2025-HX05-COS30045-Data Visualisation (HCMC)
Visualisation Assignment (Week 5)

Instructor: Mr.Tristan Nguyen
Student name: NGUYEN ANH KHOA
Student ID: 104856650
Word count: 2434
Submitted date: 08/06/2025

Contents

PART A - VISUALISATION CRITIQUE

1. Introduction

Environmental innovation—particularly through green technology patents—is a key indicator of a country's commitment to sustainable development. These patents signal technological capacity, economic opportunity, and policy direction. OECD patent statistics offer a credible foundation for analysing such trends across countries and over time.

This critique examines three visualisations that present OECD data on environmental innovation, each tailored to a specific audience and using a distinct visual strategy: a spatial map of green patents in Nordic countries, a global scatter plot showing growth patterns, and a multi-country time series. The aim is to assess how effectively these charts convey insights using core visualisation principles such as clarity, accuracy, perceptual design (e.g., Gestalt principles), accessibility, and ethical representation.

By applying frameworks such as Munzner's design triangle, Gestalt laws, and Cleveland & McGill's graphical perception hierarchy, this analysis identifies the strengths and weaknesses of each visualisation. It explores how design choices impact the viewer's ability to interpret complex environmental innovation data and make informed decisions.

2. Visualisation 1: Green Patents in the Nordic Region (2006–2011)

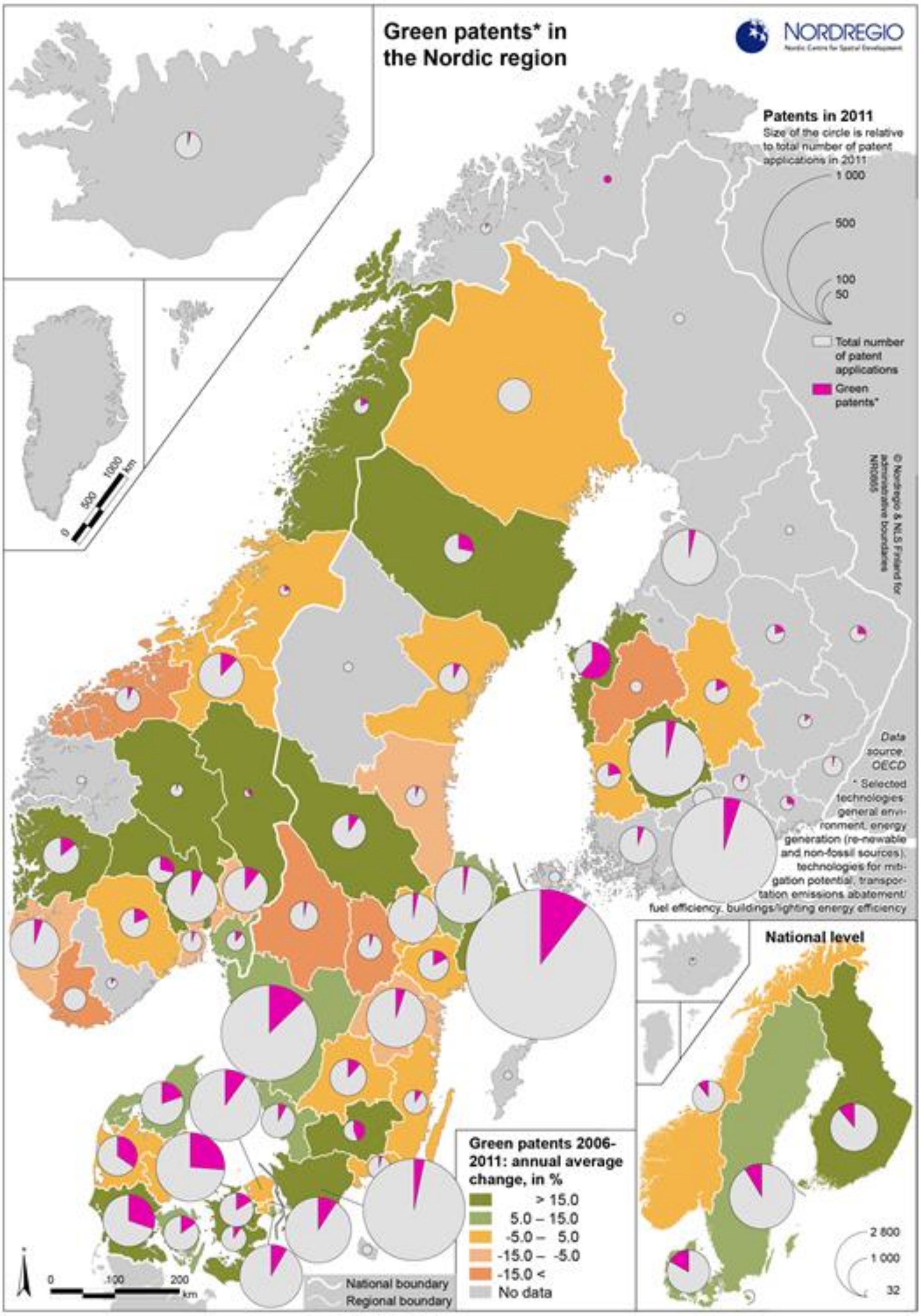


Figure 1: Regional Innovation in Green Technologies

- **Source:** Nordregio (https://nordregio.org)

- **Target Audience:** Policymakers, regional planners, innovation agencies in Nordic countries
- **Purpose:** Show spatial and thematic distribution of green patents by region, category, and growth rate

Data Representation	This visualisation integrates multiple data elements, including: - Patent counts for the year 2011 - Average annual growth rates from 2006 to 2011 - Regional breakdown at the sub-national level - Technological composition displayed through pie charts
Visualisation Type & Encoding Techniques	The visualisation combines several graphical methods: - A choropleth map uses color gradients to indicate regional growth rates - Proportional circles reflect total patent volumes based on their size - Embedded pie charts within each circle represent the share of different green technology categories through segmented color coding
Strengths	• ***High Information Density:*** The integration of spatial, numerical, and categorical data provides a multidimensional perspective on innovation. • ***Clear Geographic Context:*** The regional mapping effectively identifies areas with concentrated green technology development. • ***Effective Legend Usage:*** The legend clearly supports all encoding types, aiding interpretation.

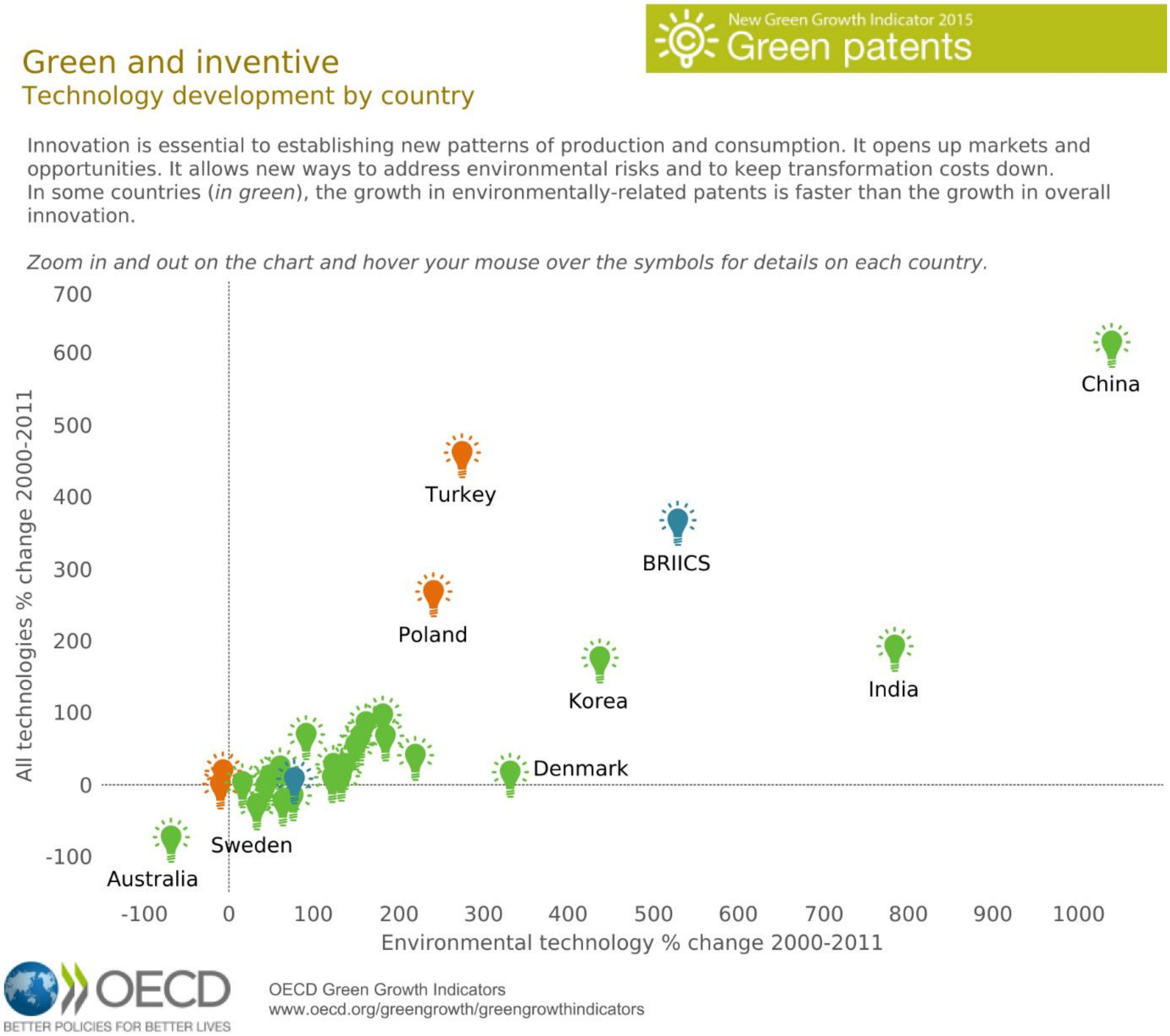
	• ***Color Gradient Intuition:*** The use of light-to-dark shading to represent growth levels is visually intuitive and aids rapid understanding.
Weaknesses	• ***Visual Overload:*** The use of multiple encoding techniques—color, size, and pie segmentation—can overwhelm viewers and increase cognitive effort. • ***Unreadable Small Pie Charts:*** In regions with limited space, pie charts become too small to discern individual segments, reducing the effectiveness of categorical data representation. • ***Color Interference:*** The background map shading can clash with pie chart colors, which may reduce clarity—especially for color-blind viewers. • ***Unclear Scaling:*** The proportional size of the circles does not convey quantitative differences clearly without referring back to the legend.
Application of Gestalt Principles	• ***Proximity:*** The alignment of pie charts with proportional circles strengthens the connection between quantity and composition. • ***Similarity:*** Consistent use of color within pie segments supports recognition of shared technology types. • ***Figure-Ground Separation:*** While the base map and overlays are distinguishable, visual clarity could be

	enhanced through improved layering or contrast adjustments.

- **Overall Evaluation**

This visualisation delivers a rich and sophisticated portrayal of environmental innovation patterns suitable for expert analysis. However, the simultaneous use of multiple encoding dimensions may reduce accessibility for non-expert audiences. Adding interactive features (such as zoom or filtering) or simplifying the number of visual elements could significantly improve user engagement and comprehension without sacrificing analytical depth.

3. Visualisation 2: Green and Inventive – Technology Growth by Country (2000–2011)



- **Source**: OECD Green Growth Indicators
- **Intended Audience**: Policy analysts, researchers in comparative innovation
- **Objective**: To assess and compare national growth trends in environmental technologies relative to overall technological development

Data Representation	This chart visualizes four key data elements: - The growth rate of environmental technologies (positioned along the X-axis)

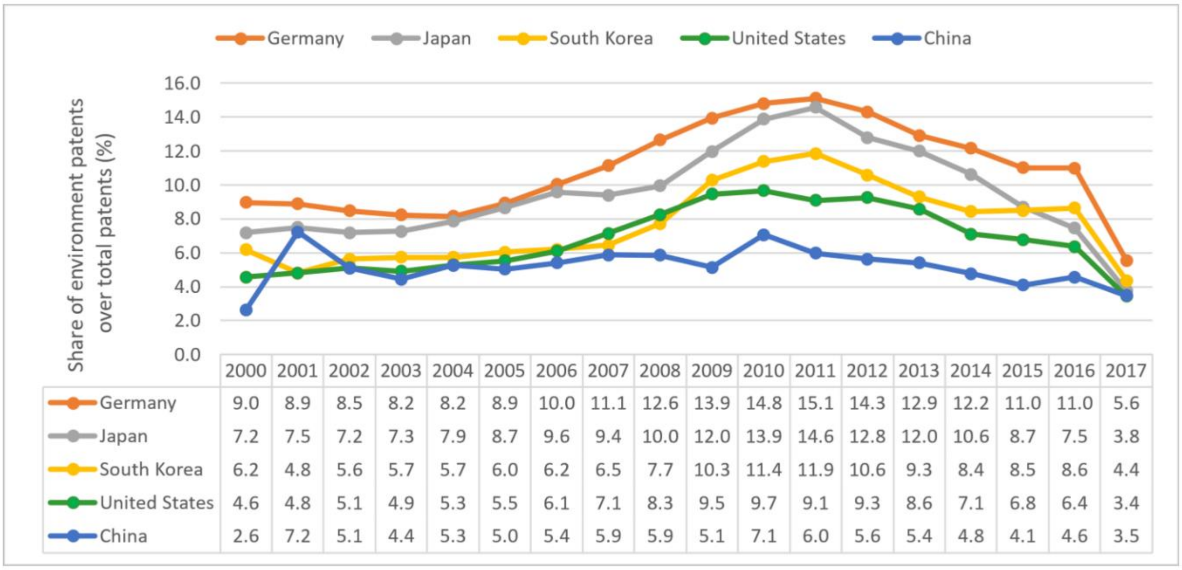
	- The growth rate of total technological innovation (Y-axis) - Country identifiers displayed as text labels - A visual emphasis on nations where environmental innovation exceeds general technological growth
Visualisation Type & Encoding Techniques	The visualisation uses a scatter plot format to depict two-dimensional growth comparisons: - Lightbulb icons serve as data markers, symbolising technological innovation - Color coding: Green icons indicate countries where green technology growth is higher than total innovation growth - Labels: Country names are added to identify each data point
Strengths	• ***Effective Quantitative Comparison***: The scatter plot structure is highly appropriate for visualising relationships between two continuous variables. • ***Thematic Symbolism***: The use of lightbulb symbols reinforces the central theme of innovation, making the visual more engaging. • ***Visual Emphasis***: Green-colored icons quickly draw the viewer's attention to countries leading in green innovation. • ***Supportive Annotation***: Explanatory text or captions clearly outline the purpose and findings, aiding comprehension.

Weaknesses	• ***Label Clutter:*** In areas where data points are densely clustered—especially within Europe—overlapping labels reduce readability and visual clarity. • ***Poor Alignment:*** The positioning of country names often fails to clearly correspond to their respective data points. • ***Imbalanced Axis Scaling:*** The uneven scale between the X and Y axes may lead to distorted interpretations of comparative growth. • ***Lack of Interactivity:*** While interactive features like tooltips or filters are suggested, the static design limits deeper exploration of the data.
Design Principles in Practice	• ***Accuracy:*** The chart communicates its core message accurately but suffers from reduced interpretability due to label congestion. • ***Preattentive Processing:*** Use of color and iconography enables rapid identification of key patterns and outliers. • ***Visual Aesthetics:*** The overall design maintains a clean layout and adequate white space but lacks interactive or dynamic elements to improve user engagement.

- **Overall Evaluation**

This visualisation demonstrates a clear and conceptually strong design for comparing national innovation trends. Its thematic consistency and use of visual cues are commendable. However, the execution is hindered by practical usability issues, particularly label overcrowding. Enhancing the chart with interactive elements (e.g., tooltips, zooming) or applying label decluttering strategies (such as grouping or toggles) would greatly improve its readability and usefulness for policy decision-making.

4. Visualisation 3: Share of Environmental Patents (2000–2017)



- **Source:** MDPI, Technological Eco-Innovation (https://www.mdpi.com)
- **Target Audience:** Researchers in innovation, economists, and policy advisors
- **Purpose:** To illustrate the evolution of the share of environmental patents over time in five major economies

Data Representation	This visualisation encompasses the following key elements: - A time span covering the years 2000 to 2017 - Five countries: Germany, Japan, South Korea, the United States, and China - The core metric displayed is the percentage share of environmental patents - A supplemental data table beneath the chart provides exact numerical values for each year and country

Visualisation Type & Encoding Techniques	The visualisation uses a scatter plot format to depict two-dimensional growth comparisons: - Line chart: Tracks annual trends in environmental patent shares across countries - Color-coded lines: Each country is represented by a distinct line color - Legend: Identifies each country's line via color association - Data table: Serves as a precise numerical reference for each plotted data point
Strengths	• ***Effective Use of Line Chart:*** The chart type is well-suited to depict longitudinal trends and facilitates comparison across countries. • ***Dual Modal Presentation:*** The inclusion of a numerical table enhances precision and supports users who prefer raw data. • ***Temporal Consistency:*** Regular annual intervals support cognitive tracking of changes over time. • ***Clear Distinction Through Color:*** Country trends are differentiated through a consistent color legend.
Weaknesses	• ***Limited Accessibility:*** Some color combinations (e.g., orange/yellow, blue/green) may pose difficulties for viewers with color vision deficiencies.

	• ***Visual Clutter from Line Intersections:*** During periods of close convergence (notably between 2008 and 2012), intersecting lines can obscure individual trends. • ***Lack of Contextual Explanation:*** Significant shifts—such as the sharp decline in several countries after 2012—are presented without annotation or contextual interpretation. • ***Suboptimal Legend Placement:*** Positioning the legend at the top of the chart disrupts the natural left-to-right reading flow; a side alignment would support smoother interpretation.
Application of Perceptual and Ethical Principles	• ***Gestalt Principles – Continuity and Similarity:*** Uniform line formatting reinforces a sense of continuity and comparability between countries. • ***Ethical Representation:*** The chart maintains an honest scale and displays the full data range, avoiding any manipulation of trends. • ***Cognitive Load:*** Generally manageable, though overlapping lines may slightly increase the effort required for interpretation during dense periods.

- **Overall Evaluation**

This visualisation provides a robust and transparent depiction of temporal trends in environmental innovation. Its combination of graphical and tabular elements caters to different user preferences, from exploratory to precise analysis. However, its accessibility and interpretability could be significantly improved through better color differentiation, strategic legend placement, and the inclusion of contextual annotations that explain key changes or anomalies in the data.

5. Conclusion and Comparative Analysis

The three visualisations, based on OECD patent data, aim to convey insights into environmental innovation using well-chosen chart types: a choropleth map for spatial patterns, a scatter plot for comparative growth, and a line chart for temporal trends. While these formats suit their analytical goals, effectiveness varies depending on design execution and audience needs.

Color is consistently used to highlight trends, but none of the charts address accessibility for color-blind users. Visual complexity is also a recurring issue: the first visual is rich but overwhelming, the second suffers from label clutter, and the third lacks explanatory annotations and uses visually similar lines that hinder clarity.

A key limitation is the lack of interactivity. Features like tooltips, zoom, or filters would improve user experience and insight. Contextual notes—such as explanations for post-2012 declines—would also enhance understanding.

In sum, effective visualisation goes beyond accurate representation. It requires thoughtful design, accessible encoding, and contextual framing. Future improvements should focus on interactivity and inclusivity to better engage diverse audiences.

PART B - Interactive Dashboard Implementation

1. Conclusion and Comparative Analysis

- **<u>Data Source and Overview</u>**

The dataset utilized in this project was sourced from the OECD STI Micro-data Lab, specifically from the patent statistics section. It includes information on priority patent filings categorized by inventor, country, and technology domain, spanning the years 2010 to 2021. The original file name was:

"*OECD.STI.PIE,DSD_PATENTS@DF_PATENTS,1.0+.A...PRIORITY...INVENTOR..._T.csv*
"

This dataset provided the foundation for all visual analyses and served as the input for further data preprocessing.

```
# Read the original data from the OECD CSV file
df = pd.read_csv("OECD.STI.PIE,DSD_PATENTS@DF_PATENTS,1.0+.A...PRIORITY...INVENTOR..._T.csv")
```

- **Preprocessing Workflow**

The raw data was prepared using a custom Python script titled preprocess_patent_data.py. The script performed the following data-cleaning and transformation steps:

✓ *Column Selection and Renaming*

To enhance clarity and usability, only relevant columns were retained and renamed for interpretability:

Original Column Name	New Label
TIME_PERIOD	Year
OBS_VALUE	Patent_Count
PATENT_AUTHORITIES	Patent_Office
OECD_TECHNOLOGY_PATENT	Tech_Domain
REF_AREA	Country
MEASURE	Measure_Type

```python
# Select and rename important columns
df_cleaned = df[[
    'TIME_PERIOD',
    'OBS_VALUE',
    'PATENT_AUTHORITIES',
    'OECD_TECHNOLOGY_PATENT',
    'REF_AREA',
    'MEASURE'
]].rename(columns={
    'TIME_PERIOD': 'Year',
    'OBS_VALUE': 'Patent_Count',
    'PATENT_AUTHORITIES': 'Patent_Office',
    'OECD_TECHNOLOGY_PATENT': 'Tech_Domain',
    'REF_AREA': 'Country',
    'MEASURE': 'Measure_Type'
})
```

- ✓ *Missing Value Handling*

To maintain data reliability, rows with missing values in critical fields—**Year**, **Patent_Count**, **Country**, or **Tech_Domain**—were removed using the dropna() function.

```python
# Remove missing values
df_cleaned.dropna(subset=['Year', 'Patent_Count', 'Country', 'Tech_Domain'], inplace=True)
```

- ✓ *Data Type Conversion and Filtering*

Ensured Year and Patent_Count columns were converted to numeric types for accurate computation.

A filter applied only records from 2010 onwards.

Removed entries with negative patent counts to exclude invalid or erroneous data.

```python
# Convert data types and filter the dataset
df_cleaned['Year'] = pd.to_numeric(df_cleaned['Year'], errors='coerce')
df_cleaned['Patent_Count'] = pd.to_numeric(df_cleaned['Patent_Count'], errors='coerce')
df_cleaned = df_cleaned[(df_cleaned['Year'] >= 2010) & (df_cleaned['Patent_Count'] >= 0)]
```

✓ *<u>Output File Generation</u>*

The final cleaned dataset was exported as **cleaned_oecd_patent_data.csv**, which served as the primary input for the interactive visualisation dashboard.

```
# Export to CSV file
df_cleaned.to_csv("cleaned_oecd_patent_data.csv", index=False)

print("✅ Data has been cleaned and saved to cleaned_oecd_patent_data.csv")
```

2. Interactive Dashboard Development

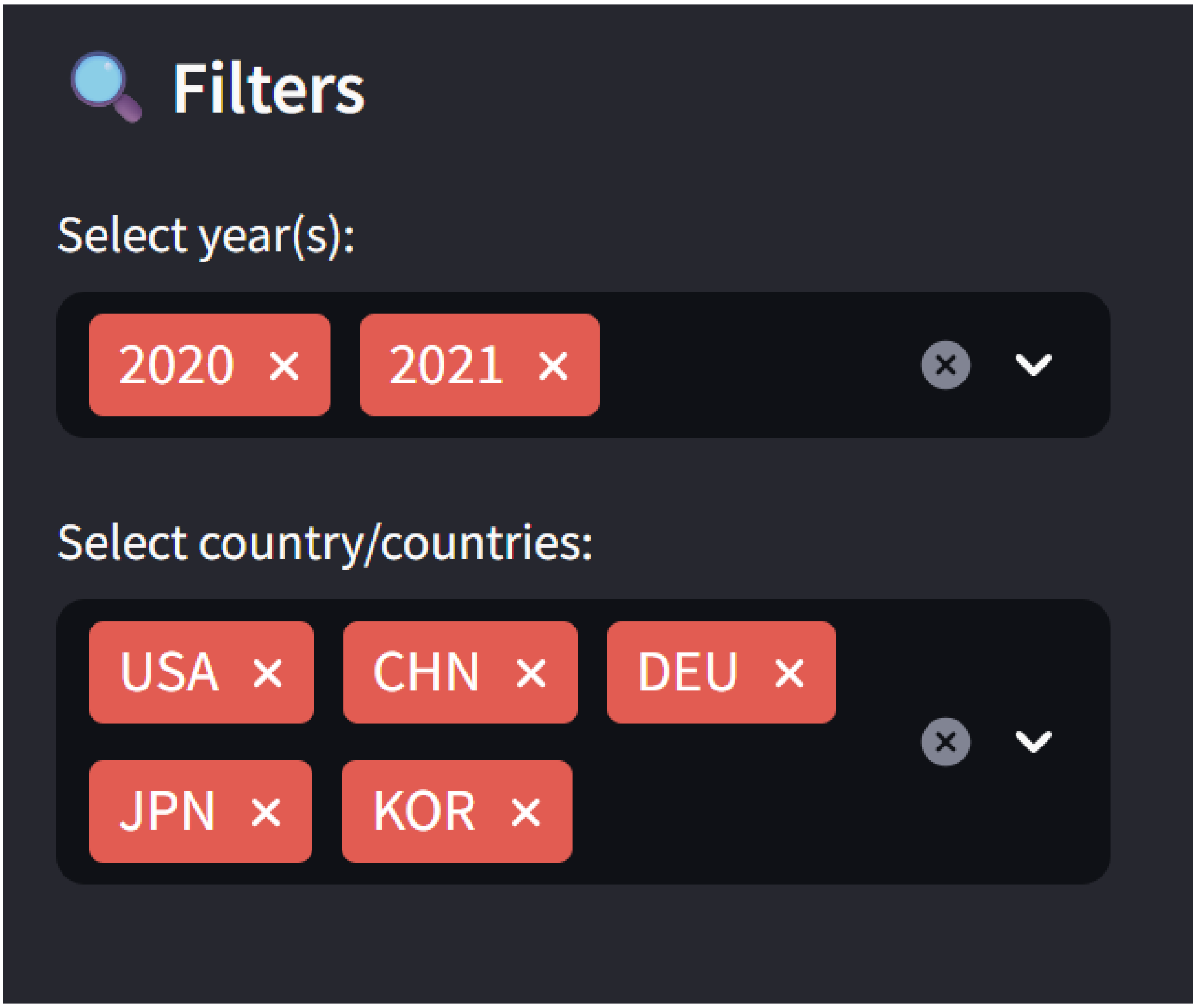
- **<u>Overview</u>**

An interactive dashboard was developed using Python, combining Streamlit for user interface design and Plotly Express for rich visualisation components. The implementation is contained within the script dashboard_final.py, which integrates six interlinked and interactive visualisations to provide a comprehensive exploration of the OECD environmental patent dataset.

- **<u>Dashboard Features</u>**

A sidebar enables user-driven filtering across multiple dimensions:

- Year: Multi-select functionality for choosing one or more years
- Country: Multi-select for regional comparisons



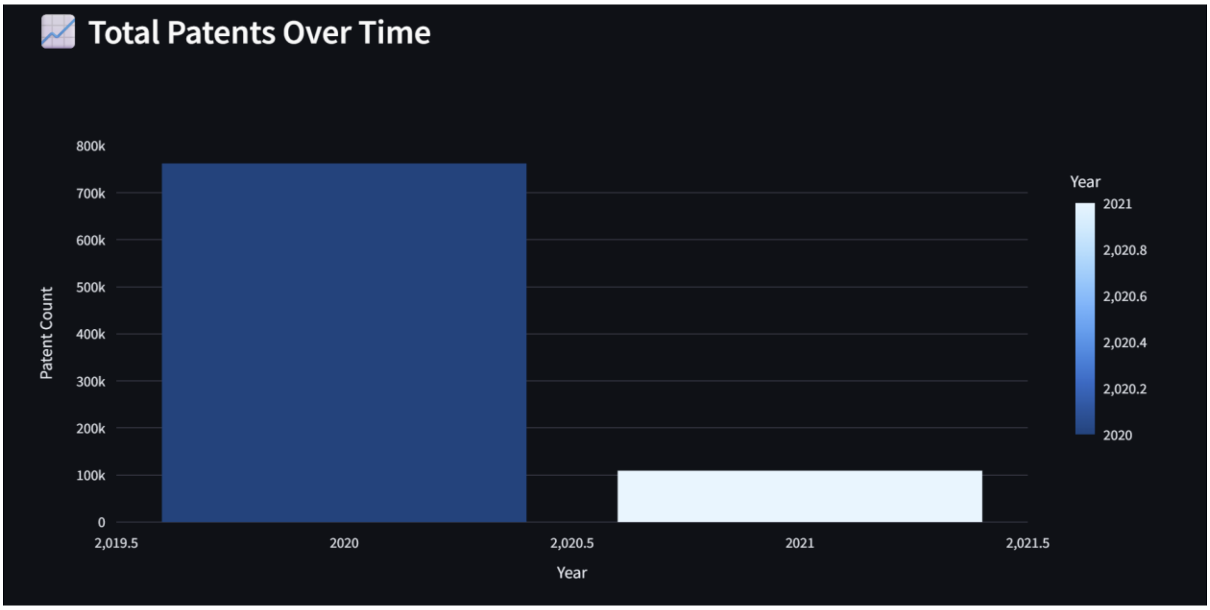
All visualisations are dynamically responsive interacting with any filter automatically updates all charts in real time, enabling linked data exploration.

- **Key Charts and Interactions**

✓ *Total Patents Over Time (Bar Chart)*

Displays yearly trends in total patent filings.

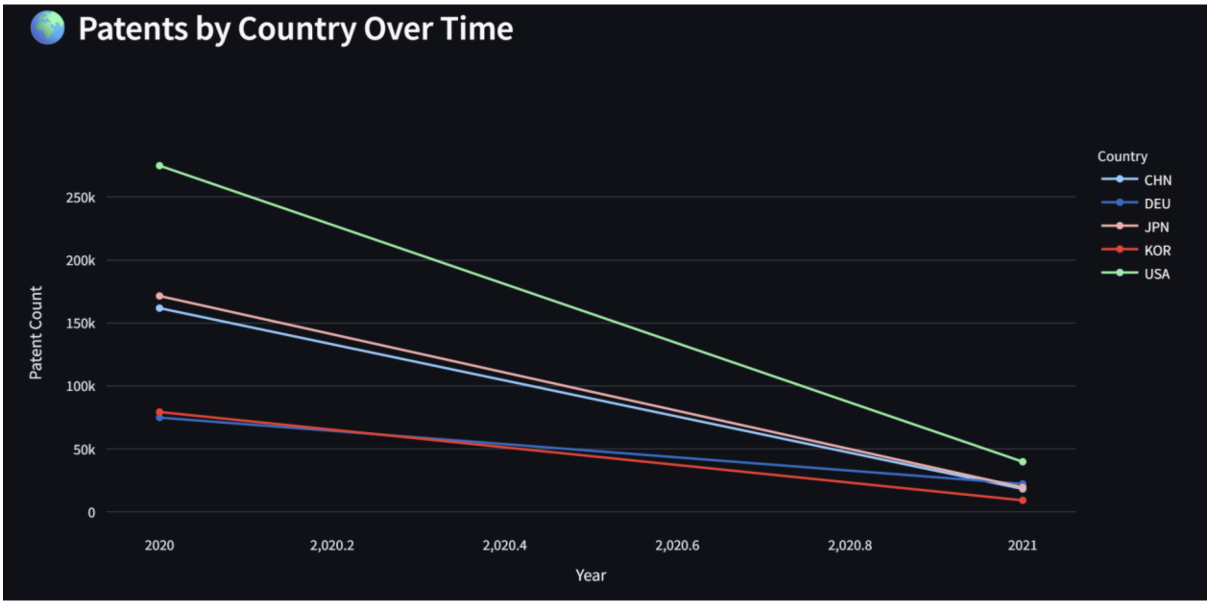
Helps identify periods of growth or decline in innovation activity.



✓ *Patents by Country Over Time (Line Chart)*

Enables direct comparison of countries' innovation trajectories.

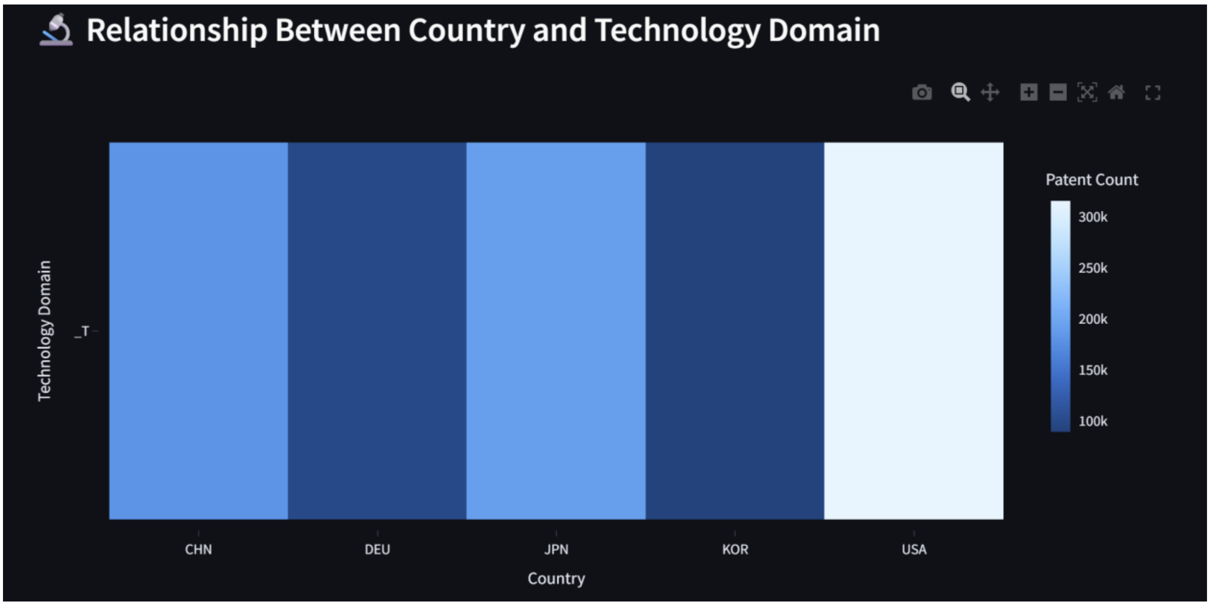
Useful for tracking performance of major economies.



✓ *Heatmap: Country vs. Technology Domain*

Visualises country-level strengths across various technology sectors.

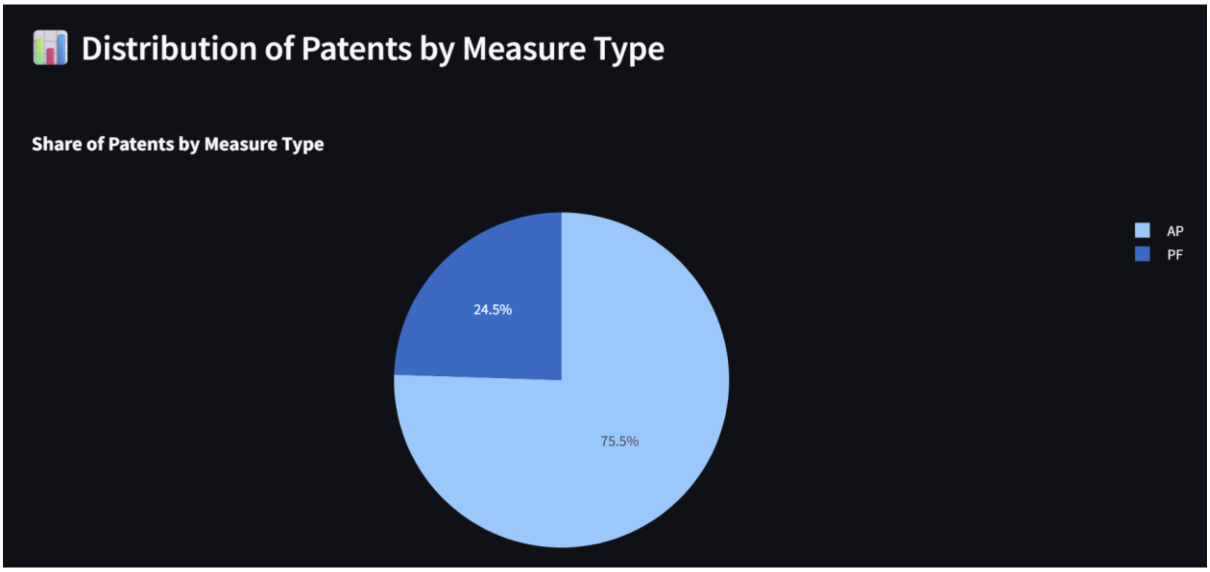
Color intensity makes outliers and dominant fields easily identifiable.



✓ *Patent Share by Measure Type (Pie Chart)*

Breaks down patent filings by measurement categories.

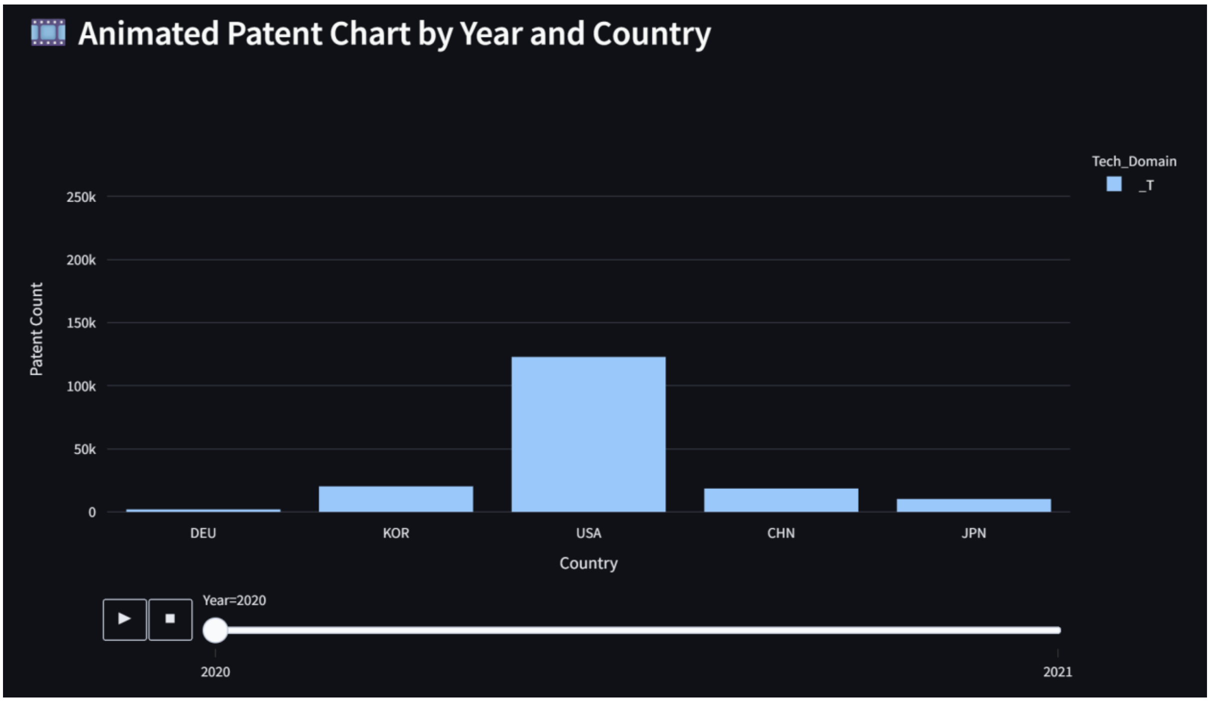
Provides a macro-level view of the types of innovation being tracked.



✓ *Animated Bar Chart (Over Time)*

An animated ranking of countries by patent volume over time.

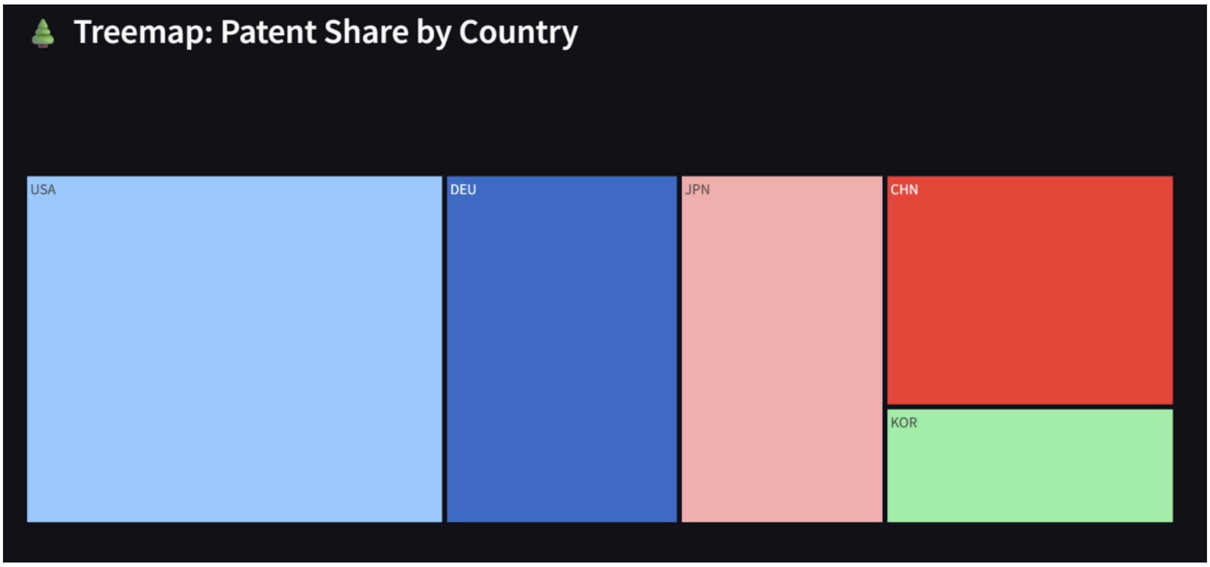
Presents trends in an engaging, time-aware format.



✓ *<u>Treemap by Country (Latest Year)</u>*

Displays each country's share of total patents for the most recent year.

Effective for showcasing current leaders in green innovation.



3. Design Justification

- **<u>User Experience and Visual Clarity</u>**

The dashboard was designed with a strong emphasis on user interaction, consistency, and responsive design to ensure a seamless analytical experience.

✓ *Interactivity:*

All visualisations respond dynamically to changes in sidebar filters, allowing users to explore and compare different subsets of the data. This enhances engagement and supports exploratory data analysis.

✓ *Consistency:*

A uniform color palette is applied across all charts to maintain country-specific color identity, aiding recognition across different visual contexts. The shared filter system ensures that all visualisations are synchronised, promoting a cohesive narrative across the dashboard.

✓ *Responsiveness:*

The dashboard is displayed in wide mode, enabling a spacious and uncluttered layout. Visual elements are set to auto-scale with the container, ensuring adaptability across various screen sizes and maintaining readability.

- **Rationale for Chart Selection**

Each chart type was intentionally chosen to match the nature of the data and the intended insight:

✓ *Line Charts:*

Used for time series to clearly illustrate longitudinal trends and enable comparison across years and countries.

✓ *Treemap and Pie Charts:*

Offer a proportional overview of data distributions, ideal for highlighting dominant contributors or categories at a glance.

✓ *Heatmap:*

Effectively represents relationships between two categorical variables (e.g., country vs. technology domain), making it easy to spot concentration areas and outliers.

✓ *Animated Bar Chart:*

Introduces a temporal storytelling element by showing changes in patent volumes across countries over time, making trends more engaging and intuitive.

4. Final Dashboard

https://cos30045-japuptm623odmkg4khnpr4.streamlit.app/

5. Code Organisation

The project follows a clear and modular code structure to enhance readability, reproducibility, and ease of maintenance. The components are as follows:

- **preprocess_patent_data.py:** This script is responsible for cleaning the raw OECD dataset and exporting a well-structured CSV file for use in the dashboard.
- **dashboard_final.py**: This is the main application script. It loads the processed data, defines sidebar filters, and generates interactive visualisations using Streamlit and Plotly Express.
- **cleaned_oecd_patent_data.csv**: The final cleaned dataset used as the input source for all dashboard visualisations.

All scripts are **thoroughly commented**, with explanations provided for each key processing step, ensuring that the code is both transparent and easily reproducible by others.

6. Conclusion

The completed interactive dashboard provides a robust and intuitive platform for exploring OECD environmental patent data. By integrating multi-dimensional filters and interactive visualisations, the tool enables users to:

- Identify leading countries and dominant technology domains
- Track changes and trends in innovation over time
- Assess the relative contributions of different innovation categories

In addition to functional performance, the dashboard design adheres to core principles of effective data visualisation—clarity, user engagement, and accessibility. The final product successfully meets the advanced requirements of the assignment and serves as a valuable tool for policymakers, researchers, and stakeholders interested in global green innovation trends.

References

1. OECD. (2015). *Green growth indicators*. OECD Publishing. http://www.oecd.org/greengrowth/greengrowthindicators
2. Nordregio. (2012). *Green patents in the Nordic region*. https://nordregio.org/maps/green-patents
3. Liu, Y., Liu, C., Zhang, N., & Dong, F. (2021). Measurements and trends in technological eco-innovation. *Sustainability, 13*(7), 3939. https://doi.org/10.3390/su13073939
4. Munzner, T. (2014). *Visualization analysis and design*. CRC Press.
5. Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Morgan Kaufmann.