

The Right UBI Data for Now and the Future

Lessons Learned Working with UBI Data

2013 CAS Special Interest Seminar: Elephants in the Room

Len Llaguno, FCAS, MAAA

Kelleen Arquette, FCAS, MAAA

September 30, 2013

© 2013 Towers Watson. All rights reserved.

TOWERS WATSON 



Antitrust Notice

- **The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.**
- **Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.**
- **It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.**

UBI data is different

- Consider a typical commuter
 - 20 minute commute

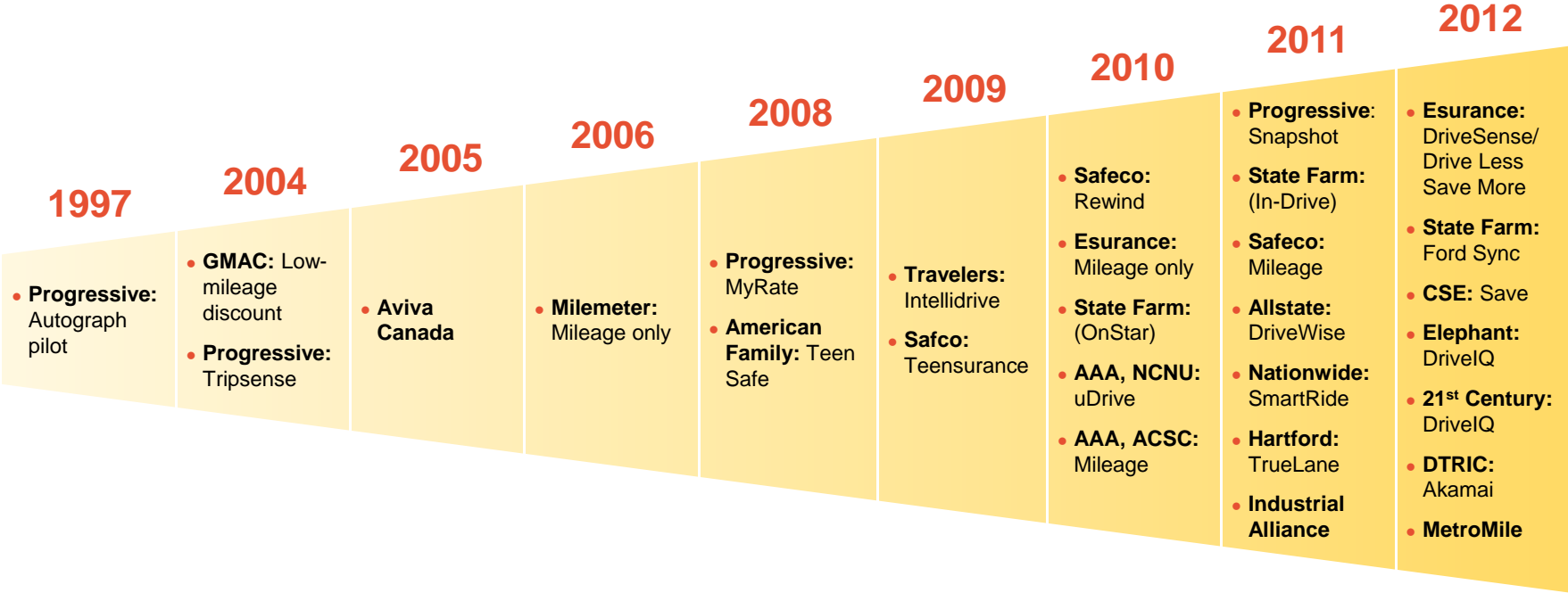
1,200 records of data

- Twice daily commute, 5 days a week, one year

500,000 records of data

That's just one vehicle!

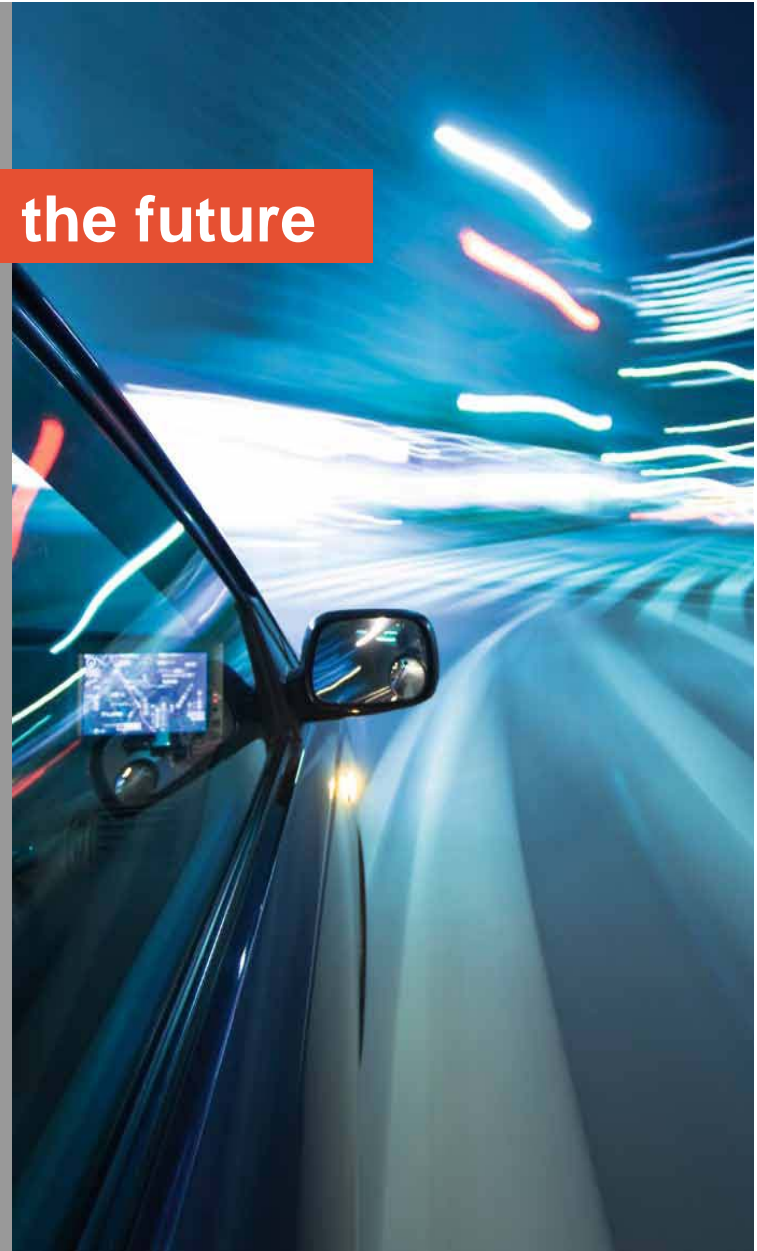
UBI programs have proliferated in North America despite concerns over data privacy



The right UBI data for now and the future

What type of data to collect?

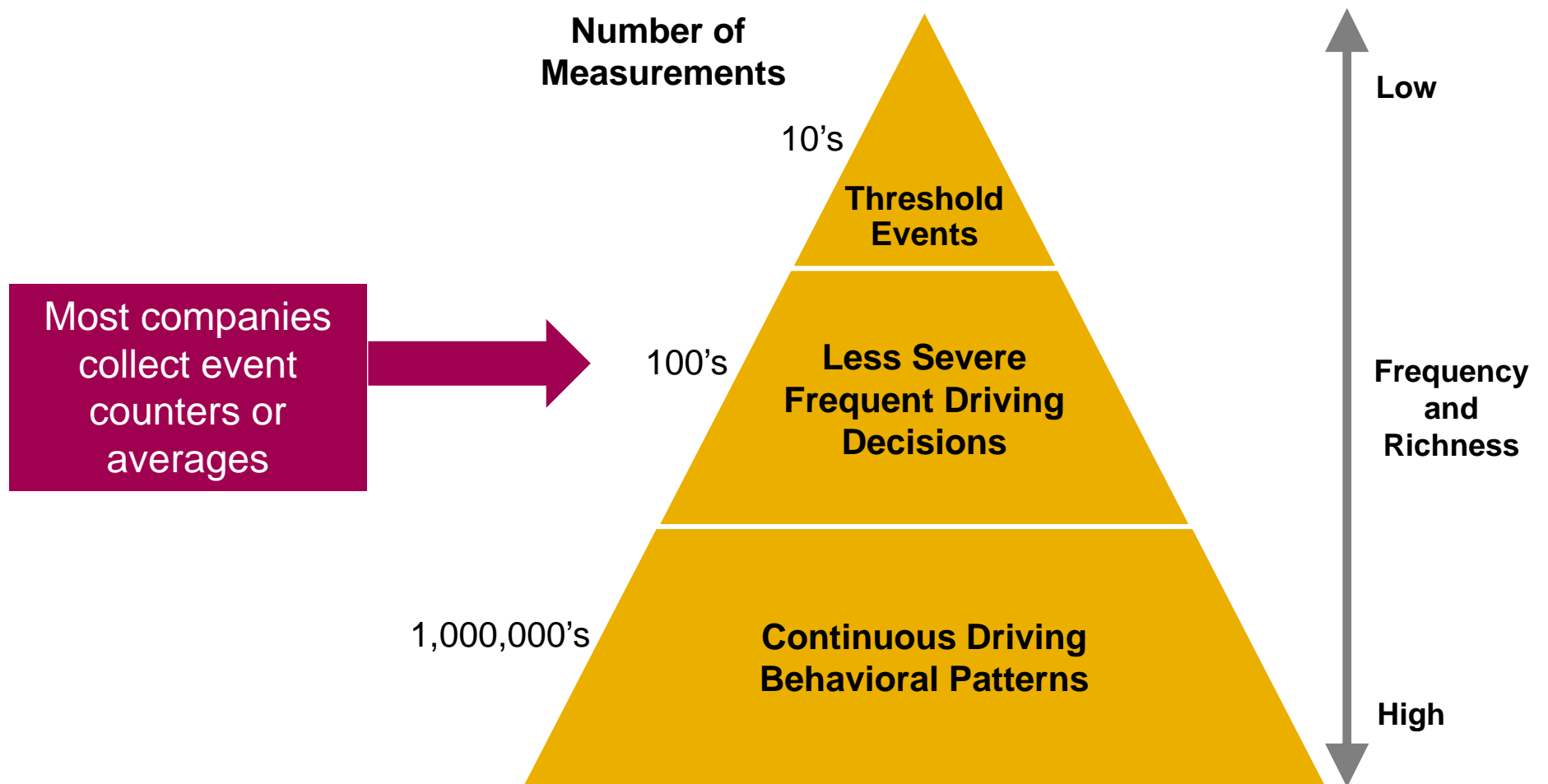
What are typical data issues?



What data to collect?

The case for granular data

There are varying degrees of granularity to UBI data



Are event counters good enough?

- Event counters and averages throw away an enormous amount of useful data
- This lack of granularity and fidelity in the data limits the knowledge that can be extracted
- To maximize the benefit of UBI data, you must collect much more granular data
 - Consider this example....

Example: Data collected every kilometer



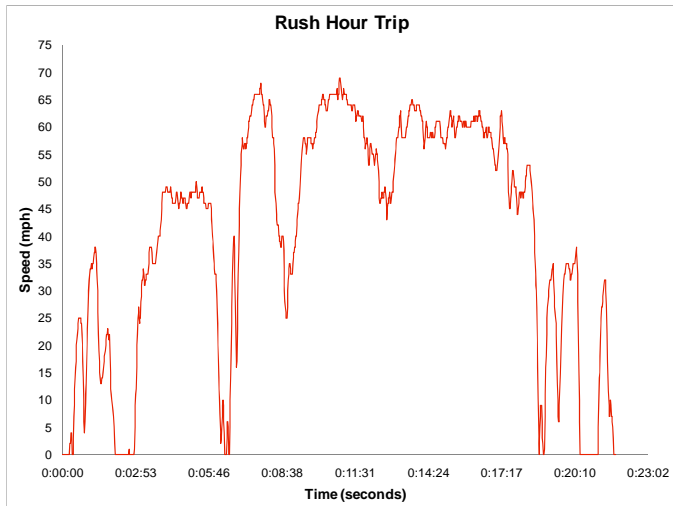
Example: Data collected every minute



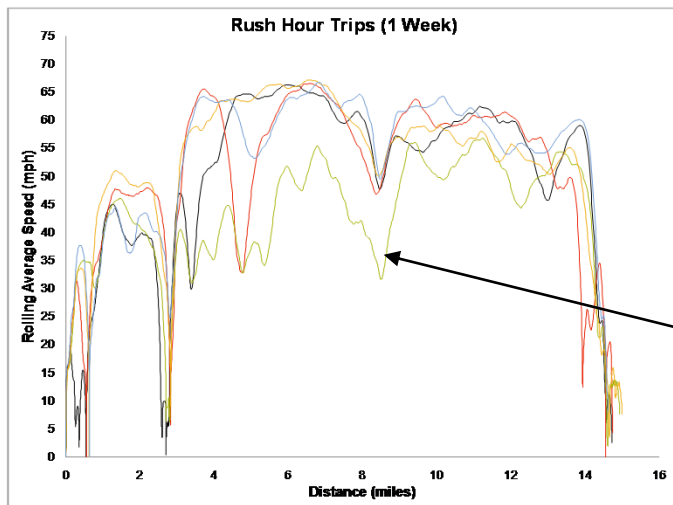
Example: Data collected every second



Granular data allows for deeper insights



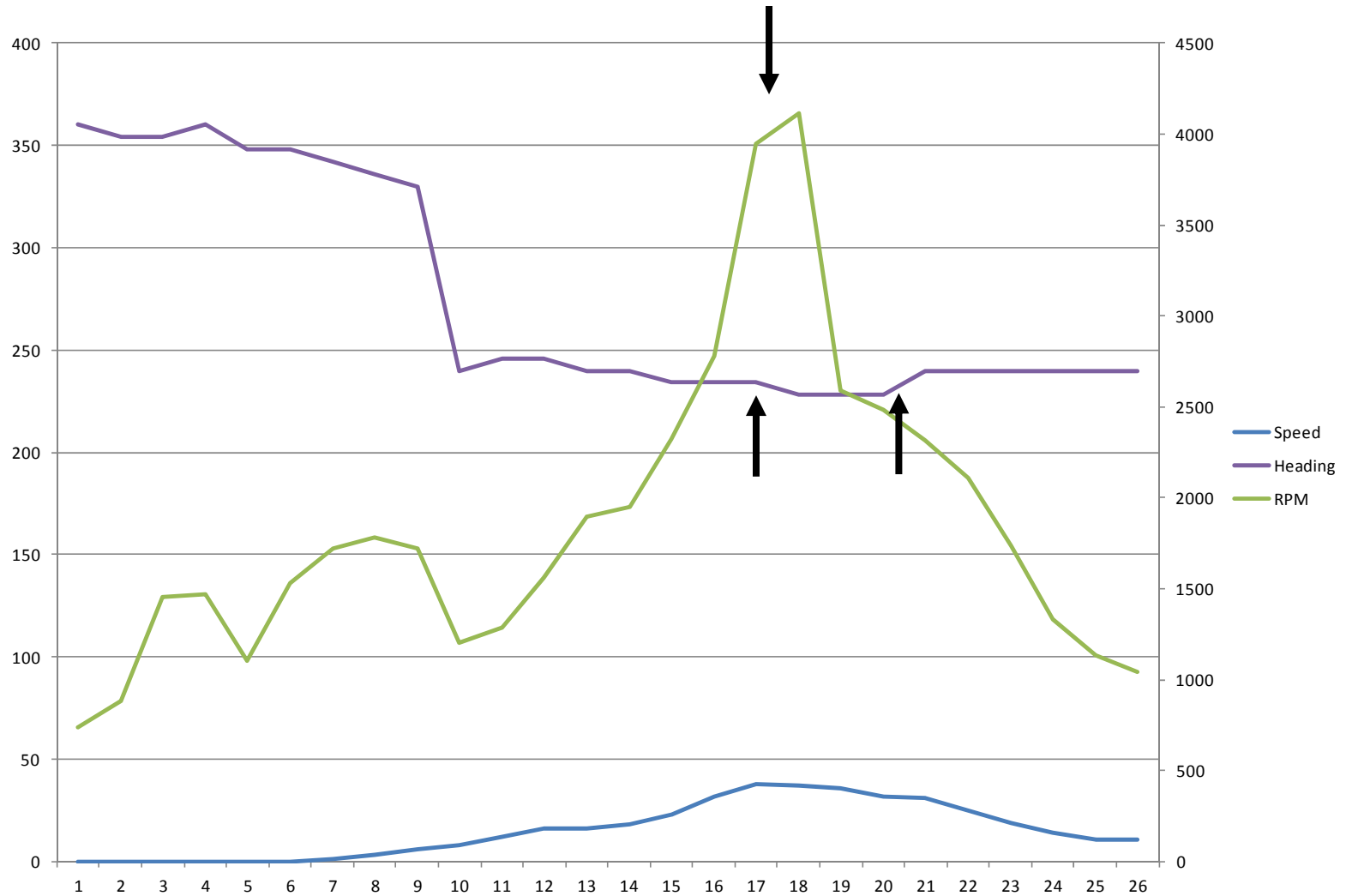
- One trip provides significant information
 - Distance
 - Time of day
 - Speed
 - Behaviors



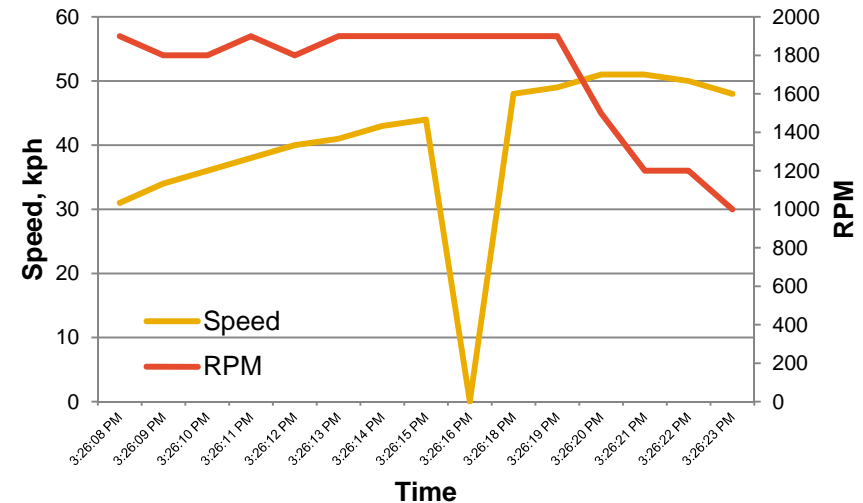
- Multiple trips begin providing patterns
- Identify risky driving behavior

Why is this trip different?

Example: identification and testing of predictive driving patterns

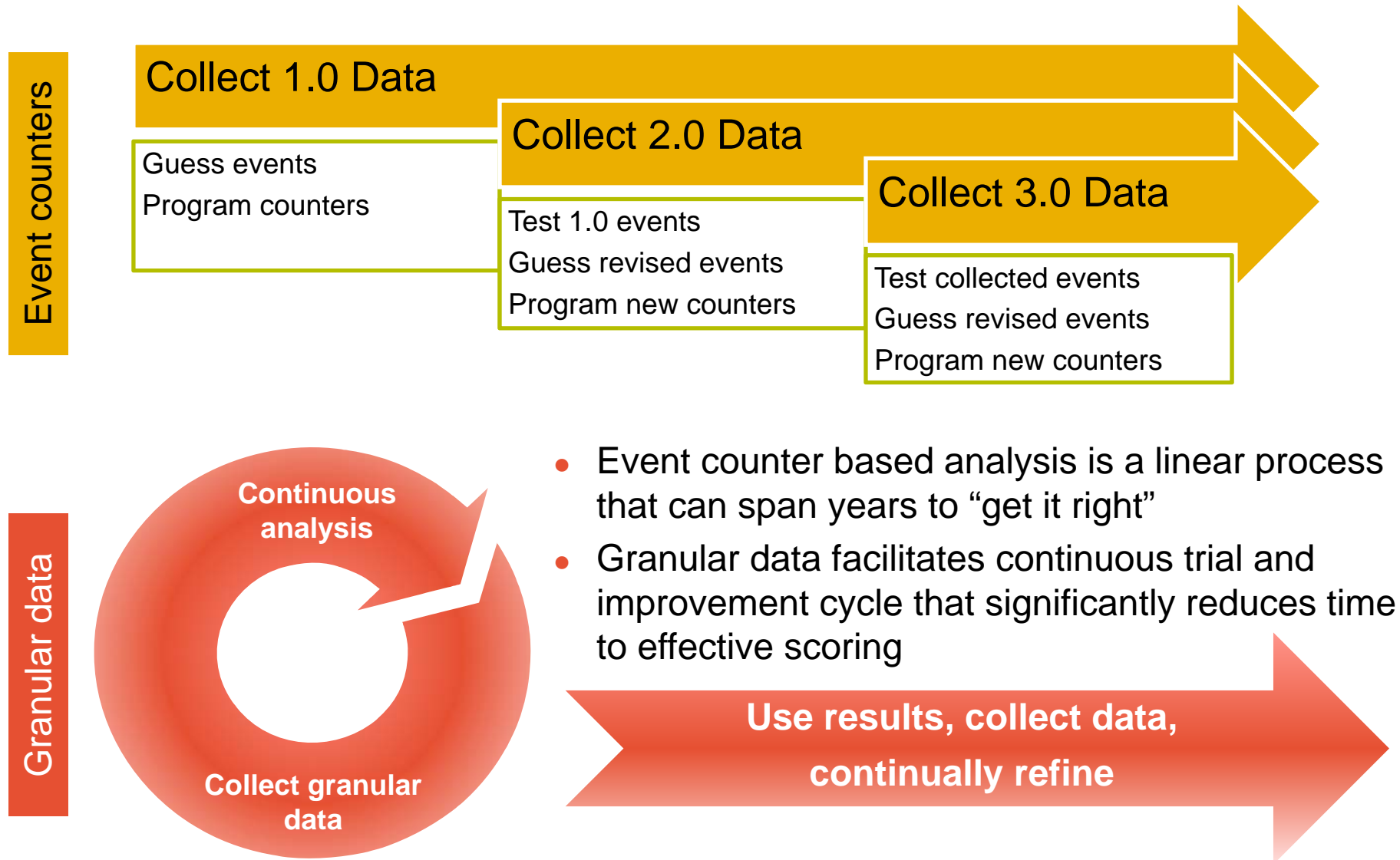


Granular data facilitates data cleansing



- Telematics data, like all data, must be scrubbed
 - Daily/weekly monitoring and scrubbing is needed to ensure completeness as data is collected
 - Our experience is that telematics data, while okay for fleet management, typically has more errors than is acceptable for pricing purposes
 - Critical to clean the data prior to the analysis to eliminate “garbage in, garbage out”
 - With granular data, possible to run scrubbing routines to minimize errors and ensure proper conclusions; more on this later

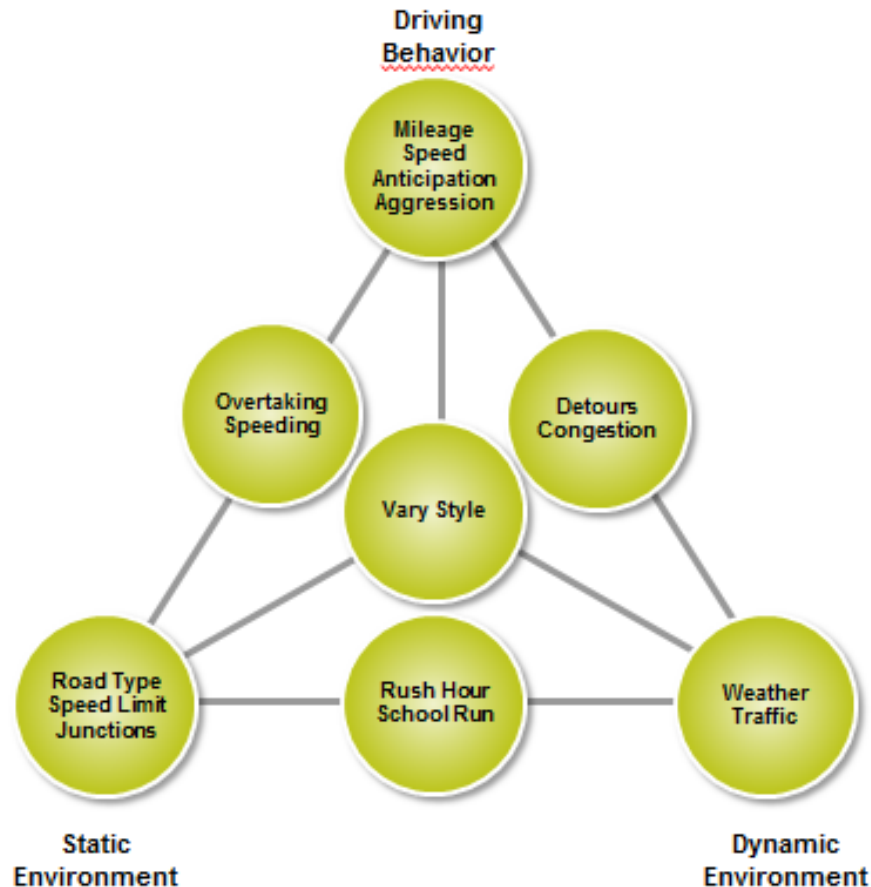
Granular data results in a better score faster



What data to collect?

The value of external data

External data allows behaviors to be put into context



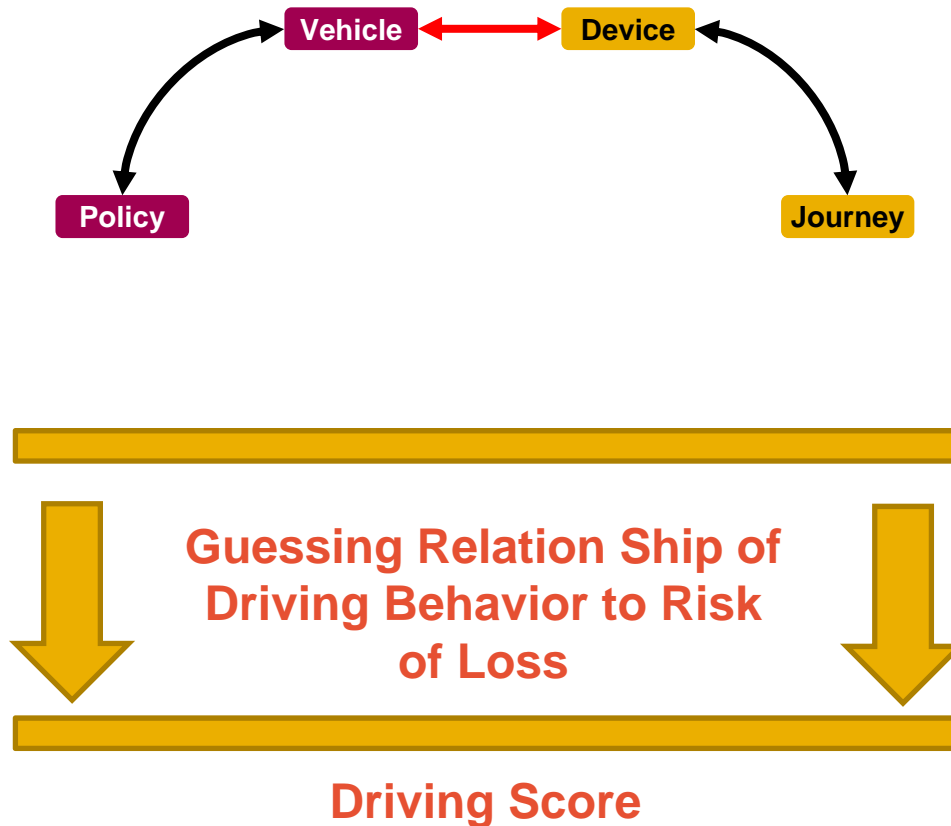
- Behaviors that are “safe” in good conditions aren’t necessarily “safe” in bad conditions
 - 60MPH on a highway is not the same as 60MPH in a neighborhood
 - 75MPH on a rainy or icy day is not the same as doing so on a nice day
- Good drivers are drivers who adjust to their environment
- Granular telematics and external data are required to do this

What data to collect?

The importance of insurance claims

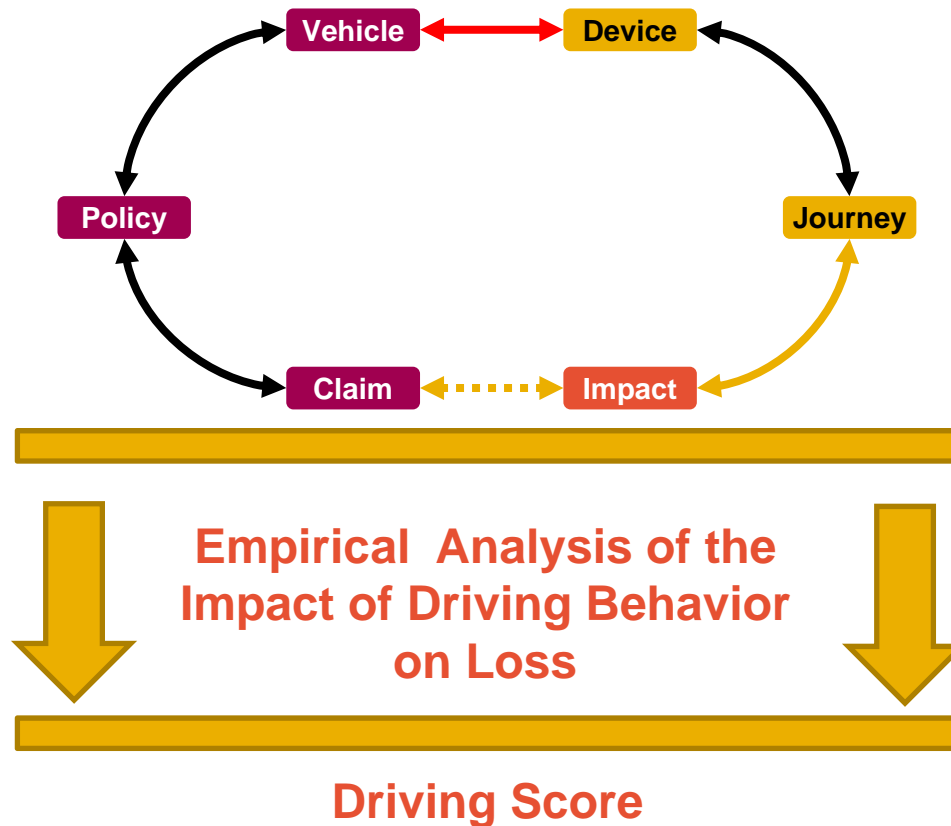
Why is it important to have claims data?

- There are some scores that are built without claims
 - Without claims, how can you quantify the relationship between driving behavior and risk of loss?



Why is it important to have claims data?

- Merging actual insurance claims to UBI data allows empirical analysis and quantification of the relationship between driving behavior and risk of loss; we can identify true causes of loss



Claim volumes

- Good News:
 - Typically participants in UBI program are risk adverse
- Bad News:
 - Need many exposures to have enough claims to build credible models



What data to collect?

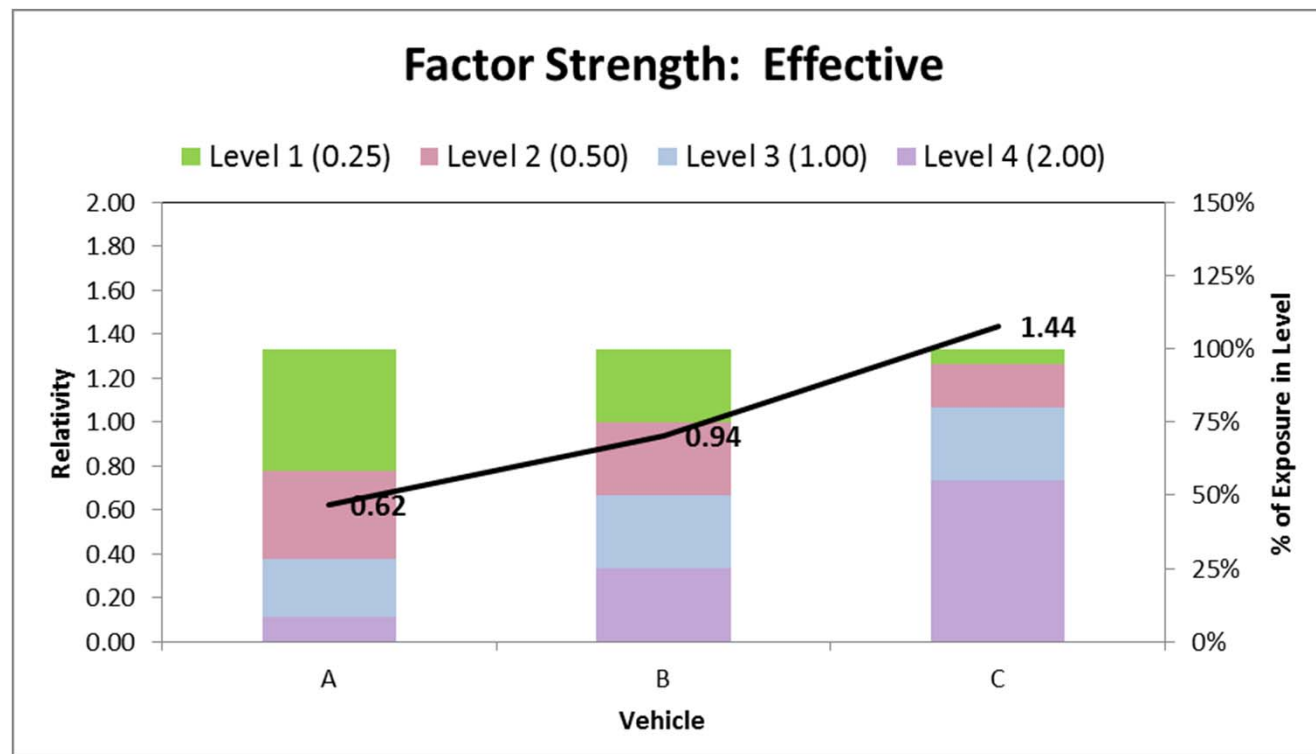
The value of collecting the right data

Multivariate analysis on UBI data is critical

- Getting complete data is only the first part of the solution
- The score should be built using multivariate analysis techniques. By doing so, the score
 - Won't cause double-counting
 - Will have maximum predictive power
 - Will be tailored to insurance use

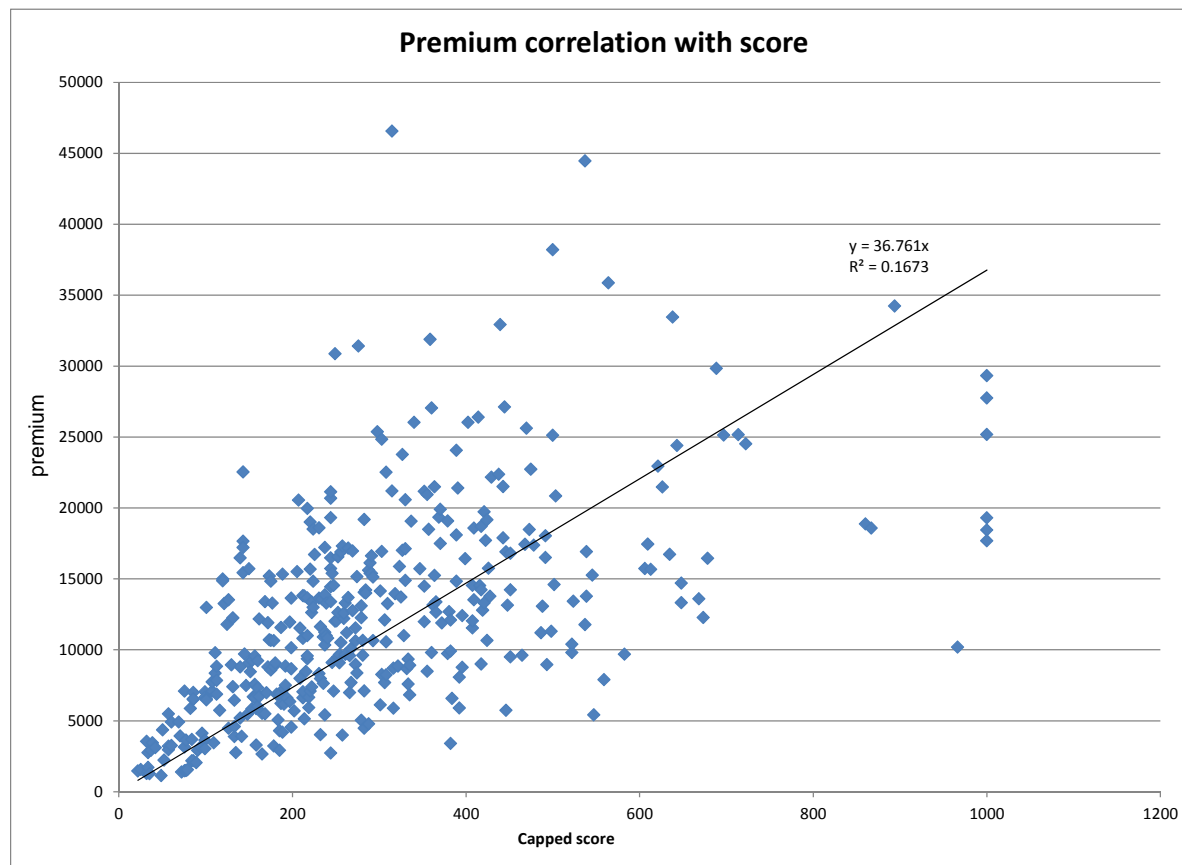
Granular data allows for searching for meaningful factors

- Are the factors strong?
 - Behavior must differentiate risk
 - Risky behavior should be more than an extremely rare event
 - Event not done equally for vehicles

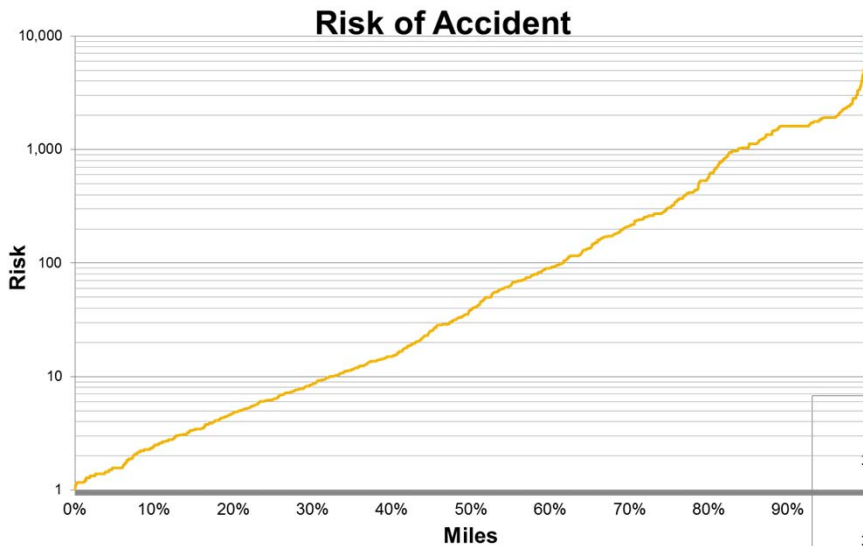


Granular data allows for searching for uncorrelated factors

- Are events “counting” the same thing as current rating factors?
 - If not, then we would see a high correlation between score and current premium

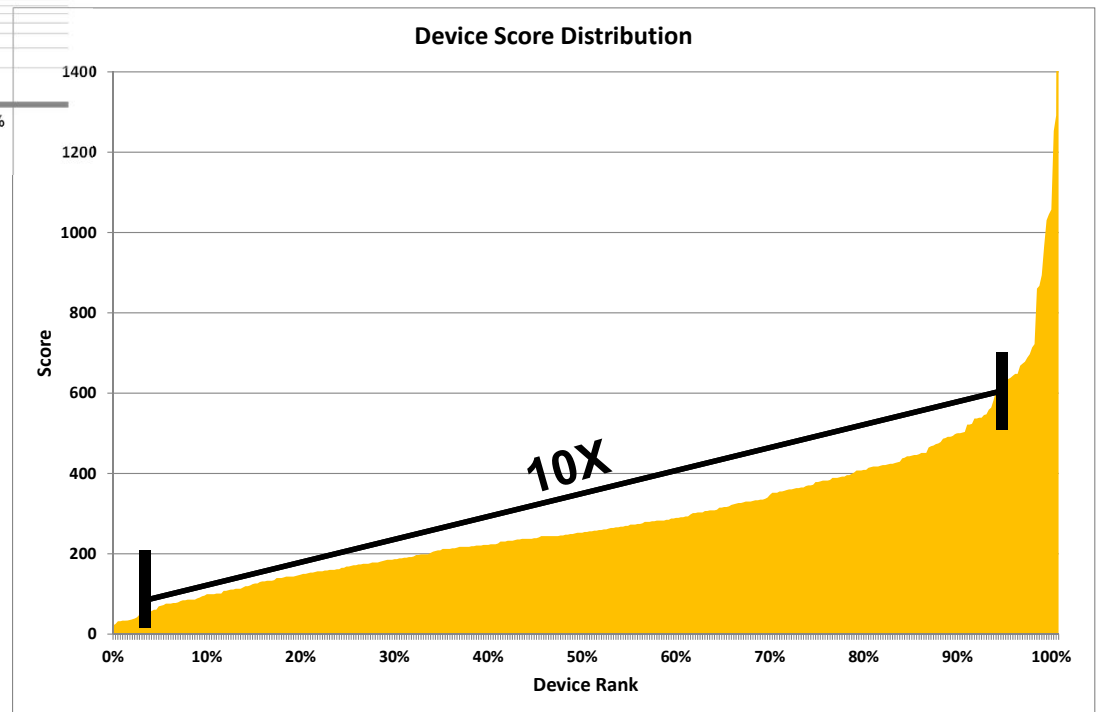


Scores built on UBI data are very predictive

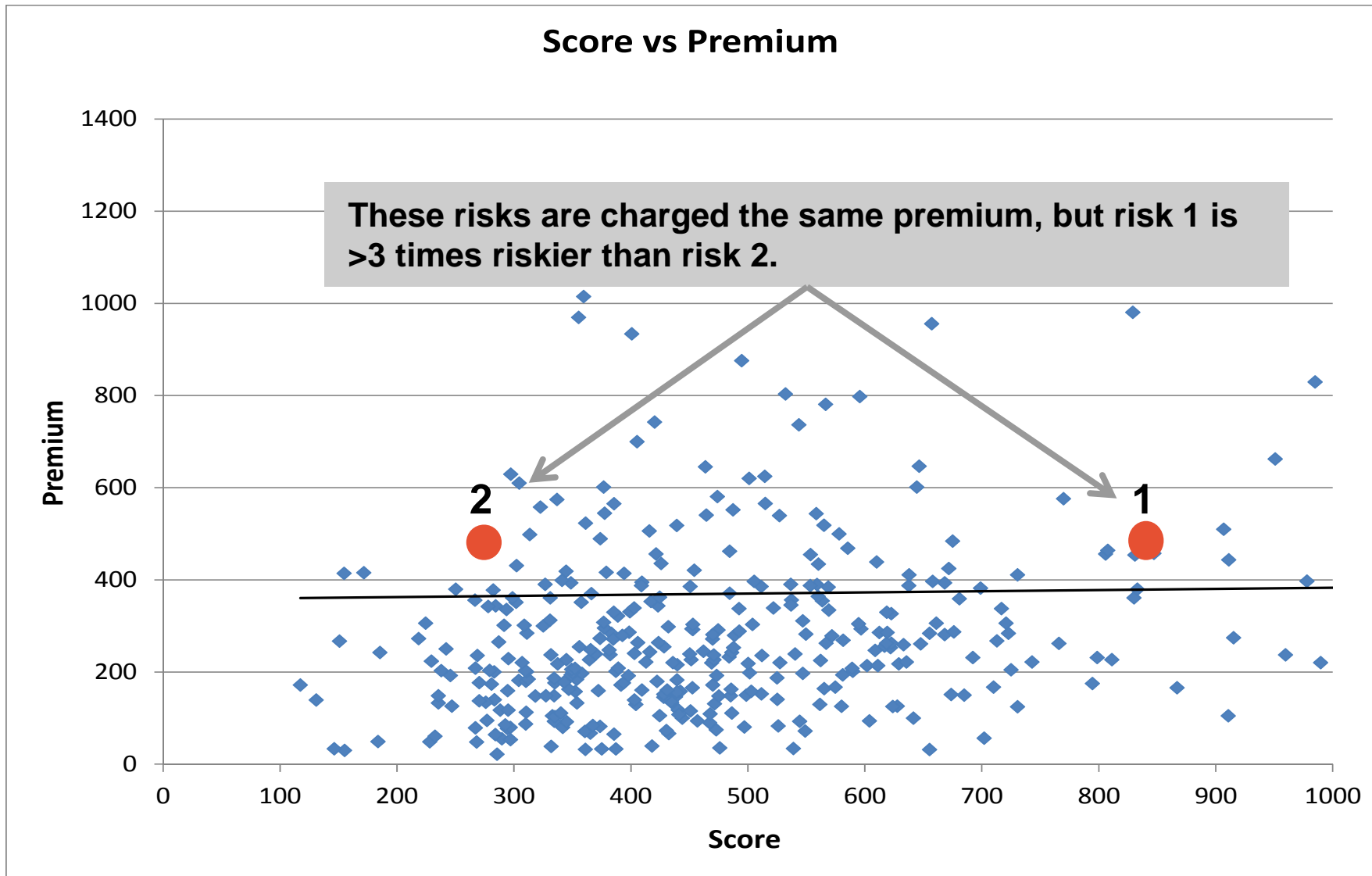


- Using our pooled data, our algorithm identifies certain “miles” as being 10,000 time riskier than others

- Aggregating miles at the vehicle level results in the shown scores
 - The highest decile of vehicles has an expected cost 10 times higher than that of the best decile



Scores can be “above and beyond” traditional factors



What are the typical data issues?

The challenges of using UBI data

Simple example UBI data for 2½ minute trip

TRIP:	1																		
DATE:	12-Jun																		
Time	MPH	Time	MPH	Time	MPH	Time	MPH	Time	MPH	Time	MPH	Time	MPH						
0:00:00	2	0:00:25	12	0:00:50	9	0:01:15	2	0:01:40	0	0:02:06	30								
0:00:01	2	0:00:26	11	0:00:51	12	0:01:16	0	0:01:41	0	0:02:07	32								
0:00:02	0	0:00:27	10	0:00:52	14	0:01:17	2	0:01:42	0	0:02:08	32								
0:00:03	0	0:00:28	9	0:00:53	15	0:01:18	5	0:01:43	0	0:02:09	33								
0:00:04	0	0:00:29	9	0:00:54	14	0:01:19	7	0:01:44	0	0:02:10	33								
0:00:05	2	0:00:30	9	0:00:55	12	0:01:20	9	0:01:46	0	0:02:11	34								
0:00:06	6	0:00:31	9	0:00:56	12	0:01:21	11	0:01:47	0	0:02:12	35								
0:00:07	7	0:00:32	10	0:00:57	11	0:01:22	13	0:01:48	0	0:02:13	35								
0:00:08	9	0:00:33	11	0:00:58	9	0:01:23	15	0:01:49	0	0:02:14	35								
0:00:09	9	0:00:34	12	0:00:59	8	0:01:24	17	0:01:50	0	0:02:15	35								
0:00:10	8	0:00:35	12	0:01:00	6	0:01:25	18	0:01:51	1	0:02:16	35								
0:00:11	8	0:00:36	14	0:01:01	5	0:01:26	19	0:01:52	7	0:02:17	33								
0:00:12	7	0:00:37	14	0:01:02	5	0:01:27	19	0:01:53	11	0:02:18	30								
0:00:13	7	0:00:38	15	0:01:03	5	0:01:28	17	0:01:54	12	0:02:19	28								
0:00:14	7	0:00:39	14	0:01:04	4	0:01:29	15	0:01:55	13	0:02:20	24								
0:00:15	7	0:00:40	12	0:01:05	4	0:01:30	14	0:01:56	13	0:02:21	21								
0:00:16	7	0:00:41	11	0:01:06	4	0:01:31	13	0:01:57	12	0:02:22	17								
0:00:17	8	0:00:42	10	0:01:07	4	0:01:32	11	0:01:58	12	0:02:23	14								
0:00:18	9	0:00:43	10	0:01:08	4	0:01:33	7	0:01:59	13	0:02:24	11								
0:00:19	12	0:00:44	9	0:01:09	4	0:01:34	3	0:02:00	15	0:02:25	7								
0:00:20	13	0:00:45	7	0:01:10	2	0:01:35	0	0:02:01	18	0:02:26	5								
0:00:21	14	0:00:46	7	0:01:11	2	0:01:36	0	0:02:02	20	0:02:27	3								
0:00:22	15	0:00:47	6	0:01:12	3	0:01:37	0	0:02:03	23	0:02:28	0								
0:00:23	15	0:00:48	6	0:01:13	4	0:01:38	0	0:02:04	26	0:02:29	0								
0:00:24	14	0:00:49	7	0:01:14	5	0:01:39	0	0:02:05	28	0:02:30	0								

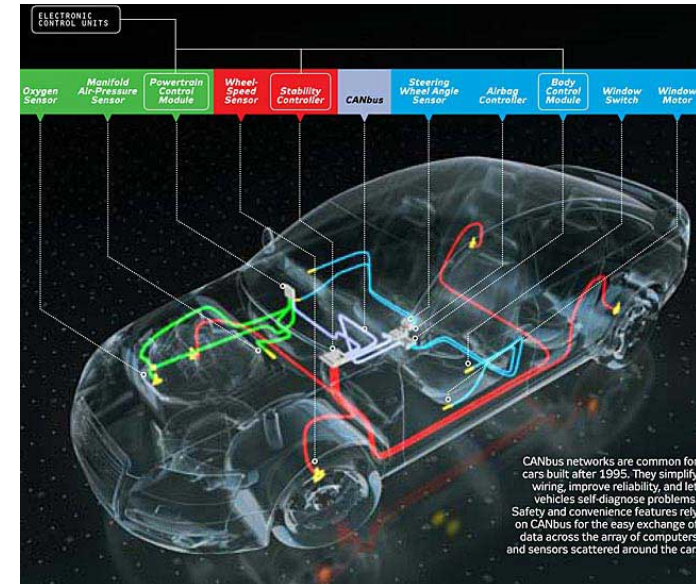
UBI data is different and exceptionally challenging

	Without UBI	With UBI
Update frequency	Semi-annual	Real time, trip, daily
Data quality	Renewal UW	Daily scrubbing
Variables	Dozens	Hundreds
Records per policy	Dozens	Millions
File size	Gigabytes	Terabytes [Petabytes?]

- What technology is needed to process this data?
- How do you clean/scrub this amount and type of data?
- What are other typical data related issues?

Understand the data through learning about the technology

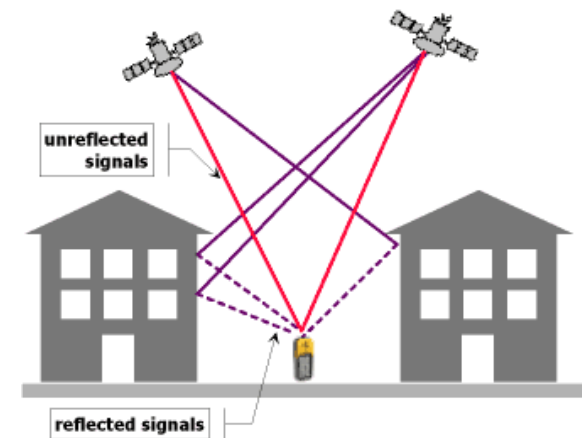
- Telematics device
 - Global Positioning System (GPS)
 - Accelerometer
- Vehicle systems and how the device interacts with them
 - On-Board Diagnostics (OBD)
 - Engine Control Unit (ECU)
 - Controller Area Network (CAN)
- Telematics Service Provider (TSP) processes for data collection and processing



Understanding the technology will also help you communicate issues with the TSP

Global Positioning System (GPS)

- Learn how it works
 - First time to fix
 - Heading
- Learn about common errors and how to identify them in the data
 - Signal propagation error (position jitter)
 - Ephemeris error
 - Clock error
 - Multipath
 - Dilution of precision
 - Space weather
- Learn about data scrubbing options
 - Comparing GPS and Vehicle Speed Sensor (VSS) data
 - Map matching



What are the typical data issues?

A new IT infrastructure

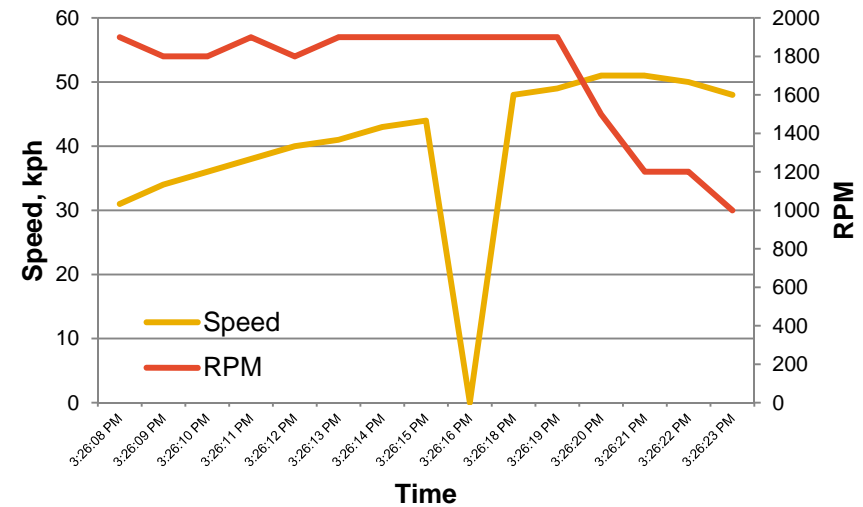
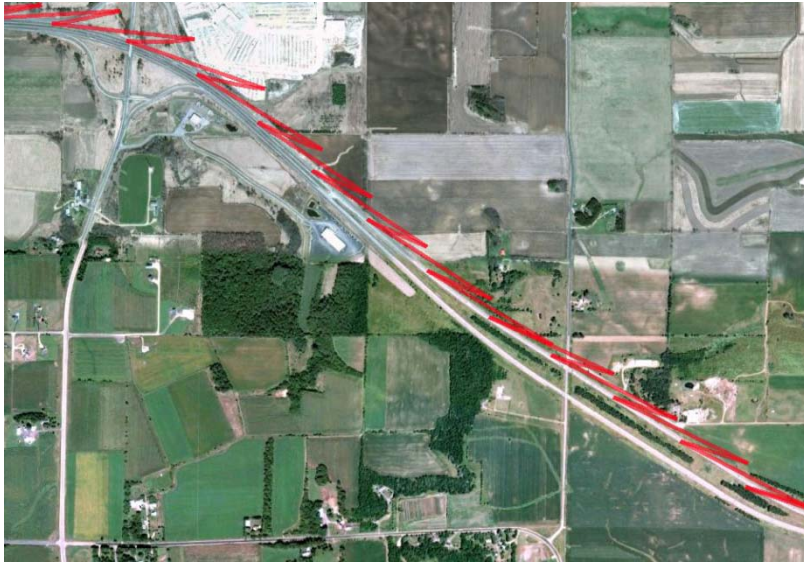
How do you process UBI data?

- Preparing UBI data for analysis is a long process; Requires many man hours to develop a streamlined process
 - Need to load constant inflow of data into main database
 - Need to clean and scrub
 - Need to merge external data
 - Need to identify patterns in the data, and program them into factors for your model
- Standard desktop database applications do not have enough horsepower to process this data
 - Powerful database servers are needed to manipulate data
 - Hardware and software must be considered
 - Powerful analytical tools are needed to extract knowledge from data

What are the typical data issues?

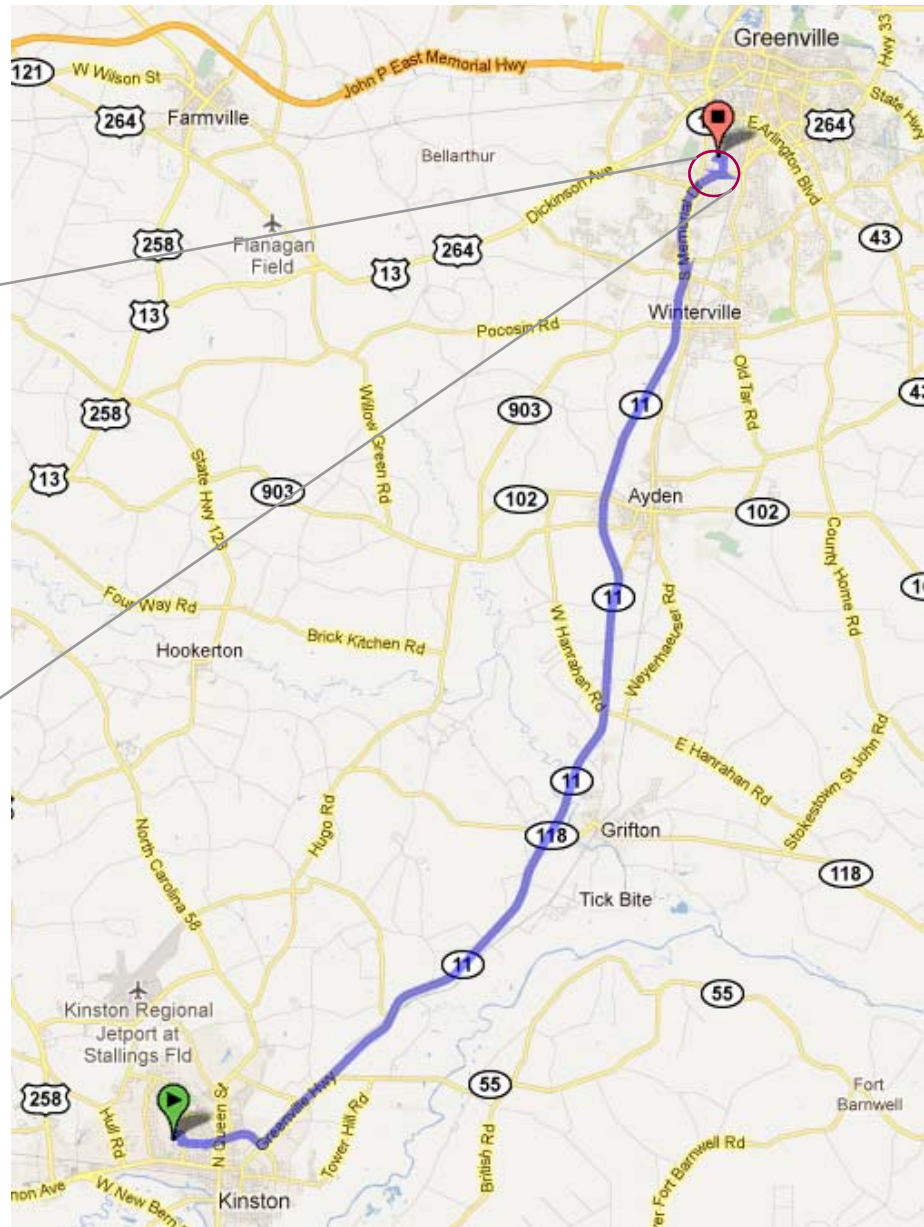
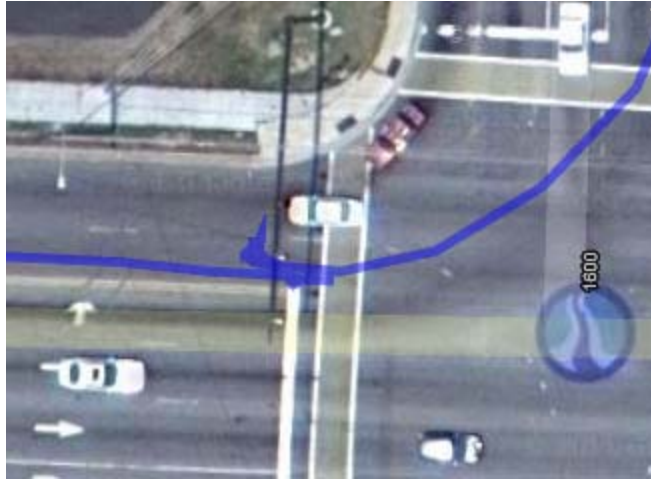
The challenges of scrubbing UBI data

How do you clean and scrub UBI data?

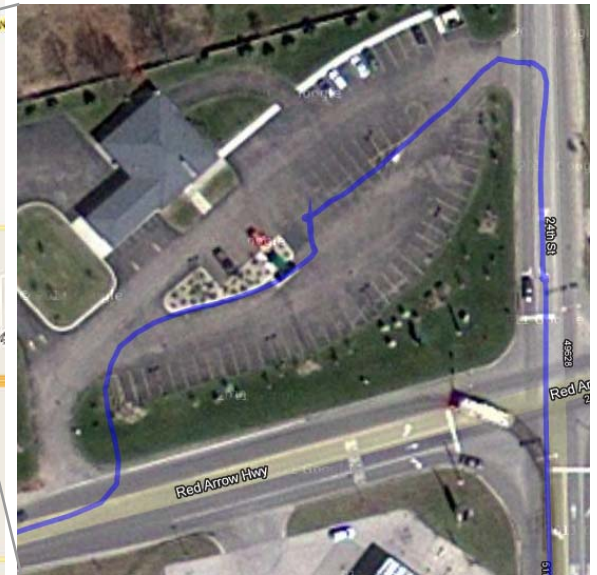
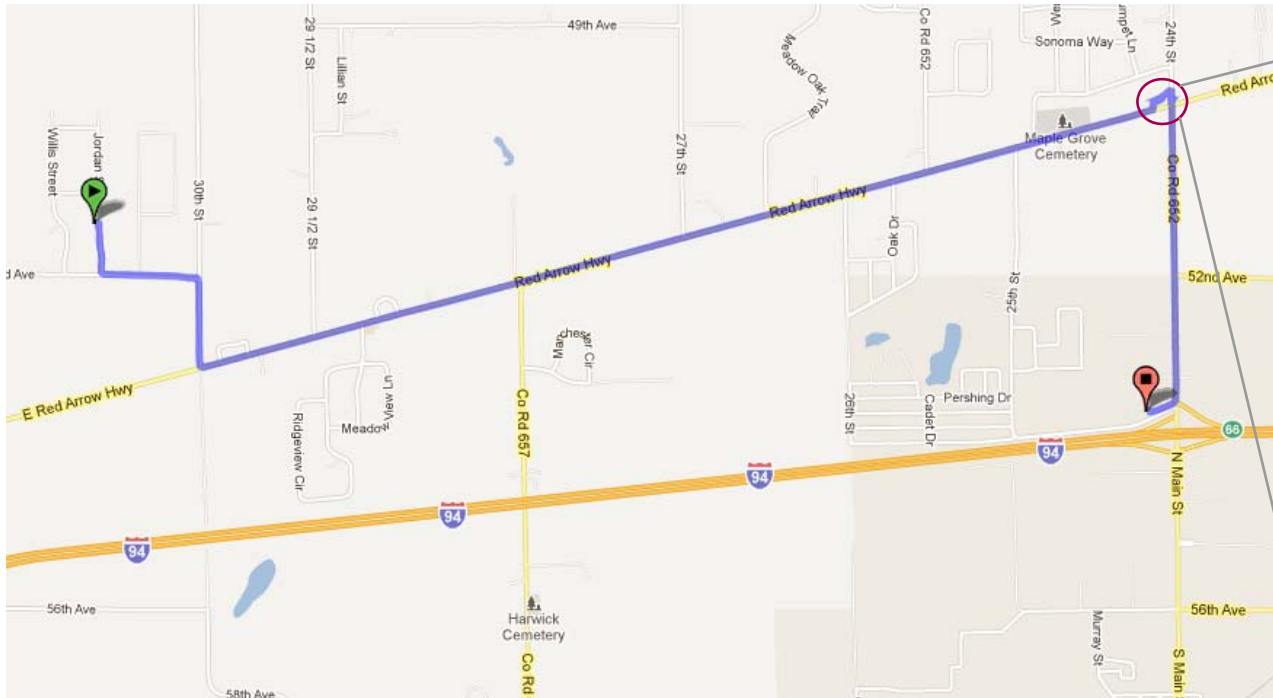


- Granular data facilitates data scrubbing
 - However, even with granular data, the task of data scrubbing is very difficult

Example journeys - 1



Example journeys - 2



Journey Data – Cleansing Checks

- Signal Skip
- Repeated time/journeys/events
- Missing minutes
- Gaps in trips
- Non-unique trips
- There are many others!

What are the typical data issues?

Other lessons learned from using UBI data

What are some other data related issues?

- Some data related challenges that we've encountered include:
 - Lossless compression is needed
 - Server capacity should be constantly monitored
 - Checks need to be created to ensure that no data is missing
 - All processes should be build to be scalable
 - Comprehensive compatible vehicle listing must be maintained
 - Process needed to ensure device was installed in enrolled vehicle
- Once a data issue is identified, how is it communicated:
 - To the TSP?
 - To the business area?
 - To the policyholder?

The right UBI data and for the future

Summary

Lesson to take with you

- Collect the right data
 - Collect granular data to facilitate cleansing and thorough analysis
 - Append external data to put driving behavior in the proper context
 - Obtain insurance policy and claim information to tailor score to insurance context
- Budget time to build the necessary processes
 - Be prepared to build a new IT infrastructure
 - Budget many, many, many hours to develop data scrubbing processes
 - Build checks and balances in your systems to monitor and anticipate issues; they will occur!