**DS210 Final Project Writeup: Exploring Global Socioeconomic Dynamics**

By Abigail Merage

**Dataset Overview:**

The dataset utilized in this project comprises country-specific information sourced from the Global Socio-Economic & Demographic Insights dataset, which was derived from Kaggle. Each entry includes the country's name, region, subregion, population density, and urban population growth, among other indicators. The geographical and demographic data are crucial for understanding the socio-economic dynamics between countries. The dataset allows for a nuanced exploration of global trends, offering a foundation for meaningful analysis.

Link to Dataset: <https://www.kaggle.com/datasets/samybaladram/databank-world-development-indicators/data>

**Code Implementation:**

The project consists of seven main modules and a test:

1. **Data Loading:** The data\_loading module handles the loading of country data from a CSV file. Each country's information includes its name, region, subregion, population density, and urban population growth. The load\_data function reads this data from the specified CSV file and returns a vector of CountryData structs.
2. **Country Graph**: The country\_graph module is an implementation for representing and managing a graph structure, specifically tailored for analyzing relationships between countries. It utilizes a HashMap to store an adjacency list, capturing the connections between countries. The module provides methods to add edges, retrieve neighbors for a given country, check the existence of a country, and initialize a new instance of the graph. This abstraction serves as a fundamental component exploring connections between countries and conducting Six Degrees of Separation analyses.
3. **Country Data:** The country\_data module defines a struct representing individual country-specific data. It encapsulates essential socio-economic indicators, including the country's name, region, subregion, population density, and urban population growth. The module includes a constructor method, new, facilitating the creation of new CountryData instances with specified attributes. This struct is designed to encapsulate detailed information about countries, serving as a foundation for organizing and processing data within the broader project context.
4. **Graph Processing:** The graph\_processing module builds a graph representing relationships between countries. The relationships are established based on geographical proximity, specifically the shared subregion. The build\_graph function creates a graph with countries as nodes and edges connecting nearby countries. The module also provides functions for checking if countries are nearby and performing Six Degrees of Separation analysis starting from a given country.
5. **Visualization:** The visualization module utilizes the plotlib crate to create scatter plots for visualizing the relationship between population density and urban population growth. The visualize\_data function takes a vector of CountryData and generates a scatter plot, where each point represents a country.
6. **Main:** The main.rs file serves as the entry point for the project, orchestrating various functionalities to analyze and visualize socio-economic and demographic data. It begins by importing necessary modules, including data\_loading, graph\_processing, and visualization. The main function then executes a sequence of operations. Firstly, it loads the dataset from the specified CSV file using the load\_data function. Subsequently, it constructs a country graph and performs a Six Degrees of Separation analysis starting from a specific country. The chosen starting country is "Afghanistan," and the maximum degrees are set to 2. Finally, it invokes the visualize\_data function to generate visual representations of the dataset.
7. **Lib:** The lib.rs file serves as the central hub for organizing and providing access to the project's functionality. It exports essential components from three distinct modules: data\_loading, graph\_processing, and visualization. This modular approach encapsulates related functionalities, promoting code organization and maintainability. Users can leverage the functionalities exposed in this file by importing them into other parts of the project.

**Test:** The integration\_tests.rs file contains an integration test for the entire project. It utilizes the assert\_cmd crate to execute a command-line binary built from the project and ensures its success. The test initializes a command, asserts its success status, and then concludes the test. This integration test serves as a crucial component in validating the overall functionality and correctness of the project, providing a systematic way to verify its behavior.

**How to Run the Project:**

1. Ensure you have Rust installed on your system.
2. Clone the project repository: git clone <repository-url>
3. Navigate to the project directory: cd global\_socioeconomic\_analysis
4. Run the project using Cargo: cargo run

**Output and Analysis:**

Upon running the project, the console output provides information about countries within a specified degree of separation from a chosen starting country. This insightful analysis offers a glimpse into the interconnectedness of nations, emphasizing the importance of geographical proximity in shaping global dynamics.

The scatter plot output enhances the understanding of the dataset, showcasing the distribution of countries based on population density and urban population growth. Each point on the plot represents a country, and the visualization helps identify clusters and outliers, facilitating a deeper exploration of the dataset.

Sub-Saharan Africa emerges as a focal point in the analysis, exhibiting the highest overall average urban population growth rate. This finding underscores dynamic urbanization trends in the region, signifying potential economic and demographic shifts. Conversely, Eastern Europe presents a contrasting narrative with a significant negative growth rate. This observation suggests a decline in urban population, possibly influenced by demographic and economic factors unique to the region.

These regional nuances contribute to a richer interpretation of the dataset, allowing for a more comprehensive understanding of global socioeconomic dynamics. The integration of insightful console outputs and visually appealing scatter plots transforms raw data into actionable insights, making the Global Socioeconomic Analysis project a valuable tool for researchers and analysts.

**Sources:**

<https://stackoverflow.com/questions/32522870/unable-to-find-crate-that-is-listed-in-build-dependencies-section>

<https://www.kaggle.com/datasets/samybaladram/databank-world-development-indicators?select=world_development_data_imputed.csv>

<https://docs.rs/plotpy/latest/plotpy/>

<https://github.com/milliams/plotlib>