Uber vs. Taxis: Event detection and differentiation in New York City

Grant McKenzie[1], Carlos Baéz[2]

[1] Department of Geographical Sciences, University of Maryland, College Park, USA
[2] Department of Geography, University of California, Santa Barbara, USA
Email: gmck@umd.edu; carlos.baez@geog.ucsb.edu

Abstract

The recent rise of ride-sourcing services such as Uber have significantly changed the transportation landscape. This work takes a first step in differentiating Uber and taxi transportation methods through events attended by their passengers. Using a sample of Uber and taxi pick-up times and locations in New York City, we show that events can be detected within each platform. Through identification of a select few of these events, this work takes a preliminary step in showing that there is a difference in the types of events that are attended by Uber users and taxi passengers.

1. Introduction

Historically, taxicab companies have controlled the largest share of the *for-hire vehicle* (FHV) market in the United States. Over the past few years, however, alternative transportation options have arisen such as *Uber* and *Lyft* that rely on the use of online-enabled platforms to connect passengers with drivers. Together with others, these types of ride-sourcing companies, often called *Transportation Network Companies*(TNC), have significantly disrupted the traditional transport model, namely taxi service. By some accounts (Certify, 2015), TNCs now account for 46% of some U.S.-based FHV markets. This dramatic shift in the means of transportation has spurred a lot of research and discussion on its impact and significance (NRC-TRB, 2015; Hall et al., 2015). From a spatiotemporal research perspective, this shift has also lead to some important questions related to the differences between these services as well as the people that use them.

This short paper presents a *first step* in exploring the differences between traditional taxi services and TNCs as described through events[1] in New York City. **Specifically, this work addresses the following research questions**.

- Is it possible to detect events based on passenger pick-up times and locations in publicly available Uber and taxi data?
- Do events detected in the Uber dataset differ from those detected in the taxi dataset?
- Do these findings support existing research showing that there are differences between TNC users and taxi riders? We approach this question through identifying a select sample of detected events.

As stated in this last question, existing work in this area indicates that there are differences in the demographics of taxi and TNC passengers. Specifically, TNC user surveys suggest that, relative to taxi users, TNC passengers are younger and posses a higher average level of education (Rayle et al., 2016). Our work continues on this

[1] See work by Worboys (2005) in discussing and defining events.

thread by exploring differences in the types of events that are attended by these two groups of passengers. Furthermore, this research builds off of work by Zhang et al. (2015) on detecting events in Chinese taxi data, although we take it several steps further in comparing taxi-based events with those discovered in TNC data.

2. Event Detection

Data for this work was accessed via the New York City Taxi & Limousine Commission. In total, 80,295,320 Yellow taxi pick-up locations[2] and 4,534,327 Uber pick-up locations[3] were used for this research (drop-off locations are not available for the Uber data). The data is spatially bounded by the extent of New York City and temporally bounded between April 1 and September 30, 2014. The attributes used in this work include geospatial coordinates and timestamps for passenger pick-ups.

For each of the two datasets, the following analysis took place. The timestamps for each pick-up were rounded to the nearest hour and aggregated to counts by intersecting with the New York city census tract spatial data from 2014. The mean and standard deviation for the number of pick-ups per census tract, aggregated to the hour of a typical week (24 values) were calculated. This produced mean pick-ups and standard deviation values for 168 hours in a week in 2,162 census tracts. Setting a minimum threshold mean of 10 pick-ups per hour, per census tract significantly reduced this value to 114 census tracts and 152 hours of the week.

Events were detected in each dataset by comparing the number of pick-ups on any given day, time and census tract with the amount of pick-ups typical for that hour of the week (mean count). An *event* was recorded if the number of pick-ups was above three standard deviations from the mean. Using this approach, 485 events were discovered in the Uber dataset and 2,671 in the taxi data.

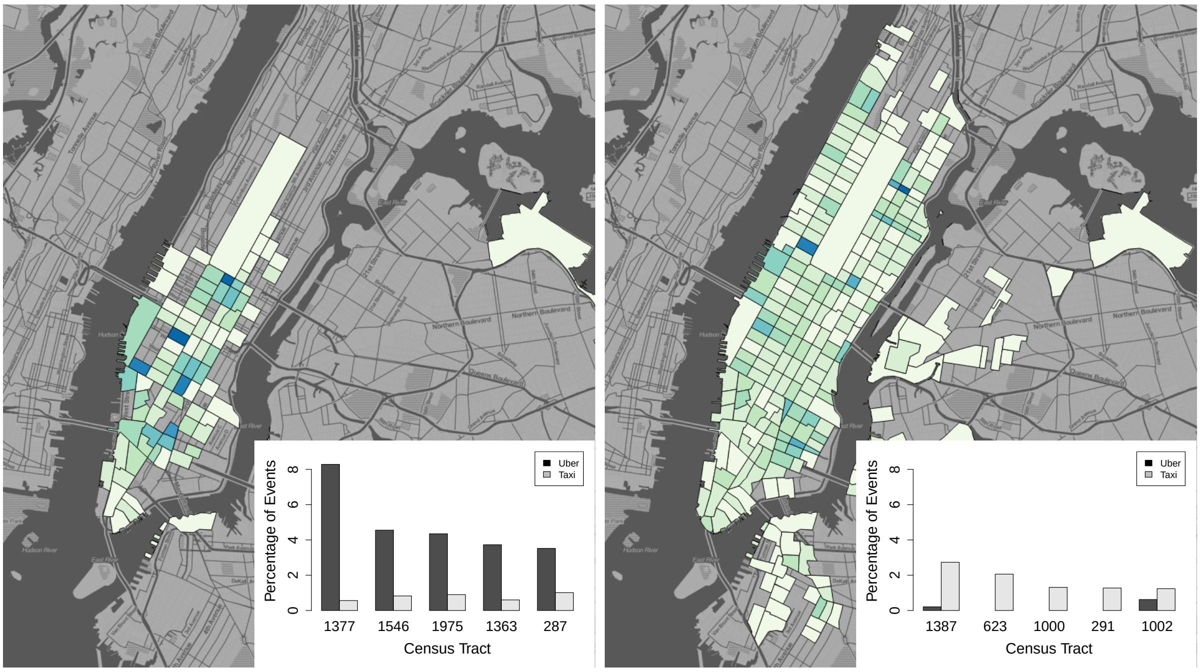
3. Event Differentiation

Given the events detected in both datasets, the next step was to identify events that were detected in both datasets and those that were specific to the taxi or Uber data. Setting a temporal buffer of two hours before and after an Uber-identified event, we searched for events in the taxi data that occurred in the same census tract within this specified temporal window. This resulted in 17 events identified in both the taxi and Uber datasets, meaning the majority of events were identified exclusively in the taxi or Uber data and not in both. Figure 1 depicts detected events through two chloropleth maps (quantile breaks) ranging from a high (blue) to low (light green) number of events for both Uber and taxi data. The inset bar plots show the top five census tracts by percentage of overall events detected for each mode of transportation. Notably, each dataset's highest event counts occurred in different census tracts. This suggests that certain types of events are aligned with Uber users while others lend themselves to taxi passengers. The difference in census tracts also suggests that there may be regional influences, although further investigation is beyond the scope of this short paper.

Given these common and mode-specific events, we manually investigated a number of the events based on their location and temporal parameters. Using application pro-

[2] http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

[3] https://github.com/fivethirtyeight/uber-tlc-foil-response



(a) Uber

(b) Taxi

Figure 1: Choropleth map showing number of events by census tract normalized by area. Inset bar plots show top 5 census tracts by percentage of overall events detected for that type (Uber or taxi). Base map by Stamen maps.

gramming interfaces (APIs) such as *Eventful*[4] along with venue specific websites (e.g., Madison Square Gardens), we were able to identify a number of events which are shown as examples in the following sections.

3.1 Taxi-specific Events

A large number of events were identified in the taxi dataset that did not appear in the Uber data. Examples of these types of events are *Summer by the Sea* on June 9 at Rockefeller Center and *David Gray* on June 23 at Madison Square Gardens. *Summer by the Sea* is an annual celebrity chefs' tribute to city meals-on-wheels while *David Gray* is an adult-contemporary musician who performed a live concert. Arguably, both of these events appeal to an older age demographic rather than young adults or teenagers. While these are only two examples, it is interesting to note that these events were only detected in the taxi data.

3.2 Uber-specific Events

Similar to the taxi-specific events, a number of Uber identified events could not be matched to the taxi data. For example, *NBC Upfront* on May 12 at Jacob Javits Convention Center and *Eric Prydz* on September 27 at Madison Square Gardens were identified in the Uber data, but not the taxi data. *NBC Upfront* is a large television media event organized by the *NBC* network to showcase new television shows for advertisers and *Eric Prydz* is a popular electronic music recording artist who gave a live performance. One possible explanation as to why these events were only detected in the Uber data is that on

[4] http://www.eventful.com

average, attendees of these events are more technologically savvy than those that attend events via taxi. Electronic music is quite popular among young adults and large television media advertising events are attended by those that actively use the latest technology. Given TNC users tend to be younger and more technologically aware than non-users (Rayle et al., 2016; Murphy, 2016), it is notable that these events are only detected in the Uber data.

3.3 Common Events

Finally, a number of events were identified that exist in both the taxi and Uber datasets. Two examples of these include *Dinner in White* on August 25 at Battery Park City and *Bruno Mars & Pharrell* on July 14 at Madison Square Gardens. *Dinner in White* was described as "one of the biggest summer events in New York City" and *Bruno Mars & Pharrell* was a music concert event featuring two of the highest selling recording artists at the time. Both large events appeal to a wide demographic, both young and old from diverse technological backgrounds.

4. Conclusions & Next Steps

This work presents a first step in differentiating Uber and taxi transportation through events attended by their passengers. Using a sample of Uber and taxi pick-up times and locations in New York City, we showed that events can be detected. Moreover, while some events were identified in both the taxi and Uber data, a larger number were detected within only one of the datasets. Through identification of a few of these events, we have taken a preliminary step in showing that there is a difference in the types of events that are attended by TNC users and traditional taxi passengers.

Next steps in this work will be to explore event detection at a variety of spatial and temporal resolutions. Census block groups were suitable as a preliminary boundary, but additional events may be discovered through increasing or decreasing the spatial resolution. Additionally, the time-window during which events occur ranges significantly depending on the type of event. This will be explored in greater detail along with any demographic information associated with the individuals that attend these events.

Lastly, the long term goal of this work is to understand and differentiate motivating factors for human urban mobility. This will have a significant impact on transportation planning and policy, emergency management, and urban infrastructure.

References

Certify (2015). Sharing the road: Business travelers increasingly choose uber. https://www.certify.com/infograph-sharing-the-road.aspx. Accessed: 2015-04-04.

Hall, J., Kendrick, C., and Nosko, C. (2015). The Effects of Uber's Surge Pricing: A Case Study.

Murphy, C. (2016). Shared mobility and the transformation of public transit. Technical report, American Public Transportation Association.

NRC-TRB (2015). Between public and private mobility: Examining the rise of technology-enabled transportation services. Technical report, National Academy of Sciences, Washington, D.C.

Rayle, L., Dai, D., Chan, N., Cervero, R., and Shaheen, S. (2016). Just a better taxi? a survey-based comparison of taxis, transit, and ridesourcing services in san francisco. *Transport Policy*, 45:168–178.

Worboys, M. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19(1):1–28.

Zhang, W., Qi, G., Pan, G., Lu, H., Li, S., and Wu, Z. (2015). City-scale social event detection and evaluation with taxi traces. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):40.