Do Strict Open Carry Gun Laws Reduce Gun Violence

Group 12:

Eric Hardy

David Ingram

Rachel Kitabjian

Bradley Merfeld

Professor Dr. Zhang

GBA6050

Spring 2019

Table of Content

[Section 1 - Descriptive Statistics](#_Section_1:_Descriptive)…...……………………………………………………………..2

[Section 2 - Descriptive Data Mining](#_Section_2:_Descriptive)……………………………………………...……………....6

[Subsection 1: K-means clustering](#_K-Means_Clustering:)…………………………………………….…………..7

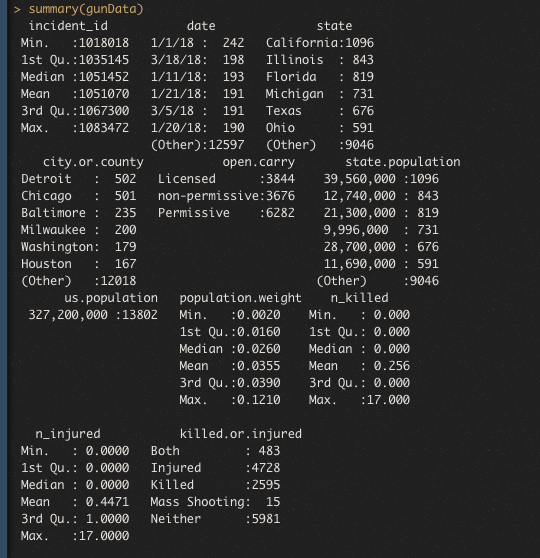
[Subsection 2: Association Rules](#_Association_Rules:)………………………………………………...………..9

[Section 3 - Updated Hypothesis, Results from Milestone 2](#_Section_3:_Revised)…..…………………………..…..…11

### Section 1: Descriptive Statistics:

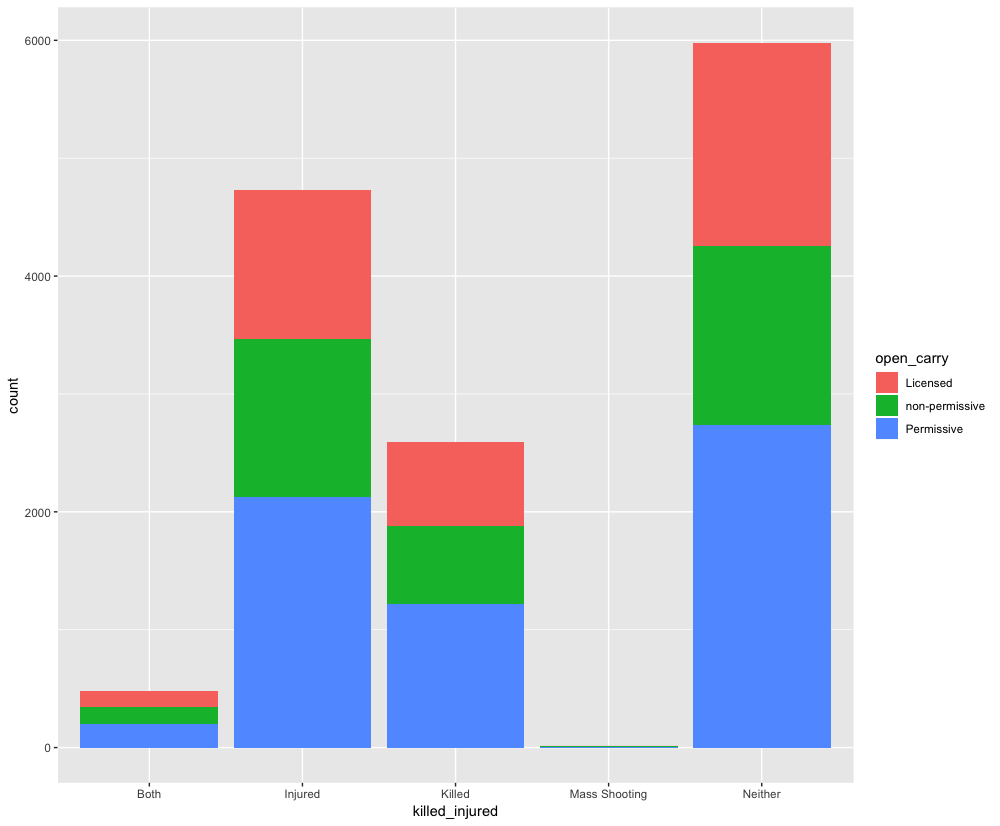
The first thing that we did was explore the data, we did this by using many different tools, mostly Excel and R. To normalize the data, we found the incidents, killed, injured, and guns per capita per state. This data is important because we wanted to see if different states have higher rates of incidents.

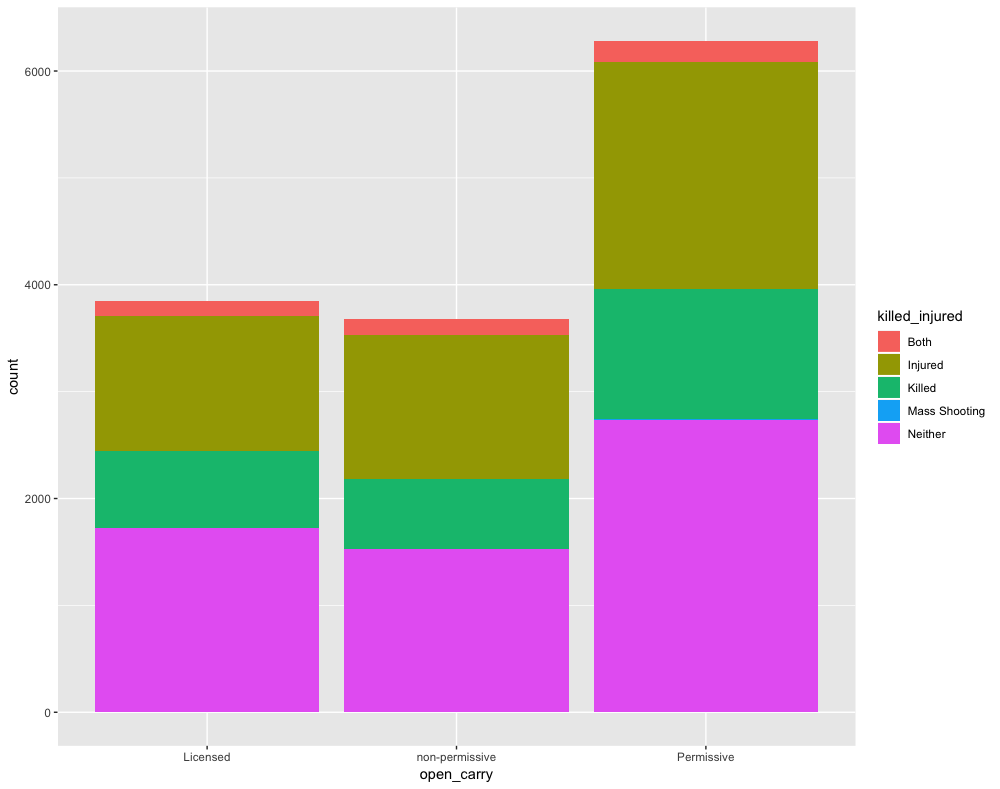
First, we pulled a summary of the data frame, in this case named, “GunData.” It is important to note that this is the raw data, not adjusted for per capita.



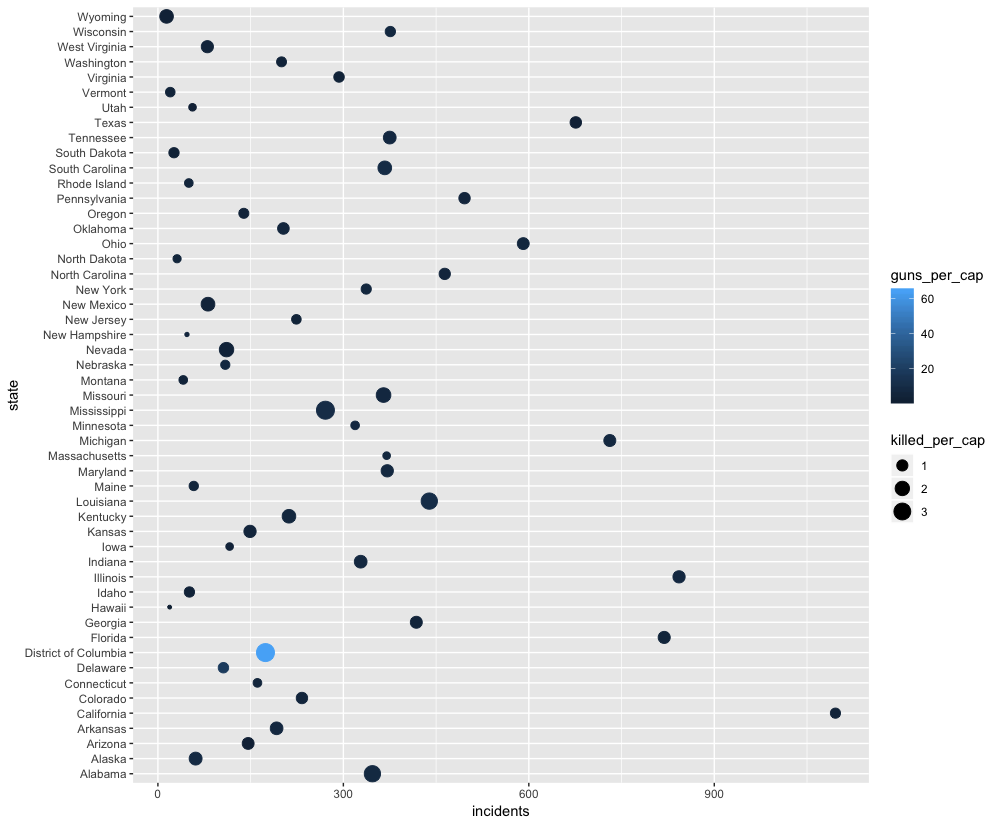
### Creating Exploratory Graphs

We then used R, specifically the ggplot2 package, which is part of the Tidyverse, to create some more graphs to explore the data, this time visually. First, we created two graphs that explored different ways at looking at open carry vs. “ amount of incidents vs. incident type.



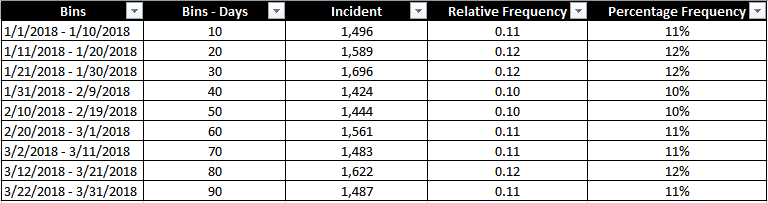


Another type of exploratory graphic that we produced was to analyze the per capita crimes. We created a point graph that shows each state, the number of incidents and how many people were killed per capita and how many guns there were per capita. A similar graph was produced for injured, which looks very similar but has bigger shapes as circle, that is omitted for brevity.



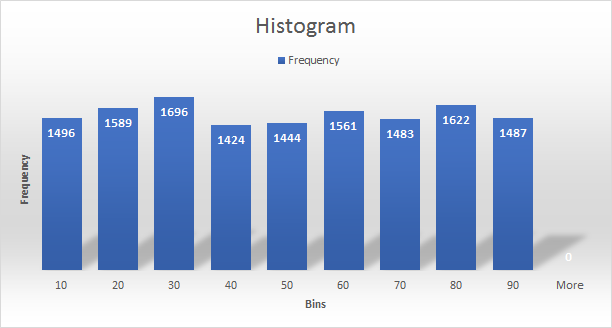
### Relative and Percentage Frequency:

Since our data set was for the first quarter of 2018 (90 days) it was decided to have each bin contain 10 days which would provide nine bins for our relative and percentage frequency analysis. After performing the relative and percentage analysis to our data set, it showed that the incident rate barely varied from bin-to-bin over the course of the 90 days. Thus, concluding that as the weather became warmer in the first quarter of 2018, it did not have an impact on the number of incidents that took place that involved a gun.



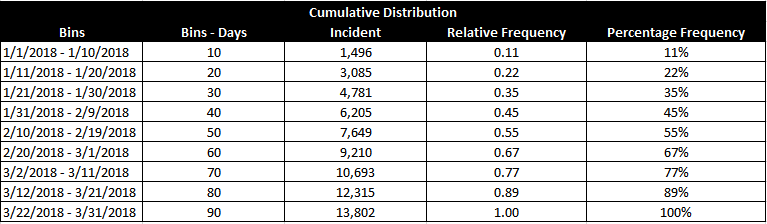
### Histogram:

The histogram below was created by utilizing the Data Analysis in Excel. The”Incident” column was used for our Input Rage and “Bins - Days” was used for the Bin Range. Thus providing the following Histogram chart. A skewness pattern could not be identified with this data set.

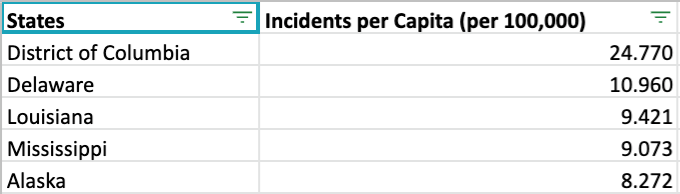


**Cumulative Frequency:**

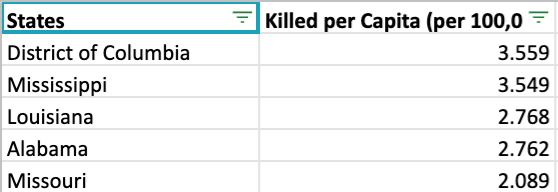
By performing the cumulative frequency analysis, it clearly supports how the weather in the first quarter of 2018 did not have an impact on the number of incidents that involved a gun.



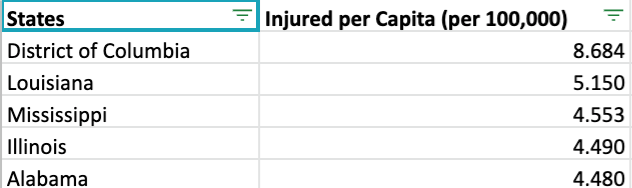
The top five states for the highest incidents per capita are as follows:



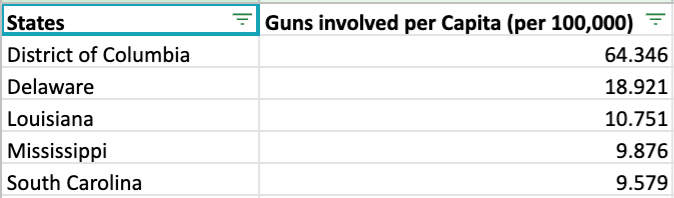
The top five states for the highest killed per capita are as follows:



The top five states for the highest injured per capita are as follows:



The top five states for the highest guns per capita are as follows:



We are going to use the data to run correlations between states that have different open carry gun laws and the incidents, killed, injured, and guns involved per capita.

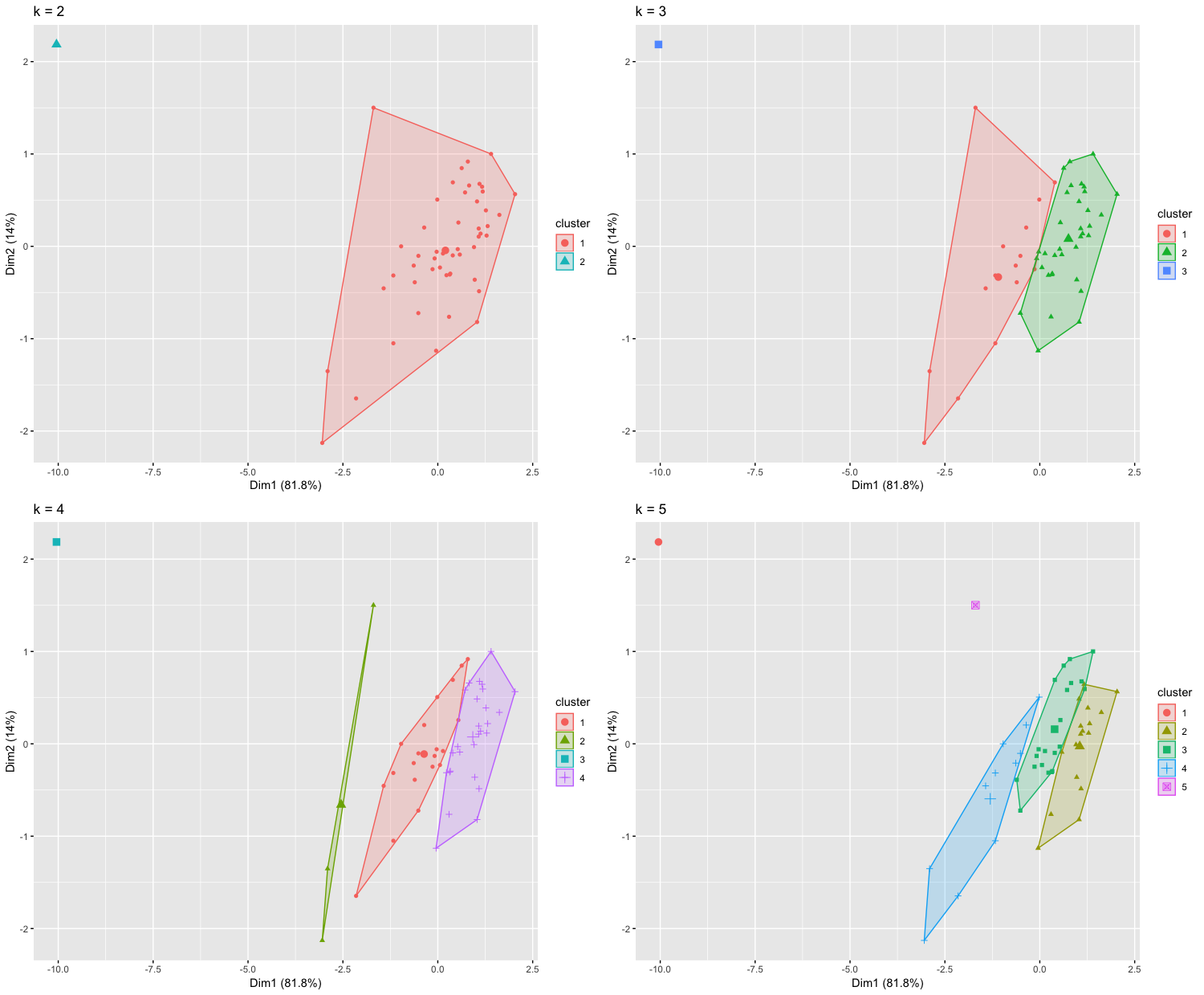
### Section 2: Descriptive Data Mining:

For data mining, we wanted to explore two types, K-means clustering, and association rules. For the K-means clustering, we made a data frame that represented the per capita crimes, this was to better understand the data as it relates to population size. The downside of doing the k-means cluster, we had to drop states, open carry, and all other categorical data. On the other hand, for association rules, we used the original data set that we created which had all of the incidents, not accounted for per capita, but still kept the open carry variable.

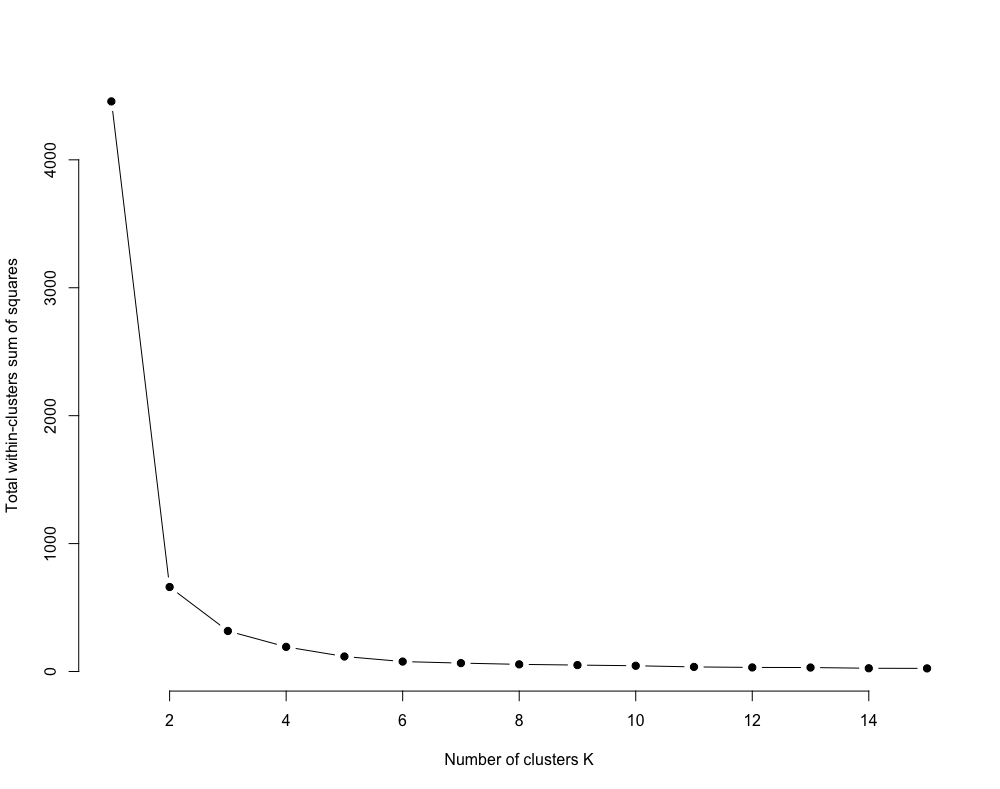
### K-Means Clustering:

While doing K-mean clustering there were some slight problems but nothing that was not able to be overcome. First, as a team of almost all Apple users, we were only able to access the cloud version of Analytics Solver, and it also did not agree with our data size. With this in mind, we moved on to other tools. After playing around with RapidMinder, Orange, BigML, and Knime, we settled on using R. The reason for this was two-fold, first, we believe that R would be useful to have a grasp on moving forward in our business careers, and two, if we get really stuck, we would be able to find some code to help us along on Github.

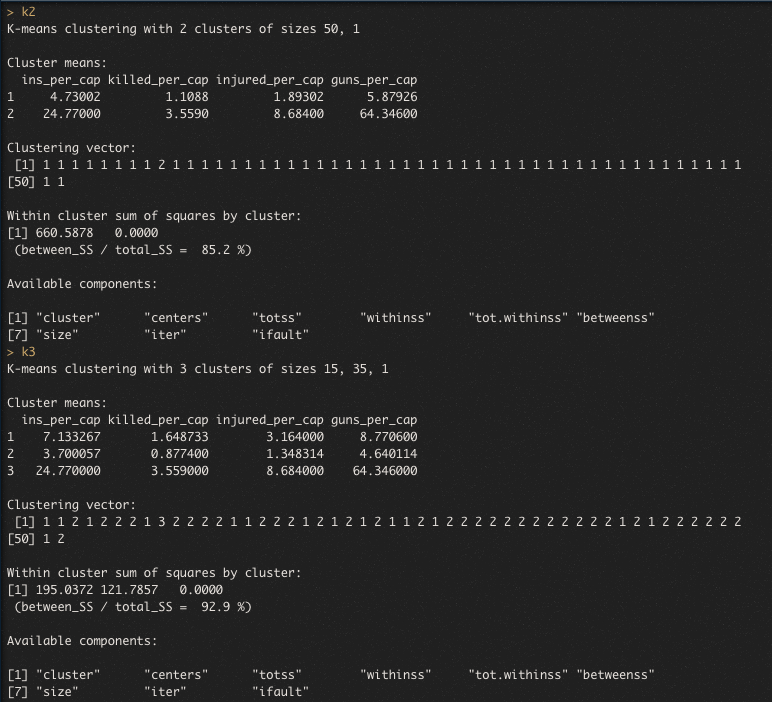
With R, we were able to complete a k-means cluster using the “kmeans” function. We completed clustering with the data set that represents the per capita incidents, killed, injured, and guns. One of the big setbacks with k-means clustering is that we had to remove the “states” column. Since running an R script allowed us to complete multiple k-means clusters, we decided to test with 2, 3, 4, and 5 centroids. After completing these clusters, they were plotted on a graph.

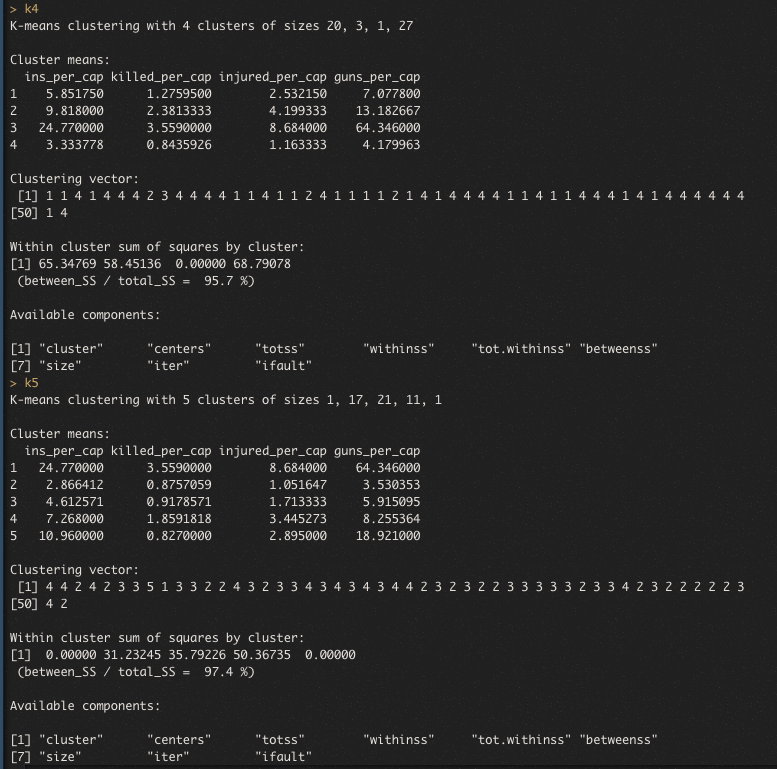


After creating the clusters visualization we created an elbow diagram to find out the optimal number of clusters, which looks to be four clusters.

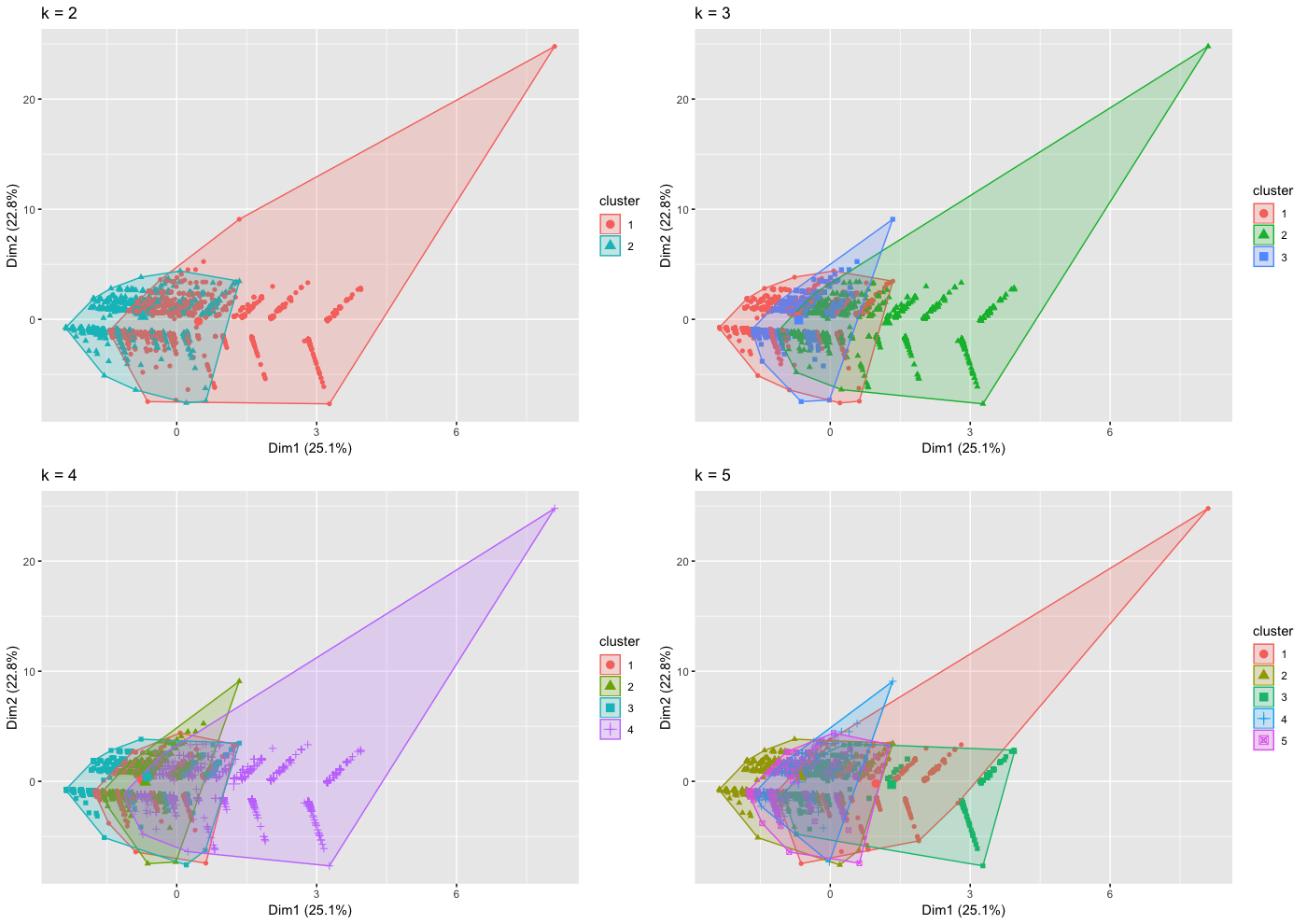


Below is a screenshot of the clusters output as taken from the console in Rstudio.



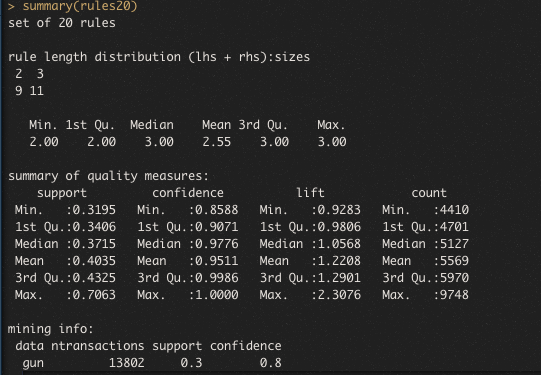


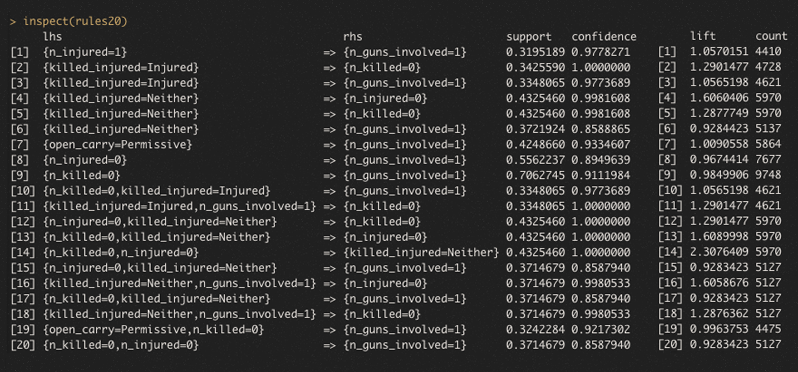
We also converted some categorical values to integers and binary values to complete a K-means cluster on data that is not adjusted to reflect per capita, ultimate though we felt that these clusters are not meaningful. Below is a screenshot of those clusters for reference.



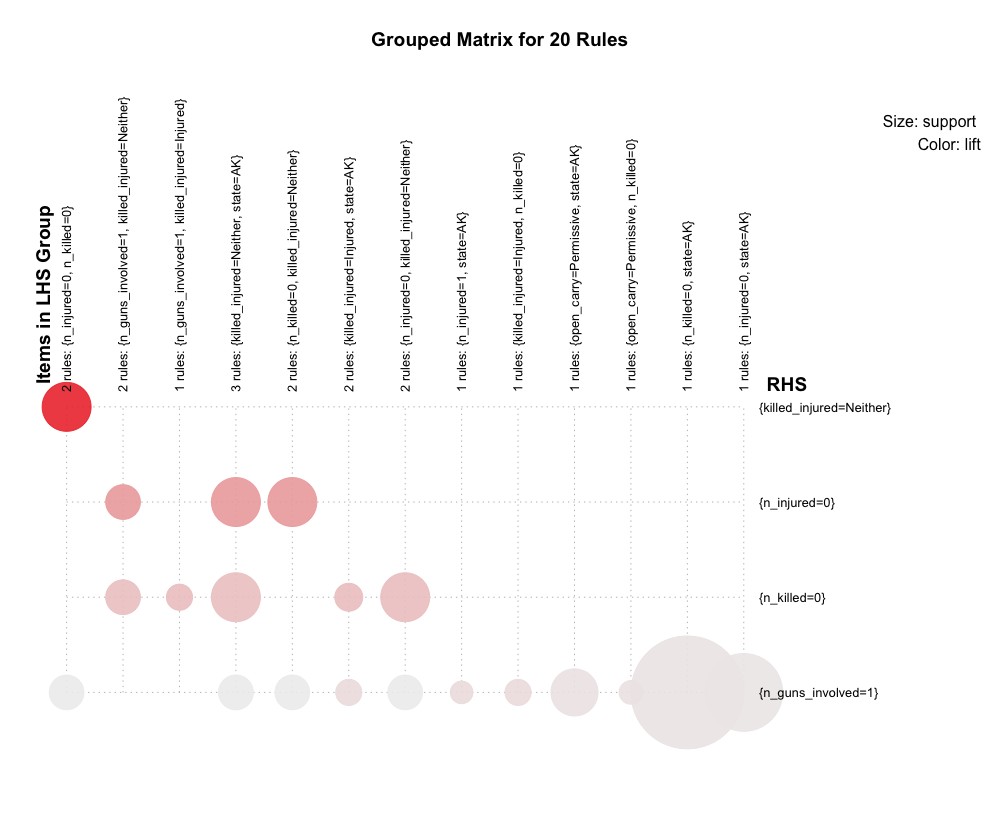
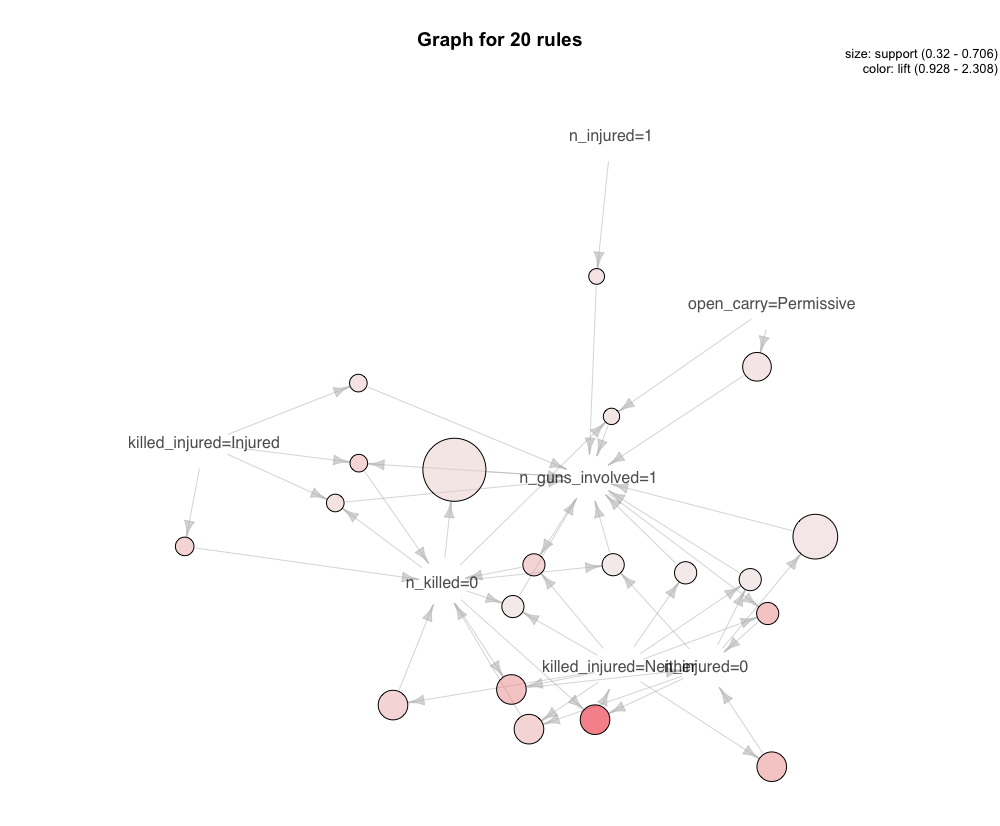
### Association Rules:

For association rules we were able to do an analysis on the original data, this included the variables like open carry, n\_killed, n\_injured, and so on. We did this in R as well using the “arules” package and an “apriori” argument. This gave us 122 rules but using “*rules20 <- apriori(gun, parameter = list(minlen = 2, maxlen = 3, supp = .3))*” we were able to reduce that to 20 rules which was easier to analyze. A screenshot of the summary of the association rules is as follows.





After this we created graphs to better analyze our results.



### Section 3: Revised Hypothesis

Hypothesis: States with non-permissive or licensed open carry gun laws will have less incidents per capita, killed per capita, injured per capita, and guns involved per capita than states without open carry laws.

### Conclusion

By utilizing Excel to normalize the data and R programming language to perform the descriptive statistic and data mining analysis we were able to provide a visualization of the data set for the first quarter of 2018.  Since our dataset was mostly comprised of categorical variables, there were some limitations to the analysis that we could perform. R allowed our team to build a script for the k-mean clustering (be converting some variables) and the association rule.  By performing the relative frequency, percentage frequency and histogram, our team was able to rule out any hypothesis that had to do with the weather. By performing these multiple tests to our dataset, it showed our team how a hypothesis could seem logical at the time but then be rejected after performing a certain analysis to the dataset and it helps determine the new null hypothesis to test and determine if it should be accepted.

From the data we decided that as of now there is no conclusive data that shows that open carry does or does not affect gun related violence. It seems as if Washington DC is the major outlier as can be seen in the per capital tables as well as the cluster plots, but most of the data otherwise is distributed in such a way that is does not seem as there is a relation between open carry laws or now.