MEASURING THE PULSE OF A CITY VIA TAXI OPERATION: A CASE STUDY

Chenguang Zhu, Corresponding Author
Microsoft
Room 4651, 14820 NE 36th St, Redmond, WA 98052
Tel: (425) 722-6329; Email: zcg.cs60@gmail.com

Balaji Prabhakar
Departments of Electrical Engineering and Computer Science
Stanford University
Room 269, 350 Serra Mall, Stanford, CA, 94305
Tel: (650) 723-5896; Email: balaji@stanford.edu

Word count: 5,117 words text + 9 tables/figures x 250 words (each) = 7,367 words

Revised on Nov. 15, 2016

ABSTRACT

Taxis are vital to modern urban transportation systems. As a great complement to public transportation, taxi services are popular for their flexibility and availability. The ever-increasing magnitude of taxi services has generated a large amount of data about vehicle locations and trips, making it possible to investigate taxi operations in detail. Previous work on this topic has addressed aspects of taxi services such as trip type clustering, visualization of taxi trips and taxi dispatching mechanisms. Based on a data set of taxi trips in New York City, in this paper we analyze a taxi system from some new perspectives. First, from records of fares collected, we investigate driver and passenger behaviors; specifically, what makes some drivers successful and how passengers tip. Second, we study the trips themselves, finding a strong periodicity in the data; for example, the number of pick-ups, the average speed and the average tip ratio are highly periodic in a location-specific manner. This periodicity suggests that each neighborhood in a city has its own "pulse" which makes the taxi system of a city behave in a predictable fashion. Third, we examine the impact of various socio-cultural factors on taxi operations. For example, the average tip ratio of taxi trips can be affected by the outcome of a sports game; fans tip more when their teams win. These results indicate a close connection between taxi systems and human activities, further connecting the pulse of a city to the operation of its taxi systems.

Keywords: Taxi Operation, Behaviors, Periodicity, External Events, Tip Ratio

1. INTRODUCTION

Taxis are vital to modern urban transportation systems. The flexibility of their reach and their round-the-clock availability makes them a great complement to public transportation. In recent years, a rapid increase in the demand for mobility, fueled by the widespread adoption of mobile application and big data technologies, has seen a boom in taxi services. This is true both in terms of the size of taxi fleets (both city taxis and those belonging to private companies), and the number of taxi trips. For example, considering just fleets of city taxis, there are about 14,000 taxis in New York City, serving around half a million passengers on an average day. The cities of Beijing, Tokyo and Seoul all have over 60,000 taxis (*1*).

In conjunction with an increase in fleets and trips, progress in location sensing technology (whether through on-board GPS sensors or phones with drivers) has resulted in a wealth of taxi trace data. Thus, taxis now create "digital footprints" during the course of their operation, either at the start and end of individual trips or through the course of an entire trip. Fortunately, several municipal taxi agencies make such "trip data" available to researchers (*2*, *3*), facilitating numerous scientific studies on taxi operations, e.g. trip destinations and fares, taxi availability, and anomalies (*4*, *5*, *6, 7*, *8, 9*).

These studies improve our understanding of taxi systems and enable innovations in fleet management, passenger service and even urban planning. Further, and closer to the spirit of this paper, an analysis of taxi trip data reveals patterns in urban traffic and the interplay between taxi services and societal dynamics. On one hand, by looking into the supply-demand relationship, the trajectories of taxi movements and other aspects of the data, one can quite precisely characterize the transportation patterns of a city. Regions of great interest are implied by the destinations of trips, bottlenecks in road networks and other congestion areas are revealed, and a shortage or excess of taxi services in different areas can be found. These findings are very useful for taxi operators, regulators and city planners for improving urban transportation systems and reducing the wastage of resources. On the other hand, numerous human activities and societal trends can exert a profound impact on taxi operations, while taxi services also influence urban life rhythms in many ways (*10*). For example, by correlating taxi trips with concurrent urban activities (sports games, concerts, exhibitions, etc.), one can establish a close connection between socio-cultural factors and taxi ridership. These findings uncover social dynamics from raw taxi trip data, facilitating a deeper understanding of how cities function and evolve. In other words, through a close examination on taxi operations, one can measure *the pulse of a city*.

In this paper, we aim to take a novel perspective to analyze and understand taxi systems in big metropolises. We start with precise visualizations[1] of the temporal and spatial distribution of taxi trips. This allows us to locate, precisely quantify and correlate taxi trips on a daily basis as well as at times of extreme anomalies such as during severe weather. We find that, on a daily basis, taxi trips exhibit strong periodicity along several dimensions, including but not limited to the number of taxi rides starting or ending in neighborhoods (or zip codes), the duration of the trips, and the percentage of tips. The periodicity is location-specific and rooted in rhythms governing human existence: we go to work or school at fixed times, and these trips have fixed origins and destinations. The periodicity is also beneficial for an accurate prediction and effective management of taxi operations. Anomalous behaviors emerge when there are socio-cultural factors such as sporting events and music concerts, and due to severe weather. We use a case study method to understand anomalies. For instance, we find that after the Nor'easter blizzard in

[1] Data analysis and visualization technologies in this paper are provided by Urban Engines company's big data system.

February, 2013, it took New York City 12 hours to recover to its normal levels in taxi operation. In another example, we obtain convincing evidence that the outcome of an NBA game can affect the percentage of tips for post-game taxi trips.

Moreover, since a typical taxi system comprises hundreds of thousands of passengers and drivers, their behaviors can profoundly shape the operational status of a taxi system. By comprehending how drivers and passengers behave and interact, one can find explanations of various phenomena in taxi operations and ways of improving taxi services. In this direction, we examine: (i) why some drivers earn more while working the same hours, and (ii) the spatial and temporal patterns of tipping behaviors. Some of these behaviors are for economic reasons, while others may be due to other factors which the data at hand doesn't allow us to explore.

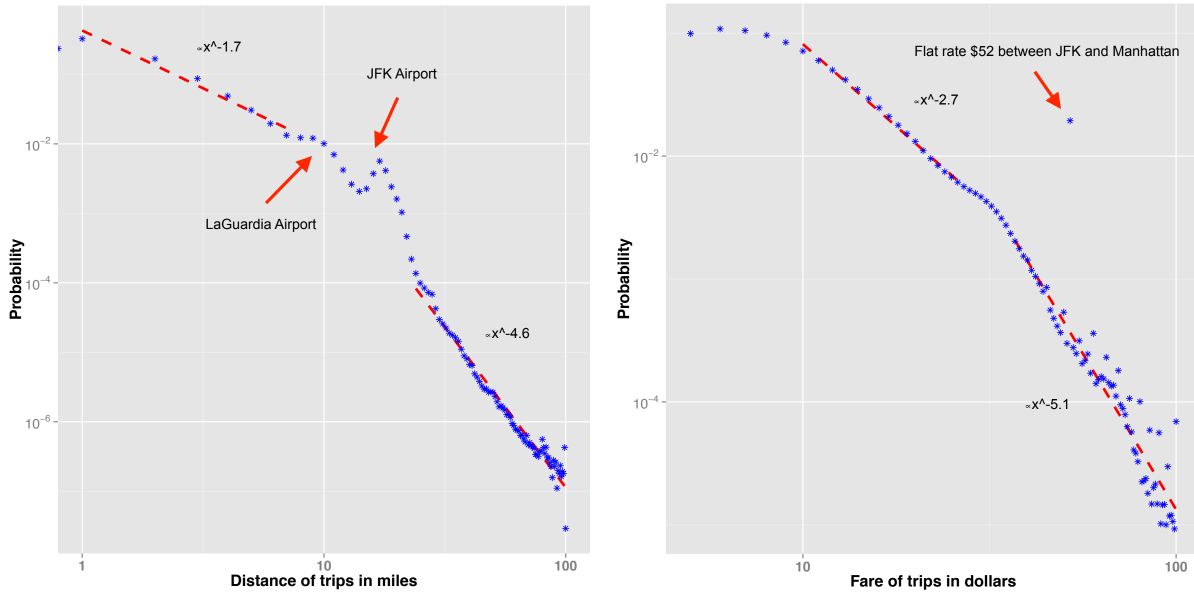
The rest of this paper is organized as follows. Section 2 reviews previous literature. Section 3 presents general statistics of New York City's yellow taxi data. Section 4 focuses on the behaviors of drivers and passengers. Section 5 presents the periodicity inherent in taxi trips. Section 6 discusses anomalous taxi operations, and Section 7 concludes the paper. Due to a limitation of space, we are only able to provide a few vignettes of New York City taxi data.

2. LITERATURE REVIEW

Correlation between urban dynamics and taxi operations. Reference (*4*) computed statistics about New York City taxi data and conducted clustering on trip types, e.g. business trips and leisure trips; reference (*11*) gave a comprehensive overview of global taxi services and various analytical methods of taxi operation, as well as its social impact such as taxi regulation, strikes and a taxi system's contribution to urban economy; reference (*5*) focused on demographic aspects of taxi ridership in New York City, including income levels and accessibility to subway; references (*6*, 7) developed methodologies to identify urban hotspots from taxi trace data; and reference (*8*) employed models to combine road network, human mobility and points of interest into a holistic framework. In (*8*), city regions were modeled as documents and the functions of regions were characterized as topics, while human mobility in regions were described as words. The model was capable of discovering intents of taxi rides and functions of urban regions simultaneously.

Anomaly detection. Reference (*9*) found outlier road segments by detecting drastic changes between current observations and historical data; reference (*12*) applied conditional random fields to detect taxis operating in anomalous modes; and reference (*13*) proposed a two-step method to explain the causes of anomalies in taxi services. In (*13*), the city was first divided into several regions, and anomalous connections between regions were identified by detecting deviations from historical norms. The authors then developed a generative traffic flow model to uncover probable sources of observed anomalies.

Compared to these works, our paper investigates previously overlooked aspects of taxi operations, including: (i) the behavior of drivers and passengers and the consequences, (ii) a strong periodicity in the usage of taxi systems, and (iii) a detailed explanation of anomalies in taxi operations from external sources such as socio-cultural factors and severe weather conditions.



(a) Distribution of trip distances (in miles) (b) Distribution of trip fares (in dollars)

FIGURE 1 Distribution of trip distances and fares of yellow taxis in New York City in 2013.

3. GENERAL STATISTICS OF NEW YORK CITY TAXI TRIP DATA

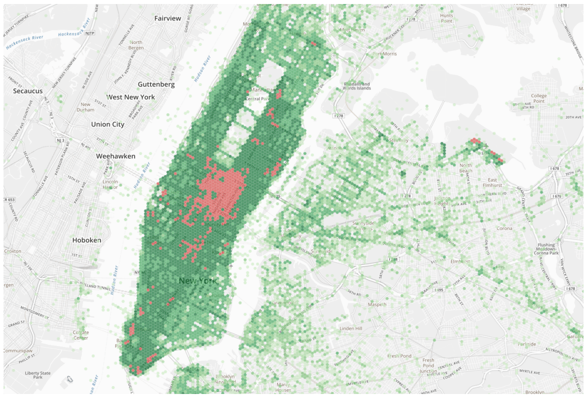
New York City is the most populous city in the United States and one of the densest urban agglomerations of the world. The transportation system in New York City supports millions of commutes each day, and yellow taxis are one of the most important parts in the city's complex transportation system.

To facilitate the management and regulation of taxicabs, the New York City Taxi & Limousine Commission recorded information about each yellow taxi trip via GPS devices and released the data for public research (*2*). The data contains a total of 173.2 million trips in the year of 2013. Each trip is described by an entry including hashed driver's license number, hashed medallion number, pickup/drop-off location and time, amount of fare, tips collected, surcharges and trip distance. These trips were made by a total of 43,191 drivers and 14,144 taxi vehicles. In 2013, yellow taxi drivers in New York City covered 500.6 million miles, collected a total fare of 2.1 billion dollars, and tips of 236.4 million dollars. Note that tips are recorded mostly for payments made with a credit card: very few of the cash payment trips (which equal 45.9% of all trips) have tip information.

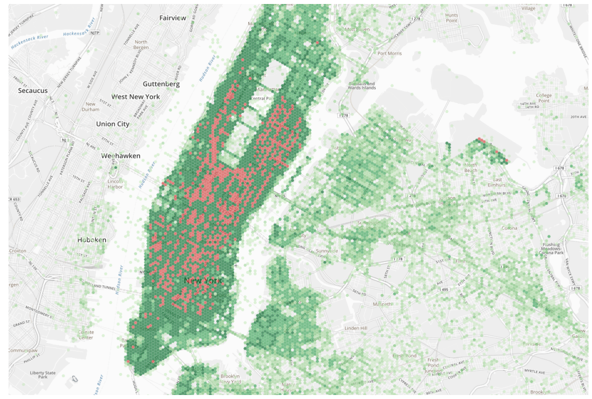
Trip distance and fare

As shown in FIGURE 1(a), the distance of taxi trips follows a power-law distribution: we can fit straight lines in the log-log plot. However, two peaks occur in the distribution. After investigation, we find out that these two peaks correspond to the two airports in New York City—LaGuardia Airport and JFK Airport. The fitted lines in red indicate that the number of trips decays more rapidly for longer distance.

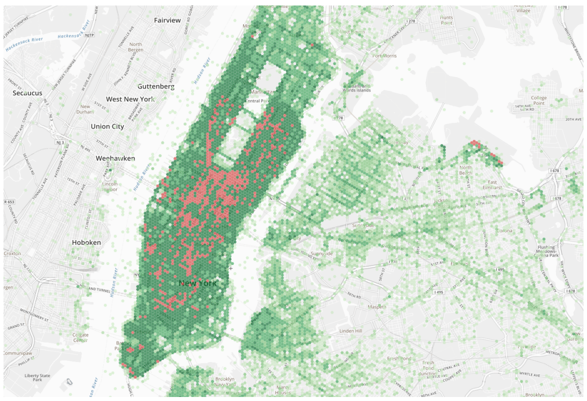
The distribution of trip fares similarly obeys power law (FIGURE 1(b)). One interesting observation is that a huge peak occurs at $52/trip. It is due to the regulation that any trip between JFK Airport and Manhattan has a flat rate of $52 (*14*). It implies that taxi trips to and from JFK Airport are very popular.



(a) Weekday mornings (6:30AM-9:30AM)



(b) Weekday evenings (5:30PM-8:30PM)



(c) Weekends (9:00AM-6:00PM)

FIGURE 2 Destination of taxi trips on weekdays and weekends. The highest density is marked by red color, followed by dark green and light green.

Ride destinations

The distribution of taxi ride destinations has a strong geo-spatial aspect, as shown in FIGURE 2. Regions with the highest density of destinations are colored red, followed by dark green and light green. From the plots, we make the following observations:

- On weekday mornings, most passengers are travelling towards midtown Manhattan, which is the central business district of New York City.
- On weekday evenings, taxi trip destinations are much more scattered, covering uptown, midtown and downtown of Manhattan. These areas include residential districts and places for entertainment: restaurants, theatres, bars, etc.
- On weekends, the distribution of taxi trip destinations resembles that on weekday evenings. Weekend hotspots include popular tourist locations or transportation hubs: Times Square, 5th Avenue, Broadway, Grand Central Terminal, etc.

To summarize, the destinations of taxi trips indicate the work-life cycle of urban dwellers and visitors.

4. SOME ASPECTS OF DRIVER AND PASSENGER BEHAVIORS

A typical taxi system comprises hundreds of thousands of passengers and drivers. Their behaviors can significantly affect taxi operations. In this section, we closely examine the behavior of taxi drivers and passengers from various perspectives and its impact on taxi systems.

Successful drivers

We observe in the data that some taxi drivers make more money while working the same hours as other drivers. Indeed, more is true: in terms of the average fare collected per busy hour (i.e., excluding the period when a taxi has no passengers inside), certain drivers perform far better than others. Since the taxi fare structure is fixed and the amount of time a taxi is idle has been accounted for, this difference in earnings is worth investigating.

We consider taxi trips in May 2013, made by 33,226 distinct taxi drivers. To remove the effect of outliers, we consider 8,314 frequent drivers: those who (i) worked on at least 20 days, (ii) made at least 600 trips, (iii) had passengers for at least 100 hours, and (iv) earned a total fare of at least $3,000. We rank the frequent drivers by their “earnings per busy hour” (EPBH). For example, if a driver makes $6,000 from trips with a total duration of 100 hours, his EPBH is $60. Surprisingly, a large difference exists between the top and bottom drivers in terms of EPBH: the top three drivers earned $106.4, $97.2, $91.8 per busy hour, while the bottom three drivers earned $41.8, $44.9 and $45.7 per busy hour. In other words, the top drivers' efficiency is more than double that of the bottom ones.

We investigate various factors affecting EPBH, including average trip distance, total trip distance, average trip duration and average trip speed. We perform a linear regression of EPBH against the above factors, finding that all factors are statistically significant at level $\alpha = 0.05$. To summarize, EPBH positively correlates with average trip distance, total trip distance and average trip speed. Conversely, EPBH negatively correlates with average trip duration.

To rank feature importance, we compute the Pearson’s correlation coefficients between these factors and EPBH. The result is: average trip speed (**0.825**), average trip distance (**0.579**), total trip distance (**0.507**) and average trip duration **(-0.458**). It follows that average trip speed is the most significant factor affecting EPBH: faster drivers earn more per unit time. This result supports previous findings that successful drivers usually take their passengers along the shortest route to their destinations (*15*). However, our conclusions are reached without any trip trajectory data; we only know durations and distances.

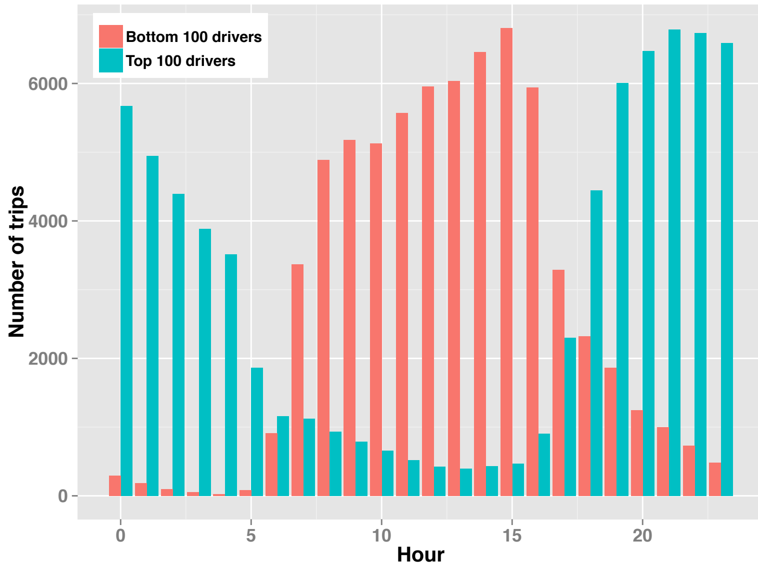


FIGURE 3 Hourly density of taxi trips made by top 100 drivers and bottom 100 drivers ranked by fares earned per busy hour (EPBH).

This finding raises another question: why can some drivers achieve higher trip speed than others? To answer this, we plot the times during which the top 100 and bottom 100 drivers ranked by EPBH make their trips. It is worth noting that the difference in earning efficiency between these two groups of drivers is huge: the top 100 drivers have an average EPBH of \$71.8, while the bottom 100 drivers have an average EPBH of \$48.7.

As shown in FIGURE 3, most top drivers work from 5PM to 5AM, whereas bottom drivers are more likely to work during the day. Therefore, one explanation of the success of the top drivers is that they work when the roads are free of congestion, thereby earning more per unit time (busy or not).

Note that in the calculations above, we have excluded the passenger surcharges of yellow taxis in New York City: \$1.00 during 4PM-8PM on weekdays and \$0.50 during 8PM-6AM every day. With these surcharges included, those top drivers actually earn even more.

Tipping

Tipping taxi drivers is a polite way to express satisfaction and gratitude for a safe and pleasant trip. In New York City taxi data, the tipping information is partially available: the tip recorded for trips with cash payment (45.9% of all trips) is mostly zero. As a result, we focus on trips with card payment in the following analysis. Here we define tip ratio as the amount of tips divided by the payment other than tips, including fares, surcharges, tolls and taxes.

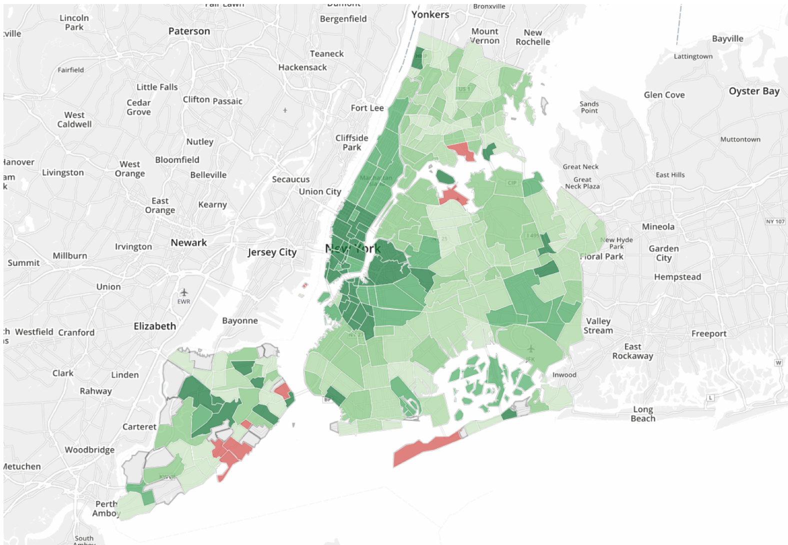
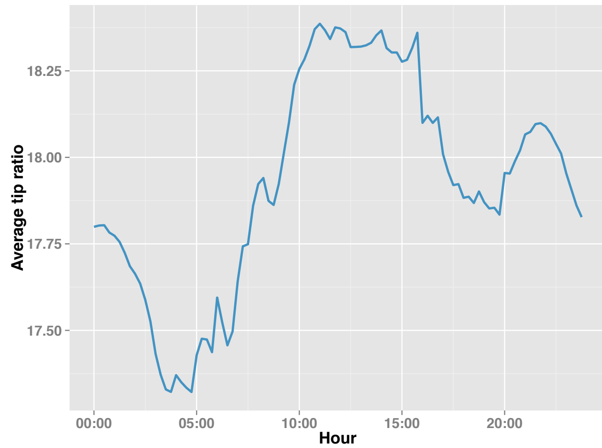


FIGURE 4 Average tip ratio for rides starting from different areas in New York City. Regions with the highest tip ratio are marked in red, followed by dark green and light green.

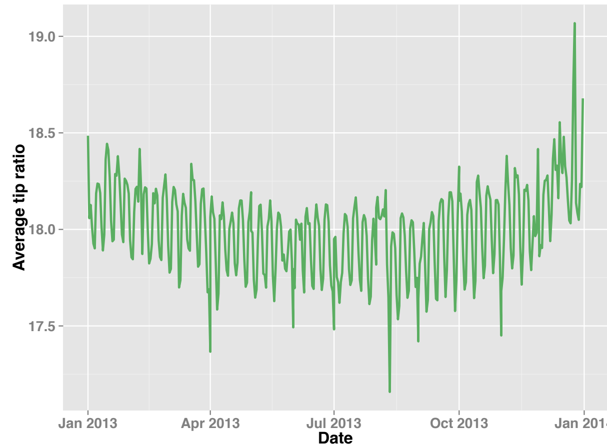
FIGURE 4 shows the average tip ratio of taxi trips starting from different regions in New York City. On average, tips are most generous from trips from south Bronx, north Queens, Long Island and eastern Staten Island. In Manhattan, trips from midtown and downtown enjoy a higher average tip ratio than those from uptown.

FIGURE 5 depicts the temporal patterns of tipping. FIGURE 5(a) shows the average tip ratio per 15 minutes over the day. As shown, passengers are more generous from 10AM to 4PM. As the evening rush hour starts, the average tip ratio drops to around 0.5%, which is probably due to the bad mood created by congestion: meters still tick up when taxis are stuck in traffic plus anxiety to get home earlier. Over the course of the whole year, tip ratio is higher around New Year, as indicated by FIGURE 5(b). In fact, the highest daily tip ratio falls on Christmas Eve (19.1%) and Christmas Day (18.8%), followed by New Year's Eve (18.7%). All this supports the believable hypothesis that tips are higher when people are happier and less stressed.

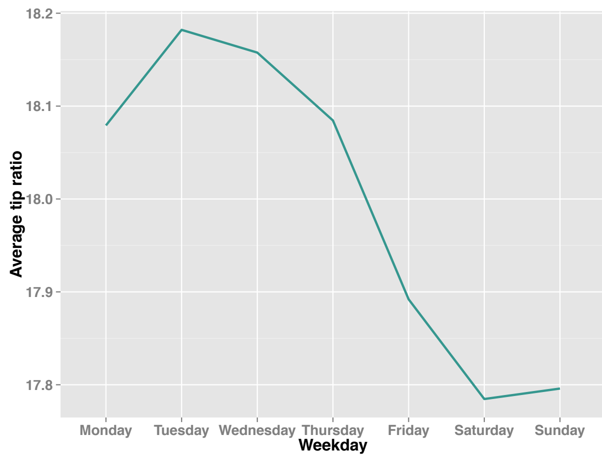
Another pattern that stands out in FIGURE 5(b) is the weekly periodicity. FIGURE 5(c) shows the average tip ratio by weekdays. While the variation across the week isn't huge, it is consistent: peaking on Tuesdays and decreasing the rest of the week. It isn't immediately obvious why this difference occurs or why it is so consistent.



(a) Tip ratio per 15 minutes over the day

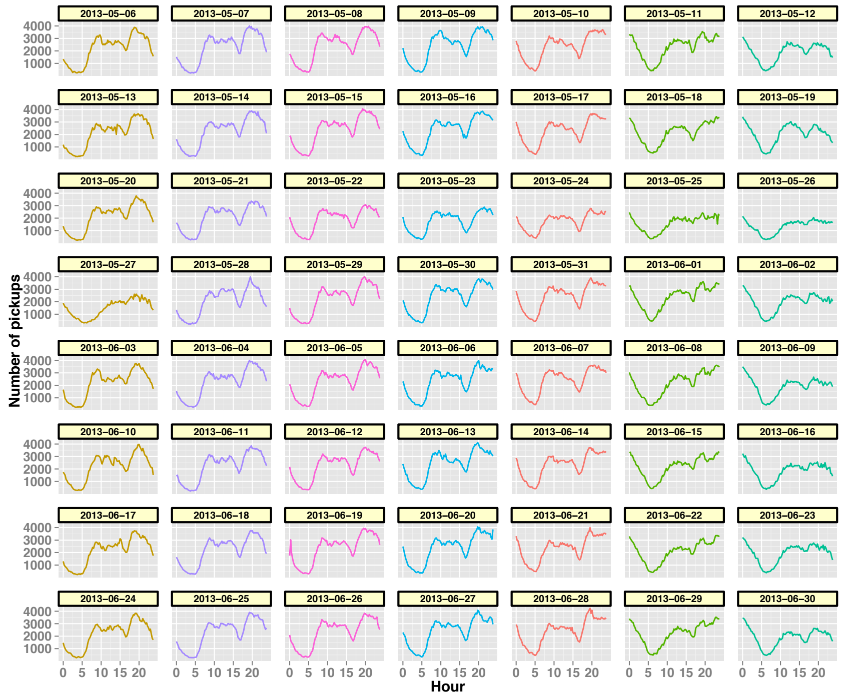


(b) Tip ratio per day



(c) Tip ratio per weekday

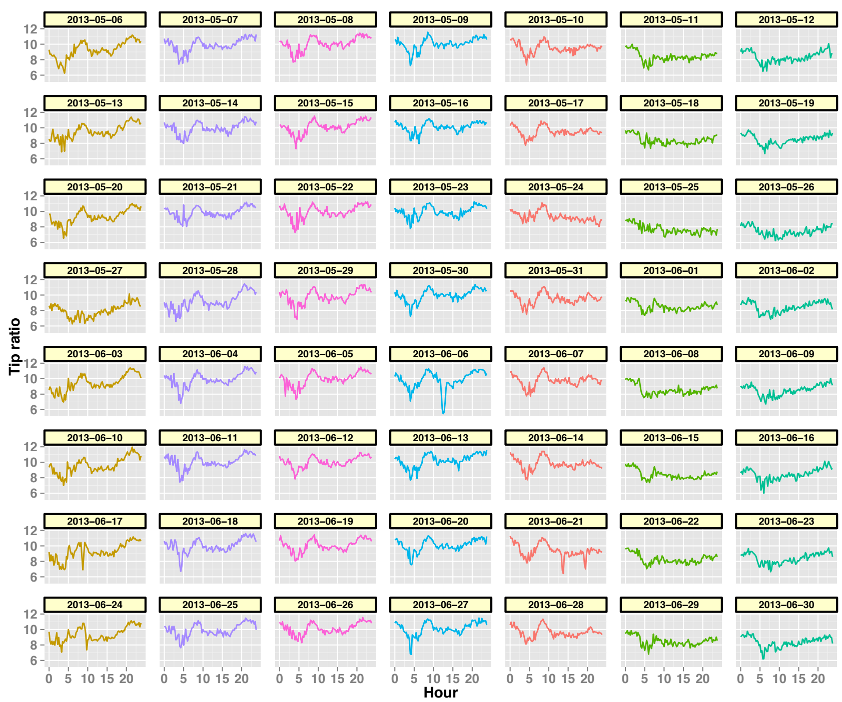
FIGURE 5 Average tip ratio (in percentage) of taxi trips in New York City in 2013, (a) per 15 minutes over different times of day; (b) per day; (c) per weekday.



(a) Number of pickups in Midtown Manhattan

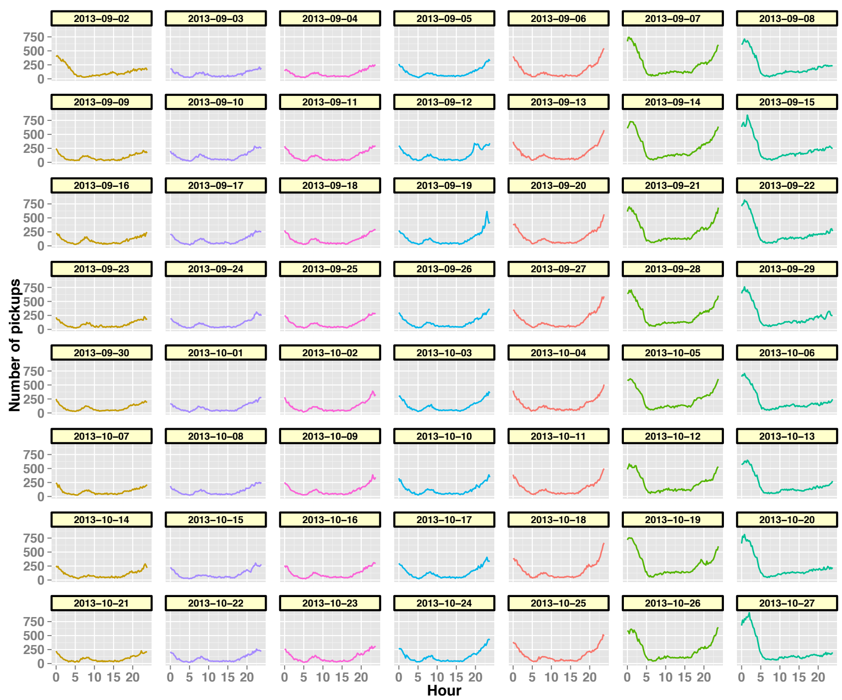
(b) Number of drop-offs in Midtown Manhattan

(c) Average trip distance (in miles) in Midtown Manhattan



(d) Average tip ratio of trips (in percentage) in Midtown Manhattan

(e) Number of pickups in Financial District



(f) Number of pickups in Brooklyn

FIGURE 6 (a) Number of taxi pickups, (b) number of taxi drop-offs, (c) average trip distance (in miles) and (d) average tip ratio of trips in percentage per 15 minutes in Midtown Manhattan and number of taxi pickups in (e) Financial District and (f) Brooklyn in May and June of 2013. From left to right, the columns correspond to Monday through Sunday.

5. PERIODICITY

Taxi passengers can be generally categorized into two types: regular commuters who go to work and return home by taxis on a frequent basis, and visitors or occasional passengers who utilize the taxi system infrequently. The first type of passenger is expected to have relatively stable trip patterns, whereas it is hard to predict the trips of the second type of passenger. However, when we look at all trip records collectively, we observe a strong daily periodicity in the taxi system.

For example, we analyzed taxi trips in Midtown Manhattan from May to June in 2013. FIGURE 6(a-d) shows the number of pickups, number of drop-offs, average trip distance and average tip ratio per 15 minutes. Each column corresponds to one weekday, the first column being Mondays. In FIGURE 6(a), daily periodicity can be clearly observed: the number of pickups in Midtown Manhattan follows nearly the same distribution on each weekday, varying slightly across each weekday. For instance, the number of pickups on Mondays drops sharply before midnight while flattening on Friday nights. One exception to the periodicity is from May 25 to May 27, the Memorial weekend, when many fewer people took taxis. Indeed, holidays are one of the main sources of anomalies in taxi operations.

Similarly, the number of taxi drop-offs follows weekly patterns, as shown in FIGURE 6(b). The same observation is made on the average trip distance in FIGURE 6(c), in which we find that long-distance taxi trips usually happen in early morning around 5AM, and trips on Monday mornings are the longest—above 6 miles on average. The average tip ratio also follows a regular pattern (FIGURE 6(d)): the tip ratio on weekdays has two peaks—around 9AM and 9PM, while the tip ratio on weekends is relatively flat. We discover similar periodicity in other regions of New York City during various times, including Financial District and Brooklyn (FIGURE 6(e-f)).

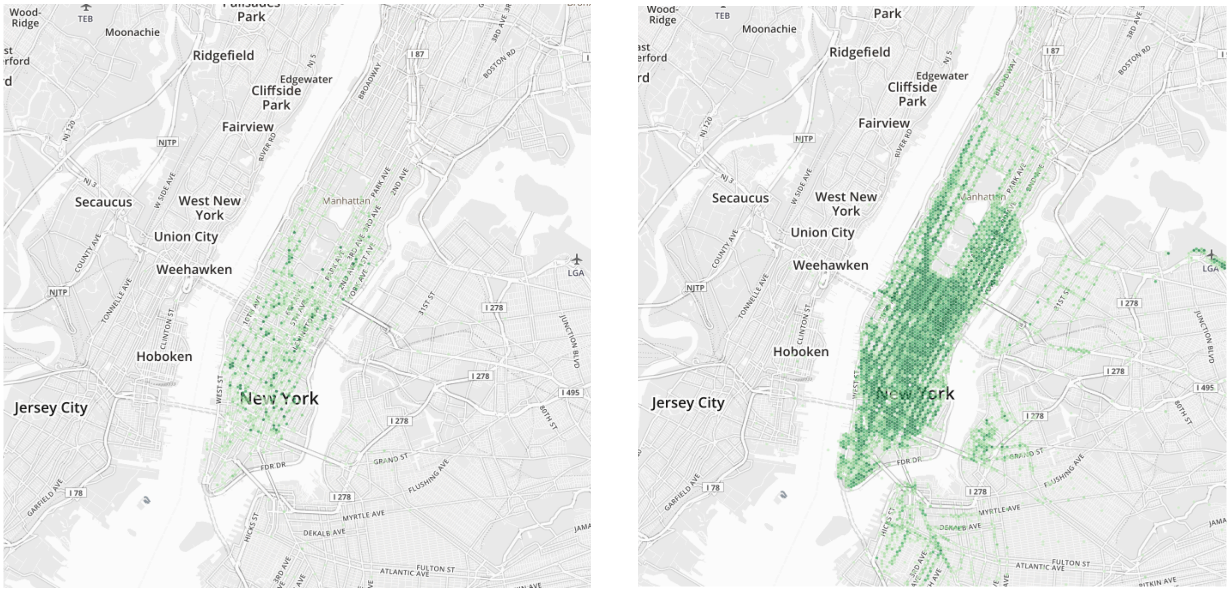
From these observations, we conclude that taxi trips possess a strong periodicity. Further, there is a strong location-specificity to these periodic patterns: trip patterns at different neighborhoods look different, but at a given neighborhood there is a strong daily periodicity, depending on the day of the week. This phenomenon is beneficial for the analysis and management of a taxi system, since snapshots of the system in a typical week describe taxi operations for a much longer time. This makes the taxi system performance more predictable and controllable. Further, since this periodicity emerges from the stable pattern of urban work and life cycles, we conjecture it holds in other urban areas.

6. CAUSES OF ANOMALIES AND IRREGULARITIES IN TAXI USAGE

In contrast to the regular patterns of taxi usage described in the previous section, external factors such as the weather and major social events (e.g., sporting events, concerts, major gatherings, strikes and holidays) give rise to anomalies and irregularities in taxi usage. Although these effects are often transient and local (except the effects of severe weather), their sudden impact may pose serious challenges to the system. In this section, we shall examine several cases to understand the causes of irregularities in taxi systems.

Impact of severe weather conditions

New York City, being on the northeastern shore of the US, is exposed to natural phenomena like blizzards and storms. While it is clear that taxi availability is impaired during bad weather, it has been hard to quantify the toll on the taxi system. In this subsection, we will investigate a blizzard, quantify its impact on the NYC taxi system, and analyze the degree of resilience of the taxi system in response to the storm.



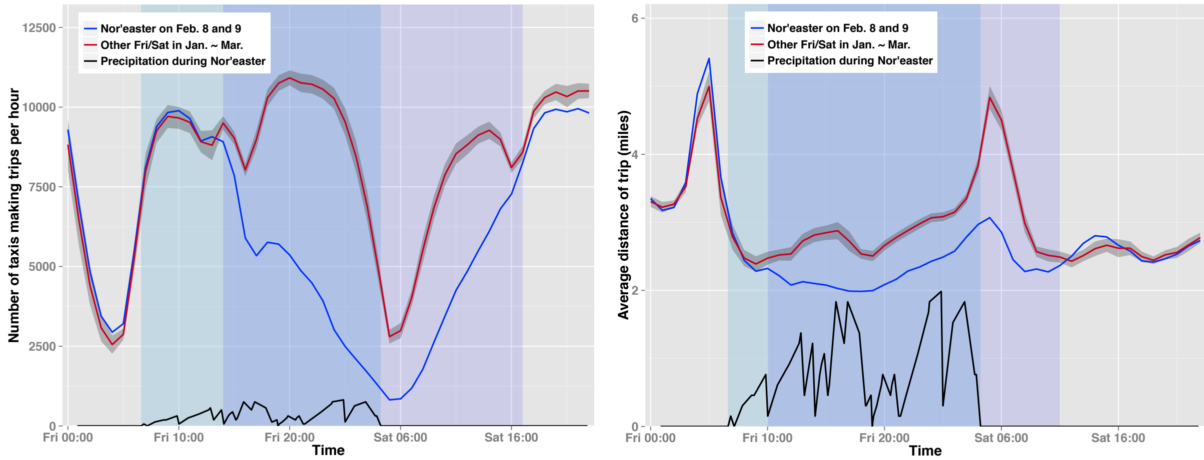
(a) Nor'easter night (Feb. 8, 2013) (b) Other Friday nights (Jan.-Mar. 2013)

FIGURE 7 Density map of taxi pickups from 9PM to midnight on Fridays. Darker green indicates more trips starting from that area. (a) Trips on Nor'easter night (February 8, 2013); (b) the average number of trips on other Friday nights from January to March 2013.

On Feb. 8-9, 2013, the February Nor'easter hit NYC, bringing heavy snowfall and hurricane-force winds. 11.4 inches (29 cm) of snow was recorded at Central Park (16). FIGURE 7 shows a considerable decrease in the number of taxi pickups on Nor'easter night, compared with other Friday nights from January to March in 2013.

FIGURE 8(a) shows the number of taxis making trips during the Nor'easter and on other Fridays and Saturdays from January to March. The gray ribbon indicates one standard deviation, while the black curve shows the amount of precipitation. As observed, after the snow started at 6:36AM on Friday, February 8, the number of taxis making trips is within the usual norm until more than seven hours later: it took a long time for the snow to accumulate to a level sufficient enough to impair driving conditions. From 2PM on Friday till 5pm on Saturday, the number of taxis dropped between 10% and 75% of the normal levels, and on average during this time by 49.2% as compared with normal levels. The anomaly lasted until 5PM on Saturday, **12 hours** after snowing had ceased. In other words, it took half a day for the NYC taxi system to recover to its normal state, a period which includes the snow melting, cleaning streets and clearing the aftermaths of traffic accidents.

Meanwhile, passengers hired taxis for nearer destinations during the Nor'easter. FIGURE 8(b) shows that the average trip distance dropped below normal levels during the snow: a taxi trip was on average 21% shorter (0.66 miles) than norms. This is likely due to the fact that people headed home instead of farther away places for Friday night entertainment, and due to safety reasons since longer journeys are much more hazardous during a storm.



(a) Number of taxis making trips (b) Average trip distance (in miles)

FIGURE 8 (a) Number of taxis making trips and (b) the average trip distance during Nor'easter (blue curve) and other Fridays and Saturdays from January to March 2013 (red curve). The gray ribbon indicates one standard deviation. The amount of precipitation is shown in black curve with scaling. The light blue rectangle indicates the period of snowing while the light purple rectangle indicates the period when the blue curve deviates from normal levels.

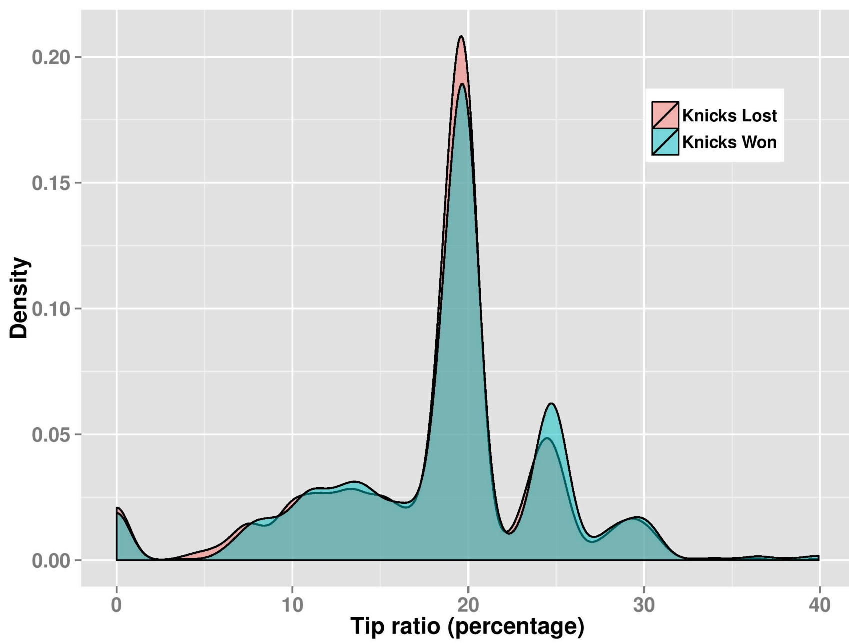


FIGURE 9 Distribution of tip ratio (in percentage) of taxi trips from Madison Square Garden after Knicks won or lost a match. When the Knicks won a game, more passengers chose higher tipping options.

As a result of both a decreasing number of taxis and shortened trips, the revenues of New York City taxis suffered from a huge loss during the Nor'easter. From 2PM on Friday to 5PM on Saturday, the total collected fare dropped 48.4%, or 3.5 million dollars, compared with the average level on other Fridays and Saturdays from January to March. Moreover, traffic accidents, vehicle maintenance costs and damaged roads also take a toll on the taxi system, although we do not have data to quantify such costs.

Impact of socio-cultural factors

Transportation is closely connected with other aspects of daily life. Various socio-cultural factors exert a profound impact on transportation, in particular the taxi system. In this subsection, we discuss a case study in which the outcome of a sports game can affect the tipping behaviors of passengers.

New York Knicks is a famous basketball team in NBA. Its home stadium is located at Madison Square Garden. When the Knicks plays at home, thousands of fans flock to the stadium to support their team. After the game, many fans take taxis to leave the stadium. These trips reveal an interesting relationship between game results and tip ratio.

In January 2013, Knicks played four games at home on weekday nights. We focus on post-game taxi trips which (i) started between 10:30PM and midnight on weekdays in January, and (ii) originated within 2 blocks of Madison Square Garden. We end up with 10,678 taxi trips and calculate the average tip ratio per trip when (i) Knicks won; (ii) Knicks lost; (iii) there was no game. We find that after Knicks won a game, the average tip increased by nearly one percent more on average compared with the case when Knicks lost: 19.0% versus 18.1%. The difference is statistically significant at level $\alpha = 0.05$ in unpaired two-sample t-tests. On nights with no games (or other events), the average tip ratio was 18.3%. A more compelling view of this extra tipping behavior is shown in FIGURE 9, which shows distribution of the tip ratio of post-game trips. First, three peaks appear at 20%, 25% and 30% because they are the suggested tipping options on the payment screen in yellow taxis. Second, when Knicks won a game, many passengers *selected* higher tipping ratios, e.g. from 20% to 25% and 30%.

This finding is quite interesting and lends credence to the hypothesis that happier customers tend to tip more, in this case even though their source of happiness is not necessarily derived from the very service (taxi ride) they are tipping!

7. SUMMARY

In this paper, we have conducted a detailed analysis of the New York City yellow taxi system. From taxi trip data over a whole year, we have obtained several insights into the behaviors of passengers and drivers, the periodicity in the taxi system during normal days and the impact of external sources such as severe weather conditions and socio-cultural factors. A limitation of space has restricted us to cover a few examples; other results may be found in (*17*). These findings depict a clear picture of the taxi system, build connections between taxi operations and urban life cycles, and shed lights on directions for further research on taxis.

ACKNOWLEDGEMENT

We would like to thank Urban Engines company for their big data system for analyzing and visualizing large-scale city taxi data.

REFERENCES

1. Y. Zheng, Y. Liu, J. Yuan, X. Xie. *Urban computing with taxicabs*. In Proceedings of the 13th international conference on Ubiquitous computing. ACM, 2011.
2. B. Donovan, D. B. Work. New York City Taxi Trip Data (2010-2013), 2014. https://uofi.app.box.com/NYCtaxidata. Accessed on Dec. 29, 2015.
3. Y. Zheng. *T-Drive trajectory data sample*. http://research.microsoft.com/apps/pubs/?id=152883. Accessed on Jan. 25, 2016.
4. X. Qian, X. Zhan, S. V. Ukkusuri. *Characterizing urban dynamics using large scale taxicab data*. Transportation Research Record, 2013.
5. X. Qian, S. V. Ukkusuri. *Spatial variation of the urban taxi ridership using GPS data.* Applied Geography, 2015.
6. Y. Zheng, L. Zhang, X. Xie, W. Y. Ma. *Mining interesting locations and travel sequences from GPS trajectories.* In Proceedings of the 18th international conference on World wide web, ACM, 2009.
7. S. Liu, Y. Liu, L. M. Ni, J. Fan, M. Li. *Towards mobility-based clustering.* In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2010.
8. J. Yuan, Y. Zheng, X. Xie. *Discovering regions of different functions in a city using human mobility and POIs.* In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012.
9. X. Li, Z. Li, J. Han, J.-G. Lee. *Temporal outlier detection in vehicle traffic data.* In Proceedings of the International Conference on Data Engineering, 2009.
10. P. S. Castro, D. Zhang, C. Chen, S. Li, G. Pan. *From taxi GPS traces to social and community dynamics: A survey.* ACM Computing Surveys (CSUR), 2013.
11. J. Cooper, M. Ray, N. John. *Taxi!: urban economies and the social and transport impacts of the taxicab.* Ashgate Publishing, Ltd., 2010.
12. Z. Liao, Y. Yu, B. Chen. *Anomaly detection in GPS data based on visual analytics.* In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, 2010.
13. S. Chawla, Y. Zheng, J. Hu. *Inferring the root cause in road traffic anomalies.* In Proceedings of the IEEE 12th International Conference on Data Mining, 2012.
14. Yellow Taxicab Rate of Fare. http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml. Accessed on Dec. 29, 2015.
15. L. Liu, C. Andris, C. Ratti. *Uncovering cabdrivers' behavior patterns from their digital traces.* Computer, Environment and Urban System, 2010.
16. February 2013 nor'easter. https://en.wikipedia.org/wiki/February_2013_nor%27easter. Accessed on Dec. 29, 2015.
17. C. Zhu. *Analysis and modeling of large-scale systems: taxis and social polling.* Ph.D. dissertation, Stanford University, 2015.