Traffic Data Collection

Luis Ureña
Napa Valley College

DDETFP poster 135
P21-20942
Wednesday, January 27,2021 2:30 PM- 4:00 PM ET

Abstract

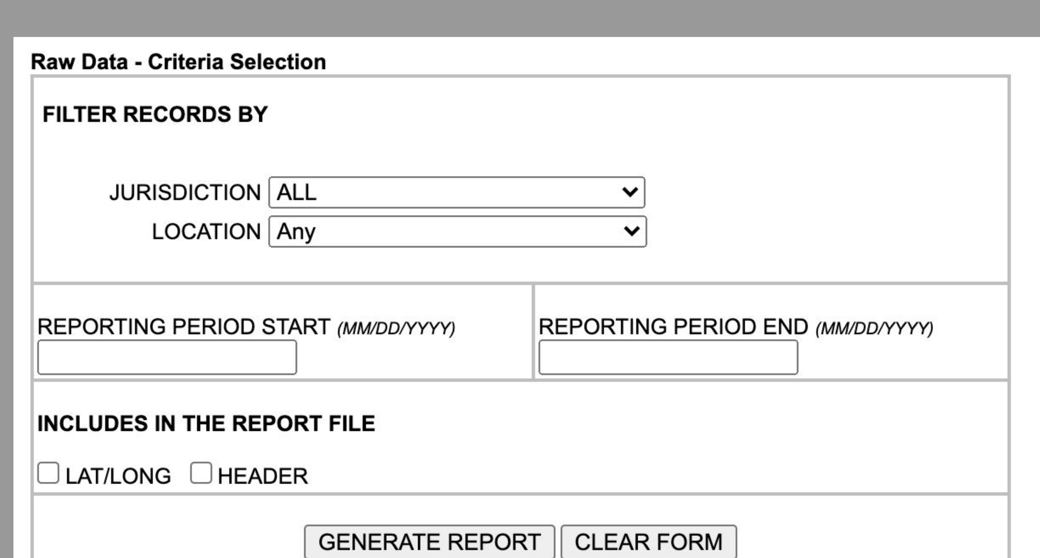
Traffic data is collected in many ways, shapes, and forms. It has the power to change our decisions of what roads we will take to work, influence our public investments, increase our public safety, and ultimately improve our well beings. This research studies the different tools that are available to analyze traffic and safety and explores the accessibility to the public of this data. This is done using a variety of sources. Interviews of the average consumer in the community to explore the everyday tools that consumers use. In addition, investigating physical hardware that exists to help consumers, and professionals to collect data. Also, it is done by searching public databases and local data-keeping entities. There exist some examples of traffic entities that allow very quick ease of access to traffic data that are prime examples of what could help generations to come. The expectation is that there are many sources of traffic data with different accessibilities to consumers. Traffic data sharing can benefit communities and their well beings by increasing their availability to the public.

Materials Used

- California Highway Patrol: Statewide Integrated Records System (SWITRS)
- City of Napa Public Works Department
- MATLAB

Procedure

There was a variety of places that had traffic data available to the public. The most useful was the SWITRS database. It allowed to RAW data to be exported for further analysis. An account needs to be created - which was really easy to make. Once you're registered, the website allows you to request data from several counties, or the entire state.



Once placing a request for data, an email will be sent to you with 4 different files for you to use

One of the files has just the report data. The second file contains the Victim data. The third has the Party involved Data. The last file contains the instruction about how to interpret the data.

01 - Driving or Bicycling Under the Influence of Alcohol or Drug
02 - Impeding Traffic
03 - Unsafe Speed
04 - Following Too Closely
05 - Wrong Side of Road
06 - Improper Passing
07 - Unsafe Lane Change
08 - Improper Turning
09 - Automobile Right of Way

Data interpretation Manual

After the data was downloaded, I noticed that there was many inconsistencies. This meant that code had to be written in order to do make up for inconsistencies

When summarizing data, it was observed that conclusions could not be made without looking at other factors that could lead to the same conclusion. For example, when looking at alcohol related incidents, the data was not filled out completely in one column, but other indicators in other columns could get us more information.

```
tbl=readtable('LosAngelesCountyCollisionRecords2019.txt');

%for i = 1:size(tbl)

    out = tbl(ismember(tbl.ALCOHOL_INVOLVED,'Y'),:);
    out2= tbl(ismember(tbl.PCF_VIOL_CATEGORY, '01'),:);
    out3= [out;out2];
    out3=unique(out3);
    %if(tbl.ALCOHOL_INVOLVED == "Y")
    %    print('found');
    %end

    writetable(out3,'LosAngelesCountyCollisionRecords2019ALCOHOL_INVOLVEDandPCF_VIOL_CATEGORY.txt')

%end
```

MATLAB Code written to summarize ALL alcohol related incidents

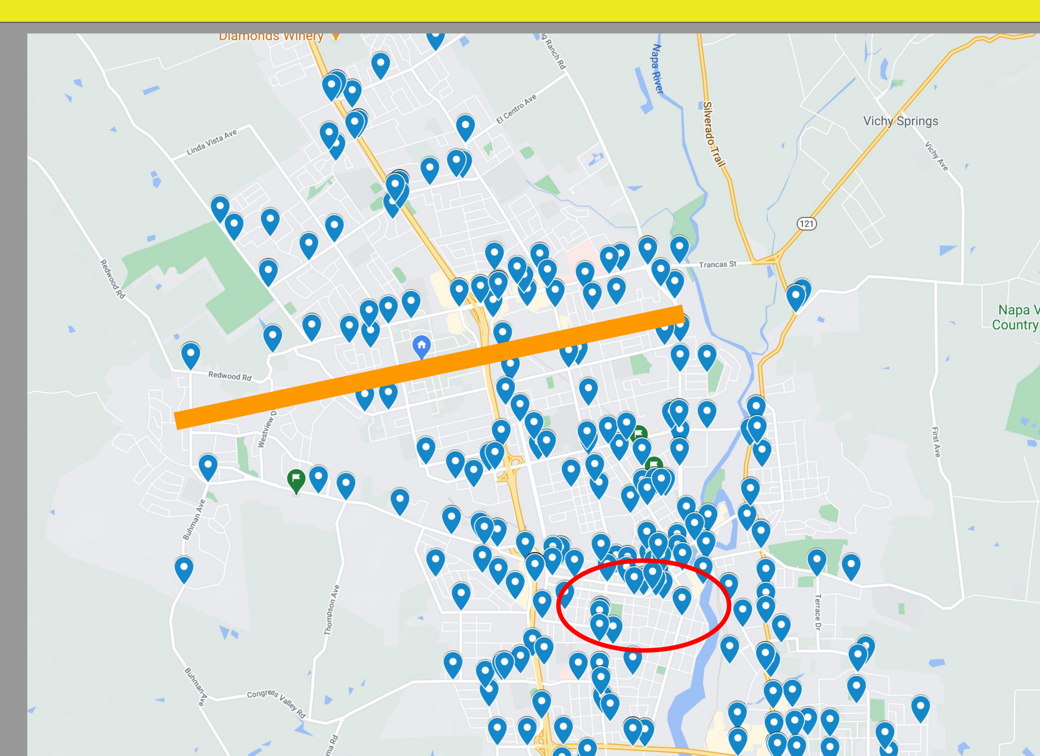
California Highway Patrol: Statewide Integrated Records System (SWITRS)

According to their website "The Statewide Integrated Traffic Records System (SWITRS) is a database that collects and processes data gathered from a collision scene. The Internet SWITRS application is a tool that leverages this database to allow California Highway Patrol (CHP) staff, members of its Allied Agencies, as well as researchers and members of the public to request various types of statistical reports in an electronic format. The application allows for the creation of custom reports requested by the user based on different categories including, but not limited to locations, dates, and collision types." (The California Highway Patrol)

The system allows for reports to be generated. We can analyze the data and use it for useful traffic data to make future decisions about our infrastructure and the needs of our community.

The SWITRS database holds a lot of information about traffic incidents. This includes: Case ID, Collision Year, Day of the Week, Direction, Road Conditions, Lighting, Alcohol Involvement, and Party Drug present, Party Age, Violation Codes and much more.

Traffic Incidents at the Local Level



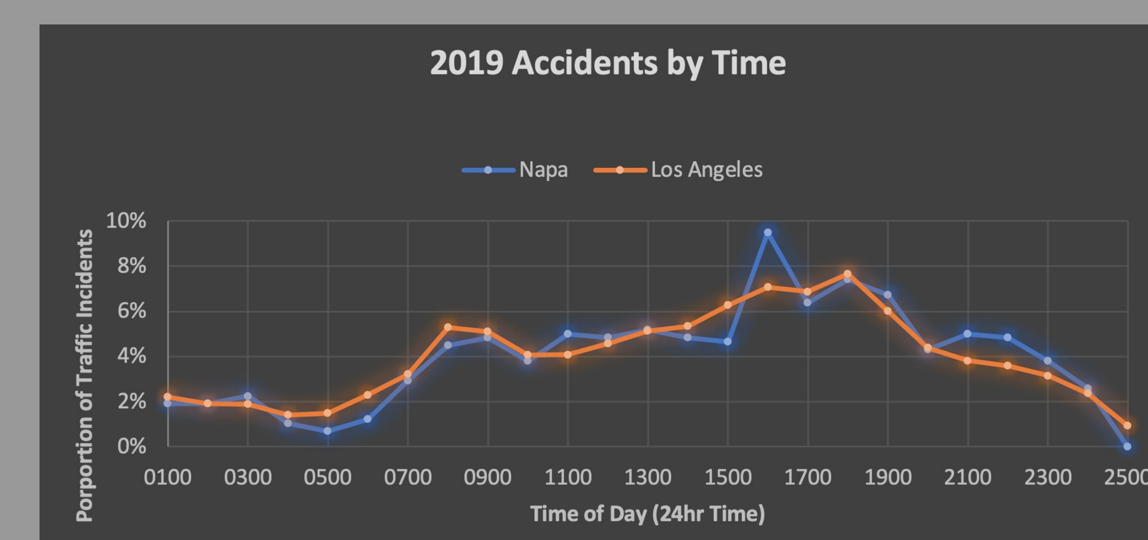
2020 Traffic Incidents, mapped, approximated

In order to map out incidents, there was a long process involved. Reason being that the Napa Police Department does not record GPS Coordinates which would be a lot easier to analyze as Napa California Highway Patrol does. So, in order to map it, I had to use a special Google Sheets add on to add GPS Coordinates based on the recorded nearest cross roads- which in return means that the map is not 100% accurate. The Napa PD information is only based on the neared crossroads.

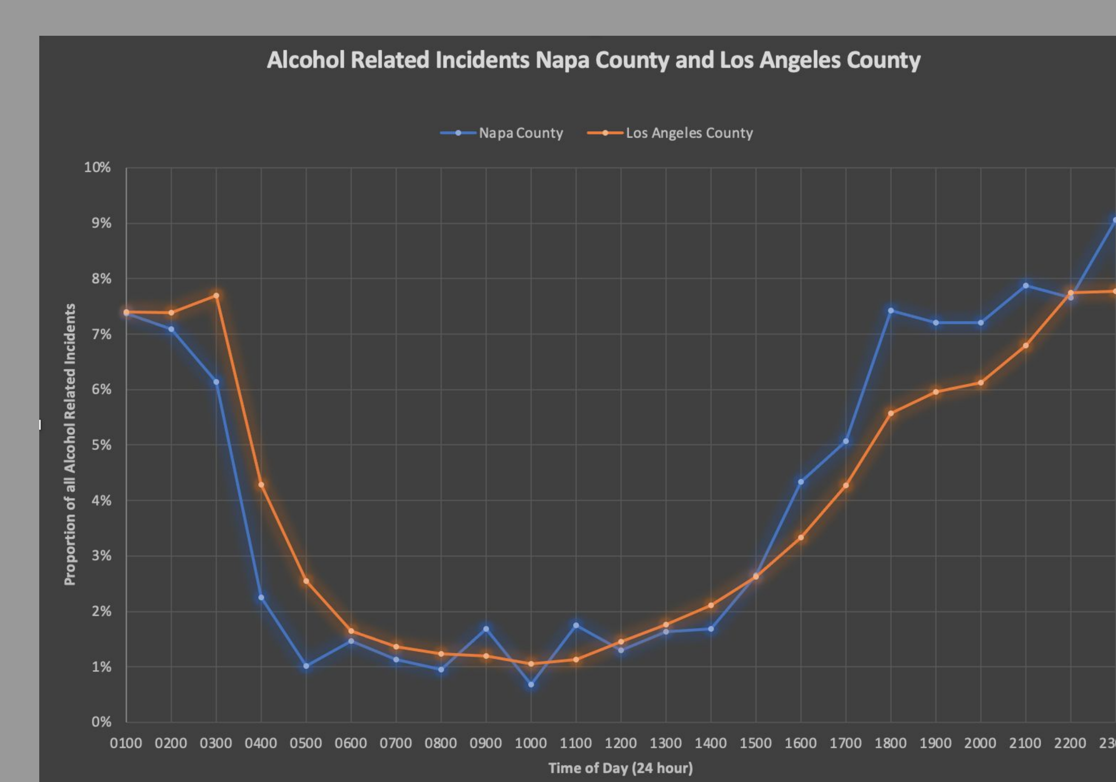
From the map, see a few hotspots. More noticeably is Redwood Rd (noted by the orange line). It is a major traffic incident road. The road connects a lot of neighborhoods to commerce and other freeways/highways which could explain the amount of accidents.

In addition, another big hotspot for accidents is the down area (noted by the red circle). It is surprising to me because the speed limit downtown is only 25 miles an hour. As Napa is a huge tourism destination, this may account for the large amount of incidents in the tourist area.

How does it compare to other California Counties?



With Data Collected from SWITRS, inputting the data into Excel for further analysis. From this Histogram, we can analyze that the most a majority of incidents arise starting at about 3pm - the time that most people get out of work.
There is another high point between 7am and 9am - the time most people commute to work.



Trying to compare Alcohol Related Incidents to the entire state was difficult. To begin, the amount of data that had to be analyzed was unsupported in Excel. In addition, MATLAB gave an error because of a inconsistency in the number of parameters - this was a fault of the file downloaded from the SWITRS system.

Another issue was that Napa County had very little Alcohol Related Incidents luckily. The graphs were hard to compare because it fluctuated compared to the abundant data from Los Angeles. Eventually I accumulated 1 year of Los Angeles County to compare a decade from Los Angeles

Results

- According to our data derived from the SWITRS System, we found that downtown Napa and Redwood Rd/Trancas Rd. were the most common places to run into accidents.
- We found that there is a huge uptick in accidents starting at 3pm - that includes Los Angeles County and Napa County.
- It seems to be that there is a large proportion of Alcohol related incidents starting at 6pm.

Conclusion

The SWITRS System is a really useful tool to get public data. In order to use it accurately there needs to be some error checking put into place before drawing any conclusions. In addition, some more data could be collected by all traffic entities in order to provide more useful and accurate data to analyze.

Limitations / Future Direction

- There are many data limitations to this to the SWITRS system. There are various instances that the data is not filled out all the way and has a lot of missing data. For example, when looking at specifically Napa County Data, the Napa Police Department does not collect GPS coordinates of incidents that happen on the road.
- On the other side, the California Highway Patrol Data does take note of the GPS coordinates of all incidents. Thus creating an inconsistency in the data collected from the County and makes it harder to analyze data.
- While having data for the public to use, it is also important to release the raw data. The raw data is a lot easier for the public to use
- More code could be written to better refine the data. For example, there was some obvious mistakes that were made my whomever write the time of incidents over 24hrs. That data was excluded and affected about .4% of data - a negligible amount but very worth of noting

```
>> onlyGrabPCFViolation
Error using readtable (line 216)
Reading failed at line 4005903. All lines of a text file must have the same
number of delimiters. Line 4005903 has 76 delimiters, while preceding lines have
75.

Note: readtable detected the following parameters:
'Delimiter', ',', 'HeaderLines', 0, 'ReadVariableNames', true, 'Format',
'%q%f%f%f%f%f%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%q%f%f%q%q%f%q%q%q%q%q%f%f%

Error in onlyGrabPCFViolation (line 1)
tbl=readtable('CaliforniaCollisionRecords20102020.txt');
```

References

- MATLAB
- Antonio Castro, Professor, Napa Valley College
- Luis Alcazar, MESA/STEM Center Associate Dean , Napa Valley College
- The California Highway Patrol. https://iswitrs.chp.ca.gov/Reports/jsp/index.jsp. Accessed 20 January 2021.