Exploring the Connection between Terrorism and Freedom Indices

Jiwon Kim, Tina Krulec
December 2020

ABSTRACT

This research project aims to provide a broad overview of the relationship between terrorism and freedom. It does not explore this complicated relationship in detail, but rather provides a broad overview of how the available data on terrorist attacks correlates with different indices for freedom. We specifically looked into the freedom of expression, religion, movement, and conflicts in general. We found weak correlations between the number of terrorist events and indices for freedom of expression, religion and movement. The strongest correlation is found between the number of events and the indices for conflict itself, which support the previously researched idea of the difficulty of defining terrorism. This data is not substantial enough or conclusive enough, due to the aggregation, averaging and lack of demographic data, to allow for explanation of any causal relation, however it does allow us to make a conclusion that terrorism and freedom are only weakly connected in the aereas explored in this paper.

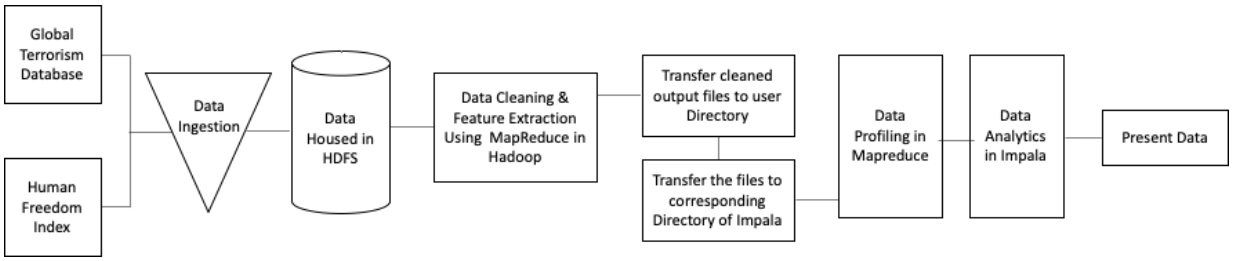
1 INTRODUCTION

This research paper provides a broad overview of the relationship between terrorism and freedom. This relationship is an important one to investigate to provide a balance to the distorted image in the popular media. We either see terrorism only occuring in countries that are considered to be more free, e.g. Western countries (Europe, USA), or in those considered to have oppressive political regimes and little freedom. This research looks at the relationship very objectively looking at correlations between different factors on different regional and time levels. The most basic of those is the year, country pairs, where we aggregate the GTD data to compare it with the HFI scores.

There has been some research done on this specific topic, but usually much more centered on specific indicators of freedom and terrorism, never this broad. We discuss some of that research below. We like to call this an exploratory project, as we believe it is important in providing a broad overview and baseline results that call for a more comprehensive research into the causal relationship between teorrorism and freedom.

Our work followed the data flow diagram below. First, We downloaded the two datasets from their perspective sources, and then ingested the data by uploading it to Dumbo and further to

HDFS. Then using MapReduce in Hadoop, we cleaned up the datasets and selected the features that we wanted to explore in specific, and we transferred the cleaned output files to our user directory and then to the corresponding directory of Impala. Given the datasets in its proper location, we ran data profiling to further our understanding of the data in MapReduce and ran analytics in Impala to explore and to gain valuable insights.



2 MOTIVATION

Ever since 9/11 terrorism has become synonymous with transnational terrorism in popular media. Furthermore, a terrorist has oftentimes become equal to a Muslim and public perception of terrorism is the one of oppression. As two international students in the Middle East we get asked a lot about our perception of Islam as a 'oppressive' religion in the context of Islam being linked to the terrorist acts in the West. Through this research paper, we wanted to get a broad overview if there is a link between terrorism and freedom out of personal curiosity. This does not exactly, address the issue of Islam being an oppressive religion, but it does look into another very important question of whether a country with more terrorist attacks is any less or more likely to score lower in freedom indices and what are the specific markers of freedom that terrorsim might impact, beyond the obvious ones, such as security, safety and conflict. Further than our own understanding, this research was important to undertake to debunk the myths of the public perception of terrorism existing only within the frames of oppressive regimes, as we were often on the receiving end of these assumptions, having to defend our choice of moving to the Middle East.

3 RELATED WORK

Eskildsen and Bjørnskov [1] have studied the relationship of civil rights and national safety and the association of those two concepts by exploring and looking at the effect in the degrees of freedom of expression on the risk of terrorist attacks. They use an older edition of the GTD dataset and find that this form of freedom of expression is directly associated with less terrorism in democracies. We implement a similar study in a more general sense, and we also found similar pattern given our negative correlation with freedom of expression with the incidents of terrorism

Greene [2] has studied the definition of terrorism in both domestic and international settings and points out a lack of a single definition for the term terrorism and how it is invariably broad to accommodate the lowest common denominator. Thus, the current scope of the definition over-exaggerated the concept of terrorism in the public and political debate. We base off this idea and explore our own version to deepen our understanding of terrorism with doubt indicator.

Lobato, Moya, Moyano, Trujillo [4] have looked at the role of oppression, radicalism, identity, and cultural intelligence in terrorism, specifically in Muslims and non-Muslims. Through two replicated studies, the data were studied in how perceived oppression influenced terrorist acts through radical intentions and found an indirect effect of perceived oppression the Muslim sample and showed that although many intervening factors influence it, the loss of significance and radical intentions seems to increase the likelihood of a process of violent disinhibition for those who are considered to be in marginal contexts.

McCann [5] have examined the reliability of "terrorism" classifications within the Global Terrorism Database (GTD) by looking at its inclusion criteria and discusses the underlying limitations of the data, namely data inclusion, the defining of terrorism, and inconsistency in labeling events. Further McCann has concluded the analyses that determine what is terrorism trickles down and affect studies that make inferences about groups, movements, or the efficacy of governmental policy.

Schmid [6] has studied to explore and to determine the causes of terrorism. He comments in the uncertainty that rises due to the lack of a single universal definition of terrorism and how different interpretation could affect the analysis of the root cause of terrorism. Although he experiments with multiple criteria, he stated that the results are rather "unsatisfactory" and concluded that the seeming causes varied but they all originate and referred back to the general idea of conflict.

4 DATASETS

Global Terrorism Database (GTD) 1970-2018

GTD is a comprehensive dataset on terrorist attacks in the world by Global National Consortium for the Study of Terrorism and Responses to Terrorism for 1970-2018. The data is extensive in the indicators and features it holds, as can be observed from its large file size of 173.5MB. Thus, in the cleaning process, we dropped various indicators, only keeping the ones we saw relevant. These include the event id number for easier identification and counting when aggregating the numbers, data on the time of the event, the country and region in which it happened. These last three are especially important to be able to draw the connection between the GTD and HFI dataset. Other indicators that we included simply because we were interested in looking at the

relationship between them and the freedom indices are the binary values indicating whether the event was a success, or if it was a suicide, string values on the types of the attacks, target type, nationality of the attcker and an integer value of how many people died in that event.

Human Freedom Index (HFI) 2019

HFI is a dataset on human freedom in the world, using 76 distinct indicators of personal and economic freedom, of 2008-2017, with the size of 672 B. All the scores are either out of 10 or they are aggregated. We dropped various indications, but still kept a lot during the cleaning process. Besides the basic information on the year, country and region, and the main scores on human freedom, and it's two components of personal and economic freedom, the column we kept are mostly related to personal freedom. You can find a full list in the Appendix one, but more specifically, we kept the scores for Rule of Law, Security and Safety, Movements, Religion, Association, Expression and Information, Identity and Relationship. From the economic freedom indicators we kept those related to international trade and labor regulations.

5 ANALYTICS STAGES

Ingestion

In the ingestion process we first changed the format of the file for the GTD dataset from xlsx to csv, and renamed both files from their original download names to gtd_data.csv and hfi_data.csv respectively. Files were transferred from the local device to the Dumbo and then to the HDFS server and furthermore shared with all the team members.

Data cleaning and profiling

Data profiling was done with map reduce in python on Hadoop. Data cleaning for HFI and GTD datasets included importing the data in csv files and only selecting specific columns that we wanted to explore. You can see those in the appendix in the data schema. In both cases our output file was a txt file. Additionally, for HFI cleaning we printed the missing data indicated as '-' as -1 to the output text file.

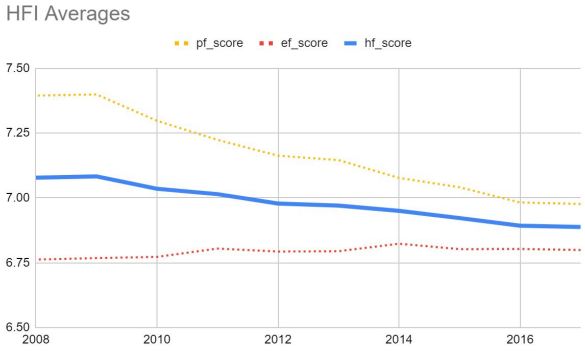
For the profiling stage we made sure that the lines of code in the input files were equal to those of the output files by writing the map reduce code. For HFI the numbers matched, however, they did not match for GTD. This is an understandable outcome given that we only choose a selected number of inputs ranging between years of 2008 and 2017, whereas the entire dataset had inputs from 1970 onwards.

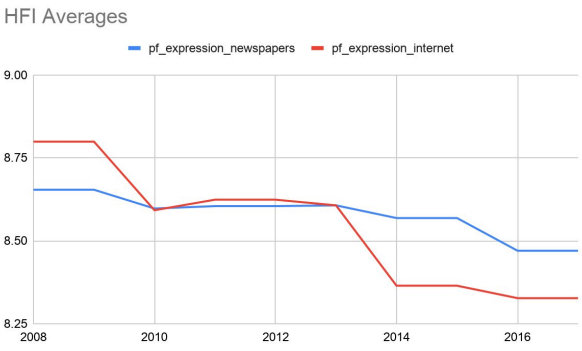
Analytics

Analytics was done in different stages - GTD separately, HFI separately, and both combined.

HFI

HFI provided only aggregate data by the year, country pairs, so we decided to look at the averages by years. Comparing the average results for Human Freedom score by the years, alongside the average scores for personal freedom and economic freedom, showed that the human freedom scored declined from 2008 to 2017, furthermore, it seems that most of the decline can be attributed to decline in personal freedom score. Alongside these, we also got averages for freedom of expression, specifically for newspapers and the internet. Following a similar pattern of the personal freedom score, both of those declined over the years, although somewhat less drastically than the personal freedom score. Due to this significant decrease of the personal freedom score relative to the increase in th economic freedom score, we decided to focus on the personal freedom indices in our combined analytics as well.





GTD

For GTD, it was a dataset less processed than HFI, as HFI was already accumulated by countries and years while GTD was organized by events of terrorism. Thus, GTD was first processed to resemble the country/year organization of HFI so that the two datasets can be combined in the later analytics. To do this, the events of terrorist attacks were counted per year per country and the other quantitative measures, such as number of fatalities (nkil/num_kill) and binary values indicating: whether there is doubt that the event was a terrorist attack (doubt_terrorist/doubt), whether the event was successful (1) or not (0) (success) and whether the event was a suicide (1) or not (0) (suicide) were summed to create an aggregate of data. We further added an additional feature to denote whether the attack was committed by someone national to the country or not by checking for a mismatch in the nationality of the terrorist and the country in which the terrorist attack took place.

Given the literature on the definition of "terrorism" being not universal, we decided to gain an understanding of the terrorism used in this specific dataset. Specifically we looked at doubt_terrorist and success as it was a measure that questioned if an event was a terrorist attack or not and success. Given the trends below, we saw high correlations with the doubt in the event was a terrorist attack and whether the event was successful, number of fatalities and whether it

had led to suicide and high correlations with success in a terrorist attack and whether the event was doubted in it being a terrorist attack, number of fatalities and whether it had led to suicide. Further, we could observe both doubts in the terrorist attack and whether an event was successful or not were not heavily but still moderately correlated to the attacker being a non-national.

	success	num_killed	suicide	Non_national terrorist
doubt_terrorist	0.857	0.750	0.716	0.540
success		0.886	0.847	0.443

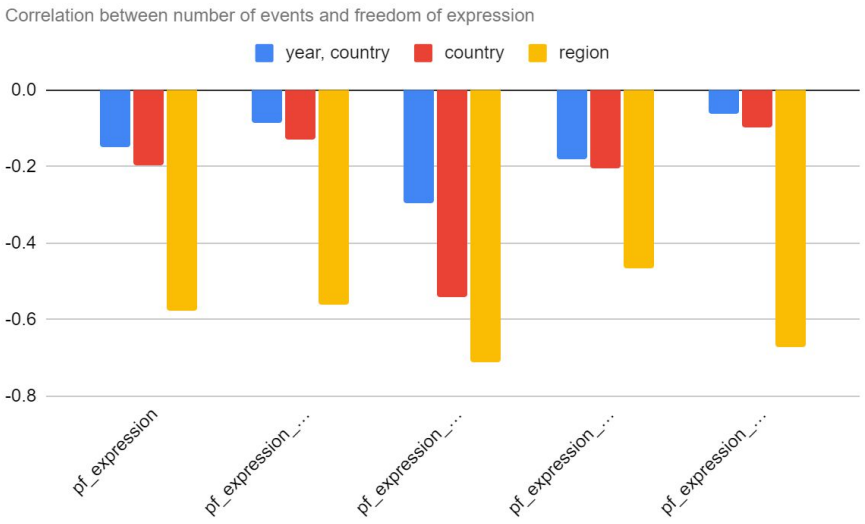
Combined Analytics

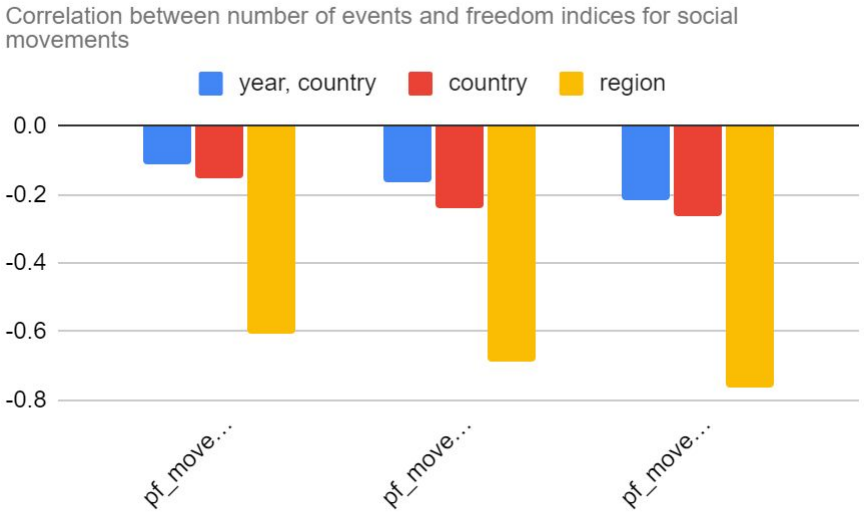
The combined analytics was done in SQL on Impala. To be able to compare the two datasets we first had to make them compatible. The main structural difference between the two is that GTD was a data set of individual events, whereas HFI has aggregate results. Due to that we had to ensure that the *year, country, region* by which all the data was classified in both GTD and HFI matched. First step was exploring what names, countries and regions, do not match. We found 12 country pairs that had different names, discovered that GTD had 3 more regions than HFI and found 3 countries for which the classification by region was switched between the two datasets. We outline this in the Appendix 2. We further explored which regions could be combined within GTD and their names changed to those in HFI to resolve the difference. All of these name changes had to be hard coded and were done through SQLs case statements while copying the data. Once the set up was done we were left with 834 unique combinations of years and countries, those included 139 unique countries, from 10 regions.

We decided to look for correlations between the data we had on terrorism and the one on freedom. In the first stage we wanted to explore the connection of aggregate data from GTD with the main scores from the HFI - human freedom, personal freedom and economic freedom. For that purpose we had to aggregate the events from the GTD, which we did by summing the number of events, number of people killed in those events, number of those that were successful, those doubted to be terrorist attacks and those that were classified as a suicide based on the year, country pair. We then ran the correlations specifically for the number of events and number of people killed.

	num_events	sum_num_killed
hf_score	-0.259	-0.282
pf_score	-0.298	-0.318
ef_score	-0.149	-0.175

In the second stage we delved into the personal freedom indices, based on the preliminary results on the HFI dataset and given that the highest correlations in the first stage were with the personal freedom score. Specifically we look at those quantifying security and safety, freedom of expression and freedom of movement. In addition to looking at the correlations between features already available in our datasets, we decided to create a few in this second stage to aid with our research process. Based on our exploration of the GTD dataset we knew that the mismatch between the nationality of the attacker and the country of the event was an important marker to consider, which is why we created a binary feature *mismatch.* Furthermore, as we are not using any demographic data, we had to find a way to balance for the varying population number across the countries and the regions. As a result we created percent based values indication how many of the total events for that year, country or region specific in the function were successful, a suicide, had a mismatch of nationality of the attacker and the country of the event, were doubted to be a terrorist attack and how many people died on average per event.
The process we undertook in this second stage was the same for all three. We created HFI subtables with only the relevant columns, and similarly to stage one aggregated the GTD data by the country, year pairs that additionally had all the percent values described above. We ran the correlations between at least the number of events and the specific freedom scores. Oftentimes also looking at the percent of mismatched events (nationality of the attacker different from the country), percent of successful events and others. We further repeated the process on the country level, aggregating the numbers over the available years for GTD and taking an average for HFI scores over the years, and did the same for regions, where scores were aggregated or averaged over all the available years and countries. For the sake of comparison, the below tables are all indicating the correlations between the number of events and the specific freedom indices we look at.





In the third stage, we delved into the freedom of religion, as it was one of our main focuses as identified in the motivation. We explored the relationship between the number of events, denoted by num_events, and total casualties, denoted by "sum_num_kil", with different freedom of religion, the aggregate score for religion, freedom on establishing and operating religious organizations, general harassment and physical hostilities, legal and regulatory restrictions, as shown in the table below.

	pf_religion	pf_religion_estop	pf_religion_harassment	pf_religion_restrictions

num_events	-0.172	-0.095	-0.050	-0.086
sum_num_kill	-0.158	-0.048	-0.091	-0.042

However we found very minimal negative correlation of both in num_events and sum_num_kill, and for specific indicators, the correlation was close to zero.

Given the minimal correlation, we delved further to see if there would be higher correlation or more insights in seeing the relationship between “pf_religion” and “num_events” in different target types, such as “Private Citizens & Property”, “Business”, “Government (General)”, “Government (Diplomatic)”, “Military”, “Religious Figures /Institutions”. As the results identified in the below table, there continued to be very minimal correlations, even less than the general score of “pf_religion”. However, we did see the highest negative correlation with target types of military and the least correlation with general government type.

target_type	num_events
Private Citizens & Property	-0.125
Business	-0.082
Government (General)	0.001
Government (Diplomatic)	0.115
Military	-0.168
Religious Figures/Institutions	-0.085

7 CONCLUSION

This research has many limitations. We are not using any demographic data to balance for the varied population in the countries which likely has an effect on the number of terrorist events. Furthemore for the main part of our analytic, the most conclusive results we get are on the regional level (refer to the graphs above), for which we only have 10 inputs, which makes the correlations more likely to occur. The increase in correlations from the year, country level, to country and regional levels are evident in all the experiments we did. While we can make some assumptions about the similarity of countries and their terrorism activity and freedom scores within the regions, we cannot draw any conclusive results from that as there many other factors that we would have to control for. The most objective results we get will still be on the year,

country pairs. Regardless of these limitations, out research does offer an important insight into the relation between terrorism and freedom. Despite being skewed toward countries with greater populations, the number of events was the factor from GTD that we used the most informative while finding correlations with the freedom indices.

We show that on the most objective level of the year, country pairs there is very little negative correlation with the main freedom indices of human freedom score, personal freedom and economic freedom. This is an important outcome of this research paper as it is clearly showing that terrorism, while related to freedom is very weakly linked to it. This supports an argument against the popular media that portrays terrism as existing only in 'free' or 'oppressed' countries. Furthemore, following a similar line of thinking we find very weak correlation between freedom of expression and terrorism. The most highly correlated index is the score of press being killed, which is expected. More terrorist attacks leads to a higher number of people killed, which would likely also include members of the press. It is this lack of significant correlation on the year, country level that is very interesting. It signals that terrorism does not relate ot the state control of the press, which usually occurs in oppressive regimes. Hence, this supports the idea that terrorism is a global phenomena, not simply limited to the specific regions. This, while very interesting to see manifest in our research, is not a new concept. The highest correlation we found on the year, country pair level was with the freedom measure for the conflict itself. This is not a surprising result, but compared to the lack of others is very telling. Especially the ones for religion. Popular media makes a point of linking terrorism with religion, so it was interesting to see the lack of any significant correlation. It confirms what popular media is not so willing to disclose, that there is no significant relation between religion and terrorism. Additionally our exploration of the GTD dataset provided quantifiable evidence to support the lack of definition of terrorism.

In conclusion, we can hardly provide a causal relationship from our exploratory research due to the limitations described above and the limited scope of it. However, what our research does is gives an overview of the relationship between the available data on terrorism and freedom indices. It will hopefully prompt further research into the relationship taking into account the restrictions and limitations that we could not. However, the lack of correlation between the data we find, is a significant evidence for the over the top and dichotomous picture that the popular media paints with regards to terrorism, especially it's link to the oppressive vs. free states and religion.

REFERENCE

1. Eskildsen LS, Bjørnskov C. Does Freedom of Expression Cause Less Terrorism? Political Studies. September 2020. doi:10.1177/0032321720950223
2. Greene A. DEFINING TERRORISM: ONE SIZE FITS ALL? International and Comparative Law Quarterly. 2017;66(2):411-440. doi:10.1017/S0020589317000070
3. Institute for Economics & Peace. Global Terrorism Index 2020: Measuring the Impact of Terrorism, Sydney, November 2020. Available from: http://visionofhumanity.org/reports (accessed 02 12 2021). https://www.visionofhumanity.org/wp-content/uploads/2020/11/GTI-2020-web-1.pdf
4. Lobato RM, Moya M, Moyano M, Trujillo HM. From Oppression to Violence: The Role of Oppression, Radicalism, Identity, and Cultural Intelligence in Violent Disinhibition. Front Psychol. 2018;9:1505. Published 2018 Aug 20. doi:10.3389/fpsyg.2018.01505
5. McCann, Wesley S. Who Said We Were Terrorists? Issues with Terrorism Data and Inclusion Criteria, Studies in Conflict & Terrorism, DOI: 10.1080/1057610X.2020.1816676
6. Schmid, Alex P. Root Causes of Terrorism: Some Conceptual Notes, a Set of Indicators, and a Model, Democracy and Security, 1:2, 127-136, DOI: 10.1080/17419160500321139

Appendix 1: Data Schema

Global Terrorism Dataset

Variable_name *variable_type* [analytics_name]

eventid *str* [event]

- A unique identification number for each terrorist event.

iyear *int* [year]

- Year of the event.

imonth *int* [month]

- The month of the event

iday *int* [day]

- Day of the event.

country_txt *str* [country]

- Country.

region_txt *str* [region]

- Region.

doubtterr - *int* [doubt_terrorist]

- A binary value indicating whether there is doubt that the event was a terrorist attack.

***int* [success]**

- A binary value indicating whether the event was successful (1) or not (0).

***int* [suicide]**

- A binary value indicating whether the event was a suicide (1) or not (0).

attacktype_txt *str* [attack_type]

- Description of the attack - shooting, etc.
- Unarmed Assault
- Hostage Taking (Barricade Incident)
- Unknown
- Assassination
- Hijacking
- Armed Assault
- Hostage Taking (Kidnapping)
- Bombing/Explosion
- Facility/Infrastructure Attack

targettype_txt *str* [target_type]

- Category of the target - educational, religious, etc.
- Telecommunication
- Tourists
- Journalists & Media
- Police
- Unknown
- Military
- Business
- Government (General)
- Other
- Educational Institution
- Utilities
- Terrorists/Non-State Militia
- NGO
- Violent Political Party
- Private Citizens & Property
- Airports & Aircraft
- Transportation
- Religious Figures/Institutions
- Food or Water Supply
- Abortion Related
- Government (Diplomatic)
- Maritime

natlty_txt *str* [nationality]

- Nationality of the attacker.

nkil *int* [num_kill]

- A number of fatalities.

Human Freedom Index

Variable_name *variable_type* [analytics_name]

Year *int* [year]

- Year if the index measured.

Countries *str* [country]

- Name of the country.

Region *str* [region]

- Region of the country.

Human Freedom (Score) *float* [hf_score]

- Aggregated score.

Rule of Law *Float* /pf_rol]

- Aggregated score out.

Disappearances *Float* [pf_ss_disappearances_disap]

- Score out of 10.

Violent Conflicts *Float* [pf_ss_disappearances_violent]

- Score out of 10.

Organised Conflicts *Float* [pf_ss_disappearances_organized]

- Score out of 10.

Terrorism Fatalities *Float* [pf_ss_disappearances_fatalities]

- Score out of 10.

Terrorism Injuries *Float* [pf_ss_disappearances_injuries]

- Score out of 10.

Disapperances, Conflicts, and Terrorism *Float* [pf_ss_disappearances]

- Score out of 10.

Female Genital Mutilation *Float* [pf_ss_women_fgm]

- Score out of 10.

Inheritance Rights: Widows *Float* [pf_ss_women_inheritance_widows]

- Score out of 10.

Inheritance Rights: Daughters *Float* [pf_ss_women_inheritance_daughters]

- Score out of 10.

Inheritance Rights *Float* [pf_ss_women_inheritance]

- Score out of 10.

Women's Security & Safety *Float* [pf_ss_women]

- Aggregated score.

***Security & Safety float* [pf_ss]**

- Aggregated score.

Domestic Movement *Float* [pf_movement_domestic]

- Score out of 10.

Foreign Movement *Float* [pf_movement_foreign]

- Score out of 10.

Movement *Float* [pf_movement]

- Aggregated score.

Establishing Religious Organizations *Float* [pf_religion_estop_establish]

- Score out of 10.

Operating Religious Organizations *Float* [pf_religion_estop_operate]

- Score out of 10.

Establishing and Operating Religious Organizations *Float* [pf_religion_estop]

- Score out of 10.

Harassment and Physical Hostilities *Float* [pf_religion_harassment]

- Score out of 10.

Legal and Regulatory Restrictions *Float* [pf_religion_restrictions]

- Score out of 10.

Religion *Float* [pf_religion]

- Aggregated score.

Association *Float* [pf_association]

- Aggregated score.

Press Killed *Float* [pf_expression_killed]

- Score out of 10.

Press Jailed *Float* [pf_expression_jailed]

- Score out of 10.

Laws and Regulations that Influence Media Content *Float* [pf_expression_influence]

- Score out of 10.

Political Pressures and Controls on Media Content *Float* [pf_expression_control]

- Score out of 10.

Access to Cable/Satellite *Float* [pf_expression_cable]

- Score out of 10.

Access to Foreign Newspapers *Float* [pf_expression_newspapers]

- Score out of 10.

State Control over Internet Access *Float* [pf_expression_internet]

- Score out of 10.

Expression & Information *Float* [pf_expression]

- Aggregated score.

Identity & Relationships *Float* [pf_identity]

- Aggregated score.

PERSONAL FREEDOM (Score) *Float* [pf_score]

- Aggregated score.

Revenue from trade taxes (% of trade sector) *Float* [ef_trade_tariffs_revenue]

- Score out of 10.

Tariffs *Float* [ef_trade_tariffs]

- Score out of 10.

Black market exchange rates *Float* [ef_trade_black]

- Score out of 10.

Foreign ownership/investment restrictions *Float* [ef_trade_movement_foreign]

- Score out of 10.

Capital controls *Float* [ef_trade_movement_capital]

- Score out of 10.

Freedom of foreigners to visit *Float* [ef_trade_movement_visit]

- Score out of 10.

Controls of the movement of capital and people *Float* [ef_trade_movement]

- Score out of 10.

Freedom to trade internationally *Float* [ef_trade]

- Aggregated score.

Labor market regulations *Float* [ef_regulation_labor]

- Score out of 10.

Business regulations *Float* [ef_regulation_business]

- Score out of 10.

Regulation *Float* [ef_regulation]

- Aggregated score.

ECONOMIC FREEDOM (Score) *Float* [ef_score]

- Aggregated score.

Appendix 2: Differences between the GTD and HFI

Regions that needed to be combined.
South Asia (HFI) → South + Southeast Asia (GTD)
Oceania (HFI) → Australasia & Oceania (GTD)
Latin America & the Caribbean (HFI) → Central America & Carribean + South America (GTD)
Caucasus & Central Asia (HFI) → Central Asia (GTD)

Countries with different names.
Yemen, Rep. (HFI) -- Yemen (GTD)
Slovak Rep. (HFI) -- Slovak Republic (GTD)
Pap. New Guinea (HFI) -- Papua New Guinea (GTD)
Kyrgyz Republic (HFI)-- Kyrgyzstan (GTD)
Timor-Leste (HFI) -- East Timor (GTD)
Korea, South (HFI) -- South Korea (GTD)
Bosnia and Herzegovina (HFI)-- Bosnia-Herzegovina (GTD)
Dominican Rep. (HFI) -- Dominican Republic (GTD)
Central Afr. Rep.(HFI) -- Central African Republic (GTD)
Czech Rep.(HFI) -- Czech Republic (GTD)
Congo, Rep. Of (HFI) --- Republic of the Congo (GTD)
Congo, Dem. R. (HFI)-- Democratic Republic of the Congo (GTD)

Countries with mismatched regions.
Cyprus: Eastern Europe (HFI) -- Western Europe (GTD) → Western Europe
Greece: Eastern Europe (HFI) -- Western Europe (GTD) → Western Europe
Mexico: Latin America & the Caribbean (HFI) -- North America (GTD) → Latin America & the Caribbean