Vishal Saravanan

Information Visualization Section 2

Flinsch

31 October, 2024

**Merging the Datasets Through Tableau**

I began by importing the 'population-and-demography'(Population) and 'number-of-teachers-across-education-levels'(Teachers) datasets into Tableau as text files (because they are comma-separated values). I intended to merge both of these datasets on both entity and year. Initially, I was unsure how to achieve this the way I wanted, so I clicked 'Learn More' to get to Tableau's website to watch an instructional video. The video showed how to make a relationship and a union. I first tried to use relationships, and I could only see education values and no population values, even though I tried to relate based on both Year and Country. Then, I used a union where I did an Inner Join on both Year and Country Name/Entity. Although there were some duplicate columns, I only used one for year and one for country names. The image below shows the final product under Data Source.

Before beginning, I wanted to double-check my data. I dragged in the column "Year" from the "Teachers" dataset (range between 1970 and 2015). I also included the "Entity" column from the "Population" dataset filtered for the world. The datasets were joined properly because the population value was correct for 2015 after mixing columns from both datasets. It did not matter whether or not I used the "Year" column from either dataset because they had the same range and resulted in the same population levels, which is another piece of evidence that shows a proper join.
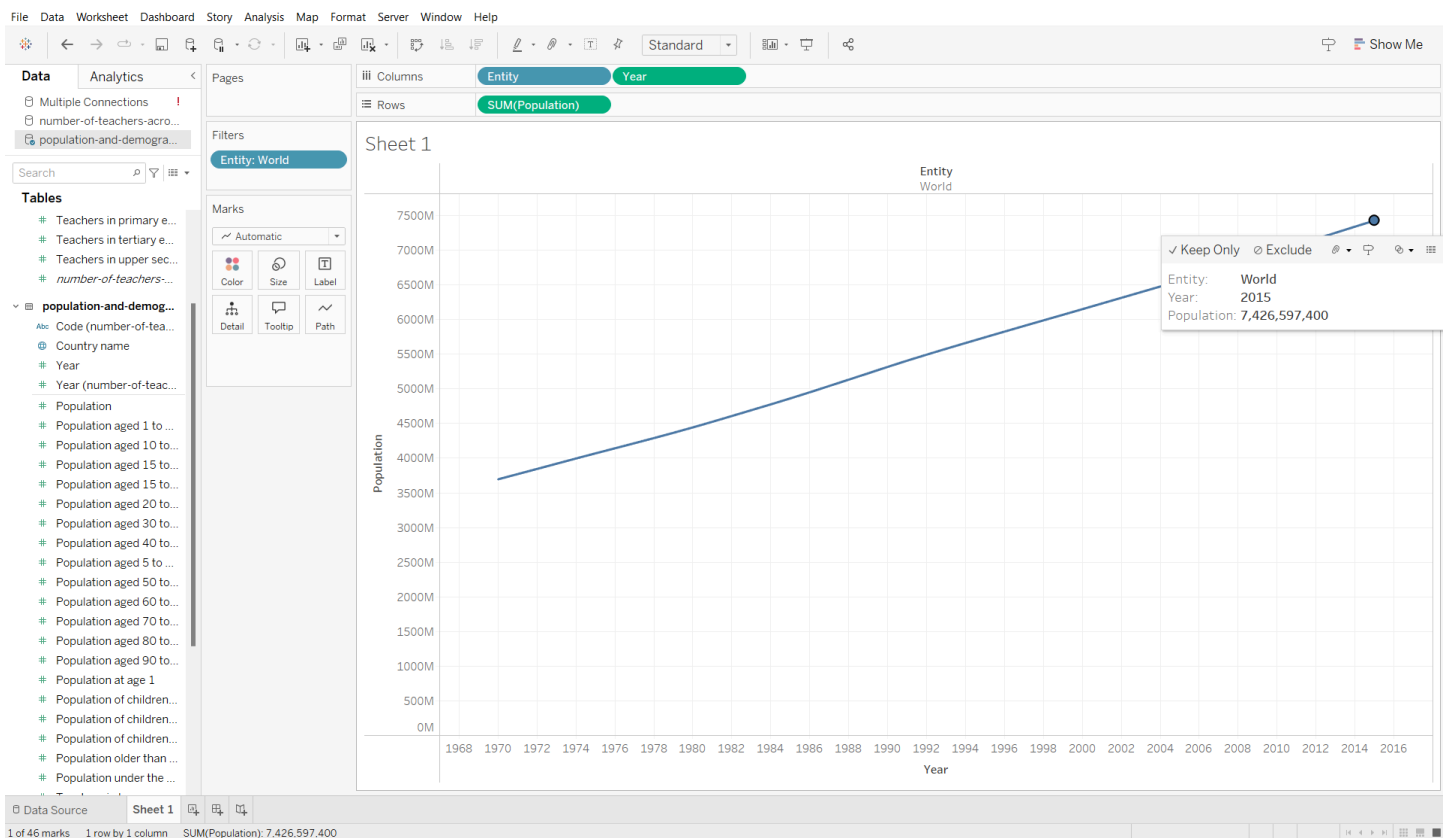
Percentage of Teachers in the US

The first sheet calculates the percentage of teachers in the US. I used the Population of people between the ages of 15-64 because it accurately represents the age of the working class in the US. I assumed that the average retirement age is around 64 years old. I created a filter for the column "Entity" that only focused on the data from the US. Then, for my calculations, I summed the columns for all the different types of teachers (lower secondary, pre-primary, primary, tertiary, upper secondary), which I then divided by the Population (aged 15 to 64 years) and multiplied by 100 (to get the percentage value). I had to find the average of each because there were many null values. I then added annotations to show the difference between the percentage of teachers in the US in 1971 and 2014. The percentage of teachers in the US significantly decreased from roughly 4% to 2.4% which indicates either significant growth in population or decrease in teachers. I changed the background color to light orange, the names of both axes, and the title.
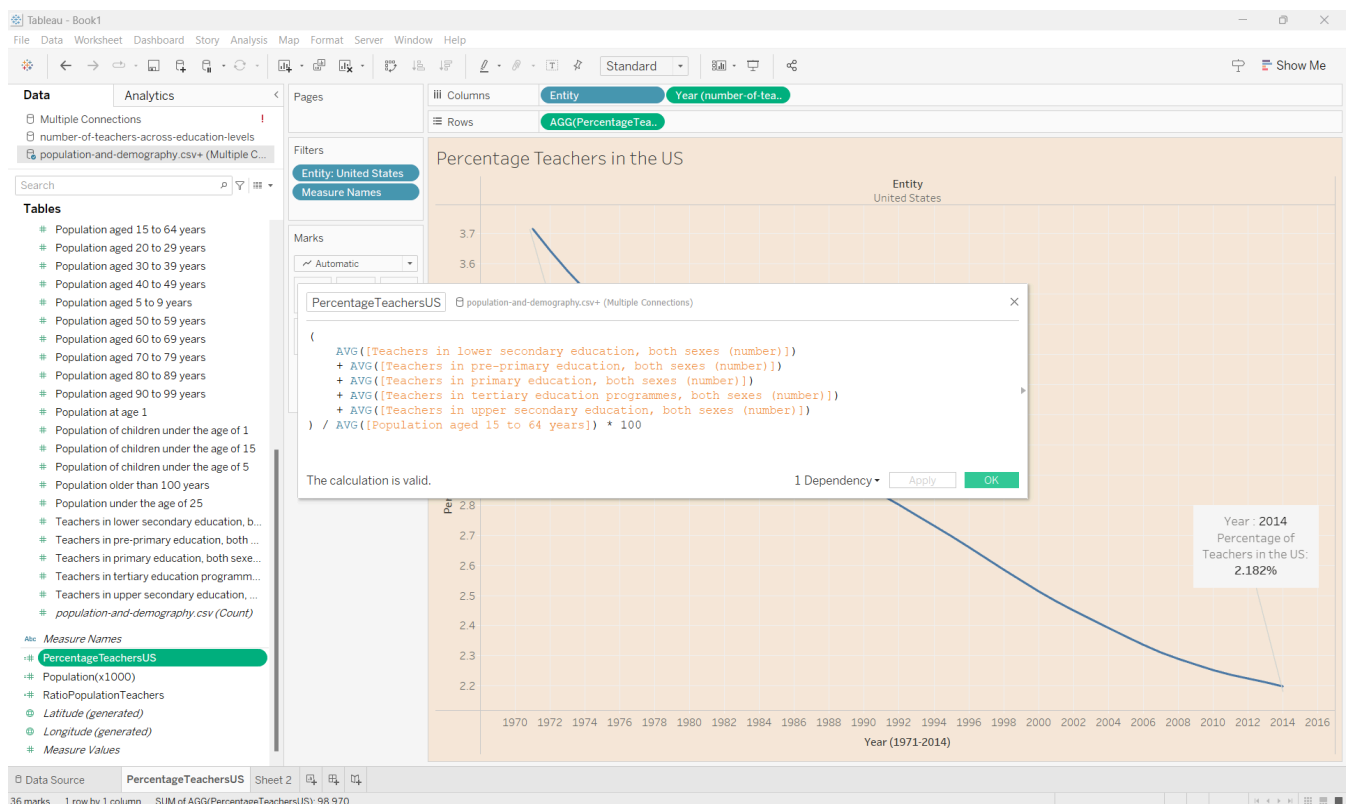
The Ratio of Students to Teachers Across Different Countries

      The second sheet focuses on the ratio of students to teachers in the countries with the

world's five highest GDP values. These countries (USA, China, Germany, Japan, India) each

have varying populations, so there was a need to take a ratio between the students and teachers to

compare each of the countries. I struggled a lot when using aggregate functions in my calculated

fields. I experimented with different methods of using SUM and AVG. I discovered many

discrepancies in the SUM aggregation because of the number of null values, so I used AVG. I

used the Show Me feature to get all five lines onto the same graph with my calculated field on

the y-axis again and the "Year" column on the x-axis. I filtered the "Entity" column to have only

the five countries I wanted. I added annotations to the data based on trends I analyzed, including

how the student-to-teacher ratio in India is significantly higher and the year in which the

student-to-teacher ratio of the US surpassed Japan. Once more, I added a light blue color to

highlight each country's line's darker colors and changed the axes' names and titles.
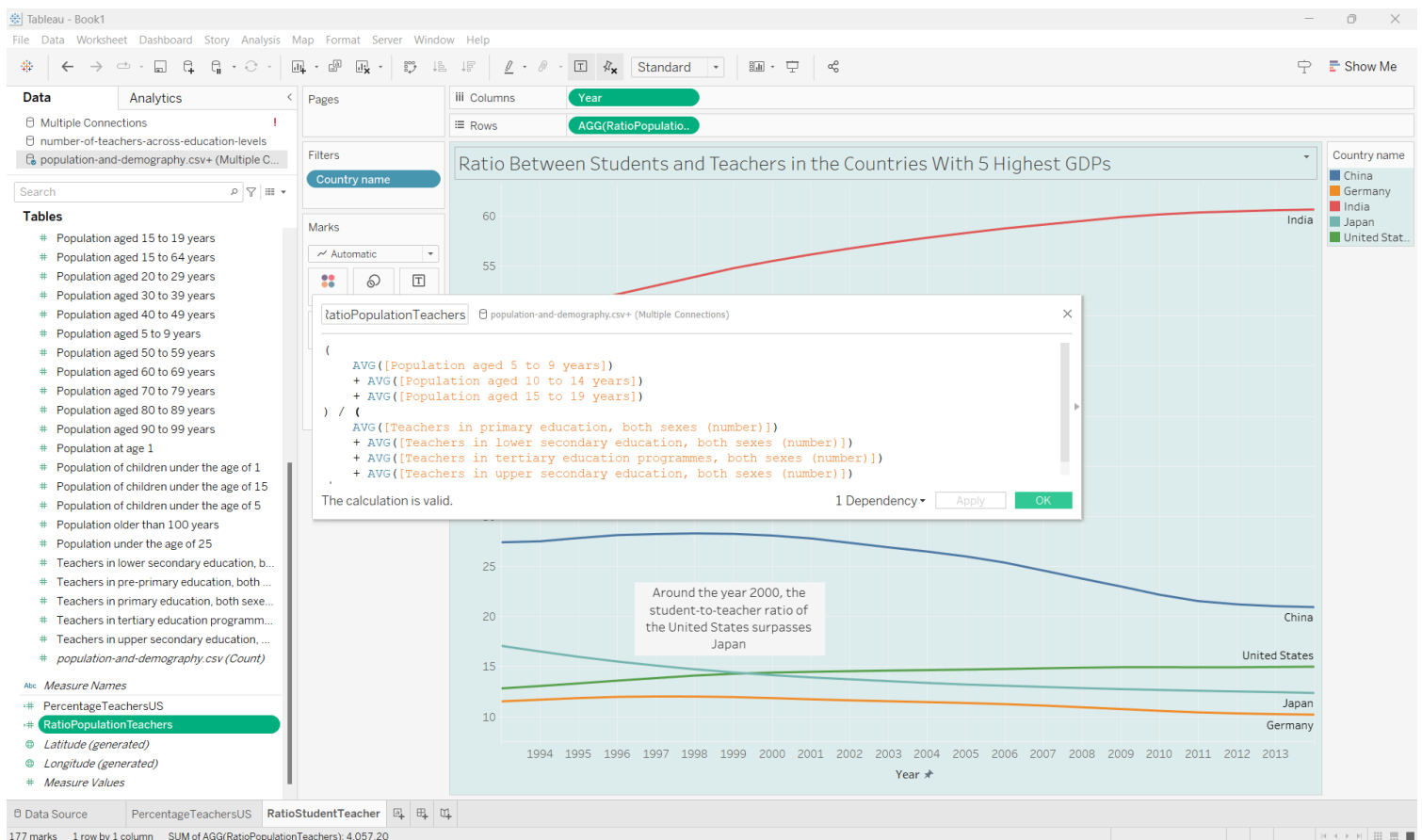
The Ratio of Students to Teachers Map

I used the calculated field described in the previous paragraph to calculate the student-to-teacher ratio. First, instead of having the "Year" column contain all years like before, I made the year 2014. Then, I used the Show Me feature to select the map option. The calculated field was used as a Text Mark and categorized by color. The darker blue shade represents a place with a larger ratio, and a lighter greenish color represents a smaller ratio. In African countries, there are much fewer teachers than in other countries, which accounts for the high ratios. A smaller ratio means that there are many teachers compared to students or that there is a small population. For example, there are many teachers in the USA, but Bermuda has a small population.

Percentage of Teachers Bar Chart

My main comparison between teachers and population was the number of teachers in a given population represented by percentage (calculated field from the first sheet). I focused on many countries worldwide, including those with the highest populations and GDP. I experimented with the 'Year' column to implement pages this time. I wanted to cycle through the years to show changes over time. An unintended consequence of going year by year was the null values. Some pages contained null values for specific countries, but we could still see the trends if played quickly. The overall percentage of teachers slowly decreased over time, which could mean that the population growth rate is higher than the teacher occupational growth rate. I implemented red bars on a gray background, changed the labels on the y-axis, and changed the title.

Population and Teacher Comparison

       I utilized filters that used the population(AVG) and number of teachers in tertiary education(AVG) in the US. I created a dual-line graph showing that the population grows much faster than the number of teachers. I created an annotation to describe this effect, adjusted the background color to a light purple to contrast with the dark lines, changed the scales not to include 0, and adjusted the titles to represent that data accurately.