INDIANA UNIVERSITY
BLOOMINGTON

INFO-I 590: TOPICS IN INFORMATICS VISUALIZATION

Analyzing U.S. Immigration and H-1B Visa Trends (2020 - 2024)

Authors:

1. Mohammad Fahad Shahul Hameed
2. Ramcharan Suresh
3. Mouryan Jayasankar
4. Kishore Kumar Ramkumar
5. Santhosh Kumaar Rajan

Abstract

This project takes a closer look at how the H-1B visa program has been working in recent years. We analysed data from 2020 to 2024, along with older records from USCIS (2009–2023) [3], [5] and consultant placement data from 2023 [1]. We pulled in data from different sources to dig into what's really going on with H-1B visas—from how much people are getting paid to where jobs are located and how policies might be shaping outcomes. The goal here isn't just to show trends—it's to give people who make decisions a clearer picture of what's happening, backed by real numbers, so they can shape policies that are actually fair and transparent.

I. Introduction

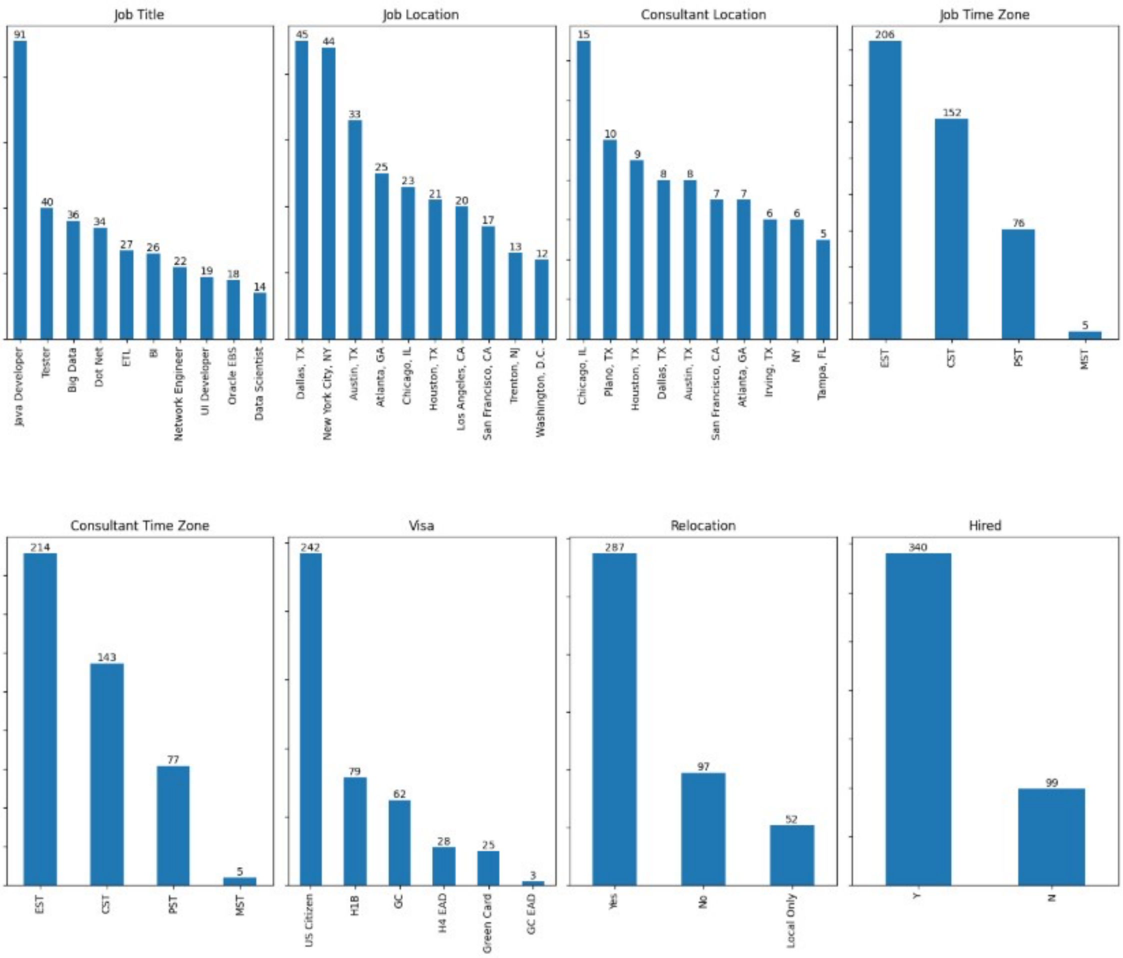
The H-1B visa has become a go-to for U.S. companies that need skilled workers—especially in areas like tech, healthcare, and engineering—where it's tough to find enough qualified people locally. But as the program has expanded, so have the questions about whether it's being used fairly and as intended. Are workers being paid fairly? Are applications being approved consistently? And how do different types of employers manage their foreign hires? This project digs into those questions by analyzing recent H-1B Labor Condition Applications, long-term USCIS data, and consultant placement records [1]. We've turned that complex data into simple, interactive visuals that highlight important trends—like wage gaps, employer behavior, and regional differences—to help support smarter immigration policies and fairer hiring practices [6], [7].

1.1. Motivation

For the H-1B system to work the way it should, fair labor practices need to be at the core. But when we looked at the data—from both USCIS records [3], [5] and consultant job reports [1], we saw red flags: big differences in offered vs. prevailing wages, inconsistent job classifications, and unclear work locations. These patterns suggest that not every employer is playing by the same rules. That's what drove us to take a closer look. Our goal is to break down these patterns—especially the differences between consulting firms and other employers—and show them clearly through careful analysis and easy-to-understand visuals. We hope this work sparks more informed conversations and support a more balanced, transparent approach to skilled immigration [6].

1.2. Existing Work

1.2.1. Consultant Job Placement Analysis – Mahipal Kumar Singh (Kaggle, 2023) [1]



This dataset explores the employment landscape for consultants, particularly those on H-1B and related visas. The visualizations provide a comprehensive breakdown across multiple dimensions such as job titles, locations, time zones, visa categories, and relocation preferences.

Techniques and Effectiveness:

- **Bar Charts** are effectively used to display categorical distributions such as job roles (e.g., Java Developer, Tester, Big Data), consultant vs. job locations, visa types (H1B, GC, OPT, etc.), and relocation willingness.
- The **use of multiple subplots** arranged in a grid layout enables side-by-side comparison, allowing viewers to spot mismatches (e.g., consultants in CST vs. job openings in EST).
- The visa-type distribution chart is particularly effective in showing that U.S. citizens and H-1B holders make up a majority of placements, supporting workforce diversity analysis.

- **Relocation and hiring bar charts** clearly demonstrate employment trends, with a significantly higher proportion of consultants being open to relocation and successfully hired.

Limitations:
While the charts are clean and interpretable, they are static. Interactivity (such as filters by visa type or job location) could enhance user-driven exploration. Nonetheless, this work lays a strong foundation for understanding job market alignment for consultants in the U.S.

1.2.2 H-1B Employer Summary Visualization – USCIS H-1B Employer Data Hub [5]

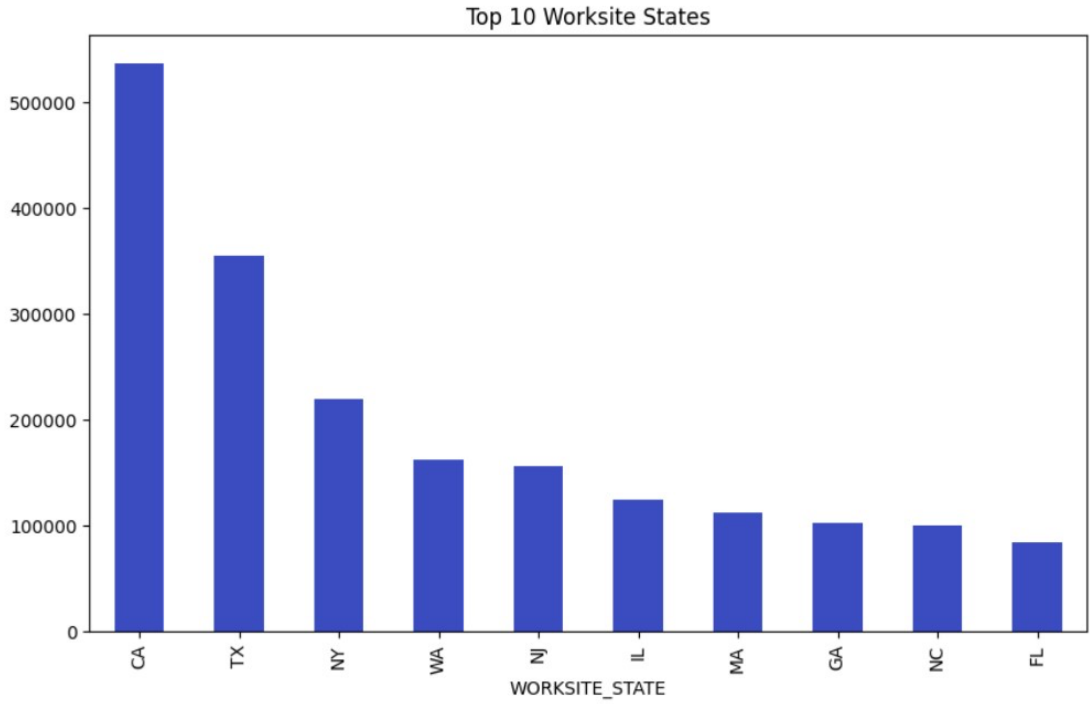
This visualization, presented through an interactive dashboard, showcases the top 100 employers with approved H-1B petitions. The choropleth-style map of the U.S. highlights state-wise concentrations, while a horizontal bar chart ranks companies by number of approved beneficiaries.

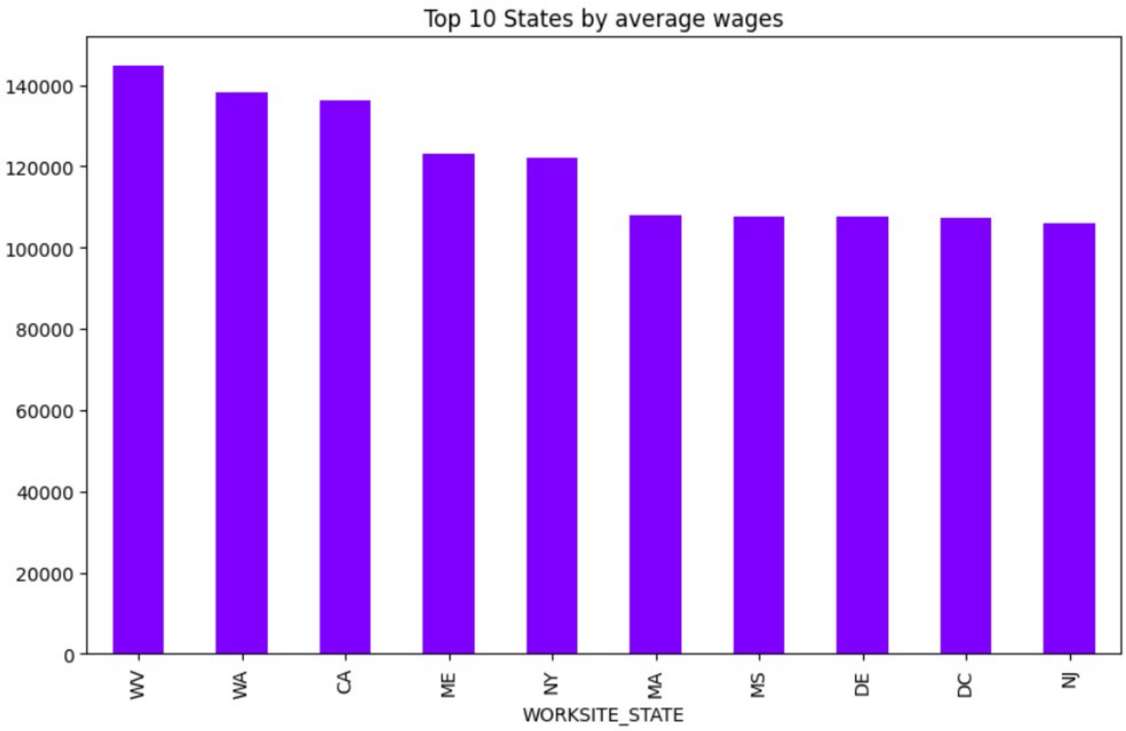
Techniques and Effectiveness:

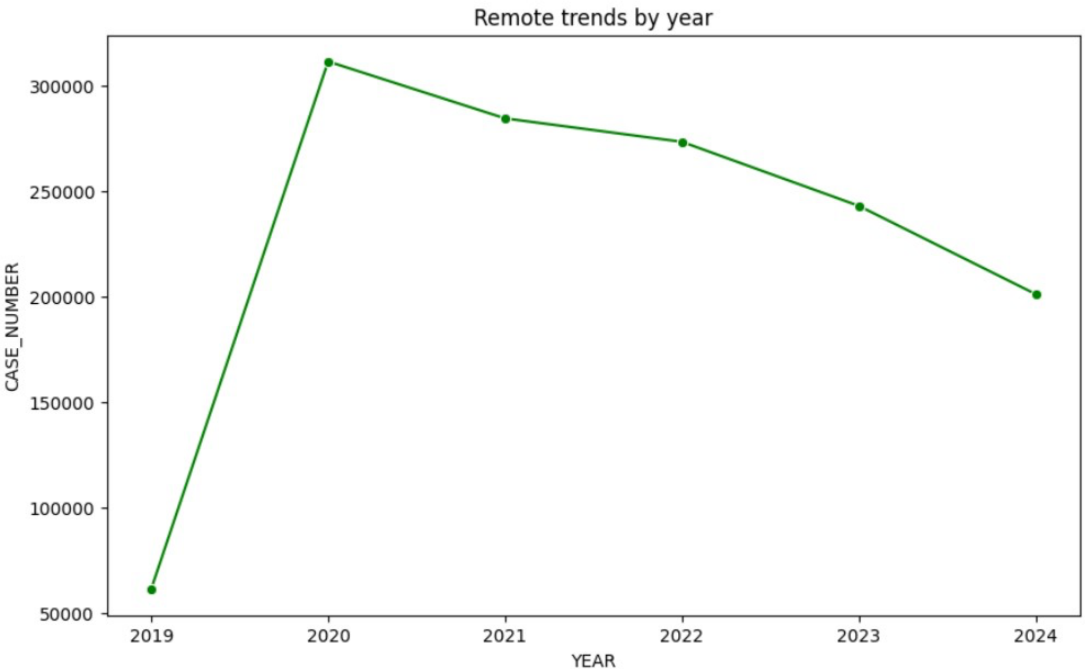
- The **interactive choropleth map** is particularly effective in visualizing geographic concentration of H-1B workers. States like California, Texas, and New York emerge as major hubs.
- The **bar chart of top petitioning employers** (e.g., Amazon, Infosys, Cognizant, Google) immediately communicates which corporations dominate the H-1B landscape.
- Tooltips and filters (available in the full interactive version) enhance exploratory analysis, allowing users to drill into individual employer histories.

Discussion:
This work is powerful in scale and authority, drawing from official USCIS approval data. It excels in usability and accessibility. However, it is limited in analytical depth it focuses more on "who" and "where" than "why" or "how," lacking contextual layers like salary trends or job roles.

1.2.3 H-1B Application Trends and Wages – Zongobian (Kaggle, 2024) [2], [3]







This analysis uses cleaned H-1B LCA filings to investigate worksite distributions, wage patterns, and remote work trends over time. The visualizations provide clear insights into both volume and value-based metrics across states and years.

Techniques and Effectiveness:

- A **bar chart of top 10 worksite states** shows California, Texas, and New York as consistent leaders in application volume, indicating high industry demand.
- Another **bar chart ranking states by average wages** highlights high-paying regions like West Virginia, Washington, and California, revealing wage disparities that might influence migration patterns or job selection.
- The **line graph of remote trends by year** effectively illustrates how remote job approvals peaked during 2020 (pandemic year) and declined steadily after, revealing a shift in employer flexibility over time.

Discussion:
These visualizations are impactful due to their simplicity and clarity. The use of consistent color schemes (e.g., blue for volume, purple for wages, green for time trends) enhances thematic coherence. Though static, these charts provide important longitudinal and geographic insights. When combined with filters by job role or employer, these visuals can guide both policy discussions and individual decision-making.

1.3. Contribution

Our project fills some important gaps in how H-1B visa trends are currently analyzed and presented. A lot of existing tools—like USCIS dashboards or Kaggle visualizations—offer broad overviews, but they often miss the deeper context. Here's how our work takes things further:

1. **Richer, Multi-Layered Analysis:** We don't just look at raw numbers. We connect wage data (offered vs. prevailing), employer types (dependent vs. non-dependent), legal representation, and filing timelines. This lets us explore how specific employer behaviors and job roles impact visa approvals—something that's usually lost in high-level summaries.

2. **Interactive, Visual-First Design:** Instead of static bar charts or basic tables, we built dynamic visuals that let users filter by employer, industry, job title, and country. Tools like choropleth maps, Sankey diagrams, and heatmaps help bring patterns to life and make the data easier to explore.

3. **New Insights You Don't See Elsewhere:** By comparing certified and denied applications across wage ranges and job categories, we surfaced salary gaps and approval trends that aren't obvious in official reports [3], [5]. We also looked at the effect of H-1B dependency and willful violator status—two factors often overlooked in most dashboards.

4. **Bridging gaps in existing works:** While most studies focus only on employers or just on countries, we connect the dots across geography, occupations, and policy. For example, we looked at how legal representation ties into wage offers, and how industries differ in their use of foreign labor. These kinds of comparisons add much-needed depth.

5. **Real-World Relevance:** Our visuals don't just look good—they can actually help with policy and oversight. They highlight potential compliance issues with wage standards, flag patterns that might suggest exploitation, and support conversations around fairer, more transparent visa practices [6], [7].

II. Data and Methods

This project relies on careful gathering, cleaning, merging, and visualization of data to identify trends in H-1B visa applications. We pulled together multiple publicly accessible datasets from sources like the U.S. Department of Labor and USCIS, thoroughly preprocessing the data and thoughtfully designing visuals to highlight key patterns. Before we created any visualizations, the team extensively prepared and standardized the data to ensure accuracy and clarity. This involved close teamwork and multiple revisions to bring all datasets into a unified analytical framework. Every visualization in this project represents careful planning, data analysis, and teamwork, built from thousands of records and numerous variables.

```python
import pandas as pd

# Columns to use
columns = [
    "CASE_STATUS", "RECEIVED_DATE", "DECISION_DATE", "VISA_CLASS",
    "JOB_TITLE", "SOC_CODE", "SOC_TITLE", "FULL_TIME_POSITION",
    "WAGE_RATE_OF_PAY_FROM", "WAGE_UNIT_OF_PAY",
    "WORKSITE_STATE", "WORKSITE_CITY", "EMPLOYER_NAME",
    "PREVAILING_WAGE",
    "NAICS_CODE"
]


file_path = "Combined_LCA_Disclosure_Data_FY2020_to_FY2024.csv"

# Read data
df = pd.read_csv(file_path, usecols=columns, low_memory=False)

# Drop missing important rows
df = df.dropna(subset=["CASE_STATUS", "WAGE_RATE_OF_PAY_FROM", "WAGE_UNIT_OF_PAY"])

# Normalize wage to annual salary
def convert_to_annual(row):
    wage = float(row["WAGE_RATE_OF_PAY_FROM"])
    unit = row["WAGE_UNIT_OF_PAY"]
    if unit == "Hour":
        return wage * 40 * 52
    elif unit == "Week":
        return wage * 52
    elif unit == "Bi-Weekly":
        return wage * 26
    elif unit == "Month":
        return wage * 12
    elif unit == "Year":
        return wage
    else:
        return None

df["ANNUAL_WAGE"] = df.apply(convert_to_annual, axis=1)

# Filter for H-1B only
df = df[df["VISA_CLASS"] == "H-1B"]

# Clean text columns
text_cols = ["CASE_STATUS", "VISA_CLASS", "JOB_TITLE", "SOC_CODE", "SOC_TITLE", "WORKSITE_STATE", "WORKSITE_CITY", "EMPLOYER_NAME"]
for col in text_cols:
    df[col] = df[col].astype(str).str.strip().str.upper()

# Remove unrealistic salaries
df = df[(df["ANNUAL_WAGE"] >= 20000) & (df["ANNUAL_WAGE"] <= 500000)]

# Save cleaned data
df.to_csv("preprocessed_lca_data.csv", index=False)
```

2.1 Dataset Description

To understand the H-1B visa landscape, we pulled together several large datasets covering applications, employer filings, wages, job roles, outcomes, and applicant locations. The main one was the LCA Disclosure Data (FY2020–FY2024) [4], with over 5 million rows and 35+ columns like CASE_STATUS, JOB_TITLE, SOC_CODE, EMPLOYER_NAME, WAGE_RATE_OF_PAY, PREVAILING_WAGE, VISA_CLASS, WORKSITE_STATE, and RECEIVED_DATE. It helped us explore wage trends, employer behavior, and job category patterns. We added the USCIS H-1B Employer Data Hub (2009–2023) [5] for historical approvals and denials by employer and state, the TRK I-129 FOIA Registration Dataset (FY2021–FY2024) to distinguish initial vs. continuing petitions, and the H-1B Country of Birth & Gender Report (FY2019) [8] for demographic context. Cleaning and merging everything took time—Santhosh and Mouryan handled de-duplication, invalid rows, and missing values, while Fahad standardized fields like RECEIVED_DATE and WAGE_RATE_OF_PAY. Kishore and Ramcharan aligned key identifiers such as SOC_CODE, WORKSITE_STATE, and EMPLOYER_NAME across datasets. Each file was thoroughly reviewed for consistency before we moved into visualization.

	CASE_STATUS	RECEIVED_DATE	DECISION_DATE	VISA_CLASS	JOB_TITLE	SOC_CODE	SOC_TITLE	FULL_TIME_POSITION	EMPLOYER_NAME	NAICS_CODE	WORKSITE_CITY	WORKSITE_STATE	WAGE_RATE_OF
0	CERTIFIED	2019-09-25	2019-10-01	H-1B	APPLICATION ENGINEER, OMS [15-1199.02]	15-1199	COMPUTER OCCUPATIONS, ALL OTHER	Y	JO-ANN STORES, INC.	451120.0	HUDSON	OH	
1	CERTIFIED	2019-09-25	2019-10-01	H-1B	BI DEVELOPER II	15-1132	SOFTWARE DEVELOPERS, APPLICATIONS	Y	DENKEN SOLUTIONS INC.	541512.0	BRENTWOOD	TN	
2	CERTIFIED	2019-09-25	2019-10-01	H-1B	QUALITY ENGINEER	17-2141	MECHANICAL ENGINEERS	Y	EPITEC, INC.	541511.0	DEARBORN	MI	
3	CERTIFIED	2019-09-25	2019-10-01	H-1B	SOFTWARE DEVELOPER, APPLICATIONS	15-1132	SOFTWARE DEVELOPERS, APPLICATIONS	Y	SYSTEMS TECHNOLOGY GROUP, INC.	541511.0	TAYLOR	MI	
4	CERTIFIED	2019-09-25	2019-10-01	H-1B	QUALITY ENGINEER LEVEL II	15-1199	COMPUTER OCCUPATIONS, ALL OTHER	Y	E-GIANTS TECHNOLOGIES LLC	541511.0	BLUE ASH	OH	

2.2 Data Processing

Cleaning the data took a big chunk of time—just prepping the LCA Disclosure dataset took 10–12 hours across a few sessions. The data came in all over the place, with mixed wage units (Hour, Week, Month, Year) and inconsistent text formats.

```python
# Check missing values
print(df.isnull().sum())

# If small amount of missing WORKSITE_CITY/STATE, you can drop them
df = df.dropna(subset=['WORKSITE_STATE', 'WORKSITE_CITY'])

# Convert RECEIVED_DATE to datetime if needed
df['RECEIVED_DATE'] = pd.to_datetime(df['RECEIVED_DATE'])

# Extract Year for time-based plots
df['YEAR'] = df['RECEIVED_DATE'].dt.year

df.head()
```

```
CASE_STATUS              0
RECEIVED_DATE            0
DECISION_DATE            0
VISA_CLASS               0
JOB_TITLE                0
SOC_CODE                 0
SOC_TITLE                0
FULL_TIME_POSITION       0
EMPLOYER_NAME            0
NAICS_CODE               0
WORKSITE_CITY            0
WORKSITE_STATE           0
WAGE_RATE_OF_PAY_FROM    0
WAGE_UNIT_OF_PAY         0
PREVAILING_WAGE       1915
ANNUAL_WAGE              0
dtype: int64
```

We started by filtering for rows where VISA_CLASS == "H-1B" and dropped entries missing key fields like WAGE_RATE_OF_PAY, WORKSITE_STATE, or CASE_STATUS. Santhosh and Fahad converted all wages to annual figures—hourly rates were multiplied by 40 × 52, monthly by 12—and removed outliers below $20,000 or above $500,000. Ramcharan cleaned up date fields using Python's datetime library and extracted RECEIVED_YEAR and DECISION_YEAR for trend analysis. Mouryan handled categorical encoding like H-1B dependency and wage tiers, while Kishore mapped NAICS codes to industry labels and expanded state abbreviations for geographic visuals. Once cleaned, the data was saved to files like preprocessed_lca_data.csv and used in PowerBI, Tableau, Python [6] (Plotly/Matplotlib), and Excel. Throughout, we ran checks like summary stats, data types, and null value heatmaps to make sure everything stayed on track.

	CASE_STATUS	RECEIVED_DATE	DECISION_DATE	VISA_CLASS	JOB_TITLE	SOC_CODE	SOC_TITLE	FULL_TIME_POSITION	EMPLOYER_NAME	NAICS_CODE	WORKSITE_CITY	WORKSITE_STATE	WAGE_RATE_OF
0	CERTIFIED	2019-09-25	2019-10-01	H-1B	APPLICATION ENGINEER, OMS [15-1199.02]	15-1199	COMPUTER OCCUPATIONS, ALL OTHER	Y	JO-ANN STORES, INC.	451120.0	HUDSON	OH	
1	CERTIFIED	2019-09-25	2019-10-01	H-1B	BI DEVELOPER II	15-1132	SOFTWARE DEVELOPERS, APPLICATIONS	Y	DENKEN SOLUTIONS INC.	541512.0	BRENTWOOD	TN	
2	CERTIFIED	2019-09-25	2019-10-01	H-1B	QUALITY ENGINEER	17-2141	MECHANICAL ENGINEERS	Y	EPITEC, INC.	541511.0	DEARBORN	MI	
3	CERTIFIED	2019-09-25	2019-10-01	H-1B	SOFTWARE DEVELOPER, APPLICATIONS	15-1132	SOFTWARE DEVELOPERS, APPLICATIONS	Y	SYSTEMS TECHNOLOGY GROUP, INC.	541511.0	TAYLOR	MI	
4	CERTIFIED	2019-09-25	2019-10-01	H-1B	QUALITY ENGINEER LEVEL II	15-1199	COMPUTER OCCUPATIONS, ALL OTHER	Y	E-GIANTS TECHNOLOGIES LLC	541511.0	BLUE ASH	OH	

2.3 Ideas, Sketches, and Prototypes

Our ideation process began with a set of research questions intended to guide the analysis, such as:

- *Which job roles and industries dominate H-1B sponsorship?*
- *Are H-1B wages generally higher than prevailing wages across states and occupations?*
- *Do employers that heavily depend on H-1B workers display different wage behaviour than others?*
- *Which U.S. states serve as hotspots for specific occupations under the H-1B program?*
- *What geographic and temporal patterns can be discerned in H1B wage offers?*
- *Does the involvement of legal representation influence wage levels?*

We began with initial sketches in Tableau and Jupyter Notebooks using sample data subsets [6] starting with basic bar charts and time series plots. Mouryan led the Tableau prototyping, testing chart types and interactivity, while Santhosh refined metric calculations and set up filters for state-wise analysis. Ramcharan and Kishore drafted Sankey diagrams and heatmaps on paper, later building them in Plotly and Seaborn. Fahad enhanced the visual narrative by suggesting KDE plots and violin charts to compare wage distributions across employer types. We went through several rounds of tweaking—looking at each chart to see if it made sense, told a clear story, and was actually useful. Anything that felt repetitive or didn't add much was cut. In the end, we landed on a solid mix of visuals: box plots, bar charts, line graphs, stacked area charts, choropleth maps, KDE plots, Sankey diagrams, treemaps, and heatmaps. Each one helped us answer a different piece of the puzzle.

III. Visualization Methods and Selections

To uncover patterns within the LCA Disclosure dataset spanning FY2020–2024, we implemented a diverse suite of visualization techniques. These methods were selected based on the nature of the data—categorical, temporal, geographic, and numerical—and were designed to maximize interpretability, scalability, and impact. The goal of our visualization strategy was to convert high-dimensional raw data into digestible narratives that reflect H-1B program behaviour over time and across regions, occupations, and employers.

In this project, we employed a diverse range of visualization techniques to extract meaningful insights from the H-1B Labor Condition Application Disclosure dataset [4] and related sources [1], [3], [5]. Each visualization method was selected to suit the specific nature of the data — whether categorical, temporal, geographic, or numerical — and was carefully matched to the research question it addressed. Below, we describe the core visualization types, how they were used, and their key strengths and limitations.

Box Plot

The box plot was one of our primary tools for examining wage distributions across multiple dimensions, including occupation categories, U.S. states, employer types, and legal representation groups. Box plots effectively summarize the spread of a dataset, showing median values, interquartile ranges, and outliers. In our study, they allowed us to highlight wage variability between heavy and light H-1B sponsors, compare wage gaps by occupation, and explore whether legal representation correlated with higher salaries. While box plots offer clear visual summaries, they are limited in that they only display summary statistics and can obscure the full shape or density of distributions, especially if extreme outliers are present.

Pros:
- Clearly shows distribution spread, median, and outliers.
- Great for comparing multiple groups side by side.
- Highlights wage gaps and variability visually.

Cons:
- Cannot show the full shape or density of distributions (only summary stats).
- Sensitive to extreme outliers if not handled carefully.

Line Chart

Line charts were extensively used to capture time-based patterns, such as trends in average wages, the number of applications filed, and approval or denial rates from 2019 to 2024. These charts are highly effective for showing both long-term trends and short-term fluctuations, making them ideal for analysing the impact of policy changes or economic shifts. However, care must be taken when including multiple series in one chart, as too many lines can lead to visual clutter, and line charts inherently assume a continuous time axis, making them unsuitable for categorical comparisons.

Pros:
- Ideal for showing trends, patterns, and fluctuations over time.
- Effective at comparing multiple series (e.g., certified vs. denied cases).
- Highlights both long-term and short-term changes.

Cons:
- Can become cluttered with too many lines.
- Assumes a continuous time axis — not ideal for categorical comparisons.

Bar Chart

For categorical data comparisons, we relied on bar charts to visualize employer rankings, top occupations, and state-level wage differences. Bar charts provide an intuitive way to compare discrete groups, clearly highlighting the largest contributors or top performers. Their flexibility allows both vertical and horizontal layouts, making them adaptable to various data scenarios. However, they can become overwhelming when displaying too many categories and are less suited for showing proportional or temporal relationships.

Pros:
- Simple, intuitive comparisons between discrete categories.
- Easily highlights top performers or largest contributors.
- Flexible for both vertical and horizontal layouts.

Cons:
- Can become overwhelming with too many categories.
- Doesn't convey temporal or proportional relationships well.

Stacked Area Chart

To explore both time trends and category breakdowns simultaneously, we used stacked area charts. These visualizations allowed us to examine how the absolute and relative shares of certified, denied, and withdrawn applications changed over time. Stacked area charts are visually appealing and compact, making it easy to see how category proportions evolve. Still, they can be harder to interpret precisely for individual categories and may mislead if the stack order obscures smaller segments.

Pros:
- Combines time trends with category breakdowns in one view.
- Highlights how the proportions of categories shift over time.
- Visually appealing and compact.

Cons:
- Harder to interpret precise values for individual categories.
- Can be misleading if the stack order hides smaller segments.

Violin Plot

When we needed to analyse the full shape of wage distributions, we turned to violin plots. These combine summary statistics with kernel density estimation, providing a detailed picture of both central tendency and distribution shape. In our case, violin plots helped us compare wage patterns between heavy and light H-1B employers, revealing not only medians and interquartile ranges but also subtle differences in wage spread and concentration. Although powerful, violin plots can be less familiar to general audiences and require careful interpretation.

Pros:
- Shows both summary stats (median, quartiles) and density shape.
- Highlights subtle differences in wage spread and concentration.

- Useful for comparing the complexity of distributions between groups.

Cons:
- Less familiar to general audiences compared to box plots.
- Requires careful interpretation to avoid misreading the density.

Choropleth Maps

For mapping geographic patterns, choropleth maps proved invaluable. We used them to visualize the global origins of H-1B applicants as well as U.S. state-level wage levels and application volumes. Choropleth maps effectively highlight regional intensity and make geographic clusters immediately visible. However, they are sensitive to scale and color choices and may mask small underlying counts, especially in sparsely populated regions.

Pros:
- Powerful for showing spatial distributions and regional intensity.
- Easy to spot geographic clusters and hotspots.
- Supports both domestic and international mapping.

Cons:
- Sensitive to geographic scale and shading choices.
- Can mask underlying small counts in sparsely populated regions.

Heatmap

To track application intensity over time across large categorical matrices, we employed heatmaps. These visualizations provided compact, color-coded matrices showing trends among top employers and states, allowing us to detect peaks, drops, and long-term patterns at a glance. While heatmaps are excellent for summarizing intensity, they do not convey exact values or trends and can be highly dependent on thoughtful colour scale design.

Pros:
- Compact and effective for large categorical matrices.
- Easily shows trends, peaks, and low points at a glance.
- Highlights patterns across both dimensions (e.g., employer × year).

Cons:
- Limited to showing intensity — no exact values or trends.
- Color scale choices can affect interpretability.

Sankey Diagram

We used Sankey diagrams to represent the flow of H-1B workers from specific occupations to destination U.S. states. These diagrams captured proportional flows in a multivariate setting, visually revealing which occupations are most concentrated in which regions. Sankey diagrams are particularly engaging and powerful for illustrating connections and bottlenecks, though they require careful labelling to avoid clutter and can become visually complex with many nodes.

Pros:

- Captures multivariate flow relationships clearly.
- Visually engaging for showing proportional connections.
- Highlights bottlenecks or dominant pathways.

Cons:

- Requires careful labelling to avoid clutter.
- Can become visually complex with too many nodes or links.

Tree Map

For showing hierarchical, part-to-whole relationships, we applied tree maps to summarize industry-level H-1B sponsorship. Each rectangle represented an industry, sized by its application volume, providing a compact, intuitive visual of which sectors rely most heavily on international talent. Tree maps are highly space-efficient and excellent for relative comparisons, but they can be less precise when reading exact values, especially when color encoding is layered on top.

Pros:

- Space-efficient for showing part-to-whole relationships.
- Visually intuitive for comparing category sizes.
- Handles many categories compactly.

Cons:

- Less precise for reading exact values.
- Color-coding can overwhelm if too many dimensions are added.

KDE Plot

We also incorporated kernel density estimation (KDE) plots to compare smoothed wage distributions between heavy and light H-1B users. KDE plots allowed us to visualize the continuous shape of wage distributions, highlighting differences in spread, peaks, and density between groups. While KDE plots provide a refined look at distributional shapes, they are sensitive to bandwidth choices and may be less interpretable to audiences unfamiliar with density estimation methods.

Pros:

- Smooths distributions for better shape comparison.
- Reveals peaks, spread, and density differences.
- Useful for comparing continuous variable distributions.

Cons:

- Sensitive to bandwidth selection (can over- or under-smooth).
- Less interpretable to audiences unfamiliar with density estimation.

Histogram

Finally, we used histograms to examine wage distributions for applications with versus without legal representation. Histograms provided a straightforward frequency-based comparison, showing where data clustered across wage bins. While histograms are intuitive and easy to interpret, they can be highly sensitive to bin width selection and generally provide a rougher, less smooth view compared to KDE plots.

Pros:

- Straightforward visual for showing frequency across bins.
- Easy to compare group-wise distributions.
- Highlights data concentration and gaps.

Cons:

- Choice of bin width can greatly affect appearance.
- Less smooth and detailed than KDE plots.

By combining these visualization techniques, we were able to build a rich, multi-dimensional picture of the H-1B program, uncovering key wage patterns, employer strategies, geographic flows, and policy-relevant trends that would have been difficult to detect using numerical tables alone.

Note: All the visualizations and heir screenshots will be in the results and interpretation section.

IV. Results and Interpretations

1. Which job roles and industries dominate H-1B sponsorship?

1.1. Top 10 H-1B Employers (Certified Petitions)

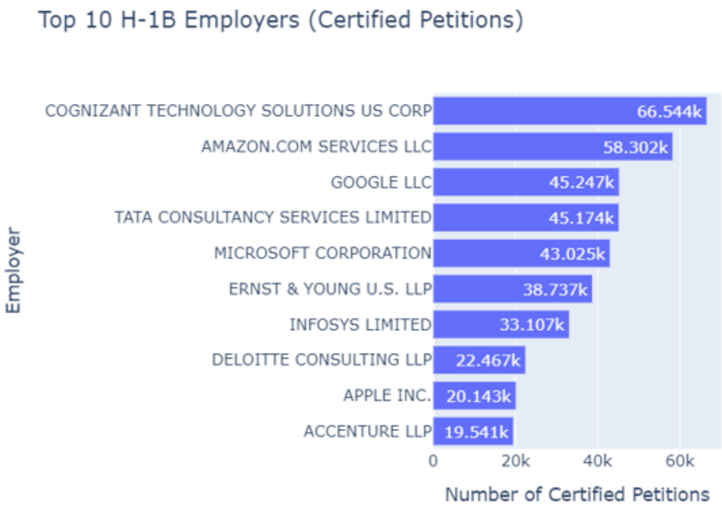


Fig. 1.1

To see which companies lead the pack in H-1B sponsorships, a horizontal bar chart does the job best. Why? Because many of these employers have long names, and a vertical chart would just look cramped and messy. A pie chart? Even worse—hard to read and not great for comparisons.

This chart highlights the top 10 companies based on the number of certified H-1B petitions. Unsurprisingly, it's mostly tech and consulting giants like Cognizant, Amazon, Google, and Tata Consultancy Services. Together, they make up a big chunk of all approved applications. That tells us one thing loud and clear: a small group of major players is doing most of the hiring when it comes to skilled foreign workers. That concentration is important. It shows just how much the H-1B program depends on the tech industry—and a handful of employers, really. If immigration policy shifts or restrictions hit, it's these big firms that feel it first, and that could ripple through the whole program. The chart itself is clean and easy to read. Names are fully visible, companies are ranked in order, and each bar is labeled with the number of petitions. It tells the story without overwhelming you.

1.2. Top 10 Job Titles in H-1B Applications

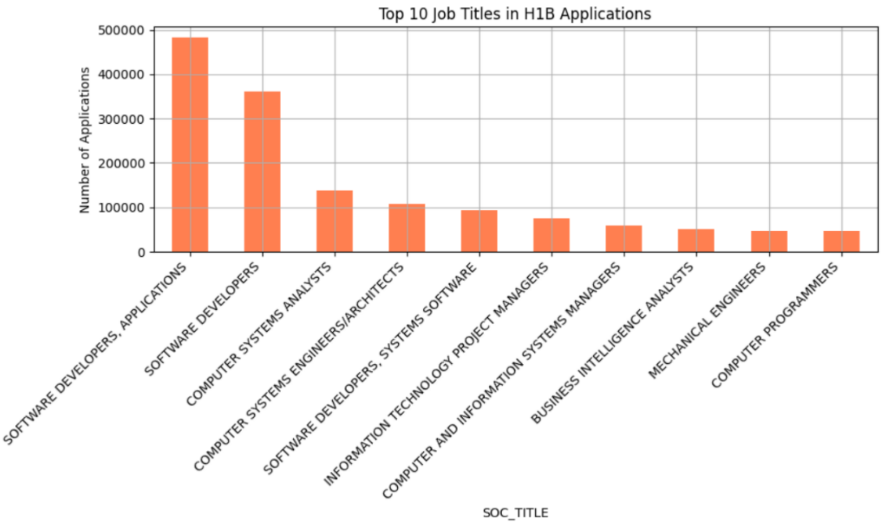


Fig. 1.2

This second chart uses a vertical bar graph to break down the most common job titles submitted in H-1B visa applications. It's a clean, simple way to show the data—especially since the job titles are short enough to line up neatly. Right away, you can see it's mostly tech jobs dominating the list. Roles like "Software Developers, Applications," "Software Developers," and "Computer Systems Analysts" make up a huge share, which isn't surprising given the tech industry's outsized role in the H-1B process. There are a few non-tech positions, but they're clearly the exception. The design choices—angled labels, bold colors, consistent layout—help keep it easy to read without feeling cluttered. Together, it backs up what the employer chart already suggested: the H-1B program leans heavily on a small number of tech-driven job types.

Results

The H-1B visa program is largely driven by the tech industry. Most certified applications come from a small group of major tech companies like Cognizant, Amazon, and Google. At the same time, the top job titles are almost all software or IT-related roles. These two visuals together show that H-1B demand is heavily concentrated in tech—both in terms of who's hiring and what roles they're hiring for.

2. Are H-1B wages generally higher than prevailing wages across states and occupations?

2.1. H-1B Offered vs. Prevailing Wages by Occupation (Bar Chart)

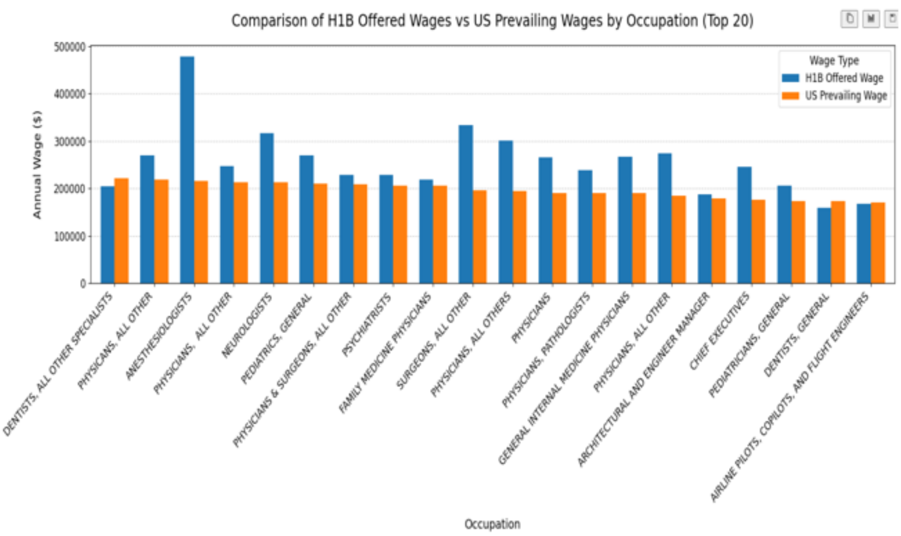


Fig. 2.1

This grouped bar chart offers a clear comparison between the wages offered to H-1B applicants and the typical prevailing wages for the same jobs across the U.S. With color-coded bars for each wage type, it's easy to see the differences side by side. The layout works especially well for this kind of data, and the tilted labels help keep longer job titles readable without crowding the chart. What really jumps out is that, for many high-demand roles—like Anaesthesiologists and Surgeons—H-1B workers are actually being paid more than the national average. That challenges the common assumption that these visas are used to cut labour costs. Instead, it suggests that employers are often offering competitive or even premium salaries, possibly to meet regulatory requirements or to attract highly specialized talent they can't find locally.

2.2. Percentage Wage Difference by State (Box Plot)

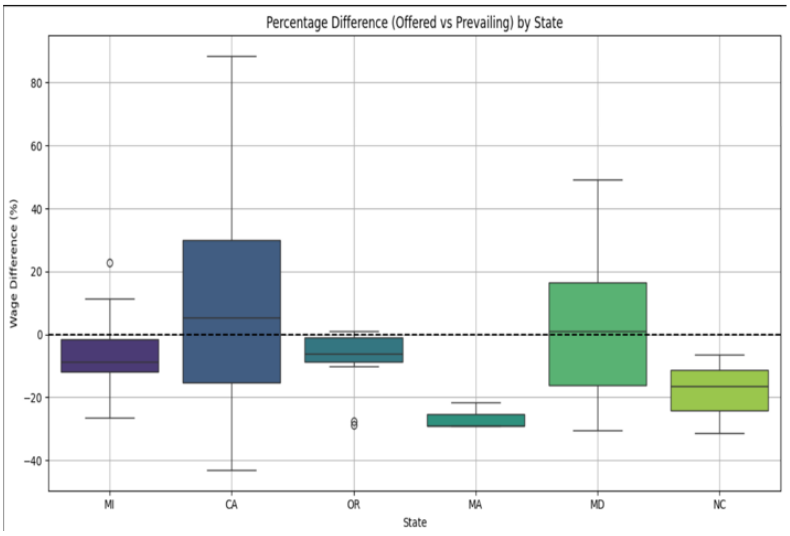


Fig. 2.2

This box plot shows how wage differences between offered and prevailing wages vary across states. By using percentage differences, it normalizes the data and lets us compare states on equal footing. The plot captures not just average differences but also the range and outliers, giving a fuller picture of wage dynamics by region. States like California and Maryland tend to offer wages above prevailing levels, while places like Michigan and North Carolina fall below. This visualization helps reveal that wage practices aren't uniform across the country. Instead, there are regional differences that could be influenced by cost of living, employer behavior, or local labor markets.

2.3. Salary Bucket Comparison (Bar Chart)

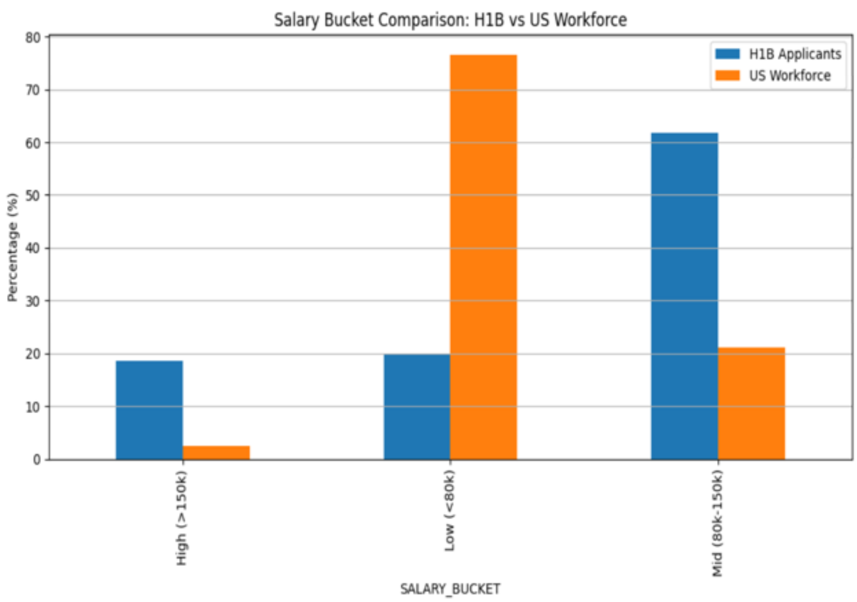


Fig. 2.3

This vertical bar chart breaks salaries into three simple categories—low, mid, and high—and compares the percentage of H-1B applicants and U.S. workers in each group. It's a clean, no-frills visual that makes the differences easy to spot. What stands out is that most H-1B applicants are landing in the mid-to-high salary range, meaning jobs paying over $80,000. On the other hand, a majority of U.S. workers are still concentrated in the lower bracket. That contrast suggests H-1B positions tend to be higher-paying and likely tied to roles that demand more specialized skills or advanced qualifications than the average job on the market.

2.4. Salary Distribution – H-1B vs. US Workforce (Density Plot)

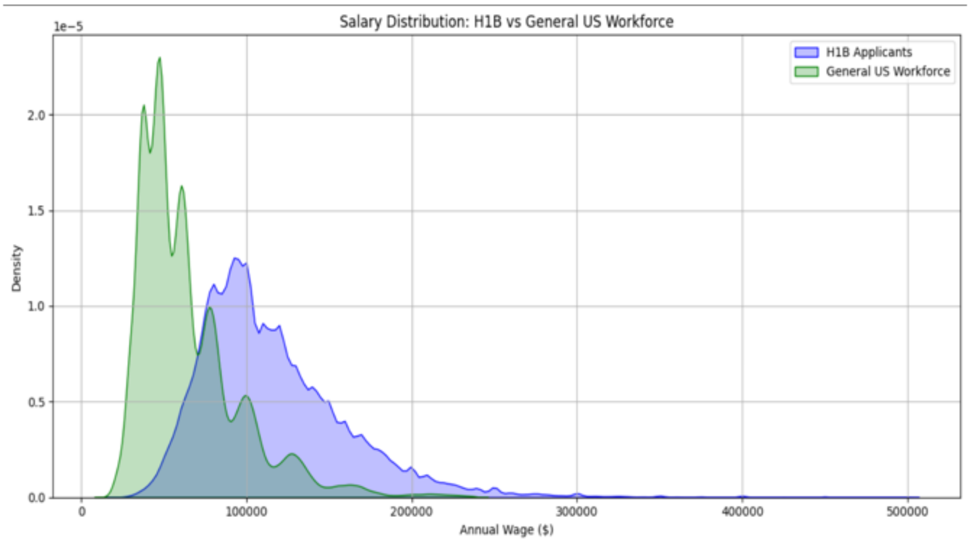


Fig. 2.4

This chart, which uses a KDE (density) format, gives a clearer picture of how salaries are spread out for H-1B applicants compared to the general U.S. workforce. Unlike a histogram, it smooths out the data, making trends easier to spot. What stands out is that H-1B wages tend to be clustered at higher levels, with a noticeable extension into very high-income territory. This seems to back up the idea that H-1B jobs are often in more specialized, better-paying roles—not the low-wage positions people sometimes assume

2.5. Top 20 Occupations by Average H-1B Wage (Bar Chart)

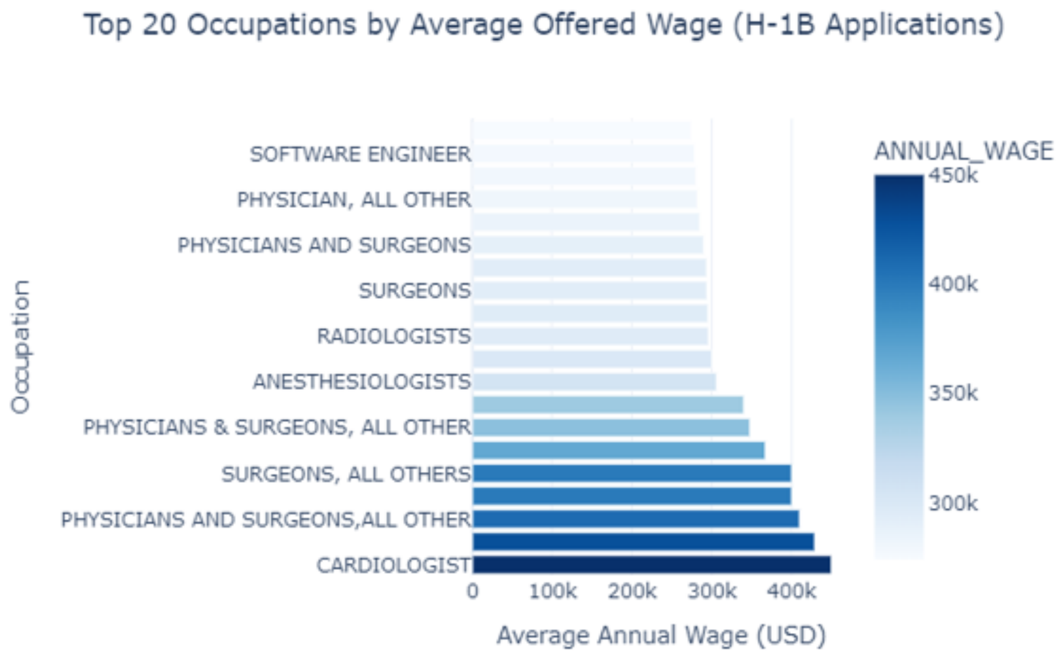


Fig. 2.5

This horizontal bar chart ranks the top 20 occupations based on their average offered wage for H-1B applicants. The use of horizontal bars and a color gradient helps visually emphasize how much higher some roles pay than others. Occupations in the medical field like Cardiologists and Surgeons clearly top the list, with average salaries exceeding \$400,000. The chart clearly shows that H-1B visas are often used to fill **critical, high-value jobs** in sectors that demand specialized expertise. It challenges the misconception that the program is dominated by low-wage labour.

Results

Looking at all the charts together, it's pretty clear that H-1B workers are usually paid more than the average U.S. worker both across different jobs and in various states. Roles in fields like healthcare and tech stand out with especially high wages. The differences from state to state show that location matters too some places pay well above average, while others are closer to the baseline. And when you look at the wage brackets, most H-1B jobs fall into mid or high salary ranges. All in all, the data paints a picture of H-1B jobs being high-skill and well-paid, not low-cost replacements.

3. Do employers that heavily depend on H-1B workers display different wage behaviour than others?

3.1. Wage Comparison – Heavy vs. Light H-1B Employers (Box Plot)

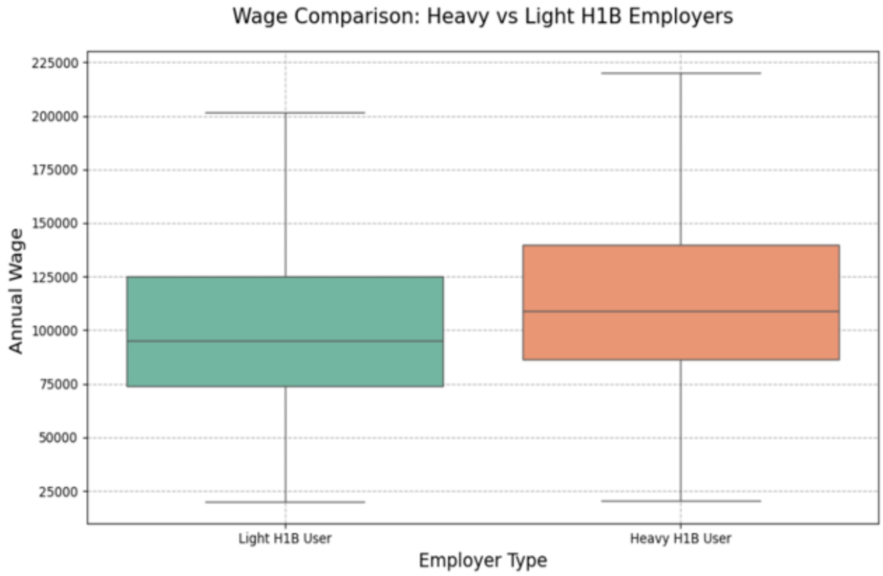


Fig. 3.1

To find out if employers that depend more on H-1B workers behave differently when it comes to wages, three charts were examined. The first one—a box plot—compares salaries between companies that hire a lot of H-1B workers and those that don't. It turns out the heavy users tend to pay slightly higher median salaries, while lighter users have a wider salary range. Some pay very well, and others don't, suggesting their pay practices are less consistent.

3.2. KDE Plot – Wage Distribution by Employer Type

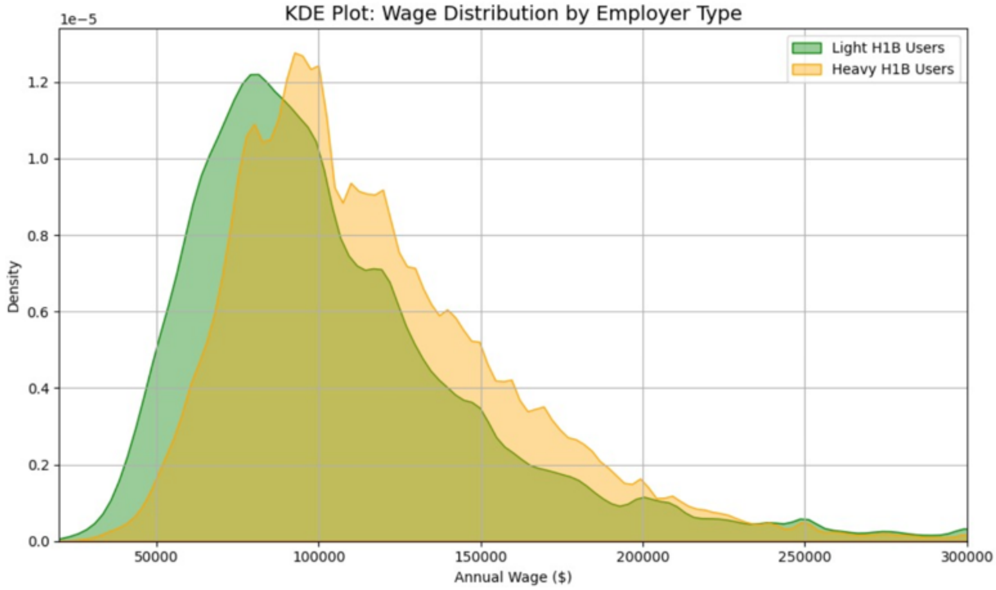


Fig. 3.2

A density plot gives a better look at how wages are spread out. The data shows that companies using more H-1B workers not only offer higher average pay but also have more employees earning at the top end. Lighter users, on the other hand, mostly stick to the mid-range.

3.3. Violin Plot – Annual Wages by Employer Type

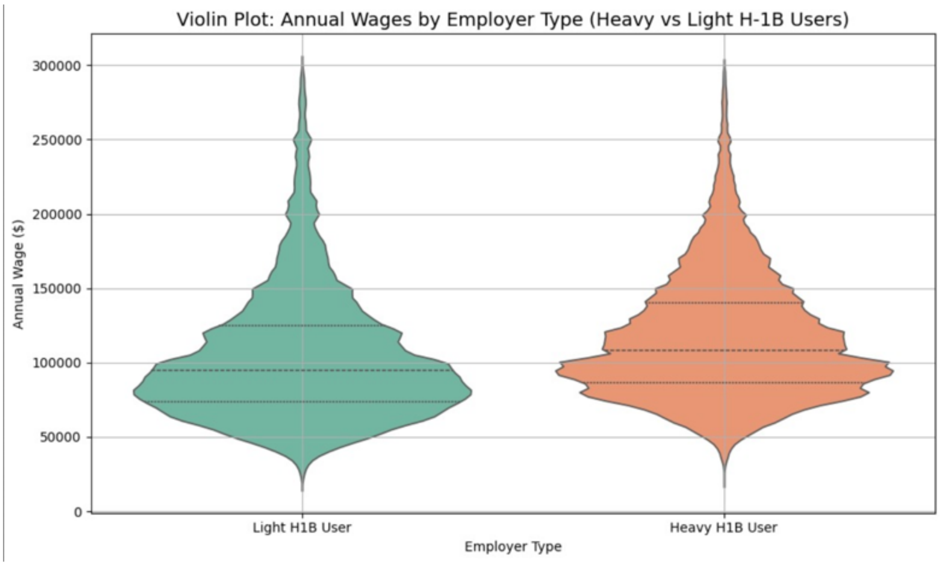


Fig. 3.3

The above chart, a violin plot, pulls everything together by showing both the range and concentration of salaries. Companies that rely heavily on H-1B talent have a noticeable bulge in the upper wage range, while lighter users are more concentrated in the lower half. Overall, these visuals clearly point to one thing: companies that hire more H-1B workers tend to offer steadier and better wages—possibly because they follow more formal pay structures or need to attract highly skilled, specialized professionals.

Results

These three charts together tell an interesting story: companies that hire a lot of H-1B workers tend to offer steadier, higher wages than those that don't. The box and violin plots show that heavy users stick to a tighter salary range, but with higher median pay. Then the KDE plots take it further, showing that some of those wages stretch deep into the high-income bracket. On the flip side, companies that only occasionally use H-1B visas have a wider spread, including more lower-paying roles. That kind of variation suggests that frequent users might have more standardized or cautious pay practices—maybe because they're under the microscope more often or have internal systems to stay on the safe side of the rules.

4. Which U.S. states serve as hotspots for specific occupations under the H-1B program?

4.1. Flow of H-1B Workers – Occupation to State (Sankey Diagram)

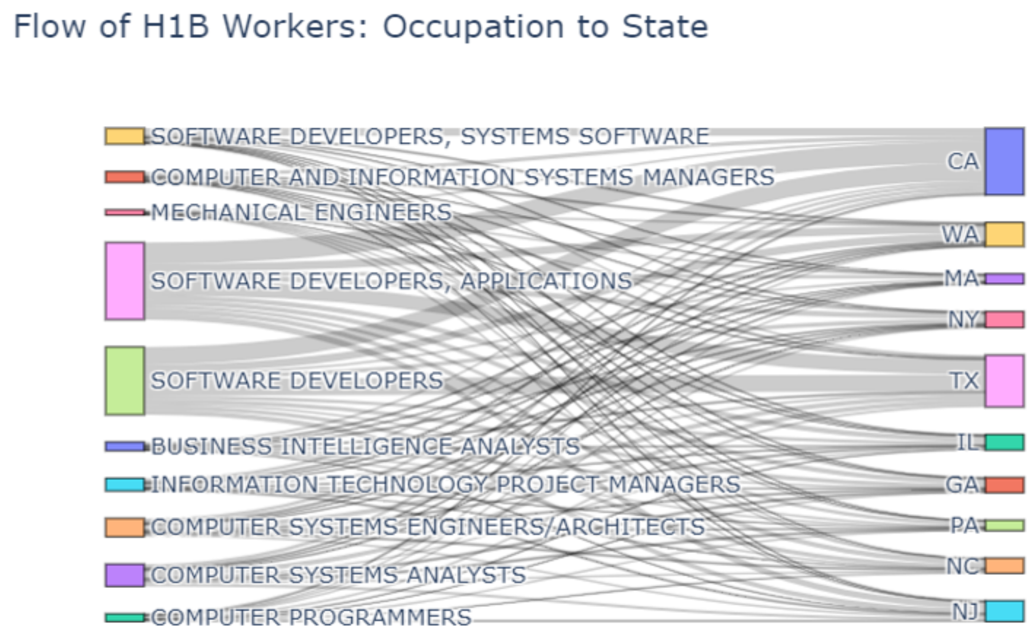


Fig. 4.1

If you're trying to figure out which states attract different kinds of H-1B job roles, a few visuals help paint the picture. One of them is a Sankey diagram that shows how specific occupations are spread across the country. For instance, software developers and systems analysts tend to end up in states like California, Texas, New York, and Washington.

Meanwhile, roles like mechanical engineers or business analysts appear more often in places such as Illinois, Pennsylvania, and Georgia. This gives a good sense that tech-heavy states are still the top choice for most H-1B workers.

4.2. Choropleth Map – H-1B Applications Distribution Across States

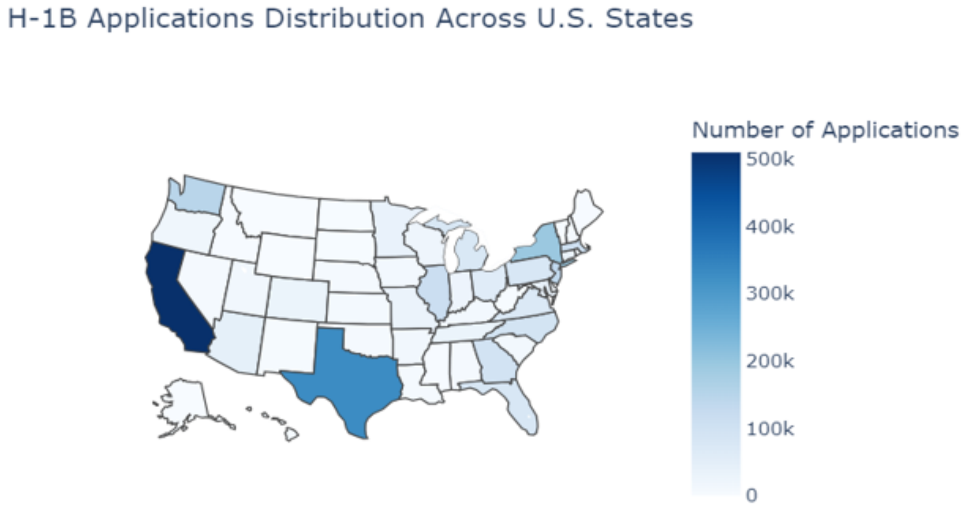


Fig. 4.2

Another helpful visual is a map of the U.S. coloured by application counts. It quickly shows that **California, Texas, and New York** get the biggest share, with a decent number also going to **Washington, New Jersey, and Illinois**. This lines up well with what we see in the Sankey chart.

4.3. Top 15 States by H-1B Applications (Bar Chart)

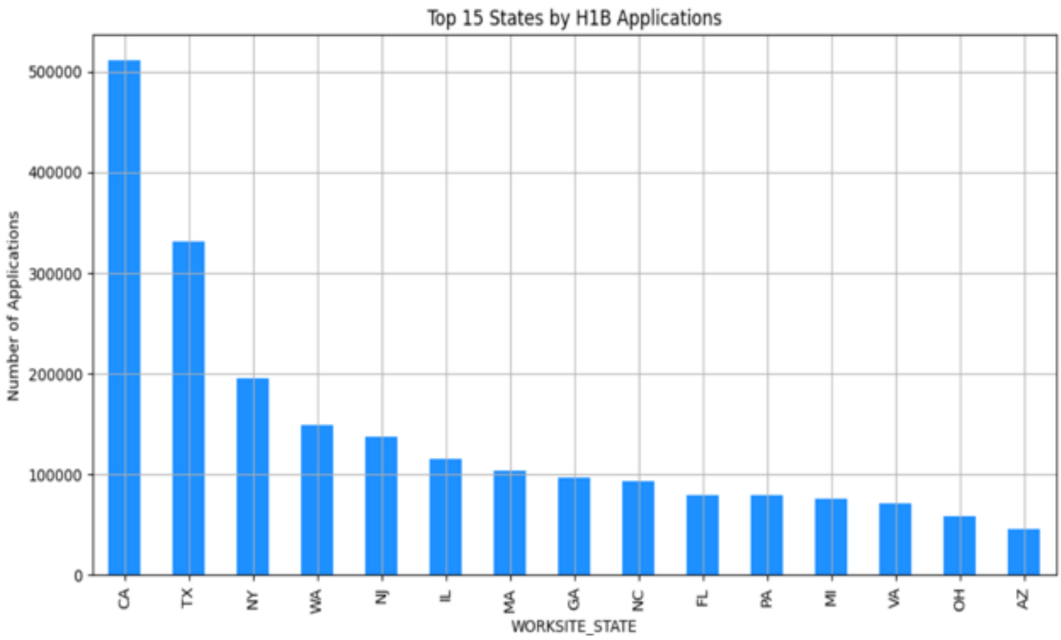


Fig. 4.3

This chart is a simple bar graph listing the top 15 states based on how many applications they've received. **California clearly leads the pack**, followed by Texas and New York. So when you put all of this together, it's pretty obvious that while H-1B workers are spread out nationwide, most of them end up in just a few key states especially for tech jobs.

Results

When you look at all three visuals together, it's pretty obvious that H-1B workers are mostly heading to California, Texas, and New York. These states aren't just popular—they dominate the numbers, especially for tech jobs like software developers and systems analysts. The Sankey diagram makes it easy to see how specific roles are flowing straight into those states. Then, the choropleth map and bar chart double down on that point, showing those three states consistently at the top in terms of applications. What's cool about this combo is that it doesn't just show where people are going—it connects the dots between the jobs and the destinations. So you're not just seeing movement, you're seeing *why* that movement is happening.

5. What geographic and temporal patterns can be discerned in H1B wage offers?

5.1. Average H-1B Annual Wage Over Years (Line Chart)

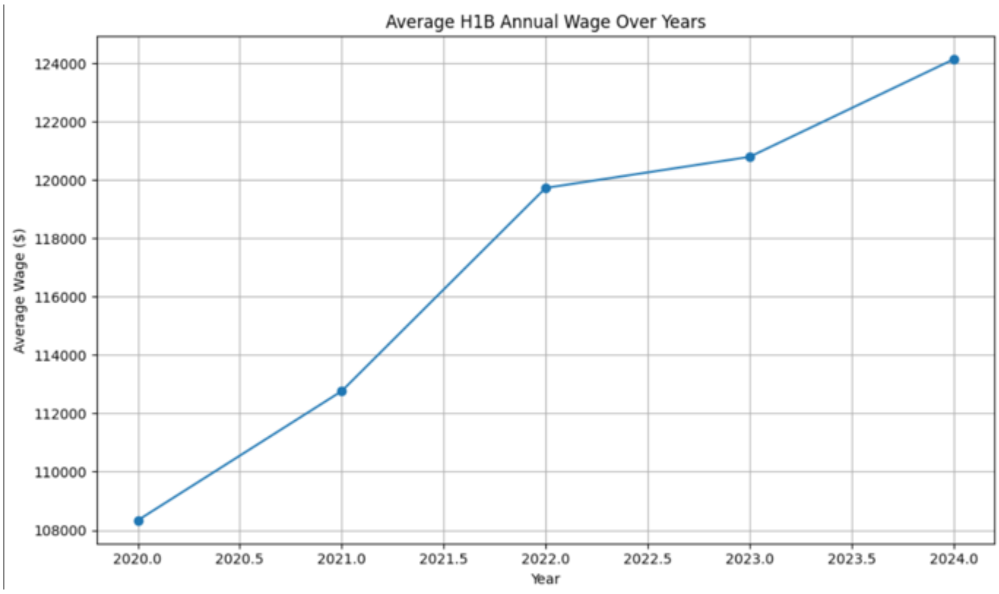


Fig. 5.1

If you want to get a feel for how H-1B wages have changed over time, the line chart from 2020 to 2024 really tells the story. Unlike bar charts that break things up year by year, the line gives you a smooth view of how salaries have climbed over the years. And they've climbed quite a bit—starting at about $108,000 in 2020 and pushing past $124,000 by

2024. That’s a solid jump. It could be a mix of things: higher demand for skilled workers, companies competing harder for talent, or just inflation catching up. The chart itself is clean and easy to follow, with clear labels and a nice year-to-year flow. Adding a few markers or notes at key points could make it even easier to read, but even as-is, it gets the point across. H-1B salaries are going up, and that trend doesn’t look like it’s slowing down.

5.2. Yearly Salary Comparison by Job Title (Line Chart)

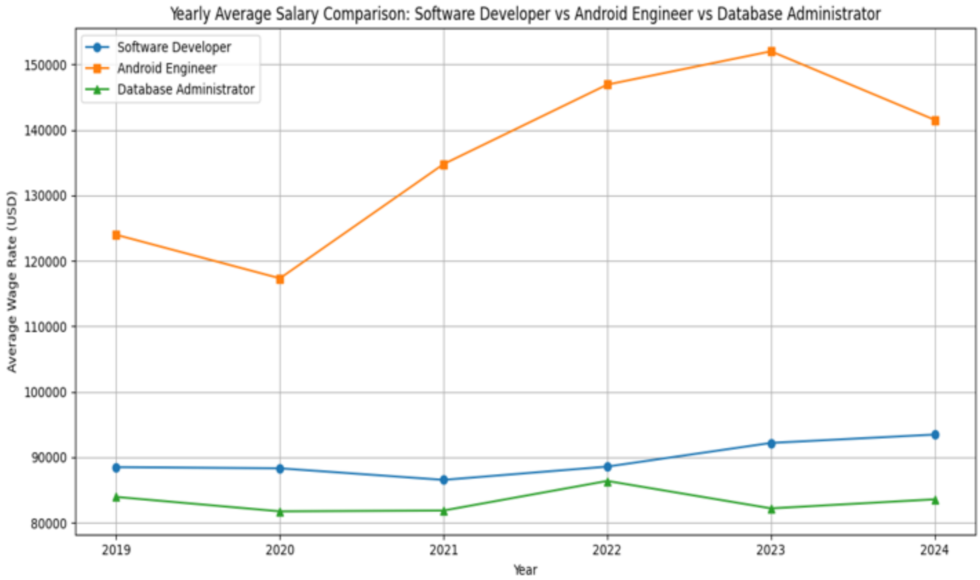


Fig. 5.2

The graph shows how pay has shifted over time for three roles—Software Developers, Android Engineers, and Database Admins. A line chart just makes more sense here. If it were bars, it’d probably get messy with all those years and job types crammed in. You can see pretty quickly where the lines start to split. What jumps out is how much more Android Engineers are making, especially around 2023 when their pay shoots past $150K. The other two jobs don’t move as much—they stay lower and more stable. So yeah, the chart isn’t just showing pay over time, it’s also showing how much your role matters. That big jump for Android work probably means mobile skills are getting more valuable. The chart’s easy to read too—clear colors, nothing too fancy. A few notes on the highs and lows could help, but even without them, it tells the story.

5.3. Annual Wage Distribution (Histogram with KDE)

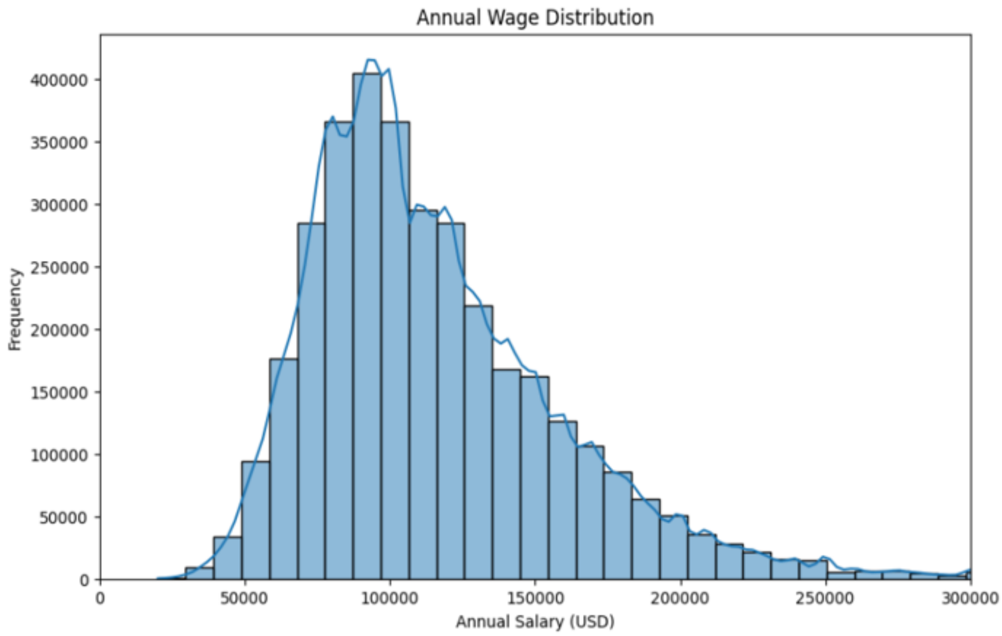


Fig. 5.3

To obtain a more intuitive understanding of what H-1B workers are really being paid, this chart combines a histogram with a KDE (Kernel Density Estimate) curve—and to great effect. Instead of showing bare raw counts or a plain average, the combination gives a more rounded sense of how the wages are spread out. The KDE smooths that out so that you can observe the shape of the distribution without the distraction of the noise. Most of the salaries bunch between \$80K and \$130K, but there is a noticeable extension past \$200K, showing that while many jobs are in the mid-to-high range, a smaller percentage of workers are getting some very high-earning jobs. That right skew shows there's considerable variation even at the high side. The plot itself is clean—good bin size, readable axes, smooth curve. It might be helpful to have a marker for median or mode, but even without them, this visualization does more than simply say "H-1B workers are well-paid." It shows what those salaries look like in practice along the range.

Results

Taken together, these visuals tell a pretty clear story: wages for H-1B workers have been climbing steadily over the past few years, especially in roles like Android Engineering, where the demand seems to be really taking off. While these charts don't show location directly, earlier choropleth maps already made it obvious that the highest-paying jobs are clustered in states like California, Texas, and New York big tech hubs. All in all, the H-1B program still leans heavily toward high-skill, high-pay roles, and the wage trends make it clear that specialization is becoming more valuable than ever.

6. Does the involvement of legal representation influence wage levels?

6.1. Violin Plot – Annual Wages With vs. Without Legal Representation

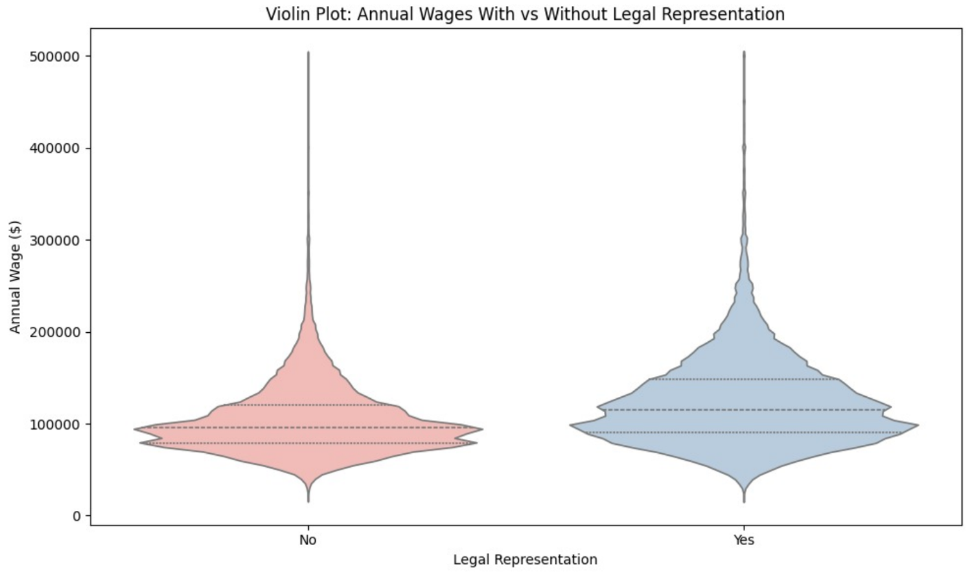


Fig. 6.1

This violin chart does a great job showing how wages vary between people who had legal help and those who didn't. It's not just about the average—it gives you a fuller picture of how salaries are spread out and where most people fall within each group. What really jumps out is that folks who had legal representation tend to earn more, not just in terms of median pay, but also when you look at the range. There's a wider stretch of higher salaries on that side, meaning they're more likely to land those top-paying roles. That means having legal help isn't just a nice-to-have—it actually seems to pay off. You can see a real difference in how much people end up earning. The chart's also pretty easy to follow, with clear colors and lines that make the pay ranges obvious. Sure, it would be nice to have some numbers on there for context, but even without them, the takeaway is clear.

6.2. Mean vs. Median Wage by Legal Representation (Grouped Bar Chart)

Fig. 6.2

This grouped bar chart breaks down the average and median wages for people who had legal support versus those who didn't. It's straightforward and easy to read—no extra clutter or confusion. What stands out right away is that both the average and median wages are higher when someone has a lawyer or legal rep involved. The difference isn't small either—we're talking about a $15,000 to $20,000 gap, which is a pretty big deal. The chart is well put together, using two different colors to separate the mean and median, so you can follow the data without any trouble. It might be even more useful if the actual wage numbers were added on top of the bars, but even without that, the message is clear: having legal support seems to make a noticeable difference.

Results

Both of these visuals tell the same story having legal help seems to make a real difference when it comes to wages. Now, we can't say it's the only reason someone might get paid more, but the numbers point in a pretty clear direction. People who had legal support often end up with better offers, maybe because their applications are stronger or their reps know how to navigate the system and push for more. These charts do a great job side by side, giving a more complete look at the wage gap and offering some insight into why it might be there.

Discussion and Conclusion

This project offers a comprehensive, visually-driven exploration of the H-1B visa landscape, highlighting wage dynamics, employer behaviors, legal influences, and regional clustering through a broad set of visual techniques. By translating complex datasets into intuitive narratives, the visualizations bridge the gap between raw data and actionable policy insights.

Insights and Effectiveness

1. Wage Disparities and Employer Practices

Box plots, violin plots, and KDEs revealed that employers heavily reliant on H-1B workers tend to offer higher and more consistent wages. In contrast, occasional sponsors showed greater variance, with more applications falling into lower wage brackets.

Legal representation emerged as a significant factor, with applicants represented by legal counsel receiving noticeably higher average and median salaries, underscoring the systemic advantage conferred by legal expertise.

2. Job Roles and Industry Dominance

Bar charts indicated that a handful of large tech companies dominate the H-1B petition space. Roles such as software developers and systems analysts repeatedly surfaced as top job titles.

These findings reinforce the narrative that the H-1B program is concentrated within a narrow band of high-skilled tech occupations, making it particularly sensitive to policy changes targeting the tech sector.

3. Temporal and Geographic Trends

Line charts and histograms showed a steady upward trend in average H-1B wages from 2020 to 2024, reflecting both inflation and increased demand for specialized talent.

Choropleth maps and Sankey diagrams captured spatial flows of H-1B workers, identifying California, Texas, and New York as dominant hubs—especially for tech-related occupations.

4. Wage vs. Prevailing Wage Analysis

Grouped bar charts and box plots comparing offered vs. prevailing wages highlighted systemic wage gaps, particularly in high-demand occupations and states with large H-1B inflows.

These differences challenge assumptions that H-1B labour is primarily used for cost savings, revealing instead a trend toward premium compensation in many sectors.

Limitations and Caveats
Demographic Incompleteness: Key applicant details like education, age, or experience were missing or inconsistent, limiting subgroup-level insights.

Static Visualizations: While static plots captured trends clearly, lack of interactivity reduced the ability for user-driven exploration—an area where tools like Tableau or Power BI could offer richer engagement.

Employer Behaviour Context: Some findings, like those on legal representation or wage variance, could benefit from qualitative context such as firm policies or industry-specific practices.

Granularity of Geography: State-level spatial resolution obscures local patterns. City or zip-code level mapping could reveal finer employment concentrations.

Summary

This project presents a deep, multi-dimensional analysis of the H-1B visa system by leveraging recent labour condition application data from FY2020 to FY2024, supported by contextual datasets on consultant placement and employer history. Through a refined visualization pipeline featuring violin plots, KDEs, choropleth maps, and Sankey diagrams the report uncovers critical insights into wage disparities, geographic clustering, employer dependency, and the effects of legal involvement on visa outcomes.

Innovative comparisons between offered and prevailing wages, as well as between heavy and light H-1B users, reveal systemic compensation gaps and behavioural trends. Each visualization is designed not only for clarity but for impact translating dense datasets into actionable insights.

By exposing wage gaps, regional skews, and compliance signals, this project supports smarter immigration policymaking and fairer labour practices. It stands as a visual and analytical framework for understanding the complexities of skilled immigration in today's U.S. job market.

V. Future Work and Recommendations

While this project successfully uncovers wage disparities, employer trends, and geographic skews in the H-1B program, several opportunities exist for deeper analysis in future work.

First, incorporating demographic dimensions such as age, education level, or years of experience could enrich understanding of wage structures and approval outcomes. Although such variables were limited in the current dataset, sourcing them from additional FOIA datasets [5], [8] or LinkedIn-derived profiles could enable more granular subgroup analysis [6].

Second, integrating **interactive dashboards** (e.g., Tableau Public or Power BI) [6] would allow users to explore trends by filtering employer type, location, job title, or decision status. This could make the project more usable by immigration analysts, journalists, or policymakers who require scenario-specific insights.

Third, clustering techniques and **predictive modelling** could be introduced to flag potentially non-compliant employers [6], [7] employers based on historical wage gaps and approval anomalies. Building a machine learning model that predicts approval likelihood based on wage, location, and legal representation would add significant applied value.

Finally, expanding the scope to include **non-H-1B visa types** such as L1 or OPT and comparing their approval trends and compensation patterns could shed light on broader labour migration strategies and how they interact with the H-1B program.

In summary, this project provides a strong foundation for visual policy analysis but can be expanded into a more dynamic, predictive, and inclusive immigration insight platform through additional data dimensions and tooling.

VI. References

1. Mahipal Kumar Singh. (2023). Consultant Job Placement Data 2023. Kaggle. Available at: https://www.kaggle.com/datasets/mahipalkumarsingh/consultant-job-placement-data-2023

2. KonradB. (2024). H1B Visa Statistics. Kaggle. Available at: https://www.kaggle.com/datasets/konradb/h1b-visa-statistics

3. Lokesh Jain. (2023). USCIS H1B Data Cleaned 2009–2023. Kaggle. Available at: https://www.kaggle.com/datasets/lokesh97jain/uscis-h1b-data-cleaned-2009-2023

4. U.S. Department of Labor. (2024). H-1B Labor Condition Application Disclosure Data (FY2020–2024). Available at: https://www.dol.gov/agencies/eta/foreign-labor/performance

5. U.S. Citizenship and Immigration Services. (2009–2023). H-1B Employer Data Hub. Available at: https://www.uscis.gov/tools/reports-and-studies/h-1b-employer-data-hub

6. Migration Policy Institute. (2024). Immigration Data Hub – Labor and Workforce. Available at: https://www.migrationpolicy.org

7. Pew Research Center. (2024). Key Findings About U.S. Immigrants. Available at: https://www.pewresearch.org

8. U.S. Department of Homeland Security. (2024). Yearbook of Immigration Statistics. Available at: https://ohss.dhs.gov/topics/immigration/yearbook