

- 1 Importation des Librairies
- 2 Importation des données
- 3 Vérification des données
- 4 Jointures et données finales
- 5 Tests de normalité et de corrélation

# Préparation, Nettoyage et Analyse exploratoire des données

2023-06-23



## POULET MONDIAL

L'entreprise "Poulet Mondial" est une entreprise française d'agroalimentaire qui souhaite se développer à l'international.

Une entreprise française d'agroalimentaire spécialisé dans la volaille souhaite se développer à l'international.

L'international, oui, mais pour l'instant, le champ des possibles est bien large : aucun pays particulier ni aucun continent n'est pour le moment choisi. Tous les pays sont envisageables !

L'objectif de l'étude est de proposer une première analyse des groupements de pays que l'on peut cibler pour exporter les poulets.

La première partie de notre travail a été de récolter les données nécessaires pour cette étude. Nous les avons ensuite nettoyés, préparés et fusionnés pour en faire une première analyse exploratoire.

## 1 Importation des Librairies

```
# Pour des opérations utilitaires
library(utils)

# Pour La création de graphiques
library(ggplot2)

# Pour des fonctionnalités statistiques avancées
library(Hmisc)

# Pour La manipulation de données
library(dplyr)

# Pour La visualisation de matrices de corrélation
library(corrplot)

# Pour créer des visualisations interactives
library(plotly)

# Pour améliorer La présentation des tableaux
library(kableExtra)
```

## 2 Importation des données

Les données utilisées sont des données issues de la FAO (Food and Agriculture Organization).

- Données la Population: : La population de chaque zone peut servir à évaluer la taille potentielle du marché.
- Données sur l'Alimentation (Disponibilité intérieure, TDI, TAS, production, importation et exportation) : Analysez ces indicateurs pour évaluer la demande intérieure. On va Rechercher les zones où la demande est élevée mais la production locale est insuffisante pour répondre à cette demande.
- Indicateur Politique (stabilité politique) Recherchez les zones avec une stabilité politique relativement élevée pour minimiser les risques potentiels pouvant affecter les opérations commerciales et garantir la sécurité des collaborateurs.
- Indicateurs économiques (PIB , inflation) ; les zones où le PIB est élevé indique un pouvoir d'achat potentiel plus élevé pour les consommateurs . Une inflation plus faible peut indiquer une économie stable, ce qui est généralement préférable pour le commerce international.

En combinant ces facteurs et en les évaluant pour chaque pays potentiel, l'étude peut identifier les groupements de pays offrant le meilleur potentiel pour les exportations de poulet de "La poule qui chante".

## 3 Vérification des données

### 3.1 Dataframe Population

- Nombre total de valeurs manquantes dans le dataframe "Population"

```
## [1] 0
```

- Nombre de doublons

```
## [1] 0
```

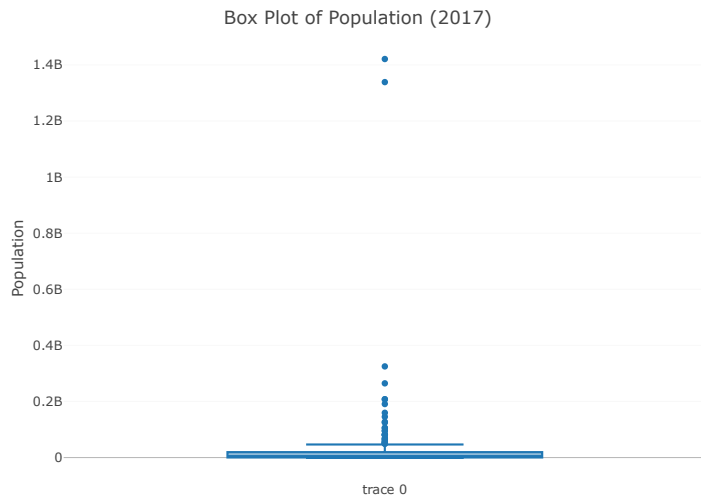
- Vérification du format des colonnes dans le dataframe "Population"

```
## 'data.frame': 4411 obs. of 4 variables:
## $ Code.zone : int 2 2 2 2 2 2 2 2 2 ...
## $ Zone : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ Année : int 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 ...
## $ Population: num 20779953 21606988 22600770 23680871 24726684 ...
```

- Résumé statistique du dataframe "Population"

Code.zone	Zone	Année	Population
Min.: 1.0	Length:4411	Min.: 2000	Min.: 7.850e+02
1st Qu.: 68.0	Class :character	1st Qu.: 2004	1st Qu.: 3.922e+05
Median :132.0	Mode :character	Median : 2009	Median : 4.765e+06
Mean :132.2	NA	Mean : 2009	Mean : 2.963e+07
3rd Qu.:195.0	NA	3rd Qu.: 2014	3rd Qu.: 1.822e+07
Max.: 299.0	NA	Max.: 2018	Max.: 1.428e+09

- Boxplot "Population"



On observe 2 pays outliers qui sont l'Inde et la Chine continentale.

Code.zone	Zone	Année	Population
41	Chine, continentale	2017	1421021791
100	Inde	2017	1338676785
231	États-Unis d'Amérique	2017	325084756
101	Indonésie	2017	264650963
165	Pakistan	2017	207906209
21	Brésil	2017	207833823
159	Nigéria	2017	190873244
16	Bangladesh	2017	159685424
185	Fédération de Russie	2017	145530082
110	Japon	2017	127502725
138	Mexique	2017	124777324
238	Éthiopie	2017	106399924
171	Philippines	2017	105172925
59	Égypte	2017	96442591
237	Viet Nam	2017	94600648
79	Allemagne	2017	82658409
250	République démocratique du Congo	2017	81398764
223	Turquie	2017	81116450
102	Iran (République islamique d')	2017	80673883
216	Thaïlande	2017	69209810

La Chine et l'Inde ont toutes deux des populations très élevées, avec 1,42 milliard de personnes en Chine et environ 1,34 milliard de personnes en Inde. Cela indique une base de consommateurs potentielle importante pour le poulet.

Cependant, nous allons attendre d'étudier d'autres facteurs par la suite pour déterminer s'il est judicieux de les garder.

Code.zone	Zone	Année	Population
94	Saint-Siège	2017	793
218	Tokélaou	2017	1300
160	Nioué	2017	1609
65	Îles Falkland (Malvinas)	2017	3068
142	Montserrat	2017	4984
190	Saint-Pierre-et-Miquelon	2017	5887
187	Sainte-Hélène, Ascension et Tristan da Cunha	2017	6008
282	Saint-Barthélemy	2017	9784
148	Nauru	2017	10573
227	Tuvalu	2017	11370
243	Îles Wallis-et-Futuna	2017	11900
258	Anguilla	2017	14584
47	Îles Cook	2017	17507
180	Palaos	2017	17808
278	Bonaire, Saint-Eustache et Saba	2017	25401
239	Îles Vierges britanniques	2017	29577
192	Saint-Marin	2017	33671
82	Gibraltar	2017	33728
281	Saint-Martin (partie française)	2017	36560
224	Îles Turques-et-Caïques	2017	37115

Certaines valeurs de population peuvent paraître basse. Cependant, après quelques vérifications, on s'aperçoit qu'elles ne sont pas erronées.

3.2 Dataframe Alimentation

Nous avons à notre disposition un dataframe contenant des données de production, importation, exportation et variation des stocks. Nous allons utiliser ces données pour calculer de nouvelles valeurs qui seront pertinentes pour notre étude.

Les nouvelles données que nous allons calculer sont les suivantes :

- 1. "TDI" (Taux de Dépendance aux Importations %)

Le TDI mesure la dépendance d'un pays par rapport aux importations. Il est calculé de la manière suivante :  $TDI = (Valeur\ des\ importations / Valeur\ de\ la\ consommation\ totale) \times 100$

Ce ratio exprime la part des importations dans la consommation totale du pays.

- 2. "TAS" (Taux d'Auto-Suffisance %)

Le TAS mesure la capacité d'un pays à produire suffisamment de produits pour satisfaire sa propre demande interne, sans dépendre des importations. Il est calculé de la manière suivante :  $TAS = (Production\ nationale / Consommation\ totale) \times 100$

Un taux d'autosuffisance élevé indique une plus grande capacité du pays à répondre à sa demande interne par sa propre production.

- 3. "Disponibilité intérieure" en milliers de tonnes

$Disponibilité\ intérieure = Production + Importation - Exportation + Variation\ des\ stocks$

3.2.1 Préparation du df "Alimentation"

- Nombre total de valeurs manquantes dans le dataframe "Alimentation"

## [1] 0

- Nombre de doublons

## [1] 0

- Vérification du format des colonnes dans le dataframe "Alimentation"

```
## tibble [170 × 10] (S3: tbl_df/tbl/data.frame)
## $ Code.zone      : int [1:170] 2 202 3 4 79 7 8 194 9 1 ...
## $ Zone           : chr [1:170] "Afghanistan" "Afrique du Sud" "Albanie" "Algérie" ...
## $ Année          : int [1:170] 2017 2017 2017 2017 2017 2017 2017 2017 2017 2017 ...
## $ Production     : num [1:170] 28 1667 13 275 1514 ...
## $ Importations - Quantité : num [1:170] 29 514 38 2 842 277 7 722 8 35 ...
## $ Variation de stock  : num [1:170] 0 0 4 0 -29 0 0 -108 0 -2 ...
## $ Exportations - Quantité : num [1:170] NA 63 0 0 646 0 0 10 207 0 ...
## $ TDI            : num [1:170] NA 24.27 74.51 0.72 49.24 ...
## $ TAS            : num [1:170] NA 78.7 25.5 99.3 88.5 ...
## $ disponibilité_intérieure: num [1:170] NA 2118 55 277 1681 ...
```

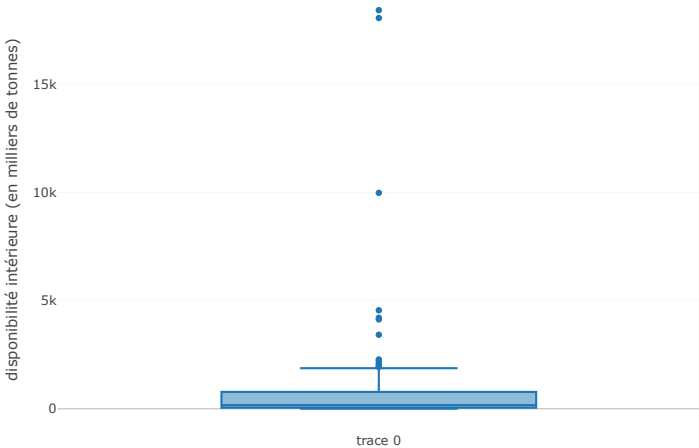
- Résumé statistique du dataframe "Alimentation"

Code.zone	Zone	Année	Production	Importations - Quantité	Variation de stock	Exportations - Quantité	TDI	TAS	disponibilité_intérieure
Min. : 1.00	Length:170	Min. :2017	Min. : 0.00	Min. : 0.00	Min. :-119.00	Min. : 0.0	Min. : 0.00	Min. : 0.00	Min. : 4.00
1st Qu.: 66.25	Class :character	1st Qu.:2017	1st Qu.: 13.75	1st Qu.: 3.00	1st Qu.: 0.00	1st Qu.: 0.0	1st Qu.: 3.99	1st Qu.: 62.72	1st Qu.: 51.25
Median :121.50	Mode :character	Median :2017	Median : 70.00	Median : 16.00	Median : 0.00	Median : 3.0	Median : 20.55	Median : 88.46	Median : 168.50
Mean :126.04	NA	Mean :2017	Mean : 725.19	Mean : 89.53	Mean : 13.67	Mean : 132.2	Mean : 38.28	Mean : 80.21	Mean : 881.24
3rd Qu.:187.25	NA	3rd Qu.:2017	3rd Qu.: 409.75	3rd Qu.: 81.25	3rd Qu.: 7.00	3rd Qu.: 32.0	3rd Qu.: 67.95	3rd Qu.: 100.00	3rd Qu.: 744.75
Max. :276.00	NA	Max. :2017	Max. :21914.00	Max. :1069.00	Max. : 859.00	Max. :4223.0	Max. :338.43	Max. :379.31	Max. :18425.00
NA	NA	NA	NA's :2	NA	NA's :1	NA's :35	NA's :35	NA's :35	NA's :36

3.2.2 Etude de la variable "disponibilité intérieure"

- Boxplot de la "disponibilité intérieure" en viande de volailles

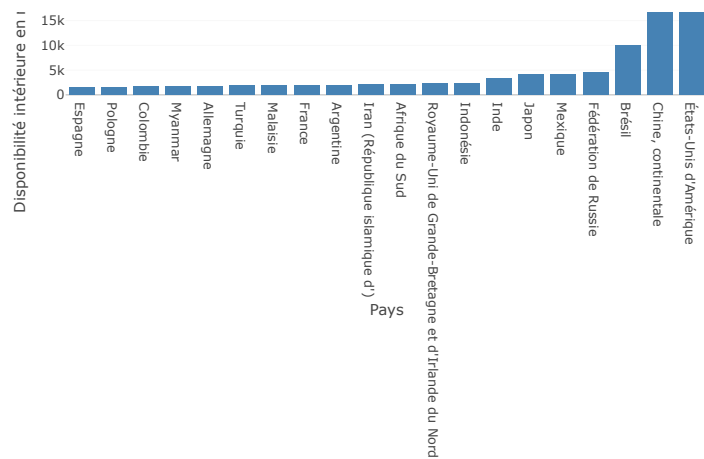
Box Plot de la disponibilité intérieure en viande de volailles



- Top 20 des pays avec la disponibilité intérieure la plus élevée

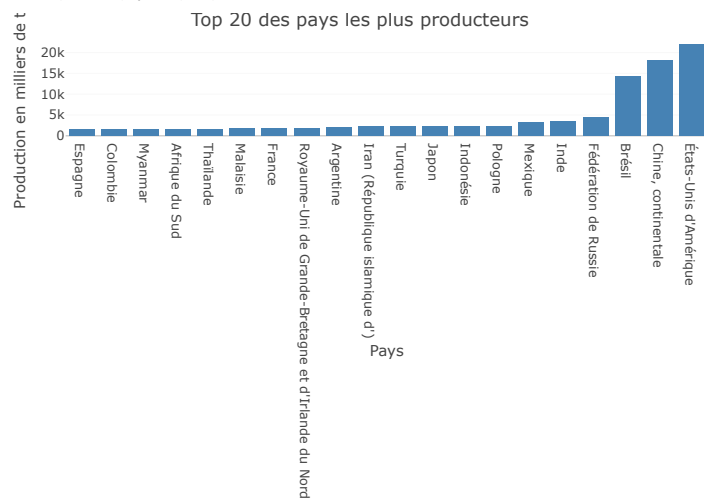
Top 20 des pays avec la disponibilité intérieure la plus élevée

millie



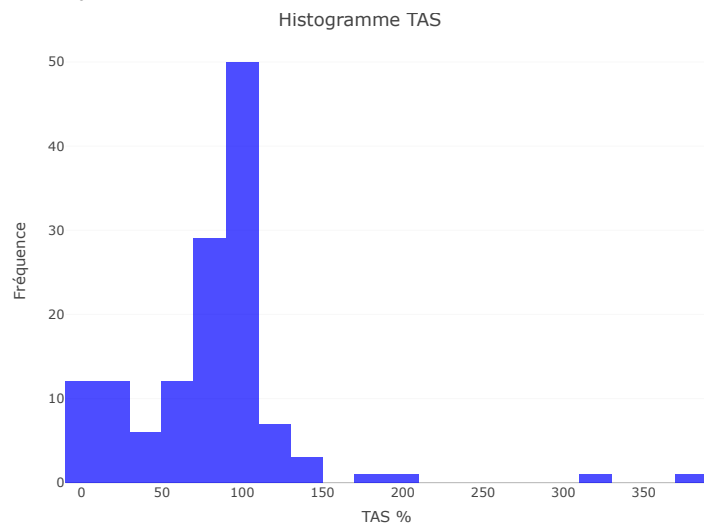
### 3.2.3 Etude de la variable Production

- Top 20 des pays les plus producteurs

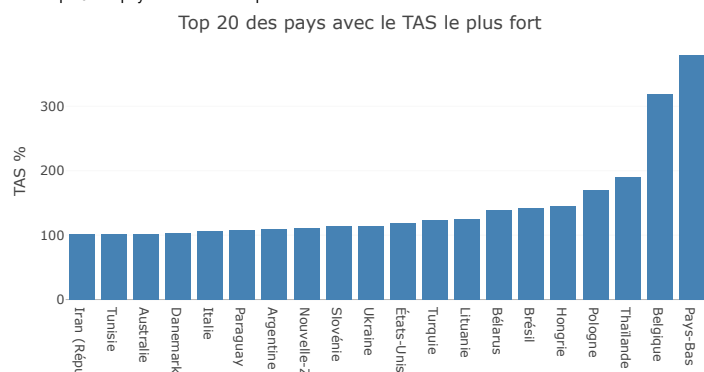


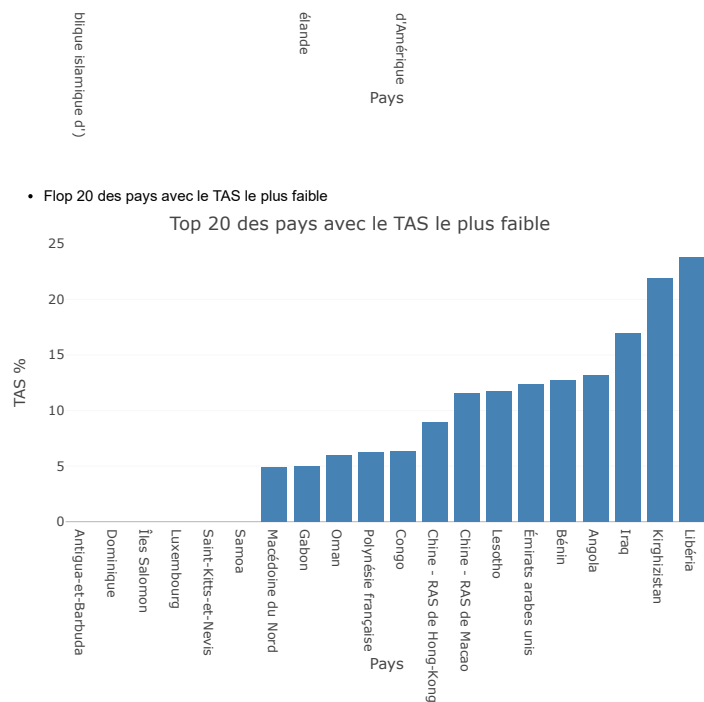
### 3.2.4 Etude de la variable "TAS"

- Histogramme du "TAS"



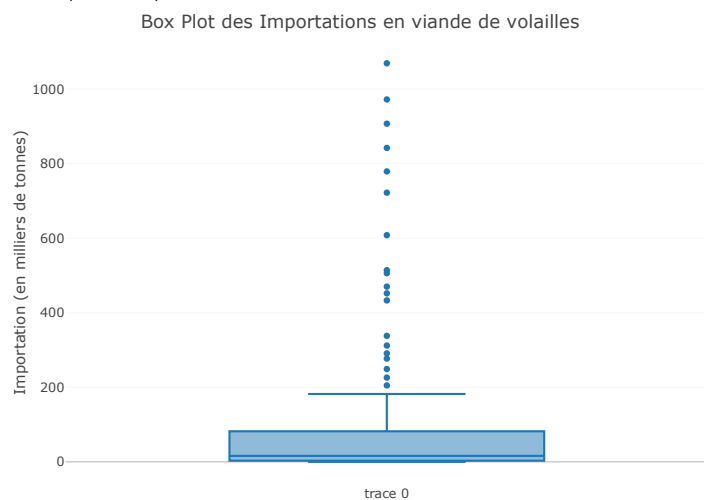
- Top 20 des pays avec le TAS le plus fort



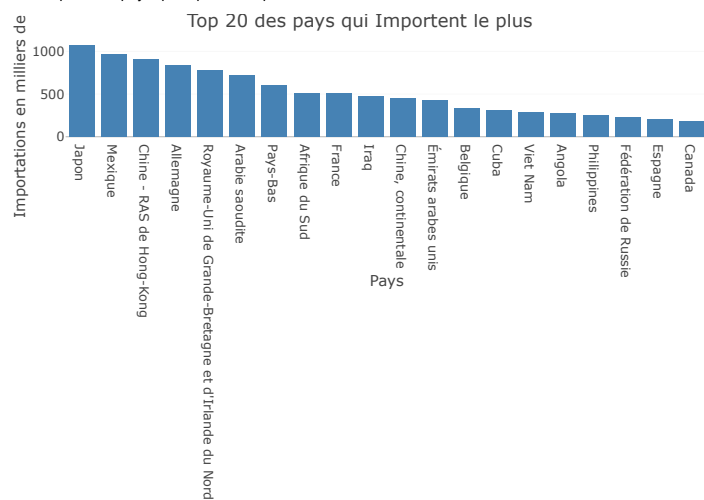


### 3.2.5 Etude de la variable “Importation”

- Boxplot de la “importation” en viande de volailles

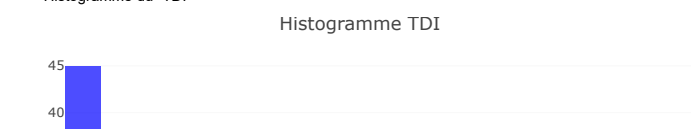


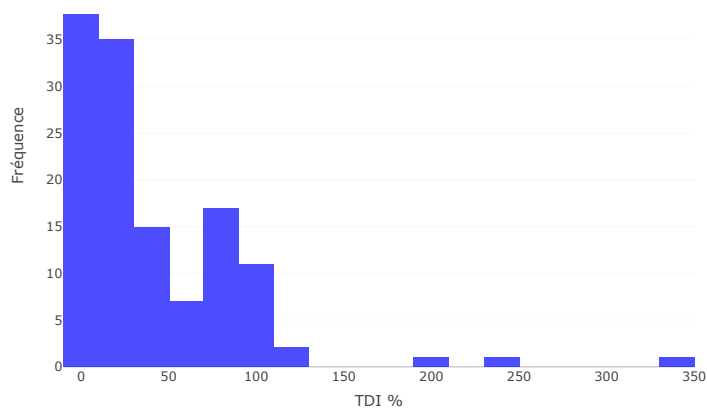
- Top 20 des pays qui importent le plus



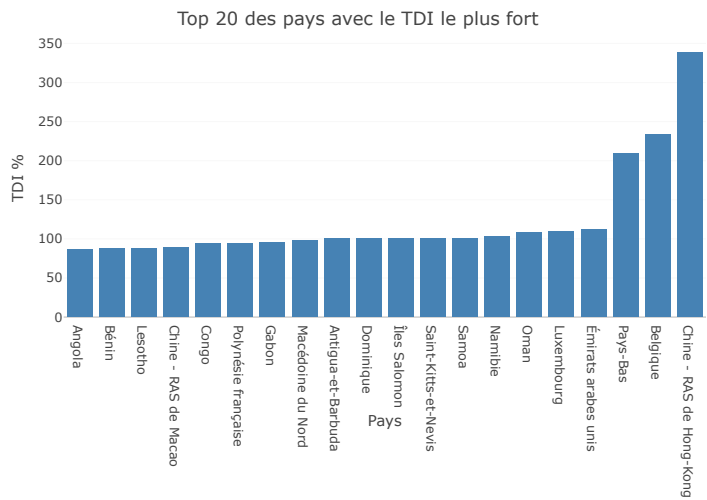
### 3.2.6 Etude de la variable “TDI”

- Histogramme du “TDI”





- Top 20 des pays avec le TDI le plus élevé



### 3.3 Dataframe PIB

PIB exprimé en \$ par habitant

- Nombre total de valeurs manquantes dans le dataframe "PIB"

```
## [1] 0
```

- Nombre de doublons

```
## [1] 0
```

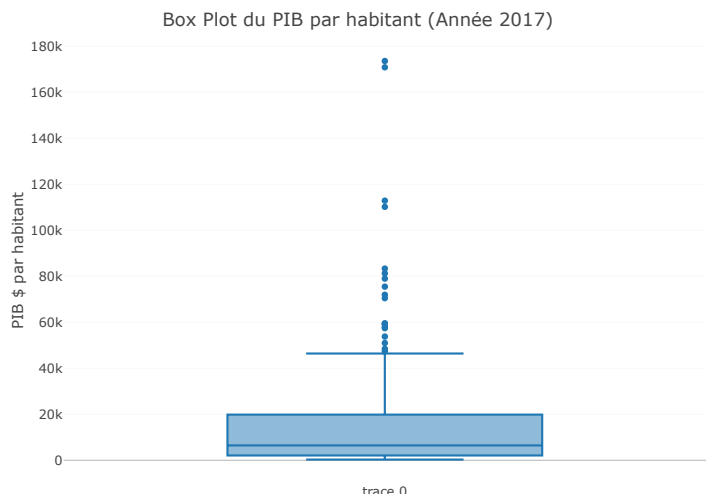
- Vérification du format des colonnes dans le dataframe "PIB"

```
## 'data.frame': 1899 obs. of 4 variables:
## $ Code.zone: int 4 4 4 4 4 4 4 4 4 4 710 ...
## $ Zone : chr "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
## $ Année : int 2013 2014 2015 2016 2017 2018 2019 2020 2021 2013 ...
## $ PIB : num 622 598 554 520 530 ...
```

- Résumé statistique du dataframe "PIB"

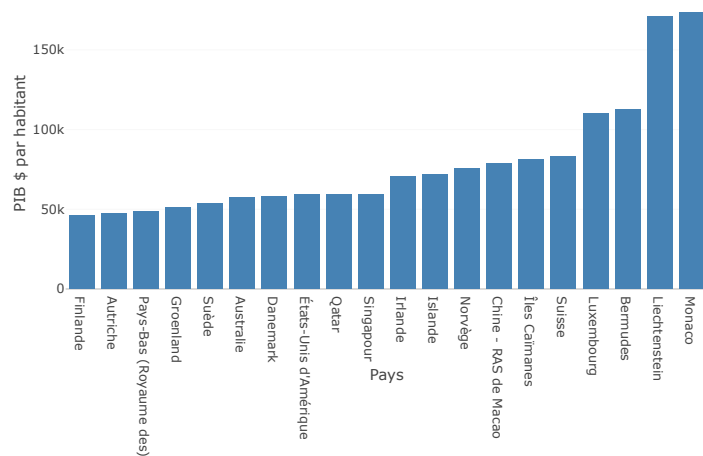
Code.zone	Zone	Année	PIB
Min.: 4.0	Length:1899	Min.: 2013	Min.: 241.6
1st Qu.:208.0	Class:character	1st Qu.:2015	1st Qu.: 2139.8
Median:430.0	Mode:character	Median:2017	Median: 6658.1
Mean:429.4	NA	Mean:2017	Mean: 17219.3
3rd Qu.:643.0	NA	3rd Qu.:2019	3rd Qu.:20182.3
Max.:894.0	NA	Max.:2021	Max.:234317.1

- Boxplot du "PIB"



- Top 20 des pays avec le PIB le plus élevé

Top 20 des pays avec le PIB le plus fort



## 3.4 Dataframe Inflation

- Nombre de doublons

```
## [1] 0
```

- Nombre total de valeurs manquantes dans le dataframe "Inflation"

```
## [1] 0
```

- Vérification du format des colonnes dans le dataframe "Inflation"

```
## 'data.frame': 4450 obs. of 4 variables:
## $ Code.zone: int 100 100 100 100 100 100 100 100 100 100 ...
## $ Zone : chr "Bulgarie" "Bulgarie" "Bulgarie" "Bulgarie" ...
## $ Année : int 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 ...
## $ Inflation: num 6.69 -0.05 -0.77 6.77 4.06 ...
```

- Résumé statistique du dataframe "Inflation"

Code.zone	Zone	Année	Inflation
Min.: 4.0	Length:4450	Min.: 2001	Min.: -18.13
1st Qu.:212.0	Class:character	1st Qu.:2006	1st Qu.: 1.65
Median :426.0	Mode:character	Median :2012	Median : 3.94
Mean :424.7	NA	Mean :2012	Mean : 49.60
3rd Qu.:638.0	NA	3rd Qu.:2017	3rd Qu.: 8.16
Max.:894.0	NA	Max.:2022	Max.:143491.55

- Boxplot inflation

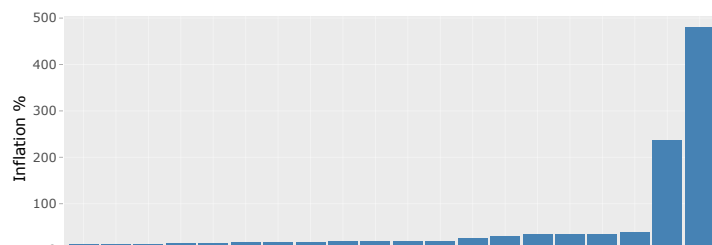
Box Plot de l'Inflation 2017

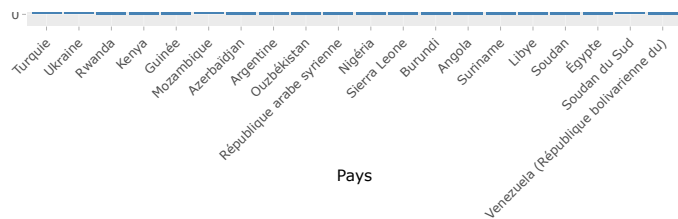


Nous observons 2 outliers.

- Top 20 des pays avec le taux d'Inflation le plus élevé

Top 20 des pays avec le taux d'Inflation le plus fort





Les outliers visualisés précédemment sont le Venezuela et le Soudan du Sud. Ils ont une inflation trop importante. L'inflation peut affecter les coûts de production et les prix des produits. Nous Privilégions les zones avec une inflation relativement faible pour maintenir la rentabilité à long terme.

Zone	Importations - Quantité	TDI	TAS
Venezuela (République bolivarienne du)	25	4	96

De plus, le Venezuela se caractérise par une importation faible en viande de volailles, un TDI très faible 4% et un TAS de 96%. Ce n'est pas un pays intéressant pour notre marché. Concernant le Soudan du sud, nous n'avons pas les données "importations", "TDI" et "TAS".

Nous décidons de supprimer ces 2 pays de l'étude.

## 3.5 Dataframe Stabilité politique

- Nombre total de valeurs manquantes dans le dataframe "Stabilité\_politique"

```
## [1] 0
```

- Nombre de doublons

```
## [1] 0
```

- Vérification du format des colonnes dans le dataframe "Stabilité\_politique"

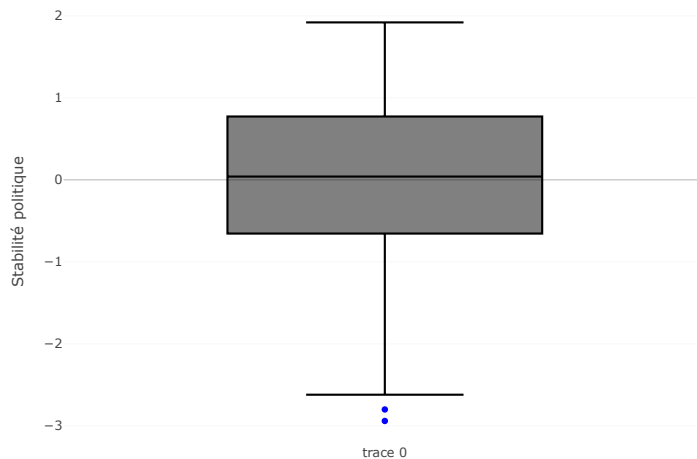
```
## 'data.frame': 3908 obs. of 4 variables:
## $ Code.zone : int 150 150 150 150 150 150 150 150 150 150 ...
## $ Zone : chr "Pays-Bas (Royaume des)" "Pays-Bas (Royaume des)" "Pays-Bas (Royaume des)" "Pays-Bas (Royaume des)" ...
## $ Année : int 2000 2002 2003 2004 2005 2006 2007 2008 2009 2010 ...
## $ Stabilité_politique: num 1.76 1.31 1.16 1.02 0.99 0.9 0.78 0.87 0.94 0.94 ...
```

- Résumé statistique du dataframe "Stabilité\_politique"

Code.zone	Zone	Année	Stabilité_politique
Min.: 1	Length:3908	Min.:2000	Min.: -3.310
1st Qu.: 61	Class :character	1st Qu.:2006	1st Qu.: -0.710
Median :127	Mode :character	Median :2011	Median : 0.055
Mean :128	NA	Mean :2011	Mean : -0.050
3rd Qu.:193	NA	3rd Qu.:2016	3rd Qu.: 0.800
Max.:299	NA	Max.:2020	Max.: 1.970

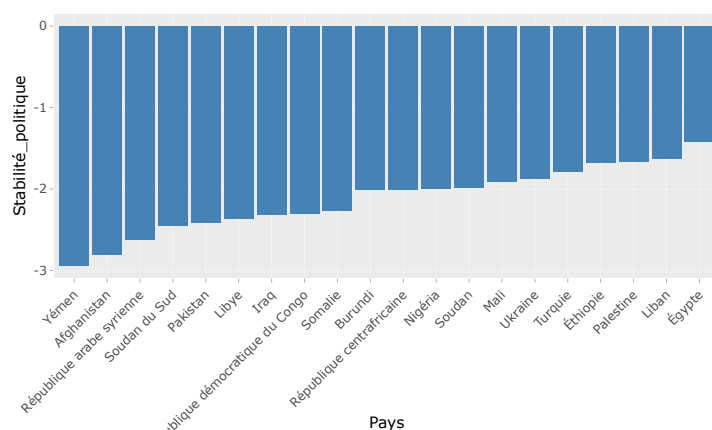
- Boxplot de la Stabilité politique

Box Plot de la Stabilité politique en 2017



- Flop 20 des pays avec la Stabilité politique la plus faible 2017

Top 20 des pays avec la Stabilité politique la plus faible 2017





Répo-

Nous observons 2 outliers qui sont "Yémen" et "Afghanistan". Nous décidons de les supprimer de l'étude.

## 4 Jointures et données finales

### 4.1 Jointures

Nous allons effectuer des jointures successives par "Zone" et par "Année" afin de rassembler toutes les données en un seul dataframe. L'objectif est de joindre les dataframes les uns aux autres et de vérifier si toutes les zones du dataframe "Alimentation" sont bien présentes dans les résultats des jointures. Si certaines zones sont manquantes, nous chercherons la raison derrière cela.

► Cliquez ici pour afficher le contenu masqué

Lorsque nous avons étudié la population, nous avons la Chine, continentale et l'Inde qui avait une population de loin supérieur aux autres.

Zone	Année	Population	Inflation	Stabilité_politique	PIB	Production	Importations	Exportations	TAS	TDI	disponibilité_intérieure
Chine, continentale	2017	1421021791	-0.19	NA	8729.14	18236	452	576	100.68	2.5	18062
Inde	2017	1338676785	1.82	-0.76	1937.92	3545	0	4	100.11	0.0	3422

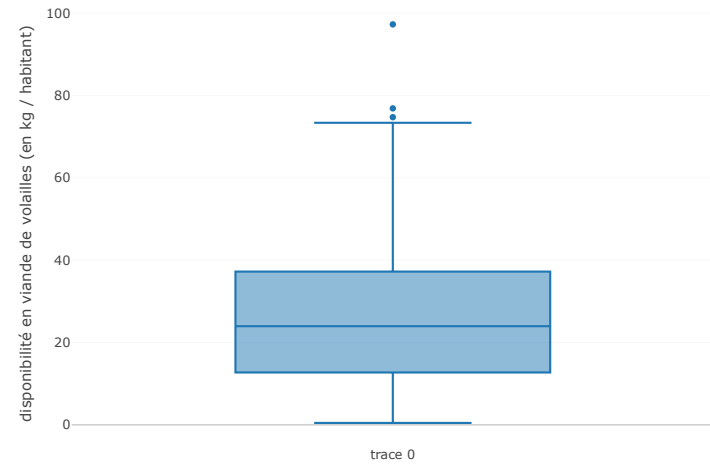
Les données montrent que l'Inde n'importe pas de poulet et est autosuffisante, ce qui suggère une forte dépendance à la production nationale. Nous décidons de supprimer l'Inde de l'étude.

Quant à la Chine, elle produit une grande quantité de poulet (18236 milliers de tonnes) et a des importations relativement élevées, ce qui indique une forte demande intérieure. Cependant, elle est autosuffisante et ne dépend pas de l'importation TDI=2.5%.

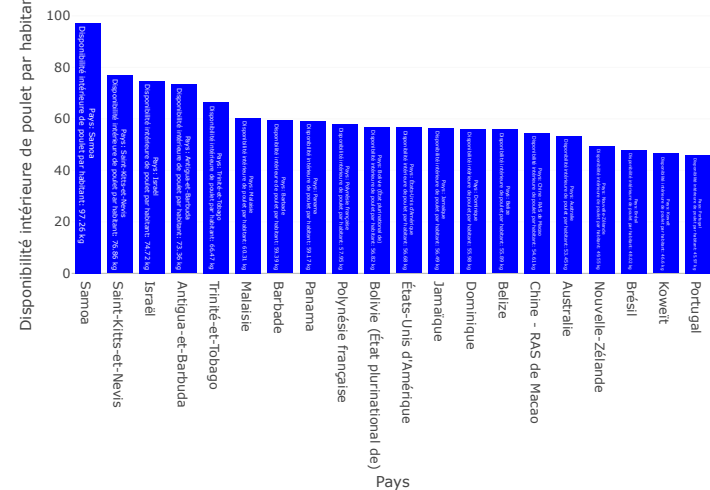
### 4.2 Nouvelle variable: disponibilité intérieure/ habitant

Nous allons utiliser ce nouveau dataframe final pour calculer la disponibilité intérieure par habitant, car cette mesure sera plus pertinente pour notre étude que la disponibilité intérieure brute. En prenant en compte la population de chaque zone, nous pourrions mieux évaluer la disponibilité réelle de poulet par personne dans chaque pays, ce qui nous permettra d'avoir des comparaisons plus significatives entre les différentes zones.

Box Plot de la disponibilité par habitant en viande de volailles



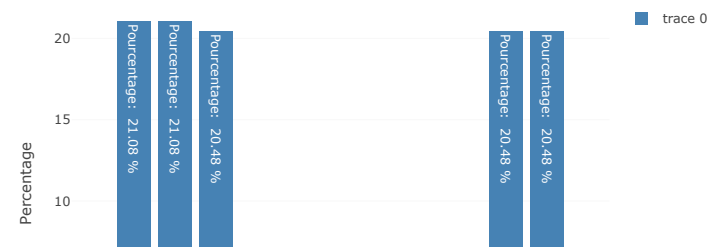
20 des pays avec la disponibilité intérieure de poulet par habitant la plus élevée

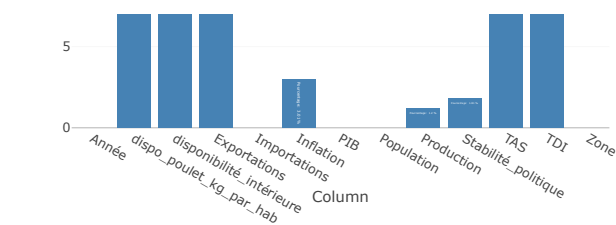


Si la disponibilité de poulet par habitant est élevée, cela peut indiquer une demande soutenue pour ce type de produit.

### 4.3 Valeurs manquantes

Pourcentages de valeurs manquantes

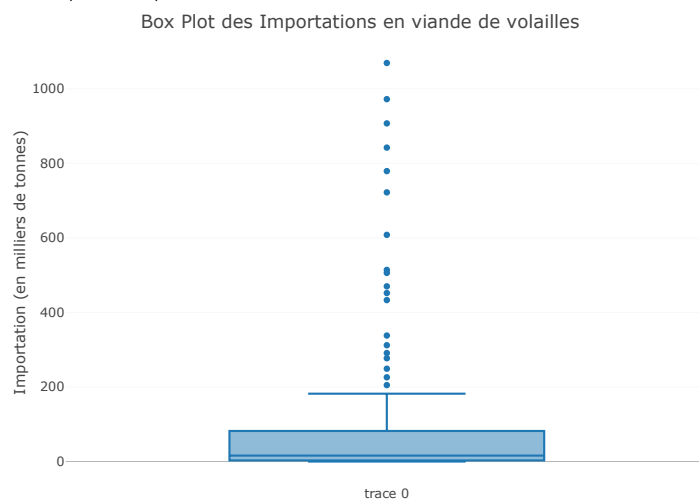




Ci-dessous le tableau récapitulatif des pays ayant des données manquantes.

Zone	Année	Population	Inflation	Stabilité_politique	PIB	Production	Importations	Exportations	TAS	TDI	disponibilité_intérieure	dispo_poulet_kg_par_hab
13 Bahamas	2017	381755	-0.12	0.99	30969.88	6	24	NA	NA	NA	NA	NA
14 Bangladesh	2017	159685424	7.17	-1.26	1785.96	249	0	NA	NA	NA	NA	NA
25 Burkina Faso	2017	19193234	2.04	-0.93	711.18	46	0	NA	NA	NA	NA	NA
26 Cabo Verde	2017	537498	-0.18	0.77	3132.62	1	12	NA	NA	NA	NA	NA
27 Cambodge	2017	16009409	3.42	0.09	1400.90	28	10	NA	NA	NA	NA	NA
33 Chine, continentale	2017	1421021791	-0.19	NA	8729.14	18236	452	576	100.68	2.50	18062	12.71
41 Cuba	2017	11339254	NA	0.65	8543.36	29	312	NA	NA	NA	NA	NA
43 Djibouti	2017	944099	1.99	-0.74	2659.85	NA	3	NA	NA	NA	NA	NA
53 Éthiopie	2017	106399924	11.29	-1.68	709.76	14	1	NA	NA	NA	NA	NA
59 Gambie	2017	2213889	8.72	-0.08	632.02	2	16	NA	NA	NA	NA	NA
63 Grenade	2017	110874	0.54	1.01	9309.27	1	7	NA	NA	NA	NA	NA
65 Guinée	2017	12067519	13.59	-0.70	843.46	13	37	NA	NA	NA	NA	NA
66 Guinée-Bissau	2017	1828145	-0.23	-0.50	718.25	3	4	NA	NA	NA	NA	NA
67 Guyana	2017	775222	NA	-0.04	6220.98	31	0	0	100.00	0.00	30	38.70
68 Haïti	2017	10982366	11.09	-0.67	1402.61	9	89	NA	NA	NA	NA	NA
85 Kiribati	2017	114158	1.80	0.88	1563.47	1	1	NA	NA	NA	NA	NA
94 Madagascar	2017	25570512	9.30	-0.30	503.50	81	0	NA	NA	NA	NA	NA
96 Malawi	2017	17670196	12.19	-0.25	500.17	130	0	NA	NA	NA	NA	NA
97 Maldives	2017	496402	5.68	0.23	10063.00	NA	12	NA	NA	NA	NA	NA
98 Mali	2017	18512430	2.20	-1.91	795.68	48	1	NA	NA	NA	NA	NA
102 Mauritanie	2017	4282570	3.10	-0.63	1634.64	5	24	NA	NA	NA	NA	NA
104 Mongolie	2017	3113786	4.51	0.79	3708.25	0	10	NA	NA	NA	NA	NA
106 Mozambique	2017	28649018	16.95	-0.93	462.70	92	24	NA	NA	NA	NA	NA
109 Népal	2017	27632681	0.36	-0.61	995.82	58	0	NA	NA	NA	NA	NA
114 Nouvelle-Calédonie	2017	277150	1.42	NA	32263.34	1	9	NA	NA	NA	NA	NA
122 Pérou	2017	31444298	3.02	-0.26	6676.31	1465	60	1	96.13	3.94	NA	NA
125 Polynésie française	2017	276102	1.82	NA	19743.96	1	15	0	6.25	93.75	16	57.95
127 République-Unie de Tanzanie	2017	54660339	9.13	-0.56	972.63	105	2	NA	NA	NA	NA	NA
128 République centrafricaine	2017	4596023	NA	-2.01	414.74	7	4	NA	NA	NA	NA	NA
132 République populaire démocratique de Corée	2017	25429825	NA	-0.50	680.53	43	0	0	100.00	0.00	43	1.69
137 Saint-Vincent-et-les Grenadines	2017	109827	1.90	0.89	8030.57	0	9	NA	NA	NA	NA	NA
138 Sainte-Lucie	2017	180954	-1.46	0.99	11270.81	1	10	NA	NA	NA	NA	NA
140 Sao Tomé-et-Principe	2017	207089	6.57	0.19	1819.87	1	2	NA	NA	NA	NA	NA
146 Soudan	2017	40813397	34.92	-1.98	3000.82	67	2	NA	NA	NA	NA	NA
151 Tadjikistan	2017	8880268	9.24	-0.67	844.37	2	38	NA	NA	NA	NA	NA
155 Timor-Leste	2017	1243258	0.95	0.07	1299.52	1	11	NA	NA	NA	NA	NA
159 Turkménistan	2017	5757667	NA	-0.13	6738.44	20	9	NA	NA	NA	NA	NA
163 Vanuatu	2017	285510	6.88	0.70	3032.15	1	4	NA	NA	NA	NA	NA
166 Zimbabwe	2017	14236595	2.53	-0.71	1494.17	69	6	NA	NA	NA	NA	NA

- Boxplot de la "importation" en viande de volailles



Parmi la liste des pays ayant des données manquantes, nous recherchons les pays qui pourraient nous intéresser pour notre marché, ceux ayant une valeur importation > q3 boxplot cad 82 milliers de tonnes.

Zone	Année	Population	Inflation	Stabilité_politique	PIB	Production	Importations	Exportations	TAS	TDI	disponibilité_intérieure	dispo_poulet_kg_par_hab
33 Chine, continentale	2017	1421021791	-0.19	NA	8729.14	18236	452	576	100.68	2.5	18062	12.71
41 Cuba	2017	11339254	NA	0.65	8543.36	29	312	NA	NA	NA	NA	NA
68 Haïti	2017	10982366	11.09	-0.67	1402.61	9	89	NA	NA	NA	NA	NA

Parmi les pays étudiés, la Chine continentale, Cuba et Haïti se démarquent comme des candidats intéressants pour notre marché en raison de leurs importations significatives de poulet.

Concernant Cuba, la disponibilité de poulet par habitant n'est pas indiquée, nous n'avons pas d'informations sur la demande potentielle. Haïti quand à elle présente une inflation élevée et une stabilité politique relativement faible. Cuba et Haïti ne sont pas intéressants pour notre étude.

Bien que la stabilité politique ne soit pas précisée, la Chine possède l'une des économies les plus importantes et en croissance au monde. Ils importent beaucoup et La disponibilité de poulet par habitant est également élevée, indiquant que la consommation de poulet par personne est importante. La Chine continentale pourrait être un marché attractif pour l'exportation de poulets.

Cependant, pour les analyses ultérieures que nous prévoyons de réaliser, telles que l'Analyse en Composantes Principales (ACP) ou le clustering, il est essentiel d'avoir des données complètes pour chaque pays. Par conséquent, nous avons pris la décision de supprimer les pays ayant des données manquantes.

En procédant ainsi, nous nous assurons d'avoir un ensemble de données complet pour mener à bien nos prochaines étapes d'analyse et d'étude de marché.

## 4.4 Dictionnaire des données

Variable	Descriptif	Unité
Population	Population totale de la zone.	Nombre de personnes
Stabilité politique	Indicateur de stabilité politique.	Mesure sans unité
PIB	Produit Intérieur Brut par habitant.	Dollars par habitant
Inflation	Taux d'inflation.	Pourcentage
Inflation décimale	Taux d'inflation décimal.	Décimal
Production	Production nationale de viande de volaille.	Milliers de tonnes
Importations	Quantité importée de viande de volaille.	Milliers de tonnes
Exportations	Quantité exportée de viande de volaille.	Milliers de tonnes
Variation de stock	Variation des stocks de viande de volaille.	Milliers de tonnes
TDI (Taux de Dépendance aux Importations %)	Pourcentage de dépendance aux importations de viande de volaille.	Pourcentage
TDI décimal	Taux de dépendance aux importations de viande de volaille décimal.	Décimal
TAS (Taux d'Auto-Suffisance %)	Pourcentage d'auto-suffisance en production de viande de volaille.	Pourcentage
TAS décimal	Taux d'autosuffisance en production de viande de volaille décimal.	Décimal
Disponibilité intérieure	Disponibilité intérieure.	Milliers de tonnes
Dispo_poulet_kg_par_hab	Disponibilité intérieure de poulet par habitant, calculée en divisant la disponibilité intérieure par la population.	Kilogrammes par habitant

## 5 Tests de normalité et de corrélation

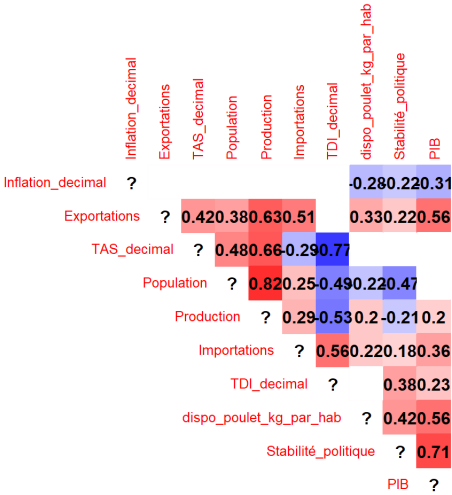
### 5.1 Normalité

Résultats des Tests de normalité	
	x
Population	Non-normal (p-value < 0.05)
Inflation_decimal	Non-normal (p-value < 0.05)
Stabilité_politique	Non-normal (p-value < 0.05)
PIB	Non-normal (p-value < 0.05)
Production	Non-normal (p-value < 0.05)
Importations	Non-normal (p-value < 0.05)
Exportations	Non-normal (p-value < 0.05)
TAS_decimal	Non-normal (p-value < 0.05)
TDI_decimal	Non-normal (p-value < 0.05)
dispo_poulet_kg_par_hab	Non-normal (p-value < 0.05)

Les variables ne suivent pas une distribution normale. Nous allons donc évaluer les relations entre les différentes variables avec le test de corrélation de Spearman.

### 5.2 Matrice de corrélation de Spearman

Pour vérifier la corrélation entre les variables, nous allons effectuer des tests de Spearman et regrouper les résultats sous forme de matrice. Ne sont affichés que les résultats statistiquement significatif p-value < 0.05 .



Cette visualisation des variables fortement corrélées nous indique qu'il existe une redondance dans les données.

Comme par exemple: la stabilité politique positivement corrélée au PIB, La population positivement corrélée à la production ou encore la production positivement corrélée au TAS et à l'exportation.

En raison donc de cette redondance, l'ACP pourra être utilisée pour réduire d'origine en un nombre plus petit de nouvelles variables.