Hoang Minh Khue – Le Tran Bao Kien

[Email address]

<https://mercury.swin.edu.au/cos30020/s104223584>

Team member: Hoang Minh Khue, Le Tran Bao Kien  
Tutor Date: Monday morning, 9 AM – 12 PM  
Spring 2024  
Word Count:

Global Migration Dynamics: Visualizing Patterns and Contexts

[Document subtitle]

Table of Contents

[1. Introduction 3](#_Toc163313056)

[1.1. Background and Motivation 3](#_Toc163313057)

[1.2. Visualization Purpose 3](#_Toc163313058)

[1.3. Project Schedule 4](#_Toc163313059)

[2. Data 5](#_Toc163313060)

[2.1. Data Source 5](#_Toc163313061)

[2.2. Data Processing 6](#_Toc163313062)

[Hand Processing: 6](#_Toc163313063)

[**World Bank( www.worldbank.org ):** 7](#_Toc163313064)

[**United Nations ( www.un.org ):** 10](#_Toc163313065)

[Code Processing: 13](#_Toc163313066)

[3. Requirements 13](#_Toc163313067)

[3.1. Must-have Features: 13](#_Toc163313068)

[3.1.1. Interactive Visualization: 13](#_Toc163313069)

[3.1.2. Additional Information: 14](#_Toc163313070)

[3.1.3. Clear Data Labeling: 14](#_Toc163313071)

[3.1.4. Color Contrast: 14](#_Toc163313072)

[3.2. Optional Features: 14](#_Toc163313073)

[3.2.1. Tooltip Integration: 14](#_Toc163313074)

[3.2.2. How To Read Section: 14](#_Toc163313075)

[4. Visualization Design: 14](#_Toc163313076)

[4.1. Conceptualize (Design Iteration) stage: 15](#_Toc163313077)

[4.2. Visualize Stage: 16](#_Toc163313078)

[4.3. Materialize Stage: 17](#_Toc163313079)

[4.4. Implementation Stage: 18](#_Toc163313080)

[4.5. Storyboard: 18](#_Toc163313081)

[5. Final Design 18](#_Toc163313082)

[6. Conclusion 18](#_Toc163313083)

[References 19](#_Toc163313084)

# Introduction

## Background and Motivation

One of the key characteristics of human history has been migration, which has shaped civilizations, economies, and cultures on many continents and during many eras. People have moved about throughout history as a necessary component of human existence, from the ancient nomadic tribes to the contemporary diasporas fueled by globalization. The topic of global migration has become increasingly important and well-known in the modern era due to a variety of causes, including political unrest, economic inequality, environmental damage, and climate change.

In a time of growing globalization and interconnection, the dynamics of human mobility are more intricate and nuanced than in the past. There are many different reasons why people move, such as to pursue better job prospects, further their education, be with family, or seek safety from persecution or violence. Furthermore, population movements are driven by environmental causes including natural catastrophes, environmental degradation, and climate change, which exacerbate already-existing vulnerabilities and inequities.

It is impossible to exaggerate the importance of the global migration issue in the modern world. As the number of individuals living in forced displacement reaches previously unheard-of heights, migration has become a significant humanitarian, social, political, and economic concern. It has significant ramifications for host communities, transit nations, and international geopolitics and has a significant impact on the rights and well-being of migrants. Furthermore, a wide range of other urgent challenges, including urbanization, sustainable development, labor markets, social cohesion, and human rights, are intertwined with migration.

## Visualization Purpose

It is becoming more and more necessary to visualize migration patterns and contexts due to the complexity and difficulties involved in global migration. A potent tool for deciphering the complex dynamics of migratory flows, exposing temporal and spatial trends, finding patterns and correlations, and locating underlying causes and effects is data visualization. Visualization helps stakeholders, policymakers, scholars, and the public better understand the scope and complexity of migration phenomena by converting complex statistics into understandable and visually appealing representations.

Furthermore, migration data visualization acts as a catalyst for comprehending and resolving the underlying reasons and effects of migration. It offers insightful information about the geographic distribution of migrants, their travel routes and corridors, the demographics of migrant communities, and the social, economic, and environmental settings in which migration takes place. Equipped with such discernments, policymakers can devise empirically grounded tactics and measures to tackle the obstacles presented by migration, alleviate its adverse effects, and capitalize on its prospective advantages for both migrant populations and receiving communities.

To summarize, the visualization of migration patterns and contexts serves two purposes: first, it improves comprehension of the intricate dynamics of global migration; second, it provides guidance and information for policy and decision-making processes that aim to address the obstacles and maximize the benefits associated with human mobility. Through the effective use of visualization, we can shed light on the paths and histories of migration, give voice to the opinions and experiences of migrants, and help create more just inclusive, and sustainable societies in a global community that is becoming more interconnected by the day.

## Project Schedule

**Week 2-4: Project Kickoff and Data Collection**

* Define project objectives and scope.
* Assign roles and responsibilities within the team.
* Identify and gather relevant data sources on global migration.
* Begin initial exploration and assessment of the collected data.

**Week 5-6: Data Cleaning and Processing**

* Conduct thorough data cleaning to address missing values, inconsistencies, and outliers.
* Standardize data formats and ensure data compatibility.
* Perform any necessary data transformations or aggregations to prepare the data for visualization.

**Week 7-8: Design and Prototyping**

* Brainstorm visualization ideas and design concepts.
* Develop initial prototypes of visualization layouts and elements.
* Gather feedback from team members and stakeholders to refine the design.

**Week 9-10: Implementation**

* Translate finalized design concepts into code using chosen visualization tools or libraries.
* Integrate cleaned and processed data into the visualization.
* Conduct iterative testing and debugging to ensure functionality and usability.

**Week 11: Refinement and Evaluation**

* Fine-tune visual elements, layouts, and interactions based on user feedback.
* Conduct thorough testing of the completed visualization across different devices and browsers.
* Evaluate the effectiveness and usability of visualization in conveying migration patterns and contexts.

**Week 11-12: Documentation and Presentation**

* Compile comprehensive documentation of the project process, including data sources, cleaning and processing methods, design decisions, and implementation details.
* Prepare a final presentation summarizing key findings, insights, and implications of the visualization.
* Submit the completed project process book and present the visualization to the instructor and class.

# Data

## Data Source

The primary sources for raw data in this project are the World Bank (worldbank.org) and the United Nations (un.org). These reputable organizations provide comprehensive datasets related to global migration, offering insights into migration trends, demographic characteristics, and socio-economic contexts across countries and regions worldwide.

* **The World Bank (** [**www.worldbank.org**](http://www.worldbank.org) **)**:
  + The World Bank offers a wealth of demographic and socio-economic data, including information on international migration patterns, immigrant populations, and migration-related variables such as country of birth, citizenship status, and migration flows. Data from the World Bank's Migration and Remittances dataset and other sources provide valuable insights into migration trends within countries and globally.
  + **Data Set 1**: Life Expectancy data from the World Bank's World Development Indicators dataset.
  + **Data Type**: The data set includes the following variables:
    1. Country Name (Categorical)
    2. Life Expectancy (Quantitative)
  + **File Type**: Raw data obtained in .csv format.
  + **Data Set 2:** GDP per Capita data from the World Bank's World Development Indicators dataset.
  + **Data Type:** The data set includes the following variables:
    1. Country Name (Categorical)
    2. GDP per Capita (Quantitative)
  + **File Type**: Raw data obtained in .csv format.
  + **Data Set 3:** Population data from the World Bank's World Development Indicators dataset.
  + **Data Type:** The data set includes the following variables:
    1. Country Name (Categorical)
    2. Population (Quantitative)
  + **File Type**: Raw data obtained in .csv format.
* **United Nations (** [**www.un.org**](http://www.un.org) **)**:
  + The United Nations is a leading source of data and research on global migration, providing comprehensive statistics and analyses on international migration trends, refugee movements, asylum seekers, internally displaced persons (IDPs), and other migration-related topics. Datasets from UN agencies such as the International Organization for Migration (IOM), the United Nations High Commissioner for Refugees (UNHCR), and the United Nations Department of Economic and Social Affairs (UN DESA) offer valuable insights into the scale, dynamics, and impacts of migration worldwide.
  + **Data Set**: Global Migration Statistics from the United Nations
  + **Data Type**: The data set includes the following variables:
    1. Country Name (Categorical)
    2. Number of Immigration People (Quantitative)
  + **File Type**: Raw data obtained in .xlsx format, converted to .csv format using [www.cloudconvert.com/csv-to-xlsx](http://www.cloudconvert.com/csv-to-xlsx) website.

These two main data sources serve as foundational pillars for this project, providing reliable and authoritative data that underpins the visualization of global migration patterns and contexts. Additional supplementary data sources may be explored to enrich the analysis and address specific research questions or objectives identified during the project's development.

## Data Processing

### Hand Processing:

Although the datasets obtained from the organizations guarantee high integrity and timeliness, data cleaning and reorganization are essential to achieve the desired level of "conformity" necessary for visualization in D3.js. The data processing procedure involves distinct steps for each of the retrieved datasets, as each one necessitates pruning, pivoting, and aggregating to a certain degree. Once the design selection is finalized, the dataset will be further filtered to align with the specifications of the data visualization design, marking the milestone of "conformity" and indicating that the data is prepared for visualization.

## **World Bank(** [**www.worldbank.org**](http://www.worldbank.org) **):**

The original data is structured as a CSV (Comma-Separated Values) file, with the first dataset for **Life Expectancy** containing rows representing different countries or regions and columns representing attributes such as the name of the country, country code, indicator (in this case, life expectancy at birth), and numerical values indicating life expectancy for each year. The second dataset was for **GDP per capita (Current US$)**, in which each row represents a country, and the columns include the country name, country code, indicator name (GDP per capita), indicator code, and GDP per capita values for different years. The years range from at least 1960 to 2022, with missing data represented by empty fields. The final dataset from World Bank is for the total population data. Each row represents a country or region, with columns indicating the name of the country or region, the country code, the type of data (in this case, total population), and population values for different years. The years span from 1960 to a more recent year, likely reflecting the most up-to-date available data. The population values are in numeric format. Additionally, there are rows representing aggregates such as "World" and "Upper middle income.". The following three figures depict the three said original datasets downloaded from the World Bank.

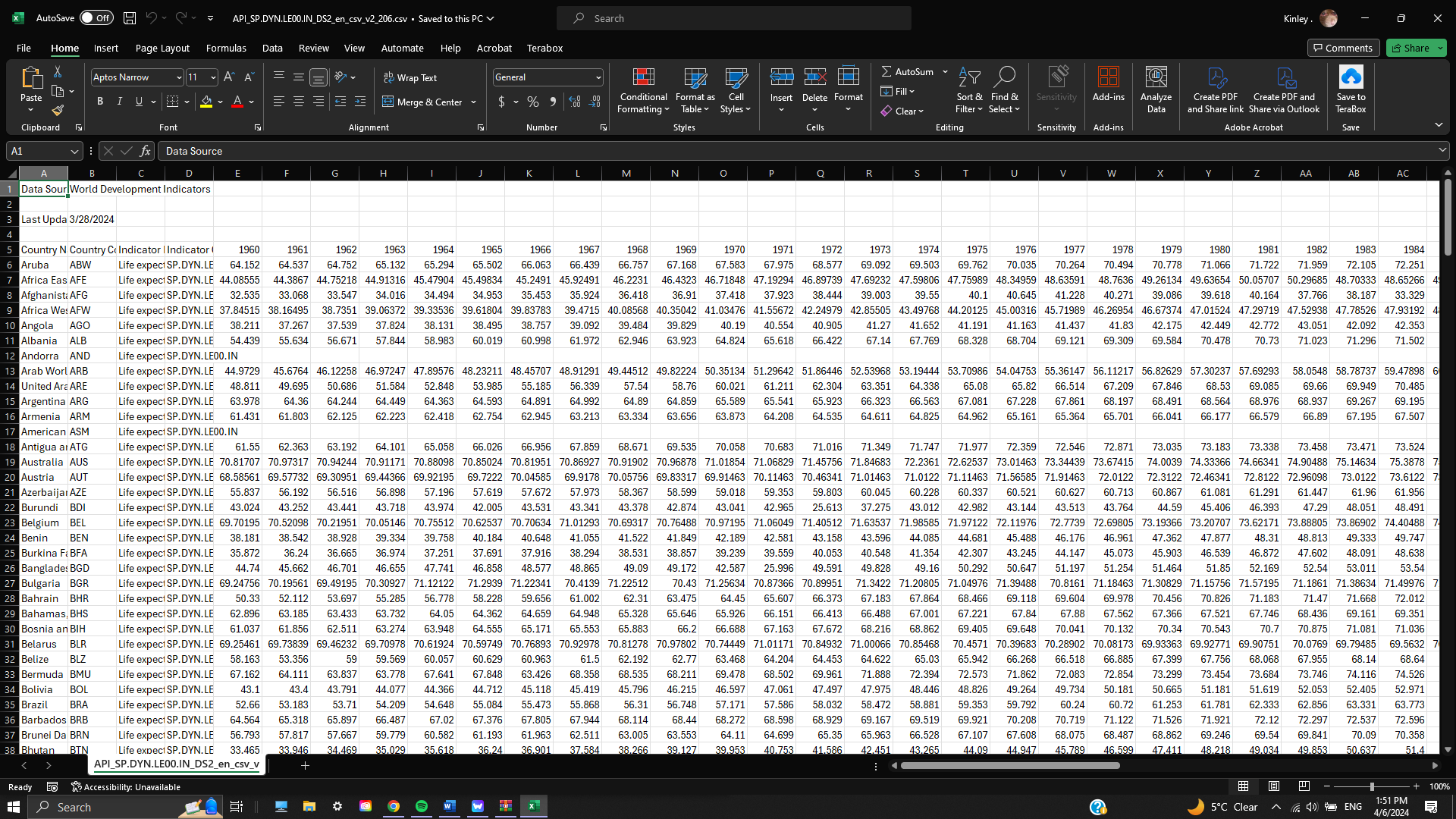


Figure 1: Life expectancy at birth from worldbank.org

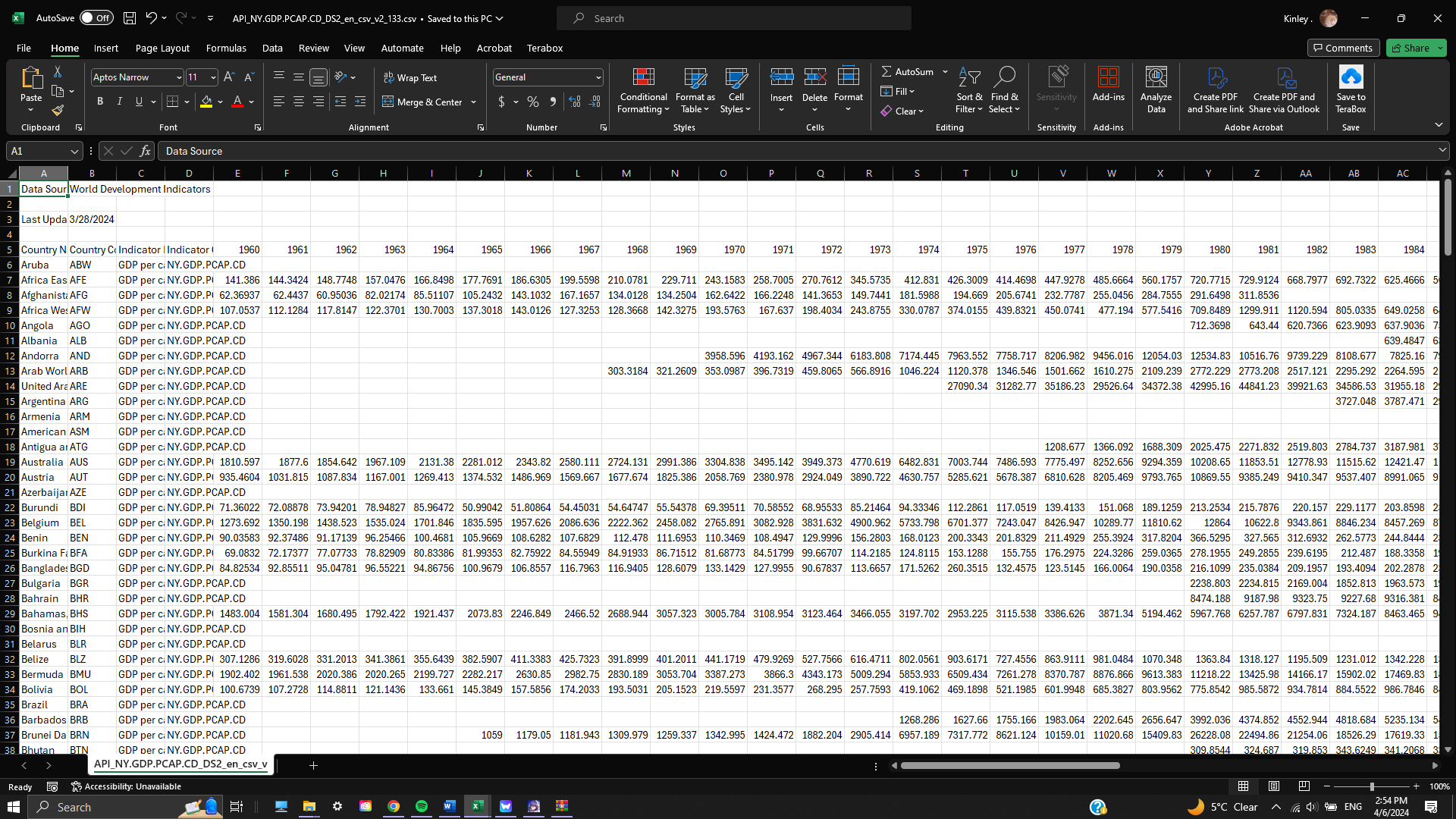


Figure 2: GDP per capita (current us$) from worldbank.org

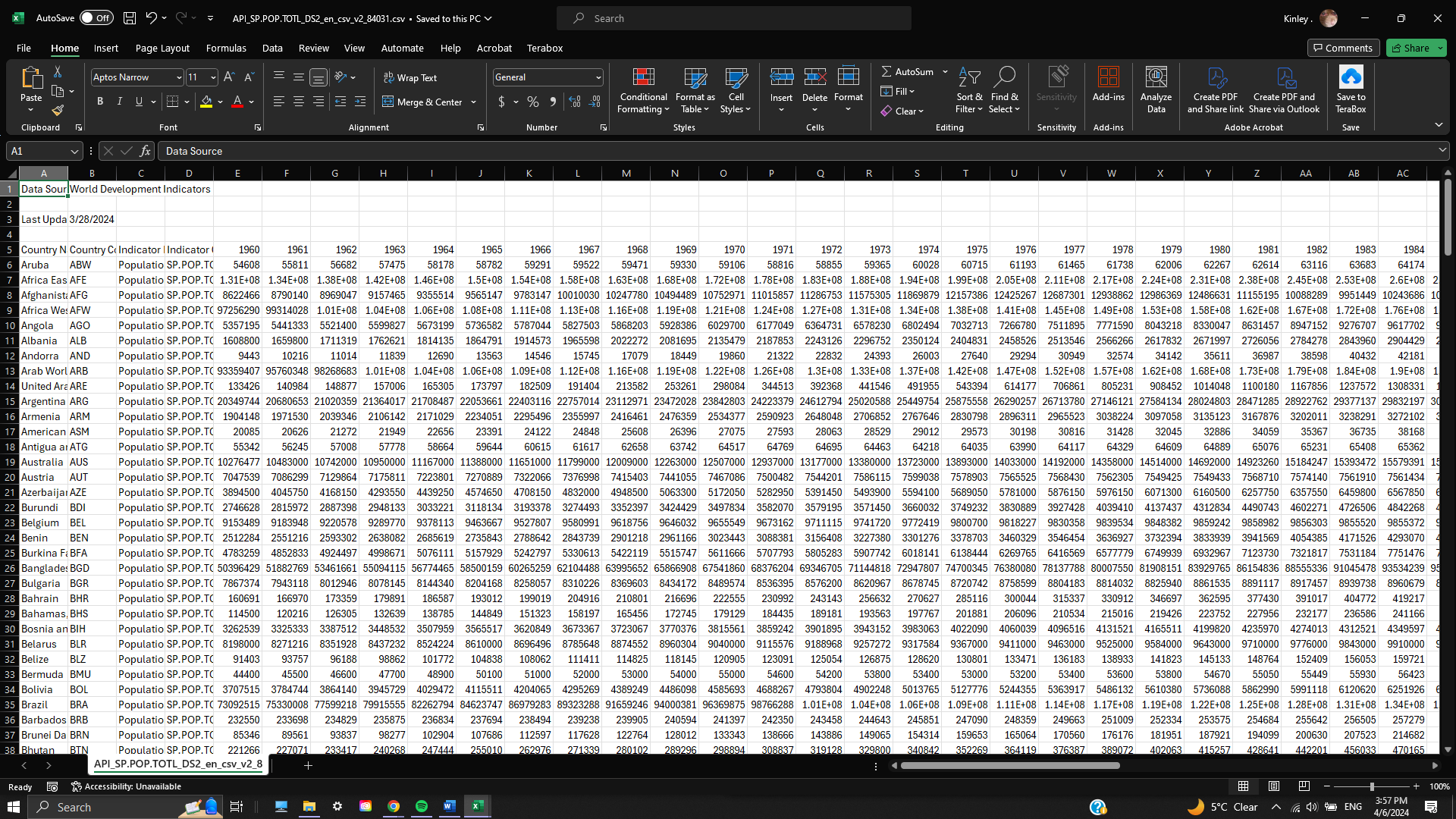


Figure 3: Population, total from worldbank.org

The first step of cleaning is to remove unnecessary rows and columns that do not contain important information needed for the visualization. This process was done by hand since all three files only contain a single table, and by cleaning and renaming the file to WB\_LifeExpect.csv, GDP.csv, and Population.csv respectively, we can better handle the file here on out. Figures 3, 4, and 5 depict the aforementioned.

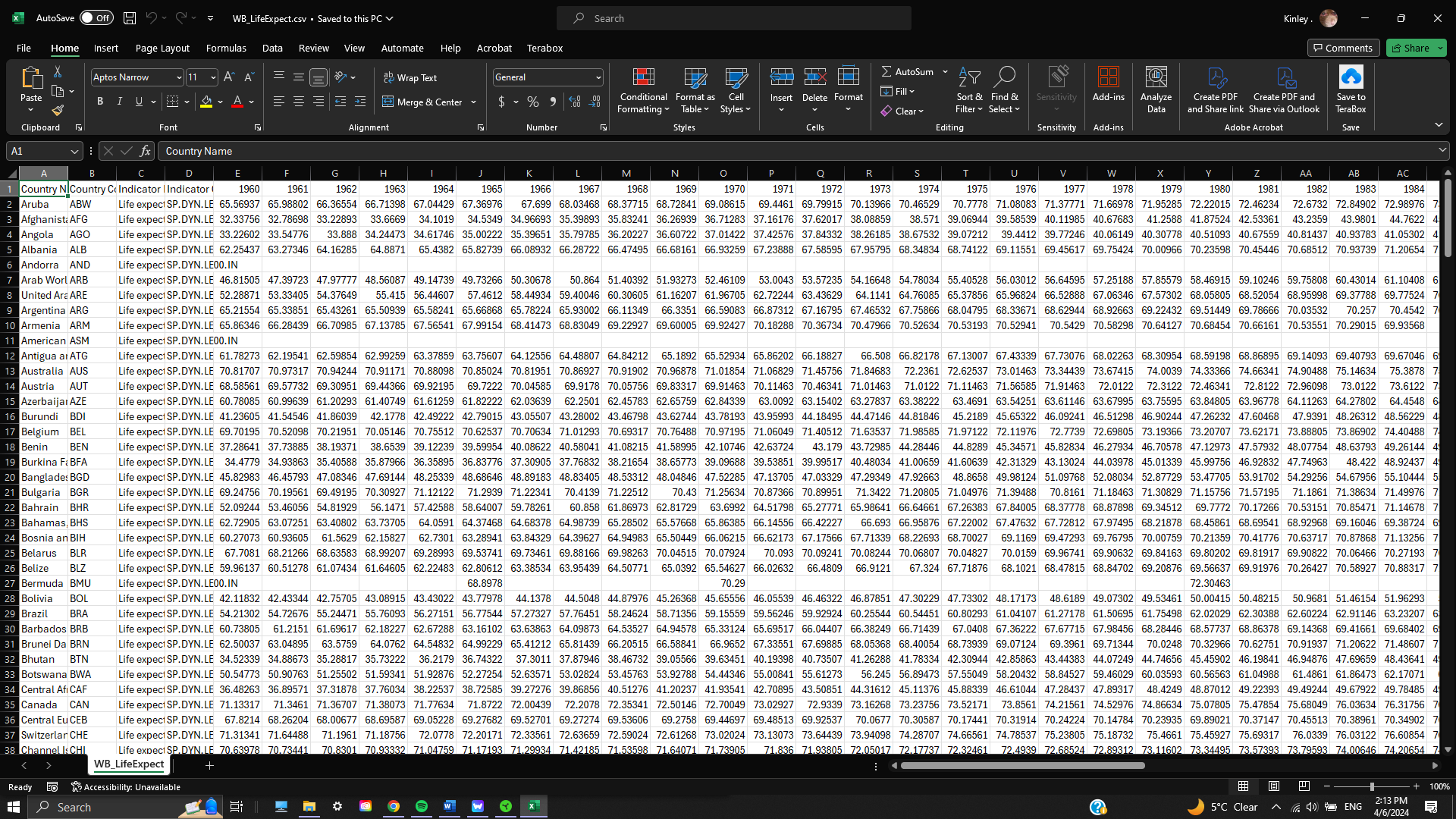


Figure 4: Life expectancy data after cleaning by hand

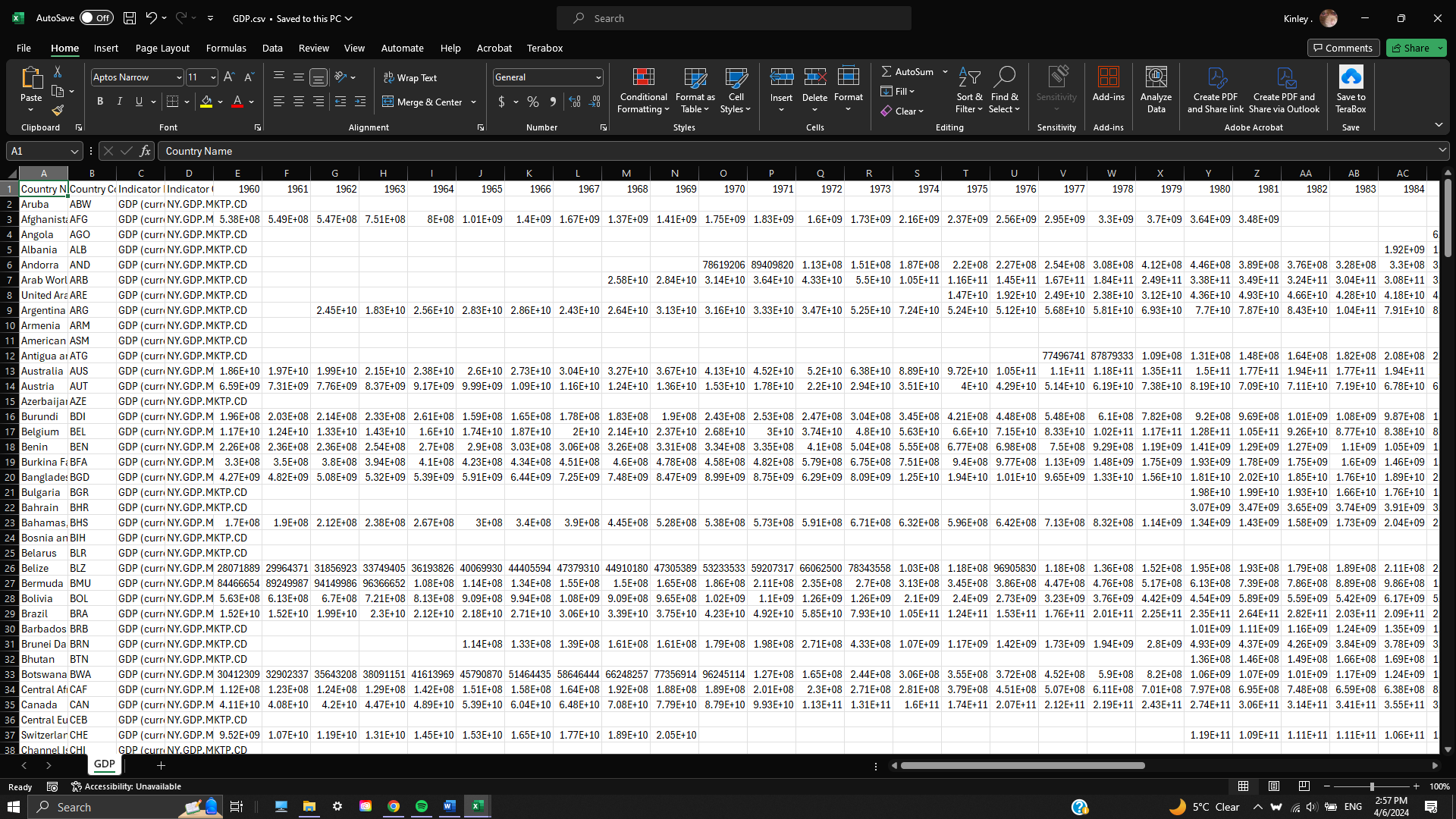


Figure 5: GDP per capita data after cleaning by hand

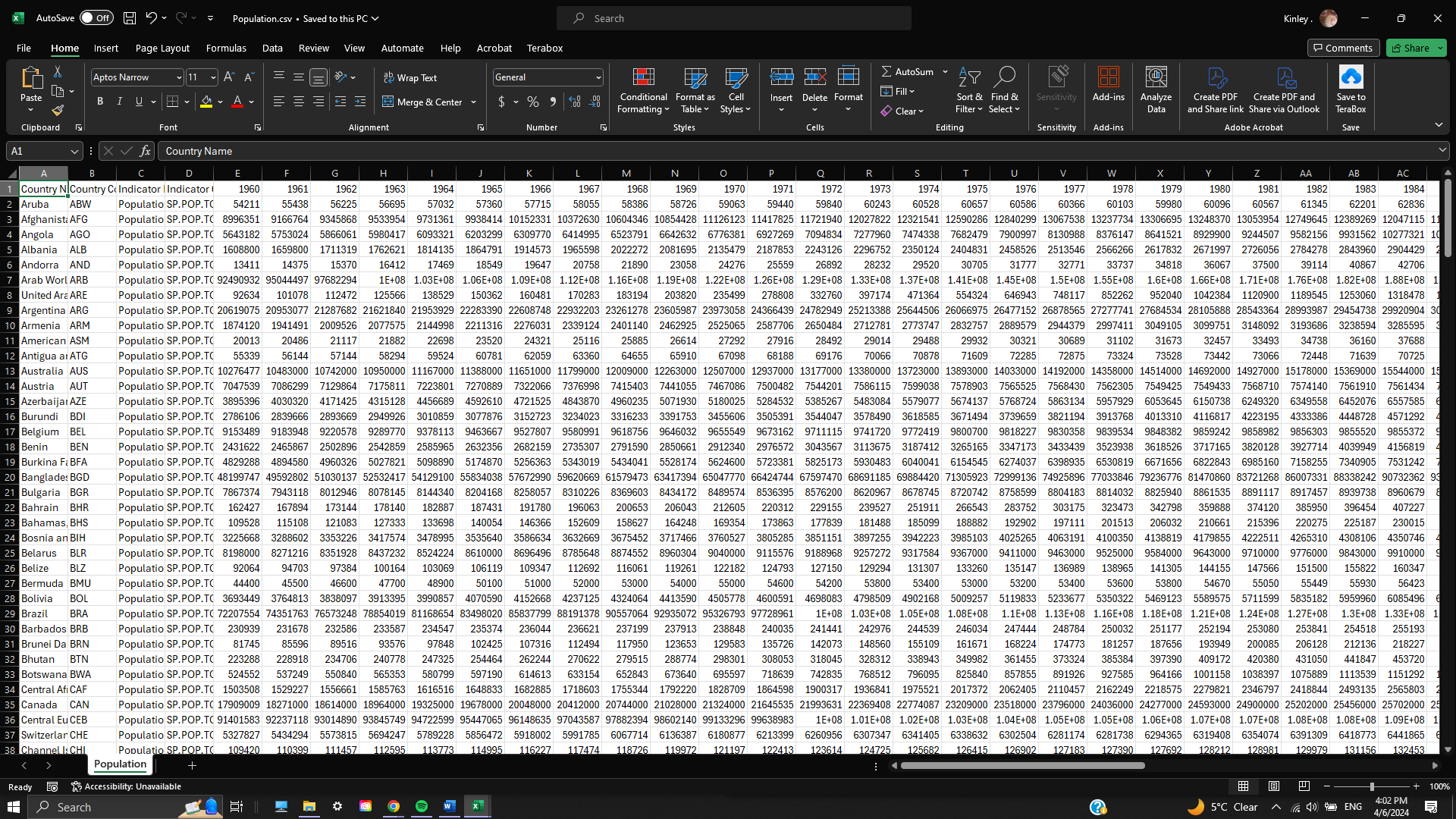


Figure 6: Total population data after cleaning by hand

## **United Nations (** [**www.un.org**](http://www.un.org) **):**

Different from the dataset gathered from the World Bank, the original data from United Nations is structured as a XLSX (Microsoft Excel Open XML Format Spreadsheet) file, with multiple tables each representing a different dataset. The following figure depicts the original data downloaded from the United Nations.

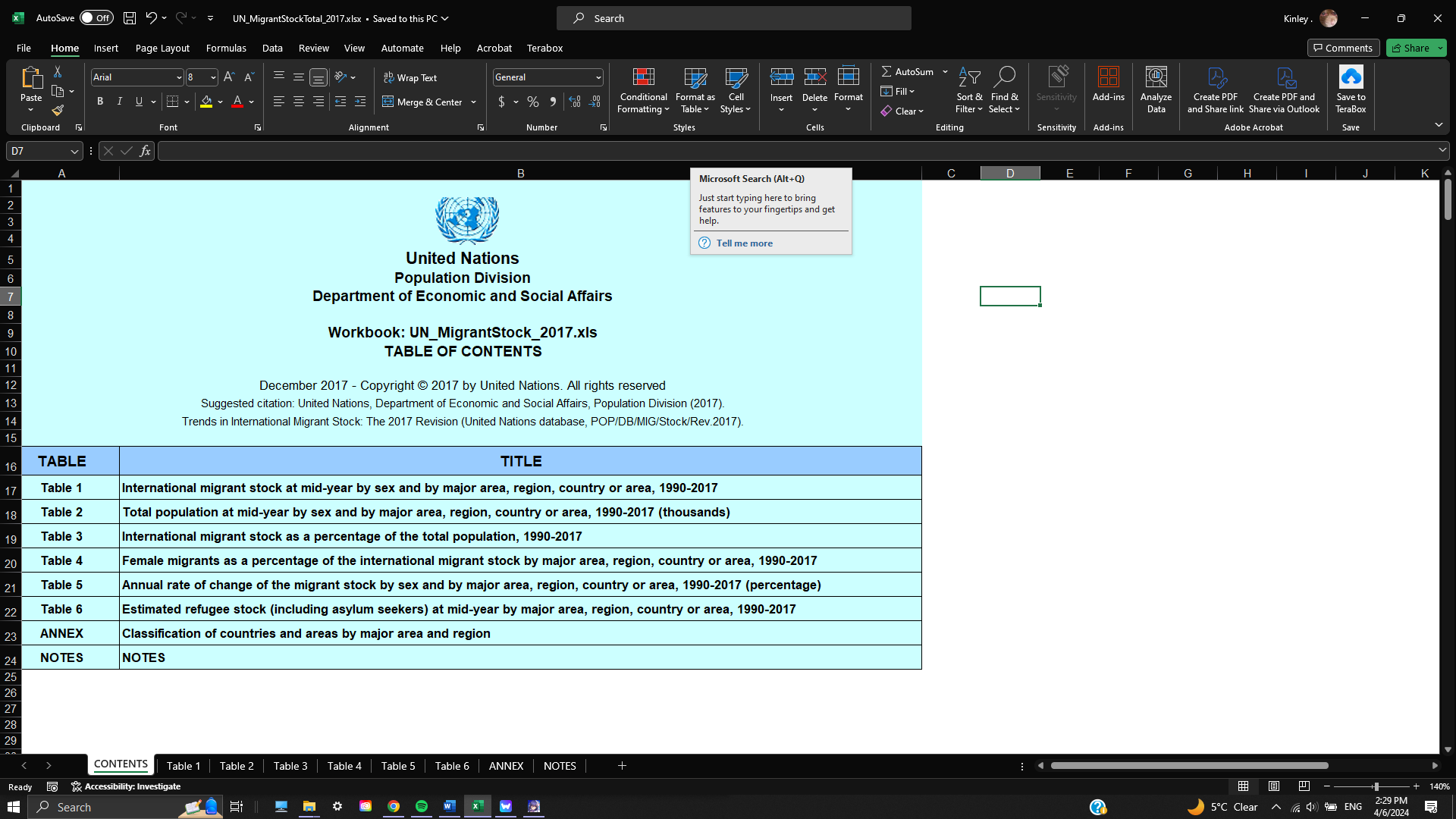


Figure 7: Raw data file downloaded from UN

With the first step of making the dataset usable for our project, the file needs to be converted to a CSV format. By using a web-based application called “cloudconvert” ([www.cloudconvert.com/csv-to-xlsx](http://www.cloudconvert.com/csv-to-xlsx) ), we were able to obtain the required file. Based on the above figure, the file contains multiple tables, each representing an aspect of immigrant and emigrant when being considered for comparison between countries. This was an obstacle for us when trying to process the data. Instead of processing the whole file, we decided to pick out the tables which hold information that were crucial to the aim of the visualization. The table we have chosen is Table 1 ( International migrant stock at mid-year by sex and by major area, region, country, or area, 1990 – 2017). With the process of handling the dataset by hand, we have split the said table into different files, with each name containing the format “Migrate\_XXXX.xlsx” where “XXXX” corelate to each respective year. The original table 1 is as follow:

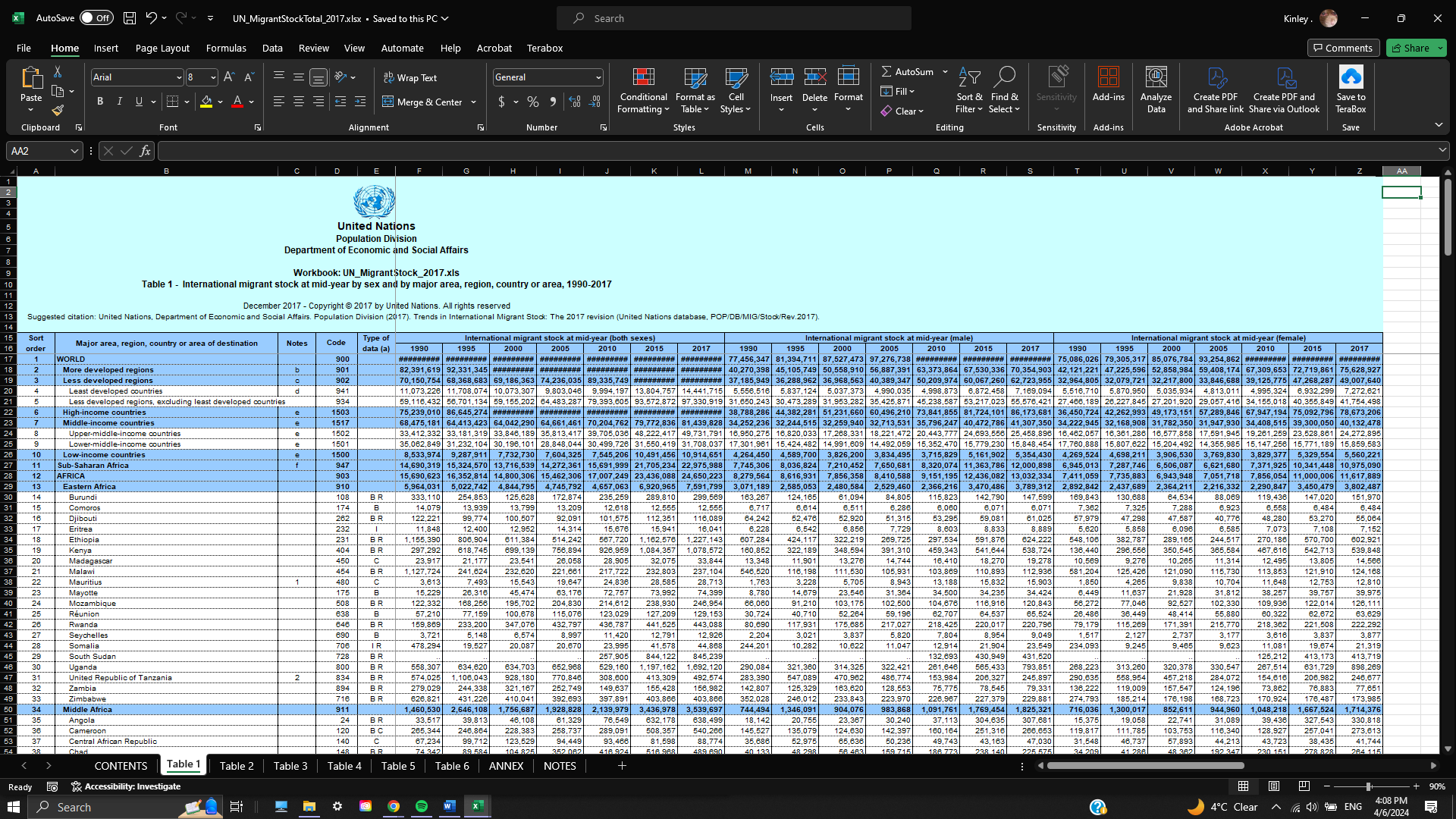


Figure 8: The original Table 1 from the UN dataset

After the process of hand-cleaning data for table 1, we have the following files:



Figure 9: Split the aforementioned table 1 into different files

Each of these files has the same format, with the only difference between each one is the year they represent. For example, the following figure is a representation of the format, more specifically the file “Migrate\_2000.csv”.

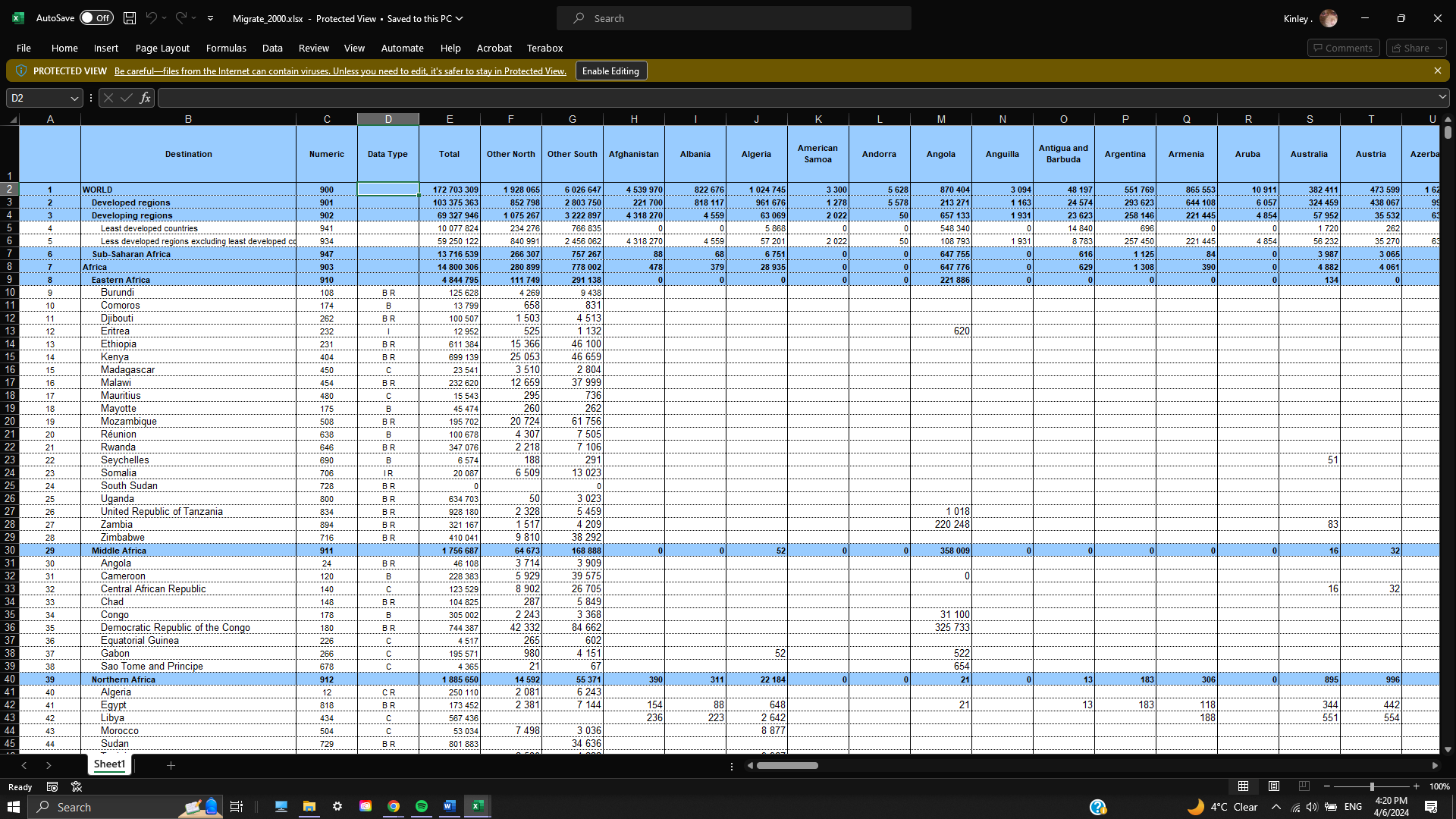


Figure 10: Migrate\_2000.csv example

All the “Migrate\_XXXX.xlsx” was formatted the same for ease of use and cleaner data representation.

### Code Processing:

With each of the previously hand-processed data file, we put them into the same directory called “raw\_data” for ease of navigation and differentiation. All the previous files were not suitable for generating a visualization, so we had to come up with a python-based program to help us process the data better and faster. The code can be explained as the following:

1. Data Processing: The program reads and processes various demographic and economic data from different sources, including data related to countries, population, life expectancy, GDP, and migration.
2. Migration Analysis: It focuses on analyzing migration data, including both immigration and emigration, for different years.
3. Statistical Calculations: The program calculates various statistics related to migration, such as positive and negative stock, net migration, and immigration fraction.
4. Data Output: After processing and analyzing the data, the program outputs the results into structured formats, including JSON and CSV files, which can be used for further analysis, visualization, or integration with other systems.



Figure 11: Python-based program for data cleaning

# Requirements

## Must-have Features:

### Interactive Visualization:

1. Hover Effect:
   1. Implemented the world map to display detailed data and explanations when hovering over specific countries or regions.
   2. Added to the map to highlight migration pathways and showcase detailed data.
2. Drill-Down:
   1. Different description box tabs are shown when clicking on specific countries, providing a How-to-read section and a summary/background section. Additionally, a Detailed data table is displayed.
3. Selectors:
   1. Implemented for the world map, allowing users to select a country and view migration patterns to and from that country.

### Additional Information:

* Providing ample context is crucial for visualization understanding, so we provide additional labeling.

### Clear Data Labeling:

* Ensures users grasp the represented data.

### Color Contrast:

* Different colors represent distinct migration flows, aiding readers' visual clarity.

## Optional Features:

### Tooltip Integration:

Tooltips offer instant context, eliminating the need to memorize information. While this functionality couldn't be implemented, an alternative has been provided to serve a similar purpose.

# Visualization Design:

We structure our visualization design process into four stages, each applicable to both initial development and refinement. These stages are as follows:

1. **Conceptualize**:

* Brainstorming & Sketching: Generate rough ideas based on project goals and data. This includes sketching layouts, exploring visual elements (maps, charts, graphs), and storytelling approaches.
* Research & Inspiration: Analyze existing migration visualizations to learn from best practices.
* Target Audience: Consider user needs (policymakers, researchers, public) for effective communication.
* Technical Feasibility: Evaluate visualization techniques considering data complexity, performance, and resources.

1. **Visualize**:

* Refine Sketches: Polish initial sketches based on feedback, using design software or prototyping tools.
* Prototype Layouts: Develop layouts balancing visual appeal and functionality. Experiment with data element arrangement, scales, and annotations.
* Color Palette & Style: Explore color schemes, typography, and graphic styles for a cohesive and engaging presentation. Apply color theory for readability and meaning.

1. **Materialize:**

* Finalize Design: Make final decisions on styling, colors, and additional visual elements. Create a style guide for consistency.
* Iterate on Feedback: Refine based on usability testing and user research, ensuring accessibility for diverse users.
* Interactive Elements: Consider tooltips, filters, and animations to enhance engagement and exploration. Design intuitive navigation to guide users.

1. **Implementation**:

* Coding the Design: Translate final design concepts into code using appropriate libraries (d3.js).
* Data Integration: Integrate cleaned and processed data, ensuring accuracy and relevance. Use data-driven visual encoding techniques (size, color, position).
* Testing & Debugging: Conduct thorough testing to identify and resolve technical issues or compatibility concerns. Test across devices, screen sizes, and browsers.
* Iterative Changes: Propose changes based on usability testing to optimize engagement and comprehension. Document the implementation process for future updates and maintenance.

## Conceptualize (Design Iteration) stage:

We want to offer a straightforward, interactive layout in the first early stages of the visualization design process that will work for the public, who is our target audience. We adopt distinct approaches to various visualizations while working with classified sorted data, all the while maintaining consistency to guarantee a user-friendly experience.

Without thoroughly reviewing the data we have discovered, we created a rough sketch for the first visualization, Migration Pattern, to convey our ideas to the reader. We concentrated on "World Map" style visualizations and thought carefully about our options because our goal is to visualize the flow of immigrants and emigrants from different countries to another.

With the figure below is our attempt to implement a “World Map” visualization, which in our opinion, was the go-to way do help users grasp the concept of Immigrant and Emigrant flows relative to countries all over the world.

A map of the world

Description automatically generated

Figure 12: An Attempt at the world map sketch

Throughout our research, we have found other visualizations alternatives, but mostly aren’t suitable for the scope our project is going for. Bar chart is a great implementation of a visualization to describe the flow of immigrations and emigrations over time, but sadly that’s only applicable for a single country for each country, since implementing for all countries over the world can make a bar chart seems inadequate when user wanting to compare between countries. Instead, we have chosen the Bar Chart as a supporting visualization for the main one, i.e. the World Map above. Another type of supporting visualization we decided to implement was a Pie Chart, for which the goal was to visualize the men and women percentage in immigrating or emigrating between countries.

## Visualize Stage:

We now produce a set of higher fidelity designs after looking into our design decisions, which aids in contextualizing the idea behind these visualizations. We would like to see a justified, centered look for the Migration Pattern visualization, with different sections representing different rates of immigration and emigration.

With the ideas in mind, a mockup for the website was made by me in an excel sheet to gather feedback from the lecturer, Dr Hoang Xuan Tung. We made suggestions on how to implement the GDP per Capita data into the visualization to help user better understand the correlation between the immigration/emigration rate and how the GDP of a country can affect those numbers. We have chosen a side-to-side Pie Chart depicted in the figure below, to help user compares between developing and developed countries. It would contain the percentage of men and women when immigrating or emigrating, and the user can specify which country to visualize.

A screenshot of a computer

Description automatically generated

Figure 13: Mockup design of the website

After receiving feedback from Dr. Hoang Xuan Tung, we have decided to implement a color system within the World Map to display the GDP per capita, making the two Pie charts below the main visualization redundant. This will be further elaborated in the next section of this report.

## Materialize Stage:

* Finalize design decisions regarding styling, coloring, and additional visual elements. Create a style guide or design system to maintain consistency across different parts of the visualization.
* Iterate on the visualization design based on feedback from usability testing and user research. Ensure that the visualization is accessible to users with diverse needs, including those with visual impairments or limited technical expertise.
* Consider the use of interactive elements such as tooltips, filters, and animations to enhance user engagement and exploration. Design intuitive navigation pathways that guide users through the visualization and encourage active interaction.

## Implementation Stage:

* Translate finalized design concepts into code using appropriate visualization libraries or frameworks such as d3.js, Vega, or Tableau.
* Integrate cleaned and processed data into the visualization, ensuring accuracy and relevance. Implement data-driven visual encoding techniques to map data attributes to visual properties such as size, color, and position.
* Conduct thorough testing and debugging to identify and resolve any technical issues or compatibility concerns. Test the visualization across different devices, screen sizes, and web browsers to ensure a consistent experience for all users.
* Propose iterative changes based on usability testing and feedback to optimize the visualization for user engagement and comprehension. Document the implementation process and any modifications made to the original design to facilitate future updates and maintenance.

# Final Design

# Conclusion

# References