PORTAL

A Case Study

Dr. Kristin Tufte
(tufte@cecs.pdx.edu)
Mark Wong
(markwkm@postgresql.org)

Linux Plumbers Conference 2009

September 23, 2009

Overview

- What is PORTAL?
- How PORTAL works
- Improving PORTAL

What is PORTAL?

Portland Oregon Regional Transportation Archive Listing (PORTAL) is an implementation of the U.S. National ITS (Intelligent Transportation Systems) Architecture's Archived Data User Service for the Portland metropolitan region.[1]

[1] http://portal.its.pdx.edu/

What is in the PORTAL Database?

Since in 2004, nearly 1 terabyte of data and over 7 million records of...

- Loop Detector Data (bulk of data size is here)
- Incident Data
- Bus Data
- Weather Data
- VMS Data

What can we do with this data?

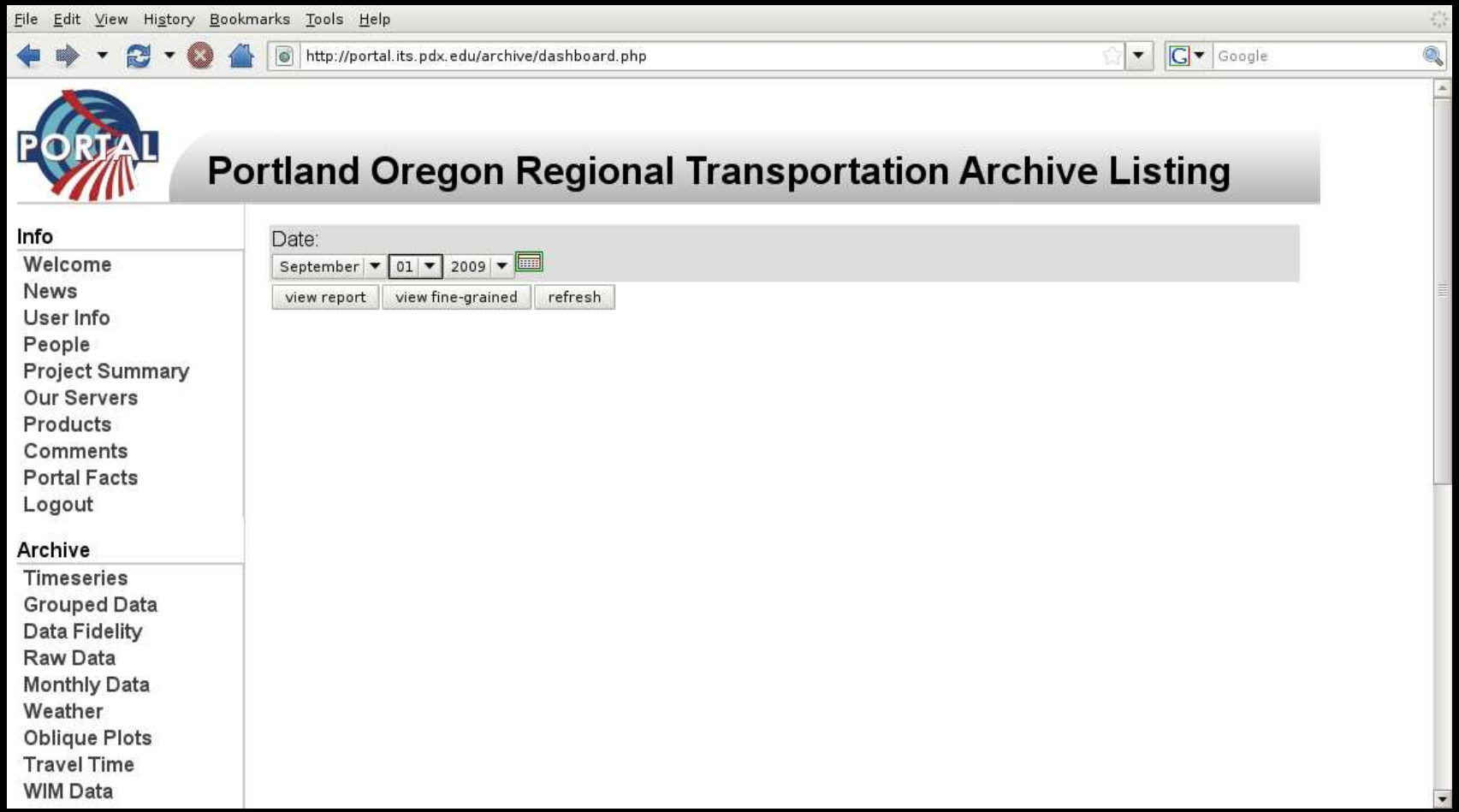
From the loop detector data:

- Timeseries plots for occupancy (length of time a vehicle is positioned over a detector)
- Traffic volume
- Traffic speed
- Vehicle Miles Traveled (VMT)
- Vehicle Hours Traveled (VHT)
- Travel time
- delay over a highway or station

PORTAL Web Site

- `http://portal.its.pdx.edu/`
- Graphical display of archived data
- Performance reports, traffic counts, freight data, ...
- Raw data

Daily Dashboard



Daily Dashboard

File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/dashboard.php?month=09&day=21&year=2009&action=view+report

Google

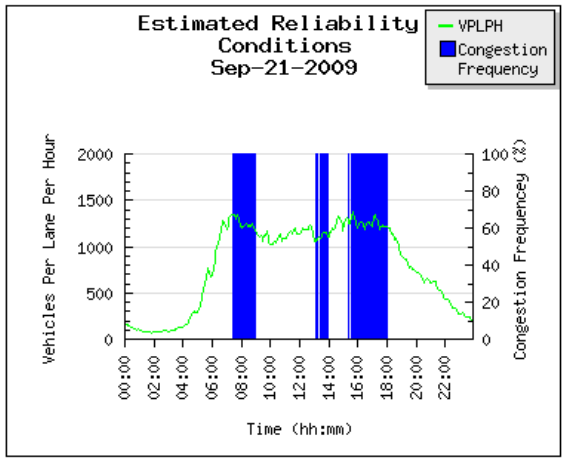
People
Project Summary
Our Servers
Products
Comments
Portal Facts
Logout

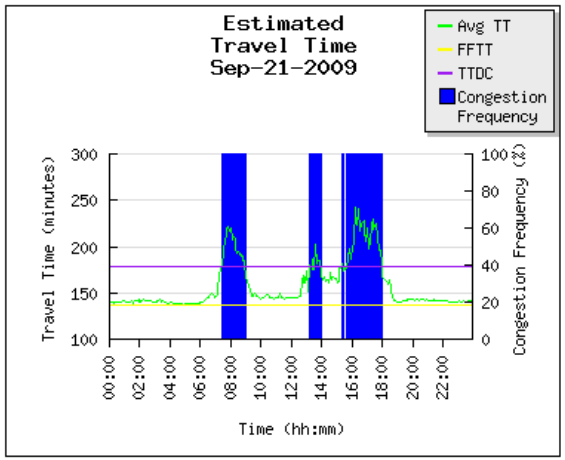
Archive

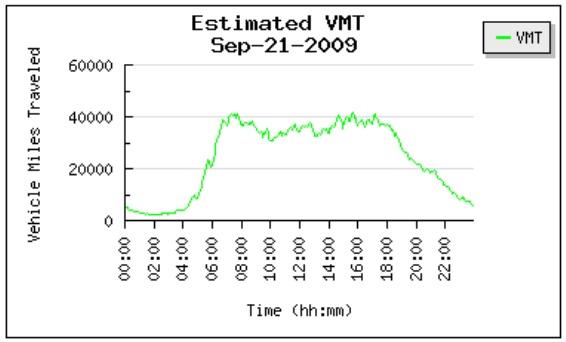
Timeseries
Grouped Data
Data Fidelity
Raw Data
Monthly Data
Weather
Oblique Plots
Travel Time
WIM Data
Performance
Dashboard
Congestion
Google Maps
SVG Maps
Bivariate Plots
Google Traffic
Vehicle Classification
Trimet Analysis
Incident Reports
Quality Control

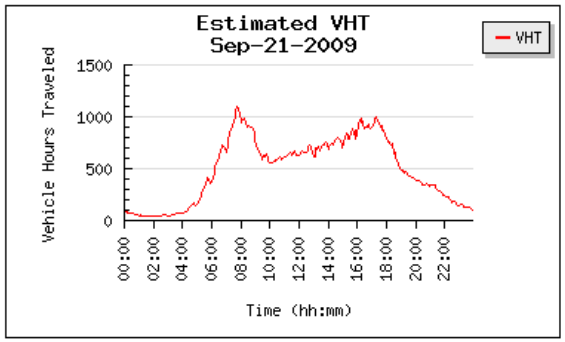
Daily Dashboard For Sep-21-2009

TOTAL VMT	TOTAL VHT	CONGESTION %	TTI
7,003,117	141,334	18.75 %	1.49
AVG TT	**AVG TS**	**95th TT**	**95th TS**
157 min	52 MPH	219 min	38 MPH

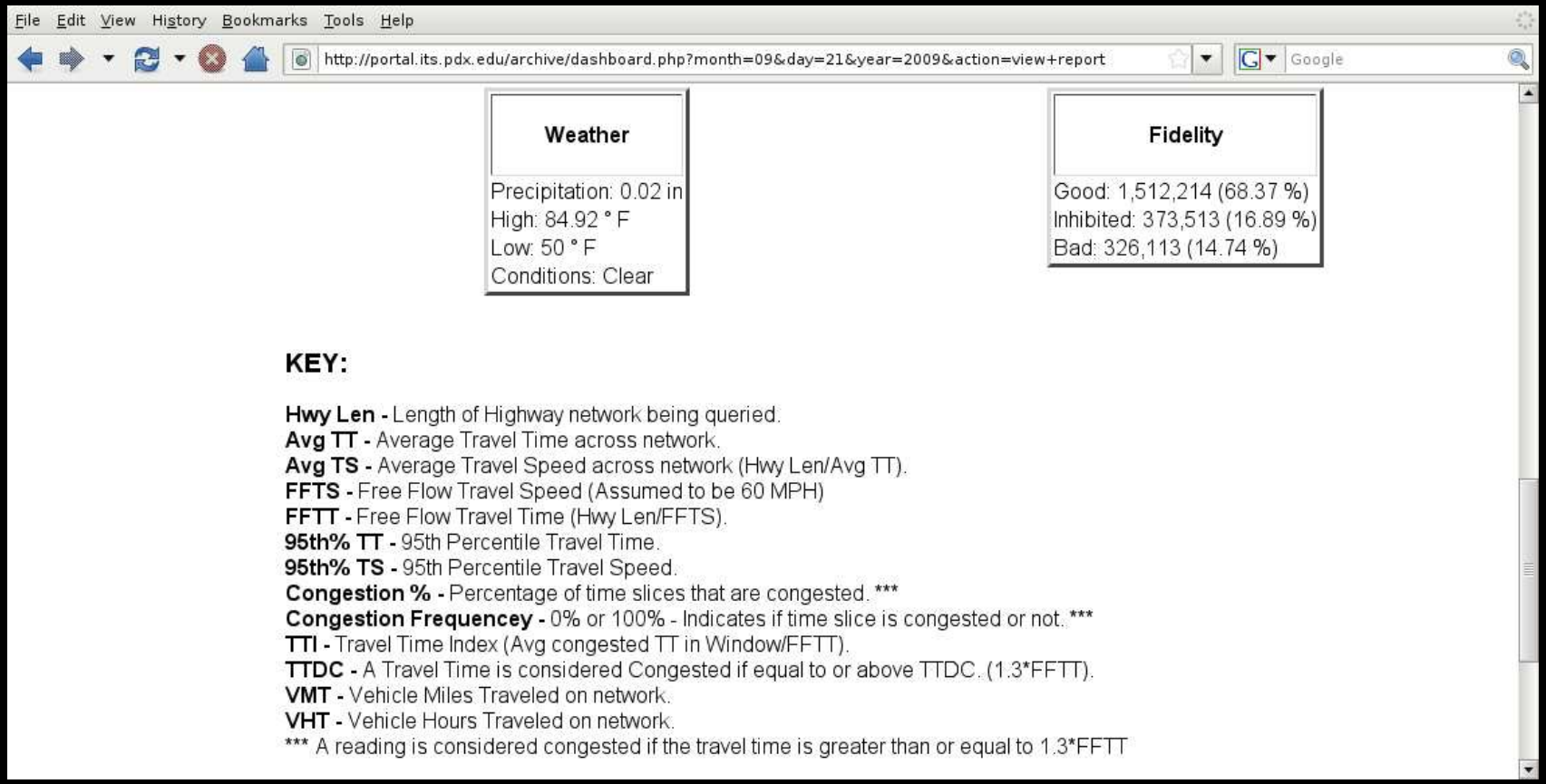




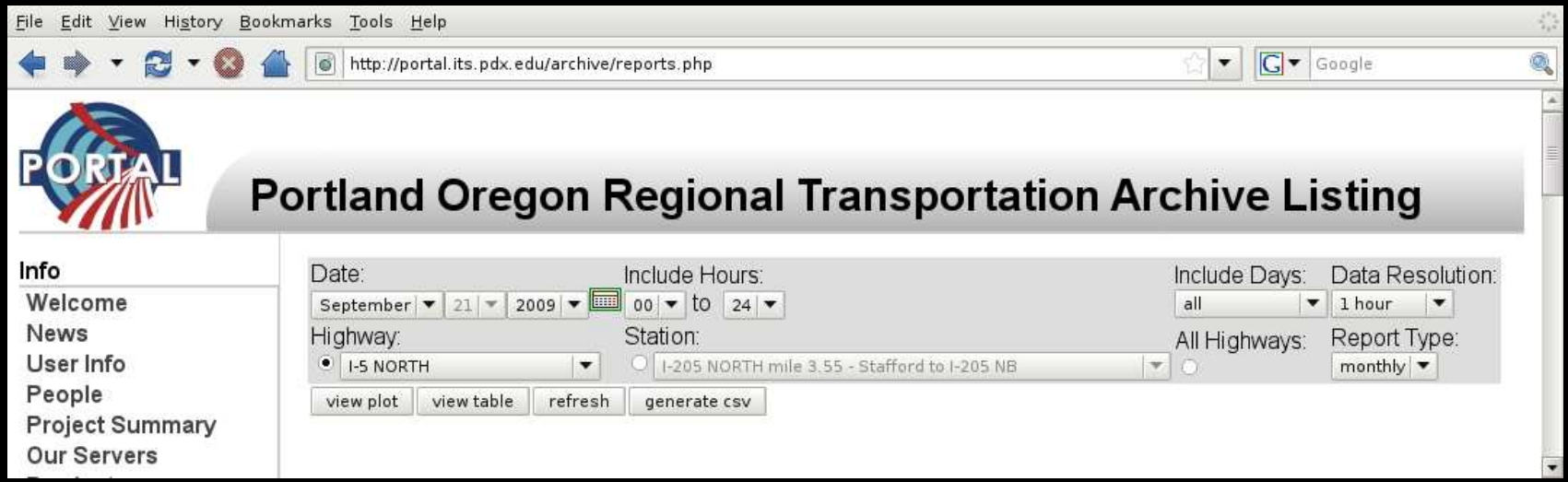




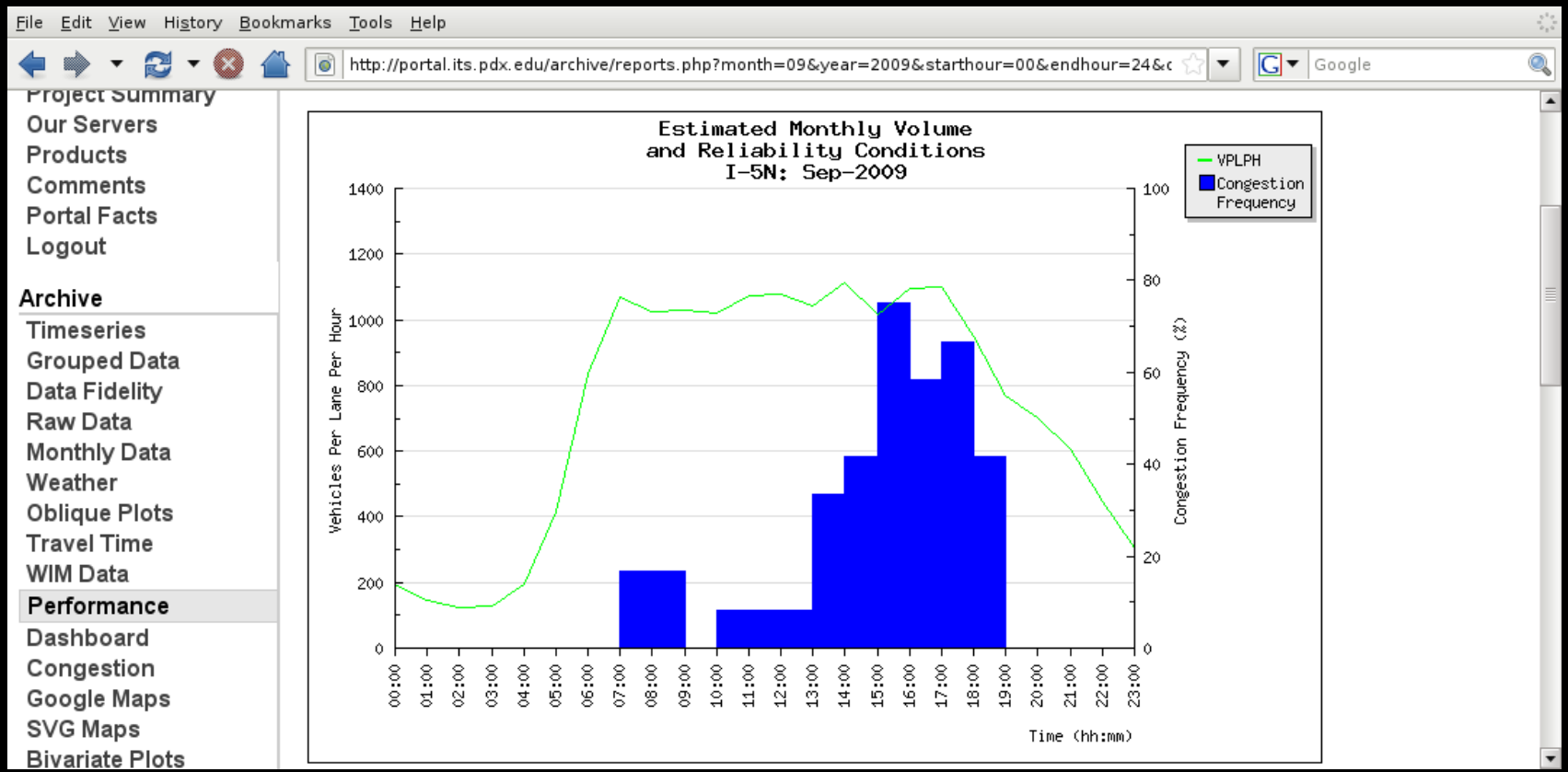
Daily Dashboard



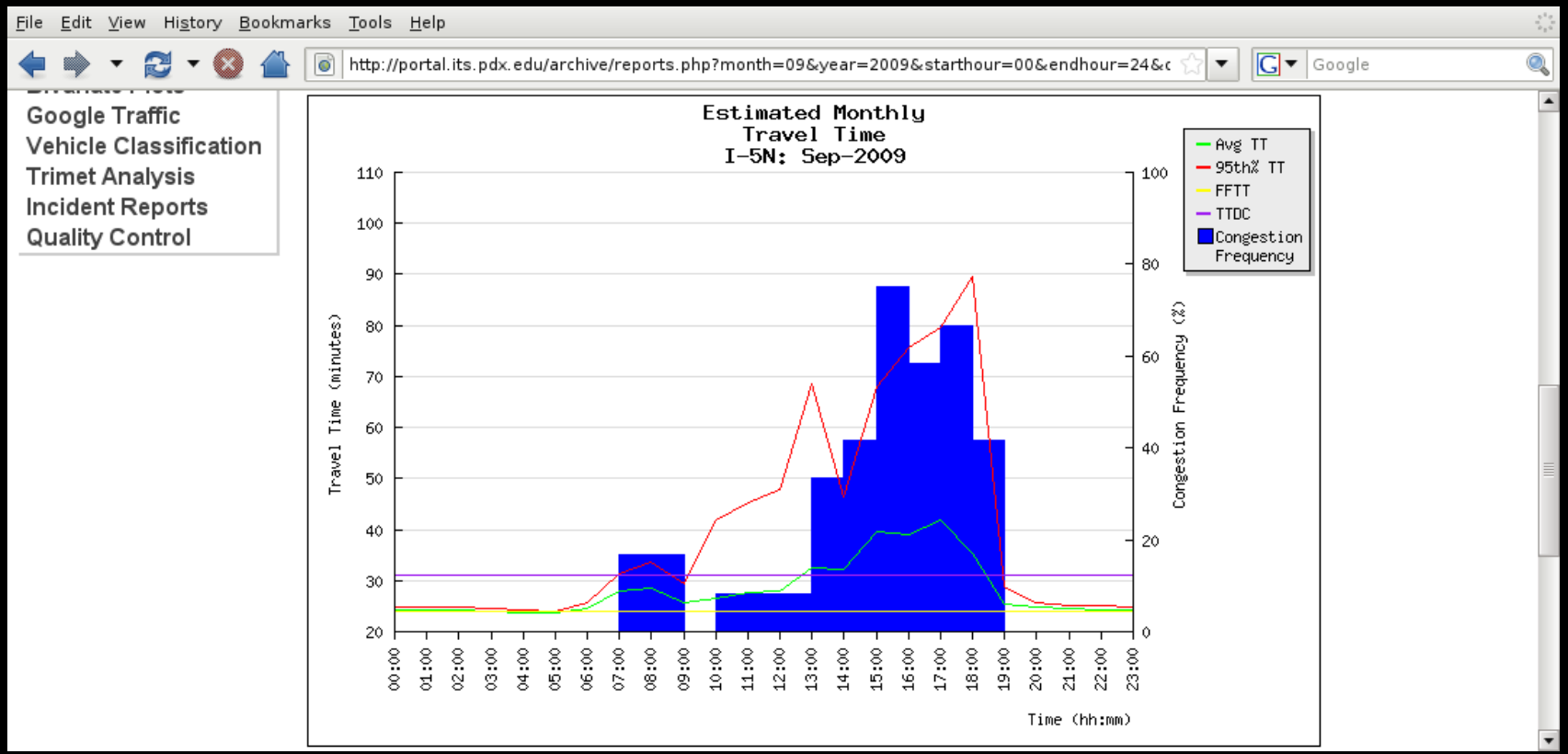
Performance Report - Reliability



Performance Report - Reliability



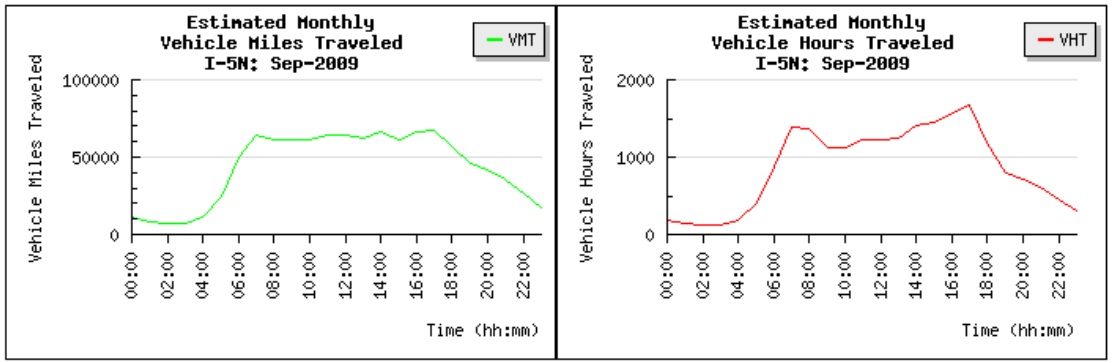
Performance Report - Reliability



Performance Report - Reliability

File Edit View History Bookmarks Tools Help

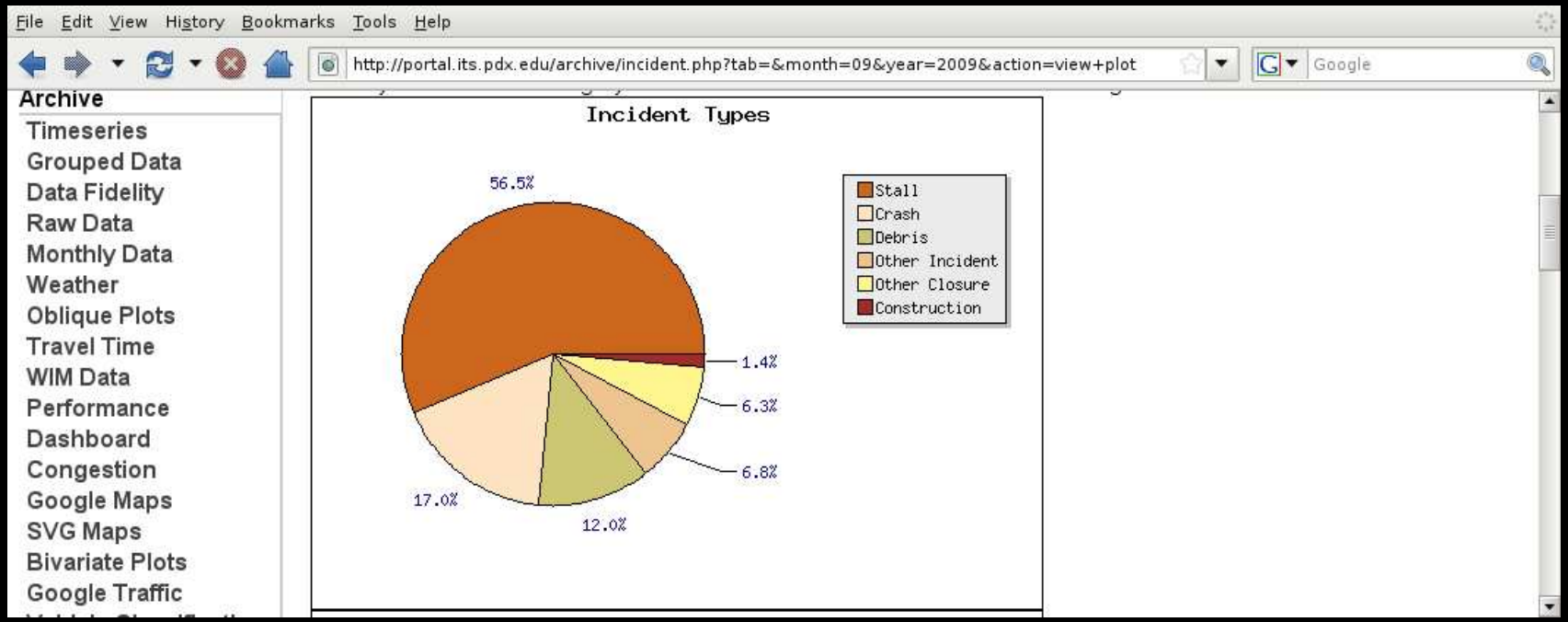
http://portal.its.pdx.edu/archive/reports.php?month=09&year=2009&starthour=00&endhour=24&c



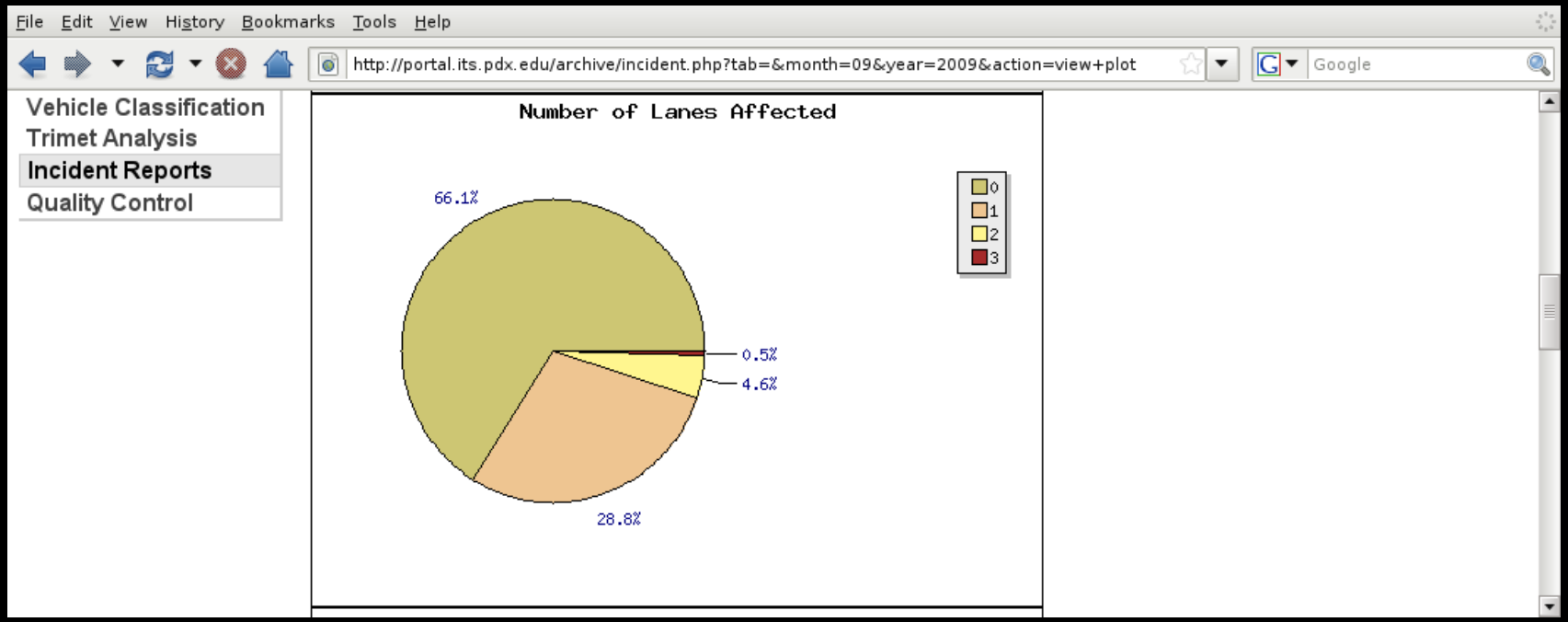
Avg TT - Average Travel Time.
95% TT - 95th Percentile Travel Time.
FFTT - Free Flow Travel Time (15th percentile Travel Time).**
FFTS - Free Flow Travel Speed (Hwy Len/FFTT).
TTDC - A Travel Time is considered Congested if equal to or above TTDC. (1.3*FFTT).
** If FFTS is less than 52 MPH, then FFTS is 52 MPH and FFTT is Hwy_Len/52 MPH.

Speed Plot with Incident Reports

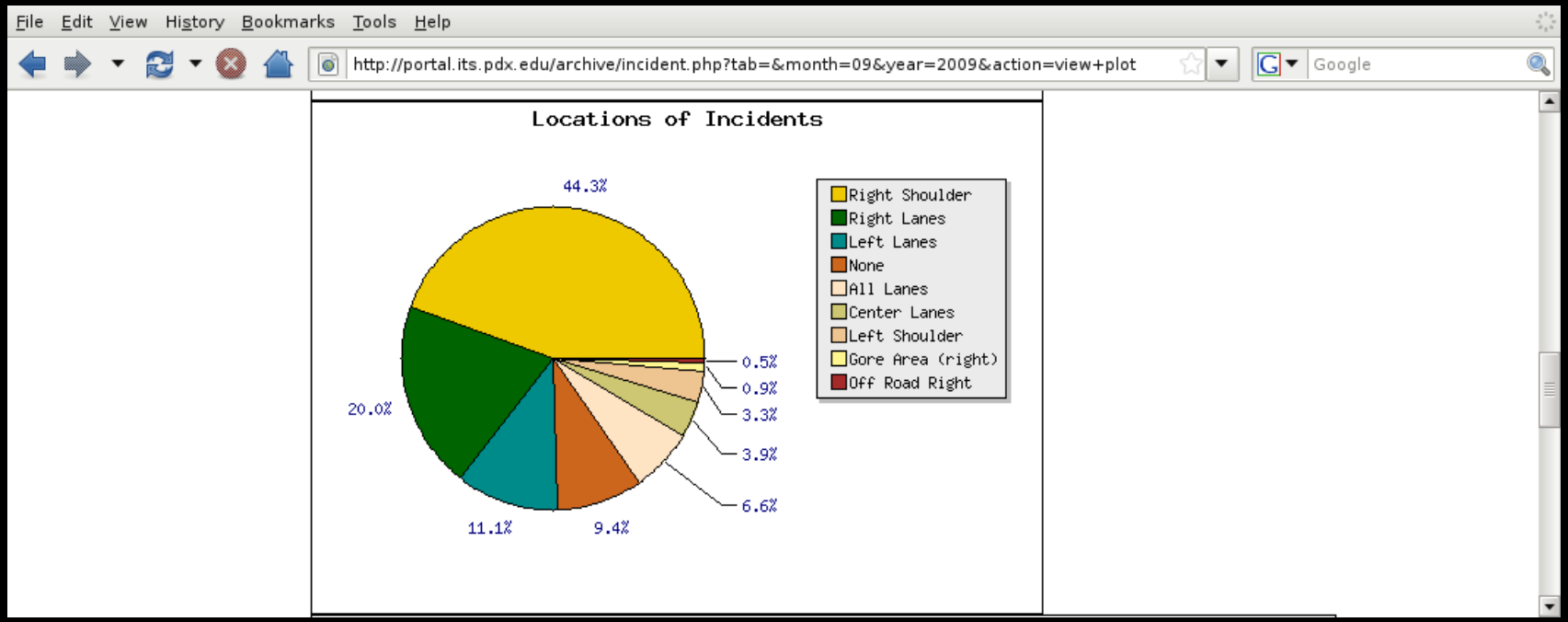
Speed Plot with Incident Reports



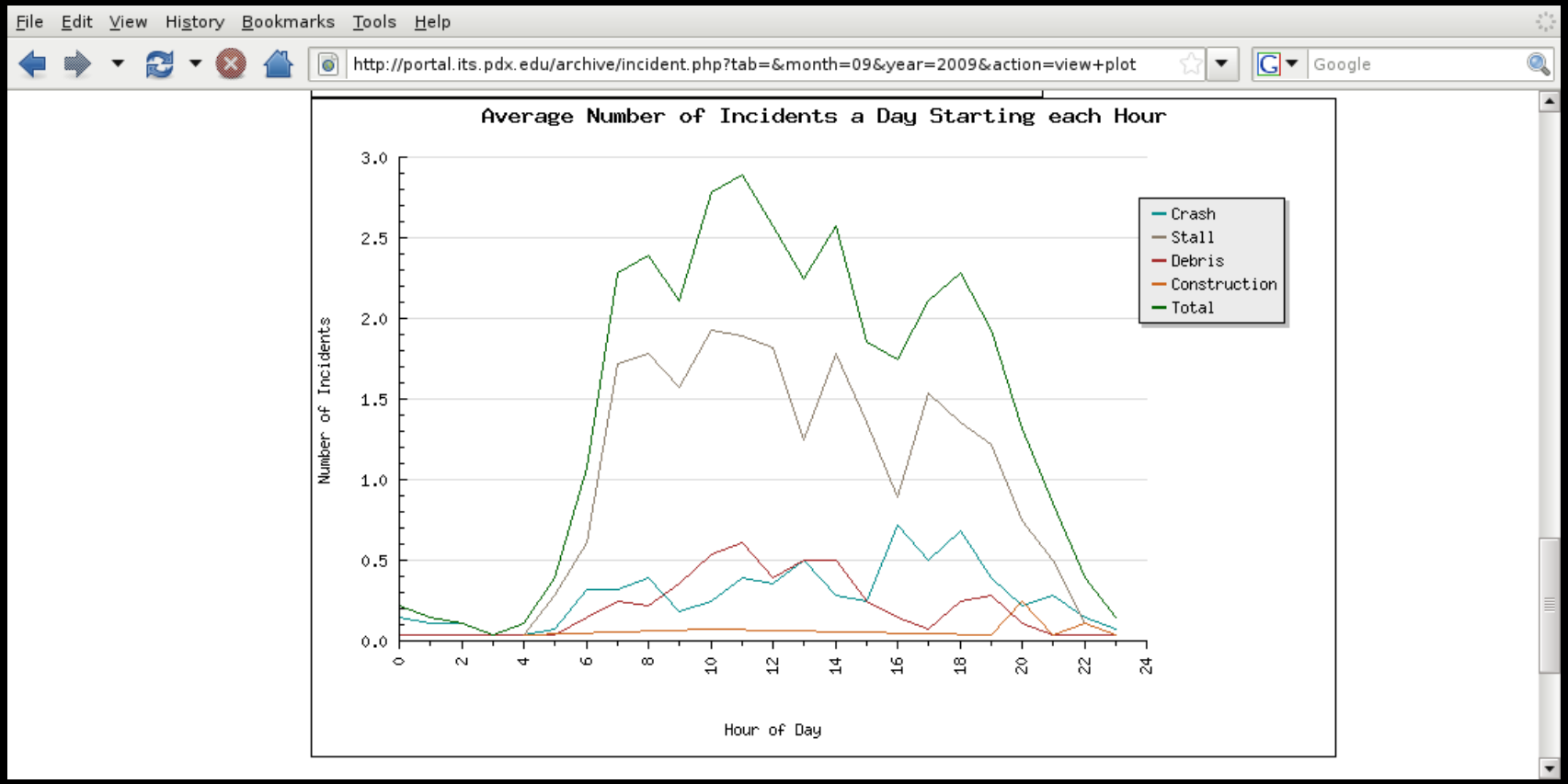
Speed Plot with Incident Reports



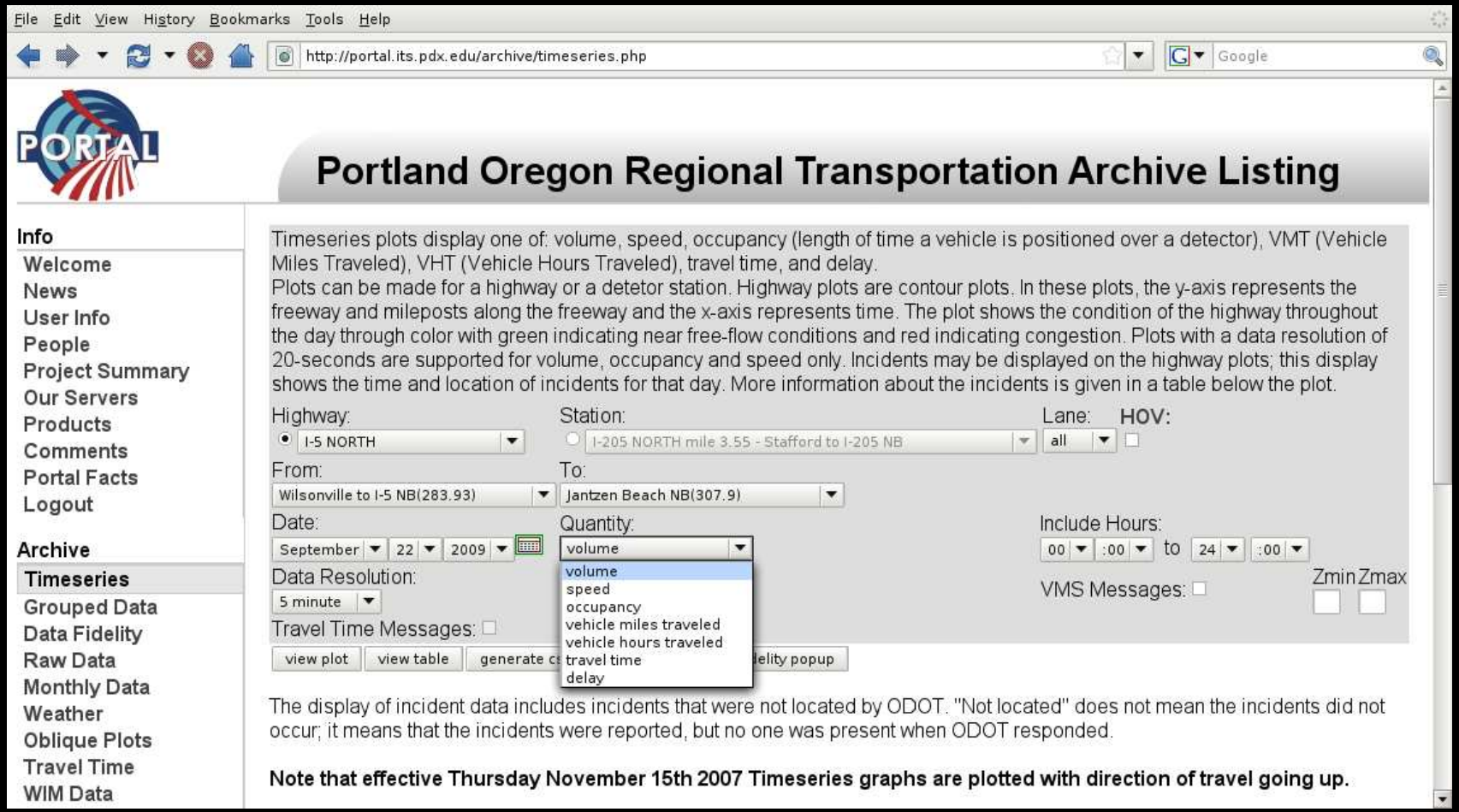
Speed Plot with Incident Reports



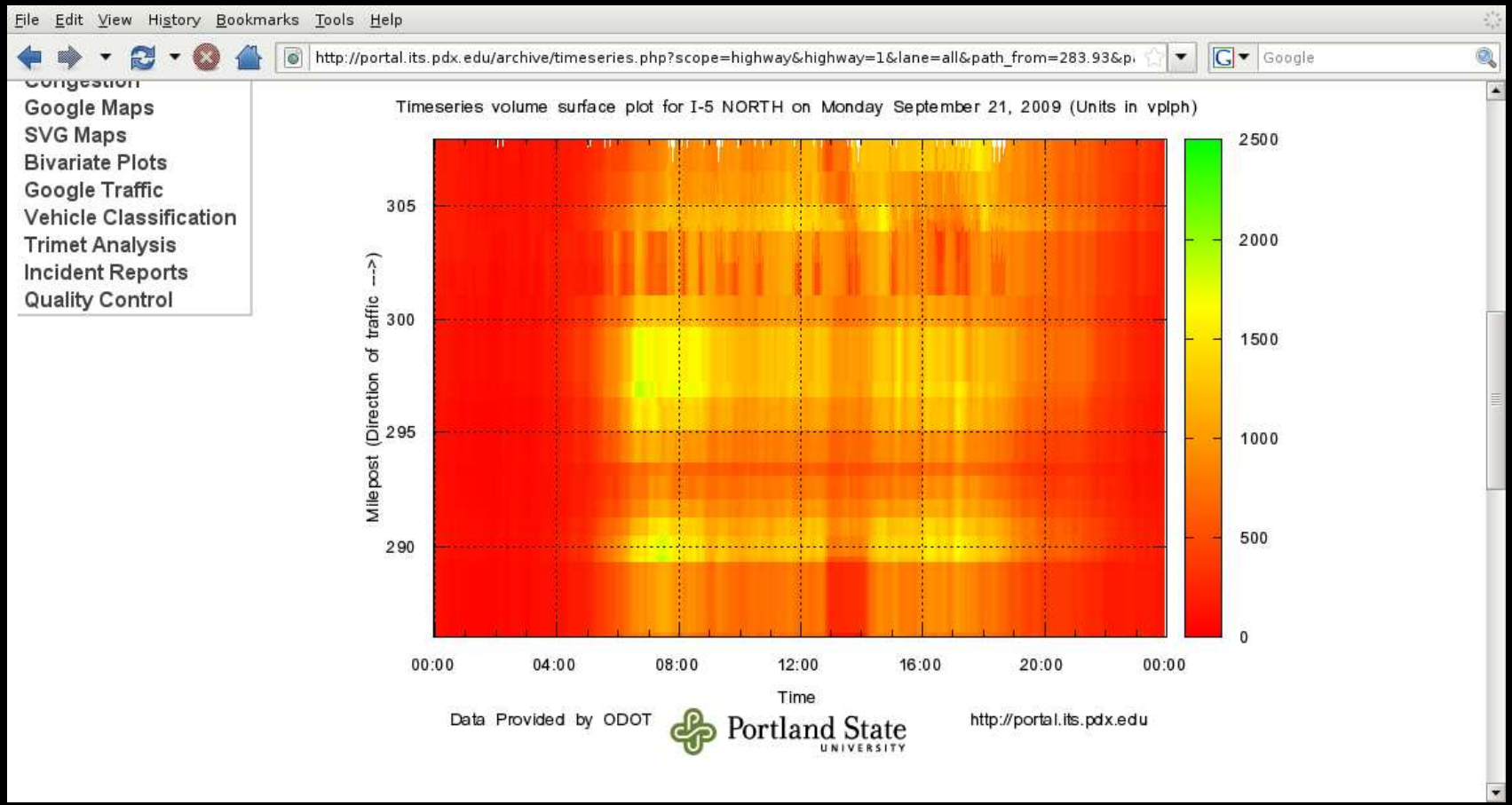
Speed Plot with Incident Reports



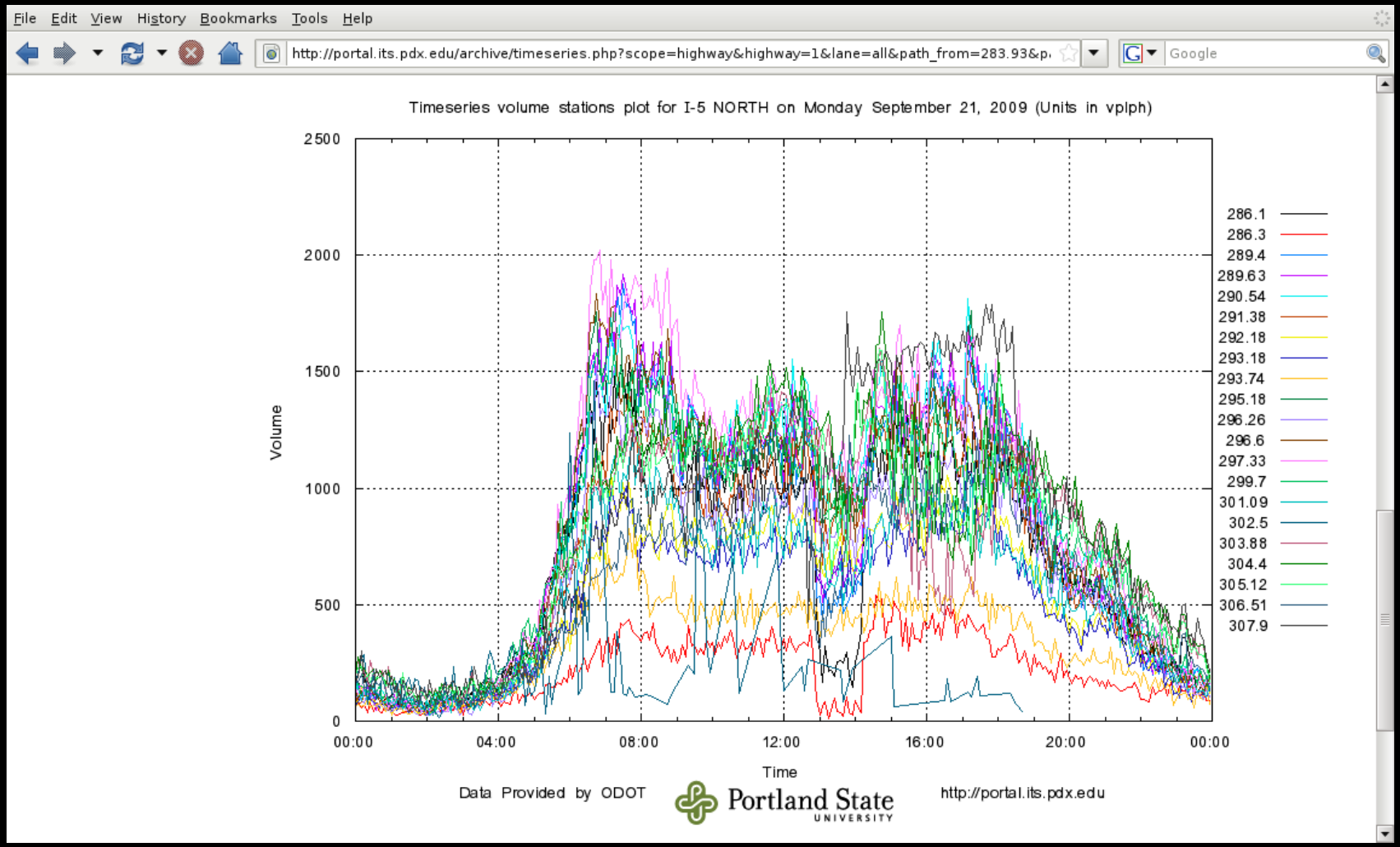
Time Series


File Edit View History Bookmarks Tools Help
http://portal.its.pdx.edu/archive/timeseries.php
Google
PORTAL
Portland Oregon Regional Transportation Archive Listing
Info
Welcome
News
User Info
People
Project Summary
Our Servers
Products
Comments
Portal Facts
Logout
Archive
Timeseries
Grouped Data
Data Fidelity
Raw Data
Monthly Data
Weather
Oblique Plots
Travel Time
WIM Data
Timeseries plots display one of: volume, speed, occupancy (length of time a vehicle is positioned over a detector), VMT (Vehicle Miles Traveled), VHT (Vehicle Hours Traveled), travel time, and delay.
Plots can be made for a highway or a detetor station. Highway plots are contour plots. In these plots, the y-axis represents the freeway and mileposts along the freeway and the x-axis represents time. The plot shows the condition of the highway throughout the day through color with green indicating near free-flow conditions and red indicating congestion. Plots with a data resolution of 20-seconds are supported for volume, occupancy and speed only. Incidents may be displayed on the highway plots; this display shows the time and location of incidents for that day. More information about the incidents is given in a table below the plot.
Highway:
I-5 NORTH
Station:
I-205 NORTH mile 3.55 - Stafford to I-205 NB
Lane:
all
HOV:
From:
Wilsonville to I-5 NB(283.93)
To:
Jantzen Beach NB(307.9)
Date:
September 22 2009
Quantity:
volume
volume
speed
occupancy
vehicle miles traveled
vehicle hours traveled
travel time
delay
Include Hours:
00 :00 to 24 :00
Data Resolution:
5 minute
VMS Messages:
Zmin Zmax
Travel Time Messages:
view plot
view table
generate c
delity popup
The display of incident data includes incidents that were not located by ODOT. "Not located" does not mean the incidents did not occur; it means that the incidents were reported, but no one was present when ODOT responded.
Note that effective Thursday November 15th 2007 Timeseries graphs are plotted with direction of travel going up.

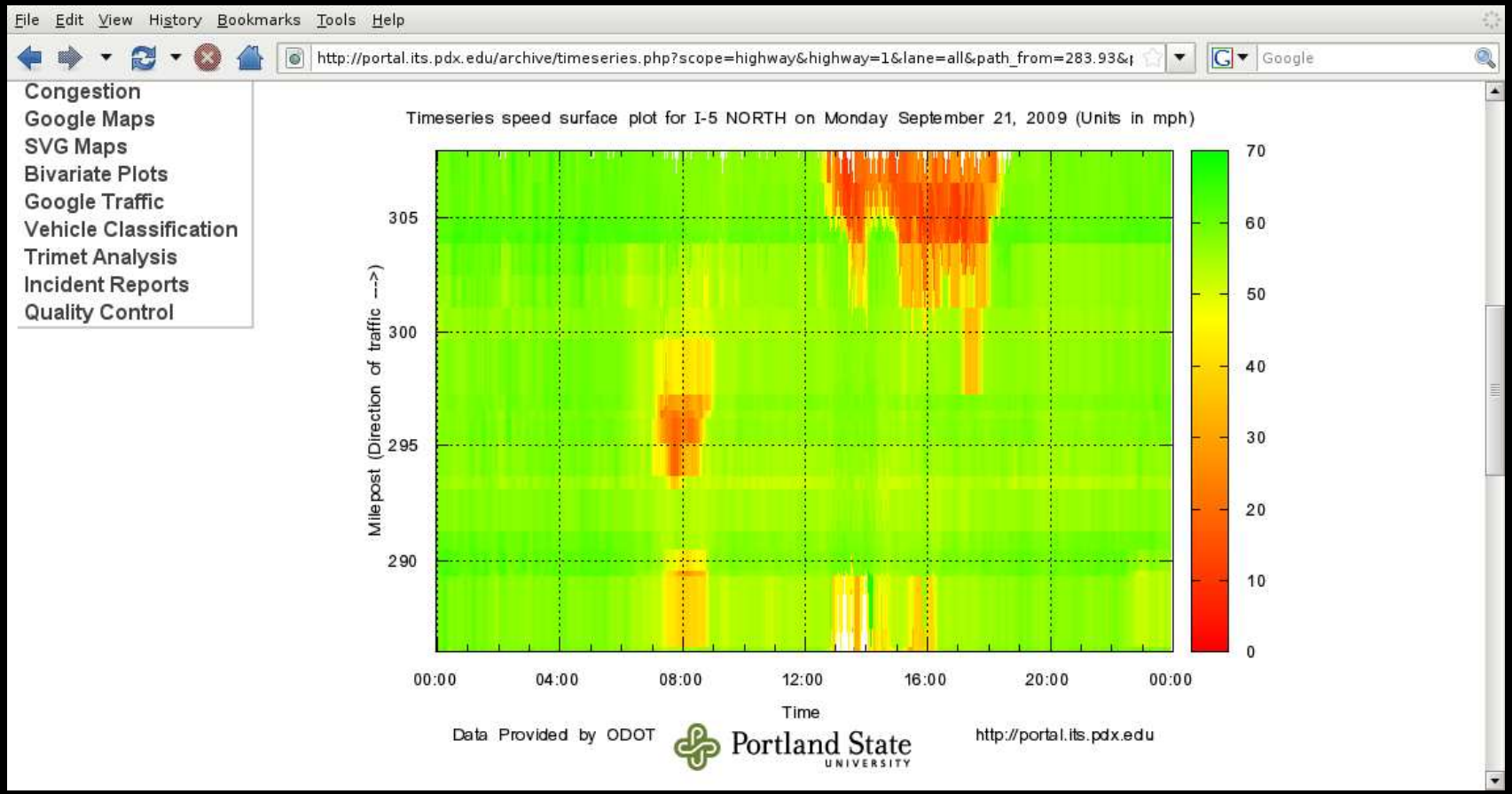
Surface Plot - Volume



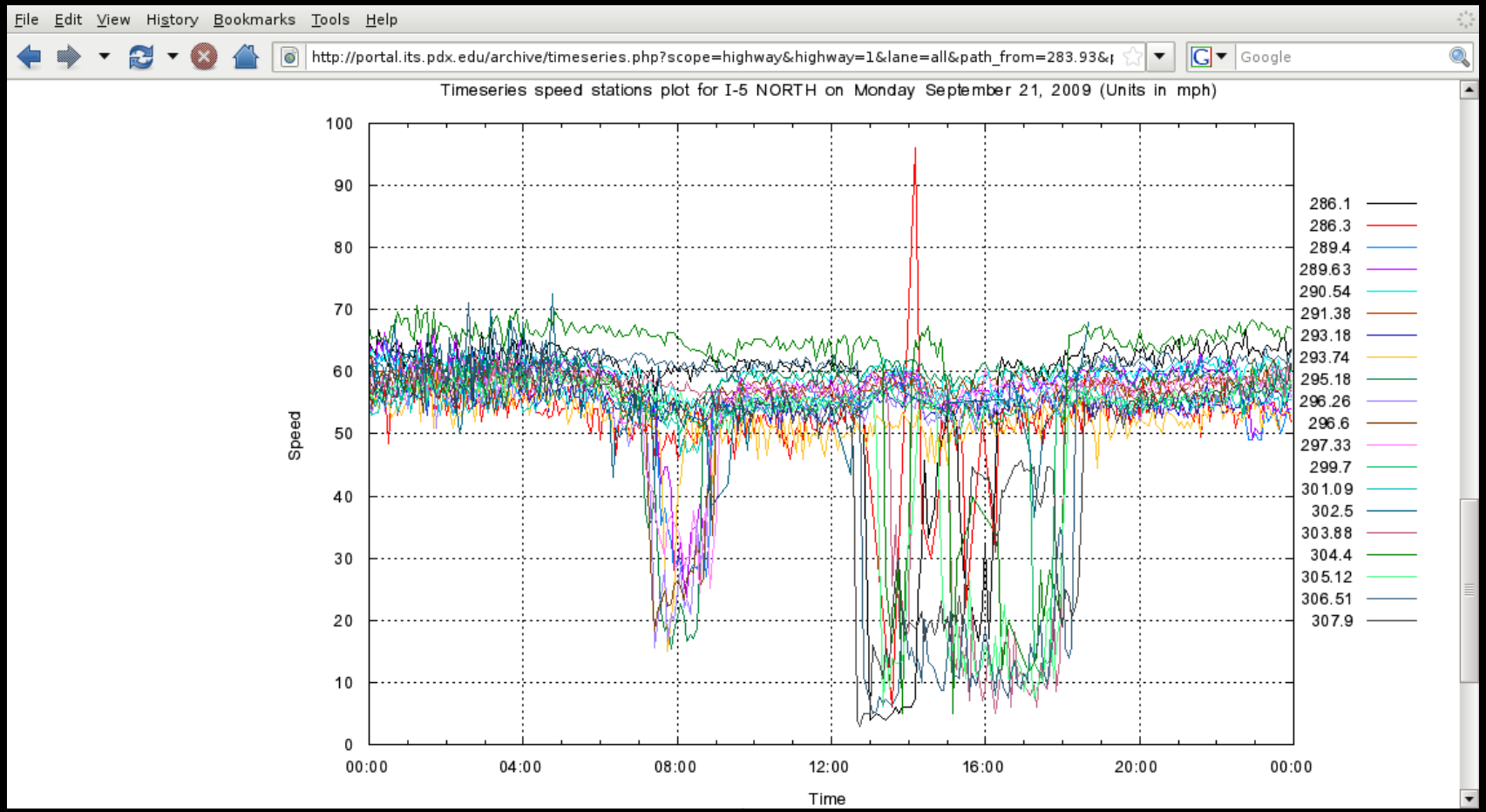
Time Series - Volume



Surface Plot - Speed

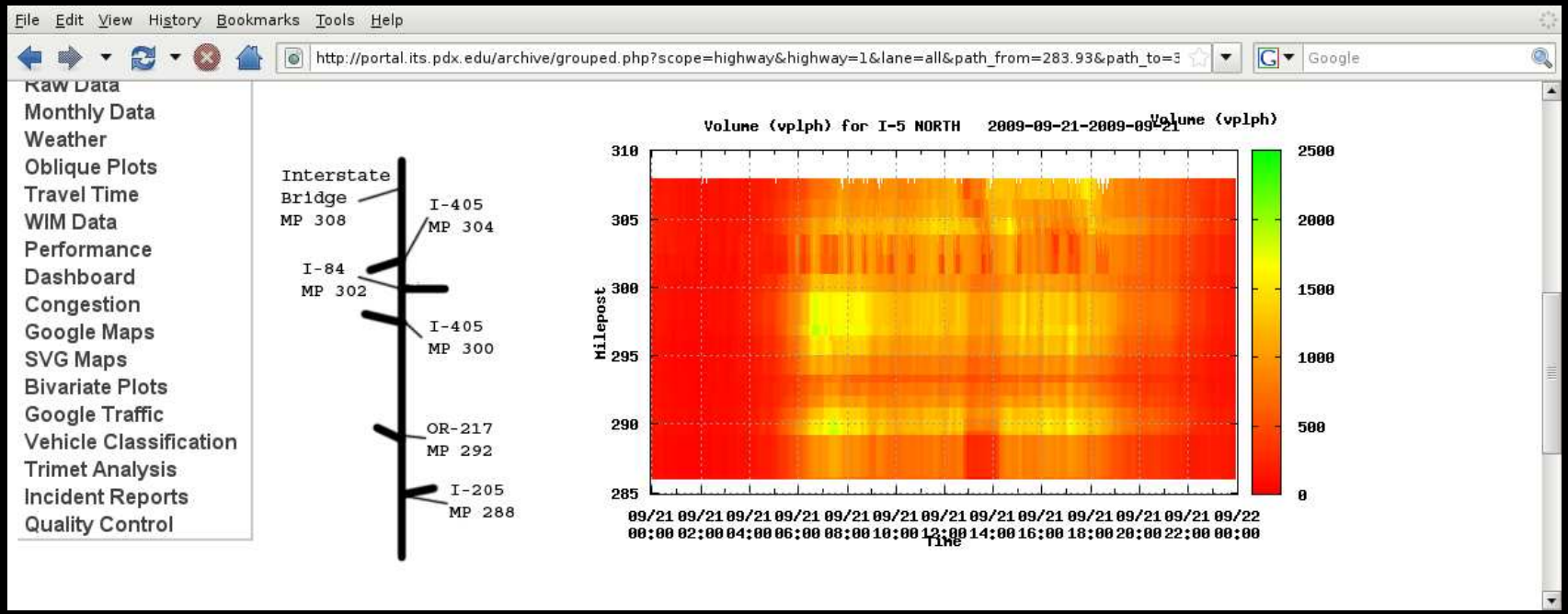


Time Series - Speed

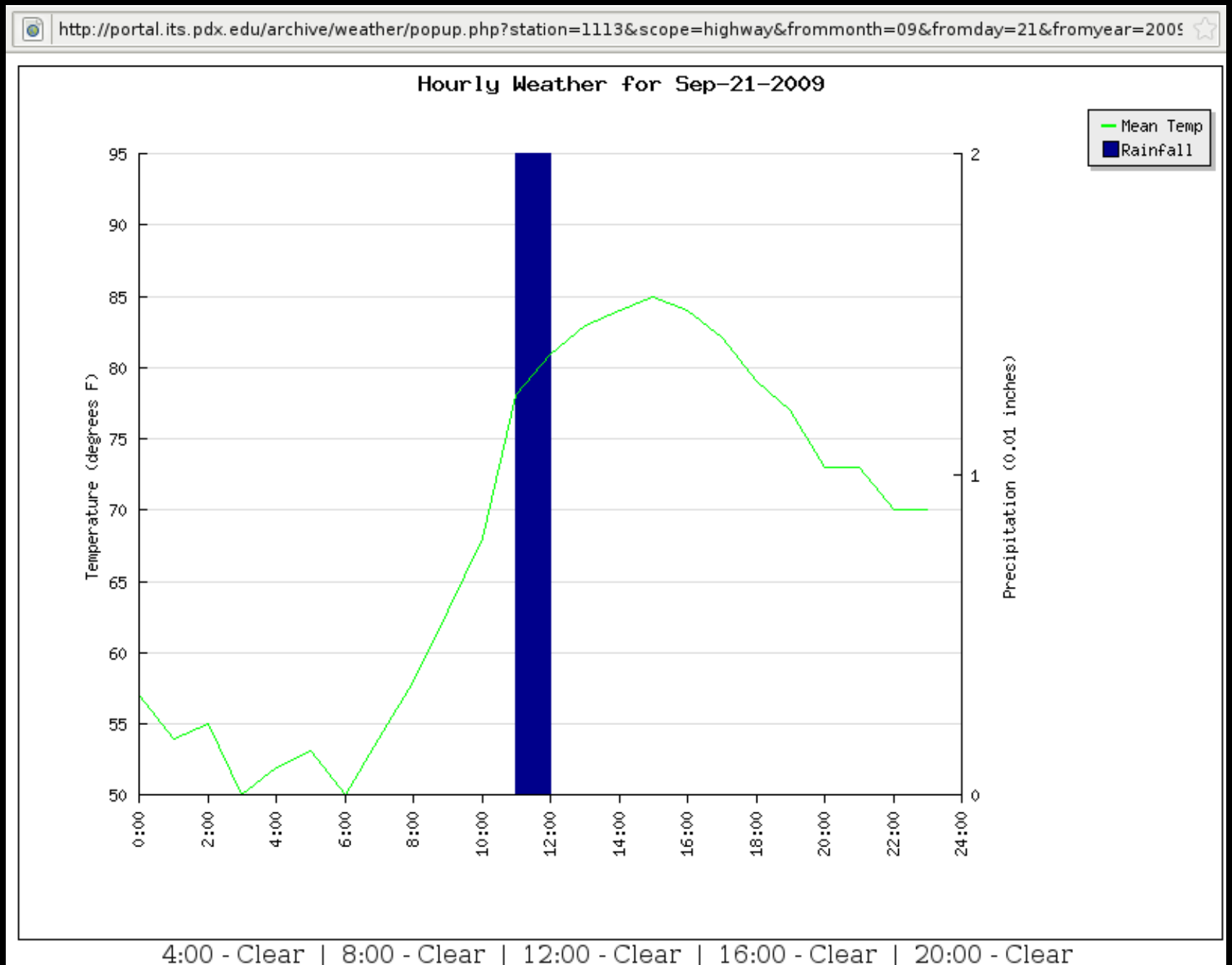


Grouped Data - Travel Time

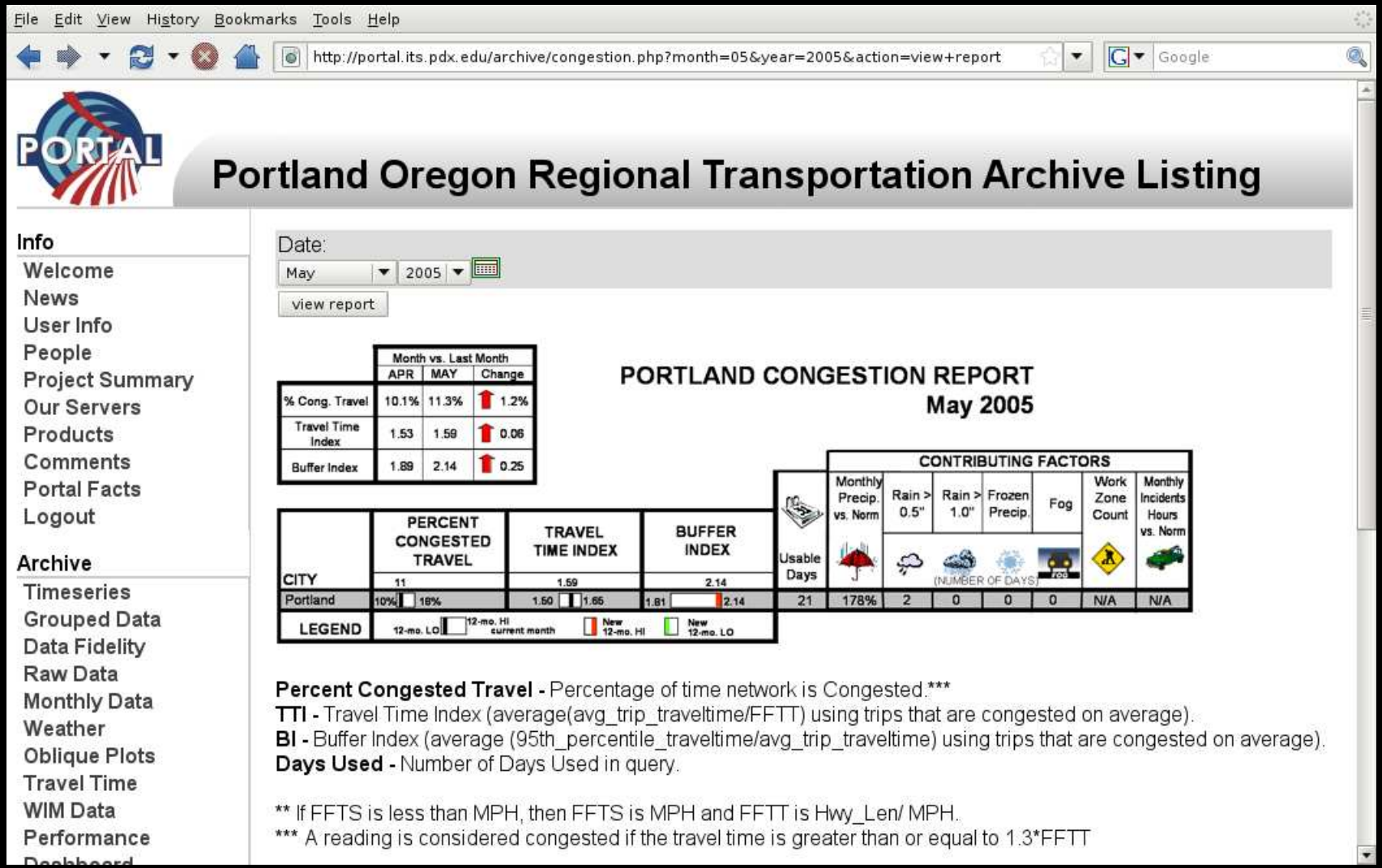
Grouped Data - Travel Time



Weather Popup



Monthly Report



File Edit View History Bookmarks Tools Help

http://portal.its.pdx.edu/archive/congestion.php?month=05&year=2005&action=view+report

Google

PORTAL

Portland Oregon Regional Transportation Archive Listing

Info

- Welcome
- News
- User Info
- People
- Project Summary
- Our Servers
- Products
- Comments
- Portal Facts
- Logout

Archive

- Timeseries
- Grouped Data
- Data Fidelity
- Raw Data
- Monthly Data
- Weather
- Oblique Plots
- Travel Time
- WIM Data
- Performance
- Dashboard

Date:

May 2005

view report

	Month vs. Last Month		
	APR	MAY	Change
% Cong. Travel	10.1%	11.3%	1.2%
Travel Time Index	1.53	1.59	0.06
Buffer Index	1.89	2.14	0.25

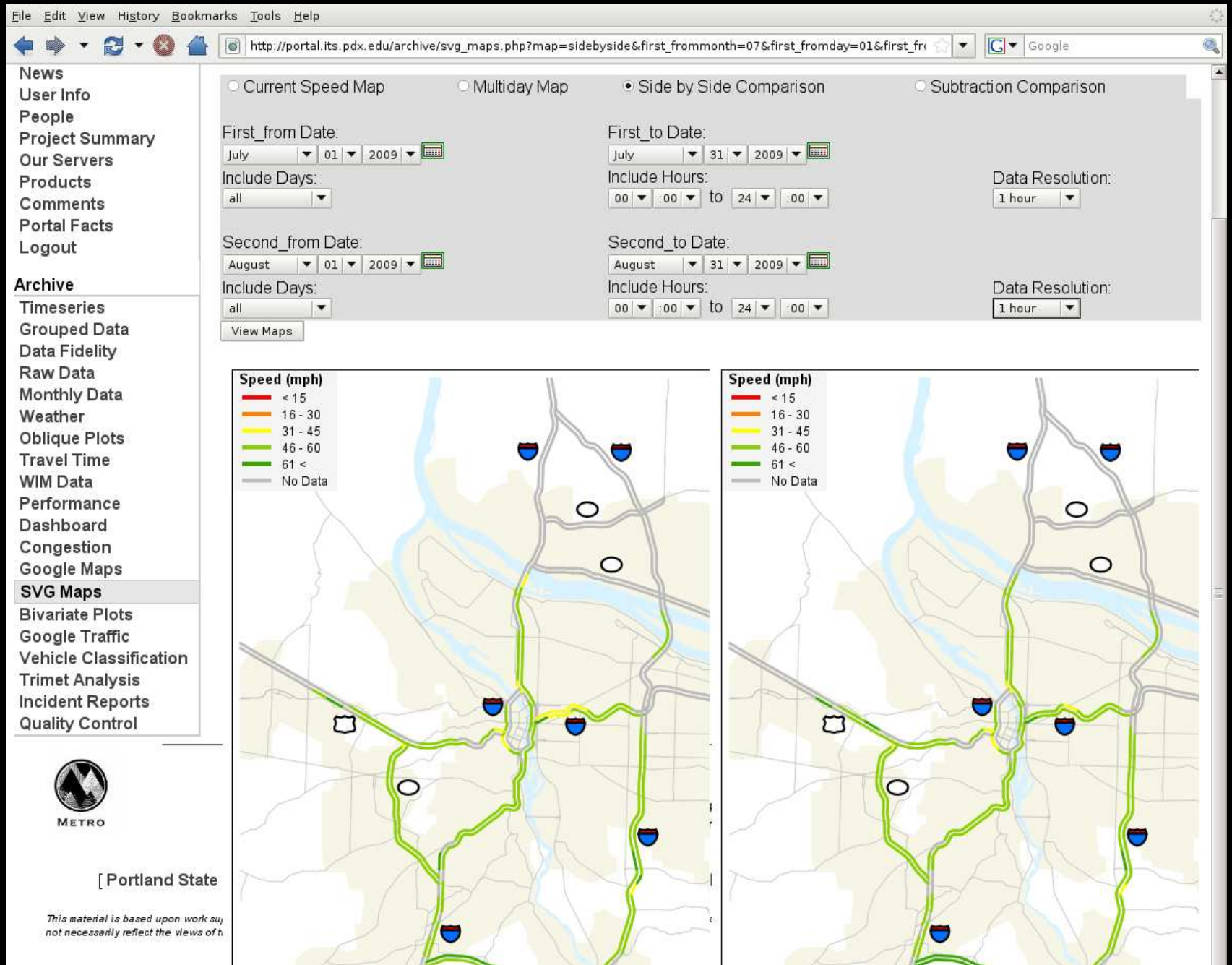
PORTLAND CONGESTION REPORT
May 2005

CITY	PERCENT CONGESTED TRAVEL	TRAVEL TIME INDEX	BUFFER INDEX	Usable Days	CONTRIBUTING FACTORS: Monthly Precip. vs. Norm	Rain > 0.5"	Rain > 1.0"	Frozen Precip.	Fog	Work Zone Count	Monthly Incidents Hours vs. Norm
	11	1.59	2.14			(NUMBER OF DAYS)					
Portland	10% 18%	1.50 1.65	1.81 2.14	21	178%	2	0	0	0	N/A	N/A
LEGEND	12-mo. LO 12-mo. HI current month		New 12-mo. HI / New 12-mo. LO								

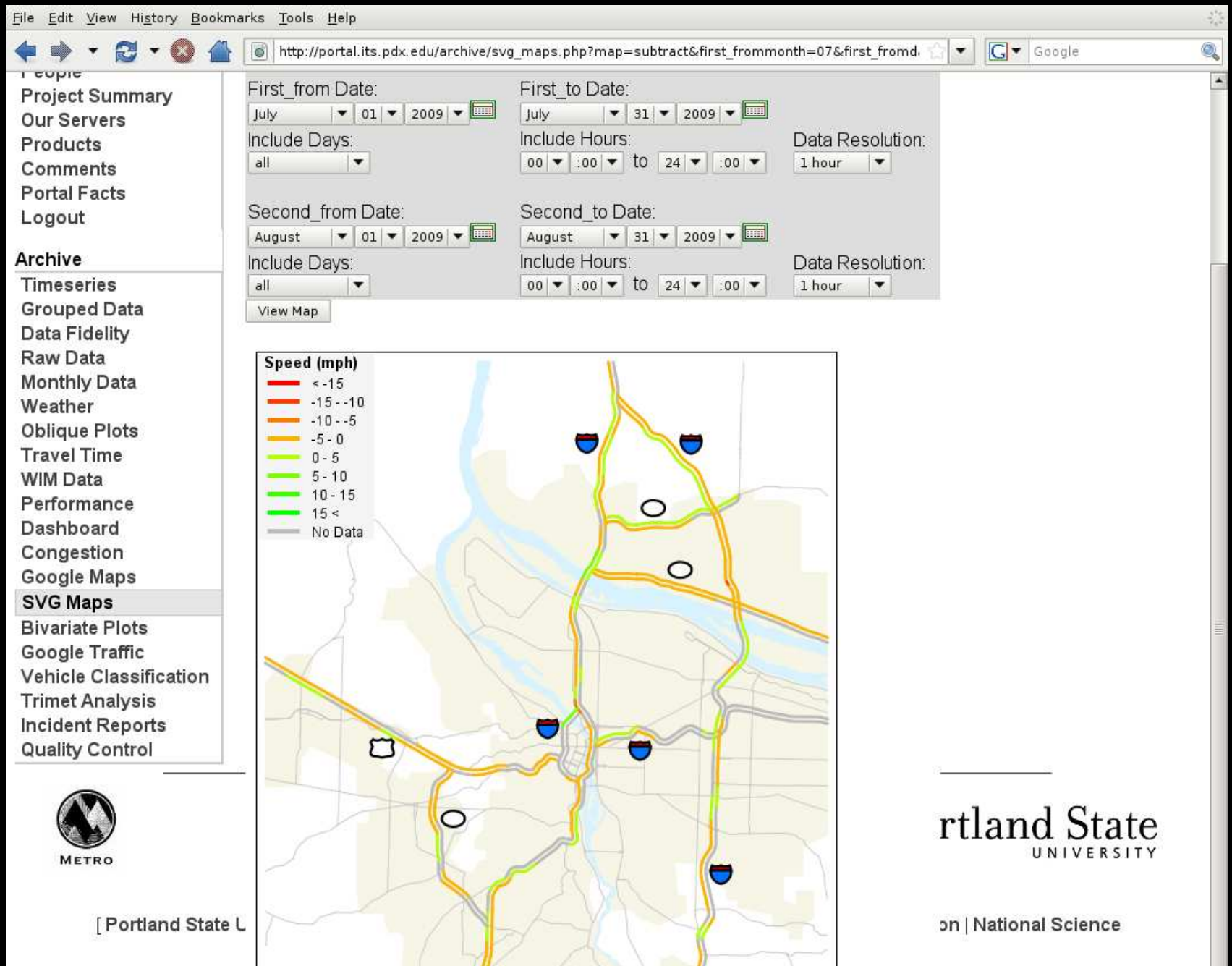
Percent Congested Travel - Percentage of time network is Congested.***
TTI - Travel Time Index (average(avg_trip_traveltime/FFTT) using trips that are congested on average).
BI - Buffer Index (average (95th_percentile_traveltime/avg_trip_traveltime) using trips that are congested on average).
Days Used - Number of Days Used in query.

** If FFTS is less than MPH, then FFTS is MPH and FFTT is Hwy_Len/ MPH.
*** A reading is considered congested if the travel time is greater than or equal to 1.3*FFTT

Mapping - Speed by Month



Mapping - Speed Subtraction



Other Uses of PORTAL

- Resource for local transportation professionals
- Metro RTP
- Projects
  - Travel Time
  - Bottleneck Identification
  - Data Quality Evaluation
  - Gap Filling
  - TriMet Data Analysis
  - Oregon Freight Data Mart
  - Incident Autopsy

How PORTAL Works

For the loop detector data:

- Data from loop detectors on the road are aggregated into 20 second intervals at the device
- The aggregated data are immediately transmitted to ODOT (Oregon Department of Transportation), then transmitted to PSU data
- The XML data is transformed into SQL statements to load into the database
- Scripts aggregating data into 5 min, 15 min and 1 hour intervals are run at regular intervals

Current PORTAL Configuration

- PostgreSQL v8.1
- Red Hat Linux
- 2 quad-core Core 2 processors
- 32 GB RAM
- 1 TB of storage on a SAN

Things to try

- Take advantage of PostgreSQL's Portland Performance Pad
- Take advantage of the way data is loaded to horizontally partition tables
- Reduce database size
  - Data is duplicated for performance for 5 min, 15 min, and 1 hour aggregates
  - Determine if additional indexes are necessary
- Experiment with newer versions of PostgreSQL (and Linux)

PostgreSQL Test System

- PostgreSQL v8.4
- Gentoo Linux
- 2 quad-core Core 2 processors (not exactly the same as the production system)
- 32 GB RAM
- 25-disk 72GB SAS array

Database Sizes

- One day's worth of loop data is approximately 165 MB
- The primary key index is approximately an additional 160 MB
- The additional index is approximately another 80 MB
- A month's worth of data and indexes approximately 12 GB of data
- A year's worth of data and indexes approximately 145 GB of data

Aggregation Size Overhead

- Total size for a month
  - table and indexes = 11,821 MB
  - table and indexes for 5 min aggregates = 493 MB, 4%
- Total size for the year
  - table and indexes = 151,935 MB
  - tables and indexes for 15 min aggregates = 1,913 MB, 1%
  - tables and indexes for 1 hour aggregates = 478 MB, 0.3%

Aggregation Performance Overhead

- 5 min aggregate table is updated every 5 minutes
- 15 min aggregate table is updated every 15 minutes
- 1 hour aggregate table is updated every hour
- 5 min, 15 min, and 1 hour aggregate tables take about 15 minutes of time to run per 1 month of data

Scaling users

Running the timeseries query:

- Users = Response Time
- 1 = 8.6s
- 2 = 9.5s
- 3 = 12.6s
- 4 = 16.6s
- 5 = 20.4s
- 6 = 24.6s
- 7 = 28.5s
- 8 = 32.7s
- 9 = 36.1s
- 10 = 39.7s

A brief look at i/o data

- Timeseries query doing approximately 0.5 MB/s reads per second
- fio testing shows 4 MB/s expected from the drive for random read
- Suggests increasing spindles per table...

Future Work

- Figure out why oprofile isn't working on the system (or what to use instead)
- Study more system characteristics when scaling up concurrent database users
- Experiment with filesystems other than ext2
- Experiment with increasing spindles

Thank you!