1: Sourcing Open Data

The WorldRiskReport, an annual technical report on global disaster risks published jointly by the Institute for International Law of Peace and Armed Conflict (IFHV) and Bündnis Entwicklung Hilft, aims to contribute to a global view of the links between natural events, climate change, development, and preparedness, and to draw forward-looking conclusions for aid, policy, and reporting.

Data Source

Kaggle Dataset by Mrinal Tyagi
(License: CC0: Public Domain)

This dataset of the WorldRiskIndex was accessed via Kaggle and originally sourced from the WorldRiskReport 2021. It includes four additional dimensions that are classifications of the WorldRiskIndex, Exposure, Vulnerability, and Susceptibility into the ordinal categories very low, low, medium, high, and very high based on their scores. These classifications and the thresholds for each category are defined in the WorldRiskReport.

Full datasets, methods, and metadata
(License: Creative Commons Attribution 4.0 International license (CC BY 4.0))

The WorldRiskIndex (WRI) is developed and maintained by organizations that specialize in disaster risks assessment and management, such as the United Nations Office for Disaster Risk Reduction (UNDRR), the World Bank, and research institutions. The basic model of the WorldRiskIndex was developed in collaboration with the United Nations University Institute for Environment and Human Security (UNU-EHS). Since 2018, the Institute for International Law of Peace and Armed Conflict (IFHV) at the Ruhr-Universität Bochum has taken over the calculation and further developed the model conceptually and methodologically. In 2022, the WorldRiskIndex was published with a completely revised model that includes 100 indicators from publicly available databases worldwide.
Data sources are carefully selected based on their reliability, consistency, relevance to disaster risk assessment, and representativeness at the national or sub-national level. Continuous efforts are made to update and improve the data used in the index to reflect changes in risk factors and emerging trends in disaster risk management. Due to the integrity of the report, the source can be considered reliable and trustworthy.

Data Collection

The data used to calculate the WorldRiskIndex are collected from a variety of national and international sources, including government agencies, non-governmental organizations (NGOs), research institutions, and international organizations. These sources provide data on factors such as exposure to natural hazards, socioeconomic vulnerability, infrastructure resilience, and adaptive capacity. Data collection methods may include surveys, administrative and other techniques such as field observations, and statistical modeling.
The statistical model used to calculate the WorldRiskIndex is constantly being expanded and updated to incorporate new aspects of disaster risk for future reports. To ensure analytical consistency of the data provided when changes or updates are made, new models require that any innovations, changes, updates, or other adjustments to the WorldRiskIndex be applied retroactively to a longitudinal dataset for the past ten years to provide users with sufficient data for trend analysis.

Data Contents

The WorldRiskIndex is calculated for each country as the geometric mean of exposure and vulnerability. Exposure represents the risk to the population from earthquakes, tsunamis, coastal and river flooding, hurricanes, droughts, and sea level rise. Vulnerability represents the social sphere and consists of three dimensions:

- **Susceptibility** describes the structural characteristics and frameworks of a society that increase the overall likelihood that populations will be harmed by extreme natural events and end up in a disaster situation.
- **Coping** encompasses various capabilities and measures of societies to counteract the negative effects of natural hazards and climate change through direct action and available resources in the form of formal or informal activities and to minimize damage in the immediate aftermath of an event.
- **Adaptation**, in contrast to coping capacity, refers to long-term processes and strategies that aim to achieve anticipatory changes in social structures and systems to counteract, mitigate, or specifically avoid future negative impacts.

In the current dataset, exposure and vulnerability together with the overall WRI score are broken down by region (181 countries worldwide) and year (2011-2021), resulting in 12 columns and 1917 rows.

Data Relevance

Overall relevance of the WorldRiskIndex dataset:

The WRI dataset is highly relevant for data analysis and provides insights into the critical global challenge of disaster risk assessment and management. In a world where natural disasters are increasing in frequency and severity, understanding and mitigating these risks is essential to the safety and well-being of communities worldwide. The dataset provides valuable tools for exploring the complex interactions between natural hazards, socio-economic vulnerabilities, and infrastructure resilience to support evidence-based decision-making and policy formulation. However, it's important to acknowledge potential limitations and biases within the dataset, such as data collection errors, reporting biases, and subjective decisions in indicator selection and weighting. Despite these limitations, analysis of the WRI dataset can help address real-world issues with global implications and improve disaster preparedness, response, and recovery efforts. By critically evaluating the data and applying rigorous analytical methods, valuable insights can be gained to guide effective risk reduction strategies and resource allocation.

Reasons for choosing the WorldRiskIndex dataset:

The WRI dataset addresses a critical global issue - disaster risk assessment and management - and provides an opportunity to work on real-world problems with practical implications. Its interdisciplinary nature encompasses a wide range of data on natural hazards, socioeconomic vulnerability, and infrastructure resilience, allowing me to explore topics as diverse as geography, economics, and public policy. Because the data are often publicly available, the relevant information is easy to access and work with, allowing me to practice various analytical techniques such as descriptive statistics and predictive modeling. In addition, analysis of the WRI dataset can lead to valuable insights that contribute to the understanding of global disaster risk patterns and foster critical thinking and problem-solving skills in addressing complex real-world challenges.

Data Cleaning

Table 1: Data cleaning actions

Variable/Row	Issues	Action
Region	Differs from official naming	Renamed to ‘Country’
Region	Inconsistent values	Unified values to official country names as of 2023

Lack of Coping Capabilities	Inconsistent naming	Renamed to ‘Lack of Coping Capacities’
Lack of Adaptive Capacities	Leading whitespace in naming	Renamed to ‘Lack of Adaptive Capacities’
Vulnerability Category	4 missing values	Inserted with data from the original report from 2016/2019
WRI Category	1 missing value	Inserted with data from the original report from 2020
Lack of Adaptive Capacities	1 missing value	Inserted with data from the original report from 2016
1858	All values misplaced	Correct all values with data from the original report from 2016
Country	Missing rows for the following countries: Antigua and Barbuda, Samoa, Sao Tome and Principe, Democratic Republic of Congo, Federated States of Micronesia, Montenegro, Saint Lucia, Maldives, Saint Vincent and the Grenadines, Dominica	The values for the missing years were reworked and could be obtained from the trend data set at a later stage of the analysis, if necessary. However, the revised data takes into account changes, adjustments, or updates in the source data. Accordingly, there is a possibility of discrepancies between the trend dataset and the individual datasets. Therefore, imputations should be handled with care.

Data Understanding

Table 2: Data profile

Variable	Time-variant/ -invariant	Structured/ Unstructured	Qualitative/ Quantitative	Binary/Nominal/Ordinal Discrete/Continuous
Region	time-invariant	Structured	Qualitative	Nominal
Year	time-variant	Structured	Qualitative	Ordinal
WRI	time-variant	Structured	Quantitative	Continuous
Exposure	time-variant	Structured	Quantitative	Continuous
Vulnerability	time-variant	Structured	Quantitative	Continuous
Susceptibility	time-variant	Structured	Quantitative	Continuous
Lack of Coping Capabilities	time-variant	Structured	Quantitative	Continuous
Lack of Adaptive Capacities	time-variant	Structured	Quantitative	Continuous

WRI Category	time-variant	Structured	Qualitative	Ordinal
Exposure Category	time-variant	Structured	Qualitative	Ordinal
Vulnerability Category	time-variant	Structured	Qualitative	Ordinal
Susceptibility Category	time-variant	Structured	Qualitative	Ordinal

Table 3: Descriptive statistics

	WRI	Exposure	Vulnerability	Susceptibility	Lack of Coping Capacities	Lack of Adaptive Capacities
count	1917	1917	1917	1917	1917	1917
mean	7.55	15.38	48.08	30.72	70.45	43.09
std	5.55	10.23	13.82	15.67	15.02	13.55
min	0.02	0.05	20.97	8.26	35.16	11.16
25%	3.74	10.16	37.04	17.78	59.33	33.17
50%	6.52	12.76	47.10	25.37	74.23	43.07
75%	9.37	16.45	60.06	42.61	83.00	53.06
max	56.71	99.88	76.47	70.83	94.36	76.11

Variable ranges:

WRI scores range from 0.02 to 56.71, indicating a wide range of variability.

Central tendency:

Mean scores for WRI, exposure, vulnerability, susceptibility, lack of coping skills, and lack of adaptive skills hover around 7.55, 15.38, 48.08, 30.72, 70.45, and 43.08, respectively, reflecting typical values.

Variability:

WRI's standard deviation is about 5.55, indicating moderate variability.

Data Limitations & Ethics

Limitations:

- **Data availability and quality:** Incomplete, outdated, or biased data, especially in regions with limited resources or political instability, can compromise risk assessments and potentially introduce bias against certain regions or demographics.
- **Spatial resolution:** WRI's risk assessments, typically at the national or subnational level, can mask regional differences within countries.

- **Temporal resolution:** The longitudinal (trend) dataset is subject to annual revisions and may reflect changes, adjustments, or updates in the source data. As a result, there may be discrepancies between the longitudinal dataset and the individual datasets.
- **Limited scope of analysis:** WRI focuses primarily on natural hazards and their impacts, neglecting other dimensions of risk such as technological hazards. This narrow focus can limit understanding of comprehensive risk dynamics and potentially bias policy decisions and resource allocations to address only certain types of risk.

Ethics:

- **Collection bias:** Bias can be introduced through the selection and weighting of indicators used in the WRI calculation, reflecting different stakeholder priorities. This process may overlook certain aspects of risk and fail to adequately capture local nuances, potentially favoring regions with more comprehensive data or those aligned with dominant perspectives.
- **Measurement bias:** The methodologies used to construct the WRI may include subjective elements, such as expert judgment or arbitrary decisions in indicator selection and aggregation. Different methodologies may lead to different interpretations of risk, potentially leading to biased assessments depending on the approach chosen.
- **Exclusion bias:** Aggregating different risk factors into a single index can oversimplify the complexity of disaster risk and miss important nuances or interactions between risk factors. Certain risk factors may be prioritized, leading to biased risk representations that favor certain risk dimensions over others.

Research Questions

Clarifying questions:

1. Which regions or countries are most exposed to natural catastrophes according to the WorldRiskIndex?
2. How do the different criteria (exposure, vulnerability, susceptibility, etc.) contribute to WorldRiskIndex? Are there specific criteria that impact the final index more than others?
3. How has the distribution of risk levels changed over time?

Adjoining questions:

1. Are there any demographic factors (population density, urbanization rates, etc.) that correlate with higher risk index scores?

2. Is there evidence of a disproportionate impact of climate change on countries with higher risk index scores?
3. How do countries with similar risk index scores differ in their disaster preparedness and response strategies?

Funneling questions:

1. Which specific countries have consistently ranked highest or lowest on the WorldRiskIndex over the years?
2. Are there common patterns or trends among countries within the same region concerning their risk index and its components?
3. How do countries with similar levels of development or geographical characteristics differ in their WorldRiskIndex scores?
4. What are the main drivers of high-risk levels in particular regions or countries?
5. Are there patterns in risk index scores based on geographic factors such as proximity to coastlines, fault lines, or climate zones?
6. Are there trends or patterns in risk index scores based on specific types of disasters (e.g., earthquakes, floods)?

Elevating questions:

1. Are there examples of countries that have successfully reduced their risk levels over time, and if so, what strategies did they use?
2. How can the results of this analysis contribute to global efforts in disaster preparedness, climate resilience, and sustainable development initiatives?

Sources

Bündnis Entwicklung Hilft. "WeltRisikoBericht." Accessed March 11, 2024. https://www.entwicklung-hilft.de/.

Institute for International Law of Peace and Armed Conflict. "WorldRiskIndex." OCHA Services. Humanitarian Data Exchange (HDX). Accessed March 11, 2024. https://data.humdata.org/dataset/worldriskindex.

Mrinal Tyagi. "World Disaster Risk Dataset." Kaggle. Accessed March 11, 2024. https://www.kaggle.com/datasets/tr1gg3rtrash/global-disaster-risk-index-time-series-dataset.