Catching the Bus

Prepared for CS7311
by Gentry Atkinson

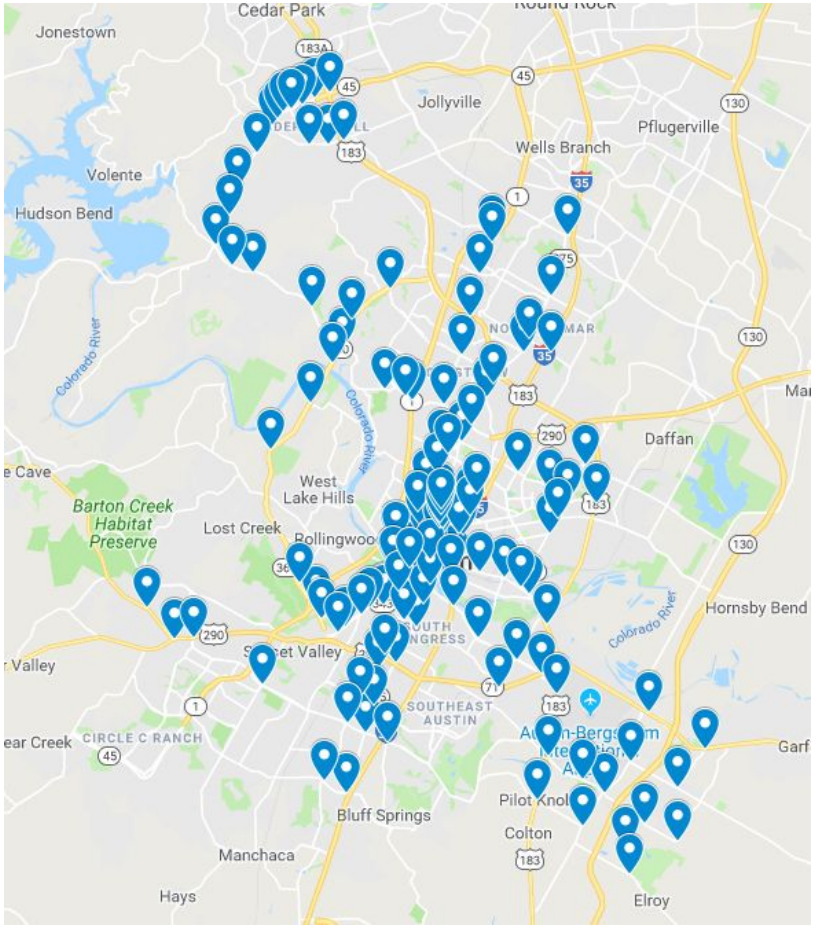
The rising STAR *of Texas*

Review of Data:

- "Hack the Traffic" dataset retrieved from: data.austintexas.gov
- 147 sensors around Austin log the presence of Bluetooth devices.
- Devices are identified with anonymized MAC addresses.
- Aggregated data gives speed and flow numbers.

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Sensors Locations:



TEXAS STATE UNIVERSITY
The rising STAR of Texas

Review of Problem:

- The city of Austin would like to know what percentage of traffic is using car pools and public transportation.
- The sensors have no way of distinguishing multi-occupant and single-occupant vehicles.
- I chose to focus on identifying buses within the data.

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Methodology:

- Calculate “walks” of devices which travel together across several sensors.
- Compare walks which occur on bus routes to the city-wide norms.
- Identify differences which may indicate that a higher percentage of a local population is on buses as compared to not.

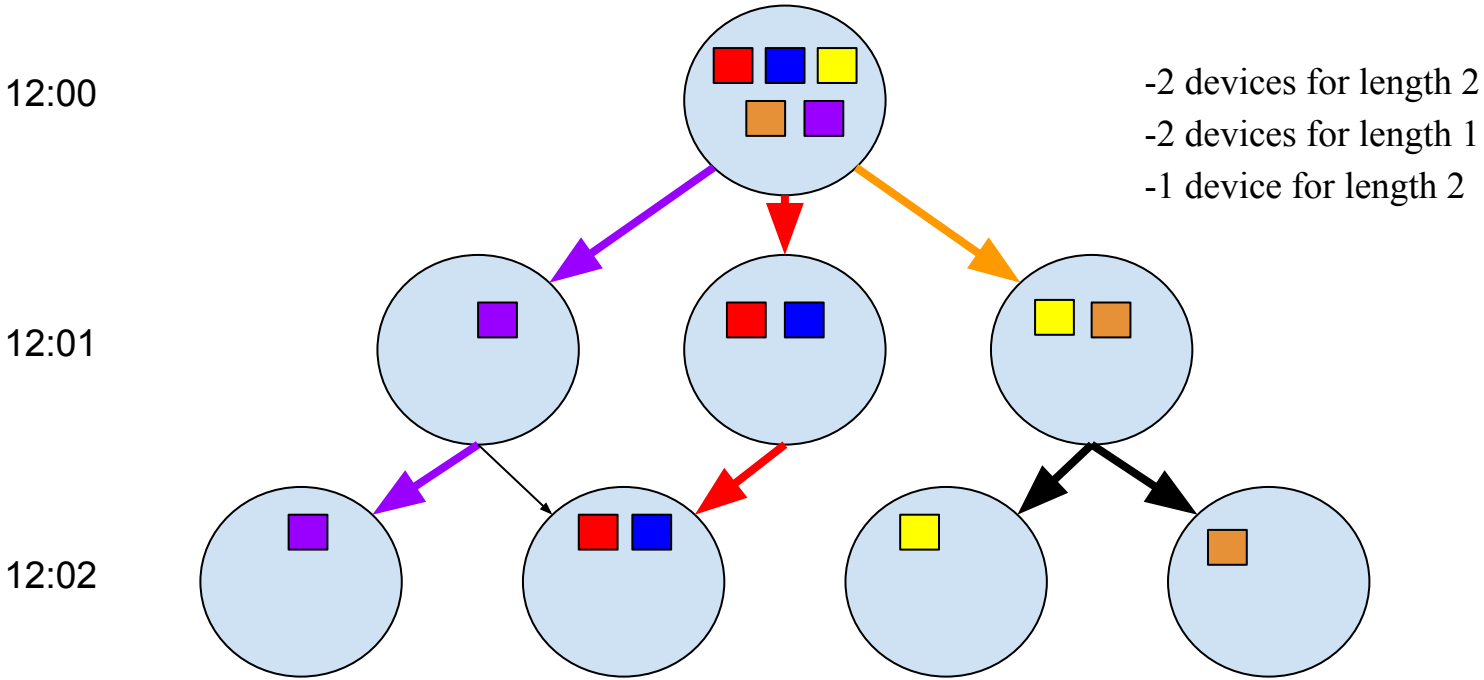
TEXAS STATE UNIVERSITY
The rising STAR of Texas

Calculating walks:

- Quantize the time values of of the data-points to account for “jitter” in their collection.
- Build a directed graph wherein each node is one sensor at one point in time.
- Visit each node and find the intersection of the set of devices present with devices at neighbors.
- Follow the subset of devices and record the distance they traveled together

TEXAS STATE UNIVERSITY
The rising STAR of Texas

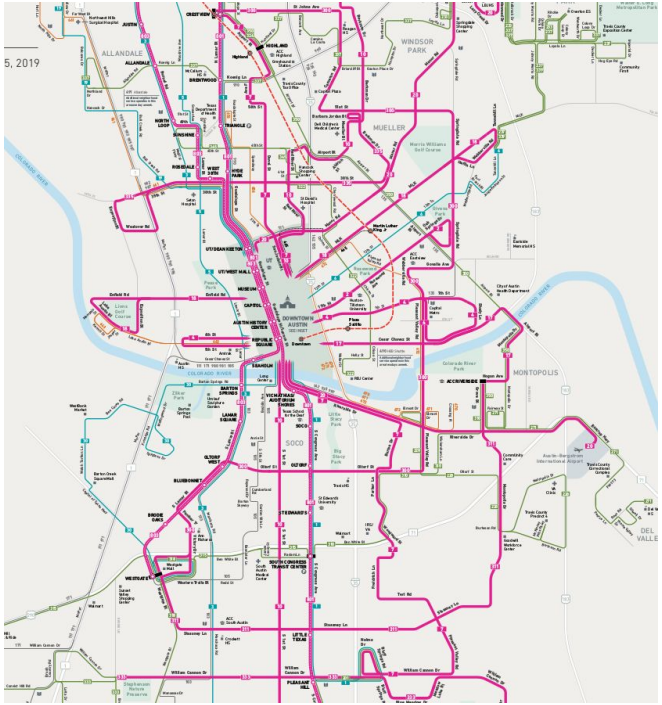
Example Walks:



TEXAS STATE UNIVERSITY
The rising STAR of Texas

Choosing Routes:

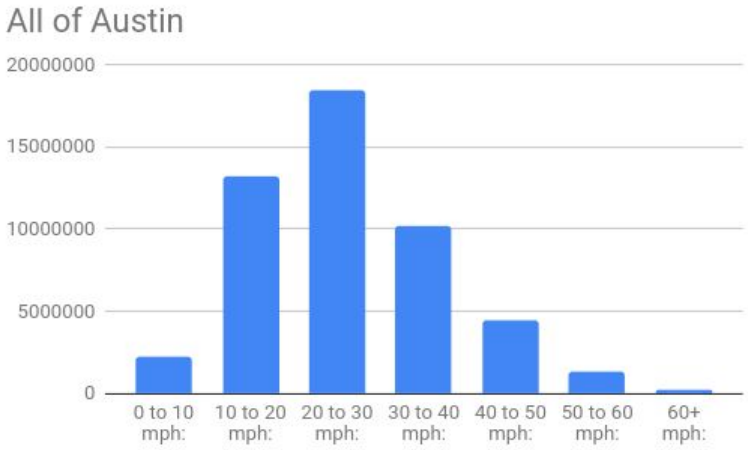
- Lamar: very dense collection of bus routes including double buses.
- Riverside: east-west travel is less common in Austin. Bus route and several sensors are present.

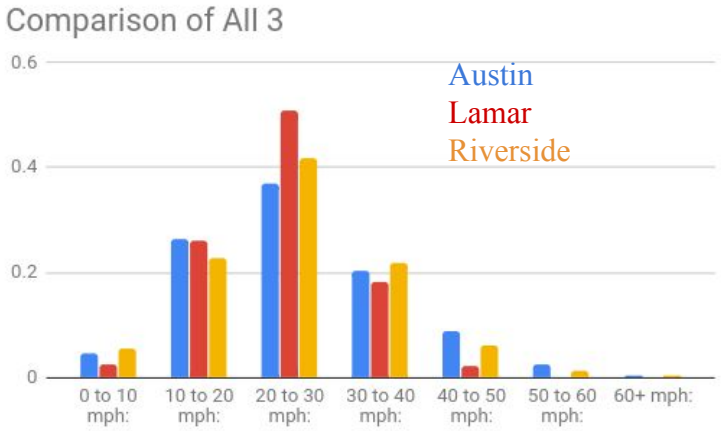


TEXAS STATE UNIVERSITY
The rising STAR of Texas

Comparison of Speeds:

Both Lamar and Riverside showed a similar distribution of speeds as compared to the city as a whole, with both skewing slightly slower.

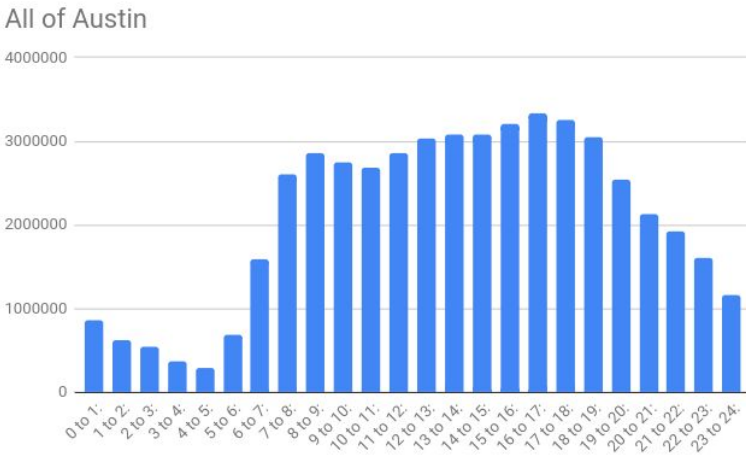


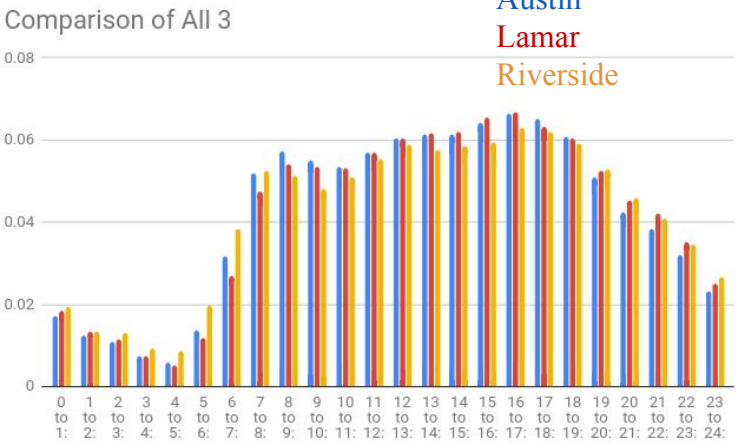


TEXAS STATE UNIVERSITY
The rising STAR of Texas

Comparison of Times of Travel:

Both Lamar and Riverside showed a similar distribution of speeds as compared to the city as a whole with Riverside having slightly more morning traffic and both having slightly more evening traffic.

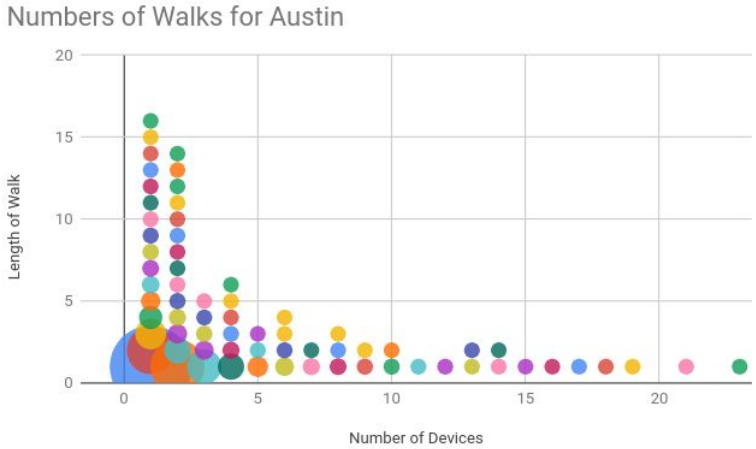


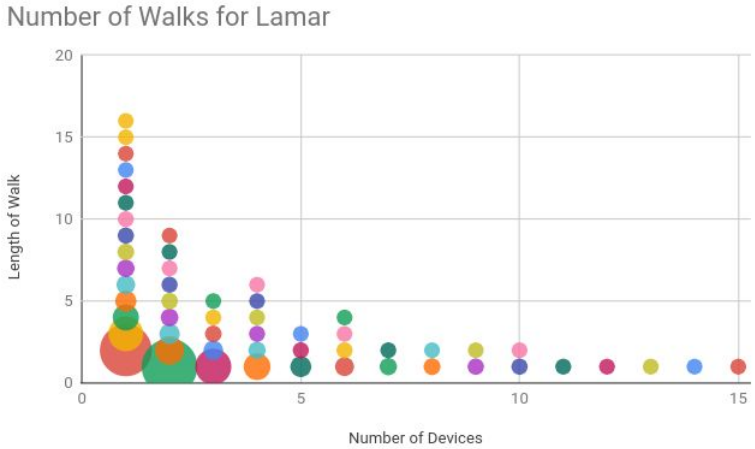


TEXAS STATE UNIVERSITY
The rising STAR of Texas

Comparison of Counts:

- Total Walks found: 50,121,162
- Total on Lamar: 12,279,128
- Similar distributions

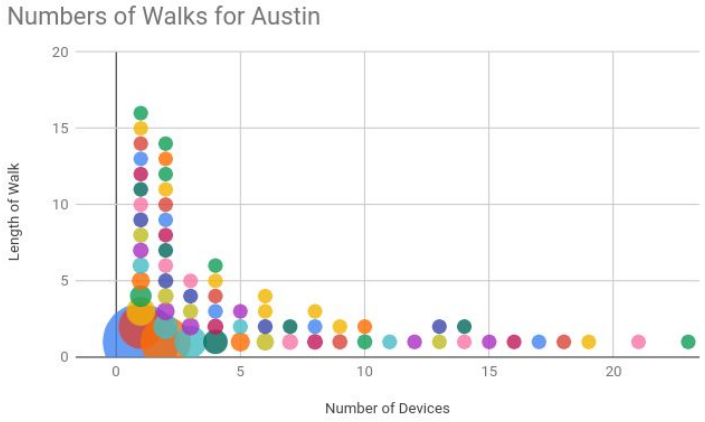


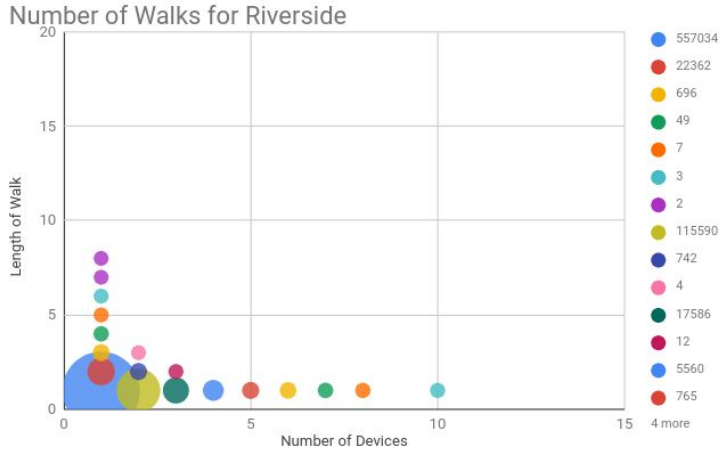


TEXAS STATE UNIVERSITY
The rising STAR of Texas

Comparison of Counts:

- Total Walks found: 50,121,162
- Total on Riverside: 721,027
- Riverside seems to skew towards lower device counts.





TEXAS STATE UNIVERSITY
The rising STAR of Texas

Review of Preliminary Results:

- The two bus routes are not immediately distinguishable from Austin as a whole.
- 17 times as many walks were found on Lamar as compared to Riverside.
- Lamar accounts for a quarter of the trips in the data.
- Short, low-device trips account for the vast majority of trips in all 3 samples. (1-device, length 1 makes up 56% of the total set)

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Considerations for Clustering:

- BIRCH was chosen for its ability to deal with noisy data.
- BIRCH also allows divides the data into clusters and sub-clusters, allowing a “granularity” to be chosen after the clusters are calculated.
- Walks with fewer than 3 devices or length 2 were excluded to prevent a single huge cluster from dominating the others.

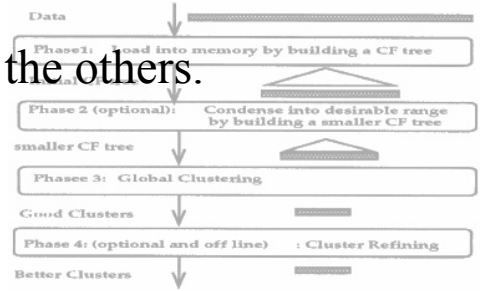
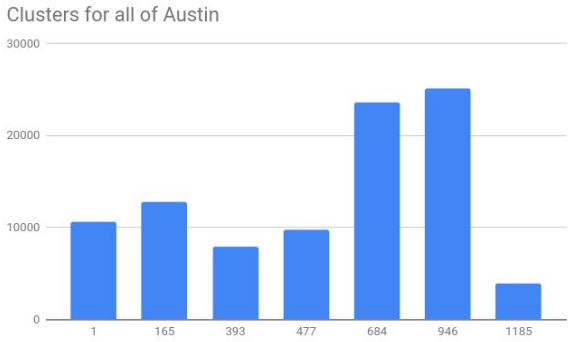


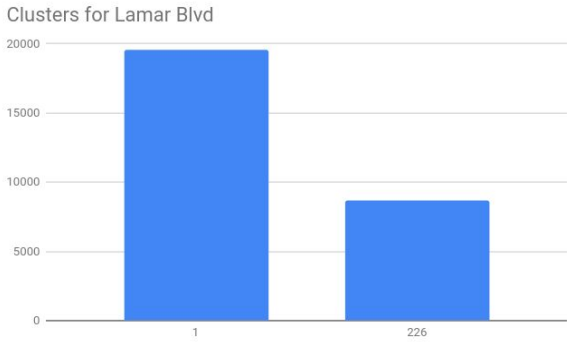
Figure 1: *BIRCH Overview*

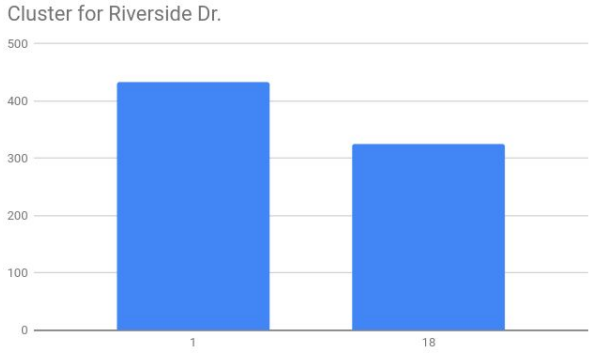
TEXAS STATE UNIVERSITY
The rising STAR of Texas

Results from Clustering:



Both Lamar and Riverside produce fewer, mor dominant cluster.

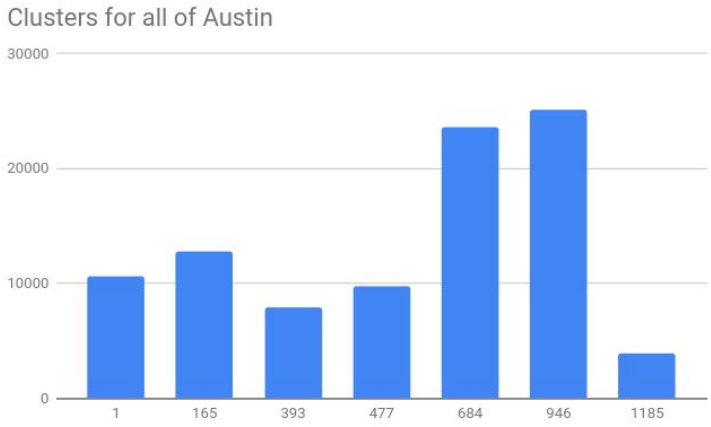


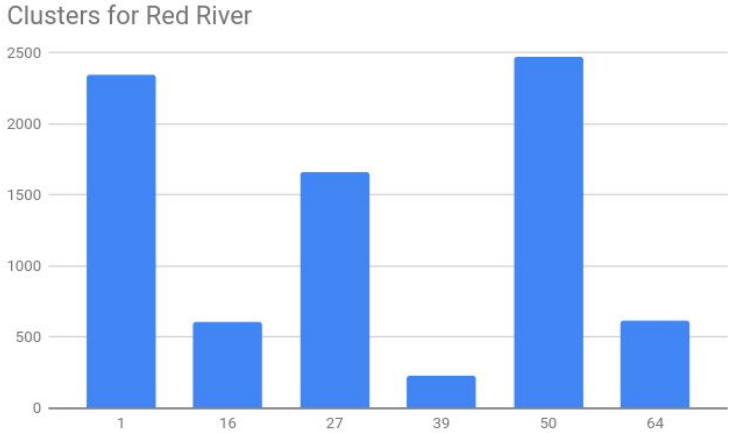


TEXAS STATE UNIVERSITY
The rising STAR of Texas

Results from Clustering:

- To test the theory that collections of fewer trips naturally produce fewer clusters, I added Red River into the mix.
- Red River (a street without many buses) produces 675,731 that divide into 6 top-level clusters.





TEXAS STATE UNIVERSITY
The rising STAR of Texas

Analysis of Clustering:

- The bus routes appear to produce fewer clusters indicating looser data or more tightly focused data.
- Additional comparisons will confirm whether or not this relationship holds.
- Other streets with many sensors but few bus routes are hard to come by.

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Conclusions:

- There is a good chance that analysis can indicate which streets have heavier bus traffic.
- There is not indication that a specific data point can be identified as single- or mass-transit.
- There is not indication that the specific percentage of traffic represented by mass-transit can be calculated.

TEXAS STATE UNIVERSITY
The rising STAR of Texas

Questions?

View the full project at: https://github.com/gentry-atkinson/traffic_analysis

TEXAS STATE UNIVERSITY
The rising STAR of Texas