# Environmental Sustainability Analysis Using the YALE-EPI Dataset and Composite indicator creation

Author: Rodions Barannikovs

1. Introduction

This project aimed at analyzing the environmental sustainability performance of countries using the Yale Environmental Performance Index (YALE-EPI) dataset. The dataset provides yearly environmental indicators from 2013 to 2022, covering aspects like air quality, biodiversity, and climate impact. My goal was to preprocess the data, explore relationships between indicators, reduce dimensionality, cluster countries based on environmental performance, and construct a composite indicator to rank countries by sustainability.

## Data Loading and Preprocessing

### 2.1

I loaded the YALE-EPI dataset (YALE-EPI.xlsx) and calculated 10-year averages

(2013–2022) for 13 key indicators, such as PM2.5 exposure and greenhouse gas

emissions. I handled missing values using linear interpolation and replaced remaining missing or zero values with medians to ensure data completeness. This step was crucial to create a dataset for analysis.

The original YALE-EPI dataset had several issues that made it difficult to use directly for composite index construction and analysis.

Indicators were listed as rows for each country instead of columns, making it unsuitable for multivariate and cluster analysis.

* Data was spread across multiple years in columns, which added unnecessary complexity.
* There were missing values in several indicators for some countries.
* The format was not normalized, and indicators had inconsistent units or scales.
* Extra metadata columns (e.g., Attribute fields, Partner) were not relevant for analysis.

To clean and restructure the dataset for use in the CA project, the following changes were made:

* Pivoted the dataset so each row represents a single country, and each column represents a single indicator.
* Removed irrelevant metadata columns.
* Identified missing values for later imputation or exclusion.
* Prepared the dataset for normalization, sub-index creation, and visualization.

**Data cleaning**

Removed the unused columns like Economy ISO3, attributes 1,2,3

Pivot the table so we only have individual countries as rows

In order to make the dataset more manageable I need to drop all the aggregate indices (e.g., Environmental Health, Air Quality) and redundant variables (e.g., NOx Exposure, Nitrous Oxides Growth Rate) to minimize multicollinearity for multivariate analysis.

The initial dataset consists of initial list of 53 variables

**Variable Selection**

To construct a composite indicator for environmental sustainability for my CA project, I started with the 53 variables in the Environmental Performance Index (EPI) dataset. After evaluation, I reduced the dataset to 13 key variables, dropping 40 others to eliminate redundancy and maintain data quality. The following sections detail the rationale for dropping variables, grouped by thematic areas, and explain the selection of retained variables.

**Dropping Aggregate Indices**

I dropped 13 aggregate indices: Environmental Performance Index, Environmental Health, Ecosystem Vitality, Climate Change, Air Quality, Biodiversity & Habitat, Ecosystem Services, Fisheries, Agriculture, Waste Management, Water Resources, Sanitation & Drinking Water, and Heavy Metals. These precomputed composites aggregate component indicators, making them redundant for building a new index from individual variables. Including them would introduce circularity and inflate correlations, complicating multivariate analysis and clustering. Focusing on underlying components ensured a granular, independent dataset.

**Streamlining Air Quality Variables**

From seven air quality variables, I retained PM2.5 Exposure, SO2 Exposure, Household Solid Fuels, and Ozone Exposure, dropping NOx Exposure, CO Exposure, and VOC Exposure. The dropped variables overlapped conceptually with PM2.5 and SO2 Exposure, which are widely studied, have significant health impacts, and offer robust data. For example, NOx Exposure measures similar air pollution effects, risking multicollinearity in regression and PCA. Ozone Exposure was retained for its unique contribution to photochemical pollution, particularly in urban areas, enhancing clustering and visualization. Household Solid Fuels was kept for its focus on indoor air pollution, ensuring comprehensive air quality coverage.

**Simplifying Climate-Related Variables**

In the climate category, I retained Adjusted Emissions Growth Rate for CO2, Adjusted Emissions Growth Rate for Methane, and Greenhouse Gas Emissions per Capita, dropping Adjusted Emissions Growth Rate for Nitrous Oxide, F-Gases, Black Carbon, Sulfur Dioxide, Nitrous Oxides, Greenhouse Gas Intensity Growth Rate, Projected GHG Emissions in 2050, Climate Change Policy Objective, and Acidification. The dropped variables were redundant (e.g., Sulfur Dioxide overlaps with SO2 Exposure), niche (e.g., F-Gases, Black Carbon), subjective (e.g., Climate Change Policy Objective), or uncertain (e.g., Projected GHG Emissions in 2050). Retaining CO2, Methane, and per capita emissions ensures robust climate trend coverage, suitable for clustering and visualization.

**Refining Water and Waste Variables**

For water and waste, I retained Unsafe Drinking Water, Unsafe Sanitation, Ocean Plastics, and Recycling, dropping Wastewater Treatment and Controlled Solid Waste. Wastewater Treatment and Controlled Solid Waste overlapped with Unsafe Sanitation and Ocean Plastics, which directly address health and environmental impacts. Recycling was retained to capture waste management practices and circular economy efforts, distinct from Ocean Plastics, enhancing the dataset’s policy relevance. This selection maintains essential health and waste metrics, ensuring interpretability for visualization.

**Consolidating Biodiversity and Ecosystem Variables**

In biodiversity and ecosystems, I kept Biodiversity Habitat Index, Terrestrial Biome Protection (Global Weights), and Tree Cover Loss, dropping Species Habitat Index, Species Protection Index, Marine Protected Areas, Protected Areas Representativeness Index, Terrestrial Biome Protection (National Weights), Grassland Loss, Wetland Loss, and Growth Rate in CO2 Emissions from Land Cover. The dropped variables were redundant (e.g., Species Habitat Index overlaps with Biodiversity Habitat Index) or specific (e.g., Marine Protected Areas). Grassland Loss, Wetland Loss, and CO2 Emissions from Land Cover were dropped as Tree Cover Loss broadly captures land-use impacts, ensuring a focused set of ecosystem variables.

**Excluding Fisheries and Agriculture Variables**

I dropped all fisheries variables (Fish Stock Status, Fish Caught by Trawling, Marine Trophic Index) and agriculture variables (Sustainable Nitrogen Management Index, Sustainable Pesticide Use). These niche variables are potentially data-sparse and add complexity without broad environmental impact. Ecosystem effects from agriculture and fisheries are indirectly captured by Tree Cover Loss and Biodiversity Habitat Index, streamlining the dataset while maintaining ecosystem health coverage.

**Retaining Heavy Metals Variable**

I retained Lead Exposure, the sole Heavy Metals variable, for its specific, health-relevant environmental and public health implications. No variables were dropped here, as Lead Exposure adds a unique dimension without redundancy.

By reducing the EPI dataset to 13 variables—PM2.5 Exposure, SO2 Exposure, Household Solid Fuels, Ozone Exposure, Unsafe Drinking Water, Unsafe Sanitation, Lead Exposure, Biodiversity Habitat Index, Terrestrial Biome Protection (Global Weights), Tree Cover Loss, Ocean Plastics, Recycling, and Greenhouse Gas Emissions per Capita—I created a streamlined dataset covering health, ecosystems, waste, and climate.

**Final Selected Variables**

1. **PM2.5 Exposure**: Key air quality metric, high health impact, widely measured.
2. **SO2 Exposure**: Complements PM2.5, captures industrial pollution.
3. **Household Solid Fuels**: Unique indoor air pollution focus, health-relevant.
4. **Ozone Exposure**: Captures photochemical pollution, enhances urban air quality insights.
5. **Unsafe Drinking Water**: Critical for health, globally comparable.
6. **Unsafe Sanitation**: Essential health metric, complements drinking water.
7. **Lead Exposure**: Specific heavy metal pollution, health-relevant.
8. **Biodiversity Habitat Index**: Broad measure of habitat quality.
9. **Terrestrial Biome Protection (Global Weights)**: Represents land-based conservation.
10. **Tree Cover Loss**: Major indicator of deforestation, climate/biodiversity link.
11. **Ocean Plastics**: Emerging waste issue, policy-relevant.
12. **Recycling**: Reflects waste management and circular economy efforts.
13. **Greenhouse Gas Emissions per Capita**: Absolute emissions, key for country comparisons.

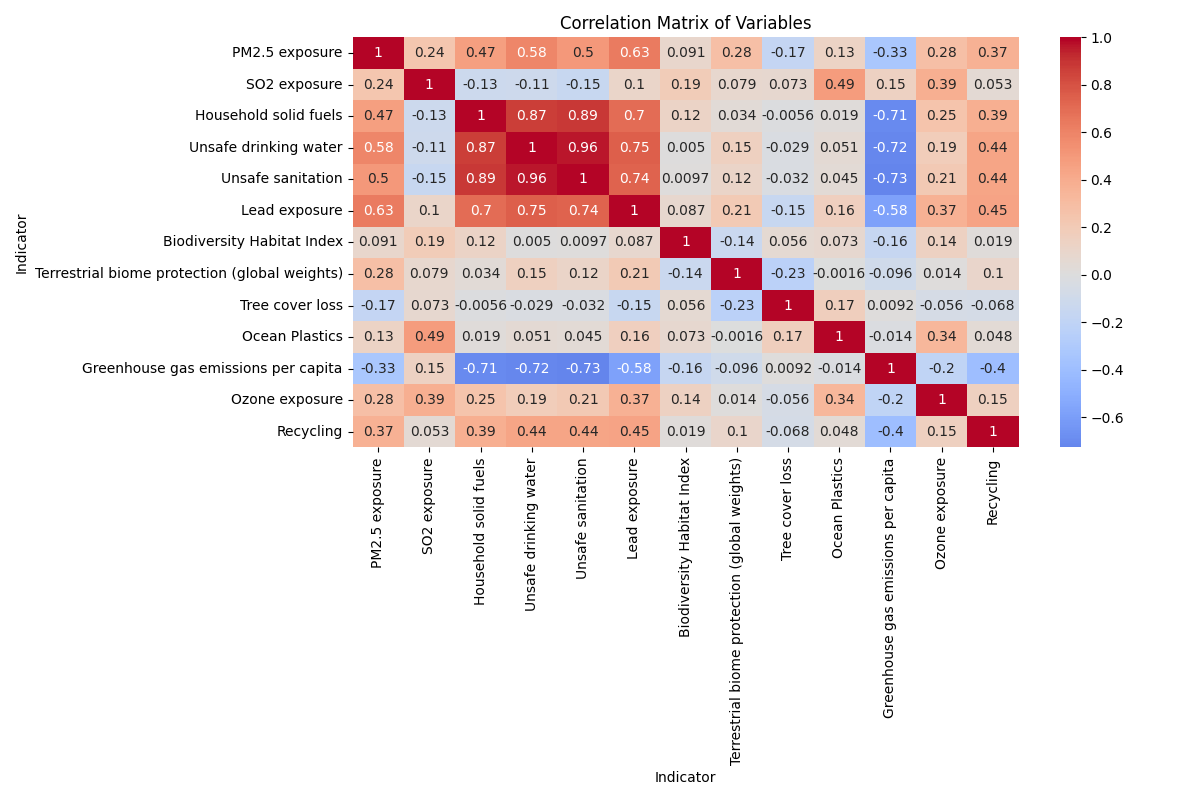
## Data Cleaning

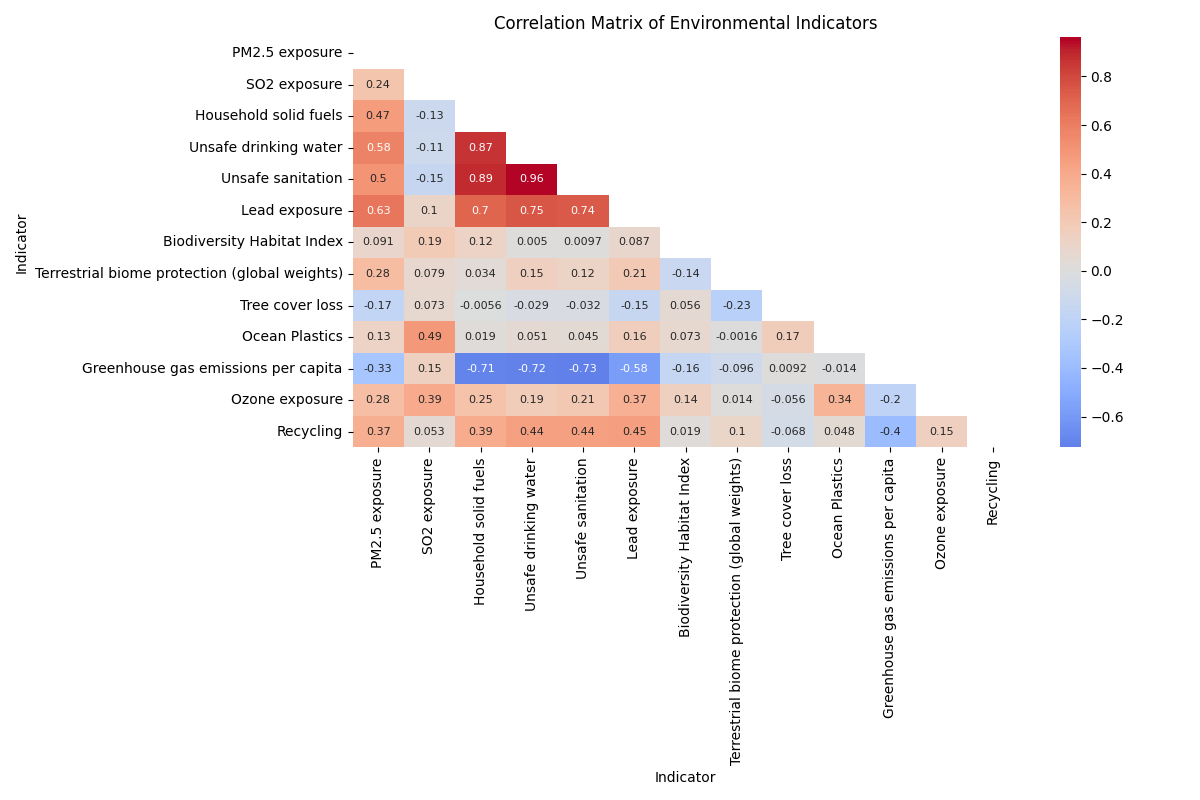
The initial\_statistics.csv report provides a comprehensive overview of the 13 environmental indicators selected from the YALE-EPI dataset, summarizing their distribution across 210 countries. Key indicators, such as PM2.5 exposure (mean: 37.15, std: 23.54), Household solid fuels (mean: 47.25, std: 32.51), and Greenhouse gas emissions per capita (mean: 51.93, std: 25.72), show significant variability, indicating diverse environmental challenges across countries. Notably, minimum values are non-zero but very small (e.g., 0.1 for PM2.5 exposure, 0.117 for Tree cover loss), suggesting that preprocessing in the load\_data method effectively replaced zeros with medians, ensuring data completeness. The high standard deviations and wide ranges (e.g., 0.1 to 100 for PM2.5 exposure) highlight the need for standardization, as performed in the PCA and clustering steps, to prevent dominant variables from skewing results.

The data\_quality\_report.csv offers a detailed assessment of data quality for a broader set of indicators, including the 13 selected ones. For these indicators, missing values are moderate (e.g., 6.67% for PM2.5 exposure, 24.29% for Ocean Plastics), and zero counts are low (e.g., 6 zeros for PM2.5 exposure, 2.857% of rows).

## Correlation Analysis

I created a correlation matrix to explore relationships between indicators. This helped identify multicollinearity, which could affect PCA.





The heatmap shows strong positive correlations, like between Unsafe drinking water and Unsafe sanitation (0.96), suggesting countries with poor drinking water quality often have sanitation issues too. Negative correlations, such as Greenhouse gas emissions per capita with Household solid fuels (-0.71), indicate that higher emissions might be linked to lower use of solid fuels, possibly due to industrial activity. The correlations like Tree cover loss with Biodiversity Habitat Index (0.06), show some indicators are less connected.

### Correlation Significance

**Strong Positive Correlations**:

Unsafe drinking water vs. Unsafe sanitation (r = 0.961, p = 3.89e-118)

Extremely high correlation, indicating redundancy.

* Household solid fuels vs. Unsafe drinking water (r = 0.868, p = 3.63e-65) and Unsafe sanitation (r = 0.885, p = 5.03e-71): Strong links to health-related indicators.
* PM2.5 exposure vs. Lead exposure (r = 0.633, p = 6.36e-25)

Unsafe drinking water (r = 0.583, p = 1.61e-20), and Unsafe sanitation (r = 0.503, p = 7.53e-15): Indicates air and health pollution connections

**Moderate Positive Correlations**:

* SO2 exposure vs. Ocean Plastics (r = 0.486, p = 7.66e-14) and Ozone exposure (r = 0.390, p = 5.01e-09)

Suggests pollution interlinkages.

* Recycling vs. PM2.5 exposure (r = 0.370, p = 3.18e-08) and Household solid fuels (r = 0.386, p = 6.96e-09): Weak but significant waste management ties.

**Strong Negative Correlations**:

* Greenhouse gas emissions per capita vs. Household solid fuels (r = -0.706, p = 5.37e-33), Unsafe drinking water (r = -0.717, p = 2.19e-34), and Unsafe sanitation (r = -0.727, p = 7.95e-36): Suggests higher emissions correlate with better health infrastructure.
* Recycling vs. Greenhouse gas emissions per capita (r = -0.400, p = 1.74e-09): Indicates recycling reduces emissions.

Based on significant relationships (p < 0.05), I identified strong correlations, such as Unsafe drinking water and Unsafe sanitation (r = 0.961), indicating redundancy, and negative ties like Greenhouse gas emissions per capita with Household solid fuels (r = -0.706), suggesting trade-offs between emissions and health infrastructure. I selected 11 indicators

**PM2.5 exposure, SO2 exposure, Household solid fuels, Lead exposure, Biodiversity Habitat Index, Terrestrial biome protection, Tree cover loss, Ocean Plastics, Greenhouse gas emissions, Ozone exposure, and Recycling, Unsafe drinking water, Unsafe sanitation**

To represent diverse environmental aspects (health, ecosystems, waste, climate) while avoiding multicollinearity. Initially, I considered excluding "Unsafe drinking water" and "Unsafe sanitation" due to their high correlation to reduce redundancy and ensure PCA captures unique variance. However, I retained these variables in the final analysis to fully represent critical health-related environmental factors, allowing that PCA can handle correlated variables by combining their variance into principal components.

Below I show examples of plots that best illustrate key correlation findings and support PCA variable selection

### PM2.5 exposure vs Unsafe drinking water (r = 0.583, p = 1.61e-20)

A graph of water and a red line

AI-generated content may be incorrect.

The scatter plot above shows a strong positive correlation between air quality and health, supporting the inclusion of PM2.5 exposure in PCA. The high r-value and p-value indicate a reliable relationship

**Household solid fuels vs Unsafe sanitation (r = 0.885, p = 5.03e-71)**

A graph of a graph with blue dots

AI-generated content may be incorrect.

Shows a very strong correlation within health indicators, this plot also supports why I retained Household solid fuels but excluded Unsafe sanitation due to redundancy with Unsafe drinking water (r = 0.961)

### Greenhouse gas emissions per capita vs Household solid fuels

A graph of gas emissions

AI-generated content may be incorrect.

Greenhouse gas emissions per capita vs Household solid fuels Illustrated a strong negative correlation between climate and health indicators, suggesting countries with lower emissions may rely more on solid fuels.

### Ocean Plastics vs SO2 exposure (r = 0.486, p = 7.66e-14):

A graph of blue dots and red line

AI-generated content may be incorrect.

Ocean Plastics vs SO2 exposure represents a moderate positive correlation between waste and air pollution,

## Principal Component Analysis (PCA)

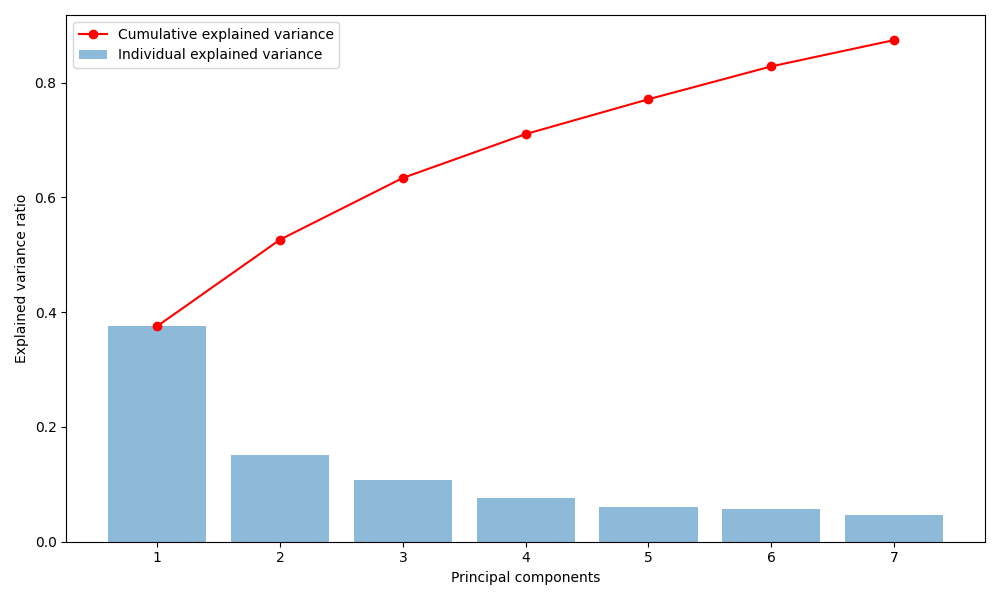
I applied PCA to reduce the dimensionality of the dataset while retaining most of its variance. I standardized the data using StandardScaler to ensure equal contribution from all indicators. PCA helped identify the most influential indicators and facilitated clustering by reducing noise. I plotted the explained variance

ratio and the first two principal components to visualize country distributions.

### Explained Variance (Individual and Cumulative)

|  |  |  |
| --- | --- | --- |
| PC | Individual | Cumulative |
| 1 | 0.375654762 | 0.375654762 |
| 2 | 0.150770897 | 0.526425659 |
| 3 | 0.107584083 | 0.634009741 |
| 4 | 0.076441413 | 0.710451154 |
| 5 | 0.060714003 | 0.771165157 |
| 6 | 0.057265637 | 0.828430794 |
| 7 | 0.045850869 | 0.874281662 |
| 8 | 0.042772742 | 0.917054405 |
| 9 | 0.028269299 | 0.945323704 |
| 10 | 0.022681225 | 0.968004929 |
| 11 | 0.019761339 | 0.987766268 |
| 12 | 0.009791819 | 0.997558086 |
| 13 | 0.002441914 | 1 |

The table shows that the first principal component (PC1) explains 37.57% of the variance, and PC2 adds 15.08%, totaling 52.64% for the first two components. By PC11 98.78% cumulative variance is reached meaning nearly all variability is captured. This is visualized better in the chart below.



## Feature Importance

Main drivers of variance included unsafe drinking water (0.423) and unsafe sanitation (0.419),

while SO2 exposure (0.006) has little influence

|  |  |
| --- | --- |
| Feature | Importance |
| Unsafe drinking water | 0.422871 |
| Unsafe sanitation | 0.418707 |
| Household solid fuels | 0.401067 |
| Lead exposure | 0.392178 |
| Greenhouse gas emissions per capita | 0.353675 |
| PM2.5 exposure | 0.309275 |
| Recycling | 0.258087 |
| Ozone exposure | 0.165061 |
| Terrestrial biome protection (global weights) | 0.09326 |
| Ocean Plastics | 0.057795 |
| Biodiversity Habitat Index | 0.049089 |
| Tree cover loss | 0.046169 |
| SO2 exposure | 0.005963 |

## Country Scores (PC1 and PC2)

Stored in /results/pca\_components

The PCA results reveal distinct environmental performance patterns across countries based on their PC1 and PC2 scores.

Countries like Iceland (PC1: 5.337, PC2: 1.664) and Finland (PC1: 5.013, PC2: 1.052) exhibit high positive PC1 scores, suggesting strong performance in key environmental factors such as access to safe drinking water and health, which dominate PC1. Conversely, nations like Yemen (PC1: -3.655, PC2: -0.982) and Lesotho (PC1: -3.635, PC2: -1.491) have strongly negative PC1 scores, indicating problems in these areas.

PC2 highlights additional variation, with countries like the Marshall Islands (PC2: 3.506) and Palau (PC2: 3.235) scoring high, potentially reflecting better outcomes in secondary factors like biodiversity or recycling, while countries such as the United Arab Emirates (PC2: -2.818) and Iran (PC2: -2.558) score low, suggesting weaker performance in these aspects. This 2D representation captures 52.64% of the total variance, providing a simplified but insightful view of environmental differences.

## 

I’ve explored the first two principal components visually but found the scatter plots less informative due to data complexity

## Clustering Analysis

I performed three clustering methods to group countries by environmental performance:

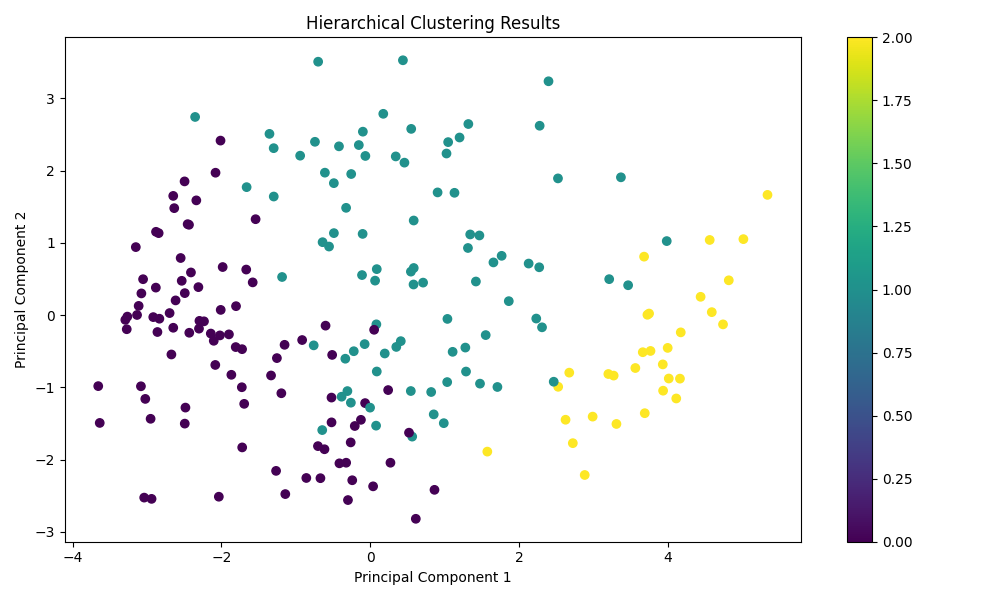
• Hierarchical Clustering: I used single, complete, and average linkage methods to create dendrograms, revealing hierarchical relationships.

• K-means Clustering: I used the elbow method to select an optimal number of clusters (set to 3) and visualized cluster assignments.

• DBSCAN: I applied density-based clustering to identify outliers and dense

clusters, using default parameters (eps=0.5, min\_samples=5).

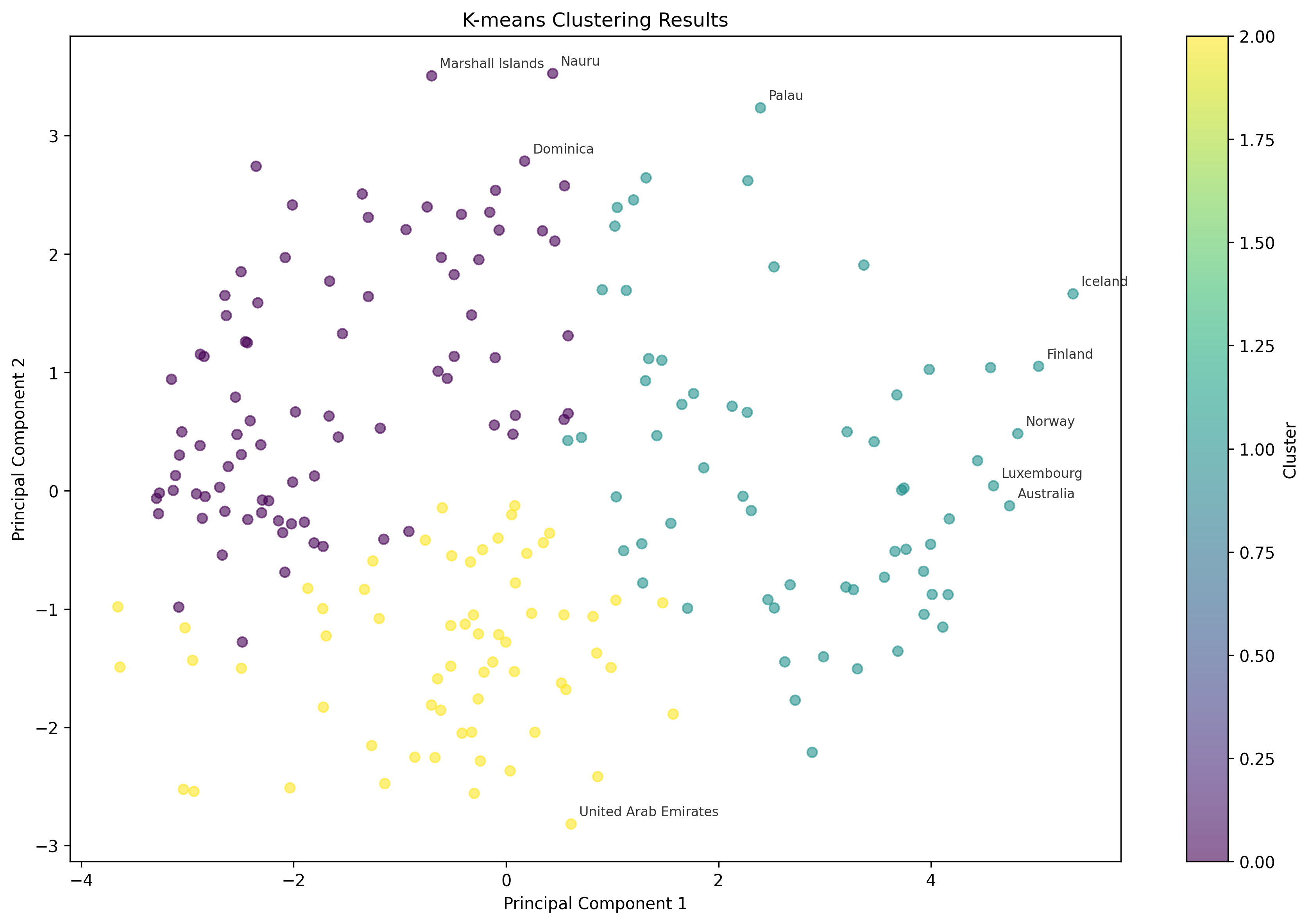
DBSCAN didn’t perform well with the parameters I used and tuning it didn’t seem worth the effort since K-means and hierarchical clustering already met the goals. If outlier detection had been a priority, I could have revisited DBSCAN and focused on tuning, but for now, it made more sense to drop it from the pipeline.



The clustering analysis used K-means and hierarchical clustering on the first 7 principal components (PCs), which together explain 87.43% of the variation in the environmental sustainability data. The first two PCs alone account for 52.64%, and the first three cover 63.40%.

The results are shown in the scatter plots where PC1 (37.57%) and PC2 (15.08%) are plotted and colored by cluster labels. Both methods found 3 main clusters, based on the elbow method and dendrogram.

From the PCA loadings (pca\_loadings.csv), PC1 is mainly about environmental health. It’s most influenced by unsafe drinking water, sanitation, household fuels, lead, and PM2.5, with a strong negative weight from greenhouse gas emissions. So, countries with high PC1 values generally struggle with pollution and health, while those with low values tend to do better, though sometimes with higher emissions due to industrialization. PC2 picks up issues related to industrial pollution like SO2, ocean plastics, and ozone—with some influence from biodiversity.



K-means and hierarchical clustering gave similar results, which suggests the clusters are stable. Cluster 0 includes countries with good environmental health (low PC1), like Samoa (score 0.918) and South Korea (0.742), though they have higher emissions. Cluster 1 has high PC1 values, indicating poorer health outcomes but usually lower emissions; examples are Benin (0.745), Sweden (0.675), and Singapore (0.534). Cluster 2 is in the middle, with balanced performance across indicators—countries like Iceland (0.878), Greenland (0.813), and Germany (0.632).

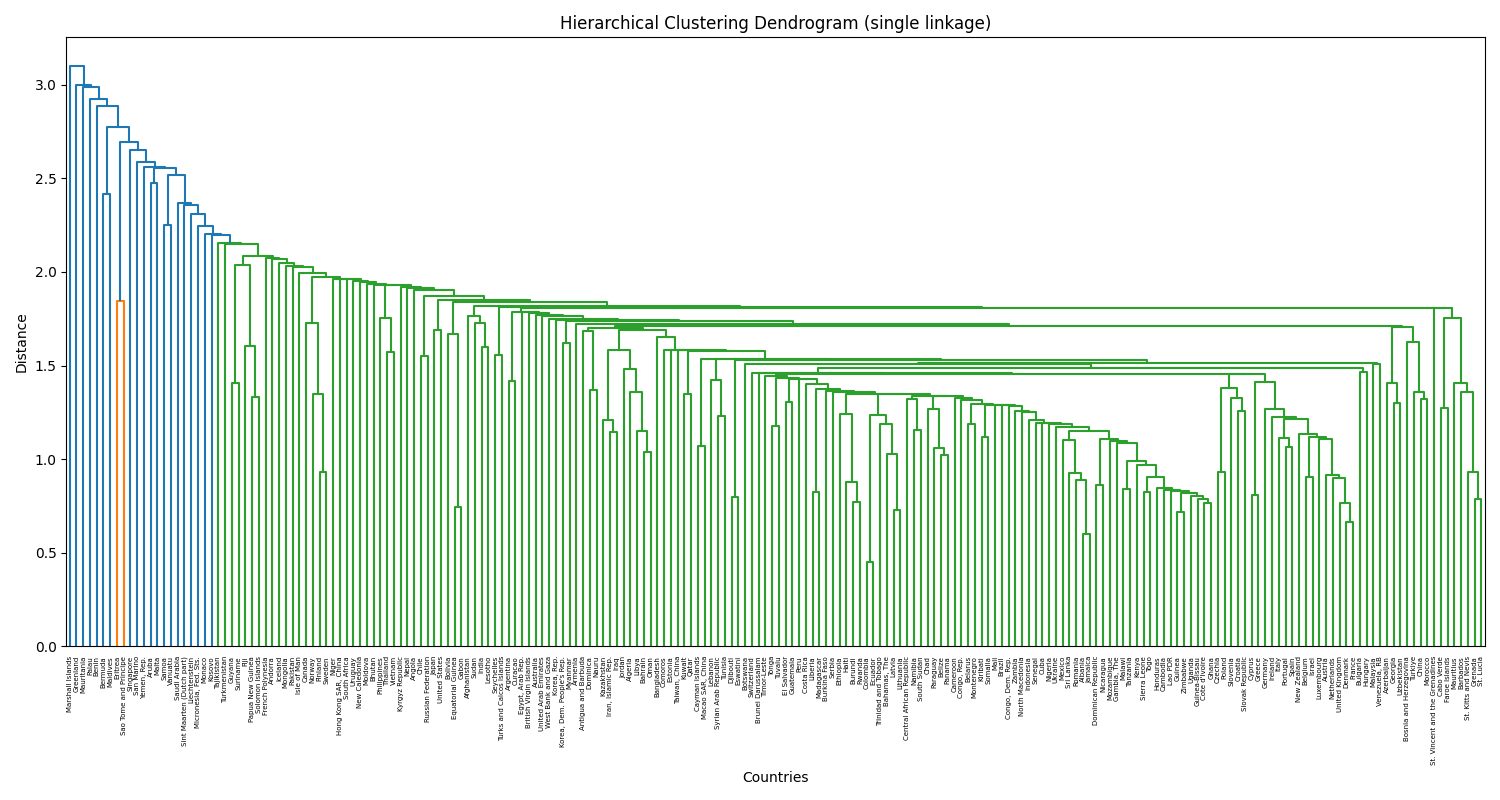
The scatter plots show PC1 does a good job separating countries by environmental health, while PC2 adds details about industrial impacts. Interestingly, countries with high composite scores show up in different clusters—Samoa in Cluster 0 and Iceland in Cluster 2—showing that strong sustainability performance can come from different profiles.

Overall, the analysis points to three clear groups: Cluster 0 with good health outcomes but higher emissions, Cluster 1 with health challenges and lower emissions, and Cluster 2 with more balanced profiles. These insights, backed by the visualizations and PCA loadings, can help guide plans like improving health systems in Cluster 1 or cutting emissions in Cluster 0.

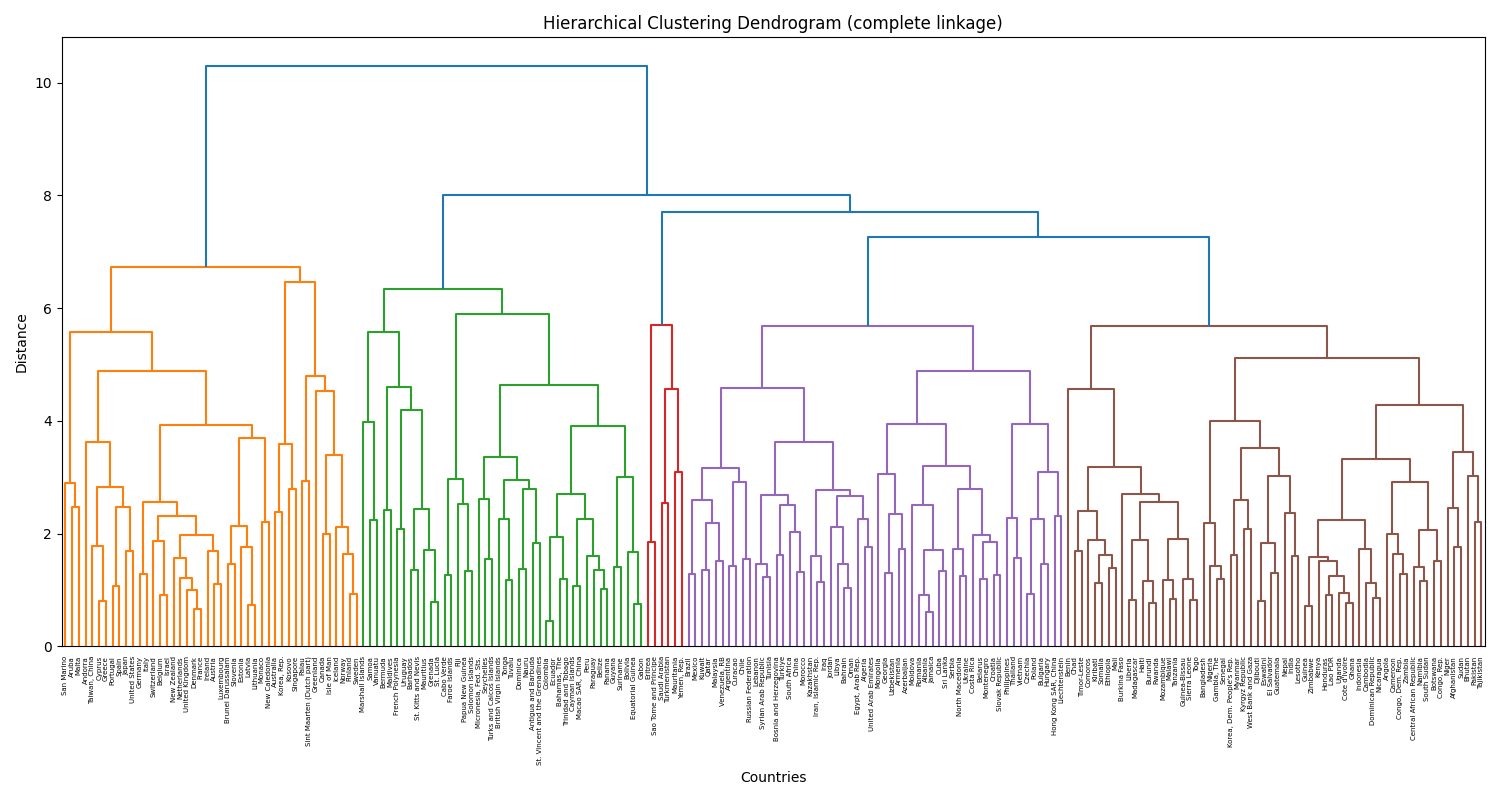
A diagram of a city

AI-generated content may be incorrect.

Average linkage Dendrogram



Single Linkage Dendrogram



Complete Linkage Dendrogram

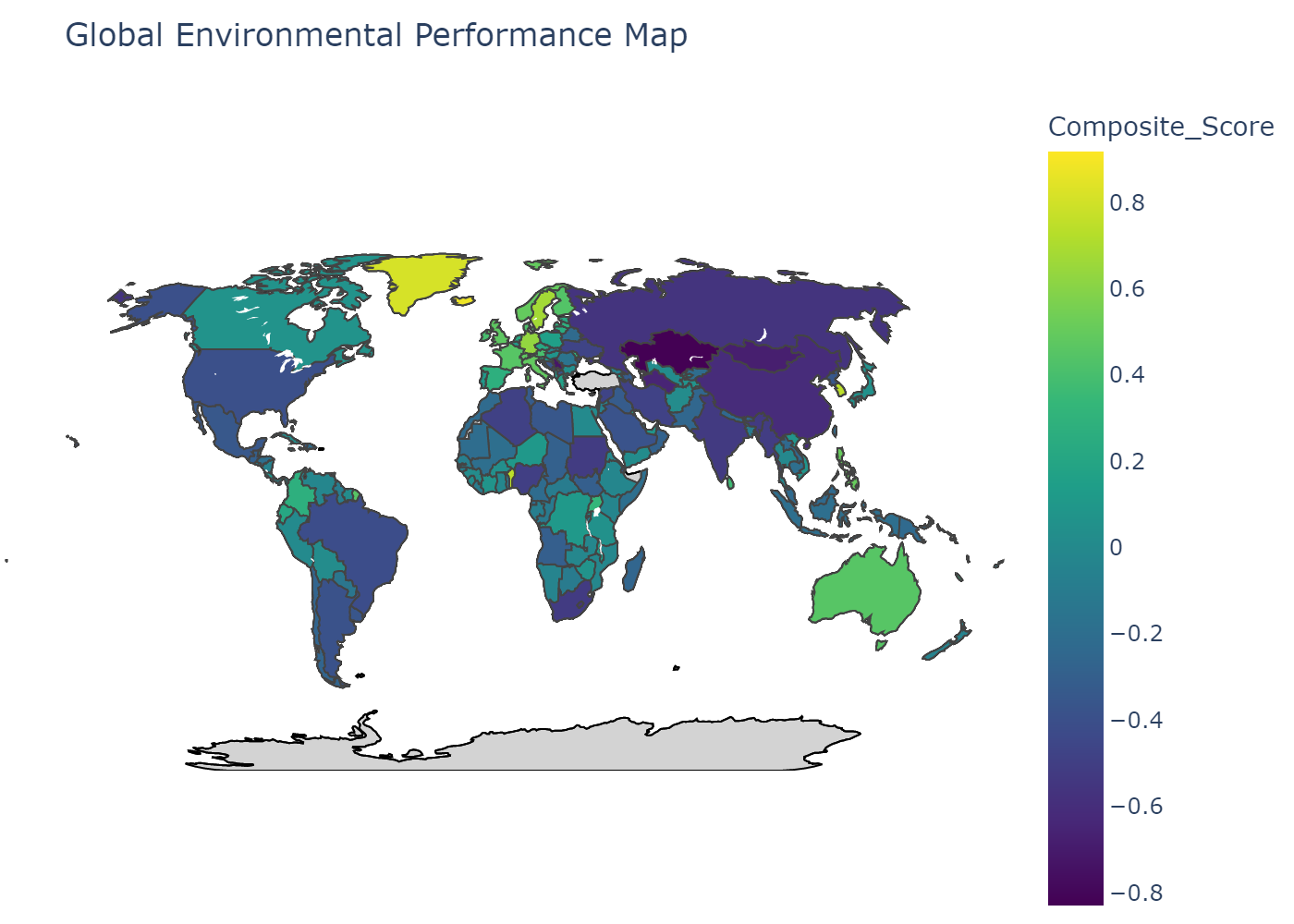
# 

DBSCAN was also tested, with results saved in dbscan\_clusters.csv and shown above. But it didn’t work well here. Out of 210 countries, 165 were marked as noise (Cluster -1), meaning DBSCAN couldn’t group them. Only a few small clusters were formed: Cluster 0 (Albania, Jamaica, etc.), Cluster 1 (Burundi, Haiti, Rwanda), Cluster 2 (mostly African countries), Cluster 3 (some Western European countries), and Cluster 4 (a few Caribbean nations).

The poor results are mainly due to how DBSCAN works—it depends a lot on its parameters (eps=1.0, min\_samples=3) and doesn't perform well in high-dimensional, standardized spaces like this one after PCA. The 7 principal components make space dense, and DBSCAN couldn’t find enough dense areas to form solid clusters, so most countries were marked as noise.

In the scatter plot, most points appear as the same color (noise), with just a few small groups, making it hard to get useful insights. Because of this, DBSCAN wasn’t helpful and was dropped from the analysis.

# Findings



Global map displaying environmental performance scores by country. Colors indicate performance levels, with brighter shades showing higher scores of the index.

I constructed a composite environmental sustainability indicator by combining sub-indices (Health, Ecosystems, Waste, Climate) weighted by PCA feature importance. I normalized the weights and calculated sub-index scores, then averaged them to obtain a final score per country. I visualized the score distribution and created a world map to show global sustainability patterns.

The resulting composite scores reveal significant variation in environmental sustainability across countries, with top performers like Samoa and Iceland demonstrating balanced strengths across multiple sub-indices, while lower-ranked nations such as Kazakhstan and Serbia face challenges in waste management and climate performance. The analysis leverages data from multiple sources, ensuring a comprehensive assessment of environmental factors. Below, I elaborate on each sub-index and the insights derived from the results.

Samoa ranked highest (0.918), mainly due to excellent waste management (2.815) and solid climate performance (0.574), even though its health score was reasonable. Iceland followed (0.878), with strong health (1.829) and waste (2.502) scores, but had a low climate score (-1.458) due to emissions. Greenland (0.813) stood out for ecosystem protection (1.504). Benin (0.745) and South Korea (0.742) showed strong performance in waste and climate. Liechtenstein (0.680) and Germany (0.632) had balanced scores in health and waste. Sweden (0.675) performed well in health (1.832), while Sao Tome and Principe (0.596) excelled in ecosystems and climate. Singapore (0.534) had solid scores in health and waste. Each country showed strengths in specific areas, though some still faced environmental challenges.

**Top 10 environmentally best countries overall**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country | Composite Score | Health | Ecosystems | Waste | Climate | Rank |
| Samoa | 0.918456 | -0.04846 | 0.333163 | 2.814893 | 0.57423 | 1 |
| Iceland | 0.877561 | 1.82905 | 0.637308 | 2.501817 | -1.45793 | 2 |
| Greenland | 0.813386 | 0.738719 | 1.504235 | 1.16951 | -0.15892 | 3 |
| Benin | 0.744819 | -1.01672 | 0.176964 | 2.389373 | 1.429663 | 4 |
| Korea, Rep. | 0.741812 | 1.246104 | 0.162321 | 2.936907 | -1.37808 | 5 |
| Liechtenstein | 0.679505 | 0.298487 | 0.395588 | 2.222498 | -0.19855 | 6 |
| Sweden | 0.674621 | 1.831763 | 0.022581 | 1.348203 | -0.50406 | 7 |
| Germany | 0.631879 | 1.481778 | 0.217606 | 1.867306 | -1.03918 | 8 |
| Sao Tome and Principe | 0.596414 | -0.63415 | 1.287264 | -0.12971 | 1.86225 | 9 |
| Singapore | 0.53373 | 1.339219 | -0.10753 | 2.046292 | -1.14307 | 10 |

**Top 10 environmentally worst countries overall**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Country | Composite Score | Health | Ecosystems | Waste | Climate | Rank |
| Sudan | -0.5185 | -1.23551 | -0.57645 | -0.8309 | 0.568852 | 201 |
| South Africa | -0.52089 | -0.61338 | -0.68236 | 0.305346 | -1.09319 | 202 |
| India | -0.52841 | -1.15152 | -1.38866 | -0.19491 | 0.621459 | 203 |
| Lesotho | -0.56002 | -1.37213 | -1.11278 | -0.60901 | 0.853828 | 204 |
| Russian Federation | -0.56707 | 0.337543 | 0.255426 | -1.18935 | -1.67191 | 205 |
| China | -0.61292 | -0.37317 | -1.06083 | -0.10968 | -0.90801 | 206 |
| Turkmenistan | -0.65024 | -0.16508 | -0.21245 | -0.23188 | -1.99156 | 207 |
| Mongolia | -0.68217 | -0.67603 | 0.213393 | -0.27869 | -1.98735 | 208 |
| Serbia | -0.72464 | 0.042192 | -0.62245 | -1.17045 | -1.14786 | 209 |
| Kazakhstan | -0.82967 | -0.22795 | 0.008404 | -1.30352 | -1.79563 | 210 |

**Health Sub-Index**  
This measured factors like PM2.5, SO2, solid fuel use, unsafe water and sanitation, lead, and ozone exposure. High scores meant better public health and lower pollution. Finland (1.929) and Sweden (1.832) scored high due to strict air quality rules and strong health systems. Lesotho (-1.372) and Sudan (-1.236) scored low, likely due to poor access to clean water and high pollution levels. The gap showed how vital environmental health was for sustainability.

**Ecosystems Sub-Index**  
This measured biodiversity, biome protection, and tree cover loss. High scores suggested good conservation. Greenland (1.504) and Sao Tome and Principe (1.287) performed well due to low deforestation and protected habitats. India (-1.389) and Lesotho (-1.113) scored poorly, struggling with habitat loss and weak policies. Results showed that protecting ecosystems needed stronger support, especially in lower-income countries

**Waste Sub-Index**

The Waste sub-index focuses on waste management practices, specifically ocean plastics and recycling rates. Countries with high scores demonstrate effective waste reduction and recycling systems. South Korea (Waste score: 2.937) and Samoa (2.815) top this category, reflecting advanced recycling infrastructure and policies to curb plastic pollution. On the other hand, nations like Kazakhstan (-1.304) and Serbia (-1.170) rank poorly, indicating inadequate waste management systems and high plastic pollution. Low-performing countries could benefit from adopting models like South Korea’s comprehensive recycling programs.

**Climate Sub-Index**  
The Climate sub-index looked at greenhouse gas emissions per person—higher scores meant lower emissions and better climate performance. Malawi (1.873) and Afghanistan (1.873) ranked high because of low emissions, mostly due to limited industrial activity. On the other hand, Turkmenistan (-1.992) and Mongolia (-1.987) ranked low because of high fossil fuel use. The data showed a tricky balance: poorer countries often had low emissions but still faced issues in areas like health or waste. Countries with high emissions need to focus more on renewables and cutting carbon to boost their scores and meet climate targets.