

## Paper RI-14

## Creating a Heatmap Visualization of 150 Million GPS Points on Roadway Maps via SAS®

Shih-Ching Wu, Virginia Tech Transportation Institute, Blacksburg, Virginia  
Shane McLaughlin, Virginia Tech Transportation Institute, Blacksburg, Virginia

### ABSTRACT

This paper introduces an approach to using SAS integrated with common geospatial tools to reduce a large amount of GPS points to heatmaps describing the frequency with which road segments are traversed. To accomplish this task, there are three steps. First, matching the large number of GPS points to known road networks which is known as a map-matching process. Second, calculating the number of trips on each road segment based on previous map-matching results. Third, coloring road segments based on the number of trips. An example is described which maps 150 million GPS points using SAS with SAS® Bridge for ESRI, ArcGIS, and two free and open-source tools - PostgreSQL and PostGIS.

### INTRODUCTION

The Virginia Tech Transportation Institute (VTTI) has been conducting several multi-terabyte collections of naturalistic driving study (NDS) data. These studies collect continuous data from vehicles as they are driven by research participants in their daily activities over a year or two years. The variety of information and large amount of data have become an important source of guidance for transportation research. Among collected sensor data, global positioning system (GPS) is a crucial data source to link naturalistic driving data to geographic elements, such as roadway segments or spatial boundaries. Conducting this association of elements is the essential step for location-based research and visualization. GPS data provides point for linking the NDS vehicle/driver measures and the broad range of roadway and geographic attributes available in geographic databases. However, associating large amount of GPS data to roadway maps requires lots of computational resources to execute complex map matching algorithms. Graphical user interface (GUI) based geographic information system (GIS) software packages lack of flexibility to implement user-developed map-matching algorithms and capability to handle large quantities of GPS data. This makes the mapping process inaccurate or infeasible within realistic time frames. This paper introduces an approach to using SAS integrated with common geospatial tools to map GPS data onto roadways and reduce a large amount of GPS points to heatmaps describing the frequency with which road segments are traversed. The mapping process described here for 150 million GPS points takes 156 computer hours to complete, which is more than two times faster than comparable methods executed within a GUI application. In the following sections, we briefly introduce PostgreSQL and PostGIS, demonstrate the implementation of a map-matching algorithm, then calculate the number of trips on each road segment based on previous map-matching results. Finally, we create a heatmap visualization and describe other benefits of this SAS integrated approach.

### OVERVIEW OF POSTGRESQL AND POSTGIS

PostgreSQL [1] is a database management system. PostGIS [2] is a library of spatial database functions, extending the utility of the PostgreSQL database. It adds to PostgreSQL spatial data types which allow PostgreSQL to be used as a spatial database. In addition, PostGIS provides functions for working with the spatial data types which allows users to create location-aware queries with SQL scripts in the SAS SQL procedure. By using spatial SQL, much of the heavy lifting that would require many manual steps in GUI-based GIS software can be scripted and automated via SAS. Following is an example of using PostGIS functions via SAS.

```
proc sql;
  connect to odbc (dsn = "PostgreSQL_Guidance"); * (A);
  title 'Point Geometry';
  select * from connection to odbc
  (select ST_SetSRID(ST_Point(-78.1391296, 42.7402077),4326) as Geometry); * (B);

  title 'Created Point Geometry';
  select * from connection to odbc
  (select ST_AsEWKT('0101000020E610000F579D67FE78853C0AADF3B20BF5E4540')
  as GPS_Point); * (C);
  disconnect from odbc;
quit;
```

In the above example, Part A lets SAS access the PostgreSQL database via an ODBC connection [3]. In Part B, the PostGIS function, ST\_Point, creates point geometry for a GPS point and the ST\_SetSRID function specifies the spatial reference system used by the GPS receiver which was used for collecting this data. The result is shown in the upper portion of Figure 1. In Part C above, the function, ST\_AsEWKT, returns the created geometry representation in Well-Known-Text (WKT) format with a spatial reference system identifier (SRID) metadata. The result is shown in the bottom part of Figure 1 which matches the latitude and longitude coordinates input in Part B of the SAS code.

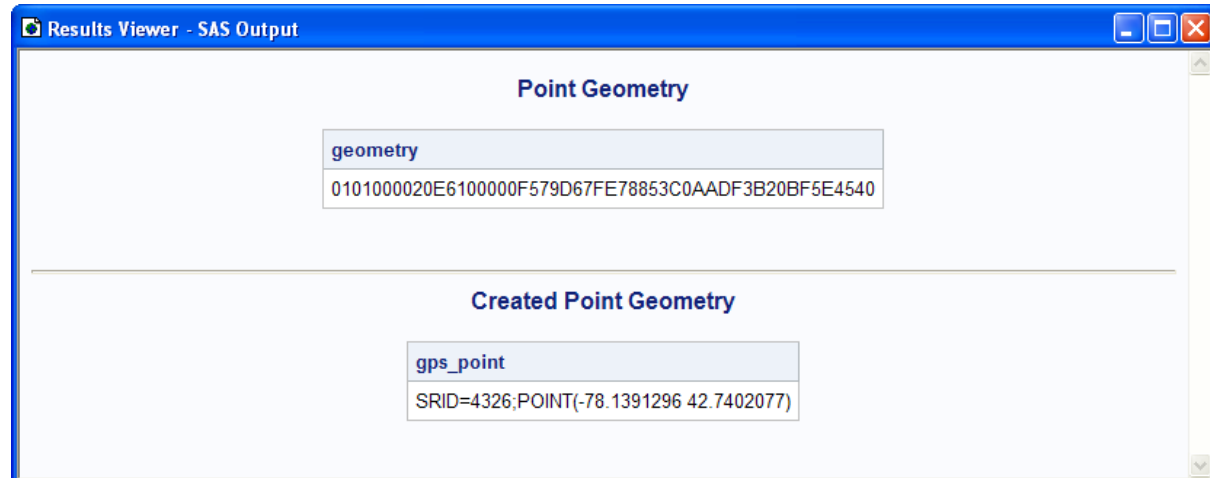


Figure 1. SAS PROC SQL Output

## PREPARATION OF DATA SETS

In order to demonstrate implementation of a map-matching algorithm, two data sets need to be introduced. First is the GPS data from the SHRP 2 NDS [4]. Figure 2 is an example of partial GPS data which contains time-stamped latitude, longitude and heading values from a trip which is assigned by a unique "file\_id." Later in the next section, these five GPS points are used for demonstration of map-matching techniques. Second is roadway data. Figure 3 is roadway data of New York state [5]. The attribute "objectid" is the unique identifier for each road segment and another attribute "shape" describes linestring geometry of each road segment. These two data sets are stored into VTTI's PostgreSQL database and later the two data sets will be referred by the table names "shrp\_2\_gps\_data" and "roads\_ny\_simplifiedstreetsegmentann\_wgs84." Figure 4 overlays the two data sets together. Arrows in Figure 4 indicate the direction of travel from one GPS point to the next using the "heading\_gps" value. At this point in the process (shown in Figure 4), no relationship between the two data sets has been established, namely we have no knowledge regarding on which road segments these five GPS points were likely collected.

file_id	timestamp	latitude	longitude	heading_gps
3827593	3683134	42.74020004	-78.13957214	89.209999084473
3827593	3684134	42.74020004	-78.13935852	88.019996643066
3827593	3685135	42.74020767	-78.13912964	89.040000915527
3827593	3686135	42.74020004	-78.13890839	90.680000305176
3827593	3687135	42.74021149	-78.13869476	90.480003356934

Figure 2. GPS Data

objectid	fullstname	shield	hwy_num	fcc	acc	jurisdict	county_nam	mcd_name	shape
8697	Airport Rd	CT	69	A35	4	02	Broome	Dickinson	0105000020E61000000600000001020000000B000000303E44B75FFC52C0B83482E
8699	Alden St	S	166	A30	3	01	Otsego	Cherry Valley	0105000020E610000002000000010200000003000000E4E3A1EF6EB052C0F049FFE
8716	Alder Bend Rd	C	12	A41	5	02	Clinton	Altona	0105000020E61000000A000000010200000003000000B857B7BE2B6D52C098F82B
8717	Alder Bend Rd	C	12	A41	5	02	Clinton	Mooers	0105000020E610000002000000010200000003000000845CC4052D6D52C068C3F4
8718	Alder Bottom Rd	CT	4	A41	5	02	Chautauqua	Sherman	0102000020E61000000220000004CF60AC991E553C060234B7E350D454088A3926A
8728	Alder Creek Rd	CT	54	A41	5	02	Ulster	Hardenburgh	0105000020E610000003000000010200000047000000788AB8400FAD52C040D01FA
9029	8th St	S	32	A21	3	01	Albany	Watervliet	0102000020E610000004000000C4DA6ACC076D52C090EABF64A95B454060FC6BE
9396	9th St	S	32	A21	3	01	Saratoga	Waterford	0102000020E61000000D00000028195B63EB6B52C010C256223365454068967AFF
9597	Abbey Rd	C	212	A41	5	02	Onondaga	Onondaga	0105000020E61000000200000001020000001A00000060DBC5DBDA0D53C07048E2
9629	Abbott Hill Rd	C	426	A41	5	02	Erie	Concord	0105000020E610000003000000010200000013000000BC6DB76EF4AC53C0A06350
9648	Abbott Rd	C	4	A31	4	02	Erie	Hamburg	0105000020E610000005000000010200000002000000905952280AB353C0901012F
9649	Abbott Rd	C	4	A31	4	02	Erie	Orchard Park	0105000020E61000001600000001020000000200000084D680C493B253C0E8DCB4
9650	Abbott Rd	C	4	A31	4	02	Erie	Lackawanna	0105000020E610000010000000010200000002000000EC5E62EA76B353C0E06BD29

Figure 3. Roadway Map Data

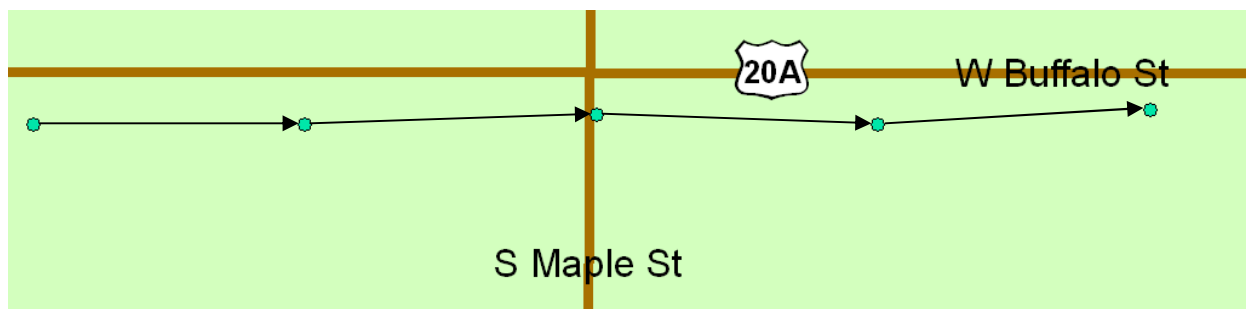


Figure 4. GPS Data on Roadway Map

## MAP MATCHING

Map matching is a process of associating GPS traces on road segments. A simple and intuitive way is mapping GPS data to a nearby road. The following SAS sample code illustrates how to search for the closest road segment for one GPS point (the first point shown in Figure 2).

```
proc sql;
  connect to odbc (dsn = "PostgreSQL_Guidance");
  title 'Map Matching Result (Nearest Neighbor)';
  create table mm_result as
  select * from connection to odbc
  (select gps.file_id, gps.timestamp, gps.latitude, gps.longitude,
  gps.heading_gps,
  road.objectid, road.fullstname, road.shield, road.hwy_num, road.fcc, road.acc,
  road.jurisdiction, road.county_name, road.mcd_name
  from scwu.shrp_2_gps_data as gps,
  scwu.roads_ny_simplifiedstreetsegmentann_wgs84 as road
  where file_id = 3827593 and timestamp = 3683134 /* (D) */
  order by ST_DISTANCE(road.shape, ST_SetSRID(ST_Point(gps.longitude,
  gps.latitude),4326)) /* (E) */
  limit 1);
  disconnect from odbc;
quit;
```

In Part D, the WHERE statement selects one GPS point from the sample data (Figure 2). Part E, the PostGIS ST\_DISTANCE function calculates distance between two geometries which are the collected GPS point and New York roadway map. Since results are ordered by distance, the first record is always the nearest road segment to GPS data. Results of applying the same map-matching algorithm to all five GPS points are shown in the following figure.

file_id	timestamp	latitude	longitude	heading_gps	objectid	fullstname	shield	hwy_num	fcc	acc	jurisdiction	county_name	mcd_name
3827593	3683134	42.7402	-78.1396	89.209999	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw
3827593	3684134	42.7402	-78.1394	88.019997	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw
3827593	3685135	42.74021	-78.1391	89.040001	276725	S Maple St			A41	5	13	Wyoming	Warsaw
3827593	3686135	42.7402	-78.1389	90.680000	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw
3827593	3687135	42.74021	-78.1387	90.480003	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw

Figure 5. Map-Matching Results

For clarity, the above code conducts the mapping for one GPS point. In practice, this code uses SAS Macro to iterate through as many GPS points as are in the data set.

From Figure 4, these five GPS points actually travel on W Buffalo Street and, from Figure 2, GPS heading values (~90 deg) also indicate the vehicle moves toward the East. However, the map-matching result has one GPS point mapped on S Maple Street. Using shortest distance or nearest neighbor as mapping criteria might be inaccurate

when there are multiple roads near a GPS point such as at intersections. This mistaken classification can be fixed if GPS heading values are considered. In this case, PostGIS and SAS come in handy to implement algorithm improvements with user-developed techniques. The following SAS sample code is an extension from the previous SAS sample code. It changes focus on the third GPS point, measures bearing of road segments, and calculates the difference between GPS heading of the vehicle and bearing of road segments.

```
proc sql;

    connect to odbc (dsn = "PostgreSQL_Guidance");
    title 'Map Matching Result (Nearest Neighbor & Match Heading)';
    create table mm_result as
    select * from connection to odbc
    (select gps.file_id, gps.timestamp, gps.latitude, gps.longitude,
    gps.heading_gps,
    road.objectid, road.fullstname, road.shield, road.hwy_num, road.fcc, road.acc,
    road.jurisdicti, road.county_nam, road.mcd_name,
    ST_Azimuth(endpoint(road.shape), startpoint(road.shape))/(2*pi())*360
    as road_bearing, /* (F) */
    abs(ST_Azimuth(endpoint(road.shape), startpoint(road.shape))/(2*pi())*360 -
    gps.heading_gps) as heading_diff /* (G) */
    from scwu.shrp_2_gps_data as gps,
    scwu.roads_ny_simplifiedstreetsegmentann_wgs84 as road
    where file_id = 3827593 and timestamp = 3685135
    order by ST_DISTANCE(road.shape, ST_SetSRID(ST_Point(gps.longitude,
    gps.latitude),4326))
    limit 3;);
    disconnect from odbc;

quit;
```

Part F uses the PostGIS function, ST\_Azimuth, to measure bearing of road segments based on two nodes of each road segment. Part G calculates difference between GPS heading and calculated bearing values from Part F. The SAS output in Figure 6 shows the top three closest road segments and only the W Buffalo Street bearing matches direction of what GPS heading indicates.

file_id	timestamp	latitude	longitude	heading_gps	objectid	fullstname	shield	hwy_num	fcc	acc	jurisdicti	county_nam	mcd_name	road_bearing	heading_diff
3827593	3685135	42.74021	-78.1391	89.040001	276725	S Maple St			A41	5	13	Wyoming	Warsaw	180.6039	91.56389
3827593	3685135	42.74021	-78.1391	89.040001	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw	86.04846	2.991538
3827593	3685135	42.74021	-78.1391	89.040001	212400	N Maple St			A41	5	13	Wyoming	Warsaw	180.6227	91.58273

Figure 6. Map-Matching Results

```
proc sql;

    connect to odbc (dsn = "PostgreSQL_Guidance");
    title 'Map Matching Result (Nearest Neighbor & Match Heading)';
    create table mm_result as
    select * from connection to odbc
    (select gps.file_id, gps.timestamp, gps.latitude, gps.longitude,
    gps.heading_gps,
    road.objectid, road.fullstname, road.shield, road.hwy_num, road.fcc, road.acc,
    road.jurisdicti, road.county_nam, road.mcd_name,
    ST_Azimuth(endpoint(road.shape), startpoint(road.shape))/(2*pi())*360
    as road_bearing,
    abs(ST_Azimuth(endpoint(road.shape), startpoint(road.shape))/(2*pi())*360 -
    gps.heading_gps) as heading_diff
    from scwu.shrp_2_gps_data as gps,
    scwu.roads_ny_simplifiedstreetsegmentann_wgs84 as road
    where file_id = 3827593 and timestamp = 3685135 and
    abs(ST_Azimuth(endpoint(road.shape), startpoint(road.shape))/(2*pi())*360 -
    gps.heading_gps) <= 25 /* (H) */

quit;
```



```

order by ST_DISTANCE(road.shape, ST_SetSRID(ST_Point(gps.longitude,
gps.latitude),4326))
limit 1;);
disconnect from odbc;
quit;

```

The above sample code adds Part H as an additional map-matching criterion which filters out cases with difference between GPS heading and bearing of road segments is greater than 25 degrees. Results are shown in Figure 7. These five GPS coordinates are visually checked in Google map (Figure 8) which matches the SAS results.

Results Viewer - SAS Output

Map Matching Result (Nearest Neighbor & Match Heading)

file_id	timestamp	latitude	longitude	heading_gps	objectid	fullstname	shield	hwy_num	fcc	acc	jurisdict	county_nam	mcd_name	road_bearing	heading_diff
3827593	3683134	42.7402	-78.1396	89.209999	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw	86.04846	2.991538
3827593	3684134	42.7402	-78.1394	88.019997	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw	86.04846	2.991538
3827593	3685135	42.74021	-78.1391	89.040001	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw	86.04846	2.991538
3827593	3686135	42.7402	-78.1389	90.680000	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw	86.04846	2.991538
3827593	3687135	42.74021	-78.1387	90.480003	350402	W Buffalo St	U	20A	A21	3	01	Wyoming	Warsaw	86.04846	2.991538

Figure 7. Map-Matching Results

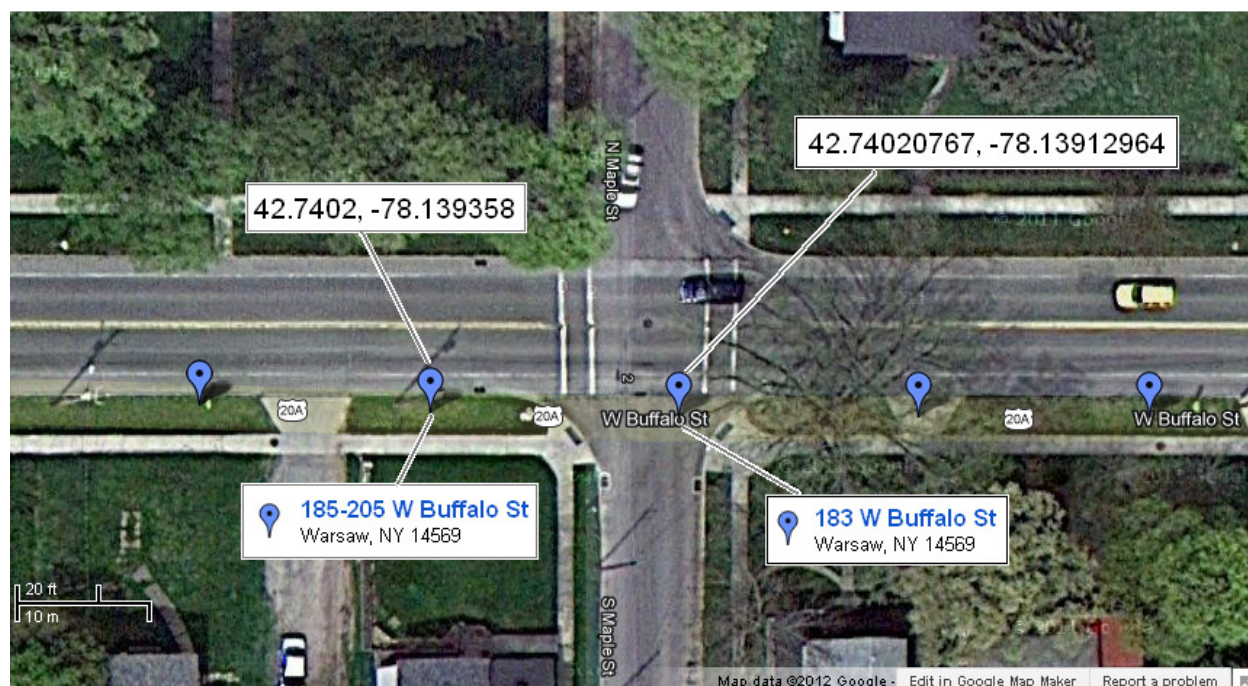


Figure 8. Visual Validation of Map Match Results on Google Map

When working with large scale data sets, it is obviously not feasible to visually check or even portray all data points. Validation methods and the amount of algorithm testing will depend on factors such as research needs, cost of misclassifications, and opportunities to detect misclassifications later in the process.

## CREATION OF A HEATMAP VISUALIZATION

The idea of a heatmap visualization uses graduated colors to describe frequency with which road segments are traversed. After the map-matching process, GPS coordinates have been converted to a sequence of road segments as shown in Figure 7. A summary of trips counts on each road segment can be created via the following SAS PROC SQL procedure and partial results are shown in Figure 9.

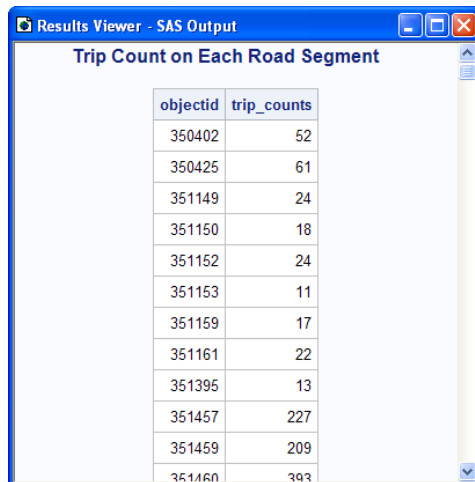
```

proc sql;
title 'Trip Count on Each Road Segment';

```

```
create table trip_counts as
select objectid, count(distinct file_id) as trip_counts
from mm_result
group by objectid;
```

```
quit;
```



objectid	trip_counts
350402	52
350425	61
351149	24
351150	18
351152	24
351153	11
351159	17
351161	22
351395	13
351457	227
351459	209
351460	293

Figure 9. SAS PROC SQL Output

SAS Bridge for ESRI [6] enables ArcGIS [7] to read a SAS data set which allows ArcGIS to create heatmaps based on the summary in Figure 9. Through the ArcGIS GUI, first, add the SAS data set and the New York roadway map into ArcMap. Then join the two imported data sets via the attribute “objectid” as shown in Figure 10. After ArcGIS spatial joins trip counts to the map data using “objectid,” information of trip counts is added to the layer of roadway map. The user can then modify layer properties and use quantity “trip\_counts” to specify the color in which road segments with different numbers of trips will be portrayed (Figure 11). Finally, a heatmap visualization has been created.

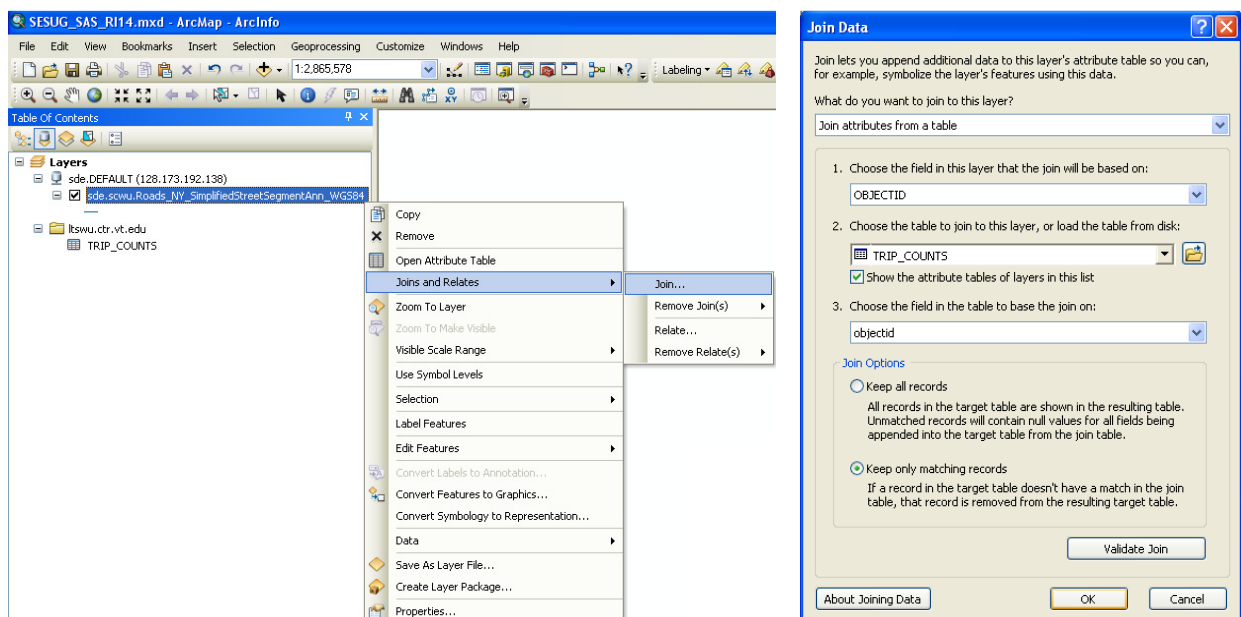
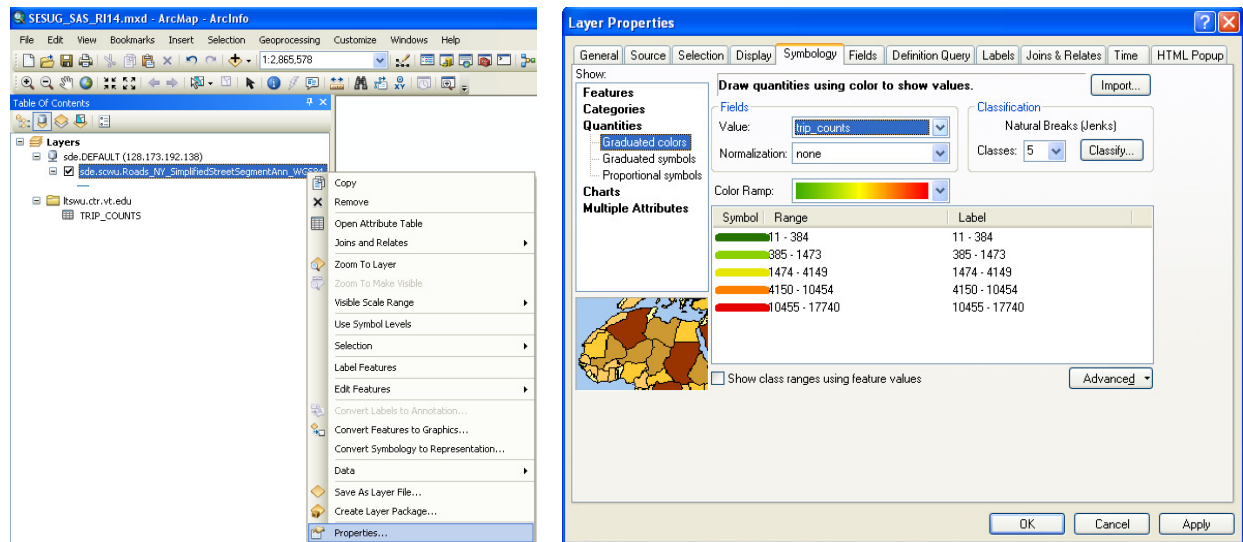
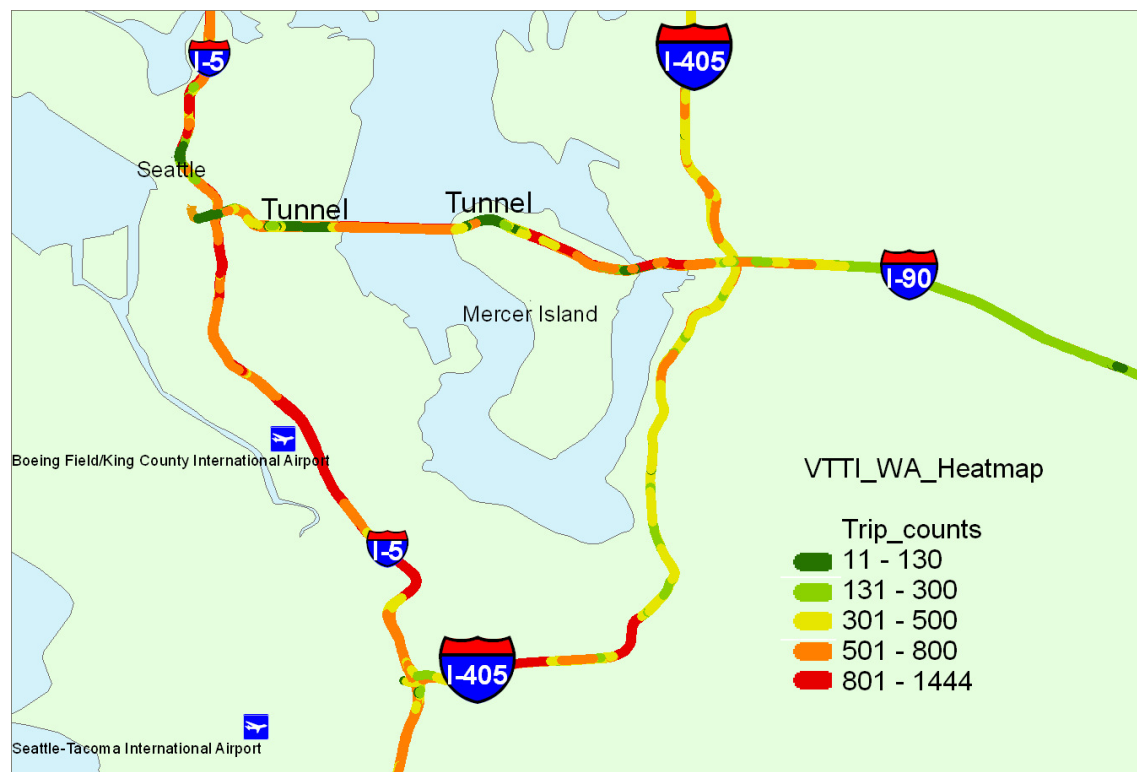


Figure 10. Creation of Heatmap Visualization in ArcGIS

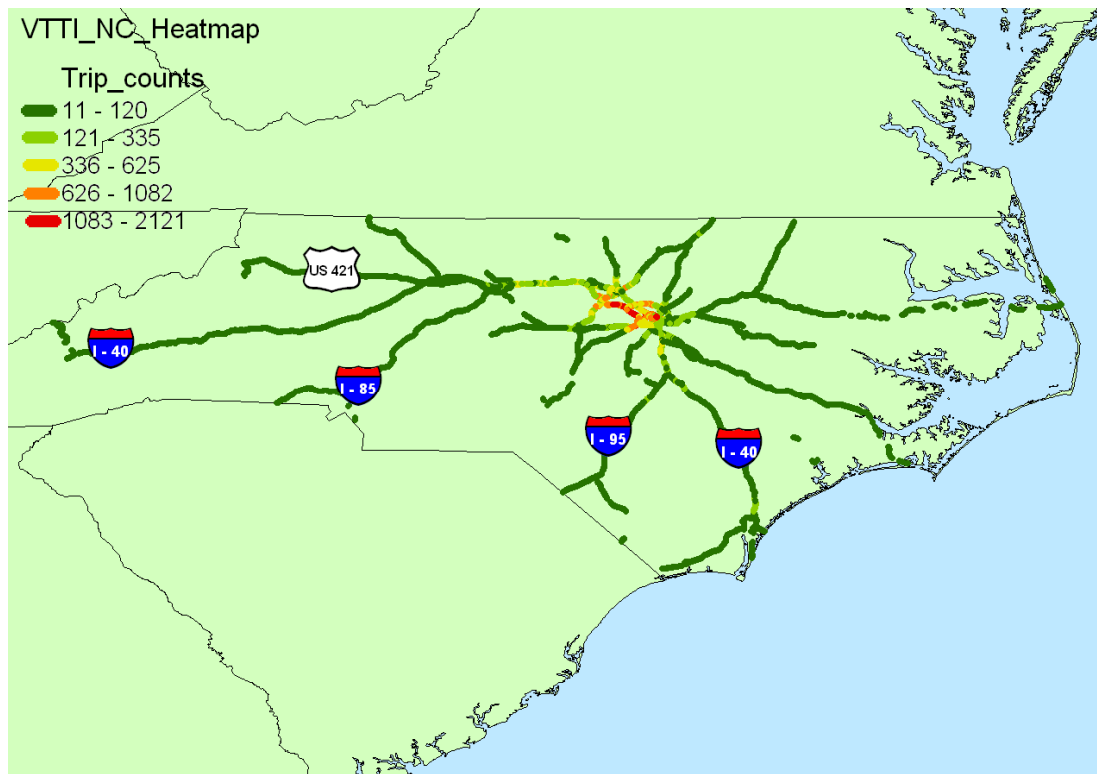
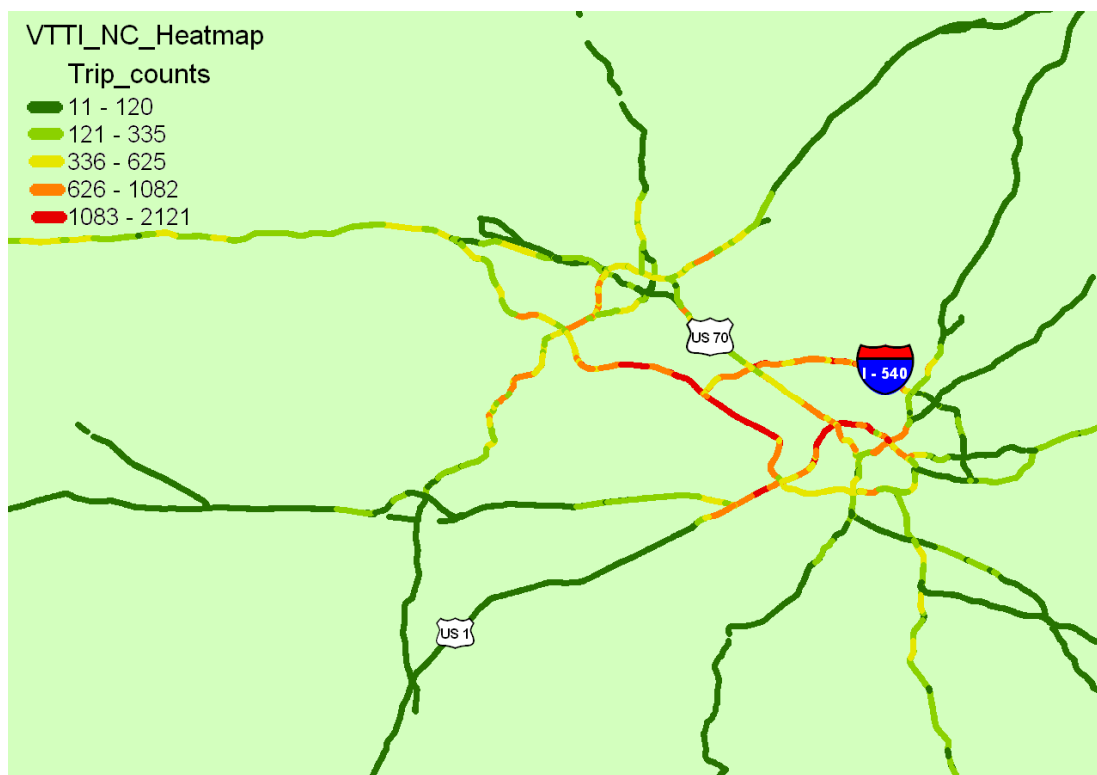


**Figure 11. Creation of Heatmap Visualization in ArcGIS**

The following figures are sample heatmap visualizations from some of the SHRP 2 NDS data. Figure 12 shows trips traveled only on interstates in the Seattle Washington area. Figure 13 shows trips that have been associated with interstates and U.S. highways in North Carolina. Figure 14 zooms in to show mapping of these data points to road segments in the Raleigh area.



**Figure 12. Heatmap Visualization**

**Figure 13. Heatmap Visualization****Figure 14. Heatmap Visualization**



## ADDITIONAL BENEFITS

There are a number of direct and ancillary benefits of the described approach. As mentioned in the introduction, the process described here is much faster than GUI based methods. More specifically though, GUI methods would require a user to manually break up the task into subtasks and monitor completion of each portion. The process described here lends itself to being divided up in an automated way across multiple computers, thereby reducing the overall time to completion [8]. For example, the 156 hour example could be divided by four if four computers are used. The method also permits process monitoring and graceful recovery in the event of a process interruption. For example, the user can continuously monitor the accumulation of results in the database and identify when an interruption occurred.

## CONCLUSION

Attributes included in GISs are important for transportation research, but the capability of efficiently processing large data in GUI-based GIS software is limited. Integrating SAS with PostgreSQL, PostGIS, and ArcGIS provides usable solutions to accomplish location-based associations of large quantities of data. This paper demonstrates an efficient approach to associate GPS data to roadway maps and create informative heatmap visualizations.

## REFERENCES

1. PostgreSQL. <http://www.postgresql.org/>.
2. PostGIS. <http://postgis.refractory.net/>.
3. SAS Institute, 2011, SAS/ACCESS® 9.3 for Relational Databases Reference. Cary, NC: SAS Institute, Inc. <http://support.sas.com/documentation/cdl/en/acrelldb/63144/PDF/default/acrelldb.pdf>.
4. Strategic Highway Research Program 2. <http://www.shrp2nds.us/>.
5. New York State Geographic Information Systems (GIS) Clearinghouse. <http://gis.ny.gov/gisdata/inventories/details.cfm?dsid=932&nysgis=#>.
6. Massengill, Darrell. 2007. "Making Business Decisions Using SAS Mapping Technologies". Proceedings of the SAS Global Forum 2007 Conference. Cary, NC. SAS Institute Inc. <http://support.sas.com/rnd/papers/sgf07/sgf2007-map.pdf>.
7. ESRI ArcGIS. <http://www.esri.com/>.
8. Wu, Shih-Ching and McLaughlin, Shane. 2012. "Tips for Using SAS to Manipulate Large-scale Data in Databases". the NorthEast SAS Users Group (NESUG) 2012 Conference.

## ACKNOWLEDGMENTS

The authors would like to acknowledge the research participants and the external contributions which made this work possible. The data used in this study were collected as part of the SHRP 2 Naturalistic Driving Study, which is sponsored by the Transportation Research Board of the National Academies. The methods developed and described in this paper were developed with funds from the National Surface Transportation Safety Center for Excellence.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Shih-Ching Wu  
Virginia Tech Transportation Institute  
3500 Transportation Research Plaza  
Blacksburg, VA, 24061  
540-231-1091  
[swu@vtti.vt.edu](mailto:swu@vtti.vt.edu)  
<http://www.vtti.vt.edu>

Shane McLaughlin  
Virginia Tech Transportation Institute  
3500 Transportation Research Plaza  
Blacksburg, VA, 24061  
540-231-1077  
[smclaughlin@vtti.vt.edu](mailto:smclaughlin@vtti.vt.edu)  
<http://www.vtti.vt.edu>

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.