MONASH University

Data Visualization Report

Refugee Unemployment

FIT5147 Data Exploration and Visualization

Semester 1 2023

Name: Nazmus Sakib

Student ID: 33361991

Tutor: Niranjan Nanjunda

Table of Contents

Introduction

Refugee unemployment refers to the challenge faced by individuals who have been forced to flee their home countries due to persecution, conflict, or violence, and subsequently struggle to secure stable employment in their host countries. This issue is prevalent in many regions across the globe, as refugees often face numerous barriers and obstacles when it comes to finding suitable employment opportunities. The high rates of unemployment among refugees can have profound implications for their overall well-being, integration into their new communities, and long-term economic stability. Understanding the complexities of refugee unemployment is essential for policymakers, organisations, and society as a whole to develop effective strategies and support systems that can empower refugees to overcome these challenges and build sustainable livelihoods.

To gain a better understanding of the refugee situation, this project aims to visualise the flow of refugees in a country and compare it with job demand. By collecting data on the number of refugees entering the country and the availability of job opportunities in different regions or industries, we can analyse the distribution of refugees and the potential economic impact of their resettlement.

The projects display heatmaps, proportional symbol graphs, line charts to visualise the data for any statistician or employee of the Bureau of Statistics.

Design

The type of graphs used in the Five Design Sheets are line charts, proportional symbol graph, heat maps, correlation matrix, bar charts and scatter plots.

As most of the data collected is a time series so the best graphs that could display the information were the line charts. With line charts our data from different years can be easily viewed. Line charts enable the visualisation of trends and patterns in the data by connecting the data points with a continuous line. This allows for a clear understanding of how the data changes and evolves over the given time period. Time series data often exhibit patterns and seasonality, where certain trends or fluctuations repeat in a predictable manner over specific time intervals. Line charts also facilitate the comparison of multiple time series on a single chart. By using different coloured lines or markers, it becomes easier to distinguish between different variables or categories, enabling the viewer to understand the relationships and interactions between them.

The scatter plots were used because there were too many variables to display at once for the work status and the PR status. Scatter plots are most useful for examining and understanding the relationship between two continuous variables. Scatter plots provide a means to examine the correlation between two variables. Correlation measures the strength and direction of the relationship between the variables. It also displays individual data points as dots on a graph, with one variable represented on the x-axis and the other variable on the y-axis. This visual representation helps in identifying patterns, trends, and relationships between the variables. Scatter plots can reveal linear relationships (positive, negative, or no

correlation), non-linear relationships, clusters, outliers, or lack of association between the variables.

Heatmaps are a powerful data visualisation tool that uses colour gradients to represent the intensity or density of values in a matrix or table. As the project also had geospatial data or geographical data, heatmaps was the best option to display the information on a geographical plane. Heatmaps are an excellent way to summarise and condense large amounts of data into a compact and visually appealing format. By aggregating and displaying data in a grid format, heatmaps provide a concise overview of the entire dataset, allowing for quick insights and analysis. Heatmaps enable the simultaneous comparison of multiple variables. By mapping different variables to the rows and columns of a heatmap, it becomes possible to identify relationships and correlations between them.

A correlation matrix helps determine the strength and direction of relationships between variables. By calculating correlation coefficients, such as Pearson's correlation coefficient, it quantifies the degree of linear association between variables. Positive values indicate a positive correlation, negative values indicate a negative correlation, and values close to zero suggest no significant correlation. It also helps in identifying patterns and dependencies within a dataset. High positive correlations indicate variables that tend to increase or decrease together, while high negative correlations suggest variables that have an inverse relationship. These patterns can provide insights into the underlying dynamics and connections between variables. Correlation matrices are useful for detecting multicollinearity, which occurs when two or more independent variables in a regression model are highly correlated. High multicollinearity can lead to unstable and unreliable regression coefficients. By examining the correlation matrix, one can identify variables with high correlations and take appropriate actions, such as removing redundant variables or using regularisation techniques.

Proportional symbol graphs, also known as proportional symbol maps, are a type of data visualisation that uses symbols of varying sizes to represent quantities or values associated with specific geographic locations. These graphs are commonly used in cartography and spatial analysis to depict data that has a spatial component. Proportional symbol graphs can be extended to represent multiple variables simultaneously. By using different symbols or colours for each variable, the graph can convey additional information about multiple aspects of the data. This enables the exploration of relationships and interactions between variables within a spatial context.

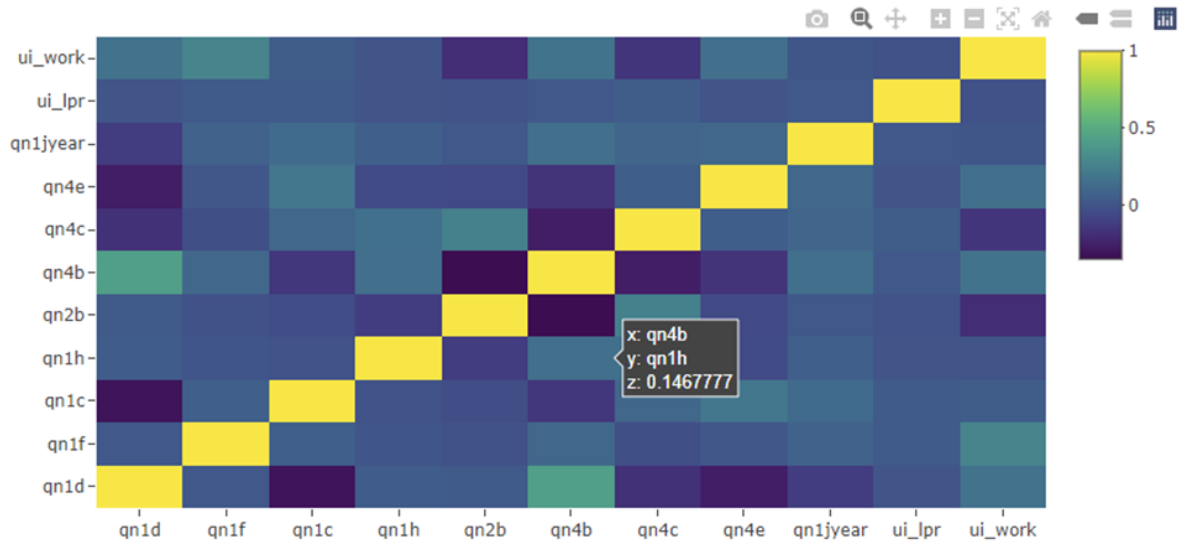
Most of the colours used are simple and easy to identify as colours are used in graphs that have multiple variables. So the final design was chosen for heatmaps, line graphs, proportional symbol graphs because the data collected from the sources are mostly multi variables, geospatial and time series data. So these graphs could give the best outputs and the message gets clear when anyone would see the graph.

Implementation

The libraries used in the implementation are:

- library("data.table")
- library(dplyr)
- library(ggplot2)
- library(shiny)
- library(jsonlite)
- library(tidyverse)
- library(maps)
- library(plotly)
- library(gganimate)
- library(animation)
- library(gapminder)
- library(ggthemes)
- library(readr)
- library(reshape2)

These libraries are used to view all the graphs that are in the Visualization project. The maps produced all require the libraries, some graphs might not need all the libraries but to view all the graphs all these libraries are needed.

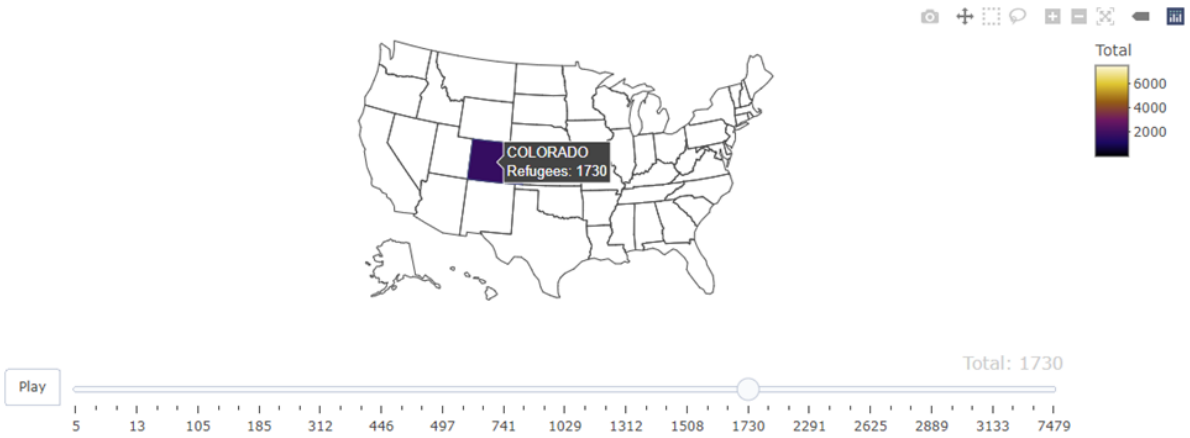


This is the correlation between all the variables that have been collected from the ASR dataset. The link to the dataset is:

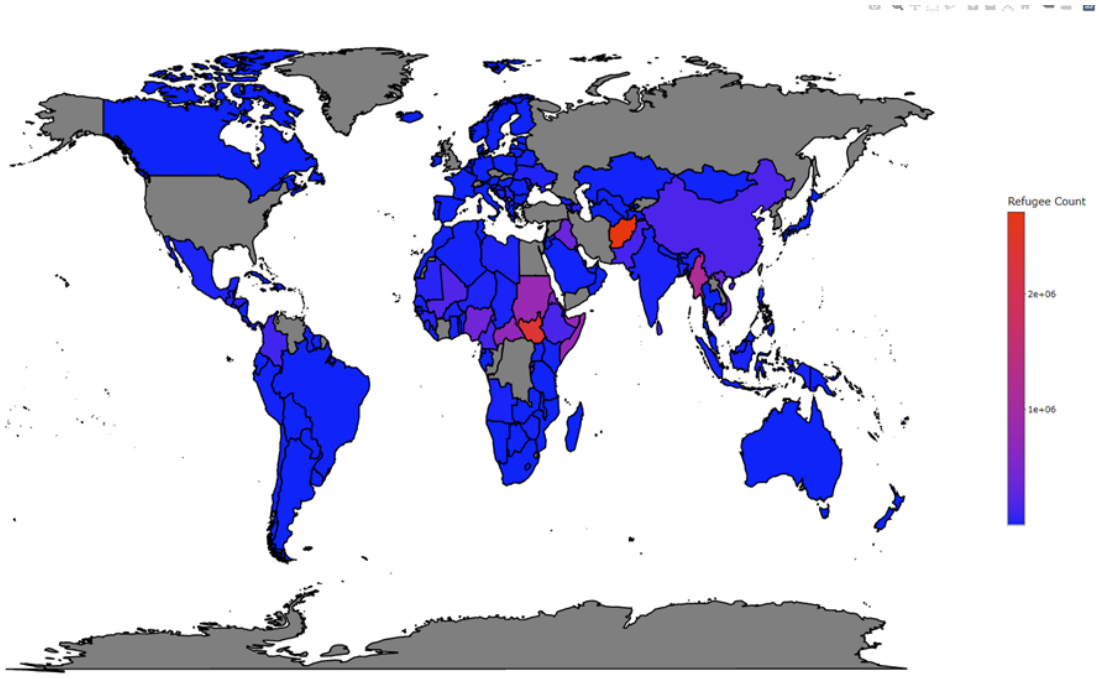
Refugee information in the US region - https://microdata.unhcr.org/index.php/catalog/681/study-description#metadata-data_collection

The correlation matrix is built with cor() function and the graph is plotted with the plotly library.

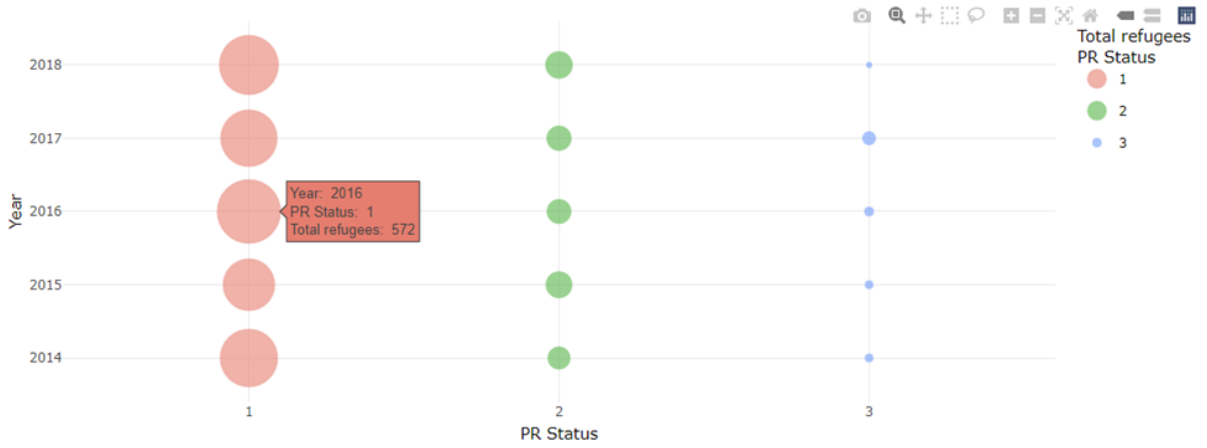
This graph displays the unemployment rate of the USA for refugees. First the unemployment csv file is gathered from the U.S Bureau of Statistics then wrangled to view the refugee data. Ggplotly is used to view the graph, transition_reveal and frames are used to animate, reshape2 is used to reshape the data.

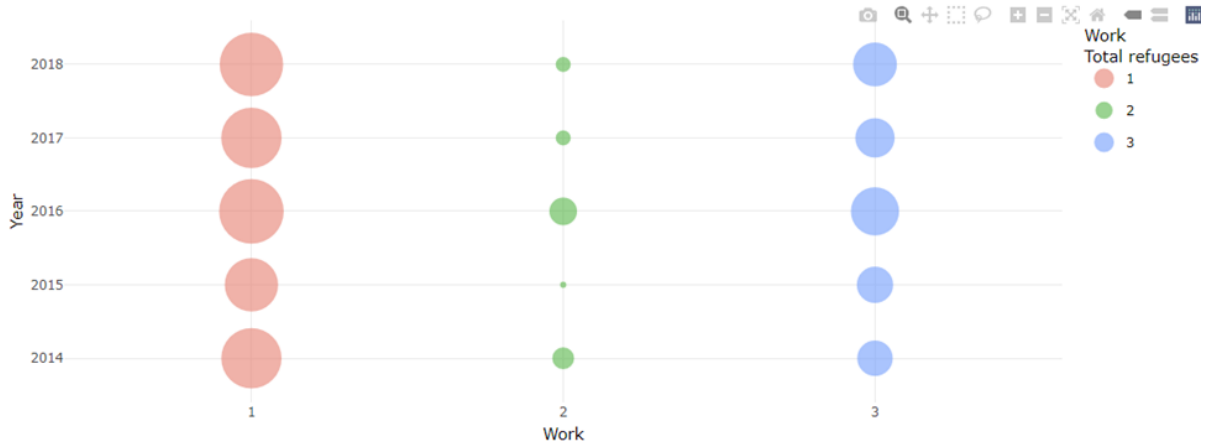


This graph displays the number of refugee intake per state and json data is used. To view the map plot_geo function is used with locationmode and frame to animate it.

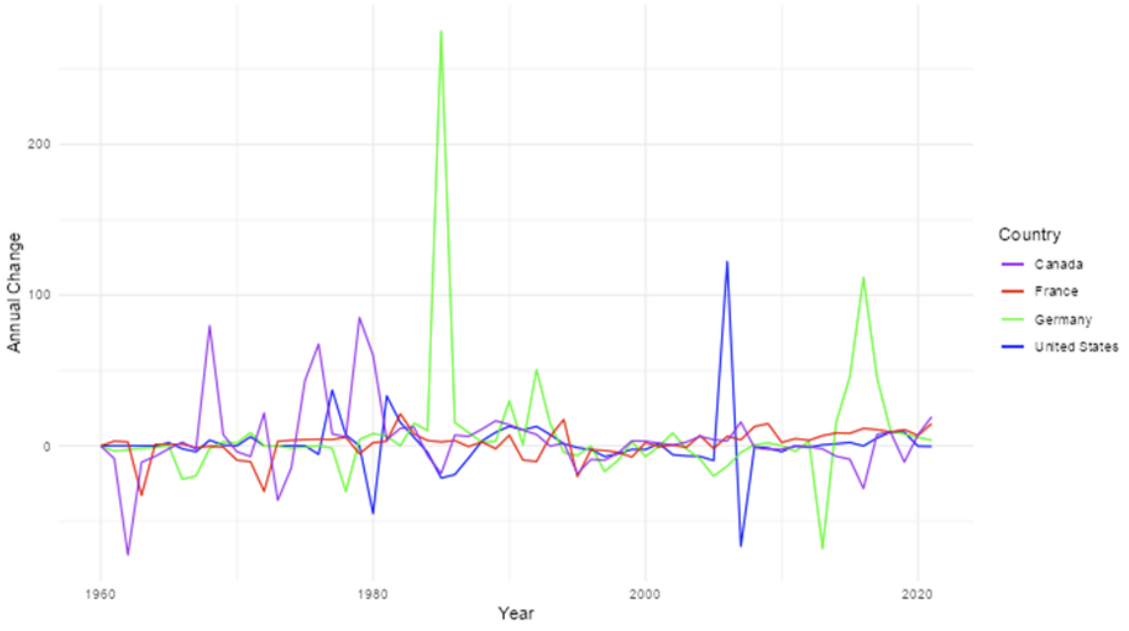


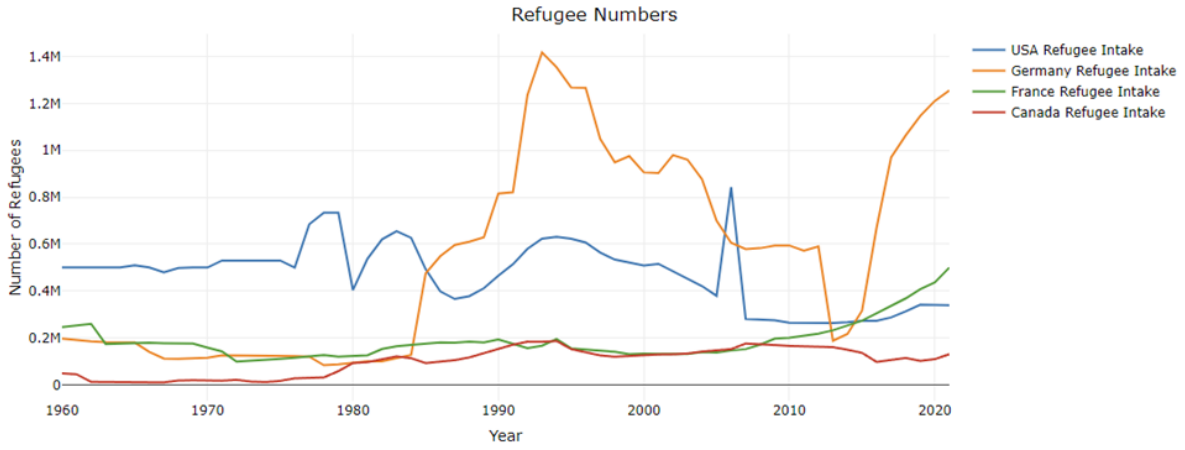
This map shows the refugees that flew out of the country in the specific year. Ggplotly is used to view the graph and map_data is combined heatmap data to view this map. The user can also select the year that he wants to view.





For the above two graphs ggplotly is used to view the graphs. The variables are turned into factors before creating the graph. Scale_size_continuous() is used to measure the size of the bubbles that is used to display information. These data are gathered from the refugee data or the ASR Dataset.





These two graphs are the simplest graphs which display the refugee intake per year per country and the annual change in refugee per year per country. Plotly() function is used to display the graphs.

This project was challenging because there is not enough data to come to a conclusion. The data collection was very tough and most of the data was confidential so I could not collect them. The wrangling of the data was very crucial and difficult. Gathering accurate and comprehensive data on refugee populations can be a complex task. Refugee flows involve multiple countries, organisations, and agencies, each with their own data collection methods and reporting systems. Coordinating and standardising data collection across different sources can be difficult, leading to inconsistencies and gaps in the data. It is often dispersed among various organisations, including government agencies, international bodies, and non-governmental organisations. Each entity may collect and report data based on their specific mandates and criteria, resulting in fragmented and decentralised data sources. There is often a lack of standardised data formats, definitions, and classifications when it comes to refugee statistics. Different organisations may use varying terminology and categorizations, making it difficult to compare and combine data from different sources. Refugee statistics may suffer from data quality issues, such as incomplete or missing data, data entry errors, and discrepancies. The nature of refugee situations, which involve movement, displacement, and often volatile contexts, can make data collection challenging and prone to inaccuracies. The datasets were very large and it had too many variables that were not relevant to unemployment of refugees. Among the 250+ variables only 12 were selected and one of them had more than half of the data missing. Still these 12 variables are not that useful that I thought they would be.

Overall, the complexity of refugee situations, the diversity of data sources, and the inherent challenges in data collection, standardisation, and quality assurance contribute to the difficulty in wrangling annual refugee statistics data. Overcoming these challenges requires close collaboration among relevant stakeholders, the establishment of standardised data collection mechanisms, and ongoing efforts to improve data quality and accessibility.

Links:

- https://www.macrotrends.net/countries/DEU/germany/refugee-statistics
- https://www.macrotrends.net/countries/USA/united-states/refugee-statistics

- https://www.macrotrends.net/countries/FRA/france/refugee-statistics
- https://www.macrotrends.net/countries/CAN/canada/refugee-statistics
- https://data.bls.gov/PDQWeb/ln
- https://microdata.unhcr.org/index.php/catalog/681/study-description#metadata-data_collection
- https://www.bls.gov/charts/job-openings-and-labor-turnover/opening-hire-seps-level.htm

User Guide

The project displays the graphs used to figure out the Refugee Unemployment Rate.

1. The datasets are in the folder called “Datasets”.
2. The user should adjust the locations of the datasets before running the code.
3. Change the location of all the datasets according to the user’s device in the code.
4. Install the required libraries that are mentioned in the code.
5. Uncomment the lines install.packages() to install the libraries.
6. The user can select from the dropdown list and if not then it will display all the countries together in the first graph.
7. The second graph displays a tooltip when hovered over any point in the line to show year and number of refugee intake for that country.
8. The user can click on play to view the animation on the third graph and also slide the slider to view information of the refugees settled in that state.
9. When the mouse hovers over the state it will display the State Name and the number of refugees that settled there.
10. The user can hover over the third graph to display Year, Work Status and Total Refugees in a tooltip.
11. The user can hover over the fourth graph to display Year, PR Status and Total Refugees in a tooltip.
12. The user can click on “Show/Hide All Years Together” in the animated line graph to display all country’s data together and click it again to view the per year graph.

13. The user can slide the slider to view the year's information.

14. The user can click on the "Play" button to view the animation made per year.

15. The user can hover over the animated graph to display Year, Month and Unemployment rate in a tooltip.

16. The user can hover over the correlation graph to display the names of the variable and the relation density in a tooltip.

17. The user can hover over the heatmap graph to display Country and Refugee Count in a tooltip.

18. The user can also select the year that he wants to display in the Select Year Box for the heatmap.

Conclusion

A distinct conclusion cannot be made as more data is required for that analysis. We could only see that refugees granted asylum and Annual Change of US, the correct number of total refugees and the total number of refugees that are working now is yet to be determined. The unemployment rate is discovered but that does not say anything about refugees, so those graphs are for refugees as well as citizens. A refugee employment dataset is needed for the proper analysis.

What I learned about the project is that the datasets that I chose do not actually give enough information to come to a rigid conclusion for the project. I need to find out more datasets and combine some of the datasets from the US Bureau of Labor Statistics together to find out the answer to the second question. The website has a lot of different datasets, and those datasets need to be combined to a common pivot point to come to any conclusion.

Bibliography

· UNHCR Microdata Library. (n.d.). Refugee and Returnee Data [Data set]. UNHCR. https://microdata.unhcr.org/index.php/catalog/681/study-description

· Inter-university Consortium for Political and Social Research. (n.d.). Eurobarometer 67.2: Citizenship in the European Union, October-November 2007 [Data set]. OpenICPSR. https://www.openicpsr.org/openicpsr/project/168641/version/V2/view;jsessionid=C85B31ADCB 9FE7D38A5B2BB055ACD1D1

· Inter-university Consortium for Political and Social Research. (2019). 2019 American Society of Reproductive Medicine (ASRM) Public Use File [Data set]. OpenICPSR. https://www.openicpsr.org/openicpsr/project/168641/version/V2/view;jsessionid=C85B31ADCB 9FE7D38A5B2BB055ACD1D1?path=/openicpsr/168641/fcr:versions/V2/2019-ASR_Public_Use_File.dta&type=file

· U.S. Bureau of Labor Statistics. (n.d.). Job Openings and Labor Turnover: Total Nonfarm [Graph]. Bureau of Labor Statistics. https://www.bls.gov/charts/job-openings-and-labor-turnover/unemp[1]per-job-opening.htm

· U.S. Bureau of Labor Statistics. (n.d.). Labor Force Statistics from the Current Population Survey [Data set]. Bureau of Labor Statistics. https://data.bls.gov/PDQWeb/ln

· U.S. Bureau of Labor Statistics. (n.d.). Survey Output [Data set]. Bureau of Labor Statistics. https://data.bls.gov/pdq/SurveyOutputServlet

· Macrotrends. (n.d.). U.S. Refugee Statistics - Historical Data [Data set]. MacroTrends. https://www.macrotrends.net/countries/USA/united-states/refugee[1]statistics#:~:text=U.S.%20Refugee%20Statistics%20-%20Historical%20Data%20%20,%20%209.09%25%20%2058%20more%20rows%20

· Hauser, C. (2023, January 20). US State Department creates new refugee program called "Welcome Corps". The Guardian. https://www.theguardian.com/us-news/2023/jan/20/us-state[1]department-new-refugee-program-welcome-corps

· The New York Times. (n.d.). Refugees and Displaced People [Webpage]. The New York Times. https://www.nytimes.com/topic/subject/refugees-and-displaced-peopl

Appendix