

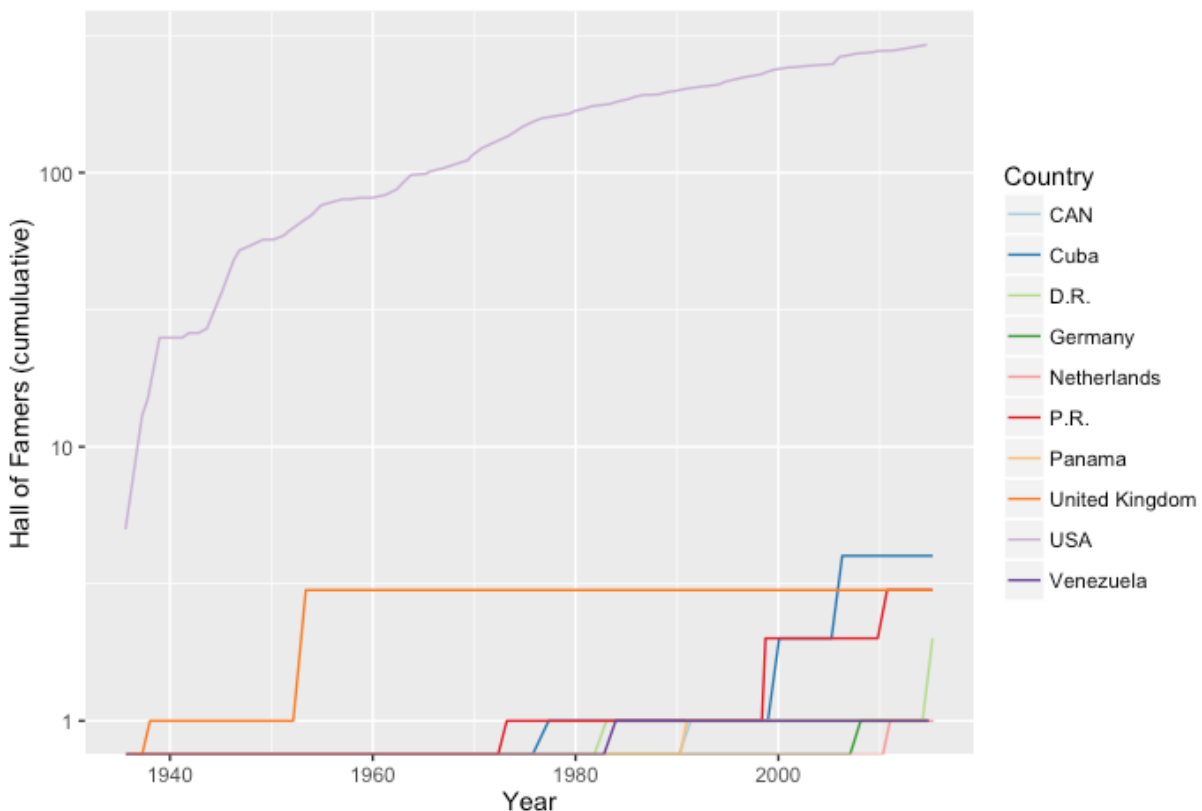
### SLEI Research Analyst Data Task

**Instructions:** For this task, you may use any tools that you like. Please share all of your results, answers, and code in a way that is clear and professional. Furthermore, be prepared to edit and run some of your code “on the fly” with screen sharing enabled.

Here are the data: [https://drive.google.com/file/d/1OMiPRSNuLjogCu0GNAKK66o-4hA8pM65/view?usp=share\\_link](https://drive.google.com/file/d/1OMiPRSNuLjogCu0GNAKK66o-4hA8pM65/view?usp=share_link)  
It is Sean Lahman's Baseball Databank, with information like salary, players who have been admitted into the Hall of Fame, player weight, and of course, baseball performance.

#### Q1

You've been asked to evaluate the Hall of Fame status of players as a result of country of origin for SLEI Faculty Advisor, Paul Oyer (not likely, but he does love baseball). You decide to make a plot of players in the hall of fame over time, grouped by each country. Your end result should look something like this:



#### Q2

After making this graph and giving it back to Paul, what questions and/or observations would you bring up? Are there any interesting research questions or findings you would explore further?

#### Q3

The data in the “Baseball Databank” currently exist as multiple csv files. Create a plan to centralize these data. What data management tools would you leverage (or build) in order to house and easily manage all of these growing data?