DATA ANALYSIS

Exploring the connection between tariffs and economic development

The Pythonistas

By Alessia Barel, Matteo Naccari, Ceren Özgür and Elisa Pascon

DO RICHER COUNTRIES HAVE LOWER TARIFFS THAN POORER ONES?

STRUCTURE

1. Reference
2. Data exploring
3. General Analysis
4. In-depth analysis
5. Final Considerations

1. Introduction

Product import tariffs

Country GDP

Trade percentage of GDP

```
heading("LOADING DATA")

base_url = "https://raw.githubusercontent.com/Elisa-PSC/CPDM_project_20252026/
 ↪main/data"

# Load complete datasets (all years)
df_tariffs_all = pd.read_csv(base_url + "/WITS-Product_bycountry_all.csv")
df_gdp_all = pd.read_csv(base_url + "/WITS-Country-GDP_USD_all.csv")
df_trade_all = pd.read_csv(base_url + "/WITS-Trade_percentage_of_GDP_all.csv")

print(f"Complete tariff data: {df_tariffs_all.shape[0]} rows, {df_tariffs_all.
 ↪shape[1]} columns")
print(f"Complete GDP data: {df_gdp_all.shape[0]} rows, {df_gdp_all.shape[1]}␣
 ↪columns")
print(f"Complete trade data: {df_trade_all.shape[0]} rows, {df_trade_all.
 ↪shape[1]} columns")
```

```
============================================================
LOADING DATA
============================================================

Complete tariff data: 189 rows, 40 columns
Complete GDP data: 190 rows, 37 columns
Complete trade data: 175 rows, 37 columns
```

2. Data exploring

Data preparation

Our datasets are in "wide" format (years as columns), but **we need "long" format** (years as rows) for easier analysis and visualization.

To achieve that, we will:

1. reshape all three datasets from wide to long format;
2. merge them into a single comprehensive dataset;
3. handle missing values appropriately;
4. create derived variables (log-transformed GDP, wealth quartiles).

```python
# Reshape trade data
trade_historical = df_trade_all.melt(
    id_vars=["Country Name", "Indicator Name"],
    value_vars=[str(year) for year in range(1988, 2023)],
    var_name="Year",
    value_name="Trade_Pct_GDP",
)
trade_historical["Year"] = trade_historical["Year"].astype(int)

# Merge all datasets
merged_all = tariff_historical.merge(
    gdp_historical[["Country Name", "Year", "GDP_USD"]],
    left_on=["Reporter Name", "Year"],
    right_on=["Country Name", "Year"],
    how="left",
)
merged_all = merged_all.merge(
    trade_historical[["Country Name", "Year", "Trade_Pct_GDP"]],
    left_on=["Reporter Name", "Year"],
    right_on=["Country Name", "Year"],
    how="left",
    suffixes=("", "_trade"),
)
merged_all = merged_all.drop(columns=["Country Name", "Country Name_trade"])

countries_before = set(merged_all["Reporter Name"].unique())

# Cleaning
merged_all = merged_all.dropna(subset=["GDP_USD", "Tariff_Rate"])

countries_after = set(merged_all["Reporter Name"].unique())
```

```
============================================================
PREPARING DATA
============================================================


Complete merged dataset (1988-2022): 3589 observations
Number of countries: 179
Years covered: [1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997,
1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]
```

2. Data exploring

Filter for 2000-2022

Restricting the **analysis to 2000–2022** ensures consistently large and diverse samples (typically exceeding 100 countries per year) allowing **for stable estimates and meaningful statistical inference**

```
# NOW FILTER FOR 2000-2022 analysis
merged = merged_all[merged_all["Year"].between(2000, 2022)].copy()

# Create wealth quartiles based on the filtered dataset
labels = ["Poorest 25%", "Lower-Middle 25%", "Upper-Middle 25%", "Richest 25%"]
merged["Wealth_Quartile"] = pd.qcut(merged["GDP_USD"], q=4, labels=labels,␣
 ↪duplicates='drop')

print(f"\nFiltered dataset (2000-2022): {merged.shape[0]} observations")
print(f"Number of countries: {merged['Reporter Name'].nunique()}")
print(f"Years covered: {sorted(merged['Year'].unique().tolist())}")
```

```
Filtered dataset (2000-2022): 3212 observations
Number of countries: 177
Years covered: [2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009,
2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022]
```

3. General Analysis

Outlier Analysis

This analysis helps us **understand** not just the average relationship, but **the full range of policy choices countries make**

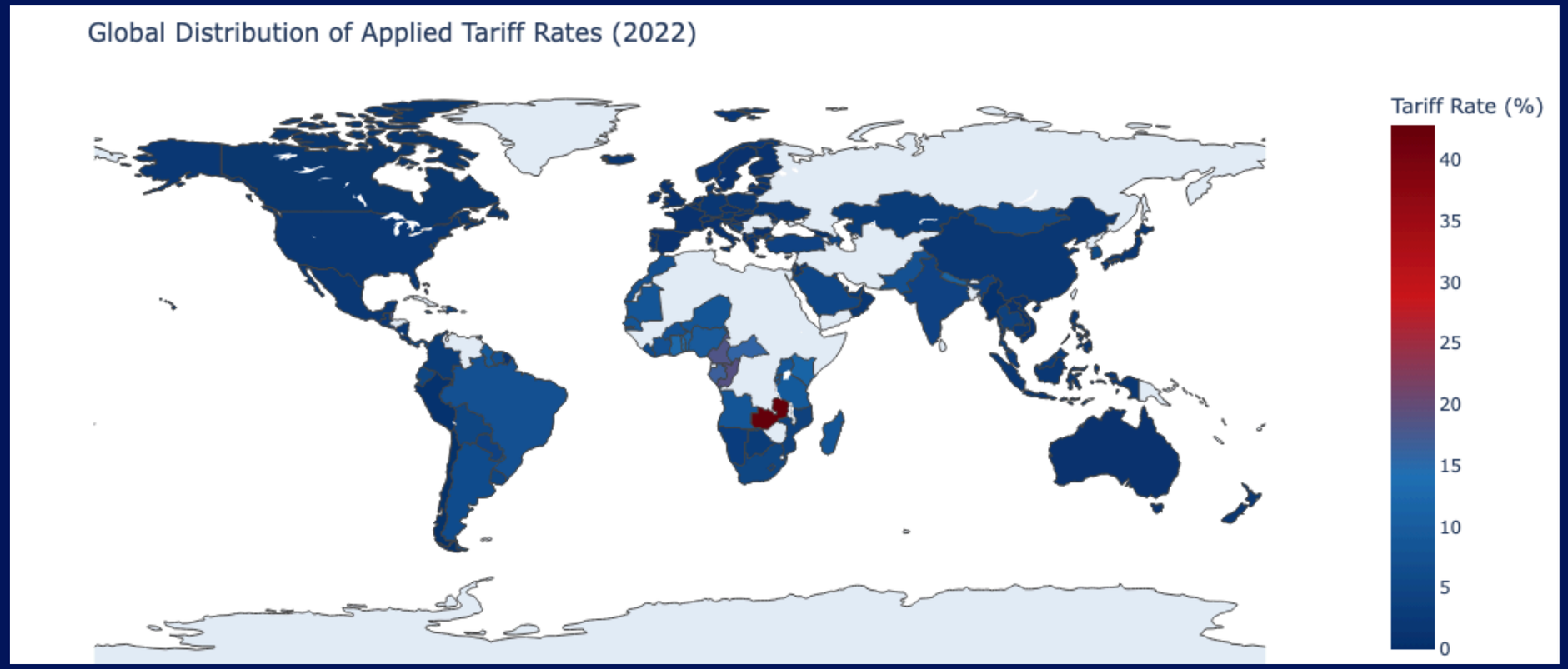
Reporter Name	Year	Tariff_Rate	Wealth_Quartile	Tariff_ZScore
Comoros	2016	274.07%	Poorest 25%	35.6016
Bermuda	2019	103.17%	Poorest 25%	12.8808
Fiji	2014	91.59%	Poorest 25%	11.3416
Zambia	2014	77.87%	Lower-Middle 25%	9.5167
Fiji	2005	65.89%	Poorest 25%	7.9246
Zambia	2022	42.84%	Lower-Middle 25%	4.8603
Solomon Islands	2015	35.65%	Poorest 25%	3.904
Palau	2018	34.63%	Poorest 25%	3.7684
Seychelles	2005	32.6%	Poorest 25%	3.4988
Seychelles	2006	32.58%	Poorest 25%	3.4961

Reporter Name	Count
India	5
Seychelles	5
Bermuda	5
Tunisia	5
Central African Republic	4
Morocco	4
Fiji	3
Bangladesh	2
Zambia	2
Solomon Islands	2

3. General Analysis

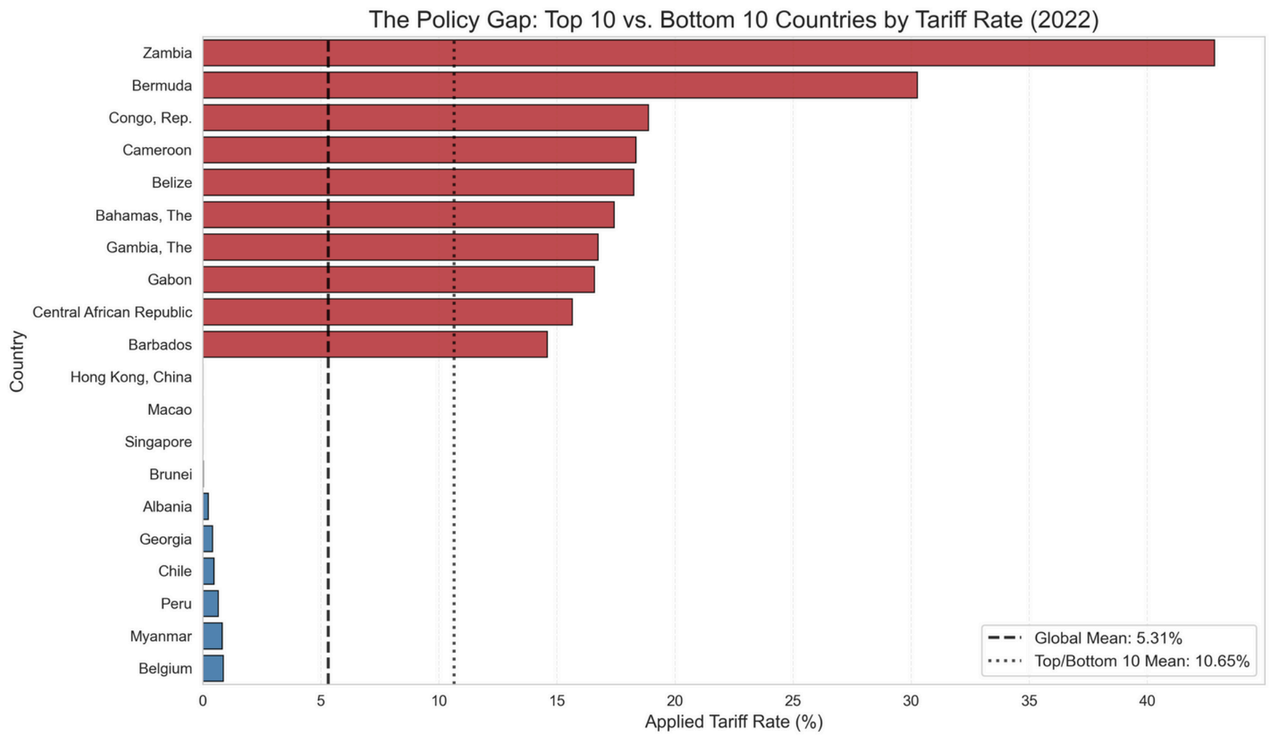
Average tariffs by Country

The map illustrates the tax burden in the world through a chromatic scale: shades close to **red** indicate **higher taxation**, while those tending towards **blue** indicate a l**ower tax burden**



3. General Analysis

Average tariffs by Country



The plot reveals that while average global tariffs appear moderate, they conceal **deep heterogeneity**, with a small group of countries maintaining highly protectionist regimes and another group operating at the opposite extreme

3. General Analysis

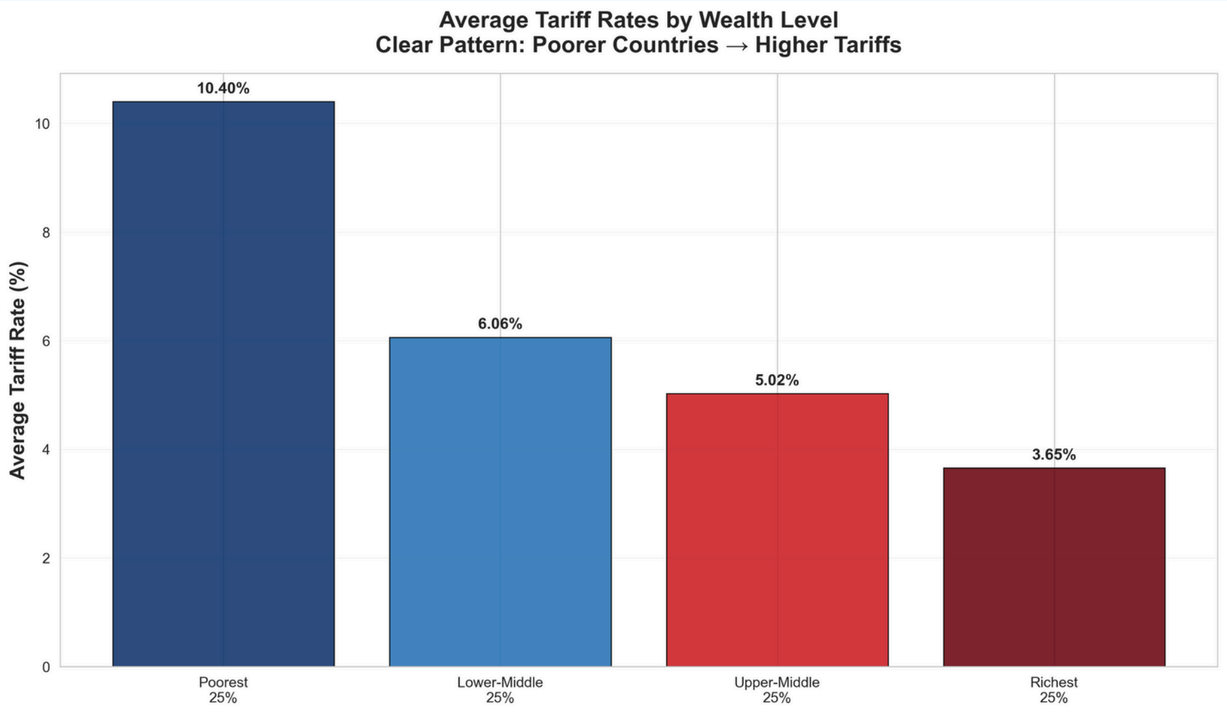
Tariff statistics by wealth level

Most importantly, the mean difference of 6.74 percentage points between the poorest and richest quartiles - representing a 185% higher tariff rate among poor countries -demonstrates that these groups occupy fundamentally different trade policy regimes. This substantial gap provides strong preliminary evidence **that GDP and tariff policy are indeed meaningfully connected**, with wealthier nations consistently maintaining more liberal trade policies than their poorer counterparts.

Wealth_Quartile	('Tariff_Rate', 'count')	('Tariff_Rate', 'mean')	('Tariff_Rate', 'median')	('Tariff_Rate', 'std')	('Tariff_Rate', 'min')	('Tariff_Rate', 'max')	('GDP_USD', 'mean')
Poorest 25%	803	10.40%	9.67%	11.93%	0.00%	274.07%	4,077,350,317
Lower-Middle 25%	803	6.06%	4.96%	5.42%	0.00%	77.87%	20,007,442,802
Upper-Middle 25%	803	5.02%	3.61%	4.40%	0.00%	25.42%	112,137,664,833
Richest 25%	803	3.65%	2.39%	3.17%	0.00%	26.51%	1,703,106,510,585

3. General Analysis

Tariff statistics by wealth level



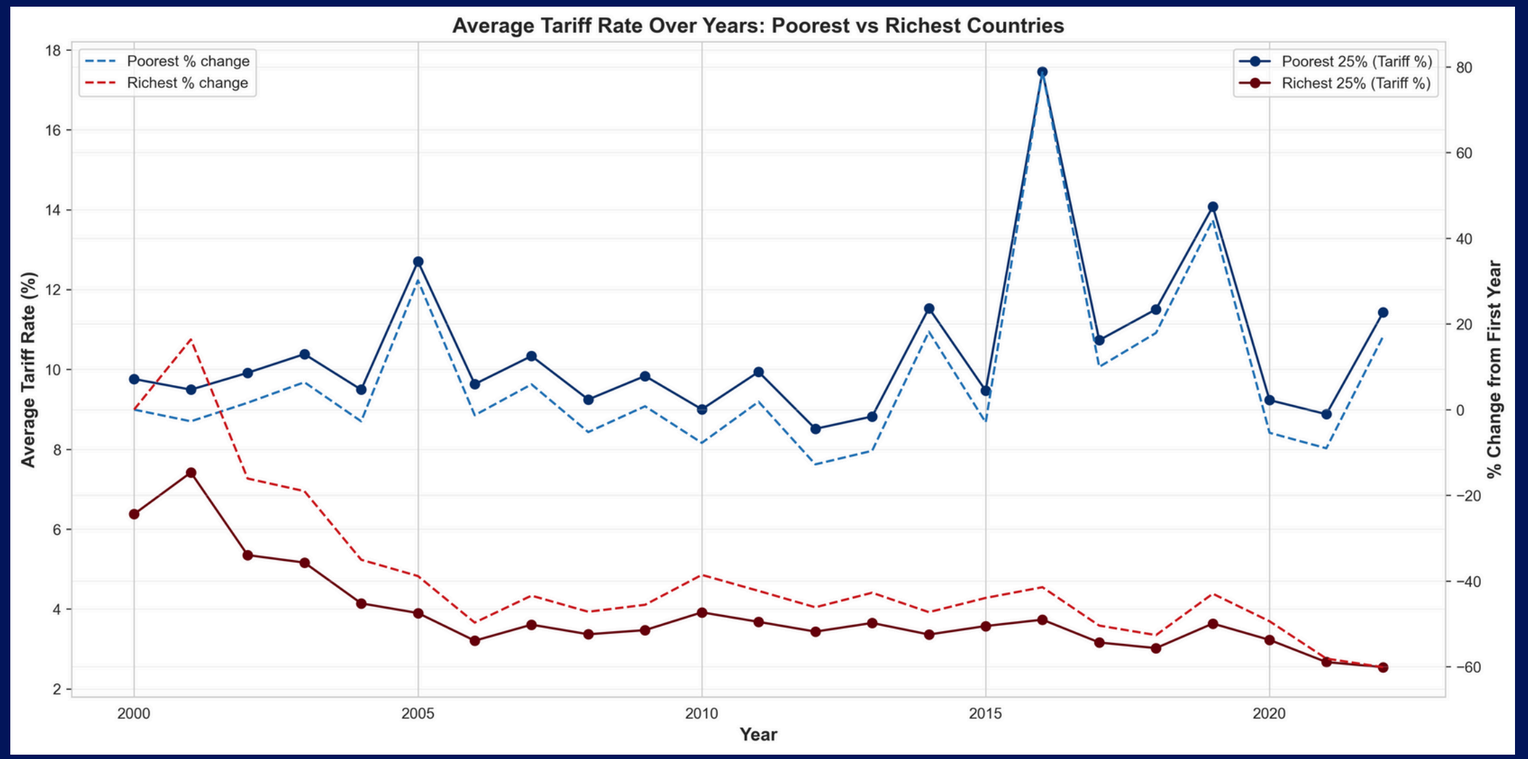
The chart shows **average tariff rates for each wealth quartile**, making the relationship visible at a glance.

A clear **descending pattern** from left to right confirms our hypothesis visually.

3. General Analysis

Time series comparison between wealth groups

The figure suggests that **trade liberalization has been uneven across income groups**. Rather than converging, tariff policies have followed divergent trajectories, with poorer countries remaining locked into higher and more volatile tariff regimes, r**einforcing the long-term wealth–tariff gap**



3. General Analysis

Volatility Analysis

Some countries maintain remarkably consistent tariff policies year after year, while others frequently adjust rates in response to economic or political pressures. We identified which **countries have the most and least volatile tariff policies**

Reporter Name	Mean_Tariff	Std_Dev	Min	Max	Wealth_Group	Range
Comoros	26.66	74.36	4.24	274.07	Poorest 25%	269.83
Fiji	20.27	22.12	8.35	91.59	Poorest 25%	83.24
Bermuda	25.58	21.09	15.27	103.17	Poorest 25%	87.9
Zambia	12.18	17.77	3.29	77.87	Lower-Middle 25%	74.58
Seychelles	14.75	14.27	1.07	32.6	Poorest 25%	31.53
Palau	14.24	12.39	6.4	34.63	Poorest 25%	28.23
Solomon Islands	20.04	10.79	7.17	35.65	Poorest 25%	28.48
Bhutan	15.74	8.16	9.97	21.51	Poorest 25%	11.54
Morocco	10.48	7.77	3.02	25.42	Upper-Middle 25%	22.4
India	10.66	7.25	4.47	26.51	Richest 25%	22.04

Reporter Name	Mean_Tariff	Std_Dev	Min	Max	Wealth_Group	Range
Samoa	10.14	0.38	9.71	10.85	Poorest 25%	1.14
Germany	1.91	0.37	1.43	2.99	Richest 25%	1.56
France	1.91	0.35	1.2	2.87	Richest 25%	1.67
Ireland	1.94	0.35	1.34	2.78	Upper-Middle 25%	1.44
Finland	1.66	0.35	1.06	2.52	Richest 25%	1.46
Canada	1.57	0.28	1.31	2.46	Richest 25%	1.15
Afghanistan	5.39	0.27	5	5.63	Lower-Middle 25%	0.63
United States	1.72	0.19	1.48	2.16	Richest 25%	0.68
Singapore	0.07	0.09	0	0.39	Richest 25%	0.39
Uzbekistan	2.6	0.05	2.56	2.63	Upper-Middle 25%	0.07

4. In-depth analysis

Correlation analysis

Correlation coefficients quantify the **strength** and **direction** of relationships between variables.

Research Objectives

We analyze the link between **GDP** and **Tariff Rates** through three specific lenses:

- **Raw Correlation**: Identifying the direct statistical baseline between the two variables
- **Log-Transformed GDP**: Adjusting for country size to ensure that vast economic differences do not skew the results
- **Trade Openness vs. Tariffs**: Testing the hypothesis that nations more integrated into global commerce tend to impose lower trade barriers

Interpreting correlations in economics:

Absolute correlation	Interpretation
0.0 - 0.2	Very weak
0.2 - 0.4	Weak to moderate
0.4 - 0.6	Moderate
0.6 - 0.8	Strong
0.8 - 1.0	Very strong

Correlation ≠ causation

we are measuring association, not proving that one variable causes changes in another

Open economies generally **benefit more** from reduced trade barriers and have **stronger incentives to maintain** them at low levels.

```
heading("CORRELATION ANALYSIS")

# Calculate correlations
corr_gdp = merged["GDP_USD"].corr(merged["Tariff_Rate"])
corr_log_gdp = merged["GDP_Log"].corr(merged["Tariff_Rate"])
corr_trade = merged.dropna(subset=["Trade_Pct_GDP"])["Trade_Pct_GDP"].corr(
    merged.dropna(subset=["Trade_Pct_GDP"])["Tariff_Rate"]
)
```

```
============================================================
CORRELATION ANALYSIS
============================================================

CORRELATION COEFFICIENTS:
GDP vs Tariff Rate: -0.1112
Log(GDP) vs Tariff Rate: -0.3284
Trade Openness vs Tariff Rate: -0.1900
```

4. In-depth analysis

Regression analysis

Regression analysis tells us **how much** one variable changes when another changes

$$\text{Tariff Rate} = \beta_0 + \beta_1 \times \log(\text{GDP}) + \epsilon$$

What to look for in the output:

- **R^2** : what percentage of tariff variation is explained by GDP?
- **Coefficient B1**: how much do tariffs change for each unit increase in log(GDP)?
- **P-value**: is this relationship statistically significant, or could it be due to change?

```
heading("REGRESSION ANALYSIS")

# Prepare data for regression
X = merged[["GDP_Log"]].values
y = merged["Tariff_Rate"].values

# Fit the model
model = LinearRegression()
model.fit(X, y)

# Calculate statistics
slope, intercept, r_value, p_value, std_err = stats.linregress(
    merged["GDP_Log"], merged["Tariff_Rate"]
)
r2 = r_value ** 2
```

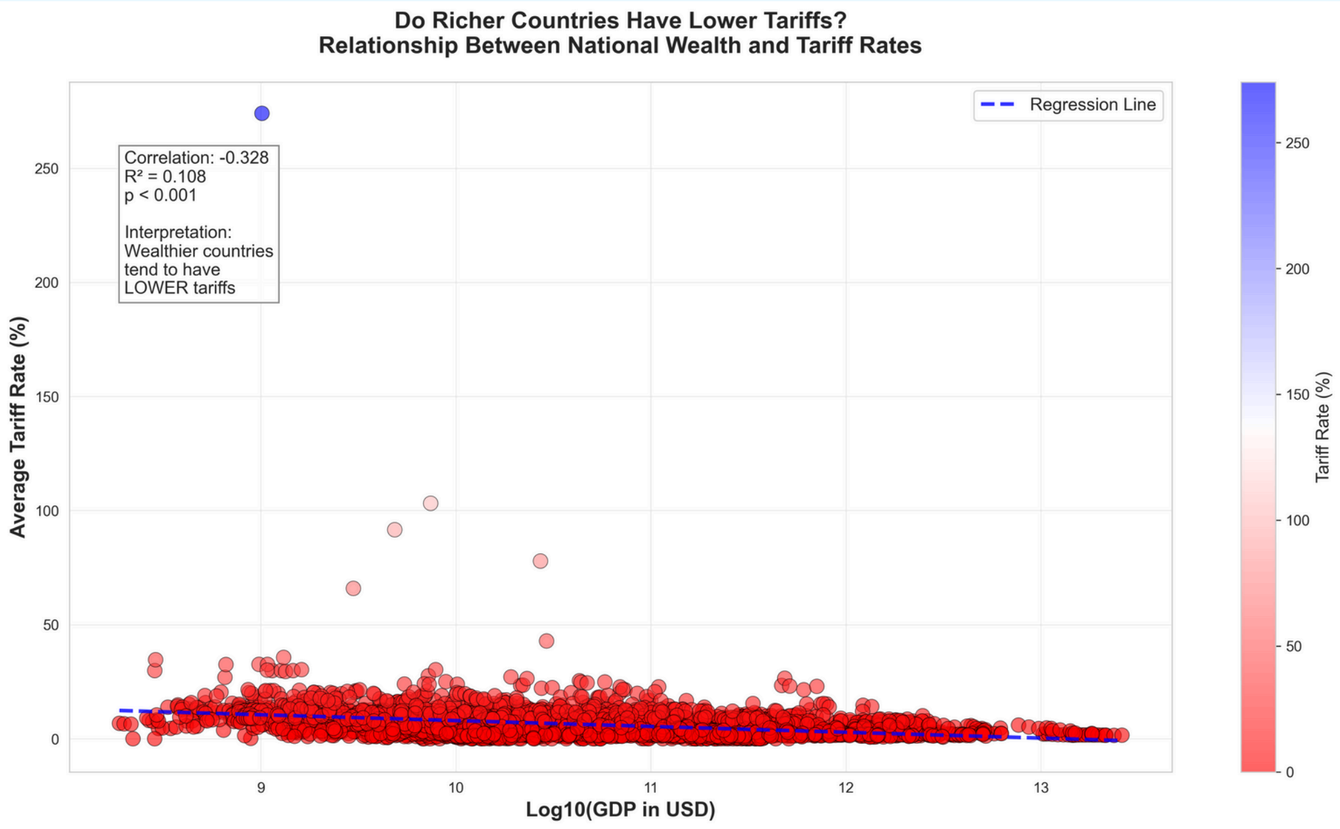
```
============================================================
REGRESSION ANALYSIS
============================================================

REGRESSION RESULTS (Predicting Tariff Rate from Log GDP):

R-squared (R2): 0.1079

Coefficient (Slope): -2.5568

Intercept: 33.5396

P-value: 1.21e-81
```

Model Performance & Explanatory Power

- **R^2 Value: 0.1079**
- Approximately **10.79% of the variation** in tariff rates is explained by GDP alone
- For a single predictor in an economic context, this represents a substantial amount of explanatory power

The Impact of Economic Scale

- **Estimated Coefficient (Slope): -2.5568**
- **Practical Effect:** A one-unit increase in log GDP is associated with a 2.56 percentage point decrease in tariffs.
- **Statistical Precision**: The result is not only statistically precise (p-value: 1.21e−81) but also **economically meaningful**



The scatter plot reveals a clear **negative relationship** between national wealth and average tariff rates

4. In-depth analysis

Hypothesis testing

To directly test our core hypothesis (that the poorest and richest countries occupy fundamentally different trade policy regimes) we need a more **targeted statistical test** comparing these specific groups.

Used a **two-sample t-test** to determine if the difference between poorest and richest countries is statistically significant or could have occurred by chance.

Our Hypotheses:

$$H_0 \text{ (Null hypothesis)}: \quad \mu_{\text{poorest}} = \mu_{\text{richest}} \quad \Rightarrow \quad \text{no difference in mean tariffs}$$

$$H_1 \text{ (Alternative hypothesis)}: \quad \mu_{\text{poorest}} > \mu_{\text{richest}} \quad \Rightarrow \quad \text{poorest countries have higher tariffs}$$

Decision Rule:

If p-value $< 0.05 \Rightarrow$ reject H_0 (statistically significant higher tariffs in poorest)

If p-value $\geq 0.05 \Rightarrow$ fail to reject H_0 (insufficient evidence of higher tariffs in poorest)

```
============================================================
HYPOTHESIS TESTING
============================================================

RESEARCH QUESTION:
Do the poorest countries have significantly higher tariffs than the richest?

HYPOTHESES:
H0 (Null): Poorest and richest countries have equal average tariffs

H1 (Alternative): Poorest countries have higher average tariffs

TWO-SAMPLE T-TEST RESULTS:
T-statistic: 15.4803
P-value: 0.000000
Sample sizes: Poorest n=803, Richest n=803

------------------------------------------------------------
DECISION:
REJECT the null hypothesis (p < 0.05)
```

The results indicate a **t-statistic of 15.4803 and a p-value effectively equal to zero**, with equal sample sizes of 186 countries in each group.

Given this **extremely low p-value**, we **reject the null hypothesis** at the conventional 5% significance level.

The evidence strongly supports the conclusion that poorer countries face higher tariffs on average, whereas wealthier countries tend to maintain lower tariff levels

4. In-depth analysis

Trade openness analysis

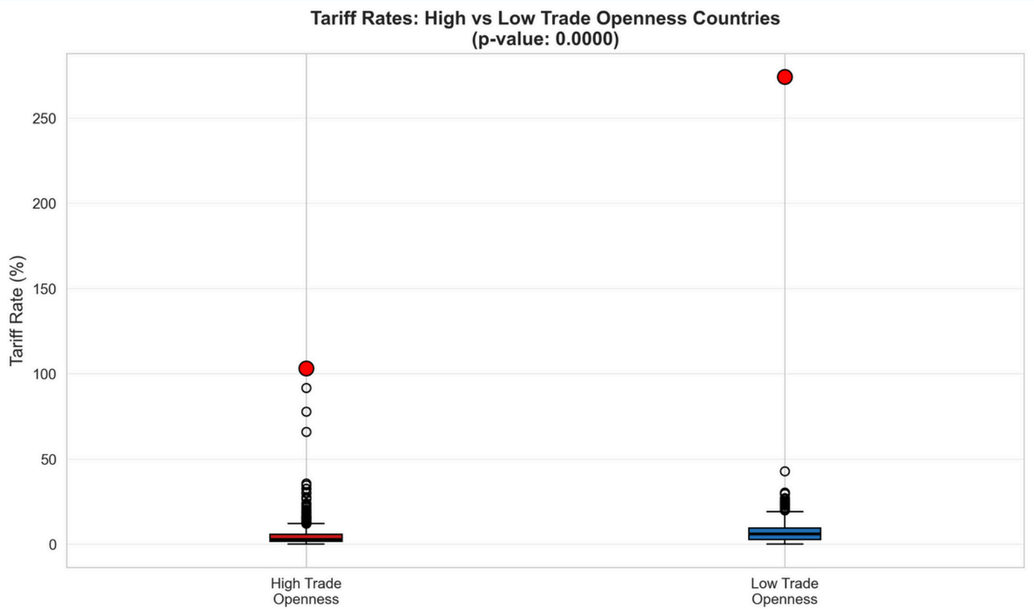
Is the wealth-tariff relationship we've found actually just a wealth-trade relationship? Or is trade openness an independent factor?

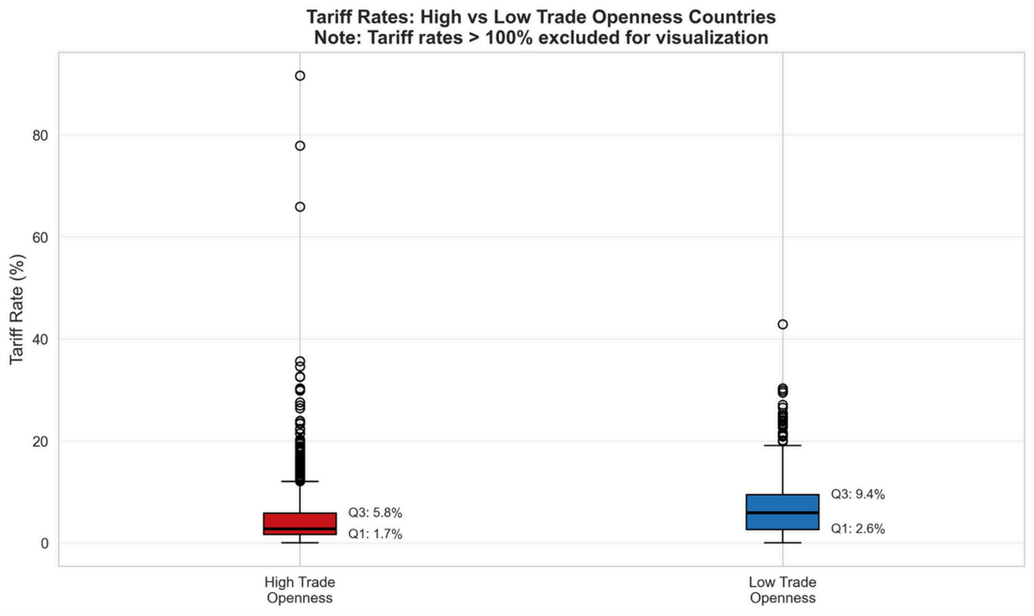
```
heading("TRADE OPENNESS ANALYSIS")

df_test = merged.dropna(subset=["Trade_Pct_GDP", "Tariff_Rate"])
median_trade = df_test["Trade_Pct_GDP"].median()
high_trade = df_test[df_test["Trade_Pct_GDP"] >= median_trade]["Tariff_Rate"]
low_trade = df_test[df_test["Trade_Pct_GDP"] < median_trade]["Tariff_Rate"]

t_stat_trade, p_value_trade = stats.ttest_ind(high_trade, low_trade)
```

```
============================================================
TRADE OPENNESS ANALYSIS
============================================================

COMPARING HIGH-TRADE vs LOW-TRADE COUNTRIES:

High-trade countries (n=1460): Mean tariff = 4.83%
Low-trade countries (n=1460): Mean tariff = 6.98%
Difference: 2.15 percentage points

T-statistic: -7.655
P-value: 0.0000

Result: STATISTICALLY SIGNIFICANT (p < 0.05)
```

There is a **systematic divergence** in tariff policies between the two groups ($p<0.0001$).
The t-value of -7.655 **confirms** the strength of the evidence

Metric	High-Trade Countries	Low-Trade Countries
Mean Tariff	4.83%	6.89%





The boxplot doesn't just confirm that tariffs are lower (as seen in the median line), it also demonstrates much greater **consistency**. The narrower box for 'High-trade' countries indicates that their policies are more **aligned**, in contrast to the **high volatility** we observe among less open economies

5. Final Considerations

The analysis clearly indicates that richer countries maintain lower tariffs than poorer countries, and the evidence strongly supports this conclusion

Countries in the **poorest quartile** impose tariffs of: **10.40%**

Countries in the **richest quartile** impose tariffs of: **3.65%**

Poorer nations **significantly higher** tariffs of: **184.5%**

1. The **negative correlation of -0.3284 confirms** an **inverse relationship** between income and tariffs, with an **R^2 of 0.1079**, meaning that **income explains** about **11% of the variation in tariffs** across countries.

2. The **p-value** is far **below** conventional **significance thresholds**, indicating that **this relationship is extremely unlikely to be due to chance!**

3. This pattern has persisted consistently over time. Between **2000** and **2022**, the **wealth - tariff gap remained stable**, suggesting that **it reflects a structural feature of the global economy** rather than a temporary fluctuation.

The implications are significant:

Tariff policies are closely tied to economic development, with countries tending to lower tariffs as they grow wealthier.

This reinforces that:

Global trade inequality is multifaceted, extending beyond mere tariff levels, and highlights the need for development strategies that recognize and address the persistent link between wealth and trade policy.

THANK YOU!