



**VNiVERSiDAD
D SALAMANCA**



UNIVERSIDAD DE SALAMANCA

Departamento de Estadística

**Máster en Análisis Avanzado de Datos Multivariantes y Big
Data**

Trabajo Fin de Máster

**Análisis multivariante espacio-temporal de
variables agrarias y socioeconómicas
vinculadas al riesgo de incendio en España
peninsular**

Autora: Marta Cuerva del Álamo

Tutora: María José Fernández Gómez

Co-tutora: María Olga Viedma Sillero

2024

D.^a “*María José Fernández Gómez*”, profesora del Departamento de Estadística de la Universidad de Salamanca,

D.^a “*María Olga Viedma Sillero*”, profesora del Departamento de Ecología de la Universidad de Castilla-La Mancha,

HACEN CONSTAR:

Que el trabajo titulado “*Análisis multivariante espacio-temporal de variables agrarias y socio-económicas vinculadas al riesgo de incendio en España peninsular*”, que se presenta, ha sido realizado por “*Marta Cuerva del Álamo*” y constituye la memoria del trabajo realizado para la superación de la asignatura Trabajo de Fin de Máster en esta Universidad.

Salamanca, a 15 de septiembre de 2024



VIEDMA
SILLERO
MARIA
OLGA -
21487132A

Firmado
digitalmente por
VIEDMA SILLERO
MARIA OLGA -
21487132A
Fecha:
2024.09.16
09:30:06 +02'00'



Fdo.: M.^a José Fernández Gómez Fdo.: M.^a Olga Viedma Sillero Fdo.: Marta Cuerva del Álamo

Resumen

Los incendios forestales están condicionados por factores climáticos y naturales, pero las actividades humanas son responsables del 90 % de las igniciones actuales en los bosques europeos. La pirogeografía emerge como una herramienta fundamental para el desarrollo de estrategias de gestión del fuego adaptadas a las características específicas de cada región. Este estudio explora la evolución de la estructura de covariación entre variables agrarias y socioeconómicas desde 1982 hasta 2020, utilizando métodos estadísticos multivariantes como STATIS-Dual, análisis de Co-inercia y el modelo de Regresión PLS.

El análisis STATIS-Dual reveló una notable estabilidad en las variables agrarias a lo largo de los períodos 1982, 1989, 1999 y 2009. Sin embargo, se detectó un cambio significativo en 2020, atribuido a los efectos disruptivos de la pandemia de COVID-19. El análisis de Co-inercia entre las variables agrarias y socioeconómicas de los años 2009 y 2020, mostró una correlación moderada, con una concentración significativa de variabilidad en un único eje, indicando una asociación limitada entre los datos agrarios y socio-económicos durante este período. Estos análisis se complementaron mediante sendos análisis de clusters, que permitieron caracterizar cartográficamente las trayectorias de los municipios durante en los periodos estudiados. El modelo PLSR destacó que las variables asociadas con superficies agrícolas pequeñas y medianas son esenciales para la predicción del régimen de incendios, aunque la capacidad predictiva del modelo fue moderada.

Estos resultados resaltan la relevancia de analizar las particularidades agrarias para obtener una comprensión más detallada y eficiente del riesgo de incendios en distintos contextos locales y regionales.

Palabras clave: Pirogeografía; estadística multivariante; clustering; variables agrarias; variables socioeconómicas; península Ibérica

Summary

Forest fires are influenced by climatic and natural factors, but human activities are responsible for 90 % of current ignitions in European forests. Pyrogeography emerges as a fundamental tool for developing fire management strategies adapted to the specific characteristics of each region. This study explores the evolution of the covariation structure between agricultural and socioeconomic variables from 1982 to 2020, using multivariate statistical methods such as STATIS-Dual, Co-inertia analysis, and the PLS regression model.

The STATIS-Dual analysis revealed notable stability in agricultural variables across the periods of 1982, 1989, 1999, and 2009. However, a significant change was detected in 2020, attributed to the disruptive effects of the COVID-19 pandemic. The Co-inertia analysis between the agricultural and socioeconomic variables from 2009 and 2020 showed a moderate correlation, with a significant concentration of variability in a single axis, indicating a limited association between agricultural and socioeconomic data during this period. These analyses were complemented by cluster analyses, which allowed the cartographic characterization of municipal trajectories over the studied periods. The PLSR model highlighted that variables associated with small and medium-sized agricultural areas are essential for predicting fire regimes, although the predictive capacity of the model was moderate.

These results underscore the importance of analyzing specific agricultural characteristics to gain a more detailed and efficient understanding of fire risk in different local and regional contexts.

Keywords: Pyrogeography; multivariate statistics; clustering; agrarian data; socioeconomic data; Iberian Peninsula

Índice

Resumen	2
Summary	3
1. Introducción	9
Justificación	13
Objetivos	13
2. Material y métodos	14
2.1. Área de estudio	14
2.2. Variables	19
2.2.1. Pre-procesamiento de las variables agrarias	22
2.2.2. Pre-procesamiento de las variables socioeconómicas	24
2.2.3. Obtención de las variables finales	25
2.2.3.1. Variables agrarias	25
2.2.3.2. Variables socioeconómicas	29
2.2.4. Unión de datos geospaciales con la base de datos	31
2.2. Estadística	34
2.2.1. Estadística descriptiva	34
2.2.2. Estadística Multivariante	36
STATIS-Dual	36
Etapas 1, 2 y 3	37
Etapas 4 y 5	38
Etapas 6 y 7	39
Análisis de Co-Inercia	40
Análisis de Componentes Principales	41
Definición del CoIA	41
Inercia y Coinercia	41
Clustering	42
Clustering K-means	42
Elbow Method o Método del codo	43
Silhouette Method o el Método de la Silueta	44

Regresión Parcial por Mínimos Cuadrados (PLSR)	45
3. Resultados	48
3.1. Estadística descriptiva	48
3.1.1. Diagrama de cajas y bigotes	52
3.1.2. Gráfico de densidad	59
3.1.3. Análisis de correlación entre variables	64
3.2. Estadística multivariante	68
3.2.1. STATIS-Dual	68
3.2.2. Co-Inercia	75
3.2.3. Clustering K-means	81
Cluster K-means en STATIS-Dual	83
Cluster K-means en Co-inercia	87
3.2.4. Regresión Parcial por Mínimos Cuadrados (PLSR)	91
4. Conclusiones	97
Bibliografía	99
Anexo	105
Scripts de RStudio	105
Cálculo de las variables para el año 1989	105
Cálculo de las variables para el año 2020	113
Estadística descriptiva de las variables socioeconómicas y agrarias	121
Cálculo del STATIS-Dual de las bases de datos de los años 1982-89-99-09-20	130
Cálculo de la Co-inercia de las bases de datos del 2009 y 2020	138
Cálculo del número óptimo de k-clusters	141
Cálculo del PLSR	144

Índice de tablas

1	Clasificación de variables y autores del marco teórico.	11
2	Clasificación climática de Köppen.	16
3	Enlaces de las fuentes de datos de las variables utilizadas.	21
4	Descripción de las variables agrarias finales.	27
5	Descripción de las variables socioeconómicas finales.	31
6	Descripción de las columnas presentes en la capa de atributos del shape de municipios.	33
7	Resumen estadístico de las variables agrarias.	49
8	Resumen estadístico de las variables socioeconómicas.	51
9	Resultados estadísticos de la correlación del censo agrario.	66
10	Matriz de correlaciones entre variables socioeconómicas.	67
11	Coeficientes de correlación vectorial del STATIS-Dual de los años 1982-89-99-09-20.	70
12	Coseno cuadrado del STATIS-Dual de los años 1982-89-99-09-20.	70
13	Pesos para las k-tablas del STATIS-Dual de los años 1982-89-99-09-20.	70
14	Valores propios para la inter-estructura del STATIS-Dual de los años 1982-89-99-09-20.	70
15	Valores propios para el compromiso del STATIS-Dual de los años 1982-89-99-09-20.	71
16	Coeficientes de correlación vectorial del STATIS-Dual de los años 1982-89-99-09.	74
17	Coseno cuadrado del STATIS-Dual de los años 1982-89-99-09.	74
18	Pesos para las k-tablas del STATIS-Dual de los años 1982-89-99-09.	74
19	Valores propios para la inter-estructura del STATIS-Dual de los años 1982-89-99-09.	74
20	Valores propios para el compromiso del STATIS-Dual de los años 1982-89-99-09.	75
21	Resultados estadísticos de la co-inercia del 2009.	77
22	Resultados estadísticos de la co-inercia del 2020.	80
23	Media de variables por cluster del STATIS-Dual.	87
24	Media de variables por cluster de la Co-inercia.	91

Índice de figuras

1	Mapa del área de estudio.	14
2	Clasificación climática en la península Ibérica.	15
3	Vegetación potencial de la península Ibérica.	18
4	Resumen de las variables estudiadas: agrarias y socioeconómicas.	22
5	Flujo del proceso de análisis de datos agrarios.	29
6	Procedimientos del STATIS-Dual.	37
7	Interestructura de las tablas en el STATIS-Dual.	38
8	Esquema del análisis de Co-inercia.	42
9	Flujo del algoritmo K-means.	43
10	Diagrama de flujo del Método del Codo.	44
11	Procedimientos del PLSR.	45
12	Box-plot de las superficies agrarias del censo agrario.	52
13	Box-plot de la edad y formación de los jefes del censo agrario.	54
14	Box-plot del régimen de tenencia del censo agrario.	55
15	Box-plot de la densidad de explotaciones y ganadadera del censo agrario.	56
16	Box-plot de los afiliados a la Seguridad Social de la socioeconomía.	57
17	Box-plot de la deuda pública de los Ayuntamientos para la socioeconomía.	57
18	Box-plot de la Tasa de Crecimiento Vegetativo para la socioeconomía.	58
19	Box-plot de la deuda pública de los Ayuntamientos para la socioeconomía.	59
20	Diagrama de densidad de las superficies agrarias del censo agrario.	60
21	Diagrama de densidad de la edad y formación del jefe de explotación del censo agrario.	61
22	Diagrama de densidad de los regímenes de tenencia del censo agrario.	61
23	Diagrama de densidad de afiliación a la Seguridad Social.	62
24	Diagrama de densidad de la deuda pública en la socioeconomía.	63
25	Diagrama de densidad de la Tasa de Crecimiento Vegetativo.	63
26	Diagrama de densidad del paro en la socioeconomía.	64
27	Mapa de calor de la correlación del censo agrario.	65
28	Mapa de calor de la correlación socioeconómica.	67
29	Matrices estudiadas en el STATIS-Dual.	68

30	Gráfico del STATIS-Dual para el censo agrario en los años 1982-89-99-09-20.	69
31	Gráfico del STATIS-Dual para el censo agrario en los años 1982-89-99-09.	73
32	Matrices estudiadas en la Co-inercia.	76
33	Gráfico resultante del análisis de co-inercia del 2009.	78
34	Gráfico resultante del análisis de co-inercia del 2020.	81
35	Gráficos del Método del Codo para el censo agrario.	82
36	Gráficos del Método de la Silueta para el censo agrario.	83
37	Mapa del clustering K-means con resultados del STATIS-Dual.	86
38	Mapa del clustering K-means con resultados de la Co-inercia.	90
39	Mapa con los incendios totales a nivel municipal.	92
40	Procedimientos del PLSR en el estudio.	93
41	Gráfico MSEP de la PLSR.	95
42	Gráfico de los coeficientes de la PLSR.	96
43	Gráfico de barras con los coeficientes del análisis PLSR.	96

1. Introducción

La “piro-geografía” constituye una línea de investigación de gran relevancia en el estudio de los incendios forestales, centrándose en la definición, delimitación y caracterización de unidades espaciales homogéneas en términos de riesgo de incendio (Morgan et al., 2001; Galizia et al., 2022; Rodrigues et al., 2020; Curt et al., 2014; Alcasena et al., 2019). Hasta el momento, en la piro-geografía predominan dos enfoques principales. El primero se basa exclusivamente en las estadísticas de incendios para definir “pirorregiones”, utilizando únicamente variables propias del régimen de incendios para segmentar o identificar regiones homogéneas (Bradstock, 2010; Archibald et al., 2013; Moreno et al., 2014; Jiménez-Ruano, 2015; 2019). El segundo enfoque, más reciente, integra factores explicativos del régimen de incendios como la vegetación, el clima, la topografía y factores humanos (Curt et al., 2014; Alcasena et al., 2019; Syphard y Keeley, 2020; Galizia et al., 2022). Este enfoque permite agrupar unidades espaciales que reflejan una amplia gama de características ecosistémicas que, combinadas, se manifiestan en patrones espaciales y temporales diferenciados de incendios forestales (Syphard y Keeley, 2020). Cuando estas variables biofísicas incluyen aspectos como ecorregiones, condiciones climáticas y composición vegetal, se utilizan los términos “eco-pirorregión” (Rodrigues et al., 2019; Curt et al., 2014; Gallant et al., 2004; Blasi et al., 2014). Por otro lado, el término “socio-eco-pirorregión” se aplica cuando se incorporan variables socioeconómicas para explicar el régimen de incendios. Entre las variables socioeconómicas más utilizadas figuran la densidad de carreteras, la proximidad a núcleos poblacionales densos, la densidad de interfaces urbano-forestales (WUI: Wildland Urban Interfaces) y el porcentaje de cambio en el uso del suelo (Archibald et al., 2013; Clarke et al., 2019).

Diversos estudios han identificado variables socioeconómicas y ambientales clave en relación con el régimen y la frecuencia de incendios (**Tabla 1**):

- En el ámbito de la **demografía y la población**, el envejecimiento ha sido identificado como un factor relevante por autores como De Diego et al. (2022), Chuvieco et al. (2023), Carlucci et al. (2019), Bar-Massada et al. (2023) y Alcasena et al. (2021). Además, variables como el número de extranjeros y el equilibrio vegetativo han sido analizadas por De Diego et al. (2022) y Carlucci et al. (2019), quienes destacan su posible impacto en la dinámica de los incendios forestales. De Diego et al. (2022) señalan que la atracción

hacia paisajes inflamables, ya sea por el valor visual de los entornos naturales o el bajo costo del terreno, incrementa el riesgo de incendios, especialmente en áreas con hoteles rurales, casas aisladas y zonas de interfaz urbano-forestal (WUIs). Además, consideran la presencia de individuos con tendencias sociopáticas o criminales como un factor de ignición significativo.

- En el ámbito de la **economía y el empleo**, De Diego et al. (2022) han explorado la influencia de variables socioeconómicas como la tasa de jubilación, la tasa de desempleo, la renta per cápita y el Producto Interior Bruto (PIB), un área en la que también han trabajado Chuvieco et al. (2023) y Bar-Massada et al. (2023). Carlucci et al. (2019) han abordado las importaciones y exportaciones, mientras que Wunder et al. (2021) se ha centrado en el impacto del turismo en los incendios.
- En cuanto a la **accesibilidad**, la presencia de infraestructuras como líneas de tren y carreteras ha sido destacada por autores como Bowman et al. (2013), Krawchuk et al. (2009), Ruano (2019), Chuvieco et al. (2023), Carlucci et al. (2019), Bar-Massada et al. (2023) y Rodrigues et al. (2022). Bowman et al. (2013) discuten cómo las actividades humanas, incluidas las igniciones químicas y eléctricas, los incendios agrícolas, y la expansión de infraestructura; como líneas de tren y de electricidad, incrementan la accesibilidad a áreas remotas y, por ende, el riesgo de incendios forestales. La investigación de Krawchuk et al. (2009) indica que la alta densidad de población; que demanda una mayor cantidad de tierras agrícolas, junto con la tendencia a reducir la explotación agrícola; posiblemente debido al surgimiento de macro-granjas, podrían estar relacionadas con el incremento en la importación de frutas y verduras. Este fenómeno, a su vez, podría contribuir al abandono de tierras agrícolas.
- Desde una perspectiva **ambiental**, la presencia de áreas protegidas y las interfaces urbano-forestales (WUI) han sido objeto de evaluación por Ruano (2019), Bar-Massada et al. (2023), Chuvieco et al. (2023), Wunder et al. (2021) y Rodrigues et al. (2022). Rodrigues et al. (2022) también han investigado las interfaces agrarias (WAI), subrayando la complejidad de los incendios en estas zonas.
- En cuanto a las **igniciones**, tanto accidentales como intencionadas, los fuegos agrícolas y las causas químicas y eléctricas han sido objeto de análisis por Bowman et al. (2013), Wunder et al. (2021), Krawchuk et al. (2009), Ruano (2019) y Rodrigues et al. (2022),

quienes destacan cómo estas causas antropogénicas contribuyen significativamente al inicio y propagación de incendios.

- Finalmente, en términos de **eficacia** contra incendios, la capacidad de supresión ha sido estudiada por Curt et al. (2014), Chuvieco et al. (2008) y Alcasena et al. (2021). Curt et al. (2014) subrayan la importancia de factores como las inversiones económicas, el número de dispositivos de supresión, y la ubicación estratégica de estaciones de bomberos. Estos autores han desarrollado un índice espacial para evaluar la efectividad de la supresión de incendios en diferentes paisajes. Asimismo, Chuvieco et al. (2008) enfatizan el impacto significativo de las actividades humanas en los regímenes de incendios a través del índice de Huella Humana (*Human Footprint Index*), que mide la presión poblacional, el uso de la tierra, la infraestructura y el acceso, correlacionando estos factores con la ocurrencia de incendios. Por otro lado, Alcasena et al. (2017) y Alcasena et al. (2021) han analizado la probabilidad de ignición y han realizado análisis de exposición para entender mejor la dinámica y riesgos asociados a los incendios forestales.

No obstante, un factor menos estudiado pero de gran importancia en el régimen de incendios mediterráneos, es la actividad agraria (Bowman et al., 2013; Roos et al., 2014; Viedma et al., 2018; Rodrigues et al., 2018). Junto con este factor y las variables socioeconómicas a nivel municipal de la Península, se pudo llevar a cabo esta investigación.

Tabla 1: Clasificación de variables y autores, proporcionando una visión exhaustiva del marco teórico y empírico en relación con el régimen de incendios. Elaboración propia.

Clasificación	Variables	Autor
Demografía y población	Envejecimiento de la población	de Diego et al., 2022; Chuvieco et al., 2023; Carlucci et al., 2019; Bar-Massada et al., 2023; Alcasena et al., 2021
	Número de extranjeros	de Diego et al., 2022; Carlucci et al., 2019
	Equilibrio vegetativo	de Diego et al., 2022; Carlucci et al., 2019

Continúa en la siguiente página

Continúa de la página anterior

Clasificación	VARIABLES	Autor
Economía y empleo	Tasa de jubilación	de Diego et al., 2022
	Tasa de desempleo	de Diego et al., 2022
	Renta per cápita	de Diego et al., 2022; Alcasena et al., 2021
	Producto Interior Bruto	de Diego et al., 2022; Chuvieco et al., 2023; Bar-Massada et al., 2023
	Importaciones	Carlucci et al., 2019
	Exportaciones	Carlucci et al., 2019
	Turismo	Wunder et al., 2021
Accesibilidad	Líneas de tren y transporte	Bowman et al., 2013; Krawchuk et al., 2009; Ruano, 2019; Chuvieco et al., 2023; Carlucci et al., 2019; Bar-Massada et al., 2023; Rodrigues et al., 2022
Ambiente	Áreas protegidas	Ruano, 2019; Bar-Massada et al., 2023
	Wild-Urban Interfaces	Ruano, 2019; Chuvieco et al., 2023; Wunder et al., 2021; Rodrigues et al., 2022
	Wild-Agrarian Interfaces	Rodrigues et al., 2022
Igniciones intencionadas y accidentales	Fuegos agrícolas	Bowman et al., 2013; Wunder et al., 2021
	Químicas y eléctricas	Bowman et al., 2013; Krawchuk et al., 2009; Ruano, 2019; Wunder et al., 2021; Rodrigues et al., 2022
Eficacia contra incendios	Capacidad de supresión de incendios	Curt et al., 2014; Chuvieco et al., 2008; Alcasena et al., 2021
	Probabilidad de ignición	Alcasena et al., 2017
	Análisis de exposición	Alcasena et al., 2021

Justificación

Aunque el régimen de incendios a nivel global y regional está fuertemente influenciado por el clima y otras variables naturales (Pechony y Shindell, 2010; Abatzoglou y Williams, 2016), las actividades humanas; especialmente en regiones donde la intervención es más intensa (Amatulli et al., 2006; Archibald et al., 2010) son responsables del 90 % de las igniciones de incendios en los bosques europeos desde la década de los 70 (FAO, 2007).

La necesidad de este estudio radica en la importancia de comprender las posibles relaciones entre las variables agrarias y socioeconómicas y su influencia en el régimen de incendios en la España peninsular. La creciente variabilidad climática y los cambios en el uso de la tierra han generado una presión adicional sobre los sistemas agrarios y forestales, lo que hace indispensable un análisis profundo de estas interacciones para el desarrollo de estrategias de gestión más efectivas y sostenibles.

Por tanto, se debe analizar conjuntamente estas variables en relación con los incendios utilizando un enfoque multivariante.

Objetivos

El objetivo principal es la caracterización en zonas geográficas homogéneas y diferenciadas en cuanto a variables agrarias y socioeconómicas dentro de la España peninsular en los años 1982, 1989, 1999, 2009 y 2020, para así establecer relaciones variables-régimen de incendios.

A través de esta caracterización, se busca establecer relaciones claras entre estas variables y el régimen de incendios, permitiendo así una mejor comprensión de los factores que pueden estar impulsando los patrones de incendios en diferentes regiones.

Para alcanzar este objetivo, se han llevado a cabo varios análisis multivariantes:

- **Análisis STATIS-Dual:** este análisis se utilizó para explorar la evolución de las variables agrarias en los periodos 1982, 1989, 1999, 2009 y 2020. El enfoque STATIS-Dual permitió comparar la estructura de datos agrarios en estos diferentes años y detectar patrones o cambios significativos a lo largo del tiempo, lo que es esencial para identificar tendencias agrarias.

- **Análisis de Coinercia:** se aplicó este análisis para examinar la relación conjunta entre las variables socioeconómicas y agrarias en los años 2009 y 2020. Este enfoque proporcionó una visión más profunda de cómo se relacionan entre sí estas dos dimensiones.
- **Análisis de Regresión Parcial por Mínimos Cuadrados (PLSR):** este método se utilizó para modelar la influencia combinada de las variables agrarias y socioeconómicas en el régimen de incendios. El PLSR permite identificar los factores que tienen un mayor peso en la ocurrencia y comportamiento de los incendios, integrando múltiples variables en un único modelo predictivo.

2. Material y métodos

2.1. Área de estudio

El área de estudio está configurado por los municipios de la España peninsular; siendo alrededor de 8.100 (**Figura 1**), los cuales han sufrido importantes variaciones a lo largo del tiempo.

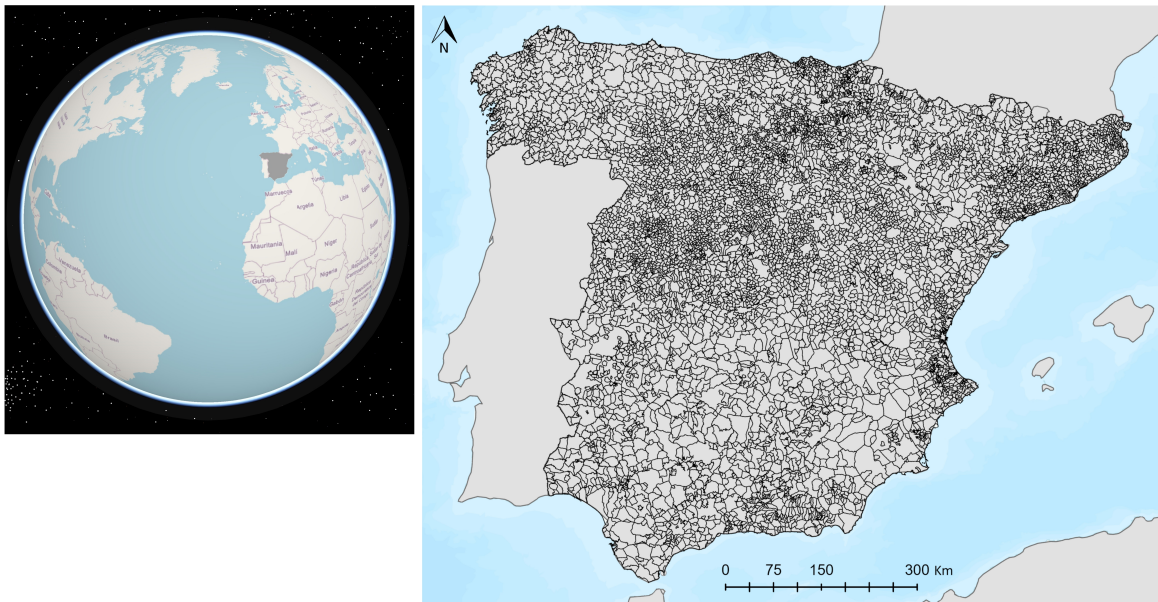


Figura 1: Situación de los municipios de la península Ibérica sobre el mundo. Capa municipal extraída del Centro de Descargas del IGN (Instituto Geológico Nacional). Datum: WGS84; sistema de coordenadas: UTM. Elaboración propia a partir de ArcGis Pro.

Este territorio está acotado aproximadamente entre los paralelos 36" y 44° lat. Norte, con un área de unos 50 millones de hectáreas y con un rango climático, topográfico y biogeográfico muy amplio (Jiménez-Ruano et al., 2015). La posición latitudinal pone en contacto a dos zonas climáticamente muy diferenciadas: la atlántica y la mediterránea. Según la clasificación de Köppen, la mayoría del centro-sur y las regiones de costa mediterráneas (40% del territorio) poseen el tipo Csa; clima con veranos secos o cálidos, con excepciones en algunas zonas localizadas en el suroeste que son BWh; clima desértico cálido. La región del centro-norte, mantiene el tipo Cfa; que es un clima con estación seca y veranos cálidos, y las demás montañosas, poseen tipos Dfb-c y Dsb-f; climas más fríos (MITECO, 2012; Castro et al., 2005) (**Figura 2, Tabla 2**). Entre estas dos zonas climáticas se encuentra una transición a la que se denomina zona subatlántica, donde el clima es atlántico pero “mediterraneizado” y viceversa (MITECO, 2012).

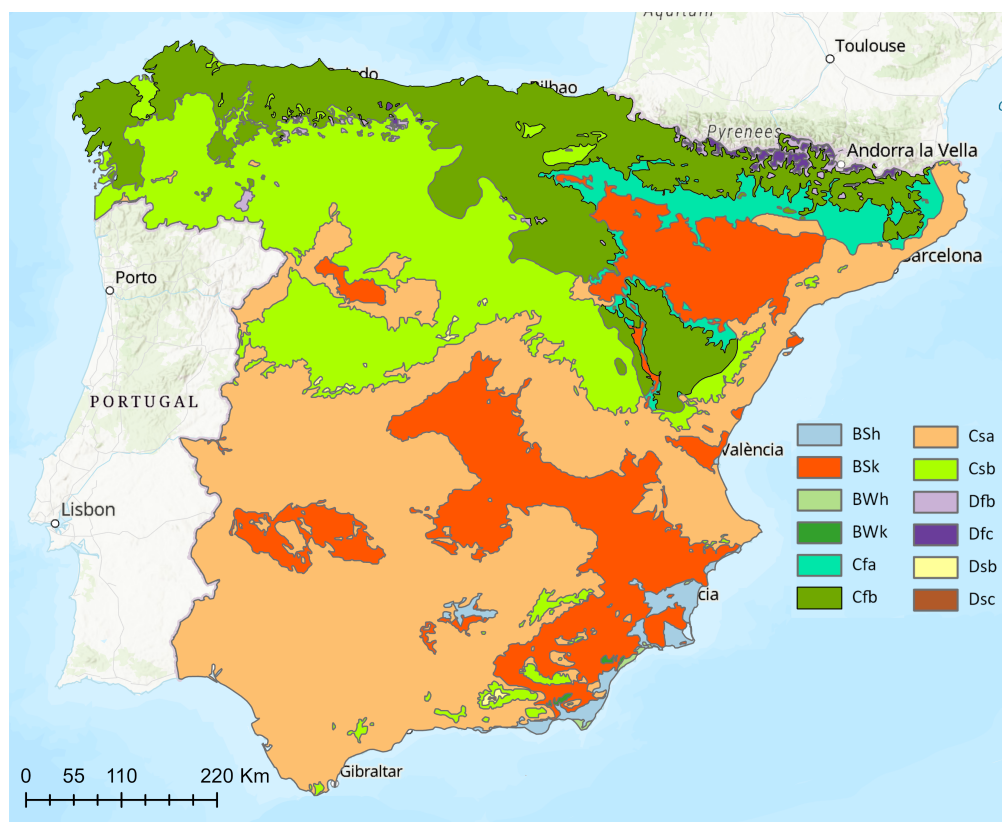


Figura 2: Mapa de los climas según la clasificación de Köppen para la península Ibérica. Capa extraída del Centro de Descargas del IGN (Instituto Geológico Nacional). Datum: WGS84; sistema de coordenadas: UTM. Elaboración propia a partir de ArcGis Pro.

Tabla 2: Clasificación climática de Köppen. Elaborado a partir de Chen y Chen, 2013.

Grupo climático	Símbolo	Tipo de clima	Región natural
A	Af	Cálido con lluvias todo el año	Selva o bosque tropical
	Am	Cálido con lluvias de monzón	Selva o bosque tropical
	Aw	Cálido con lluvias en verano	Sabana
B	BS	Seco estepario	Estepa
	BW	Seco desértico	Desierto
C	Cf	Templado con lluvias todo el año	Bosque mixto
	Cw	Templado con lluvias en verano	Pradera
	Cs	Templado con lluvias en invierno	Bosque mediterráneo
D	Df	Frío con lluvias todo el año	Bosque de coníferas
	Dw	Frío con lluvias en verano	Bosque de coníferas
E	ET	Polar de tundra	Tundra
	EF	Polar de hielos perpetuos	Hielos perpetuos
	EB	Polar de alta montaña	Hielos perpetuos

En cuanto a la litología, se pueden distinguir tres unidades básicas: zona silíceo-caliza y arcillosa. La primera, se compone por las montañas más antiguas y está situada por el oeste de la península, abarcando la región de Galicia, Asturias, una gran parte de Castilla y León, Castilla-La Mancha, Andalucía, Cantabria y la totalidad de Extremadura. La segunda, está constituida por rocas menos duras y montañas más jóvenes, siendo frecuente en la mitad

oriental de España, sobre todo en las cordilleras Béticas, Ibérica, Pirineos y Cantábrica. Y en la tercera, se incluyen las rocas más blandas y sedimentarias, encontrándose en las grandes depresiones del Ebro y Guadalquivir, a la vez que en las cuencas sedimentarias de la submeseta Central Norte y Central Sur (Martín, 1992).

El relieve, influenciado por la litología, se caracteriza por la forma maciza que posee, ya que ninguna masa de agua salada penetra hacia el interior, ocasionando a su vez una diferenciación climática en la costa y en el interior. En la Meseta, que supera los 600 metros de altitud media y que ocupa gran parte del interior, se disponen alrededor las principales unidades de relieve, lo que dificulta la entrada de aire húmedo hacia el interior. Al mismo tiempo, ya que la inclinación de la Meseta se dirige hacia el océano Atlántico, muchos grandes ríos desembocan en la zona noroeste (Martín, 1992).

La vegetación varía en función de los tres factores anteriores (**Figura 3**), por lo que se reparte en diferentes regiones biogeográficas. En la Península, encontramos a la región eurosiberiana y mediterránea. La vegetación eurosiberiana se caracteriza por adaptaciones a temperaturas bajas y precipitaciones altas, como por ejemplo el abeto (*Abies alba*), pino negro (*Pinus nigra*), enebro (*Juniperus communis*), haya (*Fagus sylvatica*) y abedul (*Betula pendula*). Se sitúan en gran parte de Galicia, la cornisa Cantábrica y los Pirineos. La vegetación mediterránea o esclerófila, soporta altas temperaturas y sequías severas. Ejemplos de esta vegetación son muchos de los *Quercus*: alcornoque (*Quercus suber*) o encina (*Quercus ilex*), y matorrales como las retamas (*Retama sphaerocarpa*) (Rodriguez et al., 2019). La zona noroeste se cubre aproximadamente de un 68% de bosques, arbustos o pastizales, mientras que en la zona mediterránea un 54% es zona silvestre, habiendo sufrido un abandono agrícola progresivo. En la costa mediterránea tan solo un 22% es masa forestal debido a la presencia de una gran interfaz forestal-urbana (WUI) (Moreno et al., 2014).

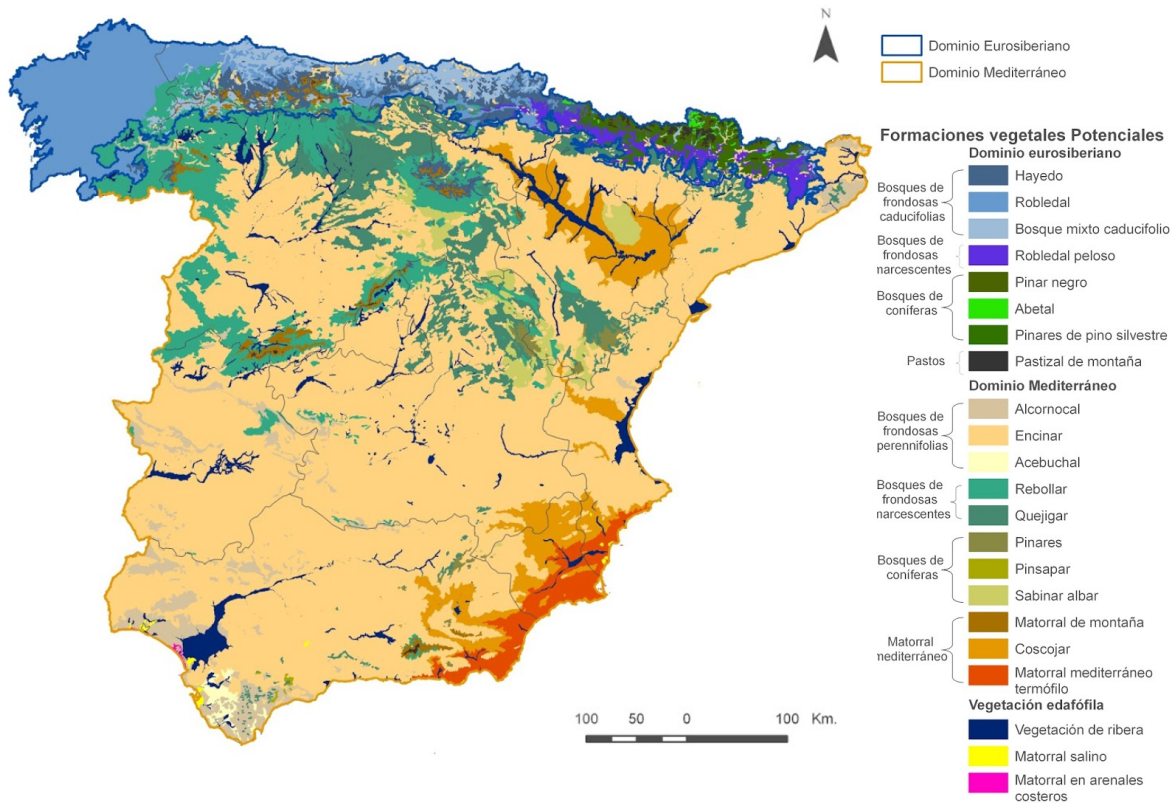


Figura 3: Mapa según la vegetación potencial en la península Ibérica. Generada a partir de la capa “Series de Vegetación de España” del Centro de Descargas del IGN (Instituto Geológico Nacional). Datum: WGS84; sistema de coordenadas: UTM. Departamento de Geografía y Ordenación del Territorio de la Universidad de Zaragoza.

En lo que se refiere a la estructura poblacional, según el Instituto Nacional de Estadística (INE), la población en España aumentó de 35,4 millones en 1982 a 47,4 millones en 2022, concentrándose en zonas urbanas. Se ha observado un envejecimiento demográfico y una disminución en la tasa de natalidad, pasando de una esperanza de vida de 72 años en 1982 a 82,2 años en 2020. Esta estructura demográfica ha ido cambiando a lo largo de los años debido a la evolución socioeconómica que ha sufrido el territorio.

Respecto a la estructura socioeconómica del país, en 1979 el comercio interior representaba un sector crucial, con un 12,4% de la población activa empleada; destacándose por una alta participación femenina (42%). A pesar de las distintas crisis económicas sufridas a lo largo de estos 40 años, este sector ha mantenido una gran estabilidad laboral, consolidándose como un pilar importante del mercado de trabajo. Por otra parte, entre 1980 y 2018 el sector financiero fue el más dinámico en términos de Valor Añadido Bruto (VAB), con un

incremento del 161,29%, seguido por la industria de la automoción, que creció un 77,23%. En contraste, sectores como la agricultura y la industria alimentaria experimentaron una disminución significativa en su peso relativo, aunque siguen siendo relevantes en el contexto europeo. Durante los años 90, la industria turística española se enfrentó a la competencia de nuevos destinos en el Mediterráneo y el Caribe, lo que motivó una mejora en la calidad de su oferta y una intensificación de la promoción en el extranjero. A partir de 1996, el turismo en España vivió un período de crecimiento sostenido, diversificándose en áreas como el turismo de congresos, salud, bienestar, deportivo y ambiental. Entre 1995 y 2020, el sector servicios lideró el crecimiento del empleo, con un aumento del 99,48%, seguido por el sector energético y el turismo. En este período, la internacionalización de la economía española fue constante tras la entrada de España en la Unión Europea en 1986. Ésto provocó un aumento de las exportaciones, que representaron el 41% del PIB en 2022. Finalmente, en este mismo año, el sector primario fue el único en aumentar su PIB en comparación con 2019, en un contexto en el que los sectores industrial y de servicios aún no habían recuperado los niveles previos a la pandemia de COVID-19 (Marín y Mínguez, 2023).

Es fundamental comprender la estructura física de un territorio, incluyendo el clima, la litología y la vegetación, así como el contexto histórico y político, para interpretar adecuadamente su estructura agraria y el nivel socioeconómico.

2.2. Variables

Este estudio se enfoca en un análisis detallado de las variables agrarias de la península Ibérica, considerando los años 1982, 1989, 1999, 2009 y 2020, así como las variables socioeconómicas correspondientes a los años 2009 y 2020. La selección de estas variables es fundamental para la comprensión del régimen de incendios en contextos de países industrializados, donde las transformaciones en el uso del suelo y las dinámicas socioeconómicas desempeñan un papel determinante en la incidencia y comportamiento de los incendios forestales.

Para llevar a cabo este análisis, se ha recurrido a diversas fuentes de datos (**Tabla 3**), que incluyen microdatos y bases de datos en formato .xlsx. Los microdatos, definidos como aquellos que poseen una dimensión adecuada para albergar datos precisos, actualizados y útiles para la realización de estudios estadísticos, son de particular relevancia en este contexto. Estos microdatos pueden clasificarse en datos estructurados, no estructurados y semiestruc-

turados, de acuerdo con su formato (Dupriez y Boyko, 2010; Haşiloğlu y Bali, 2017). En este trabajo, los microdatos utilizados son de tipo ASCII (*American Standard Code for Information Interchange*), los cuales se presentan en un formato estructurado, organizados en tablas o bases de datos. En estas tablas, cada fila representa una unidad de dato y cada columna una variable que describe una característica específica de dicha unidad. Estos datos se distribuyen en formato de texto plano (.txt), y contienen varios tipos de registros que agrupan diferentes variables. Todos los niveles o tipos de registro comparten una serie de variables comunes, como el código de provincia y el código de municipio, lo que permitió la integración de los datos.

Tabla 3: Fuentes de datos relacionadas con variables socioeconómicas, demográficas y agrarias.

Categoría	Enlace
Censos poblacionales de los años 1981-2009	Resultados detallados: Censo de Población y Viviendas 1981, 1991, 2001, 2009
Censo anual de población 2020	Censo 2020
Búsqueda de municipios	Búsqueda de municipios por código
Censos agrarios	Censos agrarios
Tasa de paro por sectores, edad y sexo	Tasa de paro por sectores, edad y sexo
Distribución porcentual ocupados	Distribución porcentual de los ocupados por sector económico
Empresas por municipio y actividad	Empresas por municipio y actividad principal
Nacimientos por municipio	Nacimientos por municipio
Defunciones por municipio	Defunciones por municipio
Afiliaciones seguridad social 2009	Afiliaciones seguridad social 2009
Afiliaciones seguridad social 2020	Afiliaciones seguridad social 2020
Deuda Ayuntamientos	Deuda Ayuntamientos
Incendios por municipio 2006-2015	Incendios por municipio de 2006 a 2015

Entre las variables en formato de microdatos se incluyen aquellas provenientes del censo agrario, así como datos de nacimientos y defunciones, todos proporcionados por el Instituto Nacional de Estadística (INE). En el caso específico del censo agrario, los registros se clasifican en tres tipos: Tipo 1, Tipo 2 y Tipo 3.

Por otro lado, las variables en formato .xlsx incluyen datos sobre afiliados a la Seguridad Social por sectores, deuda pública de los Ayuntamientos y tasas de desempleo desglosadas por sexo, edad y sectores, todos ellos pertenecientes al grupo socioeconómico. La **Figura 4** presenta un resumen de todas las variables analizadas en este estudio, distinguiendo entre variables agrarias y socioeconómicas, proporcionando así una visión clara de los diferentes aspectos evaluados en relación con el régimen de incendios en la península Ibérica.

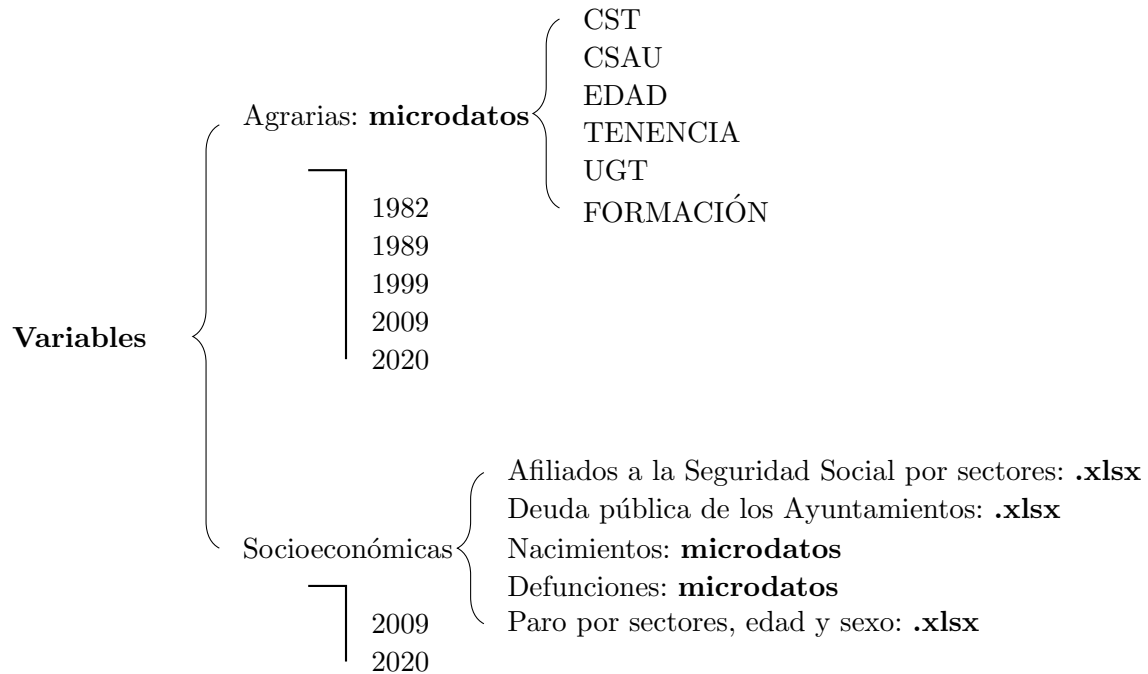


Figura 4: Diagrama de las variables agrarias y socioeconómicas con el año estudiado, la tipología de éstas y las subvariables escogidas en cada grupo.

2.2.1. Pre-procesamiento de las variables agrarias

Los censos agrarios estudiados corresponden a la situación agrícola y ganadera de la península Ibérica en los años 1982, 1989, 1999, 2009 y 2020, siendo las variables escogidas: superficie agraria total (CST), superficie agraria utilizada (CSAU), edad de los jefes de explotación (Edad), sistema de tenencia de las explotaciones (Tenencia), Unidades ganaderas totales (UGT) y nivel de estudios de los jefes de explotación (Formación).

CST: Código de Superficie Total

La superficie agraria total (CST) es un código numérico que identifica las distintas parcelas o unidades de cultivo de una explotación agraria. Recoge información acerca de la superficie

total de la explotación; tanto de las productivas como de las improductivas. Presenta 21 clasificaciones en base al tamaño de las explotaciones, empezando con “sin tierras” hasta “más de 2500 ha”.

CSAU: Código de Superficie Agraria Utilizada

Por su parte, la superficie agraria utilizada (CSAU) es otro código numérico que se utiliza para recoger información sobre la superficie de cada parcela o unidad de cultivo que se utiliza exclusivamente para la actividad agraria, donde se excluyen las parcelas de uso forestal y las improductivas. El “CSAU” permite obtener información más precisa sobre la superficie cultivable de cada explotación agraria. Como en el caso del “CST”, se divide en 21 clasificaciones en base al tamaño, abarcando desde “sin tierras” a “más de 2500 ha”.

EDAD

La variable “Edad” se refiere a la edad de los jefes de las explotaciones agrarias. Esta variable es importante para entender la estructura demográfica de la población rural, así como las tendencias y los cambios en la edad de los agricultores y trabajadores del campo. Se clasifica en distintos grupos: menores de 25 años, de 25 a 34 años, de 35 a 44 años, de 45 a 54 años, de 55 a 64 años, de 65 a 74 años y mayores de 75 años.

TENENCIA

La “Tenencia” se refiere al tipo de propiedad de una explotación, pudiendo ser de una persona o entidad. Tiene cuatro clasificaciones:

- **Propiedad:** la explotación agraria es propiedad del titular de la explotación o de su unidad familiar.
- **Arrendamiento:** la explotación agraria se arrienda o alquila a largo plazo (más de un año) a un tercero.
- **Explotaciones en aparcería o cesión temporal:** la explotación agraria se cede temporalmente a otro agricultor, ya sea a través de un acuerdo verbal o escrito.
- **Otros:** cualquier otra forma de tenencia que no se ajuste a las categorías anteriores, como la propiedad compartida, régimen comunal o la propiedad por parte del gobierno.

UGT

Las unidades ganaderas se refieren a la cantidad de animales que se crían en una explotación. La UGT se calcula en función del tipo de animal que se cría en la explotación, de su edad y peso, y utilizando un sistema de coeficientes de conversión establecido por el Ministerio de Agricultura, Pesca y Alimentación.

FORMACIÓN

La variable “Formación” mide el nivel de formación y educación del titular de la explotación agraria. Se clasifica en:

- **Sin estudios:** para las personas que no han cursado ningún tipo de estudio.
- **Educación primaria:** para las personas que han completado el primer nivel de educación formal.
- **Educación secundaria:** para las personas que han completado el segundo nivel de educación formal.
- **Formación profesional:** para las personas que han completado un programa de educación técnica o profesional.
- **Educación superior:** para las personas que han completado un programa de educación superior, como una carrera universitaria.

2.2.2. Pre-procesamiento de las variables socioeconómicas

Las variables socioeconómicas utilizadas en este trabajo son aquellas a las que se ha podido tener acceso a nivel municipal. Se ha trabajado con los años 2009 y 2020, ya que para los anteriores no se ha podido recabar la información necesaria.

Las bases de datos socioeconómicas relativas a la afiliación a la Seguridad Social, deuda pública y paro por sectores se mantuvieron en su formato original tras la descarga, sin requerir ningún tipo de procesamiento adicional, tal y como se hizo en los microdatos. Esto implica que la estructura y el contenido de los datos eran adecuados para el análisis previsto sin la necesidad de ajustes adicionales.

En el conjunto de datos original de las variables demográficas -nacimientos y defunciones- (microdatos), cada registro individual está asociado a un evento específico, como un nacimiento o un fallecimiento, y se identifica mediante el código del municipio (COD_MUNI). Esto resultó en múltiples entradas repetidas para un mismo municipio, cada una correspondiente a diferentes eventos ocurridos dentro de ese territorio.

Por tanto, con el objetivo de simplificar y resumir la información, se procedió a agrupar los datos por el código de municipio. Mediante este proceso de agregación, se contabilizó el número total de nacimientos y fallecimientos ocurridos en cada municipio durante el periodo de estudio; 2009 y 2020, dando una estructura de datos más manejable y coherente para posteriores análisis estadísticos y geográficos.

Por tanto, las variables con las que se trabajaron en el ámbito socio-económico fueron las siguientes:

- **Afiliados a la Seguridad Social por sectores económicos:** número de personas afiliadas a la Seguridad Social, desglosado por sectores económicos (agricultura, industria, servicios, etc.).
- **Deuda pública de los Ayuntamientos:** total de deuda acumulada (€) por los gobiernos locales, incluyendo préstamos y bonos emitidos.
- **Nacimientos:** cantidad de personas nacidas registrados en un período específico.
- **Defunciones:** Cantidad de defunciones registradas en un período específico.
- **Paro por sectores, edad y sexo:** número de habitantes desempleados por sectores económicos, grupos de edad y sexo.

2.2.3. Obtención de las variables finales

2.2.3.1. Variables agrarias A partir de las variables estudiadas del censo agrario, se derivaron otras nuevas en forma de porcentajes o densidades, siendo más manejables con vistas a los análisis estadísticos. Cada variable se dispone en una posición y tiene una longitud diferente según cada censo agrario, por lo que hubo de tenerse en cuenta para su correcta extracción y correspondiente cálculo. El cálculo de las variables definitivas fue realizado en el software *RStudio*:

- **Superficies agrarias**

Las superficies agrarias son las variables que hacen referencia al número de hectáreas de una explotación total y usada (ST_5ha... y SAU_5ha... respectivamente). El INE proporciona 12 clasificaciones, pero en este caso, han sido cuatro clases las calculadas: de 5 hectáreas, 5-50 ha, 50 a 500 ha y 500 ha. A partir de estas clases, se transformaron los datos en porcentaje por tipo de clasificación.

- **Edad de los jefes de la explotación**

La edad de los jefes de explotación se utilizó para observar el envejecimiento del sector primario en el país, por tanto, solo se consideró a aquellos mayores de 55 años (MAYORES_55). Esta variable midió en porcentaje de jefes de la explotación mayores de 55 años.

- **Estudios de los jefes de la explotación**

Se estudió la educación en los jefes de la explotación, considerando exclusivamente a aquellos que solo tenían estudios básicos. Al igual que en la anterior variable, esta nueva variable se expresó en porcentaje de jefes de explotación sin estudios.

- **Régimen de tenencia de las explotaciones**

Para el régimen de tenencia, se calcularon otras tres variables relativas, pudiendo ser explotaciones propias, arrendadas u otro tipo de contrato (PROPIEDAD, ARRENDAMIENTO y OTROS respectivamente). De nuevo, se calculó un total para luego anotar el porcentaje de cada régimen.

- **Densidad ganadera y de explotaciones**

Con los datos de población de cada año proporcionados igualmente por el INE, se calcularon las densidades de unidades ganaderas como de explotaciones (D_UGT y D_EXPLOT respectivamente).

En total, se recalcularon 15 variables agrarias: ST_5ha, ST_5_50ha, ST_50_500ha, ST_500ha, SAU_5ha, SAU_5_50ha, SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, PROPIEDAD, ARRENDAMIENTO, OTROS, D_EXPLOT y D_UGT. En la **Tabla 4** puede observarse a partir de qué variables se derivan las recalculadas, las unidades y el nombre identificativo que aparece en la nueva base de datos.

Tabla 4: Descripción de qué variable es derivada y qué unidades tienen las variables recalculadas.

Variable	Unidades	Variable recalculada
ST	Superficie Agraria Total: tamaño de las explotaciones (%)	<ul style="list-style-type: none"> ▪ ST_5ha ▪ ST_5_50ha ▪ ST_50_500ha ▪ ST_500ha
ST	Densidad de explotaciones (n° explotaciones/km ²)	D_EXPLOT
SAU	Superficie Agraria Utilizada: tamaño de las explotaciones (%)	<ul style="list-style-type: none"> ▪ SAU_5ha ▪ SAU_5_50ha ▪ SAU_50_500ha ▪ SAU_500ha
EDAD	Proporción de jefes mayores de 55 años (%)	MAYORES_55
TENENCIA	Porcentaje de explotaciones según el régimen de tenencia (%)	<ul style="list-style-type: none"> ▪ PROPIEDAD ▪ ARRENDAMIENTO ▪ APARCERIAyOTROS
UGT	Densidad de unidades ganaderas (UGT/km ²)	D_UGT
FORMACIÓN	Proporción de jefes sin estudios (%)	NO_ESTUDIOS

El cálculo de las nuevas variables se podría resumir en los siguientes pasos (**Figura 5**):

1. Establecer el entorno de trabajo y cargar las librerías necesarias.
2. Definir las variables y sus características para leer el archivo de datos (establecer el tipo de registro, el diseño de columnas y posiciones).
3. Leer los datos y convertirlos en un DataFrame para su correcta manipulación.
4. Revisar los códigos municipales y arreglarlos conforme al mapa de referencia (capa *.shp* del año 2015). En total, se revisaron más de 1500 municipios; con 561 modificaciones, y se seleccionó el código municipal correspondiente.
5. Seleccionar las variables específicas para el análisis y recalcularlas: superficies de explotaciones (ST_5ha... y SAU_5ha...), jefes de la explotación mayores de 55 años (MAYORES_55) y sin estudios superiores (NO_ESTUDIOS), régimen de tenencia de la explotación (PROPIEDAD, ARRENDAMIENTO, OTROS), y densidades de explotaciones, ganaderas y de población (D_EXPLOT, D_UGT y D_POB respectivamente), junto con el código municipal y el año correspondiente. y
6. Guardar los resultados de cada análisis en archivos *.txt* para todos los años.

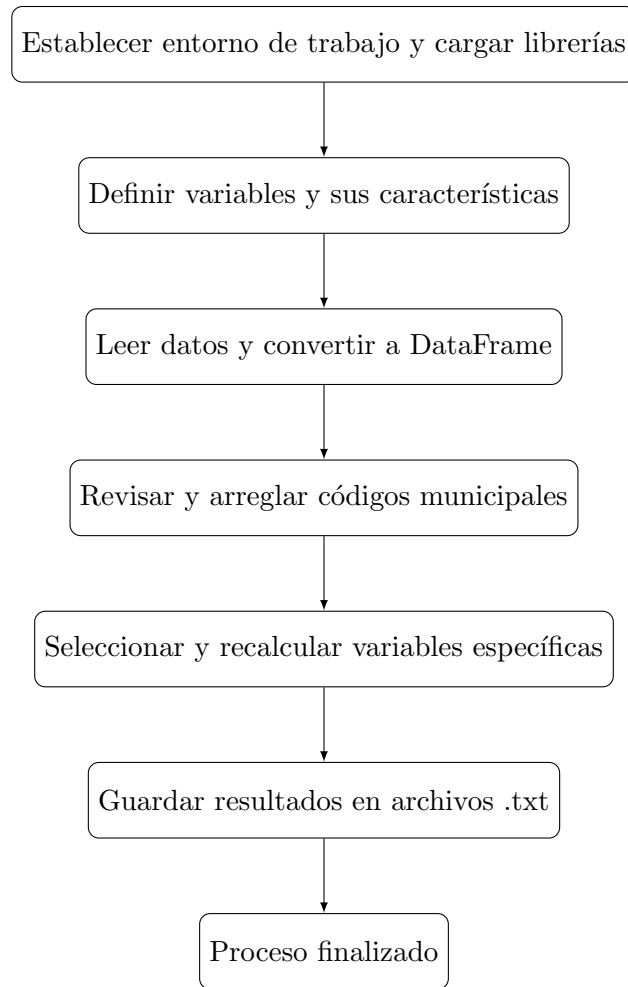


Figura 5: Flujo de trabajo para establecer entorno, procesar datos y obtener resultados de variables agrarias en RStudio.

Toda la metodología para el cálculo de las variables de los censos 82-89-99-09-20 se puede consultar en *Scripts de RStudio* del **Anexo**.

2.2.3.2. Variables socioeconómicas Una vez disponibles las variables deseadas, se recalcularon las variables socioeconómicas finales utilizadas en este estudio. Se operaron de la misma manera las variables inmersas en el grupo de:

- **afiliación a la Seguridad Social** (GENERAL, AGRARIO, MAR, HOGAR, AUTÓNOMOS, CARBÓN, TOTAL)
- **deuda pública de los Ayuntamientos** (deuda_miles_euros)

- **paro por sectores, edad y sexo** (total_paro, paro_hombre_25, paro_hombre_25_45, paro_mujer_25, paro_mujer_25_45, paro_mujer_45, paro_agricultura, paro_industria, paro_construccion, paro_servicios, paro_sin_empleo_anterior)

La operación consistió en dividir a la variable entre los habitantes por municipio y luego multiplicarlo por cien, para así obtener un porcentaje de densidad:

$$P + variable = \frac{variable}{población (n^o de habitantes)} \times 100 \quad (1)$$

Por tanto, estas nuevas variables adoptaron su nombre inicial, añadiendo una P al principio (referencia a “porcentaje”).

Este recálculo estandarizó variables en relación a la población, lo que facilita el análisis comparativo entre distintos años, eliminando el sesgo que podría introducirse por diferencias en el tamaño poblacional.

Por otra parte, la variable final referente a cambios en el número de habitantes considerada (NACIMIENTOS-DEFUNCIONES) fue la Tasa de Crecimiento Vegetativo:

$$Tcv = \left(\frac{n^o nacimientos - n^o de fallecimientos}{población} \right) \times 1000 \quad (2)$$

A partir de las variables Nacimientos y Defunciones, se pudo obtener el crecimiento vegetativo; como se muestra en la Ecuación 1, que no es más que el aumento o descenso de la población producido por la diferencia entre nacidos vivos y defunciones.

Para calcular el crecimiento vegetativo; o tasa de crecimiento vegetativo inicialmente se dividió el resultado de la resta entre el número de nacimientos y defunciones registrados en cada municipio entre la población total correspondiente, para después ajustar estos valores a una escala por mil habitantes.

Esta estandarización es crucial para garantizar que el crecimiento vegetativo proporcione una representación precisa de las dinámicas poblacionales locales, teniendo en cuenta tanto las nuevas incorporaciones a la población como las pérdidas derivadas de la mortalidad.

La **Tabla 5** muestra los nombres finales de las variables socioeconómicas recalculadas.

Tabla 5: Relación entre bases de datos socioeconómicas y variables socioeconómicas finales.

Bases de datos socioeconómicas	Variables finales
Afiliados a la Seguridad Social por sectores	PGENERAL PAGRARIO PMAR PHOGAR PAUTÓNOMOS PCARBÓN PTOTAL
Deuda pública de los Ayuntamientos	Pdeuda_miles_euros
Nacimientos	TCV (Crecimiento vegetativo)
Defunciones	
Paro por sectores, edad, sexo y sectores	Ptotal_paro Pparo_hombre_25 Pparo_hombre_25_45 Pparo_hombre_45 Pparo_mujer_25 Pparo_mujer_25_45 Pparo_mujer_45 Pparo_agricultura Pparo_industria Pparo_construccion Pparo_servicios Pparo_sin_empleo_anterior

2.2.4. Unión de datos geospaciales con la base de datos

Para observar estas variables geoespacialmente, fue necesario un estudio cartográfico municipal llevado a cabo a través de la columna COD_MUNI del censo agrario y COD_MUNI de la capa vectorial descargada del Centro de Descargas del Instituto Geográfico Nacional (IGN). Esta capa describe a todos los municipios de la Península, representado con el sistema geodésico de coordenadas Datum WGS84 zona UTM; permitiendo localizar cualquier punto

de la Tierra por medio de tres unidades dadas (x,y,z). Los husos de las coordenadas UTM de la Península se sitúan en el intervalo [29, 31]. Estos datos del IGN son de tipología *shapefile*, siendo un formato vectorial sencillo y no topológico que se utiliza para almacenar la ubicación geométrica y la información de atributos de las entidades geográficas. Las entidades geográficas de un *.shp* se pueden representar por medio de puntos, líneas o polígonos -áreas-. Es un formato multiarchivo, generado por varios ficheros informáticos (ESRI, 2021). El número mínimo requerido es 3, teniendo las siguientes extensiones:

1. *.shp*: es el archivo que almacena las entidades geométricas de los objetos.
2. *.shx*: es el archivo que almacena el índice de las entidades geométricas.
3. *.dbf*: es la base de datos, en formato dBASE, donde se almacena la información de los atributos de los objetos.

Igualmente, se pueden utilizar otros para mejorar el funcionamiento en las operaciones de consulta a la base de datos geográfica:

1. *.prj*: es el archivo que guarda la información referida al sistema de coordenadas en formato WKT.
2. *.sbn* y *.sbx*: almacena el índice espacial de las entidades.
3. *.fbn* y *.fbx*: almacena el índice espacial de las entidades para los shapefiles que son inalterables (solo lectura).
4. *.ain* y *.aih*: almacena el índice de atributo de los campos activos en una tabla o el tema de la tabla de atributos.
5. *.xml*: almacena los metadatos del *shapefile* (ESRI, 1998).

En este caso, los municipios están representados por polígonos, donde la capa de atributos almacena las siguientes columnas (**Tabla 6**):

Tabla 6: Descripción de los ítems presentes en el shapefile de municipios.

Ítem	Descripción
OBJETID	Identificador único del objeto en el archivo shapefile.
Km2	Superficie del municipio en kilómetros cuadrados.
Shape_Leng	Longitud del perímetro del municipio.
Shape_area	Área total del municipio en la proyección utilizada.
FID	Identificador único de la entidad en la tabla de atributos del shapefile.
Shape*	Geometría del municipio, es decir, la forma que define el contorno del municipio en el mapa.
INSPIREID	Identificador europeo (INSPIRE) asignado al municipio para asegurar la interoperabilidad de los datos geoespaciales.
COUNTRY	Código del país al que pertenece el municipio (en este caso, España).
NATLEV	Nivel de la unidad administrativa (por ejemplo, municipio, provincia, etc.).
NATLEVNAME	Nombre del nivel de la unidad administrativa (por ejemplo, "Municipio", "Provincia").
NATCODE	Código nacional que identifica de manera única la unidad administrativa en el país.
NAMEUNIT	Nombre oficial del municipio.
CODNUT1	Código de la primera división NUTS (Nomenclatura de Unidades Territoriales Estadísticas) que agrupa varias regiones en Europa.
CODNUT2	Código de la segunda división NUTS, que es una subdivisión de la NUTS 1, normalmente corresponde a la comunidad autónoma en España.
CODNUT3	Código de la tercera división NUTS, correspondiente a la provincia en España.

Extrayendo las últimas cinco cifras de la columna NATCODE con el código !NATCODE![-5:] en *Field Calculator* de ArcGis Pro con Python, se calculó la columna utilizada para realizar

la unión *-join-* entre matrices de datos: “COD_MUNI”.

Los COD_MUNI de la capa *.shp* son los que se pueden visualizar en un mapa cartográfico; ya que poseen información geoespacial, mientras que los COD_MUNI del censo agrario simplemente son un identificador de municipios. Para que los COD_MUNI del censo agrario puedan visualizarse cartográficamente, se les realizó un depurado y se les unió a los COD_MUNI cartográfico: a lo largo de los años, los municipios han cambiado de nomenclatura, se han segregado y formado otros nuevos, o se han unido. Ésto, no lo refleja el COD_MUNI del censo agrario; teniendo a sus COD_MUNI en la situación política y social de su respectivo año, mientras que la capa *.shp* es la última actualización (2015). Por tanto, estos cambios que no se reflejan en el censo se tuvieron que analizar uno por uno sobre los COD_MUNI no coincidentes. A través de bibliografía (MAP, 2008) y de buscadores de municipios del INE, se unieron relacionamente para que encajasen los del censo agrario con los del *.shp*.

Para alcanzar el objetivo de este estudio, se integraron las matrices de las variables agrarias de los años 1982-89-99-09-20 y las socioeconómicas de los años 2009-20, trabajando con ratios de cambio y con la técnica estadística multivariante STATIS-Dual. Además, se utilizaron dos técnicas estadísticas adicionales: la PLSR y la Co-inercia, con el fin de analizar la relación entre variables agrarias, socioeconómicas y de incendios. Se realizaron trayectorias y clasificaciones geoespaciales basadas en clústers para visualizar y comprender las tendencias y patrones en la distribución de variables agrarias y socioeconómicas.

2.2. Estadística

2.2.1. Estadística descriptiva

El primer paso en el análisis estadístico descriptivo fue generar un resumen numérico básico de las variables seleccionadas en los datos socio-económicos y del censo agrario utilizando la función *summary()* en *RStudio*. Esta función calcula la media, mediana, desviación estándar, cuartiles, y valores extremos para cada variable. Estos resúmenes proporcionan una visión general de la tendencia central y la dispersión de los datos, permitiendo identificar rápidamente las características clave de las variables analizadas.

Se necesitaron visualizar a los datos más en profundidad por lo que se realizaron gráficos Boxplot, de densidad y de correlación.

■ **Boxplot o Diagrama de Cajas y Bigotes**

En *RStudio*, se generaron boxplots para las variables seleccionadas de los conjuntos de datos socioeconómicos y del censo agrario. Se utilizó la función *ggplot* para crear estos gráficos, que visualizan la distribución de cada variable mediante su mediana, cuartiles, y valores atípicos. Los boxplots permiten comparar visualmente la variabilidad y la dispersión de los datos entre diferentes años o grupos, identificando posibles valores atípicos que podrían influir en el análisis posterior¹.

■ **Diagrama de densidad**

Usando la función *geom_density()* dentro del paquete *ggplot2*, se pudieron realizar los gráficos de densidad. Este tipo de gráfico es útil para observar la forma de la distribución, destacando áreas de alta concentración de datos y permitiendo comparaciones entre años².

■ **Mapa de Calor de Correlación**

El análisis de correlación se realizó en *RStudio* utilizando la función *cor()* para calcular la matriz de correlación entre las variables de los conjuntos de datos socio-económicos y del censo agrario. Posteriormente, se visualizó esta matriz mediante un mapa de calor utilizando el paquete *corrplot*. El mapa de calor muestra visualmente la fuerza y la dirección de la relación lineal entre las variables, con colores que representan la magnitud y el signo de la correlación. Este análisis es crucial para identificar relaciones fuertes entre variables que podrían afectar los análisis multivariantes posteriores³.

¹Código consultado en: <https://r-charts.com/ggplot2/axis/>

²Código consultado en: <https://r-charts.com/ggplot2/themes/>

³Código consultado en: <https://r-charts.com/correlation/corplot/>

2.2.2. Estadística Multivariante

STATIS-Dual

El STATIS-Dual (Abdi et al., 2012) puede verse como una generalización del Análisis de Componentes Principales (ACP). Éste método privilegia las posiciones relativas de las J variables, pues como hemos apuntado con anterioridad, las tablas tienen las mismas J variables y los individuos pueden diferir entre tablas. Se parte de las K matrices \mathbf{X}_k (de dimensión $j \times i$), donde $k = 1, 2, \dots, K$.

En la estructura del análisis se diferencian:

- **Matrices de pesos:** considerando la matriz de pesos \mathbf{M} para las variables y \mathbf{D}_k para los individuos de cada tabla \mathbf{X}_k .
- **Matrices de Varianzas y Covarianzas:** para cada tabla \mathbf{X}_k , se calcula la matriz de covarianza \mathbf{S}_k de tamaño $J \times J$, definida como $\mathbf{S}_k = \mathbf{X}_k^T \mathbf{D}_k \mathbf{X}_k$.
- **Normalización de \mathbf{S}_k :** se utiliza la norma definida por el producto escalar de Hilbert-Schmidt ($\|\mathbf{S}_k\|_{\text{HS}}$) para normalizar \mathbf{S}_k .

Este análisis se lleva a cabo en cuatro etapas:

Etapas 1. Estudio de la interestructura: procede a una comparación global de la estructura de las matrices de datos.

Etapas 2. Construcción de la matriz compromiso: búsqueda de una estructura común a las matrices estudiadas.

Etapas 3. Estudio de la intraestructura: representación de la estructura de cada matriz de datos en un espacio de baja dimensión. En esta etapa, se analizan las trayectorias de los individuos, representándolas gráficamente en el espacio euclidiano del compromiso.

En la **Figura 6** se muestran los pasos que se siguen en este análisis estadístico multivariante.

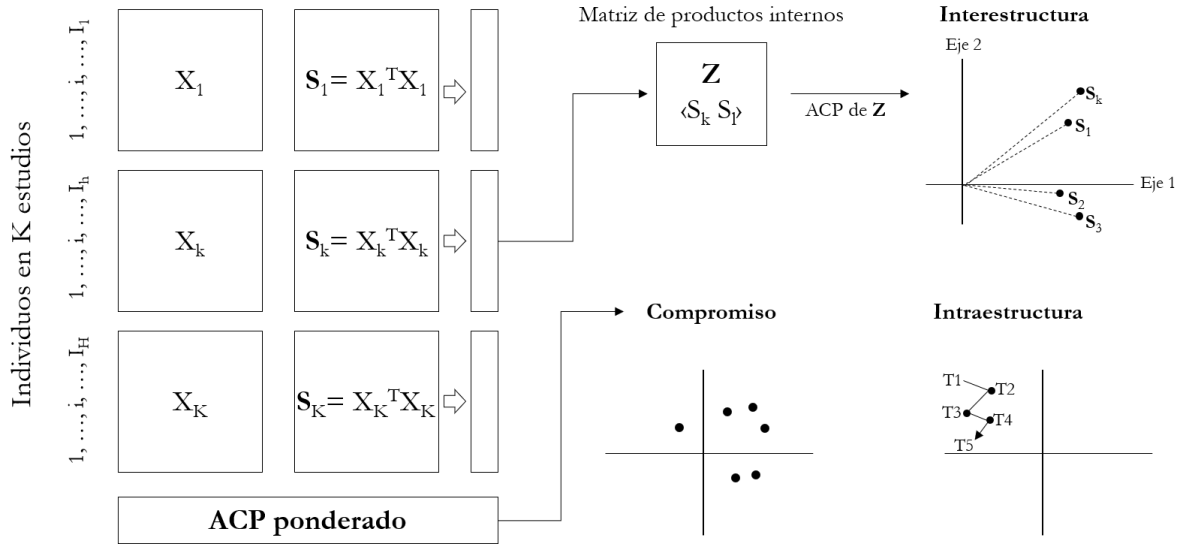


Figura 6: Esquema de los procedimientos del STATIS-Dual. Elaborado a partir de Rodríguez-Martínez (2020).

Etapa 1. Interestructura A partir de los productos escalares entre los objetos \mathbf{S}_k , se genera la matriz \mathbf{Z} ; de dimensión $K \times K$.

Es decir, para analizar las relaciones entre las distintas K tablas, se construye a la matriz \mathbf{Z} de tamaño $K \times K$, donde cada elemento $\mathbf{Z}_{kl} = \langle \mathbf{S}_k, \mathbf{S}_l \rangle_{\text{HS}}$.

- **Construcción de la imagen euclídea**

Se busca un subespacio que reproduzca los productos escalares entre los objetos \mathbf{S}_k es decir, que reproduzca las matrices de varianzas y covarianzas de las K tablas. Se diagonaliza \mathbf{Z} para obtener su descomposición espectral:

$$\mathbf{Z} = \mathbf{V}_Z \Lambda_Z^2 \mathbf{V}_Z'$$

La diagonalización proporciona una imagen euclídea de la interestructura: se calculan los vectores y valores propios de la matriz de correlaciones vectoriales para reducir así su dimensionalidad y representarla en forma euclídea.

En esta descomposición, cuando los datos están normalizados, cada elemento de esta matriz \mathbf{Z} es el coeficiente de correlación vectorial (RV) entre pares de tablas. Estos coeficientes -acotados entre 0 y 1- tienen una interpretación interesante, ya que cuanto mayor es su valor, mayor será la similitud entre dos tablas y al contrario. Por tanto,

su interpretación es análoga a la interpretación del valor absoluto del coeficiente de correlación lineal entre dos variables.

Las coordenadas de los objetos $\mathbf{S}_1, \dots, \mathbf{S}_k$ en la imagen euclídea se contienen en la matriz:

$$\mathbf{A}_{SK} = V_Z \Lambda_Z$$

El resultado de este análisis se puede representar gráficamente en un espacio bidimensional (**Figura 7**), en el que cada tabla viene representada mediante un punto que puede unirse con el origen de coordenadas. Se puede interpretar visualmente la relación entre ellas, cuanto más cercanos estén entre sí y con respecto al primer eje, habrá una estructura común a las tablas que podrá describirse adecuadamente por el consenso, es decir, cuanto más pequeños sean los ángulos que forman entre sí y con el primer eje y mayor sea la longitud de los vectores (y por tanto las coordenadas de cada tabla sobre el primer eje).

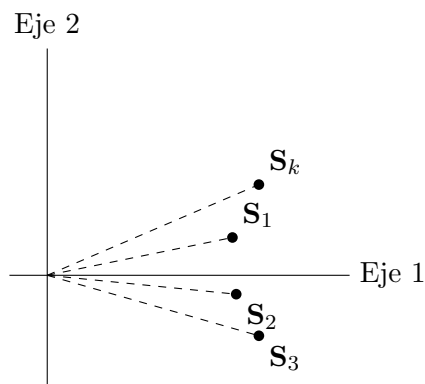


Figura 7: Situación en el estudio de la interestructura de la configuración de las tablas. Se observan puntos en el plano; $(S_{k,1,2,3})$, que representan a cada una de las tablas.

Etapa 2. Construcción de la matriz consenso o matriz compromiso El análisis de la interestructura proporciona información sobre la similitud entre las distintas K tablas. Si se encuentran similitudes significativas, es útil construir una matriz de consenso o compromiso \mathbf{S} para representar a los estudios. Esta matriz de compromiso debe ser un buen resumen de la información contenida en todas las configuraciones.

El consenso se construye como media ponderada de los K objetos (\mathbf{S}_k), definida como:

$$\mathbf{S} = \sum_{k=1}^K \beta_k \mathbf{S}_k$$

Se utilizan como ponderaciones de cada objeto la correspondiente componente del primer valor propio normado de la descomposición espectral de la matriz \mathbf{Z} o de la interestructura. Dichos factores de ponderación se designan como β_k .

- **Consenso a partir de objetos normados**

Cuando se normalizan los objetos, el consenso se calcula como:

$$\mathbf{S} = \sum_{k=1}^K \beta_k \frac{\mathbf{S}_k}{\|\mathbf{S}_k\|_{\text{HS}}}$$

Los factores de ponderación β_k se calculan como:

$$\beta_k = \frac{1}{\lambda_{1\mathbf{S}_k} \mathbf{v}_{1ks}}$$

donde \mathbf{v}_{1ks} es la componente k del primer vector propio de la matriz \mathbf{Z} .

Asimismo, hay otra manera para la construcción del consenso a través de objetos no normados. Como en este caso los datos fueron normalizados, no se procede a la explicación de dicho proceso.

- **Análisis del compromiso**

La matriz consenso o la matriz compromiso se analiza mediante un ACP, de forma que podremos analizar la estructura común de covariación entre las variables.

Etapa 3. Intraestructura El análisis de la intraestructura consiste en examinar la estructura interna de cada matriz y su relación con la matriz compromiso. El objetivo es identificar cómo se relacionan las variables dentro de cada tabla y cómo estas relaciones se alinean con el compromiso.

Es posible proyectar las variables originales de cada tabla sobre el espacio compromiso para poder analizar el comportamiento de las primeras a través de las diferentes tablas. Ésto se

hace proyectando como elementos suplementarios las variables de cada tabla sobre el espacio compromiso.

Las trayectorias son un conjunto de puntos que se define por la proyección de cada uno de los individuos I para todas las \mathbf{K} matrices.

Al proyectar los puntos sobre el espacio de los componentes principales de la matriz compromiso \mathbf{S} , se podría observar la evolución de cada individuo en el tiempo o momento.

Así podremos observar si una variable tiene un comportamiento similar a lo largo del tiempo (entre las distintas tablas) si su trayectoria es envolvente, o si cambia su comportamiento (trayectoria excéntrica). Para observar el comportamiento de las variables a través de las distintas tablas, se consideran las variables de cada tabla como elementos suplementarios y se proyectan en la imagen euclídea consenso:

$$S_k M V_S D_S^{-1} = B_k$$

Análisis de Co-Inercia

El Análisis de Co-Inercia (CoIA) (Dolédéc y Chessel, 1994; Dray et al., 2003) describe la coestructura entre dos tablas de datos ecológicos.

El CoIA estudia la relación entre dos tablas de datos maximizando las covarianzas entre ellas; asegurando que las puntuaciones tengan un buen porcentaje de varianza explicada, similar a cómo el ACP maximiza la varianza dentro de una tabla (**Figura 8**). En este análisis, la coinercia se define como la suma de los cuadrados de las covarianzas entre las dos tablas, siendo posible cuando ambas tablas comparten las mismas filas. Es similar al Análisis Canónico de Correspondencias (CCA).

En el caso de datos ambientales, sea X la primera tabla de variables con n filas de lugares de muestreo y p columnas de variables, y Y la segunda tabla de datos de especies con las mismas n filas y q columnas de especies. D_n es la matriz diagonal ($n \times n$) de pesos de los sitios, y D_p y D_q son dos métricas en el espacio \mathbb{R}^p y \mathbb{R}^q respectivamente. Al maximizar la covarianza en lugar de la correlación, el Análisis de Coinercia asegura que las puntuaciones no tengan varianzas muy pequeñas y, por lo tanto, tengan un buen porcentaje de varianza explicada en cada espacio.

Análisis de Componentes Principales Antes del Análisis de Coinercia, se analiza cada tabla por separado utilizando un ACP generalizado de estos tripletes, que corresponde a la descomposición espectral de $X^T D_n X D_p$ y $Y^T D_n Y D_q$. Cuando los pesos de las filas D_n son uniformes y las métricas D_p y D_q son euclidianas, se realizan ACPs simples.

Definición del CoIA El Análisis de Coinercia se define mediante el diagrama de dualidad que fusiona estos dos análisis separados, siempre que las filas de ambas tablas sean idénticas. El Análisis de Coinercia es el análisis de valores propios de la matriz $X^T D_n Y D_q Y^T D_n X D_p$ y $Y^T D_n X D_p X^T D_n Y D_q$. El Análisis de Coinercia analiza una tabla de productos cruzados, maximizando la covarianza entre las puntuaciones de las filas de las dos tablas.

Inercia y Coinercia Si las columnas de ambas tablas están centradas, la inercia total de cada tabla es una suma de varianzas:

$$I_X = \text{traza}(X D_p X^T D_n) \quad \text{e} \quad I_Y = \text{traza}(Y D_q Y^T D_n)$$

La coinercia entre X y Y se calcula como:

$$\text{CoIner}_{XY} = \sum_{j=1}^p \sum_{k=1}^q (\text{Cov}(X_j, Y_k))^2$$

Este valor es alto cuando los valores en ambas tablas son altos simultáneamente o varían inversamente, y bajo cuando varían independientemente.

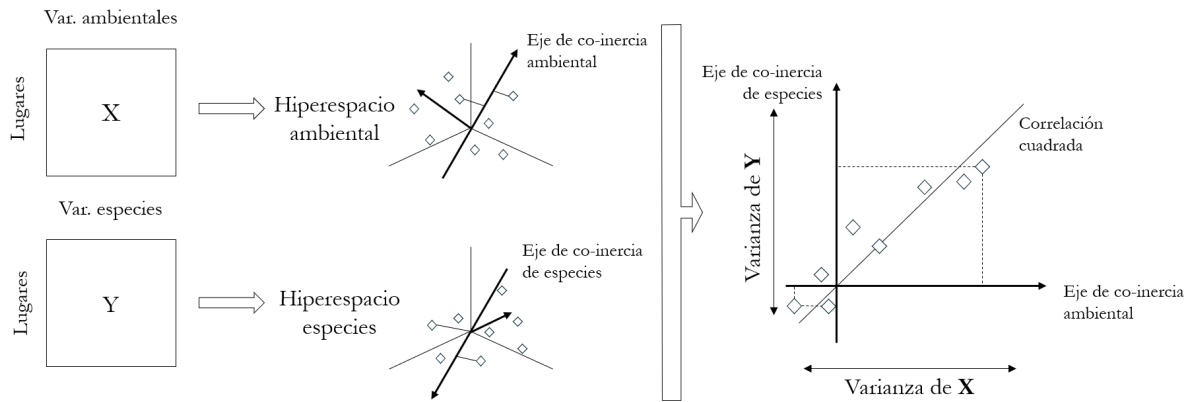


Figura 8: Principios del análisis de Coinercia (CoIA). Las tablas X e Y producen dos representaciones en los dos hiperespacios, para encontrar, mediante análisis separados, la máxima inercia en cada hiperespacio y, por tanto, encontrar los dos ejes de coinercia para proyectar los lugares. CoIA maximiza el cuadrado de la covarianza entre las proyecciones de los lugares en los ejes de coinercia. Elaborado a partir de Dufour (2009).

Clustering

El clustering es una técnica de *data mining* usada para analizar datos con variaciones y cambios en el número de individuos, agrupando a éstos en grupos o *clusters*. Cada grupo contiene datos que son lo más similares posibles y a la vez diferentes de los demás objetos cluster (Syakur et al., 2018). El *machine learning* pretende la clasificación automática de datos, pudiendo considerarse tres tipos de algoritmos:

- Clasificación supervisada
- Clasificación no supervisada

En este trabajo, el análisis de clusters se ha realizado mediante técnica no supervisada, empleándose para ayudar a la visualización de los resultados del análisis del STATIS-Dual y la Co-inercia (trayectorias de los municipios), permitiendo completar la interpretación.

Clustering K-means El Método o algoritmo K-means es una técnica no supervisada de clusterización, siendo la más simple y usada en todos los ámbitos científicos. El K-means agrupa grandes conjuntos de datos con una rápida y eficiente tiempo de computación (Ahmed et al., 2020). El algoritmo sigue tres pasos:

1. Inicialización: después de elegir el número de grupos k o clúster inicial, se establecen k centroides en el espacio de los datos aleatoriamente.
2. Asignación de datos a los centroides: cada dato es asignado a su centroide más cercano.
3. Actualización de centroides: se actualizan las posiciones de los centroides de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Los pasos 2) y 3) se repiten hasta que los centroides no cambian su posición (**Figura 9**) (UO, 2026; Pham et al., 2005).

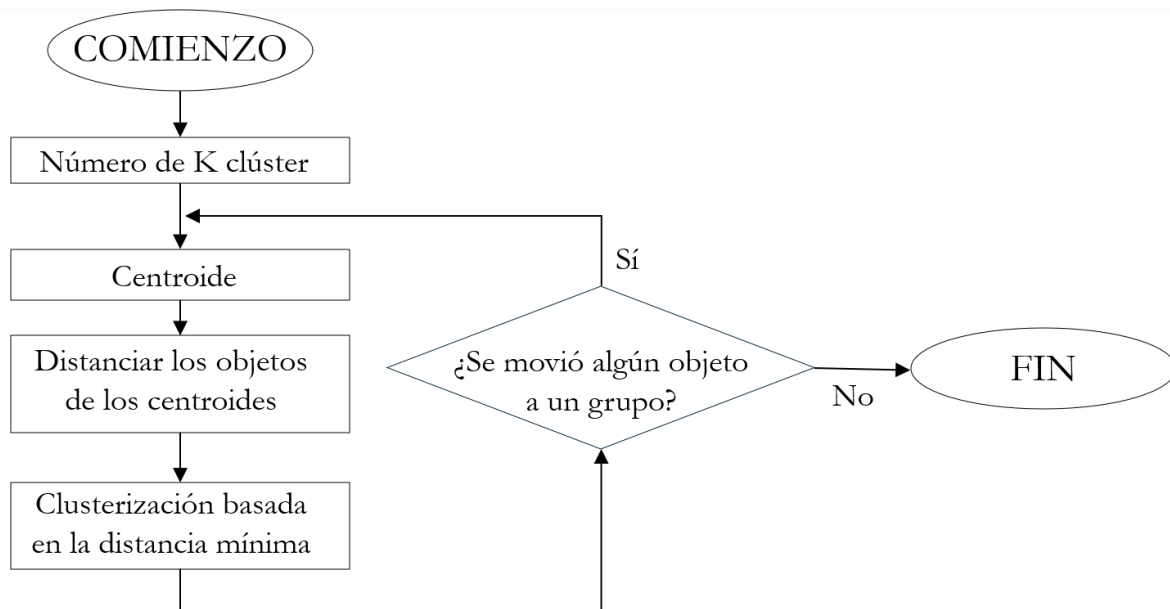


Figura 9: Diagrama de flujo del algoritmo de clustering K-means. Elaborado a partir de Bustamam et al. (2017).

Sin embargo, tiene una desventaja y es que depende del centroide del clúster inicial. Por esta razón, el Método del codo es utilizado para hallar un centroide óptimo.

Elbow Method o Método del codo El Método del codo es un método gráfico que se usa para encontrar el valor óptimo K en el algoritmo clustering K-means (**Figura 10**). El gráfico muestra los valores de la suma de cuadrados dentro del grupo (WCSS) en el eje y ; correspondiendo a los diferentes valores de K en el eje x . El valor K óptimo es el punto en el cual el gráfico forma un codo (Syakur et al., 2018).

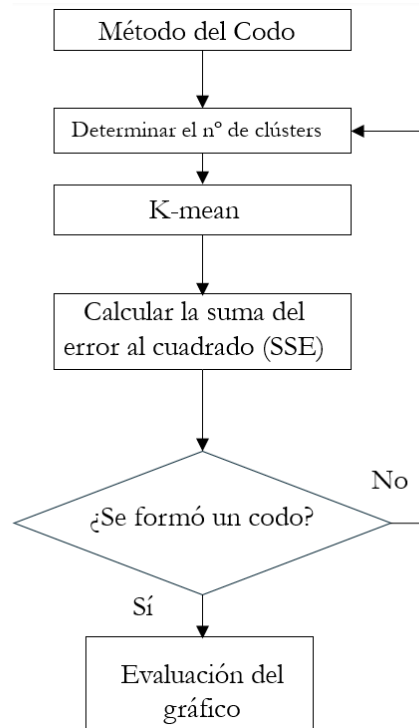


Figura 10: Diagrama de flujo del Método del Codo. Elaborado a partir de Marutho et al. (2018).

Silhouette Method o el Método de la Silueta El Método de la Silueta utiliza un coeficiente de “silueta” que combina tanto la separación como la cohesión de los clusters. Este coeficiente se determina mediante una fórmula que compara la medida de separación con la medida de cohesión, ajustando el valor en función de cuál de las dos siluetas es mayor.

La fórmula para el coeficiente de silueta es:

$$s = \begin{cases} 1 - \frac{\text{medida de cohesión}}{\text{medida de separación}} & \text{si la cohesión} < \text{separación} \\ \frac{\text{medida de separación}}{\text{medida de cohesión}} - 1 & \text{si la cohesión} > \text{separación} \end{cases} \quad (3)$$

donde:

- **la medida de cohesión** es la distancia promedio entre un punto y todos los puntos dentro del mismo cluster.
- **la medida de separación** es la distancia promedio entre el punto y todos los puntos en el cluster más cercano.

Un coeficiente de silueta más alto indica una mejor calidad del cluster, ya que refleja una mayor separación entre clusters y una mayor cohesión dentro del mismo cluster (Saputra et al., 2020)

Regresión Parcial por Mínimos Cuadrados (PLSR)

La regresión parcial por mínimos cuadrados (PLSR) (Wold, 1984) es una técnica que busca predecir, a partir de una matriz de datos $I \times J$, un conjunto de variables dependientes $\mathbf{Y} = (y_{ik})$ a partir de un conjunto de predictores $\mathbf{X} = (x_{ij})$. Por lo tanto, tenemos I individuos, J variables predictoras y K variables de respuestas (**Figura 11**). Es especialmente útil cuando hay presencia de multicolinealidad (Li et al., 2009; Anzanello y Fogliatto, 2014; Firinguetti et al., 2017).

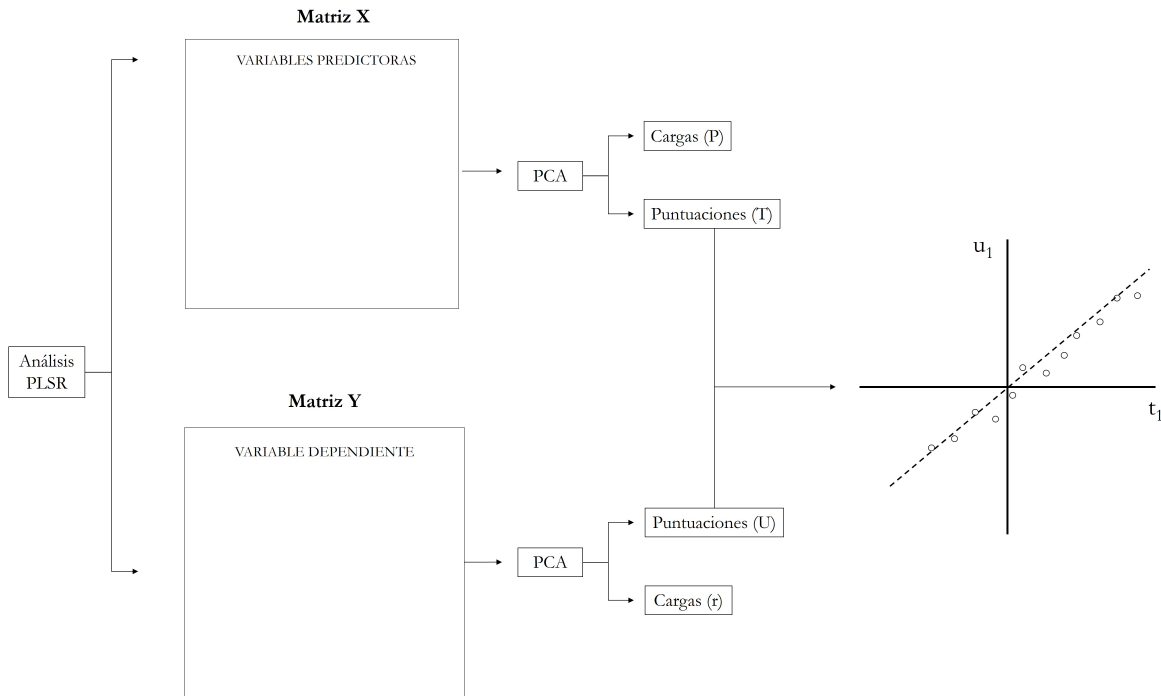


Figura 11: Diagrama que explica las matrices de datos implicadas y los procesos necesarios para construir el gráfico de regresión lineal predictivo. Elaboración propia a partir de Pelliccia (2018).

Una técnica relacionada es la regresión de componentes principales (PCR), que comienza con un análisis de componentes principales (PCA) de la matriz \mathbf{X} . El PCA descompone \mathbf{X} en sus componentes principales, que luego se utilizan como predictores en un modelo de regresión para predecir \mathbf{Y} . Este método elimina el problema de la multicolinealidad, ya que

los componentes principales son ortogonales entre sí. Sin embargo, PCR tiene una limitación importante: los componentes principales se eligen para explicar la varianza de \mathbf{X} sin considerar su capacidad para predecir \mathbf{Y} , lo que puede llevar a seleccionar componentes que no son óptimos para la predicción de \mathbf{Y} .

Descomposición simultánea de predictores y variables dependientes

A diferencia del PCR, la PLSR busca componentes en \mathbf{X} que maximicen la covarianza con \mathbf{Y} . Esto se logra descomponiendo simultáneamente \mathbf{X} y \mathbf{Y} en un conjunto común de factores ortogonales, denominados vectores latentes, y matrices de *loadings* específicos. En PLSR, \mathbf{X} se descompone como:

$$\mathbf{X} = \mathbf{TP}^T \text{ con } \mathbf{T}^T\mathbf{T} = \mathbf{I}, \quad (4)$$

donde \mathbf{T} es la matriz de puntuaciones (vectores latentes) y \mathbf{P} es la matriz de *loadings*, \mathbf{I} es la matriz identidad. Por analogía con el PCA, \mathbf{T} se llama la matriz de puntuaciones, y \mathbf{P} la matriz de *loadings* (en la regresión PLS los *loadings* no son ortogonales).

De manera similar, \mathbf{Y} se estima como:

$$\mathbf{Y} = \mathbf{TBC}^T, \quad (5)$$

donde \mathbf{B} es una matriz diagonal de pesos de regresión y \mathbf{C} es la matriz de pesos de \mathbf{Y} . La clave es que los vectores latentes se eligen para maximizar la covarianza entre \mathbf{X} y \mathbf{Y} , lo que generalmente mejora la capacidad predictiva del modelo.

Regresión PLS y Covarianza

Los vectores latentes en PLSR se determinan de manera que maximizan la covarianza entre combinaciones lineales de \mathbf{X} y \mathbf{Y} . Específicamente, se buscan pesos \mathbf{w} y \mathbf{c} para crear las combinaciones:

$$\mathbf{t} = \mathbf{Xw} \quad \mathbf{u} = \mathbf{Yc}, \quad (6)$$

de modo que la covarianza entre \mathbf{t} y \mathbf{u} sea máxima. Una vez encontrado un vector latente, se elimina su efecto de \mathbf{X} y \mathbf{Y} , y el proceso se repite hasta que \mathbf{X} se reduzca a una matriz

nula. Este procedimiento garantiza que cada vector latente capturado maximice la capacidad predictiva de \mathbf{Y} .

Algoritmo NIPALS

El algoritmo NIPALS, utilizado en PLSR, sigue un proceso iterativo para identificar los vectores latentes y sus respectivos pesos. Se parte con matrices $\mathbf{E} = \mathbf{X}$ y $\mathbf{F} = \mathbf{Y}$, centradas y normalizadas, y se inicializa un vector \mathbf{u} aleatorio. El algoritmo realiza varios pasos para calcular los pesos y puntuaciones de \mathbf{X} y \mathbf{Y} , ajustando continuamente los componentes hasta que se logra la convergencia:

1. $\mathbf{w} \propto \mathbf{E}^T \mathbf{u}$ (estimación de los pesos de \mathbf{X}).
2. $\mathbf{t} \propto \mathbf{E} \mathbf{w}$ (estimación de las puntuaciones factoriales de \mathbf{X}).
3. $\mathbf{c} \propto \mathbf{F}^T \mathbf{t}$ (estimación de los pesos de \mathbf{Y}).
4. $\mathbf{u} = \mathbf{F} \mathbf{c}$ (estimación de las puntuaciones de \mathbf{Y}).

Si \mathbf{t} no ha convergido, entonces regresa al Paso 1.

Si \mathbf{E} es una matriz nula, entonces se ha encontrado todo el conjunto de vectores latentes; de lo contrario, se puede reiterar el procedimiento desde el Paso 1.

El algoritmo NIPALS muestra una relación estrecha con la descomposición en valores singulares (SVD). Específicamente, los pesos de \mathbf{X} (\mathbf{w}) corresponden al primer vector singular por la derecha de la matriz $\mathbf{S} = \mathbf{E}^T \mathbf{F}$.

Mientras, los pesos de \mathbf{Y} (\mathbf{c}) son el primer vector singular izquierdo. Esta conexión con SVD es especialmente útil cuando el número de predictores es mucho mayor que el número de observaciones, ya que permite trabajar con matrices más pequeñas.

Predicción de las Variables Dependientes

La predicción de \mathbf{Y} en PLSR se realiza utilizando la fórmula de regresión multivariante:

$$\mathbf{Y} = \mathbf{T} \mathbf{B} \mathbf{C}^T = \mathbf{X} \mathbf{B}_{PLS} \quad (7)$$

donde \mathbf{B}_{PLS} es la matriz de coeficientes obtenida a partir de las puntuaciones y pesos. Cuando se utilizan todas las variables latentes, esta regresión es equivalente a PCR, pero con

un enfoque más orientado a la predicción de Y . Si se selecciona un subconjunto óptimo de variables latentes, la predicción de Y es óptima para ese número de predictores.

En este trabajo, las variables predictoras son las correspondientes al censo agrario de 2009 y a la socioeconomía del 2009. La variable de respuesta fue el número total de incendios de la base de datos de incendios del año 2009 a nivel municipal, proporcionada por MITECO (2021).

3. Resultados

3.1. Estadística descriptiva

VARIABLES DEL CENSO AGRARIO

Las variables estudiadas en el censo agrario se calcularon como porcentajes (%); exceptuando a la densidad de explotación y de unidades ganaderas.

En la distribución de las superficies de las explotaciones, se observan diferentes patrones dependiendo del tamaño de las parcelas. Las explotaciones menores a 5 hectáreas presentan una media de 35,22, es decir, lo más común es que en un municipio haya un 35,22% de explotaciones agrarias menores de 5 ha. Las superficies entre 5 y 50 hectáreas muestran una media de 38,10, con una distribución bastante homogénea en ambos casos. Las superficies más grandes, entre 50 y 500 hectáreas, tienen una media de 22,51, y para explotaciones mayores a 500 hectáreas, la media desciende drásticamente a 3,89, indicando que las grandes extensiones de terreno son menos comunes. Por otro lado, la superficie agrícola utilizada (SAU) sigue una tendencia similar, con una disminución progresiva de la media conforme aumenta el tamaño de las parcelas, siendo la SAU en explotaciones mayores a 500 hectáreas notablemente baja, con una media de 2,00.

En términos de características demográficas y educativas, se destaca que la media de propietarios mayores de 55 años es de 50,04 por lo que en general, la mitad de los jefes de explotación de España son mayores de 55 años. Además, el porcentaje de propietarios sin estudios es alarmantemente alto, con una media de 84,74, lo que refleja una baja formación educativa en este sector. En contraste, este porcentaje se reduce a lo largo de los años.

En cuanto a la tenencia de la explotación, la propiedad es la forma predominante, con

una media de 67,85, seguida del arrendamiento con una media de 23,87. Las otras formas de tenencia son mucho menos comunes.

Finalmente, la densidad de unidades ganaderas en el sector agrícola presenta una gran disparidad, con una media de 35.69, pero con valores extremos que alcanzan hasta 2.401,16 (**Tabla 7**)

Tabla 7: Resumen estadístico de las variables del censo agrario de los años 1982-89-99-09-20.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
ST_5ha	0.000	9.333	30.769	35.220	56.234	100.000
ST_5_50ha	0.00	21.36	38.16	38.10	53.85	100.00
ST_50_500ha	0.000	2.459	12.500	22.507	35.294	100.000
ST_500ha	0.000	0.000	0.000	3.888	1.852	100.000
SAU_5ha	0.00	13.76	39.67	42.20	67.78	100.00
SAU_5_50ha	0.00	14.41	30.83	32.56	48.21	100.00
SAU_50_500ha	0.000	1.031	9.375	20.847	32.369	100.000
SAU_500ha	0.000	0.000	0.000	1.999	0.000	100.000
MAYORES_55	0.00	38.46	50.79	50.04	62.50	100.00
NO_ESTUDIOS	0.00	78.05	93.07	84.74	98.63	100.00
D_EXPLOT	0.000	1.048	2.200	4.704	4.787	604.000
PROPIEDAD	0.00	52.09	71.45	67.85	87.12	100.00
ARRENDAMIENTO	0.000	5.753	18.237	23.870	37.205	100.000
OTROS	0.0000	0.1835	2.9015	8.6676	9.7027	199.2311
D_UGT	0.000	5.035	14.232	35.694	34.818	2401.157

Variables socioeconómicas

El análisis estadístico de las variables socioeconómicas revela una marcada asimetría en la distribución de los datos, destacando una alta heterogeneidad entre los municipios.

Las variables relativas a las personas inscritas a Seguridad Social por sectores (GENERAL, AGRARIO, MAR, HOGAR, AUTÓNOMOS, CARBÓN y TOTAL) presentan una distribución sesgada, donde la media supera significativamente la mediana, indicando la concentración de inscripciones en pocos municipios. Por ejemplo, la variable GENERAL muestra una media

de 1,604 inscritos frente a una mediana de 54, con un máximo de 1,739,449 personas.

En cuanto a la deuda de los ayuntamientos, se observa una alta variabilidad. La media es de 2.721 miles de euros mientras que la mediana es 2 €, sugiriendo que unas pocas localidades acumulan la mayor parte de la deuda total, alcanzando hasta 6.776.856 miles de euros.

La Tasa de Crecimiento Vegetativo (TCV) muestra una tendencia hacia valores cercanos a cero, lo que indica un equilibrio general entre natalidad y mortalidad, aunque con variaciones locales.

Las variables de desempleo (paro) reflejan una distribución altamente dispersa, con medias que superan considerablemente las medianas, especialmente en el sector servicios, donde el desempleo alcanza un máximo de 1,749,550 personas, subrayando la concentración del desempleo en sectores específicos y en ciertas regiones (**Tabla 8**).

Tabla 8: Resumen estadístico de las variables socioeconómicas de los años 2009 y 2020.

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
GENERAL	0.000	10	54	1604	324	1739449
AGRARIO	0.00	5	5	94	25	12076
MAR	0.000	0.000	0.000	6.078	0.000	3087.000
HOGAR	0.0	0.0	5.0	40.2	7.0	70170.0
AUTONOMOS	0.0	16.0	49.0	366.2	183.0	194069.0
CARBON	0.0000	0.0000	0.0000	0.5231	0.0000	2817.0000
TOTAL	0.000	28	108	1959	525	2007040
deuda_miles_euros	0	0	2	2721	289	6776856
TCV	-43.4705	0.0000	0.0000	0.0386	0.0000	78.8918
total_paro	0.000	64	313	4781	1689	2311807
paro_hombre_25	0.0	1.0	17.0	275.8	101.0	132307.0
paro_hombre_25_45	0	11	59	1077	333	682914
paro_hombre_45	0.0	15.0	69.0	868.1	334.0	476820.0
paro_mujer_25	0.0	0.0	13.0	220.1	82.0	99527.0
paro_mujer_25_45	0	13	72	1257	428	612737
paro_mujer_45	0.0	12.0	65.0	1083.1	345.8	595434.0
paro_agricultura	0.0	0.0	14.0	183.6	77.0	30450.0
paro_industria	0.0	4.0	32.0	555.1	199.0	171603.0
paro_construccion	0.0	5.0	36.0	669.2	220.0	392690.0
paro_servicios	0.0	38.0	180.0	3014.7	936.8	1749550.0
paro_sin_empleo_anterior	0	0	16	358	93	142616

Los datos analizados reflejan una distribución desigual y altamente asimétrica en la mayoría de las variables, con una clara presencia de valores extremos que sesgan las medias hacia valores altos. Esto sugiere una concentración significativa de la actividad económica, la deuda municipal y el desempleo en un pequeño número de municipios o regiones.

La estandarización logarítmica es una técnica útil en el análisis de datos cuando se trabaja con variables que presentan valores extremadamente distantes. Es útil para reducir la asimetría y acercar la distribución a la normalidad, mitigar el impacto de valores extremos en el análisis, traducir los datos a una escala más manejable y comparable y facilitar la interpretación de los resultados en términos de elasticidades.

Así pues, tomando la estandarización logarítmica, se generaron los siguientes gráficos:

3.1.1. Diagrama de cajas y bigotes

VARIABLES DEL CENSO AGRARIO

- *Superficies agrarias*

La **Figura 12** muestra a un conjunto de diagramas de caja y bigotes (boxplots) que desvela la distribución logarítmica de diversas categorías de superficies agrarias (SAU y ST) en diferentes rangos de tamaño a lo largo de los años 1982, 1989, 1999, 2009 y 2020.

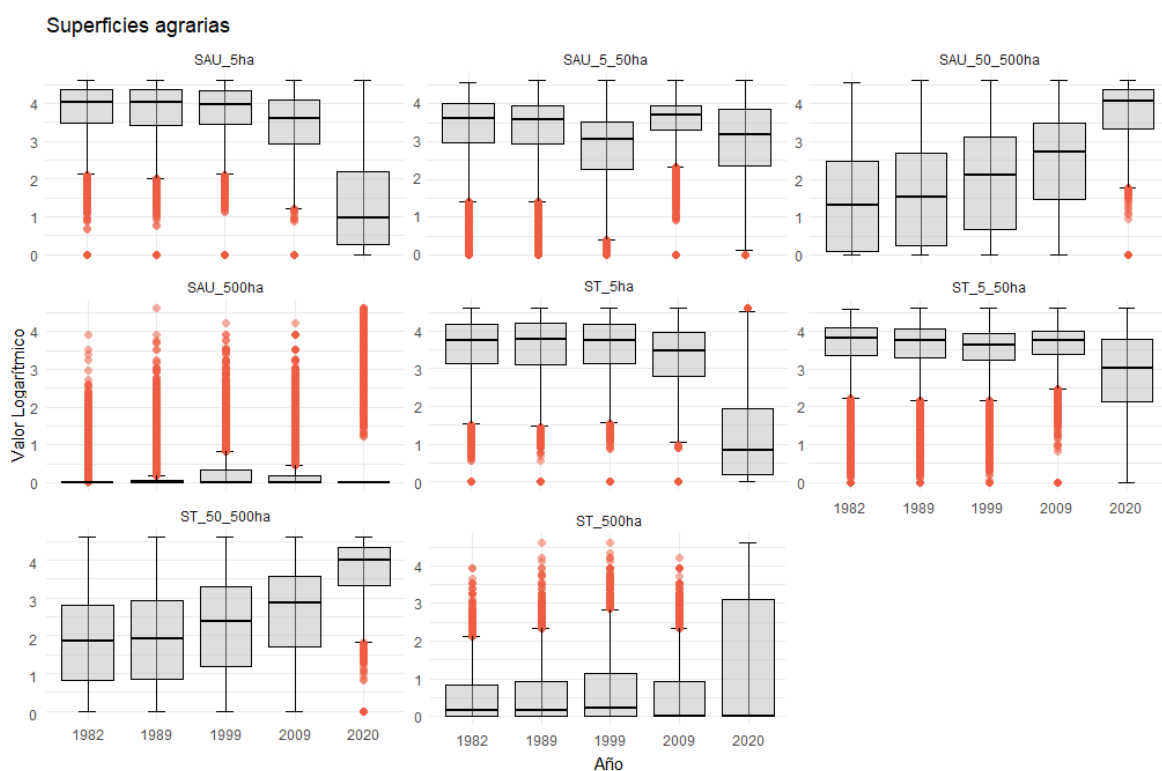


Figura 12: Diagrama de cajas y bigotes (boxplot) para las superficies agrarias en los años 1982-89-99-09-20.

En general, las superficies menores de 5 hectáreas (SAU_5ha y ST_5ha) presentan una mayor dispersión de valores en 2020, lo que sugiere una mayor variabilidad en estas pequeñas explotaciones. Las categorías intermedias (SAU_5_50ha y ST_50_500ha) también muestran un aumento en la variabilidad y en la presencia de valores atípicos en 2020, lo que podría indicar la coexistencia de explotaciones de tamaños muy dispares.

Las superficies mayores de 500 hectáreas, por su parte, tienden a concentrarse en valores logarítmicos bajos, pero en 2020 también se observa un incremento en la dispersión, especialmente en ST_500ha. Esto sugiere posibles cambios en la distribución de las grandes explotaciones agrarias.

En todas las categorías, los valores atípicos son frecuentes, particularmente en SAU_500ha y ST_500ha, lo que indica la existencia de superficies agrarias significativamente diferentes de la tendencia central.

Este análisis sugiere que, aunque la distribución general de las superficies agrarias ha sido relativamente estable, 2020 muestra un aumento en la variabilidad y en la presencia de explotaciones extremas, tanto pequeñas como grandes.

- *Edad y estudios de los jefes de explotación*

En cuanto al porcentaje de jefes de explotación mayores de 55 años, hay una tendencia al aumento de jefes jóvenes hasta 2009. En el año 2020 vuelve a subir este % de jefes mayores. En cambio, el % de jefes sin estudios decrece, es decir, cada vez los jefes están académicamente más formados (**Figura 13**).

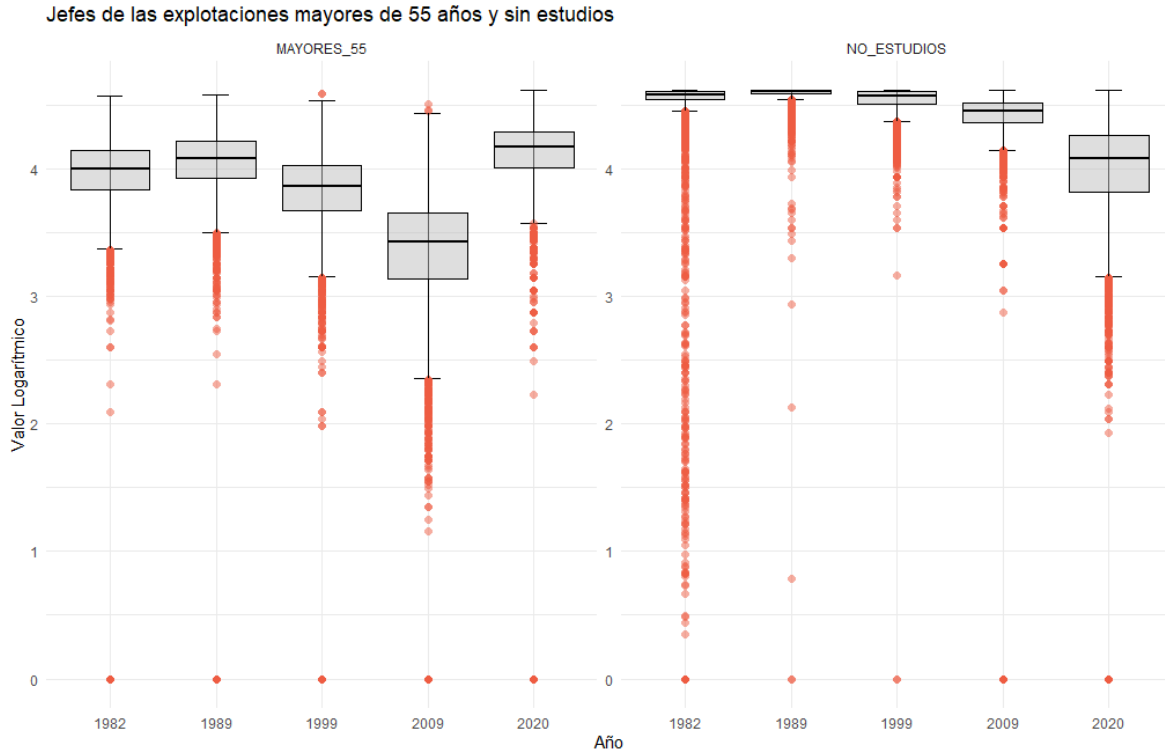


Figura 13: Diagrama de cajas y bigotes (boxplot) para el % de jefes de explotación mayores de 55 y sin estudios en los años 1982-89-99-09-20.

- *Sistema de régimen de tenencia de las explotaciones*

Por otra parte, la **Figura 14** muestra la tendencia del sistema de tenencia de las explotaciones a lo largo de estos años de estudio. En general, el arrendamiento es el sistema que ha aumentado en los dos últimos años (2009-2020), al contrario que la propiedad, que ha disminuido justo en esa etapa. El otro sistema de régimen de tenencia; en el que se incluyen el arrendamiento, alquiler, etc, se mantiene en el tiempo excepto para 2020, que decrece.

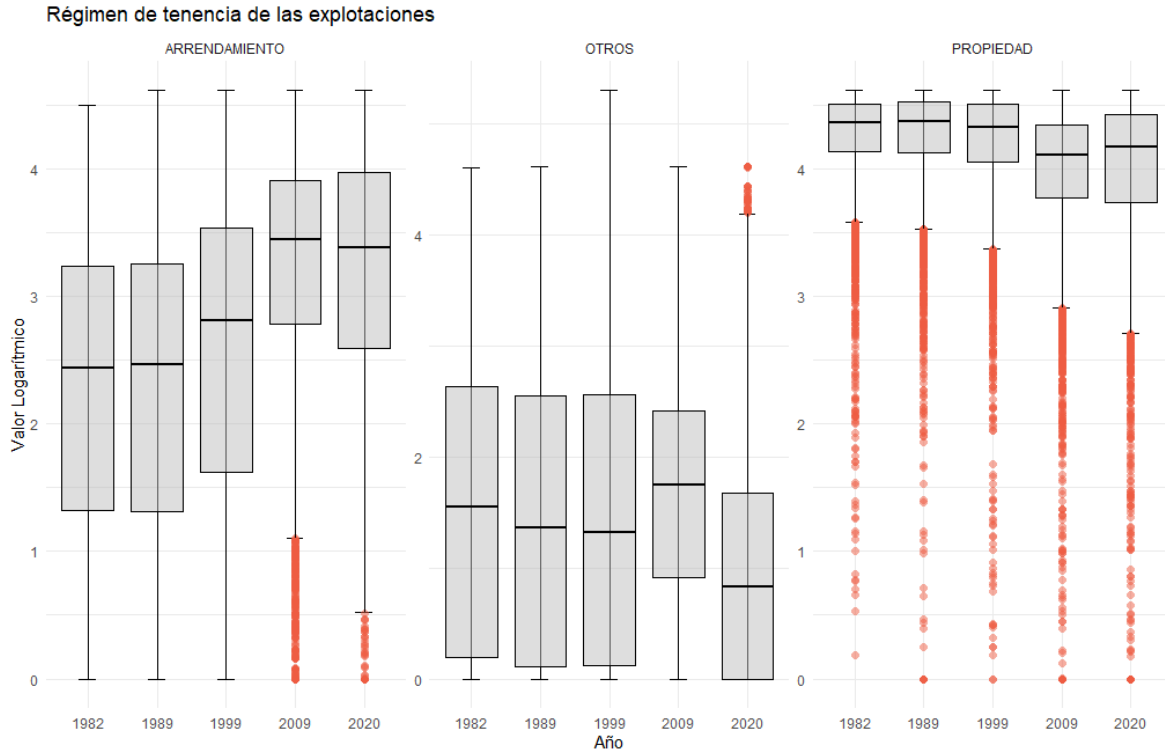


Figura 14: Diagrama de cajas y bigotes (boxplot) para los diferentes sistemas de régimen de tenencia de las explotaciones en los años 1982-89-99-09-20.

- *Densidad de explotaciones y de unidades ganaderas*

Las últimas variables del censo agrario analizadas en el Box-plot han sido la densidad de explotaciones y de unidades ganaderas. Las primeras, se mantienen en los tres primeros años (1982-89-99), para luego decrecer en los años 2009-20. La densidad de unidades ganaderas se mantienen a lo largo de los años (**Figura 15**).

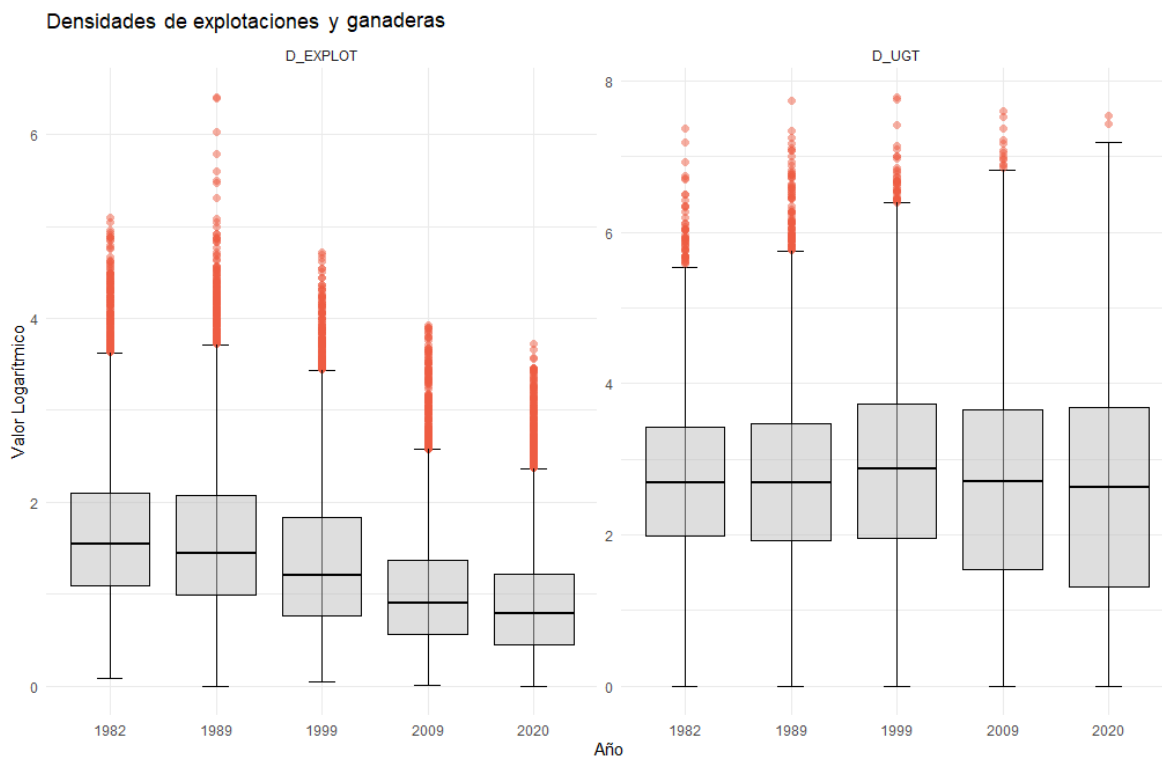


Figura 15: Diagrama de cajas y bigotes (boxplot) para la densidad de explotaciones y de unidades ganaderas en los años 1982-89-99-09-20.

Variables socioeconómicas

- *Afiliados a la Seguridad Social por sectores*

En la **Figura 16** se aprecia las tendencias de las variables socioeconómicas respecto a la afiliación a la Seguridad Social por sectores. No se aprecia una gran variación entre los dos años, aunque sí una ligera tendencia hacia una mayor afiliación en el año 2020.

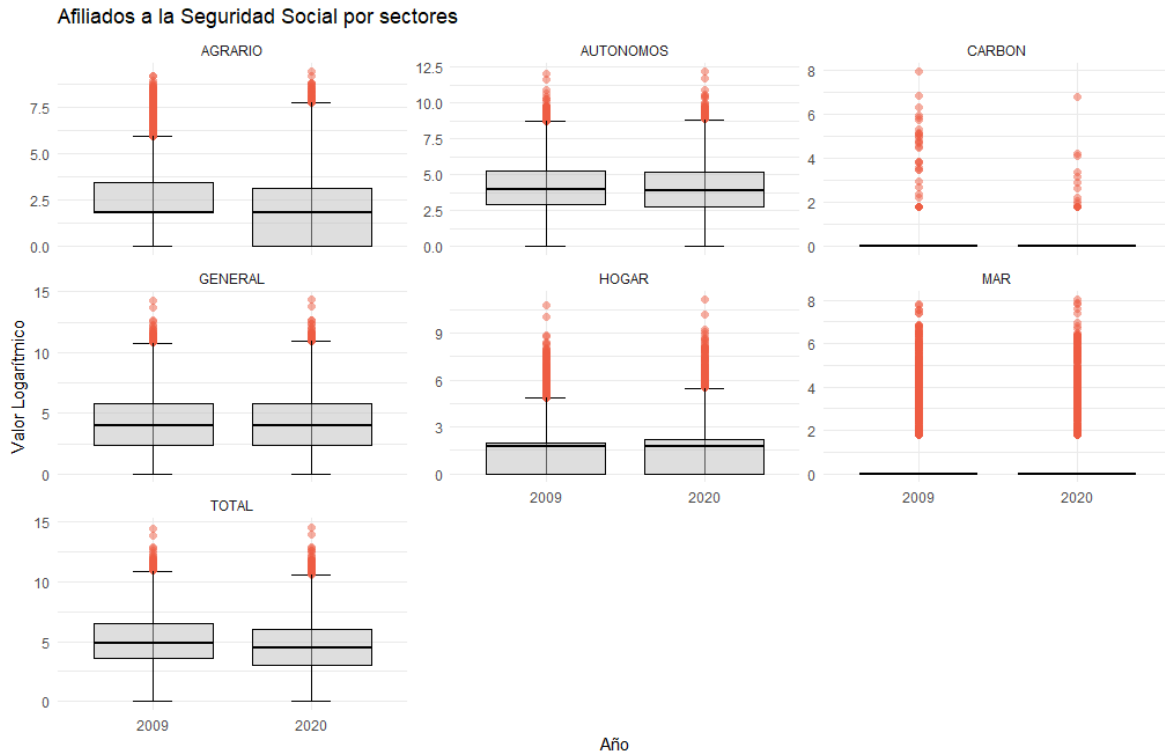


Figura 16: Diagrama de cajas y bigotes (boxplot) para los afiliados a la Seguridad Social de la socioeconomía en los años 2009-20.

■ *Deuda pública de los Ayuntamientos*

La deuda en los ayuntamientos municipales es ligeramente mayor en el año 2009 que en el 2020, gran parte explicada por la crisis económica del 2008 (**Figura 17**).

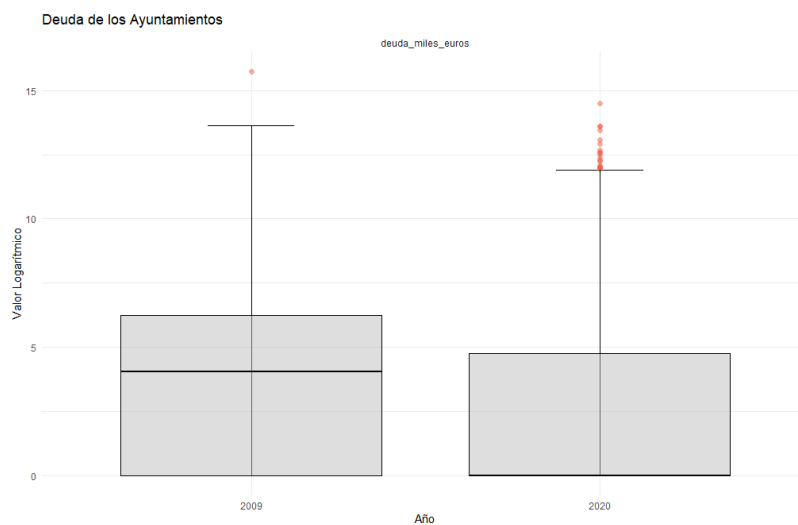


Figura 17: Diagrama de cajas y bigotes (boxplot) para la deuda pública de los Ayuntamientos de la socioeconomía en los años 2009-20.

- *Tasa de Crecimiento Vegetativo*

El Box-plot para esta variable muestra una tendencia muy similar entre ambos años (Figura 18).

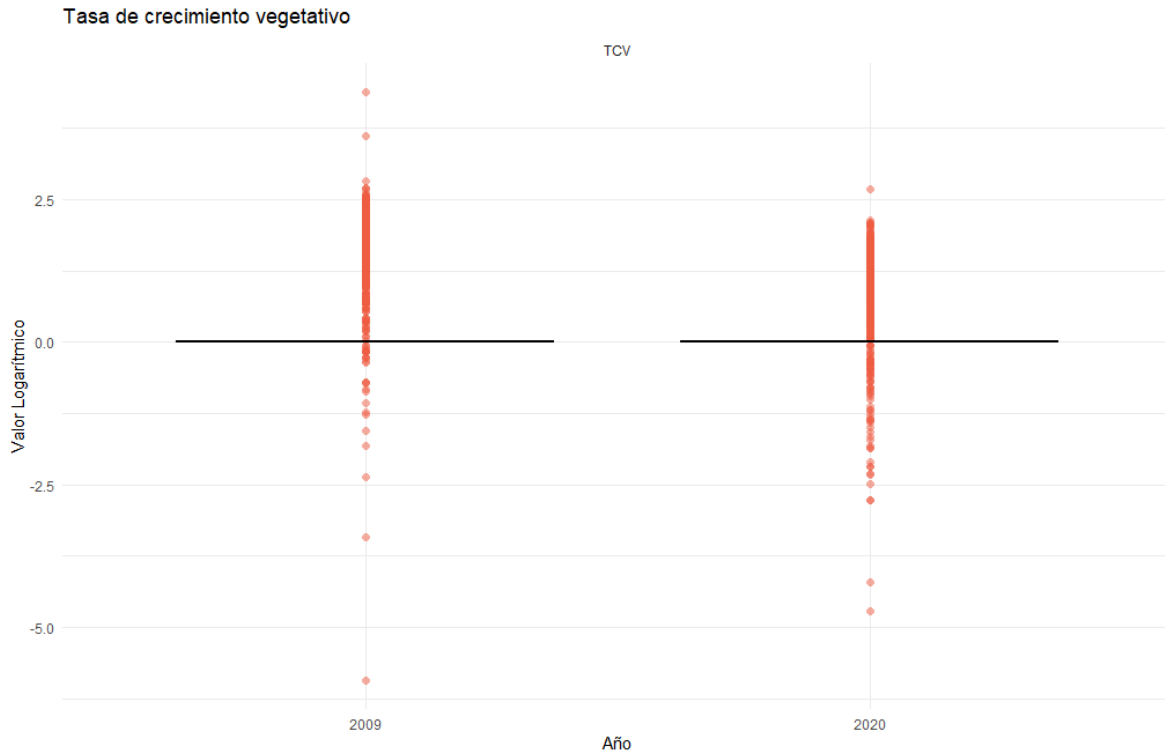


Figura 18: Diagrama de cajas y bigotes (boxplot) para la Tasa de Crecimiento Vegetativo de la socioeconomía en los años 2009-20.

- *Paro por sectores, edad y sexo*

Contrariamente a la afiliación, el paro se muestra en todas las situaciones más alto en el año 2009, siendo el paro en la construcción (paro_construccion) y el paro en los hombres jóvenes (paro_hombre_25_45) variables en donde hay mayor diferencia entre un año y otro. Una mención aparte merece el paro en la agricultura (paro_agricultura), que, contrariamente a lo dispuesto en la línea anterior, es mayor en el año 2020 (Figura 19).

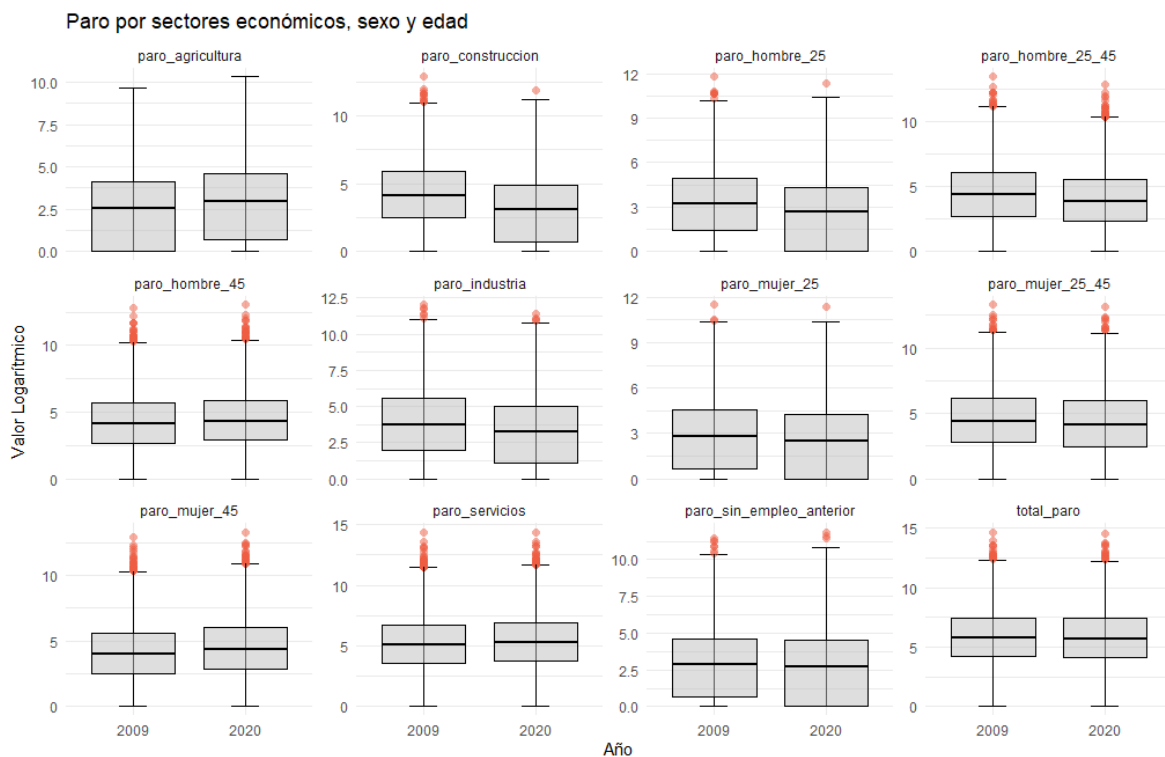


Figura 19: Diagrama de cajas y bigotes (boxplot) para la deuda pública de los Ayuntamientos de la socioeconomía en los años 2009-20.

3.1.2. Gráfico de densidad

- *Variables agrarias*

Las figuras de densidad para las explotaciones agrarias ilustran la distribución logarítmica de diversas categorías de superficies agrarias (SAU y ST) en diferentes períodos: 1982, 1989, 1999, 2009 y 2020. Cada panel corresponde a una categoría específica de superficie, y las curvas de densidad permiten observar cambios en la distribución de las explotaciones agrarias a lo largo del tiempo.

Los municipios con explotaciones menores de 5 hectáreas tienden a disminuir en el año 2020. Por otro lado, las explotaciones de 50 a 500 hectáreas y las mayores de 500 ha, aumentan en este año. Ésto sugiere una expansión espacial de explotaciones de gran tamaño, desapareciendo las explotaciones pequeñas.

Por otro lado, aumenta el % de municipios con alto porcentaje de jefes mayores de 55 años y al contrario para el nivel de estudios: cada vez hay menos municipios con jefes de explotaciones sin estudios.

En cuanto al sistema de propiedad, se confirma el aumento del % de municipios con explotaciones en régimen de arrendamiento y la reducción del % de municipios con explotaciones en régimen de propiedad (**Figura 20, Figura 21, Figura 22**).

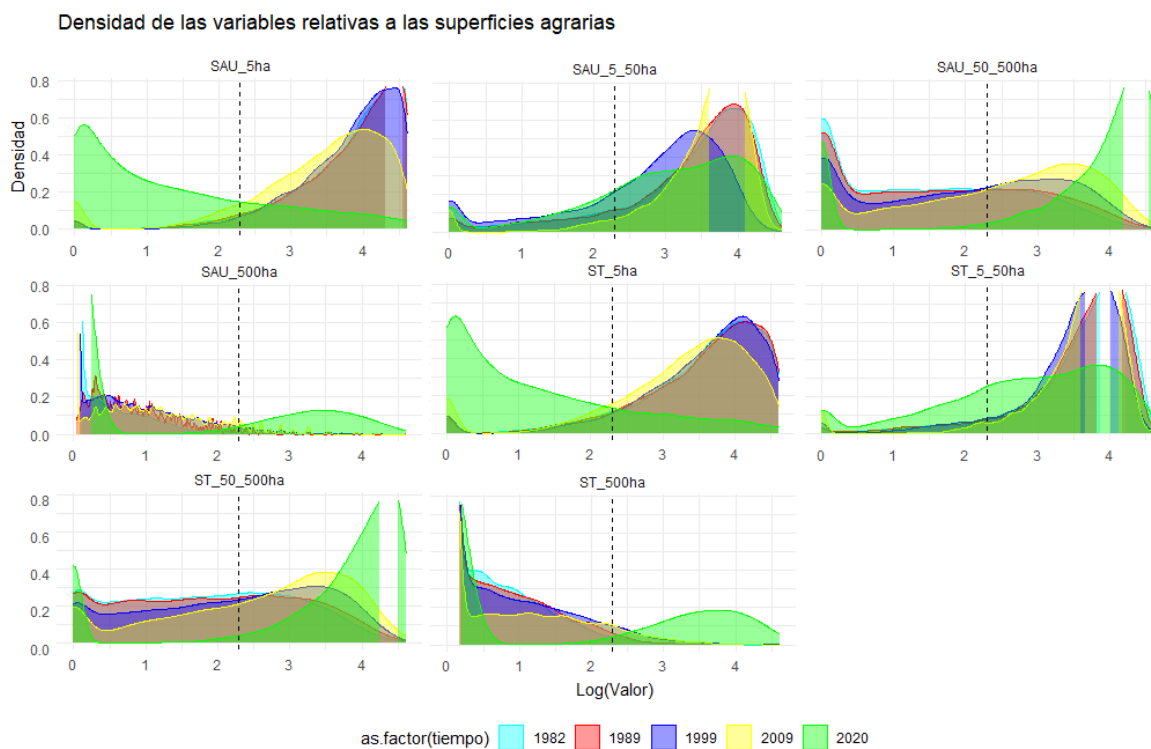


Figura 20: Diagrama de densidad para las superficies agrarias en los años 1982-89-99-09-20.

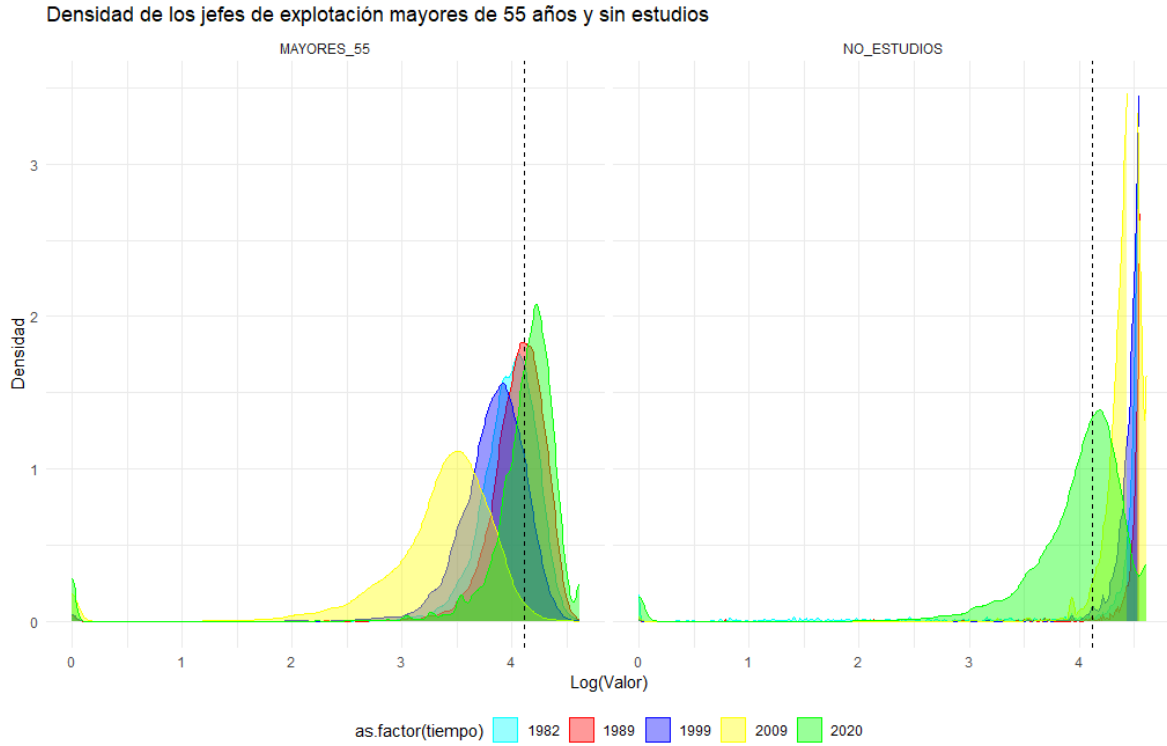


Figura 21: Diagrama de densidad para el % de jefes de explotación mayores de 55 años y % de jefes sin formación en los años 1982-89-99-09-20.

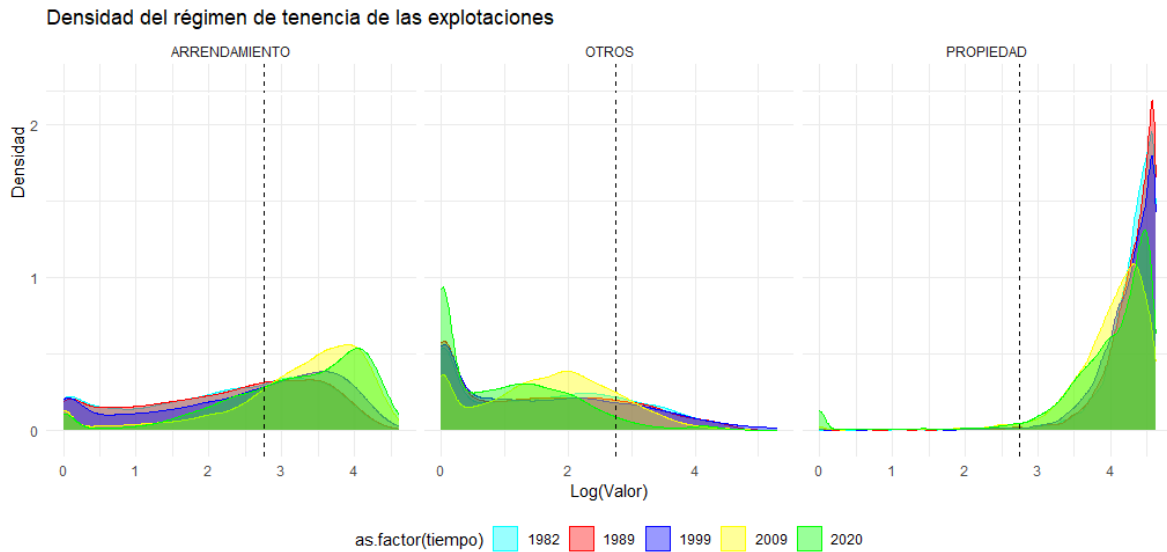


Figura 22: Diagrama de densidad para los distintos regímenes de tenencia en los años 1982-89-99-09-20.

- *Variables socioeconómicas*

En los diagramas de densidad socio-económicos, hay dos variables que difieren en el tiempo: deuda en los Ayuntamientos y Crecimiento Vegetativo (TCV). En ambas, hay un mayor valor estandarizado en 2009 con respecto a 2020. La deuda es más alta en el año 2009 debido a la crisis económica que irrumpió en España. En cuanto al gráfico de densidad relativo a TCV, en el año 2009 hubo un mayor crecimiento poblacional, o, un menor número de fallecimientos. Éste resultado se puede explicar a través de la pandemia de Covid-19, sufrida desde el año 2020 hasta principios del 2022. Durante esta pandemia, la mortalidad se disparó durante unos meses, llegando el 31 de mayo de 2020 al pico de nº de fallecidos por esta enfermedad. En el resto de variables estudiadas, los diagramas de densidad muestran una tendencia similar entre períodos (**Figura 23**, **Figura 24**, **Figura 25**, **Figura 26**).

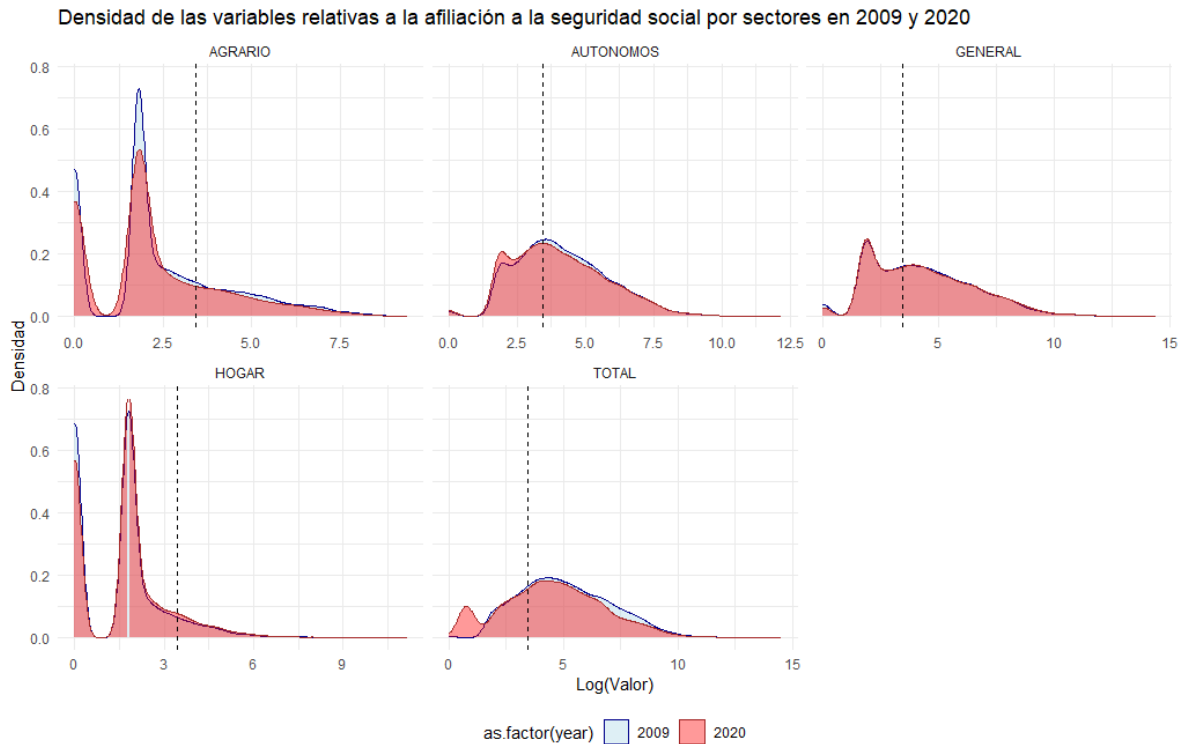


Figura 23: Diagrama de densidad de la afiliación a la Seguridad Social por sectores en la socioeconomía en los años 2009-20.

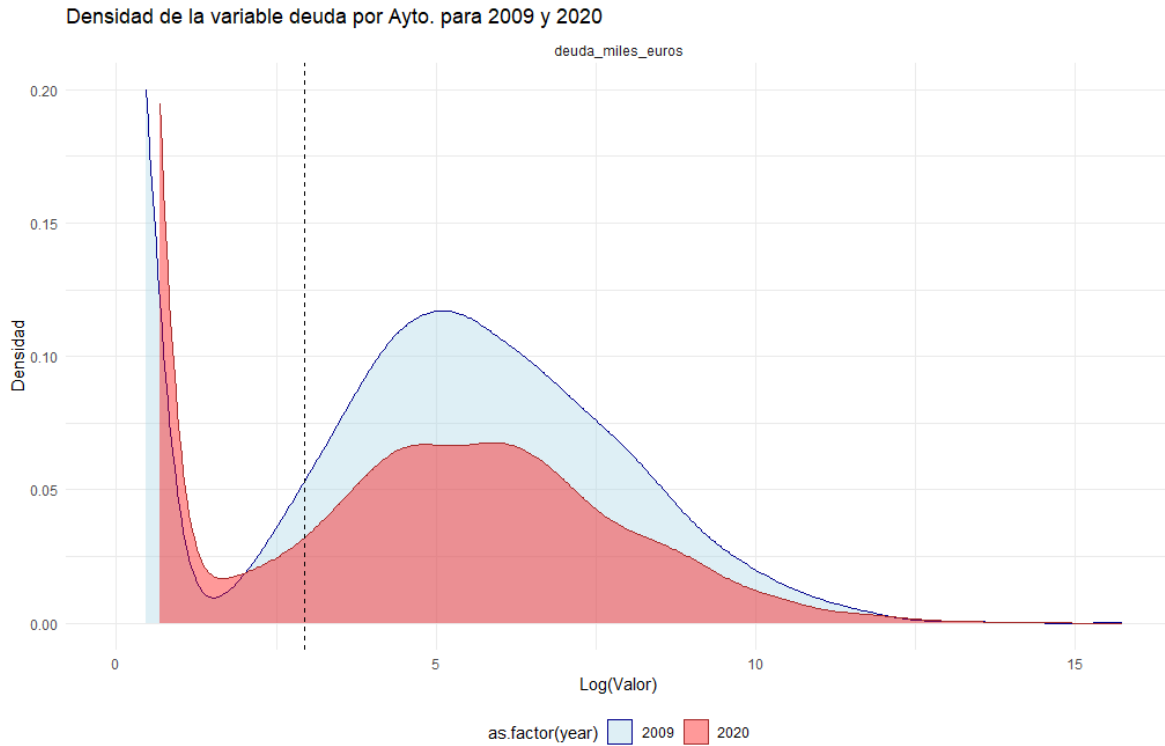


Figura 24: Diagrama de densidad para la deuda pública en los años 2009-20.

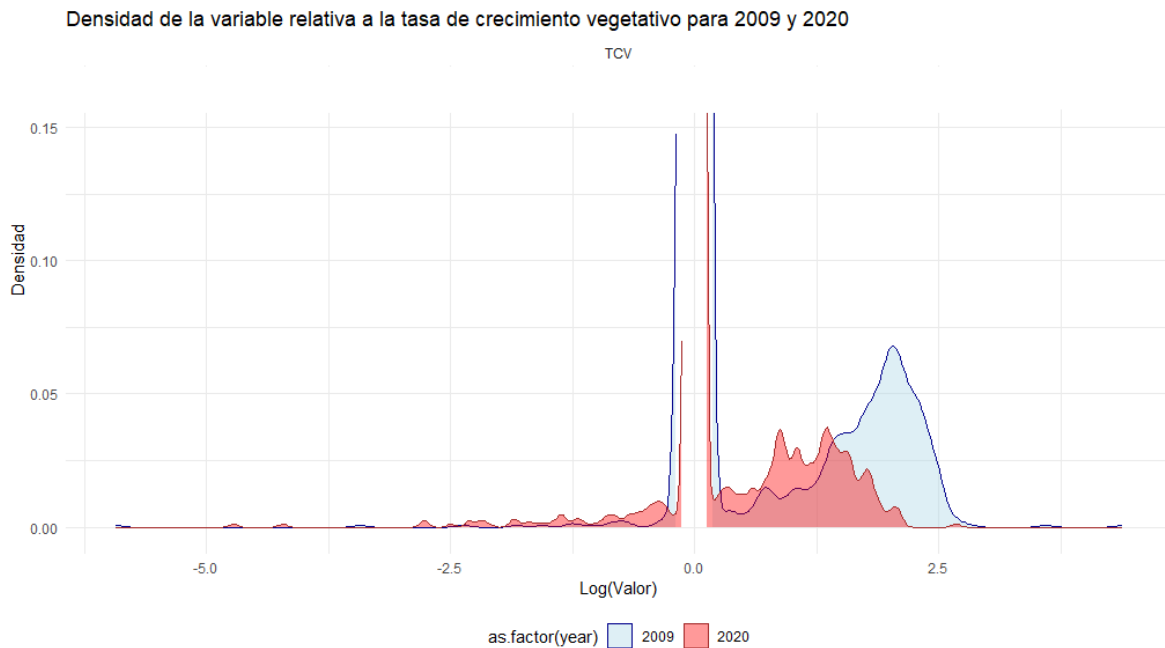


Figura 25: Diagrama de densidad de la Tasa de Crecimiento Vegetativo en la socioeconomía en los años 2009-20.

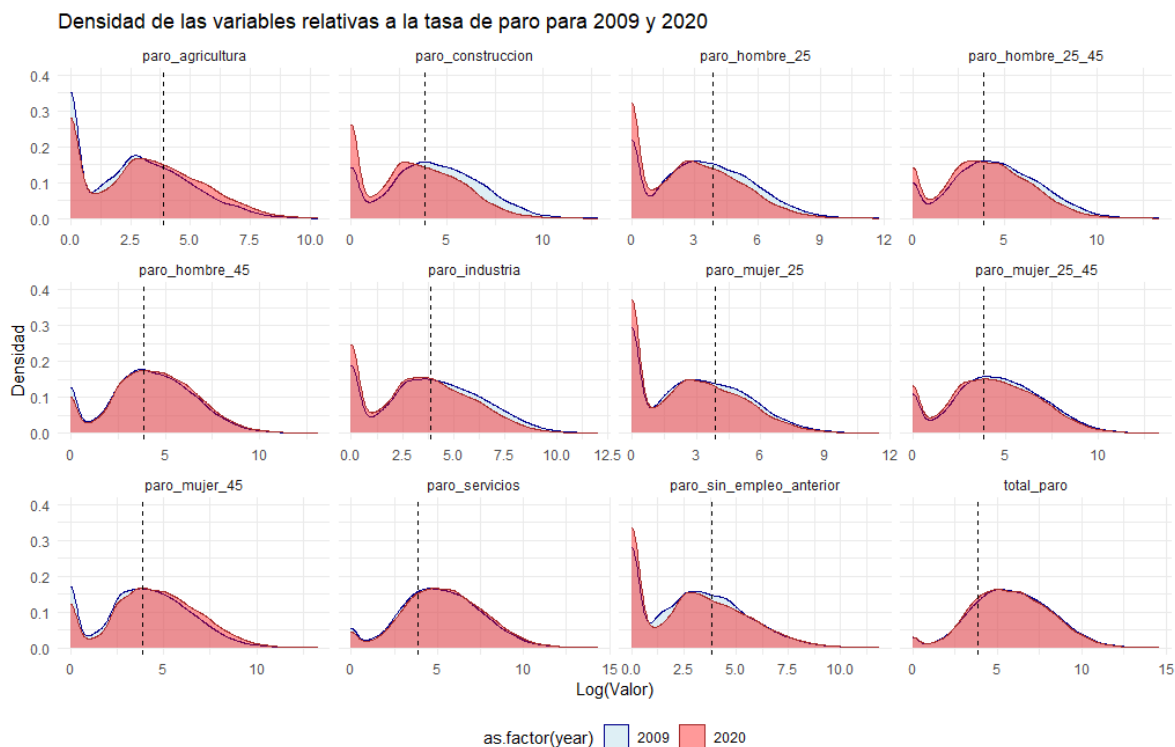


Figura 26: Diagrama de densidad para el paro por sectores, sexo y edad en los años 2009-20.

3.1.3. Análisis de correlación entre variables

El mapa de calor de correlación muestra las relaciones entre las variables estudiadas. Los colores en la matriz indican la magnitud y dirección de la correlación entre pares de variables: los tonos azulados representan correlaciones negativas, mientras que los rojizos denotan correlaciones positivas. Los valores cercanos a 1 o -1 sugieren correlaciones fuertes, mientras que los valores cercanos a 0 indican correlaciones débiles o inexistentes.

Variables del censo agrario

En las variables referentes a las superficies agrarias, SAU_50_500ha y ST_50_500ha mostraron la correlación positiva más fuerte (0,93), sugiriendo que a medida que la superficie agraria utilizada (SAU), también lo hace la superficie total (ST) en el mismo rango de tamaño. Este patrón también se observa entre ST_5ha y SAU_5ha (0,93). Al contrario que en áreas con pequeñas superficies de explotación (ST_5ha), que es menos probable encontrar grandes extensiones de explotación (ST_50_500ha) (-0,69). De manera similar ocurre entre SAU_5ha y SAU_50_500ha (-0,72).

Por otro lado, la correlación entre MAYORES_55 y PROPIEDAD (0.28) sugiere una relación positiva débil entre el envejecimiento de la población y la propiedad de las explotaciones.

La variable relativa al % de jefes sin estudios superiores (NO_ESTUDIOS) muestra una correlación positiva moderada con ST_5ha (0.45), lo que podría indicar que en áreas con menor nivel educativo, es más común encontrar pequeñas superficies de explotación.

Algunas variables, como D_UGT, presentan correlaciones muy bajas con otras, indicando una baja relación lineal (**Figura 27, Tabla 9**).

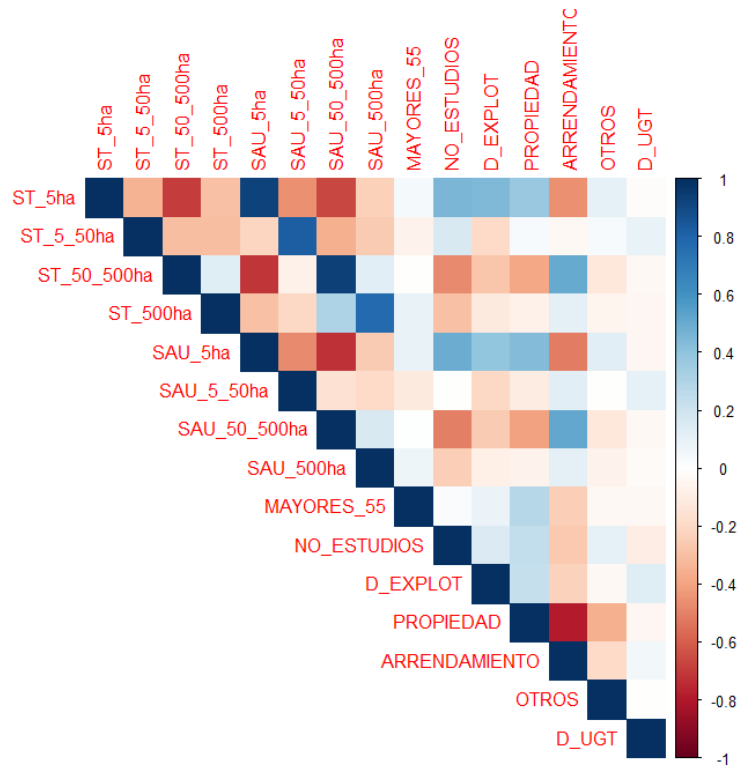


Figura 27: Mapa de calor de correlación para las variables del censo agrario en los años 1982-89-99-09-20.

Tabla 9: Matriz de correlaciones entre variables agrarias.

Variable	ST_5ha	ST_5_50ha	ST_50_500ha	ST_500ha	SAU_5ha	SAU_5_50ha	SAU_50_500ha
ST_5ha	1.0000	-0.3448	-0.6911	-0.2901	0.9290	-0.4558	-0.6692
ST_5_50ha	-0.3448	1.0000	-0.3094	-0.3059	-0.2181	0.8204	-0.3526
ST_50_500ha	-0.6911	-0.3094	1.0000	0.1357	-0.7159	-0.0781	0.9341
ST_500ha	-0.2901	-0.3059	0.1357	1.0000	-0.2953	-0.2046	0.3029
SAU_5ha	0.9290	-0.2181	-0.7159	-0.2953	1.0000	-0.4795	-0.7284
SAU_5_50ha	-0.4558	0.8204	-0.0781	-0.2046	-0.4795	1.0000	-0.1514
SAU_50_500ha	-0.6692	-0.3526	0.9341	0.3029	-0.7284	-0.1514	1.0000
SAU_500ha	-0.2303	-0.2545	0.1271	0.7729	-0.2521	-0.1918	0.1639
MAYORES_55	0.0418	-0.0667	-0.0006	0.0932	0.0999	-0.1191	0.0057
NO_ESTUDIOS	0.4546	0.1679	-0.4734	-0.2999	0.4957	-0.0073	-0.5002
D_EXPLOIT	0.4445	-0.1948	-0.2765	-0.1198	0.3905	-0.2095	-0.2575
PROPIEDAD	0.3795	0.0332	-0.3899	-0.0791	0.4339	-0.1005	-0.4035
ARRENDAMIENTO	-0.4578	-0.0316	0.5054	0.1190	-0.5187	0.1204	0.5193
OTROS	0.1097	0.0332	-0.1218	-0.0546	0.1205	-0.0089	-0.1259
D_UGT	-0.0168	0.0926	-0.0330	-0.0491	-0.0422	0.1033	-0.0357

VARIABLES SOCIOECONÓMICAS

En las variables socioeconómicas, se observó que las correlaciones más fuertes se encuentran entre aquellas relacionadas con el empleo. Por ejemplo, el total de paro muestra una fuerte correlación positiva con el paro en servicios (0.99). De manera similar, el paro en hombres de 45 años o más también tiene una alta correlación con el total de paro (0.99), sugiriendo que el desempleo en este grupo de edad contribuye significativamente al desempleo general.

En contraste, variables como TCV (Tasa de Crecimiento Vegetativo) tienen correlaciones bajas o negativas con la mayoría de las otras variables, como su relación negativa con el paro en mujeres mayores de 45 años (-0.06), lo que indica una relación lineal débil.

VARIABLES COMO CARBÓN TIENEN CORRELACIONES MUY BAJAS CON OTRAS VARIABLES, COMO CON EL TOTAL DE PARO (0.05), INDICANDO UNA BAJA RELACIÓN LINEAL ENTRE LA ACTIVIDAD MINERA DEL CARBÓN Y EL DESEMPEÑO GENERAL. ESTO SUGIERE QUE LA INFLUENCIA DE ESTE SECTOR EN EL DESEMPEÑO GENERAL ES MÍNIMA EN COMPARACIÓN CON OTROS SECTORES (**Figura 28, Tabla 10**).

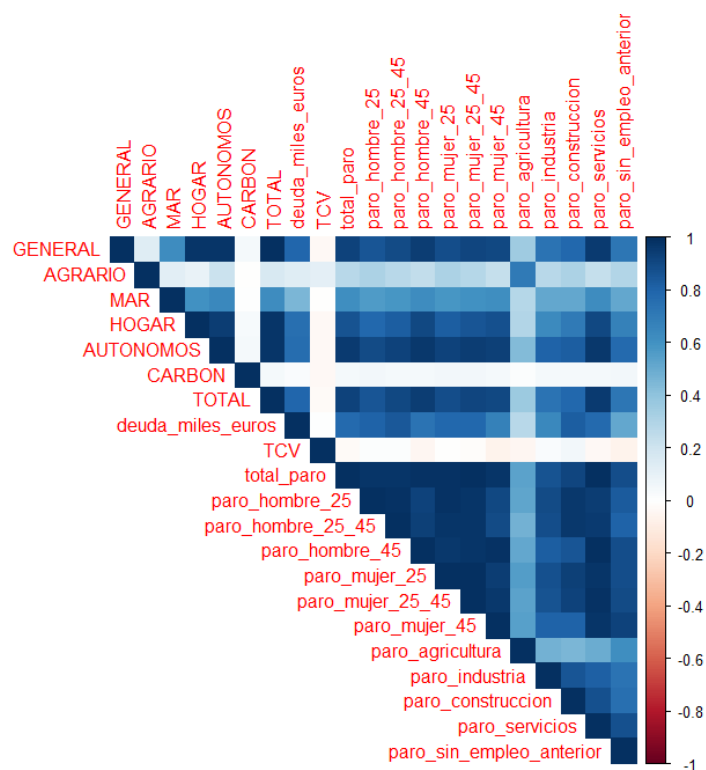


Figura 28: Mapa de calor de correlación para las variables socioeconómicas en los años 2009-20.

Tabla 10: Matriz de correlaciones entre variables socioeconómicas.

Variable	GENERAL	AGRARIO	MAR	HOGAR	AUTONOMOS	CARBON	TOTAL	deuda_miles_euros	TCV
GENERAL	1.0000	0.1383	0.6232	0.9769	0.9789	0.0459	0.9971	0.7950	-0.0349
AGRARIO	0.1383	1.0000	0.1282	0.0965	0.2171	-0.0015	0.1603	0.1338	0.1126
MAR	0.6232	0.1282	1.0000	0.6019	0.6406	0.0098	0.6292	0.4581	-0.0094
HOGAR	0.9769	0.0965	0.6019	1.0000	0.9413	0.0305	0.9741	0.7520	-0.0302
AUTONOMOS	0.9789	0.2171	0.6406	0.9413	1.0000	0.0493	0.9797	0.7626	-0.0336
CARBON	0.0459	-0.0015	0.0098	0.0305	0.0493	1.0000	0.0470	0.0218	-0.0337
TOTAL	0.9971	0.1603	0.6292	0.9741	0.9797	0.0470	1.0000	0.7928	-0.0280
deuda_miles_euros	0.7950	0.1338	0.4581	0.7520	0.7626	0.0218	0.7928	1.0000	0.0056
TCV	-0.0349	0.1126	-0.0094	-0.0302	-0.0336	-0.0337	-0.0280	0.0056	1.0000
total_paro	0.9231	0.2723	0.6153	0.8683	0.9543	0.0481	0.9254	0.7796	-0.0220
paro_hombre_25	0.8554	0.3147	0.5636	0.7822	0.8984	0.0517	0.8598	0.8035	0.0186
paro_hombre_25_45	0.8967	0.2753	0.5894	0.8223	0.9249	0.0481	0.9001	0.8428	0.0154
paro_hombre_45	0.9400	0.2470	0.6272	0.9027	0.9656	0.0415	0.9414	0.7362	-0.0409
paro_mujer_25	0.8846	0.3103	0.5814	0.8255	0.9217	0.0512	0.8881	0.7849	-0.0088
paro_mujer_25_45	0.9103	0.2852	0.6076	0.8502	0.9427	0.0537	0.9130	0.7889	-0.0146
paro_mujer_45	0.9041	0.2392	0.6128	0.8702	0.9358	0.0424	0.9048	0.6748	-0.0633
paro_agricultura	0.3512	0.7082	0.2852	0.2943	0.4311	0.0133	0.3658	0.2741	-0.0421
paro_industria	0.7301	0.2705	0.5110	0.6372	0.8065	0.0435	0.7356	0.6407	0.0235
paro_construccion	0.7836	0.3151	0.5154	0.7008	0.8220	0.0469	0.7886	0.8230	0.0587
paro_servicios	0.9508	0.2305	0.6271	0.9092	0.9696	0.0461	0.9515	0.7799	-0.0369
paro_sin_empleo_anterior	0.7233	0.2945	0.5138	0.6743	0.7717	0.0551	0.7274	0.5185	-0.0606

3.2. Estadística multivariante

3.2.1. STATIS-Dual

El STATIS-Dual se llevó a cabo en 5 k matrices correspondientes a los distintos años, con el fin; como se comentó anteriormente, de conocer la evolución de la estructura de covariación entre las variables agrarias en los años sometidos a estudio (**Figura 29**).

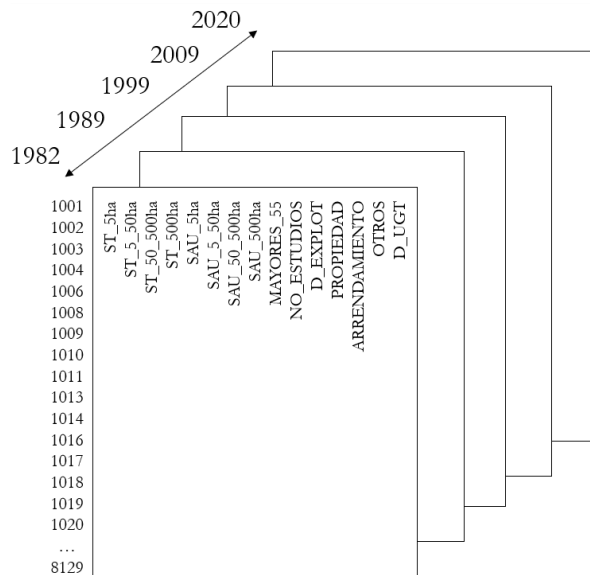


Figura 29: Matrices primarias utilizadas para la realización del análisis STATIS-Dual.

El resultado de este análisis reveló varios hallazgos clave que reflejan la evolución temporal de la estructura agraria.

Los coeficientes de correlación vectorial (RV) mostraron un alto grado de similitud entre las matrices de datos de los años consecutivos, especialmente entre 1982 y 1989 (0.987) y entre 1989 y 1999 (0.981), lo que sugiere una estabilidad en la estructura de las variables agrarias durante esos períodos. Sin embargo, se observó una disminución notable en la correlación en 2020 (0.710 con 1982), lo que indicó un cambio significativo en la estructura de los datos agrarios en ese año en comparación con los anteriores.

Los valores de coseno cuadrado (\cos^2) reflejan la calidad de la representación de las matrices originales en el espacio de compromiso, estando las matrices 1982-89-99-09 bien representadas en el compromiso, con valores (\cos^2) superiores a 0.87, mientras que los datos de 2020 tienen una representación más débil (0.728).

Tabla 11: Coeficientes de correlación vectorial (RV) entre los años 1982-89-99-09-20 del STATIS-Dual.

Año	1982	1989	1999	2009	2020
1982	1.0000	0.9874	0.9738	0.9476	0.7101
1989	0.9874	1.0000	0.9807	0.9613	0.7284
1999	0.9738	0.9807	1.0000	0.9753	0.7446
2009	0.9476	0.9613	0.9753	1.0000	0.8445
2020	0.7101	0.7284	0.7446	0.8445	1.0000

Tabla 12: Valores de coseno cuadrado (\cos^2) del STATIS-Dual de los años 1982-89-99-09-20.

Índice	Cos²
1	0.8769
2	0.8894
3	0.9073
4	0.9231
5	0.7283

Tabla 13: Pesos para las k-tablas (RV.tabw) del STATIS-Dual de los años 1982-89-99-09-20.

Índice	Peso
1	0.4555
2	0.4592
3	0.4607
4	0.4647
5	0.3917

Tabla 14: Valores propios para la inter-estructura (RV.eig) del STATIS-Dual de los años 1982-89-99-09-20.

Índice	Valor propio
1	4.5574
2	0.3868
3	0.0348
4	0.0119
5	0.0092

Tabla 15: Valores propios para el compromiso (C.eig) del STATIS-Dual de los años 1982-89-99-09-20.

Índice	Valor propio
1	1.6591
2	0.8002
3	0.6092
4	0.4875
5	0.4352
6	0.3444
7	0.3146
8	0.2673
9	0.2349
10	0.1151
11	0.0688
12	0.0518
13	0.0335
14	0.0145
15	0.0037

De esta manera, concordando con autores como Sim et al. (2006), se eliminó del análisis la K-tabla correspondiente al año 2020 y se realizó de nuevo el STATIS-Dual. El análisis se rehízo sin el año 2020 ya que indicaba una situación atípica en el panorama agrario y, en este caso, lo que se busca es que el compromiso represente mejor a todos los años.

Esta situación, fue consecuencia de la pandemia Covid-19 sufrida desde febrero de 2020 hasta principios de 2022. En un corto período de tiempo, llegó a afectar a todos los procesos socio-económicos, tales como la producción de bienes y servicios. El desempleo afectó a muchas familias, generando trastornos como la ansiedad, el estrés y depresión, moldeando la conducta de los individuos (Chaguay et al., 2020). Además, se evidenció que, en el gobierno central y otras instituciones, persistía una visión urbanocéntrica (Gutiérrez, 2020), priorizando el modelo agroalimentario convencional y ocasionando cambios en la estructura agraria española; sobre todo en explotaciones pequeñas o familiares

Si se quisieran buscar cambios, sería interesante analizar las cinco tablas de datos. Se obtuvieron así los siguientes resultados:

STATIS-Dual con los años 1982-1989-99-09

Este nuevo análisis mostró una alta correlación vectorial entre los años: los coeficientes RV oscilaron entre 0.948 y 1.000, indicando una gran similitud entre las estructuras de las diferentes etapas.

Los valores de \cos^2 indicaron que la calidad de la representación en el espacio de compromiso fue alta, con valores comprendidos entre 0.895 y 0.918.

En cuanto a los pesos asignados a las k-tablas, estos fueron bastante uniformes, con valores que oscilaron entre 0.496 y 0.502, lo que implica que cada año contribuyó de manera similar al compromiso global.

Los valores propios para la inter-estructura indicaron que la mayor parte de la variabilidad fue capturada en el primer componente, que explicó un valor de 3.913, mientras que los siguientes componentes explicaron una menor proporción de la variabilidad, con valores de 0.059, 0.017 y 0.011 respectivamente.

Por otro lado, los valores propios del compromiso mostraron una distribución decreciente, con los valores más altos en los primeros componentes: 1.597 para el primer componente, 0.666 para el segundo y 0.543 para el tercero (**Tabla 16, Tabla 17, Tabla 18, Tabla 19 y Tabla 20**).

En este caso, el compromiso proyectó las variables en el espacio definido por los dos primeros componentes principales:

- El primer eje correlaciona a las variables ST_5ha, SAU_5ha (superficies pequeñas), y D_EXPLOT (densidad de explotaciones). Además, se observa ligera correlación entre PROPIEDAD, MAYORES_55, y NO_ESTUDIOS, lo que sugiere que estas explotaciones tienden a estar en manos de jefes mayores sin educación formal. Todas estas variables correlacionan negativamente con ARRENDAMIENTO. Esto indica que el régimen de tenencia por arrendamiento es menos común en las explotaciones pequeñas.
- El segundo eje correlaciona a ST_500ha y SAU_500ha (superficies agrícolas de más de 500 hectáreas). SAU_50_500ha y ST_50_500ha covarían negativamente con las grandes superficies ubicadas en la parte superior del eje. Es decir, las superficies agrícolas medianas no tienen las características de las grandes explotaciones.

Las proyecciones de las variables en el espacio de compromiso revelaron que ST_5ha y SAU_5h mostraron altas contribuciones positivas en el primer componente, mientras que ST_-500ha y SAU_-500ha presentaron contribuciones más significativas en el segundo y tercer componente.

Por último, la proyección de componentes es similar al gráfico de la interestructura, aunque éste muestra cómo los diferentes años estudiados; 1982, 1989, 1999 y 2009, se proyectan en el espacio de los componentes principales. La cercanía de los puntos para estos años refuerza la idea de que la estructura de las variables agrarias y socioeconómicas no ha cambiado significativamente a lo largo del tiempo (**Figura 31**).

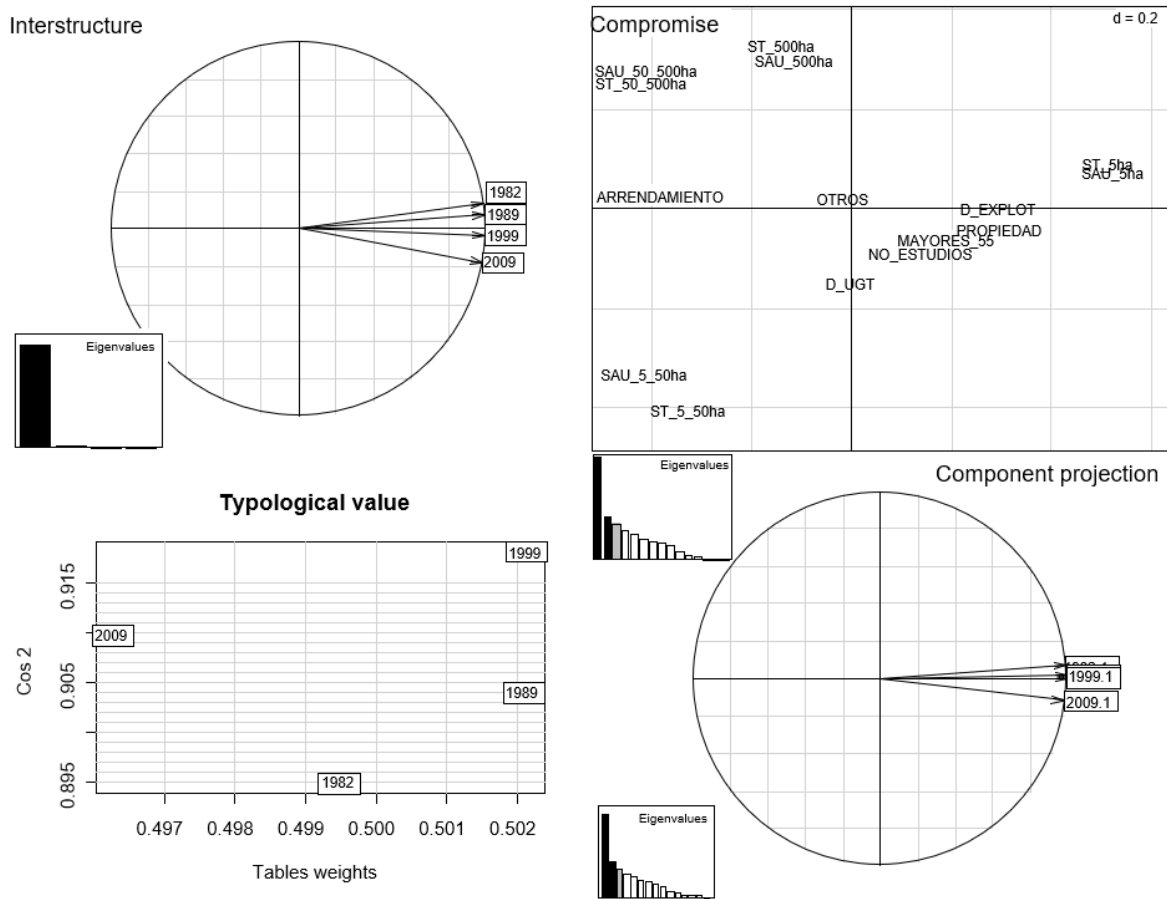


Figura 31: Gráfico resultante del análisis STATIS-Dual en donde se muestra la inter-estructura, el compromiso, los valores \cos^2 y la proyección de los componentes sin el año 2020.

Tabla 16: Coeficientes de correlación vectorial (RV) entre los años 1982, 1989, 1999 y 2009 del STATIS-Dual.

	1982	1989	1999	2009
1982	1.0000	0.9874	0.9738	0.9476
1989	0.9874	1.0000	0.9807	0.9613
1999	0.9738	0.9807	1.0000	0.9753
2009	0.9476	0.9613	0.9753	1.0000

Tabla 17: Valores de coseno cuadrado (\cos^2) del STATIS-Dual para los años 1982, 1989, 1999 y 2009.

Índice	Cos²
1	0.8948
2	0.9040
3	0.9181
4	0.9097

Tabla 18: Pesos para las k-tablas (RV.tabw) del STATIS-Dual para los años 1982, 1989, 1999 y 2009.

Índice	Peso
1	0.4995
2	0.5021
3	0.5021
4	0.4963

Tabla 19: Valores propios para la inter-estructura (RV.eig) del STATIS-Dual para los años 1982, 1989, 1999 y 2009.

Índice	Valor propio
1	3.9131
2	0.0589
3	0.0168
4	0.0112

Tabla 20: Valores propios para el compromiso (C.eig) del STATIS-Dual para los años 1982, 1989, 1999 y 2009.

Índice	Valor propio
1	1.5969
2	0.6657
3	0.5429
4	0.4389
5	0.3882
6	0.2991
7	0.2694
8	0.2463
9	0.2058
10	0.1065
11	0.0658
12	0.0281
13	0.0084
14	0.0059
15	0.0018

3.2.2. Co-Inercia

Las matrices de datos utilizadas para la realización del análisis de co-inercia fueron las bases de datos relativas a las variables del censo agrario y las bases de datos socioeconómicas de los años 2009 y 2010 (**Figura 32**), ya que, como se comentó en el apartado **2. Material y métodos**, sólo se dispuso de datos de estos dos años para las características socioeconómicas.

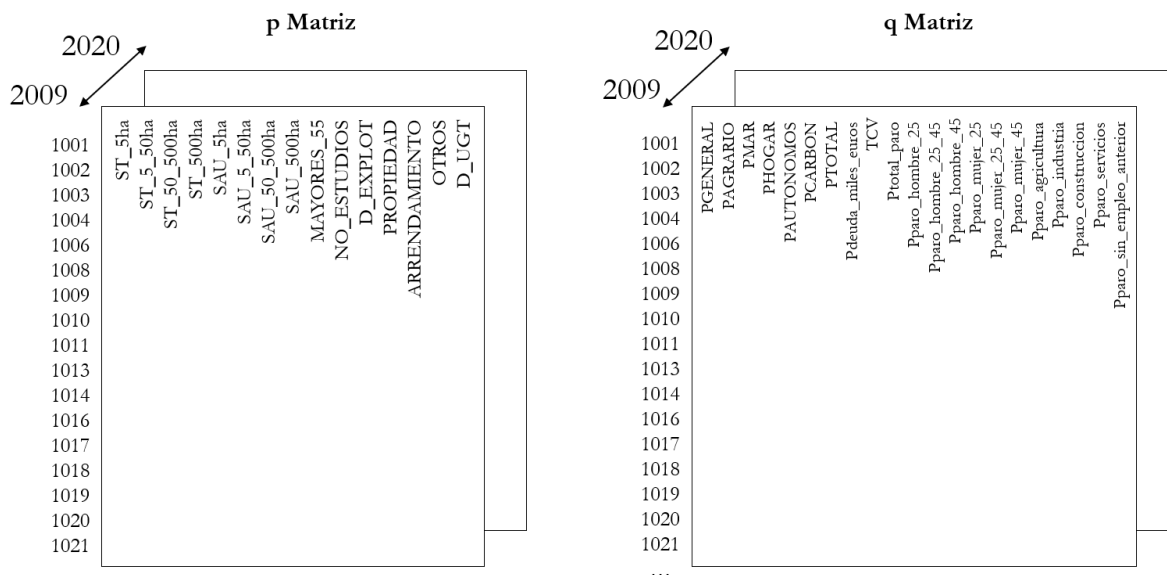


Figura 32: Matrices primarias utilizadas para la realización del análisis de la co-inercia.

Co-inercia para la base de datos del 2009

El análisis realizado reveló una inercia total de 7,382, con la retención de dos dimensiones principales que explicaron el 98,23 % de la variabilidad total, con un 96,40 % atribuido al Eje 1 y un 1,83 % al Eje 2. Las coordenadas principales indicaron que el primer eje posee un eigenvalor de 7,116, mientras que el segundo mostró un valor significativamente menor, 0,135, lo que sugirió que la mayor parte de la variabilidad entre los conjuntos de datos se capturó en el primer eje, con una contribución casi marginal del segundo. A partir del tercer eje, la contribución se volvió marginal, alcanzando una inercia proyectada acumulada del 99,69 % con los primeros cinco ejes. La inercia de las variables socioeconómicas en el primer eje fue de 6,014, con un máximo de 6,435, lo que dio una ratio de 0,935. Para las variables agrarias, la inercia en el primer eje fue de 4,950, con un máximo de 5,010 y una ratio de 0,986. El coeficiente RV, que midió la relación entre las dos tablas analizadas, fue de 0,156, lo que indicó una correlación moderadamente débil entre los conjuntos de datos socio-económicos y agrarios. Específicamente, la correlación en el Eje 1 fue de 0,48894 y en el Eje 2 de 0,23982, reforzando la idea de una asociación limitada entre las dimensiones principales de ambos conjuntos de datos. En el análisis de las contribuciones de las variables socioeconómicas en el Eje 1, ST_5ha, SAU_5ha y PROPIEDAD fueron las variables que más contribuyeron positivamente, mientras que ST_50_500ha, SAU_50_500ha y ARRENDAMIENTO presentaron contribuciones negativas. En el Eje 2, D_EXPLOIT y SAU_5_50ha contribuyeron de forma positiva, en con-

traste con MAYORES_55 y NO_ESTUDIOS, que mostraron contribuciones negativas. En el análisis de las variables agrarias, AGRARIO y paro_agricultura fueron las que más influyeron positivamente en el Eje 1, mientras que MAR y hogar tuvieron una influencia negativa. En el Eje 2, AGRARIO y TCV fueron las variables con mayores contribuciones positivas, mientras que paro_industria y paro_servicios mostraron contribuciones negativas. Estos resultados, junto con la dispersión observada en los gráficos, sugirieron la existencia de outliers, como el municipio 4286, posiblemente debido a características socioeconómicas y agrarias extremas. (Figura 33).

Tabla 21: Resultado del análisis de Co-inercia para la base de datos de 2009.

Parámetro	Eje 1	Eje 2	Eje 3	Eje 4	Eje 5
Inercia Total			7.382		
Valores Propios (Eigenvalues)	7.1164	0.1349	0.0637	0.0288	0.0152
Inercia Proyectada (%)	96.40	1.83	0.86	0.39	0.21
Inercia Acumulada (%)	96.40	98.23	99.09	99.48	99.69
Descomposición de Valores Propios					
Valor Propio	7.1164	0.1349	-	-	-
Covarianza	2.6677	0.3673	-	-	-
Desviación Estándar X	2.4524	1.1614	-	-	-
Desviación Estándar Y	2.2248	1.3187	-	-	-
Correlación	0.4889	0.2398	-	-	-
Inercia y Co-inercia X					
Inercia	6.0144	-	-	-	-
Máxima	6.4352	-	-	-	-
Ratio	0.9346	-	-	-	-
Inercia y Co-inercia Y					
Inercia	4.9496	-	-	-	-
Máxima	5.0198	-	-	-	-
Ratio	0.9860	-	-	-	-
Coefficiente RV			0.1564		

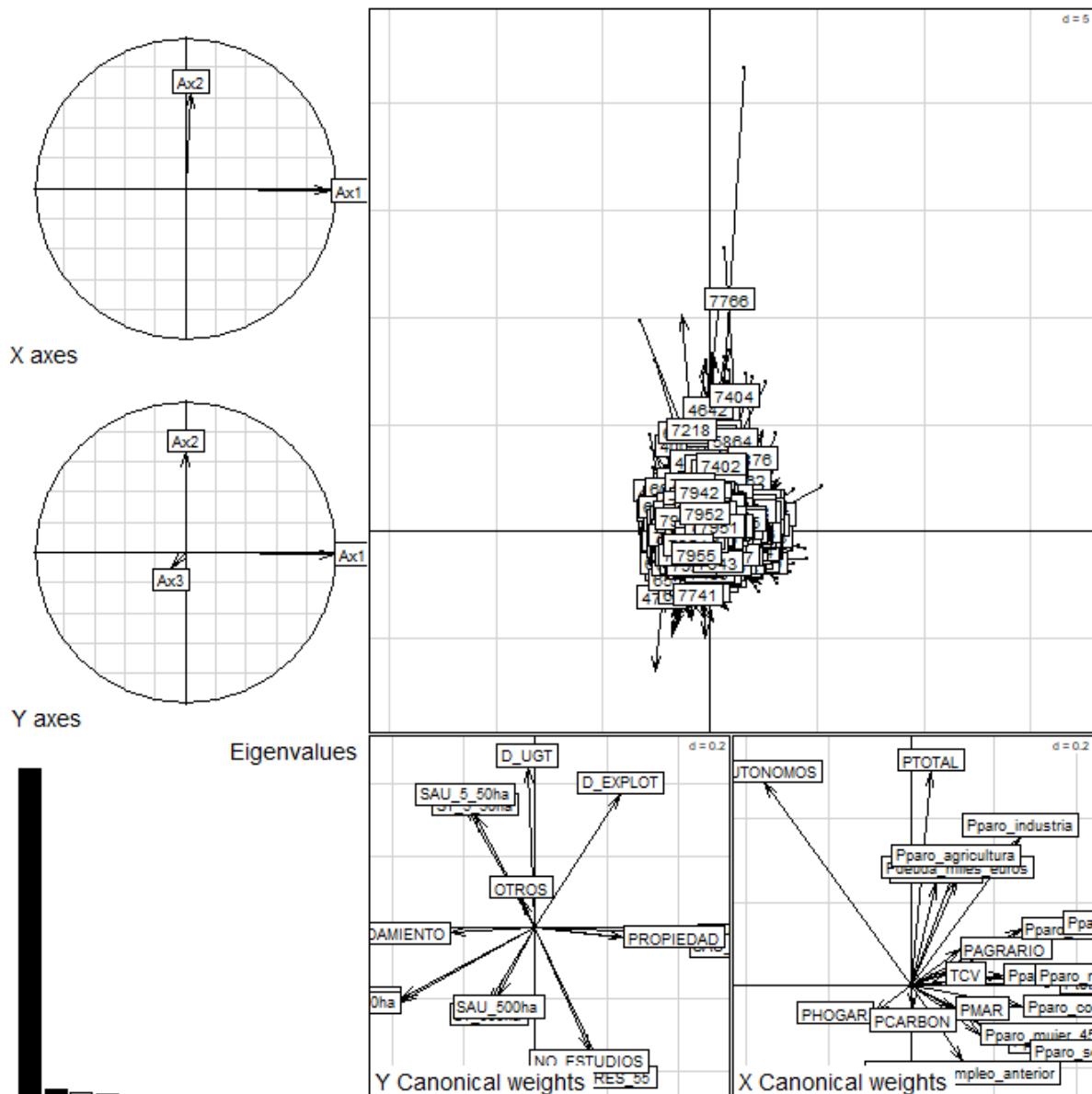


Figura 33: Resultado de la co-inercia del análisis entre las bases de datos del año 2009.

Co-inercia para la base de datos del 2020

El análisis de co-inercia realizado entre las variables socioeconómicas y el censo agrícola de 2020 mostró que la inercia total fue de 4.038, indicando la variabilidad explicada por los primeros ejes comunes. Los cinco primeros valores propios fueron 3.6202, 0.255, 0.116, 0.018 y 0.012, donde el primer eje explicó el 89.64% de la inercia total, seguido por el segundo con 6.33%, y el tercero con 2.86%, acumulando entre los cinco ejes el 99.60% de la inercia, lo que reflejó que prácticamente toda la variabilidad estaba contenida en estos ejes. En la

descomposición de los valores propios, para el primer eje, la covarianza fue de 1.902, con una desviación estándar de 2.343 para las variables socioeconómicas y 2.055 para las del censo agrícola, mostrando una correlación moderada de 0.395, mientras que en el segundo eje, la covarianza fue de 0.5054, con desviaciones estándar de 1.257 para X y 1.555 para Y, y una correlación menor de 0.258. La inercia para el conjunto de datos socio-económicos fue de 5.490, con una inercia máxima de 5.919 y un ratio de 0.927, mientras que para el censo agrícola la inercia fue de 4.224, con una inercia máxima de 4.973 y un ratio de 0.849. El coeficiente RV, que midió la relación entre las dos tablas, fue de 0.092, indicando una correlación casi nula entre sus estructuras. Estos resultados sugirieron que la mayor parte de la covariabilidad entre las variables estaba concentrada en el primer eje, aunque la correlación observada en 2020 fue menor en comparación con el análisis de 2009, lo que podría indicar cambios en la relación entre las variables analizadas (**Tabla 22**).

En el análisis del conjunto de datos X, se observa que en el Eje 1 las variables que más contribuyeron positivamente fueron ST_50_500ha y SAU_50_500ha, mientras que ST_5ha y SAU_5ha presentaron contribuciones negativas. En el Eje 2, las variables NO_ESTUDIOS y ARRENDAMIENTO contribuyeron negativamente, mientras que D_EXPLOTT y SAU_5_50ha lo hicieron de manera positiva.

Por otro lado, en el conjunto de datos Y, relacionado con el censo agrario (X), las variables AGRARIO y paro_agricultura son las que más influyeron positivamente en el Eje 1, mientras que MAR y AUTONOMOS tuvieron una influencia negativa. En el Eje 2, paro_industria y TCV fueron las variables con mayores contribuciones positivas, mientras que paro_servicios y paro_mujer_25_45 revelaron contribuciones negativas (**Figura 34**).

Tabla 22: Resultado del análisis de Co-inercia para la base de datos de 2020.

Parámetro	Eje 1	Eje 2	Eje 3	Eje 4	Eje 5
Inercia Total			4.038		
Valores Propios (Eigenvalues)	3.6202	0.2554	0.1156	0.0188	0.0123
Inercia Proyectada (%)	89.64	6.33	2.86	0.46	0.30
Inercia Acumulada (%)	89.64	95.97	98.83	99.29	99.60
Descomposición de Valores Propios					
Valor Propio	3.6202	0.2554	-	-	-
Covarianza	1.9027	0.5054	-	-	-
Desviación Estándar X	2.3430	1.2573	-	-	-
Desviación Estándar Y	2.0552	1.5553	-	-	-
Correlación	0.3951	0.2584	-	-	-
Inercia y Co-inercia X					
Inercia	5.4895	-	-	-	-
Máxima	5.9192	-	-	-	-
Ratio	0.9274	-	-	-	-
Inercia y Co-inercia Y					
Inercia	4.2239	-	-	-	-
Máxima	4.9730	-	-	-	-
Ratio	0.8494	-	-	-	-
Coefficiente RV			0.0923		

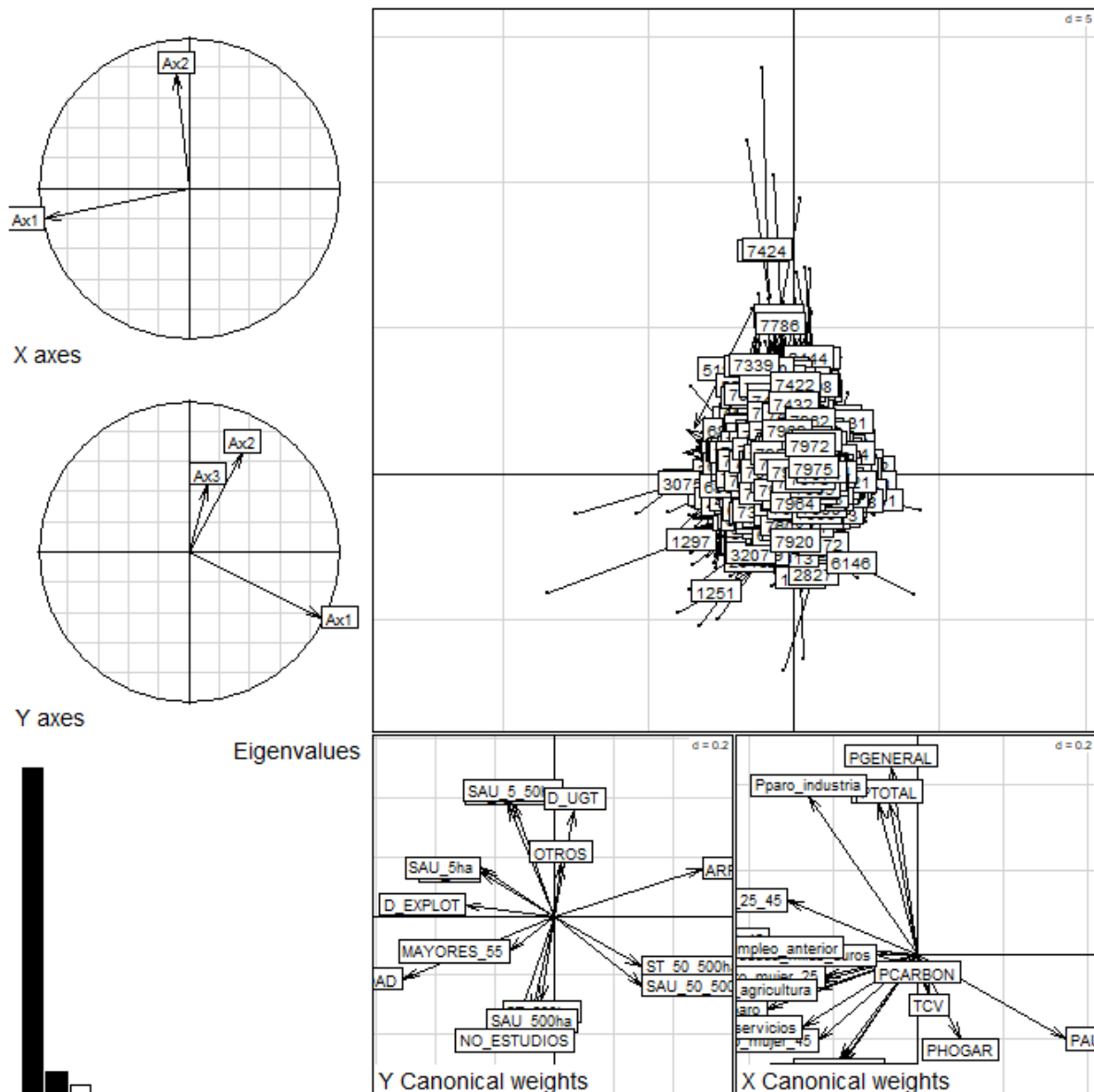


Figura 34: Resultado de la co-inercia del análisis entre las bases de datos del año 2020.

Una vez realizado el análisis STATIS-Dual y de Co-inercia, con las coordenadas obtenidas de los municipios, se procedió al análisis de cluster para determinar qué grupos conforman a las variables y sus trayectorias (cambios) en el tiempo.

3.2.3. Clustering K-means

Con ayuda de los distintos métodos estudiados, se obtuvo un número óptimo de centroides un tanto difuso. Utilizando el Método del Codo, los diagramas sugerían de 2 a 3 clústeres k,

mientras que el Método de la Silueta, parece que determina 2 como número óptimo (**Figura 35, Figura 36**).

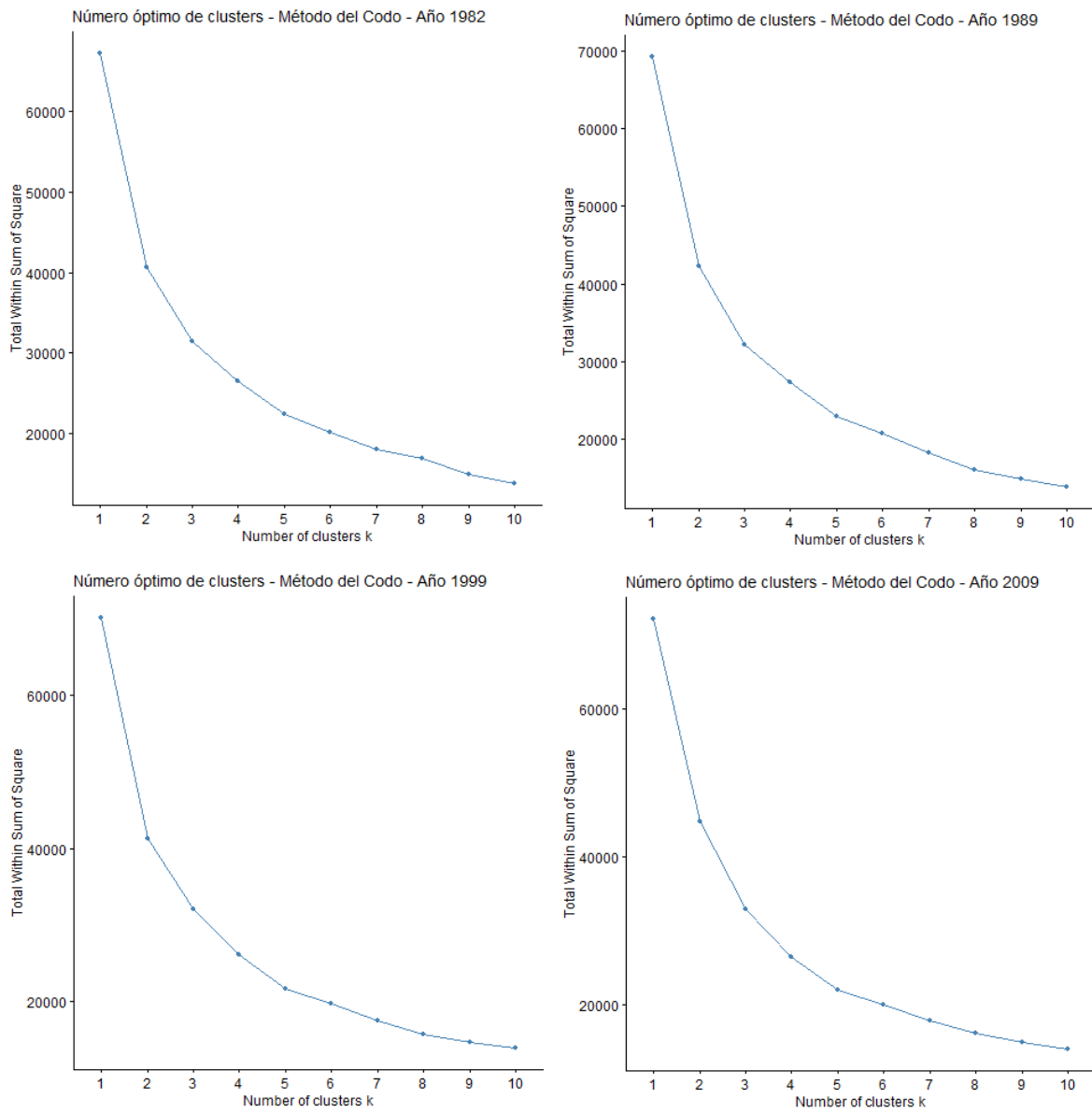


Figura 35: Gráficos de Método del Codo para los años 1982-98-99-09.

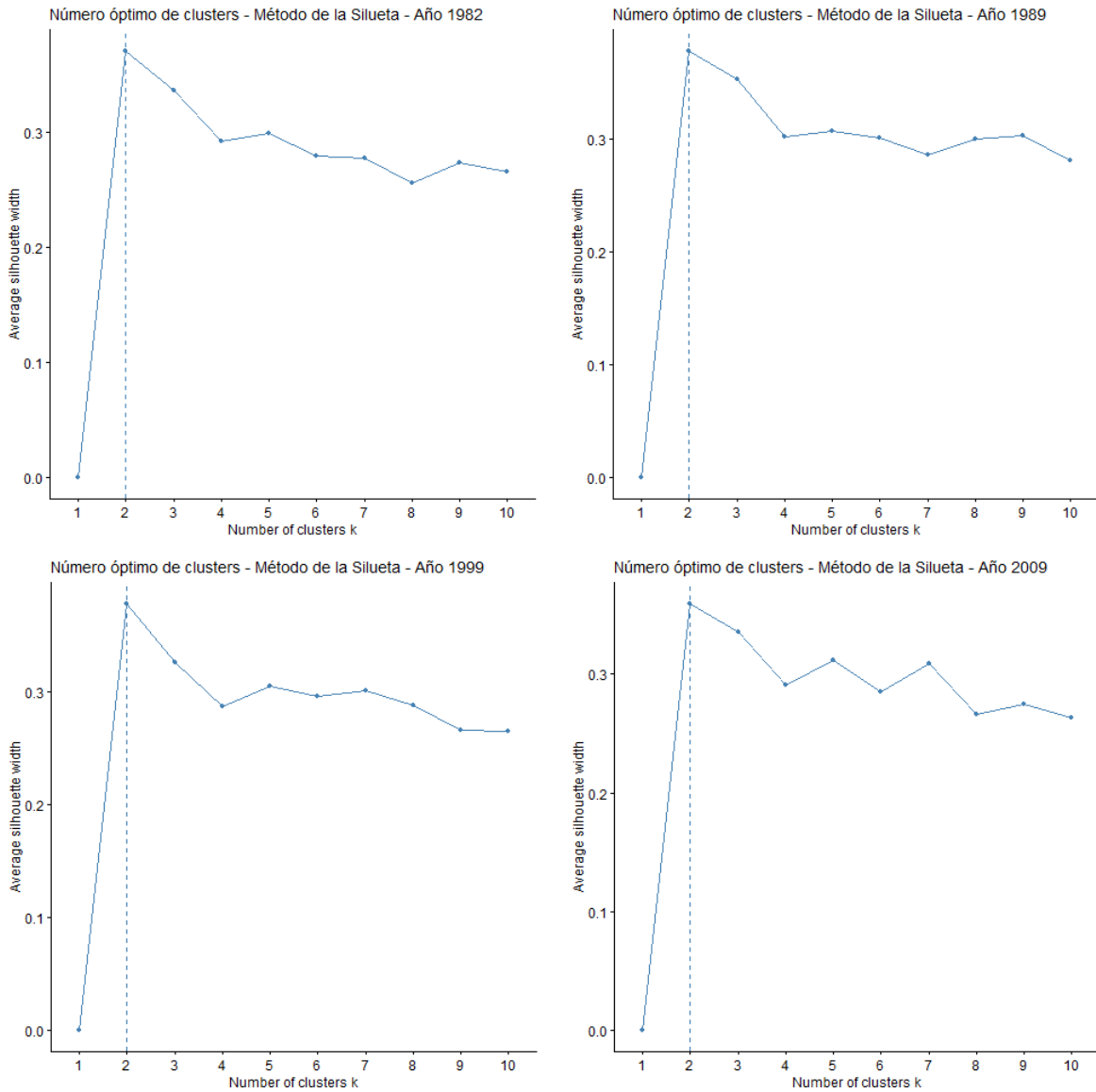


Figura 36: Gráficos de Método de la Silueta para los años 1982-98-99-09.

Se decidió realizar el K-means con 3 clústeres iniciales debido a la ausente claridad de los resultados en la anterior metodología.

Cluster K-means en STATIS-Dual : para la creación de los clusters en el STATIS-Dual, se empleó la columna *C.Co.* Este término hace referencia a un marco de datos que contiene las coordenadas de las columnas obtenidas tras proyectarlas en el espacio de las componentes principales. En este caso, se utilizaron tres componentes principales para la proyección.

Estas matrices de coordenadas son luego combinadas con los códigos municipales (COD_-

MUNI) para identificar a los correspondientes municipios. Después, se aplicó el algoritmo K-means sobre las coordenadas proyectadas, con el objetivo de agrupar los municipios en tres clusters (centers = 3).

La **Figura 37** muestra las trayectorias de los municipios en cuanto a estructura agraria en los años 1982-89-99-09. En ella, se aprecian tres agrupaciones o clusters. Cada uno, posee unas características bien definidas (**Tabla 23**):

Cluster 1

El cluster 1 (color verde) parece representar un grupo caracterizado por explotaciones de tamaño intermedio, una población relativamente envejecida, y una actividad agrícola moderada:

- **ST_5_50ha.** Las explotaciones entre 5 y 50 hectáreas representan un 59.39 %, el valor más alto en comparación con los otros clusters.
- **SAU_5_50ha.** Similarmente, el uso agrícola de explotaciones de 5 a 50 hectáreas tiene una gran relevancia, con un 53.31 %.
- **MAYORES_55.** El 52.72 % de los propietarios tienen más de 55 años, lo que sugiere una población agrícola envejecida.
- **D_UGT.** La densidad de unidades ganaderas totales es relativamente alta (31.69), lo que puede indicar una mayor actividad ganadera en comparación con otros clusters.
- **D_EXPLOT.** La densidad de explotaciones es intermedia, con un 3.80 %.

Cluster 2

El cluster 2 (color azul) representa a los municipios con alto % de explotaciones grandes y por tanto, con baja densidad total de explotaciones. El arrendamiento es el principal régimen de tenencia de estas explotaciones. Las actividades agrícolas son más intensivas en estas grandes explotaciones.

- **ST_50_500ha.** Las explotaciones entre 50 y 500 hectáreas alcanzan el 36.32 %, la proporción más alta en comparación con los otros clusters, lo que lo distingue por la presencia de explotaciones más grandes.

- **SAU_50_500ha.** Las explotaciones medianas-grandes (50-500 ha) también es elevado (31.44 %), confirmando una estructura de explotación más amplia.
- **ARRENDAMIENTO.** El 30.81 % de explotaciones es arrendado, lo que destaca la importancia del arrendamiento frente a la propiedad en este grupo.
- **D_UGT.** La densidad de unidades ganaderas totales es la más baja entre los clusters (14.86), lo que indica una actividad ganadera relativamente baja.
- **D_EXPLOT.** Este grupo tiene las densidades de explotación más bajas (1.45 %), lo que sugiere que estas explotaciones grandes podrían estar relacionadas con el tipo de agricultura o ganadería extensiva.

Cluster 3

El cluster 3 (color rojo) se caracteriza por explotaciones de muy pequeño tamaño, una explotación intensiva y una tendencia marcada hacia un sistema de tenencia en propiedad, posiblemente ligada a la necesidad de maximizar la producción en espacios limitados:

- **ST_5ha.** El 69.81 % de las explotaciones es menor a 5 hectáreas, el valor más alto entre los tres clusters. Esto sugiere predominancia de explotaciones muy pequeñas.
- **SAU_5ha.** También destaca la superficie agraria utilizada de explotaciones pequeñas, con un 81.96 %.
- **NO_ESTUDIOS.** El 94.58 % de la población no tiene estudios, lo que sugiere una menor formación educativa entre los propietarios.
- **D_EXPLOT.** Este cluster tiene el valor más alto de densidades de explotación, con un 12.10 %, lo que indica una mayor intensidad en el uso de las explotaciones; quizás debido a la fragmentación y necesidad de maximizar la producción en explotaciones pequeñas.
- **PROPIEDAD.** Las explotaciones en propiedad representan un 81.23 %, lo que sugiere una fuerte tendencia hacia la posesión directa de la explotación, frente a arrendamiento o explotación externa.

En el mapa de trayectorias, se puede observar una tendencia de expansión por parte del grupo 1, al contrario que el grupo 2, que tiende a disminuir. En el grupo 3, su crecimiento es también notable a lo largo del tiempo.

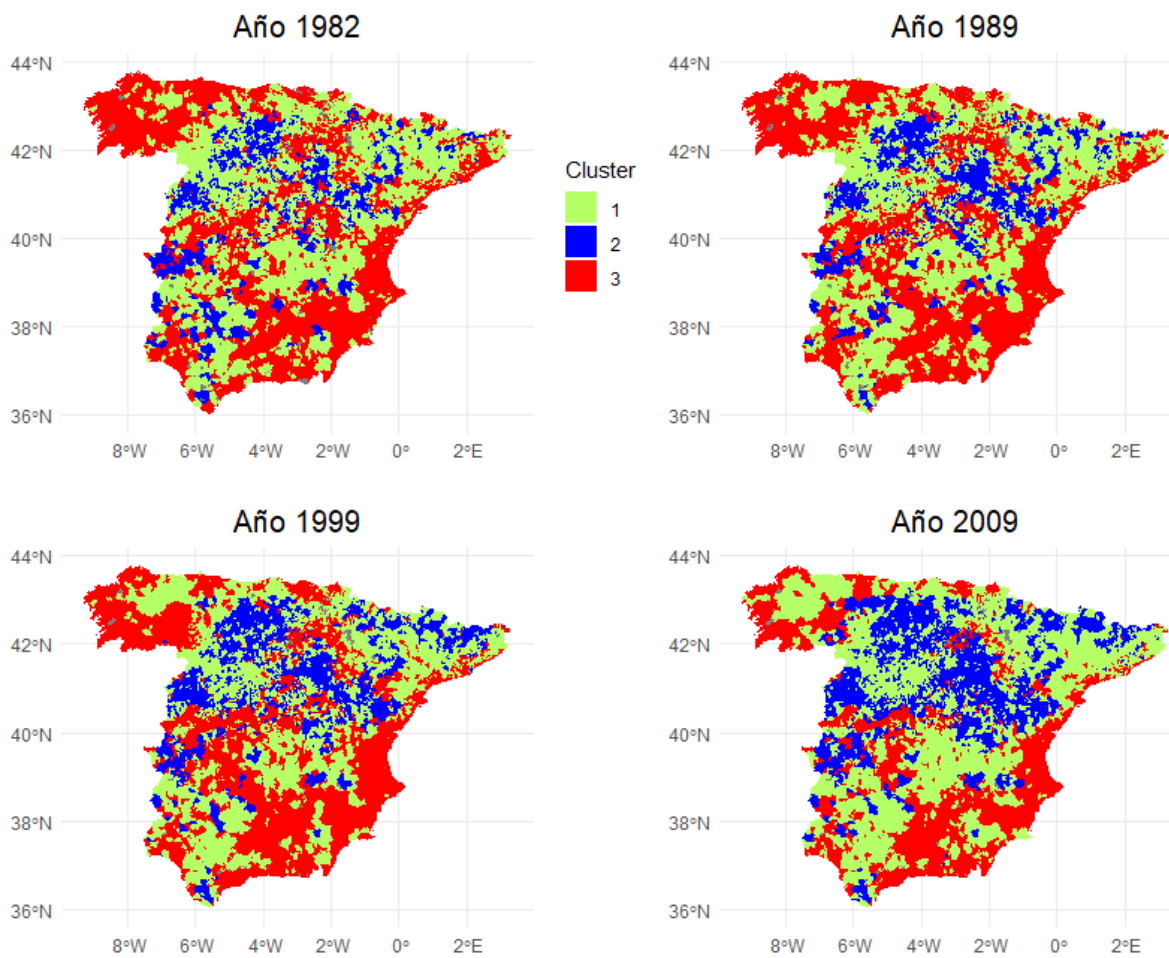


Figura 37: Mapa resultante del clustering K-means entre los 3 primeros componentes resultantes del análisis STATIS-Dual entre variables agrarias años 1982-89-99-09.

Tabla 23: Media de las variables por cluster realizado en los componentes principales del STATIS-Dual.

Variable	Cluster 1	Cluster 2	Cluster 3
ST_5ha	29.61	69.81	16.68
ST_5_50ha	59.39	26.21	43.88
ST_50_500ha	10.11	3.24	36.32
ST_500ha	0.89	0.74	3.12
SAU_5ha	39.58	81.96	23.01
SAU_5_50ha	53.31	16.35	44.45
SAU_50_500ha	6.95	1.54	31.44
SAU_500ha	0.16	0.15	1.10
MAYORES_55	52.72	57.02	45.70
NO_ESTUDIOS	91.50	94.58	91.73
D_EXPLOT	3.80	12.10	1.45
PROPIEDAD	71.57	81.23	60.42
ARRENDAMIENTO	18.47	7.74	30.81
OTROS	9.96	11.04	8.77
D_UGT	31.69	26.42	14.86

Cluster K-means en Co-inercia : en el análisis de la Co-inercia, se usaron las puntuaciones lX y lY . La primera representa las puntuaciones de las filas de la tabla dudiX (variables socioeconómicas) proyectadas en los ejes de co-inercia. Estas puntuaciones indican la posición relativa de los municipios según las variables en dudiX después de ser proyectadas en un espacio donde la inercia compartida con dudiY (variables agrarias) es maximizada. La puntuación lY es análoga a la anterior.

Del mismo modo, estas matrices de coordenadas se fusionaron con los códigos municipales (COD_MUNI) para asociar correctamente cada conjunto de puntuaciones con su municipio correspondiente y se utilizó el algoritmo de K-means basado en las coordenadas proyectadas.

La **Figura 38** muestra las trayectorias de la Co-inercia tanto en las variables del 2009 como las del 2020. La distribución de clusteres del censo agrario en el año 2009, coincide con las anteriores trayectoras del STATIS-Dual del mismo año, por lo que la explicación de dichos

clusteres quedó igual para este apartado. En las variables socioeconómicas, se ha visualizado una clara polarización entre el norte y sur de la Península.

Las clasificaciones cluster poseen unas claras características, observadas a partir del valor medio de las variables en cada uno de ellos (**Tabla 24**):

Cluster 1

El Cluster 1 (color azul celeste) se caracteriza por una mayor participación en el sector agrícola y una notable proporción de trabajadores en el sector marítimo, aunque la proporción de trabajadores autónomos es relativamente baja en comparación con otros clústeres. La deuda per cápita es la más alta en este grupo, y el paro total es elevado. Además, la tasa de crecimiento vegetativo es positiva, sugiriendo una tendencia de crecimiento poblacional.

- **PAGRARIO.** Hay una mayor afiliación al sector primario de la población en este cluster.
- **PMAR.** La proporción relativamente alta de trabajadores en el sector marítimo refleja una economía con una significativa actividad en este sector específico.
- **PAUTONOMOS.** En este cluster, la proporción de trabajadores autónomos es relativamente más baja que en los otros dos.
- **PCARBON.** La baja participación en el sector del carbono refleja una economía con una escasa dependencia de esta industria.
- **Pdeuda_miles_euros.** La deuda per cápita es la más alta en este grupo.
- **TCV.** La tasa de crecimiento vegetativo positiva indica una tendencia de crecimiento poblacional.
- **Ptotal_paro.** El paro total es alto en comparación con los otros clusters.

Cluster 2

El Cluster 2 (color amarillo) se caracteriza por una alta suma total en las afiliaciones a los distintos sectores económicos. La proporción en el sector marítimo es relativamente baja y la tasa de crecimiento vegetativo es ligeramente positiva. La deuda en los Ayuntamientos sigue teniendo importancia en este grupo.

- **PTOTAL.** La afiliación a la Seguridad Social en el Cluster 2 es alta, sugiriendo una economía diversificada y con un buen nivel de participación en distintos sectores.
- **Pdeuda_miles_euros.** La deuda per cápita relativa a los Ayuntamientos, sigue siendo alta; al igual que en el anterior cluster.
- **PMAR.** La proporción en el sector marítimo indica una economía con una débil actividad en este sector específico.
- **TCV.** La tasa de crecimiento vegetativo ligeramente positiva sugiere un leve crecimiento de la población, aunque la tendencia es hacia el mantenimiento en el tiempo.

Cluster 3

El Cluster 3 (color rosa fucsia) se diferencia por una alta participación en el sector del hogar, una significativa proporción de trabajadores autónomos y una baja deuda per cápita. Aunque la afiliación general a la Seguridad Social es baja, la proporción de paro también lo es, al igual que el crecimiento vegetativo.

- **PGENERAL.** La proporción más baja de trabajadores en general sugiere una economía con menor densidad de empleo.
- **PMAR.** La participación en el sector marítimo es menor en comparación con otros clústeres, lo que indica una menor relevancia de este sector en la economía local.
- **PHOGAR.** La mayor participación en el sector del hogar refleja una economía fuertemente enfocada en servicios domésticos.
- **PAUTONOMOS.** La mayor proporción de trabajadores autónomos indica una economía con una alta presencia de emprendedores y pequeñas empresas.
- **Pdeuda_miles_euros.** La deuda per cápita es la más baja en este grupo.
- **TCV.** La tasa de crecimiento vegetativo negativa sugiere una posible disminución de la población, pudiendo hacer referencias a zonas con fuerte despoblamiento y éxodo rural.
- **Ptotal_paro.** El bajo paro refleja una situación de empleo más favorable en comparación con otros clústeres.

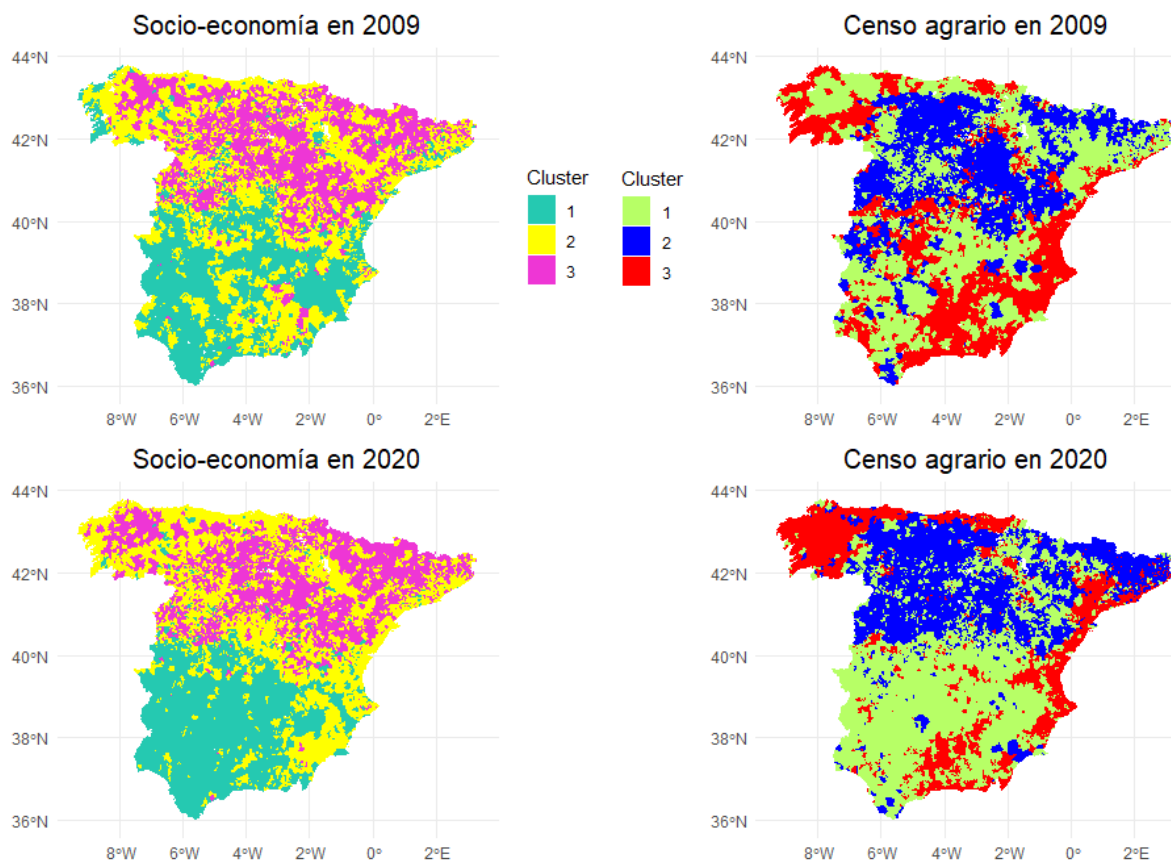


Figura 38: Mapa resultante del clustering K-means entre los 2 primeros componentes resultantes del análisis de la co-inercia entre variables agrarias y socioeconómicas en los años 2009 y 2020.

Tabla 24: Media de las variables por cluster realizado en los componentes principales de la Co-inercia.

Variable	Cluster 1	Cluster 2	Cluster 3
PGENERAL	16.9091	16.8242	10.5306
PAGRARIO	4.7622	3.4980	2.3765
PMAR	0.1064	0.0437	0.0031
PHOGAR	0.3442	0.5629	1.0425
PAUTONOMOS	6.7866	8.9614	13.0276
PCARBON	0.0046	0.0377	0.0084
PTOTAL	28.6140	29.1055	23.4644
Pdeuda_miles_euros	31.4916	28.7979	16.5172
TCV	1.1260	0.0908	-0.0003
Ptotal_paro	111.0131	65.2976	27.7828
Pparo_hombre_25	8.8105	4.8971	1.9671
Pparo_hombre_25_45	27.6013	16.1377	6.5518
Pparo_hombre_45	18.2958	12.7457	6.3032
Pparo_mujer_25	6.5023	3.2672	1.0653
Pparo_mujer_25_45	30.1577	17.0040	6.9307
Pparo_mujer_45	19.6454	11.2459	4.9648
Pparo_agricultura	5.2421	3.4139	1.6696
Pparo_industria	16.5190	10.6361	3.9411
Pparo_construccion	25.6826	13.0475	4.9782
Pparo_servicios	55.2719	34.1665	15.4609
Pparo_sin_empleo_anterior	8.2975	4.0336	1.7330

3.2.4. Regresión Parcial por Mínimos Cuadrados (PLSR)

El modelo PLSR se construyó utilizando el método kernelppls con 4289 filas, 1 variable de respuesta y 39 variables predictoras. En la **Figura 39** se observa la variable de respuesta, los incendios totales (TOTAL_INCENDIOS). Ésta, se calculó a partir de los incendios frustrados con los incendios que tomaron lugar. Los datos de las variables predictoras fueron las variables agrarias y socioeconómicas correspondientes al año 2009 (**Figura 40**).

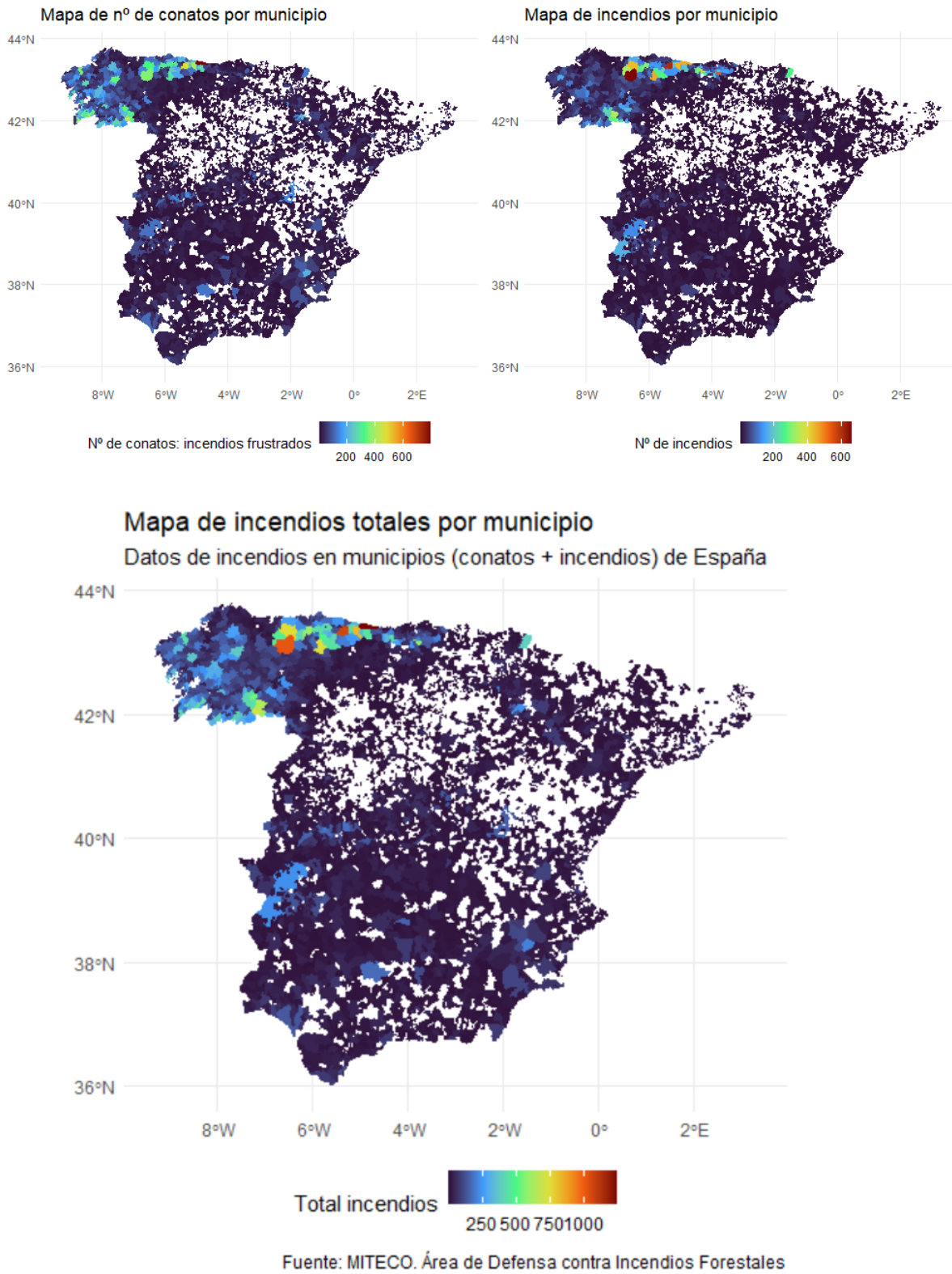
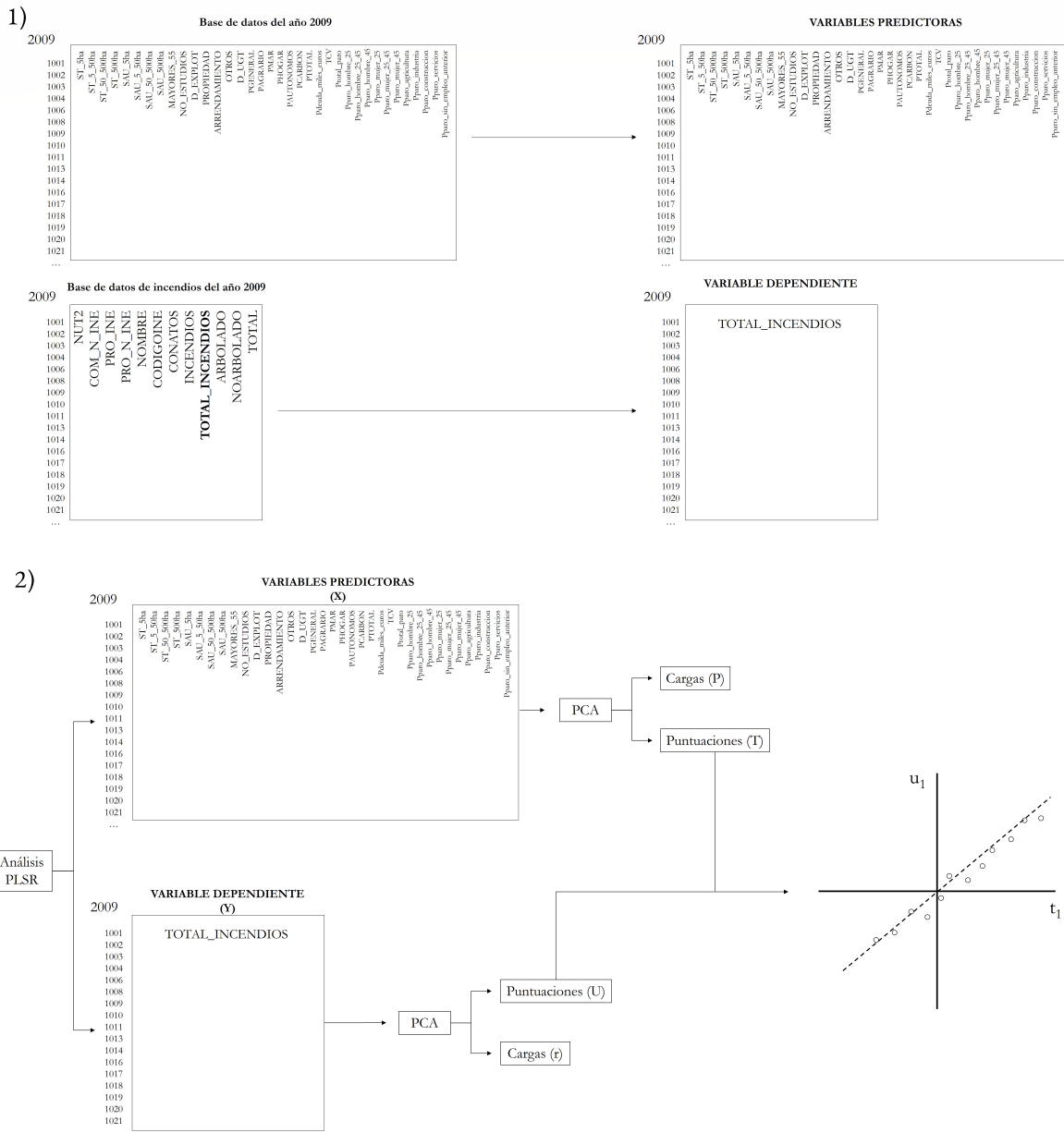


Figura 39: Mapa con los incendios totales a nivel municipal del año 2009.



Los dos primeros componentes explicaron el 19,01% y el 27,00% de la varianza en las variables predictoras, respectivamente, mientras que en la variable de respuesta, estos componentes explicaron el 6,29% y el 12,50% de la varianza. Los valores de RMSEP disminuyen con los componentes, pero la tasa de mejora comienza a estabilizarse después de dos componentes,

con un valor final de 60,60, lo que sugiere que agregar más componentes no mejora significativamente el modelo. La varianza explicada por los componentes muestra que los primeros componentes capturan la mayor parte de la varianza: Componente 1: 19,01 %, Componente 2: 7,99 %, Componente 3: 7,05 %, Componente 4: 4,87 %, y los componentes adicionales aportan solo marginalmente.

Las cargas indican cómo cada variable contribuye a cada componente; por ejemplo, SAU_5ha tiene una fuerte contribución al Componente 1 (0,366), mientras que SAU_50_500ha contribuye positivamente al Componente 2 (0,257) y UGT juega un papel importante en los Componentes 7 (0,283) y 8 (-0,642). El gráfico MSEF confirma que dos componentes son óptimos para equilibrar la complejidad y el error de predicción, y el gráfico Coefplot visualiza los coeficientes de regresión de las variables, indicando su importancia relativa (**Figura 41**, **Figura 42**).

En estadístico PLSR, las cargas indican cómo las variables X contribuyen a los componentes latentes, mientras que los coeficientes muestran cómo cada variable X afecta directamente la predicción de Y. Como el principal objetivo es la identificación de las variables que poseen mayor poder predictivo en el modelo, se usó el valor de los coeficientes.

Si las variables predictoras tienen un coeficiente positivo, esto significa que un aumento en esa variable tenderá a aumentar la predicción del total de incendios (*TOTAL_INCENDIOS*). Por lo que se refiere a las variables con coeficiente negativo, un aumento en éstas tenderá a reducir la predicción de *TOTAL_INCENDIOS*.

Las diez variables con los coeficientes absolutos mayores y, por lo tanto, las que mejor predicen el régimen de incendios según este modelo PLSR son:

- **ST_5ha:** $-1,944571 \times 10^7$
- **ST_5_50ha:** $-1,366764 \times 10^7$
- **ST_50_500ha:** $-1,530370 \times 10^7$
- **SAU_5ha:** $2,470076 \times 10^6$
- **SAU_5_50ha:** $1,657404 \times 10^6$

Aquellas que peor predicen el modelo, con coeficientes más cercanos a cero, son:

- **Pparo_hombre_25:** $4,247061 \times 10^0$
- **PAUTONOMOS:** $-3,999231 \times 10^{-1}$
- **PCARBON:** $2,701430 \times 10^{-1}$
- **PHOGAR:** $-2,776450 \times 10^{-1}$

En la (**Figura 43**) se puede apreciar un gráfico de barras resumiendo la predicción de incendios con el distinto valor de coeficientes en las variables.

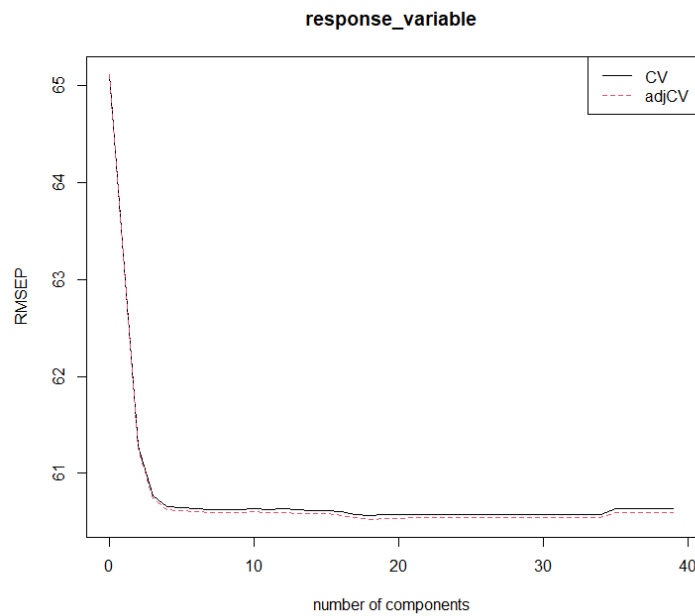


Figura 41: Gráfico MSEP de los componentes óptimos de la PLSR en el año 2009.

4. Conclusiones

1. Los resultados obtenidos a través del análisis STATIS-Dual, complementados con el análisis de la co-inercia y el modelo PLSR, proporcionan una visión comprensiva sobre la evolución de la estructura de covariación entre las variables agrarias y socioeconómicas en los años estudiados.

1.1. En primer lugar, el STATIS-Dual reveló una alta estabilidad en las estructuras de las variables agrarias durante las décadas comprendidas entre 1982 y 2009, con coeficientes de correlación vectorial (RV) que indicaron una notable similitud entre los periodos consecutivos.

1.2. Sin embargo, se detectó un cambio significativo en la estructura de 2020, año de la pandemia Covid-19, alejándose de la estructura compromiso, haciendo que este año resulte atípico, lo que motivó su exclusión del análisis.

2. El análisis de Co-inercia entre los datos agrarios y socio-económicos de 2009 y 2020 evidenció una covariación moderada:

2.1. La co-inercia se concentró mayoritariamente entre los dos primeros ejes, lo que sugiere que la mayor parte de la variabilidad entre estos dos conjuntos de datos fue capturada por una única dimensión.

2.2. Si bien algunas variables agrarias y socioeconómicas compartían un patrón de covariabilidad, la asociación general entre estos conjuntos de datos fue limitada, especialmente en el análisis de 2020.

3. Por otro lado, el modelo PLSR:

3.1. Destacó las variables agrarias relacionadas con superficies pequeñas y medianas (ST 5ha, SAU 5ha, ST 50 500ha) como las que más contribuyen a la predicción del régimen de incendios, mientras que las variables socioeconómicas relativas al paro mostraron una menor influencia.

3.2. La capacidad predictiva del modelo fue moderada, con los primeros dos componentes capturando la mayor parte de la variabilidad en las variables predictoras, pero con un margen de mejora limitado al agregar más componentes.

4. En conjunto, estos hallazgos sugieren que:

4.1. Aunque la estructura agraria y socioeconómica ha mantenido cierta estabilidad a lo largo de los cinco años estudiados, la evolución hacia 2020 refleja cambios importantes en la dinámica rural y agraria, probablemente impulsados por el cambio tan abrupto que supuso la pandemia.

4.2. Asimismo, el comportamiento de las variables predictoras en el modelo PLSR subraya la relevancia de las características agrarias en la comprensión de los fenómenos relacionados con los incendios, destacando la influencia de las pequeñas y medianas superficies agrícolas sobre el riesgo de incendios en el contexto socioeconómico rural.

5. Hemos demostrado cómo el análisis de las relaciones entre las variables socioeconómicas y agrarias a diferentes escalas geográficas y temporales proporciona información crucial para entender los factores subyacentes al régimen de incendios y, por ende, para mejorar la gestión y prevención del riesgo.

6. La pirogeografía emerge como un marco conceptual ideal para abordar el análisis comparativo del régimen de incendios a nivel regional. Esta disciplina permite identificar y caracterizar las denominadas “piorregiones”, áreas geográficas homogéneas en cuanto a sus dinámicas de incendios y que requieren de respuestas de gestión específicas, tanto en el tiempo como en el espacio. Dicho enfoque facilita la creación de estrategias de manejo del fuego más adaptadas a las características locales y regionales, optimizando los esfuerzos de prevención y control.

7. En este trabajo, se ha resaltado la importancia del estudio de las variables agrarias frente a las socioeconómicas, ya que éstas han proporcionado una mayor predicción sobre el régimen de incendios. En ellas, con ayuda del análisis multivariante STATIS-Dual, se han identificado tres regiones funcionales (clusters) en la península Ibérica con características claramente diferenciadas, manifestando la importancia de estudiar las variables agrarias a nivel municipal.

Bibliografía

- Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences*, 113(42), 11770-11775.
- Abdi, H., Williams, L. J., Valentin, D., & Bannani-Dosse, M. (2012). STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 124-167.
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295.
- Alcasena, F. J., Ager, A. A., Bailey, J. D., Pineda, N., & Vega-García, C. (2019). Towards a comprehensive wildfire management strategy for Mediterranean areas: Framework development and implementation in Catalonia, Spain. *Journal of environmental management*, 231, 303-320.
- Alcasena, F. J., Salis, M., Ager, A. A., Castell, R., & Vega-García, C. (2017). Assessing wildland fire risk transmission to communities in northern Spain. *Forests*, 8(2), 30.
- Alcasena, F., Ager, A., Le Page, Y., Bessa, P., Loureiro, C., & Oliveira, T. (2021). Assessing wildfire exposure to communities and protected areas in Portugal. *Fire*, 4(4), 82.
- Amatulli, G., Rodrigues, M. J., Trombetti, M., & Lovreglio, R. (2006). Assessing long-term fire risk at local scale by means of decision tree technique. *Journal of Geophysical Research: Biogeosciences*, 111(G4).
- Anzanello, M. J., & Fogliatto, F. S. (2014). A review of recent variable selection methods in industrial and chemometrics applications. *European Journal of Industrial Engineering*, 8(5), 619-645.
- Archibald, S., Lehmann, C. E., Gómez-Dans, J. L., & Bradstock, R. A. (2013). Defining pyromes and global syndromes of fire regimes. *Proceedings of the National Academy of Sciences*, 110(16), 6442-6447.
- Archibald, S., Nickless, A., Govender, N., Scholes, R. J., & Lehsten, V. (2010). Climate and the inter-annual variability of fire in southern Africa: a meta-analysis using long-term field data and satellite-derived burnt area data. *Global Ecology and Biogeography*, 19(6), 794-809.
- Bar-Massada, A., Alcasena, F., Schug, F., & Radeloff, V. C. (2023). The wildland–urban interface in Europe: Spatial patterns and associations with socioeconomic and demographic variables. *Landscape and Urban Planning*, 235, 104759.
- Blasi, C., Capotorti, G., Copiz, R., Guida, D., Mollo, B., Smiraglia, D., & Zavattoni, L. (2014). Classification and mapping of the ecoregions of Italy. *Plant Biosystems—An International Journal Dealing with all Aspects of Plant Biology*, 148(6), 1255-1345.
- Bowman, D. M., O'Brien, J. A., & Goldammer, J. G. (2013). Pyrogeography and the global quest for sustainable fire management. *Annual Review of Environment and Resources*, 38, 57-80.
- Bradstock, R. A. (2010). A biogeographic model of fire regimes in Australia: current and future implications. *Global Ecology and Biogeography*, 19(2), 145-158.
- Bustamam, A., Tasman, H., Yuniarti, N., Frisca, F., & Mursidah, I. (2017, July). Application of k-means clustering algorithm in grouping the DNA sequences of hepatitis B virus (HBV). In *AIP conference proceedings* (Vol. 1862, No. 1). AIP Publishing.
- Carlucci, M., Zambon, I., Colantoni, A., & Salvati, L. (2019). Socioeconomic development, demographic dynamics and forest fires in Italy, 1961–2017: a time-series analysis. *Sustainability*, 11(5), 1305.

- Castro, M. D., Martín-Vide, J., & Alonso, S. (2005). El clima de España: pasado, presente y escenarios de clima para el siglo XXI.
- Chaguay, S. L., Galeas, R. R., & Chaguay, L. A. L. (2020). Desempleo en tiempos de covid-19: efectos socioeconómicos en el entorno familiar. *Journal of Science and Research: Revista Ciencia e Investigación*, 5(4), 187-197.
- Chen, D., & Chen, H. W. (2013). Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development*, 6, 69-79.
- Chuvieco, E., Giglio, L., & Justice, C. (2008). Global characterization of fire activity: toward defining fire regimes from Earth observation data. *Global change biology*, 14(7), 1488-1502.
- Chuvieco, E., Yebra, M., Martino, S., Thonicke, K., Gómez-Giménez, M., San-Miguel, J., ... & Viegas, D. (2023). Towards an integrated approach to wildfire risk assessment: when, where, what and how may the landscapes burn. *Fire*, 6(5), 215.
- Clarke, H., Gibson, R., Cirulis, B., Bradstock, R. A., & Penman, T. D. (2019). Developing and testing models of the drivers of anthropogenic and lightning-caused wildfire ignitions in south-eastern Australia. *Journal of environmental management*, 235, 34-41.
- Curt, T., Fréjaville, T., & Bouillon, C. (2014). Characterizing pyroregions in south-eastern France.
- de Diego, J., Fernández, M., Rúa, A., & Kline, J. D. (2022). Spatializing and temporalizing socioeconomic determinants of wildfires in Galicia (Spain).
- Dolédec, S., & Chessel, D. (1994). Co-inertia analysis: an alternative method for studying species–environment relationships. *Freshwater biology*, 31(3), 277-294.
- Dray, S., Chessel, D., & Thioulouse, J. (2003). Co-inertia analysis and the linking of ecological data tables. *Ecology*, 84(11), 3078-3089.
- Dufour, A. B. (2009). Coinertia Analysis. *Logiciel R version*, 2(1), 2-11.
- Dupriez, O., & Boyko, E. (2010). Dissemination of Microdata Files: Principles Procedures and Practices. International Household Survey Network.
- ESRI (1998). ESRI Shapefile Technical Description. Environmental Systems Research Institute. Disponible online en: <https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/whitepapers/pdfs/shapefile.pdf>. Última visualización en 22 de mayo de 2024.
- ESRI (2021). ¿Qué es un shapefile? Archivo de ayuda. Disponible online en: [https://desktop.arcgis.com/es/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm#:~:text=Un%20shapefile%20es%20un%20formato.%C3%ADneas%20o%20p%0l%C3%ADgonos%20\(%C3%A1reas\)](https://desktop.arcgis.com/es/arcmap/latest/manage-data/shapefiles/what-is-a-shapefile.htm#:~:text=Un%20shapefile%20es%20un%20formato.%C3%ADneas%20o%20p%0l%C3%ADgonos%20(%C3%A1reas)). Última visualización el 9 de junio de 2024.
- FAO (2007). Fire management - global assessment 2006. A thematic study prepared in the framework of the Global Forest Resources Assessment 2005. Estudio FAO: Montes n.º 151. Roma, 2007
- Firinguetti, L., Kibria, G., & Araya, R. (2017). Study of partial least squares and ridge regression methods. *Communications in Statistics-Simulation and Computation*, 46(8), 6631-6644.
- Galizia, L. F., Barbero, R., Rodrigues, M., & Curt, T. (2022). Pyrogeography: An Alternative Zonation for Europe. *Environmental Sciences Proceedings*, 17(1), 80.
- Gallant, A. L., Loveland, T. R., Sohl, T. L., & Napton, D. E. (2004). Using an ecoregion framework to analyze land-cover and land-use dynamics. *Environmental Management*, 34, S89-S110.

- Gutiérrez-Moreno, A. (2020). Educación en tiempos de crisis sanitaria: pandemia y educación. *Praxis*, 16(1), 7-10.
- Haşiloğlu, A., & Bali, A. (2017). Statistical Methods of Confidentiality for Micro Data and Developing an Object Oriented Statistical Disclosure Control Software. *El-Cezeri*, 4(3), 319-333.
- Jiménez-Ruano, A. J. (2015). Caracterización de patrones espacio-temporales de los regímenes de incendio en la España Peninsular: pireoregiones recientes y futuras. In *Análisis espacial y representación geográfica: innovación y aplicación* (pp. 2187-2190). Departamento de Geografía y Ordenación del Territorio.
- Jiménez-Ruano, A. J. (2019). Characterizing pyroregions in mainland Spain from spatial-temporal patterns of fire regime and their underlying drivers (Doctoral dissertation, Universidad de Zaragoza).
- Krawchuk, M. A., Moritz, M. A., Parisien, M. A., Van Dorn, J., & Hayhoe, K. (2009). Global pyrogeography: the current and future distribution of wildfire. *PloS one*, 4(4), e5102.
- Li, X., Nsofor, G. C., & Song, L. (2009). A comparative analysis of predictive data mining techniques. *International Journal of Rapid Manufacturing*, 1(2), 150-172.
- MAP; Ministerio de Administraciones Públicas (2008). Secretaría General Técnica. Variaciones de los Municipios de España desde 1842 (1st ed.). Recuperado de https://mpt.gob.es/dam/es/portal/politica-territorial/local/sistema_de_informacion_local_SIL-/registro_eell/publicaciones/parrafo/02/01-Variaciones-20desde-201842.pdf. Consultado en 12 de marzo de 2024.
- Marín, S. y Mínguez, R. (2013). Evolución sectorial de la economía española en los últimos 47 años. Madrid. Estudio del Consejo General de Economistas y la Cámara de España. Disponible online en: <https://www.camara.es/evolucion-sectorial-economia-espanola--ultimos-47-anos>. Última visualización el 12 de agosto de 2024.
- Martín Escorza, C. (1992). La estructura geológica de la Península Ibérica y sus aguas termales.
- Marutho, D., Handaka, S. H., & Wijaya, E. (2018, September). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In 2018 international seminar on application for technology of information and communication (pp. 533-538). IEEE.
- MITECO; Ministerio de Transición Ecológica y Reto Demográfico (2012). El Inventario Español del Patrimonio Natural y de la Biodiversidad. Madrid. Disponible en: https://www.miteco.gob.es/es/biodiversidad/temas/inventarios-nacionales/primer_a_parte_iepnb_enero_2012_v2_tcm30-294527.pdf. Consultado en 8 de junio de 2024.
- MITECO; Ministerio para la Transición Ecológica y el Reto Demográfico (2021). Frecuencia de incendios por término municipal ocurridos entre los años 2006 a 2015 [Base de datos]. Dirección General de Biodiversidad, Bosques y Desertificación. Subdirección General de Política Forestal y Lucha contra la Desertificación. Área de Defensa contra Incendios Forestales. <https://www.opengis.net/def/crs/EPSS/0/4258>
- Moreno, M. V., Conedera, M., Chuvieco, E., & Pezzatti, G. B. (2014). Fire regime changes and major driving forces in Spain from 1968 to 2010. *Environmental Science & Policy*, 37, 11-22.
- Morgan, P., Hardy, C. C., Swetnam, T. W., Rollins, M. G., & Long, D. G. (2001). Mapping fire regimes across time and space: Understanding coarse and fine-scale fire patterns. *International Journal of Wildland Fire*, 10(4), 329-342.

- Pechony, O., & Shindell, D. T. (2010). Driving forces of global wildfires over the past millennium and the forthcoming century. *Proceedings of the National Academy of Sciences*, 107(45), 19167-19170.
- Pelliccia, D. (2018). Outliers detection with PLS regression for NIR spectroscopy in Python.
- Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 219(1), 103-119.
- Rodrigues, M., Costafreda-Aumedes, S., Comas, C., & Vega-García, C. (2019). Spatial stratification of wildfire drivers towards enhanced definition of large-fire regime zoning and fire seasons. *Science of the total environment*, 689, 634-644.
- Rodrigues, M., Jiménez-Ruano, A., & de la Riva, J. (2020). Fire regime dynamics in mainland Spain. Part 1: Drivers of change. *Science of The Total Environment*, 721, 135841.
- Rodrigues, M., Jiménez-Ruano, A., Peña-Angulo, D., & De la Riva, J. (2018). A comprehensive spatial-temporal analysis of driving factors of human-caused wildfires in Spain using Geographically Weighted Logistic Regression. *Journal of environmental management*, 225, 177-192.
- Rodrigues, M., Zúñiga-Antón, M., Alcasena, F., Gelabert, P., & Vega-Garcia, C. (2022). Integrating geospatial wildfire models to delineate landscape management zones and inform decision-making in Mediterranean areas. *Safety science*, 147, 105616.
- Rodríguez Martínez, C. C. (2020). Contribuciones a los métodos STASIS basados en técnicas de aprendizaje no supervisado.
- Roos, C. I., Bowman, D. M., Balch, J. K., Artaxo, P., Bond, W. J., Cochrane, M., ... & Swetnam, T. W. (2014). Pyrogeography, historical ecology, and the human dimensions of fire regimes. *Journal of Biogeography*, 41(4), 833-836.
- Ruano, A. J. (2019). Characterizing pyroregions in mainland Spain from spatial-temporal patterns of fire regime and their underlying drivers (Doctoral dissertation, Universidad de Zaragoza).
- Saputra, D. M., Saputra, D., & Oswari, L. D. (2020, May). Effect of distance metrics in determining k-value in k-means clustering using elbow and silhouette method. In *Sriwijaya international conference on information technology and its applications (SICONIAN 2019)* (pp. 341-346). Atlantis Press.
- Sim, K., & Hartley, R. (2006, June). Removing outliers using the L_{∞} norm. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. 1, pp. 485-494). IEEE.
- Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering* (Vol. 336, p. 012017). IOP Publishing.
- Syphard, A. D., & Keeley, J. E. (2020). Mapping fire regime ecoregions in California. *International Journal of Wildland Fire*, 29(7), 595-601.
- UO; Universidad de Oviedo (2016). El algoritmo k-means aplicado a clasificación y procesamiento de imágenes. Disponible online en: https://www.unioviado.es/compnum/laboratorios_py/kmeans/kmeans.html. Última visualización en 22 de junio de 2024.

- Viedma, O., Urbieto, I. R., & Moreno, J. M. (2018). Wildfires and the role of their drivers are changing over time in a large rural area of west-central Spain. *Scientific reports*, 8(1), 1-13.
- Wold, S., Ruhe, A., Wold, H., & Dunn, W. J. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3), 735-743.
- Wunder, S., Calkin, D. E., Charlton, V., Feder, S., Arano, I., & Moore, P. (2021). Rodríguez; Silva, F.; Tacconi, L.; Vega-García, C. Resilient landscapes to prevent catastrophic forest fires: Socioeconomic insights towards a new paradigm. *For. Policy Econ*, 128, 102458.

Anexo

En el anexo, se encuentran todos los scripts realizados en RStudio para el *datamining* de las variables, análisis estadísticos y visualización de mapas.

Scripts de RStudio

Contenido de los scripts de RStudio:

Cálculo de las variables para el año 1989

```
1      #install.packages("rstudioapi")
2      setwd(dirname(rstudioapi::getSourceEditorContext())$path))
3      library(sqldf)
4      library(dplyr)
5
6      numeros <- c(1:6, 8:34, 36:37, 39:50)
7      archivos <- paste0("P", sprintf("%02d", numeros), "A89.VDO")
8
9      # Bucle para procesar cada archivo
10     for (i in 1:length(archivos)) {
11
12         archivo_actual <- archivos[i]
13
14         #----NIVEL 1 (cts, nexplot, csau, formacion y ugt)
15         provincia <- scan(archivo_actual, what = character(), sep = "/")
16
17         provincia_temp <- provincia[grepl("^1", provincia)]
18         archivo_temporal <- paste0("P", sprintf("%02d", numeros[i]),
19             "A89_1")
20         write.table(provincia_temp, file=archivo_temporal, quote=FALSE,
21             row.names=FALSE, col.names=FALSE)
22         longitud <- c(1,2,3,6,2,2,1,1,1,1,1,1,7,2,7,2,1,1,1,1,6,6,6,6,6,
23             ,4,4,4,5,8,4,9)
24         variable <- c("NIVEL 1", "PROV", "MUNI", "NUME CUES", "RE",
25             "COMARCA", "ZONA", "PJ", "GES", "CONTABI",
26             "FORMACI", "CLAS EXPLO", "ST", "CST", "SAU", "CSAU", "STSRT",
27             "SAUSRT", "STSRT2", "SAUSRT2",
28             "UTAT", "UTAA", "UTAAF", "UTAAE", "UTAF", "UTATI", "UTAJE",
29             "UTAC", "UTAOF", "UGT", "OTE", "MBT")
30         clases<- c("character","character","character","character","cha
31             racter","character","character","character","character","charac
32             ter","character",
33             "character","character","character","character","character","ch
34             aracter","character","character","character","character","chara
35             cter",
36             "character","character", "character","character","character","c
37             haracter","character","numeric","character","character")
```

```

27     provin_nivel1_temp <- read.fwf(archivo_temporal, widths =
28     longitud, col.names = variable, colClasses = clases)
29     archivo_resultado <- paste0("provincia_", sprintf("%02d",
30     numeros[i]), "_n1.txt")
31     write.csv(provin_nivel1_temp, file = archivo_resultado, quote =
32     FALSE, row.names = FALSE)
33     #-----NEXPLOT-----
34     np <- sqldf("select PROV, MUNI, COUNT(CST) as total, CST from
35     provin_nivel1_temp group by MUNI, CST")
36     np2 <- transform(np, CODMUNICST = paste0(PROV, MUNI, CST))
37
38     expltotal<-sqldf("select PROV,MUNI,sum(total) as nexplot from
39     np2 group by MUNI")
40     expltotal<-transform(expltotal,CODMUNI=paste0(PROV,MUNI))
41     expltotal<-sqldf("select CODMUNI, nexplot from expltotal")
42
43     nexplot_resul <- paste0("NEXPLOT_", sprintf("%02d",
44     numeros[i]), ".txt")
45     write.csv(expltotal, file = nexplot_resul, quote = FALSE,
46     row.names = FALSE)
47     #-----CST-----
48     cod_mun <- unique(expltotal$CODMUNI)
49     cod_mun <- c(expltotal$CODMUNI)
50     tablamuni <- rep(c(cod_mun), each = 21)
51     grupos<-rep(c("01", "02", "03", "04", "05", "06", "07", "08", "09", "10",
52     "11", "12", "13", "14", "15", "16", "17", "18", "19", "20", "21"),length
53     (cod_mun))
54     tabla_89<-data.frame(tablamuni,grupos)
55
56     tabla_89_st <- tabla_89
57     names(tabla_89_st) <- c("CODMUNI", "CST")
58     tabla_89_st<-transform(tabla_89_st,CODMUNCST=paste0(CODMUNI,CST
59     ))
60     tablacst_89<-sqldf("select
61     tabla_89_st.CODMUNI,tabla_89_st.CST,np2.total from tabla_89_st
62     left join np2 on tabla_89_st.CODMUNCST=np2.CODMUNICST")
63     tablacst_89[is.na(tablacst_89)] <- 0
64     proba_cst<-reshape(tablacst_89, timevar="CST",
65     idvar=c("CODMUNI"), direction="wide")
66     #5/ 5 a 50/ 50 a 500/ 500 --> 01-08/ 9-12/ 13-18/ 19-21
67     proba_cst <- proba_cst %>%
68     mutate(
69     CST_5 = rowSums(select(., starts_with("total.0"))[,1:8], na.rm
70     = TRUE),
71     CST_5_50 = rowSums(select(.,
72     matches("total.0[9]$|total.1[0-2]$")), na.rm = TRUE),
73     CST_50_500 = rowSums(select(., starts_with("total."))[, -c(1:12,
74     19:21)]), na.rm = TRUE),
75     CST_500 = rowSums(select(., starts_with("total."))[, -c(1:18)],
76     na.rm = TRUE)
77     ) %>%
78     select(CODMUNI, CST_5, CST_5_50, CST_50_500, CST_500)
79     cst_resul <- paste0("CST_", sprintf("%02d", numeros[i]), ".txt")

```

```

63 write.csv(proba_cst, file = cst_resul, quote = FALSE, row.names
    = FALSE)
64 #-----CSAU-----
65 csau <- sqldf("select PROV, MUNI, COUNT(CSAU) as total, CSAU
    from provin_nivel1_temp group by MUNI, CSAU")
66 csau <- transform(csau, CODMUNCSAU = paste0(PROV,MUNI,CSAU))
67
68 tabla_89_sau <- tabla_89
69 names(tabla_89_sau) <- c("CODMUNI", "CSAU")
70 tabla_89_sau<-transform(tabla_89_sau,CODMUNCSAU=paste0(CODMUNI,
    CSAU))
71 tablacsau_89<-sqldf("select
    tabla_89_sau.CODMUNI,tabla_89_sau.CSAU,csau.total from
    tabla_89_sau left join csau on
    tabla_89_sau.CODMUNCSAU=csau.CODMUNCSAU")
72 tablacsau_89[is.na(tablacsau_89)] <- 0
73 proba_csau<-reshape(tablacsau_89, timevar="CSAU",
    idvar=c("CODMUNI"), direction="wide")
74 #5/ 5 a 50/ 50 a 500/ 500 --> 01-08/ 9-12/ 13-18/ 19-21
75 proba_csau <- proba_csau %>%
76 mutate(
77 CSAU_5 = rowSums(select(., starts_with("total.0"))[,1:8], na.rm
    = TRUE),
78 CSAU_5_50 = rowSums(select(.,
    matches("total.0[9]$|total.1[0-2]$")), na.rm = TRUE),
79 CSAU_50_500 = rowSums(select(.,
    starts_with("total.")[, -c(1:12, 19:21)], na.rm = TRUE),
80 CSAU_500 = rowSums(select(., starts_with("total."))[, -c(1:18)],
    na.rm = TRUE)
81 ) %>%
82 select(CODMUNI, CSAU_5, CSAU_5_50, CSAU_50_500, CSAU_500)
83 csau_resul <- paste0("CSAU_", sprintf("%02d", numeros[i]),
    ".txt")
84 write.csv(proba_csau, file = csau_resul, quote = FALSE,
    row.names = FALSE)
85 #-----FORMACION-----
86 #sin estudios --> 1/// 2, 3 y 4 son formaciones educativa. PJ=
    1; jefe expl
87
88 formacion<-sqldf("select PROV, MUNI,count(FORMACI)as total,
    FORMACI from provin_nivel1_temp group by MUNI, FORMACI")
89 titularesmuni<-sqldf("select MUNI,sum(total) AS ntitulares from
    formacion group by MUNI")
90 unir<-sqldf("select formacion.PROV,
    formacion.MUNI,FORMACI,total, ntitulares from formacion left
    join titularesmuni on formacion.MUNI=titularesmuni.MUNI")
91 selectunion<-sqldf("select * from unir where FORMACI='1'")
92 selectunion["sin_est"] <- NA
93 selectunion$`sin_est`<-
    (selectunion$total*100)/selectunion$ntitulares
94 selectunion <- selectunion %>%
95 mutate(CODMUNI = paste0(PROV, MUNI)) %>%
96 select(CODMUNI, sin_est)

```

```

97     formacion_resul <- paste0("FORMACION_", sprintf("%02d",
98     numeros[i]), ".txt")
99     write.csv(selectunion, file = formacion_resul, quote = FALSE,
100     row.names = FALSE)
101
102     #-----UGT-----
103     uni_gan <- sqldf("SELECT PROV, MUNI, UGT from
104     provin_nivel1_temp")
105     uni_gan$`UGT_decimal` <- uni_gan$UGT/1000
106     uni_ganaderas <- transform(uni_gan, CODMUNI = paste0(PROV,
107     MUNI))
108     ugt_89 <- sqldf("select CODMUNI, sum(UGT_decimal) as UGT from
109     uni_ganaderas group by CODMUNI")
110     ugt_resul <- paste0("UGT_", sprintf("%02d", numeros[i]), ".txt")
111     write.csv(ugt_89, file = ugt_resul, quote = FALSE, row.names =
112     FALSE)
113
114     #-----EDAD-----
115     #provincia es donde tengo el scan del archivo
116     df <- data.frame(provincia)
117     df1 <- sqldf("select provincia as tipo from df where tipo LIKE
118     '1%' or tipo LIKE '4800%'")
119     write.table(df1, file = "df1", quote = FALSE, row.names =
120     FALSE, col.names = FALSE)
121     longitud_edad <- c(17,1)
122     variable_edad <- c("REGISTRO", "PJ")
123     clases_edad <- c("character", "character")
124     df2 <- read.fwf("df1", widths = longitud_edad, col.names =
125     variable_edad, colClasses = clases_edad)
126
127     df3 <- sqldf("select REGISTRO as tipo1 from df2 where PJ = 1 or
128     REGISTRO like '4%' ")
129
130     dftipo1 <- sqldf("select tipo1 from df3 where tipo1 like '1%'")
131     write.table (dftipo1,
132     file="dftipo1",quote=FALSE,row.names=FALSE,col.names= FALSE)
133     longitudt1<- c(1,5)
134     variabelt1<- c("TIPO","CODMUNI")
135     clasest1<- c("character","character")
136     dfcodmuni<- read.fwf("dftipo1",widths = longitudt1, col.names =
137     variabelt1, colClasses= clasest1)
138     dfcodmuni["n"] <- 1:nrow(dfcodmuni)
139     dfcodmuni<-sqldf("select CODMUNI, n from dfcodmuni")
140
141     dftipo4<-sqldf("select REGISTRO as tipo4 from df2 where REGISTRO
142     like '4%' ")
143     write.table (dftipo4,
144     file="dftipo4",quote=FALSE,row.names=FALSE,col.names= FALSE)
145     longitudt4<- c(5,2)
146     variabelt4<- c("BORRAR","EDAD")
147     clasest4<- c("character","character")
148     dfedad<- read.fwf("dftipo4",widths = longitudt4, col.names =
149     variabelt4, colClasses= clasest4)
150

```

```

136     dfedad$EDAD[dfedad$EDAD >= 55]<- 1
137     dfedad$EDAD[dfedad$EDAD > 1]<- 0
138     dfedad["n"] <- 1:nrow(dfedad)
139     dfedad<-sqldf("select EDAD, n from dfedad")
140
141     dfmerge<- merge(dfcodmuni,dfedad,"n")
142     dfmerge<-sqldf("select CODMUNI, EDAD from dfmerge")
143
144     edad89<-sqldf("select CODMUNI,COUNT(EDAD) as total, EDAD from
dfmerge group by CODMUNI, EDAD")
145     proba<-reshape(edad89, timevar="EDAD", idvar=c("CODMUNI"),
direction="wide")
146     proba["totalEADADES"] <- (proba$total.0' + proba$total.1')
147     proba["PGT55Y89"] <- (proba$total.1'*100)/proba$totalEADADES
148     tablaedad89<-sqldf("select CODMUNI,PGT55Y89 from proba")
149     edad_resul <- paste0("EDAD_", sprintf("%02d", numeros[i]),
".txt")
150     write.csv(tablaedad89, file = edad_resul, quote = FALSE,
row.names = FALSE)
151
152
153     #-----TENENCIA-----
154     ten <- provincia[greg("1|^211",provincia)]
155     longitud<- c(14)
156     variable<- c("NIVEL2")
157     clases<- c("character")
158     nombre_archivo_temporal_ten <- paste0("P", sprintf("%02d",
numeros[i]), "_ten")
159     write.table (ten, file = nombre_archivo_temporal_ten,
quote=FALSE,row.names=FALSE,col.names= FALSE)
160     ten1 <- read.fwf(nombre_archivo_temporal_ten, widths =
longitud, col.names = variable, colClasses= clases)
161     write.table (ten1, file = nombre_archivo_temporal_ten,
quote=FALSE,row.names=FALSE,col.names= FALSE)
162     lines <- readLines(nombre_archivo_temporal_ten)
163     new_lines <- c(lines[1])
164     for (j in 2:length(lines)) {
165         if (grepl("[0-9]$", new_lines[length(new_lines)]) &&
substr(lines[j], 1, 1) == "2") {
166             new_lines[length(new_lines)] =
paste0(new_lines[length(new_lines)], "x",
lines[j])
167         } else {
168             new_lines <- c(new_lines, lines[j])
169         }
170     }
171     writeLines(new_lines, nombre_archivo_temporal_ten)
172
173     longitud<- c(1,5,10,3,10,2,3,10,2,3,10,2,3,10,2,3,10)
174     variable<- c("NIVEL","CODMUNI","DELETE1","CODIGO1","DATO1","DEL
ETE2","CODIGO2","DATO2","DELETE3","CODIGO3","DATO3","DELETE4","
CODIGO4","DATO4","DELETE5","CODIGO5","DATO5")

```

```

175     clases<- c("character","character","character","character","num
eric","character","character","numeric","character","character"
,"numeric","character","character","numeric","character","chara
cter","numeric")
176     reg_ten1<- read.fwf(nombre_archivo_temporal_ten, widths =
longitud, col.names = variable, colClasses= clases)
177
178     reg_ten1[is.na(reg_ten1)] <- 0
179     r1<-sqldf("select CODMUNI, CODIGO1, DATO1 from reg_ten1")
180     names(r1)<-c("CODMUNI","codigo","dato")
181     r2<-sqldf("select CODMUNI, CODIGO2, DATO2 from reg_ten1")
182     names(r2)<-c("CODMUNI","codigo","dato")
183     r3<-sqldf("select CODMUNI, CODIGO3, DATO3 from reg_ten1")
184     names(r3)<-c("CODMUNI","codigo","dato")
185     r4<-sqldf("select CODMUNI, CODIGO4, DATO4 from reg_ten1")
186     names(r4)<-c("CODMUNI","codigo","dato")
187     r5<-sqldf("select CODMUNI, CODIGO5, DATO5 from reg_ten1")
188     names(r5)<-c("CODMUNI","codigo","dato")
189
190     r_merge<-rbind(r1,r2,r3,r4,r5)
191     r_merge<-sqldf("select CODMUNI, codigo, dato from r_merge where
codigo>0")
192     r_merge$`superficieha`<- r_merge$dato/100
193     sumaryze<-sqldf("select CODMUNI, codigo, sum(superficieha) from
r_merge group by CODMUNI, codigo")
194     proba<-reshape(sumaryze, timevar="codigo", idvar=c("CODMUNI"),
direction="wide")
195     names(proba)<-c("CODMUNI", "PROPIEDAD", "ARRENDAMIENTO", "APARCERIA", "OTROS")
196     proba[is.na(proba)] <- 0
197
198     proba$`PROPIEDAD` <- round((proba$PROPIEDAD*100)/(proba$PROPIE
DAD+proba$ARRENDAMIENTO+proba$APARCERIA+proba$OTROS),
2)
199     proba$`ARRENDAMIENTO` <- round((proba$ARRENDAMIENTO*100)/(prob
a$PROPIEDAD+proba$ARRENDAMIENTO+proba$APARCERIA+proba$OTROS),
2)
200     proba$`APARCERIA` <- round((proba$APARCERIA*100)/(proba$PROPIE
DAD+proba$ARRENDAMIENTO+proba$APARCERIA+proba$OTROS),
2)
201     proba$`OTROS` <- round((proba$OTROS*100)/(proba$PROPIEDAD+prob
a$ARRENDAMIENTO+proba$APARCERIA+proba$OTROS),
2)
202     probaselect<-sqldf("select CODMUNI, PROPIEDAD, ARRENDAMIENTO,
APARCERIA, OTROS from proba")
203
204     tenen_resul <- paste0("TENENCIA_", sprintf("%02d", numeros[i]),
".txt")
205     write.csv(probaselect, file = tenen_resul, quote = FALSE,
row.names = FALSE)
206
207     #-----MAQUINARIA-----
208
209     maq <- provincia[grep("^1|^275|^2760|^2761",provincia)]

```

```

210     longitud<- c(14)
211     variable<- c("NIVEL2")
212     clases<- c("character")
213     nombre_archivo_temporal_maq <- paste0("P", sprintf("%02d",
214     numeros[i]), "_maq")
215     write.table (maq, file=nombre_archivo_temporal_maq,quote=FALSE,
216     row.names=FALSE,col.names=
217     FALSE)
218     maq1 <- read.fwf(nombre_archivo_temporal_maq, widths =
219     longitud, col.names = variable, colClasses= clases)
220     write.table (maq1, file = nombre_archivo_temporal_maq,
221     quote=FALSE,row.names=FALSE,col.names= FALSE)
222
223     linesm <- readLines(nombre_archivo_temporal_maq)
224     new_linesm <- c(linesm[1])
225     for (k in 2:length(linesm)) {
226         if (grepl("[0-9]$", new_linesm[length(new_linesm)]) &&
227         substr(linesm[k], 1, 1) == "2") {
228             new_linesm[length(new_linesm)] =
229             paste0(new_linesm[length(new_linesm)], "x",
230             linesm[k])
231         } else {
232             new_linesm <- c(new_linesm, linesm[k])
233         }
234     }
235     writeLines(new_linesm, nombre_archivo_temporal_maq)
236
237     longitud<- c(1,5,10,3,10,2,3,10,2,3,10,2,3,10,2,3,10,2,3,10,2,3,
238     ,10,2,3,10,2,3,10,2,3,10,2,3,10,2,3,10,2,3,10,2,3,10,2,3,10)
239     variable<-
240     c("NIVEL", "CODMUNI", "DELETE1", "CODIGO1", "DATO1", "DELETE2", "CODI
241     GO2", "DATO2", "DELETE3", "CODIGO3", "DATO3", "DELETE4", "CODIGO4", "D
242     ATO4", "DELETE5", "CODIGO5", "DATO5", "DELETE6", "CODIGO6", "DATO6", "
243     DELETE7", "CODIGO7", "DATO7", "DELETE8", "CODIGO8", "DATO8", "DELETE9
244     ", "CODIGO9", "DATO9", "DELETE10", "CODIGO10", "DATO10", "DELETE11", "
245     CODIGO11", "DATO11", "DELETE12", "CODIGO12", "DATO12", "DELETE13", "C
246     ODIGO13", "DATO13", "DELETE14", "CODIGO14", "DATO14")
247     clases<-
248     c("character", "character", "character", "character", "numeric", "ch
249     aracter", "character", "numeric", "character", "character", "numeric
250     ", "character", "character", "numeric", "character", "character", "nu
251     meric", "character", "character", "numeric", "character", "character
252     ", "numeric", "character", "character", "numeric", "character", "char
253     acter", "numeric", "character", "character", "numeric", "character",
254     "character", "numeric", "character", "character", "numeric", "charac
255     ter", "character", "numeric", "character", "character", "numeric")
256     maquin<- read.fwf(nombre_archivo_temporal_maq,widths =
257     longitud, col.names = variable, colClasses= clases)
258     maquin[is.na(maquin)] <- 0
259     m1<-sqldf("select CODMUNI, CODIGO1, DATO1 from maquin")
260     names(m1)<-c("CODMUNI", "codigo", "dato")
261     m2<-sqldf("select CODMUNI, CODIGO2, DATO2 from maquin")
262     names(m2)<-c("CODMUNI", "codigo", "dato")
263     m3<-sqldf("select CODMUNI, CODIGO3, DATO3 from maquin")

```

```

239     names(m3)<-c("CODMUNI","codigo","dato")
240     m4<-sqldf("select CODMUNI, CODIGO4, DATO4 from maquin")
241     names(m4)<-c("CODMUNI","codigo","dato")
242     m5<-sqldf("select CODMUNI, CODIGO5, DATO5 from maquin")
243     names(m5)<-c("CODMUNI","codigo","dato")
244     m6<-sqldf("select CODMUNI, CODIGO6, DATO6 from maquin")
245     names(m6)<-c("CODMUNI","codigo","dato")
246     m7<-sqldf("select CODMUNI, CODIGO7, DATO7 from maquin")
247     names(m7)<-c("CODMUNI","codigo","dato")
248     m8<-sqldf("select CODMUNI, CODIGO8, DATO8 from maquin")
249     names(m8)<-c("CODMUNI","codigo","dato")
250     m9<-sqldf("select CODMUNI, CODIGO9, DATO9 from maquin")
251     names(m9)<-c("CODMUNI","codigo","dato")
252     m10<-sqldf("select CODMUNI, CODIGO10, DATO10 from maquin")
253     names(m10)<-c("CODMUNI","codigo","dato")
254     m11<-sqldf("select CODMUNI, CODIGO11, DATO11 from maquin")
255     names(m11)<-c("CODMUNI","codigo","dato")
256     m12<-sqldf("select CODMUNI, CODIGO12, DATO12 from maquin")
257     names(m12)<-c("CODMUNI","codigo","dato")
258     m13<-sqldf("select CODMUNI, CODIGO13, DATO13 from maquin")
259     names(m13)<-c("CODMUNI","codigo","dato")
260     m14<-sqldf("select CODMUNI, CODIGO14, DATO14 from maquin")
261     names(m14)<-c("CODMUNI","codigo","dato")
262
263     maquin_merge<-rbind(m1,m1,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12,m13,
264     m14)
265     maquin_merge<-sqldf("select CODMUNI, codigo, dato from
266     maquin_merge where codigo>0")
267
268     maquin_merge$`nmaquinaria`<- maquin_merge$dato/100
269     machin<-sqldf("select CODMUNI, sum(nmaquinaria) as nmachin from
270     maquin_merge group by CODMUNI")
271     machin_resul <- paste0("MAQUINARIA_", sprintf("%02d",
272     numeros[i]), ".txt")
273     write.csv(machin, file = machin_resul, quote = FALSE, row.names
274     = FALSE)
275
276 }
277
278 #-----
279 library(dplyr)
280
281 variables <- c("CST", "CSAU", "NEXPLOT", "FORMACION", "EDAD",
282 "TENENCIA", "MAQUINARIA", "UGT")
283 numeros <- c(1:6, 8:34, 36:37, 39:50)
284
285 # Parte 1: Combina los archivos que comparten el mismo número
286 for (numero in numeros) {
287
288     # Comenzar con un dataframe vacío para cada número
289     df_final_numero <- data.frame()
290
291     for (variable in variables) {

```

```

286     archivo_name <- paste0(variable, "_", sprintf("%02d",
287     numero), ".txt")
288
289     if (file.exists(archivo_name)) {
290         df_actual <- read.csv(archivo_name,
291         stringsAsFactors = FALSE)
292
293         # Si el df_final_numero est vaco, copia
294         df_actual en l
295         if (nrow(df_final_numero) == 0) {
296             df_final_numero <- df_actual
297         } else {
298             # Une el df_actual a df_final_numero
299             usando CODMUNI como clave
300             df_final_numero <-
301             left_join(df_final_numero, df_actual,
302             by = "CODMUNI")
303         }
304     }
305
306     # Guarda el archivo combinado para este número
307     write.csv(df_final_numero, file = paste0("FINAL_",
308     sprintf("%02d", numero), ".txt"), quote = FALSE, row.names =
309     FALSE)
310 }
311
312 # Parte 2: Combina todos los archivos "FINAL_**" en uno
313 df_global <- data.frame()
314
315 for (numero in numeros) {
316     archivo_final_name <- paste0("FINAL_", sprintf("%02d", numero),
317     ".txt")
318
319     if (file.exists(archivo_final_name)) {
320         df_actual <- read.csv(archivo_final_name,
321         stringsAsFactors = FALSE)
322
323         # Agrega el df_actual al df_global
324         df_global <- bind_rows(df_global, df_actual)
325     }
326 }
327
328 # Guarda el archivo global
329 write.csv(df_global, file = "datos_1989.txt", quote = FALSE, row.names
330 = FALSE)

```

Cálculo de las variables para el año 2020

```

1     #Carga de la base de datos
2     #install.packages("rstudioapi")

```

```

3 setwd("C:/TFM/censo_agrario/2020/codigo_final_censo_2020")
4 #setwd(dirname(rstudioapi::getSourceEditorContext()$path))
5
6 #paquetes
7 #install.packages("writexl")
8 #install.packages("tidyverse")
9 #install.packages("readxl")
10
11 #librerías
12 library(dplyr)
13 library(writexl)
14 library(tidyverse)
15 library(tidyr)
16 library(sqldf)
17 library(readxl)
18 #-----
19
20 variable <- c("IDEXP", "ESTRA1", "ESTRA2", "M1", "M2", "M3", "FACT1",
21 "FACT2", "FACT3", "OTE", "PET", "CCAA",
22 "PROV", "COM", "MUNI", "PJUR", "JEANAC", "JESEXO", "JEUTATR", "JEAINI",
23 "JEFOR", "JECUR", "ST",
24 "SAU", "SAURTPRO", "SAURTARR", "SAURTOTR", "SAURTCOM", "SAUO", "TA",
25 "TL", "HERBAC", "CEREAL",
26 "V210", "V211", "V212", "V213", "V214", "V215", "V216", "V217", "V218",
27 "V219", "LEGUM", "V220",
28 "V221", "TUBER", "V222", "V223", "V224", "INDUS", "TEXTIL", "V230",
29 "V231", "V232", "V233", "OLEAG",
30 "V234", "V235", "V236", "V237", "V238", "V239", "V240", "V241", "V242",
31 "V243", "CVERDE", "V225",
32 "V226", "V227", "V228", "V229", "HORTAL", "V244", "V245", "V246",
33 "V247", "V248", "V250", "PERMA",
34 "CITRICO", "V260", "V261", "V262", "V263", "V264", "FRUOT", "FRUPE",
35 "V265", "V266", "V267", "FRUHU",
36 "V268", "V269", "V270", "V271", "V272", "V274", "V275", "V276", "UVA",
37 "V277", "V278", "UVAVINO",
38 "V279", "V280", "V281", "V282", "V283", "V284", "PASTOS", "V294",
39 "V299", "V300", "HUERTOS", "OT",
40 "NOSAU", "FOREST", "V304", "ERAS", "HONGOS", "SAUI", "V200", "V201",
41 "V202_3", "V204", "V205",
42 "SAUECO", "SAUECOCER", "SAUECOCON", "SAUOECO", "TAECO", "CEREALECO",
43 "V352", "V353", "V354", "V355",
44 "TUBERECO", "V356", "V357", "V358", "INDUSECO", "V359", "V360", "V361",
45 "CVERDEECO", "V362", "V363",
46 "V364", "V365", "V367", "V368", "PERMANECO", "V369", "V370", "V371",
47 "V372", "V373", "PASTOSECO",
48 "V374", "V375", "V376", "V366", "V379", "RIEGO", "BOVINOS", "V402",
49 "V403", "BOHEMBRA", "V406", "B012",
50 "V404", "V405", "VACAS", "LECHERAS", "OTRVACAS", "BUFALAS", "OVINOS",
51 "V407", "V408", "CAPRINOS", "V409",
52 "V410", "PORCINOS", "V412", "V413", "V414", "AVES", "V415", "V416",
53 "V417", "V418", "V419", "V420", "V421",
54 "CONEJAS", "COLMENAS", "V411", "BOVECO", "V430", "V431", "V433",
55 "OVIECO", "CAPECO", "PORECO", "AVEECO",

```

```

38 "V437", "V438", "V439", "TITSEXO", "TITANAC", "TITUTATR", "V600",
    "FAMHPC1T24", "FAMHPC25T49", "FAMHPC50",
39 "FAMHPC51T74", "FAMHPC75T99", "FAMHPC100", "FAMMPC1T24", "FAMMPC25T49",
    "FAMMPC50", "FAMMPC51T74",
40 "FAMMPC75T99", "FAMMPC100", "NOFAMHPC1T24", "NOFAMHPC25T49",
    "NOFAMHPC50", "NOFAMHPC51T74", "NOFAMHPC75T99",
41 "NOFAMHPC100", "NOFAMMPC1T24", "NOFAMMPC25T49", "NOFAMMPC50",
    "NOFAMMPC51T74", "NOFAMMPC75T99", "NOFAMMPC100",
42 "V551", "V553", "UTATOT", "UTATI", "UTAFAM", "UTAFAMC", "UTAFAMRF",
    "UTANOFAM", "UTARA", "UTATEM",
43 "UTASUBC", "V561", "V562", "V563", "V564", "V565", "V566", "V567",
    "V568", "V569", "V570", "V571",
44 "V572", "V526", "V575", "V576", "V577", "V578", "V527", "V579", "V580",
    "DR1", "DR2", "DR3", "DR4",
45 "DR5", "DR6", "DR8", "DR9", "DR10", "DR11", "DR12", "DR13", "DR14",
    "V464", "V450", "V451", "V452",
46 "V453", "V454", "V455", "V456", "V466", "V468", "V465", "V457", "V458",
    "V459", "V460", "V461", "V462",
47 "V463", "V467", "V469", "V482", "V470", "V471", "V472", "V473", "V474",
    "V475", "V484", "V483", "V476",
48 "V477", "V478", "V479", "V480", "V481", "V485", "V493", "V486", "V487",
    "V488", "V489", "V490", "V491",
49 "V492", "V816", "V817", "V823", "V822", "V821", "V820", "V818", "V824",
    "V825", "V826", "V827", "V828",
50 "V829", "V830", "V800", "V801", "V802", "V803", "V804", "V805", "V806",
    "V807", "V808", "V809", "V810",
51 "V811", "V812", "V813", "UGT")
52 longitud <- c(10,5,5,1,1,1,9,9,9,3,13,2,2,4,5,3,4,1,1,4,1,1,9, 9, 9, 9,
    9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,
53 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,
    9, 9, 9, 9, 9, 9, 9, 9, 9,
54 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,
    9, 9, 9, 9, 9, 9, 9, 9, 9,
55 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9,
    9, 9, 9, 9, 9, 9, 9, 9, 9,
56 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 9, 6,
    6, 6, 6, 6, 6, 6, 6, 6, 6,
57 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
    6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,
58 4, 1, 1, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5,
    5, 5, 5, 6, 6, 8, 5, 8, 8, 8,
59 8, 8, 8, 8, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 1, 2, 2,
    1, 1, 1, 1, 1, 1, 1, 1, 1,
60 1, 1, 1, 7, 7, 7, 7, 7, 7, 7, 7, 2, 1, 7, 7, 7, 7, 7, 7, 7, 7, 2, 1, 7,
    7, 7, 7, 7, 7, 2, 7, 7, 7,
61 7, 7, 7, 7, 1, 7, 7, 7, 7, 7, 7, 7, 7, 9, 9, 9, 9, 9, 9, 9, 9, 3, 3, 3, 3,
    3, 3, 3, 3, 3, 3, 3, 3, 3,
62 3, 3, 2, 2, 2, 2, 2, 10)
63 tipo <- c("character", "character", "character", "character",
    "character", "character", "numeric", "numeric", "numeric", "character",
64 "numeric", "character", "character", "character", "character",
    "character", "numeric", "character", "character", "character",
65 "character", "character", "numeric", "numeric", "numeric", "numeric",
    "numeric", "numeric", "numeric", "numeric", "numeric",

```



```

91     "numeric", "numeric", "numeric", "numeric", "numeric", "numeric",
92     "numeric", "numeric", "numeric", "numeric", "numeric", "numeric",
93     "numeric",
94     "numeric", "numeric", "numeric", "numeric", "numeric", "numeric",
95     "numeric", "numeric", "numeric", "numeric", "numeric", "numeric",
96     "numeric",
97     "numeric")
98
99 censo2020 <- read.fwf("MICRODAT_230316.txt", widths = longitud,
100 col.names = variable, colClasses = tipo)
101 View(censo2020)
102 censo2020_dfa <- as.data.frame(censo2020)
103 censo_agrario <- censo2020_dfa %>%
104 select(MUNI, PROV, COM, ST, SAU, JEANAC, JEFOR, SAURTPRO, SAURTARR,
105 SAURTOTR, SAURTCOM, UGT)
106
107 str(censo_agrario)
108 ls(censo_agrario)
109 mode(censo_agrario)
110
111 #####SUPERFICIE TOTAL
112
113 #Hacer clasificacion ST_5; ST_5_50; ST_50_500; ST_500
114
115 breaks_ST <- c(-Inf, 5, 50, 500, Inf)
116 labels_ST <- c("ST_5", "ST_5_50", "ST_50_500", "ST_500")
117
118 #Crear las columnas
119
120 censo_ST <- subset(censo_agrario, select = c("MUNI", "ST"))
121 head(censo_ST)
122
123 censo_ST$categoria_ST <- cut(censo_ST$ST, breaks = breaks_ST, labels =
124 labels_ST, right = FALSE, include.lowest = TRUE)
125 #help(cut)
126 View(censo_ST)
127
128 CST <- censo_ST %>%
129 group_by(MUNI, categoria_ST) %>%
130 summarise(total_ha = sum(ST)) %>%
131 pivot_wider(names_from = categoria_ST, values_from = total_ha,
132 values_fill = 0)
133 head(CST)
134 View(CST)
135
136 #Guardar data.frame en tabla
137
138 write_xlsx(CST, "CST.xlsx")
139
140 #####NEXPLOT. Es el conteo en cada muni de explotaciones
141
142 explotaciones <- subset(censo_agrario, select = c("MUNI", "ST", "PROV",
143 "COM"))

```

```

136 n_explot <- sqldf("select PROV, COM, MUNI, COUNT(ST) as total, ST from
137 explotaciones group by MUNI, ST")
138 View(n_explot)
139
140 NEXPLOT <- n_explot %>%
141 group_by(MUNI) %>%
142 summarise(explot = sum(total))
143 View(NEXPLOT)
144
145 # Añadir las columnas "PROV" y "COM" a NEXPLOT
146 NEXPLOT <- n_explot %>%
147 group_by(MUNI) %>%
148 summarise(explot = sum(total)) %>%
149 left_join(n_explot %>% distinct(MUNI, PROV, COM), by = "MUNI")
150 View(NEXPLOT)
151
152
153 write_xlsx(NEXPLOT, "NEXPLOT.xlsx")
154
155
156 #####CSAU
157
158 #Hacer clasificacion SAU_5; SAU_5_50; SAU_50_500; SAU_500
159
160 breaks_SAU <- c(-Inf, 5, 50, 500, Inf)
161 labels_SAU <- c("SAU_5", "SAU_5_50", "SAU_50_500", "SAU_500")
162
163 #Crear las columnas
164
165 censo_SAU <- subset(censo_agrario, select = c("MUNI", "SAU"))
166 head(censo_SAU)
167
168 censo_SAU$categoria_SAU <- cut(censo_SAU$SAU, breaks = breaks_SAU,
169 labels = labels_SAU, right = FALSE, include.lowest = TRUE)
170 View(censo_SAU)
171
172 CSAU <- censo_SAU %>%
173 group_by(MUNI, categoria_SAU) %>%
174 summarise(total_ha2 = sum(SAU)) %>%
175 pivot_wider(names_from = categoria_SAU, values_from = total_ha2,
176 values_fill = 0)
177 View(CSAU)
178
179 #Guardar data.frame en tabla
180
181
182 write_xlsx(CSAU, "CSAU.xlsx")
183
184 #####EDAD DEL JEFE
185
186 nacimiento <- subset(censo_agrario, select = c("MUNI", "JEANAC"))

```

```

187
188     edad <- nacimiento %>%
189     mutate(edad_jef = 2020 - JEANAC) %>%
190     select(MUNI, edad_jef)
191     View(edad)
192
193     p_mayor_55 <- edad %>%
194     mutate(mayor_55 = ifelse(edad_jef > 55, 1, 0)) %>%
195     select(MUNI, mayor_55)
196     View(p_mayor_55)
197
198     M55 <- p_mayor_55 %>%
199     group_by(MUNI) %>%
200     summarise(total_edad = length(mayor_55), total_m55 = sum(mayor_55)) %>%
201     select(MUNI, total_edad, total_m55)
202     View(M55)
203
204     P_M55 <- M55 %>%
205     mutate(Porcen_M55 = (total_m55 /total_edad) * 100) %>%
206     select(MUNI, Porcen_M55)
207     View(P_M55)
208
209
210     write_xlsx(P_M55, "EDAD.xlsx")
211
212     #####FORMACION. Hay que seleccionar el 1: solo exp. practica agraria
213     (sera como SIN ESTUDIOS)
214
215     formacion <- subset(censo_agrario, select = c("MUNI", "JEFOR"))
216     View(formacion)
217
218     estud <- formacion %>%
219     mutate(JEFOR = as.numeric(as.character(JEFOR))) %>%
220     group_by(MUNI) %>%
221     summarise(total_estu = sum(JEFOR), total_sin_estu = sum(JEFOR == 1))
222     View(estud)
223
224     P_sin_est <- estud %>%
225     mutate(Porcen_sinEst = (total_sin_estu/total_estu) * 100) %>%
226     select(MUNI, Porcen_sinEst)
227     View(P_sin_est)
228
229     write_xlsx(P_sin_est, "FORMACION.xlsx")
230
231     #####PORCENTAJE DE REGIMENES DE TENENCIA "SAURTPRO", "SAURTARR",
232     "SAURTOTR" Y "SAURTCOM"; rgimen comunitario
233
234     reg_tenencia <- subset(censo_agrario, select = c("MUNI", "SAURTPRO",
235     "SAURTARR", "SAURTOTR", "SAURTCOM"))
236
237     regimen <- reg_tenencia %>%
238     mutate(tenencia_total = (SAURTPRO + SAURTARR + SAURTOTR + SAURTCOM)) %>%
239     mutate(propiedad = (SAURTPRO/tenencia_total) * 100) %>%

```

```

238 mutate(arrendam = (SAURTARR/tenencia_total) * 100) %>%
239 mutate(otros = (SAURTOTR/tenencia_total) *100) %>%
240 select(MUNI, propiedad, arrendam, otros)
241
242 regimen_agrupado <- regimen %>%
243 group_by(MUNI) %>%
244 summarize(propiedad_total = sum(propiedad, na.rm = TRUE),
245 arrendam_total = sum(arrendam, na.rm = TRUE),
246 otros_total = sum(otros, na.rm = TRUE),
247 tenencia_total = sum(propiedad_total, arrendam_total, otros_total,
na.rm = TRUE))
248
249 TENENCIA <- regimen_agrupado %>%
250 mutate(propiedad = (propiedad_total/tenencia_total) * 100) %>%
251 mutate(arrendam = (arrendam_total/tenencia_total) * 100) %>%
252 mutate(otros = (otros_total/tenencia_total) *100) %>%
253 select(MUNI, propiedad, arrendam, otros)
254 View(TENENCIA)
255
256 write_xlsx(TENENCIA, "TENENCIA.xlsx")
257
258
259 #####UNIDADES GANADERAS
260
261 uni_gan <- subset(censo_agrario, select = c("MUNI", "UGT"))
262 uni_gan$UGT <- uni_gan$UGT / 1000
263 View(uni_gan)
264
265 ugt <- uni_gan %>%
266 group_by(MUNI) %>%
267 summarise(UGT = sum(UGT))
268
269 View(ugt)
270
271 write_xlsx(ugt, "UGT.xlsx")
272
273
274 #-----juntarlos todos en un mismo excel
275 list.files(pattern = "\\\\.xlsx$")
276 lista_df <- c("CSAU.xlsx", "CST.xlsx", "EDAD.xlsx", "FORMACION.xlsx", "NEXP_
LOT.xlsx", "TENENCIA.xlsx", "UGT.xlsx")
277 data_frames <- lapply(lista_df, function(x) read_excel(x))
278
279 df_final <- Reduce(function(x, y) merge(x, y, by = "MUNI", all = TRUE),
data_frames)
280 View(df_final)
281
282 write_xlsx(df_final, "CENSO_AGRARIO_2020.xlsx")

```

Estadística descriptiva de las variables socioeconómicas y agrarias

```
1      # Instalar y cargar paquetes necesarios
2
3      library(ggplot2)
4      library(FactoMineR)
5      library(factoextra)
6      library(dplyr)
7      library(tidyr)
8      library(pheatmap) # Para el mapa de calor
9      library(reshape2) # Para la matriz de correlación
10     library(SingleCellExperiment) # Asumiendo que se usa el formato SCE
11     library(corrplot)
12     library(readr)
13     library(readxl)
14     library(gridExtra)
15     library(grid)
16     library(sf)
17     library(ggpubr)
18     #install.packages("plsr")
19     #install.packages("pls")
20     library(pls)
21     #install.packages("sf")
22     #install.packages("cowplot")
23     #install.packages("viridis")
24
25     library(ggplot2)
26     library(cowplot)
27     library(sf)
28
29
30     ##### ESTADSTICA DESCRIPTIVA SOCIOECONOMA
31     #####
32
33     base_datos_2009 <- as.data.frame(read.table("C:/TFM/ANALISIS_EST_DESCRI_
34     PT/base_datos_09.txt", head =
35     TRUE))
36     base_datos_2020 <- as.data.frame(read.table("C:/TFM/ANALISIS_EST_DESCRI_
37     PT/base_datos_20.txt", head =
38     TRUE))
39
40     base_datos_2009$year <- 2009
41     base_datos_2020$year <- 2020
42
43
44     # Seleccionar las variables de interés
45     selected_vars <- c("GENERAL", "AGRARIO", "MAR", "HOGAR", "AUTONOMOS",
46     "CARBON", "TOTAL", "deuda_miles_euros",
47     "TCV", "total_paro", "paro_hombre_25", "paro_hombre_25_45",
48     "paro_hombre_45", "paro_mujer_25", "paro_mujer_25_45", "paro_mujer_45",
49     "paro_agricultura", "paro_industria", "paro_construccion",
50     "paro_servicios", "paro_sin_empleo_anterior")
```

```

45
46
47   # Seleccionar las columnas usando all_of()
48   socio_2009_selected <- base_datos_2009 %>%
49   select(all_of(selected_vars), year)
50
51   socio_2020_selected <- base_datos_2020 %>%
52   select(all_of(selected_vars), year)
53
54
55   # Combinar los datos de ambos años
56   socio_combined <- bind_rows(socio_2009_selected, socio_2020_selected)
57
58
59   write.table(socio_combined, file =
60   "C:/TFM/variables_piroregiones/variables
61   socioeconomicas/ANALISIS_ESTADISTICOS/socio_09_20_porcent.txt",
62   row.names = FALSE)
63
64
65   # Convertir a formato largo para ggplot
66   socio_long <- gather(socio_combined, key = "variable", value = "value",
67   -year)
68
69
70   #####BOXPLOTS
71   # Lista de variables para cada grupo de boxplots
72   grupo1 <- c("PGENERAL", "PAGRARIO", "PMAR", "PHOGAR", "PAUTONOMOS",
73   "PCARBON", "TOTAL")
74   grupo2 <- c("deuda_miles_euros")
75   grupo3 <- c("TCV")
76   grupo4 <- c("total_paro", "Pparo_hombre_25", "Pparo_hombre_25_45",
77   "Pparo_hombre_45",
78   "Pparo_mujer_25", "Pparo_mujer_25_45", "Pparo_mujer_45",
79   "Pparo_agricultura", "Pparo_industria", "Pparo_construccion",
80   "Pparo_servicios", "Pparo_sin_empleo_anterior")
81
82   # Estandarización logarítmica
83
84   socio_long_scaled_log <- socio_long %>%
85   mutate(value_log = log(value + 1)) # Agrega 1 para evitar log(0)
86
87
88   crear_boxplot <- function(vars, title) {
89     socio_long_scaled_log %>%
90     filter(variable %in% vars) %>%
91     ggplot(aes(x = factor(year), y = value_log, fill =
92     factor(year))) +
93     stat_boxplot(geom = "errorbar", width = 0.25, color = "black") +
94     geom_boxplot(
95     fill = "grey",
96     colour = "black",
97     alpha = 0.5,
98     outlier.colour = "tomato2",
99     outlier.shape = 16,
100    outlier.size = 2,
101    outlier.stroke = 0.5,

```

```

90         coef = 1.5
91     ) +
92     facet_wrap(~ variable, scales = "free_y") +
93     labs(title = title, x = "Año", y = "Valor Logarítmico") +
94     theme_minimal() +
95     theme(
96         text = element_text(family = "Garamond"),
97         legend.position = "none"
98     )
99 }
100
101 # Crear boxplots
102 boxplot1 <- crear_boxplot(grupo1, "Afiliados a la Seguridad Social por
103 sectores")
104 print(boxplot1)
105
106 boxplot2 <- crear_boxplot(grupo2, "Deuda de los Ayuntamientos")
107 print(boxplot2)
108
109 boxplot3 <- crear_boxplot(grupo3, "Tasa de crecimiento vegetativo")
110 print(boxplot3)
111
112 boxplot4 <- crear_boxplot(grupo4, "Paro por sectores econmicos, sexo y
113 edad")
114 print(boxplot4)
115
116 #####Histogramas de densidad
117
118 grupo1 <- c("GENERAL", "AGRARIO", "HOGAR", "AUTONOMOS", "TOTAL")
119 grupo2 <- c("MAR", "CARBON")
120 grupo3 <- c("deuda_miles_euros")
121 grupo4 <- c("TCV")
122 grupo5 <- c("total_paro", "paro_hombre_25", "paro_hombre_25_45",
123 "paro_hombre_45",
124 "paro_mujer_25", "paro_mujer_25_45", "paro_mujer_45",
125 "paro_agricultura", "paro_industria", "paro_construccion",
126 "paro_servicios", "paro_sin_empleo_anterior")
127
128 # Filtrar datos para grupo1 para los años 2009 y 2020
129 socio_group1 <- socio_long_scaled_log %>%
130 filter(variable %in% grupo1, year %in% c(2009, 2020))
131
132 socio_group2 <- socio_long_scaled_log %>%
133 filter(variable %in% grupo2, year %in% c(2009, 2020))
134
135 socio_group3 <- socio_long_scaled_log %>%
136 filter(variable %in% grupo3, year %in% c(2009, 2020))
137
138 socio_group4 <- socio_long_scaled_log %>%
139 filter(variable %in% grupo4, year %in% c(2009, 2020))
140
141 socio_group5 <- socio_long_scaled_log %>%

```

```

141 filter(variable %in% grupo5, year %in% c(2009, 2020))
142
143 # Crear el gráfico de densidad para los grupos:
144
145 ggplot(socio_group1, aes(x = value_log, color = as.factor(year), fill =
146 as.factor(year))) +
147 geom_density(alpha = 0.4) + # Densidad con
148 transparencia
149 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
150 color = "black") + # Línea de la media
151 labs(title = "Densidad de las variables relativas a la afiliación a la
152 seguridad social por sectores en 2009 y 2020",
153 x = "Log(Valor)", y = "Densidad") +
154 scale_fill_manual(values = c("2009" = "lightblue", "2020" = "red")) + #
155 Colores personalizados
156 scale_color_manual(values = c("2009" = "darkblue", "2020" = "brown")) +
157 # Colores de líneas
158 theme_minimal() +
159 ylim(0, 0.77) + # Límite del eje y
160 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
161 variable, escalas independientes en el eje x
162 theme(legend.position = "bottom") # Leyenda en la parte
163 inferior
164
165 ggplot(socio_group2, aes(x = value_log, color = as.factor(year), fill =
166 as.factor(year))) +
167 geom_density(alpha = 0.4) + # Densidad con
168 transparencia
169 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
170 color = "black") + # Línea de la media
171 labs(title = "Densidad de las variables CARBON y MAR para 2009 y 2020",
172 x = "Log(Valor)", y = "Densidad") +
173 scale_fill_manual(values = c("2009" = "lightblue", "2020" = "red")) + #
174 Colores personalizados
175 scale_color_manual(values = c("2009" = "darkblue", "2020" = "brown")) +
176 # Colores de líneas
177 theme_minimal() +
178 ylim(0, 3.5) + # Límite del eje y
179 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
180 variable, escalas independientes en el eje x
181 theme(legend.position = "bottom") # Leyenda en la parte
182 inferior
183
184 ggplot(socio_group3, aes(x = value_log, color = as.factor(year), fill =
185 as.factor(year))) +
186 geom_density(alpha = 0.4) + # Densidad con
187 transparencia
188 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
189 color = "black") + # Línea de la media
190 labs(title = "Densidad de la variable deuda por Ayto. para 2009 y 2020",
191 x = "Log(Valor)", y = "Densidad") +

```

```

176 scale_fill_manual(values = c("2009" = "lightblue", "2020" = "red")) + #
    Colores personalizados
177 scale_color_manual(values = c("2009" = "darkblue", "2020" = "brown")) +
    # Colores de lineas
178 theme_minimal() +
179 ylim(0, 0.2) + # Límite del eje y
180 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
    variable, escalas independientes en el eje x
181 theme(legend.position = "bottom") # Leyenda en la parte
    inferior
182
183
184 ggplot(socio_group4, aes(x = value_log, color = as.factor(year), fill =
    as.factor(year))) +
185 geom_density(alpha = 0.4) + # Densidad con
    transparencia
186 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
    color = "black") + # Línea de la media
187 labs(title = "Densidad de la variable relativa a la tasa de crecimiento
    vegetativo para 2009 y 2020",
188 x = "Log(Valor)", y = "Densidad") +
189 scale_fill_manual(values = c("2009" = "lightblue", "2020" = "red")) + #
    Colores personalizados
190 scale_color_manual(values = c("2009" = "darkblue", "2020" = "brown")) +
    # Colores de lineas
191 theme_minimal() +
192 ylim(0, 0.2) + # Límite del eje y
193 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
    variable, escalas independientes en el eje x
194 theme(legend.position = "bottom") # Leyenda en la parte
    inferior
195
196
197 ggplot(socio_group5, aes(x = value_log, color = as.factor(year), fill =
    as.factor(year))) +
198 geom_density(alpha = 0.4) + # Densidad con
    transparencia
199 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
    color = "black") + # Línea de la media
200 labs(title = "Densidad de las variables relativas a la tasa de paro para
    2009 y 2020",
201 x = "Log(Valor)", y = "Densidad") +
202 scale_fill_manual(values = c("2009" = "lightblue", "2020" = "red")) + #
    Colores personalizados
203 scale_color_manual(values = c("2009" = "darkblue", "2020" = "brown")) +
    # Colores de lineas
204 theme_minimal() +
205 ylim(0, 0.4) + # Límite del eje y
206 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
    variable, escalas independientes en el eje x
207 theme(legend.position = "bottom") # Leyenda en la parte
    inferior
208
209

```

```

210 #####
211 #####ANÁLISIS EXPLORATORIO CENSO AGRARIO
212 # Seleccionar las variables de interés
213 selected_vara <- c("ST_5ha", "ST_5_50ha", "ST_50_500ha", "ST_500ha",
214 "SAU_5ha", "SAU_5_50ha",
215 "SAU_50_500ha", "SAU_500ha", "MAYORES_55", "NO_ESTUDIOS", "D_EXPLOTT",
216 "PROPIEDAD", "ARRENDAMIENTO", "OTROS", "D_UGT")
217
218 censo_agrario <- as.data.frame(read.table("C:/TFM/ANALISIS_EST_DESCRIP_T_
/censo_agrario.txt", head =
TRUE))
219 #tiempo = year
220
221 # Convertir a formato largo para ggplot
222 agra_long <- gather(censo_agrario, key = "variable", value = "value",
-tiempo)
223
224 #####BOXPLOTS
225 # Lista de variables para cada grupo de boxplots
226 grupo1 <- c("ST_5ha", "ST_5_50ha", "ST_50_500ha", "ST_500ha",
227 "SAU_5ha", "SAU_5_50ha",
228 "SAU_50_500ha", "SAU_500ha")
229 grupo2 <- c("MAYORES_55", "NO_ESTUDIOS")
230 grupo3 <- c("PROPIEDAD", "ARRENDAMIENTO", "OTROS")
231 grupo4 <- c("D_UGT", "D_EXPLOTT")
232
233 # Estandarización logarítmica
234
235 agra_long_scaled_log <- agra_long %>%
236 mutate(value_log = log(value + 1)) # Agrega 1 para evitar log(0)
237
238
239
240 crear_boxplot <- function(vars, title) {
241   agra_long_scaled_log %>%
242   filter(variable %in% vars) %>%
243   ggplot(aes(x = factor(tiempo), y = value_log, fill =
factor(tiempo))) +
244   stat_boxplot(geom = "errorbar", width = 0.25, color = "black") +
245   geom_boxplot(
246   fill = "grey",
247   colour = "black",
248   alpha = 0.5,
249   outlier.colour = "tomato2",
250   outlier.shape = 16,
251   outlier.size = 2,
252   outlier.stroke = 0.5,
253   coef = 1.5
254   ) +
255   facet_wrap(~ variable, scales = "free_y") +
256   labs(title = title, x = "Año", y = "Valor Logarítmico") +
257   theme_minimal() +

```

```

258         theme(
259           text = element_text(family = "Garamond"),
260           legend.position = "none"
261         )
262     }
263
264     # Crear boxplots
265     boxplot1 <- crear_boxplot(grupo1, "Superficies agrarias")
266     print(boxplot1)
267
268     boxplot2 <- crear_boxplot(grupo2, "Jefes de las explotaciones mayores de
269     55 aos y sin estudios")
270     print(boxplot2)
271
272     boxplot3 <- crear_boxplot(grupo3, "Rgimen de tenencia de las
273     explotaciones")
274     print(boxplot3)
275
276     boxplot4 <- crear_boxplot(grupo4, "Densidades ganaderas y de
277     explotaciones")
278     print(boxplot4)
279
280     #####Histogramas de densidad
281
282     grupo1 <- c("ST_5ha", "ST_5_50ha", "ST_50_500ha", "ST_500ha",
283     "SAU_5ha", "SAU_5_50ha",
284     "SAU_50_500ha", "SAU_500ha")
285     grupo2 <- c("MAYORES_55", "NO_ESTUDIOS")
286     grupo3 <- c("PROPIEDAD", "ARRENDAMIENTO", "OTROS")
287
288     # Filtrar datos para grupo1 para los años 2009 y 2020
289     agra_group1 <- agra_long_scaled_log %>%
290     filter(variable %in% grupo1)
291
292     agra_group2 <- agra_long_scaled_log %>%
293     filter(variable %in% grupo2)
294
295     agra_group3 <- agra_long_scaled_log %>%
296     filter(variable %in% grupo3)
297
298     # Crear el gráfico de densidad para los grupos:
299
300     ggplot(agra_group1, aes(x = value_log, color = as.factor(tiempo), fill
301     = as.factor(tiempo))) +
302     geom_density(alpha = 0.4) + # Densidad con
303     transparencia
304     geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
305     color = "black") + # Lnea de la media
306     labs(title = "Densidad de las variables relativas a las superficies
307     agrarias",
308     x = "Log(Valor)", y = "Densidad") +

```

```

302 scale_fill_manual(values = c("1982" = "cyan", "1989" = "red", "1999"
    = "blue", "2009" = "yellow", "2020" = "green")) + # Colores
    personalizados
303 scale_color_manual(values = c("1982" = "cyan", "1989" = "red", "1999"
    = "blue", "2009" = "yellow", "2020" = "green")) + # Colores de líneas
304 theme_minimal() +
305 ylim(0, 0.77) + # Límite del eje y
306 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
    variable, escalas independientes en el eje x
307 theme(legend.position = "bottom") # Leyenda en la parte
    inferior
308
309
310 ggplot(agra_group2, aes(x = value_log, color = as.factor(tiempo), fill
    = as.factor(tiempo))) +
311 geom_density(alpha = 0.4) + # Densidad con
    transparencia
312 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
    color = "black") + # Línea de la media
313 labs(title = "Densidad de los jefes de explotación mayores de 55 años y
    sin estudios",
314 x = "Log(Valor)", y = "Densidad") +
315 scale_fill_manual(values = c("1982" = "cyan", "1989" = "red", "1999"
    = "blue", "2009" = "yellow", "2020" = "green")) + # Colores
    personalizados
316 scale_color_manual(values = c("1982" = "cyan", "1989" = "red", "1999"
    = "blue", "2009" = "yellow", "2020" = "green")) + # Colores de líneas
317 theme_minimal() +
318 ylim(0, 3.5) + # Límite del eje y
319 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
    variable, escalas independientes en el eje x
320 theme(legend.position = "bottom") # Leyenda en la parte
    inferior
321
322 ggplot(agra_group3, aes(x = value_log, color = as.factor(tiempo), fill
    = as.factor(tiempo))) +
323 geom_density(alpha = 0.4) + # Densidad con
    transparencia
324 geom_vline(aes(xintercept = mean(value_log)), linetype = "dashed",
    color = "black") + # Línea de la media
325 labs(title = "Densidad del régimen de tenencia de las explotaciones",
326 x = "Log(Valor)", y = "Densidad") +
327 scale_fill_manual(values = c("1982" = "cyan", "1989" = "red", "1999"
    = "blue", "2009" = "yellow", "2020" = "green")) + # Colores
    personalizados
328 scale_color_manual(values = c("1982" = "cyan", "1989" = "red", "1999"
    = "blue", "2009" = "yellow", "2020" = "green")) + # Colores de líneas
329 theme_minimal() +
330 ylim(0, 3.5) + # Límite del eje y
331 facet_wrap(~ variable, scales = "free_x") + # Facetas para cada
    variable, escalas independientes en el eje x
332 theme(legend.position = "bottom") # Leyenda en la parte
    inferior
333

```

```

334
335 #####ANÁLISIS EXPLORATORIO
336
337 # Análisis Exploratorio de Datos
338 # Resumen descriptivo
339
340
341 censoagrario2 <- censo_agrario %>% select(-D_POB)
342 censoagrario2 <- censoagrario2 %>% select(-tiempo)
343 censoagrario2 <- censoagrario2 %>% select(-CODE_FIN)
344
345
346 socio_economia <- subset(socio_combined, select = -year)
347
348
349 summary(socio_economia)
350 summary(censoagrario2)
351
352 #####Correlación
353 correlation_matrix <- cor(socio_economia, use = "complete.obs")
354 corrplot(correlation_matrix, method = "color", type = "upper")
355 correlation_matrix
356 summary(correlation_matrix)
357
358 correlation_matrix <- cor(censoagrario2, use = "complete.obs")
359 corrplot(correlation_matrix, method = "color", type = "upper")
360 correlation_matrix
361
362
363 #####VISUALIZACIÓN ESPACIAL DE LAS VARIABLES
364
365 municipios <- st_read("C:/TFM/ANALISIS/municipios_wgs84_Z30N.shp")
366 municipios$COD_MUNI <- as.character(municipios$COD_MUNI)
367
368 censo_agrario <- as.data.frame(read.table("C:/TFM/ANALISIS_EST_DESCRIPT_
/censo_agrario.txt", head =
TRUE))
369
370 #1982
371 censo_1982 <- subset(censo_agrario, tiempo == 1982)
372 censo_1982$CODE_FIN <- as.character(censo_1982$CODE_FIN)
373
374 mapa_1982 <- municipios %>%
375 left_join(censo_1982, by = c("COD_MUNI" = "CODE_FIN"))
376
377
378 # Definir las variables a graficar
379 variables <- c("ST_5ha", "ST_5_50ha", "ST_50_500ha", "ST_500ha",
380 "SAU_5ha", "SAU_5_50ha", "SAU_50_500ha", "SAU_500ha",
381 "MAYORES_55", "NO_ESTUDIOS", "D_EXPLOTT", "PROPIEDAD",
382 "ARRENDAMIENTO", "OTROS", "D_UGT")
383
384 # Crear mapas y guardarlos
385 for (var in variables) {

```

```

386     p <- ggplot(data = mapa_1982) +
387       geom_sf(aes_string(fill = var)) +
388       scale_fill_viridis_c(option = "plasma", na.value = "white") +
389       labs(title = var, fill = var) +
390       theme_minimal()
391
392       # Guardar cada mapa como un archivo temporal
393       archivo_temp <- paste0("mapa_", var, ".png")
394       ggsave(archivo_temp, p, width = 8, height = 8) # Ajustar el
395       tamaño si es necesario
396     mapas_archivos[[var]] <- archivo_temp
397   }
398
399   # Cargar los archivos de imagen y crear una lista de plots
400   mapas <- lapply(mapas_archivos, function(archivo) {
401     ggdraw() + draw_image(archivo)
402   })
403
404   # Combinar los mapas en una sola imagen
405   imagen_final <- plot_grid(plotlist = mapas, ncol = 3)
406
407   # Mostrar la imagen combinada
408   print(imagen_final)
409
410   # Guardar la imagen final con todos los mapas
411   ggsave("C:/TFM/ANALISIS_EST_DESCRIP/imagenes/censo_agrario/mapas_1982_
412   combined.png", imagen_final, width = 24, height = 24) # Ajustar tamaño
413   si es necesario
414
415   #----- Eliminar archivos temporales
416   file.remove(unlist(mapas_archivos))

```

Cálculo del STATIS-Dual de las bases de datos de los años 1982-89-99-09-20

```

1
2   setwd("C:/TFM/censo_agrario/Estadísticas/datos_definitivos")
3
4   #install.packages("readxl")
5   #install.packages("cluster")
6   #install.packages("dendextend")
7   library(ade4)
8   library(readxl)
9   library(dplyr)
10  library(sqldf)
11  library(kableExtra)
12  library(knitr)
13  library(openxlsx)
14  library(stats)
15  library(sf)
16  library(dplyr)

```

```

17     library(ggplot2)
18     library(RColorBrewer)
19     library(openxlsx)
20     library(gridExtra)
21     library(ggrepel)
22     library(grid)
23     library(cluster)
24     library(dendextend)
25     library(factoextra)
26
27     # Abrir mis datos
28
29     datos_1982 <-
30     as.data.frame(read.table("all_variables_censo82_lim_admin_FINAL.txt",
31     head = TRUE))
32     datos_1989 <-
33     as.data.frame(read.table("all_variables_censo89_lim_admin_FINAL.txt",
34     head = TRUE, sep = "\t"))
35     datos_1999 <-
36     as.data.frame(read.table("all_variables_censo99_lim_admin_FINAL.txt",
37     head = TRUE))
38     datos_2009 <-
39     as.data.frame(read.table("all_variables_censo09_lim_admin_FINAL.txt",
40     head = TRUE))
41     datos_2020 <-
42     as.data.frame(read.table("all_variables_censo20_lim_admin_FINAL.txt",
43     head = TRUE, sep = "\t"))
44
45     datos_1982 <- sqldf("select tiempo, CODE_FIN82, ST_5ha, ST_5_50ha,
46     ST_50_500ha, ST_500ha, SAU_5ha, SAU_5_50ha,
47     SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOT, PROPIEDAD,
48     ARRENDAMIENTO,
49     OTROS, D_UGT, D_POB from datos_1982")
50     datos_1982 <- datos_1982 %>% rename(CODE_FIN = CODE_FIN82)
51
52     datos_1989 <- sqldf("select tiempo, CODE_FIN89, ST_5ha, ST_5_50ha,
53     ST_50_500ha, ST_500ha, SAU_5ha, SAU_5_50ha,
54     SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOT, PROPIEDAD,
55     ARRENDAMIENTO,
56     OTROS, D_UGT, D_POB from datos_1989")
57     datos_1989 <- datos_1989 %>% rename(CODE_FIN = CODE_FIN89)
58     # Eliminar las filas donde CODE_FIN89 es igual a 0 y poner NA's como 0
59     datos_1989 <- subset(datos_1989, CODE_FIN != 0)
60     datos_1989[is.na(datos_1989)] <- 0
61     datos_1989[datos_1989 == "Inf"] <- 0
62
63     datos_1999 <- sqldf("select tiempo, CODE_FIN99, ST_5ha, ST_5_50ha,
64     ST_50_500ha, ST_500ha, SAU_5ha, SAU_5_50ha,
65     SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOT, PROPIEDAD,
66     ARRENDAMIENTO,
67     OTROS, D_UGT, D_POB from datos_1999")
68     datos_1999 <- datos_1999 %>% rename(CODE_FIN = CODE_FIN99)

```

```

55
56
57 datos_2009 <- sqldf("select tiempo, CODE_FIN09, ST_5ha, ST_5_50ha,
58 ST_50_500ha, ST_500ha, SAU_5ha, SAU_5_50ha,
59 SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOT, PROPIEDAD,
60 ARRENDAMIENTO,
61 OTROS, D_UGT, D_POB from datos_2009")
62 datos_2009 <- datos_2009 %>% rename(CODE_FIN = CODE_FIN09)
63
64 datos_2020 <- sqldf("select tiempo, CODE_FIN, ST_5ha, ST_5_50ha,
65 ST_50_500ha, ST_500ha, SAU_5ha, SAU_5_50ha,
66 SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOT, PROPIEDAD,
67 ARRENDAMIENTO,
68 OTROS, D_UGT, D_POB from datos_2020")
69
70 #Creación censo agrario (sin el 2020)
71
72 censo_agrario <- bind_rows(datos_1982, datos_1989, datos_1999,
73 datos_2009)
74 censo_agrario$tiempo <- as.factor(censo_agrario$tiempo)
75 censo_agrario$CODE_FIN <- as.factor(censo_agrario$CODE_FIN)
76 sum(is.na(censo_agrario))
77 write.table(censo_agrario, file = "censo_agrario_82_89_99_09.txt")
78
79 #Unión censo agrario + 2020
80 censo_mas_20 <-
81 as.data.frame(read.table("censo_agrario_82_89_99_09.txt", head = TRUE))
82 censo_agrario <- bind_rows(censo_mas_20, datos_2020)
83 View(censo_agrario)
84 write.table(censo_agrario, file =
85 "C:/TFM/censo_agrario/Estadsticas/datos_definitivos/censo_agrario.txt",
86 row.names = FALSE)
87
88 ##ESTADÍSTICA DESCRIPTIVA
89 resúmenes <- list(
90 "1982" = summary(datos_1982),
91 "1989" = summary(datos_1989),
92 "1999" = summary(datos_1999),
93 "2009" = summary(datos_2009),
94 "2020" = summary(datos_2020)
95 )
96
97 # Mostrar los resúmenes
98 for (año in names(resúmenes)) {
99   cat(paste("\nResumen estadstico para datos_", año, ":\n", sep =
100   ""))
101   print(resúmenes[[año]])
102 }
103
104 #-----
105 #####;Análisis Statis Dual sin D_POB (censo_agrario)
106
107

```

```

98 censo_agrario_inicial <- as.data.frame(read.table("C:/TFM/censo_agrario_
/Estadsticas/datos_definitivos/censo_agrario.txt", head =
TRUE))
99
100 censo_agrario <- censo_agrario_inicial %>%
101 select(-D_POB)
102
103
104 censo_agrario$CODE_FIN <- as.factor(censo_agrario$CODE_FIN)
105 censo_agrario$tiempo <- as.factor(censo_agrario$tiempo)
106
107 kta1_D <- ktab.within(withinpca(censo_agrario[,3:17],
censo_agrario$tiempo,
108 scaling = c("partial"), scann = FALSE))
109
110 stasis1_D <- stasis(kta1_D, scann = FALSE)
111 #mostrar otras dimensiones
112 plot(stasis1_D, xax = 1, yax = 2)
113
114 stasis1_D$RV#coeficientes de correlación vectorial RV
115 stasis1_D$cos2#coseno2
116 stasis1_D$RV.tabw#pesos para las k-tablas
117 stasis1_D$RV.eig#valores propios para la inter-estructura
118 stasis1_D$C.eig#valores propios para el compromiso
119 stasis1_D$C.Co
120
121
122 #####Análisis Stasis Dual sin año 2020
123
124 censo_agrario_inicial_2 <- as.data.frame(read.table("C:/TFM/censo_agrar_
io/Estadsticas/datos_definitivos/censo_agrario.txt", head =
TRUE))
125
126 censo_agrario_inicial_2 <- censo_agrario_inicial_2 %>%
127 select(-D_POB)
128
129 censo_agrario_2 <- subset(censo_agrario_inicial_2, tiempo != 2020)
130
131 censo_agrario_2$CODE_FIN <- as.factor(censo_agrario_2$CODE_FIN)
132 censo_agrario_2$tiempo <- as.factor(censo_agrario_2$tiempo)
133
134 kta1_D2 <- ktab.within(withinpca(censo_agrario_2[,3:17],
censo_agrario_2$tiempo,
135 scaling = c("partial"), scann = FALSE))
136
137 stasis1_D2 <- stasis(kta1_D2, scann = FALSE)
138 #mostrar otras dimensiones
139 plot(stasis1_D2, xax = 1, yax = 2)
140
141 stasis1_D2$RV#coeficientes de correlación vectorial RV
142 stasis1_D2$cos2#coseno2
143 stasis1_D2$RV.tabw#pesos para las k-tablas
144 stasis1_D2$RV.eig#valores propios para la inter-estructura
145 stasis1_D2$C.eig#valores propios para el compromiso

```

```

146   stasis1_D2$C.li
147
148
149   #####Cluster con C.Co (sin 2020)
150
151   componentes_sin20 <- as.data.frame(stasis1_D2$C.Co)
152   #write.xlsx(componentes_sin20, file =
153     "C:/TFM/censo_agrario/Estadsticas/datos_definitivos/componentes2.xlsx")
154
155   conteo_tiempo <- table(censo_agrario_2$tiempo)
156   conteo_tiempo
157
158   if (nrow(componentes_sin20) == nrow(censo_agrario_2)) {
159     # Copiar las columnas necesarias
160     componentes_sin20$tiempo <- censo_agrario_2$tiempo
161     componentes_sin20$CODE_FIN <- censo_agrario_2$CODE_FIN
162     componentes_sin20$ST_5ha <- censo_agrario_2$ST_5ha
163     componentes_sin20$ST_5_50ha <- censo_agrario_2$ST_5_50ha
164     componentes_sin20$ST_50_500ha <- censo_agrario_2$ST_50_500ha
165     componentes_sin20$ST_500ha <- censo_agrario_2$ST_500ha
166     componentes_sin20$SAU_5ha <- censo_agrario_2$SAU_5ha
167     componentes_sin20$SAU_5_50ha <- censo_agrario_2$SAU_5_50ha
168     componentes_sin20$SAU_50_500ha <- censo_agrario_2$SAU_50_500ha
169     componentes_sin20$SAU_500ha <- censo_agrario_2$SAU_500ha
170     componentes_sin20$MAYORES_55 <- censo_agrario_2$MAYORES_55
171     componentes_sin20$NO_ESTUDIOS <- censo_agrario_2$NO_ESTUDIOS
172     componentes_sin20$D_EXPLOTA <- censo_agrario_2$D_EXPLOTA
173     componentes_sin20$PROPIEDAD <- censo_agrario_2$PROPIEDAD
174     componentes_sin20$ARRENDAMIENTO <- censo_agrario_2$ARRENDAMIENTO
175     componentes_sin20$OTROS <- censo_agrario_2$OTROS
176     componentes_sin20$D_UGT <- censo_agrario_2$D_UGT
177   } else {
178     stop("El nmero de filas en 'componentes_sin20' y
179         'censo_agrario_2' no coincide.")
180   }
181
182   # Filtrar datos por año y crear nuevos data frames
183   componentes_sin20_82 <- componentes_sin20[componentes_sin20$tiempo ==
184     1982,
185     c("C1", "C2", "C3", "CODE_FIN",
186       "ST_5ha", "ST_5_50ha", "ST_50_500ha", "ST_500ha",
187       "SAU_5ha", "SAU_5_50ha", "SAU_50_500ha", "SAU_500ha",
188       "MAYORES_55", "NO_ESTUDIOS", "D_EXPLOTA", "PROPIEDAD",
189       "ARRENDAMIENTO", "OTROS", "D_UGT")]
190
191   componentes_sin20_89 <- componentes_sin20[componentes_sin20$tiempo ==
192     1989, c("C1", "C2", "C3", "CODE_FIN")]
193   componentes_sin20_99 <- componentes_sin20[componentes_sin20$tiempo ==
194     1999, c("C1", "C2", "C3", "CODE_FIN")]
195   componentes_sin20_09 <- componentes_sin20[componentes_sin20$tiempo ==
196     2009, c("C1", "C2", "C3", "CODE_FIN")]
197
198   if (!require("extrafont")) install.packages("extrafont")

```

```

194 library(extrafont)
195 font_import() # Importar todas las fuentes instaladas en el sistema
196 loadfonts()
197
198 #####HACER CLUSTERING CON 2 AGRUPACIONES
199
200 # Definir una funcin para realizar el k-means clustering y aadir la
201 # columna 'cluster'
202 perform_kmeans_clustering <- function(data, centers = 2) {
203   # Realizar el clustering k-means
204   kmeans_result <- kmeans(data[, c("C1", "C2")], centers =
205     centers)
206   # Añadir la columna de clusters al data frame
207   data$cluster <- as.factor(kmeans_result$cluster)
208   return(data)
209 }
210
211 # Aplicar la función a cada data frame
212 componentes_sin20_82 <- perform_kmeans_clustering(componentes_sin20_82,
213   centers = 2)
214 componentes_sin20_89 <- perform_kmeans_clustering(componentes_sin20_89,
215   centers = 2)
216 componentes_sin20_99 <- perform_kmeans_clustering(componentes_sin20_99,
217   centers = 2)
218 componentes_sin20_09 <- perform_kmeans_clustering(componentes_sin20_09,
219   centers = 2)
220
221 # Seleccionar las columnas de interés y almacenar en nuevas variables
222 CLUSTER_82 <- componentes_sin20_82[, c("CODE_FIN", "cluster")]
223 CLUSTER_89 <- componentes_sin20_89[, c("CODE_FIN", "cluster")]
224 CLUSTER_99 <- componentes_sin20_99[, c("CODE_FIN", "cluster")]
225 CLUSTER_09 <- componentes_sin20_09[, c("CODE_FIN", "cluster")]
226
227
228 municipios <- st_read("C:/TFM/censo_agrario/Estadsticas/datos_definitiv_
229 os/municipios_wgs84_Z30N.shp")
230 municipios$COD_MUNI <- as.character(municipios$COD_MUNI)
231
232 # Función para crear los mapas después del join
233 create_map <- function(cluster_data, year, colors) {
234   municipios_cluster <- municipios %>%
235     left_join(cluster_data, by = c("COD_MUNI" = "CODE_FIN"))
236
237   ggplot(data = municipios_cluster) +
238     geom_sf(aes(fill = cluster), color = "transparent", size = 0.1)
239     +
240     scale_fill_manual(values = colors, name = "Cluster") +
241     theme_minimal() +
242     labs(title = paste("Año", year),
243       fill = "Cluster") +
244     theme(
245       plot.title = element_text(hjust = 0.5, size = 15, family =
246         "Garamond"), # Fuente Garamond para el ttulo

```

```

238         text = element_text(family = "Garamond") # Fuente Garamond para
239         todo el texto en el grfico
240     )
241 }
242
243 # Definir colores personalizados para cada año
244 colors_1982 <- c("1" = "#8DD822", "2" = "#2260FB")
245 colors_1989 <- c("1" = "#2260FB", "2" = "#8DD822")
246 colors_1999 <- c("1" = "#2260FB", "2" = "#8DD822")
247 colors_2009 <- c("1" = "#8DD822", "2" = "#2260FB")
248
249 # Crear y mostrar mapas para cada año con los colores personalizados
250 map_1982 <- create_map(CLUSTER_82, 1982, colors_1982)
251 map_1989 <- create_map(CLUSTER_89, 1989, colors_1989)
252 map_1999 <- create_map(CLUSTER_99, 1999, colors_1999)
253 map_2009 <- create_map(CLUSTER_09, 2009, colors_2009)
254
255 # Mostrar todos los mapas en una sola visualización
256 grid.arrange(map_1982, map_1989, map_1999, map_2009, ncol = 2)
257
258
259 #####HACER CLUSTERING CON 3 AGRUPACIONES
260
261 # Definir una funcin para realizar el k-means clustering y aadir la
262 columna 'cluster'
263 perform_kmeans_clustering <- function(data, centers = 3) {
264     # Realizar el clustering k-means
265     kmeans_result <- kmeans(data[, c("C1", "C2")], centers =
266         centers)
267     # Añadir la columna de clusters al data frame
268     data$cluster <- as.factor(kmeans_result$cluster)
269     return(data)
270 }
271
272 # Aplicar la función a cada data frame
273 componentes_sin20_82 <- perform_kmeans_clustering(componentes_sin20_82,
274     centers = 3)
275 componentes_sin20_89 <- perform_kmeans_clustering(componentes_sin20_89,
276     centers = 3)
277 componentes_sin20_99 <- perform_kmeans_clustering(componentes_sin20_99,
278     centers = 3)
279 componentes_sin20_09 <- perform_kmeans_clustering(componentes_sin20_09,
280     centers = 3)
281
282 # Seleccionar las columnas de interés y almacenar en nuevas variables
283 CLUSTER_82 <- componentes_sin20_82[, c("CODE_FIN", "cluster")]
284 CLUSTER_89 <- componentes_sin20_89[, c("CODE_FIN", "cluster")]
285 CLUSTER_99 <- componentes_sin20_99[, c("CODE_FIN", "cluster")]
286 CLUSTER_09 <- componentes_sin20_09[, c("CODE_FIN", "cluster")]
287
288
289 ## MEDIAS POR CLUSTER. HECHO SOLO EN EL AO 82 PARA SABER COMO SON LOS
290 CLUSTERS: QU CARACTERSTICAS TIENEN, ETC.

```

```

284 medias_por_cluster <- aggregate(. ~ cluster, data =
componentes_sin20_82, FUN = mean)
285
286 medias_por_cluster
287 #####
288
289 municipios <- st_read("C:/TFM/censo_agrario/Estadstic/datos_definitiv_
os/municipios_wgs84_Z30N.shp")
290 municipios$COD_MUNI <- as.character(municipios$COD_MUNI)
291
292 # Función para crear los mapas después del join
293 create_map <- function(cluster_data, year, colors) {
294     municipios_cluster <- municipios %>%
295     left_join(cluster_data, by = c("COD_MUNI" = "CODE_FIN"))
296
297     ggplot(data = municipios_cluster) +
298     geom_sf(aes(fill = cluster), color = "transparent", size = 0.1)
+
299     scale_fill_manual(values = colors, name = "Cluster") +
300     theme_minimal() +
301     labs(title = paste("Año", year),
302     fill = "Cluster") +
303     theme(
304     plot.title = element_text(hjust = 0.5, size = 15, family =
"Garamond"), # Fuente Garamond para el ttulo
305     text = element_text(family = "Garamond") # Fuente Garamond para
todo el texto en el grfico
306     )
307 }
308
309 # Definir colores personalizados para cada año
310 colors_1982 <- c("1" = "#8DD822", "2" = "#2260FB", "3" = "pink")
311 colors_1989 <- c("1" = "pink", "2" = "#2260FB", "3" = "#8dd822")
312 colors_1999 <- c("1" = "#2260FB", "2" = "#8DD822", "3" = "pink")
313 colors_2009 <- c("1" = "#2260FB", "2" = "#8DD822", "3" = "pink")
314
315 # Crear y mostrar mapas para cada año con los colores personalizados
316 map_1982 <- create_map(CLUSTER_82, 1982, colors_1982)
317 map_1989 <- create_map(CLUSTER_89, 1989, colors_1989)
318 map_1999 <- create_map(CLUSTER_99, 1999, colors_1999)
319 map_2009 <- create_map(CLUSTER_09, 2009, colors_2009)
320
321
322 # Mostrar todos los mapas en una sola visualización
323 grid.arrange(map_1982, map_1989, map_1999, map_2009, ncol = 2)
324

```

Cálculo de la Co-inercia de las bases de datos del 2009 y 2020

```
1
2 socioeconomia_2009 <- sqldf("select PGENERAL, PAGRARIO, PMAR, PHOGAR,
3 PAUTONOMOS, PCARBON,
4 PTOTAL, Pdeuda_miles_euros, TCV, Ptotal_paro, Pparo_hombre_25,
5 Pparo_hombre_25_45, Pparo_hombre_45, Pparo_mujer_25, Pparo_mujer_25_45,
6 Pparo_mujer_45,
7 Pparo_agricultura, Pparo_industria, Pparo_construccion, Pparo_servicios,
8 Pparo_sin_empleo_anterior
9 from bbdd_09")
10
11 socioeconomia_2020 <- sqldf("select PGENERAL, PAGRARIO, PMAR, PHOGAR,
12 PAUTONOMOS, PCARBON,
13 PTOTAL, Pdeuda_miles_euros, TCV, Ptotal_paro, Pparo_hombre_25,
14 Pparo_hombre_25_45, Pparo_hombre_45, Pparo_mujer_25, Pparo_mujer_25_45,
15 Pparo_mujer_45,
16 Pparo_agricultura, Pparo_industria, Pparo_construccion, Pparo_servicios,
17 Pparo_sin_empleo_anterior
18 from bbdd_20")
19
20 censo_agrario_2009 <- sqldf("select ST_5ha, ST_5_50ha, ST_50_500ha,
21 ST_500ha, SAU_5ha, SAU_5_50ha,
22 SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOTA,
23 PROPIEDAD, ARRENDAMIENTO, OTROS, D_UGT from bbdd_09")
24 censo_agrario_2020 <- sqldf("select ST_5ha, ST_5_50ha, ST_50_500ha,
25 ST_500ha, SAU_5ha, SAU_5_50ha,
26 SAU_50_500ha, SAU_500ha, MAYORES_55, NO_ESTUDIOS, D_EXPLOTA,
27 PROPIEDAD, ARRENDAMIENTO, OTROS, D_UGT from bbdd_20")
28
29 # Crear objetos de tipo 'dudi' para los datos de socioeconomia y censo
30 agrario
31 dudi_socio_2009 <- dudi.pca(socioeconomia_2009, scannf = FALSE, nf = 2)
32 dudi_censo_ag_2009 <- dudi.pca(censo_agrario_2009, scannf = FALSE, nf =
33 3)
34
35 dudi_socio_2020 <- dudi.pca(socioeconomia_2020, scannf = FALSE, nf = 2)
36 dudi_censo_ag_2020 <- dudi.pca(censo_agrario_2020, scannf = FALSE, nf =
37 3)
38
39 # Realizar el analisis de coinercia entre los datos de socioeconomia y
40 censo agrario
41 coinertia_2009 <- coinertia(dudi_socio_2009, dudi_censo_ag_2009, scannf
42 = FALSE)
43 coinertia_2020 <- coinertia(dudi_socio_2020, dudi_censo_ag_2020, scannf
44 = FALSE)
45
46 # Resumen de los resultados
47 summary(coinertia_2009)
48 summary(coinertia_2020)
```

```

38     coinertia_2009$RV
39     coinertia_2009$nf
40     coinertia_2009$eig
41     coinertia_2009$li
42     coinertia_2009$l1
43     coinertia_2009$c1
44     coinertia_2009$co
45     #coinertia_2020£LX
46     #coinertia_2020£mX
47     #coinertia_2020£LY
48     #coinertia_2020£mY
49     coinertia_2009$aX
50     coinertia_2009$aY
51
52     coinertia_2020$RV
53     coinertia_2020$nf
54     coinertia_2020$eig
55     coinertia_2020$li
56     coinertia_2020$l1
57     coinertia_2020$c1
58     coinertia_2020$co
59     #coinertia_2020£LX
60     #coinertia_2020£mX
61     #coinertia_2020£LY
62     #coinertia_2020£mY
63     coinertia_2020$aX
64     coinertia_2020$aY
65
66
67     # Plot del análisis de coinerencia para 2009
68     plot(coinertia_2009, main = "Análisis de Coinercia para 2009")
69
70     options(expressions = 100000) # Aumenta el lmite de expresiones
71     permitidas
72
73     # Plot del análisis de coinerencia para 2020
74     plot(coinertia_2020, main = "Análisis de Coinercia para 2020")
75
76     setwd("C:/TFM/ANALISIS")
77
78     # Guardar el gráfico en un archivo PDF
79     pdf("coinertia_2009_plot.pdf")
80     plot(coinertia_2009, main = "Análisis de Coinercia para 2009")
81     dev.off()
82
83     # Guardar el gráfico en un archivo PDF
84     pdf("coinertia_2020_plot2.pdf")
85     plot(coinertia_2020, main = "Análisis de Coinercia para 2020")
86     dev.off()
87
88     #????????????????????????????????????????????????? lX (dudi.x) o ly (dudi.y)
89

```

```

90   # El primer dudi (x) en ambos aos es socioeconomia y el segundo dudi (y)
91   es el censo agrario
92   s_09 <- coinertia_2009$1X
93   ca_09 <- coinertia_2009$1Y
94
95   s_20 <- coinertia_2020$1X
96   ca_20 <- coinertia_2020$1Y
97
98   municipios <- st_read("C:/TFM/ANALISIS/municipios_wgs84_Z30N.shp")
99   municipios$COD_MUNI <- as.character(municipios$COD_MUNI)
100
101   cod_muni_09 <- bdd_09$COD_MUNI
102   cod_muni_20 <- bdd_20$COD_MUNI
103
104   s_09 <- cbind(COD_MUNI = cod_muni_09, s_09)
105   ca_09 <- cbind(COD_MUNI = cod_muni_09, ca_09)
106   s_20 <- cbind(COD_MUNI = cod_muni_20, s_20)
107   ca_20 <- cbind(COD_MUNI = cod_muni_20, ca_20)
108
109   kmeans_s_09 <- kmeans(s_09[, -1], centers = 3)
110   kmeans_ca_09 <- kmeans(ca_09[, -1], centers = 3)
111   kmeans_s_20 <- kmeans(s_20[, -1], centers = 3)
112   kmeans_ca_20 <- kmeans(ca_20[, -1], centers = 3)
113
114   s_09$cluster <- kmeans_s_09$cluster
115   ca_09$cluster <- kmeans_ca_09$cluster
116   s_20$cluster <- kmeans_s_20$cluster
117   ca_20$cluster <- kmeans_ca_20$cluster
118
119   create_map <- function(map_data, title, colors) {
120     ggplot(data = map_data) +
121       geom_sf(aes(fill = factor(cluster)), color = NA) +
122       scale_fill_manual(values = colors) +
123       labs(title = title, fill = "Cluster") +
124       theme_minimal() +
125       theme(
126         plot.title = element_text(hjust = 0.5, size = 15, family =
127           "Garamond"), # Fuente Garamond para el ttulo
128         text = element_text(family = "Garamond") # Fuente Garamond para
129           todo el texto en el grfico
130       )
131   }
132
133   map_s_09 <- merge(municipios, s_09, by.x = "COD_MUNI", by.y =
134     "COD_MUNI")
135   map_ca_09 <- merge(municipios, ca_09, by.x = "COD_MUNI", by.y =
136     "COD_MUNI")
137   map_s_20 <- merge(municipios, s_20, by.x = "COD_MUNI", by.y =
138     "COD_MUNI")
139   map_ca_20 <- merge(municipios, ca_20, by.x = "COD_MUNI", by.y =
140     "COD_MUNI")
141
142   colors_s09 <- c("1" = "#25C9B1", "2" = "yellow", "3" = "#EE36D5")

```

```

137     colors_s20 <- c("1" = "#25C9B1", "2" = "#EE36D5", "3" = "yellow")
138
139     colors_ca09 <- c("1" = "#B7FF66", "2" = "blue", "3" = "red")
140     colors_ca20 <- c("1" = "#B7FF66", "2" = "blue", "3" = "red")
141
142
143
144     map_s_09_plot <- create_map(map_s_09, "Socio-economía en 2009",
145                               colors_s09)
146     map_ca_09_plot <- create_map(map_ca_09, "Censo agrario en 2009",
147                               colors_ca09)
148     map_s_20_plot <- create_map(map_s_20, "Socio-economía en 2020",
149                               colors_s20)
150     map_ca_20_plot <- create_map(map_ca_20, "Censo agrario en 2020",
151                               colors_ca20)
152
153     grid.arrange(map_s_09_plot, map_ca_09_plot, map_s_20_plot,
154                 map_ca_20_plot, ncol = 2)

```

Cálculo del número óptimo de k-clusters

```

1
2     #####Elegir nº de clústeres
3
4     #1.Calcular la matriz de distancias
5     #2.Realizar el clustering jerárquico usando el método de Ward
6     #3.Visualizar el dendrograma
7     #4.Decidir el número de clusters basado en el dendrograma
8
9
10    # Función para realizar el análisis y crear el dendrograma
11    analyze_clusters <- function(data, year, file_name) {
12        # Calcular la matriz de distancias
13        dist_matrix <- dist(data[, c("C1", "C2", "C3")], method =
14            "euclidean")
15
16        # Realizar el clustering jerárquico usando el método de Ward
17        hc <- hclust(dist_matrix, method = "ward.D2")
18
19        # Visualizar el dendrograma
20        dend <- as.dendrogram(hc)
21        dend_colored <- color_branches(dend, k = 3) # Ajustar el valor
22        de k basado en la inspección visual
23
24        # Guardar el dendrograma como archivo PNG
25        png(file_name)
26        plot(dend_colored, main = paste("Dendrograma - Año", year))
27        dev.off()
28
29        return(hc)

```

```

28     }
29
30     # Aplicar la función para cada año y guardar los plots como PNG
31     hc_82 <- analyze_clusters(componentes_sin20_82, 1982, "C:/TFM/censo_agr_
32     ario/Estadsticas/datos_definitivos/dendrograma_1982.png")
33     hc_89 <- analyze_clusters(componentes_sin20_89, 1989, "C:/TFM/censo_agr_
34     ario/Estadsticas/datos_definitivos/dendrograma_1989.png")
35     hc_99 <- analyze_clusters(componentes_sin20_99, 1999, "C:/TFM/censo_agr_
36     ario/Estadsticas/datos_definitivos/dendrograma_1999.png")
37     hc_09 <- analyze_clusters(componentes_sin20_09, 2009, "C:/TFM/censo_agr_
38     ario/Estadsticas/datos_definitivos/dendrograma_2009.png")
39
40     ##### OTRO CDIGO PARA EL Mtodo de Ward -
41     DENDROGRAMAS #####3
42
43     # Función para realizar el análisis y crear el dendrograma
44     analyze_and_save_dendrogram <- function(data, year, file_name) {
45         # Calcular la matriz de distancias
46         dist_matrix <- dist(data[, c("C1", "C2", "C3")], method =
47         "euclidean")
48
49         # Realizar el clustering jerárquico usando el método de Ward
50         hc <- hclust(dist_matrix, method = "ward.D2")
51
52         # Visualizar el dendrograma
53         png(file_name)
54         plot(hc, main = paste("Dendrograma - Ao", year), xlab = "",
55         ylab = "", sub = "", cex.main = 1.5)
56         dev.off()
57     }
58
59     #####
60
61     # Aplicar la función para cada año y guardar los plots como PNG
62     analyze_and_save_dendrogram(componentes_sin20_82, 1982, "C:/TFM/censo_a_
63     grario/Estadsticas/datos_definitivos/dendrograma_1982.png")
64     analyze_and_save_dendrogram(componentes_sin20_89, 1989, "C:/TFM/censo_a_
65     grario/Estadsticas/datos_definitivos/dendrograma_1989.png")
66     analyze_and_save_dendrogram(componentes_sin20_99, 1999, "C:/TFM/censo_a_
67     grario/Estadsticas/datos_definitivos/dendrograma_1999.png")
68     analyze_and_save_dendrogram(componentes_sin20_09, 2009, "C:/TFM/censo_a_
69     grario/Estadsticas/datos_definitivos/dendrograma_2009.png")
70
71     #Para determinar el nmero ptimo de clusters Podemos utilizar varias
72     tcnicas, como el mtodo del codo (Elbow Method), el criterio de la
73     silueta (Silhouette Method), o el criterio de la varianza explicada.
74     Aqu te muestro cmo hacerlo utilizando el mtodo del codo y el criterio de
75     la silueta:
76     #-Método del Codo (Elbow Method):

```

```

64     #Este mtodo se basa en la suma de las distancias cuadradas dentro del
        cluster (within-cluster sum of squares, WSS). Elegimos el nmero de
        clusters en el punto donde la disminucin de WSS se hace menos
        pronunciada.
65     #-Criterio de la Silueta (Silhouette Method):
66     #Este mtodo evala la cohesin dentro de los clusters y la separacin entre
        los clusters. La silueta media se calcula para cada posible nmero de
        clusters, y elegimos el nmero de clusters con el valor ms alto de la
        silueta media.
67
68
69     library(factoextra)
70
71     # Funcin para visualizar y guardar los grficos del mtodo del codo y del
        criterio de la silueta
72     determine_and_save_optimal_clusters <- function(data, year, elbow_file,
        silhouette_file) {
73         # Calcular la matriz de distancias
74         dist_matrix <- dist(data[, c("C1", "C2", "C3")], method =
        "euclidean")
75
76         # Mtodo del Codo
77         plot1 <- fviz_nbclust(data[, c("C1", "C2", "C3")], kmeans,
        method = "wss") +
78         labs(title = paste("Nmero ptimo de clusters - Mtodo del Codo -
        Ao", year))
79         print(plot1) # Verifica que el grfico se genera
80
81         # Guardar el grfico del mtodo del codo
82         png(filename = elbow_file)
83         print(plot1) # Asegurate de que se imprime correctamente en el
        archivo
84         dev.off()
85
86         # Mtodo de la Silueta
87         plot2 <- fviz_nbclust(data[, c("C1", "C2", "C3")], kmeans,
        method = "silhouette") +
88         labs(title = paste("Nmero ptimo de clusters - Mtodo de la
        Silueta - Ao", year))
89         print(plot2) # Verifica que el grfico se genera
90
91         # Guardar el grfico del mtodo de la silueta
92         png(filename = silhouette_file)
93         print(plot2) # Asegurate de que se imprime correctamente en el
        archivo
94         dev.off()
95     }
96
97     # Aplicar la funcin para cada ao y guardar los grficos como PNG
98     determine_and_save_optimal_clusters(componentes_sin20_82, 1982,
        "C:/TFM/censo_agrario/Estadsticas/datos_definitivos/elbow_1982.png", "C_
        :/TFM/censo_agrario/Estadsticas/datos_definitivos/silhouette_1982.png")

```

```

99     determine_and_save_optimal_clusters(componentes_sin20_89, 1989,
    "C:/TFM/censo_agrario/Estadsticas/datos_definitivos/elbow_1989.png", "C_
100    /TFM/censo_agrario/Estadsticas/datos_definitivos/silhouette_1989.png")
    determine_and_save_optimal_clusters(componentes_sin20_99, 1999,
    "C:/TFM/censo_agrario/Estadsticas/datos_definitivos/elbow_1999.png", "C_
101    /TFM/censo_agrario/Estadsticas/datos_definitivos/silhouette_1999.png")
    determine_and_save_optimal_clusters(componentes_sin20_09, 2009,
    "C:/TFM/censo_agrario/Estadsticas/datos_definitivos/elbow_2009.png", "C_
102    /TFM/censo_agrario/Estadsticas/datos_definitivos/silhouette_2009.png")
103

```

Cálculo del PLSR

```

1     ##### Abrir .txt con incendios
2
3     # Instala el paquete readr si no está instalado
4     if (!require(readr)) install.packages("readr")
5     library(readr)
6
7     incendios <- as.data.frame(read_delim("C:/TFM/ANALISIS_EST_DESCRIPT/inc_
    endios_2006_2015.txt", delim = "\t", locale = locale(encoding =
    "UTF-16")))
8     incendios
9     incendios$COD_MUNI <- as.character(incendios$COD_MUNI)
10
11    datos_2009 <- as.data.frame(read.table("C:/TFM/ANALISIS_EST_DESCRIPT/bb_
    dd_09_def.txt", head =
12    TRUE))
13    datos_2009$COD_MUNI <- as.character(datos_2009$COD_MUNI)
14
15    base_completa <- merge(incendios, datos_2009, by = "COD_MUNI")
16
17    # Seleccionar variables de interés para la regresión
18    response_variable <- base_completa$TOTAL_INCENDIOS
19    predictor_variables <- base_completa[, c("SAU_5ha", "SAU_5_50ha",
    "SAU_50_500ha", "SAU_500ha",
20    "ST_5ha", "ST_5_50ha", "ST_50_500ha", "ST_500ha",
    "MAYORES_55", "NO_ESTUDIOS", "D_EXPLOT", "PROPIEDAD",
21    "ARRENDAMIENTO", "OTROS", "D_UGT", "D_POB", "UGT",
    "NEXPLOT", "TCV", "PGENERAL", "PAGRARIO", "PMAR",
22    "PHOGAR", "PAUTONOMOS", "PCARBON", "PTOTAL",
    "Pdeuda_miles_euros", "Ptotal_paro", "Pparo_hombre_25",
23    "Pparo_hombre_25_45", "Pparo_hombre_45", "Pparo_mujer_25",
    "Pparo_mujer_25_45", "Pparo_mujer_45", "Pparo_agricultura",
24    "Pparo_industria", "Pparo_construccion", "Pparo_servicios",
    "Pparo_sin_empleo_anterior")]
25
26
27
28
29
30    modelo_plsr <- plsr(response_variable ~ ., data = predictor_variables,
    scale = TRUE, validation = "CV")
31

```

```

32     summary(modelo_plsr)
33     explained_variance <- explvar(modelo_plsr)
34     explained_variance # Muestra la varianza explicada por cada componente
35
36     plot(RMSEP(modelo_plsr), legendpos = "topright")
37     coefplot(modelo_plsr, ncomp = 2) # Grfico de los coeficientes de
    regresin para los dos primeros componentes
38
39     validationplot(modelo_plsr, val.type = "MSEP") # Grfico del error
    cuadratico medio de prediccin para cada nmero de componentes
40
41     modelo_final <- pls(rresponse_variable ~ ., data = predictor_variables,
    scale = TRUE, ncomp = 2)
42     summary(modelo_final)
43
44     loadings(modelo_plsr)
45     scores(modelo_plsr)
46     coefficients(modelo_plsr)
47
48     validationplot(modelo_plsr, val.type = "MSEP")
49     plot(modelo_plsr, ncomp = 2, asp = 1, line = TRUE)
50
51     loadings(modelo_plsr)
52     scores(modelo_plsr)
53     coefficients(modelo_plsr)
54
55
56     #-----GRAFICAR RESULTADOS
57
58     # Extraer los coeficientes del modelo
59     # Obtener los coeficientes del modelo PLSR
60     coefficients <- as.vector(coef(modelo_plsr)) # Asegurarse de que sea un
    vector
61
62     # Ver las primeras filas de las cargas (por si acaso es necesario)
63     # head(loadings)
64
65     # Ver los coeficientes del modelo
66     print(coefficients)
67
68     # Obtener los nombres de las variables predictoras
69     variable_names <- rownames(coef(modelo_plsr))
70
71     # Ordenar los coeficientes junto con los nombres de las variables
72     sorted_indices <- order(coefficients, decreasing = TRUE) # Ordenar de
    mayor a menor
73     sorted_coefficients <- coefficients[sorted_indices]
74     sorted_variable_names <- variable_names[sorted_indices]
75
76     # Crear un grfico de barras para los coeficientes
77     # Usar diferentes colores para coeficientes positivos y negativos
78     bar_colors <- ifelse(sorted_coefficients > 0, "lightblue", "lightcoral")
79
80     # Crear el grfico

```

```

81     barplot(
82     sorted_coefficients,
83     names.arg = sorted_variable_names,
84     las = 2,                # Rotar las etiquetas del eje X
85     cex.names = 0.7,       # Tamaño del texto de las etiquetas
86     main = "Importancia de las variables para predecir al n total de
      incendios",
87     ylab = "Coeficientes",
88     col = bar_colors,      # Colores dependiendo del signo del
      coeficiente
89     border = "blue"
90     )
91
92
93     # Ver en qué munis hay incendios
94
95     municipios <- st_read("C:/TFM/ANALISIS/municipios_wgs84_Z30N.shp")
96     municipios$COD_MUNI <- as.character(municipios$COD_MUNI)
97     incendios
98     incendios_map <- merge(municipios, incendios, by = "COD_MUNI", all.x =
      TRUE)
99
100    # Filtrar las geometrías vacías si las hay
101    incendios_map <- incendios_map %>% filter(!st_is_empty(geometry))
102
103    # Crear el mapa
104
105    total_incendios_map <- ggplot(data = incendios_map) +
106    geom_sf(aes(fill = TOTAL_INCENDIOS), color = NA) +
107    scale_fill_viridis_c(option = "turbo", na.value = "white", name = "Total
      incendios") +
108    labs(title = "Mapa de incendios totales por municipio",
109    subtitle = "Datos de incendios en municipios (conatos + incendios) de
      Espaa",
110    caption = "Fuente: MITECO. rea de Defensa contra Incendios Forestales")
111    +
112    theme_minimal() +
113    theme(legend.position = "bottom")
114
115    #MINISTERIO PARA LA TRANSICIN ECOLGICA Y EL RETO DEMOGRFICO. Direccin
      General de Biodiversidad, Bosques y Desertificacin. Subdireccin General
      de Poltica Forestal y Lucha contra la Desertificacin. rea de Defensa
      contra Incendios Forestales.
116
117    conatos_map <- ggplot(data = incendios_map) +
118    geom_sf(aes(fill = conatos), color = NA) +
119    scale_fill_viridis_c(option = "turbo", na.value = "white", name = "N de
      conatos: incendios frustrados") +
120    labs(title = "Mapa de nº de conatos por municipio") +
121    theme_minimal() +
122    theme(legend.position = "bottom")
123
124    numero_incendios_map <- ggplot(data = incendios_map) +
      geom_sf(aes(fill = incendios), color = NA) +

```

```
125     scale_fill_viridis_c(option = "turbo", na.value = "white", name = "N de
126     incendios") +
127     labs(title = "Mapa de incendios por municipio") +
128     theme_minimal() +
129     theme(legend.position = "bottom")
130
131     incendios_totales_map <- ggplot(data = incendios_map) +
132     geom_sf(aes(fill = ARBOLADO, color = NA) +
133     scale_fill_viridis_c(option = "viridis", na.value = "white", name =
134     "Arbolado") +
135     labs(title = "Mapa de incendios totales por municipio",
136     subtitle = "Arbolado en los municipios en la península Ibérica",
137     caption = "Fuente: MITECO. rea de Defensa contra Incendios
138     Forestales.") +
139     theme_minimal() +
140     theme(legend.position = "bottom")
141
142     grid.arrange(conatos_map, numero_incendios_map, total_incendios_map,
143     ncol = 3)
```
