**Final Journal**

**Crime Data Analysis**

By

Usha Khetan - [khetan.u@husky.neu](file:///C:\Users\Jonathan%20Ajit\Desktop\khetan.u@husky.neu)

Jonathan Ajit Solomon - [solomon.jo@husky.neu.edu](mailto:solomon.jo@husky.neu.edu)

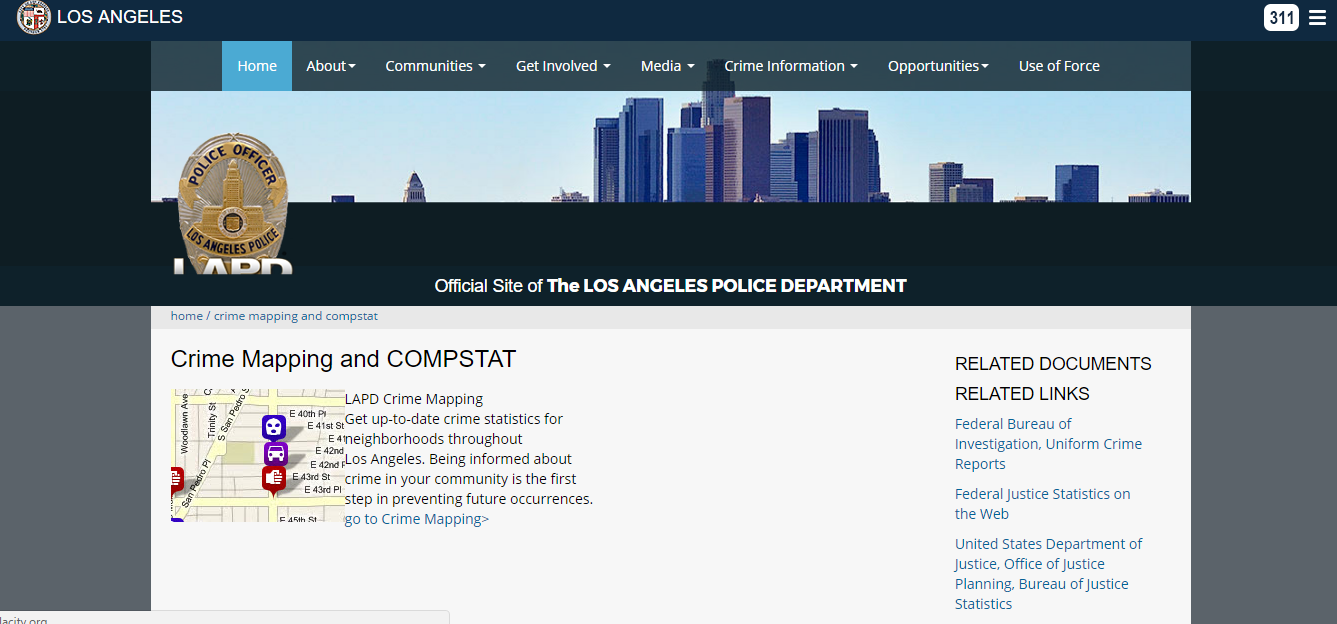


Data Analytics DA 5020  
334 Snell Engineering Center  
360 Huntington Avenue  
Boston, MA 02115

***Discussion; Date: March 09, 2017; Time: 02:00 p.m.***

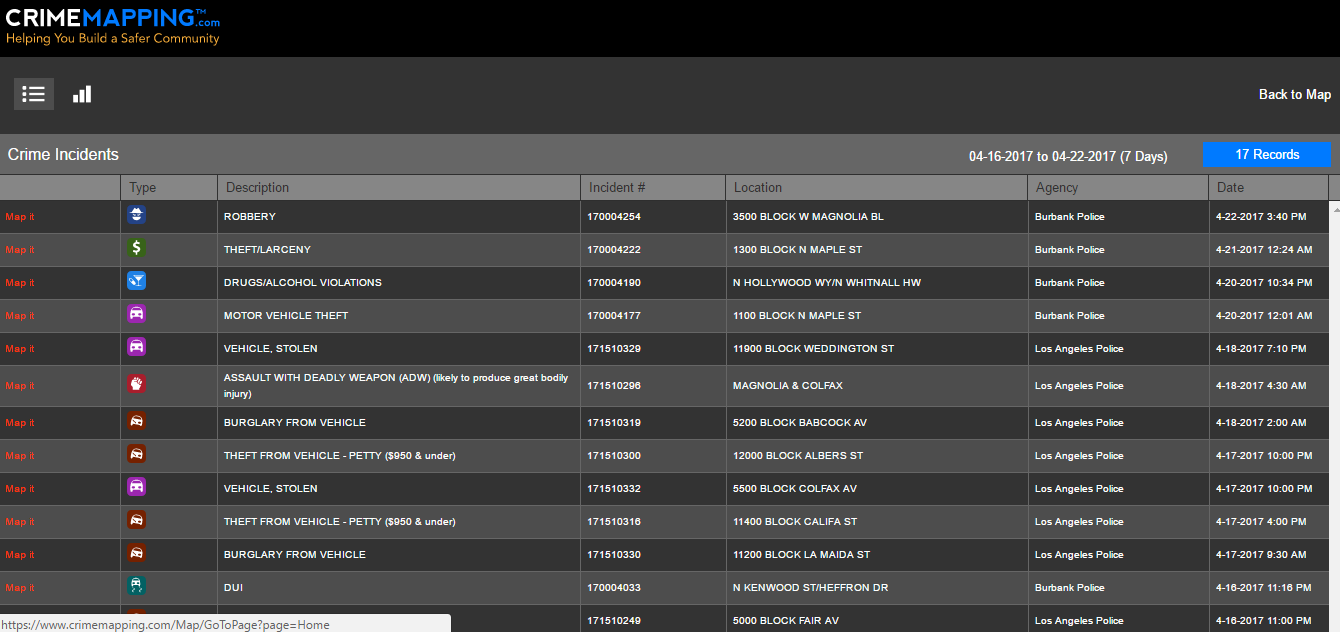
Usha and me decided to do the R project. We were given the liberty to choose our project for analysis. Pondering over what would be best to present an exploratory analysis for the final project and which would contain many attributes to propose an efficient model, we settled on the crime data analysis for the Los Angeles, California using the LAPD website

(<http://www.lapdonline.org/crime_mapping_and_compstat>)

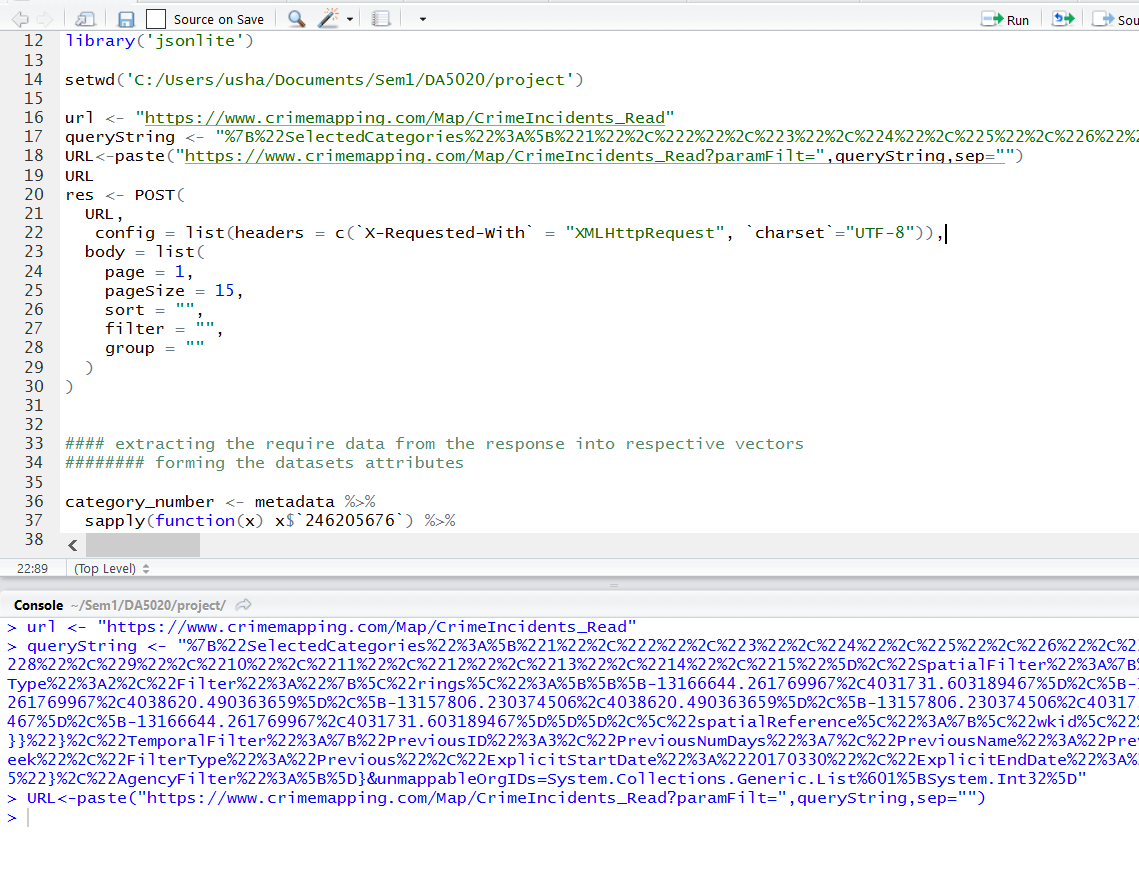


***Discussion; Date: March 17, 2017; Time: 03:00 p.m.***

Following the spring break, we then decided to go ahead and scrape data from the website. This was a dynamic website that had data updated once every week. This is how it looked;



As you can see in the screenshot shown, the query string is encoded and hence this was a hurdle we can to go across.



We then met up to put together our presentation which we had scheduled for March 24, 2017.

***Discussion; Date: March 24, 2017; Time: 03:00 p.m.***

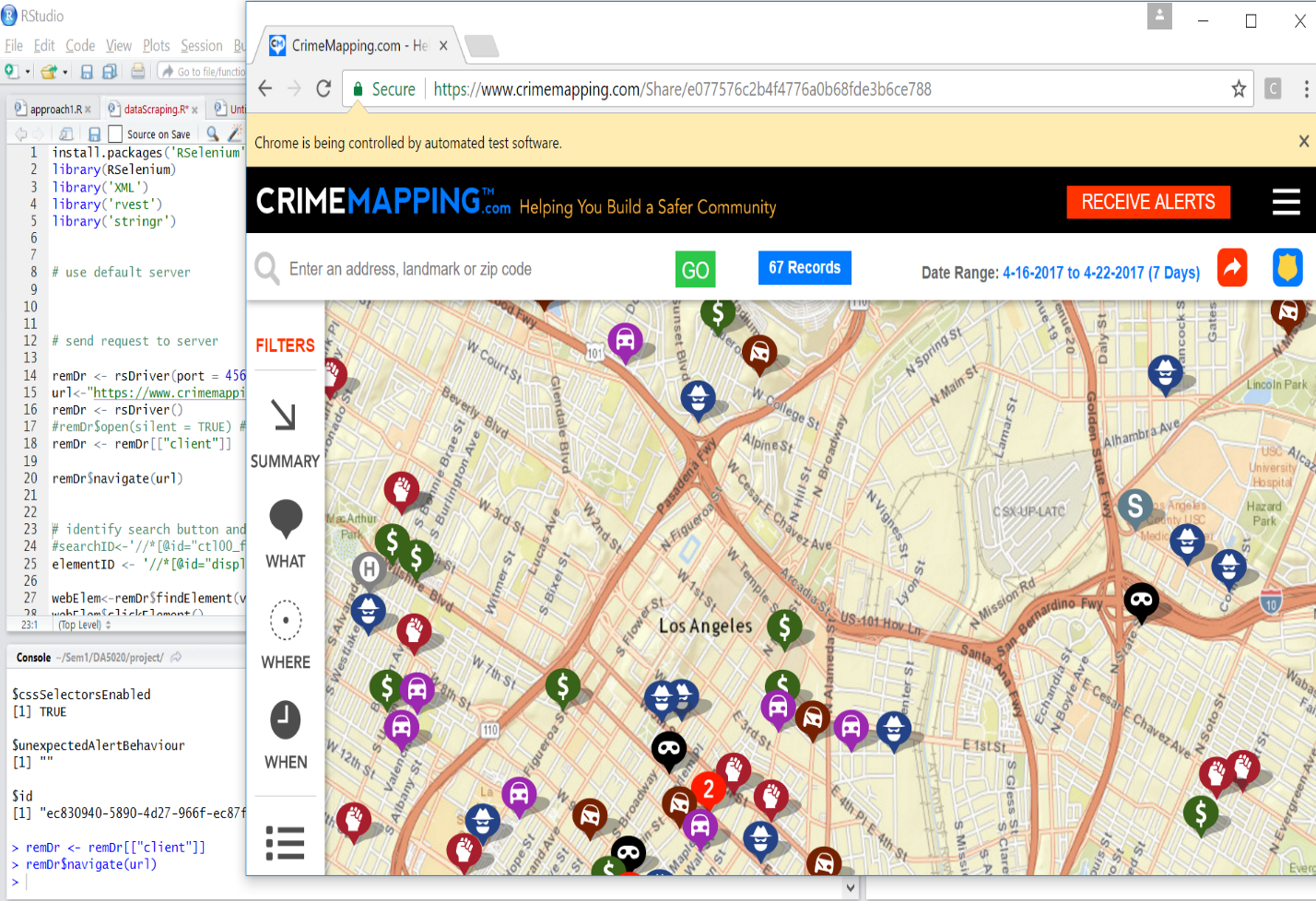
We had our presentation scheduled on this day, and here is the link to our ppt.

<https://prezi.com/mlsexgjjb51d/crime-report/?utm_campaign=share&utm_medium=copy>

***Discussion; Date: March 27, 2017; Time: 10:00 a.m.***

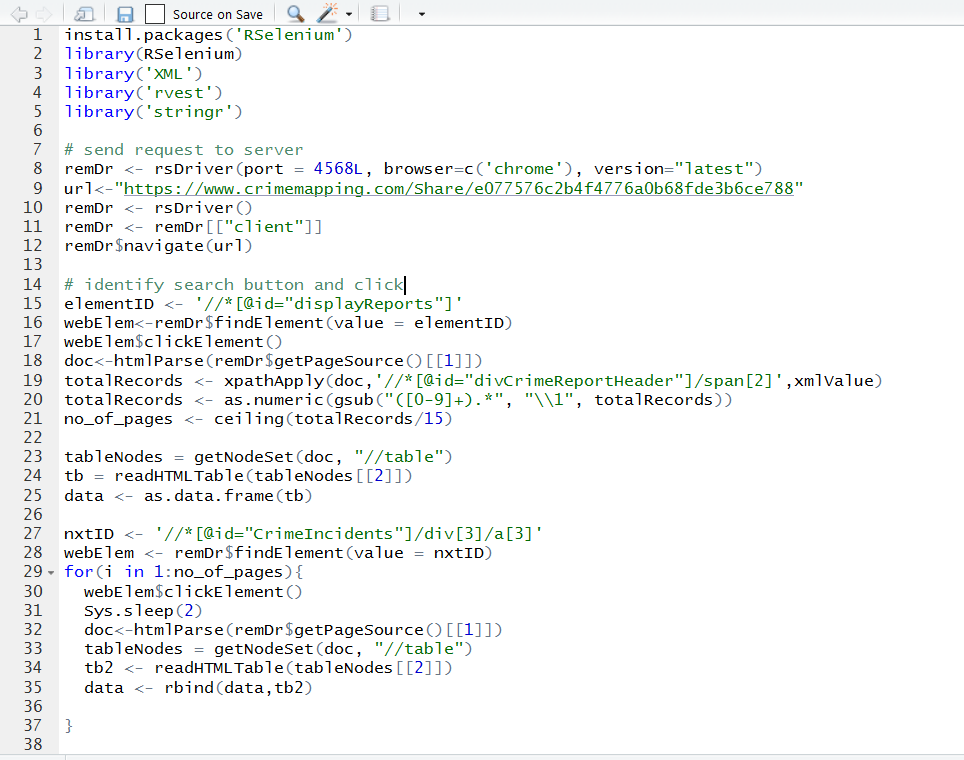
Having faced problems, the previous week, we decided to go ahead using another R package to extract data from a dynamic web page. On researching, we came across RSelenium that we felt could solve our issue of scraping.

We managed to scrape the data for the first few pages since we knew that each page has a total of 15 rows. This was how we went about our approach using RSelenium.



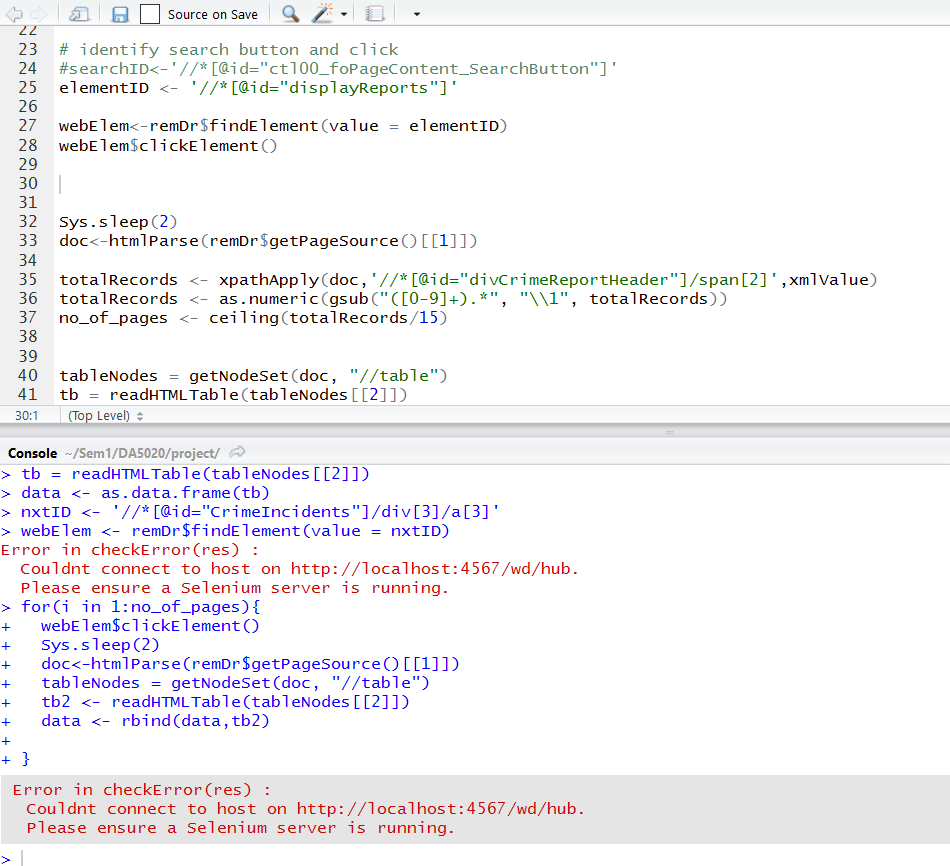
***Discussion; Date: March 29, 2017; Time: 05:00 p.m***

However, we hit a road block we figured out every time the proxy request a server, the data varies. This was our code though,



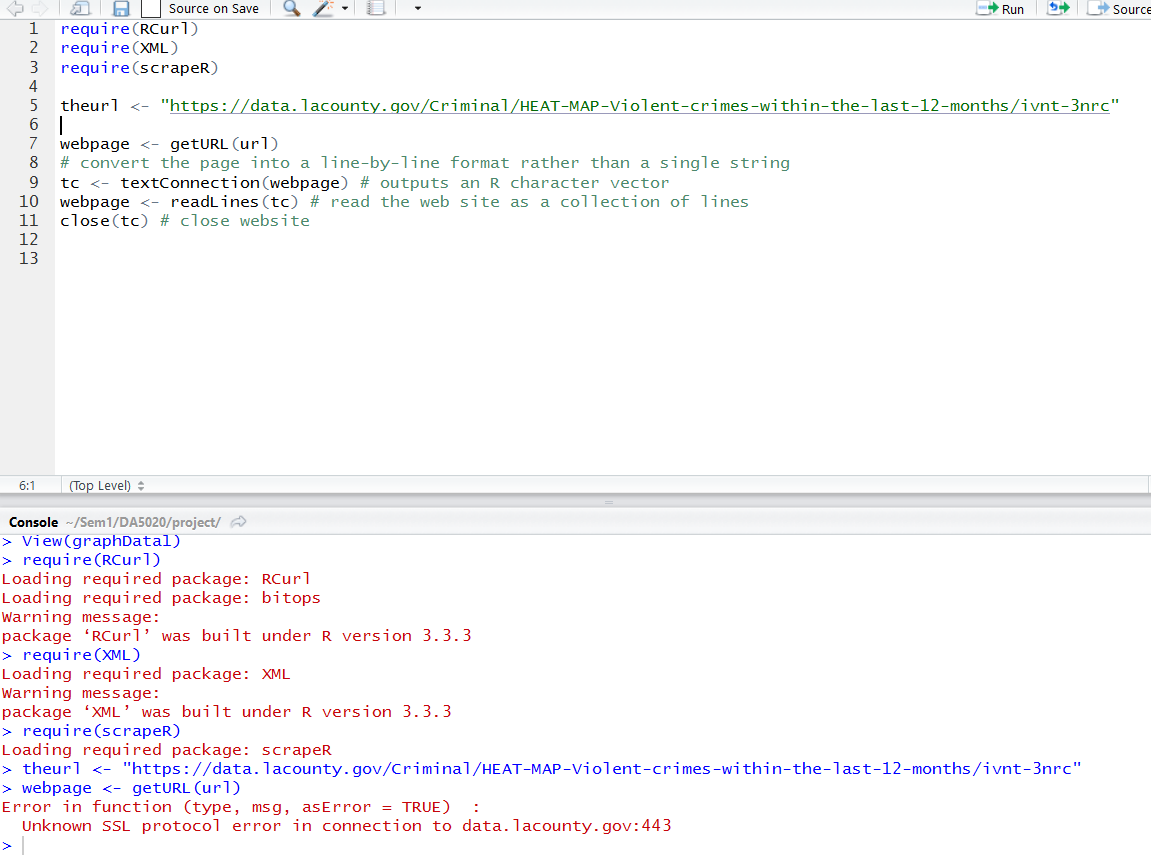
This was the error that we incurred,

Error: Server automatically shuts down after a while even before the entire data has been scraped.



***Discussion; Date: April 03, 2017; Time: 02:00 p.m.***

After trying to scrape the data in different methods, and having discussed the same with the professor and TA, we then decided to switch over to the LA county department open source website to scrape our similar data. When doing so, we came across an SSL error;



We found out that this occurs due to the website having some security norms. Henceforth, the error.

Therefore, we went ahead in our approach by using ***Postman API tool*** for the server request.

Finally, the Httr package in R was incorporated by us to make http easy.

Hence, our final approach and code was documented as mentioned in the R code uploaded and the documentation.