**TRANSBORDER FREIGHT DATA ANALYSIS**

Ebenezer Omari– omarieben7@gmail.com

### **Introduction**

This document outlines the analysis conducted on the TransBorder Freight data provided by the Bureau of Transportation Statistics (BTS).

The primary objectives were to:

* uncover freight movement patterns.
* identify inefficiencies.
* analyze environmental impacts.
* safety and risk assessment.
* economic disruptions.
* propose actionable recommendations to improve cross-border freight transportation.

### Key Analytical Questions

1. **What are the top modes of transportation by freight volume and value?**
2. **How does freight movement vary across states and countries?**
3. **What is the trend of freight charges over time?**
4. **Which states contribute the most to freight weight ?**
5. **Which country contribute the most to freight value**
6. **What patterns exist between shipment weight and charges?**
7. **What is the correlation between Freight Value, Shipment Weight, and Freight Charges for each year from 2020 to 2024?**
8. **How has the total freight value changed over the years from 2020 to 2024?**

### **CRISP-DM Framework**

The analysis followed the CRISP-DM methodology, which includes the following stages:

### 1. Business Understanding

The objectives were defined and the above analytical questions were formulated to guide the analysis process.

### 2. Data Understanding

The dataset consisted of 5,242,875 records and 14 variables. Key variables included:

* TRDTYPE: Transaction type.
* USASTATE, MEXSTATE, CANPROV: Regions involved in freight movement.
* VALUE: Value of goods in USD.
* SHIPWT: Shipment weight.
* FREIGHT\_CHARGES: Associated transportation costs.
* MONTH, YEAR: Temporal attributes.

The dataset covered multiple years and regions, with data on road, rail, air sand water transport modes.

Summary statistics across the dataset (2020 – 2024) revealed:

* **Total Value:** $14.57 trillion USD
* **Total Shipment Weight:** 6.2 trillion kilograms
* **Total Freight Charges:** $191.93 billion USD

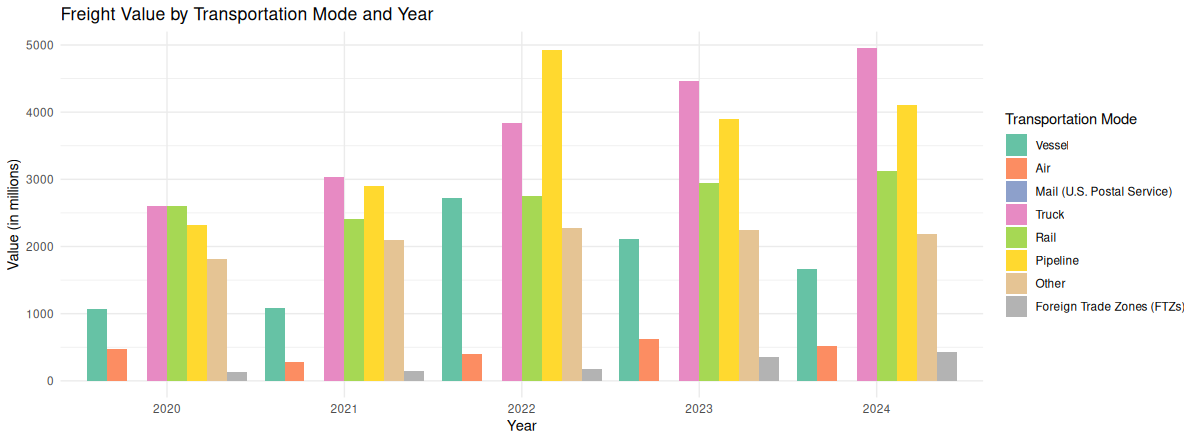
### 3. Data Preparation

* Empty values in columns such as MEXSTATE, CANPROV, DF were identified and addressed.
* The data was cleaned and aggregated to provide meaningful insights at state, country and mode-of-transport levels.
* Converting categorical and numerical variables to appropriate data types.

### 4. Analysis and Visualization

#### Key Findings from the analytical questions:

#### ****What are the top modes of transportation by freight volume and value?****

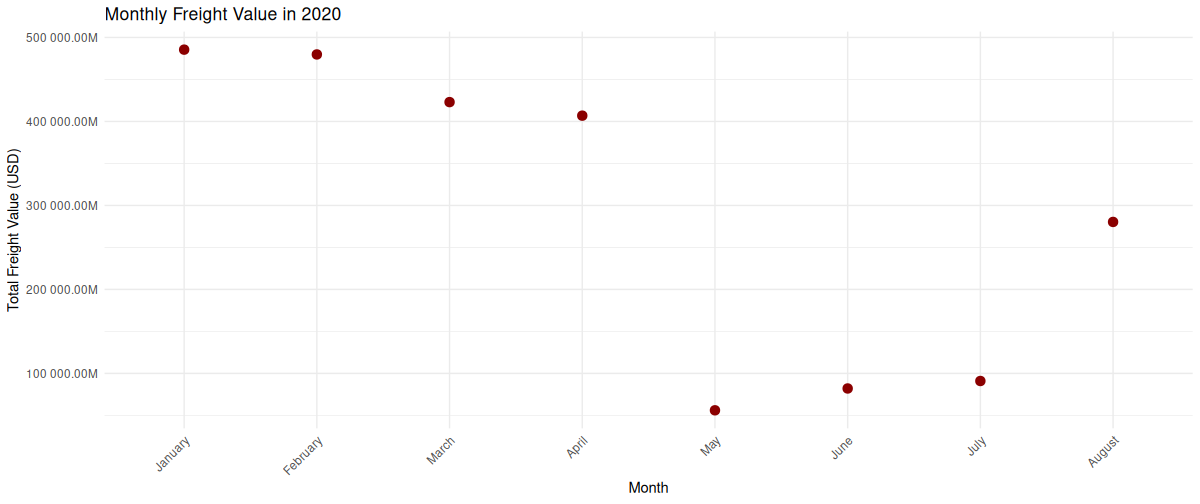


The bar chart below shows the distribution of freight value (in millions of USD) by transportation mode for each year. Each mode is represented by a distinct colour:

**Key Observations:**

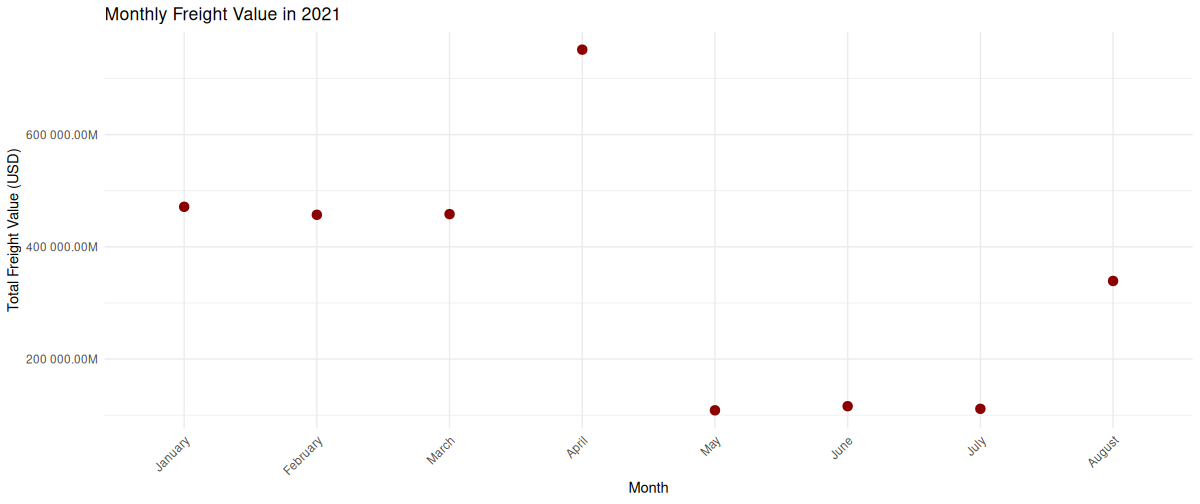
1. Trucks dominate the freight value across most years, indicating their importance in domestic and regional transportation.
2. Air transportation, while less significant in terms of volume, consistently contributes a notable portion of the freight value, highlighting its role in high-value shipments.
3. Pipelines and vessels also contribute significantly, particularly for specific years where bulk commodities like oil or goods with high shipping requirements were involved.

#### ****What is the trend of freight charges over time?****



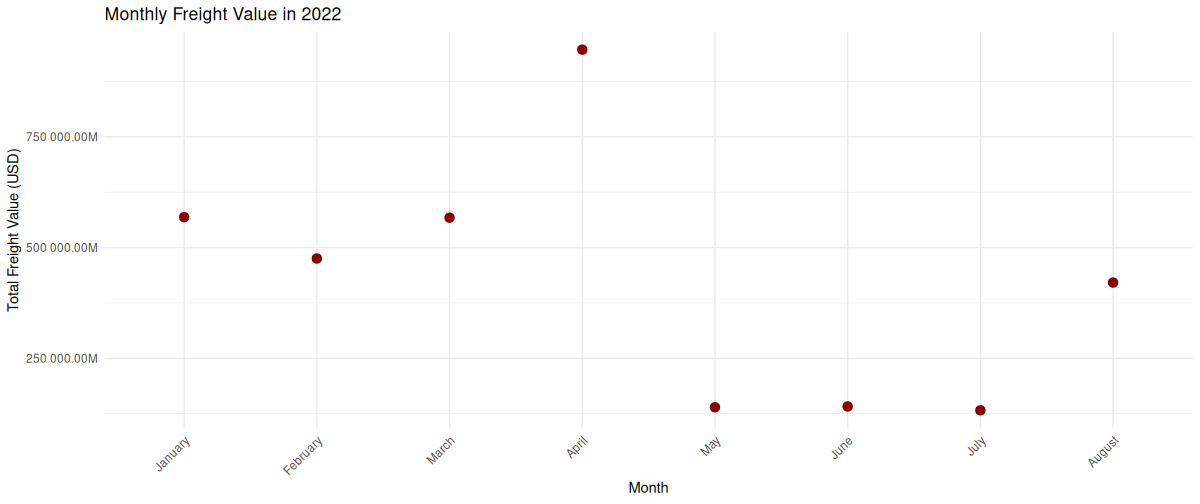
**Analysis for 2020:**

* **High Point:** Freight value peaks in **January**, likely due to increased shipment demand before global COVID-19 lockdowns.
* **Low Point: Freight value dips in May, corresponding to the start of significant international shipping restrictions.**



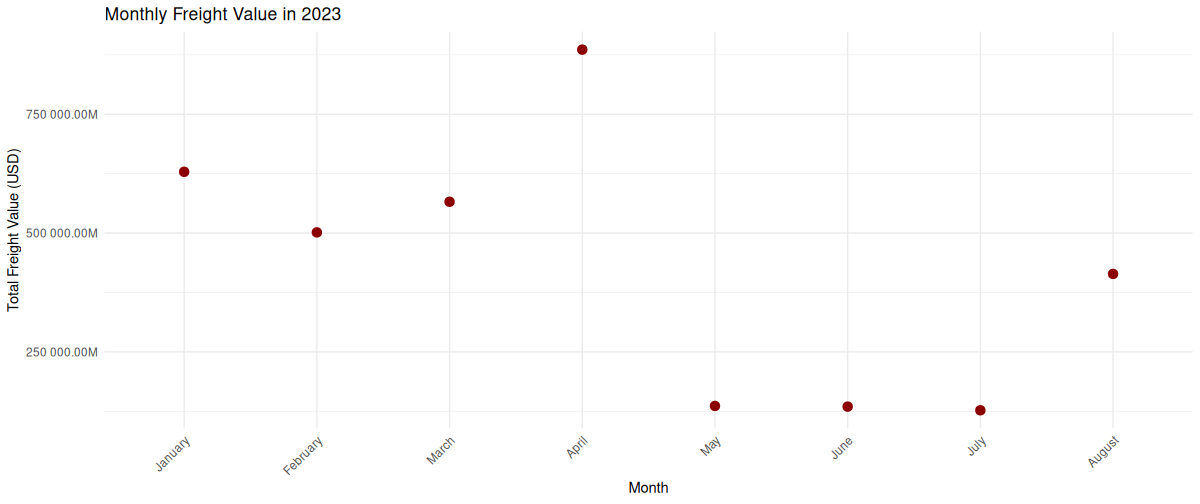
**Analysis for 2021:**

* **High Point: April** has an unusually high freight value.
* **Low Point: **May, June, July** recorded low freight values.**



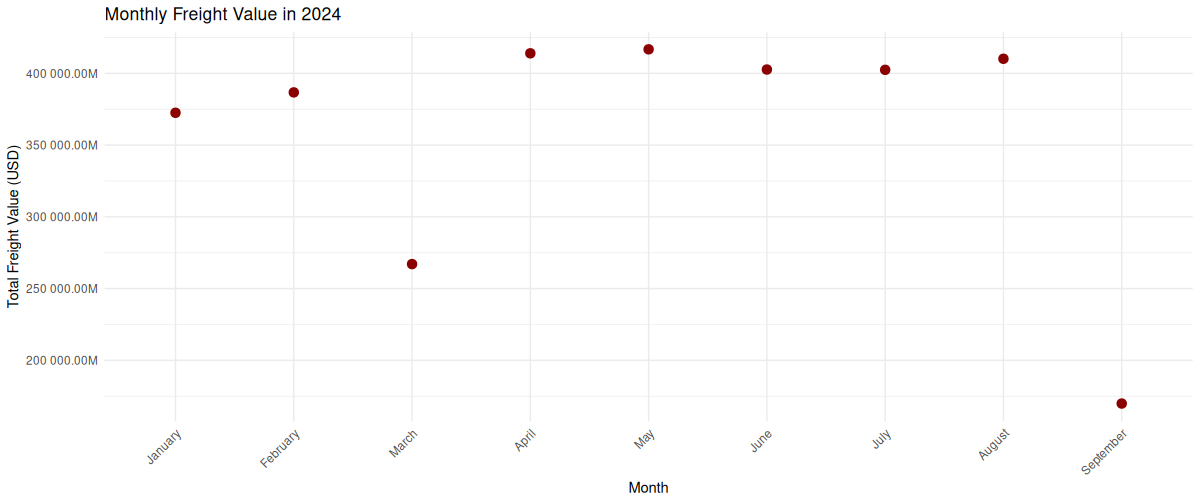
#### ****Analysis for 2022:****

* **Observation:** **May** exhibits a significant rise in freight value, possibly due to easing pandemic restrictions and a rebound in global trade.
* **Low Point:** ****May, June, July** recorded low freight values,** which might reflect the continuation of supply chain disruptions from 2021.



#### ****Analysis for 2023:****

* **Observation:** **April** has a high freight value, possibly indicating strong summer trade activity.
* **Low Point:** ****May, June, July** recorded low freight values.**



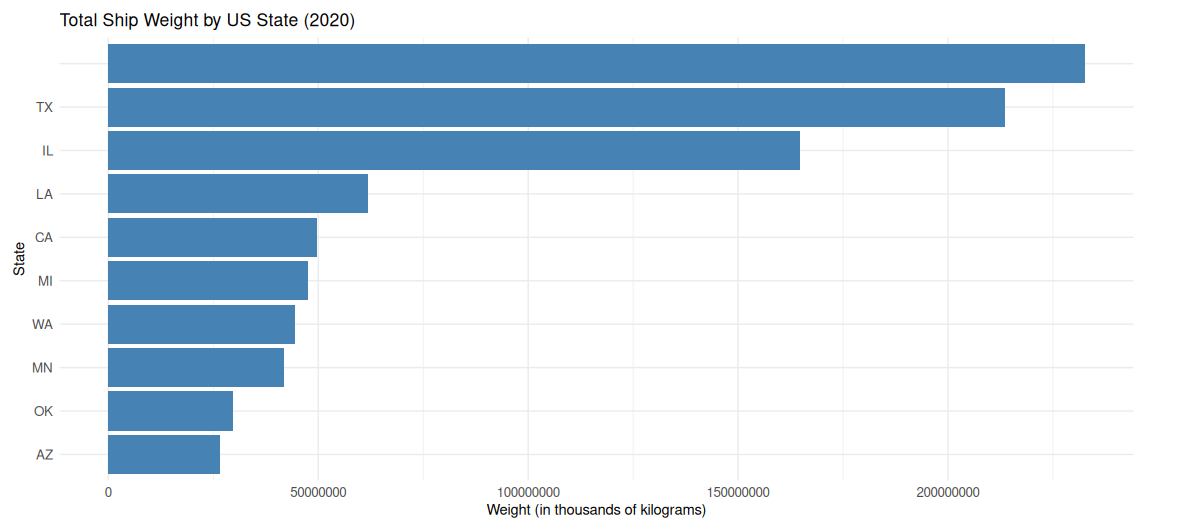
#### ****Analysis for 2024:****

* **Observation:** **April, May, June, July, August** sees a notable increase.
* **Low Point:** Freight values are relatively low in **September.**

**The data reveals that freight values are influenced by seasonal patterns, economic events and unexpected global disruptions (e.g., COVID-19).**

* **2020 and 2021** exhibit clear impacts from the pandemic, while **2022 onward** shows recovery trends with occasional fluctuations tied to trade cycles.

### ****Which states contribute the most to freight weight?****

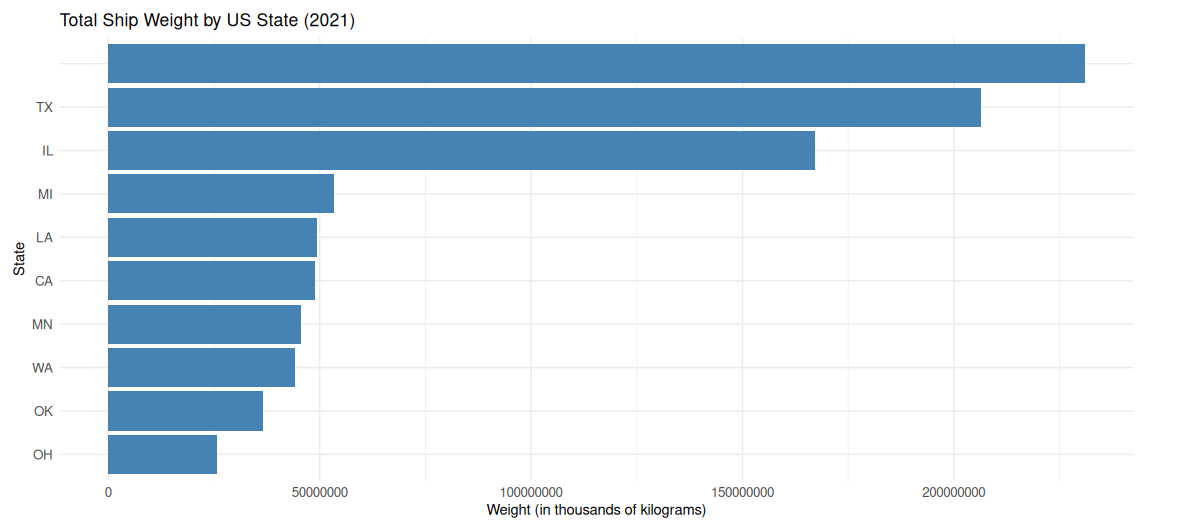


#### ****Analysis for 2020****

The following bar chart illustrates the top 10 states contributing to freight weight in 2020:

**Key Observations:**

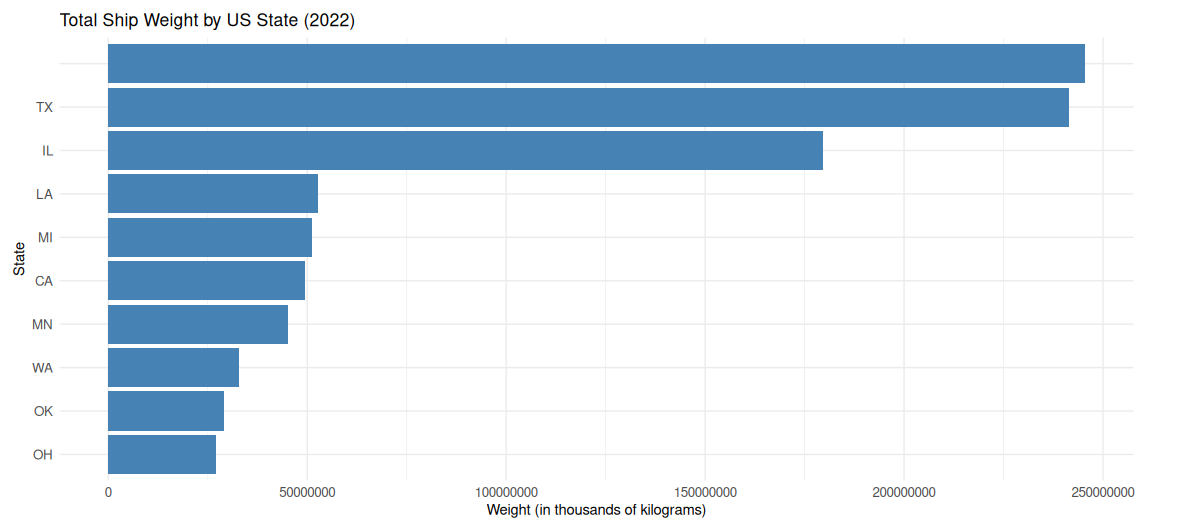
* The states with the highest freight weight include Texas, California and Illinois.
* Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.
* States like Los Angeles and California also feature prominently, likely due to their large populations and economic activity.



****Analysis for 2021****

**Key Observations:**

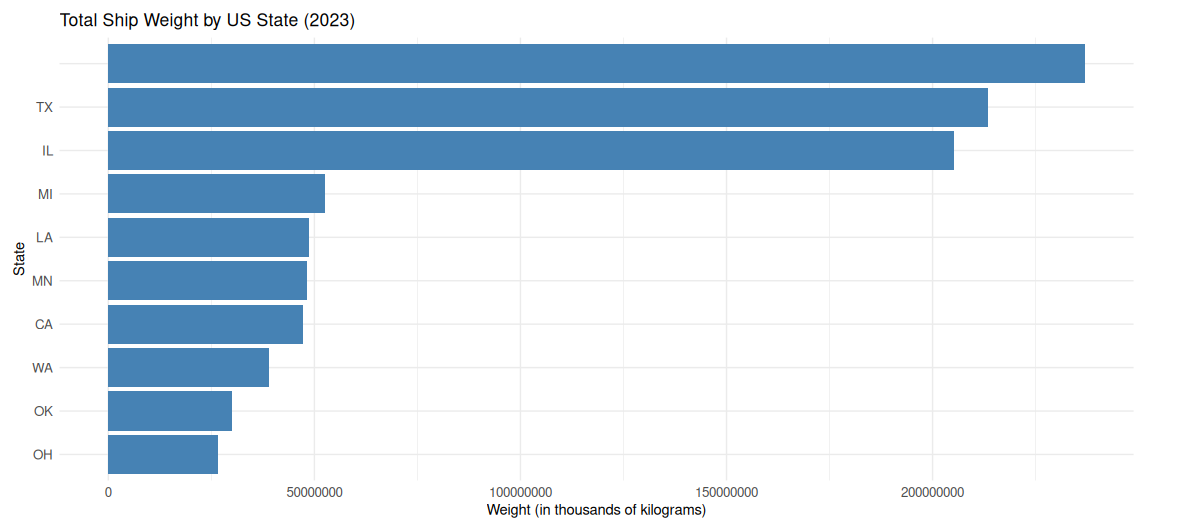
* The states with the highest freight weight include Texas, Illinois and Miami.
* Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.
* **States like Los Angeles and California also feature prominently, likely due to their large populations and economic activity.**



****Analysis for 2022****

**Key Observations:**

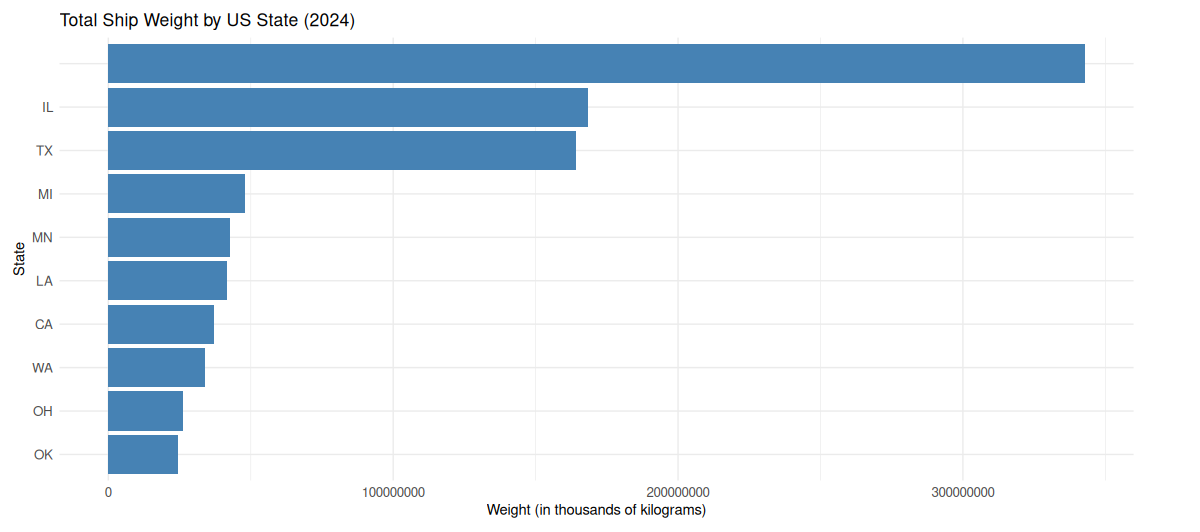
* The states with the highest freight weight include Texas, Illinois and Los Angeles.
* Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.



****Analysis for 2023****

**Key Observations:**

* The states with the highest freight weight include Texas, Illinois and Miami.
* **Texas leads by a significant margin, reflecting its central role in freight logistics, possibly due to its major ports and highways.**

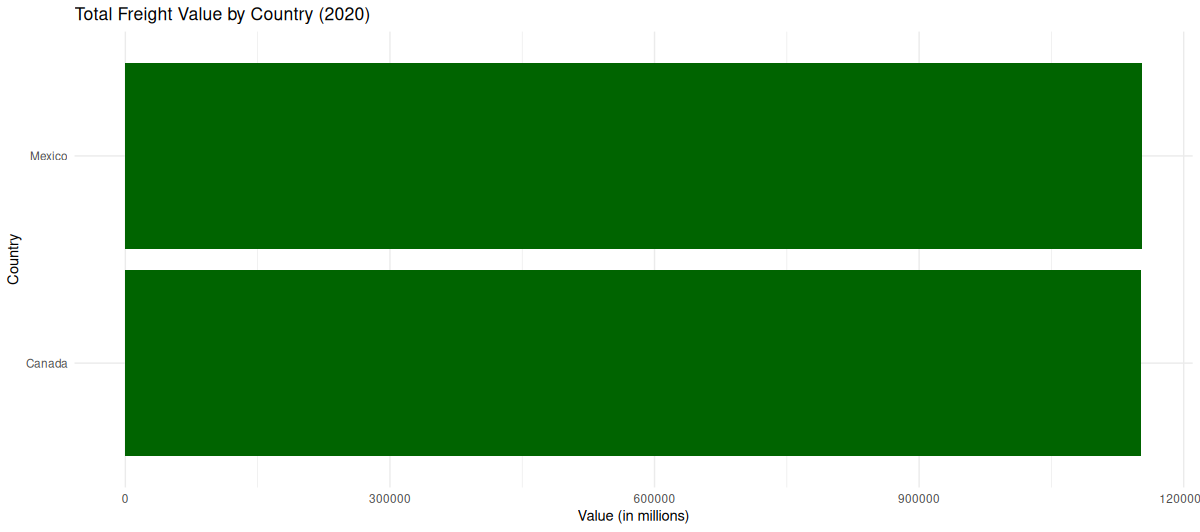


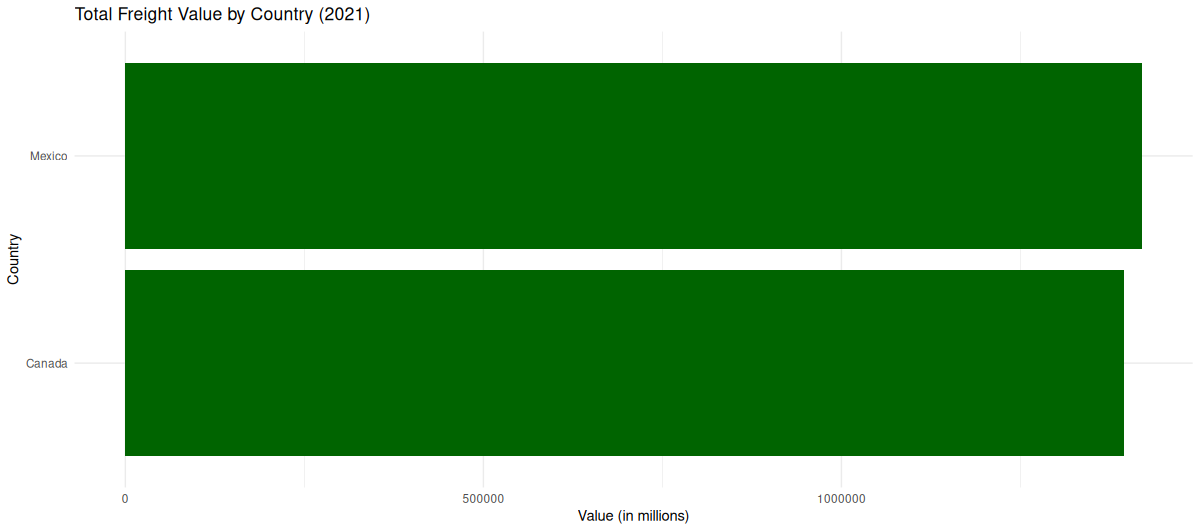
****Analysis for 2024****

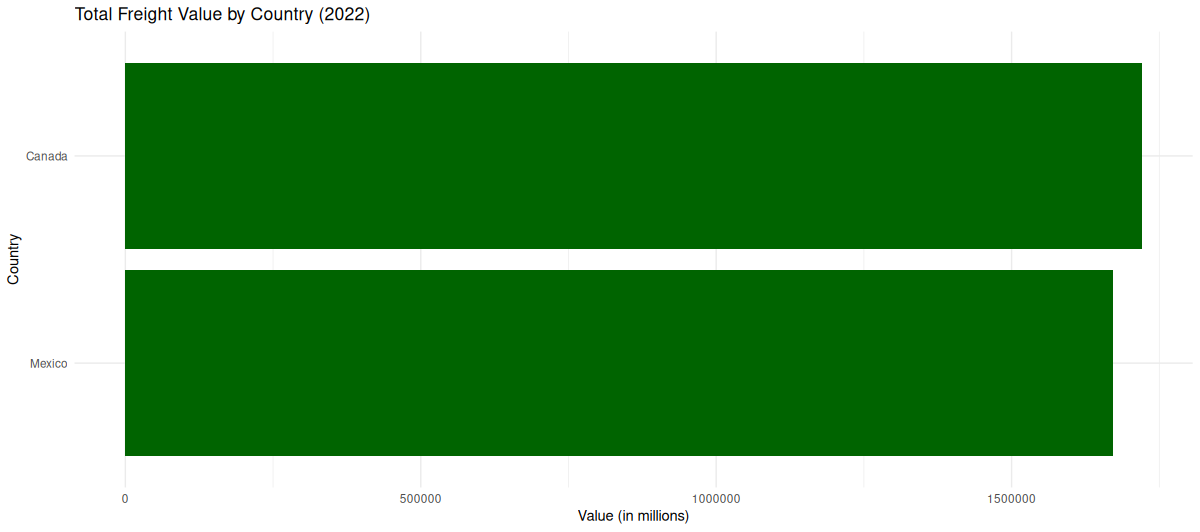
**Key Observations:**

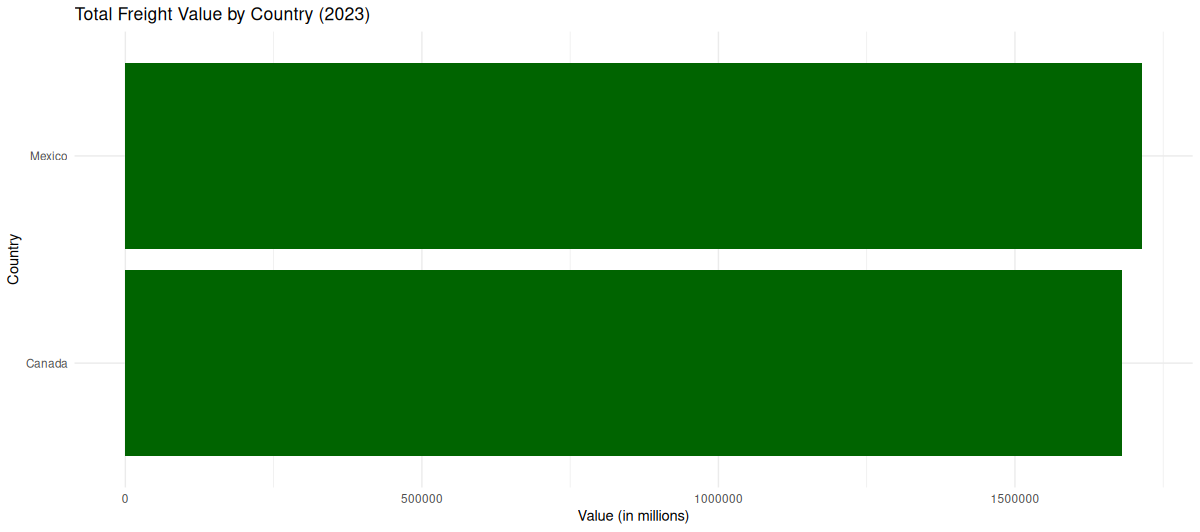
* The states with the highest freight weight include Illinois, Texas and Miami.
* **Illinois leads Texas not by a significant margin.**

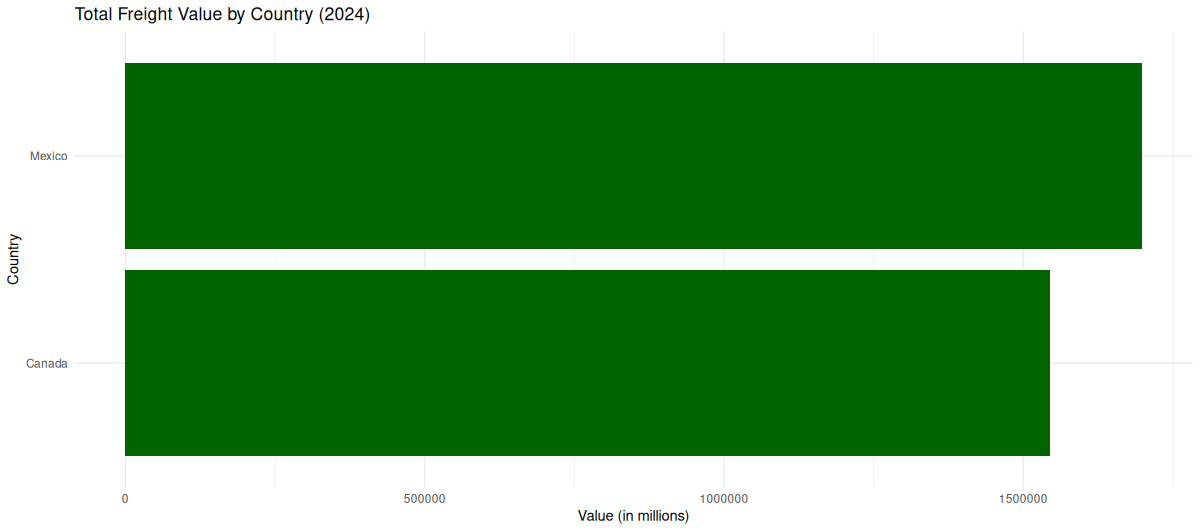
### ****Which country contribute the most to freight value?****

****

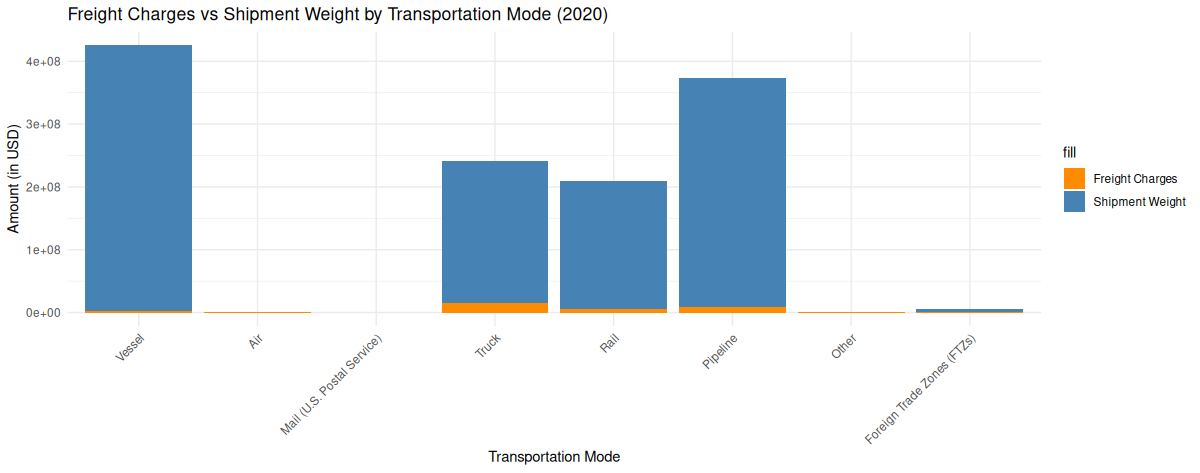
****

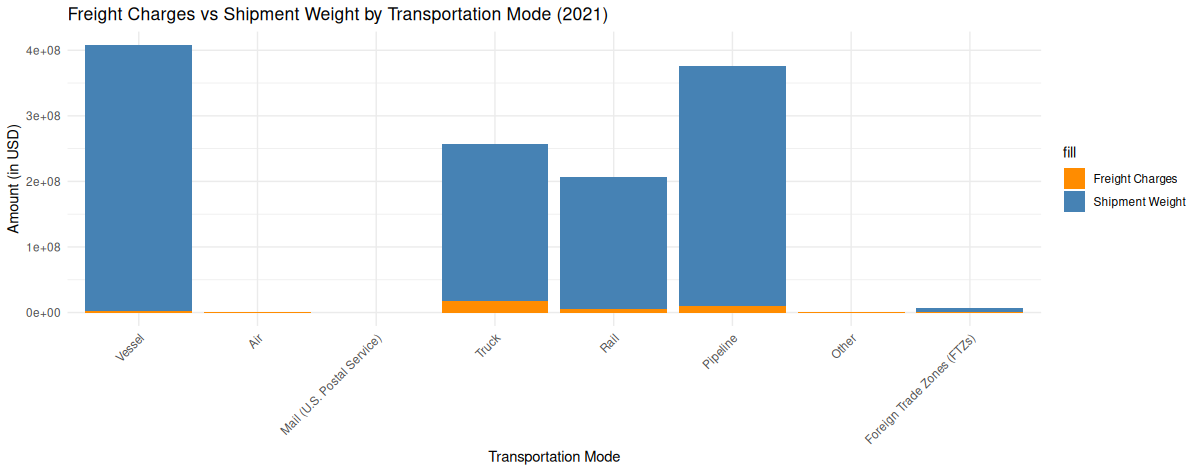
****

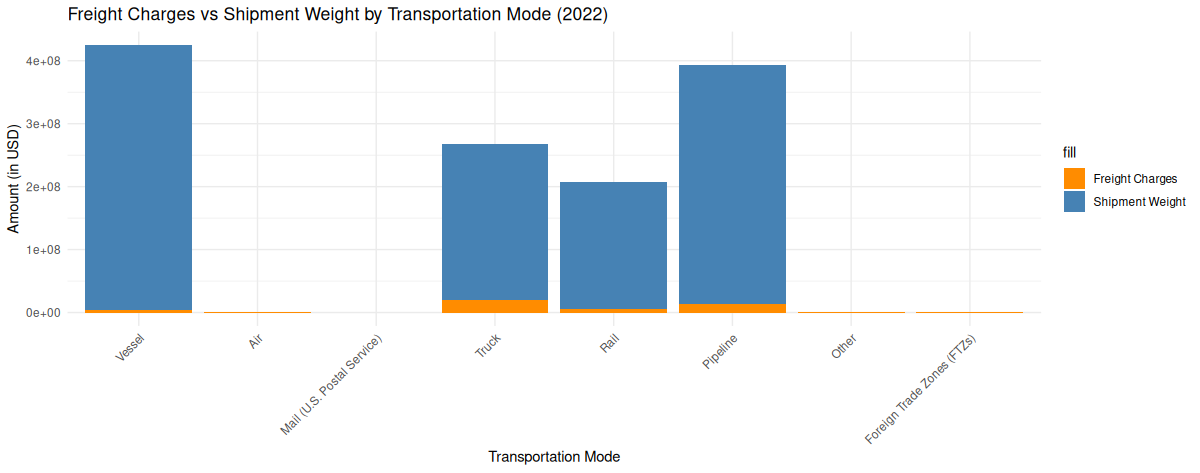
****

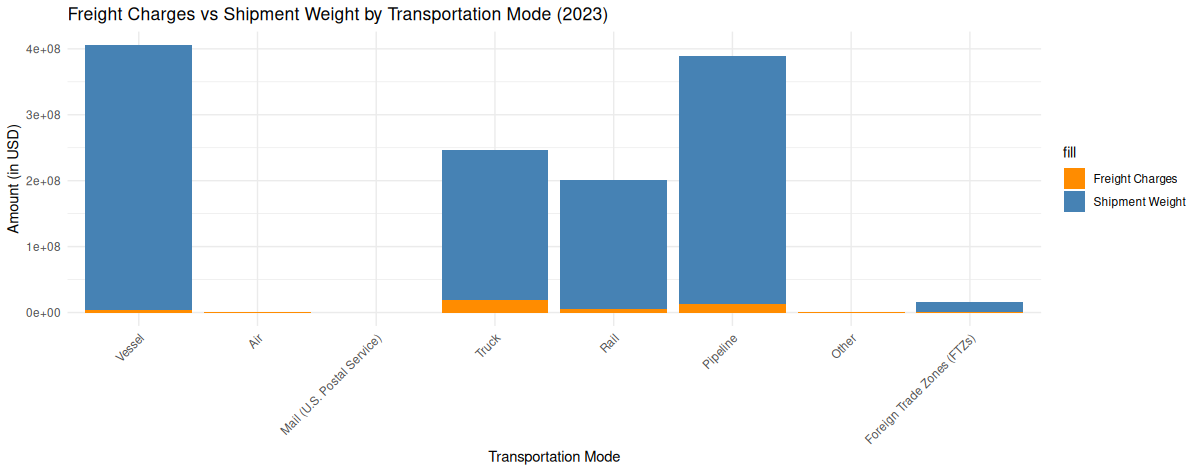
****

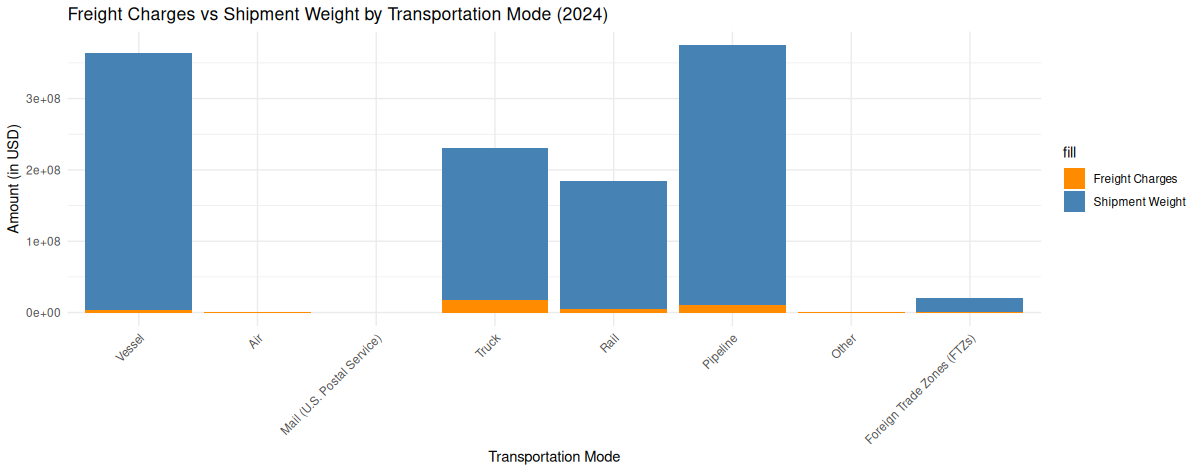
### **What patterns exist between shipment weight and charges?**

****

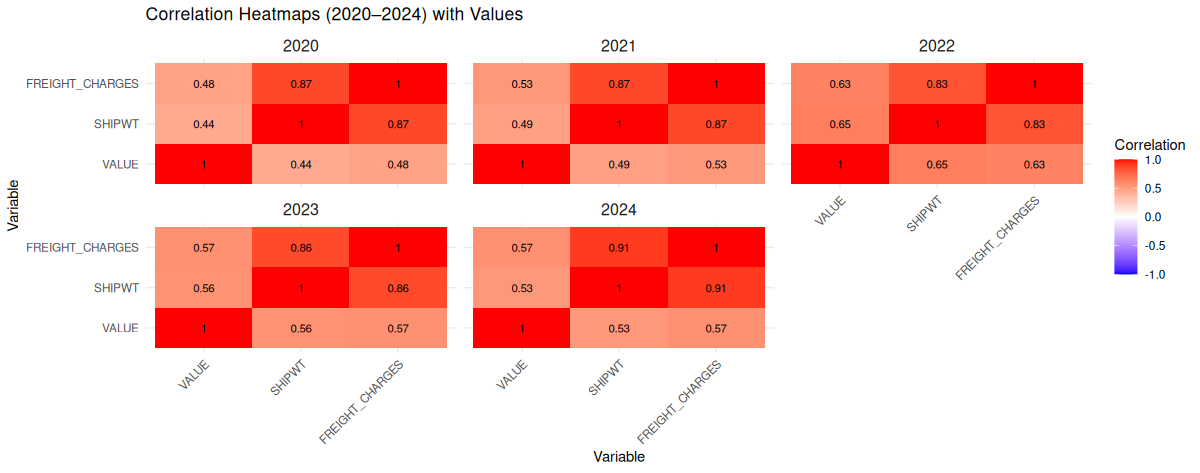
****

****

****

****

### ****What is the correlation between Freight Value, Shipment Weight and Freight Charges for each year from 2020 to 2024?****



**Analysis**

#### **2020:**

* **Freight Value vs. Shipment Weight (0.44):** A moderate positive correlation indicates that higher shipment weight is associated with an increase in freight value, though the relationship is not strong.
* **Freight Value vs. Freight Charges (0.48):** A moderate positive correlation suggests that higher freight value is moderately associated with higher freight charges.
* **Shipment Weight vs. Freight Charges (0.87):** A strong positive correlation implies that heavier shipments tend to incur higher freight charges.

#### **2021:**

* **Freight Value vs. Shipment Weight (0.49):** A moderate positive correlation, slightly stronger than in 2020, shows an increasing association between shipment weight and freight value.
* **Freight Value vs. Freight Charges (0.53):** A moderate positive correlation, also stronger than in 2020, highlights a growing relationship between freight value and freight charges.
* **Shipment Weight vs. Freight Charges (0.87):** The strong positive correlation remains consistent, indicating a continued link between shipment weight and charges.

#### **2022:**

* **Freight Value vs. Shipment Weight (0.65):** A stronger positive correlation reflects a significant relationship between the value of goods and their shipment weight in this year.
* **Freight Value vs. Freight Charges (0.63):** The correlation strengthens, showing a notable association between freight value and charges.
* **Shipment Weight vs. Freight Charges (0.83):** The strong positive correlation decreases slightly but remains substantial.

#### **2023:**

* **Freight Value vs. Shipment Weight (0.56):** A moderate positive correlation, lower than 2022, indicates the relationship between value and shipment weight weakened slightly.
* **Freight Value vs. Freight Charges (0.57):** A consistent moderate positive correlation reflects a stable association between freight value and charges.
* **Shipment Weight vs. Freight Charges (0.86):** The strong positive correlation indicates the relationship remains solid between shipment weight and charges

#### ****2024:****

* **Freight Value vs. Shipment Weight (0.53):** A moderate positive correlation, slightly lower than in 2023, suggests a declining association.
* **Freight Value vs. Freight Charges (0.57):** The consistent correlation indicates that freight value and charges remain moderately linked.
* **Shipment Weight vs. Freight Charges (0.91):** A very strong positive correlation indicates an almost linear relationship, suggesting that heavier shipments are nearly directly proportional to higher freight charges.

### ****Overall Interpretation:****

1. **Shipment Weight and Freight Charges:** Across all years, there is a consistently **strong positive correlation (above 0.83)**, with the strongest in 2024 (0.91). This indicates that shipment weight is a critical determinant of freight charges.
2. **Freight Value and Shipment Weight:** The correlation is generally moderate, ranging from 0.44 in 2020 to a peak of 0.65 in 2022, and then slightly declining afterward. This suggests that heavier shipments tend to have higher values, but the relationship is not as strong or consistent as with freight charges.
3. **Freight Value and Freight Charges:** The correlation is moderate across all years, ranging from 0.48 to 0.63, indicating that while higher-value shipments often result in higher charges, other factors also contribute.

### ****Key Insights:****

* Shipment weight is the dominant factor influencing freight charges.
* While freight value contributes to determining charges, its impact is moderate compared to shipment weight.
* The year 2022 shows the strongest correlations across all variables, possibly indicating unusual or specific market conditions.

### ****How has the total freight value changed over the years from 2020 to 2024?****

### 

### ****Analysis of Freight Value Trends (2020-2024):****

### From the data:

* + The freight value increased steadily from 2020 to 2023.
  + There was a notable jump in 2021 and 2022, with 2022 to 2023 having the highest freight value.
  + Slight decline in 2024.
* This suggests consistent growth in freight activities or higher-value shipments over the years.

### ****Insights:****

* The significant year-on-year increase from 2020 to 2024 could indicate:
  + Increased global trade.
  + An expansion in freight services, especially after 2020 (potentially post-pandemic recovery)

### 5. Deployment

Insights and visualizations were shared via:

* **GitHub Repository: Includes documentation, R script and data files.**