Documentation

Rajvardhan Jhanwar

2023-10-11

1. Background

Life expectancy is a critical indicator of the overall health and well-being of a population. It reflects the average number of years a person can expect to live, and it is influenced by a myriad of factors, ranging from healthcare systems and economic conditions to environmental sustainability and government policies. Understanding the determinants of life expectancy is essential for policymakers and public health professionals to make informed decisions aimed at improving the quality and longevity of life for individuals around the world.

2. Project purpose

The purpose of this project is to develop a Shiny web application that enables users to explore and visualize data related to life expectancy in different countries. The application provides a user-friendly interface to interact with the data, offering insights into factors that influence life expectancy and facilitating data-driven decision-making.

3. Data description + How was the data collected

One primary dataset and three auxiliary datasets are used to explore the relationship between various factors and life expectancy in countries across the world between 2000 and 2015. The datasets were transformed into long format datasets and merged into one to facilitate comprehensive analysis.

a. Primary Dataset on Life Expectancy Source – Kaggle, sourced from the World Health Organization

b. Dataset on Unemployment Rate of Countries Source – The World Bank Open Data

c. Dataset on Inflation Rate of Countries by Annual Consumer Prices Source – The World Bank Open Data

d. Dataset on Land Area of Countries Source – The World Bank Open Data

Variables Selected – i. Demographics: Country, Year, Life Expectancy, Population ii. Environmental: CO2 Emissions, Forest Area iii. Health: Health Expenditure, Obesity, Open Defecation, Potable Water Access iv. Economic: GDP per Capita, Beer Consumption Per Capita

During the period from 2000 to 2015, the global landscape witnessed significant changes in economic, governmental, demographic, and environmental factors. This era marked a time of rapid globalization, technological advancement, and shifting geopolitical dynamics, all of which had profound implications for the health and well-being of populations. Against this backdrop, it becomes imperative to investigate the multifaceted factors that may have influenced life expectancy in countries across the globe during this period especially the impact of increased CO2 emissions and decreased forest area percentage on the life expectancy.

4. Who are the users that this Shiny App was made for?

This Shiny App is designed for a broad audience, including researchers, policymakers, healthcare professionals, educators, and anyone interested in exploring the factors affecting life expectancy worldwide. The application caters to both experts and the general public, providing a user-friendly interface for users with varying levels of expertise in data analysis. The filters provided intend at spoiling the users with choice and saving their time.

5. What questions are you trying to answer?

The project aims to address various questions like:

a. What is the global distribution of life expectancy, and how has it changed over time?
b. How do various Demographics, Environmental, Health, and Economic factors influencing the life expectancy?
c. Are there regional disparities in life expectancy, and if so, what factors contribute to these disparities?
d. What is the correlation between different variables and life expectancy, and which variables have the most significant impact?
e. How have these relationships evolved over time?

6. What insights did you get from your data?

Some of the key insights obtained from the data analysis include:

a. Life expectancy has increased globally over the years, indicating improvements in healthcare and living conditions.
b. Higher GDP per capita and increased healthcare expenditure are generally associated with longer life expectancy while military expenditure decreases it.
c. There are significant regional disparities in life expectancy, with some regions showing consistently higher or lower life expectancies.
d. The impact of various factors on life expectancy has evolved over time, with some variables becoming more influential especially in recent years.

7. Desired Improvements

a. Conducting more in-depth statistical analyses to uncover nuanced insights from the data.
b. Incorporating machine learning models to predict future life expectancy trends based on historical data.

8. Sources and References

9. Kaggle. 2017. “Life Expectancy and Related Factors” Dataset. https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who?resource=download.

10. The World Bank. 2022. “Inflation, consumer prices (annual %)” Dataset. https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG

11. The World Bank. 2022. “Unemployment, total (% of total labor force)” Dataset. https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS

12. The World Bank. 2021. “Land Area (sq. km)” Dataset. https://data.worldbank.org/indicator/AG.LND.TOTL.K2

13. Web Application Development with R Using Shiny – Beeley C

Process and development –>

1. What

In this project, the main task was to create a Shiny web application for visualizing and analyzing life expectancy data. The goal was to provide users with an interactive platform to explore and gain insights into factors influencing life expectancy across different countries and regions.

2. Why

- Tidying Data: Data tidying involved transforming raw data into a clean, structured format suitable for analysis. This was done to ensure data quality, consistency, and ease of use. I addressed missing values and outliers to maintain data integrity, reformatted date and time variables to ensure they were in a standardized format, and consolidated data from various sources to create a comprehensive dataset for analysis.
- Encoding/Mapping Decisions: The encoding and mapping decisions were made to represent data visually and convey information effectively. I considered the following aspects:

a. Color Mapping: I used color to represent categorical variables such as continents or regions. Color choices were made to provide visual distinction between categories while ensuring accessibility for all users (considering colorblindness).
b. Size Mapping: Size was used to encode quantitative variables such as population, allowing users to quickly grasp the relative magnitudes of data points.
c. Position Encoding: I used position encoding to represent time on the x-axis, making it easy to understand the temporal aspect of the data.

3. How

- Tidying Data: Data tidying was performed using the tidyverse package in R, which includes tools for data manipulation and cleaning. I followed a systematic process, including:

a. Using functions like mutate and filter to clean and transform variables.
b. Ensuring documentation and code comments to explain each data transformation step.
c. Using version control to track changes and maintain data reproducibility.

- Encoding/Mapping Decisions: The encoding and mapping decisions were implemented using the ggplot2 package in R for data visualization. The package provides a consistent and reproducible way to create visualizations. Specific implementation steps included:

a. Defining aesthetics within the ggplot() function to specify the mapping of data to visual properties (e.g., color, size, position).
b. Using best practices for color choice to ensure readability and accessibility.
c. Applying consistent and well-documented themes and formatting to visualizations for reproducibility.

By adhering to these practices, I aimed to make the project's data tidying and visualization steps both reproducible and understandable to others. This would hopefully make it possible for other researchers and collaborators to reproduce and extend the work while maintaining data quality and visualization consistency.