Introduction and Course Overview

CEE412 / CET 522

Transportation Data Management and Visualization

Winter 2020

Instruction Team

Instructor:

Zhiyong Cui

Office: More 101 (STAR Lab)

Email: zhiyongc@uw.edu

Office hours: 3:30 - 4:30 PM on Monday

Teaching Assistant:

Sam Ricord

Office: More 101 (STAR Lab)

Email: samuelsr@uw.edu

Office hours: 2:30 - 3:30 PM on Thursday

Instruction Time: 8:30 - 9:50 AM on Wednesday and Friday

Computer Lab: typically on Friday (MOR 001)

Today's Outline

Why is this course offered?

What is covered in this course?

How is the course organized?

Course survey

Why Is This Course Offered?

Why Is This Course Offered?

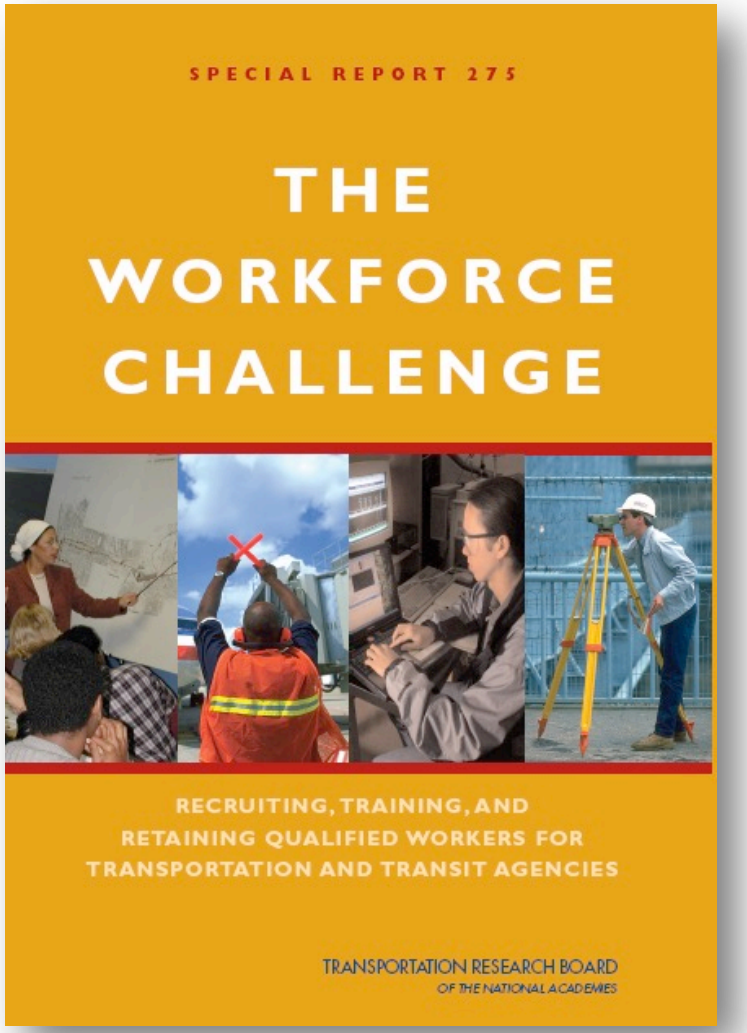
Origin of this course

Handling and managing data is necessary

Someone mentioned data science job will disappear.

The lines blur between data science and product management and other related roles at a lot of companies.

Transportation Workforce Challenge

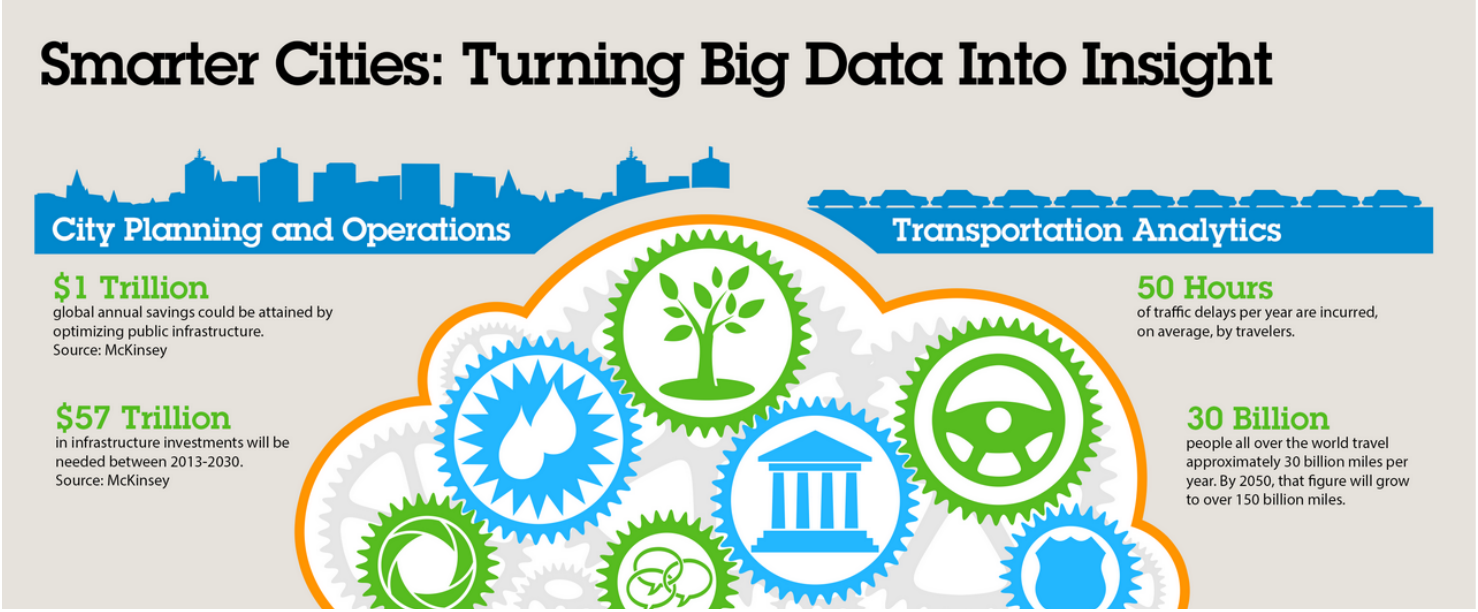


As of 2003:

- More than 50 percent of the state transportation agency workforce will be eligible to retire in the next 10 years.
- Meanwhile those qualified for transportation jobs have been decreasing over years.
- Data management and analysis are among the top desired skills for transportation workforce.

Now…

Big Data Analytics in Transportation[1]



Commercial and research oriented smart cities initiatives[2]

1. Big Data Analytics: Driving Value Beyond the Hype - Roundtable Summary. (2012). Volpe & The U.S. Department of Transportation Research and Innovation Technology Administration.
2. IBM Smarter Cities Initiative

Congestion Is a Nationwide Problem

Congestion is not just a problem for big cities:

- Congestion in small urban area peak hour travel grew from 21% to 27% of total VMT between 2000 - 2011

46% of peak period travel is congested, 64% in Seattle[1]

Seattle area fuel use attributable to congestion: more than 47 million gallons/year[1]

4.3 billion hours wasted

Indirect effects:

- air pollution, travel cost, etc.

1. Texas A&M 2012 Annual Urban Mobility Report

Data Is Critical for Modern Transportation

Modern Transportation Requires **Intelligent Transportation Systems (ITS)**

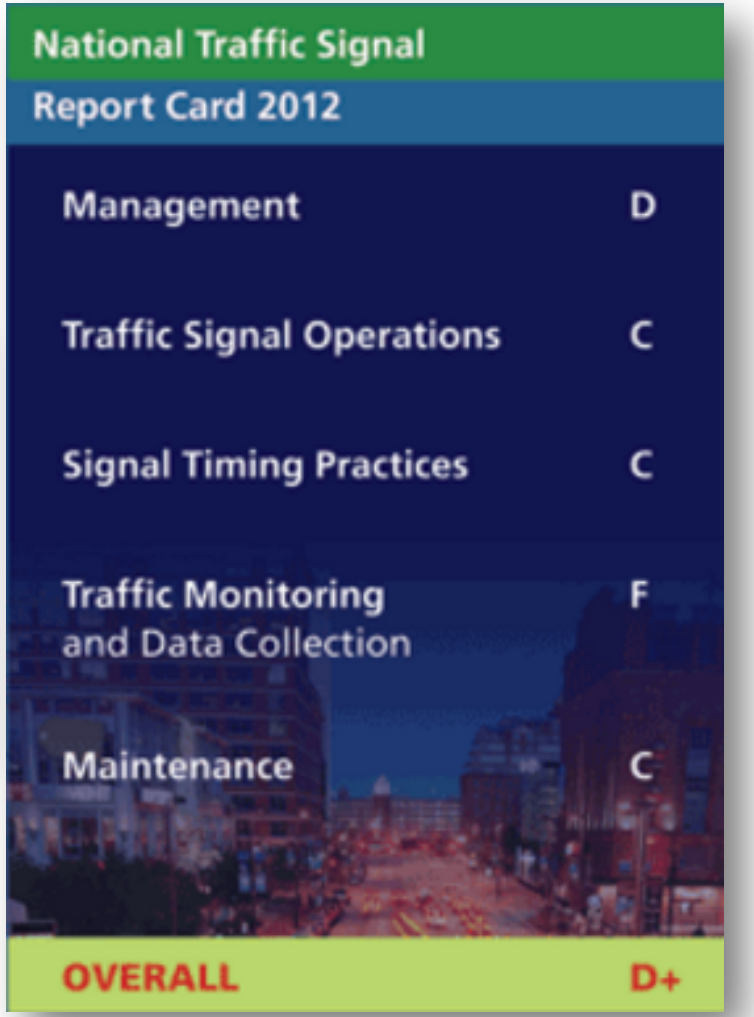
ITS Requires Intelligence → Intelligence Requires Information → Information Requires Data

Smart cities and smart transportation means:

- Ubiquitous sensing
- Data analytics
- Open data sharing
- Data-driven decision making

Data Acquisition, Management, and Interpretation Require Combination of Technologies from Transportation and IT Fields.

We Need Traffic Information



We Need Traffic Information

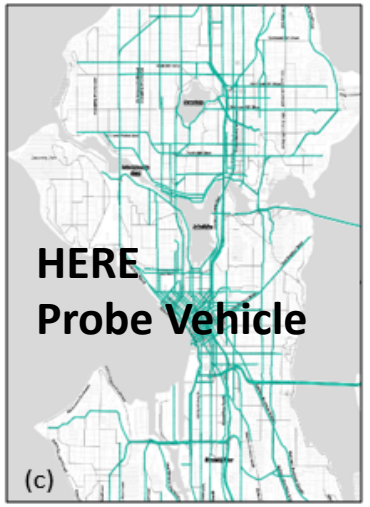
Key Arterial Performance Measurement

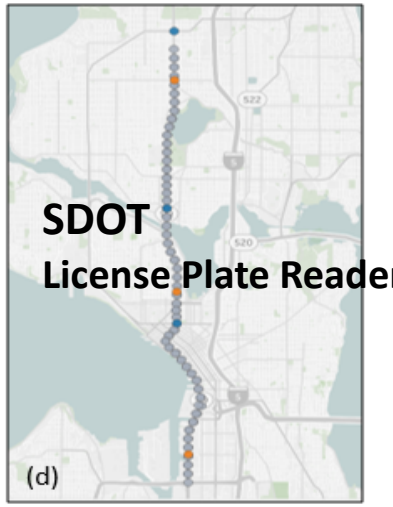
- On SR-99
- Participants:

SDOT
Seattle Department of Transportation

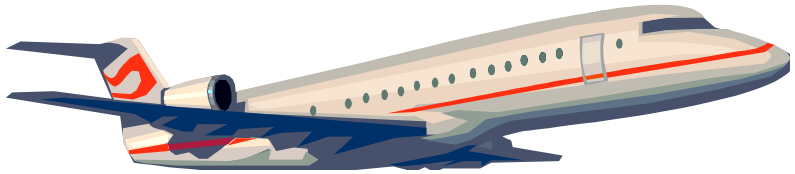
verizon

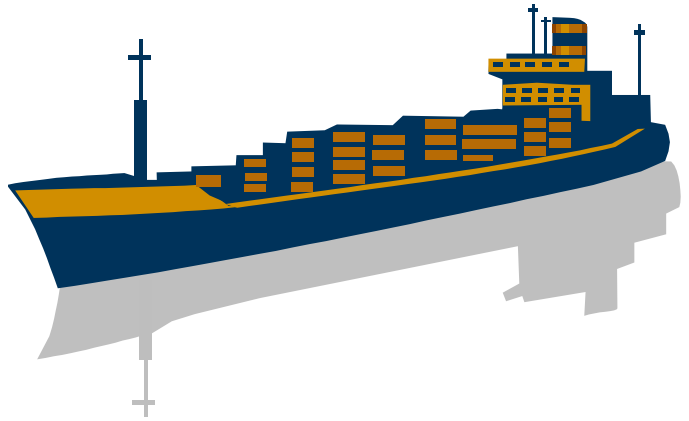
STARLab
Smart Transportation Applications & Research

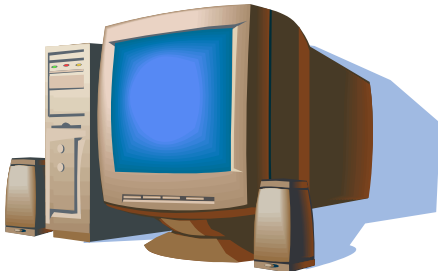




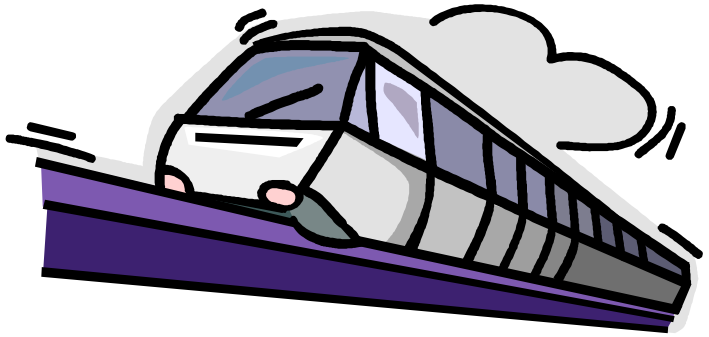
We Need Traffic Information



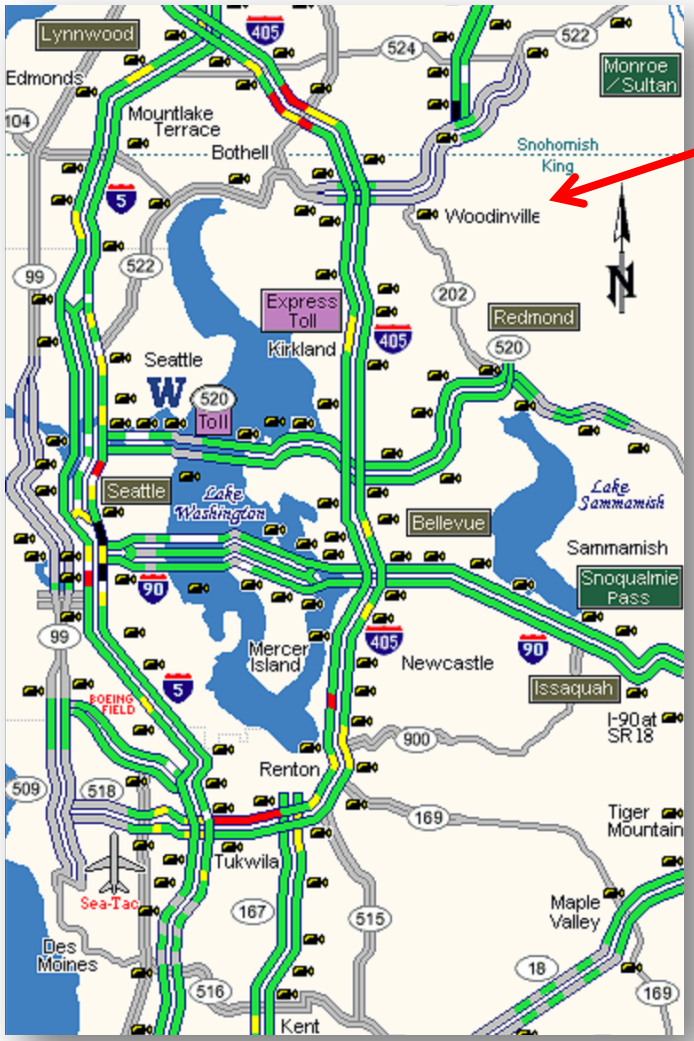




Transportation Database



Advanced Traveler Information Systems



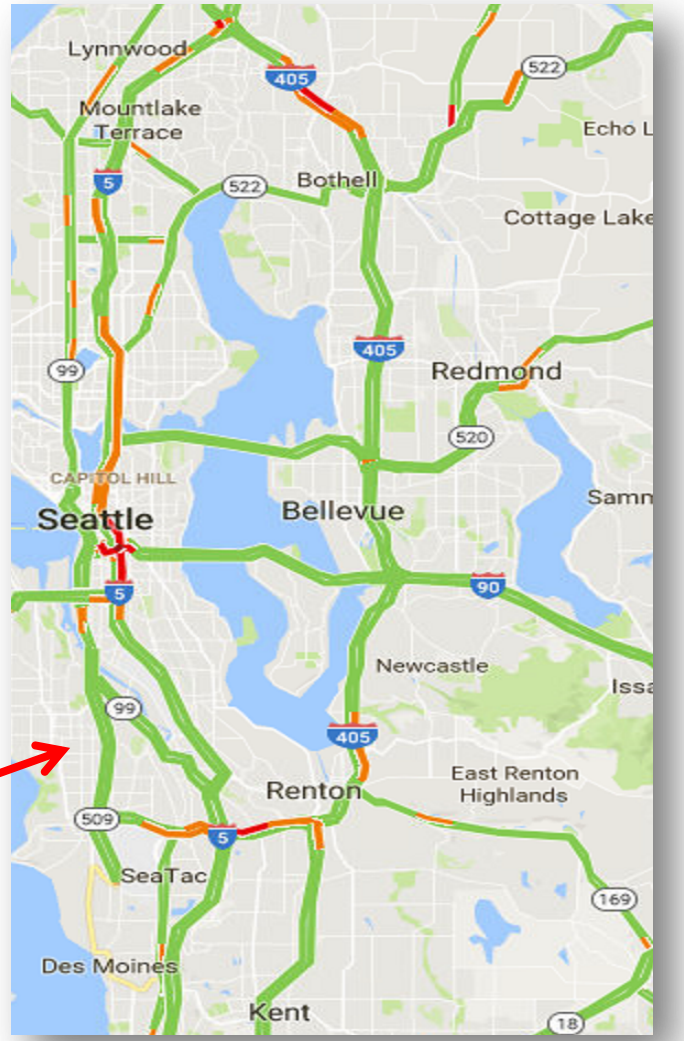
WSDOT Traffic Map

http://www.wsdot.com/traffic/seattle/default.aspx

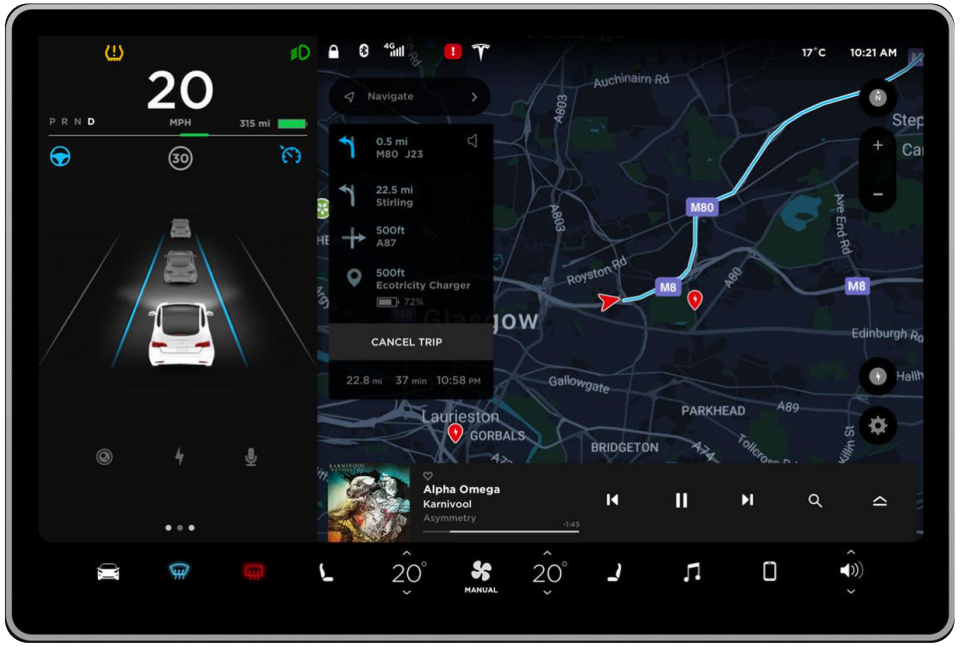
Bellevue Traffic Cameras

http://www.cityofbellevue.org/trafficcam/

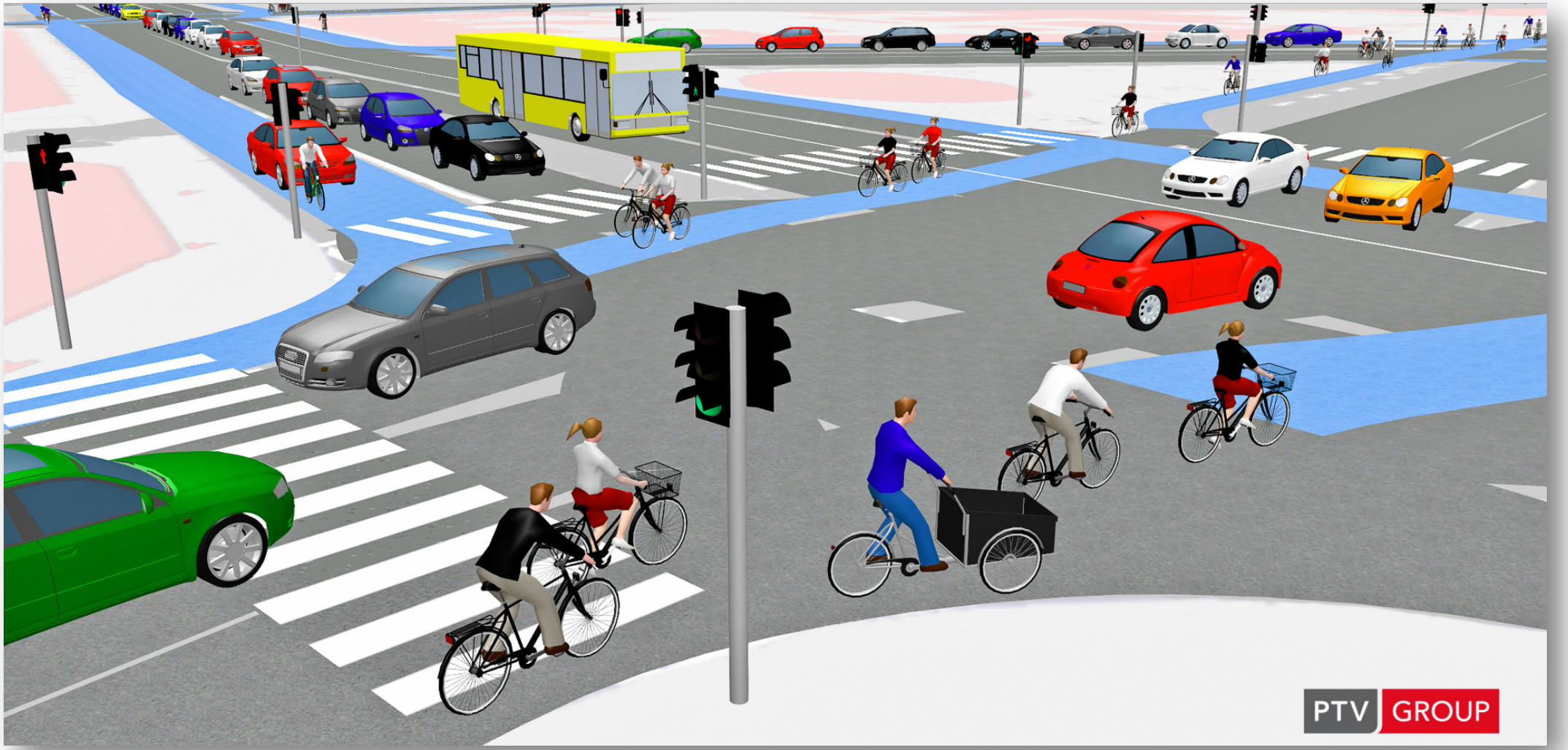
Google Traffic Map



Importance of Transportation Data



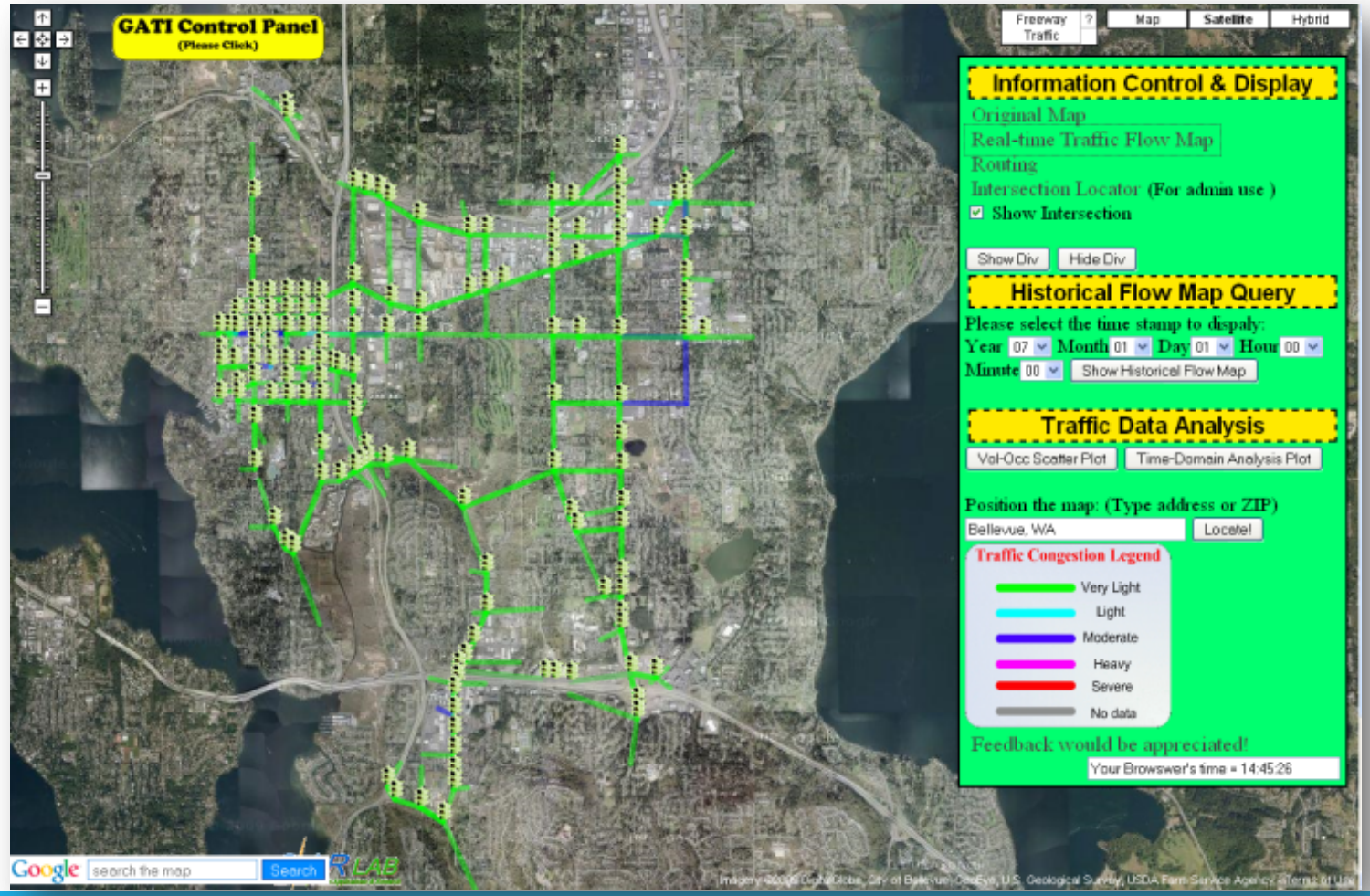
Importance of Transportation Data



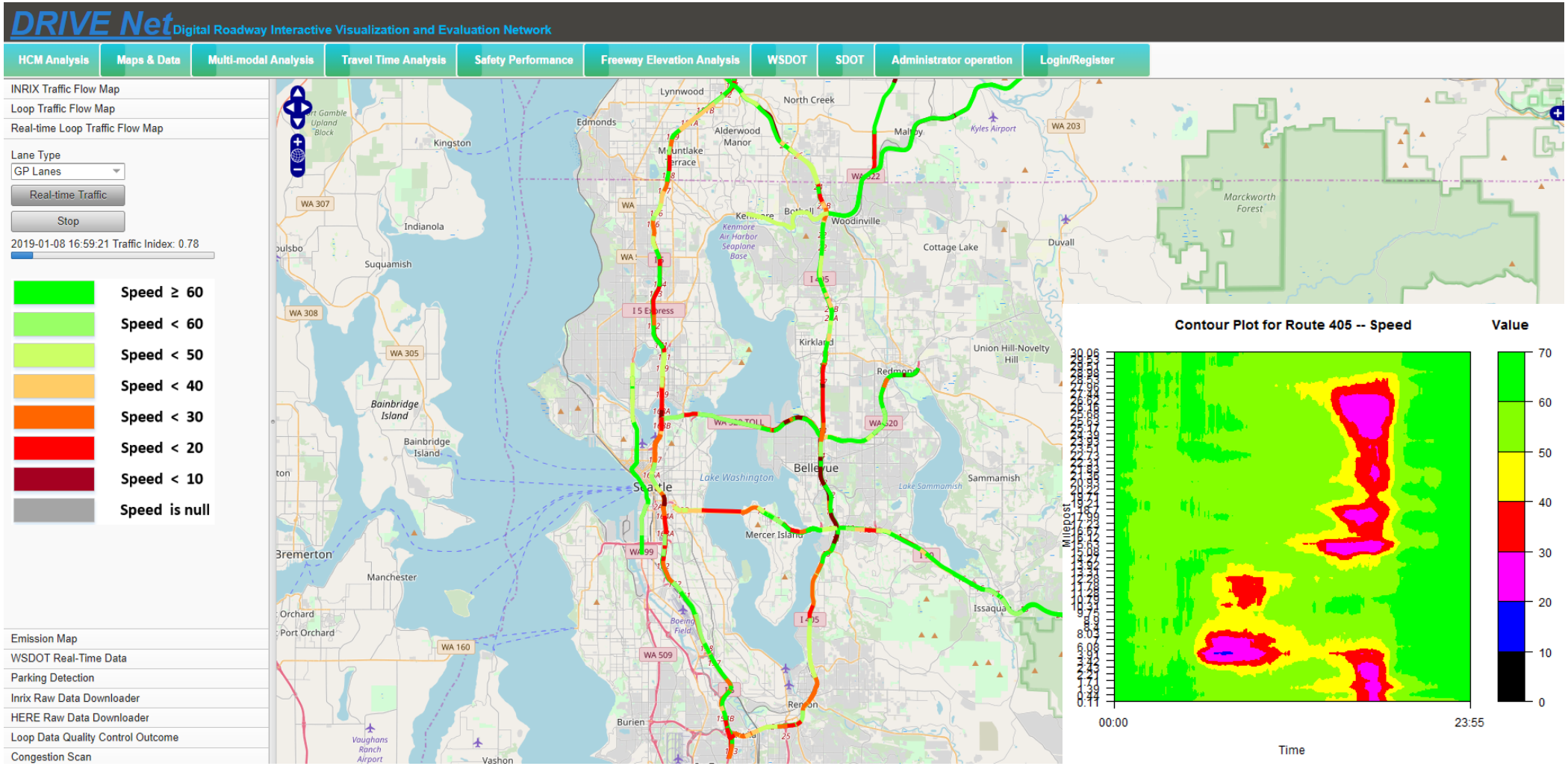
Importance of Transportation Data

Importance of Transportation Data

Google Map Based Arterial Traffic Performance System



Importance of Transportation Data

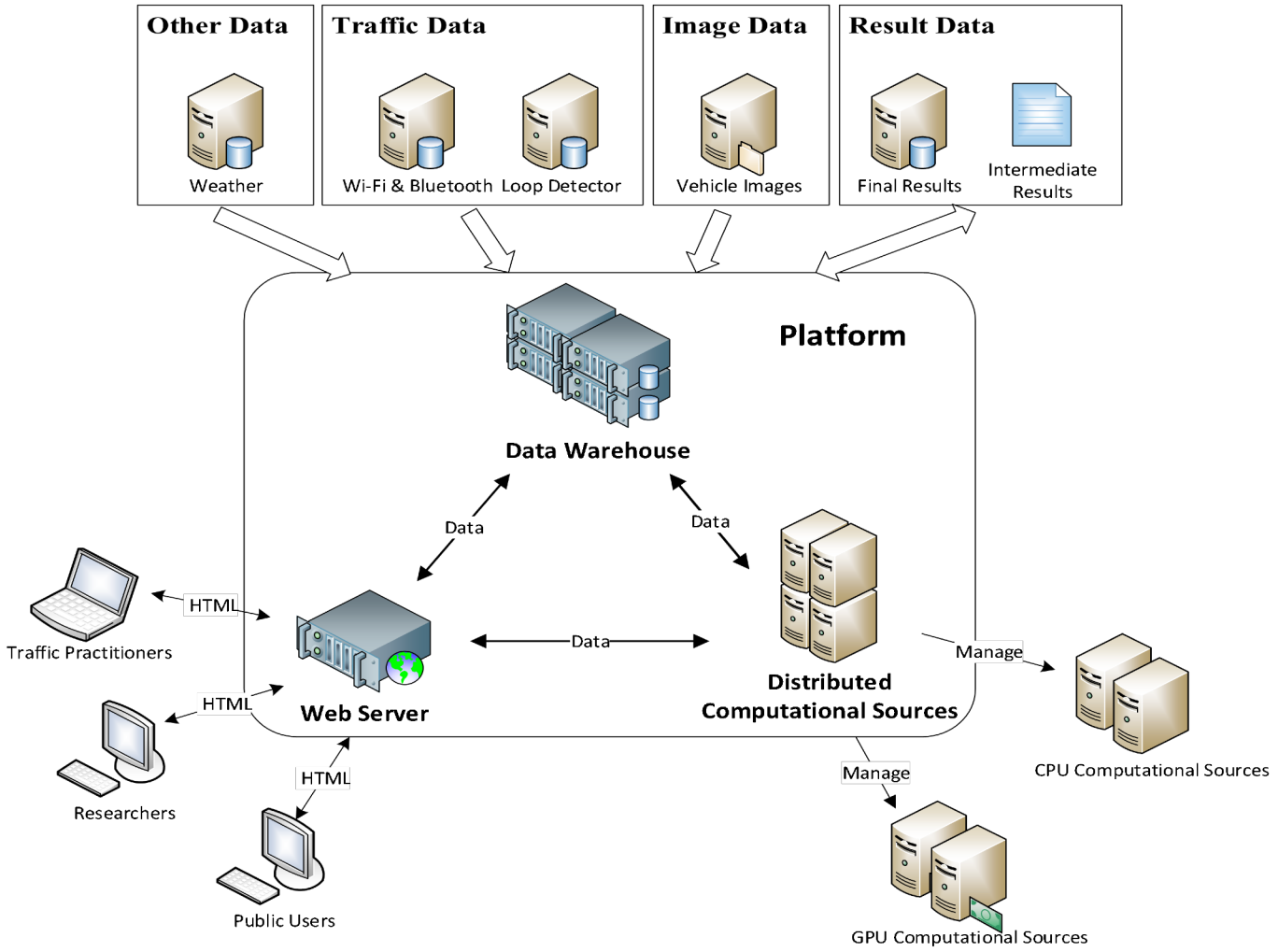


Source: http://www.uwdrive.net/

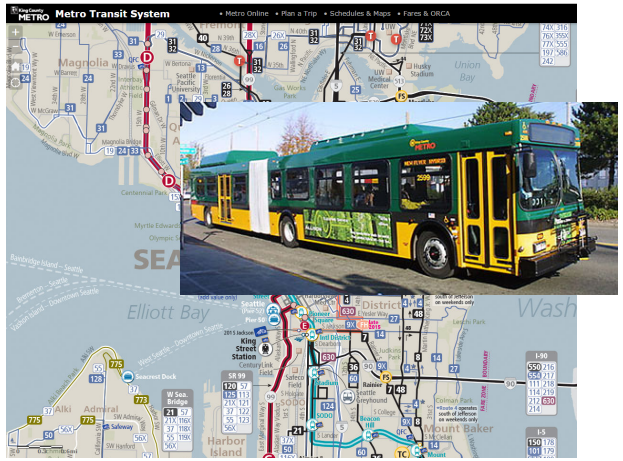
Importance of Transportation Data

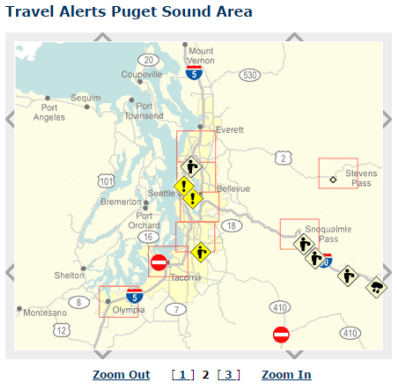
Artificial Intelligence
Transportation Platform

- Collect
- Process
- Input the data into proper AI models
- Visualize

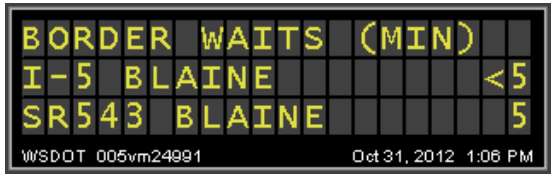


Data Is Crucial for the Success of ITS

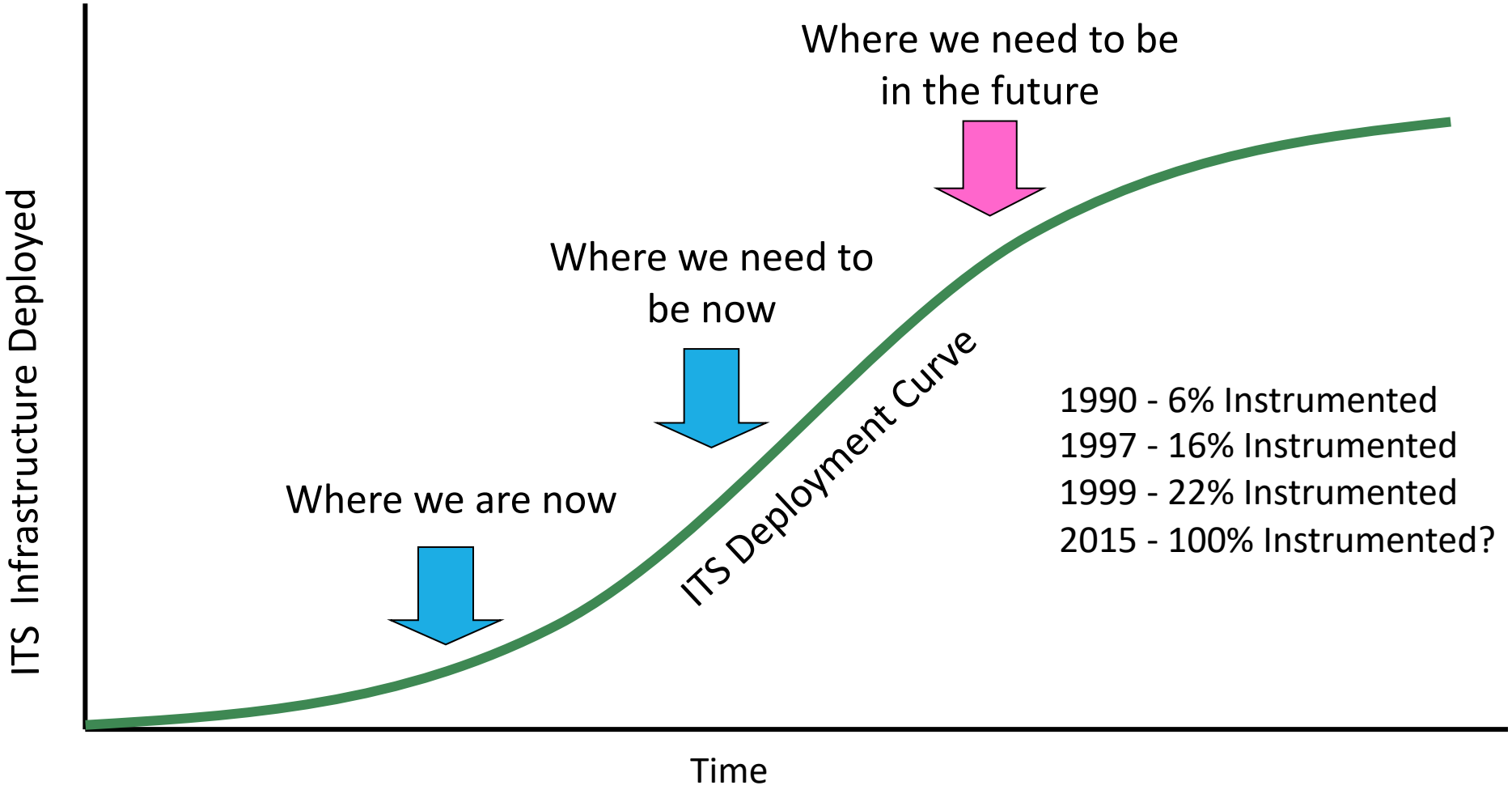




Quality Data

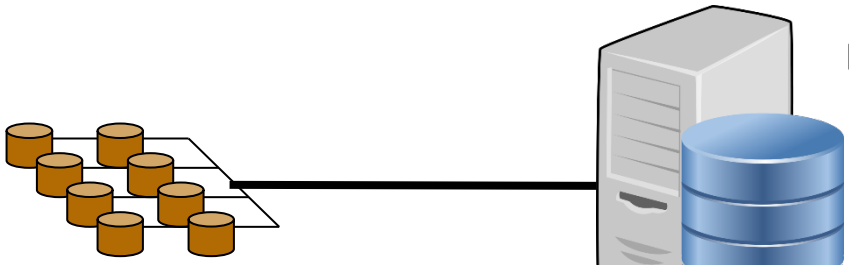


When will We Have a Fully Instrumented System?



Source: Christine Johnson, 2001

Where Are the Data From?



Data Server

Data Archive

Conventional sources
Manual collection
Accident records
Construction records
GIS Data

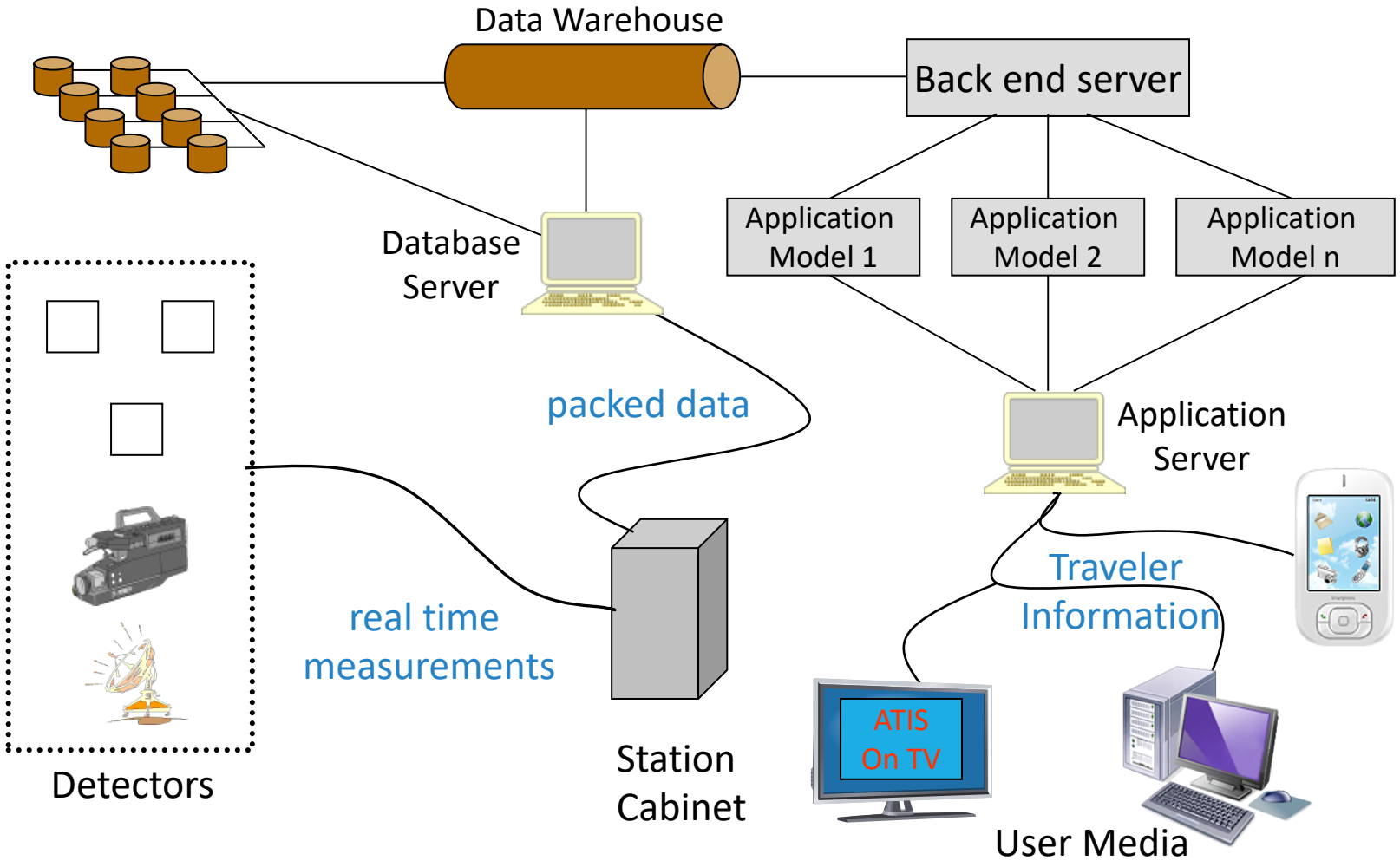
Traffic Detection Controller

Sensors

Non-conventional sources:
Mobile devices
Commercial data companies
Self-reporting
Remote sensing

Example: Architecture for ATIS

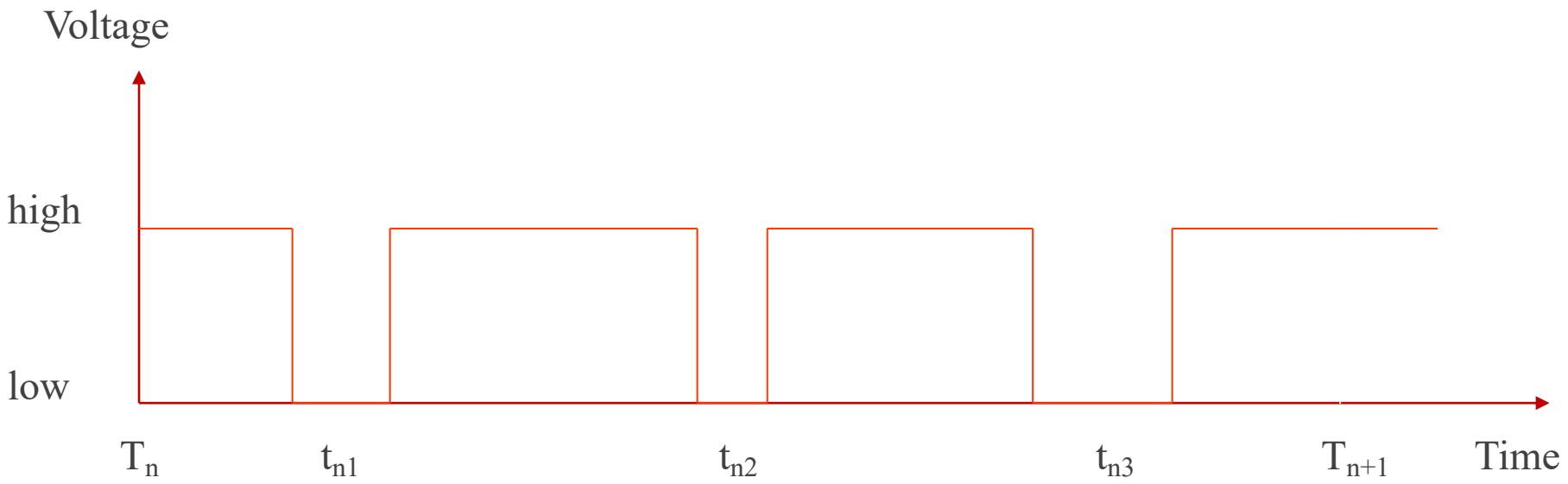
ATIS: Advanced Traveler Information System



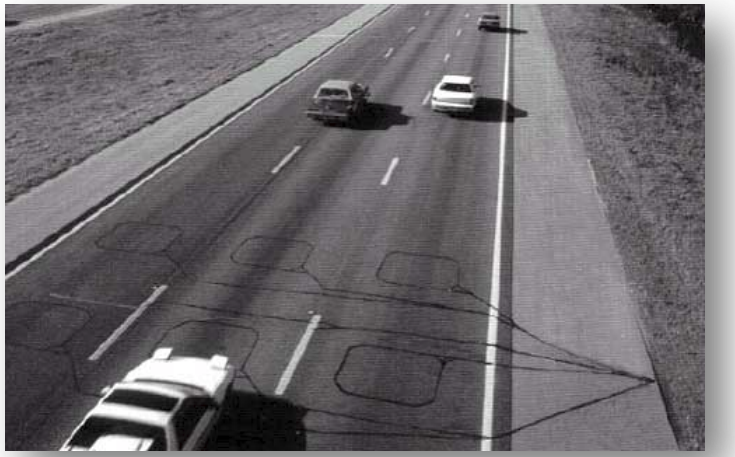
Why Have Intermediate Procedures?

Direct outputs of transportation sensors may not be in a form easy to understand.

For example, a loop detector typically outputs nothing but a series of “high” or “low” voltage that can be represented in binary form.



Single Loop Measurements Are Very Useful



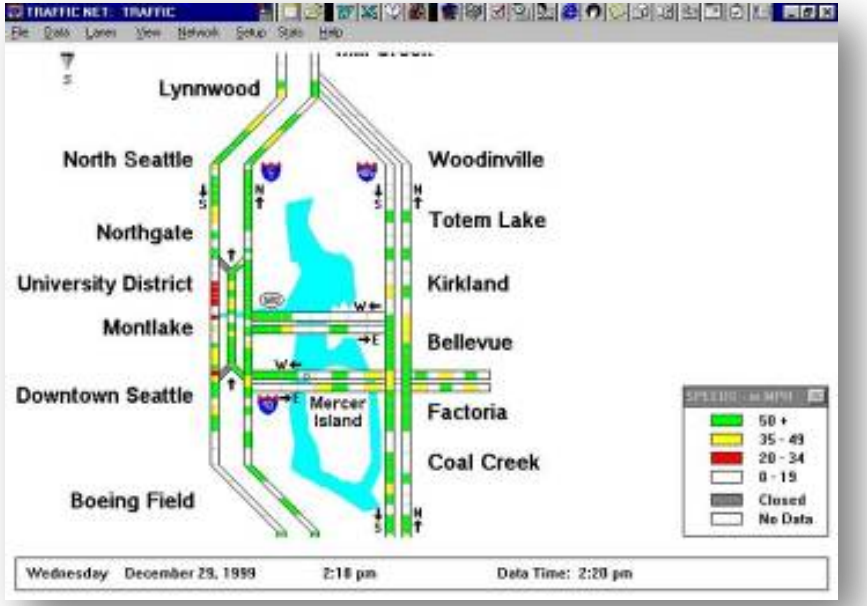
Example of Raw Data:

00001010
00001010
00001000
00001000
00001100
00000100

Processed Loop Data:

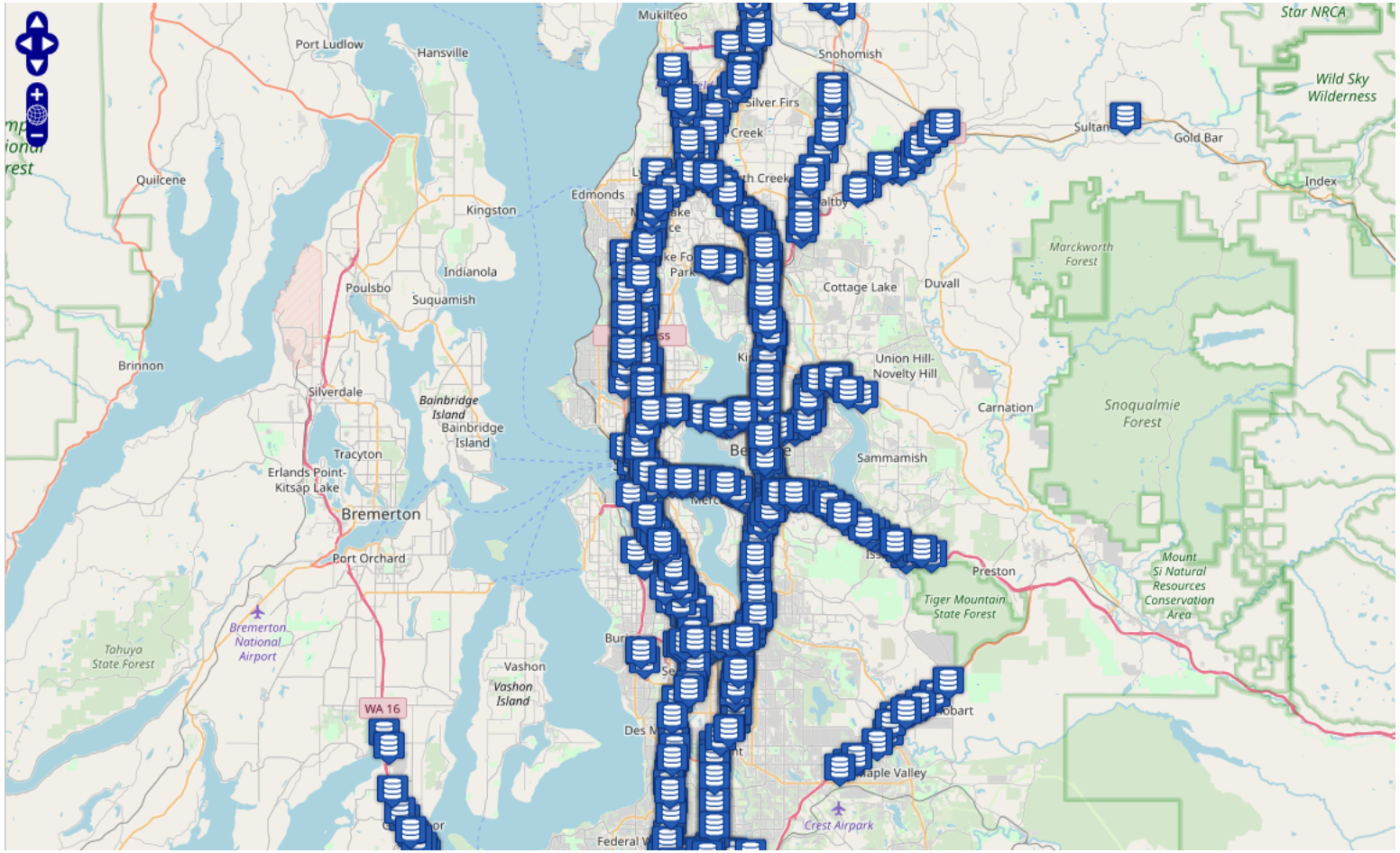
Time	Volume	Occupancy	Speed
00:00:00	2	1.36%	68.2
00:00:20	1	0.06%	70.2
00:00:40	3	1.90%	67.4

Many On-Line ATIS are based on single loop data!

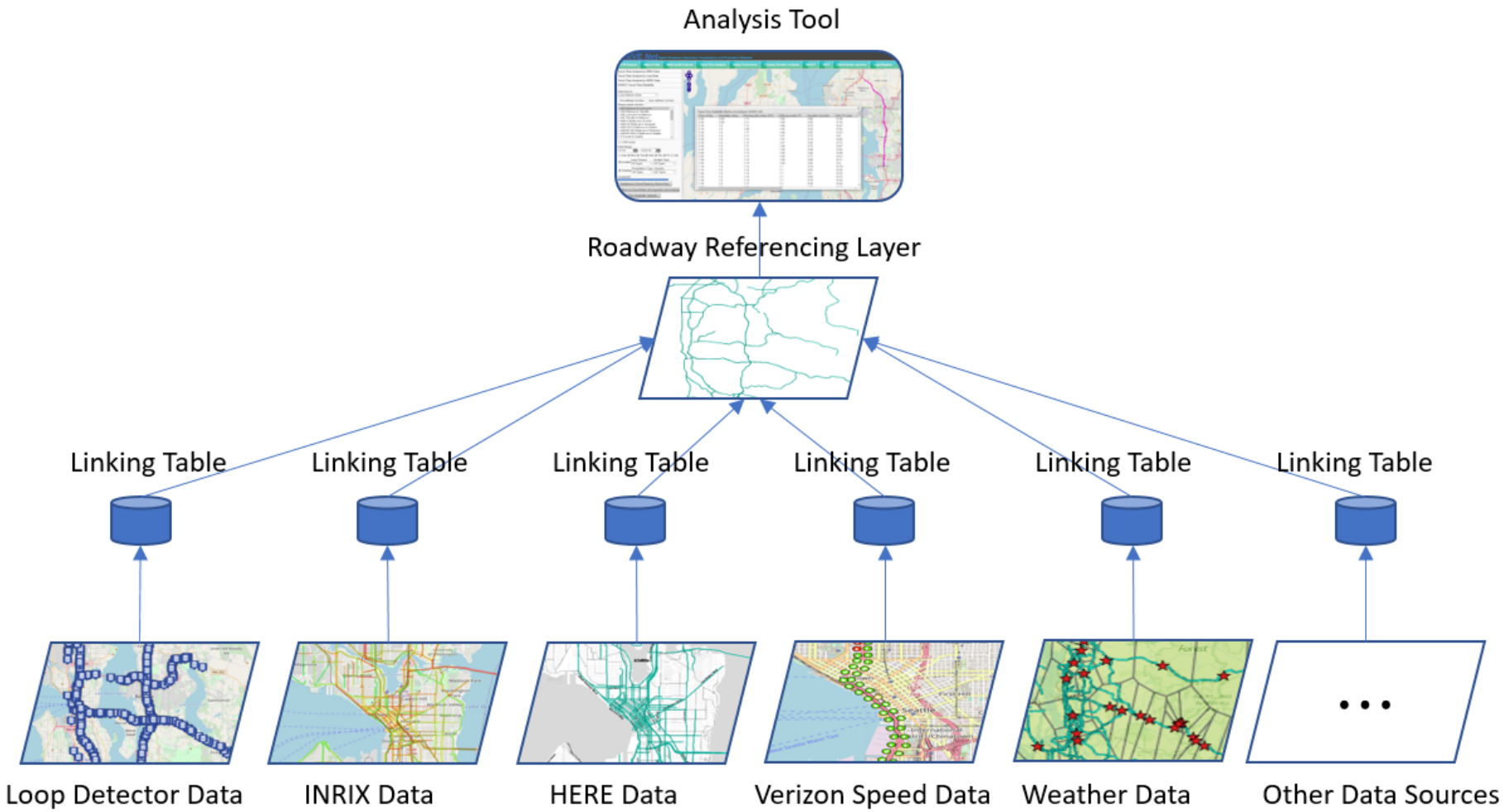


Loop Detectors in Greater Seattle Area

Over 7000 loop detectors in WSDOT's northwest region alone.



Combine Other Data Sources



How To Handle These Data?

If loop measurements are archived every 20 seconds, and each record requires 20 bytes to store it, then 8,000 single loop detectors require

24 * 3600 / 20 * 20 * 8000 = 660 MB storage space per day!

Remember this is just data from loop detectors on freeways. We have also probe vehicle data, accident data, freight data, video data, …, to deal with!

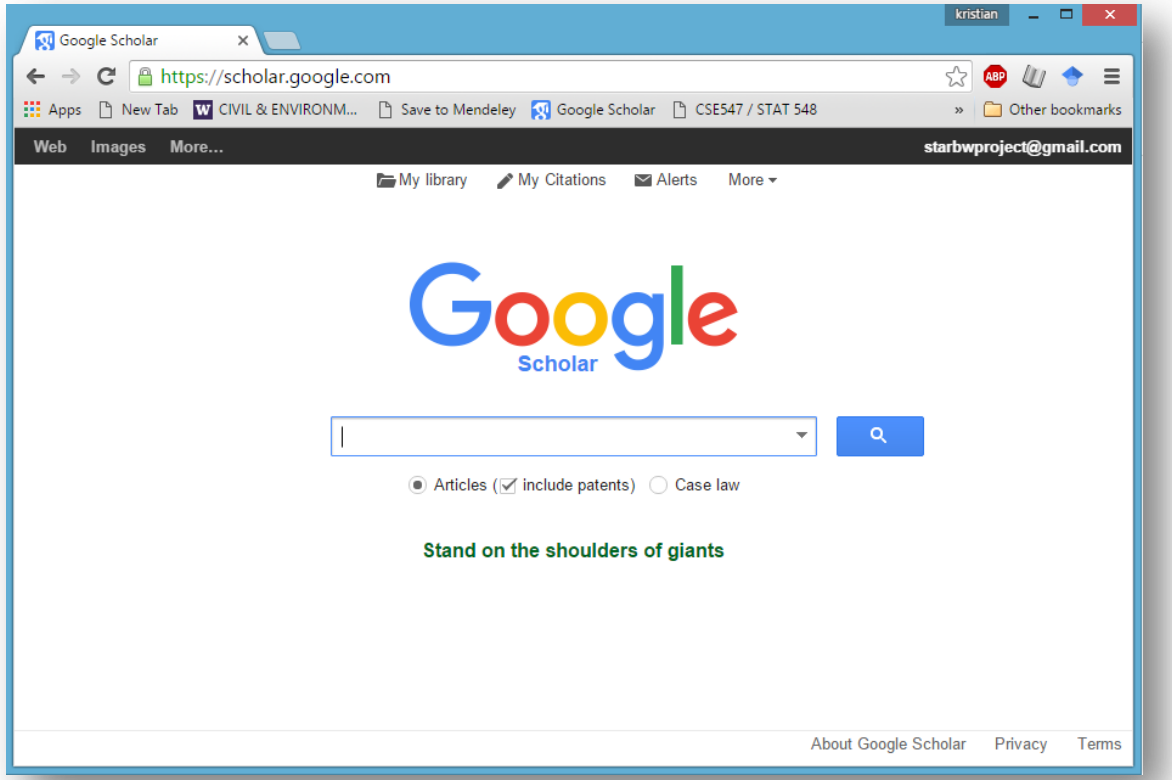
What about Other Sources of Data?

Example: INRIX

- 175 million devices
- 4 million miles of road
- 40 countries
- Terabytes of data per day
- Real time traveler information, analytics, and decision support

Source: https://Inrix.com

Benefits of Well-Managed Data?



Efficiency for data processing and query can be very different.

Benefits of Well-Managed Data?

DRIVE Net:

- WSDOT has massive data resources, but struggles to apply much of it in performance reporting and decision support
- Organization, data integration, spatial conflation, data sharing, etc. are disjointed and often done “as needed” for specific tasks
- To WSDOT, DRIVE Net represents a an automated data management and integration platform

Well managed data: Maintenance and analysis is simplified

Poorly managed data: often the work associated with making it useable outweighs the benefit of using it

Extracting Useful Information

Transportation Data Analysis

- Data analytical methods for transportation data analysis

Transportation Data Visualization

- Proper data visualization tools to generate charts, graphs, maps, etc.

Data Visualization

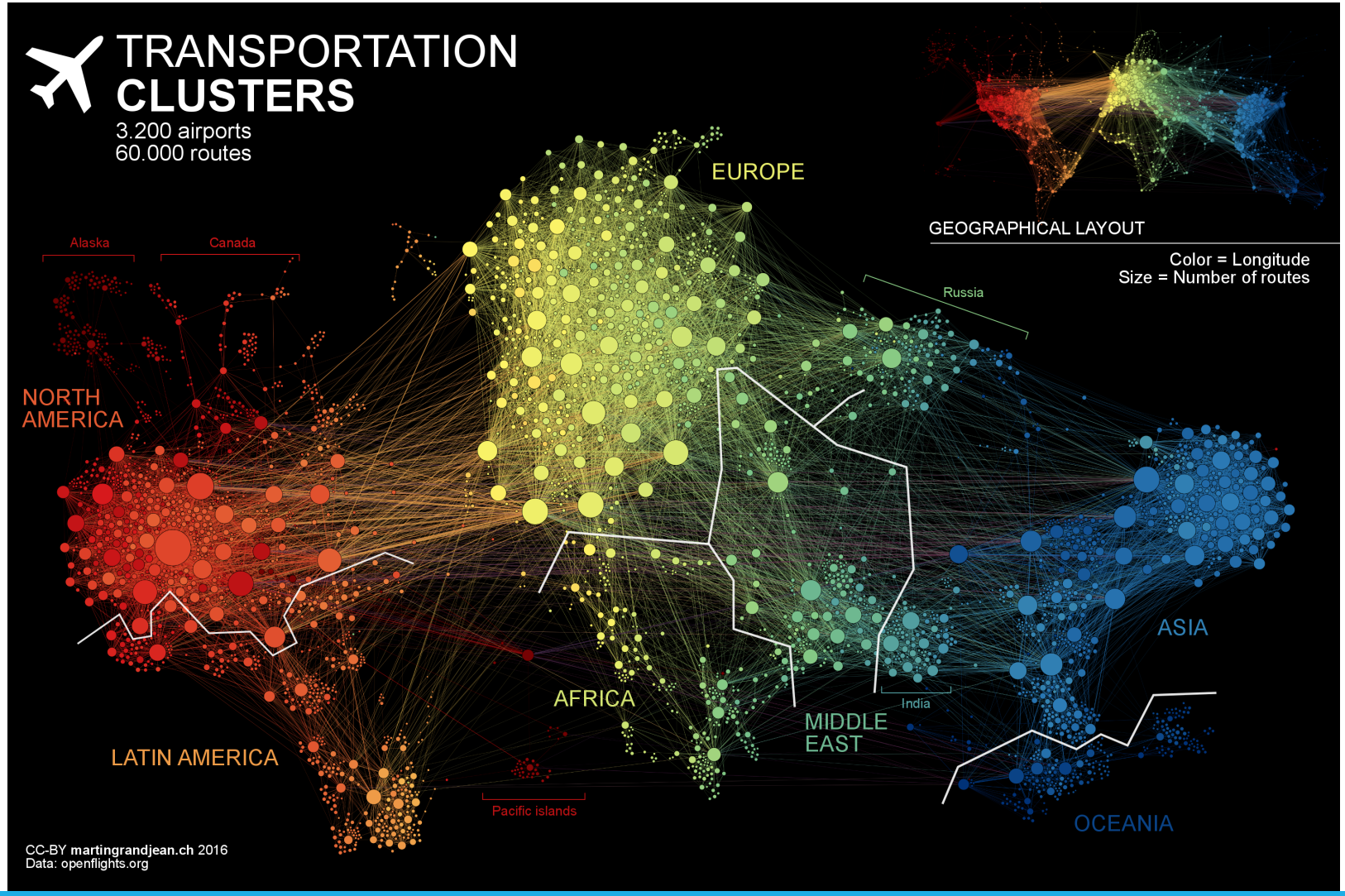
Data Visualization

- Graphical representation of information and data
- Common types:
  - Charts, Tables, Graphs, Maps, Infographics, Dashboards

Benefits from Data Visualization

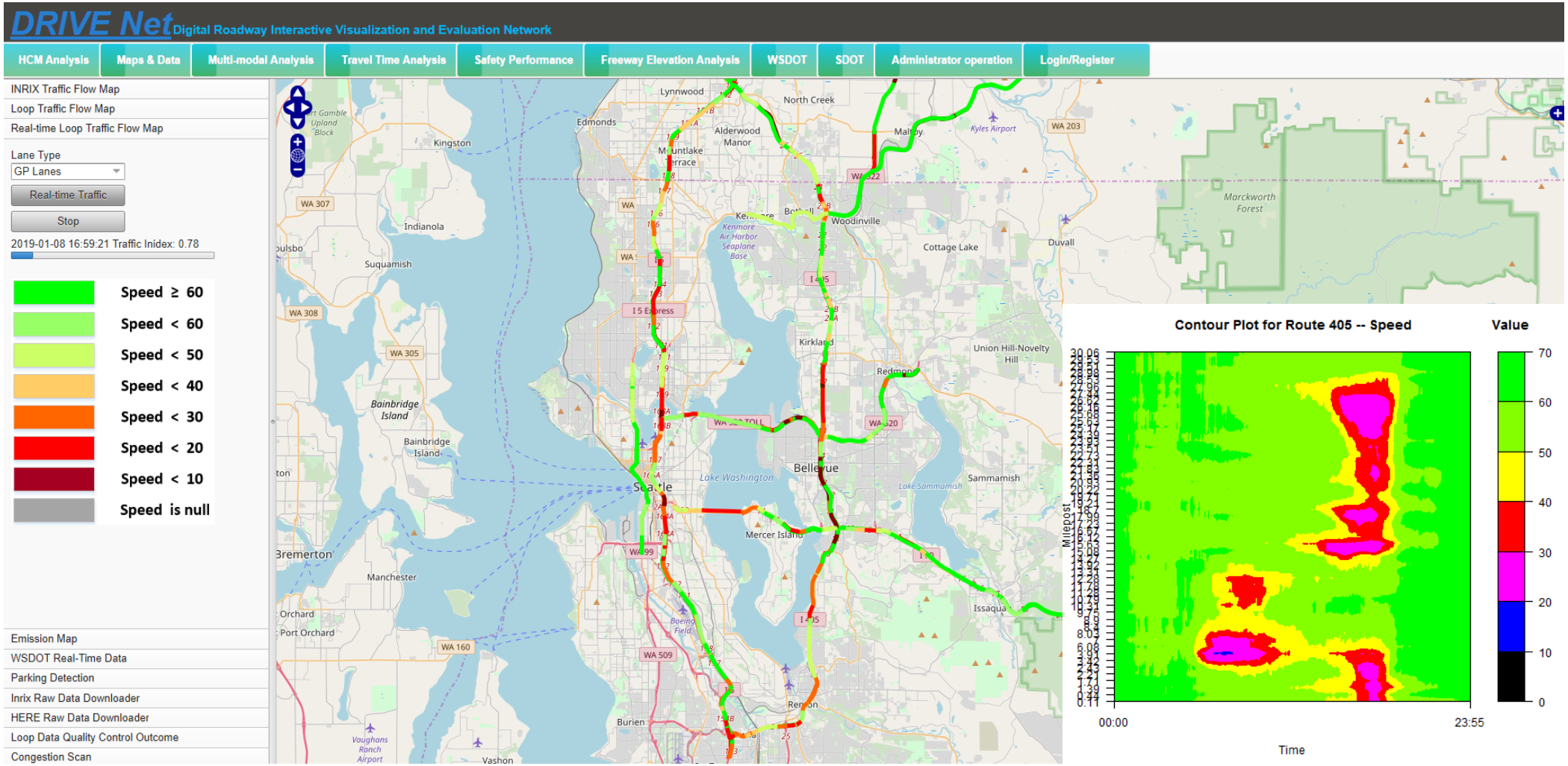
- Pattern discovery
  - Trends, outliers, patterns in data
- Efficient to see the big picture
- Support the decision-making process

Importance of Data Visualization



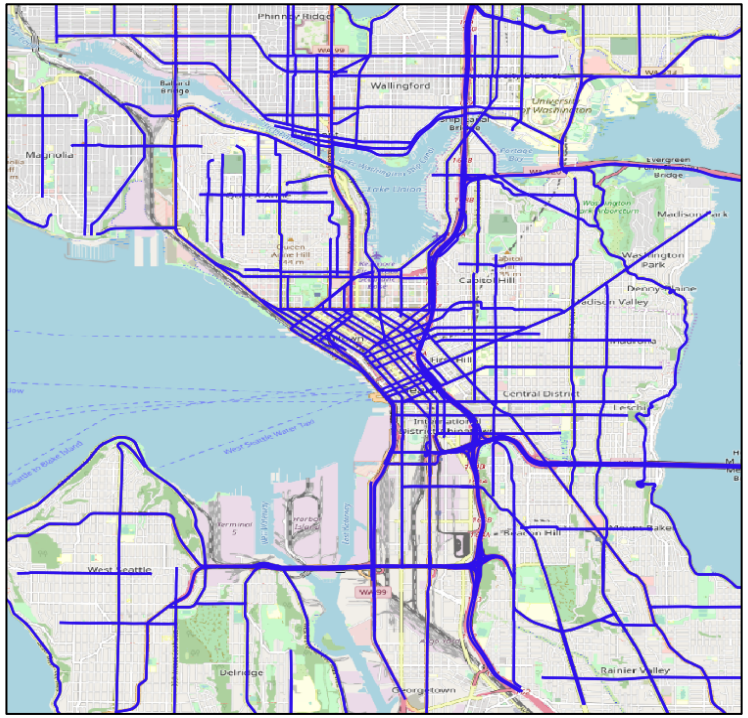
https://www.visualcapitalist.com/air-traffic-network-map/

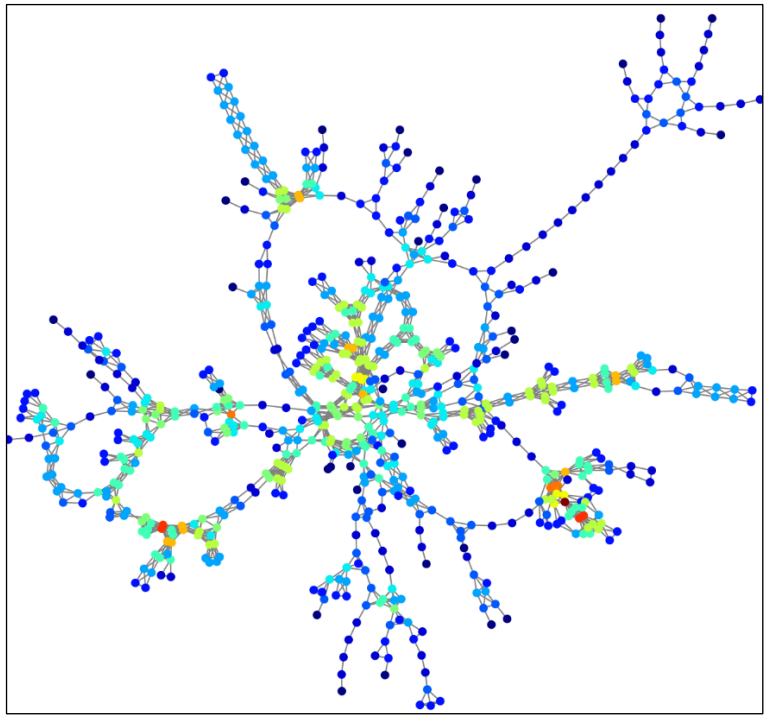
Importance of Data Visualization



Source: http://www.uwdrive.net/

Importance of Data Visualization





That's Why We Offer CEE 412/CET 522

- The quantity and variety of data sources available for transportation management and decision making is rapidly increasing
- Storing and retrieving transportation data efficiently is crucial
- Proper visualization of processed data and analysis results can benefit the decision-making process

What Is Covered in This Course?

Data Management Concepts and Tools

- Relational Database Design and Management
- Structured Query Language (SQL) - special purpose database programing language
- Microsoft SQL Server – Enterprise database management system

Data Visualization

- R - a software package used for statistical analysis
- Data visualization concepts and tools
- Data Sharing and visualization using R Shiny package
- Creating a data pipeline: analysis, visualization, and communication

How Is This Course Organized?

Week	Day	Date	Topic	Room	Notes	Homework
1	Wed	Jan. 8	Introduction and Course Overview	More 220	Course Survey	
	Fri	Jan. 10	Excel Data Model Practice	More 220		A#1 Out
2	Wed	Jan. 15	No Class (TRB Meeting)	More 220	TRB Week	
	Fri	Jan. 17	Introduction to Databases	More 220		
3	Wed	Jan. 22	E/R Diagram	More 220		A#1 Due, A#2 Out
	Fri	Jan. 24	Relational Data Model	More 220		
4	Wed	Jan. 29	Structured Query Language (SQL) I	More 220		A#2 Due, P#1 Out
	Fri	Jan. 31	SQL Practice	More 001	Exercise	
5	Wed	Feb. 5	Structured Query Language (SQL) II	More 220		
	Fri	Feb. 7	SQL Practice	More 001	Exercise	A#3 Out
6	Wed	Feb. 12	Midterm 1	More 220		
	Fri	Feb. 14	Advanced SQL & Introduction to R	More 220		P#1 Due, P#2 Out
7	Wed	Feb. 19	Transportation Data Analysis	More 220		
	Fri	Feb. 21	R Practice	More 001	Exercise	A#3 Due, A#4 Out
8	Wed	Feb. 26	Data Visualization	More 220		
	Fri	Feb. 28	Data Visualization in R & Shiny	More 220		
9	Wed	Mar. 4	Data Visualization & Shiny Practice	More 001	Exercise	A#4 Due
	Fri	Mar. 6	Guest Lecture - Data Management & Visualization	More 220		
10	Wed	Mar. 11	Midterm 2	More 220		
	Fri	Mar. 13	Shiny Practice	More 001	Exercise	
11	Wed	Mar. 18	Final Project Presentations (8:30-9:50 AM)	More 220		P#2 Due

Assignments and Projects

Written assignments

- Assignment #1: Data Analysis Using Excel (exercises)
- Assignment #2: Database Modeling (SQL basics)
- Assignment #3: Database Modeling (Advanced SQL)
- Assignment #4: Data Analysis (SQL + R)

Around 6 in-class exercises

Submit online through Canvas as separate files (please don't submit zip files)

Projects: Team projects with functional products

- Project #1: Excel + SQL Server Project
- Project #2 (Final project): SQL + R + Shiny Application Development

Policy

No late assignment/project will be accepted.

If extreme circumstances come up, it needs to be discussed before the assignment due date.

Class attendance is expected, not required.

Grading:

- Assignments: 30%
- Midterm Exam: 30%
- Projects: 40%

Other Class Notes

Canvas – everything will be uploaded before lectures/exercises

- Except for the Question & Answer discussion board: Piazza

Readings will be assigned as needed

No class on next Wednesday (because of the annual TRB meeting)

- Optional online quiz for extra credits.

Assignment 1

- Complete and turn in Exercise 1 (with some simple questions).

Instruction Team

Instructor:

Zhiyong Cui

Office: More 101 (STAR Lab)

Email: zhiyongc@uw.edu

Office hours: 3:30 - 4:30 PM on Monday

Teaching Assistant:

Sam Ricord

Office: More 101 (STAR Lab)

Email: samuelsr@uw.edu

Office hours: 2:30 - 3:30 PM on Thursday

Instruction Time: 8:30 - 9:50 AM on Wednesday and Friday

Computer Lab: typically on Friday (MOR 001)

When you send emails to the instructor/TA, I would recommend you name the email begin with [CEE412] or [CEE412_CET522].