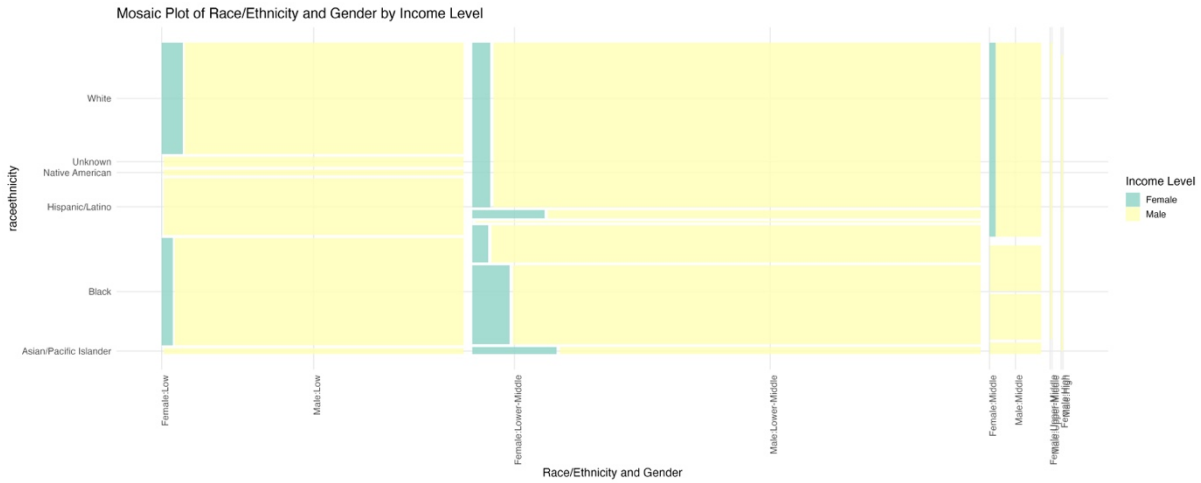
DSC 465
Homework 4

Question 1:

a.



Steps to create the visualization:

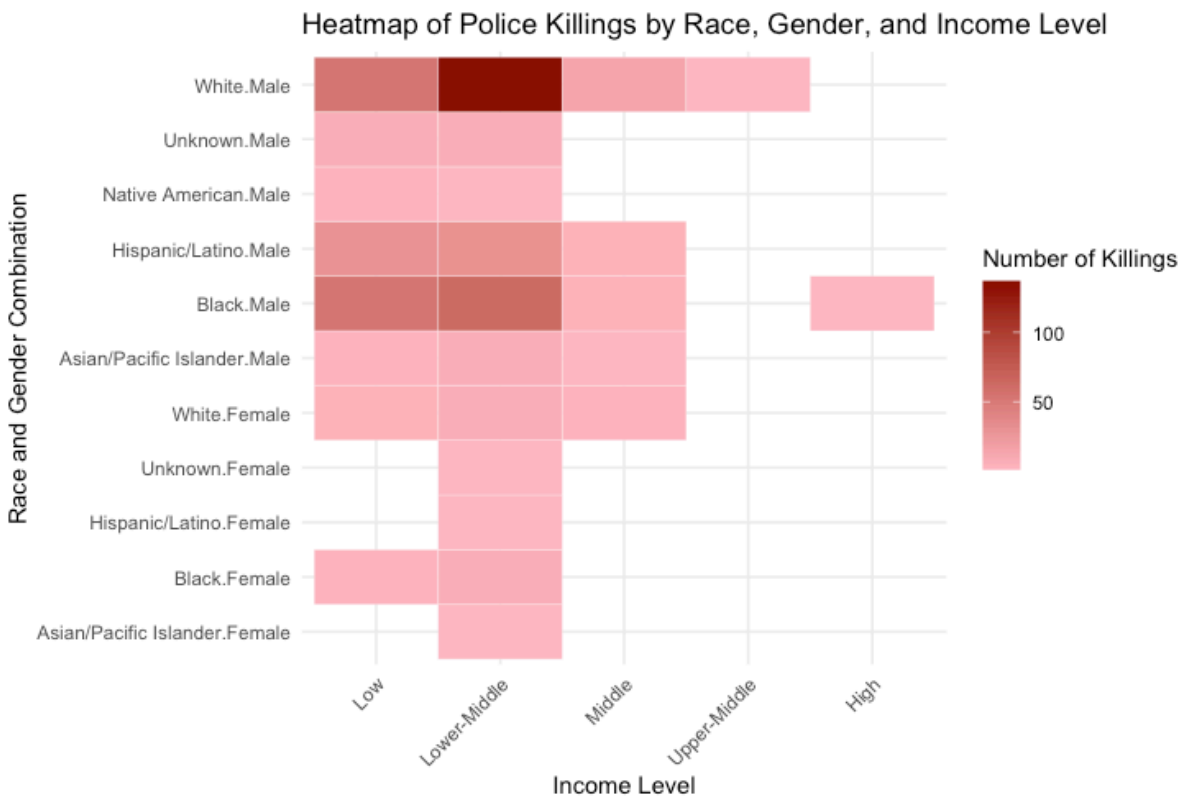
1. The police killings dataset was imported.
2. Ensure categorical variables are factors
3. Income was binned into these categories-"Low", "Lower-Middle", "Middle", "Upper-Middle" and “high”.
4. Using the Mosaic plot from ggmosaic library in R.

R code: mentioned in the RMD file attached.

Observation:

Mosaic plots are excellent for visualizing multi-dimensional categorical data, showing the relationships and proportions among categories. They allow you to see how different combinations of factors (in this case, race/ethnicity, income level, and gender) contribute to the overall distribution, making them useful for identifying patterns and imbalances.The data appears to reflect demographic categories in relation to income levels. The mosaic plot highlights how different racial/ethnic groups, income levels, and genders are distributed. By showing the compounded vulnerabilities or socioeconomic disparities among various groups, the visualization can be a powerful tool for intersectional analysis—identifying which groups are disproportionately represented in specific income brackets or by other metrics.

Question 2:



Steps followed:

1. Imported the police killings csv file to R
2. Income was binned into these categories-"Low", "Lower-Middle", "Middle", "Upper-Middle" and “high”.
3. Summarize the data- group by the race, income level and gender and count the number of kills.
4. Generate the heatmap with a light red to dark red color gradient
5. Rotate the x-axis labels for visibility

R code: mentioned in the RMD file attached.

Observation: The heatmap visualizes the relationship between race, gender, and income levels, with the y-axis representing combinations of race and gender—such as profiles of White males, Black males, and Hispanic males—while the x-axis categorizes income levels, including Low, Lower-Middle, and some Middle-income brackets. The color intensity within the heatmap indicates the number of killings, where darker shades represent higher kill counts, making high-risk profiles easy to identify. Notably, the data highlights White males, Black males, and Hispanic males within Low and Lower-Middle income categories as the most significant high-risk profiles.