NBER WORKING PAPER SERIES

RACIAL AND GENDER DISCRIMINATION IN TRANSPORTATION NETWORK
COMPANIES

Yanbo Ge
Christopher R. Knittel
Don MacKenzie
Stephen Zoepf

ABSTRACT

Passengers have faced a history of discrimination in transportation systems. Peer transportation companies such as Uber and Lyft present the opportunity to rectify long-standing discrimination or worsen it. We sent passengers in Seattle, WA and Boston, MA to hail nearly 1,500 rides on controlled routes and recorded key performance metrics. Results indicated a pattern of discrimination, which we observed in Seattle through longer waiting times for African American passengers—as much as a 35 percent increase. In Boston, we observed discrimination by Uber drivers via more frequent cancellations against passengers when they used African American-sounding names. Across all trips, the cancellation rate for African American sounding names was more than twice as frequent compared to white sounding names. Male passengers requesting a ride in low-density areas were more than three times as likely to have their trip canceled when they used a African American-sounding name than when they used a white-sounding name. We also find evidence that drivers took female passengers for longer, more expensive, rides in Boston. We observe that removing names from trip booking may alleviate the immediate problem but could introduce other pathways for unequal treatment of passengers.

Yanbo Ge
University of Washington
Department of Civil and
  Environmental Engineering
yanboge@uw.edu

Christopher R. Knittel
MIT Sloan School of Management
100 Main Street, E62-513
Cambridge, MA 02142
and NBER
knittel@mit.edu

Don MacKenzie
University of Washington
Department of Civil and
  Environmental Engineering
dwhm@uw.edu

Stephen Zoepf
Stanford University
Center for Automotive Research
    at Stanford (CARS) szoepf@stanford.edu

A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/1689

1 Introduction

As technological and market advances of the peer economy continue to outpace regulation, enforcement of responsible behavior within peer economies has been left to internal feedback mechanisms and, occasionally, to the courts. One area of behavior with important social equity implications is that of discrimination in the provision of services. In the transportation sector, Transportation Network Companies (TNCs, also referred to as "ride sourcing" or "e-hailing" companies), such as Uber and Lyft, match individual travelers with drivers in real time.

Discrimination by taxi drivers has long been an acknowledged social issue, providing plenty of pop-culture fodder. Taxi drivers in most cities are required to pick up any passenger while on duty, and taxi drivers are reminded of this obligation (Harshbarger (2015)). Despite periodic high-profile incidents of taxi discrimination (e.g., Donnelly (2015); Gonen (2015); Glanville (2015)), equal-service provisions are difficult to enforce.

The relationship between TNCs and discrimination is a complex one. At a high level, a study funded by Uber (Smart et al. (2015)) found that UberX provided lower fares and shorter waits than taxis in areas of Los Angeles with below-average incomes. Hughes and MacKenzie (2016) found that expected waiting times for an UberX ride were shorter in Seattle-area neighborhoods that had lower average incomes and more minorities, even after adjusting for differences in residential and employment density. While Uber's advertising certainly includes minority customers (e.g., Figure 1), popular press articles (e.g., Nicholson (2013)) have noted that TNC services are unavailable to customers without access to a credit card, who are more likely to be lower-income and members of minorities.

While these systemic concerns may be valid, decisions about where to drive and who to pickup are choices made by drivers—not by the TNCs themselves. Drivers are required to maintain high overall measures of performance including a high star rating from passengers, high ride acceptance rate, and low cancellation rate. However, even these driver-specific performance measures may not detect whether a driver behaves differently to passengers based on their race or gender.

In this paper, we test for discrimination in the peer-to-peer transportation domain. The primary question we seek to answer is whether the rapidly-growing TNC market treats customers of all races and genders equally. Plainly put, is a taxi driver's decision, made in public view, not to stop for a African American passenger being eliminated? Or is it just being replaced by a TNC driver's screen swipe, made in private, that has the same effect? To do this, we undertook two large-scale randomized control trials in Boston, MA and Seattle, WA. We had research assistants (RAs) serve as confederates, summoning rides in both regions. In the Seattle experiment, we hired RAs of different racial backgrounds. Each RA requested rides along pre-determined routes where the assignment to those routes was randomized across racial backgrounds. In the Boston experiment, each RA requested rides under two different names: using the nomenclature of Fryer and Levitt (2004), one

name was "white sounding" while the other was a "distinctively black name."[1]

For each trip, the RAs gathered data using four screenshots per trip: (i) just before requesting a trip (with expected wait time), (ii) just after a trip is accepted by a driver (with a new wait time), (iii) when a driver arrives for pickup, and (iv) when the vehicle stops at the end of the trip. Using the data gathered from these screenshots, we evaluate waiting times, travel times, drivers' cancellation rates, costs, and (where applicable) ratings awarded by drivers to the travelers.

We find significant evidence of racial discrimination in both experiments. In the Seattle experiment trip requests coming from African American riders take between 16 to 28 percent longer to be accepted for both UberX and Lyft services. For UberX this translates into African American riders waiting between 29 and 35 percent longer for a ride; the effect of race for Lyft is too imprecise to draw any conclusions. Interestingly, the timing of information available to *drivers* differs across UberX and Lyft. A Lyft driver sees both the name and a photo of the passenger prior to accepting or denying a ride, while UberX drivers see this information only after accepting a request.

In the Boston experiment, we leverage the variation in information timing and collected data on acceptance times, wait times, *and* driver cancellations. We find that the probability a driver accepts a ride, but then subsequently cancels the ride, more than doubles for African American riders of UberX. As expected, we find no effect on cancellations for African American riders of Lyft because, we surmise, that given that names and photos are visible to the driver prior to acceptance, any discrimination occurs prior to accepting the initial request. We also show that this increase in the rate of cancellations is concentrated among African American males; their cancellation rates are three times that of white males. Furthermore this cancellation effect is concentrated in low population density areas, perhaps because drivers in those areas self select to reduce their interaction with African Americans. We do not, however, find that the increase in cancellation rates manifests itself into increases in wait times, presumably because of the denser network of drivers in the Boston area. While our estimates of the time it takes for the request to be initially accepted are imprecise, we find that acceptance times for African American males on Lyft are longer. This is consistent with some Lyft drivers passing up requests from African American males, while UberX drivers are forced to accept and then cancel these requests.

Discrimination by TNC drivers may occur in at least four primary ways, highlighted in Figure 2: (i) drivers could elect not to drive in or near certain types of neighborhoods, (ii) drivers could decline to accept reservations from certain types of passengers or could cancel a pickup once the passenger's identity becomes revealed, (iii) drivers could leave low ratings for passengers based on race, gender, or perceived socioeconomic status, (iv) drivers could choose non-ideal routes based on the same factors, increasing costs and/or travel time. Our analysis focuses on the second, fourth, and to a lesser extent the third, of these channels.

---

[1]Bertrand and Mullainathan (2004) found that when stereotypically white names and stereotypically African American names were randomly assigned to job applications, those with the white names received 50% more callbacks for interviews.

Our results fit within existing work that has demonstrated evidence of profound discrimination in the peer economy. Pope and Sydnor (2011) showed that loan requests were significantly less likely to be successful when associated with a African American person than when associated with a white person of comparable credit history. Discrimination can occur even without seeing a person. Ongoing work by researchers at Harvard Business School has demonstrated racial discrimination on both sides of a peer-to-peer transaction on the AirBnB platform. (Edelman and Luca (2014); Edelman et al. (2016)).

Our work also fits within an existing literature that has identified discrimination in the transportation sector. For example, Goddard et al. (2015) finds discrimination behavior of drivers toward pedestrians of different races, with African American pedestrians at a crosswalk being passed by twice as many cars and waiting 30% longer for a car to stop than white pedestrians. Other research has shown that perceived differences in social class affected vehicle-vehicle and vehicle-pedestrian yielding behavior (Piff et al. (2012)). Given the presence of discrimination in more traditional parts of the transportation network, we do not claim that TNC networks are "worse" than the status quo. In fact, in the Seattle experiment we had RAs hail taxis and count the the number of empty taxis passing them by. There is a clear difference in the acceptance rate of traditional taxes. The first taxi stopped nearly 60% of the time for white RAs, but less than 20% of the time for African American RAs. The white RAs never had more than four taxis pass them before one stopped, but the African American RAs watched six or seven taxis pass them by in 20% of cases.

Our work highlights several options for reducing discrimination within TNCs. For example, TNC networks could omit personal information about potential riders completely. Names and photos of both passengers and drivers could be replaced with user-specific numbers. Confirmation of these numbers by the driver and passenger could then occur prior to the trip. While this would reduce the type of discrimination found in this paper, other channels of discrimination would remain. Most notably, it would not eliminate the ability of drivers to discriminate in where they choose to drive.

The paper unfolds as follows: In section 2 we describe the set up and results of the Seattle experiment. In section 3, we do the same for the Boston experiment. Section 4 concludes the paper.

2 Experiment 1: Seattle

Our first study tested for differences in the quality of services received by African American and white passengers using TNCs and taxis in Seattle, Washington. African American and white research assistants (RAs) used UberX, Lyft, Flywheel (app-based taxi hailing), and taxis hailed from the curb to traverse assigned routes within the city of Seattle over six weeks in August – September, 2015. We tested for differences between races and sexes in several measures that could indicate discrimination. These included measures of the speed of service, the directness of the route taken by the driver, and the cumulative "star ratings"

received by the passengers.

2.1 Experimental Design

We designed seven tours around the city of Seattle, each starting and ending at the University of Washington's Seattle campus and comprising a sequence of pre-determined stop locations linked by individual trips. The stops were located to generate variability in neighborhood characteristics (population density, percentage of residents who are African American, income level), while limiting the individual trips to roughly the distance corresponding to the UberX and Lyft minimum fares. The routes are mapped in Appendix Figure A.1, which also shows selected socioeconomic characteristics at the census block group level. The tours generally took between one and three hours, and were completed following the evening rush hours on Monday through Thursday evenings.

In the first four weeks of the study, we assigned the RAs to rotate between services in the general order UberX-Lyft-Flywheel, after starting each tour with a randomly specified service. At selected stops in downtown Seattle, they were directed to hail a passing taxi from the curb (hailing a taxi from the curb is not feasible in other areas of Seattle, due to a low density of taxis). In the final two weeks of the study, we stopped collecting data on Flywheel, and the RAs alternated between UberX and Lyft (while still hailing taxis from the curb at specified downtown stops).

To avoid confounding the effects of race, sex, and other variables, we generated a fractional factorial experimental design. The variables and levels used in the experimental design are summarized in Table 1. In this way, we produced a list of tours to be completed on specific days of the week, by travelers of a particular race and sex, beginning with a specified service.

2.2 Data Collection

We began data collection with eight RAs: two African American females, two African American males, two white females, and two white males. All of the RAs were University of Washington undergraduate students. We presented the RAs with a list of dates on which the experimental design dictated that a traveler of their race and sex should travel, and they signed up for specific travel days. Each RA completed no more than one tour in a day.

The RAs used smartphones to request rides from UberX, Lyft, and Flywheel, and to log data. We issued each RA an identical smartphone using the same mobile carrier and data plan to minimize variation in factors such as communication latency. The RAs set up passenger accounts with Uber, Lyft, and Flywheel on these smartphones, and included their name and a profile photo with each account (Flywheel did not support profile photos). Profile photos were taken during the RA training session, and consisted of a headshot of each RA with a neutral facial expression, in front of a plain white background.

The RAs logged key information by taking screenshots on their smartphones. We installed

an app on each smartphone that displayed the time including seconds, so we could easily read the precise time in each screenshot. For each trip, we instructed them to take four screenshots:

1. Immediately before requesting a trip. This captures the time when a trip was requested, and the estimated waiting time for the passenger to be picked up (displayed in the TNC app).

2. Immediately after a trip request was accepted by a driver. This captures the time when the trip was accepted and a revised estimated waiting time for the passenger to be picked up.

3. When the driver arrives to pick up the RA. This provides the actual pickup time.

4. When the car stops to drop the RA off at the requested destination. This captures the actual dropoff time.

The RAs took notes on additional relevant information that arose in the course of their tours, such as deviations from the prescribed experimental plan, cancellations by drivers, problems with data collection, or practical challenges with the prescribed stop locations. For taxis hailed from the curb, the RAs took screenshots when they began trying to hail a cab and when a cab stopped for them. They also kept a count of how many taxis passed by them before one stopped and took a note of this on their phone.

The RAs transcribed key data from their smartphones into a spreadsheet at the end of each tour. The screenshots were deleted from the smartphones after transcription because some of them contained personally identifiable information (names and photos) of the drivers. In addition, we obtained the price and distance traveled for each trip from receipts that were automatically emailed to the research team at the end of each trip. We also deleted the emails containing driver names and photos.

To minimize differences in how the RAs conducted the data collection, we conducted a two hour group training session the week before data collection began. The RAs familiarized themselves with the smartphones and the Uber, Lyft, and Flywheel apps. We provided instructions on the data collection process and on how to interact with drivers. We instructed the RAs to sit in the back seat, minimize their interactions with the drivers, and not indicate that they were collecting data. We also instructed them to enter their destination information via the apps after the driver arrived to pick them up, and if asked, to request that the driver simply follow the navigation app linked to their TNC platform.

There were nevertheless some sources of variability and deviations from the original experimental design. First, due to the RAs' scheduling constraints, we allowed the them to complete some tours on different days than originally prescribed (e.g., a day earlier or later), but all the trips were completed between Monday and Thursday. Second, we asked the RAs to begin their tours at approximately 6:30 PM each day, but allowed them flexibility

to start 30 minutes before or after this time. Third, we gave the RAs some latitude to adapt their data collection on the fly. For example, if dynamic pricing ("surge pricing" for Uber or "primetime" for Lyft) increased TNC fares to more than 2.5 times their base rate, we asked the RAs to switch to another service for that trip. The RAs also switched modes if they faced repeated cancellations or were otherwise unable to complete a trip by the prescribed service. Finally, approximately three weeks into the study, one of the African American male RAs was unable to continue with the project. We hired another African American male RA to take over his smartphone and TNC accounts (after replacing the profile photo with a photo of the new RA) and complete his remaining trips.

2.3   Empirical Strategy

Our empirical strategy focused on identifying differences in the quality of service received by the African American and white RAs when traveling by different transportation services. Here, we explain our various measures of quality of service, and how we tested for differences between travelers of different races.

We used several different metrics to quantify the quality of service received by the RAs in this study, and which we would expect to be different if discrimination were occurring. Several measures of time are illustrated in Appendix Figure A.2, and include:

- **Acceptance time** is the length of time that passed between when an RA sent a trip request and when that request was accepted by a driver.

- **Estimated waiting time 1** is the waiting time that was displayed in the app immediately before the RA requested a ride.

- **Estimated waiting time 2** is the waiting time that was displayed in the app after the trip request was accepted.

- **Actual waiting time** is the length of time that passed between when a trip was accepted and when the driver arrived to pick up the RA.

- **Travel time** is the length of time that passed between when the driver arrived to pick up the RA and when the RA was dropped off at the end of the trip.

We might expect to see differences in *acceptance time* between races if drivers discriminate. In the case of Lyft, which showed the passenger's profile photo to the driver as part of the trip request, a driver's likelihood of declining the request might be affected by the traveler's apparent race. In the case of Uber, which showed the passenger's first name to the driver *after* the trip is accepted, this seems less likely. However, it is possible that some drivers might accept and then quickly cancel a trip after seeing the passenger's name, behavior that some drivers have advocated in online forums (UberPeople.Net, 2015). Indeed, we adapted the Boston experimental design to be able to directly test for this. In either

case, the request would then be passed to another driver, which would lead to a longer delay between the time the passenger requested a trip and when that trip was ultimately accepted.

We might also see differences in *estimated waiting time 2* and *actual waiting time* if drivers discriminate. Let us first assume that a trip request is reliably routed to the nearest driver and then to progressively further drivers if the closer driver(s) decline (or accept then cancel) the request. If at least some drivers tend to decline requests from certain groups, then on average those groups would be matched with drivers who are further away, and would end up waiting longer for a car to pick them up. On the other hand, if the matching of requests to drivers is fuzzier, this effect would be attenuated, since declined requests would sometimes be rerouted to drivers who are actually closer than the original driver. We would expect the effect to be strongest for actual wait times, which is our focus below.

The purpose of this experiment is to test whether racial discrimination is present in TNC services. The "treatment" in this context is race. We used two strategies to analyze the Seattle data. First, we used linear regression to test for differences in several measures of quality of service. Second, we constructed matched pairs of trips by African American and white RAs and tested for differences in the quality of services between these matched groups. In all cases, we analyzed the data for UberX, Lyft, and Flywheel separately.

The general form of our regression model is shown in equation 2.1, for some outcome variable y (e.g., Acceptance Time, Wait Time, etc.) for RA i at location j in situation t. $Race_i$ is a dummy variables indicating RA i's race, and $\mathbf{X_{ijt}}$ is a vector of covariates unique to the individual, location, and/or situation. For example, $\mathbf{X_{ijt}}$ can include neighborhood characteristics or the estimated waiting time associated with the trip request.[2] This yields:

$$y_{ijt} = \beta_0 + \beta_1 Race_i + \gamma \mathbf{X}_{ijt} + \epsilon_{ijt}. \tag{2.1}$$

As a robustness check we also present results that control for covariates through balancing. Specifically, we constructed datasets comprising of pairs of observations (one by an African American RA, one by a white RA) that were matched on location, time of day, and the RA's sex. We required exact matching on sex and used Mahalanobis distance (Mahalanobis, 1936) to measure the closeness on time of day and location. We then matched each observation with the nearest observation of a trip by an RA of the same sex and the other race. We also included "calipers" that put an absolute maximum on the differences between the time of day and location of matched pairs. We checked to see that covariates were balanced between the matched sets, and tested for differences in service quality using paired t tests.

---

[2]In particular, we include demographics from the Census Block Group of the pickup point using 2010 data. We include three indicator variables. High Income is defined as median household income greater than $75k annually (top 25%), High Pop Density is defined as fewer than 6750 people per square mile (lower 25%), and High AA Pop is defined as more than 1890 African American people per square mile (top 25%).

2.4 Results

We collected data on 581 app-hailed trips in total: 208 Uber trips, 222 Lyft trips and 143 Flywheel trips. Among these trips, 155 were finished by African American female students, 152 by white female students, 129 by African American males, and 145 by white males. The fact that fewer trips were taken by African American males compared to other groups is due to the aforementioned replacement of one African American male RA with another halfway through the data collection process.[3,4] We first present the results and then interpret them as a whole at the end of the sub-section.

Table 2 provides summary statistics for acceptance time, estimated waiting times, and actual wait times of Uber, Lyft and Flywheel. The number of observations for the acceptance time and actual waiting time of the three modes are different because some observations were deleted due to the following reasons: (1) the RAs did not take a screenshot immediately before requesting a trip, immediately after a trip request was accepted or immediately when the driver arrived; (2) the times on the screenshots are missing or inaccurate because the clock was fully or partly blocked on the phone; (3) the RAs made typos during data entry (e.g., the time right before requesting a trip is later than the time right after the request was accepted). Some observations of the estimated waiting time 1 were deleted because the estimated waiting time did not show up when the service was busy. Comparing the three services, Lyft requests are accepted more quickly, on average, but Uber has the shortest average waiting time. Flywheel is the slowest when it comes to both trip acceptance time and waiting time.

We first test whether there are systematic differences in the *estimated* wait times reported by the services' apps. This is a test of whether there were systematic differences in the types of trips that our African American research assistants hailed. Table 3 summarizes linear regression models of the log of *estimated waiting time 1* before trip requests were initiated, for UberX, Lyft, and Flywheel. The results indicate that before trips were requested, there were no significant differences in the estimated waiting times presented to African American and white RAs.

Table 4 summarizes linear regression models of the acceptance times in Seattle for UberX, Lyft, and Flywheel. We first present a simple test across means, then we control for the *estimated waiting time 1* (the waiting time displayed in the app prior to the trip being requested) to control for situation-specific factors that would lead to a longer or shorter response time by drivers; then, we control for the income, population density, African American density,

---

[3]Appendix Figure A.3 shows the balance in the days, times, and locations (tours) of the experiments actually done by the African American and white RAs. The distributions of the day of week and time of day are similar between the races, but white males completed more trips on tour 4 than African American males. Since tour 4 goes through downtown Seattle and the Capitol Hill neighborhood, where travel demand is higher and Uber service supply is presumably higher, this imbalance in routes could lead to differences in waiting times.

[4]We also collected data on 36 taxi trips in which the taxi was hailed from the curb. We come back to this at the end of the paper.

crime rate, and the tours that the trips were on. Note that to conserve on power we include tour fixed effects instead of specific pickup-dropoff pairs. We analyze the log of the time measures in order to improve the normality and homoskedasticity of residuals, and allow for the interpretation of effects as fractional changes in acceptance time.

For each of the services, African American riders had longer acceptance times compared to white riders. The effects are statistically significant for each of the four Uber specifications and the four specifications for Lyft. Only when tour and area covariates are controlled for do we see a significant difference of Flywheel acceptance time between African American and white RAs.

Next we turn to the actual waiting times, defined as the time that passes between when the trip request is accepted and when the car arrives to pick up the traveler. Table 5 presents the results of the (log) actual waiting time for UberX, Lyft, and Flywheel. African American travelers waited approximately 30% longer to be picked up than did white travelers when using UberX, even after adjusting for differences in estimated waiting time. However, there is no significant difference in waiting times between African American and white travelers using Lyft or Flywheel.

We also analyze whether the behavior of drivers varied across the race of the passenger after the trip had started. Specifically we look at both the travel time and travel distance. Table 6 shows regression models of the log of travel time for UberX, Lyft, and Flywheel. For each service, we show the ordinary least squares (OLS) regression as well as quantile regressions for the 75th and 95th percentiles. The quantile regressions test for whether the effects are concentrated on fewer very long trips. For UberX and Lyft, the travel time as recorded on the receipt is used as the dependent variable. Flywheel receipts did not include travel time or distance, so the travel time was calculated based on the pickup and dropoff times recorded by the RAs. The travel time between the assigned origin and destination was determined using the Google Maps API. The results indicate that the mean UberX travel time was about 8% longer for African American RAs than for white RAs, after adjusting for differences in the length of the assigned trip. Moreover, the 95th percentile travel time was about 16% longer for the African American RAs than for the white RAs. This is consistent with a pattern in which a small number of drivers is causing the longer travel times for the African American RAs. Travel times by Lyft and Flywheel were not significantly different for African American and white travelers. However, the 95th percentile travel time for female travelers in Flywheel was significantly longer than for males. This suggests that some taxi drivers may have taken female passengers for longer rides than males on similar trips although there is no consistent evidence to this effect.

Table 7 summarizes models of travel distance for individual trips by African American and white travelers, based on the distances noted on Uber and Lyft receipts. As with travel time, the expected distance from origin to destination was based on the Google Maps API. There was no significant difference in travel distances between African American and white travelers, except for the 95th percentile distance for UberX. We did not analyze trip distances for Flywheel since Flywheel receipts do not include trip distance. We come back

9

to the difference across platforms below.

The linear regression models control for covariates through adjustment, which assumes a linear relationship between the dependent and independent variables, which may or may not be justified. As a robustness check, we constructed matched pairs of trips by African American and white travelers of the same sex, requested not more than one hour apart, from locations not more than one mile apart. Using this approach, we estimated the effects of race on acceptance time and actual waiting time. The results are summarized in Appendix Tables A.1 through A.4, and indicate that our African American RAs waited significantly longer than their white counterparts for UberX trips to be accepted and for an UberX vehicle to show up.[5]

2.5 Interpretation and Lessons Learned

We summarize the above results as follows. There is some evidence that it takes longer for trip requests from African Americans to be accepted across all of the platforms. The results for *estimated* wait times suggest that this is not driven by systematic differences in routes or the timing of requests. We also find statistically significant longer wait times for African Americans riding UberX. These effects are as large as 35 log points. We find no significantly different wait times for either Lyft or Flywheel. Furthermore, we can rule out effects as large as those estimated for UberX. Finally, African American travelers had (marginally) significantly longer travel times than white travelers using UberX, but travel distances were not statistically significantly different. Neither travel times nor distances were significantly different for African American and white Lyft travelers.

Taken as a whole, it would appear that there is evidence of racial discrimination among UberX drivers, some evidence of discrimination about Lyft drivers, but no evidence of discrimination among Flywheel drivers. What can explain the stark difference among Flywheel drivers? There are two plausible explanations. The first is that the Flywheel service does not include photos of travelers in their profiles and so cannot present these to drivers. Therefore racial discrimination would have to be based on the names of riders. While some of the names of our African American RAs may have provided drivers with a signal of race, we did not design the experiment to specifically test for this. Second, Flywheel works with existing taxi drivers. It is possible that the subset of taxi drivers who opt into using it are less inclined to discriminate than those who do not opt in. Perhaps taxi drivers inclined to discriminate find it easier to do so by looking at would-be passengers on the street.

We studied the behavior of taxi drivers by having RAs hail a taxi from the curb in downtown Seattle. We instructed RAs to count the number of empty taxis that passed

---

[5]We also analyzed the cumulative star ratings of each RA for UberX and Lyft. Appendix Figures A.4 and A.5 present these results. The average star ratings given to African American and white travelers are very similar, indicating that the drivers who accepted the trips and provided star ratings did not provide better or worse ratings based on the RA's race. This may not be surprising if discrimination is operating through the cancellation of trips. If this is the case the ultimate driver is not responsible for the longer wait times.

by. Appendix Figure A.6 shows the distribution of the number of taxis that passed. The first taxi stopped nearly 60% of the time for white RAs, but less than 20% of the time for African American RAs. The white RAs never had more than four taxis pass them before one stopped, but the African American RAs watched six or seven taxis pass them by in 20% of cases. Based on the Wilcoxon rank-sum test, the difference in the number of taxis passing by African American and white RAs is highly significant ($W = 48$, $p = 0.0007$).

Our results are consistent with both UberX and Lyft drivers discriminating in the process of accepting ride requests. On Lyft, which shows drivers a photo of the traveler along with the trip request, the opportunity for discrimination is obvious. With Uber, which shows only a name, and only after a request has been accepted, the opportunity is less obvious. However, it is possible that some drivers might accept a trip request and then quickly cancel it after seeing the passenger's name, as some Uber drivers have advocated in online forums. In fact, we altered the experimental design in Boston to directly test for this. We discuss this below.

While we find statistically significantly different wait times for UberX riders, we did not find differences in wait times for Lyft. The longer wait time for an UberX is consistent with the longer acceptance time. If some drivers are accepting then canceling rides, then new drivers have to be assigned, which takes more time. If Uber initially sends the trip request to the nearest driver, then the reassigned driver will tend to be farther away, increasing the waiting time. Why, then, do we not see African American travelers also waiting longer for a Lyft pickup? Admittedly, this is still an open question. One possibility is that Uber and Lyft use different algorithms to match drivers and passengers. If drivers are assigned based on criteria beyond just proximity to the passenger (e.g., star ratings, amount of time driver has been idle) then the reassigned driver may actually be closer than the original driver, attenuating the effect of discrimination on waiting time. Another possibility is that Lyft gets any discrimination out of the way immediately, so the request can be passed quickly on to another driver. However, Uber drivers might be waiting for some time before ultimately canceling, and only then does the trip request has to go to a new driver, creating an additional delay in pickup.

3   Experiment 2: Boston

3.1   Experimental Design

Our experimental design in Seattle revealed a number of potential limitations to the experiment, which we used to inform the design of the data collection in the Boston study. Some of these we have already discussed. One that we have not discussed is that it is conceivable that differences in measured acceptance times or waiting times might be due to differences in how individual RAs logged their data and this was somehow correlated with race. For example, perhaps African American RAs simply took an extra second or two between taking

11

their screenshots and sending the trip request, and between trip acceptance and taking their second screenshot. We doubt this is the case since we would expect to see this consistently across all platforms, yet we did not see any difference in acceptance times between African American and white travelers when using Flywheel. Moreover, this cannot explain the larger difference in average waiting time (roughly 90 seconds) observed between African American and white passengers using UberX.

A second limitation is that it is possible that differences between African American and white passengers using UberX were due to some drivers having trouble identifying the African American passengers at the pickup points. If, for example, drivers were not expecting an African American passenger, then it might take them longer to see the passenger and drive up to them. This could explain why the pickup times were longer for African American passengers on UberX, even if there were no overt discrimination; although the driver arrives for the pickup in the same amount of time, they might spend more time looking around for the passenger. This might also explain why the travel times for UberX were longer, although the travel distances were not; UberX might record the trip as starting when the driver arrives at the pickup location, even if drivers sometimes spend a little extra time looking around for their passengers. This could also explain why these effects were not detected on Lyft; the drivers have a photo of the passenger from the outset so they know exactly who they are looking for.

Finally, as was mentioned, we did not design the experiment to understand the precise mechanism for discrimination by drivers receiving an UberX trip. As noted, drivers do not receive any information about the passenger until after they accept the request, so it would seem that if they were discriminating, they would need to cancel an accepted trip. Since we did not originally anticipate the possibility of drivers accepting then canceling trips, we did not provide the Seattle RAs with clear instructions about logging cancellations. Sometimes the RAs noted that a cancellation had occurred, but we are not confident that they did so in all cases. It is possible that a driver could cancel and UberX could assign a new driver, without the RA noticing.

We made two major changes to the experimental design for the Boston study. The first is that we designed our study in Boston to use within-RA variation in race to eliminate differences in data collection practices between travelers. To accomplish this requires that the same individual register for two different UberX profiles and two different Lyft profiles: one with an "African American sounding" first name and one with a "white sounding" first name.[6] Furthermore, we recruited students with a range of ethnic backgrounds, but whose appearance allowed them to plausibly travel as a passenger of either race. The second change was that we instructed the RAs to watch vigilantly for cancellations. As noted above there is active discussion on driver forums (e.g., UberPeople.Net (2015)) about whether cancellations

---

[6]Students were issued two identical phones, each with UberX and Lyft applications installed and with a travel profile under the assigned pseudonym. To reduce the likelihood that students behaved differently under one profile or another, neither pseudonym was related to the traveler's true name. This had the additional benefit of preserving the travelers' anonymity for the duration of the project.

that are performed quickly are shown to a customer. If drivers can cancel quickly and not appear on a customers screen, then measurements of cancellations by students should be treated as a lower bound, and actual cancellations could be higher than those reported. A third, less substantive change was that due to the increased focus on cancellations, we turned our focus to the largest TNC services UberX and Lyft, and did not perform tests of FlyWheel or street hails in Boston.

3.2 Data Collection

Eight individuals traveled as part of the experiment in Boston, using the pseudonyms shown in Table 8. These pseudonyms were taken from lists of names developed by Bertrand and Mullainathan (2004) that had been strongly identified as African American or white by panels of observers. Paying for the rides presented a challenge, since travel profiles were not real individuals and did not have credit cards. Payment was initially made using institutional purchase cards, but rapid accumulation of Uber and Lyft transactions on a few credit cards raised fraud flags by issuing banks, Uber, and Lyft. To avoid account suspension and potentially stranding travelers in remote parts of the city, we extended payment methods to include their personal credit cards, and eventually PayPal accounts under the pseudonyms issued to the traveling students. However, even this became a challenge as both the accounts under the pseudonyms were canceled by Uber. Ultimately, these challenges led the experiment to be terminated prior to our the planned termination date. The experiment ran from Nov. 5, 2015 to March 3, 2016.

3.3 Site and Route Choices

As in Seattle, we developed specific travel tours that consisted of a list of stops with specific pickup and drop off points. These tours ensured that we could control for the demographics of the pickup location and the expected distance and duration of each trip. We performed a GIS analysis of the Boston area, shown in Figure A.7. In determining which location attributes to include in the analysis, we hypothesized that drivers would accept or reject rides based on a mixture of fear and greed. For example "What is my expected revenue for this and future rides, and what is my fear that harm will be done to me by this passenger or others in the pickup location?" We examined pickup locations for household income, population density, measures of minority population density, and the presence of a transit stop at the point of pickup. We used the demographics at each location to help design tours as described below.

The final research design used five unique tours constructed with multiple objectives in mind: (i) start and end near the same place, (ii) limit the distance and duration of each trip to near the minimum fare threshold to reduce overall cost, (iii) induce as much variation in pickup location demographics as possible, (iv) minimize overlap with other tours to reduce likelihood of multiple travelers affecting performance for each other, (v) where possible, to

ensure pickup locations were within sight of a business open until at least 9pm.

Rather than indicating the pickup location on a map, students were instructed to enter the complete address of the pickup location and destination to prevent small variations in pickup location based on individuals. Since students were traveling late in the day and sometimes in bad weather, in a few instances students requested to end the last trip of the tour at their home address. These trips, and trips where students realized they had entered the origin or destination address incorrectly, were excluded from trip distance and duration analyses.[7]

3.4 Traveler Instructions

While the research design includes fixed-RA effects to control for individual variations in behavior (using the UberX and Lyft apps or at the point of pickup) student travelers followed specific instructions in requesting and taking rides in a similar fashion to our Seattle experiment. These instructions were to: (i) enter the complete address of both the pickup location and the destination when requesting a ride to minimize variation in origins and destinations, (ii) minimize interaction with the driver to (A) lower the chances that a driver would recognize them, and (B) to minimize any differences in behavior that would change star ratings, (iii) not to cancel rides unless ride was input incorrectly, (iv) not to give drivers directions unless requested, and (v) to leave five-star ratings unless they felt the driver was threatening or dangerous.

After several initial travel days looking for cancellations, students identified an unanticipated form of driver behavior. Some drivers would accept a ride and then apparently not attempt to pickup the passenger—either not to move at all or even to drive in the direction opposite the traveler. Since this scenario could result in students waiting indefinitely for a ride in unfamiliar locations, we established a threshold of 20 minutes. If after 20 minutes the driver had not made any indication of attempting to pickup the passenger—either contacting the passenger or driving measurably closer to the pickup location—students canceled the ride and flagged the first attempt as a *de facto* cancellation by the driver.

---

[7]We were aided by the Seattle experiment for calculating statistical power. We calculated the sample size required to detect a variety effects on the means of the variables described above using the summary statistics from Seattle. We calculate the samples sizes required to detect a 5, 10, and 15 percent increase in the variables of interest. We used the UberX sample for our calculations and assume an α of 0.05 and power equal to 0.90. These calculations implied sample sizes of roughly 1,700, 400, and 200 for detectable effects with actual wait times of 5, 10, and 15 percent, respectively. The full results are reported in Appendix Table A.5. To err on the conservative side, we performed these calculations without including any covariates. In the next section we describe what covariates we included to gain some statistical power. Based on these calculations we targeted 1000 data points for the Boston experiment. Because of the early termination of the data collection in Boston we ended with 911 trip requests.

3.5 Empirical Strategy and Analysis

As with the Seattle experiment, the purpose of this experiment is to test for the presence of racial discrimination. In the Boston experiment, treatment is whether the rider has what Fryer and Levitt (2004) refer to as "distinctively black names." Since each of our student travelers used two pseudonyms are able to include RA-specific fixed effects in our econometric specifications. The starting point for our empirical analysis is to estimate differences in the within-RA means of the variables described above across implied race. More specifically, for a given outcome, y (e.g., acceptance time, waiting time, etc.), for RA i, using name k, at location j, at time t, our regressions take the form of:

$$y_{ikjt} = \beta_0 + \beta_1 ImpliedRace_{ik} + \mu_i + \epsilon_{ikjt}. \tag{3.1}$$

In some specifications we will omit the RA fixed effects to test for differences across genders. This yields the following variant of equation 3.2:

$$y_{ijkt} = \beta_0 + \beta_1 ImpliedRace_{ik} + \beta_2 ImpliedGender_i + \epsilon_{ijkt}. \tag{3.2}$$

We will also estimate specifications that control for covariates.[8] These covariates pick up location, trip, and day-of-week effects. We therefore augment equation 3.2, yielding:

$$y_{ijkt} = \beta_0 + \beta_1 ImpliedRace_{ik} + \gamma \mathbf{X_j} + \mu_i + \epsilon_{ijkt}, \tag{3.3}$$

$$y_{ijkt} = \beta_0 + \beta_1 ImpliedRace_{ik} + \beta_2 ImpliedGender_i + \gamma \mathbf{X_j} + \epsilon_{ijkt}. \tag{3.4}$$

3.6 Results

We collected data on 911 total trip requests in Boston: 460 trips with Lyft and 451 trips with UberX. Drivers canceled 66 requests (57 actual cancellations, and 9 *de facto* cancellations in which the driver started the trip without the student or made no apparent pickup attempt). Students canceled 6 trips after recognizing they made input errors in the requests. After excluding cancellations, 839 completed trips remained.

Of the 839 completed trips, for estimates of acceptance and waiting times we removed samples where travelers failed to capture a screenshot of one or more timestamps during the trip. 80 trips were removed, leaving 759 trips in which all valid timestamps were observed—372 using UberX, and 387 using Lyft. Of these 759 observations, 46 trips had at least one

---

[8] In particular we include indicator variables for whether the pickup location has a high African American population, a high median income, and a low population density. Demographics were assigned by Census Block Group of the pickup point using 2010 data. Upper Black Pop Quartile is defined as 20% or greater African American population, Upper Income Quartile is defined as median household income greater than $60k annually, and Lower Population Density Quartile is defined as fewer than 12800 people per square mile.

prior cancellation, leaving 713 observations for Accept Time and Wait Time analysis, 364 using Lyft and 349 using Uber. Table 9 provides summary statistics for request times, wait times, and estimated waiting times for UberX and Lyft in Boston. In cases where a driver cancellation occurred, students recorded the time of the first trip request as the request time. As a result, both Lyft and UberX show large standard deviations in Acceptance Times, and median Acceptance Times are far lower than means for both services.

Table 9 and Figure A.8 shows a descriptive comparison of Acceptance Time, Estimated Waiting Time and Actual Waiting Time between African American and White passengers for the UberX and Lyft services. In contrast to results of the Seattle study, waiting times and acceptance times were not substantially longer for African American passengers than for white passengers in either UberX or Lyft; Acceptance Time, Actual Waiting Time 1 and Waiting Time Increase measures were broadly similar across racial groups, and slightly shorter for African American passengers using Uber.

As with the Seattle experiment, we begin by confirming that estimated wait times prior to calling a ride do not systematically vary by name. Table 10 confirms this. Next we turn to cumulative distributions of Acceptance Time, Actual Waiting Time and Waiting Time increase in Figure A.8. As with the basic statistical measure, CDFs do not include broadly different behavior between demographic groups. Table 11 shows the results of regressions for Acceptance Time in Boston, and Table 12 shows the results of Actual Waiting Time in the Boston experiment.

As predicted by initial sample size estimates, few coefficients of statistical significance can be found in either analysis. For some specifications, Acceptance Times are longer for African American males but with large standard errors. As expected, Actual Waiting Times are highly dependent on Estimated Waiting Times, and all else equal Actual Waiting Times will be lower when the service provider is UberX.

New to the Boston experiment was the ability to analyze cancellations. All 911 observations were used in evaluating cancellation rates – 460 observations using UberX and 451 observations using Lyft. Since UberX drivers only see the name of passengers after accepting a ride, we expect to observe differences in behavior only in UberX. If Lyft drivers were to discriminate based on the name, they would simply ignore the ride request, an action which is not directly observable.

We first tabulate the rate of cancellations of trips by service, race, and gender. The results are shown in Table 13. The simple summary statistics suggest that cancellations against passengers using African American names are substantially higher for trips with UberX, but not for Lyft. For all passengers on UberX, those using African American-sounding names face more than double the cancellation rate than when the same passengers use white-sounding names (10.1% vs. 4.9%). Furthermore, this effect appears larger for African-American male names than for African American female names.

When using UberX, the difference is greater for males than females. For males, those using African American-sounding names face a cancellation rate more than twice as high as those same individuals when using white-sounding names – 11.2% cancellation rates for males

when using African American-sounding names compared to 4.5% when the same individuals use white-sounding names. The difference is nearly as bad for female riders: the cancellation rate for females is 8.4% when using a African American-sounding name and 5.4% when using a white-sounding name. For Lyft, males face approximately the same cancellation rate, whereas females actually face a lower cancellation rate when using African American-sounding names.

Next we use regression to gauge statistical significance and to control for a broad array of demographic variables for the pickup locations, individual fixed effects for the passengers, and fixed effects for the day of the week. The results from these regressions are shown in Table 14. We start by analyzing only the means. For Lyft, we find no statistical or economically significant difference across names. The treatment effect for UberX is significant and suggests a more than doubling of the probability of a cancellation for riders using African-American sounding names consistent with the cross-tabs above.

We expand the analysis in two dimensions. First, we include a variety of controls, such as demographics of the pickup location and day of week. Second, we interact the treatment with gender. For completeness we allow the caucasian baseline cancellation probability to also vary by gender. Column 2 presents these results for Lyft. Again, we find little evidence of an increase in cancellations across race or gender. This again serves as a useful null test since drivers can observe the prospective passenger's name prior to accepting. Column 4 reports the results for UberX. We find that discrimination appears to be focused on African American males. The cancellation rate faced by travelers using an African American male name more than doubles (139% increase) relative to a white-male sounding name. To gauge the robustness of this result to RA-specific behavior, Column 5 includes RA fixed effects. The estimate on African American male dummy is robust to the inclusion of these.

We also investigated whether discrimination was isolated to specific pick-up areas (columns 6 & 7). We also interact the African American male indicator variable with the low population density dummy variable. Interestingly, we find that the increase in cancellations is concentrated in low population density areas. In fact, the cancellation rate of riders with African American names is no different outside of these areas. The effects in low population density areas is startling. In these areas, males with African American-sounding names face an increase in the cancellation rate of 15.7 percentage points; this is a more than tripling of the cancellation rate faced by white males.

We next turn to travel time and distance. Of the 839 completed trips, for estimates of travel distance, time, and cost we removed samples where travelers had entered an incorrect origin or destination address. 19 trips were removed for incorrect origin or destination addresses as entered by the traveler, leaving 820 samples with correct requested trip length and time. We compared the actual travel distance, time and cost as billed on the receipt compared to the travel distance and travel time as estimated on Google Maps between the origin and destination of each scheduled trip.

Table 15 shows the results of a regression of the log of actual trip distance (as billed on the UberX or Lyft receipt) compared to the log of estimated travel distance between

the origin and destination as predicted by Google Maps. Regressions are estimated using Ordinary Least Squares (OLS), Least Absolute Deviation (LAD) and Quantile regression using values of Tau of 0.75 and 0.95.

Google Maps estimates of distance for each route are strongly significant in all cases, with a coefficient very close to a value of 1. The Surge multiplier (or PrimeTime in Lyft trips) is weakly significant, indicating that drivers will travel farther with a passenger if the passenger is paying more than the standard rate.

Results indicate that female travelers are driven farther. OLS regression using the log of travel distance indicates that female travelers are driven approximately 5% farther. Quantile regressions using Tau of 0.75 and 0.95 indicate that the impact on female travelers increases monotonically with statistical significance on all but Tau of 0.95.

Table 16 shows the results of a regression of the log of actual trip duration in seconds (as billed on the UberX or Lyft receipt) vs the log of estimated travel time for the origin and destination as predicted by Google Maps. Regressions are estimated using Ordinary Least Squares (OLS) and Quantile regression using values of Tau of 0.75 and 0.95.

No coefficients are strongly significant, although as with travel distance coefficient on female travelers is positive in all cases, with marginal significance. Coefficient on surge multiplier is strongly positive, indicating that trips tend to last longer when the passenger is paying more than the standard rate.

This behavior was confirmed by anecdotal evidence from female passengers. Some excessive fares were the result of drivers who started the trip before picking up the passenger or ended the trip after dropping off the female passenger. Other female riders reported "chatty" drivers who drove extremely long routes, on some occasions even driving through the same intersection multiple times. As a result, the additional travel that female riders are exposed to appears to be a combination of profiteering and flirting to a captive audience.[9]

4 Conclusions

Transportation sharing networks such as UberX and Lyft bring a number of benefits to the economy, ranging from cheaper transportation options to employment opportunities for drivers. The digital platform of sharing networks can also reduce discrimination because it can control what drivers know about the passenger prior to pickup.

To better understand the degree of discrimination present in current ride sharing networks, we designed two large-scale randomized control trials, one in Seattle and one in Boston, where we varied the race or perceived race of passengers.

We found that African American travelers in Seattle experienced statistically significantly longer delay waiting for a trip request through UberX or Lyft to be accepted. African

---

[9]We also analyzed the cumulative star ratings of each RA for UberX and Lyft. Appendix Figures A.4 and A.5 present these results. As in Seattle, average star ratings given to African American and white travel profiles are very similar, indicating that the drivers who accepted the trips and provided star ratings did not provide better or worse ratings based on the RA's race.

American travelers also waited significantly longer for an UberX vehicle to arrive and pick them up than did white travelers, but waiting times were the same for African American and white travelers using Lyft. Finally, we note that UberX drivers see only a passenger's location and star rating before accepting or declining a trip request, and see the passenger's name after accepting. Lyft drivers can see the passenger's name and photo before accepting or declining the request.

Based on these findings, we theorize that at least some drivers for both UberX and Lyft discriminate on the basis of the perceived race of the traveler. Because UberX drivers see a passenger's name and photo only after accepting a trip, UberX drivers discriminate by canceling after accepting the trip and seeing the passenger's name whereas Lyft drivers simply choose not to accept a trip. The added step associated with discrimination among UberX drivers increases wait times.

Given these results, we adapted the Boston experiment to more closely focus on cancellations. We found that UberX drivers are nearly three times as likely to cancel a ride on a male passenger upon seeing that he has a "black-sounding" name. This effect is robust across numerous model specifications and seems to be driven primarily by behavior in areas with low population densities. In these extreme cases drivers are more than four times as likely to cancel on a African American male passenger than on a white male passenger. As expected, given that Lyft driver see passenger information prior to accepting a ride, we find no effect of race on Lyft cancellations. We also found that UberX drivers were more likely to cancel trips for passengers being picked up near subway stops, perhaps because a passenger at a subway stop is either a low-income passenger, or a subway stop indicates a multi-modal journey with a lower expected revenue.

Using the most direct measure (observed cancellations in Boston) there appears to be evidence that African American passengers receive worse service, compared to white riders, in TNC or ride-hailing based services such as Uber and Lyft. This discrimination is not the result of any policy by ride hailing providers, but rather the behavior of individual TNC drivers. It is important to note, however, that we compare service quality across African American and white passengers *within TNC services*. That is, we do not compare the relative amount of discrimination across TNC and traditional taxi-cab services. In our Seattle experiment we found that discrimination within traditional taxi-cab also exists. The first taxi stopped nearly 60% of the time for white RAs, but less than 20% of the time for African American RAs. Furthermore, white RAs never had more than four taxis pass them before one stopped, but the African American RAs watched six or seven taxis pass them by in 20% of cases. These differences are statistically significant.

5   Potential Modifications for TNC Service Providers

Aggregate measures of service delivery by TNC providers appear to do little to dissuade drivers from discriminating against passengers to whom they do not wish to deliver service,

or from driving differently based on the gender of the passenger. TNC providers could consider a number of additional measures to ensure fairness in service delivery:

1. Do not use names to identify passengers and drivers. Instead, provide a unique passcode that is given to the passenger and the driver to confirm the correct identity at the time of pickup. However, this could lead to unintended consequences for passengers' star ratings and subsequent ability to request rides, if discriminating drivers who would otherwise decline or cancel a request are instead compelled to transport the passenger.

2. To avoid extended rides given to female passengers, fixed fares could be offered between pickup and drop off points. This would eliminate additional fares paid by female passengers and provide a disincentive for drivers to travel farther than necessary.

3. Since cancellations likely cost travelers more time than declined requests, increasing the disincentives for driver cancellations, including *de facto* cancellations where the driver makes no attempt at a pickup, may be appropriate.

4. Perform periodic audits of driver behavior that appears to be discriminatory in nature.

5. Counterintuitively, if one accepts that a certain number of drivers are inclined to discriminate, and it is difficult to police such behavior, the Lyft model of providing a name and photo immediately might lead to less negative outcomes for travelers. This is because a declined request leads to a small increase in acceptance time, but is quickly re-routed to another nearby driver. Although this presumably leads to small (though undetectable in our data) increases in waiting time, this effect is likely smaller than the time wasted on a cancellation, which is the likely outcome if passenger information is not revealed until after the trip is accepted.

6. Ultimately, passengers might resort to creating accounts with nicknames that de-identify their race. This study identified the relative ease with which Uber and Lyft accounts with arbitrary names can be created. This simply requires an email address and a form of payment, the name of which doesn't necessarily need to match that of the rider account.

7. Finally, TNC services have a treasure trove of data that could also be analyzed for evidence of discrimination across the channels we investigate as well as others. While these data obviously would not be generated from an experimental setting, it would be possible to analyze both cancellation and wait times across passengers. These data could also speak to the heterogeneity in these effects across geography and driver characteristics. Finally, researchers could also analyze the areas where drivers choose to serve and how this affects wait times.

6   Tables and Figures

6.1   Figures

Figure 1: TNC ad featuring an African American rider

Figure 2: Potential sources of discrimination

	Before Ride	Requests	Fares	After Ride
Motives	*Maximize long-term revenue potential*	*Maximize short-term revenue; minimize per-trip risk*	*Maximize short-term revenue; obtain good rating*	*Advise other drivers of experience*
Information	• Surge Prices • Neighborhood Demographics • Hearsay	• Location • Star Rating • First Name	• First Name • Interaction with Passenger	• First Name • Interaction with Passenger • Gratuity
Action	Where do I drive?	Reject or Cancel	Route & Service	Star Rating
	Smart et al. (2015) Hughes & MacKenzie (2016)	Smart et al. (2015) Edelman et al. (2016) This Work	This Work	This Work

6.2 Tables

6.2.1 Seattle Experiment

Table 1: Attribute variables of experiment design

Attributes	Attribute levels
Routes	7 levels: 7 routes
Mode for first trip	3/2 levels: UberX, Lyft, & Flywheel in stage 1; UberX & Lyft in stage 2
Race	2 levels: black, white
Sex	2 levels: female, male
Day of the week	4 levels: Monday, Tuesday, Wednesday, Thursday
Week of data collection	4/2 levels: 4 weeks in stage 1, 2 weeks in stage 2

Notes: This table shows the number of routes table by research assistants, the mode selection, research assistant race and gender, and the frequency of trips taken for the Seattle experiment.

Table 2: Descriptive analysis of the Acceptance time, Estimated waiting time 1 and Actual waiting time of Uber, Lyft and Flywheel in the Seattle experiment

Service	Attributes	Acceptance time(s)	Est. waiting time 1 (min)	Actual waiting time (s)
Uber	Num.obs	190	200	187
	Mean	25	4.1	275
	S.D.	21	1.9	212
	Median	18	4	225
	Black Mean	29	4.3	316
	White Mean	21	3.9	240
Lyft	Num.obs	199	199	202
	Mean	23	3.0	284
	S.D.	24	1.8	156
	Median	16	3	270
	Black Mean	23	3.2	284
	White Mean	19	2.7	284
Flywheel	Num.obs	130	124	124
	Mean	44	6.2	446
	S.D.	69	3.5	263
	Median	28	5	381
	Black Mean	35	6.4	447
	White Mean	35	6.1	444

Notes: This table shows the number of trips taken and summary statistics for the time until the request was accepted (Acceptance time), the *estimated* wait time reported by the app when the trip was requested (Est. wait time 1), and the actual wait time from when the trip was requested (Actual wait time) for the Seattle experiment. In addition, we report the means of these three variables by race.

Table 3: Relationship between race and the log of *estimated* waiting times prior to requesting the trip in the Seattle experiment

	UberX	Lyft	Flywheel
(Intercept)	1.27***	0.85***	1.68***
	(0.04)	(0.06)	(0.08)
African American Traveler	0.08	0.12	-0.04
	(0.06)	(0.09)	(0.11)
R^2	0.01	0.01	0.00
Adj. R^2	0.00	0.01	-0.01
Num. obs.	188	183	121
RMSE	0.43	0.59	0.62

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Notes: This table shows the affect of race on *estimated* wait time prior to the trip request in the Seattle experiment. We find no systematic relationship between the service's estimated wait times and race.

Table 4: Models of the log of acceptance time for UberX, Lyft, and Flywheel in the Seattle experiment

	UberX			Lyft			Flywheel		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
African American	0.28***	0.24***	0.24***	0.24***	0.16*	0.17**	0.06	0.19	0.27**
	(0.05)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.14)	(0.12)	(0.13)
Log(ETA1)		0.22***	0.15*		0.15**	0.11		0.28***	0.30***
		(0.08)	(0.09)		(0.07)	(0.08)		(0.10)	(0.11)
High Income			−0.19**			−0.17*			0.26*
			(0.09)			(0.10)			(0.15)
Low Pop Density			0.08			0.29**			−0.16
			(0.11)			(0.12)			(0.20)
High AA Pop			−0.09			−0.27*			0.13
			(0.11)			(0.14)			(0.21)
Route2			−0.46**			−0.22			−0.25
			(0.20)			(0.22)			(0.30)
Route3			−0.47**			−0.10			0.23
			(0.18)			(0.21)			(0.29)
Route4			−0.38**			−0.12			0.08
			(0.18)			(0.20)			(0.29)
Route5			−0.56***			0.06			0.26
			(0.20)			(0.22)			(0.30)
Route6			−0.41**			−0.09			0.47
			(0.18)			(0.19)			(0.30)
Route7			−0.67***			−0.20			0.53*
			(0.19)			(0.22)			(0.32)
(Intercept)	2.92***	2.63***	3.30***	2.80***	2.68***	3.02***	3.34***	2.73***	2.24***
	(0.05)	(0.11)	(0.23)	(0.06)	(0.08)	(0.23)	(0.10)	(0.19)	(0.40)
R^2	0.07	0.10	0.18	0.04	0.05	0.13	0.00	0.08	0.16
Adj. R^2	0.07	0.09	0.13	0.04	0.04	0.07	−0.01	0.06	0.07
Num. obs.	190	188	186	199	183	182	130	121	120
RMSE	0.49	0.48	0.47	0.59	0.56	0.55	0.81	0.68	0.68

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Notes: This table shows regression models of the log of the time it takes until a trip is accepted for the Seattle experiment. We split the regressions by service (UberX, Lyft, and Flywheel). The first model is a simple difference in mean times across African Americans and white. The second model controls for the estimated wait time prior to the trip being requested. Model 3 includes route fixed effects, where routes are defined as a pre-determined chain of trips, and pickup-location characteristics to control for characteristics of the trip. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Income is defined as median household income greater than \$75k annually (top 25%), Low Pop Density is defined as fewer than 6750 people per square mile (lower 25%), and High AA Pop is defined as more than 1890 African American people per square mile (top 25%).

Table 5: Models of the log of actual waiting time for UberX, Lyft, and Flywheel in the Seattle experiment

	UberX			Lyft			Flywheel		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
African American	0.36***	0.29***	0.31***	0.00	−0.05	−0.06	−0.00	0.04	0.03
	(0.11)	(0.09)	(0.09)	(0.09)	(0.08)	(0.08)	(0.12)	(0.11)	(0.11)
Log(ETA1)		0.96***	0.88***		0.53***	0.48***		0.60***	0.53***
		(0.10)	(0.11)		(0.07)	(0.07)		(0.09)	(0.10)
High Income			−0.02			0.01			0.21
			(0.11)			(0.09)			(0.13)
Low Pop Density			−0.17			0.10			0.09
			(0.13)			(0.11)			(0.17)
High AA Pop			0.05			0.09			−0.11
			(0.13)			(0.13)			(0.18)
Route2			−0.34			−0.30			−0.23
			(0.24)			(0.21)			(0.25)
Route3			−0.27			−0.30			−0.28
			(0.22)			(0.20)			(0.25)
Route4			−0.30			−0.29			−0.28
			(0.21)			(0.19)			(0.24)
Route5			−0.17			−0.35			−0.07
			(0.24)			(0.22)			(0.27)
Route6			−0.21			−0.24			−0.20
			(0.21)			(0.18)			(0.25)
Route7			−0.25			−0.08			−0.01
			(0.23)			(0.21)			(0.27)
(Intercept)	5.21***	3.99***	4.32***	5.48***	5.02***	5.20***	5.91***	4.89***	5.15***
	(0.07)	(0.14)	(0.28)	(0.06)	(0.08)	(0.22)	(0.09)	(0.17)	(0.34)
R^2	0.06	0.36	0.37	0.00	0.26	0.29	0.00	0.30	0.35
Adj. R^2	0.05	0.35	0.33	−0.00	0.25	0.25	−0.01	0.28	0.27
Num. obs.	186	185	183	202	187	186	124	113	112
RMSE	0.72	0.59	0.59	0.62	0.53	0.53	0.69	0.57	0.58

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: This table shows regression models of the log of actual wait time from when the trip was requested to the time the research assistant was picked up in the Seattle experiment. We split the regressions by service (UberX, Lyft, and Flywheel). The first model is a simple difference in mean times across African Americans and white. The second model controls for the estimated wait time prior to the trip being requested. Model 3 includes route fixed effects, where routes are defined as a pre-determined chain of trips, and pickup-location characteristics to control for characteristics of the trip. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Income is defined as median household income greater than $75k annually (top 25%), Low Pop Density is defined as fewer than 6750 people per square mile (lower 25%), and High AA Pop is defined as more than 1890 African American people per square mile (top 25%).

Table 6: Models of the log of travel time for UberX, Lyft, and Flywheel in the Seattle experiment

	UberX			Lyft			Flywheel		
	OLS	Tau=.75	Tau=.95	OLS	Tau=.75	Tau=.95	OLS	Tau=.75	Tau=.95
African American	0.08**	0.09*	0.16**	−0.01	−0.06	−0.05	0.07	0.03	0.03
	(0.04)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.04)
Female	−0.05	−0.01	0.02	0.05	0.11**	−0.06	0.07	0.06	0.12***
	(0.03)	(0.05)	(0.03)	(0.05)	(0.05)	(0.12)	(0.05)	(0.05)	(0.01)
Log(Estimated travel time)	0.87***	0.81***	0.79***	0.87***	0.87***	0.84***	0.80***	0.86***	0.83***
	(0.06)	(0.09)	(0.09)	(0.07)	(0.05)	(0.19)	(0.08)	(0.09)	(0.04)
High Income	−0.05	0.01	−0.15**	−0.16**	−0.22***	−0.05	0.02	−0.01	−0.01
	(0.04)	(0.05)	(0.07)	(0.06)	(0.07)	(0.11)	(0.06)	(0.03)	(0.04)
Low Pop Density	−0.09*	−0.11*	−0.23**	−0.07	−0.16**	0.05	−0.11	−0.13*	−0.24***
	(0.05)	(0.06)	(0.11)	(0.07)	(0.07)	(0.14)	(0.08)	(0.06)	(0.05)
High AA Pop	0.06	0.04	0.19*	−0.01	−0.02	−0.16	0.04	0.02	0.11
	(0.05)	(0.07)	(0.11)	(0.08)	(0.08)	(0.12)	(0.09)	(0.06)	(0.10)
Route2	0.03	0.11	0.06	0.11	0.08	0.16	−0.34***	−0.39*	−0.05
	(0.09)	(0.20)	(0.10)	(0.13)	(0.33)	(0.27)	(0.11)	(0.21)	(0.08)
Route3	−0.09	0.07	−0.23	−0.02	−0.11	−0.11	−0.35***	−0.37***	−0.47***
	(0.09)	(0.17)	(0.16)	(0.13)	(0.25)	(0.15)	(0.11)	(0.10)	(0.10)
Route4	−0.05	0.07	−0.11	−0.04	0.01	0.08	−0.25**	−0.17**	−0.28***
	(0.08)	(0.16)	(0.12)	(0.11)	(0.24)	(0.16)	(0.10)	(0.07)	(0.08)
Route5	−0.12	0.01	−0.28*	0.03	0.01	0.11	−0.21*	−0.21*	−0.18**
	(0.09)	(0.17)	(0.17)	(0.13)	(0.25)	(0.15)	(0.12)	(0.10)	(0.09)
Route6	0.05	0.19	0.06	0.06	0.07	−0.28*	−0.12	−0.05	0.15
	(0.09)	(0.18)	(0.13)	(0.12)	(0.25)	(0.16)	(0.12)	(0.07)	(0.10)
Route7	0.03	0.19	0.10	0.07	0.04	−0.08	−0.08	−0.10	−0.05
	(0.09)	(0.18)	(0.21)	(0.13)	(0.25)	(0.20)	(0.13)	(0.14)	(0.13)
(Intercept)	4.41***	4.49***	4.85***	4.30***	4.54***	5.08***	4.74***	4.83***	4.97***
	(0.12)	(0.22)	(0.11)	(0.18)	(0.27)	(0.39)	(0.18)	(0.16)	(0.16)
R²	0.73			0.62			0.63		
Adj. R²	0.71			0.59			0.59		
Num. obs.	175	175	175	180	180	180	118	118	118
RMSE	0.23			0.32			0.27		

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Notes: This table shows regression models of the log of the travel time. We split the regressions by service (UberX, Lyft, and Flywheel). For Uber and Lyft, travel times are those recorded on receipts. For Flywheel travel times are calculated based on RAs' recorded pickup and drop off times. The estimated travel times of the three services are calculated using Google API according to the assigned origins and destinations. The first model is ordinary least squares regression. We control for both race and gender, and include route fixed effects, where routes are defined as a pre-determined chain of trips, and pickup-location characteristics to control for characteristics of the trip. The next two models are quantile regressions of the 75th and 95th quantile, respectively. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Income is defined as median household income greater than $75k annually (top 25%), Low Pop Density is defined as fewer than 6750 people per square mile (lower 25%), and High AA Pop is defined as more than 1890 African American people per square mile (top 25%).

27

Table 7: Models of the log of travel distance for UberX, Lyft, and Flywheel in the Seattle experiment

	UberX			Lyft		
	OLS	Tau=.75	Tau=.95	OLS	Tau=.75	Tau=.95
African American	0.03	0.05	0.20***	0.01	-0.00	-0.12
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.14)
Female	-0.06*	-0.02	-0.01	-0.01	0.02	0.01
	(0.03)	(0.01)	(0.03)	(0.03)	(0.02)	(0.16)
Log(Estimated distance)	0.90***	0.95***	0.98***	1.04***	1.01***	1.18***
	(0.06)	(0.02)	(0.05)	(0.05)	(0.03)	(0.17)
High Income	-0.04	-0.09**	-0.02	-0.00	-0.07***	-0.01
	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)	(0.18)
Low Pop Density	-0.05	0.01	0.07*	-0.07	-0.08***	-0.24
	(0.05)	(0.05)	(0.03)	(0.05)	(0.03)	(0.17)
High AA Pop	0.01	-0.03	-0.09	0.00	0.03	0.13
	(0.05)	(0.04)	(0.06)	(0.05)	(0.03)	(0.16)
Route2	0.03	0.05	0.23**	0.04	-0.03	-0.06
	(0.09)	(0.14)	(0.12)	(0.08)	(0.12)	(0.18)
Route3	-0.06	-0.05	-0.05	0.01	-0.07	-0.41*
	(0.08)	(0.10)	(0.13)	(0.08)	(0.12)	(0.22)
Route4	-0.14*	-0.01	-0.02	-0.00	-0.06	-0.41
	(0.08)	(0.10)	(0.11)	(0.07)	(0.10)	(0.30)
Route5	-0.02	-0.03	0.23	-0.03	-0.09	-0.60**
	(0.09)	(0.11)	(0.20)	(0.08)	(0.11)	(0.28)
Route6	0.04	-0.01	-0.11	-0.04	-0.06	-0.41**
	(0.09)	(0.10)	(0.11)	(0.08)	(0.11)	(0.17)
Route7	-0.07	-0.10	-0.17	-0.01	-0.08	-0.41
	(0.10)	(0.11)	(0.11)	(0.09)	(0.11)	(0.26)
(Intercept)	0.18**	0.23**	0.30**	0.02	0.15	0.60
	(0.09)	(0.11)	(0.12)	(0.08)	(0.11)	(0.37)
R²	0.81			0.84		
Adj. R²	0.79			0.83		
Num. obs.	185	185	185	194	194	194
RMSE	0.23			0.20		

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: This table shows regression models of the log of the travel distance. We split the regressions by service (UberX, Lyft, and Flywheel). For Uber and Lyft, travel times are those recorded on receipts. For Flywheel travel times are calculated based on RAs' recorded pickup and drop off times. The estimated travel times of the three services are calculated using Google API according to the assigned origins and destinations. The first model is ordinary least squares regression. We control for both race and gender, and include route fixed effects, where routes are defined as a pre-determined chain of trips, and pickup-location characteristics to control for characteristics of the trip. The next two models are quantile regressions of the 75th and 95th quantile, respectively. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Income is defined as median household income greater than $75k annually (top 25%), Low Pop Density is defined as fewer than 6750 people per square mile (lower 25%), and High AA Pop is defined as more than 1890 African American people per square mile (top 25%).

28

Table 8: Subject name and race assignment in the Boston experiment

Individual	Gender	Race Assigned	Name
1	F	Black	Aisha
		White	Allison
2	F	Black	Ebony
		White	Kristen
3	M	Black	Hakim
		White	Brendan
4	M	Black	Darnell
		White	Brad
5	F	Black	Keisha
		White	Anne
6	M	Black	Kareem
		White	Greg
7	M	Black	Rasheed
		White	Todd
8	F	White	Laurie

Research assistant names used in the Boston experiment. The first name is meant to reflect a stereotypical African American name, while the second is meant to project the research assistant as caucasian.

Table 9: Descriptive analysis of the Acceptance time, Estimated waiting time 1 and Actual waiting time of Uber and Lyft in the Boston experiment

	Attributes	Acceptance Times	Est. waiting time 1 (min)	Actual waiting time (s)
Uber	Num.obs	372	372	372
	Mean	48	3.3	218
	S.D.	102	1.4	148
	Median	25	3	196
	Black Mean	45.9	3.3	211
	White Mean	50.7	3.3	226
Lyft	Num.obs	387	387	387
	Mean	52.4	2.8	297
	S.D.	168	1.4	211
	Median	19	2	243
	Black Mean	55	2.7	296
	White Mean	49	2.8	299

Notes: This table shows the number of trips taken and summary statistics for the time until the request was accepted (Acceptance time), the *estimated* wait time reported by the app when the trip was requested (Est. wait time 1), and the actual wait time from when the trip was requested (Actual wait time) for the Boston experiment. In addition, we report the means of these three variables by race.

Table 10: Relationship between race and the log of *estimated* waiting times prior to requesting the trip in the Boston experiment

	UberX	Lyft
Intercept	5.214***	5.013***
	(.034)	(.038)
African American Name	$-.032$	$-.038$
	(.045)	(.051)
R^2	.001	.001
Adj. R^2	-.001	-.001
Num. obs.	372	387
RMSE	.432	.498

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Notes: This table shows the affect of race on *estimated* wait time prior to the trip request in the Boston experiment. We find no systematic relationship between the service's estimated wait times and race.

Table 11: Models of the log of Acceptance Time for UberX and Lyft in the Boston experiment

	UberX			Lyft		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
African American	.000	.003	−.011	.017	.017	.011
	(.059)	(.059)	(.057)	(.060)	(.060)	(.058)
log(ETA1)		.075	.106		.009	.049
		(.069)	(.068)		(.061)	(.059)
High Income			.034			.072
			(.059)			(.060)
Low Pop Density			.039			.089
			(.060)			(.060)
High AA Pop			.097			−.024
			(.059)			(.059)
Individual1			−.506***			−.348**
			(.148)			(.153)
Individual2			−.318**			.063
			(.160)			(.162)
Individual3			−.339**			.099
			(.137)			(.144)
Individual4			−.617***			−.428***
			(.147)			(.157)
Individual5			−.251*			−.068
			(.142)			(.150)
Individual6			−.393***			−.036
			(.149)			(.156)
(Intercept)	3.285***	2.896***	3.048***	3.000***	2.953***	2.797***
	(.044)	(.363)	(.373)	(.045)	(.310)	(.316)
R^2	.000	.003	.086	.000	.000	.124
Adj. R^2	-.003	-.002	.056	-.003	-.005	.097
Num. obs.	349	349	349	364	364	364
RMSE	.546	.546	.530	.570	.571	.541

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Notes: This table shows regression models of the log of the time it takes until a trip is accepted for the Boston experiment. We split the regressions by service (UberX and Lyft). The first model is a simple difference in mean times across African Americans and white. The second model controls for the estimated wait time prior to the trip being requested. Model 3 includes route characteristics and research assistant fixed effects. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Black Pop is defined as 20% or greater African American population, High Income is defined as median household income greater than $60k annually, and Lower Population Density is defined as fewer than 12800 people per square mile.

Table 12: Models of the log of actual waiting time for UberX and Lyft in the Boston experiment

		UberX			Lyft	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
African American	−.120	−.102	−.076	−.042	−.014	−.008
	(.078)	(.073)	(.073)	(.072)	(.065)	(.065)
log(ETA1)		.609***	.625***		.612***	.611***
		(.086)	(.087)		(.065)	(.067)
High Income			.032			.042
			(.075)			(.068)
Low Pop Density			.123			.016
			(.076)			(.068)
High AA Pop			−.023			−.076
			(.075)			(.067)
Individual1			.201			.235
			(.188)			(.172)
Individual2			.494**			.028
			(.203)			(.182)
Individual3			.082			.051
			(.174)			(.162)
Individual4			.091			−.010
			(.187)			(.176)
Individual5			.192			.052
			(.180)			(.169)
Individual6			−.063			−.093
			(.190)			(.175)
(Intercept)	5.213***	2.045***	1.762***	5.478***	2.416***	2.372***
	(.058)	(.452)	(.473)	(.054)	(.331)	(.356)
R^2	.007	.132	.175	.001	.196	.218
Adj. R^2	.004	.127	.148	-.002	.192	.193
Num. obs.	349	349	349	364	364	364
RMSE	.727	.680	.672	.679	.610	.610

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Notes: This table shows regression models of the log of actual wait time from when the trip was requested to the time the research assistant was picked up in the Boston experiment. We split the regressions by service (UberX and Lyft). The first model is a simple difference in mean times across African Americans and white. The second model controls for the estimated wait time prior to the trip being requested. Model 3 includes route characteristics and research assistant fixed effects. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Black Pop is defined as 20% or greater African American population, High Income is defined as median household income greater than $60k annually, and Low Population Density is defined as fewer than 12800 people per square mile.

Table 13: Comparisons of cancellation rates between black and white passengers in the Boston experiment

Passengers	Uber		Lyft	
	Black	White	Black	White
All	10.1%	4.9%	6.0%	7.7%
Males	11.2%	4.5%	8.4%	8.7%
Females	8.4%	5.4%	2.1%	6.4%

Notes: This table reports the mean cancellation rates across the two services, as well as the cancellation rates across the four gender·race pairs in the Boston experiment.

Table 14: Models of cancellations in the Boston experiment

	Lyft Means	Lyft Weekday Effects	Uber Means	Uber Weekday Effects	Uber Rider Effects	Uber Transit Effects	Demographic Interaction
African American Name	−.017		.052**				
	(.024)		(.025)				
African American Male		.000		.075**	.072**	.073**	.006
		(.032)		(.034)	(.035)	(.035)	(.041)
African American Female		−.064*		.037	−.041	−.019	−.021
		(.036)		(.039)	(.143)	(.142)	(.141)
White Female		−.028		.012	−.058	−.030	−.037
		(.037)		(.039)	(.137)	(.136)	(.135)
High AA Pop		−.012		−.008	−.007	−.008	−.011
		(.025)		(.026)	(.026)	(.026)	(.026)
High Income		−.005		−.065**	−.067**	−.061**	−.058**
		(.025)		(.027)	(.027)	(.027)	(.026)
Low Pop Density		−.008		.055**	.055**	.053**	−.002
		(.025)		(.026)	(.026)	(.026)	(.032)
WeekdayTuesday		−.023		−.097	−.139	−.116	−.125
		(.107)		(.124)	(.127)	(.127)	(.125)
WeekdayWednesday		−.016		−.022	−.025	−.017	−.017
		(.047)		(.050)	(.051)	(.051)	(.051)
WeekdayThursday		.010		.027	.043	.032	.042
		(.050)		(.053)	(.057)	(.056)	(.056)
WeekdayFriday		.010		.043	.046	.055	.050
		(.051)		(.054)	(.056)	(.056)	(.055)
WeekdaySaturday		.022		−.035	−.042	−.052	−.048
		(.043)		(.048)	(.049)	(.049)	(.048)
WeekdaySunday		−.038		−.081	−.068	−.075	−.071
		(.054)		(.056)	(.057)	(.057)	(.056)
Individual1					.122	.105	.109
					(.128)	(.127)	(.126)
Individual2					.031	.019	.020
					(.130)	(.130)	(.128)
Individual3					−.019	−.005	−.005
					(.066)	(.065)	(.065)
Individual4					.027	.046	.045
					(.068)	(.067)	(.067)
Individual5					.068	.057	.065
					(.125)	(.125)	(.124)
Individual6					.026	.041	.047
					(.068)	(.068)	(.067)
Pickup at Transit						.100***	.097***
						(.038)	(.038)
Black Male * Low Pop Density							.157***
							(.054)
(Intercept)	.077***	.097***	.049***	.054*	.051	.024	.044
	(.017)	(.029)	(.019)	(.031)	(.068)	(.068)	(.068)
R²	.001	.014	.009	.039	.050	.065	.083
Adj. R²	-.001	-.013	.007	.013	.010	.024	.041
Num. obs.	460	460	451	451	451	451	451
RMSE	.251	.253	.267	.266	.266	.265	.262

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: This table shows linear probability models for cancellations in the Boston experiment. We first show the results for Lyft requests only. Lyft provides the name and photo of the prospective rider prior to the Lyft driver accepting the ride. Therefore, we would not expect to see an impact of name on the probability of a cancellation since any discrimination that could occur would happen prior to acceptance. An UberX driver does not see the name of the prospective passenger until after accepting the ride. One channel of discrimination, therefore, would be for the driver to accept the ride, receive a signal about the rider's race, and then cancel. The first model is a simple difference in the average rate of cancellations across implied race. We then add day of week and pickup location characteristics as controls and allow for gender differences. The third model for Uber adds research assistant fixed effects. The fourth model adds an indicator variable for whether the pickup location is a public transit location in case the attractiveness of these rides is different. The final model interacts the African American male name with an indicator variable for whether the location is in a low-population density area. For route characteristics, we use demographics from the Census Block Group of the pickup point using 2010 data. High Black Pop is defined as 20% or greater African American population, High Income is defined as median household income greater than $60k annually, and Low Population Density is defined as fewer than 12800 people per square mile.

Table 15: Models of the log of travel distance for UberX and Lyft in the Boston experiment

	UberX			Lyft		
	OLS	Tau = .75	Tau = .95	OLS	Tau = .75	Tau = .95
Female	.056**	.079***	.089	.050	.027	.126*
	(.022)	(.030)	(.097)	(.031)	(.039)	(.076)
African American Name	.002	.014	−.068	−.022	.035	−.178*
	(.022)	(.033)	(.069)	(.031)	(.037)	(.101)
log(Google Est. Dist)	.886***	.837***	.599***	.929***	.861***	.608***
	(.033)	(.041)	(.103)	(.047)	(.058)	(.115)
Surge Multiplier	.063	.076	−.033	.159	.180*	.249
	(.043)	(.055)	(.088)	(.099)	(.107)	(3.054)
(Intercept)	−.001	.069	.554***	−.120	−.041	.337
	(.049)	(.073)	(.116)	(.106)	(.117)	(3.061)
R^2	.641			.494		
Adj. R^2	.638			.489		
Num. obs.	405	405	405	415	415	415
RMSE	.220			.314		

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Notes: This table shows regression models of the log of the travel distance. We split the regressions by service (UberX and Lyft). For Uber and Lyft, travel times are those recorded on receipts. The estimated travel times of the three services are calculated using Google API according to the assigned origins and destinations. The first model is ordinary least squares regression. We control for both race and gender, and whether surge pricing is in effect. The next two models are quantile regressions of the 75th and 95th quantile, respectively.

Table 16: Models of the log of travel time for UberX and Lyft in the Boston experiment

	UberX			Lyft		
	OLS	Tau = .75	Tau = .95	OLS	Tau = .75	Tau = .95
Female	.011	.021	.009	.069*	.017	.113
	(.033)	(.042)	(.104)	(.039)	(.055)	(.120)
African American Name	.041	.036	.124	−.024	−.002	−.079
	(.033)	(.042)	(.117)	(.039)	(.055)	(.131)
log(Google Est. Time)	.623***	.560***	.441***	.767***	.647***	.595***
	(.048)	(.067)	(.140)	(.057)	(.081)	(.219)
Surge Multiplier	.211***	.274**	.378**	.215*	−.026	.427
	(.063)	(.125)	(.150)	(.124)	(.149)	(1.994)
(Intercept)	2.045***	2.526***	3.502***	1.150***	2.322***	2.581
	(.293)	(.419)	(.824)	(.350)	(.483)	(2.431)
R^2	.311			.318		
Adj. R^2	.304			.311		
Num. obs.	405	405	405	415	415	415
RMSE	.326			.391		

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Notes: This table shows regression models of the log of travel time. We split the regressions by service (UberX and Lyft). For Uber and Lyft, travel times are those recorded on receipts. The estimated travel times of the three services are calculated using Google API according to the assigned origins and destinations. The first model is ordinary least squares regression. We control for both race and gender, and include route fixed effects, and whether surge pricing is in effect. The next two models are quantile regressions of the 75th and 95th quantile, respectively.

References

BERTRAND, M. AND S. MULLAINATHAN (2004): "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*, 94, 991–1013.

DONNELLY, A. (2015): "Aboriginal actor Jack Charles 'refused taxi twice in three days'," *BBC News*.

EDELMAN, B. AND M. LUCA (2014): "Digital Discrimination: The Case of Airbnb.com," Tech. rep., Harvard University.

EDELMAN, B., M. LUCA, AND D. SVIRSKY (2016): "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment," Tech. rep., Harvard University.

FRYER, R. G. AND S. D. LEVITT (2004): "The Causes and Consequences of Distinctively Black Names," *The Quarterly Journal of Economics*, 119, 767–805.

GLANVILLE, D. (2015): "Why I Still Get Shunned by Taxi Drivers," *The Atlantic*, October 24.

GODDARD, T., K. B. KAHN, AND A. ADKINS (2015): "Racial bias in driver yielding behavior at crosswalks," *Transportation Research Part F: Traffic Psychology and Behaviour*, 33, 1–6.

GONEN, Y. (2015): "Puerto Rican mayor booted from NYC cab | New York Post," *NY Daily News*.

HARSHBARGER, R. (2015): "City puts biased drivers who refuse rides on notice with new video," *New York Post*, June 9th.

HUGHES, R. AND D. MACKENZIE (2016): "Transportation network company wait times in Greater Seattle, and relationship to socioeconomic indicators," *Journal of Transport Geography*, 56, 36–44.

MAHALANOBIS, P. C. (1936): "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*, 2, 49–55.

NICHOLSON, E. (2013): "Southern Dallas Leaders Say Uber is Profiling Customers, and They Want City Hall to Act," *Dallas Observer*, October 21.

PIFF, P. K., S. DM, C. S, M.-D. R, AND K. D (2012): "Higher Social Class Predicts Increased Unethical Behavior," *Proceedings of the National Academy of Science*, 109, 4086–4091.

POPE, D. G. AND J. R. SYDNOR (2011): "What's in a Picture? Evidence of Discrimination from Prosper. com," *Journal of Human Resources*, 46, 53–92.

SMART, R., B. ROWE, A. HAWKEN, M. KLEIMAN, N. MLADENOVIC, P. GEHRED, AND C. MANNING (2015): "Faster and Cheaper: How Ride-Sourcing Fills a Gap in Low-Income Los Angeles Neighborhoods," Tech. rep., BOTEC Analysis Corporation.

UBERPEOPLE.NET (2015): "Uber Cancellation Abuse - Mini FAQ," *Uber Drivers Forum*.

A   Appendix

Figure A.1: Routes, pick-up locations and neighborhood characteristics of Seattle experiment.


(a) Population density (per km²)   (b) Proportion of black population(%)   (c) Average income($)

under 1500
1501 - 2500
2501 - 3500
3501 - 4500
4501 - 5500
5501 - 6500
over 6500

under 5%
5%-15%
15%-25%
over 25%

<25k
25k-50k
50k-75k
75k-100k
100k-150k
>150k

● Assigned pick-up locations

Route 1   Route 2   Route 3   Route 4   Route 5   Route 6   Route 7
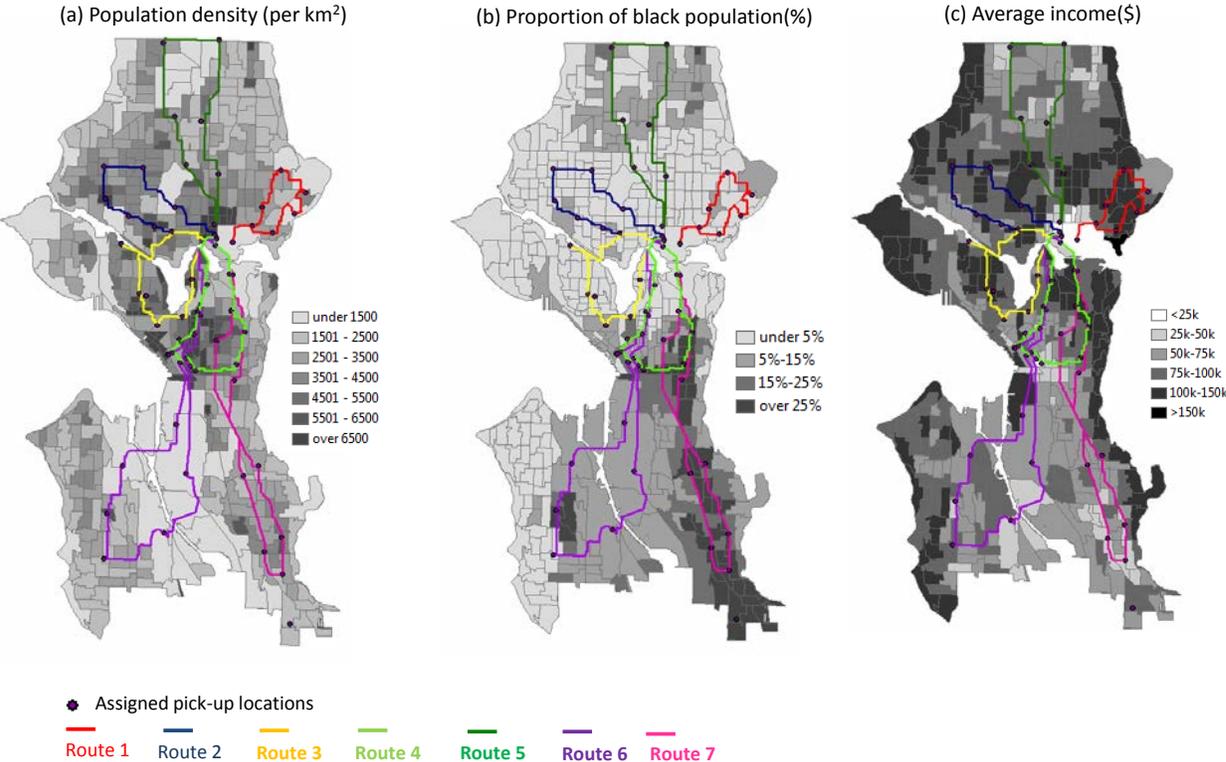
**Figure A.2:** Data collection process illustration.
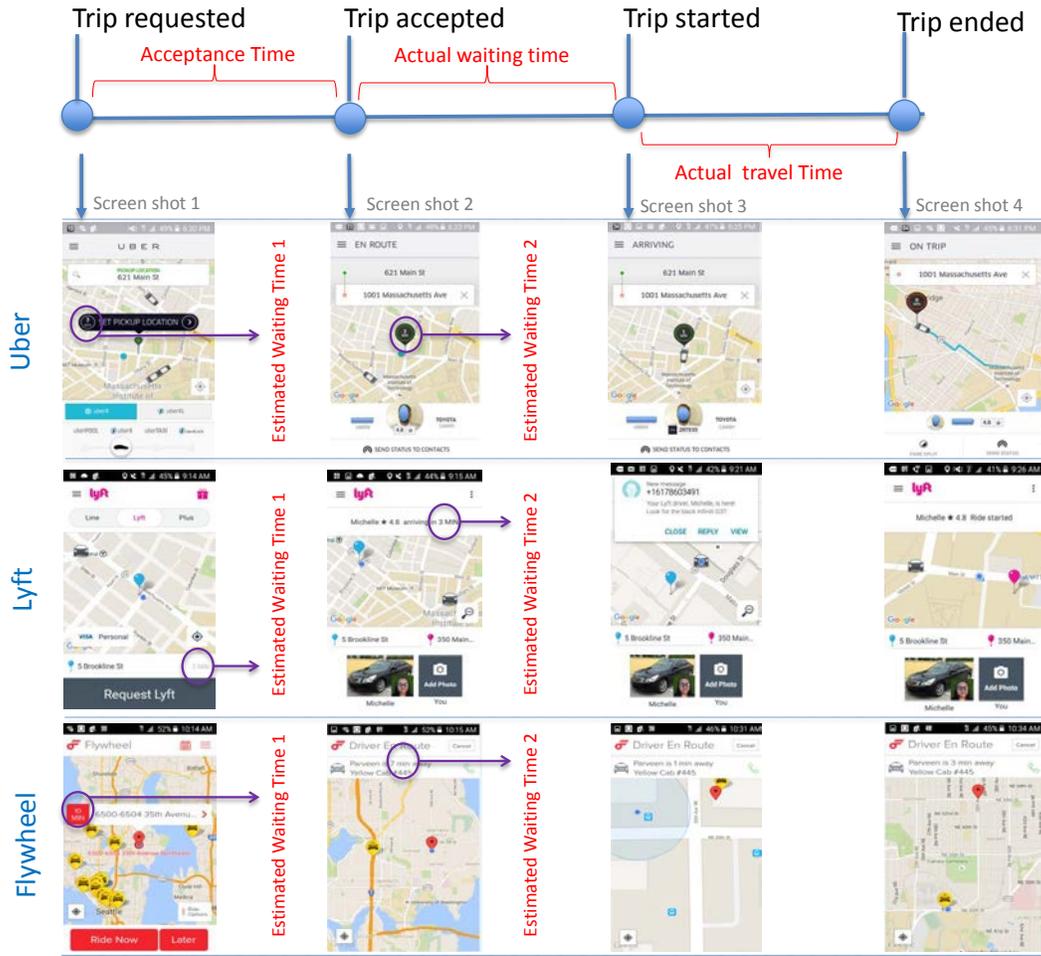

Figure A.3: Balance in days, times, and tour IDs for trips taken by black and white RAs.


2
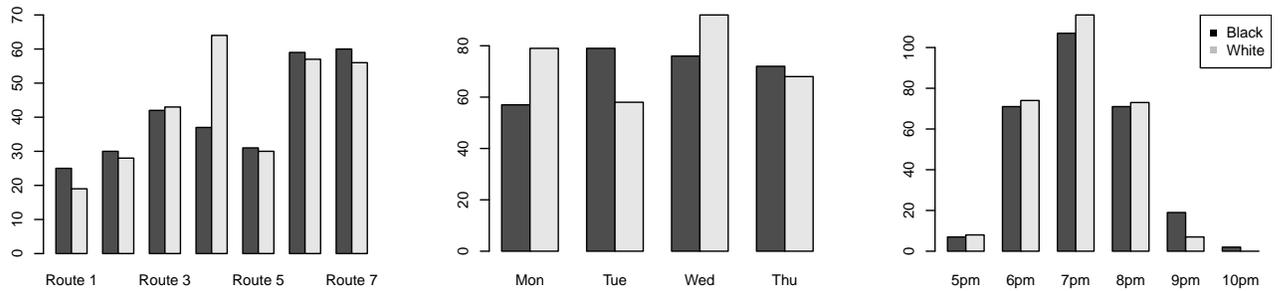
**Figure A.4:** Cumulative Star Ratings for Black and White RAs Using UberX in Seattle.
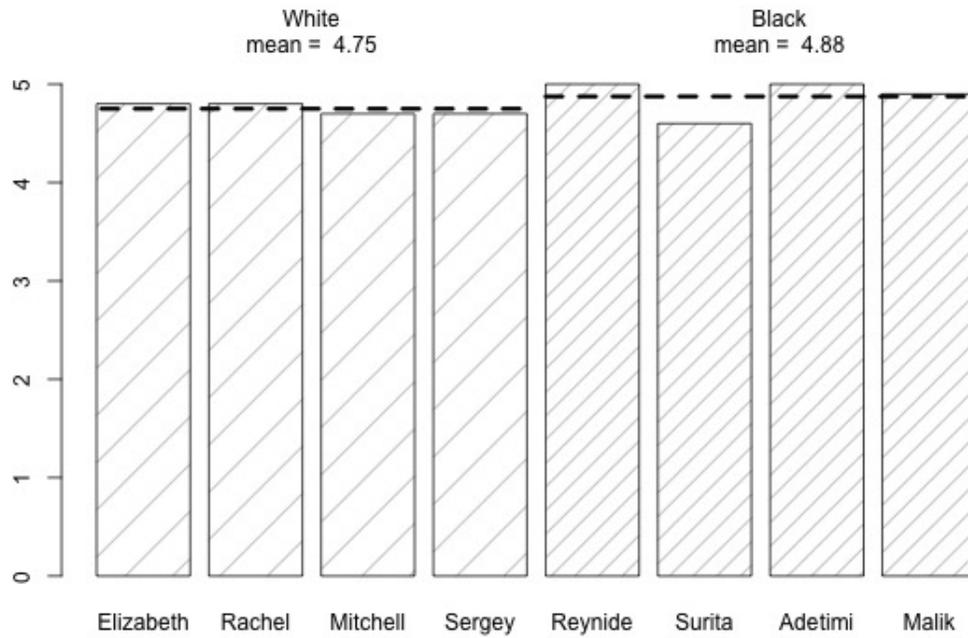
Figure A.5: Cumulative Star Ratings for Black and White RAs Using Lyft in Seattle.
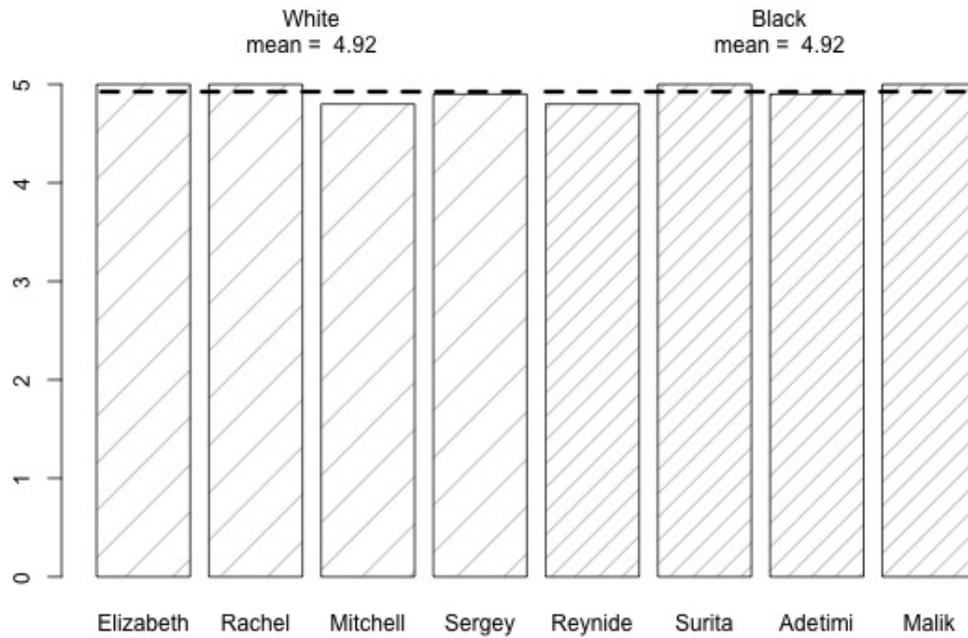
Figure A.6: Number of taxis passing by black and white RAs while hailing from the curb

Number of Taxis Passing Waiting Traveler
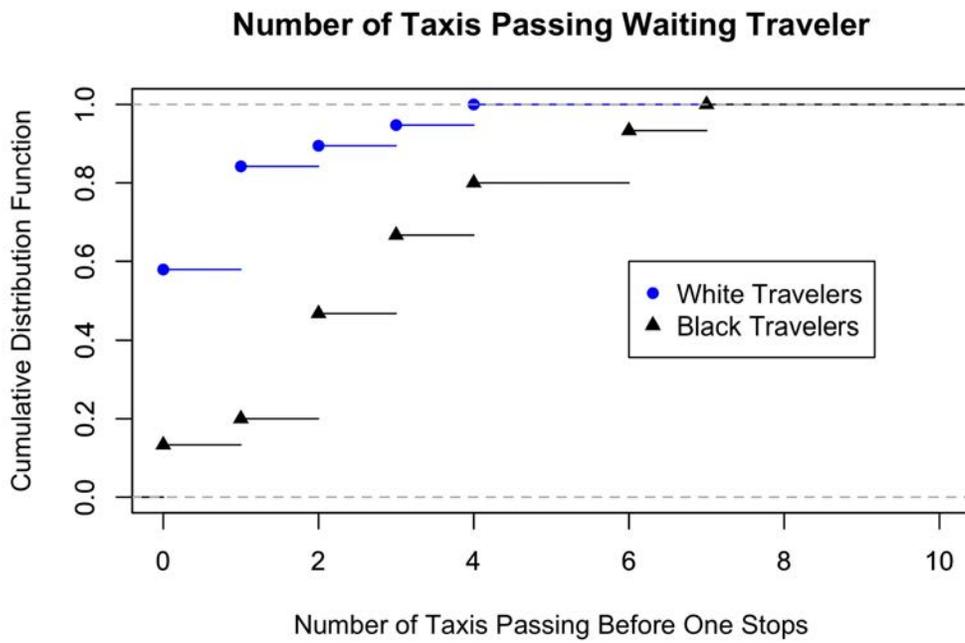
Figure A.7: GIS Analysis of Boston-area Census Block Groups by % Black Residents, Average Household Income, and Population Density

(a) Population Density, Thousands / Square Mile
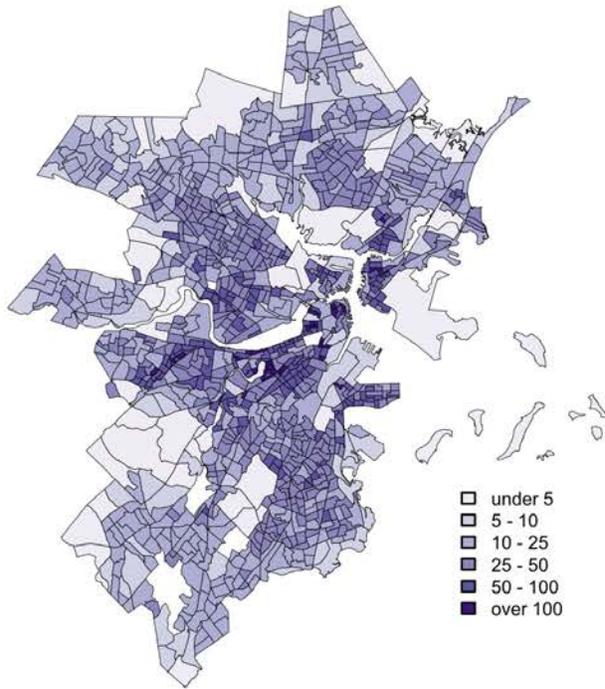
(b) Median Household Income
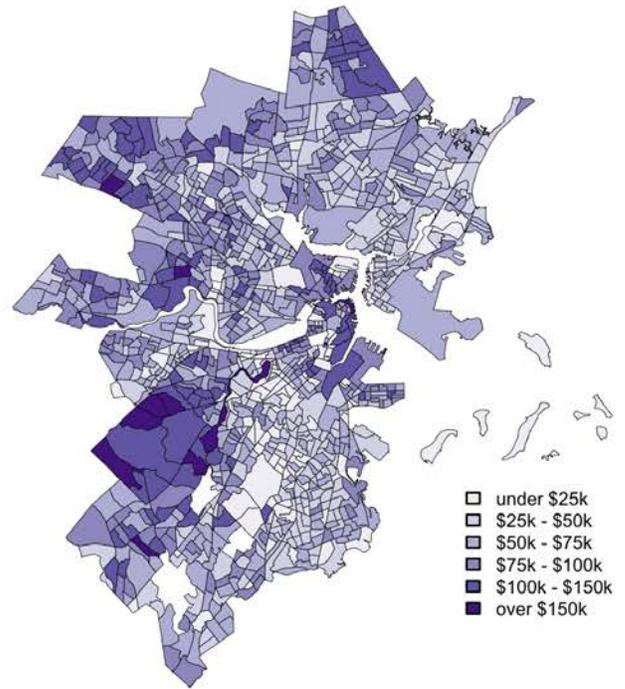
(c) Share Non-English Speaking Households
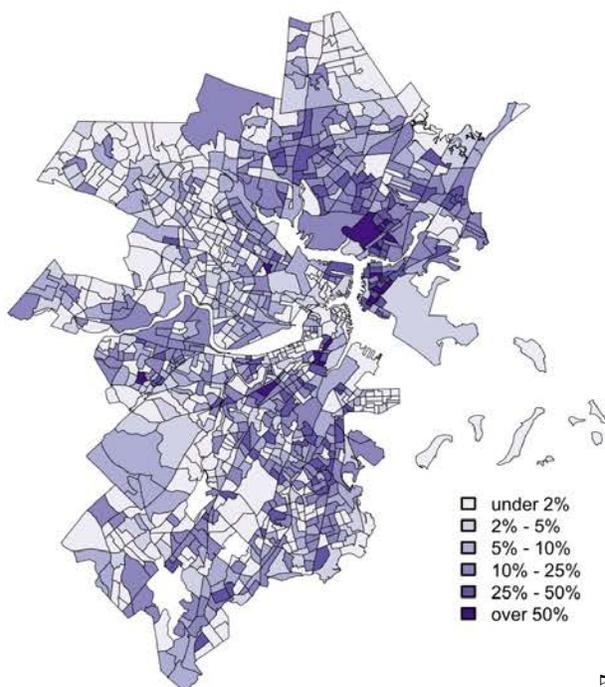
(d) Black Share of Population
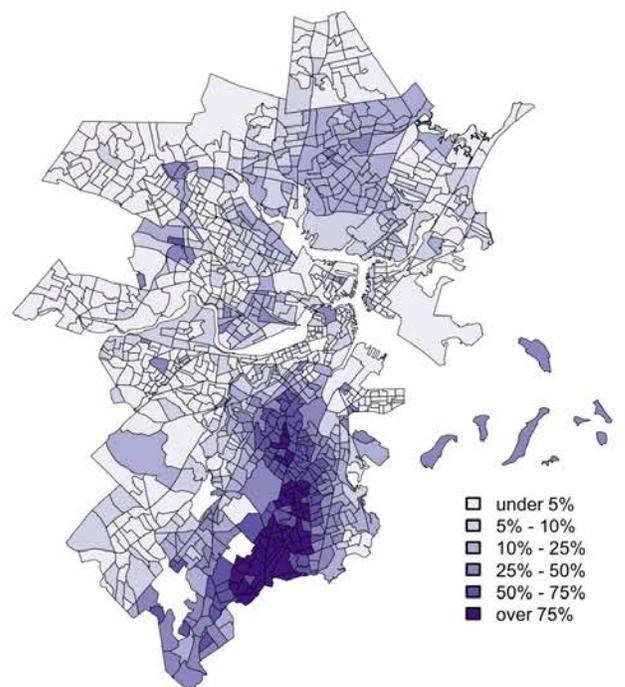
Figure A.8: Empirical Cumulative Distribution of Acceptance Time, Actual Waiting Time 1 and Waiting Time increase.
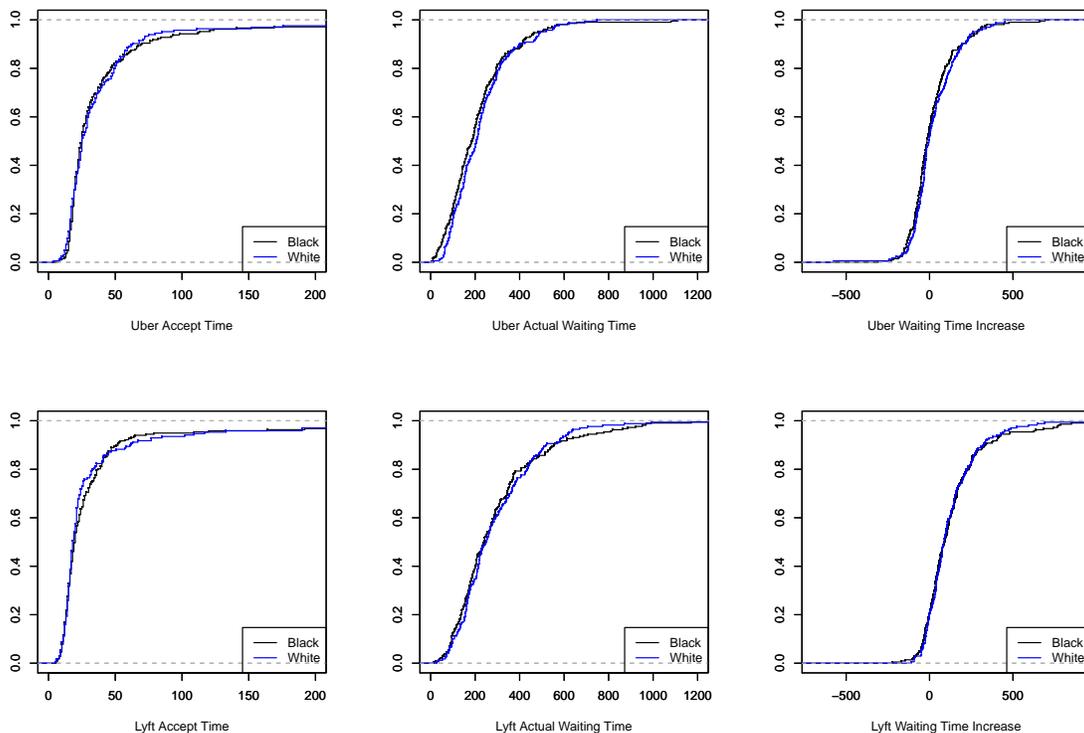
Figure A.9: Relationships between estimated and actual waiting times for UberX, Lyft, and Flywheel in the Seattle experiment
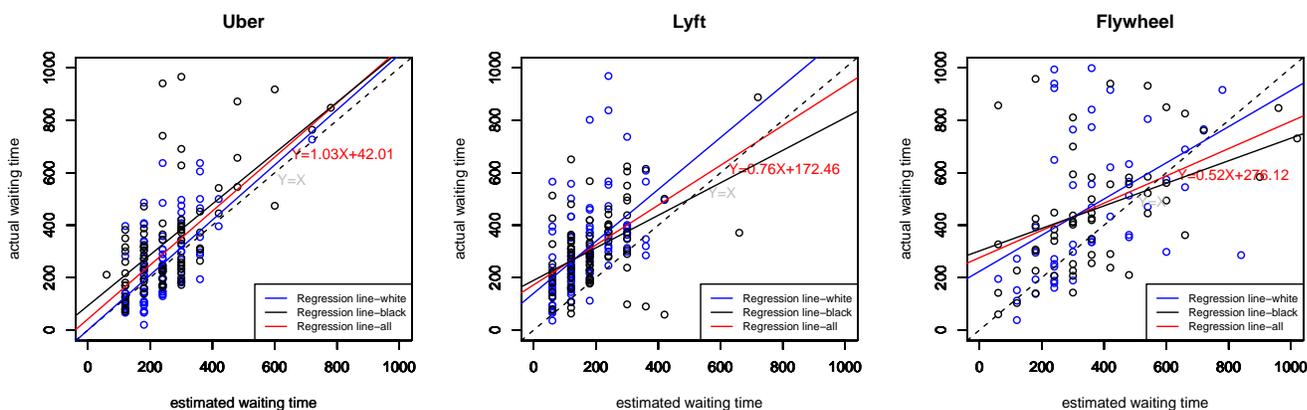
Table A.1: Matched pair comparisons of Acceptance Time for Uber, Lyft, and Flywheel.

|	Service	UberX	Lyft	Flywheel
$log(AcceptanceTime)$	# of Matched pairs	50	39	21
	ATT	0.25	0.078	0.047
	AI SE	0.08	0.066	0.080
	t	3.20	1.18	0.59
	p	0.001	0.24	0.56
$log(\frac{ActualWaitingTime}{EstimatedWaitingTime1})$	# of Matched pairs	49	37	19
	ATT	0.156	0.053	-0.102
	AI SE	0.066	0.060	0.073
	t	2.38	0.87	-1.40
	p	0.017	0.38	0.16

Table A.2: Matched pair comparison of Waiting times for UberX in Seattle.

Match based on	Gender	No limit	Exact	Exact	Exact	Exact	Exact
	Time of request diff limit	No limit	1h	1h	1h	1h	1h
	location diff limit	No limit	1 mile	1000 m	0.5 mile	500 m	200 m
Request Time	Num of Matched pairs	83	62	51	50	46	36
	ATT	5.5	6.6	6.08	6.7	6.4	9.8
	AI SE	4.4	4.2	4.1	4.0	4.0	3.6
	t-stat	1.25	1.59	1.53	1.64	1.59	2.70
	p.Val	0.21	0.11	0.12	0.10	0.11	0.007
Waiting Time Increase	Num of Matched pairs	82	61	50	49	45	34
	ATT	30.5	41.8	41.44	38.3	38.6	54.9
	AI SE	26.8	21.1	20.4	20.2	16.8	16.4
	t-stat	1.14	1.98	2.03	1.90	2.30	3.35
	p.Val	0.255	0.047	0.042	0.058	0.022	0.001

Table A.3: Matched pair comparison of Waiting times for Lyft in Seattle.

Match based on	Gender	No limit	Exact	Exact	Exact	Exact	Exact
	Time of request diff limit	No limit	1h	1h	1h	1h	1h
	location diff limit	No limit	1 mile	1000 m	0.5 mile	500 m	200 m
Request Time	Num of Matched pairs	84	50	41	39	36	23
	ATT	6.7	6.0	8.0	8.3	7.1	8.3
	AI SE	4.6	3.8	3.6	3.6	3.7	3.6
	t-stat	1.52	1.60	2.21	2.29	1.89	2.32
	p.Val	0.14	0.12	0.03	0.02	0.06	0.02
Waiting Time Increase	Num of Matched pairs	84	53	43	37	31	21
	ATT	15.5	-0.4	2.2	-1.8	-13.7	-21.6
	AI SE	24.2	20.1	18.4	17.3	17.1	12.9
	t-stat	0.64	-0.02	0.11	-0.10	-0.80	-1.67
	p.Val	0.52	0.98	0.91	0.92	0.42	0.09

Table A.4: Matched pair comparison of waiting times for Flywheel in Seattle.

Match based on	Gender	No limit	Exact	Exact	Exact	Exact	Exact
	Time of request diff limit	No limit	1h	1h	1h	1h	1h
	location diff limit	No limit	1 mile	1000 m	0.5 mile	500 m	200 m
Request Time	Num of Matched pairs	58	30	22	21	20	13
	ATT	10.7	4.6	0.05	-1.7	-1.2	3.5
	AI SE	5.4	3.2	2.9	2.7	2.6	1.6
	t-stat	2.00	1.42	0.02	-0.63	-0.44	2.25
	p.Val	0.05	0.16	0.99	0.53	0.66	0.02
Waiting Time Increase	Num of Matched pairs	55	25	20	19	18	11
	ATT	10.3	-77.2	-102.2	-114.1	-134.3	-133.2
	AI SE	56.1	33.9	33.5	33.8	35.4	33.2
	t-stat	0.18	-2.28	-3.07	-3.38	-3.79	-4.0
	p.Val	0.85	0.02	0.00	0.00	0.00	0.00

Table A.5: Required sample sizes

Variable	5 percent	10 percent	15 percent
Actual Wait Time	1,736	434	193
Estimated Wait Time	903	226	101
Ratio of Actual To Estimated Wait Time	1,378	345	154
Δ Actual and Estimated Wait Time	37,672	9,418	4,186
Cancellation	NA	NA	NA
Actual Drive Time	1,056	264	118
Estimated Drive Time	NA	NA	NA
Ratio of Actual to Estimated Drive Time	NA	NA	NA
Δ Actual and Estimated Drive Time	NA	NA	NA