

How Airbnb's data hid the facts in New York City

Murray Cox
Founder, *Inside Airbnb*
insideairbnb.com
murray@murraycox.com

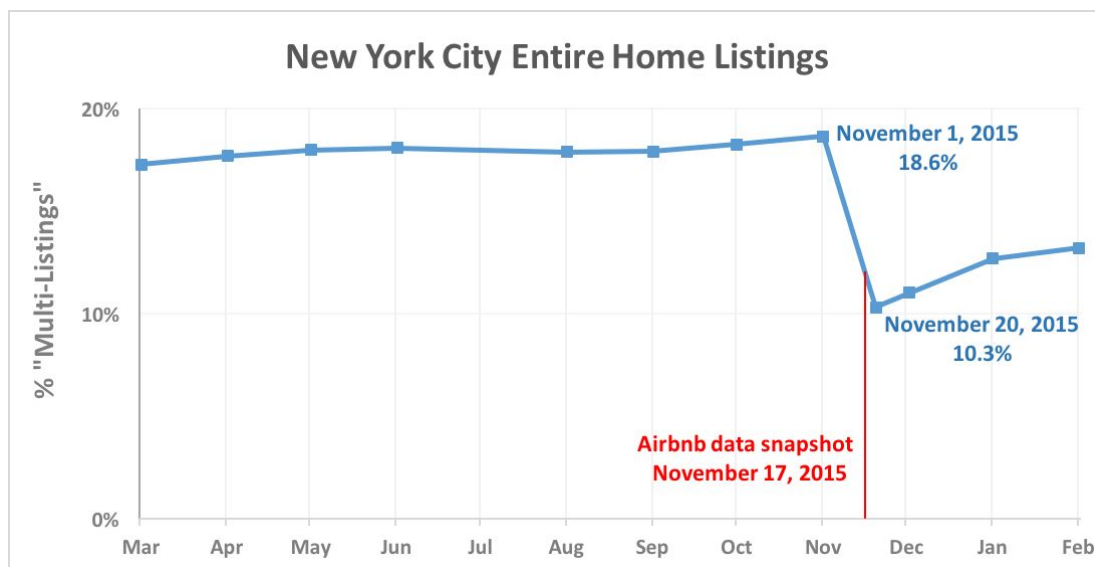
Tom Slee
Author, *What's Yours is Mine* ([OR Books](#))
tomslee.net
tom@tomslee.net

February 10, 2016

On December 1 2015, Airbnb made data available about its business in New York City, with much fanfare. This report shows that the Airbnb data release misled the media and the public.

Airbnb's data release was presented as "the first time Airbnb has voluntarily shared city data on a wide scale on how its hosts use the online platform".¹ This report shows that the data was photoshopped: Airbnb ensured it would paint a flattering picture by carrying out a one-time targeted purge of over 1,000 listings in the first three weeks of November. The company then presented November 17 as a typical day in the company's operations and mis-represented the one-time purge as a historical trend.

Over a thousand New York listings vanished from the Airbnb website just before the company took a snapshot it used for publicity.



Summary

As part of its recent transparency initiative, Airbnb made public a snapshot of its New York City business as of November 17, 2015.¹ The company used the snapshot to paint a picture of its business, claiming that “95% of our entire home hosts share only one listing”,² and forecasting that the revenue earned by hosts with three or more listings will fall from 25% of the total to 7% for 2015-2016.¹

This report shows that the snapshot was photoshopped. In the days leading up to November 17, Airbnb ensured a favorable picture by carrying out a one-time targeted purge of over 1,000 listings. The company then misrepresented November 17 as a typical day in the company’s operations.

The graph on the front page is a centerpiece of the report. It shows that any decrease in listings belonging to multiple-listing hosts can be attributed to a single event, immediately before November 17, when such listings plummeted from 19% to 10% of the total.

This report also shows that the November purge is not part of Airbnb’s regular enforcement activities: no similar intervention took place on the owners of multiple “Private room” listings in New York, and data from other major markets in North America and elsewhere show no remotely comparable intervention. The report shows that the purge was limited to the exact data set that Airbnb presented to the public, and on which it based the claims it made to major news outlets.

The intervention was so specific, and the timing so close to the date of the New York City snapshot, that the conclusion is inescapable: Airbnb removed listings from its site so that its data set would paint a more attractive picture of its business, to better influence media and public opinion. It was a cover-up, not a move to transparency.

In presenting the data, Airbnb hid the purge of unwanted listings by claiming that its forecasts were based on “historical trends”. This data manipulation by Airbnb, in pursuit of its public relations goals and as part of its resistance to regulation, shows that the

¹ Mike Isaac, *Airbnb releases trove of New York City home-sharing data*, New York Times, December 1, 2015
<http://www.nytimes.com/2015/12/02/technology/airbnb-releases-trove-of-new-york-city-home-sharing-data.html>

² Airbnb, *Data on the Airbnb Community in NYC*, Dec 1, 2015
<https://www.airbnbaction.com/data-on-the-airbnb-community-in-nyc/>

company has no commitment to transparency. An “open and transparent company” does not sweep part of its business under the carpet before letting outsiders look around; a company that truly co-operates with cities would be open about the changes on its platform.

Key Facts

Background: city officials in New York and elsewhere are particularly concerned with commercial hosts renting out multiple Entire Home listings on Airbnb.

- **Airbnb purged over 1,000 “Entire Home” listings from its site just days before it prepared a data snapshot of its business.**
- **Airbnb used the data snapshot to paint a misleading picture of its business:**
 - **Airbnb’s message was that only 10% of Entire Homes listings belonged to hosts with multiple listings. The true number had been close to 19% for all of 2015.**
 - **Airbnb’s message was that “95% of our entire home hosts share only one listing”. The claim was true for less than two weeks of the year.**
 - **Airbnb’s rosy projections about the future of its business were not objective analyses based on historical trends. The company extrapolated from an artificial and unrepresentative one-time event.**
- **Airbnb’s one-time purge was a PR effort, and does not indicate a change of heart for the company:**
 - **No similar event took place in other cities in North America or elsewhere.**
 - **Contrary to Airbnb projections, levels of multiple-listing entire homes have already jumped back to 13% of the total, only two months after the purge.**
 - **Despite claiming that it wants to “work with cities”, Airbnb carried out its purge without disclosure or consultation.**
 - **Airbnb did not kick illegal hosts off the site; many commercial hosts still have listings on the site, but the purge made them appear, briefly, to be single-listing hosts.**

Background

In December 2015, Airbnb made two data sets available for inspection in New York City. One data set included all those listings in New York City that had a guest during the period November 1, 2014 – November 1, 2015; the other was a snapshot of all active listings in the city as of November 17, 2015. The event was part of Airbnb's Community Compact and the data sets were made available "so that policymakers can make informed decisions about home sharing in their communities."³

Airbnb did not release the data sets to external parties: instead it presented a summary on its web site and allowed individuals to inspect the data under supervision at an office in New York.

Airbnb's current market valuation is set at \$25 billion: fortunes will be made or lost based on the company's ability to make its business acceptable to the cities and countries in which it operates. The company's lobbying and PR efforts in cities across North America and around the world talk repeatedly of regular people occasionally sharing their own home (hence hosts with just one listing, or "single-lister hosts").

Airbnb's New York City 2014-2015 data set, which runs to November 1, shows that multiple-lister hosts earned 41% of the complete Entire Home revenue during the 2014-2015 year. This number is in line with what critics have been claiming for the last two years: a disproportionate amount of Airbnb's revenue is gained from commercial operators offering listings that are more likely to disrupt neighbors and displace long-term residents, and many of which are illegal under New York State law (for rentals under 30 days where the primary resident is not present). It is not a number that suits Airbnb in its battle for credibility.

Airbnb focused attention instead on projections for 2015-2016, based on its November 17 data snapshot, that the share of income going to multiple-lister hosts will shrink by two thirds, to a mere 14% of the complete Entire Home revenue. The projection is a claim that their business is becoming more, rather than less, about personal and informal "sharing".

³ Airbnb, *Data on the Airbnb Community in NYC*, Dec 1, 2015
<https://www.airbnbaction.com/data-on-the-airbnb-community-in-nyc/>

Airbnb emphasized to reporters that its projection reflects a continuing and underlying trend in their business. A spokesperson told Sarah Kessler of Fast Company that “the number of listings posted by hosts with three or more entire homes on the site ‘continues to decrease as we go forward’”.⁴ Airbnb General Manager for NYC Wrede Petersmeyer told Quartz’s Alison Griswold: “the first big step in making this compact real is to show to New York—the regulators, the public, the press—who our community actually is in the city, how they’re operating, and to dispel a lot of the myths that have been swirling around over the last few years”.⁵ He emphasized that the projection, and hence implicitly the November 17 snapshot, reflected “historical trends”.

Data sets

This report is based on two data sets of Airbnb listings in New York City, assembled separately by the authors.

- Murray Cox has collected data most months from early 2015 as part of his *Inside Airbnb* initiative. This data set is labeled MC.
- Tom Slee has collected data going back to late 2013. This data set is labeled TS.

Both data sets were assembled by programmatically compiling public information from Airbnb’s website, but they were implemented and collected independently. Both data sets attempt to locate all the listings within a city, and then visit the page for each listing to collect listing data, including the host ID. The host ID allows an analysis of the number of listings posted by a single host.

The data sets are downloadable from the authors’ websites:

- MC data set: <http://insideairbnb.com/how-airbnb-hid-the-facts-in-nyc>
- TS data set: <http://tomslee.net/how-airbnb-hid-the-facts-in-nyc>

A comparison with the released Airbnb data set shows that both data sets reflect the overall number of listings and their breakdown in the Airbnb market in New York City with high accuracy; this validation is presented in the Appendix.

⁴ Sarah Kessler, *Just how many Airbnb hotels are there in NYC*, Fast Company, Dec 6, 2015. <http://www.fastcompany.com/3054286/just-how-many-airbnb-hotels-are-there-in-nyc>

⁵ Alison Griswold, *Airbnb has a really convoluted process for being “transparent” with New York City*, December 15, 2015, Quartz. <http://qz.com/571165/airbnb-is-sharing-more-data-but-in-a-really-convoluted-way/>

In the following sections, graphs from both data sets are presented side by side. The MC data set was chosen for the opening statistics because there were additional snapshots collected covering the period of the Airbnb data release, but the similarity of the independently-collected TS data set reinforces the conclusions.

Airbnb removed listings immediately before November 17

Airbnb listings are of three types: “Entire home/apt” make up approximately 57% of New York City listings, “Private room” listings make up 40%, and “Shared room” listings are 3%. Airbnb restricted its claims to the “Entire home/apt” listings and so we do the same.

In Airbnb’s own presentation the very first “highlight from the data” is that “95% of our entire home hosts share only one listing”.³ Figure 1 shows that such a statement was true for only two weeks out of the last year, immediately after the company purged its listings. The percentage of multiple-listers has been climbing slowly since mid 2014, and is now climbing again. Airbnb misrepresents the state of its business.

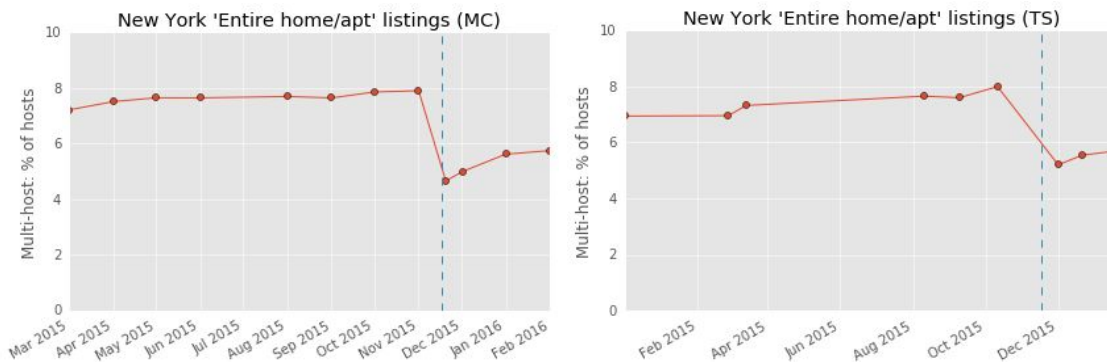


Figure 1: Percentage of Entire Home hosts sharing more than one listing. The blue line represents November 17, 2015, the day of the snapshot.

Airbnb told journalist Alison Griswold that “The company’s analysts have examined historical trends to project that, by this time next year, entire-space hosts with one or two listings will be earning 93% of revenue.” Figure 2 shows that such a projection cannot be reasonably drawn from historical trends in the data. The only fall in listings is just before the November 17 snapshot, when the percentage of multi-host listings plummets from almost 19% of the total to 10%. The conclusion is inescapable: Airbnb removed listings in

New York City belonging to multi-hosts so that the November 17 data set would fit the message the company wanted to deliver.

As an aside: Figure 2 suggests that the percentage of multiple-lister listings is again climbing after November 17. If sustained, the climb will emphasize the artificiality of Airbnb's November 17 snapshot.



Figure 2: Percentage of Entire Home Airbnb listings in New York City belonging to hosts with multiple listings. The blue line represents November 17, 2015.

Figure 3 shows that Airbnb's purge is also reflected in the relative number of visits made to single-lister and multi-lister hosts, as estimated by the number of reviews left on the site. The percentage of traffic going to multiple-listers has been a constant 30% of the total during all 2015 until the days before the November purge, when it dropped to only 20%.



Figure 3: Percentage of Entire Home Airbnb visits in New York City to hosts with multiple listings, estimated based on the number of reviews added to the site.

Airbnb's purge was PR, not a new direction

Might it be true that Airbnb has made a new resolution in November 2015? The purge is a first for the company, but may it reflect a new commitment from Airbnb to remove multi-listing hosts from its platform? Three pieces of evidence show that Airbnb's purge was not part of any such change of heart.

Private Rooms in NYC

Figure 4 shows the same graph as Figure 2, but for Private Room listings rather than for Entire Homes. For fair visual comparison, the graph has the same vertical scale as for Figure 2, covering a range of 20%. The graph shows that there is little discontinuity in the Private Room listings: the percentage belonging to multi-listers has been staying constant or even slowly increasing over time. Airbnb did not highlight, and so did not target, hosts renting out multiple Private Rooms.



Figure 4: The percent of Private Room listings belonging to multiple-lister hosts.

Entire Home listings in Los Angeles and Miami

Figure 5 shows the Entire home/apt listing market in Airbnb's second biggest North American market of Los Angeles and in another major market, Miami. Figure 5 shows that multi-hosts in Los Angeles have a share of the market that is twice as large as in New York City (until November 2015) and that in Miami they make up the majority of listings. The figure also shows that in Los Angeles their share has been slowly increasing, while in

Miami it has been steady. In neither case is there a discontinuity in November 2015 or at any other time.



Figure 5: Entire home/apt listings in Los Angeles. These graphs both use the TS data set.

Entire Home listings in European cities

Much of Airbnb’s business is outside North America. Figure 6 shows the multi-host component of Entire home/apt listings in Rome, London, and in Berlin.

Airbnb’s business varies in cities around the world. All three are large markets (Berlin has 15,000 listings; Rome 18,000; London 34,000), and the characteristics vary. What all three have in common is that the composition does not show anything like the sudden change that is seen in New York City.: the proportion of the business belonging to multi-hosts varies by no more than 3% over the time of the surveys.

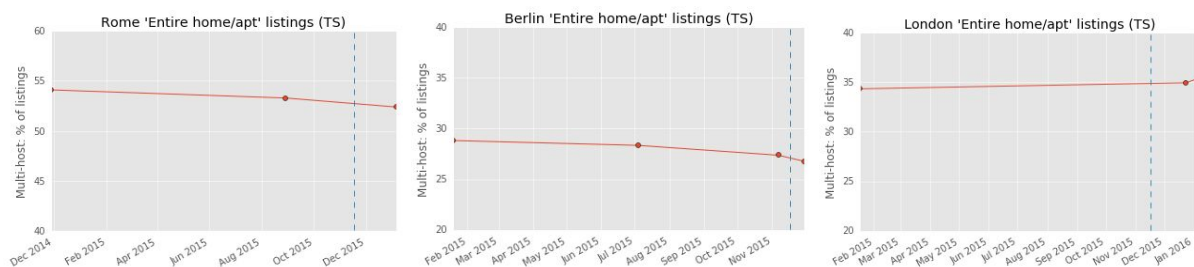


Figure 6: The multi-host percentage of Entire home/apt listings in Rome, Berlin, and London. These graphs employ the TS data set.

Taken together, these observations show that Airbnb’s intervention in its user base in New York City does not reflect an across-the-board tightening of Airbnb’s policies. It is a one-time, targeted intervention that is intended to shape the New York City data release.

In its public statements, Airbnb pledges itself to co-operation with cities:

We will also continue to work with cities that have established home sharing policies on these matters. Our community has demonstrated that it can self-regulate. And per our commitment to being open and transparent, the information released will provide a way to show that our community is doing right by the cities where they are sharing their homes.⁶

The removal of listings was not done in consultation with the city, and was not communicated to the city. Such a unilateral action goes against the commitment to openness and transparency spelled out in the Airbnb Community Compact.

Listings that vanished in November 2015

It is difficult to distinguish listings that were removed by Airbnb from listings that left the platform by choice, because there is a continual flux of Airbnb listings onto and off the platform. But by comparing the November 1-20 period to the previous two months, the scale of the intervention is made clear.

How many listings were removed?

Table 1 shows the Entire home/apt listings that vanished from the Airbnb site between November 1 and 20, compared to those that left the previous two months. More than 1,000 listings belonging to multiple-lister hosts were removed from the platform by Airbnb in the November 1 to November 20 time period, an action affecting over 500 hosts.

	Sept 1 - Oct 1	Oct 1 - Nov 1	Nov 1 - Nov 20 (actual)
Listings that left the Airbnb platform	2,274	1,934	2,627
Listings belonging to multiple-lister hosts that left the Airbnb platform	321	261	1,438
Hosts affected	206	176	665

Table 1: Entire home/apt listings in New York City that left the Airbnb platform in three time periods of 2015 (MC data set).

⁶ The Airbnb Community Compact, November 11, 2015. <http://publicpolicy.airbnb.com/compact/>

Where were the listings that were removed?

Figure 7 shows the location of the Entire Home listings that were removed from Hosts with multiple Entire Home listings between November 1 and November 20, 2015.

The purge does not appear to be concentrated in any particular neighborhoods; however the majority of removed listings were in Manhattan.

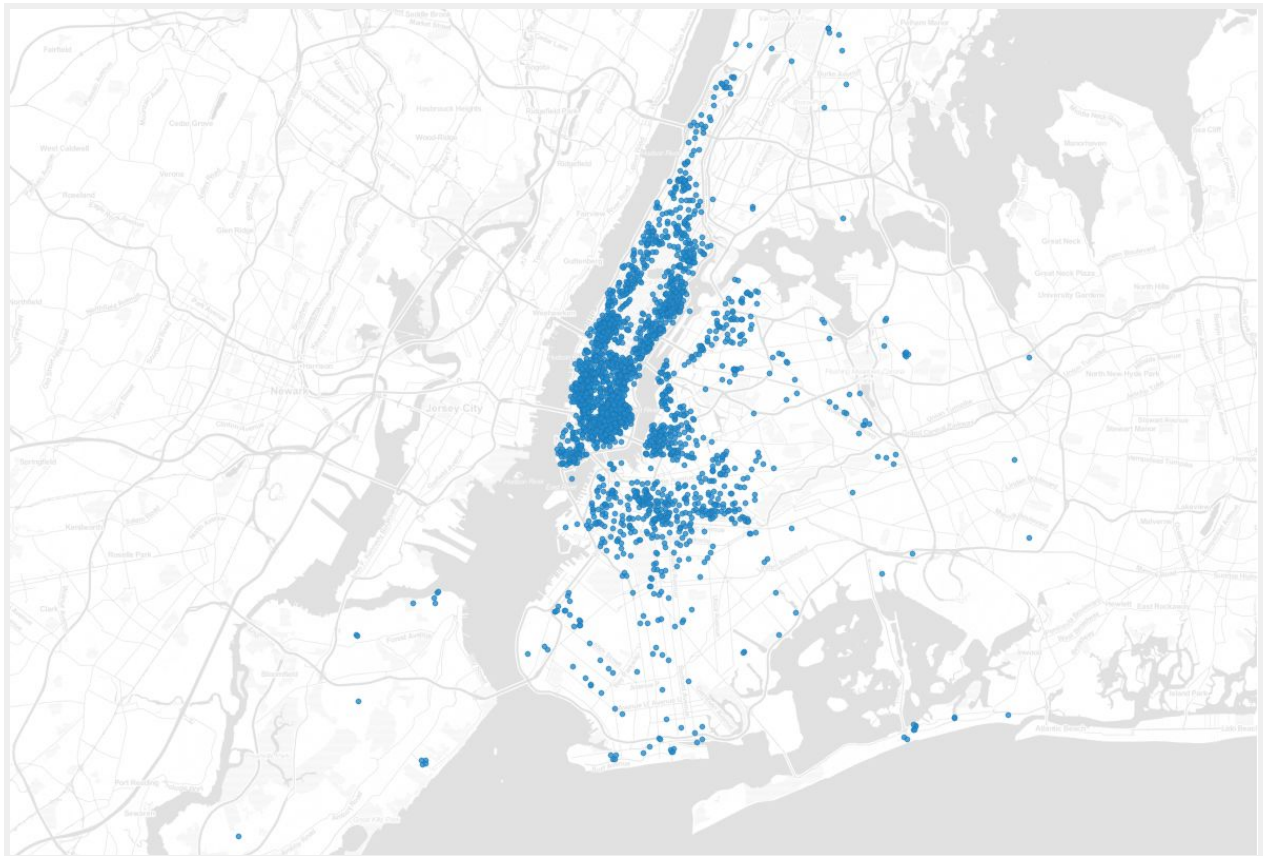


Figure 7: Location of Entire home/apt multiple-listings that left the Airbnb platform between November 1 and November 20, 2015 (MC).

How many listings were removed for each host?

Table 2 shows that, on average, the hosts affected by the purge lost 71.4% of their Entire Home inventory.

On average, most hosts were left with 0.8 Entire Home listings, although the three hosts with the most Entire home listings (with 10, 11 and 12 Entire Homes at November 1, 2015) lost all of their Entire home listings by November 20, 2015.

Most hosts affected by the purge were left with some “inventory” on the Airbnb site, indicating that Airbnb did not kick the “worst actors” off the platform. Airbnb might forego some temporary revenue from the listings it removed, but it would appear that they still like to do business with these clearly identified commercial operators.

Number of Entire Homes per host at Nov 1	Number of Hosts	Number of Listings removed between Nov 1 and Nov 20	Average % of Entire Homes removed between Nov 1 and Nov 20	Average Number of Entire Homes left at Nov 20
2	348	463	66.5%	0.7
3	162	347	71.4%	0.9
4	86	278	80.8%	0.8
5	37	155	83.8%	0.8
6	13	69	88.5%	0.7
7	9	50	79.4%	1.4
8	7	43	76.8%	1.9
10	1	10	100.0%	0.0
11	1	11	100.0%	0.0
12	1	12	100.0%	0.0
Total	665	1,438	71.4%	0.8

Table 2: A breakdown of the listings that left the site during Nov 1 to 20, belonging to “multi-lister” hosts. (MC).

Implications

Airbnb’s manipulation of its user base in preparation for the New York City data release has several implications:

- Airbnb’s data releases are part of a deliberate PR and lobbying strategy, intended to influence local governments, the media, and the broader public debate. Data

has an appearance of objectivity, but in Airbnb's case it should not be seen as an objective description of the company's business.

- The "Airbnb Community Compact", in which the company pledges itself to be more open and transparent, is also a PR and lobbying strategy. If Airbnb was truly transparent, it would not seek to manipulate its data before letting people see it. If Airbnb was truly interested in "working with cities" it would inform cities of its actions rather than hiding them.
- The Airbnb platform makes it easy for hosts to break the law. Airbnb could choose to institute relatively simple in-platform compliance to city laws, but instead makes money from hosts who break the law, only expel them when their presence is a nuisance to the company.

Attempts such as this report, to present a fair picture of the company using public data, are inevitably partial. As a privately-owned company, there is currently no mechanism for holding Airbnb accountable for its own actions and the public's ability to see the truth behind Airbnb's selected data releases is limited. Airbnb's so-called transparency initiatives are no substitute for genuine audits or for genuine accountability.

Appendix

For those who wish to verify the claims for themselves, the data sets are available for download from <http://insideairbnb.com/how-airbnb-hid-the-facts-in-nyc> and <http://tomslee.net/how-airbnb-hid-the-facts-in-nyc>. This section summarizes the reasons to believe that the data sets accurately represent the Airbnb listings in New York City.

Methodology

The first stage in data collection is to visit the Airbnb website search pages to locate as many listings as possible in New York City. Any missing listings were not visible despite many different searches, and so are very unlikely to be active listings.

Each listing has its own page on the Airbnb website. As the second part of the data collection for each data set, the programs visit each page to collect the details about that listing. Therefore, every listing described in the data sets was visible on the Airbnb website at the time the survey is taken.

Each listing page contains a variety of information, of which the following are most important for this report.

- **host_id:** Most importantly for this report, the host is unambiguously identified on each page by a unique identifier.
- **room_type:** The listing unambiguously identifies each listing as an “Entire home/apt”, “Private room”, or “Shared room”.
- **location:** Each listing has a latitude and longitude on its listing page. These were automatically compared to GIS maps of New York City to ensure that all the listings are inside the city boundaries. Some listing pages do have imprecise locations, but the accuracy for aggregate purposes, as in this report, is high.
- **reviews:** In the absences of internal Airbnb data, the number of reviews provides an indication of the number of visits to a listing. Again, this indicator may be imprecise for any one individual listing, but in aggregate it is widely used as a method for estimating the relative numbers of visits for populations of listings.

Data from both Murray Cox and Tom Slee have been used regularly over the last year by city analysts, journalists, academics and hospitality analysts.

Completeness

Airbnb claimed that "as of November 17, 2015 there were 35,966 active Airbnb listings in New York."³ The TS data set showed 33,918 verified listings on December 1, correct to within 6%. The MC data set of November 20 includes 34,918 verified listings, correct to within 3%. It should be noted that many Airbnb listings never get a visit from a guest: listings that are not easily visible do not play a significant part in the business. We believe we have captured at least 95% of all active listings in the city.

About the authors

Murray Cox

Murray Cox is a Brooklyn based independent digital storyteller, community activist and technologist. He founded the [Inside Airbnb](#) project, which provides data and tools to help understand Airbnb's impact on residential communities.

For further inquiries please contact: murray@murraycox.com, +1 917-592-9700

Tom Slee

Tom Slee has a PhD in theoretical chemistry, a long career in the software industry, and his 2006 book *No One Makes You Shop at Wal-Mart* has been used in university economics, philosophy and sociology courses. His new book, *What's Yours is Mine: Against the Sharing Economy* ([OR Books](#)), grew out of data-driven investigations of Airbnb and Uber. He lives in Waterloo, Canada.

For further inquiries please contact: tom@tomslee.net, +1 519-574-5948

Design by John Morris.

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.