Spatial and Social Frictions in the City: Evidence from Yelp*

Donald R. Davis[†]
Columbia University and NBER

Jonathan I. Dingel[‡]
Chicago Booth

Joan Monras[§]
Sciences Po

Eduardo Morales[¶]
Princeton University and NBER

July 2014
Preliminary and Incomplete

Abstract

We employ user-generated data from the social website Yelp to estimate how spatial and social frictions combine to shape consumption choices within cities. Travel time matters for consumption choices, more strongly from work than home. Social frictions also play a large role. Individual are less likely to visit venues in places demographically different from their own neighborhood. Judged by their choices, men and women inhabit different cities. Women, but not men, avoid high crime areas. And women, but not men, are particularly unlikely to visit venues deep in the interior of a neighborhood whose racial and ethnic mix differs from that of their home neighborhood. We can use our estimates to gauge how much greater these frictions would be if crime were at recent historical peaks and to show how transportation projects would alter the consumption value of a city.

*We thank Ben Eckersley, Hadi Elzayn, Benjamin Lee, and Charlene Lee for research assistance. Monras thanks the the Banque de France Sciences Po partnership for financial support.

[†]drdavis@columbia.edu
[‡]jdingel@chicagobooth.edu
[§]jm3364@columbia.edu
[¶]ecmorale@princeton.edu

1 Introduction

Most people reside in cities in the first instance to work. But cities are about more than work, they are also about consumption opportunities (Glaeser, Kolko, and Saiz, 2001). Consumption in cities often requires travel. Since travel is costly, not all consumption venues are equally accessible, and so price indices vary by location inside the city (Couture, 2014). Home and work are our primary locations within the city. And so home, work, and the commute between the two, become primary bases from which our consumption is launched.[1]

Spatial frictions fragment the city, making it less valuable to us. But frictions are not only spatial, they are also social. Nearly a half century after the passage of fair housing laws, cities in the United States continue to have high levels of *de facto* segregation. If this segregation by ethnicity, race, and income reduces our willingness to take advantage of consumption opportunities, then the value of the city is diminished. Concerns over safety may also limit our willingness to take advantage of consumption opportunities. If this is so, then we should care not only about the direct effects of crime, but also the effect it has in diminishing the consumption value of the city to us. These costs of social frictions need not fall equally on all members of society. They could vary by ethnicity, race, or gender.

In this paper, we explore empirically how spatial and social frictions govern our use of the city. In doing so, we will make a number of methodological and substantive contributions. We make the first use of data generated by a mobile social networking application to address spatial questions in economics. We make the first use of home and work locations, and the commuting path, to examine consumption of heterogeneous, quality-differentiated goods. We also go beyond the existing literature on social frictions within the city. Most importantly, we use a rich set of venue-specific controls to distinguish social frictions from the characteristics of venues that may also influence individuals' travel and consumption decisions.

Our empirical work characterizes the consumption decisions of individuals living in Manhattan who use Yelp.com, a website where users review local businesses. The data we collected from Yelp identify users' home and work locations and describe the venues they patronize. We combine this information with additional data linked to locations within New York City and estimate a discrete-choice model of the venue-visit decision to characterize consumer preferences. Our estimation procedure makes use of the choice-set construction technique suggested by McFadden (1978) and Train, McFadden, and Ben-Akiva (1987) to address the computational burden of the fact that Yelp users may choose amongst New York City's tens of thousands of restaurants.

Our (preliminary and incomplete) estimates show that both spatial and social frictions influence the geography of consumption. As one would expect, visits to venues fall with travel times. They fall more sharply with minutes from work than from home or the commuting path, perhaps reflecting a higher valuation of time at work.

Social frictions matter as well. Users make fewer visits to venues located in neighborhoods with racial or ethnic demographics different from those of their own neighborhood. Crime in a neighborhood reduces visits, and controlling for crime does not eliminate the effect of ethnicity and race on visits. Gender also matters. Women are significantly less likely than

[1] Houde (2012) analyzes spatial competition in gasoline retail market while defining consumers' locations as their commuting paths. He finds that the demand model incorporating commuting paths better matches observed sales than a model estimated assuming all consumer trips originate at their home locations.

men to visit venues in neighborhoods with more crime or with racial and ethnic demographics different from those of their own neighborhood.

The remainder of this paper proceeds as follows. We first provide more background on the substantive concerns of the paper. Then we describe the data and our empirical methodology. This is followed by a discussion of our principal empirical results. Future versions of this paper will also implement counterfactual exercises evaluating the consequences of introducing a new transit line or returning to the higher crime rates of the past.

1.1 Geography of consumption

Trade economists have long studied how geography and distance influence the production, consumption, and exchange of tradable goods. A recent literature has documented how the sets of tradable goods available for consumption differ across cities (Handbury and Weinstein, 2011; Handbury, 2012). Though a large share of consumer expenditure is devoted to non-tradable goods and services, and geographic variation in these non-tradables is posited to shape the relative attractiveness of cities (Glaeser, Kolko, and Saiz, 2001), we know relatively little about the geography of consumption within cities.

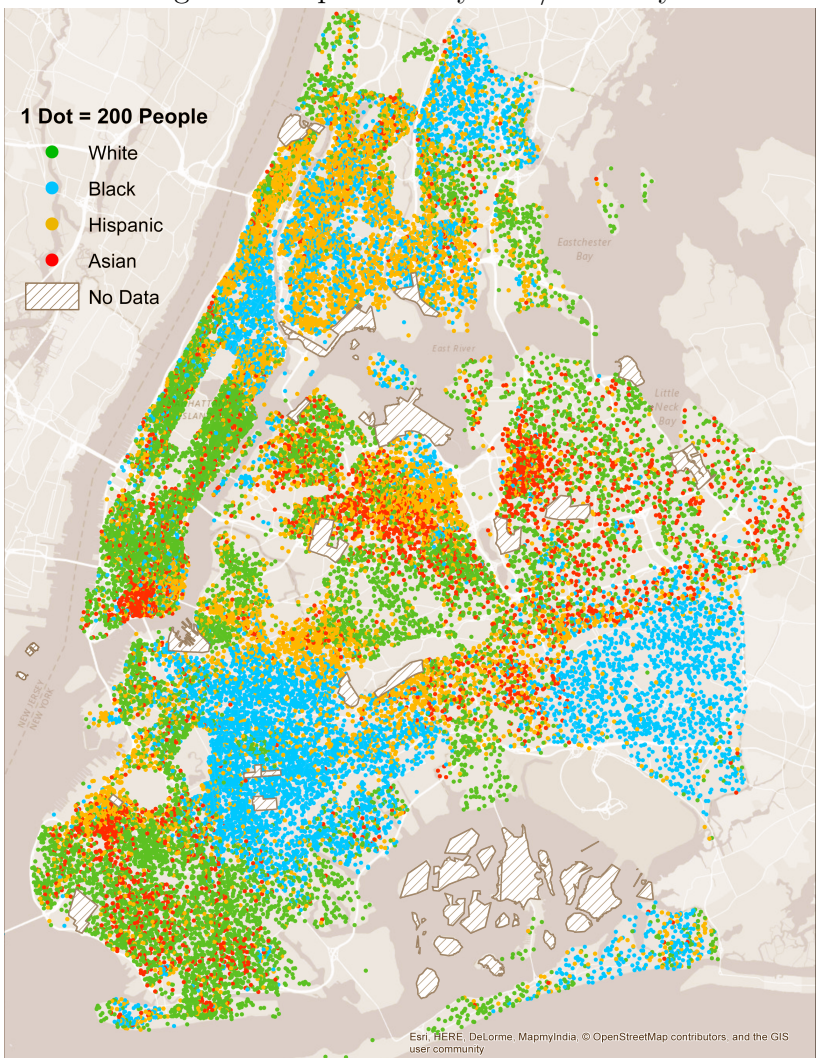
This dimension of economic life has grown increasingly important in recent decades. The share of food spending devoted to food prepared away from home grew from less than 26% in 1970 to more than 43% in 2012. While the number of daily commuting trips has stayed relatively steady for decades, trips for family/personal business and social/recreational purposes have steadily grown (source).

Prior studies of consumption within the city have been limited by data constraints. Couture (2014) infers the benefits of variety due to urban density from the time individuals spend traveling from their homes to restaurants. In his data, he observes the trip origin and purpose, as well as the set of venues at a given time-radius, but he does not know the venue visited. Thus he cannot relate consumer decisions to locational characteristics like crime rates or demographics. Luca (2011) shows the effect that Yelp ratings have on restaurant revenues using a regression-discontinuity design. His data do not describe individual consumers, so these results describe aggregate demand without incorporating spatial or social frictions. In our data, we observe both individuals' home and work locations and the venues they patronize. This allows us to estimate the roles of both spatial and social frictions and characterize their relative importance.

1.2 Ethnicity and Race

Since individuals must travel to consume non-tradable goods and services, their consumption decision depends not only on the non-tradable's characteristics and the travel time required but also the whole experience associated with the consumption trip. The characteristics of the neighborhood in which a business venue is located likely influence its popularity with consumers. For example, ethnic and racial composition might influence consumer behavior if individuals prefer to consume in neighborhoods whose demographics resemble those of their own neighborhood or if individuals use ethnic or racial composition as a proxy for other neighborhood characteristics.

Figure 1: Population by race/ethnicity



These issues matter because New York City remains *de facto* quite segregated. Figure 1 depicts the population of New York City using colored dots to represent people of four different demographic groups – whites (green), blacks (blue), Hispanics (orange), and Asians (red).[2]

Much of the literature has described ethnic and racial fragmentation in terms of residential segregation (Cutler, Glaeser, and Vigdor, 1999; Echenique and Fryer, 2007; Bayer and McMillan, 2008). Unfortunately, this measure can only describe the geographic segregation of residential locations and not the segregation of individuals' economic and social interactions.[3] Some of the mechanisms posited to explain residential segregation would also predict economic and social segregation. For example, between countries, lower bilateral trust reduces trade and investment (Guiso, Sapienza, and Zingales, 2009), and closer cultural ties can fa-

[2]This illustration was inspired by a *New York Times* project called "Mapping America: Every City, Every Block."

[3]An exception is Echenique and Fryer (2007), who study students' friendship networks within schools.

cilitate economic exchange (Rauch, 2001). A large literature documents merchant-customer racial/ethnic frictions (Lee, 2000; Ayres, 2001; Schreer, Smith, and Thomas, 2009). The perception of disparate treatment in consumer transactions due to race or ethnicity affects all groups, if not equally (Antecol and Cobb-Clark, 2008). In our study, we use information on the demographic composition of residents surrounding venues and users' residential locations, as well as the ethnicity or race of the user, to investigate whether ethnic or racial dissimilarity reduces economic interactions within New York City. We know that racial and ethnic segregation by neighborhood remain quite strong in New York City (Logan, 2011); we are in effect asking if the gap in user visits is reduced even further than that suggested by the spatial separation.

1.3 Crime

Another locational characteristic that likely influences the attractiveness of a business venue is the local crime rate. Much of the literature on urban crime has studied how crime affects local population size (Cullen and Levitt, 1999) and local housing prices (Gibbons, 2004; Linden and Rockoff, 2008; Pope, 2008). In the same manner that households change residences to avoid high-crime areas, they are presumably less likely to travel through and patronize businesses in high-crime locations. In fact, GPS devicemakers have developed technologies to incorporate crime statistics and demographic information when giving users navigational directions. Using data on consumers' choices and local crime rates, we are able to measure the degree to which individuals avoid patronizing Yelp venues in high-crime areas.

Crime rates, fear of crime, and residential segregation are complex, interrelated phenomena (O'Flaherty and Sethi, 2007, 2010). The fraction of a neighborhood's population that is black is significantly associated with its residents' perception of a crime problem, even after controlling for neighborhood crime rates (Quillian and Pager, 2001). Survey respondents generally overestimate the risk of criminal victimization, and white respondents more strongly associate racial composition with perceived risk (Quillian and Pager, 2010). Prior studies have characterized how racial composition and actual victimization rates relate to residents' fear of crime in their neighborhoods and estimated risk of burglary of their homes. Our data on consumer choices allow us to examine the influence of ethnic and racial composition and criminal activity on the decisions of non-residents to patronize various neighborhoods.

1.4 Gender

Men and women may use the city quite differently. In principle, such differences might arise from many aspects of the differences in life experiences. Some differences, such as differences in labor force participation, are partially eliminated by the fact that our sample includes only those for whom we could identify both a home and work location.

A potentially significant difference in how men and women use the city concerns safety and perceived safety. Men are more than twice as likely as women to be the subject of a violent victimization and three-and-a-half times as likely to be robbed by strangers (data for 2012 from National Crime Victimization Survey). Rape is overwhelmingly a crime committed against women (more than a factor of fifteen relative to men) and they are also victims of other sexual violence at twice the rate of men. While strangers commit only one-in-eight

rapes, they commit little shy of half of all other reported acts of sexual violence Black, Basile, Breiding, Walters, Chen, and Stevens (2011, Tables 2.1, 2.2, 2.5, 2.6). One has to keep in mind that these victimization rates are equilibrium phenomena, so it is quite likely that these differences are themselves produced in part by differential ways of using the city by men and women, which we will empirically examine.

There is a vast literature on the fear of crime (Doran and Burgess, 2011). A large majority of this literature is based on surveys such as the General Social Survey and the National Crime Survey (Ferraro, 1996). These studies uniformly find that fear of crime is higher among women. Pain (1991, p.416), for example, argues that

> ... the effects of crime have a greater impact on women's lives. Men commonly protect their property in response to the threat of crime, but as women are most concerned about sexual crimes, they are more likely to make adaptations to lifestyle and behaviour. These often involve, for example, self-imposed restrictions such as staying indoors at night, not walking alone and avoiding certain parts of town, which have implications for personal freedom and decrease general quality of life.

The surveys also indicate that a very large fraction of women adjust their behavior in regard to these fears, avoiding risky areas in ways that affect their social and leisure activities (Pain, 1997; Lorenc, Clayton, Neary, Whitehead, Petticrew, Thomson, Cummins, Sowden, and Renton, 2012).

Recently studies have begun to use GPS technologies to track movements in space (Raanan and Shoval, 2014). These are quite valuable, but have the drawbacks that participants cannot help but be aware they are part of a study and we do not know the purposes of trips.

The present study complements the older survey and time diary studies, as well as the newer GPS study. A first advantage is that the data were generated in the course of normal daily transactions, rather than being explicitly part of a study. We simply can't know how participation in a study affected either reporting of spatial activity or the activities themselves. A second major advantage of our data is that we have a very rich set of venue controls. It is very hard to know what frictions we are observing in people's use of the city if we don't directly observe why they visit different locations, as captured in the characteristics of the venues visited. The Yelp data are excellent in this dimension.

2 Data

We combine data from individual Yelp users with information on New York City to describe how these individuals use the city.

2.1 Yelp

Yelp.com is a website where users review local businesses, primarily restaurants and bars. The website describes a venue in terms of its address, hours of operation, average rating, user reviews, and a wide variety of other characteristics. In addition to assigning a rating of one

to five stars, users are encouraged to write reviews describing their personal experience with a business. These reviews vary greatly both in tone and length, ranging from a few sentences to many paragraphs. Crucial for our purposes is that users often disclose information in their reviews about their residential and work locations.[4] This provides us a novel description of the links between the spatial consumption choices of individuals and a rich set of location- and venue-specific characteristics.

2.2 Yelp user data

We examine users' Yelp reviews to identify their consumption choices and their home and work locations. In 2011, we downloaded all the reviews written by about 50,000 Yelp users who had reviewed a venue in the five boroughs of New York City. We randomly selected for study 25% of these users, who had written nearly 650,000 Yelp reviews. Between 1 January 2005 and 14 June 2011, these users wrote about 230,000 reviews of venues in New York and New Jersey. To identify these users' residential and work locations, we searched through the text of reviews in New York and New Jersey that contained at least one of 26 key phrases involving location, such as "close to me," "block away," and "my apartment."[5] Our phrase-matching process flagged 16,425 instances of these phrases. Research assistants then examined flagged user reviews to confirm that these identify the users' home and work location. Twenty-one percent of these flagged reviews contained information identifying a user's home location, and 11% of them identified a workplace. Thus, we obtained sets of venues revealing users' residential and work locations. Within each set, we estimated the residential and work locations using the average of all relevant latitude-longitude coordinates. This process identified about 1500 users with a residential location, 575 users with a workplace, and 450 users with both home and work locations.[6]

We have restricted our estimation sample to users with home and work locations in Manhattan. There are about 200 such users. We also restrict our estimation sample to users who do not reveal a change in residence or workplace.[7] This yields an estimation sample with 137 non-moving users. Their home and work locations are depicted in Figure 2. Consistent with broader patterns, our identified users have a high concentration of employment in Manhattan below 59th Street and a more dispersed residential pattern.

Figure 3 displays all the Yelp venues visited by the 137 users in our estimation sample. Unsurprisingly, visits are heavily concentrated in lower Manhattan. But these users also visit venues in other boroughs of New York City.

Figure 4 illustrates the nature of these data for two individuals in our data, "User 7" and "User 11." Red dots denote Yelp venues reviewed by this user. The "H" denotes the average coordinates of those venues identified as residential locations in the text of this user's

[4] Another example of using the information disclosed in reviews is searching Yelp to detect outbreaks of food poisoning unreported to NYC health authorities (NPR).

[5] See data appendix A.1 for the 26 phrases.

[6] Thus, we found locational information for nearly one-fifth of the Yelp users we examined. The median user for which we obtained locational information had reviewed twenty venues in New York and New Jersey, while the median for which we obtained no information had reviewed five venues. Amongst users with more than ten reviews of NY/NJ venues, we obtained locational information for about 40%.

[7] See data appendix A.1 for details.

Figure 2: Located non-moving users who live and work in Manhattan

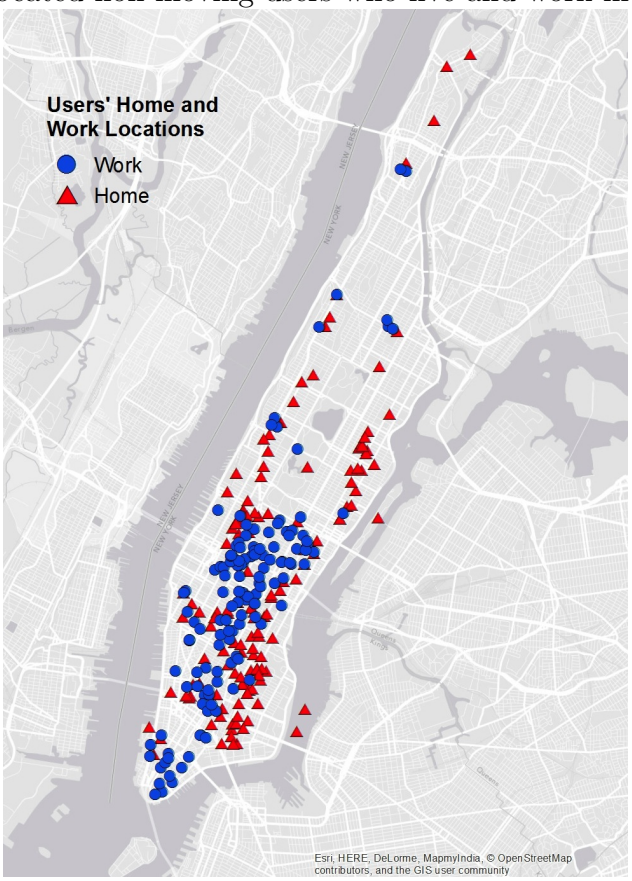


Figure 3: Yelp venues visited by users in estimation sample

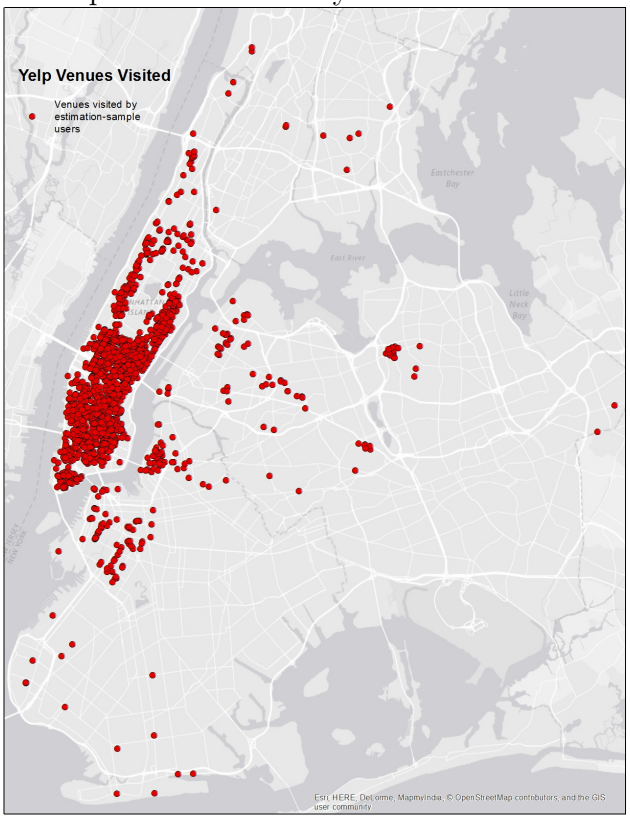
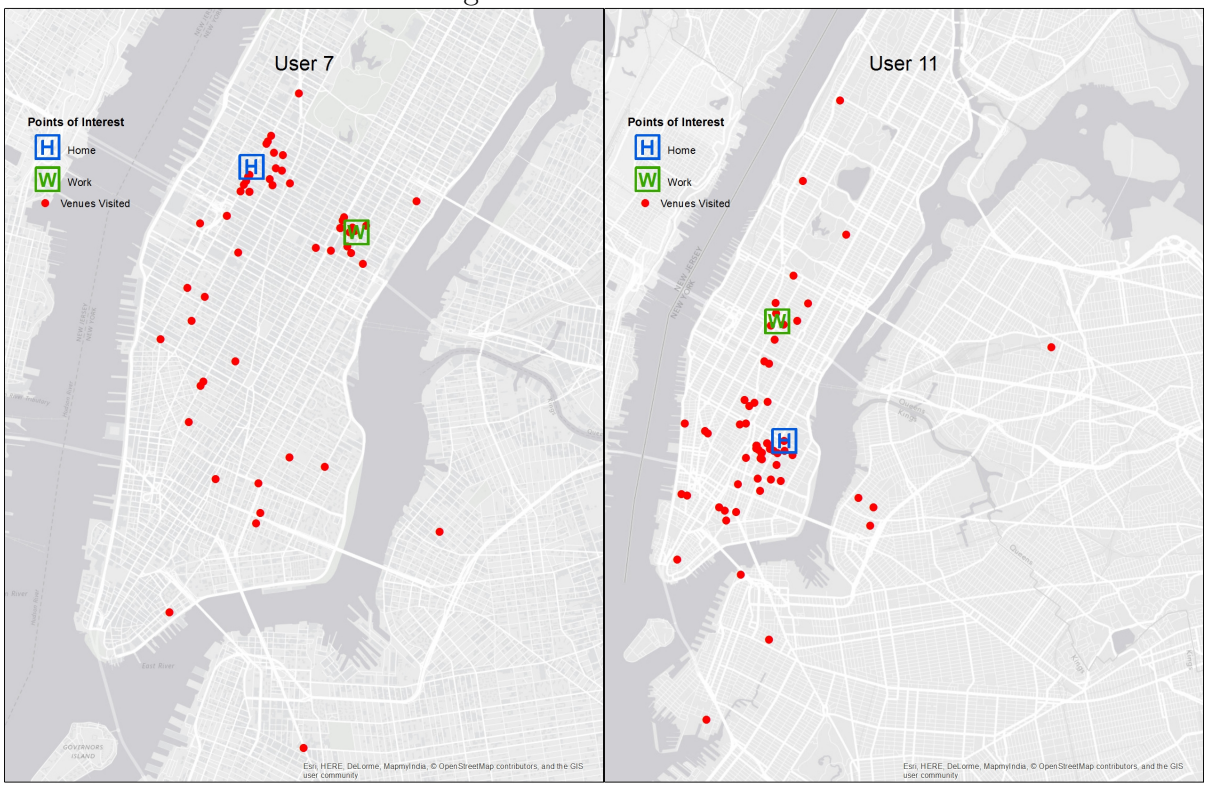


Figure 4: Users 7 and 11



reviews. The "W" denotes the similarly defined work location. User 7 lives and works in midtown Manhattan. User 11 works in midtown Manhattan and resides in a southeastern Manhattan development called Stuyvesant Town.

At a glance, the maps suggests that both proximity and venue characteristics influence user behavior. Both users primarily visits venues that are near their home or work locations. Both users visit more downtown venues than uptown venues, which may reflect differences in the quantity or quality of venues in these areas. We explore these issues by estimating a model of consumer choice, described in section 3.

Yelp users may also post information about themselves to their profile, such as a photo, their location, and hometown. In some estimation exercises, we use a user's apparent gender and ethnicity or race, inferred from their user photo, to examine how consumer behavior depends on the interaction of user characteristics and venue characteristics.[8] A previous study compared such observational measures of ethnicity and race inferred from photos with administrative data and found a high degree of accuracy in partitioning subjects into three groups: Asian, Black, and White or Hispanic (Mayer and Puller, 2008).

Table 1: User-level summary statistics

	Users with Manhattan home and work		Estimation sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Number of reviews	36.74	41.361	24.701	24.719
Female	0.593	0.492	0.577	0.496
Male	0.353	0.479	0.394	0.49
White	0.309	0.463	0.321	0.469
Hispanic	0.025	0.155	0.022	0.147
White or Hispanic	0.157	0.365	0.146	0.354
Asian	0.23	0.422	0.234	0.425
Black	0.034	0.182	0.022	0.147
Race not determined	0.245	0.431	0.255	0.438
Mover	0.328	0.471		
Total population, home census tract			6586.518	2988.61
Median income, home census tract			58495.299	28262.725
Number of observations	204		137	

Table 1 reports summary statistics for these users' gender, demographic, and locational information. The first two column describe all users, including those who revealed changes in residence and workplace locations. The latter two columns describe the users in our estimation sample, which is restricted to non-movers. Females constitute more than half of our users.[9] Very few users were identified as non-white Hispanics or blacks. Asians are

[8]While users may choose "male" or "female" for their gender on their Yelp profile, this information is not publicly displayed on their profile. All our empirical results describing differences between men and women in fact describe differences between users who were classified by the researchers based on their gender presentation in their profile photo. Some profile photos do not present a gender (e.g. cartoon graphics, photos of animals).

[9]Users with profile photos for which we could not classify the gender have both male and female dummy

overrepresented relative to their share of the general population of Manhattan. These Yelp users are presumably younger than the average Manhattan resident, though we have not attempted to infer user age. The users in our estimation sample tend to live in Manhattan census tracts with median incomes higher than typical of Manhattan as a whole. The average for tract median income in our estimation sample exceeds $58,000, while the population-weighted mean is about $49,500.

2.3 NYC crime, demographic, and transportation data

We combine this information on individual Yelp users' trips to venues with data describing locations' crime rates, income levels, and residential racial or ethnic composition and estimates of the time required to travel between locations.

Much of this information is reported at the level of census tracts. Census tracts are geographic units defined by the US Census Bureau based on population. In New York City, census tracts have an average population of about 3600, though some tracts have zero population because they cover parks or waterfronts. In 2000, Manhattan was made up of 296 census tracts and New York City was made up of 2216 census tracts.

To describe crime rates, we scraped the NYC Crime Map for robberies in 2013. These data provide a block-by-block account of robberies, which we aggregated to the level of year 2000 census tracts.[10] Historical crime data for prior years are only available at level of precincts, which are much larger geographic units. We use robberies as our crime measure because these are the most common and relevant threat to individuals visiting a Yelp venue. By contrast, burglaries are more relevant for residents than visitors; rapes, assaults and murders are certainly of real interest as potential fears, but their measures are variably polluted by the fact that the attacker may frequently or even predominantly be known to the victim, so these are potentially not clean measures of the threats that deters visits to venues.

Our data on income levels and residential racial/ethnic composition come from the 2000 Census of Population, made available via the National Historical Geographic Information System (NHGIS). These data describe each census tract's median income, total population, and population counts for whites, blacks, Asians, and Hispanics.[11] These are summarized in Table 2; the population counts are depicted in Figure 1.

To measure racial/ethnic demographic distances between two census tracts, we calculate the Euclidean distance between the two tract's population shares for five demographic groups. Stacking these population shares in a five-element vector $shares_{tract}$, the "Euclidean demographic distance" between origin and destination tracts is

$$\|shares_{origin} - shares_{destination}\|,$$

where $\|\cdot\|$ indicates the L^2 norm and $\|shares_{tract}\| = 1$ by definition. Figure 5 illustrates this for two census tracts, one in Morningside Heights with a diverse population that is similar to

variables equal to zero.

[10] We assigned each incident to a census tract based on the distances to tracts' latitude-longitude centroids.

[11] To be precise, we divide the population into five racial/ethnic groups and use the population counts of non-Hispanic whites, non-Hispanic blacks, non-Hispanic Asians, and all Hispanics. The remainder of the population includes Native Americans, Hawaiians, other races, and mixed-race categories.

Table 2: Census tract summary statistics

Variable	**Mean**	**Std. Dev.**	**N**
Median income	41056.006	19714.431	2163
Total population	3651.746	2447.35	2193
Share whites	0.356	0.325	2193
Share blacks	0.262	0.313	2193
Share Asian	0.093	0.122	2193
Share Hispanic	0.231	0.209	2193
Robberies	8.448	8.287	2193
Robberies per capita	0.016	0.389	2193
Robberies per 1000 sq feet	0.005	0.006	2193

This table describes year 2000 NYC census tracts with positive population. The robberies data are year 2013.

Table 3: Census tract pair summary statistics

	Mean	Std. Dev.	Observations
Geodetic distance	13.57	6.687	644742
Travel time (minutes)	57.261	22.701	644073
Euclidean demographic distance	0.636	0.346	644742

This table describes pairs of year 2000 NYC census tracts in which both have positive population and at least one of the tracts is located in Manhattan. The mass-transit travel time comes from Google Maps.

most tracts and one in Manhattan Chinatown that is overwhelmingly Asian and thus quite demographically distant from most tracts.

To measure segregation, we calculate the Echenique and Fryer (2007) spectral segregation index for the modal race/ethnicity in each census tract. This particular index has the property that a census tract is more segregated if it is surrounded by more segregated tracts.

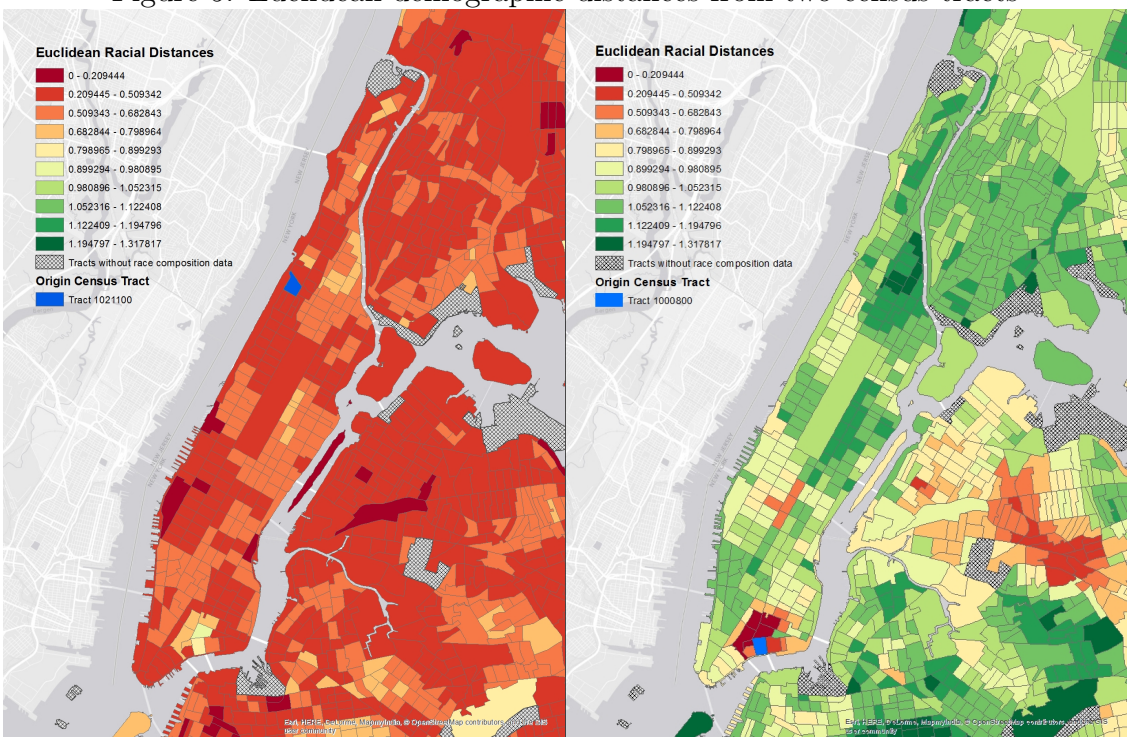
To estimate travel times, we requested the public-transit travel time between the centroids of New York City census tracts from Google Maps. This yielded a 2216×2216 symmetric matrix of estimated travel times. We used tract centroids to avoid the computational burden of making hundreds of thousands of Google Maps queries. We restricted our estimation sample to users with home and work locations in Manhattan in order to mitigate the issue of transport-mode choice, since the large majority of Manhattan residents use public transit.

In many specifications, we use the travel time to a Yelp venue that would be incurred by deviating from a commuting path between a user's home and work locations. We calculate the commuting-path travel time $time_{cj}$ as

$$time_{cj} = \frac{1}{2}\left(time_{hj} + time_{wj} - time_{hw}\right)$$

where $time_{lj}$ is the travel time from location l to venue j and $time_{hw}$ is the travel time from

Figure 5: Euclidean demographic distances from two census tracts



home to work.

3 Empirical approach

We describe individuals' decisions to visit Yelp venues using a standard discrete-choice model.

3.1 Behavioral model

Individuals are continuously taking decisions about where to dine: individuals must choose whether to visit a venue and, if they do, which venue to visit. We index individuals by i, venues by j, and we index by t the occasions in which i needs to decide on whether to visit a venue. We denote the set of venues considered as potential choices at period t by individual i as J_{it}. The choice set J_{it} is specific to individual i and period t and a subset of the set of all venues active at t, J_t. We denote the outside option of not visiting any venue as $j = 0$, and we assume that it belongs to every set J_{it}.

When visiting a venue, individuals need to travel from their current location to the venue. We index origin locations by l and assume that trips to venues may be of three types: trips from home ($l = h$), trips from work ($l = w$), or deviations from the commuting path between home and work ($l = c$).

We adopt the standard random-utility representation of preferences and assume that the utility for individual i of visiting venue j in period t from location l may be written as

$$U_{ijlt} = V_{ijlt} + \nu_{ijlt} = \beta X_{ijlt} + \nu_{ijlt}, \tag{1}$$

where the X denotes covariates that are observable to the econometrician, and ν is a scalar that is unobserved to the econometrician. We assume that the utility attached to the outside option is

$$U_{i0lt} = \nu_{i0lt}.$$

We define a variable d_{ijlt} that takes value 1 if individual i choses to travel to venue j from location l at period t:

$$d_{ijlt} = \mathbb{1}\{U_{ijlt} \geq U_{ij'l't}; \forall\ j' \in J_{it},\ l' \in \{h, w, c\}\}.$$

Analogously, we define a variable d_{ijt} that ignores the point of origin and takes value 1 if individual i chooses alternative j at period t:

$$d_{ijt} = \sum_{l \in \{h,w,c\}} d_{ijlt}.$$

For every individual i and period t, we observe the value of d_{ijt} for every venue $j \in J_{it}$. Given $d_{ijt} = 1$, we do not observe whether $d_{ijht} = 1$, $d_{ijwt} = 1$ or $d_{ijct} = 1$.

We assume that the vector of unobserved utilities for individual i at period t, ν_{it}, is independent across individuals and time periods and has joint cumulative distribution function

$$\exp\Big(-\sum_{j=1}^{J_{it}}\Big(\sum_{l=\{h,w,c\}} \exp(-(\nu_{ijlt}/\lambda))\Big)^{\lambda}\Big).$$

This distribution yields a nested-logit discrete-choice model, where each nest corresponds to a venue $j \in J_{it}$. For any two different venues j and j', this model assumes that the unobserved portion of the utility function is uncorrelated: $cov(\nu_{ijt}, \nu_{ij't}) = 0$. For a given venue, the unobserved components of utility that correspond to trips from different origins may be positively correlated. This correlation accounts for unobservable characteristics of venues. The parameter λ is a measure of the degree of independence across unobservables that correspond to the same venue. A higher value of λ implies greater independence.[12] Given our distributional assumptions, the probability that individual i visits venue j from location l at period t is

$$P(d_{ijlt} = 1|X_{it}, J_{it}; \beta, \lambda) = \frac{\exp(\beta X_{ijlt}/\lambda)}{\sum_{l'=\{h,w,c\}} \exp(\beta X_{ijl't}/\lambda)} \frac{\left(\sum_{l'=\{h,w,c\}} \exp(\beta X_{ijl't}/\lambda)\right)^{\lambda}}{\sum_{j' \in J_{it}} \left(\sum_{l'=\{h,w,c\}} \exp(\beta X_{ij'l't}/\lambda)\right)^{\lambda}},$$

where X_{it} is the vector of observable characteristics for every venue in the choice set J_{it} and every possible origin. The probability that individual i visits venue j at period t is simply

$$\begin{aligned} P(d_{ijt} = 1|X_{it}, J_{it}; \beta, \lambda) &= \sum_{l \in \{h,w,c\}} P(d_{ijlt} = 1|X_{it}, J_{it}) \\ &= \frac{\left(\sum_{l=\{h,w,c\}} \exp(\beta X_{ijlt}/\lambda)\right)^{\lambda}}{\sum_{j' \in J_{it}} \left(\sum_{l=\{h,w,c\}} \exp(\beta X_{ij'lt}/\lambda)\right)^{\lambda}}. \end{aligned} \tag{2}$$

In order to identify the parameter vector (β, λ), we a random sample of reviews posted on Yelp, rather than information from a random sample of the population of interest. This implies that it is a random sample of the subset of elements of the population that verify two conditions: (a) individual i decided at t to go to some venue (i.e. she did not choose the outside option $j = 0$) (b) individual i decided at t to write a review on Yelp that reveals her venue choice. Conditional on having visited a venue, individuals make the binary decision of writing a Yelp review or not. We introduce a dummy variable d^r_{ijt} taking value 1 when individual i writes a review on Yelp about a venue that was visited at period t. We will assume that the probability that an individual i writes a review of a venue visited at period t does not depend on the identity of the venue.

Under these assumptions, we can write the probability that we observe a review about a choice j by individual i at period t as

$$\begin{aligned} P(d^r_{ijt} = 1|X_{it}, R_{it}; \beta, \lambda) &= P(d^r_{ijt} = 1|d_{ijt} = 1, X_{it}, R_{it}; \beta, \lambda) P(d_{ijt} = 1|X_{it}, R_{it}; \beta, \lambda) \\ &= w_{it} \mathbb{1}\{j \neq 0\} P(d_{ijt} = 1|X_{it}, R_{it}; \beta, \lambda), \end{aligned} \tag{3}$$

[12] In particular, when λ is equal to 1, all the error terms corresponding to the same venue and time period become independent:

$$G(\nu_{ijt}) = G(\nu_{ijht}, \nu_{ijwt}, \nu_{ijct}) = G(\nu_{ijht}) G(\nu_{ijwt}) G(\nu_{ijct}).$$

In this case, the nested logit model becomes a simple multinomial logit.

where the first equality assumes that individuals only write reviews about choices they actually made, and the second equality assumes that the probability that individual i writes a review about a choice j is 0 when such choice was the outside option and some constant when such choice was to visit a venue. This restriction implies that the probability that individual i writes a Yelp review about a venue visited at some period t may be different across individuals and time periods. However, for each individual and time period, it must be constant across venues. The constant w_{it} denotes the probability that individual i writes a review at period t.

3.2 Identification and estimation

Using the conditional probability in equation (3) for identification of the parameter vector (β, λ) requires knowing the choice set J_{it} for every i and t. We do not directly observe this choice set. However, as McFadden (1978) and Train, McFadden, and Ben-Akiva (1987) show, under the assumptions imposed above, the correct specification of the choice set J_{it} is not necessary for the identification of the parameters (β, λ).[13]

For each individual i and time period t, we define a set $\mathcal{J}_{it}$ that is assumed to be a subset of the true consideration set $\mathcal{J}_{it} \subseteq J_{it}$.[14] We denote as R_{it} the set of alternatives that we will assign to individual i at period t in our estimation procedure. Following McFadden (1978) and Train, McFadden, and Ben-Akiva (1987), we construct R_{it} by including i's observed choice at period t plus a random subset of the other alternatives included in the set $\mathcal{J}_{it}$. Each of the choices included in R_{it} other than the actual choice of i are selected from $\mathcal{J}_{it}$ with equal probability.[15] We denote by $\pi(R_{it}|d^r_{ijt} = 1)$ the probability of assigning the subset R_{it} to an individual i at period t who wrote a review about venue j. The conditional probability of an individual i writing a review about venue j at period t, given a sample R_{it} randomly drawn by the econometrician, is:

$$\begin{aligned} P(d^r_{ijt} = 1|X_{it}, J_{it}, R_{it}; \beta, \lambda) &= \frac{P(R_{it}|d^r_{ijt} = 1, X_{it}, J_{it})P(d^r_{ijt} = 1|X_{it}, J_{it}; \beta, \lambda)}{\sum_{j' \in J_{it}} P(R_{it}|d^r_{ij't} = 1, X_{it}, J_{it})P(d^r_{ij't} = 1|X_{it}, J_{it}; \beta, \lambda)}, \\ &= \frac{\pi(R_{it}|d^r_{ijt} = 1)P(d^r_{ijt} = 1|X_{it}, J_{it}; \beta, \lambda)}{\sum_{j' \in J_{it}} \pi(R_{it}|d^r_{ij't} = 1)P(d^r_{ij't} = 1|X_{it}, J_{it}; \beta, \lambda)}, \\ &= \frac{\pi(R_{it}|d^r_{ijt} = 1)P(d^r_{ijt} = 1|X_{it}, J_{it}; \beta, \lambda)}{\sum_{j' \in R_{it}} \pi(R_{it}|d^r_{ij't} = 1)P(d^r_{ij't} = 1|X_{it}, J_{it}; \beta, \lambda)}, \end{aligned} \tag{4}$$

as long as $j \in R_{it}$, and 0 otherwise. The first equality comes from a strict application of Bayes' rule. The second equality accounts for the fact that our procedure to draw samples of venues only depends on the actual choice of individual i at period t. Finally, the third

[13] McFadden (1978) shows that one can identify the parameters of a multinomial logit model from a likelihood function that assigns to each agent a choice set that is strictly smaller than their actual set of alternatives. Train, McFadden, and Ben-Akiva (1987) extend this result to the nested-logit case and shows that one may identify the parameters from a likelihood function that restricts the set of nests from which agents may choose.

[14] In our baseline estimation, we assume that $\mathcal{J}_{it}$ is the set of Yelp venues visited by at least one of the users in the estimation sample and does not include the outside option $j = 0$.

[15] This assignment mechanism satisfies the positive conditioning property (see McFadden 1978).

equality takes into account that: (a) $R_{it} \subseteq J_{it}$, and; (b) $\pi(R_{it}|d_{ij't} = 1) = 0$ for every $j' \notin R_{it}$.[16]

Combining equations (2), (3) and (4), we obtain that, for every $j \in J_{it}$ and $j \neq 0$,

$$P(d^r_{ijt} = 1|X_{it}, R_{it}; \beta, \lambda) = \frac{\pi(R_{it}|d^r_{ijt} = 1) w_{it} \Big(\sum_{l=\{h,w,c\}} \exp(V_{ijlt}/\lambda) \Big)^{\lambda}}{\sum_{j' \in R_{it}} \Big\{ \pi(R_{it}|d^r_{ij't} = 1) w_{it} \Big(\sum_{l=\{h,w,c\}} \exp(V_{ij'lt}/\lambda) \Big)^{\lambda} \Big\}}$$

$$= \frac{\pi(R_{it}|d^r_{ijt} = 1) \Big(\sum_{l=\{h,w,c\}} \exp(V_{ijlt}/\lambda) \Big)^{\lambda}}{\sum_{j' \in R_{it}} \Big\{ \pi(R_{it}|d^r_{ij't} = 1) \Big(\sum_{l=\{h,w,c\}} \exp(V_{ij'lt}/\lambda) \Big)^{\lambda} \Big\}}, \qquad (5)$$

where the second equality divides by w_{it} in the numerator and denominator.

Note that this probability does not depend on the true unobserved choice set J_{it}. We can simplify this expression further. Given that our procedure to randomly generate R_{it} is such that R_{it} includes the chosen alternative j with probability 1 and all the other remaining alternatives in $\mathcal{J}_{it}$ with identical probability, then

$$\pi(R_{it}|d^r_{ijt} = 1) = \pi(R_{it}|d^r_{ij't} = 1), \qquad \text{for any } j, j' \in R_{it}.$$

McFadden (1978) defines this as the uniform conditioning property. Under this property, equation (5) simplifies to

$$P(d^r_{ijt} = 1|X_{it}, R_{it}; \beta, \lambda) = \frac{\Big(\sum_{l=\{h,w,c\}} \exp(V_{ijlt}/\lambda) \Big)^{\lambda}}{\sum_{j' \in R_{it}} \Big\{ \Big(\sum_{l=\{h,w,c\}} \exp(V_{ij'lt}/\lambda) \Big)^{\lambda} \Big\}}, \qquad (6)$$

for every $j \in R_{it}$, and 0 otherwise. Note that the resulting likelihood is identical to that implied by a nested logit model in which individual i at period t is assumed to have a consideration or choice set identical to the randomly chosen set R_{it}.

Given a sample of Yelp reviews $t = 1, \ldots, T_i$ and individuals $i = 1, \ldots, I$, we can easily write a likelihood function for this sample by combining the expression in equation (6) with the assumption that the probability that we observe a review for venue j written by individual i at period t is independent of the other reviews we observe in our sample, conditional on the vector of observable characteristics for individuals and venues.

4 Estimation results

This section reports the results of estimating discrete-choice models of the form described in section 3 using the data described in section 2. We systematically expand the set of included covariates, X_{ijlt}, by introducing measures that describe travel time, demographic differences, crime rates, and individual characteristics to arrive at our preferred specification that we use to construct counterfactuals.

[16] Given that our procedure to define the set R_{it} includes the chosen alternative with probability one, the probability of drawing R_{it} given a choice not included in R_{it} is zero.

In every model we estimate, the choice set R_{it} contains 20 elements, the venue chosen by the individual at that time and 19 randomly drawn venues. We have experimented with larger choice sets and found that they do not meaningfully alter the estimated coefficients nor significantly improve their precision.

A user making a trip to a Yelp venue may start that trip from her home location, work location, or during her commute, as described in section 3. Thus, while R_{it} contains 20 elements, each user chooses one of 60 possible choices when there are three possible points of origin.

In this draft, we estimate multinomial-logit models, restricting $\lambda = 1$. We will relax this assumption in future work.

4.1 Travel time

Our first set of model specifications describes how travel times influence individuals' consumption choices.

Table 4 presents estimates for five logit models, one model per column. In the first column, we assume that all trips originate from home and include the log of the number of minutes from home to the venue as the sole element of X_{ijlt}, where $l = h$. In the second column, we incorporate both home and workplace origins, $l \in \{h, w\}$, using the log travel time from each. In the third and succeeding columns, we also incorporate the commuting path, $l = c$, as a third origin point.

Spatial frictions matter. We see that travel time plays a first-order role in individuals' consumption choices within the city. Yelp users are less likely to visit venues that are more distant from their home and work locations in terms of mass-transit travel time. In the first column, we find a large, negative coefficient on travel time from home. This finding is consistent with the patterns illustrated in Figures 4 . In the second column, we obtain a similar coefficient on travel time from home and a more negative coefficient on travel time from work. This suggests that individuals are more time-sensitive when travel originates at their workplace, which may reflect a difference in the nature of these venue trips (e.g. a short lunch break during the workday versus a social occasion in the evening). In the third column, we introduce the travel time to the venue from the user's commuting path. This also exhibits a significant, negative coefficient.

The fourth and fifth columns of Table 4 examine the robustness of these travel-time results. In the fourth column, we introduce venue characteristics, prices and average rating, as additional regressors. In the fifth column, we introduce dummy variables for the venues' neighborhoods to control for unobserved locational characteristics. The results of the first three columns are robust to the introduction of these additional controls.

In the fourth column, we control for observable venue characteristics. The travel-time coefficients are largely unchanged from the third column. Individuals are less likely to visit venues located in higher-income census tracts. Individuals are more likely to visit venues that have higher ratings on Yelp. This is unsurprising. Individuals are more likely to choose a venue that Yelp lists as more expensive. This suggests that venues' prices and qualities are positively correlated, and that the average rating on Yelp does not fully capture venue quality. This is unsurprising, both because we have not controlled for cuisine types and because we believe that Yelp users often assign ratings within cuisine-price categories. For

Table 4: Travel time results

Log of travel time, Home	-.82735	-.81869	-.59469	-.63912	-.85996
Log of travel time, Home (t-stat)	(-35.237)	(-37.875)	(-17.958)	(-15.545)	(-14.916)
Log of travel time, Work		-1.6841	-2.9976	-2.6191	-1.9283
Log of travel time, Work (t-stat)		(-15.145)	(-1.7896)	(-2.6863)	(-5.2965)
Log of travel time, Path			-1.0566	-1.0120	-.92713
Log of travel time, Path (t-stat)			(-9.7115)	(-9.4300)	(-16.549)
Dummy for $$				.272650	.250449
Dummy for $$ (t-stat)				(5.74090)	(5.19229)
Dummy for $$$				.301990	.347079
Dummy for $$$ (t-stat)				(5.11929)	(5.75400)
Dummy for $$$$				.250640	.396210
Dummy for $$$$ (t-stat)				(2.70740)	(4.18970)
Average rating on Yelp				.312590	.243249
Average rating on Yelp (t-stat)				(8.48239)	(6.47329)
Number of points of origin	1	2	3	3	3
Number of choices per observation	20	20	20	20	20
Neighborhood dummies					Yes
Log-Likelihood	-2.8303	-2.7830	-2.8027	-2.7827	-2.7623
Pseudo R^2	.055236	.070974	.064414	.071074	.077922
Akaike Information Criterion	7.66050	9.56620	11.6059	19.5659	49.5250
Number of observations	3384	3384	3384	3384	3384
Number of users	137	137	137	137	137
Number of venues	1961	1961	1961	1961	1961
Logit type	multinomial	multinomial	multinomial	multinomial	multinomial
Column build date	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014

example, both El Rey Del Taco Truck and Per Se have 4.5 stars on Yelp, with the former selling $2 tacos and the latter offering a $295 tasting menu.

In the fifth column, we introduce dummies for venue neighborhoods.[17] Our qualitative findings are unchanged, but the travel-time coefficients exhibit some meaningful differences compared to the previous columns. The coefficient on travel time from home is more negative, while the coefficient for work is less negative.

4.2 Ethnicity and race

Our second set of model specifications describes how demographic differences influence individuals' consumption choices. Table 5 presents estimates for six logit models, one model per column.

The first two columns show that users are less likely to visit venues located in census tracts with demographics different from those of their home census tract. In the first column, there is a large, negative coefficient on the Euclidean demographic distance between the census tract in which the user resides and the census tract in which the restaurant venue is located. The second column shows that this Euclidean demographic distance has a particularly large, negative effect when the venue is located in a census tract that is demographically isolated, as measured by the spectral segregation index.

[17]We aggregate Yelp-defined neighborhoods to partition New York City into 17 neighborhoods, 13 in Manhattan and four "neighborhoods" that are the four other boroughs. See appendix section A.2.

The third and fourth columns introduce introduces travel times as additional regressors, which significantly reduces the coefficient on the Euclidean demographic distance. Part of the reason that users are less likely to visit venues in demographically distant tracts is that these venues are also physically distant. Residential segregation in physical space plays a clear role in segregating economic interactions. Nonetheless, even after accounting for travel times, there is a large, negative coefficient on the Euclidean demographic distance. Users are less likely to visit restaurant venues in demographically different census tracts than would be predicted by the geography of residential segregation. This holds true when we also account for observable venue characteristics (column 5) and unobservable neighborhood characteristics (column 6).

These findings may capture at least two distinct scenarios. First, users of all races may prefer to spend time in places with demographics similar to those in their residential tract. Second, individuals may prefer to spend time in places with demographics similar to their own. An example of the first would be if a white individual living in Harlem were more likely to visit black and Hispanic neighborhoods than a white individual living elsewhere. An example of the second would be if black and Hispanic individuals are more likely to visit black and Hispanic neighborhoods and Yelp users living in Harlem are more likely to be black or Hispanic. In later versions of the paper, we will attempt to distinguish between these scenarios by using individual users' demographics inferred from their appearance in their user profile photos.

Table 5: Ethnicity and race results

Log of travel time, Home			-.51030	-.52346	-.61972	-.82498	-.61610	-.82880	-.63993	-.81853
Log of travel time, Home (t-stat)			(-16.836)	(-14.527)	(-14.539)	(-13.260)	(-12.435)	(-13.185)	(-11.164)	(-13.074)
Log of travel time, Work			-2.1747	-2.9879	-2.4479	-1.8467	-2.5411	-1.8674	-2.4059	-1.7863
Log of travel time, Work (t-stat)			(-2.9890)	(-.82919)	(-3.0539)	(-5.2975)	(-2.7191)	(-5.0559)	(-2.9419)	(-5.5100)
Log of travel time, Path			-1.2898	-1.1087	-.98768	-.87240	-.92864	-.85478	-.89381	-.86489
Log of travel time, Path (t-stat)			(-6.1012)	(-6.0718)	(-9.3267)	(-14.937)	(-8.5661)	(-15.348)	(-8.4457)	(-14.659)
Euclidean demographic distance between h and l_j	-1.3516	-.99466	-.48348	-.40996	-.33131	-.61523	-.28742	-.57126	-.25655	-.58579
Euclidean demographic distance between h and l_j (t-stat)	(-15.442)	(-8.5516)	(-4.7441)	(-2.9863)	(-2.8273)	(-4.9468)	(-2.3659)	(-4.4622)	(-2.1575)	(-4.5451)
Spectral segregation index in l_j		.035884		-.15250	-.13312	-.08281	-.12185	-.08925	-.12646	-.08388
Spectral segregation index in l_j (t-stat)		(.875209)		(-2.8013)	(-3.0292)	(-1.4642)	(-2.7537)	(-1.5709)	(-2.8787)	(-1.4720)
Interaction of SSI with Euclidean demographic distance		-.59367		-.31330	-.38784	-.22938	-.41016	-.23255	-.35251	-.23278
Interaction of SSI with Euclidean demographic distance (t-stat)		(-5.0661)		(-2.2353)	(-3.2265)	(-1.9114)	(-3.3499)	(-1.9122)	(-2.8678)	(-1.8838)
Dummy for $$					.23917	.212929	.249860	.220229	.256119	.235159
Dummy for $$ (t-stat)					(5.02549)	(4.39580)	(5.26399)	(4.56069)	(5.38089)	(4.85739)
Dummy for $$$					.294569	.304760	.286100	.302269	.293309	.368640
Dummy for $$$ (t-stat)					(4.96700)	(5.03739)	(4.82079)	(4.99669)	(4.92409)	(6.08080)
Dummy for $$$$					.235029	.373389	.301230	.380070	.271400	.412209
Dummy for $$$$ (t-stat)					(2.51729)	(3.95510)	(3.20510)	(3.99920)	(2.89269)	(4.35370)
Average rating on Yelp					.312400	.248799	.302940	.239859	.314099	.246309
Average rating on Yelp (t-stat)					(8.47459)	(6.62779)	(8.19110)	(6.36660)	(8.59790)	(6.57189)
Absolute difference of median incomes $(h - l_j)$							-.09971	.013183		
Absolute difference of median incomes $(h - l_j)$ (t-stat)							(-.94976)	(.123199)		
Percentage difference of median incomes $(h - l_j)$									-.11163	-.02748
Percentage difference of median incomes $(h - l_j)$ (t-stat)									(-2.5792)	(-.61033)
Number of points of origin	0	0	3	3	3	3	3	3	3	3
Number of choices per observation	20	20	3	3	20	20	20	20	20	20
Neighborhood dummies						Yes		Yes		Yes
Log-Likelihood	-2.9551	-2.9493	-.98110	-.97582	-2.7667	-2.7493	-2.7739	-2.7537	-2.7704	-2.7495
Pseudo R^2	.013543	.015492	.106959	.111769	.076445	.082271	.074017	.080764	.075181	.082158
Akaike Information Criterion	7.91029	11.8990	9.96220	13.9519	25.5330	55.4990	27.5480	57.5079	27.5410	57.4990
Number of observations	3384	3384	3384	3384	3384	3384	3384	3384	3384	3384
Number of users	137	137	137	137	137	137	137	137	137	137
Number of venues	1961	1961	1961	1961	1961	1961	1961	1961	1961	1961
Logit type	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial
Column build date	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014

4.3 Crime

Our third set of model specifications describes how crime rates influence individuals' consumption choices.

Table 6 presents estimates for 9 logit models, one model per column.

Users are less likely to visit venues in areas where more robberies occur. The first column reveals that there is no statistically significant correlation between consumption choices and robbery rates when we omit all other regressors. If anything, the positive coefficient suggests that people are more likely to visit venues in places where more robberies occur. The second column shows that there is a large, negative coefficient on the number of robberies in the venue's census tract after we control for travel times and demographics. The third column shows that the estimated coefficient is robust to the introduction of observable venue characteristics and neighborhood dummies.

Users' consumption choices are responsive to the number of robberies, not robberies per capita.[18] The fourth column shows that there is no statistically significant correlation between consumption choices and per-capita robbery rates when we omit all other regressors. This non-relationship is also present in the next two columns, where we control for travel times and demographics and introduce venue characteristics and neighborhood dummies.

There is some evidence that users are less likely to visit venues in places where there is a greater density of robberies. While the seventh column shows that individuals are more likely to visit venues in places where more robberies per square foot occur, the next two columns show that there is a negative effect of robbery density on visits after controlling for travel times, demographics, venue characteristics and neighborhood dummies.

[18]There is no consensus about the proper measure of crime relevant for economic decisions. Ihlanfeldt and Mayock (2010) observe that studies have used the number of crimes or have deflated this using census population or the physical area. Crime per capita is at first appealing, as it may appear to be a suitable measure of the local incidence of crime. However, Ihlanfeldt and Mayock note that this has been criticized by criminologists on the grounds that it exaggerates risks in central business districts with very large inflows of daily visitors and a relatively low resident population. This concern is certainly relevant for a study of Manhattan (a concern echoed by the New York Police Department here). Bowes and Ihlanfeldt (2001)argue that crimes per unit area is highly relevant when we are interested in the knowledge of crimes in a resident's own neighborhood. Here we are interested in crimes in neighborhoods near and far. In summary, we decided to use total robberies without deflation by population or area because the former would likely compound concerns relevant for central business districts and also because total crimes may have greater salience in the public consciousness that governs choices of where to visit when this includes areas afield from one's own residence.

Table 6: Crime results

Log of travel time, Home		-.61084	-.83671		-.59641	-.83411		-.60326	-.83236
Log of travel time, Home (t-stat)		(-14.942)	(-13.050)		(-13.333)	(-12.951)		(-12.451)	(-12.902)
Log of travel time, Work		-2.5603	-1.8919		-2.7100	-1.8233		-2.6442	-1.8906
Log of travel time, Work (t-stat)		(-2.7669)	(-4.7677)		(-2.2787)	(-5.3404)		(-2.3392)	(-4.9135)
Log of travel time, Path		-1.0068	-.85187		-.95037	-.86796		-.92447	-.85018
Log of travel time, Path (t-stat)		(-9.1696)	(-15.468)		(-8.6056)	(-14.819)		(-8.5025)	(-15.196)
Euclidean demographic distance between h and l_j		-.21769	-.46924		-.23804	-.60394		-.23632	-.57849
Euclidean demographic distance between h and l_j (t-stat)		(-1.7912)	(-3.6240)		(-1.9868)	(-4.7350)		(-1.9219)	(-4.4433)
Spectral segregation index in l_j		-.13403	-.09952		-.09289	-.06346		-.10727	-.09292
Spectral segregation index in l_j (t-stat)		(-3.0034)	(-1.7483)		(-2.0924)	(-1.1231)		(-2.4261)	(-1.6318)
Interaction of SSI with Euclidean demographic distance		-.39851	-.25646		-.45271	-.18716		-.43830	-.18524
Interaction of SSI with Euclidean demographic distance (t-stat)		(-3.2067)	(-2.0975)		(-3.6370)	(-1.5463)		(-3.5174)	(-1.5203)
Dummy for $$			.218830			.232360			.226080
Dummy for $$ (t-stat)			(4.52370)			(4.79479)			(4.68520)
Dummy for $$$			.322499			.321530			.300779
Dummy for $$$ (t-stat)			(5.32679)			(5.30480)			(4.97760)
Dummy for $$$$			.372370			.330080			.360430
Dummy for $$$$ (t-stat)			(3.92350)			(3.46350)			(3.80320)
Average rating on Yelp			.234410			.255860			.257970
Average rating on Yelp (t-stat)			(6.23640)			(6.78870)			(6.85769)
Robberies per capita in venue tract in 2013				.218370	.184729	.427480			
Robberies per capita in venue tract in 2013 (t-stat)				(1.00300)	(.746399)	(1.74870)			
Robberies in venue tract in 2013	.047008	-1.0473	-.84788						
Robberies in venue tract in 2013 (t-stat)	(.185530)	(-3.7234)	(-2.7244)						
Robberies in 2013 per 1000 sq ft in venue tract							2.25609	-7.8364	-7.8975
Robberies in 2013 per 1000 sq ft in venue tract (t-stat)							(.656740)	(-1.9017)	(-1.7835)
Percentage difference of median incomes ($h - l_j$)		-.12938	-.04556		-.09211	-.03966		-.11428	-.06529
Percentage difference of median incomes ($h - l_j$) (t-stat)		(-2.9872)	(-1.0047)		(-2.1577)	(-.87505)		(-2.6336)	(-1.4290)
Number of points of origin	0	3	3	0	3	3	0	3	3
Number of choices per observation	20	20	20	20	20	20	20	20	20
Neighborhood dummies			Yes			Yes			Yes
Log-Likelihood	-2.9956	-2.7785	-2.7530	-2.9955	-2.7857	-2.7492	-2.9956	-2.7850	-2.7527
Pseudo R^2	1.69500	.072470	.080989	.000045	.070114	.082282	.000020	.070344	.081107
Akaike Information Criterion	7.99149	21.5569	59.5060	7.99119	21.5709	59.4980	7.99130	21.5699	59.5060
Number of observations	3384	3384	3384	3384	3384	3384	3384	3384	3384
Number of users	137	137	137	137	137	137	137	137	137
Number of venues	1961	1961	1961	1961	1961	1961	1961	1961	1961
Logit type	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial	multinomial
Column build date	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014	31 Jul 2014

4.4 Gender

These user responses sometimes differ by gender. Tables 7 through 10 revisit the topics of Tables 4 through 6 while interacting the regressors of interest with a dummy indicating the user's gender.

In Table 7, there is weak evidence of gender-related differences in users' responses to travel times. When there are three origin points, there are no statistically significant differences between men and women in the coefficients on home and work travel times. However, we obtain a positive coefficient on the interaction of commuting-path travel time and the female dummy, indicating that women are less likely to visit venues that are proximate to their commuting path relative to men.

In Table 8, we find gender-related differences in user responses to demographics differences. Women show no significant differential aversion to visiting neighborhoods with a different racial or ethnic composition *per se*. But they are less likely to visit venues deep inside areas whose racial and ethnic composition is different, as shown by the triple interaction of the female dummy with the spectral segregation index and the Euclidean demographic distance.

In Table 9, we find gender-related differences in user responses to criminal activity. Female users are more averse to venues located in census tracts with more robberies than male users. In fact, our estimates show that the negative effect of robberies on users' propensities to visit are entirely attributable to the negative effects on female users. There is no statistically significant effect for male users. As before, we do not detect these effects when we use robberies per capita as the measure of criminal activity.

In Table 10, we confirm that each of these gender-specific responses is robust to including the other gender-specific covariates.

Table 7: Gender-differential results for travel times

	(1)	(2)	(3)	(4)	(5)
Log of travel time, Home	-.95551	-.90434	-.65131	-.67733	-.86246
Log of travel time, Home (se)	.041324	.035464	.038676	.039857	.072572
Log of travel time, Home (t-stat)	-23.121	-25.5	-16.840	-16.993	-11.883
Interaction of log of travel time with female dummy, Home	.139709	.143779	.039078	.038485	.032356
Interaction of log of travel time with female dummy, Home (se)	.050469	.044865	.074904	.077446	.097585
Interaction of log of travel time with female dummy, Home (t-stat)	2.76819	3.20469	.521719	.496919	.331559
Log of travel time, Work		-1.9081	-2.5443	-2.4960	-2.1538
Log of travel time, Work (se)		.197919	1.06980	1.00429	.649119
Log of travel time, Work (t-stat)		-9.6405	-2.3782	-2.4853	-3.3180
Interaction of log of travel time with female dummy, Work		.352820	-4.1192	-2.4888	.077033
Interaction of log of travel time with female dummy, Work (se)		.241740	55.5299	31.3570	.934960
Interaction of log of travel time with female dummy, Work (t-stat)		1.45949	-.07418	-.07937	.082392
Log of travel time, Path			-1.3393	-1.3401	-1.1323
Log of travel time, Path (se)			.193430	.189820	.131410
Log of travel time, Path (t-stat)			-6.9239	-7.0605	-8.6160
Interaction of log of travel time with female dummy, Path			.475259	.477050	.271510
Interaction of log of travel time with female dummy, Path (se)			.227139	.220589	.145830
Interaction of log of travel time with female dummy, Path (t-stat)			2.09240	2.16260	1.86189
Dummy for $$				.269699	.240639
Dummy for $$ (se)				.048140	.048973
Dummy for $$ (t-stat)				5.60220	4.91370
Dummy for $$$				.285860	.308039
Dummy for $$$ (se)				.059702	.061138
Dummy for $$$ (t-stat)				4.78800	5.03830
Dummy for $$$$				.303009	.406879
Dummy for $$$$ (se)				.093622	.095224
Dummy for $$$$ (t-stat)				3.23650	4.27279
Average rating on Yelp				.324200	.251989
Average rating on Yelp (se)				.037464	.038212
Average rating on Yelp (t-stat)				8.65369	6.59450
Number of points of origin	1	2	3	3	3
Number of choices per observation	20	20	20	20	20
Neighborhood dummies					Yes
Log-Likelihood	-2.8243	-2.7816	-2.7969	-2.7790	-2.7567
Pseudo R2	.057188	.071451	.066343	.072318	.079741
Akaike Information Criterion	9.64879	13.5629	17.5939	25.5580	55.5139
Number of observations	3298	3298	3298	3298	3298
Number of users	133	133	133	133	133
Number of venues	1922	1922	1922	1922	1922

Table 8: Gender-differential results for demographic differences

	(1)	(2)	(3)	(4)	(5)	(6)
Log of travel time, Home			-.56563	-.60773	-.63457	-.81362
Log of travel time, Home (se)			.036828	.045781	.047084	.060208
Log of travel time, Home (t-stat)			-15.359	-13.274	-13.477	-13.513
Log of travel time, Work			-3.0079	-2.7713	-2.6758	-1.9887
Log of travel time, Work (se)			1.70790	1.26750	1.0977	.434859
Log of travel time, Work (t-stat)			-1.7612	-2.1865	-2.4377	-4.5731
Log of travel time, Path			-1.0061	-.93660	-.94308	-.87287
Log of travel time, Path (se)			.112980	.104019	.100010	.058159
Log of travel time, Path (t-stat)			-8.9062	-9.0043	-9.4298	-15.008
Euclidean demographic distance between census_h and census_j	-1.2118	-1.0559	-.32978	-.24516	-.25275	-.47347
Euclidean demographic distance between census_h and census_j (se)	.150470	.202399	.145510	.200179	.199589	.200230
Euclidean demographic distance between census_h and census_j (t-stat)	-8.0542	-5.2175	-2.2664	-1.2247	-1.2663	-2.3645
Spectral segregation index of restaurant tract		-.01564		-.14338	-.15578	-.10615
Spectral segregation index of restaurant tract (se)		.072589		.079308	.078869	.085928
Spectral segregation index of restaurant tract (t-stat)		-.21555		-1.8079	-1.9752	-1.2353
Interact Euclidean demographic distance with female dummy	-.23172	.160989	-.28134	-.01399	-.02517	-.04124
Interact Euclidean demographic distance with female dummy (se)	.186939	.248229	.177900	.245499	.244550	.238839
Interact Euclidean demographic distance with female dummy (t-stat)	-1.2395	.648530	-1.5814	-.05698	-.10294	-.17271
Interaction of SSI with female dummy		.11479		.055491	.052094	.038571
Interaction of SSI with female dummy (se)		.087861		.095161	.094702	.093286
Interaction of SSI with female dummy (t-stat)		1.30649		.583119	.550090	.413480
Interaction of SSI with Euclidean demographic distance		-.26148		-.20534	-.17966	-.02892
Interaction of SSI with Euclidean demographic distance (se)		.187380		.197019	.194450	.190530
Interaction of SSI with Euclidean demographic distance (t-stat)		-1.3954		-1.0422	-.92394	-.15178
Triple interaction of SSI with Euclidean demographic distance with female dummy		-.66455		-.53374	-.49715	-.43138
Triple interaction of SSI with Euclidean demographic distance with female dummy (se)		.241390		.252469	.249369	.241720
Triple interaction of SSI with Euclidean demographic distance with female dummy (t-stat)		-2.7530		-2.1140	-1.9936	-1.7847
Dummy for $$					.253419	.228560
Dummy for $$ (se)					.048225	.049042
Dummy for $$ (t-stat)					5.25489	4.66050
Dummy for $$$					.286269	.304919
Dummy for $$$ (se)					.059953	.061129
Dummy for $$$ (t-stat)					4.77500	4.98799
Dummy for $$$$					.317699	.404449
Dummy for $$$$ (se)					.094028	.095230
Dummy for $$$$ (t-stat)					3.37879	4.24700
Average rating on Yelp					.312339	.252099
Average rating on Yelp (se)					.037572	.038309
Average rating on Yelp (t-stat)					8.31289	6.58069
Number of points of origin	0	0	3	3	3	3
Number of choices per observation	20	20	20	20	20	20
Neighborhood dummies						Yes
Log-Likelihood	-2.9547	-2.9461	-2.7934	-2.7809	-2.7639	-2.7465
Pseudo R2	.013669	.016521	.067535	.071718	.077370	.083162
Akaike Information Criterion	9.90960	17.8920	15.5869	23.5620	31.5279	57.4930
Number of observations	3298	3298	3298	3298	3298	3298
Number of users	133	133	133	133	133	133
Number of venues	1922	1922	1922	1922	1922	1922

Table 9: Gender-differential results for crime

	(1)	(2)	(3)	(4)	(5)	(6)
Robberies in venue tract in 2013	1.08529	.031245	.069737			
Robberies in venue tract in 2013 (se)	.413819	.435270	.459259			
Robberies in venue tract in 2013 (t-stat)	2.62269	.071782	.151840			
Interaction of female dummy with robberies in venue tract in 2013	-1.4765	-1.4975	-1.4739			
Interaction of female dummy with robberies in venue tract in 2013 (se)	.522899	.543990	.547839			
Interaction of female dummy with robberies in venue tract in 2013 (t-stat)	-2.8236	-2.7530	-2.6905			
Robberies per capita in venue tract in 2013				-.38539	-.82975	-.46656
Robberies per capita in venue tract in 2013 (se)				.524039	.662440	.592130
Robberies per capita in venue tract in 2013 (t-stat)				-.73543	-1.2525	-.78793
Interaction of female dummy with robberies per capita				.659049	1.08780	.996540
Interaction of female dummy with robberies per capita (se)				.581120	.718689	.646099
Interaction of female dummy with robberies per capita (t-stat)				1.13409	1.51359	1.54240
Log of travel time, Home		-.62194	-.82055		-.60572	-.81766
Log of travel time, Home (se)		.044496	.059363		.044744	.060268
Log of travel time, Home (t-stat)		-13.977	-13.821		-13.536	-13.567
Log of travel time, Work		-2.7335	-1.9934		-2.7685	-1.9735
Log of travel time, Work (se)		1.19260	.433699		1.27369	.427650
Log of travel time, Work (t-stat)		-2.2920	-4.5966		-2.1735	-4.6149
Log of travel time, Path		-.95295	-.88484		-.94508	-.87348
Log of travel time, Path (se)		.099848	.057914		.104500	.057599
Log of travel time, Path (t-stat)		-9.5439	-15.279		-9.0439	-15.164
Euclidean demographic distance between census_h and census_j		-.20424	-.46527		-.26300	-.51560
Euclidean demographic distance between census_h and census_j (se)		.120499	.125560		.119050	.124619
Euclidean demographic distance between census_h and census_j (t-stat)		-1.6950	-3.7056		-2.2091	-4.1374
Spectral segregation index of restaurant tract		-.12482	-.09839		-.10733	-.07740
Spectral segregation index of restaurant tract (se)		.044094	.057131		.044105	.057149
Spectral segregation index of restaurant tract (t-stat)		-2.8306	-1.7223		-2.4333	-1.3545
Interaction of SSI with Euclidean demographic distance		-.52236	-.29216		-.52223	-.27954
Interaction of SSI with Euclidean demographic distance (se)		.122120	.121919		.122759	.121610
Interaction of SSI with Euclidean demographic distance (t-stat)		-4.2774	-2.3963		-4.2540	-2.2987
Dummy for $$			.221469			.22755
Dummy for $$ (se)			.049139			.049081
Dummy for $$ (t-stat)			4.50689			4.63630
Dummy for $$$			.297890			.304340
Dummy for $$$ (se)			.061315			.061257
Dummy for $$$ (t-stat)			4.85830			4.96829
Dummy for $$$$			.400499			.401540
Dummy for $$$$ (se)			.095344			.095397
Dummy for $$$$ (t-stat)			4.20050			4.20909
Average rating on Yelp			.254170			.254990
Average rating on Yelp (se)			.038375			.038345
Average rating on Yelp (t-stat)			6.62340			6.64970
Number of points of origin	0	3	3	0	3	3
Number of choices per observation	20	20	20	20	20	20
Neighborhood dummies			Yes			Yes
Log-Likelihood	-2.9944	-2.7794	-2.7453	-2.9955	-2.7818	-2.7469
Pseudo R2	.000412	.072223	.083603	.000086	.071396	.083042
Akaike Information Criterion	9.98900	21.5590	59.4910	9.99090	21.5639	59.4939
Number of observations	3298	3298	3298	3298	3298	3298
Number of users	133	133	133	133	133	133
Number of venues	1922	1922	1922	1922	1922	1922

Table 10: Gender-differential results

Log of travel time, Home	-.67997	-.83856
Log of travel time, Home (t-stat)	(-14.555)	(-11.090)
Female × log of travel time, Home	.072596	.052326
Female × log of travel time, Home (t-stat)	(.879920)	(.501990)
Log of travel time, Work	-2.1531	-1.9062
Log of travel time, Work (t-stat)	(-3.1164)	(-3.8515)
Female × log of travel time, Work	-.83039	.062020
Female × log of travel time, Work (t-stat)	(-.27052)	(.087058)
Log of travel time, Path	-1.2662	-1.0960
Log of travel time, Path (t-stat)	(-6.6645)	(-8.1353)
Female × log of travel time, Path	.407359	.272960
Female × log of travel time, Path (t-stat)	(1.79649)	(1.79309)
Euclidean demographic distance between h and l_j	.009581	-.31371
Euclidean demographic distance between h and l_j (t-stat)	(.045607)	(-1.4689)
Spectral segregation index in l_j	-.19249	-.14403
Spectral segregation index in l_j (t-stat)	(-2.3259)	(-1.6189)
Interaction of SSI with Euclidean demographic distance	-.04601	.065692
Interaction of SSI with Euclidean demographic distance (t-stat)	(-.22213)	(.323729)
Female × SSI	.090121	.081789
Female × SSI (t-stat)	(.918370)	(.845640)
Female × Euclidean demographic distance	-.23362	-.22197
Female × Euclidean demographic distance (t-stat)	(-.89977)	(-.87453)
Female × SSI × Euclidean demographic distance	-.64828	-.61035
Female × SSI × Euclidean demographic distance (t-stat)	(-2.4923)	(-2.4193)
Dummy for $$	.225799	.201079
Dummy for $$ (t-stat)	(4.66270)	(4.08239)
Dummy for $$$	.293150	.312240
Dummy for $$$ (t-stat)	(4.85470)	(5.07870)
Dummy for $$$$	.292389	.364419
Dummy for $$$$ (t-stat)	(3.08540)	(3.80419)
Average rating on Yelp	.302810	.246920
Average rating on Yelp (t-stat)	(8)	(6.40810)
Robberies in venue tract in 2013	-.19471	-.10177
Robberies in venue tract in 2013 (t-stat)	(-.41545)	(-.20881)
Female × robberies in venue tract in 2013	-.99150	-.94296
Female × robberies in venue tract in 2013 (t-stat)	(-1.6871)	(-1.5970)
Percentage difference of median incomes ($h - l_j$)	-.30520	-.20232
Percentage difference of median incomes ($h - l_j$) (t-stat)	(-3.4691)	(-2.3025)
Female × percentage difference of median incomes ($h - l_j$)	.235389	.226579
Female × percentage difference of median incomes ($h - l_j$) (t-stat)	(2.32800)	(2.27880)
Number of points of origin	3	3
Number of choices per observation	20	20
Neighborhood dummies		Yes
Log-Likelihood	-2.7588	-2.7400
Pseudo R^2	.079076	.085345
Akaike Information Criterion	45.5180	75.4800
Number of observations	3298	3298
Number of users	133	133
Number of venues	1922	1922
Logit type	multinomial	multinomial
Column build date	31 Jul 2014	31 Jul 2014

4.5 Our results in light of the prior literature

Our results touch on a series of issues in the prior literature. The first concerns the role of spatial frictions in the city. In the urban literature, this is founded in the classic Alonso-Muth-Mills formulation, which links spatial frictions to land prices, particularly in regard to the home-work commute. More recently these spatial frictions have been considered in light of the role of highways in promoting suburbanization, as in Baum-Snow (2007), or in raising speed, as in Couture, Duranton, and Turner (2012). Finally, our work is related to Houde (2012), who considers the role of consumption choice for a homogeneous good (gasoline) along a path between home and work, and Couture (2014), who considers willingness to travel given a composite of opportunities in traveling from home. In our setting, we consider the consumption of highly heterogeneous goods, restaurant meals, in a setting in which we know the venues visited as well as characteristics of venues not chosen. In this setting, we find the expected negative coefficients of visits on travel time. But we also find that trips originating from a workplace are more sensitive to travel time costs.

There is an enormous literature on residential segregation by race and ethnicity. And we know that the implied physical distances between ethnic and social groups can have important economic consequences, as in the spatial-mismatch hypothesis (Kain, 1992). But we know little empirically about how porous these barriers are once we have controlled for distance, less about how this changes once we control for the type of places one wants to visit, and virtually nothing about how race and ethnicity, crime, and gender interact when we properly control for distance and the choices of venues. We establish that, even accounting for travel time between locations, there is a clear aversion to entering neighborhoods that differ ethnically or racially. This aversion to demographically different locations increases with the degree of demographic isolation, as measured by the spectral segregation index. This, perhaps, is not surprising. What may be more surprising is that this effect is entirely driven by the aversion of women, not of men, to enter isolated neighborhoods with different ethnic and racial demographics. Crime, indexed here by the number of robberies in a census tract, is a clear deterrent to visits. This is so, even accounting for racial and ethnic differences and isolation of neighborhoods. Again, this is perhaps not surprising. What is, once again, surprising is the role that gender plays in these results. Women essentially account for all of the differentially low visits to high-crime neighborhoods. These results are underscored in our results that consider our demographic, crime, and gender variables simultaneously. Racial and ethnic differences across neighborhoods matter for everyone. But women respond significantly and differentially to crime and to isolation in neighborhoods of a different race and ethnicity, strongly reducing their visits.

We can summarize our results quite simply. Spatial frictions matter, as we would expect. Racial and ethnic differences across neighborhoods matter. But the most surprising finding is that the role of crime and isolation of neighborhoods in discouraging visits is overwhelmingly due to differential responses by women. This underscores that the crime reduction of the last twenty years, in which index crime rates, including robbery, declined by approximately 85 percent in New York City, has been a boon for its citizens, but it seems to have particularly benefited women, who will feel freer to move through a safer city.

One of the great benefits of having estimates of both spatial and social frictions is that we can implement counterfactuals to explore the costs and gains of various social policies,

including extensions of transit that make the city more accessible and reductions in crime, which make citizens, especially women, more able to benefit from the many different neighborhoods in the city.

5 Counterfactuals

[Incomplete]

6 Conclusions

This is work in progress. Our preliminary and incomplete results show that both spatial and social frictions influence the geography of consumption. Our use of a novel data source and identification of users' home and work locations allowed us to characterize how consumption in the city depends on travel times, demographic differences, crime rates, and user characteristics. As expected, travel time to venues discourages visits. An important contribution of the present work is to establish the existence and magnitudes of the social frictions that also limit our use of the city. Crime and demographic differences, as well as the particular concerns of women in these regards, are all factors that limit the benefits of the city.

In future work, we will explore these in more detail. We will develop a quantitative framework based on our estimates that will allow us to answer questions such as how a return to the peak crime levels of the early 1990s would reduce the benefits of the city or to evaluate particular infrastructure projects from the perspective of non-commuting trips this would allow us to make. Spatial and social frictions divide the city, but our work can thus also make clear the potential benefits of various efforts to reduce such frictions.

References

Antecol, H., and D. A. Cobb-Clark (2008): "Racial and ethnic discrimination in local consumer markets: Exploiting the army's procedures for matching personnel to duty locations," *Journal of Urban Economics*, 64(2), 496–509.

Ayres, I. (2001): *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination*. University of Chicago Press.

Baum-Snow, N. (2007): "Did highways cause suburbanization?," *The Quarterly Journal of Economics*, pp. 775–805.

Bayer, P., and R. McMillan (2008): "Distinguishing Racial Preferences in the Housing Market: Theory and Evidence," in *Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation*, ed. by A. Baranzini, J. Ramirez, C. Schaerer, and P. Thalmann, pp. 225–244. Springer.

Black, M. C., K. C. Basile, M. J. Breiding, S. G. S. M. L. Walters, M. T. M. J. Chen, and M. R. Stevens (2011): *The National Intimate Partner and Sexual Violence Survey: 2010 Summary Report*. Centers for Disease Control and Prevention.

Bowes, D. R., and K. R. Ihlanfeldt (2001): "Identifying the impacts of rail transit stations on residential property values," *Journal of Urban Economics*, 50(1), 1–25.

Couture, V. (2014): "Valuing the Consumption Benefits of Urban Density," .

Couture, V., G. Duranton, and M. A. Turner (2012): "Speed," .

Cullen, J. B., and S. D. Levitt (1999): "Crime, Urban Flight, And The Consequences For Cities," *The Review of Economics and Statistics*, 81(2), 159–169.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107(3), 455–506.

Doran, B., and M. Burgess (2011): *Putting Fear of Crime on the Map: Investigating Perceptions of Crime Using Geographic Information Systems*, Springer Series on Evidence-Based Crime Policy. Springer.

Echenique, F., and R. G. Fryer (2007): "A Measure of Segregation Based on Social Interactions," *The Quarterly Journal of Economics*, 122(2), 441–485.

Ferraro, K. F. (1996): "Women's Fear of Victimization: Shadow of Sexual Assault?," *Social Forces*, 75(2), 667–690.

Gibbons, S. (2004): "The Costs of Urban Property Crime," *Economic Journal*, 114(499), F441–F463.

Glaeser, E. L., J. Kolko, and A. Saiz (2001): "Consumer city," *Journal of Economic Geography*, 1(1), 27–50.

Guiso, L., P. Sapienza, and L. Zingales (2009): "Cultural Biases in Economic Exchange?," *The Quarterly Journal of Economics*, 124(3), 1095–1131.

Handbury, J. (2012): "Are Poor Cities Cheap for Everyone? Non-Homotheticity and the Cost of Living Across U.S. Cities," Columbia University working paper.

Handbury, J., and D. E. Weinstein (2011): "Is New Economic Geography Right? Evidence from Price Data," NBER Working Papers 17067, National Bureau of Economic Research, Inc.

Houde, J.-F. (2012): "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, 102(5), 2147–82.

Ihlanfeldt, K., and T. Mayock (2010): "Panel data estimates of the effects of different types of crime on housing prices," *Regional Science and Urban Economics*, 40(2), 161–172.

Kain, J. F. (1992): "The spatial mismatch hypothesis: three decades later," *Housing policy debate*, 3(2), 371–460.

Lee, J. (2000): "The Salience of Race in Everyday Life: Black Customers' Shopping Experiences in Black and White Neighborhoods," *Work and Occupations*, 27, 353–376.

Linden, L., and J. E. Rockoff (2008): "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," *American Economic Review*, 98(3), 1103–27.

Logan, J. R. (2011): "Separate and unequal: The neighborhood gap for blacks, Hispanics and Asians in Metropolitan America," *Project US2010 Report*, pp. 1–22.

Lorenc, T., S. Clayton, D. Neary, M. Whitehead, M. Petticrew, H. Thomson, S. Cummins, A. Sowden, and A. Renton (2012): "Crime, fear of crime, environment, and mental health and wellbeing: Mapping review of theories and causal pathways," *Health & place*, 18(4), 757–765.

Luca, M. (2011): "Reviews, Reputation, and Revenue: The Case of Yelp.com," Harvard Business School Working Paper No. 12-016.

Mayer, A., and S. L. Puller (2008): "The old boy (and girl) network: Social network formation on university campuses," *Journal of Public Economics*, 92(1-2), 329–347.

McFadden, D. (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, pp. 75–96. North Holland, Amsterdam.

O'Flaherty, B., and R. Sethi (2007): "Crime and segregation," *Journal of Economic Behavior & Organization*, 64(3-4), 391–405.

——— (2010): "The racial geography of street vice," *Journal of Urban Economics*, 67(3), 270–286.

Pain, R. (1991): "Space, sexual violence and social control: Integrating geographical and feminist analyses of women's fear of crime," *Progress in Human Geography*, 15(4), 415–431.

——— (1997): "Social Geographies of Women's Fear of Crime," *Transactions of the Institute of British Geographers*, 22(2), pp. 231–244.

Pope, J. C. (2008): "Fear of crime and housing prices: Household reactions to sex offender registries," *Journal of Urban Economics*, 64(3), 601–614.

Quillian, L., and D. Pager (2001): "Black Neighbors, Higher Crime? The Role of Racial Stereotypes in Evaluations of Neighborhood Crime," *American Journal of Sociology*, 107(3), pp. 717–767.

——— (2010): "Estimating Risk: Stereotype Amplification and the Perceived Risk of Criminal Victimization," *Social Psychology Quarterly*, 73(1), 79–104.

Raanan, M. G., and N. Shoval (2014): "Mental maps compared to actual spatial behavior using {GPS} data: A new method for investigating segregation in cities," *Cities*, 36(0), 28 – 40.

Rauch, J. E. (2001): "Business and Social Networks in International Trade," *Journal of Economic Literature*, 39(4), 1177–1203.

Schreer, G. E., S. Smith, and K. Thomas (2009): ""Shopping While Black": Examining Racial Discrimination in a Retail Setting," *Journal of Applied Social Psychology*, 39(6), 1432–1444.

Train, K. E., D. L. McFadden, and M. Ben-Akiva (1987): "The Demend for Lcoal Telephant Service: a Fully Discrete Model of Residential Calling Patterns and Service Choices," *The Rand Journal of Economics*, 18(1), 109–123.

A Data

A.1 Yelp users data

To identify residential and workplace locations, research assistants examined the text of reviews written by Yelp users (in the 25% sample) of venues in New York and New Jersey that contained at least one of 26 key phrases. Those phrases are: I work, coworker, colleague, lunch break, my office, my work, I live, my apt, my apartment, my building, my neighborhood, my house, my local, my block, my place, my job, my hood, delivery, laundr, block away, block from m, blocks from m, close to me, close to my, minutes from m, street from m.

We identified individual who changed their residential and workplace locations via two means. First, research assistants reported any moves identified in the text of reviews containing the 26 key phrases above. Second, we reviewed the text of reviews containing at least of four key phrases: we moved, I moved, moving into, moving here.

A.2 Yelp venue data

Yelp assigns venues to neighborhoods that it defines. Venues may be assigned to more than one neighborhood, so these are not a partition of a city's geography. We aggregated Yelp-defined neighborhoods into 17 neighborhoods that partition the five boroughs of New York City: Washington Heights, Harlem, Upper West Side, Upper East Side, Midtown West, Midtown East, Union Square, West Village, Soho/Noho, East Village, Chinatown, South Village, Financial District, Brooklyn, Queen, Bronx, Staten Island.