Reeya Pimple
ID: 2122995

DATA 511: A4

Exploratory Visual Analysis

Introduction

To understand any complex dataset, one can conduct exploratory visual analysis on it to identify the underlying trends and patterns in the data. For this assignment, we will analyze the World Development Indicators dataset which is published by The World Bank.

Over half of the refugees in 2018 were children under the age of 18. The refugee crisis isn't just a humanitarian issue, it is a highly politicized one which has citizens divided throughout the world. Over the last few years, numerous elections have been conducted in developed countries, wherein certain politicians promised to "curb" this issue. Thereby, making it clear to the masses that they do not support giving asylum to the refugees as they believe this will rob their citizens of the resources that the country has to offer.

Using the data and analyzing the various metrics provided in it, we can answer the following question, "Is the refugee crisis supported equitably by all countries?" To answer this question, we will first Clean and Manipulate the dataset using Tableau Prep, and Visualize it using Tableau Desktop.

Data Profile

The dataset, World Development Indicators, is released quarterly by The World Bank. The World Bank is a reputable international financial institution that was founded in 1944 with the aim of providing financial assistance to developing countries. The dataset is compiled from officially recognised international sources which proves its veracity.

We will use the latest dataset, updated in October 2021. It is contained in the file "WDIData.csv" of size 204.9MB. Before cleaning, the dataset contains 66 fields and 384K rows. It covers Development Indicators of countries from the period beginning 1960 to 2020.

The fields contained in it include Country Name, Country Code, Indicator Name and Indicator Code which contain values of the Nominal DataType. The remaining fields denote the Years from 1960 to 2020, containing values of the Ratio DataType.

We have to Clean and Shape the data before conducting any analysis. The dataset contains many NULL values under the years columns, which range from 1960-2020. The Country Name column also contains

many values such as “Africa Eastern and Southern” and “Euro Area” which might have been previously used to analyze specific regions. We will filter out 41 of those values in Tableau Prep, as given in Figure 1.

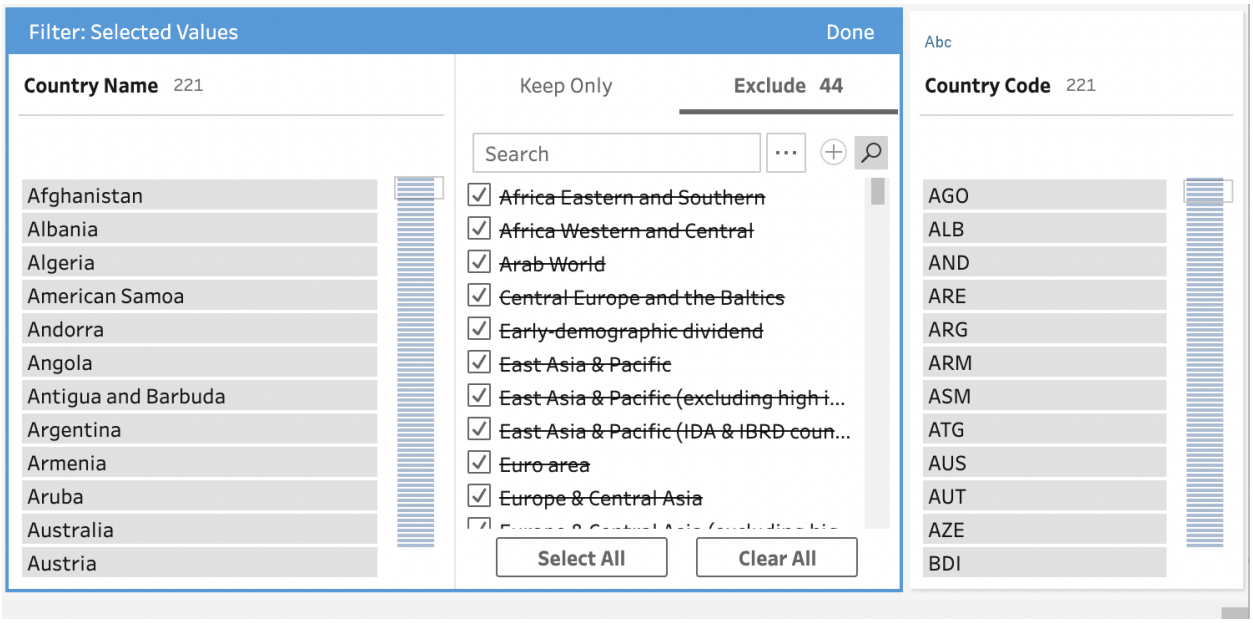

Figure 1. Using Tableau Prep’s Clean step, to filter out unwanted Country Name values

Question Exploration

Using the World Development Indicators dataset, we will answer the question “Is the refugee crisis supported equitably by all countries?”

The United Nations High Commissioner for Refugees (UNHCR) organized the first Global Refugee Forum in December 2019. The forum served as a platform to announce financial support and had discussions aimed towards countries sharing the burden more equitably. Answering the question “Is the refugee crisis supported equitably by all countries” can have a two fold impact, wherein it will depict the current situation and also the future potential of the countries.

Using Tableau Prep for Data Cleaning, as stated earlier, we removed all the aggregated Country Names which still leaves us with more values than the actual number of countries, but these values are filtered out again in the Visualization step. We use the Pivot function to convert all the years starting from 1960 to 2020 to a single row. Similarly, we convert the Indicator row to Columns wherein it is filled with all of the years’ data. We can download the refined dataset from Tableau Prep as a csv file and upload the same in Tableau Desktop for visualization.

To answer the question, we will create Sheet 1 in Tableau Desktop using the variables: [Refugee Population by Asylum Country], [Country Name] and [Year]. We will filter out the NULL values from the variable Refugee Population by Asylum Country, and we will filter the Year variable such that it only contains the latest data, i.e. from the year 2020. The type of visualization used is Maps, so that it doesn't require a lot of cognitive capacity to understand the correlation between the Countries and Refugee populations. If we directly display the actual Refugee population of each country, we will not be able to make direct comparisons, hence we will depict the Refugee Population by Asylum Country as a Percentage of the Total.

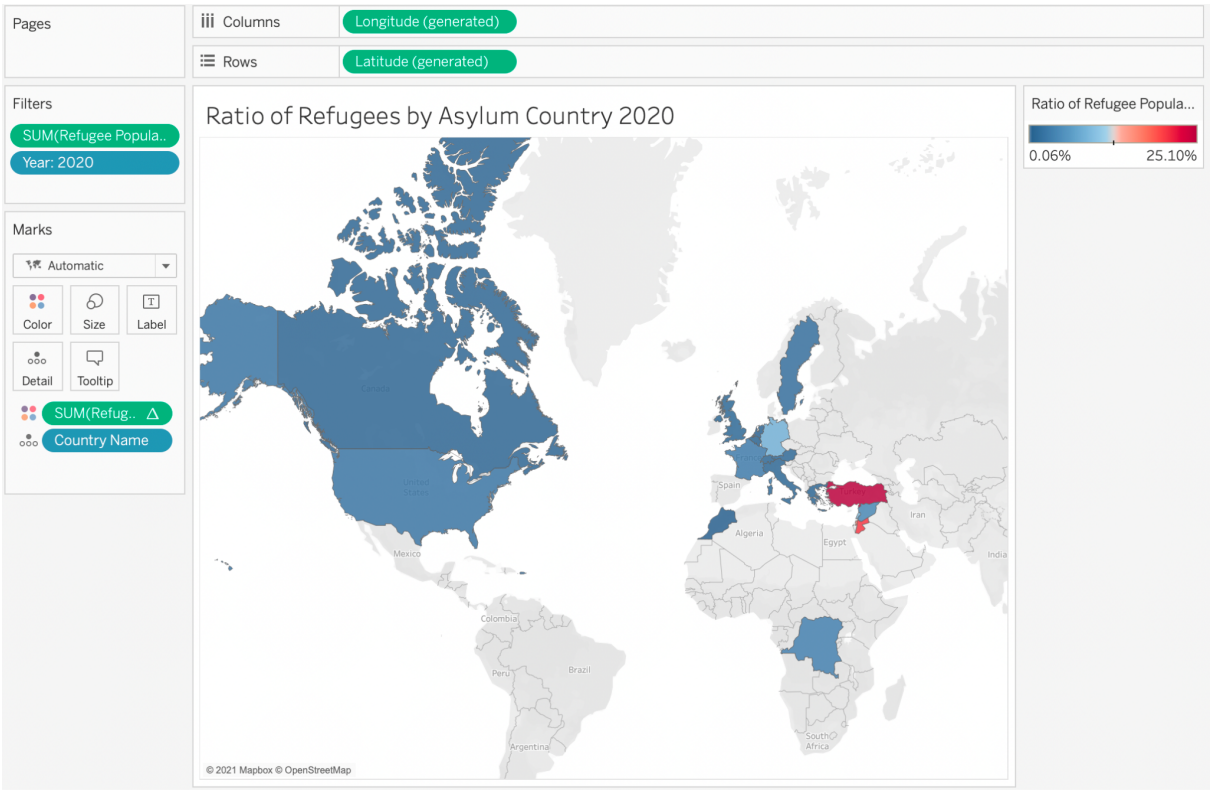

Figure 2. Visualization I Representing Ratio of Refugees by Asylum Country (2020)

Thus we can say that for Data Encoding in Visualization I, given in Figure 2, we rightly use a map to represent the Country Names, as the perceptual properties of humans help them distinguish between Position first. The Refugee ratio is represented using Hue and Saturation both. If we use only Color Saturation, it is difficult to distinguish between the neighbouring countries representing different values. Hence, we make the use of the Red-Blue Diverging color palette. Here, we have used the color red as it signifies urgency or danger.

Using Tableau aids us in navigating the data, adding annotations and looking at key features of it in-depth. For example, we can observe that many Refugees are hosted by European countries and neighboring countries. To visualize what appears to be a “hotspot” in the map, we can zoom-in to the area.

In Figure 3, we can observe that Turkey has the maximum number of Refugees, at almost 25%. Among the European Nations, Germany appears to have the highest number of Refugees at around 8.32%.

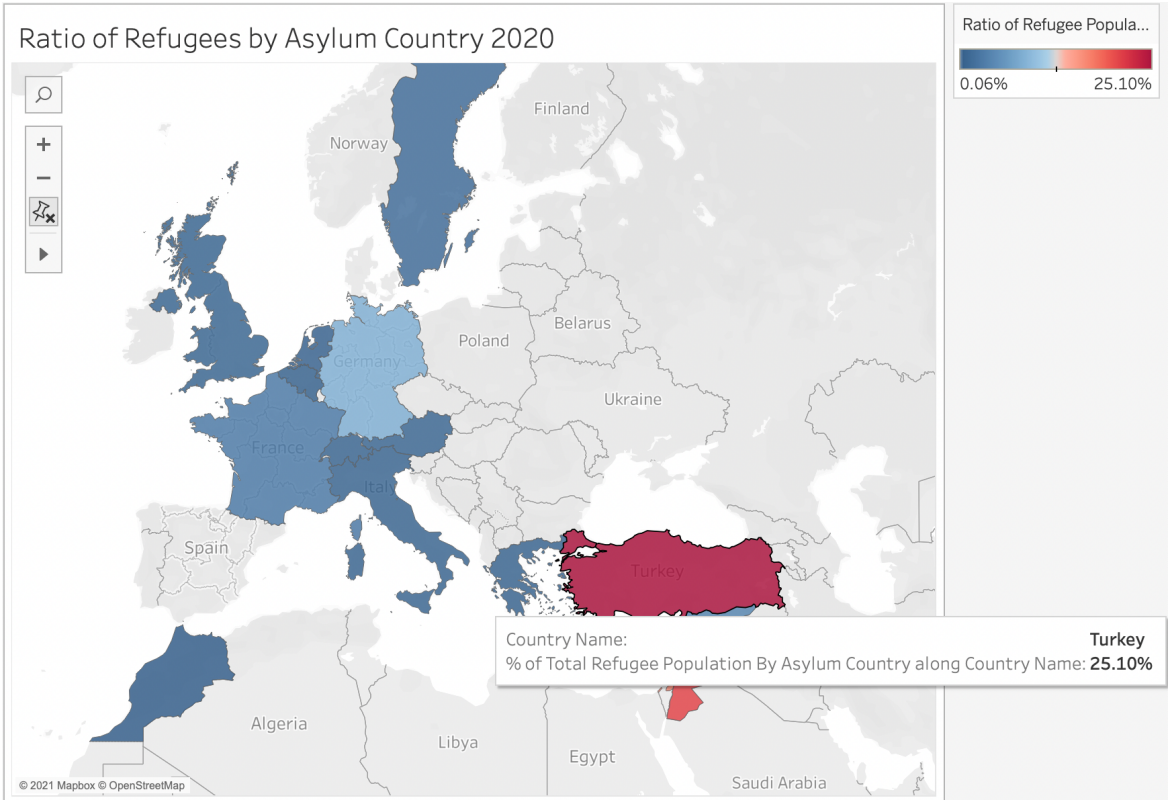

Figure 3. Zoomed-in View of Visualization I

While Visualization I depicts the Refugee Crisis, we can try to correlate equitable distribution with the Gross Domestic Product (GDP) of a country. We have to understand that GDP might represent the prosperity of a country, but its advantages are only observed if every citizen gets the benefits. Thus, for Visualization II, we will present the GDP per Capita of countries, to understand if the transfer of wealth is equitable.

For Data Encoding in Visualization II, given in Figure 4, we again represent the Country Name variable on the map, and use Color Saturation to represent the GDP per Capita. In this instance, we use the color Green as it is easy to distinguish and we are representing dollar amounts. Also, it is clear that due to the high disparity of GDP per capita among different nations, the difference in Color Saturation is easy to detect in this instance.

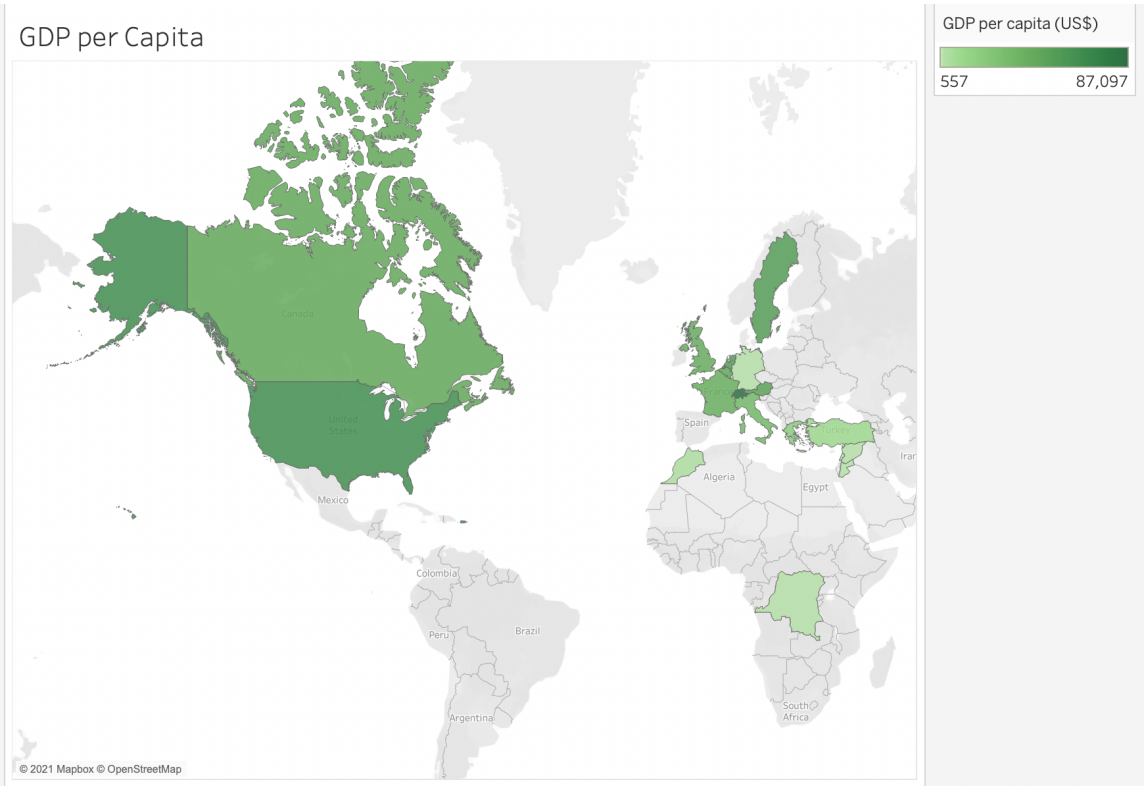

Figure 4. Visualization II Representing GDP per Capita in US$ (2020)

In Visualization II, we can observe a similar trend, wherein the developed are on one end of the spectrum and the developing countries are on the other. For comparison, we can again zoom-in to the same region.

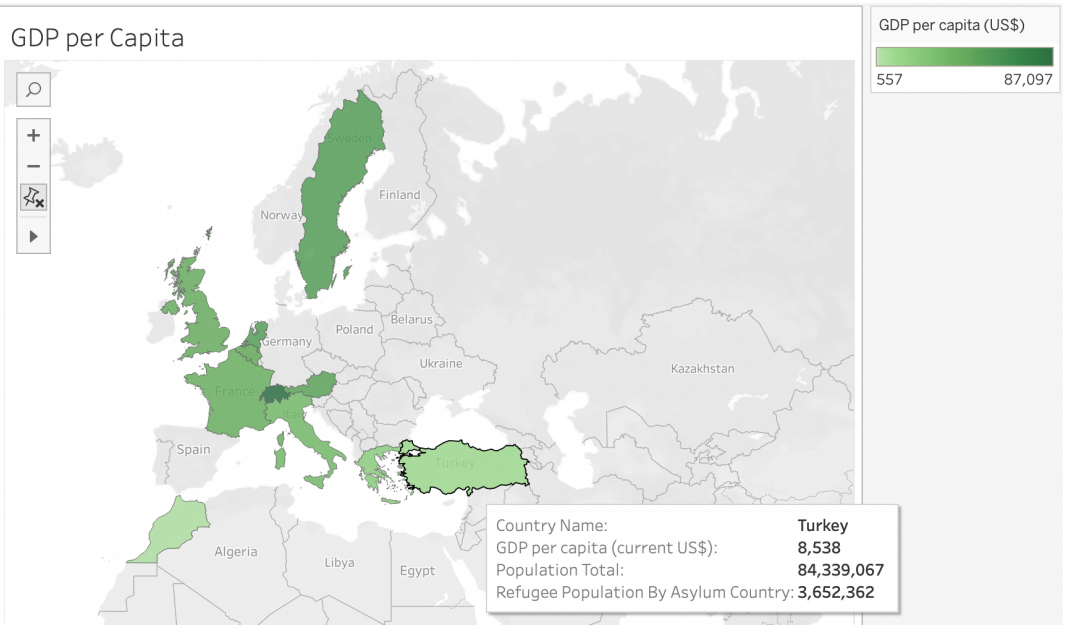

Figure 5. Zoomed-in View of Visualization II

We can further analyze the data by creating a Crosstab of the two countries in Figure 6. When considered as Percent of Total, while France makes up for 82.05% of the GDP per Capita($US) , it is Turkey that hosts 89.33% of the Refugees.

GDP per Capita

Country Name	% of Total GDP per capita (current US$) a..	% of Total Refugee Population By Asylum..
France	82.05%	10.67%
Turkey	17.95%	89.33%

Figure 6. Comparison of France and Turkey

It is clear upon observing all of the Figures, that as per the latest data of 2020 provided by The World Bank, the Refugee crisis is not equitably supported by all countries. Presently, there are countries which are stretched thin on resources which are housing a large number of refugees, such as Turkey. Upon further research, we realise that there are many reasons for this such as Turkey being surrounded by war-torn Syria and it being a “transit country” to Europe for many asylum seekers and refugees.

While data for all the countries is not present, we observe that there are many countries which can help in shouldering this responsibility of at least creating temporary shelters for the people who are fleeing war and persecution. The GDP per Capita data of these countries are testament to the fact that they are viable options.

Conclusion

We can thus answer the question “Is the refugee crisis supported equitably by all countries?”, after analyzing the World Development Indicators dataset. It is clear that the Refugee Crisis is not equitably supported by all countries. Infact, it is highly skewed. Making this data public can educate the public about the current state of affairs, and help them make a sound judgement in times when politicians make false claims. This will help in improving the reputation of refugees is not impugned.

It is also important to note that the missing data results in an incomplete picture. For instance, while Germany hosted over 1.1m refugees in 2020, Pakistan hosted over 1.4m refugees, mainly fleeing persecution in Afghanistan. This is another instance of a developing country hosting refugees way beyond its economic limits, compared to a developed country. This re-affirms the idea put forward by the visualization, as Turkey is not an outlier being the only developing country to house most of the refugees.