Assignment 3: Human Factors

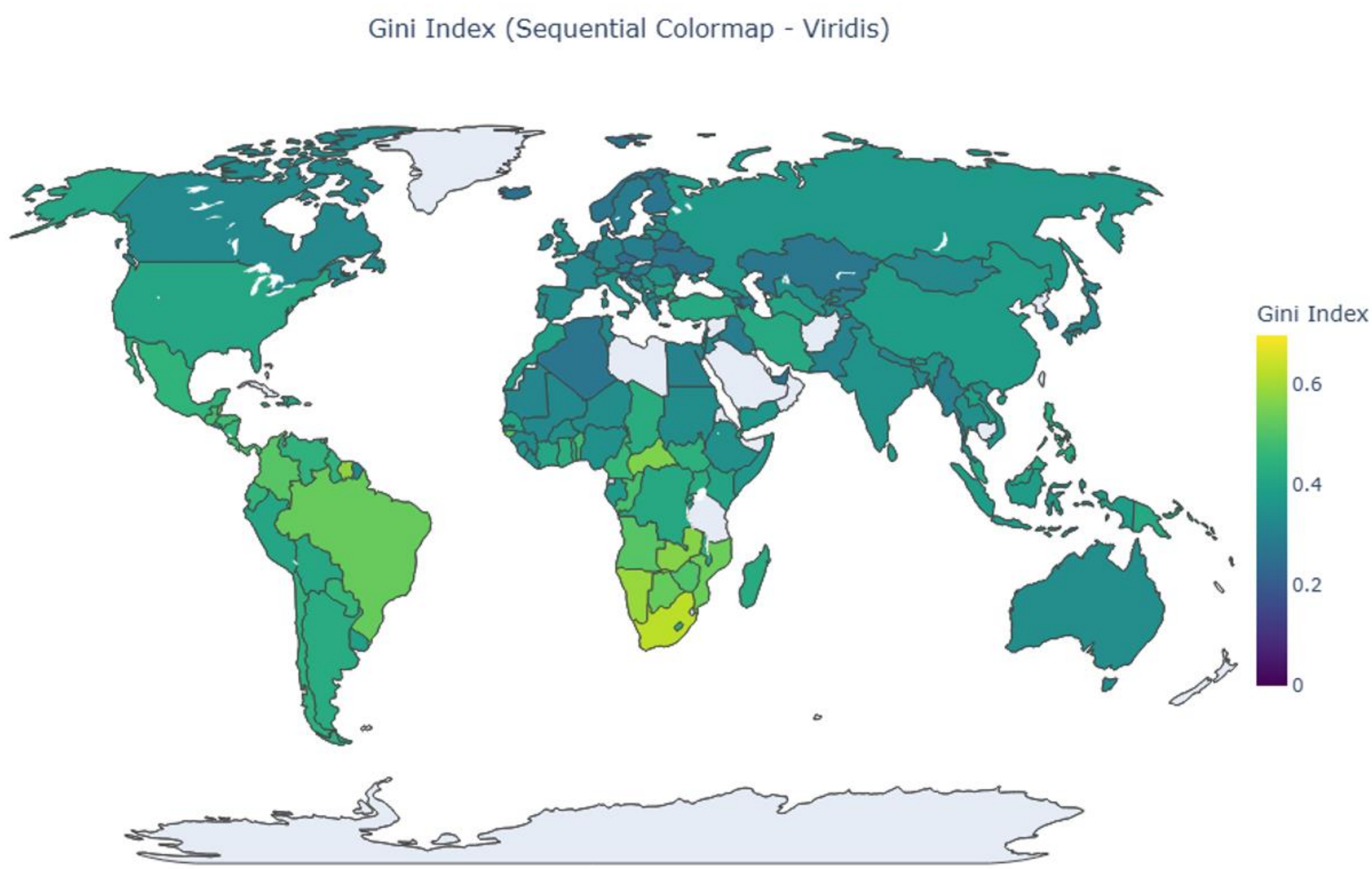
Submission

Description: My chosen SDG in my submission is ***Reduced Inequalities***.

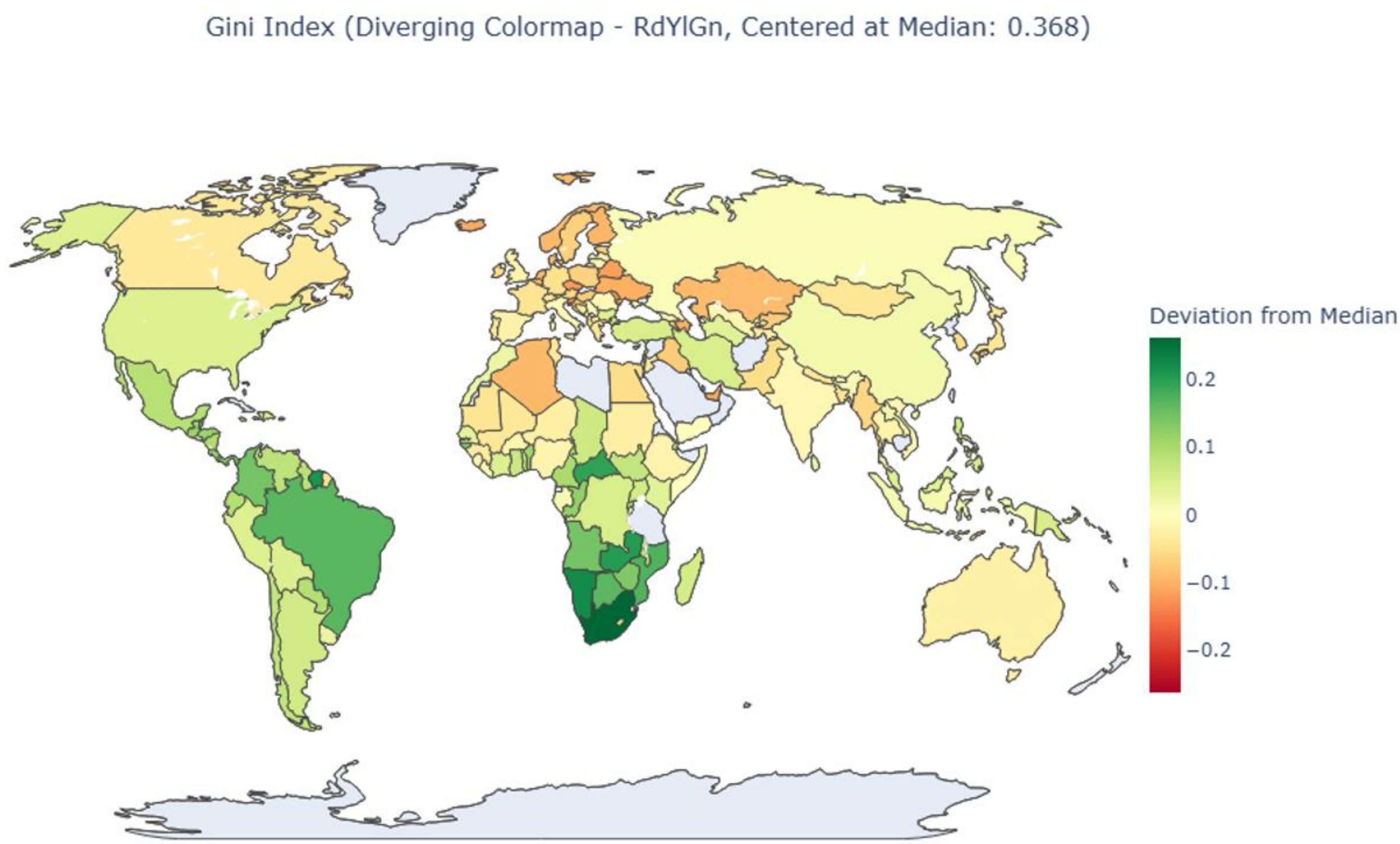
Task 1: Maps and Colormaps

- **Dataset:** Economic Indicators: GDP and Gini Index
- **Link to the dataset:** https://www.kaggle.com/datasets/shahriarkabir/economic-indicators-gdp-and-gini-index?select=EconomicData.csv
- **Visualization tools:** Python with Matplotlib, Seaborn libraries, and plotly.express.

Visualization:



Visualization 1



Visualization 2

These two choropleth maps visualize the **Gini Index**, which measures the income inequality ranging from 0 (perfect equality) to 1 (perfect inequality) all over the world.

For visualization 1 **Sequential Colormap** (*Viridis*), I use the sequential color palette *Viridis*, changing from dark purple (low inequality) to bright yellow (high inequality), with color spanning from 0 to approximately 0.6. Sequential colormap is suitable for ordered and continuous data because Gini Index shows an increase from low inequality to high inequality. The *Viridis* colormap is perceptually uniform, which means that colors have perceptually the same ordering as the scale. I chose *Viridis* palette because the transition from dark color to bright color enhances the ability to distinguish the small differences in Gini Index, making it easy to track the inequality levels globally. Moreover, this colormap emphasizes different levels of inequality, allowing viewers to identify regions with the highest and lowest values without reference to a midpoint. This design supports a clear distribution of inequality. The color bar (0 – 0.6) provides a clear quantitative reference and clear titles improve viewers' interpretation, ensuring that the map's purpose is apparent. I use white background, which is contrast with colormap, to highlight the distribution of Gini Index across nations. Countries are colored from dark purple (0 – 0.2) to bright yellow (0.4 – 0.6). Scandinavia countries (such as Finland), Cananda, and Kazakhstan are in dark purple, indicating the low inequality. South Africa and Latin America (such as Brazil) show bright yellow, indicating the high inequality. Most of other countries in North America, Asia, and Western Europe show mid-range green (0.2 – 0.4). High-inequality regions (yellow) like Southern Africa and Latin America suggest persistent challenges, requiring targeted solutions. Low-inequality areas (purple) like Northern Europe indicate progress, likely due to strong social policies. This map highlights where inequality is most outstanding, aligning with SDG 10's call for action in inequality.

For visualization 2 **Diverging Colormap** (*RdYlGn*), I use the diverging color palette *RdYlGn*, centered at the median Gini Index (0.368), with red indicating lower inequality (below the median), yellow indicating near the median, and green indicating higher inequality (above the median). The deviation scale is from – 0.2 to 0.2. I use the diverging sequences: zero has neutral color yellow and opposite ends use opponent colors red – green. This choice is to have the light response of green and red color to shift toward the other, which reduces range of trichromatic perception and can have variable effects on color vision. This colormap emphasizes outliers—countries significantly above or below the median—making it effective for assessing overall comparison. The median was chosen over the mean to avoid skewing by extreme values, ensuring a representative midpoint. At the median value 0.368, countries below the median (such as Scandinavia countries, Canada, Kazakhstan) are red (deviations approximately – 0.1 to – 0.2). Countries near the median (such as Russia, China, and Australia) are yellow and countries above the median (such as South Africa and Brazil) are dark green (deviations approximately 0.1 to 0.2). Green regions (high inequality) are clear priorities for SDG 10, showing significant deviation above the global standard. Red regions demonstrate success in reducing inequality and yellow areas suggest moderate progress. This map provides a good visualizaition at comparing countries to the median, revealing both successes and gaps. The regions in gray color such as Greenland and Antarctica have the data not recorded.

Perceptual differences between the two choropleth maps: The sequential colormap is good at showing the full distribution of inequality, ideal for global review, while the diverging colormap focuses on deviations, better for identifying inequalities relative to a benchmark. For diverging colormap, the transition through yellow may be less intuitive for a strictly increasing metric like the Gini Index, and red-green distinctions could challenge colorblind viewers. Perceptually, *Viridis* palette shows clarity and accessibility, whereas *RdYlGn* prioritizes highlighting extremes at the cost of some intuitiveness.

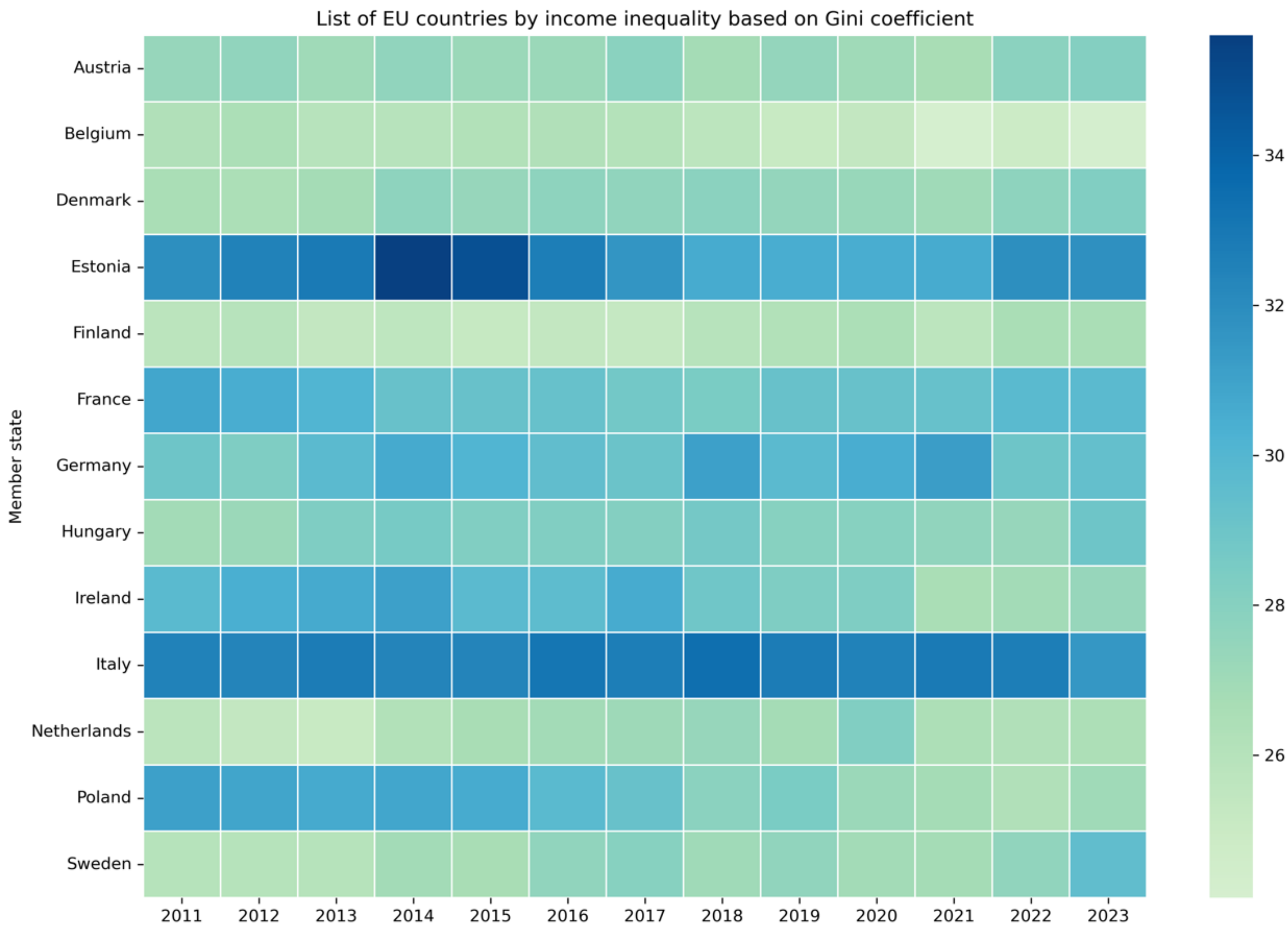
In summary, the sequential map provides a broad, absolute view of inequality, while the diverging map highlights relative deviations, both offering an helpful tool for assessing SDG 10 progress and inspiring action.

The challenges: I do not know how to visualize a map so I have to follow the tutorial video on YouTube, which helps me a lot in completing this task. I have learnt how to code map by using *plotly.express*. I also find it difficult to find some datasets which are different from ones in previous submissions. I want to explore new datasets for each part of projects so searching for information on many websites is really important for completing this project.

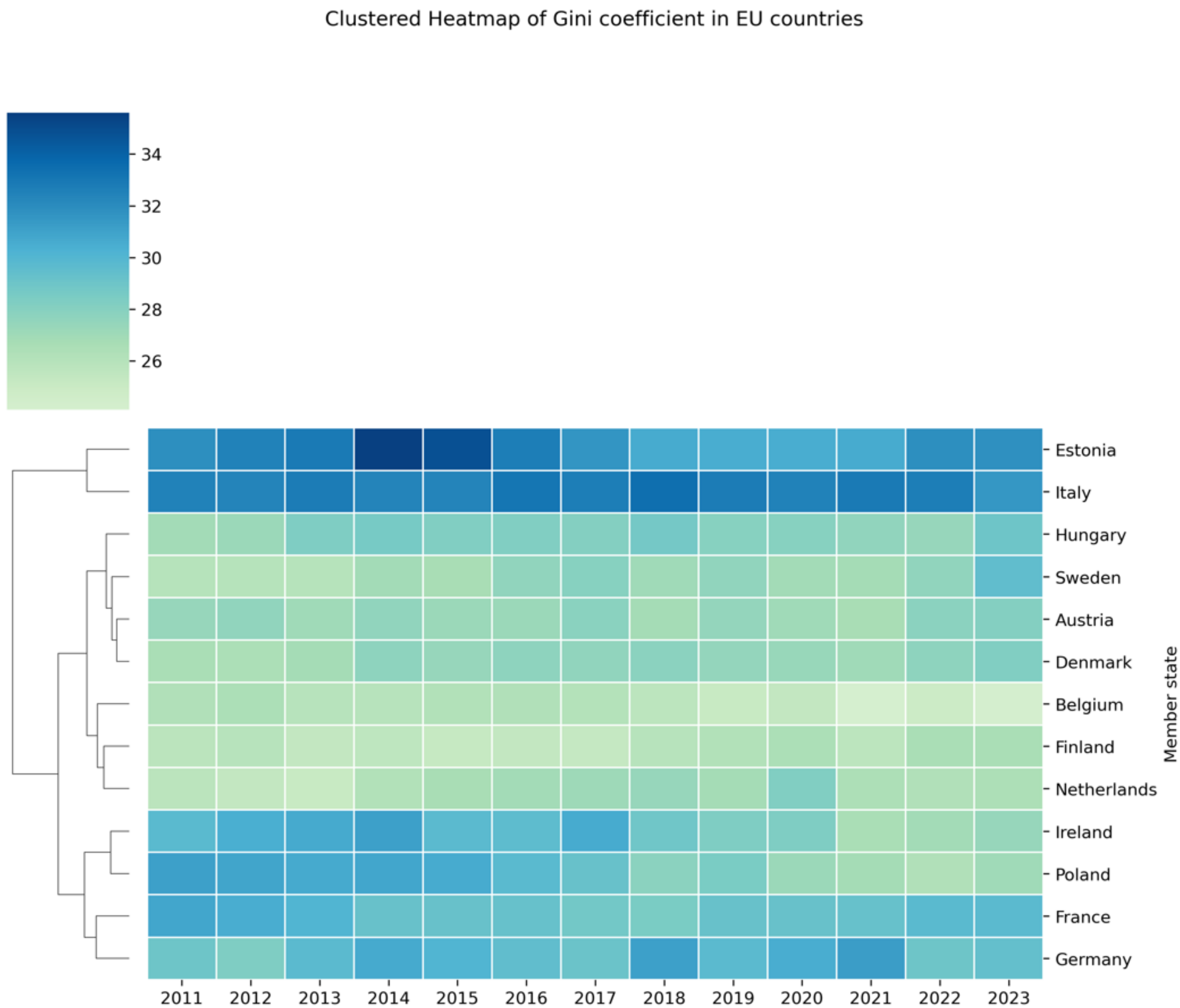
Task 2: Heatmaps and Clustermaps

- **Dataset:** List of EU countries by income inequality based on Gini coefficient
- **Link to the dataset:** [List of countries by income inequality - Wikipedia]
- **Visualization tools:** Python with Matplotlib, Seaborn libraries, and plotly.express.

Visualization:



Visualization 3



Visualization 4

For visualization 3 ***Heatmap***, the map visualizes Gini coefficients, which measures the income inequality, across 13 different EU countries for a period of 13 years from 2011 to 2023. The rows are alphabetically ordered countries and the columns are sequentially ordered timestamps. Each cell's color represents the Gini coefficients for a specific country and year, with values ranging from approximately 26 (low inequality) to around 33 (high inequality). I chose a sequential colormap with the palette *GnBu* which transitions from light green (low Gini values) to dark blue (high Gini values). This choice aligns with the colormap principles for ordered and continuous variables. Because the Gini coefficients are continuous data, a sequential colormap ensures that increases in inequality are represented by a smooth transition from light colors to dark ones. The palette *GnBu* avoids perceptual jumps, helping viewers distinguish small differences. Lighter shades (green) typically suggest better outcomes like less inequality, while darker shades (blue) suggest worse outcome like more inequality, which aligns with human perception. I use alphabetically ordered variable in this visualization because it allows viewers to find specific countries as fast as possible. I also include the scale bar on the right of the map to be used as a reference for viewers. In addition, I add the linewidth 0.5 between each adjacent cell to help viewers distinguish which cell belongs to a specific country and year more easily. Noticeably, Estonia and Italy are colored in dark blue (32 – 34), which remains stable from 2011 to 2023, indicating that both countries have high income inequality. Other countries such as France, Germany, Hungary, and Ireland have the average Gini coefficients around 28 – 30 for the period we are considering. Scandinavia countries such as Finland and Sweden have low Gini coefficients in light green (around 26), which indicates the high equality due to strong economic policies. We also find that from 2011 the Gini coefficient of Poland decreases, which shows that there is an improvement in dealing with income inequality in this country.

For visualization 4 ***Clustermap***, I builds on the heatmap by applying hierarchical clustering to the rows (countries) based on their Gini coefficient trends over time, while preserving the original temporal structure (2011 – 2023) in the columns. A dendrogram on the left visualizes the clustering hierarchy, which is a special feature of clustermap. I use the palette *GnBu*, which is the same colormap as ***Heatmap***, to remain the same interpretation across visualizations, reducing the cognitive load. This choice offers exact perception of clustered patterns, with light colors for high equality and dark ones for high inequality. Observation at the visualization shows that Estonia and Italy may group together due to their consistently high Gini coefficients (dark blue, 32 – 34). Netherlands and Finland may cluster due to lower, stable Gini coefficients (light green, 26 – 28). Within clusters, stable low-inequality groups such as Nordic countries contrast with others showing darkening trends. For dendrogram, shorter branches between Estonia and Italy suggest strong similarity, while longer branches to Finland and Netherlands indicate divergence, which corresponds to SDG interpretation. Stable low-inequality clusters such as Nordic countries show effective strategies, while divergent or darkening clusters show areas needing actions.

To create visualization 4, I use *clustermap* function in *seaborn documentation* with several parameters. Firstly, I use metric parameter, a distance used to measure the dissimilarity between nations' Gini coefficients. This parameter is effective for continuous data like Gini coefficients. Secondly, I use linkage method to calculate clusters, forming balanced clusters that hightlight groupings of countries with similar inequality trends. Thirdly, I use row clustering to reorder countries based on the similarity, emphasizing the relationships between variables. I do not apply column clustering to maintain the original temporal structure (2011 – 2023). The dendrogram is also included to show the hierarchical relationships between the countries. I do not apply standardization because I want to keep the same interpretation as ***Heatmap*** and Gini coefficients are very good for direct comparison and SDG interpretation.

Perceptual Differences Between Heatmap and Clustermap: ***Heatmap*** requires manual pattern identification, prioritizing reference over insight while ***Clustermap*** makes relationships between variables immediately outstanding but may complicate tracking a single country without rechecking labels. In general, the ***Heatmap*** is an excellent choice for quick reference, while the ***Clustermap*** improves analytical depth by revealing hidden structures.

In summary, ***Heatmap*** provides a clear visualization of inequality trends of countries based on Gini coefficients, while ***Clustermap*** reveals deeper group-level patterns. Together, they highlight successes and challenges in achieving SDG 10 ***Reduced Inequalities***.

The challenges: I have to differentiate the differences between heatmaps and clustermaps because they seem to have the same visualization. Therefore, after familiarizing myself with the idea of a hierarchically clustered heatmap in *seaborn documentation*, I know how to visualize these different maps in an effective way. In general, this task seems to be easier than the previous task.