Sydney Hoy

6.1: Sourcing Open Data

Data Source

Dataset: “**Life Expectancy based on Geographic Locations**”

Summary

Source: <https://www.kaggle.com/datasets/saurabhbadole/life-expectancy-based-on-geographic-locations/data>

Type: External

Owner: World Health Organization

Trustworthiness: World Health Organization is public health centered international organization, so it is regarded as being a reliable data source.

Collection

Type: Administrative

Method: Data is collected directly from countries that work with the World Health Organization.

Contents

This data set contains life expectancy data and identifiers from 182 countries.

Reason

As someone with a background in healthcare and biology I chose a topic that could relate to my experience, but also branch out and look into data that could relate to other career paths . Public health is a very interesting topic to me, and I’ve always been intrigued to learn about the differences in our world.

Data Profile

<https://docs.google.com/spreadsheets/d/1IlwiqShqSN0UZ4-Jd4daEmZU0Em-NBmeoiZVEY5yycQ/edit?usp=sharing>

Limitations

This data depends on data from countries official health resources to report the data to WHO so that it can be published. This leads to gaps in data if all countries do not report the same type of data or report consistently. I was able to collect some missing data from other reliable resources, but the data may have been collected differently and have slightly different numbers. Some countries data was also averaged if only some of the years were missing from the variable.

Ethics

This dataset does not contain and personal data or sensitive information, however it does contain information that can be considered sensitive because it is dealing with topics such as death and infections of a country’s population.

Bias

Reporting bias may be present here as there are inconsistencies in reporting across the countries which could affect the statistics.

Questions

1. Has life expectancy increased or decreased from 2000 – 2014?
2. Does GDP indicate a high life expectancy?
3. Does HDI indicate a high life expectancy?
4. Does population effect life expectancy positively or negatively?