A Study on Pedestrian and Bicycle Crashes near Schools of Chula Vista

GEOG 580 Final Project

Written by Bita Etaati

San Diego State University

1. Abstract

Active school transportation (walking or cycling as the means of transportation) is highly associated with providing numerous physical and mental health benefits for students. Given the rising rate of obesity in children, walking and cycling to school can provide their daily recommended dose of exercise. In addition, the affordability of active transportation makes it a perfect way to transport in less socioeconomically advantaged neighborhoods. However, the active school traveling trend is drastically decreasing, impacted by different factors, including traffic safety. Traffic accidents are the fourth most common cause of death in the US. Hence, identifying safe routes for walking and cycling can be a crucial mechanism for enhancing active transportation in students.
Safe routes to school (SR2S) is a federally funded program that promotes active transportation to school through 6 E's (Engagement, Education, Encouragement, Engineering, Evaluation, and Equity). The city of Chula Vista, being the second most populated city in San Diego County and home to almost 270,000 people, 25% of whom are children under eighteen, has been an active participant in this project since 2007. Focusing on 2-5 schools in this city, this project will investigate the impact of the SR2S program and propose a hotspot analysis to implement for future improvements. This project aims to identify the contributing factors to bicycle and pedestrian accidents in the city of Chula Vista and and schools located in this city.

2. Background and Problem Statement

Although the Safe Routes to School program advocates active transportation for children K-12 to ensure achieving its physical and health benefits of activity for students, a question arising in this project is whether streets are safe enough for active school travel. Since Chula Vista has been participating in this program since 2007 and has held numerous educational activities throughout the years, and given the high percentage of under 18 children population in the city, this city was chosen for pedestrian and bicycle accident analysis to investigate the impact of contributing factors leading to these accidents. Also, the higher rate of child obesity (20%) of the city compared to average San Diego county (16.4%) is another reason for the importance of the SR2S project in this city.
This project will focus on identifying schools located in higher accident risk areas through buffer analysis. In addition, by defining different indicators, this project aims to identify the impact of each of these factors in occurrence of the crashes.

3. Research Questions:

This project will focus to provide answer to the following questions:

1. Is Chula Vista Safe for children to ride/cycle in general?
2. What are the contributing factors leading to these accidents?
3. Where and when accidents happen in this city?
4. Are any of the schools located in high risk accident spots?

4. Data and Tools Used in This Project:

Accidents data used in this project is based on police reports collected from [Transportation Injury Mapping System (TIMS) website]() and indicates all bicycle, pedestrian accidents happened in this Chula Vista from 2010 to 2021. This website provides 3 tables including: crashes, parties and victims for all the accidents. All the 3 tables were used for the means of this project. In addition to PostGres and PostGIS, ArcGIS online was used for the visualization purposes.

1. Crashes table:

This table has 80 columns, many of which containing missing values. Case_ID, which is the unique identifier for each accident report, is chosen as the primary key in this project.

Crashes table columns without major missing values:

Table "final.crashes"

Column	Type	Collation	Nullable	Default	Storage	Stats target	Description
case_id	text		not null		extended		
accident_year	text				extended		
proc_date	timestamp without time zone				plain		
collision_date	timestamp without time zone				plain		
collision_time	integer				plain		
officer_id	text				extended		
day_of_week	numeric				main		
special_cond	numeric				main		
primary_rd	text				extended		
intersection	text				extended		
weather_1	text				extended		
collision_severity	numeric				main		
ped_action	text				extended		
road_surface	text				extended		
road_cond_1	text				extended		
lighting	text				extended		
alcohol_involved	text				extended		
latitude	numeric				main		
longitude	numeric				main		
county	text				extended		
city	text				extended		
point_x	double precision				plain		
point_y	double precision				plain		

Picture 1 - Crash table data types

2. Parties Table:
This table contains personal and vehicle information of all parties involved in the accident. Case_id is the foreign key for this table:

Table "final.parties"

Column	Type	Collation	Nullable	Default	Storage	Stats target	Description
case_id	text				extended		
party_number	numeric				main		
party_type	text				extended		
at_fault	text				extended		
party_sex	text				extended		
party_age	numeric				main		
party_sobriety	text				extended		
vehicle_year	numeric				main		
vehicle_make	text				extended		
stwd_vehicle_type	text				extended		
race	text				extended		
inattention	text				extended		
accident_year	numeric				main		

Picture 2- Parties table data types

3. Victims Table:
This table shows information of the accident victims. Case_id is foreign key for this table:

Table "final.victims"

Column	Type	Collation	Nullable	Default	Storage	Stats target	Description
case_id	text				extended		
party_number	numeric				main		
victim_number	numeric				main		
victim_role	numeric				main		
victim_sex	text				extended		
victim_age	numeric				main		
victim_degree_of_injury	numeric				main		
county	text				extended		
city	text				extended		
accident_year	numeric				main		

Picture 3- Victims table data types

4. Schools Table:
This table shows the School name, latitude and longitude of all Chula Vista schools which have participated in the SR2S program.

5.SANDAG Shapefiles: In addition to the TIMS dataset, several shapefiles has been implemented in this study including:

- Bike Routes
- Street Lights
- Intersections
- Streets

5. Methodology:

- **Importing shapefiles and creating tables:**

A new schema (Final) under class server and my personal database (Etaati) was created. All 4 tables were imported into postgres with the create table command under the Final Schema. After importing the tables, primary and foreign keys were defined as below:

```
alter table final.crashes add primary key (case_id);
alter table final.parties add foreign key (case_id) REFERENCES final.crashes(case_id);
alter table final.victims add foreign key (case_id) REFERENCES final.crashes(case_id);
```

Picture 4- Adding primary and foreign keys

Shp2pgsql command was used to import Shapefiles into Postgis:

```
shp2pgsql -s 2230 /Users/bitaetaati/Desktop/ShapeFile/BIKE_ROUTES/BIKE_ROUTES.shp final.bike_routes | /Library/PostgreSQL/13/bin/psql -h 130.191.118.187 -U etaati_admin Etaati
shp2pgsql -s 2230 /Users/bitaetaati/Desktop/ShapeFile/street_lights/street_lights.shp final.street_light | /Library/PostgreSQL/13/bin/psql -h 130.191.118.187 -U etaati_admin Etaati
shp2pgsql -s 2230 /Users/bitaetaati/Desktop/ShapeFile/streets/streets.shp final.streets | /Library/PostgreSQL/13/bin/psql -h 130.191.118.187 -U etaati_admin Etaati
shp2pgsql -s 2230 /Users/bitaetaati/Desktop/ShapeFile/Roads_Intersection/Roads_Intersection.shp final.Roads_Intersection | /Library/PostgreSQL/13/bin/psql -h 130.191.118.187 -U etaati_admin
```

Picture 5- Importing shapefiles

After importing all required tables and shapefiles, 5 indexes were created for each shapefile and the crash dataset as per below:

```
create index idx_crashes_geom on final.crashes using gist(geom);
create index idx_bikeroutes_geom on public.bike_routes using gist(geom);
create index idx_streetlights_geom on final.street_light using gist(geom);
create index idx_streets_geom on final.streets using gist(geom);
create index idx_intersection_geom on final.Roads_Intersection using gist(geom);
```

Picture 6 - Creating indexes

Also since TIMS tables and School table only had latitude and longitude information, geom column was created for these tables:

```
ALTER TABLE final.crashes ADD COLUMN geom geometry;
UPDATE final.crashes SET geom = ST_SetSRID(ST_MakePoint(point_x, point_y), 4326);

ALTER TABLE final.schools ADD COLUMN geom geometry;
UPDATE final.schools SET geom = ST_SetSRID(ST_MakePoint(LONGITUDE,latitude), 4326);
```

Picture 7 - Adding geom column to tables

- **Statistical Analysis of Each Columns:**

For non-spatial columns, a statistical analysis was performed using simple sql queries including count(*). These analysis were done for the following columns:

- Crashes Table: Number of accidents, collision severity, accident time and week day frequency, road surface, road condition, weather and lighting
- Victims Table: (number of accident per victim's age, number of accident per victim's gender)

```
SELECT LIGHTING, count(*) ,
case when lighting = 'A' then 'Daylight'
when lighting = 'C' then 'Dark and Street light'
 when lighting = 'B' then 'Dusk - Dawn'
when lighting = 'D' then 'Dark and No Street light' end as light
from final.crashes
group by LIGHTING
order by count desc
```

Picture 8 - Query example of statistical approaches

- **Spatial Analysis of the indicators:**

➢ Number of accidents and collision severity around each school:

To Analyze the number of accidents near each school in Chula Vista, a buffer analysis approach was used. Based on the previous studies on pedestrian and bicycle traffic safety[8][9][10], 3 buffer sizes (0.25 mile, 0,5 mile and 1 mile) were chosen.
For this analysis, the nearest neighbor analysis approach was implemented using PostGIS queries such as ST_BUFFER and ST_INTERSECTS.

```
with cte as(
	select distinct on (c.geom) c.geom, s.geom, s.school_name,
	ST_intersection(st_buffer((st_transform(s.geom, 2230)),5280), (st_transform(c.geom, 2230))) from final.schools s
	 inner join final.crashes c on ST_intersects(st_buffer((st_transform(s.geom, 2230)),5280), (st_transform(c.geom, 2230)))
group by s.school_name, s.geom, c.geom)
select cte.school_name, count(*) from cte
group by cte.school_name
order by count desc
```

Picture 9 - Query example of buffer analysis of schools - 0.5 mile buffer size

Outputs:

1 Mile Buffer

	school_name text	number_of_accident bigint
1	Vista Square Elementary School	146
2	Mueller Charter (Robert L)	117
3	Feaster (Mae L) Charter	113
4	Lauderbach	91
5	Harborside Elementary School	90
6	Rice Elementary School	78
7	Rosebank Elementary School	67
8	Montgomery (John J) Elementary School	58
9	Hilltop Drive Elementary School	50
10	Otay Elementary School	36

0.5 Mile Buffer

	school_name text	number_of_accident bigint
1	Vista Square Elementary School	125
2	Harborside Elementary School	96
3	Feaster (Mae L) Charter	92
4	Mueller Charter (Robert L)	86
5	Rice Elementary School	70
6	Lauderbach	56
7	Montgomery (John J) Elementary School	44
8	Otay Elementary School	29
9	Loma Verde Elementary School	28
10	Hilltop Drive Elementary School	22

0.25 Mile Buffer

	school_name text	number_of_accident bigint
1	Vista Square Elementary School	47
2	Feaster (Mae L) Charter	37
3	Mueller Charter (Robert L)	31
4	Rice Elementary School	19
5	Montgomery (John J) Elementary School	14
6	Lauderbach	12
7	Otay Elementary School	11
8	Loma Verde Elementary School	10
9	Hilltop Drive Elementary School	9
10	Castle Park Elementary School	9

Picture 10 - Query results, number of accident in different buffers around schools

➢ Number of intersections and accidents around each school:

Some previous studies have linked the proximity of intersections to the number of accidents[12][11]. For this purpose using Intersection shapefile, the number of intersections in a buffer of 0.25 mile of each school has been calculated. For achieving this goal ST_DWithin and spatial join queries have been used.

```
with aa as (select i.school_name, count(*) as number_of_accident
from final.intersect i
inner join final.crashes c on ST_DWithin((st_transform(i.geom, 2230)), (st_transform(c.geom, 2230)), 1320)
group by i.school_name, ST_DWithin((st_transform(i.geom, 2230)), (st_transform(c.geom, 2230)), 1320)
order by number_of_accident desc)

select distinct aa.school_name, i.count as number_of_intersection, aa.number_of_accident
from aa inner join final.intersect i
on aa.school_name = i.school_name
order by number_of_intersection desc
```

Picture 11 - Query example of number of intersections and accidents

School Name	Number of Intersections	Number of Accidents
Montgomery (John J) Elementary School	72	16
Valle Lindo Elementary School	65	6
Wolf Canyon Elementary School	60	4
Veterans	59	6
Casillas Elementary School	59	7
Feaster (Mae L) Charter	59	38
Harborside Elementary School	52	4
Otay Elementary School	48	11
Mueller Charter (Robert L)	45	33
Parkview Elementary School	45	4

Picture 12 - Query results, schools surrounded by highest number of intersections in 0,25 buffer

➢ Number of accidents and collision severity relation with speed limit:

Based on StreetWidths website, the average US street width is 50 feet. Considering that ST_Distance and join queries were used on Street shapefile to calculate the total number of accident and also number of severe/fatal accidents per speed limit.

```
with cte as (SELECT distinct on (c.geom) c.geom, s.geom, s.speedlimit
 from final.crashes c, final.streets s
where ST_Distance((ST_transform(c.geom,2230)),s.geom) < 50)
select cte.speedlimit, count(*) from cte
group by cte.speedlimit
```

Picture 13 - Query example, number of accidents per speed limit

Total Accidents

	speed_limit smallint	number_of_accidents bigint
1	25	347
2	30	225
3	35	507
4	40	113
5	45	31
6	50	68
7	60	60

Severe/ Fatal Accidents

	speed_limit smallint	number_of_accidents bigint
1	25	44
2	30	23
3	35	94
4	40	13
5	45	6
6	50	10
7	60	22

Picture 14 - Query results, number of accidents per speed limit

➢ Number of accidents and collision severity relation with bike routes:

```
SELECT route_clas,count(*) from final.crashes c
inner join public.bike_routes b on St_intersects(ST_Buffer(ST_transform(c.geom,2230), 2), b.geom)
where route_clas not like '%Bikeways Coming Soon%'
group by route_clas
```

Picture 14 - Query example, number of accidents per bike route

	route_type character varying (254)	number_of_accidents bigint
1	Bike Route	155
2	Bike Lane	93
3	Other Suggested Routes	4
4	Multi-Use Path	2

Picture 15 - Query results, number of accidents per bike route

➢ Number of accidents and collision severity relation with streetlights:

```
from final.crashes c
where c.case_id not in (SELECT c.case_id from final.crashes c
inner join final.street_light b on ST_intersects((St_buffer(ST_transform(c.geom,2230),75)), b.geom)
where ST_intersects((St_buffer(ST_transform(c.geom,2230),75)), b.geom))
and c.lighting not like 'A'
group by  c.collision_severity
order by count desc
```

Picture 14 - Query example, number of accident at night and street lights

Total accidents

lighting text	count bigint	light text
A	947	Daylight
C	389	Dark and Street light
B	61	Dusk - Dawn
D	45	Dark and No Street light

Collision Severity at night - with street light

	collision_severity text	count bigint
1	Fatal	27
2	Injury (Severe)	64
3	Injury (Other Visible)	131
4	Injury (Complaint of Pain)	125

Collision Severity at night - no street light

collision_severity text	count bigint
Injury (Other Visible)	58
Injury (Complaint of Pain)	44
Injury (Severe)	28
Fatal	22

Picture 15 - Query results number of accident at night and street lights

6. Results and Conclusions:

As previously mentioned in the methodology section, this report intends to identify schools located in high accident risk spots in the city. As per the results of the buffer analysis, the following schools have the highest number of accidents in different buffer sizes:

1. Vista Square Elementary School,
2. Feaster Charter School,
3. Muller Charter
4. Harbor Side

However, no visible association between number of intersections and number of accidents were identified.

Another purpose of this report was to analyze different indicators and their impact on the number of accidents and collision severity. Here is a summary of the findings:

1. Accidents were more frequent on weekdays and working/school hours.
2. 35% of victims were under 18.
3. In 50% of these accidents, pedestrians or cyclists were at fault.
4. Most accidents happen on streets with a speed limit less than 40 m/h, however the number of severe/fatal accidents drastically drops compared to higher speeds.
5. Only 17% of accidents happened at bike routes.
6. Most accidents happen during the daytime, however for night accidents, the ratio of severe/fatal accidents are higher in streets without proper street lights.

7. Project Challenges and Limitations:

To investigate the importance of each attribute in this project, we need to use machine learning approaches (classifiers). Since this project did not include any classification methods, it's not possible to identify the main contributors of accident occurrences in Chula Vista.
Another challenge of the project was the abundance of missing values. Plenty of columns of the TIMS dataset were almost useless due to this issue.
In addition, analyzing shapefiles required a considerable processing time which added to the challenges of the project.

8. References:

1. https://postgis.net/docs/manual-1.5/ch08.html
2. https://postgis.net/workshops/postgis-intro/
3. https://chulavista-cvgis.opendata.arcgis.com
4. https://www.sandag.org/index.asp?subclassid=100&fuseaction=home.subclasshome
5. https://tims.berkeley.edu
6. https://p12cdn4static.sharpschool.com/UserFiles/Servers/Server_411950/File/1-CVESD%20SRTS%20Master%20Plan.pdf
7. https://www.census.gov/quickfacts/chulavistacitycalifornia
8. Abdel-Aty, Mohamed, Sai Srinivas Chundi, and Chris Lee. "Geo-spatial and log-linear analysis of pedestrian and bicyclist crashes involving school-aged children." *Journal of safety research* 38.5 (2007): 571-579.
9. Clifton, Kelly J., and Kandice Kreamer-Fults. "An examination of the environmental attributes associated with pedestrian–vehicular crashes near public schools." Accident Analysis & Prevention 39, no. 4 (2007): 708-715.
10. Hasani, M., Jahangiri, A., Sener, I. N., Munira, S., Owens, J. M., Appleyard, B., ... & Ghanipoor Machiani, S. (2019). Identifying high-risk intersections for walking and bicycling using multiple data sources in the city of San Diego. Journal of advanced transportation, 2019.
11. Wang, Y., & Nihan, N. L. (2004). Estimating the risk of collisions between bicycles and motor vehicles at signalized intersections. Accident Analysis & Prevention, 36(3), 313-321.
12. Zegeer, C. V., Opiela, K. S., & Cynecki, M. J. (1982). Effect of pedestrian signals and signal timing on pedestrian accidents(No. HS-033 770).