Will You Use Ride Sharing Application?

A lesson from the 2017 National Household Travel Survey

Dongxu He[1]; Junfeng Jiao, Ph.D[2].

1. Introduction

With the development of mobile Internet and rapidly increasing use of smartphone, recent ten years have witnessed the skyrocketing usage of shared transport applications, and we have entered a so-called mobile Internet era since 2010, the birth year of iPhone 4.

In terms of smartphone usage, it is known that all shared transport applications could be and only be realized in the field of mobile Internet, or in the form of smartphone applications, rather than conventional cell phones, and only a few in personal computer. As mentioned in user guide[3], "the 2017 NHTS collected data on respondents' opinions and travel experiences" about the topic of technology "such as internet use through personal electronic devices (PED's), smartphone app use, and internet purchases". HOUSEHOLD data file shows that about three quarters of US households use smartphone every day, which gives shared transport applications to play their roles, as well as provides large potential demands.

It is exciting that 2017 NHTS records data about usage of three shared transport applications – bike sharing, ride sharing and car sharing, as well as household Internet use and delivery service to household, because these data reflects that 2017 NHTS embraces with the era of mobile Internet and keeps up with the frontier of the era. It is meaningful and interesting to analyze share transport application usage on the basis of NHTS.

The 2017 NHTS contains four data files: HOUSEHOLD, VEHICLE, PERSON and TRIP, sample size in each data file is 129696, 256115, 264234 and 923572, respectively. Data about shared transport application usage are presented in PERSON data file for the first time, and they record the number of times that each person uses each shared transport application in a 30-day period. It is worth noting that NHTS not just collects surveyed data, but provides weighted estimates of the totals[4]. This report uses weighted estimation to analyze shared transport application usage and impact factors of ride sharing usage, and uses raw sample surveyed data to conduct logistic regression analysis.

2. An overview of shared transport application

On the whole, Figure 1 shows that ride sharing obtains the largest usage, over 110 million, i.e., there is over 110M ride sharing application usage in a 30-day period throughout US; bike sharing the next, about 23M and car sharing the least, only 6.8M. In the supply side, shared transport application usage is greatly related to their convenience, accessibility as well as program launch time, and household car ownership will also greatly affect it. Specifically, ride sharing is

---

[1] Research assistant, Urban Information Lab, School of Architecture, University of Texas at Austin, 310 Inner Campus Dr., Stop B7500, Austin, TX 78712-1009. E-mail: henry_hdx@utexas.edu

[2] Corresponding author, Assistant Professor, Urban Information Lab, School of Architecture, University of Texas at Austin, 310 Inner Campus Dr., Stop B7500, Austin, TX 78712-1009. E-mail: jjiao@austin.utexas.edu

[3] U.S. Department of Transportation, Federal Highway Administration, 2017 NHTS Data User Guide. URL: https://nhts.ornl.gov/assets/2017UsersGuide.pdf

[4] As mentioned in user guide, "the weights reflect the selection probabilities and adjustments to account for eligibility, nonresponse and undercoverage". The NHTS users "can obtain estimates of the totals by multiplying each data value by the appropriate weight and summing the results".

considered the most convenient and accessible application among three (you just need to click in the app on your smartphone, ride sharing driver will pick you up in your location) and it is launched the earliest; meanwhile, high car ownership also provides many ride sharing drivers, so its largest usage is obvious. Despite the latest launch, with the aim to solving the so-called issue of "the first and last mile" in human mobility, bike sharing meets user's requirement of traveling in a short and middle distance to the utmost extent by providing a large number of bikes and guaranteeing their easiness to use. Oppositely, high car ownership in the US[5], difficulty in driving and small number of shared cars, result in a low car sharing usage.

In the demand side, endogenous travel demand and certain trip purposes would be related to shared transport application use. This report emphasizes on two key factors - **household income** and **age** that would have great impacts on endogenous travel demand. The question whether or not you use shared transport application is essentially an economic-related issue, and some other factors such as education level and worker status, are highly related to household income, so it is intuitive to choose household income as one of the key factors. On the other hand, shared transport application is a production of the mobile Internet era, so people in different ages will hold different attitudes towards them, and young people may be more interested in these new creations; meanwhile, health condition related to age may affect shared transport application usage.

Complete age distributions for usage of three shared transport applications, which is shown in Figure 2, provides details about the relationship between shared transport application usage and age. Note that ride sharing and car sharing require a user to be at least 16 years old, so their distribution starts from 16; since persons less than 5 years old are not included into PERSON data file, bike sharing distribution starts from 5. Findings based on the age distribution are shown as follow:

1. Car sharing has a simplest age distribution of usage, with only several significant usage peaks among people in age 25, 28 and 32. Possible reasons for the higher usage in these ages lie in that they don't have their own vehicles but need to travel frequent long distances in a short time, and car sharing costs less than ride sharing and is much more convenient than public transit.
2. Local maximum for bike sharing usage are located in age 32, 15, 37, 8 and 29 according to age distribution, that is, there are two peaks in age between 25 and 40, and age "<=15". Age "<=15" has the second largest bike sharing usage, because biking is a common travel conveyance and exercise for children no greater than 15. Large usage among people in age 15 results probably from their extremely growing travel demands, but they are prohibited to driving car independently, so bike is the best conveyance for them; however, there is an outstanding decrease in bike sharing usage for users with age from 16 to 20, which is likely because they have other travel options like driving cars and they are also curious and eager to drive, or biking is not convenient for them to travel.
3. The low usage of shared transport applications among the elderly (age >= 65) is mainly related to 1) their high car ownership due to enough wealth accumulation; 2) their small travel demands, for most of them are retired and they don't need to commute; 3) their incapability of moving due to poor health condition; 4) they hardly use Internet and smartphone, and neither know how to use shared transport applications.
4. For ride sharing, the distribution is evidently structured: the highest usage level focuses on age between 23 and 33, the second level lies in age between 43 and 60, and the lowest for the elderly. More details will show in the next section.

[5] 1.883 per household according to HOUSEHOLD data file

Figure 3 shows the relationship between household income level and share transport application usage. Noting that income data is extremely sensitive, it is a common desensitized way for household income data to be recorded in 11 levels using integer from 1 to 11. The household income spectrum is divided into 11 intervals by following 10 critical income points: $10000, $15000, $25000, $35000, $50000, $75000, $100000, $125000, $150000 and $200000. Findings based on Figure 3 are shown as follow:

1. There is no obvious relationship between household income and car sharing usage, for people in all income levels have similar car sharing usage except income between $15000 and $24999. It is likely because people with low household income don't have their own cars, but they should earn enough to try car sharing applications, then it is reasonable for people with income between $15000 and $24999 to have highest car sharing usage.
2. Higher bike sharing usage may be related to middle income level (household income between $35000 and $125000), while people with less bike sharing usage are more likely to have household income in two extremes, either in highest income level, or in lowest income level.
3. Interestingly, people with high ride sharing usage are more likely to obtain high income, so we can say that for ride sharing users, ride sharing usage is related to their household income level. To dig deeper into the relationship between income and age and high ride sharing usage, we choose two income levels ("$50000 - $74999" and ">= $200000") with first and second highest ride sharing usage to draw age distributions, and find two totally different age distributions: 1) The age distribution for people in level "$50000 - $74999" is more extreme, for the ages of people with highest ride sharing usage are focused on age between 23 and 40, especially for age group "23 -33", while the middle and old aged with middle income levels are less likely to use ride sharing. 2) For people with highest household income, both the young aged (age 32) and the middle aged (age 50) have highest ride sharing usage; while for people in age 19, their highest ride sharing usage is likely to be related to their parents' high income. To conclude, the young people with middle and high household income are becoming the main source to use ride sharing.

3. Will you use ride sharing application?

This section will emphasize on ride sharing. To dig deeper into the features of persons who may use ride sharing, we first identify some main factors, and then conduct both exploratory data analysis (EDA) and logistic regression.

3.1 Factors associated with ride sharing usage

First of all, we check all variables in NHTS, especially for HOUSEHOLD and PERSON data files, then we list all possible factors and categorizes them into four: personal characteristics, household characteristics, spatial characteristics and information and communication technologies (ICTs).

Table 1 shows the detail about selected factors that would have impacts on people's ride sharing usage, these factors will also be used as independent variables in the following logistic regression analysis. Specifically, there are 8 factors in personal characteristics category, including age, sex, race, education level, worker status, driver status, health condition and distance to work (since there is only one travel characteristics, it is listed in personal characteristics). Household characteristics consist of household income, household available vehicles per driver and number of adults in household. Spatial characteristics involve population density (measured by 1000 persons per square mile in census tract level) and urban area. ICTs include frequency of Internet use and smartphone use in household, number of times purchased online for delivery in last 30 days, as well as work from home.

EDA is mainly focus on the relationship between ride sharing usage and personal characteristics. Figure 4 shows age distribution for ride sharing usage by sex, work status, education level and health condition, and Table 2 shows the relationship between ride sharing usage and driver status. Findings are shown as follow:

1. For ride sharing users, male use ride sharing application a little more than female in a whole picture, while female use ride sharing more than male for age between 16 and 30, and male use more than female for age between 31 and 45, and in some age (age 36, 37, 38), male usage of ride sharing is nearly twice of female usage.
2. For ride sharing users, workers use ride sharing application much more than non-workers, so we can say that people who use ride sharing application is probably a worker. While for people in age less than 20 or more than 65, non-workers use more than workers.
3. Ride sharing users are more likely to obtain college or higher degrees, especially for people with age no younger than 22. Moreover, higher education level and worker status are related to higher household income, which is consistent with the finding that people with high ride sharing usage are more likely to obtain high income.
4. Most of ride sharing users think that they are in a good or better health condition. But there is survivorship bias in this finding, because those who are in poor health condition are very unlikely to use smartphone and then to use ride sharing application, or to travel alone by shared ride.
5. Interestingly, ride sharing users are probably drivers. In fact, there are only 5.5% of non-drivers use ride sharing according to survey data shown in Table 2, less than 7.5% for drivers. Possible reason is that non-drivers have less endogenous travel demand, especially for the middle and old aged non-drivers. This finding may conflict our intuitive perceptions, because it is natural to think that a non-driver is more likely to use ride sharing application. However, the truth is that the confliction can't stand because this finding and the intuitive perception are in different statistical inference - this finding is based on the sample of all ride sharing users, while intuitive perception stands in a circumstance where the ride sharing use status of the sample is unknown.

3.2 Logistic regression

Based on the analysis above, the relationship between age and three shared transport applications, the relationship between household income and three shared transport applications, as well as the relationship between ride sharing usage and personal characteristics, logistic regression will be conducted to explore the factors that would have impacts on ride sharing usage in three different age groups: all ages, age between 23 and 33, and age no less than 65 (">=65").

Data source: PERSON file provides personal sociodemographic characteristics, spatial characteristics, usage of ICTs and personal health condition. HOUSEHOLD file provides characteristics of surveyed households, such as household sociodemographic characteristics, as well as household Internet use.

Data cleaning: Since ride sharing requires a user to be at least 16 years old, individuals under 16 are excluded; person without ride sharing and age data are also excluded. Finally, 235716 valid records are selected to conduct logistic regression for all ages. Logistic regression for age between 23 and 33 contains 28773

Variable recoding: There are four variables that need to be recoded, *HHINC, HEALTH, INTERNET* and *SMARTPHONE.*

HHINC is recoded by the mean of two endpoints of each interval, measured by thousands, for example, *HHINC*=3 represents household income between $15000 and $25000, the mean is $20000, so it is recoded to *HHINC*=20; while *HHINC*=11 is recoded as *HHINC*=300.

HEALTH, SMARTPHONE and *INTERNET* are recoded in opposite measurement, for example, *INTERNET*=1 represents frequency of daily internet use, it is recoded to *INTERNET*=5, which is consistent with quantity perceptions.

Regression results: Table 3 shows the results for three logistic models, the main results are shown as follow: except *WHITE* in model for age “>=65”, other variables share same relationship direction in all three models, where positive coefficient means higher probabilities to use ride sharing, and negative coefficient means less probabilities to use ride sharing. Among all variables, *AGE, FEMALE, DRIVER, VEHAVAL, NUMADULT, INTERNET* have negative coefficients. It is intuitive that for a person with his or her ride sharing use status unknown, older age, female, driver status, more available vehicles will decrease this person’s probability of using ride sharing application, which is consistent with analysis above. For *INTERNET*, regression result may conflict intuitive perception, but possible reason is that people who daily surf the Internet are more likely to stay at home, then their endogenous travel demand will decrease, and the probability of using ride sharing application will correspondingly decrease. *WHITE* in model for age “>=65” has significant negative impacts, which means white people older than 65 are less likely to use ride sharing.

Other 11 variables except *WHITE* for age “>=65” have positive coefficients, which means that for a person with his or her ride sharing use status unknown and keeping other variables invariant, higher degree, worker status, longer work distance, better health condition, higher household income, living in area with higher population density or urban area, working from home, higher frequent use of smartphone and online shopping delivery service, each of these conditions will increase a person’s probability of using ride sharing application. For *WRK_HOME* with positive coefficient, it is perhaps because people working from home will do things rather work, such as going to bank, visiting doctor or dentist, so their non-commute travel demand may increase, and the probability of using ride sharing application will correspondingly increase.

Furthermore, *COLLEGE* has less positive effect for people in age “>=65”; *WORKER* has more positive impact for people in age between 23 and 33, which is consistent with the relationship between ride sharing usage and worker status; while *WRK_HOME* has less positive influence for people in age between 23 and 33, but higher positive influence for people in age “>=65”.

4. Conclusion

This report explores the factors that would have impacts people’s ride sharing use, which finds that older age, female, driver status, more available vehicles will decrease a person’s probability of using ride sharing application; and higher degree, worker status, longer work distance, better health condition, higher household income, living in area with higher population density or urban area, working from home, higher frequent use of smartphone and online shopping delivery service, will increase a person’s probability of using ride sharing application. To conclude, your decision to use ride sharing application will be comprehensively affected by these factors.

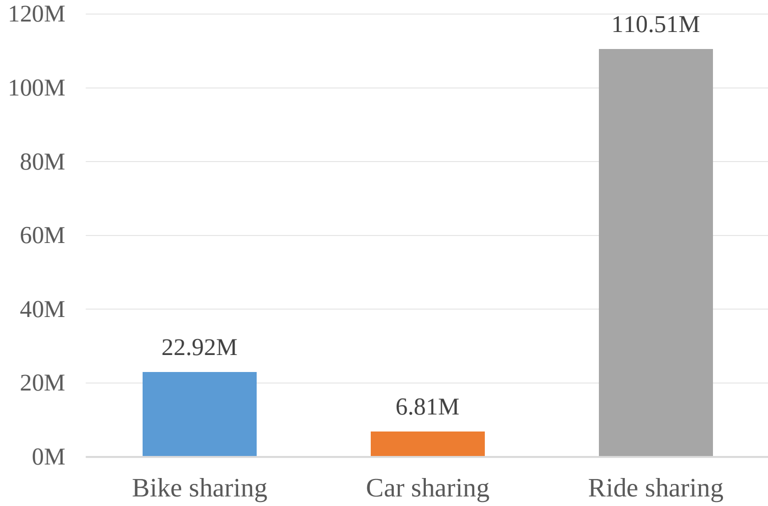

Figure 1: Total usage of three shared transport applications (weighted estimations in a 30-day period)

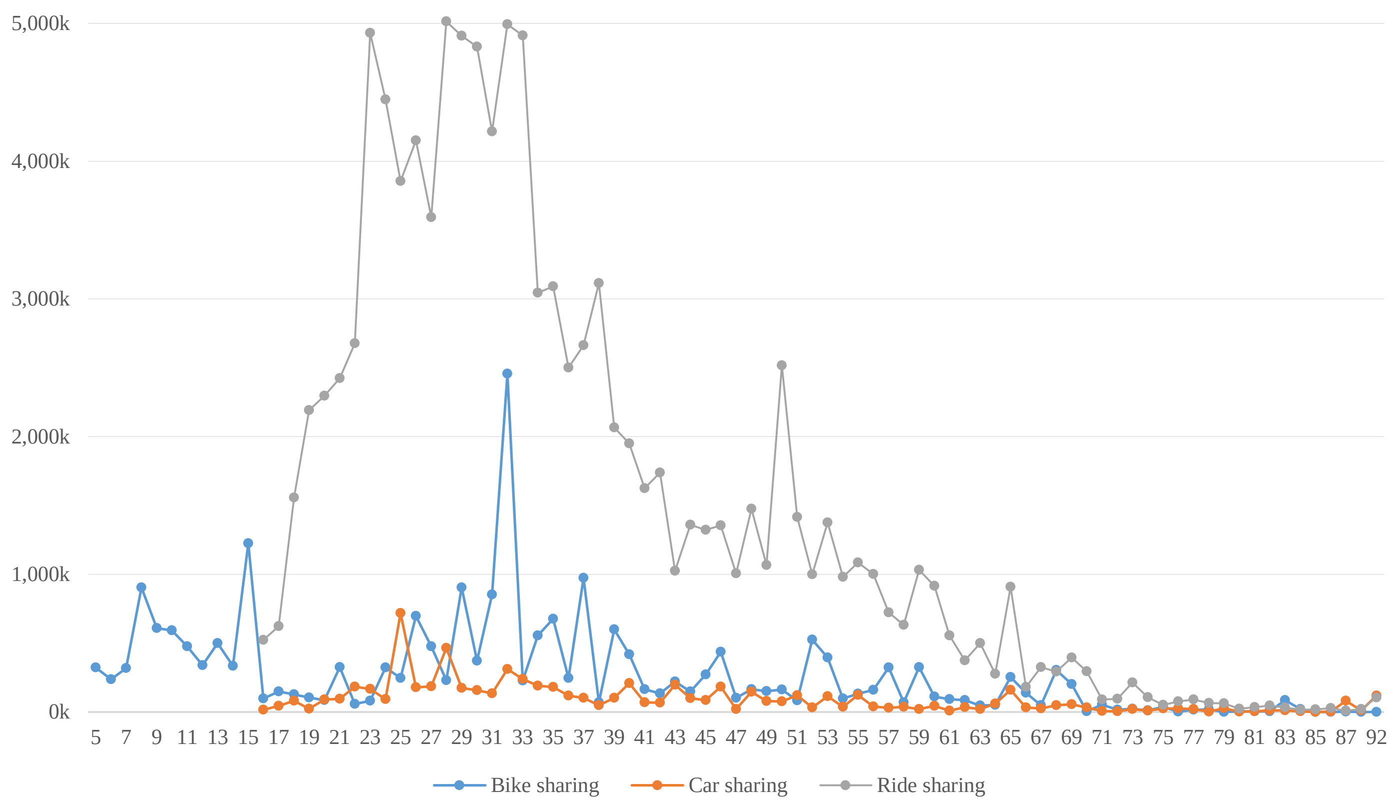

Figure 2: Age distributions for usage of three shared transport applications

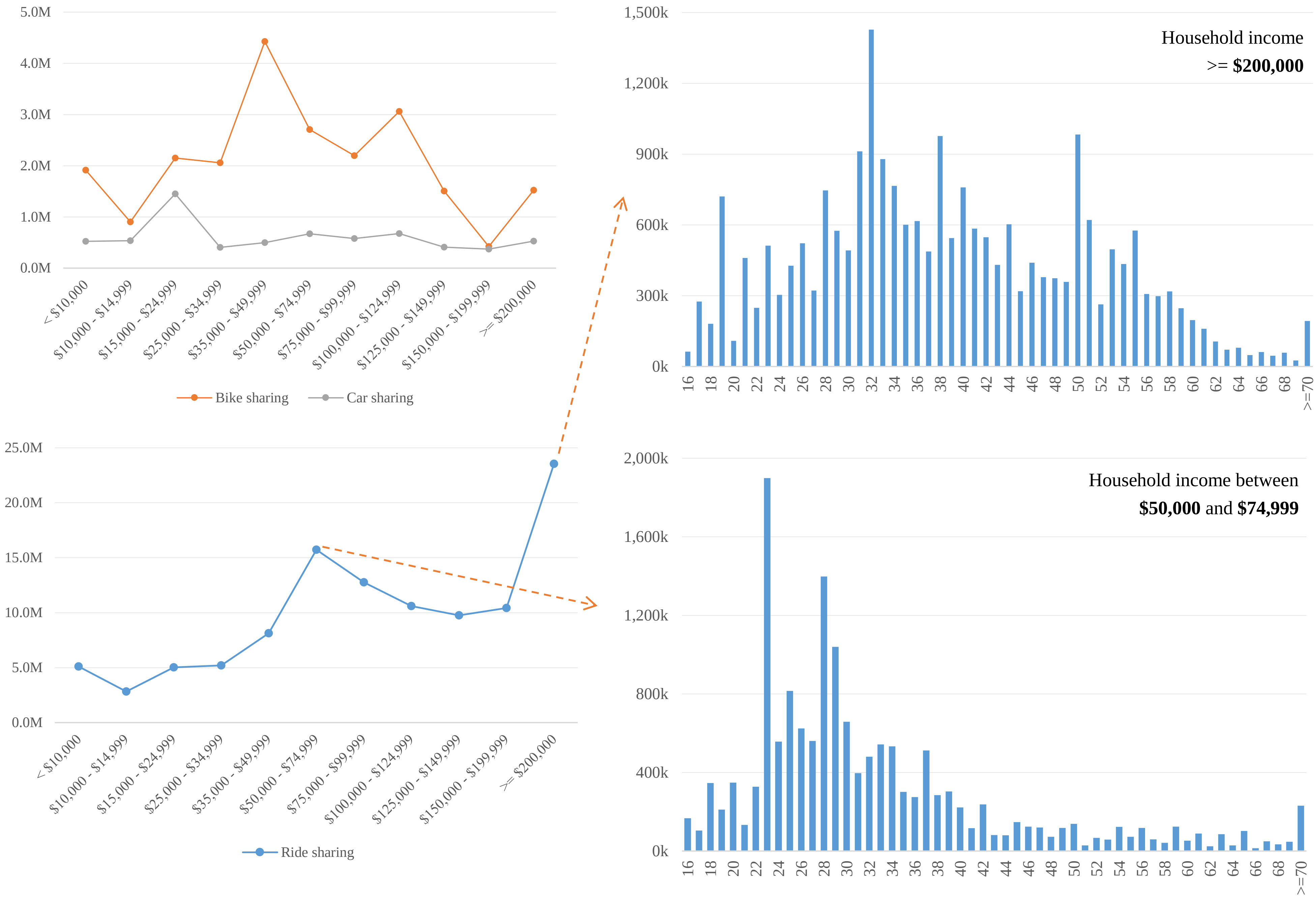

Figure 3: Household income and usage of three shared transport applications

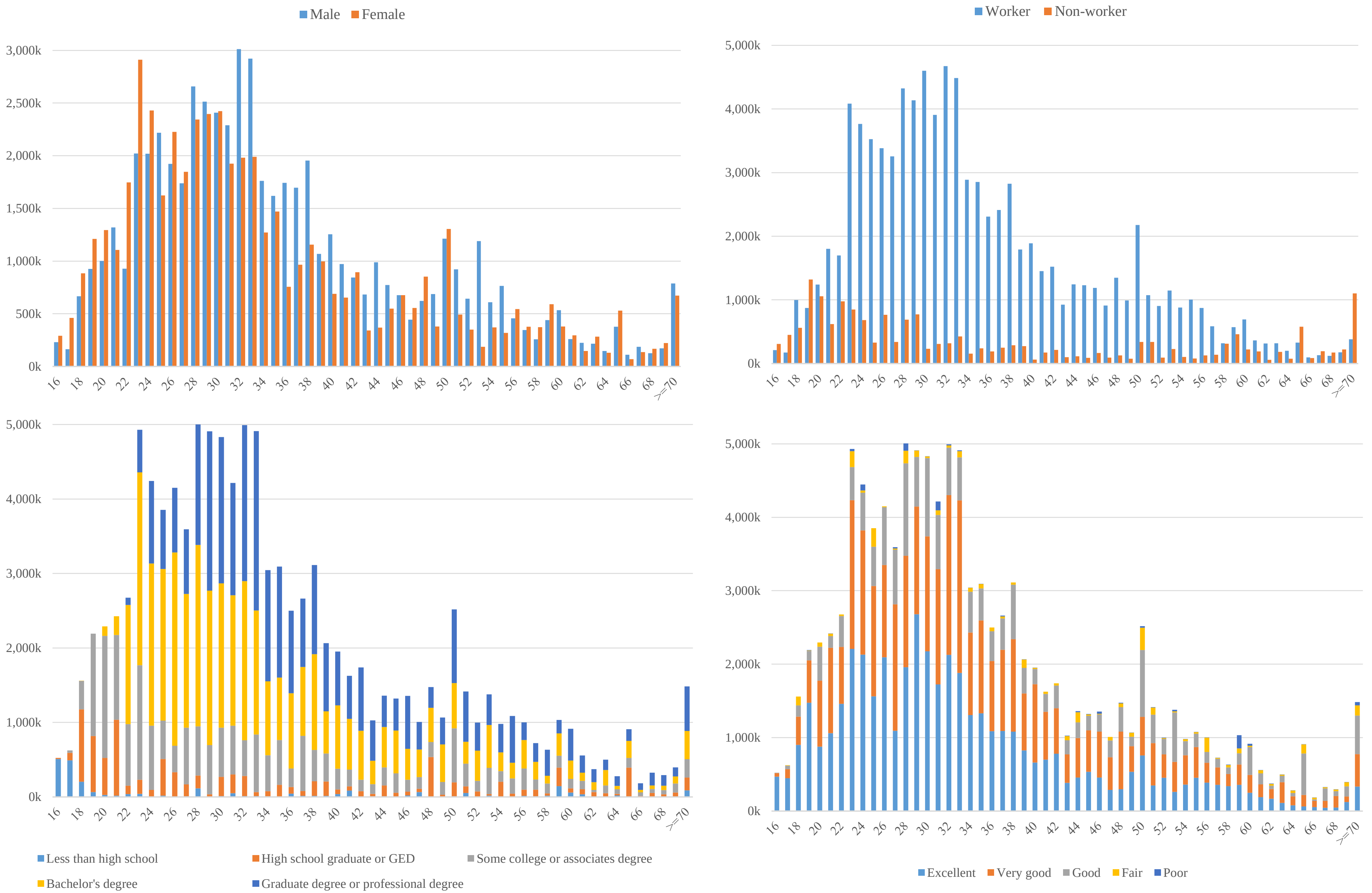

Figure 4: Age distribution for **ride sharing** usage by sex, worker status, education level and health condition

Table 1. Variable description

Category	Variable	Description	Variable type
Personal characteristics	*AGE*	Person's age	Numeric
	FEMALE	If a person is female, then 1, otherwise 0	Binary
	COLLEGE	If a person has college or higher degree, then 1, other 0	Dummy
	WHITE	If a person is white, then 1, otherwise 0	Dummy
	WORKER	If a person is a worker, then 1, otherwise 0	Binary
	DRIVER	If a person drives, then 1, otherwise 0	Binary
	WORKDIST	Road network distance in miles, between respondent's home location and work location	Numeric
	HEALTH	Personal opinion of health, 5 = excellent, 4 = very good, 3 = good, 2 = fair, 1 = poor	Ordinal
Household characteristics	*HHINC*	Level of household income	Numeric
	VEHAVAIL	Vehicles available per driver in household	Numeric
	NUMADLT	Number of adults in household	Numeric
Spatial characteristics	*POPDEN*	1000 persons per square mile in census tract level	Numeric
	URBAN	If a household is in urban area, then 1, otherwise 0	Binary
Information and communication technologies	*INTERNET*	Frequency of Internet use, 5 = daily, 4 = a few times a week, 3 = a few times a month, 2 = a few times a year, 1 = never	Ordinal
	SMARTPHONE	Frequency of smartphone use to access the Internet, same measurement with INTERNET	Ordinal
	DELIVER	Number of times purchased online for delivery in last 30 days	Numeric
	WRK_HOME	If a person can work from home, then 1, otherwise 0	Binary

Table 2. Driver status and ride sharing usage

	Use ride sharing?		
	Yes	No	
Driver	16287 7.49%	201165 92.51%	217452 100.00%
Non-driver	1189 5.51%	20400 94.49%	21589 100.00%
Total	17476	221565	239041

Table 3. Logistic regression results

Variables	All ages			Age between 23 and 33			Age at 65 or more		
	Coefficient	Wald χ^2	Significance	Coefficient	Wald χ^2	Significance	Coefficient	Wald χ^2	Significance
AGE	-0.038	3499.972	0.000	-0.073	158.495	0.000	-0.015	8.780	0.003
FEMALE	-0.213	146.999	0.000	-0.159	21.496	0.000	-0.224	15.678	0.000
COLLEGE	1.126	1160.414	0.000	1.124	219.426	0.000	0.673	45.008	0.000
WORKER	0.232	103.352	0.000	0.434	71.277	0.000	0.028	0.143	0.705
DRIVER	-0.371	84.596	0.000	-0.416	29.189	0.000	-0.633	30.450	0.000
WHITE	0.002	0.011	0.915	0.003	0.006	0.939	-0.287	11.988	0.001
WORKDIST	0.001	89.314	0.000	0.000	3.858	0.050	0.001	12.141	0.000
HEALTH	0.190	320.401	0.000	0.238	120.944	0.000	0.192	39.437	0.000
HHINC	0.007	2786.395	0.000	0.007	576.928	0.000	0.008	350.927	0.000
VEHAVAIL	-0.225	143.899	0.000	-0.338	60.556	0.000	-0.179	12.233	0.000
NUMADLT	-0.472	1299.675	0.000	-0.525	443.892	0.000	-0.640	144.751	0.000
POPDEN	0.079	3275.564	0.000	0.091	1087.021	0.000	0.062	216.439	0.000
URBAN	0.787	561.118	0.000	1.047	178.307	0.000	0.441	27.481	0.000
INTERNET	-0.206	49.964	0.000	-0.014	0.022	0.883	-0.163	13.020	0.000
SMARTPHONE	0.387	618.862	0.000	0.328	68.699	0.000	0.434	236.431	0.000
WRK_HOME	0.522	370.580	0.000	0.295	21.538	0.000	0.694	49.709	0.000
DELIVER	0.046	954.952	0.000	0.048	223.061	0.000	0.053	125.688	0.000
(Constant)	-3.671	680.987	0.000	-3.611	53.040	0.000	-4.160	72.297	0.000
N		235716			28773			73373	
χ^2		29668.575			5656.618			2316.309	