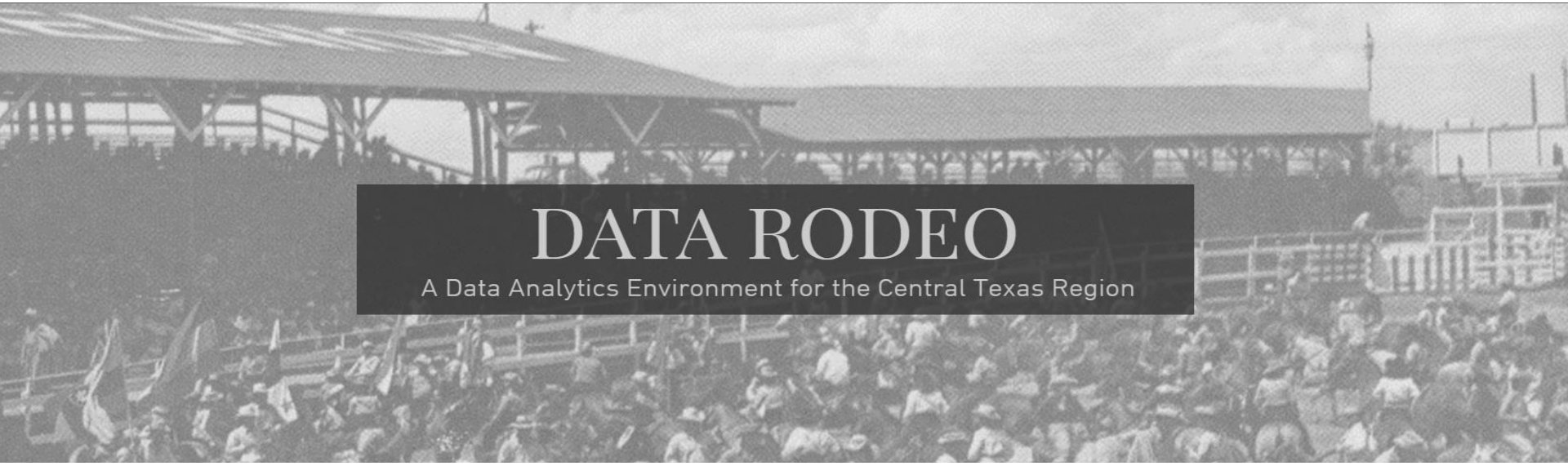
Sharing Novel Data Sources to Promote Innovation Through Collaboration: Case Studies in Austin, TX.

UTFI Workshop: Visualizing Transportation Data for Efficient Decision Making
December 8th 2017



Natalia Ruiz Juri, PhD

Outline

- Overview of the Data Rodeo
- Bluetooth/WiFi Data analysis
- RideAustin data analysis
- Video Data analysis

The Data Rodeo

Data Access & Analysis

- Facilitate the access to and analysis of real transportation data from multiple sources for research and real-world applications.

Data-Centric Research

- Facilitate the use of real transportation data in research to support high-impact research products.

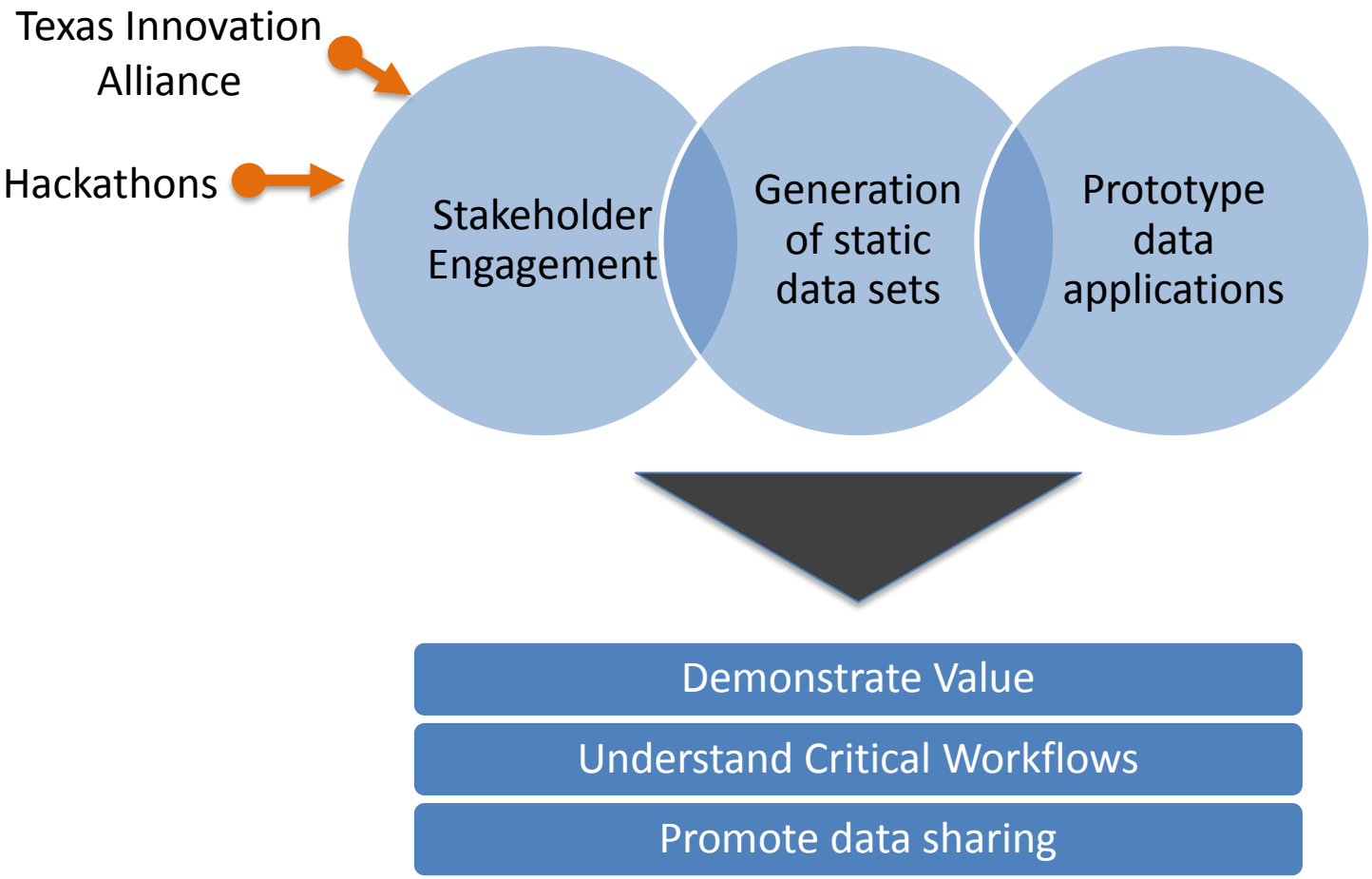
Replicability & Transferability

- Promote replicable and transferable research
- Promote cross disciplinary collaboration

Technology Transfer

- Promote stakeholder communications
- Promote faster adoption of research products

Approach


Texas Innovation Alliance
Hackathons
Stakeholder Engagement
Generation of static data sets
Prototype data applications
Demonstrate Value
Understand Critical Workflows
Promote data sharing

Working with the CoA

Phase I

Small prototypes to demonstrate value and engage the community

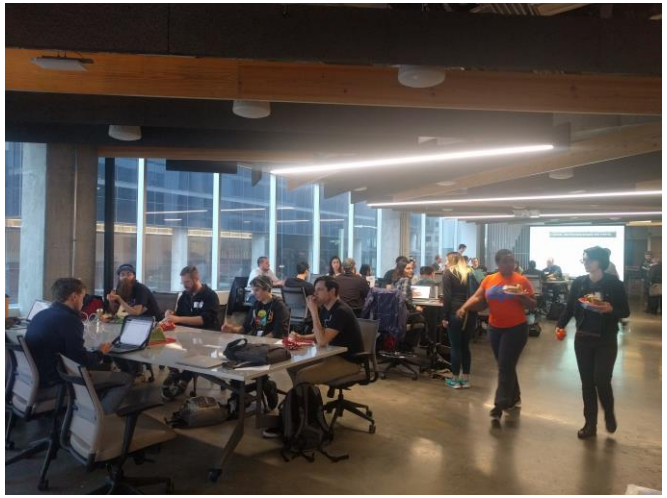
- Bluetooth data Hackathon
- Wi-Fi data Hackathon
- Video analytics using AI prototype

Phase II

Applications using sustainable data architecture

- Connected corridor pilot
- Supporting Vision Zero using video analytics
- Using data to prioritize signal re-timing

Bluetooth Data Hackathon



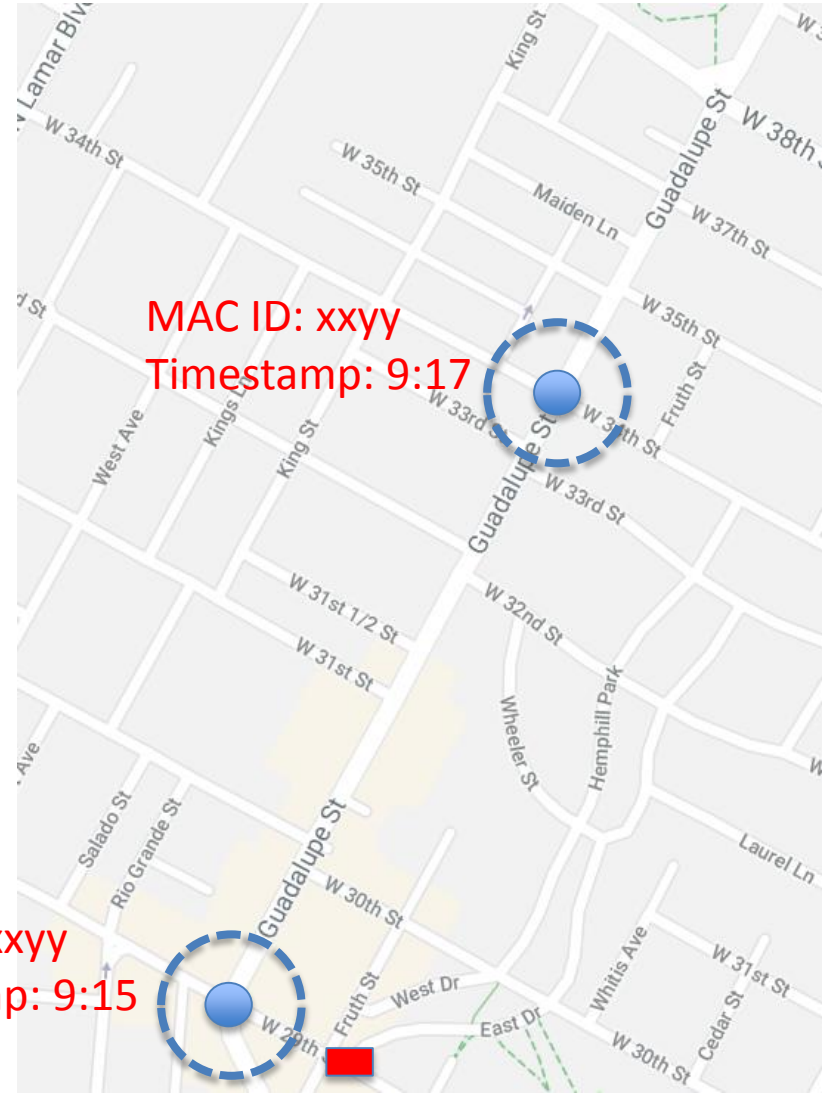
- Our role
  - Help prepare the dataset for sharing
  - Facilitate hosting
  - Developed some prototypes
- Outcomes
  - Large attendance & interest
  - Some progress by attendees during the event
  - Data available for future use
- Challenges
  - Developing valuable products requires familiarity with data and applications
  - Effort required to continue building on the prototypes developed during the event

Bluetooth data

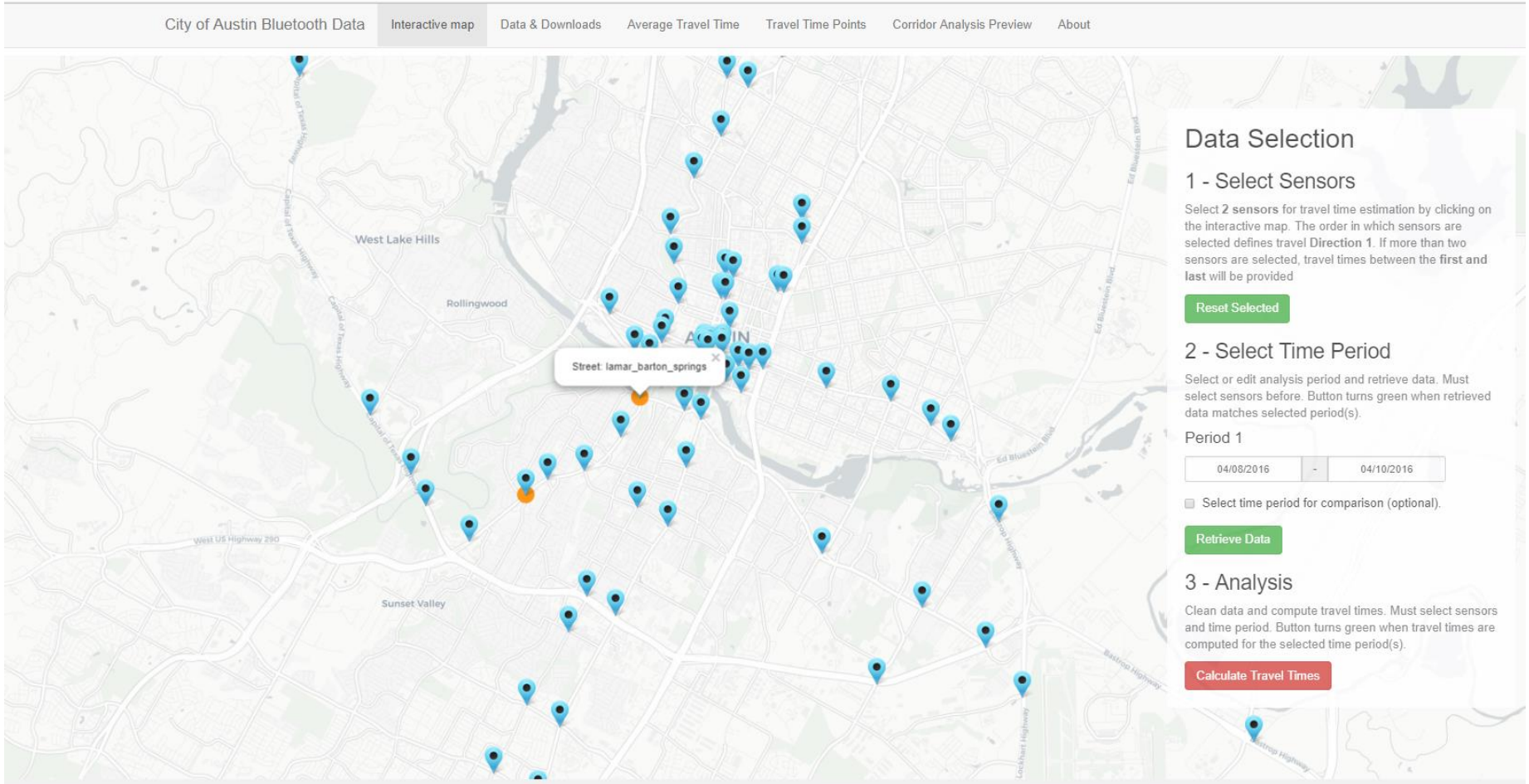
Sensor

Sensing radius

Bluetooth-enabled device

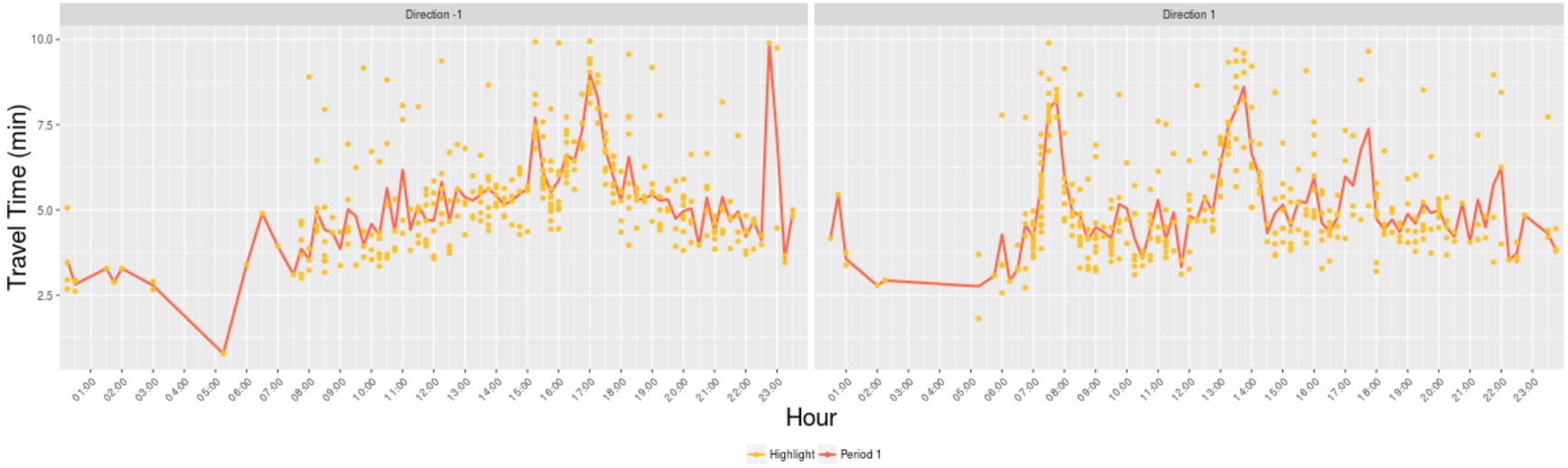


Data Analysis Tool

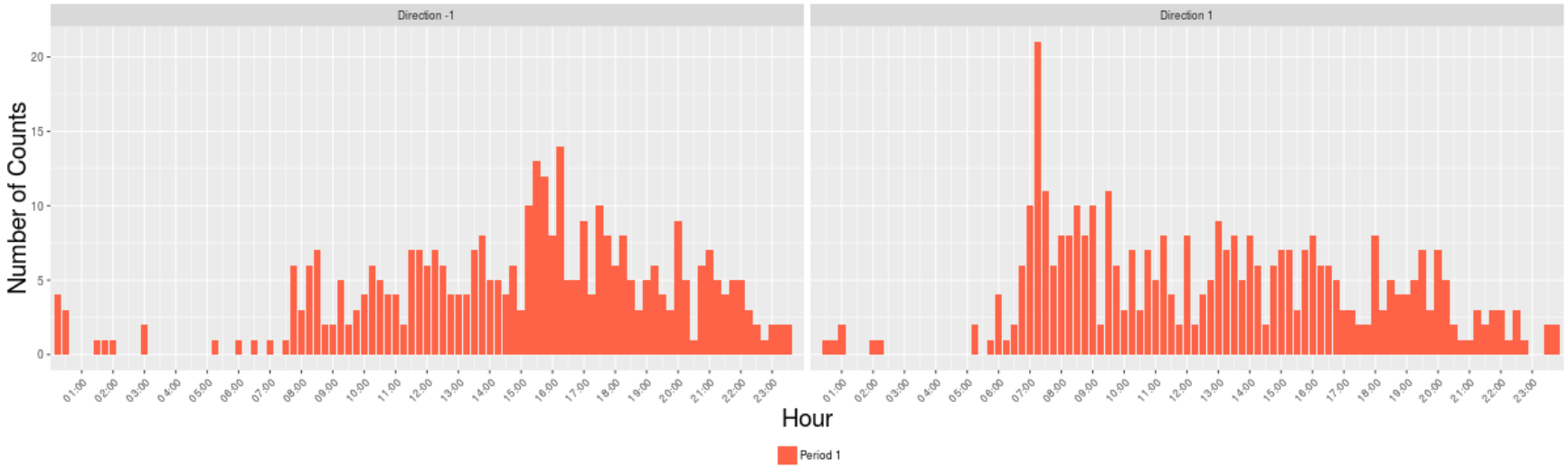


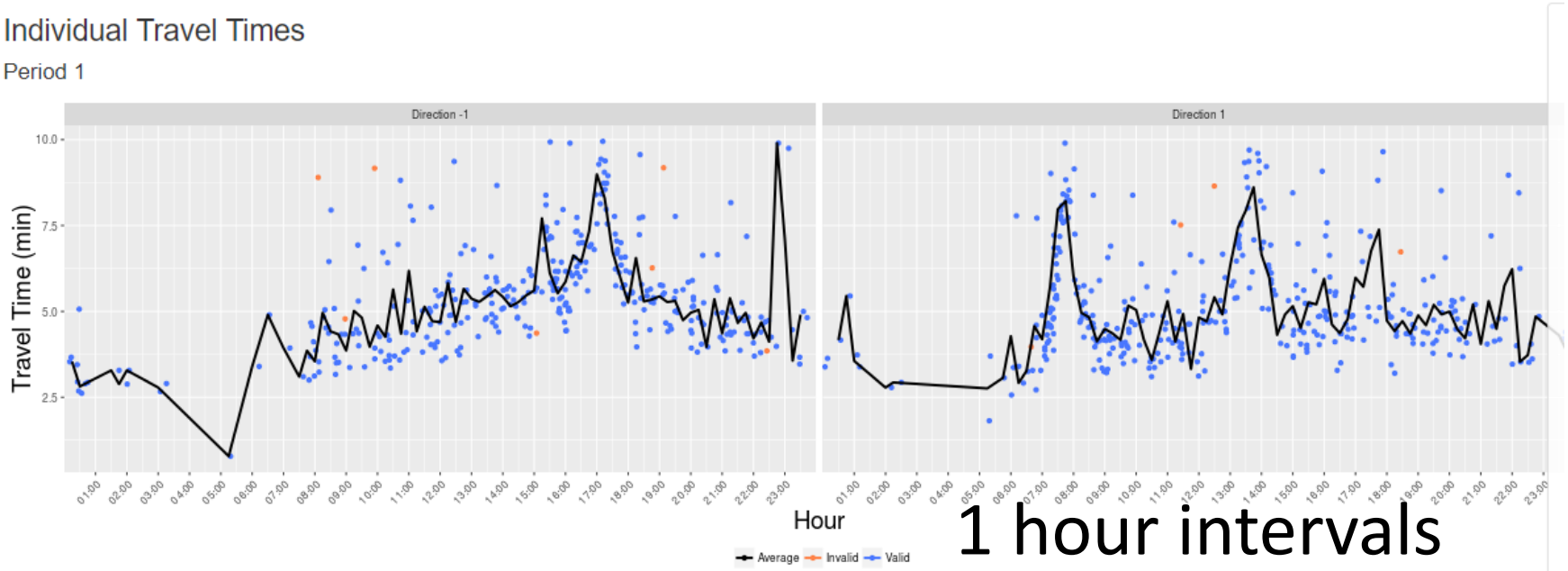
Average Travel Time Analysis

Travel Times

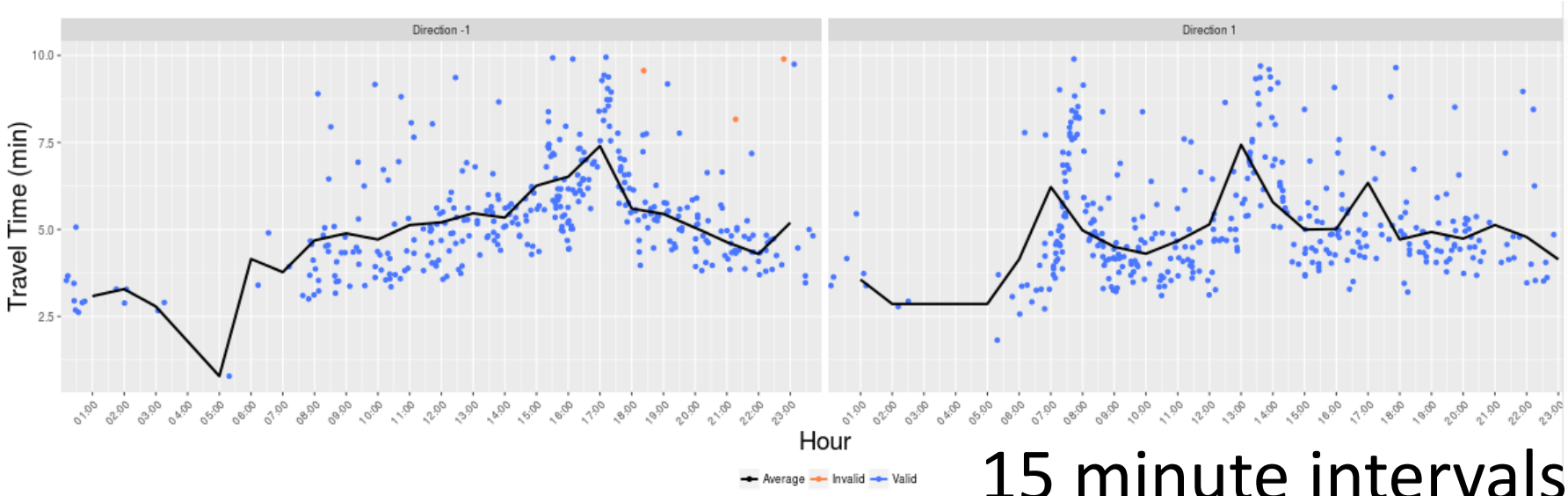


Number of Valid Data Points





1 hour intervals

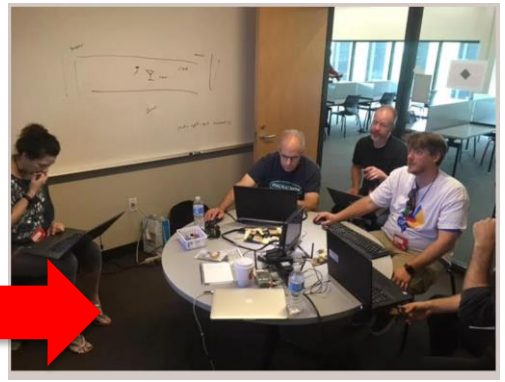


15 minute intervals

Summary

- Easier to access than sensors software
- Explore impact of data cleaning and aggregation
- Decision making
  - Validate use of sensor data for travel time estimation
  - Motivated application to support prioritization of signal re-timing
  - Pending question: sensors/probe data
    - Cost, maintenance, coverage.

Raspberry PI Hackathon

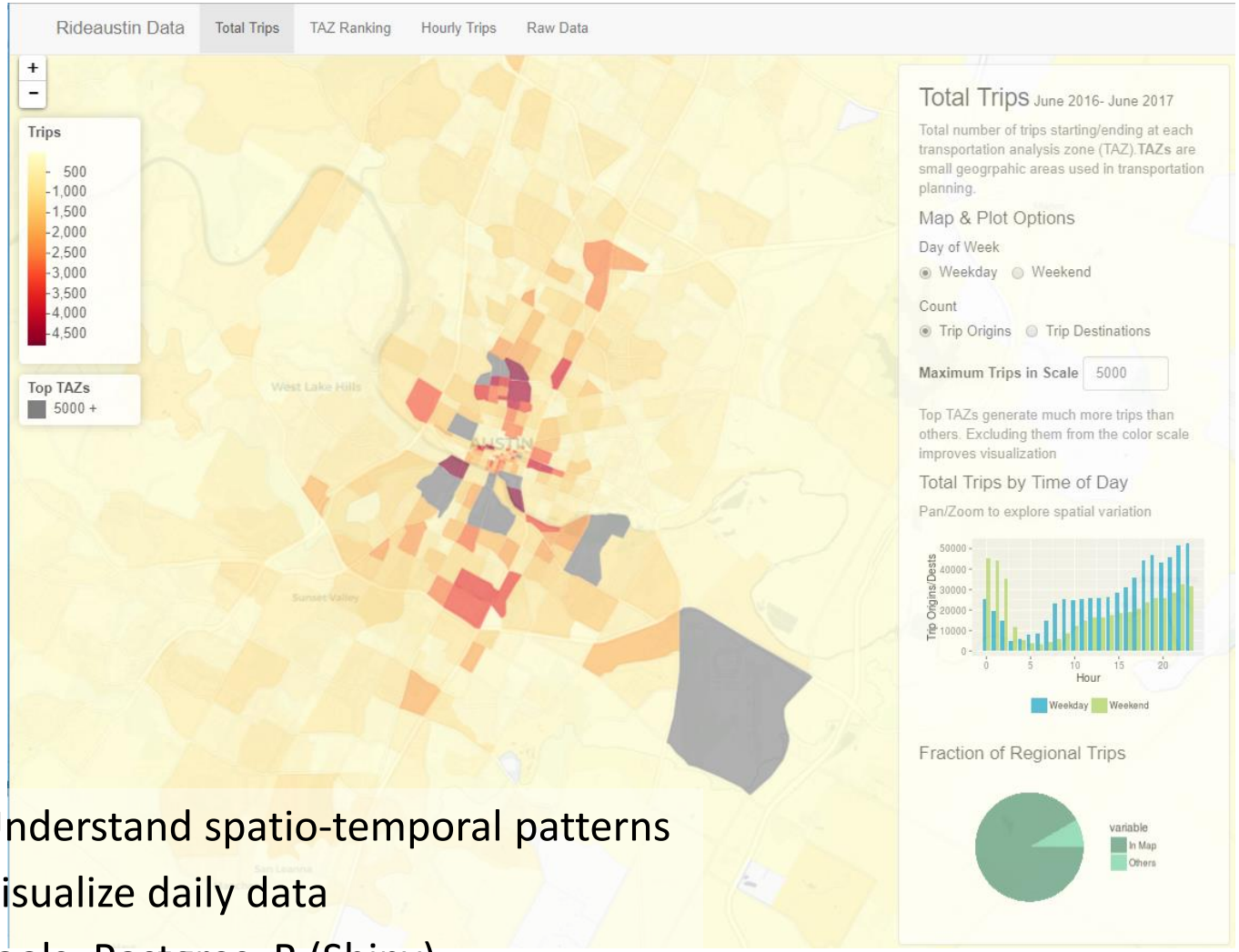


RideAustin Data Analysis

- TNC operating in Austin during 2016
- Released a large data set including pick-up and drop-off location of all rides 

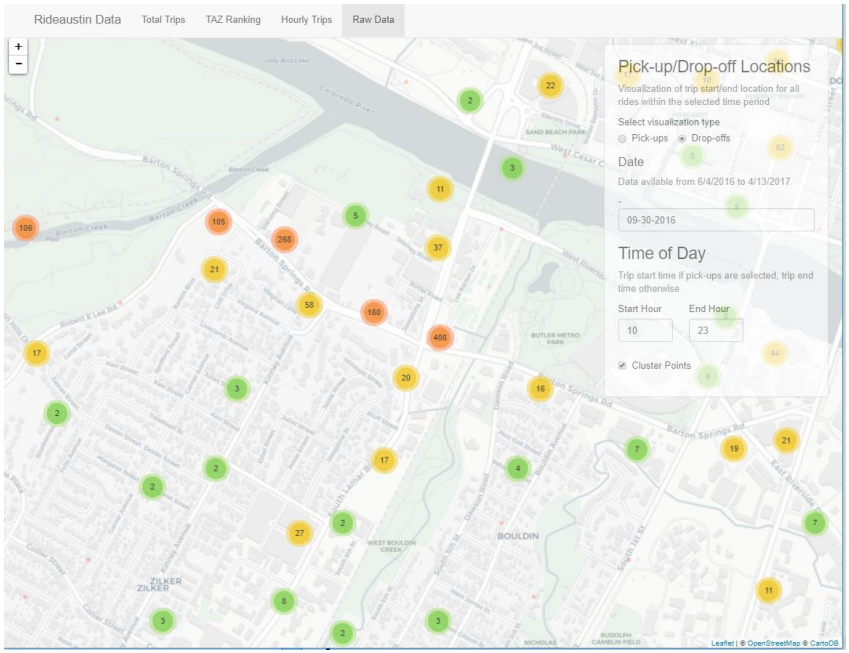
- Valuable insights
  - Planning: Behavior of drivers and travelers.
  - Operations: Change of patterns during special events, impact of designating drop-off locations.

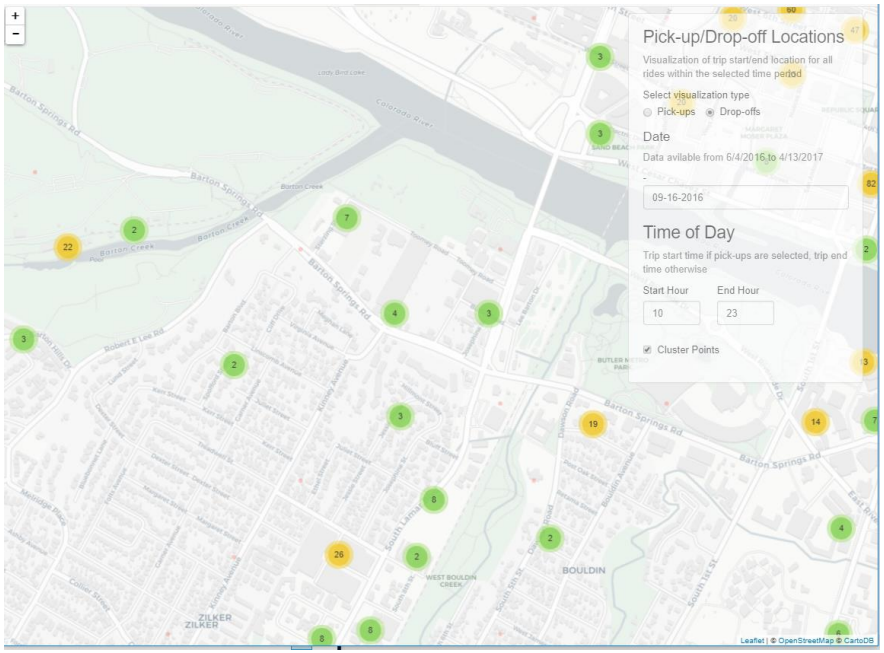
Web Application



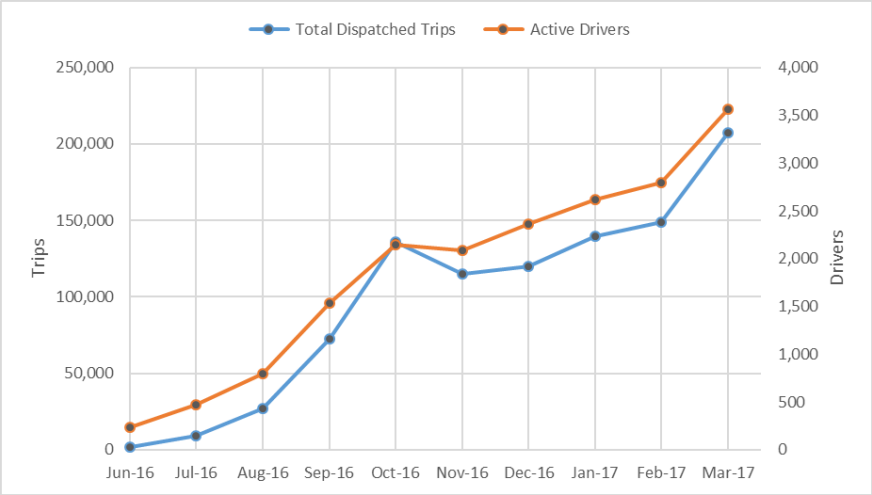
- Understand spatio-temporal patterns
- Visualize daily data
- Tools: Postgres, R (Shiny)

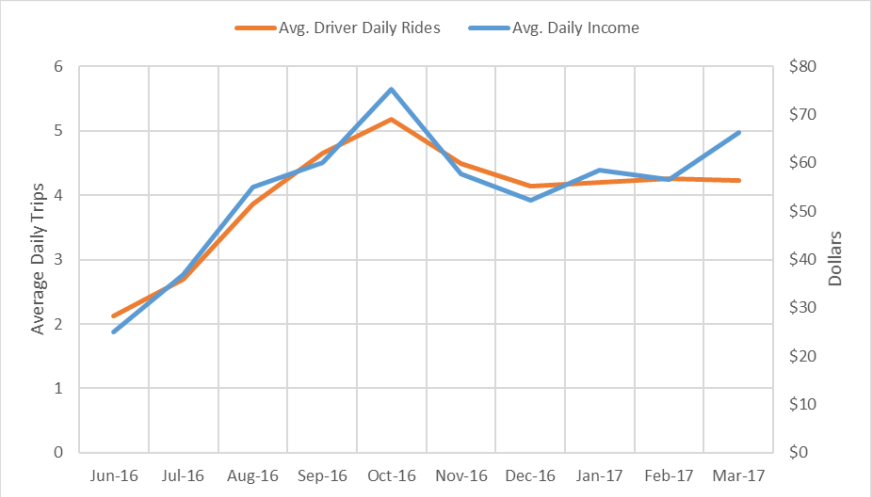
Special Events





Drivers

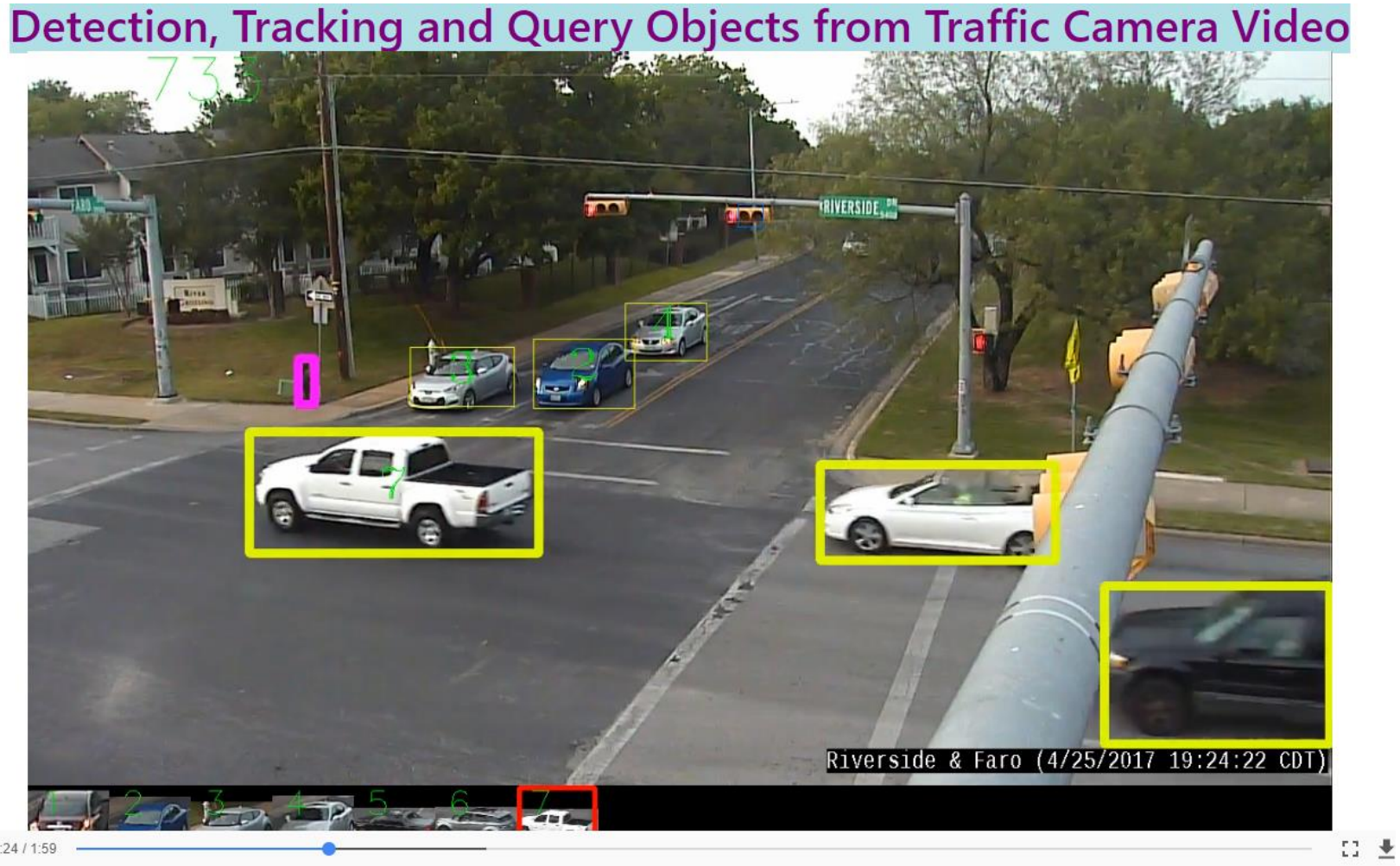

Total Dispatched Trips
Active Drivers
250,000
200,000
150,000
100,000
50,000
0
Trips
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
Drivers
Jun-16
Jul-16
Aug-16
Sep-16
Oct-16
Nov-16
Dec-16
Jan-17
Feb-17
Mar-17


Avg. Driver Daily Rides
Avg. Daily Income
6
5
4
3
2
1
0
Average Daily Trips
$80
$70
$60
$50
$40
$30
$20
$10
$0
Dollars
Jun-16
Jul-16
Aug-16
Sep-16
Oct-16
Nov-16
Dec-16
Jan-17
Feb-17
Mar-17

Video Analytics

- Goal: take advantage of existing network of traffic monitoring cameras to generate traffic data on demand.
- Approach:
  - Artificial Intelligence library for object recognition.
  - Algorithms for object tracking.
  - Innovative approach: store recognized objects using a data model that allows for flexible queries

Object Recognition



Data Structure

	Description
f_id	Index of time frame in the video file
t_id	Index of recognized object/trajectory identified through the video
c_id	Class label prediction of the object in this trajectory
s	Confidence measure of the class prediction
x_{min}	Minimum x coordinate of the bounding box in the current frame
y_{min}	Minimum y coordinate of the bounding box in the current frame
x_{max}	Maximum x coordinate of the bounding box in the current frame
y_{max}	Maximum y coordinate of the bounding box in the current frame
dx	Offset of the current center of the object from the previous location along
dy	Offset of the current center of the object from the previous location along

- Object recognition using Stampede2
- Using HiveQL to query data efficiently
- Data stored using a Hadoop distributed system or Spark

Data & Queries Samples

Raw video: video_1

Video Metadata

ID	Case 1
Description	Riverside & Faro (4/25/2017)
Length	2
FPS	30
Total Frames	3582
Resolution and Codecs	1280 x 738 H.264

Object Recognition and Tracking Results from Video

Labeled Video of video_1

Extracted background image

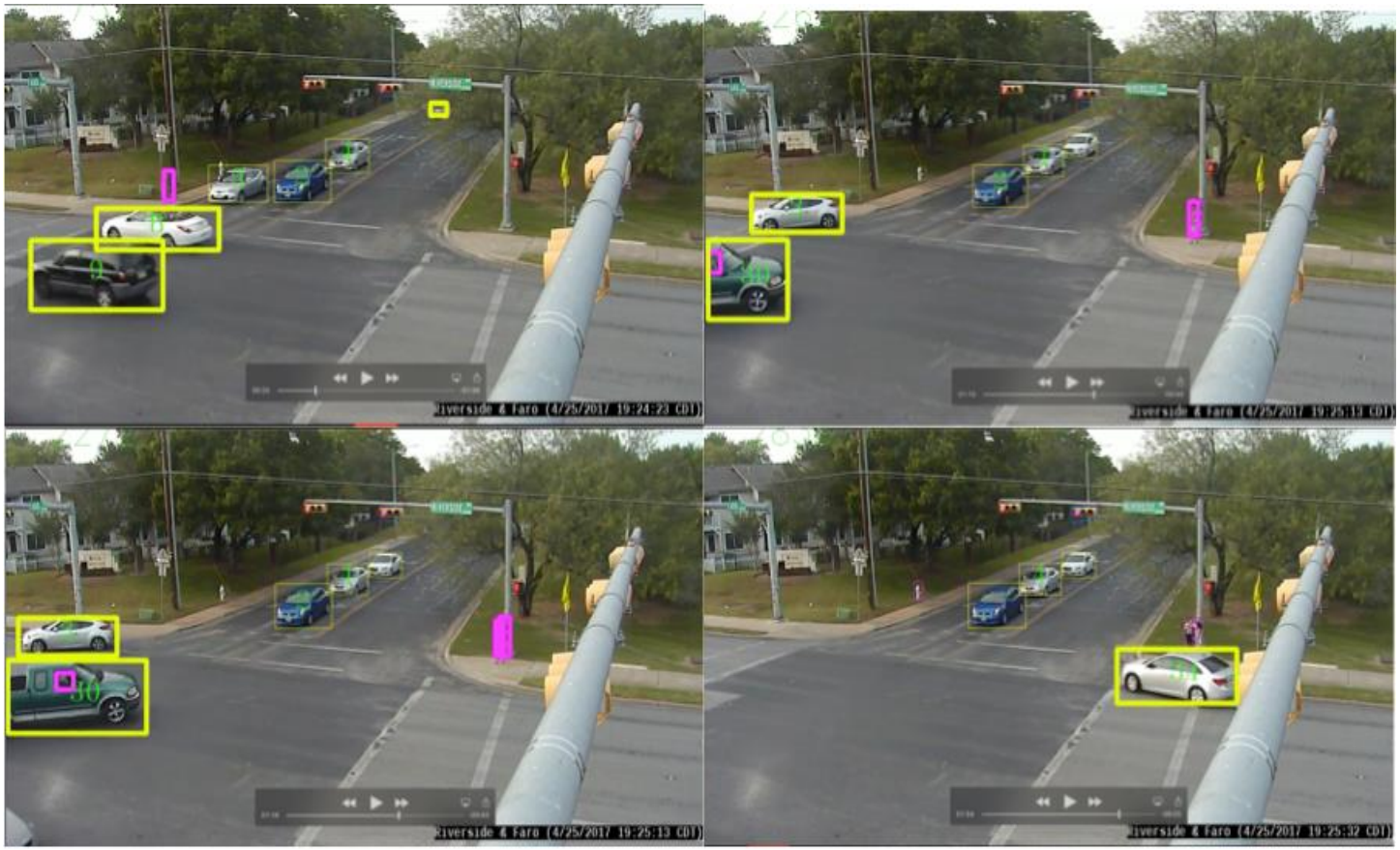
Search Objects in Video

Select: `car_id, isMoving, min(frame) as start_frame, max(frame) as last_frame`

Filter: `car_id>0 group by car_id, isMoving order by start_frame, car_id`

Search

Pedestrians & Vehicles



Outcomes & Potential Use

- Methodologies to count vehicles, possibly by vehicle type and turning movement.
- Identify frames where pedestrians in close proximity to vehicles.
- Demo website
- Applications
  - On-demand traffic counts.
  - Vision-zero: improve understanding of driver-pedestrian interactions, analyze impact of various crosswalk treatments.

Lessons learned

- Collaboration is critical
  - Understand value of data
  - Identify best collaboration schemes to accelerate innovation
  - Understand challenges and barriers
- Promoting open data and efficient data sharing policies is important to engage stakeholders and maximize data use
- Familiarity with technologies that support analysis and visualization of large datasets is needed for data exploration and innovation
- Open-data and open source may be needed for increased transferability and replicability