Human-Computer Interaction Lab
University of Maryland

LifeFlow Case Study:

Comparing Traffic Agencies' Performance from Traffic Incident Logs

Introduction

Vehicle crashes remain the leading cause of death for people between the ages of four and thirty-four. In 2008, approximately 6-million traffic accidents occurred in the United States. This resulted in nearly 40,000 deaths, 2.5 million injuries, and losses estimated at $237 billion. While traditional safety and incident analysis has mostly focused on incident attributes data, such as the location and time of the incident, there are other aspects in incident response that are temporal in nature and are more difficult to analyze.
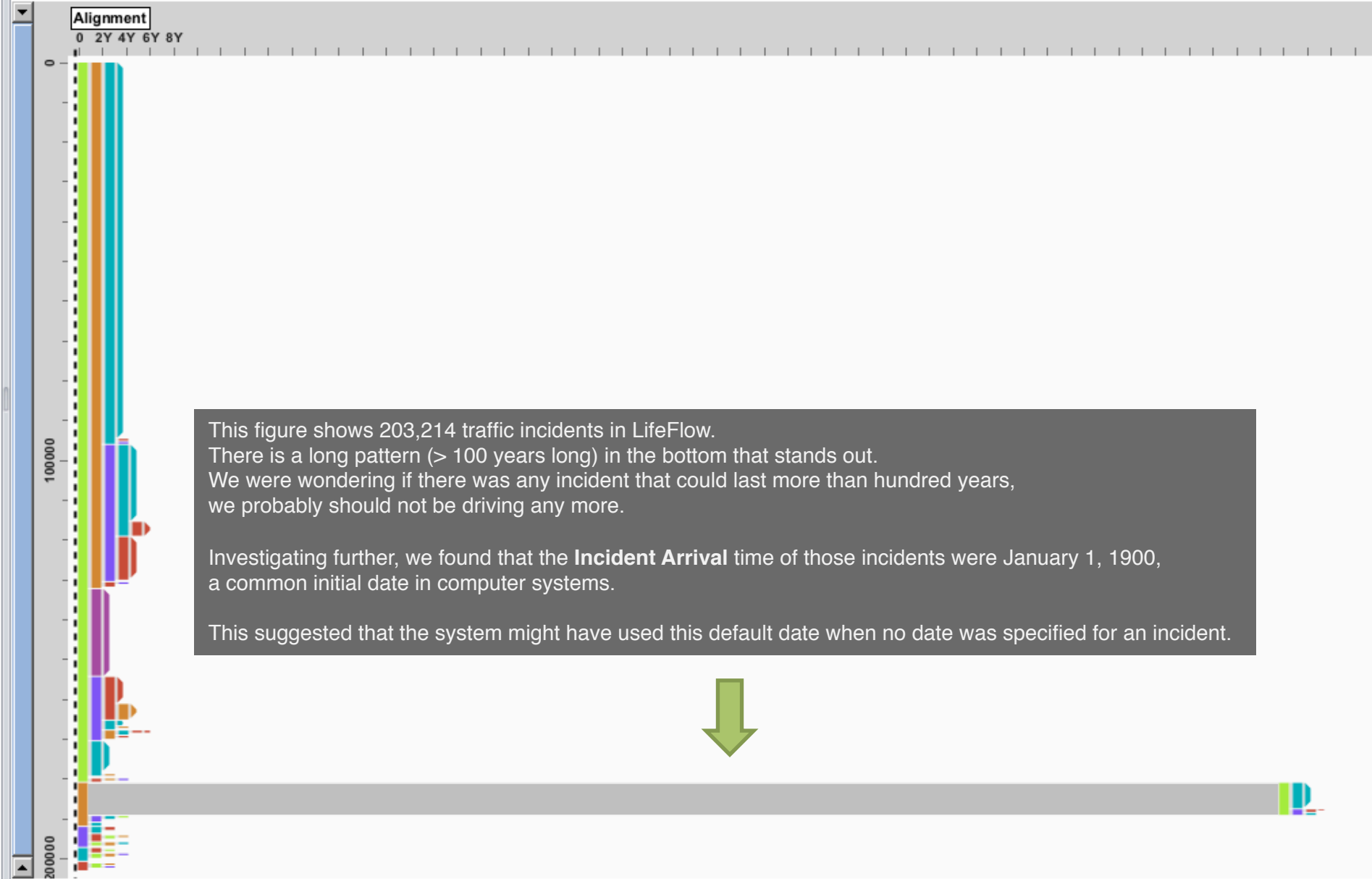
We used LifeFlow to examine a dataset from the National Cooperative Highway Research Program (NCHRP) that includes 203,214 traffic incidents from 8 agencies. Each incident record includes a sequence of incident management events:

- **Incident notification** --when the agency is first notified of the incident
- **Incident Arrival** --when the emergency team arrives the scene
- **Lane Clearance** --when the lanes are opened, but the incident scene may be not completely cleared
- **Incident cleared, Incident clearance**, and **Return to normal** --all denote the end of incidents. For ease of analysis, we aggregated all three into the new event type **Return to normal (aggregated)**.

A typical sequence should start with **Incident Notification** and finish with **Return to normal (aggregated)**, with the possibility of having **Incident Arrival** and **Lane Clearance** in between.

In addition, the traffic incidents data include two attributes:

- **Agency** (represented with a letter from A to H for anonymity)
- **Type of incident** (e.g. "Disabled Vehicle", "Fatal Accident", etc.)

This figure shows 203,214 traffic incidents in LifeFlow.
There is a long pattern (> 100 years long) in the bottom that stands out.
We were wondering if there was any incident that could last more than hundred years,
we probably should not be driving any more.

Investigating further, we found that the **Incident Arrival** time of those incidents were January 1, 1900,
a common initial date in computer systems.

This suggested that the system might have used this default date when no date was specified for an incident.

Clean the data!

File Dataset Selection View Help

Data Manager **Control**

untitled (CATTLab_003_all_...

showing 195504/195504 records

0 selected

Remove them   Remove others

All **Overview** **Timeline**

Align Filter

Please specify how to align:

◉ No alignment

○ 1 st ▲ Incident notific ▼
   from the beginning ▼

○ All ▲ Incident notific ▼

(no advanced option selected)

⚒ Advanced Options...

☑ automatic update   **Align**

Legend **Attribute**

Select all   Select none

☑ ☑ ▲ Incident notification
☐ ☐ ▲ Incident arrival
☐ ☐ ▲ Incident clearance
☑ ☑ ▲ Return to normal (Aggregated)
☐ ☐ ▲ Incident cleared
☐ ☐ ▲ Lane clearance
☐ ☐ ▲ Return to normal

▶ ◀ Measurement Tool

Zoom: 1 Hours ▼

Alignment
0

After cleaning the data, we used the time from when the agencies were notified to the final clearance of the incidents as a performance measure.

The time when the agency was notified can be indicated by the **Incident Notification** event and the final clearance is indicated by the **Return to normal (Aggregated)** event.

File Dataset Selection View Help

Data Manager **Control**

untitled (CATTLab_003_all_...

showing 195504/195504 records

0 selected

Remove them   Remove others

All **Overview** **Timeline**

Align Filter

Please specify how to align:

◉ No alignment

○ 1 st ▲ Incident notific ▼
   from the beginning ▼

○ All ▲ Incident notific ▼

(no advanced option selected)

✕ Advanced Options...

☑ automatic update   **Align**

Legend **Attribute**

☑ Agency
☐ Incident Type

◻ Measurement Tool

Zoom: 2 ⊕ Hours ▼

Alignment
0   1h   2h

G

B

H

A

E

C
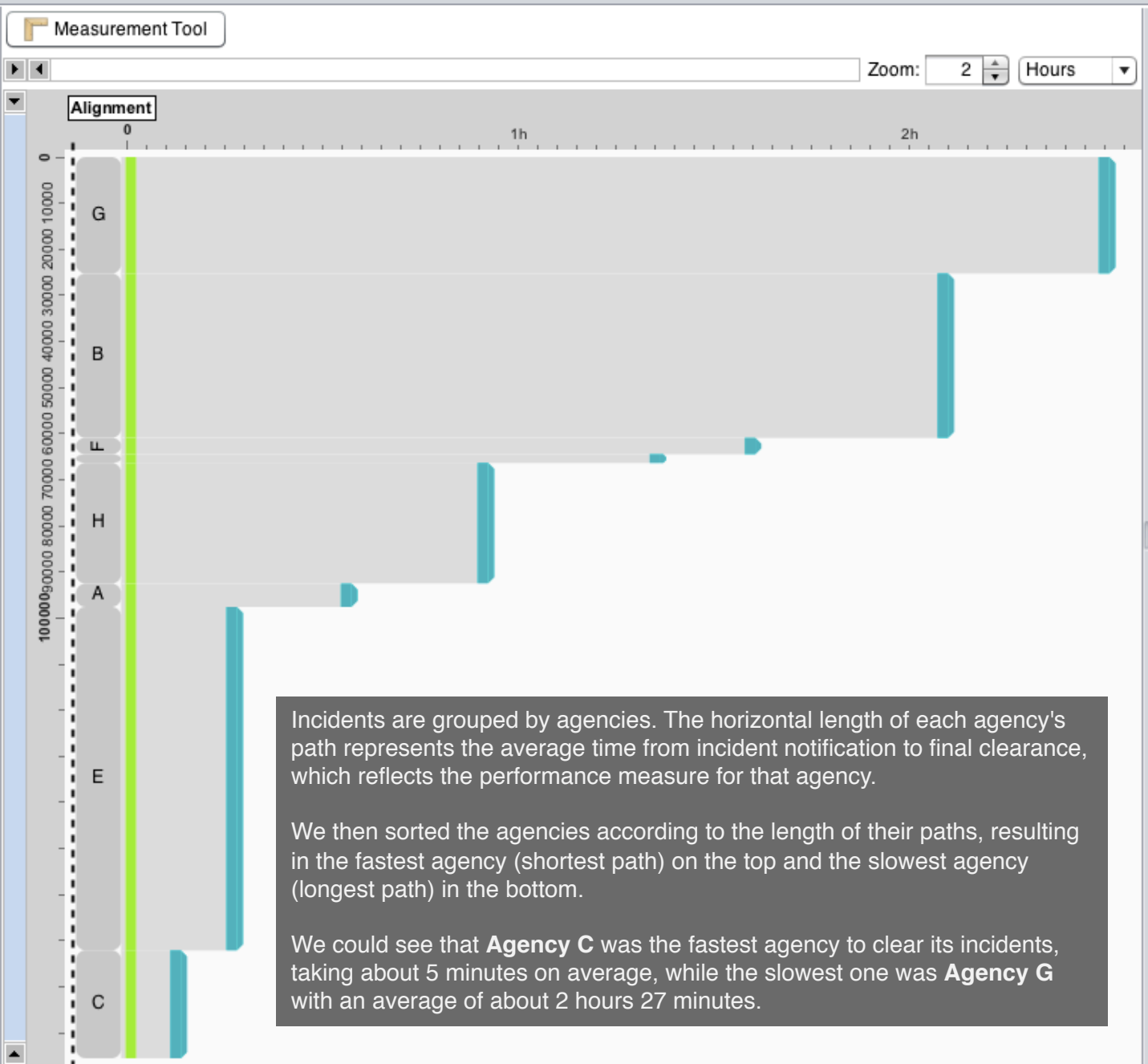
Incidents are grouped by agencies. The horizontal length of each agency's path represents the average time from incident notification to final clearance, which reflects the performance measure for that agency.

We then sorted the agencies according to the length of their paths, resulting in the fastest agency (shortest path) on the top and the slowest agency (longest path) in the bottom.

We could see that **Agency C** was the fastest agency to clear its incidents, taking about 5 minutes on average, while the slowest one was **Agency G** with an average of about 2 hours 27 minutes.

File   Dataset   Selection   View   Help

Data Manager   **Control**

untitled (CATTLab_003_all_...

showing 48672/195504 records

0 selected

Remove them   Remove others

All   **Overview**   **Timeline**

Align   Filter

Please specify how to align:

◉ No alignment

○ 1 st ▲ Incident notific ▼
from the beginning ▼

○ All ▲ Incident notific ▼

(no advanced option selected)

✖ Advanced Options...

☑ automatic update   **Align**

Legend   **Attribute**

☑ Agency
☐ Incident Type

Measurement Tool

Zoom: 2   Hours ▼

Alignment
0   1h   2h

G

C
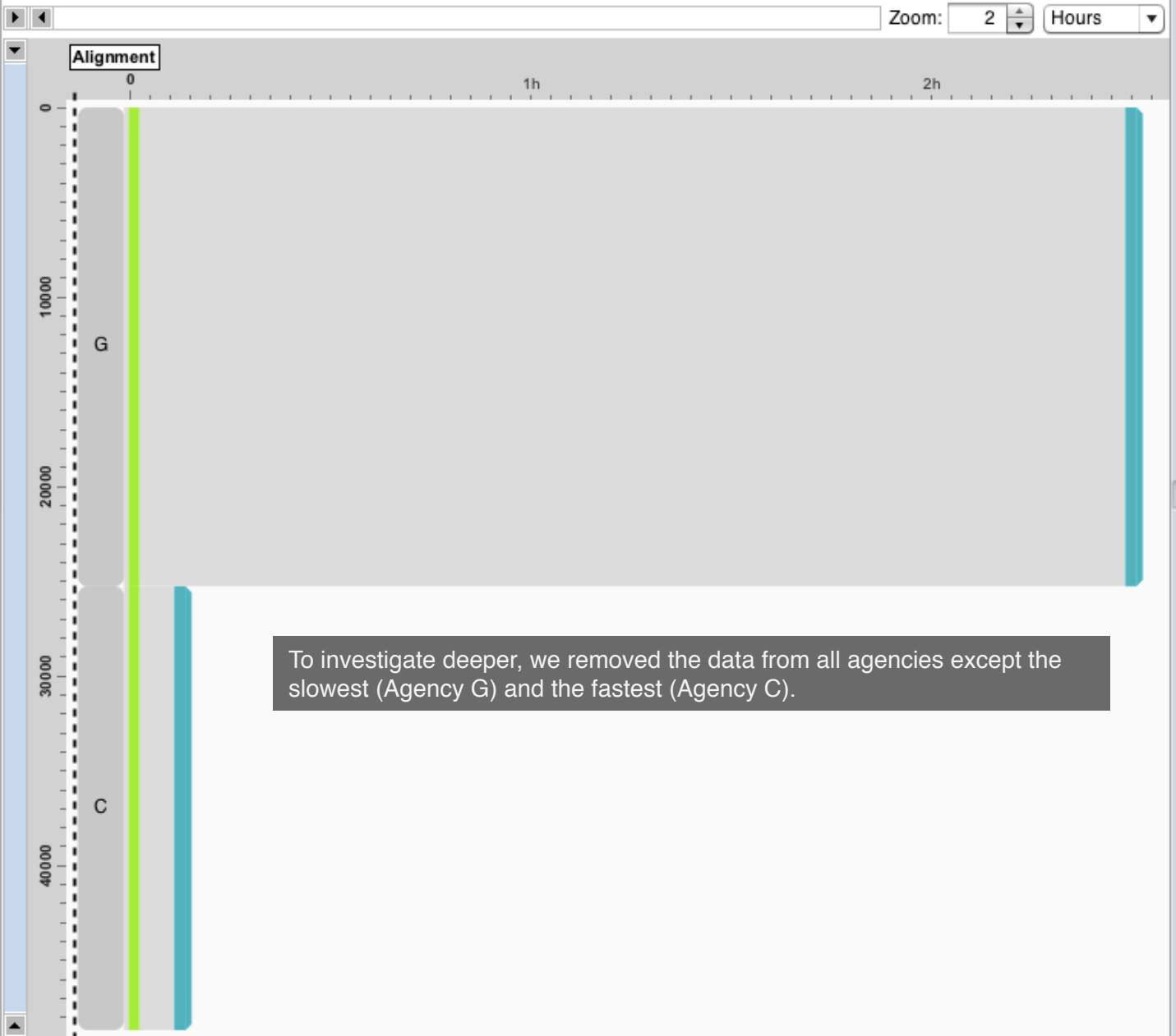
To investigate deeper, we removed the data from all agencies except the slowest (Agency G) and the fastest (Agency C).

File Dataset Selection View Help

Data Manager **Control**

untitled (CATTLab_003_all_...

showing 48672/195504 records

0 selected

Remove them   Remove others

23410 RECORDS (48.1%)*
0/23410 selected
12% of original data
* = percent of displayed data

⬡ Agency = C

| 0. Incident notification

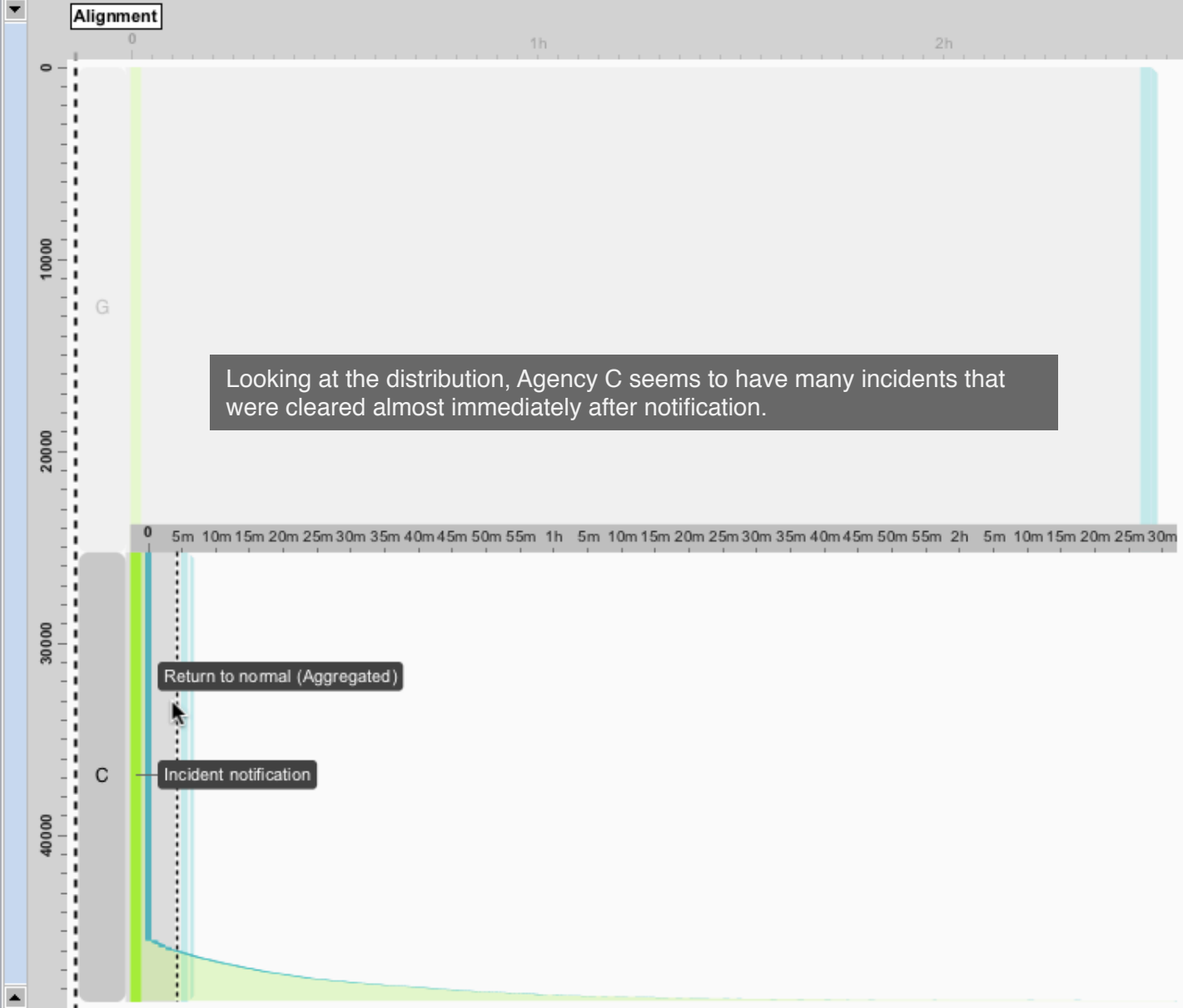Mean: 0:04 hrs 32.894 s
Median: 0
SD: 0:29 hrs 59.877 s

| **1. Return to normal (Aggregated)**

Average time to the end:
0:04 hrs 32.894 s

Press 'F2' to freeze tooltip.

Legend **Attribute**

☑ Agency
☐ Incident Type

📐 Measurement Tool

Zoom: 2 ⬍ Hours ▾

Alignment

0   1h   2h

0

10000

20000

G

Looking at the distribution, Agency C seems to have many incidents that were cleared almost immediately after notification.

0 5m 10m 15m 20m 25m 30m 35m 40m 45m 50m 55m 1h 5m 10m 15m 20m 25m 30m 35m 40m 45m 50m 55m 2h 5m 10m 15m 20m 25m 30m

30000

Return to normal (Aggregated)

C   Incident notification

40000

File Dataset Selection View Help

Data Manager   Control

untitled (CATTLab_003_all_...

showing 48672/195504 records

0 selected

Remove them   Remove others

25251 RECORDS (51.9%)*
0/25251 selected
12.9% of original data
* = percent of displayed data

⬡ Agency = G

0. Incident notification

Mean: 2:27 hrs 0.615 s
Median: 0:32 hrs 8 s
SD: 1 day 1:21 hrs

1. Return to normal (Aggregated)

Average time to the end:
2:27 hrs 0.615 s

Press 'F2' to freeze tooltip.

Legend   Attribute

☑ Agency
☐ Incident Type

Measurement Tool

Zoom: 2 ⬍ Hours ▾

Alignment

0   5m 10m 15m 20m 25m 30m 35m 40m 45m 50m 55m 1h   5m 10m 15m 20m 25m 30m 35m 40m 45m 50m 55m 2h   5m 10m 15m 20m 25m 30m

0

10000

20000

30000

40000

G

C

Return to normal (Aggregated)

Incident notification

On the other hand, Agency G does not have many incidents with very short clearing time.

File   Dataset   Selection   View   Help

Data Manager   **Control**

untitled (CATTLab_003_all_...

showing 48672/195504 records

0 selected

Remove them   Remove others

All   **Overview**   **Timeline**

Align   Filter

Please specify how to align:

◉ No alignment

○ [1 ▲▼] st [▲ Incident notific ▼]
   [from the beginning ▼]

○ All [▲ Incident notific ▼]

(no advanced option selected)

✕ Advanced Options...
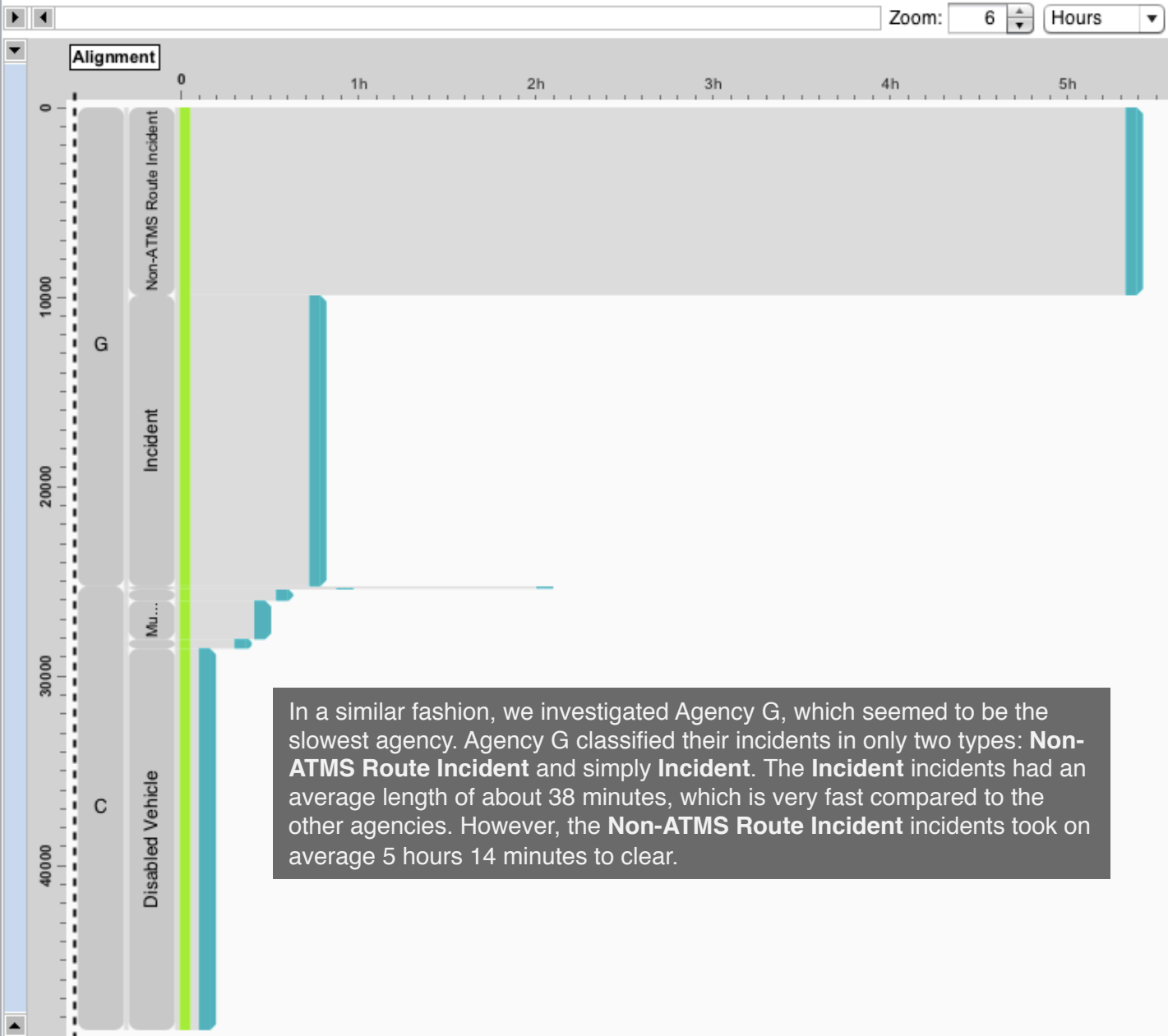
☑ automatic update   **Align**

Legend   **Attribute**

☑ Agency
☑ Incident Type

Measurement Tool

▶ ◀ [   ]   Zoom: [6 ▲▼] [Hours ▼]

Alignment

0   1h   2h   3h   4h   5h

Non-ATMS Route Incident

G

Incident

Mu...

C

Disabled Vehicle

We looked into the different incident types reported and found that most of the incidents that Agency C reported are **Disabled Vehicles** which had about 1 minute clearance time on average. A large number of the incidents reported **Clearance** immediately after **Incident Notification**. This observation made us wonder if there is any explanation for these immediate clearances, and encouraged further analysis.

Many incidents were cleared within 1 minute.

LifeFlow 3.5 – untitled (CATTLab_003_all_that_start_with_incident_notification_return_to_normal_aggregated.txt)

File   Dataset   Selection   View   Help

Data Manager   **Control**

untitled (CATTLab_003_all_...

showing 48672/195504 records

0 selected

Remove them   Remove others

All   **Overview**   **Timeline**

Align   Filter

Please specify how to align:

◉ No alignment

○ 1 ⬍ st ▲ Incident notific ▼
   from the beginning ▼

○ All ▲ Incident notific ▼

(no advanced option selected)

✖ Advanced Options...

☑ automatic update   **Align**

Legend   **Attribute**

☑ Agency
☑ Incident Type

Measurement Tool

Zoom: 6 ⬍   Hours ▼

Alignment

0   1h   2h   3h   4h   5h

G — Non-ATMS Route Incident — Incident — Mu...

C — Disabled Vehicle

In a similar fashion, we investigated Agency G, which seemed to be the slowest agency. Agency G classified their incidents in only two types: **Non-ATMS Route Incident** and simply **Incident**. The **Incident** incidents had an average length of about 38 minutes, which is very fast compared to the other agencies. However, the **Non-ATMS Route Incident** incidents took on average 5 hours 14 minutes to clear.

Conclusions

These evidences showed that the agencies reported their incident management timing data in inconsistent ways, e.g., Agency C's incidents with 1 minute clearance time. Beside timing information, the same problem occurred with incident types as the terminology was not consistent across agencies or even local areas. Difference might stem from differences in data entry practices, terminology variations, or missing data. Policies are needed to decide if, and how, to normalize the data and tools will be needed to manage this normalization process and provide audit trails documenting the transformations. At the time of this analysis, these inconsistencies among agencies restricted further meaningful performance comparison.

Although our data analysis in the case study was limited and preliminary, domain experts from the CATT Lab were conducting a more formal analysis of the data. They reviewed our work and stated that they wished LifeFlow was available earlier on when they started their own analysis. They confirmed the existence of anomalies that we had found in the data, and stated that their elimination was non-trivial when using SQL because they had to expect the errors in advance and be careful to exclude them from their analysis. However excluding all the possible erroneous sequences in a SQL query would be very difficult. In the end, they needed to review the results of SQL queries to ascertain that there were no longer any errors. Without LifeFlow, this kind of review and identification of unexpected sequences would be almost impossible.

Finally, they mentioned that LifeFlow would allow them to ask more questions faster, and probably richer questions about the data. LifeFlow was also able to reveal unexpected sequences that may have been overlooked, but the tool also suggested that their prevalence is limited. We believe that using LifeFlow can assist analysts explore large datasets, such as the NCHRP traffic incident information, in ways that would be very difficult using traditional tools and might allow analysts to find richer results in less time.