Instituto Superior de Agronomia

Analysis of agricultural activities: temporal trends

UC - Analysis and Visualization of Complex Agro-Environmental Data

Diana Santos Martins, 20900
Inês Laborinho Schwartz, 128119

2024

Contents

Introduction

In this project we aim to explore how agriculture practices have changed over time; and see how these changes of practices are related with changes in socio-economic indicators in Portugal. This exploration will ideally give us more information about how socio-economic indicators relate with agricultural practices and production values. Given that food systems involve everything from holding yield quantities, holding area, production value, and indicators such as education level and labour type, it is fascinating to explore the relations of these indicators.

The source of data for this project comes from the National Institute de Estatística (INE) database, a Portuguese database containing a volume of Portuguese census information.
For our purposes we were interested in gathering information on agricultural activities and socio-economic indicators. The INE database contains agricultural activities involving permanent, temporary, grassland crops and livestock grouped by the 2702 individual 'freguesias' or towns in Portugal with data from the years 1989, 1999, 2009, and 2019. Each of these topics also included information indicated in Table 1 and 2.

Furthermore, we wanted socio-economic indicators which we decided would be production values, education levels, and labor types and counts also grouped by each town in the 1989, 1999, 2009, and 2019 (refer to table 2). With information about both the agricultural activities happening in each town over the course of 40 years and information on how socio-economic conditions have changed over 40 years, we could demonstrate a lot about the changes that have taken place over time and make further predictions about how such practices have changed and relate to socio-economic factors. This information is both necessary and interesting to analyze given the environmental changes and socio-economic changes that have taken place in the last 40 years. The world is only becoming more fast-paced and more extreme environmental changes are taking place. Because of this, farmers will need to adapt their practices. Conversely, at the beginning of this project we were under the assumption that socio-economic conditions have improved in our societies. Therefore, this project acts as an interesting way to verify this assumption and explore what improved "socio-economic" conditions actually imply.

Specifically in this project we aim to explore the following questions:

a) How the number of agricultural holdings and production value of a holding are related over time.
b) Understand how the number of holdings increases or decreases and relate it with education level.
c) Understand how production value increases or decreases with education level.
d) Understand how production value increases or decreases with different labour type.

Data

We used the INE database, which was normalized and compiled during Data Management Systems with Professor Rui Figueira to build our dataset. In SQL, using DBeaver we created a large table that joined together information about the agricultural activities, labour types, production types, and education types all ordered by region at the 'freguesia' (town) level.
The metadata are described in Table 1 and 2.

Along with the INE database, Professor Pedro Segurado provided us with a spreadsheet that contained the information from the INE database. This spreadsheet included a table organized by 'Freguesia', and columns that described each type of agricultural activity, labour type, production type and education type along with other variables. We modified this spreadsheet for our variables of interest—production information, labour information, education information and agricultural activity information.

While the two datasets offered essentially the same information, they had different schemas. The spreadsheet provided by Professor Pedro offered detail on the exact counts of each variable (as an integer), while the dataset that was created using SQL gave a better general overview because it did not organize each crop type and socio-economic indicator type as a column.
During the coding aspect of the project, using both tables to address different analysis and visualization needs was useful.

Table 1: Agricultural Activity Data

About	**Type of agricultural activity**
Agricultural holdings with permanent crops (number of holdings per freguesia)	Citrus, fruit, fresh fruit, nuts, olives, vineyards, and other permanent crops
Agricultural holdings with temporary crops (number of holding per freguesia)	Cereals, Dried Pulses, Potatoes, Sugarbeets, Fresh vegetables, Flowers and Ornamental Plants, Industrial crops, Other temporary crops
Agricultural holdings with permanent crops (number of hectares per freguesia)	Citrus, fruit, fresh fruit, nuts, olives, vineyards, and other permanent crops
Agricultural holdings with temporary crops (number of hectares per freguesia)	Cereals, Dried Pulses, Potatoes, Sugarbeets, Fresh vegetables, Flowers and Ornamental Plants, Industrial crops, Other temporary crops

Table 2: Socio-Economic Indicator Data

Socio-Economic Indicator	**About**
Production value (euro)	Value of total standard production of agricultural holdings (in euro)
Production value (hectare/eur)	Average value of total standard output by hectare of utilized agricultural area (hectare/euro)
labour_regular	Volume of agricultural labour force that are Regular
Labour_non_regular	Volume of agricultural labour force that are non regular
Labur not hired	Volume of agricultural labour force that are workers not hired by the holder
Labour family	Volume of agricultural labour force that are workers that are from the labour force's family
Edu_none	Number of farmers without education by freguesia
Edu_basic	Number of farmers with basic education by freguesia
Edu_secondary	Number of farmers with secondary education by freguesia
Edu_superior	Number of farmers with superior education by freguesia

Data Analyses

Summary Statistics:

To begin our exploratory data analysis, we began by looking at how our data was distributed and which characteristics most stand out. The main aspects of the data that we were looking at were the temporary and permanent agricultural activities broken down by the count of holding per freguesia (town) and area of holding per freguesia (in hectares) and the socio-economic factors ---labour type, education level, and production value, in each county all divided into the years 1989, 1999, 2009, and 2019. We decided to look at the data from a national scale to get a generalized perspective of the trends happening in Portugal. We used stacked bar and other plots to demonstrate the data distributed over time for each category of interest.

Stacked Bar Plot Showing Education Types by Year

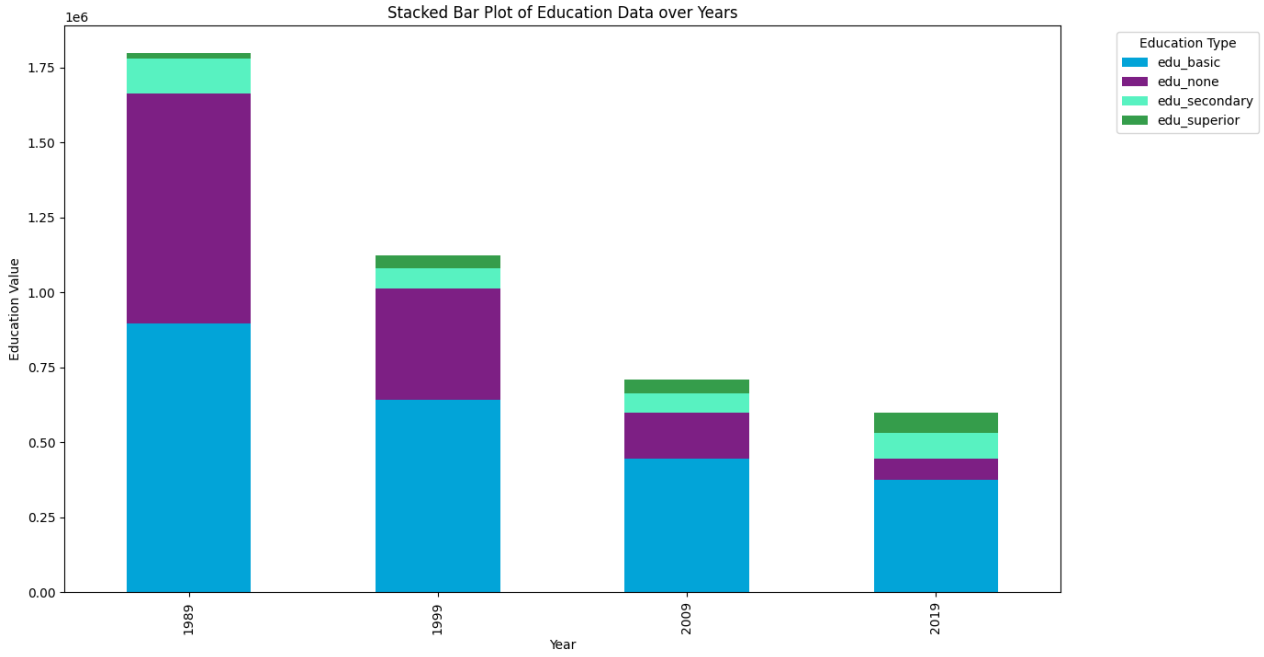

Fig 1 Education data over years

This graphic demonstrates how the education level values (number of farmers per freguesia) is distributed over the four different years. The x axis separates the data by year and the y axis, 'Education Value', shows us the proportion of farmers with each education type indicated in the key. As one would assume, there were the most farmers with no education per freguesia in 1989, indicated by the large amount of purple in the stacked bar plot. And there is an increase in secondary and superior education by 2009 and 2019, as indicated by the smaller presence of the blue and purple bars bars and the higher portion of the green bars. The large decrease in the bar plots is also interesting as it implies that there are fewer individual farmers in Portugal since 1989.

Stacked Bar Plot Showing Labor Types by Year

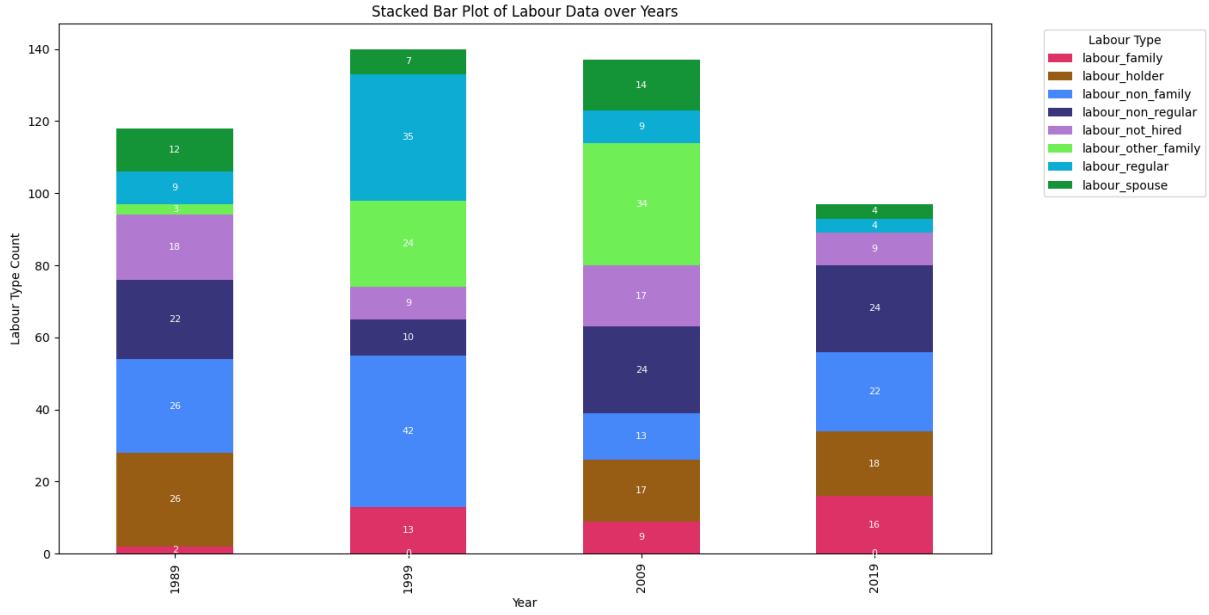

Fig 2 Labor data over years

Similarly, we can see how the average labor type count varies over time. It is possible to notice a decrease in the labor force as well as a decrease in the labor allocated to the family members. From year 1989 to 2019, there really are no outstanding changes in the labor type counts. Fortunately, the one category of 'labour_not_hired' did see a decrease from an average of 18 per freguesia in 1989 to nine in 2019. And similar to the trends in education levels, this graphic tells us interesting information about the decline in farmers in Portugal, no matter the labour type.

To better understand the generalized tendencies in agricultural activities over time, we divided the parameters only in temporary and permanent crops, graphics are as follow:

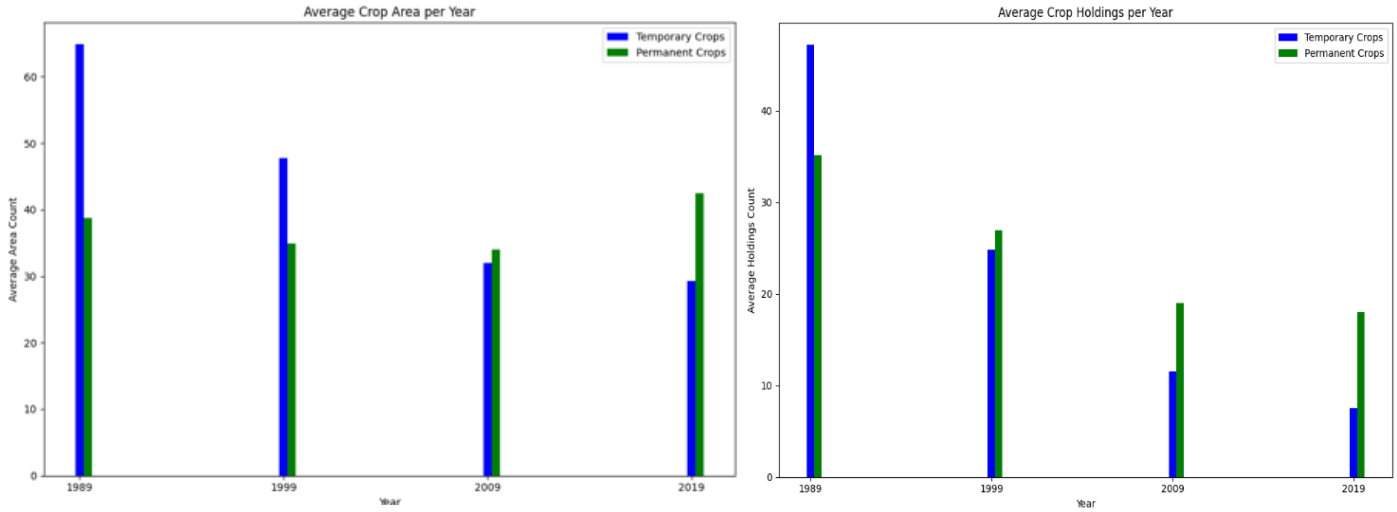

Fig 3 Average crop area per year and Average crop holding per year

This plot offers a very simplified visualization of the average crop holdings and average crop area per freguesia per year for the temporary and permanent crops. Using this graphic we can see that the number of temporary crop holdings have decreased across Portugal. In 1989, the average area of temporary crops was about 65 hectares, and it decreased to 29 hectares in 2019. The permanent crop holdings however, have slightly increased over time in Portugal. In 1989 the average area of permanent crops per freguesia was 39 hectares and increased to 42 hectares in 2019. The number of crop holdings show a similar tendency with a significant decrease in temporary crops.

Line Plots showing temporal trends in crop holdings:

Considering the general tendencies observed previously, it is also important to decompose the graphic and analyze the individual tendencies of all groups of cultures. For that we displayed some graphics (line plots) that detail the information obtained with the previous graphics.

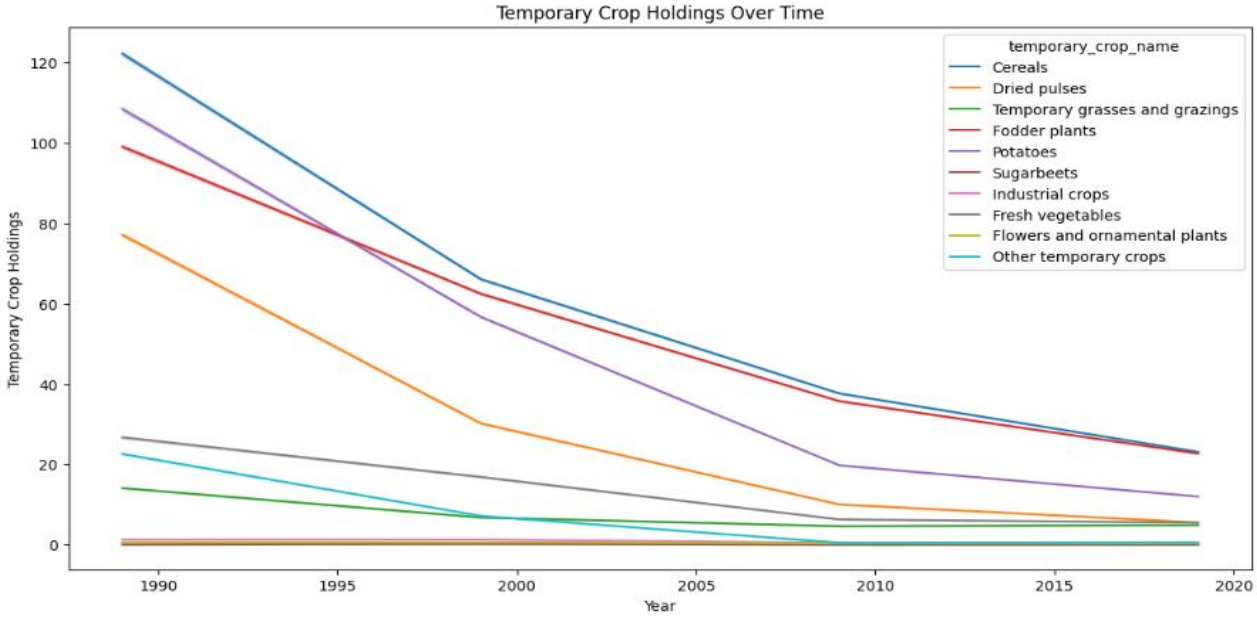

Fig 4Temporary crop holdings over time

The line graph helps us to understand that the number of temporary crop holdings has greatly decreased over the years. The line graph gives a clear temporal view of the changes in the data and tells us the specific crops. Looking at this line plot we can see the number of cereals holdings per freguesia decreased from a count of 122 in 1989 to a count of about 38 in 2019. The number of potato holdings also decreased vastly from 108 to about 20 from 1989 to 2019.

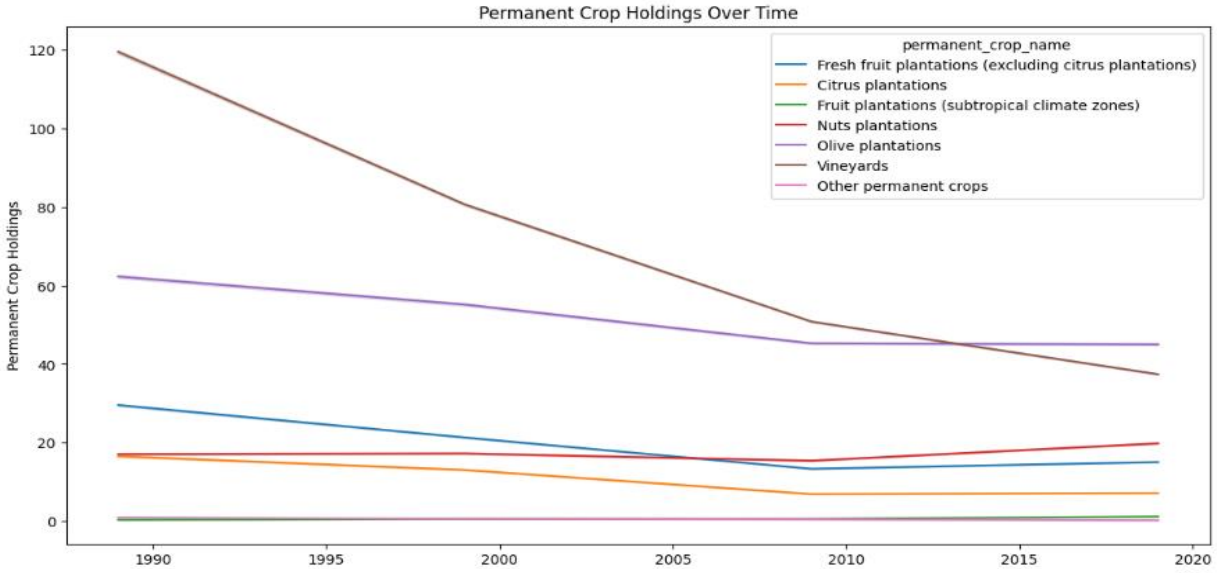

Fig 5Temporary crop holdings over time

The permanent crop holdings also saw a general decline, especially in vineyards which decreased from an average holding count of 120 in 1989 to 37 in 2019. The only value that increased in holding counts are nut plantations which had an average of 2 more holdings in 2019 per freguesia than in 1989 which had an average of 17 holdings per freguesia. Perhaps this is due to the increasing almond industry in Portugal.

Line Plots showing temporal trends in crop area:

And to understand how permanent and temporary crop area has increased or decreased over time we made more line plots. This data generally follows the same trend as the holding count data.

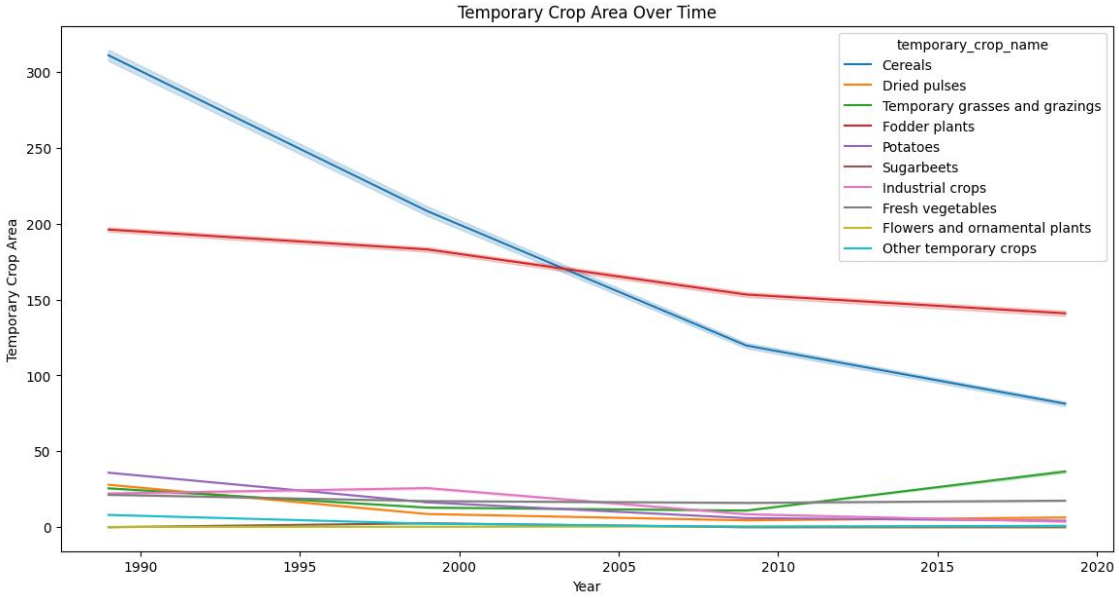

Fig 6 Temporary crop area over time

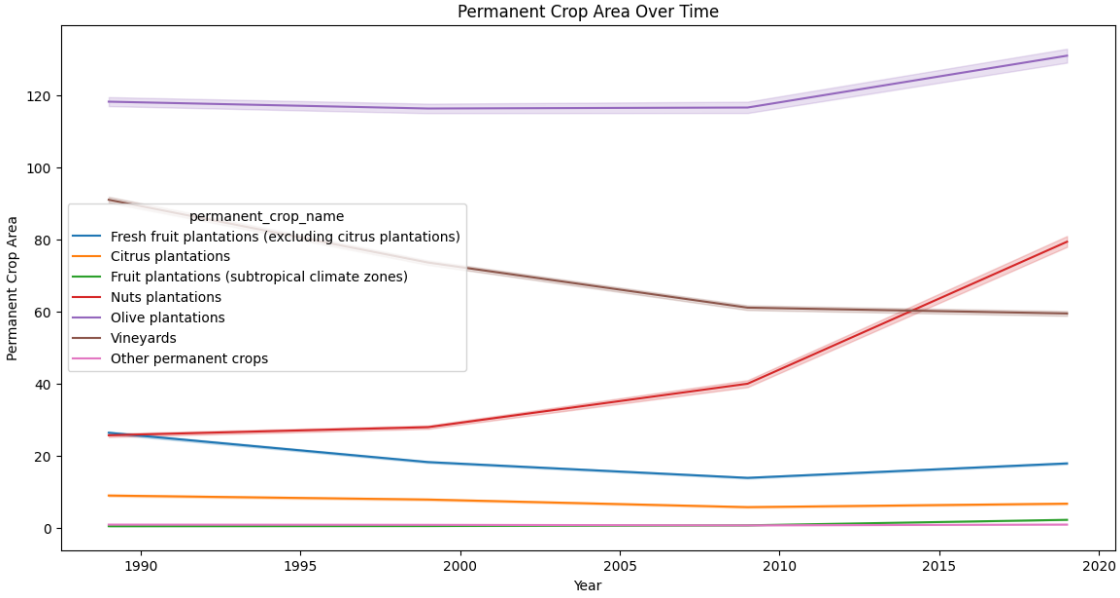

Fig 7 Permanent crop area over time

Production Summary Statistics:

Furthermore, it is also important to observe the production area and production value developments over time. These first graphics give us some very general summaries of how production area and value have increased over time.

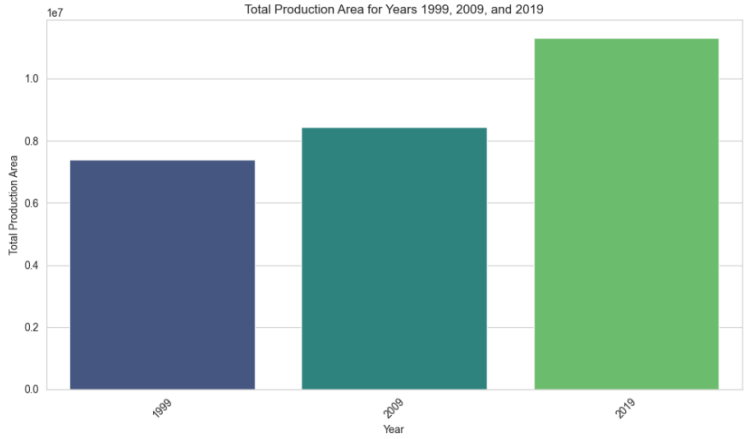

Fig 8Total production area over time

Fig 9Total production value over time

The graphics here show us that the total production area and value has increased in the last 40 years. Considering that labour force has decreased, these are very interesting results.

Education Levels and Production Analysis:

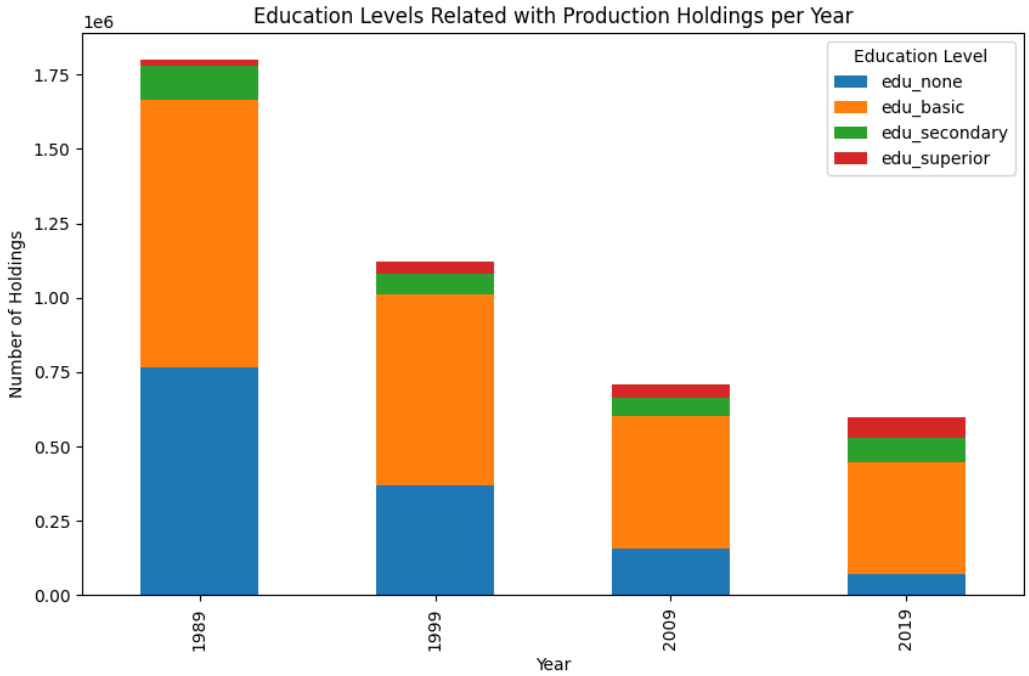

Fig 10: Total education levels related with production holdings per year

Here it is possible to see that the total number of agricultural holdings have decreased over the years. It is also noticeable that the number of farmers without education or with only basic education has been declining, while in recent years there has been an increase in the number of holdings with farmers having higher education. It is interesting not to forget the previous graph which shows that the generated values increased in the year 2019, coinciding here with a lower number of holdings and an increase in the education level.

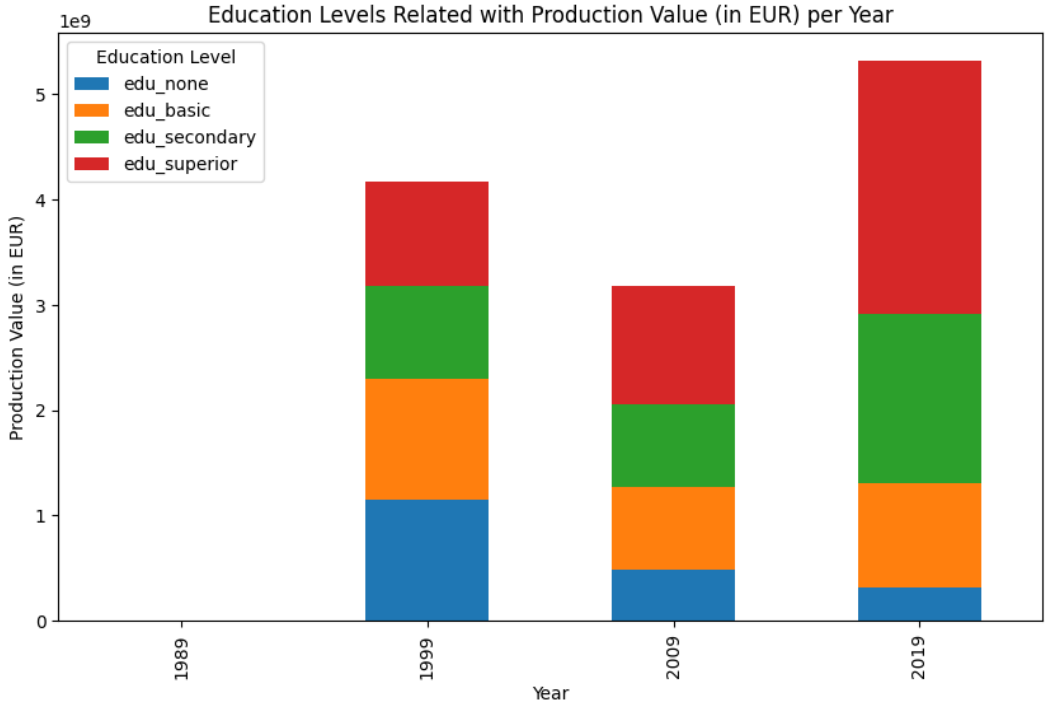

Fig 11Total education levels related with production value per year

Despite the absence of data for the year 1989, it is possible to verify that the value in euros produced over the years has increased significantly. Lower levels of education are associated with decreases in productivity in euros, whereas higher levels of education demonstrate higher productivity in euros.

Production and Education Regression Analysis

To further verify this we ran a generalized regression analysis looking at how education levels affect production value per holdings in each freguesia (refer to production_and_edu.ipyn in Annex). The Poisson regression analysis reveals compelling insights into the relationship between education levels among farmers and crop holding production values in euros. The intercept coefficient of 13.7161 signifies the baseline production value when all education variables are zero. Education plays a pivotal role: while farmers without formal education show a slight negative association with production values, those with basic education contribute positively. However, the impact escalates significantly with higher education levels. Farmers

with secondary education exhibit a notable increase in production values, and those with superior education show the most substantial positive influence.

Generalized Linear Model Regression Results

Dep. Variable:	production_eur	No. Observations:	8544
Model:	GLM	Df Residuals:	8539
Model Family:	Poisson	Df Model:	4
Link Function:	Log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-9.6108e+09
Date:	Wed, 19 Jun 2024	Deviance:	1.9222e+10
Time:	18:26:35	Pearson chi2:	3.97e+10
No. Iterations:	9	Pseudo R-squ. (CS):	1.000
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
intercept	13.7161	1.21e-05	1.14e+06	0.000	13.716	13.716
edu_none	-0.0009	1.32e-07	-7133.931	0.000	-0.001	-0.001
edu_basic	0.0012	1e-07	1.19e+04	0.000	0.001	0.001
edu_secondary	0.0071	5.22e-07	1.35e+04	0.000	0.007	0.007
edu_superior	0.0084	4.96e-07	1.69e+04	0.000	0.008	0.008

Fig 12Regression Analysis results

Prediction production value by education levels based on Poisson regression results

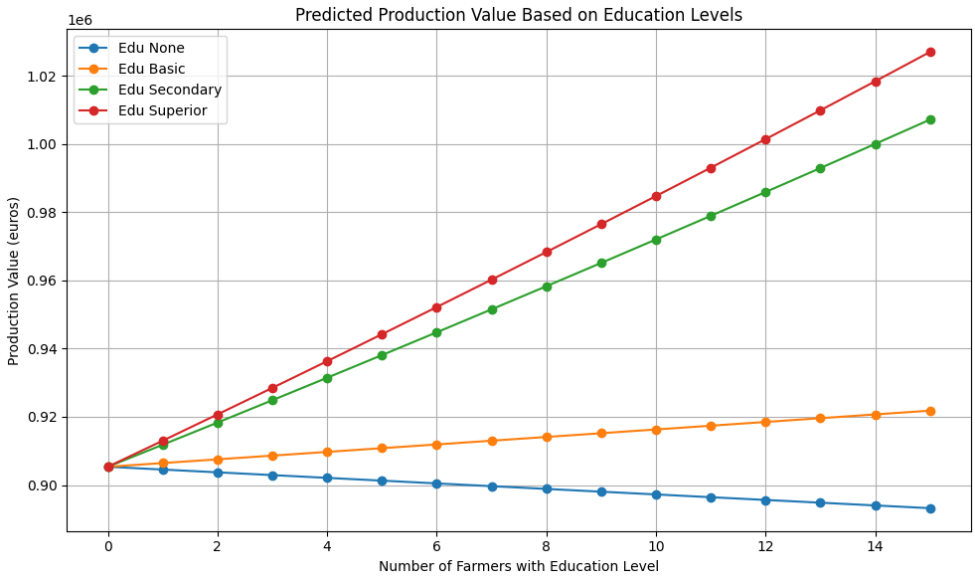

Fig 13Regression Analysis Predictions

The predictions shown in the visualization are based on a Poisson regression model that examines how different levels of farmer education influence crop holding production values in euros. The coefficients used in the model, specifically coef_intercept, coef_edu_none, coef_edu_basic,

coef_edu_secondary, and coef_edu_superior, quantify the expected change in production values for each unit increase in the respective education category, with all other variables held constant.

To generate the predictions, we selected a range of education levels (edu_levels) based on the observed data. These coefficients were used to compute the logarithm of predicted production values (log_production_value), which were then exponentiated to obtain the actual predicted production values (predicted_production_value). The resulting visualization illustrates how changes in education levels among farmers correlate with predicted changes in crop holding production values, providing insights into the influence of education on agricultural outcomes as modeled by the Poisson regression analysis.

Based on the regression analysis results provided earlier, the coefficients for education levels reveal significant insights into their impact on production value per holding in euros. Specifically, the negative coefficient for edu_none suggests that an increase in the number of individuals without education correlates with a decrease in production value per holding, indicating a detrimental effect on agricultural output. Conversely, positive coefficients for edu_basic, edu_secondary, and edu_superior indicate that higher education levels are associated with higher production values per holding. These findings underscore the importance of educational attainment in enhancing agricultural productivity, highlighting how investing in education at various levels can potentially yield economic benefits by improving production outcomes per agricultural holding.

Labor Type and Production Analysis:

We next tried to analyze to what degree the type of labor happening in each freguesia in Portugal over the last 40 years was related to the production value.

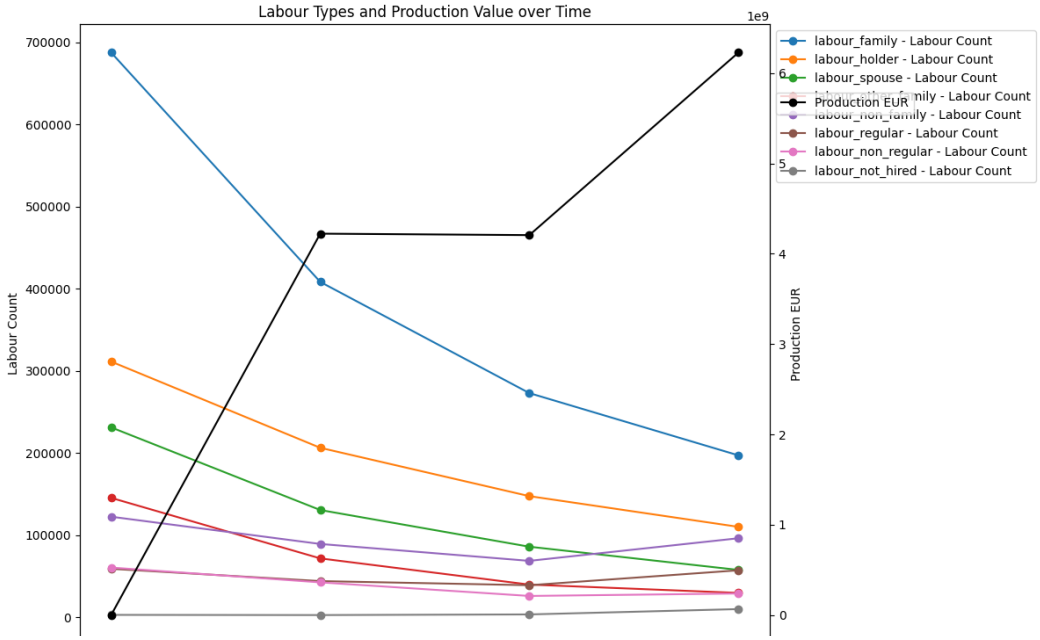

Fig 14Labor Types and Production Value over time

Unfortunately, the data only shows decreasing counts for each type of labor with increasing trends in production value. This trend can be attributed to a variety of interrelated factors. Technological advancements and mechanization have revolutionized farming practices, allowing tasks to be completed more efficiently with fewer laborers. Shifts in agricultural techniques, such as precision farming and the adoption of less labor-intensive crop varieties, also contribute to reduced labor requirements. Economic incentives and policies promoting efficiency gains might further drive this trend, as farmers optimize labor use to enhance productivity and reduce costs.

Additionally, evolving labor market dynamics, educational attainment, and government policies influencing farm management and subsidies play crucial roles. These multifaceted factors underscore the complex dynamics shaping labor counts in agriculture amidst rising production values. Overall, the decrease in laborers highlights the need for holistic approaches in agricultural policy and practice to sustainably meet future food production challenges.

Association between each labor type and production value:

To understand how well associated these variables were, we ran Chi-square tests and Spearman correlation coefficients.
The Chi-square tests and Spearman correlation coefficients provide valuable insights into the relationship between labor type counts and production values across different categories. The Chi-square tests assess the association between categorical variables, such as different types of labor (e.g., family, non-family, regular, non-regular, etc.), and production values. For instance, significant Chi-square statistics with low p-values suggest that certain types of labor significantly influence production values, indicating a non-random association.

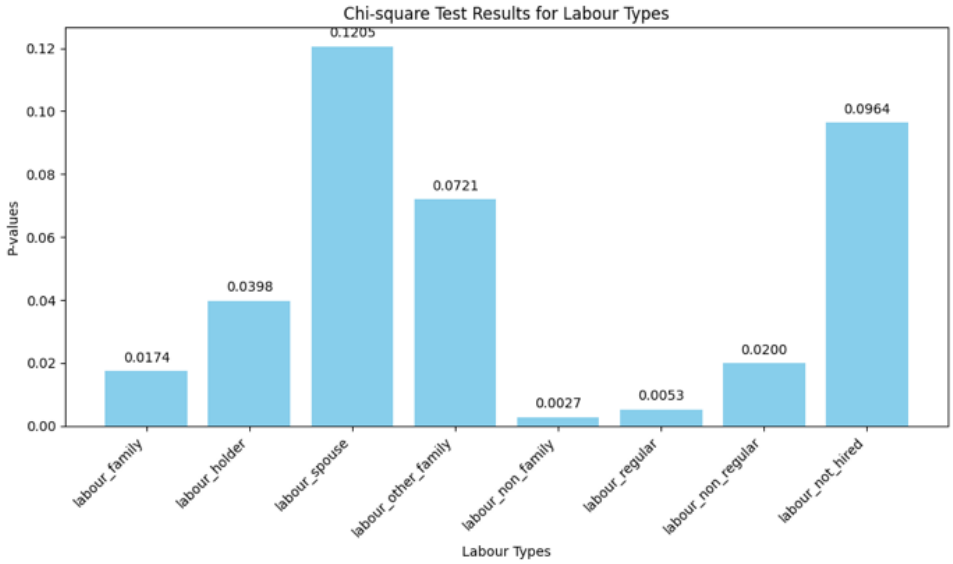

Fig 15Labour Types and Production Value Chi-Square results

On the other hand, Spearman correlation coefficients quantify the strength and direction of monotonic relationships between labor type counts and production values. Higher Spearman

correlation coefficients indicate stronger positive correlations, implying that increases in labor type counts are associated with increases in production values. Conversely, negative coefficients indicate an inverse relationship.

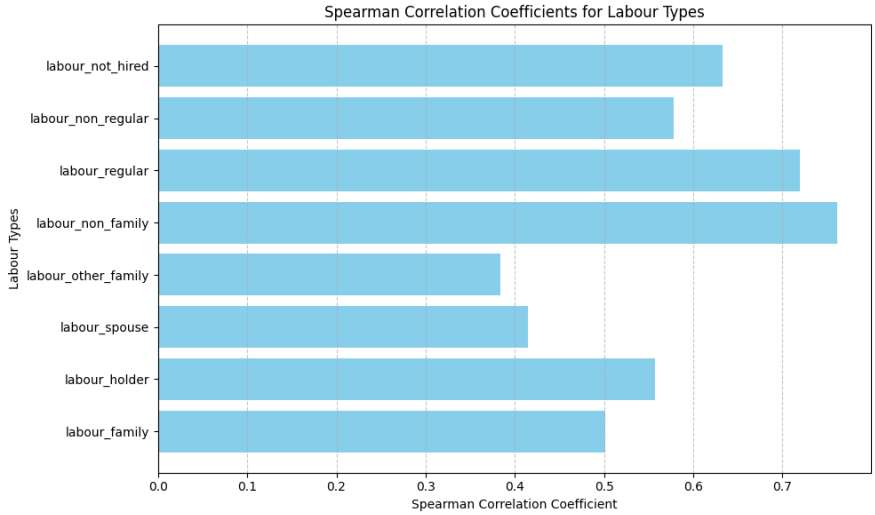

Fig 16Labour Types and Production Value Spearman Correlation results

In our analysis, we observed generally positive Spearman correlations across various labour types (e.g., family, holder, spouse, etc.), indicating that higher counts of these labour types tend to correspond with higher production values. This suggests that these labour categories play crucial roles in enhancing agricultural production within the studied context. Additionally, significant Chi-square results underscore the importance of these labour types in driving variations in production values, highlighting potential areas for targeted interventions or further exploration in agricultural management strategies.

Putting it all together:

The results from the Chi-square tests and Spearman correlation coefficients show the dynamic relationship between labor type counts and production value amidst overall trends. Despite the general decrease in labor type counts across various categories (e.g., family, non-family), the findings indicate that the remaining labor force is becoming more efficiently utilized or skilled, as evidenced by the positive correlations with increased production values. This suggests that while fewer individuals might be involved in agricultural labor overall, those remaining are likely more productive or efficient. The significant associations identified through Chi-square tests underscore that changes in labor distribution across different categories are not random but potentially strategic, aligning with efforts to optimize production efficiency and output in agriculture. Therefore, these results imply a nuanced adjustment in labor management strategies, potentially reallocating resources towards more productive labor types, like technology and mechanization to sustain or enhance overall production values despite fewer laborers.

Another look at the number of holdings and area of holdings per crop per year:

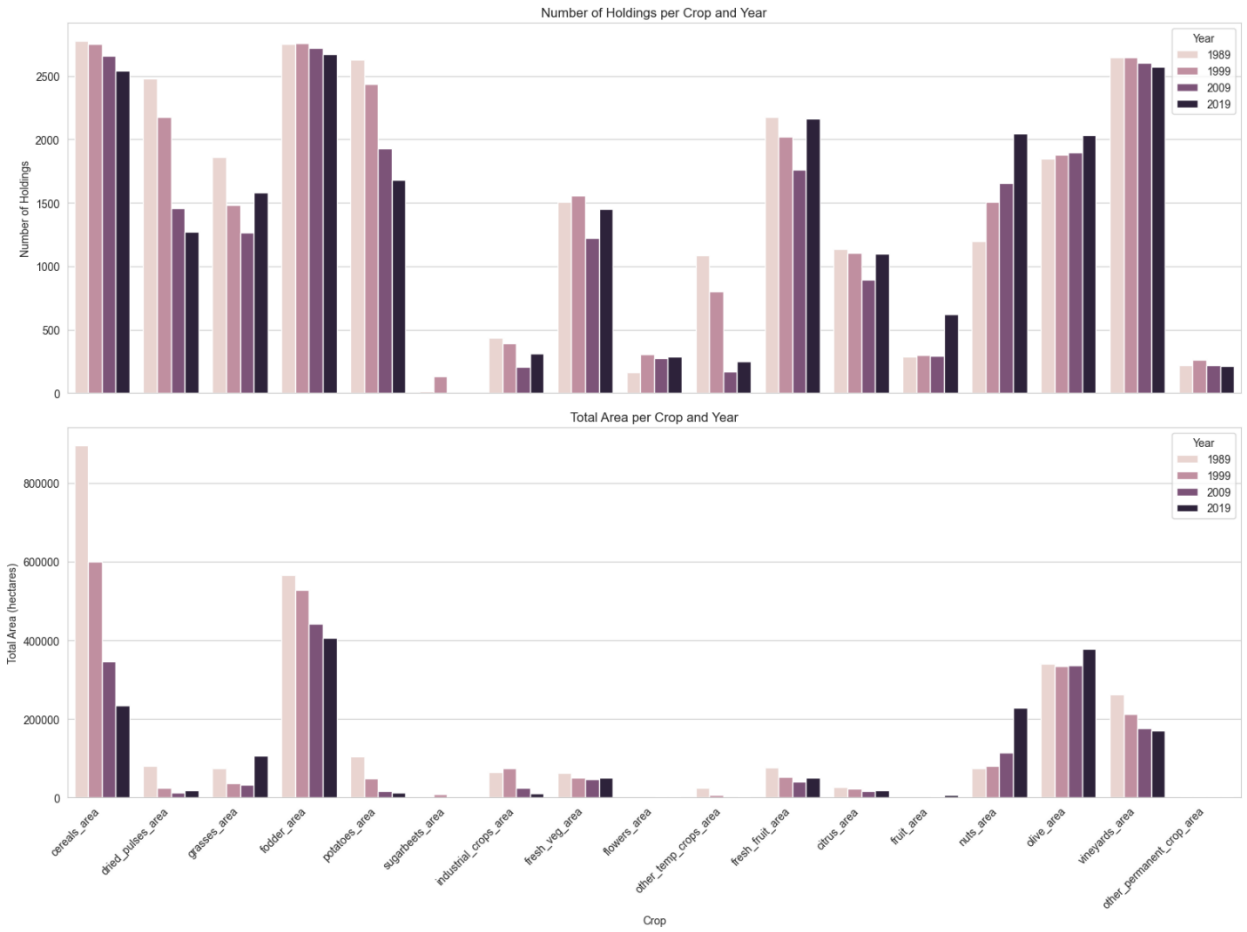

Fig 17Number of holdings and total area per crop per year

This graphic shows some very interesting information on how the total number of crop holdings is increasing or decreasing along with the total area of crop holdings. For all but the olive, nuts, and grasses; we see a decrease in total holdings and total area. That said, it must be noted that despite an overall decrease in crop holding counts and areas, the production value is going up!

Further exploration of temporal trends between agricultural activity and socio-economic factors:

First we ran a Principal component analysis to look how much our agricultural activity data varied by year. Overall, we see that there is a good degree variance in the agricultural activity data in 1989, but nothing stands out too much.

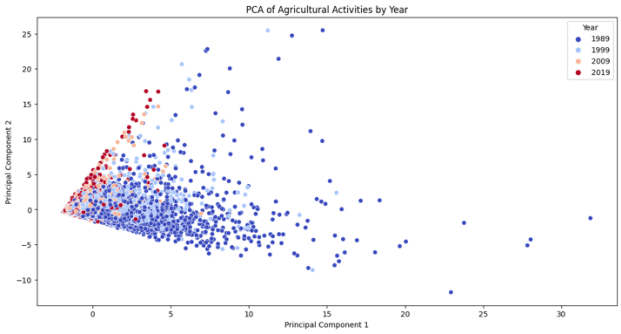

Fig 18 PCA of agricultural activities by year

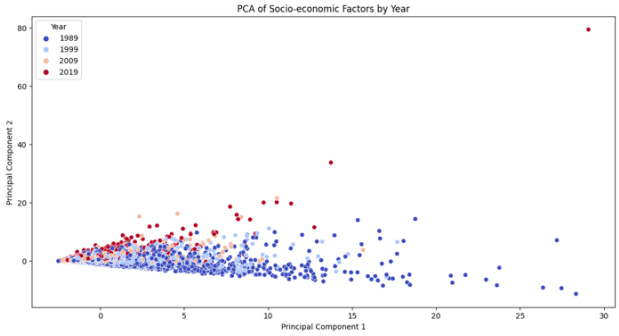

Fig 19PCA of socio-economic factors by year

The socio-economic factors shows a bit more variance over the years, again especially in 1989, but again this does not give us any big insights in the temporal trends of agricultural activity and socio-economic factors.

We then ran a Linear Discrimant Analysis to see to what degree our data can be separated by year. Like we found from the poor PCA results, the LDA results show that the agricultural activities and socio-economic factors cannot be well separated on a temporal scale.
It is interesting to note that potato count and cereal count seem to be the biggest causes of separation amongst the data.

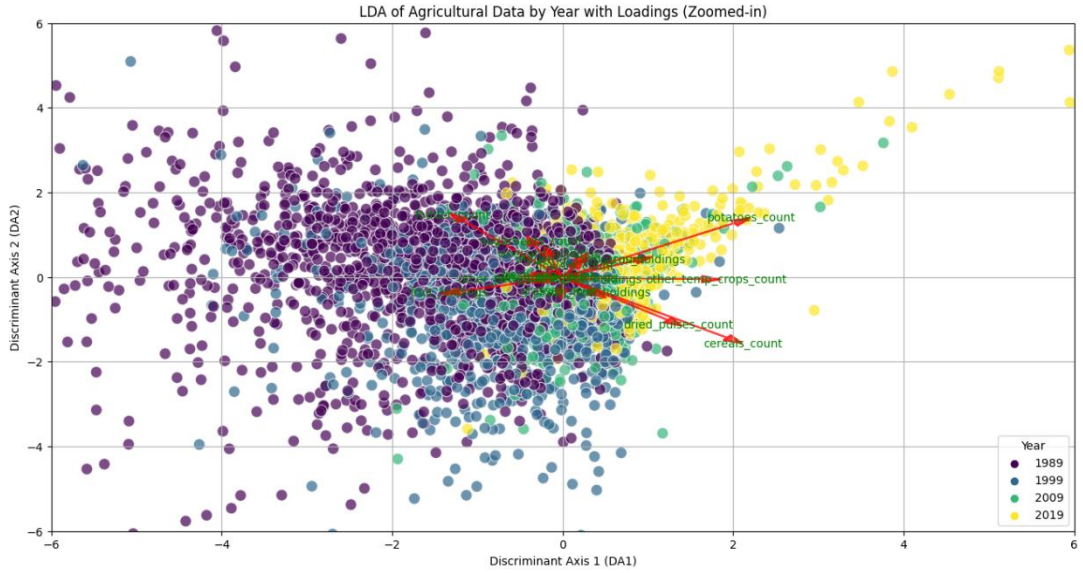

Fig 20LDA of agricultural Data by year with Loadings

The socio-economic data indicates a larger degree of separation between the years 1989 and 2019. While it doesn't show us anything too conclusive, again it is interesting to note that labor regular and labor non_family seem to be the biggest cause of separation amongst the socio-economic data set.

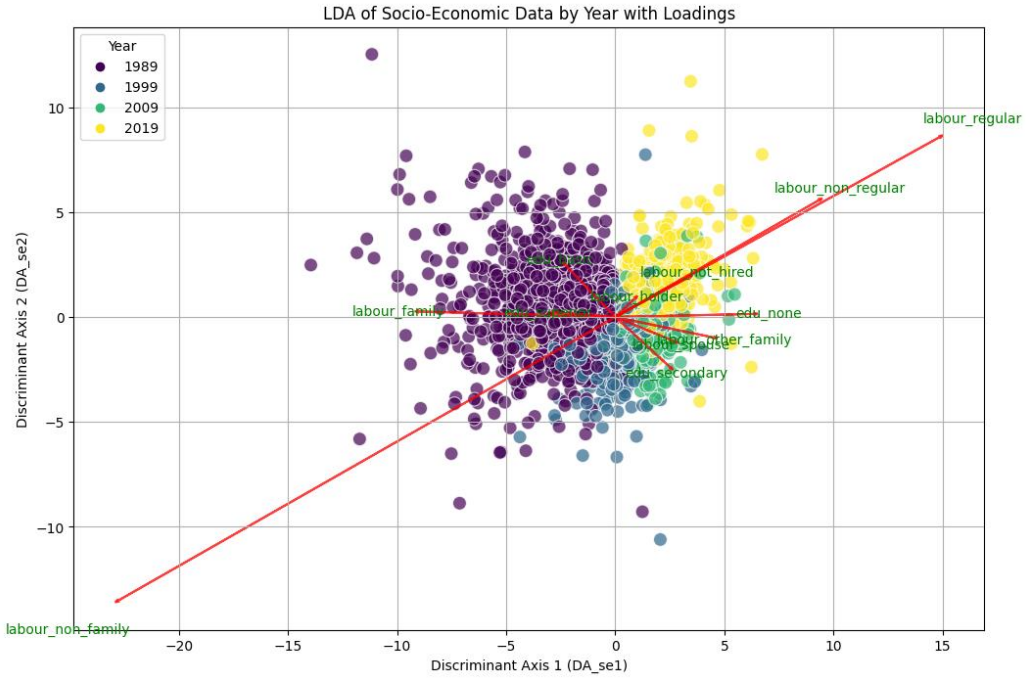

Fig 21LDA of socio-economic data by year with loadings

Discussion

In this final section of our report, we reflect on the insights gained from our analysis of agricultural activities and socio-economic indicators over the past four decades in Portugal. Our primary objective was to uncover relationships between changes in agricultural practices and socio-economic factors, aiming to provide meaningful conclusions and implications for future agricultural policies and practices.

Relationships Between Agricultural Activities and Socio-Economic Factors
Our exploratory data analysis revealed intriguing insights into how agricultural activities and socio-economic factors have evolved over time. We observed significant trends such as a decrease in the number of agricultural holdings while production values have generally increased. This suggests a shift towards larger, more productive agricultural operations, potentially influenced by technological advancements and changes in agricultural policies.

Education and Its Impact on Agricultural Productivity
A critical finding of our analysis was the positive correlation between education levels among farmers and production values per holding. Specifically, farmers with higher education levels—secondary and superior—tended to exhibit higher production values. This underscores the importance of education in enhancing agricultural productivity and economic outcomes. Policies that promote higher education among farmers could lead to further improvements in agricultural efficiency and sustainability.

Temporal Trends and Their Implications
Despite observing temporal changes in both agricultural activities and socio-economic factors, our linear discriminant analysis indicated that these changes do not neatly separate over time. This suggests that while there have been shifts in agricultural practices and socio-economic conditions, these changes do not necessarily align chronologically. This finding calls for a nuanced approach in analyzing the dynamic interactions between agricultural activities and socio-economic factors, considering various contextual factors that influence these trends.

Implications for Policy and Future Research
Our study provides valuable insights for policymakers and stakeholders in agriculture and rural development. The positive association between education and productivity highlights the potential benefits of investing in educational programs tailored to farmers. Moreover, the observed trends in agricultural practices underscore the need for adaptive strategies that can enhance agricultural resilience in the face of evolving socio-economic and environmental challenges.

Conclusion

In conclusion, our project aimed to explore and understand the complex interplay between agricultural practices and socio-economic indicators across all ‘freguesias’ in Portugal over a 40-year period. Through comprehensive data analysis and visualization techniques, we identified key trends and relationships that contribute to our understanding of how agricultural productivity and socio-economic conditions evolve over time. Moving forward, continued research and policy efforts should focus on leveraging education and technological advancements to sustainably enhance agricultural productivity and resilience.

By addressing these challenges, stakeholders can foster a more resilient and productive agricultural sector that contributes to economic growth and environmental sustainability in Portugal and beyond.

References

Figueira, Rui. "Data Management and Storage." Data Management and Storage, Institute Superior de Agronomia, Sala 48, 2023.

Stamer, John. "**StatQuest: Linear Discriminant Analysis (LDA) clearly explained**." StatsQuest, YouTube, 10 July. 2010, https://www.youtube.com/watch?v=azXCzI57Yfc.

ANNEX – Python code

Projeto2.ipynb: code for summary statics, regression analysis and some PCA and LDA's

SQL_analysis.ipynb: information on how we extracted the INE database from SQL, analyses on the agricultural activity totals and averages, analyses on the temporal trends of the the agricultural activies

holdings_edlevel_regression.ipynb: Regression analysis on agricultural holdings and education level.

lda.ipynb: LDA for agricultural activities and socio-econ factors

projeto.ipynb: more bar plots for education and agricultural activities