**Data Story Telling Competition – 31st Jan 2020**

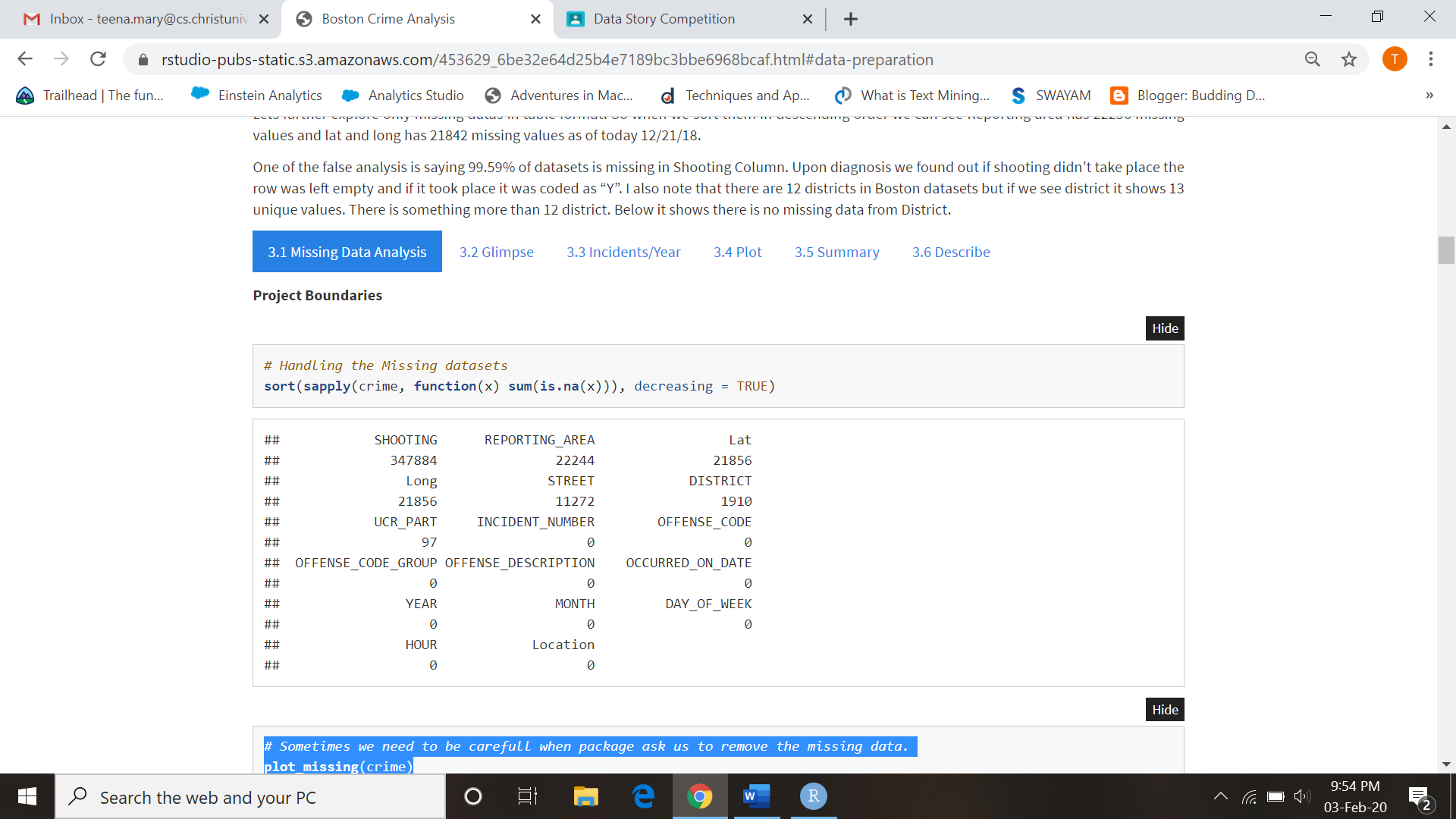
**Teena Mary 1948058**

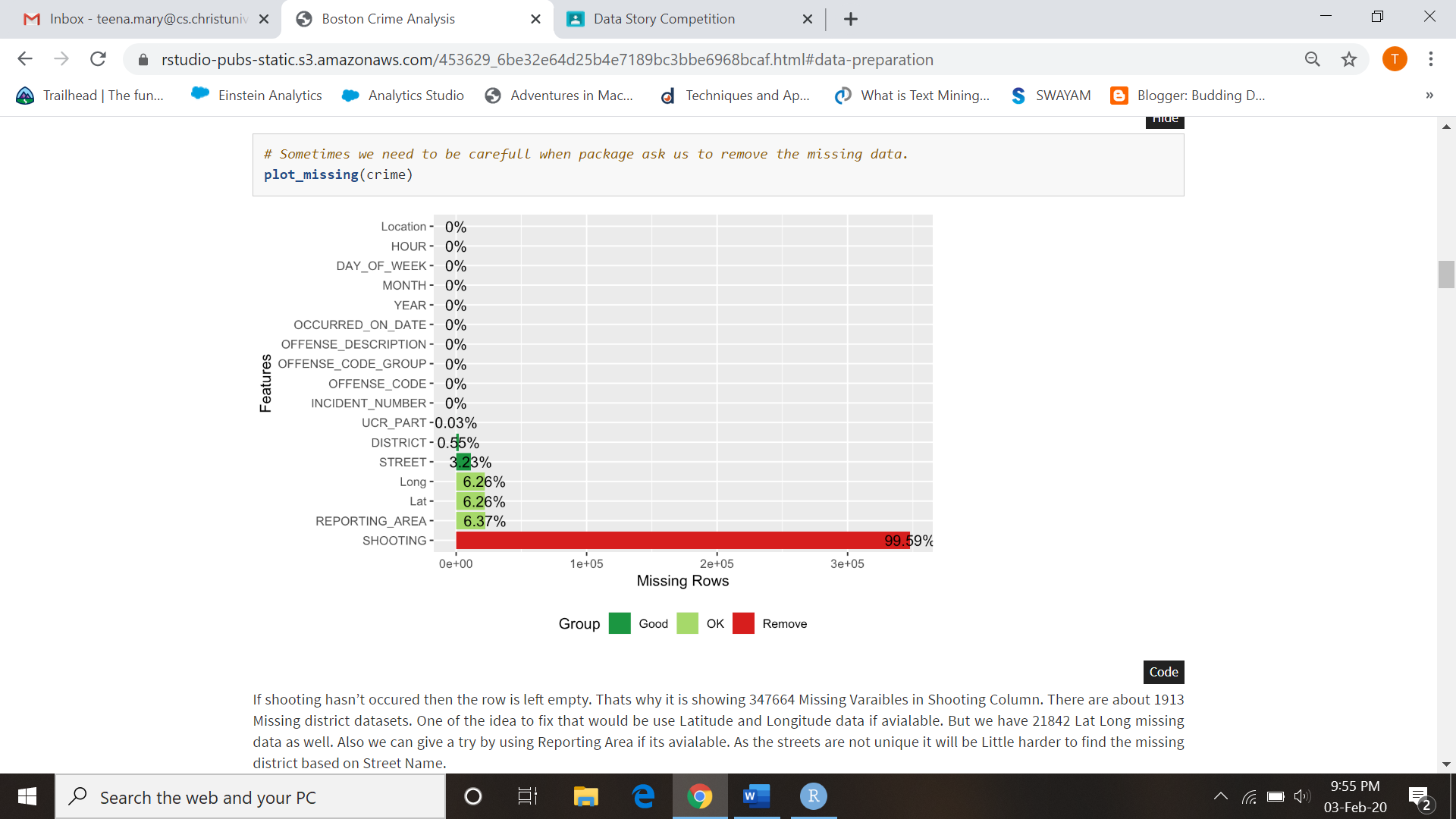
**Boston Crime Data set**

Public safety is vital to public health and happiness, and a state’s safety can be a crucial factor in deciding where to relocate your family. The map shows that Mattapan, Roxbury, and Dorchester have had more aggravated assaults than other neighborhoods. Roxbury, downtown, and Dorchester had more robberies than other sections of the city. Back Bay, the South End, and downtown suffered more larcenies. In 2017 the rate of violent crime in Massachusetts fell for the sixth year in a row and dropped below national levels for the second time since 2016, according to FBI statistics.

In 2017, there were 358 incidents of violent crime per 100,000 people in Massachusetts, compared to the national rate of 394 per 100,000 people, according to data published in late September. The FBI defines violent crime as homicide, rape, robbery, and aggravated assault. The bureau said it calculates crime rates using figures voluntarily provided by law enforcement agencies, and tallies estimates when it doesn’t receive a complete year’s worth of data.

**Missing Data Analysis:**



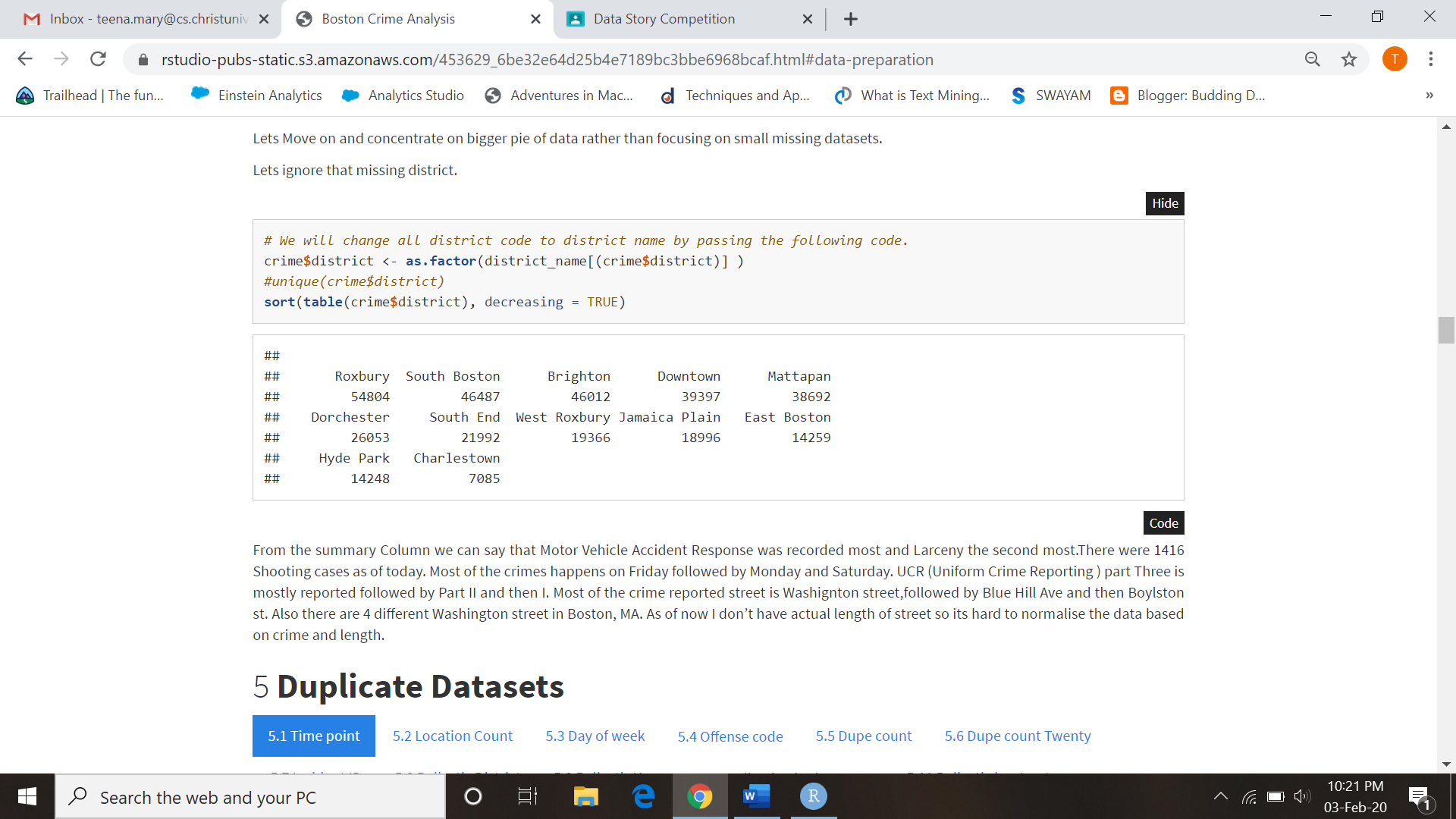


If shooting hasn’t occured then the row is left empty. Thats why it is showing 347664 Missing Varaibles in Shooting Column. There are about 1913 Missing district datasets. One of the idea to fix that would be use Latitude and Longitude data if avialable. But we have 21842 Lat Long missing data as well. Also we can give a try by using Reporting Area if its avialable. As the streets are not unique it will be Little harder to find the missing district based on Street Name.

Just knowing the district codes is hard to know where the crime was reported in Boston. People do know the neighborhood name rather than District name so our first step will be to recode the district code to District name.

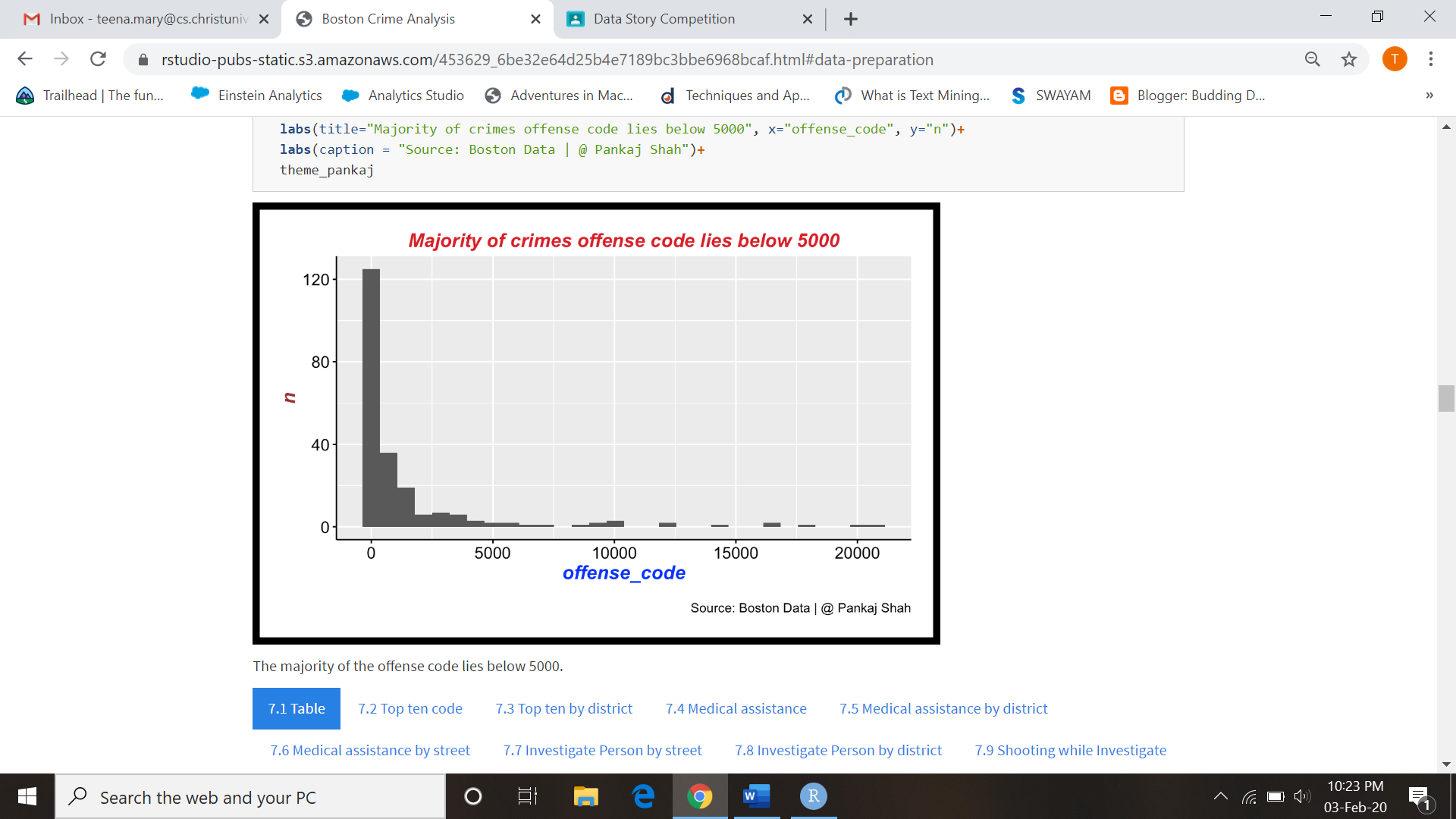
We could clean up these missing datas in many ways. Some of the ideas are by using reporting area to figure out which district these missing district are coming from in total there are only 40 columns with reporting. Latitude and Longitude position will be helpful to fix these missing districts. We have almost 1593 observations. It will be tedious work to chek each rows to fix and geolocation doesnt range by district as the district are not rectangle or square. There could be some efficient way to do these but will deal with these later.

For now my plan is to ignore these 1909 which is 0.0054% which shouldn’t have any affect on our datasets. While shooting it might fluctuate so I check that variable and found out there were 2 case where district is missing. I can fix by reporting area too. Most of the missing data are from 2018, almost 708, it could be still uploading as of today. I even break it down by month to see if certain month have whole missing data for district. But it was spread evenly between months as well.

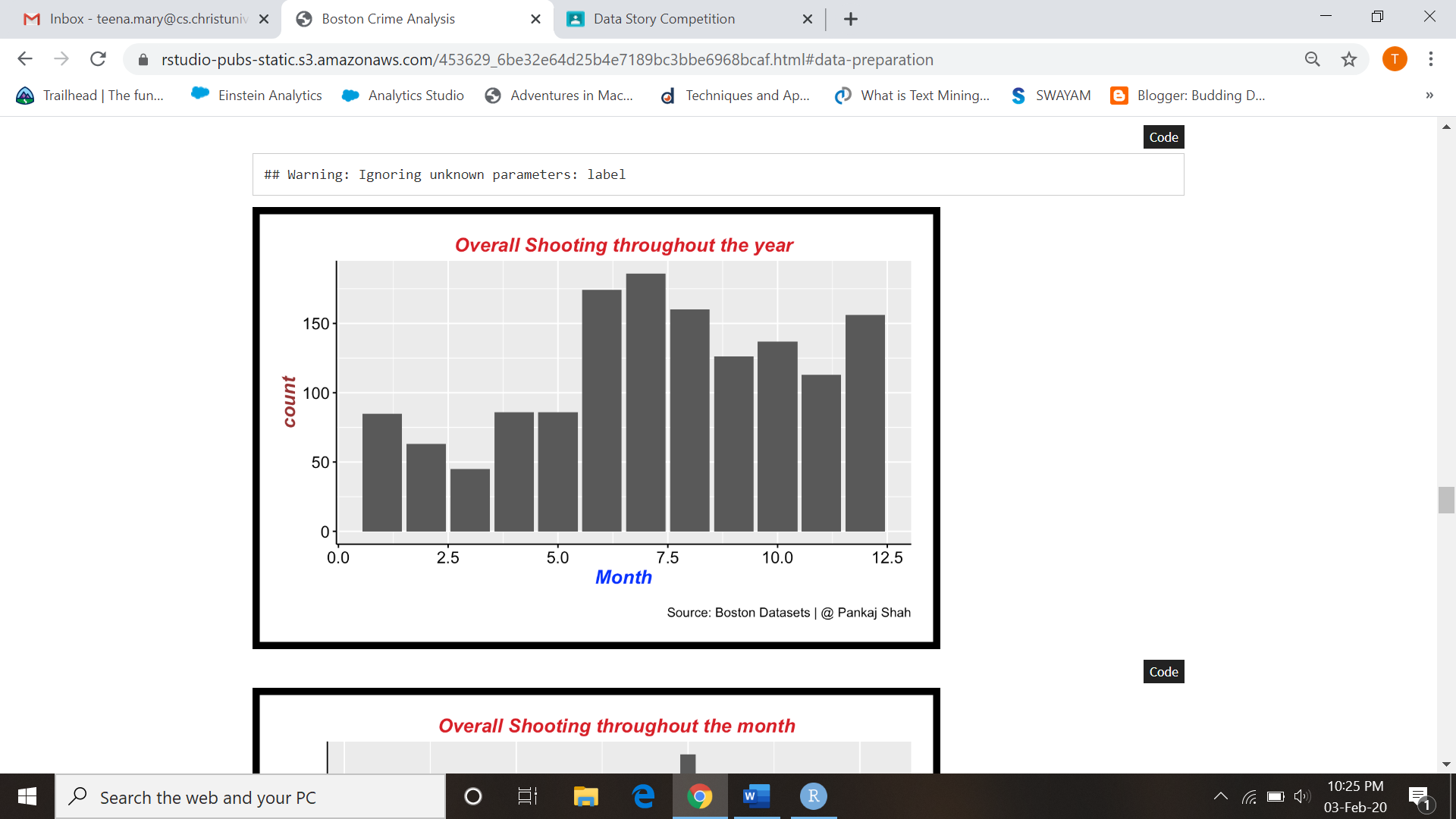


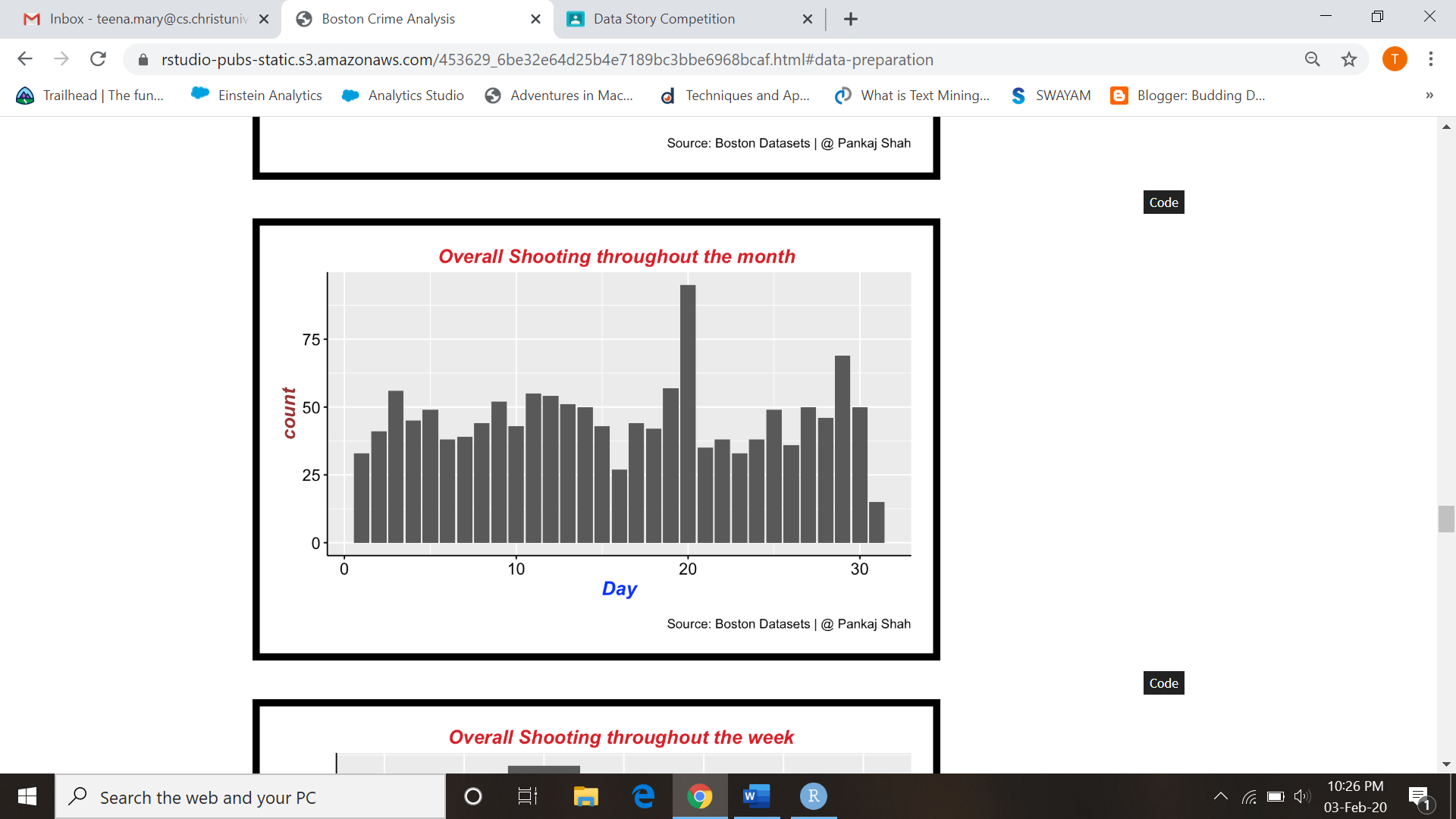
From the summary Column we can say that Motor Vehicle Accident Response was recorded most and Larceny the second most.There were 1416 Shooting cases as of today. Most of the crimes happens on Friday followed by Monday and Saturday. UCR (Uniform Crime Reporting ) part Three is mostly reported followed by Part II and then I. Most of the crime reported street is Washignton street,followed by Blue Hill Ave and then Boylston st. Also there are 4 different Washington street in Boston, MA. As of now I don’t have actual length of street so its hard to normalise the data based on crime and length.

**Offense code distribution:**

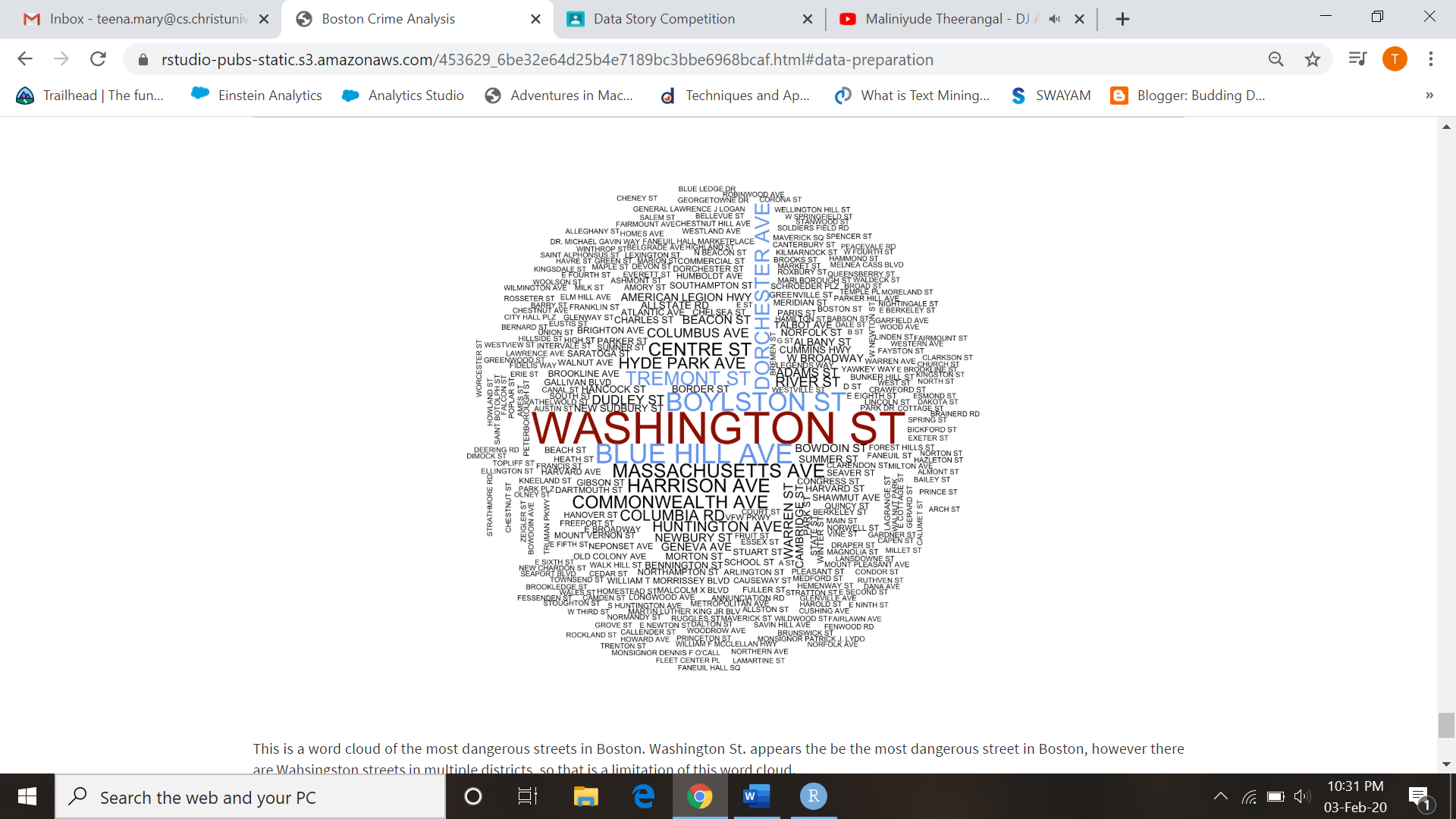


**Shooting:**





# ****Street Crimes: Most Dangerous Street****



This is a word cloud of the most dangerous streets in Boston. Washington St. appears the be the most dangerous street in Boston, however there are Washington streets in multiple districts, so that is a limitation of this word cloud.

# ****Conclusion****

After performing in-depth analysis on the Boston Crime Data Set, we can clearly see the trends and relations between the types of crimes, location and the occurance of the crime. Some of the notable takeaways from the analysis are mentioned below:

1. The most affected districts in Boston are Dorchester, South-End, Roxbury.
2. It can be noted that the highest number of the crimes were reported during summer months of July and Auguest.
3. Motor-Vehicle accident response were the highest number of report registered with the Boston Police.

This analysis can help Boston Police act accordingly and try to reduce the crimes frequently occurring in the city of Boston.

Frequency of Incidents by District: Dorchester, Roxbury, and South End districts have the highest cumulative incidents across the dataset. These three districts are also neighbors with each other, spanning from the South to the Southeast of Boston.