Analyzation of the progress towards SDG goal No.7 though various visualisation techniques

Dataset I used:

1. Share of electricity generation from fossil fuels, renewables and nuclear (Our World in Data)
https://ourworldindata.org/grapher/electricity-fossil-renewables-nuclear-line?tab=table&country=~OWID_ASI

2. Electricity production by source, Measured in terawatt-hours. (From Our World in Data)
https://ourworldindata.org/grapher/electricity-prod-source-stacked?tab=table&country=OWID_WRL~OWID_SAM~OWID_EUR~OWID_ASI~OWID_NAM

3. Share of electricity generation from fossil fuels, renewables and nuclear
https://ourworldindata.org/grapher/electricity-fossil-renewables-nuclear-line?tab=table&country=~OWID_ASI

4. Share of people who support policies to tackle climate change, 2023 (From Our World in Data)
https://ourworldindata.org/grapher/support-policies-climate

5. Primary energy consumption per GDP (From Our World in Data)
https://ourworldindata.org/grapher/energy-intensity?time=2021

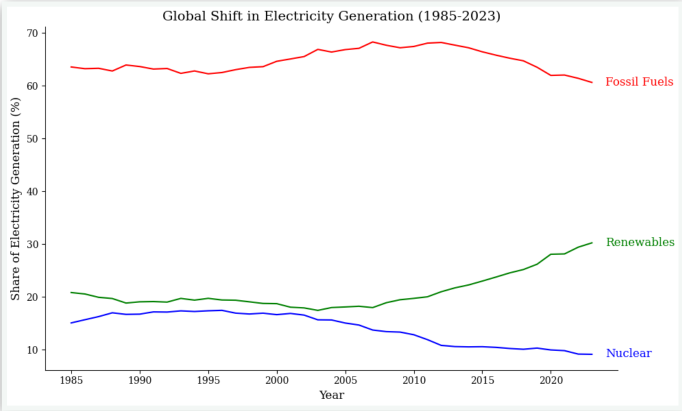
6. CO_2 emissions per capita vs. share of electricity generation from renewables, 2023 (From Our World in Data)
https://ourworldindata.org/grapher/co2-per-capita-vs-renewable-electricity

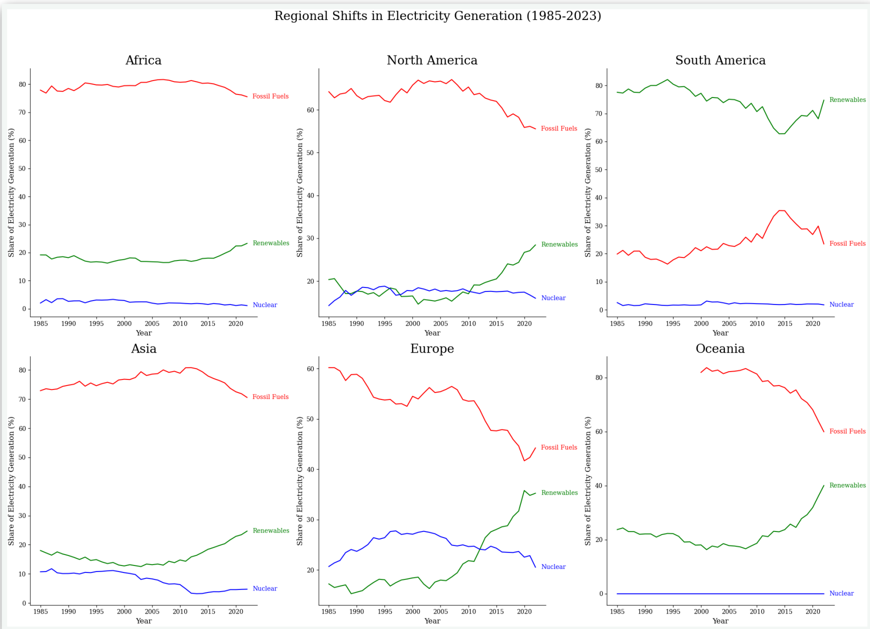
7. Share of electricity generated by renewables (From Our World in Data)
https://ourworldindata.org/grapher/share-electricity-renewables?tab=table

8. Global primary energy consumption by source (From Our World in Data)
https://ourworldindata.org/grapher/global-energy-substitution?time=1970..latest

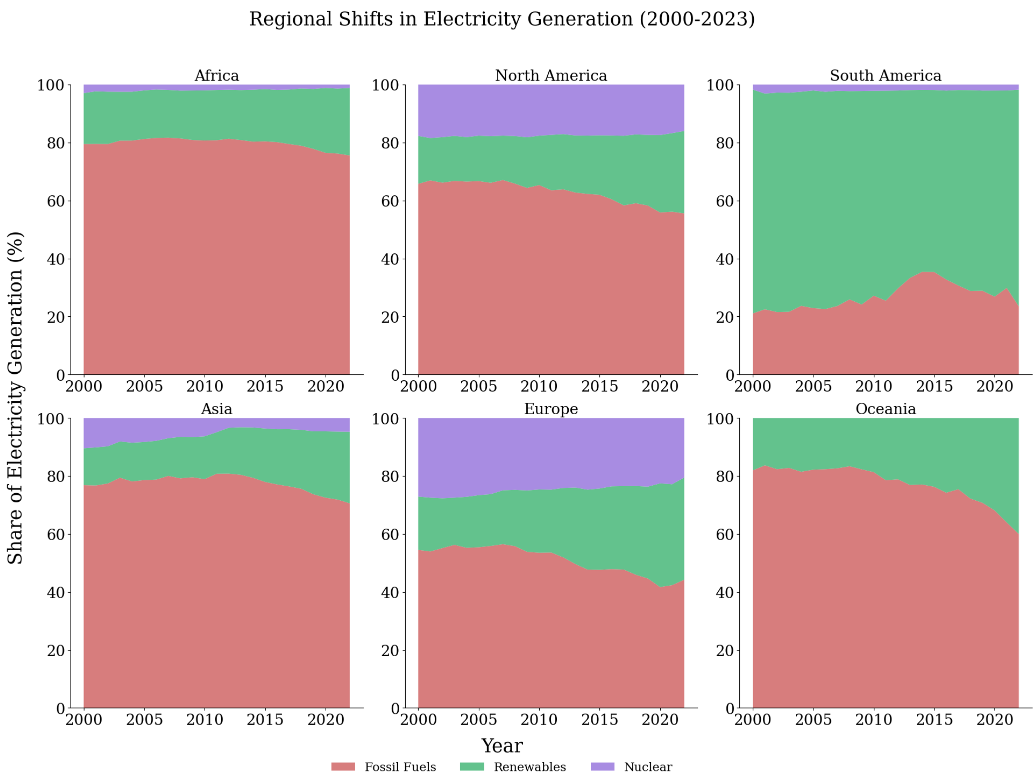
9. Global Data on Sustainable Energy, 2000-2020 (Kaggle)
https://www.kaggle.com/datasets/anshtanwar/global-data-on-sustainable-energy/data

Visualization:





Small multiples approach with six regional panels



stacked area chart

Interpretation:

Positive Indicators:

- Global renewable share has increased by ~10 percentage points
- All regions show positive renewable growth trajectories since 2010
- Fossil fuel dependency has begun meaningful decline globally
- Acceleration in renewable adoption is evident in multiple regions

Concerning Patterns:

- Africa shows minimal transition progress, indicating energy poverty risks
- Decline in nuclear capacity (low-carbon source) partially offsets renewable gains
- Regional disparities suggest uneven progress toward SDG 7 targets
- Even with progress, fossil fuels still constitute majority of global electricity generation

Regional SDG 7 Achievements:

- Europe and South America demonstrate most significant progress
- North America and Oceania show accelerating transition
- Asia displays promising recent momentum
- Africa requires concentrated investment to achieve targets

Interesting phenomenon:

South America stands out dramatically from all other regions as an exceptional renewable energy leader, maintaining the highest renewable electricity generation percentage consistently across the entire 1985-2023 period.

Despite experiencing a notable dip in renewable share around 2010-2015 (dropping to approximately 60-65%), South America quickly rebounded to approximately 75% renewable electricity generation by 2023 - more than double the renewable percentage of any other region except Europe. This exceptional pattern warrants further investigation into the unique geographical, infrastructural, and policy factors that have enabled South America to maintain such a dramatically different electricity generation profile compared to global norms.

Design Choices:

Line Chart

- Ideal for visualizing trends over time. Allowing viewers to easily identify changes in the share of electricity generation for fossil fuels, renewables, and nuclear energy across decades. Each region is represented in a separate subplot to avoid clutter and ensure clarity when comparing trends.

- Top and right borders are removed from each subplot to reduce unnecessary visual elements while maintaining essential axes for context. Gridlines are also omitted to avoid distractions, as the line plots themselves provide sufficient information.

- Labels are placed next to the lines at their endpoints instead of using a separate legend. Reducing visual clutter and makes it easier for viewers to associate labels with their corresponding lines.

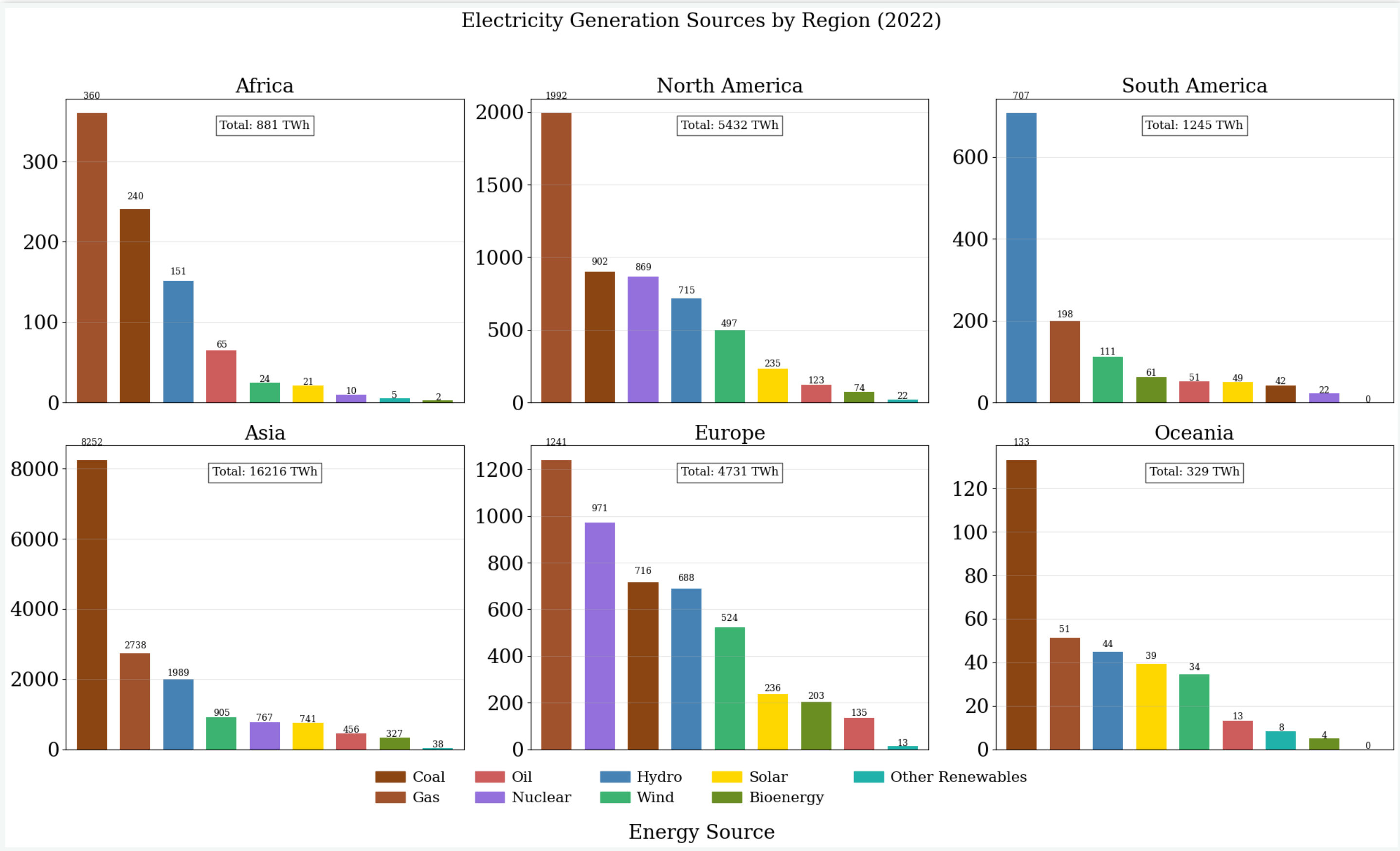
- Use the plt.tight_layout() function to minimize whitespace between subplots while ensuring that titles, labels, and lines do not overlap. This optimizes space usage without sacrificing readability.

- Avoids clutter while enabling cross-regional comparison by using small multiples layout

Stacked Area Chart

- Provides visual emphasis proportional to each energy source's significance

Visualisation:



Interpretation:

Positive Indicators:

South America demonstrates exemplary renewable leadership with ~75% clean energy
All regions show diversification with multiple renewable sources contributing to their energy mix
Wind energy has achieved significant scale in North America and Europe
Solar deployment is becoming substantial across multiple regions, reaching 235+ TWh in larger economies

Concerning Patterns:

Asia's electricity generation remains overwhelmingly coal-dependent (8252 TWh, ~51% of total)
Africa's total generation (881 TWh) is disproportionately low compared to population, indicating energy access challenges
Coal still dominates in 5 of 6 regions, representing a significant decarbonization challenge
Oil remains a notable contributor in Africa, highlighting infrastructure limitations

Regional SDG 7 Achievements:

South America: Exceptional clean energy profile with 75% renewable electricity, primarily hydro
Europe: Most balanced generation mix with substantial contributions from all renewable categories
North America: Leading in absolute wind deployment (497 TWh) with strong nuclear contribution (869 TWh)
Oceania: Despite small scale (329 TWh total), maintains balanced renewable integration

Interesting Phenomenon:

South America's hydropower dominance (707 TWh, ~57% of regional total) reflects a decades-long infrastructure investment strategy that predates modern climate concerns

Nuclear energy shows dramatic regional variation: substantial in North America (869 TWh) and Europe (971 TWh), minimal in South America (22 TWh), and entirely absent in Oceania

Design Choices:

1. Grouped Bar Chart Approach

- Categorical comparison: Effectively shows absolute contribution of each energy source within regions Side-by-side arrangement: Enables cross-regional pattern recognition while maintaining distinct regional contexts

2. Multi-panel Structure

- Small multiples layout: Maintains regional identity while enabling cross-comparison
- Consistent color scheme: Energy sources retain identical colors across all panels

3. Data Presentation Elements

- Direct data labeling: Numerical values displayed atop each bar for precise reading
- Total generation indicators: Box callouts showing regional totals provide context
- Light gridlines: Subtle horizontal references aid in value estimation without visual distraction

4. Visual Hierarchy

- Sorted ordering: Energy sources consistently arranged across all panels
- Value-based visibility: Bar heights proportional to generation volumes for intuitive magnitude comparison
- Title prominence: Clear regional identification with standardized positioning

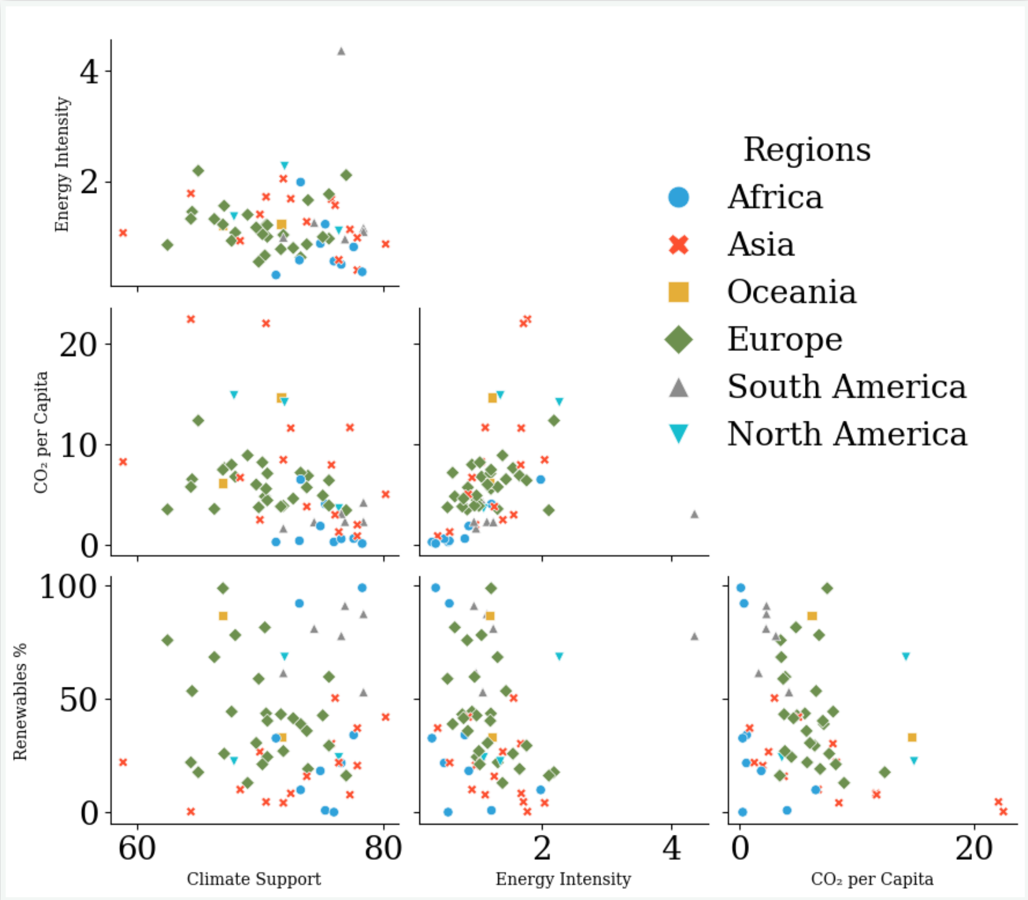
5. Color Strategy

- Semantic grouping: Browns for fossil fuels, blues for hydro, greens for bioenergy and wind
- Visual separation: Distinct hues for nuclear (purple) and solar (yellow) enhance categorical identification
- Consistent legend: Single comprehensive legend serves all panels

Chosen Variable:

1. Renewable Energy Share: The proportion of electricity generated from renewable sources.
2. CO_2 Emissions per Capita.
3. Share of People Supporting Policies to Tackle Climate Change.
4. Energy Intensity: Primary energy consumption per unit of GDP.

Visualization:



Interpretation:

Positive Indicators:

Clear inverse relationship between renewable energy adoption and CO_2 emissions per capita
Countries with strong climate policy support tend to have higher renewable energy integration
Most European countries demonstrate balanced profiles with moderate-to-high renewable adoption and climate support

Concerning Patterns:

Substantial global inequality in emissions (Africa averaging <5 CO_2/capita vs. North America >15)
High energy intensity strongly correlates with elevated emissions across all regions
Several Asian countries show concerning combination of high energy intensity and low renewable penetration
Climate support does not consistently translate to renewable adoption in all regions

Regional Characteristics:

Africa: Low emissions but extreme variation in renewable adoption (0-100%)
Europe: Most consistent renewable progress with strong climate policy support (60-80%)
North America: High emissions despite moderate renewable adoption
Asia: Generally higher emissions with lower renewable percentages, scattered distribution

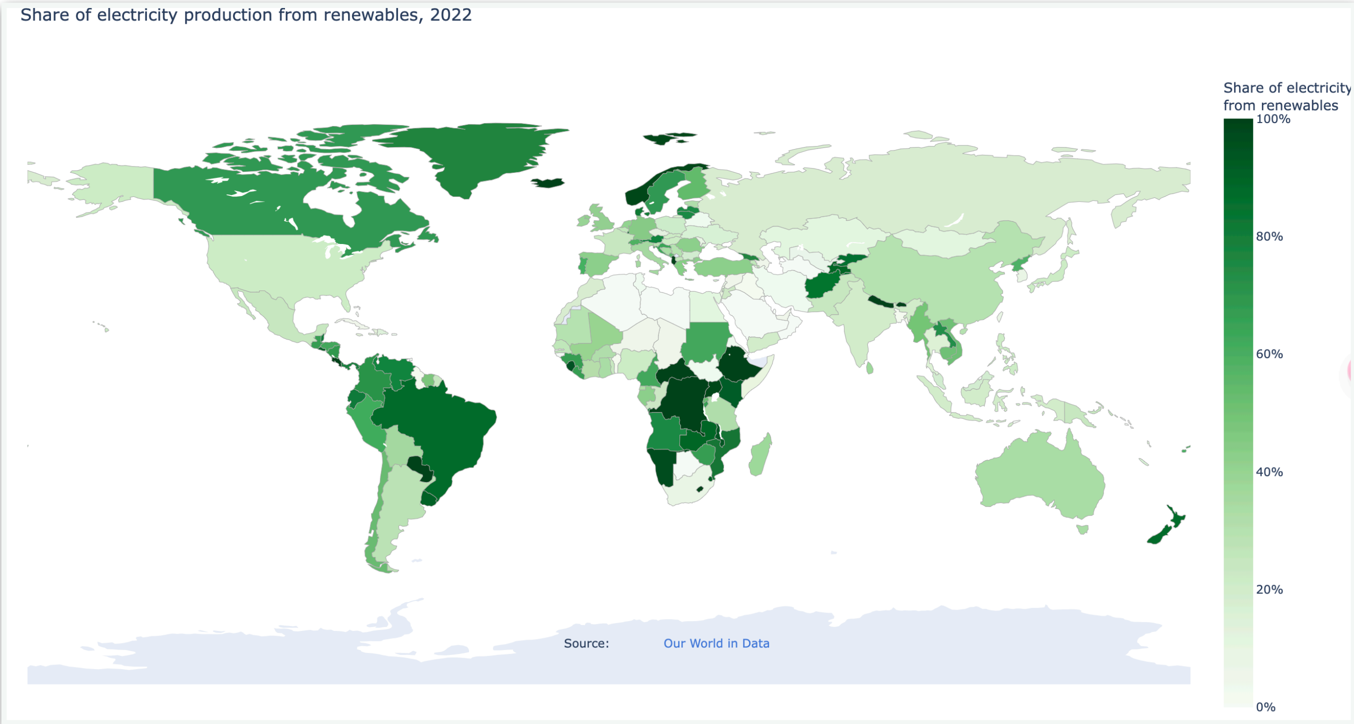
Interesting Phenomenon:

South America emerges as a distinctive renewable energy leader, with multiple countries achieving 70-90% renewable electricity despite moderate climate support scores. This unusual pattern contrasts with other regions where climate support more strongly correlates with renewable adoption. South America's unique geographical advantages (abundant hydropower potential) likely explain this outlier status, warranting further investigation into how this model might be partially adapted elsewhere.

Challenges and Solutions:

- Overplotting in regions with a high density of data points (Europe) -> reducing marker sizes and incorporating jitter to enhance visibility without compromising the underlying trends.

Visualisation on sequential colormap:



Interpretation:

Positive Indicators:

Several countries across South America and Africa have achieved 80-100% renewable electricity generation
Canada demonstrates strong renewable leadership in North America (~60-70%)
Nordic countries in Europe show robust renewable integration (60-80%)

Concerning Patterns:

Stark global inequality in renewable electricity integration is evident across continents
Most of North America, Australia, and large parts of Asia remain below 40% renewable electricity
Middle East and North African countries (excluding a few) show minimal renewable adoption

Regional SDG 7 Achievements:

South America: Continental leadership with multiple countries exceeding 70% renewable electricity
Europe: Moderate-to-high renewable integration with notable Nordic performance
Asia: Limited bright spots (e.g., Tajikistan) amid generally lower renewable penetration

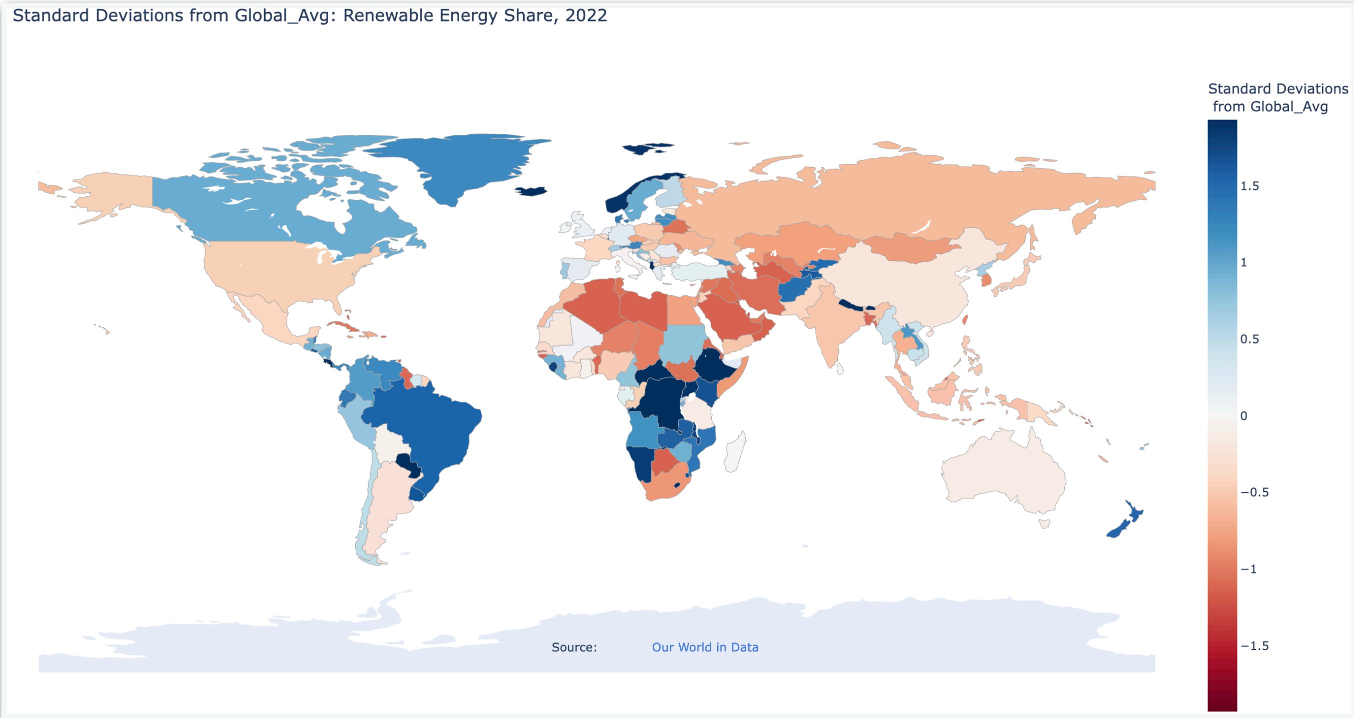
Design Choices:

1. Choropleth Map Implementation
Geospatial relationships: Reveals geographic patterns and regional clusters of renewable electricity adoption
Country-level granularity: Provides appropriate level of detail for national policy assessment
Equirectangular projection: Standard world map view that preserves familiar country shapes

2. Color Encoding Strategy
Sequential green palette: Intuitive association of deeper green with higher renewable percentages
Perceptually appropriate steps: Color gradation corresponds to meaningful percentage intervals

3. Legend Design
Continuous scale: Shows smooth transition from 0% to 100% renewable share
Right-side placement: Standard position for accessibility without obscuring key data
Percentage markers: Clear indicators at 20% intervals for precise reading

Visualisation on Diverging colormap:



Design Choices:

1. Diverging Color Scale Implementation

Statistical emphasis: Shows deviations rather than absolute values, highlighting relative performance
Centered palette: Red-white-blue spectrum with white at zero deviation creates intuitive reference point
Normalized comparison: Standard deviation metric enables fair assessment across economies of different sizes

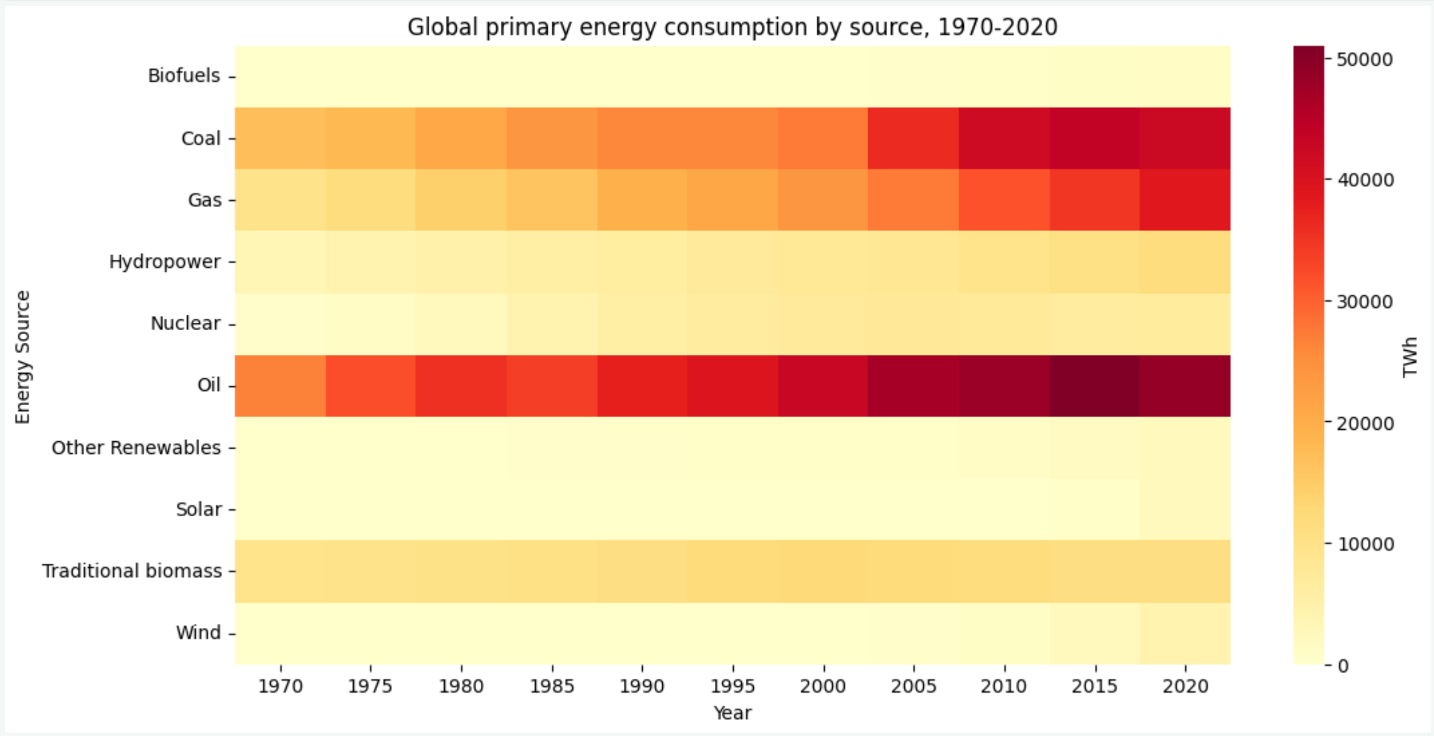
2. Statistical Approach

Standard deviation metric: Reveals country performance relative to global distribution
Z-score visualization: Effectively identifies statistical outliers in both directions
Normal distribution reference: Implicit assumption that renewable adoption follows statistical distribution

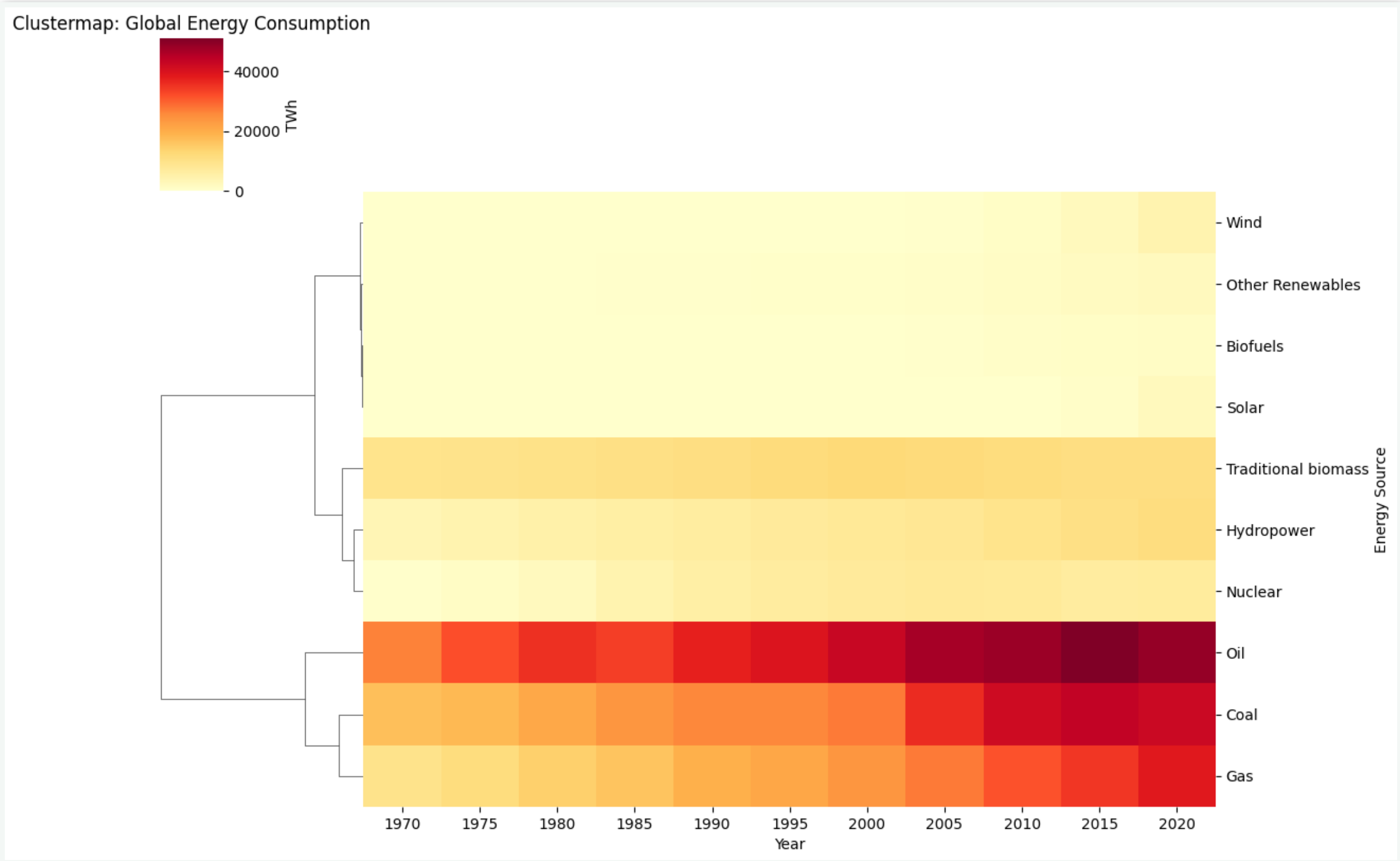
3. Color Strategy

Intuitive polarity: Blue for positive (above average), red for negative (below average)
Proportional intensity: Color saturation increases with deviation magnitude
Zero-centered design: White represents global average baseline

Visualisation on Heatmap:



Visualisation on Clustermap:



Interpretation:

Positive Indicators:

Gradual growth in low-carbon sources like nuclear (emerging in 1970s-1990s)
Recent decades show visible emergence of wind and solar energy

Concerning Patterns:

Oil consumption increased dramatically, reaching ~50,000 TWh by 2020 (darkest red)
Coal shows steady growth, particularly accelerating after 2000
Traditional biomass remains significant, indicating continued energy poverty
Solar and wind remain minimal contributors despite recent growth (pale yellow)

Energy Transition Assessment:

Fossil Fuels: Maintain overwhelming dominance throughout the entire period
Nuclear: Shows growth through 1990s, then plateaus in recent decades
Hydropower: Demonstrates modest but consistent growth
Modern Renewables: Appear only in recent years with minimal absolute contribution

Interesting Phenomenon:

Despite growing climate awareness and falling renewable costs, the heatmap reveals fossil fuels have not just maintained but increased their absolute consumption levels over 50 years. This persistence of high-carbon energy sources represents a significant challenge to SDG 7's clean energy targets. Most notably, coal consumption (indicated by the shift from orange to dark red) accelerated after 2000—precisely when climate mitigation became a global priority, suggesting economic growth has consistently outpaced decarbonization efforts.

Design Choices:

1. Heatmap Implementation

Temporal evolution: Effectively shows consumption changes across five decades
Color intensity encoding: Consumption magnitude represented through intuitive color gradient

2. Color Strategy

Sequential palette: Yellow-orange-red spectrum corresponds to increasing energy consumption

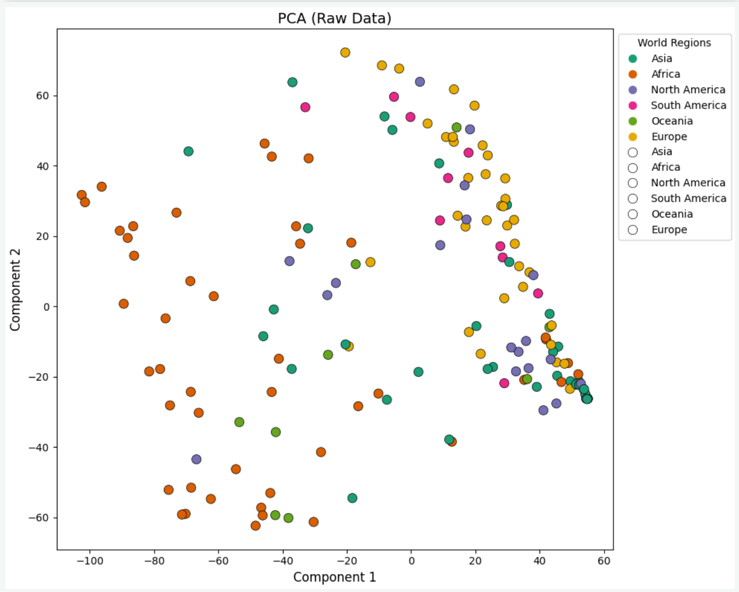
1. Clustermap Implementation

Dual information encoding: Combines consumption heatmap with hierarchical similarity analysis
Dendrogram integration: Left-side tree structure reveals energy source relationships

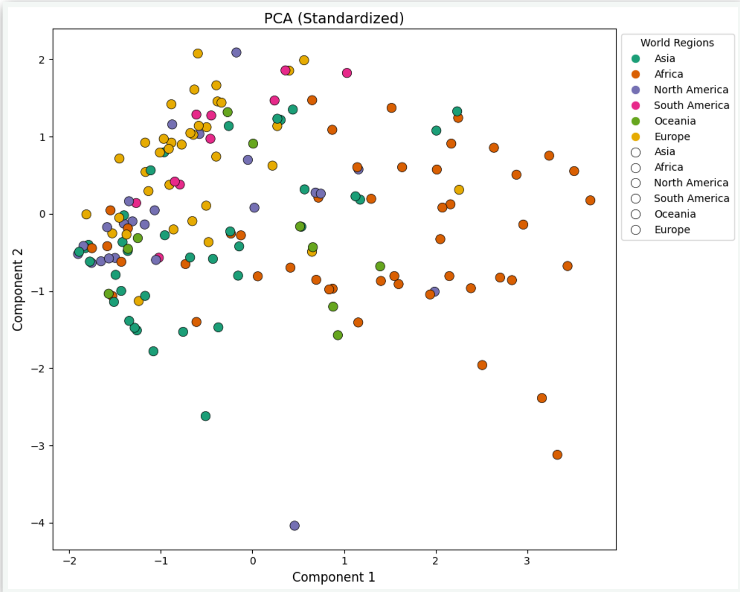
2. Hierarchical Clustering Approach

Energy source grouping: Algorithm identifies similarity patterns in consumption trends
Cluster distance visualization: Branch lengths indicate degree of similarity between sources
Natural categorization: Sources cluster based on data patterns rather than predefined categories

Visualization using PCA:

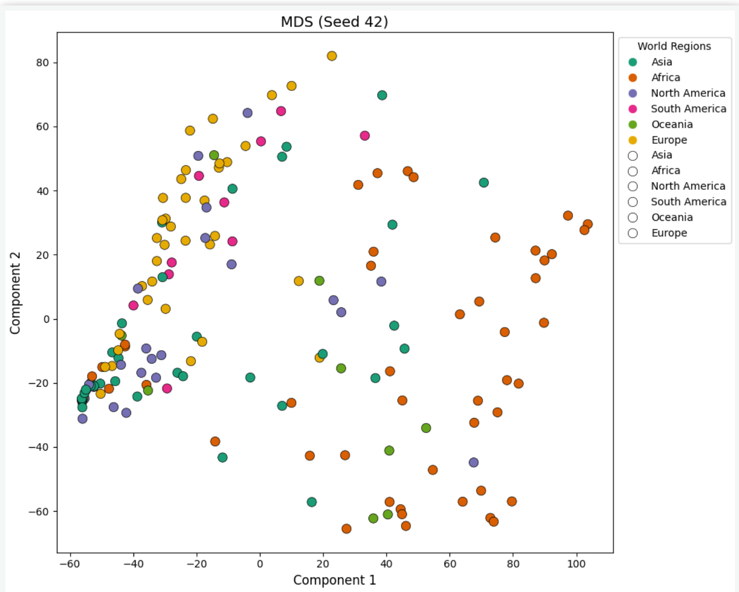


PCA without standardization

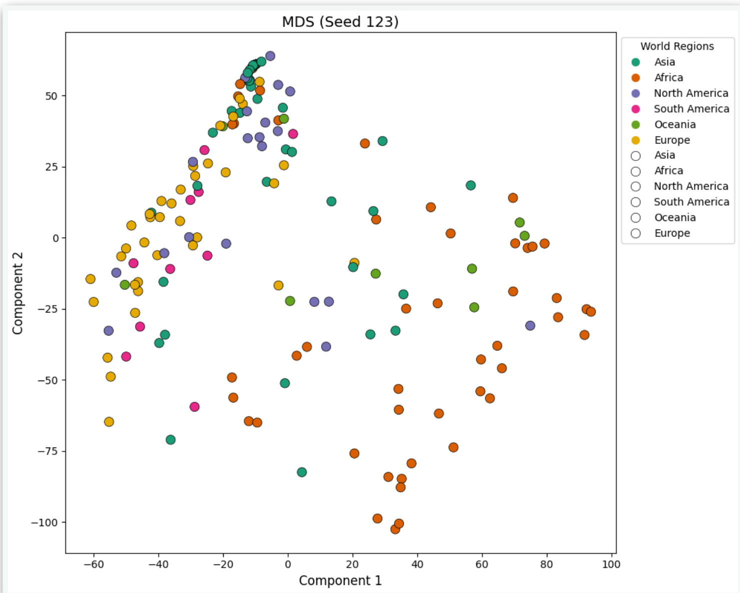


PCA with standardization

Visualization using MDS:



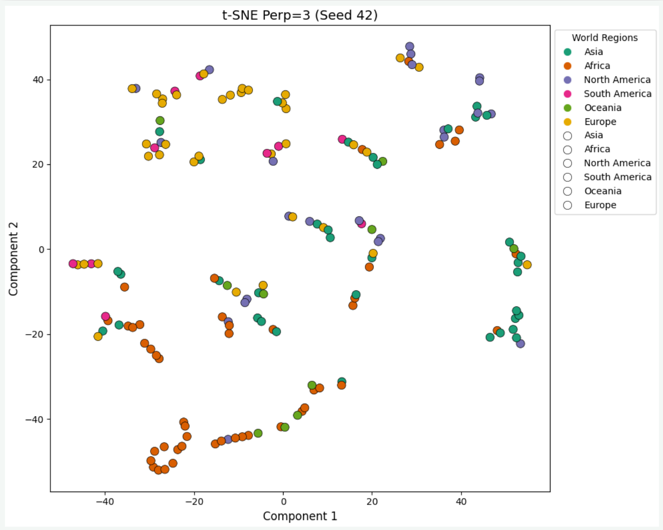
MDS with random seed 42



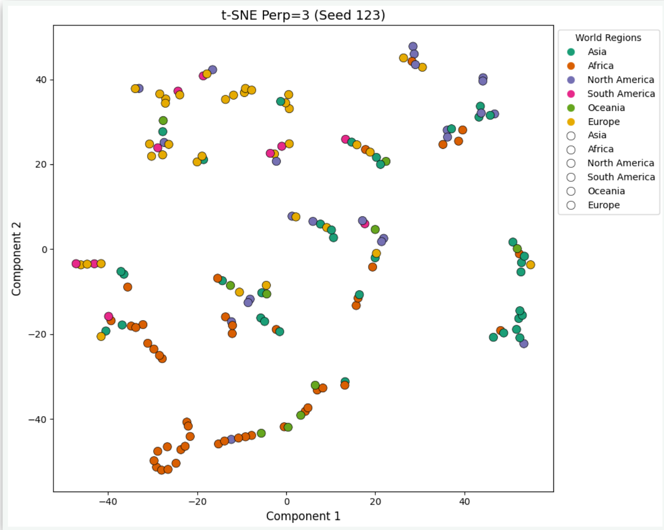
MDS with random seed 123

Visualization using t-SNE:

Perplexity value: 3

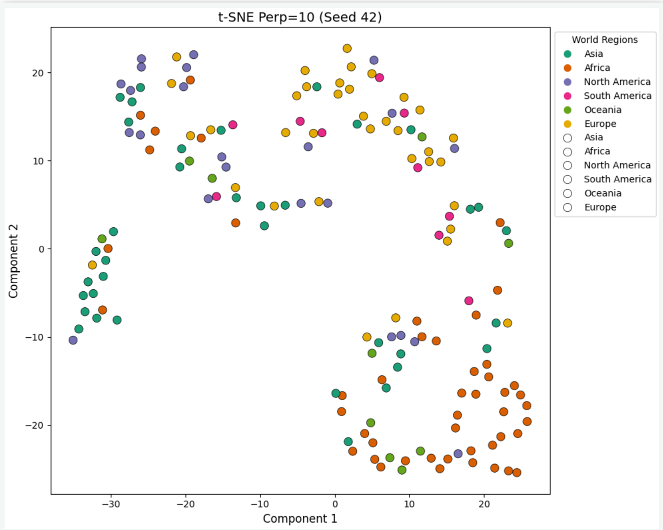


Radom seed: 42

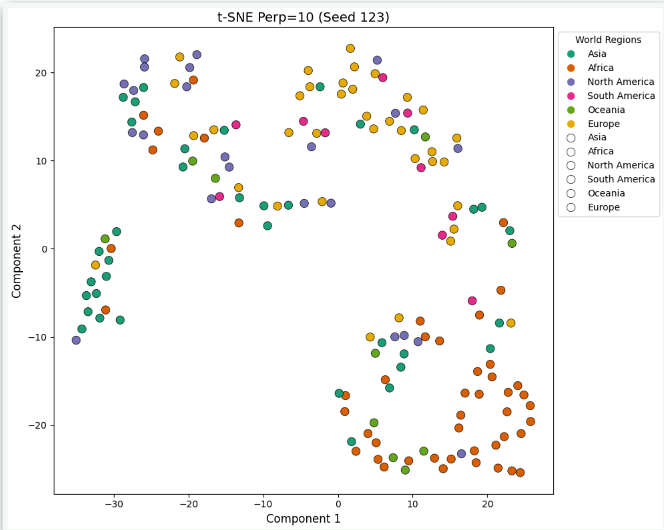


Radom seed: 123

Perplexity value: 10

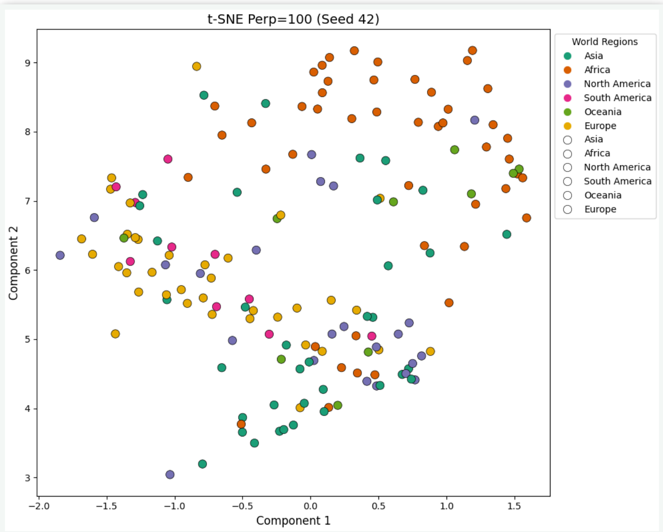


Radom seed: 42

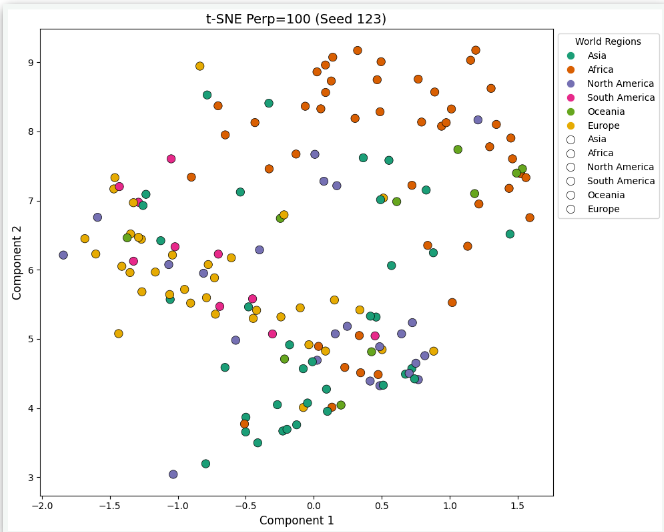


Radom seed: 123

Perplexity value: 100



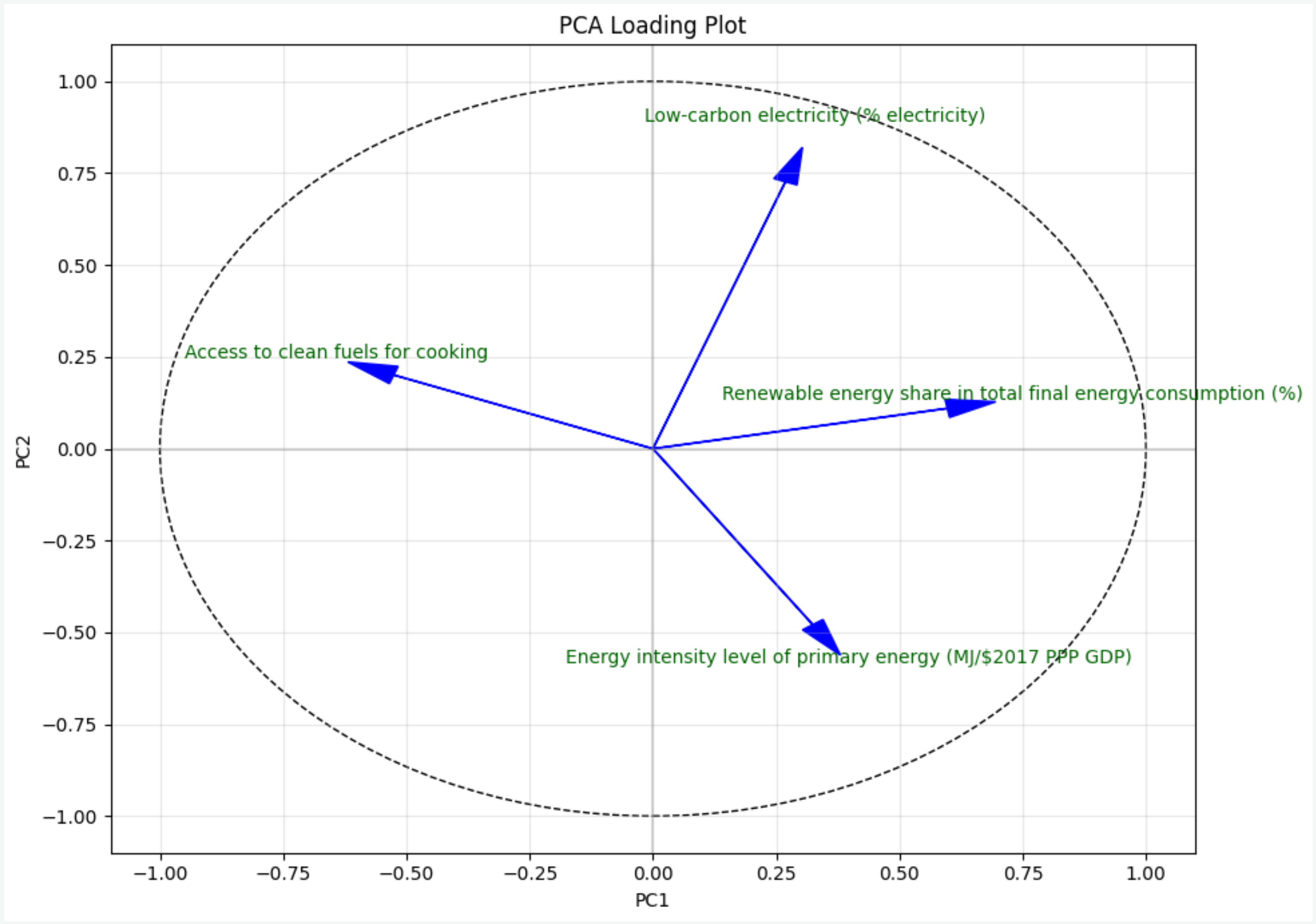
Radom seed: 42



Radom seed: 123

Four distinct variables:

- 'Access to clean fuels for cooking'
- 'Renewable energy share in the total final energy consumption (%)'
- 'Energy intensity level of primary energy'
- 'Low-carbon electricity (% electricity) : Percentage of electricity from low-carbon sources (nuclear and renewables)'




```
Results of PCA Loadings:
                                                     PC1       PC2
Access to clean fuels for cooking               -0.584099  0.222996
Renewable energy share in total final energy co...  0.663351  0.122006
Energy intensity level of primary energy (MJ/$2...  0.360159 -0.532253
Low-carbon electricity (% electricity)           0.298462  0.807523
Explained variance: PC1 0.51, PC2 0.27
```

Interpretation:

In the PCA visualizations, we can observe distinct regional patterns that reflect varying levels of energy development. The raw data PCA reveals a significant east-west distribution with African countries predominantly positioned on the left side (negative Component 1 values), while European countries cluster on the right. When standardized, this pattern shifts notably, with African countries showing greater spread, highlighting the importance of scale in interpreting energy development.

The MDS visualizations reinforce these regional distinctions while preserving distance relationships between countries, maintaining consistent clustering patterns despite random seed variations. This suggests robust underlying relationships in energy development metrics across regions.

The t-SNE visualizations, particularly at higher perplexity values, reveal more nuanced clusters that emphasize both geographic proximity and similarity in energy development stages. Countries with similar access to clean cooking fuels, renewable energy adoption rates, and energy intensity levels cluster together regardless of geographic location, highlighting convergent energy development pathways.

I also create the PCA loading plot to observe the insights into how different energy indicators contribute to the principal components. PC1 has 51% of variance, with Renewable energy share has the strongest positive loading (0.663), indicating this variable significantly drives movement along the positive PC1 axis. Access to clean fuels for cooking has a strong negative loading (-0.584), pointing in the opposite direction

This suggests PC1 primarily represents a contrast between renewable energy adoption and access to clean cooking fuels. Interestingly, countries with higher renewable energy shares often have lower access to clean cooking fuels. This could reflect a development paradox where some countries (particularly developing nations) may have high renewable percentages due to traditional biomass use, yet lack modern clean cooking technologies.

These visualizations collectively demonstrate that progress toward SDG 7 follows regional patterns but with significant intra-regional variations. European and North American nations cluster together, reflecting their more advanced position in clean energy adoption and energy efficiency. The Asian countries display the widest distribution across all visualization techniques, indicating highly varied progress within the region – from advanced energy systems comparable to European standards to developing systems similar to those in Africa.

Impact of Methodological Choices:

The stark contrast between raw and standardized PCA results demonstrates how scaling decisions fundamentally alter our perception of energy development relationships. In the raw data PCA, the scale ranges from approximately -100 to 60 for Component 1 and -60 to 60 for Component 2, creating a widely spread visualization. After standardization, these scales dramatically compress to a range of approximately -2 to 3 for Component 1 and -4 to 2 for Component 2, creating a more balanced distribution of countries.

PCA preserves global variance structure but assumes linear relationships between energy variables, which may not fully capture the complex interactions between energy access, renewable adoption, and intensity metrics. MDS better preserves distance relationships between countries but lacks the interpretable component structure of PCA, making it more challenging to identify which energy indicators drive the observed patterns.

T-SNE excels at revealing local structures and clusters but sacrifices global relationships and is highly sensitive to the perplexity parameter. The dramatic differences observed across perplexity values (3, 10, 100) demonstrate how this hyperparameter shapes our understanding of energy development clusters – low values highlight fine-grained relationships between individual countries, while higher values reveal broader regional patterns in energy development. The consistency of certain patterns across all techniques (such as the general separation of African countries from European ones) strengthens confidence in those findings.

Challenges Encountered:

The initial dataset contained 19 variables ranging from energy access metrics to economic indicators and geographic information. The first major challenge involved selecting appropriate variables from this extensive dataset. I needed to determine which variables were most relevant to SDG 7 while also being sufficiently independent from one another to provide meaningful insights. After careful consideration, I selected four key variables (access to clean fuels for cooking, renewable energy share, energy intensity, and low-carbon electricity percentage) that represented different facets of energy sustainability. Some variables in the original dataset had highly uneven distributions focused on particular scales, making them difficult to visualize effectively in a reduced-dimensional space.

Also. the visualization of 162 countries created potential overcrowding in plots, making individual country identification difficult. This was addressed by using color coding for world regions and ensuring sufficient figure size to maintain readability.

Data-modeling process & Edge Definition Reasoning:

The relationships between SDG goals were derived from the paper "SDG Synergies: An approach for coherent 2030 Agenda implementation" by the Stockholm Environment Institute [1]. Specifically, I concentrated on the strongest positive synergies (depicted as 3 green chevrons in the cross-impact matrix) between SDGs. These synergies represent cases where progress in one goal significantly promotes advancement in another. By analyzing the interaction strength patterns in the matrix, I identified 33 significant positive relationships among the 17 SDGs. I manually created a file named sdg_edges.csv to store all the edge relation pairs. I also grouped the SDG goals into six thematic clusters as they may have same amount of goals based on the article, “Categorize SDGs by Groups for More Impact”[2]:

1. Reducing Overall Inequality (Goals 1, 2, and 10)
2. Access to Safe Conditions (Goals 3, 6, and 7)
3. Sustainable Growth (Goals 8, 9, and 16)
4. Equality Through Education (Goals 4 and 5)
5. Sustainable Partnerships (Goals 11, 12, and 17)
6. Holistic Climate Action (Goals 13, 14, and 15)

Edge List

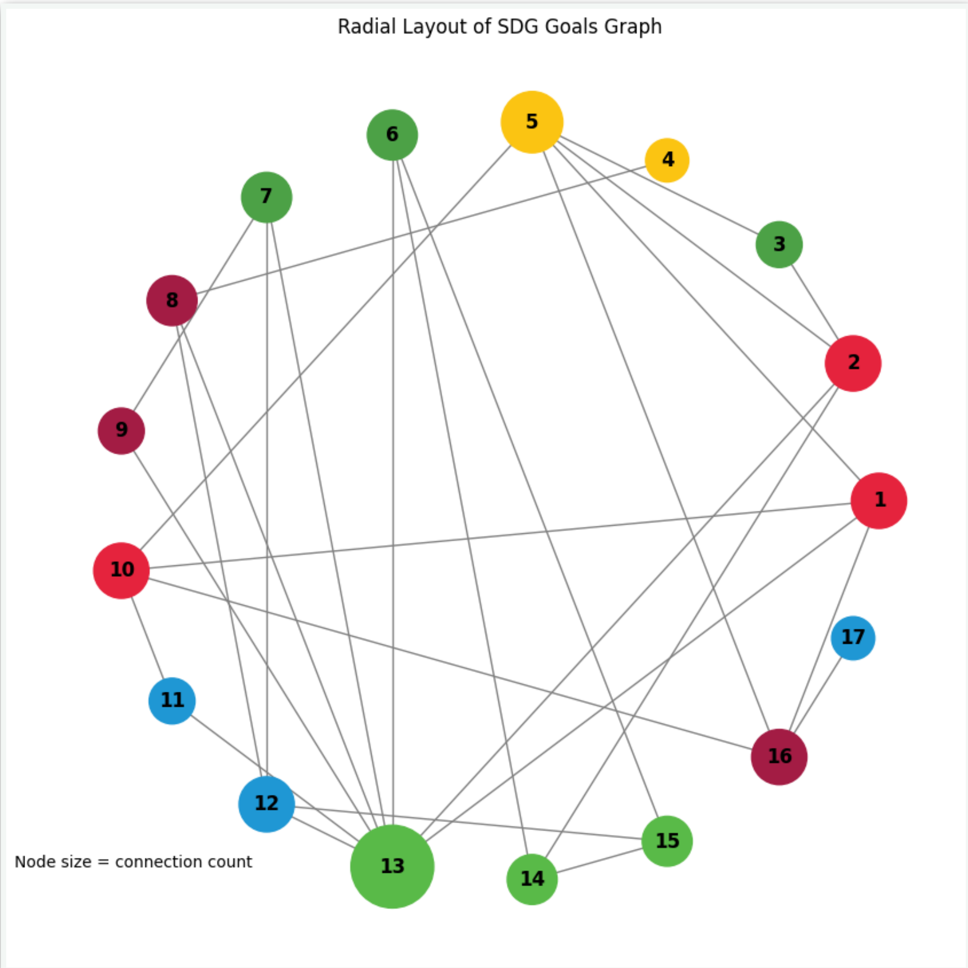
Node1	Node2
1	10
2	3
2	14
4	8
5	1
5	2
5	3
5	10
6	14
7	9
7	12
7	13
8	12
9	13
10	1
11	10
11	13
12	8

[1] Weitz, N., Carlsen, H. and Trimmer, C. (2019). SDG Synergies: An approach for coherent 2030 Agenda implementation. Stockholm Environment Institute. https://www.sei.org/publications/sdg-synergies-factsheet/

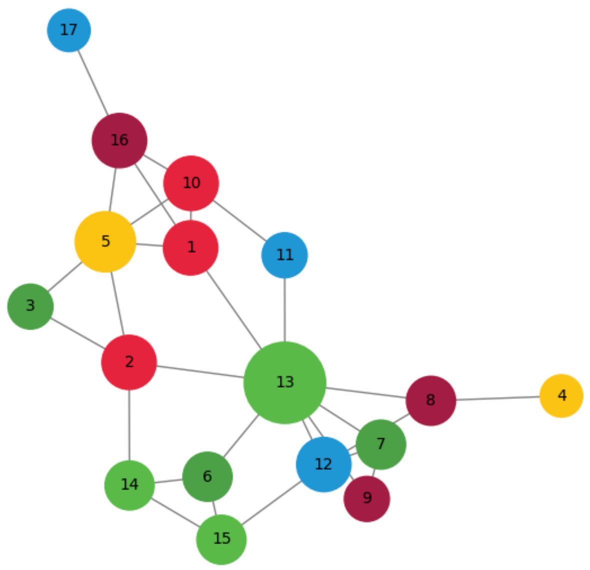
[2] Categorize SDGs by Groups for More Impact, https://abundantearthfoundation.org/categorize-sdgs-by-groups-for-more-impact/

Node1	Node2
12	13
12	15
13	1
13	2
13	6
13	7
13	8
14	2
14	6
15	6
15	14
16	1
16	5
16	10
16	17

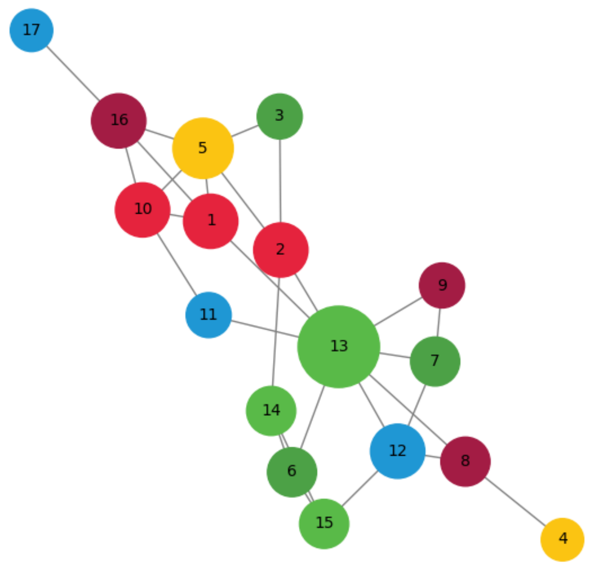
Visualizations:



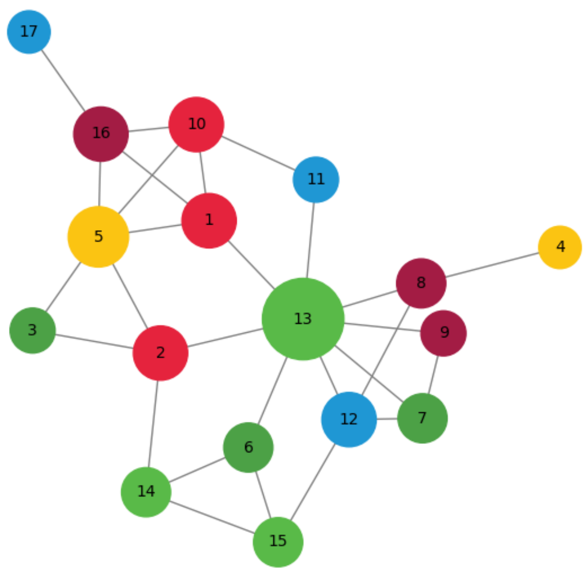
Fruchterman-Reingold Layout (Seed 42)



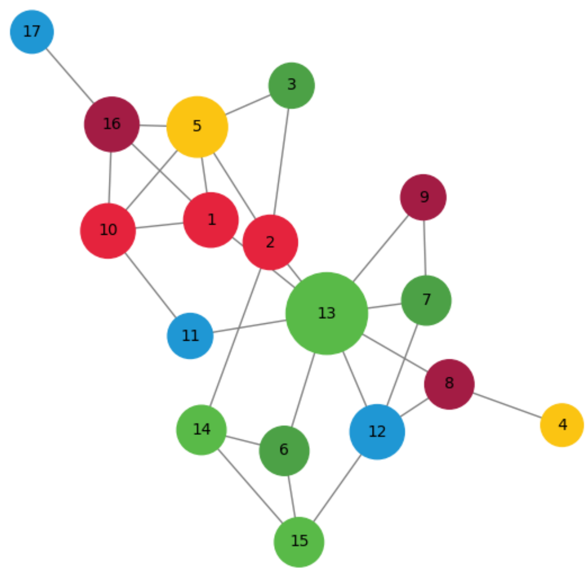
Fruchterman-Reingold Layout (Seed 100)



Kamada-Kawai Layout (Seed 42)



Kamada-Kawai Layout (Seed 100)



Interpretation:

In the radial layout visualization, I arranged all nodes in a circular pattern following numerical order. To enhance readability, I designed the nodes with sizes reflecting their connectivity – the more connections a goal has, the larger its corresponding node appears. This design choice immediately reveals that node 13 (Climate Action) has numerous connections, suggesting its pivotal role in SDG synergies. The layout effectively displays relationships across different thematic areas, with edges traversing the circle to connect various SDGs.

The Kamada-Kawai and Fruchterman-Reingold layouts produce remarkably similar clustering patterns across both seed variations (42 and 100). Node 13 consistently emerges as a central node with high connectivity, demonstrating how climate action interconnects with multiple sustainable development areas. The green-colored nodes (representing environmental sustainability goals and one from access to safe conditions) form a distinct cluster, particularly nodes 13, 14, 15, and 6.

Across all visualizations, it's evident that SDG 13 (Climate Action) plays a central role in connecting various sustainability goals. This suggests that addressing climate action challenges can significantly contribute to progress in other areas, including SDG 7 (Affordable and Clean Energy), which was my primary focus area.

Impact of Methodological Choices:

The radial layout offers a structured and predictable view with nodes arranged numerically around a circle. While visually clean, this approach somewhat obscures natural clustering patterns. The connections appear more distributed, making it harder to identify central nodes without the visual cue of node sizing.

In contrast, the Kamada-Kawai and Fruchterman-Reingold layouts naturally reveal clusters by positioning connected nodes closer together. Despite using different random seeds (42 vs 100), which produce variations in exact node positioning, both algorithms maintain consistent overall structural patterns. Node 13 reliably appears as central across all force-directed visualizations, confirming that this finding is robust rather than an artifact of visualization parameters.

My choice to size nodes based on connection count effectively highlights which SDGs have the most synergies (particularly node 13), while the color-coding by thematic groups helps identify patterns of connectivity within and between different sustainability domains.

Challenges Encountered and Solutions:

Working with a network of 17 nodes and more than 30 connections presented significant visualization challenges. The primary concern was excessive visual clutter that could obscure important patterns. To address this, I implemented several strategic solutions. First, I used meaningful color grouping by assigning nodes to six thematic clusters rather than arbitrary coloring. This approach created immediate visual patterns that help viewers mentally organize the complex network structure.

Second, I scaled node sizes proportionally to their connection count, making highly connected nodes (especially node 13) immediately distinguishable. By using exponential scaling rather than linear scaling, I ensured that differences in connectivity remained visually apparent even between nodes with similar connection counts.