SWINBURNE UNIVERSITY OF TECHNOLOGY
HO CHI MINH CAMPUS

COS30045 – Data Visualization

Process Book

Agri-environmental Indicators Dashboard

KURMOMI GROUP

Team Name: KUROMI

Team Members:
Khai Vinh Han – 104222222 (SWS00291)
Bui Thanh Ngan Pham – 104217354 (SWS00522)
Mai Ngoc Anh Doan – 104220776 (SWS00350)

Website Link: **https://kuromi-j2wg.onrender.com/**
When you want to run, **PLEASE CONTACT US**, so we can make sure it works
Best experienced on Desktop

GitHub Repo (to run locally if you want): **https://github.com/rayraurray/KUROMI**

Lecturer/Instructor: Dr. Tristan Nguyen
Due Date: August 6th, 2025

Table of Contents

1. Introduction

1.1. Background and Motivation

Agriculture sits at the intersection of economic productivity and environmental stewardship. As farming systems intensify to meet global food demand, pressures on soil, water, biodiversity and climate escalate.

Policymakers, industry stakeholders and researchers alike require a clear, unified perspective on how agricultural practices translate into environmental outcomes, both to recognise emerging challenges and to calibrate effective interventions.

Through the distillation of decades of agri-environmental data into an intuitive visual format, we empower decision-makers to appreciate long-term trends, benchmark performance across regions, and anticipate areas where policy adjustments will be most impactful.

1.2. Primary Audience

Agricultural and environmental policy analysts, researchers, and government decision-makers will be the primary audience for this visualization. In particular:

- OECD and EU policy teams monitoring progress toward environmental targets (e.g., nutrient surpluses, greenhouse-gas emissions).
- National ministries of agriculture and environment seeking to benchmark their performance against peer countries and identify “hot spots.”
- Academics and consultants exploring long-term trends in agri-environmental pressures and outcomes.

1.3. Purpose of Dashboard

The dashboard will serve as an analytical lens through which users can survey the evolving environmental footprint of agriculture. Rather than posing narrowly defined queries, it will offer an open-ended platform for exploration. These users will want to:

- Track time-series of key indicators (e.g., pesticide use, water abstractions) across 1985 until the latest time frame.
- Compare across countries and country-groups (OECD, EU, emerging economies).
- Identify emerging challenges and assess where policy responses may be needed.
- *etc.*

Presenting these data visually is important because it transforms large, multidimensional tables into intuitive charts and maps that support faster insights, better communication with stakeholders, and more evidence-based policy decisions.

Additionally, the dashboard will enable users to answer questions such as:

- Which countries have experienced the greatest reduction in fertilizer input per hectare since 2000?
- How have greenhouse-gas emissions from agriculture evolved in emerging economies compared to OECD members?
- Where are the current "hot spots" for nutrient surpluses or irrigation intensity?
- Are current trends sufficient to meet established environmental targets by 2030?
- *etc.*

2. Project Code

2.1. Data Processing

Our analysis begins by sourcing the complete Agri-Environmental Indicators dataset directly from the OECD Data Explorer, which provides a comprehensive, annual time series for each indicator across all reporting countries. The OECD's agri-environmental indicators (AEIs) database delivers the most up-to-date and all-encompassing collection of metrics for 38 OECD member nations, five non-OECD EU countries, and 11 emerging economies.

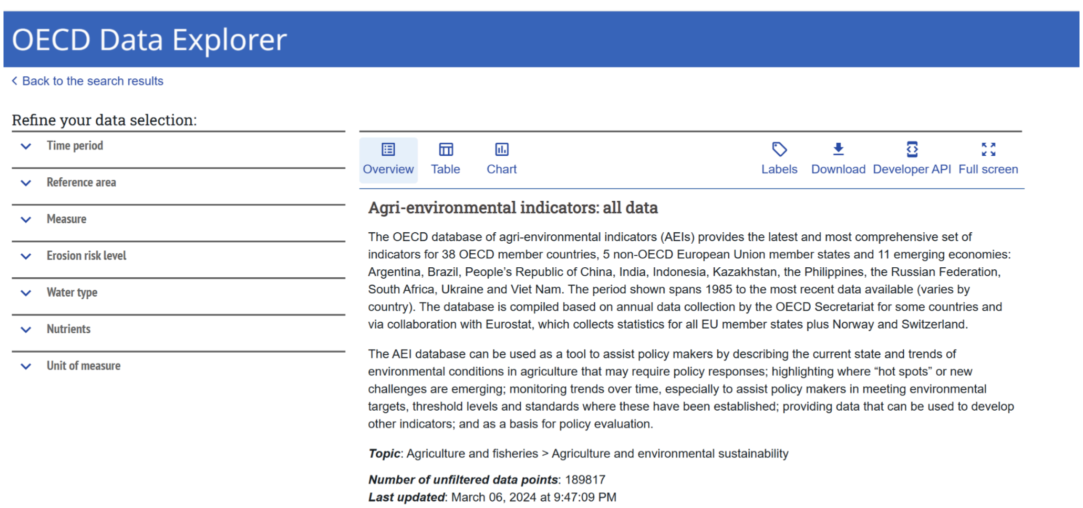


Image of OECD Data Explorer

The raw download arrives as a wide-form CSV table, comprising thirty columns - many of which duplicate descriptive labels or carry metadata footnotes that, while useful for documentation, do not contribute to the core visual narrative.

```
... <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 189817 entries, 0 to 189816
    Data columns (total 30 columns):
     #   Column                    Non-Null Count   Dtype
    ---  ------                    --------------   -----
     0   STRUCTURE                 189817 non-null  object
     1   STRUCTURE_ID              189817 non-null  object
     2   STRUCTURE_NAME            189817 non-null  object
     3   ACTION                    189817 non-null  object
     4   REF_AREA                  189817 non-null  object
     5   Reference area            189817 non-null  object
     6   FREQ                      189817 non-null  object
     7   Frequency of observation  189817 non-null  object
     8   MEASURE                   189817 non-null  object
     9   Measure                   189817 non-null  object
     10  EROSION_LEVEL             189817 non-null  object
     11  Erosion risk level        189817 non-null  object
     12  WATER_TYPE                189817 non-null  object
     13  Water type                189817 non-null  object
     14  NUTRIENTS                 189817 non-null  object
     15  Nutrients                 189817 non-null  object
     16  UNIT_MEASURE              189817 non-null  object
     17  Unit of measure           189817 non-null  object
     18  TIME_PERIOD               189817 non-null  int64
     19  Time period               0 non-null       float64
    ...
     28  BASE_PER                  1305 non-null    float64
     29  Base period               0 non-null       float64
    dtypes: float64(5), int64(3), object(22)
```

Raw Dataset Structure

To prepare the data for visualization, we first distilled the table down to thirteen essential fields: country identifiers, calendar year, indicator codes and names, measured values, and units.

```python
rename_map = {
    'REF_AREA': 'country_code',
    'Reference area': 'country',
    'Observation status': 'observation_status',
    'OBS_STATUS': 'obs_status',
    'MEASURE': 'measure_code',
    'Measure': 'measure_category',
    'Unit of measure': 'measure_unit',
    'Erosion risk level': 'erosion_risk_level',
    'Water type': 'water_type',
    'Nutrients': 'nutrients',
    'TIME_PERIOD': 'year',
    'OBS_VALUE': 'obs_value',
    'Unit multiplier': 'unit_multiplier',
}
```
✓ 0.0s

```python
rename_map = {k:v for k,v in rename_map.items() if k in df_clean.columns}
df_clean = df_clean.rename(columns=rename_map)
```
✓ 0.0s

Code for removing unused columns, and keeping useful ones

Additionally, removing extraneous label fields and footnote columns, we create a lean, tidy dataset that aligns one row per country-year-indicator. Consistency checks then harmonize country names (e.g. normalizing “China (People’s Republic of)” to “China” and collapsing multiple European Union variants into a single “EU” category), while parsing of numeric types ensures that year and observation values can be manipulated without error.

```
def clean_country(name):
    name = name.strip()

    # Manual fixes
    replacements = {
        "China (People’s Republic of)": "China",
        "China (Peopleâ€™s Republic of)": "China",
        "Macau (China)": "Macau",
        "Hong Kong (China)": "Hong Kong",
        "Democratic Peopleâ€™s Republic of Korea": "North Korea",
        "TÃ¼rkiye": "Turkey",
        "CÃ´te dâ€™Ivoire": "Côte d'Ivoire",
        "RÃ©union": "Reunion",
        "European Union (27 countries)": "EU",
        "European Union (28 countries)": "EU",
        "Euro area (19 countries)": "EU",
        "European Union (27 countries from 01/02/2020)": "EU",
        "European Union": "EU",
    }

    # Replace corrupted or long forms
    for key, value in replacements.items():
        if name == key:
            return value

    return name
✓ 0.0s
```

Code for renaming country names

Missing values are addressed through a two-stage process: explicit filtering of any rows lacking a valid observation.

Data types for `obs_value` and `year` are returned to a more suitable format for analysis. NA values for the `obs_value` are dropped.

```
df_clean["year"] = pd.to_numeric(df_clean["year"], errors="coerce").astype('Int64')
df_clean["obs_value"] = pd.to_numeric(df_clean["obs_value"], errors="coerce").astype('Float64')
✓ 0.0s
```

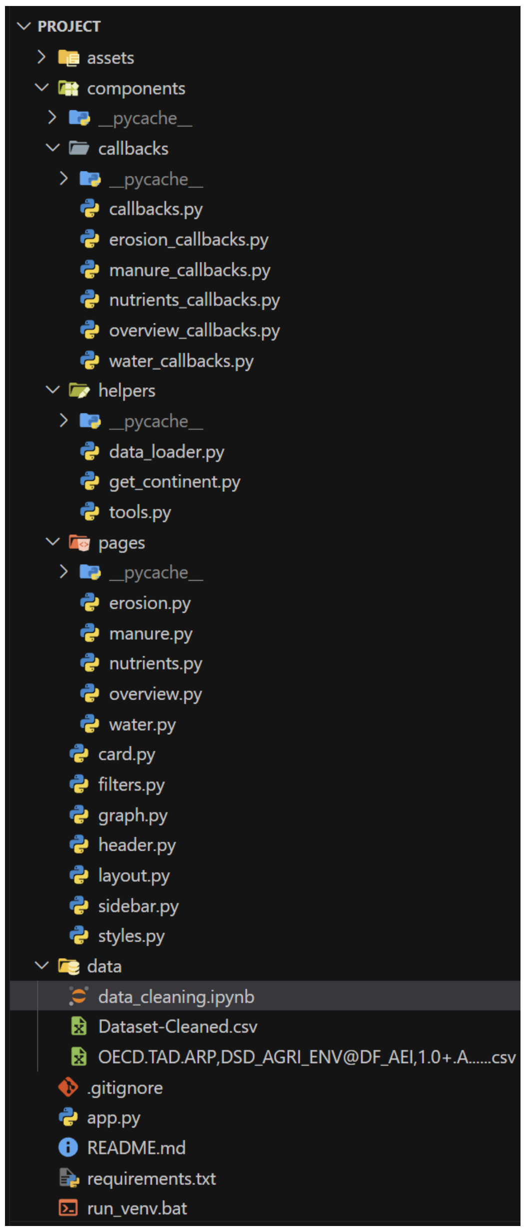
```
df_clean = df_clean.dropna(subset=["obs_value", "year"])
✓ 0.0s
```

Code for cleaning observation values and year columns

The cleaned and enriched dataset is finally serialized to an efficient Parquet format, enabling rapid loading into our dashboard framework.

This end-to-end pipeline not only streamlines data maintenance as new annual releases become available but also lays a transparent foundation for every chart and table, ensuring that our visualizations faithfully reflect the underlying agri-environmental dynamics.

2.2. Project Structure



Project Folder Structure

The dashboard application is built in Python using the Dash framework and follows a modular, folder-based organization that promotes maintainability, reusability, and clear separation of concerns.

The `assets/` folder contains static assets that Dash automatically serves at app startup:

- **Background GIFs** and other images used for headers or page decorations.
- **Custom CSS file**, which complements Dash's built-in styles and ensures consistent theming across all components.

The `components/` folder is an all reusable building blocks for the application live here, organized into three subfolders:

- The `components/callbacks/` directory holds Python files that define the callback functions that power interactivity for a specific section of the dashboard.
- This `components/helpers/` folder hosts pure-function utilities that perform shared tasks, such as:
    - `data_loader.py`: Abstracts reading and caching of CSV or JSON data sources.
    - `get_continent.py`: Maps country codes to continent names for grouping.
    - `tools.py`: Miscellaneous helper functions used across callbacks and page code (e.g., date parsers, legend formatting).
- Each page of the Dash application corresponds to a standalone Python script in the `components/pages/` directory . These scripts define:
    - **Page layouts**: Arrangement of graphs, filters, and descriptive text.
    - **Initial data fetch**: Which helper functions to call on page load.
    - **Linkage to callbacks**: Which inputs/outputs are connected via the `@app.callback` decorators.

Everything related to data ingestion and preparation lives in the `data/` here:

- `data_cleaning.ipynb`: A Jupyter notebook showing the end-to-end cleaning workflow.
- **The original dataset file** directly downloaded from source.
- `Dataset-Cleaned.csv`: The finalized, cleaned dataset used by the Dash app.

Finally, files in the root directory define key functions, including:

- `app.py`: The main Dash application file. It initializes the Dash server, sets global configurations (e.g., external stylesheets), and registers the page modules.
- `README.md`: High-level project overview, setup instructions, and usage examples.
- `requirements.txt`: Pinpointed Python dependencies needed to run the application in a virtual environment.
- `run_venv.bat`: Windows batch scripts for launching the app with or without activating a virtual environment.

This structured approach ensures that each part of the dashboard, from data loading through to interactive visualization, can be developed, tested, and maintained independently, yet integrated seamlessly into the overall application.

2.3. Project Frameworks and Libraries

2.3.1. D3

D3.js (Data-Driven Documents) is an open-source JavaScript library for creating dynamic, interactive data visualizations in web browsers. It binds arbitrary data to the DOM and then applies data-driven transformations to the document, allowing fine-grained control over every visual element. Key strengths of D3.js include:

- **Low-level manipulation**: Direct DOM and SVG control enables bespoke visual designs - everything from custom axes and scales to animated transitions.
- **Rich ecosystem**: A mature community contributes reusable modules for layouts (e.g. force-directed graphs), geographic projections, and data parsing.
- **Performance**: Efficient update patterns (enter–update–exit) make it suitable for handling large datasets in real time.

In this project, D3.js is integrated as a complementary visualization layer alongside Plotly, implemented through Dash’s clientside callbacks to leverage client-side rendering capabilities. The implementation follows a hybrid architecture where D3 visualizations are embedded within specific dashboard pages (Overview, Nutrients, and Water pages) to provide enhanced interactivity beyond standard charting libraries.

The technical implementation utilizes a three-stage pattern: first, server-side Python callbacks process and normalize the agricultural data (including logarithmic normalization by agricultural land area), then pass structured JSON data to dcc.Store components. Second, clientside JavaScript callbacks dynamically load D3.js from CDN if not already present, ensuring the library is available without affecting initial page load performance. Finally, custom D3 rendering functions create visualizations with features such as animated bar transitions, hover tooltips with detailed country statistics, and color-coded risk levels.

2.3.2. Dash

Dash is a high-level Python framework built on top of Flask, React, and Plotly.js that streamlines development of interactive, analytical web applications with minimal JavaScript. Its core features include:

- **Declarative Components**: Pre-built UI elements (graphs, sliders, dropdowns, tables) that are configured in Python.
- **Reactive Callbacks**: A simple decorator pattern (`@app.callback`) wires user inputs directly to output updates, enabling fluid interactivity without manual AJAX or REST wiring.
- **Python-First Workflow**: Data manipulation, statistical analysis, and model integration can be performed entirely in Python - leveraging the PyData stack (pandas, NumPy, scikit-learn, etc.) - and then immediately visualized.
- **Theming & Layout**: Built-in support for external CSS and Bootstrap themes via the `assets/` folder, alongside flexible layout primitives (`html.Div`, `dcc.Graph`, etc.) for responsive design.

In our agri-environmental dashboard, Dash serves as the application backbone:

1. **Layout Definition** (in `components/pages/`): Each page's structure and styling are declared in Python.
2. **Data Binding** (via `components/helpers/`): Cleaned datasets are loaded and transformed on demand.
3. **Interactive Callbacks** (in `components/callbacks/`): User selections (e.g., country filters, year sliders) automatically trigger graph redraws without manual page reloads.

3. Visualisation Design

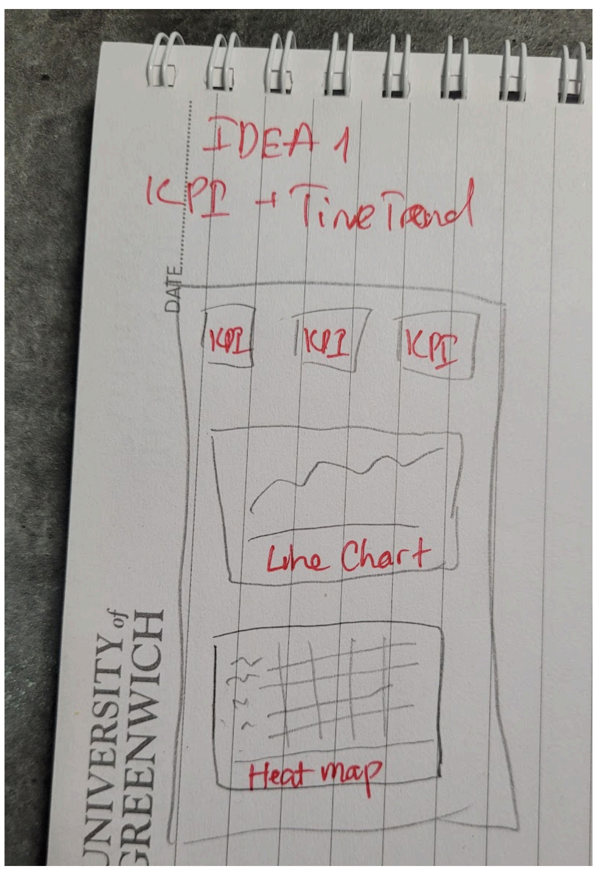
3.1. Overview Dashboard

3.1.1. Visualization Design (Proposal)

The proposal stage of visualization design focused on providing a comprehensive high-level overview of the whole dataset, but this stage also includes the availability of drilling down the detailed datasets for users. Below are ideas of layout that have been considered.

Idea 1: KPI + Time Trend Focus

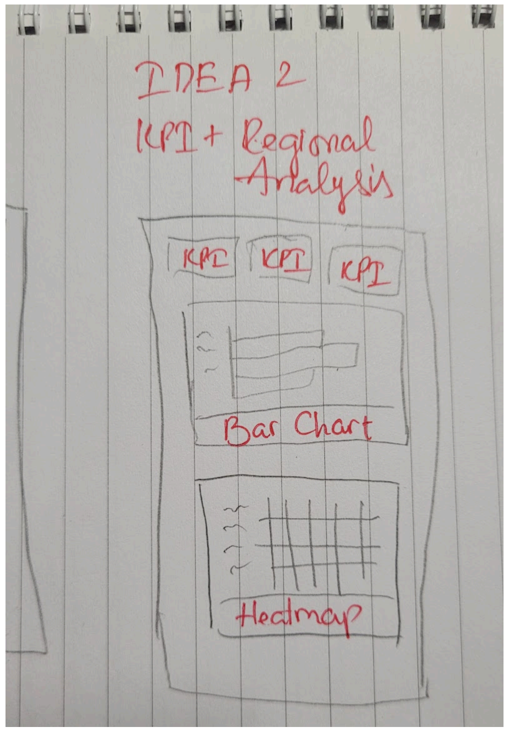
- **Structure:** Beginning with the KPI cards on top of the page, below are a line chart for nutrient trends and heatmap for country/year comparison.
- **Goal:** To emphasize the time trend.
- **Strength:** Highlight the long term trend of nutrient balance.



The draft layout for idea 1 of Overview page

Idea 2: KPI + Regional Analysis

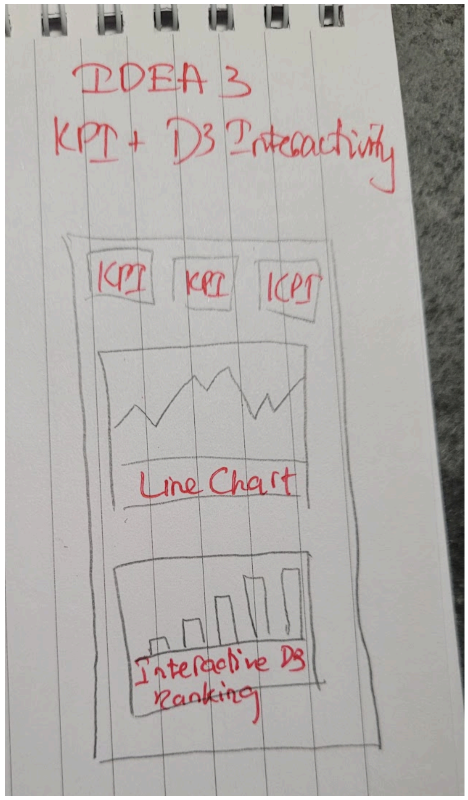
- **Structure:** Beginning with the KPI cards on top of the page, below are a horizontal bar chart (Top 10 countries) and a heatmap.
- **Goal:** To provide an overview of the dataset using country comparison, before applying the time range.
- **Strength:** Highlight the region specific data.



The draft layout for idea 2 of Overview page

Idea 3: KPI + D3 Interactivity

- **Structure:** Beginning with the KPI cards on top of the page, below are a line chart for nutrient trends and a D3 interactive chart for ranking the average nutrient balance.
- **Goal:** To combine the `dash.plotly` main framework with the D3 Javascript of this course.
- **Strength:** Highlight the variety of frameworks, also aligning with the assignment requirements of using D3 framework.



The draft layout for idea 3 of Overview page

3.1.2. Visualization Design (Progress)

In the progress stage of visualization design including the iterative design, which iterate continuously between wireframing and implementation:

Stage 1: Sketching

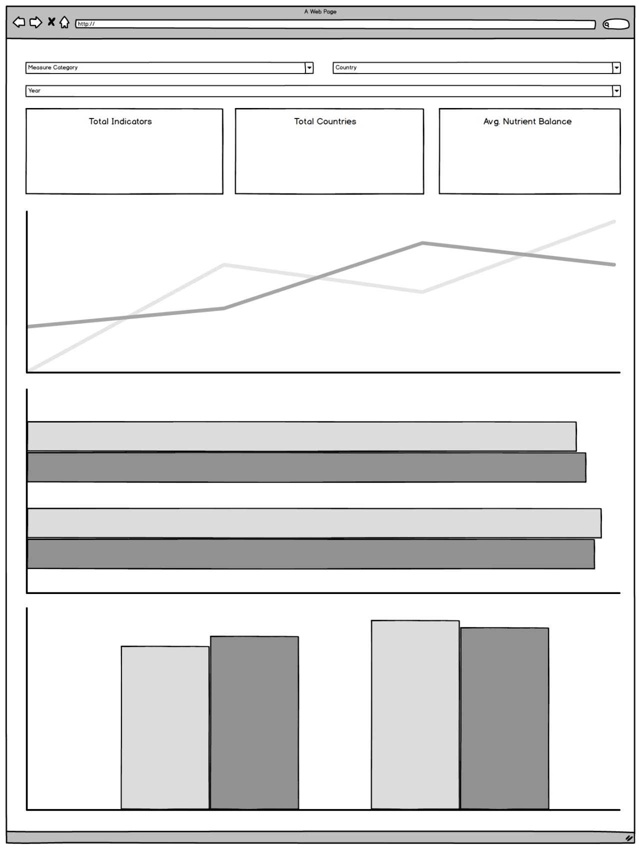
- Firstly, the filters are sketched on top of KPI cards, followed by the vertical stacked bar chart.
- Try changing between heatmap and bar chart to explore country-level views, and find the appropriate figure.

Stage 2: Wireframing

- Decide moved to the grid-based layout, which ensure the clean structure:
    - Top row: Filters
    - Second row: KPI cards
    - The rest: Below are trend chart, heatmap, horizontal bar chart, and D3 interactive chart.
- Under each chart, there will be a box for explanation on how to interact with the figure.

Stage 3: Pre-final Implementation

- Applying a dark theme to align with other pages, also symbolizes Kuromi - name of our team.
- Integrating Plotly for KPI cards, trend chart, heatmap, and bar chart.
- Finally, adding a D3 interactive chart for ranking countries based on the average nutrient balance intensity.



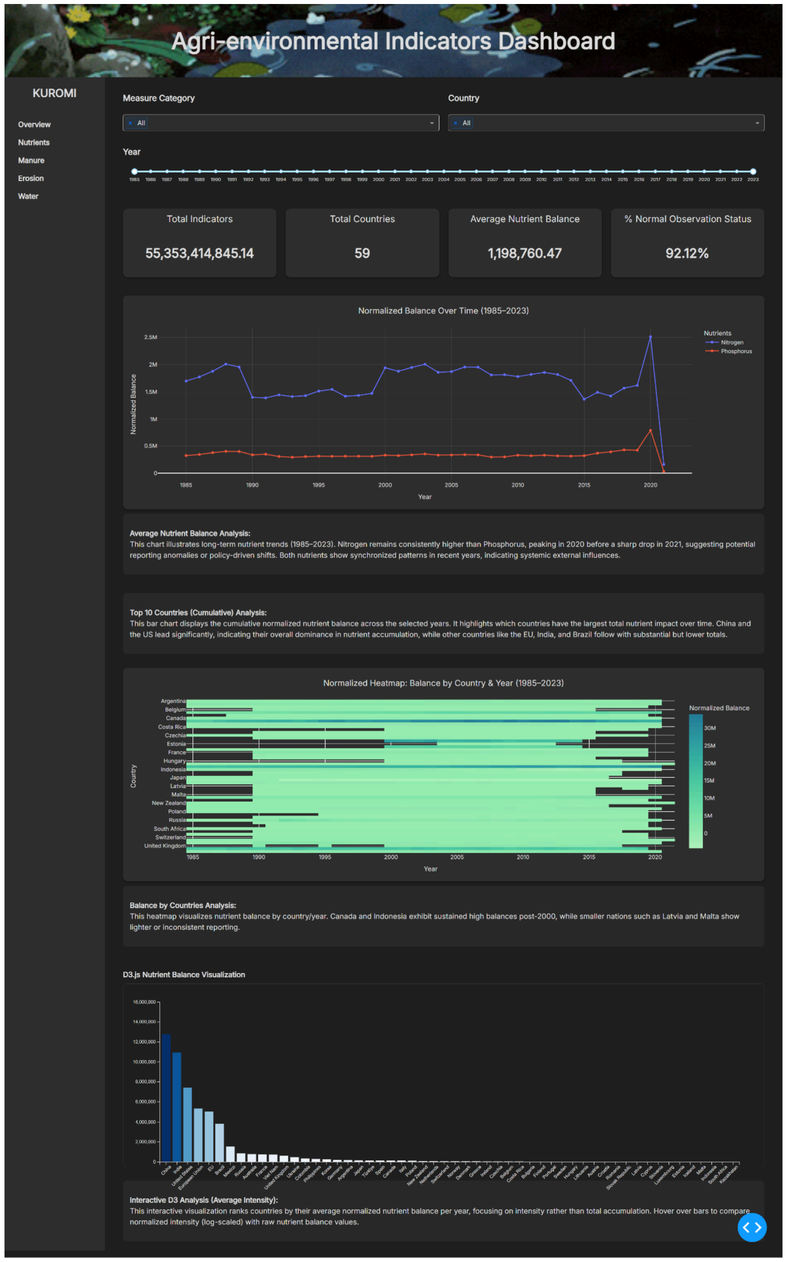
Wireframe of Overview page's final draft

3.1.3. Visualization Design (Final)

Final Dashboard Layout:

- **Top row:** Filter Controls (Measure Category, Country, Year Range)
- **Second row:** KPI Cards that describe Total Indicators, Total Countries, Average Nutrient Balance, and % Normal Status.
- **Visualisations:**
  1. **Trend Chart (Plotly):** Nutrient balance trends over time.
  2. **Heatmap (Plotly):** Country vs. Year nutrient balance over time.

3. **Bar Chart (D3):** Interactive chart of ranking countries based on the average nutrient balance.



Final Dashboard Layout of Overview page

Justification of Visual Encodings:

- **KPIs:** Cards showing big numbers to attract the user's attention.
- **Line Chart:** Showing trend of nutrient balance.
- **Heatmap:** Showing gaps between countries over time.
- **Bar Chart:** Showing ranking between countries based on nutrient balance.

The reason of this working designs:

- This layout is ideal because of its top-down flow, it ensures coherence for users to follow.
- Showing a variety of knowledge when combining several frameworks, which are Plotly and D3.
- Provide explanations for each chart to enhance user understanding.

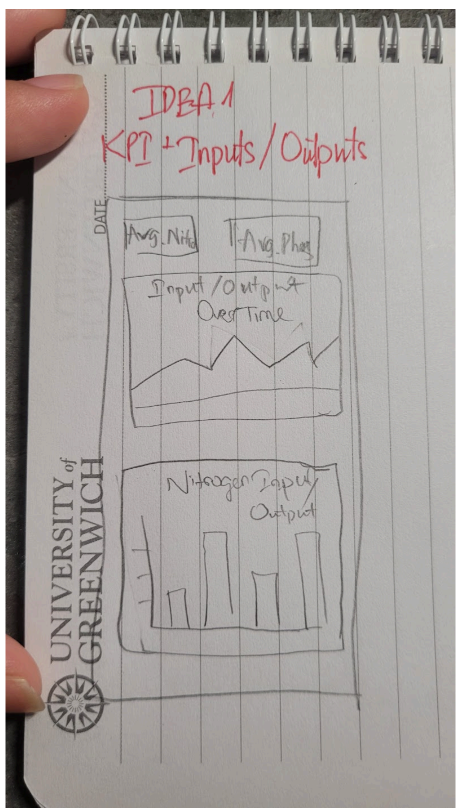
3.2 Nutrients Dashboard

3.2.1 Visualization Design (Proposal)

In this Nutrients page, the inputs/outputs and balances across country/year range are provided. There are 3 ideas proposed for this page:

Idea 1: KPI + Inputs/Outputs

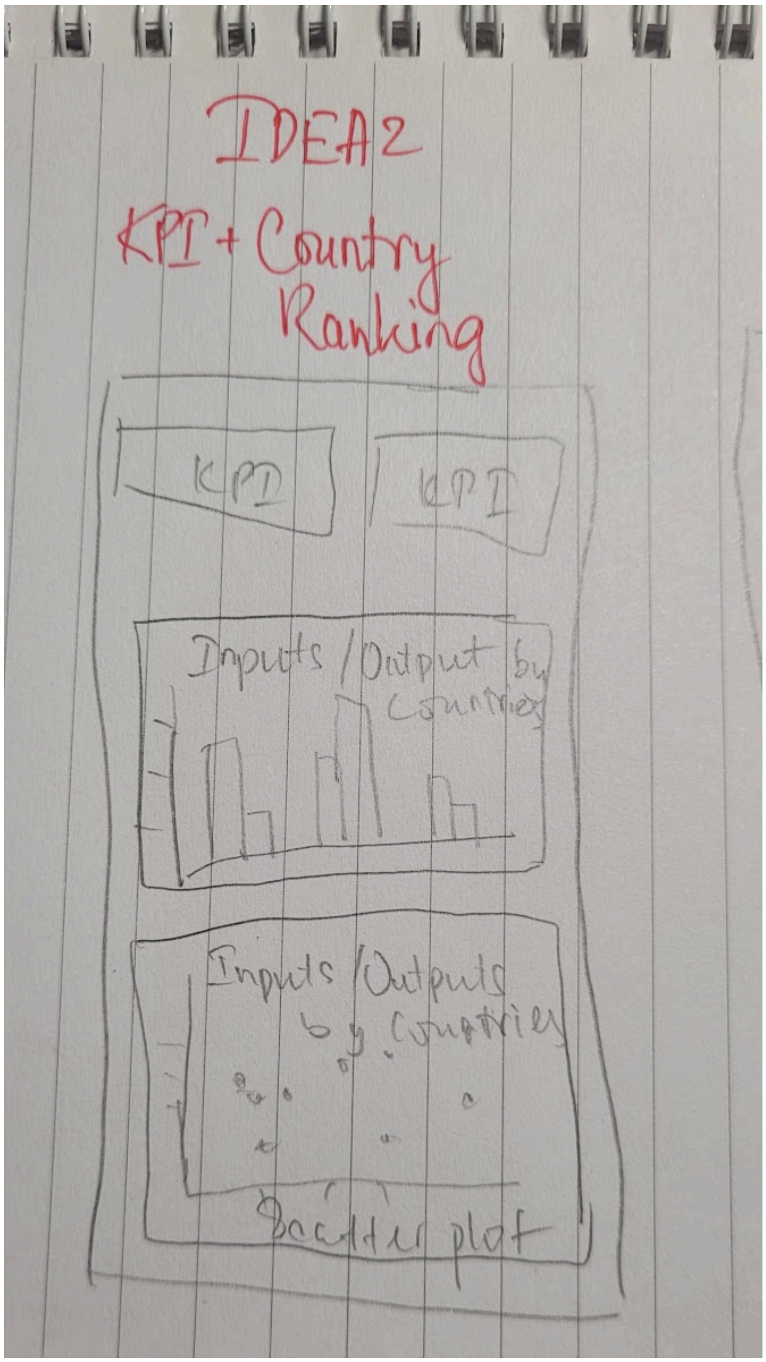
- **Structure:** Beginning with the KPI cards of average Nitrogen and Phosphorus balances, below are inputs/outputs line chart, bar chart of country by balance, and scatter plot for inputs/outputs.
- **Strength:** Highlight the flow of nutrients (inputs/outputs).



The draft layout for idea 1 of Nutrients page

Idea 2: KPI + Country Ranking

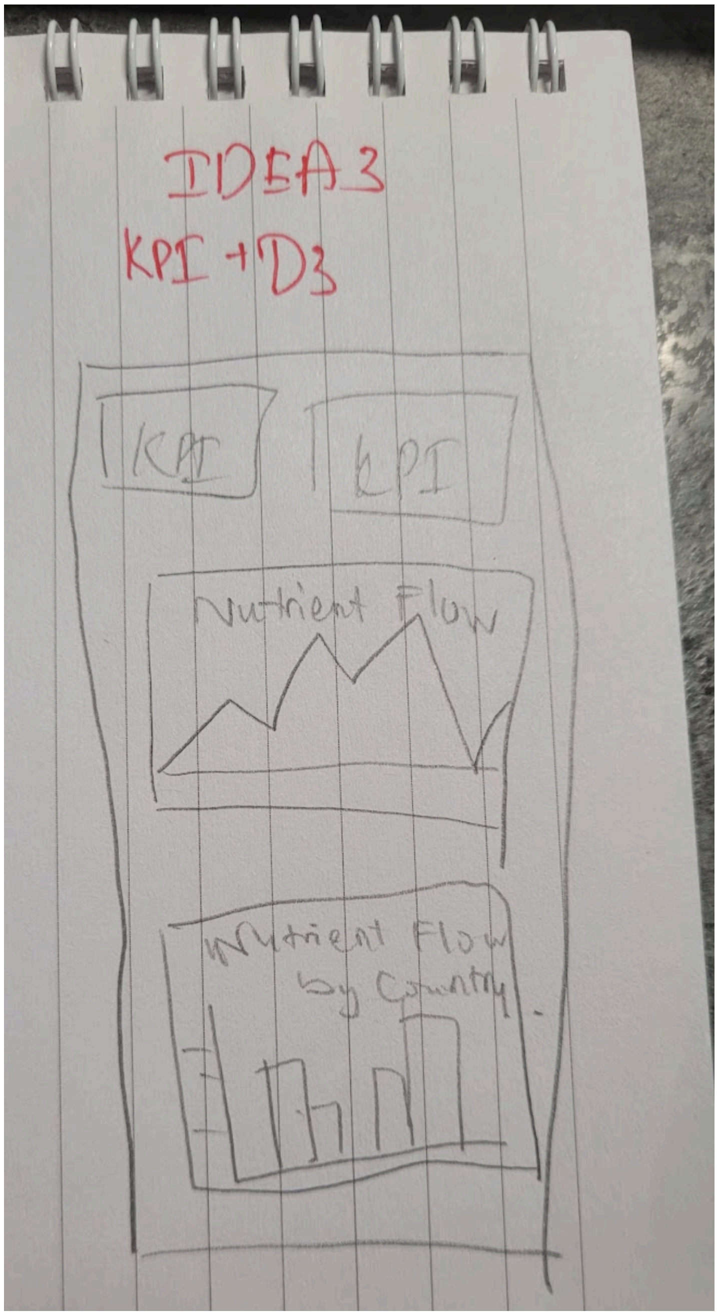
- **Structure:** Beginning with the KPI cards, below are a grouped bar chart of inputs/outputs by countries, and a scatter plot with the same topic.
- **Strength:** Highlight the country-level indicators, provide users an overview of nutrients based on the country first.



The draft layout for idea 2 of Nutrients page

Idea 3: KPI + D3 Interactive Chart:

- **Structure:** Beginning with the KPI cards, below are a trend chart for nutrient flow, and interactive D3 grouped bar chart for comparing nutrient flow between countries.
- **Strength:** Highlight the difference between Plotly and D3 framework in generating figures with the same topic.



The draft layout for idea 3 of Nutrients page

3.2.2 Visualization Design (Progress)

In the progress stage of visualization design including the iterative design, which iterate continuously between wireframing and implementation:

Stage 1: Wireframe

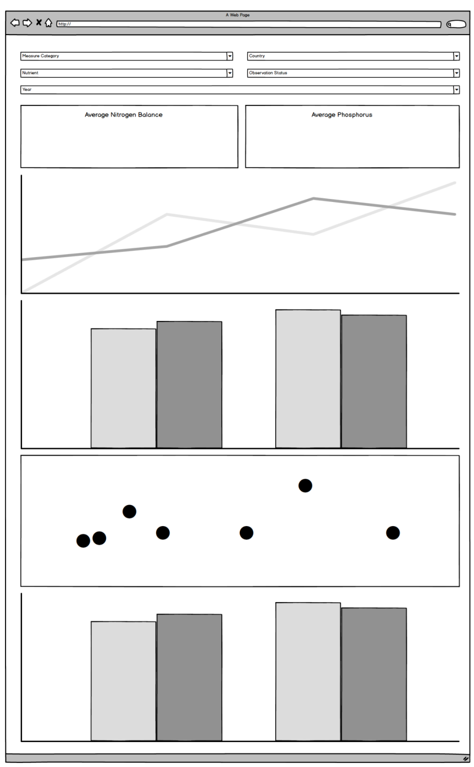
- Top row: Filter Controls (Measure Category, Country, Nutrient, and Observation Status)
- Second row: KPI Cards of Average Nitrogen and Phosphorus balance
- The early chart is only the trend chart of inputs/outputs over time.

Stage 2: Implementation

- Add a bar chart for average balance per nutrient by countries.
- Add a scatter plot of Nitrogen inputs/outputs by countries.

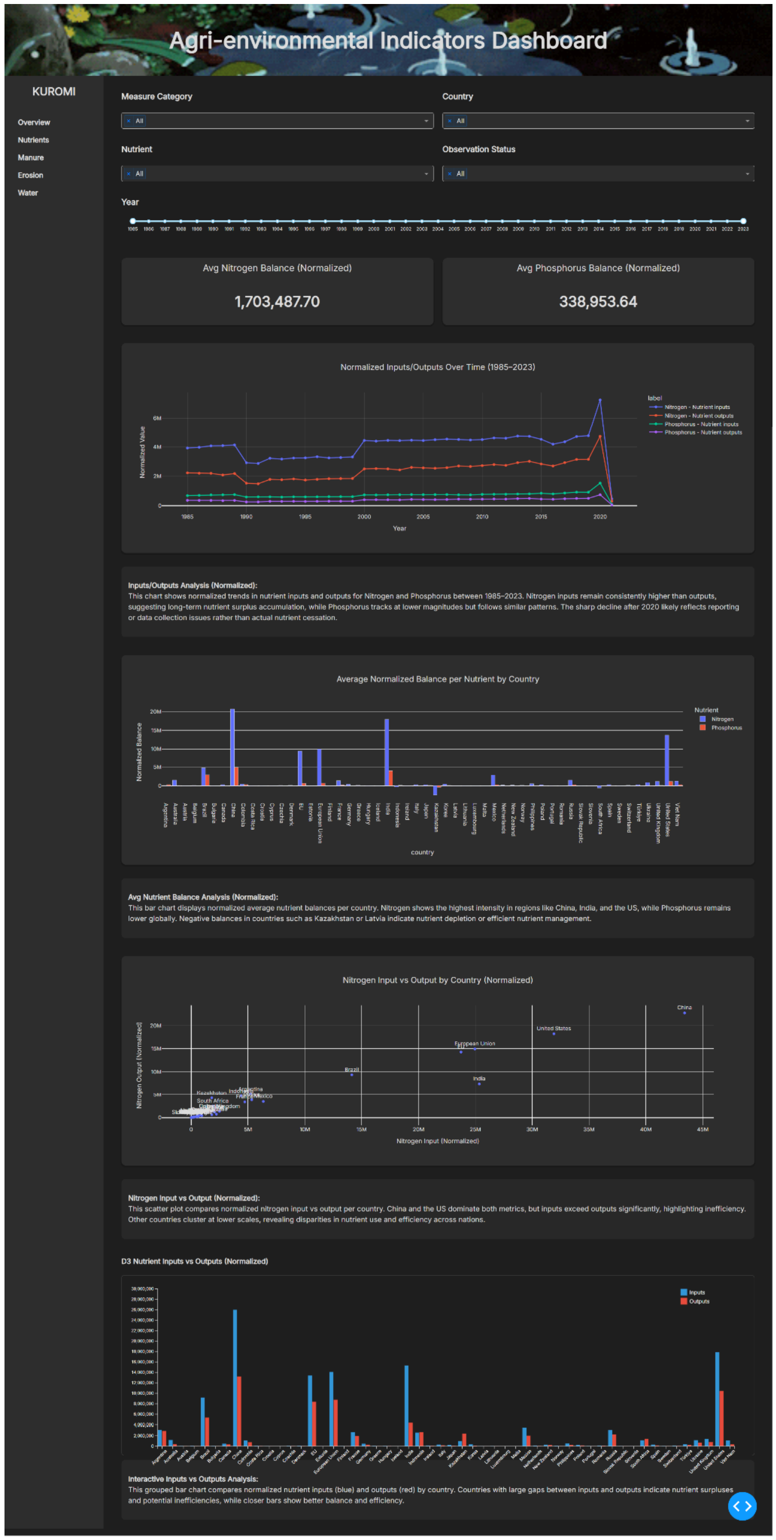
Stage 3: Final Touch

- Integrate the D3 grouped bar chart for comparing nutrient flow by countries.
- Applying a dark theme to align with other pages, also symbolizes Kuromi - name of our team.
- Add boxes of explanatory notes under each chart for comprehensive user guide.



Wireframe of Nutrients page's final draft

3.2.3 Visualization Design (Final)



Final dashboard layout of Nutrients page

Final Dashboard Layout:

- **Top row:** Filter Controls (Measure Category, Country, Nutrients, and Observation Status)
- **Second row:** KPI Cards of Average Nitrogen balance and Average Phosphorus balance.
- **Visualisations:**
  1. **Trend Chart:** Inputs/Outputs Over Time
  2. **Grouped Bar Chart:** Average Balance per Nutrient by Country
  3. **Scatter Plot:** Nitrogen Input vs. Output per Country
  4. **Grouped Bar Chart:** D3 chart of Nutrient Inputs/Outputs by Country

Justification of Visual Encodings:

- **KPIs:** Show average balance of Nitrogen and Phosphorus for quick comparison.
- **Trend Chart:** Show the Nutrient Flow over time.
- **Grouped Bar Chart:** Show the average balance of Nitrogen vs. Phosphorus by Country.
- **Scatter Plot:** Show the Nitrogen flow by Country.
- **Grouped Bar Chart:** Show the use of D3 in comparing nutrient flow by Country.

The reason of working design:

- Beside top-down flow, this design also provides the sequential analysis, which is KPIs → trend → country rankings → comparing efficiency.
- Consistent dark theme across the whole project.
- Provide explanations for each chart to enhance user understanding.

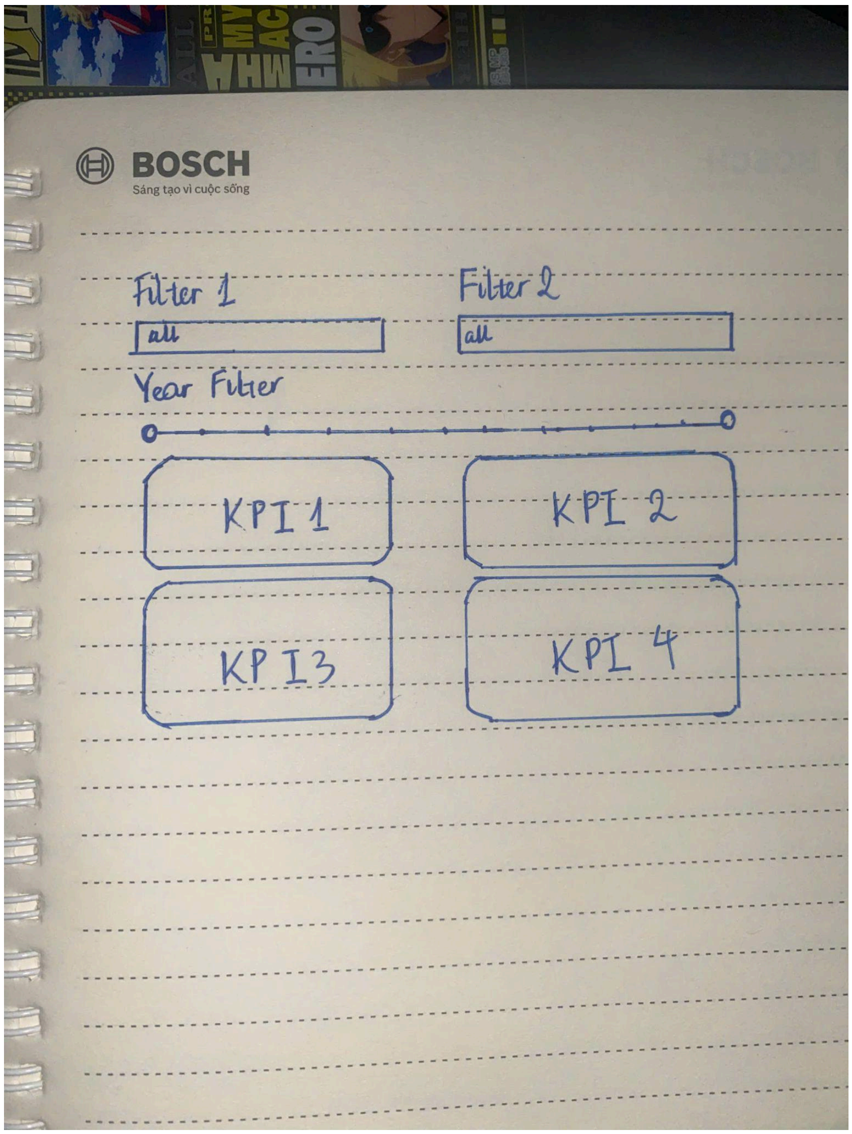
3.3. Manure Dashboard

3.2.1. Visualization Design (Proposal)

The Manure page was envisioned as a filter-driven, multi-layered exploration tool combining high-level metrics with geographic, comparative, distributional, and compositional views. Building on our wireframe, I sketched out the following coordinated components:

- **Global Filters:**
  1. **Country Dropdown** (default = All)
  2. **Nutrient Dropdown** (default = All)
  3. **Year Slider** (spanning full range of 1985 - 2023)

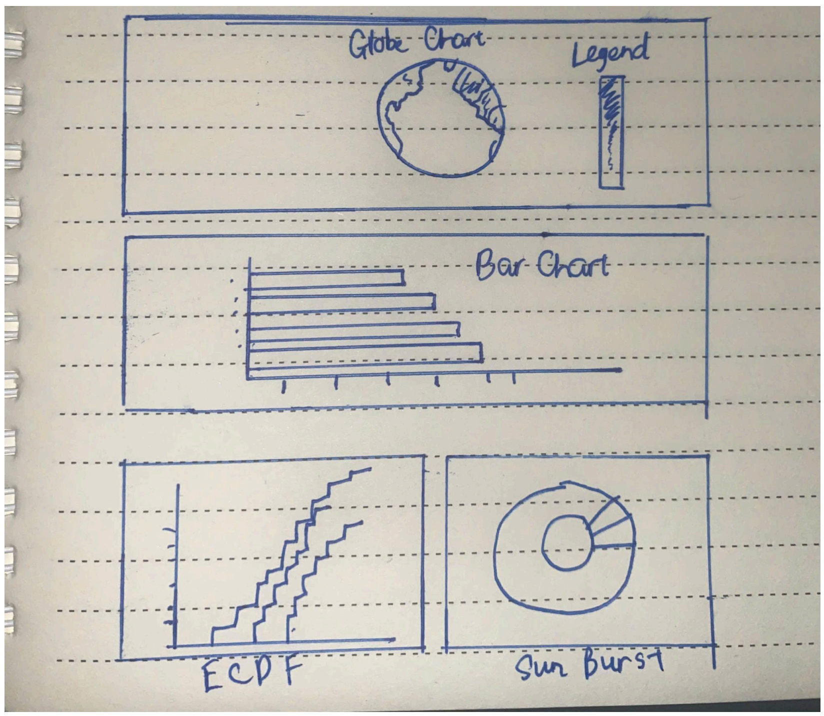
- **KPI Cards -** Four prominent cards to surface at-a-glance values:
  1. Total Livestock Manure Production
  2. Average Net Input of Manure
  3. % Manure-related Categories
  4. Top Country by Net Input



Wireframe of Manure page with filters and KPI cards

- **Coordinated Charts:**
  1. **Globe Chart** - a rotatable world view, coloring countries by aggregated manure indicators.

2. **Bar Chart** - horizontal bars of the top 10 countries normalized by agricultural land (log scale).
3. **ECDF Plot** - overlaid step lines showing cumulative distributions of manure categories over time.
4. **Sunburst Chart** - a multi-level breakdown of manure system components radiating from a central node.



Sketches of Globe, Bar, ECDF, and Sunburst charts

These early designs ensured we covered four analytical angles - spatial, ranking, distributional, and compositional - while keeping the interface intuitive and filter-centric.

3.2.2. Visualization Design (Progress)

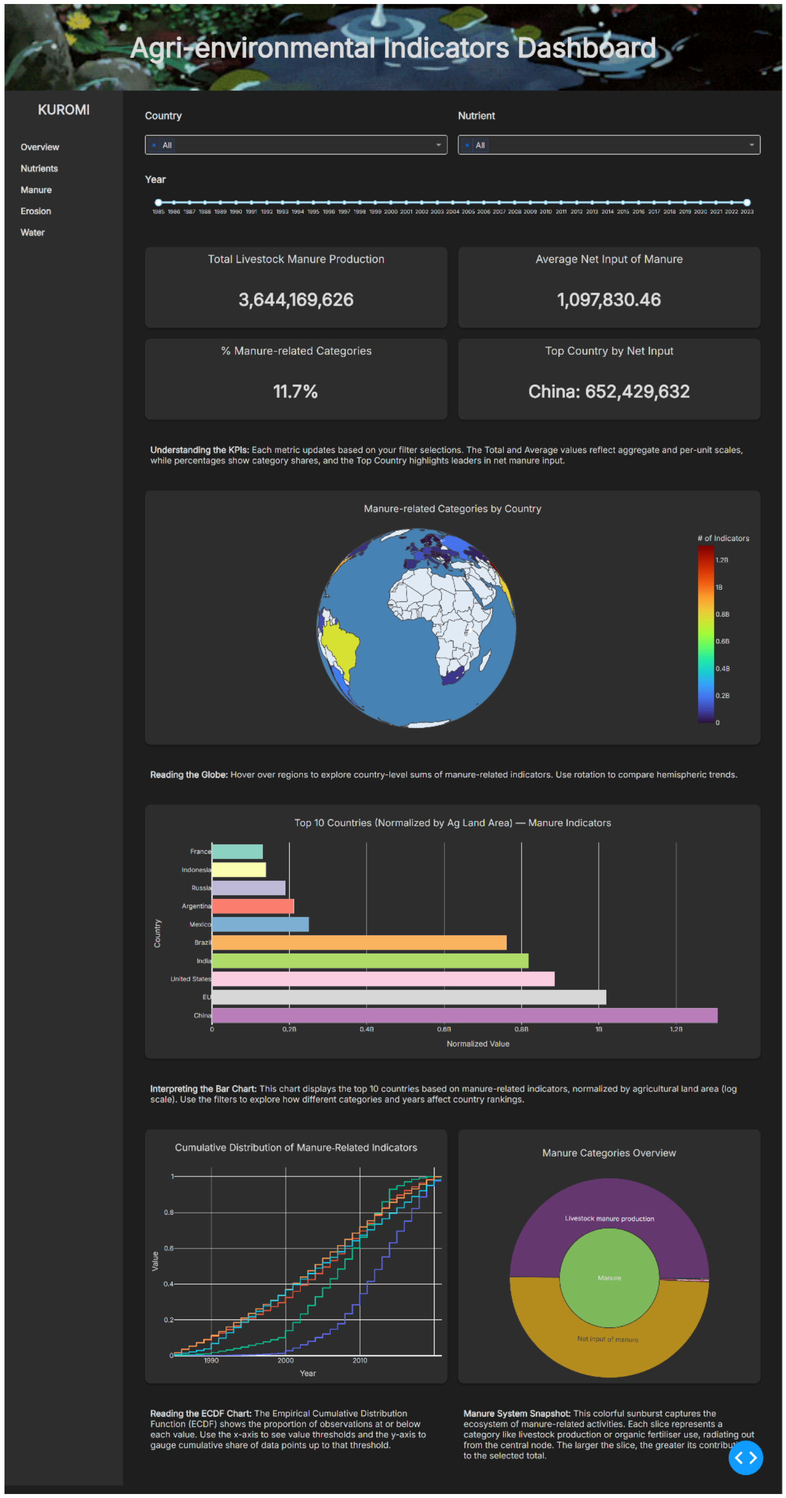
With the proposal and sketches finalized, I translated the designs into a live Dash application:

1. **Layout & Components**
   - Used Dash's `dcc.Dropdown` and `dcc.Slider` for filters, placing them in a responsive row at the top. Made `filters.py` to make these components reusable.

   - Implemented KPI cards using `dbc.Card` components, styled with left-aligned titles and large numeric headings.
2. **Charting with Plotly**
   - **Globe Chart**: Leveraged `px.choropleth` with an orthographic projection in `dcc.Graph`, enabling drag-to-rotate and hover tooltips.
   - **Bar Chart**: Built a horizontal `px.bar` chart normalized by agricultural land, rendered on log scales with dynamic axis tick formatting.
   - **ECDF**: Computed empirical CDFs in Python callbacks, then plotted as `px.ecdf`.
   - **Sunburst**: Used `px.sunburst` fed by a hierarchical DataFrame of manure categories and values.
3. **Interactivity & Callbacks**
   - A single multi-output callback listens to all three filters and updates the cards and four charts simultaneously.
     Hover and click events surface additional details (e.g. exact values, category descriptions).
4. **Descriptive Text & Guidance**
   - Inserted Markdown `html.P` paragraphs under each visualization, outlining how to interpret it - e.g. explaining the ECDF’s axes or the meaning of each sunburst slice.
   - Added a brief “Understanding the KPIs” note beneath the cards to clarify aggregate vs. per-unit metrics.
5. **Styling & Theming**
   - Applied a dark theme, ensuring visual consistency and focus on data.
   - Consistent font sizes and spacing were set with Dash Bootstrap utilities.

This Dash-based implementation concretely realizes our design intent, marrying responsive controls, coordinated analytics, and explanatory text to guide users through the manure data story.

3.2.3. Visualization Design (Final)



Final Manure Dashboard

In the final Manure page, four coordinated charts form the core analytical experience.

The **globe chart** sits top-left, using an orthographic projection to color each country by its manure indicator; users can drag to rotate, click to zoom on regions, and watch country hues animate as the year slider moves.

Below it, a **horizontal bar chart** dynamically ranks the top ten countries (normalized by agricultural land), automatically resorting to filter changes and offering a log-scale toggle for deeper comparison.

Below these, the **ECDF plot** overlays step lines for each manure category - each with a distinct dash style and includes a draggable vertical line that instantly highlights the cumulative percentage at that threshold.

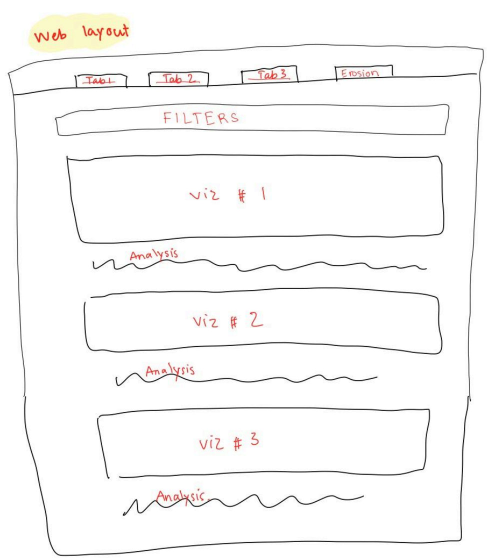
Finally, the **sunburst chart** breaks down manure system components into concentric rings, with hover tooltips that show both absolute values and percent-of-total, and a clickable breadcrumb trail for intuitive drill-up and drill-down navigation.

Each chart updates in real time as users adjust country, nutrient, or year, delivering a seamless, multi-perspective view of global manure dynamics.

3.4 Erosion Dashboard

3.4.1. Visualization Design (Proposal)

Here is the initial layout sketch for erosion dashboard:

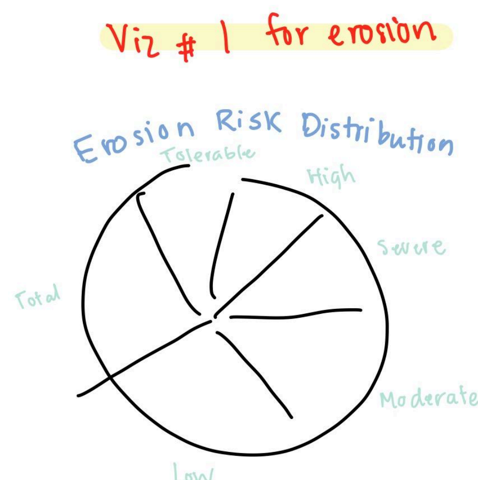


Initial Erosion Dashboard Layout

The erosion dashboard features three visualization designs that provide comprehensive analysis of global agricultural erosion risks. Interactive filters positioned at the top of the dashboard allow users to adjust parameters with all visualizations updating simultaneously to reflect the applied filters.

Each visualization is accompanied by analytical commentary positioned directly below the chart, offering contextual interpretation and actionable insights derived from the data patterns. This integrated design ensures users can understand both what the data shows and its implications for agricultural policy and erosion management strategies.

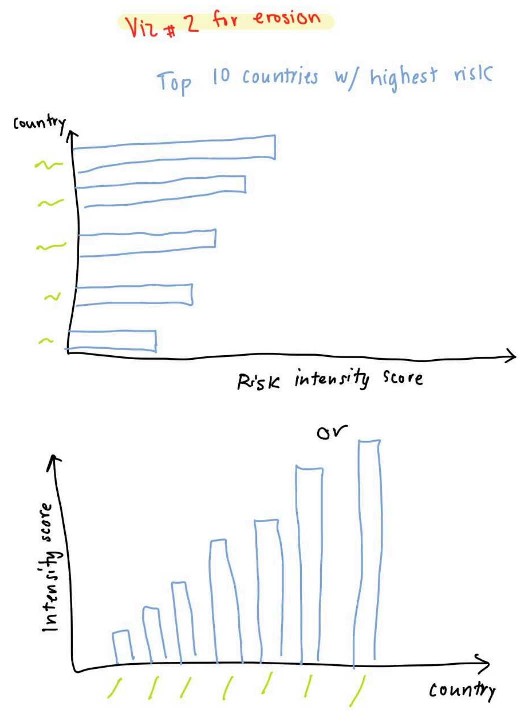
The first visualization design is a pie chart to provide an immediate, intuitive understanding of the overall erosion risk landscape across different severity levels (low, moderate, tolerable, high, server, and total)



Visualization #1: Initial sketch for erosion risk distribution

The second visualization design is in a horizontal bar chart format to effectively rank and compare erosion intensity across countries, with the horizontal orientation allowing for easy country name readability without text rotation issues.

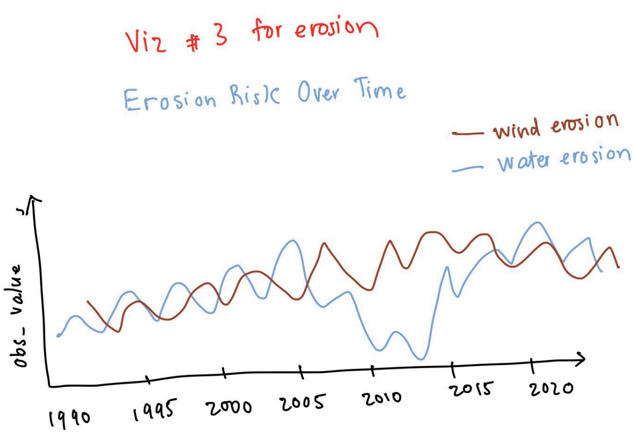
This visualization enables clear comparative analysis of which nations face the most severe erosion challenges, with bar lengths providing immediate visual ranking. This analysis will focus on identifying erosion hotspots globally, allowing policymakers and researchers to prioritize resource allocation and international cooperation efforts toward the countries most critically affected by agricultural soil degradation.



Visualization #2: Initial sketch for top 10 countries with highest erosion risk

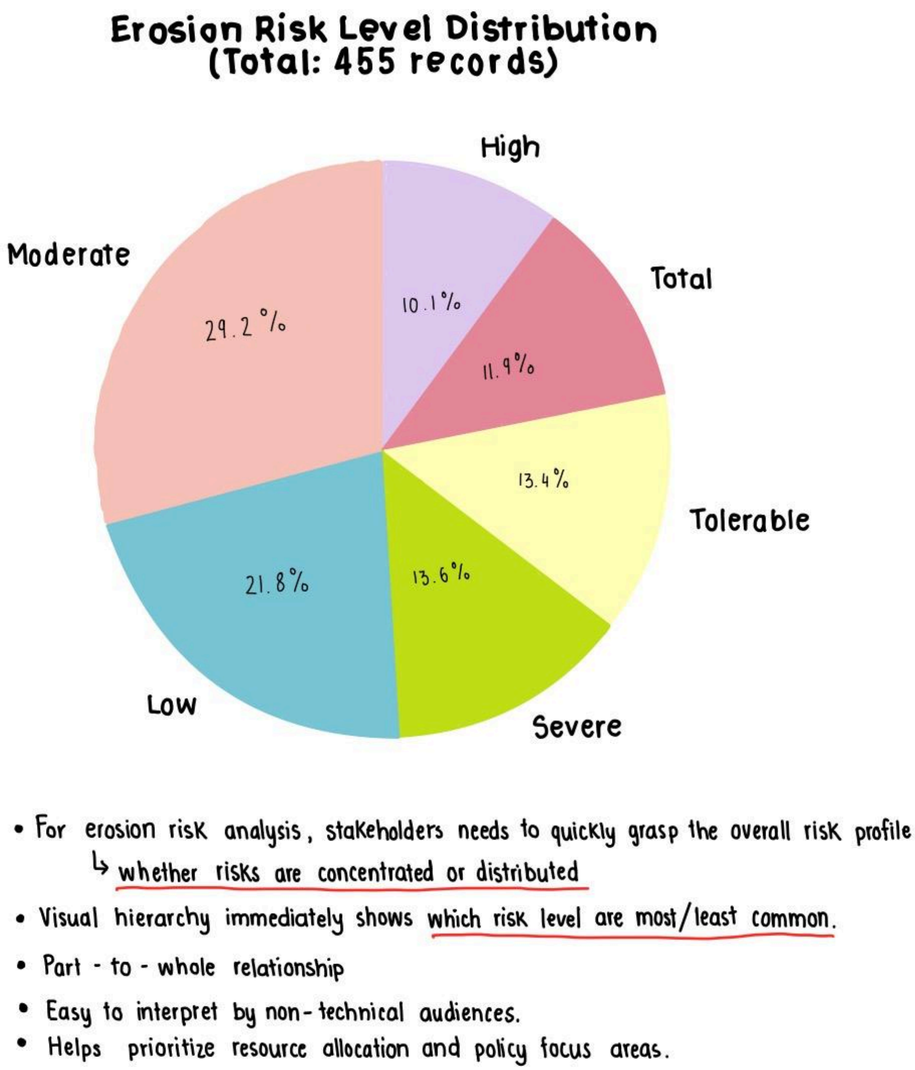
The third visualization design is a dual-line time series to reveal temporal patterns and trends in both wind and water erosion from 1990 to 2020, allowing for comparative analysis of how these two distinct erosion types evolve differently over decades.

This analysis will focus on understanding whether erosion risks are worsening over time, identifying periods of particular vulnerability, and determining if wind and water erosion follow similar patterns or require different seasonal/cyclinal management approaches.



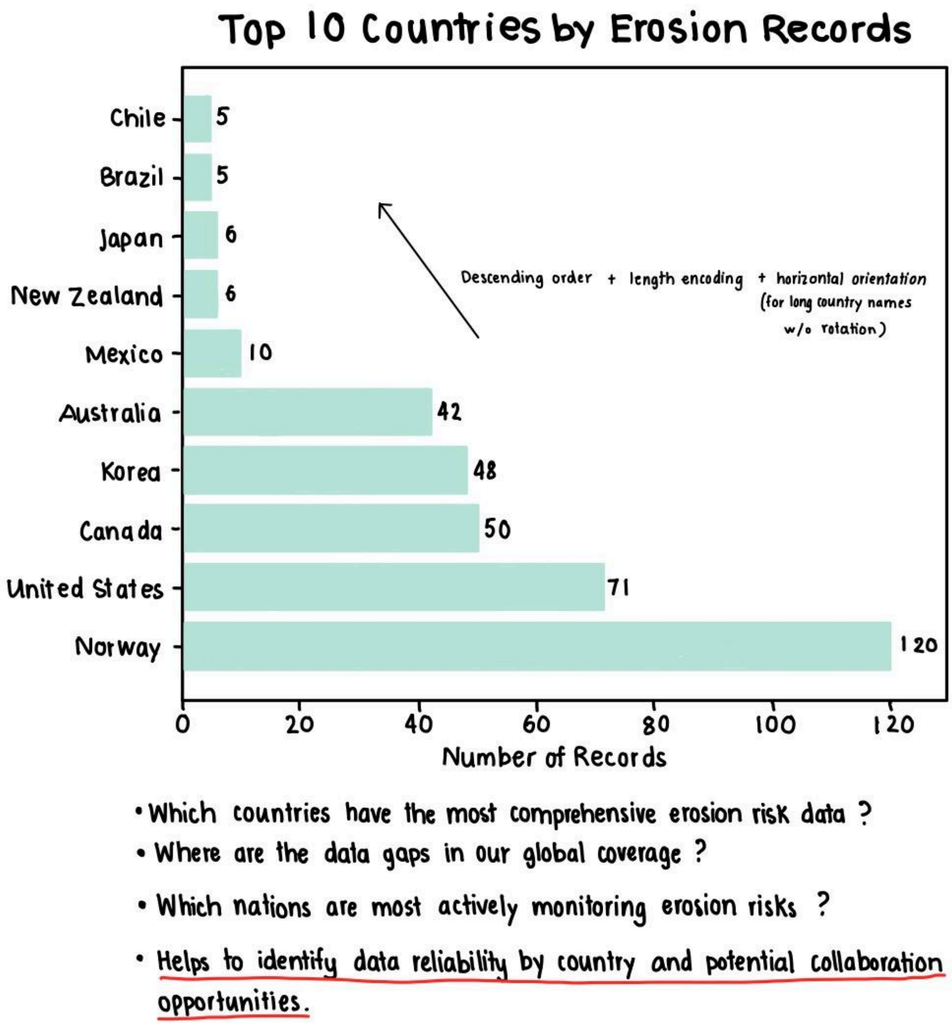
Visualization #3: Initial sketch for erosion risk over time

3.4.2. Visualization Design (Progress)



Updated visualization design #1

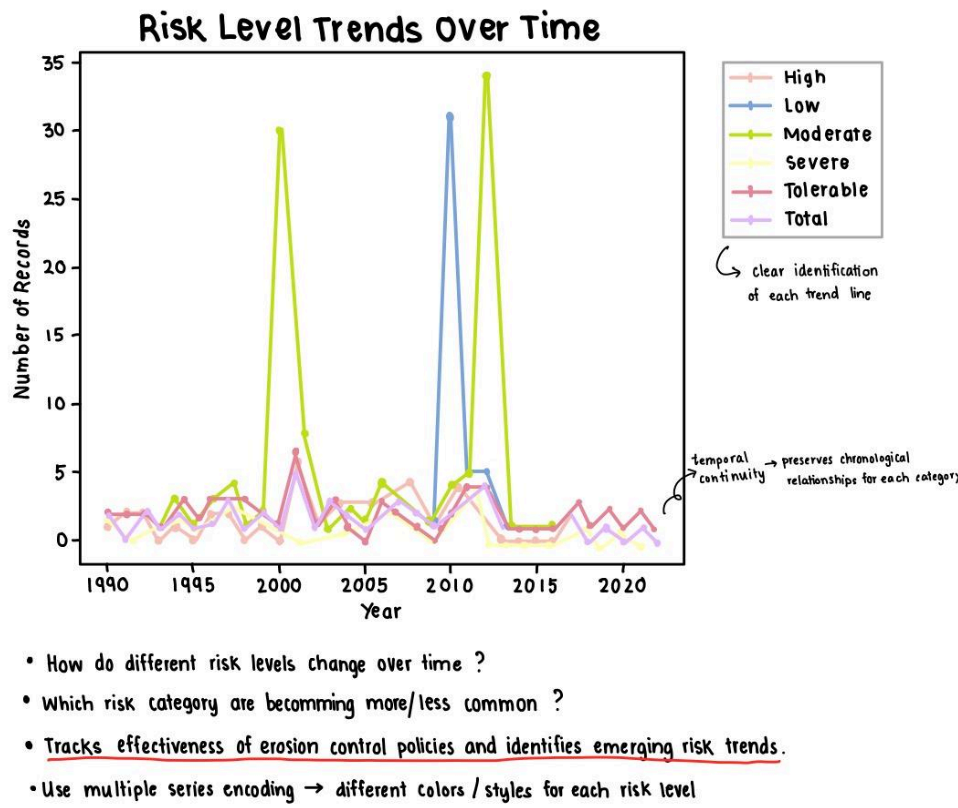
The updated visualization design #1 evolved to include detailed percentage breakdowns for each risk category, total record count, and explanatory bullet points that articulate the visualization's analytical purpose (will be used as the outline for the analysis below the chart).



Updated visualization design #2

The updated design #2 demonstrates a pivot from risk intensity ranking to data coverage. While maintaining the horizontal bar format for optimal country name readability, the focus shifted from displaying "highest risk countries" to "countries with most erosion records," revealing data reliability and monitoring comprehensive rather than solely risk severity.

The transformation converts the visualization from a risk assessment instrument into a data quality and research infrastructure evaluation tool, enabling identification of countries with robust erosion monitoring systems versus those requiring enhanced surveillance capabilities.



Updated visualization design #3

The visualization design #3 represents a fundamental conceptual shift from the original dual-line sketch focusing on wind versus water erosion types. The progress prototype expanded to track all six risk levels simultaneously across the 1990-2020 timeline, creating a more comprehensive temporal analysis framework.

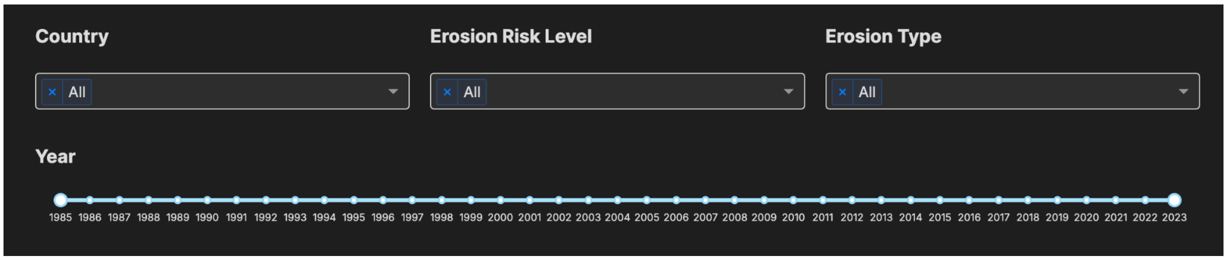
The design emphasizes visual differentiation through distinct colors and line styles for each risk category, with a legend system that ensures immediate category identification.

3.4.3. Visualization Design (Final)

The erosion dashboard incorporates a comprehensive filter system designed to enable dynamic exploration of agricultural erosion data across multiple dimensions. The filter interface consists of three multi-select dropdown menus positioned horizontally across the top: Country, Erosion Risk Level, and Erosion Type, each defaulting to “All” selections to provide complete dataset visibility upon initial load.

Below these categorical filters, a year range slider spans from 1985 to 2023, allowing users to focus their analysis on specific periods or examine long-term trends across the entire dataset timespan.

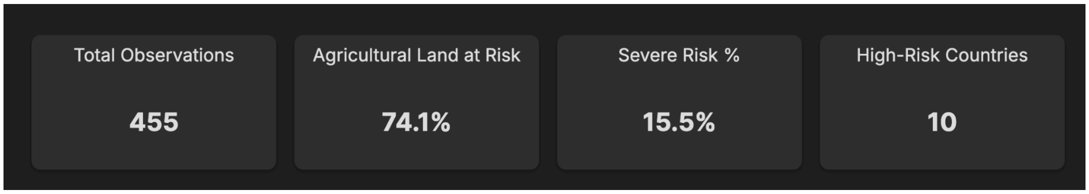
The filter system operates with real-time interactivity, meaning all visualizations and KPI cards update immediately when using the filters, ensuring cohesive and synchronized dashboard behavior.



Erosion Dashboard Filters

The dashboard features four designed Key Performance Indicator (KPI) cards that provide immediate quantitative insights into erosion risk status. The "Total Observations" card displays the complete count of erosion monitoring records (455 totals), serving as data coverage indicator that helps users understand the robustness of the underlying dataset, with hover details revealing breakdowns by erosion type (water vs. wind erosion), risk level distribution, year range coverage, and unique country counts to help users understand dataset composition and reliability. The "Agricultural Land at Risk" card presents the percentage of monitored agricultural areas experiencing erosion concerns (74.1%), providing a critical metric for understanding the scope of agricultural vulnerability, with hover information showing top countries by land risk percentage, total countries monitored, and data source methodology (whether calculated from risk levels or direct measurement). The "Severe Risk %" card quantifies the proportion of observations classified as severe risk (15.5%), highlighting the most critical erosion situations that require immediate intervention, with hover details providing country-specific severe risk counts, comparisons between severe and total observations, and geographic distribution of severe risk areas. The "High-risk Countries" card counts the number of nations with elevated erosion concerns (10 countries), enhanced with hover functionality that lists complete country names with their specific risk levels and percentage, summary statistics separating severe versus high-only risk countries.

Each KPI card updates dynamically based on applied filters, ensuring that both the primary metrics and details hover information reflect the specific analytical focus selected by the user.

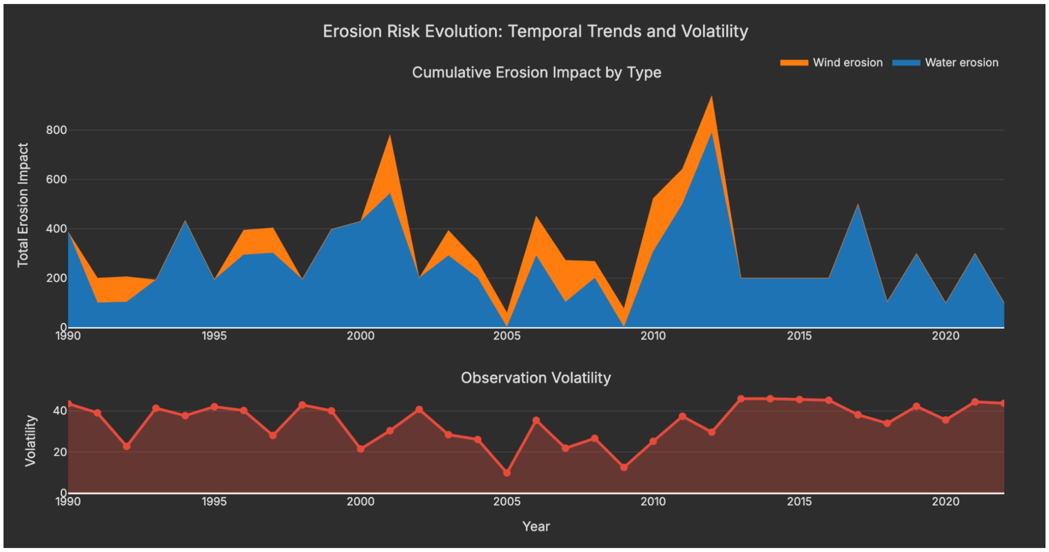


KPI Cards in the Erosion Dashboard

KPI Cards with hover details

The first erosion visualization is implemented using Plotly's make_subplots functionality to create a dual-panel layout with a 70/30 height ratio, where the upper panel displays cumulative erosion impact through stacked area charts and the lower panel shows volatility trends as a filled line chart. The implementation processes filtered erosion data by grouping observations by year and measure category, calculating total erosion impact, observation counts, and volatility metrics (standard deviation), then renders distinct erosion types using a predefined color palette (blue for water erosion, orange for wind erosion).

This visual design follows several key principles: hierarchical information architecture through the two-panel structure that separates primary trends from supporting volatility data, visual differentiation using contrasting colors and fill patterns to distinguish between water and wind erosion types, and progressive disclosure where users can examine broad temporal patterns in the upper chart while accessing detailed uncertainty metrics in the lower area visualization that emphasizes the relative contribution of each erosion type over time, while the volatility chart uses fill-to-zero to create visual weight that represents data uncertainty.



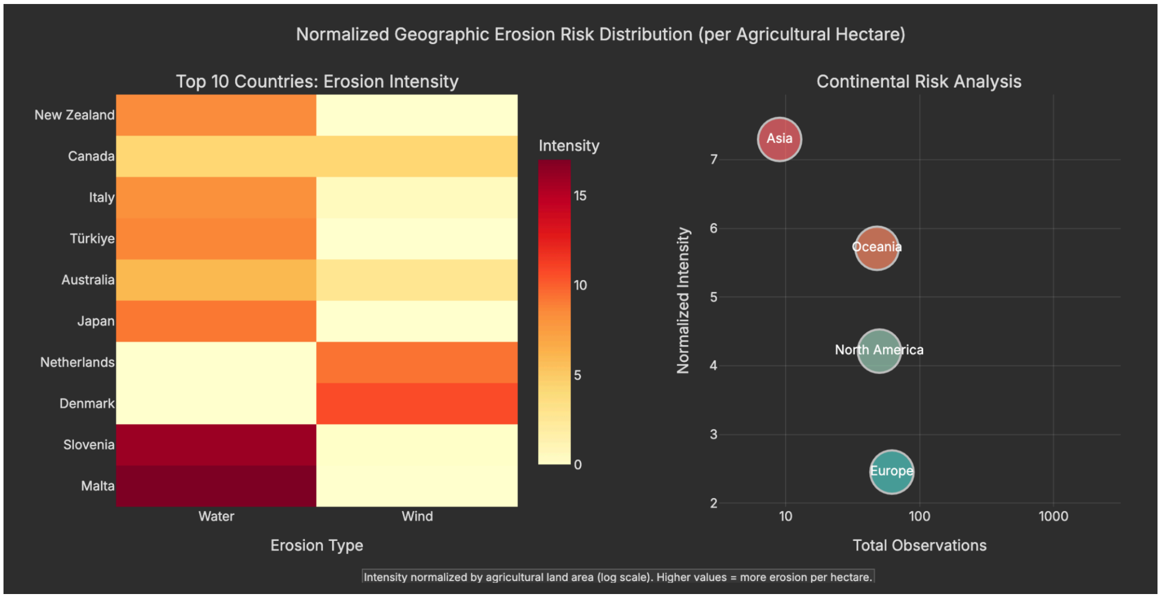
Final Visualization Design #1 for Erosion Dashboard

The second visualization design is a geographic risk distribution with a 1x2 layout featuring a heatmap and a bubble chart, incorporating logarithmic normalization by agricultural land area through the `normalize_by_agricultural_land()` function to ensure fair cross-country comparisons per hectare rather than absolute values.

The implementation filters countries to only include those with valid agricultural land data, aggregates erosion data by country and measure category, then applies the normalization before creating a pivot table for the heatmap matrix and continental summaries for the bubble chart.

The visual design employs several principles: data normalization transparency through clear labeling and explanatory annotations that inform users about the logarithmic scaling methodology, dual-encoding strategies where the heatmap uses both position and color intensity while the bubble chart combines positions, size, and color to represent multiple complexity principles by presenting detailed country-level data in the left heatmap for focused analysis, then providing broader continental patterns in the right scatter plot for contextual understanding.

Visual hierarchy is established through the color scale that intuitively maps light-to-dark with low-to-high risk intensity, while the bubble chart employs proportional encoding where bubble sizes represent risk scores and positions indicate observation coverage versus intensity relationships.



Final Visualization Design #2

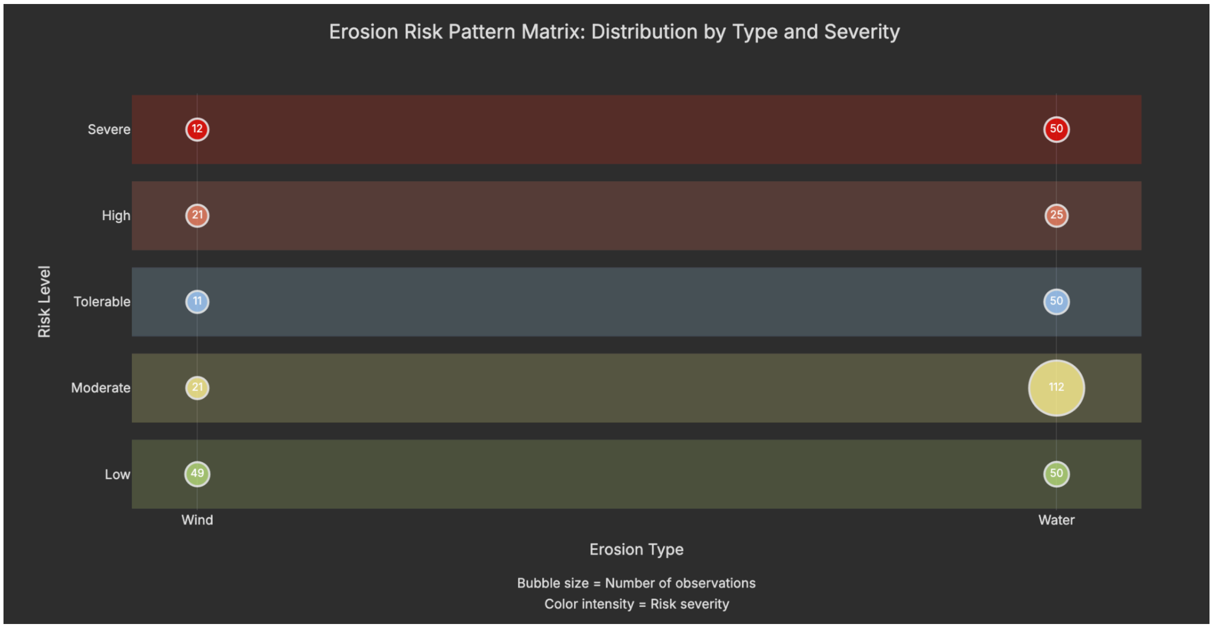
The risk pattern matrix visualization is implemented as a single scatter plot where the data is filtered to exclude "Total" risk levels, then grouped by erosion risk level and measure category to calculate average intensity, observation counts, and standard deviations.

The implementation creates bubble markers positioned at discrete x-coordinates (erosion types) and y-coordinates (risk levels), with bubble sizes scaled proportionally to observation counts and colors mapped to a predefined risk severity palette.

This visual design employs matrix layout principles by organizing data in a structured grid where categorical variables (erosion type and risk level) create natural intersections for displaying quantitative relationships.

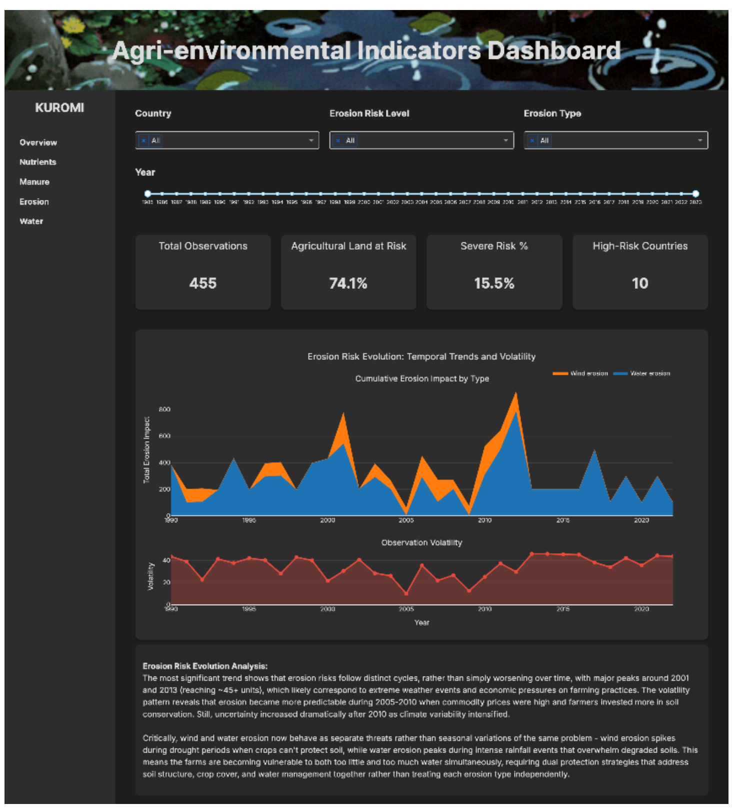
Multi-dimensional encoding is achieved through the use of position, size, color, and text simultaneously - position indicates the type-severity combination, bubble size represents data volume (observation frequency), color intensity communicates risk severity through an intuitive gradient from green (low) to red (severe), and embedded text labels provide precise quantitative values.

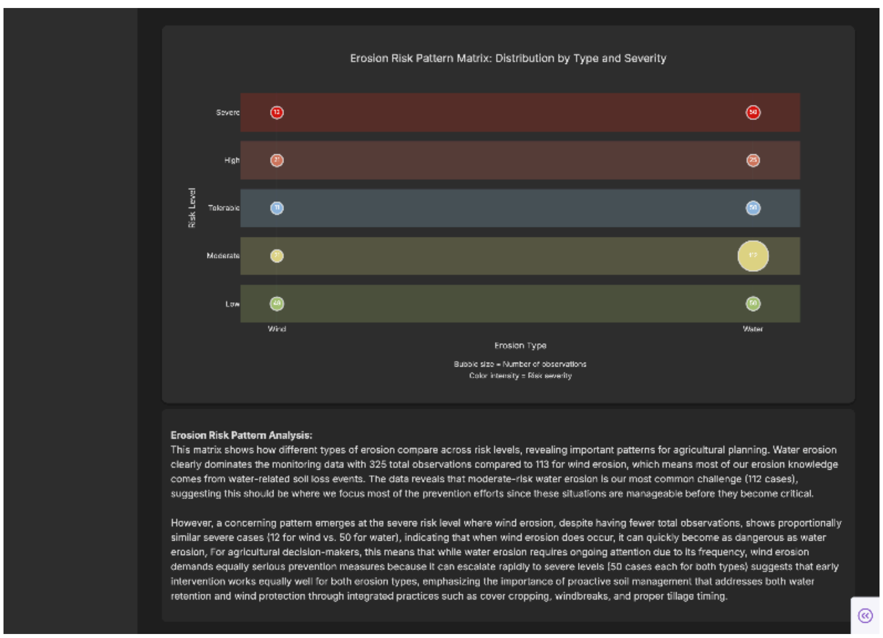
The design follows perceptual hierarchy principles by using background color bands that subtly reinforce the risk level categories while allowing the bubble markers to remain the primary focus. Cognitive load management is addressed through the dual annotation system at the bottom that explicitly explains the encoding scheme (bubble size = observations, color = severity), ensuring users can correctly interpret the multiple data dimensions.



Final Visualization Design #3

Here is the whole erosion dashboard including filters, KPI cards, and three final visual designs:





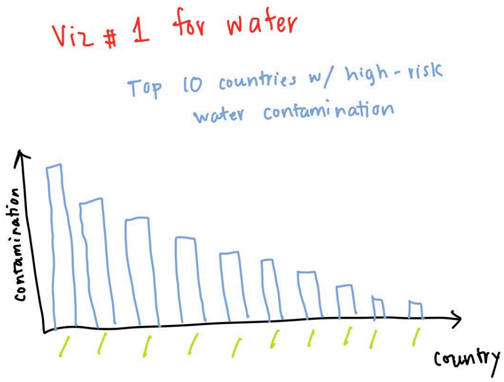
3.5. Water Contamination Dashboard

3.5.1. Visualization Design (Proposal)

The first initial sketch is a vertical bar chart visualization to provide a clear ranking of countries with the most severe agricultural water contamination issues, allowing for immediate identification of water quality hotspots that require urgent attention.

The descending bar format creates an intuitive hierarchy that makes it easy to compare contamination levels across nations and identify which countries are most at risk.

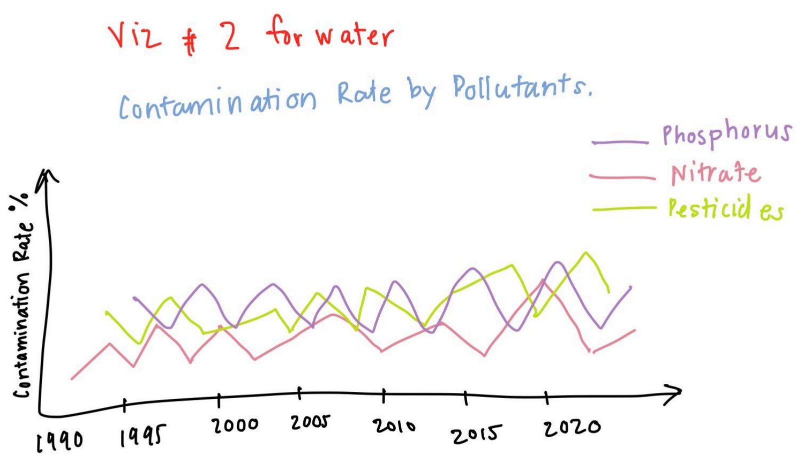
This analysis will focus on understanding the global distribution of water contamination severity, helping policymakers and international organizations prioritize resource allocation for water quality improvement programs.



Initial Sketch for Visualization Design #1

The next initial sketch is a multi-line time series format to reveal temporal trends and patterns for different pollutant types (Phosphorus, Nitrate, Pesticides) from 1990 to 2020, enabling comparative analysis of how various contamination sources evolve independently over decades.

This type of visualization allows for the identification of seasonal patterns, long-term degradation trends, and the effectiveness of pollution control policies across whether specific pollutants are becoming more problematic over time, identifying critical periods where contamination accelerated.

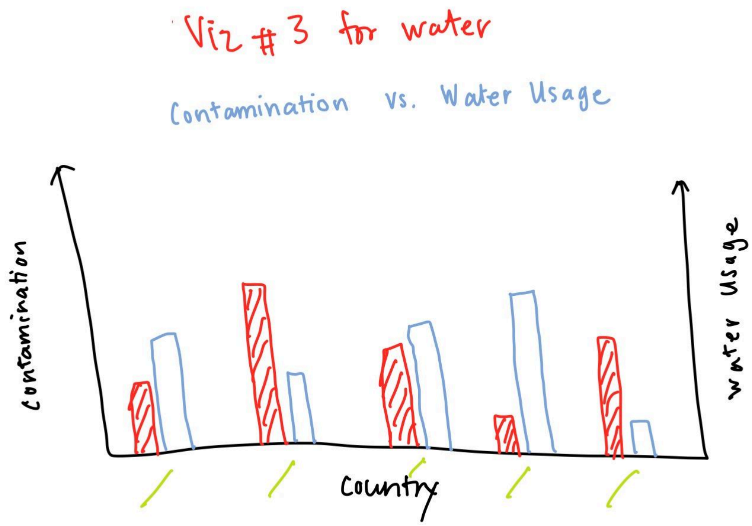


Initial Sketch for Visual Design #2

The initial sketch for grouped bar chart design was chosen to explore the relationship between water contamination levels and agricultural water consumption patterns across different countries, using dual y-axes to compare these two related but distinct metrics.

The grouped format allows for direct visual comparison of whether high water usage correlates with increased contamination rates, or if some countries achieve high usage with better contamination control.

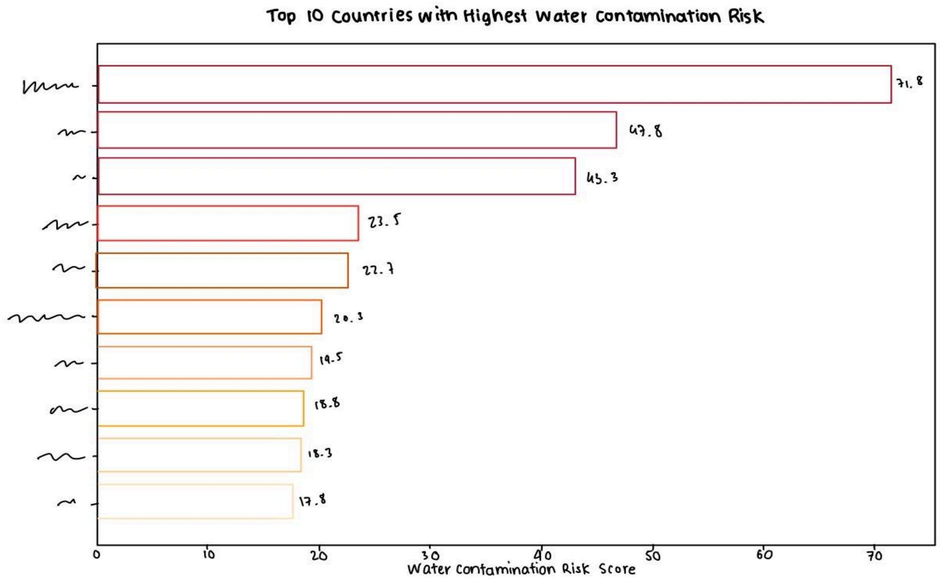
This analysis will focus on identifying water efficiency patterns, understanding whether intensive agricultural water use necessarily leads to higher contamination, and discovering best practices from countries that maintain low contamination despite high water usage.



Initial Sketch for Visual Design #3

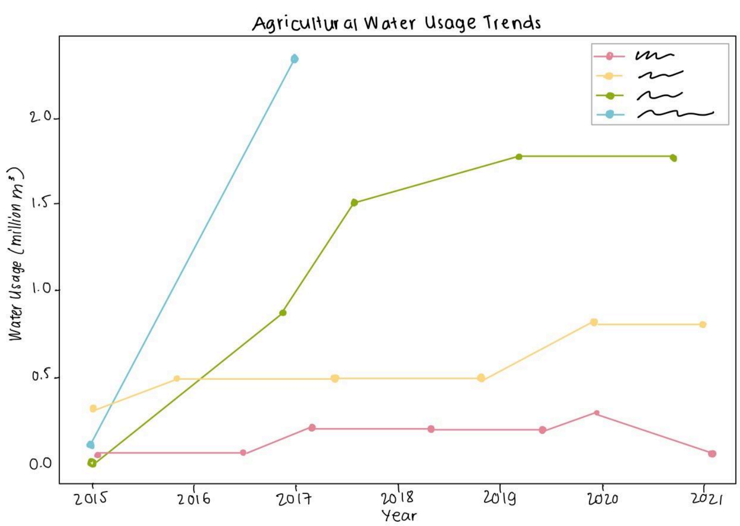
3.5.2. Visualization Design (Progress)

The most significant evolution from the initial visual design #1 is the shift to a horizontal orientation that can improve readability and data presentation. This change allows for full country name display without text rotation issues, while the color gradient from light yellow to deep red creates an intuitive risk severity mapping.



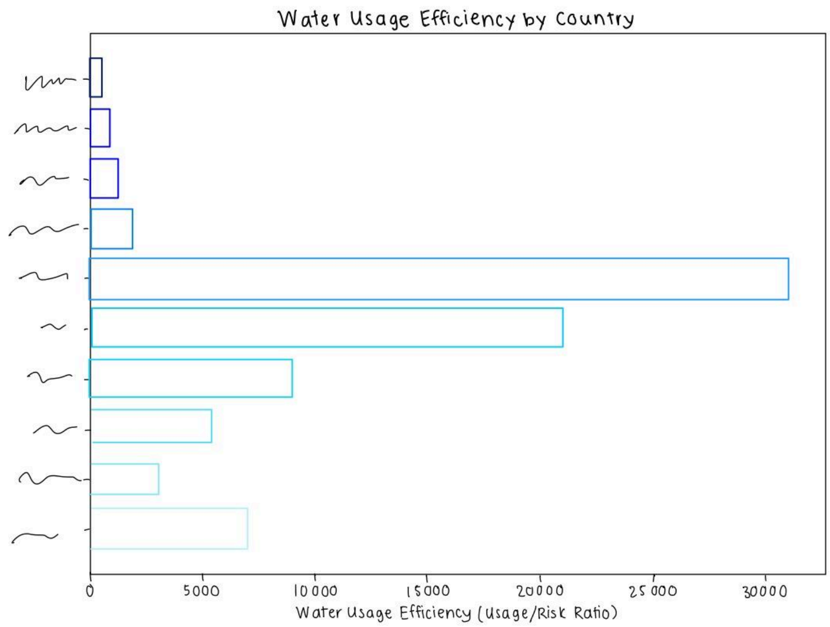
Updated Visual Design #1 for Water Dashboard

The updated visualization design #2 pivots from the contamination-by-pollutants sketch to focus exclusively on water usage patterns across different countries over time.



Updated Visual Design #2 for Water Dashboard

The updated visual design #3 shows a complete analytical reframing from the original contamination versus usage correlation concept to a focused efficiency assessment using a calculated Usage/Risk ratio metric.



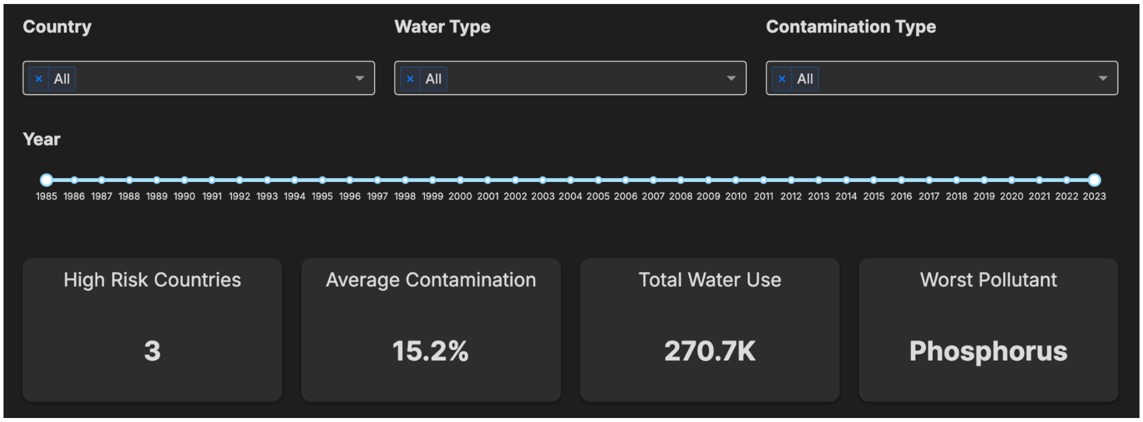
Updated Visual Design #3 for Water Dashboard

3.5.3. Visualization Design (Final)

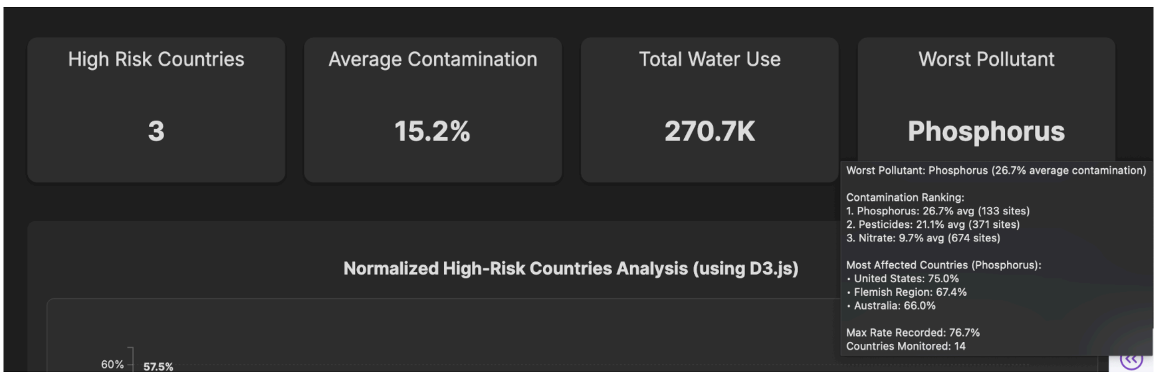
The water dashboard features four KPI cards that provide insights into agricultural water contamination and usage patterns, utilizing the same interactive filter system as other dashboards but adapted specifically for water quality metrics.

The “High-risk Countries” cards display the count of nations with contamination rates exceeding 30%, serving as an urgent alert system for identifying countries requiring immediate water quality intervention. The “Average Contamination” card presents the mean contamination percentage across all monitoring sites (15.2%), providing a baseline indicator of global water quality status in agricultural areas. The “Total Water Use” card quantifies agricultural freshwater abstraction (270.7K cubic meters), enabling resource consumption monitoring and sustainability assessment. The “Worst Pollutant” card identifies the most problematic contamination type (Phosphorus), helping prioritize pollution control strategies.

Similar to other dashboard sections, these cards respond dynamically to the Country, Water Type, Contamination Type, and Year range filters. Each KPI card incorporates hover functionality that reveals detailed breakdowns such as country-specific contamination rankings, pollutant type distribution, water source classifications (surface vs. groundwater), and temporal trends, ensuring that users can drill down from high-level metrics to actionable details for water quality management and policy formulation.



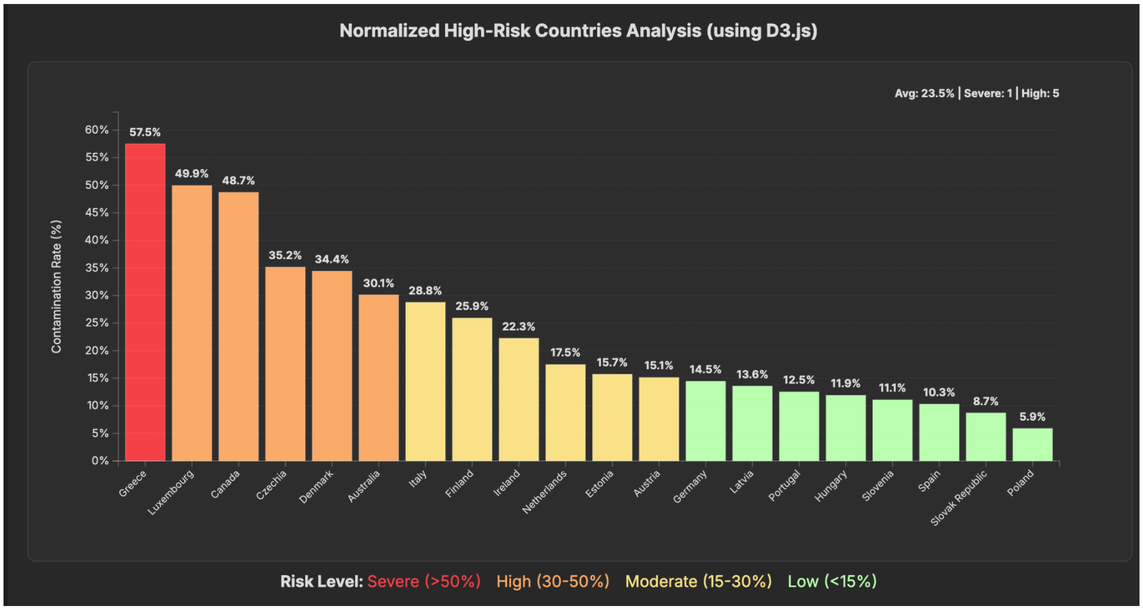
Filters and KPI Cards in the Water Dashboard



KPI Card with Hover Details

The final interactive high-risk countries visualization is implemented through a hybrid Dash-D3.js architecture where Python callbacks first apply logarithmic normalization by agricultural land area using the normalize_by_agricultural_land() function, then filter countries with valid agricultural data and contamination rates above 0.1 (adjusted for log scaling) before passing structured JSON data to a clientside callback that loads D3.js and renders the interactive bar chart.

The visual design employs semantic color encoding with a four-tier risk classification system where colors intuitively progress from red (severe > 50%) through orange (high 30-50%) and yellow (moderate 15-30%) to green (low < 15%), creating immediate visual risk assessment without requiring numerical interpretation. The design implements progressive disclosure principles by providing overview ranking at first glance, detailed hover tooltips with country-specific information including monitoring sites and main pollutant types, and contextual animations that respond to filter changes. Accessibility and clarity are ensured through high contrast between background and text, clear country name labeling with 45-degree rotation for readability, and the explanatory legend that explicitly defines the color-coded risk thresholds.

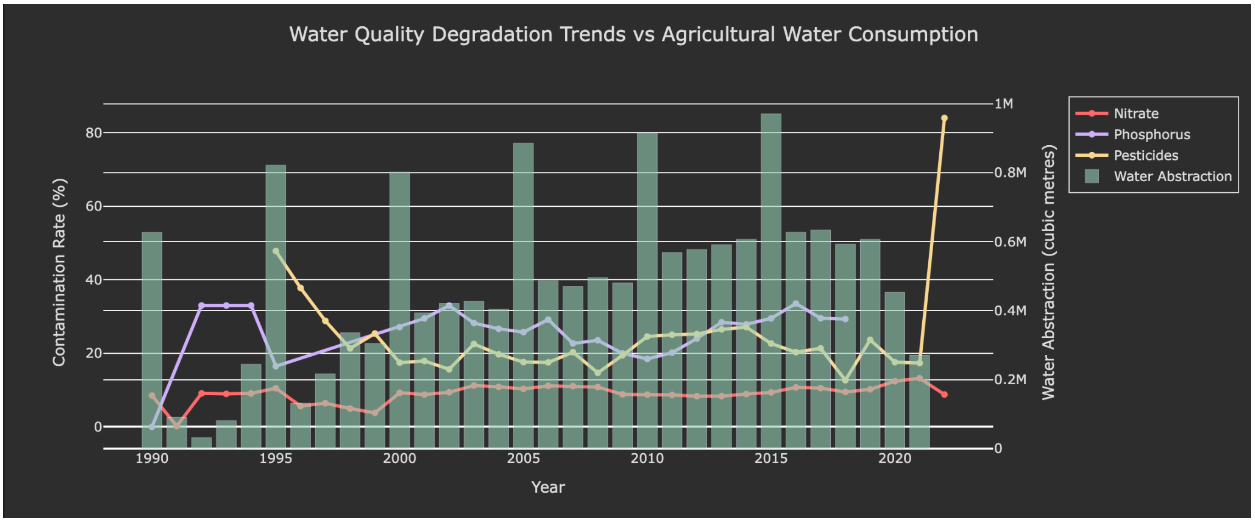


Final Visual Design #1 for Water Dashboard

The second visualization design is a dual-axis visualization where contamination data (nitrate, phosphorus, pesticides) is processed through yearly averaging and plotted as line traces on the primary y-axis, while agricultural water abstraction data is aggregated annually and rendered as semi-transparent bar charts on the secondary y-axis.

The visual design implements layered information architecture through the combination of line and bar chart elements, where the background bars provide contextual water consumption data while foreground lines emphasize contamination trend patterns over the 1990-2020 timeline.

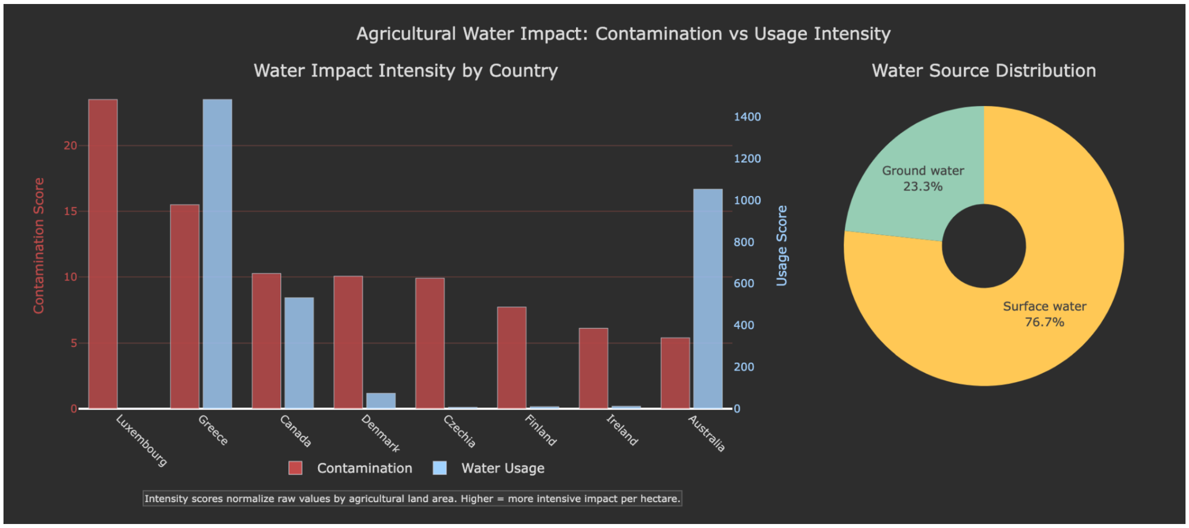
This design also follows dual-encoding principles by using distinct visual channels for related but different data types - temporal line progression for contamination rates versus discrete yearly bars for consumption volumes - enabling correlation analysis between water quality degradation and usage patterns. The unified x-axis approach with coordinated hover functionality ensures temporal alignment between contamination and consumption data, while dual y-axis labeling clearly differentiates percentage-based contamination metrics from volumetric water abstraction measurements, preventing scale confusion while enabling direct visual comparison of sustainability trends.



Final Visualization Design #2 for Water Dashboard

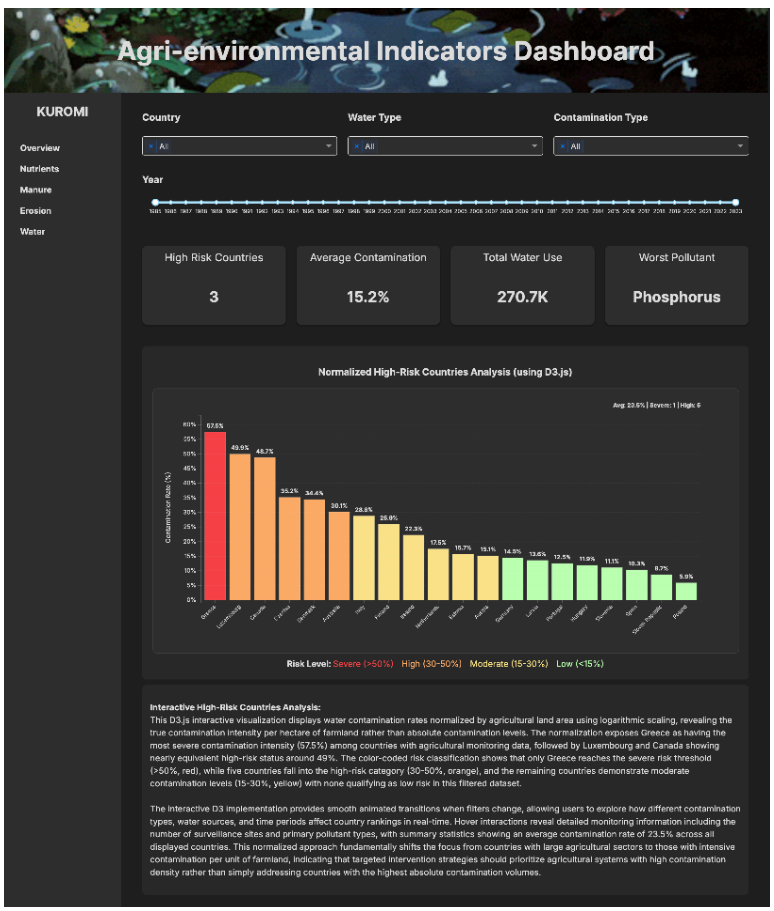
The last visualization for the water dashboard is a dual-panel water impact analysis with a 1x2 layout combining grouped bar charts and pie charts, where contamination and abstraction data undergo logarithmic normalization by agricultural land area through the normalize_by_agricultural_land() function before being merged and filtered to include only countries with complete normalized datasets.

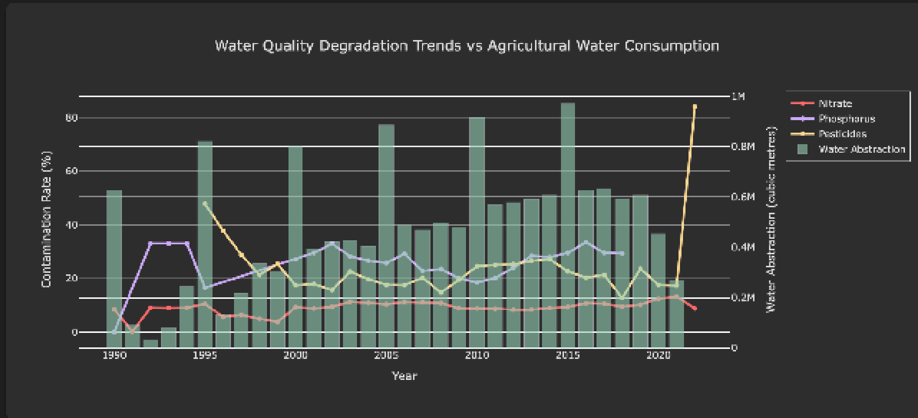
The visual design employs comparative analysis principles through the use of grouped bars with distinct color coding (red bars for contamination scores and blue bars for water usage score) enabling direct side-by-side comparison of normalized intensity metrics across the top 8 countries. Information density management is achieved through the dual-panel approach where the left panel provides detailed country-specific comparisons while the right panel offers contextual overview through a simplified donut chart showing surface water (76.7%) versus groundwater (23.3%) distribution.



Final Visualization Design #3 for Water Dashboard

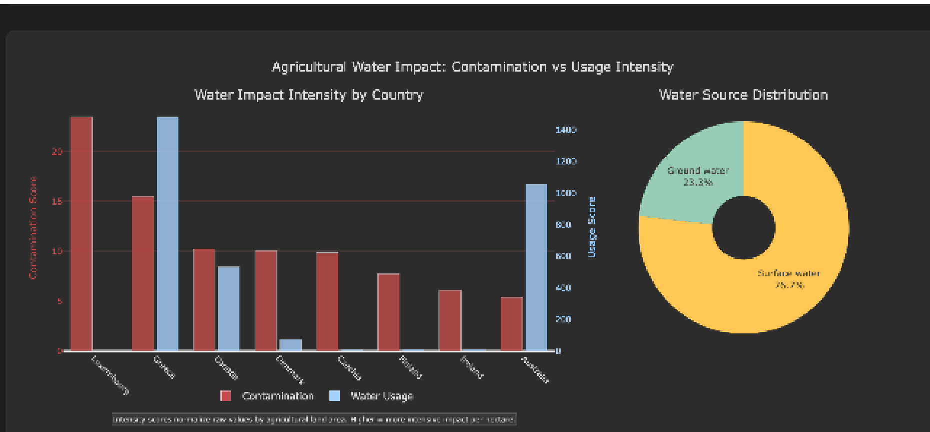
Here is the overall water dashboard with filters, KPI cards, and three final visualizations (including the analysis for each visualization):





Dual-Axis Temporal Analysis with Water Use Correlation:
This visualization shows a concerning disconnect between agricultural water consumption patterns and contamination monitoring that exposes critical gaps in environmental governance over the past three decades. The dramatic spike in pesticide contamination detection around 2020 (reaching nearly 100%) coincides with relatively stable water abstraction levels, suggesting this surge reflects enhanced monitoring capabilities and stricter detection standards rather than a sudden environmental catastrophe - indicating that pesticide contamination was likely severely underreported in earlier decades. The persistence of phosphorus contamination at elevated levels (25-35%) throughout the 2000s and 2010s, despite fluctuating water usage patterns, points to legacy pollution from accumulated fertilizer applications that creates long-term groundwater contamination independent of current agricultural intensity.

Most significantly, the relatively stable nitrate contamination levels (5-15%) against a backdrop of increasing water abstraction from the 1990s through 2015 suggest either effective nitrogen management practices were implemented or that nitrate monitoring protocols remained inadequate to capture the true extent of fertilizer-related contamination. The lack of clear correlation between water consumption peaks (mid-1990s, 2015) and immediate contamination spikes indicates that agricultural water pollution operates on delayed timescales, where current contamination rates reflect historical farming practices rather than present-day water usage, making this visualization a powerful reminder that today's agricultural decisions will determine water quality challenges for decades to come.



Water Quality vs Usage Efficiency Analysis:
The normalized intensity analysis reveals Luxembourg as having the highest contamination score (~22) despite minimal water usage, indicating extremely concentrated pollution per agricultural hectare that likely stems from intensive farming practices rather than excessive water consumption. Greece emerges as a critical case study with high contamination intensity (~15) paired with exceptionally high normalized water usage (approaching 2500 on the usage score), suggesting a dual crisis where both contamination density and water consumption efficiency require immediate intervention. Australia presents an interesting contrast with the highest normalized water usage score but relatively moderate contamination intensity (~5), indicating potential for optimization through improved water treatment and agricultural practices rather than usage reduction.

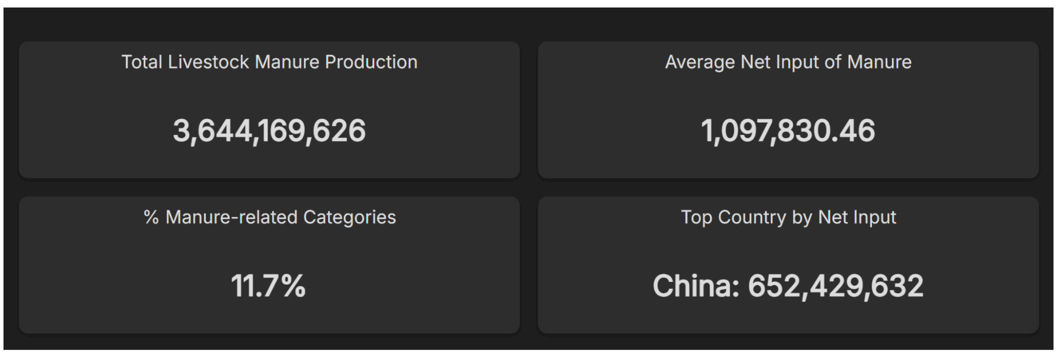
The water source distribution reveals a critical infrastructure insight: 76.7% reliance on surface water versus 23.3% groundwater creates systemic vulnerability to agricultural runoff contamination across the monitored countries. Countries like Denmark, Czechia, and Finland demonstrate similar contamination intensity scores (8-10 range) despite varying water usage patterns, suggesting that contamination issues may be more closely linked to agricultural intensity and source water management than to consumption volume. This dual-axis analysis indicates that effective water quality improvement strategies must address both contamination intensity per hectare and usage efficiency simultaneously, with particular attention to surface water protection measures given the heavy reliance on this more vulnerable water source across agricultural systems.

4. Validation

We evaluated our dashboard through three complementary lenses: *Data Abstraction & Semantics, Design Guidelines & Integrity Principles, and Perceptual & Cognitive Factors*. These lenses ensure every chart communicates OECD IP data both accurately and intuitively.

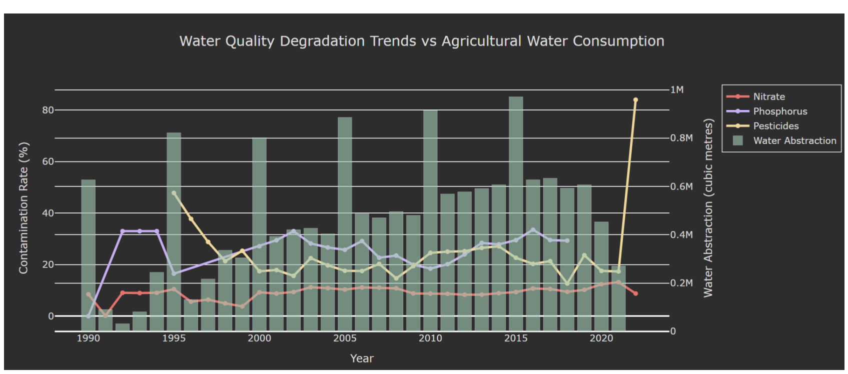
4.1. Data Abstraction & Semantics

Each visualization begins with a clear understanding of what is shown and how it's measured. The KPI Cards present single ratio-scale values, like withTotal Patents card literally the sum of patents in the selected domain and years - using oversized, centered text that matches its singular magnitude.



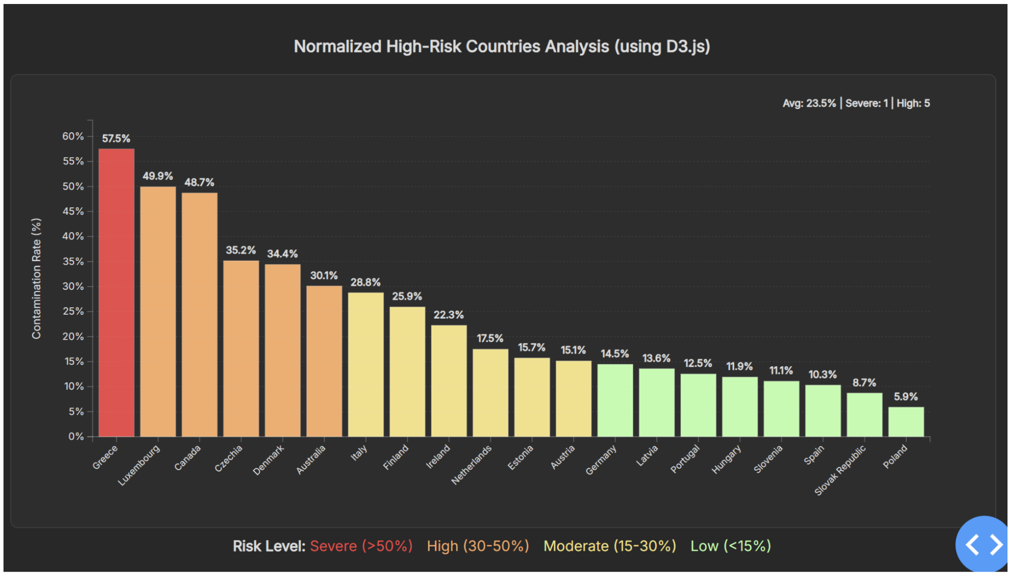
KPI Cards

The Trend Chart plots patent counts over evenly spaced years (ordinal/interval) for the top five countries (nominal), with line height encoding the ratio-scale counts. Similarly, the Globe Chart maps each country's patent total (ratio) onto its geography (nominal) via a sequential color gradient.



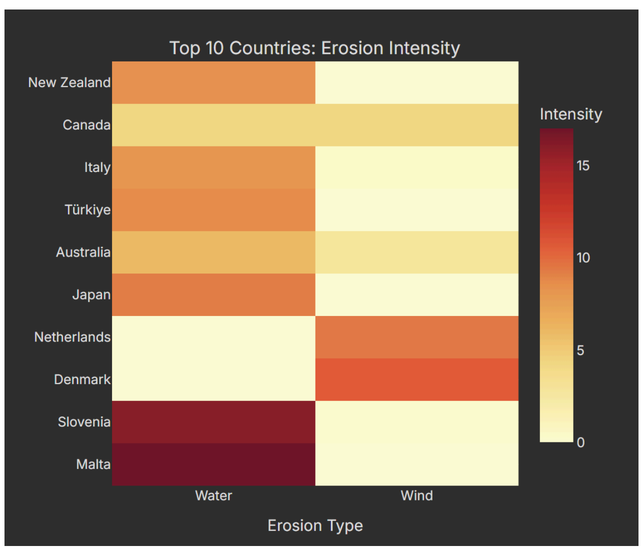
Trend Chart with even spacing and sequential colors

While the Race Chart orders the top ten countries (nominal) by bar length (ratio), supplemented by direct labels.



Race Chart ordering the top ten countries

Finally, the Erosion Intensity Heatmap groups patents by discrete categories (nominal) and uses tile area and color saturation - both ratio channels - to reinforce each category's share.



Heat map with grouping and coloring

4.2. Design Guidelines & Integrity Principles

We applied Tufte’s insistence on no distortion: all bars and lines begin at zero, axes are evenly spaced, and no deceptive scaling is used.

To maximize data-ink ratio, extraneous gridlines and backgrounds were removed from the Race and Treemap views, and the Trend Chart retains only the lightest necessary guides.

Every chart carries explicit labels - axis titles, direct labels on bars and tiles, or a clear legend - to provide context and units. Following Munzner’s critique, we avoided unnecessary 3D effects everywhere except the Globe Chart, where the orthographic projection adds meaningful spatial context.

We also manage cognitive load by favoring static, side-by-side views (Trend and Race) and direct labeling, rather than forcing users to hunt through animated transitions or hidden legends. Our color choices respect Gestalt principles: categorical elements use distinct, color-blind-safe hues, while sequential views employ a single-hue ramp.

4.3. Perceptual & Cognitive Factors

The dashboard leverages pre-attentive features: the Total Card’s large, bold number “pops out” immediately, and the Race Chart’s direct labels allow viewers to read values without scanning to a legend.

Gestalt grouping appears in how bars are sorted by rank and treemap tiles share contiguous areas, reinforcing category relationships. To prevent change blindness, filter updates redraw all views statically - no fading or sliding animations - so users never miss a shift in the data.

Taken together, this systematic, three-lens self-analysis confirms that our dashboard not only looks polished but also empowers users to extract reliable, actionable insights from OECD IP statistics.

5. References

Murray, S & O'reilly Media 2017, Interactive data visualization for the web : an introduction to designing with D3, O'reilly Media, Sebastopol, Ca I Pozostałe.

Environmental Performance of Agriculture in OECD Countries 2025, OECD, viewed 6 August 2025, <https://www.oecd.org/en/publications/environmental-performance-of-agriculture-in-oecd-countries_2679ba38-en.html>.

object Object 2015, 'Overview on agricultural landscape indicators across OECD countries', Core.ac.uk, viewed 6 August 2025, <https://core.ac.uk/reader/29290290>.

Evans, AE, Mateo-Sagasta, J, Qadir, M, Boelee, E & Ippolito, A 2019, 'Agricultural water pollution: key knowledge gaps and research needs', Current Opinion in Environmental Sustainability, vol. 36, pp. 20–27.

Shortle, J.S., Abler, D.G. and Ribaudo, M. (2001). Agriculture and water quality: the issues. Environmental policies for agricultural pollution control, pp.1–18. doi:https://doi.org/10.1079/9780851993997.0001.