What data does and does not tell us about the sharing economy

Tom Slee

August 21, 2018

Introduction

The sharing economy has arrived with its fair share of legal and policy controversy, much centering on the two leading companies: Uber and Airbnb. A prerequisite for sensible policy and legal decisions is an agreement concerning what kind of business these companies are, but over the last several years even that seemingly-elementary question has been repeatedly disputed. In a phrasing repeated frequently by the company, Chris Lehane of Airbnb says that "Home sharing offers Canadians a way to earn a little extra income to help pay the bills". Meanwhile Thorben Weiditz of Fairbnb, a Toronto-based organization that has called for increased regulation of Airbnb's business, presents a different picture: "Most of Airbnb's short-term rental business takes place in what the company calls Toronto's 'conventional hotel market'". So which is it?

How much do hosts earn? How much of Airbnb's business comes from "commercial" hosts? These would seem to be quantitative questions that could be answered by straightforward data from the company, anonymized if necessary. This is one reason why Teresa Scassa recently argued for a public interest role in reliable and comprehensive data concerning the Sharing Economy.[1] But the data-driven part of the policy debate has been just as heated as disputes over other aspects of the sharing economy. Airbnb itself and its critics have repeatedly presented data-driven portraits of the company and its business. What is surprising is that the pictures are so different.

While each separately-rentable space is sometimes still called a "room" on Airbnb, we shall see that this is no longer appropriate, and so I use the generic term "listing" to indicate a space provided for rent.

Airbnb itself does not release listing-level data for any city. By listing-level data I mean a spreadsheet or database with a row for each of the 15,000 or so rooms, apartments, and houses being offered for rent in Toronto, containing information such as the approximate location, the number of nights it has been occupied, and an identifier for the host. Airbnb limits itself instead to aggregating the data before release, providing summary statements and tables in its reports. This pre-aggregation means that people outside Airbnb cannot use the data to ask their own questions, but are limited to Airbnb's presentation. For example, in a 2016 report on "The Airbnb Community in Ontario", Airbnb states that "88 percent of hosts have one entire home listing", but say nothing about the number of listings those other 12 percent have.

Independently-collected data is an alternative source of data, and most of this data relies on "scraping" the Airbnb web site. That is, running a program that goes repeatedly to the Airbnb public web site, runs the same kind of searches that human users run, and analyzes the web pages that the searches return to pick out the key items. I have been running this kind of data collection since 2013, and the data I have collected has been used by journalists, academics, and public interest groups investigating the Airbnb phenomenon. Others who carry out data scraping include *Inside Airbnb*, a one-person effort that provides "an independent, non-commercial set

of tools and data that allows you to explore how Airbnb is really being used in cities around the world", and AirDNA, a company that sells "Airbnb data & insights to succeed in the sharing economy" both to prospective hosts (who may wish to set prices, for example) and to academics and cities. In addition, some academics have written their own code and done their own data collection.

Airbnb is dismissive of scraped data. Here is an excerpt from an Airbnb email to the City of Toronto, obtained by Fairbnb under a freedom-of-information request. Airbnb discusses the reports it provides, and continues...

> "Scraped data" is inherently unreliable, as it relies upon extracting a limited set of data from our public-facing platform, then using this data to build models that make predictions and conclusions about the entirety of the home sharing community. These erroneous findings are a major contributor to the myths about home sharing.
>
> Although reports using "scraped data" have generated significant media attention, more and more attention is now being paid to the shortfalls and inaccuracy of the scraped data. Some of this is a result of Airbnb's continuing commitment to making data available about our host and guest community...
>
> The individuals and firms that generate scraped data also speak about its limitations. Tom Slee, in an interview with the Washington Post, made it clear that there are "obvious problems with the verifiability of the data" and that it "may be inaccurate".

In this talk I review what the data I have collected from Airbnb's site tells us, and show that the broad outlines are consistent with Airbnb's own reports. The differences are not so much in the data itself as in its presentation. Readers are invited to make up their own mind about which presentation is more relevant to public policy debates.

Total Listings

We start with an uncontroversial question: what is the total number of listings in a city or other region?[1] Unsurprisingly, the number has been increasing in many places. Figure 1 shows the number of listings in Toronto over the years 2015–2017, showing that it has increased steadily from around 5,000 in early 2015 to over 15,000 at the end of 2017. These figures match reasonably well with Airbnb's own statements.

Figure 2 shows the total number for a selection of other Canadian cities and towns. The picture is, again, one of steady growth and matches statements made by Airbnb where available.

Figure 3 shows that the data collectino is good enough to capture some exceptional events. The spike in listings in Washington DC coincides with Donald Trump's inauguration (we cannot tell if the listings were rented out, and if so whether to supporters or opponents), and the spike in Ottawa coincides with Canada's 150th birthday celebrations.

[1] The term "city" is used for convenience here to describe an area investigated by a scrape, even though some searches cover regions or whole countries.

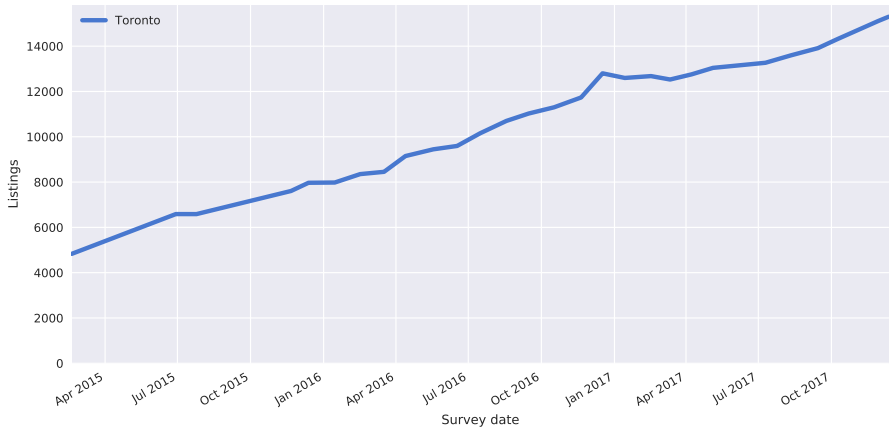

Figure 1: The number of Airbnb listings in Toronto, showing a regular and linear +CAPTION: growth rate.

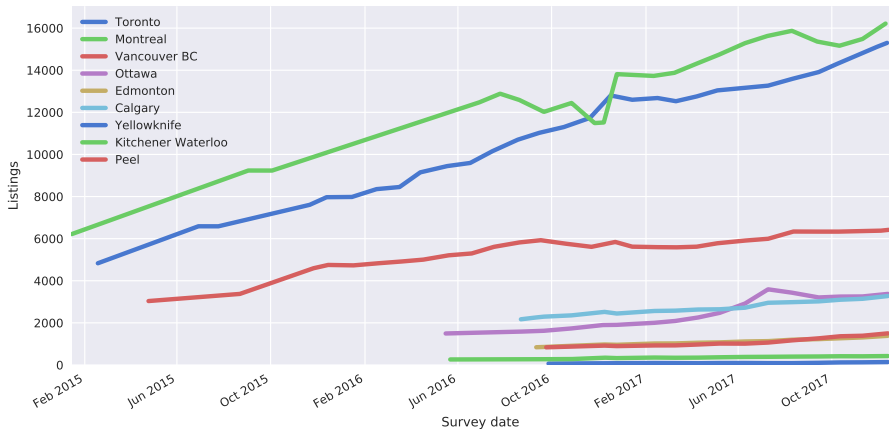

Figure 2: The number of Airbnb listings in Canadian cities, showing a regular and +CAPTION: linear growth rate in many cases. The minor unevennesses may be the +CAPTION: result of survey mistakes or variations in the data. The top two cities +CAPTION: are Montreal and Toronto, followed by Vancouver.

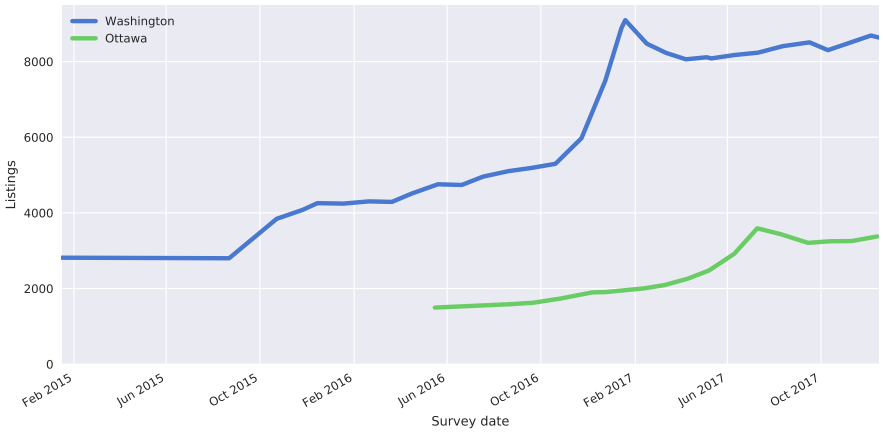

Figure 3: The number of Airbnb listings in Washington DC and Ottawa, showing spikes corresponding to Donald Trump's inauguration and to Canada's 150th birthday celebrations.

Across the world, the pattern is one of steadily increasing listing counts. Figure 4 shows data taken from several major tourist cities. The one major drop – in Berlin – is the result of a policy change where residents were not permitted to list entire homes on the site. The policy was not strongly enforced, and after a short break many listings reappeared.

Listing types

A second necessary set of data points is the relative proportions of different listing types. Listings on Airbnb can be one of three "room types", a phrase that is used in the technical codes on the site but which is now outdated.

One type is the "shared room", which corresponds to the original Airbnb story, when the founders of the company were newly graduated design students struggling to pay the rent. Seeing a conference coming to town, they rented out airbeds in their apartment as "Airbed and Breakfast". A mockup of the original apartment has been recreated in the company's headquarters.

A second listing type is the "private room", which is what many people think of when they think of Airbnb: it corresponds to renting out a spare room.

The third type is an "Entire home/apt", which means that a guest has sole use of a living space.

Figure 5 shows the number of each type in Toronto. It is immediately obvious that the original "shared room" plays almost no part in the company's business model today, comprising well under 5% of listings on the site. Despite being the way Airbnb is commonly described, "Private room" listings make up only about a third of the total, and "Entire home/apt" listings are the majority

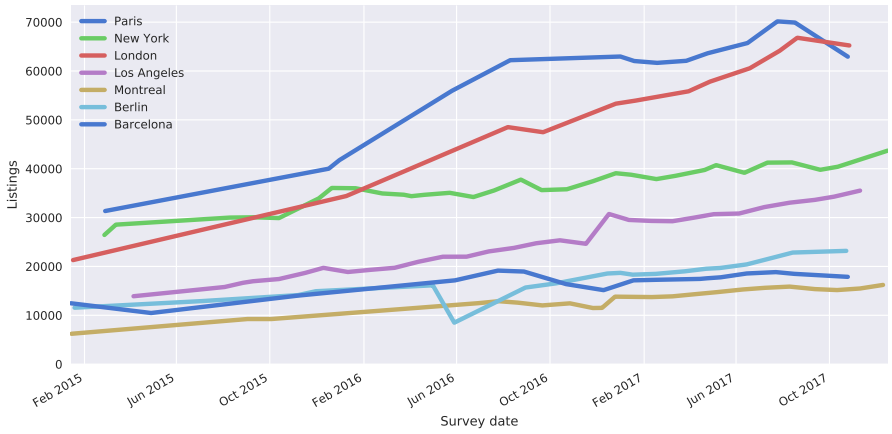

Figure 4: The number of Airbnb listings in several cities

on the site. Despite the growth in total number of listings, the proportion has stayed relatively constant over the years.

Figure 6 and Figure 7 show the same data sets for Montreal and Vancouver, respectively. Each show that the bulk of the business is Entire homes. And of course as they earn more, they are a bigger part of the business than these charts show.

Figure 8 shows the relative proportion of different listing types for a selection of Canadian cities, in late 2017. Smaller cities such as Yellowknife have a much bigger representation of private rooms, but the big tourist destinations are dominated, as we have seen, by entire home listings.

Figure 9 shows a comparison for cities around the world. The picture is varied, and the reasons for the presence of more or less of one type or another would require a knowledge of the conditions and history of each location. It is left as an exercise for the reader.

Multiple ownership

It is when we turn to multiple ownership that the controversy starts. Here is one lens where it seems like there has been a big difference between Airbnb's view of the world and that of its critics.

Airbnb statements about the incidence of multiple ownership in their major cities are consistent, emphasizing the high percentage of home-sharing hosts as opposed to more commercial hosts with multiple listings:

- "87 percent of Airbnb hosts rent out the home they live in" (New York, 2014)
- "95% of our entire home hosts share only one listing" (New York, 2016)

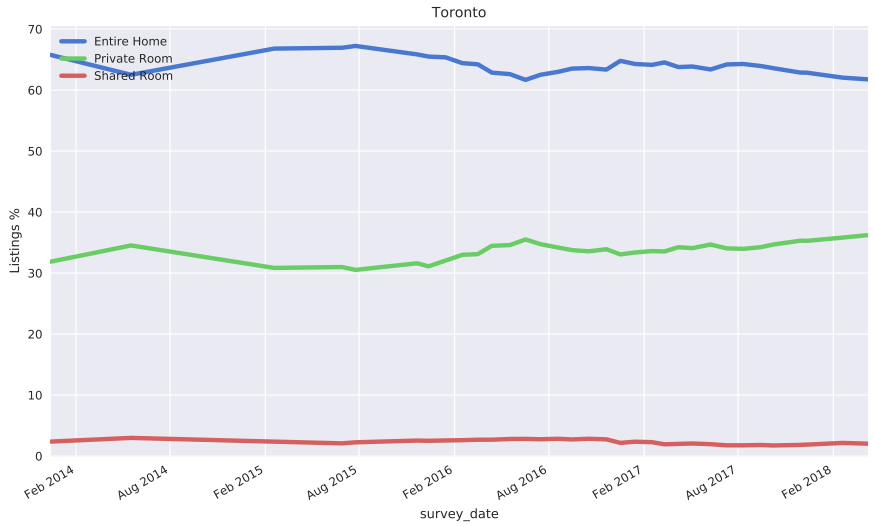

Figure 5: Listing types in Toronto

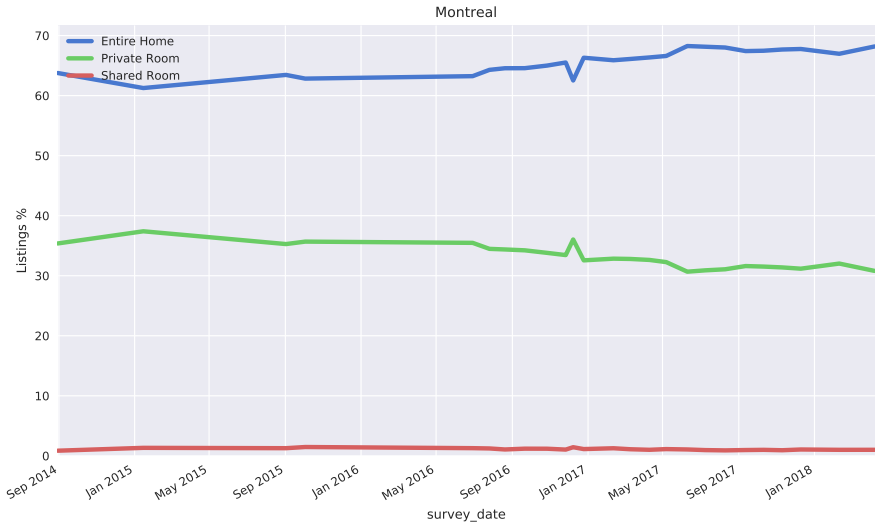

Figure 6: Listing types in Montreal

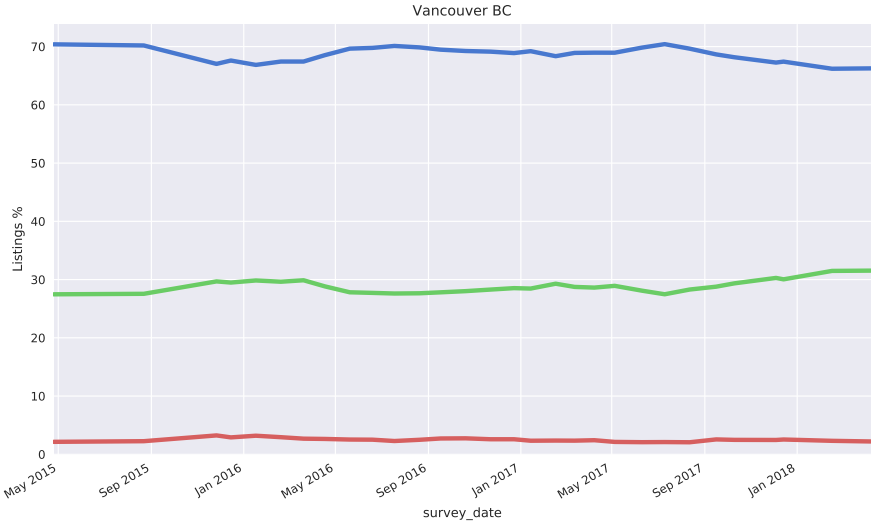

Figure 7: Listing types in Vancouver

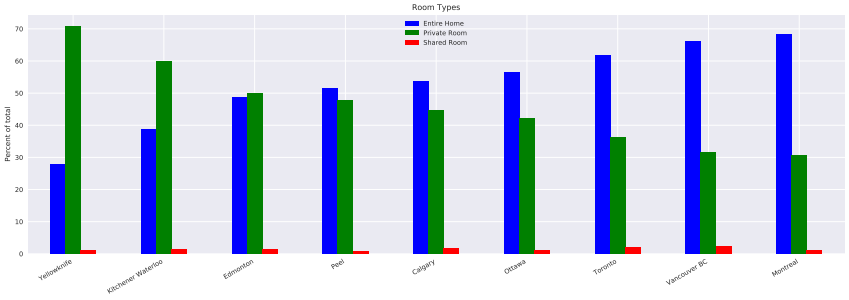

Figure 8: Listing types in a selection of Canadian cities

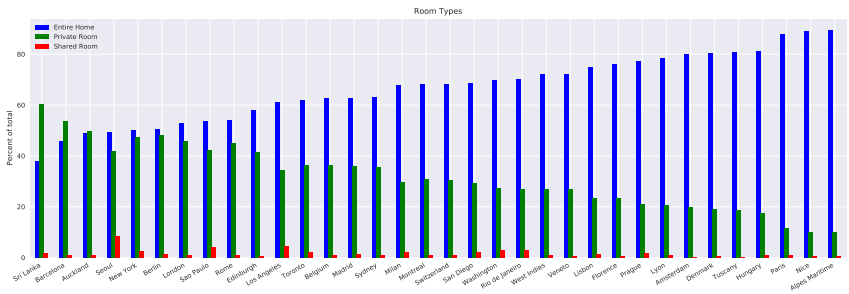

Figure 9: Listing types in cities and regions around the world.

- "87 percent of hosts rent the homes they live in" (Amsterdam, 2014)
- "88% of Airbnb hosts in Washington, D.C. are sharing space in their permanent home" (Washington 2016)
- "88%: Hosts who share their primary home" Dublin 2016
- "73 percent of hosts share just one listing" Barcelona 2016

Meanwhile, critics convey a different message: for example a 2016 report on the Airbnb market in New York City by myself and Murray Cox of Inside Airbnb claimed that "The percentage of traffic going to multiple-listers has been a constant 30% of the total during all 2015".

At first glance it seems that these two sets of figures are different, but a moment's thought shows that they are consistent. A simple example shows what is happening. Consider a small Airbnb city with four hosts and six listings:

- Aslam, Rose, and Alex each have a single listing, renting out their own homes.
- Sophie is running a small Airbnb hosting business, and has 3 listings on the site.

For this city, both the following statements are true:

1. Seventy five percent of hosts just rent out the home in which they live.
2. Commercial offerings make up half of the listings in the city.

Statement 1 is Airbnb's favoured presentation of the distribution, while critics tend to emphasize the second. For most public and community purposes the second data point – the percentage of listings, not the percentage of hosts – is the one that matters: a visitor to the city browsing the Airbnb web site will see six listings, half of which are commercial offerings, and a city government concerned about what is happening on the ground cares about the percentage of commercial listings, not the percentage of hosts.

Figure 10 shows the pattern of multiple listing ownership in Toronto, as obtained by the author. The percentage of hosts who rent out a single listing is around 12%, which is very close to Airbnb

statements. Meanwhile, the percentage of listings belonging to those hosts is more like 30%, as critics have claimed. There is no contradiction in the data after all, simply in the presentation. The top line in the figure is an estimate of revenue for single and multiple listers, based on overall review counts (see below) and shows an even higher figure, with perhaps half of Airbnb's revenue in Toronto coming from "commercial" listings.

The conclusion is perhaps surprising given Airbnb's repeated insistence on the inaccuracy of scraped data, but nevertheless it holds in other cities too. The data collected by independent "scrapes" provides a reasonably accurate description of the Airbnb listings in a city, at least as far as ownership goes.

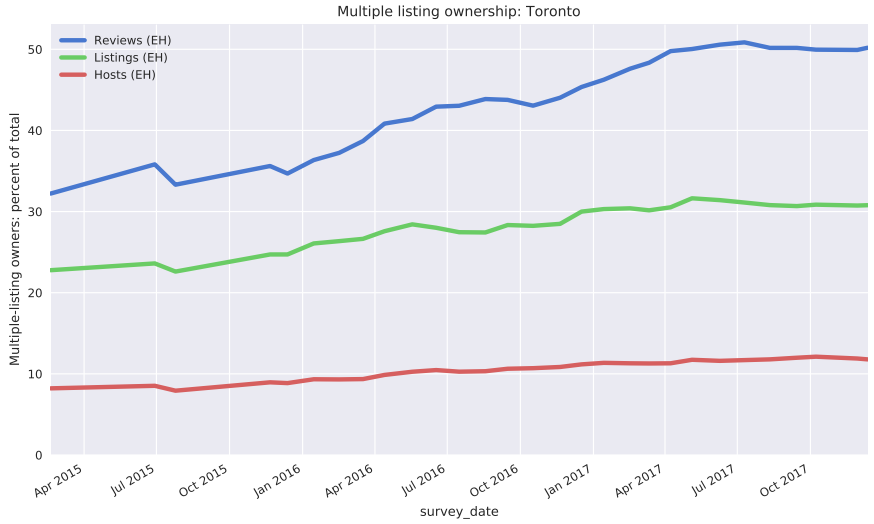

Figure 10: Multiple listing ownership in Toronto, as a percentage of listings. The top line is an estimated percentage of revenue, the middle line is the percentage of listings, and the bottom line is the percentage of hosts.

Figures 11 and 12 show the same data for Montreal and Vancouver, respectively. The pattern is similar, as is the agreement with Airbnb's claims. Again, the percentage of commercial hosts is relatively small, at around 10% of the total, but the percentage of listings is higher. What's more, using the total number of reviews for each listing as a proxy for the total visits to the listings, the commercial listings are busier than those owned by casual hosts. Perhaps this is not surprising, as a commercial owner has an incentive to seek out popular locations and to rent out the listing as often as possible, while a casual home-sharer can only rent out their own home, no matter whether it is in a popular location or not. As a result, it seems that multiple-listing hosts make up about half of Airbnb's total business in Canada's three largest cities.

Finally, Figure 13 shows an estimate of multiple listing presence in a number of cities and regions

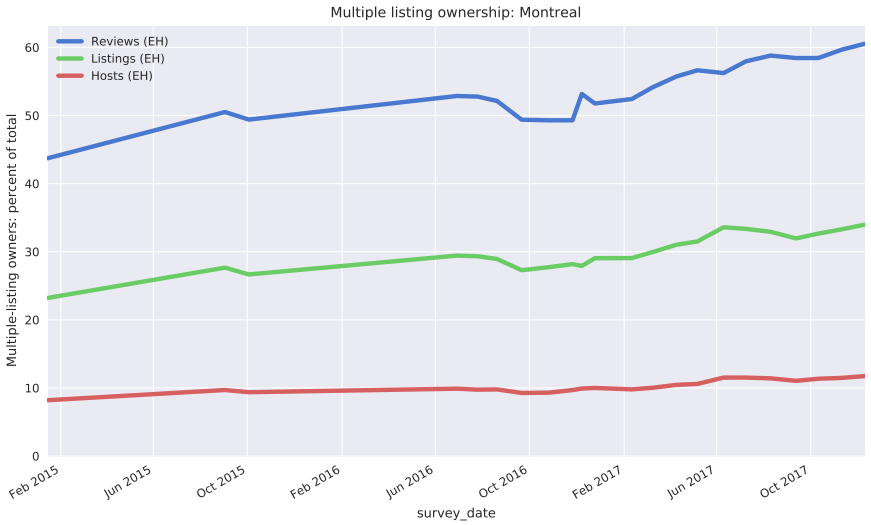

Figure 11: Multiple listing ownership in Montreal

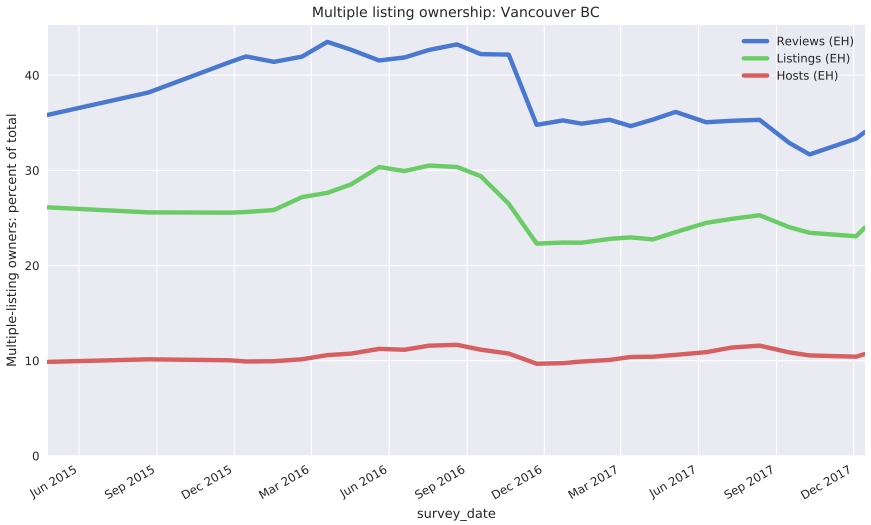

Figure 12: Multiple listing ownership in Vancouver

around the world. The diversity is striking, with some cities being much closer to a single-listing population, while others are dominated by hosts with multiple listings. Again, interpretation of these values is left as an exercise for the reader.

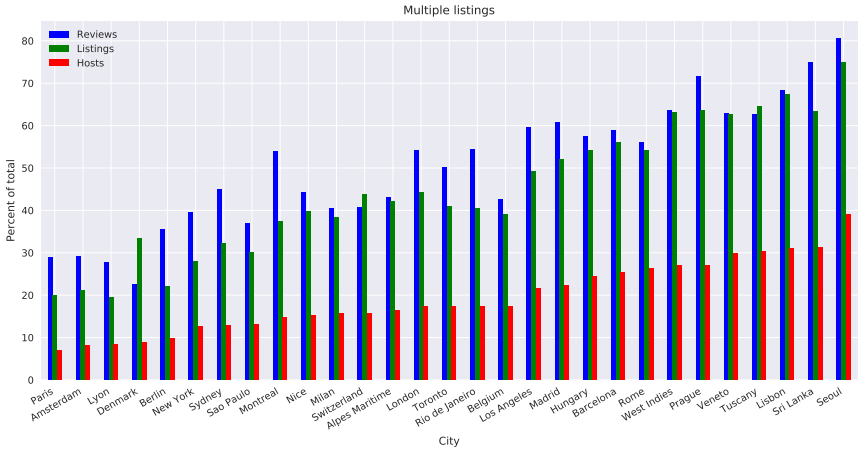

Figure 13: Multiple listing ownership in cities and regions around the world

Occupancy

While independent data on multiple listing ownership has been reliable, data on occupancy is more uncertain. There is no public way of seeing whether a listing is rented out for a particular night. Even if a listing is marked as unavailable, it may be simply that the host is not making it available to the public on that day. While AirDNA bases its estimates on calendar entries, I have used an even more approximate method of using review counts as a proxy for relative number of visitors.

Over large numbers of listings, this should give an idea of the broad trends. For example, Figures 14, 15, 16, and **??** show that using reviews as a proxy for visits captures the expected seasonal fluctuations in travel.

Again, the picture of occupancy painted by Airbnb and by critics differs. Airbnb emphasizes that listings are occupied only occasionally, with typical hosts renting out for maybe one night in ten:

- Ontario: "typical hosts are earning $3,900 annually from renting out their own primary residence for three to four nights a month" (Alex Dagg, 2016)
- New York: "the typical host earns $7,530 per year" (Airbnb Economic Impact)
- New Zealand: "the typical Airbnb host earns NZD$80 per week or NZD$4,200 a year" (Airbnb, 2018)

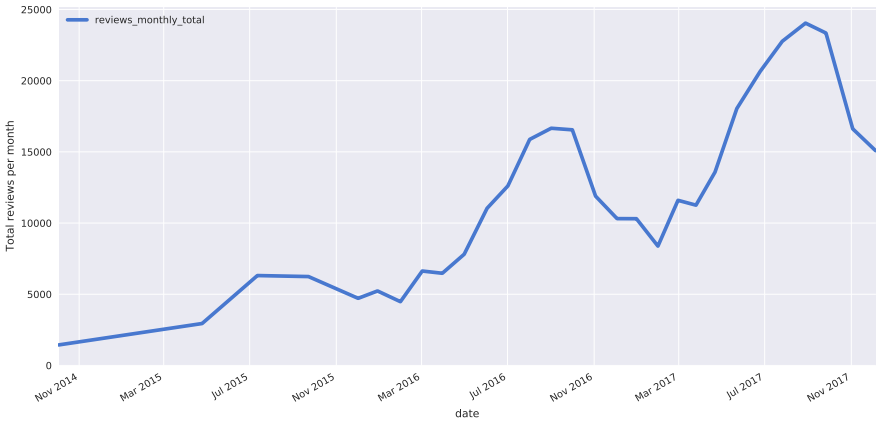

Figure 14: Visits to Toronto, using reviews as a proxy

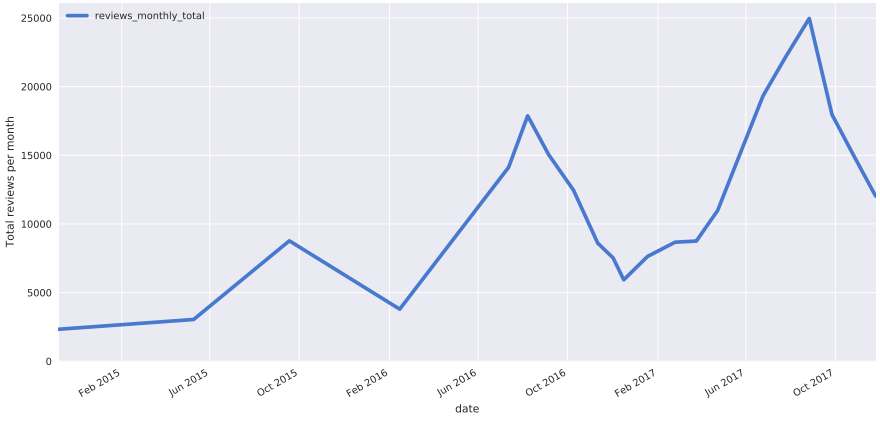

Figure 15: Visits to Montreal, using reviews as a proxy

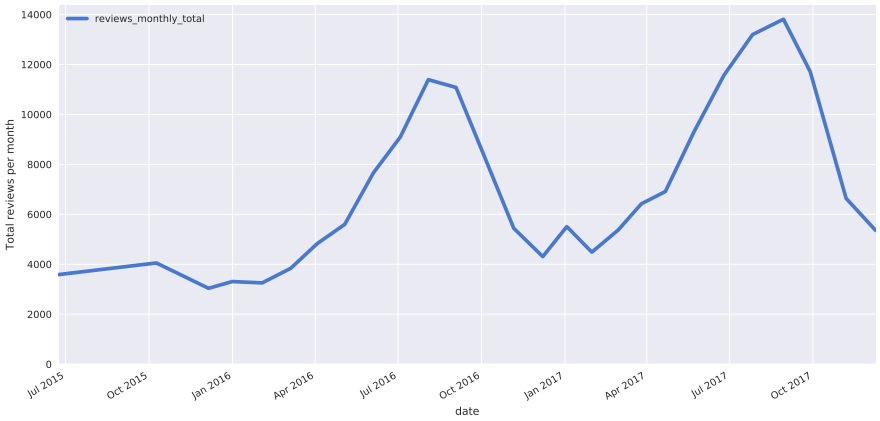

Figure 16: Visits to Vancouver, using reviews as a proxy

- Most of these listings are only shared occasionally – nearly 60% of "entire home" listings are rented less than 90 days per year. (Washington DC)

A small city example is useful again to make sense of how these numbers are presented. Consider the case where our four hosts each have a single listing, but while three (Aslam, Rose, Alex) are renting only occasionally, Sophie is renting more intensively.

- Aslam's listing is occupied for 0 nights per year
- Rose's listing is occupied for 30 nights
- Alex's listing is occupied for 42 nights
- Sophie's listing is occupied for 180 nights

The following statements are both true:

1. The typical host rents for only 36 nights in the year. The "typical host" is the median value, which for a population of four is midway between Rose and Alex.
2. Most of the guest traffic (over 2/3) happens at Sophie's "ghost hotel".

Again, the second figure is the one that is important for most public policy purposes and is the figure that neighbourhood residents will be concerned with. Condo owners are concerned with the number commercial visitors in their building, not with the number of listings on a web site. Aslam's listing, which is never rented, is no concern to anybody (is a host with no guests still a host?) and discussing "typical hosts" gives an unjustifiably high importance to low occupancy hosts.

In many cities, a limit of 90 days has been proposed as a reasonable upper limit for the kind of

casual business that Airbnb claims to be. This time we actually have some data from Airbnb, courtesy of a freedom of information request by Fairbnb. Figure 17 shows that 60% of listings rent out for less than 90 days a year, and the median value is about 60 nights, which matches Airbnb's statements about "typical" hosts.

Figure 17: Airbnb occupancy estimate.

But reworking that again: where are most nights spent? A few hosts with heavily-occupied listings can count for a large proportion of the nights spent in Airbnb listings. Figure 18 shows the data points from Figure 17 from two perspectives. On the left is an interpolation of the values given, which confirms the typical host occupancy (the red line) at 60 days. The figure on the right is the **cumulative** occupancy. It shows that well over 80% of the nights spent in Airbnb listings are spent in places rented out over 60 days a year, and about 75% of nights are spent in listings rented out over 90 days a year.

Again, the data from critics and Airbnb is broadly comparable, despite the surface differences, but the presentation by Airbnb is a misdirection.

Conclusions

My goal in this talk has been to clarify some of the disputes around Airbnb in Canada's cities. One thread has been that, despite Airbnb's grumbles, scraped data can give a reasonable picture of Airbnb's business, albeit with gaps on the topic of occupancy. The disagreements are more a matter of presentation than the underlying data quality: data can be viewed from multiple perspectives, and letting Airbnb decide which perspective to choose is not in the public interest.

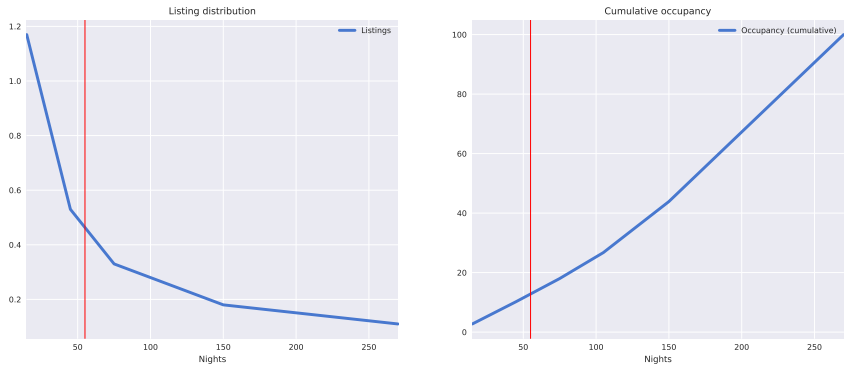

Figure 18: Computed occupancy

A second goal is to show that, while there are many hosts on the Airbnb web site who fit the archetype of a casual host, occasionally renting out the home in which they live, the company's business is driven significantly by "commercial" hosts with multiple listings and by high-occupancy listings that are rented frequently. The picture of "ghost hotels" has a good deal of truth to it, and lawmakers and policy discussions should take this into account in their debates.

References

[1] Teresa Scassa. Sharing data in the platform economy: a public interest argument for access to platform data, 50-4, 2017, 1017–1071. *UBC Law Review*, 50(4):1017–1071, 2017.

