

A Review of GeoStats' Efforts with Visualization of NHTS Data

Marcelo Simas Oliveira, GeoStats, Atlanta GA



GeoStats

Presentation Outline

- Introduction
- Background
- **GeoStats & Visualization**
- **Visualizing NHTS Data**
 - ✓ R and ggplot2
 - ✓ Google Fusion Tables
- **Final Remarks**

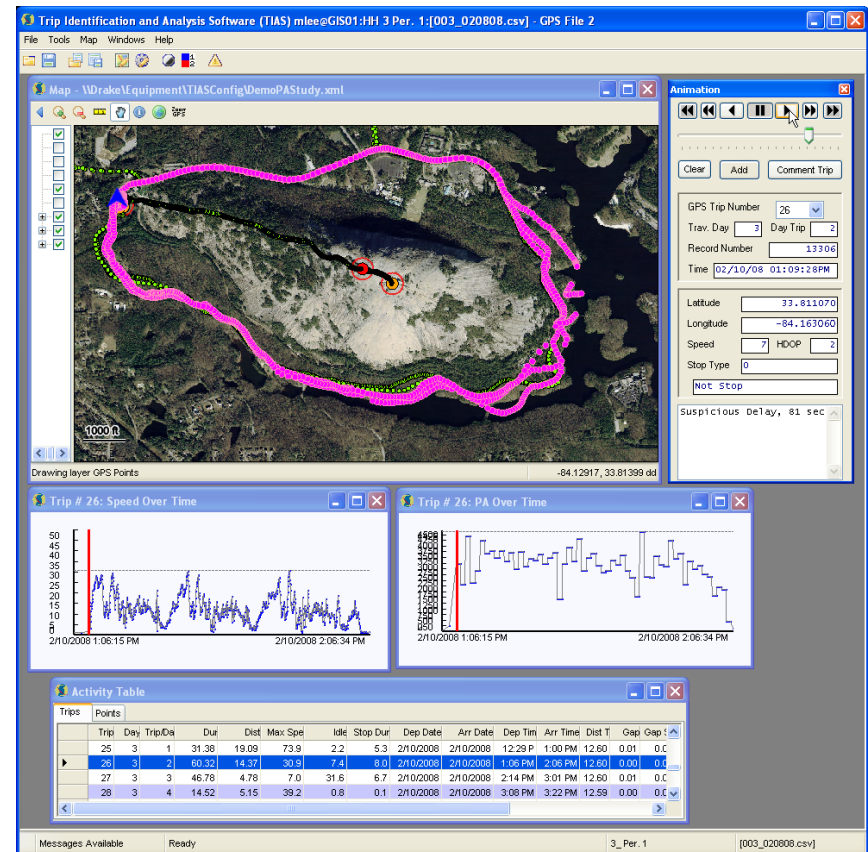
GeoStats Background

- ✓ GeoStats, a women-owned business established in 2000, is a leader and innovator in the development and use of the Global Positioning System (GPS) and Geospatial Systems for the collection, analysis, and reporting of transportation data. GeoStats is located in Atlanta, Georgia – USA
- ✓ Leader in GPS Household Travel Surveys
 - Developed processing tools for all modes of GPS travel survey
 - Now working on 100% GPS travel surveys (Jerusalem and Cleveland)
- ✓ Transportation System Performance
- ✓ Physical Activity Studies
- ✓ Technology for transportation data collection
 - Software Development



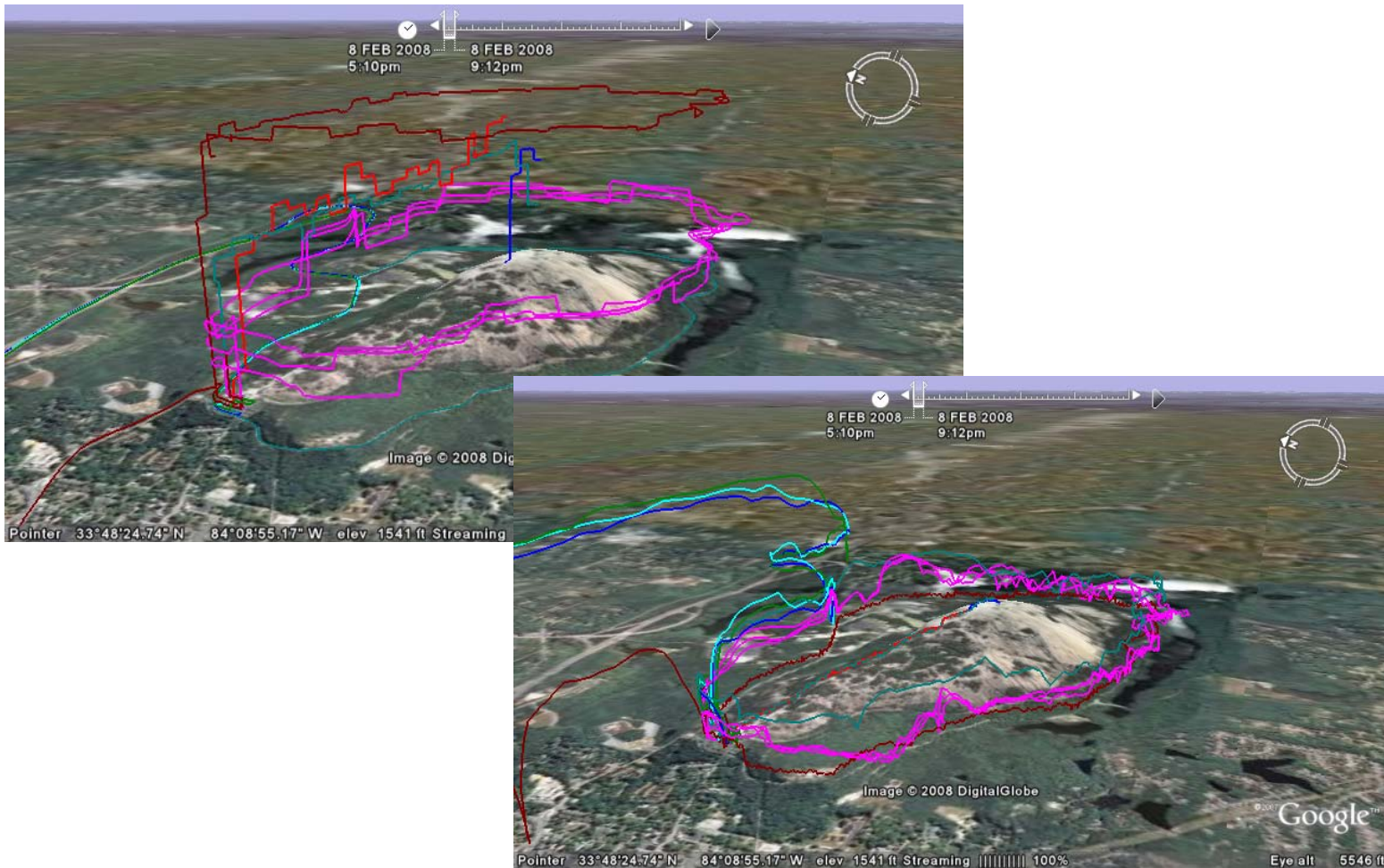
GeoStats & Visualization – GPS Processing

- ✓ GPS Data => Wealth of data
 - Strong location component
 - Processing uses tools with built in visualization
 - Visualization helps to parse through the large volume of data and to deal with the multiple dimensions in the data



GeoStats & Visualization – Google Earth

- ✓ Example of a combined GPS and accelerometer data

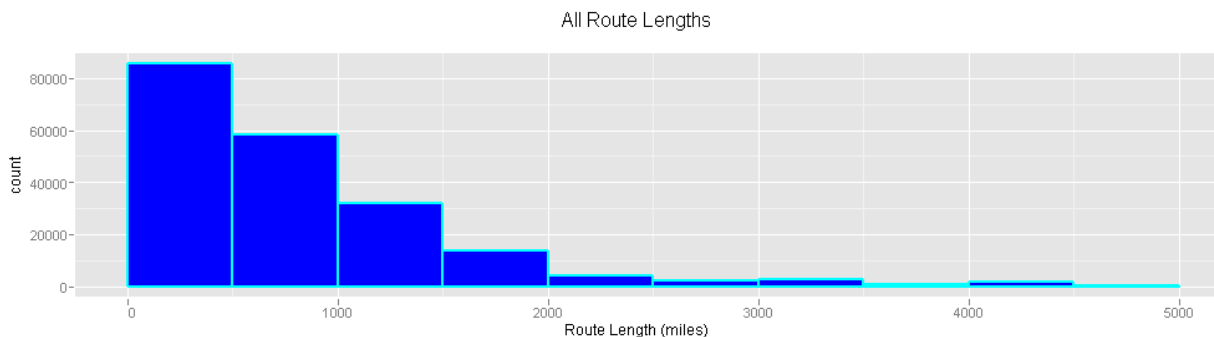
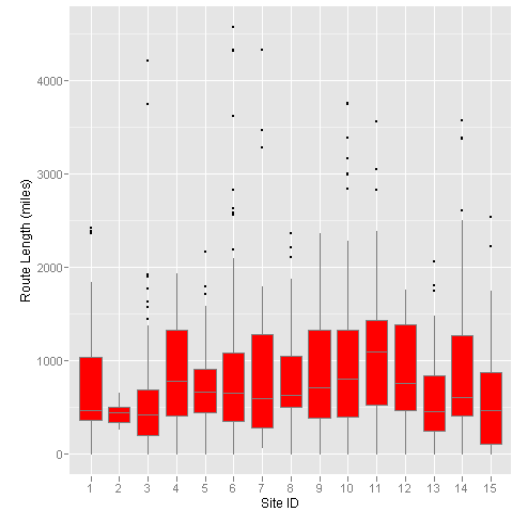


Using NHTS Data

- ✓ Using only OSS or “free” tools
- ✓ Wanted to use what’s available – so downloaded files from website
- ✓ Downloaded and read documentation 😊
 - Made sure I used weights!
- ✓ Basic set of steps:
 - First tried to bring in the SAS Transport files directly into R – did not work, resorted to downloading the CSV files
 - Created database schema (set of tables that mapped to the original CSV files)
 - Imported file into a database (SQLite)
 - Wrote queries to bring in and summarize data into R
 - Generated charts in R
 - Exported data aggregations to CSV for upload to Google Fusion Tables

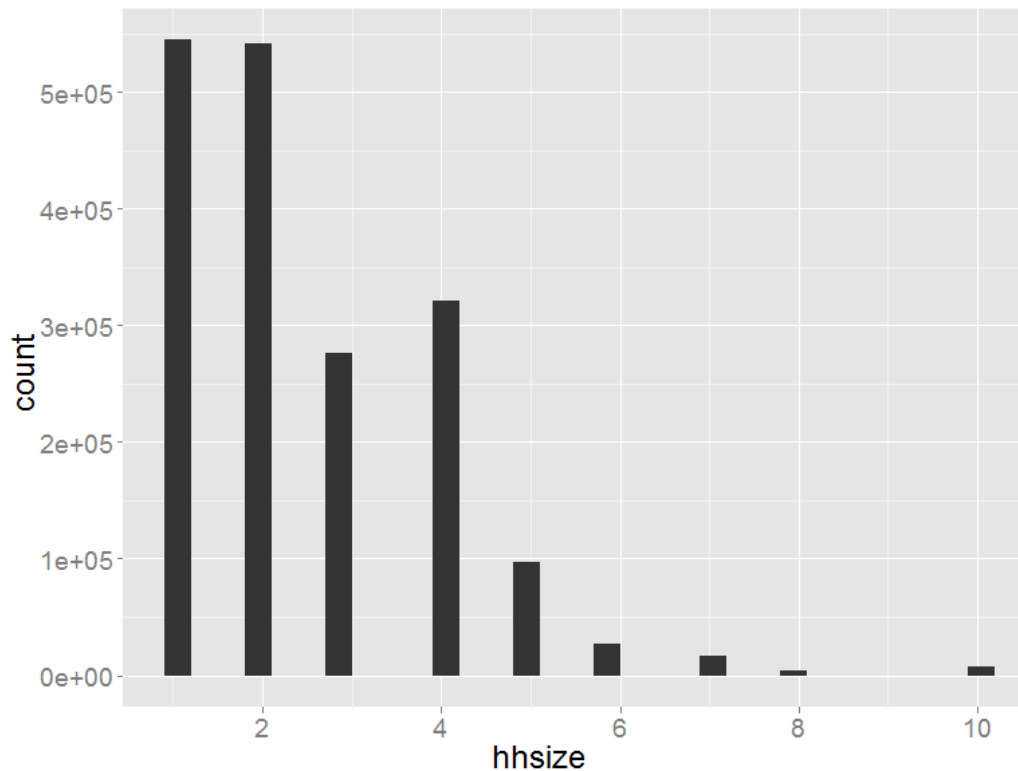
NHTS Data using R and ggplot2

- ✓ R is a OSS implementation of the S-Plus language
 - Command-line oriented
 - kind of steep learning curve but the payoff is worth it
 - Can be used to automate visualizations
- ✓ The ggplot2 library generates high quality graphics in R
 - <http://had.co.nz/ggplot2>
 - Supports using weights
- ✓ Allows user to overlay graph information using the concept of “aesthetics”



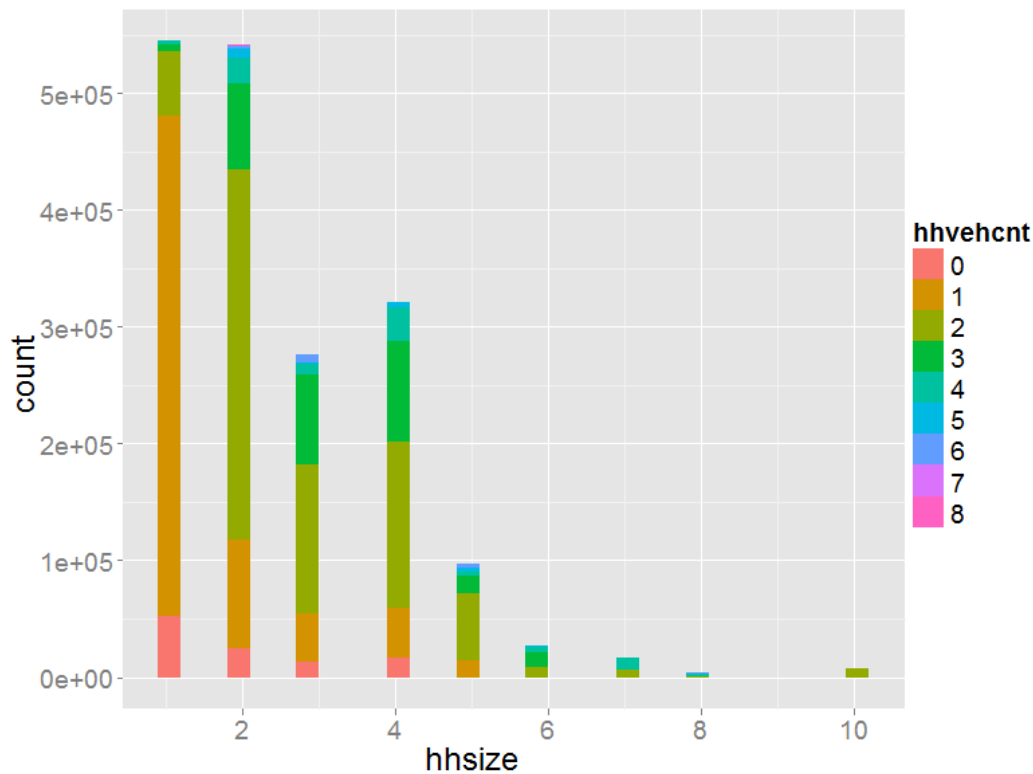
NHTS Data using R and ggplot2

- ✓ Used data for the Atlanta MSA Area (0502), household size distribution.
 - Start with a simple plain bar chart.



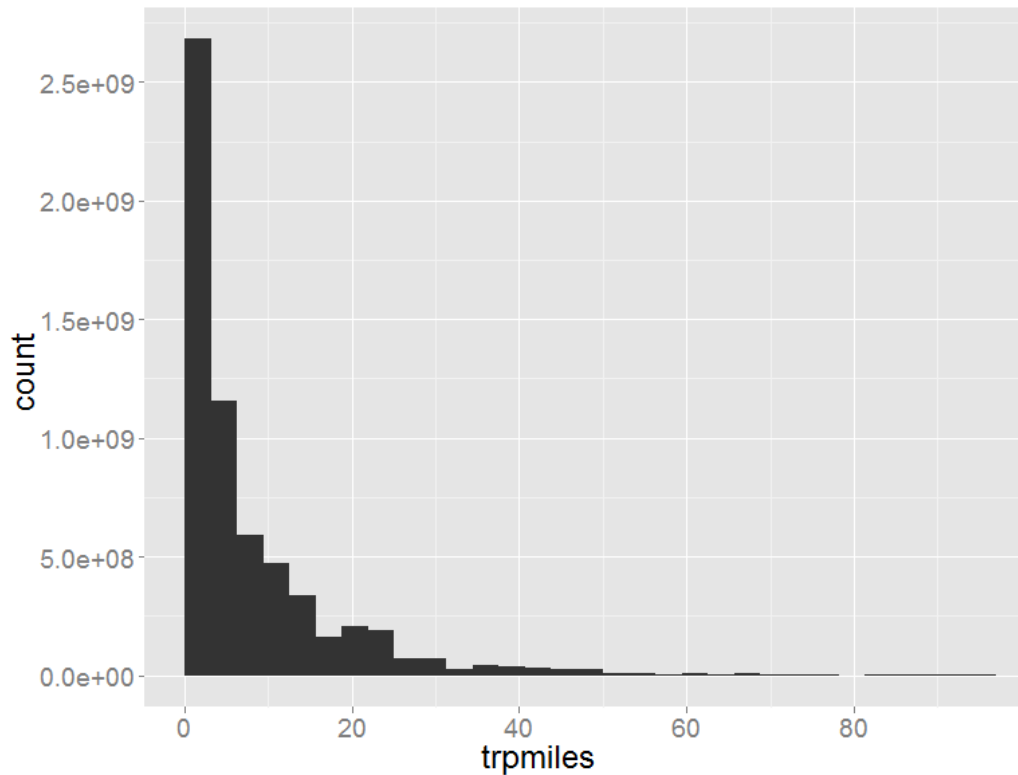
NHTS Data using R and ggplot2

- ✓ Household size distribution, color reflects number of vehicles



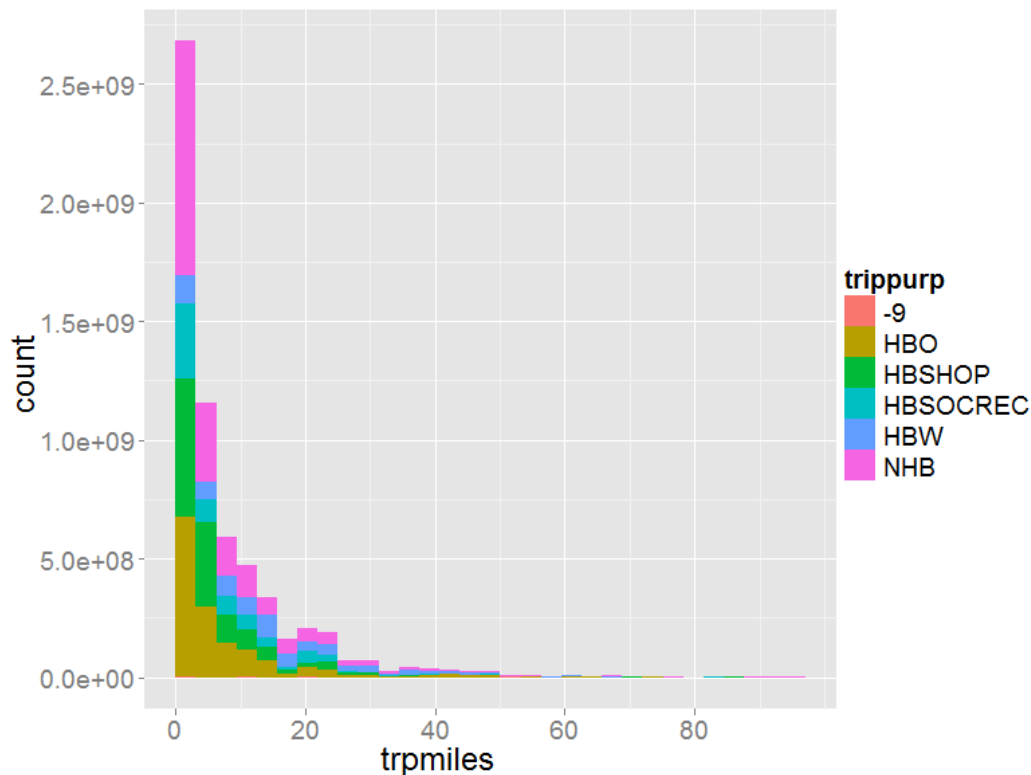
NHTS Data using R and ggplot2

- ✓ Trip length distribution for trips shorter than 100 miles, removed negative values (codes for missing data)



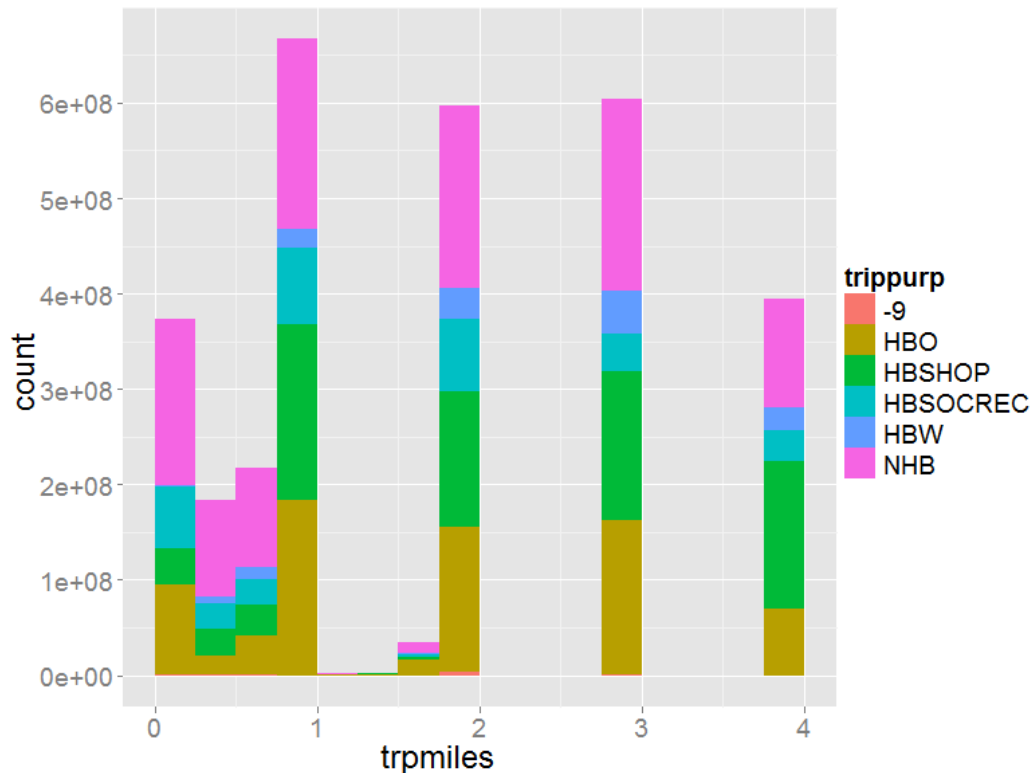
NHTS Data using R and ggplot2

- ✓ Trip length distribution for trips shorter than five miles, color reflects gen. purpose



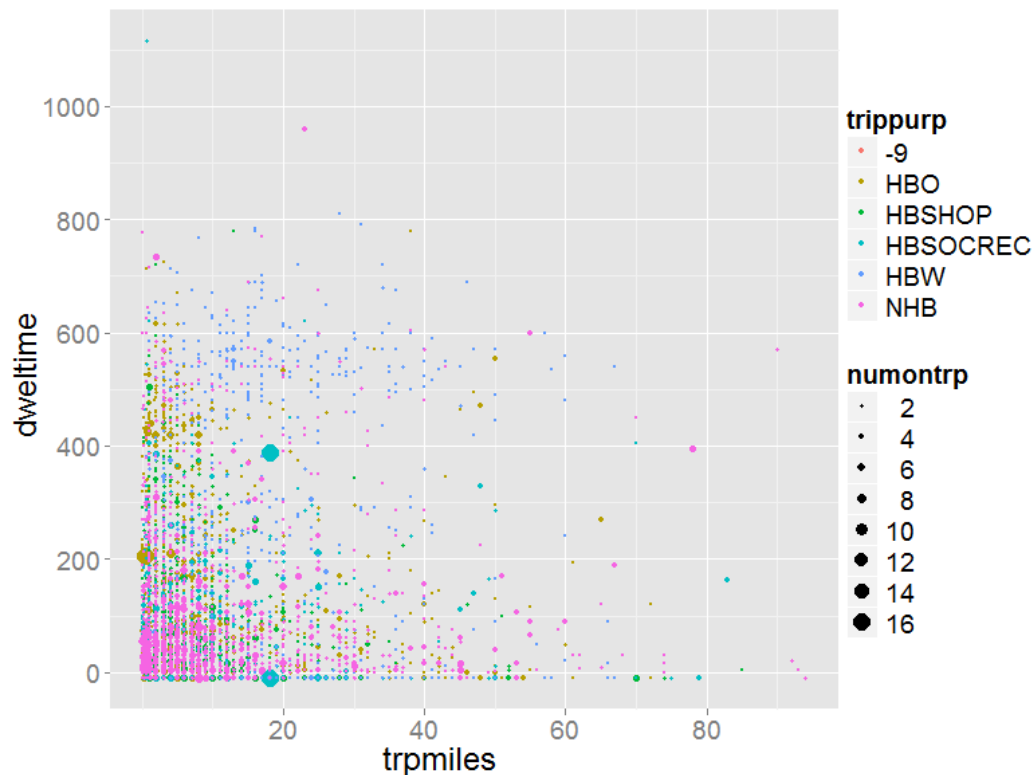
NHTS Data using R and ggplot2

- Trip length distribution for trips shorter than five miles, color reflects gen. trip purpose (bin width set to 0.25) – shows evidence of rounding



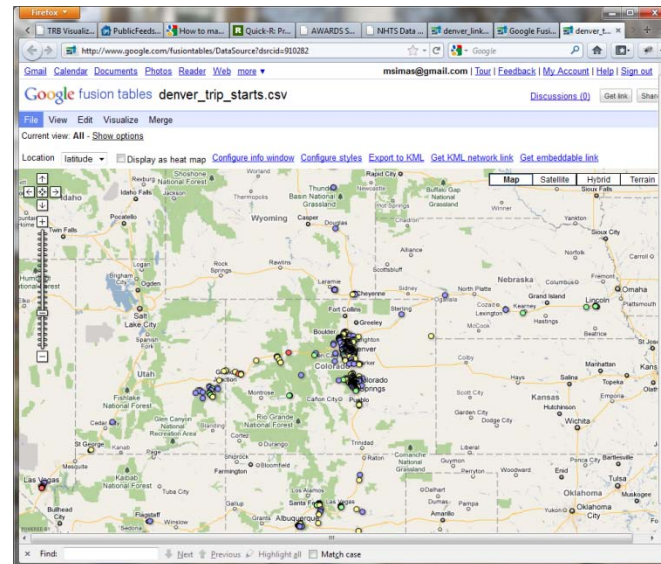
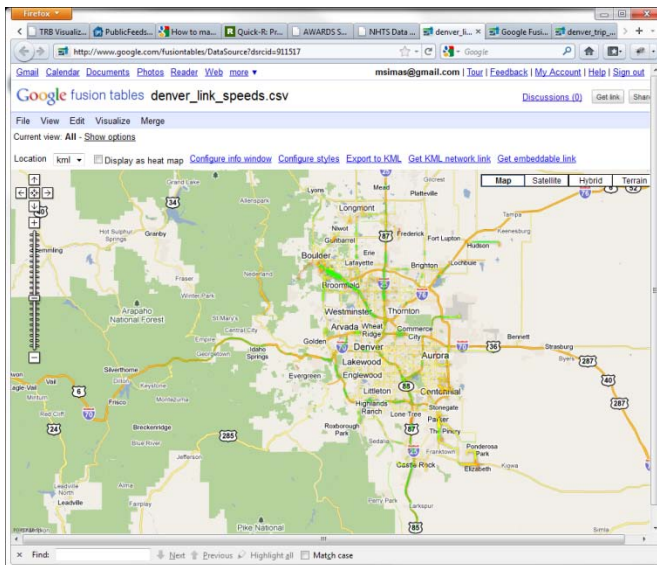
NHTS Data using R and ggplot2

- ✓ Trip length for trips shorter than 100 miles as a function of dwell time (minutes), color reflects gen. purpose and dots are sized as a function of the number of persons on the trip



Google Fusion Tables

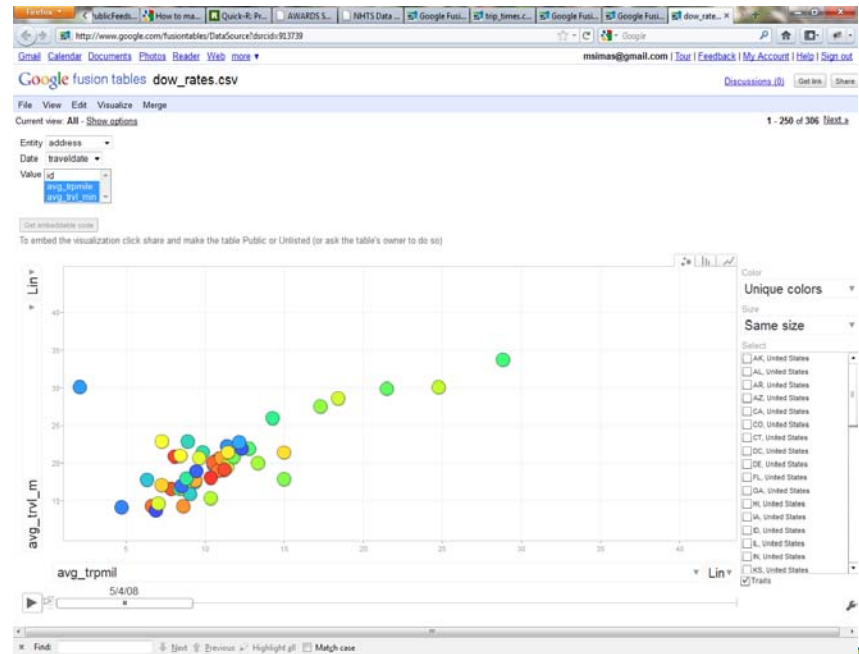
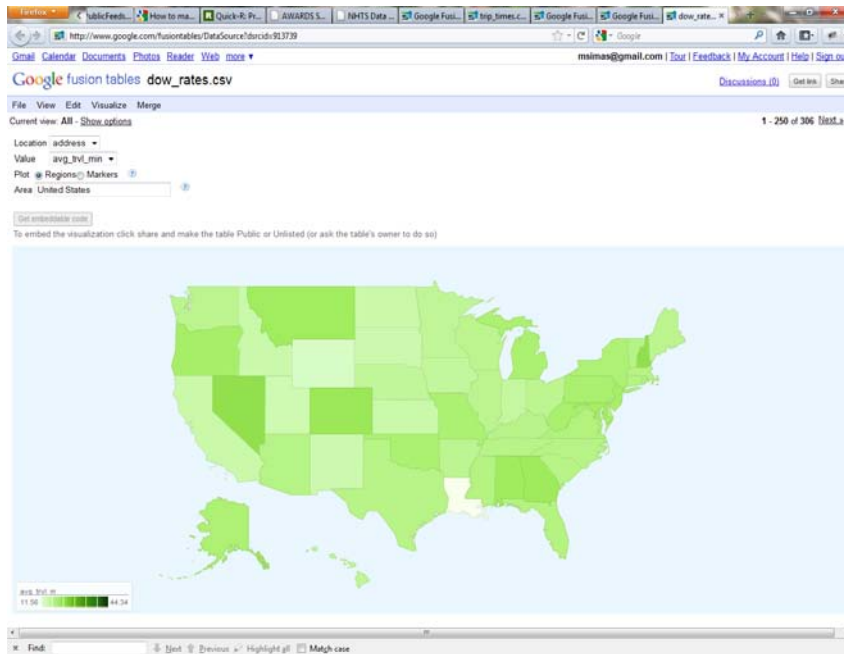
- ✓ Cloud-based data manipulation, storage and visualization tools -> easy to share
- ✓ Basic analytical capabilities + joining datasets
- ✓ Quick and easy visualization
 - DRCOG Vehicle GPS Data Shown (33,078 trips, 673 households, 1328 vehicles)
 - GPS trip traces matched to the model networks and trip starts



NHTS Data using Google Fusion Tables

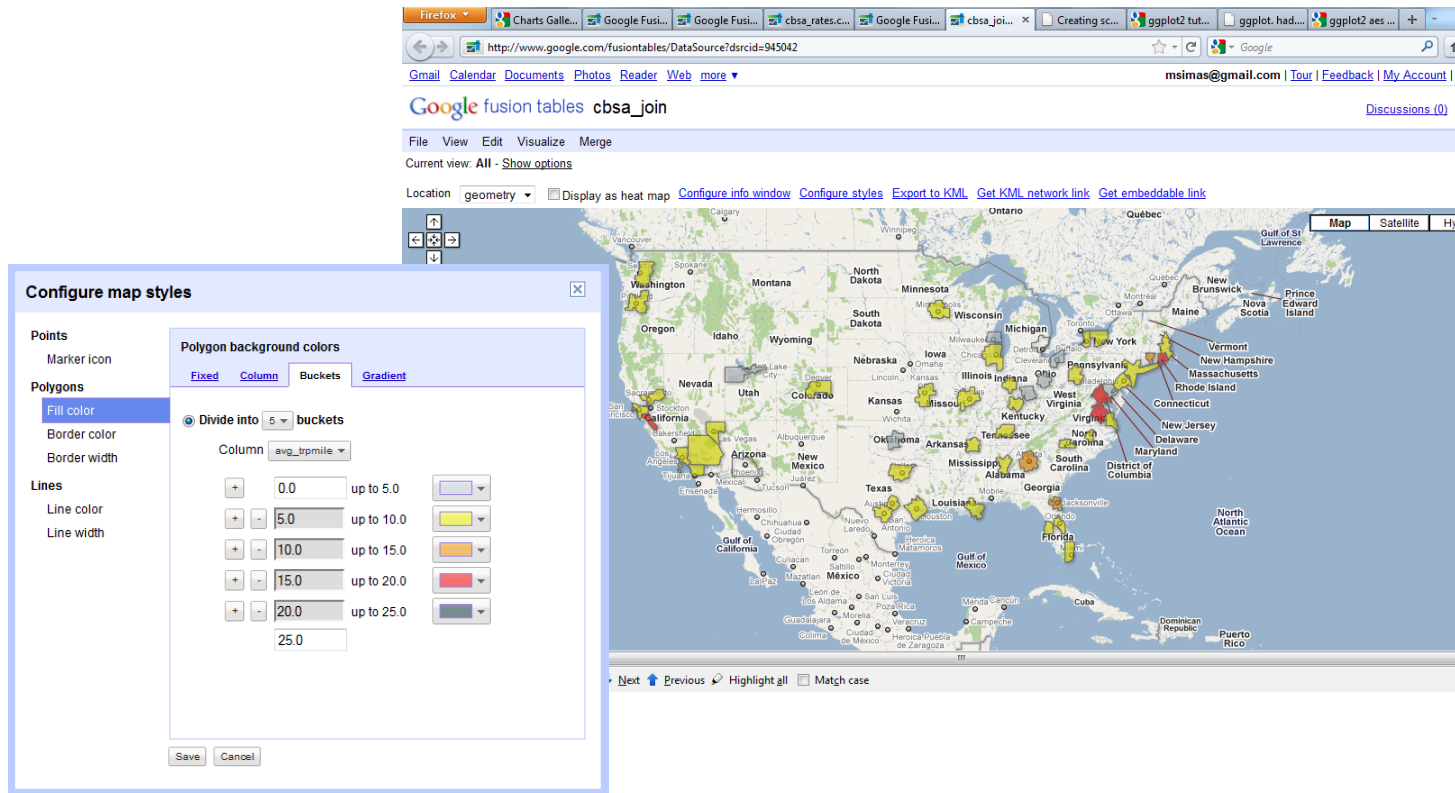
✓ Example visualizations

- Intensity Maps – average trip distance by state
- Motion – trip distance vs. travel time over an artificial week (constructed using DOW information in the trip records)



NHTS Data using Google Fusion Tables

- ✓ Downloaded MSA Shapefile from the Census website, converted it to KML uploaded and merged it with NHTS data
 - Color shows average trip length by MSA



Final Remarks

- ✓ Spatial information in the general NHTS files do not allow for very interesting mapping of the data
- ✓ It would be helpful if the NHTS data were to be provided in a format that supports code-sets and which R can understand

- ✓ Sample R code and SQL queries can be provided upon request
- ✓ Same goes for the Google Fusion tables – made all them public

- ✓ Thank you!
 - Marcelo Simas Oliveira (moliveira@geostats.com)