**7-2 Milestone Three: Data Structures and Brief Report**

Michael Surdek

Southern New Hampshire University

DAT-500 Data & Information Management

Dr. Arman Kanooni

August 9, 2020

**Findings:**

In the first few sections of Milestone Three, we used the twelve variables that we created in Milestone Two to create a vector and a 1-column matrix, then we combined those objects into another matrix, and finally we turned that matrix into a data frame and added a calculated column that showed the difference between our police and firefighter variables. I was able to use the calculated column in the data frame to find the differences between the variables for firefighters and police officers. For most of the variables, the differences found in the calculated column show that overtime pay and total compensation is higher for firefighters than for police officers. The difference between minimum overtime pay is $0.00, because there was both a police officer and a firefighter who earned $0 from overtime. The difference between minimum total compensation is $88.35, which was the only positive value in the column, but it is not very significant information, because the minimum earners of each occupation are both close to $0.

**Informed Conclusion:**

In Milestone One, we calculated statistics for multiple variables of police and firefighter compensation data. In Milestone Two, we learned methods of calculating these variables in RStudio. In this milestone, Milestone Three, we organized these variables in RStudio in data structures such as vectors, matrices, and data frames, and we added an additional piece of information that contained the difference between our police and firefighter variables.

In the first two milestones, when we were asked about what we discovered from the exercises, I brought up the fact that firefighters seemed to have larger means, mins, and maxes than police officers. The main finding from Milestone Three confirms my initial discoveries. In the last step, we printed df, which contained a column that calculated the difference between police officers and firefighters for each variable. In that column, the differences between means and maxes are negative values, which means that the police officer value is less than the firefighter value.

For example, the average firefighter earned $173,389 in total compensation, whereas the average police officer earned $110,966. Therefore, the difference in the third column of df is -62423.00. Adding this third column to df confirmed my initial findings that firefighters earned more overtime pay and total compensation than police officers.

**Looking Ahead:**

By using the data analysis methods that we developed in this milestone, it would be easily possible to include data from additional occupations and compare results such as overtime pay and total compensation. For example, if I had similar data on EMT compensation, I would follow the same steps to read in the data, create min, max, and mean variables, and organize them into a vector. I could then bind that vector into my compensationM matrix. Finally, I could include the EMT variables in my data frame and create calculated columns that could directly compare EMTs to firefighters or police officers.

I could even further the analysis using these constructs by adding more than the original 6 elements to each vector. I could calculate the min, max, and mean of other variables from the original data sets and use the same constructs to analyze the differences between occupations. I would only have to make small additions to my code to calculate these additional variables such as firefighter\_benefits\_min, etc. I could then add these variable names to the lines of code that create my vectors, and they would then be included in the matrix and data frame when I re-run the code. In the data frame, it would automatically calculate the difference between firefighters’ and police officers’ benefits pay.