10499
3201

COS30045 – Data Visualisation

Data Visualisation Report: Environmental Technology Patents Using OECD Statistics

VISUALISATION ASSIGNMENT

MAI HOANG DAI VY | 104993201

May - 2025

Executive Summary

This report provides a detailed investigation into the use of OECD environmental statistics, with a specific emphasis on patent data related to innovations in environmental technologies. The primary focus is on how such data can be effectively visualised to reveal meaningful insights and patterns. By analysing a selection of visualisation types, including bar charts, tree map, and column chart, the report seeks to evaluate how well these graphical representations communicate key trends over time and across different countries. Each visualisation is carefully critiqued based on established data visualisation principles, such as clarity, accuracy, simplicity, and audience accessibility. These principles are used to assess the overall effectiveness of each visualisation, identify its strengths and shortcomings, and propose actionable improvements.

The dataset used for this analysis is sourced from the OECD Data Explorer, with a specific focus on the number of patents filed in various environmental technology domains. The report examines contributions from both OECD and non-OECD economies, covering a time span from 2015 to 2022. The purpose of this temporal and regional comparison is to understand how different economies are progressing in terms of environmentally oriented innovation and where global leadership in this space may be emerging or shifting.

Introduction

In the context of accelerating climate change and the global push toward carbon neutrality, innovation plays a pivotal role in shaping sustainable technological solutions. The report “World Corporate Top R&D Investors: Paving the Way for Climate Neutrality,” jointly published by the European Commission’s Joint Research Centre (JRC) and the OECD, provides a comprehensive analysis of the world’s top 2,000 corporate R&D investors and their contributions to climate-related innovation. Central to this analysis is the use of patent data, which serves as a proxy for inventive activity and technological leadership.

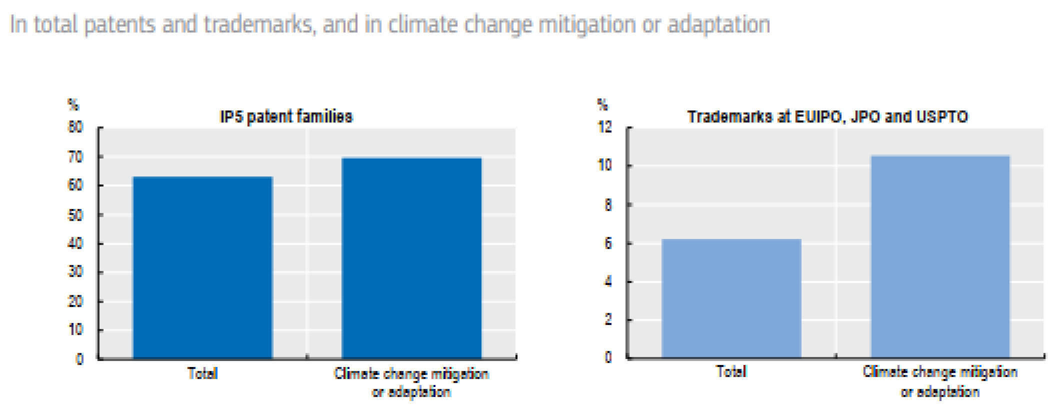
This document focuses on distinct visualisations from the report that highlight different dimensions of patent activity. These include a bar chart comparing patent ownership, a choropleth map showing the geographic distribution of inventors, and a time series column chart tracking sectoral patent trends. Together, these visualisations offer a multi-faceted view of how leading firms and regions are driving innovation in climate change mitigation and adaptation technologies.

This report uses three different visualisations to explore the OECD patent dataset and evaluate them using principles of good design such as clarity, accuracy, simplicity, and engagement. Each visual is critiqued in terms of its ability to communicate complex data effectively and fairly. The audience for this report includes policy-makers, researchers, educators, and students interested in innovation, environmental policy, and visual storytelling with data.

The overarching goal of the report is to illustrate how data visualisations can aid in understanding international trends in technological innovation, particularly in the context of environmental advancements. Effective visual communication of this data is critical for researchers, policymakers, and the general public to grasp the pace and scope of innovation efforts worldwide. In this context, visual tools are not merely aesthetic choices, but essential components of knowledge dissemination and evidence-based decision-making.

Part A: Visualisation Critique (Basic Requirement)

1. Visualisation - Bar Chart: Share of Patents and Trademarks Owned by Top R&D Investors



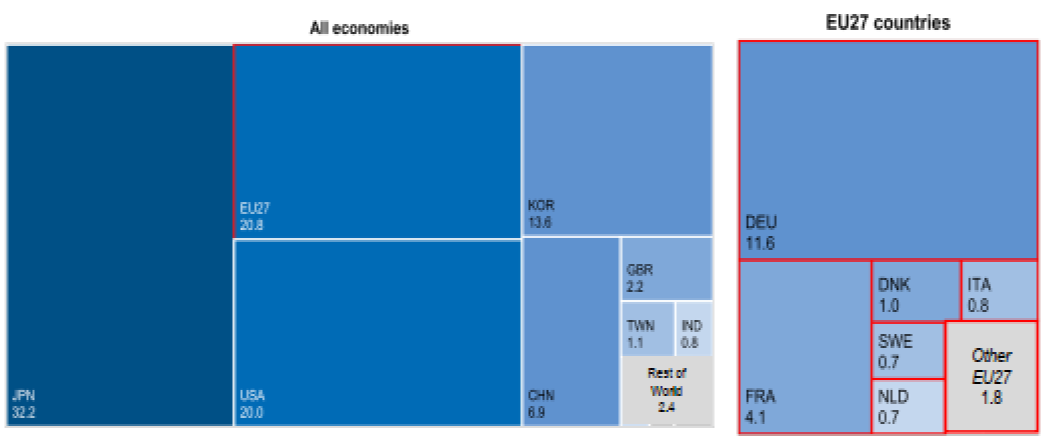
Source: JRC-OECD, COR&DIP© database v.3, 2021.

This bar chart presents a comparative overview of the share of global business-funded R&D, patents, and trademarks held by the top 2,000 R&D investors versus all other companies. The data reveals that these top firms account for 87% of global business R&D, 63% of global patents, but only 6% of global trademarks. This stark contrast highlights the concentration of innovation capacity in a relatively small number of firms, particularly in the domain of invention, while suggesting a more distributed landscape in the commercialisation phase.

The bar chart is a strong example of effective categorical comparison. It presents four bars, two for patents (total and climate-related) and two for trademarks (total and climate-related) to show the disproportionate contribution of the top 2,000 R&D investors to climate-related innovation. The use of a vertical bar chart is appropriate here, as it allows for easy comparison of discrete categories. The chart adheres to Tufte's integrity principles by starting the bars at zero, ensuring that the visual representation is proportional to the actual data values. This avoids the "lie factor," where graphical exaggeration misrepresents the underlying data (Tufte, 2001). The chart also avoids chartjunk, maintaining a high data-ink ratio by eliminating unnecessary visual elements such as 3D effects, excessive gridlines, or decorative shading.

However, while the chart is clean and accurate, it could be improved in several ways. First, the data density is relatively low, only four data points are shown. According to Tufte's principle of maximizing data density, the chart could be expanded to include more granular breakdowns, such as by region or sector, or by showing trends over time. Additionally, the chart lacks direct data labels on the bars, which forces the viewer to estimate values based on the y-axis. Including exact percentages would reduce cognitive load and improve accessibility, especially for non-expert audiences. From Munzner's perspective, the visual encoding (bar length) is appropriate for quantitative comparison, but the chart could benefit from layering, such as using color or annotations to emphasize the key insight that climate-related patents are more concentrated among top R&D investors than trademarks. Overall, the chart is effective but could be enhanced with more detail and visual cues to support interpretation.

2. Visualisation - Tree Map: Location of Inventors of Climate-Related Patents



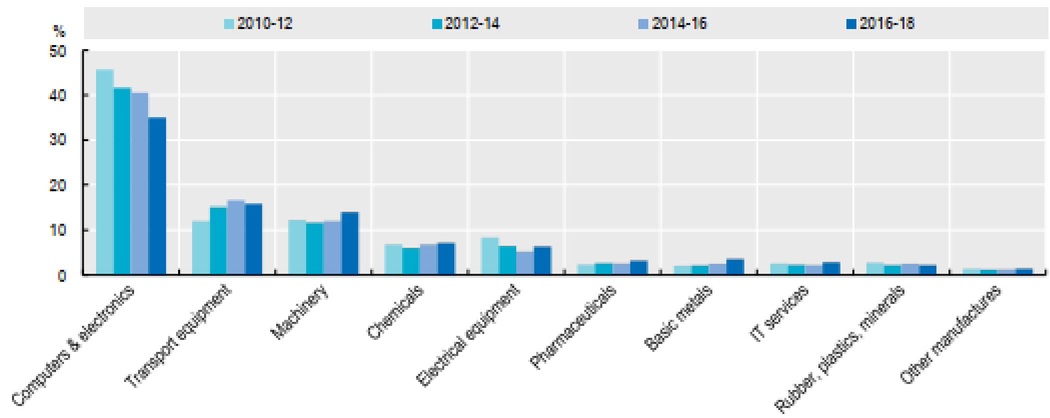
Source: JRC-OECD, COR&DIP© database v.3, 2021.

The tree map showing the geographic distribution of inventors of CCMA patents owned by the top 2,000 R&D investors. Japan leads with 32.2% of patents, followed by the EU27 with

20.8%, the United States with 20.0%, Korea with 13.6%, and China with 6.9%. Within the EU, Germany alone accounts for over half of the patents invented. This distribution highlights the regional concentration of climate innovation and the global leadership of Asian economies, particularly Japan and Korea. Japan's dominance is driven by its automotive and electronics sectors, which are heavily engaged in electric vehicle and battery technologies. The data also shows that Europe's contribution is broad but uneven, with Germany as the primary driver. Meanwhile, China's relatively low share in CCMA patents, despite its growing innovation capacity, suggests a gap between general innovation and climate-specific R&D.

Despite its strengths, the map has several limitations that affect its interpretability and effectiveness. One major issue is the lack of a detailed legend or numeric scale, which makes it difficult to quantify the differences between countries. This violates Tufte's principle of clear and thorough labeling, which is essential to prevent ambiguity and misinterpretation. Additionally, the map suffers from low data density: large countries with relatively few patents (e.g., Russia or Brazil) occupy significant visual space, potentially misleading viewers about their importance. This is a common issue with choropleth maps, where geographic area does not always correlate with data magnitude. To address this, the map could be supplemented with a ranked bar chart or numeric labels to provide more precise comparisons. Furthermore, the use of a single color gradient may obscure mid-range differences, which could be resolved by using a segmented or diverging color scale. From a design perspective, the map could also benefit from interactivity (in a digital format), allowing users to hover over countries to see exact values. Overall, while the map is visually engaging and informative at a glance, it lacks the precision and clarity needed for detailed analysis.

3. Visualisation – Column Chart: Patents by Sector Over Time



Source: JRC-OECD, COR&DIP© database v.3, 2021.

This is a time-series column chart that tracks the distribution of patent activity across sectors from 2010 to 2018 among the world's top R&D investors. The "Computers & Electronics" sector consistently leads in patent volume, contributing around 35% in 2016–18, although its share has declined slightly over time. "Transport Equipment" and "Machinery" remain strong, with the top three sectors accounting for over 60% of all patents. This reflects the technological intensity and innovation focus of these sectors and their central role in global innovation. The slight decline in "Computers & Electronics" may indicate a shift toward software-based innovation, which is less frequently patented. The chart also highlights the persistent concentration of innovation in a few sectors, raising concerns about the diversification of technological capabilities needed for a broad-based green transition. Notably, sectors like energy, construction, and agriculture, which are critical for climate mitigation, remain underrepresented in patent output.

However, the chart could be significantly improved in terms of clarity and accessibility. One issue is the color palette: the colors used to differentiate sectors are not sufficiently distinct, making it difficult to distinguish between them, especially for viewers with color vision deficiencies. This violates the principle of effective layering and separation, which is essential for multi-category visualisations (Tufte, 2001). Additionally, the chart lacks annotations or callouts to highlight key insights, such as the decline in the "Computers & electronics" sector or the rise of "Transport equipment." Including such annotations would guide the viewer's attention and support storytelling. The chart also does not include a legend directly on the chart, requiring viewers to refer to external documentation to interpret the color coding. From a data density perspective, the chart performs moderately well, but it could be enhanced by integrating trend lines or interactive filters in a digital format. Finally, the chart could benefit from a brief narrative or summary box that explains the significance of the observed trends, especially for non-expert audiences. In summary, while the chart is structurally sound and informative, it requires better visual differentiation and contextual support to maximize its effectiveness.

Conclusions

The data reveals that the world's top R&D investors dominate climate-related innovation, holding 70% of CCMA patents but only a small share of trademarks. This highlights their strength in invention but limited role in commercialisation. Geographically, innovation is concentrated in Japan, the EU, and the U.S., with Japan leading due to its industrial focus on electronics and transport. Sectorally, most patents come from high-tech industries, while critical sectors like energy and agriculture remain underrepresented. To accelerate the green transition, broader participation across regions and industries is essential.

Visualisations based on OECD's environmental patent data are valuable tools for understanding international efforts in green technology development. This report demonstrated how three types of visuals bar chart, treemap and column chart can be used to interpret trends and make comparisons. However, improvements are necessary to make them more informative, engaging, and accurate. The application of visual design principles significantly enhances the interpretability of data. As the urgency of climate change grows, clear communication of environmental innovation trends through improved visuals will support informed decision-making and public awareness.

References

Few, S. (2009). *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. https://doi.org/10.1080/10543401003641225

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Cheshire, CT: Graphics Press. https://doi.org/10.1515/9783050093833-035

Pilat, D. (2021, November). *World corporate top R&D investors: Paving the way for climate neutrality* [Conference presentation]. 8th European Conference on Corporate R&D and Innovation (CONCORDi 2021), Seville, Spain. European Commission, Joint Research Centre. https://iri.jrc.ec.europa.eu/sites/default/files/inline-files/Pilat_CONCORDi2021.pdf

Amoroso, S., Aristodemou, L., Criscuolo, C., Dechezleprêtre, A., Dernis, H., Grassano, N., … & Tübke, A. (2021). *World Corporate Top R&D Investors: Paving the Way for Climate

Neutrality (EUR 30884 EN). Publications Office of the European Union. https://doi.org/10.2760/49552

Amoroso et al. (2021). *World Corporate Top R&D Investors: Paving the Way for Climate Neutrality*. https://doi.org/10.2760/49552

Part B: Interactive Dashboard Implementation (Advanced Requirement)

Data preprocessing step by step:

```
Chart.py > ...
# Please check the OECD_Dashboard_Instrucitons.txt for instructions on how to run this code.
# This code creates an interactive dashboard using Plotly Dash to visualize OECD patent data.
import pandas as pd
import numpy as np
import dash
from dash import dcc, html, Input, Output
import plotly.express as px
from plotly.subplots import make_subplots
import plotly.graph_objects as go


# Data Preprocessing and Cleaning for OECD Patent Dataset
# ---------------------------------------------------
# 1. Load the CSV data
# 2. Handle missing values
# 3. Standardise formats (e.g., country names, years)
# 4. Transform data for dashboard use (pivot, aggregation)
# 5. Save cleaned data for dashboard use

# Load the data
df = pd.read_csv('OECD.STI.PIE,DSD_PATENTS@DF_PATENTS,1.0+.A...PRIORITY...INVENTOR..._T.csv', skiprows=0)

# Select relevant columns for analysis
columns_of_interest = [
    'REF_AREA', 'Reference area', 'TIME_PERIOD', 'Time period',
    'OBS_VALUE', 'Observation value', 'UNIT_MEASURE', 'Unit of measure',
    'OECD_TECHNOLOGY_PATENT', 'Selected OECD technology domains',
    'PATENT_AUTHORITIES', 'Patent authorities'
]
# Some columns may be missing in some rows, so use .get for safety
df = df[[col for col in columns_of_interest if col in df.columns]]

# Rename columns for easier access
df = df.rename(columns={
    'REF_AREA': 'Country_Code',
    'Reference area': 'Country',
    'TIME_PERIOD': 'Year',
    'Time period': 'Year2',
    'OBS_VALUE': 'Value',
Ln 19, Col 1   Spaces: 4   UTF-8   CRLF   {} Python   3.11.5   Go Live
```

1. Load Data: Reads a CSV file containing OECD patent data.
2. Select & Rename Columns: Keeps relevant columns and renames them for clarity.

Chart.py > ...

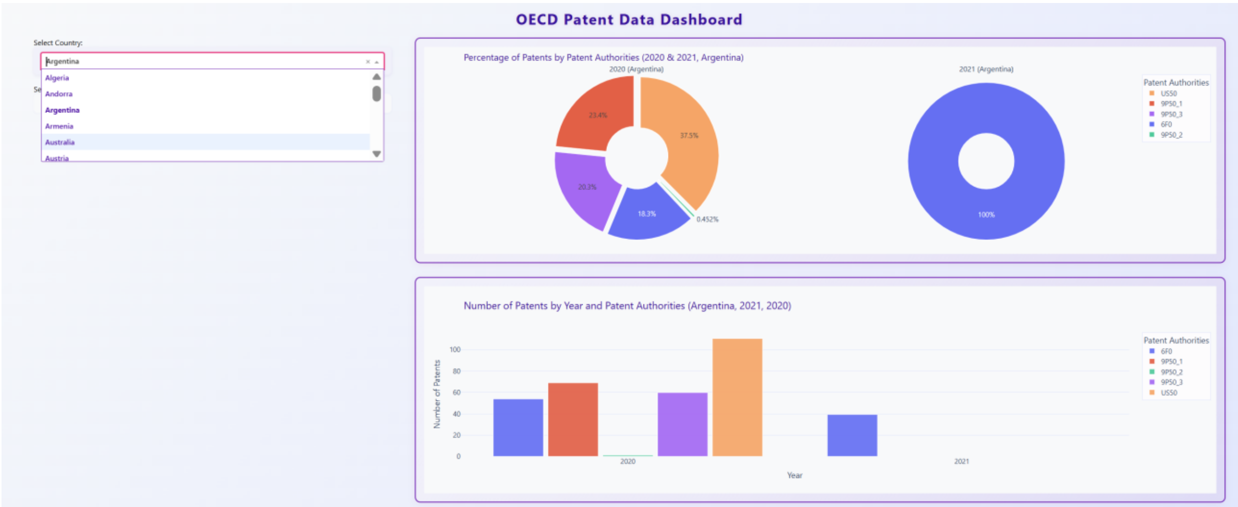
```python
# Rename columns for easier access
df = df.rename(columns={
    'REF_AREA': 'Country_Code',
    'Reference area': 'Country',
    'TIME_PERIOD': 'Year',
    'Time period': 'Year2',
    'OBS_VALUE': 'Value',
    'Observation value': 'Value2',
    'UNIT_MEASURE': 'Unit',
    'Unit of measure': 'Unit2',
    'OECD_TECHNOLOGY_PATENT': 'Tech_Domain',
    'Selected OECD technology domains': 'Tech_Domain2',
    'PATENT_AUTHORITIES': 'Patent_Office',
    'Patent authorities': 'Patent_Office2'
})

# Prefer non-null values between duplicate columns
df['Country'] = df['Country'].combine_first(df['Country_Code'])
df['Year'] = df['Year'].combine_first(df['Year2'])
df['Value'] = df['Value'].combine_first(df['Value2'])
df['Unit'] = df['Unit'].combine_first(df['Unit2'])
df['Tech_Domain'] = df['Tech_Domain'].combine_first(df['Tech_Domain2'])
df['Patent_Office'] = df['Patent_Office'].combine_first(df['Patent_Office2'])

# Drop duplicate columns
df = df.drop(columns=['Country_Code', 'Year2', 'Value2', 'Unit2', 'Tech_Domain2', 'Patent_Office2'], errors='ignore')

# Handle missing values
# Convert Value to numeric, coerce errors to NaN
df['Value'] = pd.to_numeric(df['Value'], errors='coerce')
# Drop rows with missing country, year, or value
df = df.dropna(subset=['Country', 'Year', 'Value'])

# Standardise year to integer
df['Year'] = df['Year'].astype(int)

# Standardise country names (strip whitespace)
df['Country'] = df['Country'].str.strip()
```

3. Handle Duplicates: Merges similar columns (e.g., 'Time period' and 'TIME_PERIOD') using non-null values.
4. Clean Data:
    - Converts values to numeric.
    - Drops rows with missing essential data (Country, Year, or Value).
    - Ensures Year is integer and within 2020–2021.
5. Standardize Formats: Strips whitespace from country names.

```python
# Standardise country names (strip whitespace)
df['Country'] = df['Country'].str.strip()

# Only keep data for years 2020 and 2021 (no decimals)
df = df[df['Year'].isin([2020, 2021])]
df = df[df['Year'] == df['Year'].astype(int)]
df['Year'] = df['Year'].astype(int)

# Optional: Aggregate by country and year (sum values)
df_agg = df.groupby(['Country', 'Year', 'Patent_Office', 'Tech_Domain', 'Unit'], as_index=False)['Value'].sum()

# Save cleaned data for dashboard use
df_agg.to_csv('cleaned_patent_data.csv', index=False)
```

6. Aggregate: Groups data by Country, Year, Patent_Office, Tech_Domain, and Unit to sum Value.
7. Save Cleaned Data: Outputs the cleaned, aggregated dataset to a CSV for dashboard use.

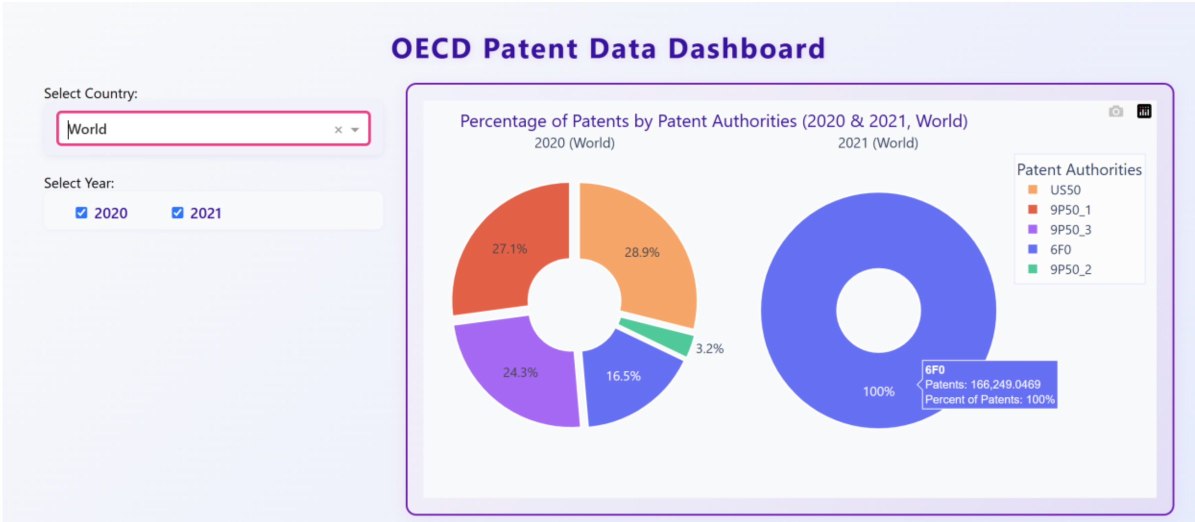
Overview of OECD Patent Data Dashboard



Design choice for the Dashboard (please make sure that you do the instruction file .txt in the .zip file to open this dashboard)

1. Dropdown Selector: Implemented a country selection dropdown to filter visualizations interactively, allowing users to focus on individual countries or global (World) data.
2. Color Coding: Each patent office is assigned a consistent color across charts to maintain visual continuity and simplify comparison.
3. Donut Charts: Used for percentage distribution of patents by office in 2020 and 2021. This format makes relative contributions easy to grasp at a glance.
4. Bar Chart: Plots the total number of patents by year and office to provide temporal trends and volume comparisons.
5. Minimalist, Modern Style: Soft background and pastel tones ensure data readability. The use of rounded chart containers and shadow effects adds to the professional and user-friendly look.
6. Tooltips: Tooltips on hover (as seen in the pie chart) offer precise values and percentages for detailed insight without cluttering the view.

Analyzing the pie charts:



Pie charts in OECD Patent Data Dashboard

Explain the legend:

US50: United States Patent and Trademark Office

6F0: European Patent Office

9P50_1: World Intellectual Property Organization

9P50_2: Triadic patent family

9P50_3: IP5 patent family

The OECD Patent Data Dashboard reveals a significant shift in the distribution of patents by patent authority between 2020 and 2021 at the global level. In 2020, patent filings were relatively balanced among several authorities: 9P50_1 (28.9%), US50 (27.1%), 9P50_3 (24.3%), and 6F0 (16.5%), with 9P50_2 contributing a smaller portion (3.2%). However, in 2021, all patents are attributed exclusively to the 6F0 patent authority, accounting for 100% of the filings. This dramatic change suggests a potential data reporting issue, reclassification of patent authorities, or an anomaly in how the 2021 data is displayed or collected. A deeper investigation into the raw data is necessary to understand whether this shift reflects actual changes in global patent activity or if it is a result of dashboard configuration or data processing errors.

In conclusion, the dashboard highlights a notable change in patent distribution from 2020 to 2021, with 6F0 emerging as the sole contributor in 2021. This unusual shift calls for further investigation to determine whether it reflects actual patent activity or potential data inconsistencies.