Codebook: Development Indicators Dataset

Colby Pastel

2024-02-28

Overview:

The data presented in this codebook describe important international development indicators of individual countries over time. In total, there are 17 variables and 13,671 observations. The period of record begins in 1960 and ends in 2022. Please note, not all variables are available for all years. Most variables contain numeric data, with a couple containing character data. There are also two factor variables presented: Continent and GINI. This dataset was designed with linear regression analysis in mind to yield insights into the development of countries around the world.

Sources and Methodology:

All variables come from the World Bank's World Development Indicators (WDI) database except for ID, Continent, GDPC (which were generated), and HAQ, which comes from the Institute for Health Metrics and Evaluation (accessed through Our World in Data). Data was saved from the World Bank's DataBank in a .csv file and imported into RStudio. Data was tidied from its original form by lengthen and widening pivots, variables were added, and variable names were renamed to be more user friendly. Data for the HAQ indicator was saved from Our World in Data as a .csv file and imported into RStudio. HAQ data was then merged with the modified World Bank dataset. Notably, the World Bank WDI database does not include an indicator for access to healthcare, so the Healthcare Access and Quality Index (HAQ) from the Institute for Health Metrics and Evaluation was used instead. Finally, the Continent and GINI variables were converted to factor variables for simplicity.

Itemized Variables:

Unique Identifier

Variable name: ID

Variable type: integer

Description: Unique identifier given to each observation. Author Generated.

Country Name

Variable name: Country

Variable type: character

Description: Name of countries in dataset.

Country Code

Variable name: Country_Code

Variable type: character

Description: Three letter country code given to each country.

Continent

Variable name: Continent

Variable type: factor

Description: Continent which country belongs to. Author generated.

Value	Label	Freq
1	Africa	3402
2	Americas	2898
3	Asia	3150
4	Europe	3024
5	Oceania	1197

Year

Variable name: Year

Variable type: numeric

Description: Years of observations.

min	mean	median	max
1960	1991	1991	2022

Population

Variable name: Pop

Variable type: numeric

Description: Provides the number of people.

min	mean	median	max
2646	24964013	4242050	1417173173

Gross Domestic Product

Variable name: GDP

Variable type: numeric

Description: Quantifies the value of goods and services produced by a country. Provides an estimate of the strength of the economy. In current $US.

min	mean	median	max
14658402	415631142463	39875503508	3.03371e+13

Gross Domestic Product per Capita

Variable name: GDPC

Variable type: numeric

Description: Gross domestic product expressed as a ratio of population. In current $US.

min	mean	median	max
287.7	15952.7	8412.5	163219.6

Poverty Ratio

Variable name: Pov_Ratio

Variable type: numeric

Description: Percentage of the population living on less than $2.15 (2017 PPP).

min	mean	median	max
0	30.5	23.7	91.5

Gini Index

Variable name: GINI

Variable type: factor

Description: Measure of the equality of the distribution of income or consumption of individuals and households within an economy.

Value	Label	Freq
1	High Equality	406
2	Low Equality	700
3	Middle Equality	85
4	NA	12480

International Aid Received

Variable name: Int_Aid

Variable type: numeric

Description: Net amount of developmental assistance or international aid received by a country. In current $US.

min	mean	median	max
-989940002	388535565	130180000	22057089844

Educational Attainment

Variable name: Edu_Aid

Variable type: numeric

Description: Percentage of the population age 25 and up which finished primary school.

min	mean	median	max
0	0.7	0.8	1

Water Access

Variable name: Water_Acc

Variable type: numeric

Description: Percentage of the population with access to a basic drinking water source.

min	mean	median	max
0.2	0.8	0.9	1

Sanitation Access

Variable name: San_Acc

Variable type: numeric

Description: Percentage of the population with access to a basic sanitation system.

min	mean	median	max
0	0.7	0.9	1

Maternal Mortality Rate

Variable name: Mat_Mort

Variable type: numeric

Description: Number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination. Measured per 100,000 live births.

min	mean	median	max
1	491.6	434	1687

Child Mortality Rate

Variable name: Child_Mort

Variable type: numeric

Description: Number of newborns who die before reaching the age of 5. Measured per 1,000 live births.

min	mean	median	max
1.7	143.5	132.9	429.1

Healthcare Access and Quality Index

Variable name: HAQ

Variable type: numeric

Description: Index measure of access to healthcare, with a 0 (worst) to 100 (best) scale. Index based upon death rates from 32 causes of death which can be avoided with timely medical care.

min	mean	median	max
23.1	58	58	94.6