**Module 5 Project  
  
COMMUNICATION AND VISUALIZATION  
(Storytelling with Application)  
NORTHEASTERN UNIVERSITY**

**SUBMITTED TO: PROFESSOR JUSTIN GROSZ**

**SUBMITTED BY: REETIKA CHATURVEDI**

**[NEU ID-001081374]**

**Introduction**

This week, we extended our learning of making dashboards by making a dashboard using R-shiny. R-shiny helps to create visualizations as it is an open package in R studio. It is easy to plot our data and display it effectively.

**Research Questions**

I took traffic collision data that has the information of collisions that took place from 2010 to 2016. This data was published by Massachusetts Police Department. The questions that I have are as follows:

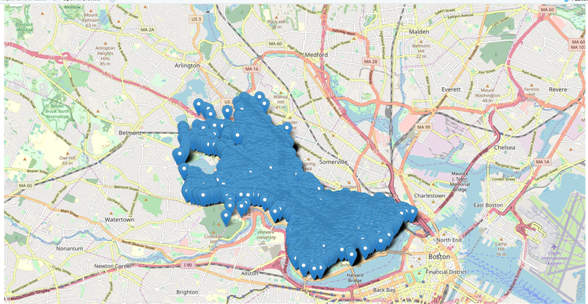
* Where are most of the collisions taking place?
* When are the collisions taking place? (which day)
* At what time the collisions are taking place?

**Analysis**

The data includes a lot of parameters including the date and time coordinates, name of street and location. The type of collisions is also given like by two wheelers or by auto. To calculate all this data needs to be sorted and cleaned.

The data required cleaning. Some fields were missing so I removed the null values using the calendar function. With the help of this function I tried to fill in the day of the week.

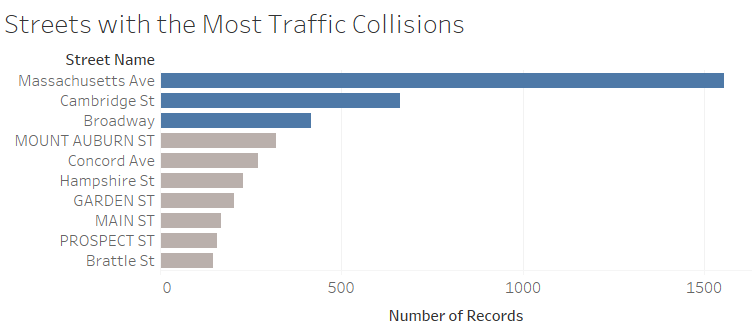
**Figure 1** – In order to find out the locations where most of the collisions took place, I used a map and plotted longitude, latitude and the number of different accidents. A pattern can be seen in the graph.



**Figure 1**: Six high-collision areas in Cambridge, MA

* Trying to find where most of the collisions took place (used Tableau for this)

**Figure 2** – The plotting of number of records by street name was done and the following was observed:



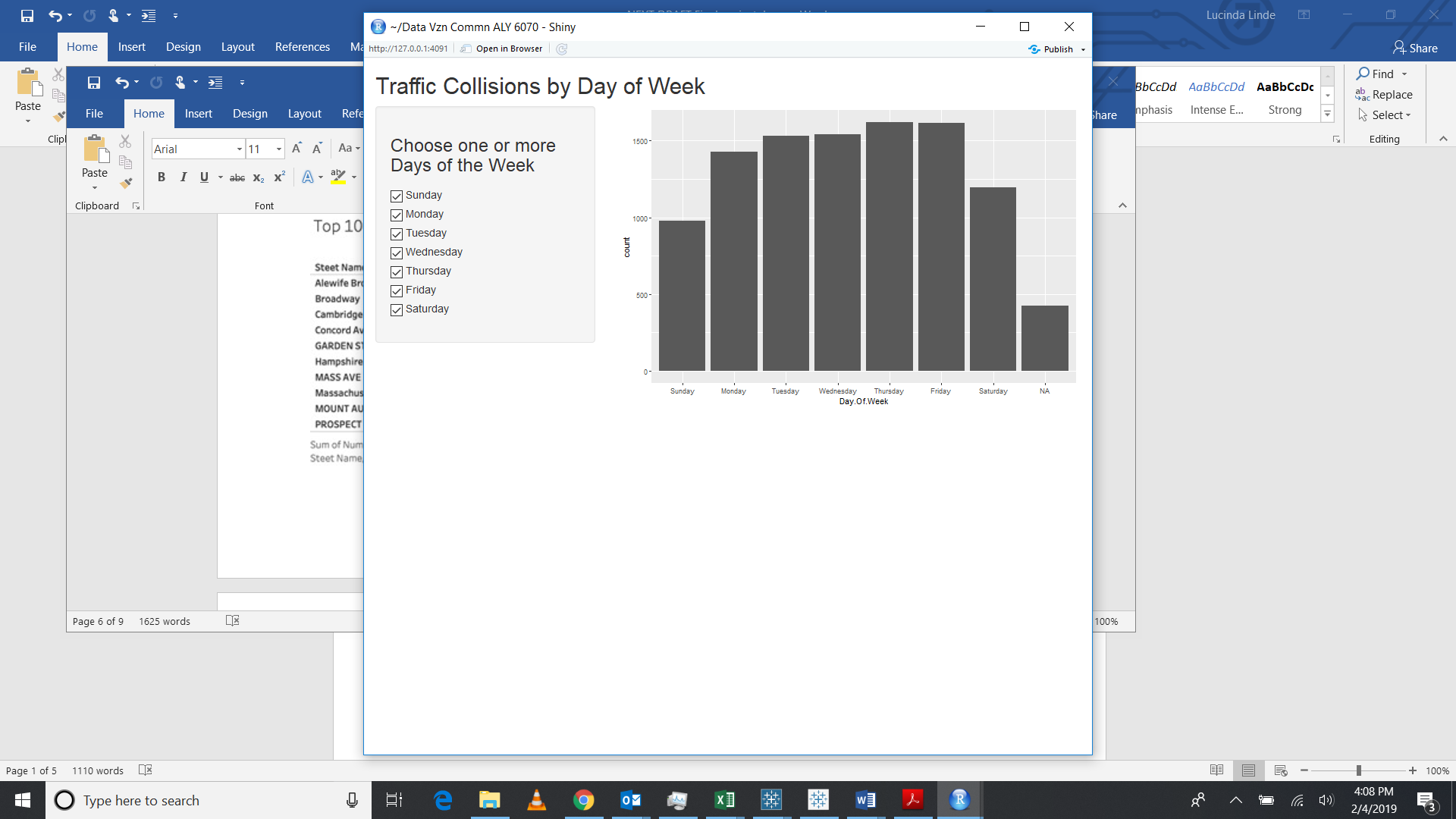
Mass Ave., Cambridge Street and Broadway have most of the traffic collisions taking place.

From the above figure it can be seen that the maximum collisions took place in Mass Ave. Also, the collision prone areas on Mass Ave are- Central Square, Porter Square, Harvard square and MIT.

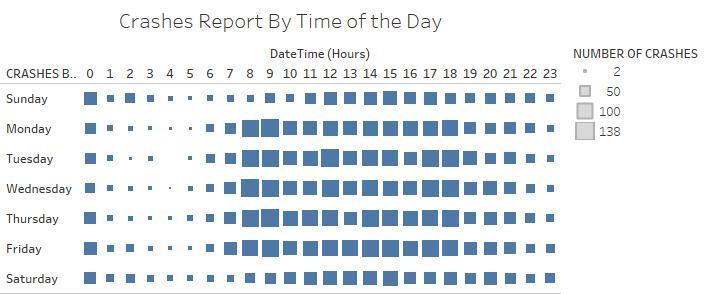
* When did the most collisions take place specifically the days and the time?

Once I was able to figure out the hit spots of the accidents, I tried to figure out that when these accidents actually took place.

After understanding the hot spots where most of the accidents happened, I focused on the time to analyze when most accidents occurred.

**Figure 3:** Number of Collisions by Day of the Week (made with R Shiny)

The x- axis has the days of the week and the Y axis has the count of the total accidents that were reported in the Cambridge area. It can be seen that the maximum collisions took place on Thursday and Friday. Cambridge area has many schools and offices so, it can be one of the reasons that this area is accident-prone.

I made a heat map. It shows the time in hours when the crashes took place. It can be analyzed that during weekdays the crashes are more between 8-10 and evening 5-6 pm. On the contrary, on the weekends the accidents are more in afternoon. **Figure 4:** Number of Collisions by Time of the Day

This figure shows the number of accidents that occurred in the top ten streets. It can be clearly seen that Massachusetts avenue has the most accidents and the frequency is more between 7am and 6pm. Broadway and Cambridge shows a somewhat similar pattern. The other streets have more frequency in the afternoon which can lead us to think that these places have restaurants where people might go for lunches.

