TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

Group 10 - 4CSD

Baylon, Karyle Zhienelle V.

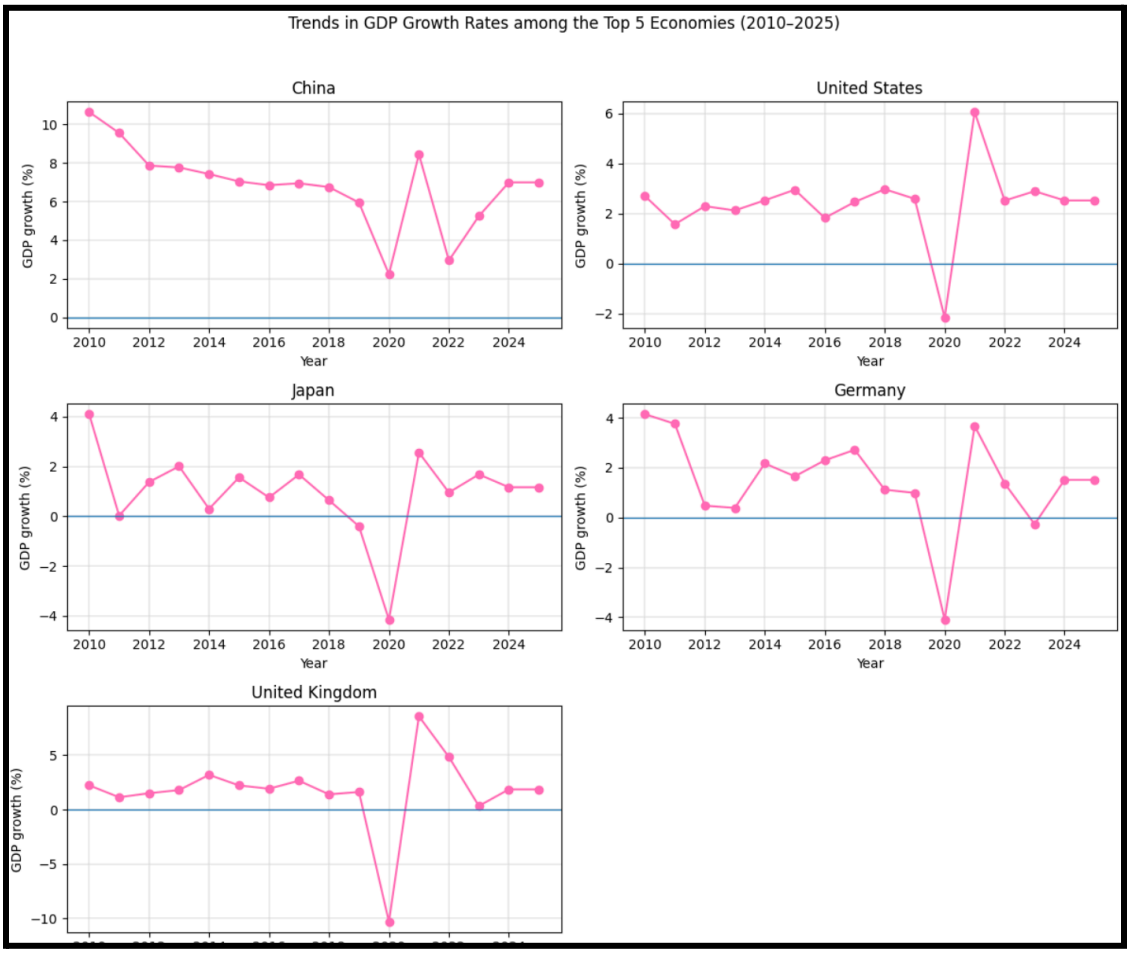
Francisco, Leann Joy Y.

Magtanong, Ralph Daven M.

Use Python to analyze time series and distribution patterns in a dataset. You'll explore trends over time and variability in key variables, choose appropriate chart types, and critically evaluate your visualizations.

QUESTIONS:

1. **How have GDP growth rates of the top 5 economies evolved over time?**



TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

Chart type: *Small Multiples*

Encoding:

- **X-axis:** Year (2010–2025, continuous time scale)
- **Y-axis:** GDP growth rate (%) (quantitative)
- **Color:** Same color for all countries
- **Facet:** Country (each country in its own subplot)

JUSTIFICATION: **Small multiples** are effective as they allow us to view each country's growth trend separately while keeping scales consistent for fair comparison. This approach reduces clutter compared to overlaying multiple lines in one graph and highlights the unique trajectory of each economy. Presenting them side by side also supports easy cross-country analysis, helping us see both common shocks (like 2020 downturns) and country-specific differences.

- Main takeaway in one sentence
    - China showed the strongest and most stable growth trend, while the U.S., Germany, Japan, and the U.K. displayed more volatility, especially sharp declines around 2020.

- One design decision and its benefit
    - I used a consistent y-axis scale across all panels. This prevents misleading impressions that could occur if each panel had its own automatic scaling, where small fluctuations might look exaggerated in some countries compared to others. Keeping the scale the same ensures that comparisons across countries are fair and accurate, since the visual size of growth swings is directly comparable.

TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

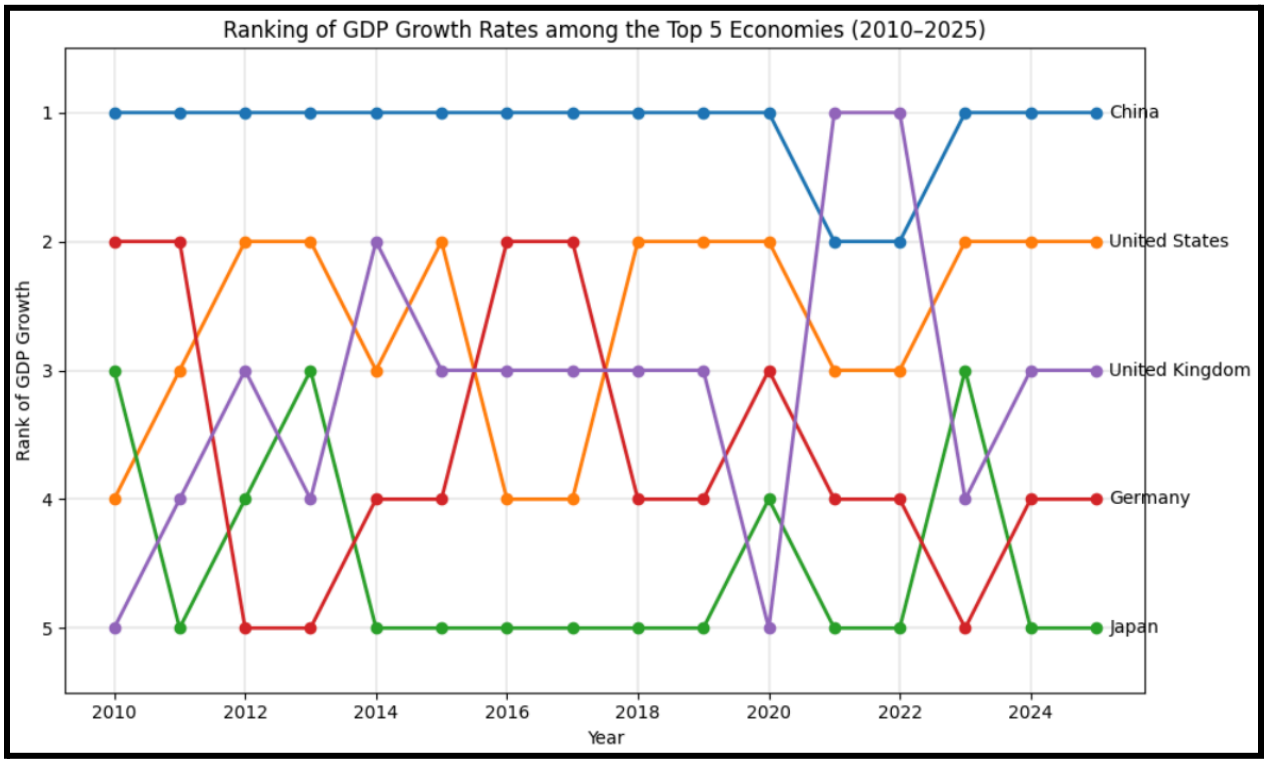


Chart type: *Bump Chart*

Encoding:

- **X-axis:** Year (2010–2025, continuous time scale)
- **Y-axis:** Rank of GDP Growth (1 = fastest growth) (ordinal)
- **Color:** Country
- **Facets:** Year (2010–2025)

JUSTIFICATION: **Bump charts** are useful because they emphasize the relative ranking of GDP growth among countries over time rather than the raw values. This makes it easy to identify which nations consistently lead in growth and when shifts in positions occur. By focusing on rank, we can quickly detect patterns of dominance, competition, or decline without being distracted by the exact percentage levels.

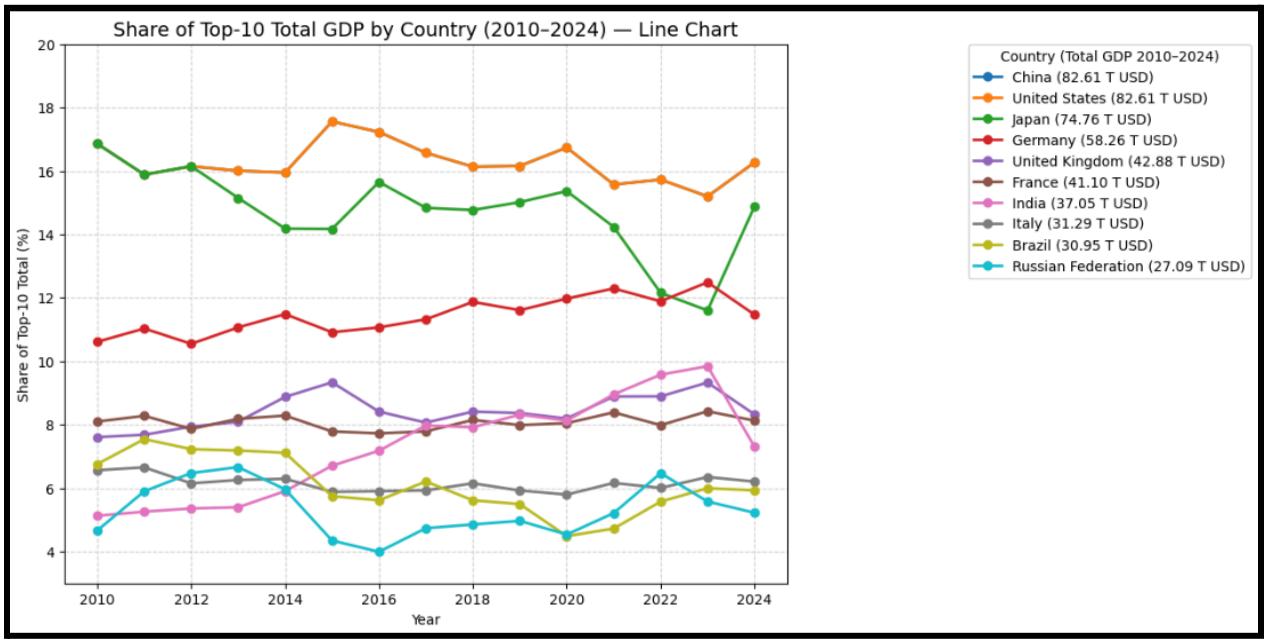
TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

- Main takeaway in one sentence
    - China consistently ranked first in GDP growth among the top five economies, while the other countries frequently shifted positions between 2nd and 5th place.

- One design decision and its benefit
    - I chose to invert the y-axis so that rank 1 is at the top. This makes the highest-ranking country feel visually dominant, which matches how people naturally interpret rankings, with the “best” placed above the rest. Without this adjustment, the chart might look counterintuitive because rank 1 would appear at the bottom.

2. **“Which countries contributed the most to total GDP from 2010 to 2024, and how did their contributions change over time?”**



TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

Chart type: *LINE CHART*

Encoding:

X-axis: *Year (2010–2024, continuous time scale)*
Y-axis: *Share of Top-10 Total GDP* (quantitative)
Color: *Country (Top 10 by total GDP contribution)*
Marks: *Lines with circular markers to emphasize yearly values*

JUSTIFICATION: The line chart is best suited for showing **how contributions evolve over time**, since it highlights changes in each country's share of global GDP among the top 10. Instead of using a full 0–100% scale, the axis was **restricted to 3%–20%**, which magnifies the subtle but important differences between countries and makes the visualization much clearer and more interpretable.

- Main takeaway in one sentence
  - From 2010 to 2024, the United States and China consistently dominated the top-10 GDP shares, while other countries maintained smaller but distinguishable contributions

- One design decision and its benefit
  - I included **total GDP contributions in the legend labels** (e.g., "United States (300.45 T USD)"), which provides immediate context for each line and helps viewers connect the relative percentage shares with absolute economic size, enhancing both comparability and interpretability. Also, restricting the y-axis to **3%–20%**, the chart balances both relative trends and absolute proof, enabling viewers to quickly compare countries without losing detail in the lines.

TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

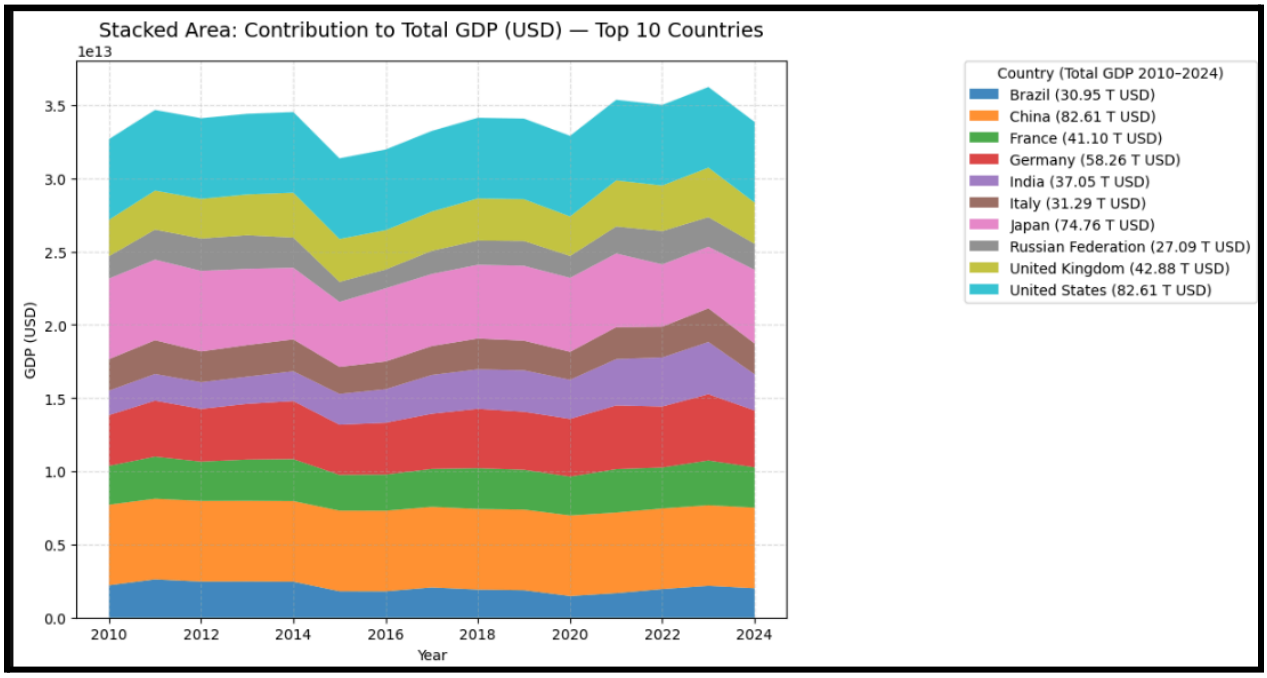


Chart type: *Stacked Area Chart*

Encoding:

- ***X-axis:*** *Year (2010–2024, continuous time scale)*
- ***Y-axis:*** *Total GDP (USD)* (quantitative)
- ***Color:*** *Country (Top 10 by total GDP contribution).Ih ang bangis*

JUSTIFICATION: This chart is ideal because it simultaneously shows the overall growth in global GDP and each share of the top contributor over time. It makes it easy to understand and visualize which countries' economies dominate the total and how their contribution evolves across years.

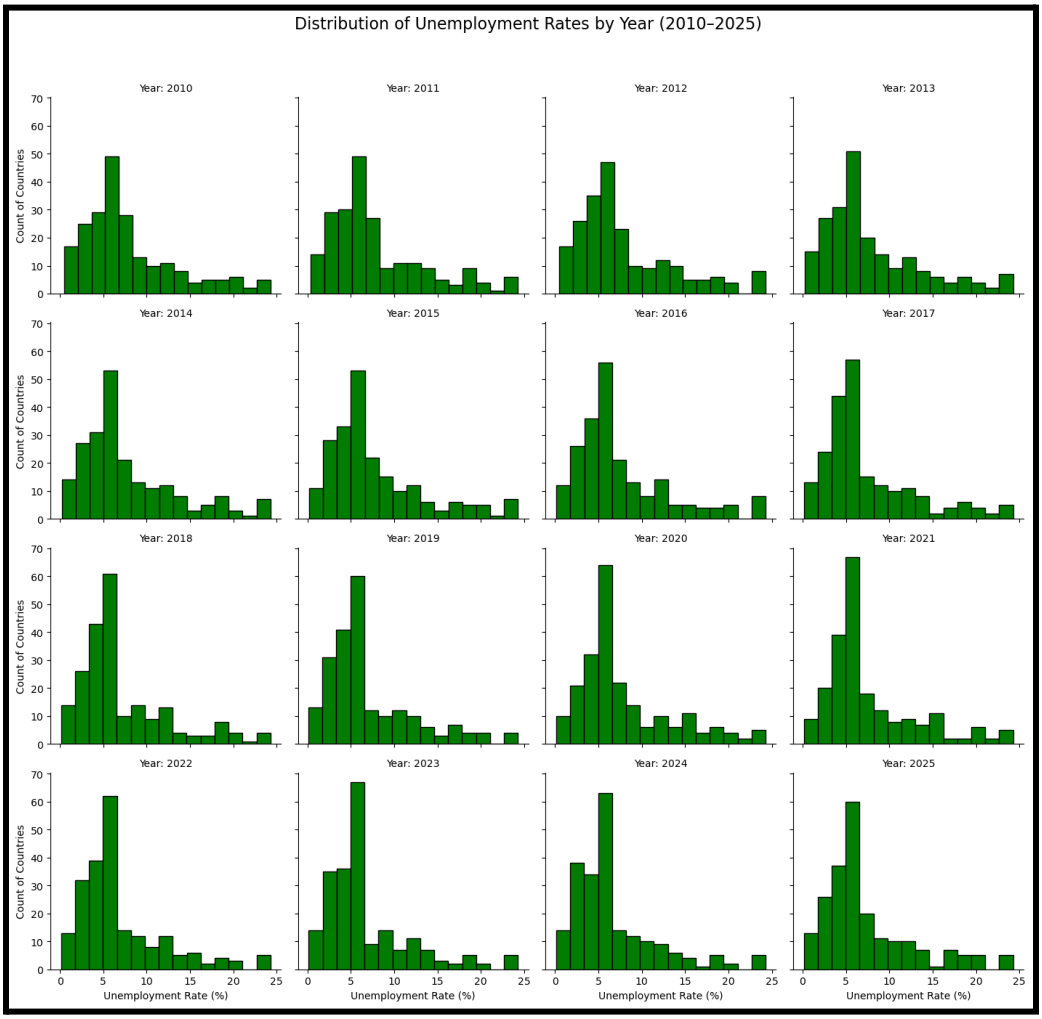
- Main takeaway in one sentence
    - From 2010 to 2024, the top 10 economies consistently made up the largest share of global GDP, with US and China maintaining the most dominant share and contribution.

TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

- One design decision and its benefit
    - I added **country labels with their total GDP contribution in trillions** directly in the legend; this enhances clarity by letting viewers immediately see both the relative share in the chart and the absolute magnitude across the full period, making the visualization both comparative and quantitative.

3. What is the distribution of the unemployment rate per year across countries?



TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

Chart type: *Histogram*

Encoding:

- **X-axis:** Unemployment Rate (%) (quantitative)
- **Y-axis:** Count of Countries (quantitative)
- **Facets:** Year (2010–2025)

JUSTIFICATION: ***Histograms*** are useful considering they illustrate not just the data's shape but also its average, enabling us to determine whether unemployment is dispersed, skewed, or clustered around specific levels. We can also monitor if the distribution gets tighter or wider over time by breaking it up into yearly panels, which provides information on consistency and variability across nations.

- Main takeaway in one sentence
    - The vast majority of countries have unemployment rates between 3 to 8%, though their distribution differs slightly from year to year with slight changes in the tail.

- One design decision and its benefit
    - Rather than putting all the years into a single plot, I decided to facet the histogram by year. This choice avoids states with exceptionally high unemployment rates in particular years from visually controlling the entire chart and facilitates cross-temporal comparisons. On top of that, it also provides a better understanding as to whether unemployment distributions are constant or changing between 2010 until 2025.

TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

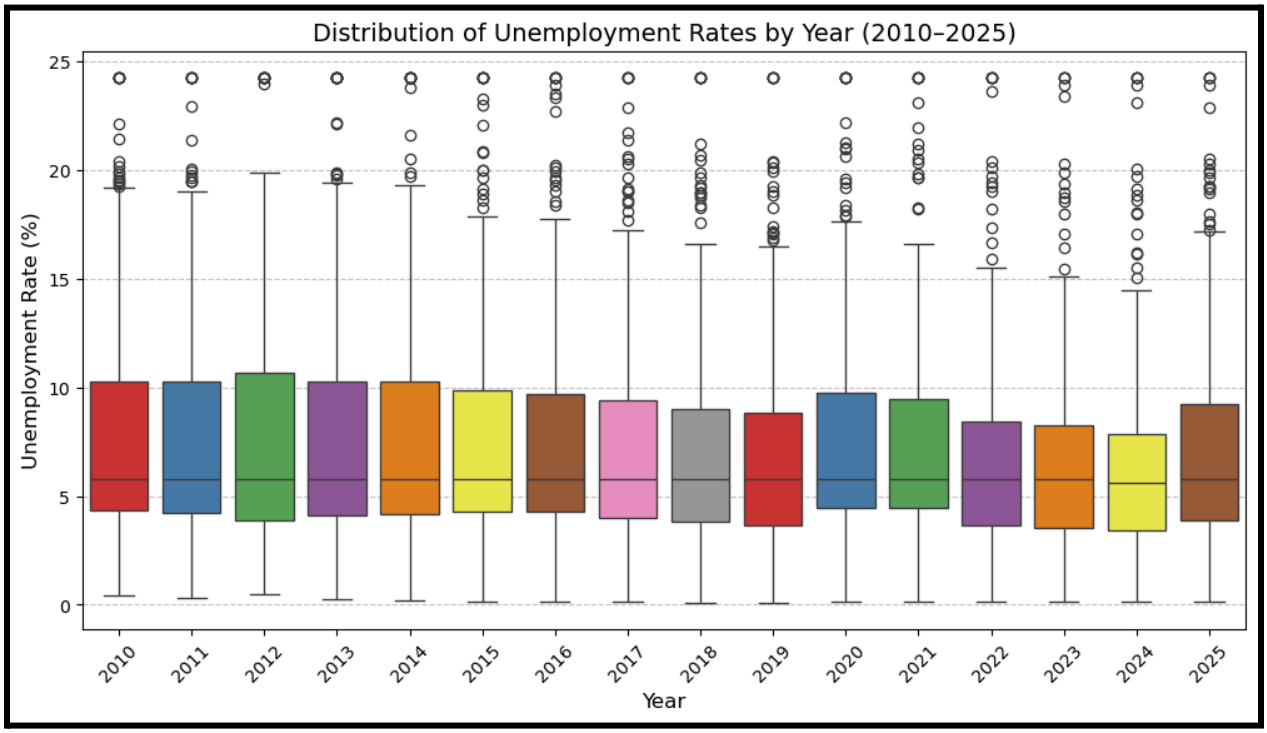


Chart type: *Boxplot*

Encoding:

- **X-axis:** Year (2010–2025)
- **Y-axis:** Unemployment Rate (%) (quantitative)
- **Color:** Different colors for each year's box for separation

JUSTIFICATION: The unemployment rate for each year is easily expressed by ***boxplots***, which highlight extreme outliers and display medians and quartiles. This greatly facilitates comparing changes in the spread from year to year and identifying years with abnormally high spikes in unemployment. This type of graph makes the data simpler so that trends and extremes are easily identified.

TIME SERIES & DISTRIBUTION ANALYSIS USING PYTHON - LAB ACT 4

DATA ANALYSIS AND VISUALIZATION (CS ELEC 3C)

- Main takeaway in one sentence
    - From 2010 to 2025, the median unemployment rate stayed relatively constant at 5–6%, although extreme outliers exceeding 15–20% are regularly observed in many countries.

- One design decision and its benefit
    - Instead of using the same colors for every year's boxplot, I added distinctive color combinations. This keeps the graphs from merging and helps the reader quickly distinguish between years especially when scanning the chart. Moreover, it also makes patterns of consistency and variation more noticeable.