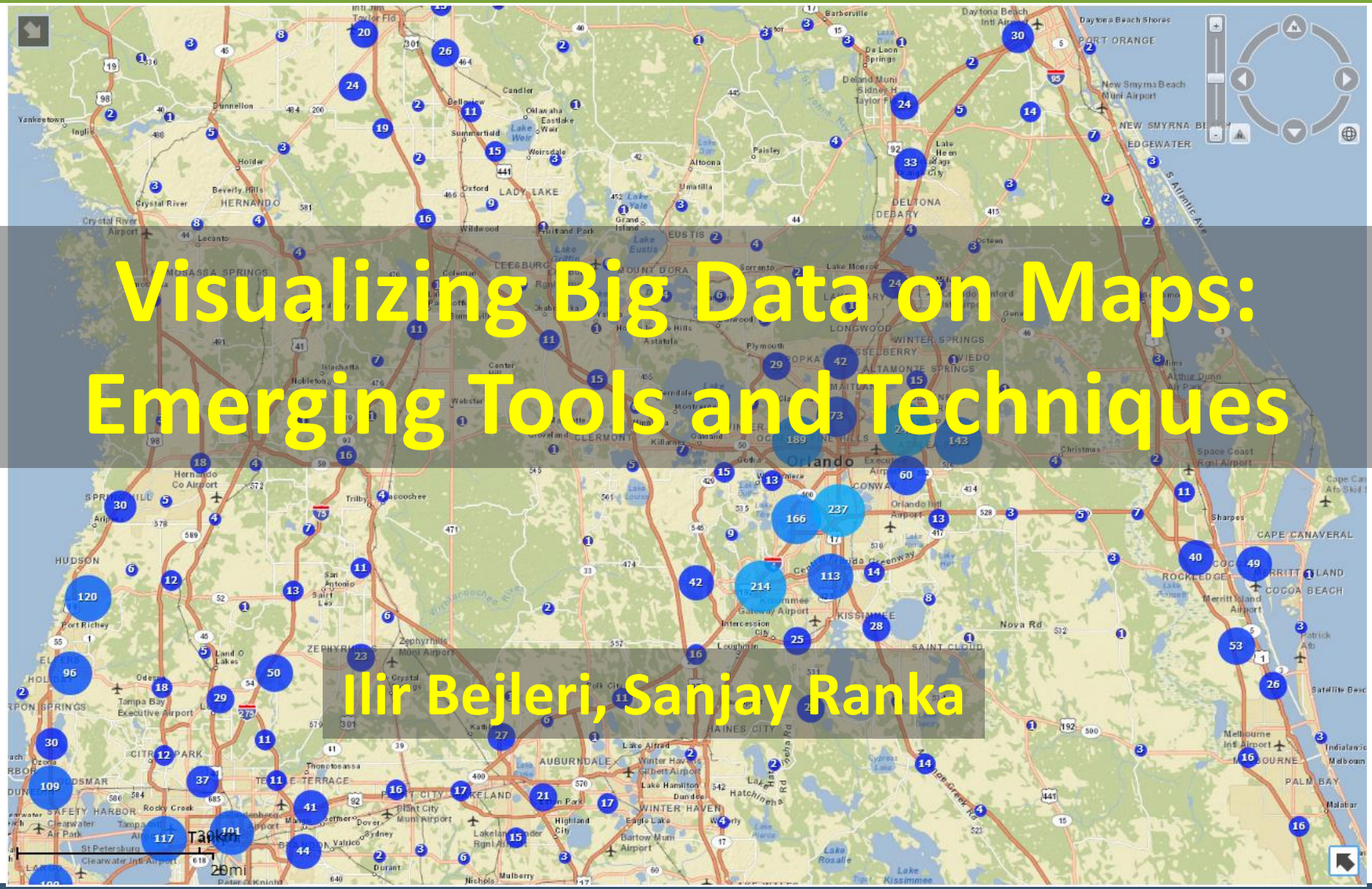

Visualizing Big Data on Maps:
Emerging Tools and Techniques
Ilir Bejleri, Sanjay Ranka

Topics

- Web GIS Visualization
- Big Data GIS Performance
- Maps in Data Visualization Platforms

Next:

- **Web GIS Visualization**
- Big Data GIS Performance
- Maps in Data Visualization Platforms

Web GIS Visualization Platforms

- Claud-based mapping platforms e.g. ArcGIS Online
- Maps created on desktop & loaded on the cloud
- Self-service capabilities and ready-to-use maps
- Include spatial analytics capabilities
- Possibilities for customization

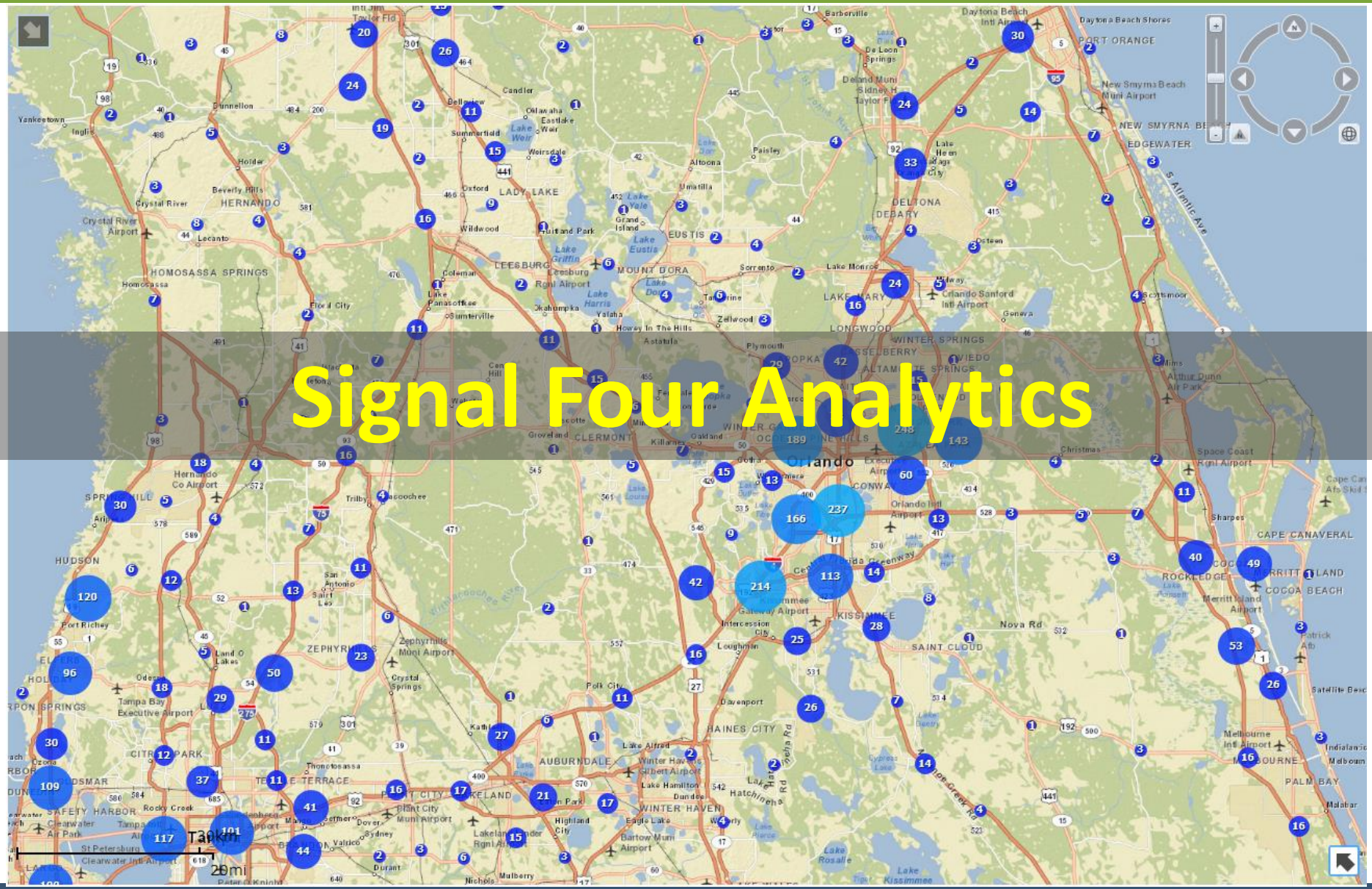
Visualization in Maps

- Smart mapping
  - relevant information
    at appropriate scale
- Vector Tiles
  - high-performance
  - re-styled to work in any map
- 3D
  - add context to the story

Custom Web GIS Visualization

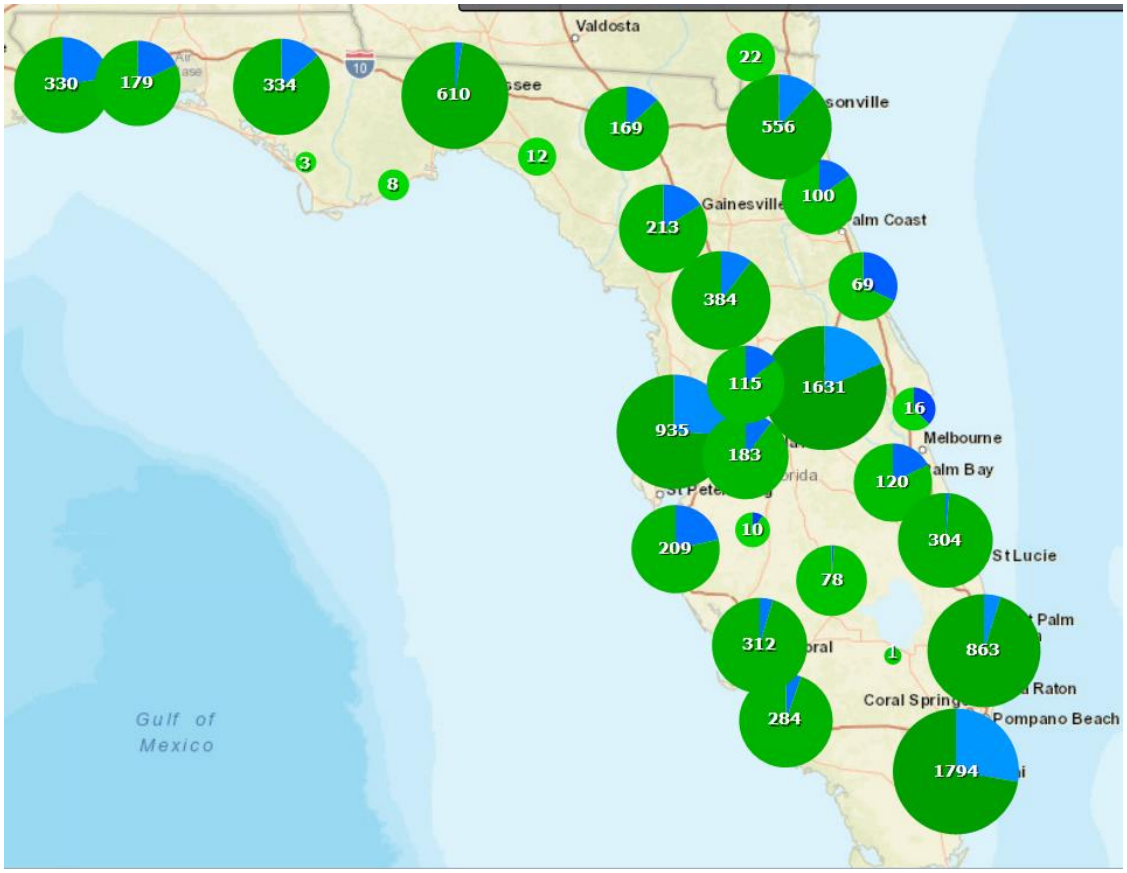
- More flexible
- Mix and match various technologies
- Integrate with relational and other databases
- But require more time to build

- Florida **Signal Four Analytics** – a web-based GIS centric crash data system


Signal Four Analytics

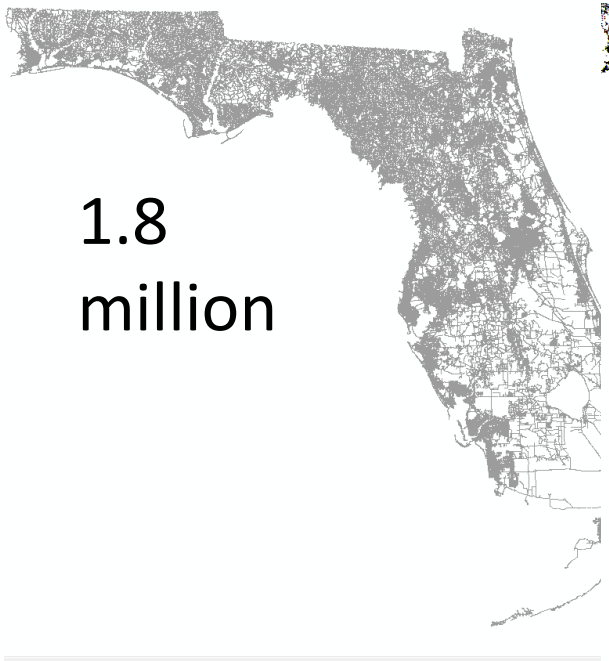
Crash and Traffic Citations Database

Over 10 million crashes & citations



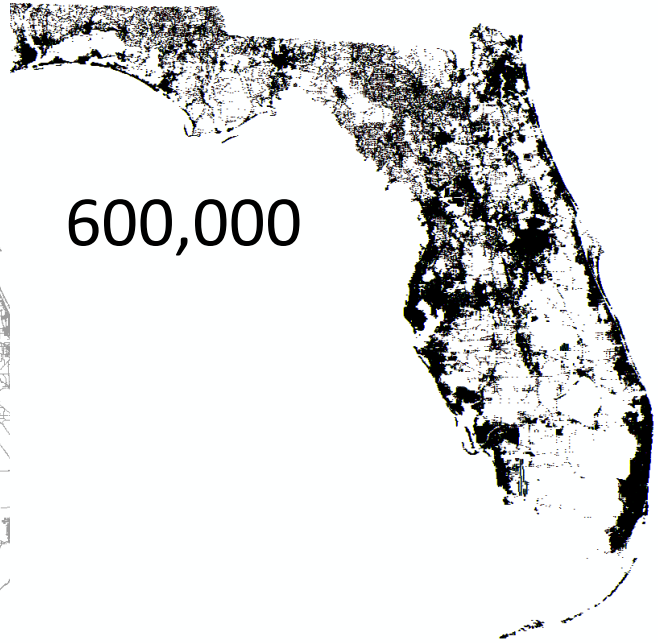
Street Network– Florida Unified Basemap

LINKS



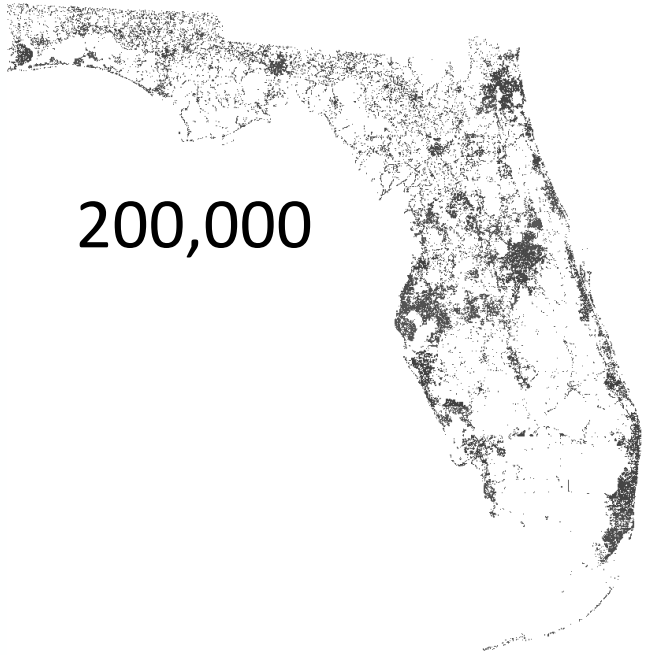
1.8 million

INTERSECTIONS



600,000

CURVES



200,000

SIGNAL FOUR ANALYTICS

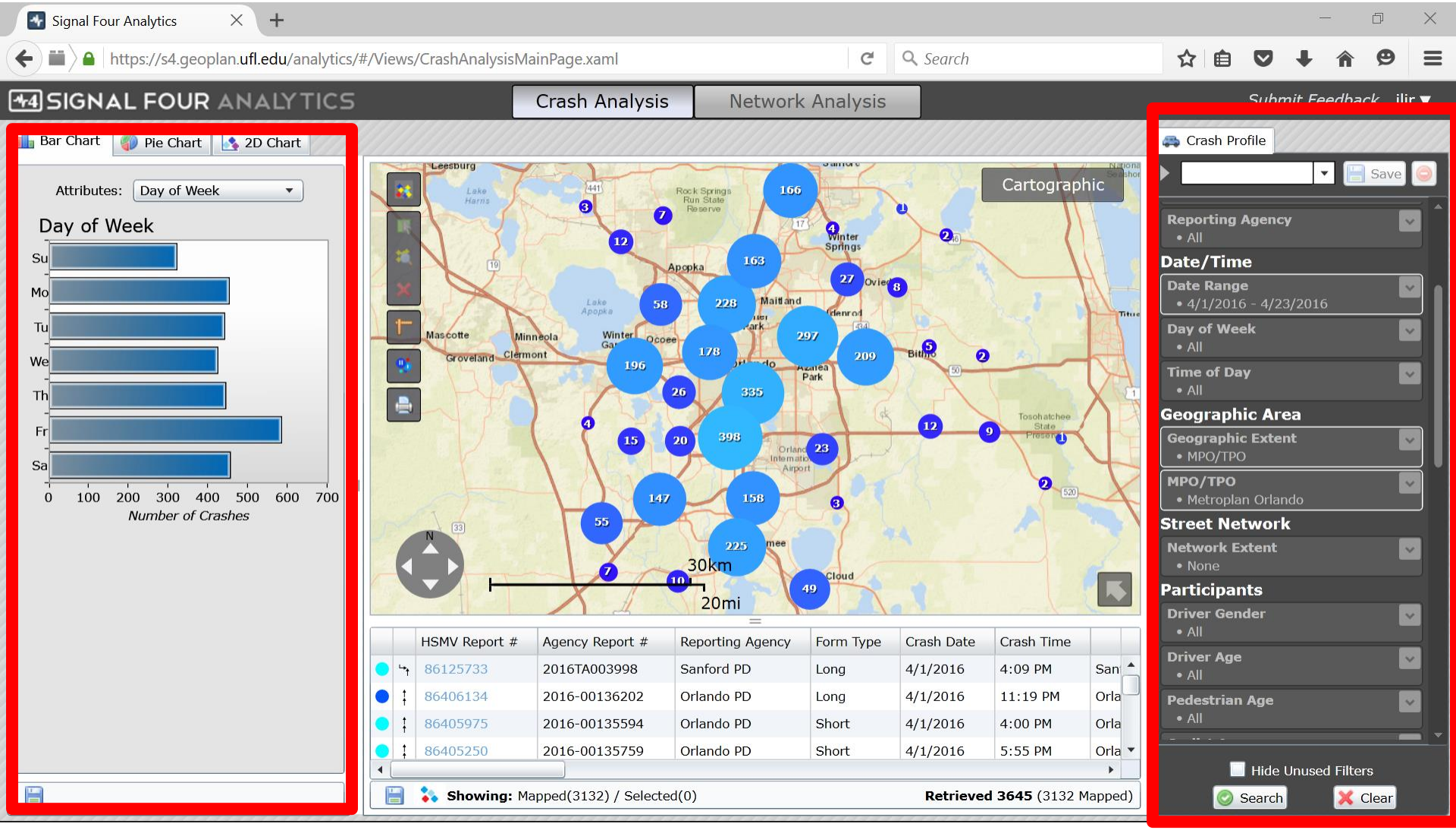
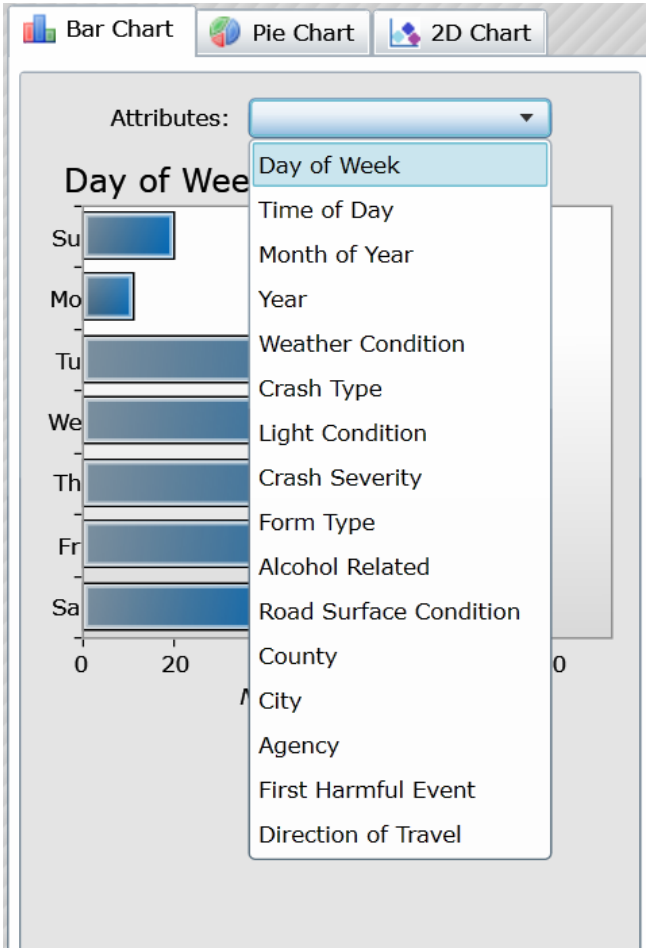
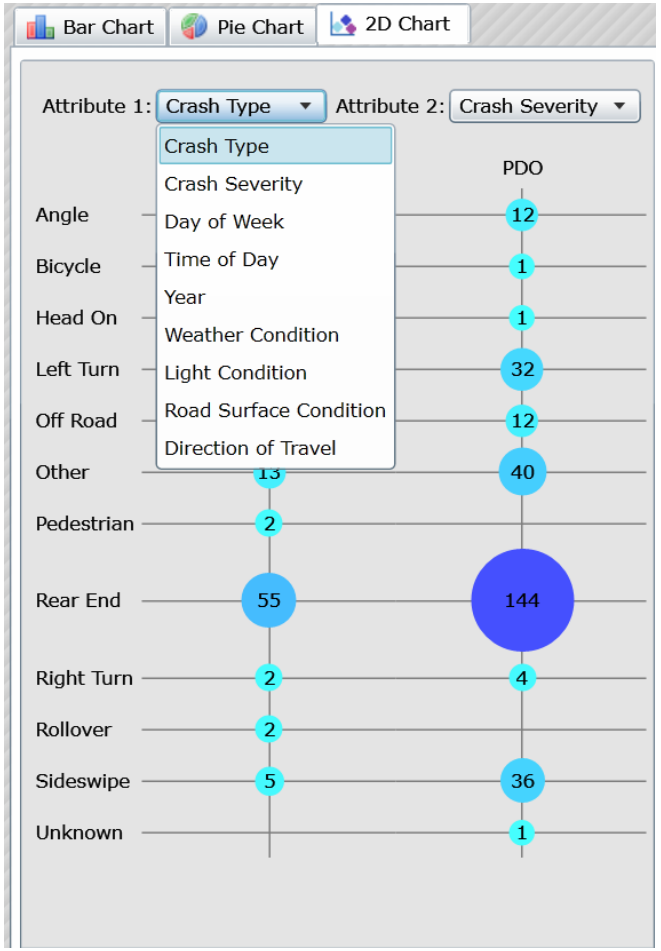
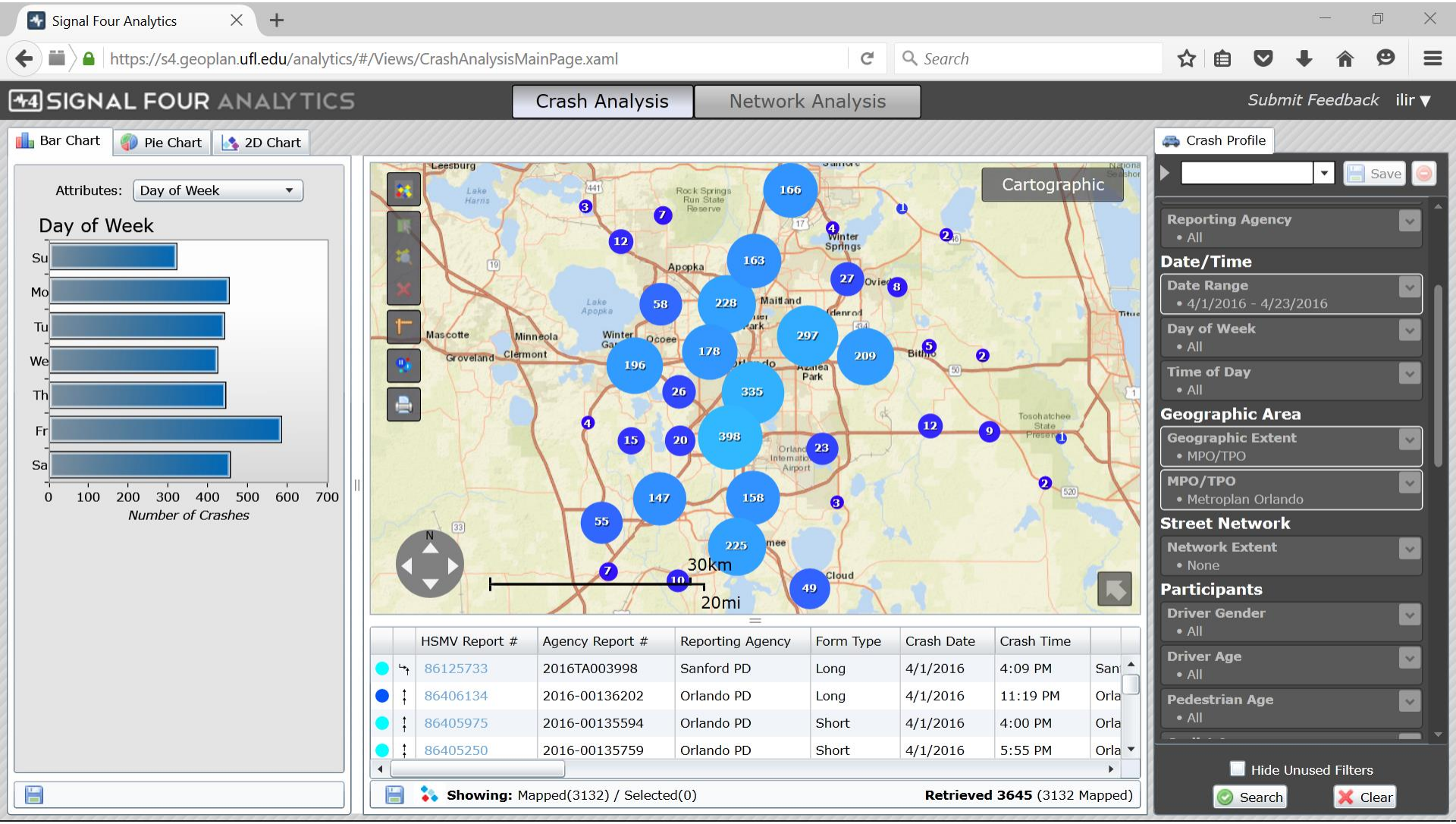


Chart Attributes

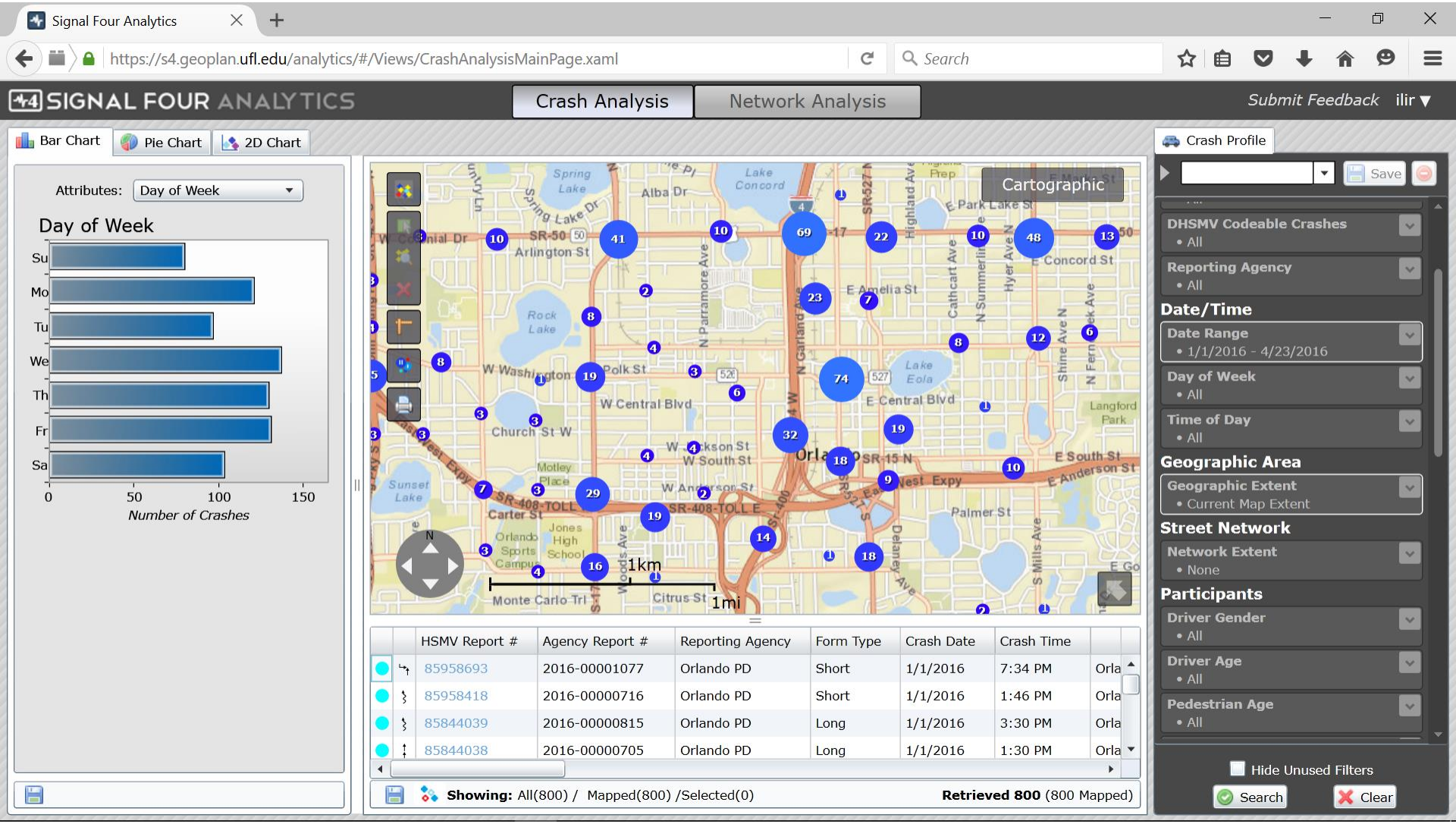




SIGNAL FOUR ANALYTICS

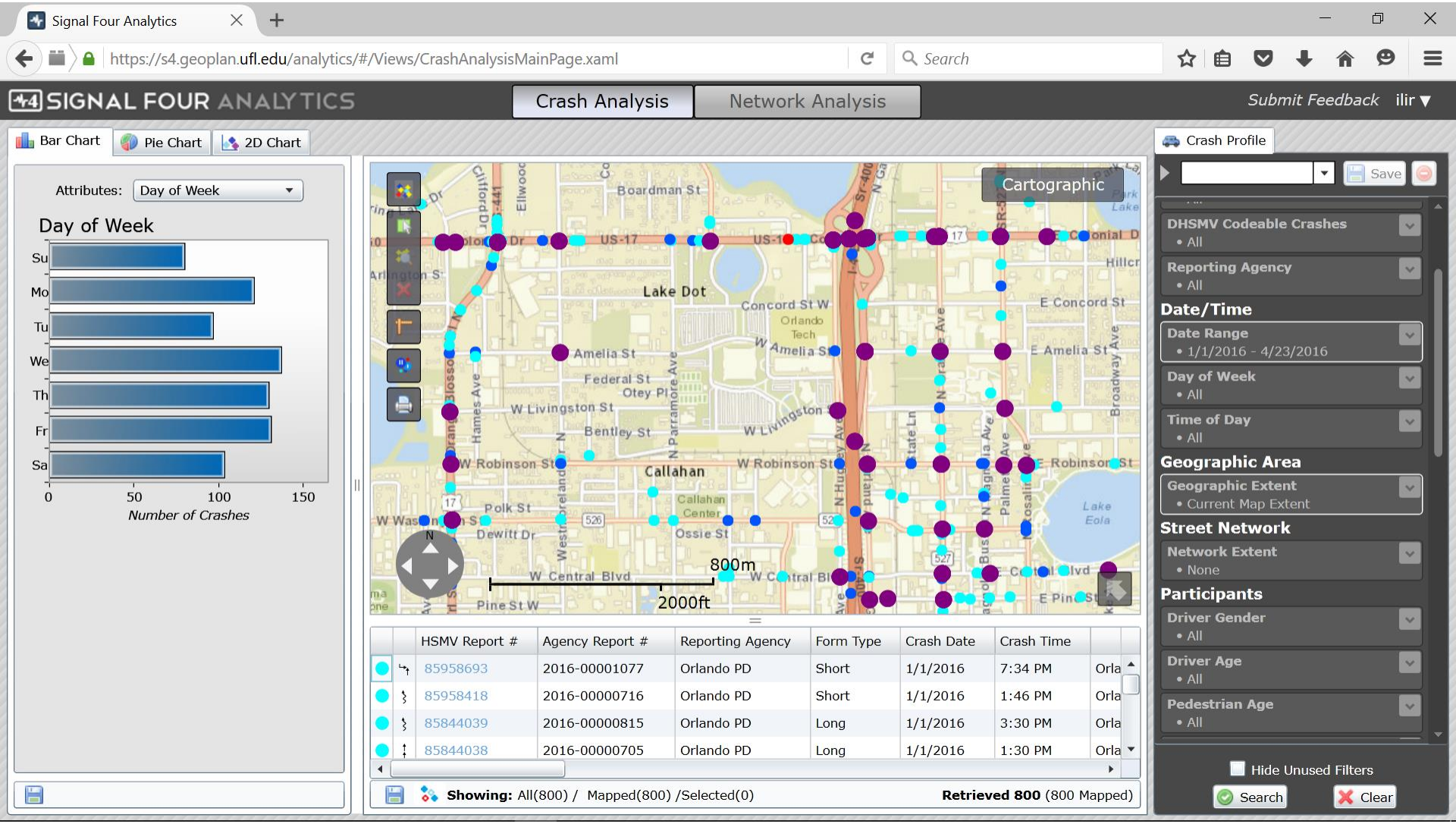


SIGNAL FOUR ANALYTICS

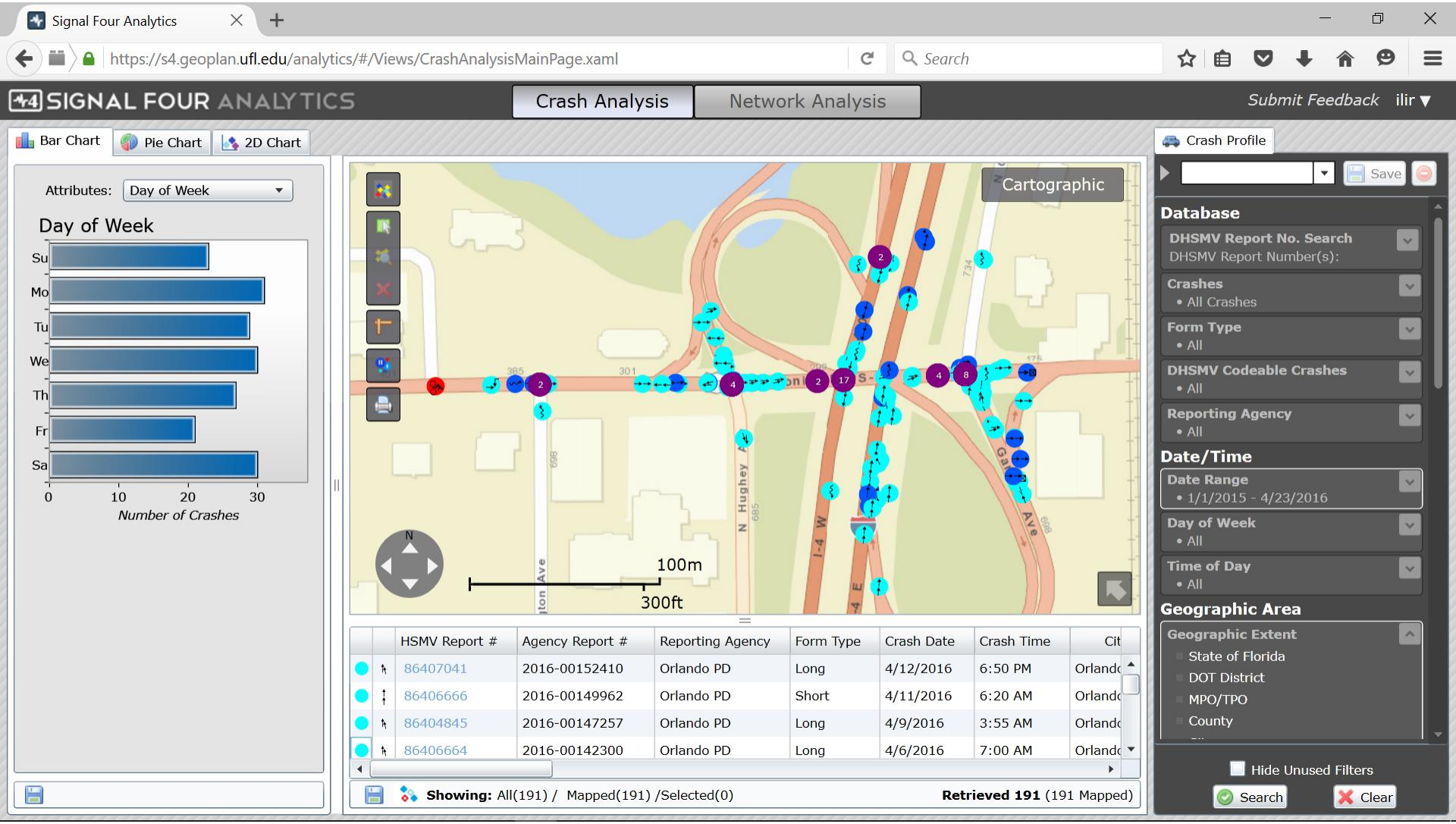


HSMV Report #	Agency Report #	Reporting Agency	Form Type	Crash Date	Crash Time	
85958693	2016-00001077	Orlando PD	Short	1/1/2016	7:34 PM	Orla
85958418	2016-00000716	Orlando PD	Short	1/1/2016	1:46 PM	Orla
85844039	2016-00000815	Orlando PD	Long	1/1/2016	3:30 PM	Orla
85844038	2016-00000705	Orlando PD	Long	1/1/2016	1:30 PM	Orla

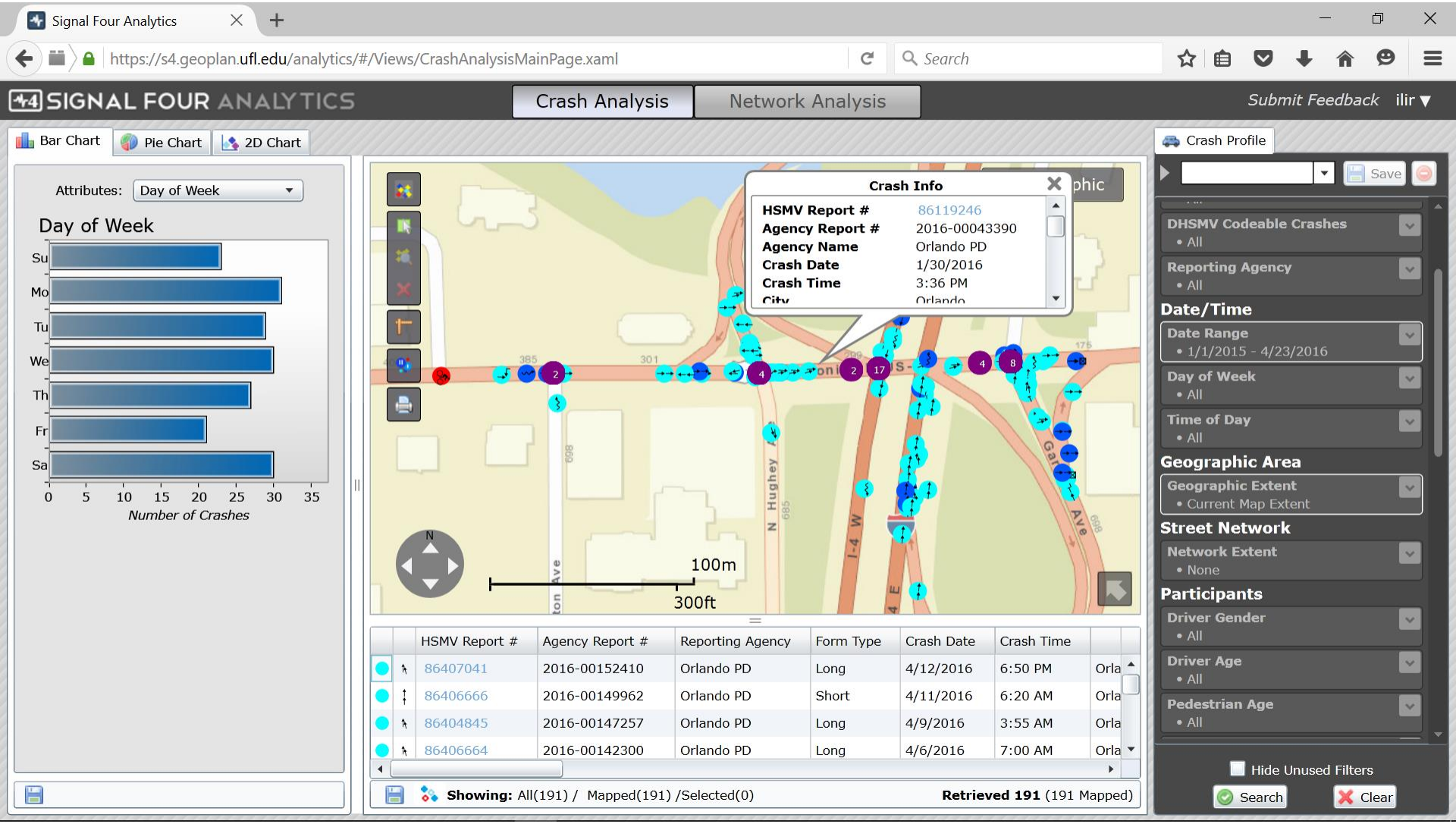
SIGNAL FOUR ANALYTICS



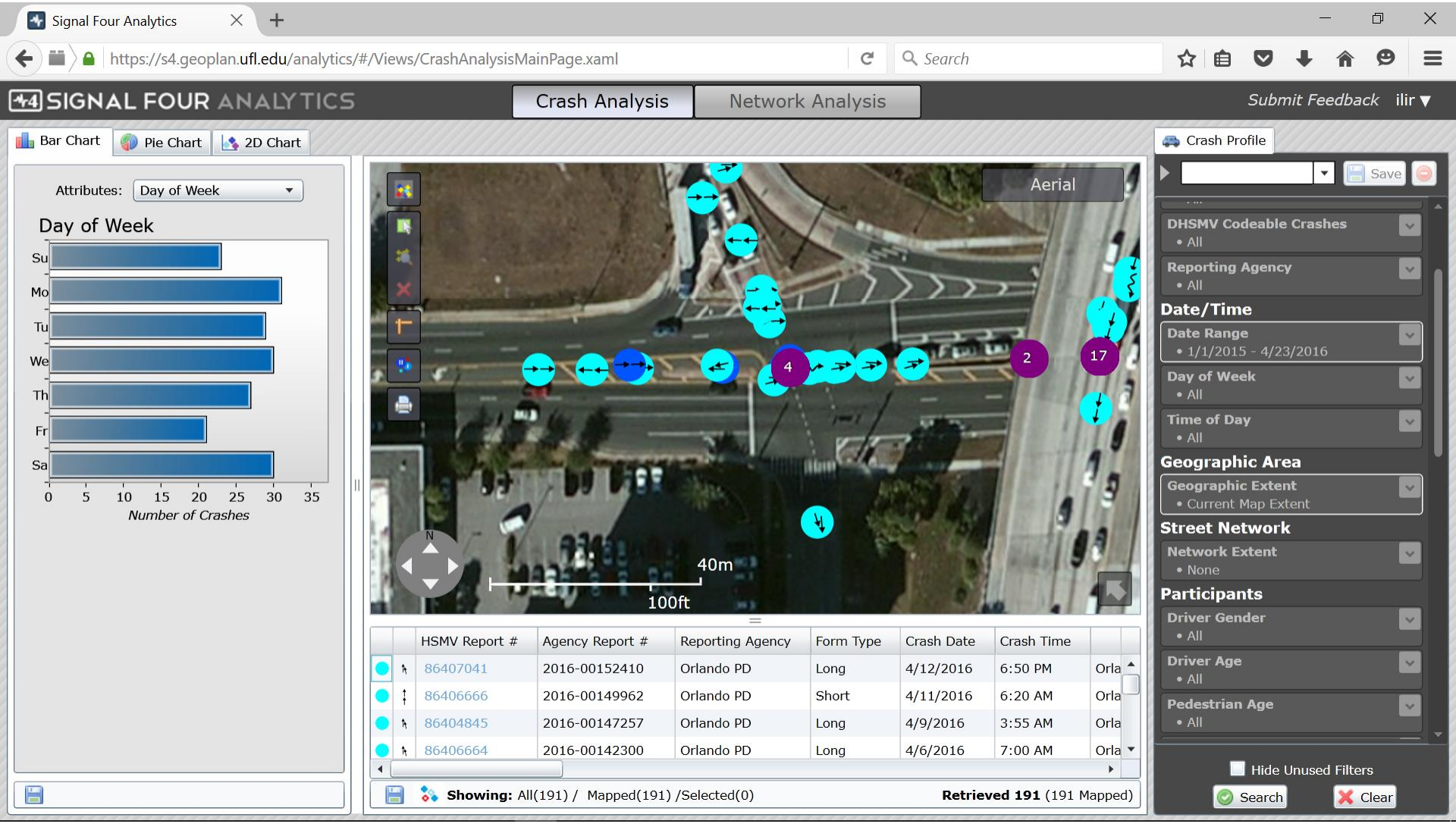
SIGNAL FOUR ANALYTICS



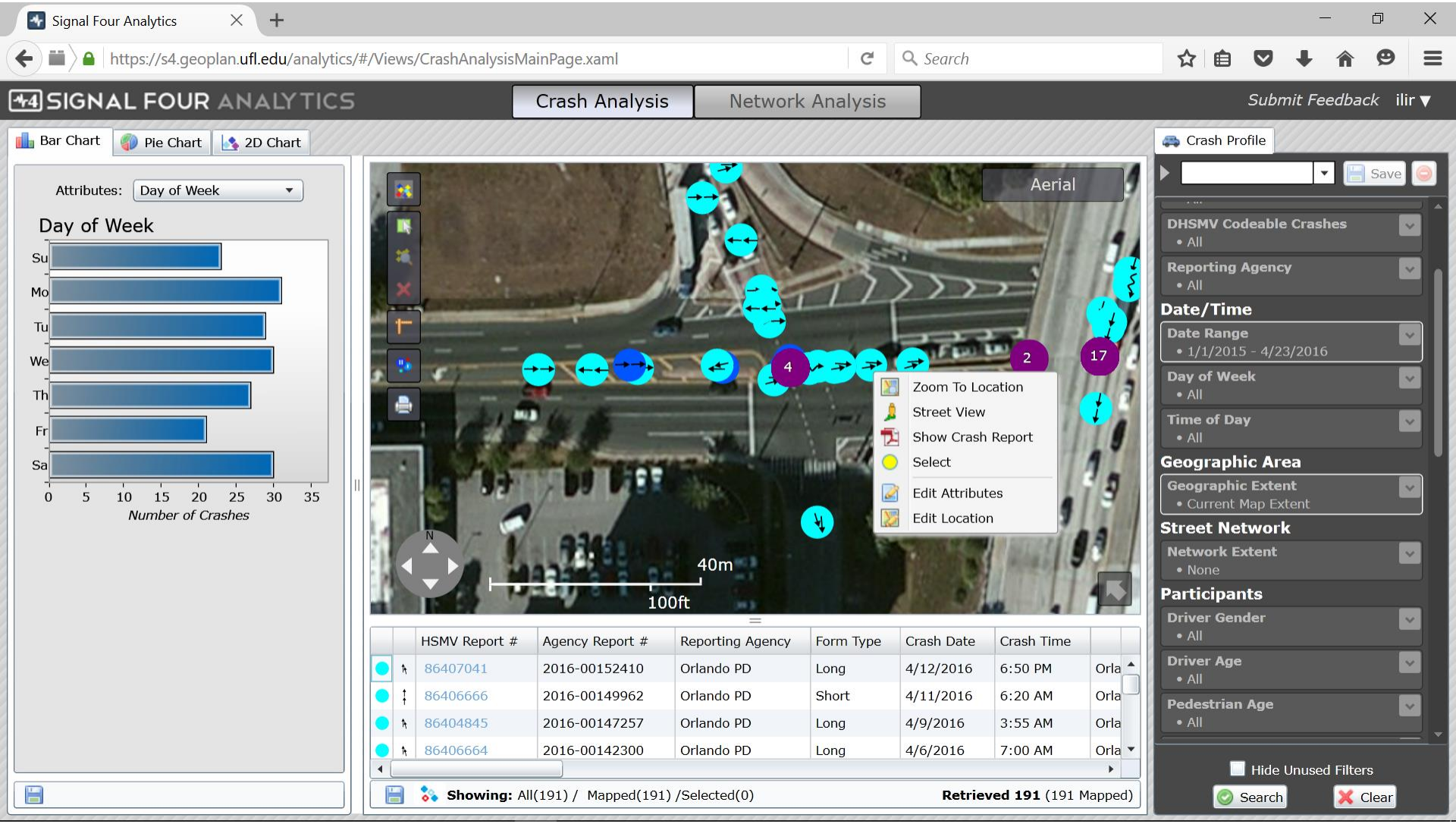
SIGNAL FOUR ANALYTICS



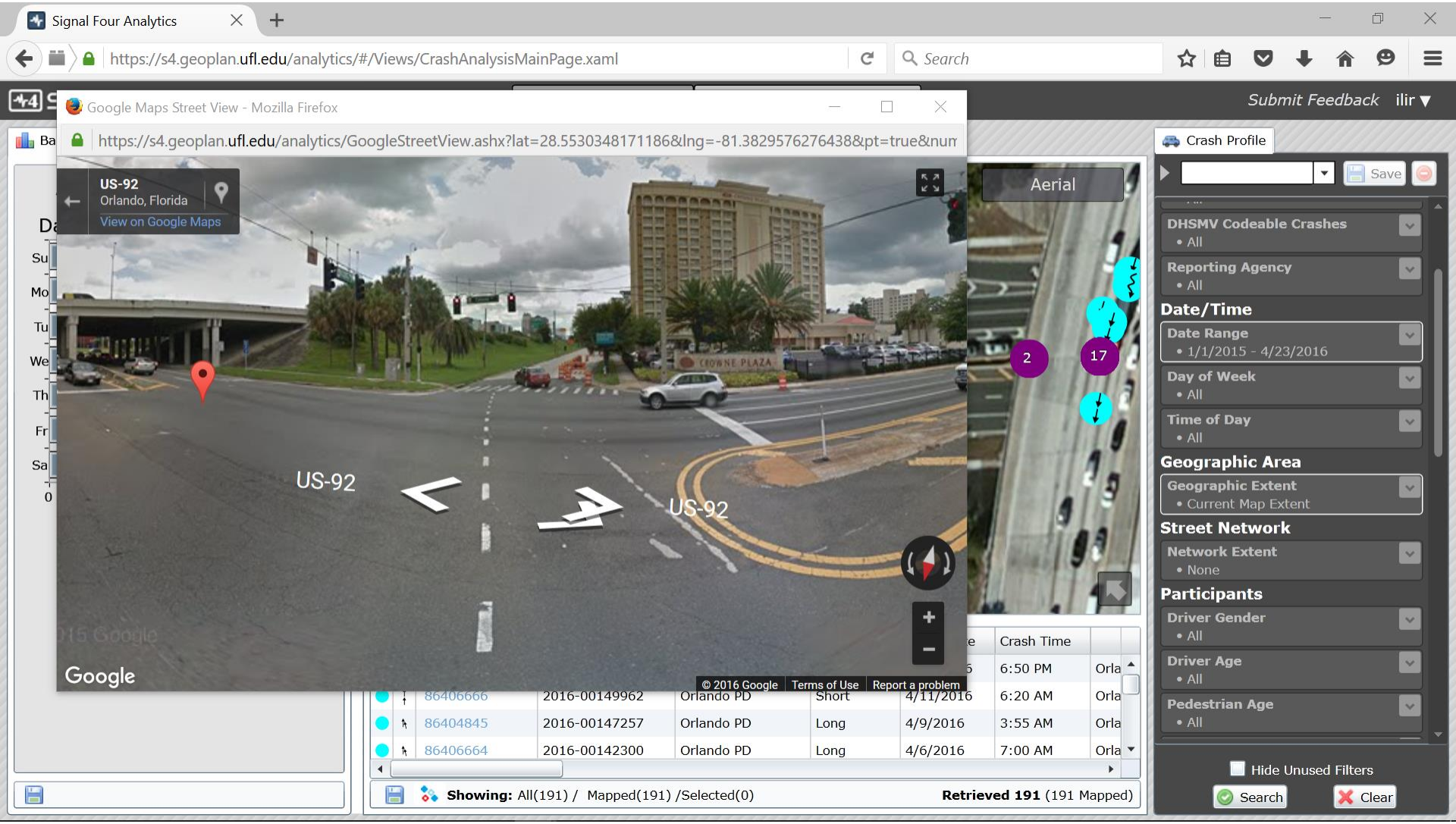
SIGNAL FOUR ANALYTICS



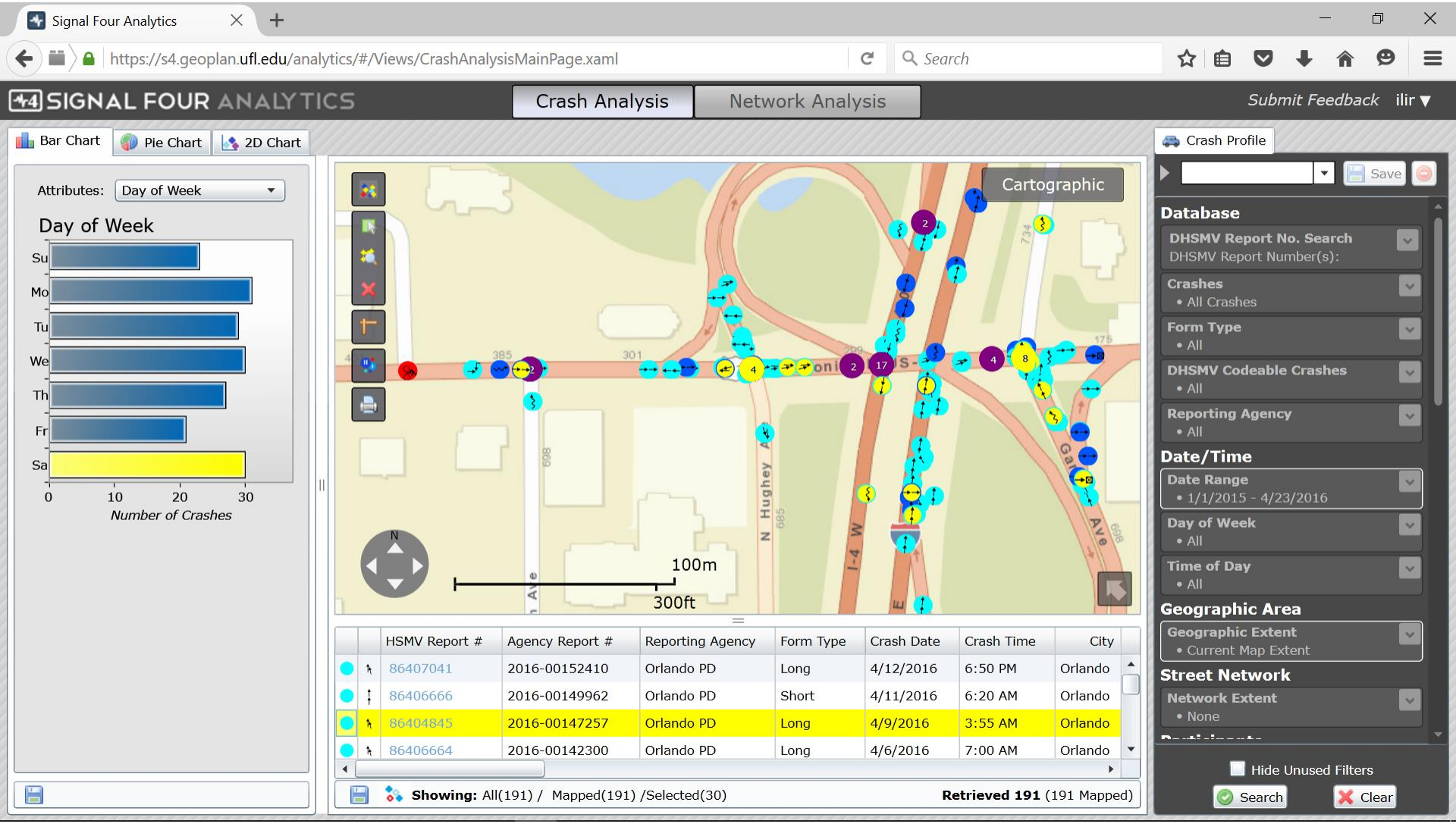
SIGNAL FOUR ANALYTICS



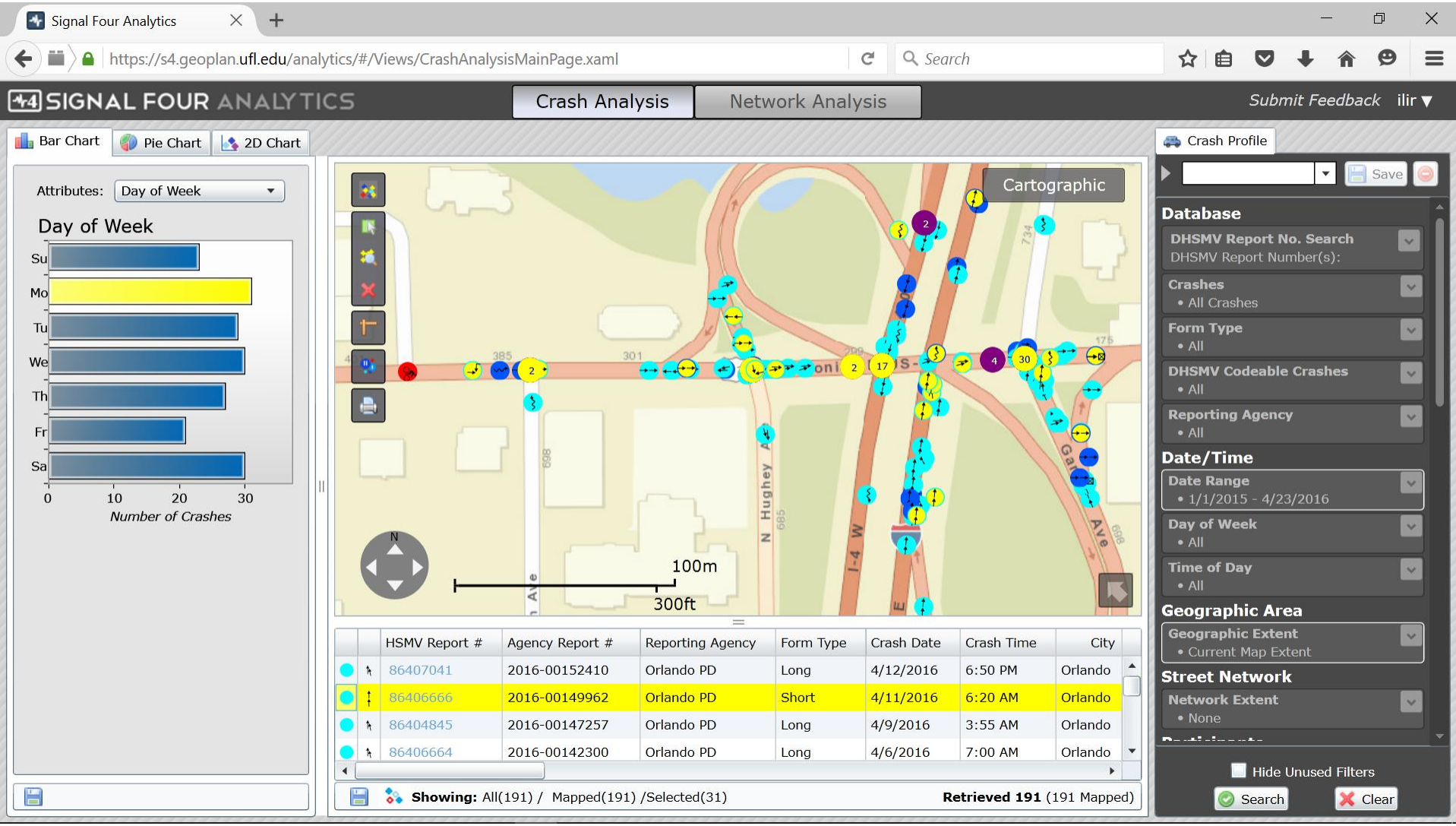
SIGNAL FOUR ANALYTICS



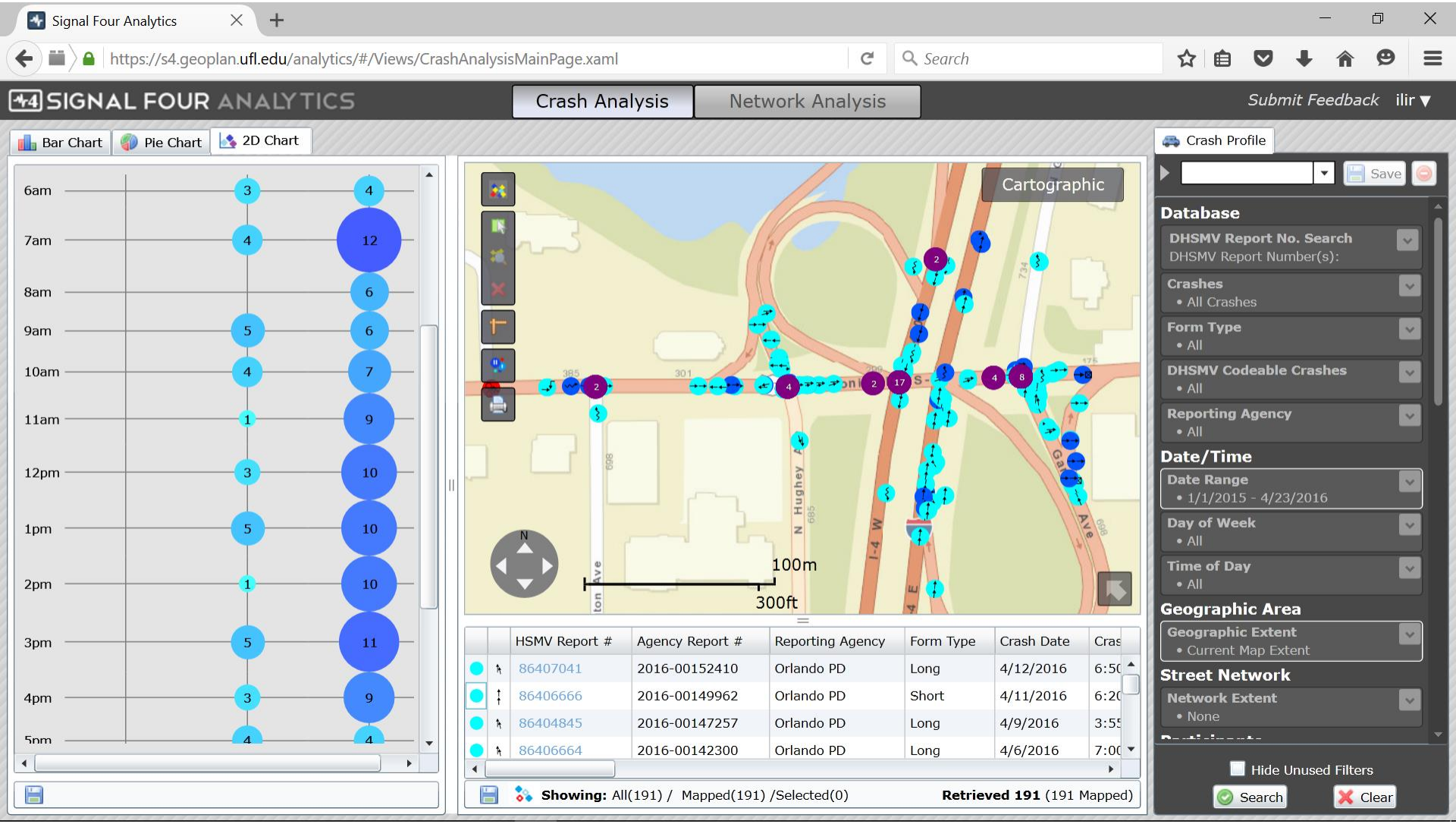
SIGNAL FOUR ANALYTICS



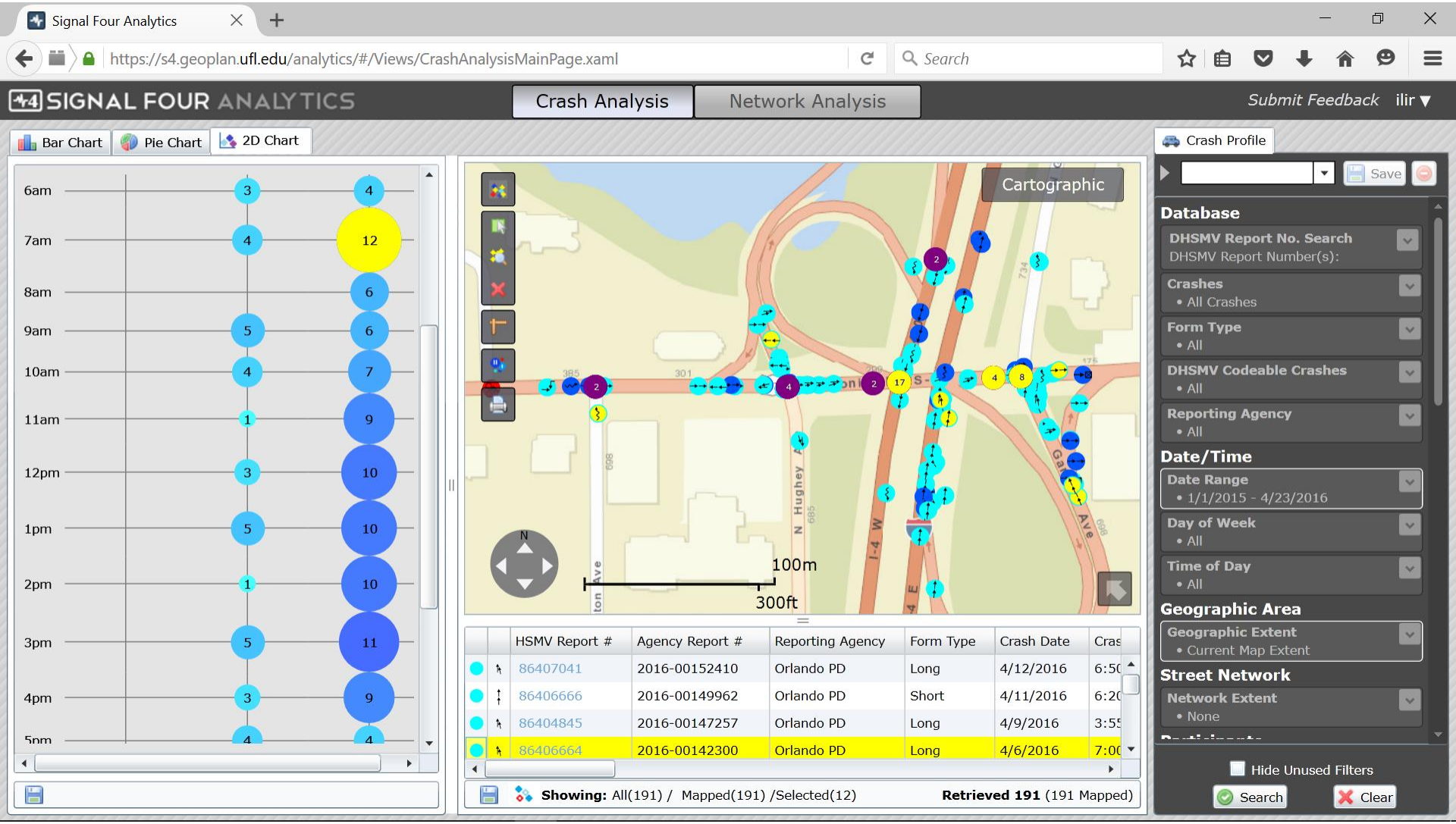
SIGNAL FOUR ANALYTICS



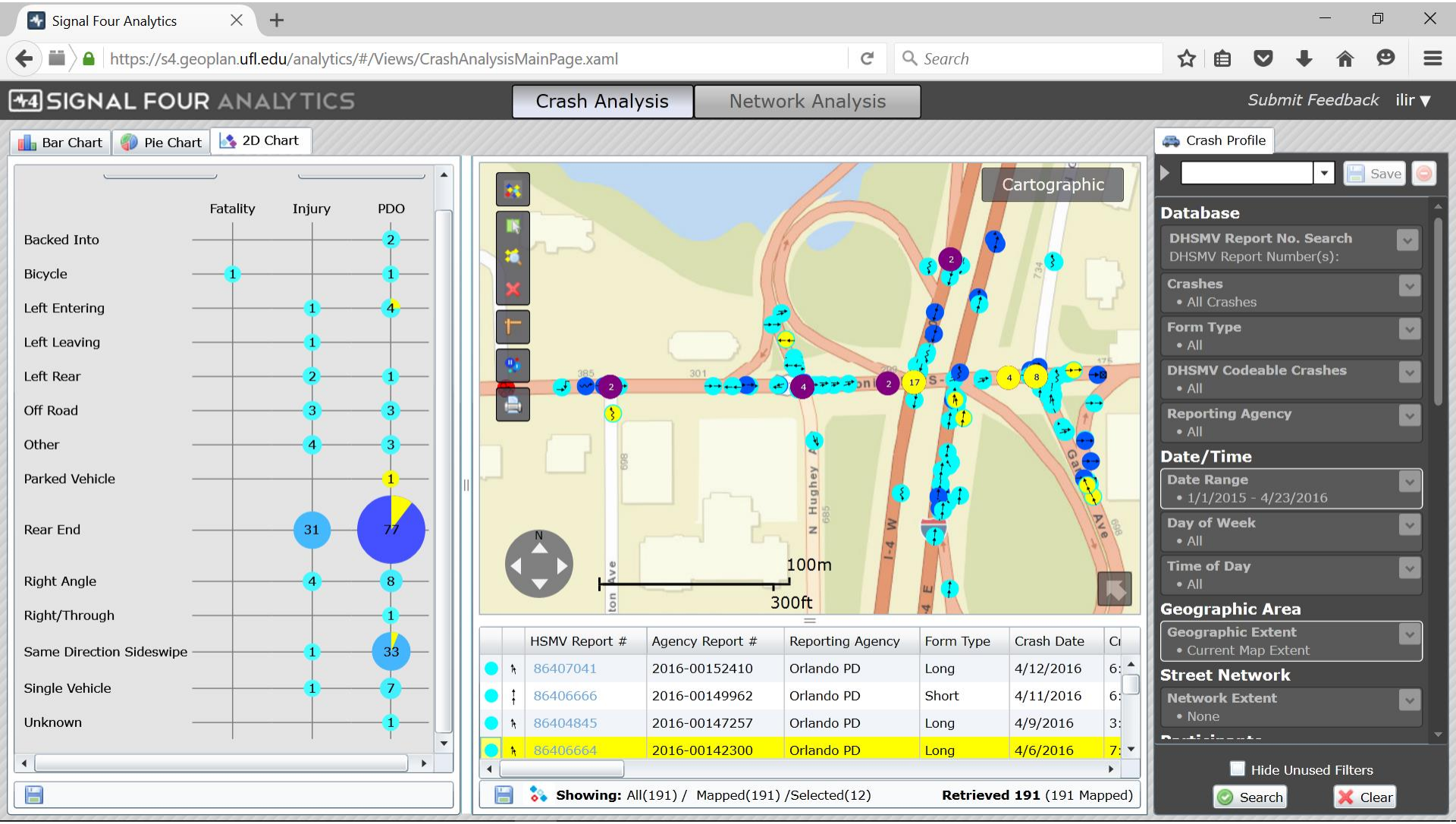
SIGNAL FOUR ANALYTICS



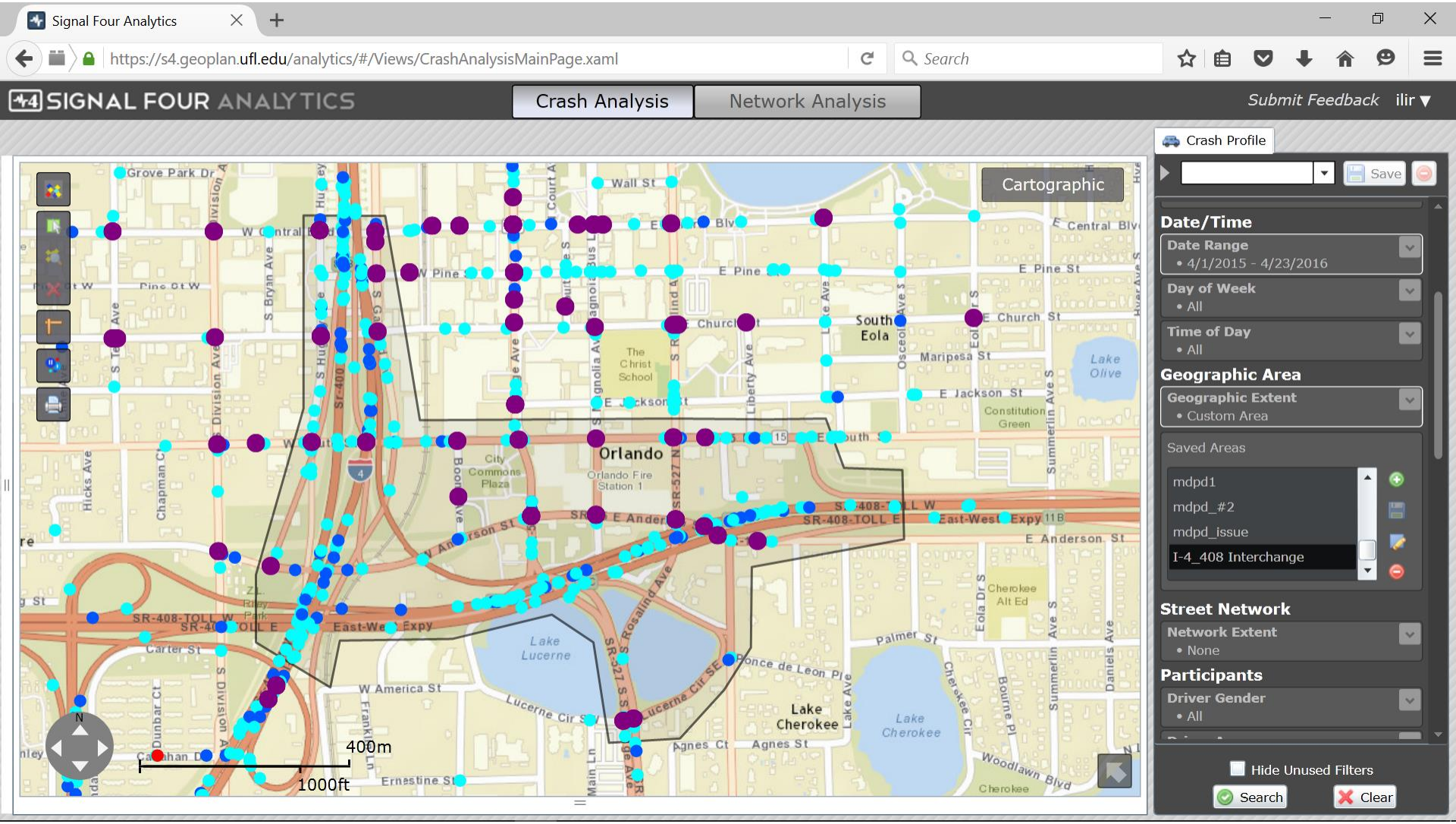
SIGNAL FOUR ANALYTICS



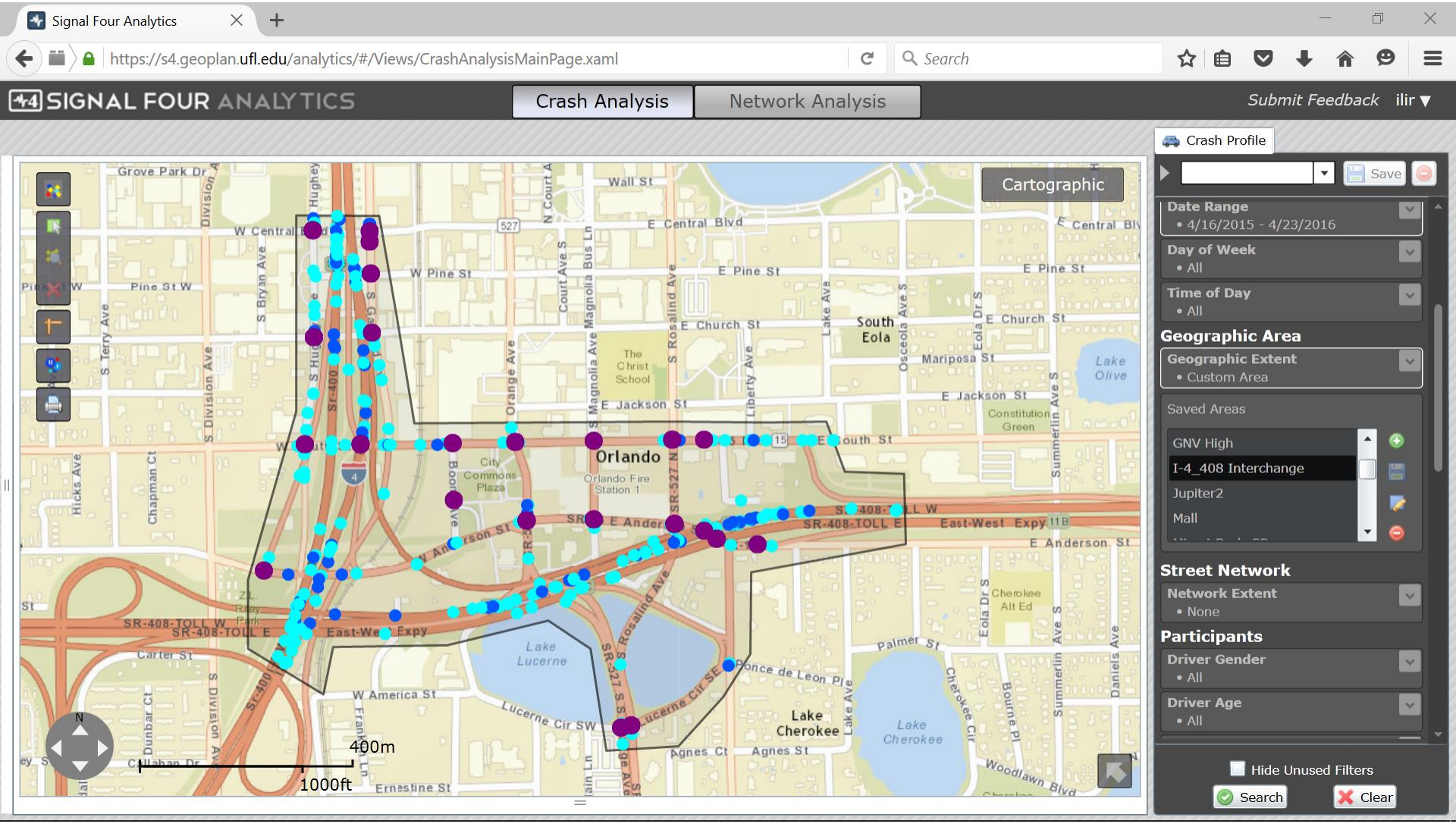
SIGNAL FOUR ANALYTICS



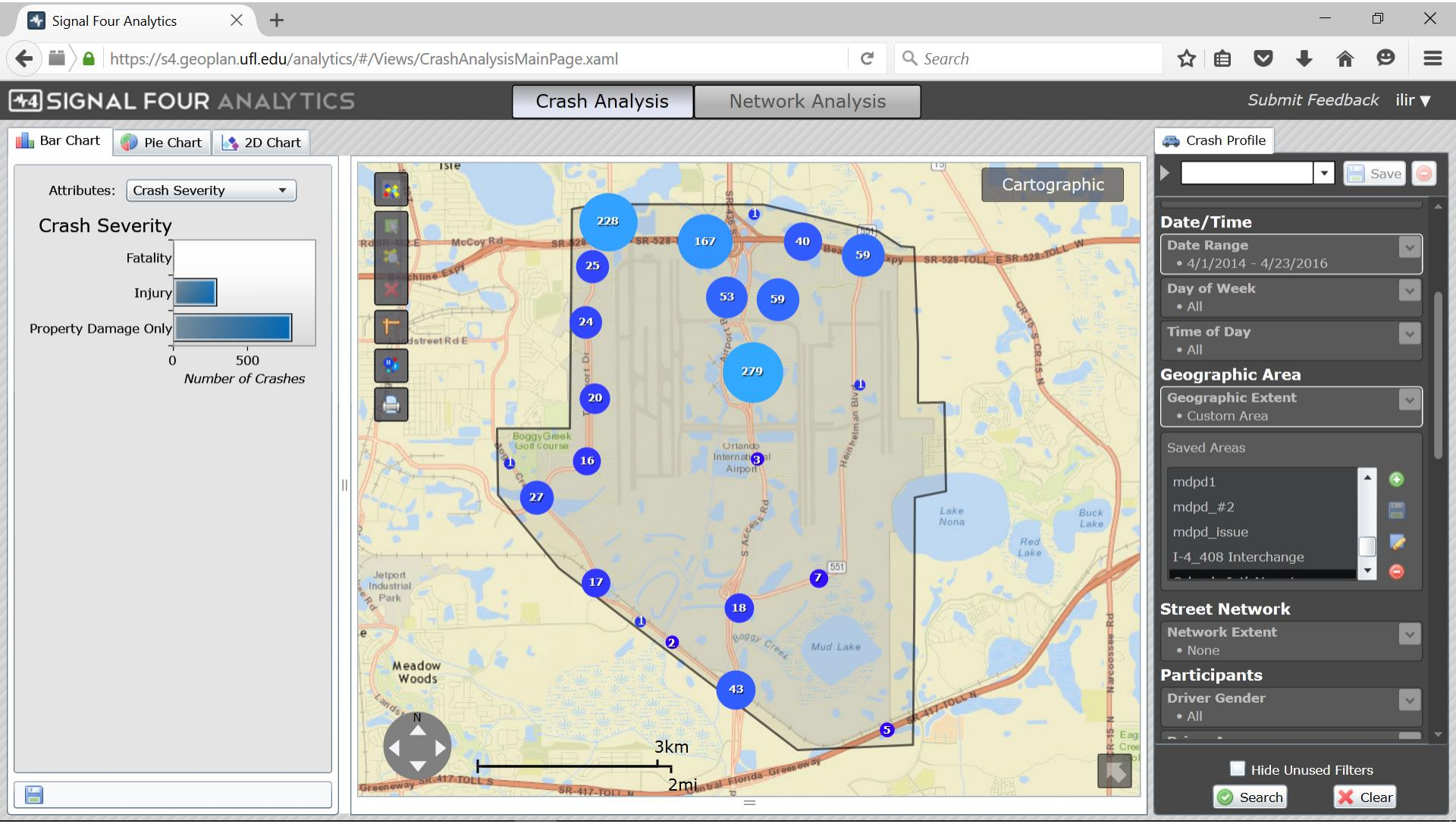
SIGNAL FOUR ANALYTICS



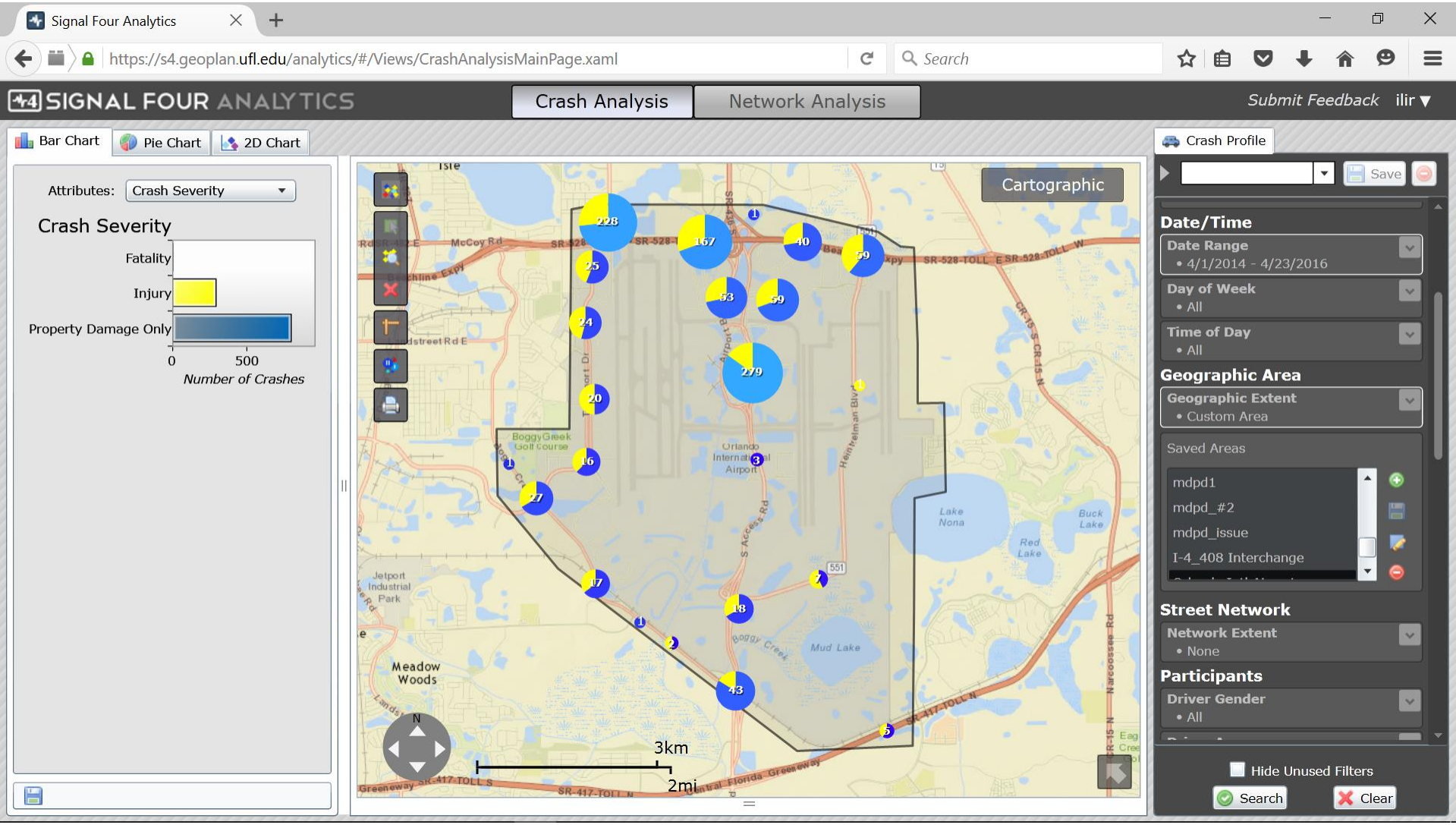
SIGNAL FOUR ANALYTICS



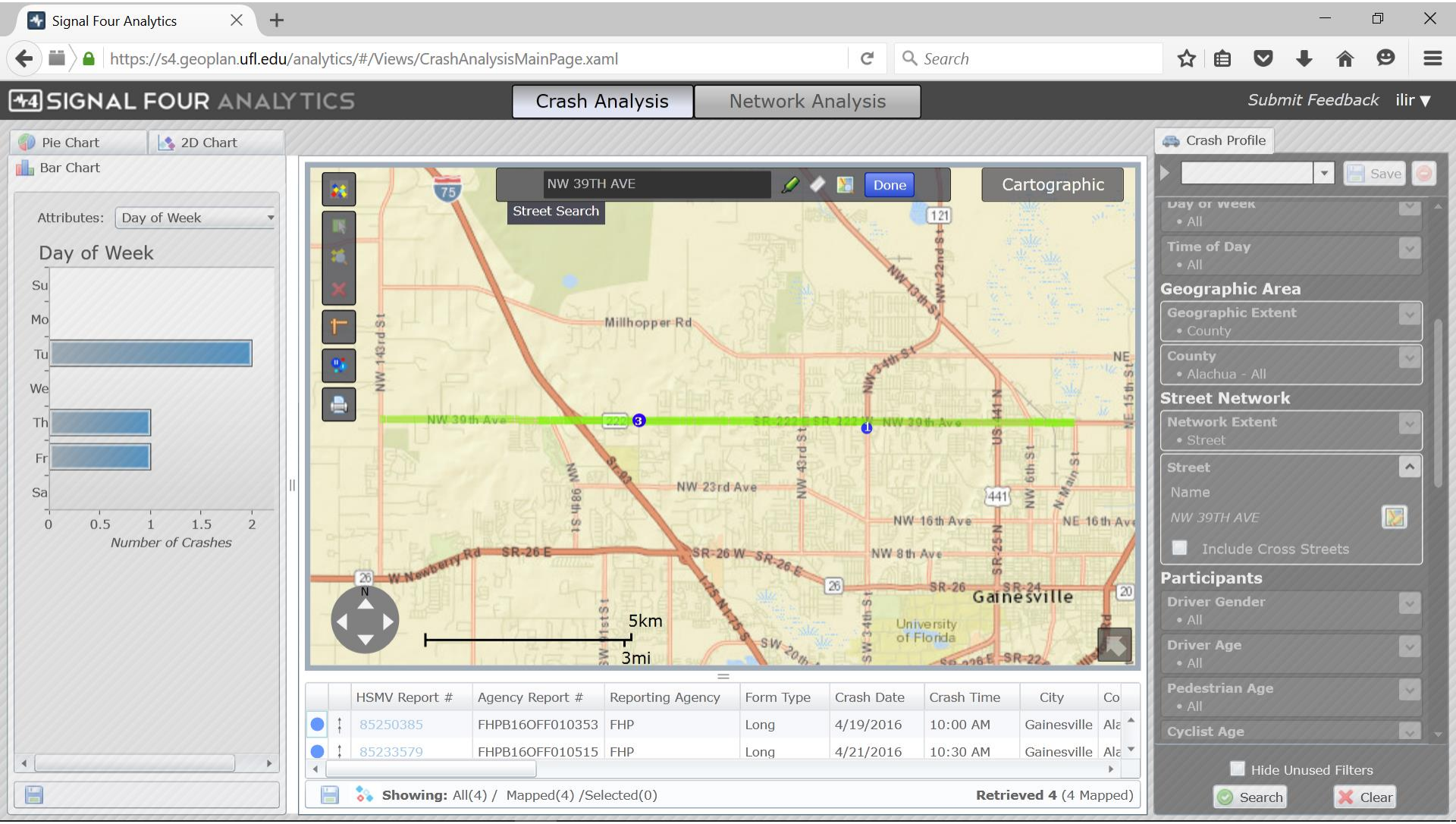
SIGNAL FOUR ANALYTICS



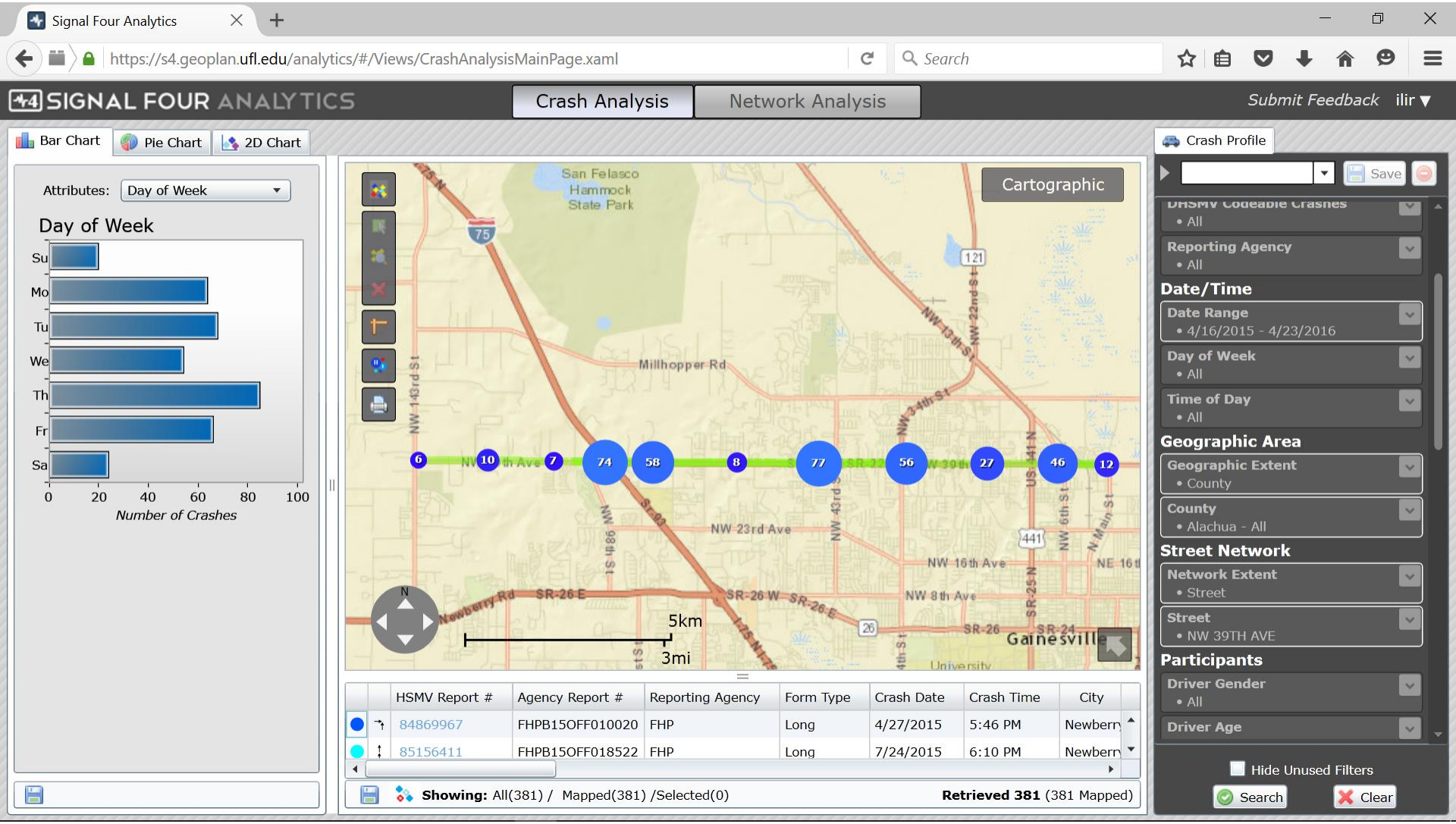
SIGNAL FOUR ANALYTICS



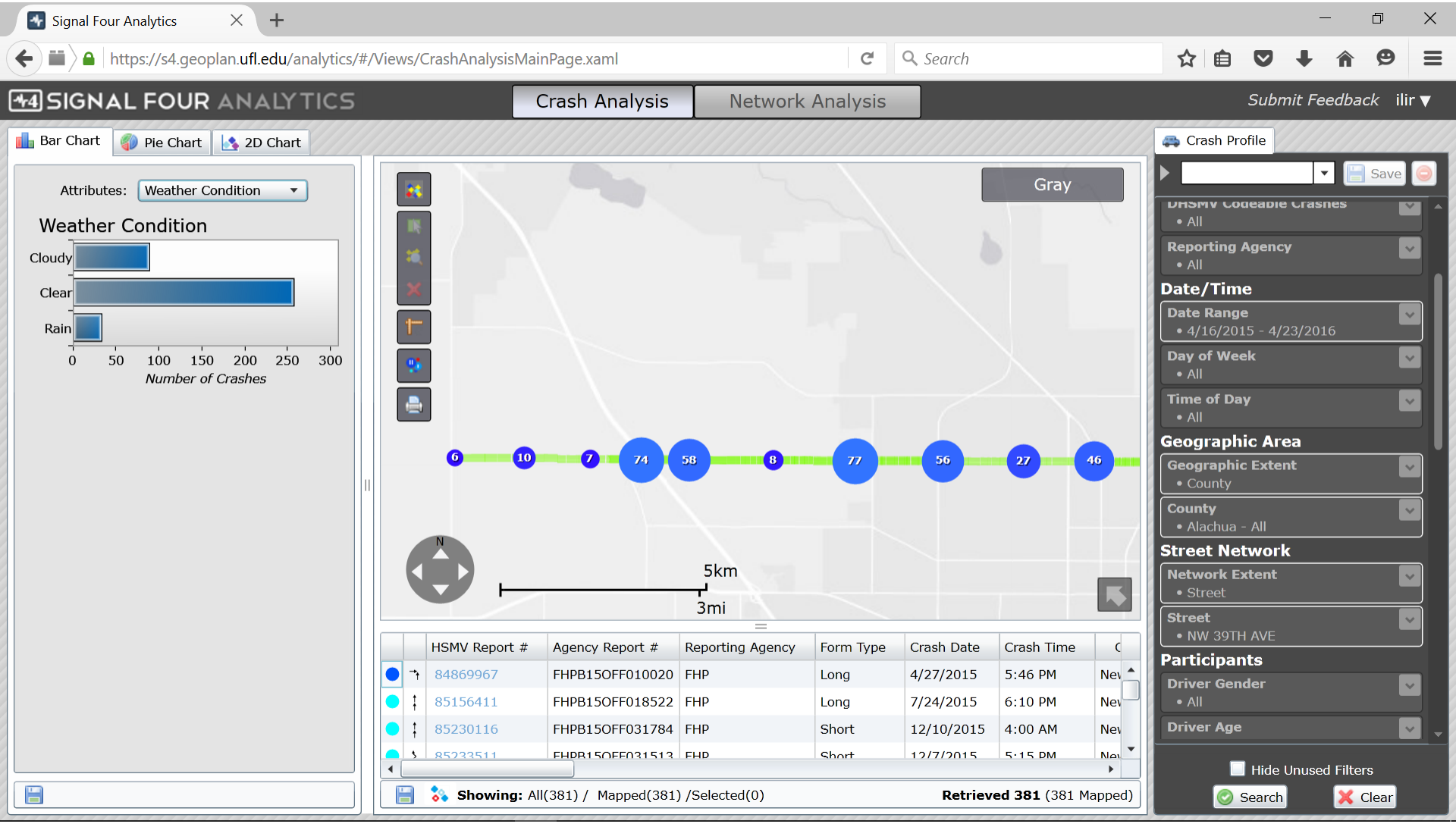
SIGNAL FOUR ANALYTICS



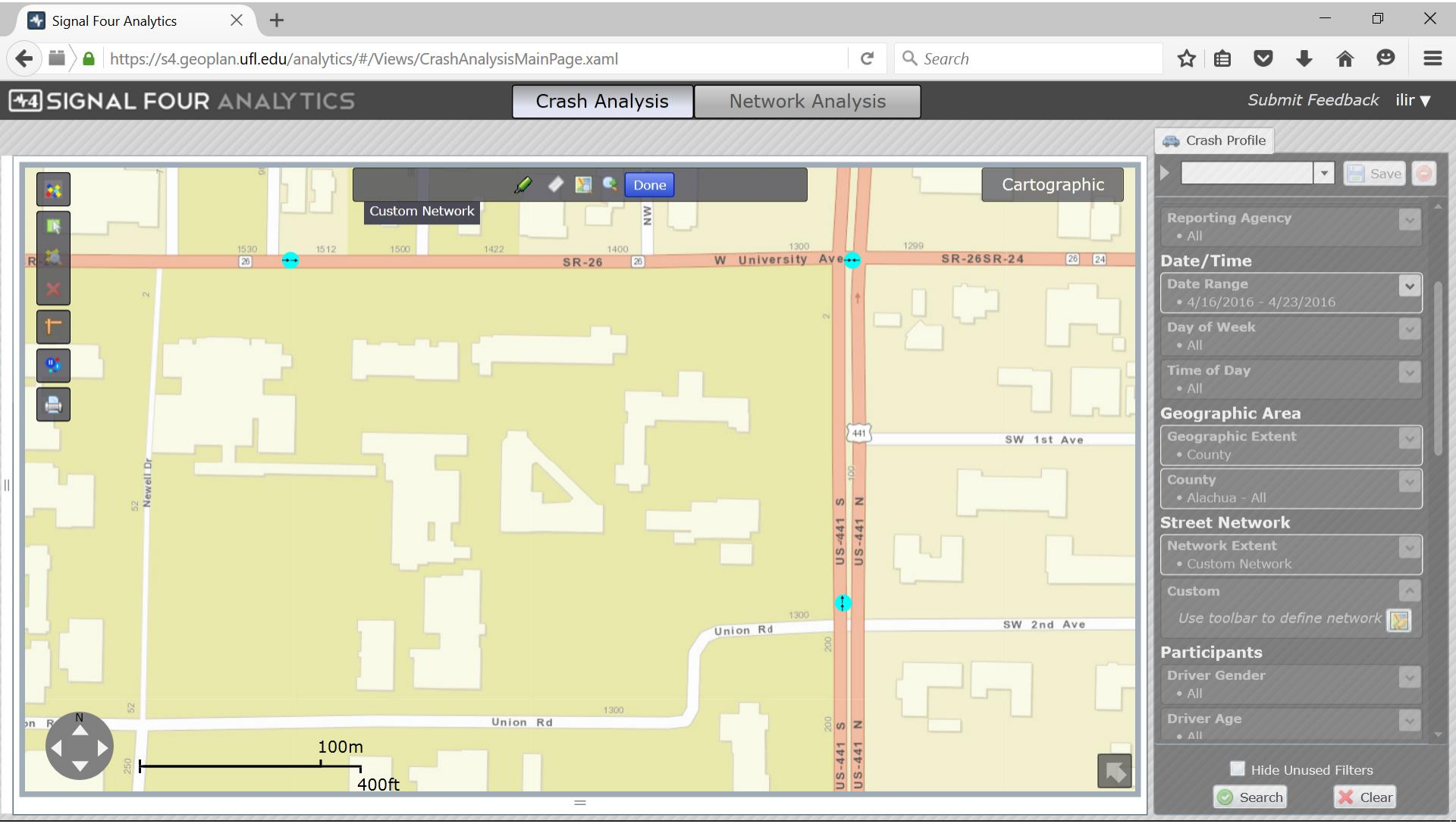
SIGNAL FOUR ANALYTICS



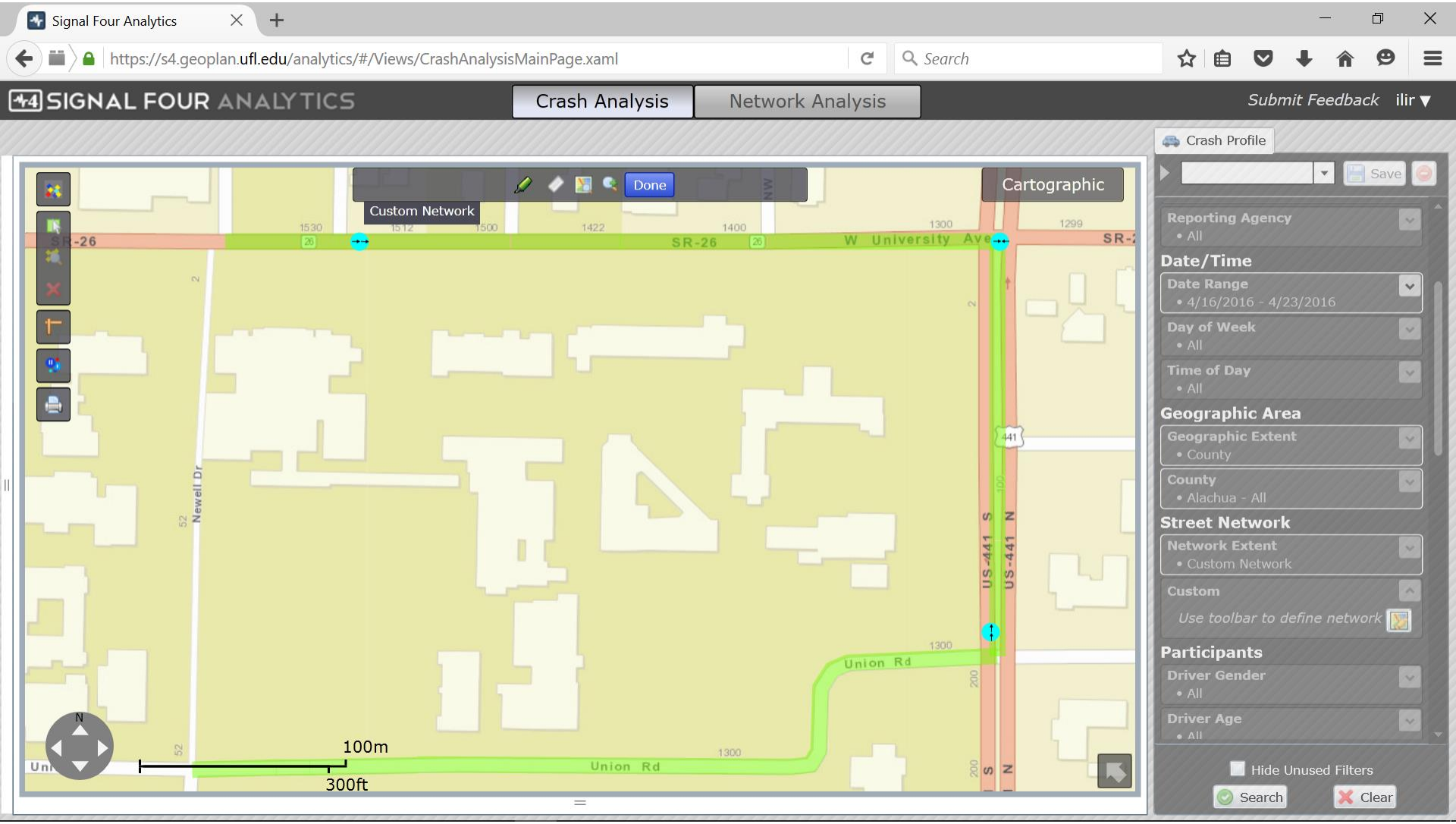
SIGNAL FOUR ANALYTICS



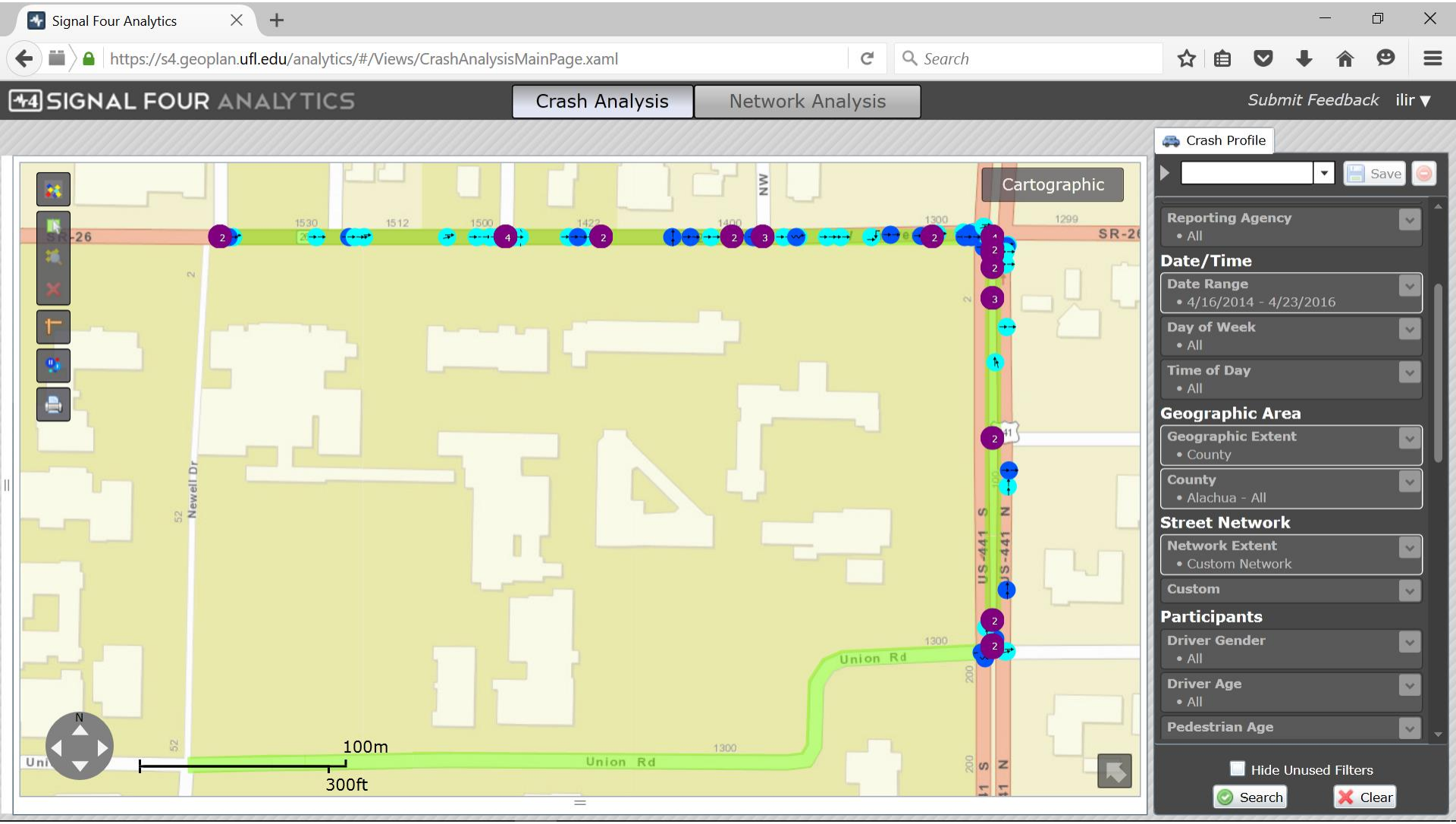
SIGNAL FOUR ANALYTICS

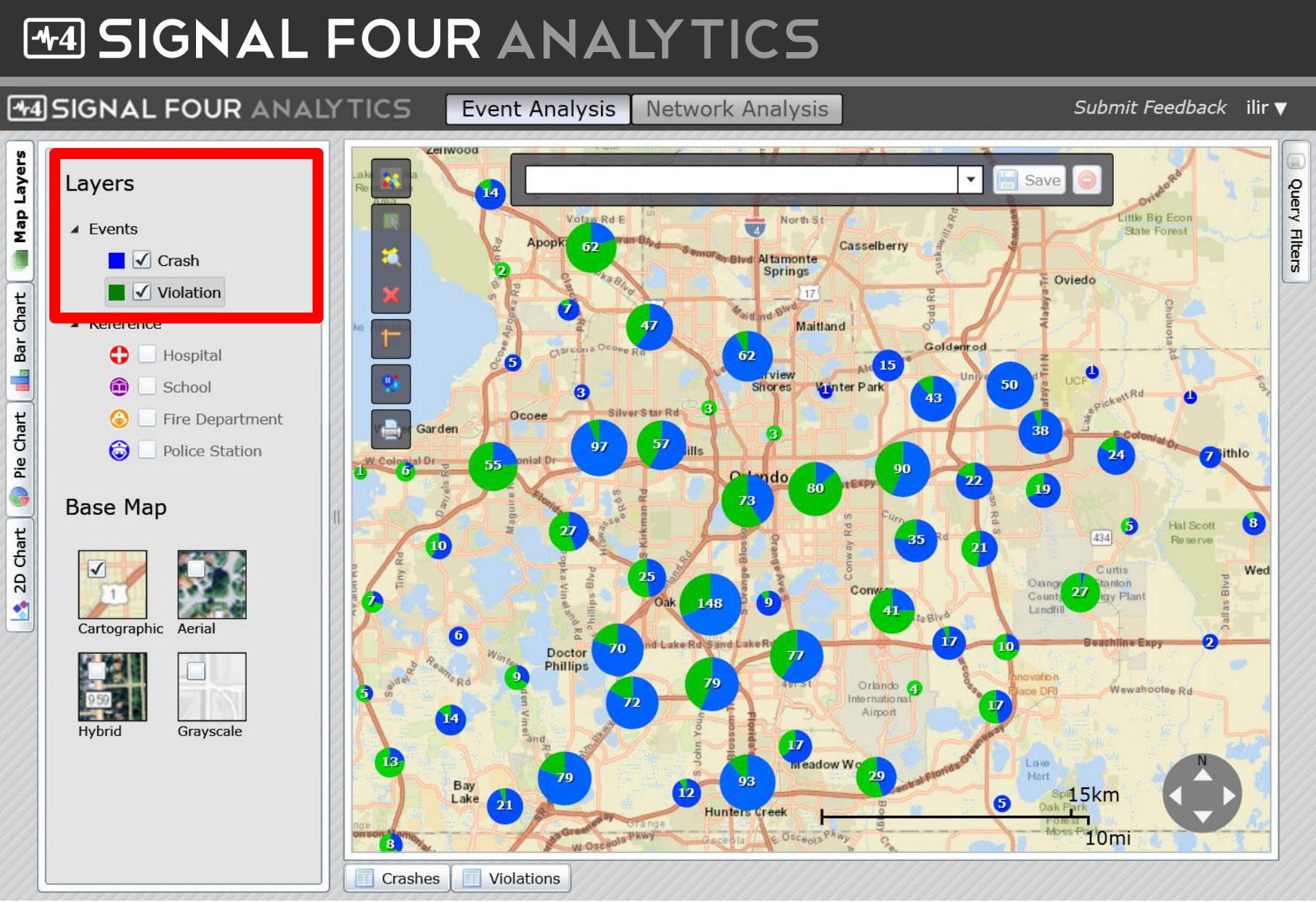


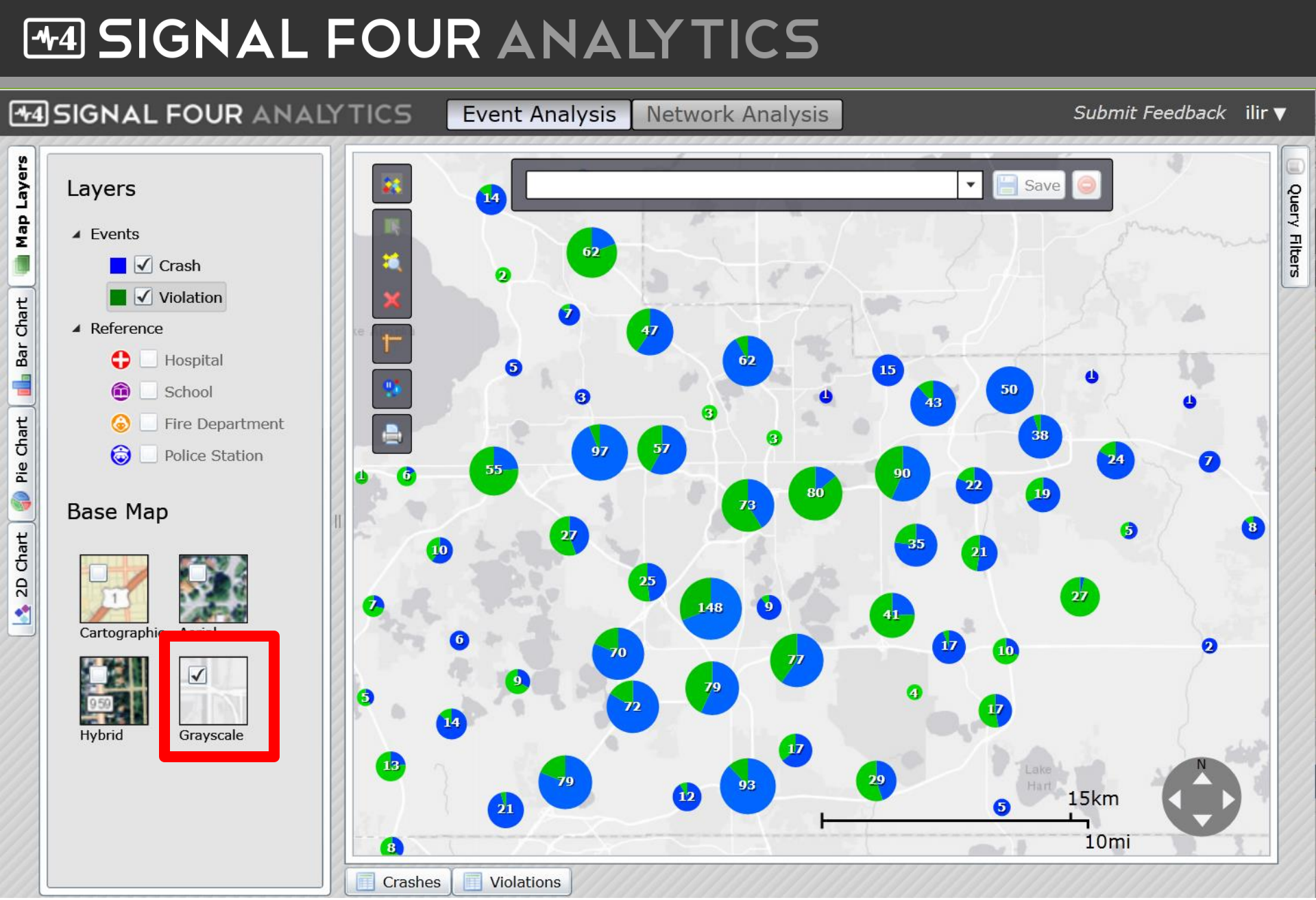
SIGNAL FOUR ANALYTICS



SIGNAL FOUR ANALYTICS




SIGNAL FOUR ANALYTICS
SIGNAL FOUR ANALYTICS
Event Analysis
Network Analysis
Submit Feedback
ilir
Map Layers
Bar Chart
Pie Chart
2D Chart
Layers
Events
Crash
Violation
Reference
Hospital
School
Fire Department
Police Station
Base Map
Cartographic
Aerial
Hybrid
Grayscale
Save
Query Filters
Crashes
Violations
15km
10mi
N


SIGNAL FOUR ANALYTICS
SIGNAL FOUR ANALYTICS
Event Analysis
Network Analysis
Submit Feedback
ilir
Map Layers
Bar Chart
Pie Chart
2D Chart
Layers
Events
Crash
Violation
Reference
Hospital
School
Fire Department
Police Station
Base Map
Cartographic
Hybrid
Grayscale
Save
Query Filters
15km
10mi
N
Crashes
Violations

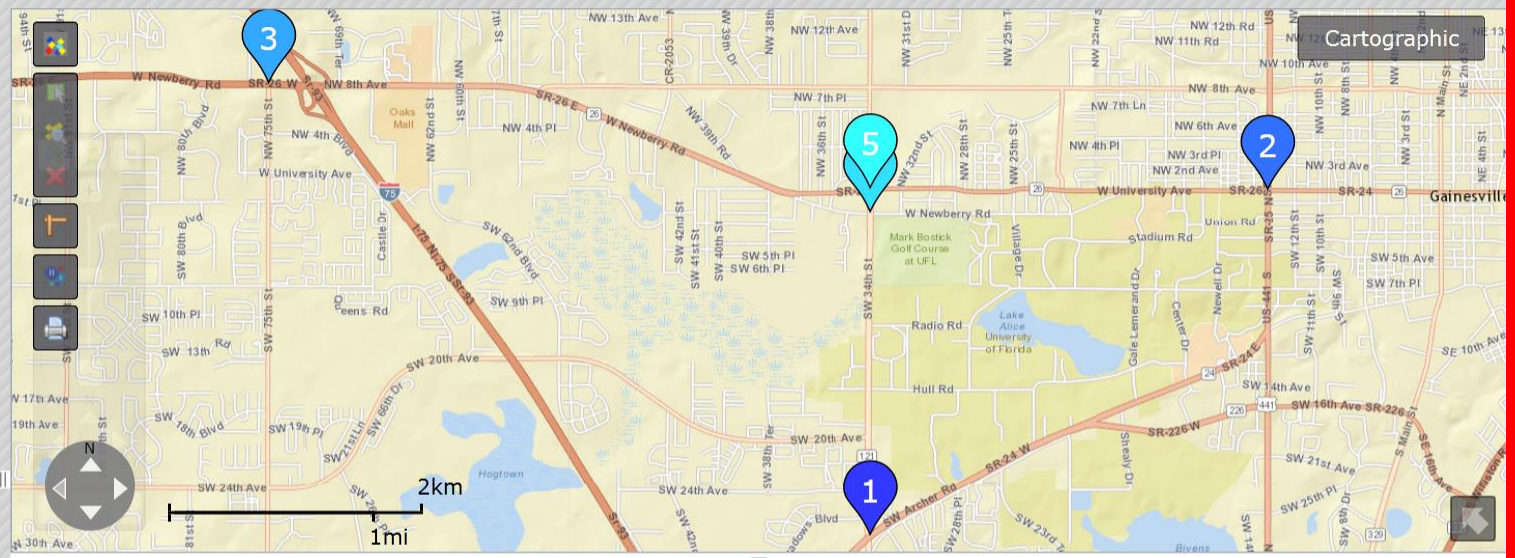
Network Analysis

SIGNAL FOUR ANALYTICS

Signal Four Analytics

https://s4.geoplan.ufl.edu/analytics/#/Views/FeatureAnalysisMainPage.xaml

SIGNAL FOUR ANALYTICS | Crash Analysis | Network Analysis | Submit Feedback | ilir



Date Range

Date Range: 2011 To 2015

Geographic Area

Geographic Extent: • County

County: • Alachua

Network Features

Show: Top 5 | Intersection

Intersection Offset: 250 Feet

Include Ramp Intersections

Ranked By: Crash count

Minimum Crash Count: 10

Roadway Classification

Roadway Classification: • All

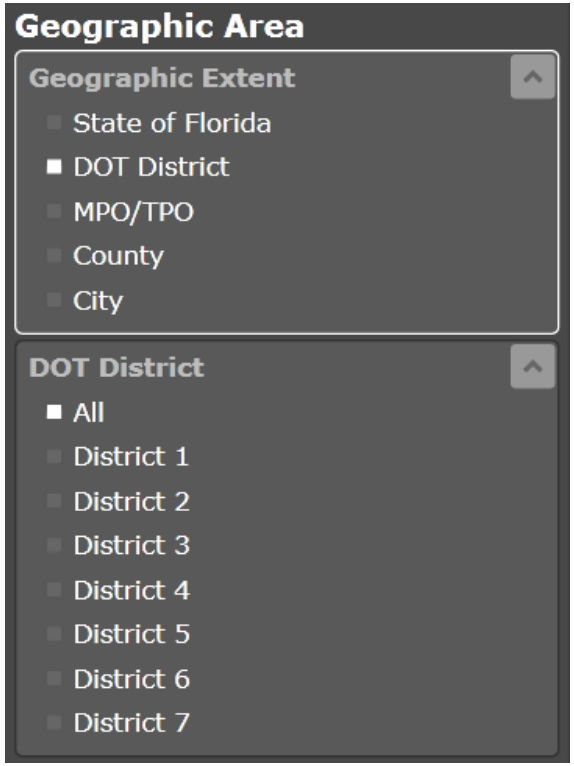
Search | Clear

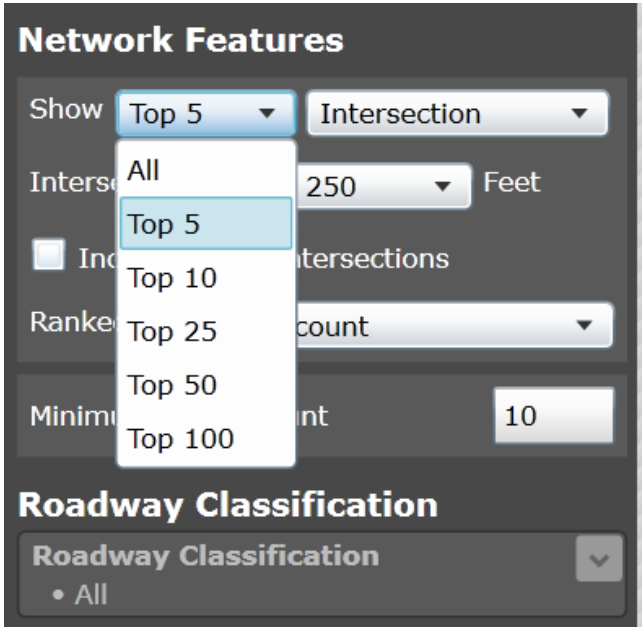
Rank	Intersection Name	Crash Count	Crash Severity	Crash Rate	Fatal Crashes	Fatal & Incapcitating Injury Crashes	Injury
1	SR-121 & SR-24 & SW 34TH ST	253	1.81	n/a	0	6	
2	SW 13TH ST & W UNIVERSITY AVE & NW 13TH ST	218	2.13	n/a	0	5	
3	W NEWBERRY RD & NW 75TH ST	212	1.55	n/a	0	0	
4	SW 34TH ST & W NEWBERRY RD	195	2.11	n/a	0	6	
5	NW 34TH ST & SR-26 & SW 34TH ST & W UNIVERSI	190	1.96	n/a	0	6	

Showing: All(5) / Mapped(5) /Selected(0)

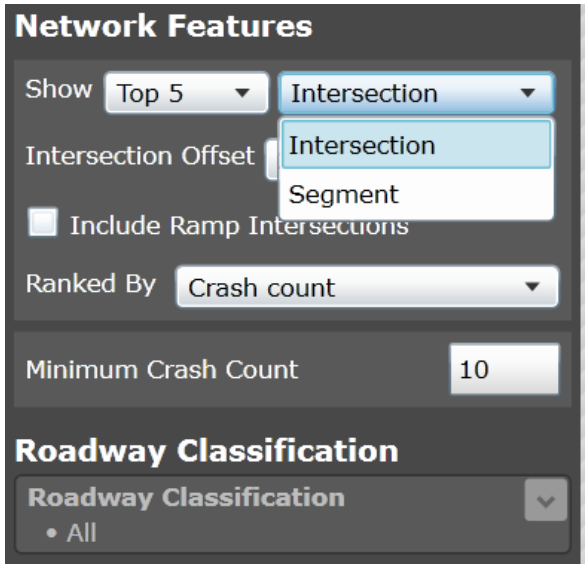
Intersection | Crashes

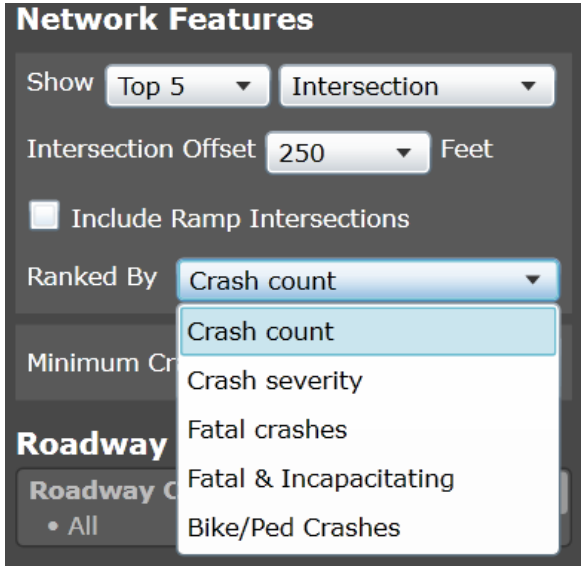
Network Analysis Filters

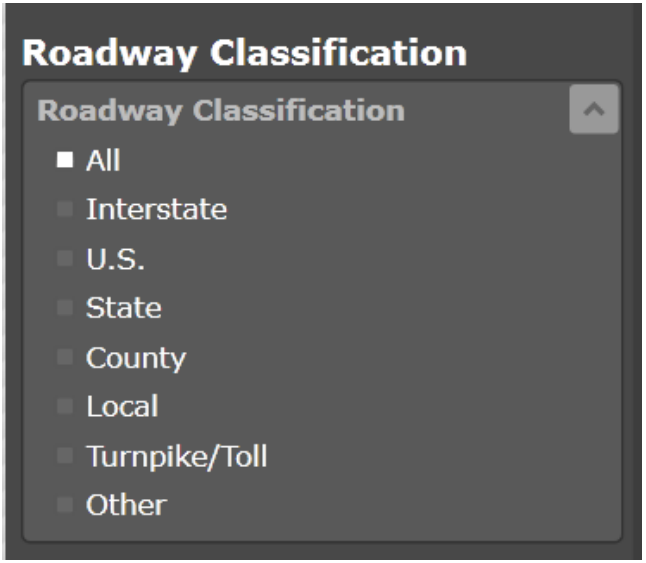




Network Analysis Filters





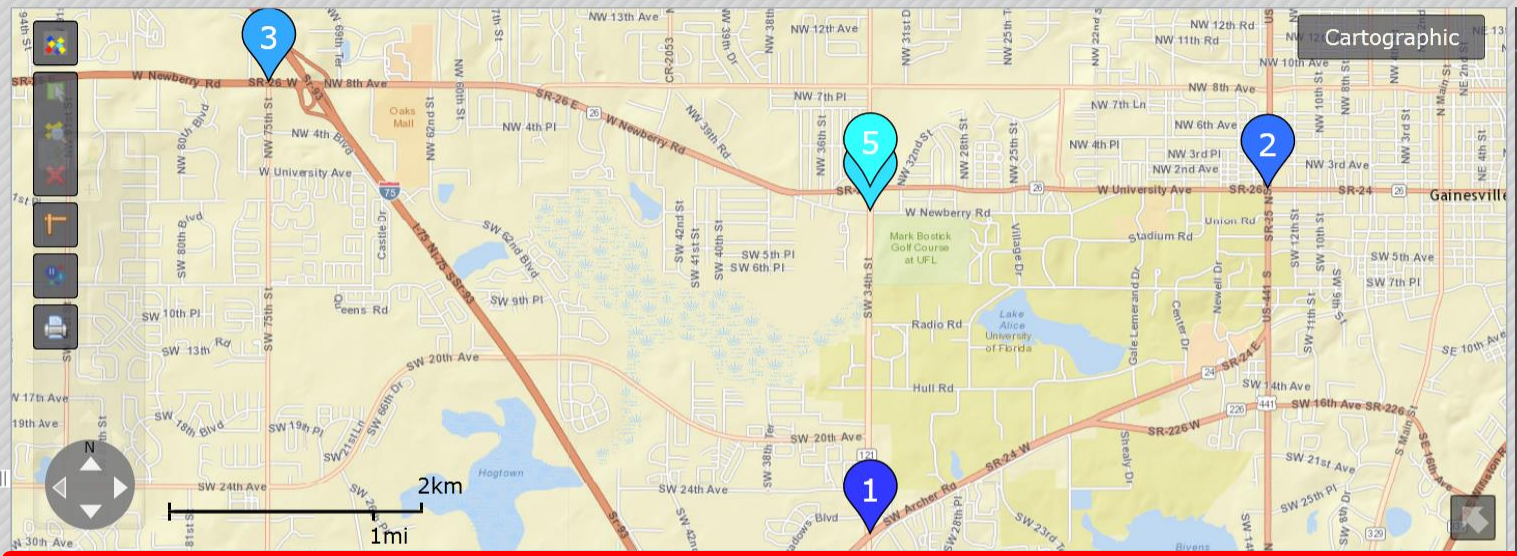


SIGNAL FOUR ANALYTICS

Signal Four Analytics

https://s4.geoplan.ufl.edu/analytics/#/Views/FeatureAnalysisMainPage.xaml

SIGNAL FOUR ANALYTICS | Crash Analysis | Network Analysis | Submit Feedback | ilir



Rank	Intersection Name	Crash Count	Crash Severity	Crash Rate	Fatal Crashes	Fatal & Incapcitating Injury Crashes	Injury
1	SR-121 & SR-24 & SW 34TH ST	253	1.81	n/a	0	6	
2	SW 13TH ST & W UNIVERSITY AVE & NW 13TH ST	218	2.13	n/a	0	5	
3	W NEWBERRY RD & NW 75TH ST	212	1.55	n/a	0	0	
4	SW 34TH ST & W NEWBERRY RD	195	2.11	n/a	0	6	
5	NW 34TH ST & SR-26 & SW 34TH ST & W UNIVERSI	190	1.96	n/a	0	6	

Showing: All(5) / Mapped(5) /Selected(0)

Intersection | Crashes

Date Range

Date Range: 2011 To 2015

Geographic Area

Geographic Extent: • County

County: • Alachua

Network Features

Show: Top 5 Intersection

Intersection Offset: 250 Feet

Include Ramp Intersections

Ranked By: Crash count

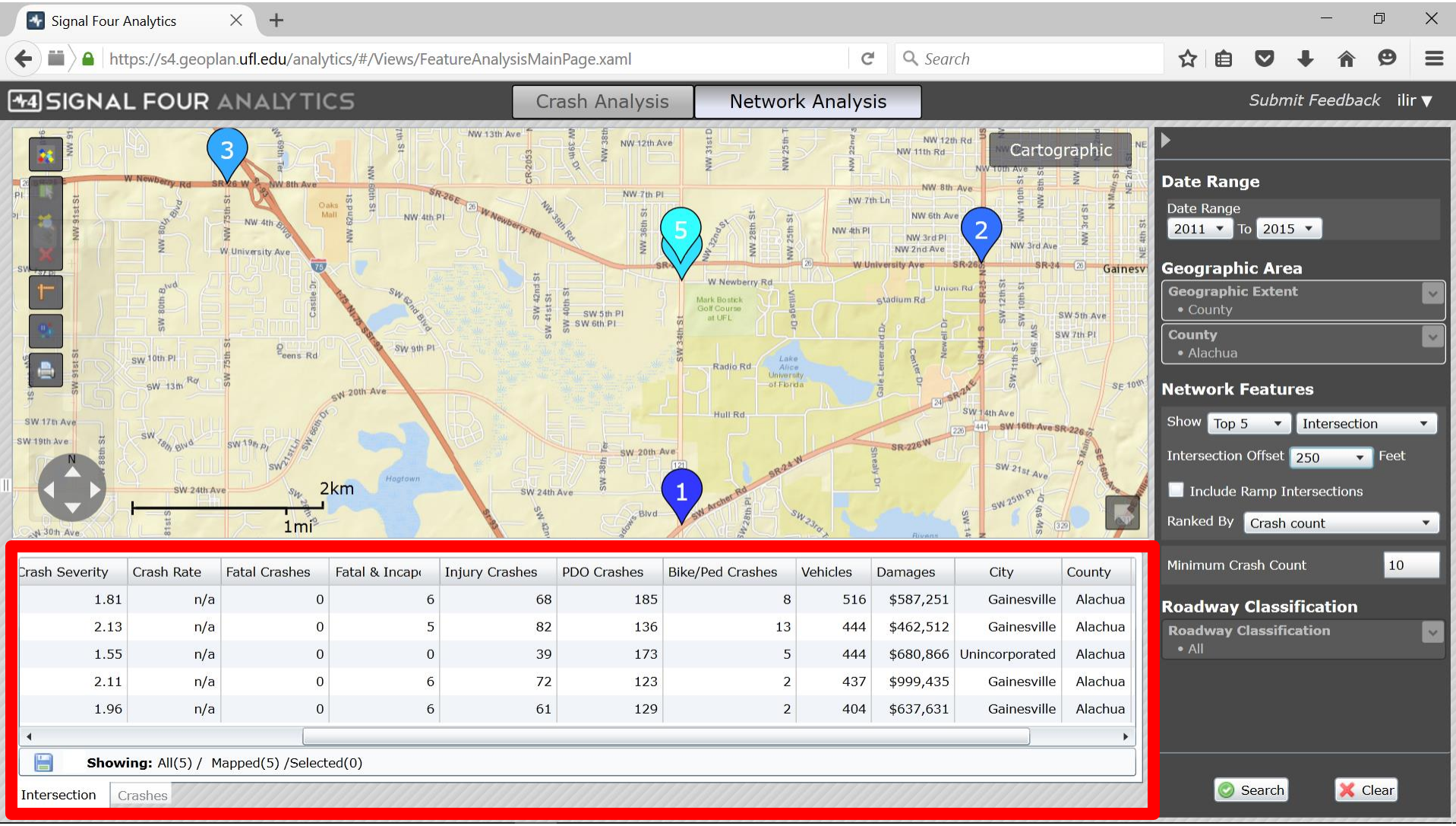
Minimum Crash Count: 10

Roadway Classification

Roadway Classification: • All

Search | Clear

SIGNAL FOUR ANALYTICS



Crash Severity	Crash Rate	Fatal Crashes	Fatal & Incap	Injury Crashes	PDO Crashes	Bike/Ped Crashes	Vehicles	Damages	City	County
1.81	n/a	0	6	68	185	8	516	$587,251	Gainesville	Alachua
2.13	n/a	0	5	82	136	13	444	$462,512	Gainesville	Alachua
1.55	n/a	0	0	39	173	5	444	$680,866	Unincorporated	Alachua
2.11	n/a	0	6	72	123	2	437	$999,435	Gainesville	Alachua
1.96	n/a	0	6	61	129	2	404	$637,631	Gainesville	Alachua

SIGNAL FOUR ANALYTICS

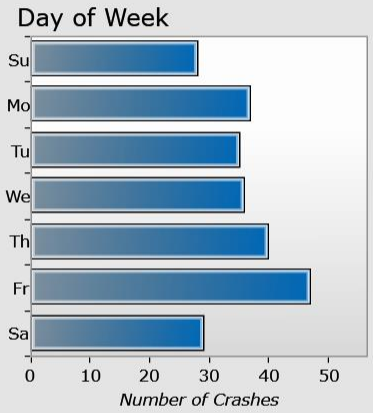
Signal Four Analytics

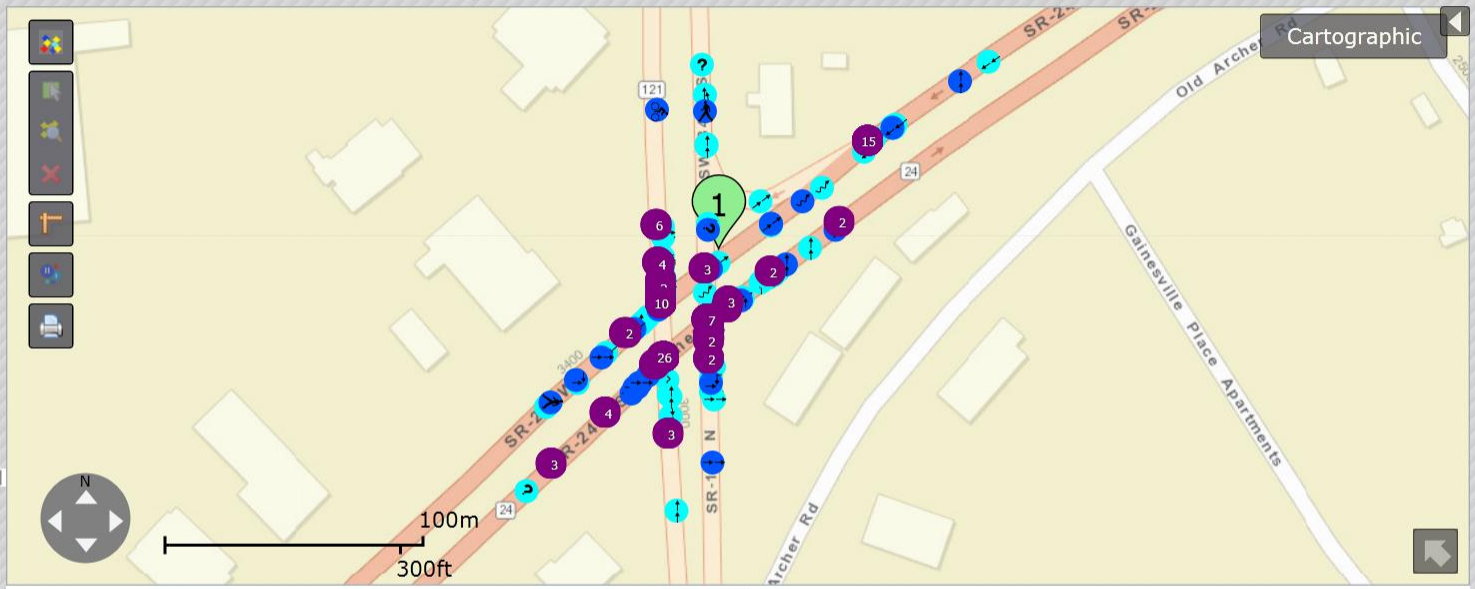
https://s4.geoplan.ufl.edu/analytics/#/Views/FeatureAnalysisMainPage.xaml

SIGNAL FOUR ANALYTICS | Crash Analysis | Network Analysis | Submit Feedback | ilir

Bar Chart | Pie Chart | 2D Chart

Attributes: Day of Week





Intersection 1: SR-121 & SR-24 & SW 34TH ST

HSMV Report #	Agency Report #	Reporting Agency	Form Type	Crash Date	Crash Time	City	County	Crash Street	Intersecting Str
84425366	0214006464	Gainesville PD	Short	4/5/2014	12:58 PM	Gainesville	Alachua	SR-121	SR-24
82122418	0213003032	Gainesville PD	Long	2/13/2013	11:22 AM	Gainesville	Alachua	SW 34TH ST	SW ARCHER RD
82115775	0211016959	Gainesville PD	Long	8/20/2011	1:05 PM	Gainesville	Alachua	SW 34TH ST	SW ARCHER RD

Showing: All(252) / Mapped(252) /Selected(0) **Retrieved 252** (252 Mapped)

Intersection | Crashes

SIGNAL FOUR ANALYTICS

Signal Four Analytics

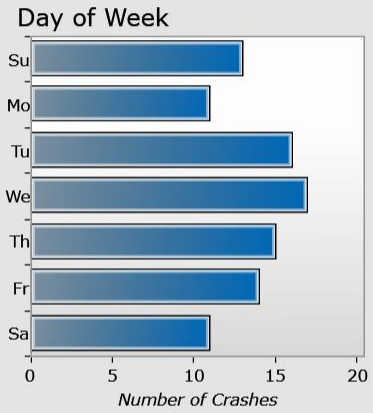
https://s4.geoplan.ufl.edu/analytics/#/Views/FeatureAnalysisMainPage.xaml

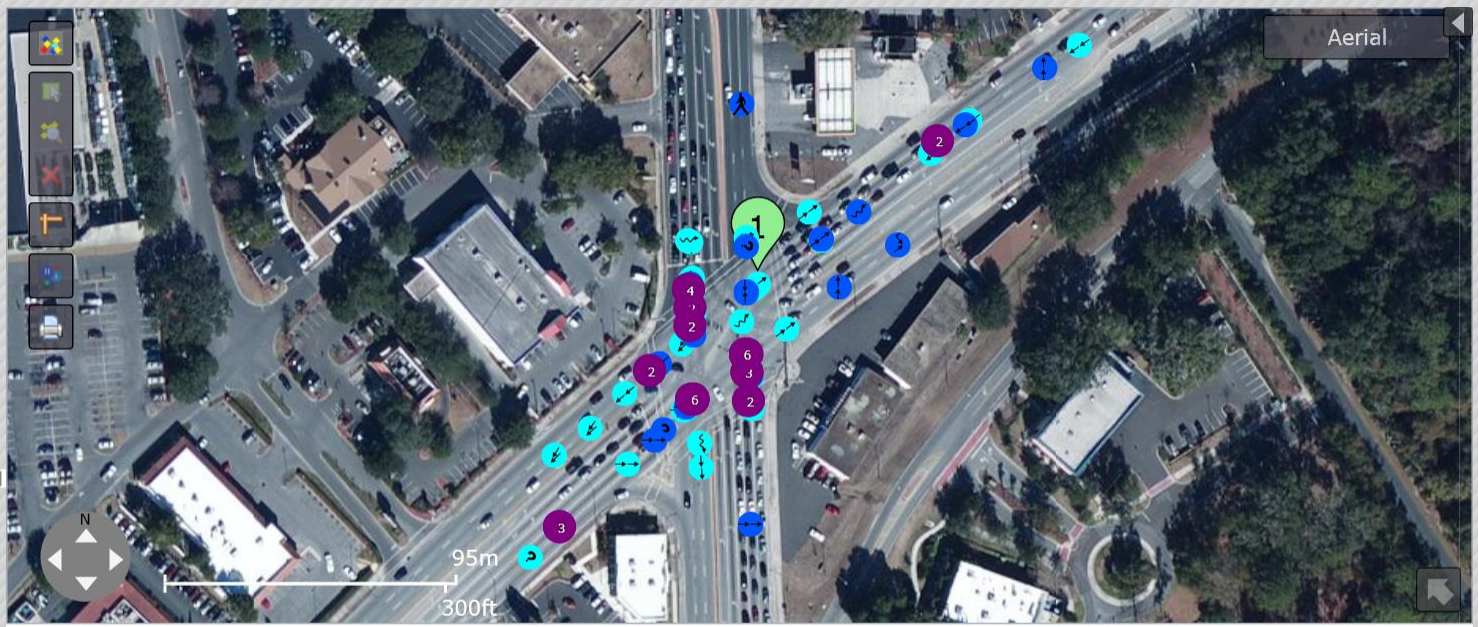
Search

SIGNAL FOUR ANALYTICS | Crash Analysis | Network Analysis | Submit Feedback | ilir

Bar Chart | Pie Chart | 2D Chart

Attributes: Day of Week





Intersection 1: SR-121 & SR-24 & SW 34TH ST

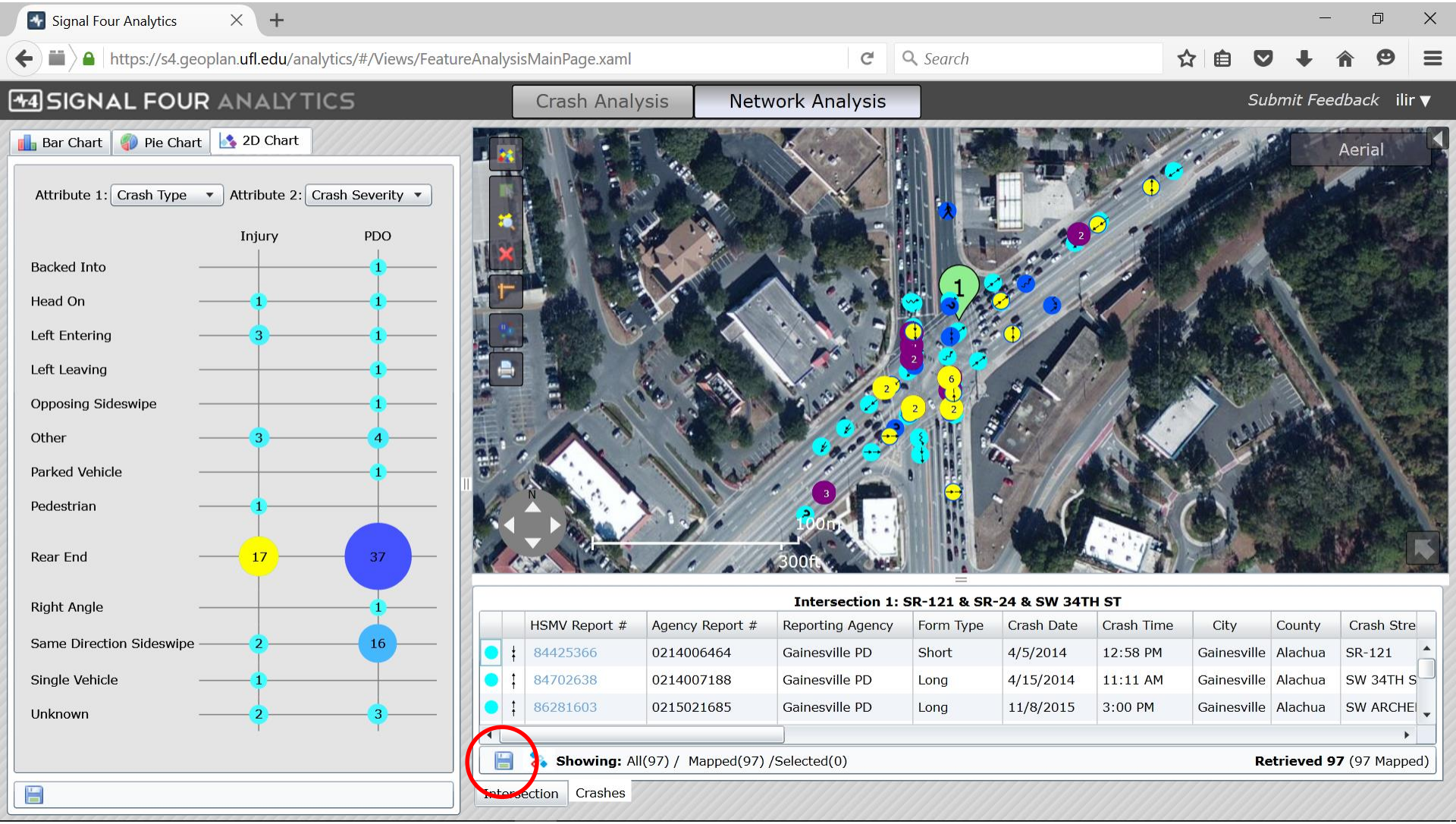
HSMV Report #	Agency Report #	Reporting Agency	Form Type	Crash Date	Crash Time	City	County	Crash Street	Intersecting St
84424537	0214002293	Gainesville PD	Long	2/5/2014	12:18 PM	Gainesville	Alachua	SR-24	SW 34TH ST
84703382	0214010969	Gainesville PD	Long	6/6/2014	3:45 PM	Gainesville	Alachua	SR-24	SW 34TH ST

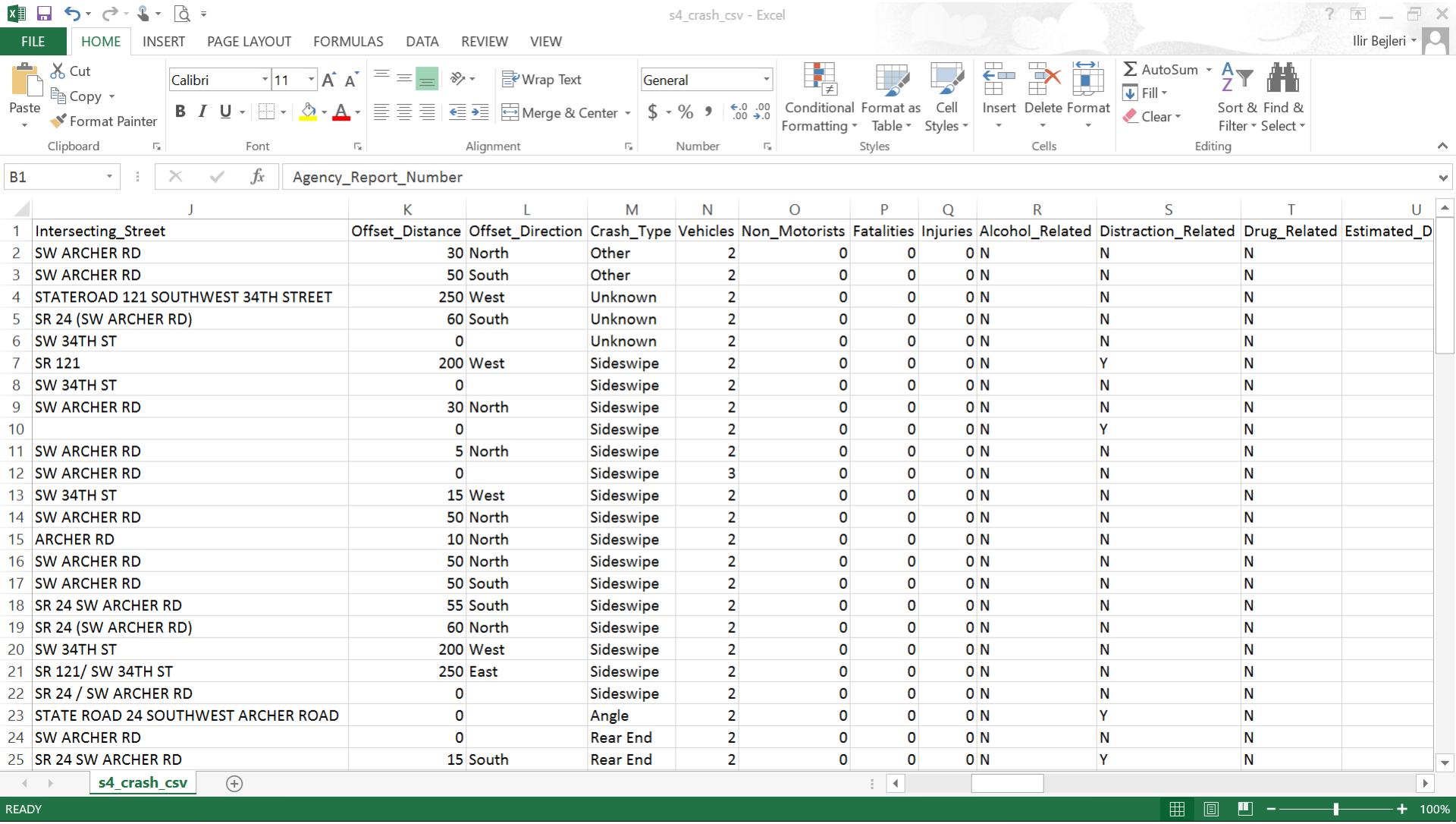
Showing: All(97) / Mapped(97) /Selected(0) **Retrieved 97** (97 Mapped)

Intersection | Crashes

Extract data out of the system

SIGNAL FOUR ANALYTICS





s4_crash_csv - Excel

FILE HOME INSERT PAGE LAYOUT FORMULAS DATA REVIEW VIEW

Ilir Bejleri

Paste Cut Copy Format Painter — Clipboard

Calibri 11 — Font

Wrap Text Merge & Center — Alignment

General — Number

Conditional Formatting Format as Table Cell Styles — Styles

Insert Delete Format — Cells

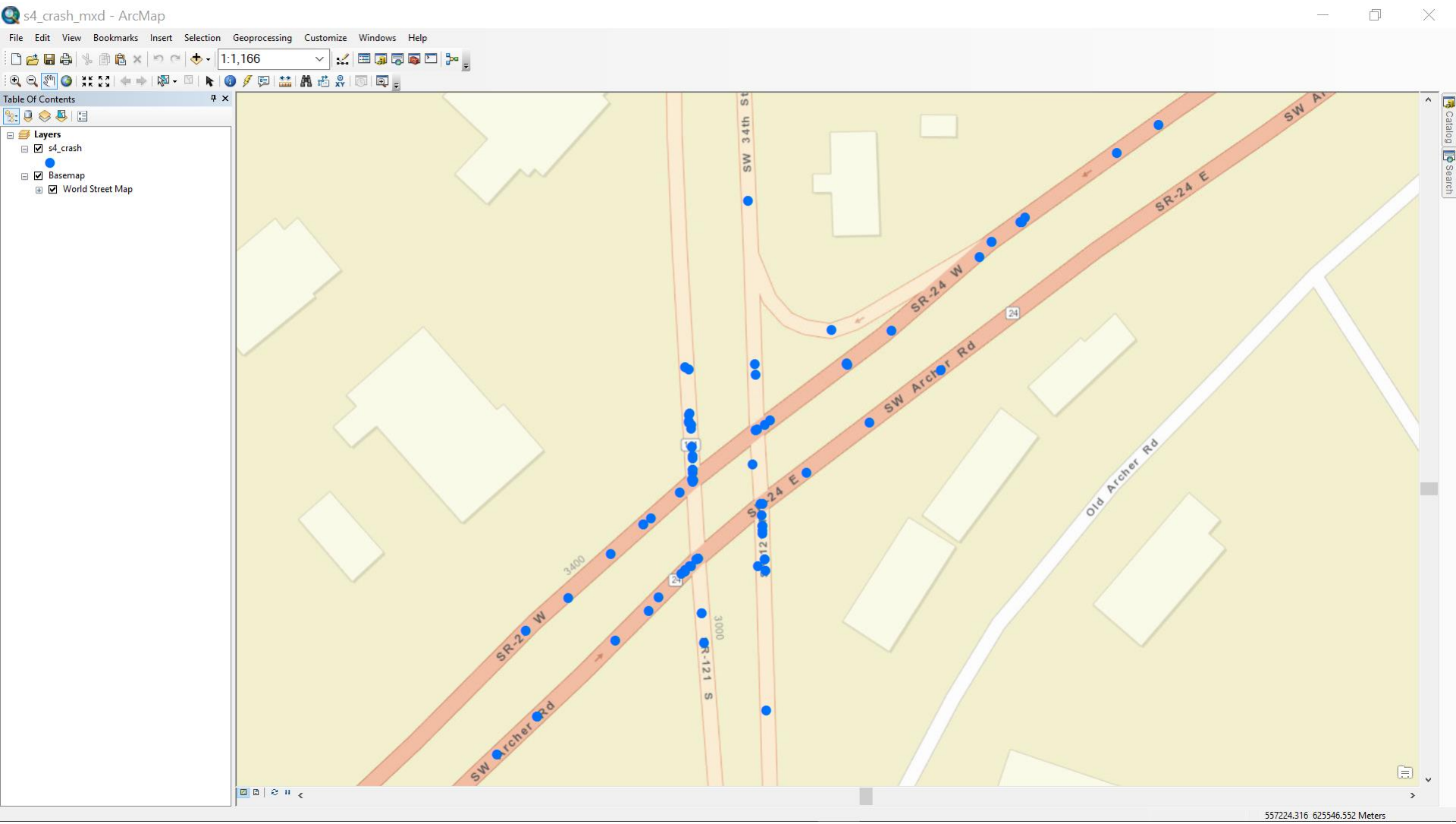
AutoSum Fill Clear Sort & Filter Find & Select — Editing

B1 Agency_Report_Number

	J	K	L	M	N	O	P	Q	R	S	T	U
1	Intersecting_Street	Offset_Distance	Offset_Direction	Crash_Type	Vehicles	Non_Motorists	Fatalities	Injuries	Alcohol_Related	Distraction_Related	Drug_Related	Estimated_D
2	SW ARCHER RD	30	North	Other	2	0	0	0	N	N	N	
3	SW ARCHER RD	50	South	Other	2	0	0	0	N	N	N	
4	STATEROAD 121 SOUTHWEST 34TH STREET	250	West	Unknown	2	0	0	0	N	N	N	
5	SR 24 (SW ARCHER RD)	60	South	Unknown	2	0	0	0	N	N	N	
6	SW 34TH ST	0		Unknown	2	0	0	0	N	N	N	
7	SR 121	200	West	Sideswipe	2	0	0	0	N	Y	N	
8	SW 34TH ST	0		Sideswipe	2	0	0	0	N	N	N	
9	SW ARCHER RD	30	North	Sideswipe	2	0	0	0	N	N	N	
10		0		Sideswipe	2	0	0	0	N	Y	N	
11	SW ARCHER RD	5	North	Sideswipe	2	0	0	0	N	N	N	
12	SW ARCHER RD	0		Sideswipe	3	0	0	0	N	N	N	
13	SW 34TH ST	15	West	Sideswipe	2	0	0	0	N	N	N	
14	SW ARCHER RD	50	North	Sideswipe	2	0	0	0	N	N	N	
15	ARCHER RD	10	North	Sideswipe	2	0	0	0	N	N	N	
16	SW ARCHER RD	50	North	Sideswipe	2	0	0	0	N	N	N	
17	SW ARCHER RD	50	South	Sideswipe	2	0	0	0	N	N	N	
18	SR 24 SW ARCHER RD	55	South	Sideswipe	2	0	0	0	N	N	N	
19	SR 24 (SW ARCHER RD)	60	North	Sideswipe	2	0	0	0	N	N	N	
20	SW 34TH ST	200	West	Sideswipe	2	0	0	0	N	N	N	
21	SR 121/ SW 34TH ST	250	East	Sideswipe	2	0	0	0	N	N	N	
22	SR 24 / SW ARCHER RD	0		Sideswipe	2	0	0	0	N	N	N	
23	STATE ROAD 24 SOUTHWEST ARCHER ROAD	0		Angle	2	0	0	0	N	Y	N	
24	SW ARCHER RD	0		Rear End	2	0	0	0	N	N	N	
25	SR 24 SW ARCHER RD	15	South	Rear End	2	0	0	0	N	Y	N	

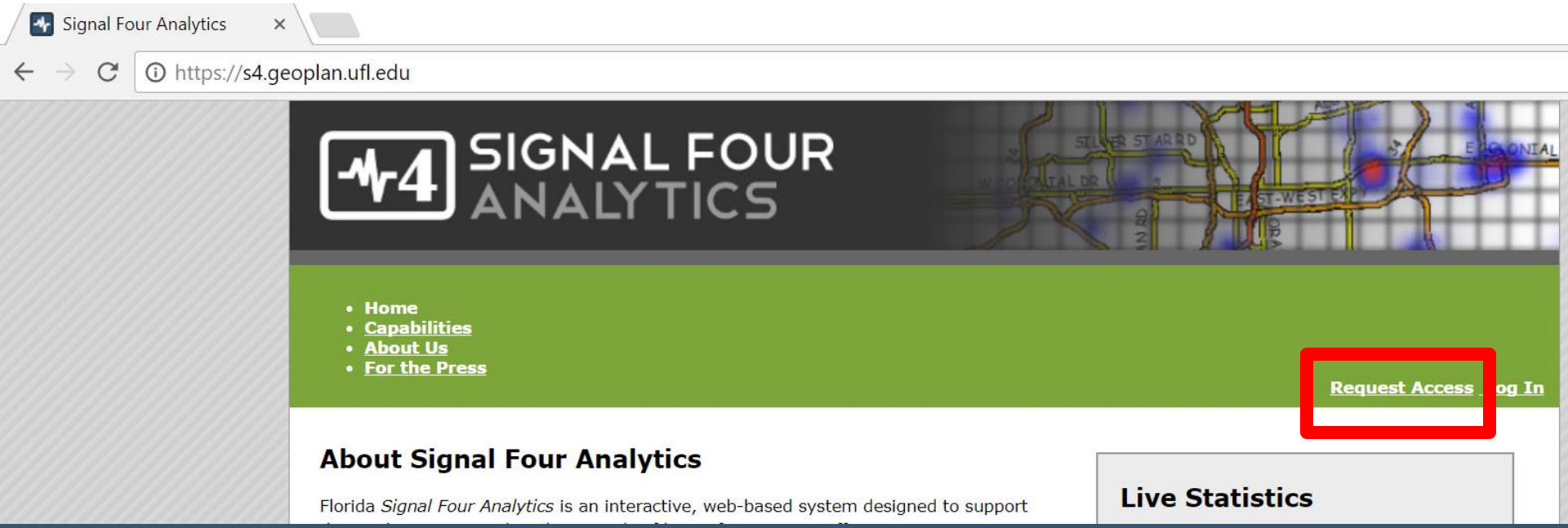
s4_crash_csv

READY 100%


s4_crash_mxd - ArcMap
File Edit View Bookmarks Insert Selection Geoprocessing Customize Windows Help
1:1,166
Table Of Contents
Layers
s4_crash
Basemap
World Street Map
SW 34th St
SR-24 W
SR-24 E
SW Archer Rd
Old Archer Rd
3400
3000
557224.316 625546.552 Meters

To Get Access:

- **s4.geoplan.ufl.edu** -> Request Access
- Contact **ilir@ufl.edu**



Next:

- GIS Visualization
- **Big Data GIS Performance**
- Maps in Data Visualization Platforms

Performance Evaluation of Big-Data GIS systems

- Increase in use of networked, location aware, pervasive computing devices for planning and monitoring the transportation infrastructure has resulted in a huge increase in the size of location-aware datasets.

Bluetooth enabled RSU

- New solutions for large-scale processing data try to exploit scalability provided by parallel, distributed computing systems.
- We try to evaluate the effectiveness of these distributed, parallel systems for spatial queries.
- There are two ways in which the performance of traditional GIS systems can be improved.
  - Traditional GIS systems are single threaded, tasks cannot be parallelized.
  - Cannot run on distributed systems.

Methodology

- Use both real and simulated data sets and a set of queries typical in the transportation industry on the following tools:
  - **Arc GIS**
  - **Magellan**: Open source platform, built on top of Apache Spark.
  - **Spatial-Spark**
  - **GIS Tools for Hadoop** -by ESRI

Machine Name	CPU Cores	Memory (GB)	I.P. Address	Node Number
freenode-X9DRG-HF	32	132	128.227.170.246	Node 1
spark1	2	4	128.227.176.142	Node 2
firestone	4	16	128.227.176.188	Node 3

Software	Version
Operation System	Ubuntu 16.10 (master)
Java environment	JDK version: 1.7.0_91
Hadoop	Hadoop 2.6.0
Spark	Spark1.6.1
Scala	Scala2.11.7
Magellan	1.0.3-s_2.10
Spatio-Spark	2.10
GIS Tools for Hadoop	2.16.0

Sample queries, dataset and corresponding primitives

Queries

1. Query which county or land use zone each crash lies in.
2. Query all the bus stops within 100 meters of a crash.

Dataset

- Point data: Crash Data & Simulated Point data
  - Crash data: Set of 120 thousand points. Represent crashes in Orange County, Florida.
  - Simulated points data: 500 thousand, 5 million, 50 million and 100 million random points
- Polygon Data: Land use (50 thousand polygons) & Parcels data (500 thousand polygons) for Orange County, Florida
- Other data: NYC Taxi Cab data composed of recorded of Taxi trips in the city from January 2009 to June 2015. It has a total of 1.1 billion records. Used for binning.

GIS Primitives

Query 1

- It involves one primitive operation: "contains"
- Given a point and a polygon, it answers whether the point lies within that polygon.

Query 2

- 2 primitives are needed.
- "buffer": It draws a 100 meters buffer around a given point (crash).
- "contain": answer which points, lines or polygons (Bus stops, Bus routes) lie inside the buffer (polygon). Similar to query 1

Completion time comparison for small datasets

Query which county each crash lies in (Query 1)

Input	Spatial Spark	Magellan	ArcGIS
18MB (120K Points)	144 s	3.294 s	26.68 s
37MB	282 s	6.464 s	52.72 s
52MB	396 s	9.740 s	78 s
73MB	540 s	13.090 s	104 s
110MB	840 s	20.133 s	157 s
146MB	1080 s	26.049 s	209 s

Query all bus stops within 100 meters of a Crash (Query 2)

Input	Spatial Spark	Magellan	ArcGIS
18MB	780 s	NA	138 s
37MB	1,500 s	NA	366 s
52MB	1,920 s	NA	444 s
73MB	2,820 s	NA	648 s
110MB	4,320 s	NA	1,014 s

Completion Time for Big Datasets (ArcGIS vs GIS Tools for Hadoop)

Input	Operation	GIS Tools for Hadoop	ArcGIS
500 Thousand Random Points	Buffer	61 s	49 s
5 million Random Points	Buffer	441 s	518 s
50 million Random Points	Buffer	3,658 s	4,920 s
100 million Random Points	Buffer	7,296 s	13, 218 s
NYC Trip Data (14 Million Points)	Binning	60.45 s	NA
NYC Trip Data (30 Million Points)	Binning	170 s	NA

Note: Query 2 → Buffer + Contains.

Possible Improvements

- Primitive 'Contains' *(Query 1 → Contains, Query 2 → Buffer + Contains).*
- GIS Tools for Hadoop, matches every given point against every polygon.

Input 1	Input 2	Operation	GIS Tools for Hadoop	GIS tools for Hadoop - Custom	ArcGIS
500 Thousand Random Points	Landuse	Contains	15,000 s	3,700 s	250 s
500 Thousand Random Points	Parcels	Contains	47,000 s	15,000 s	700 s

- Custom code on Hadoop to only try and match points to polygons when they are spatially "near". ~4x gain.

Discussion

- The performance of the systems evaluated can be significantly better (than traditional systems) for larger data sets. Both *GIS Tools for Hadoop* and *Magellan* outperform *ArcGIS* in some operations.
- But only limited support for spatial operations (*Magellan*) and/or incomplete implementation (*GIS tools for Hadoop*).
- In the future, parallel and distributed computing techniques can lead to significant gains. Both in terms of volume and the speed of Geo-spatial data processing.

Next:

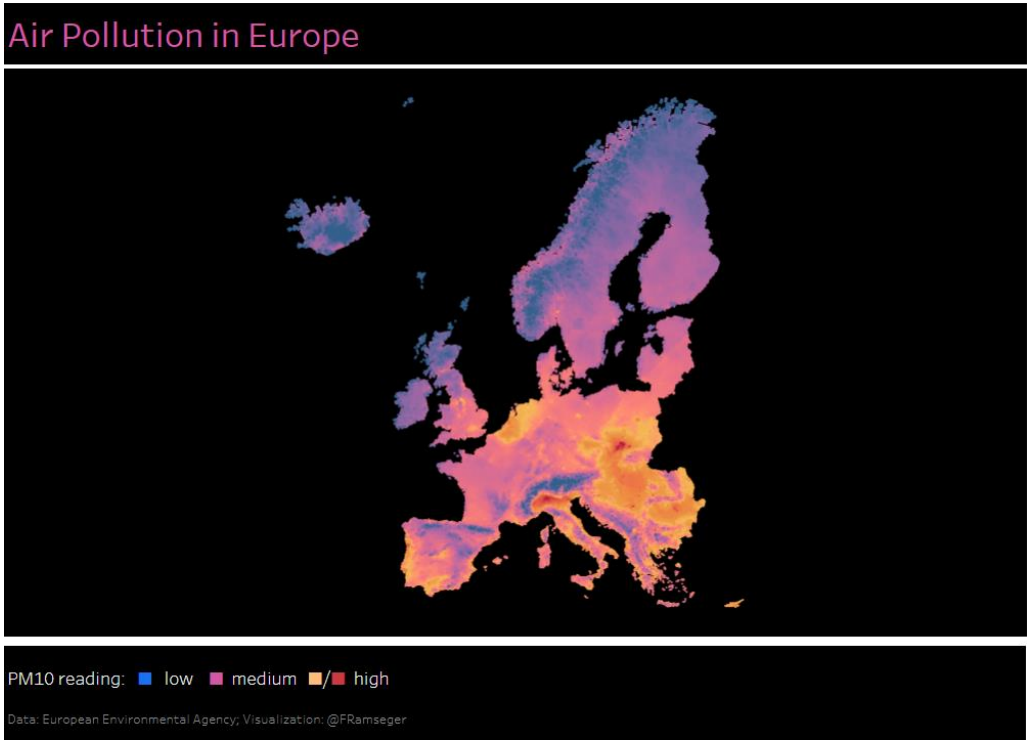
- Web GIS Visualization Platforms
- Big Data GIS Performance
- **Use of Maps in Data Visualization**

Tableau based spatial data visualization

Tableau

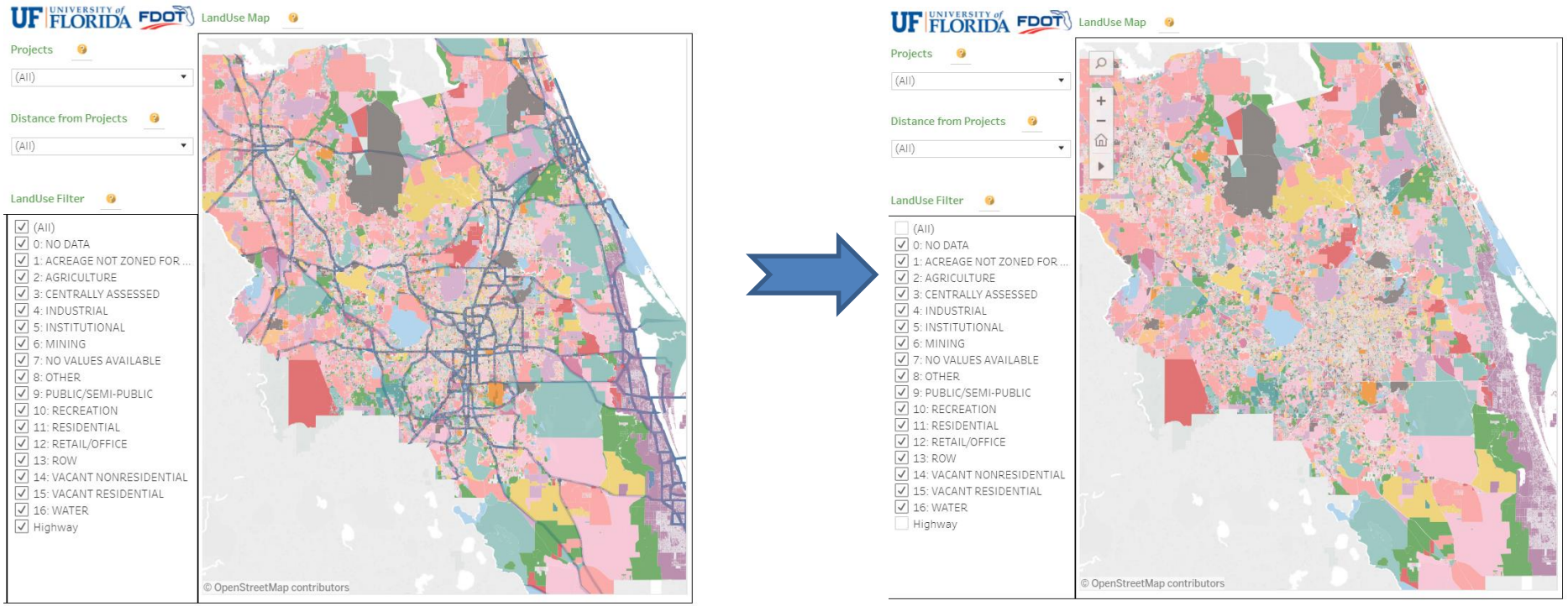
- A suite of software (desktop, server & cloud) that allows users to explore and visualize their data.
- Recently (Q1 2017) started adding support for spatial maps and files.
- Specifically: GeoJSON, Map-Info and .shp.
- But no support for spatial queries is natively built in.

Static maps



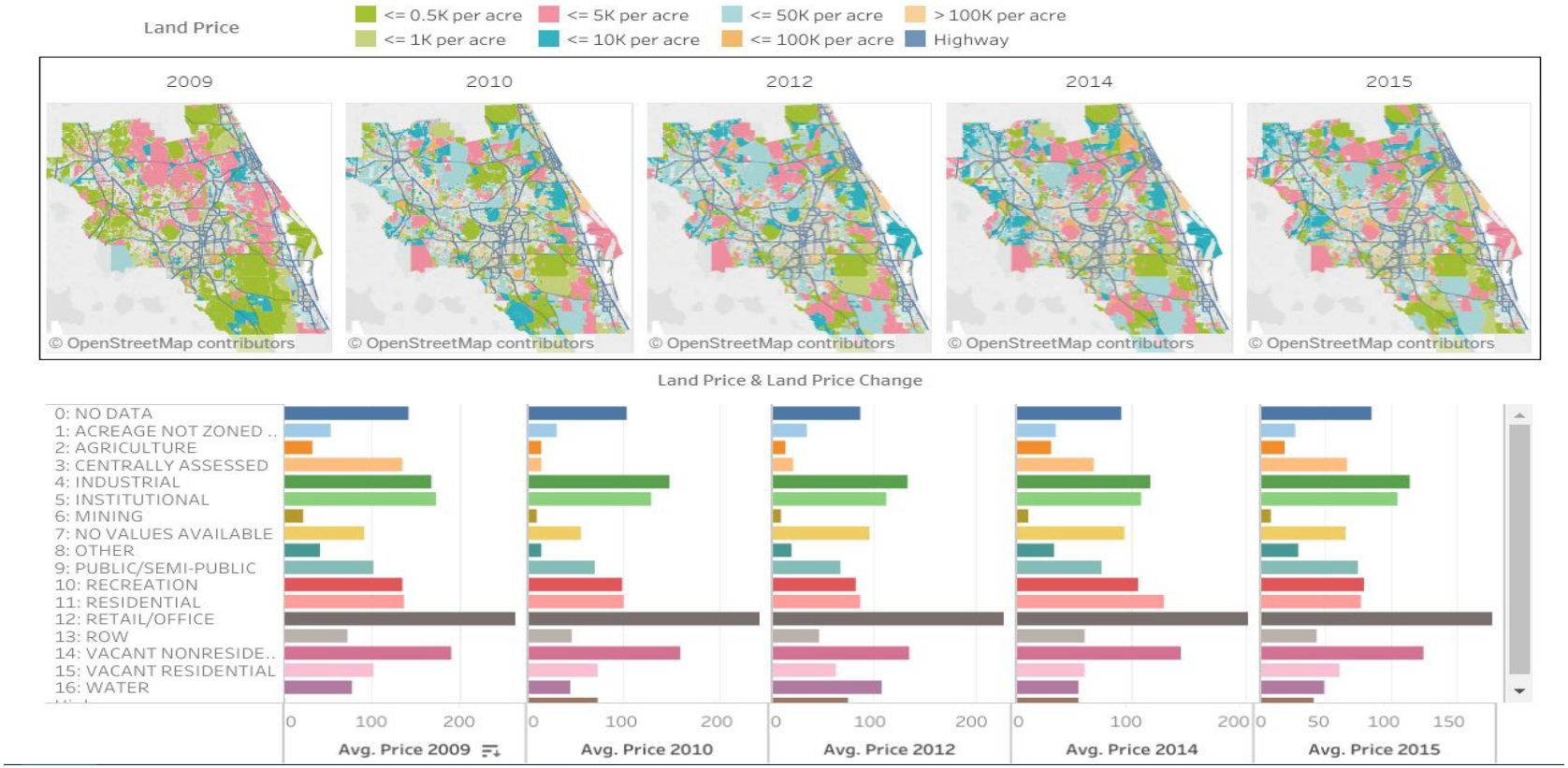
- From tableau.com

Interactive maps



Land use parcels for D5, Florida (Courtesy FDOT)

Temporal trends with maps



Conclusions

- No support for Lines, Multi lines
- No support for joins, can’t overlay sets of geometries in a single map.
- The final dataset (Shapefile/Map Info) prepared after preprocessing should be a single flat file.
  - i.e. If different sets of shapes or geometries are to displayed, they should be merged into a single file.

Questions/Comments ?