"phase2_data.csv"

EDA Final Project Data, Team 9

Description

The following is a codebook that serves as a data dictionary for all the variables used for Phase 2 of the final EDA project by Team 9. The complete data is presented as a table in the "phase2_data.csv" file. The primary key for this table is the Country variable. The data has been collected from multiple different Internet sources (please see the references page in the final report for proper citations). Several cleaning steps involving conditional R programming were needed to produce the data in its final form. The code for this data wrangling can be found in the "data_scraping_and_combining.R" file.

Variable	Description
Country	Each observation is one of the countries of the world. This serves as the primary key.
Climate_Zone	Stores the codename for the dominant climate type for each country, as per the Köppen classification. *
Avg_Temp_F	The average temperature for each country in the year 2019, in °F.
Avg_Temp_C	The average temperature for each country in the year 2019, in °C.
Freedom_Score	A quantitative (numerical) measure of how free (democratic/non-authoritarian) a country is overall, with scores ranging from 1 to 100. The higher the score, the more freedom the country has.
Freedom_Status	A qualitative (categorical) measure of how free (democratic/non-authoritarian) a country is, which is related to the Freedom_Score. There are 3 possible levels (factors): "Not Free", "Partly Free" and "Free".
Political_Rights	A quantitative (numerical) measure of how free a country is in terms of electing their leaders and organizing new political parties. The scores range from -3 to 40. The higher the score, the more freedom the country has.
Civil_Liberties	A quantitative (numerical) measure of how free a country is in terms of independent media, speech, religion, and assembly. The scores range from 2 to 60. The higher the score, the more freedom the country has.
Population	The total population for each country in the year 2019.

GDP	The total Gross Domestic Product (GDP) for each country in the year 2019, in United States Dollars (USD).
HDI	The Human Development Index (HDI) for each country in the year 2019.
Heart_Disease_Deaths	The total number of deaths from ischemic/coronary heart disease for each country in the year 2019.
Healthcare_Spending	The government healthcare expenditure per capita for each country in the year 2019, in United States Dollars (USD).

* Below are more details regarding the Climate_Zone variable:

- Each country is assigned one climate zone. This is the dominant climate zone for that country, meaning that the majority of the country's territories have this type of climate.
- The categories for the different climate zones come from the Köppen climate classification scheme, which classifies climates into 5 main groups:
    - A (Tropical)
    - B (Dry)
    - C (Temperate)
    - D (Continental)
    - E (Polar)
- These 5 main groups are then further divided into different sub-categories. Our data includes 18 of these sub-categories.
- Instead of the full climate type name, the main data only stores the codenames for the sake of brevity. The corresponding climates of these codenames are outlined below:
    - AF: Tropical rainforest
    - AM: Tropical monsoon
    - AW: Tropical Savannah
    - BWH: Hot desert
    - BWK: Cold desert
    - BSH: Hot semi-arid (steppe)
    - BSK: Cold semi-arid (steppe)
    - CWA: Dry-winter humid subtropical
    - CWB: Dry-winter subtropical highland
    - CFA: Humid subtropical
    - CFB: Oceanic
    - CSA: Mediterranean hot summer
    - CSB: Mediterranean warm/cool summer
    - DFB: Cold continental with no dry season, warm summer
    - DFC: Subarctic
    - DSB: Warm, dry-summer continental
    - DSC: Dry-summer subarctic
    - ET: Tundra