MAPPING FOREIGN DIRECT INVESTMENT IN VIETNAM: IMPLEMENTED PROJECTS, SECTOR DISTRIBUTION, & PROVINCIAL TRENDS

1. Overview and Motivation

Over the past few decades, Vietnam has emerged as a prominent destination for Foreign Direct Investment (FDI), with over $524 billion in registered capital across more than 36,000 projects. While this influx of investment has significantly contributed to the country's GDP and export strength, particularly in high-tech and manufacturing industries, FDI remains unevenly distributed, heavily concentrated in a handful of provinces like Bac Ninh, Binh Duong, and Ba Ria–Vung Tau. This regional imbalance raises critical questions about the factors that drive investment location decisions and the broader implications for equitable economic growth. The motivation behind our project is to visualize and analyze these disparities through the lens of not just raw investment data, but also key governance, infrastructure, and socio-economic indicators. By integrating FDI data with public service quality and institutional performance (sourced from PAPI), we aim to uncover patterns that can explain why certain provinces attract more capital than others.

---

2. Related Work

Our project draws inspiration from a mix of public dashboards and visualization practices introduced in CS 571. One of our primary references is the Vietnam Provincial Governance and Public Administration Performance Index (PAPI) dashboard (https://papi.org.vn/du-lieu-papi/), which presents province-level performance across governance dimensions like transparency, service delivery, and citizen participation. While PAPI offers well-organized metrics and filtering options, it lacks visual integration with economic indicators such as FDI, which our project aims to bridge.

We were also influenced by the OECD Regional Statistics Explorer (https://www.oecdregionalwellbeing.org/), which features interactive choropleths and coordinated views that compare regional disparities in economic and development metrics. These platforms exemplify clean visual encoding and intuitive navigation that helped shape the interaction design of our tool.

From class, we were especially influenced by concepts discussed in the Week 7 lecture on Visual Encodings – such as the effectiveness of position and color for quantitative variables – and Week 8's Design Guidelines, including maximizing data-ink ratio, avoiding 3D or dual-axis charts, and maintaining perceptual accuracy. We also referred back to the Five Design-Sheet Methodology

introduced in the project proposal phase to ensure our design process explored multiple alternatives before finalizing our interactive dashboard.

In terms of structure and aesthetics, we also looked at popular data journalism platforms like Information is Beautiful (https://informationisbeautiful.net/), which effectively combine storytelling with clean data visualization. These sources influenced our layout decisions, especially for layering time series and map-based views in a way that maintains visual clarity.

---

3. Questions

- **Starting Point:** Where is Foreign Direct Investment (FDI) going in Vietnam, and which provinces are attracting the most?

- **Midway Question:** Is there a connection between FDI and the quality of local government or public services, as shown by PAPI scores?

- **Final Question:** Can we use visualizations to help explain why some provinces receive less FDI despite having high PAPI scores?

---

4. Data

Sources:

- FIA (FDI statistics) – the problem with the links leading to the FDI numbers is that there are no direct links. However, for each year, there is an Excel file that we collected from 2015 to 2023
- PAPI (governance metrics) – https://papi.org.vn/du-lieu-papi/.

The FDI data was extracted from a multi-sheet Excel file. A major challenge we faced was that each year's data was formatted differently – column headers changed, some sheets had merged cells, and data positions varied. This made it very difficult to directly parse the file with Python. To resolve this, we first manually reformatted each yearly sheet to follow a unified layout, aligning columns like "Province," "Year," "FDI," etc. Once standardized, we used Python and pandas to extract and clean the data:

```python
df_city = pd.read_excel('data/FDI_processed.xlsm', sheet_name='City')
df_extracted = df_city.loc[1:].copy().rename(columns={
    'Unnamed: 1': 'Year',
    'Unnamed: 2': 'Province',
    'Unnamed: 11': 'FDI'
})
df_result = df_extracted[['Province', 'Year', 'FDI']]
df_result.to_csv('data/extracted_data.csv', index=False)
```

We also addressed province name inconsistencies, harmonizing names across both datasets (e.g., converting "TP. HCM" to "Ho Chi Minh City") to ensure proper joins during merging.

The PAPI dataset consisted of yearly governance reports stored in separate sheets. We used a custom parser to detect the location of the "Province" column, extract each of the eight governance dimensions, and group sub-dimensions when necessary. Since some dimensions were split across multiple columns, we used a predefined structure by year to accurately reassemble them:

```python
for sheet_name in process_sheets:
    year = ''.join(filter(str.isdigit, sheet_name))
    df = pd.read_excel(papi_file, sheet_name=sheet_name, header=None)
    # Extract header, group dimensions, clean province entries
    ...
    with open(output_file, 'w') as f:
        json.dump(data, f, indent=2)
```

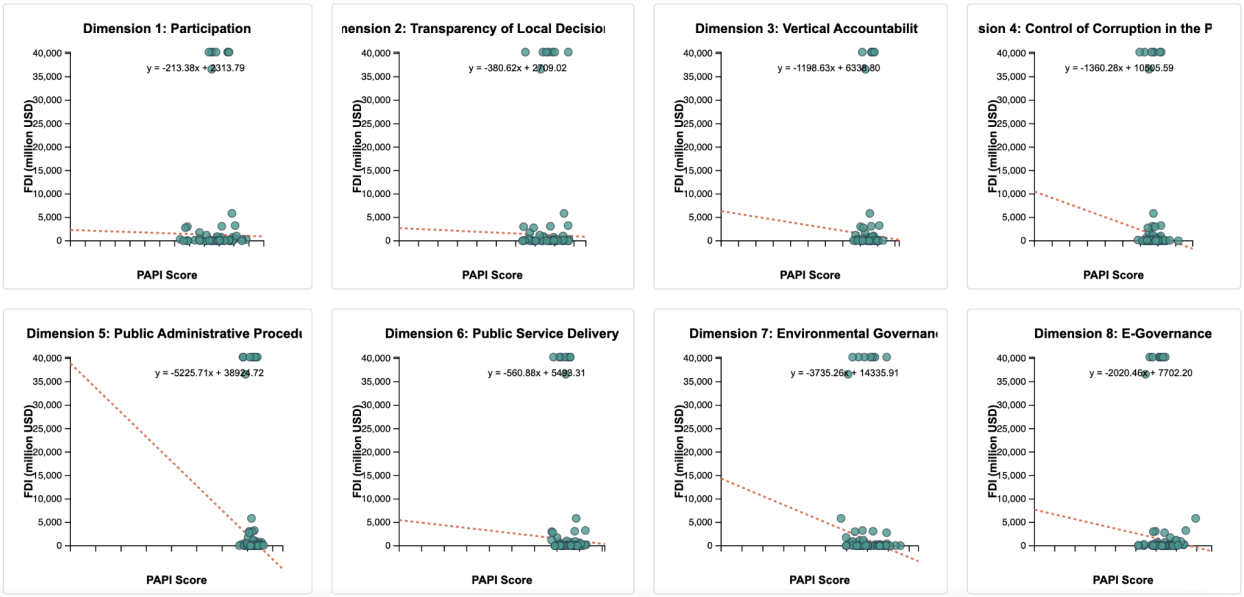
To enable consistent comparisons, we normalized both FDI and PAPI values to a 0–100 scale using this function in script.js:

```javascript
function normalizeValue(value, min, max) {
  if (min === max) return 50;
  return ((value - min) / (max - min)) * 100;
}
```

After processing, we merged the datasets by province and year, enabling interactive filtering, time-series comparisons, and regional governance investment analysis.

5. Exploratory Data Analysis

To better understand our dataset, we began with bar charts to visualize FDI inflows by province. These revealed that a few provinces, such as Ho Chi Minh City, Hanoi, and Ba Ria–Vung Tau, received disproportionately high investment. These outliers distorted the overall scale, making it difficult to assess patterns among lower-tier provinces.



We then used line plots to explore FDI trends over time across provinces. This helped identify regions with consistent growth, irregular fluctuations, or sudden investment surges. Some provinces showed steady inflows, while others experienced sharp rises or drops, which guided our decision to include time-based filtering in the final design.

One key insight was that some provinces with strong governance performance still attracted low FDI, suggesting other factors, such as industrial zoning or investment incentives, may play a larger role. Additionally, we observed that FDI values were highly skewed, whereas PAPI scores were relatively stable across provinces. To surface more meaningful trends and comparisons, we decided to remove or separately visualize extreme outliers, enabling clearer exploration of mid and low-performing regions. These findings directly influenced our visualization approach, including log scaling and interactivity to support detailed, province-level comparisons.

6. Issues Encountered & Design Evolution

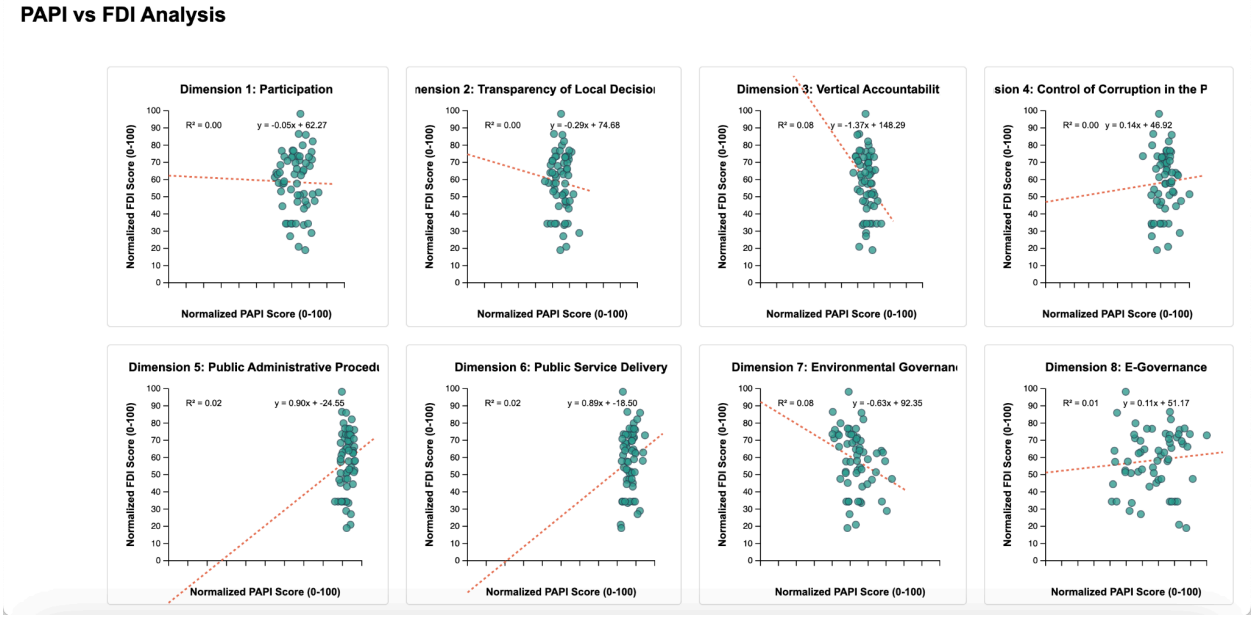
During development, we faced several technical and design challenges. One major issue was overplotting on the choropleth map, where smaller provinces like Bac Kan and Ninh Binh were nearly invisible or indistinguishable, making it difficult to interpret data at a glance. Additionally, our early use of rainbow color gradients caused misinterpretation due to poor luminance contrast and lack of perceptual uniformity, especially for users with color vision deficiencies. Another key challenge was performance lag – the combination of multiple dropdown filters, time data updates, and SVG rendering introduced delays and reduced interactivity responsiveness.

To address these problems, our design went through three major iterations. In Design Iteration 1, we used a simple choropleth map with FDI values and a bar chart shown on hover. This version was functional but too cramped – important details were unreadable, especially for small provinces. In Iteration 2, we restructured the layout to separate the map and chart views. Selecting or hovering over a province dynamically updates the corresponding time series and comparison charts, improving clarity. In the final iteration, we added dropdown filters for year and sector, implemented perceptually uniform color scales, and introduced small multiples to allow side-by-side trend comparisons. These changes improved readability, performance, and the overall analytical value of the dashboard.

We also have some trouble evaluating and presenting potential relationships between governance and investment. We analyzed correlations between PAPI scores (across all eight dimensions) and normalized FDI levels by province. As shown in the scatterplots, most dimensions exhibit very weak or no correlation with FDI ($R2 = 0.00–0.08$). Even dimensions with slightly positive slopes, such as Public Service Delivery and Public Administrative Procedures, still show large variances and no strong linear pattern.

In order to further investigate, we will try log transformations and multilevel modeling, aiming to reduce skewness and account for unobserved province-level effects. However, these approaches also may fail to produce significant associations. One reason may be the compressed range of PAPI scores (most provinces score between 40–60), compared to the much wider spread of FDI values. Additionally, FDI may be influenced more heavily by infrastructure, historical zoning

policies, or strategic geography than by governance metrics alone.



---

7. Implementation

Our project was built using D3.js for interactive visualizations, HTML/CSS for layout and styling, and Python (pandas, NumPy) for data preprocessing and transformation. The final product is deployed via GitHub Pages for public access and ease of version control.

The application uses a modular architecture with three main components:

- Choropleth Map: Displays FDI distribution across provinces. Provinces are colored based on FDI values, with tooltips showing detailed information on hover. Clicking a province updates other views. (Figure 1)
- Control Panel: Located at the bottom of the map, includes filter by year. Filters trigger updates across all linked visualizations. (Figure 1)
- Dynamic Plots: Line and bar charts from 2015 to 2024. They show temporal trends and province-specific performance. (Figure 2a-c)

Key features implemented:

- Hover tooltips for province-level FDI stats.
- Dropdown filters for years with smooth event handling.
- Animated transitions to highlight year-to-year changes when users switch filters.

- Responsive layout using flex-box to support various screen sizes, ensuring usability on desktops and tablets.

---

8. Evaluation

Our data analysis reveals that (1) the top five provinces account for over 50% of total FDI, raising important questions about regional disparities. This concentration suggests that PAPI scores at the provincial level may not fully explain investment behavior. It may be more insightful to aggregate provinces into broader regions to uncover higher-level patterns. (2) Additionally, while transparency and infrastructure indicators showed relatively stronger correlations with FDI compared to income levels, the relationships remain unclear, indicating the need for more robust re-modeling or alternative data features.

Several limitations were identified that can inform future improvements. (1) First, our current methodology lacks clarity on how provinces are compared within a region, which limits interpretability – this should be addressed with an updated regression framework. (2) Second, our dataset does not yet account for the sectoral composition of FDI; with 19 different sectors represented, we plan to introduce a bar chart (including sectors with negative net inflows) to capture this dimension. (3) Finally, the current filtering logic can be overwhelming for users. We aim to streamline the interface to present information more intuitively, while also stabilizing the map interaction to ensure a smoother and more reliable user experience.

To address several design and methodological limitations, we introduced targeted improvements: (1) Initially, our platform lacked clarity on how provinces were being compared within a region. This made comparison charts difficult to interpret. To solve this, we added region-level highlighting—when users hover over a province, all other provinces within the same region are outlined in a distinct color. This visual grouping provides important geographic context and improves the interpretability of comparisons. The logic was implemented using the following code:

```javascript
function highlightFeature(e) {
  const layer = e.target;
  layer.setStyle({ fillColor: "#264653", fillOpacity: 0.7 });

  const provinceName = layer.feature.properties.Name;
  const region = findRegionForProvince(provinceName);
```

```
  if (region) {
    const regionFeatures = [];
    geojsonLayer.eachLayer(l => {
      if (vietnamRegionsEnglish[region].includes(l.feature.properties.Name)) {
        regionFeatures.push(l.feature);
      }
    });

    if (regionHighlightLayer) {
      map.removeLayer(regionHighlightLayer);
      regionHighlightLayer = null;
    }

    regionHighlightLayer = L.geoJSON(
      { type: "FeatureCollection", features: regionFeatures },
      {
        style: {
          color: "#e76f51",
          weight: 1.5,
          fill: false
        },
        interactive: false
      }
    ).addTo(map);
  }
}
```

---

9. Future Work

To enhance the analytical depth and accessibility of the platform, we plan to incorporate several key features. These include integrating related news articles that correspond with major shifts in FDI to provide real-world context behind the data. Having sentiment analysis from investor news sources would allow users to assess market outlooks beyond raw numbers. To support domestic audiences, we also aim to develop a multi-language user interface, starting with Vietnamese.

From a design and interactivity standpoint, several improvements can strengthen usability. For instance, making the choropleth map clickable would allow users to access a sector-level FDI breakdown for each province, providing more granular insights into regional economic strengths. Furthermore, while hover-triggered comparison charts are currently implemented, users unfamiliar with Vietnam's geography may find them difficult to interpret. To address this, we

propose visually grouping provinces within the same region when hovered – using distinct background colors or borders – to provide clearer spatial context and enable more intuitive comparisons. These enhancements will deepen user engagement and make the platform a more powerful tool for exploration and analysis.

Appendix

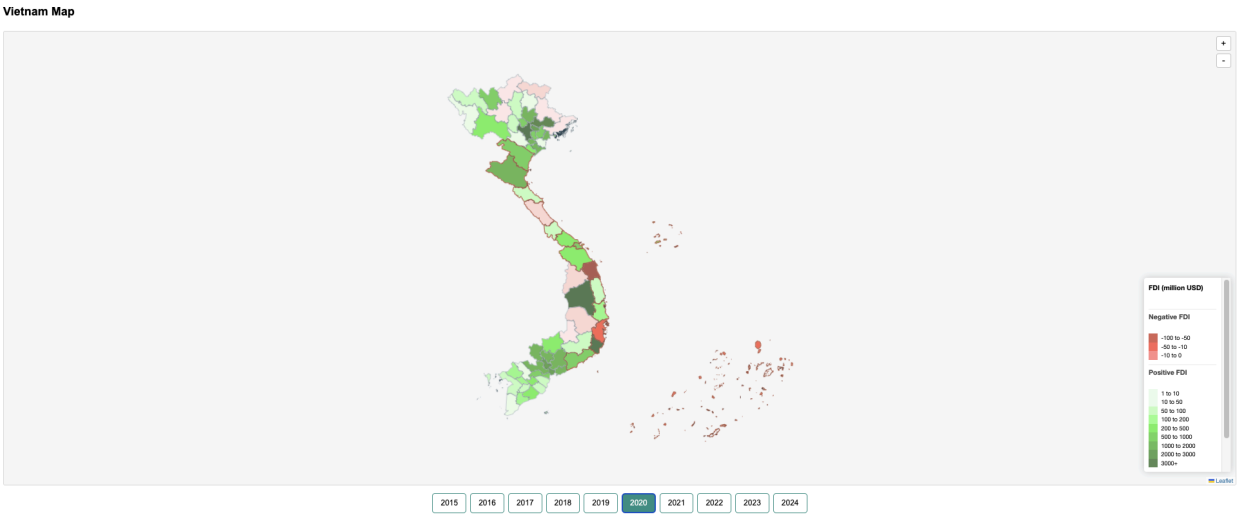


Figure 1. Choropleth Map of FDI by Province in Vietnam (2020)

This figure shows a choropleth map of Vietnam representing the distribution of Foreign Direct Investment (FDI) across provinces in the year 2020. Provinces are color-coded based on the amount of FDI received, ranging from negative inflows (shaded in red) to high positive inflows (shaded in darker green). The map highlights significant regional disparities, with a small number of provinces attracting the bulk of investment while several others experienced little to no FDI, or even net outflows. A legend on the right provides detailed investment ranges in million USD.

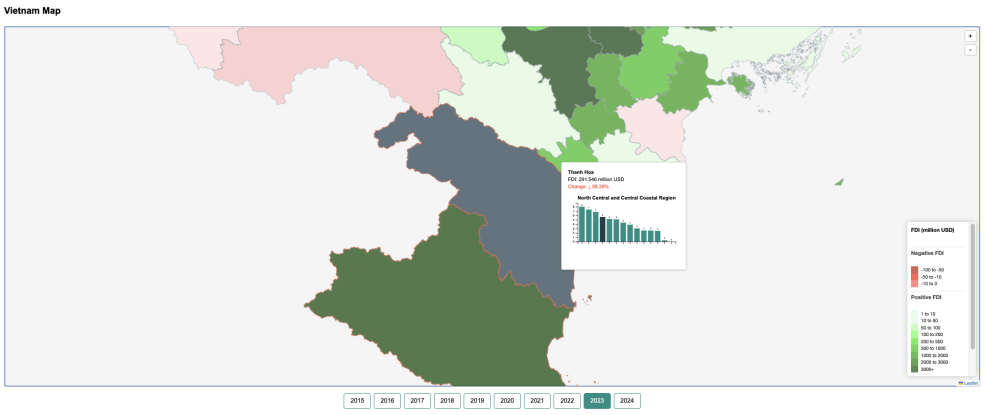


Figure 2a.

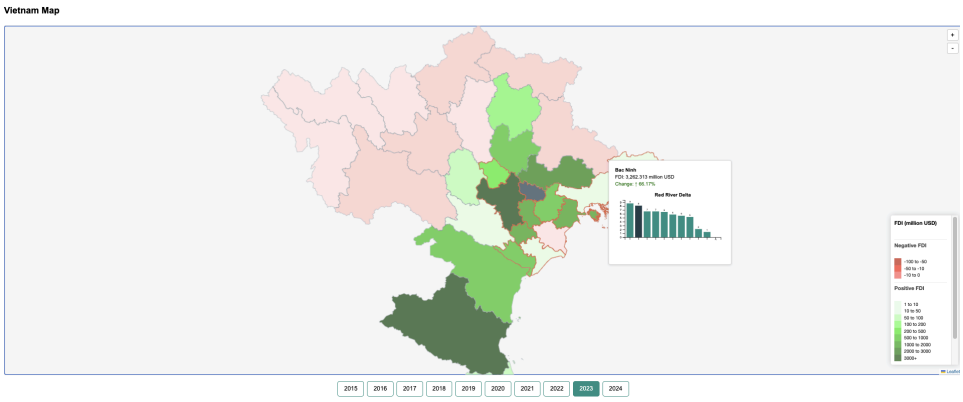


Figure 2b.

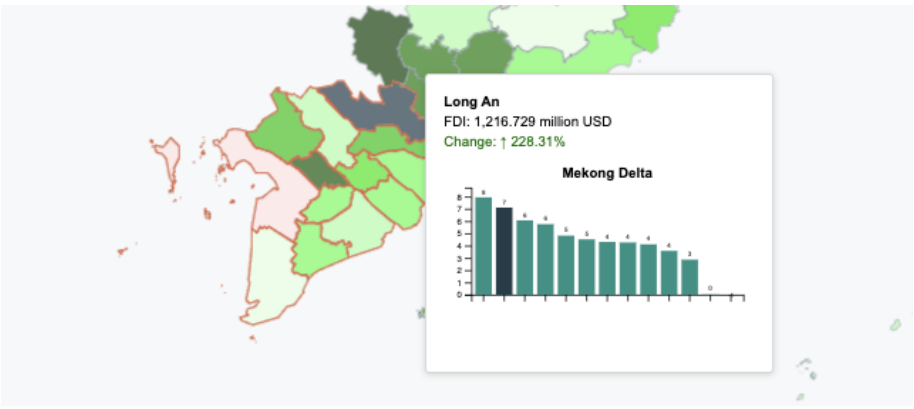


Figure 2c.

Figure 2a–2c. Interactive Map Displaying Provincial FDI and Regional Comparisons

These figures present an interactive choropleth map of Vietnam's provinces in a specified year, where users can explore Foreign Direct Investment (FDI) data by selecting individual provinces. Upon interaction, a popup displays the total FDI value, year-over-year change, and a regional bar chart comparing FDI across neighboring provinces within the same region. This design allows users to gain both detailed local insights and a broader regional context. The color scale visually represents the range of FDI inflows, from negative (red) to high positive (green), helping users quickly identify geographic investment patterns.

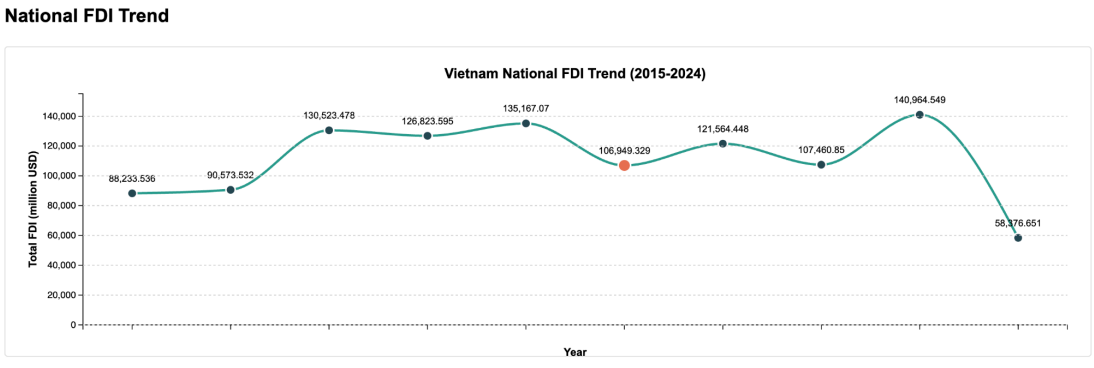


Figure 3. National FDI Trend in Vietnam (2015–2024)

This line chart presents the annual Foreign Direct Investment (FDI) inflows into Vietnam from 2015 to 2024. The data shows steady growth with notable peaks in 2019 and 2023, reaching over 135 and 140 billion USD, respectively. A dip in 2020 reflects

the global economic impact of the COVID-19 pandemic. The sharp decline in 2024, with FDI recorded at approximately 68.8 billion USD, represents only the first six months of the year and does not reflect the full annual total. This chart offers a macro-level view of FDI fluctuations and growth trends over the past decade.