# Visual Analysis of a Traffic Light Data for COSC 6344

Nabeel Bacchus
*Natural Science and Mathematics Department*
*University of Houston*
*Houston, Texas*

***Abstract*—Traffic lights are an important part in managing traffic and intersections, and are susceptible to being compromised. In this paper, we provide a pre-attentive way to show when and where traffic light anomalies happen by using a timelapse video and color scale. The system successfully identifies traffic light anomalies and shows what cycle, what time, and where the anomaly happens.**

## 1. Introduction

The objective of this paper is to improve on visualization techniques that were provided in the paper ”A Visual Analytics Approach for Anomaly Detection from a Novel Traffic Light Data” [1]. We seek to provide a more pre-attentive way in visualizing traffic light anomalies by showing when and where they happen. To do so, we obtain the spatial coordinates via Google Maps. Additionally, we calculate the traffic light cycle to better detect the anomalies. The ratio between red light time and green light is calculated for each cycle and outliers of this feature are considered anomalies. This is all visualized with a timelapse video showing where anomalies happen, what time they happen, and what cycle they occur at.

The ”RSE25” dataset has 194,000 rows and 22 columns and is based on the MMITSS sample data. It includes the epoch time, the current indicator at the intersection for major and minor road, and the duration of the indicator. The indicator represents the whether the traffic light state is green, yellow, or red. The duration is measured in seconds and records how long each state lasts for. The other columns were not used in this paper.

## 2. Related Works

This paper is an extension of ’A Visual Analytics Approach for Anomaly Detection from a Novel Traffic Light Data’ by Glenn Turner, Guoning Chen, and Yunpeng Zhang [1]. In that paper, they clean the data and use various visualization techniques to show anomalies within traffic lights. We seek to improve on the visualization techniques that have been used and use the cleaned data that was provided.

### 2.1. Data Cleaning

The data cleaning that occured in the paper by Glenn Turner et al. [1] include removing all duplicate records and sorting the data based on epochtime and indicator. Additionally, they calculated the traffic light cycles and normalized the red light and green light times.

### 2.2. Visual Techniques

The visualization techniques that were used include showing a table of statistics, stacked bar graphs, histograms, line plots, and scatter plots. These graphs are interactable and can show the metadeta of the specific point. The stacked bar graphs were used to represent the hourly accumulated signal time for the intersection. It displayed the red and green light signal time for both the major and minor road. Histograms were used to show the red light and green light signal time distributions. Line plots were used to show red light signal time and green light signal time over elapsed time. Scatter plots were used to display the green versus red light signal times.

## 3. Methodology and Implementation

In this paper, we must solve three problems that occur in order to achieve the visualization we want. In order to show where anomalies happen, we must get the spatial coordinates of the intersection. Before anomalies can be detected, the traffic light cycles must be calculated from the data. Finally, we can detect anomalies by comparing the signal time between red lights and green lights.

### 3.1. Spatial Coordinates

In order to show where anomalies happen, we must get the spatial coordinates of the intersections. However, the spatial coordinates are not provided in the data. This is something that we must manually get by knowing where the intersection is. We use a Google Maps image of the intersection and will update that image using the ”OpenCV” library. In Figure 1, the map Anthem, Arizona displays all the intersection that were present in the paper[1]. But, as

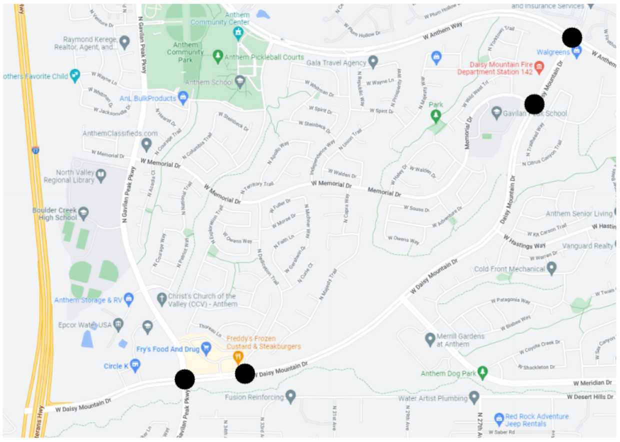



Figure 1: Google Map of where intersections are located in Anthem, Arizona. Each black circle represents an intersection whose data was used in the paper [1]. The ”RSE25” intersection is the leftmost circle.

previously mentioned, we have only been provided with the data for intersection ”RSE25” or the leftmost circle in Figure 1. Thus, the only intersection that is visualized is ”RSE25”

### 3.2. Traffic Light Cycles

To properly calculate the traffic light cycle time for the major street and minor street, we only use the timestamp, current indication, and duration variables within the data set. The timestamp shows the current epochtime of the data entry. We convert all epochtime to the ”datetime” format. The indicator describes what state the traffic light is in for the row. The three state of the traffic light is green light, yellow light, and red light. The duration shows how long the indication has lasted for in seconds. There are 322 cycle for the major road and 318 cycles for the minor road.

Each cycle will contain the cycle number, the green light signal time, yellow light signal time, red light signal time, and the timestamp the cycle ends. The traffic light times are in seconds. The pseudocode for calculating the cycles is displayed in Algorithm 1.

### 3.3. Anomaly Detection

To detect anomalies in the traffic light data, we create a new variable called *Ratio*. The purpose of this new variable is to determine the percentage of time that the cycle is green. This would make it easier to visualize with a color scale. The equation for this is displayed below

$$Ratio = \frac{greentime}{greentime + redtime} \tag{1}$$

where $greentime$ represents the total time that the cycle was green and $redtime$ represents the total time that the cycle was red.

When $Ratio = 1$, then the entire cycle would be a green light. When $Ratio = 0$, then the entire cycle would be a green light. We do not use the yellow light signal time since it does small amount influence when determining if a traffic light cycle contains an anomaly.

**Algorithm 1** Calculate Cycle Time Pseudocode. Assume that cycleCompleted() monitors if we have traversed the three states of the traffic light. It returns a boolean value.

```
Require: Data must be in MMITSS form.
  Initialize greentime, yellowtime, redtime, and cyclenum
  for each row in Data do
    if change in indicator is true then
      if cycleCompleted() is true then
        Insert the timestamp, cyclenum, greentime, yellowtime, and redtime into the cycles DataFrame.
        Reset variables again for new cycle.
        Increment cyclenum.
      end if
    end if
    Update greentime, yellowtime, or redtime based on indicator.
  end for
```

To calculate if an anomaly occurs within a cycle, we check for any outliers in the cycles data using the z-score metric. If an outlier has been found, then the cycle contains a traffic light anomaly.

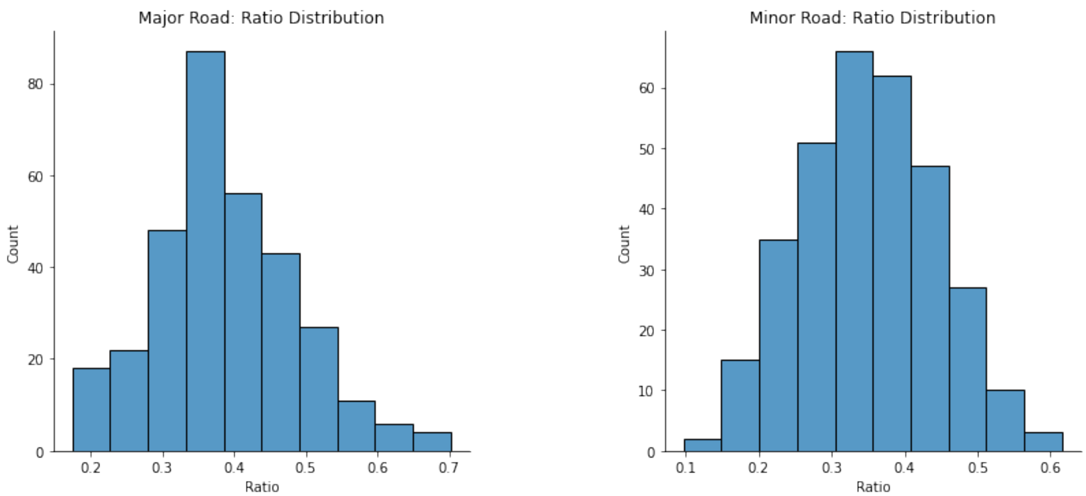
## 4. Results and Discussion

In addition to the timelapse of the cycles, two histogram graphs and a line plot have been created using the ”Matplotlib” Python library. These were used to provide a better understand of the cycles, the *Ratio* value, and how they relate to each other.

In Figure 2, the *Ratio* value distribution for both the major and minor road is displayed. The *Ratio* value tends to be around the 0.4 value for the major road and 0.35 for the minor road. This suggests that the traffic light cycles are red light for more time than they are green light. The major road is slightly skewed right which would suggest there are possible anomalies for the major road. The minor road has a normal distribution meaning that it has no anomalies.

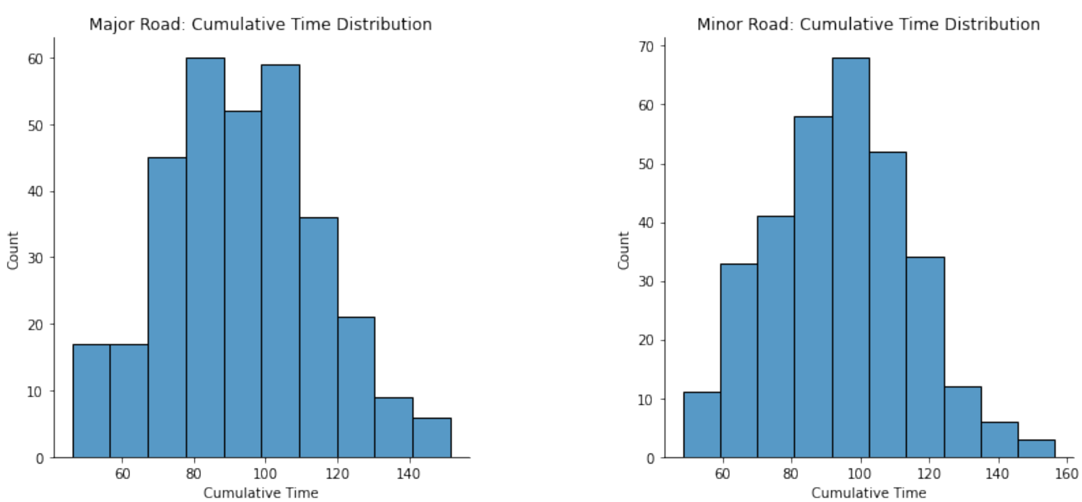
In Figure 3, the cumulative time distribution for both the major and minor road is displayed. The cumulative time shows how long each cycle lasts for. For the major road, most cycles tend to last between 80 to 100 seconds. Cycles for the minor road tend to last for 100 seconds. None of the cycles last less than 45 seconds and last more than 3 minutes.

In Figure 4, a line plot displays *Ratio* as the cycle number increases through time. The red dashed line represents the upper and lower bounds of *Ratio*. If the line goes beyond the bounds, then an anomaly are exists at that intersection. The upper bound represents that the cycle is a green light for an abnormal amount of time. The lower bound represents that the cycle is a red light for an abnormal amount of time. The bounds are different values for the




(a) (b)

Figure 2: The "Ratio" distribution for the major road (left) has the mean around 0.39 and slightly skewed right. The "Ratio" distribution for the minor road (right) has the mean of 0.35 and a normal distribution as well.




(a) (b)

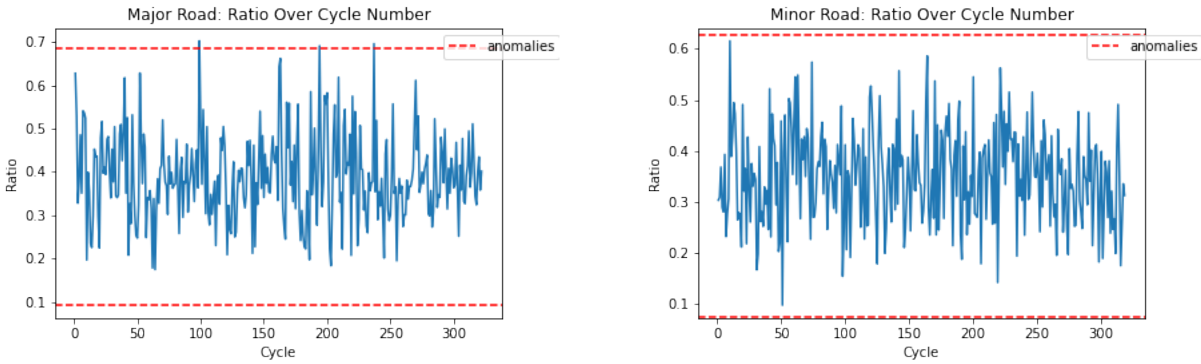
Figure 3: The cumulative time distribution for the major road (left) appears to have two peaks at around 80 and 100. The cumulative time distribution for the minor road (right) has only one peak at around 100. They both appear to have a normal distribution .

major road and the minor road. The major road contains three anomalies at cycles 99, 194, and 237. The minor road at no point crosses the bounded. The $Ratio$ value fluctuates around 0.4 for the major road and 0.35 for the minor road.

## 4.1. Timelapse Visualization

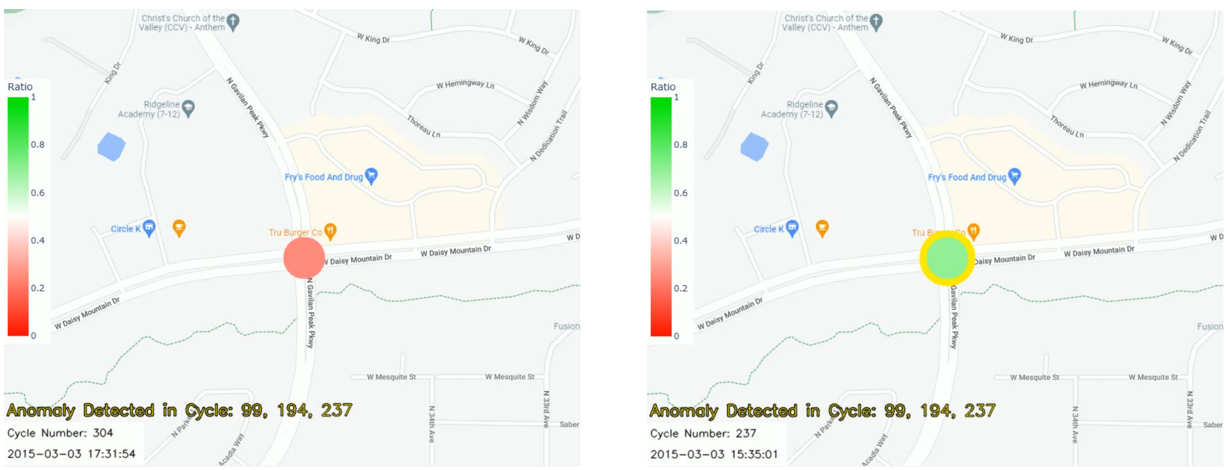
To create the timelapse, the "OpenCV" Python library was used. It was used to display the color value of the intersection based on $Ratio$, the cycle number, date, and time at the bottom right. The color scale used to represent $Ratio$ for the intersection is a divergent color scale. Red is used to represent when the $Ratio = 0$ and green is used when $Ratio = 1$. The color scale is also displayed at the left hand side. A notification appears displaying the cycle number the anomalies were detected in. Additionally,a yellow ring appears around the colored circle for the cycle an anomaly occurs in. Every frame of the video represents a new cycle and the color value of the intersection updates. Figure 5 is a screenshot of what the final timelapse video looks like.

Through watching the timelapse video, we are able to




(a) (b)

Figure 4: A line plot showing "Ratio" as the cycle number increases. The red-dashed line represent the upper and lower bounds of "Ratio". Any point beyond the bounds represents an anomaly in the traffic light data. In the major road plot (left), we see that there are anomalies detected at cycles 99, 194, and 237. The minor road plot( right) does not show any anomalies detected.




(a) (b)

Figure 5: Two screenshots of the timelapse video for the major road of intersection "RSE25". The *Ratio* value for the specific cycle is displayed with the colored circle at the intersection. The color scale is displayed on the left hand side. The cycle number, date, and time is displayed at the bottom left. If an anomaly is detected, a notification pops up showing which cycle contains an anomaly. The photo on the right displays the yellow ring that appears when an anomaly is detected in the cycle. The photo on the left shows is when there is no anomaly for the cycle.

identify where or which intersection is being analysed for traffic light anomalies. We are able to see that a majority of the time, cycles tend to be a red light longer than a green light for both the major and minor road. Only the major road has anomalies detected and those anomalies happen when the cycle is a green light significantly longer than the red light. The anomalies happen at cycles 99, 194, and 237.

## 5. Conclusion and Future Work

Overall, the visualization technique is successful at show when and where an anomaly happens. We were able to display where the intersection is without being provided the spatial coordinates in the data. We were also able to calculate the traffic light cycles and detect anomalies using the *Ratio* value.

However, there are severe limitation with the current implementation. We can only display an intersection if we have an image containing an over-head view of the intersection. Additionally, only one road of an intersection can be displayed at a time since cycles do not start and end at the same time. This can be solved if seconds were used instead of cycles for the timelapse. Future work includes creating a graphical user interface to display all visualizations in a single program. It would also make it easier to add functionality to better traverse through the timelapse. It would be easier to jump to a specific cycle number and obtain the $Ratio$ value instead of scrolling through a video and guessing the numerical value based on the color.

## References

[1] Glenn Turner, Guoning Chen, Yunpeng Zhang, " A Visual Analytics Approach for Anomaly Detection from a Novel Traffic Light Data ," IS&T Electronic Imaging, Visualization and Data Analysis (VDA) 2021, virtual event, January, 2021.