Franklin Gaspard

4/28/2022

Crime in Boston

The analysis of crime is essential in ensuring that our society is safe. Through crime analysis we can locate areas where safety needs to be improved and this analysis can assist in developing strategies and policies which can prevent future crimes. The results from crime analysis can also be used to educate the public. People who are unfamiliar with an area are reliant on accurate crime data and analyses when they decide to visit or relocate to a new area. Additionally, crime analysis helps create accurate budgets by revealing where more resources are needed.

According to the National Criminal Justice Reference Service (NCJRS), “cities with a larger population generally have higher crime rates than suburban or rural cities”. This observation motivated me to analyze crime in an urban area/city because crime is more prevalent in cities in comparison to suburban or rural areas. I choose to analyze crime in Boston, Massachusetts because I am unfamiliar with Boston and I wanted to gain a better understanding on Boston’s crime patterns. Additionally, the data was easily accessible and one that I could trust given that the dataset was directly from the Boston Police Department. In this report I will evaluate what types of crime occur in Boston as well as when these crimes occur. I will also evaluate the overall crime rate in each census tract in Boston.

Past relevant research and analysis of crime in Boston, Massachusetts includes a crime report created by the Crime Analysis Unit at the Cambridge Police Department. Although this analysis focused on crime in Cambridge instead of Boston, this analysis is relevant because according to CityTownInfo, “The city of Cambridge is part of the Greater Boston area of Massachusetts”. The crime report from the Cambridge Police Department analyzed the crime in Cambridge from the first three months of 2022. Throughout the report the results of the analysis were often compared to the results of crime analysis in previous years.

The Crime Analysis Unit at the Cambridge Police Department first analyzed serious (part one) crimes which include murder, burglary, rape, larceny, robbery, and aggravated assault. They further divided these crimes into violent crime and property crime. Next, they compared the crime analysis from the first quarter of 2022 to the crime analysis from the first quarter of 2021. They also computed a five-year average change for these crimes. It was found that part one crime had decreased by 8% in comparison to 2021. However, violent crime had stayed the same and the 8% decrease in part 1 crime is a result of a 9% decrease in total property crime. Additionally, there was a 3% increase in the five-year average crime index. Lastly, the report created a graph for murder, rape, aggravated assault, robbery, burglary, larceny, and auto theft with the intention of mapping out the patterns of these crimes over five years. Larceny from motor vehicle had a sporadic pattern but no patterns were found for the rest of the crimes.

For the analysis of crime in Boston, I used data from a crime incident report found on data.boston.gov. The data from the crime incident dataset is provided by the Boston Police Department. This analysis will be of crime in the year 2018. The crime incident dataset has 98,888 observations and 17 variables. The variables describe the crime and state when and where the crime was committed. Only 3 of the 17 variables have any missing data. REPORTING\_AREA had missing data in 6746 observations however this was not a problem because REPORTING\_AREA was not needed to find the location of the crime so I did not use this variable in my analysis. Lat and Long both had missing data in 6429 observations. These two variables had missing data in all the same rows so I decided to remove the rows with missing data. Although removing these rows may slightly affect the analysis, given that the dataset is large, excluding these rows shouldn’t drastically affect the quality of the analysis.

In addition to the crime incident report, I used census data from 2010. I obtained this data through the Census Bureau’s TIGER/Line program. I used the census tract data for Suffolk County, Massachusetts and this is the county which Boston is located in. There are 204 observations in the census tract shapefile and 13 variables. This data contains the geographical information on each census tract in Suffolk County. There is no missing data in the dataset and I am confident that the data is accurate because the data is from the United States Census Burau.

In my final project proposal, I stated that I would be using the 2020 version of the crime incident report as well as 2020 census data but I changed it to the 2018 version of the crime incident report and 2010 census data. This due to the missing data in the 2020 crime incident report. The dataset had missing data in every observation in the OFFENSE\_CODE\_GROUP and UCR\_PART variables. This missing data would make it extremely difficult to group the crimes based on severity given that these variables state what category each crime falls under. I did not realize that would be an issue when I wrote my final project proposal.

To analyze this dataset, I first read both the crime incident report and the census data into R. Next, I removed the missing observations from the Long and Lat variables after checking that the same observations were missing for both variables. I then used the st\_as\_sf function to change the crime data into a shapefile because this would allow me to then join the crime data with the census tract data. Before I could join the data I also used the st\_transform function in order to change the projection of the census tract to match with the crime data. After I joined the data I had 92122 observations and 28 variables.

Next, I created individual tables for part one, two, and three crimes with the filter function. 17204 observations were part 1 crimes, 27371 observations were part 2 crimes, and 47120 observations were part 3 crimes. Therefore, 427 crimes didn’t have a category. Since this is not a significant portion of the data I did not analyze these crimes separately. I thought it was important to split the crime data into these three categories because the severity of crimes varies drastically and its beneficial to analyze each category of crimes separately. Part 1 crimes include homicide, burglary, larceny, robbery, larceny, auto theft and aggravated assault. Part 2 crimes are less serious and include counterfeiting, harassment, drug violations, firearm violations, fraud, vandalism and other non-violent crimes. Part 3 crimes include medical assistance, towing, warrants, verbal disputes as well investigating people and properties.

Now that the data is separated, I then created a table for each part using the OFFENSE\_CODE\_GROUP. These tables stated how much of each crime had occurred. However, when I attempted to graph the results from these tables I received the following error; “Error: `data` must be a data frame, or other object coercible by `fortify()`, not an S3 object with class table”. I found the solution to this problem on Stack Overflow. To solve the error I used the as.data.frame function to convert the data into a data frame. I then created bar graphs in order to visualize the occurrence of each type of crime in each part.

Next, I created a table for each part which included what day of the week each crime occurred. I created a stacked bar graph for part 1 to visualize the occurrence of crime on each day of the week. I only created this graph for part 1 crimes because after I created this graph I realized that it was difficult to see which day of the week had the most crime on the graph. So instead I created tables using the group\_by and summarize functions in order to calculate the occurrence of crime in each day of the week. Furthermore, I repeated this step in order to calculate the occurrence of crime in each month out of the year.

After I created these tables, I retrieved information from the 2018 5-year American Community Survey. First, I used the load\_variables function to create a data frame for all the variables in the dataset. This data frame had 26997 observations and 3 variables. Next, I wrote the data frame as a csv file using the write\_csv function so that I could open the data in Excel. I then used the get\_acs function to retrieve the specific information that I wanted from the 2018 5-year American Community Survey which included population information from the Suffolk County, Massachusetts census tracts during the year 2018.

Next, I created another data frame by using the group\_by and summarize functions which listed each Suffolk County census tract and its population. I then used the left\_join function to join the data frame that I just made with the census data. I joined this data in order to have access to both a census tract’s population and GEOID. Without joining the data I would not have been able to access both variables. Before I could use the GEOID for my next calculation I had to use the rename function to rename GEOID to GEOID10 so that it would match with my original dataset. Next, I used another left\_join to combine all the data I have gathered including the crime data and census data.

In the last few steps of my R code, I used a group\_by and summarize function to calculate how many crimes occurred in each census tract. After joining this data frame with another data frame, I was able to calculate the crime rate in each census tract. I divided the total crime in each census tract by the total population in each census tract and multiplied that number by 100.

I believe that the results of my analysis should affect policy and decision-making. One of my first outcomes of my analysis is finding out that in 2018, 17204 crimes were part 1 crimes, 27371 crimes were part 2, and 47120 crimes were part 3. This information can help the Boston Police Department prepare for upcoming years because they can estimate how much of each type of crime they can expect in future years. However, my results are inconclusive because the crime data from data.boston.gov during the years 2019-2021 are completely missing the OFFENSE\_CODE\_GROUP and UCR\_PART variables. This makes it very difficult to categorize the crimes and without this data we can’t see the present trend of part 1, 2, and 3 crime in Boston. This will be the biggest limitation of all the results of my analysis.

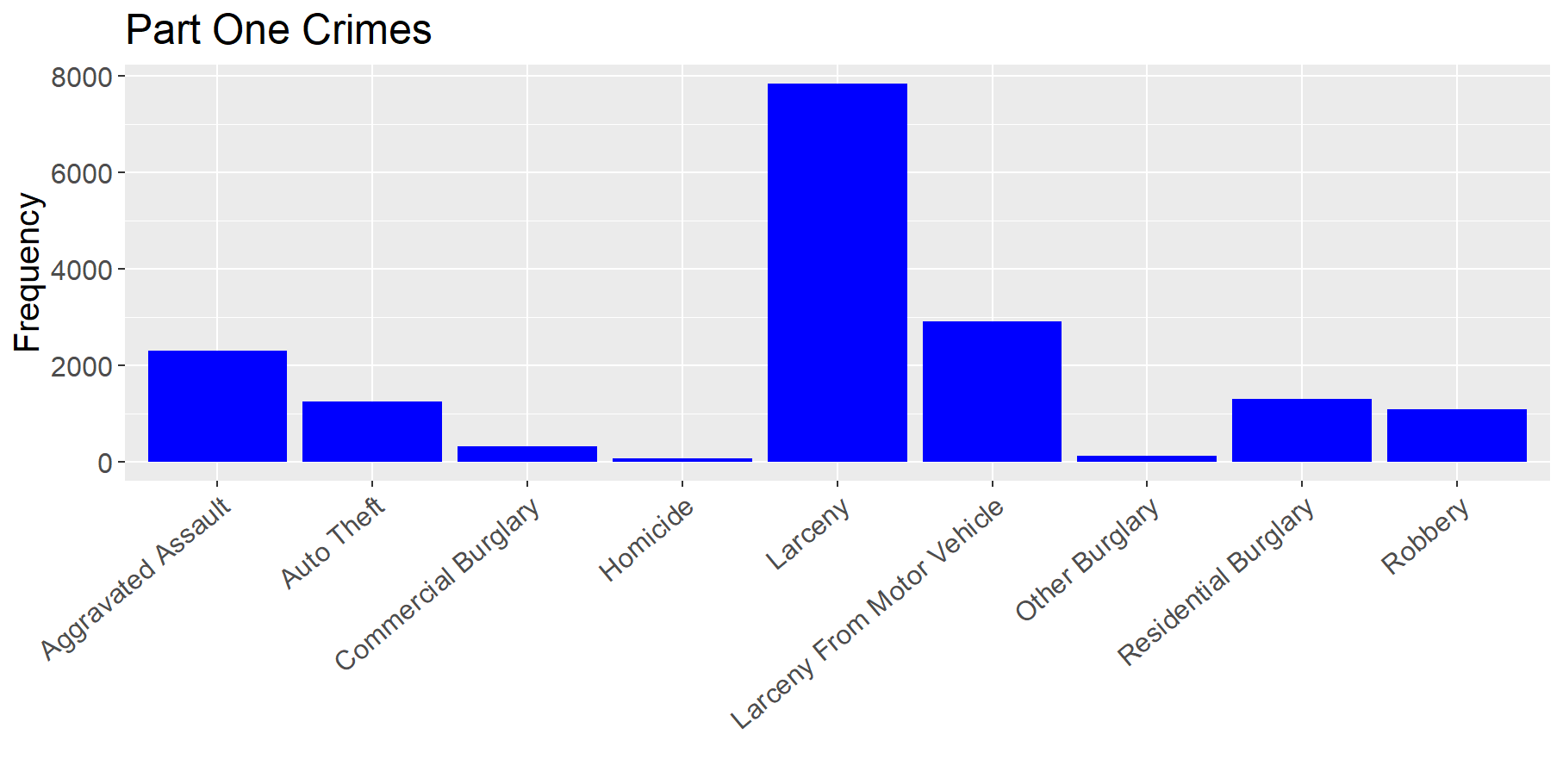
Next, I found that the part 1 crime that happens most often by far is larceny as seen in figure 1. I believe that this is valuable information to the public. According to United Federation of Teachers, larceny can be prevented by carrying only what you need, carrying small amounts of cash, not leaving valuable items visible in a vehicle, and by watching for distractions. As seen in figure 2, simple assaults, vandalism, drug violations, and other are the most common part 2 crimes. It is concerning that such a high number of part 2 crimes were labeled other, however the analysis can still be used to affect policy and decision-making. Boston can look into having better lighting around the city because this would help with vandalism and assaults. Additionally, with such high drug violations maybe Boston should look into educating the public on the negative effects of drugs as well as fund drug treatment centers. As seen in figure 3 motor vehicle accidents responses and medical assistance are the most common part 3 crimes. This result is inconclusive. I would encourage others to further analyze traffic patterns and accidents to see if Boston’s infrastructure needs to be safer. Additionally, I would encourage others to evaluate how proficient medical assistance is and if more funding needs to go to hospitals and other treatment facilities especially those who specialize in emergencies.

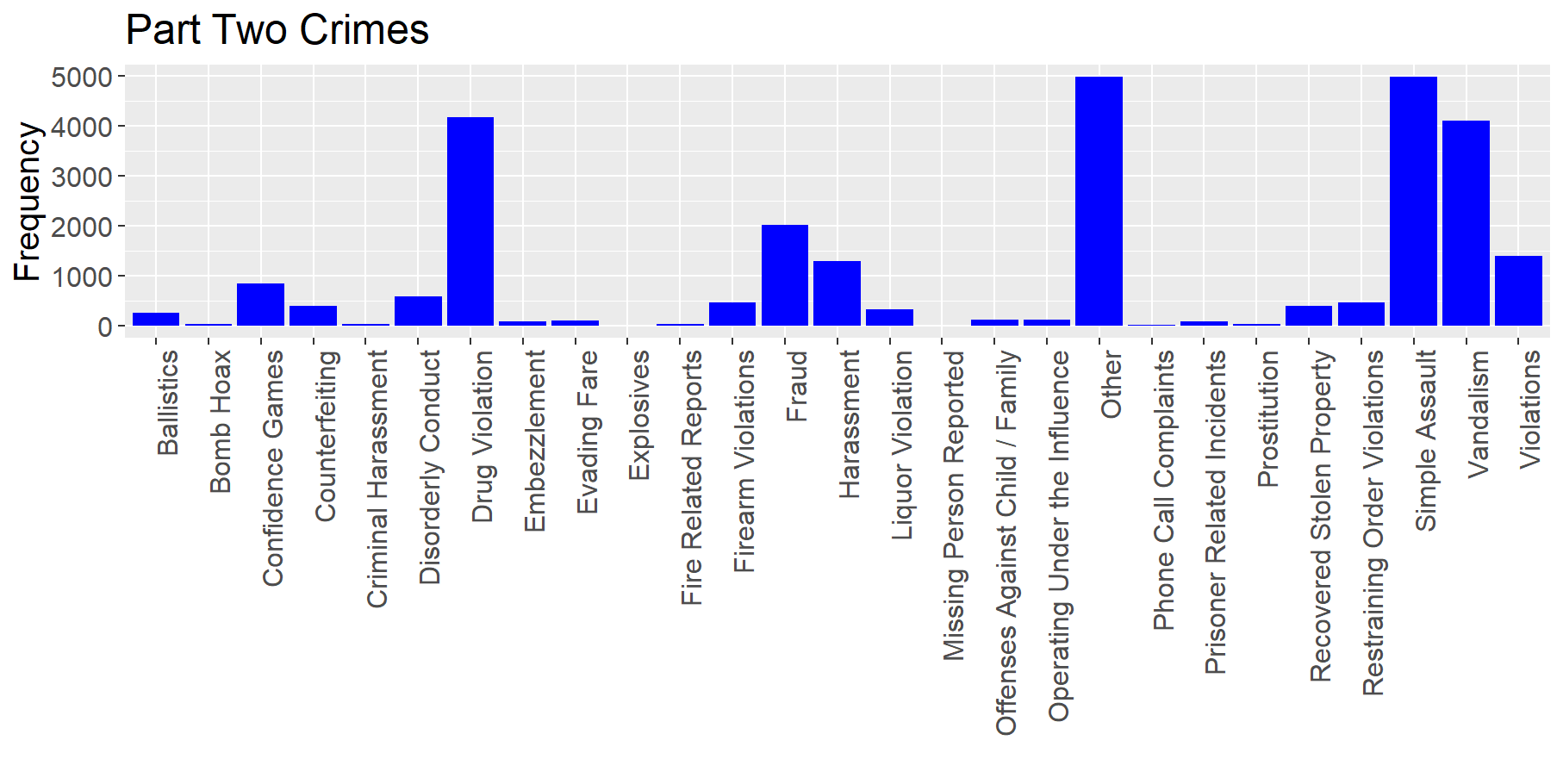
Another set of results from my analysis which can affect policy and decision-making are the results on how often crime occurs on each day of the week. Figure 4 shows that part 1 crime is pretty consistent throughout the week with a slight increase during Fridays and Saturdays. Figure 5 shows that part 2 crime is more common on Wednesdays, Thursdays, and Fridays and the least common on Sundays. Overall part 2 crime is pretty consistent Monday through Friday. Figure 6 shows that part 3 crime is pretty consistent Monday through Saturday and it the least common on Sundays. This information can help the Boston Police Department determine how many police officers should be working each day of the week.

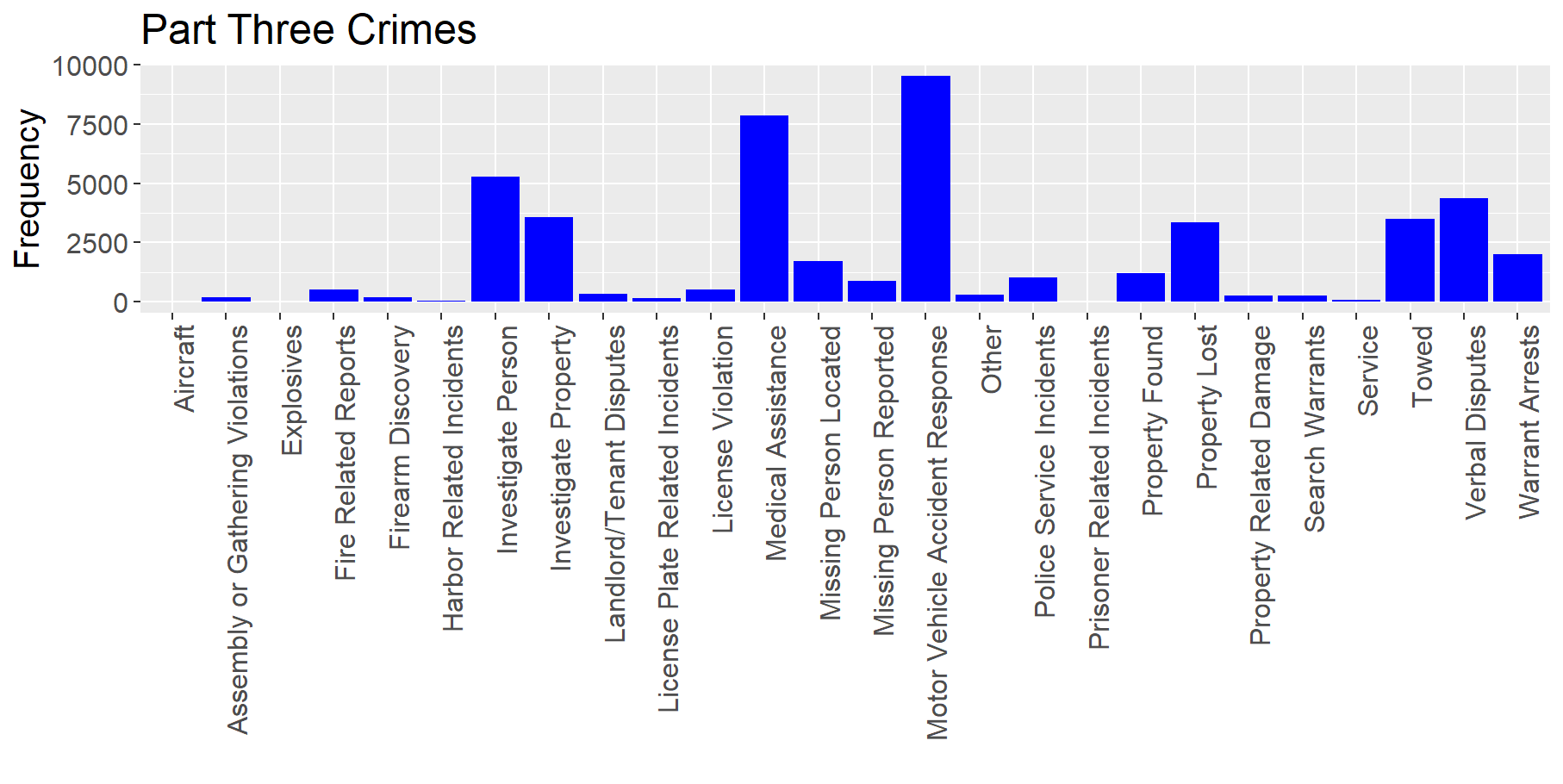
Next, I analyzed how often crime occurs each month. As seen in figure 6, 7 and 8 crime in part 1, 2 and 3 follow a similar pattern. They tend to be consistent throughout the year but they peak during the summer months. The slight increase in crime during the summer months can be a result of warmer weather. During the warmer months people are more likely to go out and participate in various activities making them more likely to experience or commit a crime. I think this information could also help the Boston Police Department determine how many police officers should be working each month.

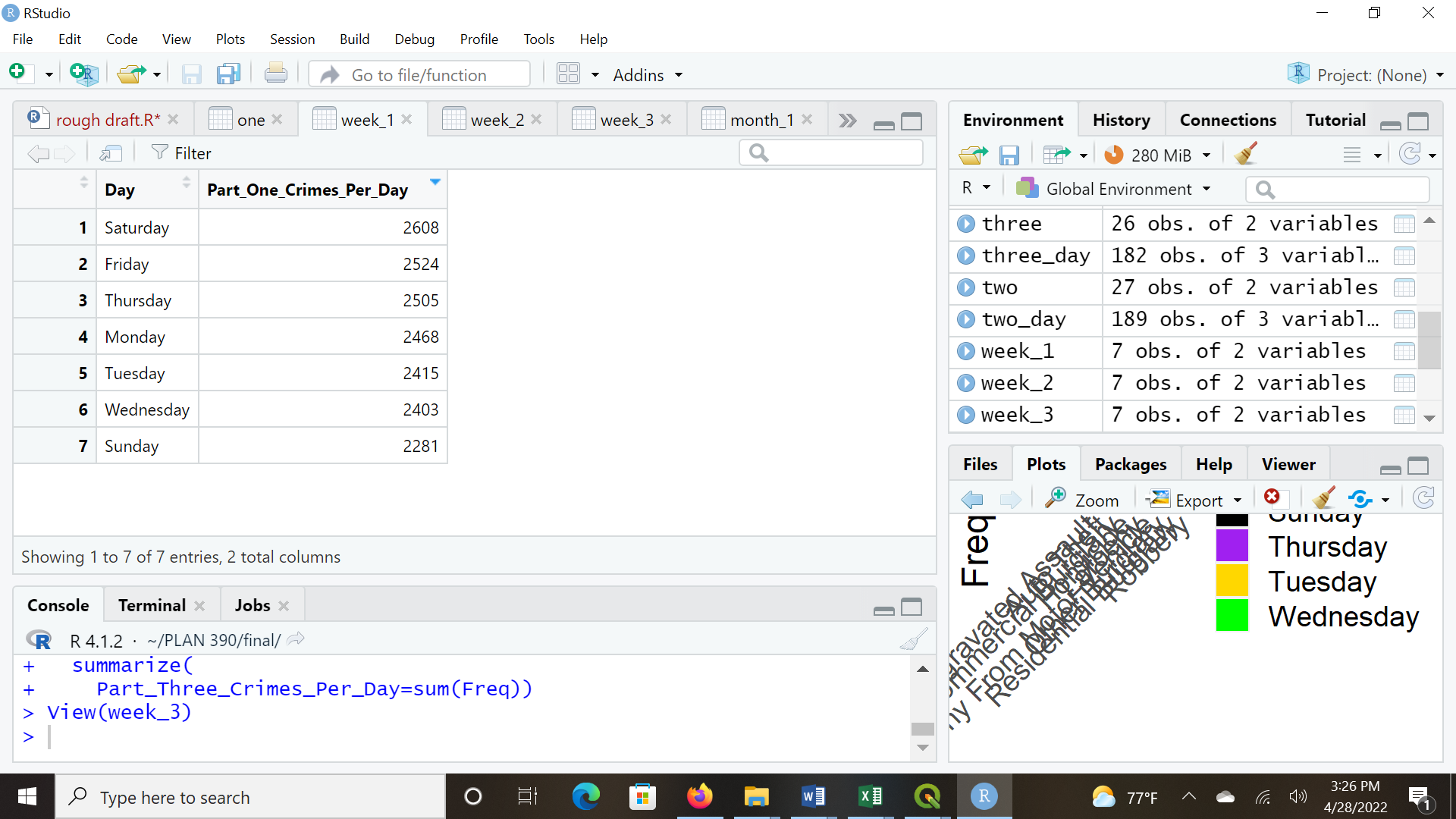
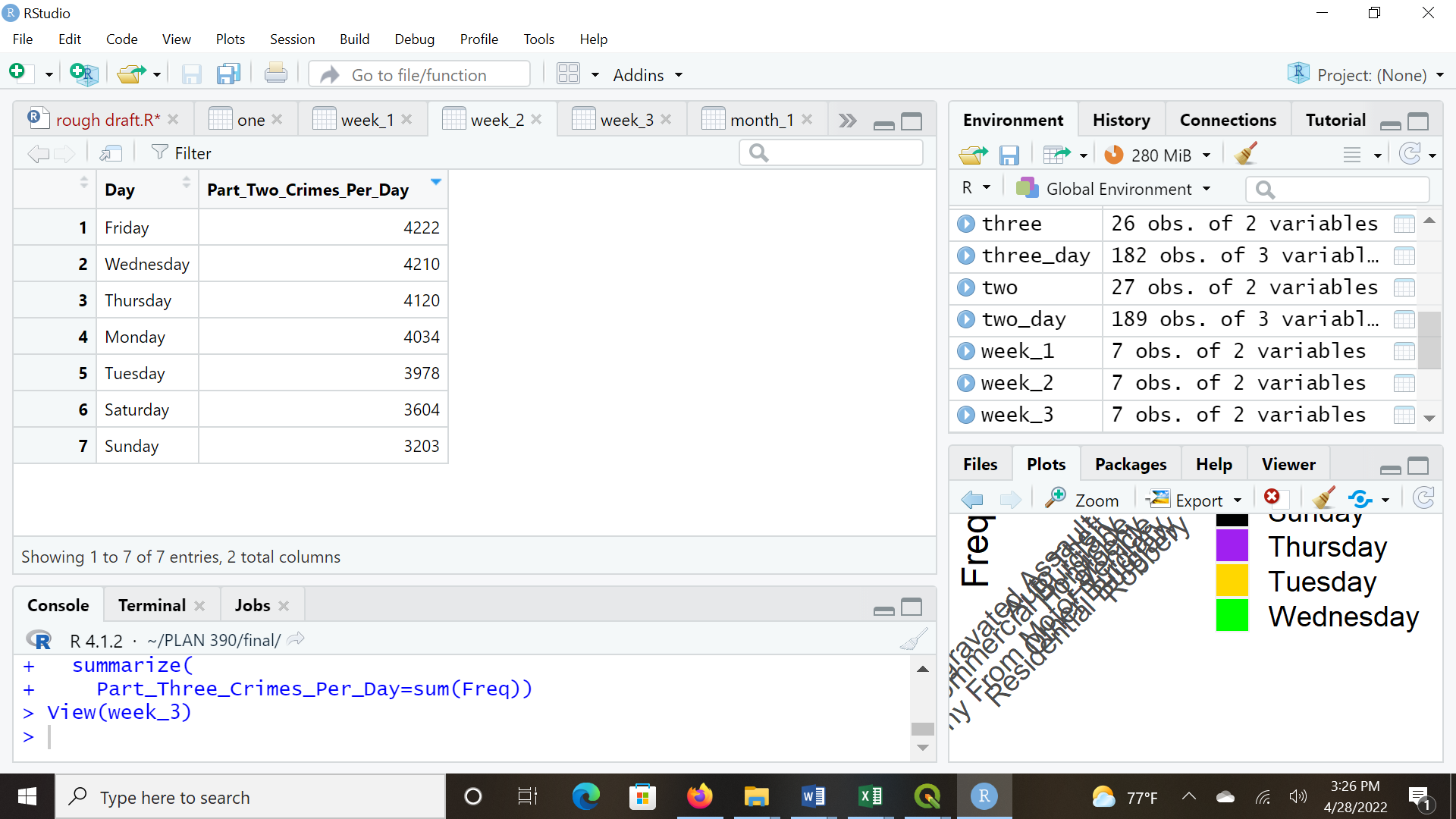
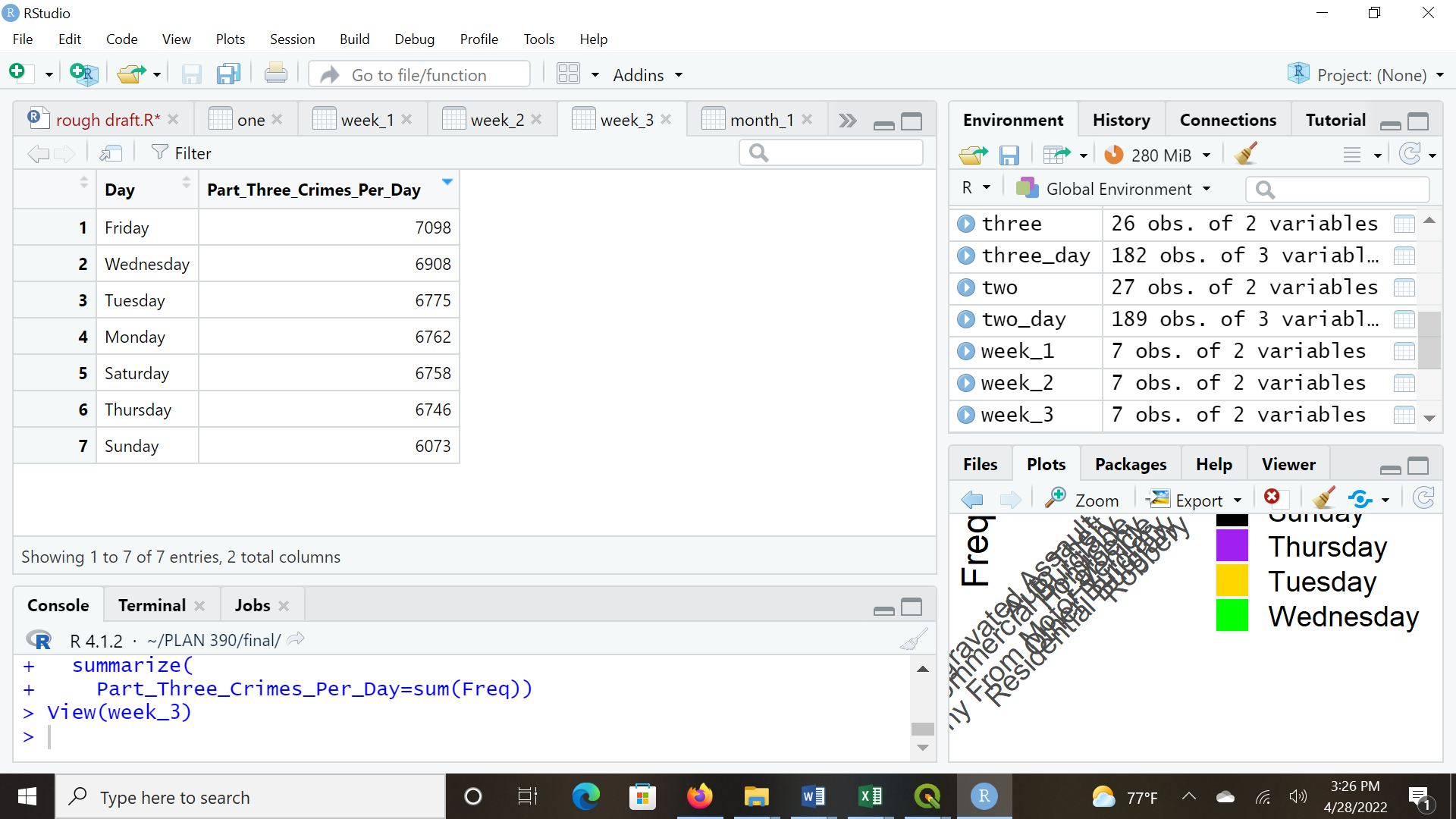
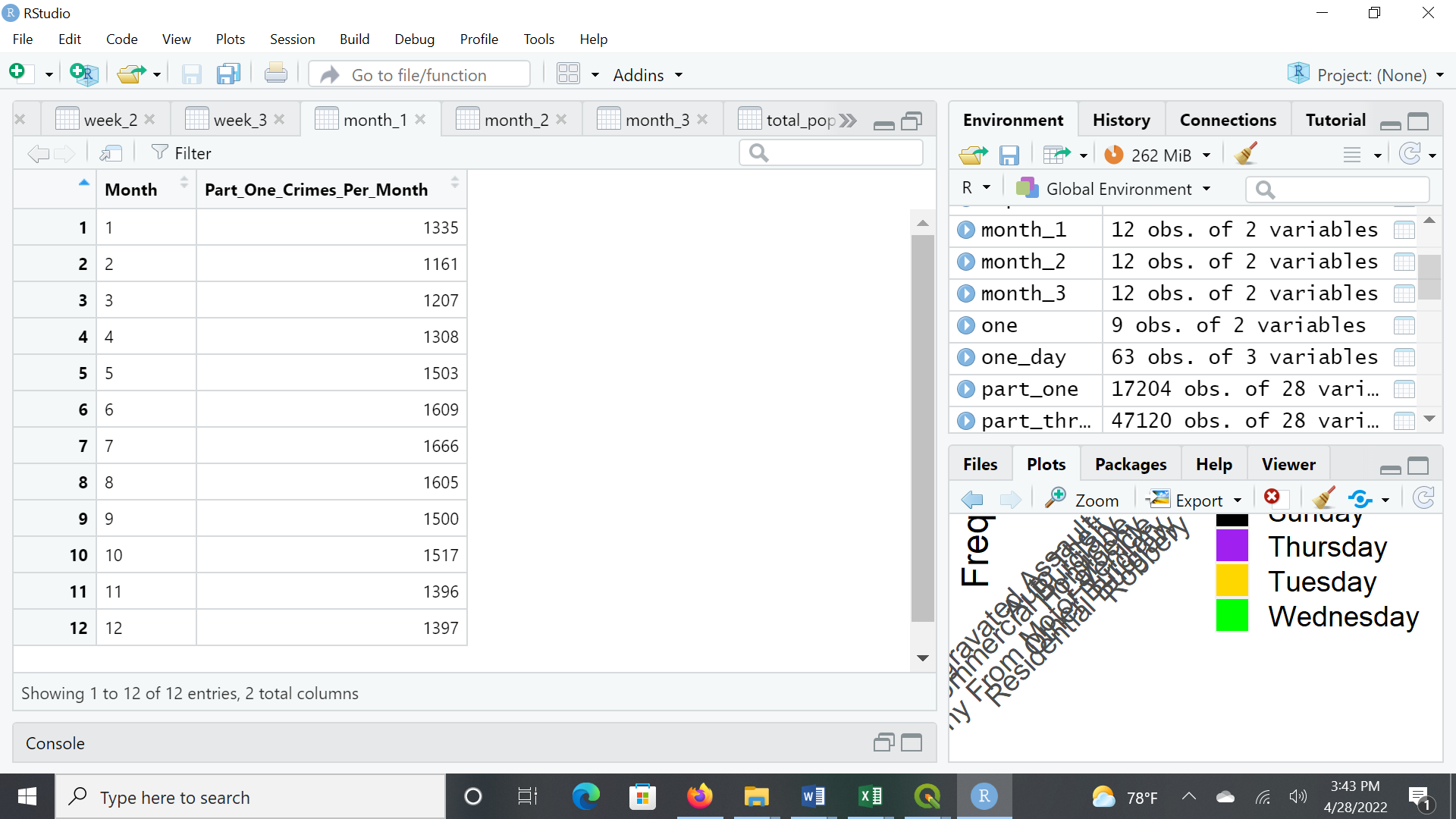
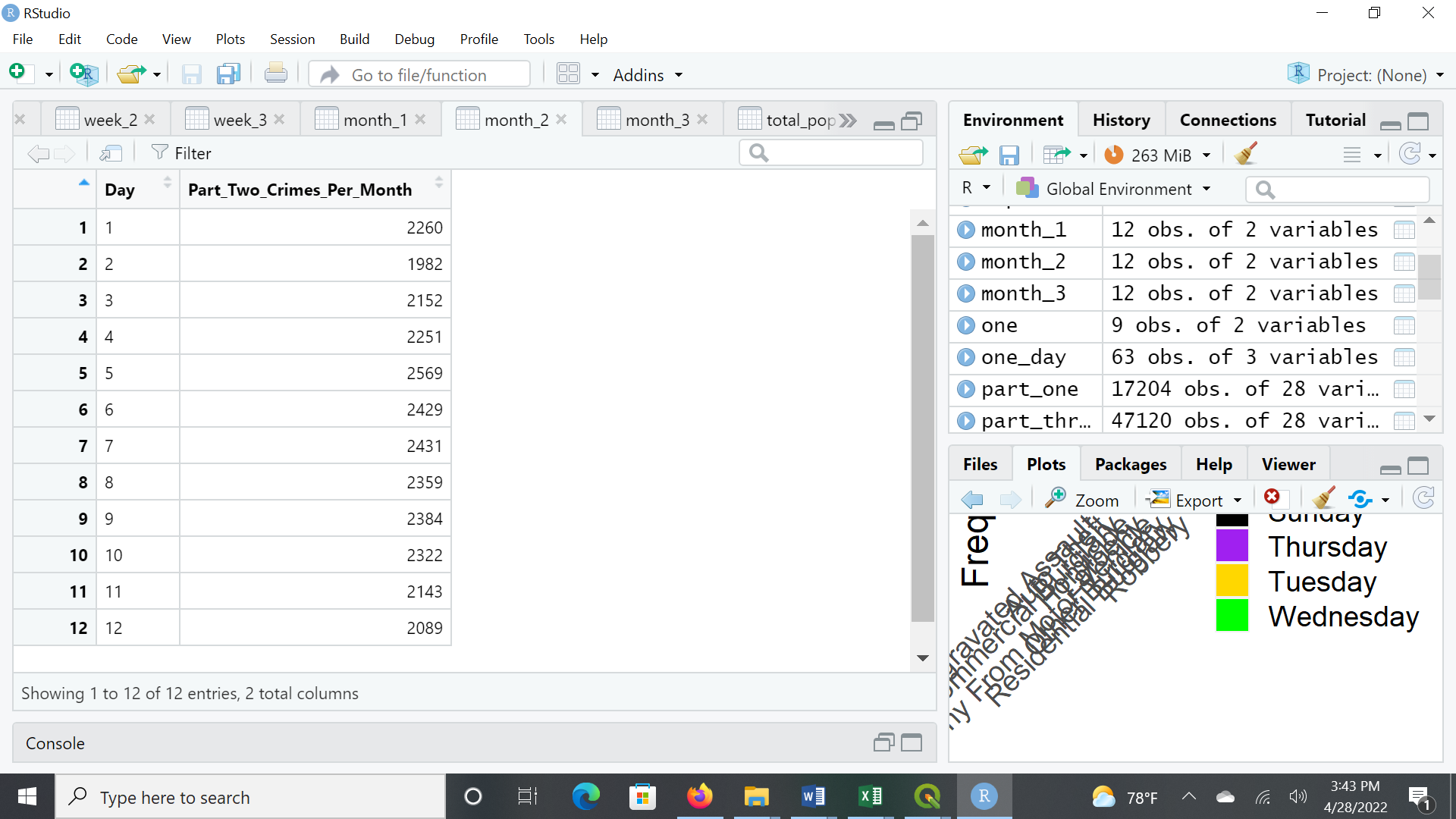
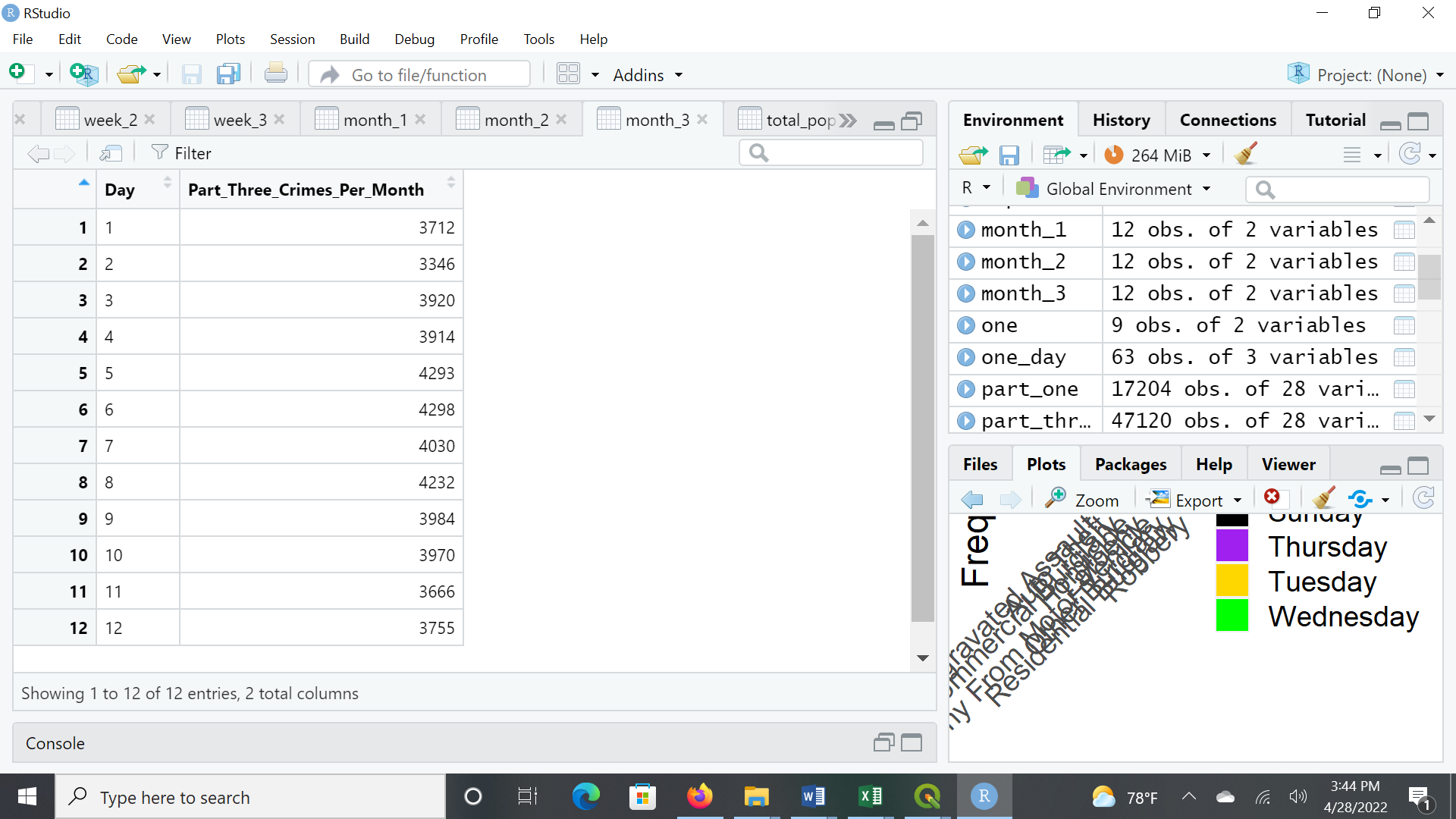
Lastly, I calculated the overall crime rate for each census tract. This information is extremely important for a multitude of reasons. Knowing where crime rates are high can allow citizens to make informed decisions when navigating in different parts of Boston. The crime rates can also help show where more resources are needed. Increased police force, improved infrastructure, improved lighting, more surveillance cameras, and services which help former prisoners find a job and stable housing are a few ways Boston can combat crime in census tracts with high crime rates. (Rao, Shebani & La Vigne N)

In conclusion the analysis of crime not only can be used to evaluate the past but it can positively shape our future. By evaluating past crime patterns, we can create and adjust policies to improve safety in the United States of America. The results from crime analysis also allows us to make informed decisions as we navigate this country and the world.

Figure 1

figure 2

figure 3

 figure 4 figure 5figure 6  Figure 7  Figure 8 Figure 9

Works Cited

Cambridge Police Department. “Cambridge Police Department BridgeStat” *CambridgeMA* https://www.cambridgema.gov/~/media/Files/policedepartment/BridgeStat/BridgeStat\_March2022\_FINAL.ashx?la=en

CityTownInfo. “Cambridge, Massachusetts” *CityTownInfo* https://www.citytowninfo.com/places/massachusetts/cambridge

National Criminal Justice Reference Service. “Urban and Rural Victimization” *NCJRS* https://www.ncjrs.gov/ovc\_archives/ncvrw/2017/images/en\_artwork/Fact\_Sheets/2017NCVRW\_UrbanRural\_508.pdf

Rao, Shebani & La Vigne N. “Five Ways to Reduce Crime” *Urban Institute* 7 May 2013. https://www.urban.org/urban-wire/five-ways-reduce-crime

Stack Overflow. “How to convert a table to a data frame” https://stackoverflow.com/questions/10758961/how-to-convert-a-table-to-a-data-frame

United Federation of Teachers. “Larceny & Theft Prevention Tips” *UFT* https://www.uft.org/your-rights/safety-health/school-safety/larceny-theft-prevention-tips

United States Census Burau. “2010 TIGER/Line® Shapefiles: Census Tracts” https://www.census.gov/cgi-bin/geo/shapefiles/index.php?year=2010&layergroup= Census+Tracts