



U.S. Securities and Exchange Commission

Analyzing IAR Registration and Employment Trends

- Presented by: James Rios
- [Project Website](#)





Project Overview

- Goal: Analyze and visualize data from the IAPD system.
- Data Source: XML files from Investment Adviser Public Disclosure (IAPD).
 - Focus Areas:
 - Employment history
 - Registration status
 - Exam results
 - Disclosures



Data Engineering Workflow

- Python Pipeline:
 - Parsed XML data, normalized tables (employment, registration, exams, etc.)
 - R Data Wrangling:
 - Cleaned employer names, removed punctuation/fillers
 - Grouped by root token frequency



Standardizing Organization Names

- Removed suffixes, punctuation, filler words
- Grouped names by root tokens and frequency
- Example: 'Morgan Stanley Wealth Management LLC' → 'morgan stanley'

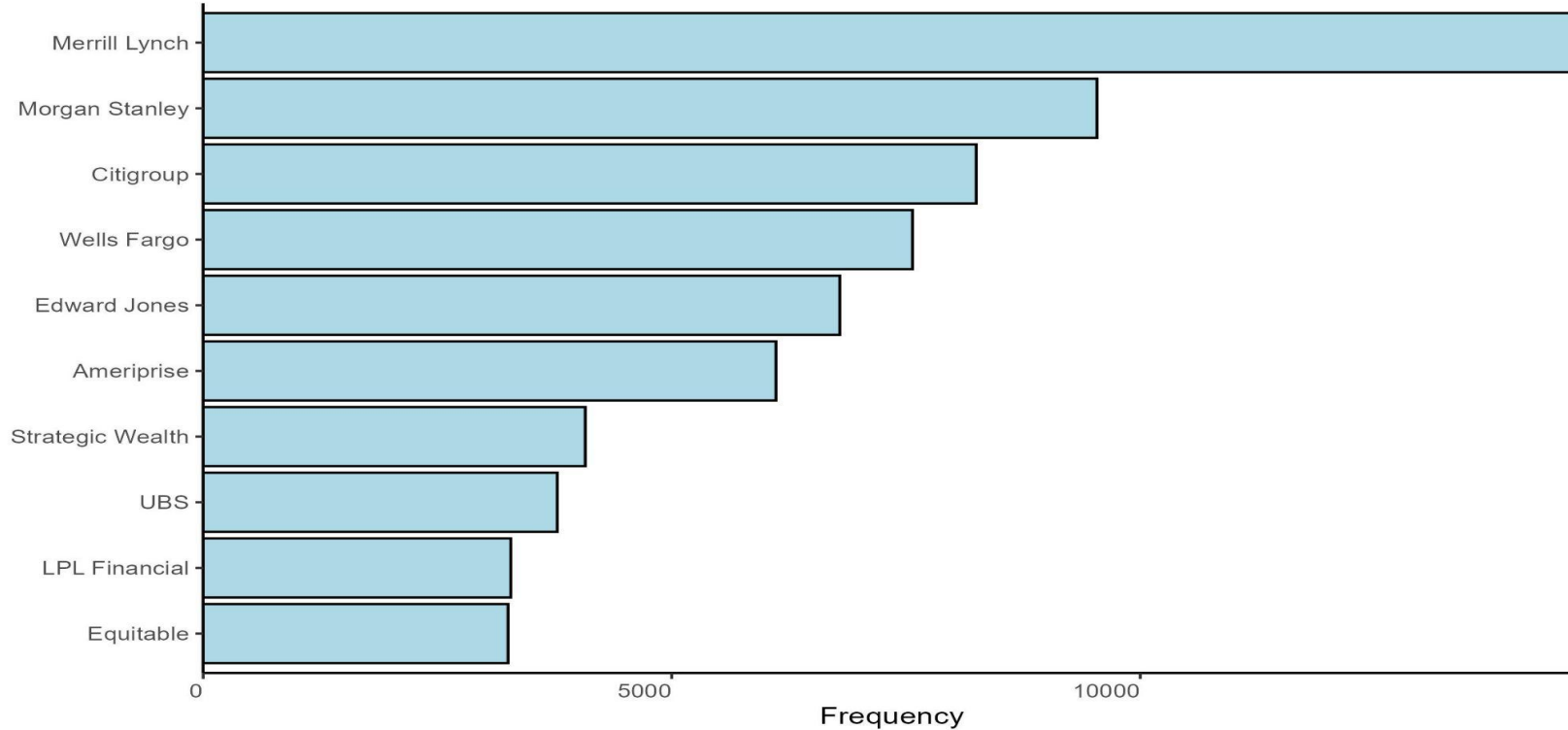


Data Enrichment

Derived Features:

- Avg. registration duration
- Total registrations per firm/individual
- City and state coverage
- Mapped state info to census regions

Top 10 Most Common Organizations

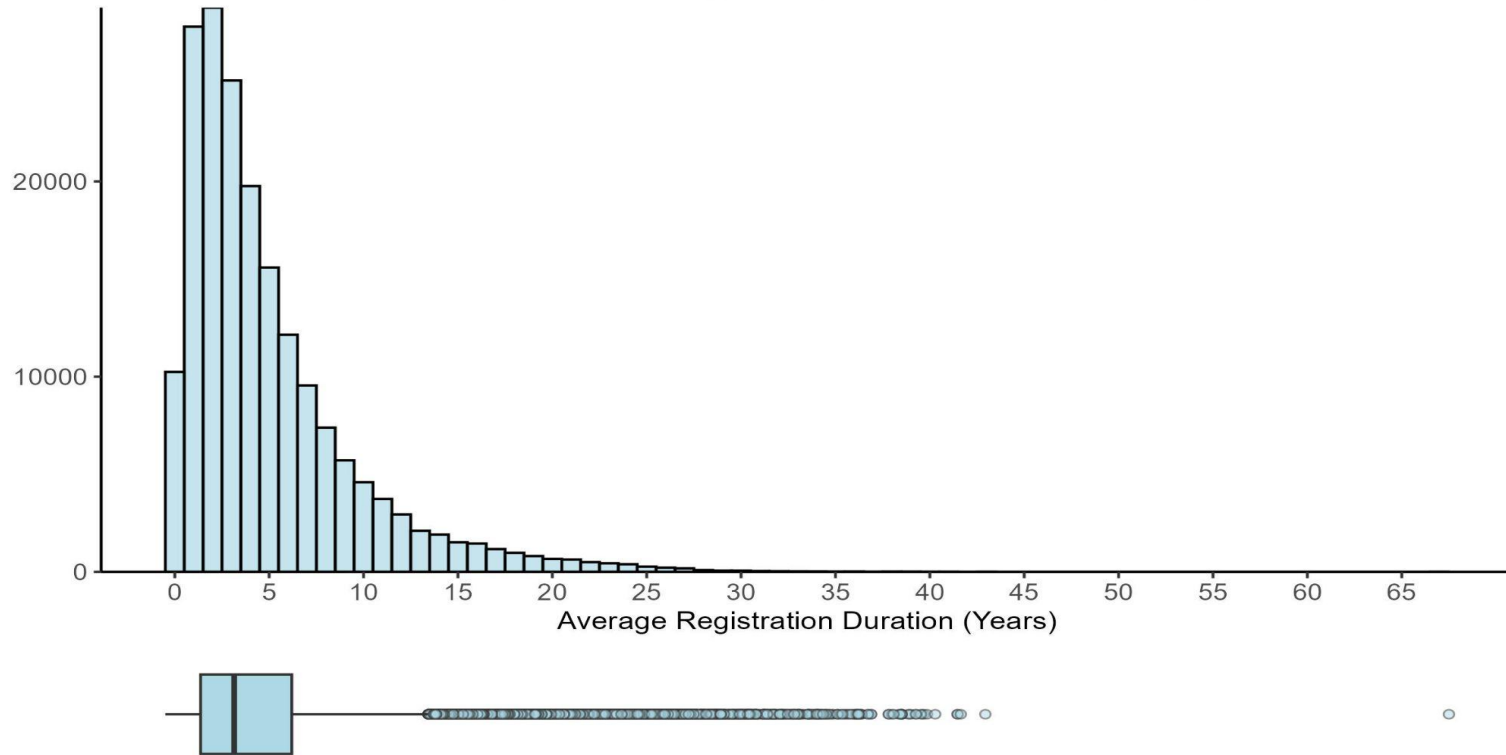


Graph 1: Bar Chart

This bar chart presents the ten investment advisory firms with the highest number of registered Investment Adviser Representatives (IARs). It provides a straightforward frequency distribution, showing raw counts of representatives per firm.



Histogram of Average Registration Duration (Years)

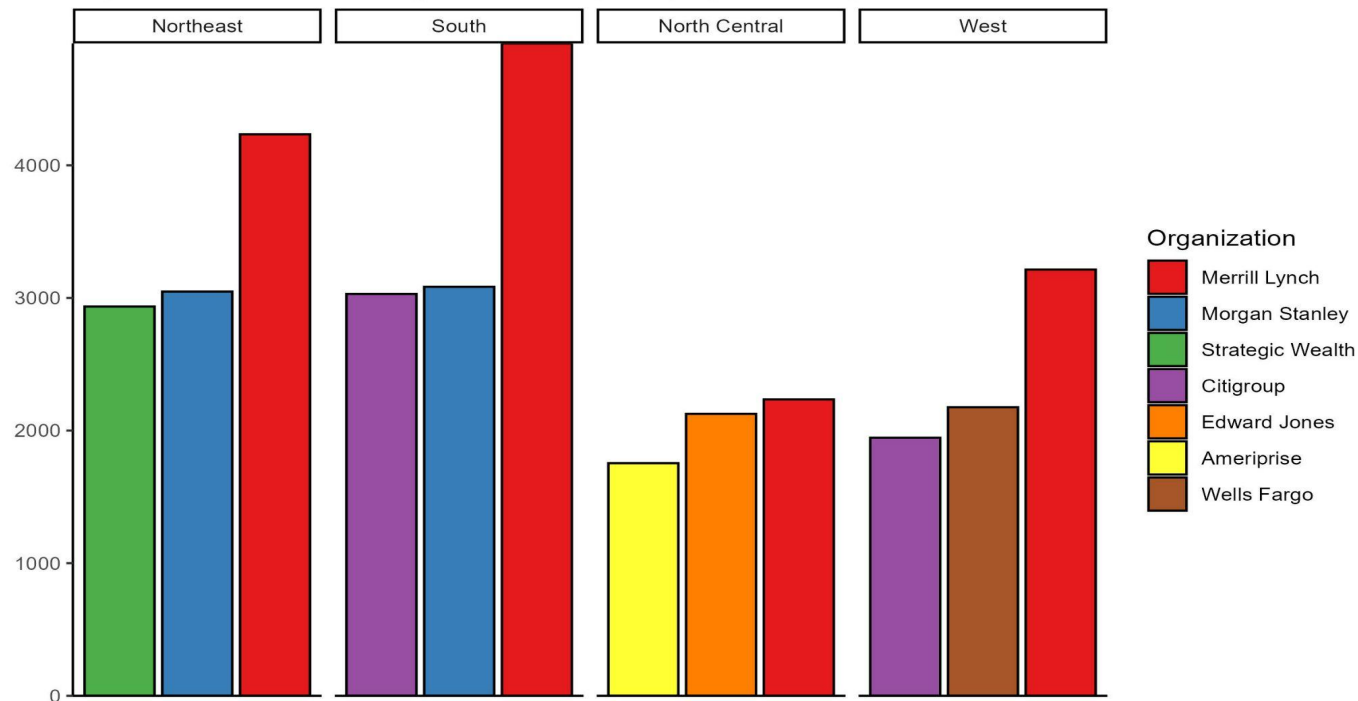


Graph 2: Histogram + Boxplot

This combined plot includes a histogram and a boxplot to analyze how long individuals remain registered as IARs. The histogram shows the distribution of tenure durations, revealing common patterns such as clustering around early-career years. The median is 3.58 years, mean is 5 years, and the maximum observed tenure is 66.83 years.



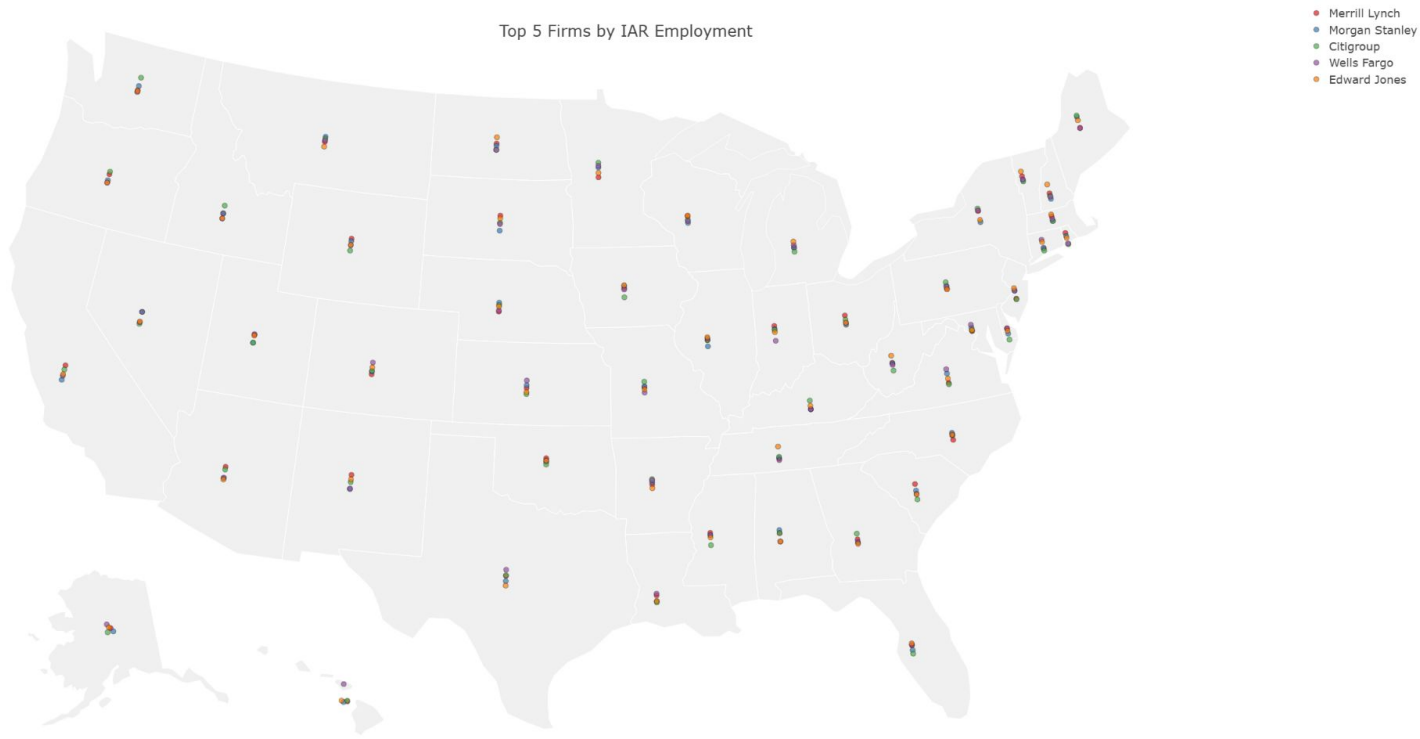
Top 3 Most Common Organizations by Employment Region



Graph 3: Bar Chart by Region

This bar chart shows how Investment Adviser Representatives (IARs) are distributed across the four U.S. Census regions—**Northeast, South, North Central, and West**—with counts broken down by the top firms in each region. The **South** leads in overall IAR representation, with a **median firm count of 3,084** and a **mean of 3,678**, suggesting both strong presence and a few very large orgs pulling the average higher. The **Northeast** follows closely with a **median of 3,048** and **mean of 3,406**. In contrast, the **North Central** and **West** regions are, with medians of **2,126** and **2,176** respectively, and means of **2,038** and **2,445**.





Graph 4: [Bubble Map](#) (Plotly)

The interactive bubble map uses geographic coordinates and bubbles to represent the distribution of IARs employed by the top five firms across U.S. states. Tooltips on the map display statistics such as the total registrations and city coverage for each firm in a state.





Resources & Link

- [Data Source](#)
- [Project Website](#)

Thank You!