NEW YORK CITY DOT

2018 Citywide Mobility Survey
Data User Guide

Research Overview

The New York City Department of Transportation conducted a mixed methodology citywide survey in conjunction with PSB, an independent market research firm, over the course of eight weeks. The objectives of this research were to:

- Track results from last year's survey to investigate year-over-year trends regarding transportation preferences and usage patterns
- Understand the factors and experiences that drive transportation choices for New York City residents
- Assess views on the current state of transportation within the City
- Measure reactions and perceptions to relevant trends and topics in New York City transportation

A total of 3,301 New York City residents age 18 and over participated in the survey. The phone sample was sized at 2,500 respondents, recruited over the phone using RDD dialing based on area code and a purchased sample list. Each cell phone number was called back 3 times and each landline number 5 times to maximize response rates. Respondents had the option to take the survey in Spanish and Chinese, with call-backs in additional languages available upon request (e.g. Russian or French Creole). The online sample was sized at 801 respondents, recruited via sample lists vetted by the PSB team and identified by zip code.

Sample Comparison

	2017 (Wave 1)	2018 (Wave 2)
Sample size	N=3,600 total • N=1,800 phone • N=1,800 online	N=3,300 total • N=2,500 phone • N=800 online
Oversamples	Oversample in the 10 survey zones to reach ~n=350 per zone	Oversample in the 10 survey zones to reach ~n=300 per zone
Languages	Directly offered in English, Spanish, Chinese	Directly offered in English, Spanish, Chinese

The overall incidence rate for the survey, showing completion rate out of total survey starts, was 46%. The response rate for completes out of the total sample *contacted* was 1%. The response rate for completes out of the total sample that was *reached* was 8%. The distinction between *contacted* and *reached* was delineated on whether or not the potential respondent was both contacted and eligible for the survey. For example, if the dialed number was disconnected or there was no answer it was not calculated in to the *reached* response rate. However, if the dialed respondent refused to take part in the survey they were included in the *reached* response rate.

Response Rate - Contacted

Instrument	Completed	Contacted	Response Rate
Online	801	77,121	1%
Landline	1,500	115,706	1%
Cell	1,000	52,191	2%
Total	3,301	245,018	1%

Response Rate - Reached

Instrument	Completes	Reached	Response Rate
Online	801	9,707	8%
Landline	1,500	18,226	8%
Cell	1,000	14,464	7%
Total	3,301	42,397	8%

The 2018 survey fielded from May 3, 2018 through June 26, 2018. The survey is conducted in the spring/early summer of each year. The timing was chosen based on the following conditions:

- The weather enables the largest cross-section of modal usage
- There are fewer holidays in May/June than in other potential parts of the year
- Public schools are in session until the end of June (the last day of public schools coincides with the end of our field period)

During this period, there were two days in which temperatures rose above 90 degrees, and precipitation never exceeded one inch. Notable holidays and events that occur during this period include Mother's Day, Father's Day, Memorial Day, the Israel Day Parade, the Puerto Rican Day Parade, the Mermaid Parade, Eid al-Fitr, and the Pride Parade.

Data Aggregation and Data Package

The survey is divided into two sections: the main survey and the trip diary. These two sections yielded separate data sets, which were analyzed independently. The main survey tracks respondent level data, assessing behaviors, attitudes, and perceptions of transportation throughout New York City. The trip diary looks at the data by aggregating trips as opposed to respondents, recording each trip that respondents had taken the previous day. A trip is defined as any one-way journey that starts in one location and ended in another. In 2018, we further refined the definition to exclude deliveries made for work. We included additional questions to capture the number of deliveries made by mode.

The main survey data set is aggregated by respondent, and percentages indicate the share of New York City residents with each characteristic. The trip diary data set is aggregated by trip, and percentages indicate the share of trips with each characteristic.

- **Any data points at the 'All NYC' or 'All NYC Trip' level, including relevant demographic sub-groups, refer only to phone data,** as it is representative of population distribution throughout New York City**.**
- **Data at the survey zone level (e.g. Manhattan Core, Inner Brooklyn, etc.) is combined, referring to both the online and phone data.** This combined data is in line with demographic trends within each survey zone based on publicly available census data.
    - *Data at the non-Staten Island borough level (e.g. Manhattan, Brooklyn, Queens, or the Bronx etc.) is not displayed in the report. However, it can be analyzed within the case level data at the 'All NYC' or 'All NYC Trip' level (phone only). Due to the larger sample size of non-Staten Island boroughs vs. survey zones, online data is not necessary to achieve a sufficient n-size for analysis. Please note: Since Staten Island is both a borough and its own survey zone, combined phone and online data should be referenced to ensure a reliable sample size.*

The data package contains the full codebook for each dataset as well as CSV files for both the main survey and the trip-tracking data. Unique identifiers act as a match key between the two datasets. The CSV files have been cleaned up to include all coding and fielding data per the specifications in the questionnaire. The weighting variable is included as "allwt" at the end of each dataset. In addition, Excel crosstabs with data broken by audiences and variables specified by the NYCDOT team are included, as well as the data (including anonymized GPS data) used to create the maps, together with the map packages for all files created in GIS. The data package also includes a code book which will help navigate the questions throughout the data sets.

Sample and Weighting

The sample size for the main survey data set is n=3,301 respondents, 2,500 of whom completed the survey by phone and 801 of whom completed the survey online. Of the phone sample, 60% of interviews were completed via landline and 40% via cell phone. Sampling occurred every day of the week to ensure an even distribution of completes across days (for the integrity of the trip-tracking module which asked about the day before the interview). While PSB had not placed daily quotas on the calling, vendors were instructed to release sample in a consistent manner, and weekly quotas were implemented on the online sample to ensure even fielding.

The sample size for the trip diary data set is n=7,977 trips: 6,063 of which were captured by phone and 1,914 of which were online.

The margin of error for the phone data set is ±1.96%, smaller than last year because of the increased phone sample size. Data cuts with a base size below 50 are directional and are considered unreliable. A base size above 250 is considered a solid, statistical read.

The 2018 phone sample was weighted to match the 5 year (2012-2016) American Community Survey average based on the following factors: age, gender, educational attainment, race, income, and borough. The 2017 sample was also weighted to ACS 5 year average, but was one year removed (2011-2015).

The online sample is an oversample of populations in certain neighborhoods that are difficult to reach by phone, and is in line with the demographics of each of those neighborhoods. The online sample was weighted to match the phone sample.

Differences to Wave 1 (2017)

For the second wave of research, the following changes were made:

- In 2017, probes were added into the survey script mid-way through fielding and respondents were re-contacted with those probes included. In 2018 additional probes were added, including automatic probes if they indicated they did not start or end their day at home. In addition, the day had been broken out into morning, afternoon, and evening for the second wave, and was asked for each of those discrete times. This tripled the number of probes asked if respondents were not recalling every trip from the prior day.
    - Due to additional probing included in survey instrument, Zero Trippers (18%) were down 8 points compared to 2017 (26%) and Single Trippers (11%) were down 6 points compared to 2017 (17%).
- Adjustments were made to how multi-punch, perceptual questions were asked to minimize the modal affect. Those taking the survey online were more likely to select more than one option than those on the phone. Thus, multi-punch questions (outside of the transportation mode-tree questions) were asked as yes/no metrics for each answer choice.
    - In addition, on the multi-punch, perceptual questions, significance was not reported on between the phone and online samples, as the absolute modal bias is not directly measurable with the methodology.
- Captured address of origin for each trip, in addition to the address of each trip destination.
- Captured how long it took to get to/from the transit boarding location.
- Removed the “return home” choice for trip purpose and instituted coding “other” responses based on origin and destination.
- Previously captured addresses in the survey (home, work, school) were automatically coded to help make the section less tedious on respondents.
- Respondents were asked to enter their addresses differently online versus over the phone. For online, a custom tool was developed that allowed respondents to drop a pin at their location (or the closest cross street), which automatically coded GPS location, NTA, survey zone, zip code, and borough. Over the phone, respondents were asked for the address, and if refused, asked for the nearest cross street and then the zip code. Coding for the NTA and survey zones were done by hand on the backend.

NYC DOT Citywide 2018 Survey – Re-Contact Methodology

Upon review by the client, three errors were identified that were rooted back to the questionnaire stage concerning logic or typos. The following steps were taken to rectify the situation:

Three Errors Identified

Change 1: In the Ride Hail section, online respondents were shown a question with "Gett" as a possible response. The correct response was "Chariot," which phone respondents were asked. According to the final questionnaire, Chariot should have been included in both phone and online responses. It was asked correctly over the phone but incorrectly online.

> In order to rectify this mismatch, PSB went back into field to re-contact the N=800 respondents from the affected portion (i.e. online portion) of the survey.

Change 2: In the Bicycle section, a series of questions (Q25f-j) were incorrectly filtered to be asked of only bike owners, but should have been asked of all respondents. This was correctly asked of all respondents online, but only asked of bicycle owners on the phone.

> In order to rectify this logic error, PSB went back into field to re-contact the N=2,500 respondents from the affected portion (i.e. phone portion) of the survey.

Change 3: In the State of Public Transportation section, among those who say they are using the bus less, Q28F asks them what mode of transportation they are using instead of the bus. C2 was mistakenly listed as "bus" instead of "subway".

> In order to rectify this mismatch, PSB went back into field to re-contact the N=307 respondents from the affected portions (both phone and online) of the survey.

Full Data Audit

PSB completely reviewed all aspects of the data, including making sure base sizes, filters, skip logic and weights are all correct; re-checked that all aspects of the final questionnaire match the implemented survey; reviewed case level data in the SPSS data file; and made sure that phone and online data were merged together correctly. PSB also generate updated codebooks that were more detailed and user friendly.

Re-Contact

The re-contact process took place from Thursday 8/30 through Monday 9/17, resulting in 19 days of fielding. Re-contact took place for both phone and online with revised questions (detailed above). Overall we were able to re-contact 797 respondents, reaching 562 over the phone and 235 online.

For the ride hail section, we attempted to reach as many of the 800 online survey respondents as possible to rebuild the data for the ride-hailing question (substituting "Chariot" for "Gett" as a response) with the expectation of completing ~200 interviews. We ended up with 210 completes.

For the bicycle section, we attempted to reach as many of the 1,597 phone survey respondents as possible to rebuild the data for the bicycle questions (fixing the logic so that all respondents were able to answer, not just bike owners) with the expectation of completing ~600 interviews. We ended up with 531 number of completes.

For the state of public transportation section, we attempted to reach as many of the 307 phone and online survey respondents as possible to rebuild the data for the bus usage question (substituting "Subway" for "Bus" as a response) with the expectation of completing ~100 interviews. We ended up with 88 number of completes.

Weighting

PSB's marketing sciences team performed a series of statistical analyses to investigate 1) whether the re-contact data are demographically skewed and 2) if the data are skewed, whether the skew was due to the demographic make-up of the re-contact respondents.

The team found marginal (non-significant) skewing in the re-contact data, but the strength of the relationship is very weak and most of the skew can be attributed to the marginal skewing we see in the overall sample due to under/over sampling certain demographic categories (e.g. a bit light on younger cases, a bit heavy on older cases). In other words, we are not concerned that the missing data will cause any significant distortions of the data, nor do we believe the missing data will require us to apply any extra weighting on either the re-contact cases themselves or the dataset at large.

We also had our marketing sciences team run some preliminary analyses to investigate whether the skewing was affecting any attitudinal variables, and they were able to confirm that the marginal skewing is confined to demographic variables, so none of the attitudinal or behavioral questions will be affected.