

**Crash Data Analysis with Visualization using Qlik**

Yi Yang

**ALY6070 Communication and Visualization for Data Analytics**

**Summer 2020**

**Professor Jack Bergersen**

**08/02/2020**

**INTRODUCTION**

Crashes involving motor vehicles are a part of our reality. We learn about them in the news as well as experience them in our personal lives. Analyzing crash data can not only effectively help decision-makers discover the causes and understand the results of crashes but can also lead to designing strategies to reduce or eliminate collisions and their outcomes.

The crash data we are using is an Excel file obtained from the Cambridge Police Department in Massachusetts. It contains a total 10335 observations of crash information and 12 attributes including Crash Number, Date Time, Day of Week, Object 1, Object 2, Street Number and Name, Cross Street, Location, Latitude, Longitude and Coordinates. In particular, the content of object 1 and 2 contains Auto, Pedestrian, Motorcycle, Parked Vehicle, Bicycle and some commercial vehicles like Taxi, Truck and etc.

Furthermore, we will utilize Qlik to explore two additional research questions and visualize our findings to better help the police department eliminate or reduce the number of crashes.

**Research Question 1: Do large vehicles pose a significant safety issue in Cambridge?**

***Step 1: Understand the context***

The rapid pace of development in Cambridge requires large vehicles transporting building materials along city streets.

One might expect that growth in development in a dense area such as Cambridge would result in more accidents and reduce safety. An examination of the Police Department dataset will determine if large vehicles like trucks and busses are having such an effect in Cambridge.

***Step 2: Data Cleansing***

The data was previously cleansed in preparation for the research using Tableau. The goal here is to examine crashes involving large vehicles in attributes Object 1 and Object 2. We define “Large\_V\*” in the dataset as:

* Bus(other)
* Bus (more than 15 people)
* MBTA Bus
* School Bus
* Truck

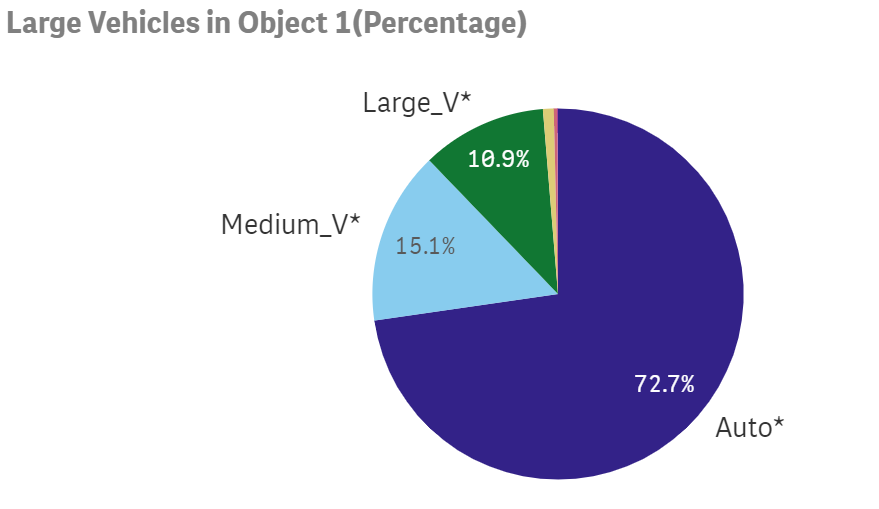
Multiple values with the term “Truck” or “Tractor” were converted into one category named “Truck”. This process changed 136 “Object 1” and 88 “Object 2” values.

To make the dataset more consistent, we define the following terms:

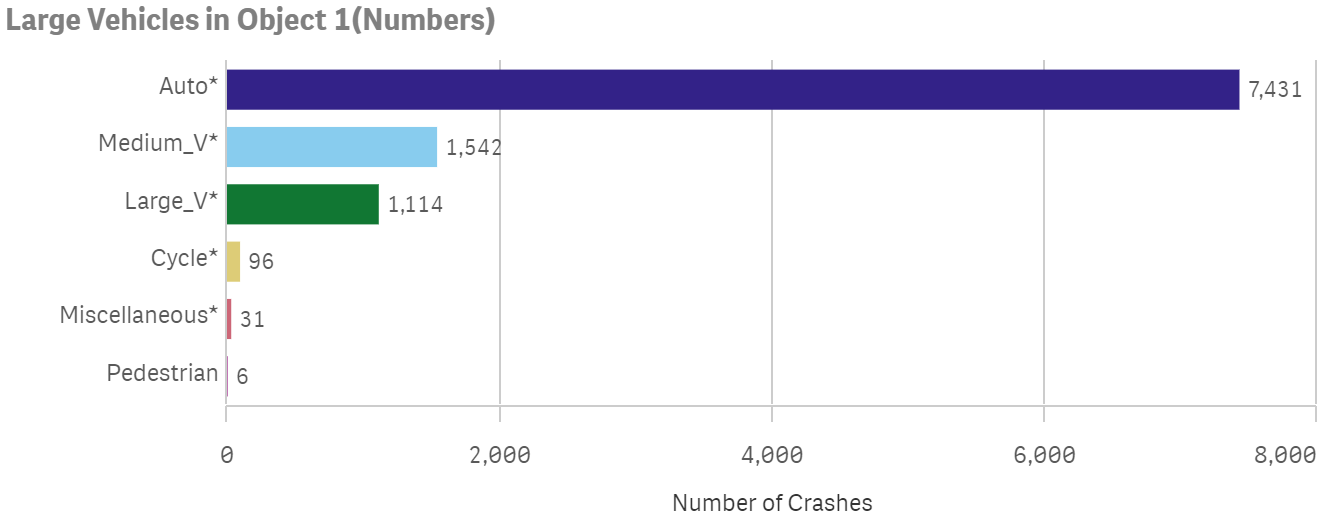
* “Medium\_V\*”:
  + Passenger Car
  + Van
  + Bus with 7-15 seats
* “Auto\*”
  + Auto
  + Taxi
* “Miscellaneous\*”
  + Miscellaneous
  + Other
* “Cycle\*”
  + MOTORCYCLE
  + Motorcycle
  + Bicycle

***Step 3: Choose a Display***

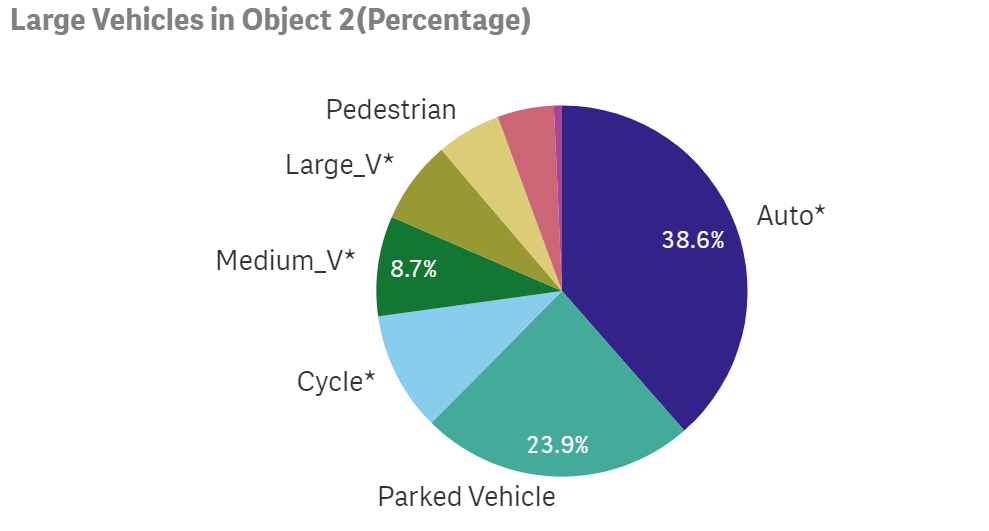
We created an app in Qlik to upload and store our data, then we built up some visualizations. We developed a sheet that displays a pie chart and a bar chart to illustrate the proportion of each vehicle type occupied in the total number of crashes involving “Object 1” and “Object 2” but that also enables the user to drill down and obtain counts for any selected “Object 1” and “Object 2” values.



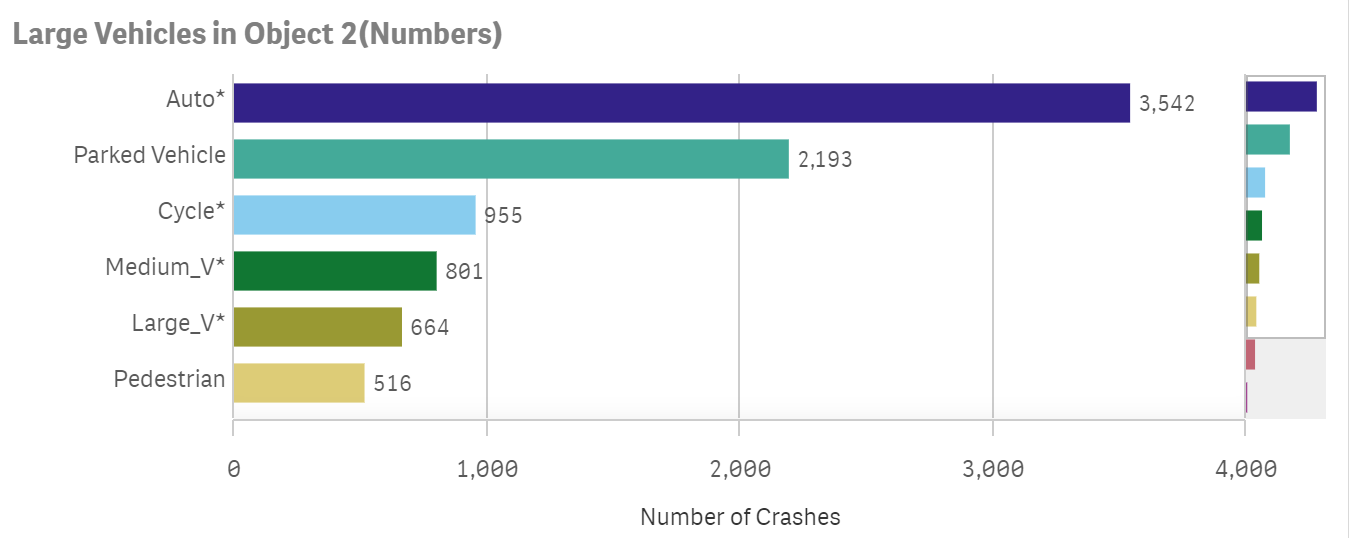
*Figure 1.* Pie Chart of Number of Crashes vs Large Vehicles in Object 1



*Figure 2.* Bar Chart of Number of Crashes vs Large Vehicles in Object 1



*Figure 3.* Pie Charts of Number of Crashes vs Large Vehicles in Object 2



*Figure 4.* Bar Chart of Number of Crashes vs Large Vehicles in Object 2

***Step 4: Summary***

Based on the proportions of crashes, large vehicles are not a significant safety issue in Cambridge. The number of crashes involving large vehicles in attribute “Object 1” takes around 10.9% (1114) of all accidents recorded. Crashes in “Object 2” categorized as large vehicles take 7.2% (664).

Of the 10,335 crashes, 3372 (45.4%) involved two automobiles (both “Object 1” and “Object 2” are automobiles). Another 2072 (27.9%) crashes occurred between an automobile (“Object 1”) and a parked vehicle (“Object 2”). The third largest number involved automobiles (“Object 1”) and cycled vehicles, 904 (12.2%), and the fourth greatest number occurred between automobiles(“Object 1”) and pedestrians 490 (6.6%).

Certainly, any collision is serious, and in fact, serious injuries or fatalities involving large vehicles such as trucks have taken place in Cambridge. Our research finds, however, that the percentage of accidents involving large vehicles is much smaller than other crash types such as Auto, which is more commonly found on Cambridge streets.

**Research Question 2: What year and month that crashes happen most?**

***Step 1. Understand the Context***

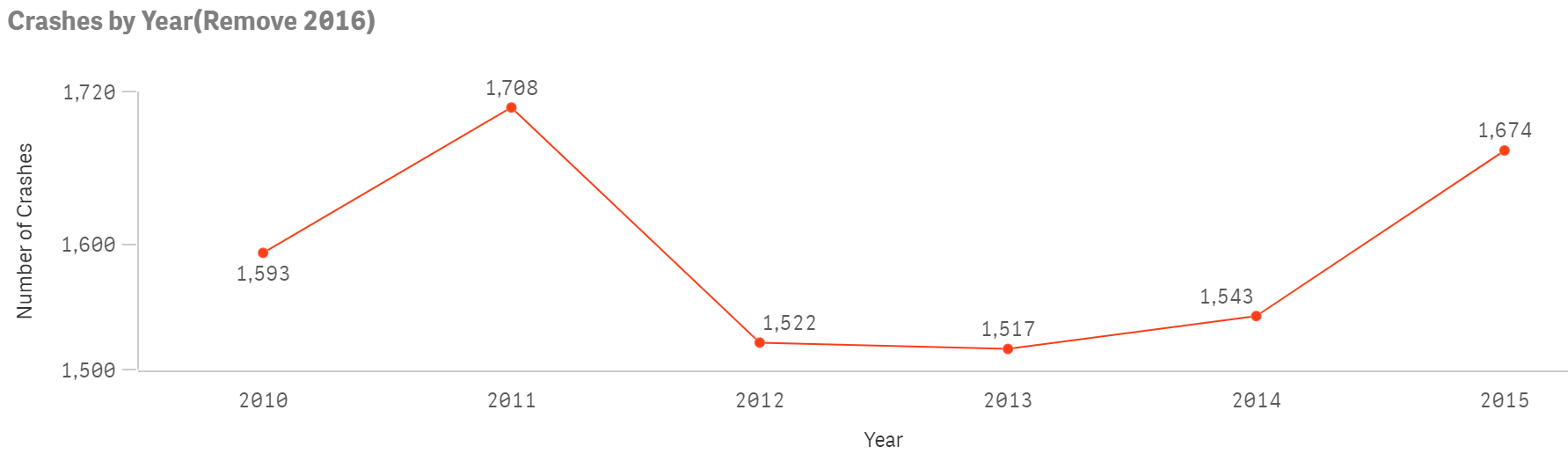
Based on our findings from the Tableau assignment that Thursday is the most frequent day for crashes occurring, we will further visualize the distribution of crashes by Year and Month.

***Step 2. Data Cleansing***

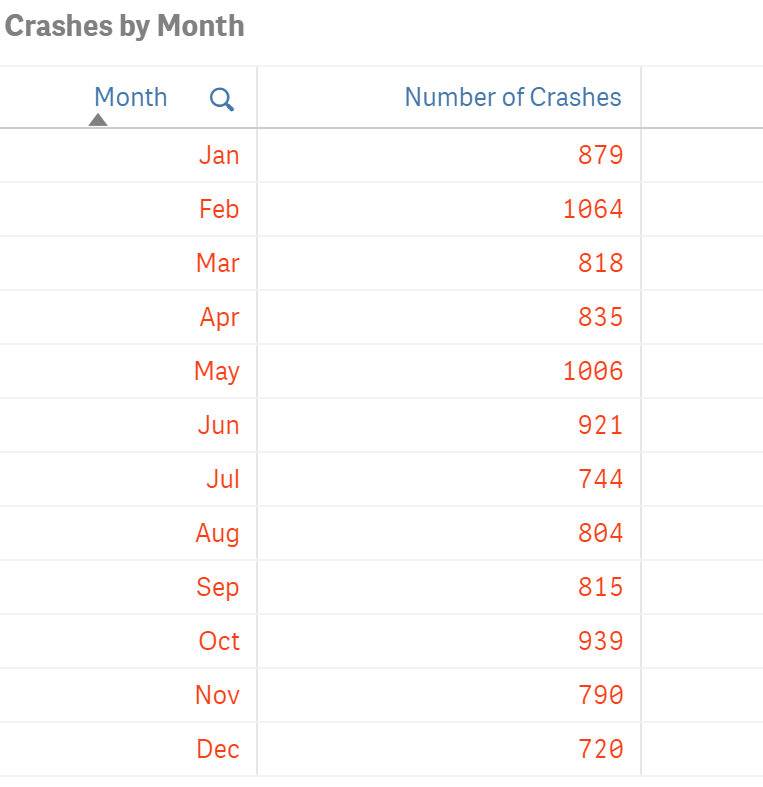
The attributes “Date Time” and “Crash Number” are used.  Neither attribute has null values. Since the year 2016 includes only six months of data, we exclude it for more consistent results.

***Step 3. Choose a Display***

We created another sheet in Qlik which includes one line chart and one table below showing how “Crash Number” (number of crashes) is distributed by different years and different months.



*Figure 5.* Line Chart of Number of Crashes vs. Year (Exclude 2016)



*Figure 6.* Table of Number of Crashes vs Month

***Step 4. Summary***

From Figures 5 and 6, we can tell that the largest number of crashes happened in 2011 followed by 2015.  Also, February and May seem to be the most dangerous of all months.

**CONCLUSION**

According to the above analysis, we summarize some results. The analysis tells us that instead of the large vehicles, the regular cars are the main object involving accidents since the total number is way more than other objects. Also, the American Community Survey of 2011-2013 shows 7% of residents commuting by bike during rush hours which increased two more times compared with the previous year.

However, Cambridge was having a lot of construction in 2014, the total number of the cycled vehicle took dip down.

Therefore, we realize it did help decrease crashes when people commute by cycled vehicles. The crash number dramatically increased when the number of cycled vehicles decreased since the construction.

As a result, the architecture should put attention on how to ensure the safety of the cycled vehicle rider. The number of bicycles has an increasing trend. How to separate the bike and the vehicle during construction becomes the biggest concern, and whether the route is big enough to handle the increasing number of bikes or will there be any safety issue while pedestrians and bike riders share the road and so on.

Based on the insight, we can see where the issue was and why it happened, so we can find a solution and overcome it effectively.

**REFERENCE**

(n.d.). Cambridge CDD@344 BICYCLE TREND. Retrieved 08/02/2020 from <https://www.cambridgema.gov/CDD/Transportation/gettingaroundcambridge/bikesincambridge/biketrends>