Codebook for Afghanistan WorldBank Data

Emory Bayer

2024-02-28

Overview of Data:

This data focuses on measuring World Development Indicators for the country of Afghanistan from the years 1960 to 2022. It encompasses a variety of social, economic, environmental, and institutional variables within Afghanistan. For this project, the data has been cleaned to include only variables where 95% of the responses for each year were recorded or were not NA values. The variables chosen to be included within the cleaned data comprise life expectancy broken down by gender, population totals, population distributions between urban and rural areas, fertility rates, and net migration."

Sources and Methodology:

This data is sourced from the World Bank and was collected on a yearly basis following the fiscal year schedule, wherein the reporting period concluded on March 20th each year. The World Bank employed data from various sources, including the United Nations Population Division, the United Nations Statistical Division Population and Vital Statistics Report, census reports, and other statistical publications from Afghan statistical offices, as well as the Secretariat of the Pacific Community: Statistics and Demography Programme.I have included the link below where the data was retrieved from:

https://data.worldbank.org/country/afghanistan

Note on missing values:

Missing values in this dataset were initially recorded as 'NA.' However, I have recoded all these missing values as 0. In the summary statistics for variables, I have indicated which values represent NAs and from which year they originate.

Itemized Presentation of Variables:

Variable name : *Year*

Variable type : Numeric

Description:

The 'year' variable records the timeframe during which the data was captured. The data spans from the years 1960 to 2022.

Variable name: *Country*

Variable type: Character

Description: The 'country' variable represents the country in which the data was recorded. For this dataset, all values pertain to Afghanistan.

Variable name: *Population.total*

Variable Type: Numeric

Description: This variable represents a midyear estimate of the number of people residing in Afghanistan in the year the data was collected, regardless of legal status or citizenship.

Mean Population	Median Population	Minimum Population	Maximum Population
18410104	12938862	8622466	41128771

Variable name: *Urban.population*

Variable Type: Numeric

Description: This variable represents the percentage of the population living in urban areas, ranging from 0 to 100%, divided by the total population.

Mean Urban Population	Median Urban Population	Minimum Urban Population	Maximum Urban Population
18.86106	21.266	8.401	26.616

Variable name: *Population.largest.city*

Variable type: Numeric

Description: This variable represents the percentage of a country's urban population living in the country's largest metropolitan area. In the case of Afghanistan, this metropolitan area is Kabul.

Mean Percentage living in Largest City	Median Percentage living in Largest City	Minimum Percentage living in Largest City	Maximum Percentage living in Largest City
18.86106	21.266	8.401	26.616

Variable name: *Rural.population.growth*

Variable type: Numeric

Description: This variable represents the population living in areas defined by the national statistics office as rural areas. It is calculated by computing the difference between the total population and the urban population.

Mean Rural Population	Median Rural Population	Minimum Rural Population	Maximum Rural Population
2.127972	2.017564	-11.95256	14.85123

Variable name: ***Fertility.rate***

Variable type: Numeric

Description: This variable represents the number of children that a woman would have if she lives through her full childbearing years, based on fertility rates from that specific year. For the year 2022, no data was collected for this variable, which is indicated by the minimum summary statistic of 0.

Mean Fertility.rate	Median Fertility.rate	Minimum Fertility.rate	Maximum Fertility.rate
6.888079	7.4	0	7.719

Variable name: ***Net.migration***

Variable type: Numeric

Description: This variable represents the net total of migrants during each fiscal year calendar. It accounts for both citizens and non-citizens and is calculated by subtracting the number of emigrants from the number of immigrants.

Mean Net.migration	Median Net.migration	Minimum Net.migration	Maximum Net.migration
-39753.65	2029	-2127145	1834556

Variable name: ***Male.life.expectancy***

Variable type: Numeric

Description: This variable represents the lifespan of a male newborn if patterns of mortality at the time of their birth stayed the same throughout their whole lifetime. For the year 2022, no data was collected for this variable, which is indicated by the minimum summary statistic of 0.

Mean Male life expectancy	Median Male life expectancy	Minimum Male life expectancy	Maximum Male life expectancy
45.53148	43.709	0	61.193

Variable name: ***Female.life.expectancy***

Variable type: Numeric

Description: This variable represents the lifespan of a female newborn if patterns of mortality at the time of her birth stayed the same throughout her whole lifetime. For the year 2022, no data was collected for this variable, which is indicated by the minimum summary statistic of 0.

Mean Female life expectancy	Median Female life expectancy	Minimum Female life expectancy	Maximum Female life expectancy
48.92089	48.397	0	66.677

Variable name: *Population.category*

Variable type: Factor

Description: This variable takes on the value of "Low Population" for years where the population total was under 1,000,000, "Medium Population" for years where the population was between 1,000,000 and 2,000,000, and "High Population" for years where the population was greater than 2,000,000.

Population.category	freq	mean_population
Low Population	8	9274297
Medium Population	34	12882963
High Population	21	30839118

Variable name: *F.life.expectancy.category*

Variable type: Factor

Description: This variable takes on the value of "Lowest life expectancy" for specific years where women's life expectancy is less than 40 years, "Lower life expectancy" for specific years where women's life expectancy is between 40 and 50 years, "Medium life expectancy" for specific years where women's life expectancy is between 50 and 60 years, and "Highest life expectancy" for specific years where women's life expectancy is over 60 years.

F.life.expectancy.category	freq	mean.life.expectancy
Lowest life expectancy	17	34.55341
Lower life expectancy	16	43.25000
Medium life expectancy	14	55.80014
Highest life expectancy	16	63.83787

Variable name: *M.life.expectancy.category*

Variable type: Factor

Description: This variable takes on the value of "Lowest life expectancy" for specific years where a man's life expectancy is under 30 years, "Lower life expectancy" for specific years where a man's life expectancy is between 30 and 40 years, "Medium life expectancy" for specific years where a man's life expectancy is between 40 and 50 years, "Higher life expectancy" for specific years where a man's life expectancy is between 50 and 60 years, and "Highest life expectancy" for specific years where a man's life expectancy is greater than 60 years.

M.life.expectancy.category	freq	mean.life.expectancy
Lowest life expectancy	3	19.89633
Lower life expectancy	25	36.18668
Medium life expectancy	7	43.94543
Higher life expectancy	21	55.80533

M.life.expectancy.category	freq	mean.life.expectancy
Highest life expectancy	7	60.65671