

# What is Data **Science**?

# Digital Sensors everywhere



# Two data analysis tasks

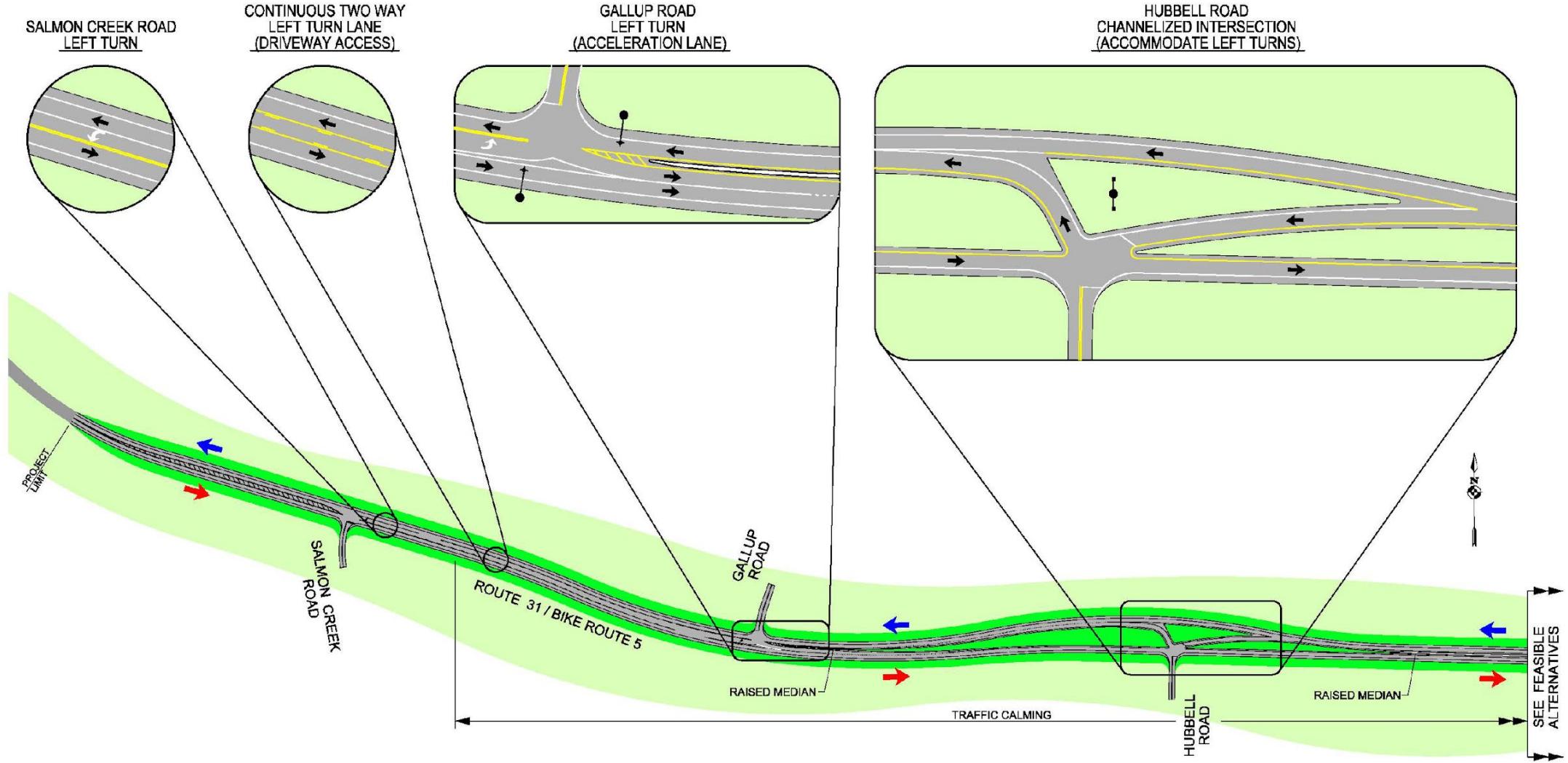
## 1. Establishing facts:

1. The water in City X is safe/unsafe to drink.
2. The average daily time spent on a smartphone is Y
3. The number of almost accidents per-day on highway 5 is Z

## 2. Making predictions:

1. Selecting ads to maximize clicks.
2. Automating/assisting medical diagnostics.
3. Choosing candidates for Loans.





LEGEND

- ← RT 31 WESTBOUND
- RT 31 EASTBOUND
- POTENTIAL HIGH MAST LIGHTING
- ↔• POTENTIAL COBRA HEAD STREET LIGHTING



A real life example  
highway traffic in California.

# Caltrans: PeMS (Performance measurement system)

Welcome, yoav ▾ [Home](#) [Help](#) [Logout](#)

**PeMS 18.0**

**State of California Every Sunday from 9 a.m. - 7 p.m., maintenance activities are performed that can impact PeMS. Please be aware that these activities can sometimes prevent users from logging in, and sometimes interfere with report generation, causes reports to be slow or time out. We apologize for any inconvenience.**

**Current Location**

Overview ▾ Facilities & Devices ▾ Performance ▾ Data Quality ▾ Events ▾

**Status Check**

**Delay by Day of Week**

Day	Last Year (veh-hrs)	Last Week (veh-hrs)	This Week (veh-hrs)
Su	~400k	~400k	~400k
M	~800k	~800k	~800k
T	~1000k	~1000k	~1000k
W	~1100k	~1100k	~1100k
Th	~1200k	~1200k	~1200k
F	~1300k	~1300k	~1300k
S	~500k	~500k	~500k

**Report Finder**

**Freeways** **Devices**

Select a freeway  Direction   
Restrict Location - Optional   
Select a report

**Freeway Details**

Directional Distance	<a href="#">30,602.1 mi</a>
Controllers	<a href="#">6,960</a>
Stations	<a href="#">18,326</a>
Detectors	<a href="#">45,147</a>
Traffic Census Stations	<a href="#">16,527</a>

**Travel Time Reliability**

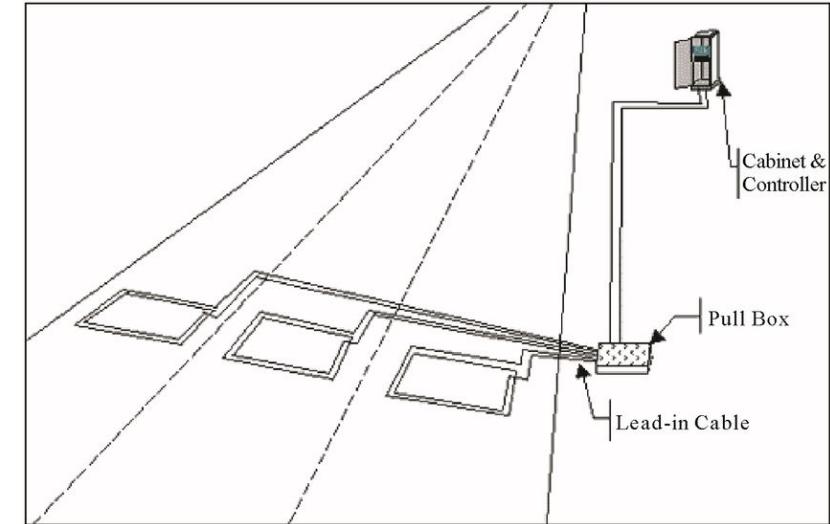
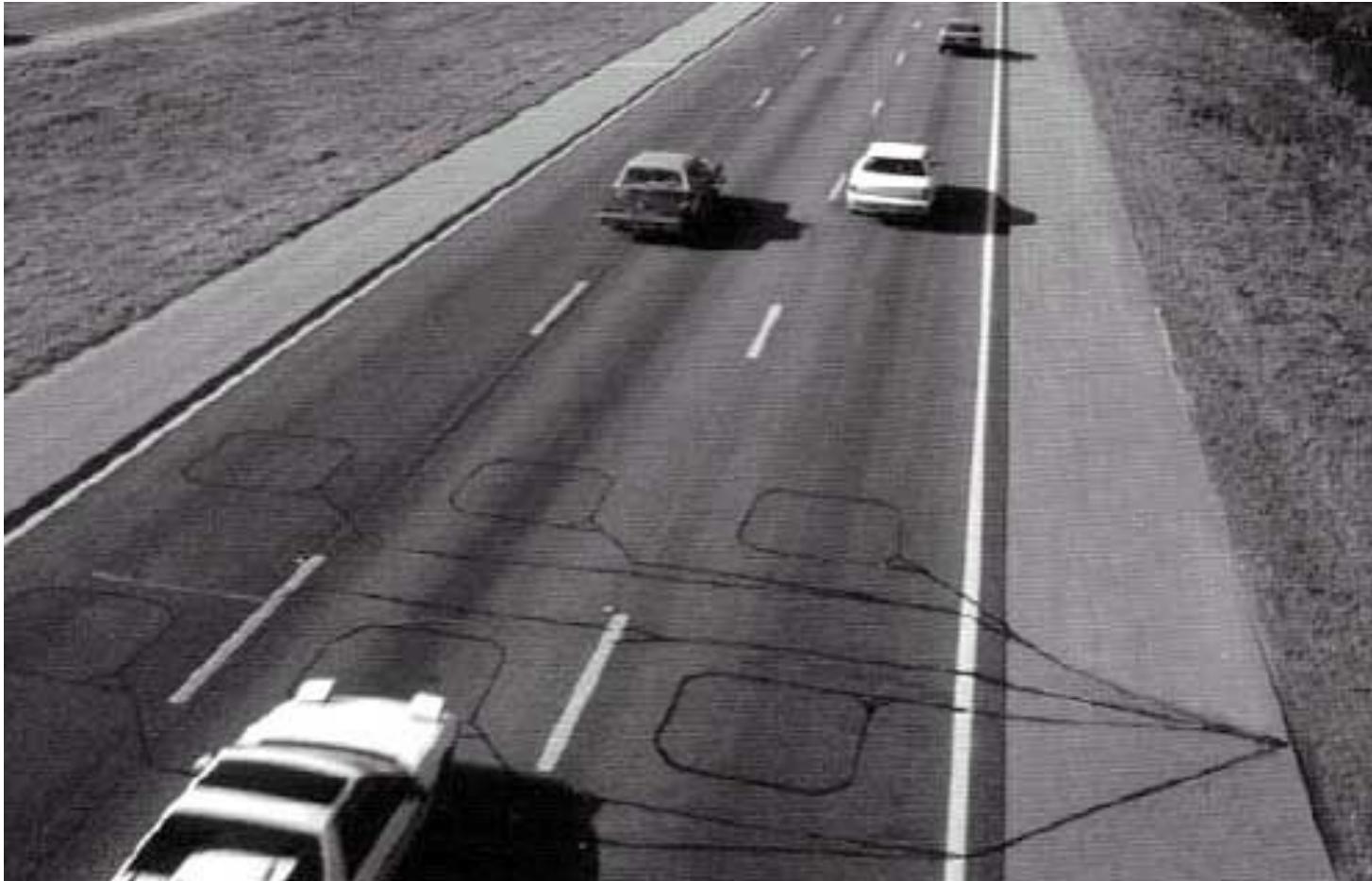
0 5-10 AM 100%

**Announcements**

**Database Improvements**  
September 12, 2018

All districts except D7 have been moved to the new

# Car Loop Detectors



- For each passing car:
  - **flow** : number of cars per unit time.
  - **Occupancy**: fraction of time that there is a car above the coil.
- ~45,000 individual detectors.
- Each detector generates a data packet every 30 seconds.

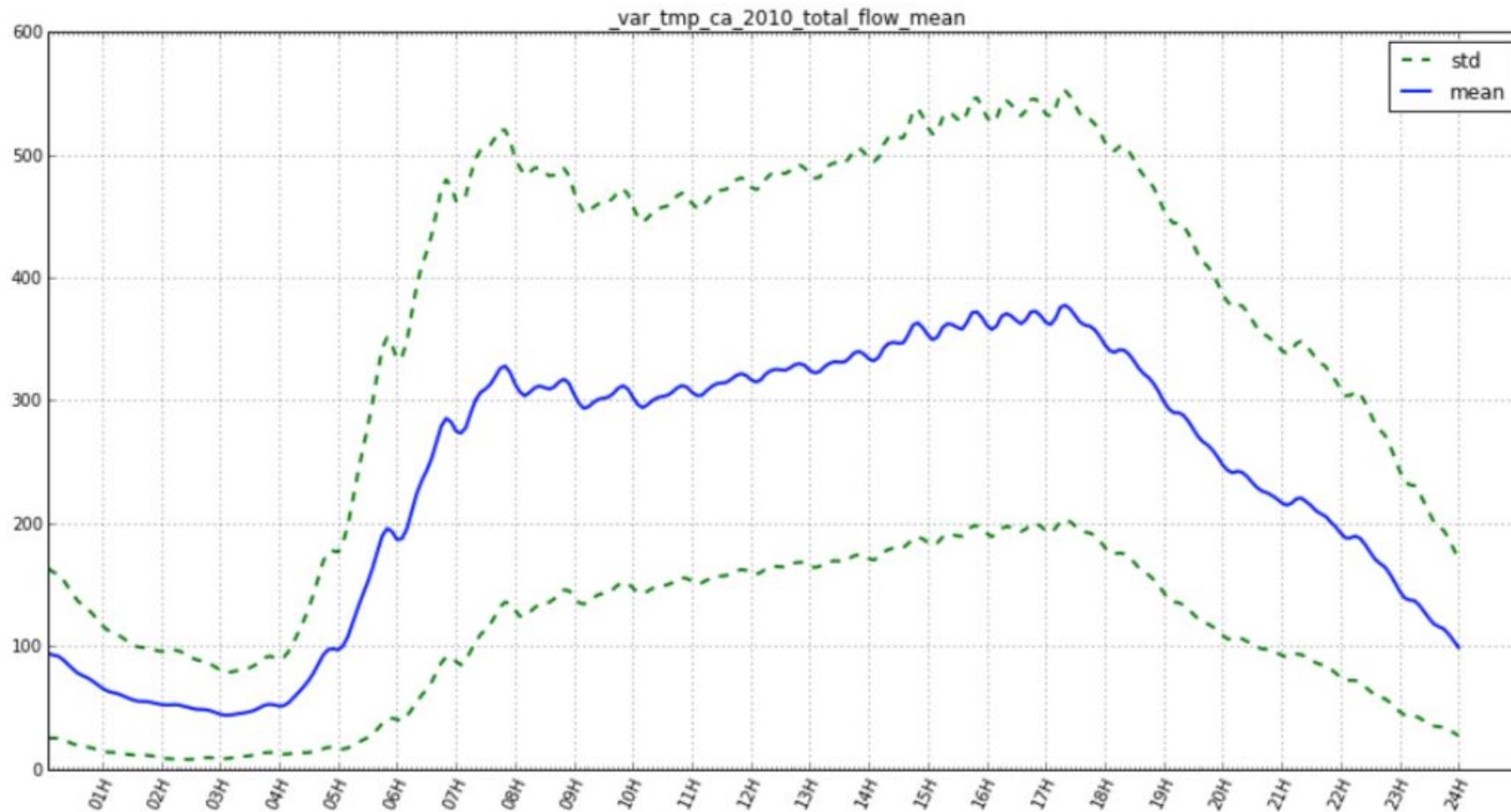
# MAS-DSE Capstone Project, 2016

Kevin Dyer, John Gill III, Conway Wong

## Caltrans PeMS

The primary data source chosen originates from the California Department of Transportation (CalTrans) Performance Measurement System ([PeMS](#)). The PeMS system is an Archived Data User Service (ADUS) site providing more than 10 years of historical traffic data collected from more than 39,000 sensors across California highways and major roadways. The site archives CalTrans loop detector data from traffic management centers (TMCs) throughout the state. The TMC's main purpose is to provide a portal which exposes analytical capabilities for various use cases such as supporting freeway operations, travel, and research. Additionally, it hosts a publically available data warehouse called the Data Clearinghouse. The Data Clearinghouse provides freely available historical data collected from sensors all across the state. The Data Clearinghouse was the source of the majority and most important data sets used for this Capstone project.

# Mean and STD of total flow (cars per minute)



# Top Eigen-vectors

