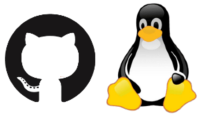
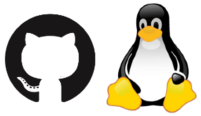
Linux & GitHub

Overview



1. Context
2. GitHub
3. Research & Predictions
4. Data
5. Explorative Analysis
6. Conclusion
7. Limitations
8. Takeaway Questions



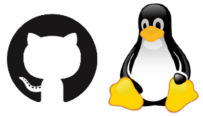
University of Exeter

1. Context

1: Context

- UN Comtrade Global Trade dataset
  - 2014-2025
  - 200+ Countries
  - Merchandise/Goods Trade
  - Imports & Exports by year

- Most reputable source for trade analysis, used by researchers and policymakers



University of Exeter

2. GitHub

2: GitHub

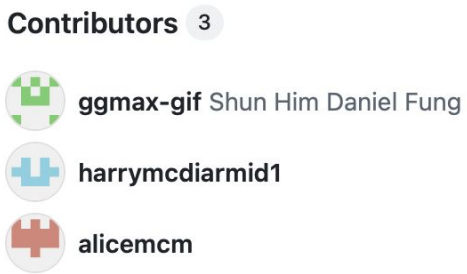
GitHub was used to ensure our analysis was collaborative, transparent, and reproducible.

Why GitHub?

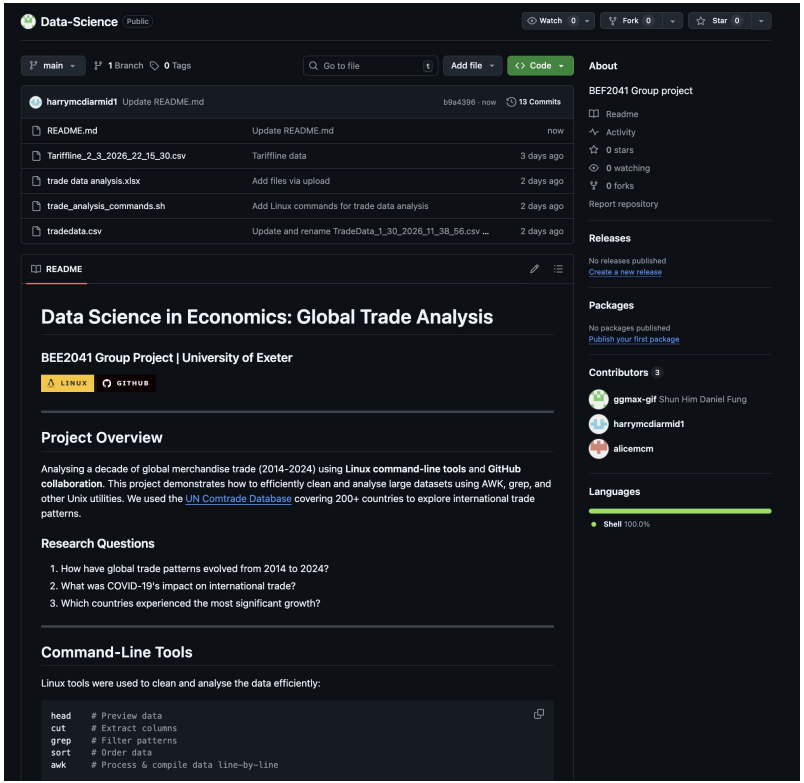
- Easy version control and collaboration
  - Codespaces allow everyone to work on the project simultaneously without overwriting each other
- Complete project history visible at a glance
  - Able to restore and backup work easily
- GitHub Desktop removes the need to memorise commands
- Built in LLM assistant (co-pilot) for troubleshooting
- Community of 100M users (larger than any competitor)
  - Plenty of public datasets and forum discussions

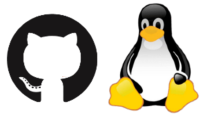
Key Git commands used:

- git push
- git pull
- git commit



2: GitHub


Data-Science
Public
Watch 0
Fork 0
Star 0
main
1 Branch
0 Tags
Go to file
Add file
Code
About
BEF2041 Group project
Readme
Activity
0 stars
0 watching
0 forks
Report repository
harrymcdiarmid1 Update README.md
b9a4396 · now
13 Commits
README.md Update README.md now
Tariffline_2_3_2026_22_15_30.csv Tariffline data 3 days ago
trade data analysis.xlsx Add files via upload 2 days ago
trade_analysis_commands.sh Add Linux commands for trade data analysis 2 days ago
tradedata.csv Update and rename TradeData_1_30_2026_11_38_56.csv ... 2 days ago
README
Data Science in Economics: Global Trade Analysis
BEE2041 Group Project | University of Exeter
LINUX
GITHUB
Project Overview
Analysing a decade of global merchandise trade (2014-2024) using Linux command-line tools and GitHub collaboration. This project demonstrates how to efficiently clean and analyse large datasets using AWK, grep, and other Unix utilities. We used the UN Comtrade Database covering 200+ countries to explore international trade patterns.
Research Questions
1. How have global trade patterns evolved from 2014 to 2024?
2. What was COVID-19's impact on international trade?
3. Which countries experienced the most significant growth?
Command-Line Tools
Linux tools were used to clean and analyse the data efficiently:
head # Preview data
cut # Extract columns
grep # Filter patterns
sort # Order data
awk # Process & compile data line-by-line
Releases
No releases published
Create a new release
Packages
No packages published
Publish your first package
Contributors 3
ggmax-gif Shun Him Daniel Fung
harrymcdiarmid1
alicemcm
Languages
Shell 100.0%



University of Exeter

3. Research & Predictions

3: Research Questions & Predictions

1. How have global trade patterns changed from 2014 to 2024?

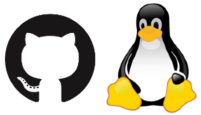
- Higher volume (+30% from 2024 vs 2014) with trade being more concentrated among the wealthier countries.

2. What was COVID-19's impact on global trade?

- Sharp decline (15-20% drop from 2019 levels) with recovery after 2021.

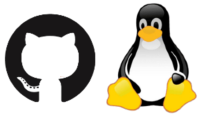
3. Which countries trade grew the most over the decade?

- China, Vietnam, USA, India, Germany. Mixture of established trading powerhouses and strong emerging economies.



4. Data

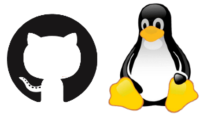
4.1: Data - Exploration



University of Exeter

1. `head -5 tradedata.csv`
2. `wc -l tradedata.csv`
3. `head -1 tradedata.csv | awk -F',' '{print NF}'`
4. `tail -n +2 tradedata.csv | cut -d',' -f4 | sort -u`
5. `tail -n +2 tradedata.csv | cut -d',' -f4 | sort | uniq -c`
6. `tail -n +2 tradedata.csv | cut -d',' -f4 | sort -u | wc -l`

4.1: Data - Exploration

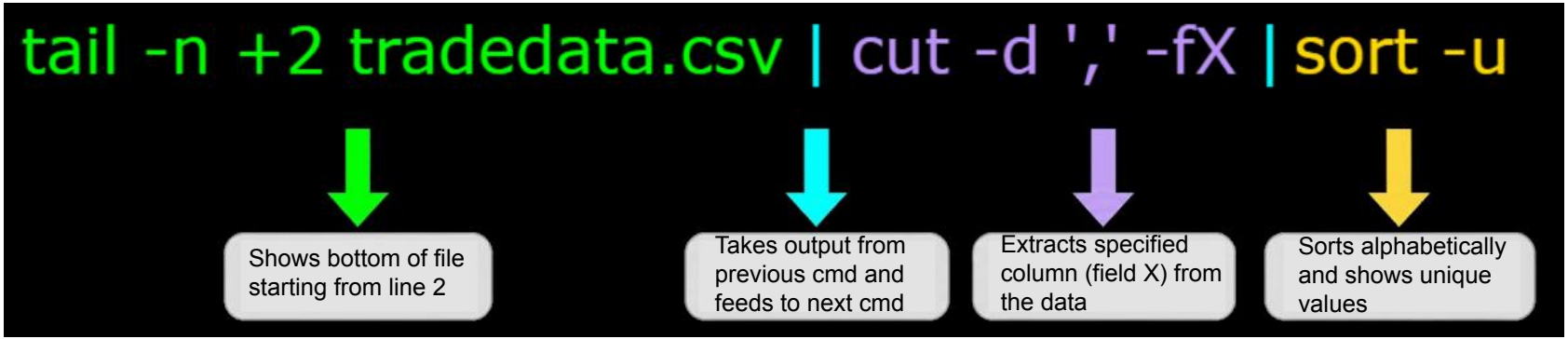


University of Exeter

```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ head -5 tradedata.csv
typeCode,freqCode,refPeriodId,refYear,refMonth,period,reporterCode,reporterISO,reporterDesc,flowCode,flowDesc,partnerCode,partnerISO,partnerDesc,partner2Code,partner2ISO,partner2Desc,classificationCode,classificationSearchCode,isOrigin
alClassification,cmdCode,cmdDesc,aggrLevel,isLeaf,customsCode,customsDesc,mosCode,motCode,motDesc,qtyUnitCode,qtyUnitAbbr,qty,isQtyEstimated,altQtyUnitCode,altQtyUnitAbbr,altQty,isAltQtyEstimated,netWgt,isNetWgtEstimated,grossWgt,isGro
ssWgtEstimated,cifvalue,fobvalue,primaryValue,legacyEstimationFlag,isReported,isAggregate
C,A,20140101,2014,52,2014,4,AFG,Afghanistan,M,Import,0,W00,World,0,W00,World,H2,HS,TRUE,TOTAL,All Commodities,0,FALSE,C00,TOTAL CPC,0,0,TOTAL MOT,-1,N/A,,FALSE,-1,N/A,,FALSE,,FALSE,,FALSE,7697178170,,7697178170,0,TRUE,FALSE
C,A,20140101,2014,52,2014,4,AFG,Afghanistan,X,Export,0,W00,World,0,W00,World,H2,HS,TRUE,TOTAL,All Commodities,0,FALSE,C00,TOTAL CPC,0,0,TOTAL MOT,-1,N/A,,FALSE,-1,N/A,,FALSE,,FALSE,,FALSE,,570534007,570534007,0,TRUE,FALSE
C,A,20140101,2014,52,2014,8,ALB,Albania,M,Import,0,W00,World,0,W00,World,H4,HS,TRUE,TOTAL,All Commodities,0,FALSE,C00,TOTAL CPC,0,0,TOTAL MOT,-1,N/A,,FALSE,-1,N/A,,FALSE,,FALSE,,FALSE,5229972238,,5229972238,0,TRUE,FALSE
C,A,20140101,2014,52,2014,8,ALB,Albania,X,Export,0,W00,World,0,W00,World,H4,HS,TRUE,TOTAL,All Commodities,0,FALSE,C00,TOTAL CPC,0,0,TOTAL MOT,-1,N/A,,FALSE,-1,N/A,,FALSE,,FALSE,,FALSE,,2430723644,2430723644,0,TRUE,FALSE
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ wc -l tradedata.csv
3636 tradedata.csv
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ head -1 tradedata.csv | awk -F',' '{print NF}'
47
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f4 | sort -u
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f4 | sort | uniq -c
    336 2014
    340 2015
    347 2016
    354 2017
    350 2018
    338 2019
    332 2020
    332 2021
    326 2022
    320 2023
    258 2024
      2 2025
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f4 | sort -u | wc -l
12
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$
```

4.2: Data - Cleaning

Variable Name	Description
refYear	Year of Trade
reporterDesc	Country Name
flowDesc	Import/Export direction
primaryValue	Trade value (USD)



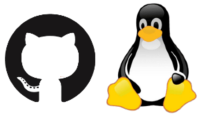
4.2: Data - Cleaning

```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f1 | sort -u
C
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f2 | sort -u
A
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f5 | sort -u
52
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f12 | sort -u
0
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f14 | sort -u
World
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f21 | sort -u
TOTAL
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f22 | sort -u
All Commodities
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f23 | sort -u
0
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f24 | sort -u
FALSE
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f29 | sort -u
TOTAL MOT
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f30 | sort -u
-1
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f32 | sort -u

0
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ tail -n +2 tradedata.csv | cut -d',' -f38 | sort -u

0
```

4.2: Data - Cleaning



University of Exeter

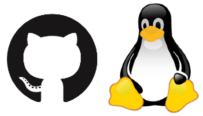
7. cut -d',' -f4,9,11,44 tradedata.csv > cleantradedata.csv

8. head -5 cleantradedata.csv

9. grep ",," cleantradedata.csv | wc -l

4.2: Data - Cleaning

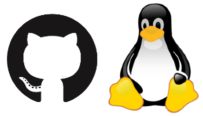

```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ cut -d',' -f4,9,11,44 tradedata.csv > cleantradedata.csv
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ head -5 cleantradedata.csv
refYear,reporterDesc,flowDesc,primaryValue
2014,Afghanistan,Import,7697178170
2014,Afghanistan,Export,570534007
2014,Albania,Import,5229972238
2014,Albania,Export,2430723644
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ grep ",," cleantradedata.csv | wc -l
0
```



University of Exeter

5. Explorative Analysis

5.1: Explorative Analysis



University of Exeter

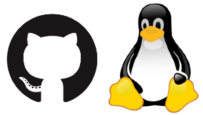
10. awk -F',' '$3 ~ /Export/ && $1 ~ /2014/ {sum+=$4} END {print sum}' cleantradedata.csv

11. awk -F',' '$3 ~ /Export/ && $1 ~ /2024/ {sum+=$4} END {print sum}' cleantradedata.csv

12. awk -F',' '$3 ~ /Import/ && $1 ~ /2014/ {sum+=$4} END {print sum}' cleantradedata.csv

13. awk -F',' '$3 ~ /Import/ && $1 ~ /2024/ {sum+=$4} END {print sum}' cleantradedata.csv

5.1: Explorative Analysis



Processes line by line

Sets comma as field separator (CSV)

Finds data containing specified trade direction in column 3

Finds data containing year in column 1

```
awk -F',' '$3 ~ /TRADEDIRECTION/ && $1 ~ /YEAR/ {sum+=$4} END {print sum}' cleantradedata.csv
```

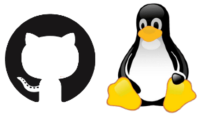
Adds column 4 value to total

Prints total after all rows processed

5.1: Explorative Analysis

```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ awk -F',' '$3 ~ /Export/ && $1 ~ /2014/ {sum+=$4} END {print sum}' cleantradedata.csv
1.76371e+13
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ awk -F',' '$3 ~ /Export/ && $1 ~ /2024/ {sum+=$4} END {print sum}' cleantradedata.csv
2.0711e+13
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ awk -F',' '$3 ~ /Import/ && $1 ~ /2014/ {sum+=$4} END {print sum}' cleantradedata.csv
1.76182e+13
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ awk -F',' '$3 ~ /Import/ && $1 ~ /2024/ {sum+=$4} END {print sum}' cleantradedata.csv
2.12474e+13
```

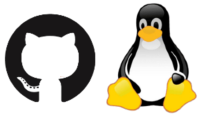
5.2: Explorative Analysis



University of Exeter

```
14. for year in 2014 2024; do
    echo "$year Imports"
    tr -d "" < cleantradedata.csv | \
    awk -F',' -v y="$year" '$1==y && $3=="Import" {t+=$4; c[$2]+=$4} END {for(x in c) print c[x]"|"x"|"t}' | \
    sort -t'|' -k1 -rn | head -10 | \
    awk -F'|' '{printf "%2d. %-20s %.2f%%\n", NR, $2, ($1/$3)*100}'
    echo ""
done
```

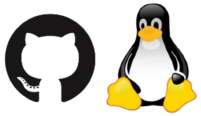
5.2: Explorative Analysis




```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ for year in 2014 2024; do
orts"
    tr -d>     echo "$year Imports"
>     tr -d '"' < cleantradedata.csv | \
>     awk -F',' -v y="$year" '$1==y && $3=="Import" {t+=$4; c[$2]+=$4} END {for(x in c) print c[x]"|"x"|"t}' | \
>     sort -t'|' -k1 -rn | head -10 | \
>     awk -F'|' '{printf "%2d. %-20s %.2f%%\n", NR, $2, ($1/$3)*100}'
>     echo ""
> done
2014 Imports
 1. USA                  13.68%
 2. China                11.12%
 3. Germany              6.90%
 4. Japan                4.61%
 5. United Kingdom       3.94%
 6. France               3.75%
 7. Rep. of Korea        2.98%
 8. Netherlands          2.88%
 9. Italy                2.69%
10. Canada               2.63%

2024 Imports
 1. USA                  15.80%
 2. China                12.17%
 3. Germany              6.48%
 4. United Kingdom       3.81%
 5. France               3.59%
 6. Japan                3.50%
 7. India                3.28%
 8. Rep. of Korea        2.97%
 9. Mexico               2.95%
10. Netherlands          2.92%
```

5.3: Explorative Analysis

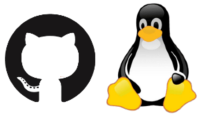



```
15. for year in 2016 2017 2018 2019 2020 2021 2022 2023; do
    exports=$(awk -F',' -v y="$year" '$1 ~ y && $3 ~ /Export/ {sum+=$4} END {print sum}' cleantradedata.csv)
    imports=$(awk -F',' -v y="$year" '$1 ~ y && $3 ~ /Import/ {sum+=$4} END {print sum}' cleantradedata.csv)
    echo "$year - Exports: $exports, Imports: $imports"
done
```

5.3: Explorative Analysis

```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ for year in 2016 2017 2018 2019 2020 2021 2022 2023; do
rts=$(>     exports=$(awk -F',' -v y="$year" '$1 ~ y && $3 ~ /Export/ {sum+=$4} END {print sum}' cleantradedata.csv)
>     imports=$(awk -F',' -v y="$year" '$1 ~ y && $3 ~ /Import/ {sum+=$4} END {print sum}' cleantradedata.csv)
>     echo "$year - Exports: $exports, Imports: $imports"
> done
2016 - Exports: 1.49276e+13, Imports: 1.50748e+13
2017 - Exports: 1.64395e+13, Imports: 1.66962e+13
2018 - Exports: 1.80993e+13, Imports: 1.84016e+13
2019 - Exports: 1.75872e+13, Imports: 1.78977e+13
2020 - Exports: 1.62805e+13, Imports: 1.64952e+13
2021 - Exports: 2.05601e+13, Imports: 2.08645e+13
2022 - Exports: 2.24738e+13, Imports: 2.3507e+13
2023 - Exports: 2.15988e+13, Imports: 2.21178e+13
```

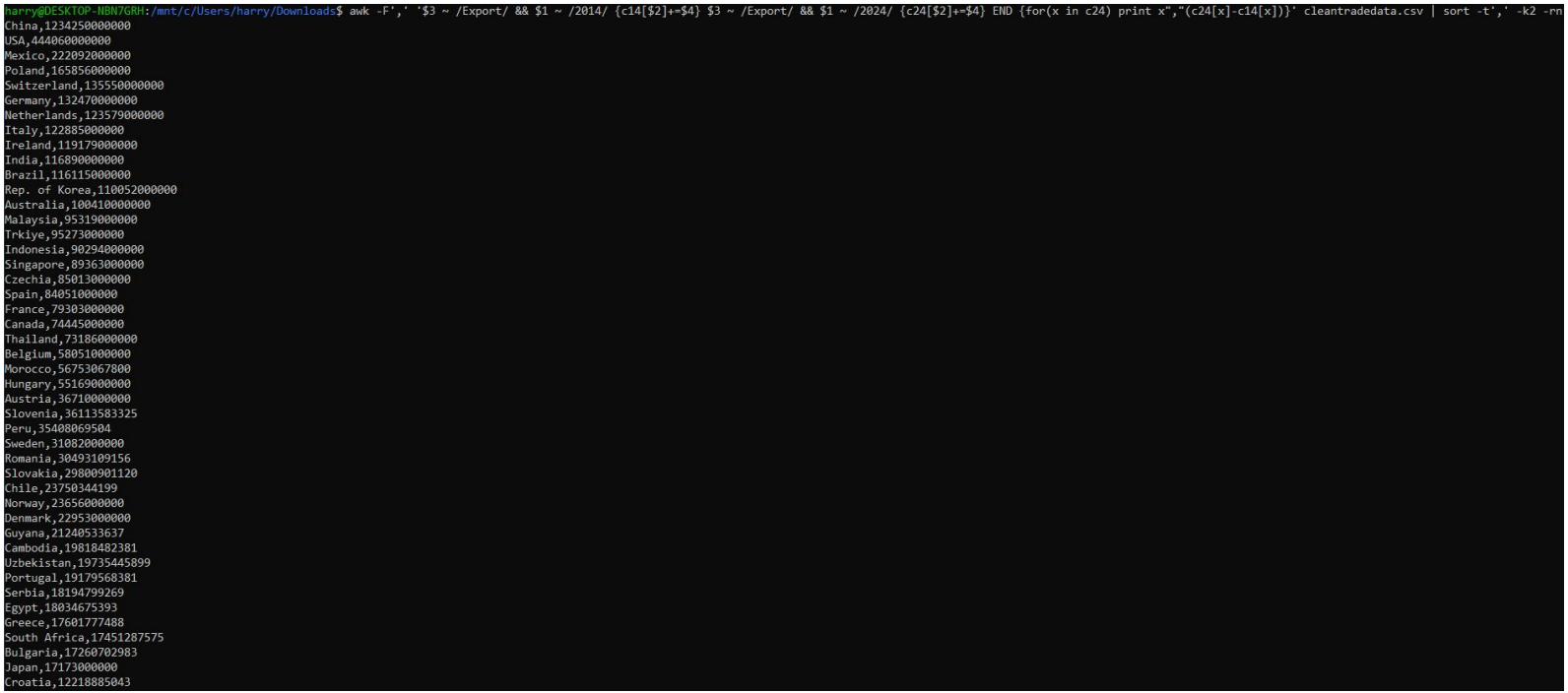
5.3: Explorative Analysis



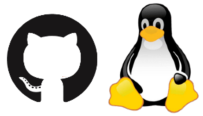
16. awk -F',' '$3 ~ /Export/ && $1 ~ /2014/ {c14[$2]+=$4} $3 ~ /Export/ && $1 ~ /2024/ {c24[$2]+=$4} END {for(x in c24) print x","(c24[x]-c14[x])}' cleantradedata.csv | sort -t',' -k2 -rn

17. awk -F',' '$3 ~ /Import/ && $1 ~ /2014/ {c14[$2]+=$4} $3 ~ /Import/ && $1 ~ /2024/ {c24[$2]+=$4} END {for(x in c24) print x","(c24[x]-c14[x])}' cleantradedata.csv | sort -t',' -k2 -rn

5.3: Explorative Analysis



```
harry@DESKTOP-NBN7GRH:/mnt/c/Users/harry/Downloads$ awk -F',' '$3 ~ /Export/ && $1 ~ /2014/ {c14[$2]+=$4} $3 ~ /Export/ && $1 ~ /2024/ {c24[$2]+=$4} END {for(x in c24) print x","(c24[x]-c14[x])}' cleantradedata.csv | sort -t',' -k2 -rn
China,1234250000000
USA,444060000000
Mexico,222092000000
Poland,165856000000
Switzerland,135550000000
Germany,132470000000
Netherlands,123579000000
Italy,122885000000
Ireland,119179000000
India,116890000000
Brazil,116115000000
Rep. of Korea,110052000000
Australia,100410000000
Malaysia,95319000000
Trkiye,95273000000
Indonesia,90294000000
Singapore,89363000000
Czechia,85013000000
Spain,84051000000
France,79303000000
Canada,74445000000
Thailand,73186000000
Belgium,58051000000
Morocco,56753067800
Hungary,55169000000
Austria,36710000000
Slovenia,36113583325
Peru,35408069504
Sweden,31082000000
Romania,30493109156
Slovakia,29800901120
Chile,23750344199
Norway,23656000000
Denmark,22953000000
Guyana,21240533637
Cambodia,19818482381
Uzbekistan,19735445899
Portugal,19179568381
Serbia,18194799269
Egypt,18034675393
Greece,17601777488
South Africa,17451287575
Bulgaria,17260702983
Japan,17173000000
Croatia,12218885043
```



6. Conclusion

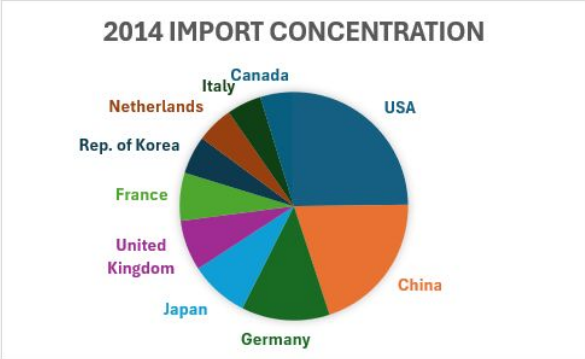
6.1: Conclusion: Evolution of Trade

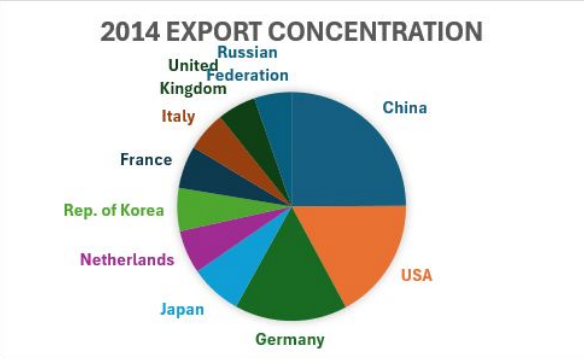
	Exports ($B)	Imports ($B)	Trade Balance ($B)
2014	17,637.1	17,618.2	18.9
2024	20,711.0	21,247.4	(536.4)
% Change	17.4%	20.6%	
Predicted % Change	30%	30%	
Variance	**(12.57%)**	**(9.40%)**	

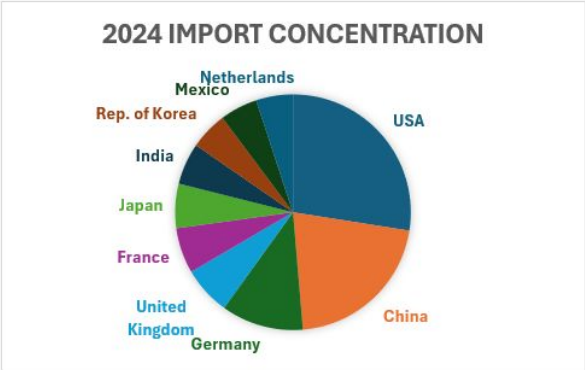
6.1: Conclusion: Evolution of Trade

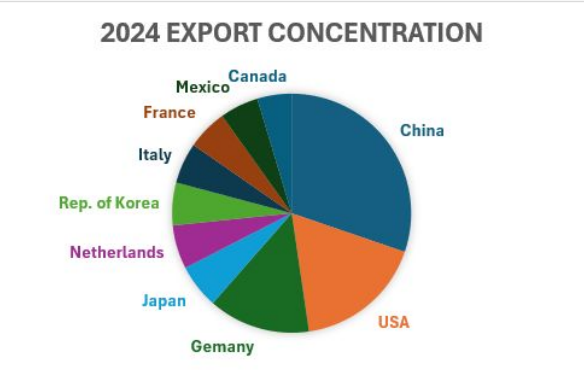
	Import Concentration		Export Concentration	
2014	USA	13.68%	China	13.28%
	China	11.12%	USA	9.18%
	Germany	6.90%	Germany	8.49%
	Japan	4.61%	Japan	3.91%
	United Kingdom	3.94%	Netherlands	3.26%
	France	3.75%	Rep. of Korea	3.25%
	Rep. of Korea	2.98%	France	3.21%
	Netherlands	2.88%	Italy	3.00%
	Italy	2.69%	United Kingdom	2.90%
	Canada	2.63%	Russian Federation	2.82%
	Total	**55.18%**	**Total**	**53.30%**
2024	Country			
	USA	15.80%	China	17.27%
	China	12.17%	USA	9.96%
	Germany	6.48%	Gemany	7.87%
	United Kingdom	3.81%	Japan	3.42%
	France	3.59%	Netherlands	3.38%
	Japan	3.50%	Rep. of Korea	3.30%
	India	3.28%	Italy	3.15%
	Rep. of Korea	2.97%	France	3.12%
	Mexico	2.95%	Mexico	2.99%
	Netherlands	2.92%	Canada	2.65%
	Total	**57.47%**	**Total**	**57.11%**

6.1: Conclusion: Evolution of Trade







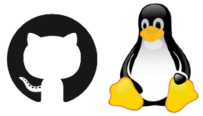


6.2: Conclusion: Impact of Covid-19

Year	Global Exports ($B)	Global Imports ($B)	Exports YoY Growth	Imports YoY Growth
2016	14,927.6	15,074.8		
2017	16,439.5	16,696.2	10.1%	10.8%
2018	18,099.3	18,401.6	10.1%	10.2%
2019	**17,587.2**	**17,897.7**	**(2.8%)**	**(2.7%)**
2020	**16,280.5**	**16,495.2**	**(7.4%)**	**(7.8%)**
2021	20,560.1	20,864.5	26.3%	26.5%
2022	22,473.8	23,507.0	9.3%	12.7%
2023	21,598.8	22,117.8	(3.9%)	(5.9%)

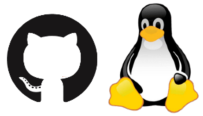
- Decline in trade growth 2019-2020 as expected. Not as sharp as the 15-20% prediction.
- Global trade decline seen in 2023 as well, suggesting unsustainable covid recovery growth

6.3: Conclusion: Growth Comparison



Country	Export Growth ($B)
China	1,234.3
USA	444.1
Mexico	222.1
Poland	165.9
Switzerland	135.6
Germany	132.5
Netherlands	123.6
Italy	122.9
Ireland	119.2
India	116.9

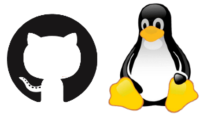
Country	Import Growth ($B)
USA	946.0
China	625.9
India	238.4
Mexico	225.9
Poland	162.8
Germany	162.0
Italy	127.9
United Kingdom	114.9
Netherlands	112.9
Rep. of Korea	106.2



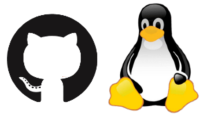
University of Exeter

7. Limitations

7: Limitations

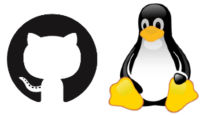


- **Missing data** - Several smaller countries have very little or no data present after 2019
- **Limited data granularity** - No product breakdown or ability to hypothesis test based on initial claims
- **Potential unreliable reporting** - Data from corrupt states may not be accurately reported
- **Digital trade not captured** - Goods/merchandise trade only, excludes services
- **Absolute growth only** - Growth data not showing percentage growth
- **Limited Analysis scope** - Dataset too large to perform analysis on every country, restricted to only top 10
- **Outdated data** - No 2025 data available
- **Currency conversion effects** - All values in USD not local currency, distorting true trading volumes



8. Takeaway Questions

8. Takeaway Questions



University of Exeter

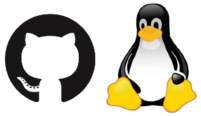
1. What key factors have made GitHub the dominant platform for developer collaboration compared to competitors?

- User base size
- Microsoft acquisition
- Pricing

3. How does the command-line efficiency used compare to Python/R for datasets of this size?

- Processing speed
- Memory usage
- Scalability
- Complexity of Python vs command-line

8. Takeaway Questions



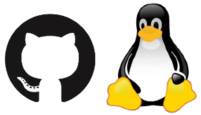
4. How does the choice of base year affect measured trade growth, e.g. comparing growth from 2014-2024 vs 2016-2024?

- Base year effects (recession/boom)
- Use awk to sum trade value for each year to compare

5. Several emerging economies (India, Vietnam, Poland) increased their trade significantly. What economic policies or structural changes drove such rapid growth?

- Free trade agreements
- Exchange rate movements
- Infrastructure investment

8.2 Takeaway Questions



University of Exeter

6. Which countries switched from trade surplus to deficit (or visa versa) over the decade?

- Use awk to calculate trade balances (exports - imports)
- Filter for countries where the sign changed and use sort to sort by magnitude

7. Theoretically every import is another countries export yet global trade balance for 2024 was -$536.4bn. What creates this imbalance and how can it be reconciled?

- Trade timing differences
- Valuation methods
- Currency conversion timing differences

Q&A

Github