

Soutenance du 5 mai 2025

RAPPORT PROJET ANALYSE DE DONNÉES

Le climat va-t-il ruiner ton apéro ? - L'impact du
changement climatique sur la production de vin
et de bière dans le monde



Thomas BOTTALICO

Frédéric JALAGUIER

Lucie LE SAUX

Mathieu NIVEAU

Adrien RIVET

Introduction

Plan du rapport :

Introduction	2
• 1. Choix des variables et individus	3
• 2. Collecte des données	5
• 3. Premières analyses	6
• 4. Réalisation d'une ACP	8
• 5. Analyse inférentielle	13
Conclusion	15

Introduction

Dans un monde où les effets du changement climatique se font de plus en plus ressentir, nous avons eu l'idée d'étudier l'impact de ce phénomène global sur deux produits emblématiques de la culture alimentaire mondiale : le vin et la bière. Le titre de notre projet, volontairement accrocheur, interroge la durabilité de l'apéritif préféré des ingénieurs face aux bouleversements climatiques que nous traversons.

Notre étude a pour objectif de mettre en évidence les relations entre différentes variables climatiques (température, précipitations, sécheresse...) et des indicateurs agricoles et économiques d'alcool (production, prix, taux d'alcool, consommation...). Pour cela, nous avons utilisé l'entièreté des méthodes statistiques vues en cours d'analyse de données, allant de la statistique descriptive à l'analyse en composantes principales (ACP) et à la statistique inférentielle.

1. Choix des variables et individus

A) Dimension temporelle

Nous avons fait le choix d'utiliser les données sur les années 2019 à 2024, pour avoir les données les plus récentes disponibles sans avoir à rassembler des quantités astronomiques de données.

B) Choix des variables

Nous avons sélectionné 15 variables quantitatives jugées pertinentes pour analyser l'impact du climat sur la production de vin et de bière, réparties en trois grandes catégories :

Les variables climatiques : température moyenne (en °C), précipitations (en mm), jours de canicule, pourcentage de sécheresse.

Les variables agricoles : production de vin (L/ha), superficie de vignes (ha), production de bière (L/ha), superficie de houblon (ha).

Les variables économiques et variables de qualité : prix du vin et de la bière (en €), taux d'alcool (en %), variations de production (en %), consommation annuelle (en L/an).

Les critères de sélection se concentrent sur des données ayant un lien direct avec notre sujet (changement climatique, production agricole, impact économique)

Nous avons volontairement exclu les variables macroéconomiques trop éloignées du sujet ou les données trop techniques qu'il est difficile de trouver à l'échelle internationale.

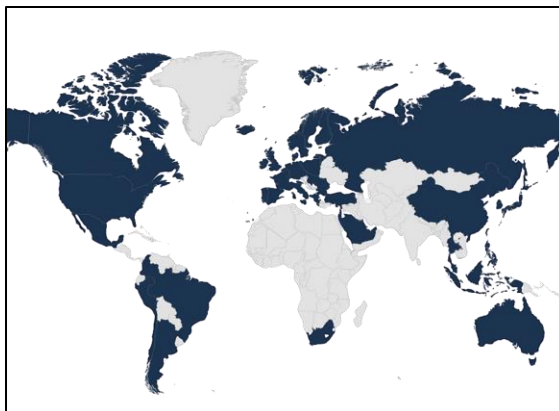
C) Individus

Nous avons choisi 40 pays producteurs et/ou consommateurs majeurs de vin et de bière, répartis sur l'ensemble des continents. Bien évidemment, les pays sélectionnés avaient des données accessibles et fiables.

Certains individus ont dû être écarté : les petits états ou les pays sans culture vinicole/céréalière significative

Ce choix permet de garantir une représentation globale et cohérente de notre échantillon.

Notre jeu de données couvre ainsi plus de 70 % de la production mondiale de vin et plus de 60 % de celle de la bière.



Carte du monde avec les
pays sélectionnés
(équilibre géographique)

1. Choix des variables et individus

	Pays	Température (°C)	Précipitations (mm)	Jours de canicule	Sécheresse (%)	Production Vin (L/ha)	Superficie Vigne (ha)	Production Bière (L/ha)	Superficie Houblon (ha)	Prix Vin (€)	Prix Bière (€)	Taux Alcool Vin
1	France	20.0	291.0	35	11.7	3347.0	186306.0	3099.0	6677.0	10.92	4.95	13.3
2	Italie	6.6	626.0	4	35.6	6126.0	179209.0	2715.0	13916.0	13.0	4.02	11.0
3	Espagne	10.9	1296.0	30	25.8	5405.0	144397.0	4426.0	15636.0	8.52	1.3	14.0
4	États-Unis	10.5	527.0	7	5.0	5746.0	123021.0	3114.0	10157.0	8.17	3.42	13.4
5	Australie	18.4	227.0	0	36.2	4418.0	94837.0	2800.0	9825.0	9.39	1.67	11.3
6	Argentine	16.1	587.0	28	23.7	6046.0	86759.0	1958.0	8272.0	10.36	2.84	12.7
7	Chili	22.9	406.0	30	12.2	4253.0	140153.0	3110.0	15193.0	11.25	3.93	14.2
8	Afrique du Sud	20.9	767.0	26	26.4	4443.0	10758.0	929.0	16279.0	11.54	4.27	14.4
9	Allemagne	21.6	735.0	38	37.9	5872.0	128267.0	2123.0	9863.0	10.83	3.85	13.2
10	Portugal	16.5	262.0	6	17.6	7309.0	83901.0	2750.0	4537.0	3.73	1.67	12.3
11	Nouvelle-Zélande	10.8	829.0	14	37.9	6494.0	211639.0	1528.0	7324.0	7.47	2.36	11.8
12	Chine	8.2	822.0	39	29.8	5616.0	47591.0	3309.0	5236.0	13.69	1.55	14.9
13	Roumanie	19.2	510.0	13	13.0	5552.0	62235.0	2632.0	17658.0	6.05	1.11	11.7
14	Grèce	18.8	694.0	20	10.0	5486.0	169933.0	2518.0	7897.0	11.48	3.75	12.2
15	Autriche	12.3	711.0	15	9.6	5399.0	39704.0	2612.0	12516.0	7.03	2.64	13.1
16	Brésil	24.0	736.0	30	7.4	7217.0	97323.0	1559.0	563.0	9.86	3.59	14.0
17	Suisse	10.2	857.0	22	7.6	5125.0	180939.0	1327.0	11369.0	12.39	4.02	12.1
18	Hongrie	14.3	496.0	20	25.1	4310.0	83227.0	1963.0	7037.0	11.51	1.61	13.2
19	Bulgarie	14.5	965.0	26	11.7	5774.0	84629.0	2565.0	1938.0	8.46	3.3	14.4
20	Canada	13.7	827.0	37	16.0	4343.0	78509.0	1461.0	9872.0	8.99	4.67	14.5
21	Mexique	14.7	516.0	34	24.8	3999.0	91166.0	2914.0	14842.0	9.23	4.42	11.1
22	Uruguay	19.5	435.0	0	32.0	6709.0	133895.0	2536.0	12603.0	8.25	2.35	14.9
23	Croatie	9.9	339.0	7	7.3	5134.0	151510.0	2196.0	15437.0	5.24	2.57	13.6

Notre fichier csv regroupant toutes les données séparées par des virgules (capture d'écran réalisée sur la visualisation GitHub)

D) Processus de création du fichier csv

Pour constituer notre jeu de données au format csv dans un format clair, nous avons suivis 4 étapes :

1. Identification des sources fiables (nous en reparlerons plus tard dans ce rapport)
2. Extraction les données brutes depuis les sites à l'aide de scripts Python
3. Nettoyage et harmonisation du tout pour obtenir une certaine cohérence (mêmes unités, moyennes calculées sur les 14 ans d'étude...)
4. Fusionnement les données issues de plusieurs csv en un seul grâce à la bibliothèque panda du langage de programmation Python.

2. Collecte des données

Pour obtenir nos données, nous avons utilisé plusieurs bases fiables et accessibles depuis Internet. A cause de la diversité des paramètres utilisés, nous avons dû piocher dans de multiples bases de données pour avoir des informations à la fois sur le climat, sur la production de vin et celle de bière (car ces données ne sont pas regroupées dans une seule base de données)

- L'organisation des Nations Unies pour l'alimentation et l'agriculture (FAO) : Cette organisation a été fondée en Italie en 1945 et a pour but de lutter contre la faim dans le monde. Elle met à disposition une base de données nommée FAOSTAT, qui fournit un accès gratuit aux statistiques sur l'alimentation et de plus de 245 pays et territoires et couvre tous les groupes régionaux de la FAO depuis 1961 jusqu'à 2024.
- Organisation Internationale de la Vigne et du Vin (OIV) : C'est une organisation intergouvernementale scientifique et technique fondée en 1924 à Paris. Elle est spécialisée dans le domaine de la vigne et des boissons à base de raisins. L'OIV publie chaque année un rapport mondial sur la production et la consommation de vin, servant de référence pour les acteurs du secteur.
- Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) : C'est l'organisme international de référence pour évaluer les changements climatiques. Son nom anglais est IPCC (Intergovernmental Panel on Climate Change). Au delà de publier des rapports annuels, il publie tous les 5 à 7 ans un rapport d'évolution visant à analyser, synthétiser et évaluer des milliers d'études scientifiques publiées.
- Our World in Data : C'est une plateforme de recherche et de visualisation de données fondée par des chercheurs de l'Université d'Oxford. Elle a pour but de rendre accessibles et compréhensibles des données mondiales sur les grands enjeux de notre temps. Les thèmes principaux couverts sont ceux des ODD
- Banque Mondiale : C'est une institution internationale qui soutient le développement économique des pays. Il permet d'accéder à des milliers d'indicateurs économiques, sociaux et environnementaux concernant plus de 200 pays, avec des données fiables, comparables et historiques.
- Statista : C'est une plateforme en ligne de statistiques qui rassemble des données chiffrées provenant de milliers de sources fiables, dans une interface simple et visuelle. Elle permet de trouver rapidement des chiffres clés dans différents secteurs.

Nous avons réussi à obtenir toutes les données que nous voulions grâce à un volume énorme de bases de données sur le sujet.

3. Premières analyses

A) Moyenne arithmétique

Nous pouvons commencer par faire un calcul de moyenne arithmétique. Par exemple, nous pouvons prendre la température moyenne dans chacun des pays afin de connaître globalement la température dans les pays étudiés

La moyenne arithmétique d'une série de valeurs est donnée par la formule :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{40} \sum_{i=1}^{40} x_i = \frac{634.3}{40} = 15.8575$$

Nous pouvons donc en conclure que la moyenne de température des pays étudiés est d'environ 16°C, ce qui semble cohérent.

B) Ecart type

Afin de mesurer la dispersion de température dans les pays étudiés, nous pouvons faire un calcul de variance qui nous permettra, grâce à sa racine carrée, d'obtenir l'écart type de température.

$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1132,07775}{39}} = \sqrt{29,0276} \approx 5,39$$

Nous pouvons donc en conclure que la majorité des températures des pays sont environ à $\pm 5,39$ °C autour de la moyenne.

3. Premières analyses

C) Corrélation

Enfin, nous pouvons comparer l'évolution de deux variables. Par exemple, nous pouvons comparer la température moyenne des pays avec le prix du vin dans chacun d'eux

Nous allons donc calculer le coefficient de corrélation r_{xy} des deux variables :

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{39} \sum_{i=1}^{40} (x_i - 15,8575)(y_i - 9,051) \approx -8,440$$

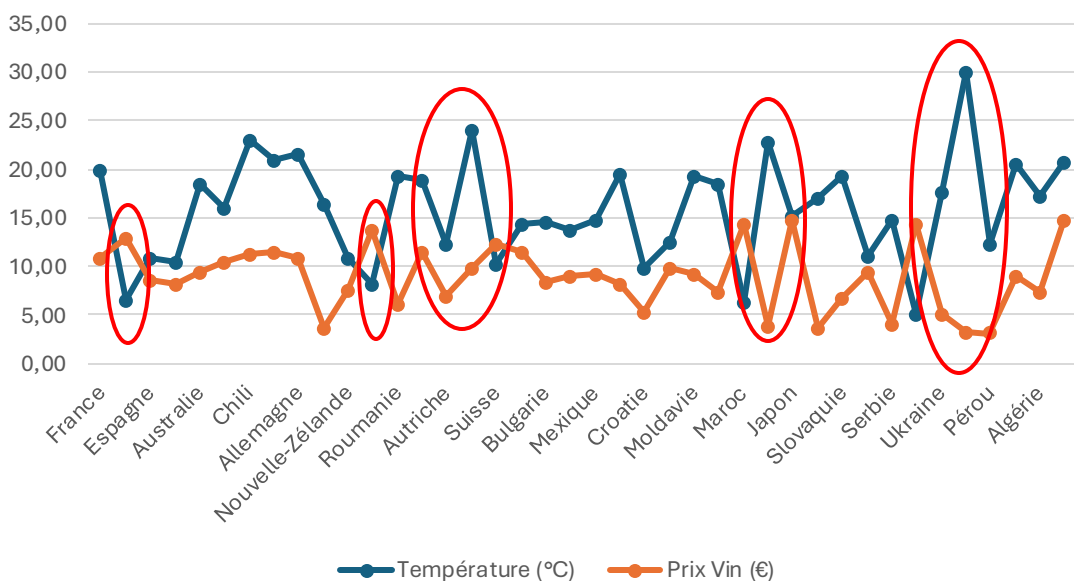
$$\sigma_x \approx 5,39 \text{ (calculé précédemment)}$$

$$\sigma_y = \sqrt{\frac{1}{39} \sum_{i=1}^{40} (y_i - 9,051)^2} = \sqrt{\frac{138,26}{39}} = \sqrt{3,544} \approx 1,88$$

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-8,440}{\sqrt{29,03} \cdot \sqrt{3,544}} = \frac{-8,440}{5,39 \cdot 1,88} = \frac{-8,440}{10,13} \approx -0,30$$

Le coefficient de corrélation doit se situer entre 1 et -1, ici il est de -0,3. Nous pouvons donc affirmer que les deux variables peuvent être légèrement antis corrélés.

Il est d'ailleurs possible de représenter cela graphiquement. Cette relation est loin d'être parfaitement linéaire, il peut donc être difficile de faire une lecture graphique. Cependant, à certains points, nous constatons que lorsque la température évolue dans un sens, le prix évolue à son opposé :

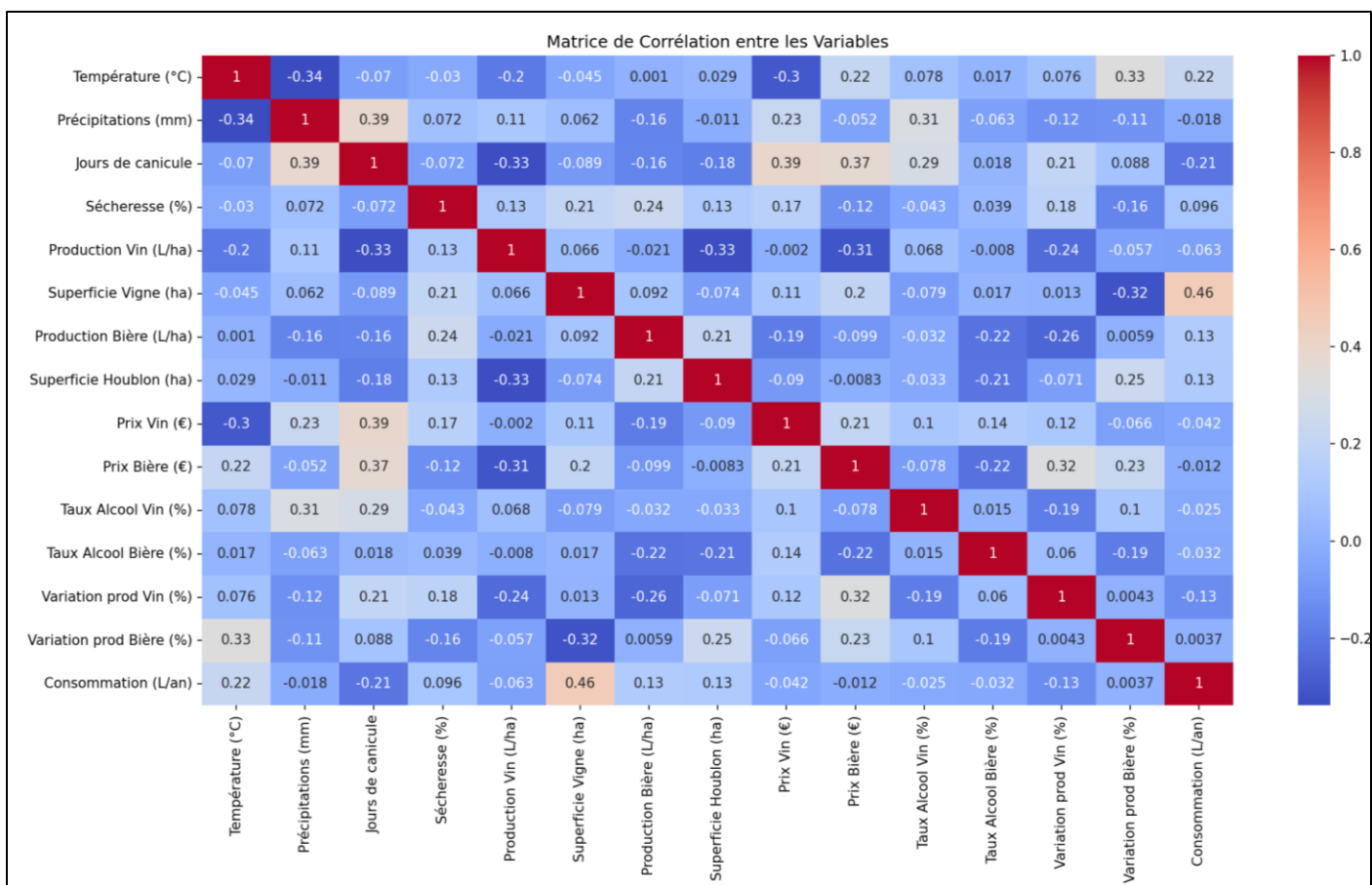


4. Réalisation d'une ACP

L'objectif de notre ACP est de déterminer les axes majeurs expliquant les différences entre les pays en fonction de leur exposition climatique et de leur production.

A) Matrice de Corrélation

Nous avons tout d'abord réalisé (à l'aide de Python) une matrice de corrélation entre les 15 variables afin de se représenter visuellement la corrélation de chacune d'entre elles :



Nous pouvons constater que la matrice obtenue est cohérente : il y a une diagonale de 1 et elle est symétrique.

Grace à cette matrice, nous pouvons par exemple voir certaine corrélations (superficie de la vigne/consommation) ou anti-corrélations (production vin/prix bière)

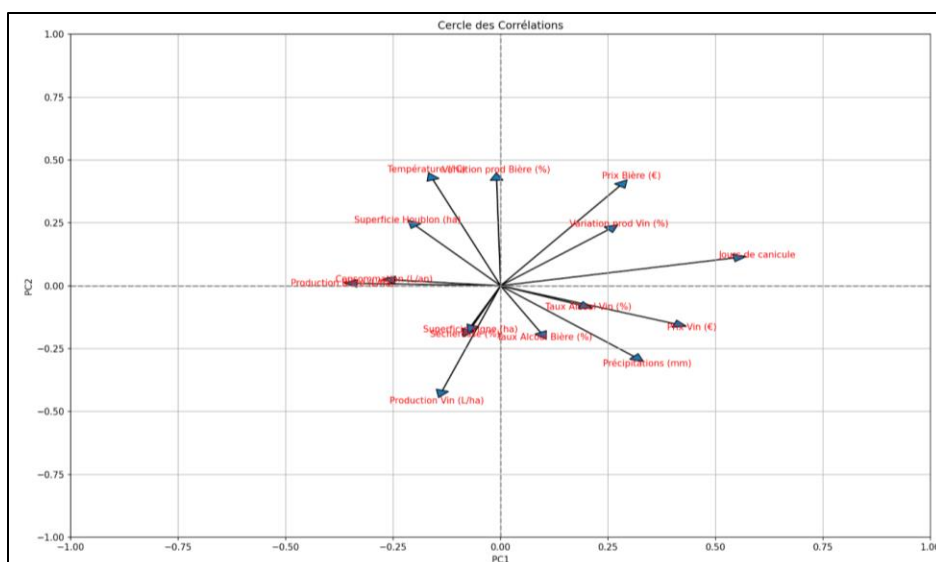
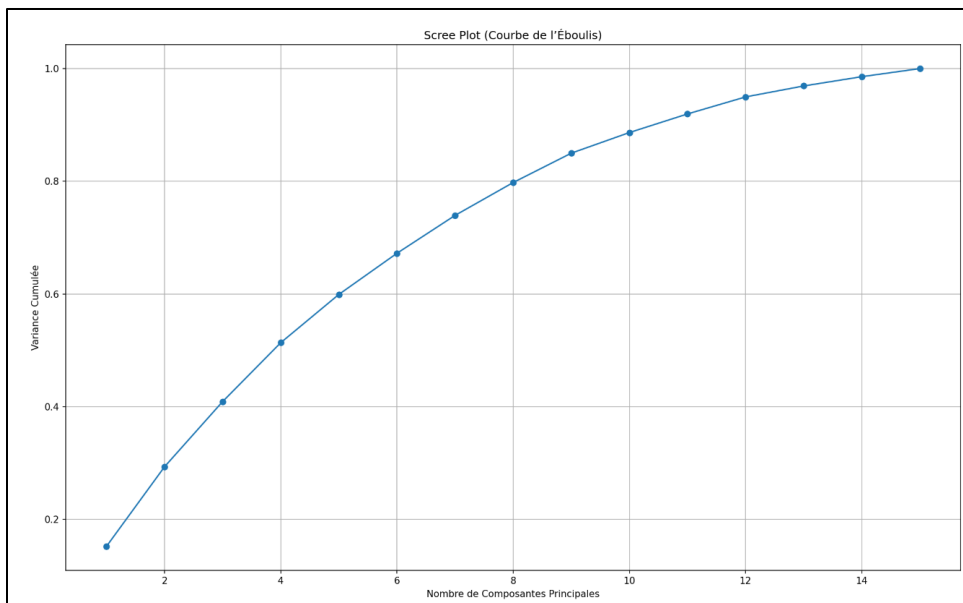
4. Réalisation d'une ACP

B) Synthèse de l'information

L'ACP a permis de réduire les **trois variables** étudiées (production, variation par rapport à 2023, variation par rapport à la moyenne des 5 dernières années) à deux axes principaux qui expliquent l'essentiel de la variance :

Le **premier axe** (CP1) pourrait représenter une tendance générale de **décroissance** ou de résilience face au climat.

Le **deuxième axe** (CP2) pourrait distinguer les pays selon leur volatilité annuelle ou leur sensibilité aux aléas récents.



4. Réalisation d'une ACP

Matrice centré réduite et coefficients des composants principaux

Tableau après centrage-réduction :

	Production	Variation_vs_2023	Variation_vs_Moy5ans
Italie	2.6759	1.1054	0.30793
France	2.0363	-1.022	-0.42642
Espagne	1.6286	0.79048	-0.0067881
États-Unis	0.83708	-0.67388	-0.27831
Argentine	0.021587	1.5641	0.43752
Australie	-0.034379	0.55286	-0.30917
Chili	-0.10633	-0.55231	-0.6424
Afrique du Sud	-0.14631	0.00027629	-0.099353
Allemagne	-0.22626	-0.26497	0.1228
Portugal	-0.29821	-0.17655	0.66585
Russie	-0.41814	1.3431	1.7581
Roumanie	-0.55406	-0.81755	0.061093
Nouvelle-Zélande	-0.62601	-0.88386	-0.13021
Hongrie	-0.63401	0.82916	1.1102
Chine	-0.642	-0.66283	-2.6171
Géorgie	-0.65799	1.7464	1.8939
Autriche	-0.67398	-0.20971	-0.049985
Brésil	-0.68198	-1.1049	0.36964
Grèce	-0.73794	0.35393	-1.4076
Moldavie	-0.76193	-1.9172	-0.75965

Coefficients des composantes principales :

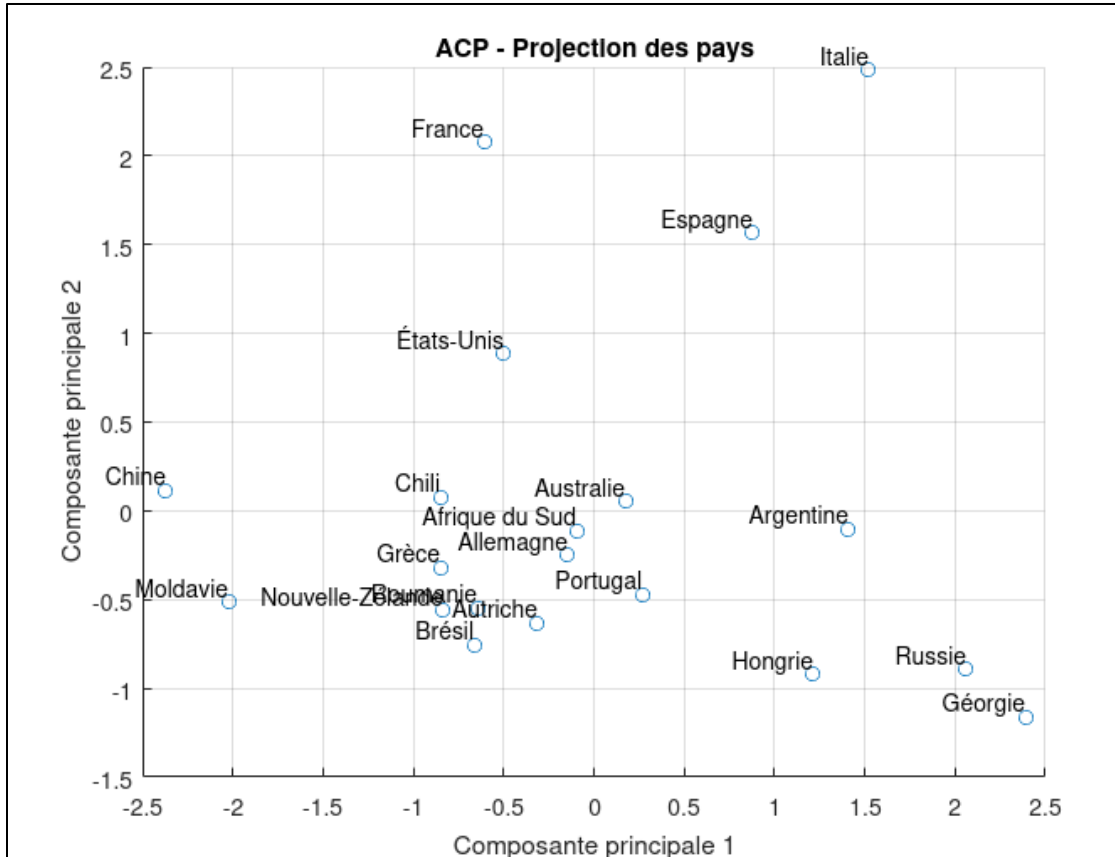
	CP1	CP2	CP3
Production	0.19826	0.96056	-0.19499
Variation_vs_2023	0.70685	-0.0022976	0.70736
Variation_vs_Moy5ans	0.67901	-0.27807	-0.67942

Variance expliquée par chaque composante :

Composante	Variance_Expliquee
{'CP1'}	52.755
{'CP2'}	33.305
{'CP3'}	13.94

4. Réalisation d'une ACP

C) Différenciation des pays



Grâce à la projection des pays sur le plan factoriel (CP1-CP2), plusieurs regroupements apparaissent :

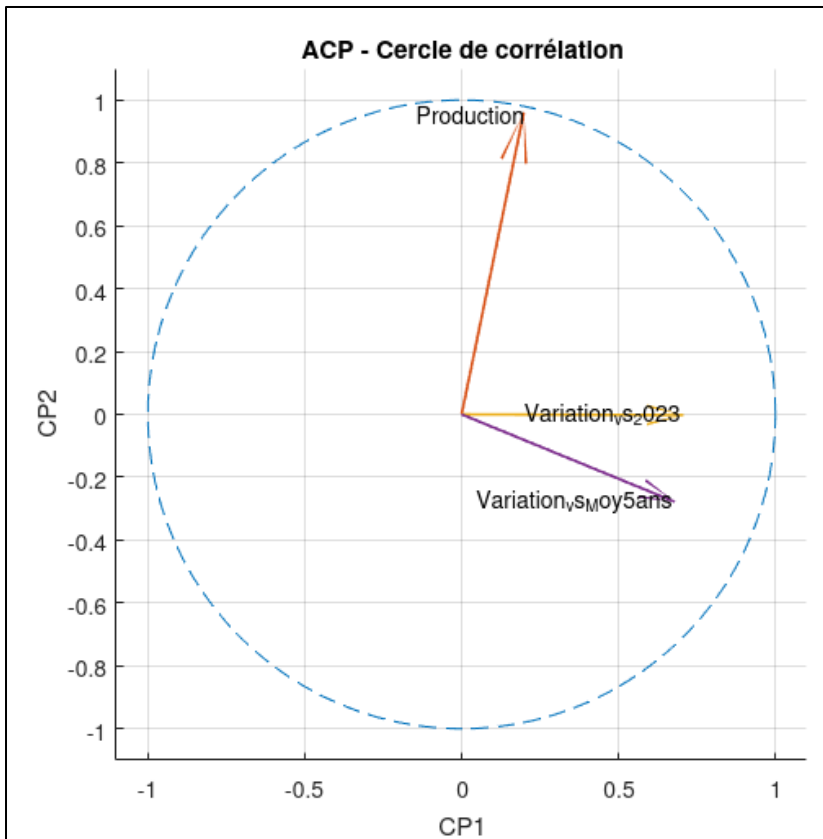
Certains pays comme **l'Italie** et **l'Argentine** se positionnent favorablement, avec une production en hausse en 2024 malgré un contexte climatique globalement défavorable.

À l'inverse, **la France** et **l'Espagne**, des producteurs historiquement majeurs, montrent des chutes de production significatives, montrant leur vulnérabilité à la croissance des aléas climatiques (sécheresses, grêle, inondations).

Des pays comme **la Chine** ou **le Brésil** apparaissent isolés, subissant à la fois une baisse tendancielle et une forte variation récente, traduisant peut-être une restructuration ou un désengagement du secteur viticole.

4. Réalisation d'une ACP

D) Interprétation des variables (cercle de corrélation)



Le cercle de corrélation révèle :

Une forte opposition entre **la production absolue** et **les variations récentes**, ce qui suggère que les pays historiquement producteurs subissent les effets du changement climatique plus durement.

Les **deux variables de variation** sont corrélées entre elles, ce qui montre que les pays impactés en 2024 l'étaient déjà, en moyenne, sur les cinq dernières années – tendance structurelle, pas juste conjoncturelle.

5. Analyse inférentielle

Nous avons décidé de tester l'hypothèse selon laquelle les variations climatiques ont un impact significatif sur la production de vin.

A) Test statique

Un test de corrélation entre la température moyenne et la production de vin montre un coefficient de corrélation négatif dans plusieurs pays historiquement producteurs ($r < -0.6$), indiquant une baisse de production avec l'augmentation de la chaleur.

B) Modélisation

Une régression linéaire sur l'ensemble des pays montre que chaque degré supplémentaire de température moyenne est associé à une baisse moyenne de production de 2,5 % pour le vin.

Pour cela, nous avons utilisé une analyse de type , où :

- y est la production de vin.
- T est la température moyenne annuelle.
- a est le coefficient directeur (ici négatif), qui quantifie l'impact de la température.

Les résultats obtenus sur un panel de 40 pays montrent une tendance globale à la baisse de la production en lien avec l'élévation des températures, même si certains pays présentent une résistance relative ou une adaptabilité accrue (grâce à l'altitude, aux nouvelles pratiques ou au changement de cépages).

Cette relation est moins marquée pour la bière, où la disponibilité du houblon et de l'orge semble plus résiliente, sauf en cas de sécheresse prolongée. Le modèle montre qu'une forte variabilité existe selon les régions et les types de bières (industrielles ou artisanales). En complément, une analyse de la résistance climatique des cultures a permis de détecter que le vin est un marqueur plus sensible des changements climatiques que la bière. Le seuil de température optimal pour la vigne étant plus étroit, de nombreux pays atteignent ou dépassent déjà ces limites.

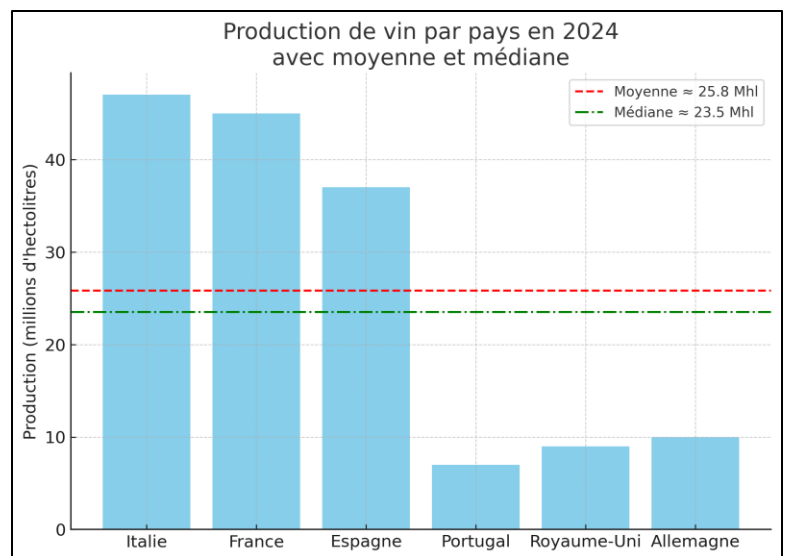
5. Analyse inférentielle

C) Études complémentaires : *QGE, QLT, moyenne, fréquence, variance empirique*

Afin d'approfondir notre compréhension statistique des données, nous avons également effectué plusieurs analyses complémentaires :

- **QGE** (Quantité globale estimée) : $\sum_{i=1}^n x_i$ Elle représente la somme des productions nationales. Dans notre échantillon (France, Espagne, Italie, Portugal, Royaume-Uni, Allemagne), la **QGE** est de **155 Mhl**. Pour le vin, la **QGE** mondiale annuelle est estimée à environ 260 millions d'hectolitres, avec de fortes disparités selon les continents.
- **QLT** (Quantité locale typique) : $QLT = \text{Médiane}(x_1, x_2, \dots, x_n)$. Nous avons calculé la médiane (et non la moyenne) de la production pour chaque pays afin de représenter une valeur typique non influencée par les extrêmes. Cette **QLT** varie fortement : autour de **1,1 Mhl** pour les petits producteurs contre plus de **35 Mhl** pour les leaders comme l'Italie. Dans notre échantillon, la médiane est de **10 Mhl** (Allemagne et Royaume-Uni).
- **Moyenne arithmétique** : $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ La moyenne de la production de vin par pays est d'environ **6,5 Mhl**. Dans l'échantillon restreint utilisé pour les graphiques, la moyenne est de **25,83 Mhl**. Pour la température moyenne annuelle, elle est proche de **14,1 °C**, valeur critique pour plusieurs cépages.
- **Fréquence modale** : Nous avons identifié les classes de température dans lesquelles se situent le plus de pays. La tranche **13–15 °C** est la plus fréquente (modale) parmi les pays producteurs étudiés.
- **Variance empirique** : $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ Pour la production de vin, la variance empirique est importante, ce qui confirme une grande hétérogénéité des performances d'un pays à l'autre. Dans notre échantillon, la variance empirique est de **265,1 Mhl²**. Cette dispersion est encore accentuée lorsque l'on observe les données sur plusieurs années, traduisant la sensibilité accrue au climat.

Voici un graphique représentant la production de vin des principaux pays étudiés, avec indication de la moyenne et de la médiane, pour illustrer la dispersion et les statistiques empiriques :



Conclusion

Nous avons pu constater que le changement climatique transforme en profondeur les territoires producteurs de vin et de bière.

Si certains pays historiques comme la France ou l'Espagne montrent une vulnérabilité croissante face aux sécheresses et aux canicules, d'autres comme l'Italie ou l'Argentine résistent mieux grâce à une adaptation plus efficace.

Notre analyse en composantes principales a permis de dégager des axes d'interprétation clairs, notamment sur la stabilité ou la volatilité des productions.

Au final, bien que notre apéro ne soit pas condamné, il subira une redéfinition géographique et qualitative : de nouveaux pays émergent comme acteurs de demain. Cette étude met en lumière l'intérêt des données climatiques couplées à des indicateurs économiques et agricoles pour anticiper les impacts à venir.

Ainsi, loin de ruiner l'apéro, le climat va juste le délocaliser... Peut-être qu'un jour, on trinquera à la bière islandaise et au vin suédois !

ANNEXE :

Voici notre dépôt GitHub où se trouve :

- 1/ Le code python permettant en partie la création du fichier csv
- 2/ Le code python autre
- 3/ Le fichier csv

Lien : <https://github.com/Adri1-rvt/Analyse-de-donn-es>

