**OECD States of Fragility - Review**

Institute for Economics and Peace (IEP)

09 June 2022

Contents

[**Background and purpose of this report** 1](#_Toc105686631)

[**Issues identified in scripts** 1](#_Toc105686632)

[**APPENDIX A** 5](#_Toc105686633)

[**APPENDIX B** 6](#_Toc105686634)

## **Background and purpose of this report**

OECD has requested IEP to review the OECD States of Fragility framework to ensure the munge and src code is outputting results correctly. OECD also requests specific variables in the master list be scrutinised for correct measures and transformations.

IEP has reviewed each munge script individually and added a number of checks to each file. These include

* Checking for the existence of duplicate country/year pairs
* Checking for the correct conversion of country to iso3c codes
* Checking for consistent coverage of countries in the latest year
* Checking for NA in columns

The checks are outputted to a log file (a text file detailing the precise issue and the script where the issue occurred). Where appropriate, duplicate entries or data missing for the latest year has also been outputted to an excel file.

## **Issues identified in scripts**

**01-WDI.R**

* Multiple countries are missing data for the latest year for variables.

**02-exchange rate volatility**

* **Does the measure of volatility make sense?** 
  + The use of a rolling average of standard deviation as a measure of volatility is intuitive and defensible from a statistical perspective. An alternative method of calculating volatility is to weigh the standard deviations over the years with greater emphasis given to the most recent years. This is a similar approach to how IEP weights the impact of terrorism over time.
* **Issues identified by log**
  + The logfile identified that multiple countries could not successfully be converted to isocodes. The issue seems to be caused by changes in country names over time. For example, “Yemen Arab Rep.” and “Yemen People's Dem. Rep. of”.
  + The logfile also identified missing values in the final data frame. This reflected missing data for early years and unsuccessful conversion of isocodes.

**03-aid-volatility**

* **Does the measure of volatility make sense with the adjustment of CPA.GNI or should we use a different coefficient of variation?**
  + The rolling mean of the cpa\_gni is calculated over 3 years and the standard deviation is calculated based on a rolling average of 5 years. Is this what is intended?
  + There are currently two different functions used to calculate the rolling mean. Specifically, `roll\_mean` (from the roll package) and the `rollmeanr` function (from the zoo package). Both appear to produce the same results and are functionally equivalent. However, the `rollmean` (from the zoo package) is functionally different to `roll\_mean` (from the roll package). This difference may result in undetected errors when changing code.
* **Issues Identified by log**
  + The gni-raw dataset contains values for South Sudan (SSD) and the other\_raw also contains values for gni for South Sudan (SSD). When rbinding the datasets, this results in duplicate values for SSD for certain years.
  + A temporary fix has been implemented to group by isocode, year and variablename and then take the mean of the values in the dat\_final. The issue can be fixed by selecting the desired values from the appropriate dataset.
  + The gni values for SSD in 2011 are either 0.5x the GNI in 2020 or 1.5x it depending on which value is used.

04- aid **dependency**

* ???

**05-debt-to-GDP-ratio**

* **Does the exclusion of estimates make sense?** 
  + The exclusion of estimated data does make sense.
* **Issues identified by log**
  + The log identified that only Dominica has data for 2021. Additionally, only 58 countries have data until 2020.

**06-inflation**

* **Issues identified by log**
  + The log identified missing data in 2022 for several countries as well as missing values for several countries in 2021. Based on the filename of the csv, it appears to be world bank data: <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG> The most recent dataset has been updated with 2021 data for several countries.

**07-corruption**

* **Issues identified by log**
  + The log identified missing data for Brunei Darussalam in 2021. This is not expected to impact the analysis significantly.

**08-coefficient-ofhuman-inequality**

* Checks have been implemented

**09-INFORM**

* The most recent data is in 2020. However, in the link below, time series data for the INFORM index includes 2021 data (though it is labelled as “TEMP”). <https://drmkc.jrc.ec.europa.eu/inform-index/INFORM-Risk/Results-and-data/moduleId/1782/id/433/controller/Admin/action/Results#inline-nav-1>

**10-Political Rights**

* **Issues identified by log**
  + Puerto Rico is missing data for the latest year. This is not expected to impact analysis significantly.
  + There are duplicate isocodes for Somalia/Somaliland (SOM) and west bank/gaza (PSE). A temporary fix has been implemented which takes the mean of the respective values in the affected years.

**12-Income inequality**

* **Issues identified by log**
  + The log identified that Serbia and Montenegro were not converted properly.
  + The log identified a large amount of missing data for 100+ countries in 2019. There is only data for 51 countries for that year in the raw data.

**14- criminal networks**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions.

**15 – VDEM**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions.

**16 – media-freedom**

* **Issues identified by log**
  + The log identified missing data for Puerto Rico in 2021
  + The log identified duplicate isocodes for Somalia/Somaliland (SOM) and west bank/gaza (PSE). A temporary fix has been implemented which takes the mean of the respective values in the affected years.

**17 – SPIRI**

* **Check directionality. Higher values should be more fragility**
  + The interpretation of higher values indicating more fragility is acceptable as a heuristic. However, it is difficult to objectively state whether more or less military spending is a good thing. This interpretation can lead to strange conclusions such as Singapore being more fragile than Zimbabwe. This interpretation does not take into account reasons for military expenditure. Some countries may spend large amounts on money on military, while others may not because they need to spend it on other things.
  + A suggestion may be to look at standard deviation away from the regional/global mean.
* **Issues identified by log**
  + The log identified 13 missing countries in the latest year. The data does not appear to exist for these countries in the raw data.

**18 dietary energy supply**

* **Issues identified by log**
  + The log identified missing data for Bermuda in 2019. This is unlikely to significantly impact analysis
  + There is no data for 2020 or 2021

**19 water stress**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions. However, there is no 2020 or 2021 data.

**20 womens participation in UNFCCCR**

* **Issues identified by log**
  + The log identified missing data for San Marino in 2019
* **NOTE:**
  + I had a question regarding the meaning of the 1’s and 0’s in the data. For example, Algeria has a value of 34.6% in 2008 COP14 is the row value 0 and Angola has a value of 25% in the row v value 1. If these rows represent quantitatively different things, then this may cause issues.

**21 environmental displacement**

* **NOTE:** 
  + The log did not identify any significant issues with duplicates, data overage or isocode conversions. However, there is no 2021 data.
  + **According to the website, data should have been updated on 23rd May 2022.**

**22 interpersonal trust**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions.

**23 forced displacement by asylum**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions. However, there is no 2020 or 2021 data.

**24 forced displacement by country of origin**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions. However, there is no 2020 or 2021 data.

**25 Digital Security**

* **Check dimensionality higher values indicates less fragility.** 
  + The variable seems to include the number of arrests from online content. If this is indeed the case, then higher values likely indicate greater levels of fragility. This will impact whether it is appropriate to use the minimum or maximum value.
* The values for some countries across most/all years appear to be the same (e.g. Australia). This may reflect the raw data, but it might cause issues for certain countries if they are considered fragile (e.g. Afghanistan).
* The log did not identify any significant issues with duplicates, data overage or isocode conversions.

**26 ND gain index**

* **NOTE:** 
  + The raw data only contains data until 2019.
  + In the final dataset, year is a character rather than an integer.

**27 Non renewable resource crime**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions. However, there is only data for 2021 and there is no time series.

**28 Social protection**

* **Issues identified by log**
  + The log identified 11 countries with missing data in the latest year.
* **NOTE:**
  + The data only contains values until 2020

**29 EPI**

* **NOTE:**
  + The raw data only contains values for 2020.

**30 World Governance Indicators**

* **Issues identified by log**
  + The log identified missing data for 4 countries in latest year (2020).

**31 renewable energy**

* **Check if there are any missing or unusual values in the source data (e.g Central African Republic).** 
  + There are multiple instance of very high values. For example. 100% for Albania, Bhutan, Central African Republic, Iceland, Lesotho, Nepal, Paraguay, Ethiopia, as well as very high values for Congo, Namibia, Norway etc.
  + Likewise, there are multiple instance of very low values. For example,0% for Benin, Botswana, Chad, Eritrea, Estonia, Gambia (the). Etc
* **Issues identified by log**
  + The log identified missing data for around 138 countries in the latest year (2021).

**32 sigi**

* **Is it possible to create a composite value of the violence against women sub indicators related to law, attitudes, prevalence instead of only attitudes?** 
  + Code has been added to calculate a composite measure based on the combination of attitudes, practice/prevalence variables. Current weightings are arbitrary but can be changed depending on what weightings are desired.
* **Issues identified by log**
  + The log identified 46 null values in the value column. These may impact some countries that are classified as extremely fragile or other fragile.
* **NOTE:**
  + The dataset only has data for 2019. No time series.

**33 womens employment in agricultural sectors**

* **NOTE:**
  + The dataset only contains data until 2019. It may be worth checking exactly where the data is from to confirm proper country/year coverage. <https://tcdata360.worldbank.org/indicators/d87c3f0a?country=BRA&indicator=28109&viz=line_chart&years=2000,2019VS> <https://tcdata360.worldbank.org/indicators/5133645c?country=BRA&indicator=28111&viz=line_chart&years=1990,2013>

**34 tax revenue**

* **Issues identified by log**
  + The log identified large amounts of missing data (100 countries) for 2020.

**35 non state and one sided conflict**

* **NOTE:** 
  + The log calculation is implemented correctly. It might be useful to implement a similar approach to how IEP calculates the GTI score. Setting the log base to be the maximum value allows you to set the maximum score as the maximum value and should give more accurate results than simply taking the log.
    - max\_score = max(dat\_final$value)
    - log\_base = (max\_score+1)^(1/10)
    - dat\_final$value <- log(dat\_final$value + 1, log\_base)
  + The rolling mean seems to be implemented correctly and takes the average of the current and the previous 2 observations.

**36 battledeaths**

* The log did not identify any significant issues with duplicates, data overage or isocode conversions.

**999-data-finishing**

* Note the missing values for indicators in the data.matrix variable (lines 350)
* The directionality appears to be appropriately implemented (line 363)
* Is the data.matrix.rankings (line 367) seems to be taking the data for the latest year only. Is this intended?

**01-SFR Calculation**

* There may be an issue with the dimensional.pca.models.
  + The drops variable always seems to be 0 for each dimension. This may be fine and working as intended due to the value of the correlations not exceeding a certain threshold. However, on line 109, there may be a syntax error “all.drops <<-“ the conditional for length(drops) may not be executing properly when there are highly correlated indicators.
  + On line 100, should year == max(year)? This results in different values of x and y (the actual principal components), but same rotation values for each principal component across all years within each dimension (human, economic, political etc). is this intended?

**02 – SFR biplots**

* Given the addition of the new variables (aid dependency and 08-human inequality), does this impact the number of clusters per dimension? (i.e. should they all be 6)?
* Note: there are warnings when running the code on line 144 to line 152. These appear to relate to missing values

**03-SFR- Snai;**

* No apparent issues here

**04-SFR-cluster**

* Line 27 temp = temp %>% filter(year == max(year)) may not be working as expected. The maximum year is 2022 for the economic domain. (and others). For many datasets this might represent imputed data (calculated in the 999-data-finishing script near line 410. s this what is intended?

**05-Analysis**

* There appears to be an issue with the 05-Analysis Script. When it is run as a source script, it produced some errors around line 250 when renaming iso\_alpha3\_codes to is03c in the un\_m49\_codes. dataframe/datatable. However, when running it in the console, there were no issues.
  + The issue seems to stem from how windows vs other systems handles the filename of the UNSD Methodology csv. The actual filename in windows explorer has the character “—“ which R seems to have trouble interpreting as “-“ when running as source, but not when running through console. The issue seems to be solved by renaming the file and replacing the “—“ with an underscore “\_”. I have made a copy of the original csv file, renamed it and amended the import of the UNSD\_Methodology.csv in the 05-Analysis script to reflect this change.
* In the 05-Analysis script, there was a fix needed in line 1726 where Cote d'Ivoire due to a utf-8 character that could not be converted to asci. I have removed that particular character and changed the conversion for Cote d'Ivoire
* The code appears to run without overt errors, but there are warnings throughout.
* Sys.time() was altered from the 01-calculation, and at various points in scripts 04, cluster and 05- analysis overwrite = T was added to ensure the scripts could run properly.

## **APPENDIX A**

Below is a general summary of the changes to each script

* An additional variable (iso3c\_CONVERT) was added when converting iso3c codes to determine which countries were able to be converted successfully or not. The change was such that a new variable is created rather than iso3c being overwritten in most cases.
* Logging tests for checking duplicate country/year pairs, checking the number of NA’s, checking the country coverage were also added.
* Code was added to explicitly select the iso3c, value, year and variablename columns.
* Many comments were added throughout (and can be removed)
* In certain scripts 10-political rights and 16-media freedom, a group\_by summarise was added to explicitly fix the issue of country year pairs being duplicated.
* In 32-sigi code was changed to calculate a composite measure (the original code remains, but is commented out). The weightings are completely arbitrary and can be changed as desired.

## **APPENDIX B**

A comparison of the before and after rankings with the lowest score having the highest ranking. Most rankings change by 0,1 or 2 places in either direction. One exception is Palestine which improved from 39 in the original data to 44 in the new data. Importantly, no country that was considered extremely fragile or other fragile changed classifications.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **iso3c** | **country** | **original\_rank** | **new\_rank** | **difference\_rank** |
| SOM | Somalia | 1 | 2 | 1 |
| YEM | Yemen | 2 | 1 | -1 |
| AFG | Afghanistan | 3 | 3 | 0 |
| CAF | Central African Republic (the) | 4 | 4 | 0 |
| SSD | South Sudan | 5 | 6 | 1 |
| COD | Congo (the Democratic Republic of the) | 6 | 5 | -1 |
| TCD | Chad | 7 | 7 | 0 |
| SYR | Syrian Arab Republic (the) | 8 | 8 | 0 |
| ERI | Eritrea | 9 | 9 | 0 |
| COG | Congo (the) | 10 | 10 | 0 |
| BDI | Burundi | 11 | 11 | 0 |
| HTI | Haiti | 12 | 12 | 0 |
| SDN | Sudan (the) | 13 | 13 | 0 |
| IRQ | Iraq | 14 | 14 | 0 |
| GNQ | Equatorial Guinea | 15 | 15 | 0 |
| NER | Niger (the) | 16 | 17 | 1 |
| MLI | Mali | 17 | 16 | -1 |
| LBY | Libya | 18 | 19 | 1 |
| LBR | Liberia | 19 | 18 | -1 |
| AGO | Angola | 20 | 20 | 0 |
| CMR | Cameroon | 21 | 21 | 0 |
| MDG | Madagascar | 22 | 22 | 0 |
| NGA | Nigeria | 23 | 25 | 2 |
| UGA | Uganda | 24 | 24 | 0 |
| GIN | Guinea | 25 | 23 | -2 |
| ZMB | Zambia | 26 | 26 | 0 |
| MOZ | Mozambique | 27 | 27 | 0 |
| PNG | Papua New Guinea | 28 | 28 | 0 |
| GNB | Guinea-Bissau | 29 | 29 | 0 |
| COM | Comoros (the) | 30 | 30 | 0 |
| ZWE | Zimbabwe | 31 | 31 | 0 |
| MRT | Mauritania | 32 | 32 | 0 |
| MMR | Myanmar | 33 | 33 | 0 |
| BFA | Burkina Faso | 34 | 34 | 0 |
| TJK | Tajikistan | 35 | 35 | 0 |
| BGD | Bangladesh | 36 | 36 | 0 |
| PAK | Pakistan | 37 | 37 | 0 |
| GMB | Gambia (the) | 38 | 39 | 1 |
| PSE | Palestine, State of | 39 | 44 | 5 |
| TLS | Timor-Leste | 40 | 38 | -2 |
| ETH | Ethiopia | 41 | 40 | -1 |
| DJI | Djibouti | 42 | 42 | 0 |
| SLE | Sierra Leone | 43 | 41 | -2 |
| KHM | Cambodia | 44 | 43 | -1 |
| VEN | Venezuela (Bolivarian Republic of) | 45 | 47 | 2 |
| CIV | CÃ´te d'Ivoire | 46 | 46 | 0 |
| SLB | Solomon Islands | 47 | 45 | -2 |
| LAO | Lao People's Democratic Republic (the) | 48 | 49 | 1 |
| SWZ | Eswatini | 49 | 48 | -1 |
| GTM | Guatemala | 50 | 51 | 1 |
| KEN | Kenya | 51 | 50 | -1 |
| NIC | Nicaragua | 52 | 54 | 2 |
| TGO | Togo | 53 | 53 | 0 |
| TZA | Tanzania, the United Republic of | 54 | 52 | -2 |
| LSO | Lesotho | 55 | 55 | 0 |
| HND | Honduras | 56 | 57 | 1 |
| TKM | Turkmenistan | 57 | 56 | -1 |
| PRK | Korea (the Democratic People's Republic of) | 58 | 59 | 1 |
| BEN | Benin | 59 | 58 | -1 |
| IRN | Iran (Islamic Republic of) | 60 | 60 | 0 |
| GAB | Gabon | 61 | 61 | 0 |
| MWI | Malawi | 62 | 62 | 0 |
| AZE | Azerbaijan | 63 | 65 | 2 |
| RWA | Rwanda | 64 | 63 | -1 |
| EGY | Egypt | 65 | 67 | 2 |
| NPL | Nepal | 66 | 66 | 0 |
| SEN | Senegal | 67 | 64 | -3 |
| IDN | Indonesia | 68 | 68 | 0 |
| PHL | Philippines (the) | 69 | 69 | 0 |
| IND | India | 70 | 70 | 0 |
| PRY | Paraguay | 71 | 71 | 0 |
| LBN | Lebanon | 72 | 73 | 1 |
| STP | Sao Tome and Principe | 73 | 74 | 1 |
| DZA | Algeria | 74 | 80 | 6 |
| SLV | El Salvador | 75 | 75 | 0 |
| KGZ | Kyrgyzstan | 76 | 77 | 1 |
| UZB | Uzbekistan | 77 | 82 | 5 |
| ZAF | South Africa | 78 | 76 | -2 |
| COL | Colombia | 79 | 78 | -1 |
| OMN | Oman | 80 | 72 | -8 |
| VUT | Vanuatu | 81 | 79 | -2 |
| GHA | Ghana | 82 | 81 | -1 |
| SAU | Saudi Arabia | 83 | 83 | 0 |
| TUR | Turkey | 84 | 85 | 1 |
| PER | Peru | 85 | 84 | -1 |
| MEX | Mexico | 86 | 86 | 0 |
| BHR | Bahrain | 87 | 88 | 1 |
| SUR | Suriname | 88 | 87 | -1 |
| ECU | Ecuador | 89 | 89 | 0 |
| NAM | Namibia | 90 | 90 | 0 |
| DOM | Dominican Republic (the) | 91 | 91 | 0 |
| VNM | Viet Nam | 92 | 93 | 1 |
| GUY | Guyana | 93 | 92 | -1 |
| MDV | Maldives | 94 | 94 | 0 |
| LKA | Sri Lanka | 95 | 95 | 0 |
| BTN | Bhutan | 96 | 96 | 0 |
| PAN | Panama | 97 | 97 | 0 |
| KWT | Kuwait | 98 | 103 | 5 |
| JOR | Jordan | 99 | 99 | 0 |
| THA | Thailand | 100 | 98 | -2 |
| BIH | Bosnia and Herzegovina | 101 | 101 | 0 |
| BWA | Botswana | 102 | 100 | -2 |
| FJI | Fiji | 103 | 102 | -1 |
| BOL | Bolivia (Plurinational State of) | 104 | 104 | 0 |
| ALB | Albania | 105 | 105 | 0 |
| KAZ | Kazakhstan | 106 | 107 | 1 |
| QAT | Qatar | 107 | 106 | -1 |
| CPV | Cabo Verde | 108 | 109 | 1 |
| CHN | China | 109 | 113 | 4 |
| BRA | Brazil | 110 | 110 | 0 |
| MAR | Morocco | 111 | 111 | 0 |
| CUB | Cuba | 112 | 108 | -4 |
| MYS | Malaysia | 113 | 118 | 5 |
| BLZ | Belize | 114 | 112 | -2 |
| JAM | Jamaica | 115 | 115 | 0 |
| MKD | North Macedonia | 116 | 114 | -2 |
| ARE | United Arab Emirates (the) | 117 | 121 | 4 |
| UKR | Ukraine | 118 | 117 | -1 |
| MUS | Mauritius | 119 | 119 | 0 |
| SRB | Serbia | 120 | 120 | 0 |
| ARG | Argentina | 121 | 116 | -5 |
| MNG | Mongolia | 122 | 122 | 0 |
| RUS | Russian Federation (the) | 123 | 123 | 0 |
| TTO | Trinidad and Tobago | 124 | 124 | 0 |
| BLR | Belarus | 125 | 125 | 0 |
| MDA | Moldova (the Republic of) | 126 | 126 | 0 |
| GEO | Georgia | 127 | 127 | 0 |
| TUN | Tunisia | 128 | 128 | 0 |
| ARM | Armenia | 129 | 129 | 0 |
| CRI | Costa Rica | 130 | 131 | 1 |
| MNE | Montenegro | 131 | 130 | -1 |
| BGR | Bulgaria | 132 | 132 | 0 |
| BRN | Brunei Darussalam | 133 | 133 | 0 |
| CHL | Chile | 134 | 135 | 1 |
| HUN | Hungary | 135 | 136 | 1 |
| HKG | Hong Kong | 136 | 134 | -2 |
| ROM | Romania | 137 | 137 | 0 |
| HRV | Croatia | 138 | 139 | 1 |
| BRB | Barbados | 139 | 138 | -1 |
| SYC | Seychelles | 140 | 140 | 0 |
| SVK | Slovakia | 141 | 141 | 0 |
| CYP | Cyprus | 142 | 142 | 0 |
| POL | Poland | 143 | 143 | 0 |
| URY | Uruguay | 144 | 144 | 0 |
| MLT | Malta | 145 | 145 | 0 |
| GRC | Greece | 146 | 146 | 0 |
| SVN | Slovenia | 147 | 147 | 0 |
| CZE | Czechia | 148 | 148 | 0 |
| ITA | Italy | 149 | 149 | 0 |
| KOR | Korea (the Republic of) | 150 | 150 | 0 |
| ISR | Israel | 151 | 151 | 0 |
| LTU | Lithuania | 152 | 152 | 0 |
| IRL | Ireland | 153 | 154 | 1 |
| USA | United States of America (the) | 154 | 153 | -1 |
| JPN | Japan | 155 | 155 | 0 |
| SGP | Singapore | 156 | 156 | 0 |
| LVA | Latvia | 157 | 157 | 0 |
| ESP | Spain | 158 | 158 | 0 |
| PRT | Portugal | 159 | 159 | 0 |
| GBR | United Kingdom | 160 | 160 | 0 |
| FRA | France | 161 | 161 | 0 |
| EST | Estonia | 162 | 162 | 0 |
| DEU | Germany | 163 | 163 | 0 |
| CHE | Switzerland | 164 | 164 | 0 |
| AUT | Austria | 165 | 166 | 1 |
| NLD | Netherlands (the) | 166 | 165 | -1 |
| AUS | Australia | 167 | 167 | 0 |
| BEL | Belgium | 168 | 168 | 0 |
| CAN | Canada | 169 | 169 | 0 |
| NOR | Norway | 170 | 170 | 0 |
| ISL | Iceland | 171 | 171 | 0 |
| NZL | New Zealand | 172 | 172 | 0 |
| SWE | Sweden | 173 | 173 | 0 |
| LUX | Luxembourg | 174 | 174 | 0 |
| FIN | Finland | 175 | 175 | 0 |
| DNK | Denmark | 176 | 176 | 0 |