Robert Cheadle

6.1: Sourcing Open Data

**Data Source:**

**Data Set:** Life Expectancy at Birth, Total (Years)

**Source:** "Life Expectancy at Birth, Total (Years)", Published online at OurWorldInData.org. Retrieved from: 'https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators' [Online Resource]

**Data Collection:**

“Life Expectancy at Birth, Total (Years)” is an external and . The data since 1961, which will be used in this analysis is available in the World Development Indicators (WDI) published by the World Bank. This data is collected on an annual basis by the United Nations Population Division. This data can be considered trustworthy.

**Content:**

The contents of this data set include location data describe by ‘Entity’(country) and ‘Code’(country code). Date data described as ‘Year’ and quantitative data ‘Life expectancy at birth, total (years).’

Life expectancy at birth is highly sensitive to the rate of death in the first few years of life. This analysis will only be considering life expectancy at birth not at different age groups. Also, it is important to note that period and cohort life expectancy estimates are statistical measures, and they do not consider any person-specific factors such as lifestyle choices.

Relevance:

This data set is relevant to the project hypothesis and objective presented because life expectancy will give insight into how a countries gross domestic product (GDP), population, carbon-dioxide emissions from production overtime may or may not show significant correlation or other relationships.

Data Set: CO₂ and Greenhouse Gas Emissions

Source: Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "CO₂ and Greenhouse Gas Emissions". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions' [Online Resource]

Data Collection:

This dataset is a collection of key metrics maintained by Our World in Data. It is updated regularly and includes data on CO2 emissions (annual, per capita, cumulative, and consumption-based), other greenhouse gases, energy mix, and other relevant metrics. This dataset is a compilation of data from other data sources such as the Global Carbon Project and BP.

Content:

The contents of this data set include location data described as ‘country’ and ‘iso\_code,’ a 3-letter country code. As well as quantitative data relevant to this project in the form of population, GDP, and co2 (annual total production-based emissions), measured in million tons.

Limitations of this data are we do not know the data collection and quantitative methods of the original sources. As well as any potential bias those sources may have in their collection and measurements.

Relevance:

This data set is relevant to the project hypothesis and objective presented because a countries gross domestic product (GDP), population, carbon-dioxide emissions from production overtime may or may not show significant correlation to a country’s life expectancy at time of birth.

**Data Profile:**

Table

Description automatically generated

* The above table displays consistency checks performed on raw and cleaned datasets.

Table

Description automatically generated

Table

Description automatically generated

* The above table shows all columns dropped from raw and cleaned datasets, also includes column name changes.

Table

Description automatically generated

* The above table includes information about each variable in the dataset and the type of data each variable consists of.

Chart

Description automatically generated with medium confidence

* The above table shows that the frequency of each value is consistent.

Table

Description automatically generated

* The above table shows some basic descriptive information about each variable found in the data set.

Chart

Description automatically generated

* The above tables show the correlation between the different variables of the data set. This information will help to determine how the variables relate to one another.

**Questions to explore:**

1. How does a country’s GDP/capita effect life expectancy? Derive new column ‘GDP per capita’
2. How does a country’s co2(production)/capita effect life expectancy? Derive new column ‘co2 per capita’
3. Do countries with high co2 production and GDP have higher life expectancy at a significant level?
4. Do countries with medium co2 production and GDP have higher life expectancy at a significant level?
5. Do countries with low co2 production and GDP have lower life expectancy at a significant level?
6. How does the rate of change from year to year for (GDP, co2, and life\_expectancy) compare? Derive rate of change variables for GDP, co2, and life expectancy.
7. Does a decrease in co2 production necessarily result in decreased GDP?
8. Does increased co2/capita have any effect on a country’s life expectancy?