

Project Report

R against the Machine

4/27/2019

Problem Statement and Background

Since 1975, the United States government has welcomed over 3 million refugees for resettlement from all over the world, allowing them to build new lives and contribute to the economies of all 50 states (UNHCR, 2018). In recent years, sharp cuts to refugee resettlement quotas have sparked debate over the program's national security implications as well as the ability of refugees to integrate into communities across the country. Hence, the key question we ask is: what does the most recent available data say about refugee integration outcomes in the United States?

We attempt to answer this question by looking at the Annual Survey of Refugees 2016 (ASR), a nationally representative survey of refugees who were resettled in the US between FY 2011 and FY2015. The survey was carried out by the Office of Refugee Resettlement at the U.S. Department of Health and Human Services (HHS) and it offers a window into respondents' first five years in the US and shows the progress that refugee families made towards learning English, participating in the workforce, and establishing permanent residence. The dataset can be downloaded as a STATA file and includes information on 1,500 households and more than 4,000 individuals.

All survey data is close-ended and multiple choice question types to capture demographics and frequency calculations (No, Yes, Not relevant, Refused).

We believe the survey data to be relevant to inform decision-making surrounding the future of the refugee resettlement program in the United States which will help consider what programs are the most effective and for what particular group of people.

Presently, there is a huge gap in capturing refugee data with very few agencies collecting information; other than ASR, the only other data sources are the Census Bureau data: American Community Survey (ACS) and the New Immigrant Survey.

Methods

Our team downloaded the data, cleaned and filtered appropriate variables that allowed us easily and quickly tell the story of refugee integration across the United States. Using data visualization tools, we created a dashboard to provide a general idea of the geographic dimension of refugee resettlement and the characteristics of welcoming communities across the country. Our data visualizations and tables essentially "tell a story" about the outcomes of refugees resettled in the US in recent years. We created an interface that allows users to navigate the different visualizations and explore the various dimensions of refugee integration, comparing and contrasting the outcomes attained in these areas. The dashboard is a dynamic tool that provides concise explanations and guides the user/reader to understand the answers we have found to the research question.

Tools

We used the following packages: ggplot2, recipes, tidyverse, pdp, utf8, skimr, foreign, readr, haven, scales for data manipulation and visualization. Additionally, we used shiny, shindashboard, and shinythemes for our dashboard.

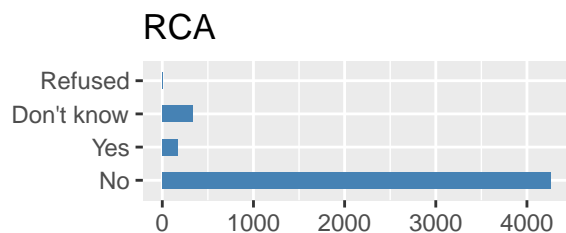
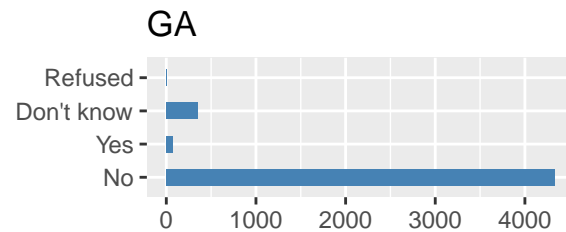
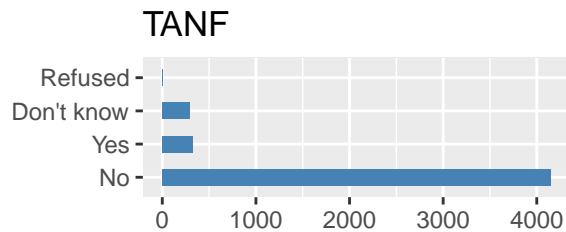
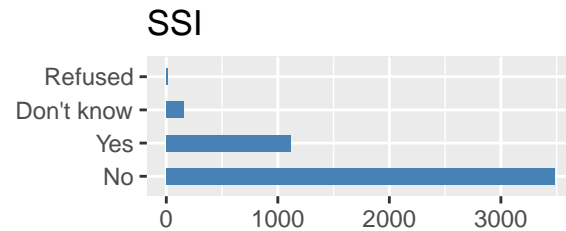
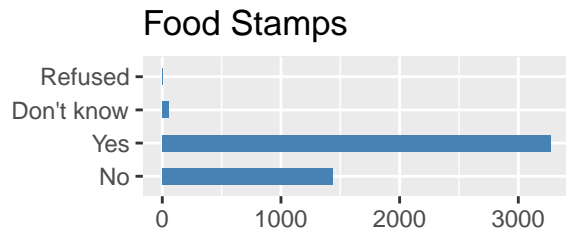
Describe the tools that your team used and why. Justify the tools used in terms of the problem itself and the methods your team was aiming to utilize. Tools can include anything from packages used for data wrangling

and visualization to machine learning and statistical processing. – How did you employ the tools used? What features worked well and what did not? – Describe any tools that you tried and ended up not using. What was the problem? Briefly, what could be improved in these packages to make them more functional?

Results

Employment and Schooling

Benefits



Demographics

– Give a detailed summary of the results of your work. Here is where you specify the exact performance measures you used. Usually there will be some kind of accuracy or quality measure. There may also be a performance (runtime or throughput) measure.

– Please use visualizations and tables whenever possible. Include links to interactive visualizations or websites if you built them.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.