**Title**: Understanding of World Population

**Member**: Alex Kimball (u1195729)

Faisal Karim (u1418923) [faisal.karim@utah.edu](mailto:faisal.karim@utah.edu)

**Github repository**: <https://github.com/Kimballa05/DataVisGlobalWarming>

**Motivation and Background**: We want to create a spike map of the world with population. Population of male and female. The spike map will also have data like birth rate, death rate, child mortality. These options will all be interactive. Like when you chose one option it will be shown in the spike map. We are hoping that this might give people some idea about other countries’ populations.

**Data**: We collected data from [OurWorldInData](https://ourworldindata.org/population-growth#:~:text=It%20was%20only%20a%20century,number%20of%20people%20ever%20born.). It will require some sorting and cleaning. For example, these datasets don’t contain data of latitude and longitude of the countries. We are planning to add it there. Also, I am not yet sure of using one dataset (combined) or different datasets (female, male, childbirth etc.). We will decide that when working with them. As these data sets are large, we don’t think it will be feasible to do it by hand. I will either use python to work on it or use chatgpt to do the operations. Though ChatGPT’s recent advance data analysis is quite helpful yet to see if it can handle large datasets.

**Analysis Questions**: There are a multitude of questions that can be explored with this data and its visualization. A few questions would be how the world’s population is distributed across different regions and countries and how male and female populations are distributed. This can help in seeing which areas have high or low population densities and as well as detect gender imbalances and investigate the reasons as to the imbalances. With birth and mortality rate data included, we could also tackle questions related to growth rate and highlight possible regions with socioeconomic disparities that usually come with higher birth and death rates. The general benefits of making this visualization would be the ability to offer insights into global and regional demographic patterns as well as addressing various challenges related to population, healthcare, and social development.

.

**Visualization Design**:We are planning to do a 3D map of the world showing data as spike map. It will be interactive as choosing different options (ex: female population etc.) will be shown in the plot.



Figure : Our main design plan created with DALLE 3. The final project won't have this much of graphics. But something similar.

Alternative design:

1. Tree map
2. Vertical bar chart
3. Heatmap.

These alternative designs are effective to make people understand world population distribution but each of them lacks some features. For example, **tree maps** are a good way of communicating information about different countries. But it won’t tell people where the country is located and won’t give much idea about the region. Even though a **heatmap** will tell where the country is located but it won’t give a perception of population as there isn’t any spike to express it. Just numbers or description might make many readers turn away from the visualization. On the other hand, **vertical bar** chart is a good way of visualization for quantitative analysis, but it will be very hard to put all the country in one page. Also, no data of the region of the country.

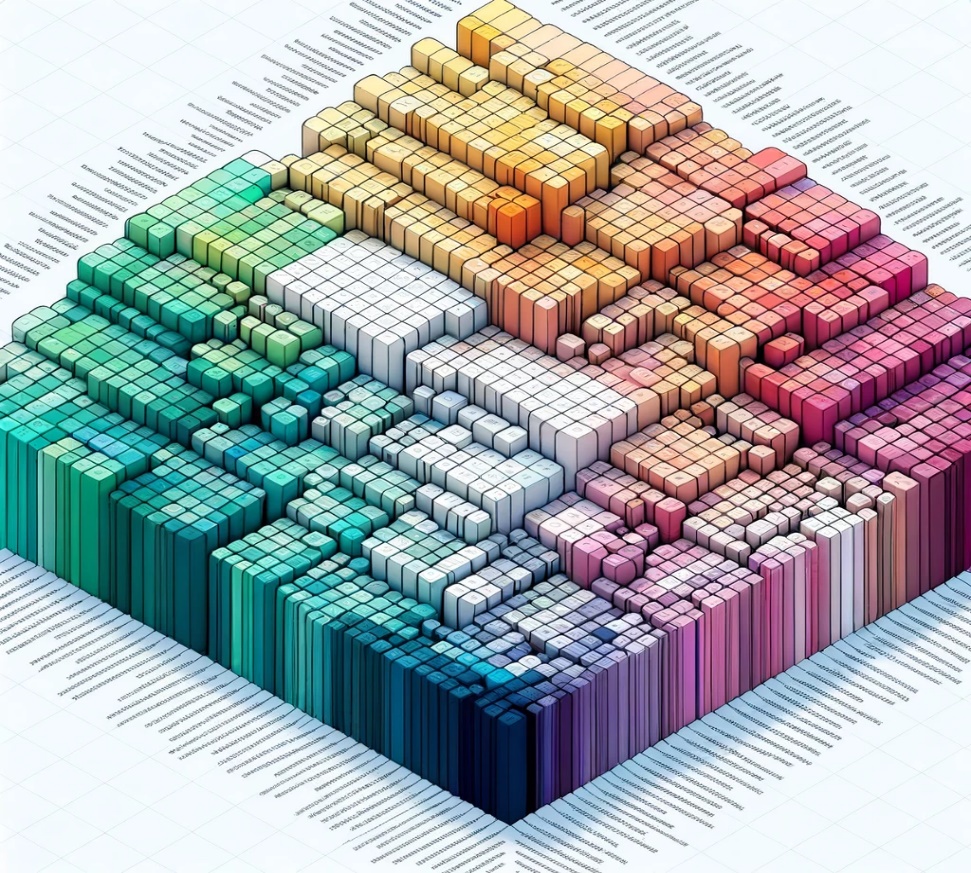


Figure : Tree map of world population.

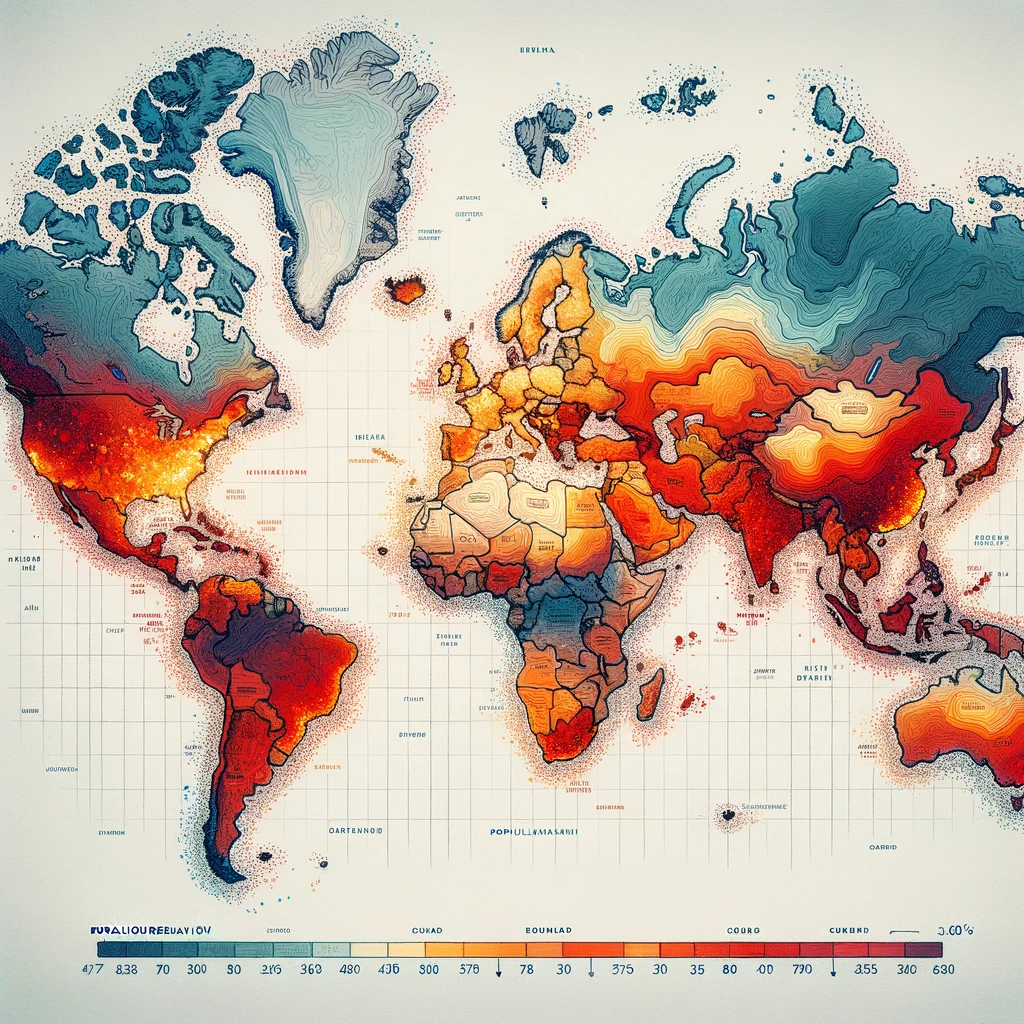


Figure : Heat map of world population

Considering all other options and the goal of my project, I believe my spikemap will be best for the qualitative understanding of world population and it’s distribution.

**Must have features:**

1. A spike map with options for different data for display
2. Selecting different countries will show its historical population change.

**Optional features:**

1. When choosing a certain data like child mortality rate or male population, in a side window it will show the top 10 countries.
2. Add more data sets to it like income per capita etc.

**Project Schedule:**

**10/30/23 – 11/4/23**

* Retrieve necessary data/datasets

**11/5/23 – 11/11/23**

* Complete data acquisition
* Create Exploratory Jupyter Notebook
* Start data wrangling

**11/12/23 – 11/18/23**

* Complete data wrangling
* Finalize data structures to prepare for visualizations
* Work on exploratory Jupyter Notebook for Milestone due date
* Start on explanatory Jupyter Notebook

**11/19/23 – 11/25/23**

* Focus on visualizations for explanatory Notebook using our processed data
  + Start on our main visualizations and work on alternatives if needed

**11/26/23 – 12/2/23**

* Clean up visualizations
* Practice our presentation

**12/3/23**

* Final Project due