CEVP Impact Analysis

Studying the impact of San Jose Fire's Centralized Emergency Vehicle Pre-Emption system on emergency vehicle travel time

Fire Department Data Story

Albert Gehami, Data Scientist

May 24, 2019

Table of Contents

Introduction and Overview

In October of 2018 San Jose introduced an emergency vehicle pre-emption system into their traffic lights.[1] As an emergency vehicle approaches an intersection, the stop light will turn green for the emergency vehicle, and red for the opposing traffic to clear the intersection for the emergency vehicle to pass through. The traffic light pre-emption is designed to get vehicles to emergencies faster. Anecdotally, this system has been a game-changer for emergency vehicles, but there has been no analysis on how much time is saved from the traffic light pre-emption system.

If the traffic light pre-emption system can save even a few seconds, that could translate into additional lives saved. Anupam et al (2017) found that a 4.4-minute increase in emergency vehicle travel time led to a 13.3% jump in 30-day mortality rate for heart attack victims – from 24.9% mortality to 28.2%. Often a fire truck is the first vehicle on the scene for medical emergencies, and serve to stabilize the situation before additional resources arrive. For a home fire, rescue may be impossible 4-5 minutes after the fire began, and only 2-3 minutes after the smoke alarm goes off.[2,3,4]

Numerous factors determine an emergency vehicle's travel time. The distance a vehicle has to travel, the speed limits of the roads, and other factors can drastically affect travel time. Every intersection will lead to an increased travel time, but after accounting for many factors, the traffic light pre-emption system appears to reduce average travel time by 5-7 seconds per intersection (Figure 1). Overall, intersections with CEVP fully implemented seem to add almost no additional time to an emergency vehicle's trip, which is consistent with the anecdotal observations of fire officers in the field.

[1] "San Jose Integrates Emergency Vehicle Pre-Emption with CAD System." *Radio Resource*, 1 Oct. 2018, www.rrmediagroup.com/News/NewsDetails/NewsID/17424.
[2] "Home Fires." Home Fires | Ready.gov, www.ready.gov/home-fires.
[3] "How Quickly Does Fire Spread?" *Disaster Company*, 3 Oct. 2017, www.disastercompany.com/quickly-fire-spread/.
[4] Robert, Crandall. "How Fast Is Fire?" *Fire Event Timeline | Home Fire Drill | Prevention 1st Foundation*, 2005, www.homefiredrill.org/?p=fire-event-timeline.
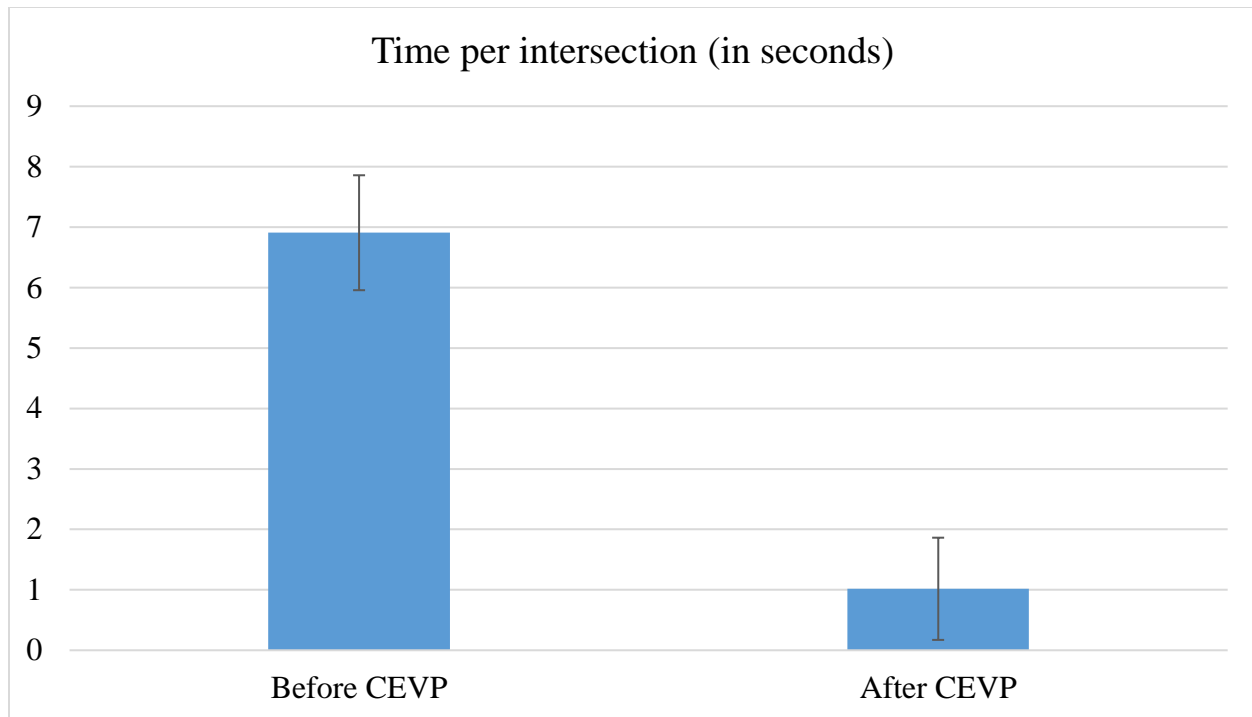
Figure 1: Seconds added to overall travel time for each San Jose-owned intersection before and after CEVP was fully implemented.

One of the more common emergency trips for a fire truck is from the Fire Station near Tully and Senter Rd to the area near Little Orchard and Cimino Street (Figure 2). This trip's distance is 1.5 miles, and includes 6 intersections owned by San Jose. It was taken 455 times as a code 3 (highest emergency) trip from January 1st 2018 to April 31st 2019, with 451 verified start and end times. Before CEVP, this trip would take, on average, 8 minutes and 38 seconds. After CEVP was fully implemented, this trip would take on average 6 minutes and 51 seconds, a reduction of 107 seconds.
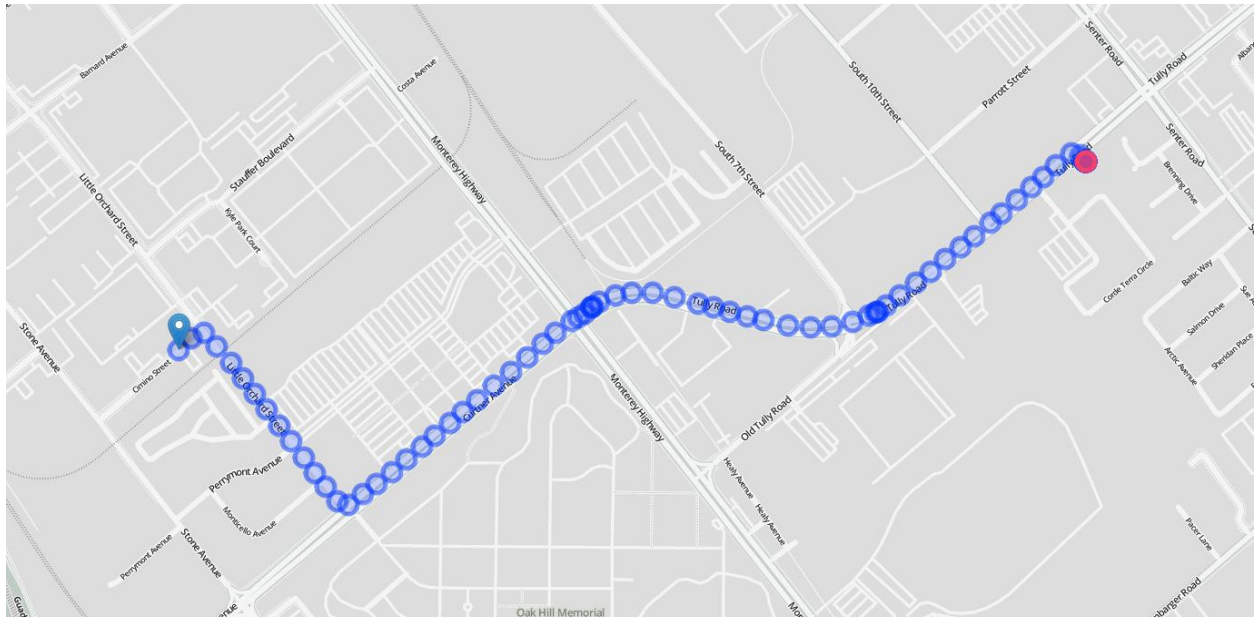
Another finding was the impact of intersections not owned by San Jose on an emergency vehicle's travel time. Intersections not owned by San Jose, such as intersections on County-owned roads, do not have CEVP. Following the implementation of CEVP, County intersections increase a vehicle's average travel time by around 10 seconds per intersection compared to CEVP intersections. It is likely that some of that additional 10 seconds per intersection comes from the lack of CEVP at County intersections.

This report outlines in detail the data science work done to evaluate the impact of CEVP on emergency vehicle travel time post CEVP implementation. Section 2 covers a brief background on the data collected. Section 3 covers the data elements used for analysis, and any calculations involved. Section 4 covers the methodology behind evaluating the impact of CEVP. Section 5 discusses results. Section 6 summarizes the report, explains limitations, and suggests avenues for building on this work.

Background on Data

The Fire department collects rigorous, automated sensor data on their vehicles as they travel to an emergency. An Automated Vehicle Log (AVL) system tracks the vehicle en route to its destination. The system is designed to collect a data point every few seconds or 30 meters traveled. Each data point can be thought of as a "timestamp" as the vehicle is en route, containing information such as the vehicle's current position, the time (hours, minutes, and seconds) and date, the vehicle's ID number, the ID number of the emergency (the "trip ID"), and the vehicle's transit status ("Status", where ER means "En Route" and AR or AD means "Arrived"). With these five pieces of information, it is possible to identify the road the vehicle

is currently traveling on, any intersections it passes, and the vehicle's overall travel time. When the vehicle reaches its destination, the system is supposed to mark its arrival and stop the timer. However, this system sometimes fails to stop the timer, so the data had to be cleaned before analysis. More of the cleaning process will be explained in section 3.

Every trip, which consists of many timestamps, is aggregated into a single data point which tracks a trip's total distance, travel time, and other variables. In short, the data provided looks like Figure 3, and the data used for analysis looks like Figure 4. The data elements in Figure 4 will be explained in section 3. 161,353 trips occurred over the sample period. Following cleaning, 12,546 trips before CEVP were studied, 80,820 trips were studied while CEVP was being installed, and 12,342 trips were studied after CEVP was fully operational.

Vehicle ID	Incident ID	Latitude	Longitude	Status	Date/Time	Time Difference (from last time stamp)
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	0
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	4
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	5
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	5
...	..	...	...	...	...	...
E26	F191199067	37.29798	-121.837	ER	4/29/2019 9:01	5
E26	F191199067	37.29798	-121.837	ER	4/29/2019 9:01	5
E26	F191199067	37.29798	-121.837	AR	4/29/2019 9:01	5

Figure 3: Example of the raw Data from the Automated Vehicle Log (AVL) system, including a calculated variable "Time Difference" which calculates the seconds that passes between the current time stamp and the prior time stamp.

Full ID	E26_F191199067
Start Latitude	37.30641
Start Longitude	-121.84895
End Latitude	37.29798
End Longitude	-121.83731
Travel time	465
Start Date/Time	4/29/2019 8:55
Distance	2135
CEVP Intersections	3
Non CEVP Intersections	1
CEVP Status	after
Average Speed Limit	32.45
Average Number of other Emergency Vehicles on Road	3.24

Figure 4: Example observation after aggregating a trip into a single data point

Data Elements and Cleaning

Once a vehicle is en route to an emergency, the Automatic Vehicle Log (AVL) system automatically begins recording data such as the vehicle's speed, location, and transit status (ER = en route, and AD or AR = arrived). Every few seconds or 30 meters travelled (whichever occurs first), a new recording is made, herein referred to as "timestamps." Through the AVL timestamps an emergency vehicle's entire journey can be tracked. For example, below is the picture of a single emergency vehicle's journey (Figure 5). By overlaying the AVL data with a map of major roads, the average speed limit on an emergency vehicle's journey can also be tracked.
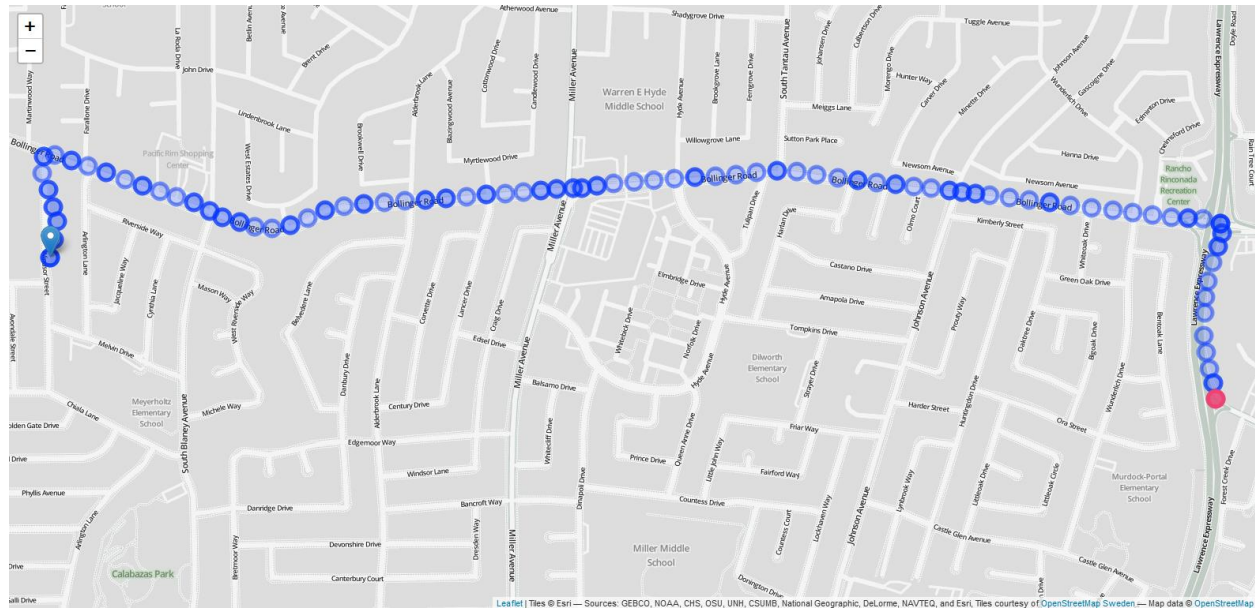

Figure 5: A trip taken by an emergency vehicle. Each dot represents another timestamp along the trip.

This section explains all data elements used in the analysis, along with any cleaning done to the data prior to analysis.

Data Elements

Each data element shown in Figures 3 and 4 above are explained below.

Full ID

This is the unique ID for each observation. It combines the Vehicle ID and Incident ID to create the unit of observation – one vehicle trip.

Start Latitude/Longitude

The starting coordinates for the vehicle. Usually this is the fire station the vehicle is dispatched from.

End Latitude/Longitude

The ending coordinates for the vehicle. Usually the site of the emergency.

Travel Time

The total travel time (in seconds) for the trip. Initially calculated with the AVL data, but ultimately Calculated using the verified Computer Aided Dispatch (CAD) dataset, explained in the Data Cleaning subsection.

Start Date/Time

The date and time the trip began. This information is used to mark if the trip occurred before, during, or after CEVP implementation.

Travel Distance

The approximate distance (in meters) traveled during the trip. This is calculated by summing the straight-line (or Euclidean) distance between each time-stamp. This is not the straight-line distance from the starting coordinates to the ending coordinates, since that would significantly underestimate the total distance traveled.

CEVP/San Jose Intersections

The number of lighted intersections the vehicle passes through that are, as of April 2019, using the CEVP pre-emption technology. Even if a trip is in early 2018, prior to CEVP, intersections that will eventually use CEVP are counted. This includes all networked intersections owned by San Jose, which is all but 8 lighted intersections owned by San Jose.

Non-CEVP/Non-San Jose Intersections

County lighted intersections and other lighted intersections not owned by San Jose do not have the same traffic preemption system that City intersections have. They will not turn green as a fire truck passes through, so this could lead to additional slow-down that would not have happened when passing through a CEVP-enabled City intersection. How much Non-CEVP intersections slow down emergency vehicles is explored later in this paper, and is accounted for as another factor potentially affecting an emergency vehicle's travel time.

CEVP Status

Marks whether this trip began before any CEVP implementation, during CEVP implementation or after CEVP was fully implemented. March 1st, 2018 is used as the date CEVP implementation began. March 1st, 2019 is used as the day CEVP implementation was completed. The trip's CEVP status is marked based on

when the trip begins, so if a trip began on February 28th, 2018, but ended March 1st, 2018, the trip is marked as occurring before CEVP implementation.

Average Speed Limit

For every timestamp, the emergency vehicle is matched to the major road it is likely on, and the speed limit of that road is recorded. The timestamp location data can be slightly off, so if two streets are around 20 meters of each other, there could be more than one potential road the vehicle is on. If there are multiple potential major roads the emergency vehicle could be on, we default to recording the higher speed limit.[5] If the vehicle is not on a major road, then the timestamp's speed limit will be recorded as 25mph, since most residential, non-major streets have a 25mph speed limit.

To get the average speed limit across the entire journey, the speed limits are averaged, weighting each speed limit by the time between each timestamp. Mathematically, this looks like this:

$$\frac{seconds_{timestamp0} * speed\ limit_{timestamp0} + seconds_{timestamp1} * speed\ limit_{timestamp1} + \cdots}{Total\ number\ of\ seconds\ in\ trip}$$

If a journey had three timestamps, like below (Figure 6):

Speed Limit	Date/Time	Time Difference
35	4/29/2019 8:55	4
30	4/29/2019 8:55	4
50	4/29/2019 8:55	2

Figure 6: Example journey

Then the weighted average speed limit would be calculated as follows:

$$\frac{35 * 4 + 30 * 4 + 50 * 2}{4 + 4 + 2} = \frac{360}{10} = 36mph$$

This means that for any second on the vehicle's trip, the average speed limit was 36mph.

Average Number of other Emergency Vehicles on Road

A higher demand for emergency vehicles could affect travel times. To account for peak demand hours, the number of other fire vehicles currently en route are included. Like the average speed limit, the average number of other emergency vehicles on the road is calculated from a weighted average based on the seconds (see Time Difference) at each timestamp.

---

[5] In practice this happens in less than 1% of cases, and has a negligible effect on the data.

Constant

This term will be seen in the regression results table (Table 3). The constant term (usually included in a regression model as β_0) serves as a baseline for the study, from which the other variables will stand on. Its value is not relevant, but excluding the constant would drastically skew the results of the model.

Data Cleaning

Because this data was machine-recorded, it was already mostly clean. Only three actions were conducted to clean the data. The first involved fixing the start time, end time, and travel time variables. The initial log data, which provided location, timestamp, and other travel information, came from a reliable but unverified system. The Automated Vehicle Log (AVL) system provided accurate location data, but had two issues regarding travel time data – 1) Occasionally the log data would continue to record time stamps after arrival (Figure 7) and 2) Occasionally the log would not accurately reflect when the vehicle had arrived (Figure 8). To address these issues, we replace the arrival time, start time, and travel time variables with a verified Computer Aided Dispatch (CAD) dataset with more reliable start, arrival, and travel times.

Status	Date/Time	Time Difference
ER	4/29/2019 8:55	0
ER	4/29/2019 8:55	4
ER	4/29/2019 8:55	5
ER	4/29/2019 8:55	5
...	...	...
ER	4/29/2019 9:01	5
AR	4/29/2019 9:01	5
AR	4/29/2019 9:01	5
AR	4/29/2019 9:01	5

Figure 7: Example Automated Vehicle Log (AVL) data of a trip in which Arrival was marked for multiple time stamps. The accurate time of initial arrival is recorded in the verified CAD dataset.

Vehicle ID	Incident ID	Latitude	Longitude	Status	Date/Time	Time Difference
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	0
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	4
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	5
E26	F191199067	37.30641	-121.849	ER	4/29/2019 8:55	5
...	..	...	...	...	...	...
E26	F191199067	37.29798	-121.837	ER	4/29/2019 9:01	5
E26	F191199067	37.29798	-121.837	ER	4/29/2019 9:01	5
E26	F191199067	37.29798	-121.837	ER	4/29/2019 9:01	5
E26	F191199067	37.29867	-121.837	AR	4/29/2019 9:20	1151

Figure 8: Example of the Automated Vehicle Log system not accurately reflecting arrival status. This particular example should have marked arrival 19 minutes earlier. The accurate arrival time is recorded in the verified CAD dataset.

This analysis only used CAD data with a verified arrival time. Many dispatched vehicles never arrive to an emergency because another emergency vehicle arrived earlier and deemed an additional unit

unnecessary. In these cases, the log data recorded a trip, but the vehicle never arrived, thus never have a CAD-recorded arrival time. 53,931 trip logs never arrived because the situation was under control before that vehicle arrived, so those observations were dropped.

Upon examining the verified CAD times, 3 observations had negative recorded travel times. These trips trip began before daylight savings time, but finished after daylight savings time, effectively arriving an hour earlier than they started. These observations were all recorded within 15 seconds of one another, and account for less than .0003% of the observations, so they were removed as well. Overall, 53,934 observations, or 33% of the overall dataset, were removed.[6]

The second cleaning step was to remove non-existent trips. Occasionally the AVL log would start recording a trip but fail to make timestamps along the journey. The verified CAD data provides accurate start and end times for these trips, but without AVL route information the trip lacks crucial variables such as distance traveled and number of intersections passed. These "route-less trips" accounted for 1,711 observations, or 1.5% of the remaining dataset, and were removed.

Following the first regression, an additional cleaning step was taken to remove outlier observations. Outliers tend to be data points which abnormally affect the model because their numbers are unexpectedly different from the rest of the data. For this study, "outlier" data points are defined as observations with a Cook's Distance 4 times greater than the mean Cook's Distance.[7] Cook's Distance is one of the common ways to mathematically identify outliers in a dataset. 68 observations were deemed outliers, and had a high influence on the model built to study over 100,000 observations. A second regression was conducted removing the 68 outlier observations. Both regressions are shown in the Results section.

Summary statistics on the cleaned data are shown (Tables 1 and 2). Variables are excluded if the data would be irrelevant. For example, the median latitude at which a vehicle begins or ends its trip does not provide any useful information, however the average number of intersections an emergency vehicle passes through in a trip might be useful information.

---

[6] In prior editions of this paper, the non-CAD verified data was used, including these 53,934 observations, and the results were similar.

[7] Information on Cook's distance can be found here: https://newonlinecourses.science.psu.edu/stat501/node/340/

Table 1: Summary Statistics for Select Variables

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Number of other Fire Trucks on Road	0.0000	0.5456	1.2421	1.8518	2.4138	18.7086
Average Speed Limit t	25.00	27.28	30.00	30.41	32.96	55.00
Travel Distance	0	1271	2006	2399	2880	473386
Non-SJ Intersections	0.0000	0.0000	0.0000	0.3496	0.0000	28.0000
CEVP/San Jose Intersections	0.00	2.00	4.00	3.96	5.00	37.00
Travel Time	1.0	333.0	410.0	467.4	506.0	68737.0

Table 2: Number of Trips Before, During, and After CEVP Implemenetation

Before	During	After
12,546	80,820	12,342

Methodology

The analysis uses a panel linear regression which compares the impact to travel time an intersection has before the CEVP preemption system was installed to the impact to travel time an intersection has after CEVP preemption was installed, while controlling for other factors such as average speed limit, distance traveled, and the number of other emergency vehicles on the road.

The panel regression is similar to a normal linear regression, except that it looks at multiple windows, or "panels" of time. In the case of this analysis, there are three panels of time: before CEVP implementation, during, and after full implementation. For each panel of time, the average seconds that a CEVP-marked intersection adds to travel time is calculated. Through this regression, we can make statements such as "Since CEVP was fully implemented, intersections with CEVP add 5-7 seconds fewer to overall travel time compared to what those same intersections added prior to CEVP."

The model for the panel linear regression can be displayed mathematically:

$$Travel\ Time = \beta_0 + \beta_1 * Distance + \beta_2 * Average\ Speed\ Limit + \beta_3 * \#\ of\ Other\ Vehicles + \beta_4$$
$$* \#\ of\ NonCEVP\ Interesections + \beta_5$$
$$* \#\ of\ CEVP\ Intersections\ before\ implementation + \beta_6$$
$$* \#\ of\ CEVP\ Intersections\ during\ implementation + \beta_7$$
$$* \#\ of\ CEVP\ intersections\ after\ implmentation + \varepsilon$$

A further explanation of a linear regression can be found in the appendix. The difference in β_6 and β_8 will be the difference in seconds added to the travel time per intersection before and after CEVP was

implemented. In other words, if β_8 is 4 less than β_6, then our model would suggest that each CEVP intersection saves 4 seconds of travel time compared to our baseline (intersections before CEVP).

With the model defined, the regression can be run and the final results can be displayed.

Results

Below are the results from the regressions. For robustness, we check the results with and without the outlier observations, as defined by having a Cook's distance 4 or more time greater than the mean cook's distance for an observation. See Table 3.

Table 3: Regression Results

	Dependent variable:	
	Travel Time	
	(Outliers Included)	(Outliers Removed)
Travel Distance	0.059***	0.075***
	(0.001)	(0.0005)
Non-SJ Intersections	66.458***	11.201***
	(2.451)	(0.892)
Average Speed Limit	−7.120***	−5.656***
	(0.614)	(0.211)
Number of other Fire Trucks on Road	−0.005	2.325***
	(1.171)	(0.407)
CEVP Intersections BEFORE Implementation	21.974***	6.907***
	(1.651)	(0.578)
CEVP Intersections DURING Implementation	18.002***	5.176***
	(0.848)	(0.308)
CEVP Intersections AFTER Implementation	8.011***	1.017**
	(1.473)	(0.514)
Constant	449.955***	419.206***
	(18.425)	(6.331)
Observations	105,708	105,640
R^2	0.075	0.273
Adjusted R^2	0.075	0.273
Residual Std. Error	742.387 (df = 105700)	254.936 (df = 105632)
F Statistic	1,221.422*** (df = 7; 105700)	5,671.616*** (df = 7; 105632)

Note: *p<0.1; **p<0.05; ***p<0.01

Before implementation, passing through a CEVP intersection added 7-22 seconds to the overall travel time per intersection. After implementation, a CEVP intersection added only 1-8 seconds to the

overall travel time. In other words, emergency vehicles are now saving on average 6-14 seconds per intersection likely because of CEVP. See Figure 9.
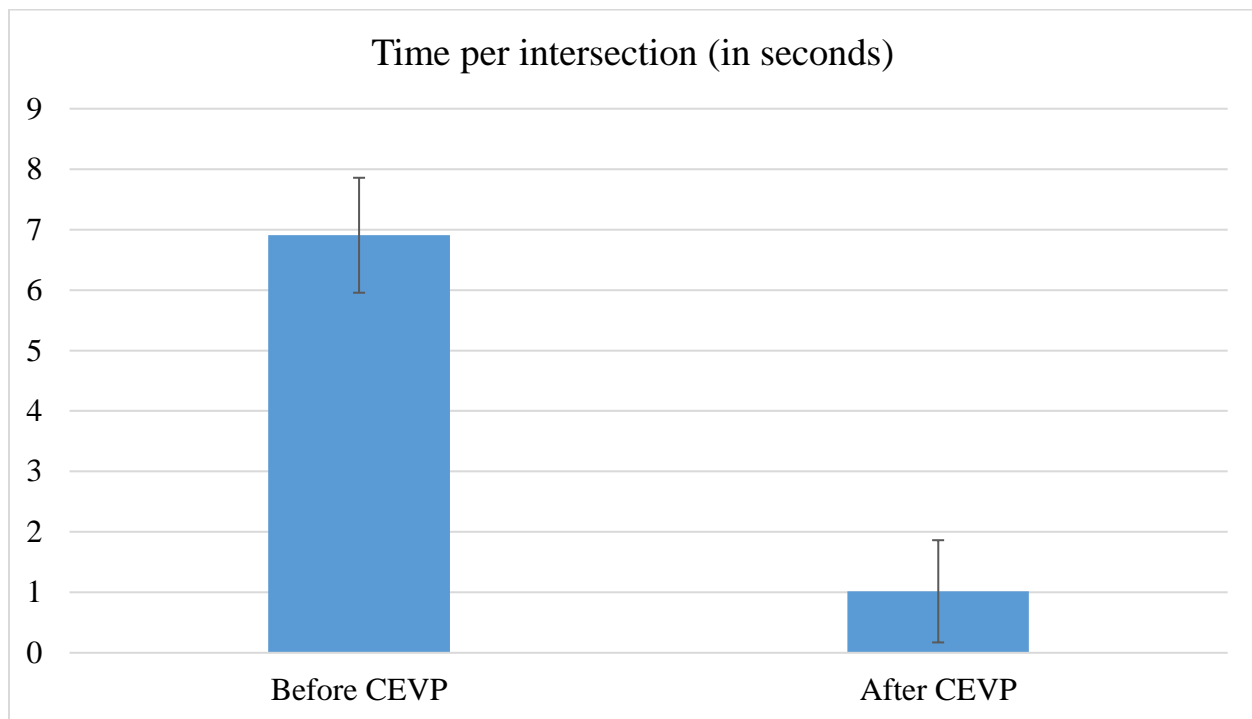

Figure 9: Seconds added to travel time per intersection before and after CEVP implementation. Based on the results after removing outliers.

Interestingly, the number of other Fire vehicles on the road may have a small but significant increase to travel time. While it is unclear why, this may be due to an increased demand for emergency vehicles slowing down the emergency travel.

The final finding of note is the amount of time a non-San Jose intersection (mostly intersections owned by the County) adds to a trip. A non-San Jose intersection adds, excluding outliers, around 11 seconds per intersection, or 10 additional seconds compared to an intersection owned by San Jose. It is likely that some of this additional time from non-San Jose intersections comes from the fact that County intersections do not have CEVP.

Conclusion, Limitations, and Future Work

There is compelling evidence that implementing CEVP reduced average travel times for emergency vehicles. While the analysis required some assumptions to clean the data properly, overall intersections seem to add less time to a trip since CEVP has been implemented.

The biggest limitation to this study is the short window of time to study before and after CEVP. The Automated Vehicle Log (AVL) system did not collect granular route data until the beginning of 2018. Looking at AVL data prior to 2018 will have limited information on the roads a vehicle traveled, and the on the actual distance traveled. Because of this, there will only ever be three months of quality data on trips before CEVP began implementation. Moreover, only 2 months have passed since CEVP's implementation was completed. While the results would likely be similar, it would be best to re-run this study in a year once more data has been collected.

If a secondary analysis confirms these results, then there is one policy suggestion – convince the County to adopt CEVP. From the results, County intersections are adding substantial travel time to emergency vehicles following CEVP implementation. Because of this, we suggest the County explore implementing CEVP or a similar system.

Overall, the CEVP system seems to save, on average, 5-7 seconds in travel time per intersection passed through. With the average emergency vehicle trip passing through 4 intersections, an average of 20-28 seconds is saved per trip because of CEVP. Based on Anupam's work, 24 seconds can lead to a 1-2% decrease in mortality rate for heart attack victims. For trips that cover substantially more intersections, the reduced travel time can save even more lives. Anecdotally, the CEVP system has been reported to reduce overall travel time, and it seems the data reflects this belief.

Appendix: An Introduction to Regressions – No math required

A well-designed regression can respond to criticisms a graph or chart could not. A graph could show that after CEVP was implemented travel times dropped by 10 seconds. Is this an argument that CEVP is working, or that emergency vehicles have had, by chance, shorter trips following CEVP? Another graph could show that vehicles have a lower average speed wherever CEVP intersections exist. Is this an argument that CEVP slows vehicles down, or just that intersections with CEVP tend to also be crowded and with lower speed limits? Without a deeper analysis, arguments would go in circles.

Regression analysis replaces this arguing with answers. Regressions can account for the fact that some emergency trips are shorter in distance than others, and for the fact that some streets have faster traffic than others. Regressions home in on the issue of interest—is CEVP having reducing travel time?

Linear regressions are common for understanding the magnitude of the relationship between two things. They allow researchers to make claims such as "For every additional year of education, a person's average annual income increases by $5,000". Linear regressions make claims about the quantity of an impact. The linear regression models in this study show how passing through an intersection before or after CEVP affects the expected travel time.

This is a panel linear regression because the result we are looking for is not if CEVP intersections reduce travel time in general, but if a fully-implemented CEVP intersection reduces travel time compared to intersections without CEVP. Intersections by their nature can clog streets and slow an emergency vehicle down. The important question is if these intersections slow the vehicle down less once CEVP is implemented. The results for the regression models can be found in the results section.