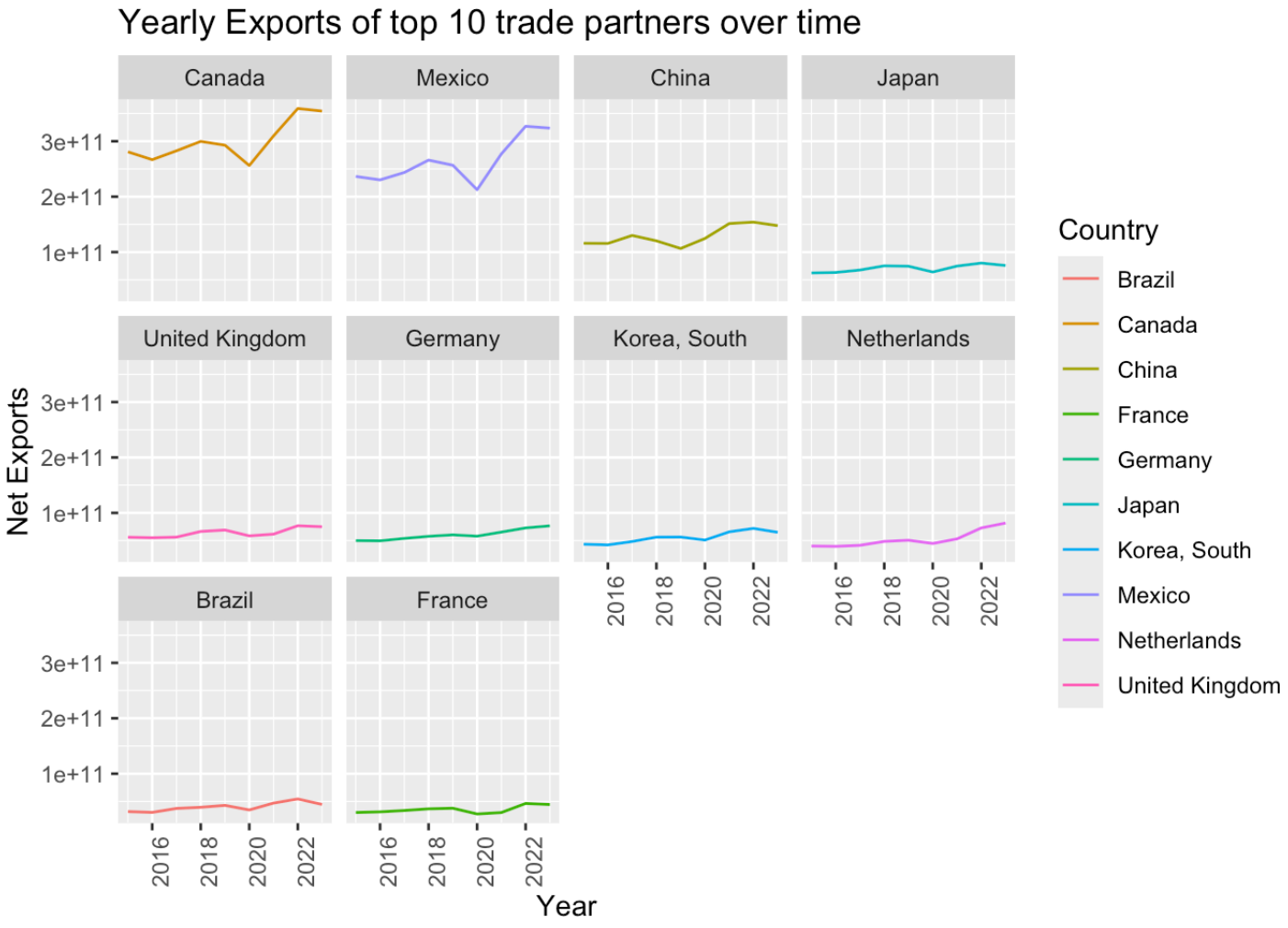
dc_sample_report

2025-10-24

1. Introduction and Motivation.

The analysis explores patterns in U.S. trade with partner countries, aiming to understand both cross-country and within-country determinants of trade volume. Specifically, it investigates two main questions:
How do U.S. trade volumes differ across regions and countries?
What country-specific factors help explain changes in U.S. exports over time?
Initial visualizations reveal that U.S. trade values vary substantially across continents and countries. This motivates a more formal statistical approach to determine whether regional groupings alone can explain trade differences, or if additional economic and political characteristics play a stronger role.

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```



2. Hypothesis Test: Do Countries within the Same Region Differ?

We begin by testing whether average U.S. trade volumes are equal across countries within each region.
Null hypothesis (H_0): Each country's mean U.S. net exports equal its regional mean.
Alternative hypothesis (H_1): At least one country differs significantly from its regional mean.

```
## # A tibble: 5 × 5
##   Continent Country       estimate statistic   p.value
##   <chr>     <chr>            <dbl>     <dbl>     <dbl>
## 1 Africa    Algeria      14389510.      5.16 3.14e-  7
## 2 Africa    Benin         4657484.     -2.53 1.18e-  2
## 3 Africa    Botswana      1244522.    -23.4  2.05e- 80
## 4 Africa    Burkina Faso   953956.    -63.7  6.79e-265
## 5 Africa    Burundi        218688.   -135.   4.56e-246
```

The results reject the null ($p < 0.05$ for many countries), indicating substantial heterogeneity within regions. For example, within Africa, Botswana shows significantly higher U.S. trade intensity than the regional average, while Burkina Faso and Comoros are substantially lower. This suggests that regional averages mask strong country-level variation, motivating a country-specific panel model.

3. Data Construction and Panel Setup

To explore these differences, a panel dataset was constructed at the country × commodity × year level. Variables were sourced from the World Bank, U.S. trade statistics, and supplementary databases:

Variable	Description	Notes
`sum_EX , sum_IM`	Total U.S. exports/imports (USD)	Nominal value
`Population`	Partner-country population	Log-transformed
`GDP`	Gross Domestic Product	USD, current prices
`politic_rank`	Political stability index	Higher = more stable
`capital_distance`	Distance from Washington D.C. (km)	Time-invariant
`CPI_Annual`	Consumer Price Index	Used for deflation

```
##                                                Country              Commodity Year
## Afghanistan.1111 Oilseeds & Grains-2015 Afghanistan 1111 Oilseeds & Grains 2015
## Afghanistan.1111 Oilseeds & Grains-2016 Afghanistan 1111 Oilseeds & Grains 2016
##                                         sum_IM  sum_EX capital_distance
## Afghanistan.1111 Oilseeds & Grains-2015   2916 7246528         11155.07
## Afghanistan.1111 Oilseeds & Grains-2016   4236       0         11155.07
##                                          Population politic_rank CPI_Annual
## Afghanistan.1111 Oilseeds & Grains-2015 -0.05559540    -1.756111   237.0170
## Afghanistan.1111 Oilseeds & Grains-2016 -0.05037155    -1.756111   240.0072
##                                         GDP_Current_USD real_EX  real_IM
## Afghanistan.1111 Oilseeds & Grains-2015     1.82950e+13 3057387 1230.291
## Afghanistan.1111 Oilseeds & Grains-2016     1.88049e+13       0 1764.947
##                                                              country_commodity
## Afghanistan.1111 Oilseeds & Grains-2015 Afghanistan.1111 Oilseeds & Grains
## Afghanistan.1111 Oilseeds & Grains-2016 Afghanistan.1111 Oilseeds & Grains
```

This structure allows for both within-country variation over time and cross-country differences. For instance, we can examine whether growth in a country's population or political stability corresponds to increased U.S. exports to that market.

4. Fixed-Effects Model: Within-Country Determinants

To study how changes inside a country relate to changes in U.S. exports to that country, we estimate a country fixed-effects model:

$$Exports_{it} = \alpha_i + \beta_1 Population_{it} + \beta_2 PoliticRank_{it} + \varepsilon_{it}$$

where αi captures each country's baseline trade potential with the U.S.

Result Summary

```
##
## t test of coefficients:
##
##               Estimate Std. Error t value  Pr(>|t|)
## Population   460024345   78075785  5.8920 3.826e-09 ***
## politic_rank -23948066    8763546 -2.7327  0.006283 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results (exports as the dependent variable):
Population: Positive and statistically significant. As a country's population grows, U.S. exports to that country tend to rise.
Interpretation: bigger market → more demand for U.S. goods.
Political stability: Negative and significant in this specification. One interpretation is that as some countries stabilize and mature, they may substitute U.S. imports with more domestic production or alternative suppliers. Another interpretation is that the U.S. sometimes exports more aggressively into less-stable but rapidly growing markets. This effect will weaken once we add controls.
Residual plots for this model show non-constant variance across countries and some non-normality, indicating heteroskedasticity. We address this in Section 6.

5. Random-Effects Model:

Fixed effects are great for causal interpretation of within-country changes, but they drop time-invariant variables. In trade, one of the most important time-invariant factors is distance.
To bring distance in, we estimate a random-effects (RE) model:

$$Exports_{it} = \alpha + u_i + \beta_1 Population_{it} + \beta_2 PoliticRank_{it} + \beta_3 CapitalDistance_i + \varepsilon_{it}$$

ui is a random country-specific intercept.
CapitalDistance enters here because it does not change over time and would be absorbed by fixed effects.
We also include commodity-category controls so that we're comparing trade within similar product classes rather than mixing, say, raw agricultural goods with aircraft parts.

```
## # A tibble: 3 × 5
##   term              estimate std.error statistic  p.value
##   <chr>                <dbl>     <dbl>     <dbl>    <dbl>
## 1 Population       119776515. 15308367.      7.82 5.15e-15
## 2 politic_rank      -5433871.  3383451.     -1.61 1.08e- 1
## 3 capital_distance    -23073.     2514.     -9.18 4.47e-20
```

Key findings:
Population: Still positive. Larger / growing markets tend to import more from the U.S.
Capital distance: Strongly negative. Countries farther from the United States trade less with the U.S., holding market size constant. This matches standard gravity-model logic: trade falls with distance due to shipping cost, logistics frictions, and weaker supply-chain integration.
Political stability: Once population and distance are included, stability becomes much weaker as a direct predictor of U.S. exports. That suggests stability may work indirectly (through GDP growth, investment climate, etc.) rather than directly determining trade flows.
Economically, this tells a consistent story: U.S. export intensity scales with market size and decays with distance.

6. Model Diagnostics and Robustness.

The fixed-effects model assumes each country can have its own baseline level of trade with the U.S. and does not assume that baseline is “unrelated” to the regressors.
The random-effects model is more efficient, but only if the country-specific effect is not correlated with the regressors. We formally compare the two using a Hausman test:
Null (RE is valid): The unobserved country component is uncorrelated with the regressors → random effects is consistent. Alternative (FE preferred): The unobserved country component is correlated with the regressors → only fixed effects is consistent.
In our data, the Hausman test rejects the null ($p < 0.05$). This means there is evidence that country-specific unobservables are correlated with things like population and political stability. In plain language: “who the country is” is not random.
Implication:
We treat the fixed-effects model as the primary causal specification for interpreting within-country changes. We use the random-effects / gravity model as a complementary descriptive model that allows us to analyze the role of time-invariant drivers like distance.

```
##
##  Hausman Test
##
## data:  sum_EX ~ Population + politic_rank + capital_distance + factor(Commodity)
## chisq = 193.05, df = 2, p-value < 2.2e-16
## alternative hypothesis: one model is inconsistent
```

7. Diagnostics and Inference

We ran diagnostics primarily on the fixed-effects model, since Hausman indicates that is our main consistent estimator. Findings:
Heteroskedasticity: Residual variance differs across countries and over time.
Non-normal residuals: Residuals are skewed for some country-years, suggesting occasional large shocks (e.g. one-off defense or energy contracts).
Within-country correlation: Errors within a country can be correlated over time.
Response:
We report heteroskedasticity-robust standard errors clustered at the country level.
Clustering lets each country have its own error variance pattern and serial correlation.
We also compared naive vs robust standard errors for key variables (Population, Distance) and found that significance and sign remained stable. So the main conclusions are not driven by fragile inference.

8. Interpretation

Putting the models together:

1. Market size matters.. Within a given country, as its population increases, U.S. exports to that country tend to increase. The U.S. sells more into larger and growing markets.

2. Geography still matters.. Even after controlling for market size and product category, physical distance from Washington, D.C. is strongly negatively related to export value. Distance is still a friction in global trade.

3. Political stability is secondary. Stability shows significance in the simple fixed-effects model but weakens once we control for market size and distance. This suggests that “being politically stable” alone doesn’t guarantee strong U.S. trade flows if the country is small or far away. Instead, stability may operate through other channels (e.g. it helps GDP grow, which then drives imports). Policy-wise: The U.S. tends to export most to large, nearby, integrated economies. Distance and scale still dominate.

10. Conclusion

We find evidence consistent with a gravity model of trade: U.S. exports are higher to partners that are (i) larger markets and (ii) closer in distance. Political stability alone is not a dominant predictor once market size and distance are considered.

The Hausman test indicates that fixed effects are the appropriate main specification for inference, since unobserved country traits are correlated with observed regressors. Random effects are still useful descriptively because they allow us to include time-invariant distance and see the classic distance penalty.