Data Visualization COS30045

OECD Health Statistic Project Process Book

Topic: Economic effect on life expectancy

Team Member: Davy Sung, 103535521 - Toan Nguyen, 104180605

Tutorial Class: Monday 4.30pm Semester 1 2024

Tutor: Raihaneh Aghaie

Group: CL04_T03

Word count: 8181

Website: https://toannguyenswinadventure.github.io/COS30045-DataVisualisation/HomePage/index.html

MercuryLink: https://mercury.swin.edu.au/cos30045/s103535521/Assignment2/HomePage/index.html

Table of Content

I. Introduction....4
1. Background and Motivation....4
Background....4
Motivation....4
2. Project Purpose....5
II. Data....5
1. Data Source....5
2. Data Collecting....6
3. Data Type....7
Choropleth Chart....7
Bubble Chart....7
Density Chart....8
4. Data processing....8
Life expectancy data processing....9
GDP data processing....12
GDP per Capita data processing....15
GDP and GDP per capita Merging data processing....16
Continent data processing....19
Population data processing....22
III. Visualisation Design....26
Design Guideline....26
Color Theory....27
1. Prototype and Sketch....28
2. Implementation....35
a. Choropleth....35
Key Features....35
Configuration Settings:....36
Steps for Implementation:....36
b. Bubble Chart....37
Key Features:....38
Configuration Settings:....38
Steps for Implementation:....38
c. Density Chart....40
Key Features:....41
Configuration Settings:....41
Steps for Implementation:....41

IV. Validation....42
Usability Testing Task.... 43
V. Conclusion....44
References....46
APPENDIX A: Usability Evaluation test materials....47
APPENDIX B: Data Collected Through Evaluation Testing....56
1. Observation Data....56
2. Informed Consent Data Collection....62
Agreement to Participate.... 62
Acknowledgement....63
3. Demographic Data Collection....64
4. Post-Questionnaire Data Collection....66
APPENDIX C: Low-fidelity Design Validation....75

I. Introduction

1. Background and Motivation

Background

Economic growth profoundly influences various facets of human life, including education, healthcare, and overall quality of life. As countries experience economic development, they often see improvements in their educational systems, healthcare infrastructure, and general living conditions. These enhancements contribute significantly to the life expectancy of their populations. Understanding the relationship between economic growth and life expectancy is essential for policymakers, researchers, and public health officials aiming to foster equitable and sustainable development.

Motivation

The modern era of promising better decision making through access to real world data, with hundreds or millions of data surrounding us. Therefore, according to Cowan (2005), Human's brain can be full and might not have enough capacity to load all of the data to generate relationships between them. This is the motivation for creating visualizations that help in identifying and comprehending complex relationships and its correlations by clearly depicting the correlation between economic growth (measured by GDP and GDP per capita) and life expectancy. This allows audiences to effectively identify key trends and patterns in the data, such as how different levels of economic growth impact life expectancy in various regions or countries. In addition, interactive visualizations allow users to explore data dynamically, providing an understanding of the relationships between variables by identifying specific factors and scenarios that contribute to variations in life expectancy.

2. Project Purpose

This project seeks to assist users in understanding the relationship between the economic metric GDP and GDP per capita with life expectancy in OECD countries. Understanding how health care expenditures influence life expectancy is crucial. Through visualization, we aim to delve deeper into this relationship by examining life expectancy, GDP, and healthcare resources. Users can explore the following questions with this visualization:

1. *Are there regions that have shown significant improvements or declines in life expectancy in recent years?*
2. *How has life expectancy changed over the years?*
3. *Is there a correlation between a country's GDP per capita and its life expectancy?*
4. *Do individuals with higher incomes tend to have longer life expectancies?*
5. *How does life expectancy density change over the year?*

The benefit of this visualization for OECD countries is that healthcare leaders can analyze the data to identify areas for improvement in their healthcare systems. Governments can use this data to compare life expectancies and invest in their healthcare systems to enhance life expectancy for their citizens.

II. Data

1. Data Source

Data sources are taken from the Organisation for Economic Co-operation and Development (OECD). OECD, founded in 1961, consists of 38 member countries and serves as a vital platform for international cooperation on economic and social policies. Renowned for its integrity and rigorous analysis, the OECD gathers and analyzes data across a range of areas, including GDP, employment, education, health, and the environment. According to the former Finnish Prime Minister, Sanna Marin (2020) states

that the OECD stands as an invaluable platform for exchanging information, facilitating peer learning, and providing evidence-based analysis for nearly six decades. Through Sanna Marin's statement, the OECD provides valuable insights into global trends and fosters collaboration among member countries to address common challenges and promote sustainable development and prosperity worldwide.

In addition to the OECD, data sources are also taken from Our World in Data (OWID). Our World in Data is an open-access, online publication that focuses on large global problems, providing reliable data and research to address critical issues facing the world today. Founded by economist Max Roser at the University of Oxford, OWID offers comprehensive and up-to-date datasets covering a wide range of topics, including health, poverty, education, and environmental sustainability. The platform is renowned for its commitment to transparency, providing detailed explanations of data sources and methodologies, and enabling users to explore the data through interactive visualizations. By aggregating data from trusted institutions and offering in-depth analyses, OWID empowers policymakers, researchers, and the public with the information needed to make informed decisions and drive progress towards global development goals.

2. Data Collecting

In this data visualization project, the focus revolves around the economic factors that impact a country's life expectancy. Thus, the required data sources include the life expectancy figures for each country. Additionally, to examine the influence of economics on life expectancy, access to each country's GDP is also needed. Furthermore, analyzing the GDP per capita of each country is crucial to understanding its correlation with health expenditure per capita, thereby assessing the level of public interest and investment in healthcare within the population.

As demanded, GDP and GDP per capita data are sourced from the Organisation for Economic Co-operation and Development (OECD). On the other hand, population,

continent and life expectancy at birth and population data is sourced from Our World in Data (OWID).

3. Data Type

For the visualization project, we will outlined three different types of charts (section 3): a choropleth, a bubble chart, and a density chart with the data requirements and the types of attributes needed for each chart:

Choropleth Chart

- **Geometry Data**: This will be the spatial data representing the geographic boundaries of countries. This data type is **categorical** since it involves distinct regions or locations.
- **Life Expectancy**: This will be the quantitative data used to color the map. This data type is **numerical (continuous)** because life expectancy is a continuous variable measured in years.

Bubble Chart

- **Life Expectancy**: This attribute is **numerical (continuous)**.
- **GDP**: This attribute is also **numerical (continuous)** because it represents the gross domestic product, a continuous measure of economic output.
- **GDP per Capita**: This attribute is **numerical (continuous)** as it represents the GDP divided by the population, another continuous measure.
- **Population**: This attribute is **numerical (continuous)**, representing the total population of each country.
- **Continent**: This attribute is **categorical**, representing the continent to which each country belongs.

Density Chart

- **Life Expectancy**: The density chart will convey the distribution of life expectancy across all countries. The attribute type for life expectancy in this context remains **numerical (continuous)**, as it involves plotting the distribution of a continuous variable.

4. Data processing

Processing the dataset is essential to ensure its usability and accuracy for analysis. Knime, an open-source platform for data analytics, is utilized for its user-friendly drag-and-drop interface, making it accessible to both technical and non-technical users. Knime's support for various data processing tasks through modular data pipelining enhances efficiency and effectiveness. The data, in CSV format, is chosen for its simplicity, compatibility, and ease of use, facilitating seamless integration with Knime. This compatibility streamlines data workflows, enhances performance, and promotes collaboration, ensuring that the data is ready for accurate and insightful visualization. Here are the techniques with glossary we used to handle the data processing:

Glossary:

Knime Analytics Platform contains various nodes with its purpose. In this project, the processing part uses the following Knime nodes in order to process the data:

1) ***CSV Reader****: This node used to let Knime read the CSV file.*
2) ***CSV Writer****: After processing data, the new processed data is exported out as a new processed csv file.*
3) ***Column Filter****: This node is used to filter out unnecessary attributes (columns) or not related to the purpose of visualisation.*
4) ***Row Filter*** *&* ***Rule - based Row Filter****: This node is used to filter the row/rows based on regex.*

5) ***Math Formula****: This node allows the processing to make calculations using math.*
6) ***String Manipulation****: This node is to change the form of a string to another.*
7) ***Column Rename****: This node is to rename the attribute to avoid conflict or duplication when merging dataset or make the name more clear.*
8) ***Joiner****: This node is to merge two dataset with the same constraint. (Same purpose as Join query of SQL)*

In the project, GDP and other economic indicators that used as **PPP** currency exchange method:

PPP: This method is the purchasing power parity (**PPP**) exchange rate. According to the former Assistant Director in the IMF's Communication Department, Callen (2023), this method ensures methodological balance by equalizing the purchasing power gap among different currencies, eliminating price level differences across countries. By focusing solely on GDP measured by PPPs, we adhere to established standards for economic analysis, ensuring accuracy and reliability in our data-driven insights.

Life expectancy data processing

Processing Life expectancy data involves several steps to ensure the data is clean, consistent, and relevant:

1. **CSV Reader**: Start by reading the life expectancy CSV file using the "CSV Reader" node to load the data into Knime.
2. **Column Rename**: Use the "Column Rename" node to rename the life expectancy value attribute from "Value" to "Life_expectancy" for clarity. *(Figure 2)*
3. **Math Formula**: Apply the "Math Formula" node to round the life expectancy values to two decimal places, keeping the values concise. *(Figure 3)*
4. **CSV Writer**: Finally, use the "CSV Writer" node to construct a new CSV file with the processed data, ready for use in subsequent analysis and visualization.

By following these steps, the life expectancy data is clean, well-structured, and formatted for effective analysis and visualization.

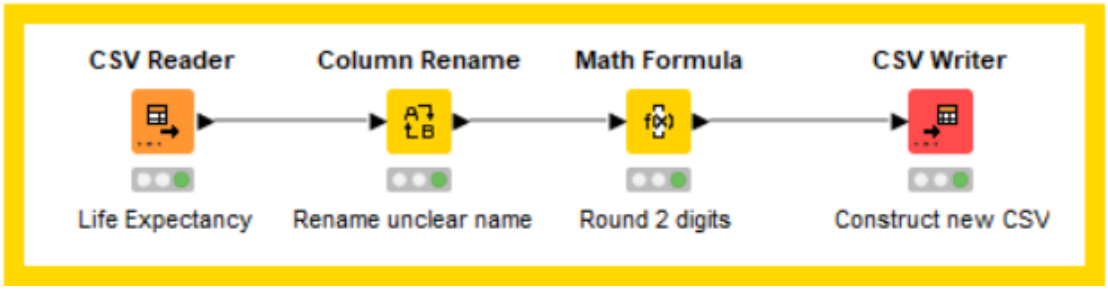


Figure 1: Life Expectancy Workflow

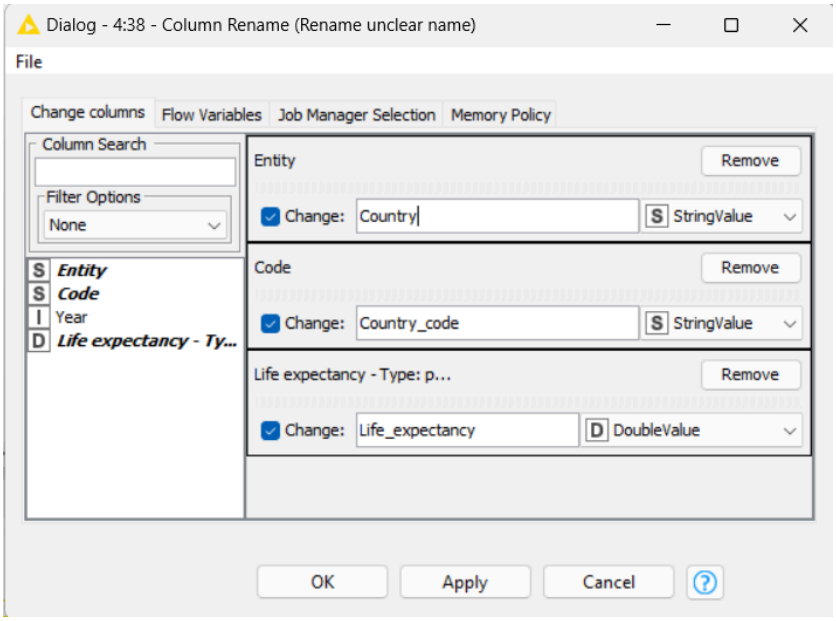


Figure 2: Column Rename

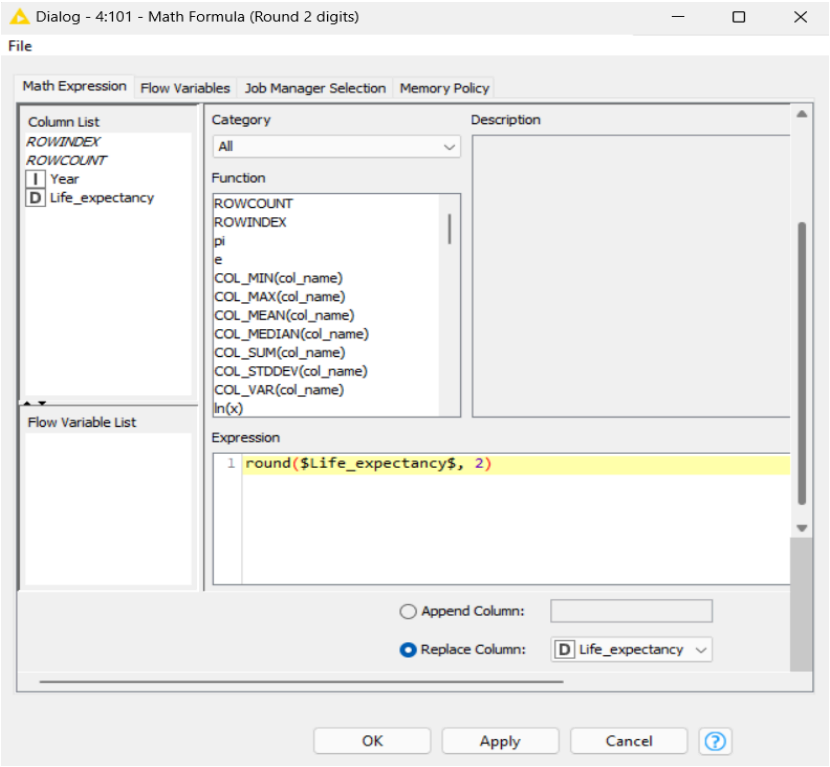


Figure 3: Value rounding

Row ID	S Country	S Countr...	I Year	D Life_expectancy
Row1023	Australia	AUS	2008	81.46
Row1024	Australia	AUS	2009	81.77
Row1025	Australia	AUS	2010	82.06
Row1026	Australia	AUS	2011	82.13
Row1027	Australia	AUS	2012	82.35
Row1028	Australia	AUS	2013	82.6
Row1029	Australia	AUS	2014	82.57
Row1030	Australia	AUS	2015	82.66
Row1031	Australia	AUS	2016	82.87
Row1032	Australia	AUS	2017	83
Row1033	Australia	AUS	2018	83.39
Row1034	Australia	AUS	2019	83.11
Row1035	Australia	AUS	2020	84.32
Row1036	Australia	AUS	2021	84.53
Row1037	Austria	AUT	1947	61.54
Row1038	Austria	AUT	1948	63.21
Row1039	Austria	AUT	1949	63.47
Row1040	Austria	AUT	1950	64.81
Row1041	Austria	AUT	1951	65.22
Row1042	Austria	AUT	1952	66.75
Row1043	Austria	AUT	1953	67.25
Row1044	Austria	AUT	1954	67.29
Row1045	Austria	AUT	1955	67.55
Row1046	Austria	AUT	1956	67.68

Figure 4: Data Final

GDP data processing

Processing GDP data involves several steps to ensure the data is clean, consistent, and relevant:

1. **CSV Reader**: Read the GDP CSV file using the "CSV Reader" node to load the data into Knime.
2. **Row Filter**: Use the "Row Filter" node to keep only the rows with GDP measured by purchasing power parity (PPP).
3. **Column Filter**: Use the "Column Filter" node to retain only the necessary attributes: country code (COU), country name (Country), year (Year), and GDP value (Value). This step removes any irrelevant attributes from the dataset.
4. **Column Rename**: Use the "Column Rename" node to update unclear column names to more descriptive labels. For instance, rename "COU" to "Country_code".
5. **CSV Writer**: Save the cleaned and processed GDP data into a new CSV file for use in subsequent analysis and visualization.

By following these steps, both the life expectancy and GDP data are prepared and formatted for effective analysis and visualization, ensuring accuracy and relevance in the project's outcomes.

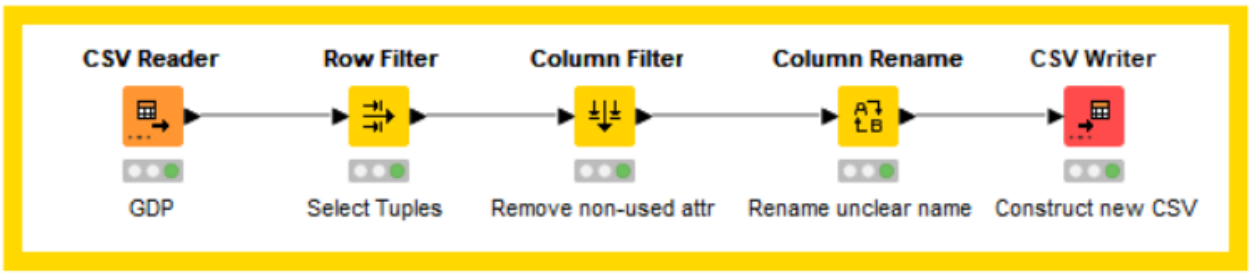


Figure 5: GDP Processing workflow

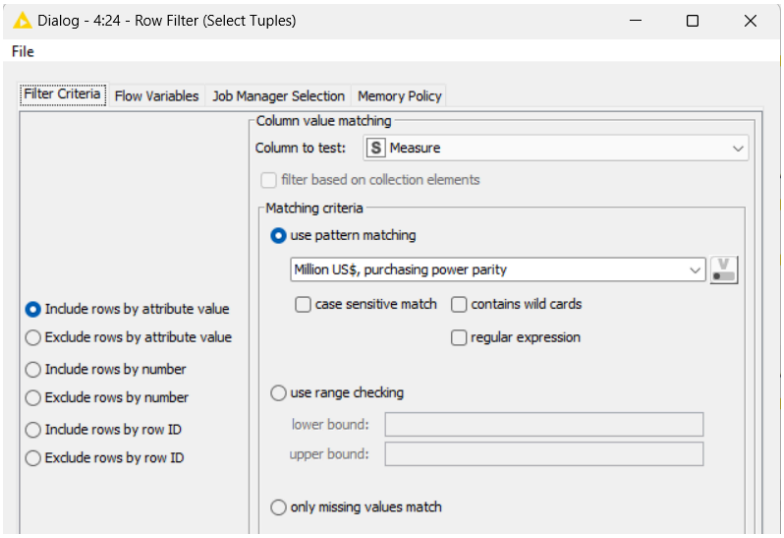


Figure 6: Kept tuples with PPP method of measurement

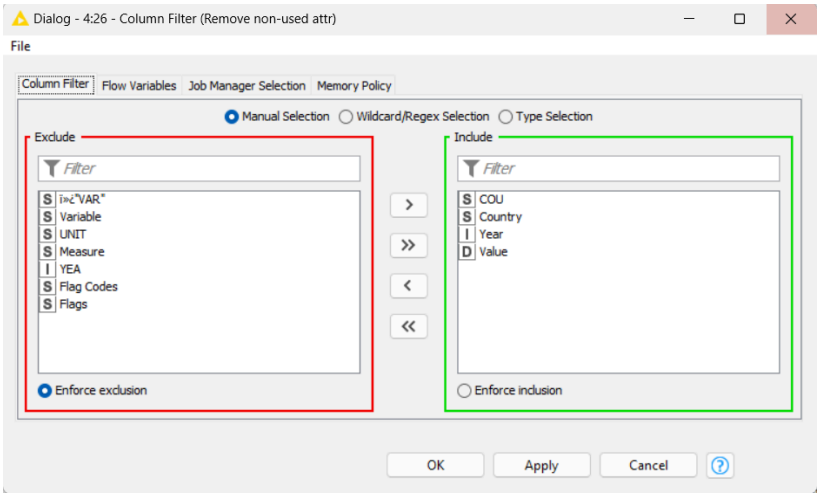


Figure 7: Keep necessary attributes.

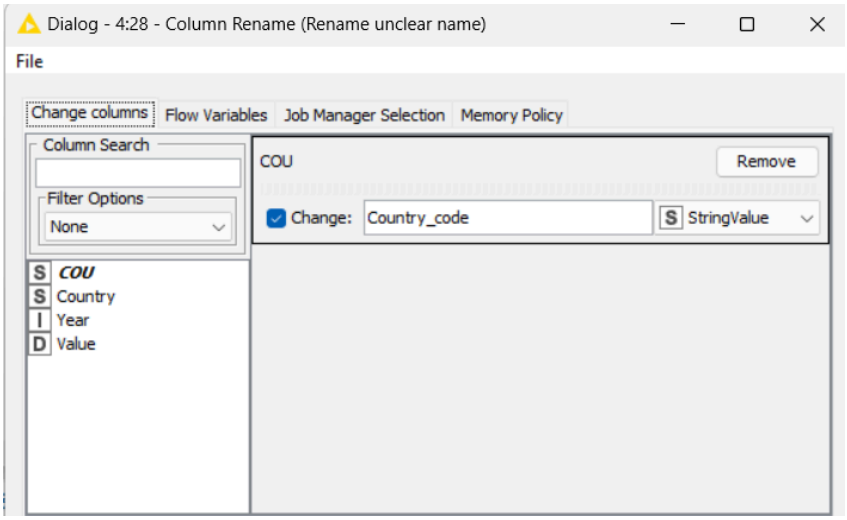


Figure 8: Rename the unclear attributes' name

Row ID	S Country_code	S Country	I Year	D Value
Row3544	AUS	Australia	2010	943,322.9
Row3545	AUS	Australia	2011	992,802.4
Row3546	AUS	Australia	2012	997,651.5
Row3547	AUS	Australia	2013	1,104,672.5
Row3548	AUS	Australia	2014	1,117,602.3
Row3549	AUS	Australia	2015	1,124,752.7
Row3550	AUS	Australia	2016	1,212,854.3
Row3551	AUS	Australia	2017	1,247,085.6
Row3552	AUS	Australia	2018	1,323,863.6
Row3553	AUS	Australia	2019	1,337,588.9
Row3554	AUS	Australia	2020	1,430,866
Row3555	AUS	Australia	2021	1,595,113.4
Row3556	AUS	Australia	2022	1,785,955.8
Row3557	AUT	Austria	2010	351,324
Row3558	AUT	Austria	2011	373,031.5
Row3559	AUT	Austria	2012	391,635
Row3560	AUT	Austria	2013	406,370.5
Row3561	AUT	Austria	2014	417,059.7
Row3562	AUT	Austria	2015	430,976
Row3563	AUT	Austria	2016	460,282.8
Row3564	AUT	Austria	2017	476,590.4
Row3565	AUT	Austria	2018	503,361.1
Row3566	AUT	Austria	2019	530,168.7
Row3567	AUT	Austria	2020	510,568

Figure 9: Data Final

GDP per Capita data processing

Processing GDP per Capita data involves several steps that mostly the same with GDP processing, with a slightly different changed in configuration of Row Filter node:

1. **Row Filter**: Use the "Row Filter" node to keep only the rows with GDP per Capita measured by purchasing power parity (**PPP**).

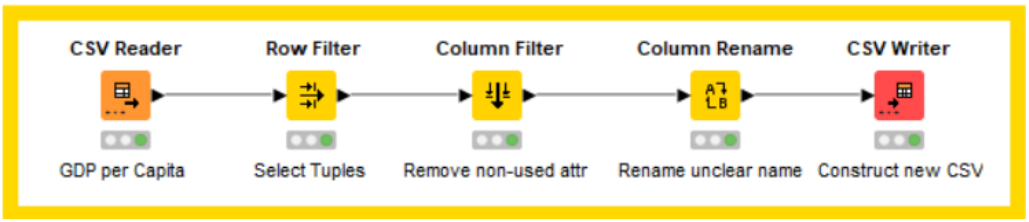


Figure 10: GDP per Capita workflow

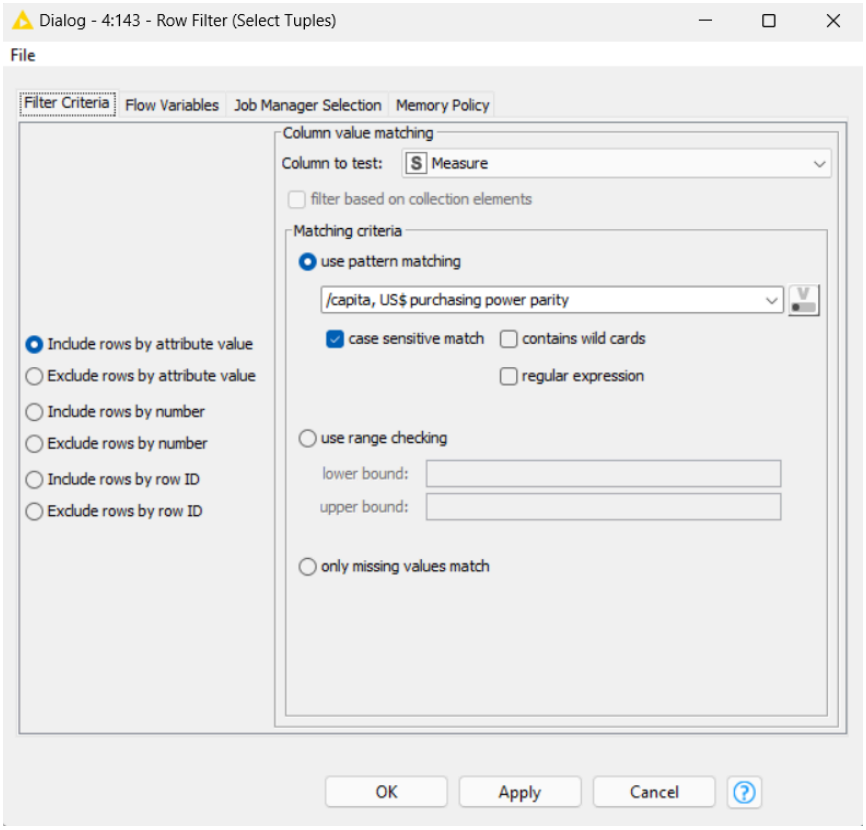


Figure 11: Row Filter with PPP and /capita

GDP and GDP per capita Merging data processing

The process of data for GDP and GDP per capita is shown as above, in this step, to merge the both of the datasets with the following steps:

1. **Math Formula**: The GDP value in US dollars is quite large and might represent a long text for the user. So in this step, the GDP value is converted to millions of US dollars and rounded to two decimal places for better readability.
2. **Joiner**: After processing both datasets, use the "Joiner" node to combine them based on country code (Country_code) and year (Year). This ensures that the merged dataset contains all relevant information for each country and year.

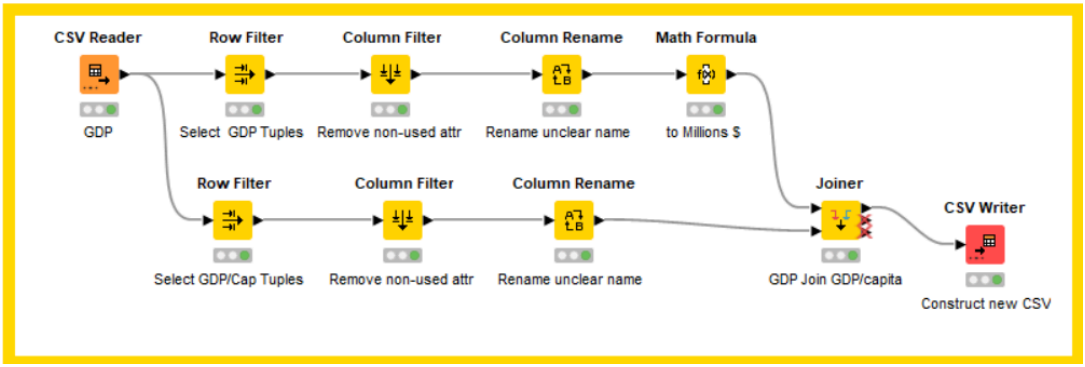


Figure 12: Merging Workflow

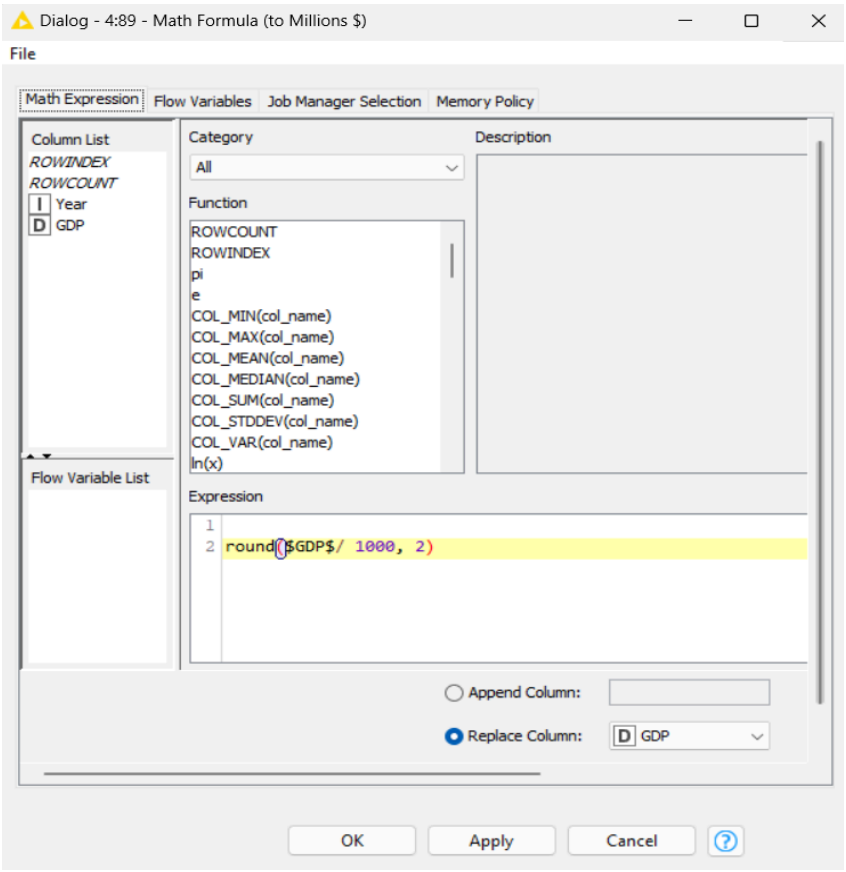


Figure 12: Convert GDP to GDP in Millions $US

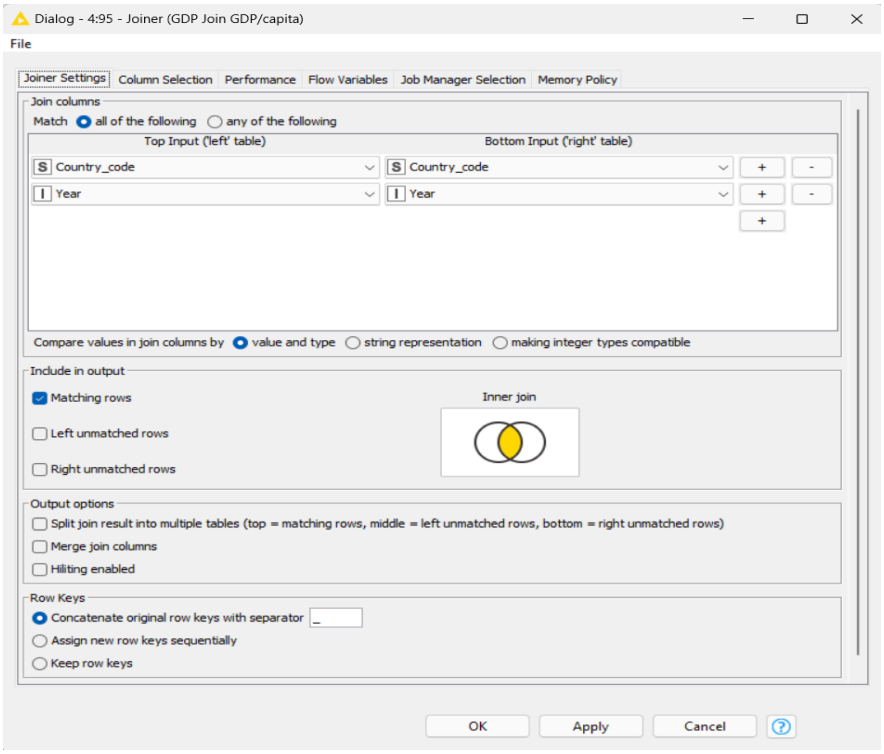


Figure 13.1: Joining GDP per capita $US with GDP in Millions $US

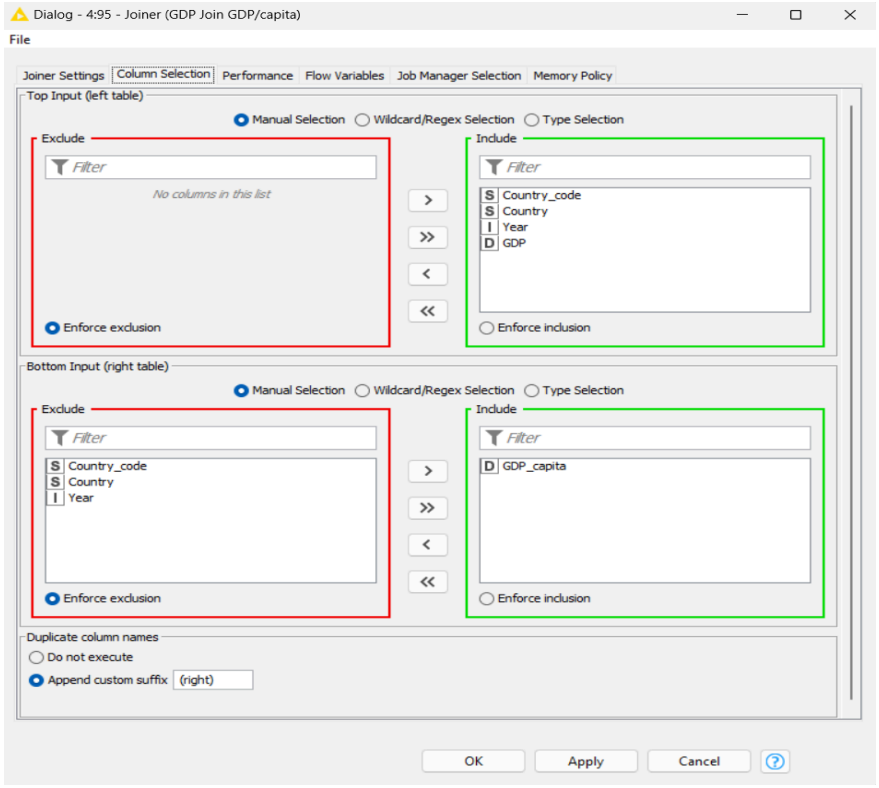


Figure 13.2: Joining GDP per capita $US with GDP in Millions $US

Row ID	Countr...	Country	Year	GDP	GDP_capita
Row3544_Ro...	AUS	Australia	2010	943.32	42,816.5
Row3545_Ro...	AUS	Australia	2011	992.8	44,440.5
Row3546_Ro...	AUS	Australia	2012	997.65	43,884.7
Row3547_Ro...	AUS	Australia	2013	1,104.67	47,763.2
Row3548_Ro...	AUS	Australia	2014	1,117.6	47,606.8
Row3549_Ro...	AUS	Australia	2015	1,124.75	47,226.8
Row3550_Ro...	AUS	Australia	2016	1,212.85	50,136.8
Row3551_Ro...	AUS	Australia	2017	1,247.09	50,706.5
Row3552_Ro...	AUS	Australia	2018	1,323.86	53,025.3
Row3553_Ro...	AUS	Australia	2019	1,337.59	52,785.2
Row3554_Ro...	AUS	Australia	2020	1,430.87	55,772.7
Row3555_Ro...	AUS	Australia	2021	1,595.11	62,095.5
Row3556_Ro...	AUS	Australia	2022	1,785.96	68,746.3
Row3557_Ro...	AUT	Austria	2010	351.32	42,007.3
Row3558_Ro...	AUT	Austria	2011	373.03	44,452.7
Row3559_Ro...	AUT	Austria	2012	391.64	46,457.3
Row3560_Ro...	AUT	Austria	2013	406.37	47,922
Row3561_Ro...	AUT	Austria	2014	417.06	48,799.7
Row3562_Ro...	AUT	Austria	2015	430.98	49,865.9
Row3563_Ro...	AUT	Austria	2016	460.28	52,684
Row3564_Ro...	AUT	Austria	2017	476.59	54,173
Row3565_Ro...	AUT	Austria	2018	503.36	56,937.9
Row3566_Ro...	AUT	Austria	2019	530.17	59,704.2
Row3567_Ro...	AUT	Austria	2020	510.57	57,258.7
Row3568_Ro...	AUT	Austria	2021	537.01	59,962.8
Row3569_Ro...	AUT	Austria	2022	622.67	69,347.4
Row3570_Ro...	BEL	Belgium	2010	434.06	39,838

Figure 14: Data Final

Continent data processing

The original data contains population data for various range of age, however in our used of this data, we only need the total population of the country in a specific year

1. **CSV Reader**: Read the GDP CSV file using the "CSV Reader" node to load the data into Knime.
2. **String manipulation (x2)**: In this dataset, the Americas continent is separated as North America and South America, and since they are both belong to Americas so this part used “String Manipulation” node to convert both to Americas

3. **Column Rename**: Use the "Column Rename" node to update unclear column names to more descriptive labels. For instance, rename "Code" to "Country_code".
4. **CSV Writer**: Save the cleaned and processed Continent data into a new CSV file for use in subsequent analysis and visualization.

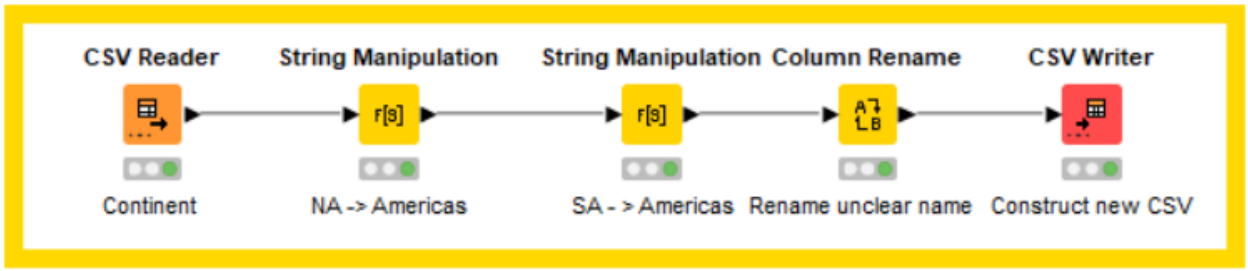


Figure 15: Continent Workflow

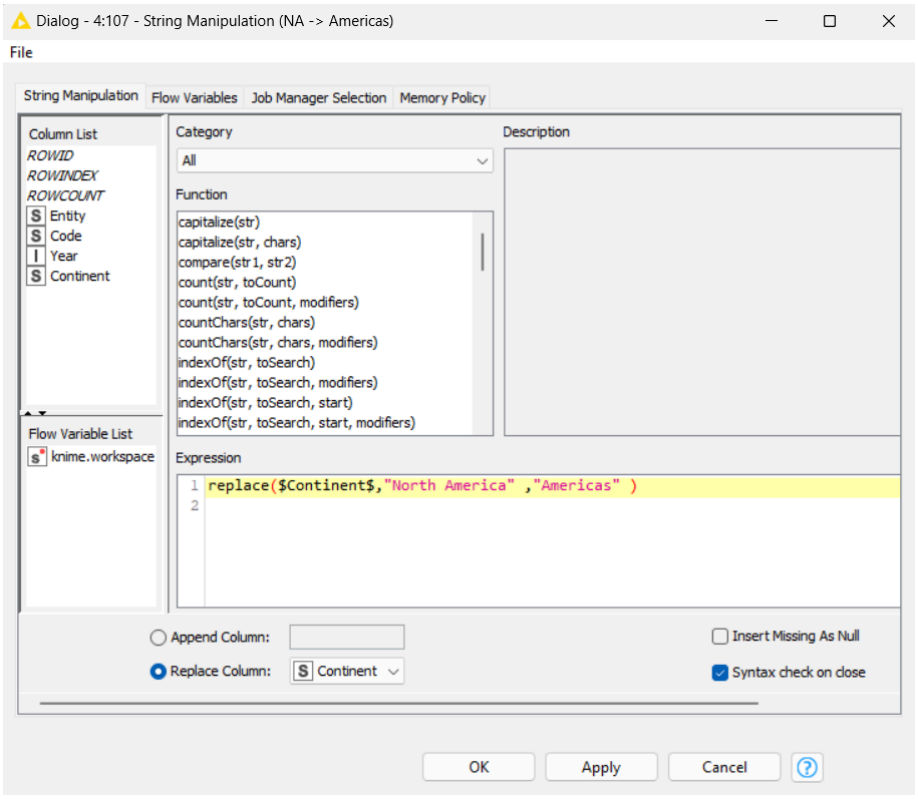


Figure 16: String Manipulation NA -> Americas

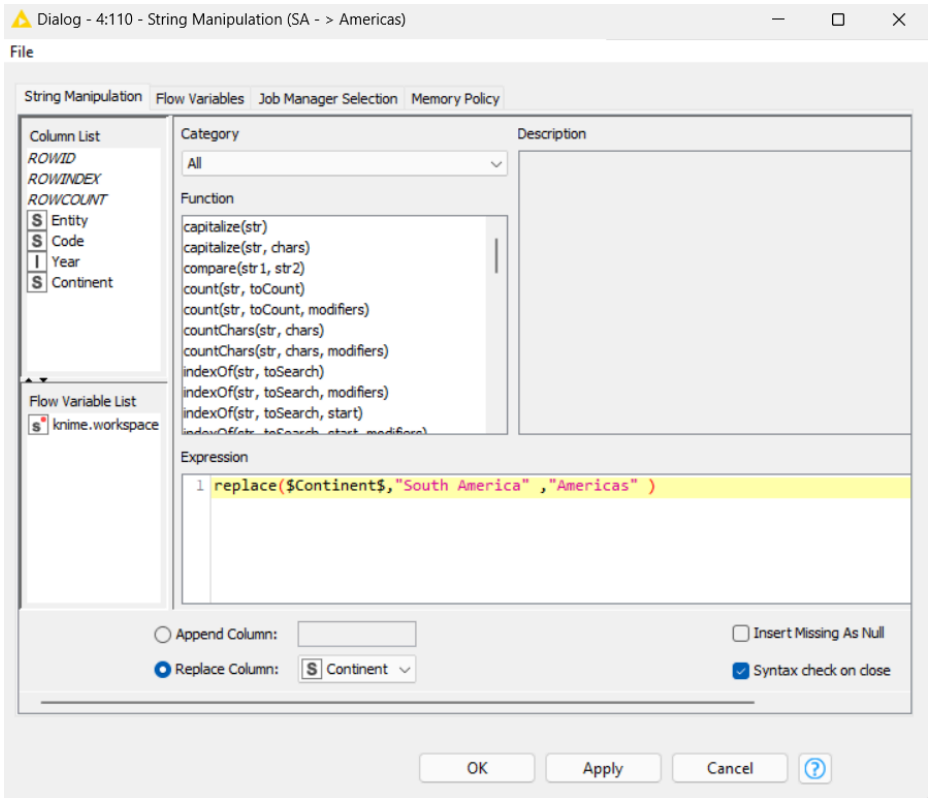


Figure 17: String Manipulation SA -> Americas

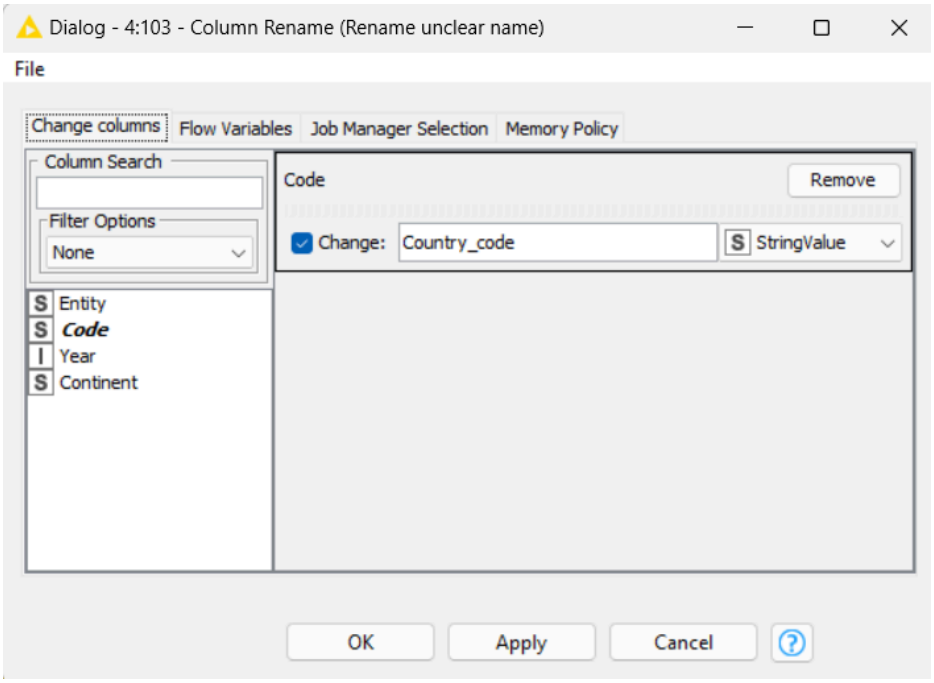


Figure 18: Changing Code to Country_code

Row ID	S Entity	S Countr...	I Year	S Continent
Row0	Abkhazia	OWID_ABK	2015	Asia
Row1	Afghanistan	AFG	2015	Asia
Row2	Akrotiri and ...	OWID_AKD	2015	Asia
Row3	Aland Islands	ALA	2015	Europe
Row4	Albania	ALB	2015	Europe
Row5	Algeria	DZA	2015	Africa
Row6	American Sa...	ASM	2015	Oceania
Row7	Andorra	AND	2015	Europe
Row8	Angola	AGO	2015	Africa
Row9	Anguilla	AIA	2015	Americas
Row10	Antarctica	ATA	2015	Antarctica
Row11	Antigua and...	ATG	2015	Americas
Row12	Argentina	ARG	2015	Americas
Row13	Armenia	ARM	2015	Asia
Row14	Aruba	ABW	2015	Americas
Row15	Australia	AUS	2015	Oceania
Row16	Austria	AUT	2015	Europe
Row17	Austria-Hun...	OWID_AUH	2015	Europe
Row18	Azerbaijan	AZE	2015	Asia
Row19	Baden	OWID_BAD	2015	Europe
Row20	Bahamas	BHS	2015	Americas
Row21	Bahrain	BHR	2015	Asia

Figure 19: Data Final

Population data processing

The original data contains population data for various range of age, however in our used of this data, we only need the total population of the country in a specific year

1. **CSV Reader**: Read the GDP CSV file using the "CSV Reader" node to load the data into Knime.
2. **Column Filter**: Use the "Column Filter" node to retain only the necessary attributes: country name (Country), year (Year), and total population (Population). This step removes any irrelevant attributes from the dataset.
3. **Joiner**: Since this dataset is not having the “Country Code”, but it has the country name. So we have to add the Country code for the dataset by doing an inner join with another dataset.

4. **Column Rename**: Use the "Column Rename" node to update unclear column names to more descriptive labels. For instance, rename "Code" to "Country_code".
5. **CSV Writer**: Save the cleaned and processed Population data into a new CSV file for use in subsequent analysis and visualization.

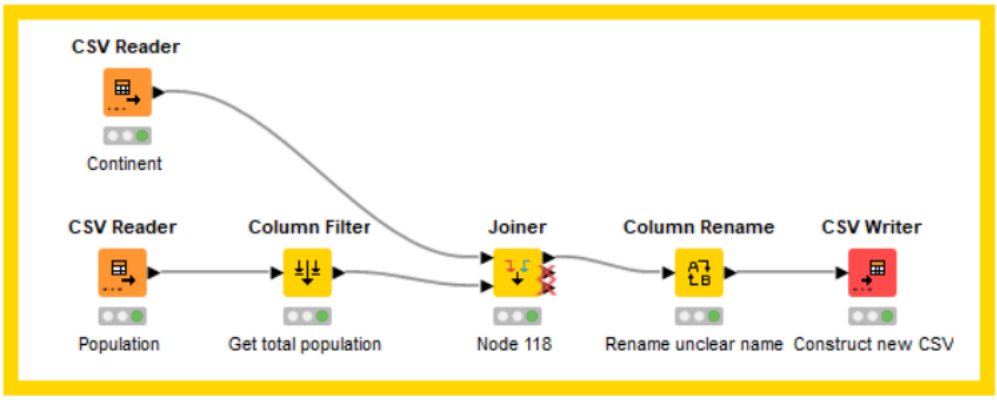


Figure 20: Population workflow

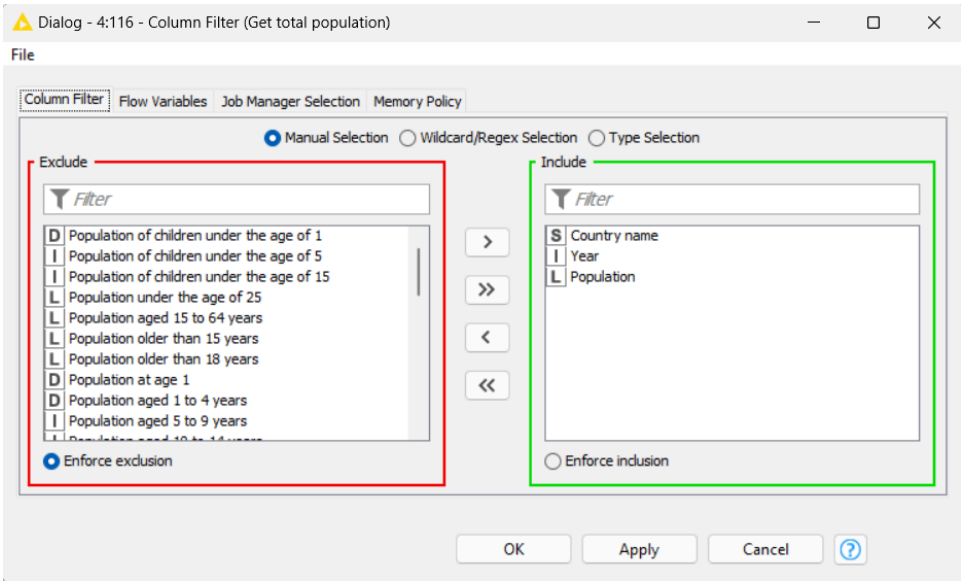


Figure 21: Filter total population attribute

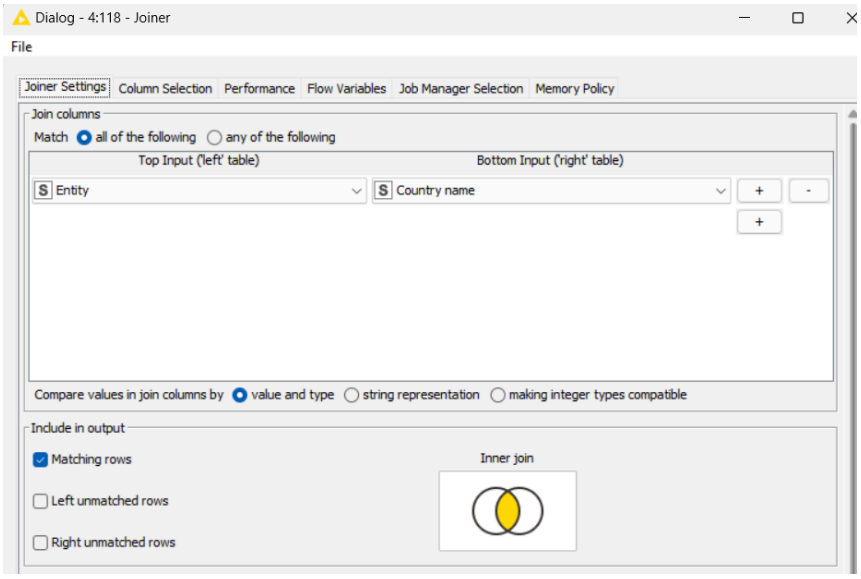


Figure 22.1: Join constraint as Country name

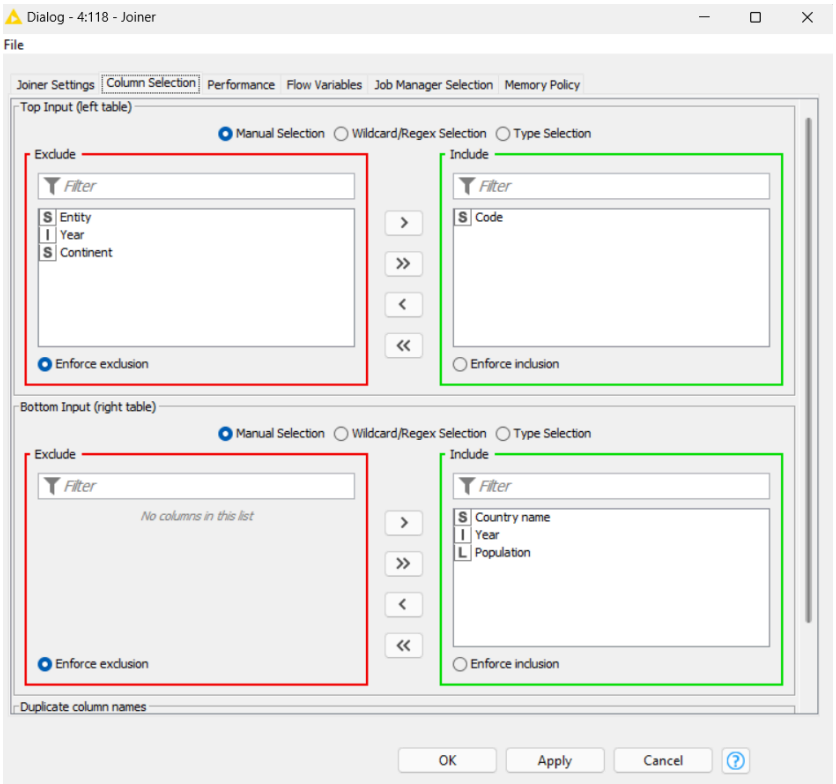


Figure 22.2: Join Country code to population dataset

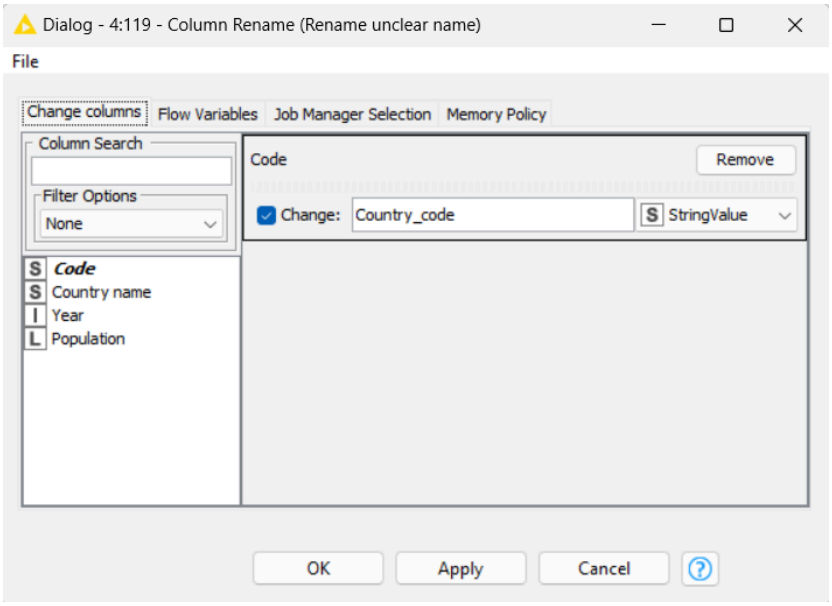


Figure 23: Rename country code attribute

Row ID	S Countr...	S Countr...	I Year	L Population
Row1_Row0	AFG	Afghanistan	1950	7480464
Row1_Row1	AFG	Afghanistan	1951	7571542
Row1_Row2	AFG	Afghanistan	1952	7667534
Row1_Row3	AFG	Afghanistan	1953	7764549
Row1_Row4	AFG	Afghanistan	1954	7864289
Row1_Row5	AFG	Afghanistan	1955	7971933
Row1_Row6	AFG	Afghanistan	1956	8087730
Row1_Row7	AFG	Afghanistan	1957	8210207
Row1_Row8	AFG	Afghanistan	1958	8333827
Row1_Row9	AFG	Afghanistan	1959	8468220
Row1_Row10	AFG	Afghanistan	1960	8622473
Row1_Row11	AFG	Afghanistan	1961	8790140
Row1_Row12	AFG	Afghanistan	1962	8969055
Row1_Row13	AFG	Afghanistan	1963	9157463
Row1_Row14	AFG	Afghanistan	1964	9355510
Row1_Row15	AFG	Afghanistan	1965	9565154
Row1_Row16	AFG	Afghanistan	1966	9783153
Row1_Row17	AFG	Afghanistan	1967	10010037
Row1_Row18	AFG	Afghanistan	1968	10247782
Row1_Row19	AFG	Afghanistan	1969	10494491
Row1_Row20	AFG	Afghanistan	1970	10752973
Row1_Row21	AFG	Afghanistan	1971	11015853
Row1_Row22	AFG	Afghanistan	1972	11286753
Row1_Row23	AFG	Afghanistan	1973	11575308
Row1_Row24	AFG	Afghanistan	1974	11869881
Row1_Row25	AFG	Afghanistan	1975	12157390

Figure 24: Final dataset

III. Visualisation Design

The main topic of this project is the economic effects on life expectancy. To showcase the data visualization to the audience, we have developed three main charts for the audience. The first chart will show the life expectancy data at birth. The second chart discussed how GDP, GDP per capita affects life expectancy. The final chart will show the distribution of life expectancy. Each chart is designed to highlight different aspects of the relationship between economic indicators and life expectancy.

Design Guideline

In designing the visualization tools for our website, we adhered to the principles outlined by Edward R.Tufte or known as Tufte's Design Principle (Tufte, 1983) and MUNZNER guidelines (Munzner, 2014). These guidelines ensure that our visualizations are not only aesthetically pleasing but also effectively convey complex data. Tufte's guidelines emphasize clarity, precision and efficiency in the visual presentation of data. Additionally, we incorporated elements of color theory to enhance readability and user engagement particularly using shades of blue to symbolize health.

Tufte's principles:

- **Data-Ink Ratio:** Tufte emphasizes maximizing the data-ink ratio, which means reducing non-essential ink that does not convey information. In our visualizations, we focused on minimizing gridlines, backgrounds, and other non-data elements to highlight the actual data.
- **Chartjunk:** We avoided chartjunk—unnecessary or distracting decorations in graphs. All elements in our charts serve a functional purpose, ensuring clarity and focus on the data itself.
- **Small Multiples:** We applied the concept of small multiples for comparative analysis, particularly in the choropleth maps. This allows users to compare multiple maps side by side, facilitating an easy comparison of life expectancy across different regions and time periods.

- **Precision and Clarity:** Our charts are designed to present data with high precision. We used clear labeling and scaled axes to ensure users can interpret the data accurately.

Munzner's Guidelines:

- **Task Abstraction:** According to Munzner (Munzer 2014), understanding the tasks that users will perform with the visualization is critical. We identified key tasks such as exploring geographical distribution (choropleth maps), analyzing relationships between variables (bubble charts), and understanding data distribution (density plots).
- **Data Transformation:** We transformed raw data into visual formats that align with the tasks. For instance, in the bubble charts, we scaled bubble sizes to represent population size and used axes to show life expectancy and GDP per capita, different color shades of bubble encode the continent.
- **Visual Encoding:** Munzner emphasizes effective visual encoding. We used position, size, and color to encode data dimensions. For example, in the bubble charts, position encodes life expectancy and GDP per capita and size encodes population.·
- **Interaction Techniques:** We incorporated interaction techniques such as tooltips, zooming, and filtering. These allow users to explore the data in more detail and interactively adjust the parameters to gain deeper insights.

Color Theory

Blue is often associated with health, trust and tranquility which makes it suitable for this visualizations related to life expectancy and economics. It evokes a sense of calmness and reliability, which is essential for conveying serious and impactful data (Ware, 2012; Kosslyn, 2006; Few, 2006). In our design, we will use a monochrome blue color to maintain perceptual uniformity, ensuring that variations in color intensity are easily distinguishable. By using a monochrome color palette, it provides consistent design and ensures visual consistency across different types of charts, creating a cohesive and professional look for the website (Kosslyn, 2006). Moreover, it simplifies the design, reducing cognitive load on users.

However, this theory will only apply to two of our main charts which is the choropleth map and density chart. For bubble charts, it is necessary to use different color shades to group data. The purpose of using different colors is to enhance data differentiation. Different colors are important to distinguish between multiple variables and categories. Each color can represent a distinct data set, making it easier for users to identify and compare data points. According to Few (2006), color can be an effective tool for differentiating data series which is particularly useful in complex visualizations like bubble charts. In bubble charts, where size and position already encode information, using different colors helps to prevent visual overlap and clutter. It ensures that each data point is easily distinguishable, even when bubbles overlap. This principle is supported by Evergreen (2016), who highlights the importance of using color to reduce visual complexity and enhance clarity. Colors facilitate quicker cognitive processing by allowing users to instantly recognize and categorize information. Ware (2012) emphasizes that our visual system is highly attuned to color differences, which helps in rapidly identifying patterns and anomalies in data visualizations.

1. Prototype and Sketch

a. Life Expectancy at birth chart design

Alternative Design:

A bar chart is widely used for displaying and comparing data, making it an excellent choice for showing life expectancy across all countries in a given year. Its clear and straightforward design allows the audience to easily identify which countries have the highest and lowest life expectancy. Each bar represents a country, with its length proportional to the life expectancy value, enabling precise reading and easy focus on individual data points. However, due to the extensive amount of countries data involved in this topic, consequently, the amount of bars for countries would be over scale from the screen of users so this would not be considered as the final design.

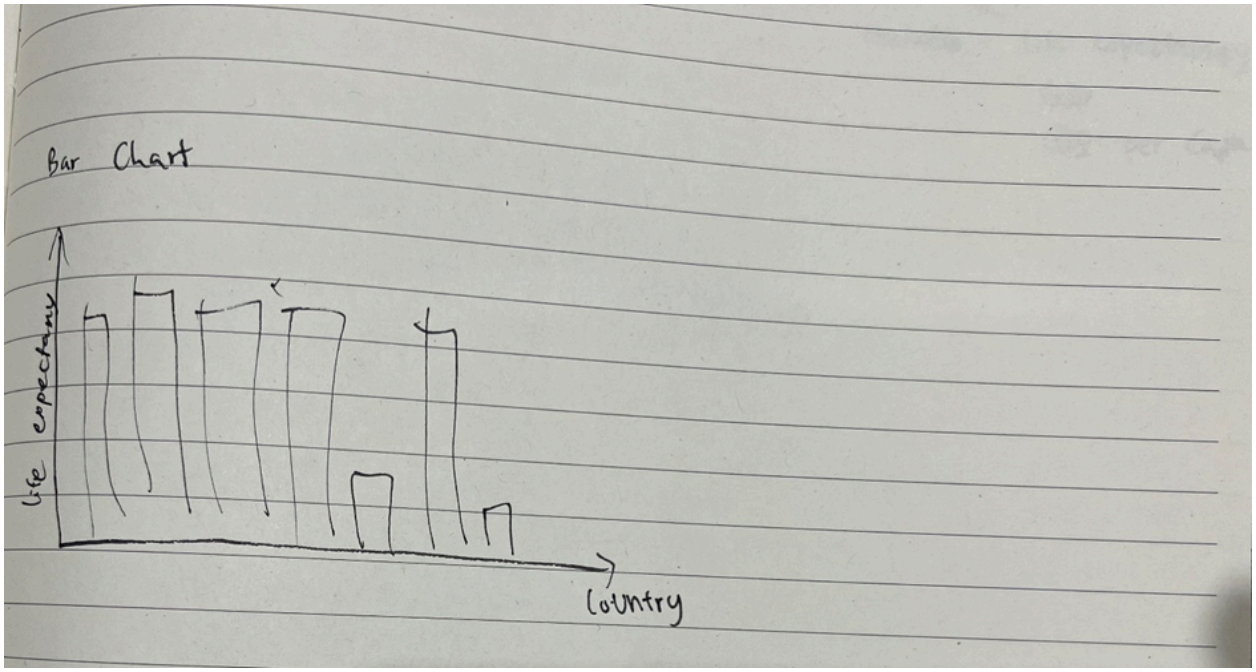


Figure 25: Life Expectancy at birth

Final Design:

Choropleth map is a more effective choice for visualizing life expectancy. Although it does not allow for precise value reading (bubble charts in charge of showing precise values), it provides a clear geographical context by showing variations in life expectancy on a map. Color gradients represent data values, with darker colors indicating higher life expectancy and lighter colors indicating lower life expectancy, making it easy for users to quickly grasp the information. Maps are inherently engaging and can capture viewers' attention more effectively than other chart types, making the data more meaningful and impactful. Additionally, choropleth maps effectively display large datasets in a compact form, conveying information through a single image using color coding. Therefore, a choropleth map is a superior design to a bar chart for providing a visually appealing and impactful visualization of life expectancy data across different regions.

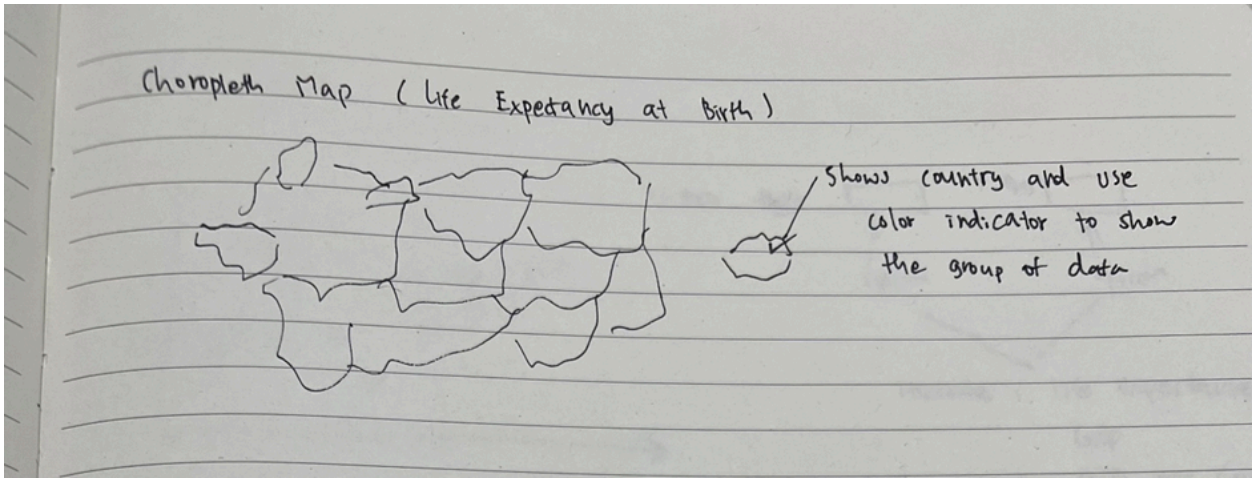


Figure 26 : Life Expectancy at birth

Purpose Alignment:

The design of the choropleth map enables audiences to effectively address *questions 1 and 2 from the Project Purpose in the Introduction*, by using varying shades of color to represent life expectancy values across different countries and regions, the choropleth map clearly highlights geographical disparities and temporal trends. This visual representation allows users to easily pinpoint areas with notable changes and observe life expectancy progression or regression over specific time periods, providing a comprehensive overview of global health patterns.

b. Economic effect on life expectancy chart design

Alternative Design:

The first initial design is using a scatter plot chart to show the impact of GDP on life expectancy. The scatter plot takes two data which is GDP or GDP per capita and compares it with life expectancy. This design is to the relationship between GDP and life expectancy with a clear, intuitive and effective design to visualize and analyze how economic correlates with health outcomes across different counties.

Since this chart only conveys two variables with the same color, the advantage of this is this could allow users to identify the overall trend of them over the world. However, this

chart is not effective showing data variation to identify more information for users because of limitations in identifying countries.

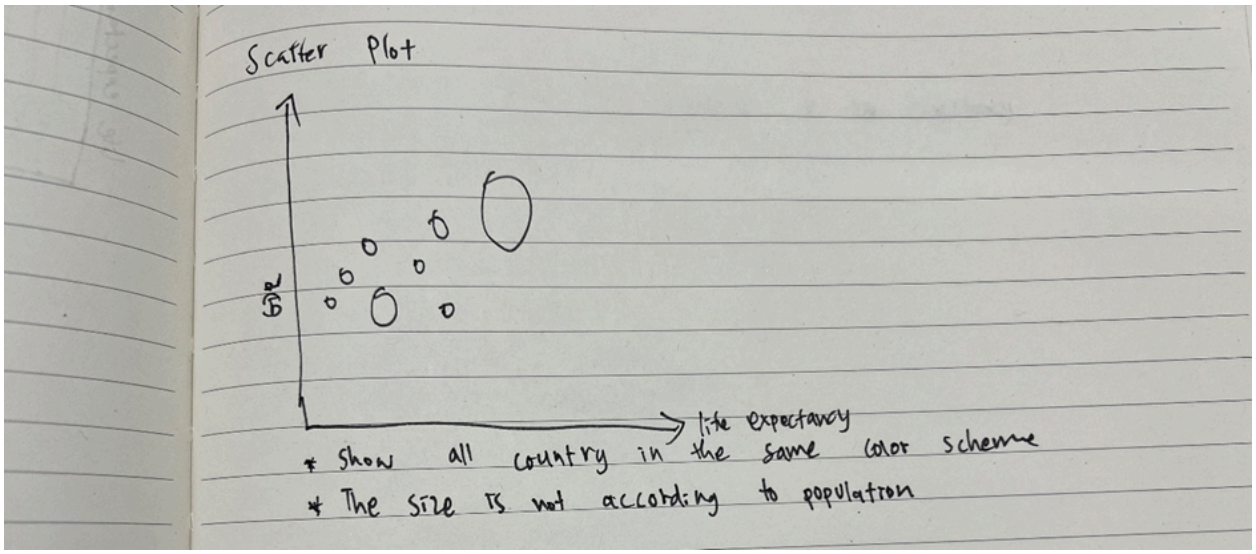


Figure 27: GDP effect on life expectancy

Final Design:

On the other hand, by inheriting the principle of “data variation” from Tufte design guidelines to enhance the visualization the chart takes in another variable which is the population data. So, to use three data variables at the same time, we could use a bubble chart. The size of the bubble is calculated using population data. This data can provide insights and more context to the chart. The population size can indicate the relative significance or weight of each data point.

This chart enhances the usual scatter plot by incorporating bubble sizes to represent population, providing deeper insights into the relationship between economic growth and life expectancy. Users can identify anomalies, such as countries with high GDP but low life expectancy, and explore how population size impacts GDP per capita and, consequently, life expectancy. This visualization helps in understanding how population dynamics can influence economic and health outcomes.

For instance, a country can have high GDP but might have low life expectancy, users could not identify this remarkableness by usual scatter plot. By leveraging the bubble

size with population, users could have more insight of how this could lead to lower the GDP per capita and can draw to their conclusion or further research about the significant changes. In addition to the previous example, a small population country might have a high GDP per capita and will have less impact than the same observation in a highly populated country. This also highlights the trends and patterns with the size of the countries and the influence of the trend.

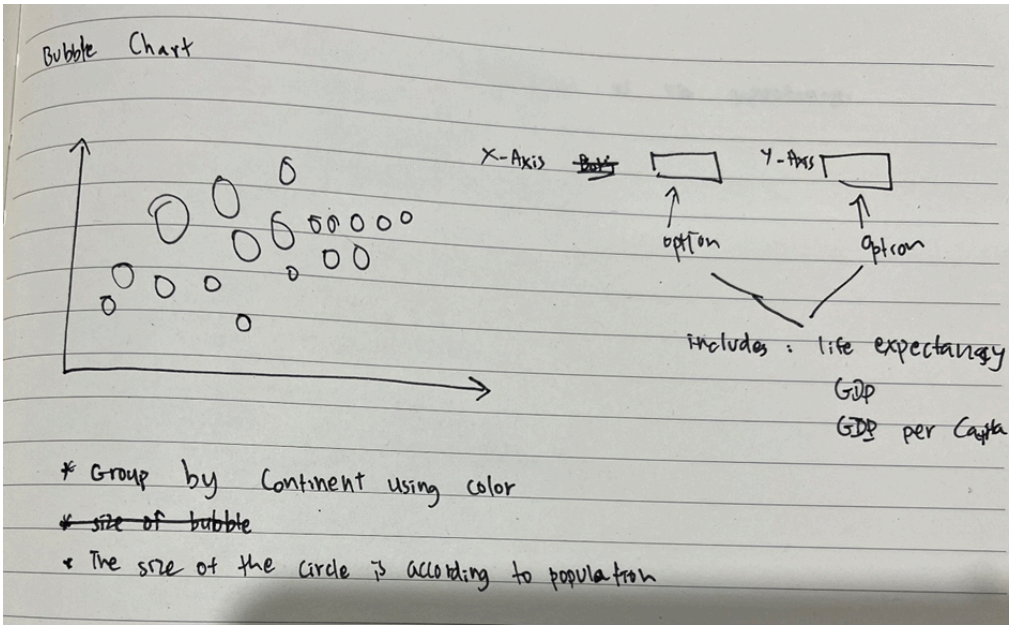


Figure 28: GDP effect on life expectancy

Additional Features:

In the chart above, the user could choose the data that they want to compare. In the option box, the user could choose to compare life expectancy data with GDP or GDP per capita. Moreover, if both users chose the same axis data as both axis shows life expectancy, it will compare the data across the region. In addition, the color of the bubble is grouped using the country region so bubbles that have the same color are located in the same region.

Purpose Alignment:

From the design of the bubble chart, audiences can answer *question 3 and 4 of Project Purpose in the Introduction* by identifying correlations between a country's GDP per

capita and its life expectancy, highlighting whether higher income levels are associated with longer life spans.

c. Life Expectancy distribution chart design

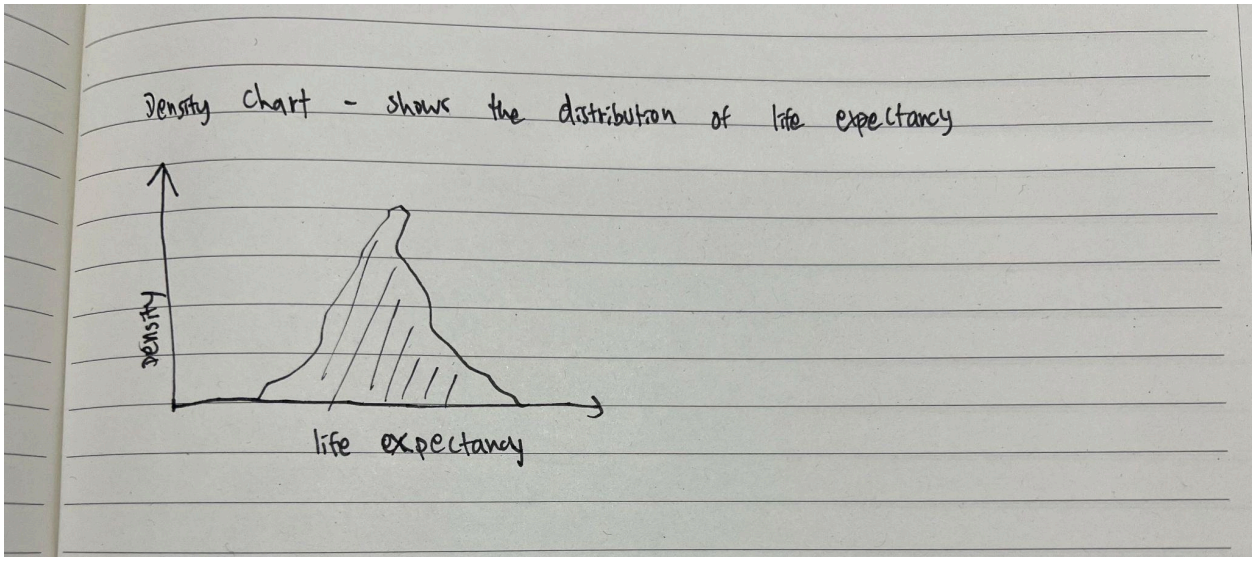


Figure 29: Density Chart on life expectancy

Final Design:

In Figure 29, we use a density chart to show the concentration of data points rather than individual points. Using a density chart to visualize life expectancy data offers several advantages, especially when dealing with large datasets.

Density charts effectively manage and represent extensive data by showing the concentration of data points rather than individual points, reducing overplotting and providing a clearer visual summary of distribution patterns. This visualization method highlights clusters of countries with similar life expectancies and helps identify anomalies or outliers that may require further investigation.

Furthermore, density charts can be customized to include multiple dimensions, such as using color to represent different groups or time periods, providing a richer and more nuanced understanding of the data. This flexibility makes density charts a powerful tool for analyzing life expectancy and its underlying factors.

Additional Features:

Density chart designed to let the user click on any country on a choropleth map, it will show the distribution of life expectancy in that country with mean statistical data of the country. Otherwise, it will calculate all the data of all countries and show the density and mean in the chart.

Purpose Alignment:

The density charts are excellent in answering the *question 5 of Project Purpose in the Introduction*, by comparing distributions with average mean across different countries, enabling a straightforward comparative analysis. They also allow for the visualization of trends over time, making it easier to see how life expectancy has evolved. The intuitive and visually appealing nature of density charts offers immediate insights, helping viewers quickly understand the overall distribution and central tendencies of the data.

2. Implementation

a. Choropleth

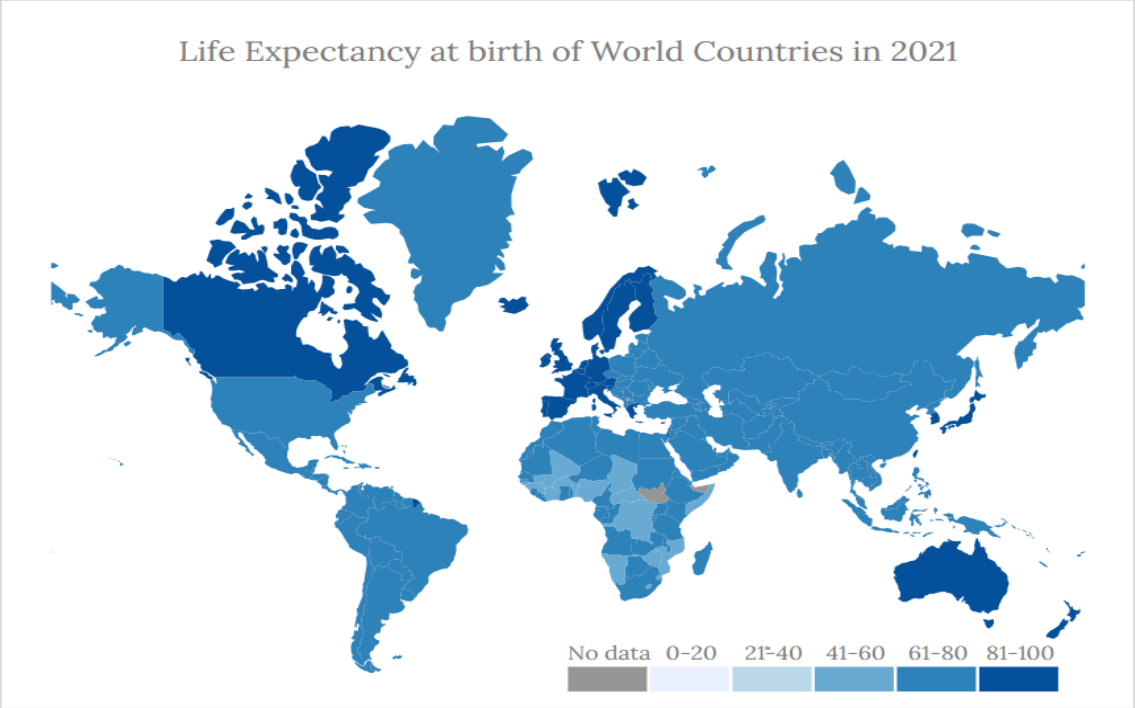


Figure 30: Implementation of choropleth map using life expectancy data

The choropleth manages to illustrate the life expectancy value for each individual country. The dataset for the chart is quantitative attributes for determined geometry for each country in the world**.** The choropleth chart visualizes life expectancy data across different countries of the world. It allows users to explore the data for different years using an interactive slider.

Key Features

1. *Color Scale: A color scale to represent the life expectancy values.*
2. *Tooltip: An interactive tooltip that displays the country name and life expectancy value when a user hovers over a country.*
3. *Legend: A legend explaining the color scale.*
4. *Title: A descriptive title at the top of the chart.*

5. *Interactive Slider: Allows users to select different years to view the corresponding life expectancy data.*
6. *Density Chart Integration: Clicking on a country displays a radar chart for more detailed information (integration with density.js).*

Configuration Settings:

- *SVG Dimensions: Scale with screen of user (Appendix D)*
- *Projection: Mercator projection centered at [0, 55], scaled and translated to fit within the SVG dimensions.*
- *Colors: Blue color scheme for the choropleth map with a range from light blue to dark blue.*

Steps for Implementation:

1. *SVG Creation:*
    - *Create the main SVG container within a div with id #choropleth.*
    - *Append an SVG element and set its dimensions considering padding.*
1. *Scales and Projection:*
    - *Define the color scales for the map and legend.*
    - *Set up the Mercator projection.*
2. *Load Data:*
    - *Load the life expectancy CSV data and GeoJSON file.*
    - *Filter data for the selected year and merge with GeoJSON.*
3. *Plotting Countries:*
    - *Append paths to the SVG for each country, setting the fill color based on life expectancy.*
4. *Tooltip:*
    - *Create a tooltip div and define functions to show and hide it on hover.*
5. *Legend and Title:*
    - *Append a legend and title to the SVG.*
6. *Interactive Slider:*

- *Add an event listener to update the map when the slider value changes.*

7. *Country Click Integration:*
    - *Define a function to handle the country data when a country is clicked.*
8. *Export and Initialization:*
    - *Ensure the necessary modules are exported and the choropleth map is initialized on window load.*

b. Bubble Chart

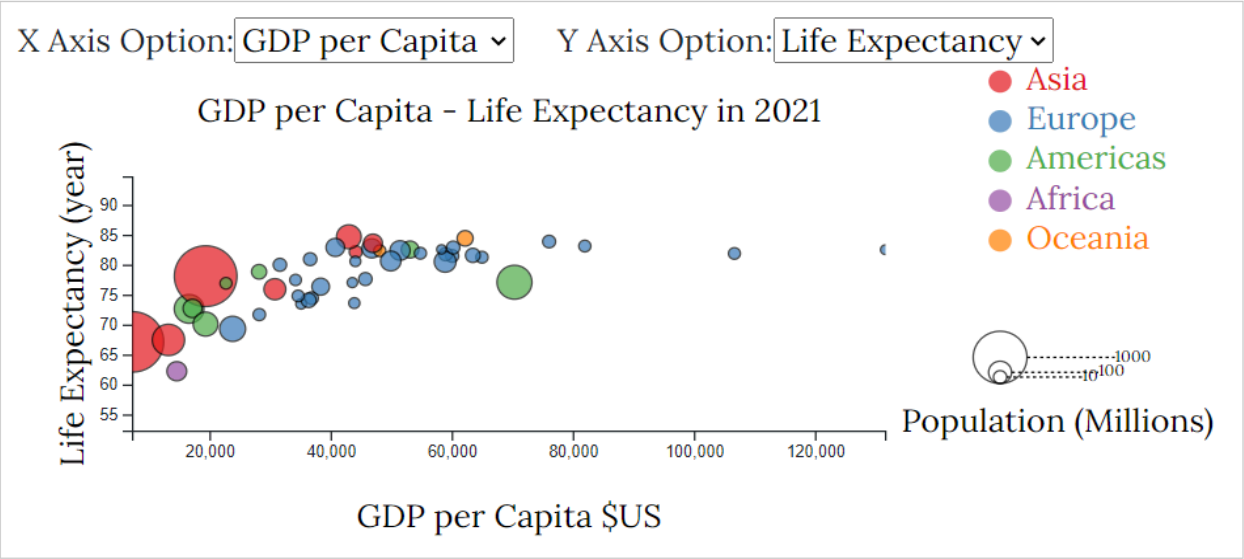


Figure 31: Implementation of bubble chart using life expectancy, GDP and GDP per capita data

The scatter plot chart provides an interactive exploration experience, enabling users to select different variables for the X and Y axes and observe the impact of these variables over time. It employs bubble sizes and colors to represent population and continents, facilitating insightful visual comparisons across different regions. Additionally, the chart supports dynamic updates, reflecting data changes in real-time as users adjust the year or variable selections, ensuring an engaging and informative visualization.

Key Features:

1. *SVG Container: The main area where the scatter plot will be drawn.*
2. *Axes: X and Y axes to represent the selected variables.*
3. *Bubbles: Circles representing countries, sized by population and colored by continent.*
4. *Tooltip: Provides detailed information when hovering over a bubble.*
5. *Legend: Explains bubble sizes and colors, helping users understand the representation.*
6. *Zoom and Pan: Allow users to zoom in and out and pan across the plot for a closer look at specific areas.*

Configuration Settings:

- *Width (cfg.w): Scale with screen of user*
- *Height (cfg.h): Scale with screen of user*
- *Padding (cfg.padding): 50px around the chart for axis labels and legends.*
- *Bubble Radius (cfg.radius): Initial radius for bubbles.*
- *Border (cfg.border): 1px*
- *Border Color (cfg.bordercolor): Black*

Steps for Implementation:

1. *SVG Creation:*
    - *Create the main SVG container within a div with id #plot.*
    - *Append an SVG element and set its dimensions considering padding.*
2. *Scales and Axes:*
    - *Define linear scales for X and Y axes.*
    - *Set up color scale for continents using d3.schemeSet1.*
    - *Add event listeners to update the chart based on user selection of X and Y axis variables.*
3. *Data Handling:*

- *Create a function get_data_from_year to fetch and merge data for the selected year.*
- *Extract and filter relevant data columns dynamically.*

4. *Plotting Bubbles:*
    - *Define scales for X, Y, and bubble size based on population.*
    - *Use D3 to append circles to the SVG for each country, sized by population and colored by continent.*
    - *Add transition animations for smooth updates.*
5. *Tooltip:*
    - *Create a tooltip div for displaying detailed information on hover.*
    - *Define handleMouseOver and handleMouseOut functions to show and hide the tooltip.*
6. *Legend:*
    - *Append circles and lines to the SVG to explain bubble sizes (population).*
    - *Add text labels for the legend.*
    - *Include a legend for continent colors, with interactivity to highlight specific continents.*
7. *Zoom and Pan:*
    - *Implement zoom and pan functionality using D3's zoom behavior.*
    - *Update scales and re-draw bubbles and axes during zoom/pan events.*
8. *Dynamic Updates:*
    - *Ensure the chart updates when users change the year or variable selections.*
    - *Use D3's enter, update, and exit pattern to handle data binding and transitions.*

c. Density Chart

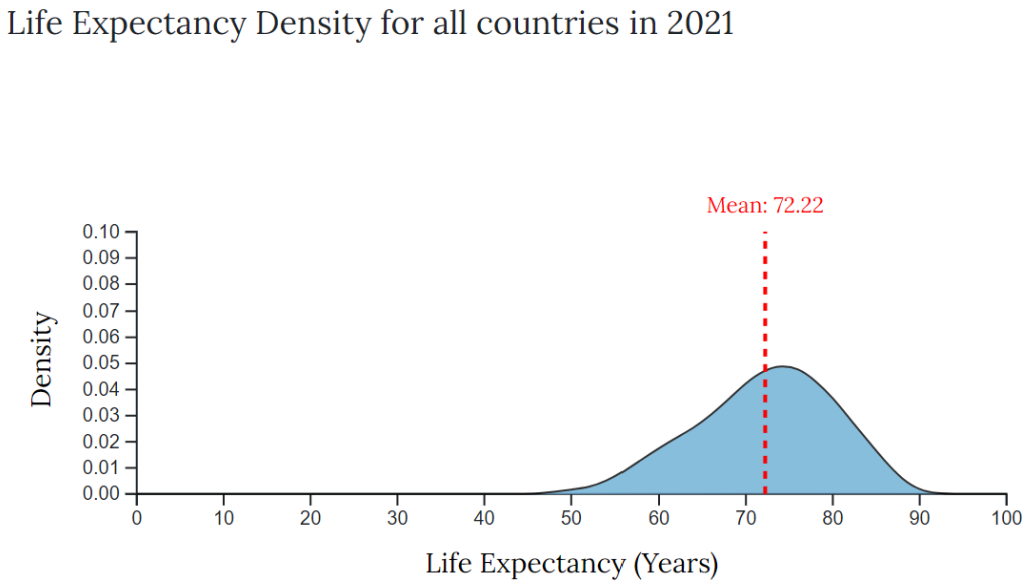


Figure 32: Implementation of density chart using life expectancy data

Density plots visualize the distribution of a numeric variable. It uses a kernel density estimate to show the probability density function of the variable. In our chart, the density chart is used to show the life expectancy of every country developed over the year where data points are concentrated, indicating the likelihood of different values. The x-axis represents the range of life expectancy and the y-axis represents the density of data points at each value of the x-axis. Moreover, as the chart shows the life expectancy of each country per year, the audience could see the overall change of life expectancy. The chart also illustrates a median value of life expectancy as well. The chart interacts with the choropleth where it will display the density of any country when the user clicks on the choropleth map. The chart provides dynamic updates, allowing users to interactively explore changes over time and view specific statistics such as mean and standard deviation.

Key Features:

- **SVG Dimensions**: Scale with screen of user
- **Dynamic Updates:** The chart updates in response to year selection and country clicks on choropleth map.
- **Statistics:** Displays mean and standard deviation for the selected dataset.
- **Tooltip:** Provides highlight detailed area of statistics when hovering over the chart.
- **Labels:** Clearly labels the axes for the density plot.

Configuration Settings:

- ***Margin:*** *Based on window size, adjusting dynamically.*
- ***Width:*** *One-third of the window's width.*
- ***Height:*** *One-fourth of the window's height.*
- ***Font Size:*** *Scales with the width of the chart.*

Steps for Implementation:

1. ***Define Margins and Dimensions:***
    - *Set initial margins and dimensions based on the window size.*
    - *Update dimensions dynamically when the window is resized.*
2. ***Create Density Chart Container:***
    - *Remove existing elements and create a new SVG container for the density chart.*
3. ***Load Data and Set Event Listeners:***
    - *Load life expectancy data and set up event listeners for country clicks and year slider changes.*
4. ***Initialization:***
    - *Initialize the density chart on window load and resize.*
5. ***Distribution Calculation:***

- ***Kernel Density Estimator (KDE):*** *Utilizes a kernel function (Epanechnikov kernel) to estimate the probability density function of the data.*
- ***Kernel Function:*** *The Epanechnikov kernel, a common choice, defined as:*

$$K(u) = ¾ (1 - u^2) \; support \; |u| \leq 1$$

- ***Density Calculation:*** *For each value in the domain (X), the KDE computes the average of the kernel function applied to the difference between the domain value and each data point.*

6. ***Draw the Curve:***
    - *Use a combination of the kernel density estimator and D3's line generator to draw the density curve.*

IV. Validation

Validation testing is crucial in software development as it ensures that the software meets the targeted audience's needs and requirements. It verifies that the product adheres to specifications and serves the intended purpose of the webpage. Various types of testing are employed, including Unit Testing, Integration Testing, System Testing, and User Acceptance Testing. For simplicity, this project focuses on user acceptance testing, which involves end-users checking the software's usability, functionality, and compatibility from their perspective.

The goal of this usability evaluation is to assess the final product's usability quality and to gather feedback from real users. There are two main types of evaluations: formative and summative. Formative evaluation occurs during the design phase, helping identify and refine the best design through sketches and prototypes. This type of evaluation ensures that the design is intuitive and aligns with user needs. In contrast, summative evaluation is conducted on the final product to determine if usability goals have been

met. This evaluation uses qualitative metrics such as error rates and time taken to complete tasks.

Usability evaluation methods can be categorized into two types:

1. Inspection/Analytical/Expert Methods: These methods do not require user involvement and include design walkthroughs, reviews, heuristic evaluations, and cognitive walkthroughs.

2. Observation/Empirical Methods: These methods involve users and include techniques like Rapid Iterative Testing and Evaluation (RITE), quick and dirty methods, controlled evaluations, remote evaluations, and field evaluations.

Our validation is divided into two stages. First stage of the validation, we have conducted a low-fidelity evaluation to help identify the best design for our visualization. The data collected is shown in Appendix C. In the last stage of development, we have conducted usability testing using a remote evaluation method. Remote Evaluation method could be done by using a software to monitor the user action or having a team member observing the action of the user. In our project, we arranged a time with the user and conducted the moderated remote evaluation where each of the members observed the action of the user and took note of the user's behavior during the testing. The testing includes collecting ten end-users to test the usability of our website. Each of the users has to sign a consent form that they agree for us to collect data and participation in this project. Our team has created an informed consent form for the user to sign and questions about themselves (See Appendix A). The user is asked to perform tasks related to our usability validation. At the end of the testing, the user is asked to provide feedback using a post-study form (See Appendix A). The collected feedback and data can be found in Appendix B.

Usability Testing Task

1) Determine the website objective to understand the goal of the website

2) Determine the project objective to identify the project's specific objective
3) Locate the visualization page
    a) Locate the choropleth map
        i) Observe the change of life expectancy over the year
        ii) Identify the life expectancy data of the country on the map
    b) Locate the bubble chart
        i) Observe the effect of GDP on life expectancy
        ii) Observe the effect of GDP per capita on life expectancy
        iii) Compare the data of life expectancy all around the world
        iv) Compare the data of GDP all around the world
        v) Compare the data of GDP per capita all around the world
    c) Locate the density chart
        i) Observe the change of life expectancy distribution over the year
        ii) Identify the density of any selected country from the choropleth map

V. Conclusion

In summary, this data visualization project has explored the relationship between economic factors and life expectancy by analyzing data on GDP, GDP per capita and life expectancy across various countries. The visualizations provide insightful conclusions and highlight the significant impact of economic conditions on the health and longevity of populations. In addition, our analysis revealed a strong positive correlation between a country's GDP and life expectancies through the bubble chart. This correlation suggests that economic prosperity enables better healthcare infrastructure, access to medical services, improved nutrition, and overall better living conditions, contributing to longer lives. Moreover, GDP per capita showed an even more pronounced correlation with life expectancy. It is noticeably that countries with higher GDP per capita get a better living condition which results in a better life expectancy. The visualizations also highlighted the stark disparities between countries with different economic standings. Low-income countries generally exhibited lower life expectancies,

underlining the challenges they face in providing adequate healthcare, clean water, and nutritious food. This disparity emphasizes the need for international support and sustainable development strategies to improve health outcomes in less affluent nations. Through the choropleth map, we could see the transition of life expectancy over the year across the world. The density chart highlights the distribution of life expectancy development of individual countries.

The analysis underscores the profound influence of economic factors on life expectancy. While higher GDP and GDP per capita are strongly associated with longer life expectancies, the interplay of economic resources with other social determinants of health cannot be overlooked. To further improve global health outcomes, policies should focus not only on economic growth but also on equitable distribution of resources, strengthening healthcare systems, and addressing social and environmental determinants of health. By leveraging economic prosperity to enhance public health infrastructure and services, countries can ensure that their populations enjoy longer, healthier lives.

References

Callaghan, D. (2023, May 22). *Health Economy Reporting: the ROI of health investments*. WifOR Institute. https://www.wifor.com/en/roi-of-health-investments/

Callen, T. (2023). *Purchasing Power Parity: Weights Matter*. International Monetary Fund. https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/Purchasing-Power-Parity-PPP

Cowan, N. (2005). *Working Memory Capacity*. Psychology Press. https://doi.org/10.4324/9780203342398

Clean Blog Angular - Bootstrap Blog Theme. (n.d.). Start Bootstrap. https://startbootstrap.com/theme/clean-blog-angular

Evergreen, S. D. H. (2019). Effective Data Visualization: The Right Chart for the Right Data. In *Google Books*. SAGE Publications. https://books.google.com.au/books?id=ilWKDwAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Evergreen, S. D. H. (2021). Effective Data Visualization: The Right Chart For The Right Data. Stephanie D. H.Evergreen. Thousand Oaks, CA: Sage, 2020, 328 pages, $58.00 paperback. *Personnel Psychology*. https://doi.org/10.1111/peps.12439

Few, S. (2006). *Information dashboard design : the effective visual communication of data*. O'reilly.

Hartson, R., & Pyla, P. S. (2012). UX Evaluation Introduction. *The UX Book*, 427–465. https://doi.org/10.1016/b978-0-12-385241-0.00012-9

Holtz, Y. (n.d.). *Density plot with slider in d3.js*. D3-Graph-Gallery.com. https://d3-graph-gallery.com/graph/density_slider.htm

Holtz, Y. (n.d.). *colored bubble plot in d3.js*. D3-Graph-Gallery.com. Retrieved June 5, 2024, from https://d3-graph-gallery.com/graph/bubble_color.html

Marin, S. (2020, December 14). *Prime Minister Sanna Marin | OECD 60th Anniversary*. Www.youtube.com; Finnish Government. https://youtu.be/0gu8vtIw7NI

Max Roser. (2015). *About*. Our World in Data. https://ourworldindata.org/about

Munzner, T. (2014). *Visualization Analysis and Design* (pp. 117–142). A K Peters/CRC Press Zurich. University of the Arts. https://doi.org/10.1201/b17511

Schmidt, K. J. (2019, April 10). *Quick fix to make your d3.js graph responsive*. Medium. https://medium.com/@kj_schmidt/quick-fix-to-make-your-d3-js-graph-responsive-bb63520cfba8

Stephen Michael Kosslyn. (2006). *Graph design for the eye and mind*. Oxford University Press.

Stone, D. L., Caroline, J., Mark, W., & Shailey, M. (2009). *User interface design and evaluation*. Elsevier.

Tufte, E. R. (1983). *The visual display of quantitative information* (pp. (ch.2, pp. 53-77)). Graphic Press. https://archive.org/details/visualdisplayofq0000edwa

Vipul @CanvasFlip. (2016, December 5). *Creating Task Scenarios that improves Usability Test Results*. Medium. https://weekdayhq.medium.com/creating-task-scenarios-that-improves-usability-test-results-eece56959d19

Ware, C. (2012). *Information visualization : perception for design* (3rd ed.). Elsevier/Morgan Kaufman.

APPENDIX A: Usability Evaluation test materials

1. Informed Consent Form: obtain a written consent from participants for their involvement in the study

Usability Evaluation Explanatory Statement

COS30045 Data Visualisation Project

Project Title: OECD Life Expectancy Visualisation

Principal Investigator(s): Davy Sung and Toan Nguyen

The aim of the project is to visualize and analyse the economic of life expectancy using OECD Data Statistic.

This project is being undertaken by the Swinburne University of Technology student team studying COS30045 Data Visualisation whose names appear at the top of this form.

Should you agree to participate in this study you will be asked to perform task regarding to the usability and of the website. You will be asked to work through a number of tasks using this system.

During the study we will ask you to fill out two questionnaires. The first is to collect information about you and feedback regarding the website. This will help us interpret the data we collect. The second is to ask you for your opinion about how you found using the system you just tried. This will help us understand how well you liked the system.

We anticipate that the study will take 10 minutes to complete.

The collected data and feedback will only used to prepare student reports for COS30045 Data Visualisation and will only be view by the student team member, tutor and convenor. You will not be personally identified by name in any of the data reported.

You do not need to answer all questions. If, for any reason, you feel uncomfortable with a question, do not answer it.

If you have any questions about this document or the consent form, please do not hesitate to ask.
We ask that you next read and sign our consent form.
We thank you for helping us in this project.

If you have any concerns please do not hesitate to contact us.

Please add your email address below for a copy of this information for your own reference.

* Indicates required question

1. Email *

Agreement to Participate

2. I am over the age of 18 and consent to participate in the project named above. I have been provided a copy of the project Explanatory Statement to which this consent form relates and any questions I have asked have been answered to my satisfaction. *

Mark only one oval.

- ◯ Yes
- ◯ No

3. 2. In relation to this project, please check the box with your response to the following: *

Check all that apply.

- ☐ I agree to be observed by the researchers
- ☐ I agree to allow the researchers to collect data for their report
- ☐ I agree to make myself available for further information if required

3. I acknowledge that:

(a) my participation is voluntary and that I am free to withdraw from the project at any time without explanation;

(b) the Student project is for the purpose of research and not for profit;

(c) any identifiable information about me which is gathered in the course of and as the result of my participating in this project will be (i) collected and retained for the purpose of this project and (ii) accessed and analysed by the student team, their tutor and the unit convenor for the purpose of conducting and assessing this project;

(d) my anonymity is preserved and I will not be identified in publications or otherwise without my express written consent.

By signing this document I agree to participate in this project.

4. Signature - Your name and date *

Figure 32: User Informed Consent Form

2. Demographic Form: this form is to gather information about the user

Pre-Study Questionnaire

COS30045 Data Visualisation Student Project
Project Title: OECD Life Expectancy Visualisation

Principal Investigator(s): Davy Sung and Toan Nguyen
Tutorial Day, Time, Semester, Year: Monday 4.30pm, Semester1 2024

* Indicates required question

1. ID. Participant ID (we will tell you what to put for this Question) *

_____

2. Which of the following includes your age?

Mark only one oval.

- ◯ 18 - 24
- ◯ 25 - 34
- ◯ 35 - 49
- ◯ 50 - 64
- ◯ 65+

3. What is your gender

Mark only one oval.

- ◯ Female
- ◯ Male
- ◯ Prefer not to say
- ◯ Other: _____

4. Have you heard of data visualisation?

Mark only one oval.

- ◯ Yes
- ◯ No

5. What is best describe you?

Mark only one oval.

- ◯ Student
- ◯ Professor
- ◯ Doctor
- ◯ Researcher
- ◯ Other: _____

6. What is your highest level of education?

Mark only one oval.

- ◯ High School
- ◯ Bachelor Degree
- ◯ Master Degree
- ◯ PHD

Figure 33: Demographic Questionnaire form

3. Post-Task Questionnaire: this form is to gather feedback after all are completed

COS30045 - Project Validation

B *I* U

Our data visualization project has reached its final stage, and we now need your help with validation. Please visit our webpage using the following link:

https://toannguyenswinadventure.github.io/COS30045-DataVisualisation/HomePage/index.html

For the best performance, we recommend using a desktop or laptop. Additionally, please use an email with your name or your student email for feedback.

Thank you for taking time to assist us!

Email *

Valid email

This form is collecting emails. Change settings

3. Do you find our website's interface overall attractive? *

Mark only one oval per row.

	Need improvement	Acceptable	Very nice 💯
Home Page	◯	◯	◯
About Project Page	◯	◯	◯
Group Project Page	◯	◯	◯
Visualisation Page	◯	◯	◯

4. Related to our above question, if you Like/Dislike any part of it, can you provide any specific elements you found appealing * or areas you think could be improved. Your feedback is valuable to us!

_____

_____

_____

_____

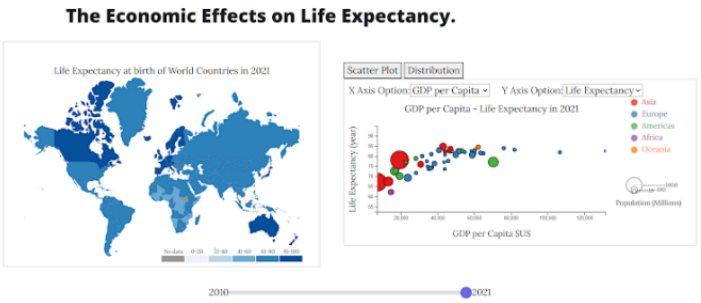
_____

5. This question is for the content of each page: *
Do you find the content on each page concise and easy to understand regarding the purpose of our project? If yes, tick to delivery your answer.

Check all that apply.

- ☐ Home Page
- ☐ About Page
- ☐ Group Project Page
- ☐ Visualisation Page
- ☐ Other: _____

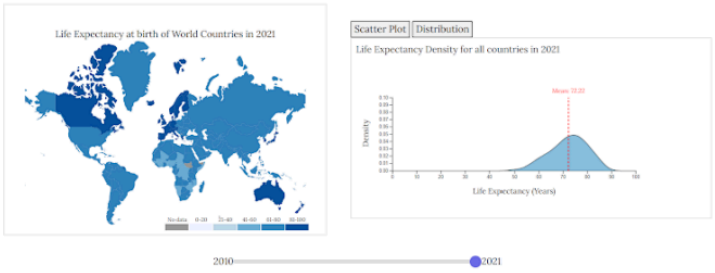
6. Does this screen show a map and a bubble chart like this when click on Visualisation Page? *



Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Other: _____

7. When click on distribution button, does the density chart pop up like the screen below? *



Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Other: _____

8. This question is related to our Visualisation Page. *
List of suggestion checklist for you:

Choropleth:
- *Functionality:*
  + *Does the country show as parital geometry with its corresponding color?*
- *Interaction:*
  + *When hovering on country, does it show tooltip (detail information)?*
  + *Does the chart title align with the year slider?*
- *Content delivery:*
  + *Does the chart allow you to understand the data?*

Scatter plot:
- *Functionality:*
  + *Does it show plots with color and size scalling with population correctly?*
  + *Does the plot point scales correctly with the coordinates?*
- *Interaction:*
  + *When hovering on country plot point, does it show tooltip (detail information)?*
  + *Can you zoom the chart (scrolling) and moving plot (hold left-click and slide the chart)*
  + *When hovering on continent legend, does it highlight the country belongs to that continent?*
  + *Does the chart title align with the year slider?*
- *Content delivery:*
  + *Does the chart allow you to understand the data?*

Density chart:
- *Functionality:*
  + *Does the chart show the density of life expectancy value?*
- *Interaction:*
  + When hovering on a chart's area, does the color becomes more intense?
  + *When clicking a country from the choropleth map, does the chart change the value according to the country?*
  + *Does the chart title align with the year slider?*
- *Content delivery:*
  + *Does the chart allow you to understand the data?*

*Tick the box if the above is working correctly

Check all that apply.

	Functionality	Interaction	Content delivery
Choropleth map	☐	☐	☐
Scatter Plot chart	☐	☐	☐
Density chart	☐	☐	☐

9. When using the slider, do all the charts change the value?

Mark only one oval.

- ◯ Yes
- ◯ No

10. Does the visualization allow you to gain insights from it? *

Mark only one oval.

- ◯ Yes
- ◯ No
- ◯ Other: _____

11. I find the interface of this website is easy to interact

Mark only one oval.

1	2	3	4
◯	◯	◯	◯

12. I feel very confident using this website

Mark only one oval.

1	2	3	4
◯	◯	◯	◯

13. Is there any feedback for the visualisation? *

_____

_____

_____

_____

_____

Figure 34: Post-Study Question Form

APPENDIX B: Data Collected Through Evaluation Testing

1. Observation Data

Through the observation by the team member, we have collected qualitative data:

Task 1

Participant ID	Time Spent	Error Encounter	Number of click	Comment
1	1 minute	0	2	The user finds it hard to understand the objective of the task that is why the user takes long to find the correct button
2	30 seconds	0	1	The user understands the task clearly
3	20 seconds	0	1	The user understands the task clearly
4	10 seconds	0	1	The user easily understand the task
5	20 seconds	0	1	The user understands the task clearly
6	25 seconds	0	1	The user understand the task clearly
7	30 seconds	0	1	The user understand the task clearly
8	20 seconds	0	1	The user understand the task clearly
9	39 seconds	0	1	The user takes time to understand the task
10	5 seconds	0	1	The user understand the task clearly

Table1: Result from evaluation testing observation task1

Task 2

Participant ID	Time Spent	Error Encounter	Number of click	Observation
1	20 seconds	0	1	The user understand the task clearly
2	15 seconds	0	1	The user understand the task clearly
3	10 seconds	0	1	The user understand the task clearly
4	30 seconds	0	1	The user understand the task clearly
5	13 seconds	0	1	The user understand the task clearly
6	10 seconds	0	1	The user understand the task clearly
7	20 seconds	0	1	The user understand the task clearly
8	30 seconds	0	1	The user understand the task clearly
9	29 seconds	0	1	The user understand the task clearly
10	10 seconds	0	1	The user understand the task clearly

Table 2: Result from evaluation testing observation task2

Task 3.1

Participant ID	Time Spent	Error Encounter	Number of click	Observation
1	2 minute 49 seconds	0	1	The user does not need any assistance in completing the task.
2	3 minute 10 seconds	0	1	The user does not need any assistance in completing the task.
3	2 minutes 30 seconds	0	1	The user does not need any assistance in completing the task.
4	2 minutes	0	1	The user does not need any assistance in completing the task.
5	3 minutes	0	2	The user does not need any assistance in completing the task.
6	2 minutes 4 seconds	0	1	The user does not need any assistance in completing the task.
7	3 minutes 20 seconds	0	2	The user could not find a way to see the data

8	2 minutes 30 seconds	0	1	The user does not need any assistance in completing the task.
9	3 minutes	0	1	The user does not need any assistance in completing the task.
10	3 minutes 35 seconds	0	1	The user does not need any assistance in completing the task.

Table 3: Result from evaluation testing observation task3.1

Task 3.2

Participant ID	Time Spent	Error Encounter	Number of click	Observation
1	4 minutes	1	2	The user does not need any assistance in completing the task.
2	3 minutes 30 seconds	0	1	The user does not need any assistance in completing the task.
3	2 minutes 45 seconds	0	1	The user does not need any assistance in completing the task.
4	3 minutes 20 seconds	0	1	The user does not need any

				assistance in completing the task.
5	2 minutes 49 seconds	2	1	The user does not need any assistance in completing the task.
6	3 minutes 45 seconds	1	1	The user does not need any assistance in completing the task.
7	4 minutes 10 seconds	1	2	The user takes long time to read the task and understand the interface of the website
8	3 minutes 40 seconds	0	1	The user does not need any assistance in completing the task.
9	3 minutes 5 seconds	0	1	The user does not need any assistance in completing the task.
10	4 minutes	1	1	The user does not need any assistance in completing the task.

Table 4: Result from evaluation testing observation task3.2

Task 3.3

Participant ID	Time Spent	Error Encounter	Number of click	Observation
1	1 minute	0	3	The user does not need any assistance in completing the task.
2	1 minute 22 seconds	0	3	The user does not need any assistance in completing the task.
3	2 minutes	0	3	The user does not need any assistance in completing the task.
4	1 minute 15 second	0	3	The user does not need any assistance in completing the task.
5	3 minutes	0	3	The user does not need any assistance in completing the task.
6	2 minutes	0	3	The user does not need any assistance in completing the task.
7	1 minute 10 seconds	0	3	The user does not need any assistance in completing the task.
8	2 minute 20 seconds	0	3	The user does not need any

				assistance in completing the task.
9	1 minute 30 seconds	0	3	The user does not need any assistance in completing the task.
10	2 minutes	0	3	The user does not need any assistance in completing the task.

Table 5: Result from evaluation testing observation task3.3

2. *Informed Consent Data Collection*

Agreement to Participate

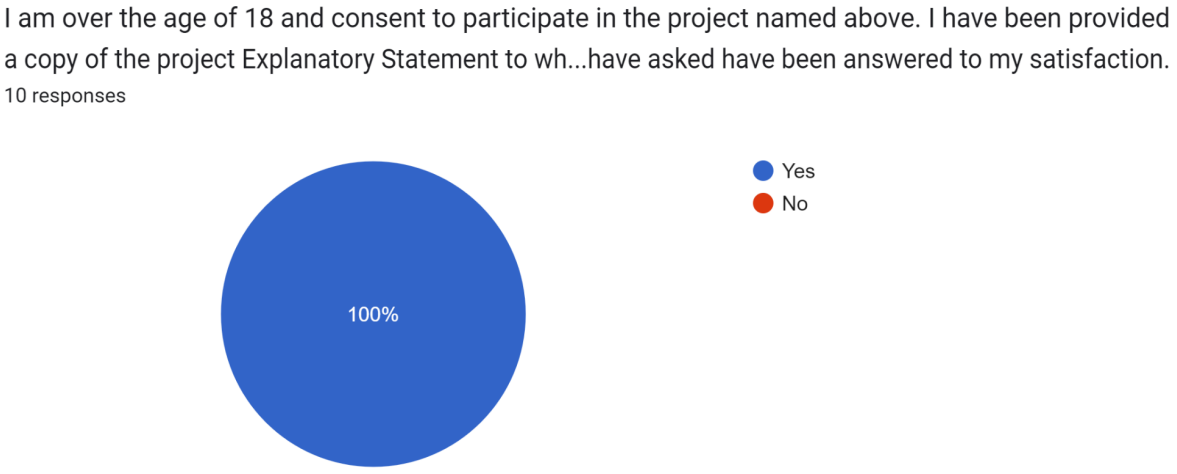


Figure 36: Result from Informed Consent Form

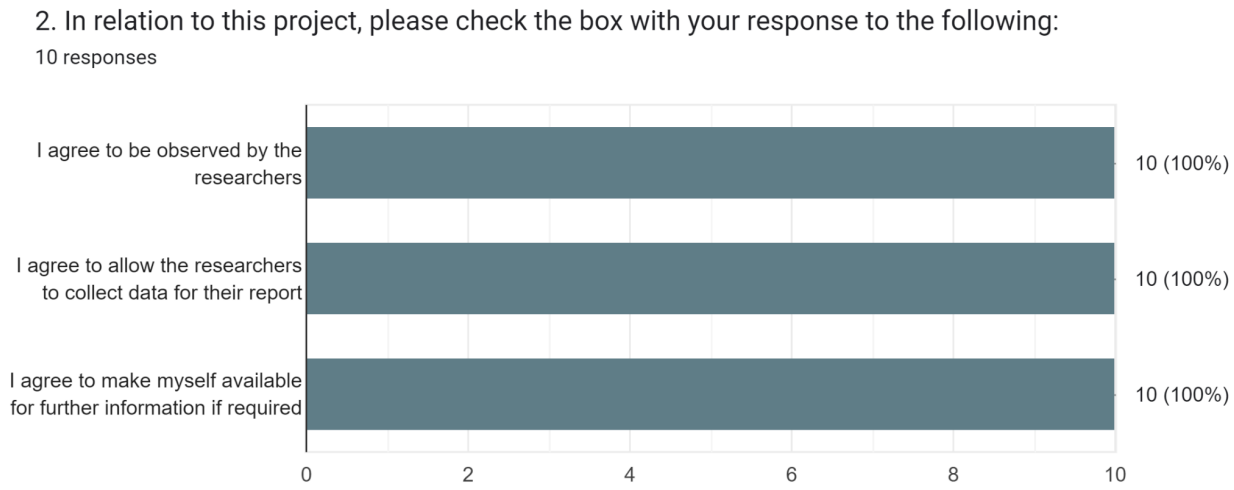


Figure 37: Result from Informed Consent Form

Acknowledgement

All users have signed to agree to participate in this survey and use the data for further study. However, to protect the participant's privacy, we will not include their signature in this document.

3. Demographic Data Collection

Question1

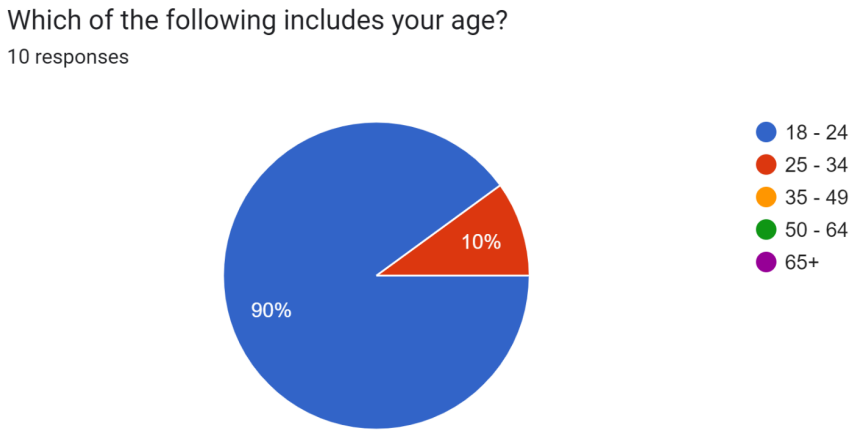


Figure 38: Result from Demographic Questionnaire

Question 2

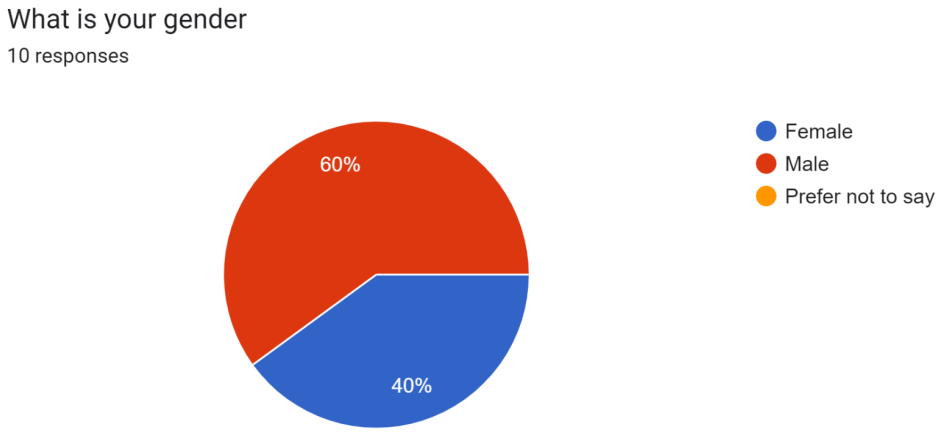


Figure 39: Result from Demographic Questionnaire

Question 3

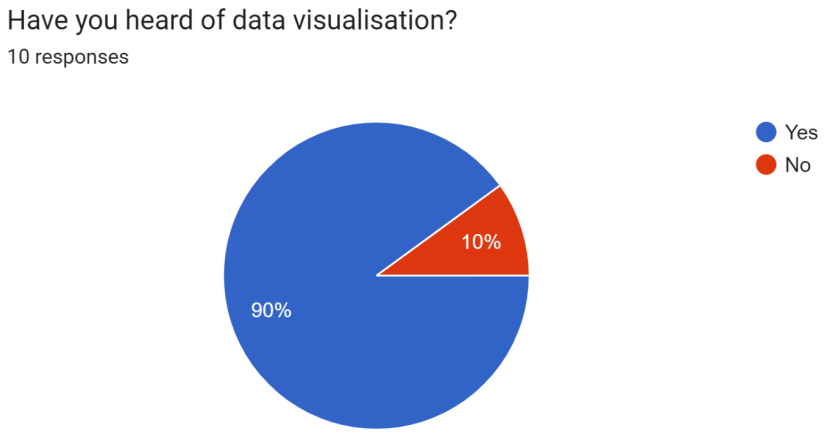


Figure 40: Result from Demographic Questionnaire

Question 4

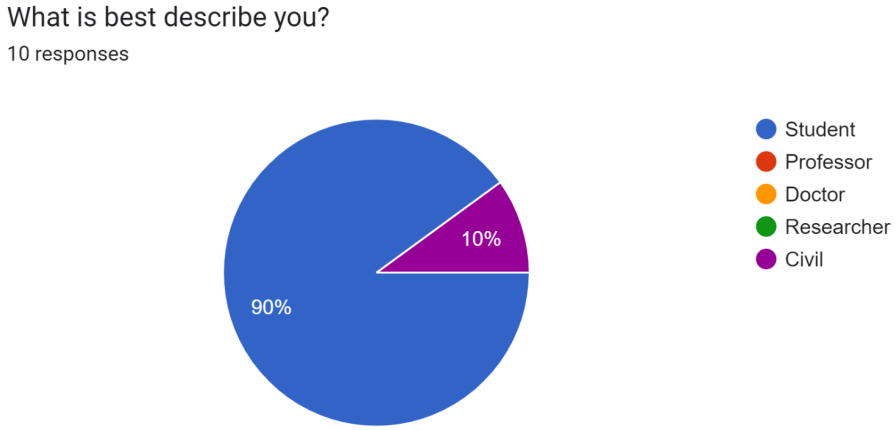


Figure 41: Result from Demographic Questionnaire

Question 5

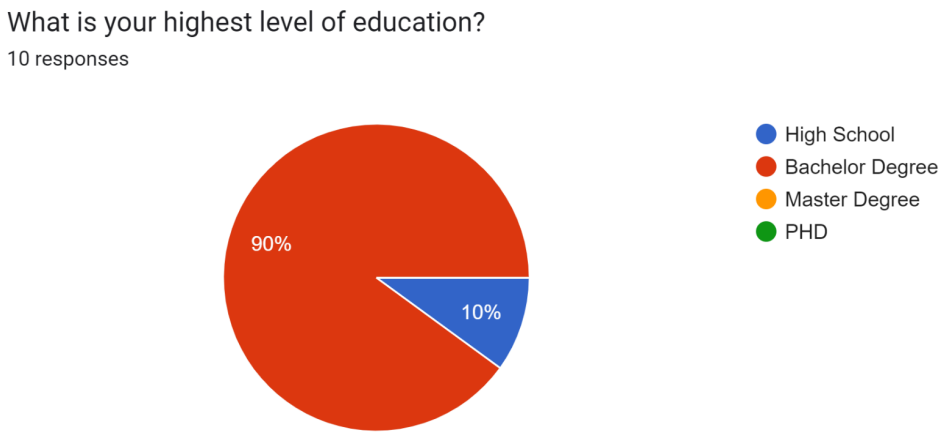


Figure 42: Result from Demographic Questionnaire

4. Post-Questionnaire Data Collection

Question1

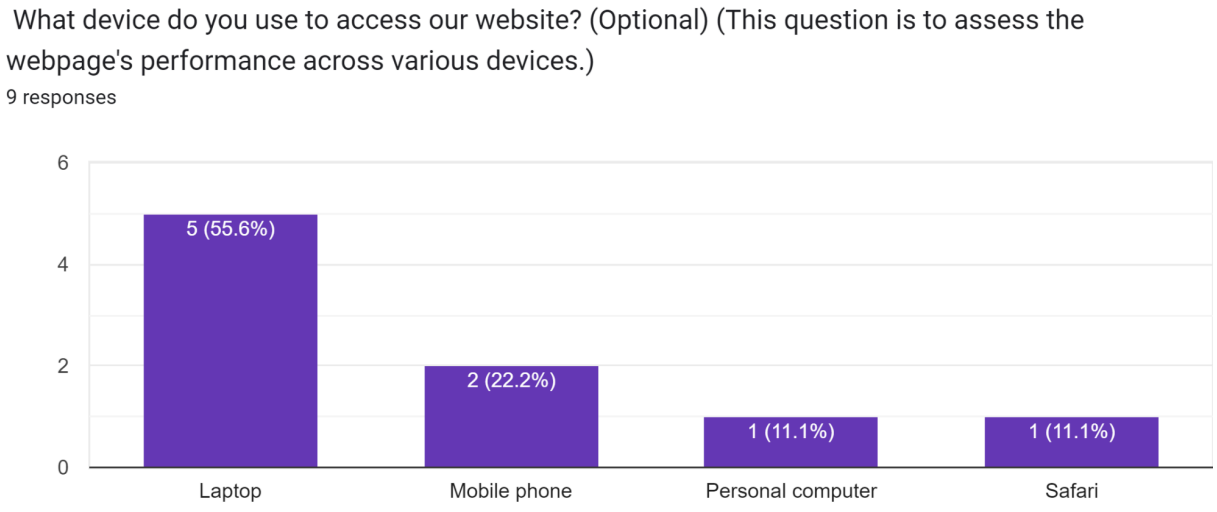


Figure 43: Result from post-study questionnaire

Question2

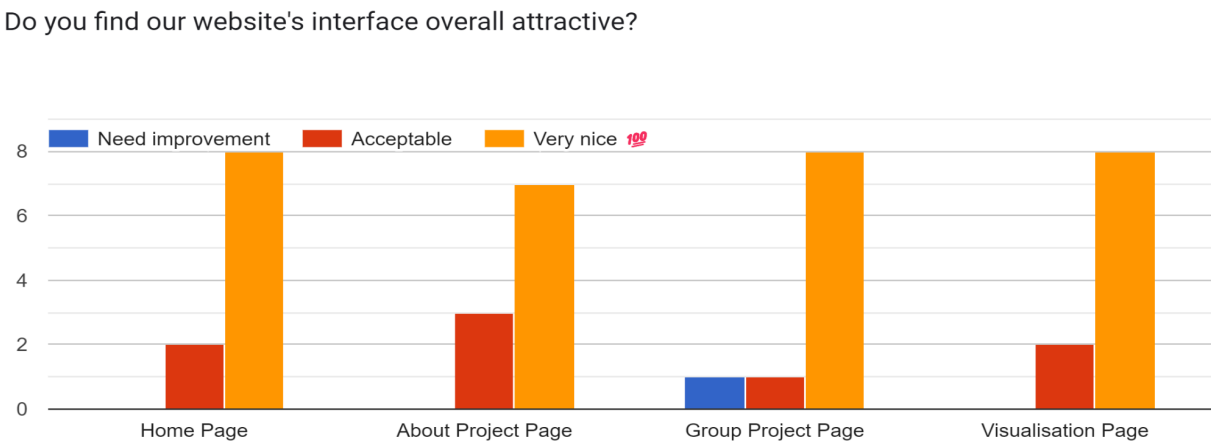


Figure 44: Result from post-study questionnaire

Question3

Related to our above question, if you Like/Dislike any part of it, can you provide any specific elements you found appealing or areas you think could be improved. Your feedback is valuable to us!

10 responses

None

Colour revision

the Visualization Page is clear and very insightful

All pages are made with consistent design and easy to navigate. Visualizations are highly aesthetic, delivering insightful and comprehensive information. This allows me to better understand the data and any important insights from the charts.

easy to use

too many words, need to explain everything briefly.

Visual was unique.

I like the interface of the website. It is very clear and precise and easy to navigate.

Actually, the homepage is quite good, however, I wish this page would have more images or more ways to access links to other parts of the website. As for the remaining pages, in general, they are quite impressive.

appealing: interactivity of scatter plot and distribution charts

Figure 45: Result from post-study questionnaire

Question4

This question is for the content of each page: Do you find the content on each page concise and easy to understand regarding the purpose of our project? If yes, tick to delivery your answer.

10 responses

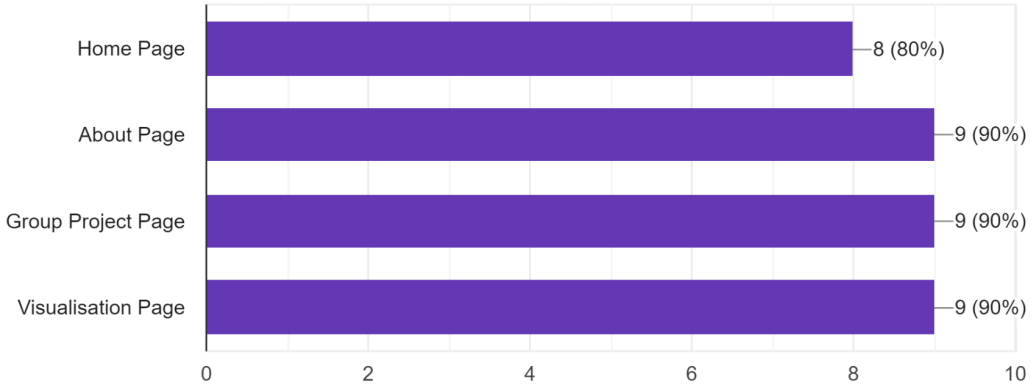


Figure 46: Result from post-study questionnaire

Question5

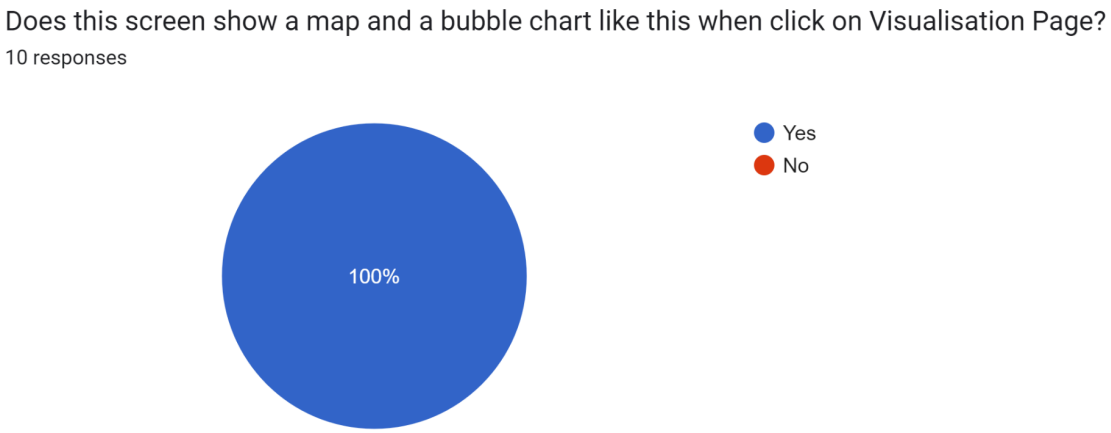


Figure 47: Result from post-study questionnaire

Question 6

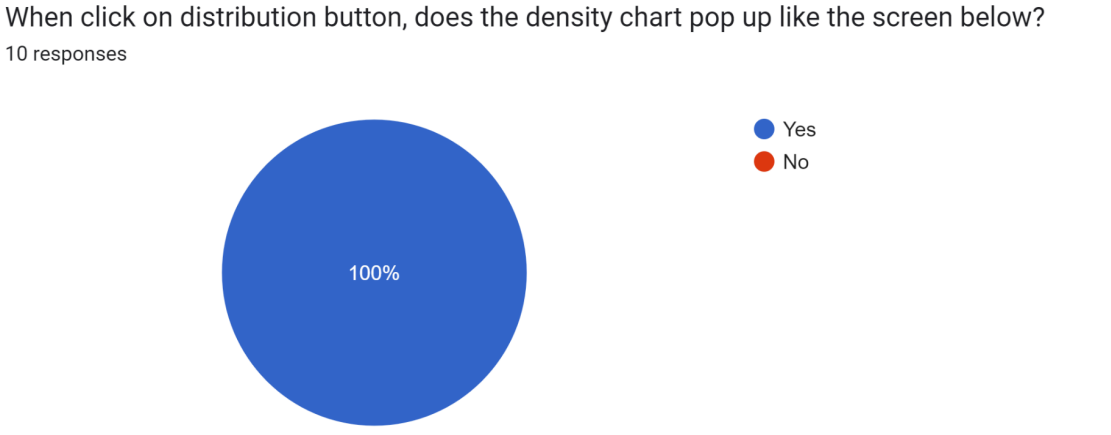


Figure 48: Result from post-study questionnaire

Question7

This question is related to our Visualisation Page. List of suggestion checklist for you: Choropleth: - Functionality: + Does the country show as parita...ata? *Tick the box if the above is working correctly

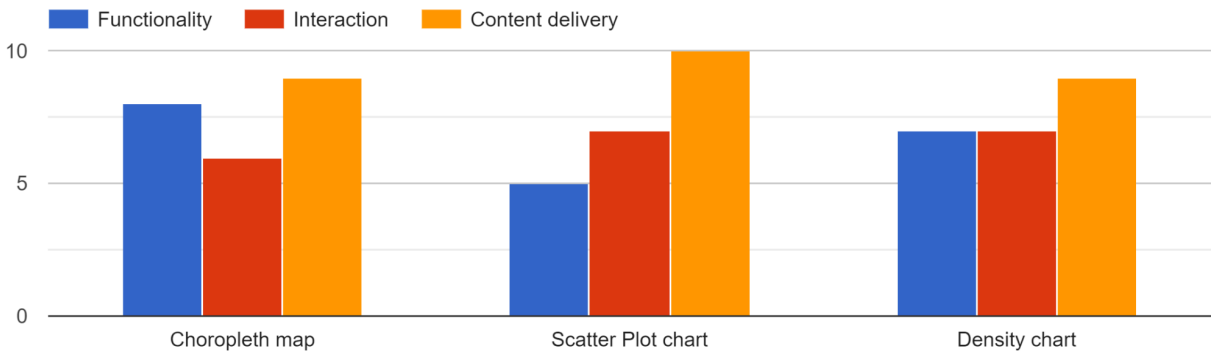


Figure 49: Result from post-study questionnaire

Question 8

When using the slider, do all the charts change the value?

10 responses

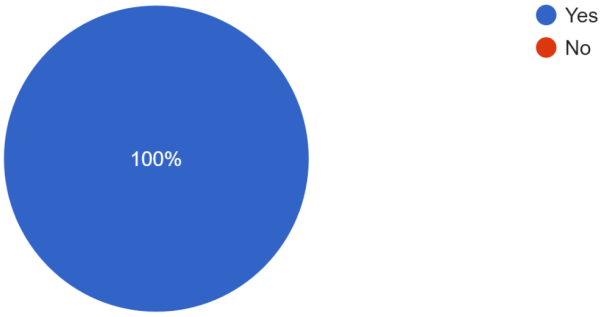


Figure 50: Result from post-study questionnaire

Question 9

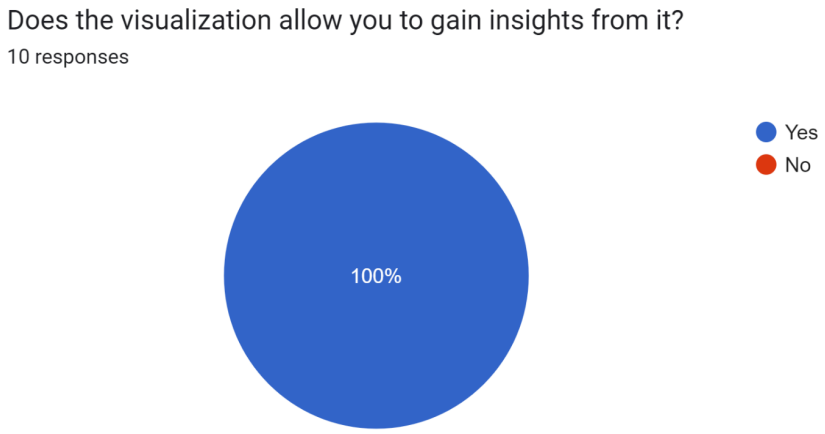


Figure 51: Result from post-study questionnaire

Question 10

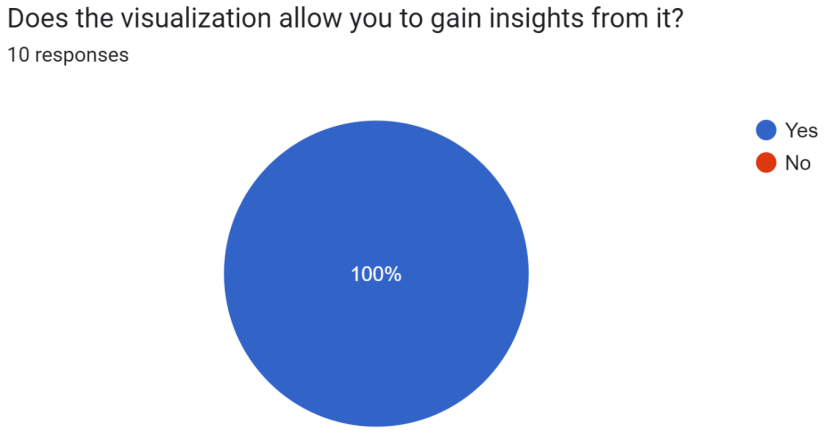


Figure 52: Result from post-study questionnaire

Question 11

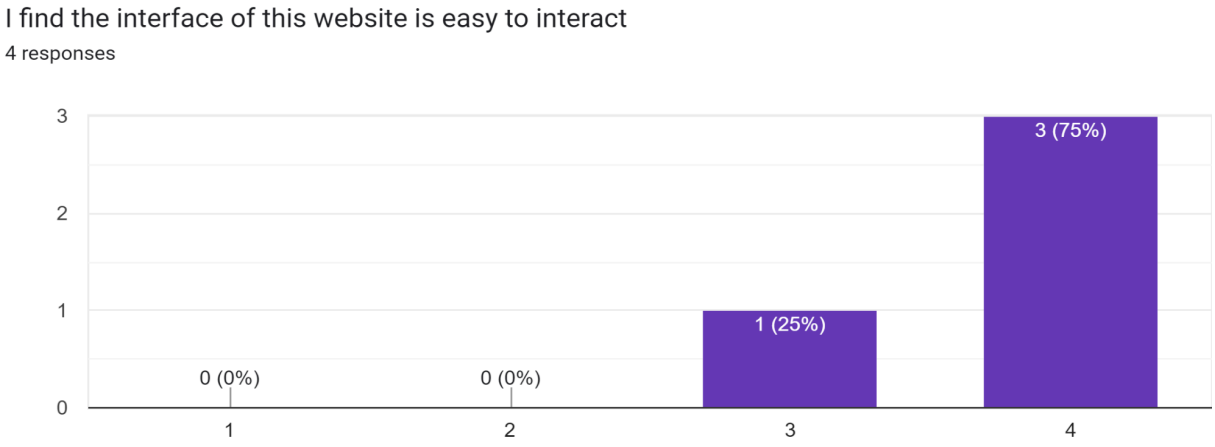


Figure 53: Result from post-study questionnaire

Question 12

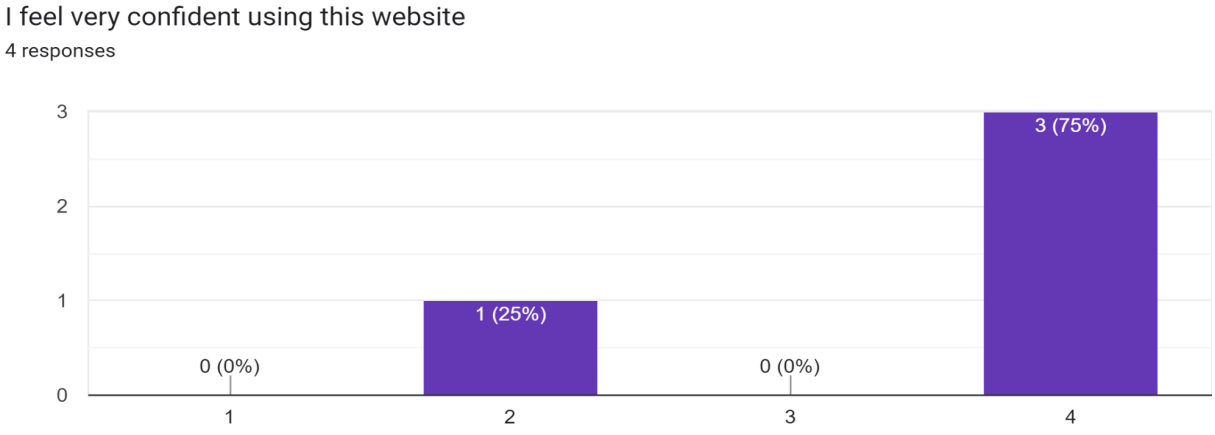


Figure 54: Result from post-study questionnaire

Question 13

Is there any feedback for the visualisation?

10 responses

Overall okay
It's not bad
no, it is presented clearly and nothing seems go wrong. Work perfectly fine
As said, I love how the visualizations in the project deliver valuable insights into how economic impacts on life expectancy from time to time. No error found in the charts, only what I found is the dedication and hard work committed by the developers <3
no
No
Great presentation
There could be room for improvement on the home page by combining home page and about page together.

Figure 55: Result from post-study questionnaire

APPENDIX C: Low-fidelity Design Validation

The objective of user validation for low-fidelity designs is to gather initial feedback on the basic structure, layout, and functionality of the visualization tools on our website. This stage is crucial for identifying major usability issues and ensuring that the design concept meets user needs before investing in high-fidelity prototypes. We have asked five users to give feedback on our sketches about their honest opinion on which is the best design when creating visualization. This testing helped us to design the best design for our visualization. The users have given a set of scenarios to navigate through the process and then provide feedback immediately of their idea.

Scenario

1. Use choropleth map to find the information about life expectancy across different countries
2. Use the bar chart to find the information about life expectancy across different countries
3. Use scatter plot to analyze the correlation between GDP and life expectancy across the continent
4. Use bubble chart to analyze the correlation between life expectancy, GDP and population across the continent

Result of the validation for choropleth map and bar chart

Participant ID	Choropleth Map	Bar Chart	Comment
1	✓		The map looks more attractive
2		✓	The bar chart is easier to see data and compare.
3	✓		Map is easier to see the whole data.

4		✓	It is easier to see data in a bar chart.
5	✓		Map is easier to navigate.

Table 6: Data collected through low-fidelity testing

Result of the validation for bubble chart and scatter plot

Participant ID	Bubble Chart	Scatter Plot	Comment
1	✓		The bubble chart shows more attribute than scatter plot
2	✓	✓	Both charts look about the same.
3	✓	✓	Both charts look the same.
4	✓		The bubble chart can display more data.
5	✓		The bubble chart is more interesting.

Table 7: Data collected through low-fidelity testing