Domestic & International Adoption Trends

General Assemby – Data Science 6

By Matthew Reese

**Problem Statement**

Every year, tens of thousands of children around the world are put up for adoption. There are myriad reasons why this occurs, but ultimately, the crux of the issue is finding these children new homes that enable them to develop optimally.

**Hypothesis**

There are a number of countries that have created international adoption policies that facilitate moving children to environments where they can achieve their full potential. These countries can provide a model of best practices.

Domestically, there are a number of states that may also serve as models for adoption practices throughout the United States.

**Data**

International adoption statistics are taken from the United States Department of State website. Files were obtained in PDF format and parsed into CSV files both manually.

Domestic adoption statistics are taken from the United States Department of Health and Human Services website. Again, data was obtained in PDF format, but the structure of the files enabled them to be translated into a CSV file through a PDF parser.[[1]](#footnote-1)

United States 2013 population estimates are taken from the 2010 United States census. The census bureau offers an API that presents selected data in a CSV format.

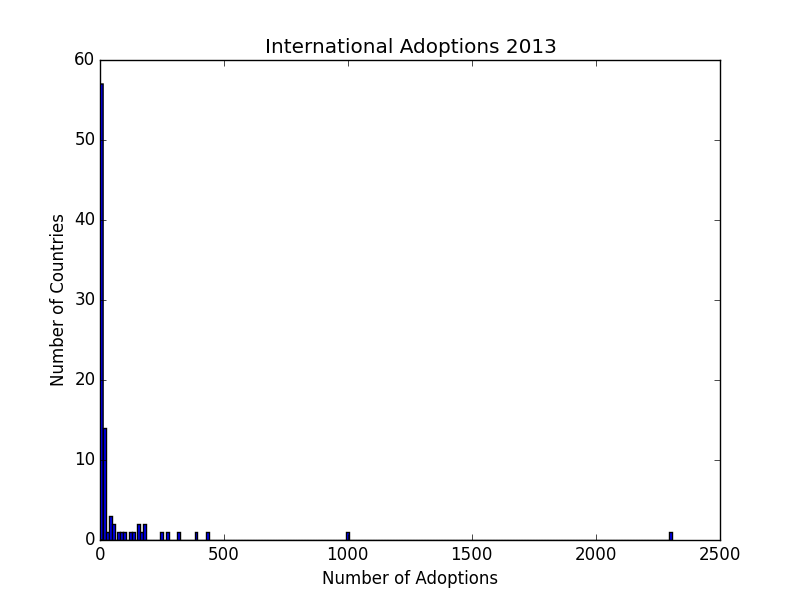
All CSVs were read into Python Pandas Dataframes.

**Data Exploration**

In exploring the international adoption statistics, I quickly learned that there was main distinction: some countries participated in the Hague Convention on the Protection of Children and Cooperation in Respect of Intercountry Adoption, hereafter referred to simply as the Hague Convention, while others did not. The Hague Convention standardizes practices such as itemizing fees and providing the prospective adoptee’s medical records, but adoptions from all countries must follow the same standards in regards to parent suitability and immigration eligibility (state.gov citation).

Ultimately, the data available for international adoptions was relatively limited. The entire sample had data for the country name and whether or not the adoption was finalized domestically or abroad. Countries that participate in the Hague Convention also have data concerning the average time to completion for adoptions from a certain country and the median associated fees. I also added a column calculating what proportion of adoptions were finalized domestically versus abroad.

Interestingly, the majority of countries listed had a relatively small number of adoptions. Indeed, 75% of the countries had less than 24 adoptions to the United States in 2013.

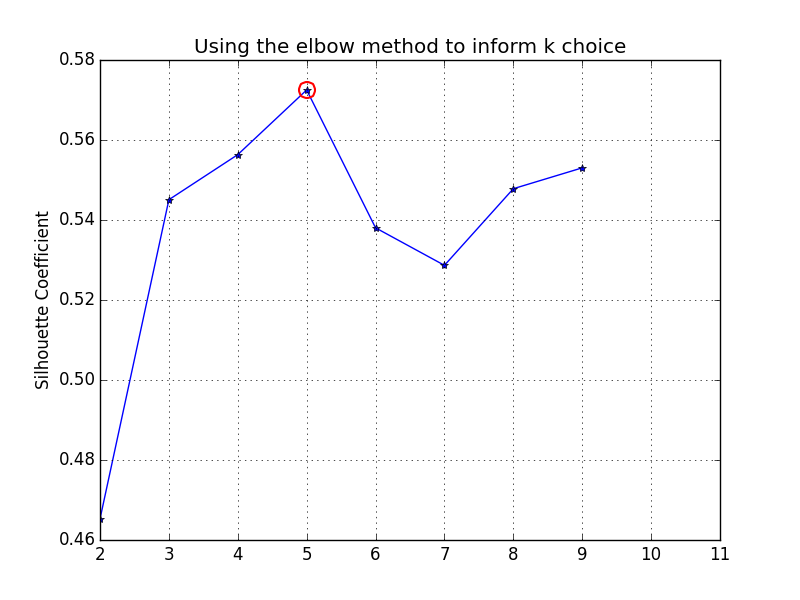


The above graph shows two obvious outliers: China and Ethiopia. China clearly has a much larger population than any other country on this list, so a more in depth analysis would standardize the data using total population figures. However, I did not expect to see Ethiopia which such a relatively high number.

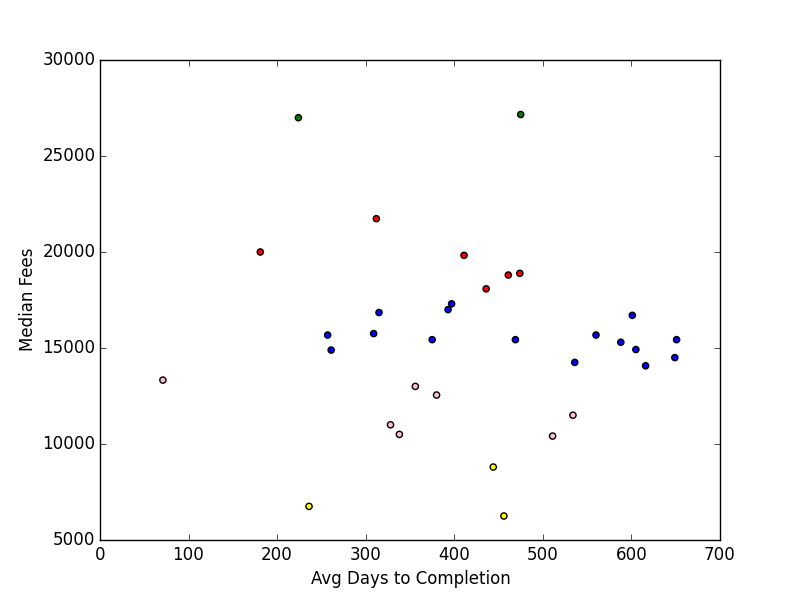
A quick search led me to a number of articles decrying Ethiopia’s international adoption standards. The opening line of E.J. Graff’s article puts things in perspective: “When large amounts of money are exchanged between a wealthy country and a poor country, here’s what happens.” Graff goes on to describe instances where birth parents were unaware they were giving up their children completely; rather, they were under the impression that their children would be given a better upbringing in the United States, but would then return home to support their biological family.[[2]](#footnote-2)[[3]](#footnote-3)

These stories highlight the importance of money in international adoption. Unfortunately, I could only find data related to adoption fees for countries participating in the Hague Convention. Still, I wanted to see if there was a commonality in regards to fees. As such, I thought a KMeans Cluster analysis would be appropriate.

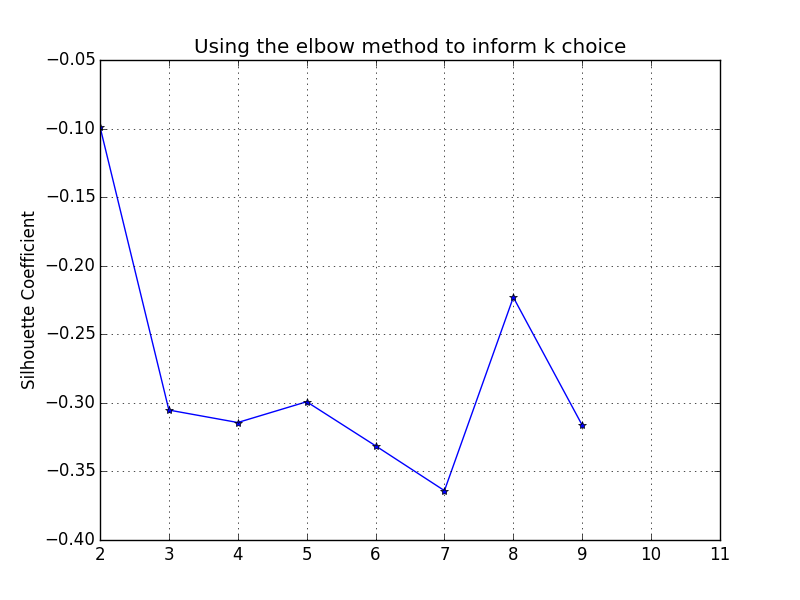
To start, I wanted to see what the optimal number of clusters would be. I subsequently modeled the cluster analysis based on the proportion of adoptions finalized abroad vs. domestically, the median fees, how long the process took, and the total number of adoptions. I compiled silhouette scores for k values between 2 and 9 and produced the following graph:



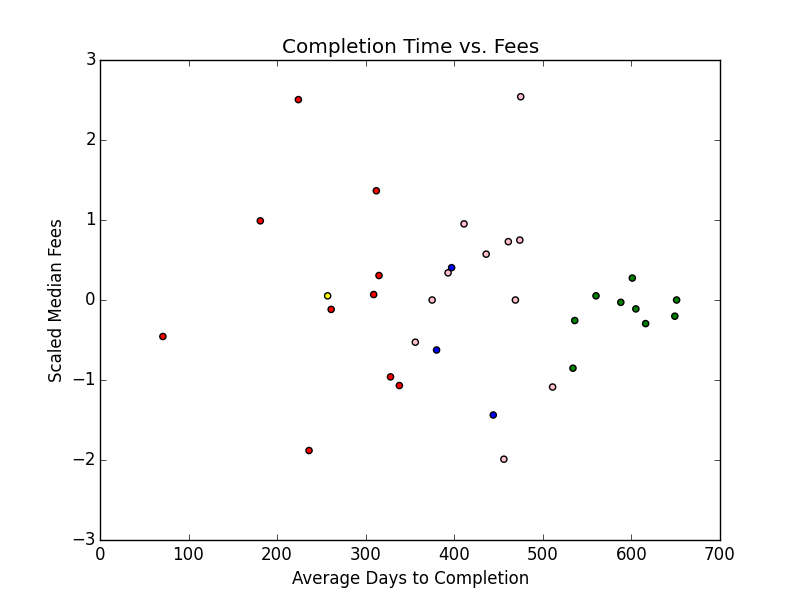
A K value of 5 seemed to produce the best silhouette coefficient. Using 5 clusters produced the following visualization:



A quick look at the data, however, made me realize that since fees were such a higher magnitude than any other information, they likely skewed the data. As a result, I decided to center and scale the fees. Upon doing so, I produced the following silhouette coefficient graph:



Unfortunately, the negative silhouette coefficients imply that there is significant overlap between the clusters:



Ultimately, I had to conclude that based on the information at hand, I was unable to determine any significant groupings amongst countries that participate in the Hague Convention. As a result, I decided to refocus my efforts on domestic adoption.

1. <https://pdftables.com>. [↑](#footnote-ref-1)
2. <http://prospect.org/article/dont-adopt-ethiopia> [↑](#footnote-ref-2)
3. <http://www.psmag.com/politics-and-law/they-steal-babies-dont-they-international-adoption-schuster-institute-95027> [↑](#footnote-ref-3)