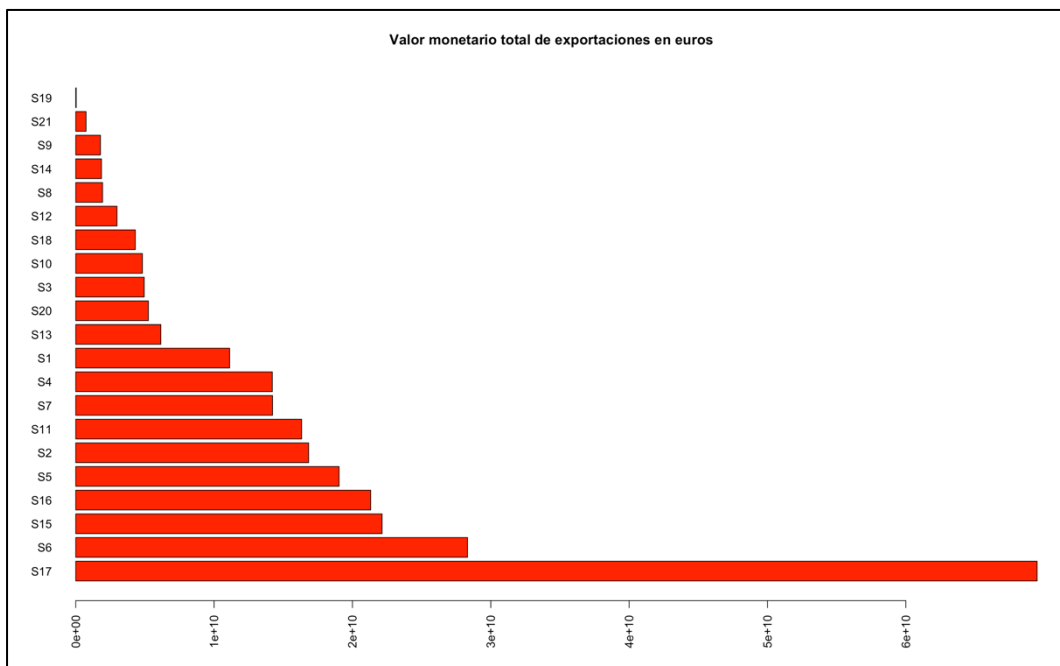


Ilustración 27. Gráfico barras exportaciones secciones

Secciones de mercancías con mayor valor de las exportaciones:

- 1) Sección 17. Material de transporte.
- 2) Sección 6. Productos de las industrias químicas o de las industrias conexas.
- 3) Sección 15. Metales comunes y manufacturas de estos metales.

Secciones de mercancías con menor valor de las exportaciones:

- 1) Sección 19. Armas, municiones, y sus partes y accesorios.
- 2) Sección 21. Objetos de arte o colección y antigüedades.
- 3) Sección 9. Madera, carbón vegetal y manufacturas de madera; corcho y sus manufacturas; manufacturas de espartería o cestería.

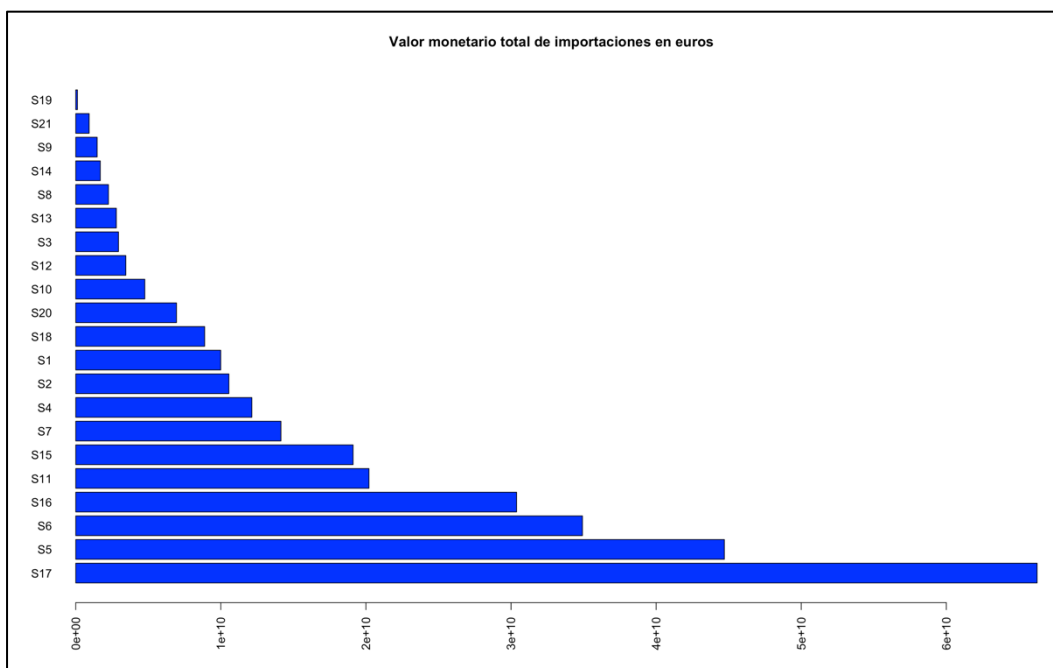


Ilustración 28. Gráfico barras importaciones secciones

Secciones de mercancías con mayor valor de las importaciones:

- 1) Sección 17. Material de transporte.
- 2) Sección 5. Productos minerales.
- 3) Sección 6. Productos de las industrias químicas o de las industrias conexas.

Secciones de mercancías con menor valor de las importaciones:

- 1) Sección 19. Armas, municiones, y sus partes y accesorios.
- 2) Sección 21. Objetos de arte o colección y antigüedades.
- 3) Sección 9. Madera, carbón vegetal y manufacturas de madera; corcho y sus manufacturas; manufacturas de espartería o cestería.

Hay que destacar que las secciones de mercancías que más se exportan, son las que a la vez más se importan.

Algo muy similar pasa con las secciones de mercancías que menos se exportan e importan.

En la tabla 9 se puede observar un resumen de la información que se presenta en los mapas anteriores con las secciones de mercancías que más se exportan e importan para cada una de las provincias.

Tabla 9. Secciones punteras

PROVINCIA	Sección con más exportaciones	Sección con más importaciones
Álava	S17. Material de transporte.	S17. Material de transporte.
Albacete	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de grabación o reproducción de sonido e imágenes y sus accesorios.
Alicante	S12. Calzado y sombreros.	S12. Calzado y sombreros.
Almería	S2. Productos del reino vegetal.	S5. Productos minerales.

Ávila	S17. Material de transporte.	S2. Productos del reino vegetal.
Badajoz	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S15. Metales comunes y manufacturas de estos metales.
Baleares	S17. Material de transporte.	S17. Material de transporte.
Barcelona	S17. Material de transporte.	S17. Material de transporte.
Burgos	S6. Productos de las industrias químicas o de las industrias conexas.	S6. Productos de las industrias químicas o de las industrias conexas.
Cáceres	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de grabación o reproducción de sonido e imágenes y sus accesorios.
Cádiz	S5. Productos minerales.	S5. Productos minerales.
Castellón	S13. Manufacturas de piedra, productos cerámicos.	S5. Productos minerales.
Ciudad Real	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S17. Material de transporte.
Córdoba	S15. Metales comunes y manufacturas de estos metales.	S15. Metales comunes y manufacturas de estos metales.
A Coruña	S11. Materias textiles y sus manufacturas.	S5. Productos minerales.
Cuenca	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S17. Material de transporte.
Girona	S1. Animales vivos y productos del reino animal.	S6. Productos de las industrias químicas o de las industrias conexas.
Granada	S2. Productos del reino vegetal.	S6. Productos de las industrias químicas o de las industrias conexas.
Guadalajara	S11. Materias textiles y sus manufacturas.	S17. Material de transporte.
Guipúzcoa	S16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de grabación o reproducción de sonido e imágenes y sus accesorios.	S15. Metales comunes y manufacturas de estos metales.
Huelva	S5. Productos minerales.	S5. Productos minerales.
Huesca	S1. Animales vivos y productos del reino animal.	S6. Productos de las industrias químicas o de las industrias conexas.
Jaén	S17. Material de transporte.	S17. Material de transporte.
León	S17. Material de transporte.	S16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de

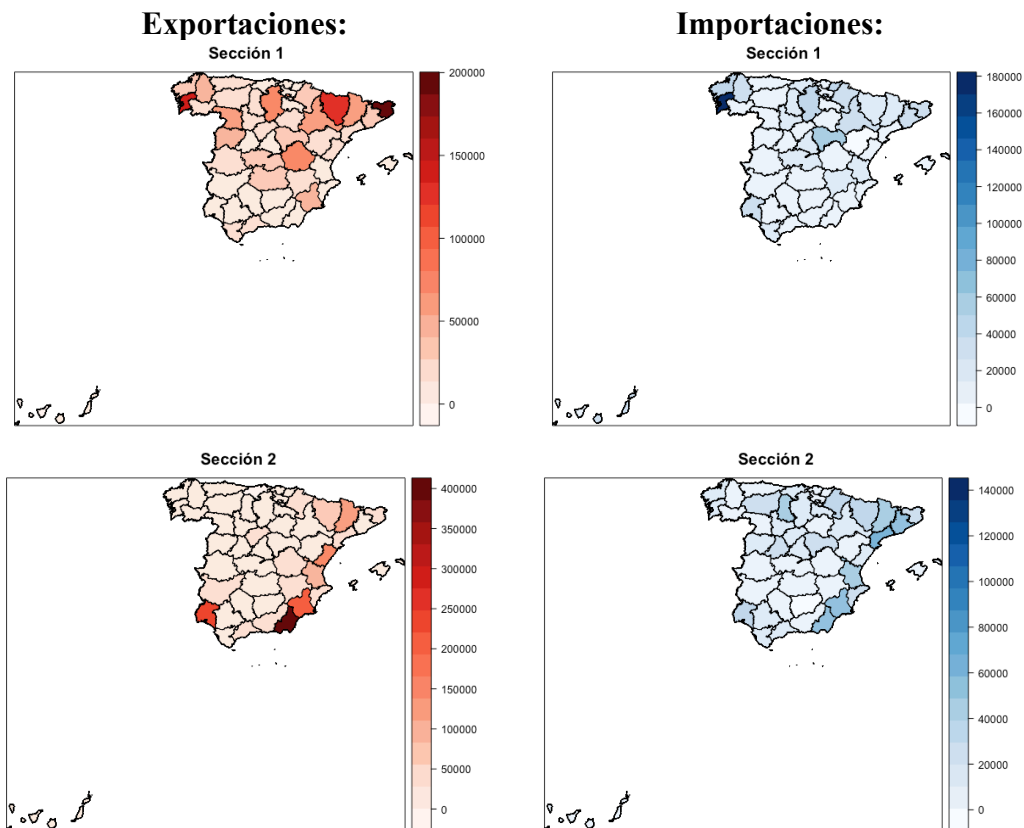
		grabación o reproducción de sonido e imágenes y sus accesorios.
Lleida	S2. Productos del reino vegetal.	S2. Productos del reino vegetal.
La Rioja	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.
Lugo	S15. Metales comunes y manufacturas de estos metales.	S1. Animales vivos y productos del reino animal.
Madrid	S17. Material de transporte.	S17. Material de transporte.
Málaga	S2. Productos del reino vegetal.	S17. Material de transporte.
Murcia	S5. Productos minerales.	S5. Productos minerales.
Navarra	S17. Material de transporte.	S17. Material de transporte.
Ourense	S13. Manufacturas de piedra, productos cerámicos.	S11. Materias textiles y sus manufacturas.
Asturias	S15. Metales comunes y manufacturas de estos metales.	S5. Productos minerales.
Palencia	S17. Material de transporte.	S17. Material de transporte.
Las Palmas	S5. Productos minerales.	S5. Productos minerales.
Pontevedra	S17. Material de transporte.	S17. Material de transporte.
Salamanca	S1. Animales vivos y productos del reino animal.	S6. Productos de las industrias químicas o de las industrias conexas.
Tenerife	S5. Productos minerales.	S5. Productos minerales.
Cantabria	S15. Metales comunes y manufacturas de estos metales.	S15. Metales comunes y manufacturas de estos metales.
Segovia	S2. Productos del reino vegetal.	S7. Plástico, caucho y sus manufacturas.
Sevilla	S17. Material de transporte.	S16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de grabación o reproducción de sonido e imágenes y sus accesorios.
Soria	S17. Material de transporte.	S17. Material de transporte.
Tarragona	S5. Productos minerales.	S5. Productos minerales.
Teruel	S17. Material de transporte.	S5. Productos minerales.
Toledo	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.	S6. Productos de las industrias químicas o de las industrias conexas.
Valencia	S17. Material de transporte.	S17. Material de transporte.
Valladolid	S17. Material de transporte.	S17. Material de transporte.
Vizcaya	S15. Metales comunes y manufacturas de estos metales.	S5. Productos minerales.

Zamora	S1. Animales vivos y productos del reino animal.	S4. Productos de las industrias alimentarias; bebidas alcohólicas y vinagre; tabaco.
Zaragoza	S17. Material de transporte.	S17. Material de transporte.
Ceuta	S7. Plástico, caucho y sus manufacturas.	S5. Productos minerales.
Melilla	S17. Material de transporte.	S2. Productos del reino vegetal.

4.2.2 Valor monetario de exportaciones/importaciones en euros por habitante

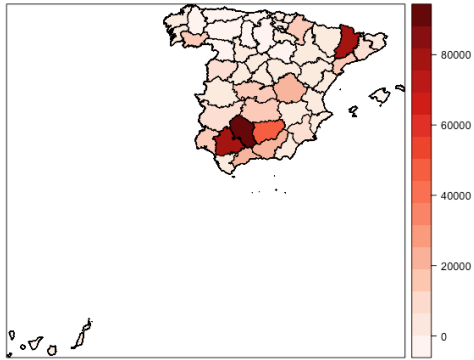
Se muestran unos mapas de España por provincias, para cada una de las 21 secciones de mercancías. Aparecerán, mediante una escala de color, con un color más oscuro aquellas provincias que más dinero exporten o importen por habitante, según corresponda y en un color más claro las que menos [11].

Si los mapas se realizaran con los valores en euros sin tener en cuenta el tamaño de la población, siempre aparecería en primer lugar Barcelona y Madrid apareciendo muy poco o casi nada representadas las demás provincias, especialmente aquellas con poca población. Por ser más informativos, los mapas se han realizado utilizando el conjunto de importaciones/exportaciones por habitante, dividiendo los euros entre el número de habitantes de cada provincia obtenido del INE.

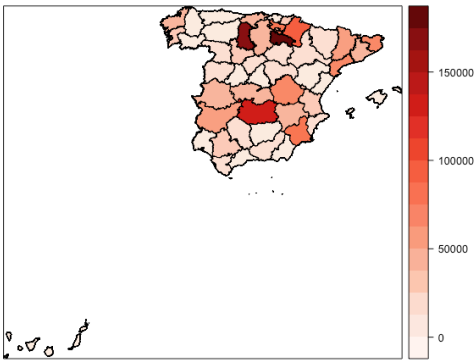


Exportaciones:

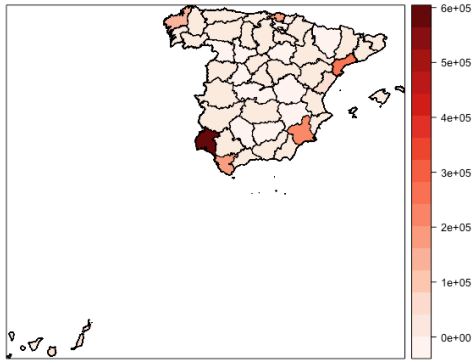
Sección 3



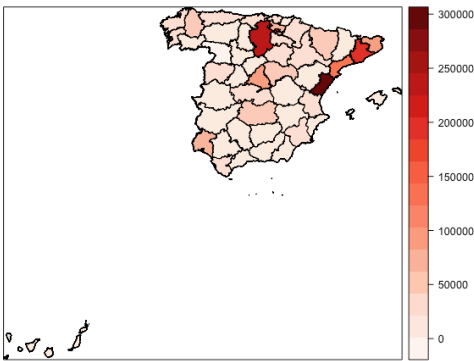
Sección 4



Sección 5

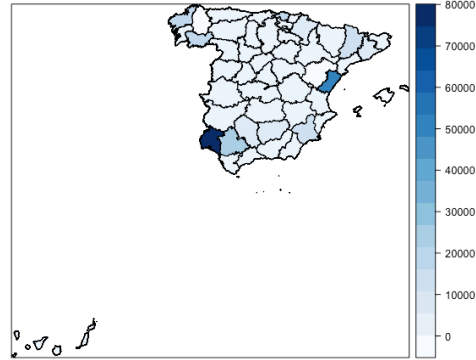


Sección 6

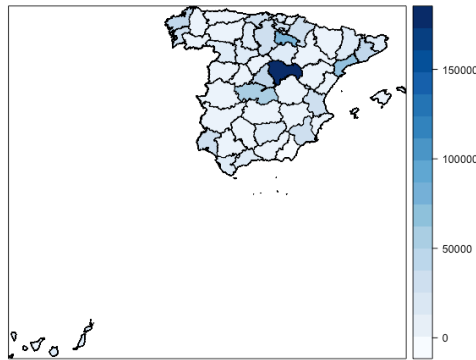


Importaciones:

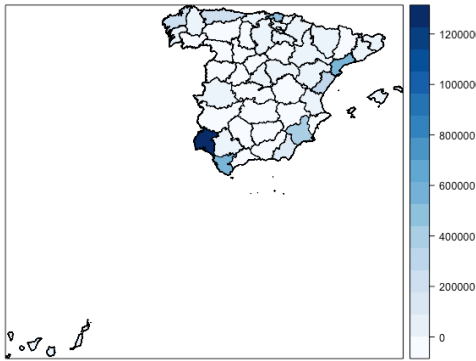
Sección 3



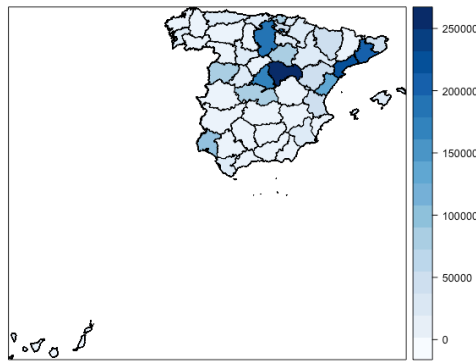
Sección 4



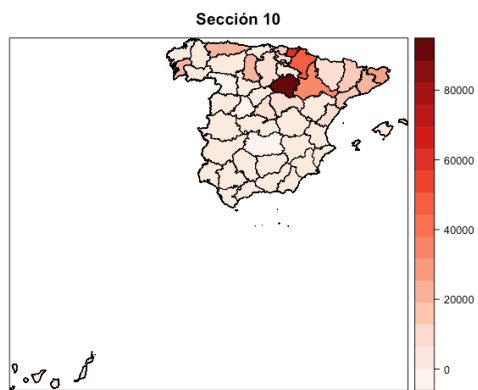
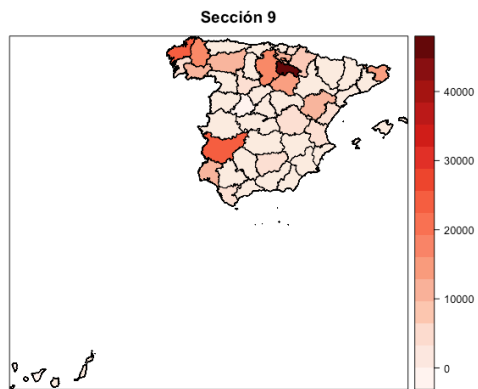
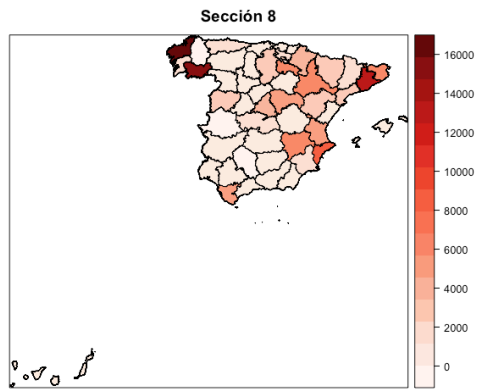
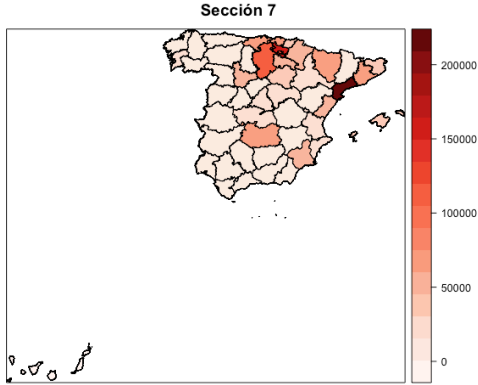
Sección 5



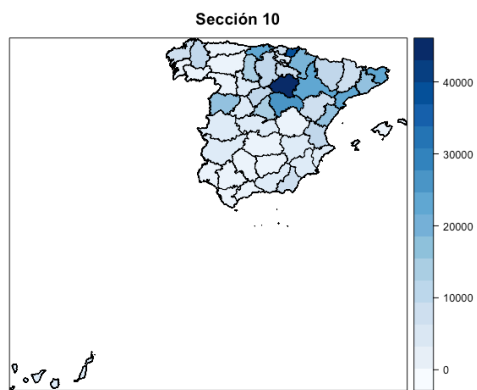
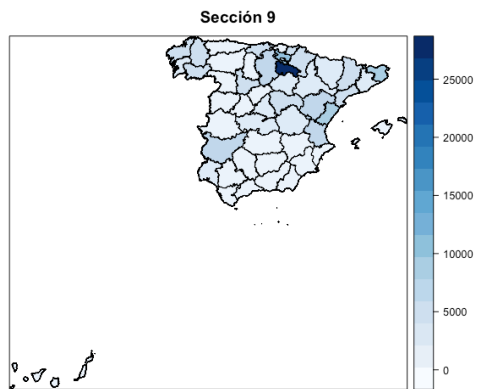
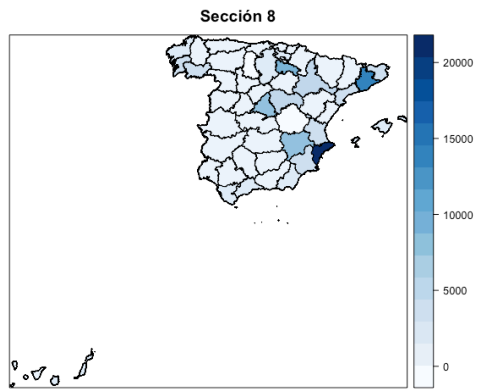
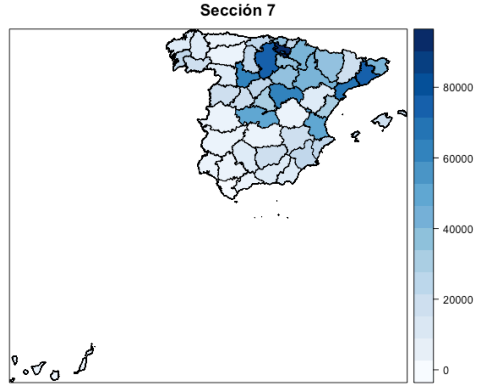
Sección 6



Exportaciones:

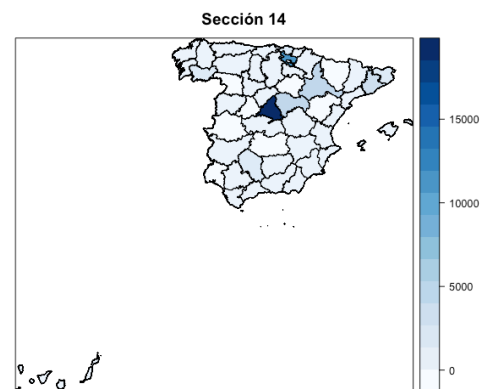
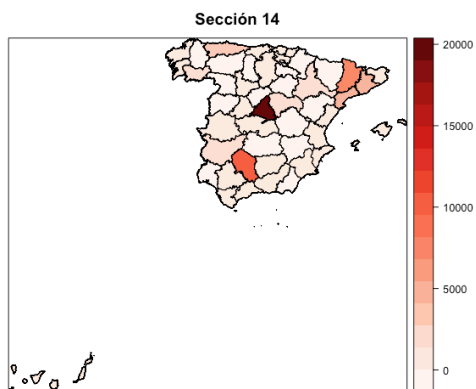
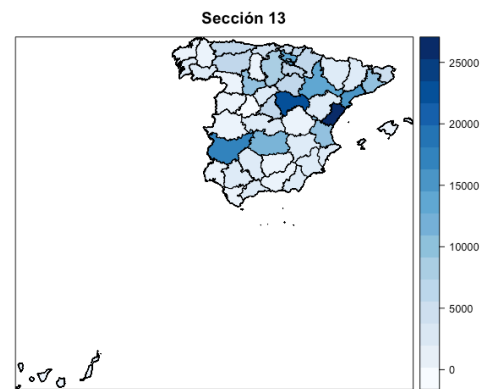
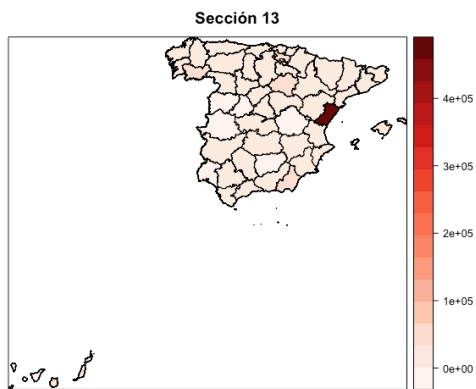
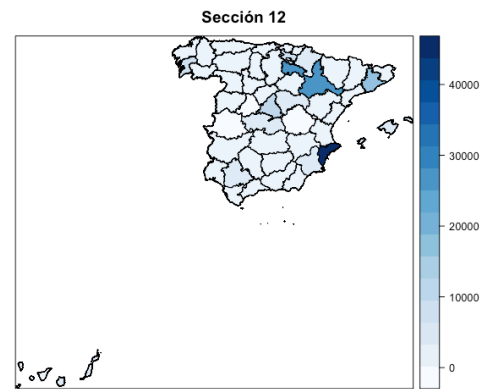
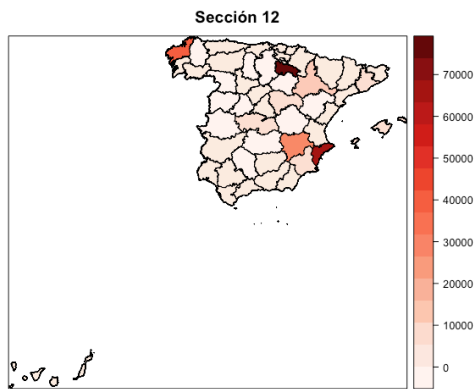
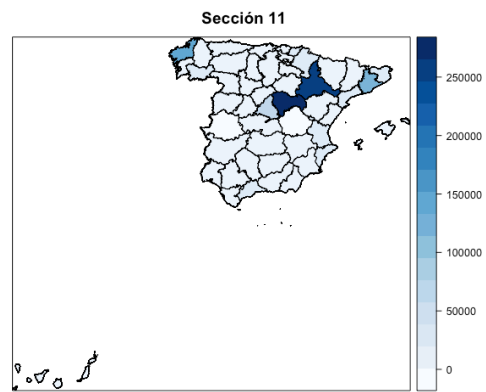
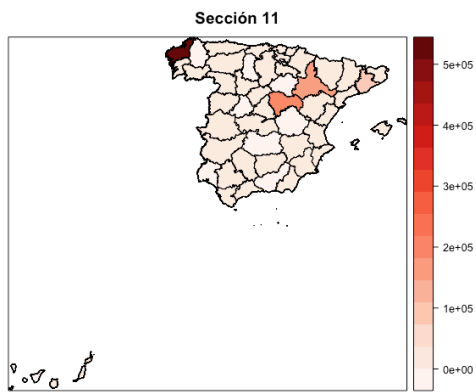


Importaciones:



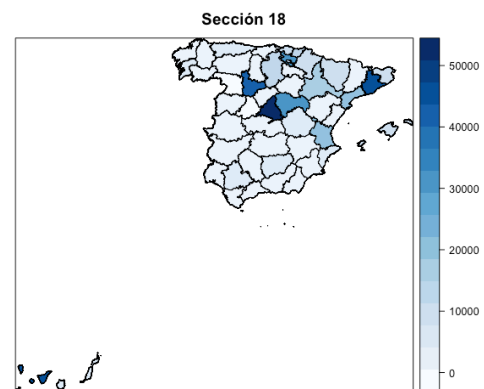
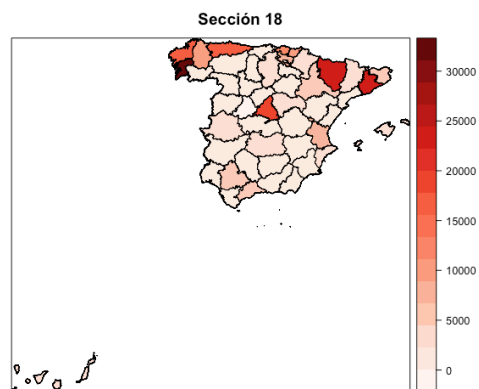
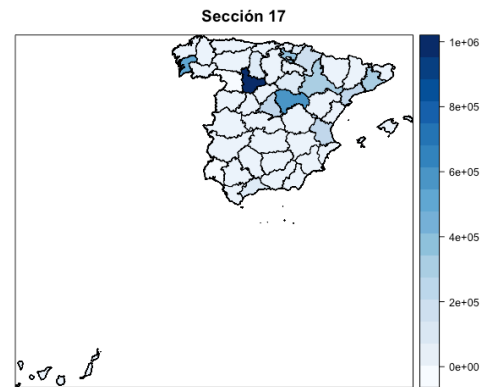
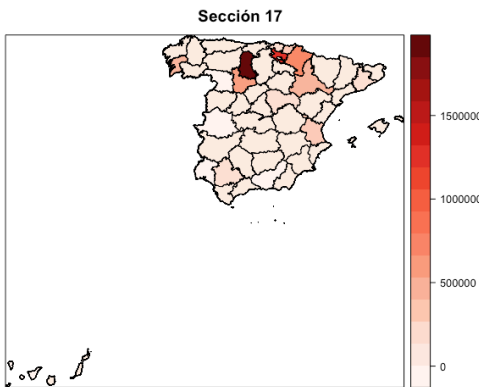
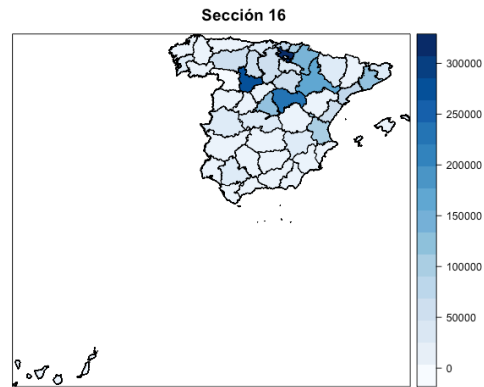
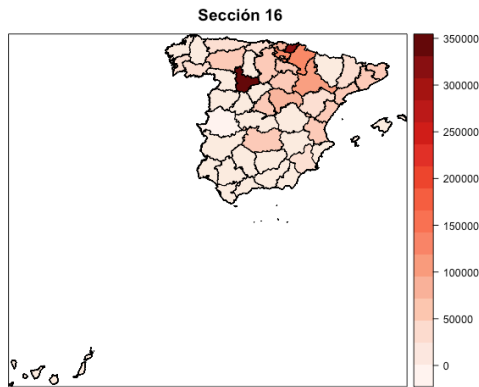
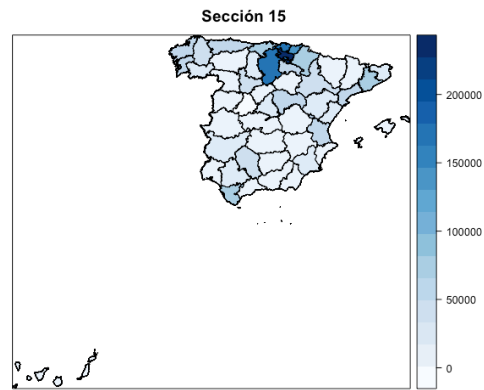
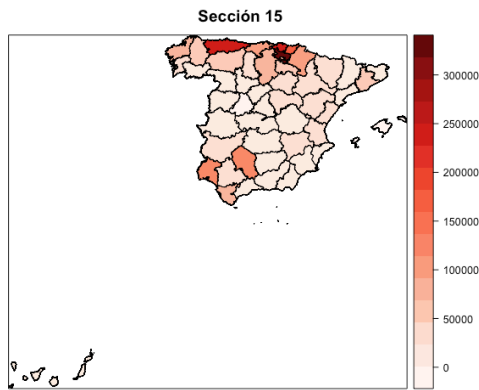
Exportaciones:

Importaciones:

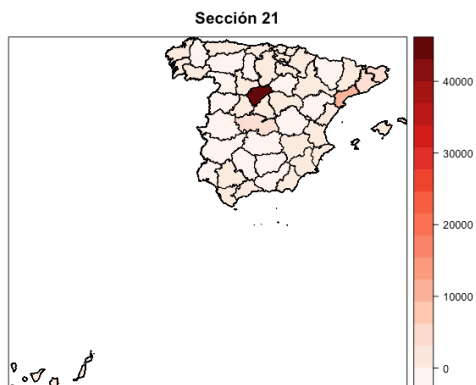
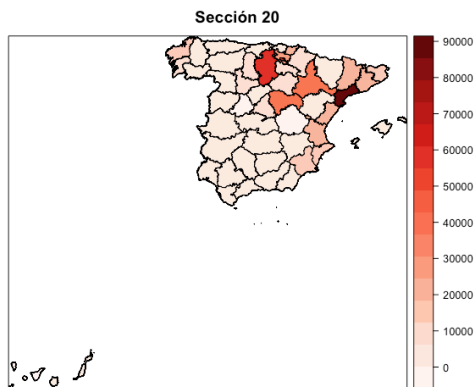
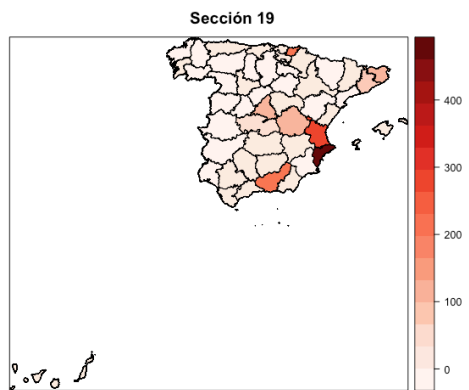


Exportaciones:

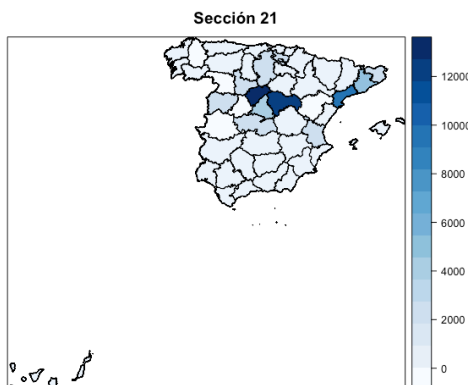
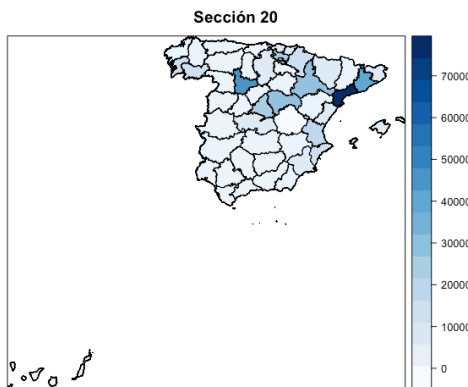
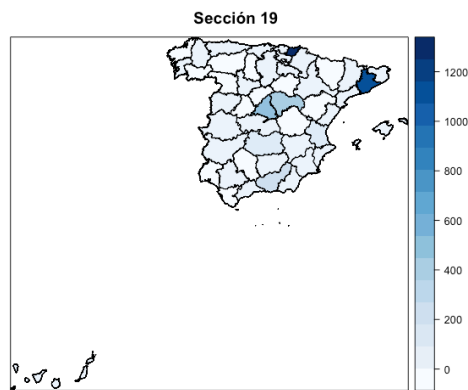
Importaciones:



Exportaciones:

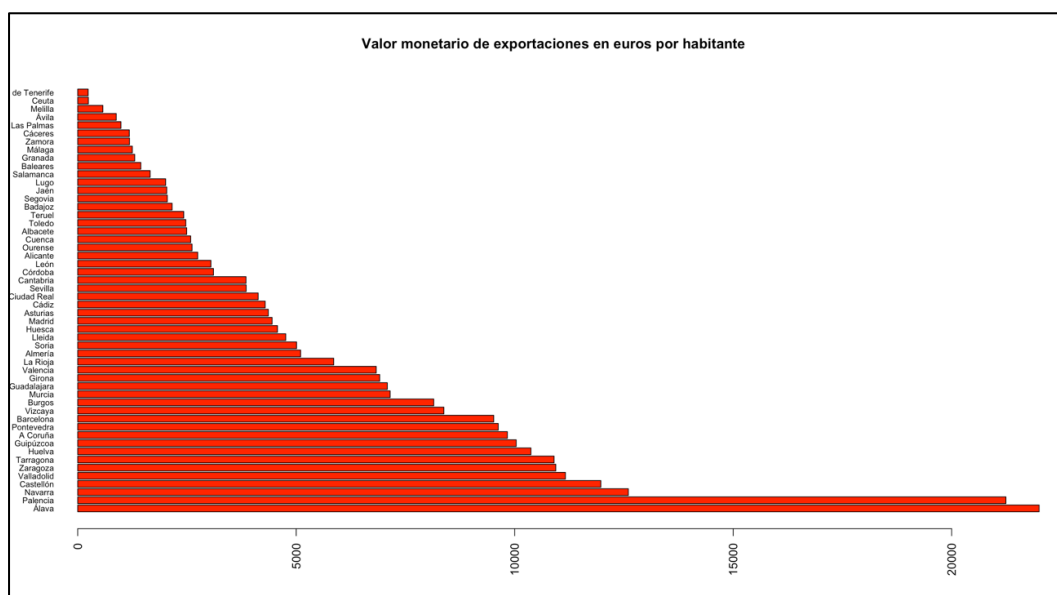


Importaciones:



Observando los mapas, se puede apreciar que provincias con una población baja en comparación con otras como puede ser Segovia, es la provincia que más exportaciones por habitante realiza de la sección 21 que trata sobre antigüedades. Otro ejemplo sería la provincia de Palencia, que es la que más exportaciones realiza de la sección 17 que trata sobre material de transporte.

Los gráficos de barras con los datos por habitante muestran los siguientes resultados:



Ilustraci n 29. Gr fico barras exportaciones por habitante

Provincias con mayor valor de exportaciones por habitante:

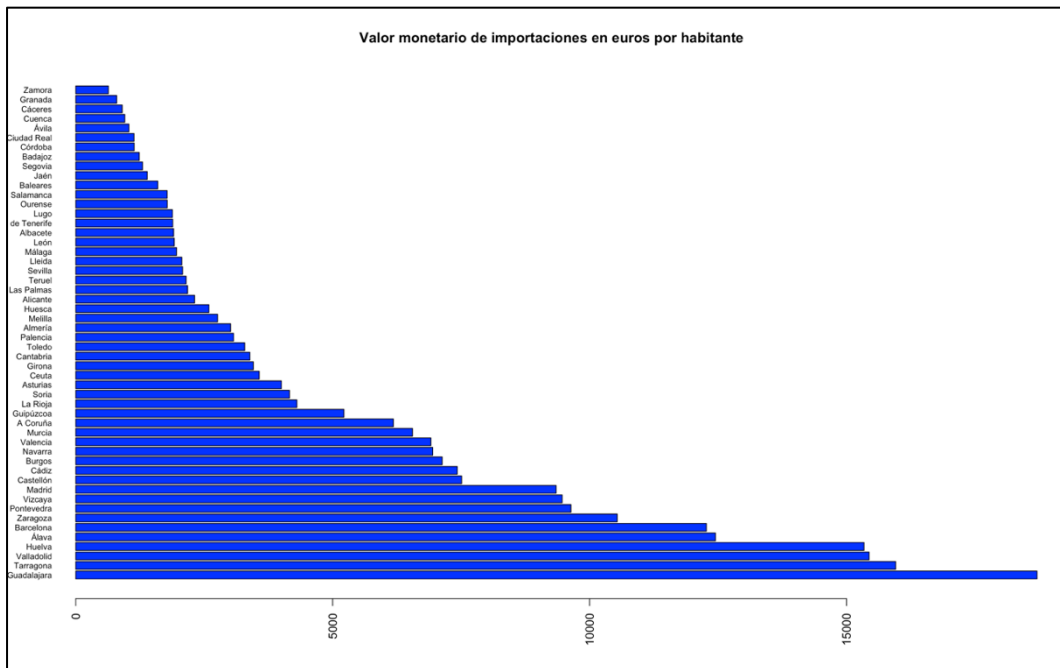
- 1)  lava
- 2) Palencia
- 3) Navarra
- 4) Castell n
- 5) Valladolid

Provincias con menor valor de exportaciones por habitante:

- 1) Santa Cruz de Tenerife
- 2) Ceuta
- 3) Melilla
- 4)  vila
- 5) Las Palmas

Ninguna de las 5 provincias que mayor valor de exportaciones por habitante presentan est n entre las provincias m s pobladas. Esto se debe principalmente a la existencia de una fuerte industria en una zona con no demasiada poblaci n.

En cuanto a las provincias con menor valor de exportaciones por habitante, se se ala su situaci n geogr fica ya que se trata de las dos ciudades aut nomas y las dos provincias canarias, por lo que la localizaci n influye.



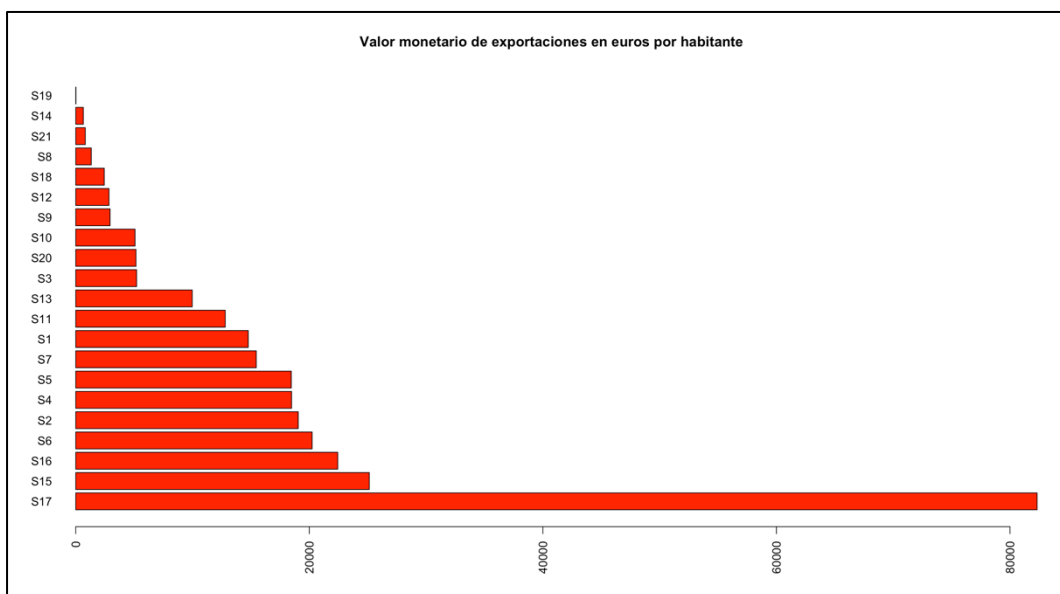
Ilustraci n 30. Gr fico barras importaciones por habitante

Provincias con mayor valor de importaciones por habitante:

- 1) Guadalajara
- 2) Tarragona
- 3) Valladolid
- 4) Huelva
- 5)  lava

Provincias con menor valor de importaciones por habitante:

- 1) Zamora
- 2) Granada
- 3) C ceres
- 4) Cuenca
- 5)  vila



Ilustraci n 31. Gr fico barras exportaciones por habitante para cada secci n

Secciones de mercancías con mayor valor de exportaciones por habitante:

- 1) Sección 17. Material de transporte.
- 2) Sección 15. Metales comunes y manufacturas de estos metales.
- 3) Sección 16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de grabación o reproducción de sonido e imágenes y sus accesorios.

Secciones de mercancías con menor valor de exportaciones por habitante:

- 1) Sección 19. Armas, municiones, y sus partes y accesorios.
- 2) Sección 14. Perlas finas (Naturales) o cultivadas, piedras preciosas o semipreciosas, metales preciosos, chapados de metal y manufacturas de estas materias. Bisutería y monedas.
- 3) Sección 21. Objetos de arte o colección y antigüedades.

La Sección 17 (material de transporte) destaca enormemente sobre las demás, se debe a la existencia de grandes fábricas automovilísticas en provincias con poca población.

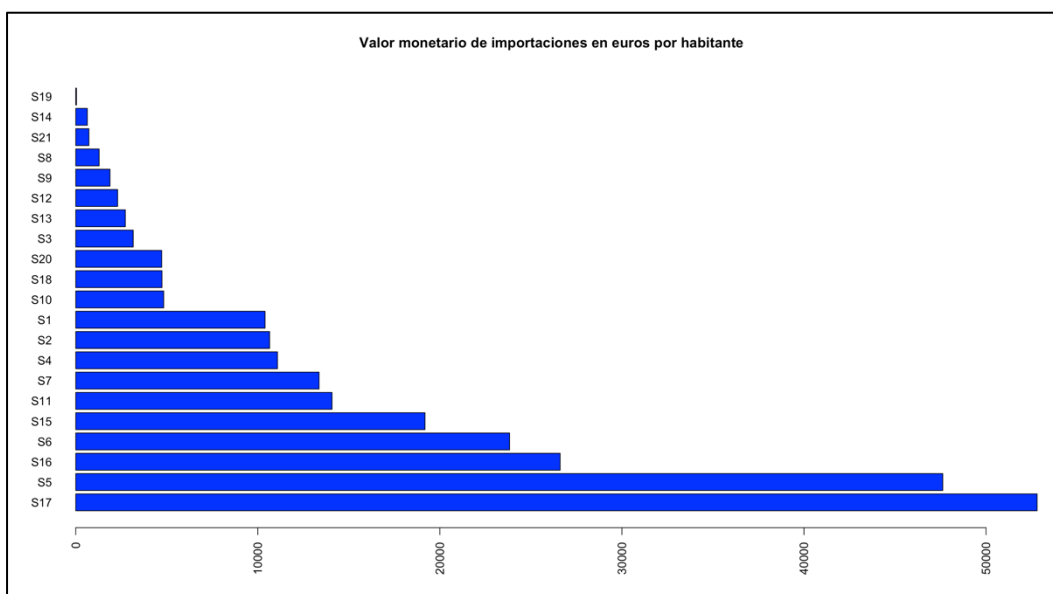


Ilustración 32. Gráfico barras importaciones por habitante para cada sección

Secciones de mercancías con mayor valor de importaciones por habitante:

- 1) Sección 17. Material de transporte.
- 2) Sección 5. Productos minerales
- 3) Sección 16. Máquinas y aparatos, material eléctrico y sus partes. Aparatos de grabación o reproducción de sonido e imágenes y sus accesorios.

Secciones de mercancías con menor valor de importaciones por habitante:

- 4) Sección 19. Armas, municiones, y sus partes y accesorios.
- 5) Sección 14. Perlas finas (Naturales) o cultivadas, piedras preciosas o semipreciosas, metales preciosos, chapados de metal y manufacturas de estas materias. Bisutería y monedas.
- 6) Sección 21. Objetos de arte o colección y antigüedades.

La sección 17 (material de transporte) aparece como la que más valor de importaciones por habitantes realiza, seguramente para obtener el material necesario para la industria automovilística.

4.3 Correlación de las variables

En esta sección se muestran las correlaciones entre las variables de los distintos conjuntos de datos agregados en los capítulos anteriores.

Para cada uno de los tres conjuntos de datos, se calcula la correlación de Pearson entre los pares de secciones de mercancías.

- Valor monetario total de exportaciones/importaciones en euros.

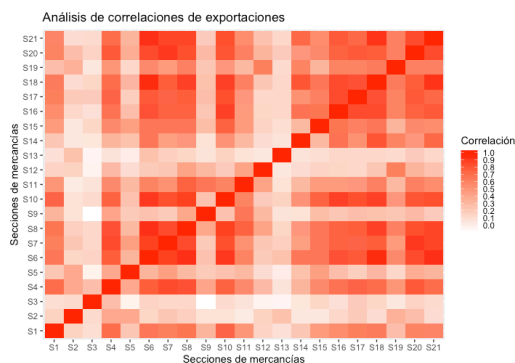


Ilustración 33. Correlaciones exportaciones

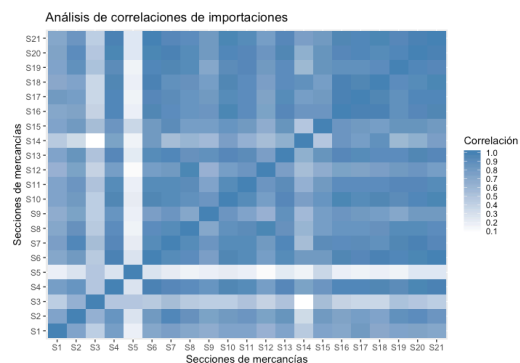


Ilustración 34. Correlaciones importaciones

Las correlaciones tanto para las exportaciones como para las importaciones son siempre positivas y pueden considerarse altas especialmente para las importaciones.

En las exportaciones, las variables que menos correlación tienen con las demás son las secciones 2 (Productos del reino vegetal), 3 (Grasas y aceites), 12 (Calzado y sombreros) y 13 (Manufacturas de piedra, yeso y cemento). Mientras que para las importaciones las variables con menor correlación con las demás son la sección 3 (Grasas y aceites) y 5 (Productos minerales).

- Valor monetario de exportaciones/importaciones en euros por habitante.

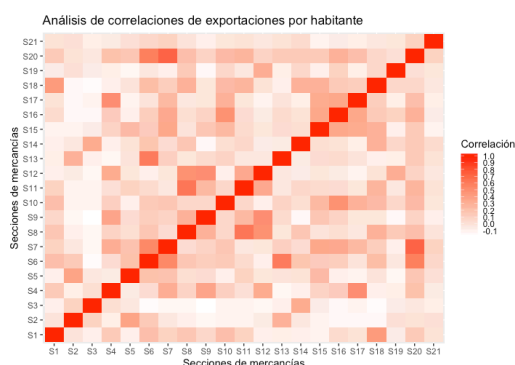


Ilustración 35. Correlación exportaciones por habitante

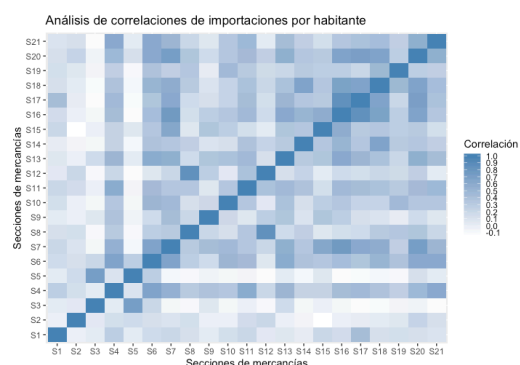


Ilustración 36. Correlación importaciones por habitante

Se observa como en comparación con los datos absolutos en euros, sin tener en cuenta la población, es decir, con el conjunto de datos visto anteriormente, los valores de las correlaciones son bastante más bajos. Incluso se puede apreciar alguna correlación negativa. Las correlaciones para las Importaciones aún siendo bajas, son ligeramente más altas que para las exportaciones.

- Porcentaje del valor de exportaciones/importaciones de cada sección de mercancías por provincia.

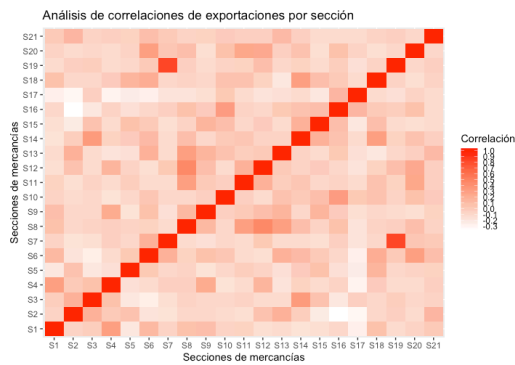


Ilustración 37. Correlaciones exportaciones sección/provincia

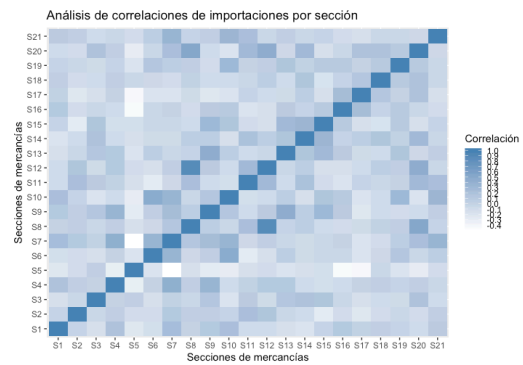


Ilustración 38. Correlaciones importaciones sección/provincia

Los resultados obtenidos con este conjunto de datos son unos valores de correlaciones entre las secciones de mercancías muy bajos. Al igual que en el caso visto justo antes, los valores de las correlaciones son ligeramente más altos para las importaciones que para las exportaciones.

5 Análisis de componentes principales

Una vez que se ha realizado el procesamiento de los datos y se tiene un conjunto de datos adecuado al objetivo de este TFG se procede a realizar el análisis multivariante. Dada la naturaleza de los datos no es posible hacer distinción entre variables explicativas y explicada. Se utilizan métodos descriptivos.

5.1 Metodología

El análisis en componentes principales es una técnica multivariante que persigue reducir la dimensión de una tabla de datos excesivamente grande por el elevado número de variables x_1, x_2, \dots, x_n . Se queda con unas pocas variables C_1, C_2, \dots, C_p combinación de las iniciales perfectamente calculables y que sintetizan la mayor parte de la información contenida en estas nuevas variables, denominadas componentes principales. Inicialmente se tienen tantas componentes como variables [12][13]:

$$\begin{aligned} C_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ &\dots \\ C_p &= a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pn}x_n \end{aligned}$$

Se retienen las p componentes principales que explican un porcentaje alto de la variabilidad de las variables iniciales (C_1, C_2, \dots, C_p). Las componentes deben de ser suficientes para resumir la mayor parte de la información contenida en las variables originales.

Asimismo, cada variable original puede expresarse en función de las componentes principales, de modo que la varianza de cada variable original se explica completamente por las componentes cuya combinación lineal la determinan.

$$\begin{aligned} x_1 &= r_{11}C_1 + r_{12}C_2 + \dots + r_{1n}C_n \\ &\dots \\ x_n &= r_{n1}C_1 + r_{n2}C_2 + \dots + r_{np}C_n \end{aligned} \quad r_{ij} = \sqrt{\lambda_i} a_{ij}$$

r_{ij} es el coeficiente de correlación entre la componente C_i y la variable x_j y se calcula multiplicando el peso a_{ij} de la variable en esa componente por la raíz cuadrada de su valor propio λ_i . Cada componente principal C_i se asocia con el valor propio i -ésimo de la matriz A que contiene los coeficientes a_{ij} .

En los siguientes apartados se mostrarán los resultados obtenidos al realizar el análisis de componentes principales con los conjuntos de datos de importaciones/exportaciones.

Para la valoración de los resultados, se considera la proporción de la varianza total explicada por una o varias componentes principales y el gráfico biplot.

Gráficos y criterios utilizados:

- Scree plot: Representa de forma ordenada los valores propios, desde el más grande hasta el más pequeño. Los valores propios de la matriz de correlación son iguales a las varianzas de los componentes principales.
- Biplot: Muestra dos gráficos simultáneamente. El primero representa las puntuaciones del segundo componente principal versus las puntuaciones del primer componente principal. El segundo gráfico proyecta los coeficientes de cada variable para el primer componente versus los coeficientes para el segundo componente.
- Regla del codo: La idea es buscar un “codo” en el scree plot, es decir, un punto a partir del cual los valores propios son aproximadamente iguales.

5.2 Resultados

5.2.1 ACP del valor monetario total de exportaciones en euros

En la Ilustración 39 se muestra el resumen obtenido al aplicar el ACP a los valores de exportaciones.

```
> summary(x_pca_E)
Importance of components:
PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10
Standard deviation  3.3051 1.28582 1.18787 1.16356 1.04279 1.01365 0.87147 0.82299 0.76277 0.63360
Proportion of Variance 0.5202 0.07873 0.06719 0.06447 0.05178 0.04893 0.03616 0.03225 0.02771 0.01912
Cumulative Proportion 0.5202 0.59891 0.66610 0.73057 0.78236 0.83128 0.86745 0.89970 0.92741 0.94652
PC11      PC12      PC13      PC14      PC15      PC16      PC17      PC18      PC19
Standard deviation  0.53576 0.48680 0.40516 0.36979 0.30701 0.26242 0.21858 0.21539 0.14857
Proportion of Variance 0.01367 0.01128 0.00782 0.00651 0.00449 0.00328 0.00228 0.00221 0.00105
Cumulative Proportion 0.96019 0.97148 0.97929 0.98581 0.99029 0.99357 0.99585 0.99806 0.99911
PC20      PC21
Standard deviation  0.11860 0.06820
Proportion of Variance 0.00067 0.00022
Cumulative Proportion 0.99978 1.00000
```

Ilustración 39. Summary ACP exportaciones

Se puede ver que la varianza explicada por la primera componente es de sólo un 52%. Las correspondientes a la segunda y tercera componente, respectivamente, son 7.8% y 6.7%. La proporción acumulada para las tres primeras componentes es de tan sólo un 66% de la varianza total. Esto supone que, al quedarse con 3 componentes, se pierde un tercio de la variabilidad. Tratándose de datos reales se puede considerar que una sola variable explique más de la mitad de los datos es un resultado aceptable.

Si se quisiera un resultado con mayor proporción de variabilidad explicada, habría que elegir las 6 primeras componentes principales que representarían un 83% de la variabilidad total. Sin embargo, la representación gráfica sería mucho más compleja.

En el siguiente gráfico (Ilustración 40), se puede apreciar lo comentado anteriormente, una primera componente que explica la mitad de la variabilidad, mientras que las demás variables aportan muy poco individualmente.

Por la regla del codo, se invitaría a pensar en la elección de las 2 primeras componentes, que representan aproximadamente un 60% de la variabilidad.

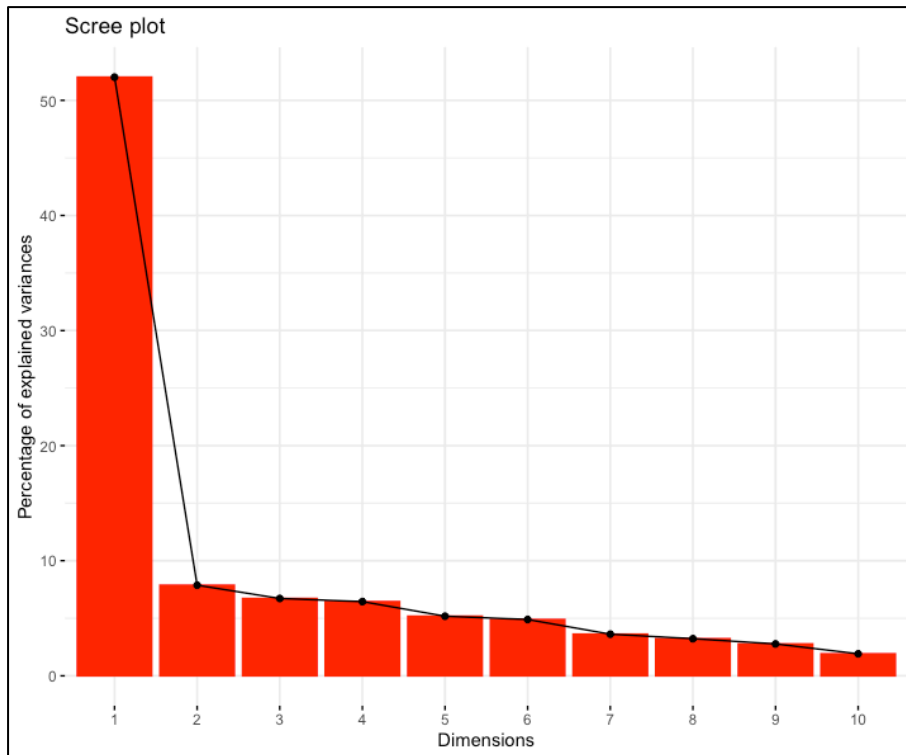


Ilustración 40. Scree plot ACP exportaciones

En el biplot que se muestra a continuación (Ilustración 41,) se puede ver como la mayoría de las provincias se encuentran en el centro del gráfico, apenas alejándose del punto (0,0). Aunque hay algunas provincias alejadas del mismo.

Barcelona es la provincia más alejada con una representación de la primera componente importante y apenas un valor muy bajo para la segunda componente. Barcelona, quedaría entonces, prácticamente representada, por la primera componente.

Madrid tiene un resultado inferior que Barcelona en cuanto a la primera componente, pero mayor para la segunda componente.

No se puede comparar de la misma forma la primera componente con la segunda, ya que la diferencia en cuanto a la variabilidad explicada es enorme (52% frente a 7.8%). Por esta razón, se aprecia como las dos provincias más importantes que son Barcelona y Madrid, son las que tienen un mayor valor para la primera componente.

También, hay una serie de provincias que se alejan del centro, como son A Coruña, Alicante, Murcia y Valencia. Estas provincias ya se ha visto con anterioridad que son las que ocupan las siguientes posiciones en cuanto a exportaciones tras Madrid y Barcelona. Se trata de cuatro provincias costeras.

En cuanto a las secciones de mercancías se puede observar como todas ellas se encuentran en la parte izquierda del gráfico, es decir, con un resultado negativo de la primera componente, que es la más importante.

Con estos análisis, se puede intuir que la primera componente está relacionada con el volumen de mercancía transportada, ya que prácticamente las provincias aparecen ordenadas de mayor a menor volumen en total de exportaciones. Para la segunda componente, podría ser la importancia con algunas secciones de mercancías como son especialmente la 9 (Madera y carbón vegetal) y la 12 (Calzado y sombreros), donde A Coruña y Alicante tienen los mejores resultados. También afectaría, aunque en menos proporción las secciones 2 (Productos del reino vegetal), 5 (Productos minerales) y 11 (Materias textiles).

Además, hay secciones de mercancías que apuntan únicamente a las provincias de Madrid como es especialmente la sección 14 (Perlas y piedras preciosas), donde Madrid supera a Barcelona.

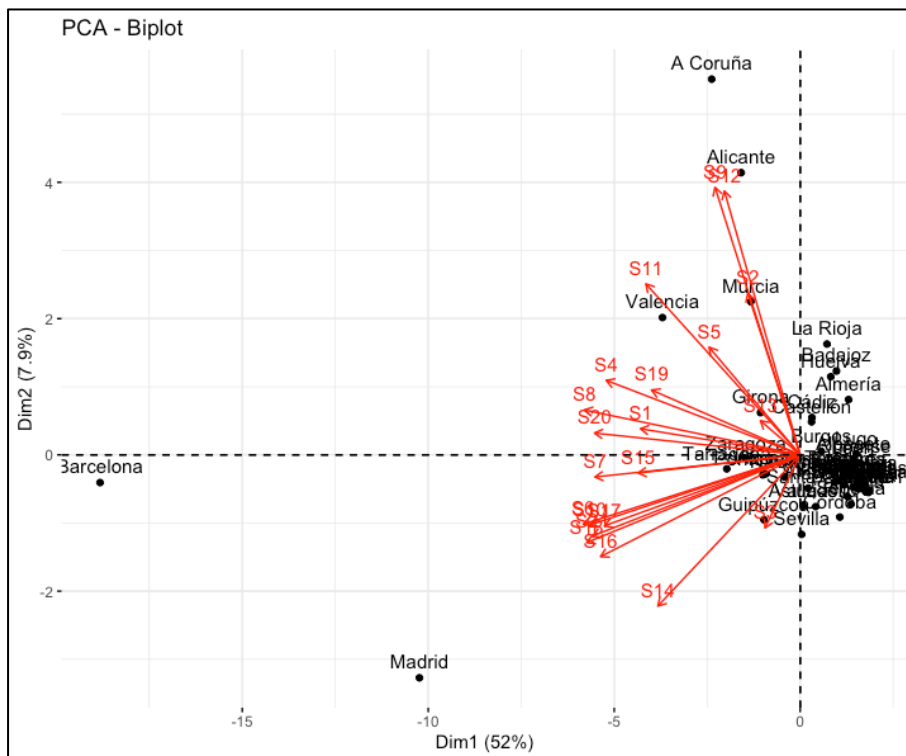


Ilustración 41. Biplot ACP exportaciones

5.2.2 ACP del valor monetario total de importaciones en euros

En la Ilustración 42 se muestra el resumen obtenido al aplicar el ACP a los valores de importaciones.

```
> summary(x_pca_I)
Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10
Standard deviation  4.0482  1.21490  0.98371  0.74525  0.63932  0.5940  0.5324  0.42649  0.31871  0.30146
Proportion of Variance  0.7804  0.07029  0.04608  0.02645  0.01946  0.0168  0.0135  0.00866  0.00484  0.00433
Cumulative Proportion  0.7804  0.85068  0.89676  0.92321  0.94267  0.9595  0.9730  0.98163  0.98647  0.99079
      PC11      PC12      PC13      PC14      PC15      PC16      PC17      PC18      PC19      PC20
Standard deviation  0.22274  0.20748  0.17949  0.1651  0.10917  0.10037  0.09273  0.06220  0.05476  0.04371
Proportion of Variance  0.00236  0.00205  0.00153  0.0013  0.00057  0.00048  0.00041  0.00018  0.00014  0.00009
Cumulative Proportion  0.99316  0.99521  0.99674  0.9980  0.99861  0.99909  0.99950  0.99968  0.99982  0.99991
      PC21
Standard deviation  0.04237
Proportion of Variance  0.00009
Cumulative Proportion  1.00000
```

Ilustración 42. Summary ACP importaciones

Se puede ver que la varianza explicada por la primera componente es de un 78%. Los valores correspondientes a la segunda y tercera componente, respectivamente, son 7% y 4.6%. La proporción acumulada para las dos primeras componentes es de un 85% de la varianza total.

A diferencia del resultado obtenido en el análisis para las exportaciones, ahora la proporción de la variabilidad explicada es un resultado mucho mayor, ya que la primera componente ha pasado de explicar la mitad de los datos a explicar más de tres cuartas partes.

Con las dos primeras componentes se tiene prácticamente la totalidad de la información explicada, mientras que, para obtener la misma información para las exportaciones, se necesitaban 5 componentes.

En la ilustración 43, se puede apreciar lo comentado anteriormente, una primera componente que explica la mayor parte de la variabilidad, mientras que las demás aportan muy poco.

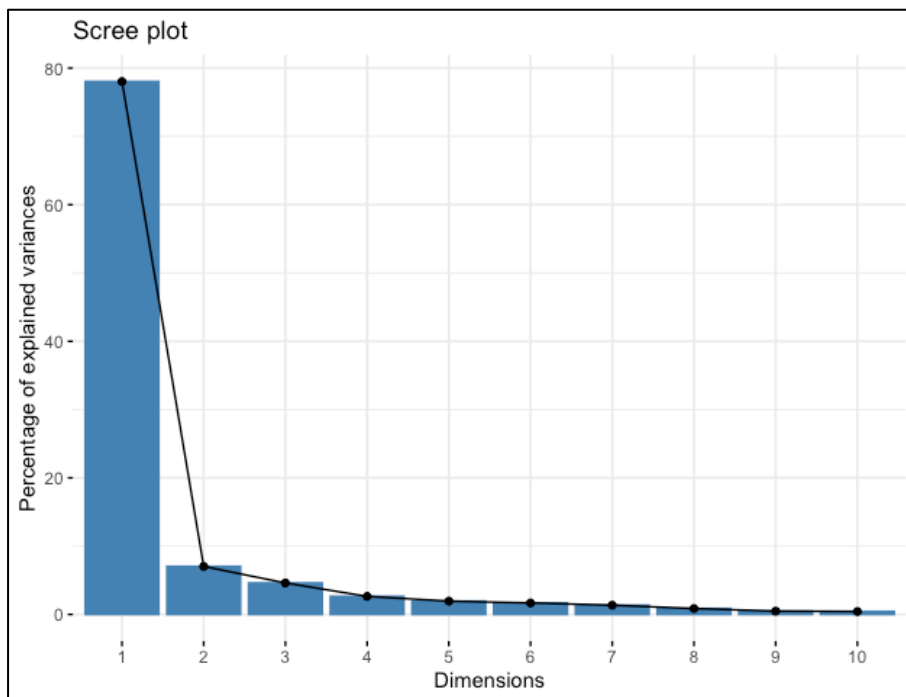


Ilustración 43. Scree plot ACP importaciones

Por la regla del codo, se invitaría a pensar en la elección de las 2 primeras componentes. Viendo que la proporción explicada es un resultado bastante aceptable con un 85% parece la elección más idónea.

En cuanto a la representación de las provincias se puede apreciar como la mayor parte de las provincias se encuentran en el centro del gráfico, apenas alejándose del punto (0,0). Aunque sí hay algunas provincias que sí que se alejan del centro del gráfico.

Barcelona es la provincia más alejada con una representación de la primera componente importante y un valor bajo de la segunda componente. Barcelona, quedaría prácticamente representada entonces, por la primera componente. Madrid tiene un resultado inferior que Barcelona en cuanto a la primera componente, pero mayor para la segunda componente.

No se puede comparar de la misma forma la primera componente con la segunda, ya que la diferencia en cuanto a la variabilidad explicada es enorme (78% frente a 7%). Por esta razón,

se aprecia como las dos provincias más importantes que son Barcelona y Madrid, son las que tienen un mayor valor para la primera componente.

También, hay una serie de provincias que se alejan ligeramente del centro como son Huelva, Murcia, Cádiz, A Coruña, Castellón, Sevilla, Tarragona y Valencia. De estas provincias, únicamente Valencia y Tarragona tienen unos valores ligeramente altos para la primera componente.

Se aprecia como todas las secciones de mercancías se encuentran en la parte izquierda del gráfico, es decir, con un resultado negativo de la primera componente, que es la más importante.

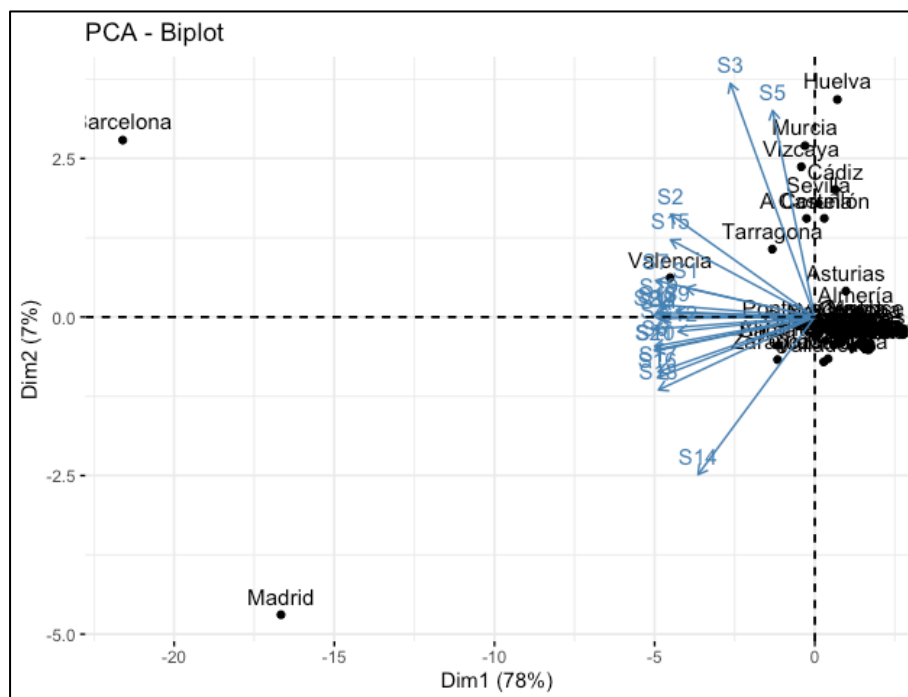


Ilustración 44. Biplot ACP importaciones

En el gráfico (Ilustración 44) conjunto de las provincias y las secciones de mercancías se puede ver como la mayoría de las secciones de mercancías apuntan hacia la primera componente, siendo solamente unas pocas las que se ven influenciadas por la segunda componente como son principalmente las secciones 3 (Grasas y aceites),5 (Productos minerales) y 14 (Perlas y piedras preciosas).

Una vez visto el análisis de componentes principales tanto para las exportaciones como para las importaciones, se puede llegar a la conclusión de que la primera componente es la que indica el volumen de mercancía transportada, ya que destacan las provincias que mayores valores tienen como son Barcelona, Madrid y Valencia. Para la segunda componente, estaría relacionada con alguna sección en especial, como se ha comentado en los gráficos por secciones. Destacando para las exportaciones la secciones 9 (Madera y carbón vegetal),12 (Calzado y sombreros) y 14 (Perlas y piedras preciosas), mientras que, para las importaciones, las secciones 3 (Grasas y aceites),5 (Productos minerales) y 14 (Perlas y piedras preciosas).

5.2.3 ACP del Valor monetario de exportaciones en euros por habitante

En la Ilustración 45 se muestra el resumen obtenido al aplicar el ACP a los valores de exportaciones por habitantes.

```
> summary(x_pca_Ehab)
Importance of components:
      PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8    PC9    PC10
Standard deviation  1.8497 1.5319 1.43332 1.32764 1.18133 1.14081 1.13064 1.0609 1.02189 0.93818
Proportion of Variance 0.1629 0.1118 0.09783 0.08393 0.06645 0.06197 0.06087 0.0536 0.04973 0.04191
Cumulative Proportion 0.1629 0.2747 0.37251 0.45644 0.52289 0.58487 0.64574 0.6993 0.74907 0.79098
      PC11   PC12   PC13   PC14   PC15   PC16   PC17   PC18   PC19
Standard deviation  0.89896 0.89470 0.74561 0.69658 0.61786 0.58570 0.54899 0.49448 0.47091
Proportion of Variance 0.03848 0.03812 0.02647 0.02311 0.01818 0.01634 0.01435 0.01164 0.01056
Cumulative Proportion 0.82946 0.86758 0.89406 0.91716 0.93534 0.95168 0.96603 0.97767 0.98823
      PC20   PC21
Standard deviation  0.39766 0.29838
Proportion of Variance 0.00753 0.00424
Cumulative Proportion 0.99576 1.00000
```

Ilustración 45. Summary ACP exportaciones por habitante

Al ser los resultados de las correlaciones de las variables relativas a exportaciones por habitante, mostradas en la sección 4.3 de este TFG, bastante bajas no se podía esperar un buen resultado con el análisis en componentes principales.

La proporción de la varianza explicada por la primera componente es de tan sólo un 16%, muy lejos del 52% y del 78% que se habían obtenido en los casos previos.

Contando las dos primeras componentes, tan sólo representan un 27%. Para obtener al menos la mitad de la variabilidad explicada se necesitarían seleccionar las 5 primeras componentes. Por tanto, se puede concluir que el análisis no reporta un buen resultado.

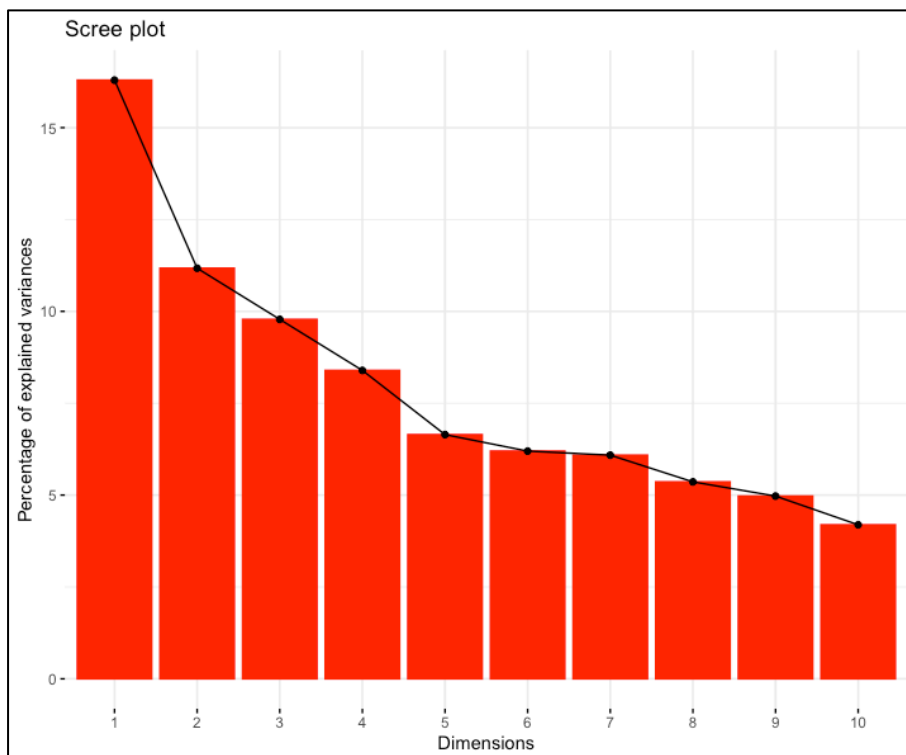


Ilustración 46. Scree plot ACP exportaciones por habitante

En el scree plot (Ilustración 46) se observa como no hay una gran diferencia entre la primera componente con respecto a las demás. Se puede ver como la variabilidad explicada por cada una de las componentes es bastante similar, el problema de este análisis es que ninguna de ellas tiene un resultado importante que destaque.

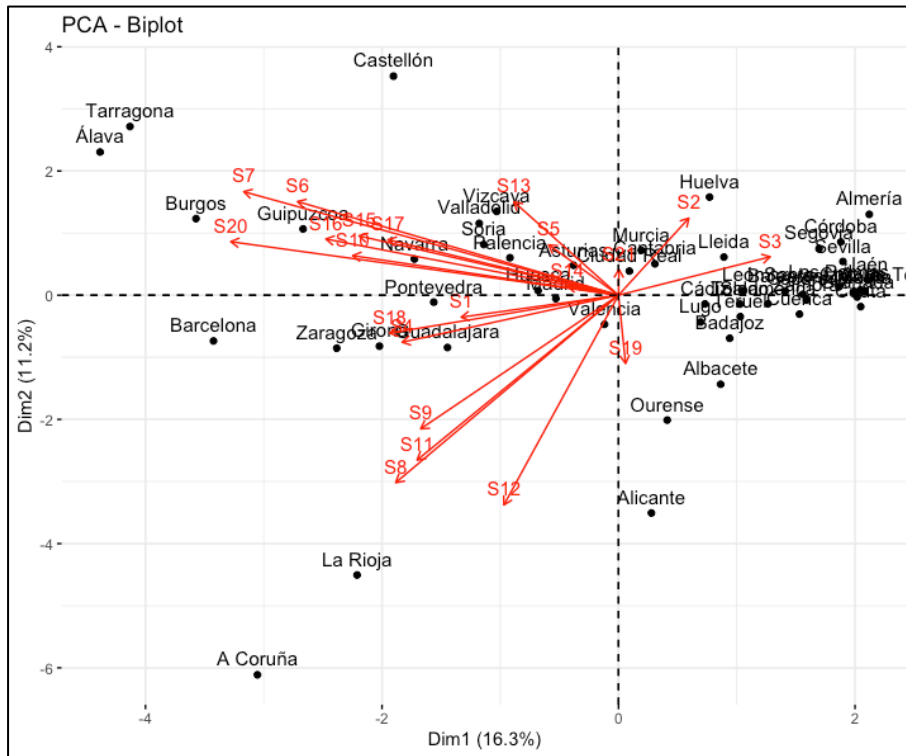


Ilustración 47. Biplot ACP exportaciones por habitante

En el biplot (Ilustración 47) se aprecia como ya aparecen las provincias más distribuidas por el gráfico, solo quedando un grupo con una representación positiva de la primera componente. También se observa como no todas las secciones se encuentran en un lado o en otro, hay tres secciones que apuntan hacia la derecha como son las secciones 2 (Productos del reino vegetal), 3 (Grasas y aceites) y 19 (Armas y municiones). No obstante, esta interpretación no tiene gran interés por la escasa varianza explicada.

5.2.4 ACP del valor monetario de exportaciones en euros por habitante

En la Ilustración 48 se muestra el resumen obtenido al aplicar el ACP a los valores de importaciones por habitantes.

```
> summary(x_pca_Ihab)
Importance of components:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9      PC10
Standard deviation  2.6757  1.44249  1.36723  1.21139  1.1281  1.09744  1.06111  0.95505  0.85306  0.81556
Proportion of Variance 0.3409  0.09909  0.08901  0.06988  0.0606  0.05735  0.05362  0.04343  0.03465  0.03167
Cumulative Proportion 0.3409  0.44000  0.52902  0.59890  0.6595  0.71685  0.77047  0.81390  0.84855  0.88023
      PC11      PC12      PC13      PC14      PC15      PC16      PC17      PC18      PC19      PC20
Standard deviation  0.72334  0.65556  0.61469  0.54628  0.51759  0.44534  0.3917  0.29848  0.29311  0.24379
Proportion of Variance 0.02491  0.02046  0.01799  0.01421  0.01276  0.00944  0.0073  0.00424  0.00409  0.00283
Cumulative Proportion 0.90514  0.92561  0.94360  0.95781  0.97057  0.98001  0.9873  0.99156  0.99565  0.99848
      PC21
Standard deviation  0.17864
Proportion of Variance 0.00152
Cumulative Proportion 1.00000
```

Ilustración 48. Summary ACP importaciones por habitante

La proporción de la primera componente es del 34%, lo que resulta un mejor resultado que con las exportaciones por habitante, sin embargo, sigue siendo un valor bastante bajo. Con las dos primeras componentes, se explica aproximadamente un 45% de la información. Por tanto, al igual que con las exportaciones se puede concluir que el resultado de este análisis es poco útil.

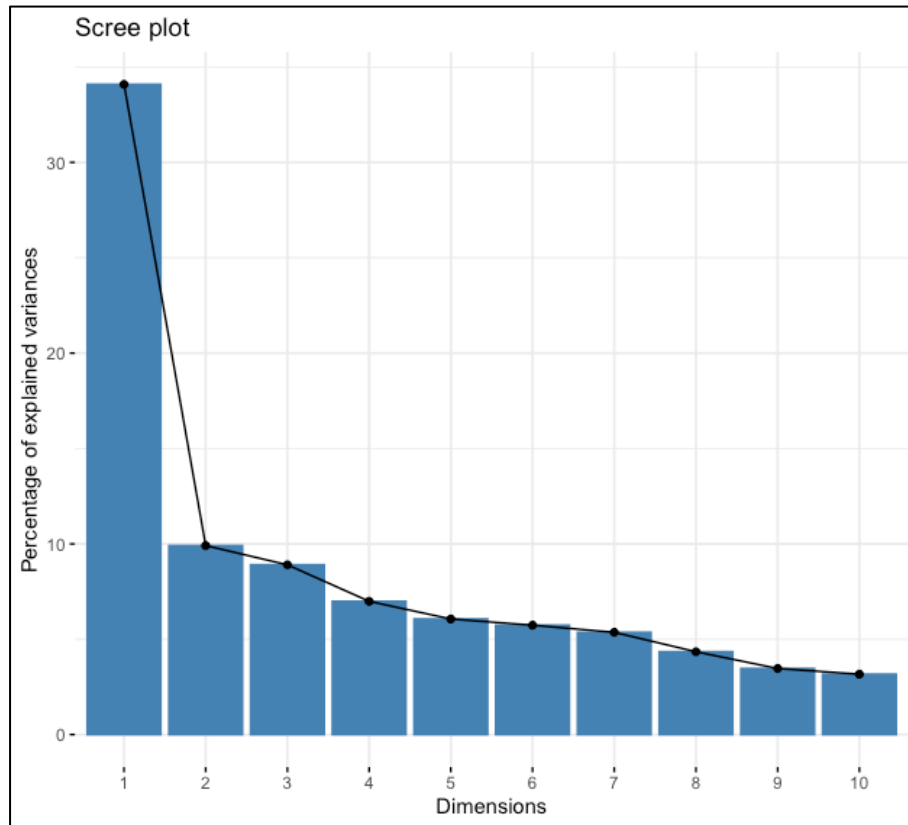


Ilustración 49. Scree plot ACP importaciones por habitante

En el scree plot (Ilustración 49), a diferencia de las exportaciones, sí que destaca la primera componente sobre las demás, aunque no lo suficiente para que este análisis sea de gran utilidad.

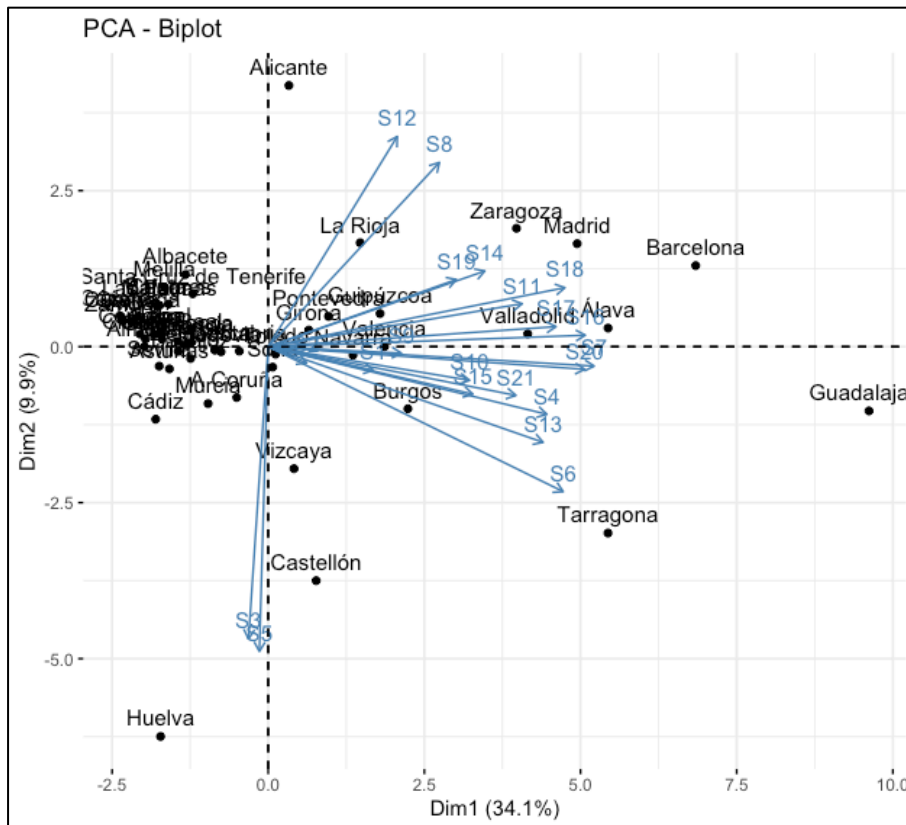


Ilustración 50. Biplot ACP importaciones por habitante

En el biplot (Ilustración 50) se aprecia un grupo de provincias agrupadas cerca del punto (0,0), mientras que hay bastantes provincias fuera de ese grupo, destacando por ejemplo Guadalajara en la primera componente o Alicante y Huelva para la segunda.

6 Clasificación de las provincias

En este capítulo se aborda el principal objetivo del presente TFG que consiste en la clasificación de las provincias en base a la estructura de su comercio exterior. Para lograr tal fin, se utilizará el análisis clúster. En primer lugar, se explicará la metodología que se va a utilizar y por último los resultados obtenidos para cada uno de los conjuntos de datos agregados que se describen en el capítulo 3 de esta memoria.

6.1 Metodología

El análisis clúster o de conglomerados es una técnica estadística multivariante de clasificación automática. A partir de una tabla de casos-variables se trata de situar todos los casos en grupos homogéneos no conocidos de antemano, pero sugeridos por la propia esencia de los datos, de manera que individuos que puedan ser considerados similares sean asignados a un mismo clúster, mientras que individuos diferentes se sitúen en clústeres distintos.

La creación de conglomerados basados en similitud de casos exige una definición de similitud o de su complementario (distancia entre individuos). Existen muchas formas de medir estas distancias y diferentes reglas matemáticas para asignar los individuos a distintos grupos, dependiendo del fenómeno estudiado y del conocimiento previo.

Los métodos de clúster se pueden dividir dos tipos: Jerárquico y no jerárquico.

- Clúster no jerárquico:

El algoritmo de k-means es un método de agrupamiento que tiene como objetivo la partición de un conjunto de n observaciones en k grupos en el que cada observación pertenece al grupo cuyo valor medio es más cercano.

- Clúster jerárquico:

Los clústeres se crean de manera que tengan un orden predeterminado, es decir, una jerarquía. Esta jerarquía de clústeres puede crearse de arriba a abajo o viceversa. Existen a su vez dos tipos: Divisivo y Aglomerativo. Para medir la distancia o similitud entre dos grupos de observaciones, en este TFG se han utilizado los siguientes métodos descritos a continuación:

- Simple: Se considera la distancia mínima entre pares.

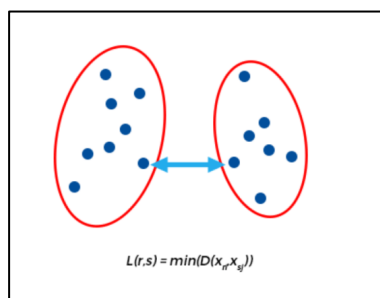


Ilustración 51. Clúster jerárquico método simple

- Completo: Se utiliza la distancia del par más lejano.

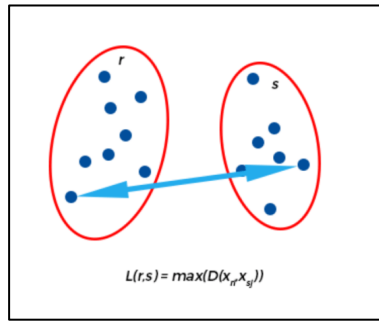


Ilustración 52. Clúster jerárquico método completo

- Media: Se utiliza la distancia promedio entre miembros

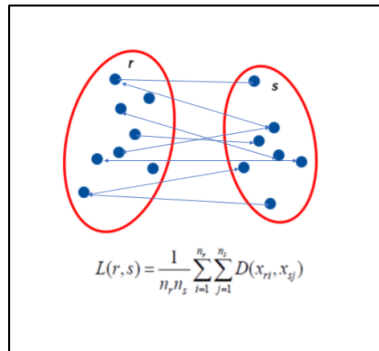


Ilustración 53. Clúster jerárquico método promedio

- Ward: Minimiza la varianza total dentro del grupo. Se fusionan en cada paso el par de clústeres con distancia mínima entre ellos.

En este TFG se ha realizado una comparación entre los cuatro métodos aglomerativos y el divisivo. Para ello se ha obtenido el coeficiente de aglomeración o de división de cada uno de ellos. El coeficiente se calcula de la siguiente forma:

Si para cada observación “i” denotamos “m(i)” al cociente entre la similaridad con el primer clúster al que es añadido y la similaridad del último agrupamiento, el coeficiente de agrupamiento es la media de todos los valores $1 - m(i)$. Los valores de este coeficiente, cuanto más cercanos a 1 sugieren una estructura de agrupación más fuerte.

6.2 Resultados

En esta sección se muestran los resultados de aplicar técnicas clúster jerárquicas y no jerárquicas sobre los conjuntos de datos de importaciones y exportaciones de las provincias de España. [14]

6.2.1 Análisis clúster del valor monetario total de exportaciones en euros

6.2.1.1 Análisis clúster jerárquico

A partir de los datos de la matriz de exportaciones se han calculado los coeficientes de aglomeración y divisivo de cada método para determinar qué técnica brinda el mejor resultado.

Tabla 10. Coeficientes análisis clúster exportaciones

Método	Average	Single	Complete	Ward	Divisive
Coeficiente	0.8374915	0.8182931	0.8653282	0.9059913	0.8548867

En la Tabla 10 se muestra el coeficiente de aglomeración para los distintos procedimientos. En esta tabla se muestra que el método que tiene el coeficiente de aglomeración más alto es el método Ward. A continuación, se muestran los resultados de aplicar este método.

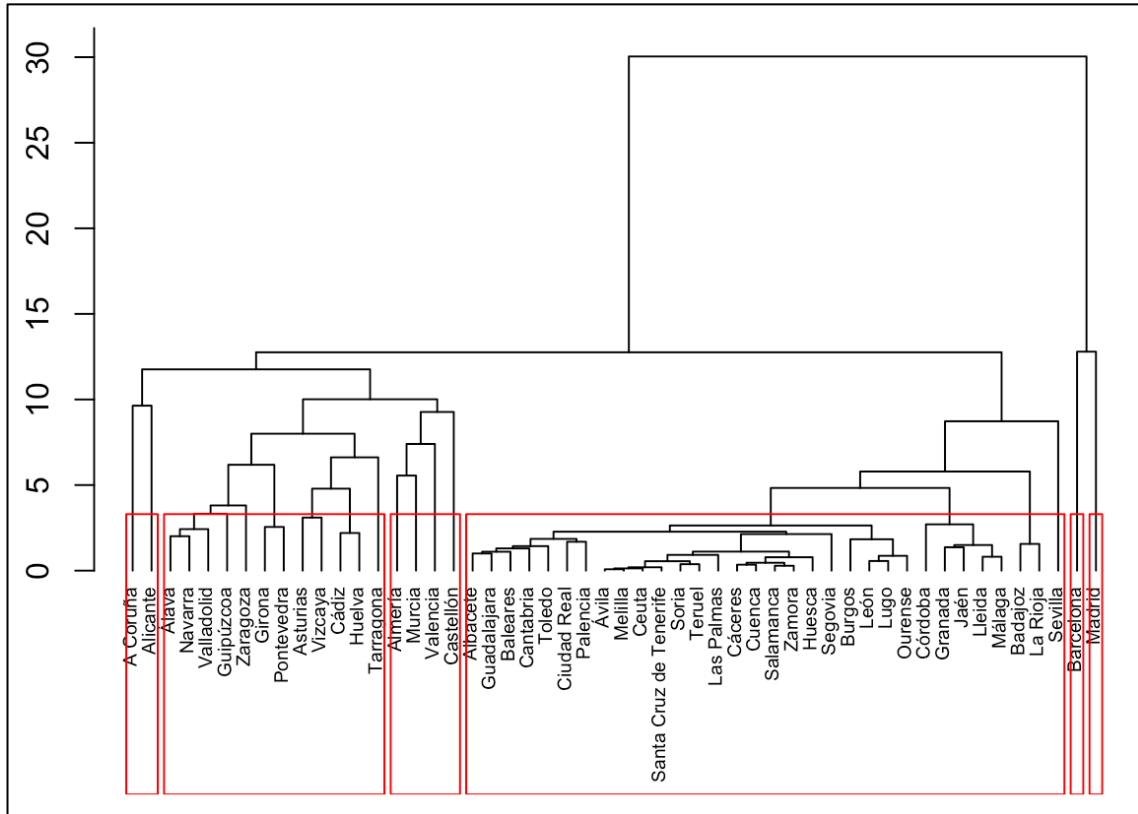


Ilustración 54. Dendrograma Ward exportaciones

Se muestra en el biplot (Ilustración 55) y en un mapa de España por provincias (Ilustración 56), cada uno de los grupos creados, con distintos colores.

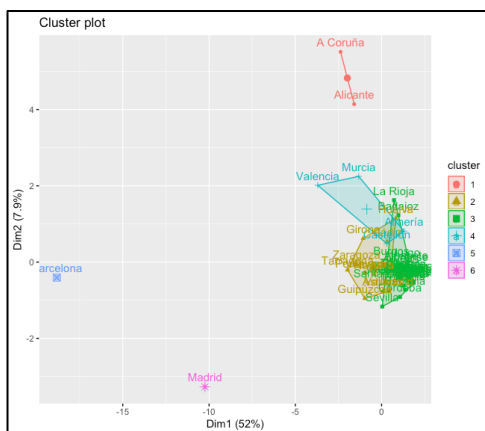


Ilustración 55. Grupos Ward exportaciones

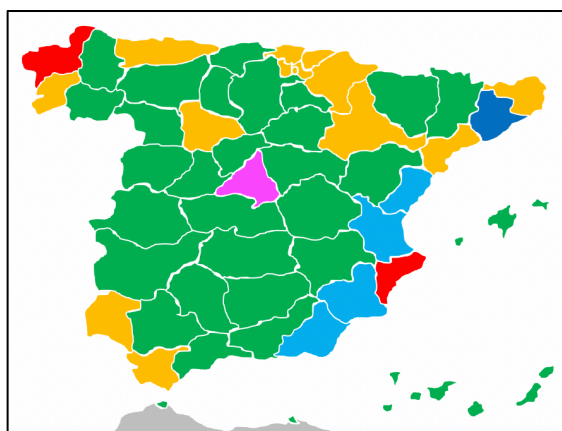


Ilustración 56. Mapa Ward exportaciones

- En el dendograma (Ilustración 54) se ha dividido el conjunto de las provincias en 6 grupos realizando el análisis clúster con estos mismos.
- Las provincias de Barcelona y Madrid se encuentran muy diferenciadas de las demás, tanto en el gráfico de ACP como en el dendograma. Se debe al volumen de transporte de mercancías, muy superior al del resto de provincias.
- Existe más diferencia entre Madrid y Barcelona, que entre cualquiera de las demás provincias. Debido a esto, aparecen cada una en un grupo único.
- Se observa como hay un grupo (Verde) muy numeroso formado por más de la mitad de las provincias. En el mapa (Ilustración 56) se aprecia como este grupo está formado principalmente por provincias de interior. Hay algunas excepciones como Valladolid por ejemplo que pertenece al grupo amarillo.
- Además, en el mapa se ve como las provincias costeras se encuentran en otros grupos. La Coruña y Alicante aparecen en un mismo grupo (Rojo), ya que son dos provincias que destacan en las mismas secciones con un volumen de mercancía muy similar.
- También se puede apreciar como prácticamente las provincias costeras mediterráneas están en otro grupo (Azul claro).
- En otro grupo (Amarillo) se encuentran agrupadas varias provincias del norte de España junto con Huelva y Cádiz.

6.2.1.2 Análisis clúster no jerárquico

A partir de los datos de la matriz de exportaciones se ha aplicado el método de k-means. A continuación, se muestran los resultados de aplicar este método en el biplot (Ilustración 57) y en un mapa de España por provincias (Ilustración 58), cada uno de los grupos creados, con distintos colores.

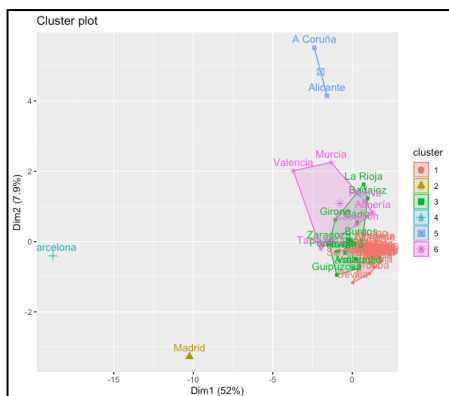


Ilustración 57. Grupos k-means exportaciones

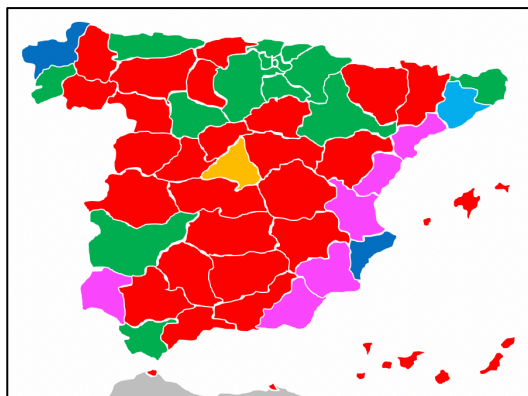


Ilustración 58. Mapa k-means exportaciones

En el siguiente gráfico (Ilustración 59) se representa para cada sección, el valor de su respectivo centroide, de cada uno de los grupos formados, indicados con su color respectivo.

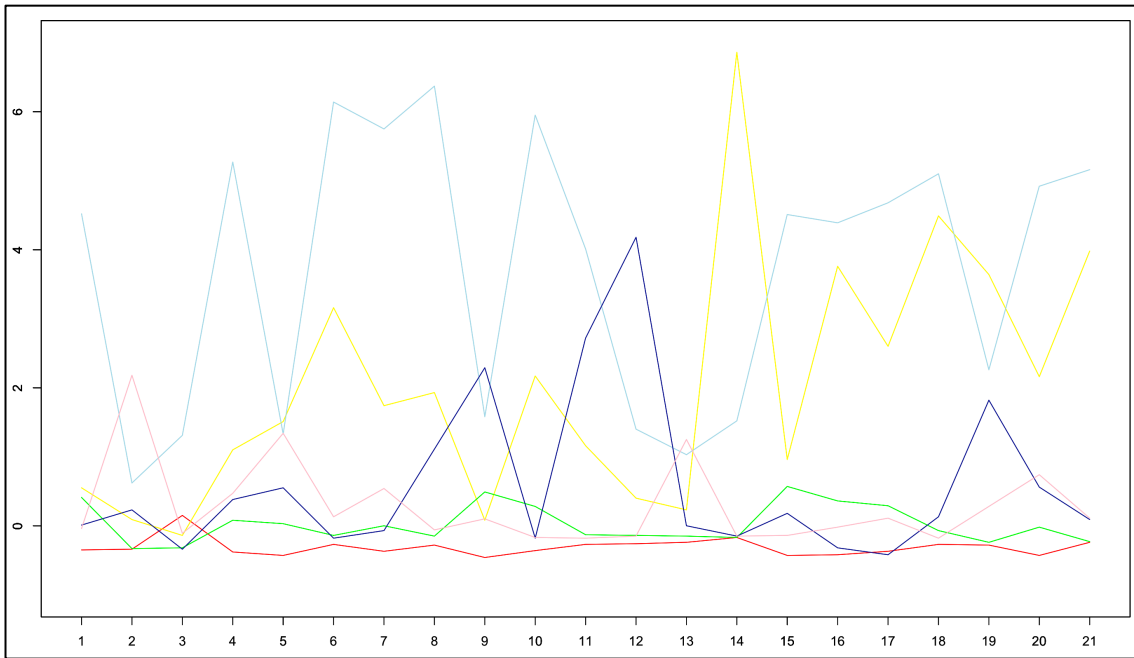


Ilustración 59. Centroides k-means exportaciones

En comparación con el análisis clúster jerárquico de la sección anterior se destacan los siguientes resultados:

- La Coruña y Alicante también forman un mismo grupo (Azul oscuro).
- Madrid y Barcelona siguen en un grupo cada una.
- Existe un grupo (Rosa) con las provincias mediterráneas, a las que se une la provincia de Huelva.
- Aparece un grupo (Verde) con las provincias del norte de la península junto con Cádiz y Badajoz.
- Se mantiene un grupo (Rojo) muy numeroso con la mayoría de las provincias de interior, las islas, Ceuta y Melilla.

6.2.1.3 Conclusiones de todos los procedimientos clúster para el valor monetario total de exportaciones en euros

- Con el análisis clúster jerárquico y no jerárquico se puede interpretar que la división de grupos se realiza principalmente atendiendo a la representación basada en el estudio del ACP, donde tiene una gran importancia la primera componente principal que está relacionada con la cantidad de exportaciones.
- El número óptimo de clústeres para el análisis de las exportaciones totales es de 2, en el que Barcelona y Madrid se encuentran en un grupo, mientras que todas las demás provincias están en el otro. Sin embargo, se ha representado una división mayor con 6 grupos.
- Las dos provincias que más exportaciones realizan (Barcelona y Madrid) están cada una en un grupo diferente.

- En un término medio de exportaciones, se encuentra un grupo de provincias que se corresponden principalmente con las provincias mediterráneas como Valencia, Castellón o Murcia, que se encuentran muy igualadas entre ellas, en cuanto al número de exportaciones.
- Además de los grupos anteriores tenemos un grupo formado principalmente por provincias del norte de España, que se encontraría en un tercer escalón en cuanto al número de exportaciones.
- Por último, las provincias que menos exportan se encuentran todas ellas en un grupo.
- También hay que tener en cuenta, aunque en menor medida, la segunda componente principal del ACP, que ayuda a la creación de un grupo formado únicamente por La Coruña y Alicante. La segunda componente principal ya no se basa en la cantidad de exportaciones realizadas sino en la influencia de estas provincias en algunas secciones.
- El gráfico con los centroides (Ilustración 59), indica qué secciones tienen más influencia en cada uno de los clústeres. Se puede apreciar como para cada grupo, las secciones que más influencia tienen en él son las siguientes:
 - Barcelona (Grupo Azul): Aparece en primera posición de 11 secciones, teniendo una influencia muy alta prácticamente en todas las secciones.
 - Madrid (Grupo Amarillo): Similar a Barcelona, aunque por debajo de ella. Destaca en la sección 14 de “Perlas finas”.
 - Alicante y La Coruña (Grupo Azul oscuro): Tienen la primera posición en la sección 9 de “Calzados y sombreros” y en la sección 12 de “Manufacturas de Madera y Carbón Vegetal”.
 - Provincias Mediterráneas (Grupo Rosa): Destacan en la sección 2 de “Productos del reino vegetal” y en la sección 13 de “Manufacturas de piedra”.
 - Para los grupos más numerosos en cuanto al número de provincias, que son el verde y el rojo, se puede ver como no destaca ninguna sección, no apareciendo ningún pico representativo en el gráfico de centroides.

6.2.2 Análisis clúster del valor monetario total de importaciones en euros

6.2.2.1 Análisis clúster jerárquico

A partir de los datos de la matriz de importaciones se han calculado los coeficientes de aglomeración y divisivo de cada método para determinar qué técnica brinda el mejor resultado.

Tabla 11. Coeficientes análisis clúster Importaciones

Método	Average	Single	Complete	Ward	Divisive
Coeficiente	0.9155577	0.8921372	0.9225159	0.9521023	0.9202538

En la Tabla 11 se muestra el coeficiente de aglomeración para los distintos procedimientos. En esta tabla se muestra que el método que tiene el coeficiente de aglomeración más alto es el método Ward. A continuación, se muestran los resultados de aplicar este método.

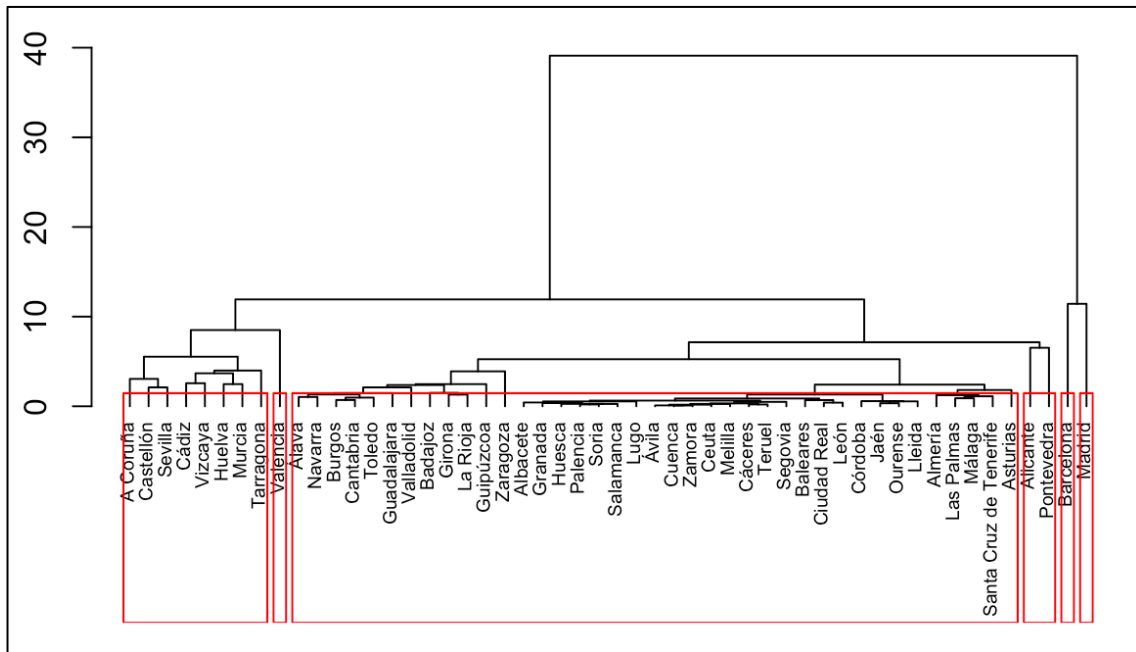


Ilustración 60. Dendrograma Ward importaciones

Se muestra en el biplot (Ilustración 61) y en un mapa de España por provincias (Ilustración 62), cada uno de los grupos creados, con distintos colores.

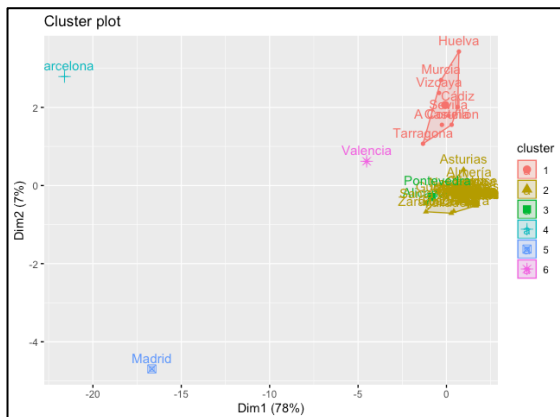


Ilustración 61. Grupos Ward importaciones

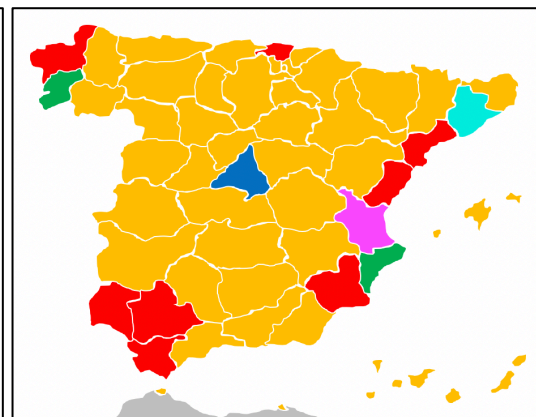


Ilustración 62. Mapa Ward importaciones

- En el dendrograma (Ilustración 60) se ha dividido el conjunto de las provincias en 6 grupos realizando el análisis clúster con estos mismos.
- Respecto al análisis de las exportaciones totales realizado en la sección anterior, con las importaciones totales se evidencia una mayor similitud entre Madrid y Barcelona, aunque siguen estando muy diferenciadas de las demás provincias, como ya se había visto con el análisis descriptivo.

- La provincia de Valencia aparece, al igual que Madrid y Barcelona en un único grupo, diferenciándose de las demás provincias y resultando ser la tercera provincia que más importaciones realiza.
- Además, se encuentran agrupadas en un grupo (Rojo) provincias que en su mayoría son costeras que en cuanto al volumen de importaciones se encuentran por detrás de la provincia de Valencia.
- Las provincias de Pontevedra y Alicante forman un grupo (Verde), diferenciándose del grupo más numeroso, en cuanto al número de provincias.
- Por último, prácticamente dos tercios de las provincias se encuentran en único grupo (Naranja) donde aparecen todas las provincias de interior menos Madrid junto con una gran representación de las provincias costeras.

6.2.2.2 No jerárquico

A partir de los datos de la matriz de importaciones se ha aplicado el método de k-means. A continuación, se muestran los resultados de aplicar este método en el biplot (Ilustración 63) y en un mapa de España por provincias (Ilustración 64), cada uno de los grupos creados, con distintos colores.

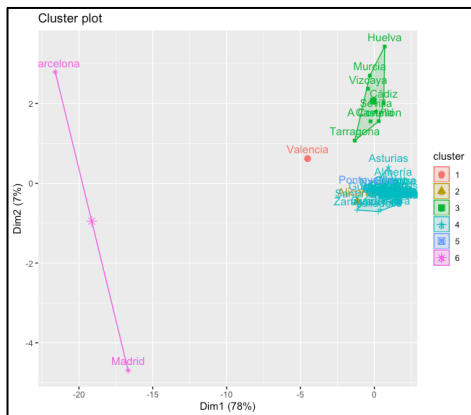


Ilustración 63. Grupos k-means importaciones

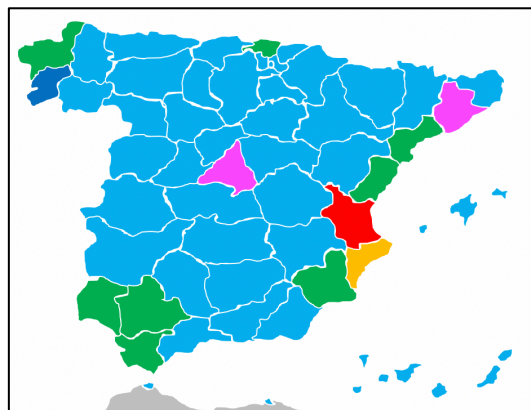


Ilustración 64. Mapa k-means importaciones

En el siguiente gráfico (Ilustración 65) se representa para cada sección, el valor de su respectivo centroide, de cada uno de los grupos formados, indicados con su color respectivo.

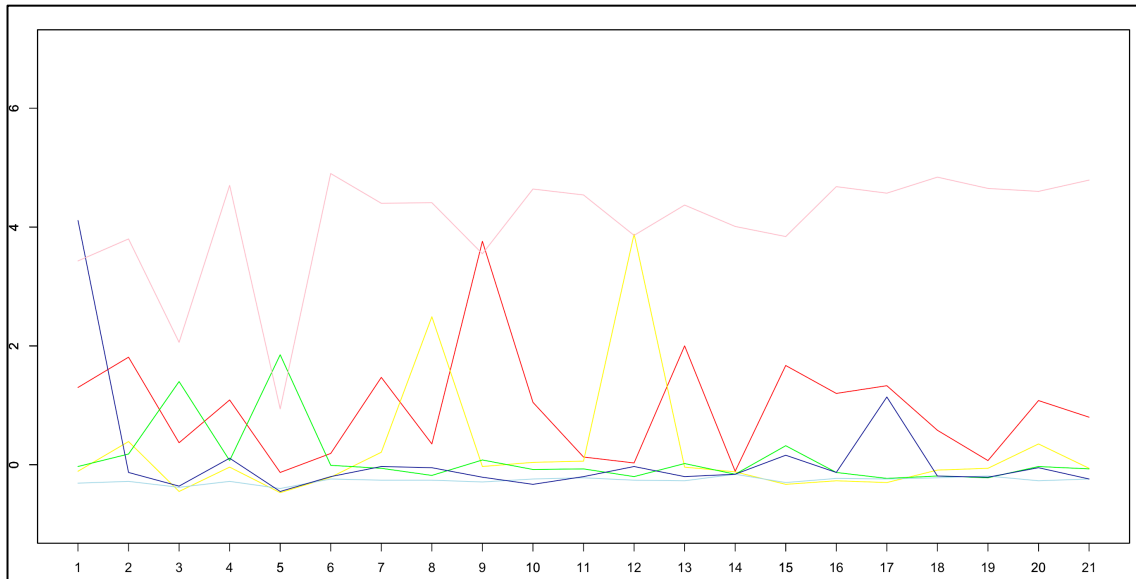


Ilustración 65. Centroides k-means importaciones

En comparación con el análisis clúster jerárquico de la sección anterior se destacan los siguientes resultados:

- Barcelona y Madrid aparecen en un único grupo.
- Pontevedra aparece separada de Alicante, formando un grupo cada una.

6.2.2.3 Conclusiones de todos los procedimientos clúster para el valor monetario total de exportaciones en euros

- En el análisis clúster jerárquico y en el no jerárquico se puede apreciar una interpretación bastante similar a la de las exportaciones totales. Es decir, grupos formados en función de la primera componente principal relacionado con la cantidad y el volumen de importaciones realizadas, con Barcelona y Madrid en grupos alejados de los demás.
- Existe una mayor similitud entre Madrid y Barcelona, como se vio en el análisis descriptivo. Aparecen más próximas en el gráfico del ACP e incluso con k-means aparecen en un mismo grupo.
- A continuación de Madrid y Barcelona en cuanto al volumen de importaciones, aparece Valencia en solitario, que se diferencia de todas las demás provincias.
- Se forma un grupo de provincias donde vuelven a aparecer las provincias costeras mediterráneas junto las provincias del oeste de Andalucía.
- Un grupo muy numeroso se crea con más de la mitad de las provincias, tanto de interior como costeras. En este grupo se encuentran aquellas provincias que no destacan en cuanto al número de importaciones, es decir, tienen números bajos.
- Viendo el gráfico con los centroides (Ilustración 65) se puede ver qué secciones tienen más influencia en cada uno de los clústeres. Se puede apreciar como para cada grupo las secciones que más influencia tienen en él son las siguientes:

- Barcelona y Madrid (Grupo Rosa): Aparece en primera posición de 11 secciones, teniendo una influencia muy alta prácticamente en todas las secciones.
- Valencia (Grupo Rojo): Sección 9 de “Madera y carbón vegetal”.
- Alicante (Grupo Naranja): Sección 8 de “Pieles y cueros” y sección 12 de “Calzado y sombreros”.
- Pontevedra (Grupo Azul oscuro): Sección 1 de “Productos del reino animal”.
- Grupo Verde: Sección 3 de “Grasas y aceites” y la sección 5 de “Productos minerales”.
- Grupo Azul claro: Se trata del grupo más numeroso, se puede ver como no destaca ninguna sección, no apareciendo ningún pico representativo en el gráfico de centroides.

6.2.3 Análisis clúster del valor monetario de exportaciones en euros por habitante

6.2.3.1 Jerárquico

A partir de los datos de la matriz de exportaciones por habitante se han calculado los coeficientes de aglomeración y divisivo de cada método para determinar qué técnica brinda el mejor resultado.

Tabla 12. Coeficientes análisis clúster exportaciones por habitante

Método	Average	Single	Complete	Ward	Divisive
Coeficiente	0.6128012	0.6170658	0.6932473	0.7539638	0.6650777

En la Tabla 12 se muestra el coeficiente de aglomeración para los distintos procedimientos. En esta tabla se muestra que el método que tiene el coeficiente de aglomeración más alto es el método Ward. A continuación, se muestran los resultados de aplicar este método.

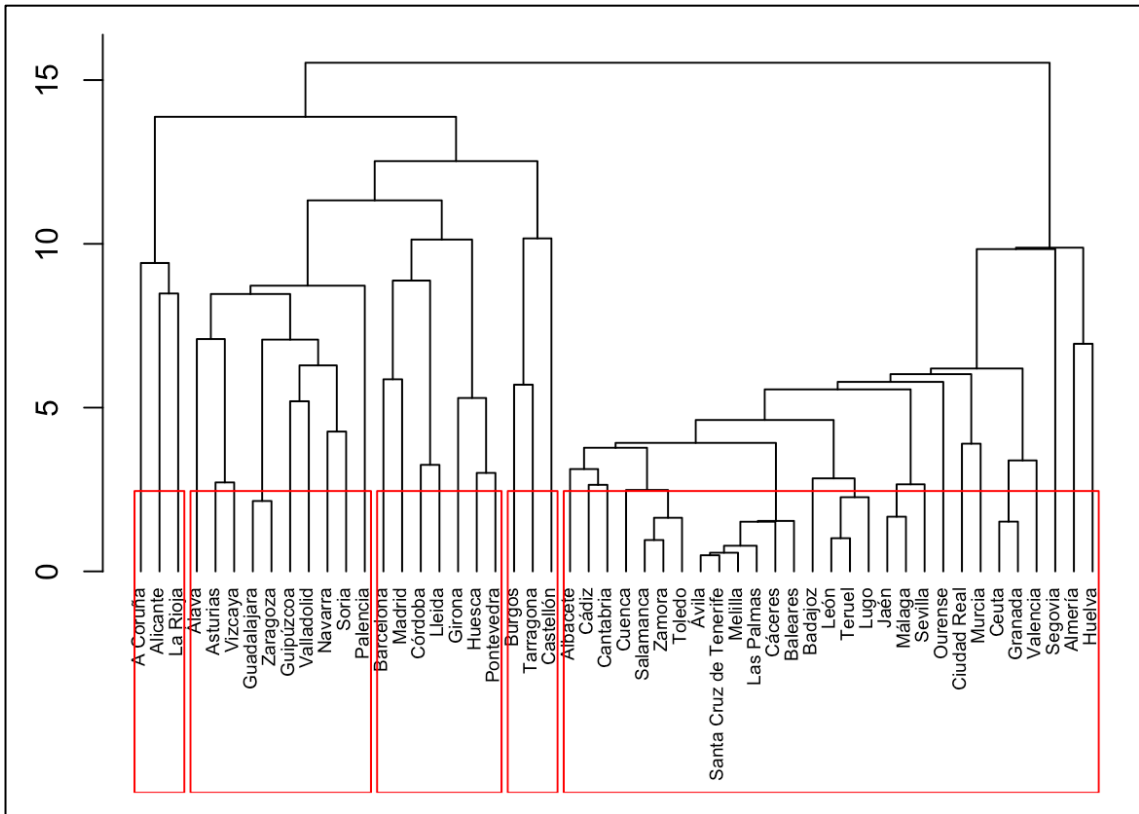


Ilustración 66. Dendrograma Ward exportaciones por habitante

Se muestra en el biplot (Ilustración 67) y en un mapa de España por provincias (Ilustración 68), cada uno de los grupos creados, con distintos colores.

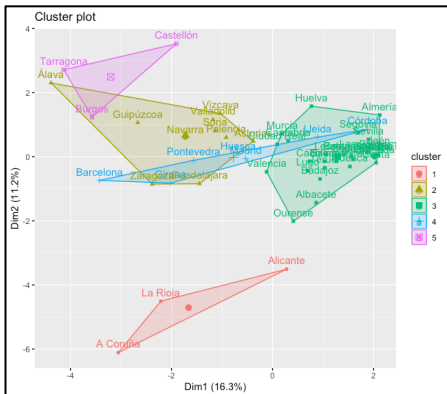


Ilustración 67. Grupos Ward exportaciones por habitante

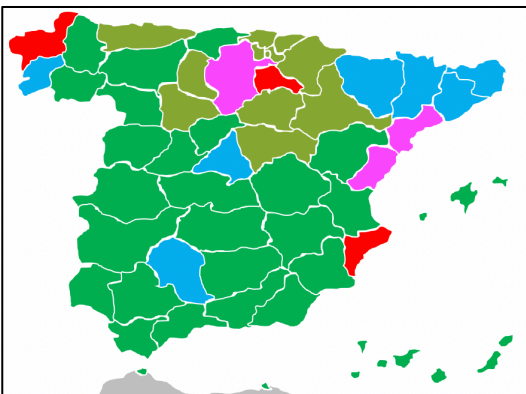


Ilustración 68. Mapa Ward exportaciones por habitante

- En el dendrograma (Ilustración 66) se ha dividido el conjunto de las provincias en 5 grupos realizando el análisis clúster con estos mismos.
- Se observa como la división de los grupos no es muy efectiva en cuanto a la representación gráfica del ACP, debido al solapamiento de unos grupos con otros.
- Se forma un grupo (Rojo) con La Coruña, La Rioja y Alicante que son 3 de las provincias mejor representadas por la segunda componente, bastante alejadas gráficamente de las demás.

- También, atendiendo a la segunda componente principal del ACP se forma un grupo (Rosa) con Tarragona, Burgos y Castellón.
- Madrid y Barcelona aparecen en el mismo grupo (Azul), junto con otras provincias como Pontevedra o Huesca y Girona.
- La mayoría de las provincias del norte de España se encuentran en otro grupo (Verde Claro).
- Geográficamente se observa un grupo (Verde) formado por prácticamente la totalidad de la mitad Sur de la península con la excepción de las provincias de Córdoba y Alicante que no se incluyen en este grupo.

6.2.3.2 No jerárquico

A partir de los datos de la matriz de exportaciones por habitante se ha aplicado el método de k-means. A continuación, se muestran los resultados de aplicar este método en el biplot (Ilustración 69) y en un mapa de España por provincias (Ilustración 70), cada uno de los grupos creados, con distintos colores.

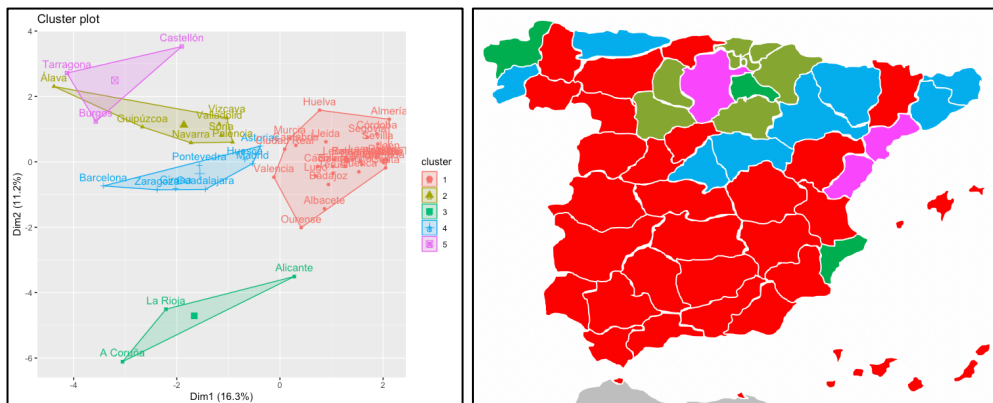


Ilustración 69. Grupos k-means exportaciones por habitante

Ilustración 70. Mapa k-means exportaciones por habitante

En el siguiente gráfico (Ilustración 71) se representa para cada sección, el valor de su respectivo centroide, de cada uno de los grupos formados, indicados con su color respectivo.

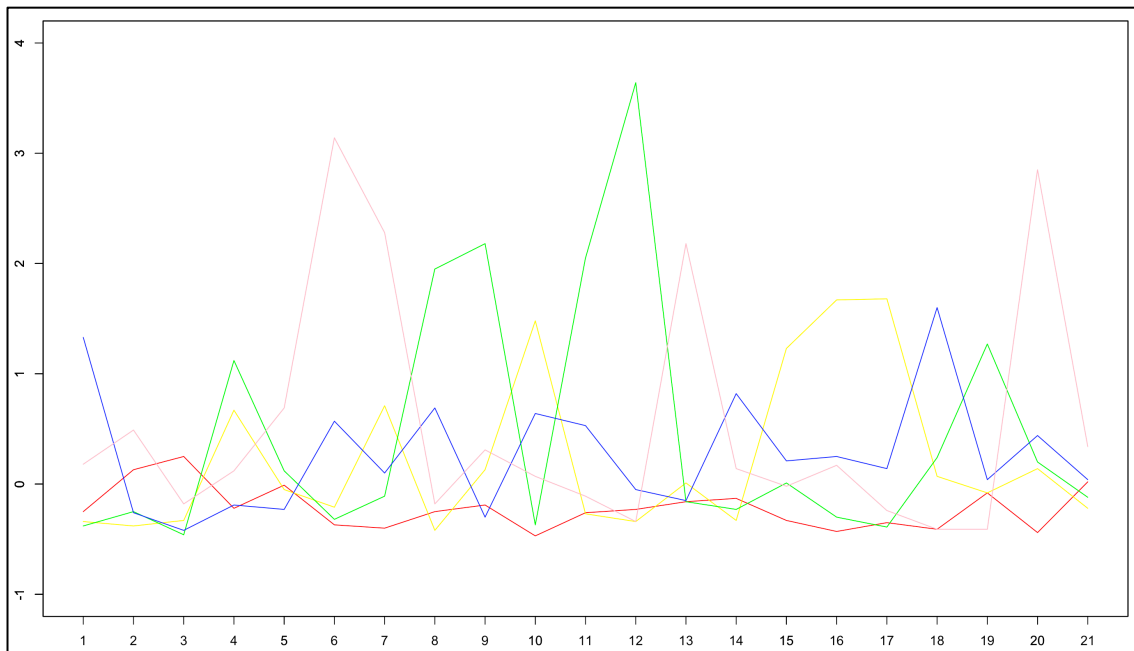


Ilustración 71. Centroides k-means exportaciones por habitante

En comparación con el análisis clúster jerárquico de la sección anterior se destacan los siguientes resultados:

- Salvo algunas provincias como Córdoba o Guadalajara entre otras, la distribución de los grupos es la misma.
- No se aprecia tanto solapamiento entre los grupos en el gráfico de ACP, por lo que se puede considerar un mejor análisis.

6.2.3.3 Conclusiones de todos los procedimientos clúster para el valor monetario total de exportaciones en euros

- No se trata de un buen análisis, ya que las dos componentes principales del ACP representan un 16,3% y 11,2% lo que no llega a una tercera parte de la variabilidad explicada.
- Se puede ver como la división de los grupos principalmente es geográfica ya que llama la atención de que prácticamente todo el noreste de España se encuentra en un mismo grupo y que la mitad sur de España se encuentra en otro. Por tanto, los dos grupos más numerosos se pueden interpretar geográficamente.
- Además, se tienen dos grupos más pequeños formados por tres provincias cada uno, formados por la segunda componente principal del ACP.
- Viendo el gráfico con los centroides (Ilustración 71) se puede ver que secciones tienen más influencia en cada uno de los clústeres. Se puede apreciar como para cada grupo las secciones que más influencia tienen en él son las siguientes:
 - Grupo Amarillo: Sección 10 de “Pasta de madera y papel”, sección 15 de “Metales comunes”, sección 16 de “Máquinas y aparatos eléctricos” y sección 17 de “Material de transporte”.

- Grupo Verde: Sección 4 de “Productos de las industrias alimentarias”, sección 8 de “Pieles y cueros”, sección 9 de “Madera y carbón vegetal”, sección 11 de “Materias textiles”, sección 12 de “Calzado y sombreros” y sección 19 de “Armas y municiones”.
- Grupo Rosa: Sección 6 de “Productos de las industrias químicas”, sección 7 de “Plástico y sus manufacturas”, sección 13 de “Manufacturas de piedra” y sección 20 de “Mercancías y productos diversos”.
- Grupo Azul: Sección 1 de “Productos del reino animal” y sección 18 de “Aparatos e instrumentos de óptica”.
- Grupo Rojo: Se trata del grupo más numeroso, se puede ver como no destaca ninguna sección, no apareciendo ningún pico representativo en el gráfico de centroides.

6.2.4 Análisis clúster del valor monetario de importaciones en euros por habitante

6.2.4.1 Jerárquico

A partir de los datos de la matriz de importaciones por habitante se han calculado los coeficientes de aglomeración y divisivo de cada método para determinar qué técnica brinda el mejor resultado.

Tabla 13. Coeficientes análisis clúster importaciones por habitante

Método	Average	Single	Complete	Ward	Divisive
Coeficiente	0.6598020	0.6246216	0.7361300	0.8438594	0.726242

En la Tabla 13 se muestra el coeficiente de aglomeración para los distintos procedimientos. En esta tabla se muestra que el método que tiene el coeficiente de aglomeración más alto es el método Ward. A continuación, se muestran los resultados de aplicar este método.

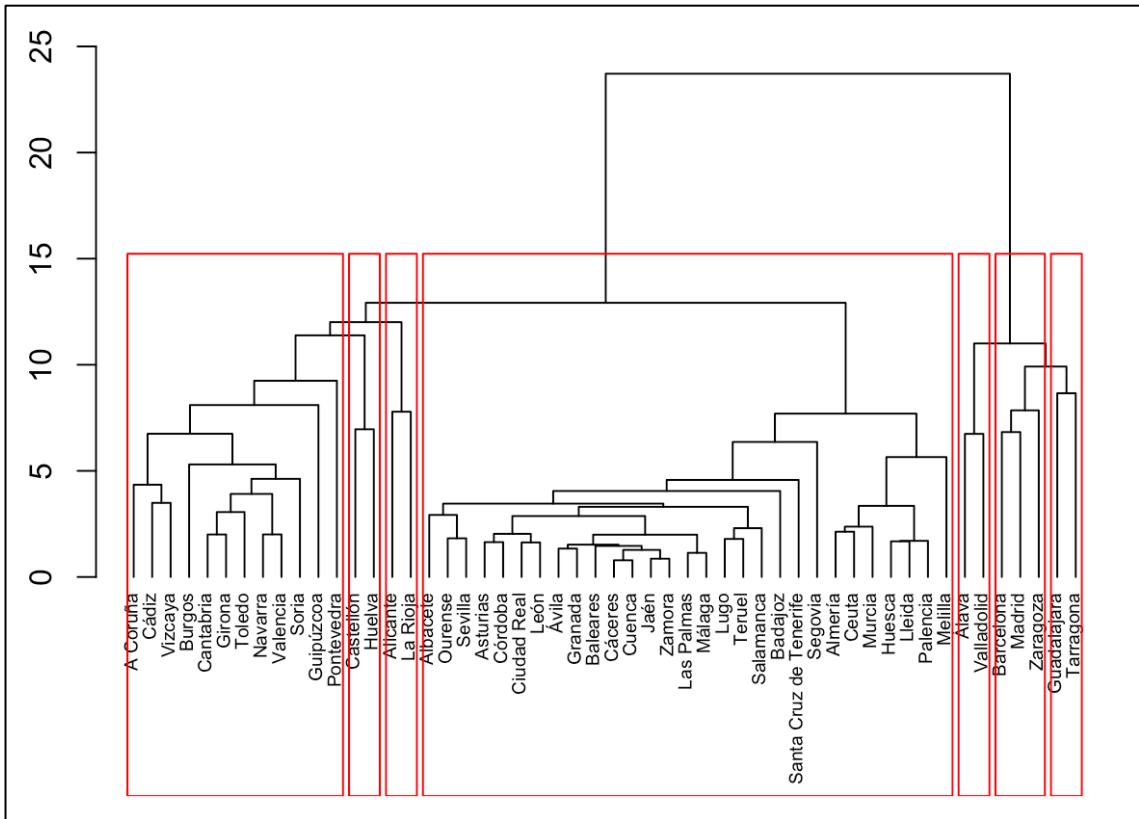


Ilustración 72. Dendograma Ward importaciones por habitante

Se muestra en el biplot (Ilustración 73) y en un mapa de España por provincias (Ilustración 74), cada uno de los grupos creados, con distintos colores.

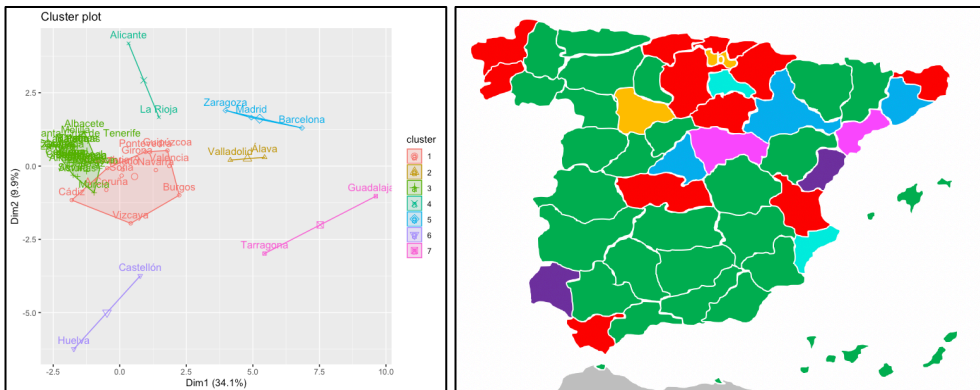


Ilustración 73. Grupos Ward importaciones por habitante

Ilustración 74. Mapa Ward importaciones por habitante

- En el dendograma (Ilustración 72) se han dividido el conjunto de las provincias en 7 grupos realizando el análisis clúster con estos mismos.
- No se aprecia una explicación geográfica para la distribución de los grupos, ya que provincias de un mismo grupo están bastante alejadas en la península, salvo en la mitad sur de España, dónde sí que predomina un único grupo.
- Madrid y Barcelona aparecen en el mismo grupo (Azul), acompañadas de la provincia de Zaragoza.

- Se puede apreciar como hay un grupo (Verde) muy numeroso y luego unos cuantos grupos (Naranja, Morado, Rosa o Cian) formados por dos y tres provincias.

6.2.4.2 No jerárquico

A partir de los datos de la matriz de importaciones por habitante se ha aplicado el método de k-means. A continuación, se muestran los resultados de aplicar este método en el biplot (Ilustración 75) y en un mapa de España por provincias (Ilustración 76), cada uno de los grupos creados, con distintos colores.

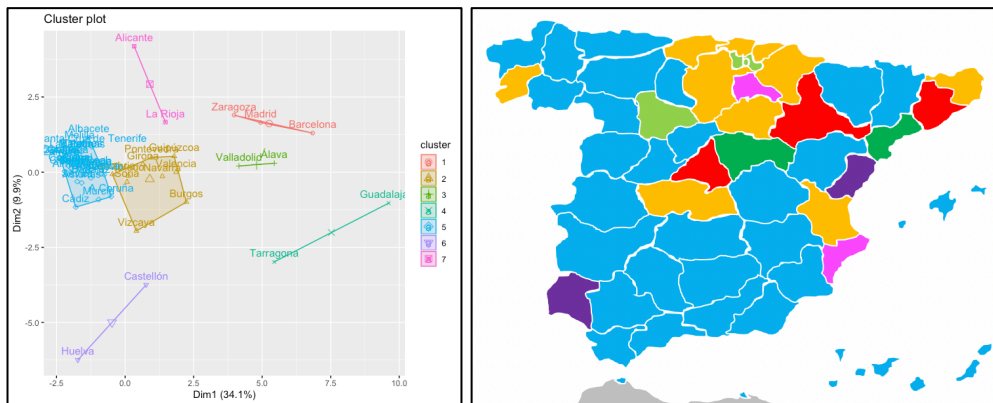


Ilustración 75. Grupos k-means importaciones por habitante

Ilustración 76. Mapa k-means importaciones por habitante

En el siguiente gráfico (Ilustración 77) se representa para cada sección, el valor de su respectivo centroide, de cada uno de los grupos formados, indicados con su color respectivo.

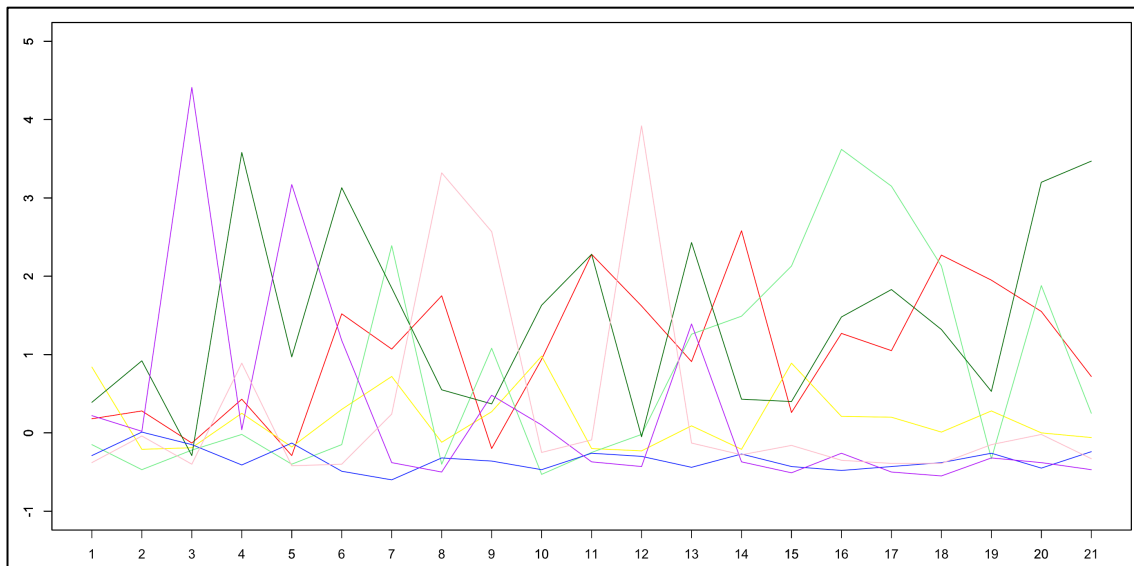


Ilustración 77. Centroides k-means importaciones por habitante

En comparación con el análisis clúster jerárquico de la sección anterior se destacan los siguientes resultados:

- Salvo algunas provincias que cambian de grupo como Murcia o Cádiz, la distribución de los grupos es la misma.

- No se aprecia ningún solapamiento entre los grupos en el gráfico de ACP, por lo que se puede considerar un mejor análisis.

6.2.4.3 Conclusiones de todos los procedimientos clúster para el valor monetario total de exportaciones en euros

- La división de los grupos se realiza mayormente debido a las componentes del ACP, puesto que se puede apreciar como en dicho análisis las provincias aparecen correctamente agrupadas en el gráfico.
- El análisis no es bueno, ya que entre las dos primeras componentes no se llega a la mitad de la variabilidad explicada.
- Geográficamente llama la atención que prácticamente la mitad sur del país pertenece a un mismo grupo.
- La primera componente principal, se encuentra relacionada con la cantidad de importaciones que las provincias realizan de las secciones 17 de “Material de Transporte”, la 16 de “Máquinas y aparatos eléctricos”, la 6 de “Productos de la industria química” y la 20 de “Mercancías y productos diversos”.
- En el gráfico con los centroides (Ilustración 77) se puede ver que secciones tienen mas influencia en cada uno de los clústeres. Se puede apreciar como para cada grupo las secciones que más influencia tienen en él son las siguientes:
 - Grupo Rojo: Sección 11 de “Materias textiles”, sección 14 de “Perlas finas” y sección 18 de “Instrumentos y aparatos de óptica”.
 - Grupo Amarillo: Sección 1 de “Productos del reino animal”.
 - Grupo Verde Claro: Sección 7 de “Plástico y sus manufacturas”, sección 15 de “Metales comunes”, sección 16 de “Máquinas y aparatos eléctricos” y sección 17 de “Material de transporte”.
 - Grupo Verde Oscuro: Sección 2 de “Productos del reino vegetal”, sección 4 de “Productos de las industrias alimentarias”, sección 6 de “Productos de las industrias químicas”, sección 11 de “Materias textiles”, sección 13 de “Manufacturas de piedra”, sección 20 de “Mercancías y productos diversos” y sección 21 de “Objetos de arte o colección”.
 - Grupo Rosa: Sección 8 de “Pieles y cueros”, sección 9 de “Madera y carbón vegetal” y sección 12 de “Calzado y sombreros”.
 - Grupo Morado: Sección 3 de “Grasas y aceites” y sección 5 de “Productos minerales”.
 - Grupo Azul: Se trata del grupo más numeroso, se puede ver como no destaca ninguna sección, no apareciendo ningún pico representativo en el gráfico de centroides.

6.2.5 Análisis clúster del porcentaje del valor de exportaciones de cada sección por provincia

6.2.5.1 Jerárquico

A partir de los datos de la matriz del porcentaje del valor de exportaciones se han calculado los coeficientes de aglomeración y divisivo de cada método para determinar qué técnica brinda el mejor resultado.

Tabla 14. Coeficientes análisis clúster exportaciones porcentaje sección

Método	Average	Single	Complete	Ward	Divisive
Coeficiente	0.6181509	0.6003217	0.6762871	0.6999833	0.6525538

En la Tabla 14 se muestra el coeficiente de aglomeración para los distintos procedimientos. En esta tabla se muestra que el método que tiene el coeficiente de aglomeración más alto es el método Ward. A continuación, se muestran los resultados de aplicar este método.

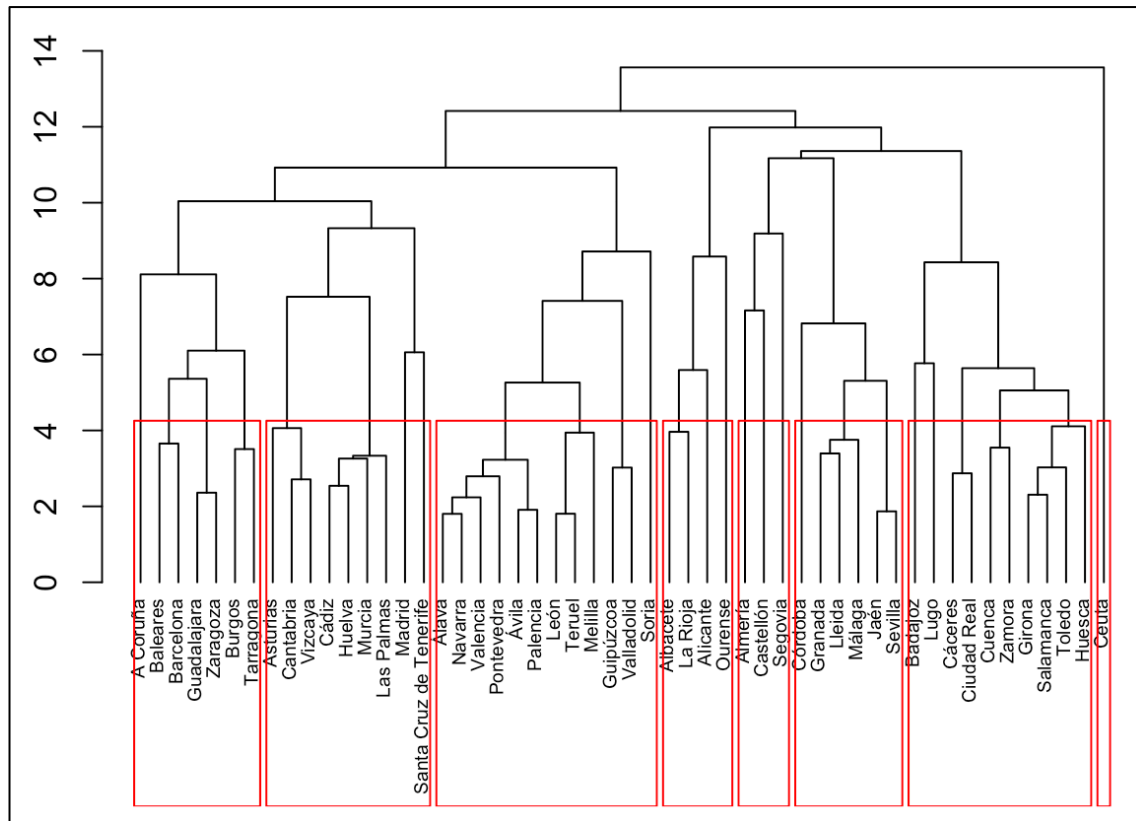


Ilustración 78. Dendrograma Ward exportaciones porcentaje sección

Se muestra en el biplot (Ilustración 79) y en un mapa de España por provincias (Ilustración 80), cada uno de los grupos creados, con distintos colores.

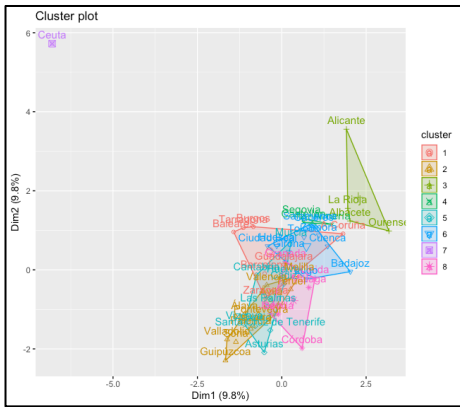


Ilustración 79. Grupos Ward exportaciones Ward porcentaje sección

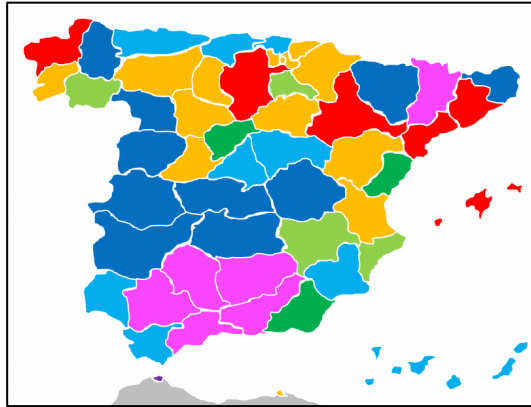


Ilustración 80. Mapa Ward exportaciones porcentaje sección

En el siguiente gráfico (Ilustración 81) se representa para cada sección, el valor de su porcentaje respecto del total de exportaciones, con el color del grupo al que pertenece. Al tratarse de proporciones se puede ver qué secciones destacan en cada uno de los grupos.

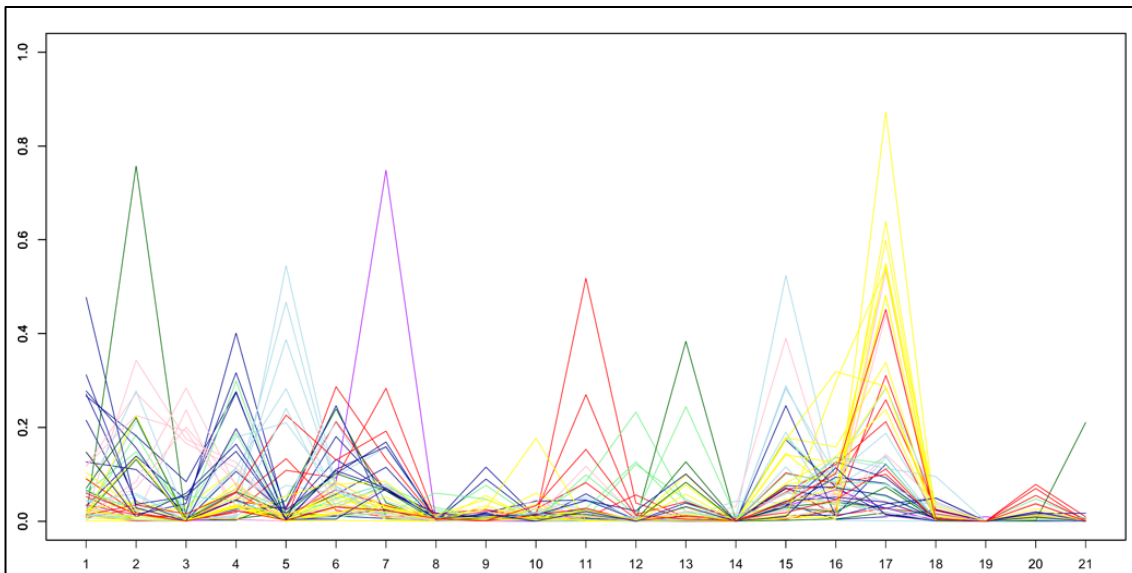


Ilustración 81. Porcentaje sección Ward clúster exportaciones

Hay una gran diferencia en la representación del análisis ACP de Ceuta con las demás provincias. Esta diferencia puede deberse a que Ceuta destaca únicamente en la Sección 7 que es la de “Plástico y sus manufacturas”.

Gráficamente la división de los grupos no parece la más adecuada puesto que se aprecia bastante solapamiento entre ellos. Fijándose en el mapa (Ilustración 80) con los 8 grupos se aprecia como sí hay una diferencia geográfica, a parte de la de Ceuta. El grupo rojo ocupa la mitad norte de España y las islas. La parte oeste de España representa el grupo azul oscuro y la mayor parte de Andalucía es determinada por el grupo rosa.

Por secciones, aunque no sea nada demasiado llamativo, se puede ver como algunos grupos destacan en determinadas secciones:

- Grupo Verde Oscuro en la sección 2 de “Productos del reino vegetal” y en la sección 13 de “Manufacturas de piedra”.

- Grupo Azul Oscuro en la sección 1 de “Animales vivos y productos del reino animal” y en la sección 4 de “Productos de industrias alimentarias”.
- Grupo Azul Claro en la sección 5 de “Productos minerales”.
- Grupo Rojo en la sección 11 de “Materias textiles” y en la sección 20 de “Mercancías y productos diversos”.
- Grupo Amarillo en la sección 17 de “Material de transporte”.

6.2.5.2 No Jerárquico

A partir de los datos de la matriz del porcentaje del valor de exportaciones se ha aplicado el método de k-means. A continuación, se muestran los resultados de aplicar este método en el biplot (Ilustración 82) y en un mapa de España por provincias (Ilustración 83), cada uno de los grupos creados, con distintos colores.

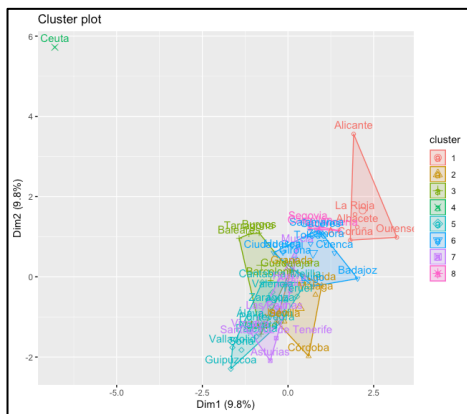


Ilustración 82. Grupos k-means exportaciones porcentaje sección

Ilustración 83. Mapa k-means exportaciones porcentaje sección

En el siguiente gráfico (Ilustración 84) se representa para cada sección, el valor de su porcentaje respecto del total de exportaciones, con el color del grupo al que pertenece.

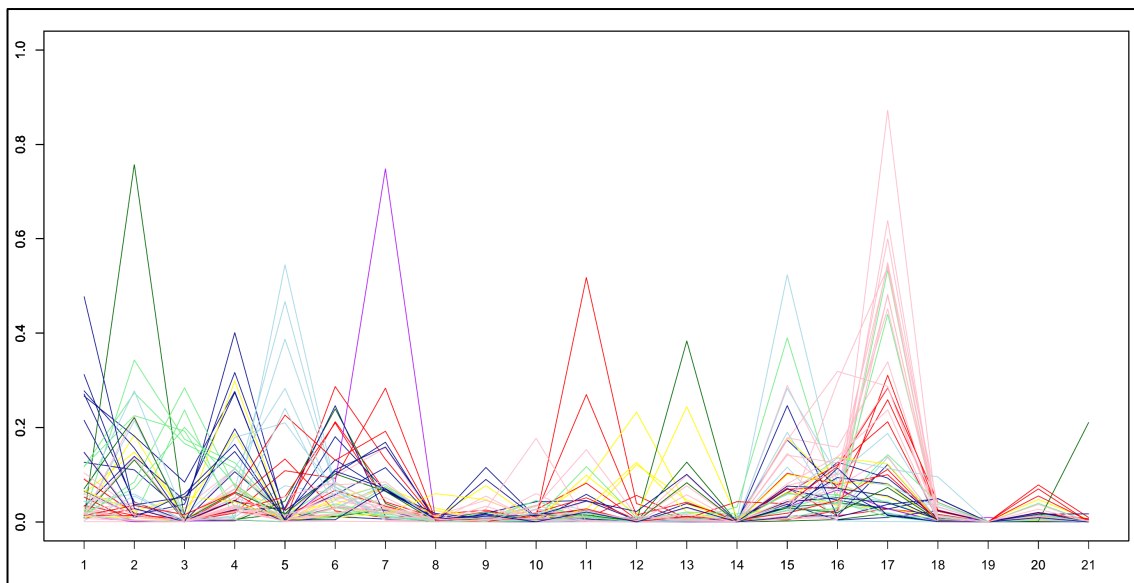


Ilustración 84. Porcentaje sección k-means clústeres exportaciones

En el siguiente gráfico (Ilustración 85) se representa para cada sección, el valor de su respectivo centroide, de cada uno de los grupos formados, indicados con su color respectivo.

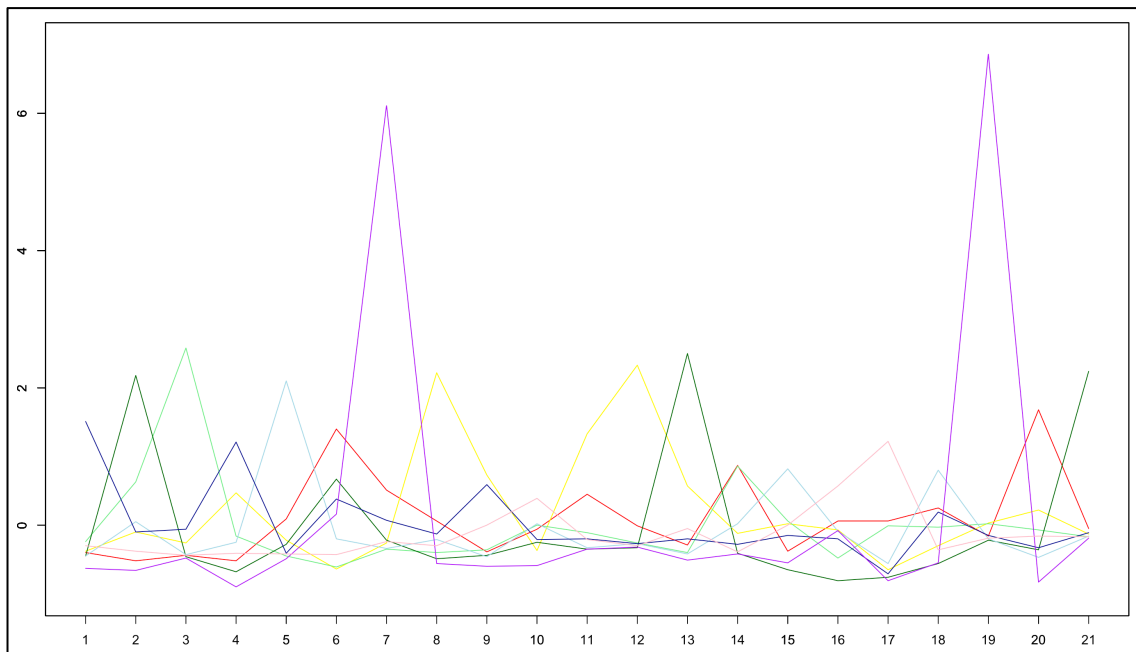


Ilustración 85. Porcentaje sección k-means centroides exportaciones

Por secciones se puede ver como cada grupo destaca en determinadas secciones:

- Grupo Rojo en la sección 6 de “Productos de las industrias químicas”, en la sección 14 de “Perlas finas” y en la sección 20 de “Mercancías y productos diversos”.
- Grupo Verde Claro en la sección 3 de “Grasas y aceites”.
- Grupo Azul Oscuro en la sección 1 de “Animales vivos y productos del reino animal” y en la sección 4 de “Productos de industrias alimentarias”.
- Grupo Morado en la sección 7 de “Plástico y sus manufacturas” y en la sección 19 de “Armas y municiones”.
- Grupo Rosa en la sección 10 de “Pasta de madera y papel” y en la sección 17 de “Material de transporte”.
- Grupo Verde Oscuro en la sección 2 de “Productos del reino vegetal” y en la sección 13 de “Manufacturas de piedra” y en la sección 21 de “Objetos de arte y colección”.
- Grupo Azul Claro en la sección 5 de “Productos minerales”, en la sección 15 de “Metales comunes” y en la sección 18 de “Instrumentos y aparatos de óptica”.
- Grupo Amarillo en la sección 8 de “Pieles y cueros”, en la sección 11 de “Materias textiles” y en la sección 12 de “Calzado y sombreros”.

6.2.6 Análisis clúster del porcentaje del valor de importaciones de cada sección por provincia

6.2.6.1 Jerárquico

A partir de los datos de la matriz del porcentaje del valor de importaciones se han calculado los coeficientes de aglomeración y divisivo de cada método para determinar qué técnica brinda el mejor resultado.

Tabla 15. Coeficientes análisis clúster importaciones porcentaje sección

Método	Average	Single	Complete	Ward	Divisive
Coeficiente	0.6178809	0.5831335	0.6670676	0.7066923	0.6657834

En la Tabla 15 se muestra el coeficiente de aglomeración para los distintos procedimientos. En esta tabla se muestra que el método que tiene el coeficiente de aglomeración más alto es el método Ward. A continuación, se muestran los resultados de aplicar este método.

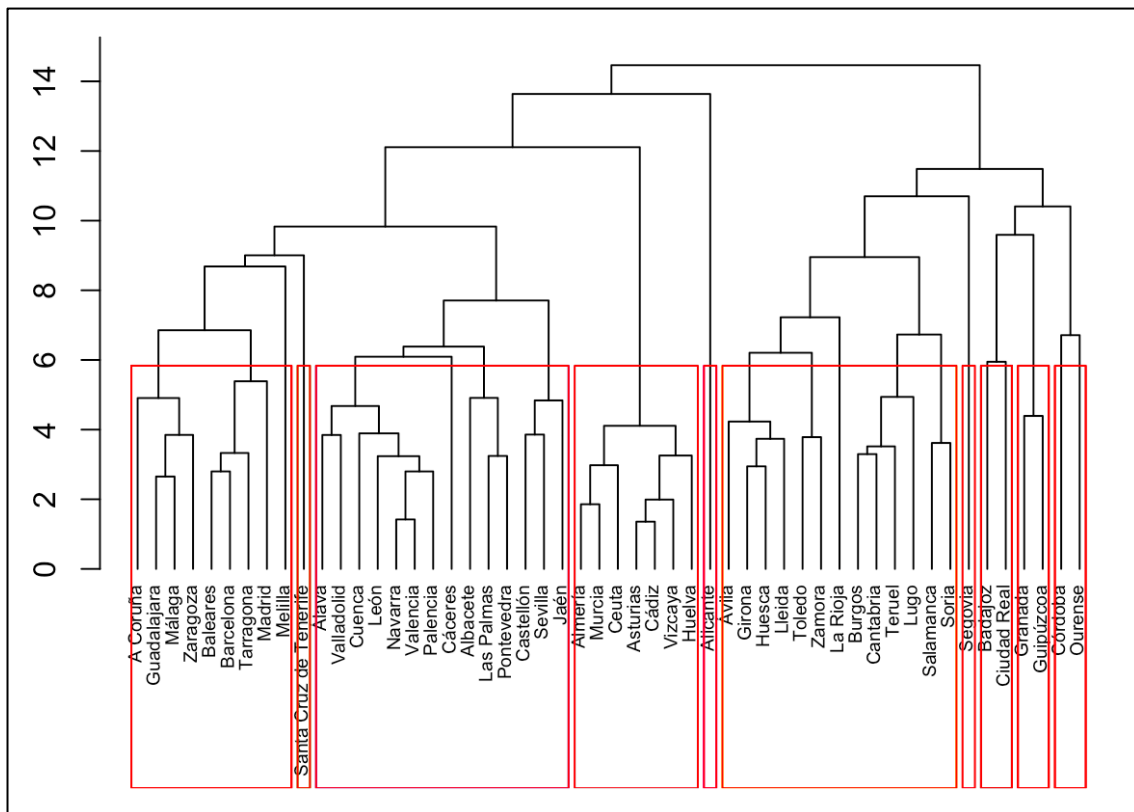


Ilustración 86. Dendrograma Ward importaciones porcentaje sección

Se muestra en el biplot (Ilustración 86) y en un mapa de España por provincias (Ilustración 88), cada uno de los grupos creados, con distintos colores.

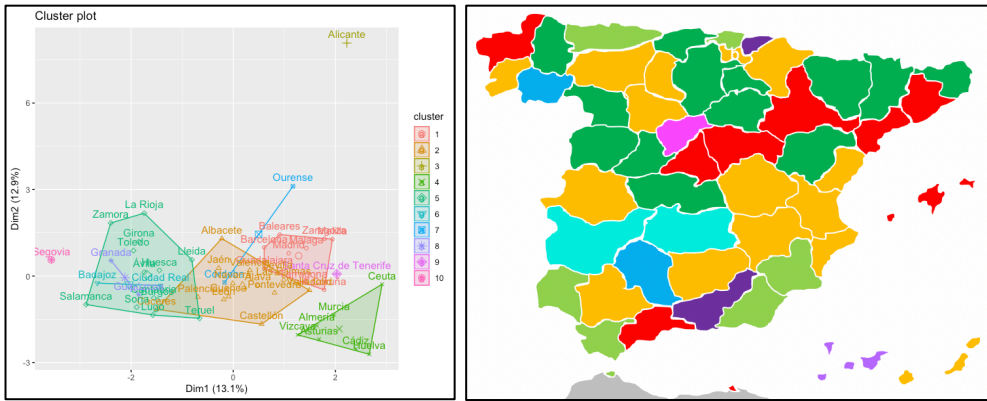


Ilustración 87. Grupos Ward importaciones porcentaje sección

Ilustración 88. Mapa Ward importaciones porcentaje sección

En el siguiente gráfico (Ilustración 89) se representa para cada sección, el valor de su porcentaje respecto del total de exportaciones, con el color del grupo al que pertenece.

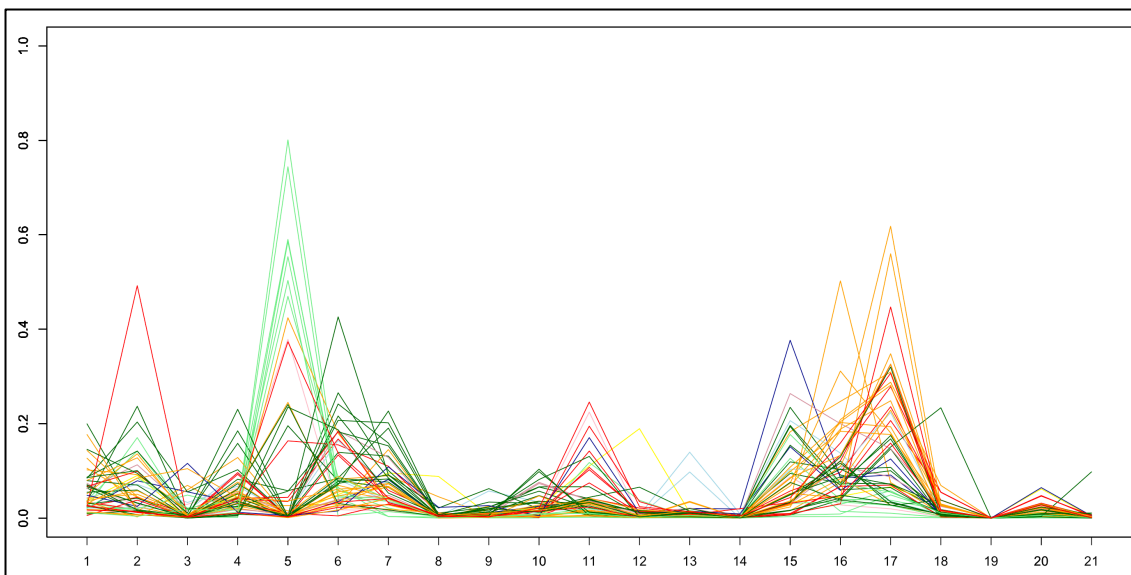


Ilustración 89. Porcentaje sección Ward clústeres importaciones

Geográficamente no se aprecia una distribución por grupos muy clara más allá de que para el mapa (Ilustración 88) con 10 grupos, el azul predomina en la parte norte y el rojo en la parte este de la península.

Por secciones se puede ver como cada grupo destaca en determinadas secciones:

- Grupo Naranja en la Sección 16 de “Máquinas y aparatos eléctricos” y en la Sección 17 de “Material de transporte”.
- Grupo Verde Claro en la Sección 5 de “Productos minerales”.
- Grupo Verde Oscuro en la Sección 4 de “Productos de industrias alimentarias”, en la Sección 6 de “Productos de las industrias químicas”, en la Sección 7 de “Plástico y sus manufacturas”, en la Sección 9 de “Madera y carbón vegetal” y en la Sección 10 de “Pasta de madera”.

- Grupo Rojo en la Sección 2 de “Productos del reino vegetal” y en la Sección 11 de “Materias textiles”.
- Grupo Azul Claro en la Sección 13 de “Manufacturas de piedra”.
- Grupo Azul Oscuro en la Sección 15 de “Metales comunes”.

6.2.6.2 No Jerárquico

A partir de los datos de la matriz del porcentaje del valor de importaciones se ha aplicado el método de k-means. A continuación, se muestran los resultados de aplicar este método en el biplot (Ilustración 90) y en un mapa de España por provincias (Ilustración 91), cada uno de los grupos creados, con distintos colores.

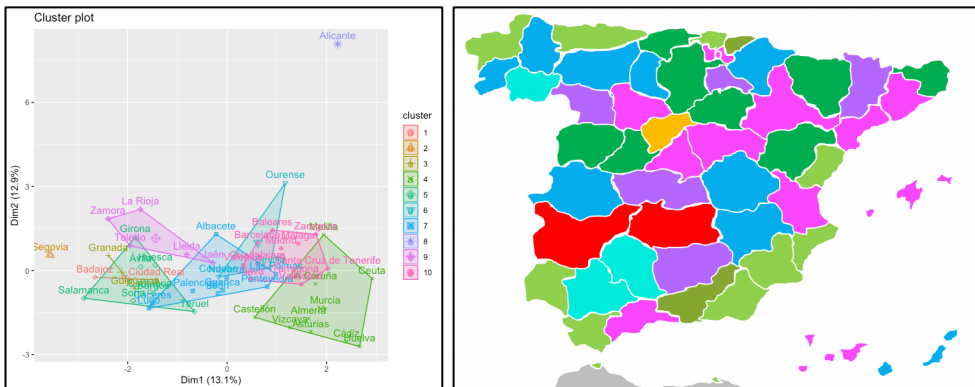


Ilustración 90. Grupos k-means importaciones porcentaje sección

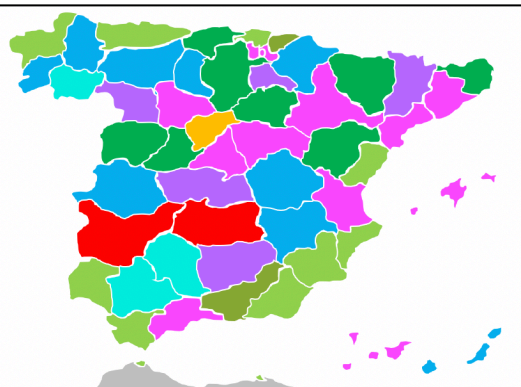


Ilustración 91. Mapa k-means importaciones porcentaje sección

En el siguiente gráfico (Ilustración 92) se representa para cada sección, el valor de su porcentaje respecto del total de exportaciones, con el color del grupo al que pertenece.

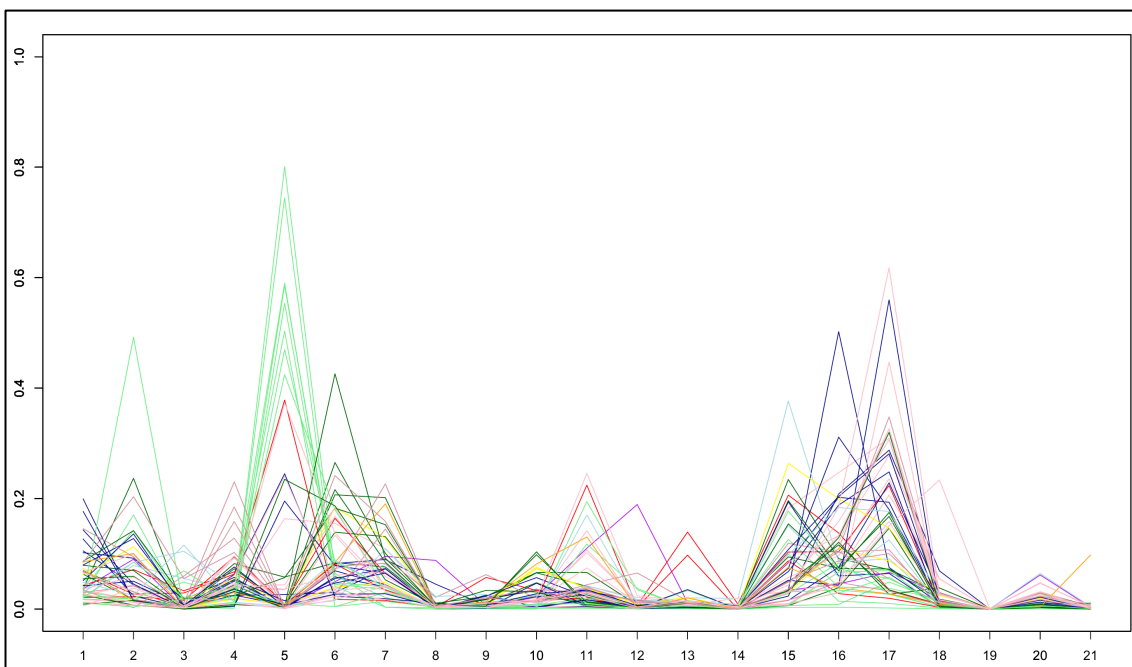


Ilustración 92. Porcentaje sección k-means clústeres importaciones

En el siguiente gráfico (Ilustración 93) se representa para cada sección, el valor de su respectivo centroide, de cada uno de los grupos formados, indicados con su color respectivo.

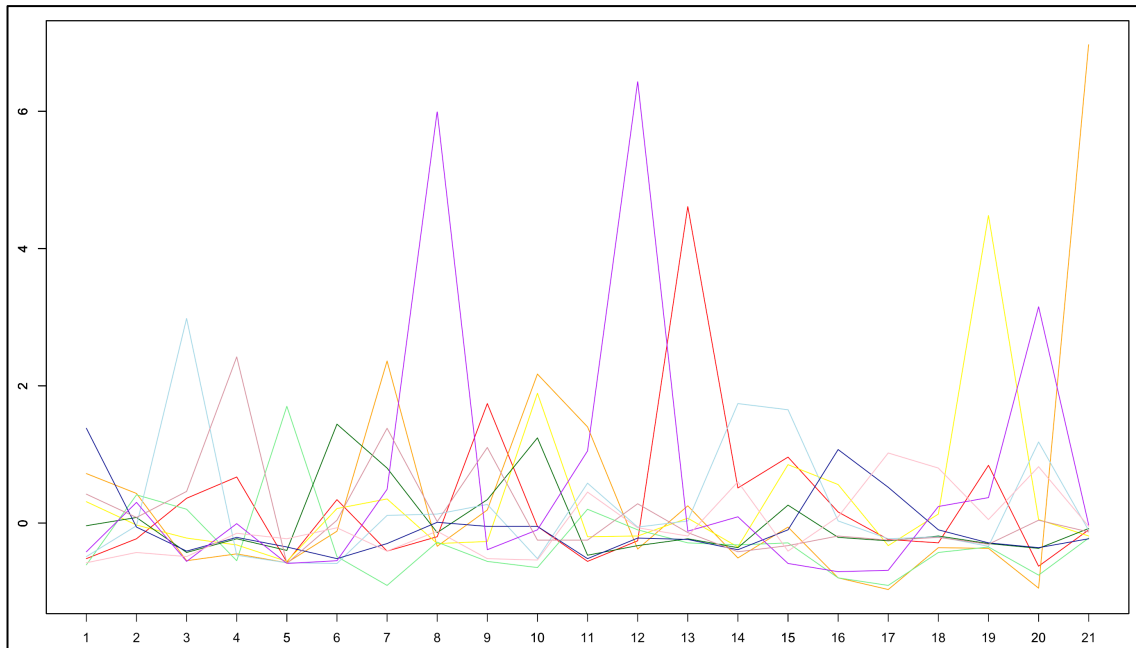


Ilustración 93. Porcentaje sección k-means centroides importaciones

Por secciones se puede ver como cada grupo destaca en determinadas secciones:

- Grupo Morado en la sección 8 de "Pieles y cueros", en la sección 12 de "Calzado y sombreros" y en la sección 20 de "Mercancías y productos diversos".
- Grupo Rojo en la sección 13 de "Manufacturas de madera".
- Grupo Naranja en la sección 21 de "Objetos de arte y colección".
- Grupo Amarillo en la sección 19 de "Armas y municiones".

7 Conclusiones

En este trabajo de fin de grado se ha conseguido realizar una clasificación de las provincias de España en base a los datos del comercio exterior correspondientes al año 2017 aplicando una serie de técnicas multivariantes.

Se han cumplido los objetivos inicialmente establecidos. Se empieza por realizar un primer proceso de tratamiento y procesamiento de los datos brutos obtenidos de la Agencia Tributaria con la finalidad de que éstos sean tratables. Tras ello se ha podido realizar la clasificación con distintas técnicas de análisis.

Se han utilizado varios conjuntos de datos obtenidos de la página web de la Agencia Tributaria con un tamaño total de 225.2 Mb. Tras realizar un procesamiento de estos se construyen distintos conjuntos de datos agregados de importaciones y exportaciones. Estos datos se utilizan para la clasificación de las provincias de España en base a su estructura de comercio exterior.

Con los análisis realizados se llega a las siguientes conclusiones:

- 1) Existe una gran diferencia entre Madrid y Barcelona, respecto a las demás provincias en cuanto a la cantidad de exportaciones e importaciones.
- 2) Barcelona es la principal provincia en el comercio exterior de España. A continuación, se encuentra Madrid, con una menor diferencia entre ambas para las Importaciones que para las Exportaciones.
- 3) Después de estas dos grandes provincias, aparece un grupo entre las que se encuentra Valencia y Alicante, en definitiva, las provincias Mediterráneas que se sitúan en un segundo escalón en volumen de comercio exterior.
- 4) Las provincias de interior apenas sobresalen en el comercio exterior, salvo algunas excepciones para ciertas secciones como puede ser Valladolid con el Material de Transporte debido a las fábricas de automóviles.
- 5) Se ha encontrado una cierta importancia de la situación geográfica en cuanto a que las provincias costeras exportan e importan más productos, por lo que permite confirmar que la existencia de un puerto es muy relevante para el comercio exterior.
- 6) Se ha visto como el principal criterio para la clasificación de provincias es el volumen de exportaciones/importaciones que realizan. Sin embargo, en algunos análisis realizados se ha podido apreciar que ciertos grupos se formaban con provincias con valores similares de exportaciones/importaciones de determinadas secciones de mercancías.
- 7) El estudio del volumen de las exportaciones/importaciones en euros por habitante ha permitido realizar un análisis donde las provincias de Barcelona y Madrid no se encontraban tan alejadas de las demás. Se han podido ver ciertas provincias con poca población que son punteras en ciertas secciones de mercancías. Sin embargo, la poca correlación existente entre las variables no permite interpretar grandes resultados.

- 8) El análisis más interesante para estudiar las características de las provincias es el que utiliza los porcentajes por sección, donde se han podido utilizar gráficos con las proporciones que exporta/importa cada provincia de una determinada sección de mercancías. Es el estudio en el que mejor se puede apreciar claramente la importancia de ciertas secciones para formar unos grupos. Esto es debido a que se comprueba como provincias con porcentajes similares en ciertas secciones pertenecen al mismo clúster.

Como trabajo futuro se puede analizar la evolución del comercio exterior en las provincias a lo largo de los años. En este trabajo se han utilizado datos correspondientes a un único año, el año 2017. En la Agencia Tributaria se pueden obtener los datos pertenecientes a otros años y se podría comparar si se mantienen estos resultados o por el contrario hay una evolución en la clasificación de provincias en cuanto al comercio exterior.

8 Referencias

1. Big Data. [Online] Available: <https://www.oracle.com/es/big-data/guide/what-is-big-data.html>
2. Comercio Exterior. [Online] Available: <https://www.significados.com/comercio-exterior/>
3. Comercio Exterior en España. [Online] Available: <http://www.cargoflores.com/cifras-record-comercio-exterior-2017/>
4. Ranking lenguajes de programación. [Online] Available: <https://blog.revolutionanalytics.com/2015/07/ieee-2015-rankings.html>
5. RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>
6. Ventajas de RStudio. [Online] Available: <https://caminosaleatorios.wordpress.com/2013/07/02/algunas-ventajas-de-rstudio/>
7. Ventanas de RStudio. [Online] Available: <https://nubededatos.blogspot.com/2013/12/usando-rstudio-por-primera-vez.html>
8. Agencia Tributaria. [Online] Available: <https://www.agenciatributaria.es>
9. Instituto Nacional de Estadística. [Online] Available: <https://www.ine.es>
10. Visualización datos en R. [Online] Available: <http://www.xavigimenez.net/blog/2012/09/visualizing-data-with-r/>
11. Mapa de España en R. [Online] Available: <https://atc1.aut.uah.es/~david/notes/2017/11/choropleth-maps-in-lattice/>
12. Pérez López, César. Técnicas de análisis multivariante de datos: aplicaciones con SPSS . Pearson-Prentice-Hall, 2004.
13. Aluja Banet, Tomas, and Alain Morineau. *Aprender de los datos: el análisis de componentes principales : una aproximación desde el data mining* . Universidad de Barcelona, 1999.
14. Plantilla Excel provincias España. [Online] Available: <https://excelyvba.com/provincias-espana-excel>
15. Pebesma, E.J., R.S. Bivand, 2005. Classes and methods for spatial data in R. R News 5 (2), <https://cran.r-project.org/doc/Rnews/>
16. Roger S. Bivand, Edzer Pebesma, Virgilio Gomez-Rubio, 2013. Applied spatial data analysis with R, Second edition. Springer, NY. <http://www.asdar-book.org/>

17. Erich Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2.
<https://CRAN.R-project.org/package=RColorBrewer>
18. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
19. Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL
<http://www.jstatsoft.org/v21/i12/>
20. Alboukadel Kassambara and Fabian Mundt (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. <https://CRAN.R-project.org/package=factoextra>
21. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K.(2019). cluster: Cluster Analysis Basics and Extensions. R package version 2.1.0.
22. Lionel Henry and Hadley Wickham (2019). purrr: Functional Programming Tools. R package version 0.3.3.
<https://CRAN.R-project.org/package=purrr>

9 Anexos

9.1 Lista de capítulos de mercancías

La siguiente tabla muestra la lista de capítulos de mercancías según la “Agencia Tributaria” tal como se explica en la sección 3.2 de esta memoria.

Capítulo
CAPÍTULO 1 – ANIMALES VIVOS
CAPÍTULO 2 – CARNE Y DESPOJOS COMESTIBLES
CAPÍTULO 3 – PESCADOS Y CRUSTÁCEOS, MOLUSCOS Y DEMÁS INVERTEBRADOS ACUÁTICOS
CAPÍTULO 4 – LECHE Y PRODUCTOS LÁCTEOS; HUEVOS DE AVE; MIEL NATURAL; PRODUCTOS COMESTIBLES DE ORIGEN ANIMAL NO EXPRESADOS NI COMPRENDIDOS EN OTRA PARTE
CAPÍTULO 5 – LOS DEMÁS PRODUCTOS DE ORIGEN ANIMAL NO EXPRESADOS NI COMPRENDIDOS EN OTRA PARTE
CAPÍTULO 6 – PLANTAS VIVAS Y PRODUCTOS DE LA FLORICULTURA
CAPÍTULO 7 – HORTALIZAS, PLANTAS, RAÍCES Y TUBÉRCULOS ALIMENTICIOS
CAPÍTULO 8 – FRUTAS Y FRUTOS COMESTIBLES; CORTEZAS DE AGRIOS (CÍTRICOS), MELONES O SANDÍAS
CAPÍTULO 9 – CAFÉ, TÉ, YERBA MATE Y ESPECIAS
CAPÍTULO 10 – CEREALES
CAPÍTULO 11 – PRODUCTOS DE LA MOLINERÍA; MALTA; ALMIDÓN Y FÉCULA; INULINA; GLUTEN DE TRIGO
CAPÍTULO 12 – SEMILLAS Y FRUTOS OLEAGINOSOS; SEMILLAS Y FRUTOS DIVERSOS; PLANTAS INDUSTRIALES O MEDICINALES; PAJA Y FORRAJE
CAPÍTULO 13 – GOMAS, RESINAS Y DEMÁS JUGOS Y EXTRACTOS VEGETALES
CAPÍTULO 14 – MATERIAS TRENZABLES Y DEMÁS PRODUCTOS DE ORIGEN VEGETAL, NO EXPRESADOS NI COMPRENDIDOS EN OTRA PARTE
CAPÍTULO 15 – GRASAS Y ACEITES ANIMALES O VEGETALES; PRODUCTOS DE SU DESDOBLAMIENTO; GRASAS ALIMENTICIAS ELABORADAS; CERAS DE ORIGEN ANIMAL O VEGETAL
CAPÍTULO 16 – PREPARACIONES DE CARNE, PESCADO O DE CRUSTÁCEOS, MOLUSCOS O DEMÁS INVERTEBRADOS ACUÁTICOS
CAPÍTULO 17 – AZÚCARES Y ARTÍCULOS DE CONFITERÍA
CAPÍTULO 18 – CACAO Y SUS PREPARACIONES
CAPÍTULO 19 – PREPARACIONES A BASE DE CEREALES, HARINA, ALMIDÓN, FÉCULA O LECHE; PRODUCTOS DE PASTELERÍA
CAPÍTULO 20 – PREPARACIONES DE HORTALIZAS, FRUTAS U OTROS FRUTOS O DEMÁS PARTES DE PLANTAS

CAPÍTULO 21 – PREPARACIONES ALIMENTICIAS DIVERSAS
CAPÍTULO 22 – BEBIDAS, LÍQUIDOS ALCOHÓLICOS Y VINAGRE
CAPÍTULO 23 – RESIDUOS Y DESPERDICIOS DE LAS INDUSTRIAS ALIMENTARIAS; ALIMENTOS PREPARADOS PARA ANIMALES
CAPÍTULO 24 – TABACO Y SUCEDÁNEOS DEL TABACO ELABORADOS
CAPÍTULO 25 – SAL; AZUFRE; TIERRAS Y PIEDRAS; YESOS, CALES Y CEMENTOS
CAPÍTULO 26 – MINERALES METALÍFEROS, ESCORIAS Y CENIZAS
CAPÍTULO 27 – COMBUSTIBLES MINERALES, ACEITES MINERALES Y PRODUCTOS DE SU DESTILACIÓN; MATERIAS BITUMINOSAS; CERAS MINERALES
CAPÍTULO 28 – PRODUCTOS QUÍMICOS INORGÁNICOS; COMPUESTOS INORGÁNICOS U ORGÁNICOS DE METAL PRECIOSO, DE ELEMENTOS RADIATIVOS, DE METALES DE LAS TIERRAS RARAS O DE ISÓTOPOS
CAPÍTULO 29 – PRODUCTOS QUÍMICOS ORGÁNICOS
CAPÍTULO 30 – PRODUCTOS FARMACÉUTICOS
CAPÍTULO 31 – ABONOS
CAPÍTULO 32 – EXTRACTOS CURTIENTES O TINTÓREOS; TANINOS Y SUS DERIVADOS; PIGMENTOS Y DEMÁS MATERIAS COLORANTES; PINTURAS Y BARNICES; MÁSTIQUES; TINTAS
CAPÍTULO 33 – ACEITES ESENCIALES Y RESINOIDES; PREPARACIONES DE PERFUMERÍA, DE TOCADOR O DE COSMÉTICA
CAPÍTULO 34 – JABÓN, AGENTES DE SUPERFICIE ORGÁNICOS, PREPARACIONES PARA LAVAR, PREPARACIONES LUBRICANTES, CERAS ARTIFICIALES, CERAS PREPARADAS, PRODUCTOS DE LIMPIEZA, VELAS Y ARTÍCULOS SIMILARES, PASTAS PARA MODELAR, «CERAS PARA ODONTOLOGÍA» Y PREPARACIONES PARA ODONTOLOGÍA A BASE DE YESO FRAGUABLE
CAPÍTULO 35 – MATERIAS ALBUMINOIDEAS; PRODUCTOS A BASE DE ALMIDÓN O DE FÉCULA MODIFICADOS; COLAS; ENZIMAS
CAPÍTULO 36 – PÓLVORA Y EXPLOSIVOS; ARTÍCULOS DE PIROTECNIA; FÓSFOROS (CERILLAS); ALEACIONES PIROFÓRICAS; MATERIAS INFLAMABLES
CAPÍTULO 37 – PRODUCTOS FOTOGRÁFICOS O CINEMATOGRÁFICOS
CAPÍTULO 38 – PRODUCTOS DIVERSOS DE LAS INDUSTRIAS QUÍMICAS
CAPÍTULO 39 – PLÁSTICO Y SUS MANUFACTURAS
CAPÍTULO 40 – CAUCHO Y SUS MANUFACTURAS
CAPÍTULO 41 – PIELES (EXCEPTO LA PELETERÍA) Y CUEROS
CAPÍTULO 42 – MANUFACTURAS DE CUERO; ARTÍCULOS DE TALABARTERÍA O GUARNICIONERÍA; ARTÍCULOS DE VIAJE, BOLSOS DE MANO (CARTERAS) Y CONTINENTES SIMILARES; MANUFACTURAS DE TRIPA
CAPÍTULO 43 – PELETERÍA Y CONFECCIONES DE PELETERÍA; PELETERÍA FACTICIA O ARTIFICIAL
CAPÍTULO 44 – MADERA, CARBÓN VEGETAL Y MANUFACTURAS DE MADERA
CAPÍTULO 45 – CORCHO Y SUS MANUFACTURAS
CAPÍTULO 46 – MANUFACTURAS DE ESPARTERÍA O CESTERÍA

CAPÍTULO 47 – PASTA DE MADERA O DE LAS DEMÁS MATERIAS FIBROSAS CELULÓSICAS; PAPEL O CARTÓN PARA RECICLAR (DESPERDICIOS Y DESECHOS)
CAPÍTULO 48 – PAPEL Y CARTÓN; MANUFACTURAS DE PASTA DE CELULOSA, DE PAPEL O CARTÓN
CAPÍTULO 49 – PRODUCTOS EDITORIALES, DE LA PRENSA Y DE LAS DEMÁS INDUSTRIAS GRÁFICAS; TEXTOS MANUSCRITOS O MECANOGRAFIADOS Y PLANOS
CAPÍTULO 50 – SEDA
CAPÍTULO 51 – LANA Y PELO FINO U ORDINARIO; HILADOS Y TEJIDOS DE CRIN
CAPÍTULO 52 – ALGODÓN
CAPÍTULO 53 – LAS DEMÁS FIBRAS TEXTILES VEGETALES; HILADOS DE PAPEL Y TEJIDOS DE HILADOS DE PAPEL
CAPÍTULO 54 – FILAMENTOS SINTÉTICOS O ARTIFICIALES; TIRAS Y FORMAS SIMILARES DE MATERIA TEXTIL SINTÉTICA O ARTIFICIAL
CAPÍTULO 55 – FIBRAS SINTÉTICAS O ARTIFICIALES DISCONTINUAS
CAPÍTULO 56 – GUATA, FIELTRO Y TELA SIN TEJER; HILADOS ESPECIALES; CORDELES, CUERDAS Y CORDAJES; ARTÍCULOS DE CORDELERÍA
CAPÍTULO 57 – ALFOMBRAS Y DEMÁS REVESTIMIENTOS PARA EL SUELO, DE MATERIA TEXTIL
CAPÍTULO 58 – TEJIDOS ESPECIALES; SUPERFICIES TEXTILES CON MECHÓN INSERTADO; ENCAJES; TAPICERÍA; PASAMANERÍA; BORDADOS
CAPÍTULO 59 – TELAS IMPREGNADAS, RECUBIERTAS, REVESTIDAS O ESTRATIFICADAS; ARTÍCULOS TÉCNICOS DE MATERIA TEXTIL
CAPÍTULO 60 – TEJIDOS DE PUNTO
CAPÍTULO 61 – PRENDAS Y COMPLEMENTOS (ACCESORIOS), DE VESTIR, DE PUNTO
CAPÍTULO 62 – PRENDAS Y COMPLEMENTOS (ACCESORIOS), DE VESTIR, EXCEPTO LOS DE PUNTO
CAPÍTULO 63 – LOS DEMÁS ARTÍCULOS TEXTILES CONFECCIONADOS; JUEGOS; PRENDERÍA Y TRAPOS
CAPÍTULO 64 – CALZADO, POLAINAS Y ARTÍCULOS ANÁLOGOS; PARTES DE ESTOS ARTÍCULOS
CAPÍTULO 65 – SOMBREROS, DEMÁS TOCADOS, Y SUS PARTES
CAPÍTULO 66 – PARAGUAS, SOMBRILLAS, QUITASOLES, BASTONES, BASTONES ASIENTO, LÁTIGOS, FUSTAS, Y SUS PARTES
CAPÍTULO 67 – PLUMAS Y PLUMÓN PREPARADOS Y ARTÍCULOS DE PLUMAS O PLUMÓN; FLORES ARTIFICIALES; MANUFACTURAS DE CABELLO
CAPÍTULO 68 – MANUFACTURAS DE PIEDRA, YESO FRAGUABLE, CEMENTO, AMIANTO (ASBESTO), MICA O MATERIAS ANÁLOGAS
CAPÍTULO 69 – PRODUCTOS CERÁMICOS
CAPÍTULO 70 – VIDRIO Y SUS MANUFACTURAS

CAPÍTULO 71 – PERLAS FINAS (NATURALES) O CULTIVADAS, PIEDRAS PRECIOSAS O SEMIPRECIOSAS, METALES PRECIOSOS, CHAPADOS DE METAL PRECIOSO (PLAQUÉ) Y MANUFACTURAS DE ESTAS MATERIAS; BISUTERÍA; MONEDAS
CAPÍTULO 72 – FUNDICIÓN, HIERRO Y ACERO
CAPÍTULO 73 – MANUFACTURAS DE FUNDICIÓN, DE HIERRO O ACERO
CAPÍTULO 74 – COBRE Y SUS MANUFACTURAS
CAPÍTULO 75 – NÍQUEL Y SUS MANUFACTURAS
CAPÍTULO 76 – ALUMINIO Y SUS MANUFACTURAS
CAPÍTULO 78 – PLOMO Y SUS MANUFACTURAS
CAPÍTULO 79 – CINC Y SUS MANUFACTURAS
CAPÍTULO 80 – ESTAÑO Y SUS MANUFACTURAS
CAPÍTULO 81 – LOS DEMÁS METALES COMUNES; CERMETS; MANUFACTURAS DE ESTAS MATERIAS
CAPÍTULO 82 – HERRAMIENTAS Y ÚTILES, ARTÍCULOS DE CUCHILLERÍA Y CUBIERTOS DE MESA, DE METAL COMÚN; PARTES DE ESTOS ARTÍCULOS, DE METAL COMÚN
CAPÍTULO 83 – MANUFACTURAS DIVERSAS DE METAL COMÚN
CAPÍTULO 84 – REACTORES NUCLEARES, CALDERAS, MÁQUINAS, APARATOS Y ARTEFACTOS MECÁNICOS; PARTES DE ESTAS MÁQUINAS O APARATOS
CAPÍTULO 85 – MÁQUINAS, APARATOS Y MATERIAL ELÉCTRICO, Y SUS PARTES; APARATOS DE GRABACIÓN O REPRODUCCIÓN DE SONIDO, APARATOS DE GRABACIÓN O REPRODUCCIÓN DE IMAGEN Y SONIDO EN TELEVISIÓN, Y LAS PARTES Y ACCESORIOS DE ESTOS APARATOS
CAPÍTULO 86 – VEHÍCULOS Y MATERIAL PARA VÍAS FÉRREAS O SIMILARES, Y SUS PARTES; APARATOS MECÁNICOS, INCLUSO ELECTROMECAÑICOS, DE SEÑALIZACIÓN PARA VÍAS DE COMUNICACIÓN
CAPÍTULO 87 – VEHÍCULOS AUTOMÓVILES, TRACTORES, VELOCÍPEDOS Y DEMÁS VEHÍCULOS TERRESTRES, SUS PARTES Y ACCESORIOS
CAPÍTULO 88 – AERONAVES, VEHÍCULOS ESPACIALES, Y SUS PARTES
CAPÍTULO 89 – BARCOS Y DEMÁS ARTEFACTOS FLOTANTES
CAPÍTULO 90 – INSTRUMENTOS Y APARATOS DE ÓPTICA, FOTOGRAFÍA O CINEMATOGRAFÍA, DE MEDIDA, CONTROL O PRECISIÓN; INSTRUMENTOS Y APARATOS MEDICOQUIRÚRGICOS; PARTES Y ACCESORIOS DE ESTOS INSTRUMENTOS O APARATOS
CAPÍTULO 91 – APARATOS DE RELOJERÍA Y SUS PARTES
CAPÍTULO 92 – INSTRUMENTOS MUSICALES; SUS PARTES Y ACCESORIOS
CAPÍTULO 93 – ARMAS, MUNICIONES, Y SUS PARTES Y ACCESORIOS
CAPÍTULO 94 – MUEBLES; MOBILIARIO MEDICOQUIRÚRGICO; ARTÍCULOS DE CAMA Y SIMILARES; APARATOS DE ALUMBRADO NO EXPRESADOS NI COMPRENDIDOS EN OTRA PARTE; ANUNCIOS, LETREROS Y PLACAS INDICADORAS LUMINOSOS Y ARTÍCULOS SIMILARES; CONSTRUCCIONES PREFABRICADAS
CAPÍTULO 95 – JUGUETES, JUEGOS Y ARTÍCULOS PARA RECREO O DEPORTE; SUS PARTES Y ACCESORIOS

CAPÍTULO 96 – MANUFACTURAS DIVERSAS
CAPÍTULO 97 – OBJETOS DE ARTE O COLECCIÓN Y ANTIGÜEDADES
CAPÍTULO 98 – CONJUNTOS INDUSTRIALES