# World Trends -

# Data Visualization Group 16

Lok Yin Wong, lowon23@student.sdu.dk
Sirintra Kunakornpaiboonsiri, sikun23@student.sdu.dk
Narit Chatchawanchokchai, nchat23@student.sdu.dk
Kasim Emre Sahin, kasah19@student.sdu.dk

## Abstract:

*The report examines world trends with the data provided by the United States Census Bureau, Numbeo and The World Bank which holds a multitude of data surrounding the 227 countries. Problem objectives in the report are made as central points in guiding the research into paths for creating data visualisations and further enlightening results over the various topics. The data visualisation is designed in such a way to adhere to the general principles and techniques outlined in the prior literature. Most of all, utilising data visualisation in creating streamlined graphs which communicate the message in a clear unbiased manner with both static and dynamic visual items.*
*Through the research, the group discovered several facts of the world regarding several topics such as fertility rate and infant mortality, gender imbalance in developing countries, living cost correlating with migration rate and more. Shedding light on these topics and concluding the world is trending in a positive direction.*

# 1. Background and Motivation

The aim of the project is examining countries and their corresponding changes through the years, thereby drawing relationships between the different variables within the dataset.
The aim of the project is examining the world trends and their corresponding changes through the years, granting knowledge of the different countries and their relationship with the various variables within the chosen dataset
The group intends to delve deep into the dataset which describes the developments of the world and unravel whether there are positive or negative trajectories depending on the chosen aspect. Most of all, the group is interested in researching several multifaceted aspects, such as examining the gender imbalance, living cost and migration rate and more. These are all key points of interest which could grant insight at the state of the world and its general trends through time.
In the end, this group will arrive at a paper which is intended for informing an audience of experts who seek further insight into the state of the world and the current positive or negative directions.

# 2. Project Objectives

For the purposes of the report, setting up clear problem objectives will be vital in guiding the project towards a purposeful goal. In this case, exploring the possible developments in the world, subregions and countries will be done by posing questions. The questions are listed below with no particular order:

- Is there a positive relationship between Fertility Rates and Infant Mortality Rate? When Fertility Rate is low it will consequently mean Infant Mortality rate will also be small? The aim of this question is gauging whether the variables will correspond to developing and developed nations.
- Is gender imbalance more serious in developing countries? What is the trend of gender disparity across different countries?
- Does the economical environment correlate with the migration rate? In other words, does a country with a lower living cost and a higher income level attract more immigrants?
- Do countries in the same income range and continent share similar groceries and renting index?
- Are there any changes in each variable when a country shifts from one income group to another?
- Inflation has risen steadily since 2010; however, after the pandemic in 2020, the inflation rose sharply resulting in the high cost of living. With the rise of the cost of living, it discourages many families from having kids(Demogr Res.2021). Therefore it is apt to examine the birth rate trend in each country.
- Population growth leads to a higher population density. According to research, it is said that Living in a high population density area gives a better access to various resources such as services, transportation, and healthcare which lead to a better living quality(Kolkowska and Roser 2023). However, simultaneously, densely populated areas can also lead to insufficient resources and increase in pollution (Yu and Gustafson 2022). So, is the higher the population density, the better the life expectancy? If not, what is the best population density?

With these problem objectives outlined the group can initiate their work in forming competent data visualisations to inform the current trends in the world. But first begins the data collection and processing part.

# 3. Data

## Data Collection

The data used are from 3 sources:

1. **International Database**

The data was acquisitioned through the The United States Census Bureau which collected information of 227 countries with various variables recorded where some were filtered out when choosing the dataset.
https://www.census.gov/data-tools/demo/idb/#/table?menu=tableViz

2. **Cost of living by city**

This is a data of cost of living and other related number in terms of index such as cost of living index, rent index and grocery index in each city
https://www.numbeo.com/cost-of-living/rankings.jsp

3. **World Bank Country and Lending Groups**

This dataset separates each country into each income range of high, upper middle, lower middle and low in each year.
https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

## Data Processing

Numbers in all data are converted to numeric type and column names are changed to appropriate names that can be consumed by R language. On cost of living by city, the data are grouped by country and then the average numbers of each country is used. Then all 3 dataset are merged using year and country as the keys.
For the Map Chart, data over the longitude and latitude points of each nation were collected from the library in R known as Tidyverse(tidyverse 2020). However, some country names did not correspond with the data collected from the The United States Census Bureau. Therefore, when the world map data is loaded in, the country names from the Tidyverse data will be modified to correspond with the data collected by the Census Bureau.

# 4. Visualisation/Dashboard

For the visualisation part of the report, the group members decided on using three main types of graphs for displaying the data, time series, map chart and scatter plot. Furthermore, the design of each graph will be elaborated on in order to justify their appearance and

communicative capability. The last part discusses how animation was incorporated into the visualisation and any optional features will be listed in the different sections.
Several requirements have been made for the visualisation, there must be at least three types of graphs, one animated and in total at least 8 graphs.

## 4.1 Link to Visualization/Dashboard

https://leolyw12138.shinyapps.io/DV_16/

## 4.2 Time Series plot

The time series plot shows data and its changes as the time passed by, with time represented on the x-axis and the variable of interest on the y-axis. In our dashboard, time series plot is used to show the trend of gender ratio in different countries. The time series help visualise and understand the temporal behaviour of the formation of population, making them valuable for analysing and decision-making on public policy.

## 4.3 Map Chart

Much of the things captured in the modern world includes a special competent which can be mapped. Whether one looks up a city address or takes a photo of a tree, both can be geocoded as points on a map (Dougherty and Ilyankou 2021).
However, it is important to ask oneself first whether the location matters to the story and problem objectives set out in the report. Just because data can be mapped does not mean it should be and where other charts should be considered. Yet, as stated in project's background and motivation, the group was interested in examining the world trends and developments which justifies the utilization of a map chart in granting more insight on the problem objectives. As with all good charts, a title and labels are a must for contextualizing the data visualization.
Furthermore, only one variable will be mapped as to not cause distraction and noise to the commutative aspect of the data visualization. By not overloading the map with information, the story becomes more streamlined and clearer with little obfuscation. Additionally, good choropleth maps must uphold correct geographic patterns which are clear to the readers, accessible in print black and white or displayed in color on a computer screen. The best maps are designed to be interpreted correctly by people with colorblindness. Therefore, a sequential palette will be chosen in showing low-to-high numeric values and which will be helpful in not confusing the viewer with a multitude of different color hues.
For the inputs, the map chart will have one which will determine which variable such as example "Fertility Rate" and the other input is the year.

## 4.4 Scatter Plot

The scatter plot is used to visualise a data point in two dimensions that represent the value in each axis. It is capable of illustrating the relationship between the two variables. Furthermore, the graph can also visualise the third variable in the graph by categorising each datapoint using different colour or shape. Adding features like trend line or tooltip can make the graph easier to read the trend and examine other information of the data point.

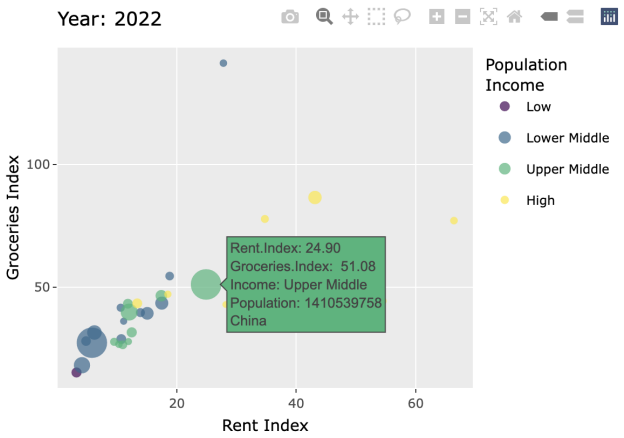



Fig 1: Example of tooltip in the scatter plot

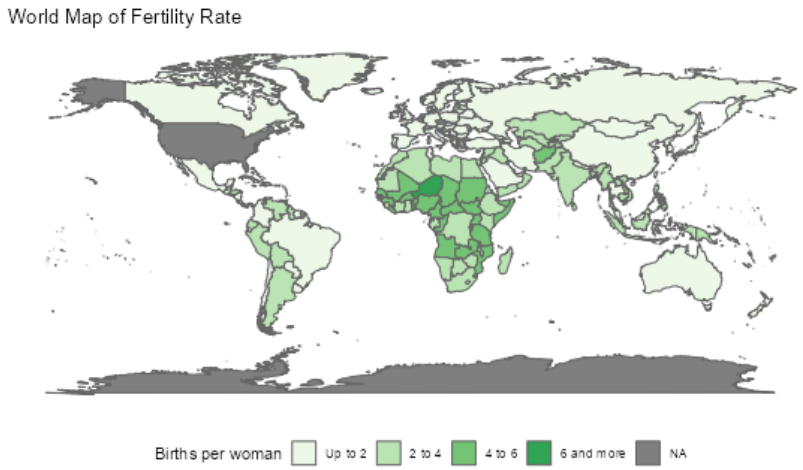
## 4.5 Animated

Our dashboard employs gganimate to generate an animated scatter plot. The primary focus is on visualizing the relationship between two variables over time, with each point representing a specific year. The plot is enhanced with animation features such as a transition effect based on the "Year" variable, shadow effects (shadow_mark), and entry animations (enter_grow and enter_fade). The renderImage function is used to create and display the resulting animated GIF. The temporary file ("outfile.gif") is generated and then saved using anim_save. Animated plots provide a dynamic and interactive way to visualize data trends and patterns over time. The animation engages users and allows them to observe changes as they unfold.

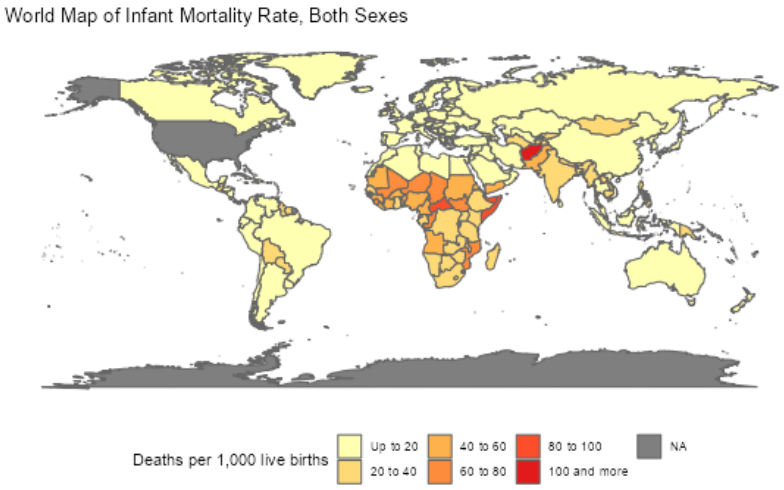
# 5. Story/Results

## 5.1 Fertility Rate and Infant Mortality

To answer the question, if there is a positive relationship between “Fertility Rate” and “Infant Mortality Rate”, a map chart was utilized in grasping the knowledge of the world in 2022. First by creating two map charts for the two variables, fertility rate and infant mortality.




(a)




(b)

*Figure 2: Two different world charts over the two different variables, "Fertility Rate" for (a) and "Infant Mortality Rate" for (b).*

As displayed in Figure 2, one can see a similarity between the two variables as countries with high fertility will generally also have a higher degree of infant mortality rate for both sexes. An overlap can be made of developing and developed countries, as the central region with a high amount in fertility rate and infant mortality is in the central parts of africa. Where Somalia has the highest degree of infant mortality rates in the continent. The other notable country in the world is Afghanistan which has a "100 and more" infant mortality rate than Somalia.

In developed countries, women are more likely to be engaged within the job market, thereby opening a multitude of options for women beyond the traditional role as a mother.

Furthermore, through the ages the general “Fertility Rate” and “Infant Mortality Rate” is dropping over the years showing a general positive trend.

## 5.2 Is gender imbalance more serious in developing countries?

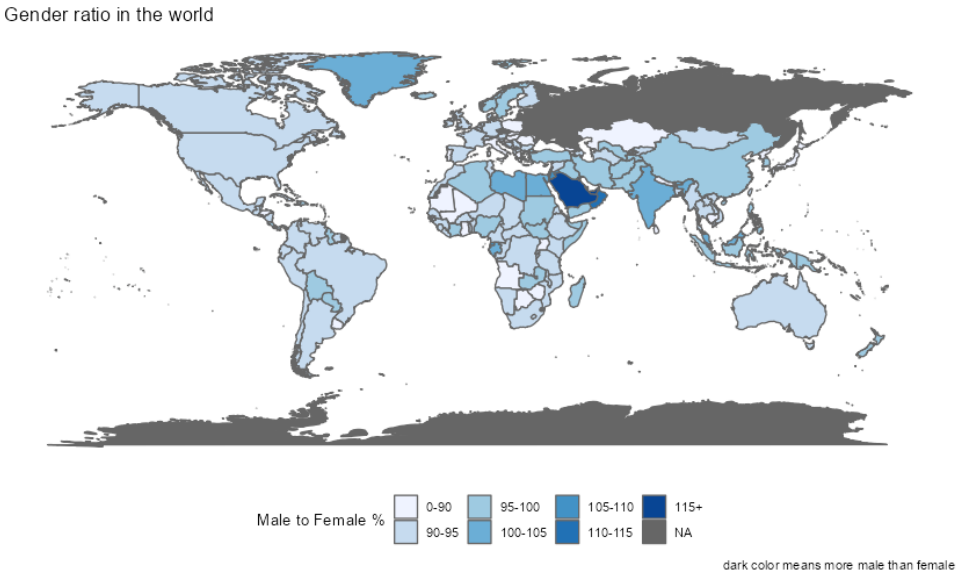
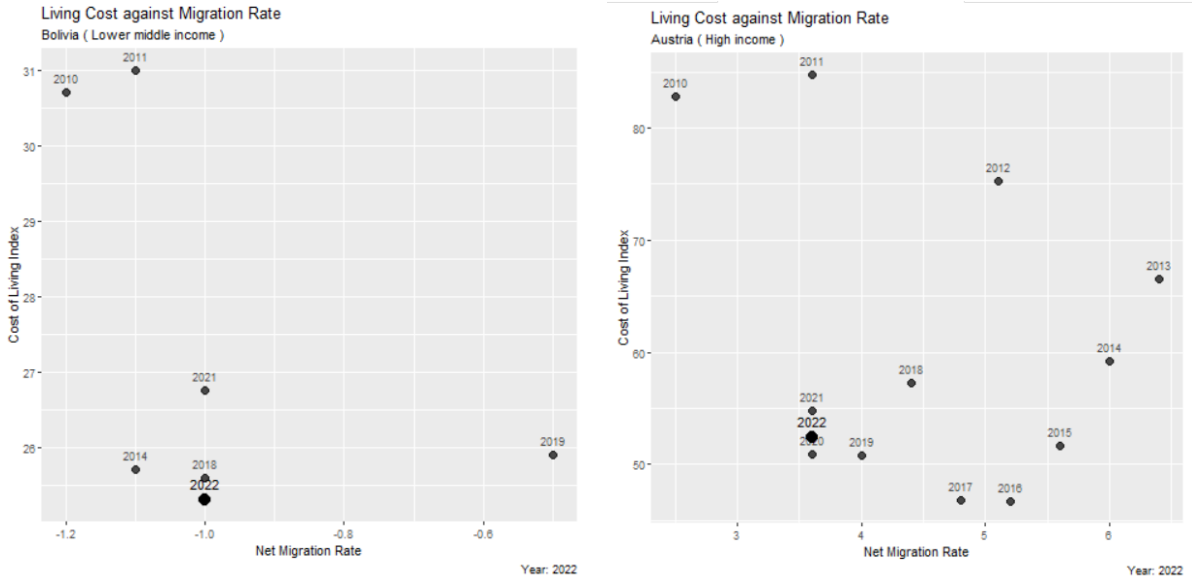



*Figure 3: World map of gender ratio in year 2016*

This is a world map showing the male to female percentage across the globe. From the map, the majority of countries have a relatively balanced gender distribution, with a male to female percentage ranging between 90 and 100, indicating a slight female prevalence. Northern Africa and the Middle East, on the other hand, show a distinct pattern, with a higher concentration of males, falling within the 100-115 range. Notably, Saudi Arabia has a particularly pronounced gender imbalance, with males outnumbering females by a factor of more than 115. Interestingly, most developed countries concentrated in America and Europe demonstrate a more equitable gender ratio, suggesting a balanced distribution between males and females. A noteworthy trend that emerges from the analysis is a general decline in the male to female ratio in high income countries, indicating a tendency towards a more balanced distribution. Even though there is also a tendency of prevalence of male in developing countries in the africa, the variation is not immediately significant. The exception here is Saudi Arabic, where the male to female ratios has increased by 4% up to 130%. This insight motivates further research into the factors influencing gender imbalances, such as cultural dynamics, social policies, and economic development, in order to better understand and address disparities around the world.

## 5.3 Does the economical environment correlate with the migration rate?




*(a)* *(b)*

*Figure 4: Two scatter charts on living cost index/migration rate of two countries with different income level,"Bolivia (lower middle income)" for (a) and "Austria (high income)" for (b).*

Comparing these two charts of countries with different economic statuses, we can observe that even the living cost (less than 31) in Bolivia is much lower than that of Austria (higher than 45), the migration rate stays negative in Bolivia in contrast of the positive migration rate in Austria. There is a hint of an inverse-proportional pattern between living cost index the migration rate. For example, in the year of 2010 and 2011, the living cost indices of both countries are in local high, while the migration rates are in local low. But focusing on the point where the migration rate is high, the cost of living is not directly correlate with the migration rate. It shows that the living cost of a country is not the primary deciding factor for migrants deciding whether to relocate to the country for a better life. Rather, higher-income areas may attract individuals seeking better economic opportunities and higher living standards. This can contribute to a positive migration rate in regions with robust economies as shown with the aid of the graphs. In terms of design, scatter plot is used to show the relationship of two continuous data on two axes. Since there are many countries with the same categorical income level, it is unwise to show all countries in the same graph which can lead to visual overload. To keep it simple, one chart is representing one country at a time. It is then animated to help visualise the trend of migration rate.

## 5.5 Do countries in the same income range and continent share similar groceries and renting index?

Yes to a certain extent, countries in the same income range and continent share similar groceries and renting index but the difference between each income range cannot be separated by groceries and renting index alone.

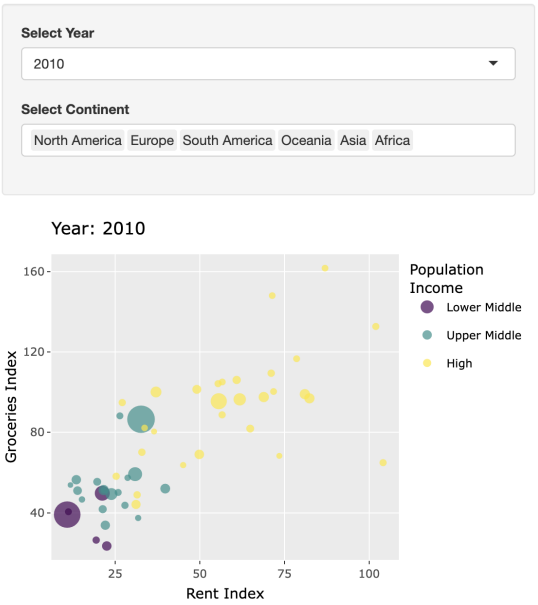



Fig 5: Scatter plot of Grocery Index against Rent Index of every continent in 2010

Overall, as can be seen on the graph above, in 2010, there are similarities in each income group except for only some countries such as Poland, Portugal and Hungary which are in high income groups but have a lower groceries and rent index compared to other counties in the same group. Or Indonesia, which is in a Lower Middle Income group but has a slightly higher grocery and rent index compared to other counties in the same group as can be seen on the picture below .

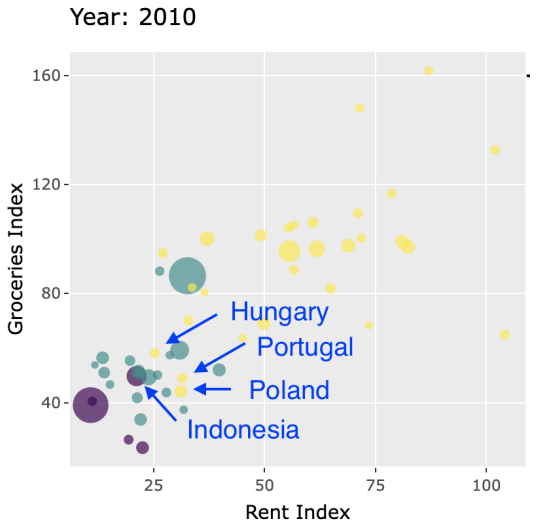


Fig 6: Groceries and Rent Index of Hungary, Portugal, Poland and Indonesia in 2010

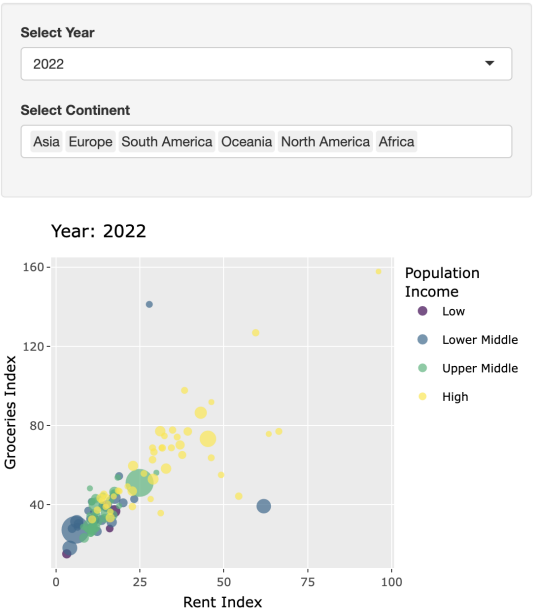



Fig 7: Scatter plot of Grocery Index against Rent Index of every continent in 2022

In 2022, the similarities are not as clear as in 2010, since the dots above overlapped a lot. When looking at each continent, Europe for example, the similarity is better, as can be seen in the picture below. Most countries are in the High income group. The countries in the lower income group have lower Groceries and Rent Index.

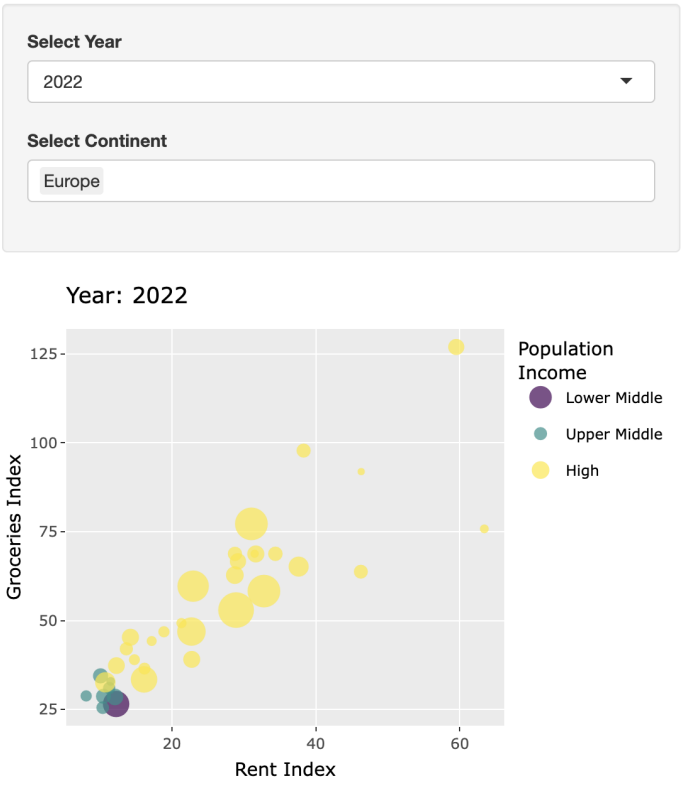



Fig 8: Scatter plot of Grocery Index against Rent Index of Europe continent in 2022

But when looking at Asia, the similarities are not as clear as in Europe since all three income groups of High, Upper Middle and Lower Middle share similar groceries and rent index.

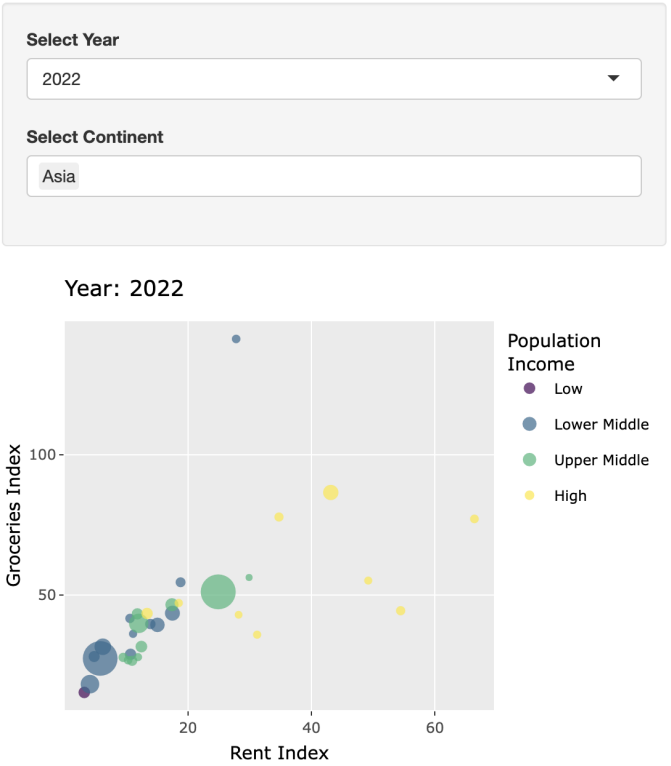



Fig 9: Scatter plot of Grocery Index against Rent Index of Europe continent in 2022

## 5.6 Are there any changes in each variable when a country shifts from one income group to another?

There is a change but it is not from the income range shift since the graph fluctuates and the shift from one group to another doesn't happen at the same time.

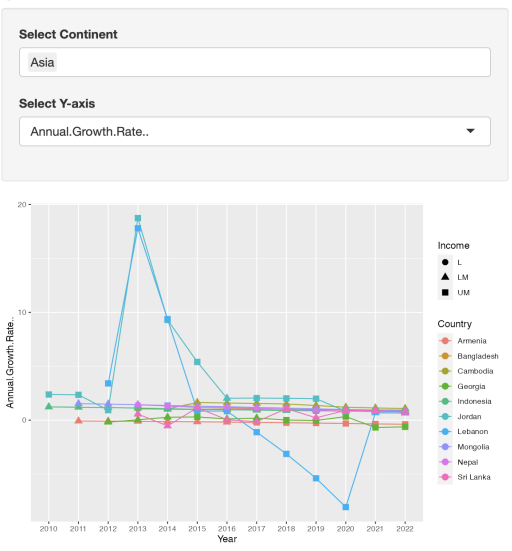



Fig 10: Graph showing Annual Growth Rate in Asian Countries where there is a shift in income range

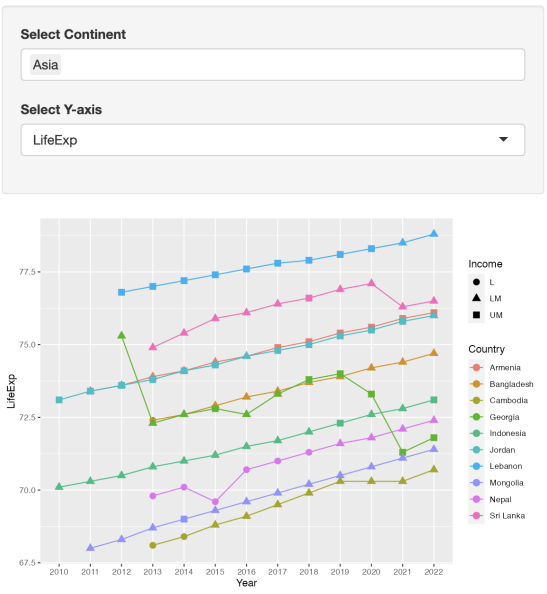



Fig 11: Graph showing Live Expectancy in Asian Countries where there is a shift in income range

Except in some cases, for example the one below, looking at Moldova, when the country shifts from triangle (Lower Middle Group) to rectangle (Upper Middle Group) the Net Migration Rate spikes up.

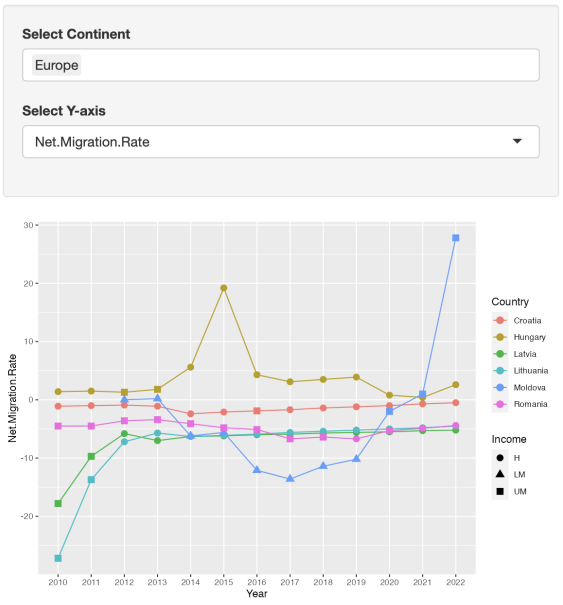



Fig 12: Graph showing Net Migration Rate in European Countries where there is a shift in income range

## 5.7. Is the higher the population density, the better the life expectancy?

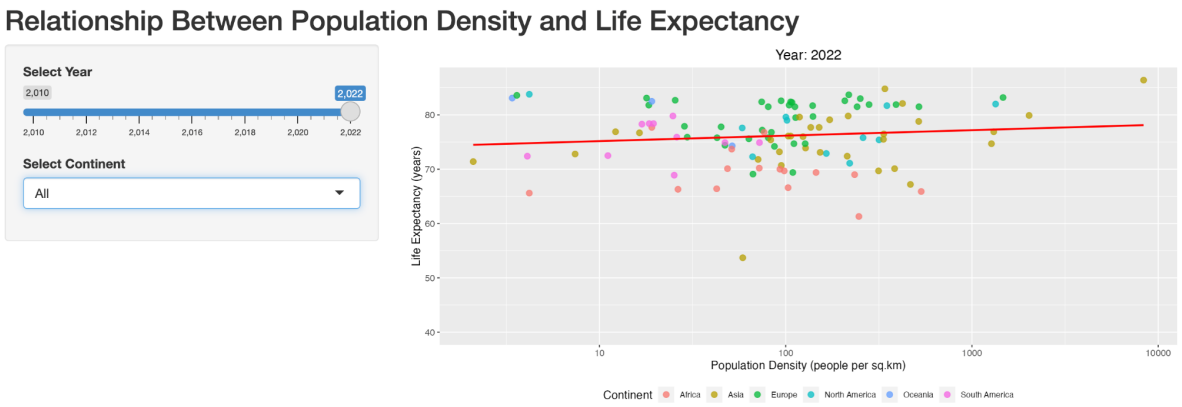



*Fig13: Relationship Between Population Density and Life Expectancy in 2022*

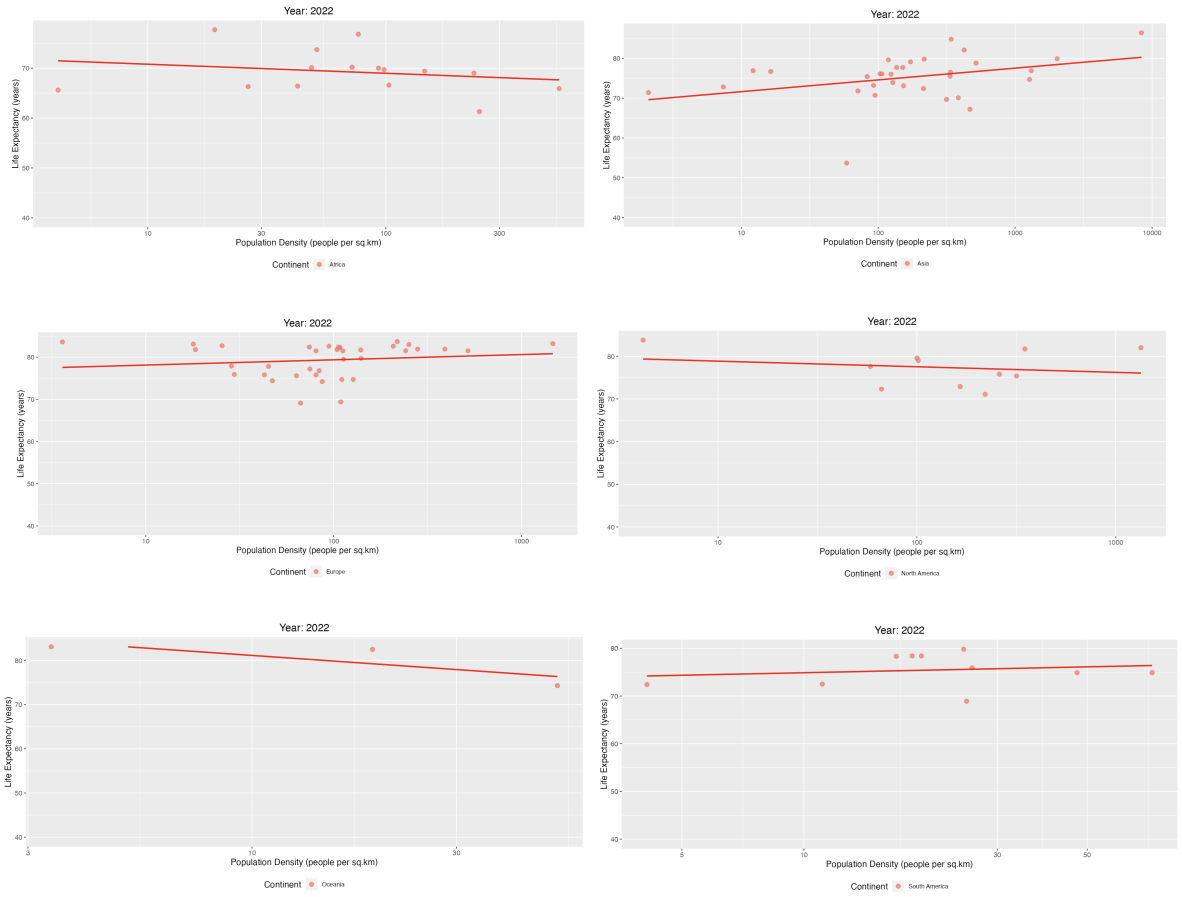



*Fig14: Relationship Between Population Density and Life Expectancy in each Continent in 2022*

According to the graph displaying a relationship between population density and life expectancy in Fig13, it is observed that countries with population density less than 1,000 people per square kilometre do not have significant relationship between life expectancy and population density. However, when considering countries that have population density over

1,000 people per square kilometre, they all have life expectancy above the average with Monaco and Singapore as the first and second highest in both population density and life expectancy. Furthermore, when the trend line is drawn, it can be observed that countries with higher population density tend to have higher life expectancy.

Nonetheless, population density is calculated using the number of population and the total area of that country instead of inhabited area which might lead to bias. Based on the assumption that countries in the same continent have a similar ratio of inhabited area, the relationship between population density and life expectancy in each continent is examined and illustrated in Fig14. As displayed in the graph, some continents with low population density, for instance, Africa, North America, and Oceania, have less life expectancy when the population density is higher. On the other hand, continents with high population density, like Asia, have a strong positive relationship between population density and life expectancy.

In conclusion, population density and life expectancy has a slightly positive relationship. Countries with higher population density are expected to have better life expectancy.

## 5.8 What is the birth rate trend in each continent and income group?

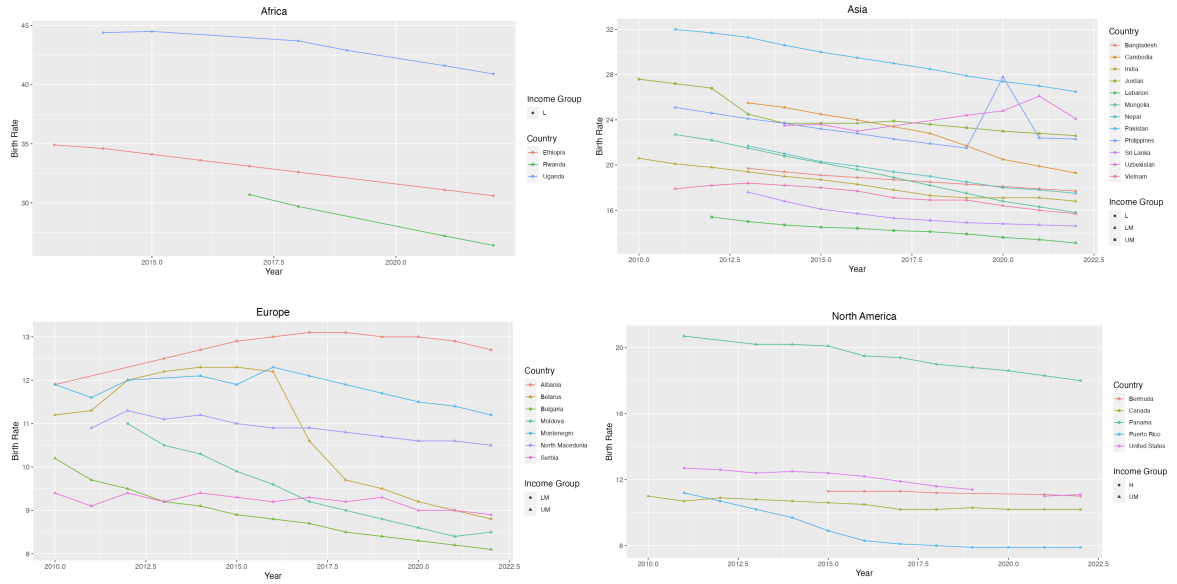



*Fig a. Birth Rate Trend of Low Class in Africa*

*Fig b. Birth Rate Trend of Lower Middle Class in Asia*

*Fig c. Birth Rate Trend of Upper Middle Class in Europe*

*Fig d. Birth Rate Trend of High Class in North America*

Since 2010, the world has had inflation resulting in the rise in cost of living which affects the birth rate in several countries. The visualisation depicts the birth rate trend from

2010 to 2022 in each continent and income class. The graphs show that the birth rate trend in almost every country is decreasing regardless of income group and continents.

# 6. Conclusion/Discussion

The report concludes the general progress of the world seems to be going towards a positive direction, which was arrived from the problem objectives stated and answered through data visulizaitons. Most of all, through considerations and design choices of the graphs, the group gained an understanding on the proper procedures and guidelines for mapping data onto graphs including the best practices for conveying a story. Then from the developed data visulizaitons, arriving at results for each of the problem objectives and discovering several facets in the world. Such as Fertility Rate and Infant Mortality Rate having a positive relationship, having a larger population density will predicate a higher life expectancy and more. Creating the graphs help enlighten the group not only in discovering new information but also going against the group member's biassed beliefs in some areas.

Yet, there is room for improvement regarding the general structure of the dashboard as an integration of the current report could provide a better method in conveying the goal and the results. Furthermore, for the map chart an additional tool-tip could be added in to make it more interactive. Where a user can hover over a specific nation providing a box over additional information. Though a better suggestion would be having an even more in-depth overview of a specific country. Such as when the user clicks on the United States of America, the world chart will change to focus on the country displaying the various states with their own gradients corresponding to the chosen value. Thereby, unlocking further potential within the world map chart and covering deeper aspects.

The group had also difficulty in coordinating with each other and implementing changes together as some did not have experience with git. In turn rendering more difficulty in orchestrating changes.

# Contribution:

| **Section:** | **Main Contributor:** |
|---|---|
| Abstract | Kasim |
| Background and Motivation | Kasim |
| Project Objectives | Kasim |
| Data - Data Collection | Sirintra |

| | |
|---|---|
| Data - Data Processing | Sirintra, Leo |
| Visualisation/Dashboard - Time Series Plot | Leo,<br>Sirintra, Narit |
| Visualisation/Dashboard - Map Chart | Kasim, Leo |
| Visualisation/Dashboard - Scatter Plot | Narit,<br>Sirintra, Leo |
| Visualisation/Dashboard - Animated | Leo |
| Story/Results -<br>Fertility Rate and Infant Mortality Rate | Kasim |
| Story/Results -<br>Is gender imbalance more serious in developing countries? | Leo |
| Story/Results -<br>Does the living cost correlate with the migration rate? | Leo |
| Story/Results -<br>Income range and groceries and renting index | Sirintra |
| Story/Results -<br>A country shifting from one income group to another | Sirintra |
| Story/Results -<br>Population Density and Life expectancy | Keen |
| Story/Results -<br>What is the birth rate trend in each continent and income group? | Keen |
| Conclusion/Discussion | Kasim |
| Code management and deployment | Leo |

References

Dougherty, Jack, and Ilya Ilyankou. 2021. “Hands-On Data Visualization.” handsondataviz. https://handsondataviz.org/.

Kolkowska, Natalia, and Max Roser. 2023. “High Population Density: Pros and Cons — Sustainable Review.” Sustainable Review. https://sustainablereview.com/high-population-density-pros-and-cons/.

studysmarter. 2020. “Rapid Population Growth: Overview and Causes.” StudySmarter. https://www.studysmarter.co.uk/explanations/geography/changing-economic-world/rapid-population-growth/.

tidyverse. 2020. “Create a data frame of map data — map_data.” ggplot2. https://ggplot2.tidyverse.org/reference/map_data.html.

Yu, Jessica, and Paul Gustafson. 2022. “Assessing Trade-Offs and Optimal Ranges of Density for Life Expectancy and 12 Causes of Mortality in Metro Vancouver, Canada, 1990–2016.” NCBI. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8910136/.

Alboukadel. “gganimate: How to Create Plots with Beautiful Animation in R” https://www.datanovia.com/en/blog/gganimate-how-to-create-plots-with-beautiful-animation-in-r/

# Appendix:

## Github Link:

https://github.com/LeoLYW12138/DV_16