

# Proyecto AED: Análisis de la Calidad de Vida en España

Daniela Meriño, Víctor Mateu, Antonio Nova

2024-12-31

## Introducción

La calidad de vida es un concepto multidimensional que abarca aspectos económicos, sociales, ambientales y subjetivos, reflejando el bienestar general de las personas en una sociedad. En este trabajo, analizaremos la calidad de vida en España, desglosada en diversas dimensiones, a lo largo de varias comunidades autónomas y un periodo de años.

El análisis se centra en múltiples dimensiones de la calidad de vida, como las condiciones materiales, la salud, la educación, el trabajo, las relaciones sociales, la seguridad, el entorno ambiental y la gobernanza, entre otras. Además, el enfoque temporal nos permitirá evaluar tendencias, identificar patrones y observar disparidades entre las diferentes comunidades autónomas.

Los datos utilizados para este estudio provienen de un conjunto estructurado que organiza información clave de calidad de vida por año, comunidad autónoma y dimensión. Este formato permitirá realizar comparaciones entre las regiones y temporales, utilizando herramientas estadísticas y visualizaciones para descubrir las relaciones y diferencias significativas entre las regiones de España.

## Importación y Ordenación de los datos

En este apartado, se procede a la importación y preparación de los datos proporcionados por el Instituto Nacional de Estadística (INE) de España. Los datos corresponden al **Indicador Multidimensional de Calidad de Vida (IMCV)**. Este indicador tiene como objetivo ofrecer una visión integral del bienestar en las distintas comunidades autónomas mediante la evaluación de múltiples dimensiones, como salud, educación, trabajo, entre otras.

Nuestro conjunto de datos se organiza en nueve dimensiones. Estas abarcan aspectos clave como condiciones materiales de vida (renta, pobreza y seguridad económica), trabajo (empleo y calidad laboral), salud (esperanza de vida y determinantes como ejercicio y tabaquismo), educación (formación y abandono escolar), ocio y relaciones sociales (satisfacción y apoyo social), seguridad física (criminalidad y percepción de seguridad), gobernanza (confianza en instituciones y participación política), entorno (contaminación y acceso a zonas verdes) y experiencia de vida (satisfacción y propósito vital). Juntas, estas dimensiones ofrecen una visión del bienestar en las diferentes comunidades autónomas de España.

La naturaleza experimental de estos datos implica que están en proceso de consolidación metodológica y pueden incluir innovaciones en las fuentes de información. En este sentido, el primer paso es garantizar que los datos se encuentren correctamente estructurados y listos para su análisis.

En primer lugar, destacar que de este conjunto de datos utilizada tiene un número muy elevado de variables por lo que nosotros nos centraremos en los resultados globales de cada dimensión por comunidad autónoma por año.

Aunque, para entender un poco mejor de cómo se han obtenido los datos del conjunto que vamos a analizar, destacar que el cálculo del **Indicador Multidimensional de Calidad de Vida (IMCV)** se basa en la selección de indicadores representativos de nueve dimensiones clave, como condiciones materiales de vida,

salud, educación y trabajo, obtenidos de fuentes estadísticas como la Encuesta de Condiciones de Vida. Los valores de cada indicador se normalizan para permitir comparaciones homogéneas mediante el método Mazziotta-Pareto Ajustado (AMPI), que reescala los datos en un rango común y penaliza la variabilidad excesiva. Posteriormente, se agregan los valores normalizados de cada dimensión en un índice representativo, y finalmente, los índices de todas las dimensiones se combinan en un único valor global mediante una media ponderada. El resultado permite analizar la evolución temporal y las diferencias territoriales en la calidad de vida en España.

Una vez tenemos una idea básica sobre el dataset que vamos a analizar, nos centraremos en la importación y ordenación de este.

El fichero descargado a través del INE se obtuvo en formato .xlsx, por lo que la importación se tuvo que realizar a través de la función `read_excel` de la librería `readxl`. Se creó un código en R de manera que cumpla las condiciones de un conjunto de datos considerado *tidy*. Esto requiere que cada variable esté representada en una columna, cada observación de una variable está en una fila distinta, la primera fila incluye los nombres de las variables y los nombres deben ser representativos de las variables que almacenan.

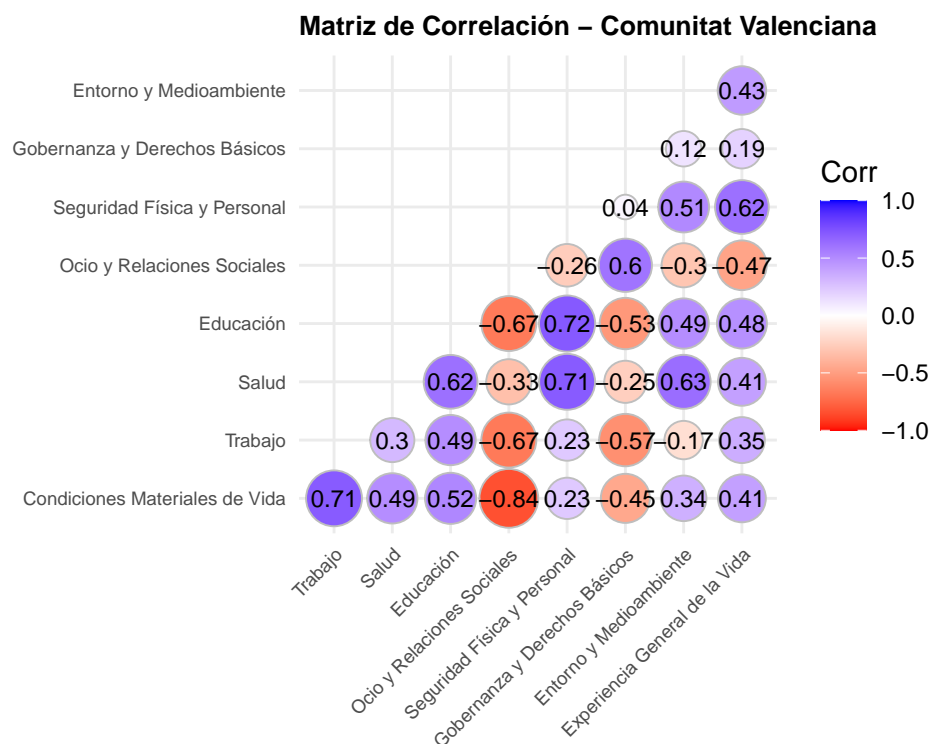
```
##           CCAA Anio  Dim      Valor
## 1 Andalucía 2008 dim1  98.19206
## 2 Andalucía 2008 dim2  95.79060
## 3 Andalucía 2008 dim3  97.64004
## 4 Andalucía 2008 dim4  96.31249
## 5 Andalucía 2008 dim5  98.45372
## 6 Andalucía 2008 dim6 100.29773
```

## Análisis de los datos

El análisis de datos es un proceso fundamental para extraer información relevante, identificar patrones y apoyar la toma de decisiones basada en evidencia. En este caso, los datos proporcionan información multidimensional que permite explorar la calidad de vida en diferentes comunidades autónomas a lo largo del tiempo. Este análisis combina técnicas descriptivas, como visualizaciones y resúmenes estadísticos, con enfoques más avanzados, como modelos estadísticos y exploraciones multivariantes.

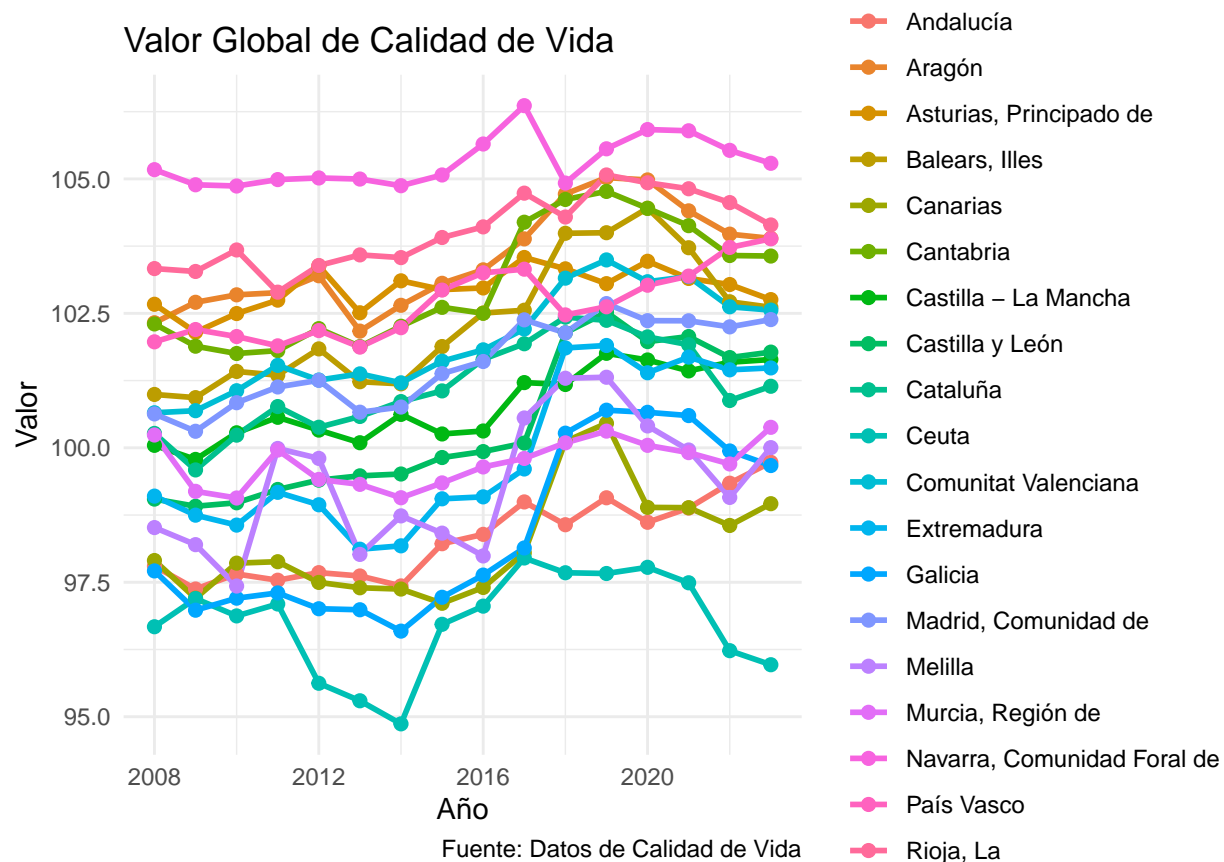
En primer lugar, notamos que no hay valores faltantes en el conjunto de datos, sin embargo sí se observó que ciertos valores para diferentes dimensiones se repetían durante varios años. Investigando esto se observó que no todos los indicadores están disponibles todos los años. Por ejemplo, el caso más extremo es el de la dimensión Gobernanza y Derechos Básicos para el que solo hay datos en el año 2013 y, parcialmente para 2022. Los valores para los otros años, que ya venían en el conjunto de datos descargados del INE, simplemente fueron repetidos el resto de años posteriores a la nueva adquisición de datos en 2023. También se puede observar la imputación de estos valores para las dimensiones Ocio y relaciones sociales y Experiencia general de la vida.

En primer lugar, vamos a analizar la relación de las diferentes dimensiones para cada comunidad autónoma. Esto se realiza calculando la matriz de correlación, a partir del método de Spearman, esta indicará una medida de cuán asociadas están las dimensiones. En la siguiente gráfica vemos un ejemplo de matriz de correlación para la Comunidad Valenciana, pero esta se realizó para todas las comunidades, se pueden ver en el Anexo, al final del documento.



El análisis de las correlaciones entre dimensiones en todas las comunidades autónomas revela que las Condiciones Materiales de Vida, el Trabajo y la Educación son las dimensiones más interrelacionadas, formando el núcleo del bienestar socioeconómico. Estas tres dimensiones están particularmente conectadas en regiones como Navarra, País Vasco y Madrid, que presentan sistemas más integrados. Por otro lado, dimensiones como el Entorno y Medioambiente y el Ocio y Relaciones Sociales tienden a ser más independientes, con menor conexión con factores económicos, aunque en algunas comunidades, como la Comunitat Valenciana y Melilla, muestran una relación moderada con la seguridad. La Experiencia General de la Vida está influenciada principalmente por las condiciones materiales y el empleo, aunque su alcance es más amplio en regiones con mayor desarrollo social. Finalmente, en comunidades como Melilla, Ceuta y Extremadura, las dimensiones muestran una mayor independencia, reflejando realidades locales más desconectadas entre sí. Este panorama resalta cómo el bienestar general varía significativamente según el contexto regional.

Siguiendo con el análisis, dado el volumen de datos disponibles, procedemos a simplificar el conjunto calculando un promedio de todas las dimensiones para obtener un valor global de calidad de vida por comunidad autónoma y por año. Este enfoque nos permite mantener la representatividad de todas las dimensiones sin perder información relevante.



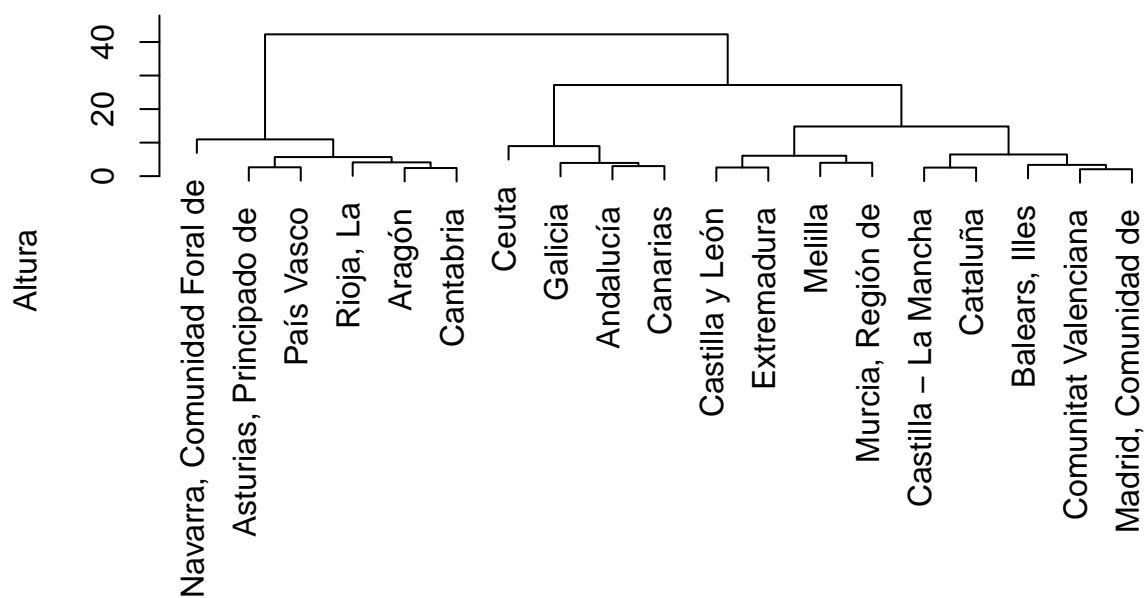
Asimismo, se puede apreciar en la gráfica que las tres agrupaciones presentan una tendencia lineal general creciente y muy similar entre ellas. Sin embargo, cabe destacar que, si bien la mayoría de las comunidades autónomas han experimentado un aumento de su índice de calidad de vida con poca variación, hay alguna comunidad autónoma que sí presenta grandes variaciones alrededor de su tendencia a lo largo del tiempo. Además, notamos que, a lo largo del tiempo, a penas hay solapamiento entre las agrupaciones, excepto en el último lustro. Este solapamiento cesa en el último año registrado.

En la gráfica anterior se puede observar el índice global de calidad de vida en las diferentes comunidades autónomas de España a lo largo del tiempo. Notamos que, en general, las comunidades y ciudades autónomas presentan una tendencia creciente para este índice (especialmente después de 2015), aunque en los últimos años se puede apreciar una ligera caída general. La mayoría de las comunidades y ciudades autónomas han experimentado estos fenómenos con poca variabilidad alrededor de su tendencia, a excepción de alguna que sí que presenta gran variabilidad.

El País Vasco y la Comunidad Foral de Navarra son las comunidades que presentan un índice más alto, que además se ha mantenido a lo largo del tiempo. Por otro lado, Ceuta presenta los valores más bajos en el conjunto, mantenidos a lo largo del tiempo. En los últimos años, se puede apreciar que Ceuta ha sufrido una caída alcanzando un índice en el último año que es notablemente más bajo que el de el resto de comunidades o ciudades autónomas. Parece haber disparidad entre ciertos grupos de comunidades y ciudades autónomas.

Debido a esto, clasificaremos las comunidades autónomas en clústeres que representen tres niveles: clase baja, media y alta, según su calidad de vida. Esto nos permitirá trabajar con menos datos de forma más manejable, conservando la gran parte de la información.

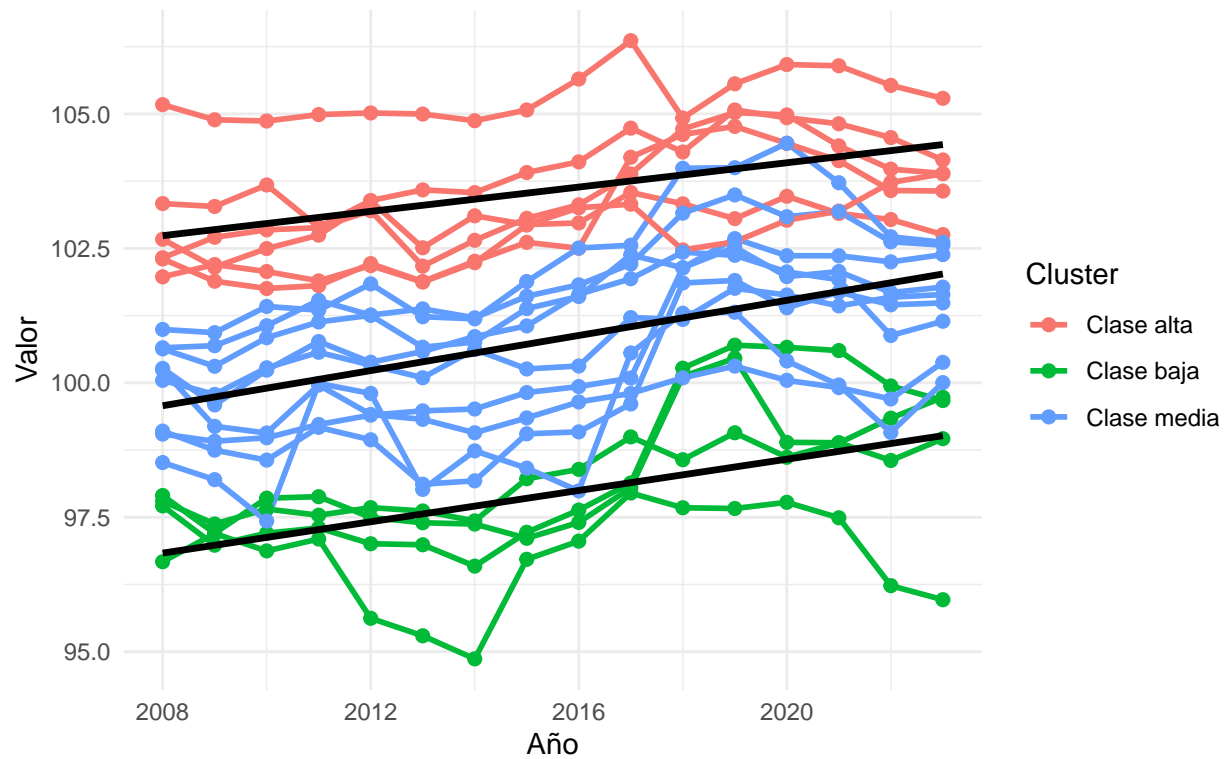
## Clustering Jerárquico de las CCAA



`hclust (*, "ward.D2")`

`## 'geom_smooth()' using formula = 'y ~ x'`

## Clustering de Calidad de Vida Global



Fuente: Datos de Calidad de Vida

## Cluster 1:

##

## Call:

## `lm(formula = Valor_Global ~ Anio, data = cluster1)`

##

## Residuals:

##	Min	1Q	Median	3Q	Max
##	-1.6747	-0.7713	-0.2502	0.7257	2.6095

##

## Coefficients:

##		Estimate	Std. Error	t value	Pr(> t )
##	(Intercept)	-124.07874	45.91165	-2.703	0.00817 **
##	Anio	0.11296	0.02278	4.959	3.14e-06 ***

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 1.029 on 94 degrees of freedom

## Multiple R-squared: 0.2073, Adjusted R-squared: 0.1989

## F-statistic: 24.59 on 1 and 94 DF, p-value: 3.14e-06

## Cluster 2:

##

## Call:

```
## lm(formula = Valor_Global ~ Anio, data = cluster2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.8902 -0.9010  0.1621  0.8919  2.9204
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -228.33700   43.36757  -5.265 5.07e-07 ***
## Anio         0.16330    0.02152   7.589 3.88e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.19 on 142 degrees of freedom
## Multiple R-squared:  0.2886, Adjusted R-squared:  0.2836
## F-statistic:  57.6 on 1 and 142 DF,  p-value: 3.88e-12

## Cluster 3:

##
## Call:
## lm(formula = Valor_Global ~ Anio, data = cluster3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.04993 -0.44938  0.04533  0.54934  2.26482
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -195.3640    59.8646  -3.263  0.00179 **
## Anio         0.1455     0.0297   4.899 7.23e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.095 on 62 degrees of freedom
## Multiple R-squared:  0.2791, Adjusted R-squared:  0.2675
## F-statistic:    24 on 1 and 62 DF,  p-value: 7.234e-06
```

Tras obtener las agrupaciones, notamos que las comunidades autónomas que forman cada clase son:

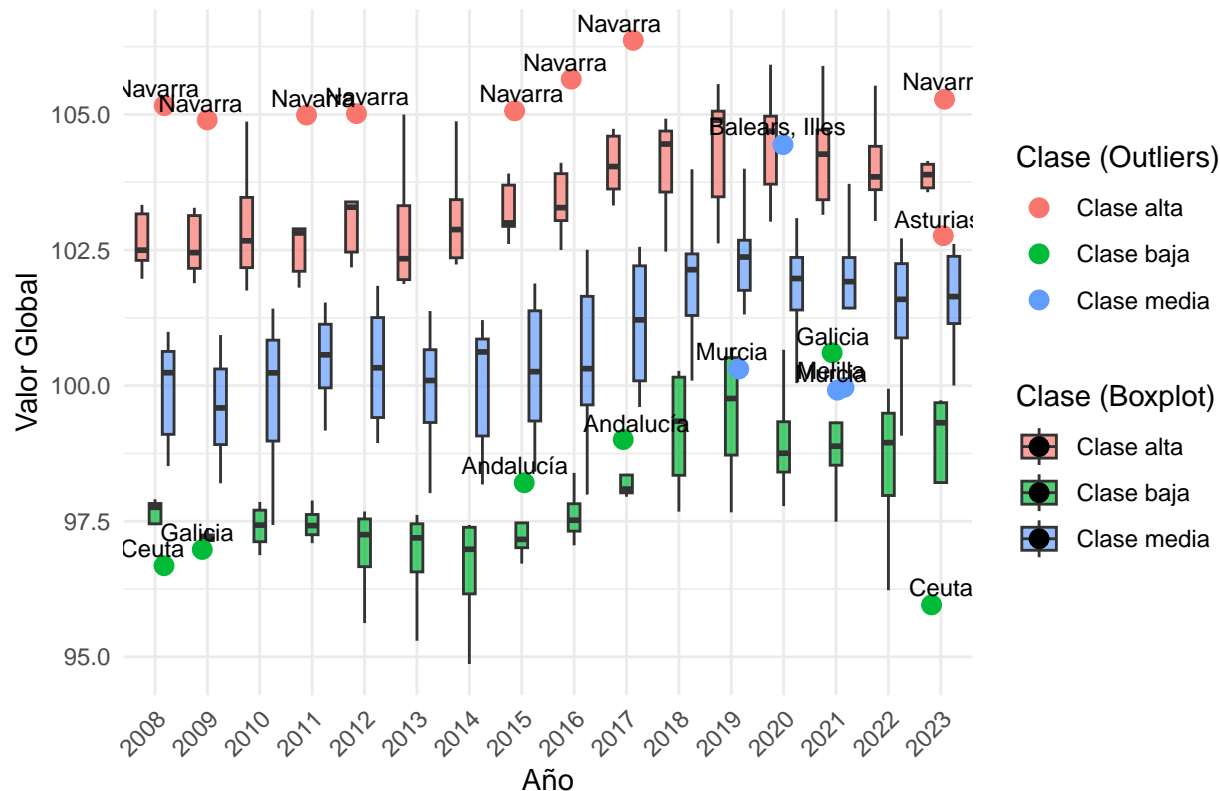
Clase alta: Aragón, Principado de Asturias, Cantabria, Comunidad Foral de Navarra, País Vasco, y La Rioja. Clase media: Illes Balears, Castilla - La Mancha, Castilla y León, Cataluña, Comunitat Valenciana, Extremadura, Comunidad de Madrid, Melilla, y Región de Murcia. Clase baja: Andalucía, Canarias, Ceuta, Galicia.

En la gráfica, se puede observar la recta de regresión ajustada a cada agrupación. Los dos coeficientes de cada recta son significativos a nivel  $\alpha = 0.05$ . Hay evidencia estadística a favor de la hipótesis sobre la tendencia lineal y creciente en los índices de las tres agrupaciones. Además, las pendientes de las rectas son muy similares. Por otro lado, la diferencia de altura entre las rectas de la clase alta y de la media es muy similar a la diferencia de altura entre las rectas de la clase media y de la baja. El valor de esta diferencia es de aproximadamente 3 puntos, i.e., hay una clara diferencia entre las clases.

Finalmente, notamos que a penas hay solapamiento entre las agrupaciones a lo largo del tiempo, excepto en el último lustro. Este solapamiento cesa en el último año registrado. La clase alta presenta un crecimiento estable a lo largo del tiempo. La clase media presenta un crecimiento generalmente estable hasta 2017, donde

experimenta un gran crecimiento relativo que se estanca decrece ligeramente posteriormente (a partir de 2019). La clase baja presenta un estancamiento genereal hasta 2015, donde experimenta un crecimiento general (que en algunos casos es superlativo), y posteriormente experimenta un decaimiento general (a partir de 2019).

## Boxplots con Outliers Identificados por Comunidad Autónoma



Observando la gráfica anterior, observamos que hay algunos valores extremos a lo largo del tiempo presentes en las agrupaciones. A través de las etiquetas asociadas a cada valor extremo, que identifican a qué comunidad o ciudad autónoma pertenece cada uno, verificamos algunas observaciones previas, y además logramos divisar nuevas observaciones.

Los valores extremos superiores de la clase alta pertenecen a la Comunidad Foral de Navarra; y el valor extremo inferior de la clase alta en el último año pertenece al Principado de Asturias. El valor extremo superior de la clase media en 2020 pertenece a les Illes Balears; y los valores extremos inferiores de la clase media pertenecen a la Región de Murcia (2019 y 2021) y a Melilla (2021). Los valores extremos superiores de la clase baja pertenecen a Andalucía (2015 y 2017) y a Galicia (2021); y los valores extremos inferiores de la clase baja pertenecen a Ceuta (2008 y 2023) y a Galicia (2009).

Habiendo observado estos valores extremos, volvemos a apreciar como la Comunidad Foral de Navarra ha liderado el conjunto con el mayor índice, con diferencia, del conjunto en su totalidad. Ceuta se ha mantenido con valores extremos alcanzando índices muy bajos comparados con los del resto de observaciones. Asimismo, Ceuta no ha experimentado ese crecimiento generalizado que sí se ha experimentado, en general, en el conjunto. Por último, notamos que Galicia ha experimentado un gran crecimiento a lo largo del período estudiado, llegando a superar a la Región de Murcia y a Melilla en algún año (como en 2021), que son miembros de la clase media.

Ahora, para poder analizar también las dimensiones, se pensó en crear las 3 macrocomunidades autónomas (clase alta, media y baja) y realizar el promedio de las diferentes dimensiones dentro de cada macrocomunidad. Sin embargo, como pueden ser independientes de cada comunidad, primero hay que hacer un análisis de cada



dimensión para saber si estas comunidades son factibles.