Project 3: Option 6
Orlando Traffic Analysis Zones

Randy Tejeda
University of Illinois At Chicago
rtejed2@uic.edu

Francisco Gonzalez
University of Illinois At Chicago
fgonza21@uic.edu

Mario Becerra
University of Illinois At Chicago
mbecer3@uic.edu

Abstract
This data visualization looks at the trip dataset for the month of April 2014 provided by Transportation Research Board. The analysis includes a map of the Orlando Metropolitan Area, divided into Traffic Analysis Zones. Each trip definition is described, in part, by its origin and destination zones which correspond to a Traffic Analysis Zone. The focus of this data visualization is to allow a user to view specific trip information. Using several filters, the user can choose to display trip information on both a map and adjacency matrix. Additionally, the top 25 trip destinations are available. This information can be used to analyze traffic patterns in the Orlando Metropolitan Area.

Introduction
The Transportation Research Board (TRB) hosted a competition to analyze and visualize an Orlando transportation dataset. Our group created three visualizations to provide complete and detailed representations of the dataset. The representations aid the user in finding specific information regarding individual zones and larger macro-patterns in the data.

Analysis Tasks

The task was to create an interactive visualization that displays trip information based on the user's filter preferences. We experimented with exposing the raw categories provided by the data and grouping some of the data into logical subsets.

The dataset was comprised of thirty, one for each day, csv files describing the transportation in the Orlando Metropolitan Area. Each csv file has over 170,000 rows. Each row represented a group of trips taken from a origin zone to a destination zone, as well as time and subscriber information.

The zone data was represented as Traffic Analysis Zones (TAZ) that have little or no relation to any other private or governmental zones such as census, city, and county zoning. TAZs are common with transportation planning models. Using the TAZ data made it difficult to correlate the zones with external information such as area socioeconomic data or population statistics.

Each trip was further categorized as follows:

- **Subscriber Class:** Describes the driver.
  - Short Term Visitor
  - Long Term Visitor
  - Home Worker
  - InBound Commuter
  - OutBound Commuter
  - Resident Worker
- **Trip Purpose:** Describes the trip.
  - Home to Work
  - Work to Home
  - Home to Other
  - Other to Home
  - Work to Other
  - Other to Work
  - Home to Home
  - Work to Work
  - Other to Other
- **Time of the Day:** Segmented hourly

The TAZ shape files were additionally included allowing a rendering of the relevant zones.

To efficiently illustrate this complex dataset, we decided to created three separate visualizations each showing a different perspective on the dataset.

Visualization Details

The first visualization is a geospatial zoning map depicting the boundaries of each zone. Clicking on any zone will show trips associated with the clicked zone meeting the filter criteria chosen by the user. Filters, can be done on the purpose of the trip, time of day, and day. The trips are color coded to indicate the relationship between the two zones. A red zone, indicates that it has a net of trips originating in the clicked zone and terminating in the red zone. A blue zone, indicates that it has a net of trips originating in the blue zone and terminating in the clicked zone. This visualization allows for a detailed look into a specific zone's relationship with other connected zones. Moreover, the visualization allows for a user to use the regions spatial/distance relationships to see what zones are proximally related or related regards of distance.

The second visualization is an adjacency matrix. The matrix allows for the user to specifically look at any category and view the data in a matrix form. The

matrix allows for the data to be filtered by day, hour, trip purpose and description. This visualization provides an overall look into how zones can change dependent on the trip's purpose or what type of trip driver. Furthermore, the matrix allows the user to see what source/destination zones are related in regards to a specific category. If a user is more familiar with zone IDs, this visualization easily allows the user to view trip information based on the TAZ ID.

The third visualization is a bar chart representing the top 25 most traveled destination zones. This allows for an overall view into the busiest areas in the region. The user can specify the day for which to display the information. Hovering over a specific bar in the chart will indicated the total count for that day's trips terminating in the indicated zone.

Expert Feedback

Our case expert was Dr. Sybil Derrible. Dr. Derrible is an Assistant Professor of Sustainable Infrastructure Systems at University of Illinois at Chicago.

Dr. Derrible provided specific insights into what these visualizations should be focusing on. Due to the enormity of the data, we were trying to represent macro views that incorporate all the data with each feature. Dr. Derrible suggested that we should aggregate some of the data to provide larger categorical relationships.

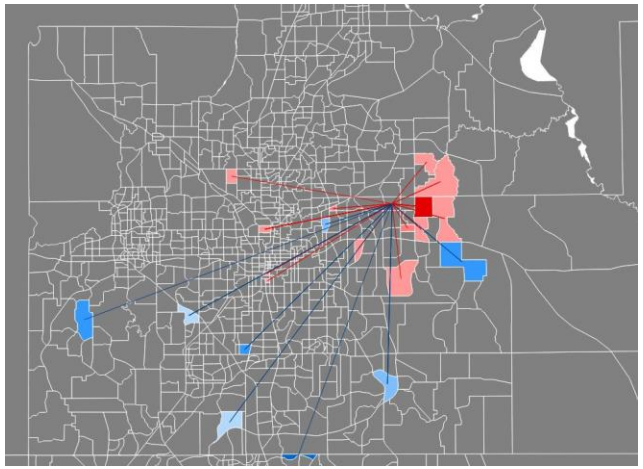
Dr. Derrible also suggested that we try to make correlations to the destinations or sources using socioeconomic data or geographical data.

Furthermore, he explained the background regarding the competition which was extremely helpful because it gave us an insight into how the data was collected and eventually represented. The sponsoring corporation is AirSage. AirSage uses cellular tower information to track and record survey information giving a geographical element to data. The TAZ provided was an aggregation of this cellular data with the user inputting the data describing the trip, such as trip purpose.

Lastly, Dr. Derrible insisted that it was better to show a smaller subset of the data if the larger subset would confuse the user. The user typically spends under a minute on a visualization so it is imperative to efficiently and rapidly disseminate the information you want known.

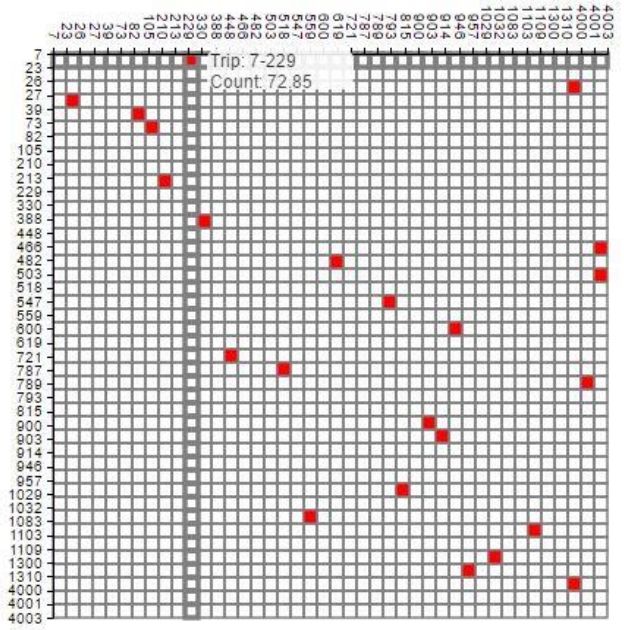
We took the Dr. Derrible's suggestions and merged many of the categories, such as hours, into larger categories allowing for high leveler comparisons. We, also, created the bar chart to show some basic data analysis and some insight into the larger transportation motifs.

Case Study

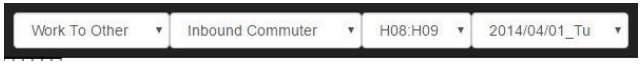


The map above is a filtered using the following categories:

The map above shows trips, represented by lines, associated with the clicked zone, the zone from which the lines originate. Red indicates trips originating in the red zone, blue indicates trips terminating in the clicked zone. The user can click on different zones and select different filter data to see specific trip information of interest.



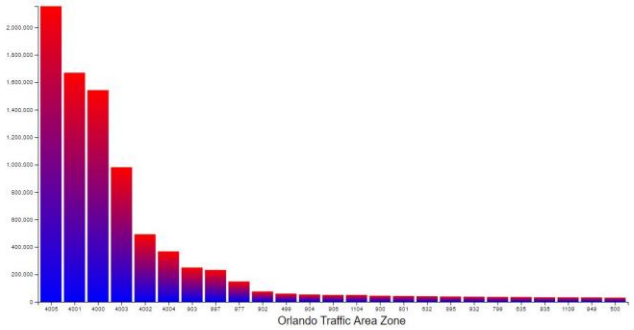
The adjacency matrix above is a filtered using the following categories:



As the user changes any of the categories, the matrix is redrawn showing only the zones that appear in a trip. The rows are source zones while the columns are destination zones.

The matrix also allows the user to hover over any cell and view the statistics regarding that trip.

The matrix gives a wide perspective regarding a zone with particular categories.



The bar chart displays the most traveled destination for the date specified by the user. Because the outer zones, zones 4000-4005, are so much larger than relative to other zones, they are always the top six destination zones. We chose the top 25 zones to show data for the smaller zones.

Conclusions and Future Work

The dataset was not easy to work with. The TAZ standard does not easily lend itself to comparison with other types of geographical zoning standards. Furthermore, it is near impossible to combine zones to aggregate the data.

The magnitude of the data was another issue. Representing large cross sections of the zones can cause the browser to crash due to lack of memory. Moreover, some matrices can load slowly due to the number of rows/columns having to be searched and rendered. As stated in the introduction, one day of trips can include over 170,000 individual trips.

As with many datasets, this one had data integrity and missing values issues. For example, the first day of trip does not have any data for the beginning five hours. Additionally, the Other category is a giant umbrella that can encompass a multitude of trip purposes and information. More comparisons could have been made if the Other category could have been split into smaller subsets.

Meeting with the expert provided many critical suggestions and recommendations. Unfortunately, we met later than expected. We were not able to incorporate a lot of the expert's suggestions because of time and the dataset's inability to be extrapolated and compared to other common forms of geographical and socioeconomic statistics.

Originally, we had an additional map that would iterate through each trip provided by the data and draw an arc between the two zones that made up the trip. Because of the large amounts of data that made this visualization cumbersome, we decided to remove efforts from this visualization and focus on a data analysis visualization.

Trying to do simple data manipulation proved impossible with such a large dataset. For example, after reading in the csv file and trying to aggregate data in the array or sorting the array often slowed down the browser or made it crash. For the bar chart, we had to do the data manipulation using excel and create new files that could easily be read by d3. This limited the data analysis we could visualize.

I recently learned about NoSQL databases that are often used for fast online queries. They are faster

because the NoSQL databases are more flexible than relational databases. Although data integrity is sacrificed for speed, the risk is minimal. In the future, we could have investigated importing the data into a NoSQL databased and querying the data from d3.

Additionally, for future works, we could try to shrink the data as much as possible. There are some categories that provide very large matrices that become increasingly more difficult to interpret. We would try to aggregate as many zones as possible to speed up computation and to show larger relationships between larger areas. If the GPS data was given, we could have included specific locations located in specific zones allowing us to speculate why an individual was going to that zone.

Related Works

1. Picozzi, Matteo, Nervo Verdezoto, Matti Pouke, Jarkko Vatjus-Anttila, and Aaron John Quigley. "Traffic visualization-applying information visualization techniques to enhance traffic planning." In GRAPP 2013 IVAPP 2013-Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications. SciTePress, 2013.
2. Gora, Paweł. "Traffic Simulation Framework-a cellular automaton-based tool for simulating and investigating real city traffic." Recent Advances in Intelligent Information Systems (2009): 641-653.
3. Neo4j from JavaScript: https://neo4j.com/developer/javascript/.