Traffic Accident Analysis, Commonalities, and Impacts on State Highways in Los Angeles County, California

8/23/2020
Geog 574
Celine Chen, Emily Johansen, and Mark Wojta

Introduction

Our objective is to understand transportation patterns and traffic accidents for California state highways. To accomplish this, we decided to do a comparison of available 2010 and 2017 data. This allows us to better understand traffic and accident patterns over a period of 7 years. An important variable to keep in mind for this study, is that according to the United States Census Bureau, the estimated population of Los Angeles County in 2010 was 9.8 million people and in 2017 was 10.1 million people. We chose 2010 data to use because it was the earliest year in which various California state agencies all had complete datasets. 2017 data was chosen because it represents the most recent year of accurate data across the various datasets available. According to the Safe Transportation Research and Education Center at the University of California, Berkeley, traffic and accident data in the Transportation Injury Mapping System (TIMS) site takes about 12 to 18 months to be input by the California Highway Patrol. Traffic data from TIMS isn't finalized for up to 2 to 3 years, making 2017 the most current data. Also, in this study, we decided to focus on mainland Los Angeles County and did not include the Channel Islands due to low incident reports

Our research focused on answering three specific questions:

- Where are accidents on state highways in California clustered?
- What is the most common collision type in LA county municipalities?
- What impact, if any, does traffic volume have on accidents on the state highway?

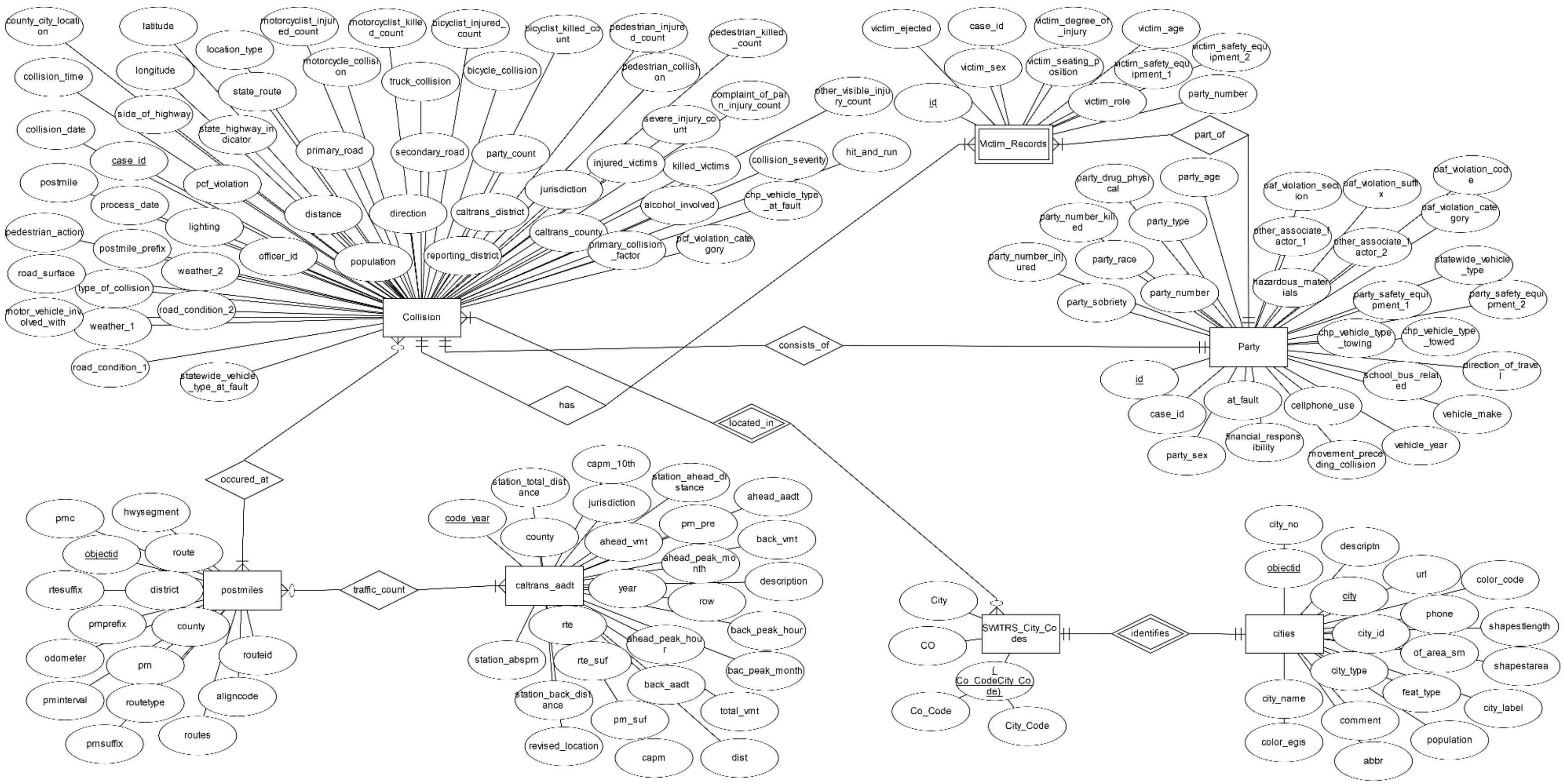
To address the first question, we used ArcGIS Pro's GeoAnalytics tools to determine clustering. For the second question, we examined most common collisions by collision type in municipal areas compared to unincorporated areas. On the third question, we looked at traffic volume on state highways to understand whether accidents were more likely to happen in congested areas.

Knowing the answers to these questions could be used critically in future research on transportation safety. Following growth patterns from this seven-year lapse could predict how transportation patterns are evolving and help determine where improvements are needed to improve safety.

Multiple datasets were gathered from Caltrans PEMS, California Highway Patrol SWITRS, Caltrans Traffic Census and Highway Performance Monitoring Systems (HPMS).

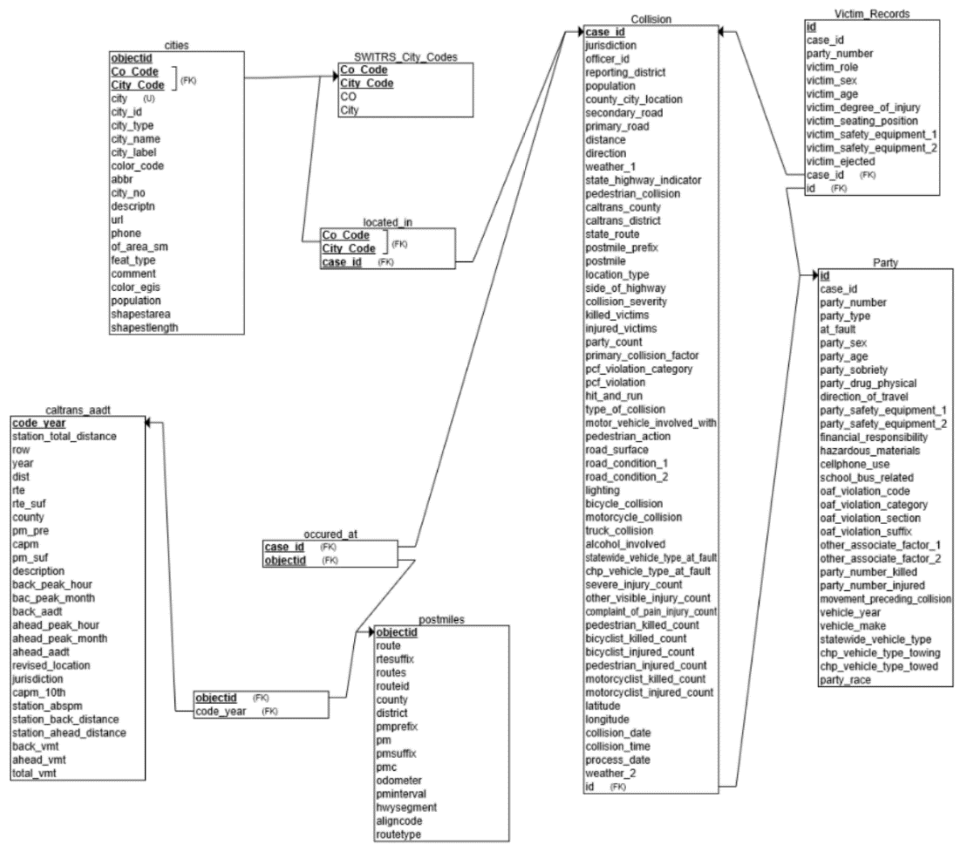
Conceptual Design: ER Diagram

The main entity needed for this project was the Collision entity. Other entities like Victim Records and Party are meant to supplement the data for individuals involved in the Collision. Entities like SWITRS_City_Codes and cities were needed to link the Collision entity geographically. The Postmiles entity was used to link traffic counts in Caltrans_AADT entity to the collision record through the postmile location recorded.



Logical Schema Diagram

Using our ER Diagram, we formed our Logical Schema Diagram which represents relations between entities. To relate query language between our entities, typically, there must be a primary key that's a unique attribute data in a parent table. This primary key attribute is used as a relationship in a child table by a foreign key, which is a field that links data from the parent table to the child table. For example, in the entity Party, the primary key is the underlined attribute "id". This primary key is linked to the Collision entity by a foreign key, indicated by FK. This allows a database user to query the "party_type" of parties the Party entity in connection with the "jurisdiction" of collisions in the Collision entity by using the "id" attribute in the Collision and Party entity.



Data Processing

To answer our research questions, datasets from California Highway Patrol SWITRS, Caltrans Traffic Census, LA County GIS Hub and Caltrans GIS Hub were used.

2010 and 2017 accident data were downloaded from SWITRS with .txt files for collisions, party records, and victims. 2010 and 2017 AADT data was downloaded in CSV format from the Caltrans Traffic Census website, while city and highway shapefiles were downloaded from LA County GIS Hub and Caltrans GIS Hub respectively. To process these datasets, we imported data with the \copy command after creating the table. The shapefiles were uploaded using the PostGIS 2.0 Shapefile and DBF Exporter.

We largely uploaded the columns in varchar datatype and then fixed the data types after deciding which columns we needed. For instance, some entries for integers or decimals had rogue spaces or character types that prevented initial upload. To fix this in pgAdmin, we used the following function to remove spaces and non-numeric characters. In the future, we could improve this process by ideally cleaning the data before uploading it through Python or another coding language.

```
--clean data: remove non-numeric characters from postmile column in
aadt_2010 table
ALTER TABLE "2010_aadt" ADD column pm_cleaned NUMERIC;
UPDATE "2010_aadt" SET pm_cleaned =
      ROUND((regexp_replace(postmile, '[^0-9/.]+', '', 'g')::NUMERIC),1)
```

Our data collection was expansive, so we needed to narrow it down to smaller datasets. Using SQL Queries in pgAdmin, we were able to do this. For example, in the Collision dataset, traffic incidents were tracked by date, starting on 2010-01-01 and continue through the years until 2012-12-31. Since we were focusing on the maximum growth in our study, we only needed data collected in 2010 and 2017. Two queries were run to create two new tables, see below, for 2010 and 2017 respectively.

```
CREATE TABLE twentyten AS
(SELECT * FROM public.collision
WHERE collision_date >= '2010-01-01' AND collision_date <= '2010-12-31');

CREATE TABLE seventeen AS
(SELECT * FROM public.collision
WHERE collision_date >= '2017-01-01' AND collision_date <= '2017-12-31');
```

The first topic researched was based on the question, “Where are accidents on state highways in California clustered?”. For this question, we connected ArcGIS Pro to the database in pgAdmin and added four layers into a new map, the two newly created data tables ‘twentyten’ and ‘seventeen’, the data table ‘cities’ that created the boundary of Los Angeles County, and the ‘postmiles’ layer, which marks mile markers on state highways. Once again there is a surplus of information so we clipped the ‘twentyten’ and ‘seventeen’ layers individually in the *Input Features* section and used the ‘cities’ layer in the *Clip Features* section to create two new layers that follow the extent of the county. Now that the data is contained by year and geography, we can use GeoAnalytics tools to analyze it.

It is important to note the difference in population growth and the collision incident growth of Los Angeles County in this seven-year span. Thanks to the 2010 Census done by the US Census Bureau, we know the population of the county in 2010 was 9,818,605. The number of traffic incidents reported is 32,643. The next census being 2020 results in the US Census Bureau giving an estimated population for 2017 of 10,100,000, and the number incidents reported is 50,614. The population in 2017 would have had to be 15,224,056 in order to account for the raise in incidents, so it should be assumed that outside factors played a part in the raise in incidents, one possibility could be the heightened use of technology while driving.

The Find Hot Spots (GeoAnalytics) tool in ArcGIS Pro identifies statistically significant hot spots and cold spots of input data using the Getis-Ord Gi* statistic. This analyzes each point by its neighboring points based on the Bin distance and Neighborhood distance you set. To cover the county, and with a little trial and error, I set the Bin distance to 2000ft and the neighborhood distance to 4500ft. 2010 and 2017 had very similar results, showing hot spots at many of the on ramps throughout the highways. (See figure 1 below.) The densest hot spot in 2010 is Downtown Los Angeles. The 2017 map shows a spread in this dense downtown area and adds a second dense location around the Santa Monica Airport and West Los Angeles.

To better understand these incident clusters, we ran the Aggregate Points (GeoAnalytics) tool in ArcGIS Pro, as well. This tool also uses bins to aggregate points based on density. For symmetry, we used 2000ft as the bin distance and aggregated into hexagons instead of squares. The results of this echoed the Hot-Spot results. (See figure 2 below.) It similarly shows the highest congestion around Downtown Los Angeles, and strong congestion throughout on ramps. 2017 echoes the growth in congestion, as well.

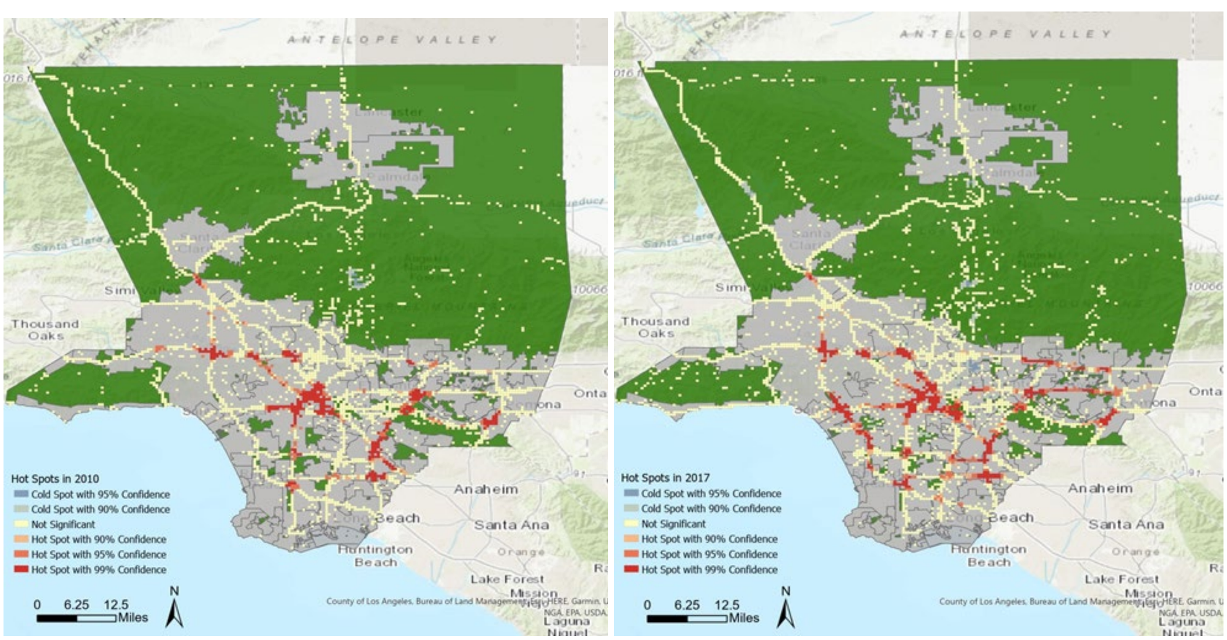


Figure 1: 2010 and 2017 Hot Spot Results

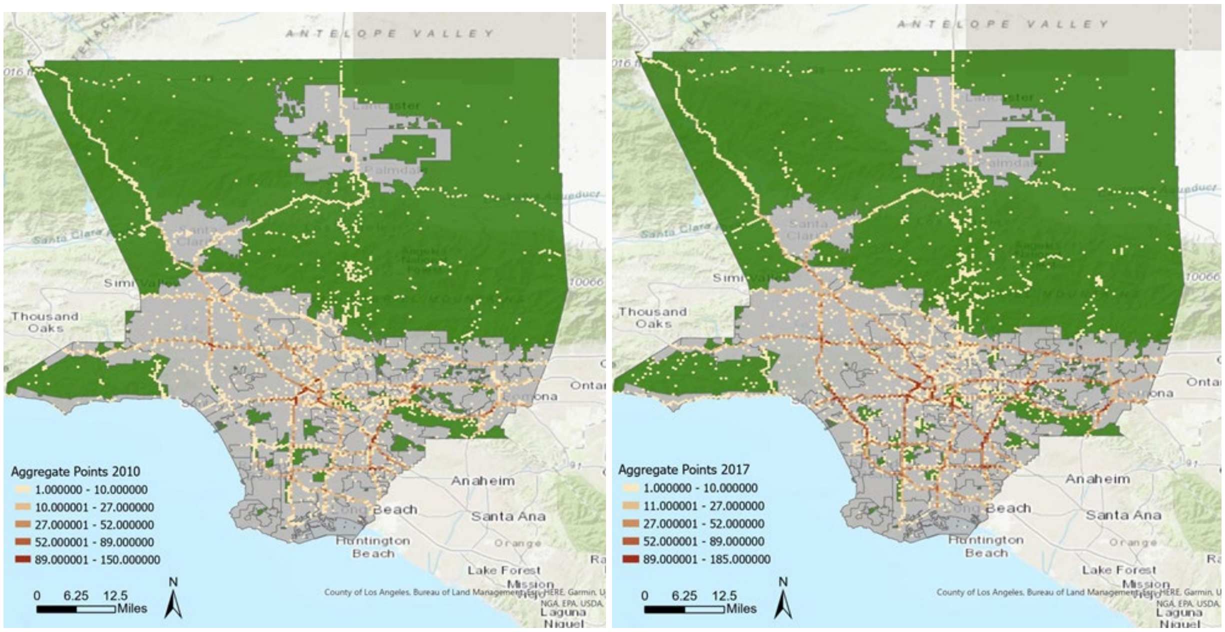


Figure 2: 2010 and 2017 Point Aggregate Results

Search Query Language was used to answer our second research question and is displayed below. The four accident attribute types were included to yield a table summarizing accidents by each type for 2017 in municipal areas. The year search includes a wild card for specific dates that are connected to the year 2017. To obtain 2010 data results, change 2017 to 2010. Accident type attributes were considered a domain with values of 1 and 0. 1 indicates that the accident was that type of accident, and A 0 indicates that the data isn't attributed to that single accident type incident. An inner join was used between the collisions and cities table so that the relation of accident types could be related to municipal or unincorporated areas. The attribute city name in the cities table is currently set to not equal values equal to unincorporated, which retrieves all municipal areas from the cities table. Remove the exclamation point to retrieve all unincorporated areas from the cities table.

Values retrieved by this query should yield values with a location. If values don't have a location, the longitude and latitude fields of the table were empty. This mainly relates to the accident by type locations as all unincorporated and urban areas have a location within Los Angeles County.

```
SELECT SUM((m.motorcycle_collision)::NUMERIC) AS motorcycle_collision,
SUM((m.truck_collision)::NUMERIC) AS truck_collision,
SUM((m.bicycle_collision)::NUMERIC) AS bicycle_collision,
SUM((m.pedestrian_collision)::NUMERIC) AS pedestrian_collision
FROM collision AS m INNER JOIN cities AS c ON ST_Within(m.geom, c.geom)
WHERE (m.collision_date LIKE '2017%' AND c.city_name != 'Unincorporated') AND (m.
motorcycle_collision = '1' OR
m.truck_collision = '1' OR m.bicycle_collision = '1' OR m.pedestrian_collision =
'1');
```

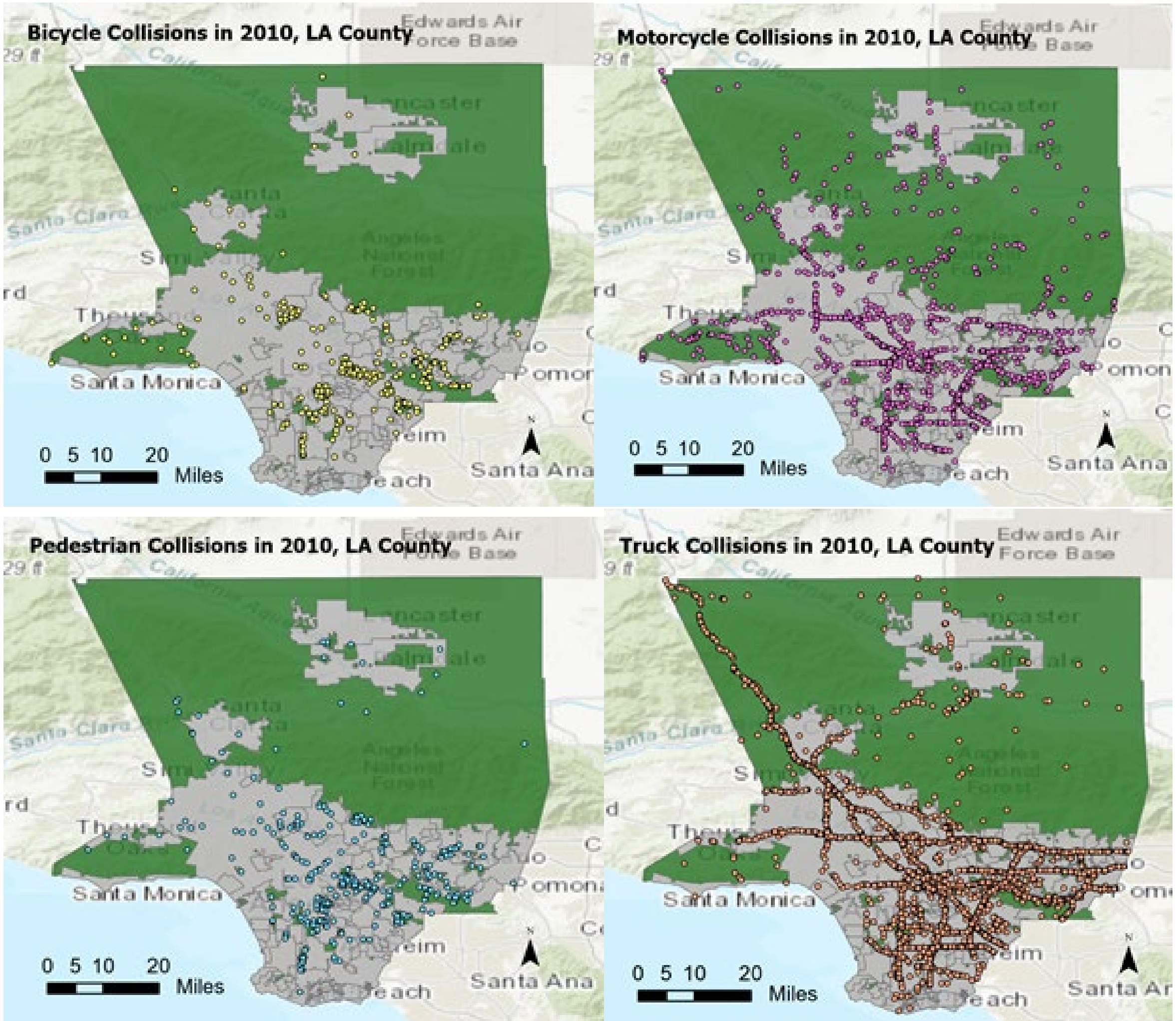


Figure 3: Bicycle, Motorcycle, Pedestrian, and Truck Involved Vehicle Collisions in 2010, Los Angeles County

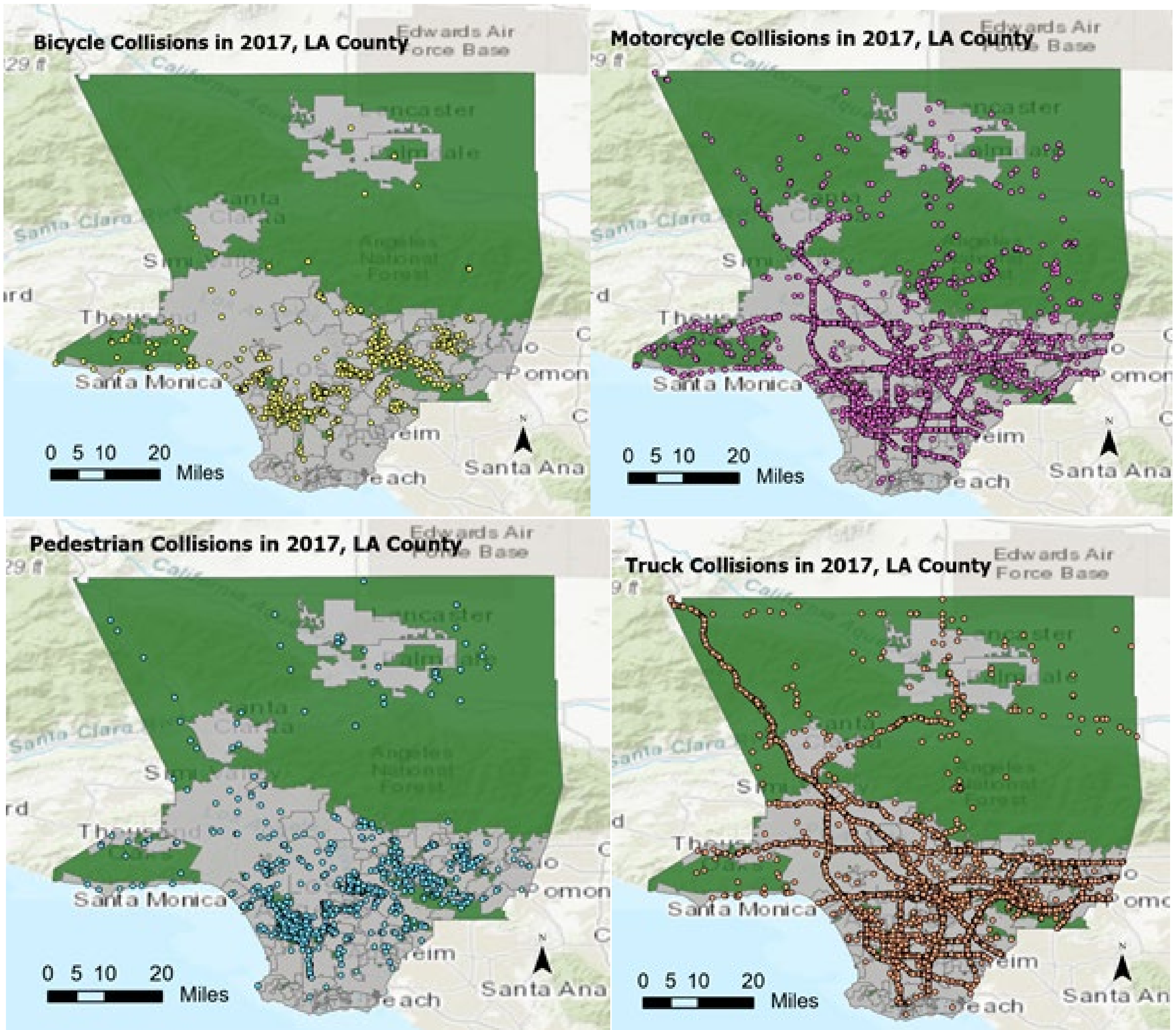


Figure 4: Bicycle, Motorcycle, Pedestrian, and Truck Involved Vehicle Collisions in 2017, Los Angeles

Both figures above are the results of vehicle collisions, by type, which we examined based on our second research question. These collision types occurred with a vehicle, but also involved bicycles, motorcycles, pedestrians or trucks. Individual points in Figures 3 and 4 represent individual collisions. The polygon of Los Angeles County is split up into two colors, green representing unincorporated areas and gray representing municipal areas. Our figure results display that the most common vehicle collision type in both 2010 and 2017 was vehicle collisions involved with trucks. In both 2010 and 2017, the second most common vehicle collision type was vehicle collisions involved with motorcycles. In both 2010 and 2017, the third most common vehicle collision type was vehicle collisions involved with pedestrians. In both 2010 and 2017, the least common vehicle collision type, in our study, was vehicle collisions involved with bicycles. Figures 3 and 4 also display that there were more accidents by each type in 2017 than in 2010, which makes sense with the population increases as noted in our introduction. To answer the second research question though, we examine the figures below.

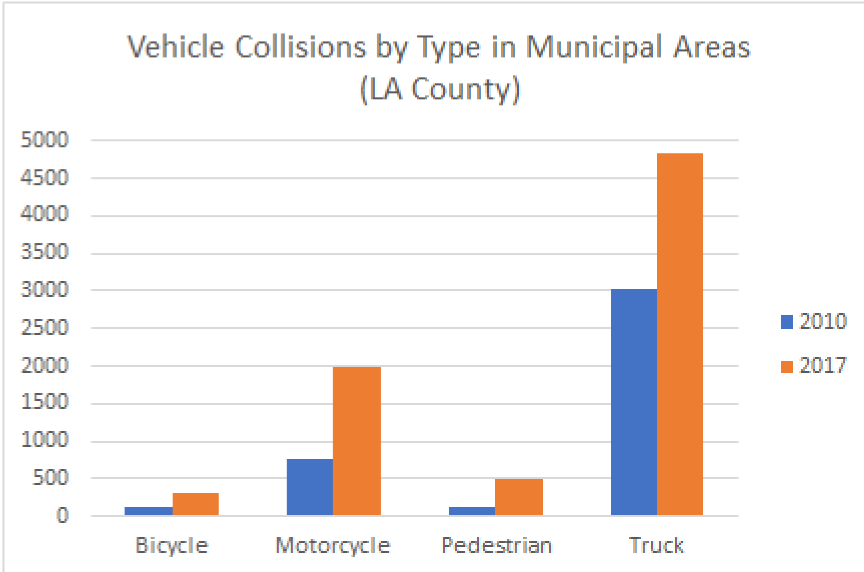


Figure 5: 2010 and 2017 Vehicle Collisions by Type in Los Angeles County Municipal Areas

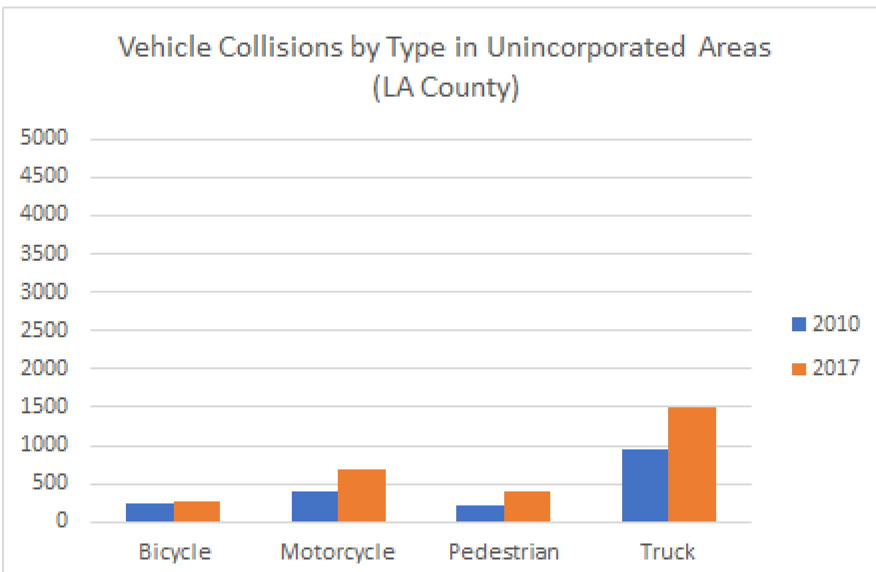


Figure 6: 2010 and 2017 Vehicle Collisions by Type in Los Angeles County Unincorporated Areas

Above, in Figures 5 and 6, collisions by type are examined in municipal areas compared to unincorporated areas, while also considering the year comparison from 2010 to 2017. The results indicate vehicle collisions by type occur more commonly in municipal areas compared to unincorporated areas. The answer to our research question though, is that the most common vehicle collision by type in municipal and unincorporated areas, and for both years examined, was truck collisions. An interesting result for the 2010 municipal data compared to the 2010 unincorporated data, is that there were more pedestrian and bicycle collisions in unincorporated areas than municipal areas. Another trend that is pointed out in Figures 3 and 4 but is much more drastic in Figures 5 and 6, is that collisions by type almost double from 2010 to 2017 across all collision types and areas examined. The only data that didn't significantly increase was bicycle collisions in unincorporated areas from 2010 to 2017.

To understand traffic volume and accidents, the common denominator between all the datasets was the Caltrans postmile for each data point. We started by joining the traffic volume tables to the postmiles table to determine the point geometry for the traffic volumes. Because some postmiles repeated, we also joined it on the route number to ensure the postmiles were linked correctly.

```
--inner join 2010 traffic volume data with postmiles data to get point geom
SELECT p.route, p.county, p.pmc, aadt_2010.descriptio,
aadt_2010.ahead_pe_1 AS peak_month, aadt_2010.ahead_aadt AS ann_aadt, p.geom
FROM postmiles p
JOIN "2010_aadt" aadt_2010
        ON p.pmc = aadt_2010.postmile
        AND p.route = aadt_2010.route
WHERE p.county = 'LA';

--inner join 2017 traffic volume data with postmiles data to get point geom
SELECT p.route, p.county, p.pmc, aadt_2017.descriptio,
aadt_2010.ahead_pe_1 AS peak_month, aadt_2017.ahead_aadt AS ann_aadt, p.geom
FROM postmiles p
JOIN "2017_aadt" aadt_2017
        ON p.pmc = aadt_2017.postmile
        AND p.route = aadt_2017.route
WHERE p.county = 'LA';
```

After cleaning the postmiles file again to ensure the correct order of the points, we used ST_MakeLine with subqueries to create linestrings from the points in the correct order of the postmiles by route. We had to clean up some lines in the downtown area where the postmile order in the data was wrong. Then we used "Join by nearest attribute" in QGIS to assign the points to the links of the lines segment after splitting the lines by the postmile points. The maps below in Figure 7 display traffic volume by postmile links by the tenth decimal.

```
SELECT p.route, ST_MakeLine(p.the_point) AS segment,
        MIN(pm) AS pmstart, MAX(pm) AS pmend
FROM (SELECT route, pm, aligncode, ST_centroid(geom) AS the_point
        FROM postmiles WHERE county = 'LA' AND aligncode = 'Left' ORDER BY route,
pm ) p
GROUP BY p.route
```

To get the table for 2010 accidents:

```
CREATE MATERIALIZED VIEW collision_2017
AS
SELECT
        case_id, year, collision_date, collision_time, postmile, pm_tenth,
        location_type, state_highway_indicator, latitude, longitude, geom
FROM    collision c
JOIN    switrs_codes sc
```

```
ON      c.county_city_location = sc.Co_CodeCity_Code
WHERE   year = 2017 AND state_highway_indicator='1' AND sc.co = 'LA'
        AND geom IS NOT NULL;
```

To get the table for 2017 accidents:

```
CREATE MATERIALIZED VIEW collision_2010
AS
SELECT
        case_id, year, collision_date, collision_time, postmile, pm_tenth,
        location_type, state_highway_indicator, latitude, longitude, geom
FROM    collision c
JOIN    switrs_codes sc
ON      c.county_city_location = sc.Co_CodeCity_Code
WHERE   year = 2010 AND state_highway_indicator='1' AND sc.co = 'LA'
        AND geom IS NOT NULL;
```

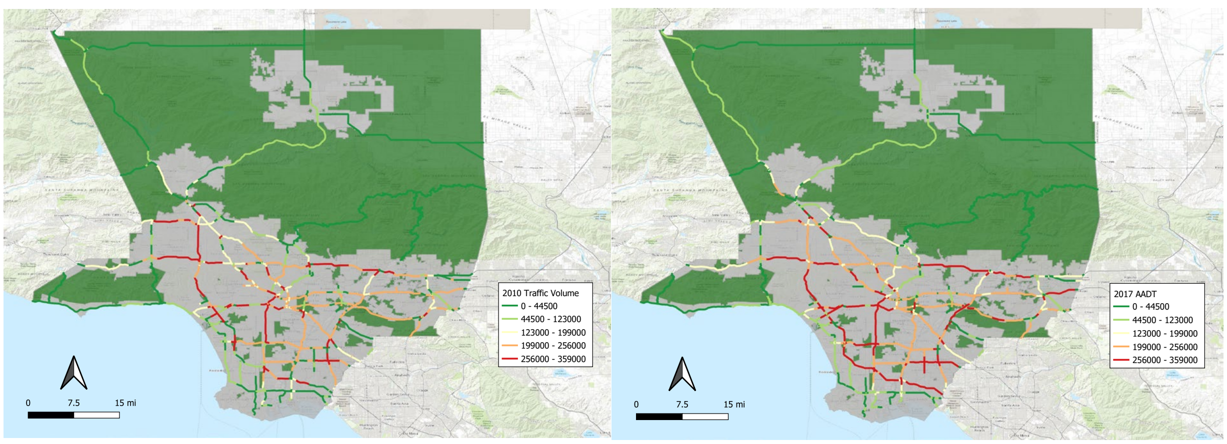


Figure 7: 2010 and 2017 Traffic Volume

Traffic volume increased in 2017 compared to 2010.

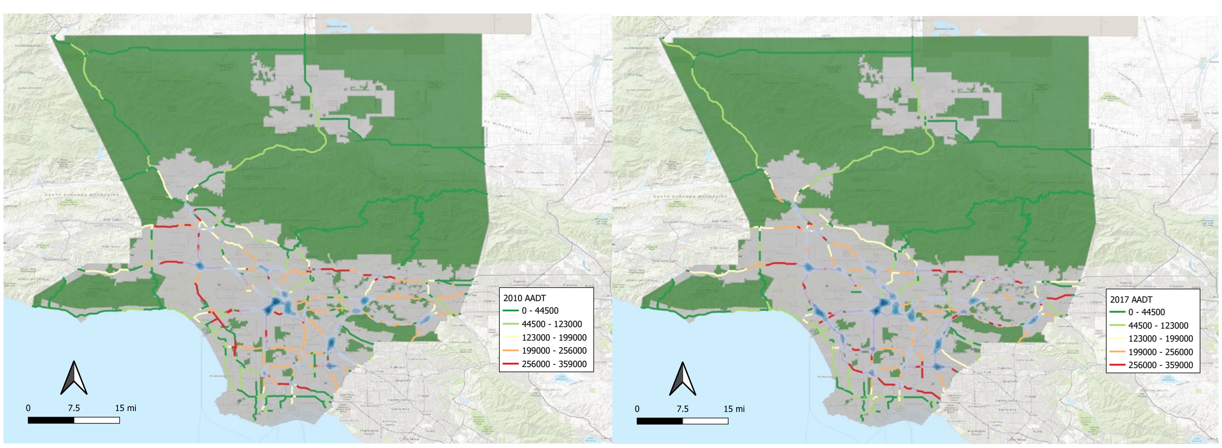


Figure 8: 2010 and 2017 Accidents with Traffic Volume

Traffic volume was classified by natural breaks, while collisions were displayed by overlaying a heatmap. A side by side comparison allows us to see where clusters of accidents occur in relation to the traffic volume between 2010 and 2017.

Discussion

A major takeaway from answering the research question on vehicle collisions by type in municipalities versus unincorporated areas is population influences on accidents. There seems to be a correlation between population increase and accident by type increase across Los Angeles County. What's so curious about the results though, is how the numbers of each type of collision almost doubled from 2010 to 2017. This substantial increase is something that we cannot answer with our research but is significant to point out. A future study should be done on why collisions by type where so much more common in 2017 compared to 2010.

Another interesting trend noted in the results is that in 2010, bicycle and pedestrian accidents were more common in unincorporated areas compared to municipal areas. All other accident types, regardless of the years we examined, were more common in municipal areas compared to unincorporated areas. We also don't have any reasonings behind these results but would like to point out that municipal areas seem have accidents occurring more often than unincorporated areas. Due to these results, we recommend that separate studies be done to improve transportation safety in cities.

In another study from 1988, revised in 1989, from Genevieve Giuliano, titled, *Incident Characteristics, Frequency, And Duration on a High Volume Urban Freeway*, she studied how incident reports effect travel, and how traffic accidents are low proportion of all incidents, yet take a high share of all incident duration. She also used the TASAS database, focusing on a 12-mile section of the I-10 Freeway going through Los Angeles. Though, since it was 1988, she had a hard time finding data, and used California Highway Patrol dispatch logs. She was able to collect enough data to study 607 days, and compared; weekday versus weekend, weather, sunlight, as well as incident report type; accidents, disablements and other. Her conclusions are left slightly open-ended, but her report was meant to start a discussion, which she was successful in doing. The research we conduct here could be a start to continuing her study.

From the data available, we can conclude that there is a trend of traffic accidents increasing as traffic volume has increased. While there are apparent accident clusters around high traffic volume areas, the frequency of accidents cannot be solely attributed to traffic volume. For instance, accidents are most frequent in the downtown area of Los Angeles County, where there are multiple interchanges and other factors that could lead to accidents. Between 2010 and 2017, traffic volume has significantly increased. This is likely due to population growth and urban sprawl. Traffic accidents have also increased in 2017 where traffic volume has increased, as clusters appear where traffic volume segments have become red.

It can be difficult to compare traffic volume in Los Angeles County where extreme congestion occurs on almost every major highway. Traffic volume and accident frequency around downtown LA can be so high that it can make volume and accident frequency on other congested corridors seem low.

Conclusion

Overall, traffic accidents of all different modes have increased from 2010 to 2017. Traffic volume has also seen an increase from 2010 to 2017. This is likely due to the increase in population, but there could be other factors involved as infrastructure changes. For instance, Los Angeles has multiple industry centers that are spread throughout the county. Job and population growth would lead to changing travel patterns that could result in more accidents and traffic volume. Methodology in collection of data also changes over time and with technology. Traffic volume data is a great example of this methodology change in estimating amount of traffic on a highway at any given time. Changes to municipal and unincorporated boundaries are also key considerations as the buying and selling of land alters boundaries, changing data aggregation results.

With COVID-19, travel demand has significantly decreased in Los Angeles County and it remains to be seen if traffic levels will return to the same levels before the pandemic. There may be drastically reduced traffic accidents as traffic volume has decreased. Teleworking policies will likely remain in some capacity as remote work has increased, meaning traffic patterns will be permanently impacted. On the other hand, bicycle use has soared and whether this has impact on the frequency of bicycle accident is still to be determined. Any future studies on traffic accidents and volumes of various modes should consider the various factors impacting travel patterns after COVID-19.

Throughout the process, we found that validating the integrity of our datasets was essential for our analysis. Statewide datasets should be assessed for quality assurance and quality control. For example, some columns for 2017 collision data were not filled to the extent that 2010 collision data was. In addition, not all the collisions had lat long coordinates or postmiles recorded. While this did not significantly impact our analysis, future studies should note that dataset from more recent years may still be changing as accident cases are settled. While our original project proposal intended on researching impacts of COVID-19, there is not enough accurate data for 2020 accidents. More accurate research on 2020 traffic patterns and accidents will likely not occur for a couple years as datasets are validated.

References

County Of Los Angeles Enterprise GIS. (2020). County of Los Angeles. https://egis-lacounty.hub.arcgis.com/

Giuliano, G. (1989). Incident characteristics, frequency, and duration on a high volume urban freeway. *Transportation Research Part A: General*, *23*(5), 387–396. https://doi.org/10.1016/0191-2607(89)90086-1

Traffic Census Program | Caltrans. (2020). CA.Gov. https://dot.ca.gov/programs/traffic-operations/census

Transportation Injury Mapping System. (2020). UC Berkeley. https://tims.berkeley.edu/help/Query_and_Map.php

U.S. Census Bureau QuickFacts: Los Angeles County, California. (2020). Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/losangelescountycalifornia/POP010210#POP010210