Parallel Judgments: A Visual Exploration of Similar Case Outcomes

Sampad Banik

u1467294

u1467294@utah.edu

K M Arefeen Sultan

u1419693

u1419693@utah.edu

Project Repository:

https://github.com/dataviscourse2024/group-project-parallel-judgments-vis.git

1. Overview and Motivation

The legal system is built on precedents, where past cases influence future decisions. However, understanding how similar cases are decided across different jurisdictions or time periods can be complex. This project aims to provide an interactive visualization tool that enables legal scholars, students, and researchers to compare and contrast the outcomes of similar court judgments. My interest in legal informatics and data visualization has driven me to explore how visual tools can make legal research more accessible and insightful. This project will also help bridge the gap between law and technology, providing a modern solution for comparing case law effectively.

2. Related Work

Several related works provide a foundation for visualizing legal case outcomes and using data visualization in the legal domain. Projects like the Caselaw Access Project (CAP) [1] and CourtListener [2] provide access to legal data and have been used in studies to visualize judicial decisions, helping researchers analyze case outcomes, precedents, and jurisdictional variations. NLP in legal studies plays a significant role, with tools like LexNLP applied to extract key legal principles and attributes from legal texts, similar to our project's data processing tasks [3]. The use of interactive data visualizations in law, such as in the RuleMap project and platforms like Docket Alarm, highlights how visual tools can simplify the understanding of complex legal information by representing legal rules and case trends using flowcharts, bar charts, and other visual methods [4]. In precedent-based decision analysis, works like The Precedent Explorer demonstrate how interactive visualizations can highlight relationships between cases and the precedents they rely on, aligning closely with our goal of comparing similar judgments across jurisdictions. Other platforms like Jureeka!, Justia, and FindLaw emphasize the importance of interactive filtering and case comparison for legal professionals[5]. Additionally, studies on legal network analysis and legal citation networks use visualization techniques to map how cases are connected through citations, illustrating the hierarchical structure of legal precedents, much like the visual comparisons. These projects collectively demonstrate the value of visualizing legal data and the potential impact of our tools in making legal research more accessible and insightful.

3. Questions

This part is from the proposal.

- **Primary Goal:** To develop an interactive visualization tool that compares the outcomes of similar court judgments.
- **Key Questions to Answer:**
    - How do legal outcomes for similar cases vary across different courts or time periods?
    - What are the key factors (jurisdiction, legal principles, or precedent cases) that contribute to these variations?
- **Benefits:**
    - Provide insights into patterns and disparities in legal decisions.
    - Help legal professionals and researchers easily compare similar cases.
    - Facilitate learning for law students through an interactive and engaging tool.

4. Data

Legal case data are sourced from publicly available two databases:

CourtListener and Caselaw Access Project

The data are collected via APIs where possible or manually downloaded from the websites. We have chosen 12 cases which are downloaded in multiple formats like: JSON, CSV etc. All the JSON files are kept in our project GitHub repository.

- **Data Preprocessing Tasks:**
    - The first step is to clean the raw legal text by removing unnecessary components such as footnotes, citations, and legal references. These elements, while important for the completeness of a legal document, often clutter the text and detract from the core content that needs to be analyzed. Additionally, different legal jurisdictions may present documents in varying formats, so standardizing the format across cases helps create a uniform dataset. This cleaning ensures that the remaining text is concise and focused on the essential legal information.Extract key case metadata such as the outcome, cited precedents, and legal principles.
    - Once the text is cleaned, the next step is to extract key case metadata, which is critical for comparison and visualization. Important metadata

includes the outcome of the case, such as whether it was dismissed or upheld, as well as cited precedents and legal principles that were key to the decision. Cited precedents help understand how prior cases influenced the current decision, and legal principles provide insight into the reasoning behind the judgment. Extracting these pieces of metadata allows the creation of structured datasets that can be filtered and compared across multiple dimensions, such as legal outcomes or the influence of particular precedents.

- The third step involves the application of Natural Language Processing (NLP) to further refine the analysis. NLP techniques such as tokenization, Named Entity Recognition (NER), and text similarity algorithms allow for the automatic extraction and comparison of key legal terms, principles, and cases. By using NLP, the tool can break down the legal text into manageable components and identify significant patterns, such as frequently cited precedents or recurring legal principles. NLP can also facilitate text similarity comparisons between cases, helping to cluster and analyze similar judgments. In some cases, sentiment analysis may be applied to assess the tone of the judgment, providing further insights into the nature of the case outcomes.

5. Exploratory Data Analysis

There are many sections in the courtlistener website like Case Law, RECAP archive, Oral Arguments, etc. We chose Case Law section and worked with the given APIs.

Case law data management is structured around four key objects: Courts, Dockets, Clusters, and Opinions, each playing a role in organizing legal information.

Court objects store details about courts, such as their name and location, while Docket objects include metadata like case initiation or termination dates and docket numbers, linking the case to a specific court. Cluster objects group related opinions, such as dissents or concurrences, when a panel makes multiple decisions.

Finally, Opinion objects hold the full text of a court's decision along with relevant metadata about the authoring panel member. This structure ensures that dockets are organized within courts, clusters contain multiple opinions, and all metadata is stored efficiently at the lowest level (within opinions), preventing redundancy in the system.

The insights from case law data management significantly informed the design of the visualization tool by emphasizing the need for structured, hierarchical data

representation. Understanding that Courts, Dockets, Clusters, and Opinions are interconnected helped shape how case data is displayed and compared in the tool. The hierarchical relationship between these objects—where dockets are filed within courts and contain clusters of opinions—influenced the design to ensure that users can seamlessly navigate from broader categories (e.g., courts or jurisdictions) to more specific details (e.g., individual case opinions). This hierarchical approach also allowed for clear filtering and comparison of cases across different courts, time periods, and legal principles. Additionally, the emphasis on organizing metadata at the opinion level guided the inclusion of interactive features, where users can access detailed case metadata by exploring opinions directly, ensuring that the design provides both high-level overviews and in-depth insights.

6. Design Evolution

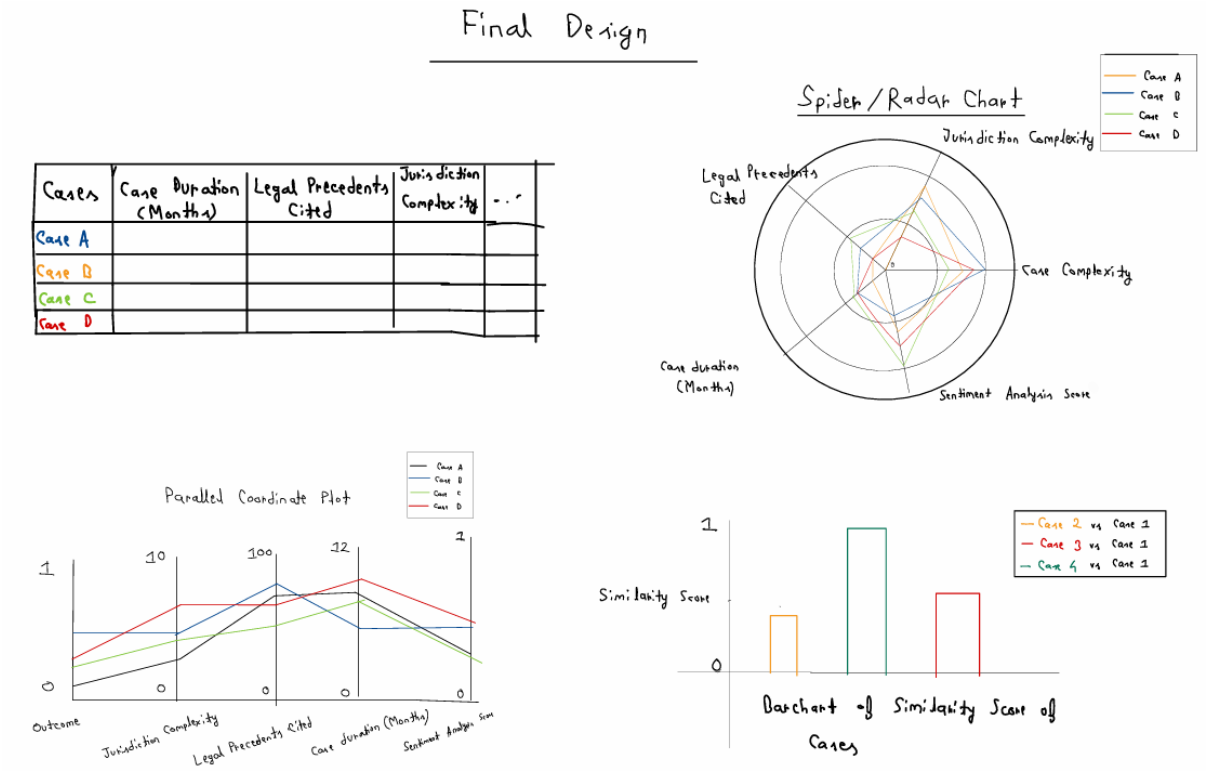
At first, we designed three Prototype Designs along key dimensions such as jurisdiction, time period, Citations, and outcomes. Sketch 1 was focused on a simple, high-level analysis. It uses a table to summarize case details like duration, legal precedents, and jurisdiction complexity. A pie chart shows a 50/50 split between favourable and unfavorable outcomes, while a flowchart illustrates the potential legal processes such as dismissal, appeal, and settlement. A bar chart compares similarity scores between cases, highlighting how closely they align based on selected criteria.

Sketch 2 was added more depth with multiple visualizations. A table provides case attributes, while a bar chart breaks down the stages of case progression, such as dismissal or appeal. A heatmap highlights the variation in case complexity, legal precedents, and duration across cases, making it easier to identify patterns. The bar chart also shows similarity scores, comparing Case 2, Case 3, and Case 4 with Case 1, providing insights into case differences.

Sketch 3 takes a more advanced, multi-dimensional approach. It includes a radar chart comparing cases A-D across five variables, such as legal precedents, complexity, and sentiment analysis. A timeline shows case milestones over time, offering a chronological view of case progression. As with the other sketches, a bar chart displays similarity scores between cases.

Final Design: The final design offers a comprehensive analysis of case attributes and similarities using multiple visualizations. The table provides key details such as case duration, legal precedents cited, and jurisdiction complexity for Cases A-D. A radar chart compares the cases across dimensions like complexity, duration, and sentiment

analysis. A parallel coordinates plot illustrates the variations in case outcomes, complexity, and other attributes across the cases. Lastly, a bar chart visualizes the similarity scores between Case 1 and the other cases (2, 3, and 4), highlighting their relative differences.



7. Implementation

(a) Data Overview:

- **Table Visualization:**

The **case comparison table** displayed in the visualization provides a structured, user-friendly way to analyze parallel judgments across multiple court cases. This table is a critical component of the overall implementation, enabling users to access detailed metadata for each case in a clear and organized format.

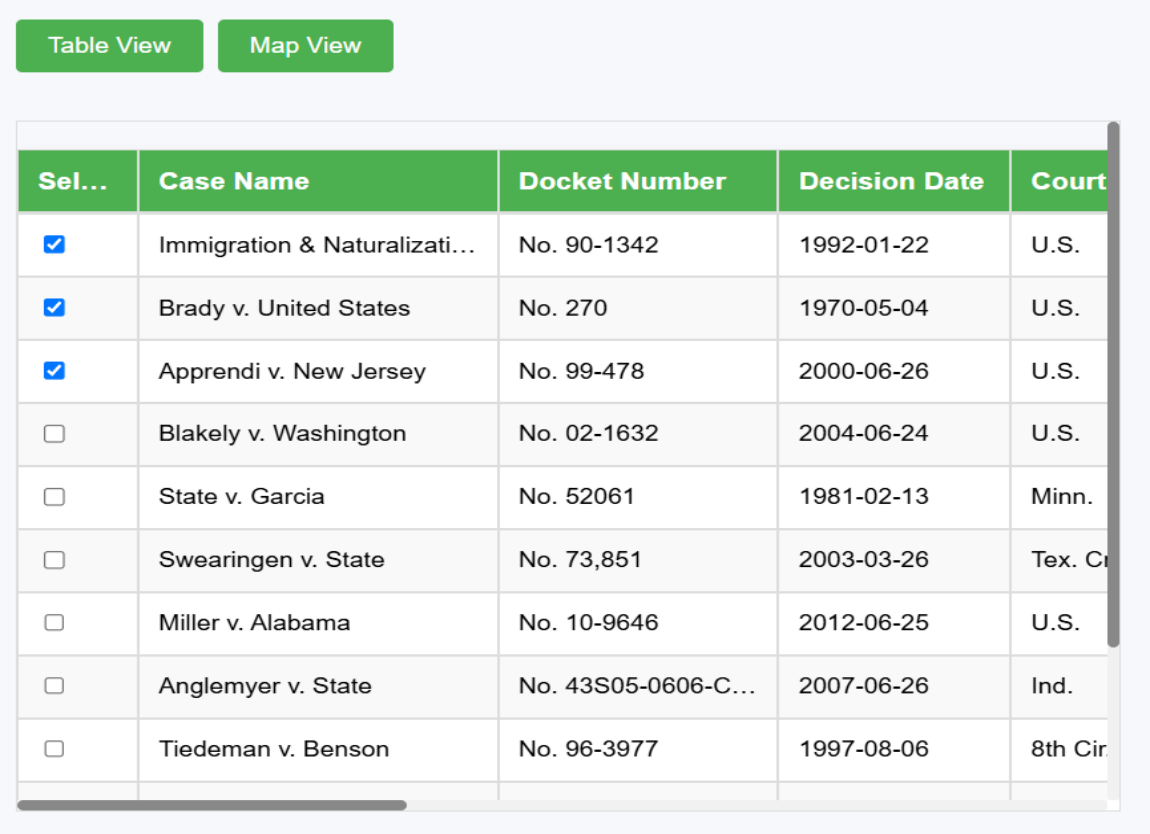
The table includes columns for the case name, decision date, jurisdiction, court, citations, and cases cited. Each of these columns offers crucial information that helps in comparing similar court decisions. The case name column allows users to quickly identify the cases being compared, while the decision date and court

columns provide context about the timing and level of the court making the judgment.

Each case name and jurisdiction is hyperlinked, allowing users to click through for additional information. These links can take users to full case details or to broader information about the court system in that jurisdiction. The citations column shows how the case was cited in future decisions, and the cited cases column reflects how past cases influenced the decision. This allows for an easy exploration of the precedents and subsequent impacts of each judgment.

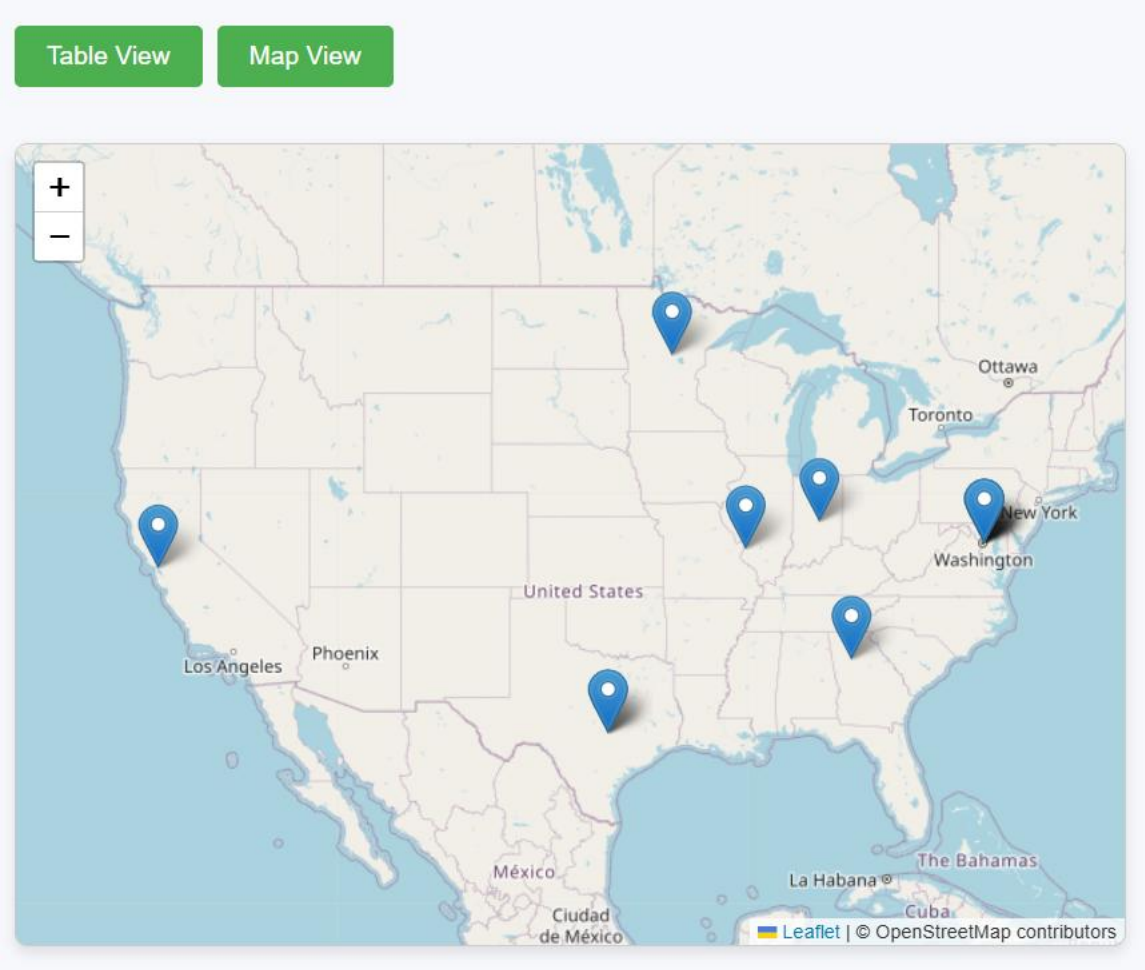
The table design itself is clean and minimalist, ensuring the focus remains on the key data. The columns are aligned for readability, and case data is presented in a way that allows for quick visual comparisons across cases. Additionally, expandable sections provide an efficient way to show extended citation details without overwhelming the viewer with text.

This table serves as a core feature of the interactive visualization tool, enabling users to engage with complex legal data in an intuitive and informative manner. By structuring and presenting the data in this way, users can easily compare key case attributes and understand the relationships between legal decisions across different jurisdictions and time periods.



- **Map Visualization:**

The **Map Visualization** provides an interactive geographic representation of court locations associated with the cases in your dataset. Each blue marker on the map corresponds to a court, dynamically placed based on its latitude and longitude, and clicking a marker reveals details like the court name, case name, and decision date. The map is interactive, allowing users to zoom, pan, and explore regional distributions of court cases, making it easier to identify patterns across jurisdictions. It dynamically adjusts to fit all markers within view, ensuring visibility of all locations. This visualization is particularly useful for analyzing trends, comparing federal and state-level court activity, and understanding the geographic spread of legal decisions, offering an accessible and insightful layer to your dataset.



(b) Radar Chart

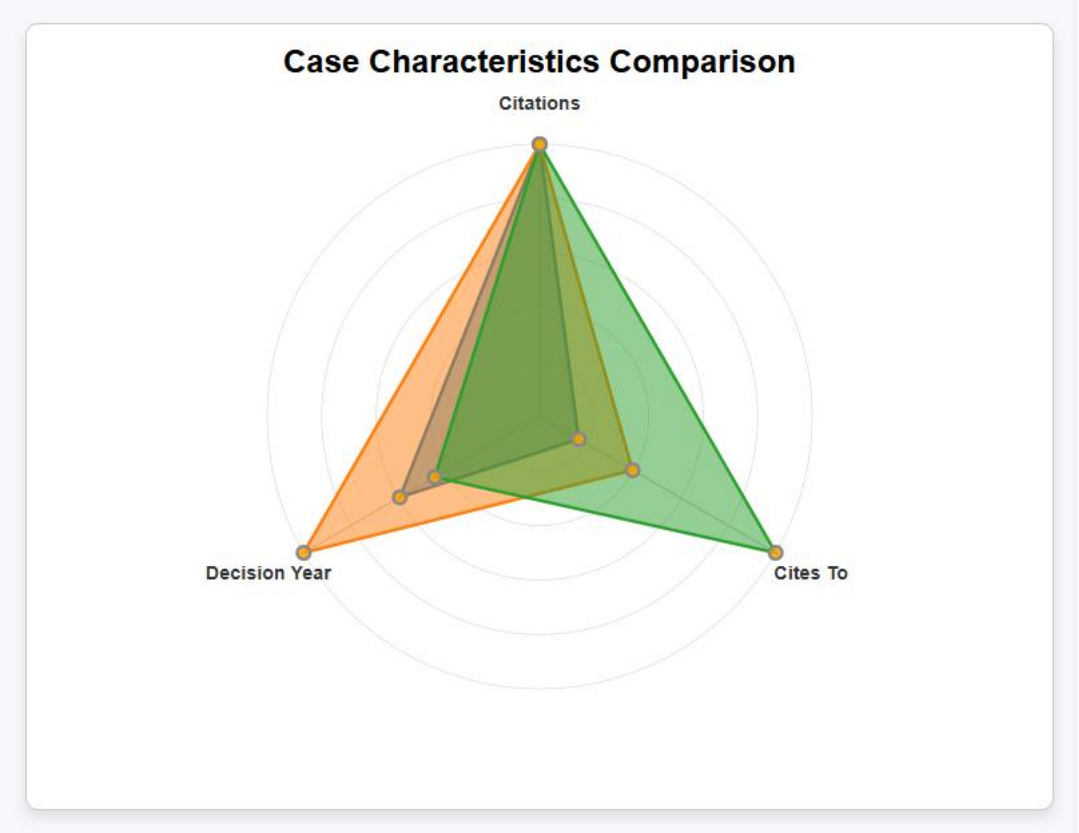
The radar chart shown in the visualization is a powerful tool for comparing multiple attributes of legal cases simultaneously across several dimensions. In this chart, each axis represents a specific variable, such as citations, decision year, and cites to, and the data for each case is plotted along these axes to create a polygonal shape. This visualization helps users easily compare cases across multiple factors in a single view.

The radar chart includes three key axes: citations, decision year, and cites to. Each axis extends from the center of the chart outward, with values plotted along these axes for each case. For instance, cases with a high number of citations will have points plotted further out along the citations axis. This allows users to see how each case performs across the different variables.

Each legal case is represented by a polygon, formed by connecting the points plotted along each axis. The shape and size of the polygon vary depending on the values of the variables. For example, a case with high values for all variables will create a large, well-rounded polygon, while a case with lower or more varied values will form a smaller, more irregular shape. This visual representation allows for quick comparison across cases, as users can instantly see how cases differ based on the shapes and sizes of the polygons.

Radar charts are particularly effective for comparing multiple cases at once. In this visualization, multiple cases are shown with different colors, enabling users to compare how cases align or diverge across the three dimensions. For example, a case with high citations but fewer cited references ("cites to") will have a distinctly different polygon shape compared to a case with a more balanced distribution of values across all three dimensions. This visual allows users to spot outliers or cases that are particularly strong or weak in certain areas.

In the context of legal case comparisons, the radar chart helps users visually assess the influence of different variables on a case. For example, it can highlight cases that have been widely cited or cases that cite a significant number of precedents. By visualizing these relationships, users can gain insights into how different factors—such as the year of the decision or the number of citations—impact the case's relevance or influence.



(c) Parallel Coordinates Plot

The parallel coordinates plot shown in the visualization provides a unique way to compare multiple dimensions of legal case data across different attributes, such as opinion length, cite ratio, and jurisdiction level. This plot is essential for identifying patterns and relationships between cases based on these variables, allowing users to assess how different factors correlate.

Each axis in the plot represents a specific attribute. In this case, the plot includes axes for opinion length, cite ratio, and jurisdiction level. These variables are scaled between 0 and 1, allowing the plot to compare data on a normalized scale. By placing these axes in parallel, users can trace how each case's attributes align across the different dimensions.

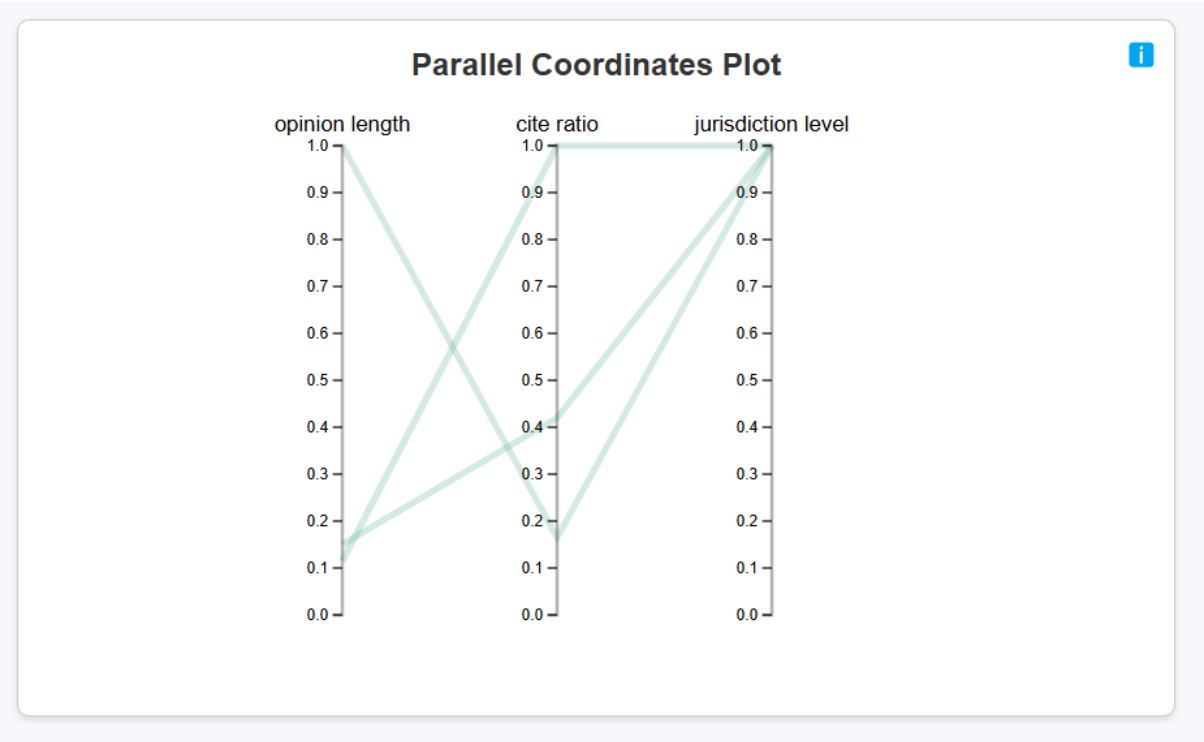
Each line in the plot represents a single legal case, and the line intersects each axis at a point corresponding to that case's value for the respective attribute. For instance, if a case has a high opinion length but a low cite ratio, the line will start near the top of the opinion length axis and dip towards the bottom of the cite ratio axis. This visual structure makes it easy to see how cases vary across the selected attributes.

The parallel coordinates plot is particularly useful for recognizing patterns. For example, users can quickly identify clusters of cases that exhibit similar trends across the three variables. A series of parallel or closely aligned lines across the axes indicates

similar behaviour among cases, whereas intersecting or diverging lines highlight outliers or differing trends.

In the context of legal case comparisons, this plot enables users to assess how factors like the length of a court opinion, its citation ratio, and the level of the jurisdiction relate to one another. It can help reveal whether, for example, cases from higher jurisdiction levels tend to have longer opinions or higher citation ratios, providing valuable insights into the nature of judicial decisions across different courts.

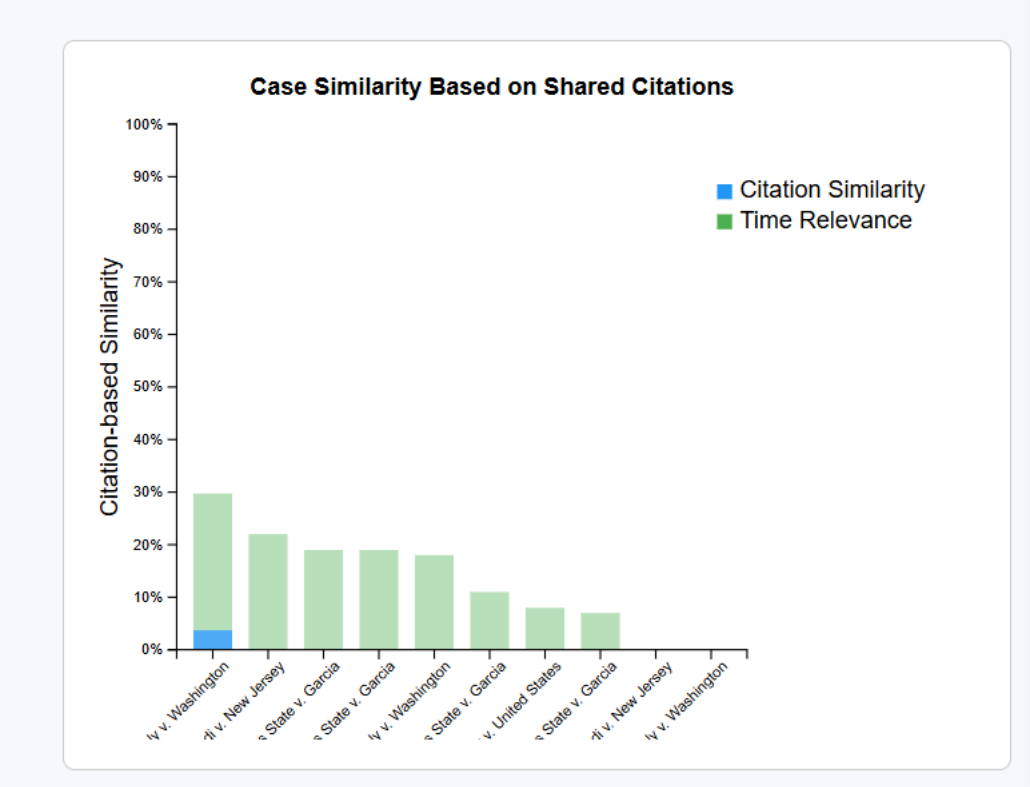
The parallel coordinates plot adds a powerful multidimensional analysis tool to the visualization, complementing other visual elements by enabling in-depth comparisons of multiple case attributes simultaneously. It allows users to explore complex relationships and trends within legal data that would be difficult to discern from a simple table or list of statistics.



Overall, the table, radar chart, and parallel coordinates plot provide complementary views of the dataset. The table offers precise case-by-case details, while the radar chart and parallel coordinates plot enable high-level comparison across multiple dimensions. These visualizations offer a range of perspectives, making it easier to analyze and understand legal cases based on textual complexity, influence, and jurisdiction. Expanding these visualizations with interactivity—such as sorting, filtering, and zooming—could further enhance data exploration capabilities."

(d) Bar Chart Plot

The “Case Similarity Based on Shared Citations bar chart” visually compares the similarity between selected legal cases by combining two factors: Citation Similarity and Time Relevance. Each bar represents a pair of cases, with the total height reflecting their overall similarity score. The blue segment denotes the Citation Similarity, which measures how many references or citations the two cases share, while the green segment shows Time Relevance, indicating how temporally close the cases are, with greater weight given to cases decided in closer years. For instance, in the chart, the first case pair ("v. New Jersey") has a higher similarity score primarily due to time relevance, as shown by the larger green bar. This visualization helps analyze relationships between cases, highlighting the significance of shared legal references and the recency of decisions in determining their similarity.



8. Evaluation

Using the visualizations, we gained significant insights into the legal case data, particularly in terms of how different cases compare across attributes like citation count, decision dates, and jurisdiction levels. The parallel coordinates plot allowed me to see patterns and relationships between variables like opinion length and citation ratio, helping to identify trends among cases. The radar chart provided a clear visual

representation of multidimensional data, making it easy to spot outliers or cases that stood out due to high or low values across different variables, such as citation influence and case references. These visualizations helped to address key questions about how similar legal cases are across different factors and which attributes contribute the most to their similarity.

The visualizations effectively answered questions regarding the variation in legal outcomes across cases and jurisdictions. For instance, by comparing multiple cases using the radar chart, we could quickly assess how the number of citations or the jurisdiction level influenced case outcomes. The case comparison table also provided a detailed breakdown of metadata, allowing for deeper analysis of how precedents and cited cases factored into the decision-making process. Each visualization worked together to provide a comprehensive understanding of the dataset, making the complex relationships between legal cases more apparent.

While the visualizations worked well overall, there are several ways they could be improved. One potential enhancement would be to introduce more interactivity, allowing users to filter cases based on specific criteria such as court type, jurisdiction, or case outcome. Additionally, improving the clarity of the parallel coordinates plot by reducing overlap or adding more color distinction between cases could help make the data easier to interpret. Another possible improvement would be to add more dynamic features, like hover tooltips or clickable elements within the radar chart, providing users with more detailed case information when exploring the plots. These changes would make the visualizations more engaging and provide users with more flexibility in analyzing the data.

9. References

[1] **CourtListener**. Free Law Project. (n.d.). *CourtListener: A Comprehensive Search Tool for U.S. Court Opinions*. Retrieved from https://www.courtlistener.com

[2] **Caselaw Access Project**. Harvard Law School. (n.d.). *Caselaw Access Project: Making All U.S. Court Decisions Freely Available*. Retrieved from https://case.law

[3] **Jackson, P., & Cunningham, H.** (2000). *Natural Language Processing for Legal Texts*. In *Handbook of Natural Language Processing*. CRC Press

[4] **Ashley, K. D.** (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge University Press.

[5] **Jensen, D., & Allen, C.** (2006). *Visualization of Legal Information: A Case Study Approach*. Law Library Journal, 98(1), 1-16.