CS/INFO 3300 Project 2

Fedinard Nyanyo (fen7), Stephan Volynets (svv6), Merry Zebro (mz289), Justin Wong (jsw345)

This project visualizes global life expectancy and population over time. This allows users to see changes in life expectancy and population in different countries from 1960 to 2022. By exploring these visualizations, users can observe how life expectancy and population size have changed across different countries in the past six decades.

Data Description:

- Life Expectancy Data: Annual life expectancy for each country from 1960 to 2022. Each entry corresponds to a country and provides data for each year within the range. The dataset for this CSV file is found on the World Bank Group website.
- Population Data: Population figures were collected in parallel with life expectancy for a similar time frame, allowing for a comparative analysis between life expectancy and population growth across different regions. The dataset for this CSV file is found on Our World In Data
- Geospatial Data: a world map representation. Dataset for the TopoJSON file found on GitHub.
- combined_data.json: a TopoJSON file representing the world map (excluding Antarctica), with life expectancy and population data for each year from 1960 to 2022 associated with each country. This data is stored in the "properties" dictionary for each country. The TopoJSON file also has a list of every possible life expectancy value to make it easier to create a color scale. Lastly, the file has values storing the maximum and minimum life expectancy to make it easier to find the domain for life expectancy. We describe how we processed the above three datasets into combined_data.json below.

Data Processing:

The processLifeExpData.js and processPopulationData.js combine life expectancy and population data respectively into the TopoJSON file representing the world map.

processLifeExpData.js - First, we create a function called "inputNewMapData" that takes a country's name, that country's index in the life expectancy CSV, and that country's index in the TopoJSON file. The function then takes data from the life expectancy CSV and creates a dictionary where each key is a year and the value is the life expectancy of that country for that year. Then, the function puts this dictionary into the "properties" variable of the corresponding country in the TopoJSON file. Also, during the creation of the dictionary, the function adds each life expectancy to a list and keeps track of the maximum and minimum life expectancy values.

After creating this function, we go through every country in the CSV and TopoJSON file and if the country name matches between the two datasets, we call the “inputNewMapData” function we just created. The big caveat we had to account for when matching the names of countries was that the same country was named differently between the datasets. For instance,

- TopoJSON = Democratic Republic of the Congo; CSV = Congo, Dem. Rep.
- TopoJSON = Côte d'Ivoire; CSV = Cote d'Ivoire
- TopoJSON =Virgin Islands (U.S.); CSV = U.S. Virgin Is.

Then, we store the list of every recorded life expectancy, the minimum life expectancy, and the maximum life expectancy in their respective variables in the TopoJSON file. Lastly, we stringify the modified TopoJSON file and print out the result to the console, so we can copy the result into combined_data.json. We chose to exclude Antarctica from the TopoJSON file as there was no life expectancy or population data associated with it, thus it took up unnecessary space in the canvas.

processPopulationData.js - makes sure that population data is aligned with life expectancy data across countries and years in the TopoJSON file. Here we load the TopoJSON file and population_data.csv. We begin by creating a populationLookup dictionary to map each country’s name to its row in the CSV for quick access. For the country name mismatches between the datasets, “countryNameMap” fixes these inconsistencies, it matches names like "Taiwan" in TopoJSON to "Taiwan, China" in the CSV. The “inputNewMapData” function then combines life expectancy and population data into each country’s properties. If both data types are available for the years 1960 to 2022, a dictionary with both life and population values is created under that year in the “properties” value for that country.

Then we have the updated TopoJSON output as a JSON string, which holds life expectancy and population data for each year. This combination allows us to access the different properties of each country and year.

Visual Design Rationale:

Map Visualization of Life Expectancy and Population

- For the life expectancy and population map visualization, we decided on a map visualization because it seemed the most intuitive to represent data associated with many countries. We initially thought to use a bar graph, but the bars were way too cluttered and it was even harder to show what country each bar represented, even with mouseover events. We chose a Mercator representation for our visualization because, although the land area is not accurate, it’s a map style that is very common and easily understood by most. Thus, we decided to use it as it looked better than other maps where the shapes of countries were more distorted

- To visually show life expectancy population data, we decided that we would use colors to represent each dataset, with varying colors representing varying values for life expectancy or population number. For both datasets we wanted to use that was intuitive, easy to distinguish, and accessible. For life expectancy, it seemed intuitive to use green to represent countries with a higher life expectancy and to, thus, use red to represent countries with a lower life expectancy as it is green's opposite on the color wheel. For population data, we wanted to use a different color scale. So, we decided to use an intense color like red to represent heavily populated countries and a calmer color like blue to represent sparsely populated countries. For both of these scales, we had to vary hue because if only saturation and/or luminosity were varied, it was hard to tell the difference between countries with similar colors, especially if the countries were on the smaller side.
- To help the viewer understand what each color meant for life expectancy or population, we created a color legend. We provided tick marks to help the viewer understand the thresholds of each color, and we provided text to help the viewer get a general understanding of what the colors mean. For the color legend associated with life expectancy, we used a linear scale as it was easy to see the associated color thresholds. Also, it was intuitive as the amount of space between the axis ticks makes proportional sense to the viewer. For example, the amount of space between the first and second tick (12 to 56) is much larger than the amount of space between the second and third (56 to 65) tick, which makes visual sense.
- For the color legend associated with the population, we used a log scale. We decided on this, because if a linear scale was used, every color representing a population threshold below 20 million was not visible at all on the legend, and the axis ticks below the legend were all smushed to the left. However, this comes with a tradeoff as the spaces between the axis ticks aren't as intuitive to the viewer as a linear scale. For instance, the amount of space between the first and second ticks (2,646 and 419,224, respectively) is visually identical to the amount of space between the second to last and last tick (20 million to 1 billion, respectively). In other words, the numerical difference between the first and second tick is 416,578, while the numerical difference between the second to last and last tick is 980,000,000. One might think the spacing for the latter should be several magnitudes larger than the former, but this is not the case when using a log scale.
- We settled on the current size of the canvas and map for the following reasons. If the canvas was too large, you would have to scroll too much between using the slider and buttons versus seeing the whole map. If the canvas was too small, it was hard to see and identify smaller countries, even with the functionality to zoom.
- We purposely positioned the slider, buttons, and other information above the map. We did this because, although these features may be out of the user's perspective if they are viewing the whole map and the slider is moving, we provided the functionality to press space so that the user can stop the slider's movement Thus, if the user is viewing the map

and the slider is moving on its own, they can press space to pause on a year of interest and then scroll up to use whatever feature they desire.
- To help the reader interpret the features above the map we provided a box to show the current year so the user knows exactly what year they are looking at. We also added tick marks below the slider to make it easier for the user to see what years they can slide to. Each tick mark represents a single year and every 5th year/tick has an associated year label.

- Marks
    - Shapes of countries
- Visual channel(s)
    - Color (hue specifically)

<u>Scatterplot Visualization of Life Expectancy vs Population</u>

- The scatterplot visualization was designed to effectively show the correlation between life expectancy and population for each country from 1960 to 2022. Each circle in the plot represents a country, with life expectancy plotted on the y-axis and population on the horizontal axis.
- We chose a scatterplot for this data because it allows users to easily compare life expectancy against population size on an individual country level, making trends and outliers visually distinct. Circles as markers effectively show differences in population size, as each circle's size is scaled according to the population of that country the circle is representing. This choice of circle size makes it easy to observe relative population sizes without it looking overwhelming.
- The population axis is log-scaled because of the wide range of population sizes in different countries. This choice allows us to have a balanced view that allows smaller countries to be visible while preventing larger countries from dominating the graph. For life expectancy, we decided on having it be a linear scale, which allows the highlighting of increases over time while maintaining intuitive proportional spacing between values.
- We assign the colors to the countries using a categorical color scale. The colors enhance the visual separation between the countries, this makes it easier for users to track each country's data throughout the years.
- The supporting design elements include gridlines on both axes, which makes it easier to read and creates a more precise-looking visual comparison between countries. The y-axis is labeled "Life Expectancy (years)," and the x-axis is labeled "Population Size," both labels help users easily understand the data being compared. There is also a size legend to show population sizes for different circle sizes, this gives users more information to be able to accurately interpret the visual they are seeing.

- Marks
    - Circles represent each country on the scatterplot
- Visual channel(s)

    - Color (hue) unique to each country
    - Aligned vertical and horizontal position of circles shows the correlation between population and life expectancy
    - Size of circle for population

Interactive Elements and their Design Rationale:

Map Visualization of Life Expectancy and Population

- Buttons to show life expectancy or population
    - We created these buttons because we wanted a way to show life expectancy and population data without making two entirely different visualizations.
    - To show life expectancy data, the user can press the “Show Life Expectancy” button. To show population data, the user can press the “Show Population” button. Both of these buttons are displayed below the list of instructions at the top of the page.
- Textbox to input a valid year; if an invalid year is inputted, nothing happens
    - We wanted a way for the user to directly choose a year of interest, rather than forcing them to use the slider, thus we added this option. This is especially useful if the user's browser is laggy and they are experiencing latency issues with the slider.
    - This textbox is visible above the textbox showing the current year. At the top of the page, we also provide instructions on how to input a valid year.
- Slider to select the year
    - We wanted a way for the user to see the total range of years they can select and be able to choose a specific year along that range, thus we decided to add a slider. The slider also allows the user to better follow the progression of years, especially when the animation that cycles through each year is active. We were split between putting the slider right above the map horizontally or to the side of the map vertically. In both cases, the user can see most of the map when they are using the slider. However, we decided on the former because we saw that having the slider positioned vertically made it awkward to use the other features provided above the map visualization. Thus, we grouped the slider with the rest of the features above the map visualization because having every feature to the side of the map made less sense.
    - This slider (with its associated axis) is very visible right above the map visualization. To reiterate, the slider was purposely put there so the user can see most of the map when they are using the slider.
- Play and stop button to control if the slider moves (in increasing order)

    - We wanted a way for the user to see the data as the years progress, so we added the ability for the slider to move on its own. When the slider moves autonomously, the year increases by 1 every two seconds, and the map changes to represent the current year. For the user to control if the slider is moving, we added a play and stop button.
    - These buttons are visible above the year slider (or below the text box telling you the current year). Also, this control functionality is made apparent through the provided instructions at the top of the page.
- Press space to start and stop the animation
    - On our site, the user may not be able to see every feature and the entire map at the same time. Thus, if the features are out of the user's perspective when they are viewing the whole map and the slider is moving on its own, we provided the functionality to press space so that the user can stop the slider's movement. Thus, if the user is viewing the map, they can press space to pause on a year of interest and then scroll up to use whatever feature they desire.
    - This functionality is made apparent through the provided instructions at the top of the page.
- Can hover over countries to see life expectancy and population
    - We wanted a way for the user to see the exact life expectancy or population number associated with each country. Thus, we added mouseover events for each country. When you mouse over a specific country, a text box on the bottom left of the canvas shows the country's name and its exact life expectancy or population number. Also, when you hover over a country, its outline is highlighted black to explicitly show the viewer what country is being hovered over.
    - This hover functionality is made apparent through the provided instructions at the top of the page.
- Can zoom in and drag the map
    - We wanted a way for the user to zoom into the smaller countries that are harder to see, as well as move the map around when they are zoomed in. Thus, we added the ability for the user to pan and zoom within the map visualization. We restricted the panning so that the user cannot completely pan away from the map, such that it is no longer visible. We restricted the zoom so that the user could zoom in just enough to easily select the Vatican and zoom out just enough to see every country on the world map.
    - This pan and zoom functionality is made apparent through the provided instructions at the top of the page.
- Can zoom in on individual countries by clicking on them
    - We wanted a way for users to easily zoom in on specific countries of interest, so we added the ability for them to click on a country to zoom in on it. However, there are some countries where this zooming function doesn't work as well

because their territories are too spread out. For instance, if you try to zoom in on Russia on the top right of the map, you barely zoom in at all, because a small portion of Russia is also on the top left of the map. To accurately show a country and all its territories, we chose to keep the current implementation to not exclude any (smaller) territories associated with a country.
    - This functionality is made apparent through the provided instructions at the top of the page.
- Reset zoom
    - After zooming in, we wanted a way for the user to easily zoom back out to the initial view of the world map (when the page was initially loaded). Thus, we added a button to reset the view of the map.
    - This functionality is made apparent by the button with the text “Reset Zoom” that is below the instructions at the top of the page

<u>Scatterplot Visualization of Life Expectancy vs Population</u>

- Year Slider
    - This allows users to select a specific year or slide over the years to see the changes over time in both life expectancy and population for each country. It is positioned so users can easily see it without taking away from the main visualization. As users move the slider, the visualization updates which allows users to see global shifts in an easy-to-understand manner. There is also the year that updates to show the selected year, this helps users stay oriented as they adjust the year slider to observe changes over time.
- Tooltip on hover
    - When users hover over each of the circles, a tooltip is displayed with information about the country that the circle represents. The name, life expectancy, population size, and year are shown. This makes it much easier for users to focus on data points without feeling overwhelmed by detailed information on everything all at once. The selected circle’s opacity increases and is outlined in black which allows users to correctly and easily select the circle they intended to pick on.
- Circle sizing
    - The circle sizing that represents the population changes as the year slider is used and as a change in population is detected in the countries. This is a smooth transition that is noticeable but does not overwhelm and take away from the overview look.

The Story

<u>Map Visualization of Life Expectancy and Population</u>

The Map Visualization shows us how each country compares to others in terms of life expectancy and population. Moving the year slider, you can choose to see data for population change over the years or life expectancy change over the years. We can see a general increase in the world's life expectancy, with shades of green increasingly dominating the map. Today, regions like parts of Africa, Afghanistan, and Bolivia still show relatively low life expectancy, but even in these areas, the life expectancy trend is moving upward. However, you can also see how life expectancy is affected by tragic events such as the Cambodian Genocide from 1975 to 1979 where the life expectancy drops to 12 years, the lowest life expectancy value in our dataset. If we click on "Show Population", we see a similar upward trend where the world has had an increase in overall population with countries turning red (high population) as the year increases.

Scatterplot Visualization of Life Expectancy vs Population

Similarly, in the scatterplot, we see all countries following a similar trend, moving upward along the y-axis, which represents an increase in life expectancy. The population size of each country also sees this growth trend. The larger the population gets, the larger the circles get. We wanted to see whether there was a correlation between a country's life expectancy and its population size, and we tracked this over time. We found that both highly populated and smaller countries tend to increase at a similar pace in terms of life expectancy.

Notably, China, which has the largest population, started with one of the lowest life expectancies, but as we move the year slider, we see it progress toward the upper range of life expectancy. Another thing that stood out was as every country's life expectancy was rising, there was a quick major dip in Cambodia from a life expectancy of 40 to the age of 12 in 1976. We found that this was because of the Genocide the country was going through.

This visualization captures both gradual improvements in life expectancy, the complexities of population growth, and the impact of certain historical events on both population and life expectancy.

Team Contributions

- Ferdinand Nyanyo (fen7)
    - Preprocessed TopoJSON file by merging map data and and population data (1 hour)
    - Created the scatterplot visualization (4 hours)
- Stephan Volynets (svv6)
    - Researched datasets (30 min)
    - Cleaned up and polished the scatterplot visualization (1 hour)
    - Minor Visual Fixes (1 hour)
- Merry Zebro (mz289)

    - Set up communication channel
    - Found population data from the website: Our World In Data (1 hour)
    - Worked on write-up (2 hour)
- Justin Wong (jsw345)
    - Researched datasets (1 hour)
        - Found life expectancy data on the World Bank Group website (30 min)
        - Found map data on GitHub (30 min)
    - Preprocessed TopoJSON file by merging map data and and population data (2 hours)
    - Created the map data visualization (10 hours)
    - Worked on write-up (3 hours)