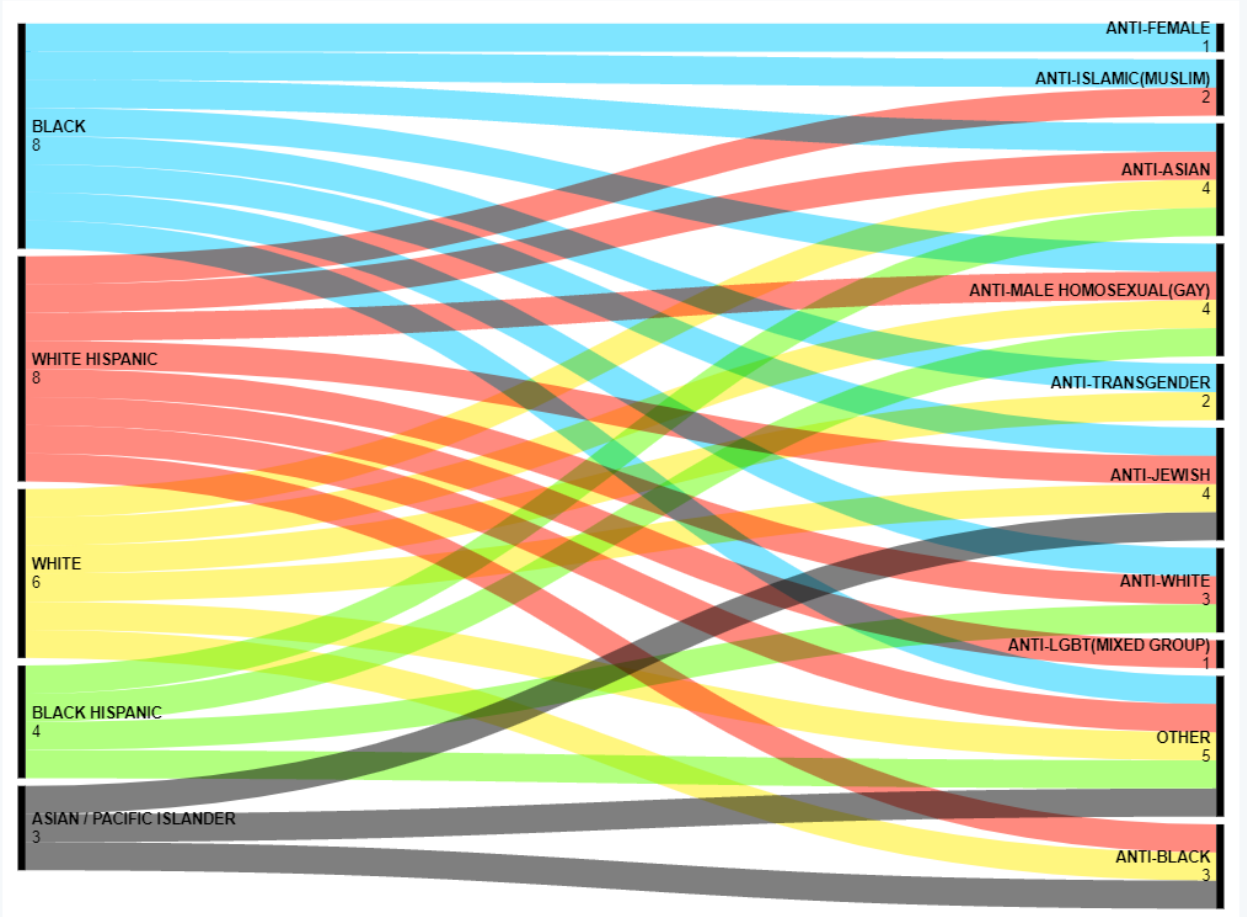
Assignment 1A – INFO 250 Data Visualization

Tapasya Sharma

Chosen dataset - *Hate Crimes in New York*

Visual 1 – *Sankey Diagram showing the relationship between race and bias-motivation for the crime*

1. Steps taken
   a. Selecting the Source Node as Race and Target Node as Bias-Motivation under the Sankey Diagram tab
   b. Selecting colors as 'Categorical 10' and individually modifying the color choice to have the most distinct ones.
   c. Labels – Show node values
2. Observations – A clean Sankey Diagram appears with all the required labels.



Visual 2 – *Tree map describing the relationship between gender and race.*

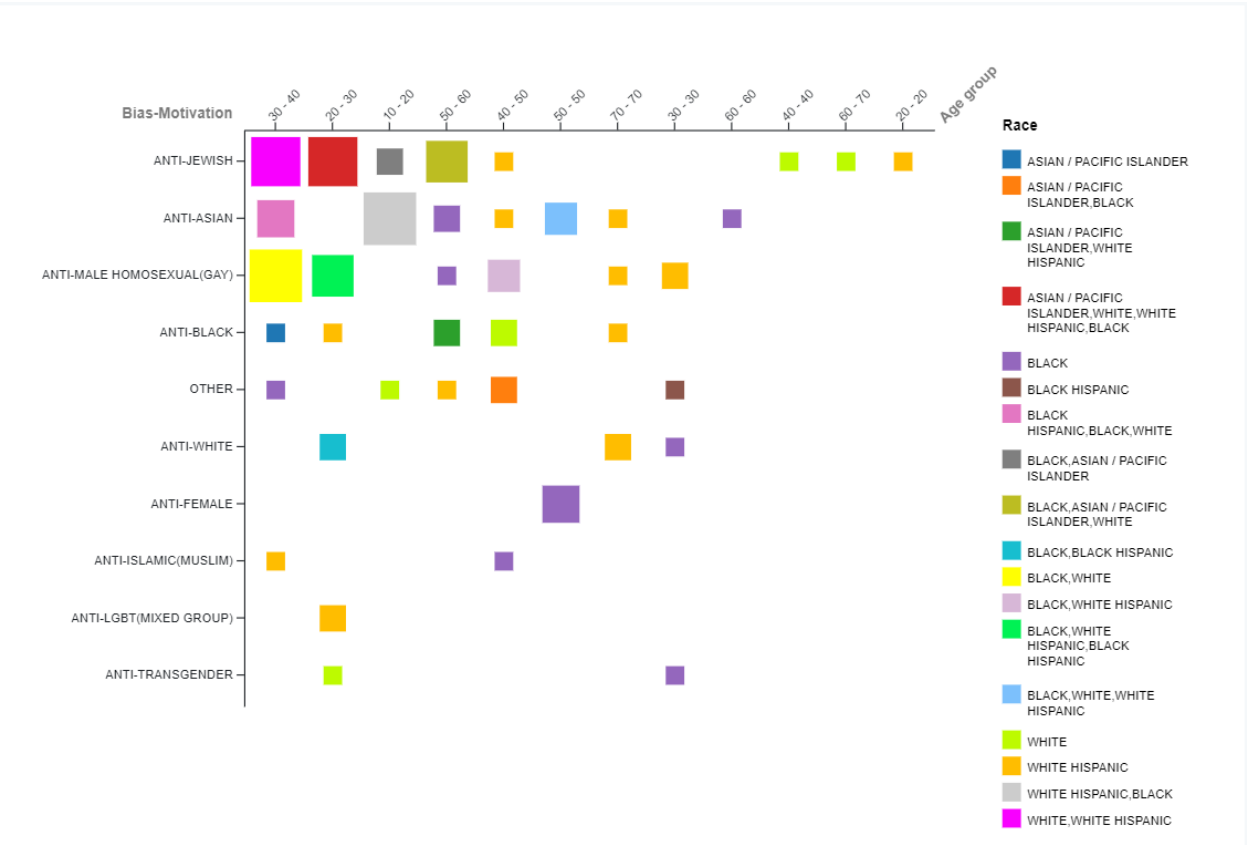
1. Steps taken

a. Selecting the Treemap step and then selecting Gender and Race under ‘Hierarchy’ and selecting Gender and Race under ‘Label’
b. Defining a new color scheme to see a distinction between Male and Females.
c. Selecting option to see hierarchy labels

2. Observations – A clean Tree Map appears with the required labels. The frequency for the boxes does not appear hence it is hard to identify how much quantity each box stands for.

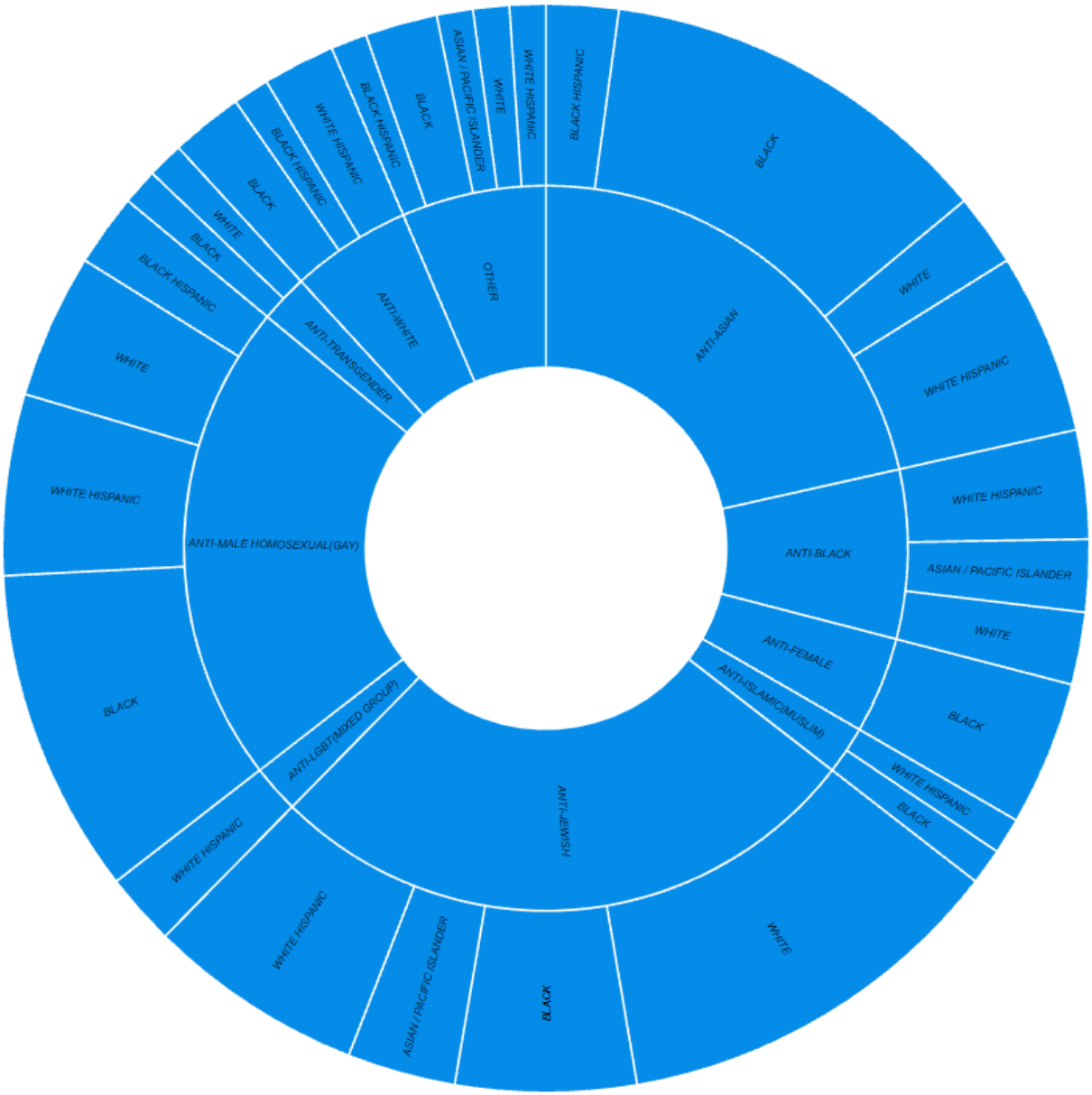
Visual 3 – Matrix Plot showing the r*elationship between age group and bias-motivation with race being visualized with colors*

1. Steps Taken
   a. Selecting Matrix Plot as the visualization and then assigning Age Group to X-axis, Bias Motivation to Y-Axis and Race to colour.
   b. Customizing and assigning all colors to Race so they are all distinct
   c. Selecting the label button to show
   d. Realigning margin to fit the figure and putting the axis in descending order
2. Observations – A clean diagram with different with sizes corresponding to frequencies appears with all the required scale and legends



Visual 4 – *Sunburst Diagram describing the relationship between Race and Bias – Motivation*

1. Steps Taken
   a. Selecting Sunburst Diagram as the Diagram to visualize and Bias-Motivation and Race as 'Hierarchy'
   b. Adjusting and customizing chart size to fit the page
2. Observations – A clean labelled sunburst diagram appears. It has numerous categories and a little difficult to understand. No option to modify colors.



Name/ Criteria	# of Visual Patterns	Does visual support data?	General Relationship	Pros	Cons
1.Sankey Diagram	3 (Lines, Colors, Labeling of Colors)	Yes	The diagram shows the relationship between Race and Bias – Motivation (2 variables only). The color helps identify this relationship further.	1. Clearly defined with distinct network of lines showing the connections between the variables 2. Intuitive	1. Not scalable to large amount of data as too many lines make it incomprehensible 2. Large data also amounts to a greater number of colors. The more colors it becomes harder to pinpoint

					what color relates to which diagram
2.Tree Map	3 (Boxes, Colors, Labels)	Yes but not enough information	The diagram shows the relationship between Race and Gender (2 variables only) with the colors drawing a distinction between the two genders.	1.Boxes are well – defined	1. Visual cannot support large data, too many boxes and colors and ultimately incomprehensible. 2. Sometimes the box size can be confusing. For example - It is hard to find which category has higher number of occurrences – Male White or Male White Hispanic 3. Does not provide enough information
3.Matrix Plot	3 (Colors, Box Sizes, Labels)	Yes	The diagram shows the relationship between age group and bias-motivation with race being visualized with colors (3 variables). On the x-axis we have the age group and the y-axis showing if the bias-motivation is present/absent and with magnitude with the square sizes. The Colors also signify the Race.	1.Easy to comprehend and intuitive since everybody is equipped with the concept and x-axis and y-axis. 2. Can graph a large amount of data by grouping into categories with three distinct variables 3 Sizes of squares are intuitive as well to understand the number of occurrences.	1.Too many categories in labels can make it a little harder to comprehend sometimes
4.Sunburst Diagram	2 (Labels, Categories)	Yes	The diagram shows the relationship between Race	1.Represents many categories in one diagram	1.Lack of distinct colors 2. Can be misleading as

			and Bias Motivation (2 variables only)		angles can be hard to understand 3. Can be crowded

Best Visualization

For this dataset the strongest competitors are the Sankey diagram and the Matrix Plot since both provide a large of information while being easy to understand and intuitive. However, the winner is the **Matrix Plot** as it visualizes 3 variables and be concise about it over Sankey plot that can only visualize 2 variables. The Matrix Plot also depicts a story with the varying colors and square sizes. It also achieves its goal in identifying distinct relationships between race, bias-motivation and age group. As a bonus, it is quite aesthetic.