Project Analysis

In recent years, we have all seen the headlines about how more African Americans are shot by police in the United States than any other race. For this project, I decided to determine if this is a valid assumption by analyzing a police shootings dataset from the Washington Post for my data analysis on this topic. I was curious about what the most prevalent race is for these shootings to see if the media is making appropriate remarks about police shootings in America. The dataset contains a variety of different elements about the suspect namely, location, race, and age, as well as some other specific elements such as being armed with a weapon, fleeing status, and if they had a mental illness or not.

For my analysis, I took the locations, longitude and latitude, and used k-means clustering to group suspects based on the distances that the shooting events occurred between one another. Once I had these clusters, I then wanted to count what race was most prevalent and what weapon they were armed with the most when they were shot by police. In my code, I made two structs for the two dimensional location of the suspect using the distance formula, as well as a struct for the weapons and race categories which will be for the counting system after the clustering. I also made a struct that will create the weighted graph for the csv file, which is also undirected as I did not care about the direction of the edges between vertices. I have the k-means clustering function that clusters each of the locations and another function called calculated_average_datapoint to create the centroids for each of these clusters. Lastly, my main function contains the csv reader, runs the weighted graph into the clustering function, and has a count system to tally the race and weapon for each cluster. The output for the analysis is the most popular race and weapon for each cluster as well as the count for each of the clusters and a horizontal bar graph for reference.

What I discovered when running the analysis is that the majority of these clusters of suspects had white people as the most popular race for police shootings, as well as being armed with a gun at the shooting event as being most prevalent. When using around 1-8 clusters, all the regions were white and armed with a gun, but when I increased k more and more, I began to see more locations with different races. The majority of clusters still tally white people as being the most prevalent, but the interesting thing is that suspects are armed with a gun in almost all of the clusters. Additionally, when there are different races other than white for some clusters, there is a close split between these clusters being African Americans and Hispanics.

In conclusion from the analysis of this dataset, the media does show a large bias towards African American shootings when the majority of police shootings are of White Americans. Also from the analysis, the majority of suspects were armed with guns when they were shot by police, which shows that there possibly a good reason as to why these police-shot these people since they could have shot the police. Even though this dataset may not tell the full story, the analysis shows that more White Americans who are armed with guns that are shot by police.