DATA ANALYSIS

TAI

CLIMATE CHANGE AND ITS IMPACT ON NATURAL DISASTERS

Harish SELVAKUMAR / Artemiy Smogunov /
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai

Table of content

Introduction

In today's world, climate change stands as a pressing global concern, with its effects increasingly evident in our environment. One of the most notable consequences is the rise in both the frequency and severity of natural disasters such as storms, floods, droughts, and wildfires.
Our project aims to analyze the impact of climate change on these events through data analysis. By examining recent trends, we seek to understand how climate change influences natural disasters and identify patterns that can inform mitigation strategies. This endeavor underscores the importance of proactive measures to address the challenges posed by climate change and ensure a more resilient future for all.

1. Descriptive statistic of one and two variables

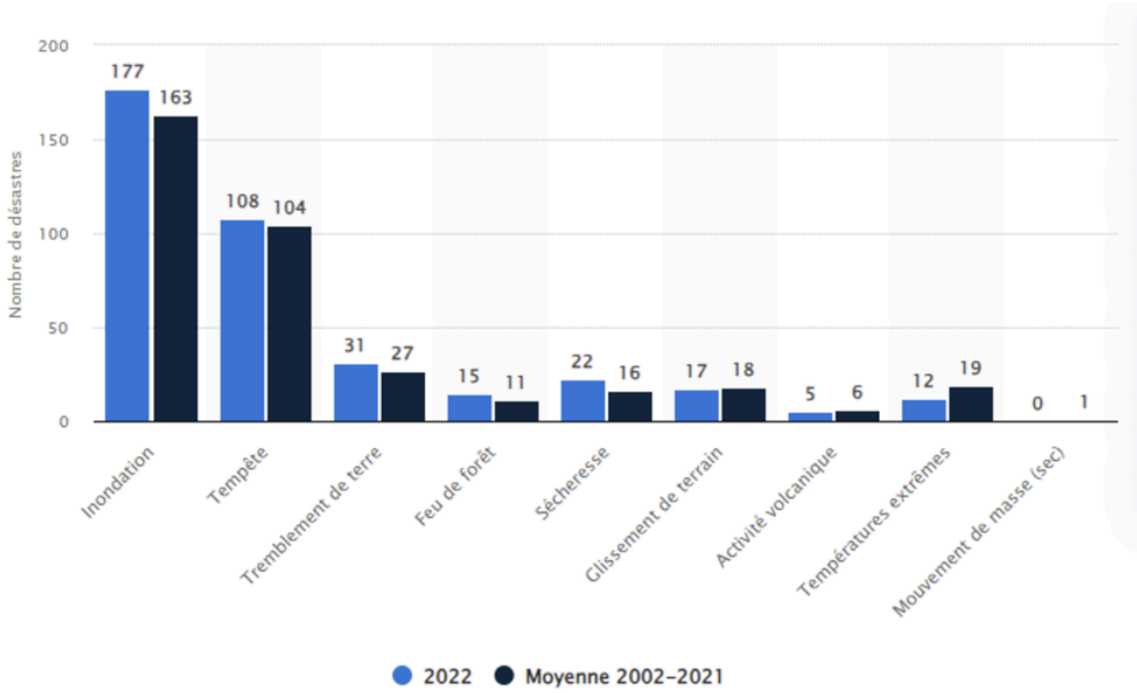
Year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Number of natural disasters	302	357	296	309	349	315	330	281	308	416

Calculation of the average of natural disasters over the last decade:

$$\bar{x} = \frac{1}{n}\sum_{i=0}^{n} x_i$$

—> X = 1/10 (302+357+…+416) = 326

Therefore, there were approximately 326 natural disasters per year from 2011 to 2020. From 2000 to 2010, the average number of natural disasters was around 300 per year, representing an increase of about 10%.
Using a database from the STATISTA website, we compare the number of natural disasters (ND) in 2022 to the average number of NDs from 2002 to 2021.



Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

Here, we observe that in the majority of cases, the year 2022 exceeds the average. Therefore, we will investigate whether temperature (climate change) is correlated with the number of natural disasters.
We understand that natural disasters are increasing every year, and we also know that Climate change is occurring. We will then try to see if these two phenomena are correlated.

2. Correlation between climate change and natural disasters

We will analyze a graph, extracted from the publication of Alternative Économique, which provides a valuable overview of the evolution of natural disasters, correlated with temperature trends observed over the past decades in less developed countries. As a renowned economic review, Alternative Économique relies on factual data and in-depth analysis to illustrate the impact of climate change on our environment and the global economy.
To verify the correlation between the number of natural disasters and the temperature change, we will calculate the covariance between these two variables. Covariance measures the joint variation of two random variables. In this context, it will allow us to assess whether the number of natural disasters and the temperature change tend to vary similarly, oppositely, or without any particular correlation over time.

$$Cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

N : number of periods
Xi : number of natural disasters
Yi : temperature of each year
$\bar{X}$: average of the number of natural disasters
$\bar{Y}$: average of the temperature

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

Year	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Nbr of natural disasters	0	13	8	10	30	35	57	60	105	100	90
Temperature (°C)	13.7	13.8	13.8	13.8	13.9	14	14.2	14.1	14.6	14.7	14.4

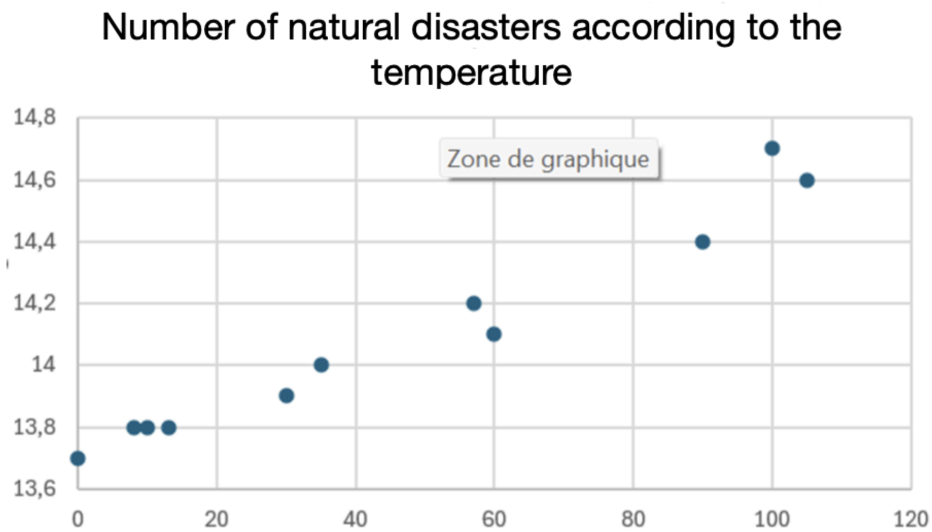
To calculate this covariance, we first need to compute the average number of natural disasters as well as the average temperature.

For this, we will use the formula for calculating the average: $\bar{x} = \frac{1}{n}\sum_{i=0}^{n} x_i$

The average number of natural disasters in the least developed countries is $\bar{x}$ = 46.2.
The average temperature is $\bar{y}$ = 14.1.

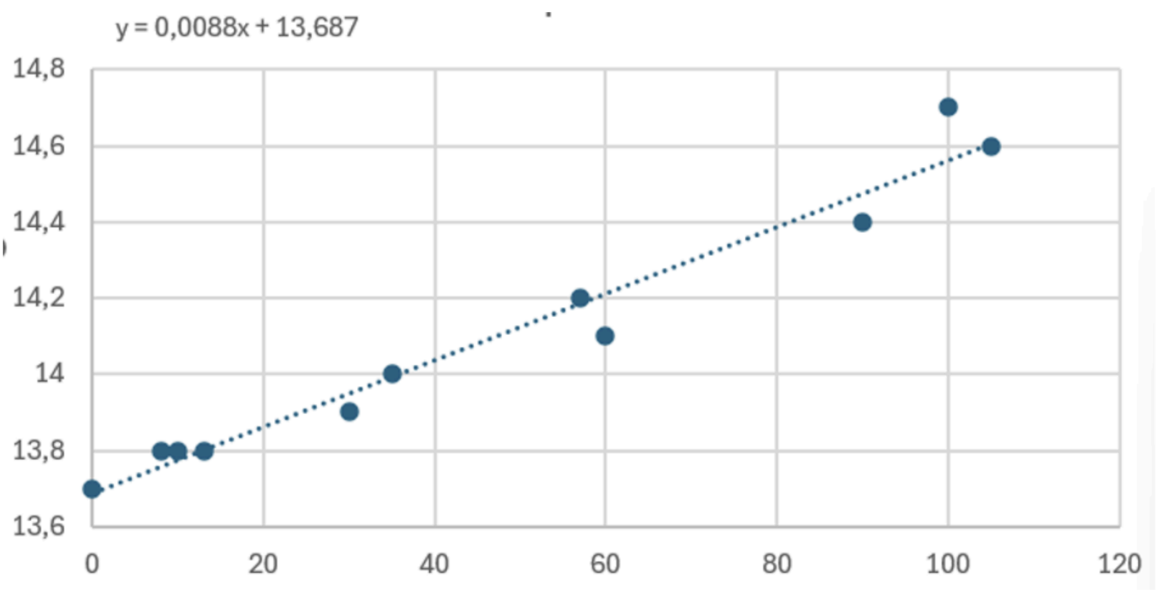
With these averages, we can now examine the covariance :

Cov(X,Y) = 1/11(0*13,7+13*13,8+…+90*14,4)-46,2*14,1 =11,91



Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

To study the evolution of the number of natural disasters according to temperature, a scatter plot was initially one of the primary methods. Through this and the least squares method, we were able to establish the regression line.



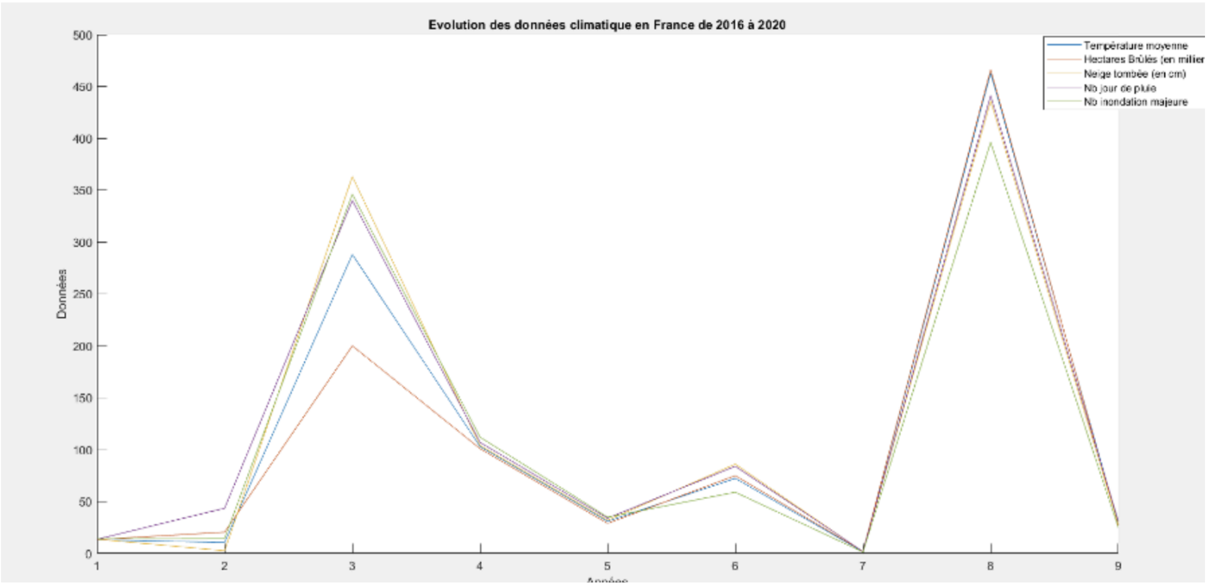
Regression line of Y with respect to X :
y = ax + b where :
a = $\sigma XY/\sigma^2 X$ and b = $\overline{y} - a\overline{x}$
Which gives us :
y = 0.0088x + 13.687

3. Correlation analysis of various data related to climate change

For this section, we sought various data related to climate change, weather, and natural disasters. We collected data from the most reliable sources available for the years 2016, 2017, 2018, 2019, and 2020. Subsequently, we created a data matrix with each parameter as a column and the results for each year as rows. Below are the gathered data along with their sources:

	2016	2017	2018	2019	2020	Sources
Température moyenne	13,2	13,4	13,9	13,7	14,1	meteofrance
hectare brûlé (k)	10,7	20,6	2,6	43,6	14,5	statista
cm de neige	288	200	363	340	346	skiinfo
nb jour de pluie	103	101	104	107	112	apc_paris
nb inondation	31	29	32	34	35	vigicrue
nb gens qui se déplace en voiture en %	72,4	75	86	84	59	statista
récolte agricole blé (M de Tonne)	1,73	2,1	1,67	1,95	1,47	insee
Emission GES (M de Tonne)	463	466	436	441	396	statista
nb météo extrême	29	27	28	31	25	meteofrance

And here is the graph representing the evolution of these over 5 years and the associated Matlab program:




```matlab
% Création d'une matrice
matriceDonnees = [13.2 10.7 288 103 31 72.4 1.73 463 29;
    13.4 20.6 200 101 29 75 2.1 466 27;
    13.9 2.6 363 104 32 86 1.67 436 28;
    13.7 43.6 340 107 34 84 1.95 441 31;
    14.1 14.5 346 112 35 59 1.47 396 25];

% Création du graphique
figure;
hold on;

% Boucle pour tracer chaque ligne de la matrice
for i = 1:size(matriceDonnees, 1)
    plot(matriceDonnees(i, :));
end

% Entrer les légendes pour chaque lignes
legendCell = {'Température moyenne', 'Hectares Brûlés (en millier)', 'Neige tombée (en cm)', 'Nb jour de pluie', 'Nb inondation majeure', 'Déplacement en voi

% Ajout de la légende
legend(legendCell);

xlabel('Années');
ylabel('Données');s
title('Evolution des données climatique en France de 2016 à 2020');

hold off; |
```

To study the correlation between the different data, we needed to create the correlation matrix, and thus calculate the standardized centered matrix and its transpose, then multiply them together using the following two formulas:

$$X^s = \left[\frac{x_{ij} - \overline{x_j}}{\sigma_j \sqrt{n}}\right]_{\substack{1 \le i \le n \\ 1 \le j \le p}} \qquad \mathrm{R} = {}^t X^s \times X^s$$

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

Given the size of the matrix, which involved numerous calculations and thus increased the risk of errors, we decided to write a MATLAB script that, starting from a given matrix, calculates the correlation matrix R.

```matlab
%Données éudiées
matriceDonnees = [13.2 10.7 288 103 31 72.4 1.73 463 29;
    13.4 20.6 200 101 29 75 2.1 466 27;
    13.9 2.6 363 104 32 86 1.67 436 28;
    13.7 43.6 340 107 34 84 1.95 441 31;
    14.1 14.5 346 112 35 59 1.47 396 25];

% Calcul de la moyenne de chaque colonne
moyennes_colonnes = mean(matriceDonnees);



% Calcul de l'écart-type de chaque colonne
%-------------------------------------------------------------

%Initialisation de la matrice contenant les ecarts-types
ecarts_types_colonnes = zeros(1, size(matriceDonnees, 2));

for i= 1 : size(matriceDonnees, 2)  %Parcour les colonnes
    variance = 0;  %Initialise la variance de la colonne à 0
    for j= 1 : size(matriceDonnees, 1)  %Parcour les lignes dans chaque colonne
        variance = variance + ((matriceDonnees(j,i) - moyennes_colonnes(i)) ^ 2);
    end
    variance = variance / size(matriceDonnees, 1);  %On divise notre valeur par le nombre de lignes
    ecart_type = sqrt(variance);    %On calcule la racine de la variance pour trouver l'ecart type
    ecarts_types_colonnes(i) = ecart_type;  % On assigne l'ecart type à la colonne où il appartient
end
%---------------------------------------------------------------


% Calcul de la matrice centrée réduite à partir de l'ensemble des autres données
matrice_centree_reduite = ((matriceDonnees - moyennes_colonnes) ./ (ecarts_types_colonnes * sqrt(size(matriceDonnees, 1))));


% Calcul de la transposée de la matrice centrée réduite
matrice_transposee_centree_reduite = transpose(matrice_centree_reduite);


% Multiplication des matrices transposée par centrée réduite
matrice_correlation = matrice_transposee_centree_reduite * matrice_centree_reduite;

% Affichage du produit
disp('Matrice de correlation');
disp(matrice_correlation);
```

Here is the result we obtained:

```
>> Projet_AnalyseDeDonnees
Matrice de correlation
```

temp moy	hectBrulé	cmNeige	nbPluie	nbInnondé	%trajVoit	récoltes	GESémis	meteoExtreme
1.0000	-0.0676	0.7086	0.7820	0.7580	-0.1809	-0.6026	-0.9179	-0.4292
-0.0676	1.0000	-0.0716	0.1713	0.2685	0.2024	0.5448	0.0803	0.5574
0.7086	-0.0716	1.0000	0.6600	0.8260	0.1183	-0.7049	-0.6886	0.1181
0.7820	0.1713	0.6600	1.0000	0.9449	-0.5443	-0.6910	-0.9419	-0.3398
0.7580	0.2685	0.8260	0.9449	1.0000	-0.2690	-0.6404	-0.8625	-0.0468
-0.1809	0.2024	0.1183	-0.5443	-0.2690	1.0000	0.5099	0.5027	0.7513
-0.6026	0.5448	-0.7049	-0.6910	-0.6404	0.5099	1.0000	0.7748	0.4864
-0.9179	0.0803	-0.6886	-0.9419	-0.8625	0.5027	0.7748	1.0000	0.5253
-0.4292	0.5574	0.1181	-0.3398	-0.0468	0.7513	0.4864	0.5253	1.0000

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

The closer the correlation between two elements is to 1, the more correlated they are; the closer it is to -1, the more they are negatively correlated; and the closer it is to 0, the less correlation they have.
Now, let's move on to the analysis of the results:
For temperature, we first observe that it has no correlation with the number of hectares burned or the percentage of car trips. However, it is strongly correlated with the number of centimeters of snowfall, the number of rainy days, and the number of days with significant floods. Greenhouse gas emissions are inversely correlated with temperature.
The number of hectares burned is slightly correlated with agricultural crops and extreme weather days.
The number of centimeters of snow is correlated with the number of rainy days and the number of days with floods, and negatively correlated with agricultural crops and greenhouse gas emissions.
The number of rainy days is logically strongly correlated with floods and strongly negatively correlated with greenhouse gas emissions. The same applies to flood days.
The percentage of car trips is moderately correlated with agricultural crops and greenhouse gas emissions, and is correlated with extreme weather days.
Agricultural crops are correlated with greenhouse gas emissions.

In conclusion, despite some inconsistencies, the results are satisfactory. These inconsistencies may arise from exceptional years, such as 2020 when COVID drastically reduced car trips and greenhouse gas emissions, and from the database containing only a few years, allowing exceptional years to have a stronger impact on the results.

4. PCA

```
donneesEtudiees = [13.2, 10.7, 288, 103, 31, 7.24, 72.4, 463, 29;
                   13.4, 20.6, 200, 101, 29, 75, 2.1, 466, 27;
                   13.9, 2.6, 363, 104, 32, 86, 1.67, 436, 28;
                   13.7, 43.6, 340, 107, 34, 84, 1.95, 441, 31;
                   14.1, 14.5, 346, 112, 35, 59, 1.47, 396, 25];
```

We need to centered-reduced the data because : different Units, orders of magnitude and different standard deviations :

```
Xs =

   -0.5641   -0.2218   -0.1310   -0.2509   -0.2248   -0.7559    0.8000    0.3598    0.2000
   -0.3188    0.0634   -0.7250   -0.4600   -0.5994    0.1752   -0.1957    0.4075   -0.2000
    0.2943   -0.4552    0.3753   -0.1464   -0.0375    0.3264   -0.2018   -0.0700         0
    0.0491    0.7259    0.2201    0.1673    0.3372    0.2989   -0.1978    0.0096    0.6000
    0.5396   -0.1123    0.2606    0.6900    0.5245   -0.0446   -0.2046   -0.7068   -0.6000
```

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

Now that we have the correlation matrix, we will find the eigenvalues with matlab through the following code :

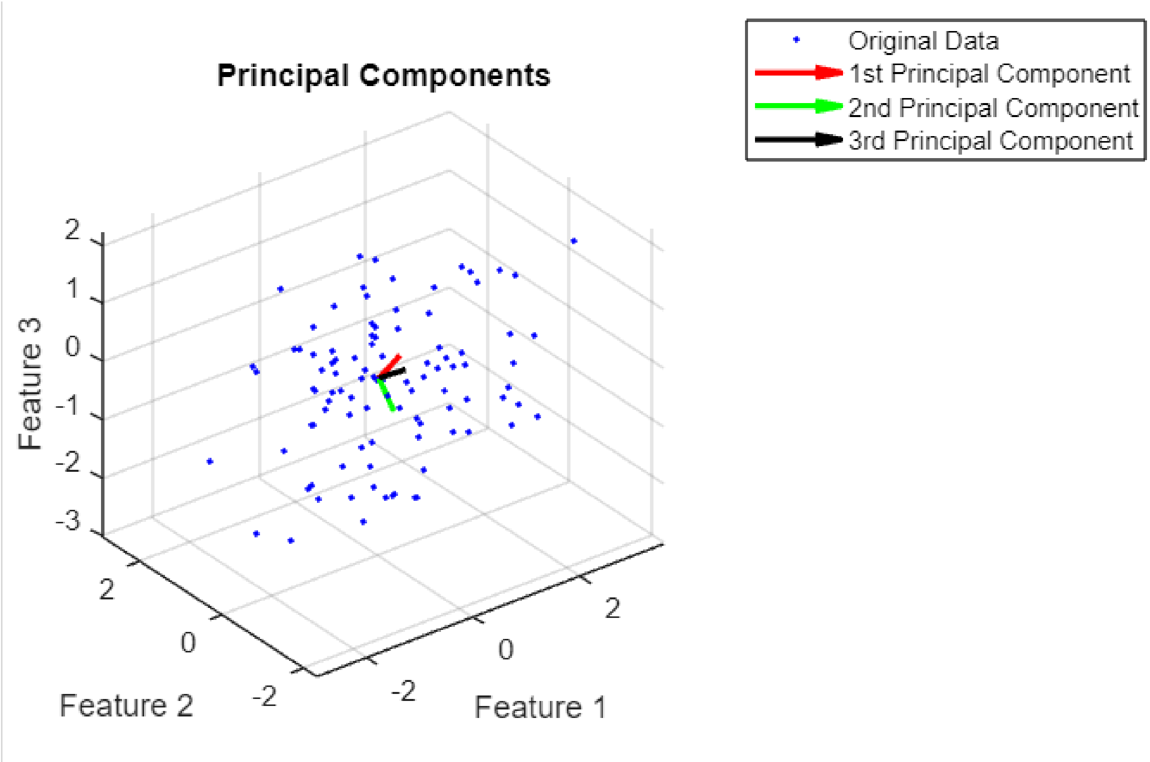
```
correlation_matrix = [1.0000, -0.0676, 0.7086, 0.7820, 0.7580, -0.1809, -0.6026, -0.9179, -0.4292;
                      -0.0676, 1.0000, -0.0726, 0.1713, 0.2685, 0.2024, 0.5448, 0.0803, 0.5574;
                      0.7086, -0.0716, 1.0000, 0.6600, 0.8260, 0.1183, -0.7049, -0.6886, 0.1181;
                      0.7820, 0.1713, 0.6600, 1.0000, 0.9449, -0.5443, -0.6910, -0.9419, -0.3398;
                      0.7580, 0.2685, 0.8260, 0.9449, 1.0000, -0.2690, -0.6404, -0.8625, -0.0468;
                      -0.1809, 0.2024, 0.1183, -0.5443, -0.2690, 1.0000, 0.5099, 0.5027, 0.7513;
                      -0.6026, 0.5448, -0.7049, -0.6910, -0.6404, 0.5099, 1.0000, 0.7748, 0.4864;
                      -0.9179, 0.0803, -0.6886, -0.9419, -0.8625, 0.5027, 0.7748, 1.0000, 0.5253;
                      -0.4292, 0.5574, 0.1181, -0.3398, -0.0468, 0.7513, 0.4864, 0.5253, 1.0000];


% Compute eigenvalues
eigenvalues = eig(correlation_matrix);

disp('Eigenvalues:');
disp(eigenvalues);
```

We ended up with these eigenvalues :

```
>> TAI
Eigenvalues:
   5.2019 + 0.0000i
   2.1192 + 0.0000i
   1.1579 + 0.0000i
   0.5211 + 0.0000i
  -0.0001 + 0.0000i
  -0.0001 - 0.0000i
  -0.0001 + 0.0000i
   0.0000 + 0.0000i
   0.0001 + 0.0000i
```

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

```
Eigenvalues:
   5.2019 + 0.0000i   2.1192 + 0.0000i   1.1579 + 0.0000i   0.5211 + 0.0000i  -0.0001 - 0.0000i  -0.0001 + 0.0000i  -0.0001 + 0.0000i   0.0001 + 0.0000i   0.0000 + 0.0000i
Eigenvectors:
   0.3823 + 0.0000i  -0.0966 + 0.0000i   0.0781 + 0.0000i  -0.6394 + 0.0000i  -0.0498 - 0.0956i  -0.0498 + 0.0956i  -0.1840 + 0.0000i   0.0035 + 0.0000i  -0.3774 + 0.0000i
  -0.0667 + 0.0000i  -0.5029 + 0.0000i  -0.6158 + 0.0000i  -0.0529 + 0.0000i  -0.5504 + 0.0712i  -0.5504 - 0.0712i  -0.1798 + 0.0000i   0.4694 + 0.0000i   0.3020 + 0.0000i
   0.3276 + 0.0000i  -0.3256 + 0.0000i   0.4225 + 0.0000i   0.1428 + 0.0000i  -0.0695 + 0.0745i  -0.0695 - 0.0745i  -0.0165 + 0.0000i  -0.0943 + 0.0000i  -0.0327 + 0.0000i
   0.4116 + 0.0000i  -0.1182 + 0.0000i  -0.2677 + 0.0000i   0.1093 + 0.0000i   0.1671 - 0.1610i   0.1671 + 0.1610i  -0.3962 + 0.0000i  -0.5518 + 0.0000i   0.3118 + 0.0000i
   0.3845 + 0.0000i  -0.3071 + 0.0000i  -0.1111 + 0.0000i   0.1787 + 0.0000i   0.2397 + 0.0753i   0.2397 - 0.0753i   0.7913 + 0.0000i  -0.1785 + 0.0000i  -0.0970 + 0.0000i
  -0.2241 + 0.0000i  -0.4279 + 0.0000i   0.4804 + 0.0000i  -0.4013 + 0.0000i  -0.1537 + 0.0143i  -0.1537 - 0.0143i   0.0280 + 0.0000i   0.0480 + 0.0000i   0.5269 + 0.0000i
  -0.3764 + 0.0000i  -0.1716 + 0.0000i  -0.3083 + 0.0000i  -0.4162 + 0.0000i   0.6002 + 0.0000i   0.6002 + 0.0000i   0.2859 + 0.0000i  -0.5517 + 0.0000i  -0.3604 + 0.0000i
  -0.4326 + 0.0000i   0.0069 + 0.0000i   0.0800 + 0.0000i   0.1903 + 0.0000i  -0.2351 - 0.0781i  -0.2351 + 0.0781i   0.0428 + 0.0000i  -0.3566 + 0.0000i   0.0901 + 0.0000i
  -0.2140 + 0.0000i  -0.5576 + 0.0000i   0.1411 + 0.0000i   0.3919 + 0.0000i   0.3087 - 0.1104i   0.3087 + 0.1104i  -0.2570 + 0.0000i  -0.0253 + 0.0000i  -0.4929 + 0.0000i
```



From these eigenvalues we will choose the first three ones which are relatively large compared to the rest. Let's compute their overall quality of explanation, which is the part of information kept when we choose these 3 axis :

$$\frac{\lambda_i}{\text{nbr eigenvalues}} \times 100$$

$\lambda_1 = 5.2019$ $\qquad \frac{5.2019}{9} \times 100 = 57,8\ \%$ of contribution

$\lambda_2 = 2.1192$ $\qquad \frac{2.1192}{9} \times 100 = 23,55\ \%$ " "

$\lambda_3 = 1.1579$ $\qquad \frac{1.579}{9} \times 100 = 12,87\ \%$ " "

$OQE = 57,8 + 23,55 + 12,87 = 94,22\ \%$

Loss of info : $5,78\ \% < 20\ \%$

Principal Component Matrix:

-0.7637 + 0.0000i	0.3845 + 0.0000i	-0.4234 + 0.0000i
-0.8825 + 0.0000i	0.5463 + 0.0000i	-0.0317 + 0.0000i
0.2242 + 0.0000i	0.0016 + 0.0000i	0.7186 + 0.0000i
0.1159 + 0.0000i	-0.9932 + 0.0000i	-0.1424 + 0.0000i
1.3060 + 0.0000i	0.0608 + 0.0000i	-0.1211 + 0.0000i

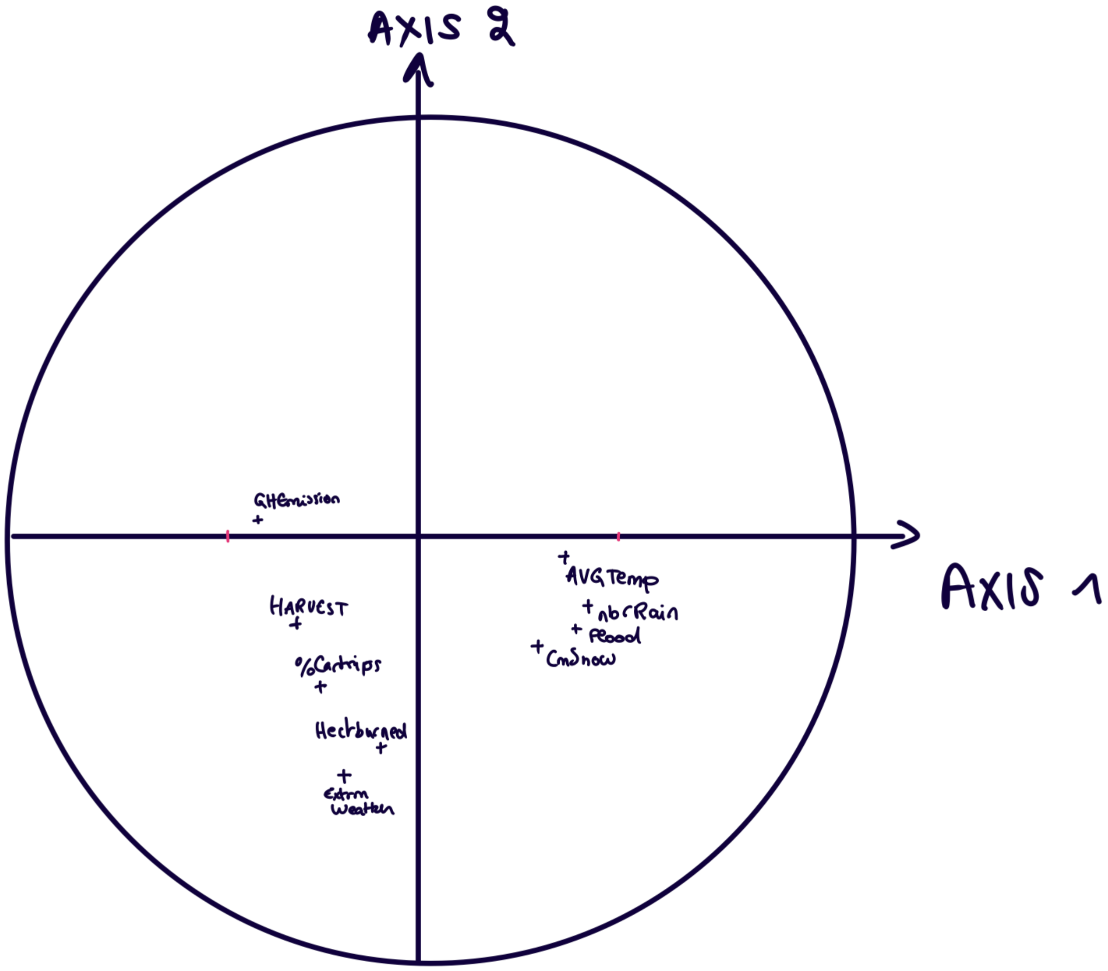
Now that we chose the axis that we will focus on, we need to create the saturation matrix with the eigenvectors. Again, we used matlab to compute it for us with the following code :

```
>> TAI
Saturation matrix (loading matrix):
```

avg temp	0.3823	-0.0966	0.0781	-0.6394	-0.0498	-0.0498	-0.1840	-0.3774	0.0035
Hect burned	-0.0667	-0.5029	-0.6158	-0.0529	-0.5504	-0.5504	-0.1798	0.3020	0.4694
cmSnow	0.3276	-0.3256	0.4225	0.1428	-0.0695	-0.0695	-0.0165	-0.0327	-0.0943
nbrRain	0.4116	-0.1182	-0.2677	0.1093	0.1671	0.1671	-0.3962	0.3118	-0.5518
Flood	0.3845	-0.3071	-0.1111	0.1787	0.2397	0.2397	0.7913	-0.0970	-0.1785
%carTrips	-0.2241	-0.4279	0.4804	-0.4013	-0.1537	-0.1537	0.0280	0.5269	0.0480
Harvest	-0.3764	-0.1716	-0.3083	-0.4162	0.6002	0.6002	0.2859	-0.3604	-0.5517
GHEmission	-0.4326	0.0069	0.0800	0.1903	-0.2351	-0.2351	0.0428	0.0901	-0.3566
ExtremeMeteo	-0.2140	-0.5576	0.1411	0.3919	0.3087	0.3087	-0.2570	-0.4929	-0.0253

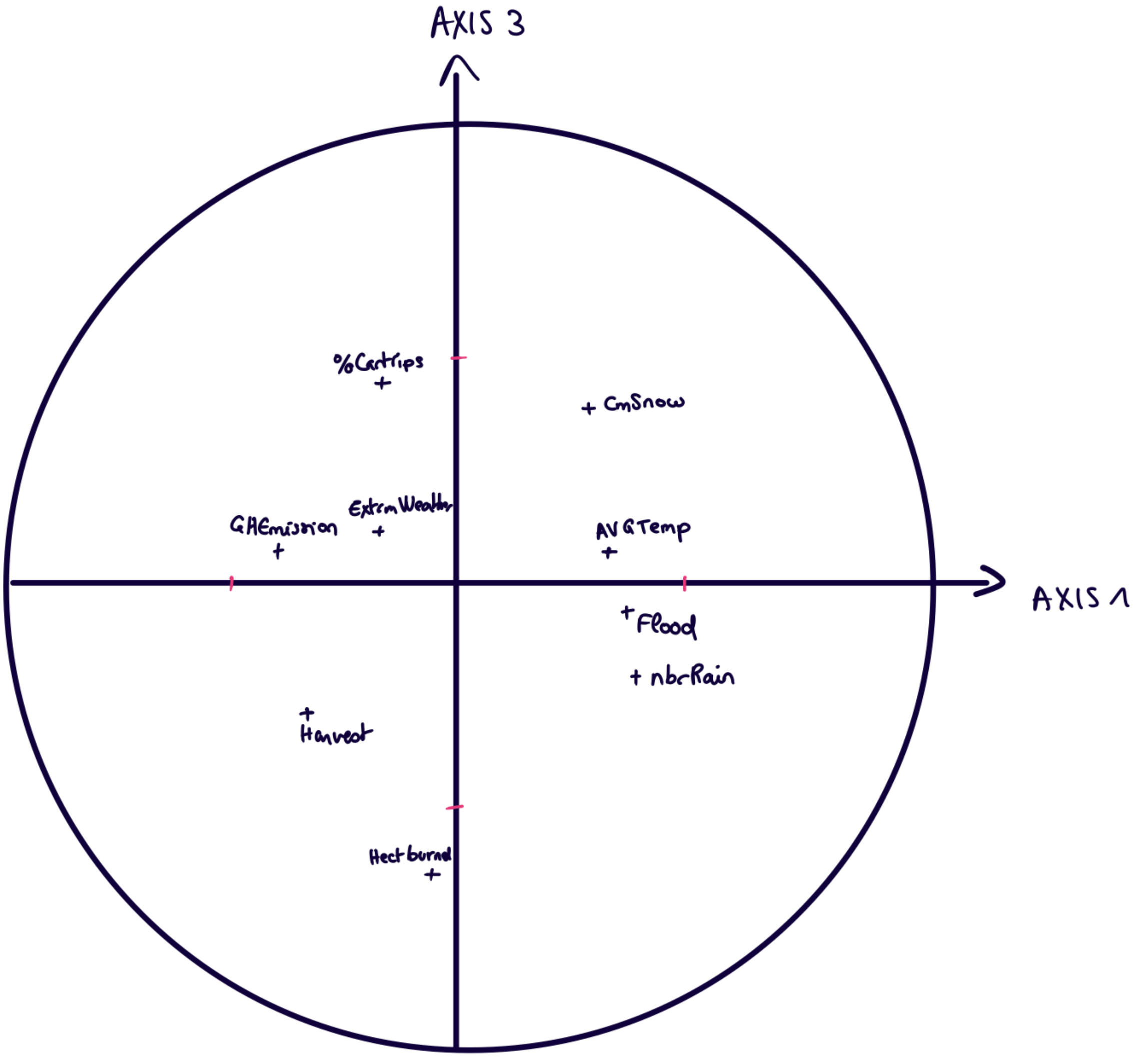
With the saturation matrix found, we can now analyze the first 3 axes. To analyze them we created correlation circles. One correlation circle for each plan.

Plan 1-2



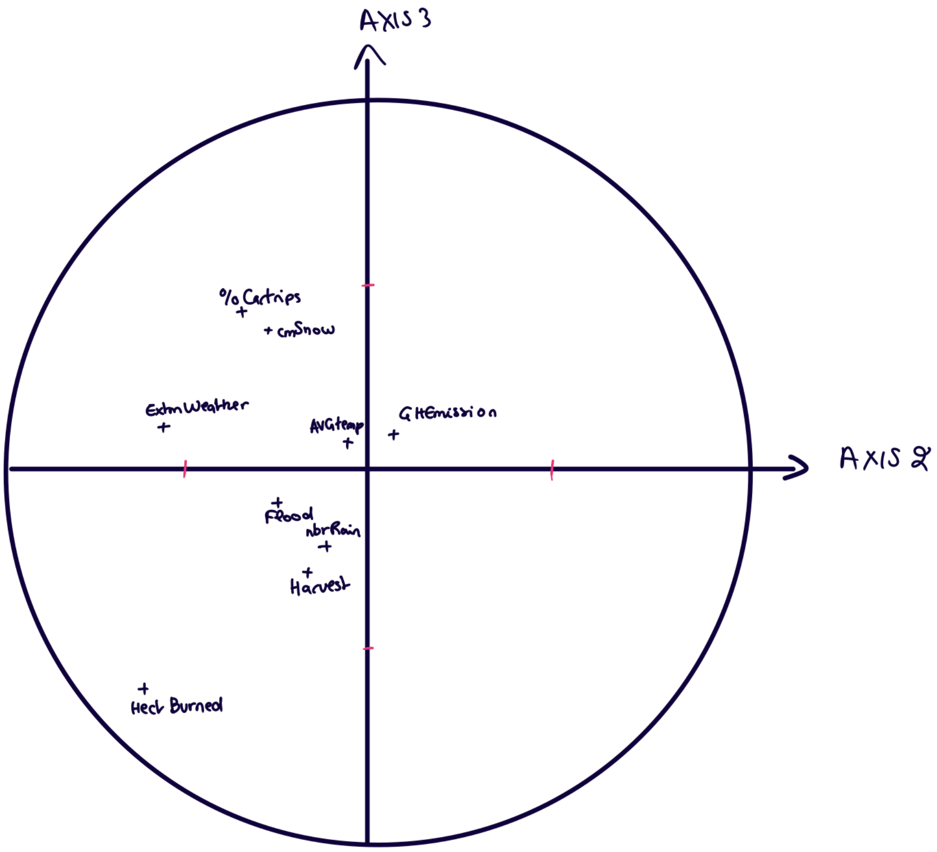
Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

PLAN 1-3 :



Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

PLAN 2-3



AXIS 1 : (+) Flood, NbrRain, AVGTemp, CmSnow

(−) GHEmission

AXIS 2 : (+) /

(−) ExtremeWeather, HectBurned, %Cartrips

AXIS 3 : (+) %Cartrips, cmSnow

(−) HectBurned

AXIS 1 : Based on the interpretation of the axis and its correlation with the variables, we can suggest a name that captures its characteristics. Since the axis is positively correlated with environmental conditions such as rainfall, temperature, snow cover, and flood occurrence, while negatively correlated with greenhouse gas emissions, Environmental Sustainability Axis.

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2

AXIS 2 : Based on the interpretation of the axis and its correlation with the variables, we can suggest a name that captures its characteristics. Since axis 2 is negatively correlated with ExtremeWeather, HectBurned, and %Cartrips, and not positively correlated with any variable, it represents a combination of factors related to vulnerability to extreme weather events, land area burned, and the percentage of car trips. Given these associations, we could name it something like "Vulnerability and Disaster Risk Axis," highlighting its role in assessing vulnerability to disasters and extreme weather events. This name encapsulates the axis's relevance to understanding vulnerability and disaster risk based on the variables included in the analysis.

AXIS 3 : Based on the interpretation of the axis and its correlation with the variables, we can suggest a name that captures its characteristics. Since axis 3 is positively correlated with %Cartrips and CMSnow, and negatively correlated with HectBurned, it represents a combination of factors related to transportation patterns, snow cover, and the extent of land area burned. We could name it something like "Transportation and Snow Cover Axis".

With the quality we have computed :

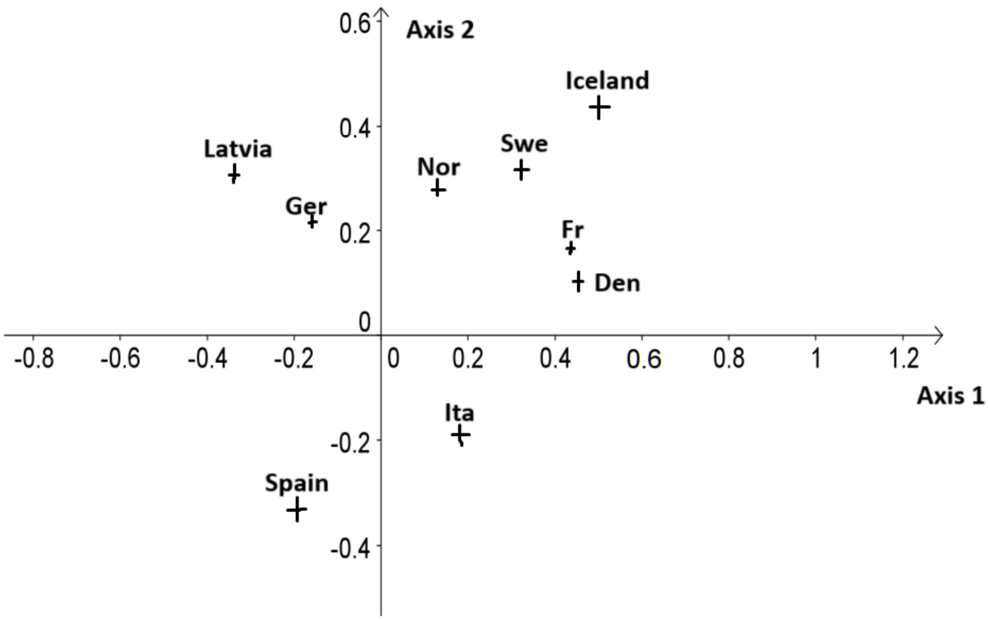
Coordinate = Score
Quality = qlt

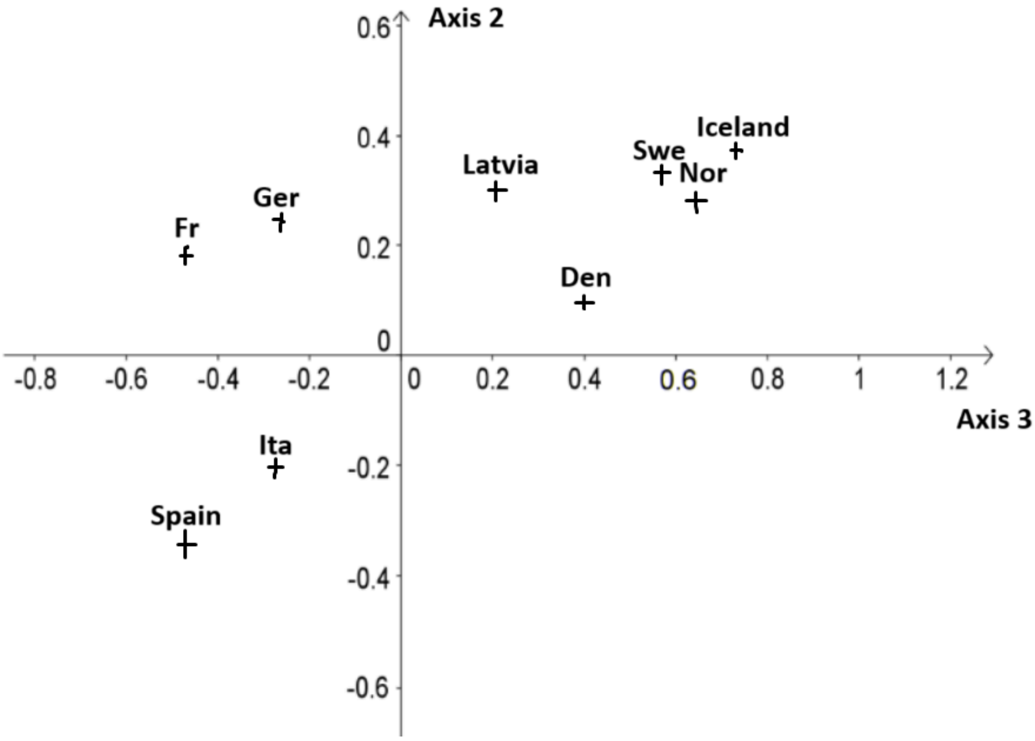
```
% Perform PCA
[~, score, ~, ~, ~] = pca(data);

% Compute quality (cosine squared)
qlt = (score.^2) ./ sum(score.^2);

% Display results
fprintf('qlt axis 1 \tqlt axis 2.\tqlt axis 3\n');
for i = 1:size(data, 1)
    fprintf('%.3f\t%.3f\t%.3f\t\n', ...
        qlt(1),qlt(2),qlt(3));
end
```

	qlt axis 1	qlt axis 2.	qlt axis 3
	0.033	0.075	0.111
France	0.631	0.028	0.002
Italy	0.154	0.114	0.266
Sweden	0.049	0.131	0.620
Latvia	0.133	0.652	0.000
Norway	0.266	0.130	0.610
Spain	0.131	0.164	0.275
Iceland	0.033	0.075	0.749

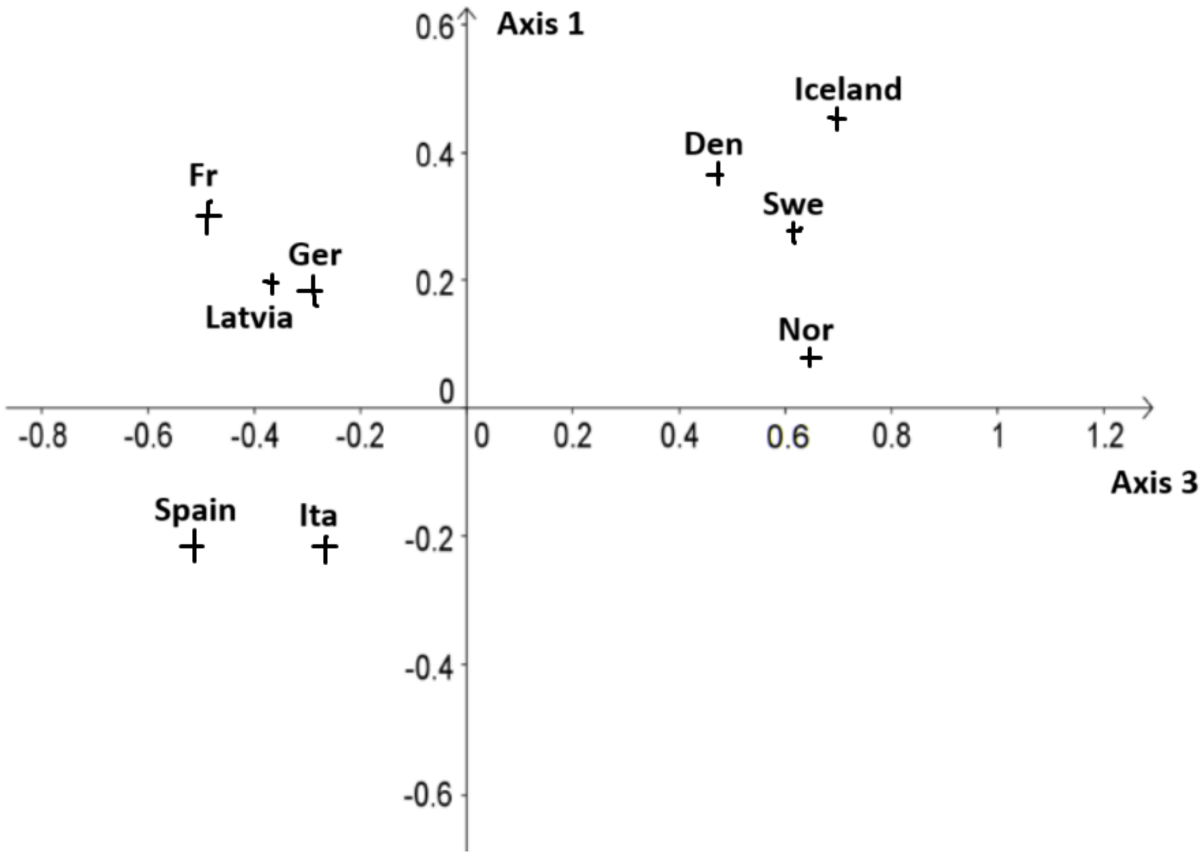
That the axis 1 represents well France
The axis 2, represents well Latvia
The axis 3, represents well Norway, Sweden and Iceland

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2





Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2



Conclusion

In conclusion, our data analysis confirms the clear link between climate change and the increase in natural disasters. The examined data reveal a significant correlation between rising temperatures and the growing incidence of extreme weather phenomena such as storms, floods, droughts, and wildfires. These results underscore the urgency of taking action to mitigate the effects of climate change and to strengthen resilience in the face of these devastating events.

Artemiy Smogunov / Harish
Nissanth / Vincent Ly
Tristan Pichard /Iness Bennai
L2-int2