INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa

Final project in Analysis and Visualization of Complex Agro-Environmental Data

AGRICULTURAL PRACTICES

geographical patterns

temporal trends

Master in Green Data Science

Professor: Pedro Segurado

Students: Joana Esteves nº 24914 & Vasco Florentino nº 23411

2022/2023 2º Semester

Table of Contents

Links for additional deliverables:

Github Repository: https://github.com/jfilipa1/AVCD_JE_VF.git

Storytelling Presentation: https://www.canva.com/design/DAFmNO8A45I/DRoJER2SUGNm0ragJfRX9g/view?utm_content=DAFmNO8A45I

Introduction

In a world where the growth of the population is projected to keep increasing, food production must keep pace. Agricultural practices are really important because of their output, by producing materials that contribute to the world's supply chain and have an effect on the economy.

In Portugal, a considered developed country, it is also interesting to understand how the distribution methods work. Where are most of the crops located? Is there a reason for having a bigger distribution in a certain region of the country?

Image 1: Wordcloud created with the names of Temporary and Permanent crops.

INE (*Instituto Nacional de Estatística*) collects data on national agricultural activity on a regular basis. Data from different years is easily accessed, which makes it easier to even understand how the crop distribution changed over time. Have agricultural practices increased in Portugal? Is there a type of crop whose production is outstanding? Are there any interesting insights revealed?

To perform an analysis of agricultural activities, we must have information about the regions in Portugal, including the region level (*Freguesias* is the region level chosen), information for different years and for the areas of crop production.

Data about Temporary and Permanent crops was also provided. Permanent crops include Citrus plantations, Olive plantations, Fresh fruit plantations (excluding citrus plantations),

Fruit plantations (subtropical climate zones), Nuts plantations, Vineyards and Other Permanent crops. Meanwhile, Cereals, Dried pulses, Temporary grasses and grazings, Fodder plants, Potatoes, Sugarbeets, Industrial crops, Fresh vegetables, Flowers and ornamental plants and Other Temporary crops are all considered Temporary crops. How are the crops geographically distributed according to its type? Is the production of Permanent and Temporary somehow related?

Our main objective is to explore agriculture practices in Portugal, concerning different types of crop and observing the evolution in the national territory, over the years. We can gain a deeper understanding of agriculture in different geographical regions, by recognizing the patterns of change.

Database description

To access data related to agricultural activities in Portugal over the years, a database was provided by Professor Rui Figueira. The data is from INE (*Instituto Nacional de Estatística*) and to select only relevant information for our project, the data was manipulated in DBeaver and OpenRefine, before any analysis takes place.

For the analysis of our project, there were attributes that were considered more relevant. Therefore, in DBeaver a SQL script was created, selecting data with the following restrictions:

- data about the NUTS and the respective regions' names
- data only about the region level called *Freguesia*
- data of the years 1989, 1999, 2009 and 2019
- data about Permanent and Temporary crops
- data about the area of the crops

After obtaining the required data, OpenRefine was needed to create a Primary Key that could help to merge the data from all the selected years. For the analysis purpose, it made more sense to transpose some cells across columns into rows, like the type of crops, and also made it possible to combine the crops, NUTS and regions' names into the Primary Key.

The final tables exported have the following schemas:

3_perm_crops (Nuts_region_perm, NutsID, region_name, Perm_crop, Area_1989, Area_1999, Area_2009, Area_2019)

3_temp_crops (Nuts_region_temp, NutsID, region_name, Temp_crop, Area_1989, Area_1999, Area_2009, Area_2019)

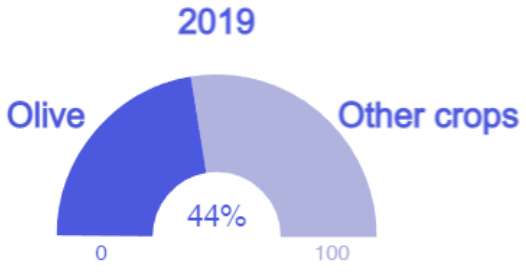
Both tables were stored in a CSV file, with 3_perm_crops containing 21645 rows and 3_temp_crops containing 30921 rows.

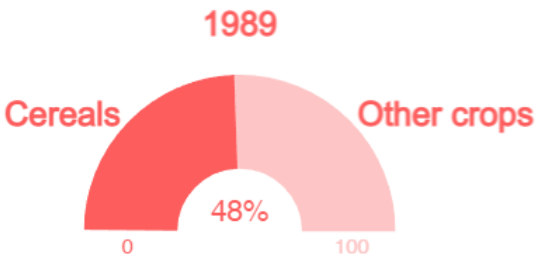
Data analysis

From 1989 to 2019, data about agricultural practices, within different *Freguesias,* was provided, allowing us to explore a journey through geographical patterns and temporal trends.

Summary statistics

Starting by analyzing the total area for every crop for every year, it is evident that there are crops that stand out. In 2019, Olive plantations represented 44% of the total area of Permanent crops in all *Freguesias*, and in 1989, 48% of the total area of Temporary crops was occupied by Cereals.

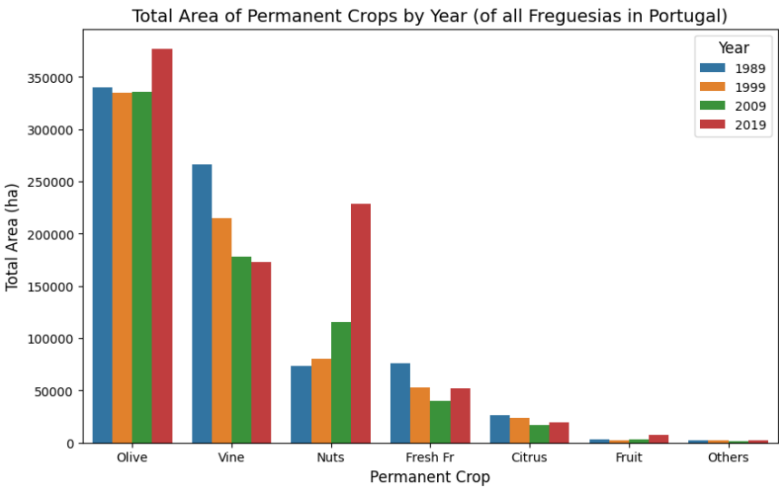




Graph 1 (left): Half pie chart representing the percentage of the total area of Permanent crops occupied by Olive plantations, in 2019. The area of Olive groves in 2019 is the biggest value we observe in the bar chart (graph 3).

Graph 2 (right): Half pie chart representing the percentage of the total area of Temporary crops occupied by Cereals, in 1989. The area of Cereals in 1989 is the biggest value we observe in the bar chart (graph 5).

For the Permanent crops, besides Olive plantations (Olive), the biggest areas of production, during these 30 years, was composed of Vineyards (Vine), Fresh fruit plantations (Fresh Fr) and Nuts plantations (Nuts).



Olive Plantations had an area of 340 484ha in 1989, having increased 10.79% in 2019, in comparison to the early year of 1989. This may be related to the rise in the area of olive groves that were cultivated intensively, and also because of the water that was provided by Alqueva in the region of Alentejo.

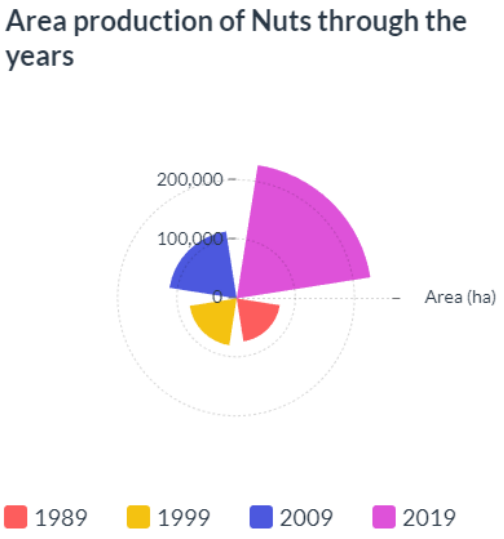
Graph 3: Bar chart representing the total area for each Permanent crop, in Portugal (for all *Freguesias*). Each bar represents a year, with the colors according to the legend.

Table 1: Variation of area (ha) for each Permanent crop, from 1989 to 2019.

Crop	Variation (hectares)
Citrus plantations	-7201.00
Olive plantations	36751.00
Fresh fruit plantations (excluding citrus plantations)	-24637.00
Fruit plantations (subtropical climate zones)	4643.00
Nuts plantations	154844.00
Vineyards	-93057.00
Other permanent crops	-167.00

Even though the Olive plantations had the biggest area in Portugal, for the Permanent crops, the Nuts plantations had the biggest variation between 1989 and 2019. It is observed a positive and negative change in cultivation of the Permanent crops, most of the crops having a negative variation (Table 1).

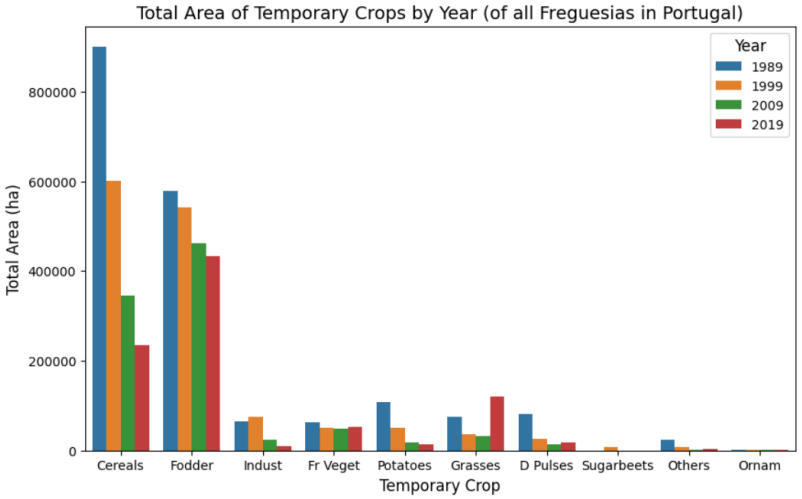
The values of the area of Nuts production were always increasing, but with small variations, between 1989 and 1999, and between 1999 and 2009. However, as soon as the value for 2019 shows up (graph 3), you realize the big Nuts revolution. In 2009, the area was of 115 127ha and in 2019 it increased to 228 699ha, which is a variation of 98.65%. This increase may be linked to the inclusion of Pine Nuts in the category of Nuts plantations, which made the total area of Nuts bigger.



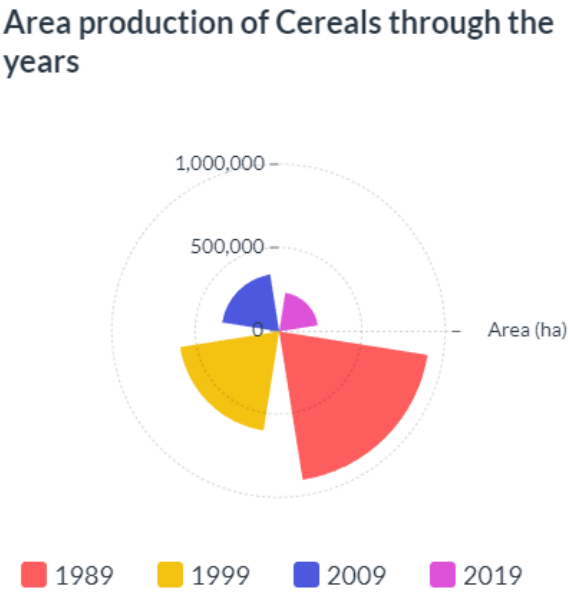
Graph 4: Polar chart representing the area of Nuts plantations over the years, for all *Freguesias* in Portugal.

Also, in the region of Alentejo there were many farmers that ended up changing their production and started having Nuts plantations, due to investment support and having Alqueva providing water.

For the Temporary production, Cereals (Cereals) and Fodder Plants (Fodder) are the biggest crops in Portugal during these 30 years.



Graph 5: Bar chart representing the total area for each Temporary crop, in Portugal (for all *Freguesias*). Each bar represents a year, with the colors according to the legend.



In 1989, the Temporary crops had Cereals as the major area of production, but there was a drastic reduction of 666 003ha in the national territory. It may be connected to the *Campanha do Trigo* (1929-1938) in Portugal, in which there was a massive wheat production all over the national territory, ordered by the government. The amount of production was so big that in 1989, its effects are still observed.

Graph 6: Polar chart representing the area of Cereals plantations over the years, for all *Freguesias* in Portugal.

The variation of the area (in hectares) of the Temporary crops was not so good, from 1989 to 2019, as seen in Table 2. Most of the crops saw their area being reduced and, even the positive variations had a minor increase.

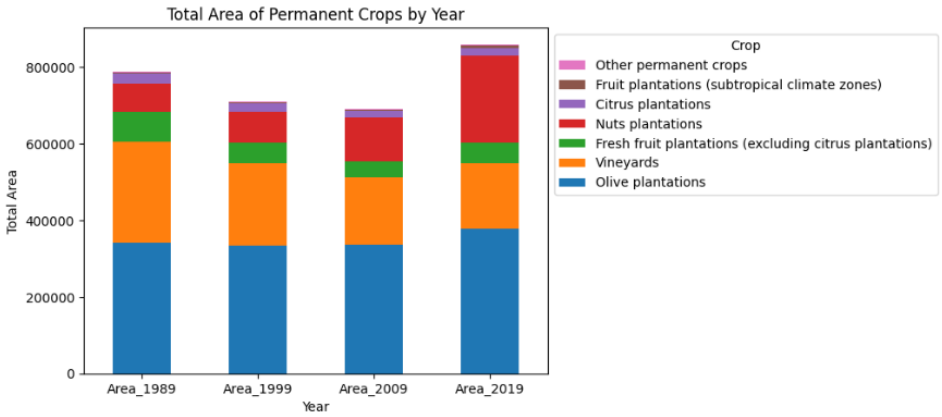
Table 2: Variation of area (ha) for each Temporary crop, from 1989 to 2019.

Crop	Variation (hectares)
Cereals	-666003.00
Dried pulses	-63297.00
Temporary grasses and grazings	46303.00
Fodder plants	-145086.00
Potatoes	-93784.00
Sugarbeets	-256.00
Industrial crops	-53540.00
Fresh vegetables	-10787.00
Flowers and ornamental plants	1287.00
Other temporary crops	-21153.00

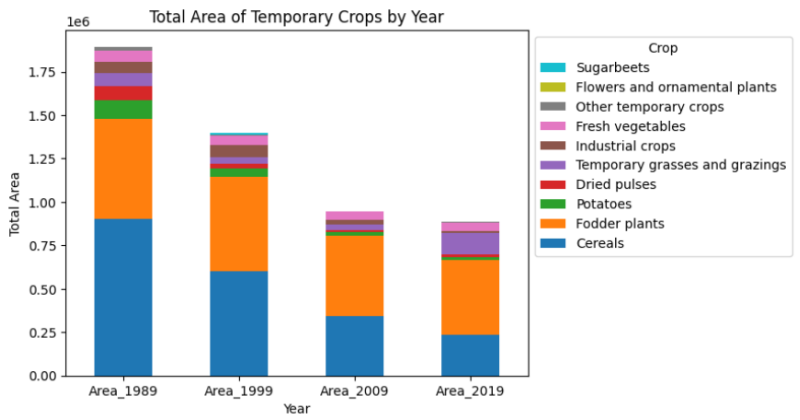
For the Permanent crops (graph 7), the total area for the year of 2019 is the biggest of 30 years. Meanwhile, in graph 8, it is shown a big decline in the cultivation of Temporary crops in 2019, in Portugal. A reduction in the total area of Temporary crops is observed every year, which may indicate that the future doesn't look promising for this type of crop.

The bars of the charts above help visualize the distribution of the area of each crop for every year. In the Permanent crops it is clear that the Nuts plantations had the biggest increase and that the Vineyards are having a decline in their area over the years. So, the Nuts plantations ended up being the reason why the area of the Permanent crops increased in 2019. Whereas, the reduction of the area of Temporary crops is mainly connected to the decrease in Cereals production.

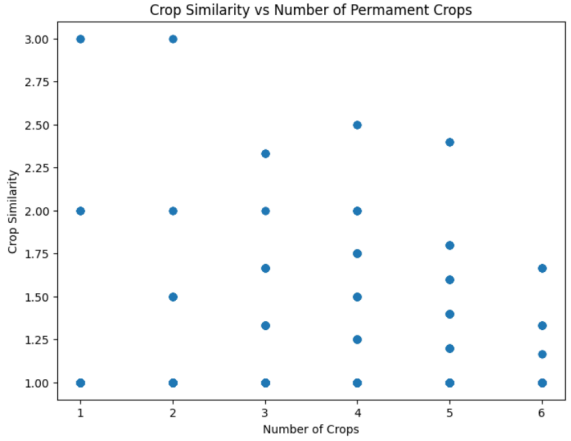
To observe the relationship between the number of crops and the crop similarity for each region in Portugal, a scatter plot was performed for both Temporary and Permanent crops (graph 9 and 10). It is possible to understand the distribution of the crops at the *Freguesia* level. In one axis, we represent the number of crops present in one *Freguesia*, and in the other the division between how many crops present by the number of total crops possible.



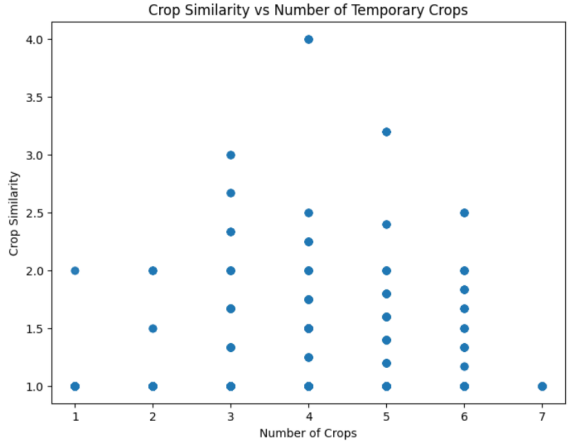
Graph 7: Bar chart representing the distribution of the area of Permanent crops for each year, with all *Freguesias* of Portugal. The units for the total area in the y-axis is hectares.



Graph 8: Bar chart representing the distribution of the area of Temporary crops for each year, with all *Freguesias* of Portugal. The units for the total area in the y-axis is 10^6 hectares.



Graph 9: Scatter plot for the Permanent crop similarity in all *Freguesias*.



Graph 10: Scatter plot for the Temporary crop similarity in all *Freguesias*.

Spatial distribution

We plotted the area for each *Freguesia* with a shapefile and associated DBF file to join the data. This way we are able to visualize the data geographically in a map of Portugal. We then associate the data of each year to each *Freguesia*, resulting in four maps for each type of Permanent and Temporary crop, such as Graph 11 and 12.

When analyzing the patterns that result from each crop, we can better understand the regions of Portugal that have the biggest area.

In the case of Citrus plantations, we can see the biggest areas are found in Algarve, which goes according to common knowledge of the ‘Laranjas do Algarve’, famous worldwide.

Regarding Olive production, the latest data shows the large areas in Alentejo Sul, where the irrigation by Alqueva allowed for greater areas than ever before seen in this dataset.

The Cereals of Portugal have been in a steady decline, and the biggest productions to date are located around Central Alentejo, Ribatejo and Setúbal.

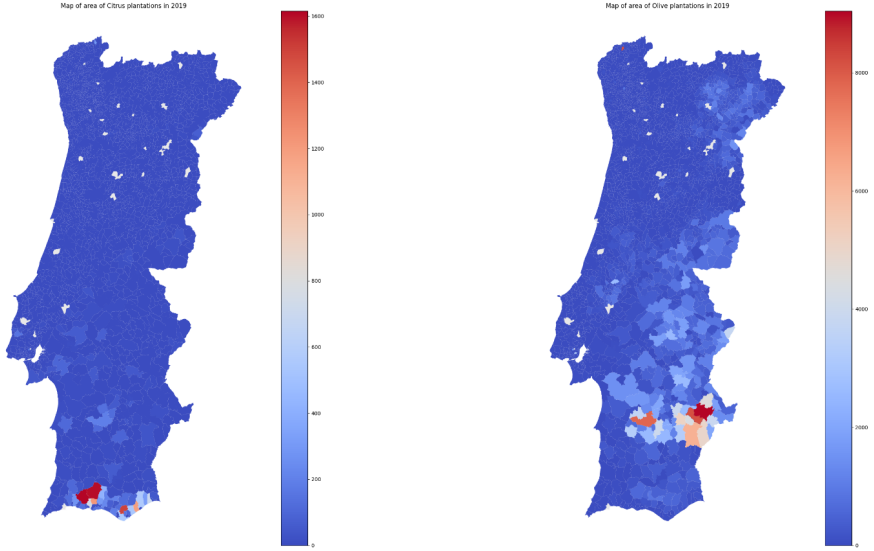
When we look at the Dried Pulses situation, we see that the large majority of area is located in a small latitude band just above the Algarve region.

In the annexes and storytelling presentation, we show the animated maps with all the years cycling and the patterns become evident, by year and by location.

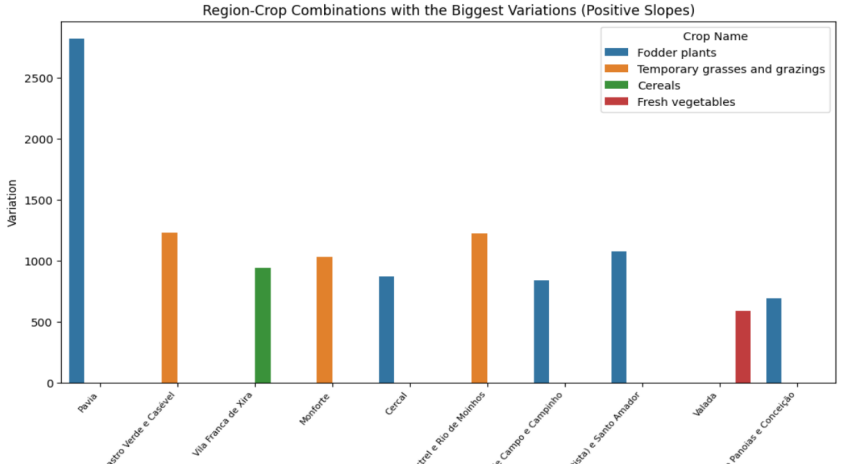
Regarding the variations of area, there are zones with large area variations, and these swings can be positive and negative. We can see those in Graph 13 and 14. The methodology for this was the creation of trendlines for each crop area of each *Freguesia* and separating the slopes into positive, negative and intermediate values. These values are separated at the 1.5 positive and negative values.

The results of this study separate the *Freguesias* into classes and we can see that Pavia is the biggest gain of area of Fodder plants in all of Portugal, with a gain of about 2300ha. The biggest loss of area in Portugal is Castelo Branco, which lost about 900ha of area of Olive plantations. The rest of the top 10 of these areas related to crops are, in Temporary crops gain, Fodder plants, Temporary grasses and grazes, Cereals and Fresh vegetables, and in Permanent crop loss, are Olive plantations, Vineyards and Fresh fruit plantations.

Permanent crops

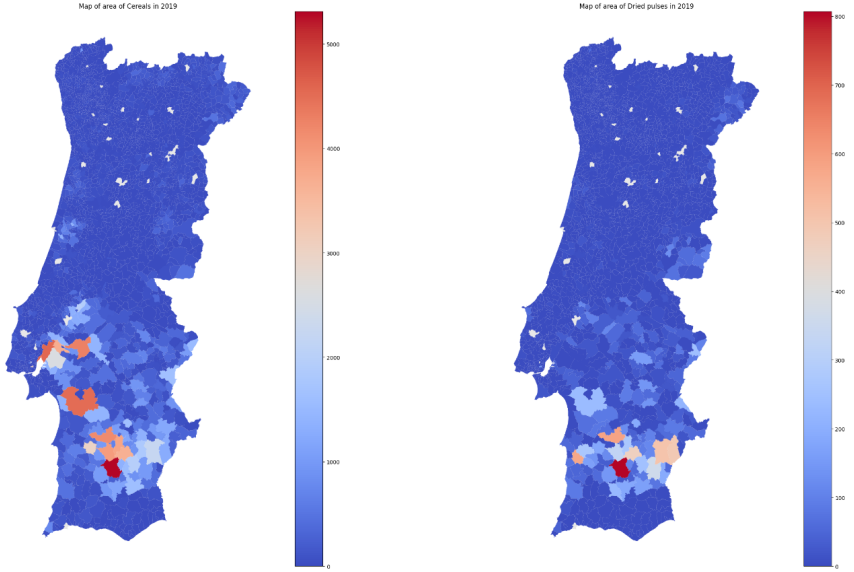


Graph 11 (left) : Map of area of Citrus and Olive plantations in 2019

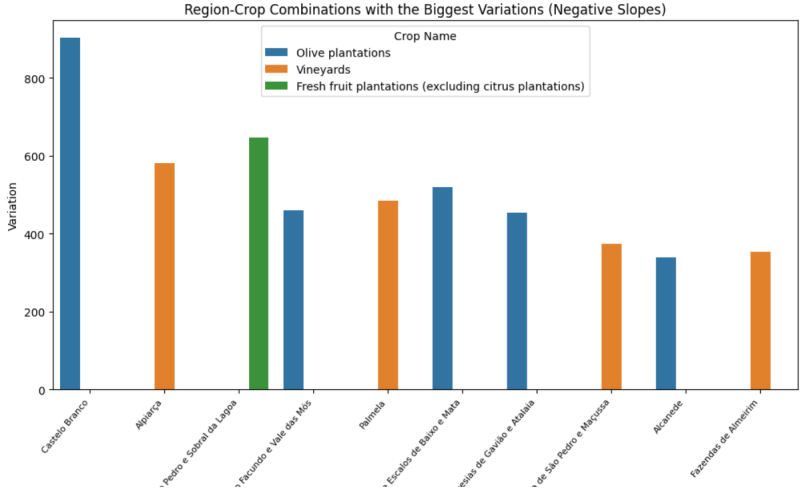


Graph 13 (left) : Region-Crop Temporary Combinations with the biggest Variations (Positive Slopes) Top 10

Temporary crop



Graph 12 (right) : Map of area of Cereals and Dried pulses in 2019



Graph 14 (right) : Region-Crop Permanent Combinations with the biggest Variations (Negative Slopes) Top 10

When analyzing the proportion of area of each crop by the area of each *Freguesia*, we discover that the propositions are very low, about 99% of all *Freguesias* have an area below 1% of crop's proportion. For this study, we did the average of all years of each crop area. The results can be found in Graph 14 and 15.

We observe that Fodder plants, with 16 *Freguesias*, and Cereals, with 12, have by far the biggest areas proportions in *Freguesias*, with one case of Industrial crops. This makes sense because most of these types of productions require large amounts of area to produce, they are traditionally extensive productions.

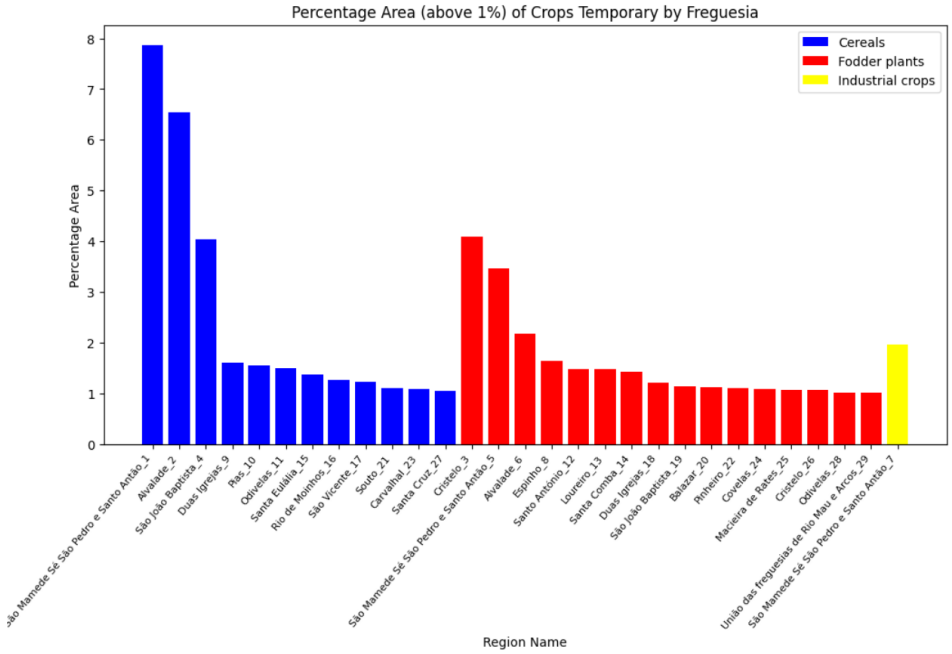
In the case of Permanent crops, we see more diversity, but Olive plantation being the dominant class, with 8 cases, followed by Vineyards, with 4, and after that just two Fresh fruit plantations and one Nut plantation.

These results are the outlines of the grand majority of Portugal's *Freguesias*, with an average of 0.02% of area occupied by crops. This means that these regions are highly dependent on these areas and outputs. The special case of São Mamede, Sé, São Pedro e Santo Antão, in Évora, is an historical example of how much the culture of Évora and Alentejo is connected to agriculture production.

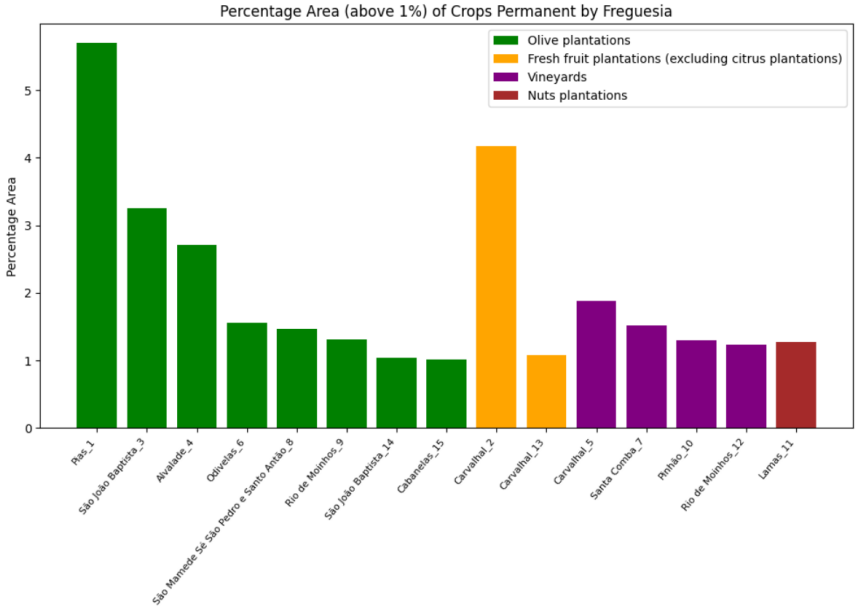
Not all places produce crops, in fact, the production of each crop is normally very clustered around certain regions of Portugal. To show visually and analytically the number of *Freguesias* with the production or each crop, we produced the Graph 16 and 17, that shows each *Freguesia* that has area in a relevant way (by the area being above 30% of the mean area of each crop). These graphs show the number of *Freguesias*, the area and year of each crop's type.

We can see how long the tail is of *Freguesias* and how tall the biggest areas get in comparison to the majority. Also, the fact that its color coded makes it easy to separate the years of each crop area. For example, in the Sugarbeets, we see that 1999 was the biggest year for this crop, and a very short tail. On the Olive plantation side, we see a long tail and the majority of the area registered in 2019.

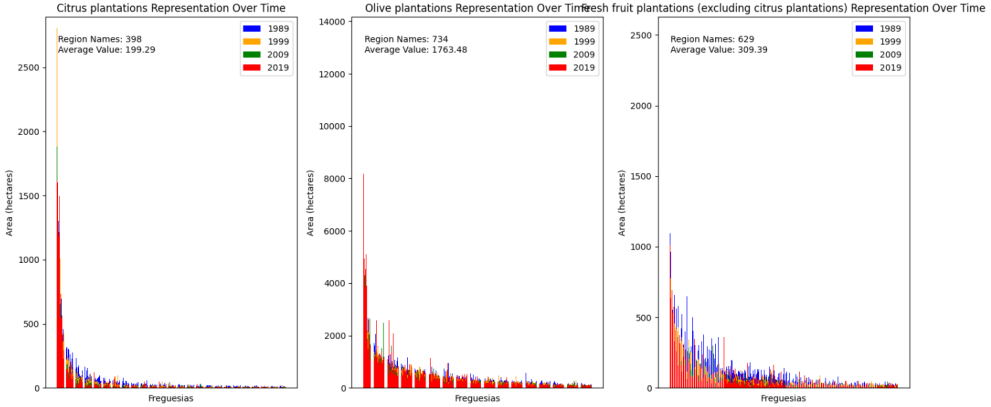
We can see some bugs in the production of these images, due to the large sample set or the small resolution, of the small image size plot, there are these small white gaps in the lines of the *Freguesias* and we acknowledge it could be fixed or presented in another way to fix this visualization bug.



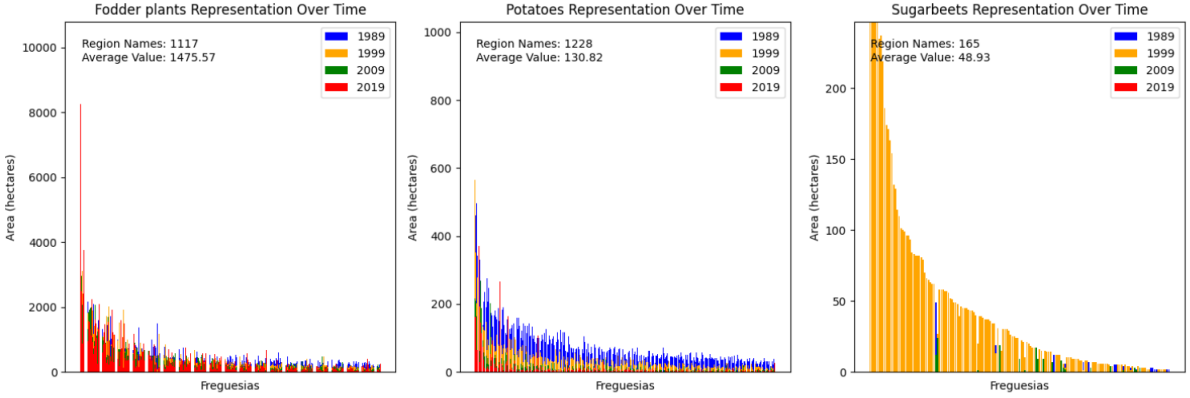
Graph 14 (left) : Percentage Area (above 1%) of Crops Temporary by *Freguesia*



Graph 15 (right) : Percentage Area (above 1%) of Crops Permanent by *Freguesia*



Graph 16 (left): Citrus, olive and fresh fruit plantations representation over time.



Graph 17 (right) :Fodder plants, Potatoes and Sugarbeets representation over time.

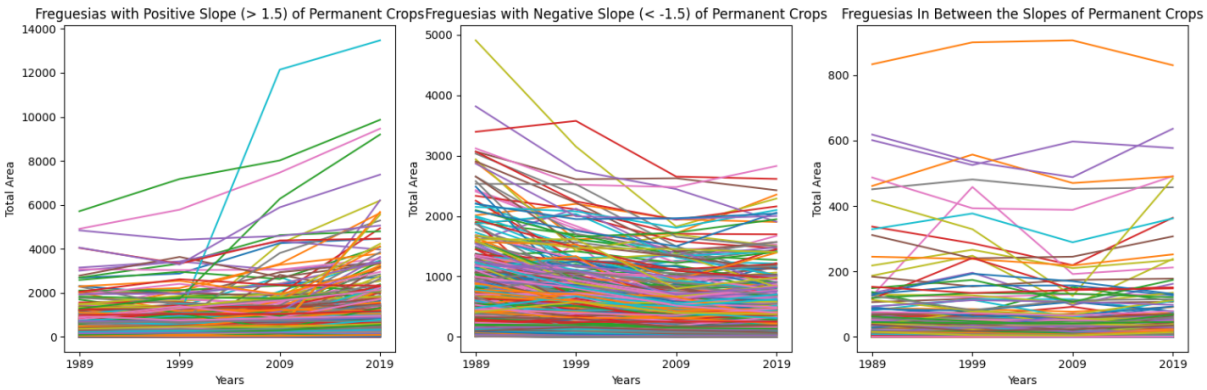
Trends

How do we know the trends in this dataset? We saw earlier in the overall area sums that the Permanent crop overall area is the biggest in 2019 and the Temporary crops have the biggest sum of area in 1989. This could indicate that Permanent crops are on the rise and Temporary crops are on a decline. But is this actually true or just a simplification of the data?

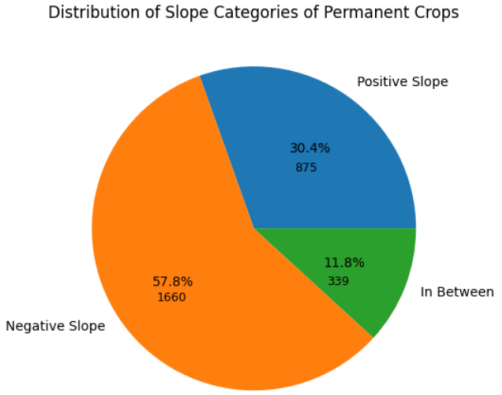
To better understand the trends, we decided to see at the *Freguesia* level each area for each crop and compare them all. We started by plotting the area evolution overtime by each crop, and then drew a trendline for each of these to better determine what the trend is. We chose a linear trend line for this analysis. Then we decided on a value to separate the trendlines, in this case it was positive and above 1.5, or negative and below -1.5, or in between these two values, and divided them into three charts and counted how many are in each class.

This resulted in the Graph 18 and 20, where we can see all the *Freguesias* divided into each class of positive, negative or neutral growth. At this stage, this is purely representative and no real information can be derived from these charts, as they have too much noise in the data.

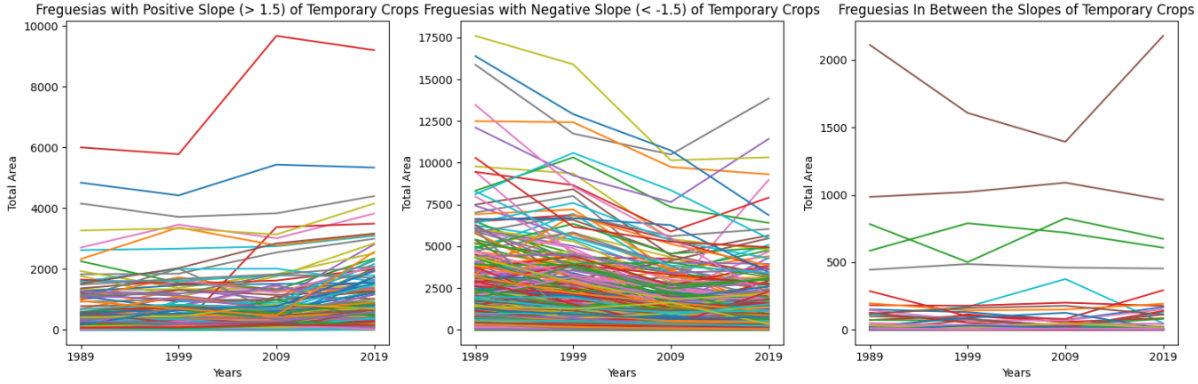
To better see how many *Freguesias* we have in each class, we count the values and display those in Graph 19 and 21. This paints a clear picture of how many *Freguesias* are in a positive or negative area trend. In both Permanent and Temporary crops, the majority class is the negative growth. The second largest class in both is the positive growth. Lastly, we have neutral growth in both crop types. So this is a much more clear indication of the overall trend, where we now see the trend in the area of crops is largely negative, and raises questions about the future of agriculture in Portugal. The Permanent crop trend is less worrying, but with 57.6% of all *Freguesias* with negative area evolution, the decrease is still to be reflected in future surveys. The trend in Temporary crops is very noticeable in the overall area sum, but with 89.5% *Freguesias* on the area decline, we can also expect a certain future decrease in crop area.



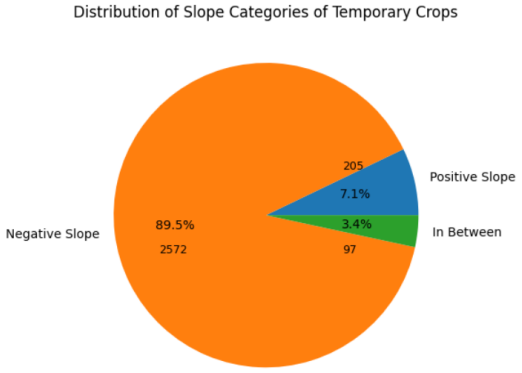
Graph 18 : *Freguesias* with positive slope(left), negative slopes(center) and neutral slopes (right) of Permanent crops.



Graph 19 (right): Distribution of slope categories of Permanent crops.



Graph 20 : *Freguesias* with positive slope(left), negative slopes(center) and neutral slopes (right) of Temporary crops.



Graph 21 (right): Distribution of slope categories of Temporary crops.

Inferences and relations

Correlation

Freguesias have their number of crops individually, but are there relations between the combinations of crops in each *Freguesia*? To know the correlation between the crops, we plotted a heatmap chart to visualize the connections in each of the Permanent and Temporary crops, as we can see in Graph 22 and 23.

We can see the table of Permanent crops is dominated by blue values, meaning each crop is exclusionary of each other, meaning that the area of one crop is not the same as the other, the only value close to 0.5 is Citrus plantations and Fruit plantation (subtropical climate). About Temporary crops, there are greater positive relations between each other, with one value about 0.7, of Cereals and Industrial crops, meaning that if a *Freguesia* has Cereals, it's probable to also find Industrial crops.

The list of correlations ordered can be found in Table 3 and 4, to better understand values that are correlated and not.

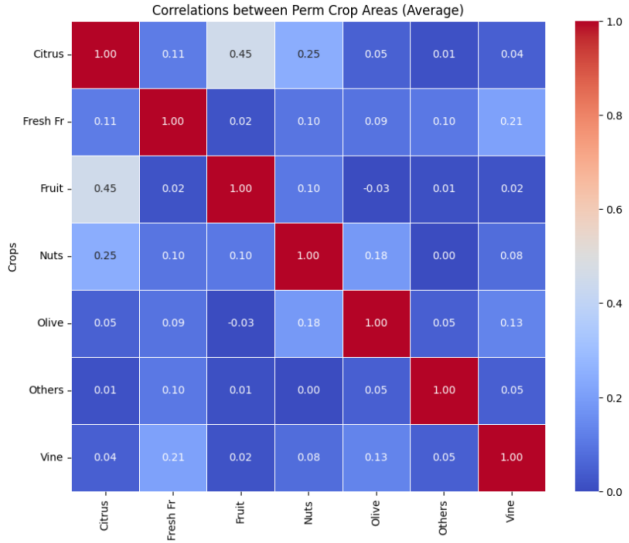
PCA

A PCA analysis was conducted on the selected crop area features ('Area_1989', 'Area_1999', 'Area_2009', 'Area_2019'). Two principal components, PC1 and PC2, were obtained, representing new axes that capture the most significant patterns and variability in the dataset. These components offer insights into the relationships between regions and different Permanent crops. The scatter plot visualization of PC1 and PC2 helps understand the distribution of data points and identify similarities and differences among regions based on their agricultural practices.

As we can see in Graph 24 and 25, the values are clustered in a triangle-like shape, with a dense starting point, with lots of values close to the origin and a more dispersed set of values further away from the origin. Also, both the other crops from Permanent and Temporary are the closest to the origin.

The Permanent crops are all closely related, except for the classes Nuts and Olives plantations, probably for having bigger overall areas.

The Temporary crops are also very closely related, but Cereals and Fodder plants are more distante to the origin, probably for also having larger areas of plantations.



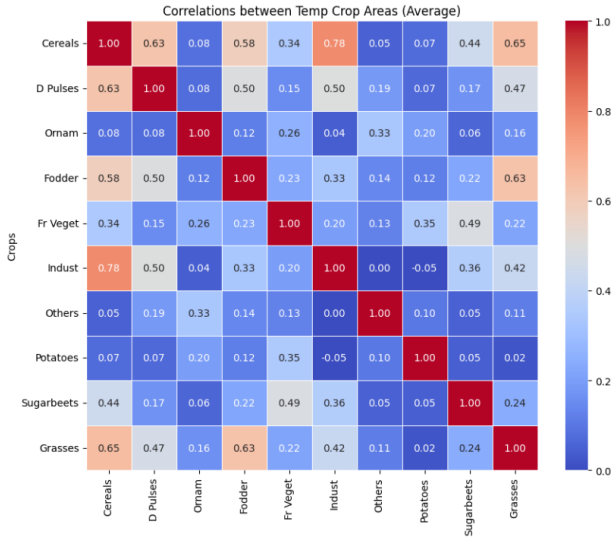


Table 3: Permanent Crop Correlation values ordered by biggest to smallest.

Perm Crop 1	Perm Crop 2	Correlation
Citrus plantations	Fruit plantations (subtropical climate zones)	0.452
Citrus plantations	Nuts plantations	0.246
Fresh fruit plantations (excluding citrus plantations)	Vineyards	0.208
Nuts plantations	Olive plantations	0.184
Vineyards	Olive plantations	0.129
Citrus plantations	Fresh fruit plantations (excluding citrus plantations)	0.112
Fresh fruit plantations (excluding citrus plantations)	Other permanent crops	0.103
Fruit plantations (subtropical climate zones)	Nuts plantations	0.102
Fresh fruit plantations (excluding citrus plantations)	Nuts plantations	0.096
Olive plantations	Fresh fruit plantations (excluding citrus plantations)	0.093
Vineyards	Nuts plantations	0.077
Citrus plantations	Olive plantations	0.051
Other permanent crops	Vineyards	0.046
Olive plantations	Other permanent crops	0.045
Citrus plantations	Vineyards	0.044
Fresh fruit plantations (excluding citrus plantations)	Fruit plantations (subtropical climate zones)	0.018
Fruit plantations (subtropical climate zones)	Vineyards	0.017
Other permanent crops	Citrus plantations	0.015
Other permanent crops	Fruit plantations (subtropical climate zones)	0.009
Nuts plantations	Other permanent crops	0.003
Fruit plantations (subtropical climate zones)	Olive plantations	-0.035

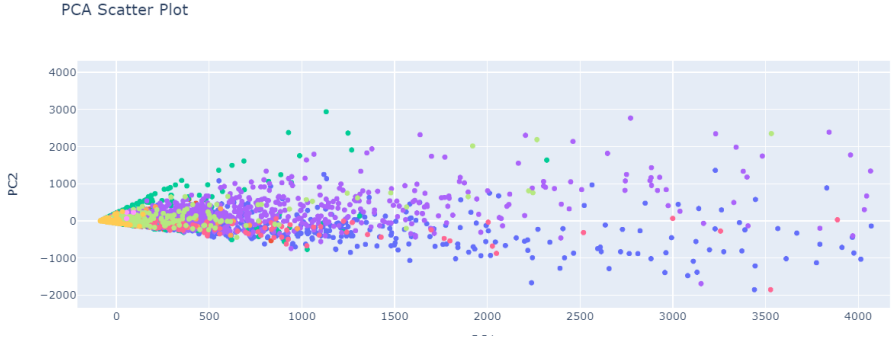


Table 4: Temporary Crop Correlation values ordered by biggest to smallest.

Temp Crop 1	Temp Crop 2	Correlation
Industrial crops	Cereals	0.784
Cereals	Temporary grasses and grazings	0.651
Temporary grasses and grazings	Fodder plants	0.632
Dried pulses	Cereals	0.627
Cereals	Fodder plants	0.583
Industrial crops	Dried pulses	0.498
Dried pulses	Fodder plants	0.496
Sugarbeets	Fresh vegetables	0.488
Temporary grasses and grazings	Dried pulses	0.465
Cereals	Sugarbeets	0.444
Industrial crops	Temporary grasses and grazings	0.422
Sugarbeets	Industrial crops	0.360
Potatoes	Fresh vegetables	0.346
Fresh vegetables	Cereals	0.342
Industrial crops	Fodder plants	0.334
Flowers and ornamental plants	Other temporary crops	0.332
Flowers and ornamental plants	Fresh vegetables	0.258
Sugarbeets	Temporary grasses and grazings	0.241
Fresh vegetables	Fodder plants	0.226
Fodder plants	Sugarbeets	0.221
Fresh vegetables	Temporary grasses and grazings	0.217
Industrial crops	Fresh vegetables	0.198
Potatoes	Flowers and ornamental plants	0.196
Other temporary crops	Dried pulses	0.185
Sugarbeets	Dried pulses	0.170
Flowers and ornamental plants	Temporary grasses and grazings	0.162
Dried pulses	Fresh vegetables	0.150
Fodder plants	Other temporary crops	0.140
Fresh vegetables	Other temporary crops	0.125
Fodder plants	Flowers and ornamental plants	0.119
Fodder plants	Potatoes	0.118
Temporary grasses and grazings	Other temporary crops	0.115
Potatoes	Other temporary crops	0.102
Cereals	Flowers and ornamental plants	0.082
Flowers and ornamental plants	Dried pulses	0.077
Potatoes	Cereals	0.072
Potatoes	Dried pulses	0.069
Sugarbeets	Flowers and ornamental plants	0.059
Other temporary crops	Sugarbeets	0.050
Cereals	Other temporary crops	0.049
Sugarbeets	Potatoes	0.047
Flowers and ornamental plants	Industrial crops	0.041
Temporary grasses and grazings	Potatoes	0.020
Industrial crops	Other temporary crops	0.002
Industrial crops	Potatoes	-0.051

Graph 22 (left): Correlations between perm crop areas (crop).

Graph 23 (right): Correlations between temp crop areas (crop).

Graph 24 (left): PCA Scatter Plot of Temporary crops.

Graph 25 (right): PCA Scatter Plot of Permanent crops.

Discussion

As we have shown, the data can be messy, but we have organized it in a way that tries to paint a clear picture of what has happened in Portugal regarding the agriculture explorations and trends in crops.

We have shown only about half of the graphs and statistical analysis produced, we only displayed the most relevant for the report. The full extent of the exploration of the data can be found in the Notebooks in the Github repository.

We have shown that there are relations found between crops and correlations with some very positive values. Also the trends shown indicate that it's possible that the area of crops in Portugal is decreasing. It's important to note that this does not indicate commercial output of quality or quality of production. The yields in agriculture have been increasing in Portugal and it's possible that the same production is being achieved with less area, but this is speculation on our part.

There could be further testing done with this dataset, such as the correlation between Permanent and Temporary crops, the statistical probability of the hypothesis of the decrease of the crops of Portugal, we analyzed 30 years of data, what could be the area in another 30 years? In a more pattern related study, the use of Markov chains could be used to determine where the areas of each crop could be located.

Conclusions

This report has provided valuable insights into the topic of areas of crops in Portugal. Through the analysis of data on the area of crops, several key topics have been explored, including summary statistics, spatial distribution, trends, and the significance of crop distribution.

Regarding summary statistics, the report has presented data on the area of Temporary and Permanent crops in Portugal, allowing for a better understanding of how crop distribution has

changed over time. The overall area of crops decreased, except for Olive Plantations and Nuts plantations.

The spatial distribution analysis has shed light on the regions in Portugal and their respective crop production areas. By examining the *Freguesia* level data, trends in crop areas have been identified and represented through trendlines. However, it is important to note that the noise in the data limits the derivation of concrete information from these charts.

In conclusion, this report highlights the importance of understanding the distribution of crop areas in Portugal. It emphasizes the need for comprehensive data analysis techniques, such as summary statistics, spatial distribution analysis, and PCA, to gain valuable insights into agricultural practices. Further research and analysis are recommended to delve deeper into the potential correlations within the dataset.

References

1. "Olival português em transformação estrutural" - Agriterra. Available at: https://www.agriterra.pt/Artigos/387751-Olival-portugues-em-transformacao-estrutural.html (Accessed: 13 of June 2023)

2. "greends-dms-exercises" - GitHub Repository. Available at: https://github.com/isa-ulisboa/greends-dms-exercises/blob/main/data/dms_INE_v2.sql.zip (Accessed: 21 of May 2023)

3. "SPSS Modeler Documentation: Writing a Description Report" - IBM. Available at: https://www.ibm.com/docs/ro/spss-modeler/18.0.0?topic=data-writing-description-report (Accessed: 13 of June 2023)

4. "Freguesias de Portugal" - Dados.gov.pt. Available at: https://dados.gov.pt/pt/datasets/freguesias-de-portugal/ (Accessed: 3 of June 2023)

5. "Campanha do Trigo" - Wikipedia. Available at: https://pt.wikipedia.org/wiki/Campanha_do_Trigo (Accessed: 15 of June 2023)

6. "Portugal perto da autossuficiência em frutos secos" - Dinheiro Vivo. Available at: https://www.dinheirovivo.pt/economia/nacional/portugal-perto-da-autossuficiencia-em-frutos-secos-14817426.html (Accessed: 14 of June 2023)

7. "Caracterização e Evolução das Culturas Permanentes da Draplvt Entre os

Recenseamentos Agrícolas de 1989 e 2009" - DRAPLVT. Available at: http://www.draplvt.mamaot.pt/DRAPLVT/CaracterizacaoInfoTecnica/Documents/20120322_dt2.pdf (Accessed:13 of June 2023)

8. "Principal Component Analysis (PCA) Explained Visually with Zero Math" - Towards Data Science. Available at: https://towardsdatascience.com/principal-component-analysis-pca-explained-visually-with-zero-math-1cbf392b9e7d#0226 (Accessed: 19 of June 2023)

Annex 1 - SQL script

The following SQL script was created to obtain data for the Temporary crops for the year of 1989. In order to have data for a different year, it must be substituted for the desired year.

```
show databases;
-- selecionar base de dados
use dms_INE_v2;
-- ver tabela que se quer processar
select * from Temporary_crop tc
   inner join Temporary_crop_name tcn on tc.tc_name_ID = tcn.tc_name_ID
   inner join region r on tc.NutsID = r.NutsID
where tc.year = 1989
and tcn.crop_name not like 'Total' and r.level_ID = 5
order by tc.NutsID;
-- seleccionar colunas
select tc.NutsID, tc.`year`, tc.`area`, tcn.crop_name, r.region_name from
Temporary_crop tc
   inner join Temporary_crop_name tcn on tc.tc_name_ID = tcn.tc_name_ID
   inner join region r on tc.NutsID = r.NutsID
where tc.year = 1989
and tcn.crop_name not like 'Total' and r.level_ID = 5
order by tc.NutsID;
-- apagar tabela se já existir
drop table if exists process_temp_crops;
-- criar tabela temporária para processar
create table dms_ine_v2.process_temp_crops select tc.NutsID, tc.`year`,
tc.`area`, tcn.crop_name, r.region_name from Temporary_crop tc
   inner join Temporary_crop_name tcn on tc.tc_name_ID = tcn.tc_name_ID
   inner join region r on tc.NutsID = r.NutsID
where tc.year = 1989
and tcn.crop_name not like 'Total' and r.level_ID = 5
order by tc.NutsID;
-- ver tabela
select * from process_temp_crops;
-- ver todas as colunas da tabela pivot, a ser usado em CASE
select distinct  crop_name from process_temp_crops;
-- criar tabela pivot (falta acrescentar as outras culturas)
select NutsID, `year`,
       max(CASE WHEN crop_name = 'Cereals' then `area` END) AS `Cereals`,
        max(CASE WHEN crop_name = 'Dried pulses' then `area` END) AS `Dried
pulses`,
          max(CASE WHEN crop_name = 'Temporary grasses and grazings' then
`area` END) AS `Temporary grasses and grazings`,
       max(CASE WHEN crop_name = 'Fodder plants' then `area` END) AS `Fodder
plants`,
       max(CASE WHEN crop_name = 'Potatoes' then `area` END) AS `Potatoes`,
```

```
            max(CASE WHEN crop_name = 'Sugarbeets' then `area` END) AS
`Sugarbeets`,
          max(CASE WHEN crop_name = 'Industrial crops' then `area` END) AS
`Industrial crops`,
          max(CASE WHEN crop_name = 'Fresh vegetables' then `area` END) AS
`Fresh vegetables`,
        max(CASE WHEN crop_name = 'Flowers and ornamental plants' then `area`
END) AS `Flowers and ornamental plants`,
        max(CASE WHEN crop_name = 'Other Temporary crops' then `area` END) AS
`Other Temporary crops`,
        region_name
FROM process_temp_crops
GROUP BY NutsID;
```

The following SQL script was created to obtain data for the Permanent crops for the year of 1989. In order to have data for a different year, it must be substituted for the desired year.

```
show databases;

-- selecionar base de dados
use dms_INE_v2;
-- ver tabela que se quer processar
select * from Permanent_crop pc
   inner join Permanent_crop_name pcn on pc.pc_name_ID = pcn.pc_name_ID
   inner join region r on pc.NutsID = r.NutsID
where pc.year = 1989
      and pcn.crop_name not like 'Total' and r.level_ID = 5
order by pc.NutsID ;
-- seleccionar colunas
select pc.NutsID, pc.`year`, pc.`area`, pcn.crop_name, r.region_name from
Permanent_crop pc
   inner join Permanent_crop_name pcn on pc.pc_name_ID = pcn.pc_name_ID
   inner join region r on pc.NutsID = r.NutsID
where pc.year = 1989
      and pcn.crop_name not like 'Total' and r.level_ID = 5
order by pc.NutsID;
-- apagar tabela se já existir
drop table if exists process_perm_crops;
-- criar tabela temporária para processar
create table dms_ine_v2.process_perm_crops select pc.NutsID, pc.`year`,
pc.`area`, pcn.crop_name, r.region_name from Permanent_crop pc
   inner join Permanent_crop_name pcn on pc.pc_name_ID = pcn.pc_name_ID
   inner join region r on pc.NutsID = r.NutsID
where pc.year = 1989
      and pcn.crop_name not like 'Total' and r.level_ID = 5
order by pc.NutsID ;
-- ver tabela
select * from process_perm_crops;
-- ver todas as colunas da tabela pivot, a ser usado em CASE
select distinct  crop_name from process_perm_crops;
```

```sql
-- criar tabela pivot (falta acrescentar as outras culturas)
select NutsID, `year`,
        max(CASE WHEN crop_name = 'Citrus plantations' then `area` END) AS
`Citrus plantations`,
         max(CASE WHEN crop_name = 'Olive plantations' then `area` END) AS
`Olive plantations`,
       max(CASE WHEN crop_name = 'Fresh fruit plantations (excluding citrus
plantations)' then `area` END) AS `Fresh fruit plantations (excluding
citrus plantations)`,
         max(CASE WHEN crop_name = 'Fruit plantations (subtropical climate
zones)' then `area` END) AS `Fruit plantations (subtropical climate
zones)`,
         max(CASE WHEN crop_name = 'Nuts plantations' then `area` END) AS
`Nuts plantations`,
            max(CASE WHEN crop_name = 'Vineyards' then `area` END) AS
`Vineyards`,
        max(CASE WHEN crop_name = 'Other Permanent crops' then `area` END)
AS `Other Permanent crops`,
       region_name
FROM process_perm_crops
GROUP BY NutsID;
```

Annex 2 - OpenRefine

This is the operation history as JSON, performed in OpenRefine for the Temporary crops to join all the years and to create a primary key.

```
[
  {
    "op": "core/column-reorder",
    "columnNames": [
      "NutsID",
      "region_name",
      "year",
      "Cereals",
      "Dried pulses",
      "Temporary grasses and grazings",
      "Fodder plants",
      "Potatoes",
      "Sugarbeets",
      "Industrial crops",
      "Fresh vegetables",
      "Flowers and ornamental plants",
      "Other Temporary crops"
    ],
    "description": "Reorder columns"
  },
  {
    "op": "core/transpose-columns-into-rows",
```

```
    "startColumnName": "Cereals",
    "columnCount": -1,
    "ignoreBlankCells": true,
    "fillDown": false,
    "separator": null,
    "keyColumnName": "Temp_crop",
    "valueColumnName": "Area_1989",
     "description": "Transpose cells in columns starting with Cereals into
rows in two new columns named Temp_crop and Area_1989"
  },
  {
    "op": "core/column-removal",
    "columnName": "year",
    "description": "Remove column year"
  },
  {
    "op": "core/fill-down",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "NutsID",
    "description": "Fill down cells in column NutsID"
  },
  {
    "op": "core/fill-down",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "region_name",
    "description": "Fill down cells in column region_name"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "NutsID",
     "expression": "grel:value + '_' + cells[\"region_name\"].value + '_' +
cells[\"Temp_crop\"].value",
    "onError": "set-to-blank",
    "newColumnName": "Nuts_region_temp",
    "columnInsertIndex": 1,
      "description": "Create column Nuts_region_temp at index 1 based on
column NutsID using expression grel:value + '_' +
cells[\"region_name\"].value + '_' + cells[\"Temp_crop\"].value"
  },
  {
    "op": "core/column-reorder",
    "columnNames": [
      "Nuts_region_temp",
```

```
      "NutsID",
      "region_name",
      "Temp_crop",
      "Area_1989"
    ],
    "description": "Reorder columns"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "Nuts_region_temp",
              "expression":   "grel:cell.cross('2.1   temp   1999   csv',
'Nuts_region_temp').cells['Area_1999'].value[0]",
    "onError": "set-to-blank",
    "newColumnName": "Area_1999",
    "columnInsertIndex": 1,
      "description": "Create column Area_1999 at index 1 based on column
Nuts_region_temp using expression grel:cell.cross('2.1 temp 1999 csv',
'Nuts_region_temp').cells['Area_1999'].value[0]"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "Nuts_region_temp",
              "expression":   "grel:cell.cross('2.1   temp   2009   csv',
'Nuts_region_temp').cells['Area_2009'].value[0]",
    "onError": "set-to-blank",
    "newColumnName": "Area_2009",
    "columnInsertIndex": 1,
      "description": "Create column Area_2009 at index 1 based on column
Nuts_region_temp using expression grel:cell.cross('2.1 temp 2009 csv',
'Nuts_region_temp').cells['Area_2009'].value[0]"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "Nuts_region_temp",
              "expression":   "grel:cell.cross('2.1   temp   2019   csv',
'Nuts_region_temp').cells['Area_2019'].value[0]",
    "onError": "set-to-blank",
    "newColumnName": "Area_2019",
    "columnInsertIndex": 1,
```

```
      "description": "Create column Area_2019 at index 1 based on column Nuts_region_temp using expression grel:cell.cross('2.1 temp 2019 csv', 'Nuts_region_temp').cells['Area_2019'].value[0]"
  },
  {
    "op": "core/column-reorder",
    "columnNames": [
      "Nuts_region_temp",
      "NutsID",
      "region_name",
      "Temp_crop",
      "Area_1989",
      "Area_1999",
      "Area_2009",
      "Area_2019"
    ],
    "description": "Reorder columns"
  }
]
```


This is the operation history as JSON, performed in OpenRefine for the Permanent crops to join all the years and to create a primary key.


```
[
  {
    "op": "core/column-reorder",
    "columnNames": [
      "NutsID",
      "region_name",
      "year",
      "Citrus plantations",
      "Olive plantations",
      "Fresh fruit plantations (excluding citrus plantations)",
      "Fruit plantations (subtropical climate zones)",
      "Nuts plantations",
      "Vineyards",
      "Other Permanent crops"
    ],
    "description": "Reorder columns"
  },
  {
    "op": "core/transpose-columns-into-rows",
    "startColumnName": "Citrus plantations",
    "columnCount": -1,
    "ignoreBlankCells": true,
    "fillDown": false,
    "separator": null,
    "keyColumnName": "Perm_crop",
    "valueColumnName": "Area_1989",
       "description": "Transpose cells in columns starting with Citrus plantations into rows in two new columns named Perm_crop and Area_1989"
  },
```

```
  {
    "op": "core/column-removal",
    "columnName": "year",
    "description": "Remove column year"
  },
  {
    "op": "core/fill-down",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "NutsID",
    "description": "Fill down cells in column NutsID"
  },
  {
    "op": "core/fill-down",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "columnName": "region_name",
    "description": "Fill down cells in column region_name"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "NutsID",
     "expression": "grel:value + '_' + cells[\"region_name\"].value + '_' +
cells[\"Perm_crop\"].value",
    "onError": "set-to-blank",
    "newColumnName": "Nuts_region_perm",
    "columnInsertIndex": 1,
      "description": "Create column Nuts_region_perm at index 1 based on
column NutsID using expression grel:value + '_' +
cells[\"region_name\"].value + '_' + cells[\"Perm_crop\"].value"
  },
  {
    "op": "core/column-reorder",
    "columnNames": [
      "Nuts_region_perm",
      "NutsID",
      "region_name",
      "Perm_crop",
      "Area_1989"
    ],
    "description": "Reorder columns"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
```

```
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "Nuts_region_perm",
              "expression":   "grel:cell.cross('2.1   perm   1999   csv',
'Nuts_region_perm').cells['Area_1999'].value[0]",
    "onError": "set-to-blank",
    "newColumnName": "Area_1999",
    "columnInsertIndex": 1,
      "description": "Create column Area_1999 at index 1 based on column
Nuts_region_perm using expression grel:cell.cross('2.1 perm 1999 csv',
'Nuts_region_perm').cells['Area_1999'].value[0]"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "Nuts_region_perm",
              "expression":   "grel:cell.cross('2.1   perm   2009   csv',
'Nuts_region_perm').cells['Area_2009'].value[0]",
    "onError": "set-to-blank",
    "newColumnName": "Area_2009",
    "columnInsertIndex": 1,
      "description": "Create column Area_2009 at index 1 based on column
Nuts_region_perm using expression grel:cell.cross('2.1 perm 2009 csv',
'Nuts_region_perm').cells['Area_2009'].value[0]"
  },
  {
    "op": "core/column-addition",
    "engineConfig": {
      "facets": [],
      "mode": "row-based"
    },
    "baseColumnName": "Nuts_region_perm",
              "expression":   "grel:cell.cross('2.1   perm   2019   csv',
'Nuts_region_perm').cells['Area_2019'].value[0]",
    "onError": "set-to-blank",
    "newColumnName": "Area_2019",
    "columnInsertIndex": 1,
      "description": "Create column Area_2019 at index 1 based on column
Nuts_region_perm using expression grel:cell.cross('2.1 perm 2019 csv',
'Nuts_region_perm').cells['Area_2019'].value[0]"
  },
  {
    "op": "core/column-reorder",
    "columnNames": [
      "Nuts_region_perm",
      "NutsID",
      "region_name",
      "Perm_crop",
      "Area_1989",
```

```
      "Area_1999",
      "Area_2009",
      "Area_2019"
    ],
    "description": "Reorder columns"
  }
]
```

Annex 3 - Python code

All the code is in Jupyter Notebooks and its available at the following repository:

https://github.com/jfilipa1/AVCD_JE_VF.git

This contains all the images produced with the code for this report.