**Analyzing the Impact of Demographic Factors on the U.S. Presidential Election Outcomes: A Case Study of North Carolina**

Xiaoyi Kuang, Yijia Sun, Peng Wang

Duke Kunshan University

Prof. Luyao Zhang

December 11, 2024



Fig. 1: Research Proposal Logical Flow (https://whimsical.com/impact-of-demographic-factors-on-elections-updated-JGC768A7tPjL993UDeu84y)

# 

# **Background and Motivation**

As the 2024 United States presidential election concluded, Donald Trump won the presidency with an advantage over his base voters and a remarkable expansion to those who were traditionally supportive of the Democratic Party (Sanders 2024). This subtle change intrigues our group, for it highlights dynamic voting behaviors and implies the significance of demographic factors in shaping election outcomes. That is to say, factors such as age, gender, race, education level, income, etc. interact comprehensively to affect individual political preferences. Understanding how each of these demographic features plays a role in voter decisions enables us to uncover trends or patterns, explore socioeconomic drivers behind these changes, make political strategies, and even predict future election outcomes.

Therefore, our project aims to analyze the impact of various demographic factors on the U.S. presidential election outcomes, with a specific focus on North Carolina, a swing state. By narrowing our scope to North Carolina, we ensure a manageable workload and gain more precise insights compared to studying nationwide data. Exploring the relationship between demographic factors and voting outcomes in North Carolina informs discussions about social, cultural, and economic factors that promote political change, sheds light on the prediction of future election outcomes, and illuminates the potential strategies behind political campaigns for this state. It also contributes to the field of political science and enhances the overall understanding of broader societal transformations.

In addition to its importance and relevance to social impact, our project also has value in the field of data visualization. By transforming data into scientific, intuitive visualizations, we aim to identify trends and patterns that are otherwise hard to perceive by looking at the raw data. These visualizations not only provide a more persuasive interpretation of our findings but also offer insight into how effectively designed visualization can fill the gap between abstract data and practical understanding.

Our group draws inspiration for the research from real-life visualizations. Fig. 1 shows the changing trends in presidential candidate preference by race or ethnicity among American people from 1980 to 2016, as introduced in a research paper authored by Alec Tyson and Shiva Maniam (Tyson and Maniam 2016).

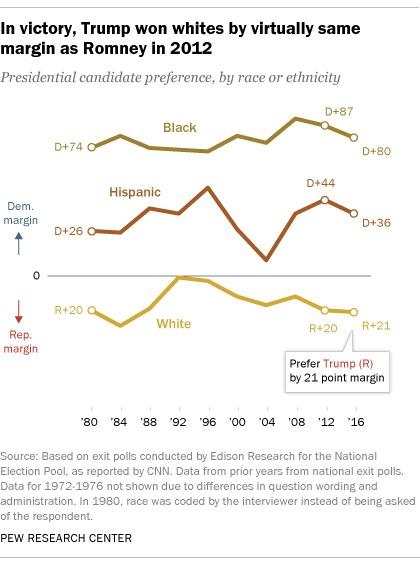


Fig. 2: Presidential Candidate Preference by Race or Ethnicity from 1980 to 2016 (from [“Behind Trump’s Victory”](https://www.pewresearch.org/short-reads/2016/11/09/behind-trumps-victory-divisions-by-race-gender-education/))

This visualization reminds us of the critical role of race or ethnicity as a prominent demographic characteristic influencing election results. From Fig. 1, we can see that most Black and Hispanic people voted for the Democratic Party while White people largely supported the Republicans. Yet, these preferences have fluctuated over time, reflecting the dynamic nature of voting behavior.

This real-life visualization connects to our research goals by encouraging us to question whether the trend remains identical in the 2024 election, particularly in North Carolina. Further, it also emphasizes the importance of demographic analysis, inspiring us to explore what influence other demographic factors, such as age, gender, and education level, have on the election. By incorporating the methodology used in this visualization, we could fulfill our research goal of gaining insight into the relationships between demographic features and political outcomes.

# **Research Question**

How do demographic factors, including age, gender, race, education, income, and occupation, impact the U.S. election outcomes in North Carolina?

# **Application Scenarios**

**1. Tailored Political Campaign Strategies**

Political campaign teams can use our findings to design more targeted strategies based on demographic insights. For example, understanding that certain age groups, races, or education levels are more likely to support Donald Trump enables his campaigns to refine messaging, advertising, and outreach methods more catering to those people to win over key segments of the population.

**2. Promoting Policy and Governance**

Policymakers can use our findings to understand how demographic factors influence political priorities and policy preferences. For instance, shifts in voting patterns among minority groups or younger voters may highlight issues that resonate with these demographics, such as education reform, healthcare access, or economic mobility. In that case, policymakers can make relatively promoted policies to diminish the negative effects of socioeconomic issues.

**3. Shaping Media and Public Narratives**

Media outlets and journalists can leverage our insights to provide nuanced reporting on election outcomes and demographic trends. Rather than broad generalizations, data visualizations showcasing how demographics influence voting patterns offer a clear and evidence-based foundation for election narratives.

## **Links to Disciplines or Industries**

***Disciplines:***

**1. Sociology and Psychology:** Exploring how demographic and behavioral factors interact to shape individual political preferences.

**2. Political Science:** Analyzing swing states, enhancing predictive models of electoral outcomes, and designing tailored campaign strategies and political communication.

**3. Data Science and Visualization:** Advancing methods for transforming complex political data into actionable insights and compelling visualizations.

***Industries:***

**1. Campaign management organizations:** Creating campaign strategies that target a certain type of voters.

**2. Data analytics companies specializing in political behavior:** Analyzing demographic and voting data for various usage, including campaign management and policy making.

**3. Media and journalism:** Reporting and communicating more credible election narratives and educating the public about the complexity of political behaviors.

**4. Local and state governments:** Delving into the study of socioeconomic drivers behind demographic shifts and making related policies.

# **Methodology**

## Data Sources and Integration

This research integrates voting data and geospatial data to create the geographic map shown above, which visualizes county-level voting trends in North Carolina for the 2024 U.S. Presidential Election. The integration combines datasets to provide a comprehensive view of voter distribution and geographic patterns of political preference.

The dataset provided by the North Carolina State Board of Elections (NC SBE 2024) includes county-level voting results, specifying the total number of votes cast and their distribution across political parties (Republican, Democratic, and Unaffiliated). This data is crucial for the analysis as it enables the visualization of voting intensity and party preferences across counties, offering quantitative insights essential for understanding electoral dynamics in North Carolina.

The NC OneMap geospatial dataset provides shapefiles that define the county boundaries within North Carolina (NC OneMap n.d.). This data is essential for accurately mapping voting results to their respective county locations, allowing for spatial analysis of electoral patterns across the state. By integrating this geospatial information with voting data, the study can visualize and analyze electoral trends within specific geographic areas.

The integration of data involves joining the voting data from the North Carolina State Board of Elections (NCSBE) with geographic shapefiles based on the common attribute of county name. This linkage enables the accurate spatial representation of voting data. The integrated dataset is then visualized as a choropleth map, where each county polygon is color-coded according to the number of votes cast, with the intensity of the color reflecting voter turnout or voting patterns. To enhance the map's interpretability, county labels, and a color legend are added, providing additional context and clarity in understanding the electoral distribution across counties.

The integration of voting data with geographic information provides valuable insights into spatial voting patterns across North Carolina. The resulting map highlights regional trends, such as urban counties like Wake and Mecklenburg, which exhibit higher voter turnout, as evidenced by their darker shades, compared to rural areas with lighter coloring. Additionally, the integration facilitates the identification of geographic clusters where voter engagement or party preferences are concentrated, offering a deeper understanding of political dynamics that is not apparent in tabular data alone. This spatial analysis allows for comparative assessments between counties, supporting both exploratory and explanatory analyses of electoral trends and behaviors across the state.

By integrating vote counts with geographic boundaries, this visualization transforms raw data into actionable insights, highlighting regional variations in voter participation and enabling deeper analysis of political trends.

## Age Distribution by Party

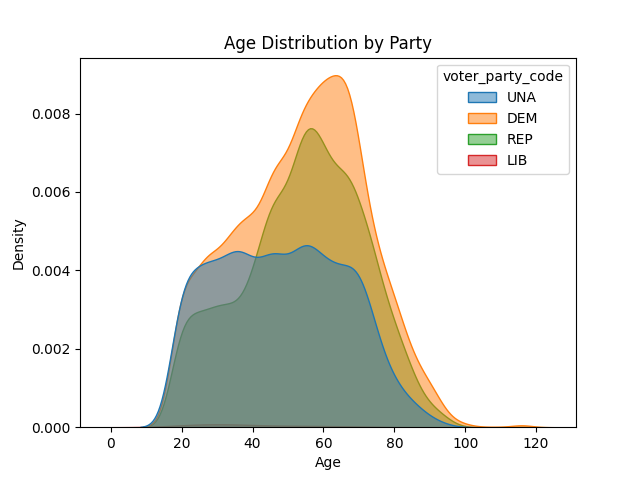


Fig. 3: Age Distribution of Voters by Party Affiliation in the 2016 U.S. Presidential Election (Created by Python)

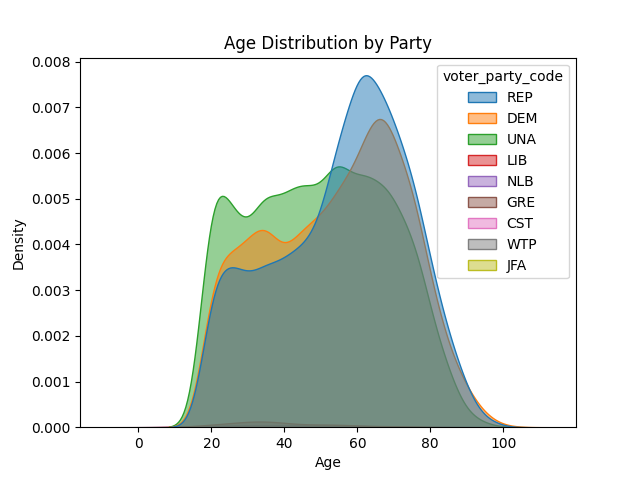


Fig. 4: Age Distribution of Voters by Party Affiliation in the 2024 U.S. Presidential Election (Created by Python)

Visualization Design

This research employs overlapping density plots to represent the *Age Distribution of Voters by Party Affiliation* effectively. As shown in Fig 3 and 4, the visualization uses:

Marks and ChannelsIn the visualization, density curves serve as area marks, representing the relative probability density of voter ages. The channels used to encode the data are as follows:

In the visualization, the position on the x-axis represents age, a continuous variable that provides a clear representation of the age distribution across the dataset. The height on the y-axis encodes density, which indicates the relative frequency of voters within specific age ranges. Additionally, color is used to differentiate party affiliation—Unaffiliated, Democratic, Republican, and Libertarian—by assigning distinct hues to each party. This color distinction not only facilitates easy identification of party preferences within age groups but also enhances the clarity of the relationships between age and political affiliation, making it easier to interpret patterns in voter behavior.

Spatial Arrangement

The data is displayed on a shared Cartesian coordinate system, with overlapping curves for each party affiliation. This spatial arrangement facilitates direct comparisons between the different parties, allowing for the identification of areas of convergence, where shared demographic characteristics are evident, and divergence, where distinct differences in party preferences and age distributions emerge. By overlapping the curves, the visualization effectively highlights both similarities and contrasts in voter behavior across political affiliations.

Data and Task Abstraction

The dataset includes both quantitative and categorical data types. The quantitative variables, age, and density, are continuous, representing the distribution of voter ages and the relative frequency of voters within each age group, respectively. The categorical variable, *voter\_party\_code*, represents party affiliation, categorizing voters into political parties such as Democratic, Republican, and Unaffiliated.

This visualization supports both exploratory and explanatory tasks. For exploratory analysis, it enables the identification of patterns, overlaps, and differences in voter age distributions across political parties, providing insights into how demographic characteristics vary. For explanatory purposes, it highlights trends such as the prevalence of older or younger voters within each party, shedding light on party-specific demographic preferences.

The use of Kernel Density Estimates (KDE) serves as an abstraction technique to smooth the data, reducing noise and offering a continuous, clearer representation of voter age patterns. This method enhances the interpretability of the data by providing a more coherent and detailed view of the relationships between age and party affiliation.

Justification of Techniques

The visualization employs marks and channels to effectively encode both continuous and categorical data, enhancing intuitive understanding. The use of Kernel Density Estimates (KDE) smooths the data, reducing visual clutter while still preserving meaningful trends, such as peaks and overlaps in the distribution of voter ages. The overlapping curves allow for simultaneous comparisons between different parties, minimizing cognitive load and enabling the viewer to quickly grasp key differences and similarities. This approach supports both exploratory analysis, by uncovering patterns and relationships, and explanatory communication, by clearly conveying insights to a diverse audience. The integration of these elements ensures that the visualization is both analytically robust and easily interpretable.

These design choices ensure accurate, interpretable, and actionable visualizations for stakeholders.

## Voter Distribution by Age Group and Party

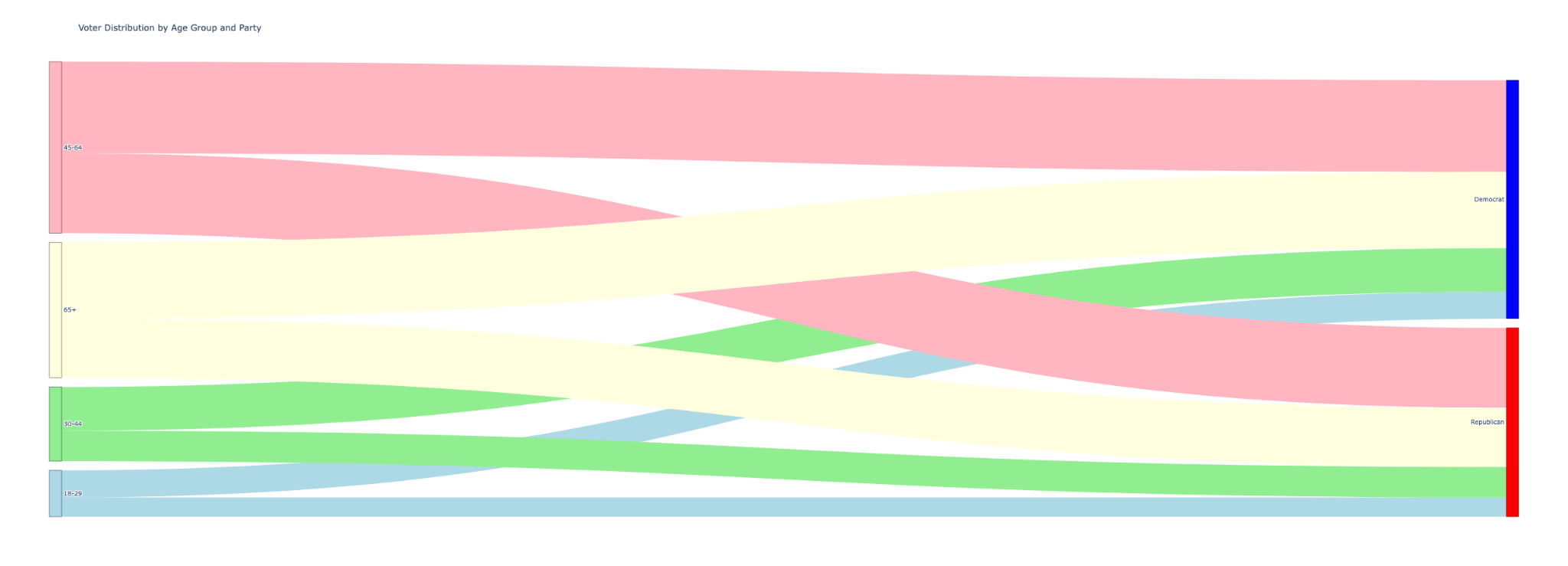
****

Fig. 5: Voter Distribution by Age Group and Party in the 2016 U.S. Presidential Election (Created by Python)

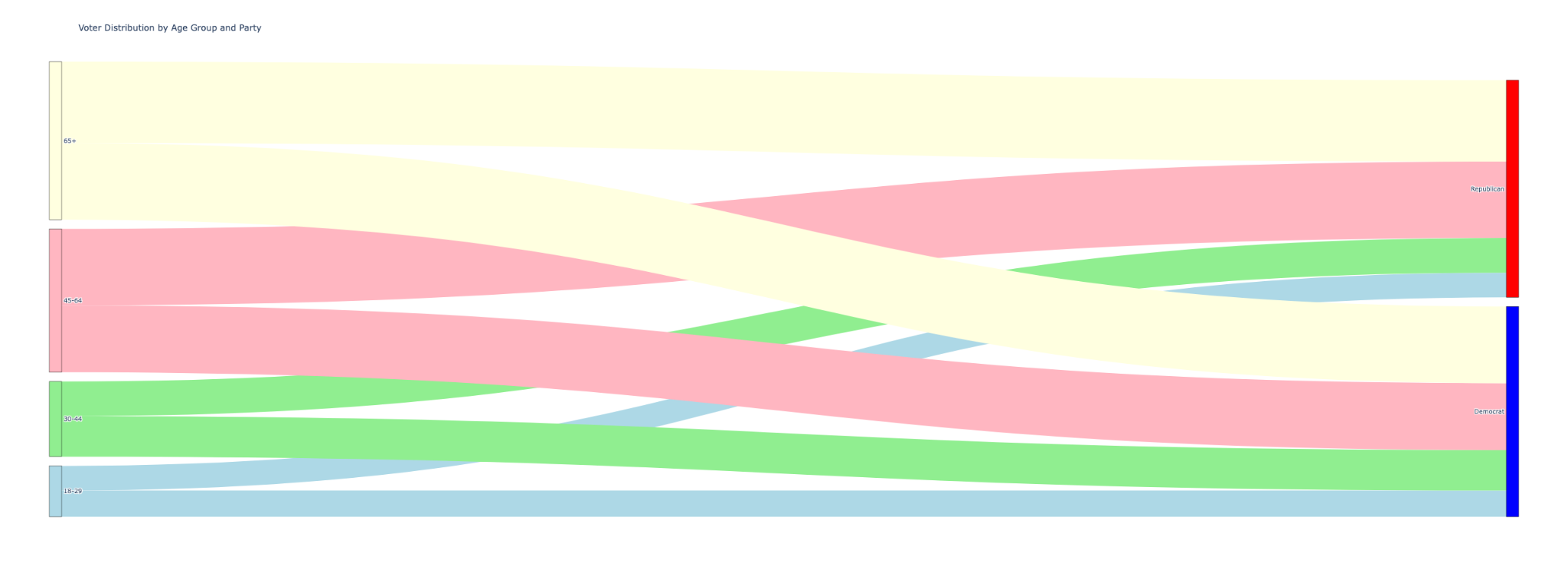
****

Fig. 6: Voter Distribution by Age Group and Party in the 2024 U.S. Presidential Election (Created by Python)

The visualization, Fig. 5 and 6, uses a Sankey diagram to represent the flow of voter age groups to political party affiliation, emphasizing both proportional relationships and transitions. Below, we detail the techniques used, focusing on Marks and Channels, Arrangement of Data, and Abstraction, with a justification for each design choice.

Visualization Design

The Sankey diagram effectively represents the relationship between age groups and political party affiliation.

In the visualization, marks are utilized to represent both age groups and political parties. The nodes correspond to specific age groups (18–29, 30–44, 45–64, 65+) positioned on the left side, and political parties (Democrat and Republican) positioned on the right side. The links between the nodes illustrate the proportion of voters transitioning from each age group to a particular party.

The channels employed to encode the data include the width of the links, which represents voter proportions. Wider links signify higher percentages of voters transitioning from a given age group to a specific party. Color is used to differentiate the age groups, with each group assigned a distinct pastel hue, while the parties are represented by conventional colors: blue for Democrats and red for Republicans. The position of the nodes is carefully arranged, with age groups vertically aligned on the left and political parties on the right, creating a hierarchical flow that facilitates the visualization of voter transitions between different age groups and political affiliations.

Width highlights proportions effectively, color enhances categorical distinction, and position emphasizes relationships between variables.

The arrangement of data in the visualization is carefully structured to enhance clarity and understanding. Node placement is based on the population size of the age groups, with larger groups placed higher in the hierarchy, while political parties are vertically aligned to maintain a clear flow. The flow direction is horizontal, with links extending from the age groups to the political parties, representing the logical transition from the independent variable (age) to the dependent variable (party affiliation). Finally, proportionality is preserved in the flow of the links, ensuring that the width of the links accurately reflects the quantitative proportions of voters transitioning from each age group to a given party. This arrangement ensures that the

This arrangement simplifies comparisons while maintaining intuitive reading patterns and preserving the hierarchical structure.

Data Abstraction

In the visualization, aggregation is applied by grouping ages into four standard demographic categories—18–29, 30–44, 45–64, and 65+—to enhance clarity and make the data more comprehensible. This categorization simplifies the age distribution, allowing for more effective comparison. To ensure proportional representation, the data is normalized, which allows for meaningful comparisons across groups of varying sizes. This normalization ensures that differences in the size of the age groups do not distort the analysis, providing an accurate representation of voter behavior within each category.

Aggregation reduces complexity while retaining interpretability, and proportional representation highlights trends effectively.

Task Abstraction

The visualization supports both exploratory and explanatory tasks. Exploratory analysis helps identify the distribution of voters across different age groups and political parties, highlighting demographic trends and uncovering key patterns in voter behavior. This allows for an in-depth understanding of how age and party affiliation interact. In terms of explanatory tasks, the visualization effectively communicates findings about voter demographics and party preferences, making these insights accessible to both technical and non-technical audiences. By presenting the data intuitively and clearly, it facilitates understanding across a wide range of viewers.

Sankey diagrams support exploration by revealing patterns and relationships while effectively communicating results in a visually engaging manner.

Justification of Techniques

The visualization leverages marks and channels to facilitate an intuitive understanding of the data. The width of the links encodes the proportions of voters, while color distinguishes the categories of age groups and political parties, making the relationships between these variables immediately clear. The hierarchical arrangement of nodes follows a left-to-right flow, aligning with natural reading patterns, which enhances the ease with which users can interpret the data. In terms of data abstraction, the use of aggregation and normalization simplifies complex data, yet preserves interpretive depth, ensuring that key insights about voter distribution are not lost. The bipartite design—with age groups on the left and political parties on the right—emphasizes the relationships and proportions between the two categories, making it particularly well-suited for demographic analysis and comparison.

## Voter Distribution by Race and Party

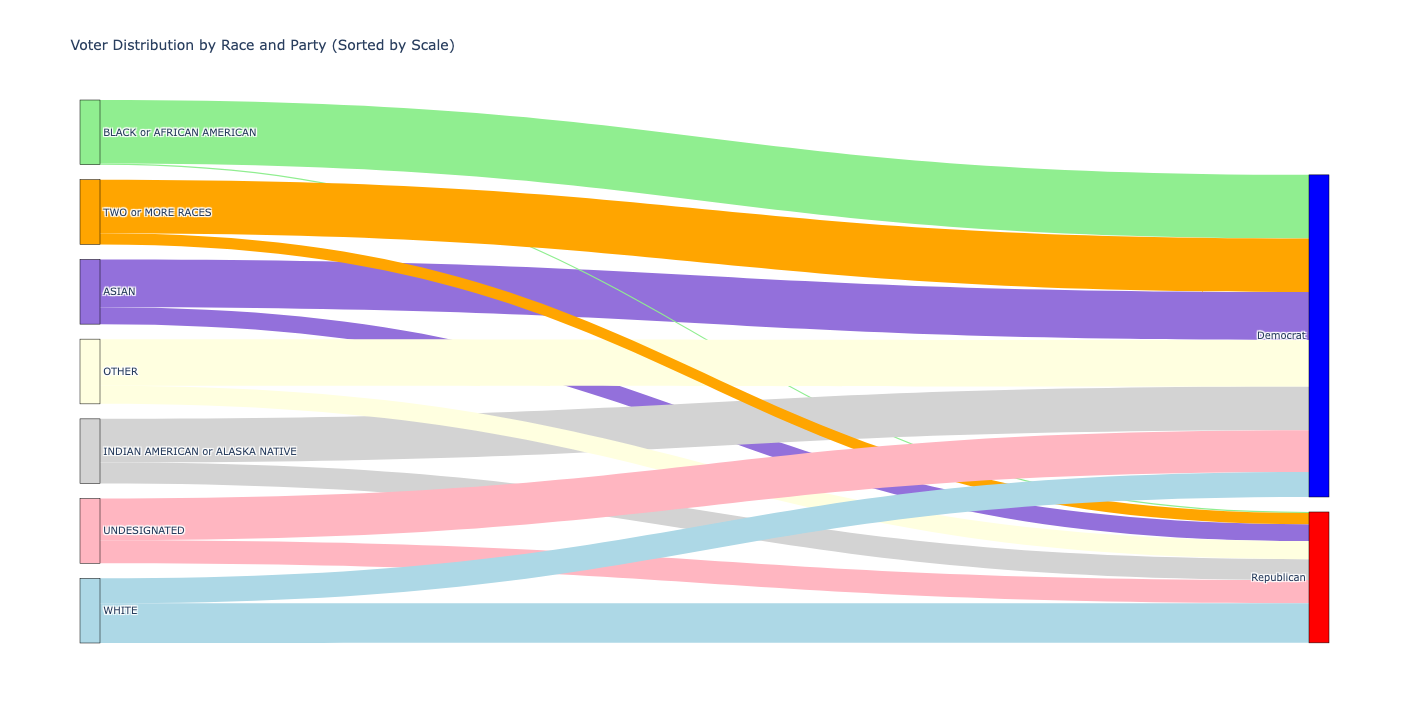


Fig. 7: Voter Distribution by Race and Party in the 2016 U.S. Presidential Election (Created by Python)

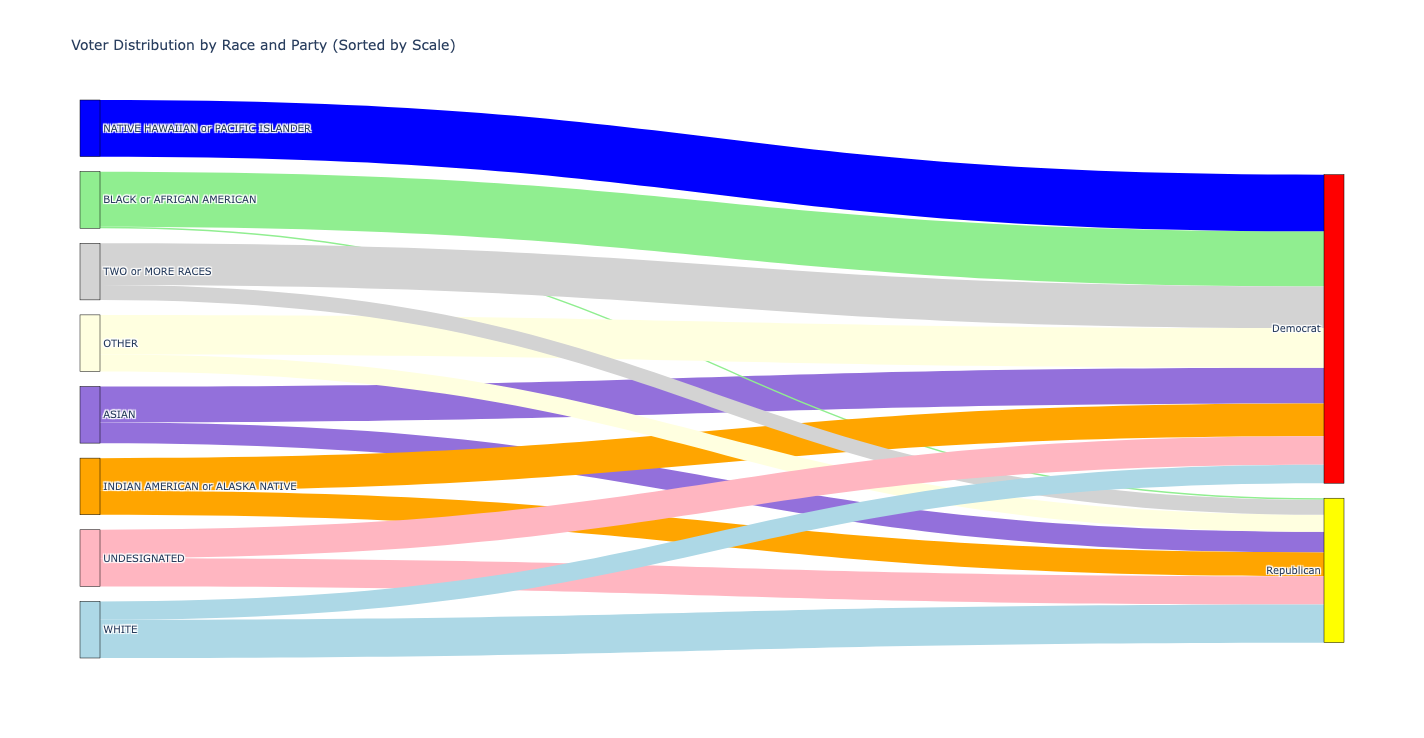
****

Fig. 8: Voter Distribution by Race and Party in the 2024 U.S. Presidential Election (Created by Python)

Visualization Design

This visualization uses a Sankey diagram to illustrate the proportional relationships between racial groups and political party affiliations (Democrat and Republican).

In the visualization, marks are used to represent both racial categories and political parties. The nodes on the left represent various racial categories, such as Black or African American, Asian, and White, while the nodes on the right represent political parties. The links between these nodes illustrate the proportion of voters from each racial group affiliating with a particular party.

The channels employed in the visualization include the width of the links, which encodes the proportion of voters, with wider links indicating larger voter groups transitioning to a specific party. Color is used to differentiate racial categories by assigning distinct hues—green for Black or African American, orange for Two or More Races—and political parties are represented by conventional colors, such as blue for Democrats and red for Republicans. The position of the nodes is arranged vertically, with racial groups aligned by population size, with the most populous groups placed at the bottom, while political parties are positioned on the right. This arrangement ensures a logical flow from the input (race) to the output (party), enhancing the clarity of the relationships between racial categories and political affiliation.

Width effectively represents proportions, color highlights categories, and nodes and links simplify the relationships for easy pattern recognition.

Arrangement of Data

In the visualization, node placement is strategically arranged so that racial groups are sorted by size, with the most populous groups placed at the bottom. Political parties are aligned vertically on the right to maintain clarity and facilitate comparison. The flow direction is horizontal, with links flowing from racial groups to political parties, aligning with natural reading patterns and making the visualization easier to follow. Proportional representation is ensured through the width of the links, which corresponds to the proportion of voters affiliating with each party. This approach accurately reflects the contributions of each racial group to the overall party affiliation, allowing for an intuitive understanding of the data.

The Sankey diagram simplifies comparisons and prioritizes significant trends by grouping data logically and proportionally.

Data Abstraction

Aggregation ensures clarity, and normalization emphasizes trends without being skewed by population differences.

Task Abstraction

The Sankey diagram supports both exploratory analysis and clear communication by presenting complex relationships in an intuitive, engaging format.

The chosen techniques in the visualization are carefully justified to enhance clarity and interpretability. Marks and channels are employed to intuitively encode both quantitative and categorical data, with width representing the proportion of voters and color differentiating racial categories and political affiliations. This dual encoding allows for a clear and immediate understanding of the relationships between race and party affiliation.

The data arrangement follows a left-to-right flow, with racial groups on the left and political parties on the right. This hierarchical sorting by size and alignment with natural reading patterns simplifies the interpretation of the data and emphasizes key trends, such as the dominant party preferences within each racial group.

In the visualization, aggregation of categories groups racial demographics into Census-standard categories, such as Black or African American, White, and others, to simplify the data and ensure consistency with commonly used demographic classifications. The focus is on the two dominant political parties, Democrats and Republicans, as they are the primary variables of interest. Normalization is applied to convert raw voter counts into proportions, allowing for fair comparisons across racial groups of varying sizes. This process ensures that differences in group sizes do not distort the analysis, providing a more accurate representation of how each racial group is affiliated with the political parties.

Data abstraction is achieved through aggregation and normalization, which reduces complexity while maintaining the richness of the data. By grouping racial demographics into

Census-standard categories and converting raw counts into proportions, the visualization enables comparison across groups of varying sizes without losing critical insights.

Finally, the bipartite design directly links racial groups to political affiliations, creating a clear visual representation of voting patterns. This design effectively highlights the relationships between race and party affiliation, making the trends in voting behavior both visible and actionable for further analysis or decision-making.

## Party Preference by County

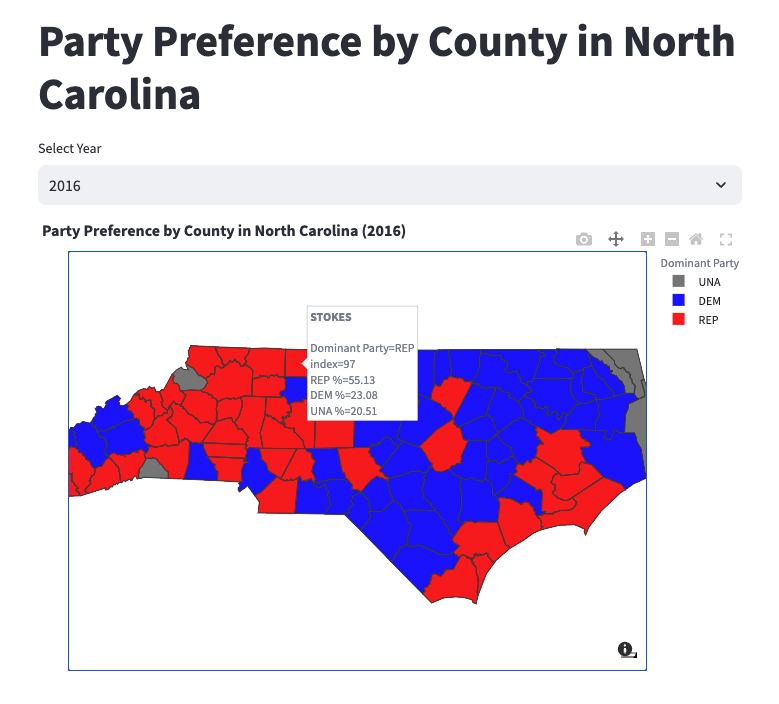
****

Fig. 9: Party Preference by County in the 2016 U.S. Presidential Election (Created by Python)

Visualization Design

This research employs an interactive geographic map to analyze and communicate Party Preference by County across the 2016, 2020, and 2024 U.S. Presidential Elections. The map dynamically visualizes spatial data to show the dominant political party in each county, with color coding used to represent party affiliations (Republican, Democrat, Unaffiliated, and No Data).

In the visualization, marks are utilized to represent spatial regions through polygons, with each polygon corresponding to the boundaries of a county. These polygons serve as the foundational elements of the choropleth map, enabling the depiction of geographically specific data.

The channels employed include color and position to encode critical information. Color is used to differentiate party affiliations, with distinct hues representing each category—red for Republicans, blue for Democrats, gray for Unaffiliated, and dark gray for regions with no data. This use of color ensures that the viewer can easily identify the dominant party or data status in each county. The position reflects the geographic arrangement of counties, with the x- and y-coordinates corresponding to their longitude and latitude. This spatial alignment preserves the geographic accuracy of the visualization, allowing the map to serve as both a data representation tool and a spatial reference.

Polygons ensure spatial accuracy, color distinguishes party categories, and geographic position aligns with viewer expectations for maps.

The interactive features of the visualization are designed to enhance user engagement and provide a comprehensive understanding of the data. One key feature is year selection, allowing users to choose from 2016, 2020, or 2024 to dynamically update the map and view the election data corresponding to the selected year. This flexibility enables users to explore changes and trends over time, offering insights into how voting patterns have evolved across North Carolina.

Another important feature is hover information, which allows users to interact with the map at a granular level. By hovering over a specific county, users can access detailed information about that region. This includes the county name, the dominant political party (Republican, Democrat, or Unaffiliated), and the proportional breakdown of votes for each party. This feature makes the data highly accessible, allowing viewers to quickly gather insights without needing additional tools.

The spatial data arrangement further enhances the clarity of the visualization. Counties are represented as polygons, accurately outlining their boundaries and enabling users to identify geographic patterns effectively. To ensure the map is intuitive, a compact legend is strategically placed, explaining the color encoding for party affiliations. The legend is designed to provide clarity without interfering with the map itself. By preserving spatial relationships and offering a concise legend, the visualization highlights regional trends while maintaining ease of interpretation for the audience.

The data abstraction in the visualization is designed to simplify complex datasets while retaining key insights. Voting data is aggregated at the county level, with each county encoding the dominant political party. This approach reduces the complexity of the data, allowing viewers to focus on regional trends rather than being overwhelmed by individual data points. Counties are further categorized into four distinct groups: Republican, Democrat, Unaffiliated, and No Data. This categorization ensures clarity and provides a straightforward framework for interpreting the data.

The task abstraction supports both exploratory and explanatory objectives. Exploratory tasks are facilitated by the map's ability to help users identify clusters of party preference, detect spatial outliers, and analyze the balance of political affiliations across regions. At the same time, the visualization enables explanatory tasks, such as communicating regional trends and highlighting areas of political dominance or competition. By supporting both exploration and explanation, the map is well-suited for diverse audiences seeking to understand voting patterns in North Carolina.

The justification of techniques underscores the effectiveness of the design. Polygons are used as marks to represent counties, and color is employed to encode categorical data, effectively differentiating between political parties. The spatial arrangement aligns counties geographically, preserving an intuitive interpretation of trends and relationships. Aggregation simplifies the data while ensuring that key insights are retained, allowing users to focus on meaningful patterns. Overall, the visualization is designed to support both detailed exploration and effective communication of findings, making it an invaluable tool for analyzing voting behaviors.

## Votes by County

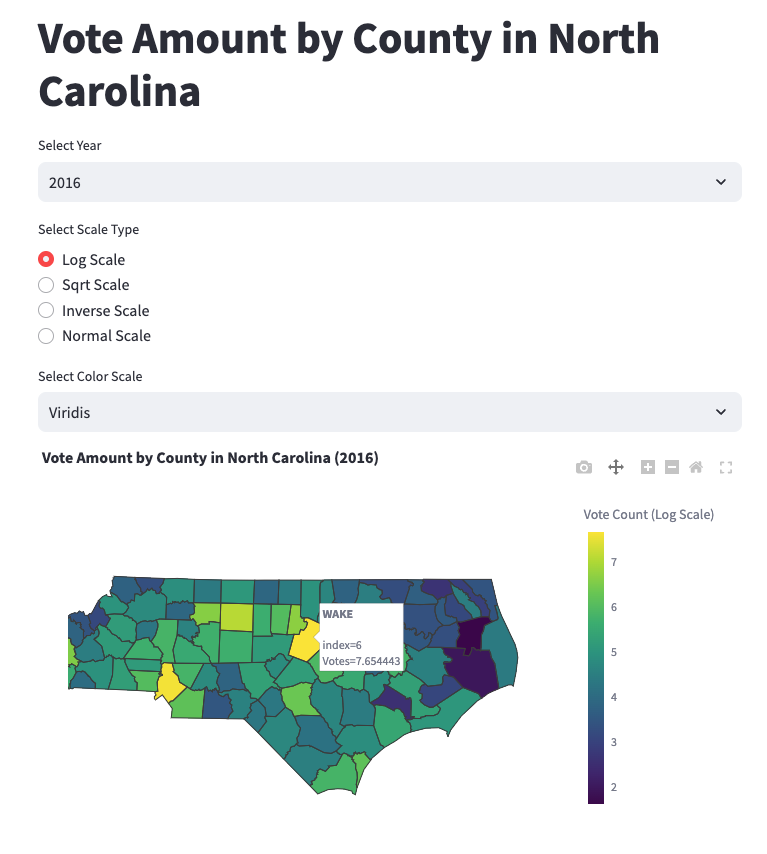
****

Fig. 10: Votes amount by County in the 2016 U.S. Presidential Election (Created by Python)

Visualization Design

This research leverages an interactive geographic map to analyze and communicate Vote Amounts by County across the 2016 to 2024 U.S. Presidential Elections. The map dynamically visualizes spatial data to show the total vote count per county, providing users with multiple customization options, including year selection, data scale, and color mapping.

In this visualization, marks take the form of polygons representing county boundaries. It encodes spatial regions and provides a geographical framework for the data.

The channels used include color and position. Color is utilized to encode vote counts, employing customizable color scales such as Viridis or Plasma to differentiate the magnitude of votes across counties. This channel effectively highlights variations in voting intensity, making patterns easily discernible. The position reflects the actual geographical arrangement of counties, using their real-world coordinates (longitude and latitude). This spatial alignment ensures that the visualization accurately corresponds to the physical geography of the represented area, enhancing interpretability and relevance.

Polygons ensure spatial accuracy, color effectively encodes numerical data (vote counts), and geographic positioning provides a familiar and intuitive layout for users to interpret.

The interactive features of the visualization are designed to provide users with flexibility and control over how they view and analyze the data. Year selection allows users to choose from 2016, 2020, or 2024 to dynamically update the map and explore vote count data for the corresponding election year. This functionality supports comparative analysis across different election cycles, helping users identify trends and shifts in voter turnout over time.

Data scale options further enhance the usability of the map by enabling users to adjust the visual representation of vote counts. The logarithmic scale compresses large vote counts, making it easier to compare smaller values. The square root scale balances mid-range values for a more equitable distribution, while the inverse scale highlights areas with lower vote counts. For an unaltered view, the normal scale displays raw vote counts. These scaling options empower users to customize the visualization based on their analytical needs.

The color map selection feature allows users to choose from a variety of color palettes, such as Viridis, Inferno, and Plasma. This flexibility ensures that users can tailor the visual style to their preferences or emphasize specific trends, improving accessibility and interpretability. Additionally, the hover information feature provides detailed insights for each county, displaying the county name and total vote count when the user hovers over a specific region on the map. This interactivity makes it easy to explore granular data without overwhelming the user.

The spatial data arrangement preserves the geographic relationships between counties by representing them as polygons, with boundaries that accurately reflect their real-world locations. A compact legend is included to explain the color encoding, ensuring users can interpret the map quickly and accurately. This geographic layout highlights regional patterns while maintaining intuitive navigation and understanding.

The data abstraction simplifies complex datasets to focus on meaningful insights. Voting data is aggregated at the county level, with each county encoding the dominant political party. This aggregation reduces complexity while retaining regional trends. Counties are categorized into four groups: Republican, Democrat, Unaffiliated, and No Data, which ensures clarity and enhances interpretability for users.

The visualization supports both exploratory tasks and explanatory tasks. Users can explore the data to identify clusters of party preference, detect spatial outliers, and analyze regional trends. Additionally, the map is effective in communicating geographic patterns of political dominance or competition, offering a clear narrative for diverse audiences. This dual functionality ensures that the visualization serves both as a tool for in-depth analysis and as a medium for effective communication.

The techniques used in the visualization are well-justified. Polygons are used as marks to represent counties, and color is employed as a channel to effectively encode categorical data, such as political party dominance. The spatial arrangement aligns with real-world geography, making the map intuitive and straightforward to interpret. The combination of aggregation and categorization simplifies the data without losing key insights, while the interactive features support both the exploration and presentation of voting patterns. Overall, the visualization is carefully designed to balance complexity and usability, making it an invaluable tool for analyzing and communicating election data.

Age Distribution of Party Preferences

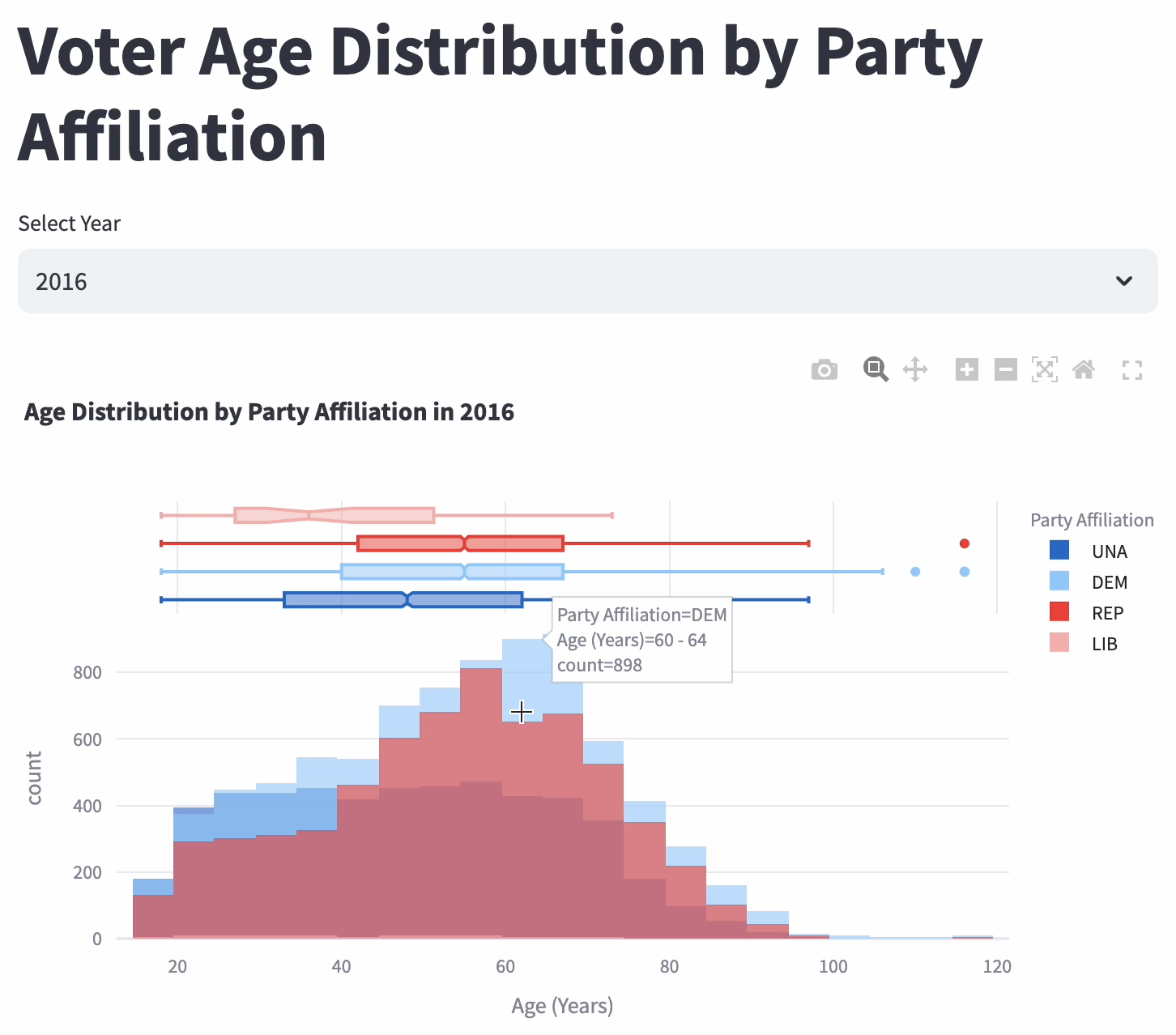


Fig. 11: Voter age distribution by party affiliation in the 2016 presidential election (created by Python).

Visualization Design

For this visualization, the focus is on analyzing and presenting voter age distributions across different political party affiliations in 2016. The design utilizes an interactive histogram with overlaid boxplots to enable both exploratory and explanatory insights into voter demographics.

In this visualization, the marks include histogram bars and boxplot elements. The histogram bars represent voter counts, while the boxplots encode statistical summaries (median, interquartile range, and outliers) of voter ages for each party affiliation. The combination of these marks allows for a detailed exploration of age distributions across parties.

The channels employed in the design are color, position, and size. Color is used to differentiate party affiliations, with distinct hues assigned to Unaffiliated, Democrat, Republican, and Libertarian voters. This helps users quickly identify trends in voter preferences by age. Position along the x-axis encodes the continuous variable of age, providing a clear visual representation of the distribution. The y-axis encodes voter counts or densities, allowing for direct comparisons between age groups. The size of the histogram bars represents the frequency of voters within each age bin.

Interactive features, such as hover-over tooltips, further enhance the utility of the visualization. When a user hovers over a specific area, detailed information about voter counts, age ranges, and party affiliations is displayed, offering immediate insights without cluttering the visualization. Additionally, users can select different years using a dropdown menu, enabling comparative analysis of voter age distributions across election cycles.

The data abstraction in this visualization simplifies raw voting data into aggregated age groups, ensuring clarity while preserving essential trends. The overlaid boxplots provide a summary of key statistical metrics, reducing noise and allowing users to focus on meaningful patterns, such as the central tendencies and variability in voter ages for each party.

This visualization supports both exploratory and explanatory tasks. For exploratory purposes, it allows users to investigate patterns, overlaps, and differences in age distributions across political affiliations. For explanatory purposes, it highlights demographic trends, such as the concentration of older voters within certain parties, providing a clear narrative for diverse audiences.

The design techniques are well-justified. Histogram bars effectively encode numerical data, while boxplots summarize statistical insights. The color differentiation and x-axis alignment simplify comparisons across party affiliations. Interactive features further enhance user engagement, making the visualization a powerful tool for analyzing voter demographics. The combination of these elements ensures that the visualization is both intuitive and informative, facilitating a deeper understanding of voter behavior across different political affiliations.

Sankey Diagram of Demographical Factors Cross-visualization

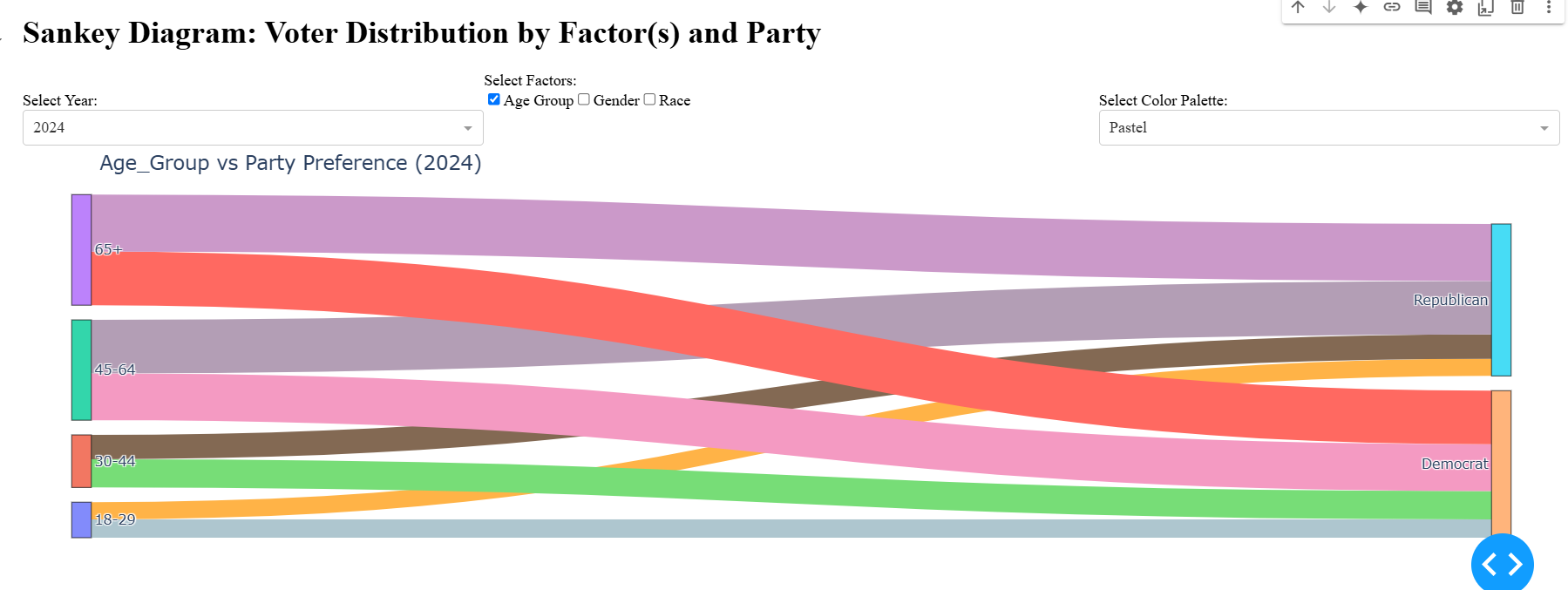


Fig. 12: Voter Distribution by Factors and Party

Visualization Design This research utilizes a Sankey diagram to analyze and communicate voter distribution across demographic factors and political affiliations for the 2024 U.S. Presidential Election. The visualization dynamically illustrates the relationships between age groups, gender, race, and party preferences, allowing users to explore how demographic characteristics influence political alignment. Users can customize the diagram through interactive options, such as selecting demographic factors and adjusting the color palette, enabling deeper analysis of the data.

In this visualization, the marks take the form of nodes and flows. Nodes represent demographic groups (e.g., age, gender, race) and political parties, while flows connecting the nodes encode the proportions of voters transitioning from each demographic group to their political preferences. The channels used include width and color. The width of the flows encodes the relative voter size for each group, emphasizing the magnitude of connections, while the color differentiates demographic groups for clarity and aesthetic appeal.

The Sankey diagram's left-to-right layout organizes the data hierarchically, with demographic groups on the left and political parties on the right. This spatial arrangement mirrors natural reading patterns and enhances the interpretability of the data flow. Interactivity further empowers users to explore the influence of different demographic combinations on voter distribution, making the visualization both an exploratory and explanatory tool.

The techniques employed ensure that complex datasets are abstracted into an intuitive and meaningful representation. Aggregating data at the demographic level simplifies interpretation without losing essential insights. The interactive options and customizable color palettes cater to user preferences and analytical needs, making the visualization a versatile tool for understanding voter dynamics in the 2024 election.

Advanced tool

The visualizations in Fig. 4 to Fig. 7 leverage real-time interactive features to enhance user engagement and analytical depth. These visualizations provide dynamic exploration of data, allowing users to zoom, filter, and hover for additional details, enabling granular insights into party preferences, voter turnout, and demographic patterns.

# **Results**

The project uses the multifactorial approach aiming to compare the results from the 2016 election and the 2024 election considering demographic factors. Below are the observations we found.

1. Demographic Impact on Voter Turnout

Age Groups:

Fig. 2 and Fig. 4 illustrate the voter distribution by age group and their party affiliation in 2016. The width of the bands indicates the number of voters in each age group. As shown in the figure, the 18–29 group contributed the smallest share in 2016. The 45–64 have the largest voter bases, which reflects their higher voter turnout compared to younger groups. The figure witnesses wider Democratic flows from younger groups, which signifies stronger Democratic appeal among younger voters in 2016. However, in Fig. 3 and Fig. 5, the percentage of the 18–29 age group who voted for Republicans is almost as even as that for Democrats. People over 65 years old share the largest contribution of total votes, exceeding that of the 45–64 age group. The age distribution charts for both elections highlight a consistent trend: younger voters (18-29) have lower turnout rates, with their density curve being substantially smaller than those of older age groups. This is especially evident in 2016 when younger voter participation was notably low. In contrast, voters aged 45-64 and 65+ demonstrate consistent and significant turnout across both elections.

Fig 11. highlights significant differences in the age distributions among political party affiliations. Older age groups (60+) show a higher concentration of voters affiliating with Republican and Democrat parties, while younger age groups (18-29) are more evenly distributed across party affiliations, with a noticeable presence in Unaffiliated voters. This trend suggests that older voters tend to have more defined party preferences, while younger voters exhibit more diversity or neutrality in political alignment.

Fig. 12 reveals how age groups are distributed across political parties in the 2024 election. The "65+" age group heavily flows towards the Republican Party, while the "30-44" and "18-29" age groups show a balanced distribution between Democrats and other affiliations. This indicates a generational divide in party preference, where older voters lean more conservative, while younger voters are more inclined towards progressive or independent positions.

Race and Ethnicity:  
 The 2024 voter distribution by race and party in Fig. 7 shows that White voters continue to dominate turnout, consistent with historical trends in Fig. 6. Minority groups such as African Americans and Hispanic/Latino populations remain underrepresented, reflecting persistent systemic barriers. These barriers may include accessibility issues, economic disparities, and voter suppression tactics in certain regions. However, there is evidence of marginal improvement from visualized data on turnout among minority groups in urban areas, likely due to targeted mobilization efforts and increased representation by Democratic candidates.

2. Geographic Clustering

Urban vs. Rural Divide:

The geographic clustering of votes demonstrates the urban-rural divide. Urban centers such as Raleigh, Durham, and Charlotte in North Carolina show robust Democratic support. At the same time, less densely populated regions overwhelmingly favor Republican candidates in Fig. 9. This divide might be rooted in cultural and socioeconomic differences, with urban areas embracing diversity and innovation. At the same time, rural communities prioritize traditional values and resist rapid societal changes.

# **Intellectual Merit and Practical Impacts**

## Academic Contribution

Innovative Visualization Techniques: By using interactive Sankey diagrams, choropleth maps, and time-series comparisons, this project showcases novel methods for visualizing complex, multi-dimensional election data. These approaches highlight relationships between variables such as ethnicity, education, economic status, and voter turnout, inspiring similar studies in other states or countries.

Integration of Temporal and Spatial Analysis: The focus on election results from both 2016 and 2024 enables researchers to explore temporal dynamics alongside spatial patterns. This combined approach enriches academic literature on longitudinal election studies and provides a framework for future analyses.

Open-Source Methodology: By using publicly available data sources and widely accessible tools like Python, Plotly, and open data repositories, this project contributes to the reproducibility and scalability of research in data visualization and political analysis.

## Practical Impacts

Informing Policy and Decision-Making: The findings from the project can help policymakers identify regions or demographic groups with lower voter turnout. This insight can guide resource allocation for voter education campaigns or infrastructure improvements, especially in underserved communities.

Enhancing Voter Engagement: By presenting clear and engaging visualizations, this project can raise public awareness of the factors influencing voter turnout and party preference. This could inspire more informed civic participation among citizens.

Improving Electoral Strategies: Political parties and advocacy groups can use the research to tailor campaign strategies, focusing on areas where specific demographic or socioeconomic factors strongly correlate with voter behavior.

**Acknowledgment**  
 This final project is submitted for INFOSCI 301: Data Visualization and Information Aesthetics, instructed by Professor Luyao Zhang at Duke Kunshan University in Autumn 2024. We extend our heartfelt appreciation to Professor Zhang for her insightful guidance and to ChatGPT for its support in generating and refining code.

# **Supplementary Materials**

Github link: <https://github.com/ClAy140/InfoSci301-DKU-Fall2024/tree/main/Group_Project>

# **Bibliography**

Munzner, Tamara. “Visualization Analysis and Design.” O’Reilly Online Learning, 2014. https://learning.oreilly.com/library/view/visualization-analysis-and/9781466508910/?sso\_link=yes&sso\_link\_from=duke-university.

Ncsbe.gov. “NC SBE Contest Results,” November 20, 2024. https://er.ncsbe.gov/?election\_dt=11/05/2024&county\_id=0&office=FED&contest=0.

SANDERS, LINLEY. “How 5 Key Demographic Groups Voted in 2024: AP VoteCast.” AP News, November 7, 2024. https://apnews.com/article/election-harris-trump-women-latinos-black-voters-0f3fbda3362f3dcfe41aa6b858f22d12.

Tyson, Alec, and Shiva Maniam. “Behind Trump’s Victory: Divisions by Race, Gender, Education.” Pew Research Center, November 9, 2016. https://www.pewresearch.org/short-reads/2016/11/09/behind-trumps-victory-divisions-by-race-gender-education/.

www.nconemap.gov. “NC OneMap,” n.d. https://www.nconemap.gov/.